



# Geometric bounds on the fastest mixing Markov chain

Sam Olesker-Taylor<sup>1</sup> · Luca Zanetti<sup>2</sup>

Received: 10 July 2022 / Revised: 17 August 2023 / Accepted: 9 December 2023  
© The Author(s) 2024

## Abstract

In the Fastest Mixing Markov Chain problem, we are given a graph  $G = (V, E)$  and desire the discrete-time Markov chain with smallest mixing time  $\tau$  subject to having equilibrium distribution uniform on  $V$  and non-zero transition probabilities only across edges of the graph. It is well-known that the mixing time  $\tau_{RW}$  of the lazy random walk on  $G$  is characterised by the edge conductance  $\Phi$  of  $G$  via Cheeger's inequality:  $\Phi^{-1} \lesssim \tau_{RW} \lesssim \Phi^{-2} \log |V|$ . Analogously, we characterise the fastest mixing time  $\tau^*$  via a Cheeger-type inequality but for a different geometric quantity, namely the vertex conductance  $\Psi$  of  $G$ :  $\Psi^{-1} \lesssim \tau^* \lesssim \Psi^{-2} (\log |V|)^2$ . This characterisation forbids fast mixing for graphs with small vertex conductance. To bypass this fundamental barrier, we consider Markov chains on  $G$  with equilibrium distribution which need not be uniform, but rather only  $\varepsilon$ -close to uniform in total variation. We show that it is always possible to construct such a chain with mixing time  $\tau \lesssim \varepsilon^{-1} (\text{diam } G)^2 \log |V|$ . Finally, we discuss analogous questions for continuous-time and time-inhomogeneous chains.

**Keywords** Mixing time · Random walks · Conductance · Fastest mixing Markov chain

**Mathematics Subject Classification** 05C81 · 60J10 · 60J20 · 60J27

---

Olesker-Taylor was supported by EPSRC Grant No. (EP/N004566/1).  
A preliminary version of this article appeared in ITCS'22 [35].

---

✉ Luca Zanetti  
lz2040@bath.ac.uk  
Sam Olesker-Taylor  
oleskertaylor.sam@gmail.com

<sup>1</sup> Department of Statistics, University of Warwick, Coventry, UK

<sup>2</sup> Department of Mathematical Sciences, University of Bath, Bath, UK

## Contents

1	Introduction	.....
2	Vertex conductance and the optimal spectral gap	.....
3	Almost mixing	.....
4	Continuous-time markov chains	.....
5	Time-inhomogeneous markov chains	.....
6	Open problems and concluding remarks	.....
	References	.....

## 1 Introduction

### 1.1 Fastest mixing Markov chain set-up and motivation

Sampling objects from a finite set is a basic primitive which has a myriad of applications. Sampling directly from such a set, however, may be computationally too expensive or even impossible, for example, if the objects are nodes of a distributed network. A common approach in these scenarios is to design a random walk (*RW*), or, more generally, a Markov chain with state space corresponding to the set from which we wish to sample and appropriate equilibrium distribution. Furthermore, to ensure our sampling procedure is computationally efficient, we desire our Markov chain to converge to equilibrium in a small number of steps, i.e., have fast mixing time.

This has wide-ranging applications: from shuffling cards [6, 15], to approximating statistical physics models [16, 25] and analysing load-balancing protocols in distributed computing [36, 44]. Furthermore, approximately sampling from the uniform distribution of a set can be used to estimate the size of the set itself [26]. This has been applied to approximating the permanent of a matrix [24, 27] and counting the number of independent sets [17], perfect matchings [18] and forests [4] in graphs.

Fundamental to these applications is a fast mixing time. Understanding in which instances fast mixing is achievable and what the intrinsic obstacles to fast mixing are is the focus of this paper. More precisely, we consider the following scenario.

- We are given a finite, undirected graph  $G = (V, E)$ : the vertex set represents the underlying state space, while the edge set  $E$  defines the transitions allowed.
- Our goal is to study the *fastest* mixing Markov chain satisfying these constraints.

We assume throughout that graphs are finite, undirected and connected.

This problem was originally introduced by Boyd, Diaconis and Xiao [12] as the *Fastest Mixing Markov Chain (FMMC)* problem. Specifically, by considering only reversible chains and optimising the *spectral gap* as a proxy for the mixing time, they recast the problem of finding the fastest mixing Markov chain on a graph as a convex optimisation problem. Analogously to Boyd, Diaconis and Xiao [12], we dedicate most of our attention to reversible, time-homogenous chains in discrete-time. We do, however, dedicate one section to questions in the continuous-time setting, first studied in [43], and one short final section to time-inhomogeneous chains. Compared with discrete-time chains, continuous-time and time-inhomogeneous chains are considerably more powerful, but perhaps less natural from an application viewpoint.

To be explicit and precise, a transition matrix  $P$  is *reversible* with respect to (w.r.t.)  $\pi$  if  $\pi(u)P(u, v) = \pi(v)P(v, u)$  for all  $u, v \in V$ . Results on the spectral gap below—both our own and those referenced—are always in the reversible set-up. We also restrict to *lazy* chains, ie chains with  $P(v, v) \geq \frac{1}{2}$  for all  $v \in V$ . This is without loss of generality, since we are interested in maximising the spectral gap and this restriction costs a factor of at most  $\frac{1}{2}$  in the optimal spectral gap.

There are a variety of choices for how “convergence to equilibrium” is measured. It is typically measured in the total variation (TV), or equivalently  $\ell_1$ , distance. Other popular measures, particularly in the statistics literature, include  $\ell_2$ , or  $\chi^2$ , distance and relative entropy, or Kullback–Leibler divergence. We also recall that  $\ell_2$  is equivalent to  $\ell_\infty$ , or uniform, distance for reversible chains.

Nevertheless, no matter which one of these measure we choose, the long-term convergence to equilibrium of a lazy, reversible Markov chain is governed by its spectral gap  $\gamma_P$ . More precisely, given a transition matrix  $P$  of a lazy, reversible Markov chain, let  $d_P(t, x)$  denote the distance between  $P^t(x, \cdot)$  and its equilibrium distribution according to any of the aforementioned measures and let  $d_P(t) := \max_{x \in V} d_P(t, x)$ . Then,

$$d_P(t)^{1/t} \rightarrow \gamma_P \text{ as } t \rightarrow \infty.$$

See [31, Theorems 12.4 and 12.5] for details. The spectral gap thus determines the asymptotic convergence to equilibrium without having to select a specific measure.

We now define formally the class of reversible Markov chains on a graph and then the spectral gap and the relaxation and mixing times.

**Definition (Markov Chains on a Graph)** Let  $G = (V, E)$  be a graph and  $\pi$  a probability measure on  $V$ . We say that a Markov chain with transition matrix  $P$  is *on*  $G$  if  $P \in [0, 1]^{V \times V}$  and  $P(u, v) > 0$  implies either  $\{u, v\} \in E$  or  $u = v$ . We denote with  $\mathcal{M}(G, \pi)$  the set of lazy transition matrices on  $G$  which are reversible w.r.t.  $\pi$ .

**Definition (Spectral Gap, Relaxation Time and Mixing Time)** Let  $G = (V, E)$  be a graph and  $\pi$  a probability measure on  $V$ . Let  $P \in \mathcal{M}(G, \pi)$ . The *spectral gap* is  $\gamma_P := 1 - \lambda_P$ , where  $\lambda_P$  is the largest non-unitary eigenvalue of  $P$ . The *relaxation time* is  $1/\gamma_P$ . The (*uniform*) *mixing time* is  $\tau_P(\xi) := \inf\{t \geq 0 \mid d_P^\infty(t) \leq \xi\}$  for  $\xi \in [0, 1]$ , where  $d^\infty$  is the  $\ell_\infty$ -distance. Write  $\tau_P := \tau_P(\frac{1}{4})$ .

There is a standard relation between the relaxation and mixing times:

$$\gamma_P^{-1} \lesssim \tau_P \lesssim \gamma_P^{-1} \cdot \log \pi_{\min}^{-1};$$

see [31, Theorems 12.4 and 12.5] for details. The “ $\lesssim$ ” symbol hides a multiplicative universal constant; we use the symbols “ $\gtrsim$ ” and “ $\asymp$ ” similarly. Typically,  $\log \pi_{\min}^{-1} \asymp \log |V|$ . So, the relaxation time is a proxy for the mixing time, as well as characterising long-term convergence to equilibrium.

We are now finally ready to formally introduce the FMMC problem.

**Definition** (*Fastest Mixing Markov Chain*) Let  $G = (V, E)$  be a graph and let  $\pi$  be a probability measure on  $V$ . The *optimal spectral gap* is defined as

$$\gamma_{\pi}^*(G) := \max\{\gamma_P \mid P \in \mathcal{M}(G, \pi)\}.$$

The *optimal relaxation time* is  $1/\gamma_{\pi}^*(G)$ . We write  $\gamma^*(G)$ , omitting the  $\pi$ , when  $\pi = \mathcal{U}_V$  is uniform.

A transition matrix  $P$  is *fast mixing* if  $1/\gamma_P$  is polylogarithmic in  $|V|$ , asymptotically. Analogously, a graph  $G$  *admits a fast mixing chain* if  $1/\gamma^*(G)$  is polylogarithmic in  $|V|$ , asymptotically.

Previous work has been mainly focussed on finding useful formulations of the problem or on solving particular cases; see Sect. 1.4 for further details. The primary aim of our work, instead, is twofold:

- (i) To control the optimal spectral gap in terms of geometric barriers in the graph;
- (ii) To find ways to overcome these geometric barriers by slightly relaxing the FMMC problem.

Finally, we centre our attention on the case where  $\pi = \mathcal{U}_V$  is the uniform distribution on  $V$ . This case was also the main focus of the original series of papers studying the FMMC problem.

## 1.2 Main results

This article includes multiple avenues of study, all on the theme of finding fast mixing Markov chains. We introduce these and the main theorems that we prove in the following subsections.

### 1.2.1 Characterisation of fast mixing on graphs

We are looking for some natural statistic of the graph  $G$  which *characterises* fast mixing: we desire necessary and sufficient conditions for  $1/\gamma^*(G)$  to be ‘small’, namely polylogarithmic in  $|V|$ .

How well-connected a graph is should, intuitively, influence how fast a chain on the graph can mix. Thus, we would like to understand what kind of connectivity measure best characterises fast mixing. A natural candidate is the *edge conductance*  $\Phi^*$  of a graph, which is defined as follows.

**Definition A.1** (*Edge Conductance*) The *edge conductance*  $\Phi^*(G)$  of a graph  $G = (V, E)$  is

$$\Phi^*(G) := \min_{S \subseteq V: 0 < \text{vol}(S) \leq \text{vol}(V)/2} \Phi(S) \text{ where } \Phi(S) := |E(S, S^c)| / \text{vol}(S) \text{ for } S \subseteq V,$$

where  $E(S, S^c)$  is the *edge boundary* of  $S \subseteq V$  and  $\text{vol}(S)$  is the *volume* of  $S \subseteq V$ :

$$E(S, S^c) := \{\{x, y\} \in E \mid x \in S, y \notin S\} \text{ and } \text{vol}(S) := \sum_{x \in S} \deg(x).$$

It is well-known the edge conductance  $\Phi^*$  characterises the spectral gap of the lazy random walk (abbreviated *RW*)  $P^{RW}$  on  $G$  via the discrete Cheeger inequality, discovered in [23, 30]:

$$\Phi^*(G)^2 \lesssim \gamma_{P^{RW}} \lesssim \Phi^*(G).$$

The lazy *RW* on a graph, however, does not have uniform equilibrium distribution, unless the graph is regular. For this reason, it is natural to consider the *uniform*, or *maximum degree*, *RW*  $P^U$  which is defined by adding the appropriate number of self-loops to each vertex so that the graph becomes regular. A simple calculation with the Dirichlet characterisation gives

$$\gamma_{P^U} \leq \gamma^*(G) \leq 2d_{\max}\gamma_{P^U};$$

see [12, §7.2] for details. Applying this along with the discrete Cheeger inequality gives

$$\Phi^*(G)^2 \cdot d_{\min}/d_{\max} \lesssim \gamma_{P^U} \leq \gamma^*(G) \lesssim d_{\max}\gamma_{P^U} \lesssim d_{\max}\Phi^*(G) \cdot d_{\max}/d_{\min}.$$

Fast mixing for low-degree graphs is thus characterised by the edge conductance  $\Phi^*(G)$ :

- (i)  $G$  admits a fast mixing chain if and only if  $P^U$  is fast mixing;
- (ii)  $P^U$  is fast mixing if and only if  $1/\Phi^*(G)$  is polylogarithmic in  $|V|$ .

Such a simple characterisation does not hold if  $d_{\max}$  is large. This may be slightly counter-intuitive at first: adding edges can only increase the optimum  $\gamma^*$ ; but the lower bound above gets worse as  $d_{\max}$  increases. A striking example is given by taking two cliques on  $n$  vertices and connecting them by a perfect matching; see Fig. 3 in §1.3. It is a regular graph with  $\Phi^* \asymp 1/n$ , but, as we will see later, it has  $\gamma^* \asymp 1$ . Informally,  $\gamma^* \asymp 1$  because we can replace the two cliques with two bounded degree expander graphs without overly damaging its connectivity properties. This shows that edge conductance is not the correct conductance measure for the FMMC problem.

This prompts us to consider an alternative notion of connectivity: the *vertex conductance*  $\Psi^*$ . It measures how well connected a set is by comparing the number of *vertices* in the boundary with its size. Contrastingly, edge conductance compares the number of edges in the boundary with the total number of edges inside the set.

**Definition A.2** (*Vertex Conductance*) The *vertex conductance*  $\Psi^*(G)$  of a finite graph  $G = (V, E)$  is

$$\Psi^*(G) := \min_{S \subseteq V: 0 < |S| \leq |V|/2} \Psi(S) \quad \text{where} \quad \Psi(S) := |\partial S|/|S| \quad \text{for} \quad S \subseteq V,$$

where  $\partial S$  is the *vertex boundary* of  $S \subseteq V$ :

$$\partial S := \{y \notin S \mid \exists x \in S \text{ s.t. } \{x, y\} \in E\}.$$

The example above in which two equisized cliques are connected by a perfect matching has vertex conductance  $\Psi^* \asymp 1$ . This agrees with our claimed optimal spectral gap  $\gamma^* \asymp 1$ .

Vertex conductance has been used to provide upper bounds on the time to spread a rumour in a graph and on the hitting times of RWs, by Giakkoupis [20] and Chandra et al [13], respectively, amongst others. Roch [38, Proposition 2] showed that vertex conductance represents a fundamental barrier to fast mixing:  $\gamma^*(G) \lesssim \Psi^*(G)$ . This can be seen directly via a simple calculation comparing the edge conductance of any reweighing of  $G$ , for which the RW on this weighted graph has uniform equilibrium distribution, with the vertex conductance.

The edge and vertex conductances are comparable for low-degree graphs:

$$\Phi^*(G) \leq \Psi^*(G) \leq d_{\max} \Phi^*(G).$$

Thus, the fact that the edge conductance  $\Phi^*(G)$  characterises fast mixing for low-degree graphs means that the same holds for the vertex conductance  $\Psi^*(G)$ :

$$\Psi^*(G)^2/d_{\max}^2 \lesssim \gamma_{P^U} \leq \gamma^*(G) \lesssim \Psi^*(G).$$

We remove this  $d_{\max}$  factor, at the cost of a  $\log |V|$  factor, thus showing that vertex conductance characterises the existence of a fast mixing chain for any graph. The graph of Fig. 3 in §1.3 shows this does not hold for the edge conductance.

**Theorem A** (Characterisation of Fast Mixing) *Let  $G = (V, E)$  be a finite graph. Then  $\gamma^*(G)$  satisfies*

$$\Psi^*(G)^2/\log |V| \lesssim \gamma^*(G) \lesssim \Psi^*(G).$$

*Thus, vertex conductance characterises fast mixing for any graph.*

The quadratic dependence on the vertex conductance in the lower bound is needed for graphs such as the cycle. This has optimal spectral gap  $\gamma^* \asymp 1/n^2$  and vertex conductance  $\Psi^* \asymp 1/n$ ; see [11]. We are not aware of a graph for which the  $\log |V|$  factor is needed, but we have reasons to believe that such a factor, or at least a factor  $\log d_{\max}$ , is necessary. We elaborate.

Louis, Raghavendra and Vempala [32] essentially showed that, under the so-called Small-Set Expansion Conjecture of Raghavendra and Steurer [37], for any  $\varepsilon > 0$ , there is no polynomial-time algorithm that can distinguish between  $\Psi^*(G) \leq \varepsilon$  and  $\Psi^*(G) \gtrsim \sqrt{\varepsilon \log d_{\max}}$  for any graph  $G = (V, E)$ . The optimal spectral gap  $\gamma^*(G)$  can be computed in polynomial time, so removing the logarithmic factor in Theorem A altogether would violate the Small Set Expansion Conjecture.

One of the most interesting aspects of the proof of Theorem A, given in Sect. 2, is that it does not directly relate the vertex conductance to the spectral gap. Rather, it relates a variational characterisation of the optimal spectral gap, due to Roch [38, Proposition 1], to a new connectivity measure for graphs which we introduce. We term it *matching conductance* and denote it  $\Upsilon^*$ . It is defined similarly to vertex conductance,

but it replaces the size of the vertex boundary of a set  $S$  in the numerator with the size of a maximum matching between  $S$  and  $S^c$  in  $E$ . A formal definition is given in Definition 2.1. It can be viewed as a measure of *fault tolerance* of a graph: a graph has small matching conductance if and only if we can remove a few vertices of the graph and split the graph into two large, disconnected subsets.

It turns out the matching conductance of a graph is essentially equivalent to its vertex conductance:  $\Upsilon^*(G) \asymp \Psi^*(G)$ , uniformly over all graphs  $G$ ; see Proposition 2.2. A specific set  $S$  of vertices, however, can have matching conductance  $\Upsilon(S)$  much smaller than its vertex conductance  $\Psi(S)$ . This fact makes using matching, rather than vertex, conductance essential in our proof.

### 1.2.2 Almost mixing

We introduced the FMMC problem to formalise our desire to construct a fast mixing Markov chain. Theorem A, however, implies there are certain graphs, namely those with small vertex conductance, for which this desire cannot be attained. It is then natural to ask if we can slightly relax the constraints we imposed to overcome this fundamental obstacle.

We answer this question affirmatively: we show that if the Markov chain is not required to have equilibrium distribution *exactly* uniform, but only sufficiently close to uniform, then all graphs with small diameter admit a fast-mixing Markov chain. Before formalising this claim, we gain some intuition by considering the following simple example, known as the *dumbbell graph*.

- Take two complete graphs  $H_{\pm} = (V_{\pm}, E_{\pm})$ , each on  $n$  vertices. Choose  $v_{\pm} \in V_{\pm}$ , respectively.
- Form  $D_{\star} = (V, E)$  by connecting both  $v_{\pm}$  to a single ‘external’ vertex  $v_{\star} \notin V_{+} \cup V_{-}$ :

$$V := V_{+} \cup V_{-} \cup \{v_{\star}\} \text{ and } E := E_{+} \cup E_{-} \cup \{\{v_{+}, v_{\star}\}, \{v_{-}, v_{\star}\}\}.$$

Since  $D_{\star}$  has vertex conductance equal to  $1/n$ , Theorem A implies that no chain with uniform equilibrium distribution can have relaxation time of smaller order than  $n$ .

In light of the above, we propose the following RW, described by a weighting on the edges of  $D_{\star}$ :

- give all edges which do not include any of  $\{v_{+}, v_{-}, v_{\star}\}$  unit weight;
- give the remaining edges weight  $\varepsilon n$ .

The RW takes steps with distribution proportional to the edge weights. It is straightforward to check that the equilibrium distribution induced is at most  $\varepsilon$  far from uniform in TV.

The fundamental barrier to fast mixing in  $D_{\star}$  is that any chain with uniform equilibrium gets stuck in one side of the graph for a time at least order  $n$  in expectation. Up-weighting the edges through the bottleneck means that the new RW transitions between the two sides with expected time order  $1/\varepsilon$ . This leads to a relaxation time order  $1/\varepsilon$ . This all comes at a cost of having invariant distribution  $\varepsilon$  far from uniform. Further details are given in Sect. 1.3.

We are able to generalise this construction to general graphs and general equilibrium distributions. The fast-mixing Markov chain we design is a RW on a carefully weighted breadth-first search (BFS) spanning tree, supplemented with self-loops. We establish an upper bound of  $12(\text{diam } G)^2/\varepsilon$  on the relaxation time when we allow the RW to have equilibrium distribution  $\varepsilon$ -far from uniform.

We now define precisely our set-up and formally state our result.

**Definition B** (*Almost-Mixed Distributions*) Let  $V$  be a finite set. Let  $\mathcal{D}(V)$  denote the set of positive probability distributions on  $V$ . For  $\varepsilon \in [0, 1]$  and  $\pi \in \mathcal{D}(V)$ , define

$$\mathcal{D}(\pi, \varepsilon) := \{\pi' \in \mathcal{D}(V) \mid \min_{x \in V} \pi'(x)/\pi(x) \geq 1 - \varepsilon\}.$$

In particular, if  $\pi' \in \mathcal{D}(\pi, \varepsilon)$ , then  $\|\pi' - \pi\|_{\text{TV}} \leq \varepsilon$ .

We actually establish a stronger result than the one described above. The above description says that *there exists some* reversible chain which is fast mixing: there exist  $\pi' \in \mathcal{D}(\pi, \varepsilon)$  and  $P \in \mathcal{M}(G, \pi')$  such that  $\gamma_P \gtrsim \varepsilon/(\text{diam } G)^2$ . We prove that *any* reversible chain can be perturbed into a fast mixing chain: for all  $\pi \in \mathcal{D}(V)$  and all  $P \in \mathcal{M}(G, \pi)$ , there exist  $\pi' \in \mathcal{D}(\pi, \varepsilon)$  and  $Q \in \mathcal{M}(G, \pi')$  such that  $\gamma_Q \gtrsim \varepsilon/(\text{diam } G)^2$  and  $Q(e) \geq (1 - \varepsilon)P(e)$  for all  $e \in E$ .

**Theorem B** (Almost Mixing) *Let  $G = (V, E)$  be a finite, connected graph and  $\pi \in \mathcal{D}(V)$ . Let  $\varepsilon \in (0, 1)$  and  $P \in \mathcal{M}(G, \pi)$ . There exist  $\pi' \in \mathcal{D}(\pi, \varepsilon)$  and  $Q \in \mathcal{M}(G, \pi')$  such that*

$$\gamma_Q \geq \frac{1}{12}\varepsilon(\text{diam } G)^{-2} \quad \text{and} \quad Q(e) \geq (1 - \varepsilon)P(e) \quad \text{for all } e \in E.$$

A consequence of this spectral gap estimate is that

$$\tau_Q \leq 24\varepsilon^{-1}(\text{diam } G)^2 \log(16\pi_{\min}^{-1}).$$

The matrix  $Q$  is obtained as a perturbation of  $P$ . Moreover, this perturbation is actually independent of  $P$ : we construct a weighted BFS tree and mix it with the weights from  $P$ . A more refined statement, making this independence of the perturbation explicit, is given in Theorem 3.1.

Many algorithms on graphs work well when applied to graphs with very good connectivity properties, such as expanders. A way to utilise these algorithms for non-expanders is to perform an expander decomposition: the graph is partitioned into disjoint expanders and few edges connecting the expanders [42]. An alternative approach, recently proposed by Bernstein, Gutenberg and Saranurak [7], re-weights the vertices of the graph in order to obtain a type of vertex expander.

Our approach is in a very similar vein: we re-weight the edges of the graph to obtain a very good edge expander, whilst increasing the total weight as little as possible. We have not checked all the details, and their definition of vertex expansion is slightly unusual, but it seems highly plausible that our algorithm, or a minor adjustment of it, could be applied in their set-up.



This diameter bound is a substantial improvement over the vertex conductance lower bound on the optimal spectral gap from Theorem A. It comes at the cost of having invariant distribution only ‘almost’ uniform—hence the name “almost mixing”. We show in the next section that passing to the continuous-time setting allows this diameter-squared bound to be maintained while having *exactly* uniform invariant distribution. We use fundamentally the same chain: it is a RW on the same weighted BFS tree, where the weights now represent the rate at which an edge is crossed.

### 1.2.3 Continuous-time markov chains

The discussion and results above all concern *discrete-time* Markov chains. It is also natural to study the question of the fastest mixing Markov chain in *continuous-time*. We restrict to the case where the target distribution  $\pi = \mathcal{U}_V$ . The question of the FMMC in continuous-time was originally raised by Sun, Boyd, Xiao and Diaconis [43] and has been studied subsequently by Sammer [39] and Montenegro and Tetali [34]. We review their work in Sect. 1.4.

A continuous-time Markov chain on a graph  $G = (V, E)$  with uniform equilibrium distribution can be represented by the RW on a weighted graph  $(G, q)$ , where  $q : E \rightarrow \mathbb{R}_+$ , as follows.

**Definition C.1** (RW on Weighted Graph) Let  $G = (V, E)$  be a graph and  $q : E \rightarrow \mathbb{R}_+$  a collection of non-negative weights. The RW on  $(G, q)$  jumps from  $x$  to  $y$  at rate  $q(\{x, y\})$  for  $x, y \in V$  with  $\{x, y\} \in E$ . The *Laplacian*  $L^q \in \mathbb{R}^{V \times V}$  of the weighting  $q$  is defined by

$$L^q_{x,y} := \mathbf{1}\{\{x, y\} \in E\} \cdot q(\{x, y\}) - \mathbf{1}\{x=y\} \sum_{z \in V: \{x,z\} \in E} q(\{x, z\}) \quad \text{for } x, y \in V.$$

The *spectral gap*, which we denote  $\gamma_q$ , is given by the second smallest eigenvalue of  $L^q$ .

The spectral gap  $\gamma_q$  is intrinsically related to the mixing time  $\tau_q$ , as in discrete-time:

$$\gamma_q^{-1} \lesssim \tau_q \lesssim \gamma_q^{-1} \cdot \log |V| \quad \text{where } \tau_q := \tau_q(\tfrac{1}{4});$$

see [1, Lemma 4.23]. Again, *fast mixing* means relaxation time  $\gamma_q^{-1}$  polylogarithmic in  $|V|$ . The above relations thus imply that a polylogarithmic relaxation time characterises fast mixing.

It is immediate to see that if all the rates are multiplied by some factor, then the spectral gap changes by that factor too:  $\gamma_{cq} = c\gamma_q$  for any  $c > 0$ . We must therefore impose some normalisation.

**Definition C.2** (Normalisation) The rate at which the walk leaves the vertex  $x$  is given by

$$q(x) := \sum_{y \in V} \mathbf{1}\{\{x, y\} \in E\} q(\{x, y\}) \quad \text{for } x \in V.$$

We call  $\max_{x \in V} q(x)$  the *maximal* leave-rate and  $|V|^{-1} \sum_{x \in V} q(x)$  the *average* leave-rate.

A natural normalisation is to require a *maximal* leave-rate of 1. It can be seen, however, that this reduces to the discrete-time case via exponential-1 waiting times. We impose instead an *average* leave-rate of 1, or, equivalently,  $q(E) \leq \frac{1}{2}|E|$ . This allows a few vertices to have abnormally large leave-rate, but rarely enough that the average is not significantly affected. This will allow the RW to exit small ‘bottlenecks’ quickly, where the discrete-time walk would remain stuck for significant time. This average leave-rate normalisation was considered in [34, 39, 43]. Montenegro and Tetali [34, §7.1] describe this normalisation as “rather powerful [compared with discrete-time]” due to the fact that the maximal leave-rate may be very large.

The main result of this section states that, for any graph, it is possible to construct a weighting with average leave-rate of 1 such that its spectral gap depends only on the diameter of the graph.

**Theorem C** (Continuous-Time) *Let  $G = (V, E)$  be a graph. There exists a weighting  $w : E \rightarrow \mathbb{R}_+$  with average leave-rate at most 1 such that the RW on  $(G, w)$  satisfies*

$$\gamma_w \geq \frac{1}{16}(\text{diam } G)^{-2} \quad \text{and} \quad \tau_w \leq 8(\text{diam } G)^2 \log(16n).$$

An upper bound on  $1/\gamma^*(G)$  of order  $(\text{diam } G)^2$  is required for graphs with diffusive behaviour, such as the cycle or the path. A lower bound of order  $\text{diam } G$ , however, is *not* necessary, in general. This is in stark contrast to the discrete-time case. Indeed, in continuous-time, a few edges can be up-weighted significantly with little affect on the average. So if the ‘typical’ distance is much less than the maximal, a relaxation time of smaller order than the diameter may well be achievable.

A highly related theorem is given in Sammer’s PhD thesis [39, §3.3]; see also [34, §7.1]. They bound the optimal spectral gap  $\gamma^*(G)$  in terms of the *spread constant*  $c(G)$ , introduced in [3], which is the maximal variance of a function that is Lipschitz on the edges of  $G$ :

$$2c(G) \leq 1/\gamma^*(G) \lesssim c(G) \log |V|.$$

The spread constant  $c(G)$  can be upper bounded by  $\frac{1}{4}(\text{diam } G)^2$ , but there are examples for which this is far from tight. Still, if a very general, easy to calculate, bound is desired, then we do not know of a better bound than  $c(G) \lesssim (\text{diam } G)^2$ , which reduces to approximately our bound. The spread constant  $c(G)$  can also be lower bounded by a type of ‘typical’ distance; see [34, Corollary 7.2].

In contrast with our result, however, theirs is non-constructive, relying on the famous, but non-constructive, Johnson–Lindenstrauss lemma [28]. Montenegro and Tetali [34, Remark 7.3] comment on the difficulty of explicitly constructing such a process: “It might be challenging and in general impractical a task to actually find such a process explicitly.” Our construction is explicit and can actually be constructed in time linear in the size of the graph.

Montenegro and Tetali [34, Remark 7.3] also comment on the existence of such a fast mixing Markov chain in continuous-time: “The key [to the existence of such a

chain]... might be that we were implicitly providing the continuous-time chain with more power... by not requiring the rates in each row to sum to 1, but only the [average rate to be 1].” This significant additional power allows bottlenecks to be traversed quickly while maintaining an *average* leave-rate of 1. Indeed, the weighting  $w$  that we construct has  $\max_{x \in V} w(x) \asymp n / \text{diam } G$ , which may be far larger than 1.

This really emphasises the strength of our ‘almost mixing’ result, Theorem B: the chain there is in discrete-time —or, equivalently, has  $\max_{x \in V} q(x) \leq 1$ — but still attains a spectral gap only order  $\varepsilon$  smaller than that attained in the continuous-time case of Theorem C. Of course, the cost is that the equilibrium distribution  $\pi'$  only satisfies  $\min_{x \in V} \pi'(x) / \pi(x) \geq 1 - \varepsilon$ , not  $\pi' = \pi$ .

We expect that our continuous-time analysis can be adjusted to handle general equilibrium distributions  $\pi$  with relatively little changed. We have not checked the details, however. We focussed on the uniform case because it is, arguably, the most important and the cleanest to present.

### 1.2.4 Time-inhomogeneous markov chains

Our attention has been so far restricted to *time-homogeneous* Markov chains, in which the transition probabilities do not change over time and are described by a single transition matrix  $P$ . A *time-inhomogeneous* Markov chain, instead, is described by a sequence  $(P_t)_{t \in \mathbb{N}}$  of transition matrices and an initial law  $\mu_0$ : the time- $t$  law  $\mu_t := \mathbb{P}(X_t \in \cdot)$  is given by  $\mu_t = \mu_0 P_1 P_2 \cdots P_t$  for  $t \in \mathbb{N}$ . A *time-homogeneous* chain has  $P_t = P$  for all  $t \in \mathbb{N}$ , for some  $P$ . Nevertheless, we close our section of results by showing that they can lead to improvements over time-homogeneous chains.

**Theorem D** *Let  $G = (V, E)$  be a connected graph and let  $\pi \in \mathcal{D}(V)$ . There exists a time-inhomogeneous Markov chain on  $G$  that perfectly mixes to  $\pi$  after  $2 \text{ diam } G$  steps:  $\mu_{2 \text{ diam } G} = \pi$ .*

It is easy to see that  $\text{diam } G$  is a lower bound on the fastest ‘perfectly mixing’ chain. If one only requires  $\|\mathbb{P}(X_t \in \cdot) - \pi\|_{\text{TV}} < \frac{1}{2}$ , then  $\frac{1}{2} \text{ diam } G$  is a lower bound. Thus the bound of  $2 \text{ diam } G$  above is tight up to a factor of at most 4.

### 1.3 Notable examples

We discuss briefly a few examples which are of particular interest. We always consider the uniform distribution, i.e.  $\pi = \mathcal{U}_V$ , unless specified to the contrary.

#### 1.3.1 Dumbbell graph

Let  $D_\star$  be the dumbbell graph with bells  $H_\pm$  of size  $n$ . The bells  $H_\pm$  need not be cliques  $K_n$ ; they can be arbitrary connected graphs on  $n$  vertices. See Fig. 1 for an illustration when  $H_\pm = K_n$ .

*Conductance Measures.* It is straightforward to see that the set with the worst vertex conductance is given by one side of the dumbbell graph:  $S = H_-$  or  $S = H_+$ . This

shows that

$$\Psi^*(D_\star) = \Psi(H_\pm) = 1/n.$$

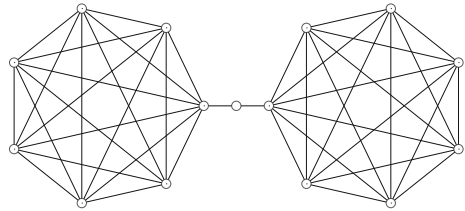
This implies that the optimal relaxation time  $1/\gamma^*(D_\star)$  satisfies  $n \lesssim 1/\gamma^*(D_\star) \lesssim n^2 \log n$ .

It is easy to find a chain attaining the correct order of  $n^{-1}$  when  $H_\pm = K_n$ , i.e. each bell is a complete graph on  $n$  vertices. Define a weighting as follows. Each vertex gets the same total weight.

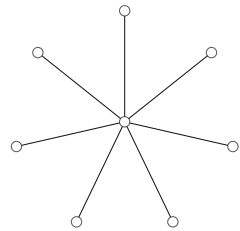
- Place unit weights on all edges which do not include the centre  $v_\star$ .
- Give the edges  $\{v_-, v_\star\}$  and  $\{v_+, v_\star\}$  weight  $n - 1$ .
- Add self-loops of weight  $2(n - 1)$  to each of the vertices in  $V \setminus \{v_-, v_+, v_\star\}$ .

The probability of stepping to  $v_\star$  from either of  $v_\pm$  is  $\frac{1}{2}$ . This gives an order- $n$  hitting time of one clique from the other. This implies that our chain has relaxation

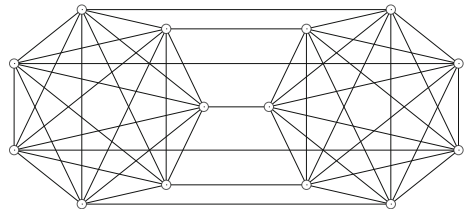
**Fig. 1** Dumbbell graph  $D_\star$  with  $n = 7$ : two cliques connected to a single external vertex



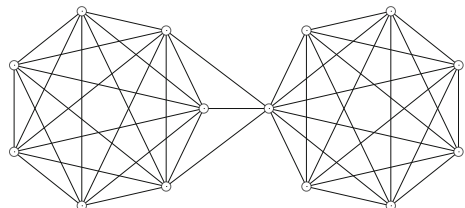
**Fig. 2** Star graph  $G_\star$  with  $n = 7$ : a central vertex connected to leaves



**Fig. 3** Matching graph  $\mathcal{M}$  with  $n = 7$ : two cliques connected via a matching



**Fig. 4** Source graph  $\Sigma$  with  $n = 7$  and  $k = 3$ : two cliques connected via a 'source'



time order  $n$ . Contrast this with the suboptimal order  $n^2$  hitting and relaxation time for the uniform RW.

*Almost Mixing.* If the graphs  $H_{\pm}$  have polylogarithmic diameter then Theorem B provides a chain with polylogarithmic relaxation time. This is a substantial improvement from linear. This is true regardless of the particular structure of  $H_{\pm}$ : it just needs  $\log \text{diam } H_{\pm} \lesssim \log \log n$ . If  $\text{diam } H_{\pm} \asymp 1$ , then we obtain relaxation time order 1, known also as an expander.

We now explain very roughly how to construct this chain for this dumbbell example  $D_{\star}$ . The general idea is to up-weight edges towards the central vertex  $v_{\star}$ , which is the bottleneck. We do this in such a way to make the distance to  $v_{\star}$  behave somewhat like an unbiased RW on  $\mathbb{Z}$ . This way it should take time order  $(\text{diam } H_{\pm})^2$  to move from  $H_{\pm}$  to  $H_{\mp}$ .

It is natural to try to achieve this bias by rooting a spanning tree  $T$  at  $v_{\star}$  and then up-weighting the vertices towards the root. This leads to a worst-case hitting time for the root  $v_{\star}$  of order  $(\text{diam } T)^2$ . We choose  $T$  to be a breadth-first search (BFS) tree since this has  $\text{diam } T \leq 2 \text{diam } G$ .

We give a more detailed overview in Sect. 3.2. We specifically chose the bottleneck vertex  $v_{\star}$  to be the root of  $T$  above. It turns out that actually any choice of root suffices. The reader may find this surprising at first; we did. More generally, suppose that  $o \in V$  is any vertex and a BFS is rooted at  $o$ ; we up-weight the edges towards  $o$ . Paths from  $v \neq o$  to  $o$  naturally go through bottlenecks. This automatically up-weights edges in bottlenecks.

### 1.3.2 Binary tree

Let  $\mathbb{T} = (V, E)$  be the complete binary tree on  $n = 2^N - 1$  vertices with depth  $N \approx \log_2 n$ .

*Conductance Measures.* It is straightforward to see that the set with the worst vertex conductance is given by one side of the tree: the root, a child and all its descendants. This gives

$$\Psi^*(\mathbb{T}) = 1/n.$$

This implies that the optimal relaxation time  $1/\gamma^*(\mathbb{T})$  satisfies  $n \lesssim 1/\gamma^*(\mathbb{T}) \lesssim n^2 \log n$ .

$\mathbb{T}$  has bounded degree, so the maximum degree chain attains the correct order relaxation time. This chain is just the simple RW, but with extra laziness at the leaves to make the invariant distribution uniform. The correct relaxation time is order  $n$ ; see [41] for details.

*Almost Mixing.* It is very natural to root the BFS tree at the root  $o$  of  $\mathbb{T}$ . The up-weighting will help pull the walk up the tree towards the root, allowing it to spread across the width of the tree more easily. The weight given to the edge from  $x \neq o$  to its parent is given by the number of vertices in the subtree rooted at  $x$ . Precisely, if  $x$  is at distance  $d \geq 1$  from the root, then the weight is  $2^{N-d} - 1$ . The up-weighting means that the distance from the root behaves roughly as an unbiased RW. The hitting time of the

root is then order  $N^2 \asymp (\log n)^2$ . Once the RW hits the root, which branch it takes after is uniformly distributed. Thus, once it hits the leaves again, it is uniform on the leaves and so approximately mixed. The total time for this is order  $N^2 \asymp (\log n)^2$ .

Our method does not know, however, that there is anything special about the root of the binary tree. Any vertex can be picked as the root of the BFS tree. Viewed from this vertex,  $\mathbb{T}$  is like a complete binary tree of some depth, but with some of the branches pruned. The depth is at most  $2N$ . The same ideas give relaxation and mixing time order  $N^2 \asymp (\log n)^2$ .

### 1.3.3 Star graph

Let  $G_\star = (V, E)$  be the star graph with centre  $v_\star$  and  $n$  leaves. See Fig. 2 for an illustration.

*Conductance Measures.* There is a simple dichotomy for the vertex boundary of a set  $S \neq \emptyset$  in the star graph  $G_\star$ : if  $v_\star \in S$ , then  $\partial S = S^c$ ; if  $v_\star \notin S$ , then  $\partial S = \{v_\star\}$ . Thus any  $S \subseteq V$  with  $|S| = \lfloor \frac{1}{2}n \rfloor$  and  $v_\star \notin S$  satisfies  $\Psi(S) = 1/|S| \asymp 1/n$ . It is straightforward to see that this gives the correct order: place unit weights on all the edges and weight- $(n-1)$  self-loops on all the non-central vertices; it is easy to see that this chain has mixing time order  $n$ .

Another measure of vertex conductance replaces the  $\partial S$  with the symmetric union  $\partial_{\text{sym}} S := \partial S \cup \partial S^c$ ; denote the vertex conductance with  $\partial_{\text{sym}}$  by  $\Psi_{\text{sym}}$ . Again, there is a simple dichotomy for  $S \neq \emptyset$ : if  $v_\star \in S$ , then  $\partial_{\text{sym}} S = S^c \cup \{v_\star\}$ ; if  $v_\star \notin S$ , then  $\partial_{\text{sym}} S = S \cup \{v_\star\}$ . Thus  $\Psi_{\text{sym}}^\star \asymp 1$ .

The difference between the two measures is that if the boundary of  $S$  is small, then that of  $S^c$  is large. The use of  $\partial S$ , as opposed to  $\partial_{\text{sym}} S$ , is thus important in Theorem A.

It is well-known that the spectral gap  $\gamma$  is characterised by a variational form. The relationship between spectral gap  $\gamma$  and the edge conductance  $\Phi^\star$  of the simple RW is given by the well-known Cheeger inequalities. A related variational form was introduced by Bobkov, Houdré and Tetali [8], which they denote  $\lambda_\infty$ . They establish various Cheeger-type relationships between  $\lambda_\infty$  and the vertex conductance  $\Psi_{\text{sym}}^\star$ . In particular, they show, for any graph, that

$$(\Psi_{\text{sym}}^\star)^2 \lesssim \lambda_\infty \lesssim \Psi_{\text{sym}}^\star;$$

see [8, Theorems 1 and 2]. It is immediate from the definitions that  $\lambda_\infty/d_{\max} \lesssim \gamma^\star \lesssim \lambda_\infty$ .

In light of [8] and our Theorem A, it is natural to wonder whether  $\lambda_\infty$  can be directly related to the optimal spectral gap  $\gamma^\star$ , without a  $d_{\max}$  factor. The example of the star graph  $G_\star$  shows that this is not possible:  $\Psi_{\text{sym}}^\star \asymp 1$  and thus  $\lambda_\infty \asymp 1$ , but  $\Psi^\star \asymp 1/n$  and so  $\gamma^\star \lesssim 1/n$ . This shows that  $\lambda_\infty$  is really not the correct parameter for the FMMC problem.

*Almost Mixing.* Obtaining an ‘almost mixing’ chain with order  $1/\varepsilon$  mixing time is simple: place  $\varepsilon$  weights on all the edges and weight-1 self-loops to all non-central vertices. The total weight of the central vertex is  $\varepsilon(n-1)$ . The remainder of the

weight is spread uniformly. Thus the distribution  $\pi$  induced by this weighting is in  $\mathcal{D}(G_\star, \varepsilon)$ . It is easy to see that the mixing time is order  $1/\varepsilon$ .

### 1.3.4 Complete graphs connected via a matching

Let  $\mathcal{M} = (V, E)$  be two  $n$ -cliques  $H_\pm = (V_\pm, E_\pm)$ , connected via a matching: enumerate  $V_\pm$  as  $\{v_\pm^1, \dots, v_\pm^n\}$  and connect  $v_+^i$  with  $v_-^i$  for all  $i \in \{1, \dots, n\}$ . See Fig. 3 for an illustration.

*Conductance Measures.* If  $v_+^i \in S$  but  $v_-^i \notin S$ , then  $v_-^i \in \partial S$ . From this,  $\Psi^\star(\mathcal{M}) \asymp 1$  follows easily. It is not too hard to show that the optimal spectral gap  $\gamma^\star(\mathcal{M})$  is of the same order: just replace the two cliques by two 3-regular expanders and leave the perfect matching in place. The lazy simple RW on this edge-induced subgraph of  $\mathcal{M}$  has order-1 spectral gap.

*Almost Mixing.* The optimal spectral gap is order 1, so there is no need for ‘almost mixing’.

### 1.3.5 Complete graphs connected via a ‘source’

Let  $\Sigma = (V, E)$  be two  $n$ -cliques  $H_\pm$  as ‘bells’, connected via a ‘source’: choose  $v_0 \in H_-$  and  $v_1, \dots, v_k \in H_+$ ; connect  $v_0$  with each of  $\{v_i\}_{i=1}^k$ . See Fig. 4 for an illustration.

*Conductance Measures.* One may think at first that this ‘source’ of  $k$  edges, rather than just a single edge, give rise to faster mixing; indeed,  $\Psi(H_-) = k/n$ . However, removing the source from the set gives  $\Psi(H_- \setminus v_0) = 1/(n - 1)$ . So in fact the vertex conductance of the source graph  $\Sigma$  is almost the same as that of the dumbbell graph  $D_\star$ .

The edge conductance of the graph does improve with  $k$ :  $\Phi(\Sigma) \asymp k/n^2$ . But this is always at most  $1/n$ . So the improvement from  $k$  is not enough to outweigh the fact that the uniform RW has spectral gap far from the optimal—unless  $k \asymp n$ .

An optimal spectral gap can be achieved by choosing an arbitrary 3-regular expander as a subgraph of each of the cliques and connecting these via a single edge. The uniform RW on this sparse subgraph then has spectral gap order  $1/n$ .

*Almost Mixing.* We can use exactly the same construction as in the dumbbell graph  $D_\star$ , picking an arbitrary edge amongst the  $k$  connecting edges.

## 1.4 Review of previous work

We now review previous related work. The FMMC question was originally introduced by Boyd, Diaconis and Xiao [12], which was the first in a series of articles [9–12, 43] by those authors along with Parrilo and Sun. It has subsequently been studied by [2, 14, 19, 21, 34, 38, 39]. We roughly collect these by theme.

### 1.4.1 Finding useful formulations

*Boyd, Diaconis and Xiao* [12]. This original work introduces the FMMC question and then primarily studies equivalent formulations. In our view, the most important contribution of that paper, beyond the introduction of the very interesting FMMC question, is their formulation of the FMMC optimisation problem as a semi-definite program (SDP). This allows the computation of an optimal solution in polynomial time via standard convex optimisation techniques. The SDP leads naturally to a dual formulation, which found use in subsequent work [34, 38, 39].

*Roch* [38]. Roch takes the dual formulation of [11] much further, writing the optimal spectral gap  $\gamma_\pi^*(G)$  as a minimisation of the variance of a certain constrained graph embedding. To quote him, “Informally, to obtain [a lower bound on the optimal spectral gap] we seek to embed the graph into  $\mathbb{R}^{|V|}$  so as to ‘spread’ the nodes as much as possible under constraints over the distances separating nodes connected by edges.” He re-derives the upper bound  $\gamma_\pi^*(G) \lesssim \Psi_\pi(G)$  using this formulation. This shows vertex conductance is a fundamental barrier to fast mixing. Our result shows that vertex conductance is essentially *the* fundamental barrier to fast mixing.

*Sun, Boyd, Xiao and Diaconis* [43]. The paper [43] is of a similar flavour to [12] but in the continuous-time set-up. We discuss it in detail in Sect. 1.4 below.

### 1.4.2 Special cases and particular examples

*Boyd, Diaconis, Sun and Xiao* [11]. The special case of the path with uniform distribution is studied in the short note [11], as a follow-on from [12]. They show that the ‘uniform chain’, i.e., the unbiased RW with  $\frac{1}{2}$ -holding at the ends, has the largest spectral gap.

*Boyd, Diaconis, Parrilo and Xiao* [9, 10]. The FMMC problem on graphs with rich symmetry properties is studied in [10]. They are able to solve various cases analytically: edge-transitive graphs, such as the cycle; Cartesian products of graphs, such as the two-dimensional torus and the hypercube; distance-transitive graphs, such as Petersen, Hamming and Johnson graphs. They then use algebraic methods to study FMMC on orbit graphs. This uses powerful representation theory arguments developed in [9].

*Cihan and Akar* [14]. Many similar scenarios, such as edge-transitive graphs, are studied in [14]. The focus is on two SDP methods. They study the degree-biased and uniform equilibria.

*Jafarizadeh and Jamalipour* [21]. Symmetric  $K$ -partite graphs and connections to sensor networks are considered in [21]. They compare numerically with a Metropolis–Hastings algorithm.

*Allison and Shader* [2]. Graphs which are overlapping unions of two cliques are studied in [2]. Here are two cliques, say of sizes  $r + s$  and  $r + t$ , respectively, and there are  $s$  overlapping vertices. The FMMC problem is solved analytically for such graphs.

*Fill and Kahn* [19]. A rather different approach is taken in [19]. Their paper is focussed on comparison inequalities and majorisation of measures. They use these to analyse



the FMMC problem. This use of majorisation allows them to study a distance, such as TV, separation or  $\ell_2$ , rather than the spectral gap, which is only a proxy for the mixing time.

### 1.4.3 Continuous-time set-up

*Sun, Boyd, Xiao and Diaconis* [43]. The study of the FMMC question in continuous-time was initiated in [43]. The structure and goals of this paper are similar to [12]. The primary contribution is a convex SDP formulation as well as some dual formulations.

Recall that a normalisation on the weights was required. Indeed, doubling all the weights doubles the spectral gap. We imposed an ‘‘average leave-rate of 1’’. A slightly more general ‘weighted average’ is considered in [43]. A number of physical interpretations of this normalisation are given.

*Sammer* [39] and *Montenegro and Tetali* [34]. The FMMC problem is considered by Sammer [39, §3.3]. It is referenced and discussed by *Montenegro and Tetali* [34, §7]. We discussed their work in detail immediately after Theorem C. We add a small caveat to that discussion.

*Montenegro and Tetali* [34, §7.1] claim to impose a scaling of  $q(V) \leq 1$  on their edge weightings  $q : E \rightarrow \mathbb{R}_+$ ; contrast this with our imposition of  $q(V) \leq n$ . Their scaling immediately implies that the relaxation time is at least order  $n$ ; this contradicts their theorem. There are a couple of other points where there seem to be issues with the scalings, in particular in application of results from [43]. It may be possible to rectify these issues, but we have not checked carefully.

## 1.5 Subsequent work

Following the release of our paper on arXiv in late 2021, two papers [22, 29] extending our work on the FMMC question were released, almost simultaneously, in early 2022.

*Kwok, Lau and Tung* [29]. We posed the open question of

$$\text{improving } \Psi^*(G)^2 / \log |V| \lesssim \gamma^*(G) \text{ to } \Psi^*(G)^2 / \log d_{\max} \lesssim \gamma^*(G)$$

in Theorem A and also proving an analogous result for general equilibrium distributions, rather than restricted to the uniform distribution. This was the inspiration for [29] (private communication). They solve both open problems and they also generalise our framework to optimal higher-order eigenvalues. The extension to non-uniform equilibrium distribution was the main challenge and, whilst their solution follows the framework of our proof, it required a significant number of new ideas.

*Jain, Pham and Vuong* [22]. The above improvement from  $\log |V|$  to  $\log d_{\max}$  was also established in [22] independently, but only in the set-up of a uniform equilibrium distribution.

## 2 Vertex conductance and the optimal spectral gap

This section is devoted to a proof of Theorem A. In Sect. 2.1, we define the matching conductance of a graph, which plays a central role in the proof of Theorem A. We also show in Proposition 2.2 that matching and vertex conductance of a graph differ by at most a universal constant factor. Sect. 2.2 contains the necessary notation and preliminaries needed in the proof of Theorem A. In Sect. 2.3, we relate the optimal spectral gap of a graph to its matching conductance. This relation is formalised in Theorem 2.10. Notice that Theorem 2.10 together with Theorem 2.2 directly imply Theorem A.

### 2.1 The matching conductance of a graph

A *matching* is a set of edges such that any pair of edges in the set do not share an endpoint. Given a set of (undirected) edges  $E$  together with a weight function  $w: E \rightarrow \mathbb{R}_{\geq 0}$ , a *maximum matching* for  $E$  is a matching with maximum total weight (if  $E$  is the edge set of an unweighted graph, we assume  $w$  is equal to one on  $E$ ). We denote with  $\nu(E)$  the weight of a maximum matching for  $E$ :

$$\nu(E) \triangleq \max_{\text{matching } F \subseteq E} \sum_{e \in F} w(e).$$

We can now define the *matching conductance* of a graph.

**Definition 2.1** Let  $G = (V, E)$  be a graph and  $\emptyset \neq S \subset V$ . The *matching conductance* of  $S$  is defined as

$$\Upsilon(S) \triangleq \frac{\nu(E(S, S^c))}{|S|}.$$

The matching conductance of  $G$  is defined as

$$\Upsilon^*(G) \triangleq \min_{S: 0 < |S| \leq |V|/2} \Upsilon(S).$$

The next proposition relates matching and vertex conductance.

**Proposition 2.2** Let  $G = (V, E)$  be a graph. Then, it holds that

$$\Upsilon^*(G) \leq \Psi^*(G) \leq 4\Upsilon^*(G).$$

**Proof.** The inequality  $\Upsilon^*(G) \leq \Psi^*(G)$  is obvious: for any  $S \subseteq V$ ,  $\nu(E(S, S^c))$  must be smaller than the size of the vertex boundary of  $S$ . Therefore,  $\Upsilon(S) \leq \Psi(S)$  for any  $S \subseteq V$ , which yields the inequality.

The proof of  $\Psi^*(G) \leq 4\Upsilon^*(G)$  is slightly more involved. In particular it's not true that  $\Psi(S) \lesssim \Upsilon(S)$  for any  $S \subseteq V$ . Figure 4 provides an example of a graph with a set with small matching conductance, but large vertex conductance. Nevertheless, the

worst vertex conductance of a set in a graph is related to the matching conductance of the graph. To prove this, consider  $|S| \leq |V|/2$  with  $\Upsilon(S) = \nu(E(S, S^c))/|S|$ . We can assume  $\Upsilon(S) \leq 1/4$ , otherwise  $\Upsilon^*(G) > 1/4 \geq \Psi^*(G)/4$ . Let  $M$  be a maximum matching for  $E(S, S^c)$ , that is  $|M| = \nu(E(S, S^c))$ , and  $V(M) \subset V$  be the set of vertices adjacent to edges in  $M$ . Now consider the set  $T = S \setminus V(M)$ . We claim  $T$  has small vertex conductance. To this end, consider  $\partial T$ . It holds that  $|\partial T| \leq |V(M)|$ . Indeed, any  $u \in S^c \setminus V(M)$  cannot be in  $\partial T$ , otherwise there would exist an edge between a vertex  $S \setminus V(M)$  and a vertex in  $S^c \setminus V(M)$ , and this would contradict the maximality of  $M$ . Therefore, since  $|V(M)| = 2|M|$  and  $|M| = \Upsilon(S) \cdot |S| \leq |S|/4$ , we have that

$$\Psi^*(G) \leq \frac{|\partial T|}{|T|} \leq \frac{2\nu(E(S, S^c))}{|S| - 2|M|} \leq \frac{2\nu(E(S, S^c))}{|S|/2} = 4\Upsilon(S) = 4\Upsilon^*(G). \quad \square$$

### 2.2 Definitions and preliminaries

Given a set of vertices  $V$ , together with a set of edges  $E$  on  $V$ , a *fractional matching* is a function  $f : E \rightarrow [0, 1]$  such that, for any  $v \in V$ ,  $\sum_{e \ni v} f(e) \leq 1$ . Moreover, the fractional matching number of  $E$ , denoted by  $\nu^*(E)$ , is the maximum total weight of a fractional matching for  $E$ :

$$\nu^*(E) = \max_{f: E \rightarrow [0,1]} \sum_{e \in E} w(e)f(e),$$

where the maximisation is over valid fractional matchings.

Notice that  $\nu^*(E)$  is the solution of a linear program which is a convex relaxation for  $\nu(E)$ . As such,  $\nu(E) \leq \nu^*(E)$ . A useful characterisation of  $\nu^*(E)$  is the following.

**Proposition 2.3** *The fractional matching number of  $E$ ,  $\nu^*(E)$ , is equal to the minimum of the following linear program.*

$$\begin{array}{ll} \min_{g: V \rightarrow \mathbb{R}_{\geq 0}} & \sum_{u \in V} g(u) \\ \text{subject to} & g(u) + g(v) \geq w(u, v) \quad \forall \{u, v\} \in E. \end{array}$$

**Proof** This simply follows from linear programming duality. □

With a slight abuse of notation we will also use  $\nu(G)$  and  $\nu^*(G)$  to denote the maximum (fractional) matching weight on the edge set of  $G$ .

Up until now we have considered only matchings in undirected graphs. For technical reasons, however, we will also need to consider matchings in directed graphs. Given a set  $\vec{E} \subseteq V \times V$  of directed edges, a *directed matching*  $\vec{M} \subseteq \vec{E}$  is a set of edges such that, if  $(u, v), (w, z) \in \vec{M}$  and  $(u, v) \neq (w, z)$ , then  $u \neq w$  and  $v \neq z$ . Alternatively, a directed matching can be seen as a subgraph where each vertex has indegree and outdegree at most one, whereas an undirected matching is a subgraph where each

vertex has degree at most one. Analogously to the undirected case, we denote with  $\nu(\vec{E})$  the weight of a maximum matching in  $\vec{E}$ .

**Definition 2.4** Given an undirected (weighted) graph  $G = (V, E, w)$ , an *orientation*  $\vec{G} = (V, \vec{E}, w)$  is a directed graph constructed by replacing each undirected edge  $\{u, v\} \in E$  with a directed edge  $(u, v)$  (with arbitrary orientation) having weight  $w(u, v)$ .

The next lemma relates the maximum matching weight in an undirected graph with the maximum matching weight of its orientation.

**Proposition 2.5** For any graph  $G = (V, E, w)$  and an orientation  $\vec{G}$ , it holds that  $\nu(\vec{G}) \leq 4\nu(G)$ .

**Proof** Let  $M \subseteq E$  be a matching returned by the greedy algorithm for finding a maximal matching on  $G$ , which works as follows: let  $e_1, \dots, e_m$  be an ordering of the edges of  $G$  such that  $w(e_1) \geq \dots \geq w(e_m)$ . Then, greedy incrementally construct  $M$  by adding  $e_i$  to it, for  $i = 1, \dots, m$ , as long as this operation maintains the property that  $M$  is a matching. Denote with  $e_{i_1}, \dots, e_{i_{|M|}}$  the edges of  $M$  ordered non-increasingly according to their weight.

Let  $\vec{M}^*$  be a maximum matching in  $\vec{G}$ . We upper bound its total weight as follows. Let  $\vec{M}_0 = \vec{M}^*$  and, for  $j = 1, \dots, |M|$ , let  $\vec{M}_j$  be the directed graph obtained from  $\vec{M}_{j-1}$  by removing all edges incident to one of the endpoints of  $e_{i_j}$ . Since  $M$  is maximal by construction,  $\vec{M}_{|M|}$  is empty. Moreover, at each iteration  $j$  we remove at most four edges, since there are at most four edges in  $\vec{M}^*$  that share an endpoint with  $e_{i_j}$ . Notice these edges cannot share an endpoint with  $\{e_1, \dots, e_{i_{j-1}}\}$  otherwise they would have been removed in a previous iteration. Therefore, they must all have weight less than or equal to  $w(e_{i_j})$ . This is because the matching  $\{e_1, \dots, e_{i_{j-1}}\}$  can be augmented by adding any one of these edges (or rather, their undirected equivalent) without breaking the property of it being a matching. But then, their weight must be less than or equal to  $w(e_{i_j})$ , since otherwise greedy would have chosen one of those instead of  $e_{i_j}$ . Hence, we have proved that  $4\nu(G) \geq \nu(\vec{M}^*)$ , from which the proposition follows.  $\square$

### 2.3 Matching conductance and the fastest mixing problem

The following result is due to Roch [38] and gives a variational characterisation of  $\gamma^*(G)$ . It follows from the fact that  $\gamma^*(G)$  can be expressed as the solution to a semidefinite program for which strong duality holds.

**Proposition 2.6** Let  $G = (V, E)$  be a graph of  $n$  vertices. Then,  $\gamma^*(G)$  is equal to the minimum of the following optimisation problem.

$$\min_{\substack{f: V \rightarrow \mathbb{R}^n \\ g: V \rightarrow \mathbb{R}_{\geq 0}}} \frac{\sum_{u \in V} g(u)}{\sum_{u \in V} \|f(u)\|^2}$$

$$\begin{aligned} \text{subject to} \quad & \sum_{u \in V} f(u) = 0 \\ & g(u) + g(v) \geq \|f(u) - f(v)\|^2 \quad \forall \{u, v\} \in E. \end{aligned}$$

**Remark 2.7** The variational characterisation actually given by Roch [38] doesn't include a non-negativity constraint for the function  $g$ . The function  $g$ , however, needs to be non-negative whenever, as in our case, Markov chains on  $G$  are allowed non-negative holding probabilities. More precisely, for any  $u \in V$  and  $P$  transition matrix of a Markov chain on  $G$ , if we allow  $P(u, u) > 0$ , then we need to require  $g(u) \geq 0$ .

The variational characterisation of  $\gamma^*(G)$  given by Proposition 2.6 requires minimising over  $n$ -dimensional embeddings of the vertices in the graph. It is often more convenient to work with one-dimensional embeddings. For this reason, we introduce the following parameter.

**Definition 2.8** Let  $G = (V, E)$  be a graph of  $n$  vertices. We denote with  $\gamma^{(1)}(G)$  the minimum of the following optimisation problem.

$$\begin{aligned} \min_{\substack{f: V \rightarrow \mathbb{R} \\ g: V \rightarrow \mathbb{R}_{\geq 0}}} & \frac{\sum_{u \in V} g(u)}{\sum_{u \in V} f(u)^2} \\ \text{subject to} & \sum_{u \in V} f(u) = 0 \\ & g(u) + g(v) \geq (f(u) - f(v))^2 \quad \forall \{u, v\} \in E. \end{aligned}$$

The following proposition shows that  $\gamma^{(1)}(G)$  is a  $O(\log n)$ -approximation of  $\gamma^*(G)$ .

**Proposition 2.9** Let  $G = (V, E)$  be a graph. It holds that

$$\gamma^*(G) \leq \gamma^{(1)}(G) \lesssim \log n \cdot \gamma^*(G).$$

The proof of this proposition uses a standard trick (see, e.g., Montenegro and Tetali [34]):

- (i) we apply the Johnson–Lindenstrauss lemma [28] to show that considering only  $O(\log n)$ -dimensional embeddings suffices to obtain a constant approximation for  $\gamma^*(G)$ ;
- (ii) we transform such  $O(\log n)$ -dimensional embedding into a one-dimensional embedding, but in doing so we will lose a  $O(\log n)$  factor.

**Proof.** The relation  $\gamma^*(G) \leq \gamma^{(1)}(G)$  follows trivially since computing  $\gamma^{(1)}(G)$  can be seen as minimising over the same set of  $n$ -dimensional embeddings as for  $\gamma^*(G)$ , with the additional constraint that only the first coordinate can be non-zero.

To prove the upper bound, let  $f: V \rightarrow \mathbb{R}^n, g: V \rightarrow \mathbb{R}_{\geq 0}$  be the minimiser achieving  $\gamma^*(G)$  in Proposition 2.6. Then, the Johnson–Lindenstrauss lemma ensures there

exists an embedding  $\tilde{f}: V \rightarrow \mathbb{R}^d$  such that  $d = O(\log n)$  and, for any  $u, v \in V$ ,

$$\frac{1}{2} \|f(u) - f(v)\|^2 \leq \|\tilde{f}(u) - \tilde{f}(v)\|^2 \leq \|f(u) - f(v)\|^2,$$

and

$$\frac{1}{2} \|f(u)\|^2 \leq \|\tilde{f}(u)\|^2 \leq \|f(u)\|^2.$$

Now let

$$i \triangleq \arg \max_{j \in \{1, \dots, d\}} \sum_{u, v \in V} (\tilde{f}(u)_j - \tilde{f}(v)_j)^2,$$

and define  $h: V \rightarrow \mathbb{R}$  as

$$h(u) = \tilde{f}(u)_i - \frac{1}{n} \sum_{v \in V} \tilde{f}(v)_i$$

for any  $u \in V$ .

By construction,

$$\sum_{u \in V} h(u) = \sum_{u \in V} \tilde{f}(u)_i - \sum_{u \in V} \frac{1}{n} \sum_{v \in V} \tilde{f}(v)_i = 0.$$

Moreover, for any  $u, v \in V$ ,

$$(h(u) - h(v))^2 = (\tilde{f}(u)_i - \tilde{f}(v)_i)^2 \leq \|\tilde{f}(u) - \tilde{f}(v)\|^2 \leq \|f(u) - f(v)\|^2.$$

Therefore,  $(h, g)$  is a feasible solution to the optimisation problem of Definition 2.8.

Finally,

$$\begin{aligned} \sum_{u \in V} h(u)^2 &= \frac{1}{2n} \sum_{u, v \in V} (h(u) - h(v))^2 \\ &= \frac{1}{2n} \sum_{u, v \in V} (\tilde{f}(u)_i - \tilde{f}(v)_i)^2 \\ &\geq \frac{1}{2nd} \sum_{u, v \in V} \|\tilde{f}(u) - \tilde{f}(v)\|^2 \\ &\geq \frac{1}{4nd} \sum_{u, v \in V} \|f(u) - f(v)\|^2 \\ &= \frac{1}{2d} \sum_{u \in V} f(u)^2, \end{aligned}$$

where we used the fact that both  $h$  and  $f$  are centred at zero. Therefore,

$$\gamma^{(1)}(G) \leq \frac{\sum_{u \in V} g(u)}{\sum_{u \in V} h(u)^2} \leq 2d \cdot \frac{\sum_{u \in V} g(u)}{\sum_{u \in V} f(u)^2} \leq 2d\gamma^*(G) \lesssim \log n \cdot \gamma^*(G). \quad \square$$

Suppose we fix  $f : V \rightarrow \mathbb{R}$  that minimises the optimisation problem above. Then, by Proposition 2.3,  $\gamma^{(1)}(G)$  can be seen as the fractional matching value of a graph  $G_f$  which is constructed from  $G$  by reweighing each edge  $\{u, v\}$  by  $(f(u) - f(v))^2$ . Together with Proposition 2.9, this hints towards a connection between the matching conductance of  $G$  and  $\gamma^*(G)$ . This connection is formalised in Theorem 2.10, which is the main result of this section.

**Theorem 2.10** *Let  $G = (V, E)$  be a graph. It holds that*

$$\Upsilon^*(G)^2 \lesssim \gamma^{(1)}(G) \lesssim \Upsilon^*(G).$$

Moreover, this implies that

$$\frac{\Upsilon^*(G)^2}{\log n} \lesssim \gamma^*(G) \lesssim \Upsilon^*(G).$$

The proof of Theorem 2.10 follows the standard template of the proof of the discrete Cheeger inequality. To upper bound  $\gamma^{(1)}(G)$  it suffices to construct test functions  $f, g$  from a set  $S$  minimising the matching conductance of  $G$ . The other direction is more complicated and, similarly to the case of the “hard direction” of the discrete Cheeger inequality, it requires using the function  $f$  that minimises  $\gamma^{(1)}(G)$  to construct *sweep sets* and analyse the matching conductance of such sets. Analysing the matching conductance of these sweep sets, however, is not as straightforward as analysing their edge conductance as in the proof of the standard discrete Cheeger inequality.

We split the proof of Theorem 2.10 in several lemmata. The first one, Lemma 2.11, relates the maximum matching of cuts in the graph to the maximum matching of a weighted directed graph appositely constructed.

**Lemma 2.11** *Let  $G = (V, E)$  be an unweighted undirected graph and let  $f : V \rightarrow \mathbb{R}_{\geq 0}$ . Let  $\vec{G}_f = (V, \vec{E}_f, w_f)$  be a directed weighted graph constructed as follows:*

- (i) *for any  $u, v \in V$ ,  $(u, v) \in \vec{E}_f$  if and only if  $\{u, v\} \in E$  and  $f(u) < f(v)$ ;*
- (ii) *for any  $(u, v) \in \vec{E}_f$ ,  $w_f(u, v) = f(v)^2 - f(u)^2$ .*

*For any  $t > 0$ , define  $S_t = \{u \in V : f(u)^2 > t\}$ . Then, it holds that,*

$$\int_0^\infty v(E(S_t, S_t^c)) dt \leq 2v(\vec{G}_f).$$

**Proof** For any  $t \in [0, \infty)$ , let  $M_t \subseteq E(S_t, S_t^c)$  be a matching achieving value  $v(E(S_t, S_t^c))$ . Notice that, for any  $t$ , there might be several distinct maximum matching

for  $E(S_t, S_t^c)$ : we just pick one of them arbitrarily. We have that

$$\int_0^\infty v(E(S_t, S_t^c)) dt = \int_0^\infty \sum_{e \in E} \mathbf{1}\{e \in M_t\} dt.$$

Notice that, for any edge  $\{u, v\} \in E$  with  $f(u) < f(v)$ ,

$$\{t \in [0, \infty) : \{u, v\} \in M_t\} \subseteq [f(u)^2, f(v)^2).$$

Let  $\vec{M} \subseteq \vec{E}_f$  output by an execution of greedy on  $\vec{G}_f$ , which works as follows. We first order the edges  $\vec{E}_f = \{e_1, \dots, e_m\}$  of  $\vec{G}_f$  such that  $w_f(e_1) \geq w_f(e_2) \geq \dots \geq w_f(e_m)$ . For  $i = 1, \dots, m$ , we incrementally construct  $\vec{M}$  by including  $e_i$  for  $i = 1, \dots, m$  as long as adding this edge doesn't break the property of  $\vec{M}$  being a directed matching. This is the same algorithm as the one described in the proof of Proposition 2.5, with the difference that we are now constructing a directed instead of an undirected matching.

Consider now  $\{u, v\} \in E$  such that  $f(u) < f(v)$  and  $\{u, v\} \in M_t$  for some  $t \geq 0$ . Then, there must exist an edge  $(u', v') \in \vec{M}$  such that  $u = u'$  or  $v = v'$  (since greedy outputs a maximal matching) and

$$w_f(u', v') = f(v')^2 - f(u')^2 \geq f(v)^2 - f(u)^2.$$

The inequality above holds because otherwise greedy would have picked  $(u, v)$  instead of  $(u', v')$ . This implies that  $[f(u), f(v)] \subseteq [f(u'), f(v')]$  and  $t \in [f(u'), f(v')]$ .

For any  $t \geq 0$ , let  $h_t : E \rightarrow E$  be the function that maps any  $\{u, v\} \in E$  such that  $f(u) < f(v)$  and  $\{u, v\} \in M_t$  to an edge  $(u', v')$  as above. Notice that for any edge  $(u', v') \in \vec{M}$  and any  $t \geq 0$ , there can be at most two edges in  $M_t$  that share an endpoint with  $(u', v')$ . Hence,  $|h_t^{-1}(u', v')| \leq 2$ .

Therefore, since  $\{t \in [0, \infty) : \{u, v\} \in M_t\} \subseteq [f(u)^2, f(v)^2]$  as noted above, we have

$$\begin{aligned} \sum_{\{u,v\} \in E} \int_0^\infty \mathbf{1}\{\{u, v\} \in M_t\} dt &\leq 2 \sum_{(u',v') \in \vec{M}} \int_0^\infty \mathbf{1}\{h_t^{-1}(u', v') \cap M_t \neq \emptyset\} dt \\ &\leq 2 \sum_{(u',v') \in \vec{M}} (f(v')^2 - f(u')^2) \\ &\leq 2v(\vec{G}_f). \end{aligned}$$

The lemma follows by observing that

$$\int_0^\infty v(E(S_t, S_t^c)) dt = \int_0^\infty \sum_{e \in E} \mathbf{1}\{e \in M_t\} dt = \sum_{e \in E} \int_0^\infty \mathbf{1}\{e \in M_t\} dt \leq 2v(\vec{G}_f),$$



where we can swap the signs of integration and summation since the matchings  $\{M_t : t \geq 0\}$  can be chosen so that we need to consider only at most  $n - 1$  different matchings (since there are at most  $n - 1$  different sets  $S_t$ ), which implies that the integral can actually be computed as a finite sum.  $\square$

The next lemma shows how to construct a set of small matching conductance given a “good” non-negative function  $f : V \rightarrow \mathbb{R}_{\geq 0}$ .

**Lemma 2.12** *Let  $G = (V, E)$  be a graph and  $f : V \rightarrow \mathbb{R}_{\geq 0}$  such that  $0 < |\{u \in V : f(u) \neq 0\}| < |V|$ . Let  $\lambda$  be the minimum of the following optimisation problem.*

$$\begin{aligned} \min_{g : V \rightarrow \mathbb{R}_{\geq 0}} & \frac{\sum_{u \in V} g(u)}{\sum_{u \in V} f(u)^2} \\ \text{subject to} & g(u) + g(v) \geq (f(u) - f(v))^2 \quad \forall \{u, v\} \in E. \end{aligned}$$

Then, there exists a set  $S \subseteq \{u \in V : f(u) > 0\}$  such that  $\Upsilon(S) \leq 8\sqrt{2\lambda}$ .

**Proof.** Let  $S_t = \{u \in V : f(u)^2 > t\}$  for  $t \geq 0$ . We have that

$$\min_{t : 0 < |S_t| < |V|} \Upsilon(S_t) \leq \frac{\int_0^\infty \nu(E(S_t, S_t^c)) dt}{\int_0^\infty |S_t| dt}.$$

First notice the denominator is equal to

$$\int_0^\infty |S_t| dt = \sum_{u \in V} f(u)^2.$$

We now upper-bound the numerator. Let  $\vec{G}_f = (V, \vec{E}_f, w_f)$  be the directed weighted graph defined in Lemma 2.11:

- (i) for any  $u, v \in V$ ,  $(u, v) \in \vec{E}_f$  if and only if  $\{u, v\} \in E$  and  $f(u) < f(v)$ ;
- (ii) for any  $(u, v) \in \vec{E}_f$ ,  $w_f(u, v) = f(v)^2 - f(u)^2$ .

Notice that this is an orientation of a graph  $G_f = (V, E_f, w_f)$  where each directed edge  $(u, v) \in \vec{E}_f$  is replaced by  $\{u, v\} \in E$ . Therefore, by Lemma 2.11 and Proposition 2.5, we have that

$$\int_0^\infty \nu(E(S_t, S_t^c)) dt \leq 2\nu(\vec{G}_f) \leq 8\nu(G_f).$$

We now want to relate  $\nu(G_f)$  to  $\lambda$ . Let  $M$  be a maximum matching in  $G_f$ . By applying Cauchy–Schwartz and the triangle inequality,

$$\begin{aligned}
 v(G_f) &= \sum_{\{u,v\} \in M} w_f(u, v) \\
 &= \sum_{\{u,v\} \in M} |f(u)^2 - f(v)^2| \\
 &= \sum_{\{u,v\} \in M} |f(u) - f(v)| \cdot |f(u) + f(v)| \\
 &\leq \sqrt{\sum_{\{u,v\} \in M} (f(u) - f(v))^2} \sqrt{\sum_{\{u,v\} \in M} (f(u) + f(v))^2} \\
 &\leq \sqrt{\sum_{\{u,v\} \in M} (f(u) - f(v))^2} \sqrt{\sum_{\{u,v\} \in M} 2(f(u)^2 + f(v)^2)} \\
 &\leq \sqrt{\sum_{\{u,v\} \in M} (f(u) - f(v))^2} \sqrt{2 \sum_{u \in V} f(u)^2},
 \end{aligned}$$

where the first inequality follows from Cauchy–Schwartz, while the second from the inequality  $(a+b)^2 \leq 2a^2+2b^2$  for any  $a, b \in \mathbb{R}$ . Notice that  $\sum_{\{u,v\} \in M} (f(u) - f(v))^2$  can be interpreted as the weight of a matching in an undirected graph obtained from  $G$  by reweighing each edge  $\{u, v\}$  with weight  $(f(u) - f(v))^2$ . Therefore, we can apply Proposition 2.3 and the definition of  $\lambda$  to show that

$$\begin{aligned}
 &\sum_{\{u,v\} \in M} (f(u) - f(v))^2 \\
 &\leq \min_{g: V \rightarrow \mathbb{R}_{\geq 0}} \left\{ \sum_{u \in V} g(u) \mid g(u) + g(v) \geq (f(u) - f(v))^2 \ \forall \{u, v\} \in E \right\} \\
 &\leq \lambda \sum_{u \in V} f(u)^2.
 \end{aligned}$$

Putting all together, we obtain

$$\min_t \Upsilon(S_t) \leq \frac{\int_0^\infty v(S_t) dt}{\int_0^\infty |S_t| dt} \leq \frac{8v(G_f)}{\sum_{u \in V} f(u)^2} \leq \frac{8\sqrt{2\lambda} \sum_{u \in V} f(u)^2}{\sum_{u \in V} f(u)^2} \leq 8\sqrt{2\lambda}. \quad \square$$

We are now finally ready to prove Theorem 2.10.

**Proof of Theorem 2.10** We start by proving the “easy side” of the Cheeger-type inequality, i.e.,  $\gamma^{(1)}(G) \lesssim \Upsilon^*(G)$ . Let  $S \subset V$  such that  $|S| \leq |V|/2$  and  $\Upsilon(S) = \Upsilon^*(G)$ . Define  $f: V \rightarrow \mathbb{R}$  as

$$f(u) = \begin{cases} \frac{1}{\sqrt{2|S|}} & \text{if } u \in S \\ -\frac{1}{\sqrt{2|V \setminus S|}} & \text{if } u \notin S. \end{cases}$$

Let  $M$  be a maximum matching for  $E(S, S^c)$ , i.e.,  $|M| = \nu(E(S, S^c))$ . Denote with  $V(M)$  the set of vertices incident to  $M$ . Define  $g: V \rightarrow \mathbb{R}$  as

$$g(u) = \begin{cases} 2/|S| & \text{if } u \in V(M) \\ 0 & \text{otherwise.} \end{cases}$$

By construction,  $f, g$  satisfy the constraints of the optimisation problem of Definition 2.8. Moreover,  $\sum_{u \in V} g(u) = 2V(M)/|S| = 4\nu(E(S, S^c))/|S|$ , while  $\sum_{u \in V} f(u)^2 = 1$ . Therefore,

$$\gamma^{(1)}(G) \leq \frac{\sum_{u \in V} g(u)}{\sum_{u \in V} f(u)^2} = 4\Upsilon(S) = 4\Upsilon^*(G).$$

We now turn the attention to the ‘‘harder side’’, i.e.,  $\Upsilon^*(G)^2 \lesssim \gamma^{(1)}(G)$ . Let  $f, g: V \rightarrow \mathbb{R}$  minimise  $\gamma^{(1)}(G)$ . We cannot directly apply Lemma 2.12 with  $f$  since  $f$  is not non-negative and might be supported over the entire  $V$ . For this reason, we define two non-negative functions  $h^-, h^+: V \rightarrow \mathbb{R}_{\geq 0}$  as follows. Let  $c$  be the median of  $f$ , i.e., order the vertices in  $V$  such that  $f(u_1) \leq f(u_2) \leq \dots \leq f(u_n)$  and set  $c \triangleq f(u_{\lceil n/2 \rceil})$ . For any  $u \in V$ , define  $h^-(u) \triangleq \max\{0, -(f(u) - c)\}$  and  $h^+(u) \triangleq \max\{0, f(u) - c\}$ . If  $\sum_{u \in V} h^-(u)^2 \geq \sum_{u \in V} h^+(u)^2$ , we define  $h \triangleq h^-$ , otherwise  $h \triangleq h^+$ . We now apply Lemma 2.12 with  $h$ .

First notice that

$$\sum_{u \in V} h(u)^2 \geq \frac{1}{2} \sum_{u \in V} (f(u) - c)^2 \geq \frac{1}{2} \sum_{u \in V} f(u)^2,$$

since  $\sum_{u \in V} f(u) = 0$  because  $f$  is a feasible solution to the optimisation problem of Definition 2.8. Moreover, for any  $\{u, v\} \in E$ ,

$$(h(u) - h(v))^2 \leq (f(u) - f(v))^2 \leq g(u) + g(v).$$

We can then apply Lemma 2.12 with

$$\lambda \triangleq \frac{\sum_{u \in V} g(u)}{\sum_{u \in V} h(u)^2} \leq \frac{2 \sum_{u \in V} g(u)}{\sum_{u \in V} f(u)^2} = 2\gamma^{(1)}(G)$$

Therefore, there exists  $S \subseteq \{u \in V : h(u) > 0\}$  such that  $\Upsilon(S) \leq 8\sqrt{2\lambda}$ . Moreover, by construction the support of  $h$  has size at most  $|V|/2$ . Hence,

$$\Upsilon^*(G) \leq \Upsilon(S) \leq 8\sqrt{2\lambda} \leq 16\sqrt{\gamma^{(1)}(G)}.$$

□

### 3 Almost mixing

#### 3.1 Set-up and main result

The previous section was devoted to estimating mixing-type statistics for the FMMC problem: we controlled the maximal spectral gap  $\gamma_P$  amongst all transition matrices  $P$  on a given graph  $G = (V, E)$  which are reversible w.r.t. the uniform distribution in terms of the vertex conductance of  $G$ . The purpose of the current section is to relax the condition that the invariant distribution of  $P$ , which we denote  $\pi_P$ , is *exactly*  $\mathcal{U}_V$ : we allow  $\pi_P$  to be  $\varepsilon$ -far from uniform in TV. We show that this can allow a *significant* speed up in the mixing time versus requiring the invariant distribution to be exactly  $\pi$ : we explicitly construct a Markov chain with spectral gap order at least  $\varepsilon/(\text{diam } G)^2$ .

Recall that we write  $\mathcal{D}(V)$  for the set of positive probability distributions on a set  $V$  and

$$\mathcal{D}(\pi, \varepsilon) = \left\{ \pi' \in \mathcal{D}(V) \mid \min_{x \in V} \pi'(x)/\pi(x) \geq 1 - \varepsilon \right\} \\ \text{for } \pi \in \mathcal{D}(V) \text{ and } \varepsilon \in [0, 1].$$

Let  $u : E \rightarrow \mathbb{R}_+$ , let  $A, B, S \subseteq V$ , let  $x \in V$  and let  $E' \subseteq E$ . We use the following notation.

$$u(x) := \sum_{y \in V} u(\{x, y\}) \quad \text{and} \quad u(S) := \sum_{x \in S} u(x); \\ E(A, B) := \{ \{x, y\} \in E \mid x \in A, y \in B \} \quad \text{and} \quad u(E') := \sum_{e \in E'} u(e).$$

Define the transition matrix  $P_u \in [0, 1]^{V \times V}$  by

$$P_u(x, y) := u(\{x, y\})/u(x) \quad \text{for } x, y \in V \text{ with } x \neq y; \\ P_u(x, x) := 1 - \sum_{y \in V \setminus \{x\}} P_u(x, y) \quad \text{for } x \in V.$$

Abbreviate the spectral gap as  $\gamma_u := \gamma_{P_u}$ . Define the probability measure  $\pi_u : V \rightarrow [0, 1]$  by

$$\pi_u(x) := u(x)/u(V) \quad \text{for } x \in V.$$

$P_u$  is the transition matrix of the RW on the weighted graph and  $\pi_u$  is its invariant distribution. It is the unique invariant distribution if the edge set  $\{e \in E \mid u(e) \neq 0\}$  is connected.

The following theorem is a refinement of Theorem B.

**Theorem 3.1** ('Almost Mixing') *Let  $G = (V, E)$  be a graph and let  $\pi \in \mathcal{D}(V)$ . There exists an edge weighting  $w_1 : E \rightarrow \mathbb{R}_+$ , depending only on  $G$  and  $\pi$ , with unit total weight, i.e.  $w_1(V) = 1$ , and the following property. Let  $\varepsilon \in (0, 1)$ . Let  $P$  be a transition matrix on  $G$  which is reversible w.r.t.  $\pi$ ; it need not be irreducible. Let*

$w_0 : E \rightarrow \mathbb{R}_+$  be the unique edge weighting of  $G$  with  $P_w = P$  and  $w_0(V) = 1$ . Define the transition matrix  $Q$  via the superposition weighting  $w := w_0 + \varepsilon w_1$ :

$$Q := P_w \text{ where } u(e) := w_0(e) + \varepsilon w_1(e) \text{ for } e \in E.$$

Then  $\min_{x \in V} \pi_w(x)/\pi(x) \geq 1 - \varepsilon$ . Let  $Q' := \frac{1}{2}(I + Q)$  denote its lazification. Then

$$\begin{aligned} \gamma_{Q'} &\geq \frac{1}{48} \varepsilon (\text{diam } G)^{-2} \text{ and } \tau_{Q'}(\xi) \\ &\leq 24\varepsilon^{-1} (\text{diam } G)^2 \log(\xi^{-2} \pi_{\min}^{-1}) \text{ for all } \xi \in (0, 1). \end{aligned}$$

Immediate consequences of these properties are  $\|\pi_w - \pi\|_{TV} \leq \varepsilon$ ,

$$\begin{aligned} Q(e) &\geq (1 - \varepsilon)P(e) \text{ for all } e \in E \text{ and} \\ \tau_{Q'} &= \tau_{Q'}(\tfrac{1}{4}) \leq 24\varepsilon^{-1} (\text{diam } G)^2 \log(16\pi_{\min}^{-1}). \end{aligned}$$

We remark briefly on the ‘independence’ of the perturbation by  $\varepsilon w_1$ .

**Remark 3.2** (Independence of Perturbation) The weighting  $w = w_0 + \varepsilon w_1$  can be seen as a perturbation of  $w_0$  by  $\varepsilon w_1$ , since we are most interested in the case where  $\varepsilon$  is very small—indeed, we want the new equilibrium distribution to be very close to  $\pi$ . We emphasise that the perturbation weighting  $w_1$  does not depend on the base weighting  $w_0$ ; rather,  $w_1$  is a function only  $G$  and  $\pi$ .

We fix a graph  $G = (V, E)$  and a probability measure  $\pi \in \mathcal{D}(V)$  throughout this section. We do not always repeat these in statements below. Also, we write  $n := |V|$ .

We start by proving a slightly weaker statement. Assume that  $w_0$  corresponds to unit-self loops:

$$\begin{aligned} w_0(\{x, y\}) &= 0 \text{ for all } x, y \in V \text{ with } x \neq y \text{ and } w_0(\{v\}) \\ &= \pi(v) \text{ for all } v \in V. \end{aligned}$$

The corresponding transition matrix  $P_{w_0}$  is diagonal and thus reversible w.r.t. any measure. We then extend the argument to handle arbitrary initial weightings  $w_0$  in the Sect. 3.5.

### 3.2 Outline and proof given later results

We start by giving a very brief outline with cross-references to the results proved in the following subsections. We then flesh out this outline, giving a more detailed description.

**Outline of Proof: Very Brief** The proof has four key steps.

- (i) We construct a weighted spanning tree; see Definition 3.11.
- (ii) We control the difference between the invariant distribution of the RW on this weighted tree and the target distribution  $\pi$ ; see Lemma 3.12.

- (iii) We estimate the conductance of this weighted tree; see Proposition 3.13.  
 (iv) We relate its spectral gap and conductance using canonical paths; see Corollary 3.14.  $\square$

**Outline of Proof: More Detailed** We now flesh out the above details somewhat for  $\pi = \mathcal{U}_V$ .

- Let  $T = (V, F)$  be a BFS spanning tree of  $G$ , rooted at  $v_*$ . We choose a weighting  $w_* : F \rightarrow (0, \infty)$  such that the weights increase towards the root  $v_*$  in such a way that  $w_*$  has edge conductance order 1. We then rescale the  $w_*$  to get  $\tilde{w}_*$  with total weight  $\tilde{w}_*(V) = \varepsilon n$ .
- Define  $w : E \rightarrow \mathbb{R}_+$  by adding unit-weight self-loops to  $\tilde{w}_*$ . The total weight  $w(V) = (1 + \varepsilon)n$  and  $w(x) \geq 1$  for all  $x$ . Thus the invariant distribution  $\pi'$  satisfies  $\min_{x \in V} \pi'(x) \geq (1 - \varepsilon)/n$ .
- It remains to analyse the edge conductance of  $w$ , which is intimately related to the original total weight  $w_*(V)$ . We can choose the weights such that  $w_*(V) \asymp n \text{diam } G$ . We then apply a Cheeger-type inequality to deduce a spectral gap lower bound of order  $\varepsilon/(\text{diam } G)^2$ .

We now describe how to choose the weighting  $w_*$ . Let  $T_x \subseteq T$  denote the subtree rooted at  $x$  and consisting of all descendants of  $x$ . We choose the weight  $w_*(\{x, \text{prt}(x)\}) := |T_x|$ , where  $\text{prt}(x)$  is the (unique) parent of  $x$ , for  $x \neq o$ . This way, the conductance of a subtree  $T_x$  in the weighted tree  $(T, w_*)$  is precisely 1. We emphasise that this is in the weighted tree  $(T, w_*)$ . We need to rescale  $w_*$  and combine it with the unit-weight self-loops to get an approximately uniform weighting.

It turns out that  $w_*(F) \asymp n \text{diam } T \asymp n \text{diam } G$ . This then gives rise to a final conductance  $\Phi^* \asymp \varepsilon/\text{diam } G$ . The standard Cheeger inequality then gives  $\gamma \gtrsim \varepsilon^2/(\text{diam } G)^2$ . We improve this to  $\gamma \gtrsim \Phi^*/(\text{diam } T)$  by applying the canonical paths method, using the fact that  $T$  is a tree.

The proof for general  $\pi$  is very similar. One gives the self-loop at  $x$  weight  $\pi(x)$  and defines  $w_*(\{x, \text{prt}(x)\}) := \pi(T_x)$ . This is the natural extension. The same arguments follow through.  $\square$

### 3.3 Preliminaries

We introduce some preliminary material which is used throughout the proof, as well as in Sect. 4. The majority of it will be familiar to a reader well-versed in RWs and mixing time analysis.

First, we generalise the notation of *edge conductance* of the graph, herein abbreviated *conductance*. We introduced this in Definition A.1 for RWs on unweighted graphs, i.e. unit weights on all edges. Reversible RWs correspond to a general weighting  $u : E \rightarrow \mathbb{R}_+$ , as in the above notation. The (edge) conductance of a reversible RW is the (edge) conductance of that weighted graph.

**Definition 3.3** (Edge Conductance) Let  $G = (V, F)$  be a graph and  $u : F \rightarrow \mathbb{R}_+$  be a weighting. The *conductance*  $\Phi_u(S)$  of a set  $S \subseteq V$  with  $\pi_u(S) > 0$  w.r.t.  $u$  is defined to be

$$\Phi_u(S) := \frac{\pi_u(F(S, S^c))}{\pi_u(S)} = \frac{u(F(S, S^c))}{u(S)}.$$

The *conductance*  $\Phi_u^*$  of  $u$  is defined to be

$$\Phi_u^* := \min_{S \subseteq V: 0 < \pi_u(S) \leq 1/2} \Phi_u(S).$$

The *adjusted conductance*  $\tilde{\Phi}_u(S)$  of a set  $S \subseteq V$  with  $0 < \pi_u(S) < 1$  is defined similarly:

$$\tilde{\Phi}_u(S) := \frac{\pi_u(F(S, S^c))}{\pi_u(S)\pi_u(S^c)} \quad \text{and} \quad \tilde{\Phi}_u^* := \min_{S \subseteq V: 0 < \pi_u(S) < 1} \tilde{\Phi}_u(S).$$

**Remark 3.4a** (Conductance: Original and Adjusted Relations) The definition of the adjusted conductance does not need the restriction  $\pi_u(S) \leq \frac{1}{2}$  as  $\tilde{\Phi}_u(S)$  is invariant under complementation:

$$\tilde{\Phi}_u(S) = \tilde{\Phi}_u(S^c) \quad \text{for all } S \subseteq V \text{ with } 0 < \pi_u(S) < 1.$$

The following inequalities between  $\Phi_u$  and  $\tilde{\Phi}_u$  are immediate:

$$\begin{aligned} \Phi_u(S) &\leq \tilde{\Phi}_u(S) && \text{for all } S \subseteq V \text{ with } 0 < \pi_u(S) < 1; \\ \frac{1}{2}\tilde{\Phi}_u(S) &\leq \Phi_u(S) && \text{for all } S \subseteq V \text{ with } 0 < \pi_u(S) \leq \frac{1}{2}. \\ \frac{1}{2}\tilde{\Phi}_u^* &\leq \Phi_u^* \leq \tilde{\Phi}_u^* && \text{whatever the graph.} \quad \triangle \end{aligned}$$

**Remark 3.4b** (Conductance: Connectivity Assumption) We may assume that  $S$  induces a connected subset  $T[S]$  when analysing  $\Phi_u^*$ . Indeed, if  $S = A \dot{\cup} B$  with  $F(A, B) = \emptyset$ , then

$$\Phi_u(S) = \frac{u(F(A, A^c)) + u(F(B, B^c))}{u(A) + u(B)} \geq \min\{\Phi_u(A), \Phi_u(B)\},$$

using the fact that  $F(S, S^c) = F(A, A^c) \dot{\cup} F(B, B^c)$  and that

$$\frac{a + b}{a' + b'} \geq \min\left\{\frac{a}{a'}, \frac{b}{b'}\right\} \quad \text{for all } a, a', b, b' > 0. \quad \triangle$$

Next, we introduce the *canonical paths method* and use it to relate the spectral gap to the conductance in trees. A proof of Proposition 3.6 can be found in [40, Theorem 5].

**Definition 3.5** (Paths) Let  $G = (V, F)$  be a graph.  $\Gamma : \{0, \dots, L\} \rightarrow V$  is an  $F$ -path from  $x$  to  $y$  if

$$\Gamma(0) = x, \quad \Gamma(L) = y \quad \text{and} \quad \{\Gamma(\ell - 1), \Gamma(\ell)\} \in F \quad \text{for all } \ell \in [1, L].$$

The *length* of a path  $\Gamma : \{0, \dots, L\} \rightarrow V$  is defined to be  $|\Gamma| := L$ .

**Proposition 3.6** (Canonical Paths: General) *Let  $G = (V, E)$  be a connected graph and  $u : E \rightarrow \mathbb{R}_+$  be a weighting. Let  $\Gamma_{x,y}$  be an arbitrary  $F$ -path from  $x$  to  $y$  for  $x, y \in V$ . The spectral gap  $\gamma_u$  of the RW on  $(G, u)$  satisfies*

$$\gamma_u \geq \min_{e \in E} \{ (u(e)/u(V)) / \sum_{x,y \in V} \mathbf{1}\{e \in \Gamma_{x,y}\} \pi_u(x) \pi_u(y) / |\Gamma_{x,y}| \}.$$

**Corollary 3.7** (Canonical Paths: Trees) *Let  $T = (V, F)$  be a connected tree and  $u : F \rightarrow (0, \infty)$  be a weighting. The spectral gap  $\gamma_u$  of the RW on  $(T, u)$  satisfies*

$$\gamma_u \geq \tilde{\Phi}_u^* / \text{diam } T.$$

**Proof** Let  $\Gamma_{x,y}$  be the shortest path between  $x$  and  $y$  for all  $x, y \in V$ . Then  $|\Gamma_{x,y}| \leq \text{diam } T$  for all  $x, y \in V$ . Removing the edge  $e = \{e_-, e_+\} \in F$  disconnects the graph, leaving two components, with  $e_- \in V$  in one component and  $e_+ \in V$  in the other. Denote the component containing  $e_{\pm}$  by  $T_{\pm}^e$ . The canonical paths method (Proposition 3.6) then implies that

$$\gamma_u \geq \frac{1}{\text{diam } T} \cdot \min_{e \in F} \frac{u(e)/u(V)}{\pi_u(T_+^e) \pi_u(T_-^e)} = \frac{\tilde{\Phi}_u(T_+^e)}{\text{diam } T} \geq \frac{\tilde{\Phi}_u^*}{\text{diam } T}.$$

This uses the fact that  $T_+^e \dot{\cup} T_-^e = V$  and  $F(T_+^e, T_-^e) = e$  for all  $e \in F$ . □

**Remark 3.8** A more general statement is proved by Miclo [33, Theorem 1] (article in French). He does not require the graph  $T$  to be a tree at the cost of replacing  $\text{diam } T$  in the denominator of the bound by the longest path in  $T$ ; eg, if  $T$  has a Hamiltonian path, then the denominator becomes  $n - 1$ . He gives two proofs, one of which uses a canonical paths style argument.

Finally, we introduce some notation for trees and prove a counting lemma. This lemma, innocuous as it may appear, is fundamental to multiple calculations. The notation and definitions above were for any graph  $T = (V, F)$ . Assume that  $T$  is a tree for the rest of this preliminary section.

**Definition 3.9** (Tree Notation) *Let  $T = (V, F)$  be a tree rooted at  $o$ .*

- Let  $\text{anc}(z)$  denote the unique shortest path from  $z$  to the root  $o$ , including both  $z$  and  $o$ .
- Let  $V_y := \{z \in V \mid y \in \text{anc}(z)\}$  and  $T_y := T[V_y]$  denote the subtree rooted at  $y$ .
- Let  $\text{prt}(x)$  denote the parent of  $x \neq o$ , ie the unique neighbour  $y$  of  $x$  satisfying  $y \in \text{anc}(x)$ .

**Lemma 3.10** (Counting Weighted Subtrees) *For all measures  $\mu$  on  $V$  and all  $x \in V$ , we have*

$$\sum_{y \in T_x \setminus \{x\}} \mu(T_y) \leq \mu(T_x) \text{diam } T.$$



**Proof.**  $z \in T_y$  if and only if  $y \in \text{anc}(z)$ ; if  $y \in T_x \cap \text{anc}(z)$ , then  $z \in T_x$ . Thus

$$\begin{aligned} \sum_{y \in T_x \setminus \{x\}} \mu(T_y) &= \sum_{y \in T_x \setminus \{x\}} \sum_{z \in T_y} \mu(z) \\ &= \sum_{y \in V} \sum_{z \in V} \mu(z) \mathbf{1}\{y \in T_x \setminus \{x\}\} \mathbf{1}\{z \in T_y\} \\ &= \sum_{z \in V} \mu(z) \sum_{y \in V} \mathbf{1}\{y \in \text{anc}(z) \cap T_x \setminus \{x\}\} \mathbf{1}\{z \in T_x\} \\ &\leq \sum_{z \in T_x} \mu(z) \text{depth } T_x = \mu(T_x) \text{depth } T_x \leq \mu(T_x) \text{diam } T. \quad \square \end{aligned}$$

### 3.4 Construction via a weighted spanning tree

First, we define a weighted spanning tree  $(T, w)$ .

**Definition 3.11** (Weighted Spanning Tree) Let  $o \in V$  and let  $T = (V, F)$  be a BFS tree rooted at  $o$ . Supplement  $F$  with a self-loop at each vertex of  $V$ . Define the following weightings  $w_{0/1} : F \rightarrow \mathbb{R}_+$ :

$$\begin{aligned} w_0(\{x, y\}) &:= \pi(x) \mathbf{1}\{x = y\} && \text{for } x, y \in V; \\ w_1(\{x, \text{prt}(x)\}) &:= \pi(T_x) && \text{for } x \in V \setminus \{o\}. \end{aligned}$$

Define the weighting  $w : F \rightarrow \mathbb{R}_+$  via a linear combination:

$$w := w_0 + \eta w_1 \quad \text{where } \eta := \frac{1}{2} \varepsilon / \text{diam } T.$$

$T$  is a BFS tree, so  $\Delta := \text{diam } T$  satisfies  $\Delta \leq 2 \text{diam } G$ .

The distribution  $\pi_w$  induced by  $w$  is close to  $\pi$  in the following sense.

**Lemma 3.12** ( $\pi_w$  Close to  $\pi$ ) *The weighted tree  $(T, w)$  and its induced distribution  $\pi_w$  satisfy*

$$w(V) \leq 1 + \varepsilon \quad \text{and} \quad \min_{x \in V} \pi_w(x) / \pi(x) \geq 1 / (1 + \varepsilon) \geq 1 - \varepsilon.$$

**Proof** First, the subtree counting lemma (Lemma 3.10) implies that

$$\begin{aligned} w_1(V) &= 2w_1(F) = 2 \sum_{e \in F} w_1(e) = 2 \sum_{x \in V \setminus \{o\}} w_1(\{x, \text{prt}(x)\}) \\ &= 2 \sum_{x \in T_o \setminus \{o\}} \pi(T_x) \leq 2\pi(T_o) \Delta = 2\Delta. \end{aligned}$$

Trivially,  $w_0(V) = \sum_{x \in V} \pi(x) = 1$ . Thus

$$w(V) = w_0(V) + \eta w_1(V) \leq 1 + (\frac{1}{2} \varepsilon / \Delta) \cdot (2\Delta) = 1 + \varepsilon.$$

Second,  $w(x) \geq w_0(x) = \pi(x)$  and thus  $\pi_w(x) = w(x) / w(V) \geq \pi(x) / (1 + \varepsilon)$  for all  $x \in V$ . □

Next, we control the conductance of this weighted spanning tree.

**Proposition 3.13** (Conductance) *Let  $(T, w)$  be the weighted spanning tree from Definition 3.11. The conductance  $\Phi_w^*$  of the RW on  $(T, w)$  satisfies*

$$\Phi_w^* \geq \frac{1}{6}\varepsilon/\Delta.$$

**Proof.** First, suppose that  $o \notin S$ . Choose  $x \in S$  with  $\text{dist}(x, o)$  minimal. We may assume that  $T[S]$  is connected, by Remark 3.4b. These together imply that  $S \subseteq V_x$  and

$$\{x, \text{prt}(x)\} = F(T_x, T_x^c) \subseteq F(S, S^c).$$

This implies that

$$\Phi_w(S) = \frac{w(F(S, S^c))}{w(S)} \geq \frac{w(F(T_x, T_x^c))}{w(T_x)} = \Phi_w(T_x).$$

The definition of  $w$  gives

$$\Phi_w(T_x) = \frac{w(\{x, \text{prt}(x)\})}{w(T_x)} = \frac{\eta\pi(T_x)}{\pi(T_x) + \sum_{y \in T_x} \eta\pi(T_y)}.$$

The tree counting lemma (Lemma 3.10), then implies that

$$\Phi_w(T_x) \geq \frac{\eta\pi(T_x)}{\pi(T_x) + \eta\Delta\pi(T_x)} = \frac{\eta}{1 + \eta\Delta} = \frac{\varepsilon}{(2 + \varepsilon)\Delta} \geq \frac{\varepsilon}{3\Delta},$$

recalling that  $\eta = \frac{1}{2}\varepsilon/\Delta$ . We have thus shown that

$$\Phi_w(S) \geq \frac{\varepsilon}{3\Delta} \quad \text{for all } S \subseteq V \text{ with } o \notin S \neq \emptyset.$$

Importantly, this inequality does not require  $\pi_w(S) \leq \frac{1}{2}$ .

Next, suppose that  $o \in S$  and  $\pi_w(S) \leq \frac{1}{2}$ . The relations of Remark 3.4a imply that

$$\Phi_w(S) \geq \frac{1}{2}\tilde{\Phi}_w(S) = \frac{1}{2}\tilde{\Phi}_w(S^c) \geq \frac{1}{2}\Phi_w(S^c).$$

But now  $S' := S^c$  satisfies  $o \notin S'$  and  $S' \neq \emptyset$ . Thus the previous case implies that

$$\Phi_w(S^c) = \Phi_w(S') \geq \frac{1}{3}\varepsilon/\Delta.$$

Importantly, this case did not require  $\pi_w(S') \leq \frac{1}{2}$ . Combining gives

$$\Phi_w(S) \geq \frac{1}{6}\varepsilon/\Delta \quad \text{for all } S \subseteq V \text{ with } 0 < \pi_w(S) \leq \frac{1}{2}. \quad \square$$

Finally, we apply the canonical paths method for trees (Corollary 3.7) to deduce a bound on the spectral gap for  $(T, w)$ .

**Corollary 3.14** (Spectral Gap) *Let  $(T, w)$  be the weighted spanning tree from Definition 3.11. The spectral gap  $\gamma_w$  of the RW on  $(T, w)$  satisfies*

$$\gamma_w \geq \frac{1}{6}\varepsilon/\Delta^2 \geq \frac{1}{24}\varepsilon/(\text{diam } G)^2.$$

**Proof** This is an immediate consequence of the canonical paths method for trees (Corollary 3.7) and Proposition 3.13, along with the relations of Remark 3.4a. Also,  $\Delta \leq 2 \text{diam } G$ . □

We have now almost proved the main result. We just need to make sure the chain is lazy and convert the spectral gap result into a mixing time result.

**Proof of Theorem 3.1 when  $w_0 = \pi$**  The Markov chain constructed is a RW on a weighted BFS tree. It is defined in Definition 3.11. Denote the invariant distribution of the RW on this weighted tree by  $\pi'$ . Lemma 3.12 establishes the claim on the invariant distribution.

The spectral gap bound is proved via Proposition 3.13 and Corollary 3.14. Precisely, Corollary 3.14 defines a reversible chain  $Q$  satisfying  $\gamma_Q \geq \frac{1}{24}\varepsilon/(\text{diam } G)^2$ .

The mixing time bound will follow from the spectral gap bound via a standard mixing time–spectral gap relation. To apply this relation, we first pass from  $Q$  to its lazy version  $Q' := \frac{1}{2}(I + Q)$ . This ensures that the *spectral gap* and *absolute spectral gap* agree.  $Q$  and  $Q'$  have the same invariant distribution and that  $\gamma_{Q'} = \frac{1}{2}\gamma_Q$ . A simple calculation establishes the mixing time claim using the spectral–mixing relation; see [1, Lemma 4.23] for details of this relation. □

It remains to handle the case of general  $w_0$ , i.e. where  $w_0$  is any unit edge weighting with  $\pi$  as its induced invariant distribution. This is done in the next subsection.

### 3.5 Perturbation to arbitrary base chain

The analysis up to this point has shown the *existence* of a fast ‘almost mixing’ chain. Precisely, we defined a weighted graph by constructing an appropriately weighted BFS tree and supplementing it with  $\pi$ -weighted self-loops. We can think of the self-loops as a ‘base’ weighting which is reversible w.r.t.  $\pi$ . We denoted the ‘base’ weighting  $w_0$  and the ‘tree’ weighting  $w_1$ ; recall Definition 3.11.

We now explain how to extend this to an arbitrary ‘base’ weighting  $w_0$ . The analysis is extremely similar to that of the self-loops case above: we simply take an arbitrary base weighting  $w_0$  and superimpose on it the same weighted BFS tree. Some small adjustments are needed, but not many.

Let  $w_0 : E \rightarrow \mathbb{R}_+$  be an arbitrary unit edge weighting of  $E$ . Define  $E' := \{e \in E \mid w_0(e) \neq 0\}$ , the edge set of the graph induced by  $w_0$ .  $\pi$  is the invariant distribution of the  $w_0$ -weighted RW.

**Construction.** We define the weighted tree  $(T, w_1)$  exactly as before in Definition 3.11:  $T = (V, F)$  is an arbitrary BFS tree and  $w_1(\{x, \text{prt}(x)\}) = \pi(T_x)$  for  $x \in V \setminus \{o\}$ ; set  $w := w_0 + \varepsilon w_1$ . The proof of Lemma 3.12 is unchanged, showing that the induced distribution  $\pi_w$  is close to  $\pi$ .  $\square$

**Canonical Paths and Adjusted Conductance.** We can no longer use the canonical paths method for trees (Corollary 3.7) since the weighting does not necessarily give rise to a tree. The ‘extra edges’—i.e. those corresponding to non-self-loops in  $w_0$ —can only increase the conductance. Intuitively, these cannot harm mixing, but we must establish this carefully.

First, we adjust the proof of the canonical paths method for trees, i.e. the deduction of Corollary 3.7 from Proposition 3.6. We use the same canonical paths  $\Gamma$ , defined by paths in the BFS tree  $T$ . The bound on the spectral gap  $\gamma$  does not require a lower bound on  $\Phi(S)$  for arbitrary  $S$ ; rather, it only needs a lower bound on  $\Phi(T_x)$  for all  $x \in V$ .

The fact that  $E' = \emptyset$  when  $w_0$  is only self-loops meant that the set of edges emanating from the set  $T_x$  was given by  $F(T_x, T_x^c) = \{x, \text{prt}(x)\}$ . More generally, it is given by  $\{x, \text{prt}(x)\} \cup E'(T_x, T_x^c)$ . But this is always a superset of  $\{x, \text{prt}(x)\}$ , so the edge conductance is always larger than if the edges from  $E'$  were ignored. Motivated by this, define the following adjustment of edge conductance:

$$\widehat{\Phi}_w(T_x) := \frac{w(\{x, \text{prt}(x)\})/w(V)}{\pi_w(T_x)\pi_w(T_x^c)} \quad \text{for } x \in V \quad \text{and} \quad \widehat{\Phi}_w^* := \min_{x \in V} \widehat{\Phi}_w(T_x).$$

This is the edge conductance where only the boundary edges in the tree  $T = (V, F)$  are considered. The same proof as for canonical paths for trees then implies that

$$\gamma_w \geq \widehat{\Phi}_w^*/\Delta, \quad \text{recalling that } \Delta = \text{diam } T. \quad \square$$

**Conductance Analysis.** The analysis of the conductance in Proposition 3.13 needs to be adjusted. We need only analyse the conductance of complete subtrees  $T_x$  and must regard the boundary as only  $F(T_x, T_x^c) = \{x, \text{prt}(x)\}$ , not the full boundary  $(E' \cup F)(T_x, T_x^c)$ . The proof of Proposition 3.13 applies almost unchanged to control  $\widehat{\Phi}_w^*$ : we obtain

$$\widehat{\Phi}_w^* \geq \frac{1}{6}\varepsilon/\Delta.$$

The only point to be noted is the establishment of the equality  $w_0(T_x) = \pi(T_x)$ . Previously, this was obvious from the self-loop weightings. It still holds here, since the invariant distribution induced by  $w_0$  is  $\pi$  and  $w_0$  has unit total weight, by assumption. Thus, in fact,  $w_0(x) = \pi(x)$  for all  $x \in V$ .  $\square$

**Conclusion.** We combine the two results above, exactly as before, to obtain

$$\gamma_w \geq \widehat{\Phi}_w^*/\Delta \geq \frac{1}{6}\varepsilon/\Delta^2 \geq \frac{1}{24}\varepsilon/(\text{diam } G)^2.$$

The conversion of this into a lazy chain and then into a mixing estimate is unchanged.  $\square$

## 4 Continuous-time markov chains

### 4.1 Set-up, main result and outline

We have been studying discrete-time Markov chains throughout this paper. It is natural to ask the same question for continuous-time chains. Our attention is devoted to continuous-time Markov chains which are reversible w.r.t. the uniform distribution. Such chains can always be represented as a RW on a weighted graph  $(G, w)$  where  $w : E \rightarrow \mathbb{R}_+$  is a weighting on the edge of  $G = (V, E)$ .

Our main result for continuous-time chains is simple to state: we impose a normalisation of  $|V|^{-1} \sum_{e \in E} w(e)$ , ie the average rate at which the RW leaves a vertex is at most 1; we define a weighting  $w$  and show an upper bound of order  $(\text{diam } G)^2$  on the spectral gap of this RW.

**Theorem 4.1** (Fast Mixing Continuous-Time Markov Chain) *Let  $G = (V, E)$  be a graph. There exists a weighting  $w : E \rightarrow \mathbb{R}_+$  with average rate  $|V|^{-1} \sum_{x \in V} w(x) \leq 1$  and such that the Markov chain induced by this weighting has spectral gap  $\gamma_w$  and mixing time  $\tau_w(\cdot)$  satisfying*

$$\gamma_w \geq \frac{1}{16}(\text{diam } G)^{-2} \quad \text{and} \quad \tau_w(\xi) \leq 8(\text{diam } G)^2 \log(\xi^{-2}|V|) \quad \text{for all } \xi \in (0, 1).$$

**Proof of Theorem 4.1: Outline** The outline is the same as in discrete-time.

- (i) We construct a weighted spanning tree; see Definition 4.2.
- (ii) We control the total weight of the spanning tree; see Lemma 4.3.
- (iii) We estimate the conductance of this weighted tree; see Proposition 4.5.
- (iv) We relate its spectral gap and conductance using canonical paths; see Corollary 4.6. □

We fix a graph  $G = (V, E)$  and always take  $\pi := \mathcal{U}_V$  to be the uniform distribution on  $V$ . We do not always repeat these in statements below. Also, we write  $n := |V|$ .

### 4.2 Proof via adjustments to discrete-time case

The proof in continuous-time is surprisingly similar to that used in discrete-time.

- We construct the same weighted tree  $(T, w)$ , except that we do not include the self-loops; contrast Definitions 3.11 and 4.2.
- The invariant distribution of a RW on a graph with weights on the edges is always uniform. Thus we do not need an analogue of Lemma 3.12. We require the *total weight* to be at most  $n$ , instead of requiring the invariant distribution to be close to a given measure.
- We use the same argument to control the conductance; cf Proposition 3.13. The only differences is that now we do not include the self-loop weight in the calculation.
- The canonical paths argument applies in continuous-time; cf Corollaries 3.7 and 3.14.

First, we define the weighted spanning tree  $(T, w)$ ; cf Definition 3.11.

**Definition 4.2** (Weighted Spanning Tree) Let  $o \in V$  and let  $T = (V, F)$  be a BFS tree rooted at  $o$ . Define the following weightings  $w : F \rightarrow \mathbb{R}_+$  by

$$w(\{x, \text{prt}(x)\}) := \frac{1}{2}|T_x|/\text{diam } T. \quad \text{for } x \in V \setminus \{o\}.$$

$T$  is a BFS tree, so  $\Delta := \text{diam } T$  satisfies  $\Delta \leq 2 \text{diam } G$ .

This is equivalent to the tree-weight in the discrete-time case; see  $w_1$  in Definition 3.11. The particular scaling is chosen so that the total weight of  $w$  is at most  $n$ , as the next lemma shows.

**Lemma 4.3** (Total Weight of  $w$ ) *We have  $w(V) \leq n$ .*

*Proof.* This is an immediate consequence of the subtree counting lemma (Lemma 3.10):

$$w(V) = 2w(F) = 2 \sum_{x \in V \setminus \{o\}} w(\{x, \text{prt}(x)\}) = \sum_{x \in T_o \setminus \{o\}} |T_x|/\Delta \leq |T_o| = n. \quad \square$$

Next, we control the conductance of this weighted spanning tree; cf Proposition 3.13. To do this, we must first give the precise definition of *conductance* in continuous-time.

**Definition 4.4** (Conductance) Let  $T = (V, F)$  be a graph and let  $u : F \rightarrow \mathbb{R}_+$  be a weighting. The *conductance*  $\Phi_u(S)$  of a set  $S \subseteq V$  with  $\pi_u(S) > 0$  w.r.t.  $u$  is defined to be

$$\Phi_u(S) := u(F(S, S^c))/|S|.$$

The *conductance*  $\Phi_u^*$  of  $u$  is defined to be

$$\Phi_u^* := \min_{S \subseteq V: 0 < \pi_u(S) \leq 1/2} \Phi_u(S).$$

**Proposition 4.5** (Conductance) *Let  $(T, w)$  be the weighted spanning tree from Definition 3.11. The conductance  $\Phi_w^*$  of the RW on  $(T, w)$  satisfies*

$$\Phi_w^* \geq \frac{1}{4} \Delta^{-1}.$$

*Proof.* The same reductions as used in Proposition 3.13 show that it suffices to show that

$$\Phi_w(T_x) \geq \frac{1}{2} \Delta^{-1} \quad \text{for all } x \in V.$$

But this is immediate from the definition of  $w$ :

$$\Phi_w(T_x) = w(\{x, \text{prt}(x)\})/|T_x| = \frac{1}{2}|T_x|\Delta^{-1}/|T_x| = \frac{1}{2}\Delta^{-1}. \quad \square$$

Finally, we apply the canonical paths method for trees (Corollary 3.7) to deduce a bound on the spectral gap for  $(T, w)$ ; cf Corollary 3.14. We must adjust this to apply in continuous-time; see Proposition 4.7 and Corollary 4.8.

**Corollary 4.6** (Spectral Gap) *Let  $(T, w)$  be the weighted spanning tree from Definition 3.11. The spectral gap  $\gamma_w$  of the RW on  $(T, w)$  satisfies*

$$\gamma_w \geq \frac{1}{4} \Delta^{-2}.$$

**Proof** This is an immediate consequence of the canonical paths method for trees in continuous-time (Corollary 4.8) and Proposition 4.5. □

It remains to adjust the method of canonical paths to continuous-time.

**Proposition 4.7** (Canonical Paths in Continuous-Time: General) *Let  $G = (V, E)$  be a graph and  $u : E \rightarrow \mathbb{R}_+$  be a weighting. Let  $\gamma_{x,y}$  be an  $F$ -path from  $x$  to  $y$  for all  $x, y \in V$ . The spectral gap  $\gamma_u$  of the RW on  $(G, u)$  satisfies*

$$\gamma_u \geq n \min_{e \in E} \{ Q_u(e) / \sum_{x,y \in V} \mathbf{1}\{e \in \gamma_{x,y}\} |\gamma_{x,y}| \}.$$

**Proof** The discrete-time case is proved in [40, Theorem 5]. It involves the variational characterisation of the spectral gap in terms of the Dirichlet form. This characterisation holds both in discrete- and continuous-time; see [1, §3.6]. The proof in [40] then passes almost unchanged to the continuous-time set-up, recalling that now the invariant distribution is uniform.

Concretely, one can rescale the weights, setting  $\tilde{w}(\cdot) := cw(\cdot)$  for some value  $c$  such that  $\max_{x \in V} \tilde{w}(x) = 1$ . This can then be realised by placing mean-1 exponential wait times between jumps of a discrete-time chain  $P$ . One then applies the canonical paths method to  $P$ . The Dirichlet form is *linear* in this scaling meaning that the scaling can be ‘undone’ at the end. □

**Remark** (Cheeger-Inequality in Continuous-Time) We remark that while a scaling argument as used above does apply for the usual discrete-time Cheeger inequality, namely  $\gamma \geq \frac{1}{2}(\Phi^*)^2$ , the bound is *quadratic* in the scaling. Thus a factor of  $c$  is lost. See [1, Theorem 4.40]. In our set-up,  $\max_{x \in V} w(x)$  may be as large as  $(n - 1)/\Delta$ . This would lead to a lower bound of  $1/(n\Delta)$  on the spectral gap, rather than  $1/\Delta^2$  as we were able to achieve using canonical paths.

Analogous arguments to those used in the special case that  $G$  is a tree, ie deducing Corollary 3.7 from Proposition 3.6, apply in the continuous-time set-up too.

**Corollary 4.8** (Canonical Paths in Continuous-Time: Trees) *Let  $T = (V, F)$  be a tree and  $u : F \rightarrow (0, \infty)$  be a weighting. The spectral gap  $\gamma_u$  of the RW on  $(T, u)$  satisfies*

$$\gamma_u \geq \Phi_u^* / \text{diam } T.$$

We now have all the ingredients required to deduce the main result.

**Proof of Theorem 4.1** The Markov chain constructed is a RW on a weighted tree. It is defined in Definition 3.11. Lemma 4.3 bounds the total weight of this tree by  $n$ , as required.

The spectral gap bound is proved via Proposition 4.5 and Corollary 4.6. The mixing time bound is then deduced from the spectral gap bound via the (continuous-time) spectral gap–mixing time relation; see [1, Lemma 4.23] for details of this relation.  $\square$

### 4.3 Hitting time of the root

The following result is not needed for the proof, but it is a nice little result and its proof is extremely simple, given a reference regarding hitting times in trees. Write  $\tau_x$  for the hitting time of vertex  $x$ .

**Lemma 4.9** (Hit Root in Diameter Squared) *Let  $G = (V, E)$  be a graph. For all  $o \in V$ , there exists a weighting  $w : E \rightarrow \mathbb{R}_+$  with average rate  $\frac{1}{n} \sum_{x \in V} w(x) \leq 1$  and such that the Markov chain induced by this weighting has worst-case expected hitting time of  $o$  satisfying*

$$\max_{x \in V} \mathbb{E}_x(\tau_o) \leq 8(\text{diam } G)^2.$$

We use the weighted spanning tree  $(T, w)$  used above, i.e. from Definition 4.2. Precisely, the tree  $T = (V, F)$  is rooted at some vertex  $o \in V$  and

$$w(\{x, \text{prt}(x)\}) := \frac{1}{2}|T_x|/\text{diam } T \quad \text{for } x \in V \setminus \{o\}.$$

Moments of hitting times of the root in reversible Markov chains on trees were investigated by Zhang [45]. The following result is a special case of [45, Theorem 1.1].

**Theorem 4.10** (Hitting Times in Trees; cf [45, Theorem 1.1]) *Let  $T = (V, F)$  be a finite tree and  $q : F \rightarrow (0, \infty)$  a weighting. Let  $o \in V$  and root the tree at  $o$ ; use the notation of Definition 3.9. Let  $\tau_o$  denote the hitting time of the root. For all  $x \in V$ , we have*

$$\mathbb{E}_x(\tau_o) = \sum_{y \in \text{anc}(x) \setminus \{o\}} |T_y|/q(\{y, \text{prt}(y)\}).$$

The hitting time result of Lemma 4.9 follows easily from this.

**Proof of Lemma 4.9** Let  $(T, w)$  be the weighted spanning tree from Definition 3.11. We have  $|T_y|/w(\{y, \text{prt}(y)\}) = 2 \text{diam } T$  for all  $y$ . Thus applying Theorem 4.10 to this weighted tree gives

$$\mathbb{E}_x(\tau_o) = |\text{anc}(x) \setminus \{o\}| \cdot 2 \text{diam } T \leq 2(\text{diam } T)^2 \leq 8(\text{diam } G)^2,$$

since  $\text{diam } T \leq 2 \text{diam } G$ . The weighting satisfies  $w(V) \leq n$  by Theorem 4.3. Theorem 4.9 follows.  $\square$



**Remark 4.11** (Inspiration) This result of Zhang [45] was the inspiration behind our choice of weighted spanning tree in both the discrete- and continuous-time set-ups. Particularly, the simplicity of the formula when one takes  $q(\{x, \text{prt}(x)\}) := |T_x|$  encouraged us to try this weighting.

We had originally tried to make the distance to the root behave roughly like an unbiased RW on the integers—this gives the right “diameter-squared” bound. This means balancing the weights on either ‘side’ of a vertex. This works for some trees, e.g. the path, rooted at one end, and the binary tree. But it does not combine well: attaching a path and binary tree, each of the same depth, at their root gives rise to a “diameter-cubed” hitting time of the root.

Although we have phrased the continuous-time proof as an adjustment to the discrete-time one, we actually developed the continuous-time argument first. Indeed, this is natural because *any* edge weighting gives rise to the uniform distribution in continuous-time, so there is no need to do any superposition with a ‘base’ weighting, such as self-loops.

## 5 Time-inhomogeneous markov chains

The content of this section is somewhat different from the previous ones. Our desire, as always, is to sample from a distribution  $\pi$  on a set  $V$  via some Markov process which only uses transitions permitted by the graph  $G$ . The difference here is that we use a *time-inhomogeneous* Markov chain.

Markov chains are typically time-homogeneous. Discrete-time chains are then described by a transition matrix  $P$  and an initial law  $\mu_0$ . The time- $t$  law  $\mu_t$  is then given by applying  $P$   $t$  times to  $\mu$ :  $\mu_t = \mu_0 P^t$ . Continuous-time chains are described in a somewhat similar manner. Time-inhomogeneous chains are allowed to use a different transition matrix at each step:  $P_1$  is used for the first step,  $P_2$  for the second and so on. The time- $t$  law  $\mu_t$  is then given by  $\mu_t = \mu_0 P_1 \cdots P_t$ . The special case where  $P_t = P$  for all  $t$ , for some  $P$ , reduces to the time-homogeneous case.

Our main result for time-inhomogeneous chains is simple to state: given a graph  $G = (V, E)$  and  $\pi \in \mathcal{D}(V)$ , we exhibit a time-inhomogeneous Markov chain which satisfies  $\mu_{2 \text{diam } G} = \pi$ .

**Theorem 5.1** *Let  $G = (V, E)$  be a connected graph and let  $\pi \in \mathcal{D}(V)$ . There exists a time-inhomogeneous Markov chain on  $G$  obtaining perfect mixing after  $2 \text{diam } G$  steps:  $\mu_{2 \text{diam } G} = \pi$ .*

We fix a graph  $G = (V, E)$  and a probability measure  $\pi \in \mathcal{D}(V)$  throughout this section. We do not always repeat these in statements below.

**Proof of Theorem 5.1** Our argument proceeds by induction on the depth of the tree.

Choose  $o \in V$  and a breadth-first search (BFS) tree  $T = (V, F)$  rooted at  $o$  arbitrarily. Given  $x \in V$ , let  $V_x$  denote the set of vertices for which  $x$  lies in the unique shortest path to the root  $o$ ; let  $T_x := T[V_x]$ . Write  $\Delta(x) := \text{depth } T_x$  for the *depth* of  $T_x$ , i.e. the maximal distance  $\max_{y \in T_x} \text{dist}(y, x)$  to the *root* of  $T_x$ . We construct the time-inhomogeneous chain inductively

Suppose that  $X_0 = o$ ; we cover the general case later. We claim that there exists a chain  $X^o$  on  $T_o$  such that  $X_{\Delta(o)}^o \sim \pi|_{V_o}$ . This is trivially true if  $T_o$  is a singleton, which has depth 0. Assume now that  $|T_o| \geq 2$ . Define the first transition matrix  $P_1$  to keep  $X$  at  $o$  with probability  $\pi(o)$  and otherwise move to  $x$  with  $\{o, x\} \in F$  with probability proportional to  $\pi(T_x)$ :

$$P_1(o, o) := \pi(o) \quad \text{and} \quad P_1(o, x) := \pi(T_x) \quad \text{for } x \in V \text{ with } \{o, x\} \in V.$$

Let  $P_k(o, o) = 1$  for all  $k \geq 2$ . Thus if  $X_1 = o$  then  $X_k = o$  for all  $k \geq 1$ .

There exists a chain  $X^x$  on  $T_x$  such that  $X_{\Delta(x)}^x \sim \pi|_{V_x}$  for each  $x \in V$ , by the inductive hypothesis. Importantly,  $T_x \cap T_y = \emptyset$  if  $\{o, x\}, \{o, y\} \in F$  and  $x \neq y$ . We can thus define sequences  $P_x := (P_k^x)_{k=1}^{\Delta(x)}$  of transition matrices for each  $x \in V$  with  $\{o, x\} \in F$  and each  $P^x$  defined on a disjoint set by induction. Define  $P_k^x := I$  for  $k > \Delta(x)$ . These can then be combined into a single sequence  $(P_k)_{k=2}^{\Delta(o)}$ : if  $X_1^o = x$ , then we use sequence  $P^x$ . Then

$$X_{\Delta(o)}^o = X_{\Delta(x)}^x \sim \pi|_{V_x} \quad \text{conditional on } X_1^o = x \neq o,$$

noting that  $\Delta(x) = \Delta(o) - 1$ . Thus, for  $y \neq o$ ,

$$\begin{aligned} \mathbb{P}_o(X_{\Delta(o)}^o = y) &= \sum_{x \in V: \{o, x\} \in F} \mathbb{P}_x(X_{\Delta(x)}^x = y) \mathbb{P}_o(X_1^o = x) \\ &= \sum_{x \in V: \{o, x\} \in F} (\mathbf{1}\{y \in V_x\} \pi|_{V_x}(y)) \cdot \pi(V_x) \\ &= \sum_{x \in V: \{o, x\} \in F} \mathbf{1}\{y \in V_x\} \cdot (\pi(y)/\pi(V_x)) \cdot \pi(V_x) \\ &= \pi(y) \sum_{x \in V: \{o, x\} \in F} \mathbf{1}\{y \in V_x\} = 1/|T_o|. \end{aligned}$$

This argument is no more than a formalisation of the following informal verbal description.

- If the walk is at  $x$ , then stay at  $x$  with probability  $\pi(x)/\pi(V_x)$ .
- Otherwise, move to the children of  $x$  with probability proportional to their  $\pi$ -measure.
- If at some point the walk stays put, then keep it at that state indefinitely.

This completes the argument when  $X_0 = o$ . It remains to consider the case that  $X_0 \neq o$ . Direct all edges towards  $o$  and run for  $\Delta(o)$  steps. Precisely, set

$$P_0(x, y) := \mathbf{1}\{x \neq o, y = \text{prt}(x)\} + \mathbf{1}\{x = y = o\},$$

where  $\text{prt}(x)$  is the unique neighbour of  $x \neq o$  on the unique shortest path from  $x$  to  $o$ . If  $\Delta(o)$  steps are made according to this matrix, then  $X_{\Delta(o)} = o$ , regardless of  $X_o$ . We then apply the construction from the case  $X_0 = o$ , all shifted by  $\Delta(o)$ .

In summary, we obtain a time-inhomogeneous Markov chain  $X$  with  $X_{2\Delta(o)} \sim \pi$ . Finally,  $\Delta(o) = \text{depth } T \leq \text{diam } G$ . Keeping  $X$  fixed during  $(2\Delta(o), 2 \text{diam } G]$  gives  $X_{2 \text{diam } G} \sim \pi$ . □

## 6 Open problems and concluding remarks

We have studied fundamental barriers to fast mixing on graphs. We have essentially shown that three geometric quantities characterise the mixing properties of a graph: edge conductance characterises fast mixing for the simple random walk and vertex conductance characterises the fastest mixing time, whilst the diameter characterises the fastest almost-mixing time.

There are a few questions left open by our work. First, Theorem A implies that the fastest *mixing time* on a graph is at most order  $(\Psi^*)^{-2} \log^2 |V|$ . It is natural to ask if this  $\log^2 |V|$  factor can be improved to  $\log |V|$ . This improvement would require departing from the framework of our proof, at least if  $d_{\max} \gg 1$  is permitted, since we believe the optimal *relaxation time*  $1/\gamma^*$  can be of order  $(\Psi^*)^{-2} \log d_{\max}$ ; see the discussion after Theorem A.

Another open problem prompted by our work is to construct a graph sparsifier, i.e., an edge-induced sparse subgraph, that approximately preserves the vertex conductance of the original graph. Our work together with a result by Batson, Spielman and Srivastava [5] implies that it is possible to construct an order  $n$ -size sparsifier of  $G$  with vertex conductance at least order  $\Psi^*(G)^2 / \log n$ . It is then natural to ask if we can obtain a better approximation.

Can our results spur new algorithmic applications? We mention two. First, we would like to design a distributed algorithm to compute a fast mixing Markov chain on  $G$ , where  $G$  also represents the topology of the distributed network. Second, is it possible to design a local algorithm, in the spirit of Spielman and Teng [42], that outputs a subset of nodes with small vertex conductance?

We discussed the application of our almost-mixing result—or, more precisely, the weighted-tree construction—to the objective of turning a non-expander into an expander; see the discussion after Theorem B. We would also like to find a more sampling-based application, but have not found a particularly satisfying one yet. In many scenarios, one has a target distribution  $\pi$  to sample from—e.g., the uniform distribution. One *approximately* samples from  $\pi$  by running a Markov chain  $X$  with equilibrium distribution  $\pi$  for its  $\varepsilon$ -mixing time  $t$ . The output  $X_t$  satisfies  $\|X_t - \pi\|_{\text{TV}} \leq \varepsilon$ . If one instead runs a Markov chain  $X'$  with equilibrium distribution  $\pi'$  for its  $\varepsilon$ -mixing time  $t'$ , then the output  $X'_{t'}$  satisfies  $\|X'_{t'} - \pi\|_{\text{TV}} \leq \|X'_{t'} - \pi'\|_{\text{TV}} + \|\pi' - \pi\|_{\text{TV}} \leq 2\varepsilon$  if  $\|\pi' - \pi\|_{\text{TV}} \leq \varepsilon$ . Thus, there is no real disadvantage in adjusting the equilibrium distribution slightly.

Many natural graph structures have small diameter and thus give rise to a fast almost-mixing Markov chain. For example, the graph structure of the Ising model on an  $n$ -graph has diameter at most  $2n$ . Thus, one gets fast—that is, polynomial-in- $n$ —almost mixing for any temperature; this can be used to approximate the partition function. Similar statements hold for sampling independent sets via the hardcore model or counting the number of proper colourings on a graph.

Unfortunately, in all these scenarios, actually calculating the weights on the superimposed tree—which need not be a BFS tree; any tree  $T$  can be used, giving a bound of  $(\text{diam } T)^2$ —is computationally infeasible since it often entails knowing the partition

function already. We hope that a useful application of this result can be found, but we are yet to find one ourselves.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Aldous, D., Fill, J.A.: Reversible markov chains and random walks on graphs. Unfinished Monograph (2002)
2. Allison, M., Shader, B.L.: Fastest mixing Markov chain problem for the union of two cliques. *Linear Multilin. Algebra* **59**(7), 801–823 (2011). <https://doi.org/10.1080/03081087.2010.538690>
3. Alon, N., Boppana, R., Spencer, J.: An asymptotic isoperimetric inequality. *Geom. Funct. Anal.* **8**(3), 411–436 (1998). <https://doi.org/10.1007/s000390050062>
4. Annan, J.D.: A randomised approximation algorithm for counting the number of forests in dense graphs. *Comb. Probab. Comput.* **3**(3), 273–283 (1994). <https://doi.org/10.1017/S0963548300001188>
5. Batson, J., Spielman, D.A., Srivastava, N.: Twice-ramanujan sparsifiers. *SIAM J. Comput.* **41**(6), 1704–1721 (2012)
6. Bayer, D., Diaconis, P.: Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.* **2**(2), 294–313 (1992)
7. Bernstein, A., Gutenberg, M.P., Saranurak, T.: Deterministic decremental SSSP and approximate min-cost flow in almost-linear time. (2021). [arXiv:2101.07149](https://arxiv.org/abs/2101.07149), <https://doi.org/10.48550/arXiv.2101.07149>
8. Bobkov, S., Houdré, C., Tetali, P.:  $\lambda_\infty$  Vertex isoperimetry and concentration. *Combinatorica* **20**(2), 153–172 (2000). <https://doi.org/10.1007/s004930070018>
9. Boyd, S., Diaconis, P., Parrilo, P., Xiao, L.: Symmetry analysis of reversible Markov chains. *Intern. Math.* **2**(1), 31–71 (2005)
10. Boyd, S., Diaconis, P., Parrilo, P., Xiao, L.: Fastest mixing Markov chain on graphs with symmetries. *SIAM J. Optim.* **20**(2), 792–819 (2009). <https://doi.org/10.1137/070689413>
11. Boyd, S., Diaconis, P., Sun, J., Xiao, L.: Fastest mixing Markov chain on a path. *Amer. Math. Monthly* **113**(1), 70–74 (2006). <https://doi.org/10.2307/27641840>
12. Boyd, S., Diaconis, P., Xiao, L.: Fastest mixing Markov chain on a graph. *SIAM Rev.* **46**(4), 667–689 (2004). <https://doi.org/10.1137/S0036144503423264>
13. Chandra, A.K., Raghavan, P., Ruzzo, W.L., Smolensky, R., Tiwari, P.: The electrical resistance of a graph captures its commute and cover times. *Comput. Compl.* **6**(4), 312–340 (1996)
14. Cihan, O., Akar, M.: Fastest mixing reversible Markov chains on graphs with degree proportional stationary distributions. *Institute of Electrical and Electronics Engineers. Trans. Autom. Control.* **60**(1), 227–232 (2015). <https://doi.org/10.1109/TAC.2014.2322942>
15. Diaconis, P., Shahshahani, M.: Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete.* **57**(2), 159–179 (1981). <https://doi.org/10.1007/BF00535487>
16. Dyer, M., Heinrich, M., Jerrum, M., Müller, H.: Polynomial-time approximation algorithms for the antiferromagnetic Ising model on line graphs. (2021). [arXiv:2005.07944](https://arxiv.org/abs/2005.07944)
17. Dyer, M., Jerrum, M., Müller, H., Vušković, K.: Counting weighted independent sets beyond the permanent. *SIAM J. Discret. Math.* **35**(2), 1503–1524 (2021). <https://doi.org/10.1137/20M1347747>
18. Dyer, M., Müller, H.: Counting perfect matchings and the switch chain. *SIAM J. Discret. Math.* **33**(3), 1146–1174 (2019). <https://doi.org/10.1137/18M1172910>
19. Fill, J.A., Kahn, J.: Comparison inequalities and fastest-mixing Markov chains. *Ann. Appl. Probabil.* **23**(5), 1778–1816 (2013). <https://doi.org/10.1214/12-AAP886>
20. Giakkoupis, G.: Tight bounds for rumor spreading with vertex expansion. *Proceedings of the twenty-fifth annual ACM-SIAM symposium on discrete algorithms*, 801–815, (2014)

21. Jafarizadeh, S., Jamalipour, A.: Fastest mixing Markov chain on symmetric K-partite sensor networks. In: 2011 7th International wireless communications and mobile computing Conference, pp. 1883–1888 (2011). <https://doi.org/10.1109/TWCMC.2011.5982616>
22. Jain, V., Pham, H.T., Vuong, T.-D.: Dimension reduction for maximum matchings and the fastest mixing Markov chain. (2022). [arXiv:2203.03858](https://arxiv.org/abs/2203.03858), <https://doi.org/10.48550/arXiv.2203.03858>
23. Jerrum, M., Sinclair, A.: Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. Comput.* **82**(1), 93–133 (1989). [https://doi.org/10.1016/0890-5401\(89\)90067-9](https://doi.org/10.1016/0890-5401(89)90067-9)
24. Jerrum, M., Sinclair, A.: Approximating the permanent. *SIAM J. Comput.* **18**(6), 1149–1178 (1989). <https://doi.org/10.1137/0218077>
25. Jerrum, M., Sinclair, A.: Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.* **22**(5), 1087–1116 (1993). <https://doi.org/10.1137/0222066>
26. Jerrum, M., Sinclair, A.: The Markov chain Monte Carlo method: an approach to approximate counting and integration. *Approximation Algorithms for NP-Hard Problems*, pp. 482–520. PWS Publishing Co. (1996)
27. Jerrum, M., Sinclair, A., Vigoda, E.: A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM* **51**(4), 671–697 (2004). <https://doi.org/10.1145/1008731.1008738>
28. Johnson, W.B., Lindenstrauss, J.: Extensions of Lipschitz mappings into a Hilbert space. *Contemporary mathematics*. 26 (1984)
29. Kwok, T.C., Lau, L.C., Tung, K.C.: Cheeger inequalities for vertex expansion and reweighted eigenvalues. To appear in FOCS (2022). [arXiv:2203.06168](https://arxiv.org/abs/2203.06168)
30. Lawler, G.F., Sokal, A.D.: Bounds on the  $L^2$  spectrum for Markov chains and Markov processes: a generalization of Cheeger’s inequality. *Trans. Amer. Math. Soc.* **309**(2), 557–580 (1988). <https://doi.org/10.2307/2000925>
31. Levin, D.A., Peres, Y., Wilmer, E.L.: *Markov chains and mixing times*. Second ed., American mathematical society, Providence, RI, USA (2017). <https://doi.org/10.1090/mbk/107>
32. Louis, A., Raghavendra, P., Vempala, S.: The complexity of approximating vertex expansion. In: 2013 IEEE 54th annual symposium on foundations of computer science, pp. 360–369 (2013)
33. Miclo, L.: Relations Entre Isopérimétrie et Trou Spectral pour les Chaînes de Markov Finies. *Probab. Theory Relat. Fields* **114**(4), 431–485 (1999). <https://doi.org/10.1007/s004400050231>
34. Montenegro, R.R., Tetali, P.: *Mathematical aspects of mixing times in Markov chains*. Now Publishers Inc. 1 (2006). <https://doi.org/10.5555/1202526>
35. Olesker-Taylor, S., Zanetti, L.: Geometric Bounds on the Fastest Mixing Markov Chain. 13th Innovations in Theoretical Computer Science Conference (ITCS 2022), Leibniz International Proceedings in Informatics (LIPIcs), M. Braverman, ed., Schloss Dagstuhl – Leibniz-Zentrum für Informatik 215 109:1–109:1, (2022). <https://doi.org/10.4230/LIPIcs.ITCS.2022.109>
36. Rabani, Y., Sinclair, A., Wanka, R.: Local divergence of Markov chains and the analysis of iterative load-balancing schemes. In: Proceedings 39th annual symposium on foundations of computer science (Cat. No.98CB36280), 694–703 (1998). <https://doi.org/10.1109/SFCS.1998.743520>
37. Raghavendra, P., Steurer, D.: Graph expansion and the unique games conjecture. In: Proceedings of the forty-second ACM symposium on Theory of computing, pp. 755–764 (2010)
38. Roch, S.: Bounding fastest mixing. *Electron. Commun. Probab.* **10**, 282–296 (2005). <https://doi.org/10.1214/ECP.v10-1169>
39. Sammer, M.D.: Aspects of mass transportation in discrete concentration inequalities. Thesis (2005)
40. Sinclair, A.: Improved bounds for mixing rates of Markov chains and multicommodity flow. *Comb. Probab. Comput.* **1**(4), 351–370 (1992). <https://doi.org/10.1017/S0963548300000390>
41. Spielman, D.: Spectral and algebraic graph theory. Manuscript available online (2019)
42. Spielman, D.A., Teng, S.-H.: A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning. *SIAM J. Comput.* **42**(1), 1–26 (2013)
43. Sun, J., Boyd, S., Xiao, L., Diaconis, P.: The fastest mixing Markov process on a graph and a connection to a maximum variance unfolding problem. *SIAM Rev.* **48**(4), 681–699 (2006). <https://doi.org/10.1137/S0036144504443821>
44. Tang, D., Subramanian, V.G.: Random Walk Based Sampling for Load Balancing in Multi-Server Systems. Abstracts of the 2019 SIGMETRICS/Performance joint international conference on measurement and modeling of computer systems, SIGMETRICS ’19, Association for Computing Machinery, pp 11–12 (2019). <https://doi.org/10.1145/3309697.3331506>

45. Zhang, Y.: Moments of first hitting times for birth–death processes on trees. *Front. Math. China* **14**(4), 833–854 (2019). <https://doi.org/10.1007/s11464-019-0784-6>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.