Random walks on randomly evolving graphs

Leran Cai  Thomas Sauerwald  Luca Zanetti

June 19, 2020

Abstract

A random walk is a basic stochastic process on graphs and a key primitive in the design of distributed algorithms. One of the most important features of random walks is that, under mild conditions, they converge to a stationary distribution in time that is at most polynomial in the size of the graph. This fundamental property, however, only holds if the graph does not change over time, while on the other hand many distributed networks are inherently dynamic, and their topology is subjected to potentially drastic changes.

In this work we study the mixing (i.e., converging) properties of random walks on graphs subjected to random changes over time. Specifically, we consider the edge-Markovian random graph model: for each edge slot, there is a two-state Markov chain with transition probabilities \( p \) (add a non-existing edge) and \( q \) (remove an existing edge). We derive several positive and negative results that depend on both the density of the graph and the speed by which the graph changes. We show that if \( p \) is very small (i.e., below the connectivity threshold of Erdős-Rényi random graphs), random walks do not mix (fast). When \( p \) is larger, instead, we observe the following behavior: if the graph changes slowly over time (i.e., \( q \) is small), random walks enjoy strong mixing properties that are comparable to the ones possessed by random walks on static graphs; however, if the graph changes too fast (i.e., \( q \) is large), only coarse mixing properties are preserved.

1 Introduction

A random walk on a network is a simple stochastic process, defined as follows. Given an undirected graph \( G = (V, E) \), the walk starts at a fixed vertex. Then at each step, the random walk moves to a randomly chosen neighbor \(^1\). Due to their simplicity and locality, random walks are very useful algorithmic primitive, especially in the design of distributed algorithms. In contrast to topology-driven algorithms, algorithms based on random walks benefit from a strong robustness against structural changes in the network.

Random walks and related works have found various applications such as routing, information spreading, opinion dynamics and graph exploration \(^8\). One key property of random walks is that, under mild assumptions on the underlying network, they converge to a stationary distribution – an equilibrium state in which every vertex is visited proportionally to its degree. The time for this convergence to happen is called mixing time, and understanding this time is crucial for many sampling or exploration related tasks. In particular, whenever a graph has a small mixing time, also its cover time (the expected time to visit all vertices of the graph) is small as well.

---

\(^*\)The second and third author acknowledge support by the ERC Starting Grant “Dynamic March”.

\(^1\)In case of a lazy random walk, the walk would remain at the current location with probability \( 1/2 \), and otherwise move to a neighbor chosen uniformly at random.
While most of the classical work devoted to understanding random walks has focused on static graphs, many networks today are subject to dramatic changes over time. Hence understanding the theoretical power and limitations of dynamic graphs has been identified as one of the key challenges in computer science [18]. Several recent works have indeed considered this problem and studied the behavior of random walks [2, 3, 11, 16, 22, 24, 25] or similar processes [4, 6, 9, 12, 15] on such dynamic graphs, and their applications to distributed computing [2, 15, 25]. Moreover, rather than a property of the underlying network itself, dynamic graphs may naturally arise in distributed algorithms when communication is performed on a changing, possibly disconnected, subgraph like a spanning-tree or a matching (see, e.g., [3]).

In this work, we study the popular evolving graph model. That is, we consider sequences of graphs $G_1, G_2, \ldots$ over the same set of vertices but with a varying set of edges. This model has been studied in, for example, [3, 16, 26]. Both [3] and later [26] proved a collection of positive and negative results about the mixing time (and related parameters), and they assume a worst-case scenario where the changes to the graph are dictated by an oblivious, non-adaptive adversary. For example, [3] proved the following remarkable dichotomy. First, even if all graphs $G_1, G_2, \ldots$ are connected, small (but adversarial) changes to the stationary distribution can cause exponential mixing (and hitting) times. Secondly, if the sequence of connected graphs share the same stationary distribution i.e., the degrees (or relative degrees) of vertices are time-invariant, then mixing and hitting times are polynomial. This assumption about a time-invariant stationary distribution is crucial in the majority of the positive results in [3, 26]. In general, in order to compensate for the adversarial changes, both works had to impose assumptions about the existence of a time-invariant stationary distribution, which means that the degree (or relative degree) of a vertex remains the same at each step. Both [3] and later [26] proved a collection of positive and negative results about the mixing time (and related parameters), assuming that the graph changes arbitrarily (i.e., by an oblivious, non-adaptive adversary). However, to compensate for the fact that the graph changes in an arbitrary way, both works had to impose assumptions about connectivity and the existence of a time-invariant stationary distribution, which means that the degree (or relative degree) of a vertex remains the same at each step.

In contrast to [3, 26], we do not impose such assumptions, but instead study a model with incremental changes. Specifically, we consider a setting where the evolving graph model changes randomly and study the so-called edge-Markovian random graph $G(n, p, q)$, which is defined as follows (see Definition 2.5 for a more formal description). For each edge slot, there is a two-state Markov chain that switches from off to on with probability $p$ and from on to off with probability $q$. This model can be seen as a dynamic version of the Erdős-Rényi random graph, and has been studied in the context of information spreading and flooding [6–8]. While these results demonstrate that information disseminates very quickly on these dynamic graphs, analysing the convergence properties of a random walk seems to require new techniques, since the degree fluctuations make the use of any “inductive” arguments very difficult – from one step to another, the distribution of the walk could become worse, whereas the set of informed (or reachable) nodes can never decrease.

In this work, we will investigate the mixing time of a random walk on such evolving graphs. It turns out that, as our results demonstrate, the mixing time depends crucially on the density as well as on the speed by which the graph changes. We remark that deriving bounds on the mixing time on $G(n, p, q)$ poses some unique challenges, which are not present in the positive results of [3, 26]. The main difficulty is that in $G(n, p, q)$, due to the changing degrees of the vertices, there is no time-invariant stationary distribution, and the traditional notion of mixing time must be adapted to our dynamic setting. Informally, what we ask, then, is how many steps the walk needs to take before the distance to a time-dependent stationary distribution becomes small enough. Furthermore, in contrast to static graphs, where the distance between the distribution of the walk and the stationary
distribution can only decrease, in dynamic graphs the distance to the time-dependent stationary distribution might increase with time. For this reason, we also ask that the distribution of the walk remains close to a time-dependent stationary distribution for a long enough interval of time (for a precise definition of our notion of mixing time, see Definition 2.7). We believe this requirement is necessary for our definition of mixing time to be useful in potential applications.

**Further Related Work.**

Recently, [16] analysed the cover time of so-called “Edge-Uniform Stochastically-Evolving Graphs”, that include our model as a special case (i.e., the history is \( k = 1 \)). Their focus is on a process called “Random Walk with a Delay”, where at each step, the walk picks a (possible) neighbor and then waits until the edge becomes present. In [16 Theorem 4], the authors also relate this process to the standard random walk, and prove a worst-case upper bound on the cover time. However, one of the key difference to [16] is that we will study the mixing time instead of the cover time.

In [27], the authors analysed a continuous-time version of the edge-Markovian random graph. However, unlike the standard random walk, they consider a slightly different process: when the random walk tries to make a transition from a vertex \( u \), it picks one of the \( n-1 \) other vertices and moves there only if the edge is present; otherwise it remains in place. For this process, they were able to derive very tight bounds on the mixing time that establish the so-called cutoff phenomena. The same random walk was also analysed on a dynamic graph model of the \( d \)-dimensional grid in [20, 21], and, more generally, in [13].

1.1 Main Results

We study the mixing properties of random walks on edge-Markovian random graphs \( G(n, p, q) \). In particular, we consider six different settings of parameters \( p \) and \( q \), which separates edge-Markovian models based on how fast graphs change over time (slowly vs. fast changing), and how dense graphs in the dynamic sequence are (sparse vs. semi-sparse vs dense).

As noted in previous works (see, e.g., [6]), a dynamic sequence sampled from \( G(n, p, q) \) will eventually converge to an Erdős-Rényi random graph \( \tilde{G}(n, \tilde{p}) \) where \( \tilde{p} = \frac{p}{\rho + q} \) (for the sake of completeness, we give a proof in Appendix A). We use the expected degree in such a random graph, which is equal to \( d = (n-1)\tilde{p} \), to separate edge-Markovian models according to their density as follows:

1. **Sparse** \( d = o(\log n) \)

2. **Semi-sparse** \( d = \Theta(\log n) \)

3. **Dense** \( d = \omega(\log n) \).

Notice that the sparse regime corresponds to random graphs with density below the connectivity threshold of Erdős-Rényi random graphs.

We further separate edge-Markovian models based on how fast they change over time. Let \( \delta = \binom{n}{2}\tilde{p}q + \binom{n}{2}(1-\tilde{p})p \) be the expected number of changes at each step, when starting from a stationary initial graph \( G_0 \sim \tilde{G}(n, \tilde{p}) \). We consider the following two opposite regimes.

1. **Fast-changing** \( \delta = \Theta(dn) \).

2. **Slowly-changing** \( \delta = O(\log n) \)
Notice that the fast-changing regime corresponds to graphs for which a constant fraction of edges change at each step in expectation.

For a cleaner exposition, we consider six different settings based on fast-changing vs. slowly-changing rates and dense vs. medium vs. sparse graphs (Table 1.1). The density of the graph refers to the expected degree of the graph sampled from the stationary graph distribution (Sec. 2.1) of the model. Moreover this quantifies the expected degree \( d = np \). For reasons of space, we will not analyse all combinations of \( p \) and \( q \). Also note that some choices of \( p \) and \( q \) are not interesting, for example, small values of \( p \) (and \( q \)) may trap the walk inside a small region for a long time.

|                  | Fast-changing          | Slowly-changing         |
|------------------|------------------------|-------------------------|
| Sparse           | \( t_{\text{mix}} = \Theta(dn) \) | \( t_{\text{mix}} = O(\log n) \) |
| \( d \in [1, o(\log n)] \) | Thm 1.1 | Proposition 1.4 |
| Semi-sparse      | Coarse mixing\(^2\) in \( O(\log n) \) | \( t_{\text{mix}} = O(\log n) \), Thm 1.3 |
| \( d = \Theta(\log n) \) | Prop 3.2 | |
| Dense            | \( t_{\text{mix}} = O(\log n) \) | |
| \( d \in [\omega(\log n), n/2] \) | Thm 1.2 | |

Table 1.1: Summary of our main results (informal). See referenced theorems for the precise and complete statements.

The main results of our work are presented in Table 1.1. Here, we assume \( G_0 \) is sampled from the stationary graph distribution \( G(n, p) \). In the fast-changing regime, as highlighted in Remark A.2, this is without loss of generality. For slow-changing models, instead, different choices of \( G_0 \) can result in drastically different outcomes with regard to the mixing time. For ease of presentation, we assume in Table 1.1 that \( G_0 \sim G(n, p) \), but this assumption can usually be relaxed, and we refer to the full statement of the corresponding results for our actual assumptions on \( G_0 \).

Next we formally state the four main results of our work. The formal definitions of mixing time for random walk on dynamic graphs will be presented in Section 2.1 (see in particular Definition 2.7 and Definition 2.8). The first theorem is a negative result that tells us that, for fast-changing and sparse edge-Markovian graphs, random walks don’t have finite mixing time. Its proof will be presented in Section 3.1.

**Theorem 1.1** (Fast-changing and sparse, no mixing). Let \( p = \Theta \left( \frac{1}{n} \right) \) and \( q = \Omega(1) \). Then, \( t_{\text{mix}}(G(n, p, q)) = \infty \).

The following theorem is, instead, a positive result that establishes fast mixing time in the dense and fast-changing regime. Its proof is presented in Section 3.2.

**Theorem 1.2** (Fast-changing and dense, fast mixing). Let \( p = \omega \left( \log \frac{n}{n} \right) \) and \( q = \Omega(1) \). Then, \( t_{\text{mix}}(G(n, p, q)) = O(\log n) \).

The only case missing in the fast-changing regime is the semi-sparse case, where nodes have average degree \( d = \Theta(\log n) \). We do not have a definitive answer on the mixing time of random walks in such case, however, we do have a partial result that guarantees at least that random walk distributions will be “well spread” over a large support after \( O(\log n) \) steps (we call this behavior

\(^2\)In this regime we are not able to prove finite mixing time. However, we show that the distribution of the walk will “flatten out” after \( O(\log n) \) steps. We refer to this behavior as coarse mixing.
coarse mixing). This statement can be made formal by considering the $\ell_2$-norm of the distribution of the walk. Because of its technical nature, we defer the formal statement to Proposition 3.2.

We now turn our attention to the slowly-changing regime, where at most $\delta = O(\log n)$ edges are created and removed at each step. Unlike the results for the fast-changing regime, where the choice of the starting graph $G_0$ does not really affect the mixing time of a random walk (see Appendix A and Remark A.2 for a discussion), in the slowly-changing regime the choice of $G_0$ will affect the properties of $G_t$ for a large number of steps $t$.

The following theorem shows that in the slowly-changing and dense regime, under mild conditions on the starting graph $G_0 = (V, E_0)$ (which are satisfied for $G_0$ drawn from the limiting distribution of dense $G(n, p, q)$), random walks will mix relatively fast. We use $E_0(S, V \setminus S)$ to indicate the set of edges in $G_0$ between a subset of vertices $S \subset V$ and its complement, and $\Phi_{G_0}$ to indicate the minimum conductance of $G_0$ (see Definition 2.2).

**Theorem 1.3** (Slowly-changing and dense, fast mixing). Let $d = \Omega(\log n)$, $p = O(\log n/n^2)$, and $q = O((\log n)/(dn))$. Let the following assumptions on the starting graph $G_0 = (V, E_0)$ be satisfied for large enough constants $c_1, c_2, c_3 > 0$.

1. $\text{deg}_0(x) = \Theta(d)$ for any $x \in V$;
2. $|E_0(S, V \setminus S)| \geq c_2 \log n|S|$, for any $S \subset V$ with $|S| \leq c_1 \log n$;
3. $\Phi_{G_0} \geq c_3 \log d/d$.

Then, $t_{\text{mix}}(G(n, p, q)) = O((\log n/\Phi_{G_0}^2))$.

Let us briefly discuss the assumptions and results of Theorem 1.3. First of all notice that the parameters $p$ and $q$ are defined so that the average degree $d = \Omega(\log(n))$ and the number of changes in the graph at each step is $\delta = O(\log(n))$. Assumption (1) just require the degree of the vertices in $G_0$ to be of the same order as the degree of the vertices in the limiting graph $\mathcal{G}(n, \bar{p})$. Assumption (2) guarantees that for any small set $S$ there are enough edges going from $S$ to the rest of the graph. Assumption (3) is a mild condition on the conductance of $G_0$. These two conditions ensure that the conductance of $G_t$ will not be much lower than the conductance of $G_0$ for a large number of steps $t$. Finally, notice that $O((\log n/\Phi_{G_0}^2))$ is a classic bound for the mixing time of a static random walk on $G_0$. Theorem 1.3 essentially states that, if the three assumptions are satisfied, the mixing time of a random walk on $\mathcal{G}(n, p, q)$ will not be much larger. In particular, all the three assumptions are satisfied for a starting graph $G_0 \sim \mathcal{G}(n, \bar{p})$ with $\bar{p} = p/(p + q)$. Furthermore, in such case $t_{\text{mix}}(\mathcal{G}(n, p, q)) = O(\log n)$. The proof of this theorem can be found in Section 4.1.

We conclude this section by stating our result in the slowly-changing and dense regime. We prove a negative result: we show that the mixing time of $G(n, p, q)$ is at least linear in $n$.

**Proposition 1.4** (Slowly-changing and sparse, slow mixing). Let $p = O(1/n^2)$ and $q = \omega(1/(n \log n))$. Consider a random walk on $\mathcal{G}(n, p, q)$ with starting graph $G_0 \sim \mathcal{G}(n, \bar{p})$ with $\bar{p} = p/(p + q)$. Then, $t_{\text{mix}}(\mathcal{G}(n, p, q)) = \Omega(n)$.

2. **Notation and Definitions**

2.1. **Random Walk and Conductance**

In this section we introduce the relevant notation and basic results about Markov chains that we will use throughout the paper. For more background on Markov chains and random walks we defer the reader to [17].
Let \( \mathcal{G} = (G_t)_{t \in \mathbb{N}} \) be a sequence of undirected and unweighted graphs defined on the same vertex set \( V \), with \( |V| = n \), but with potentially different edge-sets \( E_t \) (\( t \in \mathbb{N} \)). We study (lazy) random walks on \( \mathcal{G} \): suppose that at a time \( t \geq 0 \) a particle occupies a vertex \( u \in V \). At step \( t+1 \) the particle will remain at the same vertex \( u \) with probability 1/2, or will move to a random neighbor of \( u \) in \( G_t \). In other words, it will perform a single random walk step according to a transition matrix \( P_t \), which is the transition matrix of a lazy random walk on \( G_t \): \( P_t(u, u) = 1/2 \), \( P_t(u, v) = 1/(2 \deg_t(u)) \) (where \( \deg_t(u) \) is the degree of \( u \) in \( G_t \)) if there is an edge between \( u \) and \( v \) in \( G_t \), or \( P_t(u, v) = 0 \) otherwise.

Given an initial probability distribution \( \mu_0 : V \to [0,1] \), which is the distribution of the initial position of the walk, the \( t \)-step distribution of a random walk on \( \mathcal{G} \) is equal to \( \mu_t = \mu_0 P_1 \cdot P_2 \cdot \ldots \cdot P_t \). In particular, we use \( \mu_t^x \) to denote the \( t \)-step distribution of the random walk starting at a vertex \( x \in V \). Hence \( \mu_0^x(1) = 1 \) and \( \mu_0^x(y) = 0 \) for \( x \neq y \in V \). Furthermore, we use \( \pi_t \) to denote the probability distribution with entries equal to \( \pi_t(x) = \deg_t(x)/(2|E_t|) \) for any \( x \in V \). This distribution is stationary for \( P_t \) (i.e., it satisfies \( \pi_t P_t = \pi_t \)) and, if \( G_t \) is connected, it is the unique stationary distribution of \( P_t \). If \( G_t \) is disconnected, \( P_t \) will have multiple stationary distribution. However, unless stated otherwise, we will consider only the “canonical” stationary distribution \( \pi_t \). Finally, while any individual \( P_t \) is time-reversible (it satisfies \( \pi_t(x)P_t(x, y) = \pi_t(y)P_t(y, x) \) for any \( x, y \in V \), a random walk on \( \mathcal{G} \) may not.

Recall that if \( P \) is a transition matrix of a reversible Markov chain, it has \( n \) real eigenvalues, which we denote with \(-1 \leq \lambda_0(P) \leq \ldots \leq \lambda_1(P) = 1\). If \( P \) is the transition matrix of a lazy random walk on a graph \( G \), it holds that \( \lambda_0(P) \geq 0 \). Moreover, \( \lambda_1(P) < 1 \) if and only if \( G \) is connected.

For two probability distributions \( f, g : V \to [0,1] \), the total variation distance between \( f \) and \( g \) is defined as \( \|f - g\|_{TV} := \frac{1}{2} \sum_{x \in V} |f(x) - g(x)| \). We denote with \( \|f\|_2 = (\sum_{x \in V} f^2(x))^{1/2} \) and \( \|f\|_\infty = \max_{x \in V} |f(x)| \) the standard \( \ell_2 \) and \( \ell_\infty \) norms of \( f \). Given a probability distribution \( \pi : V \to \mathbb{R}_+ \), we also define the \( \ell_2(\pi) \)-norm as \( \|f\|_{2,\pi} := \sqrt{\sum_{x \in V} f^2(x)\pi(x)} \). By Jensen’s inequality, it holds for any \( f, g \) that \( \|f - g\|_{TV} \leq \|f - g\|_{2,\pi} \). The lemma below relates the decrease in the distance to stationarity after one random walk step to the spectral properties of its transition matrix.

**Lemma 2.1** (Lemma 1.13 in [19], rephrased). Let \( P \) be the transition matrix of a lazy random walk on a graph \( G = (V, E) \) with stationary distribution \( \pi \). Then, for any \( f : V \to \mathbb{R} \), we have that

\[
\left\| \frac{fP}{\pi} - 1 \right\|_{2,\pi}^2 \leq \lambda_2(P)^2 \left\| \frac{f}{\pi} - 1 \right\|_{2,\pi}^2.
\]

In the lemma above and throughout the paper, a division between two functions is to be understood entry-wise, while \( 1 \) refers to a function always equal to one. An important quantity which can be used to obtain bounds on \( \lambda_2(P) \) is the conductance of \( G \), which is defined as follows.

**Definition 2.2.** The conductance of a non-empty set \( S \subseteq V \) in a graph \( G \) is defined as:

\[
\Phi_G(S) := \frac{|E(S, V \setminus S)|}{\text{vol}(S)},
\]

where \( \text{vol}(S) := \sum_{x \in V} \deg(x) \) and \( E(S, V \setminus S) \) is the set of edges between \( S \) and \( V \setminus S \). The

---

3For example, it might happen that \( P_1 \cdots P_t(x,y) > 0 \) while \( P_1 \cdots P_t(y,x) = 0 \). This cannot happen in the “static” case where \( P_1 = \cdots = P_t = P \) with \( P \) reversible.
conductance of the entire graph \( G \) is defined as
\[
\Phi_G := \min_{S \subseteq V: 1 \leq \text{vol}(S) \leq \text{vol}(V)/2} \frac{|E(S, V \setminus S)|}{\text{vol}(S)}.
\]

The conductance of \( G \) and the second largest eigenvalue of the transition matrix \( P \) of a lazy random walk in \( G \) are related by the so-called discrete Cheeger inequality [1], which we state below.

**Theorem 2.3** (Cheeger inequality). Let \( P \) be the transition matrix of a lazy random walk on a graph \( G \). Then, it holds that
\[
1 - \lambda_2(P) \leq \Phi_G \leq 2\sqrt{1 - \lambda_2(P)}.
\]

Finally, we use the notation \( o(n) \) to denote any function \( f : \mathbb{N} \to \mathbb{R} \) such that \( \lim_{n \to +\infty} f(n) = 0 \). We often drop the subscript \( n \).

### 2.2 Dynamic graph models

In this section we formally introduce the random models of (dynamic) graphs that are the focus of this work. We start by recalling the definition of the Erdős-Rényi model of random (static) graphs.

**Definition 2.4** (Erdős-Rényi model). \( G = (V, E) \sim \mathcal{G}(n, p) \) is a random graph such that \( |V| = \{1, \ldots, n\} \) and the \( \binom{n}{2} \) possible edges appear independently, each with probability \( p \).

We now introduce the edge-Markovian model of dynamic random graphs, which has been studied both in the context of information spreading in networks [6, 7] and random walks [16]. This model is the focus of our work.

**Definition 2.5** (edge-Markovian model). Given a starting graph \( G_0 \), we denote with \((G_t)_{t \in \mathbb{N}} \sim \mathcal{G}(n, p, q)\) a sequence of graphs such that \( G_t = (V, E_t) \), where \( V = \{1, \ldots, n\} \) and, for each \( t \in \mathbb{N} \), any pair of distinct vertices \( u, v \in V \) will be connected by an edge in \( G_t \) independently at random with the following probability:
\[
\mathbb{P} \left[ \{u, v\} \in E_{t+1} \mid G_t \right] = \begin{cases} 1 - q & \text{if } \{u, v\} \in E_t \\ p & \text{if } \{u, v\} \notin E_t. \end{cases}
\]

Notice that different choices of a starting graph \( G_0 \) will induce different probability distributions over \((G_t)_{t \in \mathbb{N}}\). In general, we try to study \( \mathcal{G}(n, p, q) \) by making the fewest possible assumptions on our choice of \( G_0 \). Moreover, as pointed out for example in [16], \((G_t)_{t \in \mathbb{N}} \sim \mathcal{G}(n, p, q)\) converges to \( \mathcal{G}(n, \bar{p}) \) with \( \bar{p} = p/(p + q) \). We leave considerations about the speed of this convergence and how this affects our choice of \( G_0 \) to Appendix A and, in particular, Remark A.2.

### 2.3 Mixing time of random walks on dynamic graphs

One of the most studied quantities in the literature about time-homogeneous (i.e., static) Markov chains (random walks included) is the mixing time, i.e., the time it takes for the distribution of the chain to become close to stationarity. Formally, it is defined as follows.

**Definition 2.6** (Mixing time for time-homogeneous Markov chains). Let \( \mu_t^x \) be the \( t \)-step distribution of a Markov chain with state space \( V \) starting from \( x \in V \). Let \( \pi \) be its stationary distribution. For any \( \epsilon > 0 \), the \( \epsilon \)-mixing time is defined as
\[
t_{\text{mix}}(\epsilon) := \min \{ t \in \mathbb{N} : \max_{x \in V} \| \mu_t^x - \pi \|_{TV} \leq \epsilon \}.
\]
A basic fact in random walk theory states that a lazy random walk on a connected undirected graph \( G = (V, E) \) has always a finite mixing time. In particular, if \( |V| = n \), \( t_{\text{mix}}(1/4) = O(n^3) \). Moreover, considering a different \( \epsilon \) does not significantly change the mixing time: for any \( \epsilon > 0 \), \( t_{\text{mix}}(\epsilon) = O(t_{\text{mix}}(1/4) \log(1/\epsilon)) \) (see, e.g., [17]). Also, it is a well-known fact that \( \|\mu_t - \pi\|_{TV} \) is non-increasing.

However, in the case of random walks on dynamic graphs, convergence to a time-invariant stationary distribution does not, in general, happen. For this reason, other works have studied alternative notions of mixing for dynamic graphs, such as merging [23], which happens when a random walk “forgets” the vertex where it started. In this work, instead, we focus on a different approach that we believe best translates the classical notion of mixing from the static to the dynamic case. More precisely, let us consider a dynamic sequence of graphs \((G_t)_{t \in \mathbb{N}}\) with increasing stationary distributions \((\pi_t)_{t \in \mathbb{N}}\). Our goal is to establish if there exists a time \( t \) such that the distribution \( \mu_t \) of the walk at time \( t \) is close to \( \pi_t \). Moreover, to make this notion of mixing useful in possible applications, we require that \( \mu_s \) remains close to \( \pi_s \) for a reasonably large number of steps \( s \geq t \). Formally, we introduce the following definition of mixing time for dynamic graph sequences.

**Definition 2.7 (Mixing time for dynamic graph sequences).** Let \( \mathcal{G} = (G_t)_{t \in \mathbb{N}} \) be a dynamic graph sequence on a vertex set \( V, |V| = n \). The mixing time of a random walk in \( \mathcal{G} \) is defined as

\[
t_{\text{mix}} (\mathcal{G}) = \min \left\{ t \in \mathbb{N} : \forall s \leq t + \sqrt{n}, \forall x \in V, \|\mu^x_s - \pi_s\|_{TV} = o_n(1) \right\},
\]

where \( \pi_s \) is the stationary distribution of a random walk in \( G_s \), and \( \mu^x_s \) is the \( s \)-step distribution of a random walk in \( \mathcal{G} \) that started from \( x \in V \).

First observe we require that the total variation distance between \( \mu_s \) and \( \pi_s \) goes to zero as the number of vertices increases.\(^4\) This is motivated by the fact that the distance to stationarity, unlike in the static case, might never drop beyond a certain threshold: for this reason, we explicitly require that such threshold becomes smaller and smaller with an increasing number of vertices. Secondly, we require that such distance remains small for \( \sqrt{n} \) steps (recall \( n \) is the number of vertices in the graph). This is due to the fact that, for all dynamic graph models we consider, we cannot hope for such distance to stay small arbitrarily long. However, we believe that \( \sqrt{n} \) steps is a long enough period of time for mixing properties to be useful in applications.

Since our goal is to study the mixing property of \( \mathcal{G}(n, p, q) \), we now introduce a definition of mixing time for edge-Markovian models that takes into account the probabilistic nature of such graph sequences. Essentially, we say that the mixing time of \( \mathcal{G}(n, p, q) \) is \( t \) if a random walk on a dynamic sequence of graphs sampled from \( \mathcal{G}(n, p, q) \) mixes (according to the previous definition) in \( t \) steps with high probability over the sampled dynamic graph sequence.

**Definition 2.8 (Mixing time for edge-Markovian models).** Given an edge-Markovian model \( \mathcal{G}(n, p, q) \), its mixing time is defined as

\[
t_{\text{mix}} (\mathcal{G}(n, p, q)) = \min \left\{ t \in \mathbb{N} : \mathbb{P}_{G \sim \mathcal{G}(n, p, q)} \left[ t_{\text{mix}} (G) \leq t \right] \geq 1 - o_n(1) \right\}.
\]

Finally, we remark that, while in static graphs connectivity is a necessary prerequisite to mixing, random walks on sequences of disconnected dynamic graphs might nonetheless exhibit mixing properties. Examples of this behavior were studied in [26].

\(^4\) We are implicitly assuming there is an infinite family of dynamic graph sequences with increasing \( n \).
3 Results for the fast-changing case

3.1 Negative result for mixing in the sparse and fast-changing case

In this section we consider random walks on sparse and fast-changing edge-Markovian graphs. In particular, we study $G(n,p,q)$ with $0 < q = \Omega(1)$ and $p = \Theta\left(\frac{1}{m}\right)$. Since $\Omega(1)$, by Remark A.2, we can restrict ourselves to consider the case where $G_0 \sim \mathcal{G}(n,\bar{p})$ with $\bar{p} = p/(p+q)$. We will show a negative result on the mixing of random walks in this regime: no matter how small is the distance between $\mu_t$ and $\pi_t$, the total variation distance will increase to a positive constant with constant nonzero probability.

**Theorem 1.1** (Fast-changing and sparse, no mixing). Let $p = \Theta\left(\frac{1}{m}\right)$ and $q = \Omega(1)$. Then, $t_{\text{mix}}(\mathcal{G}(n,p,q)) = \infty$.

The key idea behind this result is that due to the fast-changing nature, the degrees of nodes change rapidly. In particular, for a linear number of nodes $u$, there is at least one neighbor $v_{\min} \in \Gamma_t(u)$ whose degree may change from one constant in round $t$ to, basically any other constant (this also makes use of the assumption on $p$, ensuring that the graph is sparse). The proof then exploits that, due to the "unpredictable" nature of this change, the probability mass received by $v_{\min}$ will be non-zero in round $t + 1$ is likely to cause a significant difference between $\mu_{t+1}(u)$ and $\pi_{t+1}(u)$. Since this holds for a linear number of nodes $u$, we obtain a sufficiently large lower bound on the total variation distance, and the theorem is established.

**Proof of Theorem 1.1.** We will prove that no matter how small the distance between $\mu_t$ and $\pi_t$ is, the total variation distance between $\mu_{t+1}$ and $\pi_{t+1}$ will increase to a positive constant with constant nonzero probability. This will yield the theorem.

In this regime, at a time $t$ the graph has converged to $G(n,\bar{p})$, and when the graph has isolated nodes, the stationary distribution of the random walk on a disconnected graph is not unique. Recall that in this case, we choose $\pi_t(u) = \deg(u)/2|E_t|$ as the stationary distribution.

For any node $u$, we have

$$|\mu_{t+1}(u) - \pi_{t+1}(u)| = \left| \sum_{v \in \Gamma_{t+1}(u)} \pi_t(v) \cdot \frac{1}{\deg_{t+1}(v)} - \frac{\deg_{t+1}(u)}{2|E_{t+1}|} \right|$$

where by Chernoff bound, we know $|E_{t+1}|/|E_t|$ is close to 1 and both of them are $\Theta(n)$ in this regime. Hence we can use $|E| = \binom{n}{2} \tilde{p}$ to replace either of them to get an approximation. So the distance is contributed by $\deg_t(v)/\deg_{t+1}(v)$. W.l.o.g. in our setting we assume $p = \frac{1}{2}$ and $q = \frac{1}{2}$.

For a vertex $u \in V$, let $v_{\min} \in \Gamma_t(u)$ be the neighbor of $u$ with the smallest degree in $G_t$, and assume $u$ is a vertex so that $\deg_t(v_{\min}) \leq 20$. Let us define $\alpha := \sum_{v \in \Gamma_{t+1}(u) \setminus \{v_{\min}\}} \left( \frac{\deg_t(v)}{\deg_{t+1}(v)} - 1 \right)$, the contribution by all neighbors of $u$ instead of $v_{\min}$. Also let us define $\beta := \sum_{v \in \Gamma_{t+1}(u)} 1_{\{v, v_{\min} \in E_{t+1}\}}$, the number of neighbors of $v_{\min}$ in $\Gamma_{t+1}(u)$ and $\gamma := \sum_{v \in V \setminus \Gamma_{t+1}(u)} \sum_{v \in \Gamma_{t+1}(u)} 1_{\{v, v_{\min} \in E_{t+1}\}}$, the number of
neighbors of \( v_{\text{min}} \) in \( V \setminus \Gamma_{t+1}(u) \). Let us define the following events:

\[
A := \{|E_{t+1}| = \left(1 \pm \sqrt{10 \log n/n}\right)|E|\} \cap \{|E_t| = \left(1 \pm \sqrt{10 \log n/n}\right)|E|\},
\]

\[
B := \{\beta \leq 80\},
\]

\[
C := \left\{ \left| \alpha + \left( \frac{\deg_t(v_{\text{min}})}{\beta + \gamma} - 1 \right) \right| \geq c_4 \right\},
\]

\[
D := \{v_{\text{min}} \in \Gamma_{t+1}(u)\},
\]

where \(|E| = \binom{n}{2} \tilde{p}\) is the expected number of edges of a stationary graph, and \(c_4 > 0\) is a proper constant that is defined later.

Using Chernoff bound, we have that

\[
P[\neg A] \leq \exp\left(\frac{10 \log n|E|}{2n}\right) = O(n^{-c}),
\]

for some constant \(c\).

Note that \(P[A] \geq 1 - o(1), P[B] \geq 3/4\) and \(P[D] = 1 - q > 0\) which is also a constant probability in this regime. If the event \(A\) holds, then

\[
|\mu_{t+1}(u) - \pi_{t+1}(u)| = \sum_{v \in \Gamma_{t+1}(u)} \frac{1}{2|E_t|} \left( \frac{\deg_t(v)}{\deg_{t+1}(v)} - 1 \right) \pm O(\log^2 n/\sqrt{n}).
\]

Further, assuming that \(D\) holds, we can lower bound

\[
\sum_{v \in \Gamma_{t+1}(u) \setminus \{v_{\text{min}}\}} \frac{1}{2|E_t|} \left( \frac{\deg_t(v)}{\deg_{t+1}(v)} - 1 \right) + \frac{1}{2|E_t|} \left( \frac{\deg_t(v_{\text{min}})}{\deg_{t+1}(v_{\text{min}})} - 1 \right).
\]

Now let us expose all degrees for \( v \in \Gamma_{t+1}(u) \setminus \{v_{\text{min}}\} \), subject to event \(B\) holding. Hence the above formula is equal to

\[
\frac{1}{2|E_t|} \cdot \left( \alpha + \frac{\deg_t(v_{\text{min}})}{\beta + \gamma} - 1 \right).
\]

Recall that \(\alpha\) and \(\beta \leq 80\) are arbitrary. Further, \(\gamma\) is independent of \(\alpha\) and \(\beta\). Also for any constant \(c_1 > 0\), \(P[\gamma = c_1] > c_2\) for some other constant \(c_2 > 0\). Note in order for \(|\mu_{t+1}(u) - \pi_{t+1}(u)|\) not to contribute significantly to the total variation distance, we must have

\[
\alpha + \frac{\deg_t(v_{\text{min}})}{\beta + \gamma} - 1 = 0,
\]

which is equivalent to

\[
\gamma = \frac{\deg_t(v_{\text{min}})}{1 - \alpha} - \beta.
\]

There is at most one possible (positive integer) value for \(\gamma\) that solves this equation after \(\alpha, \beta\) have been revealed. In particular, there is at least one constant \(c_3 > 0\) so that

\[
\left| \alpha + \frac{\deg_t(v_{\text{min}})}{\beta + c_3} - 1 \right| > c_4,
\]
for some other constant \( c_4 > 0 \) because \( v_{\text{min}} \leq 20 \) and \( \beta \leq 80 \). This proves \( \mathbb{P}[C \mid B] \geq c_5 \). Further,

\[
\begin{align*}
\mathbb{P}[A \land C] &\geq \mathbb{P}[A] - \mathbb{P}[-C] \\
&\geq \mathbb{P}[A] - 1 + \mathbb{P}[C] \\
&\geq \mathbb{P}[A] - 1 + \mathbb{P}[C \land B] \\
&\geq \mathbb{P}[A] - 1 + \mathbb{P}[B] \cdot \mathbb{P}[C \mid B] \\
&\geq \frac{1}{4} \cdot c_5 > 0.
\end{align*}
\]

Hence with constant probability, any vertex \( u \in V \) in \( G_t \) that has a neighbor \( v_{\text{min}} \) with \( \deg_t(v_{\text{min}}) \leq 20 \) will contribute \( \Omega(1/|E|) \) to the \( \ell_1 \)-norm with constant probability \( c_6 > 0 \). Let us now lower bound the number of such vertices, \( \mathcal{S}_t := \{u \in V(G_t) : \min_{v \in \Gamma_t(u)} \{\deg(v)\} \leq 20 \} \).

Fix any vertex \( u \in V \). With constant probability \( C_1 > 0 \), it has at least one neighbor, say \( w \). Further, that neighbor \( w \) will have at most 19 neighbors (besides \( u \)) with constant probability \( C_2 > 0 \). Hence \( \mathbb{E}(|\mathcal{S}_t|) \geq (C_1 C_2) \cdot n \). Using Markov’s inequality, it follows that

\[
\mathbb{P}[n - |\mathcal{S}_t| \geq (1 + C_1 C_2/2) \cdot (1 - C_1 C_2) \cdot n] \leq \frac{1}{1 + C_1 C_2/2}.
\]

Rearranging the above, it follows that with constant probability \( C_3 := \frac{1}{1 + C_1 C_2/2} > 0 \), we have

\[
|\mathcal{S}_t| \geq (C_1 C_2/2) \cdot n.
\]

Hence the overall contribution of all vertices in \( \mathcal{S}_t \) is at least \( \Omega(n) \cdot \Omega(1/|E|) = \Omega(1) \) (as \( |E| = \Theta(n) \) in this regime), with some constant probability \( c_7 > 0 \).

### 3.2 Positive result for mixing in the dense and fast-changing case

In this section we analyse the mixing properties of \( \mathcal{G}(n, p, q) \) for \( p = \Omega(\log n/n) \) and \( q = \Omega(1) \). Since \( q \) is large, for simplicity we will assume throughout this section that \( G_0 \sim \mathcal{G}(n, \tilde{p}) \), where \( \tilde{p} = \frac{p}{p + q} \) (see Remark A.2 for an explanation of why this is not a restriction). The main result of this section is the following theorem.

**Theorem 1.2** (Fast-changing and dense, fast mixing). Let \( p = \omega(\log n/n) \) and \( q = \Omega(1) \). Then, \( t_{\text{mix}}(\mathcal{G}(n, p, q)) = O(\log n) \).

While in this paper we study for simplicity only lazy random walks on graphs, to prove Theorem 1.2, however, we need to introduce simple random walks on graphs: given a graph \( G = (V, E) \), a simple random walk on \( G \) has transition matrix \( Q \) such that, for any \( x, y \in V \), \( Q(x, y) = 1/\deg(x) \) if \( \{x, y\} \in E \), \( Q(x, y) = 0 \) otherwise. The following lemma, whose proof is the main technical part of the section, shows that if the simple random walk on a sequence of graphs \( \mathcal{G} = (G_t)_{t \in \mathbb{N}} \) exhibits strong expansion properties, and the time-varying stationary distribution is always close to uniform, then a lazy random walk on \( \mathcal{G} \) will be close to the stationary distribution of \( G_t \) for any \( t \) large enough. We remark that a strong expansion condition on lazy random walks can never be satisfied; luckily, we just need this strong expansion condition to hold for their simple counterpart.

**Lemma 3.1.** Let \((G_t)_{t \in \mathbb{N}}\) be a sequence of graphs, and \((P_t)_{t \in \mathbb{N}}\) (resp. \((Q_t)_{t \in \mathbb{N}}\)) the corresponding sequence of transition matrices for a lazy (resp. non-lazy) random walk. Assume there exists \( 1 < C = O(1) \) such that, for any \( t \geq 1 \) and any \( x \in V \), \( 1/(C \cdot n) \leq \pi_t(x) \leq C/n \). Moreover, also assume that, for any \( t \in \mathbb{N} \), \( \max\{|\lambda_2(Q_t)|, |\lambda_n(Q_t)|\} \leq \lambda = o(1) \). Then, there exists an absolute constant \( C' \) such that, w.h.p., for any \( t \geq C' \log n \) and any starting distribution \( \mu_0 \),

\[
\left\| \frac{\mu_t}{\pi_t} - 1 \right\|^2_{2, \pi_t} \leq 10C^2(C - 1)^2,
\]

where \( \mu_t = \mu_0 P_1 \cdots P_t \).
Proof of Lemma 3.1. We first relate the \( \ell_2(\pi) \) distance to stationarity to the \( \ell_2 \) distance from the uniform distribution. We start by observing that, for any \( t \in \mathbb{N} \), by our assumptions on \( \pi_t \), it holds that

\[
\|\pi_t - 1/n\|_2^2 = \sum_{x \in V} \left( \pi_t(x) - \frac{1}{n} \right)^2 \\
\leq \max \left\{ \left( 1 - \frac{1}{C} \right)^2, (C - 1)^2 \right\} \cdot \frac{n}{n^2} \\
= \frac{(C - 1)^2}{n}.
\]

Then, for any probability distribution \( p \), we have that

\[
\left\| \frac{p}{\pi_t} - 1 \right\|_{2,\pi_t}^2 = \sum_x \left( \frac{p(x) - \pi_t(x)}{\pi_t(x)} \right)^2 \\
\leq Cn \cdot \| p - \pi_t \|_2^2 \\
\leq 2Cn \cdot \left( \| p - 1/n \|_2^2 + \| \pi_t - 1/n \|_2^2 \right) \tag{3.1} \\
\leq 2Cn \cdot \left( \| p - 1/n \|_2^2 + (C - 1)^2/n \right) \\
= 2Cn \cdot \| p - 1/n \|_2^2 + 2C(C - 1)^2, \tag{3.2}
\]

where (3.1) holds by the triangle inequality and the basic inequality \((a + b)^2 \leq 2a^2 + 2b^2\). Analogously, it holds that

\[
\left\| \frac{p}{\pi_t} - 1 \right\|_{2,\pi_t}^2 \geq (n/C) \cdot \left( \frac{1}{2} \| p - 1/n \|_2^2 - 3\| \pi_t - 1/n \|_2^2 \right) \tag{3.3} \\
\geq (n/C) \cdot \left( \frac{1}{2} \| p - 1/n \|_2^2 - \frac{3}{n} (C - 1)^2 \right) \\
\geq (n/C) \cdot \| p - 1/n \|_2^2 - 3(C - 1), \tag{3.4}
\]

where (3.3) holds by the triangle inequality and the basic inequality \((a - b)^2 \geq a^2/2 - 3b^2\).

Notice that the distance to the uniform distribution does not change if at step \( t \) we perform a lazy step, which happens with probability \( 1/2 \). Conditioning on the fact that we don’t take a lazy step, at time \( t \) we can bound the decrease in the distance to the uniform distribution as follows:

\[
\left\| \frac{\mu_{t+1}}{\pi_{t+1}} - 1 \right\|_{2,\pi_{t+1}}^2 \leq \lambda \cdot \left\| \frac{\mu_t}{\pi_{t+1}} - 1 \right\|_{2,\pi_{t+1}}^2 \\
\leq \lambda \cdot \left( 2Cn \cdot \| \mu_t - 1/n \|_2^2 + 2C(C - 1)^2 \right), \tag{3.5}
\]

where the first inequality follows from Lemma 2.1 and the second by (3.2). Moreover, we have that

\[
n \cdot \| \mu_{t+1} - 1/n \|_2^2 \leq C \left\| \frac{\mu_{t+1}}{\pi_{t+1}} - 1 \right\|_{2,\pi_{t+1}}^2 + 3C(C - 1) \\
\leq o\left( \| \mu_t - 1/n \|_2^2 + C^2(C - 1)^2 \right) + 3C(C - 1), \tag{3.6}
\]

where the first inequality follows from (3.3), and the last from (3.5) and the assumption \( \lambda = o(1) \). This implies that whenever \( n \cdot \| \mu_t - 1/n \|_2^2 \) is large enough (e.g., \( n \cdot \| \mu_t - 1/n \|_2^2 \geq 4C(C - 1) \)), if we condition on the walk not taking a lazy step at time \( t \), the distance to the uniform distribution
will shrink significantly (this follows because \( \lambda = o(1) \)). Therefore, we just need \( O(\log n) \) non-lazy steps for such distance to become small. Hence, after \( t = O(\log n) \) steps, it holds w.h.p. that

\[
n \cdot \| \mu_t - 1/n \|_2^2 \leq 4C(C - 1), \tag{3.7}
\]

which also implies by (3.2) that

\[
\left\| \frac{\mu_t}{\pi_t} - 1 \right\|_{2,\pi_t}^2 \leq 10C^2(C - 1)^2.
\]

Moreover, after \( O(\log(n)) \) steps, this distance will continue to be small. In fact, let \( \mu_t \) satisfy (3.7). If we condition on taking a lazy step at time \( t + 1 \), such distance will not change. If instead we take a non-lazy step, by (3.6),

\[
n \cdot \| \mu_t - 1/n \|_2^2 \leq 3C(C - 1) + o(C^2(C - 1)^2) \leq 4C(C - 1),
\]

and, therefore, the distance to the uniform distribution again satisfies (3.7). The lemma follows by applying (3.2) once again. \( \square \)

We now show how it can be used to derive Theorem 1.2. First recall that since we are assuming \( G_0 \sim \mathcal{G}(n, \tilde{p}) \), all graphs in the sequence \((G_t)_{t \in \mathbb{N}}\) are sampled (non-independently) from \( \mathcal{G}(n, \tilde{p}) \) (see Appendix A). Furthermore, for any \( t \in \mathbb{N} \), the assumptions of Theorem 1.2 on \( \lambda_2(Q_t) \) and \( \lambda_\pi(Q_t) \) are satisfied with probability \( 1 - o(1/n^2) \) for any graph sampled from \( \mathcal{G}(n, \tilde{p}) \) with \( \tilde{p} > 2 \log n/n \) by [14 Theorem 1.1]. Moreover, for \( \tilde{p} = \omega(\log n/n) \), by standard Chernoff bounds argument we can show that, with probability \( 1 - o(1/n^2) \), all vertices of a graph sampled from \( \mathcal{G}(n, \tilde{p}) \) have degree \( (1 + o_n(1))n\tilde{p} \). This implies that, for any \( t \), w.h.p. the stationary distribution of \( G_t \) satisfies the assumptions of Lemma 3.1 with \( C = 1 + o(1) \), which yields Theorem 1.2.

It is natural to ask if we can relax the condition on \( p \). Assume for example that \( p, q \) are such that \( \tilde{p} = p/(p + q) > 2 \log n \). By [14 Theorem 1.1], the conditions on \( \lambda \) are still satisfied. However, it only holds that \( C = \Theta(1) \). Therefore, Lemma 3.1 can only establish that the \( \ell_2(\pi_t) \)-distance to stationarity is a constant (potentially larger than 1). This, unfortunately, does not give us any meaningful bound on the total variation distance. However, if the \( \ell_2 \)-distance between two distributions \( \mu \) and \( \pi \) is small, \( \mu(x) \) cannot be much larger than \( \pi(x) \). In a sense, this result can be interpreted as a coarse mixing property. This is summarised in the following proposition.

**Proposition 3.2.** Let \((G_t)_{t \in \mathbb{N}} \sim \mathcal{G}(n, p, q)\) with \( p/(p + q) > 2 \log n/n \) and \( q = \Omega(1) \). Let \( \pi_t \) be the stationary distribution of \( G_t \). Then, there exists absolute constants \( c_1, c_2 > 0 \) such that, for any starting distribution \( \mu_0 \) and any \( c_1 \log n \leq t \leq \sqrt{n} + c_1 \log n \), it holds that

\[
\mathbb{P} \left[ \left\| \frac{\mu_t}{\pi_t} - 1 \right\|_{2,\pi_t}^2 \leq c_2 \right] \geq 1 - o_n(1).
\]

**4 Results for the slowly-changing case**

**4.1 Positive result for mixing in the dense and slowly-changing case**

The aim of this section is to prove the following theorem.

**Theorem 1.3** (Slowly-changing and dense, fast mixing). Let \( d = \Omega(\log n) \), \( p = O(\log n/n^2) \), and \( q = O(\log n/(dn)) \). Let the following assumptions on the starting graph \( G_0 = (V, E_0) \) be satisfied for large enough constants \( c_1, c_2, c_3 > 0 \).
Lemma 4.1 (Conductance lower bound). Let $d = \Omega(n \log n)$, $p = O(n \log n/n^2)$, and $q = O(n \log n/(dn))$. Assume that $G_0$ satisfies assumptions (1), (2), (3) of Theorem 1.3. Then, there exists a constant $c > 0$ such that, for any $t = O(nd \log n)$ and any vertex $v \in V$,

$$\mathbb{P}\left[\deg_{G_t}(v) \leq \frac{1}{2}\deg_0(v)\right] = O(n^{-4})$$

and

$$\mathbb{P}\left[\Phi_{G_t} \geq c \cdot \Phi_{G_0}\right] = 1 - O(n^{-4}).$$

The proof of this lemma proceeds as follows: for any $S \subset V$, when an edge is randomly added or removed from the graph, we show that the probability that $|E_t(S, V \setminus S)|$ increases is usually larger than the probability it decreases. Therefore, we model $|E_t(S, V \setminus S)|$ as a random walk on $\mathbb{N}$ with a bias towards large values of $|E_t(S, V \setminus S)|$, i.e., a birth-and-death chain. Using standard argument about birth-and-death chains, we show in Lemma 4.1 that it is very unlikely that $|E_t(S, V \setminus S)|$ becomes much smaller than $|E_0(S, V \setminus S)|$. By a similar argument, in Lemma 4.3 we also show that the degrees of all nodes in $S$ are approximately the same as their original degrees in $G_0$. This ensures that the conductance of a single set $S$ is preserved after $t = O(dn \log n)$ steps. We then use a union bound argument to show that, with high probability, the conductance of the entire graph is preserved. For certain value of $d$, however, we cannot afford to use an union bound on all the possible sets of vertices. To overcome this, we show that we need to apply this union bound only for connected sets $S$. By carefully bounding the total number of such sets with respect to the maximum degree in $G_0$, we are able to establish the lemma.

We can now give an outline of the proof of Theorem 1.3. The idea is to show that $\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - 1\right\|_{2, \pi_{t+1}}$ is smaller than $\left\|\frac{\mu_{t}}{\pi_{t}} - 1\right\|_{2, \pi_{t}}$ (unless the latter is already very small). We do this by first relating $\left\|\frac{\mu_{t}}{\pi_{t}} - 1\right\|_{2, \pi_{t}}$ with $\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - 1\right\|_{2, \pi_{t}}$. More precisely, we can use Lemma 2.1 and Lemma 4.1 to show that the latter is smaller than the former by a multiplicative factor that depends on $\Phi_{G_0}$. Then, we bound the difference between $\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - 1\right\|_{2, \pi_{t}}$ and $\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - 1\right\|_{2, \pi_{t+1}}$. In particular, by exploiting the fact that at each step only $O(\log n)$ random edges can be deleted with high probability, we are able to show that $\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - 1\right\|_{2, \pi_{t+1}}$ is not much larger than $\left\|\frac{\mu_{t+1}}{\pi_{t}} - 1\right\|_{2, \pi_{t}}$. Finally, by putting together all these argument, we show that $\left\|\frac{\mu_{t}}{\pi_{t}} - 1\right\|_{2, \pi_{t}}$ is monotonically decreasing in $t$, at least until the walk is mixed. This establishes the theorem.

In the following analysis we need a standard Markov chain called birth-and-death chain [17, Chapter 2]. It is a random walk $(Z_t)_{t \in \mathbb{N}}$ on $\mathbb{N}$ whose transition probabilities are position based.
Formally, we define

$$\Delta Z_t := Z_{t+1} - Z_t = \begin{cases} +1 & \text{w.p. } b_{Z_t} \\ 0 & \text{w.p. } r_{Z_t} = 1 - b_{Z_t} - d_{Z_t} \\ -1 & \text{w.p. } d_{Z_t} \end{cases},$$

(4.1)

where \((b_{Z_t}, r_{Z_t}, d_{Z_t})\) are the probabilities that the position of the walk increases by 1, remains, and decreases by 1. \(b_{Z_t}\) and \(d_{Z_t}\) are functions depending on the position \(Z_t\) of the walk. Furthermore, we also need the hitting time of a random walk \(\tau^a_b := \min\{t \in \mathbb{N} : Z_t = b, Z_0 = a\}\) be the time the walk requires to hit \(b\) when starting from \(Z_0 = a\).

For a birth-and-death chain moving between 0 and \(m\), the stationary distribution of this Markov chain is given as follows:

$$\pi(k) = \frac{w_k}{\sum_{j=0}^{m} w_j},$$

(4.2)

where \(w_k = \prod_{i=1}^{k} \frac{b_i}{d_i}\) and if all \(b_{Z_t} = b'\) and \(d_{Z_t} = d'\), \(w_k = (b'/d')^k\). The birth-and-death Markov chain is time-reversible \([17]\) as defined in Sec. 2.1. Hence \(\pi(x)P^t(x,y) = \pi(y)P^t(y,x)\) where \(P\) is the transition matrix of the Markov chain and \(P^t(x,y)\) is the \(t\)-step transition probability from \(x\) to \(y\).

To prove Lemma 4.2, we first lower bound the number of edges in \(E_t(S,V \setminus S)\) and then upper bound the volume of \(S\) in \(G_t\) for any \(t = O(nd\log n)\). The following lemma proves that the number of edges on the boundary of any set \(S\) will not decrease to \(\epsilon\) less than the initial number with high probability.

**Lemma 4.2 (Lower bounding \(|E_t(S,V \setminus S)|\)).** Given the assumptions in Lemma 4.1, for one set \(S\), let \(M_t = |E_t(S,V \setminus S)|\) be the number of edges on the boundary of \(S\) in \(G_t\). Let \(M_0 = m_0\) be the initial number. Denote \(\mathbb{P}[\tau^m_b \leq t]\) the probability that a walk starting from \(a\) hits \(b\) before \(t\). Then for any constant \(\epsilon \in (0,1)\), we have

$$\mathbb{P}[\tau^m_b \leq t] = \begin{cases} O(n^{-c_1|S|}) & \text{if } |S| = O(\log n) \\ O\left(\frac{\log d}{\varepsilon^2 |S|}\right) & \text{if } |S| = \omega(\log n) \end{cases},$$

where \(m_T = \lfloor(1-\epsilon)m_0\rfloor\) and \(c_1, c_2 \geq 4\) are constants.

**Proof.** By the settings of the regime and a Chernoff bound, the number of total changes in the graph is upper bounded by \(O(\log n)\), and this holds for each of the first \(n^3\) steps with probability \(1 - n^{-\Omega(1)}\). Hence in one graph changing step, the quantity \(|E_t(S,V \setminus S)|\) changes by at most \(O(\log n)\). In the following analysis, we condition on this event being true.

We study the changes of the number of edges in \(E_t(S,V \setminus S)\) by modeling it as a birth-and-death Markov chain. As mentioned above, from \(M_t\) to \(M_{t+1}\), there are at most \(O(\log n)\) possible modifications. If we “unpack” them into single changes, then each change can be regarded as a random change on the number of edges in \(E(S,V \setminus S)\). Each change on an edge slot can add an edge if there is no edge, remove an edge if there is an edge, or keep it as it is. Hence we build a birth-and-death chain \((M'_s)_{s \in \mathbb{N}}\) where \(M'_0 = M_0 = m_0 = m'_0\) and \(m_T = m'_S = \lfloor(1-\epsilon)m_0\rfloor\). \(M'_s\) is still the number of edges in \(E(S,V \setminus S)\), but the time stamp represents the number of random changes we apply on the graph. From \(M_t\) to \(M_{t+1}\), the graph changing step can contribute \(O(\log n)\) such random changes for \((M'_s)_{s \in \mathbb{N}}\). We denote \(s(t)\) be the total number of random changes before time \(t\) (included).
Hence the transition probability of this birth-and-death chain is then defined as: for all $s \in [s(t), s(t + 1)]$, 
\[
\Delta M_s' := M_{s+1}' - M_s' = \begin{cases} 
+1 & \text{w.p. } b_{M_s'} \geq \left(1 - \frac{M_s + O(\log n)}{|S|(n - |S|)}\right) \cdot p \\
0 & \text{w.p. } r_{M_s} = 1 - b_{M_s'} - d_{M_s'} \\
-1 & \text{w.p. } d_{M_s} \leq \frac{M_s + O(\log n)}{|S|(n - |S|)} \cdot q
\end{cases}
\]

When the chain is at the position $M_s'$ and one random change happens, the probability that it causes an edge addition is \(\left(1 - \frac{M_s'}{|S|(n - |S|)}\right) \cdot p\) because it first needs to choose an empty slot in \(E(S, V \setminus S)\) with probability \(\left(1 - \frac{M_s'}{|S|(n - |S|)}\right)\) and then with probability \(p\) it adds an edge. However note that from \(M_t\) to \(M_{t+1}\), the random changes are not independent because if one random change picks one edge slot, the other changes caused by the same graph changing step cannot choose the same edge slot again. Hence the actual number of empty slots is in fact lower bounded by \(|S|(n - |S|) - M_t - O(\log n)\). Hence the probability \(b_{M_s'}\) has the lower bound as shown above. This argument works similarly for \(d_{M_s'}\). Due to the assumption that the initial graph has degree \(\Omega(\log n)\), these bounds would still remain to be of the same order.

Recall that \(p = \Omega\left(\frac{\log n}{n^2}\right)\) and \(q = O\left(\frac{\log n}{|E|}\right)\), hence the ratio is
\[
\frac{b_{M_s'}}{d_{M_s'}} \geq \frac{(|S|(n - |S|) - M_t + O(\log n)) \cdot p}{(M_t + O(\log n)) \cdot q} = \Omega(|S|).
\]

The current regime can give us appropriate constants such that \(\frac{b_{M_s'}}{d_{M_s'}} \geq 4\). Even if at the beginning the ratio may not satisfy this condition, there always exists a constant threshold \(\epsilon'\) such that when \(M_s'\) falls below \((1 - \epsilon')m_0\), \(\frac{b_{M_s'}}{d_{M_s'}}\) becomes less than 4. Hence w.l.o.g. we can wait until then and assume we start with \(\frac{b_{M_s'}}{d_{M_s'}} \geq 4\). Note that \(m_0'\) depends on \(|S|\) so we do the following case analysis.

**For small sets, \(|S| \leq 100 \log n\):** By the assumption that the conductance is a constant, we have \(M_0 = \Omega(|S| \log n)\). By (4.2) \(\pi(m_0', m_S') = O(1/4^m_{m_0})\) and \(\pi(m_0') \geq 3/4\). By the definition of the reversibility of a Markov chain [17]; for any time \(t\),
\[
\pi(m_0')p_t(m_0', m_S') = \pi(m_S')p_t(m_S', m_0').
\]
where \(p_t(x, y)\) means the \(t\)-step transition probability from \(x\) to \(y\). Hence
\[
p_t(m_0', m_S') \leq \pi(m_S') \cdot \frac{p_t(m_S', m_0')}{\pi(m_0')} \leq \frac{4}{3} \pi(m_S').
\]
Then we plug in \(\pi(m'_S) = O(1/4^m_{m_0})\),
\[
P\left[\tau_{m_0} = O(nd \log n)\right] = P\left[\tau_{m_S'} = O(nd \log^2 n)\right] = \sum_{i=0}^{O(nd \log^2 n)} p_t(m_0', m_S') \leq S \cdot \frac{4}{3} \pi(m_S') = O(n^{-c_1|S|}),
\]
where \(c_1 \geq 4\) is a constant.
For larger sized sets, \(|S| \geq 100 \log n\): By our assumption that the conductance is \(\Omega(\log d/d)\). We have \(m'_0 = \Omega(|S| \log d)\). Then \(\frac{b_{M'_0}}{d_{M'_0}}\) becomes \(O(d/\log d)\). Hence \(\pi(m'_S) = O((\log d/d)^{m'_0})\) and \(\pi(m'_0) \geq 1 - O(\log d/d)\). By applying a similar analysis we have

\[
\mathbb{P}\left[\tau_{m'_S}^{m'_0} = O(nd \log^2 n)\right] = \sum_{i=0}^{O(n \log^2 n)} p_i(m'_0, m'_S) \\
\leq \frac{1}{1 - O(\log d/d)} \cdot \pi(m'_S) \\
= O((\log d/d)^{-c_2 |S|})
\]

where \(c_2 \geq 4\) is a constant.

**Lemma 4.3** (Upper bounding the volume, \(\text{vol}(S)\)). Given the assumption in Lemma 4.1, for one set \(S\), let \(N_t = \text{vol}(S)\), the sum of the degrees of the vertices of \(S\) in \(G_t\). Let \(N_0 = n_0\) be the initial volume. Denote by \(\mathbb{P}[\tau_{b}^{t} \leq t]\) the probability that a walk starting at \(b\) hits \(t\) before time \(t\). Then for any constant \(\delta \geq 0\), we have

\[
\mathbb{P}\left[\tau^{N_0}_{n_T} = O(nd \log n)\right] = O(n^{-c|S|}),
\]

where \(n_T = \lfloor (1 + \delta)n_0 \rfloor\) and \(c \geq 4\).

**Proof.** Denote \(N_t\) the volume of the set \(S\) at time \(t\) then we have another birth-and-death chain \((N'_s)_{s \in \mathbb{N}}\) similar to the previous proof where \(N'_0 = N_0 = n_0\). Let \(s(t)\) denote the number of random changes before time \(t\) (included). For all \(s \in [s(t), s(t+1)]\),

\[
\Delta N'_s := N'_{s+1} - N'_s = \begin{cases} 1 & \text{w.p. } b_{N'_s} \leq \left(1 - \frac{N_t - O(\log n)}{(n-1)|S|}\right) \cdot p \\
0 & \text{w.p. } 1 - b_{N'_s} - d_{N'_s} \\
-1 & \text{w.p. } d_{N'_s} \geq \frac{N_t - O(\log n)}{(n-1)|S|} \cdot q
\end{cases}
\]

In general this proof is similar to the previous one. The ratio between \(b\) and \(d\) is

\[
\frac{b}{d} \leq O\left(\frac{(n-1)|S| - N_t + O(\log n)) \cdot p}{(N_t - O(\log n)) \cdot q}\right).
\]

This ratio is always a constant in our regime. Similar to the previous proof, we may not have a good ratio at the beginning, but there is always a constant threshold \(\delta'\) such that when \(N'_s = (1 + \delta')N'_0\) the ratio \(b/d\) is less than 1. Then by applying the same birth-and-death chain technique, we prove the lemma. 

Now by combining the above two lemmas we can prove Lemma 4.1.

**Proof of Lemma 4.1.** We have showed that for one set \(S\), within \(O(n \log^2 n)\) steps, the number of the edges on the boundary will not be \(\epsilon\) smaller and the volume will not be \(\delta\) larger than they originally were with high probability.

We apply a union bound to bound the conductance \(\Phi_G\) of the entire graph. By Lemma B.1, \(\Phi_G\) is revealed by just looking at the connected sets. Let \(G\) be the union of all graphs from \(G_1, ..., G_t\). Then all the possible connected sets that ever exist in \(G\) can be found in \(G\). The number of all connected sets for one certain graph \(G_i\) is upper bounded by those in the union graph. By Lemma B.2, the number of connected sets of size \(|S|\) is bounded by \(n \cdot \Delta^{|S|}\) where \(\Delta\) is the maximum degree in \(G\). By applying the birth-and-death chain argument for \(|S| = 1\), the maximum degree \(\Delta\) of the
union graph should be upper bounded by $O(d)$. Hence below we use $n \cdot O(d^{2|S| - 2})$ for the union bound when needed.

Denote $E_1(S)$ the event that the number of the edges on the boundary of $S$ ever reaches the $\epsilon$ less than the beginning. By union bound and Lemma 4.2, the probability that there exists such a set is upper bounded by

$$P\left[ \bigcup_{S \subseteq V} E_1(S) \right] \leq P \left[ \bigcup_{S \subseteq V: |S| = O(\log n)} E(S) \right]$$

$$+ P \left[ \bigcup_{S \subseteq V: |S| \in [\omega(\log n), O(n/\log n)]} E_1(S) \right]$$

$$+ P \left[ \bigcup_{S \subseteq V: |S| \in [\omega(n/\log n), O(n)]} E_1(S) \right]$$

$$\leq \sum_{|S| = O(\log n)} n \cdot \Delta |S| \cdot O(n^{-c_1|S|})$$

$$+ \sum_{[\omega(\log n), O(n/\log n)]} n \cdot \Delta |S| \cdot O\left(\frac{\log d}{d} \right)^{-c_2|S|}$$

$$+ \sum_{[\omega(n/\log n), O(n)]} \left(\frac{n}{|S|}\right) O\left(\frac{\log d}{d} \right)^{-c_2|S|}$$

$$\leq O(n^{-c_3})$$

where $c_1, c_2$ are the constants used in Lemma 4.2 and $c_3 \geq 4$ is a constant.

Denote $E_2(S)$ the event that the volume of $S$ reaches $\delta$ larger than the beginning. By union bound and Lemma 4.3, the probability that there exists such a set is upper bounded by

$$P\left[ \bigcup_{S \subseteq V} E_2(S) \right] \leq \sum_{S \subseteq V} \left(\frac{n}{|S|}\right) O(n^{-c|S|}) = O(n^{-c'})$$

where $c$ is the constant used in Lemma 4.3 and $c' \geq 4$ is a constant.

By combining everything above, we can lower bound the conductance. If the initial conductance is lower bounded by $\phi$, i.e., $\Phi_G^0(S) = \frac{X_0}{Y_0} \geq \phi$ and

$$\Phi_{G_1}(S) \geq \frac{(1 - \epsilon)X_0}{(1 + \delta)Y_0} \geq \frac{(1 - \epsilon)\phi}{(1 + \delta)}$$

where $\epsilon, \delta$ are the constants used in Lemma 4.2 and Lemma 4.3. For simplicity and also because of the arguments we have made along this proof, we choose come constants to be 4 in our final statement of the lemma.

Proof of Theorem 1.3. We establish the theorem by showing unless $\|\mu_t - 1\|_{2, \pi_t}$ is already small, $\|\mu_t - 1\|_{2, \pi_t}$ will significantly decrease at each step. In particular we relate $\|\mu_t - 1\|_{2, \pi_t}$ to $\|\mu_{t+1} - 1\|_{2, \pi_{t+1}}$ in two steps:
(1) We lower bound the change between $\|\frac{\mu_t}{\pi_t} - 1\|_{2,\pi_t}$ and $\|\frac{\mu_{t+1}}{\pi_{t+1}} - 1\|_{2,\pi_t}$.

(2) We upper bound the difference between $\|\frac{\mu_{t+1}}{\pi_{t+1}} - 1\|_{2,\pi_t}$ and $\|\frac{\mu_{t+1}}{\pi_{t+1}} - 1\|_{2,\pi_{t+1}}$.

**Step 1:** For the first step, we use a spectral argument. Using Lemma\textsuperscript{2.1}

$$
\left\| \frac{\mu_t}{\pi_t} - 1 \right\|^2_{2,\pi_t} - \left\| \frac{\mu_{t+1}}{\pi_{t+1}} - 1 \right\|^2_{2,\pi_t} \geq (1 - \lambda_2^2(P_t)) \left\| \frac{\mu_t}{\pi_t} - 1 \right\|^2_{2,\pi_t},
$$

by rearranging terms we get

$$
\left\| \frac{\mu_{t+1}}{\pi_{t+1}} - 1 \right\|^2_{2,\pi_t} \leq \lambda_2^2(P_t) \left\| \frac{\mu_t}{\pi_t} - 1 \right\|^2_{2,\pi_t},
$$

where $\lambda_2(P_t)$ is the second largest eigenvalue of $P_t$, the transition matrix of $G_t$.

**Step 2:** Next step we upper bound the difference between $\|\frac{\mu_{t+1}}{\pi_{t+1}} - 1\|_{2,\pi_t}$ and $\|\frac{\mu_{t+1}}{\pi_{t+1}} - 1\|_{2,\pi_{t+1}}$.

Due to the randomness of the graph we will compute the expectation of this difference. In the following analysis we condition on the event that at any time $t$, $|E_t| \in [(1 - o(1))nd, (1 + o(1))nd]$ where $d = (n - 1)\bar{\rho}$. This event has probability $1 - o(1)$. Recall that

$$
\left\| \frac{\mu_t}{\pi_t} - 1 \right\|^2_{2,\pi_t} = \sum_{y \in V} \pi_t(y) \left( \frac{\mu(y)}{\pi_t(y)} - 1 \right)^2 = \left( \sum_{y \in V} \frac{\mu_t^2(y)}{\pi_t(y)} \right) - 1.
$$

Hence we have

$$
E \left[ \left\| \frac{\mu_{t+1}}{\pi_{t+1}} - 1 \right\|^2_{2,\pi_{t+1}} - \left\| \frac{\mu_{t+1}}{\pi_{t+1}} - 1 \right\|^2_{2,\pi_t} \right]
$$

$$
= E \left[ \left( \sum_{y \in V} \frac{\mu_{t+1}^2(y)}{\pi_{t+1}(y)} \right) - \left( \sum_{y \in V} \frac{\mu_t^2(y)}{\pi_t(y)} \right) \right]
$$

$$
= \sum_{y \in V} E \left[ \frac{\mu_{t+1}^2(y)}{\pi_{t+1}(y)} \left( \frac{1}{\pi_{t+1}(y)} - \frac{1}{\pi_t(y)} \right) \right]
$$

$$
= \sum_{y \in V} E \left[ \frac{\mu_{t+1}^2(y)}{\pi_{t+1}(y)} \left( \frac{2|E_{t+1}|}{\deg_{t+1}(y)} - \frac{2|E_t|}{\deg_t(y)} \right) \right]
$$

$$
\leq 2(1 + o(1))|E| \sum_{y \in V} E \left[ \frac{\mu_{t+1}^2(y)}{\deg_{t+1}(y)} \left( \frac{1}{\deg_{t+1}(y)} - \frac{1}{\deg_t(y)} \right) \right]
$$

$$
\leq 2(1 + o(1))|E| \sum_{y \in V} \mu_{t+1}^2(y) E \left[ \frac{1}{\deg_{t+1}(y)} - \frac{1}{\deg_t(y)} \right]
$$

$$
\leq 2(1 + o(1))|E| \sum_{y \in V} \mu_{t+1}^2(y) \frac{\epsilon \deg_t(y)}{\deg_t(y) \cdot (1 - \epsilon) \deg_t(y)} (1 - (1 - q)^{\deg_t(y)})
$$

$$
\leq \frac{2(1 + o(1))}{1 - o(1)} \sum_{y \in V} \frac{\epsilon}{\deg_t(y)} \cdot \frac{\mu_{t+1}^2(y)}{\deg_t(y)/((1 - o(1))|E|)} (1 - (1 - q)^{\deg_t(y)})
$$

(4.3)
where \(|E| = nd\) and \(d = (n-1)\tilde{p}\). The \(\epsilon\) is the constant used by Lemma 4.1, \(\deg_{t+1}(y)\) will not decrease to \(\epsilon\) smaller than \(\deg(y)\) with probability \(1 - O(n^{-c})\). From line \((4.3)\) to line \((4.4)\) we upper bound the expectation by only considering the cases where the difference is positive, i.e., \(\deg_t(y) \geq \deg_{t+1}(y)\). In line \((4.4)\), by Lemma 4.1 we know \(\deg_{t+1}(y)\) will not be smaller than \(\frac{1}{2} \cdot \deg(y)\) with probability \(1 - O(n^{-4})\). Moreover, the probability \(1 - (1-q)^{\deg_t(y)}\) is the probability that at least one of the edges connected to \(y\) at time \(t\) changes at \(t+1\). In line \((4.5)\), we hide unimportant constants in the \(O\)-notation and we use the inequality \((1-q)^{\deg_t(y)} \geq 1 - q \cdot \deg_t(y)\).

Since \(q = O(\log n/(dn))\) by assumption, we get \(O(\log n/n)\) in line \((4.5)\).

By combining the two steps above we have

\[
\mathbb{E}\left[\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - 1\right\|^2_{2,\pi_{t+1}} - \left\|\frac{\mu_t}{\pi_t} - 1\right\|^2_{2,\pi_t}\right] \\
\leq O\left(\frac{\log n}{n}\right) \left(\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - 1\right\|^2_{2,\pi_{t+1}} - (1 - \lambda_2^2(P_t)) \left\|\frac{\mu_t}{\pi_t} - 1\right\|^2_{2,\pi_t}\right) \\
\leq O\left(\frac{\log n}{n}\right) \left(\lambda_2^2(P_t) - 1\right) \left\|\frac{\mu_t}{\pi_t} - 1\right\|^2_{2,\pi_t} + O\left(\frac{\log n}{n}\right) \\
\leq \left(\frac{n + \log n}{n}\right) \lambda_2^2(P_t) - 1\right) \left\|\frac{\mu_t}{\pi_t} - 1\right\|^2_{2,\pi_t} + O\left(\frac{\log n}{n}\right).
\]

Therefore, it holds that

\[
\mathbb{E}\left[\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - 1\right\|^2_{2,\pi_{t+1}}\right] \leq \left(\frac{n + \log n}{n}\right) \lambda_2^2(P_t) - 1\right) \left\|\frac{\mu_t}{\pi_t} - 1\right\|^2_{2,\pi_t} + O\left(\frac{\log n}{n}\right).
\]

By Theorem 22.3 and the laziness of the walk,

\[
\frac{\Phi_{G_t}^2}{2} \leq 1 - \lambda_2(P_t) \leq 2\Phi_{G_t}.
\]

Since we assume the conductance is lower bounded by \(O(\log d/d)\), we have \(\lambda_2(P_t) \leq 1 - O(\log^2 d/d^2)\) and hence \((n + \log n)/n\lambda_2^2(P_t) \leq 1\). Therefore in expectation the \(\ell_2\) distance shrinks by a factor less than 1 but with an additive term. By a similar analysis for the static graph in [17], after \(O(\log n/\Phi_{G_0}^2)\) rounds, the expected distance to \(\pi_t\) is at most \(O(\sqrt{\log n/n})\). By Lemma 4.1 we know this holds for \(\text{poly}(n)\) time. Hence it suffices to apply Markov’s inequality and union bound to show that the expected distance is small with probability \(1 - O(n^{-c})\) on a polynomially long time interval as required in the mixing time notion.

### 4.2 Negative result for mixing in the sparse and slowly changing case

**Proposition 1.4** (Slowly-changing and sparse, slow mixing). Let \(p = O(1/n^2)\) and \(q = \omega(1/(n \log n))\). Consider a random walk on \(G(n,p,q)\) with starting graph \(G_0 \sim G(n,\tilde{p})\) with \(\tilde{p} = p/(p+q)\). Then, \(t_{\text{mix}}(G(n,p,q)) = \Omega(n)\).
Proof. Consider the graph $G_0 \sim G(n, \bar{p})$. Notice that $\bar{p} = o(\log n/n)$ is well below the connectivity threshold of Erdős-Rényi random graphs. Therefore, with high probability, there is at least one isolated vertex in $G_0$; call this vertex $u$ and assume the random walk starts from that vertex. The probability that $u$ remains isolated in the steps $1, 2, \ldots, t$ is at least

$$(1 - p)^{(n-1)t} \geq (1 - O(1/n^2))^{(n-1)t} \geq 1 - O(t/n).$$

Therefore, with at least constant nonzero probability, there exists a constant $c > 0$ such that, for any $t \leq c \cdot n$, $\mu_t^u(u) = 1$. Since $\pi_t(u) = 0$, this implies that $\|\mu_t^u - \pi_t\|_{TV} = 1$. \hfill \Box

Actually the proof reveals a stronger “non-mixing” property; if the random walk starts from a vertex that is isolated in $G_0$, then this vertex will remain isolated for $\Theta(1/(np))$ rounds in expectation, and in this case the random walk did not move at all!

5 Conclusion

In this work we investigated the mixing time of random walks on the edge-Markovian random graph model. Our results cover a wide range of different densities and speeds by which the graph changes. On a high level, these findings provide some evidence of the intuition that both of the two properties “high density” and “slow changes” correlate with fast mixing.

For further work, one interesting setting that is not fully understood is the semi-sparse ($d = \Theta(\log n)$) and fast-changing ($q = \Omega(1) > 0$) case. While we proved that the random walk achieves some coarse mixing in $O(\log n)$, we conjecture that strong mixing is not possible. Another possible direction for future work is, given the bounds on the mixing time at hand, to derive tight bounds on the cover time. Finally, it would be also interesting to study the mixing time in a dynamic random graph model, where not all edge slots are present (similar to the models studied in [13][16], where the graph at each step is a random subgraph of a (possibly sparse) network).

References

[1] Alon, N., and Milman, V. D. $\lambda_1$, isoperimetric inequalities for graphs, and superconcentrators. J. Combin. Theory Ser. B 38, 1 (1985), 73–88.

[2] Augustine, J., Pandurangan, G., and Robinson, P. Distributed algorithmic foundations of dynamic networks. SIGACT News 47, 1 (2016), 69–98.

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

[4] Berenbrink, P., Giakkoupis, G., Kermarrec, A.-M., and Mallmann-Trenn, F. Bounds on the voter model in dynamic networks. In Proceedings of the 43rd International Colloquium on Automata, Languages and Programming (ICALP) (July 12–15 2016), pp. 146:1–146:15.

[5] Boyd, S., Ghosh, A., Prabhakar, B., and Shah, D. Randomized gossip algorithms. IEEE Trans. Inform. Theory 52, 6 (2006), 2508–2530.

[6] Clementi, A., Crescenzi, P., Doerr, C., Fraigniaud, P., Pasquale, F., and Silvestri, R. Rumor spreading in random evolving graphs. Random Structures Algorithms 48, 2 (2016), 290–312.
[7] Clementi, A., Monti, A., Pasquale, F., and Silvestri, R. Information spreading in stationary Markovian evolving graphs. *IEEE Transactions on Parallel and Distributed Systems* 22, 9 (2011), 1425–1432.

[8] Clementi, A. E. F., Macci, C., Monti, A., Pasquale, F., and Silvestri, R. Flooding time of edge-markovian evolving graphs. *SIAM J. Discrete Math.* 24, 4 (2010), 1694–1712.

[9] Clementi, A. E. F., Silvestri, R., and Trevisan, L. Information spreading in dynamic graphs. *Distributed Computing* 28, 1 (2015), 55–73.

[10] Cooper, C. Random walks, interacting particles, dynamic networks: Randomness can be helpful. In *Structural Information and Communication Complexity - 18th International Colloquium, SIROCCO 2011* (2011), A. Kosowski and M. Yamashita, Eds., vol. 6796, Springer, pp. 1–14.

[11] Denysyuk, O., and Rodrigues, L. E. T. Random walks on evolving graphs with recurring topologies. In *Distributed Computing - 28th International Symposium, DISC 2014, Austin, TX, USA, October 12-15, 2014. Proceedings* (2014), pp. 333–345.

[12] Giakkoupis, G., Sauerwald, T., and Stauffer, A. Randomized rumor spreading in dynamic graphs. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part II* (2014), pp. 495–507.

[13] Hermon, J., and Sousi, P. Random walk on dynamical percolation. *arXiv preprint arXiv:1902.02770* (2019).

[14] Hoffman, C., Kahle, M., and Paquette, E. Spectral Gaps of Random Graphs and Applications. *International Mathematics Research Notices* (05 2019).

[15] Kuhn, F., and Oshman, R. Dynamic networks: models and algorithms. *SIGACT News* 42, 1 (2011), 82–96.

[16] Lamprou, I., Martin, R., and Spirakis, P. Cover time in edge-uniform stochastically-evolving graphs. *Algorithms (Basel)* 11, 10 (2018), Paper No. 149, 15.

[17] Levin, D. A., and Peres, Y. *Markov chains and mixing times*, vol. 107. American Mathematical Soc., 2017.

[18] Michail, O., and Spirakis, P. G. Elements of the theory of dynamic networks. *Commun. ACM* 61, 2 (2018), 72.

[19] Montenegro, R., and Tetali, P. Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.* 1, 3 (2006), x+121.

[20] Peres, Y., Sousi, P., and Steif, J. Mixing time for random walk on supercritical dynamical percolation. *Prob. Theory and Related Fields* (2019). (to appear).

[21] 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 (2015), 487–530.

[22] Saloff-Coste, L., and Zúñiga, J. Merging for time inhomogeneous finite Markov chains. I. Singular values and stability. *Electron. J. Probab.* 14 (2009), 1456–1494.
[23] Saloff-Coste, L., and Zúñiga, J. Merging and stability for time inhomogeneous finite Markov chains. In *Surveys in stochastic processes*, EMS Ser. Congr. Rep. Eur. Math. Soc., Zürich, 2011, pp. 127–151.

[24] Saloff-Coste, L., and Zúñiga, J. Merging for inhomogeneous finite Markov chains, Part II: Nash and log-Sobolev inequalities. *Ann. Probab.* 39, 3 (2011), 1161–1203.

[25] Sarma, A. D., Molla, A. R., and Pandurangan, G. Distributed computation in dynamic networks via random walks. *Theor. Comput. Sci.* 581 (2015), 45–66.

[26] 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*, vol. 132 of LIPIcs. Leibniz Int. Proc. Inform. 2019, pp. Art. No. 93, 15.

[27] Sousi, P., and Thomas, S. Cutoff for random walk on dynamical Erdos-Renyi graph. *arXiv preprint arXiv:1807.04719* (2018).
A Mixing times for the graph chain of edge-Markovian models

It is well known that the edge-Markovian graph model $G(n, p, q)$ converges to an Erdős-Rényi model $G(n, \tilde{p})$ where $\tilde{p} = \frac{p}{p+q}$, which is the stationary distribution of the original edge-Markovian model. The mixing time of the graph chain has not been proven formally in previous works. Hence, we provide a proof for the sake of completeness. We remark that since an edge-Markovian model is a time-homogeneous (i.e., static) Markov chain, the classical definition of mixing time (Definition 2.6) applies.

**Theorem A.1** (Graph chain mixing time). For an edge-Markovian model $G(n, p, q)$, the graph distribution converges to the graph distribution of the random graph model $G(n, \tilde{p})$ where $\tilde{p} = \frac{p}{p+q}$.

For any $\epsilon \in (0, 1)$, the mixing time of the graph chain $G(n, p, q)$ is $t_{\text{mix}}(\epsilon) = O\left(\frac{\log(n/\epsilon)}{\log(1/|1-p-q|)}\right)$ for $p+q \neq 1$, and $t_{\text{mix}}(\epsilon) = 1$ if $p+q = 1$.

**Proof.** Every edge slot can be represented by a two-state (close/open) Markov chain with transition matrix

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

and stationary distribution $\left(\frac{q}{p+q}, \frac{p}{p+q}\right)$. By using standard Markov chain arguments (see, e.g., [17, Chapter 1]), the distance to the stationary distribution shrinks at each step by a factor of $1 - p - q$, i.e.,

$$||\mu_{t+1} - \pi||_{TV} \leq |1 - p - q||\mu_t - \pi||_{TV}.$$ 

Therefore, when $p+q \neq 1$, the mixing time $t_{\text{mix}}(\epsilon)$ of this two-state Markov chain is $O\left(\frac{\log(1/\epsilon)}{\log(1/|1-p-q|)}\right)$ where $\epsilon < 1$. For all the $\binom{n}{2}$ edge slots, the time that all of them mix is $O\left(\frac{\log(\binom{n}{2}) + \log(1/\epsilon)}{\log(1/|1-p-q|)}\right)$. When $p+q = 1$, instead, the graph mixes immediately, which confirms the fact that in this regime the graph model is equivalent to a sequence of independent graphs from $G(n, \tilde{p})$.

**Remark A.2.** Theorem A.1 essentially tells us that, whenever at least one between $p$ and $q$ is large (e.g., $\Omega(1)$), the graph chain quickly converges to $G(n, \tilde{p})$ with $\tilde{p} = \frac{p}{p+q}$. This suggests that for a fast-changing edge-Markovian model $G(n, p, q)$ with $q = \Omega(1)$, we can consider w.l.o.g. the starting graph $G_0$ as sampled from $G(n, \tilde{p})$.

B Missing proofs in Section 4.1

**Lemma B.1.** For any graph $G$, the conductance of $G$ is equal to

$$\Phi_G = \min_{S \subseteq V, |S| \leq |G|/2} \Phi_G(S).$$

**Proof.** By definition the conductance of a graph $\Phi_G$ is

$$\Phi_G = \min_{S \subseteq V, |S| \leq |G|/2} \frac{|E(S, V \setminus S)|}{\text{vol}(S).}$$

Assuming the graph conductance is achieved by a disconnected set $D$. W.l.o.g. we assume it has two connected components $A$ and $B$. Then since there are no edges between $A$ and $B$, we have

$$\Phi_G = \Phi_G(D) = \frac{|E(A, V \setminus A)| + |E(B, V \setminus B)|}{\text{vol}(A) + \text{vol}(B)}$$
Then by a simple inequality,

\[ \Phi_G(D) \geq \min \left\{ \frac{|E(A, V \setminus A)|}{\text{vol}(A)}, \frac{|E(B, V \setminus B)|}{\text{vol}(B)} \right\} \]

where \(|A|, |B| \leq |D| \leq n/2\). By induction, this holds for all disconnected sets with more than two connected components. Hence if there exists a disconnected set \(D\) which gives the minimum conductance over the entire graph then there must exist a connected set \(S\) in the graph which also achieves this conductance.

**Lemma B.2.** Let \(G\) be a graph with \(n\) vertices and the maximum degree is \(\Delta\). The number of all the connected sets with \(k\) vertices is at most \(n \cdot \Delta^{2k-2}\).

**Proof.** For a fixed \(k > 0\), we define an encoding which can represent a path of length \(2k\) starting from a certain node. First we label \(n\) vertices from 1 to \(n\). Then when enumerating all paths of length \(2k\), we first output the label of the starting vertex, then for its neighbors we sort them based on their labels and then use the ranks in the following output.

The output should start with a label which represents the root node which is from 1 to \(n\). Then the following numbers are all less than \(\Delta\) because each of them only means the rank of a node among all the neighbors of its predecessors. Hence at most we have \(n \cdot \Delta^{2k-2}\) encodings.

There is an injective map between all the connected set of size \(k\) with the encoding. For any such set, there is a spanning tree of size \(k\). A DFS traversal of this tree would only use each edge of the spanning tree twice. Hence a not necessarily simple path of length \(2(k - 1)\) must exist. Our encoding essentially gives all possible paths of length \(2(k - 1)\). So there must be such an injective map. The number of the strings in our output is \(n \cdot \Delta^{2(k-1)}\). Hence the total number of connected sets is upper bounded by that as well. \(\square\)