J. Eur. Math. Soc. 28, 1913–1977 (2026) DOI 10.4171/JEMS/1743

Jonathan Hermon · Sam Olesker-Taylor

Cutoff for almost all random walks on Abelian groups

Received 5 February 2021; revised 13 May 2025

Abstract. Consider the random Cayley graph of a finite group G with respect to k generators chosen uniformly at random, with $1 \ll \log k \ll \log|G|$; denote it by G_k. A conjecture of Aldous and Diaconis (1985) asserts, for $k \gg \log|G|$, that the random walk on this graph exhibits cutoff. Further, the cutoff time should be a function only of k and $|G|$, to sub-leading order. This was verified for all Abelian groups in the '90s. We extend the conjecture to $1 \ll k \lesssim \log|G|$. We establish cutoff for *all* Abelian groups under the condition $k - d(G) \gg 1$, where $d(G)$ is the minimal size of a generating subset of G, which is almost optimal. The cutoff time is described (abstractly) in terms of the entropy of random walk on $\mathbb{Z}^k$. This abstract definition allows us to deduce that the cutoff time can be written as a function only of k and $|G|$ when $d(G) \ll \log|G|$ and $k - d(G) \asymp k \gg 1$; this is not the case when $d(G) \asymp \log|G| \asymp k$. For certain regimes of k, we find the limit profile of the convergence to equilibrium. Wilson (1997) conjectured that $\mathbb{Z}_2^d$ gives rise to the slowest mixing time for G_k amongst all groups of size at most 2^d. We give a partial answer, verifying the conjecture for nilpotent groups. This is obtained via a comparison result of independent interest between the mixing times of nilpotent G and a corresponding Abelian group $\bar{G}$, namely the direct sum of the Abelian quotients in the lower central series of G. We use this to refine a celebrated result of Alon and Roichman 1994: we show for nilpotent G that G_k is an expander provided $k - d(\bar{G}) \gtrsim \log|G|$. As another consequence, we establish cutoff for nilpotent groups with relatively small commutator subgroup, including high-dimensional special groups, such as Heisenberg groups. The aforementioned results all hold with high probability over the random Cayley graph G_k.

Keywords: cutoff, mixing times, random walk, random Cayley graphs, entropy.

1. Introduction and statement of results

1.1. Motivation, brief overview of results and notation

1.1.1. Motivating conjectures of Aldous, Diaconis and Wilson. We analyse properties of the random walk (abbreviated *RW*) on a *Cayley graph* of a finite group. The generators

Jonathan Hermon: Department of Mathematics, University of British Columbia, 1984 Mathematics Road, Vancouver, VC V6T 1Z2, Canada; jhermon@math.ubc.ca

Sam Olesker-Taylor: Department of Statistics, University of Warwick, Coventry CV4 7AL, UK; sam.olesker-taylor@warwick.ac.uk

Mathematics Subject Classification 2020: 05C80 (primary); 05C81, 20D15, 60B15, 60J27, 60K37 (secondary).

of this graph are chosen independently and uniformly at random. Precise definitions are given in Section 1.4.1. For now, let G be a finite group, let k be an integer (allowed to depend on G) and denote by G_k the Cayley graph of G with respect to k independently and uniformly random generators. We consider values of k with $1 \ll \log k \ll \log|G|$ for which G_k is connected with high probability (abbreviated *w.h.p.*), i.e., with probability tending to 1 as $|G|$ grows.

Since pioneering work of Erdős, it has been understood that the typical behaviour of *random* objects in some class can shed valuable light on the class as a whole. Thus, when considering some class of combinatorial objects, it is natural to ask questions such as the following:

- What does a typical object in this class 'look like'?
- If an object is chosen uniformly at random, which properties hold with high probability?

Aldous and Diaconis [1] applied this philosophy to the study of random walks on groups.

Aldous and Diaconis [1, 2] coined the phrase *cutoff phenomenon*: this occurs when the total-variation distance between the law of the RW and its invariant distribution drops abruptly from close to 1 to close to 0 in a time-interval of smaller order than the mixing time. The material in this article is motivated by a conjecture of theirs regarding 'universality of cutoff' for the RW on the random Cayley graph G_k. It is given in [1, p. 40], which is an extended version of [2].

Conjecture (Aldous and Diaconis, 1985). *For any group G, if $k \gg \log|G|$ and $\log k \ll \log|G|$, then the random walk on G_k exhibits cutoff w.h.p. Further, the cutoff time, to leading order, is independent of the algebraic structure of the group: it can be written as a function only of k and $|G|$.*

This conjecture spawned a large body of work, including [18, 19, 30, 32, 33, 44, 48]; see Section 1.3. It has been established in the Abelian set-up by Dou and Hildebrand [18, 30]; see Section 1.3.1, and Theorem 7.2 where we give a short proof. Save [33] which considers the cyclic group $\mathbb{Z}_p$ for prime p and [48] which considers $\mathbb{Z}_2^d$ (which enforces $k \geq d = \log_2|G|$), focus has been on $k \gg \log|G|$.

We establish cutoff for all Abelian groups when $1 \ll k \lesssim \log|G|$ under almost optimal conditions in terms of group-generation. We also give simple conditions under which the cutoff time is independent of the algebraic structure of the group.

The second part of this article is motivated by a conjecture of Wilson. Wilson [48] established cutoff for the RW on G_k when $G = \mathbb{Z}_2^d$ and then conjectured that $\mathbb{Z}_2^d$ is the slowest amongst all groups of size at most 2^d, asymptotically as $d \to \infty$; see [48, Theorem 1 and Conjecture 7].

Conjecture (Wilson, 1997). *For all diverging d and n with $n \leq 2^d$ and all groups G of size n, if $k - \log_2 n \gg 1$ and $\log k \ll \log n$, then $t_{\mathrm{mix}}(\varepsilon, G_k)/t_{\mathrm{mix}}(\varepsilon', H_k) \leq 1 + o(1)$ w.h.p. for all $\varepsilon, \varepsilon' \in (0, 1)$, where $H := \mathbb{Z}_2^d$, i.e., the mixing time for G_k is at most that of H_k w.h.p. up to smaller order terms.*

We establish a comparison between the mixing times for nilpotent and Abelian groups, of which Wilson's conjecture in the nilpotent set-up is an immediate consequence. Additionally, we apply our nilpotent–Abelian comparison theorem to establish cutoff for various examples of non-Abelian groups, including p-groups with 'small' commutator subgroup and Heisenberg groups.

1.1.2. Brief overview of results. Our focus is on mixing properties of the RW on the random Cayley graph G_k. We consider the limit as $n := |G| \to \infty$ under the assumption that $1 \ll \log k \ll \log|G|$. The condition $1 \ll \log k \ll \log|G|$ is necessary for cutoff on G_k for all nilpotent G; see Remark A.5 below.

We establish cutoff when G is any Abelian group, requiring only $k - d(G) \gg 1$, where $d(G)$ is the minimal size of a generating subset of G. We show that the leading order term in the cutoff time is independent of the algebraic structure of G when $d(G) \ll \log|G|$ and $k - d(G) \asymp k$, i.e., it depends only on k and $|G|$. It is the time at which the entropy of RW on $\mathbb{Z}^k$ is $\log|G|$. This extends the Aldous–Diaconis conjecture to $1 \ll k \lesssim \log|G|$. For certain k, we find the limit profile of the convergence to equilibrium.

We deduce Wilson's conjecture in the Abelian set-up, as a consequence of our cutoff results. We then extend this to the nilpotent set-up via the following result, which is of independent interest: to a nilpotent group G, we associate an Abelian group $\overline{G}$ of the same size, which is the direct sum of the Abelian quotients in the lower central series of G, and show that $t_{\text{mix}}(G_k)/t_{\text{mix}}(\overline{G}_k) \le 1 + o(1)$ w.h.p. (provided $k - d(\overline{G}) \gg 1$).

We give examples where this comparison is tight: we establish cutoff w.h.p. for the RW on G_k where G is a nilpotent group with a relatively small commutator subgroup. Examples of such groups include high-dimensional extra special or Heisenberg groups.

Lastly, we show that the random Cayley graph of a nilpotent group G is an *expander* w.h.p. whenever $k \gtrsim \log|G|$ and $k - d(\overline{G}) \asymp k$. (If G is Abelian, then $G = \overline{G}$.)

Introduced by Aldous and Diaconis [1], there has been a great deal of research into these random Cayley graphs. Motivation for this model and an overview of historical work are given in Section 1.3.

1.1.3. Notation and terminology. Cayley graphs can be either directed or undirected; we emphasise this by writing G_k^+ and G_k^-, respectively. When we write G_k or $G_k^\pm$, this means 'either G_k^- or G_k^+', corresponding to the undirected, respectively directed, graphs with generators chosen independently and uniformly at random.

Conditional on being simple, G_k^+ is uniformly distributed over the set of all simple degree-k Cayley graphs. Up to a slightly adjusted definition of *simple* for undirected Cayley graphs, our results hold with G_k replaced by a uniformly chosen simple Cayley graph of degree k; see Section 1.4.2.

Our results are for sequences $(G_N)_{N\in\mathbb{N}}$ of finite groups with $|G_N| \to \infty$ as $N \to \infty$. For ease of presentation, we write statements like 'let G be a group' instead of 'let $(G_N)_{N\in\mathbb{N}}$ be a sequence of groups'. Likewise, the quantities $d(G)$ and, of course, k appearing in the statements below all correspond to sequences, which need not be fixed (or bounded) unless we explicitly say so. In the same vein, an event holds *with high probability* (abbreviated *w.h.p.*) if its probability tends to 1.

We use standard asymptotic notation: '$\ll$' or '$o(\cdot)$' means 'of smaller order'; '$\lesssim$' or '$\mathcal{O}(\cdot)$' means 'of order at most'; '$\asymp$' means 'of the same order'; '$\eqsim$' means 'asymptotically equivalent'.

1.2. *Statements of main results*

We analyse mixing in the *total-variation* (abbreviated *TV*) distance. The uniform distribution on G, denoted π_G, is invariant for the RW. Let $S = (S(t))_{t \ge 0}$ denote the RW on G_k; its law is denoted by $\mathbb{P}_{G_k}(S(t) \in \cdot)$. For $t \ge 0$, denote the TV distance between the law of $S(t)$ and π_G by

$$d_{G_k}(t) := \|\mathbb{P}_{G_k}(S(t) \in \cdot) - \pi_G\| := \max_{A \subseteq G} \left| \mathbb{P}_{G_k}(S(t) \in A) - \frac{|A|}{|G|} \right|.$$

Throughout, unless explicitly specified otherwise, we use continuous time: $t \ge 0$ means $t \in [0, \infty)$.

1.2.1. Cutoff for all Abelian groups. We use standard notation and definitions for *mixing* and *cutoff*; see, e.g., [35, Sections 4 and 18].

Definition. A sequence $(X^N)_{N \in \mathbb{N}}$ of Markov chains is said to exhibit *cutoff* when, in a short time-interval, known as the *cutoff window*, the TV distance of the distribution of the chain from equilibrium drops from close to 1 to close to 0, or more precisely if there exists $(t_N)_{N \in \mathbb{N}}$ with

$$\limsup_{N \to \infty} d_N(t_N(1 - \varepsilon)) = 1 \quad \text{and} \quad \limsup_{N \to \infty} d_N(t_N(1 + \varepsilon)) = 0 \quad \text{for all } \varepsilon \in (0, 1),$$

where $d_N(\cdot)$ is the TV distance of $X^N(\cdot)$ from its equilibrium distribution for each $N \in \mathbb{N}$.

We say that a RW on a sequence of random graphs $(H_N)_{N \in \mathbb{N}}$ *exhibits cutoff around time* $(t_N)_{N \in \mathbb{N}}$ *w.h.p.* if, for all fixed ε, in the limit $N \to \infty$, the TV distance at time $(1 + \varepsilon)t_N$ converges in distribution to 0 and at time $(1 - \varepsilon)t_N$ to 1, where the randomness is over the random graph H_N.

To extend the Aldous–Diaconis conjecture to $1 \ll k \lesssim \log|G|$, one needs additional assumptions. For an Abelian group G, write $d(G)$ for the minimal size of a generating set of G. If $k < d(G)$, then the group cannot be generated by any choice of generators. Pomerance [43] shows that the expected number of independent, uniform generators required to generate the group is at most $d(G) + 3$. (That is, if $Z_1, Z_2, \ldots \sim^{\text{iid}} \text{Unif}(G)$ and $\kappa \in \mathbb{N}$ is minimal with $\langle Z_1, \ldots, Z_\kappa \rangle = G$, then $d(G) \le \mathbb{E}(\kappa) \le d(G) + 3$.) Thus $k - d(G) \gg 1$ is always sufficient for G to be generated by $\{Z_1^\pm, \ldots, Z_k^\pm\}$ w.h.p. (by Markov's inequality); we assume this throughout. In many cases, $k - d(G) \gg 1$ is necessary to generate the group w.h.p., and so this assumption cannot be removed.

For a characterisation of these cases and related discussion, see [25, Lemma 8.1]. The condition $k - d(G) \asymp k$ is particularly relevant for the Aldous–Diaconis conjecture; see Remark A.1.

We use an *entropic method*, which involves defining *entropic times*; see Section 1.3.5 for a high-level description of the method and Section 2.1 for the specific application. The main idea is to use an auxiliary process W to generate the walk S; one then studies the entropy of the process W. Write $Z = [Z_1, \dots, Z_k]$ for the multiset of generators of the Cayley graph; then G_k corresponds to choosing $Z_1, \dots, Z_k \sim^{\text{iid}} \text{Unif}(G)$. Here, $W_i(t)$ is, for each i, the number of times generator Z_i has been applied minus the number of times Z_i^{-1} has been applied; W is a rate-1 RW on $\mathbb{Z}^k$. Then, $S(t) = W(t) \cdot Z$ when the group is Abelian. (This auxiliary process W is key even when studying nilpotent groups.)

For undirected graphs, W is the usual simple RW (abbreviated *SRW*): a coordinate is selected uniformly at random and incremented/decremented by 1 each with probability $1/2$. For directed graphs, inverses are never applied, so a step of W is as follows: a coordinate is selected uniformly at random and incremented by 1; we term this the *directed* RW (abbreviated *DRW*).

The *entropic times* are defined in terms of the entropy of this auxiliary RW W.

Definition A. For $\gamma \in \mathbb{N} \cup \{\infty\}$, let $\tau_\gamma^\pm := \tau_\gamma^\pm(k, G)$ be the time at which the entropy of rate-1 RW (i.e., SRW or DRW, as appropriate) on $\mathbb{Z}_\gamma^k$ is $\log|G/\gamma G|$, where $\gamma G := \{\gamma g \mid g \in G\}$; we use the convention $\mathbb{Z}_\infty := \mathbb{Z}$ and $\infty G := |G| G = \{\text{id}\}$. Set $\tau_*^\pm := \tau_*^\pm(k, G) := \max_{\gamma \in \mathbb{N} \cup \{\infty\}} \tau_\gamma^\pm(k, G)$.

We establish cutoff for all Abelian groups, under almost optimal conditions on k in terms of G. This gives an affirmative answer for Abelian groups in a strong sense to the primary part of the conjecture (occurrence of cutoff) of Aldous and Diaconis [1, 2] as well as the informal question asked by Dou [13]; we discuss the secondary part (time depending only on k and $|G|$) in Remark A.1.

Cutoff has already been established for Abelian groups when $k \gg \log|G|$ with $\log k \ll \log|G|$, as mentioned above; see Section 1.3.1. We thus restrict our statements to $1 \ll k \lesssim \log|G|$. For $1 \ll k \lesssim \log|G|$, only two groups had been considered previously: $\mathbb{Z}_2^d$ in [48] and $\mathbb{Z}_p$ with p prime in [33]. Recall that $1 \ll \log k \ll \log|G|$ is necessary for cutoff for nilpotent G, e.g., Abelian G; see Remark A.5. More refined statements are given in Theorems 2.5, 3.7, and 4.2.

Theorem A. *Let G be an Abelian group and k an integer with $1 \ll k \lesssim \log|G|$. Suppose that $k - d(G) \gg 1$. Then, the RW on $G_k^\pm$ exhibits cutoff at time $\tau_*^\pm(k, G)$ w.h.p. Further, if $k - d(G) \asymp k$ and $d(G) \ll \log|G|$, then $\tau_*(k, G) \eqsim \tau_\infty(k, |G|)$, which depends only on k and $|G|$.*

Moreover, the following asymptotic relations regarding the entropic times hold:

- *If $k \ll \log|G|$, then $\tau_\infty(k, |G|) \eqsim k|G|^{2/k}/(2\pi e)$.*
- *If $k - d(G) \asymp k \asymp \log|G|$, then $\tau_*(k, G) \asymp k|G|^{2/k} \asymp k$.*
- *If $k > d(G)$, then $k|G|^{2/k} \lesssim \tau_*(k, G) \lesssim k|G|^{2/k} \log k$.*

We now give some remarks on this theorem. Further remarks are deferred to Section 1.2.4.

Remark A.1. Theorem A establishes cutoff for all Abelian groups, under the mild (almost necessary) condition $k - d(G) \gg 1$, verifying the primary part of the Aldous–Diaconis conjecture. Further, the secondary part is partially verified too: the cutoff time depends only on k and $|G|$, up to smaller order terms, when $k - d(G) \asymp k$ and $d(G) \ll \log|G|$. Cases with $k - d(G) \ll k$ or $d(G) \asymp \log|G|$ need not satisfy this, however. For example, if $k \eqsim 2\log(4^r)$, then the groups $\mathbb{Z}_2^{2r}$ and $\mathbb{Z}_4^r$ give rise to mixing times which differ by a constant factor; see also [48]. For a counterexample with $1 \ll k \ll \log|G|$, see [25, Proposition 3.2 and Theorem 3.4], where $\mathbb{Z}_p^d$, with p prime, is studied.

Remark A.2. For certain regimes of k, we find the limit profile of the convergence to equilibrium: we define entropic times τ_α and show that $d_{G_k}(\tau_\alpha) \to^{\mathbb{P}} \Psi(\alpha)$, where Ψ is the standard Gaussian tail; see Definition 2.1, Proposition 2.2 and Theorem 2.5. This holds for any Abelian group if, for example, $k - d(G) \asymp k$ and $1 \ll k \ll \log|G| / \log\log\log|G|$ or $k - d(G) \gg 1$ and $1 \ll k \ll \sqrt{\log|G| / \log\log\log|G|}$. The result holds for any $1 \ll k \ll \log|G|$ under some constraints on the group. In [25, Theorem A], we show the same for $k \asymp \log|G|$, again with some constraints on G.

Remark A.3. From the abstract entropic definition, $\mathbb{Z}_2^d$ is the slowest amongst Abelian groups:

$$\max\{\tau_*(k, G) \mid G \text{ Abelian group with } |G| \le 2^d\} = \tau_*(k, \mathbb{Z}_2^d).$$

This verifies Wilson's conjecture in the Abelian set-up; the general nilpotent set-up comes later.

Remark A.4. The entropic time $\tau_\infty(k, G)$ arises naturally; see Section 2.5 for an outline. In essence, we want

$$\mathcal{W}_t := \left\{ w \in \mathbb{Z}^k \mid \mathbb{P}(W(t) = w) \ll \frac{1}{|G|} \right\} = \{ w \in \mathbb{Z}^k \mid -\log \mathbb{P}(W(t) = w) - \log|G| \gg 1 \}$$

to satisfy $\mathbb{P}(W(t) \in \mathcal{W}_t) = 1 - o(1)$. We thus want the entropy of $W(t)$ to be at least $\log|G|$.

The emergence of the entropic times τ_L ($L \ne \infty$) is more delicate. We outline this in Section 3.5.

Cutoff in L_2, instead of TV (i.e., L_1), can also be analysed. For time $t \ge 0$, define

$$d_{G_k}^{(2)}(t) := \|\mathbb{P}_{G_k}(S(t) \in \cdot) - \pi_G\|_{2,\pi_G} := \Big(|G|^{-1} \sum_{g \in G} (|G|\mathbb{P}_{G_k}(S(t) = g) - 1)^2 \Big)^{1/2}.$$

Mixing and cutoff can then be defined with respect to L_2 analogously to TV (L_1) distance.

It turns out that L_2 mixing time may be a constant, or even more, larger than the TV. Similar considerations to those in Remark A.4 suggest that for the L_2 mixing the key condition is $\mathbb{P}(W(2t) = 0) \ll 1/|G|$. This leads us to a conjecture for the L_2 mixing time, which we state informally now. We elaborate briefly on where the proof would differ, compared with TV, in Section 7.3.

Conjecture A. *For $L \in \mathbb{Z} \cup \{\infty\}$, let $\tilde{\tau}_L^{\pm} := \tilde{\tau}_L^{\pm}(k, G)$ be the time t at which the return probability for RW on $\mathbb{Z}_L^k$ at time $2t$ is $|G/LG|^{-1}$. Set $\tilde{\tau}_*^{\pm}(k, G) := \max_{L \in \mathbb{N}} \tilde{\tau}_L^{\pm}(k, G)$. Then, under similar conditions to those of Theorem A, w.h.p., the RW on G_k exhibits cutoff in the L_2 metric at time $\tilde{\tau}_*^{\pm}(k, G)$.*

We also consider cutoff in *separation distance*. For time $t \geq 0$, define

$$s_{G_k}(t) := \max_{g \in G} \{1 - |G| \mathbb{P}_{G_k}(S(t) = g)\}.$$

Mixing and cutoff can then be defined with respect to separation distance analogously to TV.

It is standard that, under reversibility, the TV and separation mixing times differ by up to a factor 2; see, e.g., [35, Lemmas 6.16 and 6.17]. However, Hermon, Lacoin and Peres [23, Theorem 1.1] showed that TV and separation cutoff are not equivalent, and that neither one implies the other.

We show that separation cutoff occurs w.h.p. in a certain regime and, moreover, that the cutoff time is the same, up to subleading order terms, as for TV.

A more refined statement is given in Theorem 5.2.

Theorem B. *Let G be an Abelian group and k an integer. Suppose that $1 \ll \log k \ll \log|G|$ and $k - d(G) \gg \max\{((\log|G|)/k)^2, (\log|G|)^{1/2}\}$. Then, the RW on G_k exhibits cutoff in separation distance at time $\tau_*(k, G)$ w.h.p.*

Remark B. The conditions hold if $k \gtrsim (\log|G|)^{3/4}$, $\log k \ll \log|G|$ and $k - d(G) \gg (\log|G|)^{1/2}$. Analogously to Remark A.3, the slowest amongst Abelian groups for separation mixing is $\mathbb{Z}_2^d$.

1.2.2. Comparison of mixing times between different groups. The previous results concerned cutoff. The next results are of a slightly different flavour. They concern *nilpotent* groups: these are groups G whose *lower central series*, i.e., the sequence $(G_{(\ell)})_{\ell \geq 0}$ defined by $G_{(0)} := G$ and $G_{(\ell)} := [G_{(\ell-1)}, G]$ for $\ell \geq 1$, stabilises at the trivial group. The results compare the mixing times between different groups; these mixing times are random.

Definition. For $\varepsilon \in (0, 1)$ and a Cayley graph H, write

$$t_{\text{mix}}(\varepsilon, H) := \inf\{t \geq 0 \mid d_H(t) \leq \varepsilon\}.$$

For two sequences $H := (H_N)_{N \in \mathbb{N}}$ and $H' := (H'_N)_{N \in \mathbb{N}}$ of random Cayley graphs, say that $t_{\text{mix}}(H)/t_{\text{mix}}(H') \leq 1 + o(1)$ w.h.p. if there exist non-random sequences $(\gamma_N)_{N \in \mathbb{N}}$ and $(\delta_N)_{N \in \mathbb{N}}$ with $\lim_N \delta_N = 0$ such that, for all $\varepsilon, \varepsilon' \in (0, 1)$, we have

$$\lim_{N \to \infty} \mathbb{P}(t_{\text{mix}}(\varepsilon, H_N) \leq (1 + \delta_N)\gamma_N) = 1 = \lim_{N \to \infty} \mathbb{P}((1 - \delta_N)\gamma_N \leq t_{\text{mix}}(\varepsilon', H'_N)).$$

We establish Wilson's conjecture in the nilpotent set-up, as the following theorem describes.

Theorem C. *For all diverging d and n with $n \le 2^d$ and all nilpotent groups G of size n, if $k - \log_2 n \gg 1$ and $\log k \ll \log n$, then $t_{\mathrm{mix}}(G_k)/t_{\mathrm{mix}}(H_k) \le 1 + o(1)$ w.h.p., where $H := \mathbb{Z}_2^d$.*

As noted in Remark A.3, for Abelian groups this follows from our cutoff result and the abstract entropic definition of the cutoff time $\tau_*(k, G)$ for Abelian G. The extension to nilpotent groups is then established by Theorem D below, which is of independent interest. It is quite significantly stronger than Wilson's conjecture in the nilpotent set-up. We can use it to establish cutoff for a class of nilpotent groups with 'small commutator subgroup'; see Corollaries D.1–D.3.

Theorem D. *Let G be a nilpotent group. Set $\bar{G} := \bigoplus_1^L (G_{(\ell-1)}/G_{(\ell)})$, where $(G_{(\ell)})_{\ell \ge 0}$ is the lower central series of G and $L := \min\{\ell \ge 0 \mid G_{(\ell)} = \{\mathsf{id}\}\}$. Suppose that $1 \ll \log k \ll \log|G|$ and $k - d(\bar{G}) \gg 1$. Then, $t_{\mathrm{mix}}(G_k)/t_{\mathrm{mix}}(\bar{G}_k) \le 1 + o(1)$ w.h.p.*

The quotients $G_{(\ell-1)}/G_{(\ell)} = G_{(\ell-1)}/[G_{(\ell-1)}, G]$ are all Abelian, by definition of the commutator. The result says that the mixing time of G_k is at least as fast as its Abelian counterpart $\bar{G}_k$. For a group G, denote its *commutator* subgroup $G^{\mathrm{com}} := [G, G]$ and its *Abelianisation* $G^{\mathrm{ab}} := G/G^{\mathrm{com}}$.

Corollary D.1. *Let G be a finite, non-Abelian, nilpotent group and k such that $1 \ll \log k \ll \log|G|$.*

- *If $k \lesssim \log|G|$, then suppose that $k \gg d(\overline{[G,G]}) \log|[G,G]|$ and $k - d(G^{\mathrm{ab}}) \gg d(\overline{[G,G]})$.*
- *If $k \gg \log|G|$, then suppose only that $\log|[G,G]| \ll \log|G|$.*

Then, the RW on G_k exhibits cutoff at $\tau_*(k, G^{\mathrm{ab}})$ w.h.p.

For step-2 nilpotent groups, $[G,G]$ is Abelian and hence $\overline{[G,G]} = [G,G]$. The above corollary is thus particularly applicable for these groups. A prime example of such groups is *special groups* with small commutator subgroup. For a prime p, a p-group is *special* if it is step-2 and its centre $Z(G)$, Frattini subgroup $\Phi(G)$ and commutator subgroup $[G,G]$ are all equal and elementary Abelian (i.e., isomorphic to $\mathbb{Z}_p^s$ for some s). In this case, also $G^{\mathrm{ab}} \cong \mathbb{Z}_p^r$, where $r := \ell - s$ and $\ell := \log_p|G|$.

We can relax the conditions on k using this particular form of the Abelianisation and commutator subgroup. The time at which the entropy of RW on $\mathbb{Z}_p^k$ reaches $\log(p^r) = \log|\mathbb{Z}_p^r|$ is $\tau_p(k, \mathbb{Z}_p^r)$.

Corollary D.2. *Let p be prime, G be a non-Abelian, special p-group and k be such that $1 \ll \log k \ll \log|G|$. Let $r := \log_p|G^{\mathrm{ab}}|$, $s := \log_p|G^{\mathrm{com}}|$ and $\ell := r + s = \log_p|G|$. Suppose that $k \ge \ell$.*

- *If $k \lesssim \log|G|$, then suppose that $k \gg s^2 \log p$ and $k - r \gg s$.*
- *If $k \gg \log|G|$, then suppose only that $s \ll r$.*

Then, the RW on G_k exhibits cutoff at $\tau_*(k, G^{\mathrm{ab}}) = \tau_p(k, \mathbb{Z}_p^r)$ w.h.p. conditional that G_k is connected. If $(k - r)p \gg 1$, then G_k is connected w.h.p. If $k - r \asymp k$ and $p \gg 1$, then $\tau_p(k, \mathbb{Z}_p^r) \eqsim \tau_\infty(k, \mathbb{Z}_p^r)$.

Special groups are ubiquitous amongst p-groups of a given size in a precise, quantitative sense. Hence, Corollary D.2 is applicable to many groups. See Remark 6.12 for a precise statement as well as some asymptotic expressions. Sims [46] gives, for given (p, ℓ, s), a simple, explicit description of all special groups of size p^ℓ whose commutator subgroup is of size p^s.

Extra special groups satisfy $G^{\mathrm{com}} \cong \mathbb{Z}_p$ (so $d(G^{\mathrm{com}}) = 1$) and $|G| = p^{2d-3}$ for some integer $d \geq 3$. For given d and $p \neq 2$, up to isomorphism there are only two extra special groups. One of these is the *Heisenberg group*, which can be defined for p not prime also. For (not necessarily prime) $m, d \in \mathbb{N}$, the Heisenberg group $H_{m,d}$ is the set triples $(x, y, z) \in \mathbb{Z}_m^{d-2} \times \mathbb{Z}_m^{d-2} \times \mathbb{Z}_m$ with

$$(x, y, z) \circ (x', y', z') := (x + x', y + y', z + z' + x \cdot y'),$$

where $x \cdot y'$ is the usual dot product for vectors in $\mathbb{Z}_m^{d-2}$. We have

$$H_{m,d}^{\mathrm{ab}} \cong \mathbb{Z}_m^{2d-4} \quad \text{and} \quad H_{m,d}^{\mathrm{com}} \cong \mathbb{Z}_m.$$

For p prime, $H_{p,d}$ with $d \gg 1$ falls into the class analysed in Corollary D.2 with $r = 2d - 4$ and $s = 1$. The following corollary thus focuses on $H_{m,d}$ with m not (necessarily) prime. Note that $\tau_\infty(k, \mathbb{Z}_m^r)$ is the time at which the entropy of RW on $\mathbb{Z}^k = \mathbb{Z}_\infty^k$ reaches $\log(m^r) = \log|\mathbb{Z}_m^r|$.

Corollary D.3. *Let $m, d \in \mathbb{N}$ with $d \gg 1$. Suppose that $k - 2d \gg 1$, $k \gg \log m$ and $\log k \ll d \log m \asymp \log|H_{m,d}|$. Then, w.h.p., the RW on $(H_{m,d})_k$ exhibits cutoff at $\tau_*(k, H_{m,d}^{\mathrm{ab}} \cong \mathbb{Z}_m^{2d-4})$. If additionally $k - 2d \asymp k$ and $m \gg 1$, then $\tau_*(k, \mathbb{Z}_m^{2d-4}) \eqsim \tau_\infty(k, \mathbb{Z}_m^{2d-4})$.*

If m is fixed (and thus $d \gg 1$), then the condition $k \gg \log m$ is absorbed into $k \gg 1$. Thus, this corollary handles arbitrary Heisenberg groups $H_{m,d}$ with m fixed and $k - 2d \gg 1$.

We now give some remarks on Theorem D and Corollaries D.1–D.3.

Remark D.1. The bounds on $\tau_*(k, \bar{G})$, for Abelian $\bar{G}$, described in Theorem A, complement the upper bound $t_{\mathrm{mix}}(G_k) \leq t_{\mathrm{mix}}(\bar{G}_k)$ to give explicit bounds on $t_{\mathrm{mix}}(G_k)$ which hold w.h.p.

Remark D.2. In the course of proving this theorem, we prove an *exact* relation between the L_2 mixing time for the RWs on G_k and $\bar{G}_k$, namely

$$\mathbb{E}(d_{G_k}^{(2)}(t)) \leq \mathbb{E}(d_{\bar{G}_k}^{(2)}(t)).$$

As explained below, it is natural to conjecture that Theorem D does not require G to be nilpotent. The definition of the Abelian group $\bar{G}$ corresponding to G required G to be nilpotent. We extend this definition to allow general group G. (The definitions are equivalent if G is nilpotent.)

The following conjecture extends Theorem D; it contains, as a special case, Wilson's conjecture.

Conjecture D. *Let G be a group, $(G_{(\ell)})_{\ell \geq 0}$ be its lower central series and let* $L := \min\{\ell \geq 0 \mid G_{(\ell)} = \{\mathsf{id}\}\}$. *Let the prime decomposition of $|G_L|$ be $|G_L| = \prod_1^r p_j$. Set* $\bar{G} := (\bigoplus_1^L (G_{(\ell-1)}/G_{(\ell)})) \oplus (\bigoplus_1^r \mathbb{Z}_{p_j})$. *Suppose that $1 \ll \log k \ll \log|G|$ and $k - d(\bar{G}) \gg 1$. Then, $t_{\text{mix}}(G_k)/t_{\text{mix}}(\bar{G}_k) \leq 1 + o(1)$ w.h.p.*

We are showing in Theorem D, for nilpotent groups, that being non-Abelian can only speed up mixing. Finite nilpotent groups are intuitively thought of as 'almost Abelian'; this is partially because two elements having co-prime orders must commute. Removing the nilpotent property should only mean the group is 'farther from Abelian', and thus is expected to speed up mixing.

1.2.3. Expander graphs of nilpotent groups. Our last result considers the expansion properties of the random Cayley graph.

Definition E. The *isoperimetric constant* of a finite d-regular graph $G = (V, E)$ is defined as

$$\Phi_* := \min_{1 \leq |S| \leq |V|/2} \Phi(S), \quad \text{where } \Phi(S) := \frac{|[\{a, b\} \in E \mid a \in S,\, b \in S^c]|}{d|S|}.$$

Theorem E. *Let G be a nilpotent group. Set $\bar{G} := \bigoplus_1^L (G_{(\ell-1)}/G_{(\ell)})$, where $(G_{(\ell)})_{\ell \geq 0}$ is the lower central series of G and $L := \min\{\ell \geq 0 \mid G_{(\ell)} = \{\mathsf{id}\}\}$. Then, for all $c > 0$, there exists a $c' > 0$ such that if $k - d(\bar{G}) \geq c \log|G|$, then $\Phi_*(G_k) \geq c'$ w.h.p.*

Remark E. This theorem is already known when $k - \log_2|G| \asymp k$, without the nilpotent restriction: it is a celebrated result of Alon and Roichman [3]. For them, the constant c' vanished as k got closer to $\log_2|G|$. Our result removes this when $d(\bar{G})$ is not close to $\log_2|G|$, e.g., $d(\bar{G}) \leq 0.99 \log_2|G|$.

1.2.4. Further remarks on Theorem A. Here we make some remarks on Theorem A in addition to the three in Section 1.2.1.

Remark A.5. This article establishes cutoff in a variety of set-ups, but always in the regime $1 \ll \log k \ll \log|G|$. This leaves the regimes $k \asymp 1$ and $\log k \asymp \log|G|$, for which there is no cutoff for any choice of generators: when $k \asymp 1$, this holds whenever the group is nilpotent; when $\log k \asymp \log|G|$, this holds for all groups. The former result is due to Diaconis and Saloff-Coste [15]; we give a short exposition of this in [25, §4]. We prove the latter in Theorem 7.2 below; the mixing time is of order 1. Dou [19, Theorems 3.3.1 and 3.4.7] establishes a more general result for $\log k \asymp \log|G|$.

Remark A.6. Our approach lifts the walk S from the Abelian Cayley graph $G(Z)$ to a walk W on the free Abelian group with $k = |Z|$ generators. Note that the walk W is independent of Z, i.e., of *which* k generators are chosen. We study the lifted walk W, in particular its entropic profile, before projecting back from W to S. This gives us a candidate mixing time; see Sections 1.3.5 and 2.1.

Remark A.7. The theorem is established via two distinct approaches: the former applies for k not growing too rapidly; the second can be seen as a refinement of the first, optimised for larger k, where the first breaks down. We combine the two approaches to analyse an interim regime of k.

We separate the exposition of the approaches: they are given in Sections 2, 3 and 4, respectively. In the first two, a concept of *entropic times* is defined; see Sections 2.2 and 3.2. A precise statement for each approach is given; see Sections 2.4, 3.4 and 4.1, respectively. In summary, Theorem A is a direct consequence of Propositions 2.2 and 3.2 and Theorems 2.5, 3.7 and 4.2.

1.3. Historic overview

In this subsection, we give a fairly comprehensive account of previous work on mixing and cutoff for random walk on random Cayley graphs; we compare our results with existing ones. The occurrence of cutoff in particular has received a great deal of attention over the years. We also mention, where relevant, other results which we have proved in companion papers; see also Section 1.4.3.

1.3.1. Universal cutoff: The Aldous–Diaconis conjecture. Aldous and Diaconis stated their conjecture for $k \gg \log|G|$; see [1, p. 40]. A more refined version is given by Dou [19, Conjectures 3.1.2 and 3.4.5]; see also [30, 44]. An informal, more general, variant was reiterated by Diaconis [13, Chapter 4G, Question 8]; he gave some related open questions recently in [14, §5]. Towards the conjecture, an upper bound, valid for arbitrary groups, was established by Dou and Hildebrand [18, Theorem 1] and later Roichman [44, Theorems 1 and 2], who simplified their argument. A matching lower bound, valid only for Abelian groups, was given by Hildebrand [30, Theorem 3]; see also Hildebrand [32, Theorem 5]. Dou and Hildebrand [18, Theorem 4] modify the proof of [30, Theorem 3] to extend the lower bound from Abelian groups to some families of groups with irreducible representations of bounded degree. Combined, this established the Aldous–Diaconis conjecture for Abelian groups and such groups with low degree irreducible representations. Moreover, the cutoff time was determined explicitly: it is at

$$T(k,|G|) := \frac{\log|G|}{\log(k/\log|G|)} = \frac{\rho}{\rho-1}\log_k|G|, \quad \text{where } \rho \text{ is defined by } k = (\log|G|)^\rho.$$

(Having $k \gg \log|G|$ needs $\rho - 1 \gg 1/\log\log|G|$.) See also Dou [19] and Hildebrand [32].

There is a trivial diameter-based lower bound of $\log_k|G|$. If $\rho \gg 1$, i.e., k is super-polylogarithmic in $|G|$, then $T(k,|G|) \eqsim \log_k|G|$. Thus, cutoff is established for all groups for such k.

In [24, Theorem B], using the group $U_{m,d}$ of $d \times d$ unit upper triangular matrices with entries in $\mathbb{Z}_m$, we disprove the part of the conjecture concerning the independence of the cutoff time from the algebraic structure of the group: if $d \geq 3$ is fixed and $k = (\log|U_{m,d}|)^{1+1/d}$, then there is cutoff at $(2/d)T(k,|U_{m,d}|)$. In fact, $T(k,|U_{m,d}|)$ does not even capture the correct order: letting $d \to \infty$ sufficiently slowly, we still have $k =$

$(\log|U_{m,d}|)^{1+1/d} \gg \log|U_{m,d}|$ and the cutoff time is still shown to be $(2/d)T(k,|U_{m,d}|)$, which is $o(T(k,|U_{m,d}|))$.

There has been a little investigation into the regime $1 \ll k \lesssim \log|G|$, but with much less success. Hildebrand [30, Theorem 4] showed that the mixing time must be super-polylogarithmic, unlike for $k \gg \log|G|$. Wilson [48, Theorem 1] established cutoff for $\mathbb{Z}_2^d$; this naturally requires $k \geq d = \log_2|G|$. Regarding $1 \ll k \ll \log|G|$, a breakthrough came recently when Hildebrand [33, Theorem 1.7] established cutoff for $\mathbb{Z}_p$ with $1 \ll k \leq \log p / \log\log p$ and p a (diverging) prime. The techniques were specialised to their respective cases; we consider arbitrary Abelian groups.

Relatedly, Hildebrand [31] analysed the regime in which k is just above the $\log_2 n$ threshold. For $k = a \log_2 n$ with $a > 1$, he established an upper bound on the mixing time of $a \log(a/(a-1)) \log_2 n$ w.h.p.; see [31, Theorem 1]. However, this is quite far from tight when a is large:

$$a \log\left(\frac{a}{a-1}\right) \geq 1 \quad \text{for all } a > 1,$$

yet

$$\frac{T(a \log_2 n, n)}{\log_2 n} \eqsim \frac{1}{\log_2 a} \to 0 \quad \text{as } a \to \infty.$$

He also analysed $k = \log_2 n + f(n)$ with $1 \ll f(n) \ll \log n$; see [31, Theorem 2].

1.3.2. Comparison of mixing times. In the direction of comparison of mixing times, there has been much less work. The only work of note (of which we are aware) is by Pak [40]. There, he studies universal mixing bounds (i.e., ones valid for all groups), but his bounds are not tight; they are always at least a constant factor away from those conjectured by Wilson [48] (and by us above).

A related universal bound in which $\mathbb{Z}_2^d$ is the worst case is given by Pak [41]. Let $\varphi_k(G) := \mathbb{P}(G_k$ is connected), i.e., the probability that the group G is generated by k uniformly chosen generators. Then, Pak [41, Lecture 1, Theorem 6] proves that if $|G| \leq 2^d$, then $\varphi_k(G) \geq \varphi(\mathbb{Z}_2^d)$ for all k.

1.3.3. Random walks on upper triangular matrix groups. The study of random walks on Heisenberg groups and other groups of upper triangular matrices has a rich history. We give a detailed historical account in [24, §1.3.2].

As noted above, in [24] we study $d \times d$ unit upper triangular matrices with entries in $\mathbb{Z}_m$. By viewing the Heisenberg group as $d \times d$ matrices (see Section 1.1.3), these $d \times d$ unit upper triangular matrices can be seen as a supergroup of the d-dimensional Heisenberg group.

1.3.4. Expander graphs for nilpotent groups. A celebrated result of Alon and Roichman [3, Corollary 1] asserts that, for any finite group G, the random Cayley graph with at least $C_\varepsilon \log|G|$ random generators is w.h.p. an ε-expander, provided C_ε is sufficiently large in terms of ε. (A graph is an *ε-expander* if its isoperimetric constant is bounded below by ε; up to a reparametrisation, this is equivalent to having the spectral gap of the graph bounded below by ε.) There has been a considerable line of work building upon this general result of Alon and Roichman [3]. (Pak [39] proves a similar result.) Their proof was simplified and extended, independently, by Loh and Schulman [36] and Landau and Russell [34]; both were able to replace $\log_2|G|$ by $\log_2 D(G)$, where $D(G)$ is the sum of the dimensions of the irreducible representations of the group G; for Abelian groups $D(G) = |G|$. A 'derandomised' argument for Alon–Roichman is given by Chen, Moore and Russell [10]. Both [10,34] use some Chernoff-type bounds on operator valued random variables.

Christofides and Markström [11] improve these further by using matrix martingales and proving a Hoeffding-type bound on operator valued random variables. They also improved the quantification for C_ε, showing that one may take $C_\varepsilon := 1 + c_\varepsilon$ with $c_\varepsilon \to 0$ as $\varepsilon \to 0$; this means that, w.h.p., the graph is an ε-expander whenever $k \geq (1 + c_\varepsilon)\log_2 D(G)$ and $c_\varepsilon \to 0$ as $\varepsilon \to 0$. They also generalise Alon–Roichman to random coset graphs. The proofs use tail bounds on the (random) eigenvalues.

It is well known that $D(G) \geq \sqrt{|G|}$. Thus, all these results require at least $k \geq (\log_2|G|)/2$. Our result, on the other hand, applies to $k \geq c\log|G|$ for *any* constant $c > 0$, provided the underlying group is suitable – e.g., this is the case if G is Abelian and $d(G) \ll \log|G|$; another example is given by $d \times d$ unit upper triangular matrix groups with entries in $\mathbb{Z}_m$ if $m \gg 1$.

Hildebrand [33, Theorem 1.1] showed, for all diverging (sequences of) primes p, that the order of the relaxation time of the RW on the cyclic group $\mathbb{Z}_p$ is $p^{2/k}$ when $1 \ll k \leq \log p/\log\log p$.

In [27, Theorem E], we restrict to Abelian groups under the assumption $k - 2d(G) \asymp k$ and determine, via an altogether different method, the order of the relaxation time whenever $1 \ll k \lesssim \log|G|$: it is $|G|^{2/k}$ w.h.p. This extends Theorem E in the Abelian set-up to allow $1 \ll k \ll \log|G|$.

1.3.5. Cutoff for 'generic' Markov chains and the entropic method. We now put our results into a broader context. A recurrent theme in the study of mixing times is that 'generic' instances often exhibit the cutoff phenomenon. In this set-up, a family of transition matrices chosen from a certain family of distributions is shown to give rise to a sequence of Markov chains which exhibits cutoff w.h.p. A few notable examples include random birth and death chains [16, 47], the simple or non-backtracking random walk on various models of sparse random graphs, including random regular graphs [38], random graphs with given degrees [5–8], the giant component of the Erdős–Rényi random graph [7] (where the authors consider mixing from a 'typical' starting point) and a large family of sparse Markov chains [8], as well as random walks on a certain generalisation of Ramanujan graphs [9] and random lifts [9, 12].

A recurring idea in the aforementioned papers is that the cutoff time can be described in terms of entropy. One can look at some auxiliary random process which up to the cutoff time can be coupled with, or otherwise related to, the original Markov chain – often in the above examples this is the RW on the corresponding Benjamini–Schramm local limit. The cutoff time is then shown to be (up to smaller order terms) the time at which the entropy of the auxiliary process equals the entropy of the invariant distribution of the original Markov chain. It is a relatively new technique, and has been used recently in [7–9, 12]. For 'most' regimes of k, this is the case for us too; further, for the non-Abelian groups considered in [24] we use a similar idea. As our auxiliary random process, we use a SRW (resp. DRW) in the undirected (resp. directed) case.

With the exception of the very recent [28], to the best of our knowledge, in all previous instances where the entropic method was used the graphs were tree-like. This is not the case for us: in the Abelian set-up, G_k has cycles of length 4 (potentially up to the direction of edges). Admittedly, this has less of an impact on the walk since each vertex is of diverging degree.

1.3.6. Subsequent work. The release of this multi-paper project in early 2021 spurred significant interest and progress on several related problems. Salez [45] established a sufficient condition for cutoff involving an entropic concentration criterion, and verified it for RW on any undirected Abelian Cayley graph which is an expander. The collection of Cayley graphs was extended beyond expanders by Hermon et al. [22].

Regarding *random* Cayley graphs, Hermon and Huang [21] built on the ideas initiated here, hinging on multiple aspects of the proofs of Theorems A and D. They established cutoff for SRW on G_k for nilpotent G, but required bounded step and *rank* $d(G^{\mathrm{ab}})$. The generality permitted by allowing divergent rank $d(G) = d(G^{\mathrm{ab}})$ for Abelian G is one of the major improvements of Theorem A over previous work. Also, several of our examples in Corollary D, such as high-dimensional Heisenberg groups $H_{m,d}$ ($d \gg 1$), are ruled out. Under further restriction, they showed that the mixing times of the RW on G and the projection to G^{ab} are asymptotically equivalent.

Even more recently, Pedrotti and Salez [42] introduced a new criterion for cutoff for Markov chains with non-negative *curvature*; RW on an undirected Abelian Cayley graph has this property. Again, cutoff can be deduced when the graph is an expander, but also under some weaker, quantitative conditions. No estimate on the mixing time itself is given in [42, 45], though, so connection to the motivating Aldous–Diaconis conjecture is lost. Again, directed graphs are excluded.

1.4. Additional remarks

1.4.1. Precise definition of Cayley graphs. Let G be a finite group and Z a multisubset of G. We focus on mixing properties of the *Cayley graph* of G with *generators* Z. The *undirected* (resp. *directed*) *Cayley graph of G generated by Z*, denoted $G^-(Z)$ (resp. $G^+(Z)$), is the multigraph with vertex set G and edge multiset

$$[\{g, g \cdot z\} \mid g \in G,\ z \in Z] \quad (\text{resp. } [(g, g \cdot z) \mid g \in G,\ z \in Z]).$$

If the walk is at $g \in G$, then a step in $G^+(Z)$ (resp. $G^-(Z)$) involves choosing a generator $z \in Z$ uniformly at random and moving to gz (resp. one of gz or gz^{-1} each with probability $1/2$).

We focus attention on the *random* Cayley graph defined by choosing $Z_1, \dots, Z_k \sim^{\text{iid}} \text{Unif}(G)$; when this is the case, denote $G_k^+ := G^+(Z)$ and $G_k^- := G^-(Z)$. Whilst we do not assume that the Cayley graph is connected (i.e., Z may not generate G), in the Abelian set-up the random Cayley graph G_k is connected w.h.p. whenever $k - d(G) \gg 1$; see [25, Lemma 8.1]. In the nilpotent set-up, this is the case whenever $k - d(G/[G, G]) \gg 1$; see [24, Remark E.1].

The graph depends on the choice of Z. Sometimes, it is convenient to emphasise this; we use a subscript, writing $\mathbb{P}_{G(z)}(\cdot)$ if the graph is generated by the group G and multiset z. Analogously, $\mathbb{P}_{G_k}(\cdot)$ stands for the *random* law $\mathbb{P}_{G(Z)}(\cdot)$, where $Z = [Z_1, \dots, Z_k]$ with $Z_1, \dots, Z_k \sim^{\text{iid}} \text{Unif}(G)$.

1.4.2. Typical and simple Cayley graphs. The directed Cayley graph $G^+(z)$ is simple if and only if no generator is picked twice, i.e., $z_i \neq z_j$ for all $i \neq j$. The undirected Cayley graph $G^-(z)$ is simple if in addition no generator is the inverse of any other, i.e., $z_i \neq z_j^{-1}$ for all $i, j \in [k]$. In particular, this means that no generator is of order 2, as any $s \in G$ of order 2 satisfies $s = s^{-1}$ – this gives a multiedge between g and gs for each $g \in G$.

The RW on $G^-(z)$ is equivalent to an adjusted RW on $G^+(z)$, where when a generator $s \in z$ is chosen, instead of applying a generator s, either s or s^{-1} is applied, each with probability $1/2$. Abusing terminology, we relax the definition of simple Cayley graphs to allow order 2 generators.

Given a group G and an integer k, we are drawing the generators $Z_1, \dots, Z_k$ independently and uniformly at random. It is not difficult to see that the probability of drawing a given multiset depends only on the number of repetitions in that multiset. Thus, conditional on being simple, G_k is uniformly distributed on all simple degree-k Cayley graphs. Since $k \ll \sqrt{|G|}$, the probability of simplicity tends to 1 as $|G| \to \infty$. So, when we say that our results hold 'w.h.p. (over Z)', we could equivalently say that the result holds 'for almost all degree-k simple Cayley graphs of G'.

Our asymptotic evaluation does not depend on the particular choice of Z, so the statistics in question depend very weakly on the particular choice of generators for almost all choices. In many cases, the statistics depend only on G via $|G|$ and $d(G)$. This is a strong sense of 'universality'.

1.4.3. Overview of random Cayley graphs project. This paper is one part of an extensive project on random Cayley graphs. There are three main articles (including the current one and [24, 27]), a technical report [25] and a supplementary document [26] containing deferred technical proofs. *Each main article is readable independently.*

The main objective of the project is to establish cutoff for the random walk and to determine whether this can be written in a way that, up to subleading order terms, depends only on k and $|G|$; we also study universal mixing bounds, valid for all, or large classes of, groups. Separately, we study the distance of a uniformly chosen element from the identity, i.e., typical distance, and the diameter; the main objective is to show that these distances concentrate and to determine whether the value at which these distances concentrate depends only on k and $|G|$.

- Cutoff phenomenon (and Aldous–Diaconis conjecture) for general Abelian groups; also, for nilpotent groups, expander graphs and comparison of mixing times with Abelian groups (the current paper).
- Typical distance, diameter and spectral gap for general Abelian groups [27].
- Cutoff phenomenon and typical distance for upper triangular matrix groups [24].
- Additional results on cutoff and typical distance for general Abelian groups [25].

2. Total variation cutoff: Approach #1

In this section, we prove the first part of the upper bound on mixing for arbitrary Abelian groups. The main result of the section is Theorem 2.5. The outline of the section is as follows:

- Section 2.1 introduces the *entropic method*.
- Section 2.2 defines *entropic times* and states a CLT.
- Section 2.3 sketches arguments to evaluate these entropic times.
- Section 2.4 states precisely the main theorem of the section.
- Section 2.5 outlines the argument.
- Section 2.6 is devoted to the lower bound.
- Section 2.7 is devoted to the upper bound.

2.1. Entropic times: Methodology

We use an 'entropic method', as mentioned in Section 1.3; cf. [7–9, 12]. The method is fairly general; we now explain the specific application in a little more depth.

We define an auxiliary random process $(W(t))_{t \geq 0}$, recording how many times each generator has been used: for $t \geq 0$, for each generator $i = 1, \dots, k$, write $W_i(t)$ for the number of times that it has been picked by time t. By independence, $W(\cdot)$ forms a rate-1 DRW on $\mathbb{Z}_+^k$. For the undirected case, recall that we either apply a generator or its inverse; when we apply the inverse of generator i, the increment is $W_i \to W_i - 1$ (rather than $W_i \to W_i + 1$). In this case, $W(\cdot)$ is a SRW on $\mathbb{Z}^k$.

Since the underlying group is Abelian, the order in which the generators are applied is irrelevant and generator-inverse pairs cancel. Hence, we can write

$$S(t) = \sum_{i=1}^{k} W_i(t) Z_i = W(t) \cdot Z.$$

Recall that the uniform distribution is invariant, regardless of the group and generators. For an Abelian group G, we propose as the mixing time the time at which the auxiliary process W obtains entropy $\log|G|$. The reason for this is the following: take t to be slightly larger than the above entropic time; using the equivalence $-\log \mu \geq \log|G|$

if and only if $\mu \le 1/|G|$, 'typically' $W(t)$ takes values to which it assigns probability smaller than $1/|G|$; informally, this means that $W(t)$ is 'well spread out'. We can have two independent copies S and S' (using the same generators Z) with $S(t) = S'(t)$ but $W(t) \neq W'(t)$. The uniformity of the generators will show that, on average, this is unlikely. We thus deduce that $S(t)$ is well spread out, i.e., well mixed.

Contrastingly, if the entropy is much smaller than $\log|G|$, then $W(t)$ is not well spread out: it is highly likely to lie in a set of size $o(1/|G|)$. The same must be true for $S(t)$; hence, $S(t)$ is not mixed.

2.2. *Entropic times: Definition and concentration*

We now define precisely the notion of *entropic times*. Write μ_t (resp. ν_s) for the law of $W(t)$ (resp. $W_1(sk)$); so $\mu_t = \nu_{t/k}^{\otimes k}$. Define

$$Q_i(t) := -\log \nu_{t/k}(W_i(t)), \quad \text{and set } Q(t) := -\log \mu_t(W(t)) = \sum_{i=1}^{k} Q_i(t).$$

So $\mathbb{E}(Q(t))$ and $\mathbb{E}(Q_1(t))$ are the entropies of $W(t)$ and $W_1(t)$, respectively. Observe that $t \mapsto \mathbb{E}(Q(t))\colon [0,\infty) \to [0,\infty)$ is a smooth, increasing bijection.

Definition 2.1 (Entropic and cutoff times). For all $k, n \in \mathbb{N}$ and all $\alpha \in \mathbb{R}$, define $\mathfrak{t}_\alpha$ so that

$$\mathbb{E}(Q_1(\mathfrak{t}_\alpha)) = \frac{\log n + \alpha\sqrt{vk}}{k} \quad \text{and} \quad \mathfrak{s}_\alpha := \frac{\mathfrak{t}_\alpha}{k}, \quad \text{where } v := \mathbb{V}\mathrm{ar}(Q_1(\mathfrak{t}_0)),$$

assuming that $\log n + \alpha\sqrt{vk} \ge 0$. We call $\mathfrak{t}_0$ the *entropic time* and the $\{\mathfrak{t}_\alpha\}_{\alpha\in\mathbb{R}}$ *cutoff times*.

Comparing with notation in the introduction, $\mathfrak{t}_0 = \tau_\infty(G)$; see Definition A. The definition there was for *cutoff* only; the *profile* is described by the full range $(\mathfrak{t}_\alpha)_{\alpha\in\mathbb{R}}$, in the regime handled here.

Direct calculation with the Poisson distribution and SRW on $\mathbb{Z}$ gives the following relations.

Proposition 2.2 (Entropic and cutoff times). *Assume that $1 \ll k \ll \log n$. Then, for all $\alpha \in \mathbb{R}$,*

$$\mathfrak{t}_\alpha \eqsim \mathfrak{t}_0 \eqsim \frac{k \cdot n^{2/k}}{2\pi e} \quad \textit{and} \quad \frac{\mathfrak{t}_\alpha - \mathfrak{t}_0}{\mathfrak{t}_0} \eqsim \alpha\sqrt{\frac{2}{k}} \ll 1.$$

The idea is to approximate the SRW and DRW laws by a normal distribution, then calculate the entropy of this. A rigorous proof is long and tedious, requiring many careful approximations. We sketch the principal ideas below in Section 2.3. The precise details are deferred to [26, Proposition A.2].

Since $Q = \sum_k^1 Q_i$ is a sum of $k \gg 1$ iid random variables, $Q(\mathfrak{t}_0)$ concentrates around $\mathbb{E}(Q(\mathfrak{t}_0)) = \log N$. One can show that multiplying the time by a factor $1+\xi$ for any constant $\xi > 0$ increases the entropy by a significant amount; similarly, if $\xi < 0$, then the entropy decreases significantly. Further, the change is by an additive term of larger order than the standard deviation $\sqrt{\mathbb{V}\mathrm{ar}(Q(\mathrm{t}_0))}$. Thus, $Q((1+\xi)\mathrm{t}_0)$ concentrates around this new value. In particular, the following hold:

$$\mu_{(1+\xi)\mathrm{t}_0}(W((1+\xi)\mathrm{t}_0)) = e^{-Q((1+\xi)\mathrm{t}_0)} \ll \frac{1}{n} \quad \text{w.h.p.},$$
$$\mu_{(1-\xi)\mathrm{t}_0}(W((1-\xi)\mathrm{t}_0)) = e^{-Q((1-\xi)\mathrm{t}_0)} \gg \frac{1}{n} \quad \text{w.h.p.}$$

The following proposition quantifies this change in entropy and this concentration.

Proposition 2.3 (CLT). *Assume that $1 \ll k \ll \log n$. For all $\alpha \in \mathbb{R}$, we have*

$$\mathbb{P}(Q(\mathrm{t}_\alpha) \le \log n \pm \omega) \to \Psi(\alpha) \quad \textit{for } \omega := \mathbb{V}\mathrm{ar}(Q(\mathrm{t}_0))^{1/4} = (\upsilon k)^{1/4}.$$

(There is no specific reason for choosing this ω. We just need some ω with $1 \ll \omega \ll (\upsilon k)^{1/2}$.)

This follows without too much difficulty from the local CLT. Again, though, the details are technical – albeit less so than for the entropic times. We defer the proof to [26, Proposition A.3]

2.3. *Entropic times: Sketch evaluation*

In this subsection, we sketch details towards a proof of Proposition 2.2. The full, rigorous details can be found in [26, Proposition A.2], where all of the approximations below are carefully justified.

Recall that t_0 is the time t at which the entropy of $W_1(t)$, which is a rate-$1/k$ RW, is $(\log n)/k$. We need to find the variance $\mathbb{V}\mathrm{ar}(Q_1(\mathrm{t}_0))$, as this is used in the definition of t_α, given in Definition 2.1.

In the sketch below, we replace $\mathbb{V}\mathrm{ar}(Q_1(\mathrm{t}_0))$ by an approximation.

For $s \ge 0$, denote $X_s := W_1(sk)$ for $s \ge 0$ and the entropy of X_s as $H(s)$. The target entropy $\log n/k \gg 1$, and so the entropic time $\mathfrak{s}_0 \gg 1$. For $s \gg 1$, the RW X_s has approximately the normal $N(\mathbb{E}(X_s), s)$ distribution. Translating the random variable has no affect on its entropy, and so we approximate the entropy $H(s)$ of X_s by the entropy $\bar{H}(s)$ of a $N(0,s)$ random variable. Direct calculation with the normal distribution gives

$$\bar{H}(s) = \frac{1}{2}\log(2\pi e s), \quad \text{and hence } \bar{H}'(s) = \frac{1}{2s}.$$

Define $\bar{\mathfrak{s}}_\alpha$ as the entropic times for the approximation

$$\bar{H}(\bar{\mathfrak{s}}_\alpha) = \frac{\log n + \alpha\sqrt{\upsilon k}}{k},$$

where

$$\bar{\upsilon} := \mathbb{V}\mathrm{ar}(\bar{Q}_1(\bar{\mathfrak{s}}_0 k)),$$

where $\bar{Q}_1(sk)$ is the analogue of $Q_1(sk)$, except with $W_1(sk)$ replaced by $N(0,s)$. Hence,

$$\bar{H}(\mathfrak{s}_0) = \log n \quad \text{implies that } \bar{\mathfrak{s}}_0 = \frac{n^{2/k}}{2\pi e} \gg 1.$$

By direct calculation, specific to the normal distribution, one finds

$$\mathbb{V}\text{ar}(\bar{Q}_1(sk)) = \frac{1}{2}.$$

As mentioned above, for this sketch, to ease the calculation of $\mathfrak{t}_\alpha$ in Definition 2.1, we replace $\mathbb{V}\text{ar}(Q_1(\mathfrak{t}_0))$ by its approximation $1/2$, and assume the above normal distribution approximation.

In order to find the window, assuming for the moment that $\alpha > 0$, we write

$$\mathfrak{s}_\alpha - \mathfrak{s}_0 = \int_0^\alpha \frac{d\mathfrak{s}_a}{da}\, da.$$

Again, we replace $\mathfrak{s}_\alpha$ by $\bar{\mathfrak{s}}_\alpha$. By definition, $\bar{\mathfrak{s}}_\alpha$ satisfies

$$\bar{H}(\bar{\mathfrak{s}}_\alpha) = \log\frac{n}{k} + \frac{\alpha}{\sqrt{2k}}, \quad \text{and hence } \frac{d\bar{\mathfrak{s}}_\alpha}{d\alpha}\bar{H}'(\bar{\mathfrak{s}}_\alpha) = \frac{1}{\sqrt{2k}}.$$

Using the expressions for $d\bar{\mathfrak{s}}_a/da$ and $\bar{H}'(s) = 1/(2s)$ above, we find that

$$\bar{\mathfrak{s}}_\alpha - \bar{\mathfrak{s}}_0 = (2k)^{-1/2}\int_0^\alpha 2\bar{\mathfrak{s}}_a\, da \approx (2k)^{-1/2}\int_0^\alpha 2\bar{\mathfrak{s}}_0\, da = \alpha\bar{\mathfrak{s}}_0\sqrt{\frac{2}{k}},$$

since $\bar{\mathfrak{s}}_a$ only varies by subleading order terms over $a \in [0,\alpha]$. The argument is analogous for $\alpha < 0$.

We have now shown the desired result for $\bar{\mathfrak{s}}_\alpha$, i.e., when approximating $W_1(sk)$ by $N(\mathbb{E}(X_s), s)$. It turns out that this approximation is sufficiently good for the results to pass over to the original case, i.e., to apply to $\mathfrak{s}_0$ and $\mathfrak{t}_0 = \mathfrak{s}_0 k$. This is made rigorous in [26, §A] via a local CLT.

2.4. *Precise statement and remarks*

In this subsection, we state precisely the main theorem of the section. There are some simple conditions on k, in terms of $d(G)$ and $|G|$, needed for the upper bound.

Hypothesis A. *The sequence $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis A if the following hold:*

$$\lim_{N\to\infty} |G_N| = \infty, \qquad \lim_{N\to\infty} (k_N - d(G_N)) = \infty$$

and

$$\frac{k_N - d_N(G_N) - 1}{k_N} \geq 5\frac{k_N}{\log|G_N|} + 2\frac{d_N(G_N)\log\log k_N}{\log|G_N|} \quad \textit{for all } N \in \mathbb{N}.$$

Remark 2.4. Write $n := |G|$. Any of the following conditions imply Hypothesis A:

$$
\begin{aligned}
&1 \ll k \lesssim \sqrt{\frac{\log n}{\log\log\log n}} && \text{and} \quad k-d \gg 1;\\
&1 \ll k \lesssim \sqrt{\log n} && \text{and} \quad k-d \gg \log\log k;\\
&1 \ll k \ll \frac{\log n}{\log\log\log n} && \text{and} \quad k-d \geq \delta k \ \text{ for some suitable } \delta = o(1);\\
&d \ll \log \frac{n}{\log\log\log n} && \text{and} \quad k-d \asymp k \ll \log n.
\end{aligned}
$$

Throughout the proofs, we drop the subscript-N from the notation, e.g., writing k or n, considering sequences implicitly. Recall that we abbreviate the TV distance from uniformity at time t as

$$d_{G_k,N}(t) = \|\mathbb{P}_{G_N([Z_1,\ldots,Z_{k_N}])}(S(t) \in \cdot) - \pi_{G_N}\|_{\mathrm{TV}}, \text{ where } Z_1,\ldots,Z_{k_N} \sim^{\mathrm{iid}} \mathrm{Unif}(G_N).$$

We now state the main theorem of this section. Recall that Ψ is the standard Gaussian tail.

Theorem 2.5. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ of finite, Abelian groups; for each $N \in \mathbb{N}$, define $Z_{(N)} := [Z_1,\ldots,Z_{k_N}]$ by drawing $Z_1,\ldots,$ $Z_{k_N} \sim^{\mathrm{iid}}$ $\mathrm{Unif}(G_N)$. Suppose that the sequence $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis A.*

For all $\alpha \in \mathbb{R}$ and all $N \in \mathbb{N}$, write $\mathrm{t}_{\alpha,N} := \mathrm{t}_\alpha(k_N, |G_N|)$. Let $\alpha \in \mathbb{R}$. Then,

$$\frac{\mathrm{t}_{\alpha,N}}{\mathrm{t}_{0,N}} \to 1 \quad \textit{and} \quad d_{G_k,N}(\mathrm{t}_{\alpha,N}) \to^{\mathbb{P}} \Psi(\alpha) \quad \textit{(in probability) as } N \to \infty.$$

That is, w.h.p. there is TV cutoff at t_0 with profile given by $\{\mathrm{t}_\alpha\}_{\alpha\in\mathbb{R}}$: for all $\varepsilon \in (0,1)$, the difference in the mixing times $t_{\mathrm{mix}}(\varepsilon) - t_{\mathrm{mix}}(1/2)$ is given, up to smaller order terms, by $\mathrm{t}_{\Psi^{-1}(\varepsilon)} - \mathrm{t}_0$. Moreover, the implicit lower bound on the TV distance holds deterministically, i.e., for all choices of generators.

Remark. Using Proposition 2.2, we can write the cutoff statement in the form

$$\frac{t_{\mathrm{mix}}(\varepsilon) - \mathrm{t}_0}{w} \to^{\mathbb{P}} \Psi^{-1}(\varepsilon) \quad \text{w.h.p. } \forall \varepsilon \in (0,1),$$

where $\mathrm{t}_0 \eqsim k|G|^{2/k}/(2\pi e)$ is the mixing time and $w \eqsim \sqrt{k}|G|^{2/k}/(\sqrt{2}\pi e)$ the window.

Remark. The CLT, Proposition 2.3, will give the dominating term in the TV distance:

- on the event $\{Q(\mathrm{t}_\alpha) \leq \log n - \omega\}$, we lower bound the TV distance by $1 - o(1)$;
- on the event $\{Q(\mathrm{t}_\alpha) \geq \log n + \omega\}$, we upper bound the expected TV distance by $o(1)$.

Combining this with the CLT, we deduce that $d_{G_k}(\mathrm{t}_\alpha) \to \Psi(\alpha)$ in probability.

Remark. Observe that Hypothesis A does not cover the regime $k \gtrsim \log|G|$. Under fairly mild conditions on the group, we can apply a variation on the argument given below to obtain a limit profile result for any $1 \ll \log k \ll \log|G|$. The detailed analysis is carried out in [25, §2].

2.5. Outline of proof

We now give a high-level description of our approach, introducing notations and concepts along the way. No results or calculations from this section will be used in the remainder of the document. Further, we restrict attention to establishing *cutoff* only, not the *limit profile*: take $t = (1 \pm \varepsilon)\mathfrak{t}_0$.

In all cases, we show that cutoff occurs around the entropic time. As $Q(t)$ is a sum of many iid random variables, we expected it to concentrate around its mean. Loosely speaking, we show that the shape of the cutoff, i.e., the profile of the convergence to equilibrium, is determined by the fluctuations of $Q(t)$ around its mean, which in turn, by the CLT (Proposition 2.3), are determined by $\mathbb{V}\mathrm{ar}(Q(t))$, for t 'close' to $\mathfrak{t}_0$. Note that $\mathbb{V}\mathrm{ar}(Q(t)) = k\mathbb{V}\mathrm{ar}(Q_1(t))$ since the Q_i are iid.

Throughout this section (Section 2.5), we write 0 for the identity element of the Abelian group G, as is standard. We now outline the proof in more detail. We often drop t-dependence from the notation.

We start by discussing the lower bound. If Q is sufficiently small, then W, and hence also S, is restricted to a small set. Indeed, $Q \le \log n - \omega$ if and only if $\mu(W) \ge n^{-1}e^{\omega}$, and thus if this is the case, then $W \in \{w \mid \mu(w) \ge n^{-1}e^{\omega}\}$. Since we generate S via W, it is thus also the case that

$$S \in E := \{g \in G \mid \mathbb{P}(S = g) \ge n^{-1}e^{\omega}\}.$$

But clearly, $|E| \le ne^{-\omega}$. Choosing the time t slightly smaller than the entropic time $\mathfrak{t}_0$ and $\omega \gg 1$ suitably, the event $\{Q(t) \le \log n - \omega\}$ will hold w.h.p. Thus, w.h.p., $S(t)$ is restricted to a set of size $o(n)$. Hence, it cannot be mixed. This heuristic applies for *any* choice of generators.

Precisely, we show for any ω with $1 \ll \omega \ll \log n$, all t and all $Z = [Z_1, \ldots, Z_k]$, that

$$d_{G_k}(t) \ge \mathbb{P}(Q(t) \le \log n - \omega) - e^{-\omega}.$$

Thus, we are interested in the fluctuations of $Q(t)$ for t close to $\mathfrak{t}_0$. Using the CLT application above, i.e., Proposition 2.3 with $\omega := \mathbb{V}\mathrm{ar}(Q(\mathfrak{t}_0))^{1/4}$, we deduce the lower bound in Theorem 2.5.

We now turn to discussing the upper bound. The lower bound was valid for any choice of generators Z. Here, we exploit the independence and uniformity of the elements of Z.

Let $W'(t)$ be an independent copy of $W(t)$, and let $V(t) := W(t) - W'(t)$. Observe that, in both the undirected and directed case, the law of $V(t)$ is that of the rate-2 SRW in $\mathbb{Z}^k$, evaluated at time t. The standard L_2 calculation (using the Cauchy–Schwarz inequality) says that

$$2\|\zeta - \pi_G\|_{\mathrm{TV}} \le \|\zeta - \pi_G\|_2 = \sqrt{n \sum_{x \in G} \Big(\zeta(x) - \frac{1}{n}\Big)^2},$$

recalling that $\pi_G(x) = 1/n$ for all $x \in G$. A standard, elementary calculation shows that

$$\|\mathbb{P}_{G_k}(S(t) \in \cdot) - \pi_G\|_2 = \sqrt{n\mathbb{P}(V(t) \cdot Z = 0 \mid Z) - 1}.$$

Unfortunately, writing $X = (X(s))_{s \geq 0}$ for a rate-1 SRW on $\mathbb{Z}$, a simple calculation shows that

$$\mathbb{P}(V(\mathrm{t}_0) \cdot Z = 0 \mid Z) \geq \mathbb{P}(V(\mathrm{t}_0) = (0, \dots, 0) \in \mathbb{Z}^k) \\ = \mathbb{P}\Big(X\Big(\frac{2\mathrm{t}_0}{k}\Big) = 0\Big)^k \gg \frac{1}{n}.$$

(This calculation differs amongst the regimes of k.) Moreover, the L_2-mixing time can then be shown to be larger than the TV-mixing time by at least a constant factor; hence, this is insufficiently precise for showing cutoff in TV. We drop the t-dependence from the notation from now on.

This motivates the following 'modified L_2 calculation'. First, let $\mathcal{W} \subseteq \mathbb{Z}^k$, and write

$$\mathsf{typ} := \{W, W' \in \mathcal{W}\}, \quad \overline{\mathbb{P}}(\cdot) := \mathbb{P}(\cdot \mid \mathsf{typ}) \quad \text{and} \quad \overline{\mathbb{E}}(\cdot) := \mathbb{E}(\cdot \mid \mathsf{typ});$$

note that here we are (implicitly) averaging over Z. The set $\mathcal{W} \subseteq \mathbb{Z}^k$ will be chosen later so that

$$\overline{\mathbb{P}}(V = 0) = (V = 0 \mid \mathsf{typ}) \ll \frac{1}{n} \quad \text{and} \quad \mathbb{P}(W \notin \mathcal{W}) = o(1);$$

we call this *typicality*. We now perform the same L_2 calculation, but for $\overline{\mathbb{P}}$ rather than $\mathbb{P}$:

$$d_{G_k}(t) = \|\mathbb{P}_{G_k}(S \in \cdot) - \pi_G\|_{\mathrm{TV}} \leq \|\mathbb{P}_{G_k}(S \in \cdot \mid W \in \mathcal{W}) - \pi_G\|_{\mathrm{TV}} + \mathbb{P}(W \notin \mathcal{W})$$

and

$$4\mathbb{E}(\|\mathbb{P}_{G_k}(S \in \cdot \mid W \in \mathcal{W}) - \pi_G\|_{\mathrm{TV}}^2) \leq \mathbb{E}(|G|\overline{\mathbb{P}}(V \cdot Z = 0 \mid Z) - 1) \\ = |G|\overline{\mathbb{P}}(V \cdot Z = 0) - 1;$$

see Lemma 2.6. By taking expectation over Z and doing a modified L_2 calculation, we transformed the quenched estimation of the mixing time into an annealed calculation concerning the probability that a random word involving random generators is equal to the identity. This is a key step.

To have $w \in \mathcal{W}$, we impose *local* and *global typicality requirements*. The *global* part says that

$$-\log \mu(w) \geq \log n + \omega \quad \text{for all } w \in \mathcal{W},$$

where $\omega := (vk)^{1/4}$ as above; the *local* part will come later. For a precise statement of the typicality requirements, see Definition 2.7. These have the property that, when $t = (1+\varepsilon)\mathrm{t}_0$,

$$\mathbb{P}(W(t) \notin \mathcal{W}) \ll 1;$$

see Proposition 2.8. Then, since

$$-\log p \geq \log n + \omega \quad \text{if and only if} \quad p \leq n^{-1}e^{-\omega},$$

we have

$$\overline{\mathbb{P}}(V = (0, \dots, 0)) \asymp \mathbb{P}(W = W' \mid W' \in \mathcal{W}) \leq n^{-1}e^{-\omega}.$$

Of course, there are other scenarios in which we may have $V \cdot Z \equiv 0$. To deal with these, since linear combinations of independent uniform random variables in an Abelian group are uniform on their support, we have

$$v \cdot Z \sim \mathrm{Unif}(\mathfrak{g}_v G), \quad \text{where } \mathfrak{g}_v := \gcd(v_1, \ldots, v_k, n);$$

see Lemma 2.11. Then,

$$|G|\overline{\mathbb{P}}(V \cdot Z = 0,\, V \neq 0) = |G|\overline{\mathbb{E}}\Big(\frac{\mathbf{1}(V \neq 0)}{|\mathfrak{g}_V G|}\Big).$$

(Recall that V and Z are independent.) We use the *local* typicality conditions to ensure that $\max_i |W_i| \leq r_*$, for some explicit r_* which diverges a little faster than $n^{1/k}$. This allows us to consider only values $\mathfrak{g} \in [2r_*]$ for the gcd. It is here where the two approaches (Sections 2 and 3) diverge.

First (Section 2), we use a rather direct approach. Elementary group theory gives

$$|G|\overline{\mathbb{E}}\Big(\frac{\mathbf{1}(V \neq 0)}{|\mathfrak{g}_V G|}\Big) \leq \overline{\mathbb{E}}(\mathfrak{g}_V^{d(G)} \mathbf{1}(V \neq 0)) \leq 1 + \sum_{\gamma=2}^{2r_*} \gamma^{d(G)} \mathbb{P}(\mathfrak{g}_V = \gamma);$$

see Lemma 2.12. Since the law of SRW on $\mathbb{Z}$ is unimodal, for each non-zero coordinate, the probability that γ divides it is at most $1/\gamma$. Thus, in general, the probability is at most $1/\gamma$ plus the probability that the coordinate is 0, the latter of which is of order $1/\sqrt{t/k}$. This leads to

$$\overline{\mathbb{P}}(\mathfrak{g}_V = \gamma) \lesssim \Big(\frac{2}{n^{1/k}} + \frac{1}{\gamma}\Big)^k \quad \text{when } t \geq \mathfrak{t}_0;$$

see Lemma 2.14. Provided at least one of $d(G)$ or k is not too close to $\log n$, we are able to use this to control the expectation, showing $\overline{\mathbb{E}}(\mathfrak{g}_V^{d(G)} \mathbf{1}(V \neq 0)) = 1 + o(1)$ when $t \geq \mathfrak{t}_0$; see Corollary 2.15.

Combining these two analyses, we deduce that

$$n\overline{\mathbb{P}}(V \cdot Z = 0) \leq n\overline{\mathbb{P}}(V \cdot Z = 0, V \neq 0) + n\overline{\mathbb{P}}(V = 0) = 1 + o(1).$$

The modified L_2 calculation then says that the TV distance tends to 0 in probability, as required.

The only real adjustment needed to handle the profile is the use of the estimate

$$\mathbb{P}(W(\mathfrak{t}_\alpha) \notin \mathcal{W}) \eqsim \Psi(\alpha).$$

The remainder of the analysis is fairly robust to the specific value of t.

The second approach (Section 3) analyses the term $\overline{\mathbb{P}}(\mathfrak{g}_V = \gamma)$ and uses it to kill $|G/\gamma G|$ directly in

$$|G|\overline{\mathbb{E}}\Big(\frac{\mathbf{1}(V \neq 0)}{|\mathfrak{g}_V G|}\Big) = \sum_{\gamma \in \mathbb{N}} \overline{\mathbb{P}}(\mathfrak{g}_V = \gamma)|G/\gamma G|.$$

We outline the details of the adaptation in Section 3.5, including where Approach #1 breaks down.

This concludes the outline. We now move onto the formal proofs.

2.6. *Lower bound on total-variation mixing*

In this subsection, we prove the lower bound on mixing, which holds for all choices of generators.

Proof of lower bound in Theorem 2.5. For this proof only, to emphasise that the multisubset Z is fixed, not being averaged over, we add a subscript-Z to the probabilities involving Z: $\mathbb{P}_Z(S(\mathrm{t}_\alpha) \in \cdot)$.

For all $\alpha \in \mathbb{R}$, by the CLT (Proposition 2.3),

$$\mathbb{P}(\mathcal{E}_\alpha) \gtrsim \Psi(\alpha),$$

where

$$\mathcal{E}_\alpha := \{\mu(W(\mathrm{t}_\alpha)) \geq n^{-1}e^\omega\} = \{Q(\mathrm{t}_\alpha) \leq \log n - \omega\};$$

recall that $\omega \gg 1$. Fix $\alpha \in \mathbb{R}$. Consider the set

$$E_\alpha := \{x \in G \mid \exists w \in \mathbb{Z}^d \text{ st } \mu_{\mathrm{t}_\alpha}(w) \geq n^{-1}e^\omega \text{ and } x = w \cdot Z\}.$$

Then, $\mathbb{P}_Z(S(\mathrm{t}_\alpha) \in E_\alpha \mid \mathcal{E}_\alpha) = 1$ since W generates $S = W \cdot Z$. Every element $x \in E_\alpha$ can be realised as $x = w_x \cdot Z$ for some $w_x \in \mathbb{Z}^k$ with $\mu_{\mathrm{t}_\alpha}(w_x) \geq n^{-1}e^\omega$. Hence, for all $x \in E_\alpha$, we have

$$\mathbb{P}_Z(S(\mathrm{t}_\alpha) = x) \geq \mathbb{P}(W(\mathrm{t}_\alpha) = w_x) = \mu_{\mathrm{t}_\alpha}(w_x) \geq n^{-1}e^\omega.$$

Taking the sum over all $x \in E_\alpha$, we deduce that

$$1 \geq \sum_{x \in E_\alpha} \mathbb{P}_Z(S(\mathrm{t}_\alpha) = x) \geq |E_\alpha| \cdot n^{-1}e^\omega, \quad \text{and hence } \frac{|E_\alpha|}{n} \leq e^{-\omega} = o(1).$$

Finally, we deduce the lower bound from the definition of TV distance:

$$\begin{aligned}\|\mathbb{P}_Z(S(\mathrm{t}_\alpha) \in \cdot) - \pi_G\|_{\mathrm{TV}} &\geq \mathbb{P}_Z(S(\mathrm{t}_\alpha) \in E_\alpha) - \pi_G(E_\alpha) \geq \mathbb{P}(\mathcal{E}_\alpha) - \frac{1}{n}|E_\alpha| \\ &\geq \Psi(\alpha) - o(1).\end{aligned}$$

■

Remark. Given an arbitrary group G, projecting the walk from G to the Abelianisation $G^{\mathrm{ab}} = G/[G,G]$, which is Abelian, cannot increase the TV distance. Thus, $\mathrm{t}_0(k, |G^{\mathrm{ab}}|)$ is a lower bound on mixing for the projected walk on the G^{ab}, and hence for the original walk on G too.

2.7. *Upper bound on total-variation mixing*

It is often easier to control L_2 distances, rather than L_1 (i.e., TV). However, L_2 is sensitive to rare events, unlike TV. We use a 'modified L_2 calculation' to bound the TV: first, condition that W is 'typical'; then, use a standard L_2 calculation on the conditioned law. Let W' be an independent copy of W. Then, $S' := W' \cdot Z$ has the same law as S and is conditionally independent of S given Z.

Lemma 2.6. *For all $t \ge 0$ and all $\mathcal{W} \subseteq \mathbb{Z}^k$, the following inequalities hold:*

$$\begin{aligned} d_{G_k}(t) &= \|\mathbb{P}_{G_k}(S(t) \in \cdot) - \pi_G\|_{\mathrm{TV}} \\ &\le \|\mathbb{P}_{G_k}(S(t) \in \cdot \mid W(t) \in \mathcal{W}) - \pi_G\|_{\mathrm{TV}} + \mathbb{P}(W(t) \notin \mathcal{W}); \end{aligned}$$

and

$$4\mathbb{E}(\|\mathbb{P}_{G_k}(S(t) \in \cdot \mid W(t) \in \mathcal{W}) - \pi_G\|_{\mathrm{TV}}^2) \le n\mathbb{P}(S(t) = S'(t) \mid W(t), W'(t) \in \mathcal{W}) - 1.$$

We emphasise that d_{G_k} is a random variable, a function of $Z_1, \ldots, Z_k \sim^{\text{iid}} \mathrm{Unif}(G)$.

Proof. The first claim follows immediately from the triangle inequality. For the second, using the Cauchy–Schwarz inequality, we upper bound the TV distance of the conditioned law by its L_2 distance:

$$\begin{aligned} 4\|\mathbb{P}_{G_k}(S \in \cdot \mid W \in \mathcal{W}) - \pi_G\|_{\mathrm{TV}}^2 &\le n \sum_x \Big(\mathbb{P}_{G_k}(S = x \mid W \in \mathcal{W}) - \frac{1}{n}\Big)^2 \\ &= n \sum_x \mathbb{P}_{G_k}(S = x \mid W \in \mathcal{W})^2 - 1 \\ &= n \sum_x \mathbb{P}_{G_k}(S = S' = x \mid W, W' \in \mathcal{W}) - 1, \end{aligned}$$

since

$$S = W \cdot Z \quad \text{and} \quad S' = W' \cdot Z.$$

The claim follows by taking expectations over $[Z_1, \ldots, Z_k]$. ■

We now make the specific choice of the 'typical' set $\mathcal{W}$; we make a different choice for each $\alpha \in \mathbb{R}$. The collection $\{\mathcal{W}_\alpha\}_{\alpha \in \mathbb{R}}$ of sets will satisfy $\mathbb{P}(W(\mathrm{t}_\alpha) \notin \mathcal{W}_\alpha) \precsim \Psi(\alpha)$, using the CLT (Proposition 2.3); see Proposition 2.8. Recall that Ψ is the standard Gaussian tail. We show that the modified L_2 distance (given typicality) is $o(1)$ at t_α; see Proposition 2.9. Applying Lemma 2.6, we find that $d_{G_k}(\mathrm{t}_\alpha) \le \Psi(\alpha) + o(1)$ w.h.p. over Z. This matches the lower bound from Section 2.6.

By considering all $\alpha \in \mathbb{R}$, we find the shape of the cutoff. If we only desire the order of the window, then we need only consider the limit $\alpha \to \infty$; in this case, $\mathbb{P}(W(\mathrm{t}_\alpha) \notin \mathcal{W}_\alpha) \approx \Psi(\alpha) \approx 0$, which explains the use of the word 'typical' in describing $\mathcal{W}_\alpha$.

The typicality conditions will be a combination of 'local' (coordinate-wise) and 'global' ones.

Definition 2.7. For all $\alpha \in \mathbb{R}$, define the *local* and *global typicality conditions*, respectively:

$$\mathcal{W}_{\alpha,\mathrm{loc}} := \{w \in \mathbb{Z}^k \mid |w_i - E(W_1(\mathrm{t}_\alpha))| \le r_* \ \forall i = 1, \ldots, k\}, \quad r_* := \frac{1}{2} n^{1/k} (\log k)^2;$$

$$\mathcal{W}_{\alpha,\mathrm{glo}} := \{w \in \mathbb{Z}^k \mid \mathbb{P}(W(\mathrm{t}_\alpha) = w) \le n^{-1} e^{-\omega}\}.$$

Define $\mathcal{W}_\alpha := \mathcal{W}_{\alpha,\mathrm{loc}} \cap \mathcal{W}_{\alpha,\mathrm{glo}}$, and say that $w \in \mathbb{Z}^k$ is (α-)*typical* if $w \in \mathcal{W}_\alpha$.

The following propositions determine the probability that $W(\mathrm{t}_\alpha)$ lies in $\mathcal{W}_\alpha$, i.e., the probability of typicality.

Proposition 2.8a. *Let $X = (X_s)_{s \geq 1}$ be a rate-1 RW – either a SRW or a DRW – on $\mathbb{Z}$. Then,*

$$\mathbb{P}_0(|X_s - \mathbb{E}(X_s)| > r) \leq \frac{1}{k^{3/2}} \quad \textit{for all } r \geq r_* = \frac{1}{2} n^{1/k} (\log k)^2 \textit{ whenever } s \lesssim n^{2/k} \log k,$$

where the subscript 0 indicates that X starts from $X_0 = 0$. In particular, for all $\alpha \in \mathbb{R}$, we have

$$\mathbb{P}(W(\mathrm{t}_\alpha) \notin \mathcal{W}_{\alpha,\mathrm{loc}}) \leq k^{-1/2} = o(1).$$

Proof. The proof of the first part of this proposition follows from standard large deviation estimates on the RW on $\mathbb{Z}$, and the fact that $\mathrm{t}_\alpha \asymp k n^{2/k}$ for all $\alpha \in \mathbb{R}$, as stated in Proposition 2.2. The precise details are arduously technical. We defer them to [26, §C]. The second part follows immediately from the first part and the union bound over the k coordinates. ■

Proposition 2.8b. *For each $\alpha \in \mathbb{R}$, we have*

$$\mathbb{P}(W(\mathrm{t}_\alpha) \notin \mathcal{W}_{\alpha,\mathrm{glo}}) \to \Psi(\alpha).$$

Proof. This follows immediately from our CLT (Proposition 2.3). ■

Herein, we fix $\alpha \in \mathbb{R}$ and frequently suppress the time t_α from the notation, e.g., writing W for $W(\mathrm{t}_\alpha)$ or $\mathcal{W}$ for $\mathcal{W}_\alpha$. Let $V := W - W'$, so $\{W \cdot Z = W' \cdot Z\} = \{V \cdot Z = 0\}$. Write

$$D := D(\mathrm{t}_\alpha) := n\mathbb{P}(V(\mathrm{t}_\alpha) \cdot Z = 0 \mid \mathsf{typ}_\alpha) - 1,$$

where

$$\mathsf{typ} := \mathsf{typ}_\alpha := \{W(\mathrm{t}_\alpha), W'(\mathrm{t}_\alpha) \in \mathcal{W}_\alpha\}.$$

It remains to show that $D(\mathrm{t}_\alpha) = o(1)$ for all $\alpha \in \mathbb{R}$. Recall Hypothesis A, the crux of which is that

$$\frac{k-d-1}{k} - 2\frac{d \log\log k}{\log n} \geq 5\frac{k}{\log n} \quad \text{and} \quad k - d \gg 1.$$

For $r_1, \ldots, r_\ell \in \mathbb{Z} \setminus \{0\}$, we use the convention $\gcd(r_1, \ldots, r_\ell, 0) := \gcd(|r_1|, \ldots, |r_\ell|)$.

Proposition 2.9. *Suppose that (d, n, k) jointly satisfy Hypothesis A. (Recall that, implicitly, (d, n, k) is a sequence of triples of integers.) Write $\mathfrak{g} := \gcd(V_1, \ldots, V_k, n)$. Then, for all $\alpha \in \mathbb{R}$, we have*

$$0 \leq D(\mathrm{t}_\alpha) = \sum_{\gamma \in \mathbb{N}} \mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \cdot \frac{|G|}{|\gamma G|} - 1 = o(1).$$

Given this proposition, we can prove the upper bound in the main theorem, Theorem 2.5.

Proof of upper bound in Theorem 2.5 given Proposition 2.9. Fix $\alpha \in \mathbb{R}$ and consider the TV distance at time t_α. Apply the modified L_2 calculation, i.e., Lemma 2.6 and Definition 2.7, at time t_α: by Proposition 2.9, the modified L_2 distance (given typicality) is $o(1)$ in expectation; by Markov's inequality, it is thus $o(1)$ w.h.p. Proposition 2.8 says that typicality holds with probability $\Psi(t_\alpha)$ asymptotically. Combined, this all says that $d_{G_k}(t_\alpha) \le \Psi(\alpha) + o(1)$ w.h.p. ∎

It remains to prove Proposition 2.9, i.e., to bound the modified L_2 distance. The remainder of the section is dedicated to this goal. To do this, we are interested in the law of $V \cdot Z$.

Obviously, when $V = 0$, we have $V \cdot Z = 0$. The following auxiliary lemma controls this.

Lemma 2.10. *We have*

$$n\mathbb{P}(V = 0 \mid \mathsf{typ}) \le \frac{e^{-\omega}}{\mathbb{P}(\mathsf{typ})} \lesssim e^{-\omega} = o(1).$$

Proof. By direct calculation, since W and W' are independent copies,

$$\mathbb{P}(V = 0, \mathsf{typ}) = \mathbb{P}(W = W', W \in \mathcal{W}) = \sum_{w \in \mathcal{W}} \mathbb{P}(W = w)^2.$$

Recall global typicality: $\mathbb{P}(W = w) \le n^{-1} e^{-\omega}$ for all $w \in \mathcal{W}$. Thus,

$$n\mathbb{P}(V = 0 \mid \mathsf{typ}) \le n \sum_{w \in \mathcal{W}} \frac{\mathbb{P}(W = w)^2}{\mathbb{P}(\mathsf{typ})} \le \frac{e^{-\omega}}{\mathbb{P}(\mathsf{typ})}.$$ ∎

We now analyse $v \cdot Z = \sum_i v_i Z_i$ for $v \ne 0$. Sums of independent uniform random variables are uniform. Some simple technicalities take care of the fact that the v_i's need not be 1, or even equal.

Lemma 2.11 ([26, Lemma F.1]). *For all $v \in \mathbb{Z}^k$, we have*

$$v \cdot Z \sim \mathrm{Unif}(\gamma G), \quad \textit{where } \gamma := \gcd(v_1, \dots, v_k, n).$$

We now need to control $|\gamma G|$, since Lemma 2.11 implies that

$$\mathbb{P}(V \cdot Z = 0 \mid \mathsf{typ}) = \sum_{\gamma \in \mathbb{N}} \frac{\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ})}{|\gamma G|}, \quad \text{where } \mathfrak{g} := \gcd(V_1, \dots, V_k, n).$$

Lemma 2.12. *For all Abelian groups G and all $\gamma \in \mathbb{N}$, we have*

$$\frac{|G|}{|\gamma G|} \le \gamma^{d(G)}.$$

Proof. Decompose G as $\bigoplus_1^d \mathbb{Z}_{m_j}$ with $d = d(G)$ and some $m_1, \dots, m_d \in \mathbb{N}$. Then γG can be decomposed as $\bigoplus_1^d \gcd(\gamma, m_j)\mathbb{Z}_{m_j}$. Hence, $|\gamma G| = \prod_1^d (m_j / \gcd(\gamma, m_j)) \ge \prod_1^d (m_j/\gamma) = |G|/\gamma^d$. ∎

These lemmas combine to produce a simple, but key, corollary.

Corollary 2.13. *We have*

$$n\mathbb{P}(V \cdot Z = 0, V \neq 0 \mid \mathsf{typ}) \le \mathbb{E}(\mathfrak{g}^{d(G)} \mathbf{1}(V \neq 0) \mid \mathsf{typ}).$$

Proof. The conditioning does not affect Z, so the claim is immediate from the previous lemmas. ∎

We control this gcd coordinate-by-coordinate, using a crude divisibility bound.

Lemma 2.14. *Recall that $V = V(\mathrm{t}_\alpha)$. For all $\gamma \in \mathbb{N}$, we have*

$$\mathbb{P}(V_1 \in \gamma\mathbb{Z} \mid V_1 \neq 0) \le \frac{1}{\gamma} \quad \textit{and} \quad \mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \lesssim \Big(\frac{1}{\gamma} + \frac{2}{n^{1/k}}\Big)^k.$$

Proof. For both the SRW and DRW, the difference $V = W - W'$ is a rate-2 SRW on $\mathbb{Z}^k$. Hence, each coordinate is an independent rate-$2/k$ SRW on $\mathbb{Z}$, which is symmetric about 0.

It is easy to see that any non-increasing distribution on $\mathbb{N}$ can be written as a mixture of $\mathrm{Unif}(\{1, \ldots, Y\})$ distributions, for different $Y \in \mathbb{N}$. The map $m \mapsto \mathbb{P}(|V_1(2t/k)|) = m$: $\mathbb{N} \to [0, 1]$ is non-increasing for any $t \ge 0$. Hence, $|V_1|$ conditional on $V_1 \neq 0$ has such a distribution. Thus,

$$|V_1| = |V_1(\mathrm{t}_\alpha)| \sim \mathrm{Unif}\{1, \ldots, Y\} \quad \text{conditional on } V_1 \neq 0,$$

where Y has some distribution. Hence, we have

$$\mathbb{P}(V_1 \in \gamma\mathbb{Z} \mid V_1 \neq 0) = \mathbb{E}(\lfloor Y/\gamma \rfloor / Y) \le \frac{1}{\gamma}.$$

If the gcd $\mathfrak{g} = \gamma$, then $V_i \in \gamma\mathbb{Z}$ for all $i \in [k]$. By independence of coordinates, we then obtain

$$\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \le \frac{\mathbb{P}(\mathfrak{g} = \gamma)}{\mathbb{P}(\mathsf{typ})} \lesssim \mathbb{P}(V_1 \in \gamma\mathbb{Z})^k \le (\mathbb{P}(V_1 = 0) + \mathbb{P}(V_1 \in \gamma\mathbb{Z} \mid V_1 \neq 0))^k,$$

noting that $\mathbb{P}(\mathsf{typ}) \asymp 1$. Using Proposition 2.2 to argue that $\mathbb{P}(V_1 = 0) \le 2/n^{1/k}$, we deduce that

$$\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \lesssim \Big(\frac{2}{n^{1/k}} + \frac{1}{\gamma}\Big)^k.$$ ∎

From this, using the Hypothesis A, we can deduce that $\mathbb{E}(\mathfrak{g}^{d(G)} \mathbf{1}(V \neq 0) \mid \mathsf{typ}) = 1 + o(1)$.

Corollary 2.15. *Recall that time $t = \mathrm{t}_\alpha$. Given Hypothesis A, we have*

$$\mathbb{E}(\mathfrak{g}^{d(G)} \mathbf{1}(V \neq 0) \mid \mathsf{typ}) = 1 + o(1).$$

This proof is briefly deferred. First, we deduce Proposition 2.9 from the above results.

Proof of Proposition 2.9. By Lemma 2.10 and Corollaries 2.13 and 2.15, we have

$$\begin{aligned} n\mathbb{P}(V \cdot Z = 0 \mid \mathsf{typ}) &\le n\mathbb{P}(V = 0 \mid \mathsf{typ}) + n\mathbb{P}(V \cdot Z = 0, V \neq 0 \mid \mathsf{typ}) \\ &\le n\mathbb{P}(V = 0 \mid \mathsf{typ}) + \mathbb{E}(\mathfrak{g}^{d(G)} \mathbf{1}(V \neq 0) \mid \mathsf{typ}) = 1 + o(1). \quad \blacksquare \end{aligned}$$

We close the analysis of Approach #1 with the briefly-deferred proof of Corollary 2.15.

Proof of Corollary 2.15. Let $d := d(G)$. By local typicality, $\mathfrak{g} \le 2r_* = n^{1/k}(\log k)^2$ if $V \neq 0$. Thus,

$$\mathbb{E}(\mathfrak{g}^d \mathbf{1}(V \neq 0) \mid \mathsf{typ}) = \sum_{\gamma \in \mathbb{N}} \gamma^d \mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \le 1 + \sum_{\gamma=2}^{\lfloor n^{1/k}(\log k)^2 \rfloor} \gamma^d \, \mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}).$$

For $\gamma \ge 2$, we use Lemma 2.14. Let $\delta \in (0,1)$. For $2 \le \gamma \le \delta n^{1/k}$, we use the bound

$$\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \lesssim \Big(\frac{1}{\gamma} + \frac{2}{\gamma/\delta}\Big)^k = \frac{(1+2\delta)^k}{\gamma^k}.$$

For $\gamma \ge \delta n^{1/k}$, we use the slightly crude bound $(a+b)^k \le 2^k(a^k + b^k)$ for $a, b \ge 0$ to deduce that

$$\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \lesssim 2^k\Big(\frac{1}{\gamma^k} + \frac{2^k}{n}\Big) = \frac{2^k}{\gamma^k} + \frac{4^k}{n}.$$

Dividing the appropriate sum over γ into two parts according to whether or not $\gamma \le \delta n^{1/k}$ and using the above inequalities, elementary algebraic manipulations can be used to deduce that

$$\begin{aligned} \mathbb{E}(\mathfrak{g}^d \, \mathbf{1}(V \neq 0) \mid \mathsf{typ}) - 1 &\lesssim e^{2\delta k} 2^{d+1-k} + 2^k \delta^{d+1-k} n^{(d+1-k)/k} \\ &\quad + \frac{4^k n^{(d+1)/k} (\log k)^{2(d+1)}}{n}. \end{aligned}$$

This is $o(1)$, by the conditions of Hypothesis A, as we now explain. Write

$$\eta := \frac{k-d-1}{k} \in (0,1).$$

- We wish to choose δ as large as possible, but with the first term $o(1)$: set $\delta := \eta/4$.
- Hypothesis A implies that $\eta \ge 4k/\log n$, which shows that the second term is $o(1)$.
- The inequality in Hypothesis A is designed precisely so that the final term is $o(1)$. ■

Remark. We have always assumed that $k - d(G) \gg 1$. Our analysis does apply if $M := k - d(G) \ge 2$ is fixed (i.e., not diverging) too. Then, however, it is not necessarily the case that the group is generated w.h.p. – e.g., if $G = \mathbb{Z}_2^d$, then it is not. Our analysis shows that the mixing time is of order $\mathfrak{t}_0$ with probability bounded away from 0, and approaching 1 as M grows.

3. Total variation cutoff: Approach #2

Recall that Theorem A is established via two distinct approaches. In the previous section, we used one approach to deal with the case that k is 'not too large'. Here, we use a new approach to deal with the case that k is 'not too small'. The main result of the section is Theorem 3.7. The outline of the section is roughly the same as that of the previous one.

- Section 3.1 discusses the new, refined entropic methodology.
- Section 3.2 defines the new *entropic times*.
- Section 3.3 states bounds on the growth rate of the entropy and concentration.
- Section 3.4 states precisely the main theorem of the section.
- Section 3.5 outlines the differences between this argument and the previous approach.
- Section 3.6 is devoted to the lower bound.
- Section 3.7 is devoted to the upper bound.

3.1. Entropic times: New methodology and definition

The underlying principles of the method used in this section (Section 3) are the same as those of the previous one (Section 2), just adjusted slightly to deal with the cases not previously covered.

We first discuss where the previous approach broke down and how we might fix it. The primary issue was when $d(G)$ was very large. For example, consider $\mathbb{Z}_2^d$. All elements are of order 2, so instead of looking at W, a RW on $\mathbb{Z}$, we could equally have taken W mod 2. The entropy of $W_1(t)$ mod 2 is significantly smaller than that of $W_1(t)$ once $t/k \gtrsim 1$. This suggests a longer mixing time.

Now, $V \cdot Z \sim \operatorname{Unif}(\gamma G)$ when $\gcd(V_1, \ldots, V_k, n) = \gamma$. This motivates defining τ_γ to be the time at which the entropy of W_1 mod γ is $\log|G/\gamma G|$, and proposing $\tau_* := \max_{\gamma \in \mathbb{N}} \tau_\gamma$ as the upper bound.

3.2. Entropic times: Definition and concentration

In this section, we refine the definition of *entropic times*. The concept is highly analogous to that of the previous section. There is, thus, some overlap in both verbal description and notation. We have been careful, though, to set it up as not to cause confusion: we always use indices such as $\gamma \in \mathbb{N}$ in the new entropic times $\tau_0(\gamma, N)$ or τ_γ below, whilst previously we used $\alpha \in \mathbb{R}$ in $\mathfrak{t}_\alpha$.

We now define precisely the (updated) notion of *entropic times*. Let $W = (W_i(t) \mid i \in [k],\ t \geq 0)$ be a RW on $\mathbb{Z}^k$, counting the uses of the generators, as in the previous sections. This can be either a SRW on $\mathbb{Z}^k$ or DRW on $\mathbb{Z}_+^k$. As before, $S(t) = W(t) \cdot Z$. For $\gamma \in \mathbb{N} \cup \{\infty\}$, define W_γ via

$$W_{\gamma,i}(t) := W_i(t) \bmod \gamma \ \text{ for } \gamma \in \mathbb{N} \quad \text{and} \quad W_\infty := W.$$

Then, W_γ is a RW on $\mathbb{Z}_\gamma^k$. So, $W_{\gamma,i} := (W_{\gamma,i}(t))_{t \ge 0}$ forms an iid sequence of rate-$1/k$ RWs on $\mathbb{Z}_\gamma$. As before, 'mod ∞' has no effect: $w = w \bmod \infty$ for all $w \in \mathbb{Z}^k$, and $G/\infty G = G$ as $\infty G = \{\mathsf{id}\}$.

Write $\mu_{\gamma,t}$ (resp. $\nu_{\gamma,s}$) for the law of $W_\gamma(t)$ (resp. $W_{\gamma,1}(sk)$); so, $\mu_{\gamma,t} = \nu_{\gamma,t/k}^{\otimes k}$. Define

$$Q_\gamma(t) := -\log \mu_{\gamma,t}(W_\gamma(t)) \quad \text{and} \quad Q_{\gamma,i}(t) := -\log \nu_{\gamma,t/k}(W_{\gamma,i}(t)).$$

So, $Q_{\gamma,i}$ forms an iid sequence over $i \in [k]$; also,

$$Q_\gamma(t) = \sum_{k}^{i=1} Q_{\gamma,i}(t), \quad h_\gamma(t) := \mathbb{E}(Q_\gamma(t)) \quad \text{and} \quad H_\gamma(s) := \mathbb{E}(Q_{\gamma,1}(sk)).$$

So, $h_\gamma(t)$ and $H_\gamma(s)$ are the entropies of $W_\gamma(t)$ and $W_{\gamma,1}(sk)$, respectively. Note that $h_\gamma(t) = kH_\gamma(t/k)$ and that $h_\gamma\colon [0,\infty) \to [0, \log(\gamma^k))$ is a strictly increasing bijection.

Some of these expressions, such as h_γ, depend on k; we usually suppress this from the notation.

Definition 3.1. For $N < \gamma^k$, define the *entropic time*

$$\tau_0(\gamma, N) := h_\gamma^{-1}(\log N) \quad \text{and} \quad \sigma_0(\gamma, N) := \frac{\tau_0(\gamma, N)}{k} = H_\gamma^{-1}\Big(\frac{\log N}{k}\Big).$$

We are interested primarily in $N := |G/\gamma G|$ for an Abelian group G: set

$$\tau_* := \tau_*(k, G) := \max_{\gamma \in \mathbb{N} \cup \{\infty\}} \tau_0(\gamma, |G/\gamma G|).$$

This τ_* is the same as defined in the introduction; see Definition A. Comparing with notation in Approach #1, $\tau_0(\infty, |G|) = \mathfrak{t}_0$; see Definition 2.1. We establish *cutoff* here, not the *profile* as well.

Recall that $\infty G = |G|G = \{\mathsf{id}\}$ and $1G = G$. So, the maximum is achieved at some $\gamma \in [2, n]$. Below, for brevity, we write '$\gamma \ge 2$' to mean '$\gamma \in \mathbb{N} \cup \{\infty\} \setminus \{1\}$'.

Our next result determines the asymptotics of τ_*.

Proposition 3.2a. *Suppose that* $1 \ll k \lesssim \log|G|$. *The following hold:*

$$\text{if } k - d(G) \asymp k, \text{ then } \tau_*(k, G) \asymp k|G|^{2/k};$$
$$\text{if } k - d(G) \ge 1, \text{ then } k|G|^{2/k} \lesssim \tau_*(k, G) \lesssim k|G|^{2/k} \log k.$$

Proposition 3.2b. *Suppose that* $d(G) \ll \log|G|$ *and* $k - d(G) \asymp k \gg 1$. *Then,*

$$\tau_*(k, G) \eqsim \tau_0(\infty, G) = \mathfrak{t}_0.$$

As with earlier results, the rigorous proofs are technical. They boil down to comparing the RW on $\mathbb{Z}_\gamma$ with one on $\mathbb{Z}$. Precise details are deferred to [26, Propositions B.17 and B.18].

In Section 3.6, we show that $\tau_0(\gamma, |G/\gamma G|)$ is a lower bound on mixing for all γ, for all Z. Throughout this section, we work under the assumption that $k \lesssim \log|G|$. (Recall from Section 1.3 that cutoff had already been established for all Abelian groups when $k \gg \log|G|$.) As a result of this, taking $\gamma := |G|$, we see that the mixing time is at least of order k. Indeed, $|G|G = \{\mathsf{id}\}$, so the target entropy per coordinate is $\log|G|/k \gtrsim 1$, so each coordinate needs to be run for time $t/k \gtrsim 1$. Hence, there exists a $\varsigma > 0$ so that the mixing time is at least $2\varsigma k$. This holds for all Z, not just w.h.p. over Z.

Unfortunately, the γ with $\tau_0(\gamma, |G/\gamma G|) \le \varsigma k$ cause some technical difficulties. For this reason, we take the maximum with ςk in the definition of the 'adjusted' entropic time τ_γ below. Crucially,

$$\max_{\gamma \in \mathbb{N}} (\tau_0(\gamma, |G/\gamma G|) \vee \varsigma k) = \max_{\gamma \in \mathbb{N}} \tau_0(\gamma, |G/\gamma G|) \vee \varsigma k = \tau_*.$$

We emphasise that this last adjustment is purely technical. On the other hand, the entropic times in Definition 3.1 capture properties of the group to which those in Definition 2.1 are oblivious.

Definition 3.3. For $\gamma \ge 2$ and $s \ge 0$, define the *(adjusted) entropic time* and *relative entropy* via

$$\sigma_\gamma := \sigma_0(\gamma, |G/\gamma G|) \vee \varsigma, \quad \tau_\gamma := \sigma_\gamma k \quad \text{and} \quad R_\gamma(s) := \log \gamma - H_\gamma(s).$$

The maximal entropy of a distribution on $\mathbb{Z}_\gamma$ is $\log \gamma$, obtained uniquely by the uniform distribution $\mathrm{Unif}(\mathbb{Z}_\gamma)$. Hence, $R_\gamma(s) \to 0$ as $s \to \infty$, since the RW converges to $\mathrm{Unif}(\mathbb{Z}_\gamma)$.

3.3. Entropic times: Entropy growth rate and concentration

We determine the rate of change of the entropy around the entropic time and establish concentration estimates on the 'random entropy' Q_γ at a time shortly after the entropic time.

The first lemma controls the rate of change of the entropy near the entropic time.

Lemma 3.4. *There exists a continuous function* $\overline{c}\colon (0,1) \to (0,1)$ *so that, for all* $\gamma \ge 2$, *all* $\xi \in (-1,1) \setminus \{0\}$ *and all* $s \ge \varsigma$, *we have*

$$|H_\gamma(s(1+\xi)) - H_\gamma(s)| \ge 2\overline{c}_{|\xi|}(R_\gamma(s) \wedge 1).$$

Outline of proof. If $s \asymp 1$, then it is easy as all terms are order 1. If $s \ll \gamma^2$, then the fact that the RW is on $\mathbb{Z}_\gamma$, not $\mathbb{Z}$, is not significant. The entropy is thus approximately $(1/2)\log(2\pi e s)$, i.e., that of $N(0,s)$, if also $s \gg 1$. The case $s \gtrsim \gamma^2$ follows from standard (modified) log-Sobolev arguments.

Making this proof rigorous is technical. Doing so is deferred to [26, Lemma B.20]. ∎

Recall that $\tau_* = \max_{\gamma \in \mathbb{N}} \tau_\gamma$. Abbreviate $d = d(G)$. For $\gamma \in \mathbb{N}$, write

$$\zeta_\gamma := \frac{1}{k}(k - d(G)) \log \gamma.$$

Proposition 3.5. *Assume that $k > d$. There exists a continuous function $c\colon (0,1) \to (0,1)$ so that, for all $\gamma \geq 2$ and all $\varepsilon \in (0,1)$, the following hold:*

$$\mathbb{P}(Q_\gamma(\tau_*(1+\varepsilon)) \leq \log|G/\gamma G| + c_\varepsilon(\zeta_\gamma \wedge 1)k) \leq e^{-c_\varepsilon(\zeta_\gamma \wedge 1)k},$$
$$\mathbb{P}(Q_\gamma(t(1-\varepsilon)) \geq \log|G/\gamma G| - c_\varepsilon(\zeta_\gamma \wedge 1)k) = o(1) \quad \textit{for all } t \leq \tau_\gamma.$$

Outline of proof. Recall that $Q_\gamma(t) = \sum_1^k Q_{\gamma,i}(t)$ is a sum of iid terms, each of mean $H_\gamma(t/k)/k$. By the entropy growth rate (Lemma 3.4), for any $\xi \in (-1,1) \setminus \{0\}$, the change in entropy between times s and $(1+\xi)s$ is of order $R_\gamma(s) \wedge 1$, with implicit constant depending on $|\xi|$. Taking $s := \sigma_0(\gamma, |G/\gamma G|)$, recalling that $|G/\gamma G| \leq \gamma^{d(G)}$ by Lemma 2.12, gives

$$R_\gamma(s) = \log\gamma - H_\gamma(s) = \log\gamma - \frac{\log|G/\gamma G|}{k} \geq \frac{1}{k}(k - d(G))\log\gamma = \zeta_\gamma.$$

We are interested in the times σ_γ, not $\sigma_0(\gamma, |G/\gamma G|)$; this is only a minor technical complication. The quantitative concentration estimate requires first deterministically bounding $\mathbb{E}(Q_{1,\gamma}) - Q_{1,\gamma}$ from above. A (one-sided) variant of Bernstein's inequality for a sum of iid, deterministically bounded random variables, is then applied. The non-quantitative part is just an application of the Chebyshev inequality, once the variance $\mathbb{V}\mathrm{ar}(Q_{\gamma,1}(sk))$ has been uniformly bounded over $s \geq \varsigma$. Again, making argument this rigorous is technical. It is deferred to [26, Proposition B.21]. ∎

3.4. *Precise statement and remarks*

In this subsection, we state precisely the main theorem of the section. There are some simple conditions on k, in terms of $d(G)$ and $|G|$, needed for the upper bound.

Hypothesis B. *The sequence $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis B if the following hold:*

$$\limsup_{N\to\infty} \frac{k_N}{\log|G_N|} < \infty, \quad \liminf_{N\to\infty}\{k_N - d(G_N)\} = \infty, \quad \liminf_{N\to\infty} \frac{k_N}{\log(|\mathcal{H}_N|+1)} = \infty,$$

where

$$\mathcal{H}_N := \{\gamma G_N \mid \gamma \wr |G_N| \text{ and } \gamma \in [2, n_{*,N}]\} \quad \textit{and} \quad n_{*,N} := \lfloor |G_N|^{1/k_N}(\log k_N)^2 \rfloor,$$

where the notation $a \wr b$ means that a divides b, i.e., $b \in a\mathbb{Z}$, for $a, b \in \mathbb{N}$.

Remark 3.6. If $k \gg \sqrt{\log n}$, then $k \gg \log(|\mathcal{H}|+1)$, since $|\mathcal{H}| \leq n_* \leq n^{1/k}(\log k)^2$.

Throughout the proofs, we suppress the subscript-N, e.g., writing k or n, considering sequences implicitly. Recall that we abbreviate the TV distance from uniformity at time t as

$$d_{G_k,N}(t) = \|\mathbb{P}_{G_N([Z_1,\ldots,Z_{k_N}])}(S(t) \in \cdot) - \pi_{G_N}\|_{\mathrm{TV}},$$

where

$$Z_1, \ldots, Z_{k_N} \sim^{\mathrm{iid}} \mathrm{Unif}(G_N).$$

We now state the main theorem of this section. Recall that

$$\tau_* = \max_\gamma \tau_0(\gamma, |G/\gamma G|) = \max_\gamma \tau_\gamma.$$

Theorem 3.7. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ of finite, Abelian groups; for each $N \in \mathbb{N}$, define $Z_{(N)} := [Z_1, \ldots, Z_{k_N}]$ by drawing $Z_1, \ldots, Z_{k_N} \sim^{\text{iid}} \text{Unif}(G_N)$. Suppose that the sequence $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis B.*

Let $c \in (-1, 1) \setminus \{0\}$. Then,

$$d_{G_k,N}((1+c)\tau_*(k_N, G_N)) \to^{\mathbb{P}} \mathbf{1}(c < 0) \quad \textit{(in probability) as } N \to \infty.$$

That is, w.h.p., there is TV cutoff at $\tau_* = \max_\gamma \tau_0(\gamma, |G/\gamma G|)$. Moreover, the implicit lower bound on the TV distance holds deterministically, i.e., for all choices of generators.

3.5. *Outline of proof*

The general outline is analogous to that before; see Section 2.5. That approach failed once either d or k became too large or $k - d$ became too small. We outline here the ideas used to cover these cases.

For the lower bound, we project the walk from G to $G/\gamma G$. This cannot increase the TV distance. The same argument shows that $\tau_0(\gamma, |G/\gamma G|)$ is a lower bound for all γ.

For the upper bound, fundamentally, we still wish to bound the same expression

$$D(t) = \sum_{\gamma\in\mathbb{N}} \mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \cdot \frac{|G|}{|\gamma G|} - 1;$$

see Propositions 2.9 and 3.13. If $\mathfrak{g} = \gamma$, then $W \equiv W' \mod \gamma$. But $W_\gamma := W \mod \gamma$ and $W'_\gamma := W' \mod \gamma$ are simply RWs on $\mathbb{Z}_\gamma^k$. So, the same argument as in Lemma 2.10 gives

$$\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \le \mathbb{P}(W_\gamma = W'_\gamma \mid \mathsf{typ}) \ll \frac{1}{|G/\gamma G|} = \frac{|\gamma G|}{|G|} \quad \text{when } t \ge \tau_0(\gamma, |G/\gamma G|).$$

Thus, $D(t) = 1 + o(1)$ when $t \ge \tau_0(\gamma, |G/\gamma G|)$ for all $\gamma \ge 2$. Hence, the proposed upper bound of

$$\tau_* = \max_{\gamma\ge 2} \tau_0(\gamma, |G/\gamma G|).$$

The adjusted entropic times τ_γ are only introduced to alleviate a technical problem.

3.6. *Lower bound on total-variation mixing*

The idea is to quotient out by γG, and show that the walk on this quotient is not mixed at time $(1-\varepsilon)\tau_0(\gamma, |G/\gamma G|)$, as in Section 2.6. Hence, the original walk is not mixed on G either. This is achieved via the entropy growth rate and variance bounds detailed in Proposition 3.5.

Proof of lower bound in Theorem 3.7. As in Section 2.6, for this proof only, to emphasise that Z is fixed, not being averaged over, we add a subscript-Z to the probabilities involving Z: $\mathbb{P}_Z(S(t) \in \cdot)$.

Fix $\varepsilon \in (0, 1)$ and set $t := (1-\varepsilon)\tau_0(\gamma, |G/\gamma G|)$. Recall that

$$\zeta_\gamma = \frac{k - d(G)}{k} \log \gamma.$$

Then, $\mathbb{P}(\mathcal{E}) = 1 - o(1)$, where

$$\mathcal{E} := \left\{ \mu_{\gamma,t}(W_\gamma(t)) \geq \frac{\delta_\gamma^{-1}}{|G/\gamma G|} \right\} \quad \text{and} \quad \delta_\gamma := e^{-c_\varepsilon(\zeta_\gamma \wedge 1)k},$$

by the entropy concentration (Proposition 3.5). Also, $|G/\gamma G| \leq \gamma^{d(G)}$ by Lemma 2.12, and so

$$R_\gamma(\sigma_0(\gamma, |G/\gamma G|)) = \log \gamma - \frac{\log|G/\gamma G|}{k} \geq \frac{1}{k}(k - d(G)) \log \gamma = \zeta_\gamma.$$

Also, $k - d(G) \gg 1$, by assumption. Thus, $\delta_\gamma = o(1)$ uniformly in γ. Consider the set

$$E := \left\{ x \in G/\gamma G \;\middle|\; \exists w \in \mathbb{Z}_\gamma^k \text{ st } \mu_{\gamma,t}(w) \geq \frac{\delta_\gamma^{-1}}{|G/\gamma G|} \text{ and } x = (w \cdot Z)\gamma G \right\}.$$

Define S_γ to be the projection of $S = W_\infty \cdot Z$ to $G/\gamma G$. Then, $\mathbb{P}_Z(S_\gamma(t) \in E \mid \mathcal{E}) = 1$ since W_∞ generates S. Every element $x \in E$ can be realised as $x = w_x \cdot Z$ for some $w_x \in \mathbb{Z}_\gamma^k$ with $\mu_{\gamma,t}(w_x) \geq \delta_\gamma^{-1}/|G/\gamma G|$, by definition of E. Hence, for all $x \in E$, we have

$$\mathbb{P}_Z(S_\gamma(t) = x) \geq \mathbb{P}(W_\gamma(t) = w_x) = \mu_{\gamma,t}(w_x) \geq \frac{\delta_\gamma^{-1}}{|G/\gamma G|},$$

recalling that S_γ lives in the quotient $G/\gamma G$. Taking the sum over all $x \in E$, we deduce that

$$1 \geq \sum_{x \in E} \mathbb{P}_Z(S_\gamma(t) = x) \geq |E| \cdot \frac{\delta_\gamma^{-1}}{|G/\gamma G|},$$

and hence

$$\frac{|E|}{|G/\gamma G|} \leq \delta_\gamma = o(1).$$

Projecting onto $G/\gamma G$ cannot increase the TV distance, so

$$\begin{aligned} \|\mathbb{P}_{G_k}(S(t) \in \cdot) - \pi_G\|_{\text{TV}} &\geq \mathbb{P}_Z(S_\gamma(t) \in E) - \pi_{G/\gamma G}(E) \\ &\geq \mathbb{P}(\mathcal{E}) - \frac{|E|}{|G/\gamma G|} = 1 - o(1). \end{aligned}$$

Finally, recall that $\max_{\gamma \in \mathbb{N}} \tau_\gamma = \max_{\gamma \in \mathbb{N}} \tau_0(\gamma, |G/\gamma G|) = \tau_*$. This completes the proof. ∎

3.7. Upper bound on total-variation mixing

We use the same 'modified L_2 calculation' as in Section 2.7, conditioning on 'typicality'; see Lemma 2.6.

Abbreviate $d = d(G)$ and recall that $\zeta_\gamma = ((k-d)/k)\log\gamma$; set $\widehat{\zeta}_\gamma := \zeta_\gamma \wedge 1$. Fix $\varepsilon > 0$. The following depend on ε; we suppress this. Set $t := \tau_*(1+\varepsilon)$. Recall the constant $c = c_\varepsilon > 0$ from Proposition 3.5.

Definition 3.8. Set $t = \tau_*(1+\varepsilon)$. Define *global* typicality sets for $\gamma \in \mathbb{N} \cup \{\infty\}$ by

$$\mathcal{W}_{\gamma,\text{glo}} := \Big\{ w \in \mathbb{Z}_\gamma^k \;\Big|\; \mathbb{P}(W_\gamma(t) = w) \le \frac{\delta_\gamma}{|G/\gamma G|} \Big\}, \quad \text{where } \delta_\gamma := e^{-c\widehat{\zeta}_\gamma k},$$

using the convention $\widehat{\zeta}_\infty = \zeta_\infty \wedge 1 = 1$, so $\delta_\infty := e^{-ck}$. Define the *local* typicality set by

$$\mathcal{W}_{\text{loc}} := \{ w \in \mathbb{Z}_\infty^k \mid |w_i - \mathbb{E}(W_{\infty,i}(t))| \le r_* \ \forall i \in [k] \}, \quad \text{where } r_* := \frac{1}{2}|G|^{1/k}(\log k)^2.$$

When W' is an independent copy of W, define *typicality* by

$$\mathsf{typ} := \{ W_\infty(t), W'_\infty(t) \in \mathcal{W}_{\text{loc}} \} \cap \Big(\bigcap_{\gamma \in \Gamma} \{ W_\gamma(t), W'_\gamma(t) \in \mathcal{W}_{\gamma,\text{glo}} \} \Big),$$

where Γ is a specific subset of $[2, |G|]$ to be defined below in Definition 3.11.

We are going to use a union bound over $\gamma \in \Gamma$, so desire control on $\sum_{\gamma\in\Gamma} \delta_\gamma$.

Lemma 3.9. *For all* $\Gamma \subseteq \mathbb{N} \setminus \{1\}$, *we have*

$$\sum_{\gamma\in\Gamma} \delta_\gamma \le \delta_\infty |\Gamma| + 2^{-c(k-d)+1} = \delta_\infty |\Gamma| + o(1).$$

Proof. Since $\min \Gamma \ge 2$ and $k - d \gg 1$, we have $\sum_{\gamma\in\Gamma} \gamma^{-c(k-d)} \le \sum_{\gamma \ge 2} \gamma^{-c(k-d)} \le 2^{-c(k-d)+1}$. So,

$$\sum_{\gamma\in\Gamma} \delta_\gamma \le \sum_{\gamma\in\Gamma} (e^{-ck} + e^{-c\zeta_\gamma k}) = e^{-ck}|\Gamma| + \sum_{\gamma\in\Gamma} \gamma^{-c(k-d)} \le \delta_\infty|\Gamma| + 2^{-c(k-d)+1}. \quad \blacksquare$$

Proposition 3.10. *For all* $\varepsilon > 0$ *and any subset* $\Gamma \subseteq \mathbb{N} \setminus \{1\}$, *we have*

$$\mathbb{P}(\mathsf{typ}) \ge 1 - \delta_\infty |\Gamma| - o(1).$$

Proof. Suppress the time-dependence from the notation: e.g., write W_γ for $W_\gamma(t)$ and Q_γ for $Q_\gamma(t)$.

We consider global typicality first. Observe that

$$Q_\gamma = -\log \mu_\gamma(W_\gamma) \ge \log|G/\gamma G| + c\widehat{\zeta}_\gamma k \quad \text{if and only if} \quad \mu_\gamma(W_\gamma) \le \frac{e^{-c\widehat{\zeta}_\gamma k}}{|G/\gamma G|}.$$

Hence, recalling that $\delta_\gamma = e^{-c\hat{\zeta}_\gamma k}$ with $\hat{\zeta}_\gamma = \zeta_\gamma \wedge 1$, by Proposition 3.5, we have

$$\mathbb{P}(W_\gamma \notin \mathcal{W}_{\gamma,\mathrm{glo}}) \le \delta_\gamma, \quad \text{and hence } \mathbb{P}\Big(\bigcap_{\gamma\in\Gamma}\{W_\gamma \in \mathcal{W}_{\gamma,\mathrm{glo}}\}\Big) \ge 1 - \sum_{\gamma\in\Gamma}\delta_\gamma,$$

by the union bound. Recall that $\zeta_\gamma = (1/k)(k-d)\log\gamma$. Applying Lemma 3.9, we deduce that

$$\mathbb{P}\Big(\bigcap_{\gamma\in\Gamma}\{W_\gamma \in \mathcal{W}_{\gamma,\mathrm{glo}}\}\Big) \ge 1 - \delta_\infty|\Gamma| - o(1).$$

We turn to local typicality. Proposition 3.2a gives $t/k \le |G|^{2/k}\log k$. So, Proposition 2.8a gives

$$\mathbb{P}\Big(\bigcap_i\{|W_{\infty,i} - \mathbb{E}(W_{\infty,i})| \le r_*\}\Big) = 1 - o(1), \quad \text{and hence } \mathbb{P}(W_\infty \in \mathcal{W}_{\mathrm{loc}}) = 1 - o(1).$$

The claim follows by combining local and global typicality and applying the union bound. ∎

We now choose the set Γ, to make sense of typicality. Recall that $a \wr b$ means that a divides b.

Definition 3.11. Define $\Delta := \{\gamma \in [2, n_*] \mid \gamma \wr n\}$, where $n_* = \lfloor 2r_* \rfloor$. Recall that

$$\mathcal{H} := \{\gamma G \mid \gamma \in \Delta,\ \gamma G \ne G\} = \{H \mid H = \gamma G \ne G \text{ for some } \gamma \in \Delta\}.$$

Given $H \in \mathcal{H}$, write $\Gamma_H := \{\gamma \in \Delta \mid H = \gamma G\}$ and denote by γ_H the minimal $\gamma \wr n$ with $H = \gamma G$, i.e., $\gamma_H := \inf \Gamma_H$. Finally, define $\Gamma := \{\gamma_H \mid H \in \mathcal{H}\} \cup \{n\}$; so, $\Gamma \subseteq \Delta \cup \{n\} \subseteq [2, n_*] \cup \{n\}$.

The following lemma, whose proof is deferred to the end of this subsection, is also needed.

Lemma 3.12. *For all $H \in \mathcal{H}$ and all $\gamma \in \Gamma_H$, we have $\gamma_H \wr \gamma$.*

Recall that $\tau_* = \max_\gamma \tau_\gamma$. In analogy with Section 2.7 and Proposition 2.9, write

$$D := D(t) := n\mathbb{P}(V_\infty(t)\cdot Z = 0 \mid \mathsf{typ}) - 1.$$

Proposition 3.13. *Write $\mathfrak{g} := \gcd(V_{\infty,1}, \dots, V_{\infty,k}, n)$. Recall that $\varepsilon > 0$ and $t = \tau_*(1+\varepsilon)$. We have*

$$0 \le D(t(1+\varepsilon)) = \sum_{\gamma\in\mathbb{N}} \mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \cdot \frac{|G|}{|\gamma G|} - 1 \le \frac{\delta_\infty|\Gamma| + o(1)}{\mathbb{P}(\mathsf{typ})}.$$

The conditions of Hypothesis B imply immediately that this last term is $o(1)$, since $\delta_\infty = e^{-ck}$.

It is straightforward to deduce the upper bound on mixing from Propositions 3.10 and 3.13.

Proof of upper bound in Theorem 3.7. We use a modified L_2 calculation at time $(1+\varepsilon) \times \max_\gamma \tau_\gamma$.

- Condition that W satisfies typicality; see Definition 3.8 and Proposition 3.10.
- Perform the standard TV–L_2 bound on $S = W \cdot Z$ conditional that W is typical; cf. Lemma 2.6.
- Upper bound the expected L_2 distance by $(\delta_\infty |\Gamma| + o(1))/\mathbb{P}(\mathsf{typ})$; see Proposition 3.13.
- This gives an upper bound on the expected TV distance of $(\delta_\infty |\Gamma| + o(1))/\mathbb{P}(\mathsf{typ}) + \mathbb{P}(\mathsf{typ}^c)$.
- Clearly, $|\Gamma| \le |\mathscr{H}| + 1$. So, $k \gg \log(|\mathscr{H}| + 1)$ implies $\delta_\infty |\Gamma| \le \delta_\infty(|\mathscr{H}| + 1) \ll 1$, as $\delta_\infty = e^{-ck}$. Thus, $\mathbb{P}(\mathsf{typ}) = 1 - o(1)$ by Proposition 3.10. Hence, the expected TV distance is $o(1)$.
- This means that the TV distance is $o(1)$ w.h.p., by Markov's inequality. ■

We now prove Proposition 3.13. All terms are evaluated at $t = \tau_*(1+\varepsilon)$, and this is suppressed.

Proof of Proposition 3.13. Write $V_\infty := W_\infty - W'_\infty$ and $\mathfrak{g} := \gcd(V_{\infty,1}, \dots, V_{\infty,k}, n)$. If $\mathfrak{g} = \gamma$, which must have $\gamma \wr n$ as the gcd is with n, then $V_\infty \cdot Z \sim \mathrm{Unif}(\gamma G)$ by Lemma 2.11. Then,

$$D = n\mathbb{P}(V_\infty \cdot Z = 0 \mid \mathsf{typ}) - 1 = |G| \sum_{\gamma \wr n} \frac{\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ})}{|\gamma G|} - 1.$$

We consider various cases. First, combining together all γ such that $\gamma G = G$, we upper bound

$$|G| \frac{\mathbb{P}(\mathfrak{g} \in \{\gamma \mid \gamma G = G\})}{|\gamma G|} \le 1.$$

If $V_\infty = 0$ in $\mathbb{Z}^k$, then $\mathfrak{g} = \gamma = n$, which gives $\gamma G = \{\mathsf{id}\}$; using the definition of typicality,

$$|G| \frac{\mathbb{P}(V_\infty = 0 \mid \mathsf{typ})}{|\gamma G|} = |G| \mathbb{E}(\mathbb{P}(W_\infty = W'_\infty \mid W'_\infty, \mathsf{typ}) \mid \mathsf{typ}) \le \frac{\delta_\infty}{\mathbb{P}(\mathsf{typ})};$$

cf. Lemma 2.10. If $V_\infty \ne 0$, then, given (local) typicality, $\mathfrak{g} \le n_* = \lfloor 2r_* \rfloor$.

It remains to study $\gamma \in \Delta$. By Lemma 3.12, for any $H \in \mathscr{H}$, we have

$$\{V_\gamma = 0 \text{ for some } \gamma \in \Gamma_H\} \subseteq \{V_{\gamma_H} = 0\}.$$

(Recall that $V_\gamma \in \mathbb{Z}_\gamma^k$ for each γ.) This collapses the sum over all $\gamma \in \Gamma_H$ into the single term γ_H:

$$\sum_{\gamma \in \Gamma_H} \frac{\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ})}{|\gamma G|} = \frac{\mathbb{P}(\bigcup_{\gamma \in \Gamma_H} \{\mathfrak{g} = \gamma\} \mid \mathsf{typ})}{|H|}$$
$$\le \frac{\mathbb{P}(V_\gamma = 0 \text{ for some } \gamma \in \Gamma_H)}{|H|} \le \frac{\mathbb{P}(V_{\gamma_H} = 0 \mid \mathsf{typ})}{|H|} \le \frac{\delta_{\gamma_H}/|G|}{\mathbb{P}(\mathsf{typ})},$$

using typicality for the final inequality. We decompose $\sum_{\gamma \in \Delta}$ into $\sum_{H \in \mathcal{H}} \sum_{\gamma \in \Gamma_H}$:

$$|G| \sum_{\gamma \in \Delta} \frac{\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ})}{|\gamma G|} = |G| \sum_{H \in \mathcal{H}} \sum_{\gamma \in \Gamma_H} \frac{\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ})}{|\gamma G|} \le \sum_{H \in \mathcal{H}} \frac{\delta_{\gamma H}}{\mathbb{P}(\mathsf{typ})}.$$

Combining all parts and using Lemma 3.9, we deduce the proposition

$$0 \le n\mathbb{P}(V \cdot Z = 0 \mid \mathsf{typ}) - 1 = |G| \sum_{\gamma \wr n} \frac{\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ})}{|\gamma G|} - 1 \le \frac{\delta_\infty |\Gamma| + o(1)}{\mathbb{P}(\mathsf{typ})}. \quad \blacksquare$$

It remains to give the deferred proof Lemma 3.12. Recall that $a \wr b$ means that a divides b.

Proof of Lemma 3.12. Decompose $G = \bigoplus_1^r \mathbb{Z}_{m_j}$ arbitrarily. Fix $\beta \in \Gamma_H$. Then, $H = \bigoplus_1^r h_j \mathbb{Z}_{m_j}$, where $h_j := \gcd(\beta, m_j)$ for all j, since $\alpha G = \beta G$ if and only if $\gcd(\alpha, m_j) = \gcd(\beta, m_j)$ for all j. Set $\gamma_* := \operatorname{lcm}(h_1, \dots, h_r)$. We show that $\gamma_* G = H$ and $\gamma_* \wr \alpha$ for all $\alpha \in \Gamma_H$, proving the lemma.

Fix $j \in [r]$. Now, $h_j \wr \gamma_* = \operatorname{lcm}(h_1, \dots, h_r)$ and $h_j \wr m_j$ by assumption. Hence, $h_j \wr \gcd(\gamma_*, m_j)$. Conversely, if $x \wr z$ and $y \wr z$, then $\operatorname{lcm}(x, y) \wr z$, and so

$$\gamma_* = \operatorname{lcm}(h_1, \dots, h_r) \wr \beta$$

since $h_j \wr \beta$. Hence, $\gcd(\gamma_*, m_j) \wr \gcd(\beta, m_j) = h_j$. Thus, $h_j = \gcd(\gamma_*, m_j)$. Hence, $\gamma_* G = H$. Now consider any α with $\alpha G = H$; so, $h_j = \gcd(\alpha, m_j)$ for all j. Hence, $h_j \wr \alpha$ for all j, and so $\operatorname{lcm}(h_1, \dots, h_r) \wr \alpha$, i.e., $\gamma_* \wr \alpha$. $\blacksquare$

4. Total variation cutoff: Combining Approaches #1 and #2

The only regime which we have not yet covered is

$$\sqrt{\frac{\log|G|}{\log\log\log|G|}} \lesssim k \lesssim \sqrt{\log|G|} \quad \text{with } 1 \ll k - d(G) \ll k;$$

see Remarks 2.4 and 3.6 for the regimes covered by Approaches #1 and #2, respectively. We combine the approaches here, using the refined notion of entropic times (Section 3.2), to handle the rest.

4.1. *Precise statements and remarks*

There are some simple conditions on k, in terms of $d(G)$ and $|G|$, needed for the upper bound.

Hypothesis C. *The sequence* $(k_N, G_N)_{N \in \mathbb{N}}$ *satisfies Hypothesis* C *if the following hold:*

$$\liminf_{N \to \infty} \frac{k_N}{\sqrt{\log|G_N| / \log\log\log|G_N|}} > 0, \qquad \limsup_{N \to \infty} \frac{k_N}{\sqrt{\log|G_N|}} < \infty,$$

$$\liminf_{N \to \infty} (k_N - d(G_N)) = \infty, \qquad \limsup_{N \to \infty} \frac{k_N - d(G_N)}{k_N} = 0.$$

Remark 4.1. In short, the conditions of Hypothesis C say that

$$\sqrt{\frac{\log|G|}{\log\log\log|G|}} \lesssim k \lesssim \sqrt{\log|G|} \quad \text{and} \quad 1 \ll k - d(G) \ll k.$$

Throughout the proofs, we drop the subscript-N from the notation, e.g., writing k or n, considering sequences implicitly. Recall that we abbreviate the TV distance from uniformity at time t as

$$d_{G_k,N}(t) = \|\mathbb{P}_{G_N([Z_1,\ldots,Z_{k_N}])}(S(t) \in \cdot) - \pi_{G_N}\|_{\mathrm{TV}},$$

where

$$Z_1, \ldots, Z_{k_N} \sim^{\mathrm{iid}} \mathrm{Unif}(G_N).$$

We now state the main theorem of this section. Recall that

$$\tau_* = \max_{\gamma \in \mathbb{N}} \tau_0(\gamma, |G/\gamma G|).$$

Theorem 4.2. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ of finite, Abelian groups; for each $N \in \mathbb{N}$, define $Z_{(N)} := [Z_1, \ldots, Z_{k_N}]$ by drawing $Z_1, \ldots, Z_{k_N} \sim^{\mathrm{iid}} \mathrm{Unif}(G_N)$. Suppose that the sequence $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis C.*

Let $c \in (-1, 1) \setminus \{0\}$. Then

$$d_{G_k,N}((1+c)\tau_*(k_N, G_N)) \to^{\mathbb{P}} \mathbf{1}(c < 0) \quad \textit{(in probability) as } N \to \infty.$$

That is, w.h.p., there is TV cutoff at $\max_\gamma \tau_0(\gamma, |G/\gamma G|)$. Moreover, the implicit lower bound on the TV distance holds deterministically, i.e., for all choices of generators.

Remark. The TV lower bound from Section 3.6 is valid whenever $k - d(G) \gg 1$. Thus, it suffices to consider only the upper bound here. The asymptotic evaluation of τ_* depends on the regime of k.

4.2. Outline of proof

Fundamentally, we still wish to bound the same expression that we did previously:

$$\sum_{\gamma \wr |G|} \mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \cdot \frac{|G|}{|\gamma G|} - 1;$$

see Propositions 2.9 and 3.13. In Section 2.7, we used $|G/\gamma G| \le \gamma^{d(G)}$. In Section 3.7, we used typicality to get

$$\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \le \mathbb{P}(W_\gamma = W'_\gamma \mid \mathsf{typ}) \ll \frac{1}{|G/\gamma G|}.$$

The idea here, for this interim regime of k near $\sqrt{\log|G|}$, is to improve the bound $|G/\gamma G| \le \gamma^{d(G)}$ for all but $e^{o(k)}$ of the γ; for the remaining γ, we use $|G/\gamma G| \le \gamma^{d(G)}$ and the second approach.

4.3. Upper bound on total-variation mixing

Let G be an Abelian group; set $n := |G|$. One can find a decomposition $\bigoplus_1^d \mathbb{Z}_{m_j}$ of G such that $d = d(G)$, the minimal size of a generating set, and $m_i \wr m_j$ for all $i \le j$. Fix such a decomposition.

Let $\varepsilon > 0$ and let $t := (1+\varepsilon)\tau_*(k, G)$. We frequently suppress the t and ε dependence in the notation. Let $c := c_\varepsilon > 0$ be the constant from Proposition 3.5. Recall some notation from Section 4.3:

$$\zeta_\gamma = \frac{1}{k}(k-d)\log\gamma, \quad \hat{\zeta}_\gamma = \zeta_\gamma \wedge 1, \quad \delta_\gamma = e^{-c\hat{\zeta}_\gamma k} \quad \text{and} \quad r_* = \frac{1}{2}|G|^{1/k}(\log k)^2.$$

Since $k - d \gg 1$ and $k \lesssim \sqrt{\log n}$, we have $\hat{\zeta}_n = 1$; set $\hat{\zeta}_\infty := 1$. Recall that W is a RW on $\mathbb{Z}$ and we define W_γ by W mod γ; set $W_\infty := W$. We now define typicality for this section precisely.

Definition 4.3 (Cf. Definition 3.8). Define typical sets for $\gamma \in \mathbb{N} \cup \{\infty\}$ by the following:

$$\mathcal{W}_{\gamma,\text{glo}} := \Big\{ w \in \mathbb{Z}_\gamma^k \mid \mathbb{P}(W_\gamma(t) = w) \le \frac{\delta_\gamma}{|G/\gamma G|} \Big\}, \quad \text{where } \delta_\gamma = e^{-c\hat{\zeta}_\gamma k};$$

$$\mathcal{W}_{\text{loc}} := \{ w \in \mathbb{Z}^k \mid |w_i - \mathbb{E}(W_i(t))| \le r_* \ \forall\, i \in [k] \}, \quad \text{where } r_* = \frac{1}{2} n^{1/k}(\log k)^2.$$

Choose L to be the maximal integer in $[1, d]$ with $m_L \le M$, where

$$M := e^{\sqrt{\log n \log\log n}}; \qquad \text{set } \Gamma := \{ rm \mid r \in [k^{1/2}],\ m \wr m_L,\ rm \wr n \} \setminus \{1\}.$$

When W' is an independent copy of W, define *typicality* by

$$\mathsf{typ} := \{ W(t), W'(t) \in \mathcal{W}_{\text{loc}} \} \cap \Big(\bigcap_{\gamma\in\Gamma} \{ W_\gamma(t), W'_\gamma(t) \in \mathcal{W}_{\gamma,\text{glo}} \} \Big).$$

Lemma 4.4. *We have* $\log|\Gamma| \ll k$. *In particular,* $\delta_\infty |\Gamma| = o(1)$.

Proof. We have $|\Gamma| \le k^{1/2} \operatorname{div} m_L$, where $\operatorname{div} m$ is the number of divisors of $m \in \mathbb{N}$. It is a standard number-theoretic result that $\log \operatorname{div} m \lesssim \log m / \log\log m$ uniformly in $m \in \mathbb{N}$; see, e.g., [20, §18.1]. By the definition of m_L and the assumption that $k \gtrsim \sqrt{\log n / \log\log\log n}$, we obtain

$$\log \operatorname{div} m_L \lesssim \frac{\log M}{\log\log M} \lesssim \sqrt{\frac{\log n \log\log n}{\log\log n}} \ll k.$$

Thus, $\log|\Gamma| \ll k$. Finally, recall that $\log(1/\delta_\infty) = ck \asymp k$. ∎

The following result is an immediate consequence of Lemmas 3.9 and 4.4 and Proposition 3.10.

Lemma 4.5 (Cf. Lemma 3.9 and Proposition 3.10). *We have*

$$\sum_{\gamma\in\Gamma} \delta_\gamma = o(1) \quad \textit{and} \quad \mathbb{P}(\mathsf{typ}) = 1 - o(1).$$

Thus, by applying the modified L_2 calculation, as before, it suffices to prove the following result.

Proposition 4.6. *Let $\varepsilon > 0$ be fixed and set $t := (1+\varepsilon)\tau_*(k, G)$. Then,*

$$|G|\mathbb{P}(S = S' \mid \mathsf{typ}) - 1 = \sum_{\gamma\in\mathbb{N}} |G/\gamma G|\,\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) - 1 = o(1).$$

In order to prove this, we first show that $L \eqsim d \eqsim k$.

Lemma 4.7. *We have $0 \le d - L \le \sqrt{\log n/\log\log n} \ll k$. In particular, $L \eqsim d \eqsim k$.*

Proof. By definition, $L \in [1, d]$, so $L \le d$. If $L < d$, then $m_L \le M \le m_{L+1}$. Now, $n = m_1 \cdots m_d$, so then $M^{d-L} \le m_{L+1}^{d-L} \le m_{L+1} \cdots m_d \le n$. Recall that

$$k \gtrsim \sqrt{\frac{\log n}{\log\log\log n}}.$$

Rearranging,

$$d - L \le \frac{\log n}{\log M} = \sqrt{\frac{\log n}{\log\log n}} \ll k \eqsim d. \qquad \blacksquare$$

We prove Proposition 4.6 by separating the sum over γ into two parts according to Γ.

Proof of Proposition 4.6. Observe that $|G/\gamma G|\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \le 1$ when $\gamma = 1$. Also, $\mathfrak{g} \wr n$. Thus,

$$\sum_{\gamma\in\mathbb{N}} |G/\gamma G|\,\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) - 1 \le \sum_{\gamma\in\Gamma'} |G/\gamma G|\,\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) + \sum_{\gamma\in\Gamma} |G/\gamma G|\,\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}),$$

where $\Gamma' := \{\gamma \in [2, n] \mid \gamma \wr n\} \setminus \Gamma$. We analyse these sums with Approaches #1 and #2, respectively: namely, we show below that both sums are $o(1)$, when $t := (1+\varepsilon)\tau_*(k, G)$ with $\varepsilon > 0$ a constant.

#1 Suppose that $\gamma \in \Gamma'$, so $\gamma \notin \Gamma \cup \{1\}$. For each $j \in [L]$, we may write

$$\gamma = r_j \cdot \gcd(\gamma, m_j) \quad \text{and} \quad m_j = r_j' \cdot \gcd(\gamma, m_j),$$

where

$$\gcd(r_j, r_j') = 1.$$

By definition of Γ, if $\gamma = \tilde{r} \cdot m$ for some $m \wr m_j$, then $\tilde{r} > k^{1/2}$, as $\gamma \notin \Gamma$. Hence, $\gcd(\gamma, m_j) = \gamma/r_j \le \gamma/k^{1/2}$ for $j \in [L]$. Applying this to the first L terms of the product gives

$$|G/\gamma G| = \prod_1^d \gcd(\gamma, m_j) \le \frac{\gamma^d}{k^{L/2}}.$$

Let $\delta \in (0,1)$. Exactly the same analysis as in the proof of Corollary 2.15 then leads us to

$$\sum_{\gamma \in \Gamma'} |G/\gamma G| \, \mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \le e^{2\delta k} 2^{d+1-k} + 2^k \delta^{d+1-k} n^{(d+1-k)/k} + \frac{4^k (\log k)^{2(d+1)}}{k^{L/2}}.$$

Setting $\delta := (1/4)(k-d-1)/k$ makes the first two terms $o(1)$, as in Corollary 2.15. For the third term, $4^k (\log k)^{2(d+1)}/k^{L/2} \ll 1$ as $L \eqsim k \eqsim d$. Hence, the sum over $\gamma \in \Gamma'$ is $o(1)$.

#2 The typicality conditions set out in Definition 4.3 imply that

$$\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \le \mathbb{P}(W_\gamma = W'_\gamma \mid \mathsf{typ}) \le \frac{\delta_\gamma}{|G/\gamma G|};$$

cf. Lemma 2.10. Hence, combined with Lemma 4.5, the sum over $\gamma \in \Gamma$ is $o(1)$. ■

5. Separation cutoff

Recall that separation distance is defined by

$$s(t) := \max_{x,y} \left\{ 1 - \frac{P_t(x,y)}{\pi(y)} \right\} \quad \text{for } t \ge 0,$$

where $P_t(x,y)$ is the time-t transition probability from x to y and π the invariant distribution. We write $s_{G_k,N}$ when considering sequences $(k_N, G_N)_{N \in \mathbb{N}}$, analogously to $d_{G_k,N}$ for total variation.

5.1. Precise statement and remarks

As for the previous theorems, conditions are imposed on (k, G).

Hypothesis D. *The sequence $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis D if the following hold:*

$$\liminf_{N\to\infty} \frac{k_N - d(G_N)}{\max\{(\log|G_N|/k_N)^2, \, (\log|G_N|)^{1/2}\}} = \infty \quad \textit{and} \quad \limsup_{N\to\infty} \frac{\log k_N}{\log|G_N|} = 0.$$

Remark 5.1. It is easy to check that Hypothesis D is satisfied when $\log k \ll \log|G|$ and

$$\begin{aligned} &\text{either} \quad k \gtrsim (\log|G|)^{3/4} \quad \text{and} \quad k - d(G) \gg (\log|G|)^{1/2} \\ &\text{or} \quad\quad k \gg (\log|G|)^{2/3} \quad \text{and} \quad k - d(G) \asymp k. \end{aligned}$$

In particular, the latter holds whenever $k - \log_2|G| \gtrsim \log|G|$, e.g., $k \gg \log|G|$.

Theorem 5.2. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ of finite, Abelian groups; for each $N \in \mathbb{N}$, define*

$$Z_{(N)} := [Z_1, \ldots, Z_{k_N}]$$

by drawing $Z_1, \ldots, Z_{k_N} \sim^{\text{iid}} \text{Unif}(G_N)$. Suppose that the sequence $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis D. Let $c \in (-1, 1) \setminus \{0\}$. Then,

$$s_{G_k,N}((1+c)\tau_*(k_N, G_N)) \to^{\mathbb{P}} \mathbf{1}(c < 0) \quad \textit{(in probability) as } N \to \infty.$$

That is, w.h.p. there is separation cutoff at $\tau_*(k, G)$. Moreover, the implicit lower bound on the separation distance holds deterministically, i.e., for all choices of generators.

Remark. Total-variation distance is a lower bound on separation distance; see, e.g., [35, Lemma 6.16]. Hence, the lower bound on separation distance is immediate from that on TV distance.

The proof uses the previously established TV mixing time upper bound as a building block.

5.2. Upper bound on separation mixing

Preliminaries. Write $n := |G|$. We want to show, for fixed $\xi > 0$, that

$$\min_{x\in G} P_t^{\pm}(0, x) \ge \frac{1}{n}(1 - o(1)) \quad \text{for some } t \le (1 + 2\xi)\tau_*^{\pm}(k, G).$$

Abbreviate $d := d(G)$. Let $\chi = o(1)$, to be specified later. Throughout the course of the proof, we impose conditions on χ; at the end, we show that these are equivalent to Hypothesis D.

Set $k' := k - \chi(k - d)$; then, $k' \eqsim k$ and $k' - d = (1 - \chi)(k - d) \eqsim k - d \gg 1$. Let $A := [Z_1, \ldots, Z_{k'}]$ be the first k' generators and $B := [Z_{k'+1}, \ldots, Z_k]$ be the remaining $k - k' = \chi(k - d)$. Since G is Abelian, $P_t = P_{t,A}P_{t,B}$, where in $P_{t,A}$ (resp. $P_{t,B}$), we pick each generator of A (resp. B) at rate $1/k$ independently. In words, first apply the generators from A, then those from B. ■

Let $\xi > 0$ be a constant; let $t' := (1 + \xi)\tau_*(k', G)$. Since there is cutoff on $G(A)$ w.h.p. at $\tau_*(k', G)$, we can choose $\delta = o(1)$ so that t' is larger than the δ^2-TV mixing time for the rate-1 RW on $G(A)$ for a typical choice of A. In the regime $k \gg \log n$, simply having $\delta = o(1)$ will be sufficient. In the regime $k \lesssim \log n$, we quantify this to be $\delta = e^{-2c(k-d)}$. Hypothesis D implies that $k \gg \sqrt{\log n}$; combined with $k \lesssim \log n$, this means that Hypothesis B, used in Approach #2 (Section 3), is satisfied. We also compare $\tau_*(k', G)$ and $\tau_*(k, G)$, the w.h.p.-cutoff times for $G(A)$ and $G(Z)$, respectively.

We need two auxiliary lemmas, which we state now; their proofs are deferred to Section 5.3.

Lemma 5.3. *Assume Hypothesis D. Then, there exists $\delta \ll 1$ such that the δ^2-mixing time of the RW on $G(A)$ is at most $t' = (1+\xi)\tau_*(k', G)$ w.h.p. Further, when additionally $k \lesssim \log n$, there exists a constant $c > 0$ such that we may take $\delta := e^{-4c(k-d)}$.*

Lemma 5.4. *We have $\tau_*(k', G) \eqsim \tau_*(k, G)$ if and only if $\chi(k-d)k^{-2} \log n \ll 1$.*

Note that $k \gg \log n$ implies $k - d \eqsim k$, and so $(k-d)k^{-2} \log n \eqsim \log n/k \ll 1$ already; so, any $\chi \ll 1$ suffices. Assume that $\chi(k-d)k^{-2} \log n \ll 1$ so that $\tau_*(k', G) \eqsim \tau_*(k, G)$. To relate this to the rate-1 RW on $G(Z)$, rescale time by $k/|A| = 1/(1 - \chi(k-d)/k)$: set $t := t'/(1 - \chi(k-d)/k)$. Thus,

$$t \eqsim (1+\xi)\tau_*(k, G) \quad \text{as } \chi \ll 1 \le \frac{k}{k-d}; \quad \text{in particular, } t \le (1+2\xi)\tau_*(k, G).$$

By monotonicity of the separation distance with respect to time, it thus suffices to show that

$$\min_{x \in G} P_t(0, x) \ge \frac{1}{n}(1 - o(1)) \quad \text{w.h.p.}$$

Recall that the generators $Z = A \cup B$ are separated into A, the first k', and B, the remainder.

Lemma 5.5. *Assume Hypothesis D. Let $t \ge t' = (1+\xi)\tau_*(k', G)$. Assume that*

$$\frac{1}{k-d} \ll \chi \ll 1 \ \text{ when } k \lesssim \log n \quad \text{and} \quad \frac{1}{t'} \ll \chi \ll 1 \ \text{ when } k \gg \log n.$$

Let δ be as in Lemma 5.3. For $y, z \in G$, define

$$Q_B(y, z) := |B_\pm|^{-1} \sum_{b \in B_\pm} \mathbf{1}(y + b^{-1} = z);$$

i.e., Q_B is the transition matrix for the time-reversed RW on the Cayley graph $G(B)$. Suppose also that for all (deterministic) sets $D \subseteq G$ with $|G \setminus D| \le \delta |G|$ and all $x \in G$ uniformly,

$$\mathbb{P}(Q_B(x, D) \le 1 - \eta) = o\Big(\frac{1}{|G|}\Big) \quad \text{for some } \eta = o(1),$$

where $B^+ := B$ and $B^- := B \cup B^{-1}$ (as multisets). Then,

$$\min_{x \in G} P_t(0, x) \ge \frac{1}{n}(1 - o(1)) \quad \text{w.h.p.}$$

Proof. We condition on a typical realisation of A: write $\mathcal{A} := \{a \mid t_{\text{mix}}(\delta^2; G(a)) \le t'\}$ and condition on $A = a$ for a fixed $a \in \mathcal{A}$. Then, $\mathbb{P}(A \in \mathcal{A}) = 1 - o(1)$ by Lemma 5.3. Given $A = a \in \mathcal{A}$, the set

$$D := \Big\{ z \in G \ \Big| \ P_{t,a}(0, z) \ge \frac{1}{n}(1 - \delta) \Big\} \quad \text{satisfies } |D| \ge n(1 - \delta).$$

Indeed, using the distinguishing-statistic representation of total-variation distance,

$$\delta^2 \ge \pi(D^c) - P_{t,a}(0, D^c) \ge \frac{1}{n}|D^c| - \frac{1}{n}(1-\delta)|D^c| = \frac{1}{n}\delta|D^c|, \quad \text{so } \frac{1}{n}|D^c| \le \delta.$$

For the undirected case (i.e., the RW on G_k^-), by reversibility, conditional on $A = a \in \mathcal{A}$, we have

$$P_t^-(0,x) \geq P_{t,B}^-(x,D) \cdot \frac{1}{n}(1-\delta).$$

While the RW on G_k^+ is not reversible, Cayley graphs have the special property that a step 'backwards' with a generator z corresponds to a step 'forwards' with z^{-1}. Thus,

$$P_t^+(0,x) \geq Q_{t,B}^+(x,D) \cdot \frac{1}{n}(1-\delta),$$

where $Q_{\cdot,B}^+$ is the heat kernel for the RW on $G^+(B^{-1})$, where $B^{-1} := [z^{-1} \mid z \in B]$, rather than on $G^+(B)$. For the RW on G_k^-, replacing the generators with their inverses has no effect on the graph (or RW); set $Q_{\cdot,B}^- := P_{\cdot,B}^-$. We want to show that $Q_{t,B}(x,D) = 1 - o(1)$ uniformly in $x \in G$ w.h.p.

Now, $Q_{t,B}$ corresponds to a RW on $G^\pm(B^{-1})$ run for time t. By considering just the final step of this RW, we now argue that the hypothesis of the lemma is sufficient. Indeed, first note that

$$\min_x Q_{t,B}(x,D) \geq (1 - e^{-t|B|/k}) \cdot \min_x Q_B(x,D),$$

where $e^{-t|B|/k}$ is the probability that none of the generators in B are applied by time t.

- If $k \gg \log n$, then $(k-d)/k \eqsim 1$, so $t|B|/k \eqsim \chi t \gg 1$ whenever $1/t' \ll \chi \ll 1$.
- If $k \lesssim \log n$, then $t \geq t' \gtrsim k$, so $t|B|/k \gtrsim \chi(k-d) \gg 1$ whenever $1/(k-d) \ll \chi \ll 1$.

Thus, the assumptions in the lemma allow us to perform a union bound over $x \in G$:

$$\mathbb{P}(\min_x Q_{t,B}(x,D) \leq 1 - 2\eta \mid A = a) = o(1),$$

where the randomness is over the generators B, provided η decays sufficiently slowly.

For each $a \in \mathcal{A}$, we have the desired lower bound on $\min_x P_t(0,x)$ conditional on $A = a$. Finally, we average over A and use

$$\mathbb{P}(A \in \mathcal{A}) = 1 - o(1)$$

to complete the proof:

$$\begin{aligned}
&\mathbb{P}\Big(\min_x P_t(0,x) \geq \frac{1}{n}(1 - o(1))\Big) \\
&\quad \geq \mathbb{P}\Big(\min_x P_t(0,x) \geq \frac{1}{n}(1 - o(1)) \;\Big|\; A \in \mathcal{A}\Big)\mathbb{P}(A \in \mathcal{A}) = 1 - o(1).
\end{aligned}$$ ∎

We need to check that the supposition of the previous lemma is satisfiable.

Lemma 5.6. *Suppose that $\chi(k-d)^2 \gg \log n$ if $k \lesssim \log n$ and that $\chi \ll 1$ sufficiently slowly if $k \gg \log n$. Then, we can choose $\eta \ll 1$ vanishing sufficiently slowly so that for all (deterministic) sets $D \subseteq G$ with $|G \setminus D| \leq \delta|G|$ and all $x \in G$ uniformly,*

$$\mathbb{P}(Q_B(x,D) \leq 1 - \eta) = o\Big(\frac{1}{|G|}\Big).$$

Proof. Fix an arbitrary $x \in G$. We desire a proportion at least $1 - \eta$ of the generators in B to connect x to D. The generators are chosen independently, and each connects with probability $|D|/|G| \geq 1 - \delta$. Since there are $\chi(k - d)$ generators, it thus suffices to choose $\eta \ll 1$ so that

$$\mathbb{P}(\mathrm{Bin}(\chi(k-d), 1-\delta) \leq \chi(k-d)(1-\eta)) = o\Big(\frac{1}{|G|}\Big).$$

Let $L := \chi(k - d)$; then, $L \gg 1$. Direct calculation, using standard inequalities, gives

$$\begin{aligned}\mathbb{P}(\mathrm{Bin}(L, 1-\delta) \leq L(1-\eta)) &= \mathbb{P}(\mathrm{Bin}(L, \delta) \geq \eta L) \\ &\leq \binom{L}{\eta L}\delta^{\eta L} \leq \Big(\frac{\delta e}{\eta}\Big)^{\eta L} = \Big(\frac{\delta e}{\eta}\Big)^{\eta\chi(k-d)}.\end{aligned}$$

- Consider first $k \gg \log n$; necessarily, $k - d \eqsim k$. Here, we do not quantify δ: we simply assume $\delta = o(1)$. Requiring η and χ to vanish sufficiently slowly (compared with δ) gives $(\delta e/\eta)^{\eta\chi} = o(1)$. Raising this to the power $k - d \eqsim k \gg \log n$ gives super-polynomial decay.
- Consider now $k \lesssim \log n$. Here, $\delta = e^{-2c(k-d)}$. Choosing $\eta \geq e\sqrt{\delta} = e^{-c(k-d)+1}$ gives

$$\Big(\frac{\delta e}{\eta}\Big)^{\eta\chi(k-d)} = \delta^{\eta\chi(k-d)/2} = e^{-c\eta\chi(k-d)^2}.$$

We can choose $\eta \ll 1$ so that $\eta\chi(k - d)^2 \gg \log n$, giving super-polynomial decay.

These bounds are independent of x, and hence hold for all $x \in G$ uniformly, as required. ∎

On top of Hypothesis D, we need χ to satisfy the following simultaneously when $k \lesssim \log n$:

- we need $\chi(k - d)k^{-2} \log n \ll 1$ to hold for Lemma 5.4;
- we need $1/(k - d) \ll 1 \ll \chi$ and $\chi(k - d)^2 \gg \log n$ to hold for Lemmas 5.5 and 5.6.

No requirements beyond 'sufficiently slowly', in terms of δ, are imposed on χ when $k \gg \log n$. The next lemma states that having such a χ is equivalent to Hypothesis D; its proof is deferred to Section 5.3.

Lemma 5.7. *Assume Hypothesis* D *and that* $k \lesssim \log n$. *Then, we can choose* $\chi \in (0, 1)$ *satisfying*

$$\chi(k-d)k^{-2}\log n \ll 1, \quad \chi(k-d)^2 \gg \log n \quad \text{and} \quad \frac{1}{k-d} \ll \chi \ll 1.$$

In fact, Hypothesis D *is equivalent to being able to find such a* χ.

These lemmas combine easily to establish the upper bound in Theorem 5.2, as we expose now.

Proof of upper bound in Theorem 5.2. Assume that $1 \ll \log k \ll \log n$. Lemma 5.7 guarantees that a χ satisfying the conditions of Lemmas 5.4, 5.5 and 5.6 simultaneously can be found under Hypothesis D when $k \lesssim \log n$; when $k \gg \log n$, simply take $\chi \ll 1$ sufficiently slowly.

The conclusion of these lemmas is that the separation distance is $o(1)$ w.h.p. at time t', and that $t' = (1+\xi)\tau_*(k', G) \le (1+2\xi)\tau_*(k, G)$. This completes the proof of the upper bound. ■

5.3. Auxiliary lemmas

It remains to give the deferred proofs of the auxiliary lemmas: Lemmas 5.3, 5.4 and 5.7.

Proof of Lemma 5.3. Consider $k' \eqsim k \lesssim \log n$ first. We use Approach #2 (Section 3.7), applied to $G(A)$; recall that A has k' iid generators, and that $t' = (1+\xi)\tau_*(k', G)$. Quantifying the *proof of upper bound in Theorem 3.7*, using Lemma 3.9 and Proposition 3.13, gives a bound of

$$\frac{\delta_\infty(|\Gamma|+1) + 2^{-c(k-d)+1}}{\mathbb{P}(\mathsf{typ})} + \mathbb{P}(\mathsf{typ}^c)$$

on the expected TV distance at time t', under Hypothesis B with typicality given by Definition 3.8.

The proof of Proposition 3.10 shows that *global* typicality fails with probability at most $\delta_\infty|\Gamma| + e^{-c(k-d)+1}$. The *local* conditions as stated in Definition 3.8 do not fail with sufficiently low probability – only at most η, for some $\eta = o(1)$. This is proved via standard large deviation estimates from [26, §C]. Replacing r_* by $r_* \log n$ gives failure probability $\eta^{\log n} \le e^{-k}$.

Increasing r_* like this increases $|\Gamma|$, but not enough to cause any issues. Indeed, we have

$$\log|\Gamma| \le \log n_* \le \log(n^{1/k}(\log k)^2 \log n) \le \frac{1}{k}\log n + 3\log\log n.$$

But, Hypothesis D implies that $k - d \gtrsim (\log n)^{1/2}$ and that

$$k \ge k-d \gg \left(\frac{\log n}{k}\right)^2, \quad \text{so } k \gg (\log n)^{2/3}; \quad \text{hence, } \log|\Gamma| \ll (\log n)^{1/3} \ll k-d.$$

Thus, the dominating term in the upper bound is $2^{-c(k-d)}$. Adjusting c, this completes the case.

The case $k' \eqsim k \gg \log n$ is trivial, since here all Abelian groups have TV cutoff at $\tau_*(k', G)$. ■

Proof of Lemma 5.4. We have $k \eqsim k'$ and $k - d \eqsim k' - d$. Observe that $n^{2/k} \eqsim n^{2/k'}$ if and only if

$$1 \gg \left(\frac{2}{k'} - \frac{2}{k}\right)\log n = \left(\frac{2}{k-\chi(k-d)} - \frac{2}{k}\right)\log n, \quad \text{i.e., } \chi(k-d)k^{-2}\log n \ll 1.$$

The claim follows by Proposition 3.2a for $1 \ll k \lesssim \log n$. On the other hand, if $k \gg \log n$, then

$$\tau_*(k, G) \eqsim T(k,n) := \frac{\log n}{\log(k/\log n)};$$

see Section 1.3.1. Hence, $T(k,n) \eqsim T(\alpha k, n)$ for all $\alpha \in (0,\infty)$. Thus, $T(k,n) \eqsim T(k',n)$ as $k \eqsim k'$. ■

Proof of Lemma 5.7. Rearranging the conditions, they are equivalent to having

$$\sqrt{\frac{\log n}{\chi}} \ll k - d \ll \frac{k^2}{\chi \log n} \quad \text{for some } \chi \in (0,1) \text{ with } \frac{1}{k-d} \ll \chi \ll 1.$$

We reparametrise these conditions. Let $\varepsilon \in (0,\infty)$ and set

$$\chi := \frac{\varepsilon k^2}{(k-d)\log n}; \quad \text{then } \sqrt{\frac{\log n}{\chi}} = \frac{\sqrt{k-d}\,\log n}{\sqrt{\varepsilon} k}.$$

The conditions on χ then, in terms of ε, become

$$\frac{(\log n)^2}{(k-d)k^2} \ll \varepsilon \ll 1 \quad \text{and} \quad \frac{\log n}{k^2} \ll \varepsilon \ll \frac{(k-d)\log n}{k^2}.$$

We can find such an $\varepsilon \in (0,\infty)$, implicitly a sequence, if and only if

$$\max\left\{\frac{(\log n)^2}{(k-d)k^2}, \frac{\log n}{k^2}\right\} \ll \min\left\{1, \frac{(k-d)\log n}{k^2}\right\}.$$

Some case analysis shows that this condition is equivalent to the first condition of Hypothesis D. ■

6. Nilpotent groups: Mixing comparison and expansion

We compare the mixing times for a nilpotent group G with a 'corresponding' Abelian group $\bar{G}$: $t_{\mathrm{mix}}(G_k)/t_{\mathrm{mix}}(\bar{G}_k) \le 1 + o(1)$. We use this to show that G_k is an expander w.h.p. if $k - d(\bar{G}) \gtrsim \log|G|$.

To emphasise, the material in this section applies to both the un- and directed cases, simultaneously. Throughout, $(G_{(\ell)})_{\ell \ge 0}$ is the lower central series of G and $L := \min\{\ell \ge 0 \mid G_{(\ell)} = \{\mathsf{id}\}\}$.

6.1. Precise statements

We prove Theorem D, which we recall here for the reader's convenience as Theorem 6.1.

Theorem 6.1. *Let G be a nilpotent group. Set $\bar{G} := \bigoplus_1^L (G_{(\ell-1)}/G_{(\ell)})$. Suppose that $1 \ll \log k \ll \log|G|$ and $k - d(\bar{G}) \gg 1$. Let $\varepsilon > 0$ and let $t \ge (1+\varepsilon)\tau_*(k, \bar{G})$. Then, $d_{G_k}(t) = o(1)$ w.h.p.*

Remark. An equivalent bound, depending only on k and $|G| = |\bar{G}|$, valid for all groups has already been established when $k \gg \log|G|$; recall Section 1.3.1. Thus, we need only consider $1 \ll k \lesssim \log|G|$.

We use this mixing time bound to show that G_k for nilpotent G is an expander w.h.p. when $k - d(\bar{G}) \gtrsim \log|G|$. The isoperimetric constant was defined in Definition E for d-regular graphs:

$$\Phi_* := \min_{1 \le |S| \le |V|/2} \Phi(S), \quad \text{where } \Phi(S) := \frac{1}{d|S|} |[\{a, b\} \in E \mid a \in S, \, b \in S^c]|.$$

Specifically, we prove Theorem E, which we recall here for the reader's convenience as Theorem 6.2.

Theorem 6.2. *Let G be a nilpotent group. Set $\bar{G} := \bigoplus_1^L (G_{(\ell-1)}/G_{(\ell)})$. Suppose that $k - d(\bar{G}) \gtrsim \log|G|$. Then, $\Phi_*(G_k) \asymp 1$ w.h.p.*

6.2. *Outline of proof*

Consider the series of quotients $(Q_\ell := G_{(\ell-1)}/G_{(\ell)})_{\ell=1}^L$. For each $\ell \in [L]$, choose a set $R_\ell \subseteq G_{(\ell-1)}$ of coset representatives for $Q_\ell = G_{(\ell-1)}/G_{(\ell)}$. To sample $Z_i \sim \mathrm{Unif}(G)$, it suffices to sample $Z_{i,\ell} \sim \mathrm{Unif}(R_\ell)$ for each ℓ independently, and then take the product: $Z_i := Z_{i,1} \cdots Z_{i,L}$; see Lemma 6.3. Then $Z_{i,\ell} G_{(\ell)} \sim \mathrm{Unif}(Q_\ell)$ independently for each i and ℓ; see Corollary 6.4.

Suppose that M steps are taken; let $\sigma\colon [M] \to [k]$ indicate which generator is used in each step. Set $S := \prod_{m=1}^M Z_{\sigma(m)}$. For each $\ell \in [L]$, let $S_\ell := \prod_{m=1}^M Z_{\sigma(m),\ell}$; this is the projection of S to Q_ℓ. Then, each $S_\ell G_{(\ell)}$ is a RW on Q_ℓ, which is an Abelian group, but all using the same choice σ.

These are RWs on Abelian groups, so the ordering in σ does not matter. For each $i \in [k]$, let W_i be the number of times in σ that generator Z_i has been applied minus the number of times that Z_i^{-1} has been applied. Let σ' be an independent copy of σ and define S' and W' via σ' and Z; for each $\ell \in [L]$, define $S'_\ell := \prod_{m=1}^M Z_{\sigma(m),\ell}$. Then, S and S' are iid conditional on Z.

To compare the RW on the nilpotent group with one on an Abelian group, we show that

$$n\mathbb{P}(S = S' \mid (W, W')) \le n \prod_1^L \mathbb{P}(S_\ell G_{(\ell)} = S'_\ell G_{(\ell)} \mid (W, W')) = |\bar{G}/\mathfrak{g}\bar{G}|,$$

where

$$\mathfrak{g} := \gcd(W_1 - W'_1, \ldots, W_k - W'_k, n);$$

see Proposition 6.6 and Corollary 6.9. By analysing $|\bar{G}/\mathfrak{g}\bar{G}|$, we showed in Sections 2–4 that the RW on $\bar{G}_k$ is mixed w.h.p. shortly after $\tau_*(k, \bar{G})$; see specifically Lemma 2.11. From this and the inequality above, we are able to deduce that the RW on G_k is mixed w.h.p. shortly after the same time.

6.3. Reduction to Abelian-type calculations

Consider the series of quotients $(Q_\ell := G_{(\ell-1)}/G_{(\ell)})_{\ell=1}^L$. For each $\ell \in [L]$, choose a set $R_\ell \subseteq G_{(\ell-1)}$ of coset representatives for $Q_\ell = G_{(\ell-1)}/G_{(\ell)}$, i.e., a set R_ℓ with $|R_\ell| = |Q_\ell|$ and $\{rG_{(\ell)}\}_{r\in R_\ell} = G_{(\ell-1)}/G_{(\ell)} = Q_\ell$. We sample the uniform generators via uniform random variables on each of the quotients. In this way, projecting to one of the quotients, we get a RW on this quotient.

Lemma 6.3. *For each $\ell \in [L]$, let $Y_\ell \sim \mathrm{Unif}(R_\ell)$ independently. Then, $Y := Y_1 \cdots Y_L \sim \mathrm{Unif}(G)$.*

Proof. Let $r_0 \in G$ and consider the event $\{Y = r_0\}$. If $r_0 = Y_1 \cdots Y_L$, then $r_1 := Y_1^{-1} r_0 = Y_2 \cdots Y_L$. Clearly, the right-hand side is in $G_{(1)}$, and so the left-hand side is too. Hence, $r_0 \equiv Y_1 \bmod G_{(1)}$. But, $Y_1 \sim \mathrm{Unif}(R_1)$, so the probability of this is $1/|R_1| = 1/|G_{(0)}/G_{(1)}|$. Similarly, $r_2 := Y_2^{-1} r_1 = Y_3 \cdots Y_L \in G_{(2)}$, so $r_1 \equiv Y_2 \bmod G_{(2)}$, the probability of which is $1/|R_2| = 1/|G_{(1)}/G_{(2)}|$. Iterating,

$$\mathbb{P}(Y = r_0) = \prod_1^L \frac{1}{|G_{(\ell-1)}/G_{(\ell)}|} = \prod_1^L \frac{|G_{(\ell)}|}{|G_{(\ell-1)}|} \frac{|G_{(L)}|}{|G_{(0)}|} = \frac{1}{|G|},$$

since the Y_ℓ are independent. Since $r_0 \in G$ was arbitrary, we deduce that $Y \sim \mathrm{Unif}(G)$. ∎

This gives the following corollary.

Corollary 6.4. *For all $(i, \ell) \in [k] \times [L]$, sample $Z_{i,\ell} \sim \mathrm{Unif}(R_\ell)$ independently and set $Z_i := Z_{i,1} \cdots Z_{i,L}$. Then, $Z_1, \ldots, Z_L \sim^{\mathrm{iid}} \mathrm{Unif}(G)$. Further, $Z_{i,\ell} G_{(\ell)} \sim \mathrm{Unif}(Q_\ell)$ independently for all (i, ℓ).*

Proof. All the independence claims are immediate. The first claim is immediate from Lemma 6.3.

For the second claim, we have $Z_{i,\ell} \sim \mathrm{Unif}(R_\ell)$ and $|R_\ell| = |Q_\ell|$. Now, $xG_{(\ell)} = yG_{(\ell)}$ if and only if $y^{-1}xG_{(\ell)} = G_{(\ell)}$. If $X \sim \mathrm{Unif}(R_\ell)$ and $H \in Q_\ell$, say $H = yG_{(\ell)}$ with $y \in R_\ell$, then $y^{-1}X \sim \mathrm{Unif}(R_\ell)$ independently of y. So, $\mathbb{P}(XG_{(\ell)} = yG_{(\ell)}) = 1/|R_\ell|$. Hence, $XG_{(\ell)} \sim \mathrm{Unif}(Q_\ell)$. ∎

Assume that Z is drawn in this way for the remainder of the section. The next main result (Proposition 6.6) is the key element of the proof of Theorem 6.1. It reduces the problem to a collection of Abelian calculations, similar to those already considered earlier.

First, we need an auxiliary lemma, showing that $v \cdot Z = 0$ is the 'worst case'.

Lemma 6.5. *Let H be an Abelian group. Let $Z_1, \ldots, Z_k \sim^{\mathrm{iid}} \mathrm{Unif}(H)$ and $v \in \mathbb{Z}^k$. Then,*

$$\max_{h\in H} \mathbb{P}(v \cdot Z = h) = \mathbb{P}(v \cdot Z = \mathsf{id}(H)).$$

Proof. Write $0 := \mathsf{id}(H)$. Let $h \in H$. Write $A(h) := \{z \in H^k \mid v \cdot z = h\}$. If $w \in A(h)$, then

$$B := \{z - w \mid z \in A(h)\} \subseteq A(0);$$

also, clearly, $|B| = |A(h)|$, so $|A(h)| \le |A(0)|$. Hence,

$$\mathbb{P}(v \cdot Z = h) = \frac{|A(h)|}{|H|^k} \le \frac{|A(0)|}{|H|^k} = \mathbb{P}(v \cdot Z = 0). \qquad \blacksquare$$

The following theorem decomposes the probability $\mathbb{P}(S = S')$ into a product of probabilities $\mathbb{P}(S_\ell G_{(\ell)} = S'_\ell G_{(\ell)})$. The latter correspond to probabilities of RWs on *Abelian groups* Q_ℓ.

Proposition 6.6. *Let* $M, M' \in \mathbb{N}$. *Let* $\sigma\colon [M] \to [k]$ *and* $\sigma'\colon [M'] \to [k]$. *Let* $\eta \in \{\pm 1\}^M$ *and* $\eta' \in \{\pm 1\}^{M'}$. *Recall that* $(G_{(\ell)})_{\ell \ge 0}$ *is the lower central series and that* $G_{(L)} = \{\mathsf{id}\}$. *For* $\ell \in [L]$, *set*

$$S_\ell := \prod_{m=1}^{M} Z_{\sigma(m),\ell}^{\eta_m}, \quad S'_\ell := \prod_{m=1}^{M} Z_{\sigma'(m),\ell}^{\eta'_m}, \quad S := \prod_{m=1}^{M} Z_{\sigma(m)}^{\eta_m}, \quad S' := \prod_{m=1}^{M} Z_{\sigma'(m)}^{\eta'_m}.$$

For $i \in [k]$, *write*

$$v_i := \sum_{m \in [M']:\sigma'(m)=i} \eta'_m - \sum_{m \in [M]:\sigma(m)=i} \eta_m.$$

Then,

$$\mathbb{P}(S = S') \le \prod_{\ell=1}^{L} \mathbb{P}(S_\ell G_{(\ell)} = S'_\ell G_{(\ell)}) = \prod_{\ell=1}^{L} \mathbb{P}\biggl(\sum_{i=1}^{k} v_i Z_{i,\ell} G_{(\ell)} = \mathsf{id}(Q_\ell) \biggr).$$

The randomness in the above set-up comes from the choice $(Z_k)_{i=1}^k$ of *generators*, not from the RW aspect (σ, η) or (σ', η'): it is valid for *any* choices of $(\sigma, \sigma', \eta, \eta')$. In particular, it applies to both the undirected and directed Cayley graphs, the latter requiring η and η' to be all-1 sequences.

Proof. The claimed equality follows immediately from the fact that Q_ℓ is Abelian.

We now set up a little notation. Write $A_{i,\ell} := Z_{i,1} \cdots Z_{i,\ell-1}$ and $B_{i,\ell} := Z_{i,\ell+1} \cdots Z_{i,L}$; then $Z_i = A_{i,\ell} Z_{i,\ell} B_{i,\ell}$. (Here, $A_{i,1} := \mathsf{id}$ and $B_{i,L} := \mathsf{id}$.) Note that $B_{j,\ell} \in G_{(\ell)}$ for all $j \in [k]$ and $\ell \in [L]$.

Let $\mathcal{E}_\ell := \{S' S^{-1} \in G_{(\ell)}\}$. Then,

$$\mathbb{P}(S = S') = \prod_{1}^{L} \mathbb{P}(\mathcal{E}_\ell \mid \mathcal{E}_{\ell-1}).$$

Now, $[g, h] \in G_{(\ell)}$ and $hg = gh[h^{-1}, g^{-1}] = gh[g, h]^{-1}$ for all $g \in G$ and $h \in G_{(\ell-1)}$. So,

$$S' S^{-1} = M_\ell N_\ell \cdot \biggl(\prod_{m=1}^{M'} B_{\sigma'(m),\ell}^{\eta'_m} C'_{\sigma'(m),\ell} \biggr) \cdot \biggl(\prod_{m=1}^{M} B_{\sigma(M+1-m),\ell}^{-\eta_{M+1-m}} C'_{\sigma(M+1-m),\ell} \biggr)$$

for some $C_{j,\ell}, C'_{j,\ell} \in G_{(\ell)}$ and M_ℓ and N_ℓ defined as follows:

$$M_\ell := \Big(\prod_{m=1}^{M'} A_{\sigma'(m),\ell}^{\eta'_m} \Big) \cdot \Big(\prod_{m=1}^{M} A_{\sigma(M+1-m),\ell}^{-\eta_{M+1-m}} \Big),$$
$$N_\ell := \Big(\prod_{m=1}^{M'} Z_{\sigma'(m),\ell}^{\eta'_m} \Big) \cdot \Big(\prod_{m=1}^{M} Z_{\sigma(M+1-m),\ell}^{-\eta_{M+1-m}} \Big) \in G_{(\ell-1)}.$$

We thus see that

$$\mathcal{E}_{\ell-1} = \{S'S^{-1} \in G_{(\ell-1)}\}$$

holds if and only if $\{M_\ell \in G_{(\ell-1)}\}$ holds. Crucially, this implies that the indicator $\mathbf{1}(\mathcal{E}_{\ell-1})$ of this event's occurrence is independent of N_ℓ.

We claim that the following is true:

$$\text{given that } S'S^{-1} \in G_{(\ell-1)}, \text{ we have } S'S^{-1} \in G_{(\ell)} \text{ if and only if } M_\ell N_\ell \in G_{(\ell)}.$$

To prove this, we first make three observations, recalling that $G_{(\ell-1)}/G_{(\ell)}$ is Abelian:

- for all $\alpha \in G_{(\ell-1)}$, we have $\alpha G_{(\ell)} = G_{(\ell)}$ and $(\alpha\beta)G_{(\ell)} = (\alpha G_{(\ell)})(\beta G_{(\ell)})$ for all $\beta \in G$;
- $B_{j,\ell}, C_{j,\ell}, C'_{j,\ell} \in G_{(\ell)}$ for all $j \in [k]$ and $N_\ell \in G_{(\ell-1)}$;
- $S'S^{-1} \in G_{(\ell-1)}$ if and only if $M_\ell \in G_{(\ell-1)}$, and so $M_\ell N_\ell \in G_{(\ell-1)}$.

Assume that $S'S^{-1} \in G_{(\ell-1)}$. Applying these observations in the above formula above gives

$$\begin{aligned} S'S^{-1}G_{(\ell)} &= (M_\ell N_\ell G_{(\ell)}) \cdot \Big(\prod_{m=1}^{M'} (B_{\sigma'(m),\ell}^{\eta'_m} G_{(\ell)})(C'_{\sigma'(m),\ell} G_{(\ell)}) \Big) \\ &\quad \times \Big(\prod_{m=1}^{M} (B_{\sigma(M+1-m),\ell}^{-\eta_{M+1-m}} G_{(\ell)})(C_{\sigma(M+1-m),\ell} G_{(\ell)}) \Big) \\ &= M_\ell N_\ell G_{(\ell)}. \end{aligned}$$

Thus, $S'S^{-1} \in G_{(\ell-1)}$ if and only if $M_\ell N_\ell \in G_{(\ell-1)}$, as claimed.

Now, M_ℓ is independent of N_ℓ, and so N_ℓ is independent also of $\mathbf{1}(\mathcal{E}_{\ell-1})$. Thus,

$$\mathbb{P}(\mathcal{E}_\ell \mid \mathcal{E}_{\ell-1}) = \mathbb{P}(M_\ell N_\ell \in G_{(\ell)} \mid \mathcal{E}_{\ell-1}) \le \max_{x \in G_{(\ell-1)}} \mathbb{P}(x N_\ell \in G_{(\ell)}).$$

Now, $Q_\ell = G_{(\ell-1)}/G_{(\ell)}$ is Abelian, and N_ℓ is a product of generators $Z_{j,\ell}$ and $Z_{j,\ell}^{-1}$ for different $j \in [k]$. Hence, we are in the set-up of Lemma 6.5. Applying that lemma,

$$\mathbb{P}(\mathcal{E}_\ell \mid \mathcal{E}_{\ell-1}) \le \mathbb{P}(N_\ell \in G_{(\ell)}) = \mathbb{P}(S_\ell G_{(\ell)} = S'_\ell G_{(\ell)}),$$

using the definition of N_ℓ. This proves the desired inequality. ∎

6.4. Evaluation of Abelian-type calculations

The quotients Q_ℓ are Abelian, so the order in which the generators are applied does not matter. Define $W_i := \sum_{m=1}^{M} \mathbf{1}(\sigma(m) = i)$ for each i. Then, $W = (W_i)_{i=1}^k$ is the RW on $\mathbb{Z}^k$ run for M steps.

Key in analysing the Abelian-type terms are gcds: for all $w, w' \in \mathbb{Z}^k$, define

$$\mathfrak{g}_{(w,w')} := \gcd(w_1 - w_1', w_2 - w_2', \ldots, w_k - w_k', |G|).$$

We use these to evaluate the right-hand side of Proposition 6.6, culminating in Corollary 6.9.

First, we prove an auxiliary lemma akin to Lemma 2.11.

Lemma 6.7. *Let $\ell \in [L]$. For all $w, w' \in \mathbb{Z}^k$, we have*

$$\sum_{i=1}^k v_i Z_{i,\ell} G_{(\ell)} \sim \mathrm{Unif}(\mathfrak{g}_{(w,w')} Q_\ell).$$

Proof. Corollary 6.4 says that $Z_{i,\ell} G_{(\ell)} \sim \mathrm{Unif}(Q_\ell)$ independently. The quotients Q_ℓ are Abelian. Lemma 2.11 says that linear combinations of independent random variables in an Abelian group are also uniform, but on the subgroup given by the gcd of the coefficients. This proves the lemma. ∎

This leads us to a bound on $\mathbb{P}_{(w,w')}(S = S')$ in terms of a product of $|Q_\ell/\gamma Q_\ell| = |Q_\ell|/|\gamma Q_\ell|$ over $\ell \in [L]$, for some γ which is a suitable gcd. The following lemma controls this product.

Lemma 6.8. *For all $\gamma \in \mathbb{N}$, we have $\prod_{\ell=1}^L |\gamma Q_\ell| = |\gamma \bar{G}|$.*

Proof. For any Abelian groups A and B and any $\gamma \in \mathbb{N}$, we have $\gamma(A \oplus B) = (\gamma A) \oplus (\gamma B)$ and $|A \oplus B| = |A||B|$. Since $\bar{G}$ was defined to be a direct sum of the Q_ℓ, the claim now follows. ∎

Let (S', W') be an independent copy of (S, W). Combining Proposition 6.6 and Lemmas 6.7 and 6.8 gives the following corollary. For $w, w' \in \mathbb{Z}^k$, write

$$\mathbb{P}_{(w,w')}(\cdot) := \mathbb{P}(\cdot \mid (W, W') = (w, w')).$$

Corollary 6.9. *For all $w, w' \in \mathbb{Z}^k$, we have*

$$n\mathbb{P}_{(w,w')}(S = S') \le \prod_{\ell=1}^L \frac{|Q_\ell|}{|\mathfrak{g}_{(w,w')} Q_\ell|} = \frac{|\bar{G}|}{|\mathfrak{g}_{(w,w')} \bar{G}|} = |\bar{G}/\mathfrak{g}_{(w,w')} \bar{G}|.$$

Proof. Note that $|Q_\ell|$ divides $|G|$, and so $\gcd(v_1, \ldots, v_k, |Q_\ell|) \le \gcd(v_1, \ldots, v_k, |G|)$ for all $v \in \mathbb{Z}^k$. Also, for any Abelian subgroup H of G, if $\alpha \wr |H|$ and $\alpha \wr \beta$, then $\beta H \le \alpha H$. Combined with Proposition 6.6 and Lemma 6.7, this proves the inequality. To emphasise, Proposition 6.6 is valid for any choice of $(\sigma, \sigma', \eta, \eta')$, so, in particular, applying under this conditioning. The first equality follows immediately from Lemma 6.8. The second equality follows from Lagrange's theorem. ∎

The right-hand side of this corollary depends only on the Abelian group $\bar{G}$. We apply the theory developed in Sections 2–4 to bound the mixing time for this Abelian group.

Proof of Theorem 6.1. Let $\mathcal{W} \subseteq \mathbb{Z}^k$ be arbitrary for the moment. Set

$$D := n\mathbb{P}(S = S' \mid \mathsf{typ}) - 1, \quad \text{where } \mathsf{typ} := \{W, W' \in \mathcal{W}\}.$$

Abbreviate $\mathfrak{g} := \mathfrak{g}_{(W,W')}$. Applying now Corollary 6.9, we obtain

$$D \le \sum_{\gamma \in \mathbb{N}} \mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \cdot |\bar{G}/\gamma\bar{G}| - 1.$$

This latter expression is purely a statistics of the Abelian group $\bar{G}$. We established the upper bound on mixing by looking at *precisely* this quantity. Bounding it was one of the main challenges. There were three different arguments for bounding it, corresponding to different regimes of k. We briefly outline these arguments now. The choice of $\mathcal{W}$ varies from argument to argument.

- In Section 2.7, we upper bounded $|\bar{G}/\gamma\bar{G}| \le \gamma^{d(\bar{G})}$; we then used unimodality to show that $\mathbb{P}(\gamma \wr W_i \mid W_i \neq 0) \le 1/\gamma$, from which we deduced that $\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \le (1/\gamma + \mathbb{P}(W_1 = 0 \mid \mathsf{typ}))^k$.
- In Section 3.7, we analysed (W, W') taken modulo γ, for each γ; we then used entropic considerations to bound $\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \ll |\bar{G}/\gamma\bar{G}|$ in a quantitative sense.
- In Section 4.3, we combined these two approaches.

Instead of reconstructing these arguments, we reference the appropriate places in the previous sections. For each approach, there are conditions on $(k, \bar{G})$; see Hypotheses A–C. At least one of these is satisfied if

$$1 \ll k \lesssim \log|G| - \log|\bar{G}| \quad \text{and} \quad k - d(\bar{G}) \gg 1;$$

see Remarks 2.4, 3.6 and 4.1.

We need to choose the set $\mathcal{W}$; see Definitions 2.7 and 3.8 for the respective definitions, replacing G with $\bar{G}$ in those definitions. See Propositions 2.9, 3.13, and 4.6 specifically for the results bounding the above sum. The conclusion of these results is that

$$D \le \sum_{\gamma \in \mathbb{N}} \mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \cdot |\bar{G}/\gamma\bar{G}| - 1 = o(1).$$

Combined with the modified L_2 calculation of Lemma 2.6, this completes the proof. ■

6.5. *Cutoff for nilpotent groups with small commutators*

We now prove Corollaries D.1–D.3, which relate to nilpotent groups with large Abelianisation.

First, we compare $\tau_*(k, \bar{G} = G^{\mathrm{ab}} \oplus \overline{[G,G]})$ with $\tau_*(k, G^{\mathrm{ab}})$.

Proposition 6.10. *Let A and B be finite, Abelian groups and k be such that $1 \ll \log k \ll \log|A|$.*

- *If $k \lesssim \log|A \oplus B|$, then suppose that $k \gg d(B) \log|B|$ and $k - d(A) \gg d(B)$.*
- *If $k \gg \log|A \oplus B|$, then suppose only that $\log|B| \ll \log|A|$.*

(In either case, $\log|B| \ll \log|A|$.) Then, $\tau_*(k, A \oplus B) \eqsim \tau_*(k, A)$.

These conditions imply that $A \oplus B$ should be viewed as a 'small perturbation' of A. The proof is technical, relying on auxiliary results on entropy; it is deferred to [26, Proposition B.30].

Proof of Corollary D.1. The lower bound of $\tau_*(k, G^{\mathrm{ab}})$ follows by projecting onto the Abelianisation, which is an Abelian group. The argument is analogous to that of Approach #2 in Section 3.6.

The conditions of Corollary D.1 match those of Proposition 6.10 when $A = G^{\mathrm{ab}}$ and $B = \overline{[G, G]}$. Hence, $\tau_*(k, G^{\mathrm{ab}}) \eqsim \tau_*(k, \bar{G} = G^{\mathrm{ab}} \oplus \overline{[G, G]})$. The upper bound of $\tau_*(k, \bar{G})$ is immediate from Theorem D if $k \lesssim \log|G|$. It was already known for $k \gg \log|G|$; recall Section 1.3.1 or see Theorem 7.2. ∎

To prove Corollaries D.2 and D.3, we need an asymptotic evaluation of the entropic times. Recall that $\tau_\gamma(k, \gamma^r)$ is the time at which the entropy of RW on $\mathbb{Z}_\gamma^k$ reaches $\log(\gamma^r) = \log|\mathbb{Z}_\gamma^r|$.

Proposition 6.11 ([26, Proposition B.25a]). *Let $\zeta_\gamma := ((k - r)/k) \log \gamma$. Suppose that $1 \ll k \lesssim r \log \gamma$.*

$$\textit{If } \zeta_\gamma \ll 1, \quad \textit{then } \frac{\tau_\gamma(k, \gamma^r)}{k} \eqsim \frac{1}{2} \frac{\log(1/\zeta_\gamma)}{1 - \cos(2\pi/\gamma)}.$$

$$\textit{If } \zeta_\gamma \gtrsim 1, \quad \textit{then } \frac{\tau_\gamma(k, \gamma^r)}{k} \asymp \gamma^2 e^{-2\zeta_\gamma} = (\gamma^r)^{2/k}.$$

Suppose further that $1 \ll k \ll r \log m$.

$$\textit{If } \zeta_\gamma \gg 1, \quad \textit{then } \frac{\tau_\gamma(k, \gamma^r)}{k} \eqsim \frac{\gamma^2 e^{-2\zeta_\gamma}}{2\pi e} = \frac{(\gamma^r)^{2/k}}{2\pi e}.$$

Proof of Corollary D.2. The lower bound argument is exactly the same as for Corollary D.1.

For the upper bound, we slightly refine the argument used to prove Theorem D. First, we claim that $G^{\mathrm{ab}} \cong \mathbb{Z}_p^r$. Indeed, the Frattini subgroup $\Phi(G)$ satisfies $\Phi(G) = [G, G] G^p$ when G is a p-group, where $G^p := \langle g^p \mid g \in G \rangle$. By definition of being special, $\Phi(G) = [G, G]$. Thus, $G^p \le [G, G]$. In particular, the Abelianisation is of exponent p, as required. Thus, $\bar{G} \cong \mathbb{Z}_p^\ell$ as $G^{\mathrm{com}} \cong \mathbb{Z}_p^s$.

The general arguments for Abelian groups used to prove Theorem A can be specialised to the group $\mathbb{Z}_p^\ell$. We do not include the details here, but rather defer them to [25, Theorem B]. The conditions for this approach are only $k \ge \ell$, rather than $k - \ell \gg 1$ as previously.

We turn to the entropic time. We have $\gamma \mathbb{Z}_p^r = \mathbb{Z}_p^r$ unless $p \wr \gamma$, as p is prime. Thus, the worst-case γ in $\tau_*(k, \mathbb{Z}_p^r) = \max_{\gamma \wr p^r} \tau_\gamma(k, \mathbb{Z}_p^r)$ is $\gamma = p$, and so $\tau_*(k, \mathbb{Z}_p^r) = \tau_p(k, \mathbb{Z}_p^r)$.

For a nilpotent group G, for Z to generate G it suffices that Z generates G^{ab}; here, $G^{\text{ab}} = \mathbb{Z}_p^r$. We analyse Abelian group generation in [25, Lemma 8.1]. That lemma shows that $k - r \gg 1$ is always sufficient to generate the group w.h.p., but if $p \gg 1$ then merely $k - r > 0$ is sufficient.

Finally, if $k - r \asymp k$ and $p \gg 1$, Propositions 2.2 and 6.11 gives

$$\tau_p(k, \mathbb{Z}_p^r) \eqsim \tau_\infty(k, \mathbb{Z}_p^r) = \mathfrak{t}_0(|\mathbb{Z}_p^r|).$$

Indeed, in the notation there, $\zeta_p = ((k - r)/k) \log p$ is the relative entropy; so, $\zeta_p \gg 1$. ∎

For the Heisenberg group $H_{m,d}$, we have the explicit expression $H_{m,d}^{\text{ab}} \cong \mathbb{Z}_m^{2d-4}$, even for m not prime. This allows us to evaluate $\tau_*(k, H_{m,d}^{\text{ab}})$ even when m is not prime, provided $m \gg 1$.

Proof of Corollary D.3. Let $r := 2d - 4$. We have $H_{p,d}^{\text{ab}} \cong \mathbb{Z}_m^r$ and $H_{p,d}^{\text{com}} \cong \mathbb{Z}_m$. The conditions of Corollary D.3 are precisely those required to apply Corollary D.1. Hence, there is cutoff at $\tau_*(k, H_{p,d}^{\text{ab}})$ w.h.p. It remains to evaluate $\tau_*(k, H_{p,d}^{\text{ab}}) = \tau_*(k, \mathbb{Z}_m^r)$ when $k - r \asymp k$ and $m \gg 1$.

First, observe that

$$\gamma \mathbb{Z}_m^r = \mathbb{Z}_{m/\gcd(\gamma,m)}^r \quad \text{for all } \gamma.$$

Hence, by replacing γ by $\gcd(\gamma, m) \le \gamma$, we need only consider γ with $\gamma \wr m$. Next, $|\mathbb{Z}_m^r / \mathbb{Z}_{m/\gamma}^r| = \gamma^r$. Hence,

$$\tau_*(k, \mathbb{Z}_m^r) = \max_{\gamma \wr m} \tau_\gamma(k, \gamma^r).$$

In the notation of Proposition 6.11, $\zeta_\gamma = ((k - \gamma)/k) \log \gamma \asymp \log \gamma$. Hence, $\zeta_\gamma \gg 1$ if (and only if) $\gamma \gg 1$. If the maximising γ, call it γ_*, satisfies $\gamma_* \gg 1$, then Propositions 2.2 and 6.11 give

$$\max_{\gamma \wr m} \tau_\gamma(k, \gamma^r) = \tau_{\gamma_*}(k, \gamma_*^r) \eqsim \frac{k \gamma_*^{2r/k}}{2\pi e} \eqsim \tau_\infty(k, \gamma_*^r).$$

It remains to show that $\gamma_* = m$. The map $\gamma \mapsto k\gamma^{2r/k}/(2\pi e)$ is increasing. So, if $\gamma_* \gg 1$, then in fact $\gamma_* = m$. Similarly, Proposition 6.11 show that $\tau_\gamma(k, \gamma)/k \lesssim \gamma^{2r/k}$ as $\zeta_\gamma \asymp \log \gamma \gtrsim 1$. Hence,

$$\frac{\tau_\gamma(k, \gamma^r)}{k} \lesssim \gamma^{2r/k} \ll m^{2r/k} \asymp \frac{\tau_m(k, \gamma^r)}{k};$$

thus, $\gamma_* = m$. ∎

Finally, after Corollary D.2 we mentioned that special groups are ubiquitous amongst p-groups of a given size. We elaborate on this claim in the following remark.

Remark 6.12. In the classical work [29], Higman gave upper and lower bounds on the number groups of size p^ℓ for a prime p. The upper bound was later refined by Sims [46]. Together, they show that this number is equal to $p^{(2/27)\ell^3 \pm \mathcal{O}(\ell^{8/3})}$. The lower bound $p^{(2/27)\ell^3 - \mathcal{O}(\ell^2)}$ is obtained from Higman [29, Theorem 2.1] by counting step-2 groups whose Frattini group is equal to the centre and is elementary Abelian of size p^s and of index p^r, where $r = \ell - s$. It is classical that such a group is special if and only if it has exponent p, i.e., every element other than the identity has order p.

Higman in [29] showed that the number of such groups of size p^ℓ is between $p^{(1/2)sr(r+1)-r^2-s^2}$ and $p^{(1/2)sr(r+1)-s(s-1)}$ if $s \le r(r+1)/2$ and 0 otherwise. A small variant of his argument shows that the number of special groups of size p^{s+r} whose commutator subgroup is of size p^s is between $p^{(1/2)sr(r-1)-r^2-s^2}$ and $p^{(1/2)sr(r-1)-s(s-1)}$ for $s \le r(r-1)/2$. (In [29], one includes in F all elements of order p. See also the short argument in Sims [46, p. 152]; there, the change is considering the case that $b(i, j) = 0$ for all $1 \le i \le r$ and $1 \le j \le s$.) Taking r and s such that $|r - 2s| < 3$, we see that the logarithm of the number of groups of size p^ℓ is dominated by special groups.

6.6. *Expander graphs of nilpotent groups with $k \gtrsim \log|G|$*

The isoperimetric constant Φ_* can be defined for general Markov chains; see [35, §7.2]. The isoperimetric constant of a graph is that of the nearest-neighbour RW on the graph. It is easy to see that the isoperimetric constant of a Markov chain, of its time reversal and of its additive symmetrisation are all equal. But, for any generators Z, the additive symmetrisation of the RW on $G^+(Z)$ is the RW on $G^-(Z)$; so, $\Phi_*(G^+(Z)) = \Phi_*(G^-(Z))$. It thus suffices to work with undirected Cayley graphs.

We analyse the spectral gap via considering the $1/n^c$-mixing time for some $c > 0$.

Proposition 6.13. *Let G be a nilpotent group. Suppose that $k - d(\bar{G}) \asymp k \asymp \log|G|$. Let $\varepsilon > 0$ and set $t := (1+\varepsilon)\tau_*^-(k, G)$. Then, there exists a constant $c > 0$ so that $d_{G_k}^-(t) \le |G|^{-c}$ w.h.p.*

Proof. Consider first Abelian G; here, $G = \bar{G}$. Since Hypothesis D is satisfied,

$$d_{G_k}(t) \le e^{-c'(k-d(G))}$$

w.h.p. for some constant $c' > 0$, by Lemma 5.3. The claim now follows as $k - d(G) \asymp \log|G|$ here.

Consider now nilpotent G; here, $G \ne \bar{G}$. We apply our nilpotent-to-Abelian method. There, we upper bounded the modified L_2 distance for the RW on G (at time t) by the modified L_2 distance for the RW on $\bar{G}$ (at time t); see specifically Proposition 6.6, Lemma 6.7 and Corollary 6.9. For Abelian groups, we used the modified L_2 calculation (in Sections 2–4). Thus, the nilpotent case is an immediate application of the nilpotent-to-Abelian method and Abelian case. ∎

We apply Proposition 6.13 along with standard mixing-type results.

Proof of Theorem 6.2. As noted in Remark E, it suffices to consider only $k \asymp \log n$.

First, we use the discrete Cheeger inequality, for reversible Markov chains

$$\gamma \lesssim \Phi_* \lesssim \sqrt{\gamma},$$

where γ is the spectral gap; see, e.g., [35, Theorem 13.10]. Thus, it suffices to show that $\gamma \asymp 1$.

Next, we use a standard relation between the mixing time and spectral gap: for a reversible Markov chain with invariant distribution uniform on n states, mixing time t_{mix} and spectral gap γ,

$$t_{\text{mix}}\Big(\frac{1}{n^c}\Big) \asymp \gamma^{-1} \log n \quad \text{for any constant } c > 0;$$

see [35, Theorem 20.6 and Lemma 20.11]. Thus, $\gamma \asymp 1$ if $t_{\text{mix}}(1/n^c) \lesssim \log n$ for such c.

The mixing claim follows immediately from Proposition 6.13 and that $k \asymp \log|G|$. ■

7. Concluding remarks and open questions

In Section 7.1, we discuss lack of cutoff in the regime where k is a fixed constant. In Section 7.2, we give a very short proof, which is a small variant on Roichman's argument [44, Theorem 2], establishing an upper bound on mixing, for arbitrary groups and any $k \gg \log|G|$. In Section 7.3, we briefly discuss cutoff in other metrics, namely L_2 and relative entropy. In Section 7.4, we discuss some questions which remain open and give some conjectures.

Throughout this section, we only sketch details.

7.1. *Lack of cutoff when k is constant*

Throughout the paper, we have always been assuming that $k \to \infty$ as $|G| \to \infty$. It is natural to ask what happens when k does not diverge. This case has actually already been covered by Diaconis and Saloff-Coste [15], using their concept of *moderate growth*. In short, there is no cutoff.

Diaconis and Saloff-Coste establish this not only for Abelian groups, but for nilpotent groups. Recall that a group G is called *nilpotent of step at most* $L + 1$ if its lower central series terminates in the trivial group after at most L steps: $G_{(0)} := G$ and $G_{(\ell)} := [G_{(\ell-1)}, G]$ for $\ell \in \mathbb{N}$ with $G_L = \{\mathsf{id}\}$.

For a Cayley graph $G(Z)$, use the following notation. Write $\Delta := \operatorname{diam} G(Z)$ for its diameter. For the lazy simple random walk on $G(Z)$, write $t_{\text{rel}} := t_{\text{rel}}(G(Z))$ for the relaxation time (i.e., the inverse of the spectral gap) and $t_{\text{mix}} := t_{\text{mix}}(1/4; G(Z))$ for the (TV) mixing time. When considering sequences $(G_N(Z_{(N)}))_{N\in\mathbb{N}}$, add an N-sub/superscript, analogously to before.

We phrase the result of Diaconis and Saloff-Coste [15] in our language.

Theorem 7.1 (Cf. [15, Corollary 5.3]). *Let $(G_N)_{N \in \mathbb{N}}$ be a sequence of finite, nilpotent groups. For each $N \in \mathbb{N}$, let $Z_{(N)}$ be a symmetric generating set for G_N and write L_N for the step of G_N. Suppose that $\sup_N |Z_{(N)}| < \infty$ and $\sup_N L_N < \infty$. Then,*

$$\frac{t^N_{\mathrm{mix}}}{|Z_{(N)}|} \lesssim \Delta_N^2 \lesssim t^N_{\mathrm{rel}} \lesssim t^N_{\mathrm{mix}} \quad \textit{as } N \to \infty.$$

In particular, $(t^N_{\mathrm{mix}})_{N \in \mathbb{N}}$ does not exhibit the cutoff phenomenon.

We give a very brief exposition of the results of Diaconis and Saloff-Coste [15], including the definition of moderate growth, leading to this conclusion in [25, §4].

7.2. A variant on Roichman's argument

In this subsection, we give a very short argument upper bounding the mixing time for arbitrary groups and $k \gg \log|G|$; it is a small modification of Roichman's argument [44, Theorem 2], but it applies in both the undirected and directed cases. (Roichman [44, Theorem 1] deals with the directed case, but requires additional matrix algebra machinery.)

Theorem 7.2. *Let $\varepsilon > 0$. Let G be a finite group and k an integer with $k \gg \log|G|$ and $\log k \ll \log|G|$. Then, the RW on $G_k^{\pm}$ is mixed w.h.p. at time $(1+\varepsilon)T(k, |G|)$, where*

$$T(k, n) := \frac{\log n}{\log(k/n)} \quad \textit{for } n, k \in \mathbb{N}.$$

In particular, this upper bound $T(k, |G|)$ does not depend on the algebraic structure of the group.

Proof. Let $\varepsilon > 0$ and set $t := (1+\varepsilon)\log|G| / \log(k/\log|G|)$. Note that $1 \ll t \ll k$. Choose some $\omega \gg 1$, diverging arbitrarily slowly; set $t_\pm := \lfloor t(1 \pm \omega/\sqrt{t})\rfloor$ and $L := \omega\lfloor t^2/k\rfloor$. The number of generators picked at most once is at least $k - L$ w.h.p.; of these, the number picked exactly once lies in $[t_-, t_+]$ w.h.p. Take typ to be the event that these two conditions hold for two independent copies, W and W'. We use a modified L_2 calculation (see, e.g., Lemma 2.6) meaning that we need to control

$$D := |G|\mathbb{P}(S = S' \mid W = W', \mathsf{typ}) - 1.$$

Let $\mathcal{E}$ be the event that some generator is used precisely once in W and never in W', or vice versa:

$$\mathcal{E} := \bigcup_{i \in [k]} \left(\{|W_i| = 1, |W_i'| = 0\} \cup \{|W_i'| = 1, |W_i| = 0\}\right).$$

Then, $S' \cdot S^{-1} \sim \mathrm{Unif}(G)$ on $\mathcal{E}$. Indeed, if $Z \sim \mathrm{Unif}(G)$ and $X, Y \in G$ are independent of Z, then $XZY \sim \mathrm{Unif}(G)$; here, Z corresponds to one of the generators used precisely once in W and not in W' or vice versa, with the obvious choice of X and Y so that $XZY = S'S^{-1}$. Off $\mathcal{E}$, every generator picked once in W must be picked at least once in W', and vice versa. There are at most L generators which are picked more than once in W'. Thus,

$$\mathbb{P}(\mathcal{E} \mid \mathsf{typ}) \le \max_{a \in [t_-, t_+], b \le L} \frac{1}{\binom{k-b}{a-b}} = \frac{1}{\binom{k-L}{t_- - L}}.$$

An application of Stirling's approximation shows that this probability is $o(1/|G|)$ when ω diverges sufficiently slowly. Combined with the modified L_2 calculation, this proves the upper bound. ■

Finally, consider the case $k = |G|^\alpha$ for some fixed $\alpha \in (0, 1)$. The discrete-time chain cannot be mixed at time $\lceil 1/\alpha \rceil - 1$ by considering the size of its support, but noting that $\binom{k}{t} \gg |G|$ for $t := \lfloor 1/\alpha \rfloor + 1$, by the above argument we see that the walk is mixed w.h.p. after t steps.

Dou [19] proves a more general statement than this which allows the generators to be picked from a distribution other than the uniform distribution; see [19, Theorems 3.3.1 and 3.4.7].

7.3. Mixing in different metrics

One can also consider cutoff in the L_2 distance. Recall that, for a time $t \ge 0$,

$$d_{G_k}^{(2)}(t) = \|\mathbb{P}_{G_k}(S(t) \in \cdot) - \pi_G\|_{2,\pi_G} = \Big(|G|^{-1} \sum_{g \in G} (|G| \mathbb{P}_{G_k}(S(t) = g) - 1)^2 \Big)^{1/2}.$$

For reasons explained below, we *strongly believe* the following is true – and can be proved in the framework which we have developed in this article. It was stated as Conjecture A in Section 1.2.1.

Conjecture A. *For $\gamma \in \mathbb{Z} \cup \{\infty\}$, let $\tilde{\tau}_\gamma^\pm := \tilde{\tau}_\gamma^\pm(k, G)$ be the time t at which the return probability for RW on $\mathbb{Z}_\gamma^k$ at time $2t$ is $|G|^{-1}$. Set $\tilde{\tau}_*^\pm(k, G) := \max_{\gamma \in \mathbb{N}} \tilde{\tau}_\gamma^\pm(k, G)$. Then, under similar conditions to those of Theorem A, w.h.p., the RW on G_k exhibits cutoff in the L_2 metric at time $\tilde{\tau}_*^\pm(k, G)$.*

The main change is that we now no longer perform a modified L_2 calculation. Replacing $r_* := (|G|^{1/k} (\log k)^2)/2$ with $r_* := (|G|^{1/k} \log|G|)/2$, local typicality then holds with probability $1 - o(1/|G|)$; cf. Proposition 6.13. Thus, we may condition on local typicality as this can only change the L_2 distance by at most an $o(1)$ additive term. On the other hand, we no longer condition on global typicality. Instead, we must handle directly terms like $\mathbb{P}(W = W')$ or $\mathbb{P}(W_\gamma = W'_\gamma)$.

For Approach #1, we must handle a gcd. Under the assumption that $1 \ll k \lesssim \sqrt{\log|G|}$, increasing r_* as we have has little effect on the proof, in essence because $(\log n)^d = n^{o(1)}$. In Approach #2, we replace $|\mathcal{H}|$ by $|\mathcal{H}| \log|G|$, but still $k \gg \sqrt{\log|G|}$ implies that $k \gg \log(|\mathcal{H}| \log|G|)$. Lastly, the combination of the two approaches works when

$$\sqrt{\frac{\log|G|}{\log\log|G|}} \ll k \lesssim \sqrt{\log|G|}.$$

Using somewhat similar adaptations, we believe that cutoff in the relative entropy (abbreviated *RE*) distance can be established. In this case, we quantify the probability with which global typicality holds: the maximal relative entropy of a measure on G with respect to π_G is $\log|G|$; thus, naively at least, to condition on global typicality we desire it to hold with probability $1 - o(1/\log|G|)$ – for L_2 we had $1 - o(1/|G|)$. Also, one should modify local typicality as previously. This gives conditions on k and $d(G)$. Under such conditions, the RE and TV cutoff times should then be the same.

We believe that with more effort these conditions can be improved via obtaining some estimates on the relative entropy given that global typicality fails.

7.4. Open questions and conjectures

We close the paper with some questions which are left open.

(1) *Does the product condition imply cutoff?* The problem of singling out abstract conditions under which the cutoff phenomenon occurs has drawn considerable attention. For a reversible Markov chain X, write $t_{\text{mix}}(X)$ for its mixing time and $\gamma_{\text{gap}}(X)$ for its spectral gap. In 2004, Peres proposed a simple spectral criterion[1] for a sequence $(X^N)_{N \in \mathbb{N}}$ of reversible Markov chains, known as the *product condition*:

$$\text{cutoff is equivalent to } t_{\text{mix}}(X^N)\gamma_{\text{gap}}(X^N) \to \infty \text{ as } N \to \infty.$$

It is well known that the product condition is a necessary condition for cutoff; see, e.g., [35, Proposition 18.4]. It is relatively easy to artificially create counterexamples, but these are not 'natural'; see, e.g., [35, §18], where constructions due to Aldous and due to Pak are described. The product condition is widely believed to be sufficient for most 'natural' chains.

We conjecture that the product condition implies cutoff for random Cayley graph of Abelian groups. In fact, we conjecture this whenever G is *nilpotent* of bounded *step* (denoted step G), i.e., has lower central series terminating at the trivial group and this sequence is of bounded length.

Conjecture 1. *Let $(G_N)_{N \in \mathbb{N}}$ be a sequence of finite, nilpotent group and $(Z_{(N)})_{N \in \mathbb{N}}$ a sequence of subsets with $Z_{(N)} \subseteq G_N$ for all $N \in \mathbb{N}$. For each $N \in \mathbb{N}$, write t^N_{mix} (resp. γ^N_{gap}) for the mixing time (resp. spectral gap) of the SRW on $G^-_N(Z_{(N)})$. Suppose that* $\limsup_{N \to \infty} \text{step}\, G_N < \infty$ *and that the product condition holds, i.e., $t^N_{\text{mix}}\gamma^N_{\text{gap}} \to \infty$ as $N \to \infty$. Then, the sequence of SRWs exhibits cutoff.*

An equivalence between the product condition and cutoff has been established for birth-and-death chains by Ding, Lubetzky and Peres [17] and, more generally, for RWs on trees by Dasu, Hermon and Peres [4]. It is believed to imply cutoff for the SRW on transitive expanders of bounded degree, but this is known only in the case of Ramanujan graphs, due to Lubetzky and Peres [37].

[1] See https://aimath.org/WWN/mixingtimes/mixingtimes.pdf, visited on 1 February 2026.

(2) *An explicit choice of generators?* We have shown that choosing the generators Z uniformly gives cutoff w.h.p. at a time which does not depend on Z, in many regimes. In particular, this means that there is cutoff for almost all choices of generators at a time independent of the choice of generators. This 'almost universal' mixing time is given by $\tau_*(k, G)$ from Definition 3.1. A question raised to us by Diaconis (2019, private communication) is to find *explicit* sets of generators for which cutoff occurs; see also [13, Chapter 4G, Question 2].

Open Problem 1. Let G be an Abelian group and $1 \ll k \lesssim \log|G|$. Find an explicit choice of generators Z (implicitly a sequence) so that the RW on $G(Z)$ exhibits cutoff. Further, find generators so that the cutoff time is $\tau_*(k, G)$ asymptotically.

Hildebrand [33, Theorem 1.11] shows for the cyclic group $\mathbb{Z}_p$ with p prime that the choice $Z := [0, \pm 1, \pm 2, \ldots, \pm 2^{\lceil \log_2 p \rceil - 1}]$, which he describes as 'an approximate embedding of the classical hypercube walk into the cycle', gives rise to a random walk on $\mathbb{Z}_p$ which has cutoff. The cutoff time is not the entropic time, however. Although the entropic time is the mixing time for 'most' choice of generators, finding an explicit choice of generators which gives rise to cutoff at the entropic time is still open – even for the cyclic group of prime order.

Acknowledgements. The random Cayley graphs project has benefited greatly from advice, discussions and suggestions from many of our peers and colleagues. We thank Allan Sly for suggesting the underlying entropy idea for cutoff in Approach #1 (Section 2), Justin Salez for reading this paper in detail and giving many helpful and insightful comments as well as stimulating discussions ranging across the entire random Cayley graphs project, Evita Nestoridi and Persi Diaconis for general discussions, consultation and advice.

Funding. JH was partially supported by EPSRC Grant EP/L018896/1 and an NSERC Grant. SOT was partially supported by EPSRC Grants 1885554 and EP/N004566/1.

References

[1] Aldous, D., Diaconis, P.: Shuffling cards and stopping times. Technical Report 231, Department of Statistics, Stanford University (1985)

[2] Aldous, D., Diaconis, P.: Shuffling cards and stopping times. Amer. Math. Monthly **93**, 333–348 (1986) Zbl 0603.60006 MR 841111

[3] Alon, N., Roichman, Y.: Random Cayley graphs and expanders. Random Structures Algorithms **5**, 271–284 (1994) Zbl 0798.05048 MR 1262979

[4] Basu, R., Hermon, J., Peres, Y.: Characterization of cutoff for reversible Markov chains. Ann. Probab. **45**, 1448–1487 (2017) Zbl 1374.60129 MR 3650406

[5] Ben-Hamou, A., Lubetzky, E., Peres, Y.: Comparing mixing times on sparse random graphs. In: Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1734–1740 (2018) Zbl 1403.05133 MR 3775901

[6] Ben-Hamou, A., Salez, J.: Cutoff for nonbacktracking random walks on sparse random graphs. Ann. Probab. **45**, 1752–1770 (2017) Zbl 1372.60101 MR 3650414

[7] Berestycki, N., Lubetzky, E., Peres, Y., Sly, A.: Random walks on the random graph. Ann. Probab. **46**, 456–490 (2018) Zbl 1393.60077 MR 3758735

[8] Bordenave, C., Caputo, P., Salez, J.: Cutoff at the "entropic time" for sparse Markov chains. Probab. Theory Related Fields **173**, 261–292 (2019) Zbl 1480.60202 MR 3916108

[9] Bordenave, C., Lacoin, H.: Cutoff at the entropic time for random walks on covered expander graphs. J. Inst. Math. Jussieu **21**, 1571–1616 (2022) Zbl 1496.05177 MR 4476123

[10] Chen, S., Moore, C., Russell, A.: Small-bias sets for nonabelian groups: Derandomizations of the Alon–Roichman theorem. In: Approximation, randomization, and combinatorial optimization, Lecture Notes in Comput. Sci. 8096, Springer, Heidelberg, 436–451 (2013) Zbl 1405.68436 MR 3126546

[11] Christofides, D., Markström, K.: Expansion properties of random Cayley graphs and vertex transitive graphs via matrix martingales. Random Structures Algorithms **32**, 88–100 (2008) Zbl 1130.05030 MR 2371053

[12] Conchon-Kerjan, G.: Cutoff for random lifts of weighted graphs. Ann. Probab. **50**, 304–338 (2022) Zbl 1486.05284 MR 4385128

[13] Diaconis, P.: Group representations in probability and statistics. IMS Lecture Notes Monogr. Ser. 11, Institute of Mathematical Statistics, Hayward, CA (1988) Zbl 0695.60012 MR 964069

[14] Diaconis, P.: Some things we've learned (about Markov chain Monte Carlo). Bernoulli **19**, 1294–1305 (2013) Zbl 1412.60109 MR 3102552

[15] Diaconis, P., Saloff-Coste, L.: Moderate growth and random walk on finite groups. Geom. Funct. Anal. **4**, 1–36 (1994) Zbl 0795.60005 MR 1254308

[16] Diaconis, P., Wood, P. M.: Random doubly stochastic tridiagonal matrices. Random Structures Algorithms **42**, 403–437 (2013) Zbl 1278.15035 MR 3068032

[17] Ding, J., Lubetzky, E., Peres, Y.: Total variation cutoff in birth-and-death chains. Probab. Theory Related Fields **146**, article no. 61 (2010) Zbl 1190.60005 MR 2550359

[18] Dou, C., Hildebrand, M.: Enumeration and random random walks on finite groups. Ann. Probab. **24**, 987–1000 (1996) Zbl 0862.60006 MR 1404540

[19] Dou, C. C. Z.: Studies of random walks on groups and random graphs. Ph.D. thesis, Massachusetts Institute of Technology, https://www.proquest.com/docview/304017313, visited on 1 February 2026 (1992)

[20] Hardy, G. H., Wright, E. M.: An introduction to the theory of numbers. 6th ed., Oxford University Press, Oxford (2008) Zbl 1159.11001 MR 2445243

[21] Hermon, J., Huang, X.: Cutoff for random Cayley graphs of nilpotent groups. arXiv:2403.12355v1 (2024)

[22] Hermon, J., Huang, X., Pedrotti, F., Salez, J.: Concentration of information on discrete groups. arXiv:2409.16869v1 (2024)

[23] Hermon, J., Lacoin, H., Peres, Y.: Total variation and separation cutoffs are not equivalent and neither one implies the other. Electron. J. Probab. **21**, article no. 44 (2016) Zbl 1345.60077 MR 3530321

[24] Hermon, J., Olesker-Taylor, S.: Cutoff for random walks on upper triangular matrices. arXiv:1911.02974v2 (2021)

[25] Hermon, J., Olesker-Taylor, S.: Further results and discussions on random Cayley graphs. arXiv:1911.02975v2 (2021)

[26] Hermon, J., Olesker-Taylor, S.: Supplementary material for random Cayley graphs project. arXiv:1810.05130v2 (2021)

[27] Hermon, J., Olesker-Taylor, S.: Geometry of random Cayley graphs of Abelian groups. Ann. Appl. Probab. **33**, 3520–3562 (2023) Zbl 1530.05081 MR 4663490

[28] Hermon, J., Sly, A., Sousi, P.: Universality of cutoff for graphs with an added random matching. Ann. Probab. **50**, 203–240 (2022) Zbl 1486.05278 MR 4385126

[29] Higman, G.: Enumerating p-groups. I: Inequalities. Proc. Lond. Math. Soc. (3) **10**, 24–30 (1960) Zbl 0093.02603 MR 113948

[30] Hildebrand, M.: Random walks supported on random points of $\mathbf{Z}/n\mathbf{Z}$. Probab. Theory Related Fields **100**, 191–203 (1994) Zbl 0809.60011 MR 1296428

[31] Hildebrand, M.: Random lazy random walks on arbitrary finite groups. J. Theoret. Probab. **14**, 1019–1034 (2001) Zbl 0988.60045 MR 1860086

[32] Hildebrand, M.: A survey of results on random random walks on finite groups. Probab. Surv. **2**, 33–63 (2005) Zbl 1189.60016 MR 2121795

[33] Hough, R.: Mixing and cut-off in cycle walks. Electron. J. Probab. **22**, article no. 90 (2017) Zbl 1378.60073 MR 3718718

[34] Landau, Z., Russell, A.: Random Cayley graphs are expanders: a simple proof of the Alon–Roichman theorem. Electron. J. Combin. **11**, article no. 62 (2004) Zbl 1053.05060 MR 2097328

[35] Levin, D. A., Peres, Y.: Markov chains and mixing times. 2nd ed., American Mathematical Society, Providence, RI (2017) Zbl 1390.60001 MR 3726904

[36] Loh, P.-S., Schulman, L. J.: Improved expansion of random Cayley graphs. Discrete Math. Theor. Comput. Sci. **6**, 523–528 (2004) Zbl 1066.68097 MR 2180056

[37] Lubetzky, E., Peres, Y.: Cutoff on all Ramanujan graphs. Geom. Funct. Anal. **26**, 1190–1216 (2016) Zbl 1351.05208 MR 3558308

[38] Lubetzky, E., Sly, A.: Cutoff phenomena for random walks on random regular graphs. Duke Math. J. **153**, 475–510 (2010) Zbl 1202.60012 MR 2667423

[39] Pak, I.: Random Cayley graphs with $O(\log|G|)$ generators are expanders. In: Algorithms—ESA '99 (Prague), Lecture Notes in Comput. Sci. 1643, Springer, Berlin, 521–526 (1999) Zbl 0942.05061 MR 1729149

[40] Pak, I.: Random walks on finite groups with few random generators. Electron. J. Probab. **4**, article no. 1 (1999) Zbl 0918.60061 MR 1663526

[41] Pak, I.: Combinatorics, probability, and computations on groups. Lecture notes (2001)

[42] Pedrotti, F., Salez, J.: A new cutoff criterion for non-negatively curved chains. arXiv:2501.13079v1 (2025)

[43] Pomerance, C.: The expected number of random elements to generate a finite abelian group. Period. Math. Hungar. **43**, 191–198 (2001) Zbl 0980.20079 MR 1830576

[44] Roichman, Y.: On random random walks. Ann. Probab. **24**, 1001–1011 (1996) Zbl 0912.60016 MR 1404541

[45] Salez, J.: Cutoff for non-negatively curved Markov chains. J. Eur. Math. Soc. (JEMS) **26**, 4375–4392 (2024) Zbl 1548.60106 MR 4780485

[46] Sims, C. C.: Enumerating p-groups. Proc. Lond. Math. Soc. (3) **15**, 151–166 (1965) Zbl 0133.28401 MR 169921

[47] Smith, A.: The cutoff phenomenon for random birth and death chains. Random Structures Algorithms **50**, 287–321 (2017) Zbl 1361.60077 MR 3607126

[48] Wilson, D. B.: Random random walks on $\mathbb{Z}_2^d$. Probab. Theory Related Fields **108**, 441–457 (1997) Zbl 0896.60034 MR 1465637