ALEA, Lat. Am. J. Probab. Math. Stat. 21, 1915–1948 (2024)
DOI: 10.30757/ALEA.v21-73

Bounds on Mixing Time for Time-Inhomogeneous Markov Chains

Raphael Erb

Department of mathematics and computer science, University of Basel, Spiegelgasse 1, 4051 Basel, Switzerland
E-mail address: raphael.erb@unibas.ch

Abstract. Mixing of finite time-homogeneous Markov chains is well understood nowadays, with a rich set of techniques to estimate their mixing time. In this paper, we study the mixing time of random walks in dynamic random environments. To that end, we propose a concept of mixing time for time-inhomogeneous Markov chains. We then develop techniques to estimate this mixing time by extending the evolving set method of Morris and Peres (2005). We apply these techniques to study a random walk on a dynamic Erdős-Rényi graph, proving that our proposed definition of mixing time is $O(\log(n))$ when the graph is well above the connectivity threshold. We also give an almost matching lower bound.

1. Introduction

In recent years, the topic of mixing time of random walks in dynamic random environments has received considerable attention (we refer to Avena et al. (2018, 2019); Avin et al. (2018); Cai et al. (2020); Hermon and Sousi (2020); Lelli and Stauffer (2024); Mans and Pourmiri (2022); Moumeni (2024); Sousi and Thomas (2020); Peres et al. (2015, 2018, 2020); Saloff-Coste and Zúñiga (2009); Sauerwald and Zanetti (2019) for a selection of recent results). One implicit technical challenge these works have faced is that random walks in dynamic environments, in general, do not have a stationary distribution, which is required to define the mixing time. A common way around this issue, used in some of the above references (Avena et al., 2018, 2019; Avin et al., 2018; Hermon and Sousi, 2020; Lelli and Stauffer, 2024; Sousi and Thomas, 2020; Peres et al., 2015, 2018, 2020; Sauerwald and Zanetti, 2019), is to consider models where, despite the dynamic environment, the stationary distribution exists and is independent of time.

The goal of this paper is to consider situations where such time-independent stationary distributions do not exist. We will propose a candidate for a "time-dependent" stationary distribution and a corresponding definition of mixing time. To show the workability of this approach, we develop

Received by the editors December 5th, 2023; accepted October 18th, 2024.

2010 *Mathematics Subject Classification.* 60J05, 05C81.

Key words and phrases. probability theory, time-inhomogeneous markov chains, random walks, dynamic random graphs.

The author was supported by Swiss National Science Foundation through SNF project no. 200021E_193063 in the context of a joint DFG-SNF project within the DFG priority programme SPP 2265 Random Geometric Systems.

a method to find upper bounds on this new mixing time, based on the theory of evolving sets. This method was originally developed to bound mixing time of time-homogeneous Markov chains in Morris and Peres (2005). It has also been applied to bound mixing time for time-inhomogeneous Markov chains, such as random walks on dynamical percolation (see Peres et al., 2018, 2020), albeit under the assumption that a time-independent stationary distribution exists. The evolving set method has also been used to study the transience of dynamic random graphs (Dembo et al., 2017), we will briefly comment on this below.

To describe the main idea of our work, let us recall the usual definition of mixing time. For a Markov chain X on a finite state space with (time-independent) transition matrix P and unique invariant distribution π, the ε-mixing time, $\varepsilon \in (0,1)$, is defined by

$$t_{\text{mix}}^{\text{static}}(\varepsilon) = \inf\{t \geq 0 : \sup_x \|P^t(x,\cdot) - \pi(\cdot)\|_{\text{TV}} \leq \varepsilon\} \tag{1.1}$$

where $\|\mu - \nu\|_{\text{TV}} = \frac{1}{2}\sum_x |\mu(x) - \nu(x)|$ is the total variation distance and P^t denotes the t-th power of P. That is to say the mixing time is the first time where the distribution of X_t is ε-close to π, uniformly in the starting position. Furthermore (see e.g. Levin and Peres, 2017, Chapter 4), this mixing time is related to another quantity

$$\tilde{t}_{\text{mix}}^{\text{static}}(\varepsilon) \coloneqq \inf\{t \geq 0 : \sup_{x,y} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{\text{TV}} \leq \varepsilon\} \tag{1.2}$$

by the inequality

$$t_{\text{mix}}^{\text{static}}(2\varepsilon) \leq \tilde{t}_{\text{mix}}^{\text{static}}(2\varepsilon) \leq t_{\text{mix}}^{\text{static}}(\varepsilon), \qquad \text{for every } \varepsilon \leq \frac{1}{2}. \tag{1.3}$$

Let us now consider a discrete time-inhomogeneous Markov chain $(X_t)_{t\in\mathbb{Z}}$ on a finite state space whose distribution $\mathbb{P}^P$ is determined by a sequence $P = (P_t)_{t\in\mathbb{Z}}$ of transition matrices via

$$\mathbb{P}^P(X_t = y | X_{t-1} = x) = P_t(x,y). \tag{1.4}$$

For $s < t$, we will write $P^{s,t} = P_{s+1}\cdots P_t$ for the matrix product.

In order to define mixing time for $(X_t)_{t\in\mathbb{Z}}$, we propose to replace π in (1.1) with a time-dependent target distribution π_t, defined by

$$\pi_t(\cdot) \coloneqq \lim_{s\to-\infty} P^{s,t}(x,\cdot) = \lim_{s\to-\infty} \mathbb{P}^P(X_t = \cdot | X_s = x), \tag{1.5}$$

if the limit exists and is independent of x. We will see later that this is always the case under some mild irreducibility assumptions. We refer to Definition 2.1 and Proposition 2.2 for details. For $\varepsilon \in (0,1)$, we then define the ε-mixing time of the sequence $P = (P_t)_{t\in\mathbb{Z}}$ at time $s \in \mathbb{Z}$, by

$$t_{\text{mix}}^P(\varepsilon, s) \coloneqq \inf\{t \geq 0 : \sup_x \|P^{s,s+t}(x,\cdot) - \pi_{s+t}(\cdot)\|_{\text{TV}} \leq \varepsilon\}, \tag{1.6}$$

see Definition 2.6 below. It is obvious that if P_t does not depend on t, this definition coincides with (1.1). In general, (1.6) has similar properties as (1.1), e.g. an analogous statement to (1.3) holds (see Lemma 2.7).

1.1. *Main Results.* We will now describe the results of this paper. Some of them are technical and require more notation. Therefore, in this introduction we only give informal versions of those results, for precise statements we refer the reader to the later sections.

The key observation that motivated this work was that the definitions (1.5) and (1.6) lend themselves well to techniques more commonly used for time-homogeneous Markov chains, in particular the evolving set method. This method yields the first principal result of this paper, Theorem 3.2, an upper bound on the mixing time given in (1.6). Similarly as in the static case, the upper bound involves certain "isoperimetric" quantities related to the chain, namely the *bottleneck ratio* Φ_t^* (see (3.10)) which in our case is time-dependent. In the following, let

$$\pi_t^{\min} = \min_x \pi_t(x) \qquad \text{and} \qquad g_t = \min_x \frac{\pi_{t-1}(x)}{\pi_t(x)}. \tag{1.7}$$

Theorem 1.1 (restated Theorem 3.2)**.** *Fix $\varepsilon \in (0,1)$. If for some $t > 0$ it holds that*

$$\frac{1}{\sqrt{\pi_t^{\min}\pi_0^{\min}}} \prod_{s=1}^{t} \left[1 - \frac{1}{2}\left(\frac{1}{2}\frac{g_s}{1-\frac{1}{2}g_s}\Phi_s^*\right)^2\right] \leq 2\varepsilon, \tag{1.8}$$

then also

$$\sup_x \|P^{0,t}(x,\cdot) - \pi_t(\cdot)\|_{\mathrm{TV}} \leq \varepsilon.$$

Note that for *time-homogeneous* chains the above simplifies to a result originally obtained in Jerrum and Sinclair (1989) - although their proof explicitly assumes reversibility of the chain. The homogeneous result has been slightly improved since, removing the reversibility assumption, and the evolving set method offers an alternative proof (see Levin and Peres (2017, Theorem 17.10) for a full account). In the present paper, we show that time-homogeneity is not a necessary assumption for those techniques to be applicable if one defines π_t and mixing time as in (1.5) and (1.6) above.

Of particular interest to us are models where the transition matrices of the chain are not deterministic, but sampled randomly, either independently, or by a Markov chain on the space of transition matrices. Such models fall into the category of random walks in random dynamic environments. In Section 4, we apply our main result, Theorem 1.1, to prove an upper bound on the mixing time in this situation. We would like to highlight that these estimates are quenched, and we show that they hold on a set of sequences of transition matrices whose probability tends to one as the size of the state space grows (we refer to this as *with high probability*, or *w.h.p.*, see Section 4). In particular, in Corollary 4.3, we provide a set of assumptions under which the (now random) mixing time $t_{\mathrm{mix}}(\varepsilon, 0)$ grows at most logarithmically with the size of the space.

Theorem 1.2 (informal version of Corollary 4.3)**.** *Assume there exists a constant $\beta > 0$ such that $\pi_t^{\min} \geq n^{-\beta}$ for every $t \in \{0,\ldots,n\}$, w.h.p. and furthermore assume there exists a constant $c > 0$, such that $g_t\Phi_t^* \geq c$ for every $t \in \{0,\ldots,n\}$, w.h.p. Then, for every $\varepsilon \in (0,1)$,*

$$t_{\mathrm{mix}}(\varepsilon, 0) = O(\log n)$$

with high probability.

Similarly as in the static case, these assumptions require the absence of "poorly connected" regions in the space, i.e. the bottleneck ratio Φ_t^* (3.10) being large enough. In addition we require that $\pi_t^{\min}$ and g_t from (1.7) are sufficiently bounded from below for a "long enough" stretch of time.

It turns out that application of this corollary is more complex than in the static case. While the absence of bottlenecks can be shown with standard techniques, establishing the lower bound on $\pi_t^{\min}$ is not so straightforward. The definition of π_t in (1.5) as a limit makes it less accessible. To illustrate this difficulty, we study in Section 5 the example of a lazy simple random walk on a dynamic random Erdős-Rényi (ER) graph with n vertices and edge occupation probability p_n. We assume that the ER graph is independently sampled at each time step, and that $p_n = \frac{\eta \log n}{n-1}$ with $\eta > 50$, so that the graph is typically connected. We denote by $t_{\mathrm{mix}}(\varepsilon, 0)$ the random mixing time with respect to those parameters. Then the following holds:

Theorem 1.3 (informal version of Theorem 5.1)**.** *For a lazy simple random walk on ER graphs with the above parameters, for every $\varepsilon \in (0,1)$,*

$$t_{\mathrm{mix}}(\varepsilon, 0) = O(\log n) \tag{1.9}$$

with high probability.

The main and most technical step of the proof of this statement is the establishment of the bound

$$\frac{\alpha_1}{n} \le \pi_t(x) \le \frac{\alpha_2}{n}, \qquad \text{for all } x \text{ and } t \in \{0, \dots, n\}, \tag{1.10}$$

for some $\alpha_1 \in (0,1), \alpha_2 > 1$, which holds with probability converging to 1 as n grows. By computing those bounds, we can then verify the assumptions of Theorem 1.2 to arrive at (1.9). We want to point out that this proof is the only place where we use the independent sampling of those Erdős-Rényi graphs, with an alternative way of obtaining control over π_t our results could be more general.

Finally, we prove that the logarithmic upper bound (1.9) is essentially optimal. With p_n as above, the following statement holds:

Theorem 1.4 (informal version of Theorem 5.13)**.** *In the same setting as Theorem 1.3, there exists $c > 0$ such that for every $\varepsilon \in (0, 1/2)$*

$$t_{\text{mix}}(\varepsilon, 0) \ge c \frac{\log n}{\log \log n}$$

with high probability.

The proof is based on a combinatorial observation about the number of reachable vertices in a dynamic graph.

1.2. *Literature.* As noted earlier, there is a wealth of research on mixing time and related quantities for time-inhomogeneous Markov chains. We will now briefly comment on just a few of those to point out the differences to the present work.

A first category of papers, in contrast to the present work, relies on the existence of a time-independent stationary distribution. For example, the authors of Peres et al. (2018) study the mixing time of random walks on dynamical percolation on the discrete torus $\mathbb{Z}_n^d$. In dynamical percolation, edges refresh independently at some rate μ and upon refreshing are declared open with some probability p or closed with probability $1 - p$. The authors then obtain quenched bounds on mixing time of the random walk. The existence of a time-independent target distribution, despite inhomogeneity, is achieved by adjusting the waiting time distribution of the random walk depending on the state of the environment at the time. An important technique used in Peres et al. (2018) is the evolving set method, adapted to time-inhomogeneous Markov chains.

The authors of Sousi and Thomas (2020) study a model of Erdős-Rényi graphs with dynamical percolation. They achieve bounds on mixing time of a random walker on ER graphs not just for graphs that are connected, but also in the critical and sub-critical regime. The existence of a time-independent stationary distribution has to be assumed for their definition of mixing time, and in their model it is in fact guaranteed to be equal to the uniform distribution by again slowing the random walk when more edges are closed, as in Peres et al. (2018). They use different techniques than the present paper, as their ER graph is not connected, instead proving and using results on the hitting and exit times of the giant component of the graph, as well as using a coupling construction of two random walkers to obtain bounds on mixing time.

Mixing time of random walks has also been studied on a dynamic configuration model in Avena et al. (2018, 2019), that is on a sequence of random graphs with given degree sequence, where the dynamics of the underlying graph are obtained by rewiring a subset of the half-edges in each time step. By construction, the stationary distribution of the random walk for this model exists and is uniform. The results of Avena et al. (2018, 2019) are observing the walker jointly with the dynamics of the environment. We will discuss the significance of the difference to the present paper in Remark 4.5 and Remark 5.3.

In contrast to this first category of papers, there are comparatively fewer works on mixing time of random walks *without* time-independent stationary distribution. Similarly as the present paper, these papers need to introduce new concepts that replace the traditional definition of mixing time.

For example, in Saloff-Coste and Zúñiga (2009), the term *merging time* has been coined and defined as follows: For a set of transition matrices $\mathcal{Q}$, we say that $\mathcal{Q}$ has ε-merging time $t_{\text{merge}}^{\mathcal{Q}}(\varepsilon) \le T$ if for every sequence $(P_t)_{t\ge 0}$ with $P_t \in \mathcal{Q}$ for all t, we have

$$\max_{x,y} \|P^{0,T}(x,\cdot) - P^{0,T}(y,\cdot)\|_{\text{TV}} \le \varepsilon, \tag{1.11}$$

where $P^{0,T} = P_1 \cdots P_T$ as defined earlier. The requirement that this inequality (cf. (1.2)) holds for any sequence of transition matrices can lead to the merging time being significantly increased by a specific outlier. By introducing a probability distribution on $\mathcal{Q}$, we can potentially eliminate those outliers and arrive at a similar setting as in Section 4 of the present paper. Merging time and our proposed mixing time are related through

$$t_{\text{merge}}^{\mathcal{Q}}(\varepsilon) \le \sup_{P\in\mathcal{Q}^{\mathbb{Z}}} t_{\text{mix}}^{P}\Big(0, \frac{\varepsilon}{2}\Big).$$

Since Theorem 3.2 treats deterministic sequences of transition matrices, and the framework in Section 4 is quenched, the present work also implies both deterministic and probabilistic upper bounds on merging time. A crucial assumption in Saloff-Coste and Zúñiga (2009) is that of *c-stability*, which is a relaxation of the condition (1.10). In a very recent article (Moumeni, 2024), c-stability has been replaced by a new set of assumptions to be able to study a different set of inhomogeneous Markov chains. Both papers use techniques entirely different from the present paper, namely functional inequalities such as Nash and logarithmic Sobolev inequalities.

A different approach at redefining mixing time is found in Cai et al. (2020). The model is that of a random walk on Erdős-Rényi graphs with dynamical percolation, but rather than slowing down the random walk on the graph to force the existence of a stationary distribution (cf. Sousi and Thomas (2020) or the description above), the authors of that paper propose the following notion of mixing time: At each time t, consider the stationary distribution $\tilde{\pi}_t$ of a time-homogeneous Markov chain with transition matrix P_t, and then observe whether the distribution of the time-inhomogeneous Markov chain stays close to those different $\tilde{\pi}_t$. A Markov chain on a graph with n vertices is said to have mixed if it did stay close to $\tilde{\pi}_t$ for at least $\sqrt{n}$ consecutive time steps where the choice of $\sqrt{n}$ is essentially an arbitrary expression for "long enough". We do not need to make such a choice, since we will see (in Lemma 2.7) that the distance $\sup_x \|P^{s,s+t}(x,\cdot) - \pi_{s+t}(\cdot)\|_{\text{TV}}$ for our proposed π_t is non-increasing, which implies that a mixed chain will remain mixed. However, even though there are fundamental differences between the definition of mixing time in Cai et al. (2020) and the definition in the present paper, the example in Section 5 yields the same order of upper bound for mixing time as Cai et al. (2020, Theorem 1.2, Theorem 1.3).

Finally, there are papers that are more loosely related to our work, either by the models they study or the techniques they use. These works do not concern themselves with the existence of a stationary distribution. For instance, in Mans and Pourmiri (2022), the speed of rumor spreading, *spread time* is studied on dynamic graphs, including dynamic Erdős-Rényi graphs. Its authors discover an upper bound for the spread time through some isoperimetric observations, including certain bounds on the bottleneck ratio of the graph. As the spread time does not require a stationary distribution in its definition, it circumvents the above discussion in its entirety.

An interesting example of the evolving set method being used is Dembo et al. (2017). In this paper, the authors study the transience of dynamic random graphs, specifically of a time-varying conductance model. For their purposes, they make similar adjustments to the evolving set method as those presented in Section 3. However, the time-varying quantity they use for the evolving sets is related to the conductance of the edges in their model, and very different from our π_t.

1.3. *Organisation of the paper.* This paper is divided into four further sections: In Section 2, we define the proposed notion of mixing time. In particular, we give criteria under which our time-dependent target distribution exists. Technical proofs are deferred to Appendix A. In Section 3, we adapt the theory of evolving sets to the time-inhomogeneous setting. We show that our time-dependent target measure is a suitable replacement for the stationary distribution. We then use evolving sets to prove an upper bound on mixing time for time-inhomogeneous Markov chains. In Section 4, we move from deterministic time-inhomogeneous Markov chains to chains where the transition matrix for each time step is random. In Section 5, we present a concrete example of a randomly evolving Erdős-Rényi graph. After that, we make a few suggestions on potential improvements of our results.

2. Target distribution and mixing time

2.1. *Target distribution.* In this section, we will define the target distribution π_t and give an irreducibility condition that is sufficient to show its existence.

Without loss of generality, we consider Markov chains on the state space $[n] := \{1, \dots, n\}$ for a fixed $n \in \mathbb{N}$. Furthermore, let $P = (P_s)_{s\in\mathbb{Z}}$ be a sequence of transition matrices on $[n]$ and $X = (X_s)_{s\in\mathbb{Z}}$ the time-inhomogeneous Markov chain governed by the transition matrices, as in (1.4). Let $\mathbb{P}^P_{x,s}$ be the distribution of $(X_t)_{t\geq s}$ when started from x at time s.

We define a sequence of probability measures on $[n]$ that takes on the role of a stationary distribution for the Markov chain X. Set for $x, y \in [n]$, and $t \in \mathbb{Z}$,

$$\lim_{s\to-\infty} P^{s,t}(x, y) = \lim_{s\to-\infty} \mathbb{P}^P_{x,s}(X_t = y) =: Q^t(x, y) \tag{2.1}$$

for some limiting matrix Q^t, if the limit exists.

Definition 2.1. If the value of $Q^t(x, y)$ does not depend on x, the *time-dependent target distribution* π_t exists at time t, and is given by

$$\pi_t(y) = Q^t(x, y) \qquad \text{for all } x, y \in [n]. \tag{2.2}$$

To formulate conditions that imply the existence of π_t we introduce a quantity which measures the similarity between rows of a stochastic matrix P,

$$\delta(P) := \sup_{x,y} \sum_z [P(x, z) - P(y, z)]^+ = \sup_{x,y} \|P(x, \cdot) - P(y, \cdot)\|_{\mathrm{TV}} \in [0, 1]$$

where $[a]^+ := \max(0, a)$. $\delta(P) = 0$ if and only if all rows of P are equal. δ is known as Dobrushin's ergodic coefficient (see Dobrushin, 1956).

The question whether a matrix Q^t as in (2.1) exists is then related to the concepts of weak ergodicity/merging (rows of $P^{s,t}$ are approaching each other as $t \to \infty$) and strong ergodicity ($P^{s,t}$ converges to a limit as $t \to \infty$) that are described in e.g. Brémaud (2020, Chapter 12), with one fundamental difference: By taking the limit $s \to -\infty$ instead of $t \to \infty$, weak and strong ergodicity are equivalent. In that vein, we only need to find conditions such that $\lim_{s\to-\infty} \delta(P^{s,t}) = 0$ and can deduce the existence of the limit Q^t from that. The following proposition provides such conditions:

Proposition 2.2. *Let $(P_s)_{s\in\mathbb{Z}}$ be a sequence of stochastic matrices. For $t \in \mathbb{Z}$, $\varepsilon > 0$, iteratively define $t_0 = t$ and $t_k = \sup\{s \in \mathbb{Z} : s < t_{k-1}, \delta(P^{s,t_{k-1}}) \leq 1 - \varepsilon\}$, terminating when $t_k = -\infty$. Assume that for every $t \in \mathbb{Z}$, there exists $\varepsilon > 0$ such that $\{t_1, t_2, ...\}$ is an infinite set. Then, for every $t \in \mathbb{Z}$, $\lim_{s\to-\infty} P^{s,t} =: Q^t$ exists, $\delta(Q^t) = 0$, $\pi_t(y) := Q^t(1, y)$ for every $y \in [n]$ is well-defined.*

The proof of this proposition follows immediately from Lemma 2.3 and Lemma 2.4 below. These lemmas are standard results. For Lemma 2.3 we refer to Paz and Reichaw (1967, equation (2) on page 779). The proof of Lemma 2.4 is given in Appendix A.

Lemma 2.3 (Submultiplicativity)**.** *For any two stochastic matrices P, Q it holds that*

$$\delta(PQ) \leq \delta(P)\delta(Q).$$

Lemma 2.4. *Let $t \in \mathbb{Z}$. If $\lim_{s\to-\infty} \delta(P^{s,t}) = 0$, then there exists a rank 1 matrix Q^t such that*

$$\lim_{s\to-\infty} P^{s,t}(x,y) = Q^t(x,y), \qquad \text{for all } x,y \in [n].$$

Proof of Proposition 2.2: The submultiplicativity applied to $P^{t_k,t_{k-1}}$ for $k = 1,2,\dots$ yields

$$\lim_{s\to-\infty} \delta(P^{s,t}) = 0.$$

The claim follows from Lemma 2.4 with $Q^t(x,\cdot) = \pi_t(\cdot)$ for all $x \in [n]$. □

From now on, we assume that π_t exists for all $t \in \mathbb{Z}$ and state its properties.

Lemma 2.5. *$(\pi_s)_{s\in\mathbb{Z}}$ satisfies*

$$\pi_t = \pi_r P^{r,t} \tag{2.3}$$

for all $t > r \in \mathbb{Z}$.

Proof: By the Markov property it is evident that

$$\begin{aligned}\pi_{t+1}(y) &= \lim_{s\to-\infty} \mathbb{P}^P_{x,s}(X_{t+1} = y) = \lim_{s\to-\infty} \sum_z \mathbb{P}^P_{x,s}(X_{t+1} = y, X_t = z)\\ &= \lim_{s\to-\infty} \sum_z \mathbb{P}^P_{x,s}(X_t = z) P_{t+1}(z,y) = \sum_z \pi_t(z) P_{t+1}(z,y)\end{aligned}$$

for every $y \in [n]$ and for every $t \in \mathbb{Z}$, which shows $\pi_{t+1} = \pi_t P^{t,t+1}$. The claim follows by iteration. □

Lemma 2.5 allows to define a "stationary" version of X: Indeed, to this end it is sufficient to observe that by Lemma 2.5, the sequence of distributions $\mathbb{P}^P_{\pi_t,t}$ of the Markov chain started at time t with initial distribution π_t, is compatible in the sense that

$$\mathbb{P}^P_{\pi_{t-1},t-1}\big|_{[n]^{\{t,t+1,\dots\}}} = \mathbb{P}^P_{\pi_t,t}.$$

The Kolmogorov extension theorem then implies the existence of the measure $\mathbb{P}^P$ on $[n]^{\mathbb{Z}}$ under which the chain is "stationary", meaning that for all $x \in [n]$,

$$\mathbb{P}^P(X_t = y) = \lim_{s\to-\infty} \mathbb{P}^P_{x,s}(X_t = y) = \pi_t(y), \qquad t \in \mathbb{Z}, y \in [n]. \tag{2.4}$$

We call $\mathbb{P}^P$ the *law of the stationary chain.*

2.2. *Mixing Time.* With the measures $(\pi_t)_{t\in\mathbb{Z}}$ at our disposal, we define mixing time for time-inhomogeneous Markov chains, as alluded to in (1.6). Set for $s \in \mathbb{Z}$ and $t \geq 0$,

$$d(s,s+t) \coloneqq \sup_x \|P^{s,s+t}(x,\cdot) - \pi_{s+t}(\cdot)\|_{\mathrm{TV}}.$$

Definition 2.6. Let $\varepsilon \in (0,1)$ and $s \in \mathbb{Z}$. The ε-mixing time for a sequence $P = (P_t)_{t\in\mathbb{Z}}$ is defined by

$$t^P_{\mathrm{mix}}(\varepsilon,s) \coloneqq \inf\{t \geq 0 : d(s,s+t) \leq \varepsilon\}.$$

To make this definition meaningful, we need to confirm that if $d(s,s+t) \leq \varepsilon$ for some t, then $d(s,s+u) \leq \varepsilon$ for all $u \geq t$. Hence whether $d(s,s+\cdot)$ is monotonic is a question that arises naturally. As briefly discussed in the introduction, we can answer this affirmatively. Furthermore, mixing time is related to merging time (recall (1.11) in the introduction). The following lemma collects these results.

Lemma 2.7. *Let $t \geq s \geq u$. Then*

(a) $d(u,t) \leq d(u,s)$,
(b) $d(u,t) \leq d(s,t)$,

(c) $d(s,t) \le \delta(P^{s,t}) \le 2d(s,t)$.

The elementary proof of Lemma 2.7 is deferred to Appendix A.

3. Evolving sets for time-inhomogeneous Markov chains

At the beginning of this section we adapt the theory of evolving sets (as introduced by Morris and Peres (2005) for time-homogeneous chains) to time-inhomogeneous Markov chains. Note that, as explained in the introduction, the evolving set method has already been used in this context, assuming the existence of a time-independent stationary distribution (see Peres et al., 2018, 2020). We show here that the same results also hold in our new setup with π_t. We then apply the evolving set method to estimate the mixing time given by Definition 2.6. More specifically, we prove an upper bound on the mixing time in the spirit of a result in Jerrum and Sinclair (1989, Corollary 2.3). As mentioned in the introduction, this corollary can be generalized to non-reversible chains using evolving sets, and a full account of the proof can be found in Levin and Peres (2017, Theorem 17.10). While evolving sets allow for sharper bounds on stronger notions of mixing time (see Morris and Peres, 2005) in the time-homogeneous case, for this paper we content ourselves with generalizing the result presented in Levin and Peres (2017).

A crucial observation we have made is that Lemma 2.5 allows to seamlessly insert π_t into previously known results that require a stationary distribution. One example of this is the martingale property of $(\pi(S_t))_{t\ge 0}$ (see Levin and Peres, 2017, Lemma 17.13), where π is the stationary distribution and S_t is the stochastic process introduced in (3.5) below (with appropriate modifications for time-homogeneity). In Lemma 3.1, we show that $(\pi_t(S_t))_{t\ge 0}$ is also a martingale, using the same argument as Levin and Peres (2017, Lemma 17.13). Furthermore, to prove Theorem 3.2 below, we can follow the proof of Levin and Peres (2017, Theorem 17.10) with relatively small adjustments.

In the following, we will assume that for every P_s,

$$P_s(x,x) \ge \frac{1}{2}, \qquad \text{for all } x \in [n], \tag{3.1}$$

i.e. that the resulting Markov chain (X_t) is lazy. This avoids all problems with periodicity and is an assumption that is also often made in the homogeneous case. We will point out where it is used in our proofs.

Additionally, we make the following irreducibility assumption: For some $t_0 \in \mathbb{Z}$, assume that for every $x \in [n]$ there exists $s < t_0$ such that $\pi_s(x) > 0$. Together with (3.1), this implies that for every $t \ge t_0$ and for every $x \in [n]$, $\pi_t(x) > 0$.

3.1. *Evolving sets and mixing times.* Let $P = (P_s)_{s\in\mathbb{Z}}$ be a sequence of transition matrices such that π_t exists for every $t \in \mathbb{Z}$, and consider the corresponding Markov chain $X = (X_t)_{t\in\mathbb{Z}}$. The evolving set process (S_t) which we define below is a Markov chain on the space of all subsets of $[n]$. Its time until absorption in either $\emptyset$ or $[n]$ is closely linked to the mixing time of X. In preparation of its construction, recall the law of the stationary chain $\mathbb{P}^P$ from (2.4) and define for $t \in \mathbb{N}$, $A \subset [n]$, and $y \in [n]$,

$$Q_{t+1}(A,y) \coloneqq \mathbb{P}^P(X_t \in A, X_{t+1} = y) = \sum_{x\in A} \pi_t(x) P_{t+1}(x,y), \tag{3.2}$$

and furthermore for $B \subset [n]$

$$Q_{t+1}(A,B) \coloneqq \sum_{y\in B} Q_{t+1}(A,y).$$

The quantity $Q_{t+1}(A,B)$ is the "stationary flow" from A to B between time t and $t+1$. Observe that, due to Lemma 2.5, the stationary flow satifies

$$Q_{t+1}([n],y)=\pi_{t+1}(y), \qquad \text{for all } y\in[n], \tag{3.3}$$

and, by (3.2),

$$Q_{t+1}(A,[n])=\sum_{z\in[n]}Q_{t+1}(A,z)=\pi_t(A), \qquad A\subset[n]. \tag{3.4}$$

Let $(U_t)_{t\in\mathbb{Z}}$ be a family of i.i.d. uniform random variables on $[0,1]$ that is independent of $(X_t)_{t\in\mathbb{Z}}$. For a starting time $t_0\in\mathbb{Z}$ and a non-empty starting state $S=S_{t_0}\subset[n]$, iteratively set

$$S_{t+1}\coloneqq\left\{y\in[n]:\frac{Q_{t+1}(S_t,y)}{\pi_{t+1}(y)}\geq U_{t+1}\right\}, \qquad t\geq t_0. \tag{3.5}$$

This defines a time-inhomogeneous Markov chain $(S_t)_{t\geq t_0}$ on the set of subsets of $[n]$. It is easy to see that $\emptyset$ and $[n]$ are the absorbing states of this chain. Let $\mathbb{P}^P_{S,t_0}$ be the distribution of $(S_t)_{t\geq t_0}$. By (3.5) and the Markov property of $(S_t)_{t\geq t_0}$, we have

$$\frac{Q_{t+1}(S_t,y)}{\pi_{t+1}(y)}=\mathbb{P}^P_{S,t_0}(y\in S_{t+1}|S_t)=\mathbb{P}^P_{S_t,t}(y\in S_{t+1}), \tag{3.6}$$

where $\mathbb{P}^P_{S_t,t}$ is the distribution of $(S_s)_{s\geq t}$ started in the state S_t. Combining (3.2) and (3.6) yields

$$\mathbb{P}^P_{S,t}(y\in S_{t+1})\pi_{t+1}(y)=Q_{t+1}(S,y)=\sum_{z\in S}\pi_t(z)P_{t+1}(z,y). \tag{3.7}$$

We now have all the tools to prove that $(\pi_t(S_t))_{t\geq t_0}$ is a martingale. This will be later used to link the growth of S_t to the bottleneck ratio of the underlying P_t. Since the choice of starting time t_0 is essentially arbitrary, without loss of generality we assume $t_0=0$.

Lemma 3.1. *The sequence $(\pi_t(S_t))_{t\geq 0}$ is a martingale under $\mathbb{P}^P_{S,0}$ with respect to $(\mathcal{F}_t)_{t\geq 0}$, where $\mathcal{F}_t=\sigma(S_s:0\leq s\leq t)$. In particular*

$$\mathbb{E}^P_{S,t}[\pi_{t+1}(S_{t+1})]=\pi_t(S) \tag{3.8}$$

for every $S\subset[n]$ and $t\in\mathbb{N}$.

Proof of Lemma 3.1: The proof is similar to the homogeneous version Levin and Peres (2017, Lemma 17.13). Since $(S_t)_{t\geq 0}$ is a Markov chain, it suffices to condition on S_t in place of $\mathcal{F}_t$:

$$\begin{aligned}\mathbb{E}^P_{S,0}[\pi_{t+1}(S_{t+1})|S_t]&=\mathbb{E}^P_{S,0}\left[\sum_{z\in[n]}\mathbb{1}_{\{z\in S_{t+1}\}}\pi_{t+1}(z)\middle|S_t\right]\\&=\sum_{z\in[n]}\mathbb{P}^P_{S,0}(z\in S_{t+1}|S_t)\pi_{t+1}(z)\overset{(3.6)}{=}\sum_{z\in[n]}Q_{t+1}(S_t,z)\\&\overset{(3.4)}{=}\pi_t(S_t),\end{aligned}$$

hence it is a martingale. (3.8) follows immediately. □

To estimate the mixing time, we express the time-dependent connectivity structure of the underlying state space in terms of the growth of the evolving set. Before we can state the main result of this section, we introduce auxiliary notation. Let

$$g_t\coloneqq\min_{z\in[n]}\frac{\pi_{t-1}(z)}{\pi_t(z)}\in(0,1] \tag{3.9}$$

bound the rate of change between π_{t-1} and π_t. In time-homogeneous settings, or when $\pi_t = \pi_{t-1}$ for other reasons, $g_t = 1$. Furthermore denote

$$\pi_t^{\min} := \min_z \pi_t(z),$$

which will be frequently used to give quantitative lower bounds on $\pi_t(S)$ for $S \subset [n]$ based on the number of elements in S. We further define

$$\Phi_t(S) := \frac{1}{2\pi_{t-1}(S)}(Q_t(S,S^c) + Q_t(S^c,S)).$$

Note that $\Phi_t(S)$ normalizes the stationary flow between S and its complement S^c by the size of the set S under π_{t-1}. Finally, we define the *time-dependent bottleneck ratio*

$$\Phi_t^* := \inf\{\Phi_t(S) : S \subset [n] \text{ with } \pi_{t-1}(S) \le 1/2\}. \tag{3.10}$$

We now have the necessary prerequisites to state the main theorem of this section.

Theorem 3.2. *Fix $\varepsilon \in (0,1)$. If for some $t > 0$*

$$\frac{1}{\sqrt{\pi_t^{\min}\pi_0^{\min}}} \prod_{s=1}^{t} \left[1 - \frac{1}{2}\left(\frac{1}{2}\frac{g_s}{1-\frac{1}{2}g_s}\Phi_s^*\right)^2\right] \le 2\varepsilon, \tag{3.11}$$

then

$$d(0,t) \le \varepsilon.$$

Remark 3.3. Theorem 3.2 only yields that $t_{\text{mix}}^P(\varepsilon, 0) \le t$ for any t that satisfies (3.11), while giving no estimate on the t. We refine the statement in Corollary 3.8 and Corollary 4.3 to strengthen the link to mixing time.

3.2. *Proof of Theorem 3.2.* Let

$$\psi_t(S) := 1 - \mathbb{E}_{S,t-1}^P\left[\sqrt{\frac{\pi_t(S_t)}{\pi_{t-1}(S)}}\right]. \tag{3.12}$$

The proof of Theorem 3.2 will be based on the observation that we can relate $\psi_t(S)$ to $d(0,t)$. However, first we show that $\Phi_t(S)$ is closely related to $\psi_t(S)$. We do this in two lemmas below. (In the following we use the standard notation $a \wedge b := \min(a,b)$.)

Lemma 3.4. *Let $\varphi_t(S) := \frac{1}{2\pi_{t-1}(S)}\sum_{y\in[n]}(Q_t(S,y) \wedge Q_t(S^c,y))$. Then for every $S \subset [n]$*

$$1 - \psi_t(S) \le \frac{\sqrt{1+2\varphi_t(S)} + \sqrt{1-2\varphi_t(S)}}{2} \le 1 - \frac{\varphi_t(S)^2}{2}. \tag{3.13}$$

Proof: First, note that $\frac{\sqrt{1+2x}+\sqrt{1-2x}}{2} \le 1 - \frac{x^2}{2}$ holds for any real number $x \in [-1/2, 1/2]$. Hence, to show the second inequality in (3.13), it is enough to verify that $\varphi_t(S) \in [-1/2, 1/2]$. It is clear that $\varphi_t(S) \ge 0$. On the other hand

$$\sum_{y\in[n]} (Q_t(S,y) \wedge Q_t(S^c,y)) \le \sum_{y\in[n]} Q_t(S,y) \overset{(3.2)}{=} \pi_{t-1}(S)$$

which implies that $\varphi_t(S) \le 1/2$. Hence $\varphi_t(S) \in [-1/2, 1/2]$.

The first inequality in (3.13) is harder to prove. Recall that U_t denotes the uniform random variable used to generate S_t from S_{t-1}. We split the proof by conditioning on $U_t \le 1/2$ first, and $U_t > 1/2$ later. Note that conditioned on $U_t \in [0,1/2]$, U_t is uniform on $[0,1/2]$. By (3.5), it is immediate that

$$\mathbb{P}_{S,t-1}^P(y \in S_t | U_t \le 1/2) = 1 \wedge 2\frac{Q_t(S,y)}{\pi_t(y)}. \tag{3.14}$$

After multiplying both sides of (3.14) by $\pi_t(y)$, this implies

$$\pi_t(y)\mathbb{P}^P_{S,t-1}(y \in S_t | U_t \leq 1/2) = \pi_t(y) \wedge 2Q_t(S,y) \overset{(3.3)}{=} (Q_t(S,y) + Q_t(S^c,y)) \wedge 2Q_t(S,y)$$
$$= Q_t(S,y) + (Q_t(S^c,y) \wedge Q_t(S,y)).$$

Summing over all $y \in [n]$ yields

$$\mathbb{E}^P_{S,t-1}[\pi_t(S_t)|U_t \leq 1/2] = \sum_{y\in[n]} Q_t(S,y) + \sum_{y \in [n]} (Q_t(S^c,y) \wedge Q_t(S,y))$$
$$\overset{(3.4)}{=} \pi_{t-1}(S) + 2\pi_{t-1}(S)\varphi_t(S).$$

Dividing both sides by $\pi_{t-1}(S)$ and defining $R_t := \frac{\pi_t(S_t)}{\pi_{t-1}(S_{t-1})}$ results in

$$\mathbb{E}^P_{S,t-1}[R_t|U_t \leq 1/2] = 1 + 2\varphi_t(S).$$

However, by the martingale property (3.8),

$$\mathbb{E}^P_{S,t-1}[R_t] = \mathbb{E}^P_{S,t-1}[\pi_t(S_t)/\pi_{t-1}(S_{t-1})] = \frac{\pi_{t-1}(S)}{\pi_{t-1}(S)} = 1.$$

Since $\mathbb{P}^P_{S,t-1}(U_t \leq 1/2) = 1/2$, this implies that

$$\mathbb{E}^P_{S,t-1}[R_t|U_t > 1/2] = 1 - 2\varphi_t(S).$$

Note that from (3.12), $1 - \psi_t(S) = \mathbb{E}^P_{S,t-1}[\sqrt{R_t}]$, so by Jensen's inequality we can conclude

$$1-\psi_t(S) = \frac{1}{2}\Big(\mathbb{E}^P_{S,t-1}[\sqrt{R_t}|U_t \leq 1/2] + \mathbb{E}^P_{S,t-1}[\sqrt{R_t}|U_t > 1/2]\Big)$$
$$\leq \frac{1}{2}\Big(\sqrt{\mathbb{E}^P_{S,t-1}[R_t|\, U_t \leq 1/2]} + \sqrt{\mathbb{E}^P_{S,t-1}[R_t|U_t > 1/2]}\Big)$$
$$= \frac{1}{2}\Big(\sqrt{1+2\varphi_t(S)} + \sqrt{1-2\varphi_t(S)}\Big).$$

This shows the first inequality in (3.13) and completes the proof. □

We can now prove a relation between ψ_t and Φ_t. This result is inspired by Morris and Peres (2005, Lemma 3), with an additional term appearing due to time-inhomogeneity.

Lemma 3.5. *For every $t > 0$ and for every set $S \subset [n]$*

$$\psi_t(S) \geq \frac{1}{8}\frac{g_t^2}{(1-\frac{1}{2}g_t)^2}(\Phi_t(S))^2. \tag{3.15}$$

Proof: We show that

$$\varphi_t(S) \geq \frac{1}{2}\frac{g_t}{1-\frac{1}{2}g_t}\Phi_t(S). \tag{3.16}$$

From (3.16) and Lemma 3.4, (3.15) follows. To prove (3.16), consider $y \in S$. Clearly $Q_t(S,y) \geq Q_t(y,y) \geq \frac{1}{2}\pi_{t-1}(y)$, by the laziness (3.1), and thus

$$\frac{Q_t(S,y)}{\pi_t(y)} \geq \frac{1}{2}\frac{\pi_{t-1}(y)}{\pi_t(y)} \geq \frac{1}{2}\min_z \frac{\pi_{t-1}(z)}{\pi_t(z)} \overset{(3.9)}{=} \frac{1}{2}g_t. \tag{3.17}$$

Noting that

$$Q_t(S^c,y) \overset{(3.3)}{=} \pi_t(y) - Q_t(S,y) \overset{(3.17)}{\leq} \Big(1 - \frac{1}{2}g_t\Big)\pi_t(y)$$

we deduce

$$\frac{1}{2}\frac{g_t}{1-\frac{1}{2}g_t}Q_t(S^c,y) \le \frac{1}{2}g_t\pi_t(y). \tag{3.18}$$

So, combining (3.17) and (3.18) yields that for $y \in S$

$$\begin{aligned} Q_t(S,y) \wedge Q_t(S^c,y) &\ge \left(\frac{1}{2}g_t\pi_t(y)\right) \wedge Q_t(S^c,y) \\ &\ge \left(\frac{1}{2}\frac{g_t}{1-\frac{1}{2}g_t}Q_t(S^c,y)\right) \wedge Q_t(S^c,y) \\ &= \frac{1}{2}\frac{g_t}{1-\frac{1}{2}g_t}Q_t(S^c,y), \end{aligned}$$

where the last equality holds since $g_t \le 1$. On the other hand, if $y \in S^c$, swapping all instances of S and S^c, the argument above yields

$$Q_t(S,y) \wedge Q_t(S^c,y) \ge \frac{1}{2}\frac{g_t}{1-\frac{1}{2}g_t}Q_t(S,y).$$

Summing over all $y \in [n]$, we can separate the case $y \in S$ from $y \in S^c$ to get

$$\begin{aligned} \sum_{y\in[n]} Q_t(S,y) \wedge Q_t(S^c,y) &= \sum_{y\in S} Q_t(S,y) \wedge Q_t(S^c,y) + \sum_{y\in S^c} Q_t(S,y) \wedge Q_t(S^c,y) \\ &\ge \frac{1}{2}\ \frac{g_t}{1-\frac{1}{2}g_t}\sum_{y\in S} Q_t(S^c,y) + \frac{1}{2}\frac{g_t}{1-\frac{1}{2}g_t}\sum_{y\in S^c} Q_t(S,y) \\ &= \frac{1}{2}\ \frac{g_t}{1-\frac{1}{2}g_t}(Q_t(S^c,S) + Q_t(S,S^c)). \end{aligned}$$

Multiplying both sides by $\frac{1}{2\pi_{t-1}(S)}$ results in (3.16) which completes the proof. □

In order to prove Theorem 3.2, we need the following proposition (cf. Levin and Peres, 2017, Lemma 17.12) that relates $P^{0,t}$ and S_t.

Proposition 3.6. *For every $t \ge 0$, and for every $x, y \in [n]$,*

$$P^{0,t}(x,y) = \frac{\pi_t(y)}{\pi_0(x)}\mathbb{P}^P_{\{x\},0}(y \in S_t). \tag{3.19}$$

Proof: We proceed by induction. For $t = 0$ there is nothing to prove. Assume (3.19) holds up to $t-1$ for every x, y. Then for arbitrary $x, y \in [n]$

$$\begin{aligned} P^{0,t}(x,y) &= \sum_{z\in[n]} P^{0,t-1}(x,z)P_t(z,y) = \sum_{z\in[n]} \mathbb{P}^P_{\{x\},0}(z \in S_{t-1})\frac{\pi_{t-1}(z)}{\pi_0(x)}P_t(z,y) \\ &= \frac{\pi_t(y)}{\pi_0(x)}\sum_{z\in[n]} \mathbb{P}^P_{\{x\},0}(z \in S_{t-1})\frac{\pi_{t-1}(z)}{\pi_t(y)}P_t(z,y) \end{aligned}$$

by using the induction assumption. $\mathbb{E}_{\{x\},0}[\mathbb{1}_{z\in S_{t-1}}] = \mathbb{P}^P_{\{x\},0}(z \in S_{t-1})$ yields that

$$\begin{aligned}
P^{0,t}(x,y) &= \frac{\pi_t(y)}{\pi_0(x)}\mathbb{E}^P_{\{x\},0}\Big[\sum_{z\in S_{t-1}} \pi_{t-1}(z)P_t(z,y)\pi_t(y)^{-1}\Big] \\
&\overset{(3.2)}{=} \frac{\pi_t(y)}{\pi_0(x)}\mathbb{E}^P_{\{x\},0}\Big[\sum_{z\in S_{t-1}} Q_t(z,y)\pi_t(y)^{-1}\Big] \\
&= \frac{\pi_t(y)}{\pi_0(x)}\mathbb{E}^P_{\{x\},0}\big[\pi_t(y)^{-1}Q_t(S_{t-1},y)\big] \overset{(3.6)}{=} \frac{\pi_t(y)}{\pi_0(x)}\mathbb{E}^P_{\{x\},0}\Big[\mathbb{P}^P_{\{x\},0}(y\in S_t|S_{t-1})\Big] \\
&= \frac{\pi_t(y)}{\pi_0(x)}\mathbb{P}^P_{\{x\},0}(y \in S_t).
\end{aligned}$$

This shows the induction step and completes the proof. □

We can now prove the main result of the section.

Proof of Theorem 3.2: We follow the strategy outlined in Levin and Peres (2017, Theorem 17.10). For every t, we define

$$S^\#_t := \begin{cases} S_t, & \text{if } \pi_t(S_t) \le 1/2, \\ S^c_t, & \text{if } \pi_t(S_t) > 1/2. \end{cases} \tag{3.20}$$

It is useful to note in preparation that $(S^c_t)_{t\ge 0}$ is a stochastic process that has the same transition probabilities as $(S_t)_{t\ge0}$, since

$$\begin{aligned}
S^c_t &= \{y \in [n] : \tfrac{Q_t(S,y)}{\pi_t(y)} < U_t\} \\
&= \{y \in [n] : \tfrac{Q_t(S^c,y)}{\pi_t(y)} \ge 1 - U_t\} \\
&= \{y \in [n] : \tfrac{Q_t(S^c,y)}{\pi_t(y)} \ge \tilde{U}_t\},
\end{aligned}$$

where $\tilde{U}_t = 1 - U_t$ is again uniform. This also shows that S^c_t, given that $S_{t-1} = S$, behaves like an evolving set started in S^c at time $t-1$, that is to say for $S, T \subset [n]$,

$$\mathbb{P}_{S,t-1}(S_t = T) = \mathbb{P}_{S^c,t-1}(S_t = T^c)$$

and in particular for this proof

$$\mathbb{E}_{S,t-1}\Big[\sqrt{\pi_t(S^c_t)}\Big] = \mathbb{E}_{S^c,t-1}\Big[\sqrt{\pi_t(S_t)}\Big]. \tag{3.21}$$

Recall the notation $R_t = \pi_t(S_t)/\pi_{t-1}(S_{t-1})$. Observe that

$$\mathbb{E}^P_{S,t-1}\big[\sqrt{R_t}\big] \overset{(3.13)}{\le} 1 - \varphi_t(S)^2/2 \overset{(3.15)}{\le} 1 - \frac{1}{2}\Big(\frac{1}{2}\frac{g_t}{1-\frac{1}{2}g_t}\Phi_t(S)\Big)^2.$$

Fix $S_{t-1} = S \subset [n]$. If $\pi_{t-1}(S) \le 1/2$, then

$$\frac{\pi_t(S^\#_t)}{\pi_{t-1}(S^\#_{t-1})} \le \frac{\pi_t(S_t)}{\pi_{t-1}(S)} \tag{3.22}$$

because $\pi_t(S^\#_t) \le \pi_t(S_t)$ by (3.20) and $S^\#_{t-1} = S$ by assumption. Taking the expectation of (3.22) therefore yields

$$\mathbb{E}^P_{S,t-1}\Bigg[\sqrt{\frac{\pi_t(S^\#_t)}{\pi_{t-1}(S^\#_{t-1})}}\Bigg] \le \mathbb{E}^P_{S,t-1}\big[\sqrt{R_t}\big] \le 1 - \frac{1}{2}\Big(\frac{1}{2}\frac{g_t}{1-\frac{1}{2}g_t}\Phi_t(S)\Big)^2. \tag{3.23}$$

Similarly, if $\pi_{t-1}(S) > 1/2$, then

$$\frac{\pi_t(S_t^\#)}{\pi_{t-1}(S_{t-1}^\#)} \leq \frac{\pi_t(S_t^c)}{\pi_{t-1}(S^c)} \tag{3.24}$$

because $\pi_t(S_t^\#) \leq \pi_t(S_t^c)$ by (3.20) and $S_{t-1}^\# = S^c$ by assumption. Taking the expectation of the square root and starting from S^c at $t-1$, using (3.21), yields

$$\mathbb{E}_{S,t-1}^P\left[\sqrt{\frac{\pi_t(S_t^\#)}{\pi_{t-1}(S_{t-1}^\#)}}\right] \leq \mathbb{E}_{S^c,t-1}^P\left[\sqrt{\frac{\pi_t(S_t)}{\pi_{t-1}(S^c)}}\right] \leq 1 - \frac{1}{2}\left(\frac{1}{2}\frac{g_t}{1-\frac{1}{2}g_t}\Phi_t(S^c)\right)^2. \tag{3.25}$$

Note that $\pi_{t-1}(S^c) \leq 1/2$ so indeed $\Phi_t(S^c) \geq \inf\{\Phi_t(V) : V \subset [n] \text{ with } \pi_{t-1}(V) \leq 1/2\}$. Combining (3.23) and (3.25) hence results in

$$\mathbb{E}_{\{x\},0}^P\left[\sqrt{\frac{\pi_t(S_t^\#)}{\pi_{t-1}(S_{t-1}^\#)}}\middle| S_{t-1}\right] \leq 1 - \frac{1}{2}\left(\frac{1}{2}\frac{g_t}{1-\frac{1}{2}g_t}\Phi_t^*\right)^2.$$

Multiplying both sides by $\sqrt{\pi_{t-1}(S_{t-1}^\#)}$ and then taking expectations gives by the tower property that

$$\mathbb{E}_{\{x\},0}^P\left[\sqrt{\pi_t(S_t^\#)}\right] \leq \left[1 - \frac{1}{2}\left(\frac{1}{2}\frac{g_t}{1-\frac{1}{2}g_t}\Phi_t^*\right)^2\right]\mathbb{E}_{\{x\},0}^P\left[\sqrt{\pi_{t-1}(S_{t-1}^\#)}\right]. \tag{3.26}$$

Recursively applying (3.26), we arrive at

$$\mathbb{E}_{\{x\},0}^P\left[\sqrt{\pi_t(S_t^\#)}\right] \leq \sqrt{\pi_0(x)} \cdot \prod_{s=1}^t \left[1 - \frac{1}{2}\left(\frac{1}{2}\frac{g_s}{1-\frac{1}{2}g_s}\Phi_s^*\right)^2\right].$$

Clearly $\sqrt{\pi_t^{\min}} \cdot \mathbb{P}_{\{x\},0}^P(S_t^\# \neq \emptyset) \leq \mathbb{E}_{\{x\},0}^P\left[\sqrt{\pi_t(S_t^\#)}\right]$. Rearranging the terms gives

$$\mathbb{P}_{\{x\},0}^P(S_t^\# \neq \emptyset) \leq \frac{\sqrt{\pi_0(x)}}{\sqrt{\pi_t^{\min}}} \cdot \prod_{s=1}^t \left[1 - \frac{1}{2}\left(\frac{1}{2}\frac{g_s}{1-\frac{1}{2}g_s}\Phi_s^*\right)^2\right]. \tag{3.27}$$

We will now introduce four identities that relate $d(0,t)$ to (3.27). Let

$$\tau := \inf\{t \geq 0 : S_t^\# = \emptyset\} = \inf\{t \geq 0 : S_t = \emptyset \text{ or } S_t = [n]\}.$$

By the optional stopping theorem and the law of total expectation

$$\begin{aligned}\pi_0(x) &= \mathbb{E}_{\{x\},0}^P[\pi_{\tau\wedge t}(S_{\tau\wedge t})] \\ &= \mathbb{E}_{\{x\},0}^P\left[\pi_\tau(S_\tau)\middle|\tau \leq t\right]\mathbb{P}_{\{x\},0}^P(\tau \leq t) + \mathbb{E}_{\{x\},0}^P\left[\pi_t(S_t)\middle|\tau > t\right]\mathbb{P}_{\{x\},0}^P(\tau > t).\end{aligned} \tag{3.28}$$

On the other hand, by Proposition 3.6, for any $x, y \in [n]$

$$|P^{0,t}(x,y) - \pi_t(y)| = \frac{\pi_t(y)}{\pi_0(x)}\left|\mathbb{P}_{\{x\},0}^P(y \in S_t) - \pi_0(x)\right|. \tag{3.29}$$

Note that $\{y \in S_t, \tau \leq t\} = \{S_\tau = [n], \tau \leq t\}$ by definition of τ, since both $\emptyset$ and $[n]$ are absorbing states for S_t. Therefore

$$\begin{aligned}\mathbb{P}_{\{x\},0}^P(y \in S_t) &= \mathbb{P}_{\{x\},0}^P(y \in S_t, \tau > t) + \mathbb{P}_{\{x\},0}^P(y \in S_t, \tau \leq t) \\ &= \mathbb{P}_{\{x\},0}^P(y \in S_t, \tau > t) + \mathbb{P}_{\{x\},0}^P(S_\tau = [n], \tau \leq t).\end{aligned} \tag{3.30}$$

As a final piece of preparation, note that also

$$\mathbb{P}_{\{x\},0}^P(S_\tau = [n], \tau \leq t) = \mathbb{E}_{\{x\},0}^P\left[\pi_\tau(S_\tau)\middle|\tau \leq t\right]\mathbb{P}_{\{x\},0}^P(\tau \leq t). \tag{3.31}$$

Hence, by (3.28)–(3.31),

$$|P^{0,t}(x,y)-\pi_t(y)| = \frac{\pi_t(y)}{\pi_0(x)}\left|\mathbb{P}^P_{\{x\},0}(y\in S_t, \tau>t)-\mathbb{E}^P_{\{x\},0}[\pi_t(S_t)|\tau>t]\mathbb{P}^P_{\{x\},0}(\tau>t)\right|$$
$$\le \frac{\pi_t(y)}{\pi_0(x)}\mathbb{P}^P_{\{x\},0}(\tau>t) = \frac{\pi_t(y)}{\pi_0(x)}\mathbb{P}^P_{\{x\},0}(S^\#_t\neq\emptyset).$$

It is easy to show that $d(0,t)=\sup_x\|P^{0,t}(x,\cdot)-\pi_t(\cdot)\|_{\mathrm{TV}}\le\frac{1}{2}\sup_{x,y}\frac{|P^{0,t}(x,y)-\pi_t(y)|}{\pi_t(y)}$, see Levin and Peres (2017, Chapter 4.7) for a time-homogeneous version. So combining the above with (3.27) results in

$$2d(0,t)\le\sup_{x,y}\frac{|P^{0,t}(x,y)-\pi_t(y)|}{\pi_t(y)}$$
$$\le\sup_x\frac{1}{\pi_0(x)}\frac{\sqrt{\pi_0(x)}}{\sqrt{\pi_t^{\min}}}\prod_{s=1}^t\left[1-\frac{1}{2}\left(\frac{1}{2}\frac{g_s}{1-\frac{1}{2}g_s}\Phi^*_s\right)^2\right]$$
$$\le\frac{1}{\sqrt{\pi_t^{\min}\pi_0^{\min}}}\prod_{s=1}^t\left[1-\frac{1}{2}\left(\frac{1}{2}\frac{g_s}{1-\frac{1}{2}g_s}\Phi^*_s\right)^2\right].$$

In order to mix, we want to find t, such that $d(0,t)\le\varepsilon$ is implied. That is clearly the case when the right-hand side is smaller than 2ε, which completes the proof. □

Remark 3.7. In the final part of the proof we deduced that

$$\sup_{x,y}\frac{|P^{0,t}(x,y)-\pi_t(y)|}{\pi_t(y)}\le\frac{1}{\sqrt{\pi_t^{\min}\pi_0^{\min}}}\prod_{s=1}^t\left[1-\frac{1}{2}\left(\frac{1}{2}\frac{g_s}{1-\frac{1}{2}g_s}\Phi^*_s\right)^2\right].$$

Defining

$$d^{(\infty)}(0,t):=\sup_{x,y}\frac{|P^{0,t}(x,y)-\pi_t(y)|}{\pi_t(y)},$$

it is clear that Theorem 3.2, like in the time-homogeneous case, also applies to a notion of mixing stronger than mixing in total variation.

It is often impractical to compute all the individual Φ^*_s. In Section 5 we compute a uniform lower bound on Φ^*_s that does not depend on the time s. With this application in mind, the statement of Theorem 3.2 is more involved than necessary. Define

$$\Theta_t:=\min_{1\le s\le t}\left(\frac{1}{2}\frac{g_s}{1-\frac{1}{2}g_s}\Phi^*_s\right)^2. \tag{3.32}$$

Note that $\Theta_t=0$ for every $t\ge s$ if the state space becomes disconnected at time $s\ge1$. So the following simplification is only useful when studying models where the state space remains connected until the chain has mixed.

Corollary 3.8. *Fix $\varepsilon\in(0,1)$. If for some $t>0$*

$$t\ge\frac{2}{\Theta_t}\left[\log\left(\frac{1}{2\sqrt{\pi_0^{\min}\pi_t^{\min}}}\right)+\log\left(\varepsilon^{-1}\right)\right], \tag{3.33}$$

then

$$d(0,t)\le\varepsilon.$$

Proof: By definition of Θ_t,

$$\prod_{s=1}^{t}\left[1-\frac{1}{2}\left(\frac{1}{2}\frac{g_s}{1-\frac{1}{2}g_s}\Phi_s^*\right)^2\right] \leq (1-\frac{1}{2}\Theta_t)^t \leq e^{-t\Theta_t/2}.$$

If we have

$$\frac{1}{\sqrt{\pi_t^{\min}\pi_0^{\min}}}e^{-t\Theta_t/2} \leq 2\varepsilon \tag{3.34}$$

then we can apply Theorem 3.2. Taking the logarithm of (3.34) gives the sufficient condition

$$t \geq \frac{2}{\Theta_t}\left[\log\left(\frac{1}{2\sqrt{\pi_0^{\min}\pi_t^{\min}}}\right)+\log\left(\varepsilon^{-1}\right)\right]$$

which implies $d(0,t) \leq \varepsilon$ and thus completes the proof. □

Remark 3.9. Let us denote the right-hand side of (3.33) by

$$F(t) = \frac{2}{\Theta_t}\left[\log\left(\frac{1}{2\sqrt{\pi_0^{\min}\pi_t^{\min}}}\right)+\log\left(\varepsilon^{-1}\right)\right].$$

In the time-homogeneous case, $F(t) =: F$ does not depend on t. Thus, for $t = F$, Corollary 3.8 implies that $d(0,F) \leq \varepsilon$, i.e. $t_{\text{mix}}^P(\varepsilon,0) \leq F$. With only small adjustments in the proof of Theorem 3.2, Levin and Peres (2017, Theorem 17.10) and Jerrum and Sinclair (1989, Corollary 2.3) follow.

In the time-inhomogeneous case, this is unfortunately not immediate. While we do know that for every t with $t \geq F(t)$ we have $d(0,t) \leq \varepsilon$, it does not follow that $d(0,F(t)) \leq \varepsilon$, since every such t could be strictly larger than the corresponding $F(t)$. In particular, $F(t)$ is not an upper bound on mixing time. We could hope that $t = F(t)$ for some t, from which $d(0,F(t)) \leq \varepsilon$ follows, but unfortunately there is no reason to believe that such t exists.

One way around this issue is to construct $T(n), \tau(n) \in \mathbb{N}$, such that $T(n) \geq \tau(n)$ and

$$F(t) \leq \tau(n) \qquad \text{for every } t \in \{0,\dots,T(n)\}.$$

Then immediately $t_{\text{mix}}^P(\varepsilon,0) \leq \tau(n)$, by simply picking $t = \tau(n)$ and applying Corollary 3.8, yielding the desired upper bound. Since the choice of $T(n)$ and $\tau(n)$ depends on the model, we will make this more specific in the following sections after having introduced the random environment.

4. Random environment

In Sections 2 and 3, $P = (P_s)_{s\in\mathbb{Z}}$ has been a fixed sequence of transition matrices on $[n]$. In applications we have in mind, this sequence is interpreted as a random sample from an underlying distribution $\mathbf{P}$ on the sequences of transition matrices. We assume that $\mathbf{P}$ is such that the sequence $(P_s)_{s\in\mathbb{Z}}$ of transition matrices is a time-homogeneous Markov chain. For each sample $(P_s)_{s\in\mathbb{Z}}$ we consider the law of the stationary chain $\mathbb{P}^{(P_s)_{s\in\mathbb{Z}}}$ as in (2.4) and proceed with the tools from Section 3.

Let

$$\mathcal{Q}_n := \left\{P \in [0,1]^{n\times n} : \sum_{y=1}^{n} P(x,y) = 1 \quad \forall x \in [n]\right\}$$

be the space of all $n \times n$ stochastic matrices on $[n]$. Let $P = (P_s)_{s\in\mathbb{Z}}$ be some time-homogeneous Markov chain on $\mathcal{Q}_n$ with unique stationary distribution Π that starts from equilibrium, and write $\mathbf{P}$ for the law of the stationary Markov chain. In particular, we have

$$\mathbf{P}(P_t \in A) = \Pi(A), \qquad t \in \mathbb{Z}, A \subset \mathcal{Q}_n.$$

In this setting, we first give a sufficient condition under which π_t, as defined in Section 2, exists $\mathbf{P}$-almost surely.

Proposition 4.1. *Assume that there exists a measurable set $A \subset \mathcal{Q}_n$ with*

$$\varepsilon := \sup\{\delta(P) : P \in A\} < 1$$

such that

$$\mathbf{P}\left(\sum_{k=1}^{\infty} \mathbb{1}_A(P_k) = \infty\right) = 1.$$

Then π_t exists $\mathbf{P}$-a.s. for every $t \in \mathbb{Z}$.

Proof: Let $t \in \mathbb{Z}$. Since $\mathbf{P}(\sum_{k=1}^{\infty} \mathbb{1}_A(P_k) = \infty) = 1$, because of time-homogeneity and stationarity of P, for every $m \in \mathbb{N}$ there exists $T_m \in \mathbb{N}$ such that

$$\mathbf{P}\left(\sum_{k=1+t-T_m}^{t} \mathbb{1}_A(P_k) \geq m\right) = \mathbf{P}\left(\sum_{k=1}^{T_m} \mathbb{1}_A(P_k) \geq m\right) \geq 1 - 2^{-m}.$$

Since $Q \in A$ implies $\delta(Q) \leq \varepsilon$, by submultiplicativity of δ, we also have

$$\left\{\sum_{k=1+t-T_m}^{t} \mathbb{1}_A(P_k) \geq m\right\} \subset \{\delta(P^{t-T_m,t}) \leq \varepsilon^m\} \subset \{\lim_{s\to-\infty} \delta(P^{s,t}) \leq \varepsilon^m\}.$$

Hence

$$\mathbf{P}\left(\lim_{s\to-\infty} \delta(P^{s,t}) > \varepsilon^m\right) \leq \mathbf{P}\left(\sum_{k=1+t-T_m}^{t} \mathbb{1}_A(P_k) < m\right) \leq 2^{-m}, \qquad m \in \mathbb{N}.$$

By picking m arbitrarily large, $\varepsilon^m \to 0$, we conclude that $\lim_{s\to-\infty} \delta(P^{s,t}) = 0$, $\mathbf{P}$-a.s. The $\mathbf{P}$-almost sure existence of π_t follows for almost every $(P_s)_{s\in\mathbb{Z}}$ fixed separately by Lemma 2.4 and the same arguments as in Proposition 2.2. □

Remark 4.2. The condition of Proposition 4.1 essentially requires that infinitely many transition matrices are "irreducible". This can be relaxed. For the argument above, it is sufficient that products of finite length, $P^{u,u+k}$, with $\delta(P^{u,u+k}) < 1$ are visited infinitely often, and then the result follows by submultiplicativity of $\delta(\cdot)$ applied to those products instead of individual transition matrices. In other words, as long as there are infinitely many opportunities for the chain to reach every state, then π_t does exist, even if the state space is never connected at once. This can be viewed as a sort of *dynamic connectivity.*

In the following, we are interested in the behavior of mixing time as $n \to \infty$. Consider a sequence of probability spaces $(\mathcal{Q}_n, \mathcal{A}_n, \mathbf{P}_n)_{n\in\mathbb{N}}$. A sequence of events $(A_n)_{n\in\mathbb{N}}$ (where each $A_n \in \mathcal{A}_n$) is said to occur *with high probability* (w.h.p.) if

$$\lim_{n\to\infty} \mathbf{P}_n(A_n) = 1.$$

To simplify notation, we write $\mathbf{P}$ in place of $\mathbf{P}_n$. We use this concept of *high probability* in the following corollary, where we give an explicit bound on mixing time that holds w.h.p. only. Recall the definition of Θ_t from (3.32).

Corollary 4.3. *Assume there exists a constant $\beta > 0$ such that $\pi_t^{\min} \geq n^{-\beta}$ for every $t \in \{0, \dots, n\}$ w.h.p. and furthermore assume there exists a constant $\kappa > 0$, such that $\Theta_t \geq \kappa$ for every $t \in \{0, \dots, n\}$ w.h.p. Then, for every $\varepsilon \in (0, 1)$,*

$$\lim_{n\to\infty} \mathbf{P}\left(t_{\mathrm{mix}}(\varepsilon, 0) \leq 1 + \frac{2}{\kappa}\left[\log\left(\tfrac{n^\beta}{2}\right) + \log(\varepsilon^{-1})\right]\right) = 1.$$

Proof: Fix $\varepsilon \in (0, 1)$, let $n \in \mathbb{N}$ be large enough. For any $t \in \{0, \dots, n\}$, it holds that

$$\frac{2}{\Theta_t}\left[\log\left(\tfrac{1}{2\sqrt{\pi_0^{\min}\pi_t^{\min}}}\right) + \log(\varepsilon^{-1})\right] \leq \frac{2}{\kappa}\left[\log\left(\tfrac{n^\beta}{2}\right) + \log(\varepsilon^{-1})\right] \leq n \tag{4.1}$$

with high probability. Let $t = \left\lceil \frac{2}{\kappa}\Big(\log\big(\frac{n^\beta}{2}\big) + \log(\varepsilon^{-1})\Big)\right\rceil \in \{0, \dots, n\}$. By Corollary 3.8, (4.1) implies that $d(0,t) \leq \varepsilon$ with high probability. Hence

$$\lim_{n\to\infty} \mathbf{P}\bigg(t_{\mathrm{mix}}(\varepsilon, 0) \leq 1 + \frac{2}{\kappa}\Big[\log\big(\tfrac{n^\beta}{2}\big) + \log(\varepsilon^{-1})\Big]\bigg) = 1$$

which completes the proof. □

Remark 4.4. In light of the discussion in Remark 3.9, this corollary eliminates the dependency on t of the right-hand side of (3.33). For that purpose, in the notation of Remark 3.9, we chose $T(n) = n$ and $\tau(n) = \frac{2}{\kappa}\big[\log\big(\frac{n^\beta}{2}\big) + \log(\varepsilon^{-1})\big]$.

Remark 4.5. The quantity $t_{\mathrm{mix}}(\varepsilon, 0)$ is a random variable that depends on the *sequence* of transition matrices $(P_s)_{s\in\mathbb{Z}}$. As such, the randomness of the transition matrices and the randomness of the resulting chain are inherently viewed separately.

We point out that even small changes in the setup can significantly alter the interpretation of mixing time: In Avena et al. (2018), the authors consider mixing time as a random variable that depends on the *initial state* of the environment, whereas the dynamics of the environment are observed jointly with the random walk. Therefore the dynamics of the environment can speed up the mixing. As an example, if the environment is likely to undergo significant changes in every time step, the mixing time as defined in Avena et al. (2018) is essentially constant.

5. Dynamic Erdős-Rényi graphs

We now demonstrate how the results from Section 3 and their Corollary 4.3 can be applied to a concrete example, a random walk on an Erdős-Rényi graph that is independently resampled after each time step.

Let $(G_t)_{t\in\mathbb{Z}}$ be a sequence of independent Erdős-Rényi graphs of size $n \in \mathbb{N}$ with parameter $p_n \in (0,1)$. We introduce the following notation: We write $\deg_t(x)$ for the (random) degree of vertex x at time t and we denote by $x \sim_t y$ that x and y are connected by an edge in the graph G_t.

Given G_t, the transition matrix P_t corresponding to a lazy simple random walk on the graph is given by

$$P_t(x,y) = \begin{cases} 1/2 & \text{if } x = y \text{ and } \deg_t(x) \geq 1, \\ 1 & \text{if } x = y \text{ and } \deg_t(x) = 0, \\ 1/(2\deg_t(x)) & \text{if } x \neq y \text{ and } x \sim_t y, \\ 0 & \text{otherwise.} \end{cases}$$

Since the transition matrices are i.i.d., $(P_s)_{s\in\mathbb{Z}}$ is a time-homogeneous Markov chain that has a stationary distribution Π determined by the edge probability p_n. As before we write $\mathbf{P}$ for the distribution of $(P_s)_{s\in\mathbb{Z}}$.

Observe that the complete graph is sampled with positive probability and thus infinitely many G_t's are complete graphs $\mathbf{P}$-almost surely. The assumptions of Proposition 4.1 are therefore easily verified and it follows that $(\pi_t)_{t\in\mathbb{Z}}$ exists for $\mathbf{P}$-almost every sequence $(P_t)_{t\in\mathbb{Z}}$.

In the above construction we assume that each graph is independent from the previous. This assumption seems restrictive, however we will see that even under this assumption, the subsequent proofs are non-trivial. One could expand the graph dynamic to a more classical Markovian model where non-existent edges appear with probability p_n and existing edges disappear with probability q_n. This introduces correlation between π_t and P_{t+1} which makes the proof techniques we present here fail in some crucial aspects; namely, in Proposition 5.5 below. We have unfortunately not found a way to address these issues to account for dependencies.

Additionally, we make some assumptions on the parameter p_n. In particular, we choose p_n such that the graphs are very likely to be connected, with p_n significantly above the usual connectivity threshold for Erdős-Rényi graphs. We did not make an effort to optimize the constant of the connectivity requirement in the theorem, however we do not think reducing it to 1 would be straightforward. We explain two improvement ideas in Section 5.4.

Theorem 5.1. *Let $p_n = \frac{\eta \log n}{n-1}$ with $\eta > 50$. Then there exists a constant $c' = c'(\eta)$, such that for every $\varepsilon \in (0,1)$*

$$\lim_{n\to\infty} \mathbf{P}\big(t_{\text{mix}}(\varepsilon,0) \le c'[\log n + \log(\varepsilon^{-1})]\big) = 1. \tag{5.1}$$

Remark 5.2. By the stationarity of the random graphs, $t_{\text{mix}}(\varepsilon,0) \overset{d}{=} t_{\text{mix}}(\varepsilon,s)$ for all s. Therefore we only consider $t_{\text{mix}}(\varepsilon,0)$.

Remark 5.3. Before every time step, all edges in the graph are independently resampled. This leads to the tempting conclusion that the random walk can forget its starting position very quickly and $t_{\text{mix}}(\varepsilon,0) = c'(\varepsilon)$ for some constant c' independent of n. This intuition is *false* (if $p_n \ll 1$), since our concept of mixing is *quenched*: We fix the transition matrices beforehand and then view mixing time of the random walk given that fixed sequence of transition matrices. We prove a lower bound on mixing time in Theorem 5.13, demonstrating that mixing time is indeed not constant.

To prove Theorem 5.1, we apply Corollary 4.3. To this end, we show that w.h.p. for every $t \in \{0,\dots,n\}$, $\pi_t^{\min} \ge \frac{\alpha_1^*}{n}$ for some $0 < \alpha_1^* < 1$ independent of n, and $\Theta_t \ge \kappa > 0$ for some constant κ independent of n.

5.1. *Lower and upper bounds for π_t.* In this section, we work under the assumptions of Theorem 5.1. That is to say

$$p_n = \frac{\eta \log n}{n-1} \text{ with } \eta > 50. \tag{5.2}$$

The aim is to show there exist some constants $\alpha_1^*(\eta) \in (0,1)$, $\alpha_2^*(\eta) > 1$, and $\beta > 0$ independent of n, such that for all n large enough

$$\mathbf{P}\Big(\frac{\alpha_1^*}{n} \le \pi_t(x) \le \frac{\alpha_2^*}{n}, \forall t \in \{0,...,n\}, \forall x \in [n]\Big) \ge 1 - n^{-\beta}. \tag{5.3}$$

The following lemma provides a first basic, yet useful, upper bound for π_t.

Lemma 5.4. *Let $(P_s)_{s\in\mathbb{Z}}$ be a realization from $\mathbf{P}$, such that $(\pi_s)_{s\in\mathbb{Z}}$ exists. Then*

$$\pi_t(x) \le \tfrac{1}{2}, \qquad \textit{for all } x \in [n], \textit{ for all } t \in \mathbb{Z}.$$

Proof: Note that $P_t(x,x) = \frac{1}{2}$ (which holds unless x is isolated) implies that

$$\pi_t(x) \overset{(2.3)}{=} \frac{1}{2}\pi_{t-1}(x) + \sum_{y\neq x} \pi_{t-1}(y)P_t(y,x) \le \frac{1}{2}[\pi_{t-1}(x) + (1-\pi_{t-1}(x))] \le \frac{1}{2} \tag{5.4}$$

for every $t \in \mathbb{Z}, x \in [n]$. If $P_t(x,x) = 1$, then $\pi_t(x) = \pi_{t-1}(x)$. So $\pi_t(x) > \frac{1}{2}$ is only possible when $P_s(x,x) = 1$ for all $s < t$, meaning that x has been isolated for the entire history of the graph. Then π_t is not well-defined, as $\lim_{s\to-\infty} \mathbb{P}^P_{x,s}(X_t = x) = 1$, yet for any $z \neq x$, $\lim_{s\to-\infty} \mathbb{P}^P_{z,s}(X_t = x) = 0$. This contradicts the assumption that $(\pi_s)_{s\in\mathbb{Z}}$ exists. □

With the next proposition we gain good control over π_t. Its proof will take up the remainder of this section.

Proposition 5.5. *There exist constants $\alpha_1^*(\eta) \in (0,1)$, $\alpha_2^*(\eta) > 1$, and $\beta > 0$ independent of n, such that for all n large enough there exists a non-decreasing sequence $(\alpha_1^{(t)})_{0\le t\le n^2}$ and a non-increasing sequence $(\alpha_2^{(t)})_{0\le t\le n^2}$ with $\alpha_1^{(t)} = \alpha_1^*$ and $\alpha_2^{(t)} = \alpha_2^*$ for all $t \ge n^2 - n$, such that*

$$\mathbf{P}\left(\frac{\alpha_1^{(t)}}{n} \le \pi_t(x) \le \frac{\alpha_2^{(t)}}{n}, \forall t \in \{0,...,n^2\}, \forall x \in [n]\right) \ge 1 - n^{-\beta}. \tag{5.5}$$

Remark 5.6. By time-invariance of the random graph sequence, (5.3) is an immediate consequence of (5.5), with those same constants $\alpha_1^*(\eta) \in (0,1)$, $\alpha_2^*(\eta) > 1$, and $\beta > 0$.

In preparation of the proof of Proposition 5.5, we give some definitions and notation. To specify the sequences used in (5.5), we fix $\beta = 0.3$, $\alpha_1^* = 0.002$, $\alpha_2^* = 7$, and $\varepsilon = 10^{-4}$ independent of n. These constants are chosen to make the proof work, but are not optimized to achieve the best possible constant c' in Theorem 5.1. We define the sequences

$$\alpha_2^{(t)} = \begin{cases} \frac{n}{2}, & \text{if } 0 \le t \le n^{1.1}, \\ (1-\varepsilon)\alpha_2^{(t-1)}, & \text{if } n^{1.1} < t \le n^{1.2} \text{ and } (1-\varepsilon)\alpha_2^{(t-1)} \ge \alpha_2^*, \\ \alpha_2^*, & \text{otherwise,} \end{cases} \tag{5.6}$$

and

$$\alpha_1^{(t)} = \begin{cases} 0, & \text{if } 0 \le t \le n^{1.1}, \\ \frac{n}{(16\eta \log n)^{n-1}}, & \text{if } n^{1.1} < t \le n^{1.2}, \\ (1+\varepsilon)\alpha_1^{(t-1)}, & \text{if } n^{1.2} < t < n^2 - n \text{ and } (1+\varepsilon)\alpha_1^{(t-1)} \le \alpha_1^*, \\ \alpha_1^*, & \text{otherwise.} \end{cases} \tag{5.7}$$

Note that $\alpha_1^{(t)}$ and $\alpha_2^{(t)}$ depend on n.

We then define the events

$$F_{u,t} := \left\{\pi_t(x) \le \tfrac{\alpha_2^{(t)}}{n}, \forall x \in [n]\right\},$$
$$F_{l,t} := \left\{\pi_t(x) \ge \tfrac{\alpha_1^{(t)}}{n}, \forall x \in [n]\right\},$$
$$F_t := F_{u,t} \cap F_{l,t},$$

so that (5.5) is equivalent to $\mathbf{P}(\cap_{t\le n^2} F_t) \ge 1 - n^{-\beta}$. We note that, due to Lemma 5.4,

$$\mathbf{P}(F_0) = 1. \tag{5.8}$$

Furthermore, we fix

$$c_1 = \frac{11}{21}\eta, \quad c_2 = 2\eta.$$

We define

$$D_t := \{\deg_t(x) \in [c_1 \log n, c_2 \log n], \forall x \in [n]\},$$
$$C_t := \{G_k \text{ is a connected graph for all } 0 \le k < t\}.$$

It is well-known for Erdős-Rényi graphs with p_n as in (5.2) that D_1 and C_1 are events that occur with high probability. However, Proposition 5.5 requires quantitative estimates which we collect in Lemma 5.7 and Lemma 5.8 below. Their proofs are given in Appendix B.

Lemma 5.7. *There exists $\rho > 2 + \beta$, such that for every n large enough and every $t \in \mathbb{Z}$,*

$$\mathbf{P}(D_t^c) \le n^{-\rho}. \tag{5.9}$$

Lemma 5.8. *For every n large enough and for every $0 < t \le n^2$*

$$\mathbf{P}(C_t^c) \le \mathbf{P}(C_{n^2}^c) \le n^{-\frac{\eta}{2}+4}.$$

Let us now turn our attention to (5.5). We decompose the complement of the probability in (5.5), using the notation established above:

$$\begin{aligned}\mathbf{P}\left(\bigcup_{t=0}^{n^2}(F_{u,t}^c \cup F_{l,t}^c)\right) &= \mathbf{P}(F_{u,0}^c \cup F_{l,0}^c) + \sum_{t=1}^{n^2}\mathbf{P}(F_0,\dots,F_{t-1},(F_{u,t}^c \cup F_{l,t}^c)) \\ &\overset{(5.8)}{\leq} \sum_{t=1}^{n^2}(\mathbf{P}(F_0,\dots,F_{t-1},F_{u,t}^c) + \mathbf{P}(F_0,\dots,F_{t-1},F_{l,t}^c)) \\ &\leq \sum_{t=1}^{n^2}(\mathbf{P}(F_{u,t}^c,F_{u,t-1}) + \mathbf{P}(F_{l,t}^c,F_{t-1},C_t) + \mathbf{P}(F_{l,t}^c,F_{t-1},C_t^c)) \\ &\leq n^2\mathbf{P}(C_{n^2}^c) + \sum_{t=1}^{n^2}(\mathbf{P}(F_{u,t}^c|F_{u,t-1}) + \mathbf{P}(F_{l,t}^c,F_{t-1},C_t)).\end{aligned}$$

We have already established in Lemma 5.8 that $n^2\mathbf{P}(C_{n^2}^c)$ is small, so it remains to bound the final sum. To that end we will use, for $t \leq n^{1.2}$,

$$\mathbf{P}(F_{l,t}^c,F_{t-1},C_t) \leq \mathbf{P}(F_{l,t}^c,C_t) \tag{5.10}$$

and for $t > n^{1.2}$

$$\mathbf{P}(F_{l,t}^c,F_{t-1},C_t) \leq \mathbf{P}(F_{l,t}^c,F_{t-1}) \leq \mathbf{P}(F_{l,t}^c,D_t|F_{t-1}) + \mathbf{P}(D_t^c). \tag{5.11}$$

Further note that

$$\mathbf{P}(F_{u,t}^c|F_{u,t-1}) \leq \mathbf{P}(F_{u,t}^c,D_t|F_{u,t-1}) + \mathbf{P}(D_t^c). \tag{5.12}$$

To prove Proposition 5.5, it now suffices to prove that the right-hand sides of (5.10), (5.11) and (5.12) are each smaller than $n^{-\rho}$ for some $\rho > 2+\beta$. Lemma 5.7 already bounds $\mathbf{P}(D_t^c) \leq n^{-\rho}$ for a (potentially different) $\rho > 2+\beta$.

Remark 5.9. We compute bounds that have a leading term $n^{-\rho}$ for some $\rho > 2+\beta$ independent of n. To simplify and remove all multiplicative constants we reduce ρ very slightly and say the statement holds for all n large enough. This is not a limitation since Theorem 5.1 only considers the limit $n \to \infty$.

We will restate explicitly that statements hold for all n large enough when it is important to clarify which constants do not depend on n. The constant ρ varies from statement to statement, but it never depends on n.

The following lemma gives a bound on the first term appearing on the right-hand side of (5.12).

Lemma 5.10. *There exists $\rho > 2+\beta$, such that for every n large enough and for all $t \in \{0,\dots,n^2\}$,*

$$\mathbf{P}(F_{u,t}^c,D_t|F_{u,t-1}) \leq n^{-\rho}. \tag{5.13}$$

Proof of Lemma 5.10: For $t \leq n^{1.1}$, $\alpha_2^{(t)} = n/2$, so $\mathbf{P}(F_{u,t}^c|F_{u,t-1}) = 0$ since $\pi_t(x) \leq 1/2$ for every x, by Lemma 5.4. Therefore we can reduce the problem to $t > n^{1.1}$.

If the event D_t occurs, then, as a consequence of Lemma 2.5 and by definition of P_t,

$$\pi_t(x) = \frac{1}{2}\pi_{t-1}(x) + \sum_{y\neq x}\pi_{t-1}(y)\frac{\mathbb{1}_{x\sim_t y}}{2\deg_t(y)} \leq \frac{1}{2}\pi_{t-1}(x) + \sum_{y\neq x}\pi_{t-1}(y)\frac{\mathbb{1}_{x\sim_t y}}{2c_1\log n}. \tag{5.14}$$

For arbitrary $x \in [n]$,

$$\begin{aligned}\mathbf{P}(F_{u,t}^c, D_t | F_{u,t-1}) &\leq n\mathbf{P}\Big(\pi_t(x) > \frac{\alpha_2^{(t)}}{n}, D_t \Big| F_{u,t-1}\Big) \\ &\overset{(5.14)}{\leq} n\mathbf{P}\Big(\frac{1}{2}\pi_{t-1}(x) + \sum_{y\neq x} \frac{1}{2}\pi_{t-1}(y)\frac{\mathbb{1}_{x\sim_t y}}{c_1 \log n} > \frac{\alpha_2^{(t)}}{n}\Big| F_{u,t-1}\Big) \\ &\leq n\mathbf{P}\Big(\frac{1}{2}\frac{n-1}{n}\frac{\alpha_2^{(t-1)}}{\alpha_2^{(t)}} + \sum_{y\neq x}\frac{1}{2}\frac{n-1}{\alpha_2^{(t)}}\pi_{t-1}(y)\frac{\mathbb{1}_{x\sim_t y}}{c_1\log n} > \frac{n-1}{n}\Big| F_{u,t-1}\Big) \\ &\leq n\mathbf{P}\left(\sum_{y\neq x}\frac{n-1}{\alpha_2^{(t)}}\pi_{t-1}(y)\mathbb{1}_{x\sim_t y} > \frac{n-1}{n}\left(2 - \frac{\alpha_2^{(t-1)}}{\alpha_2^{(t)}}\right)c_1 \log n \middle| F_{u,t-1}\right),\end{aligned} \tag{5.15}$$

where the third line follows by multiplying by $n-1$ and dividing by $\alpha_2^{(t)}$, and using that conditioned on the event $F_{u,t-1}$, $\pi_{t-1}(x) \leq \alpha_2^{(t-1)}/n$.

Let $\Lambda_x = \sum_{y\neq x}(n-1)\pi_{t-1}(y)\mathbb{1}_{x\sim_t y}/\alpha_2^{(t)}$ denote the sum in the last formula. It holds that

$$\mathbf{E}[\Lambda_x|\pi_{t-1}] = \frac{n-1}{\alpha_2^{(t)}}\frac{\eta\log n}{n-1}(1-\pi_{t-1}(x)) = \frac{\eta \log n}{\alpha_2^{(t)}}(1-\pi_{t-1}(x)). \tag{5.16}$$

To proceed, we need a concentration result for Λ_x under $\mathbf{P}(\cdot|F_{u,t-1})$, namely for every $\delta > 0$,

$$\mathbf{P}\big(\Lambda_x > (1+\delta)\mathbf{E}[\Lambda_x|\pi_{t-1}]\big|F_{u,t-1}\big) \leq \exp\Big[(\delta - (1+\delta)\log(1+\delta))\tfrac{\eta\log n}{2\alpha_2^{(t)}}\Big]. \tag{5.17}$$

We postpone the proof of (5.17) and first complete the proof of (5.13). We set

$$\hat{\delta} := \begin{cases} \frac{11}{21}(1-3\varepsilon)\alpha_2^{(t-1)} - 1, & \text{if } n^{1.1} < t \leq n^{1.2} \text{ and } \alpha_2^{(t)} \neq \alpha_2^*, \\ \frac{77}{21}(1-3\varepsilon) - 1, & \text{if } n^{1.2} < t \leq n^2 \text{ or } \alpha_2^{(t)} = \alpha_2^*. \end{cases}$$

Then it is easily shown by recalling $\beta = 0.3$, $c_1 = \frac{11}{21}\eta$, $\alpha_2^* = 7$, $\eta > 50$, and $\varepsilon = 10^{-4}$, that

$$\frac{n-1}{n}\left(2 - \frac{\alpha_2^{(t-1)}}{\alpha_2^{(t)}}\right)c_1\log n \geq (1+\hat{\delta})\mathbf{E}[\Lambda_x|\pi_{t-1}]. \tag{5.18}$$

Moreover there exists $\rho > 2+\beta$ independent of $\hat{\delta}, t$, and n, such that

$$\exp\Big[(\hat{\delta} - (1+\hat{\delta})\log(1+\hat{\delta}))\tfrac{\eta\log n}{2\alpha_2^{(t)}}\Big] \leq n^{-\rho-1}. \tag{5.19}$$

By (5.15) and (5.17) it follows that there exists a $\rho > 2+\beta$ such that for every n large enough,

$$\mathbf{P}(F_{u,t}^c, D_t|F_{u,t-1}) \leq n^{-\rho}.$$

Since this holds for every $n^{1.1} < t \leq n^2$, (5.13) follows.

So let us now turn our attention to (5.17). The inequality resembles a standard result about weighted sums of Bernoulli random variables achieved by the means of Chernoff bounds, and some variant and discussion of it can be found in Raghavan (1988, Theorem 1). However, the expression (5.17) has a key difference to classical theory, namely that $\mathbf{E}[\Lambda_x|\pi_{t-1}]$ and the weights $(n-1)\pi_{t-1}(y)/\alpha_2^{(t)}$ are random variables themselves. If a weight can randomly be large while all other weights are comparatively small, the sum Λ_x is mainly affected by the outcome of a single Bernoulli random variable, and thus does not concentrate.

However, as it turns out, the condition on the event $F_{u,t-1}$ is the deciding ingredient. When t is small and $F_{u,t-1}$ is barely restricting the weights in size, by construction of $\alpha_2^{(t-1)}$, the right-hand side of (5.17) is close to 1. When t is large and the right-hand side tightens, the condition $F_{u,t-1}$

limits how large single weights can be, which improves concentration. In essence, with increasing t, (5.17) improves itself accordingly.

To show (5.17), we use the Markov inequality for conditional probability: For any event F with positive probability, and random variables $X \geq 0$, $Y > 0$,

$$\mathbf{P}(X \geq Y|F) \leq \frac{\mathbf{E}[\mathbb{1}_F X/Y]}{\mathbf{P}(F)}. \tag{5.20}$$

For any $\lambda > 0$, with $X = e^{\lambda \Lambda_x}$, $Y = \exp\left[\lambda(1+\delta)\eta \log n(1-\pi_{t-1}(x))/\alpha_2^{(t)}\right]$, $F = F_{u,t-1}$, the expectation on the right-hand side of (5.20) becomes

$$\mathbf{E}\Big[e^{\lambda\Lambda_x} e^{-\eta \log n(1-\pi_{t-1}(x))(1+\delta)\lambda/\alpha_2^{(t)}} \mathbb{1}_{F_{u,t-1}}\Big]. \tag{5.21}$$

We emphasize here that $\pi_{t-1}(x)$ is random, however it is independent from all the Bernoulli random variables $\mathbb{1}_{x\sim_t y}$ in Λ_x. In fact, the $\mathbb{1}_{x\sim_t y}$ are also independent from $\mathbb{1}_{F_{u,t-1}}$. The tower property yields

$$(5.21) = \mathbf{E}\left[e^{-\eta\log n(1-\pi_{t-1}(x))(1+\delta)\lambda/\alpha_2^{(t)}} \mathbb{1}_{F_{u,t-1}} \mathbf{E}\left[e^{\lambda\sum_{y\neq x}(n-1)\pi_{t-1}(y)\mathbb{1}_{x\sim_t y}/\alpha_2^{(t)}}\middle|\pi_{t-1}\right]\right]$$

since $\mathbb{1}_{F_{u,t-1}}$ is also $\sigma(\pi_{t-1})$-measurable. Exploiting the independence of the $\mathbb{1}_{x\sim_t y}$'s

$$\begin{aligned}\mathbf{E}\left[e^{\lambda\sum_{y\neq x}(n-1)\pi_{t-1}(y)\mathbb{1}_{x\sim_t y}/\alpha_2^{(t)}}\middle|\pi_{t-1}\right] &= \prod_{y\neq x}\mathbf{E}\left[e^{\lambda(n-1)\pi_{t-1}(y)\mathbb{1}_{x\sim_t y}/\alpha_2^{(t)}}\middle|\pi_{t-1}\right]\\ &= \prod_{y\neq x}\left(1-p_n+p_n e^{\lambda(n-1)\pi_{t-1}(y)/\alpha_2^{(t)}}\right),\end{aligned}$$

where the equalities hold $\mathbf{P}$-almost surely. Let $Z_y \coloneqq (n-1)\pi_{t-1}(y)/\alpha_2^{(t)}$. Using the inequality $(1+z) \leq e^z$ and choosing $\lambda = \log(1+\delta)$, the right-hand side is bounded by

$$\prod_{y\neq x}\left(1-p_n+p_n e^{\log(1+\delta)Z_y}\right) \leq \prod_{y\neq x}\exp[p_n((1+\delta)^{Z_y}-1)].$$

On the event $F_{u,t-1}$, $Z_y \in (0,1]$. We can hence apply the inequality $(1+z)^l \leq 1+lz$ for $l \in (0,1]$ and $z > 0$ to arrive at

$$\mathbb{1}_{F_{u,t-1}}\prod_{y\neq x}\exp[p_n((1+\delta)^{Z_y}-1)] \leq \mathbb{1}_{F_{u,t-1}}\prod_{y\neq x}\exp[p_n\delta Z_y].$$

Note that

$$\prod_{y\neq x}\exp[p_n\delta Z_y] = \exp[p_n\delta(n-1)(1-\pi_{t-1}(x))/\alpha_2^{(t)}].$$

Using that $p_n = \frac{\eta\log n}{n-1}$ and $1-\pi_{t-1}(x) \geq \frac{1}{2}$, the calculation above implies that for $\lambda = \log(1+\delta)$

$$\begin{aligned}(5.21) &\leq \mathbf{P}(F_{u,t-1})\left[\frac{e^\delta}{(1+\delta)^{1+\delta}}\right]^{\frac{\eta\log n}{2\alpha_2^{(t)}}}\\ &= \mathbf{P}(F_{u,t-1})\exp\left[(\delta-(1+\delta)\log(1+\delta))\frac{\eta\log n}{2\alpha_2^{(t)}}\right].\end{aligned} \tag{5.22}$$

This, together with (5.20), implies (5.17). □

We can now show that the gradually ascending lower bound $F_{l,t}$ also holds with high probability.

Lemma 5.11. *There exists $\rho > 2+\beta$ such that for every n large enough and for every $n^{1.2} < t \le n^2$,*

$$\mathbf{P}(F_{l,t}^c, F_{t-1}) \le n^{-\rho}, \tag{5.23}$$

and for every $0 \le t \le n^{1.2}$,

$$\mathbf{P}(F_{l,t}^c, C_t) \le n^{-\rho}. \tag{5.24}$$

Proof: We first show (5.23). Its proof uses similar ideas as the proof of Lemma 5.10. For $t > n^{1.2}$, intersecting $F_{l,t}^c, F_{t-1}$ with D_t and D_t^c yields

$$\mathbf{P}(F_{l,t}^c, F_{t-1}) \le \mathbf{P}(F_{l,t}^c, D_t | F_{t-1}) + \mathbf{P}(D_t^c),$$

and we know from Lemma 5.7 that $\mathbf{P}(D_t^c)$ is small. It remains to consider $\mathbf{P}(F_{l,t}^c, D_t|F_{t-1})$. We fix an arbitrary $x \in [n]$. Bounding $\pi_t(x) \ge \frac{1}{2}\pi_{t-1}(x) + \sum_{y\neq x} \pi_{t-1}(y)\frac{\mathbb{1}_{x\sim_t y}}{2c_2 \log n}$ and repeating the steps in Lemma 5.10, we arrive at

$$\begin{aligned}
\mathbf{P}(F_{l,t}^c, D_t|F_{t-1}) &\le n\mathbf{P}\Big(\frac{1}{2}\pi_{t-1}(x) + \sum_{y\neq x}\frac{1}{2}\pi_{t-1}(y)\frac{\mathbb{1}_{x\sim_t y}}{2c_2\log n} < \frac{\alpha_1^{(t)}}{n}\Big| F_{t-1}\Big) \\
&\le n\mathbf{P}\Big(\frac{\alpha_1^{(t-1)}}{n}\frac{n-1}{\alpha_2^{(t)}} + \frac{1}{c_2\log n}\sum_{y\neq x}\frac{n-1}{\alpha_2^{(t)}}\pi_{t-1}(y)\mathbb{1}_{x\sim_t y} < 2\frac{n-1}{n}\frac{\alpha_1^{(t)}}{\alpha_2^{(t)}}\Big| F_{t-1}\Big) \\
&\le n\mathbf{P}\Big(\sum_{y\neq x}\frac{n-1}{\alpha_2^{(t)}}\pi_{t-1}(y)\mathbb{1}_{x\sim_t y} < \frac{n-1}{n}\frac{2\alpha_1^{(t)} - \alpha_1^{(t-1)}}{\alpha_2^*}c_2\log n\Big| F_{t-1}\Big)
\end{aligned} \tag{5.25}$$

where the second line follows by multiplying with $(n-1)/\alpha_2^{(t)}$. Since $t > n^{1.2}$ and hence $\alpha_2^{(t)} = \alpha_2^*$, we know conditional on the event F_{t-1} that $(n-1)\pi_{t-1}(y)/\alpha_2^* \le 1$. We write the left-hand side inside the last probability in (5.25) as $M_x = \sum_{y\neq x}(n-1)\pi_{t-1}(y)\mathbb{1}_{x\sim_t y}/\alpha_2^*$. It holds that

$$\mathbf{E}[M_x|\pi_{t-1}] = \eta\log n(1-\pi_{t-1}(x))/\alpha_2^*.$$

For all $\gamma \in (0,1)$ we will show the concentration result

$$\mathbf{P}\big(M_x < (1-\gamma)\mathbf{E}[M_x|\pi_{t-1}]\ \big| F_{t-1}\big) \le \exp\Big(-\frac{\gamma^2}{2}\frac{\eta\log n}{\alpha_2^*}(1-\tfrac{\alpha_2^*}{n})\Big). \tag{5.26}$$

We postpone the proof of (5.26) and first complete the proof of (5.23). For $n^{1.2} < t \le n^2$, we fix

$$\hat\gamma := \begin{cases} 1 - 4(1+2\varepsilon)\alpha_1^{(t-1)}, & \text{if } n^{1.2} < t \le n^2 - n \text{ and } \alpha_1^{(t)} \neq \alpha_1^*, \\ 1 - 4\alpha_1^*, & \text{if } n^2 - n < t \le n^2 \text{ or } \alpha_1^{(t)} = \alpha_1^*. \end{cases}$$

If we are in the situation that $\alpha_1^{(t)} \neq \alpha_1^*$, then indeed $\alpha_1^{(t)} = (1+\varepsilon)\alpha_1^{(t-1)}$ by definition, hence

$$2\alpha_1^{(t)} - \alpha_1^{(t-1)} = (1+2\varepsilon)\alpha_1^{(t-1)}$$

By recalling the values of $\varepsilon = 10^{-4}$ and $c_2 = 2\eta$, we see immediately that $\hat\gamma \in (0,1)$ and furthermore

$$(1-\hat\gamma) \ge \frac{n-1}{n}\Big(2\alpha_1^{(t)} - \alpha_1^{(t-1)}\Big)\frac{2c_2}{\eta}. \tag{5.27}$$

In the case $\alpha_1^{(t)} = \alpha_1^*$, it is easy to see that (5.27) also holds. Furthermore, since $\alpha_2^* = 7$, it is a short computation to show that for $\rho = 2.4 > 2 + \beta$,

$$\exp\Big(-\frac{\hat\gamma^2}{2}\frac{\eta\log n}{\alpha_2^*}\big(1-\frac{\alpha_2^*}{n}\big)\Big) \le n^{-\rho-1}. \tag{5.28}$$

By (5.25)–(5.28), we conclude that

$$\mathbf{P}(F_{l,t}^c, D_t|F_{t-1}) \le n^{-\rho},$$

and hence (5.23).

For the case $n^{1.2} < t \le n^2$ it remains to show the concentration result (5.26). Similar to Lemma 5.10, we show an upper bound for

$$\mathbf{E}\Big[e^{-\gamma M_x} e^{-\eta \log n(1-\pi_{t-1}(x))(1-\gamma)\gamma/\alpha_2^*} \mathbb{1}_{F_{t-1}}\Big]. \tag{5.29}$$

By the tower property,

$$(5.29) = \mathbf{E}\Big[e^{-\eta \log n(1-\pi_{t-1}(x))(1-\gamma)\gamma/\alpha_2^*} \mathbb{1}_{F_{t-1}} \mathbf{E}\Big[e^{-\gamma \sum_{y\neq x}(n-1)\pi_{t-1}(y)\mathbb{1}_{x\sim_t y}/\alpha_2^*}\Big|\pi_{t-1}\Big]\Big].$$

We now consider the inner conditional expectation. We write

$$Z_y := (n-1)\pi_{t-1}(y)/\alpha_2^*$$

and it holds that

$$\begin{aligned}\mathbf{E}\Big[e^{-\gamma\sum_{y\neq x} Z_y \mathbb{1}_{x\sim_t y}}\Big|\pi_{t-1}\Big] &= \prod_{y\neq x}\mathbf{E}\Big[e^{-\gamma Z_y \mathbb{1}_{x\sim_t y}}\Big|\pi_{t-1}\Big] \\ &\le \prod_{y\neq x}(1-p_n+p_n e^{-\gamma Z_y}) \le \exp\Big[\sum_{y\neq x} -p_n + p_n e^{-\gamma Z_y}\Big].\end{aligned}$$

Since $0 < Z_y^2 \le Z_y \le 1$ on the event F_{t-1}, and $-1+e^{-z} \le -z + z^2/2$ for $z>0$, we obtain

$$\mathbb{1}_{F_{t-1}} \exp\Big[\sum_{y\neq x} -p_n + p_n e^{-\gamma Z_y}\Big] \le \mathbb{1}_{F_{t-1}} e^{\sum_{y\neq x} p_n(-\gamma Z_y + \gamma^2 Z_y/2)}.$$

Noting that $\sum_{y\neq x} Z_y = (1-\pi_{t-1}(x))\eta\log n/\alpha_2^*$ then finally yields

$$(5.29) \le \mathbf{P}(F_{t-1})e^{-\gamma^2\eta\log n(1-\pi_{t-1}(x))/(2\alpha_2^*)} \le \mathbf{P}(F_{t-1})e^{-\gamma^2\eta\log n(1-\alpha_2^*/n)/(2\alpha_2^*)}$$

since $\pi_{t-1}(x) \le \alpha_2^*/n$ on the event F_{t-1}. Applying (5.20) completes the proof of (5.26).

We now turn to (5.24). For $t \le n^{1.1}$, by (5.7), $\alpha_1^{(t)} = 0$, which makes $F_{l,t}^c$ trivially empty, so $\mathbf{P}(F_{l,t}^c) = 0$. For $n^{1.1} < t \le n^{1.2}$, let $O_k = \{\deg_k(x) \le 8\eta\log n,\ \forall x \in [n]\}$. With the same argument as for the upper bound in the proof of Lemma 5.7 (see Appendix B), we see that $\mathbf{P}(O_k^c) \le n^{-\rho-2}$ for some $\rho > 2+\beta$. Therefore

$$\begin{aligned}\mathbf{P}(F_{l,t}^c, C_t) &= \mathbf{P}\Big(F_{l,t}^c\ \Big|C_t, \bigcap_{k=0}^{t-1} O_k\Big)\mathbf{P}\Big(C_t, \bigcap_{k=0}^{t-1} O_k\Big) + \mathbf{P}\Big(F_{l,t}^c, C_t, \bigcup_{k=0}^{t-1} O_k^c\Big) \\ &\le \mathbf{P}\Big(F_{l,t}^c\ \Big|C_t, \bigcap_{k=0}^{t-1} O_k\Big) + n^{-\rho}.\end{aligned}$$

We now just consider

$$\mathbf{P}\Big(F_{l,t}^c\ \Big|C_t, \bigcap_{k=0}^{t-1} O_k\Big).$$

For $t > n^{1.1}$, the condition

$$C_t \cap \Big(\bigcap_{k=0}^{t-1} O_k\Big)$$

implies that the graph has been connected for at least the last $n^{1.1}$ time steps, with all degrees bounded by $8\eta\log n$. In such a graph, if $x \sim_s y$, then $P^{s,s+1}(x,y) \ge \frac{1}{16\eta\log n}$.

We say that there is a path from x to y of length k, starting at time s, if there exist $x_1,\dots,x_{k-1} \in [n]$, such that $x \sim_{s+1} x_1$, $x_1 \sim_{s+2} x_2, \dots, x_{k-1} \sim_{s+k} y$ and we write this path as

$$[x,y]_s^{s+k} = (x, x_1, \dots, x_{k-1}, y)_s^{s+k}.$$

We allow vertices on the path to be equal, that is to say $x_i = x_j$ for some $i \neq j$.

If there exists a path $[x, y]_s^{s+n-1}$, then

$$P^{s,s+n-1}(x,y) \geq \frac{1}{(16\eta \log n)^{n-1}}. \tag{5.30}$$

If we can show that (5.30) holds for all x, y on $C_t \cap \big(\bigcap_{k=0}^{t-1} O_k\big)$, then

$$\pi_t(x) = \sum_y \pi_{t-n+1}(y) P^{t-n+1,t}(y,x) \geq \frac{1}{(16\eta \log n)^{n-1}} \overset{(5.7)}{=} \frac{\alpha_1^{(t)}}{n}$$

yields that

$$\mathbf{P}\Big(F_{l,t}^c \,\Big|\, C_t, \bigcap_{k=0}^{t-1} O_k\Big) = 0.$$

So it remains to show that there is indeed a path from x to y for any two vertices $x, y \in [n]$. Consider the following set process (for simplicity we start the process at time $k = 0$, but everything can be shifted accordingly): Let $T_x^0 \coloneqq \{x\}$,

$$T_x^k \coloneqq \ \{v \in [n] : \exists z \in T_x^{k-1} \text{ s.t. } v \sim_k z\} \cup T_x^{k-1}$$

that is to say given a set of vertices T_x^{k-1}, T_x^k is the set of all vertices that are neighbors of vertices in T_x^{k-1} at time k. This makes T_x^k the set of all vertices that are reachable from x in k steps. Since the graph is connected, there exists one edge that connects T_x^k with its complement $(T_x^k)^c$ at time $k+1$ if the complement is not empty. In particular, this either implies $|T_x^{k+1}| > |T_x^k|$ or the complement is the empty set, that is to say $T_x^k = [n]$. Thus $T_x^{n-1} = [n]$, or equivalently, after $n-1$ steps the entire graph is reachable. This proves the existence of a path from x to y of length $n-1$ and completes the proof. □

The proof of Proposition 5.5 is now a mere application of the lemmas above.

Proof of Proposition 5.5: Inserting the Lemmas 5.7–5.11 into the inequalities (5.10)–(5.12) shows equation (5.5). □

5.2. *Lower bound on* Θ_t. In this section, we show that there exists a constant $\kappa > 0$, independent of n, such that $\Theta_t \geq \kappa$ for every $t \in \{0, \dots, n\}$ with high probability. This is the final step in the proof of Theorem 5.1, which is then immediate from Corollary 4.3. We recall $c_1 = \frac{11}{21}\eta$, $c_2 = 2\eta$, $\alpha_1^* = 0.002$, and $\alpha_2^* = 7$.

Proposition 5.12. *It holds that*

$$\lim_{n\to\infty} \mathbf{P}\left(\Theta_t \geq \left(\frac{\alpha_1^*}{2\alpha_2^* - \alpha_1^*} \cdot \frac{\alpha_1^*}{32(\alpha_2^*)^2}\right)^2, \forall t \in \{0, \dots, n\}\right) = 1. \tag{5.31}$$

Proof: The proof adapts a strategy presented in Durrett (2007, Chapter 6). Only the case of static graphs is considered there, but our previous bounds on the degrees and on π_t are strong enough to transfer the strategy to a dynamic situation.

In this proof, we bound Φ_t^* and g_t from below. Let

$$\mathcal{E} = \Big\{\frac{\alpha_1^*}{n} \leq \pi_t(x) \leq \frac{\alpha_2^*}{n}, \forall t \in \{0, \dots, n\}, \forall x \in [n]\Big\}$$

be the event that π_t is nearly uniform and recall from Proposition 5.5 that $\mathbf{P}(\mathcal{E}) \geq 1 - n^{-\beta}$ for some constant $\beta > 0$. Observe that for any event A that is deterministic on $\mathcal{E}$, that is to say $\mathbf{P}(A|\mathcal{E}) = 1$, we also have

$$\mathbf{P}(A) \geq \mathbf{P}(A|\mathcal{E})\mathbf{P}(\mathcal{E}) \geq 1 - n^{-\beta}. \tag{5.32}$$

By definition for every set $S \subset [n]$, $Q_t(S^c, S) \geq 0$, so on the set $\mathcal{E}$, we have

$$\Phi_t(S) \geq \frac{Q_t(S,S^c)}{2\pi_{t-1}(S)} \geq \frac{Q_t(S,S^c)}{2\alpha_2^*|S|}n. \tag{5.33}$$

In particular (5.33) holds with probability $1 - n^{-\beta}$ for all $t \in \{1, \dots, n\}$.

We now find a lower bound for $Q_t(S, S^c)$. Let $e_t(x, S^c) = \sum_{y \in S^c} \mathbb{1}_{y \sim_t x}$ be the number of edges from x into S^c at time t, and $e_t(S, S^c) = \sum_{x \in S} e_t(x, S^c)$. In analogy to (5.32),

$$\mathbf{P}(\cap_{t=1}^n D_t \cap \mathcal{E}) \geq 1 - 2n^{-\beta},$$

and therefore an event that is deterministic on $\cap_{t=1}^n D_t \cap \mathcal{E}$ will also have a probability greater than $1 - 2n^{-\beta}$. We henceforth write $\forall t$ meaning for all $t \in \{1, \dots, n\}$. For every $S \subset [n]$ and $\forall t$, it is deterministic on $\cap_{t=1}^n D_t \cap \mathcal{E}$ that

$$\begin{aligned} Q_t(S,S^c) &= \sum_{x\in S}\sum_{y\in S^c}\pi_{t-1}(x)P_t(x,y) = \sum_{x\in S}\pi_{t-1}(x)\sum_{\substack{y\in S^c\\ y\sim_t x}}\frac{1}{2\deg_t(x)} \\ &\geq \sum_{x\in S}\pi_{t-1}(x)\frac{1}{2c_2\log n}e_t(x,S^c) \\ &\geq \frac{1}{2c_2\log n}\frac{\alpha_1^*}{n}\sum_{x\in S}e_t(x,S^c) = \frac{1}{2c_2\log n}\frac{\alpha_1^*}{n}e_t(S,S^c) =: m_t(S). \end{aligned}$$

So in particular $\mathbf{P}(Q_t(S,S^c) \geq m_t(S), \forall S \subset [n], \forall t) \geq 1 - 2n^{-\beta}$.

Let us now consider $m_t(S)$. To find a lower bound, we bound $e_t(S, S^c)$. By definition of Φ_t^* we need only consider $S \subset [n]$ with $\pi_{t-1}(S) \leq 1/2$. We claim that, on $\mathcal{E}$, $\pi_{t-1}(S) \leq 1/2$ implies $|S| \leq \big(1 - \frac{1}{2\alpha_2^*}\big)n$. To that end, it suffices to observe that on $\mathcal{E}$, for any $S \subset [n]$,

$$\pi_{t-1}(S) \leq |S|\frac{\alpha_2^*}{n}.$$

If $|S^c| < \frac{n}{2\alpha_2^*}$, then $\pi_{t-1}(S^c) < 1/2$ and hence $\pi_{t-1}(S) > 1/2$. Therefore it suffices to consider $S \subset [n]$ such that $|S| \leq \big(1 - \frac{1}{2\alpha_2^*}\big)n$.

Let

$$B \ := \ \left\{S : \frac{n}{\eta\log n} \leq |S| \leq \left(1 - \frac{1}{2\alpha_2^*}\right)n\right\}.$$

We consider all $S \in B$ such that $|S| = s$ for some $\frac{n}{\eta \log n} \leq s \leq (1 - \frac{1}{2\alpha_2^*})n$. Note that $e_t(S, S^c) \sim \mathrm{Bin}(s(n-s), p_n)$ for each s. We observe that, on $\mathcal{E}$,

$$\frac{s(n-s)p}{2} \geq \frac{s\eta\log n}{4\alpha_2^*}.$$

There are $\binom{n}{s}$ sets $S \in B$ of size s, thus, using classical tail bounds for the binomial distribution (e.g. Durrett (2007, Lemma 2.8.5) with $z = p_n/2$),

$$\mathbf{P}\Big(\exists S \in B \text{ with } |S| = s, e_t(S,S^c) \leq \frac{s(n-s)p_n}{2}\Big) \leq \binom{n}{s}\exp\Big(-s(n-s)\frac{\eta\log n}{8(n-1)}\Big) \tag{5.34}$$

and since $n - s \geq \frac{1}{2\alpha_2^*}n \geq \frac{1}{2\alpha_2^*}(n-1)$ we can use $\binom{n}{s} \leq \frac{n^s}{s!} \leq n^s s^{-s} e^s$ to arrive at an upper bound of

$$\begin{aligned} (5.34) &\leq \exp\Big(-s\Big[\frac{\eta\log n}{16\alpha_2^*} + \log(s/n) - 1\Big]\Big) \\ &\leq \exp\Big(-\frac{n}{\eta\log n}\Big[\frac{\eta\log n}{16\alpha_2^*} - \log(\eta\log n) - 1\Big]\Big), \end{aligned}$$

where we used twice in the second line that $s \geq \frac{n}{\eta \log n}$. This goes to zero exponentially fast, in particular faster than $n^{-\beta-2}$. Since it holds for every s, a union bound yields that

$$\mathbf{P}\Big(e_t(S,S^c) \geq \frac{|S|\eta \log n}{4\alpha_2^*}, \forall t, \forall S \in B\Big) \geq 1 - 2n^{-\beta}.$$

Combining this with the previous bound on $Q_t(S,S)$, $c_2 = 2\eta$, and with (5.33) implies

$$\mathbf{P}\Big(\Phi_t(S) \geq \frac{\alpha_1^*}{32(\alpha_2^*)^2}, \forall t, \forall S \in B\Big) \geq 1 - 5n^{-\beta}.$$

It remains to show a similar bound for $e_t(S,S^c)$ for every $S \in A = \Big\{S : 1 \leq |S| \leq \frac{n-1}{\eta \log n}\Big\}$. On the event D_t, we observe that for every $S \subset [n]$

$$e_t(S,S^c) = \sum_{x \in S} \deg_t(x) - e_t(S,S) \geq |S| \cdot c_1 \log n - e_t(S,S).$$

So in order to arrive at a lower bound for $e_t(S,S^c)$, we prove an upper bound on $e_t(S,S)$.

A priori, note that for $|S| = s \leq \frac{n-1}{\eta \log n}$, we have $\mathbf{E}[e_t(S,S)] \leq \frac{s^2}{2} p_n \leq \frac{s}{2}$. As in the case $S \in B$, we use standard tail bounds for the binomial distribution to estimate

$$\begin{aligned}
\mathbf{P}(\exists S \in A \text{ with } |S| = s, e_t(S,S) \geq s \log\log n) &\leq C \binom{n}{s} p_n^{s \log\log n} \binom{s^2/2}{s \log\log n} \\
&\leq C \Big(\frac{ne}{s}\Big)^s \frac{(s^2/2)^{s \log\log n} p_n^{s \log\log n}}{(s \log\log n)^{s \log\log n} e^{s \log\log n}} \\
&= C \exp(q(s))
\end{aligned} \tag{5.35}$$

for $q(s) := s \log(ne/s) + s \log\log n \Big[\log s + \log\Big(\frac{\eta \log ne}{2(n-1) \log\log n}\Big)\Big]$. We show that $q(s)$ is maximized when $s = 1$. To prove this, we take the derivative with respect to s and find

$$q'(s) = \log(n) - \log(s) + \log\log n \big[\log s + \log(e\eta \log n) - \log\big(2(n-1)\log\log n\big)\big] + \log\log n.$$

Differentiating with respect to s once more yields

$$q''(s) = \frac{\log\log n - 1}{s} > 0,$$

so $q'(s)$ is increasing in s. In particular, we show that for the maximal $s = \frac{n-1}{\eta \log n}$ we have $q'(s) < 0$ which then implies that $q(s)$ is decreasing in s, hence $q(1)$ is the maximum on $1 \leq s \leq \frac{n-1}{\eta \log n}$. We compute

$$\begin{aligned}
q'\Big(\frac{n-1}{\eta \log n}\Big) &= \log\Big(\frac{n}{n-1}\Big) + \log(\eta \log n) + \log\log n \cdot \big(2 - \log(2) - \log\log\log n\big) \\
&= \log\Big(\frac{n}{n-1}\Big) + \log(\eta) + \log\log n \cdot \big(3 - \log(2) - \log\log\log n\big),
\end{aligned}$$

which is smaller than 0 for n large enough. Thus $q(s) \leq q(1)$ for every s and hence

$$\begin{aligned}
q(s) &\leq \log(ne) + \log\log n \cdot \log\Big(\frac{\eta \log ne}{2 \log\log n(n-1)}\Big) \\
&= 1 + \log n + \log\log n \cdot \Big[1 + \log(\eta) + \log\log n - \log(2(n-1)) - \log\log\log n\Big] \\
&\leq -\log n \cdot (\log\log n - 1) + 1 + \log\log n \cdot \big[1 + \log(\eta) + \log\log n\big] \\
&\leq -\log n \cdot (\log\log n - 2)
\end{aligned}$$

for n large enough, so ultimately

$$C\exp(q(s)) \le Cn^{2-\log\log n} \le n^{-\beta-2}$$

for n large enough.

In particular, using a union bound for all $s \le \frac{n-1}{\eta\log n}$ on (5.35), results in the lower bound on the probability

$$\mathbf{P}(e_t(S,S) \le |S|\log\log n, \forall t, \forall S \in A) \ge 1 - n^{-\beta-1}.$$

Since $e_t(S,S^c) \ge |S|c_1\log n - e_t(S,S)$, it follows that

$$e_t(S,S^c) \ge |S|(c_1\log n - \log\log n) \ge |S|\frac{c_1\log n}{2}.$$

Combining all of the above yields

$$\mathbf{P}\Big(\Phi_t(S) \ge \frac{\alpha_1^* c_1}{8\alpha_2^* c_2}, \forall t, \forall S \in A\Big) \ge 1 - 5n^{-\beta}.$$

Observing that, for our particular choice of constants,

$$\frac{\alpha_1^* c_1}{8\alpha_2^* c_2} > \frac{\alpha_1^*}{32(\alpha_2^*)^2}$$

that is to say $\Phi_t(S)$ can be smaller for $S \in B$ than for $S \in A$, it follows that

$$\mathbf{P}\Big(\Phi_t^* \ge \frac{\alpha_1^*}{32(\alpha_2^*)^2}, \forall t\Big) \ge 1 - 10n^{-\beta}.$$

Furthermore on $\mathcal{E}$, it holds that $\sqrt{\pi_0^{\min}\pi_t^{\min}} \ge \frac{\alpha_1^*}{n}$ and $\left(\frac{1}{2}\frac{g_s}{1-\frac{1}{2}g_s}\right)^2 \ge \left(\frac{\alpha_1^*}{2\alpha_2^*-\alpha_1^*}\right)^2$. Hence

$$\Theta_t = \min_{1\le s\le t}\left(\frac{1}{2}\frac{g_s}{1-\frac{1}{2}g_s}\Phi_s^*\right)^2 \ge \left(\frac{\alpha_1^*}{2\alpha_2^*-\alpha_1^*}\cdot\frac{\alpha_1^*}{32(\alpha_2^*)^2}\right)^2$$

with high probability and (5.31) follows. □

Theorem 5.1 is an easy consequence.

Proof of Theorem 5.1: From Proposition 5.5 giving a lower bound on $\pi_t^{\min}$ and Proposition 5.12 providing a lower bound on Θ_t, the assumptions of Corollary 4.3 are satisfied, which implies (5.1). □

5.3. *A lower bound on mixing time.* We already know from Proposition 5.5 that $\pi_t(x) \in [\alpha_1^*/n, \alpha_2^*/n]$ for all $x \in [n]$, $t \in \{0,...,n\}$ w.h.p., which also allows us to employ a classical argument for a lower bound on the mixing time. Namely, a random walk cannot mix if a substantial section of the graph is not accessible. This also shows that the intuition (see Remark 5.3) that the walk should mix in just a few steps is wrong, as long as p_n is not too large.

Theorem 5.13. *Let $p_n = \frac{\eta\log n}{n-1}$ with $\eta > 50$. There exists $c > 0$ such that for every $\varepsilon \in (0,1/2)$*

$$t_{\text{mix}}(\varepsilon,0) \ge c\frac{\log n}{\log\log n}$$

with high probability.

Proof: As in the proof of Proposition 5.12, we note: If

$$|S| > \left(1 - \frac{1}{2\alpha_2^*}\right)n$$

then (w.h.p.)

$$\pi_t(S) > \frac{1}{2}, \qquad \text{for all } 0 \leq t \leq n.$$

In particular, if only vertices in S^c are reachable within t steps from x, then

$$\|P^{0,t}(x,\cdot) - \pi_t(\cdot)\|_{\mathrm{TV}} \geq 1 - \pi_t(S^c) = \pi_t(S) > \frac{1}{2}.$$

Let T_x^k be the set of vertices reachable from x in k steps starting at time 0, in the sense that $T_x^k = \{y \in [n] : \text{there exists a path } [x,y]_0^k\}$ (see the proof of Lemma 5.11). Since every vertex (w.h.p.) has at most $c_2 \log n$ neighbors,

$$|T_x^1| \leq c_2 \log n + 1.$$

Even if all those vertices get an entirely new set of neighbors, the maximum number of reachable vertices after two steps is still bounded by $|T_x^2| \leq (c_2 \log n + 1)^2$. Iteratively $|T_x^k| \leq (c_2 \log n + 1)^k$. For k small enough, this implies that the random walk is confined to a small set of vertices. In particular, if

$$k \leq \frac{\log\left(\frac{1}{2\alpha_2^*} n\right)}{\log(1 + c_2 \log n)},$$

then $|T_x^k| \leq n/(2\alpha_2^*)$ and thus $d(0,t) > 1/2$ with high probability. We can choose $c > 0$ such that

$$t_{\mathrm{mix}}(\varepsilon, 0) \geq c \frac{\log n}{\log \log n}$$

with high probability. □

5.4. *Possible Extensions.* There is potential to make a few significant extensions to the results in this paper, which greatly vary in difficulty.

First off, we note that the condition $\eta > 50$ in Theorem 5.1 is not optimal, and more careful choice of constants in the proofs of this section can relax this condition a bit. However, in order to get close to 1, one probably needs an improved argument for a lower bound on $\pi_t^{\min}$. In particular, we argue above that if randomly sampled ER graphs satisfy certain conditions for $n^{1.1}$ time steps in a row, then $\pi_t^{\min} \geq \frac{1}{(c \log n)^{n-1}}$ (see (5.30) and the discussion surrounding it). This is an "extreme" estimate for a lower bound on π_t and in fact even with deliberately adversarial choices for a sequence of transition matrices we found it difficult to explicitly construct a sequence where $\pi_t^{\min} < \frac{1}{n^2}$; and it would be very unlikely for ER graphs to randomly produce exactly such an adversarial counterexample. Being able to improve this worst case lower bound that significantly would decrease the number of time steps our iterative arguments in Lemma 5.10 and Lemma 5.11 need to hold for, which in turn would allow us to relax our condition on η.

Secondly, as we have mentioned before, the assumption that the ER graphs sampled in this section are independent from each other is only made to prove Proposition 5.5. At a few points in the proof, we use that π_t and the state of edges $\mathbb{1}_{x \sim_{t+1} y}$ are independent from each other. This is obviously not true if we do not assume independence between graphs, so e.g. for the classical dynamical percolation model (see Sousi and Thomas, 2020). We feel that the correlation may be weak enough so our techniques could still work, but so far we have not been able to resolve the issues.

Finally, we want to make a remark on the statement of Theorem 3.2. This theorem uses the time-dependent bottleneck ratio Φ_t^*. By definition, the bottleneck ratio of a disconnected state space is always 0. So the statement of this theorem limits us to computing bounds on what are essentially time-inhomogeneous irreducible Markov chains. However, there are Markov chains that, while they are not irreducible, satisfy a kind of dynamic connectivity in the spirit of Remark 4.2. If we observe a Markov chain e.g. only every $\log \log n$ time steps, it is possible that the disconnected states of the chain disappear on this larger time-scale and the chain behaves like an irreducible chain for the purpose of Theorem 3.2. Then an argument similar to that of Theorem 3.2 should work, and due to the dynamic connectivity on this time-scale, the "scaled" bottleneck ratios are nonzero and we would obtain an upper bound for the mixing time of this scaled Markov chain. By paying a factor of $\log\log n$ and maybe some constant, this argument would then yield an upper bound for the mixing time of the original chain. For concrete applications to examples, this obviously raises the question on how to compute products of transition matrices, which is also non-trivial.

Appendix A. Existence proofs

In order to prove Lemma 2.4, we need the following simple observation on total variation.

Lemma A.1. *Let*

$$\hat{d}(s,t) \coloneqq \sup_{\mu,\nu} \|\mu P^{s,t} - \nu P^{s,t}\|_{\mathrm{TV}}$$

where the supremum is taken over all probability measures on $[n]$*. Then the supremum is attained, and in fact it holds that*

$$\hat{d}(s,t) = \delta(P^{s,t}) = \sup_{x,y} \|P^{s,t}(x,\cdot) - P^{s,t}(y,\cdot)\|_{\mathrm{TV}}.$$

Proof: See Levin and Peres (2017, Exercise 4.1) for the statement, and Appendix D therein for the proof. □

Proof of Lemma 2.4: Let $\varepsilon > 0$. By assumption, there exists $u < t$, such that $\delta(P^{u,t}) \le \varepsilon/2$. Choose any $s, r \in \mathbb{Z}$ with $s < r < u$. Then, denoting $\mu_x = P^{s,u}(x,\cdot)$ and $\nu_y = P^{r,u}(y,\cdot)$,

$$\begin{aligned}
\sup_{x,y} \|P^{s,t}(x,\cdot) - P^{r,t}(y,\cdot)\|_{\mathrm{TV}} &= \sup_{x,y} \|(P^{s,u}P^{u,t})(x,\cdot) - (P^{r,u}P^{u,t})(y,\cdot)\|_{\mathrm{TV}} \\
&= \sup_{x,y} \|\mu_x P^{u,t} - \nu_y P^{u,t}\|_{\mathrm{TV}} \\
&\le \sup_{x,y} \|P^{u,t}(x,\cdot) - P^{u,t}(y,\cdot)\|_{\mathrm{TV}} \\
&= \delta(P^{u,t}) \le \varepsilon/2.
\end{aligned}$$

where the inequality in the third line follows from Lemma A.1. In particular, for any $x, y \in [n]$ we have

$$\sum_z |P^{s,t}(x,z) - P^{r,t}(y,z)| \le \varepsilon,$$

so the matrix products $(P^{s,t})_{s\le t}$ are a Cauchy sequence as $s \to -\infty$. By completeness of the space of matrices, the Cauchy sequence converges and there exists a matrix Q^t with $\lim_{s\to-\infty} P^{s,t}(x,y) = Q^t(x,y)$. Moreover $\delta(Q^t) = 0$, so it is a rank 1 matrix. □

We now show the technical results of Lemma 2.7, that $d(s,t)$ is decreasing in t and that it is closely related to $\delta(P^{s,t})$.

Proof of Lemma 2.7: To show (a), simply consider

$$\begin{aligned}
\sup_x \|P^{u,t}(x,\cdot) - \pi_t(\cdot)\|_{\mathrm{TV}} &= \sup_x \|(P^{u,s}P^{s,t})(x,\cdot) - (\pi_s P^{s,t})(\cdot)\|_{\mathrm{TV}} \\
&\le \sup_x \|P^{u,s}(x,\cdot) - \pi_s(\cdot)\|_{\mathrm{TV}}
\end{aligned}$$

by the definition of the total variation distance and applying the triangle inequality. So $d(u,s) \le d(u,t)$ for all $u \le s \le t$.

(b) is shown by using the same arguments as in Lemma A.1 (see Levin and Peres, 2017, Exercise 4.1, Appendix D). We omit the details here because the statement is not used in the present paper.

To show (c), remark that $d(s,t) \le \delta(P^{s,t})$ (see again Lemma A.1). To prove $\delta(P^{s,t}) \le 2d(s,t)$, write

$$\begin{aligned}\sup_{x,y}\sum_z |P^{s,t}(x,z) - P^{s,t}(y,z)| &= \sup_{x,y}\sum_z |P^{s,t}(x,z) - \pi_t(z) + \pi_t(z) - P^{s,t}(y,z)| \\ &\le \sup_x \sum_z |P^{s,t}(x,z) - \pi_t(z)| + \sup_y \sum_z |P^{s,t}(y,z) - \pi_t(z)|\end{aligned}$$

which completes the proof. □

Appendix B. Erdős-Rényi graphs

In this section, we first show Lemma 5.7, giving degree bounds for connected Erdős-Rényi graphs. Similar results are standard (cf. Durrett, 2007, Lemma 6.5.2), but here we compute an explicit decay rate of the right-hand side for our choice of parameters.

Proof of Lemma 5.7: Note that $\deg_t(x) \sim \mathrm{Bin}(n-1, p_n)$. For any $\lambda > 0$, $c_2 > \eta$

$$\begin{aligned}\mathbf{P}(\deg_t(x) \ge c_2 \log n) &\le e^{\log(1-p_n+p_n e^{\lambda})(n-1)} n^{-\lambda c_2} \le e^{(-p_n+p_n e^{\lambda})(n-1)} n^{-\lambda c_2} \\ &= n^{-\eta+\eta e^{\lambda}-\lambda c_2}.\end{aligned}$$

We can minimize this expression by choosing $\lambda = \log(\frac{c_2}{\eta})$ to arrive at

$$\mathbf{P}(\deg_t(x) \ge c_2 \log n) \le n^{-\eta + c_2(1-\log(\frac{c_2}{\eta}))}.$$

Similarly, for $0 < c_1 < \eta$, we can establish the lower bound by

$$\mathbf{P}(\deg_t(x) \le c_1 \log n) \le n^{-\eta+\eta e^{-\lambda}+\lambda c_1},$$

which we can optimize with $\lambda = \log(\frac{\eta}{c_1})$ to arrive at

$$\mathbf{P}(\deg_t(x) \le c_1 \log n) \le n^{-\eta + c_1(1+\log(\frac{\eta}{c_1}))}.$$

Since $c_1 = \frac{11}{21}\eta$, $c_2 = 2\eta$, and $\eta > 50$, applying a union bound over all n vertices yields the desired result. □

We now show Lemma 5.8, that n^2 independent Erdős-Rényi graphs are all connected with high probability and again compute the decay rate.

Proof of Lemma 5.8: Proofs of connectivity are standard (cf. Durrett, 2007, Theorem 2.8.1). The argument presented here specifically uses that the graphs are far above the connectivity threshold (recall $\eta > 50$ by assumption) to get a quantitative bound on the probability.

We consider the connectivity at a time $k \in \{0, \dots, n^2-1\}$. Let K be the number of ways the graph can be divided at time k into two non-empty subgraphs that are disconnected. Then clearly $\mathbf{P}(\text{Graph not connected}) = \mathbf{P}(K \ge 1) \le \mathbf{E}[K]$. We estimate

$$\begin{aligned}\mathbf{E}[K] &= \sum_{i=1}^{n/2} (1-p_n)^{i(n-i)} \binom{n}{i} \le \sum_{i=1}^{n/2} e^{-p_n i(n-i)} n^i \\ &= \sum_{i=1}^{n/2} e^{-p_n i(n-i) + i\log n} = \sum_{i=1}^{n/2} e^{(-\frac{\eta}{n-1} i(n-i)+i)\log n} \\ &\le \sum_{i=1}^{n/2} e^{(-\eta i(\frac{n}{n-1} - \frac{n}{2(n-1)})+i)\log n} = \sum_{i=1}^{n/2} n^{-i(\frac{\eta}{2}\frac{n}{n-1}-1)} \\ &\le n^{-\frac{\eta}{2}+2}.\end{aligned}$$

Union bound for all n^2 times yields the result. □

Acknowledgements

The author would like to thank an anonymous referee for a very careful and detailed reading of a preliminary version of the manuscript, as well as their constructive feedback and many helpful suggestions.

References

Avena, L., Güldaş, H., van der Hofstad, R., and den Hollander, F. Mixing times of random walks on dynamic configuration models. *Ann. Appl. Probab.*, **28** (4), 1977–2002 (2018). MR3843821.

Avena, L., Güldaş, H., van der Hofstad, R., and den Hollander, F. Random walks on dynamic configuration models: a trichotomy. *Stochastic Process. Appl.*, **129** (9), 3360–3375 (2019). MR3985565.

Avin, C., Koucký, M., and Lotker, Z. Cover time and mixing time of random walks on dynamic graphs. *Random Structures Algorithms*, **52** (4), 576–596 (2018). MR3809689.

Brémaud, P. *Markov chains—Gibbs fields, Monte Carlo simulation and queues*, volume 31 of *Texts in Applied Mathematics*. Springer, Cham, second edition (2020). ISBN 978-3-030-45981-9; 978-3-030-45982-6. MR4174390.

Cai, L., Sauerwald, T., and Zanetti, L. Random walks on randomly evolving graphs. In *Structural information and communication complexity*, volume 12156 of *Lecture Notes in Comput. Sci.*, pp. 111–128. Springer, Cham (2020). MR4139772.

Dembo, A., Huang, R., Morris, B., and Peres, Y. Transience in growing subgraphs via evolving sets. *Ann. Inst. Henri Poincaré Probab. Stat.*, **53** (3), 1164–1180 (2017). MR3689964.

Dobrushin, R. L. Central limit theorem for nonstationary Markov chains. I. *Theory Probab. Its Appl.*, **1** (1), 65–80 (1956). DOI: 10.1137/1101006.

Durrett, R. *Random graph dynamics*, volume 20 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge (2007). ISBN 978-0-521-86656-9; 0-521-86656-1. MR2271734.

Hermon, J. and Sousi, P. A comparison principle for random walk on dynamical percolation. *Ann. Probab.*, **48** (6), 2952–2987 (2020). MR4164458.

Jerrum, M. and Sinclair, A. Approximating the permanent. *SIAM J. Comput.*, **18** (6), 1149–1178 (1989). MR1025467.

Lelli, A. and Stauffer, A. Mixing time of random walk on dynamical random cluster. *Probab. Theory Related Fields*, **189** (3-4), 981–1043 (2024). MR4771109.

Levin, D. A. and Peres, Y. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, second edition (2017). ISBN 978-1-4704-2962-1. MR3726904.

Mans, B. and Pourmiri, A. Asynchronous rumor spreading in dynamic graphs. In *25th International Conference on Principles of Distributed Systems*, volume 217 of *LIPIcs. Leibniz Int. Proc. Inform.*, pp. Art. No. 31, 20. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern (2022). MR4390266.

Morris, B. and Peres, Y. Evolving sets, mixing and heat kernel bounds. *Probab. Theory Related Fields*, **133** (2), 245–266 (2005). MR2198701.

Moumeni, N. Quantitative merging for time-inhomogeneous Markov chains in non-decreasing environments via functional inequalities. *ArXiv Mathematics e-prints* (2024). arXiv: 2404.11432.

Paz, A. and Reichaw, M. Ergodic theorems for sequences of infinite stochastic matrices. *Proc. Cambridge Philos. Soc.*, **63**, 777–784 (1967). MR214141.

Peres, Y., Sousi, P., and Steif, J. E. Quenched exit times for random walk on dynamical percolation. *Markov Process. Related Fields*, **24** (5), 715–731 (2018). MR4246022.

Peres, Y., Sousi, P., and Steif, J. E. Mixing time for random walk on supercritical dynamical percolation. *Probab. Theory Related Fields*, **176** (3-4), 809–849 (2020). MR4087484.

Peres, Y., Stauffer, A., and Steif, J. E. Random walks on dynamical percolation: mixing times, mean squared displacement and hitting times. *Probab. Theory Related Fields*, **162** (3-4), 487–530 (2015). MR3383336.

Raghavan, P. Probabilistic construction of deterministic algorithms: approximating packing integer programs. *J. Comput. System Sci.*, **37** (2), 130–143 (1988). MR979115.

Saloff-Coste, L. and Zúñiga, J. Merging for time inhomogeneous finite Markov chains. I. Singular values and stability. *Electron. J. Probab.*, **14**, 1456–1494 (2009). MR2519527.

Sauerwald, T. and Zanetti, L. Random walks on dynamic graphs: mixing times, hitting times, and return probabilities. In *46th International Colloquium on Automata, Languages, and Programming*, volume 132 of *LIPIcs. Leibniz Int. Proc. Inform.*, pp. Art. No. 93, 15. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern (2019). MR3984910.

Sousi, P. and Thomas, S. Cutoff for random walk on dynamical Erdős-Rényi graph. *Ann. Inst. Henri Poincaré Probab. Stat.*, **56** (4), 2745–2773 (2020). MR4164855.