

A Random Walk on the Rado Graph



Sourav Chatterjee, Persi Diaconis, and Laurent Miclo

Dedicated to our friend and coauthor Harold Widom.

Abstract The Rado graph, also known as the random graph $G(\infty, p)$, is a classical limit object for finite graphs. We study natural ball walks as a way of understanding the geometry of this graph. For the walk started at i , we show that order $\log_2^* i$ steps are sufficient, and for infinitely many i , necessary for convergence to stationarity. The proof involves an application of Hardy's inequality for trees.

Keywords Rado graph · Random graph · Random walk · Mixing speed · Cheeger's inequality · Hardy's inequality

Mathematics Subject Classification (2020) 05C81, 05C80, 05C63, 60J10, 60J27, 37A25, 46E39

1 Introduction

The Rado graph R is a natural limit of the set of all finite graphs (Fraïssé limit, see Sect. 2.1). In Rado's construction, the vertex set is $\mathbb{N} = \{0, 1, 2, \dots\}$. There is an undirected edge from i to j if $i < j$ and the i^{th} binary digit of j is a one (where the 0^{th} digit is the first digit from the right). Thus, 0 is connected to all odd numbers, 1 is connected to 0 and all j which are 2 or 3 (mod 4) and so on. There are many

S. Chatterjee (✉) · P. Diaconis
Department of Mathematics, Stanford University, Stanford, CA, USA

Department of Statistics, Stanford University, Stanford, CA, USA
e-mail: souravc@stanford.edu; diaconis@math.stanford.edu

L. Miclo
Toulouse School of Economics, CNRS, University of Toulouse, Toulouse, France
e-mail: miclo@math.cnrs.fr

alternative constructions. For $p \in (0, 1)$, connecting i and j with probability p gives the Erdős–Rényi graph $G(\infty, p)$, which is (almost surely) isomorphic to R . Further constructions are in Sect. 2.1.

Let $(Q(j))_{0 \leq j < \infty}$ be a positive probability on \mathbb{N} (so, $Q(j) > 0$ for all j , and $\sum_{j=0}^{\infty} Q(j) = 1$). We study a ‘ball walk’ on R generated by Q : from $i \in \mathbb{N}$, pick $j \in N(i)$ with probability proportional to $Q(j)$, where $N(i) = \{j : j \sim i\}$ is the set of neighbors of i in R . Thus, the probability of moving from i to j in one step is

$$K(i, j) = \begin{cases} Q(j)/Q(N(i)) & \text{if } i \sim j, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

As explained below, this walk is connected, aperiodic and reversible, with stationary distribution

$$\pi(i) = \frac{Q(i)Q(N(i))}{Z}, \tag{2}$$

where Z is the normalizing constant.

It is natural to study the mixing time—the rate of convergence to stationarity. The following result shows that convergence is extremely rapid. Starting at $i \in \mathbb{N}$, order $\log_2^* i$ steps suffice, and for infinitely many i , are needed.

Theorem 1 *Let $Q(j) = 2^{-(j+1)}$, $0 \leq j < \infty$. For $K(i, j)$ and π defined at (1) and (2) on the Rado graph R ,*

1. *For universal $A, B > 0$, we have for all $i \in \mathbb{N}$, $\ell \geq 1$,*

$$\|K_i^\ell - \pi\| \leq Ae^{\log_2^* i} e^{-B\ell}.$$

2. *For universal $C > 0$, if $2^{(k)} = 2^{2^{\dots^2}}$ is the tower of 2’s of height k ,*

$$\|K_{2^{(k)}}^\ell - \pi\| \geq C$$

for all $\ell \leq k$. Here $\|K_i^\ell - \pi\| = \frac{1}{2} \sum_{j=0}^{\infty} |K^\ell(i, j) - \pi(j)|$ is the total variation distance and $\log_2^ i$ is the number of times \log_2 needs to be applied, starting from i , to get a result ≤ 1 .*

The proofs allow for some variation in the measure Q . They also work for the $G(\infty, p)$ model of R , though some modification is needed since then K and π are random.

Theorem 1 answers a question in Diaconis and Malliaris [8], who proved the lower bound. Most Markov chains on countable graphs restrict attention to locally finite graphs [25]. For Cayley graphs, Bendikov and Saloff-Coste [1] begin the study of more general transitions and point out how few tools are available. See also [12, 20]. Studying the geometry of a space (here R) by studying the properties of the

Laplacian (here $I - K$) is a classical pursuit (“Can you hear the shape of a drum?”)—see [16].

Section 2 gives background on the Rado graph, Markov chains, ball walks, and Hardy’s inequalities. Section 3 gives preliminaries on the behavior of the neighborhoods of the $G(\infty, p)$ model. The lower bound in Theorem 1 is proved in Sect. 4. Both Sects. 3 and 4 give insight into the geometry of R . The upper bound in Theorem 1 is proved by proving that the Markov chain K has a spectral gap. Usually, a spectral gap alone does not give sharp rates of convergence. Here, for any start i , we show the chain is in a neighborhood of 0 after order $\log_2^* i$ steps. Then the spectral gap shows convergence in a bounded number of further steps. This argument works for both models of R . It is given in Sect. 5.

The spectral gap for the $G(\infty, p)$ model is proved in Sect. 6 using a version of Cheeger’s inequality for trees. For Rado’s binary model, the spectral gap is proved by a novel version of Hardy’s inequality for trees in Sect. 7. This is the first probabilistic application of this technique, which we hope will be useful more generally. There are two appendices containing technical details for the needed versions of Cheeger’s and Hardy’s inequalities.

2 Background on R , Markov Chains, and Hardy’s Inequalities

2.1 The Rado Graph

A definitive survey on the Rado graph (with full proofs) is in Peter Cameron’s fine article [6]. We have also found the Wikipedia entry on the Rado graph and Cameron’s follow-up paper [7] useful.

In Rado’s model, the graph R has vertex set $\mathbb{N} = \{0, 1, 2, \dots\}$ and an undirected edge from i to j if $i < j$ and the i^{th} digit of j is a one. There are many other constructions. The vertex set can be taken as the prime numbers that are $1 \pmod{4}$ with an edge from p to q if the Legendre symbol $(\frac{p}{q}) = 1$. In [8], the graph appears as an induced subgraph of the commuting graph of the group $U(\infty, q)$ —infinite upper-triangular matrices with ones on the diagonal and entries in \mathbb{F}_q . The vertices are points of $U(\infty, q)$. There is an edge from x to y if and only if the commutator $x^{-1}y^{-1}xy$ is zero. The infinite Erdős–Rényi graphs $G(\infty, p)$ are almost surely isomorphic to R for all p , $0 < p < 1$.

The graph R has a host of fascinating properties:

- It is stable in the sense that deleting any finite number of vertices or edges yields an isomorphic graph. So does taking the complement.
- It contains all finite or countable graphs as induced subgraphs. Thus, the (countable) empty graph and complete graphs both appear as induced subgraphs.
- The diameter of R is two—consider any $i \neq j \in \mathbb{N}$ and let k be a binary number with ones in positions i and j and zero elsewhere. Then $i \sim k \sim j$.

- Each vertex is connected to “half” of the other vertices: 0 is connected to all the odd vertices, 1 to 0 and all numbers congruent to 2 or 3 (mod 4), and so on.
- R is highly symmetric: Any automorphism between two induced subgraphs can be extended to all of R (this is called homogeneity). The automorphism group has the cardinality of the continuum.
- R is the “limit” if the collection of all finite graphs (Fraïssé limit). Let us spell this out. A *relational structure* is a set with a finite collection of relations (we are working in first order logic without constants or functions). For example, \mathbb{Q} with $x < y$ is a relational structure. A graph is a set with one symmetric relation. The idea of a “relational sub-structure” clearly makes sense. A class C of structures has the *amalgamation property* if for any $A, B_1, B_2 \in C$ with embeddings $A \xrightarrow{f_1} B_1$ and $A \xrightarrow{f_2} B_2$, there exists $C \in C$ and embeddings $B_1 \xrightarrow{g_1} C$ and $B_2 \xrightarrow{g_2} C$ such that $g_1 f_1 = g_2 f_2$. A countable relational structure M is *homogeneous* if any isomorphism between finite substructures can be extended to an automorphism of M . Graphs and \mathbb{Q} are homogeneous relational structures. A class C has the *joint embedding property* if for any $A, B \in C$ there is a $C \in C$ so that A and B are embeddable in C .

Theorem 2 (Fraïssé) *Let C be a countable class of finite structures with the joint embedding property and closed under ‘induced’ isomorphism with amalgamation. Then there exists a unique countable homogeneous M with C as induced substructures.*

The rationals \mathbb{Q} are the Fraïssé limit of finite ordered sets. The Rado graph R is the Fraïssé limit of finite graphs. We have (several times!) been told “for a model theorist, the Rado graph is just as interesting as the rationals”.

There are many further, fascinating properties of R ; see [6].

2.2 Markov Chains

A *transition matrix* $K(i, j), 0 \leq i, j < \infty, K(i, j) \geq 0, \sum_{j=0}^{\infty} K(i, j) = 1$ for all $i, 0 \leq i < \infty$, generates a Markov chain through its powers

$$K^\ell(i, j) = \sum_{k=0}^{\infty} K(i, k)K^{\ell-1}(k, j).$$

A probability distribution $\pi(i), 0 \leq i < \infty$, is *reversible* for K if

$$\pi(i)K(i, j) = \pi(j)K(j, i) \quad \text{for all } 0 \leq i, j < \infty. \tag{3}$$

Example With definitions (1), (2) on the Rado graph, if $i \sim j$,

$$\pi(i)K(i, j) = \frac{Q(i)Q(N(i))}{Z} \frac{Q(j)}{Q(N(i))} = \frac{Q(i)Q(j)}{Z} = \pi(j)K(j, i).$$

(Both sides are zero if $i \not\sim j$.)

In the above example, we think of $K(i, j)$ as a ‘ball walk’: From i , pick a neighbor j with probability proportional to $Q(j)$ and move to j . We initially found the neat reversible measure surprising. Indeed, we and a generation of others thought that ball walks would have Q as a stationary distribution. Yuval Peres points out that, given a probability $Q(j)$ on the vertices, assigning symmetric weight $Q(i)Q(j)$ to $i \sim j$ gives this K for the weighted local walk. A double ball walk—“from i , choose a neighbor j with probability proportional to $Q(j)$, and from j , choose a neighbor k with probability proportional to $Q(k)/Q(N(k))$ ”—results in a reversible Markov chain with Q as reversing measure. Note that these double ball walks don’t require knowledge of normalizing constants. All of this suggests ball walks as reasonable objects to study.

Reversibility (3) shows that π is a stationary distribution for K :

$$\sum_{i=0}^{\infty} \pi(i)K(i, j) = \sum_{i=0}^{\infty} \pi(j)K(j, i) = \pi(j) \sum_{i=0}^{\infty} K(j, i) = \pi(j).$$

In our setting, since the Rado graph has diameter 2, the walk is connected. It is easy to see that it is aperiodic. Thus, the π in (2) is the unique stationary distribution. Now, the fundamental theorem of Markov chain theory shows, for every starting state i , $K^\ell(i, j) \rightarrow \pi(j)$ as $\ell \rightarrow \infty$, and indeed,

$$\lim_{\ell \rightarrow \infty} \|K_i^\ell - \pi\| = 0.$$

Reversible Markov chains have real spectrum. Say that (K, π) has a *spectral gap* if there is $A > 0$ such that for every $f \in \ell^2(\pi)$,

$$\sum_i (f(i) - \bar{f})^2 \pi(i) \leq A \sum_{i,j} (f(i) - f(j))^2 \pi(i)K(i, j), \tag{4}$$

where $\bar{f} = \sum_{i=0}^{\infty} f(i)\pi(i)$. (Then the gap is at least $1/A$.) For chains with a spectral gap, for any i ,

$$4\|K_i^\ell - \pi\|^2 \leq \frac{1}{\pi(i)} \left(1 - \frac{1}{A}\right)^{2\ell}. \tag{5}$$

Background on Markov chains, particularly rates of convergence, can be found in the readable book of Levin and Peres [19]. For the analytic part of the theory, particularly (4) and (5), and many refinements, we recommend [23].

There has been a healthy development in Markov chain circles around the theme ‘How does a Markov chain on a random graph behave?’. One motivation being, ‘What does a typical convergence rate look like?’. The graphs can be restricted in various natural ways (Cayley graphs, regular graphs of fixed degree or fixed average degree, etc.). A survey of by now classical work is Hildebrand’s survey of ‘random-random walks’ [14]. Recent work by Bordenave and coauthors can be found from [4, 5]. For sparse Erdős–Rényi graphs, there is remarkable work on the walk restricted to the giant component. See [22], [11] and [3].

It is worth contrasting these works with the present efforts. The above results pick a neighbor uniformly at random. In the present paper, the ball walk drives the walk back towards zero. The papers above are all on finite graphs. The Markov chain of Theorem 1 makes perfect sense on finite graphs. The statements and proofs go through (with small changes) to show that order $\log_2^* i$ steps are necessary and sufficient. (For the uniform walk on $G(n, 1/2)$, a bounded number of steps suffice from most initial states, but there are states from which $\log_2^* n$ steps are needed.)

2.3 Hardy’s Inequalities

A key part of the proof of Theorem 1 applies Hardy’s inequalities for trees to prove a Poincaré inequality (Cf. (4)) and hence a bound on the spectral gap. Despite a large expository literature, Hardy’s inequalities remain little known among probabilists. Our application can be read without this expository section but we hope that some readers find it useful. Extensive further references, trying to bridge the gap between probabilists and analysts, is in [17].

Start with a discrete form of Hardy’s original inequality [13, pp. 239–243]. This says that if $a_n \geq 0$, $A_n = a_1 + \dots + a_n$, then

$$\sum_{n=1}^{\infty} \frac{A_n^2}{n^2} \leq 4 \sum_{n=1}^{\infty} a_n^2,$$

and the constant 4 is sharp. Analysts say that “the Hardy operator taking $\{a_n\}$ to $\{A_n/n\}$ is bounded from ℓ^2 to ℓ^2 ”. Later writers showed how to put weights in. If $\mu(n)$ and $\nu(n)$ are positive functions, one aims for

$$\sum_{n=1}^{\infty} A_n^2 \mu(n) \leq A \sum_{n=1}^{\infty} a_n^2 \nu(n),$$

for an explicit A depending on $\mu(n)$ and $\nu(n)$. If $\mu(n) = 1/n^2$ and $\nu(n) = 1$, this gives the original Hardy inequality. To make the transition to a probabilistic application, take $a(n) = g(n) - g(n - 1)$ for g in ℓ^2 . The inequality becomes

$$\sum_{n=1}^{\infty} g(n)^2 \mu(n) \leq A \sum_{n=1}^{\infty} (g(n) - g(n - 1))^2 \nu(n). \tag{6}$$

Consider a ‘birth and death chain’ which transits from j to $j + 1$ with probability $b(j)$ and from j to $j - 1$ with probability $d(j)$. Suppose that this has stationary distribution $\mu(j)$ and that $\sum_j g(j)\mu(j) = 0$. Set $\nu(j) = \mu(j)d(j)$. Then (6) becomes (following simple manipulations)

$$\text{Var}(g) \leq A \sum_{j,k} (g(j) - g(k))^2 \mu(j) K(j, k) \tag{7}$$

with $K(j, k)$ the transition matrix of the birth and death chain. This gives a Poincaré inequality and spectral gap estimate. A crucial ingredient for applying this program is that the constant A must be explicit and manageable. For birth-death chains, this is indeed the case. See [21] or the applications in [9]. The transition from (6) to (7) leans on the one-dimensional setup of birth-death chains. While there is work on Hardy’s inequalities in higher dimensions, it is much more complex; in particular, useful forms of good constants A seem out of reach. In [21], Miclo has shown that for a general Markov transition matrix $K(i, j)$, a spanning tree in the graph underlying K can be found. There is a useful version of Hardy’s inequality for trees due to Evans, Harris and Pick [10]. This is the approach developed in Sect. 7 below which gives further background and details.

Is approximation by trees good enough? There is some hope that the best tree is good enough (see [2]). In the present application, the tree chosen gives the needed result.

2.4 The \log^* Function

Take any $a > 1$. The following is a careful definition of $\log_a^* x$ for $x \geq 0$. First, an easy verification shows that the map $x \mapsto (\log x)/x$ on $(0, \infty)$ is unimodal, with a unique maximum at $x = e$ (where its value is $1/e$), and decaying to $-\infty$ as $x \rightarrow 0$ and to 0 as $x \rightarrow \infty$. Thus, if $a > e^{1/e}$, then for any $x > 0$,

$$\log_a x = \frac{\log x}{\log a} \leq \frac{x}{e \log a} < x.$$

Since \log_a is a continuous map, this shows that if we start with any $x > 0$, iterative applications of \log_a will eventually lead to a point in $(0, a)$ (because there

are no fixed points of \log_a above that, by the above inequality), and then another application of \log_a will yield a negative number. This allows us to define $\log_a^* x$ as the minimum number of applications of \log_a , starting from x , that gets us a nonpositive result.

If $a \leq e^{1/e}$, the situation is a bit more complicated. Here, $\log a \leq 1/e$, which is the maximum value of the unimodal map $x \mapsto (\log x)/x$. This implies that there exist exactly two points $0 < y_a \leq x_a$ that are fixed points of \log_a (with $y_a = x_a$ if $a = e^{1/e}$). Moreover, $\log_a x < x$ if $x \notin [y_a, x_a]$, and $\log_a x \geq x$ if $x \in [y_a, x_a]$. Thus, the previous definition does not work. Instead, we define $\log_a^* x$ to be the minimum number of applications of \log_a , starting from x , that leads us to a result $\leq x_a$. In both cases, defining $\log_a^* 0 = 0$ is consistent with the conventions. Note that $\log_a^* x \geq 0$ for all $x \geq 0$.

3 The Geometry of the Random Model

Throughout this section the graph is $G(\infty, 1/2)$ — an Erdős–Rényi graph on $\mathbb{N} = \{0, 1, 2, \dots\}$ with probability $1/2$ for each possible edge. From here on, we will use the notation \mathbb{N}_+ to denote the set $\{1, 2, \dots\}$ of strictly positive integers. Let $Q(x) = 2^{-(x+1)}$ for $x \in \mathbb{N}$. The transition matrix

$$K(x, y) = \frac{Q(y)}{Q(N(x))} \mathbb{1}_{\{y \in N(x)\}}$$

and its stationary distribution $\pi(x) = Z^{-1} Q(x) Q(N(x))$ are thus random variables. Note that $N(x)$, the neighborhood of x , is random. The main result of this section shows that this graph, with vertices weighted by $Q(x)$, has its geometry controlled by a tree rooted at 0. This tree will appear in both lower and upper bounds on the mixing time for the random model.

To describe things, let $p(x) = \min N(x)$ (p is for ‘parent’, not to be confused with the edge probability p in $G(\infty, p)$). We need some preliminaries about $p(x)$.

Lemma 1 *Let \mathcal{B} be the event that for all $x \in \mathbb{N}_+$, $p(x) < x$. Then we have that $\mathbb{P}(\mathcal{B}) \geq 1/4$.*

Proof Denote $\overline{E} := \{\{x, y\} : x \neq y \in \mathbb{N}\}$, and for any $e \in \overline{E}$, consider $B_e = \mathbb{1}_E(e)$, where E is the set of edges in $G(\infty, 1/2)$, so that $(B_e)_{e \in \overline{E}}$ is a family of independent Bernoulli variables of parameter $1/2$.

For $x \in \mathbb{N}_+$, define A_x the event that x is not linked in \mathcal{G} to a smaller vertex. Namely, we have formally

$$A_x := \bigcap_{y \in \llbracket 0, x-1 \rrbracket} \{B_{\{y,x\}} = 0\},$$

where $\llbracket 0, x - 1 \rrbracket := \{0, 1, \dots, x - 1\}$. Note that the family $(A_x)_{x \in \mathbb{N}_+}$ is independent, and in particular, its events are pairwise independent. We are thus in position to apply Kounias–Hunter–Worsley bounds [15, 18, 26] (see also the survey [24]), to see that for any $n \in \mathbb{N}_+$,

$$\mathbb{P}\left(\bigcup_{x \in \llbracket 1, n \rrbracket} A_x\right) \leq \min\left\{\sum_{x \in \llbracket 1, n \rrbracket} \mathbb{P}(A_x) - \mathbb{P}(A_1) \sum_{y \in \llbracket 2, n \rrbracket} \mathbb{P}(A_y), 1\right\},$$

where we used that $\mathbb{P}(A_1) \geq \mathbb{P}(A_2) \geq \dots \geq \mathbb{P}(A_n)$, which holds because

$$\forall x \in \mathbb{N}_+, \quad \mathbb{P}(A_x) = \prod_{y \in \llbracket 0, x-1 \rrbracket} \mathbb{P}(B_{\{y, x\}}) = \frac{1}{2^x}.$$

We deduce that

$$\mathbb{P}\left(\bigcup_{x \in \llbracket 1, n \rrbracket} A_x\right) \leq \min\left\{\sum_{x \in \llbracket 1, n \rrbracket} \frac{1}{2^x} - \frac{1}{2} \sum_{y \in \llbracket 2, n \rrbracket} \frac{1}{2^y}, 1\right\} = \frac{1}{2} + \frac{1}{4} - \frac{1}{2^{n+1}}.$$

Letting n tends to infinity, we get $\mathbb{P}\left(\bigcup_{x \in \mathbb{N}_+} A_x\right) \leq \frac{3}{4}$. To conclude, note that $\mathcal{B}^c = \bigcup_{x \in \mathbb{N}_+} A_x$. □

Remark 1 Assume that instead of $1/2$, the edges of \bar{E} belong to E with probability $p \in (0, 1)$ (still independently), the corresponding notions receive p in index. The above computations show $\mathbb{P}_p(\mathcal{B}) \geq 1 - (2 - 3p + p^2) \wedge 1$, so that $\mathbb{P}_p(\mathcal{B})$ goes to 1 as p goes to 1, but this bounds provides no information for $p \in (0, (3 - \sqrt{5})/2)$.

In fact the above observation shows that the Kounias–Hunter–Worsley bound is not optimal, at least for small $p > 0$. So let us give another computation of $\mathbb{P}_p(\mathcal{B})$:

Lemma 2 Consider the situation described in Remark 1, with $p \in (0, 1)$. We have

$$\mathbb{P}_p(\mathcal{B}) = \left(\sum_{n \in \mathbb{N}} p(n)(1 - p)^n\right)^{-1}$$

where $p(n)$ is the number of partitions of n . In particular $\mathbb{P}(\mathcal{B}) > 0$ for all $p \in (0, 1)$.

Proof Indeed, we have $\mathcal{B} = \bigcap_{x \in \mathbb{N}_+} A_x^c$, so that by independence of the A_x , for $x \in \mathbb{N}_+$,

$$\mathbb{P}_p(\mathcal{B}) = \prod_{x \in \mathbb{N}_+} \mathbb{P}(A_x^c) = \left(\prod_{x \in \mathbb{N}_+} \frac{1}{1 - (1 - p)^x}\right)^{-1} = \left(\prod_{x \in \mathbb{N}_+} \sum_{n \in \mathbb{N}} (1 - p)^{xn}\right)^{-1}$$

Let \mathcal{N} be the set of sequences of integers $(n_l)_{l \in \mathbb{N}_+}$ with all but finitely many elements equal to zero. Applying the distributive law to the above expression, we have

$$\mathbb{P}_p(\mathcal{B}) = \left(\sum_{(n_l)_{l \in \mathbb{N}_+} \in \mathcal{N}} \prod_{x \in \mathbb{N}_+} (1 - p)^{xn_x} \right)^{-1} = \left(\sum_{n \in \mathbb{N}} p(n)(1 - p)^n \right)^{-1}$$

where $p(n)$ is the number of ways to write n as $\sum_{x \in \mathbb{N}_+} xn_x$, with $(n_l)_{l \in \mathbb{N}_+} \in \mathcal{N}$. □

Consider the set of edges

$$F := \{\{x, p(x)\} : x \in \mathbb{N}_+\}$$

and the corresponding graph $\mathcal{T} := (\mathbb{N}, F)$. Under \mathcal{B} , it is clear that \mathcal{T} is a tree. But this is always true:

Lemma 3 *The graph \mathcal{T} is a tree.*

Proof The argument is by contradiction. Assume that \mathcal{T} contains a cycle, say $(x_l)_{l \in \mathbb{Z}_n}$ with $n \geq 3$. Let us direct the a priori unoriented edges $\{x_l, x_{l+1}\}$, for $l \in \mathbb{Z}_n$, by putting an arrow from x_l to x_{l+1} (respectively from x_{l+1} to x_l) if $p(x_l) = x_{l+1}$ (resp. $p(x_{l+1}) = x_l$). Note that we either have

$$\forall l \in \mathbb{Z}_n, x_l \rightarrow x_{l+1}, \text{ or } \forall l \in \mathbb{Z}_n, x_{l+1} \rightarrow x_l, \tag{8}$$

because otherwise there would exist $l \in \mathbb{Z}_n$ with two arrows exiting from x_l , a contradiction. Up to reindexing $(x_l)_{l \in \mathbb{Z}_n}$ as $(x_{-l})_{l \in \mathbb{Z}_n}$, we can assume that (8) holds.

Fix some $l \in \mathbb{Z}_n$. Since $p(x_l) = x_{l+1}$, we have $x_l \in N(x_{l+1})$, so $x_{l+2} = p(x_{l+1}) \leq x_l$. Due to the fact that $x_l \neq x_{l+2}$ (recall that $n \geq 3$), we get $x_{l+2} < x_l$. Starting from x_0 and iterating this relation (in a minimal way, $n/2$ times if n is even, or n times if n is odd), we obtain a contradiction: $x_0 < x_0$. Thus, \mathcal{T} must be a tree. □

Let us come back to the case where $p = 1/2$. The following result gives an idea of how far $p(x)$ is from x , for $x \in \mathbb{N}_+$.

Lemma 4 *Almost surely, there exist only finitely many $x \in \mathbb{N}_+$ such that $p(x) > 2 \log_2(1 + x)$. In particular, a.s. there exists a (random) finite $C \geq 2$ such that*

$$\forall x \in \mathbb{N}_+, \quad p(x) \leq C \log_2(1 + x).$$

Proof The first assertion follows from the Borel–Cantelli lemma, as follows. For any $x \in \mathbb{N}_+$, consider the event

$$A_x := \{p(x) > 2 \log_2(1 + x)\}.$$

Denoting $\lfloor \cdot \rfloor$ the the integer part, we compute

$$\begin{aligned} \sum_{x \in \mathbb{N}_+} \mathbb{P}(A_x) &= \sum_{x \in \mathbb{N}_+} \mathbb{P}(B_{\{0,x\}} = 0, B_{\{1,x\}} = 0, \dots, B_{\{\lfloor 2 \log_2(1+x) \rfloor, x\}} = 0) \\ &= \sum_{x \in \mathbb{N}_+} \frac{1}{2^{\lfloor 2 \log_2(1+x) \rfloor}} \leq \sum_{x \in \mathbb{N}_+} \frac{1}{(1+x)^2} < +\infty. \end{aligned}$$

Having shown that a.s. there exists only a finite number of integers $x \in \mathbb{N}_+$ satisfying $p(x) > 2 \log_2(1+x)$, denote these points as x_1, \dots, x_N , with $N \in \mathbb{N}$. To get the second assertion, it is sufficient to take $C := \max\{p(x_l)/\log(1+x_l) : l \in \llbracket 1, N \rrbracket\}$, with the convention that $C := 2$ if $N = 0$. □

4 The Lower Bound

The lower bound in Theorem 1, showing that order $\log_2^* i$ steps are necessary for infinitely many i is proved in [8] for the binary model of the Rado graph and we refer there for the proof. A different argument is needed for the $G(\infty, 1/2)$ model. This section gives the details (see Theorem 3 below).

Let μ be the stationary distribution of our random walk on $G(\infty, 1/2)$ (with $Q(j) = 2^{-(j+1)}$, as in Theorem 1), given a realization of the graph. Note that μ is random. For each $x \in \mathbb{N}$, let τ_x be the mixing time of the walk starting from x , that is, the smallest n such that the law of the walk at time n , starting from x , has total variation distance $\leq 1/4$ from μ . Note that the τ_x 's are also random.

Theorem 3 *Let τ_x be as above. Then with probability one,*

$$\limsup_{x \rightarrow \infty} \frac{\tau_x}{\log_{16}^* x} \geq 1.$$

Equivalently, with probability one, given any $\varepsilon > 0$, $\tau_x \geq (1 - \varepsilon) \log_{16}^ x$ for infinitely many x .*

We need the following lemma.

Lemma 5 *With probability one, there is an infinite sequence $x_0 < x_1 < x_2 < \dots \in \mathbb{N}$ such that:*

1. *For each i , x_{i+1} is connected to x_i by an edge, but not connected by an edge to any other number in $\{0, 1, \dots, 2x_i - 1\}$.*
2. *For each i , $2^{3x_i} \leq x_{i+1} \leq 2^{3x_i+1} - 1$.*

Proof Define a sequence y_0, y_1, y_2, \dots inductively as follows. Let y_0 be an arbitrary element of \mathbb{N} . For each i , let y_{i+1} be the smallest element in $\{2^{3y_i}, 2^{3y_i} + 1, \dots, 2^{3y_i+1} - 1\}$ that has an edge to y_i , but to no other number in $\{0, 1, \dots, 2y_i - 1\}$. If there exists no such number, then the process stops. Let A_i be the event that y_i exists. Note that $A_0 \supseteq A_1 \supseteq A_2 \supseteq \dots$.

Let $F(x) := 2^{3x}$, $G(x) := 2^{3x+1} - 1$, $a_0 = b_0 = y_0$, and for each $i \geq 1$, let

$$a_i := \underbrace{F \circ F \circ \dots \circ F}_{i \text{ times}}(y_0), \quad b_i := \underbrace{G \circ G \circ \dots \circ G}_{i \text{ times}}(y_0).$$

Since $2^{3y_i} \leq y_{i+1} \leq 2^{3y_i+1} - 1$ for each i , it follows by induction that $a_i \leq y_i \leq b_i$ for each i (if y_i exists). Now fix some $i \geq 1$. Since the event A_{i-1} is determined by y_1, \dots, y_{i-1} , and these random variables can take only finitely many values (by the above paragraph), we can write A_{i-1} as a finite union of events of the form $\{y_1 = c_1, \dots, y_{i-1} = c_{i-1}\}$, where $c_1 < c_2 < \dots < c_{i-1} \in \mathbb{N}$.

Now note that for any $c_1 < \dots < c_{i-1}$, the event $A_i \cap \{y_1 = c_1, \dots, y_{i-1} = c_{i-1}\}$ happens if and only if $\{y_1 = c_1, \dots, y_{i-1} = c_{i-1}\}$ happens and there is some $y \in \{2^{3c_{i-1}}, 2^{3c_{i-1}} + 1, \dots, 2^{3c_{i-1}+1} - 1\}$ that has an edge to c_{i-1} , but to no other number in $\{0, \dots, 2c_{i-1} - 1\}$. The event $\{y_1 = c_1, \dots, y_{i-1} = c_{i-1}\}$ is in $\mathcal{F}_{c_{i-1}}$, where \mathcal{F}_x denotes the σ -algebra generated by the edges between all numbers in $\{0, \dots, x\}$. On the other hand, on the event $\{y_1 = c_1, \dots, y_{i-1} = c_{i-1}\}$, it is not hard to see that

$$\mathbb{P}(A_i | \mathcal{F}_{c_{i-1}}) = 1 - (1 - 2^{-2c_{i-1}})^{2^{3c_{i-1}}}.$$

Thus,

$$\begin{aligned} &\mathbb{P}(A_i \cap \{y_1 = c_1, \dots, y_{i-1} = c_{i-1}\}) \\ &= \mathbb{P}(y_1 = c_1, \dots, y_{i-1} = c_{i-1})(1 - (1 - 2^{-2c_{i-1}})^{2^{3c_{i-1}}}) \\ &\geq \mathbb{P}(y_1 = c_1, \dots, y_{i-1} = c_{i-1})(1 - e^{-2^{c_{i-1}}}), \end{aligned}$$

where in the last step we used the inequality $0 \leq 1 - x \leq e^{-x}$ (which holds for all $x \in [0, 1]$). Note that the term inside the parentheses on the right side is an increasing function of c_{i-1} , and the maximum possible value of y_{i-1} is b_{i-1} . Thus, summing both sides over all values of c_1, \dots, c_{i-1} such that $\{y_1 = c_1, \dots, y_{i-1} = c_{i-1}\} \subseteq A_{i-1}$, we get $\mathbb{P}(A_i) = \mathbb{P}(A_i \cap A_{i-1}) \geq \mathbb{P}(A_{i-1})(1 - e^{-2^{b_{i-1}}})$. Proceeding inductively, this gives

$$\mathbb{P}(A_1 \cap \dots \cap A_i) \geq \prod_{k=0}^{i-1} (1 - e^{-2^{b_k}}).$$

Taking $i \rightarrow \infty$, we get $\mathbb{P}(B) \geq \prod_{k=0}^{\infty} (1 - e^{-2^{b_k}})$, where $B := \bigcap_{k=1}^{\infty} A_k$. Now recall that the event B , as well as the numbers b_0, b_1, \dots , are dependent on our choice of y_0 . To emphasize this dependence, let us write them as $B(y_0)$ and $b_k(y_0)$. Then by the above inequality,

$$\sum_{y_0 \in \mathbb{N}} \mathbb{P}(B(y_0)^c) \leq \sum_{y_0 \in \mathbb{N}} \left(1 - \prod_{k=0}^{\infty} (1 - e^{-2^{b_k(y_0)}}) \right),$$

where $B(y_0)^c$ denotes the complement of $B(y_0)$. Due to the extremely rapid growth of $b_k(y_0)$ as $k \rightarrow \infty$, and the fact that $b_0(y_0) = y_0$, it is not hard to see that the right side is finite. Therefore, by the Borel–Cantelli lemma, $B(y_0)^c$ happens for only finitely many y_0 with probability one. In particular, with probability one, $B(y_0)$ happens for some y_0 . This completes the proof. \square

Proof (Of Theorem 3) Fix a realization of $G(\infty, 1/2)$. Let x be so large that $\mu([x, \infty)) < 1/10$, and $\prod_{k=1}^{\infty} (1 - 2^{-a_k(x)+1}) \geq 9/10$.

Let x_0, x_1, x_2, \dots be a sequence having the properties listed in Lemma 5 (which exists with probability one, by the lemma). Discarding some initial values if necessary, let us assume that $x_0 > x$. By the listed properties, it is obvious that $x_i \rightarrow \infty$ as $i \rightarrow \infty$. Thus, to prove Theorem 3, it suffices to prove that

$$\liminf_{i \rightarrow \infty} \frac{\tau_{x_i}}{\log_{16}^* x_i} \geq 1. \tag{9}$$

We will now deduce this from the properties of the sequence.

Suppose that our random walk starts from x_i for some $i \geq 1$. Since x_i connects to x_{i-1} by an edge, but not to any other number in $\{0, \dots, 2x_{i-1} - 1\}$, we see that the probability of the walk landing up at x_{i-1} in the next step is at least

$$1 - \frac{1}{2^{-x_i}} \sum_{k=2x_i}^{\infty} 2^{-k} = 1 - 2^{-x_i+1}.$$

Proceeding by induction, this shows that the chance that the walk lands up at x_0 at step i is at least $\prod_{k=1}^i (1 - 2^{-x_k+1})$. Let μ_i be the law of walk at step i (starting from x_i , and conditional on the fixed realization of our random graph). Then by the above deduction and the facts that $x_0 > x$ and $x_k \geq a_k(x_0) \geq a_k(x)$, we have

$$\mu_i([x, \infty)) \geq \prod_{k=1}^i (1 - 2^{-x_k+1}) \geq \prod_{k=1}^i (1 - 2^{-a_k(x)+1}).$$

By our choice of x , the last expression is bounded below by $9/10$. But $\mu([x, \infty)) < 1/10$. Thus, the total variation distance between μ_i and μ is at least $8/10$. In particular, $\tau_{x_i} > i$. Now, $x_i \leq 2^{3x_{i-1}+1} - 1 \leq 16^{x_{i-1}}$, which shows that

$\log_{16}^* x_i \leq \log_{16}^* x_{i-1} + 1$. Proceeding inductively, we get $\log_{16}^* x_i \leq i + \log_{16}^* x_0$. Thus, $\tau_{x_i} > \log_{16}^* x_i - \log_{16}^* x_0$. This proves (9). \square

5 The Upper Bound (Assuming a Spectral Gap)

This section gives the upper bound for both the binary and random model of the Rado graph. Indeed, the proof works for a somewhat general class of graphs and more general base measures Q . The argument assumes that we have a spectral gap estimate. These are proved below in Sects. 6 and 7. We give this part of the argument first because, as with earlier sections, it gives a useful picture of the random graph.

Take any undirected graph on the nonnegative integers, with the property:

$$\left\{ \begin{array}{l} \text{There exists } C > 0 \text{ such that for any } j \geq 2, \\ j \text{ is connected to some } k \leq C \log j. \end{array} \right. \tag{10}$$

Let $\{X_n\}_{n \geq 0}$ be the Markov chain on this graph, which, starting at state i , jumps to a neighbor j with probability proportional to $Q(j) = 2^{-(j+1)}$. The following is the main result of this section.

Theorem 4 *Let K be the transition kernel of the Markov chain defined above. Suppose that K has a spectral gap. Let μ be the stationary distribution of the chain, and let $a := e^{1/C}$. Then for any $i \in \mathbb{N}$ and any $\ell \geq 1$,*

$$\|K_i^\ell - \mu\| \leq C_1 e^{\log_a^* i} e^{-C_2 \ell},$$

where C_1 and C_2 are positive constants that depend only the properties of the chain (and not on i or ℓ).

By Lemma 4, $G(\infty, 1/2)$ satisfies the property (10) with probability one, for some C that may depend on the realization of the graph. The Rado graph also satisfies property (10), with $K = 1/\log 2$. Thus, the random walk starting from j mixes in time $\log_2^* j$ on the Rado graph, provided that it has a spectral gap. For $G(\infty, 1/2)$, assuming that the walk has a spectral gap, the mixing time starting from j is $\log_a^* j$, where a depends on the realization of the graph. The spectral gap for $G(\infty, 1/2)$ will be proved in Sect. 6, and the spectral gap for the Rado graph will be established in Sect. 7. Therefore, this proves Theorem 1 and also the analogous result for $G(\infty, 1/2)$.

Proof (Of Theorem 4) Note that $a > 1$. Let $Z_n := \log_a^* X_n$. We claim that there is some j_0 sufficiently large, and some positive constant c , such that

$$\mathbb{E}(e^{Z_{n+1}} | \mathcal{F}_n) \leq e^{Z_n - c} \quad \text{if } Z_n > j_0, \tag{11}$$

where \mathcal{F}_n is the σ -algebra generated by X_0, \dots, X_n . (The proof is given below.) This implies that if we define the stopping time $S := \min\{n \geq 0 : X_n \leq j_0\}$, then $\{e^{Z_{S \wedge n} + c(S \wedge n)}\}_{n \geq 0}$ is a supermartingale with respect to the filtration $\{\mathcal{F}_n\}_{n \geq 0}$ (see details below). Moreover, it is nonnegative. Thus, if we start from the deterministic initial condition $X_0 = j$, then for any n ,

$$\mathbb{E}(e^{Z_{S \wedge n} + c(S \wedge n)} | X_0 = j) \leq e^{Z_{S \wedge 0} + c(S \wedge 0)} = e^{\log_a^* j}.$$

But $Z_{S \wedge n} \geq 0$. Thus, $\mathbb{E}(e^{c(S \wedge n)} | X_0 = j) \leq e^{\log_a^* j}$. Taking $n \rightarrow \infty$ and applying the monotone convergence theorem, we get

$$\mathbb{E}(e^{cS} | X_0 = j) \leq e^{\log_a^* j}. \tag{12}$$

Now take any $j \geq 1$ and $n \geq 1$. Let μ be the stationary distribution, and let $\mu_{j,n}$ be the law of X_n when $X_0 = j$. Take any $A \subseteq \{0, 1, \dots\}$. Then for any $m \leq n$,

$$\begin{aligned} \mu_{j,n}(A) &= \mathbb{P}(X_n \in A | X_0 = j) \\ &= \sum_{i=0}^m \sum_{l=0}^{j_0} \mathbb{P}(X_n \in A | S = i, X_i = l, X_0 = j) \mathbb{P}(S = i, X_i = l | X_0 = j) \\ &\quad + \mathbb{P}(X_n \in A | S > m, X_0 = j) \mathbb{P}(S > m | X_0 = j). \end{aligned}$$

But

$$\mathbb{P}(X_n \in A | S = i, X_i = l, X_0 = j) = \mathbb{P}(X_n \in A | X_i = l) = \mu_{l,n-i}(A),$$

and

$$\mu(A) = \sum_{i=0}^m \sum_{l=0}^{j_0} \mu(A) \mathbb{P}(S = i, X_i = l | X_0 = j) + \mu(A) \mathbb{P}(S > m | X_0 = j).$$

Thus, $|\mu_{j,n}(A) - \mu(A)|$ can be bounded above by

$$\sum_{i=0}^m \sum_{l=0}^{j_0} |\mu_{l,n-i}(A) - \mu(A)| \mathbb{P}(S = i, X_i = l | X_0 = j) + \mathbb{P}(S > m | X_0 = j).$$

Now, if our Markov chain has a spectral gap, there exist constants C_1 and C_2 depending only on j_0 and the spectral gap, such that

$$|\mu_{l,n-i}(A) - \mu(A)| \leq C_1 e^{-C_2(n-i)} \leq C_1 e^{-C_2(n-m)}$$

for all $0 \leq i \leq m$ and $0 \leq l \leq j_0$. Using this bound and the bound (12) on $\mathbb{E}(e^{cS}|X_0 = j)$ obtained above, we get

$$|\mu_{j,n}(A) - \mu(A)| \leq C_1 e^{-C_2(n-m)} + e^{\log_a^* j - cm}.$$

Taking $m = \lceil n/2 \rceil$, we get the desired result. □

Proof (Of inequality (11)) It suffices to take $n = 0$. Suppose that $X_0 = j$ for some $j \geq 1$. By assumption, there is a neighbor k of j such that $k \leq K \log j = \log_a j$. Assuming that j is sufficiently large (depending on K), we have that for any $l \leq k$,

$$\log_a^* l \leq \log_a^* k \leq \log_a^*(\log_a j) = \log_a^* j - 1.$$

Also, $\log_a^* l \leq \log_a^* j$ for any $l \leq j$. Thus,

$$\begin{aligned} \mathbb{E}(e^{Z_1 - Z_0} | X_0 = j) &\leq e^{-1} \mathbb{P}(X_1 \leq k | X_0 = j) + \mathbb{P}(k < X_1 \leq j | X_0 = j) \\ &\quad + \sum_{l > j} e^{\log_a^* l - \log_a^* j} \mathbb{P}(X_1 = l | X_0 = j). \end{aligned}$$

Now for any $l \geq k$,

$$\mathbb{P}(X_1 = l | X_0 = j) \leq \frac{\mathbb{P}(X_1 = l | X_0 = j)}{\mathbb{P}(X_1 = k | X_0 = j)} = \frac{Q(l)}{Q(k)} = 2^{-(l-k)}.$$

Thus,

$$\sum_{l > j} e^{\log_a^* l - \log_a^* j} \mathbb{P}(X_1 = l | X_0 = j) \leq \sum_{l > j} e^{\log_a^* l - \log_a^* j} 2^{-(l-k)},$$

which is less than $1/4$ if j is sufficiently large (since $k \leq \log_a^* j$). Next, let L be the set of all $l > k$ that are connected to j . Then

$$\mathbb{P}(X_1 > k | X_0 = j) \leq \frac{\mathbb{P}(X_1 > k | X_0 = j)}{\mathbb{P}(X_1 \geq k | X_0 = j)} = \frac{\sum_{l \in L} 2^{-l}}{2^{-k} + \sum_{l \in L} 2^{-l}}.$$

Since the map $x \mapsto x/(2^{-k} + x)$ is increasing, this shows that

$$\mathbb{P}(X_1 > k | X_0 = j) \leq \frac{\sum_{l > k} 2^{-l}}{2^{-k} + \sum_{l > k} 2^{-l}} = \frac{1}{2}.$$

Combining, we get that for sufficiently large j ,

$$\begin{aligned} \mathbb{E}(e^{Z_1 - Z_0} | X_0 = j) &\leq e^{-1} \mathbb{P}(X_1 \leq k | X_0 = j) + \mathbb{P}(X_1 > k | X_0 = j) + \frac{1}{4} \\ &= e^{-1} + (1 - e^{-1}) \mathbb{P}(X_1 > k | X_0 = j) + \frac{1}{4} \\ &\leq e^{-1} + \frac{1 - e^{-1}}{2} + \frac{1}{4} = \frac{3 + 2e^{-1}}{4} < 1. \end{aligned}$$

□

Proof (Of the Supermartingale Property) Note that

$$\begin{aligned} &\mathbb{E}(e^{Z_{S \wedge (n+1)} + c(S \wedge (n+1))} | \mathcal{F}_n) \\ &= \sum_{i=0}^n \mathbb{E}(e^{Z_{S \wedge (n+1)} + c(S \wedge (n+1))} 1_{\{S=i\}} | \mathcal{F}_n) + \mathbb{E}(e^{Z_{S \wedge (n+1)} + c(S \wedge (n+1))} 1_{\{S>n\}} | \mathcal{F}_n) \\ &= \sum_{i=0}^n \mathbb{E}(e^{Z_i + ci} 1_{\{S=i\}} | \mathcal{F}_n) + \mathbb{E}(e^{Z_{n+1} + c(n+1)} 1_{\{S>n\}} | \mathcal{F}_n). \end{aligned}$$

The events $\{S = i\}$ are \mathcal{F}_n -measurable for all $0 \leq i \leq n$, and so is the event $\{S > n\}$. Moreover, Z_0, \dots, Z_n are also \mathcal{F}_n -measurable. Thus, the above expression shows that

$$\mathbb{E}(e^{Z_{S \wedge (n+1)} + c(S \wedge (n+1))} | \mathcal{F}_n) = 1_{\{S \leq n\}} e^{Z_{S \wedge n} + c(S \wedge n)} + 1_{\{S > n\}} \mathbb{E}(e^{Z_{n+1} + c(n+1)} | \mathcal{F}_n).$$

But if $S > n$, then $Z_n > j_0$, and therefore by (11),

$$\mathbb{E}(e^{Z_{n+1} + c(n+1)} | \mathcal{F}_n) \leq e^{Z_n - c + c(n+1)} = e^{Z_n + cn}.$$

Thus,

$$\begin{aligned} \mathbb{E}(e^{Z_{S \wedge (n+1)} + c(S \wedge (n+1))} | \mathcal{F}_n) &\leq 1_{\{S \leq n\}} e^{Z_{S \wedge n} + c(S \wedge n)} + 1_{\{S > n\}} e^{Z_n + cn} \\ &= e^{Z_{S \wedge n} + c(S \wedge n)}. \end{aligned}$$

□

6 Spectral Gap for the Random Model

Our next goal is to show that the random reversible couple (K, π) admits a spectral gap. The arguments make use of the ideas and notation of Sect. 3. In particular, recall the event $\mathcal{B} = \{p(x) < x \ \forall x \in \mathbb{N}_+\}$ from Lemma 1 and the random tree \mathcal{T} with

edge set F from Lemma 3. The argument uses a version of Cheeger’s inequality for trees which is further developed in Appendix 1.

Proposition 1 *On \mathcal{B} , there exists a random constant $\Lambda > 0$ such that*

$$\forall f \in L^2(\pi), \quad \Lambda \pi[(f - \pi[f])^2] \leq \mathcal{E}(f)$$

where in the r.h.s. \mathcal{E} is the Dirichlet form defined by

$$\forall f \in L^2(\pi), \quad \mathcal{E}(f) := \frac{1}{2} \sum_{x,y \in \mathbb{N}} (f(y) - f(x))^2 \pi(x) K(x, y).$$

Taking into account that for any $f \in L^2(\pi)$, the variance $\pi[(f - \pi[f])^2]$ of f with respect to π is bounded above by $\pi[(f - f(0))^2]$, the previous result is an immediate consequence of the following existence of positive first Dirichlet eigenvalue under \mathcal{B} .

Proposition 2 *On \mathcal{B} , there exists a random constant $\Lambda > 0$ such that*

$$\forall f \in L^2(\pi), \quad \Lambda \pi[(f - f(0))^2] \leq \mathcal{E}(f). \tag{13}$$

The proof of Proposition 2 is based on the pruning of \mathcal{G} into \mathcal{T} and then resorting to Cheeger’s inequalities for trees. More precisely, let us introduce the following notations. Define the Markov kernel $K_{\mathcal{T}}$ as

$$\forall x, y \in \mathbb{N}, \quad K_{\mathcal{T}}(x, y) := \begin{cases} K(x, y) & \text{if } \{x, y\} \in F, \\ 1 - \sum_{z \in \mathbb{N} \setminus \{x\}} K_{\mathcal{T}}(x, z) & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases}$$

Note that this kernel is reversible with respect to π . The corresponding Dirichlet form is given, for any $f \in L^2(\pi)$, by

$$\mathcal{E}_{\mathcal{T}}(f) := \sum_{x,y \in \mathbb{N}} (f(y) - f(x))^2 \frac{\pi(x) K_{\mathcal{T}}(x, y)}{2} = \sum_{\{x,y\} \in F} (f(y) - f(x))^2 \pi(x) K(x, y)$$

It will be convenient to work with $\tilde{\mathcal{E}} := Z \mathcal{E}_{\mathcal{T}}$, where Z is the normalizing constant of π , as in equation (2). Define a nonnegative measure μ on \mathbb{N}_+ as

$$\forall x \in \mathbb{N}_+, \quad \mu(x) := Q(x) Q(p(x)). \tag{14}$$

Proposition 3 *On \mathcal{B} , there exists $\lambda > 0$ such that*

$$\forall f \in L^2(\mu), \quad \lambda \mu[(f - f(0))^2] \leq \tilde{\mathcal{E}}(f). \tag{15}$$

This result immediately implies Proposition 2. Indeed, due on one hand to the inclusion $N(x) \subset \llbracket p(x), \infty \llbracket$ and on the other hand to the nature of Q , we have

$$\forall x \in \mathbb{N}_+, \quad Q(p(x)) \leq Q(N(x)) \leq 2Q(p(x)). \tag{16}$$

Thus for any $f \in L^2(\mu)$,

$$\begin{aligned} \lambda\pi[(f - f(0))^2] &= \frac{\lambda}{Z} \sum_{x \in \mathbb{N}_+} (f(x) - f(0))^2 Q(x)Q(N(x)) \\ &\leq \frac{2\lambda}{Z} \sum_{x \in \mathbb{N}_+} (f(x) - f(0))^2 Q(x)Q(p(x)) \\ &= \frac{2\lambda}{Z} \mu[(f - f(0))^2] \leq \frac{2}{Z} \tilde{\mathcal{E}}(f) = 2\mathcal{E}_{\mathcal{T}}(f) \leq 2\mathcal{E}(f), \end{aligned}$$

and thus, Proposition 2 holds with $\Lambda := \lambda/2$.

The proof of Proposition 3 is based on a Dirichlet-variant of the Cheeger inequality (which is in fact slightly simpler than the classical one, see Appendix 1). For any $A \subset \mathbb{N}_+$, define $\partial A := \{\{x, y\} : x \in A, y \notin A\} \subset \bar{E}$. Endow \bar{E} with the measure ν induced by, for any $\{x, y\} \in \bar{E}$,

$$\nu(\{x, y\}) := Z\pi(x)K_{\mathcal{T}}(x, y) = \begin{cases} Q(x)Q(y) & \text{if } \{x, y\} \in F, \\ 0 & \text{otherwise.} \end{cases}$$

Define the Dirichlet–Cheeger constant

$$\iota := \inf_{A \in \mathcal{A}} \frac{\nu(\partial A)}{\mu(A)} \geq 0$$

where $\mathcal{A} := \{A \subset \mathbb{N}_+ : A \neq \emptyset\}$. The proof of the traditional Markovian Cheeger’s inequality given in the lectures by Saloff-Coste [23] implies directly that the best constant λ in Proposition 3 satisfies $\lambda \geq \iota^2/2$. Thus it remains to check:

Proposition 4 *On \mathcal{B} , we have $\iota \geq 1/2$ and in particular $\iota > 0$.*

Proof Take any nonempty $A \in \mathcal{A}$ and decompose it into its connected components with respect to \mathcal{T} : $A = \bigsqcup_{i \in \mathcal{I}} A_i$, where the index set \mathcal{I} is at most denumerable. Note that

$$\mu(A) = \sum_{i \in \mathcal{I}} \mu(A_i), \quad \nu(A) = \sum_{i \in \mathcal{I}} \nu(A_i),$$

where the second identity holds because there are no edges in F connecting two different A_i ’s. Thus, it follows that $\iota = \inf_{A \in \tilde{\mathcal{A}}} \nu(\partial A)/\mu(A)$, where $\tilde{\mathcal{A}}$ is the set of subsets of \mathcal{A} which are \mathcal{T} -connected.

Consider $A \in \tilde{\mathcal{A}}$, it has a smallest element $a \in \mathbb{N}_+$ (since $0 \notin A$). Let T_a be the subtree of descendants of A in \mathcal{T} (i.e., the set of vertices from \mathbb{N}_+ whose non-self-intersecting path to 0 passes through a). We have $A \subset T_a$, and $\partial A \supset \{a, p(a)\} = \partial T_a$, and it follows that $\nu(\partial A)/\mu(A) \geq \nu(\partial T_a)/\mu(T_a)$. We deduce that

$$\iota \geq \inf_{a \in \mathbb{N}_+} \frac{\nu(\partial T_a)}{\mu(T_a)} = \inf_{a \in \mathbb{N}_+} \frac{Q(a)Q(p(a))}{\mu(T_a)}.$$

On \mathcal{B} , we have for any $a \in \mathbb{N}_+$, on the one hand

$$\forall x \in T_a, \quad p(x) \geq p(a), \tag{17}$$

and on the other hand

$$T_a \subset \llbracket a, \infty \llbracket. \tag{18}$$

We get $\mu(T_a)$ equals

$$\sum_{x \in T_a} Q(x)Q(p(x)) \geq Q(p(a)) \sum_{x \in T_a} Q(x) \geq Q(p(a)) \sum_{x \in \llbracket a, \infty \llbracket} Q(x) = 2Q(p(a))Q(a).$$

It follows that $\iota \geq 1/2$. □

Lemma 4 can now be used to see that the ball Markov chain on the random graph has a.s. a spectral gap. Indeed, we deduce from Lemma 4 that there exists a (random) vertex $x_0 \in \mathbb{N}$ such that for any $x > x_0$, $p(x) < x$. Consider

$$x_1 := \max\{p(x) : x \in \llbracket 1, x_0 \llbracket\}.$$

It follows that for any $a > x_1$, we have, for all $\forall x \in T_a$, $p(x) < x$. (To see this, take any path a_0, a_1, \dots in T_a , starting at $a_0 = a$, so that $p(a_i) = a_{i-1}$ for each i . Let k be the first index such that $a_k \geq a_{k+1}$, assuming that there exists such a k . Then $a_{k+1} \leq x_0$, and so $a_k = p(a_{k+1}) \leq x_1$. But this is impossible, since $a_0 \leq a_k$ and $a_0 > x_1$.) In particular, we see that (17) and (18) hold for $a > x_1$. As a consequence,

$$\inf_{a > x_1} \frac{\nu(\partial T_a)}{\mu(T_a)} \geq \frac{1}{2}.$$

By the finiteness of $\llbracket 1, x_1 \llbracket$, we also have $\inf_{a \in \llbracket 1, x_1 \llbracket} \nu(\partial T_a)/\mu(T_a) > 0$. So, finally,

$$\iota = \inf_{a \in \mathbb{N}_+} \frac{\nu(\partial T_a)}{\mu(T_a)} > 0,$$

which shows that $G(\infty, 1/2)$ has a spectral gap a.s.

7 Spectral Gap for the Rado Graph

This section proves the needed spectral gap for the Rado graph. Here the graph has vertex set \mathbb{N} and an edge from i to j if i is less than j and the i th bit of j is a one. We treat carefully the case of a more general base measure, $Q(x) = (1 - \delta)\delta^x$. As delta tends to 1, sampling from this Q is a better surrogate for “pick a neighboring vertex uniformly”. Since the normalization doesn’t enter, throughout take $Q(x) = \delta^x$. The heart of the argument is a discrete version of Hardy’s inequality for trees. This is developed below with full details in Appendix 2.

Consider the transition kernel K reversible with respect to π and associated to the measure Q given by $Q(x) := \delta^x$ for all $x \in \mathbb{N}$, where $\delta \in (0, 1)$ (instead of $\delta = 1/2$ as in the introduction, up to the normalization). Recall that

$$\begin{aligned} \forall x, y \in \mathbb{N}, \quad K(x, y) &:= \frac{Q(y)}{Q(N(x))} \mathbb{1}_{N(x)}(y), \\ \forall x \in \mathbb{N}, \quad \pi(x) &= Z^{-1} Q(x) Q(N(x)), \end{aligned}$$

where $N(x)$ is the set of neighbors of x induced by K and where $Z > 0$ is the normalizing constant. Here is the equivalent of Proposition 3:

Proposition 5 *We have*

$$\lambda \geq \frac{1 - \delta}{16(2 \vee \lceil \log_2 \log_2(2/\log_2(1/\delta)) \rceil)}.$$

This bound will be proved via Hardy’s inequalities. If we resort to Dirichlet–Cheeger, we rather get

$$\lambda \geq \frac{(1 - \delta)^2}{2}. \tag{19}$$

To see the advantage of Proposition 5, let δ come closer and closer to 1, namely, approach the problematic case of “pick a neighbor uniformly at random”. In this situation, the r.h.s. of the bound of Proposition 5 is of order

$$\frac{1 - \delta}{16 \lceil \log_2 \log_2(1/(1 - \delta)) \rceil}$$

which is better than (19) as δ goes to 1–.

Here we present the Hardy’s inequalities method to get Proposition 5 announced above. Our goal is to show that K admits a positive first Dirichlet eigenvalue:

Proposition 6 *There exists $\Lambda > 0$ depending on $\delta \in (0, 1)$ such that*

$$\forall f \in L^2(\pi), \quad \Lambda \pi[(f - f(0))^2] \leq \frac{1}{2} \sum_{x, y \in \mathbb{N}} (f(y) - f(x))^2 \pi(x) K(x, y).$$

It follows that the reversible couple (K, π) admits a spectral gap bounded below by Λ given above. Indeed, it is an immediate consequence of the fact that for any $f \in L^2(\pi)$, the variance of f with respect to π is bounded above by $\pi[(f - f(0))^2]$.

The proof of Proposition 6 is based on a pruning of K and Hardy’s inequalities for trees. Consider the set of unoriented edges induced by K : $E := \{\{x, y\} \in \mathbb{N} \times \mathbb{N} : K(x, y) > 0\}$ (in particular, E does not contain the self-edges or singletons). For any $x \in \mathbb{N}_+$, let $p(x)$ the smallest bit equal to 1 in the binary expansion of x , i.e.,

$$p(x) := \min\{y \in \mathbb{N} : K(x, y) > 0\}.$$

Define the subset F of E by

$$F := \{\{x, p(x)\} \in E : x \in \mathbb{N}_+\}$$

and the function ν on F via

$$\forall \{x, p(x)\} \in F, \quad \nu(\{x, p(x)\}) := Z\pi(x)K(x, p(x)) = Q(x)Q(p(x)).$$

To any $f \in L^2(\pi)$, associate the function $(df)^2$ on F given by

$$\forall \{x, p(x)\} \in F, \quad (df)^2(\{x, p(x)\}) := (f(x) - f(p(x)))^2.$$

Finally, consider the (non-negative) measure μ defined on \mathbb{N}_+ via

$$\forall x \in \mathbb{N}_+, \quad \mu(x) := Q(x)Q(p(x)). \tag{20}$$

Then we have:

Proposition 7 *There exists $\lambda > 0$ depending on $\delta \in (0, 1)$ such that*

$$\forall f \in L^2(\mu), \quad \lambda\mu[(f - f(0))^2] \leq \sum_{e \in F} (df)^2(e)\nu(e).$$

This result implies Proposition 6. Indeed, note that by the definition of Q ,

$$\forall x \in \mathbb{N}_+, \quad Q(p(x)) \leq Q(N(x)) \leq \frac{1}{1 - \delta} Q(p(x)). \tag{21}$$

Thus, for any $f \in L^2(\mu)$,

$$\begin{aligned} \lambda\pi[(f - f(0))^2] &= \frac{\lambda}{Z} \sum_{x \in \mathbb{N}_+} (f(x) - f(0))^2 Q(x)Q(N(x)) \\ &\leq \frac{\lambda}{(1 - \delta)Z} \sum_{x \in \mathbb{N}_+} (f(x) - f(0))^2 Q(x)Q(p(x)) \end{aligned}$$

$$\begin{aligned}
 &= \frac{\lambda}{(1-\delta)Z} \mu[(f - f(0))^2] \leq \frac{1}{(1-\delta)Z} \sum_{e \in F} (df)^2(e) \nu(e) \\
 &\leq \frac{1}{2(1-\delta)} \sum_{x,y \in \mathbb{N}} (f(y) - f(x))^2 \pi(x) K(x, y)
 \end{aligned}$$

namely Proposition 6 holds with $\Lambda := \lambda(1 - \delta)$.

Note that \mathbb{N} endowed with the set of non-oriented edges F has the structure of a tree. We interpret 0 as its root, so that for any $x \in \mathbb{N}_+$, $p(x)$ is the parent of x . Note that for any $x \in \mathbb{N}$, the children of x are exactly the numbers $y2^x$, where y is an odd number. We will denote $h(x)$ the height of x with respect to the root 0 (thus, the odd numbers are exactly the elements of \mathbb{N} whose height is equal to 1).

According to [21] (see also Evans, Harris and Pick [10]), the best constant λ in Proposition 7, say λ_0 , can be estimated up to a factor 16 via Hardy’s inequalities for trees, see (23) below. To describe them we need several notations.

Let \mathcal{T} the set of subsets T of \mathbb{N}_+ satisfying the following conditions

- T is non-empty and connected (with respect to F),
- T does not contain 0,
- there exists $M \geq 1$ such that $h(x) \leq M$ for all $x \in T$,
- if $x \in T$ has a child in T , then all children of x belong to T .

Note that any $T \in \mathcal{T}$ admits a closest element to 0, call it $m(T)$. Note that $m(T) \neq 0$. When T is not reduced to the singleton $\{m(T)\}$, then $T \setminus \{m(T)\}$ has a denumerable infinity of connected components which are indexed by the children of $m(T)$. Since these children are exactly the $y2^{m(T)}$, where $y \in \mathcal{I}$, the set of odd numbers, call $T_{y2^{m(T)}}$ the connected component of $T \setminus \{m(T)\}$ associated to $y2^{m(T)}$. Note that $T_{y2^{m(T)}} \in \mathcal{T}$. We extend ν as a functional on \mathcal{T} , via the iteration

- when T is the singleton $\{m(T)\}$, we take $\nu(T) := \nu(\{m(T), p(m(T))\})$,
- when T is not a singleton, decompose T as $\{m(T)\} \sqcup \bigsqcup_{y \in \mathcal{I}} T_{y2^{m(T)}}$, then ν is defined as

$$\frac{1}{\nu(T)} = \frac{1}{\nu(\{m(T)\})} + \frac{1}{\sum_{y \in \mathcal{I}} \nu(T_{y2^{m(T)}})}. \tag{22}$$

For $x \in \mathbb{N}_+$, let S_x be the set of vertices $y \in \mathbb{N}_+$ whose path to 0 passes through x . For any $T \in \mathcal{T}$ we associate the subset

$$T^* := (S_{m(T)} \setminus T) \sqcup L(T)$$

where $L(T)$ is the set of leaves of T , namely the $x \in T$ having no children in T . Equivalently, T^* is the set of all descendants of the leaves of T , themselves included.

Consider $\mathcal{S} \subset \mathcal{T}$ the set of $T \in \mathcal{T}$ which are such that $m(T)$ is an odd number. Finally, define

$$A := \sup_{T \in \mathcal{S}} \frac{\mu(T^*)}{\nu(T)}.$$

We are interested in this quantity because of the Hardy inequalities:

$$A \leq \frac{1}{\lambda_0} \leq 16A, \tag{23}$$

where recall that λ_0 is the best constant in Proposition 7. (In [21], only finite trees were considered, the extension to infinite trees is given in Appendix 2). So, to prove Proposition 7, it is sufficient to show that A is finite. To investigate A , we need some further definitions. For any $x \in \mathbb{N}_+$, let

$$b(x) := \frac{Q(2^x)}{Q(p(x))}.$$

A finite path from 0 in the direction to infinity is a finite sequence $z := (z_n)_{n \in \llbracket 0, N \rrbracket}$ of elements of \mathbb{N}_+ such that $z_0 = 0$ and $p(z_n) = z_{n-1}$ for any $n \in \llbracket 1, N \rrbracket$. On such a path z , we define the quantity

$$B(z) := \sum_{n \in \llbracket 1, N \rrbracket} b(z_n).$$

The following technical result is crucial for our purpose of showing that A is finite.

Lemma 6 *For any finite path from 0 in the direction to infinity $z := (z_n)_{n \in \llbracket 0, N \rrbracket}$, we have $B(z) \leq C$, where $C := \sum_{l \in \mathbb{N}} \delta^{2^l - l} < +\infty$.*

Proof Note that for any $n \in \llbracket 1, N \rrbracket$, $h(z_n) = n$. Furthermore, for any $x \in \mathbb{N}_+$, we have $h(x) \leq x$ and we get $h(p(z_n)) = h(z_n) - 1 = n - 1$, so that $p(z_n) \geq n - 1$. Writing $z_n = y_n 2^{p(z_n)}$, for some odd number y_n , it follows that

$$b(z_n) = \frac{Q(2^{y_n 2^{p(z_n)}})}{Q(p(z_n))} = \delta^{2^{y_n 2^{p(z_n)}} - p(z_n)} \leq \delta^{2^{2^{p(z_n)}} - p(z_n)} \leq \delta^{2^{2^{n-1}} - n - 1}.$$

The desired result follows at once. □

We need two ingredients about ratios $\mu(T^*)/\nu(T)$. Here is the first one.

Lemma 7 *For any $T \in \mathcal{T}$ which is a singleton, we have $\frac{\mu(T^*)}{\nu(T)} \leq \frac{1}{1-\delta}$.*

Proof When T is the singleton $\{m(T)\}$, on the one hand we have

$$\nu(T) = \nu(\{p(m(T)), m(T)\}) = \mu(m(T)).$$

On the other hand, T^* is the subtree growing from $m(T)$, namely the subtree containing all the descendants of $m(T)$. Note two properties of T^* :

$$T^* \subset \{y \in \mathbb{N}_+ : y \geq m(T)\} \text{ and } \forall y \in T^*, p(y) \geq p(m(T)), \tag{24}$$

and we further have $p(y) \geq m(T)$ for any $y \in T^* \setminus \{m(T)\}$. It follows that

$$\begin{aligned} \mu(T^*) &= \sum_{y \in T^*} Q(y)Q(p(y)) \leq Q(p(m(T))) \sum_{y \geq m(T)} Q(y) = Q(p(m(T))) \sum_{y \geq m(T)} \delta^y \\ &= Q(p(m(T))) \frac{Q(m(T))}{1 - \delta} = \frac{1}{1 - \delta} \mu(m(T)). \end{aligned} \tag{25}$$

Thus, we get $\mu(T^*)/\nu(T) \leq \frac{1}{1-\delta}$. □

For the second ingredient, we need some further definitions. The length $\ell(T)$ of $T \in \mathcal{T}$ is given by $\ell(T) := \max_{x \in T} h(x) - \min_{x \in T} h(x)$, and for any $l \in \mathbb{N}$, we define

$$\mathcal{T}_l := \{T \in \mathcal{T} : \ell(T) \leq l\}$$

Lemma 8 *For any $l \in \mathbb{N}$, we have $\sup_{T \in \mathcal{T}_l} \frac{\mu(T^*)}{\nu(T)} < +\infty$.*

Proof We will prove the finiteness by induction over $l \in \mathbb{N}$. First, note that \mathcal{T}_0 is the set of singletons, and so Lemma 7 implies that $\sup_{T \in \mathcal{T}_0} \frac{\mu(T^*)}{\nu(T)} \leq \frac{1}{1-\delta}$. Next, assume that the supremum is finite for some $l \in \mathbb{N}$ and let us show that it is also finite for $l + 1$.

Consider $T \in \mathcal{T}_{l+1}$, with $\ell(T) = l + 1$; in particular, T is not a singleton. Decompose T as $\{m(T)\} \sqcup \bigsqcup_{y \in \mathcal{I}} T_{y2^m(T)}$ and recall the relation (22). Since $T^* = \bigsqcup_{y \in \mathcal{I}} T_{y2^m(T)}^*$, it follows that

$$\begin{aligned} \frac{\mu(T^*)}{\nu(T)} &= \sum_{y \in \mathcal{I}} \mu(T_{y2^m(T)}^*) \left(\frac{1}{\nu(\{m(T)\})} + \frac{1}{\sum_{y \in \mathcal{I}} \nu(T_{y2^m(T)})} \right) \\ &= \frac{\sum_{y \in \mathcal{I}} \mu(T_{y2^m(T)}^*)}{\nu(\{m(T)\})} + \frac{\sum_{y \in \mathcal{I}} \mu(T_{y2^m(T)}^*)}{\sum_{y \in \mathcal{I}} \nu(T_{y2^m(T)})} \\ &\leq \frac{\mu(\bigsqcup_{y \in \mathcal{I}} T_{y2^m(T)}^*)}{\mu(m(T))} + \sup \left\{ \frac{\mu(T_{y2^m(T)}^*)}{\nu(T_{y2^m(T)})} : y \in \mathcal{I} \right\}. \end{aligned} \tag{26}$$

Consider the first term on the right. Given $y \in \mathcal{I}$, the smallest possible element of $T_{y2^m(T)}^*$ is $y2^{m(T)}$, and we have for any $x \in T_{y2^m(T)}^*$,

$$p(x) \geq p(y2^{m(T)}) = m(T).$$

Thus we have the equivalent of (24):

$$\bigsqcup_{y \in \mathcal{I}} T_{y2^{m(T)}}^* \subset \{y \in \mathbb{N}_+ : y \geq 2^{m(T)}\}, \forall x \in \bigsqcup_{y \in \mathcal{I}} T_{y2^{m(T)}}^*, p(x) \geq m(T). \quad (27)$$

Following the computation (25), we get

$$\mu \left(\bigsqcup_{y \in \mathcal{I}} T_{y2^{m(T)}}^* \right) < \frac{1}{1 - \delta} Q(m(T)) Q(2^{m(T)}),$$

where the inequality is strict, because in (27) we cannot have equality for all $x \in \bigsqcup_{y \in \mathcal{I}} T_{y2^{m(T)}}^*$. It follows that

$$\frac{\sum_{y \in \mathcal{I}} \mu(T_{y2^{m(T)}}^*)}{\mu(m(T))} < \frac{1}{1 - \delta} \frac{Q(m(T)) Q(2^{m(T)})}{Q(m(T)) Q(p(m(T)))} = \frac{b(m(T))}{1 - \delta} \leq \frac{C}{1 - \delta} \quad (28)$$

where C is the constant introduced in Lemma 6. Since for any $y \in \mathcal{I}$, we have $T_{y2^{m(T)}} \in \mathcal{T}_l$, we deduce the desired result from the induction hypothesis. \square

We are now ready to prove Proposition 7.

Proof (Of Proposition 7) Fix some $T \in \mathcal{S}$, we are going to show that $\mu(T^*)/v(T) \leq 1 + C/(1 - \delta)$, where C is the constant introduced in Lemma 6. Due to Lemma 7, this bound is clear if T is a singleton. When T is not the singleton $\{m(T)\}$, decompose T as $\{m(T)\} \sqcup \bigsqcup_{y \in \mathcal{I}} T_{y2^{m(T)}}$ and let us come back to (26). Denote $z_1 := m(T)$ and

$$\epsilon := \frac{b(z_1)}{1 - \delta} - \frac{\sum_{y \in \mathcal{I}} \mu(T_{y2^{m(T)}}^*)}{\mu(m(T))}$$

which is positive according to (28). Coming back to (26), we have shown

$$\frac{\mu(T^*)}{v(T)} \leq \frac{b(z_1)}{1 - \delta} + \frac{\mu(T_{z_2}^*)}{v(T_{z_2})}$$

where $z_2 \in \{y2^{m(T)} : y \in \mathcal{I}\}$ is such that

$$\sup \left\{ \frac{\mu(T_{y2^{m(T)}}^*)}{v(T_{y2^{m(T)}})} : y \in \mathcal{I} \right\} \leq \frac{\mu(T_{z_2}^*)}{v(T_{z_2})} + \epsilon.$$

To get the existence of z_2 , we used that the supremum is finite, as ensured by Lemma 8.

By iterating this procedure, define a finite path from 0 in the direction to infinity $z := (z_n)_{n \in \llbracket 0, N \rrbracket}$, such that for any $n \in \llbracket 1, N - 1 \rrbracket$,

$$\frac{\mu(T_{z_n}^*)}{\nu(T_{z_n})} \leq \frac{b(z_n)}{1 - \delta} + \frac{\mu(T_{z_{n+1}}^*)}{\nu(T_{z_{n+1}})}$$

and T_{z_N} is a singleton. We have $N \leq \max\{h(x) : x \in T\}$. We deduce that

$$\frac{\mu(T^*)}{\nu(T)} \leq \frac{B(z)}{1 - \delta} + \frac{\mu(T_{z_N}^*)}{\nu(T_{z_N})} \leq \frac{C + 1}{1 - \delta},$$

as desired. □

To get an explicit bound in terms of δ , it remains to investigate the quantity C .

Lemma 9 *We have*

$$C \leq \begin{cases} 2 & \text{if } \delta \in (0, 1/\sqrt{2}], \\ 1 + \lceil \log_2 \log_2 \left(\frac{2}{\log_2(1/\delta)} \right) \rceil & \text{if } \delta \in (1/\sqrt{2}, 1). \end{cases}$$

Proof Consider $l_0 := \min\{l \in \mathbb{N}_+ : \delta^{2^{2^l - 1}} \leq 1/2\}$. Elementary computations show that

$$\forall l \geq 1, \quad 2^{2^{l+1}} - l - 1 \geq 2(2^{2^l} - l),$$

so we get

$$\sum_{l \geq l_0} \delta^{2^{2^l - 1}} \leq \sum_{n \geq 0} \frac{1}{2^{2^n}} \leq \sum_{n \geq 1} \frac{1}{2^n} = 1.$$

Since we have for any $l \in \mathbb{N}$, $2^{2^l} - l \geq 0$, we deduce

$$C \leq 1 + \sum_{l \in \llbracket 0, l_0 - 1 \rrbracket} \delta^{2^{2^l - 1}} \leq 1 + l_0.$$

It is not difficult to check that for any $l \geq 1$, $2^{2^l} - l \geq \frac{1}{2}2^{2^l}$, so that

$$\begin{aligned} l_0 &= \min\{l \in \mathbb{N}_+ : 2^{2^l} - l \geq 1/\log_2(1/\delta)\} \leq \min\{l \in \mathbb{N}_+ : 2^{2^l} \geq 2/\log_2(1/\delta)\} \\ &= 1 \vee \lceil \log_2 \log_2(2/\log_2(1/\delta)) \rceil. \end{aligned}$$

The announced result follows from the fact

$$\log_2 \log_2(2/\log_2(1/\delta)) \geq 1 \Leftrightarrow \delta \geq \frac{1}{\sqrt{2}}.$$

□

The following observations show that Q needs to be at least decaying exponentially for the Hardy inequality approach to work.

Remark 2

(a) In view of the expression of π , it is natural to try to replace (20) by

$$\forall x \in \mathbb{N}_+, \quad \mu(x) := Z\pi(x) = Q(x)Q(N(x)).$$

But then in Lemma 7, where we want the ratios $\mu(T^*)/\nu(T)$ to be bounded above for singletons T , we end up with the fact that

$$\frac{Q(N(m(T)))}{Q(p(m(T)))} = \frac{\mu(T)}{\nu(T)} \leq \frac{\mu(T^*)}{\nu(T)}$$

must be bounded above for singletons T . Namely an extension of (21) must hold: there exists a constant $c > 0$ such that

$$\forall x \in \mathbb{N}_+, \quad Q(N(x)) \leq cQ(p(x)). \tag{29}$$

Writing $x = y2^p$, with $y \in \mathcal{I}$ and $p \in \mathbb{N}$, we must have $Q(N(y2^p)) \leq cQ(p)$. Take $y = 1 + 2 + 4 + \dots + 2^l$, then we get that $p, p + 1, \dots, p + l$ all belong to $Q(N(y2^p))$, so that $Q(\{p, p + 1, \dots, p + l\}) \leq cQ(p)$, and letting l go to infinity, it follows that $Q(\llbracket p, \infty \rrbracket) \leq cQ(p)$, namely, Q has exponential tails.

- (b) Other subtrees of the graph generated by K could have been considered. It amounts to choose the parent of any $x \in \mathbb{N}_+$. But among all possible choices of such a neighbor, the one with most weight is $p(x)$, at least if Q is decreasing. In view of the requirement (29), it looks like the best possible choice.
- (c) If one is only interested in Proposition 7 with μ defined by (20), then many more probability measures Q can be considered, in particular any polynomial probability of the form $Q(x) := \frac{1}{\zeta(l)(x+1)^l}$, for any $x \in \mathbb{N}$, where ζ is the Riemann function and $l > 1$.

Acknowledgments We thank Peter Cameron, Maryanthe Malliaris, Sebastien Martineau, Yuval Peres, and Laurent Saloff-Coste for their help. S. C.’s research was partially supported by NSF grants DMS-1855484 and DMS-2113242. P. D.’s research was partially supported by NSF grant DMS-1954042. L. M. acknowledges funding from ANR grant ANR-17-EUR-0010.

Appendix 1: Dirichlet–Cheeger Inequalities

We begin by showing the Dirichlet–Cheeger inequality that we have been using in the previous sections. It is a direct extension (even simplification) of the proof of the Cheeger inequality given in Saloff-Coste [23]. We end this appendix by proving that it is in general not possible to compare linearly the Dirichlet–Cheeger constant of an absorbed Markov chain with the largest Dirichlet–Cheeger constant induced on a spanning subtree.

Let us work in continuous time. Consider L a sub-Markovian generator on a finite set V . Namely, $L := (L(x, y))_{x, y \in V}$, whose off-diagonal entries are non-negative and whose row sums are non-positive. Assume that L is irreducible and reversible with respect to a probability π on V .

Let $\lambda(L)$ be the smallest eigenvalue of $-L$ (often called the Dirichlet eigenvalue). The variational formula for eigenvalues shows that

$$\lambda(L) = \min_{f \in \mathbb{R}^V \setminus \{0\}} \frac{-\pi[fL[f]]}{\pi[f^2]}. \tag{30}$$

The Dirichlet–Cheeger constant $\iota(L)$ is defined similarly, except that only indicator functions are considered in the minimum:

$$\iota(L) = \min_{A \subset V, A \neq \emptyset} \frac{-\pi[\mathbb{1}_A L[\mathbb{1}_A]]}{\pi[A]}. \tag{31}$$

Here is the Dirichlet–Cheeger inequality:

Theorem 5 *Assuming $L \neq 0$, we have*

$$\frac{\iota(L)^2}{2\ell(L)} \leq \lambda(L) \leq \iota(L)$$

where $\ell(L) := \max\{|L(x, x)| : x \in V\} > 0$.

When L is Markovian, the above inequalities are trivial and reduce to $\iota(L) = \lambda(L) = 0$. Indeed, it is sufficient to consider $f = \mathbb{1}$ and $A = V$ respectively in the r.h.s. of (30) and (31). Thus there is no harm in supposing furthermore that L is strictly sub-Markovian: at least one of the row sums is negative. To bring this situation back to a Markovian setting, it is usual to extend V into $\overline{V} := V \sqcup \{0\}$ where $0 \notin V$ is a new point. Then one introduces the extended Markov generator \overline{L} on \overline{V} via

$$\forall x, y \in \overline{V}, \quad \overline{L}(x, y) := \begin{cases} L(x, y) & \text{if } x, y \in V, \\ -\sum_{z \in V} L(x, z) & \text{if } y = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Note that the point 0 is absorbing for the Markov processes associated to \bar{L} .

It is convenient to give another expression for $\iota(L)$. Consider the set of edges $\bar{E} := \{\{x, y\} : x \neq y \in \bar{V}\}$. We define a measure μ on \bar{E} :

$$\forall e := \{x, y\} \in \bar{E}, \quad \mu(e) := \begin{cases} \pi(x)L(x, y) & \text{if } x, y \in V, \\ \pi(x)\bar{L}(x, 0) & \text{if } y = 0, \\ \pi(y)\bar{L}(y, 0) & \text{if } x = 0. \end{cases}$$

(Note that the reversibility assumption was used to ensure that the first line is well-defined.) Extend any $f \in \mathbb{R}^V$ into the function \bar{f} on \bar{V} by making it vanish at 0 and define

$$\forall e := \{x, y\} \in \bar{E}, \quad |d\bar{f}|(e) := |\bar{f}(y) - \bar{f}(x)|.$$

With these definitions we can check that

$$\forall f \in \mathbb{R}^V, \quad -\pi[fL[f]] = \sum_{e \in \bar{E}} |d\bar{f}|^2(e)\mu(e).$$

These notations enable to see (31) as a L^1 version of (30):

Proposition 8 *We have*

$$\iota(L) = \min_{f \in \mathbb{R}^V \setminus \{0\}} \frac{\sum_{e \in \bar{E}} |d\bar{f}|(e)\mu(e)}{\pi[|f|]}.$$

Proof Restricting the minimum in the r.h.s. to indicator functions, we recover the r.h.s. of (31). It is thus sufficient to show that for any given $f \in \mathbb{R}^V \setminus \{0\}$,

$$\frac{\sum_{e \in \bar{E}} |d\bar{f}|(e)\mu(e)}{\pi[|f|]} \geq \iota(L). \tag{32}$$

Note that $|d\bar{f}|(e) \geq |d|f||(e)$ for any $e \in \bar{E}$, so without loss of generality, we can assume $f \geq 0$. For any $t \geq 0$, consider the set F_t and its indicator function given by

$$F_t := \{\bar{f} > t\} = \{f > t\} \text{ and } f_t := \mathbf{1}_{F_t}.$$

Note that

$$\forall x \in V, \quad f(x) = \int_0^{+\infty} f_t(x) dt,$$

so that by integration,

$$\pi[f] = \int_0^{+\infty} \pi[F_t] dt.$$

Furthermore, we have

$$\begin{aligned} \sum_{e \in \bar{E}} |d\bar{f}|(e)\mu(e) &= \sum_{e=\{x,y\}:\bar{f}(y) > \bar{f}(x)} (\bar{f}(y) - \bar{f}(x))\mu(e) = \sum_{e=\{x,y\}:\bar{f}(y) > \bar{f}(x)} \int_{\bar{f}(x)}^{\bar{f}(y)} \mu(e) dt \\ &= \int_0^{+\infty} \sum_{e=\{x,y\}:\bar{f}(y) > t \geq \bar{f}(x)} \mu(e) dt = \int_0^{+\infty} \mu(\partial F_t) dt, \end{aligned}$$

where for any $A \subset V$, we define

$$\partial A := \{\{x, y\} \in \bar{E} : x \in A \text{ and } y \notin A\}.$$

Note that for any such A , we have $\mu(\partial A) = -\pi[\mathbb{1}_A L[\mathbb{1}_A]]$, so that

$$\sum_{e \in \bar{E}} |d\bar{f}|(e)\mu(e) = - \int_0^{+\infty} \pi[f_t L[f_t]] dt \geq \iota(L) \int_0^{+\infty} \pi[F_t] dt = \iota(L)\pi[f],$$

showing (32). □

Proof (Of Theorem 5) Given $g \in \mathbb{R}^V$, let $f = g^2$. By Proposition 8, we compute

$$\begin{aligned} \iota(L)\pi[f] &\leq \sum_{e \in \bar{E}} |d\bar{f}|(e)\mu(e) = \sum_{e=\{x,y\} \in \bar{E}} |\bar{g}^2(y) - \bar{g}^2(x)|\mu(e) \\ &= \sum_{e=\{x,y\} \in \bar{E}} |\bar{g}(y) - \bar{g}(x)| |\bar{g}(y) + \bar{g}(x)|\mu(e) \\ &\leq \sqrt{\sum_{e=\{x,y\} \in \bar{E}} (\bar{g}(y) - \bar{g}(x))^2 \mu(e)} \sqrt{\sum_{e=\{x,y\} \in \bar{E}} (\bar{g}(y) + \bar{g}(x))^2 \mu(e)} \\ &\leq \sqrt{-\pi[gL[g]]} \sqrt{2 \sum_{e=\{x,y\} \in \bar{E}} (\bar{g}^2(y) + \bar{g}^2(x))\mu(e)} \\ &= \sqrt{-\pi[gL[g]]} \sqrt{4 \sum_{e=\{x,y\} \in \bar{E}} \bar{g}^2(x)\mu(e)} \\ &= \sqrt{-\pi[gL[g]]} \sqrt{2 \sum_{x \in V} g^2(x)\pi(x) \sum_{y \in \bar{V} \setminus \{x\}} \bar{L}(x, y)} \end{aligned}$$

$$\begin{aligned}
 &= \sqrt{-\pi[gL[g]]} \sqrt{2 \sum_{x \in V} g^2(x)\pi(x)|L(x, x)|} \\
 &\leq \sqrt{2\ell(L)}\sqrt{-\pi[gL[g]]}\sqrt{\pi[g^2]} = \sqrt{2\ell(L)}\sqrt{-\pi[gL[g]]}\sqrt{\pi[f]}.
 \end{aligned}$$

Thus, we have

$$\frac{\iota(L)^2}{2\ell(L)}\pi[g^2] \leq -\pi[gL[g]],$$

which gives the desired lower bound for $\lambda(L)$. The upper bound is immediate. \square

The unoriented graph associated to L is $\overline{G} := (\overline{V}, \overline{E}_L)$ where $\overline{E}_L := \{e \in \overline{E} : \mu(e) > 0\}$. Consider \mathbb{T} , the set of all subtrees of \overline{G} , and for any $T \in \mathbb{T}$, consider the sub-Markovian generator L_T on V associated to T via

$$L_T(x, y) := \begin{cases} L(x, y) & \text{if } \{x, y\} \in \overline{E}(T), \\ -\sum_{z \in V \setminus \{x\}} L_T(x, z) & \text{if } x = y \text{ and } \{x, 0\} \notin \overline{E}(T), \\ -\sum_{z \in V \setminus \{x\}} L_T(x, z) - \overline{L}(x, 0) & \text{if } x = y \text{ and } \{x, 0\} \in \overline{E}(T), \\ 0 & \text{otherwise,} \end{cases}$$

where $x, y \in V$ and $\overline{E}(T)$ is the set of (unoriented) edges of T .

Note that L_T is also reversible with respect to π (it is irreducible if and only if 0 belongs to a unique edge of $\overline{E}(T)$). Denote μ_T the corresponding measure on \overline{E} . It is clear that $\mu_T \leq \mu$, so we get $\iota(L_T) \leq \iota(L)$. In the spirit of Benjamini and Schramm [2], we may wonder if conversely, $\iota(L)$ could be bounded above in terms of $\max_{T \in \mathbb{T}} \iota(L_T)$. A linear comparison is not possible:

Proposition 9 *It does not exist a universal constant $\chi > 0$ such that for any L as above, $\chi \iota(L) \leq \max_{T \in \mathbb{T}} \iota(L_T)$.*

Proof Let us construct a family $(L^{(n)})_{n \in \mathbb{N}_+}$ of sub-Markovian generators such that

$$\lim_{n \rightarrow \infty} \frac{\max_{T \in \mathbb{T}} \iota(L_T^{(n)})}{\iota(L^{(n)})} = 0 \tag{33}$$

For any $n \in \mathbb{N}_+$, the state space $V^{(n)}$ of $L^{(n)}$ is $\llbracket n \rrbracket \times \{0, 1\}$ (more generally, all notions associated to $L^{(n)}$ will be marked by the exponent (n)). Denote $V_0^{(n)} := \llbracket n \rrbracket \times \{0\}$ and $V_1^{(n)} := \llbracket n \rrbracket \times \{1\}$. We take

$$L^{(n)}(x, y) := \begin{cases} \epsilon & \text{if } x \in V_i^{(n)}, y \in V_{1-i}^{(n)} \text{ with } i \in \{0, 1\}, \\ n\epsilon + 1 & \text{if } x = y \in V_0^{(n)}, \\ n\epsilon & \text{if } x = y \in V_1^{(n)}, \\ 0 & \text{otherwise,} \end{cases}$$

where $x, y \in V^{(n)}$, and $\epsilon > 0$, that will depend on n , is such that $n\epsilon < 1/2$.

Recall that 0 is the cemetery point added to $V^{(n)}$, we have

$$\forall x \in V^{(n)}, \quad \bar{L}^{(n)}(x, 0) = \begin{cases} 1 & \text{if } x \in V_0^{(n)}, \\ 0 & \text{if } x \in V_1^{(n)}. \end{cases}$$

Note that $\pi^{(n)}$ is the uniform probability on $V^{(n)}$. Let us show that

$$\iota(L^{(n)}) = n\epsilon. \tag{34}$$

Consider any $\emptyset \neq A \subset V^{(n)}$, and decompose $A = A_0 \sqcup A_1$, with $A_0 := A \cap V_0^{(n)}$ and $A_1 := A \cap V_1^{(n)}$. Denote $a_0 := |A_0|$ and $a_1 := |A_1|$. We have that ∂A is given by

$$\{\{x, y\} : x \in A_0, y \in V_1^{(n)} \setminus A_1\} \sqcup \{\{x, y\} : x \in V_0^{(n)} \setminus A_0, y \in A_1\} \sqcup \{\{x, 0\} : x \in A_0\},$$

and thus $\mu^{(n)}(\partial A) = \frac{1}{2n}(\epsilon(a_0(n - a_1) + a_1(n - a_0)) + a_0)$. It follows that

$$\frac{\mu^{(n)}(\partial A)}{\pi^{(n)}(A)} = n\epsilon + \frac{a_0(1 - 2\epsilon a_1)}{a_0 + a_1}.$$

Taking into account that $1 - 2\epsilon a_1 > 0$, the r.h.s. is minimized with respect to $a_0 \in \llbracket 0, n \rrbracket$ when $a_0 = 0$ and we then get (independently of a_1), $\mu^{(n)}(\partial A)/\pi^{(n)}(A) = n\epsilon$. We deduce (34).

Consider any $T \in \mathbb{T}^{(n)}$ and let us check that

$$\iota(L_T^{(n)}) \leq \epsilon. \tag{35}$$

Observe there exists $x \in V_1^{(n)}$ such that there is a unique $y \in V_0^{(n)}$ with $\{x, y\}$ being an edge of T . Indeed, put on the edges of T the orientation toward the root 0. Thus from any vertex $x \in V_1^{(n)}$ there is a unique exiting edge (but it is possible there are several incoming edges). Necessarily, there is a vertex in $V_0^{(n)}$ whose edge exits to 0. So there are at most $n - 1$ vertices from $V_0^{(n)}$ whose exit edge points toward $V_1^{(n)}$. In particular, there is at least one vertex from $V_1^{(n)}$ which is not pointed out by a vertex

from $V_0^{(n)}$. We can take x to be this vertex from $V_1^{(n)}$ and $y \in V_0^{(n)}$ is the vertex pointed out by the oriented edge exiting from x .

Considering the singleton $\{x\}$, we get

$$\mu_T^{(n)}(\partial\{x\}) = \mu_T(\{x, y\}) = \frac{\epsilon}{2n} \text{ and } \pi^{(n)}(x) = \frac{1}{2n}.$$

implying (35) (a little more work would prove that an equality holds there). As a consequence, we see that $\max_{T \in \mathbb{T}^{(n)}} \iota(L_T^{(n)}) \leq \epsilon$. Taking for instance $\epsilon := 1/(4n)$ to fulfill the condition $n\epsilon < 1/2$, we obtain $\frac{\max_{T \in \mathbb{T}^{(n)}} \iota(L_T^{(n)})}{\iota(L^{(n)})} \leq \frac{1}{n}$, and (33) follows. \square

Appendix 2: Hardy’s Inequalities

Our goal here is to extend the validity of Hardy’s inequalities on finite trees to general denumerable trees, without assumption of local finiteness. We begin by recalling the Hardy’s inequalities on finite trees. Consider $\mathcal{T} = (\overline{V}, \overline{E}, 0)$ a finite tree rooted in 0, whose vertex and (undirected) edge sets are \overline{V} and \overline{E} . Denote $V := \overline{V} \setminus \{0\}$, for each $x \in V$, the parent $p(x)$ of x is the neighbor of x in the direction of 0. The other neighbors of x are called the children of x and their set is written $C(x)$. For $x = 0$, by convention $C(0)$ is the set of neighbors of 0. Let be given two positive measures μ, ν defined on V . Consider $c(\mu, \nu)$ the best constant $c \geq 0$ in the inequality

$$\forall f \in \mathbb{R}^V, \quad \mu[f^2] \leq c \sum_{x \in V} (f(p(x)) - f(x))^2 \nu(x) \tag{36}$$

where f was extended to 0 via $f(0) := 0$.

According to [21] (see also Evans, Harris and Pick [10]), $c(\mu, \nu)$ can be estimated up to a factor 16 via Hardy’s inequalities for trees, see (39) below. To describe them we need several notations.

Let \mathbb{T} the set of subsets T of V satisfying the following conditions

- T is non-empty and connected (in \mathcal{T}),
- T does not contain 0,
- if $x \in T$ has a child in T , then all children of x belong to T .

Note that any $T \in \mathbb{T}$ admits a closest element to 0, call it $m(T)$, we have $m(T) \neq 0$. When T is not reduced to the singleton $\{m(T)\}$, the connected components of $T \setminus \{m(T)\}$ are indexed by the set of the children of $m(T)$, namely $C(m(T))$. For $y \in C(m(T))$, denote by T_y the connected component of $T \setminus \{m(T)\}$ containing y . Note that $T_y \in \mathbb{T}$.

We extend ν as a functional on \mathbb{T} , via the iteration

- when T is the singleton $\{m(T)\}$, we take $\nu(T) := \nu(m(T))$,
- when T is not a singleton, decompose T as $\{m(T)\} \sqcup \bigsqcup_{y \in C(m(T))} T_y$, then ν satisfies

$$\frac{1}{\nu(T)} = \frac{1}{\nu(m(T))} + \frac{1}{\sum_{y \in C(m(T))} \nu(T_y)}. \tag{37}$$

For $x \in V$, let S_x be the set of vertices $y \in V$ whose path to 0 pass through x . For any $T \in \mathbb{T}$ we associate the subset

$$T^* := (S_{m(T)} \setminus T) \sqcup L(T)$$

where $L(T)$ is the set of leaves of T , namely the $x \in T$ having no children in T . Equivalently, T^* is the set of all descendants of the leaves of T , themselves included.

Consider $\mathbb{S} \subset \mathbb{T}$, the set of $T \in \mathbb{T}$ which are such that $m(T)$ is a child of 0. Finally, define

$$b(\mu, \nu) := \max_{T \in \mathbb{S}} \frac{\mu(T^*)}{\nu(T)}. \tag{38}$$

We are interested in this quantity because of the Hardy inequality:

$$b(\mu, \nu) \leq c(\mu, \nu) \leq 16 b(\mu, \nu). \tag{39}$$

Our goal here is to extend this inequality to the situation where V is denumerable and where μ and ν are two positive measures on V , with $\sum_{x \in V} \mu(x) < +\infty$.

Remark 3 Without lost of generality, we can assume 0 has only one child, because what happens on different S_x and S_y , where both x and y are children of 0, can be treated separately.

More precisely, while V is now (denumerable) infinite, we first assume that the height of $\mathcal{T} := (\overline{V}, \overline{E}, 0)$ is finite (implying that \mathcal{T} cannot be locally finite). Recall that the height $h(x)$ of a vertex $x \in \overline{V}$ is the smallest number of edges linking x to 0. The assumption that $\sup_{x \in \overline{V}} h(x) < +\infty$ has the advantage that the iteration (37) enables us to compute ν on \mathbb{T} , starting from the highest vertices from an element of \mathbb{T} . Then $b(\mu, \nu)$ is defined exactly as in (38), except the maximum has to be replaced by a supremum. Extend $c(\mu, \nu)$ as the minimal constant $c \geq 0$ such that (36) is satisfied, with the possibility that $c(\mu, \nu) = +\infty$ when there is no such c . Note that

in (36), the space \mathbb{R}^V can be reduced and replaced by $\mathcal{B}(V)$, the space of bounded mappings on V :

Lemma 10 *We have*

$$c(\mu, \nu) = \sup_{f \in \mathcal{B}(V) \setminus \{0\}} \frac{\mu[f^2]}{\sum_{x \in V} (f(p(x)) - f(x))^2 \nu(x)}.$$

Proof Denote $\tilde{c}(\mu, \nu)$ the above r.h.s. A priori we have $c(\mu, \nu) \geq \tilde{c}(\mu, \nu)$. To prove the reverse bound, consider any $f \in \mathbb{R}^V$ and consider for $M > 0$, $f_M := (f \wedge M) \vee (-M)$. Note that

$$\sum_{x \in V} (f_M(p(x)) - f_M(x))^2 \nu(x) \leq \sum_{x \in V} (f(p(x)) - f(x))^2 \nu(x).$$

(This a general property of Dirichlet forms and comes from the 1-Lipschitzianity of the mapping $\mathbb{R} \ni r \mapsto (r \wedge M) \vee (-M)$.) Since $f_M \in \mathcal{B}(V)$, we have

$$\begin{aligned} \mu[f_M^2] &\leq \tilde{c}(\mu, \nu) \sum_{x \in V} (f_M(p(x)) - f_M(x))^2 \nu(x) \\ &\leq \tilde{c}(\mu, \nu) \sum_{x \in V} (f(p(x)) - f(x))^2 \nu(x). \end{aligned}$$

Letting M go to infinity, we get at the limit by monotonous convergence

$$\mu[f^2] \leq \tilde{c}(\mu, \nu) \sum_{x \in V} (f(p(x)) - f(x))^2 \nu(x).$$

Since this is true for all $f \in \mathbb{R}^V$, we deduce that $c(\mu, \nu) \leq \tilde{c}(\mu, \nu)$. □

Consider $(x_n)_{n \in \mathbb{N}_+}$ an exhaustive sequence of \bar{V} , with $x_0 = 0$ and such that for any $n \in \mathbb{N}_+$, $\bar{V}_n := \{x_0, x_1, \dots, x_n\}$ is connected. We denote \mathcal{T}_n the tree rooted on 0 induced by \mathcal{T} on \bar{V}_n and as above, $V_n := \bar{V}_n \setminus \{0\} = \{x_1, \dots, x_n\}$. For any $n \in \mathbb{N}_+$ and $x \in V_n$, introduce the set

$$R_n(x) := \{x\} \bigsqcup_{y \in \mathcal{C}(x) \setminus V_n} S_y.$$

In words, this is the set of elements of V whose path to 0 first enters V_n at x .

From now on, we assume that 0 has only one child, taking into account Remark 3. It follows that

$$V = \bigsqcup_{x \in V_n} R_n(x). \tag{40}$$

Let μ_n and ν_n be the measures defined on V_n via

$$\forall x \in V_n, \quad \begin{cases} \mu_n(x) := \mu(R_n(x)), \\ \nu_n(x) := \nu(x). \end{cases}$$

The advantage of the μ_n and ν_n is that they brought us back to the finite situation while enabling to approximate $c(\mu, \nu)$:

Proposition 10 *We have $\lim_{n \rightarrow \infty} c(\mu_n, \nu_n) = c(\mu, \nu)$.*

Proof We first check that the limit exists. For $n \in \mathbb{N}_+$, consider the sigma-field \mathcal{F}_n generated by the partition (40). To each \mathcal{F}_n -measurable function f , associate the function f_n defined on V_n by

$$\forall x \in V_n, \quad f_n(x) := f(x).$$

This function determines f , since for any $x \in V_n$ and any $y \in R_n(x)$, $f(y) = f_n(x)$. Furthermore, we have:

$$\begin{aligned} \mu[f^2] &= \mu_n[f_n^2] \\ \sum_{x \in V} (f(p(x)) - f(x))^2 \nu(x) &= \sum_{x \in V_n} (f_n(p(x)) - f_n(x))^2 \nu_n(x). \end{aligned}$$

It follows that

$$c(\mu_n, \nu_n) = \sup_{f \in \mathcal{B}(\mathcal{F}_n) \setminus \{0\}} \frac{\mu[f^2]}{\sum_{x \in V} (f(p(x)) - f(x))^2 \nu(x)},$$

where $\mathcal{B}(\mathcal{F}_n)$ is the set of \mathcal{F}_n -measurable functions, which are necessarily bounded, i.e., belong to $\mathcal{B}(V)$. Since for any $n \in \mathbb{N}_+$ we have $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, we get that the sequence $(c(\mu_n, \nu_n))_{n \in \mathbb{N}_+}$ is non-decreasing and, taking into account Lemma 10, that

$$\lim_{n \rightarrow \infty} c(\mu_n, \nu_n) \leq c(\mu, \nu).$$

To get the reverse bound, first assume that $c(\mu, \nu) < +\infty$. For given $\epsilon > 0$, find a function $f \in \mathcal{B}(V)$ with

$$\frac{\mu[f^2]}{\sum_{x \in V} (f(p(x)) - f(x))^2 \nu(x)} \geq c(\mu, \nu) - \epsilon.$$

Consider π the normalization of μ into a probability measure and let f_n be the conditional expectation of f with respect to π and to the sigma-field \mathcal{F}_n . Note that the f_n are uniformly bounded by $\|f\|_\infty$. Thus by the bounded martingale

convergence theorem and since π gives a positive weight to any point of V , we have

$$\forall x \in V, \quad \lim_{n \rightarrow \infty} f_n(x) = f(x).$$

From Fatou's lemma, we deduce

$$\begin{aligned} \liminf_{n \rightarrow \infty} \sum_{x \in V_n} (f_n(p(x)) - f_n(x))^2 v_n(x) &= \liminf_{n \rightarrow \infty} \sum_{x \in V_n} (f_n(p(x)) - f_n(x))^2 v(x) \\ &\geq \sum_{x \in V} \liminf_{n \rightarrow \infty} [(f_n(p(x)) - f_n(x))^2 \mathbb{1}_{V_n}(x)] v(x) = \sum_{x \in V} (f(p(x)) - f(x))^2 v(x). \end{aligned}$$

By another application of the bounded martingale convergence theorem, we get

$$\lim_{n \rightarrow \infty} \mu_n[f_n^2] = \lim_{n \rightarrow \infty} \mu[f_n^2] = \mu[f^2],$$

so that

$$\limsup_{n \rightarrow \infty} \frac{\mu_n[f_n^2]}{\sum_{x \in V} (f_n(p(x)) - f_n(x))^2 v(x)} \geq \frac{\mu[f^2]}{\sum_{x \in V} (f(p(x)) - f(x))^2 v(x)}.$$

It follows that $\lim_{n \rightarrow \infty} c(\mu_n, v_n) \geq c(\mu, v) - \epsilon$, and since $\epsilon > 0$ can be chosen arbitrary small,

$$\lim_{n \rightarrow \infty} c(\mu_n, v_n) \geq c(\mu, v).$$

It remains to deal with the case where $c(\mu, v) = +\infty$. Then for any $M > 0$, we can find a function $f \in \mathcal{B}(V)$ with

$$\frac{\mu[f^2]}{\sum_{x \in V} (f(p(x)) - f(x))^2 v(x)} \geq M.$$

By the above arguments, we end up with $\lim_{n \rightarrow \infty} c(\mu_n, v_n) \geq M$, and since M can be arbitrary large, $\lim_{n \rightarrow \infty} c(\mu_n, v_n) = +\infty = c(\mu, v)$. □

Our next goal is to show the same result holds for $b(\mu, v)$. We need some additional notations. The integer $n \in \mathbb{N}_+$ being fixed, denote \mathbb{T}_n and \mathbb{S}_n the sets \mathbb{T} and \mathbb{S} associated to \mathcal{T}_n . The functional v_n is extended to \mathbb{T}_n via the iteration (37) understood in \mathcal{T}_n . To any $T \in \mathbb{T}_n$, associate T_n the minimal element of \mathbb{T} containing T . It is obtained in the following way: to any $x \in T$, if x has a child in T , then add all the children of x in V , and otherwise do not add any other points.

Lemma 11 *We have the comparisons*

$$v_n(T) \geq v(T_n) \text{ and } \mu_n(T^*) \leq \mu(T_n^*),$$

where T^* is understood in \mathcal{T}_n (and T_n^* in \mathcal{T}).

Proof The first bound is proven by iteration on the height of $T \in \mathbb{T}_n$.

- If this height is zero, then T is a singleton and T_n is the same singleton, so that $v_n(T) = v(T_n)$.
- If the height $h(T)$ of T is at least equal to 1, decompose

$$T = \{m_n(T)\} \sqcup \bigsqcup_{y \in C_n(m_n(T))} T_{n,y}$$

where $m_n(\cdot)$, $C_n(\cdot)$ and $T_{n,\cdot}$ are the notions corresponding to $m(\cdot)$, $C(\cdot)$ and T in \mathcal{T}_n .

Note that T and T_n have the same height and decompose

$$T_n = \{m(T_n)\} \sqcup \bigsqcup_{z \in C(m(T_n))} T_{n,z}.$$

On the one hand, we have $m(T_n) = m_n(T)$ and $C_n(m_n(T)) \subset C(m_n(T))$ and on the other hand, we have for any $y \in C_n(m_n(T))$, $v_n(T_y) \geq v((T_y)_n) = v(T_{n,y})$, due to the iteration assumption and to the fact that the common height of T_y and $(T_y)_n$ is at most equal to $h(T) - 1$. The equality $(T_y)_n = T_{n,y}$ is due to the fact that $T_{n,y}$ is obtained by the same completion of T_y as the one presented for T just above the statement of Lemma 11, and thus coincides with $(T_y)_n$. It follows that

$$\begin{aligned} \frac{1}{v_n(T)} &= \frac{1}{v_n(m_n(T))} + \frac{1}{\sum_{y \in C_n(m_n(T))} v_n(T_y)} \\ &= \frac{1}{v(m(T_n))} + \frac{1}{\sum_{y \in C_n(m_n(T))} v_n(T_y)} \leq \frac{1}{v(m(T_n))} + \frac{1}{\sum_{y \in C_n(m_n(T))} v(T_{n,y})} \\ &\leq \frac{1}{v(m(T_n))} + \frac{1}{\sum_{y \in C(m(T_n))} v(T_{n,y})} = \frac{1}{v(T_n)}, \end{aligned}$$

establishing the wanted bound $v_n(T) \geq v(T_n)$. We now come to the second bound of the above lemma. By definition, we have

$$T^* = \sqcup_{x \in L_n(T)} S_{n,y},$$

where $L_n(T)$ is the set of leaves of T in \mathcal{T}_n and $S_{n,y}$ is the subtree rooted in y in \mathcal{T}_n .

Note that $L_n(T) \subset L(T_n)$ and by definition of μ_n , we have

$$\forall y \in L_n(T), \quad \mu_n(S_{n,y}) = \mu(S_y).$$

It follows that

$$\mu_n(T^*) = \sum_{x \in L_n(T)} \mu_n(S_{n,y}) = \sum_{x \in L_n(T)} \mu(S_y) \leq \sum_{x \in L(T_n)} \mu(S_y) = \mu(T_n^*).$$

□

Let $\tilde{\mathbb{S}}_n$ be the image of \mathbb{S}_n under the mapping $\mathbb{S}_n \ni T \mapsto T_n \in \mathbb{S}$. Since $\mathbb{S}_n \ni T \mapsto T_n \in \tilde{\mathbb{S}}_n$ is a bijection, we get from Lemma 11,

$$b(\mu_n, \nu_n) := \max_{T \in \mathbb{S}_n} \frac{\mu_n(T^*)}{\nu_n(T)} \leq \max_{T_n \in \tilde{\mathbb{S}}_n} \frac{\mu(T_n^*)}{\nu(T_n)} \leq b(\mu, \nu),$$

so that

$$\limsup_{n \rightarrow \infty} b(\mu_n, \nu_n) \leq b(\mu, \nu). \tag{41}$$

Let us show more precisely:

Proposition 11 We have $\lim_{n \rightarrow \infty} b(\mu_n, \nu_n) = b(\mu, \nu)$.

Proof According to (41), it remains to show that

$$\liminf_{n \rightarrow \infty} b(\mu_n, \nu_n) \geq b(\mu, \nu). \tag{42}$$

Consider $T \in \mathbb{S}$ such that the ration $\mu(T^*)/\nu(T)$ serves to approximate $b(\mu, \nu)$, namely up to an arbitrary small $\epsilon > 0$ if $b(\mu, \nu) < +\infty$ or is an arbitrary large quantity if $b(\mu, \nu) = +\infty$. Define

$$\forall n \in \mathbb{N}_+, \quad T^{(n)} := T \cap V_n.$$

Arguing as at the end of the proof of Proposition 10, we will deduce (42) from

$$\lim_{n \rightarrow \infty} \frac{\mu_n((T^{(n)})^*)}{\nu_n(T^{(n)})} = \frac{\mu(T^*)}{\nu(T)},$$

where $(T^{(n)})^*$ is understood in \mathcal{T}_n . This convergence will be the consequence of

$$\lim_{n \rightarrow \infty} \mu_n((T^{(n)})^*) = \mu(T^*), \tag{43}$$

$$\lim_{n \rightarrow \infty} \nu_n(T^{(n)}) = \nu(T). \tag{44}$$

For (43), note that

$$\mu(T^*) = \sum_{x \in L(T)} \mu(S_y),$$

and as we have seen at the end of the proof of Lemma 11,

$$\mu(T^*) = \sum_{x \in L_n(T^{(n)})} \mu(S_y).$$

Thus (43) follows by dominated convergence (since $\mu(V) < +\infty$), from

$$\forall x \in T, \quad \lim_{n \rightarrow \infty} \mathbb{1}_{L_n(T^{(n)})}(x) = \mathbb{1}_{L(T)}(x).$$

To show the latter convergences, consider two cases:

- If $x \in L(T)$, then we will have $x \in L_n(T^{(n)})$ as soon as $x \in V_n$.
- If $x \in T \setminus L(T)$, then we will have $x \notin L_n(T^{(n)})$ as soon as V_n contains one of the children of x in T .

We now come to (44), and more generally let us prove by iteration over their height, that for any $\tilde{T} \in \mathbb{T}$ and $\tilde{T} \subset T$, we have

$$\lim_{n \rightarrow \infty} \uparrow v_n(\tilde{T} \cap V_n) = v(\tilde{T}), \tag{45}$$

i.e., the limit is non-decreasing. Indeed, if \tilde{T} has height 0, it is a singleton $\{x\}$, we have $v_n(\tilde{T} \cap V_n) = v(\tilde{T})$ as soon as x belongs to V_n , insuring (45).

Assume that \tilde{T} has height a $h \geq 1$ and that (45) holds for any \tilde{T} whose height is at most equal to $h - 1$. Write as usual

$$\frac{1}{v(\tilde{T})} = \frac{1}{v(m(\tilde{T}))} + \frac{1}{\sum_{y \in C(m(\tilde{T}))} v(\tilde{T}_y)}. \tag{46}$$

Assume that n is large enough so that $C(m(\tilde{T})) \cap V_n \neq \emptyset$ and in particular $m(\tilde{T}) \in V_n$ and $m_n(\tilde{T} \cap V_n) = m(\tilde{T})$. Thus we also have

$$\begin{aligned} \frac{1}{v_n(\tilde{T} \cap V_n)} &= \frac{1}{v_n(m_n(\tilde{T} \cap V_n))} + \frac{1}{\sum_{y \in C_n(m_n(\tilde{T} \cap V_n))} v_n((\tilde{T} \cap V_n)_y)} \\ &= \frac{1}{v(m(\tilde{T}))} + \frac{1}{\sum_{y \in C_n(m(\tilde{T}))} v_n(\tilde{T}_y \cap V_n)}. \end{aligned} \tag{47}$$

On the one hand, the set $C_n(m(\tilde{T}))$ is non-decreasing and its limit is $C(m(\tilde{T}))$, and on the other hand, due to the induction hypothesis, we have for any $y \in C(m(\tilde{T}))$,

$$\lim_{n \rightarrow \infty} \uparrow \nu_n(\tilde{T}_y \cap V_n) = \nu(\tilde{T}_y).$$

By monotone convergence, we get

$$\lim_{n \rightarrow \infty} \uparrow \sum_{y \in C_n(m(\tilde{T}))} \nu_n(\tilde{T}_y \cap V_n) = \sum_{y \in C(m(\tilde{T}))} \nu(\tilde{T}_y),$$

which leads to (45), via (46) and (47). This ends the proof of (42). □

The conjunction of Propositions 10 and 11 leads to the validity of (39), when V is denumerable with \mathcal{T} of finite height.

Let us now remove the assumption of finite height. The arguments are very similar to the previous one, except that the definition of $b(\mu, \nu)$ has to be modified (μ and ν are still positive measures on V , with μ of finite total mass). More precisely, for any $M \in \mathbb{N}_+$, consider $V_M := \{x \in V : h(x) \leq M\}$. Define on V_M the measure ν_M as the restriction to V_M of ν and μ_M via

$$\forall x \in V_M, \quad \mu_M(x) := \begin{cases} \mu(x) & \text{if } h(x) < M, \\ \mu(S_x) & \text{if } h(x) = M. \end{cases}$$

By definition, we take

$$b(\mu, \nu) := \lim_{M \rightarrow \infty} b(\mu_M, \nu_M).$$

This limit exists and the convergence is monotone, since we have for any $M \in \mathbb{N}_+$, $b(\mu_M, \nu_M) = \max_{T \in \mathbb{S}_M} \frac{\mu(T^*)}{\nu(T)}$, where $\mathbb{S}_M := \{T \in \mathbb{S} : T \subset V_M\}$. Note that a direct definition of $b(\mu, \nu)$ via the iteration (37) is not possible: we could not start from leaves that are singletons.

By definition, $c(\mu, \nu)$ is the best constant in (36). It also satisfies $c(\mu, \nu) := \lim_{M \rightarrow \infty} c(\mu_M, \nu_M)$, as can be seen by adapting the proof of Proposition 10. We conclude that (39) holds by passing at the limit in

$$\forall M \in \mathbb{N}_+, \quad b(\mu_M, \nu_M) \leq c(\mu_M, \nu_M) \leq 16 b(\mu_M, \nu_M).$$

References

1. A. Bendikov, L. Saloff-Coste, Random walks on some countable groups, in *Groups, Graphs and Random Walks* (2011), pp. 77–103
2. I. Benjamini, O. Schramm, Every graph with a positive Cheeger constant contains a tree with a positive Cheeger constant. *Geom. Funct. Anal.* **7**(3), 403–419 (1997)

3. N. Berestycki, E. Lubetzky, Y. Peres, A. Sly, Random walks on the random graph. *Ann. Probab.* **46**(1), 456–490 (2018)
4. C. Bordenave, P. Caputo, D. Chafaï, Spectrum of large random reversible Markov chains: heavy-tailed weights on the complete graph. *Ann. Probab.* **39**(4), 1544–1590 (2011)
5. C. Bordenave, P. Caputo, J. Salez, Random walk on sparse random digraphs. *Probab. Theory Related Fields* **170**(3–4), 933–960 (2018)
6. P.J. Cameron, The random graph, in *The Mathematics of Paul Erdős II* (1997), 333–351
7. P.J. Cameron, The random graph revisited, in *European Congress of Mathematics* (Birkhäuser, Basel, 2001), pp. 267–274
8. P. Diaconis, M. Malliaris, Complexity and randomness in the Heisenberg groups (and beyond) (2021). arXiv:2107.02923v2
9. P. Diaconis, P.M. Wood, Random doubly stochastic tridiagonal matrices. *Random Struct. Algor.* **42**(4), 403–437 (2013)
10. W.D. Evans, D.J. Harris, L. Pick, Weighted Hardy and Poincaré inequalities on trees. *J. Lond. Math. Soc.* (2) **52**(1), 121–136 (1995)
11. N. Fountoulakis, B.A. Reed, The evolution of the mixing rate of a simple random walk on the giant component of a random graph. *Random Struct. Algor.* **33**(1), 68–86 (2008)
12. A. Georgakopoulos, J. Haslegrave, Percolation on an infinitely generated group. *Combin. Probab. Comput.* **29**(4), 587–615 (2020)
13. G.H. Hardy, J.E. Littlewood, G. Pólya, *Inequalities*, 2nd edn. (Cambridge University Press, Cambridge, 1952)
14. M. Hildebrand, A survey of results on random random walks on finite groups. *Probab. Surv.* **2**, 33–63 (2005)
15. D. Hunter, An upper bound for the probability of a union. *J. Appl. Probab.* **13**, 597–603 (1976)
16. J. Jorgenson, S. Lang, The ubiquitous heat kernel, in *Mathematics Unlimited—2001 and Beyond* (Springer, Berlin, Heidelberg, 2001), pp. 655–683
17. C.A.J. Klaassen, J.A. Wellner, Hardy’s inequality and its descendants: a probability approach. *Electron. J. Probab.* **26**, 1–34 (2021)
18. E.G. Kounias, Bounds for the probability of a union, with applications. *Ann. Math. Stat.* **39**, 2154–2158 (1968)
19. D.A. Levin, Y. Peres, *Markov Chains and Mixing Times* (American Mathematical Society, Providence, 2017)
20. S. Martineau, Locally infinite graphs and symmetries. *Grad. J. Math.* **2**(2), 42–50 (2017)
21. L. Miclo, Relations entre isopérimétrie et trou spectral pour les chaînes de Markov finies. *Probab. Theory Related Fields* **114**(4), 431–485 (1999)
22. A. Nachmias, Y. Peres, Critical random graphs: diameter and mixing time. *Ann. Probab.* **36**(4), 1267–1286 (2008)
23. L. Saloff-Coste, Lectures on finite Markov chains, in *Lectures on Probability Theory and Statistics (Saint-Flour; 1996)*. Lecture Notes in Mathematics, vol. 1665 (Springer, Berlin, 1997), pp. 301–413
24. Wikipedia contributors. Pairwise independence - Wikipedia, the free encyclopedia (2021). https://en.wikipedia.org/wiki/Pairwise_independence [Online: Accessed 18 September 2021]
25. W. Woess, *Random Walks on Infinite Graphs and Groups* (Cambridge University Press, Cambridge, 2000)
26. K.J. Worsley, An improved Bonferroni inequality and applications. *Biometrika* **69**, 297–302 (1982)