Universality of Cutoff for Random Walks on Random Cayley Graphs

Samuel Mark Thomas

PhD Supervisor: Perla Sousi

Department of Pure Mathematics and Mathematical Statistics
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

St John's College, Cambridge

July, 2020

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared below and specified in the text. It is not substantially the same as any that I have submitted, or is being concurrently submitted, for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted, for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

The entire dissertation is based on joint work with Jonathan Hermon, who was a post-doctoral researcher at the University of Cambridge and is currently at the University of British Columbia (in Vancouver). The research papers on which this these are based are joint work with Jonathan Hermon. Ideas were contributed in roughly equal measure; I did the writing.

Acknowledgements

There are two people who deserve the most acknowledgement for this work. One is my supervisor, Perla Sousi. She has supported me throughout my PhD, and I am very grateful for her guidance. In particular, the first project which I undertook was with her. I was very raw and rough around the edges as a budding research mathematician, and her support and advice was invaluable. She has dedicated enormous amounts of her time to my professional (and personal) development.

The other is Jonathan Hermon, who initially approached me with the idea of the topic which has become my entire thesis. I quote from his initial email dated early February 2018:

> *"I have a project that I think you might be interested in. [...] An elementary approach should work and hopefully this can be a short paper..."*

As of July 2020, we are just about to submit four research papers all to top quality journals (with an additional paper of results and a supplementary file on the arXiv). I think that we can all agree that *"this can be a short paper"* did not happen; of this, I am very glad. Throughout these two and a half years, Jonathan has been like a second supervisor to me. As someone who has himself just recently finished his PhD, he was able to offer insights and advice from a different point of view to that of Perla. Because of my long-term collaboration with him, my PhD thesis is *far better* than it ever would have been otherwise. To him, I am truly grateful.

The CCA crew have adopted me as one of their own, allowing me to attend the events with food and do none of the administrative stuff—really the best of both worlds; for this, particular thanks goes to Tessa. Thank you to Fritz and Kweku for your detailed lunch time discussions on how to run train networks, amongst many other less-than-productive topics, in which you pushed your socialist agenda particular in the face of Andrew's ultra-capitalist mindset. Eardi: thank you for your Italian advice, and for being on your phone most of the time during these (some would say pointless) discussions. Since becoming my officemate, Andrew and I have been known to have long discussions about fairly random stuff—perhaps neither of us would have finished our PhDs yet if it were not for this pandemic's forcing separation? A highlight was when, along with Kweku, Andrew and others, I entered the Cardboard Boat Race. This involved Andrew's buying a significant amount of DIY equipment. Many hours of drilling, sawing and screwing later, we created a (frankly superb) propulsion system. Then it came to putting the boat in the water, getting in and letting go of the bank; what happened next is anyone's guess... (youtu.be/M8We-TTgwLo).

To the quantum group (plus Ed) who invaded Pavilion D because it is demonstratively better than Pavilion E: it has been a pleasure having you. Learning Italian with Sathya made it far more interesting. Once she decided to stop teasing me at every opportunity (usually about my lack of knowledge of physics or physicists), it turned out that Mithuna is actually a really nice person!

Last but not least, to Guillaume, my officemate for six months prior to Andrew: it was superb having you in Cambridge, including the hundreds of crumpled up sheets of paper which you threw at the recycling bin to "estimate my p". (To refer to this as "threw *in*" would imply larger p-value than I feel is representative of reality.) I hope our paths cross again in the not too distant future!

To all the fellow students and researchers whom I have met throughout the years at conferences, summer schools and the like: probability is a wonderful community. I have always felt most welcome and have had an excellent time at these events; they have been some of the highlights of my PhD.

My family has always supported my quest for higher knowledge. In particular my soon-to-be wife Emily: as I searched (high and low) for a post-doctoral research position, you supported me, even though that may mean moving (and leaving your job) quite possibly for less than two years before onto a new place and a new job. As a thank you for your never-ending support, I shall only *occasionally* teasingly refer to us as "Dr and Mrs"—your retort that you are far more widely employable and earn far more than me will forever fall on deaf ears... and in our *joint* bank account.

I gratefully acknowledge the financial support of the EPSRC (Doctoral Training Award 1885554) and of St John's College. My co-author Jonathan was also supported by an EPSRC grant. Perla allowed me access to her EPSRC grant on occasions; this has been received with much gratitude.

Universality of Cutoff for Random Walks on Random Cayley Graphs

Abstract

Samuel Mark Thomas

Consider the random Cayley graph of a finite group G with respect to k generators chosen uniformly at random. This draws a Cayley graph uniformly amongst all degree-k Cayley graphs of G. A conjecture of Aldous and Diaconis [1] from the '80s asserts, for $k \gg \log|G|$, the following:

- the random walk on this (random) graph exhibits *cutoff* with high probability (*whp*);
- the cutoff time depends only on k and $|G|$ asymptotically (up to smaller order terms).

The cutoff time should not depend (strongly) on the choice of generators. In other words,

cutoff is universal for the random walk on the random Cayley graph.

Restricted to Abelian groups, this was verified in the '90s; the cutoff time $T(k,|G|)$ was found explicitly. In fact, $T(k,|G|)$ was shown to be to be an upper bound on mixing for arbitrary groups.

First we extend the conjecture to $1 \ll k \lesssim \log|G|$. Write $d(G)$ for the minimal size of a generating set of G. We establish cutoff for (the random walk on) all Abelian group under the condition $k - d(G) \gg 1$, verifying the occurrence of cutoff part of the Aldous–Diaconis conjecture. This condition is almost optimal to guarantee that the group if generated whp. For the cutoff time to depend only on k and $|G|$, not the algebraic structure of G, we show that $d(G) \ll \log|G|$ and $k - d(G) \asymp k \gg 1$ is sufficient. However, the result does not hold if $k \asymp \log|G| \asymp d(G)$; there are even regimes with $1 \ll k \ll \log|G|$ for which it does not hold if we allow $1 \ll k - d(G) \ll k$.

Next we consider the (non-Abelian) Heisenberg group $H \coloneqq H_{p,d}$ of $d \times d$ matrices with entries in $\mathbb{Z}_p$, with p prime and $d \geq 3$ not diverging too quickly. We establish cutoff for any $k \gg 1$ with $\log k \ll \log|H|$. Except for k growing super-polylogarithmically in $|G|$ (ie $\log k \gg \log\log|G|$), this is the first example where cutoff has been established for any non-Abelian group. Further, even restricting to $k \gg \log|H|$, the cutoff time cannot be written as a function only of k and $|G|$; rather, one needs $|H^{\mathrm{ab}}|$, the size of the Abelianisation, also. In fact, taking $d \to \infty$ sufficiently slowly, the mixing time is of smaller order (not just a constant smaller) than $T(k,|H|)$, the universal upper bound. When $k \gtrsim \log|H^{\mathrm{ab}}|$, we can remove the primality assumption on p.

Our next sequence of results still regards mixing, but this time determines upper bounds which hold for large classes of groups, rather than establishing cutoff. From a nilpotent group G, we construct an Abelian group $\overline{G}$ (from the lower central series of G) of the same size. We show that the mixing time for G is at least as fast (asymptotically) as that for $\overline{G}$ whp.

Wilson [77] conjectured that, amongst all groups of size at most 2^d, the group $\mathbb{Z}_2^d$ gives rise to the slowest mixing time. When restricted to Abelian groups, we deduce this from the explicit description of the mixing time which we obtain. As a corollary of the above nilpotent-to-Abelian comparison, this is extended from the Abelian to the nilpotent set-up.

The spirit of the Aldous–Diaconis conjecture is that the certain properties of the random Cayley graph should depend very weakly on the choice of generators. We apply this principle to geometric aspects of the graph. Primarily we study the *typical distance*: draw $U \sim \mathrm{Unif}(G)$ and consider $\mathrm{dist}(\mathsf{id}, U)$, where $\mathsf{id} \in G$ is the identity and dist is the graph distance.

We show that the typical distance concentrates whp for Abelian groups. We establish this for all Abelian groups when either $1 \ll k \ll \log|G|/\log\log\log|G|$ and $k - d(G) \asymp k$ or $k \gg \log|G|$; for k in the interim regime or smaller $1 - d(G)/k$, we need additional conditions. Further, the concentration value depends only on k and $|G|$ in the former cases. We study typical distance for Heisenberg groups, proving analogous results. Again, the value depends on $|H^{\mathrm{ab}}|$ as well as k and $|H|$.

For $k \gtrsim \log|G|$, we can extend the typical distance results to show that the *diameter* of the graph agrees asymptotically with the typical distance whp. (For $H_{p,d}$, we need $d \asymp 1$ for this.)

Finally, we find the order of the *spectral gap* when the underlying group is Abelian: it is $|G|^{2/k}$ whp when $1 \ll k \lesssim \log|G|$ and $k - d(G) \asymp k$. This extends, in the Abelian set-up, a celebrated result of Alon and Roichman [3] which states that for any group the random Cayley graph is an *expander*, ie has spectral gap order 1, whp when $k - \log_2|G| \asymp k$.

Table of Contents for Full Thesis

1 Introduction, Results and History

Table of Contents for Chapter 1

1.1 Random Cayley Graphs

Consider a finite group G. Let Z be a multisubset of G, called the *generators*. We consider the (nearest-neighbour) random walk (abbreviated *RW* and denoted $S = (S(t))_{t\geq 0}$) on the *Cayley graph* of (G, Z). (Here and throughout, unless otherwise specified explicitly, time is continuous.) The *undirected*, respectively *directed*, *Cayley graph of G generated by Z*, denoted $G^-(Z)$, respectively $G^+(Z)$, is the multigraph whose vertex set is G and whose edge multiset is

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

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

We focus attention on the *random* Cayley graph defined by choosing $Z_1, ..., Z_k \sim^{\text{iid}} \text{Unif}(G)$. When this is the case, we denote $G_k^+ := G^+(Z)$ and $G_k^- := G^-(Z)$. Introduced by Aldous and Diaconis [1], there has been a great deal of research into these "random random walks". Motivation for this model, along with an overview of historical work, is given in §1.5.

This procedure corresponds to choosing a Cayley graph of a given degree uniformly at random; our results then hold "for almost all Cayley graphs". See §1.6.1 for more details.

1.2 Definitions of Statistics and the Aldous–Diaconis Conjecture

Before diving into our results, we make precise the statistics we study. In particular, the results will be "with high probability over the random graph", made precise in §1.2.1. We also discuss briefly a conjecture of Aldous and Diaconis [1], which is the inspiration for this entire thesis.

1.2.1 With High Probability Over the Random Graph

For a group (or set) G, denote by π_G the uniform distribution on G. This is invariant for the RW on any Cayley graph—any generators; both directed and undirected. Further, if the Cayley graph is connected, then it is the unique invariant distribution.

The graph, clearly, depends on the choice of generators, ie of the multiset Z. Sometimes we want to emphasise this: we add a subscript, eg writing $\mathbb{P}_{G(z)}(S(t) \in \cdot)$ for the law of $S(t)$, ie the RW at time t, on the graph $G(z)$. Analogously, we write $\mathbb{P}_{G_k}(S(t) \in \cdot)$ for the *random law* corresponding to the *random choice* of $Z = [Z_1, ..., Z_k]$ with $Z_1, ..., Z_k \sim^{\text{iid}} \text{Unif}(G)$.

All the results appearing in the introduction are for sequences $(G_N)_{N\in\mathbb{N}}$ of finite groups with $|G_N| \to \infty$. (With some additional care, they can be turned into statements about a fixed group G with explicit error terms.) 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 $n := |G|$, k, d and so on appearing in the statements are all implicitly sequences; eg "$k - d \gg 1$" means that the sequence $(k_N, d_N)_{N\in\mathbb{N}}$ satisfies $k_N - d_N \to \infty$ as $N \to \infty$. Similarly, we say that an event (implicitly a sequence of events) holds *with high probability* (abbreviated *whp*) if its probability tends to 1 in the limit. Typically (but not always), our results are "whp over Z" statements: they hold with probability (over the randomness in Z) tending to 1 as $|G| \to \infty$.

1.2.2 Mixing Time and Cutoff

For two probability measures μ and π on a common (finite) space Ω, we define the *total variation* (abbreviated *TV*) *distance* between μ and π by

$$\|\mu - \pi\|_{\text{TV}} := \max_{A\subseteq\Omega} \big|\mu(A) - \pi(A)\big| = \tfrac{1}{2}\textstyle\sum_{x\in\Omega} \big|\mu(x) - \pi(x)\big|.$$

We specialise this to our application: for a multiset z, set

$$d_{G(z)}(t) := \big\|\mathbb{P}_{G(z)}\big(S(t) \in \cdot\big) - \pi_G\big\|_{\text{TV}}.$$

The ergodic theorem says that an irreducible Markov chain on a finite state spaces has a unique invariant distribution and furthermore the law of the chain converges to this invariant distribution.

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*

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

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 $(G_N)_{N\in\mathbb{N}}$ exhibits cutoff around time $(t_N)_{N\in\mathbb{N}}$ whp 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 G_N.

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

$$s_{G(z)}(t) := \max_{g\in G}\big\{1 - n\,\mathbb{P}_{G(z)}\big(S(t)=g\big)\big\}.$$

One can then define mixing and cutoff 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, eg, [49, Lemmas 6.16 and 6.17]. However, Hermon, Lacoin and Peres [41, Theorem 1.1] showed that TV and separation cutoff are not equivalent, and that neither one implies the other.

1.2.3 Typical Distance and Diameter

For a graph H, write $\mathrm{dist}_H(x,y)$ for the graph distance between two vertices $x,y\in H$.

Definition. *For a group G, generators z, $R\geq 0$ and $\beta\in(0,1)$, write*

$$\mathcal{B}_{G(z)}(R) := \{x\in G \;\big|\; \mathrm{dist}_{G(z)}(\mathsf{id},x)\leq R\} \quad \text{and} \quad \mathcal{D}_{G(z)}(\beta) := \min\big\{R\geq 0 \;\big|\; |\mathcal{B}_{G(z)}(R)|\geq \beta|G|\big\};$$

the diameter ie the maximal distance between pairs of vertices, is given by $\mathrm{diam}\,G(z) := \mathcal{D}_{G(z)}(1)$.

Investigating this typical distance for G_k when k diverges with $|G|$ was suggested to us by Benjamini [10]. Previous work concentrated on fixed k, ie independent of $|G|$; see §1.2.3.

1.2.4 On the Worst-Case Groups

We show that the group $\mathbb{Z}_2^d$ gives rise to the largest mixing time and largest diameter in the random Cayley amongst all groups of size at most 2^d, up to subleading order terms. These are random sequences. We make this precise in the following definition.

Definition. *For two random sequences $\alpha := (\alpha_N)_{N\in\mathbb{N}}$ and $\beta := (\beta_N)_{N\in\mathbb{N}}$ of reals, we say that $\alpha\leq\beta$ whp up to smaller order terms if there exist non-random sequences $(\gamma_N)_{N\in\mathbb{N}}$ and $(\delta_N)_{N\in\mathbb{N}}$ of reals with $\delta_N\to 0$ as $N\to\infty$ such that $(\{\alpha_N\leq(1+\delta_N)\gamma_N\})_{N\in\mathbb{N}}$ and $(\{(1-\delta_N)\gamma_N\leq\beta_N\})_{N\in\mathbb{N}}$ both hold whp. We say that $\alpha\approx\beta$ whp if $\alpha\leq\beta$ and $\beta\leq\alpha$ whp up to smaller order terms.*

We sometimes refer to these as subleading order terms. In either case, we abbreviate as sot.

For generators z and $\varepsilon\in(0,1)$, write $t_{\mathrm{mix}}(\varepsilon;G(z)) := \inf\{t\geq 0 \mid d_{G(z)}(t)\leq\varepsilon\}$ for the ε-mixing time (in TV) of the RW on the graph $G(z)$. We tend to suppress the ε from the notation.

1.2.5 Spectral Gap

Definition. *Consider a reversible Markov chain with (real) eigenvalues $1=\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_n\geq -1$ of its transition matrix. The usual, respectively absolute, spectral gap is defined as*

$$\gamma := \max_{i\neq 1}\{1-\lambda_i\} = 1-\lambda_2, \quad \textit{respectively,} \quad \gamma_* := \max_{i\neq 1}\{1-|\lambda_i|\} = 1-\max\{|\lambda_2|,|\lambda_n|\};$$

the usual, respectively absolute, relaxation time is defined as $t_{\mathrm{rel}} := 1/\gamma$ and $t^*_{\mathrm{rel}} := 1/\gamma_*$.

By the spectral gap or relaxation time of a graph, we mean that of the SRW on the graph.

1.2.6 Heisenberg Matrix Groups

In general we use as our underlying group one of the following two classes of groups: either general Abelian groups, or Heisenberg matrix groups. The latter, which we denote $H_{m,d}$ for integers m and d, is all $d \times d$ upper triangular matrices with 1s on the diagonal and entries in $\mathbb{Z}_m$. For a group G, denote by $G^{\text{com}} := [G, G]$ its *commutator* and by $G^{\text{ab}} := G/G^{\text{com}}$ its *Abelianisation*. We have $H_{m,d}^{\text{ab}} \cong \mathbb{Z}_m^{d-1}$, corresponding to the $d-1$ super-diagonal entries.

1.2.7 Aldous–Diaconis Conjecture

The following conjecture, made by Aldous and Diaconis in the '80s, is the underlying inspiration for the entirety of this thesis.

Conjecture (Aldous and Diaconis [1, Page 40]). *For any group G, if $k \gg \log|G|$ and $\log k \ll \log|G|$, then the random walk on G_k exhibits cutoff whp. 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 $|G|$.*

An informal, more general, variant is reiterated by Diaconis [23, Chapter 4G, Question 8]. The Aldous–Diaconis conjecture was verified for Abelian groups in the '90s, primarily by Dou and Hildebrand [34, 44]; see §1.5.2 for more details. Further, the upper bound holds for arbitrary groups. We also consider a version of the conjecture adapted from *mixing* to *typical distance* and *diameter*.

To extend the 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 via any choice of Z. A general condition which one can impose so that the group is generated whp is that $k - d(G) \gg 1$; see Pomerance [67]. We impose this condition. Note that it is a necessary condition if $d(G) \asymp \log|G|$. One could also impose $k - d(G) \asymp k$—this is particularly relevant for the Aldous–Diaconis conjecture; see Remark A.1 below.

The underlying approach is to use the *entropic method*, the main idea of which is to use an auxiliary process W to generate S; one then studies the entropic properties of the process W. For each $i \in [k]$, ie each generator index, let $W_i(t)$ be the number of times that indeed i has been chosen and the step $g \mapsto gZ_i$ is taken minus the number of times $g \mapsto gZ_i^{-1}$ is taken. (For directed graphs, the latter never happens.) Thus W is a SRW/DRW on $\mathbb{Z}^k$ in the un/directed case.

Observe that $S(t)$ is a function of $(W(t'))_{t' \le t}$. Further, if the underlying group is Abelian, then we have $S(t) = W(t) \cdot Z$. This is a projection though: we could have $W(t) \neq W'(t)$ but $S(t) = S'(t)$, even for Abelian groups. We describe in detail the entropic method, both its history and our application, in §1.4. We describe the high-level idea further in Remarks A.2 and C.2.

Relatedly, Wilson [77] conjectured that the group $\mathbb{Z}_2^d$ gives rise to the slowest mixing time.

1.3 Summarised Statements of Results

In this section we list the results in summarised form. More refined statements are given later.

1.3.1 Cutoff for Random Walks on Cayley Graphs

Our first sequence of results considers mixing properties random walks on the random Cayley graphs. We are interested in the existence of cutoff (§1.2.2). We start by analysing general Abelian groups, before considering non-Abelian Heisenberg matrix groups (§1.2.6). We contrast the results, in particular in the context of the Aldous–Diaconis conjecture (§1.2.7).

1.3.1.1 Cutoff for Almost All Random Walks on Abelian Groups

Our first result establishes cutoff for the random walk on all Abelian groups. It will be the case that whenever $k - d(G) \asymp k$ and $d(G) \ll \log|G|$, the mixing time depends, up to subleading order, only on k and $|G|$. We propose as the mixing time the following entropic time t_*.

Definition A. *For $\gamma \in \mathbb{N} \cup \{\infty\}$, let $t_\gamma^\pm := t_\gamma^\pm(k, G)$ be the time at which the entropy of rate-1 RW (ie 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 $\infty G := |G|G = \{\mathsf{id}\}$. Set $t_*^\pm := t_*^\pm(k, G) := \max_{\gamma \in \mathbb{N}} t_\gamma^\pm(k, G)$.*

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] as well as the informal question asked by Diaconis [23]; we discuss the secondary part (time depending only on k and $|G|$) in Remark A.1.

As mentioned above, cutoff has already been established for Abelian group when $k \gg \log|G|$ with $\log k \ll \log|G|$; see §1.5.2. 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 [77] and $\mathbb{Z}_p$ with p prime in [46]. More refined statements are given in Theorems 2.1.4, 2.2.6 and 2.3.1; see also Hypotheses A to C.

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, whp over Z, the RW on $G_k^\pm$ exhibits cutoff at time $t_*^\pm(k, G)$.*

Moreover, if $k - d(G) \asymp k$ and $d(G) \ll \log|G|$, then $t_*(k, G) \approx t_\infty(k, |G|) \approx k|G|^{2/k}/(2\pi e)$. If $k > d(G)$, then $t_*(k, G) \lesssim k|G|^{2/k}\log k$. If $k \asymp \log|G| \asymp d(G)$, then $t_*(k, G) \asymp k|G|^{2/k}$.

Remark A.1. When $d(G) \ll \log|G|$ and $k - d(G) \asymp k$, one can check that $t_*(k, G)$ is the same as the time at which the entropy of rate-1 RW on $\mathbb{Z}^k$ is $\log|G|$. When $1 \ll k \ll \log|G|$, this is $k|G|^{2/k}/(2\pi e)$, up to sot; see Proposition 2.2.2b. This means that the Aldous–Diaconis conjecture is verified in full for Abelian groups when $d(G) \ll \log|G|$ and $k - d(G) \asymp k$.

However, when $k \asymp \log|G| \asymp d(G)$, while cutoff is still exhibited whp, the cutoff *time* does not depend only on k and $|G|$. Eg, if $k \approx 2\log(4^r)$, then $\mathbb{Z}_2^{2r}$ and $\mathbb{Z}_4^r$ give rise to mixing times which differ by a constant factor. There are even regimes with $1 \ll k \ll \log|G|$ where the claim does not hold, provided $1 \ll k - d(G) \ll k$; see Proposition 5.2.2 and Theorem 5.2.4 where $\mathbb{Z}_p^d$ is studied.

From the definition of t_*, it is not difficult to see that amongst Abelian groups $\mathbb{Z}_2^d$ is the slowest:

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

For $k \gg \log|G|$, cutoff has been established for all Abelian groups, at an explicit time, and this time is an upper bound on mixing for arbitrary (not just Abelian) groups; see §1.5.2. It is not difficult to show the explicit time given is the same as $t_*(k, G)$; see, eg, Proposition 6.2.19. △

Remark A.2. 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, ie of *which* k generators are used. We then 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 §1.4.

Since the group is Abelian, if two walks W and W' on the free group satisfy $W(t) = W'(t)$, then the corresponding projections S and S' to the Cayley graph satisfy $S(t) = S'(t)$. However, the converse is not true. Key is to analyse $\mathbb{P}(S(t) = S'(t) \mid W(t) \neq W'(t))$ for independent W and W'. This is the only place in which we use the uniformity of the generators Z; see Lemma 2.1.11. △

Remark A.3. 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 §2.1, §2.2 and §2.3, respectively. In the first two a concept of *entropic times* is defined; see §2.1.1 and §2.2.2. A precise statement for each approach is given; see §2.1.3, §2.2.4 and §2.3.1. In summary, Theorem A is a direct consequence of Propositions 2.1.2 and 2.2.2 and Theorems 2.1.4, 2.2.6 and 2.3.1; see also Hypotheses A to C. △

Remark A.4. In the first approach, as well as establishing cutoff, we find the *limit profile*: we define entropic times t_α and show that $d_{G_k}(t_\alpha) \to^{\mathbb{P}} \Psi(\alpha)$, where Ψ is the standard Gaussian tail; see Definition 2.1.1, Proposition 2.1.2 and Theorem 2.1.4. If $k - d(G) \asymp k$, then this approach applies for all $1 \ll k \ll \log|G|/\log\log\log|G|$; see Hypothesis A for general conditions. △

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

$$s_{G(z)}(t) \coloneqq \max_{g \in G}\{1 - n\,\mathbb{P}_{G(z)}(S(t) = g)\}.$$

One can then define mixing and cutoff 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, eg, [49, Lemmas 6.16 and 6.17]. However, Hermon, Lacoin and Peres [41, Theorem 1.1] showed that TV and separation cutoff are not equivalent, and that neither one implies the other.

We analyse the regime $k - d(G) \asymp k \gtrsim \log|G|$; in this regime, we show that separation cutoff occurs, and moreover that the cutoff time is the same, up to subleading order, as for TV.

A more refined statement is given in Theorem 2.4.1; see also Hypothesis D.

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

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.1, the slowest amongst Abelian groups for separation mixing is $\mathbb{Z}_2^d$. △

1.3.1.2 Cutoff for Random Walks for Heisenberg Groups

Our next result establishes cutoff for the random walk on Heisenberg matrix groups $H_{p,d}$ (§1.2.6. We propose as the mixing time the following (adjusted) entropic time t_*.

Definition C. *Let $t_0^{\pm}(k, N)$ be the time at which the entropy of rate-1 RW (ie SRW or DRW, as appropriate) on $\mathbb{Z}^k$ is $\log N$. Define $t_*^{\pm}(k, p, d) \coloneqq \max\{t_0^{\pm}(k, |H_{p,d}^{\mathrm{ab}}|),\ \log_k|H_{p,d}|\}$.*

A description of $t_0^{\pm}(k, |H_{p,d}^{\mathrm{ab}}|)$, up to sot, can be found in Proposition 3.1.2. A more refined statement than the one below is given in Theorem 3.1.6; see also Hypothesis E.

Theorem C (Cutoff). *Let p be prime and $d \geq 3$. Let $H \coloneqq H_{p,d}$ and $A \coloneqq H_{p,d}^{\mathrm{ab}}$. Recall that $|H| = p^{d(d-1)/2}$ and $|A| = p^{d-1}$. Assume that $1 \ll \log k \ll \log|H|$ and that one of the following holds:*

- *d is fixed;*
- *$1 \ll k \leq \log|A|/\log d$ and $1 \ll d^3 \ll k$;*
- *$k \gtrsim \log|A|$ and $\log d \ll \log\log p$.*

Whp, the RW on $H_k^{\pm}$ exhibits cutoff at $t_*^{\pm}(k, p, d) = \max\{t_0^{\pm}(k, |A|),\ \log_k|H|\}$. Moreover,

$$t_*^{\pm}(k, p, d) \asymp \begin{cases} t_0^{\pm}(k, |A|) & \text{when} \quad k \leq (\log|A|)^{1+2/(d-2)}, \\ \log_k|H| & \text{when} \quad k \geq (\log|A|)^{1+2/(d-2)}. \end{cases}$$

Remark C.1. While the cutoff time for $H_{p,d}$ cannot be written as a function only of k and $|H_{p,d}| = p^{d(d-1)/2}$, the only additional information required is the size of the Abelianisation, ie $|H_{p,d}^{\mathrm{ab}}| = p^{d-1}$. In Open Question 1 we discuss potential generalisations of this phenomenon. △

Remark C.2. The Abelianisation $H_{p,d}^{\mathrm{ab}}$ is isomorphic to $\mathbb{Z}_p^{d-1}$; it corresponds to the super-diagonal of the matrices. Roughly, we split the analysis into "the mixing of the Abelianisation" and "the mixing of the commutator (ie 'non-Abelian part')". The structure of the proof is the same for all k, except in bounding one specific (combinatorial) probability.

In §2, we study cutoff when the underlying group G is an arbitrary Abelian group. The proof goes via lifting the walk S on the Cayley graph to a walk W on the free Abelian group with k generators (namely $\mathbb{Z}^k$). The mixing time is then the time at which W has entropy $\log|G|$. We perform some analysis on W before projecting back to S.

It may seem natural here, then, to lift the random walk to the free nilpotent group of class $d-1$ (ie the nilpotency class of $H_{p,d}$) and to take as the candidate mixing time the time at which this walk has entropy $\log|H_{p,d}|$. To the best of our knowledge, the idea of studying RWs on nilpotent groups by lifting the walk to a corresponding free nilpotent group was first used by Diaconis and Saloff-Coste [28]. Interestingly, though, instead we still consider a walk on the free *Abelian* group,

but now the candidate mixing time is the time at which this walk has entropy $\log|H_{p,d}^{\mathrm{ab}}|$. At this time, by our results in §2, the walk on the Cayley graph projected to the Abelianisation has mixed.

Naturally we require the mixing time to be at least $\log_k|H_{p,d}|$ so that all vertices can be reached with reasonable probability. We consider the maximum of this entropic time with $\log_k|H_{p,d}|$. △

Remark C.3. Heisenberg groups are a canonical class of nilpotent groups. Our analysis extends to other nilpotent groups; see §3.3.2 for a brief overview and §5.3 for more details. Hence this article is a first step towards establishing cutoff for other nilpotent groups. This is work in progress. △

We adapt the proof of Theorem C to prove two related results, given as Theorem D below.

D.1 When $1 \ll k \ll \log|H_{p,d}^{\mathrm{ab}}|$, we find the limit profile of the cutoff.

D.2 When $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$, we remove the condition that p is prime.

The adaptations to the proof are described in §3.1.9.

Theorem D.1. *Let p be prime and $d \geq 3$ a fixed constant. Assume that $1 \ll k \ll \log|H_{p,d}^{\mathrm{ab}}|$. There exist times $(t_\alpha)_{\alpha\in\mathbb{R}}$ satisfying*

$$t_0 \approx k \cdot \tfrac{1}{2\pi e}|H_{p,d}^{\mathrm{ab}}|^{2/k}, \quad t_\alpha - t_0 \approx \alpha\sqrt{2}t_0/\sqrt{k} = o(t_0) \quad \text{and} \quad d_{H_k}^{\pm}(t_\alpha) \approx \Psi(\alpha) \text{ whp},$$

where Ψ is the standard Gaussian tail, ie $\Psi(\alpha) := (2\pi)^{-1/2}\int_\alpha^\infty e^{-x^2/2}dx$ for $\alpha \in \mathbb{R}$.

Theorem D.2. *Let $m, d \in \mathbb{N}$ with $d \geq 3$. Suppose that $k \gtrsim \log|H_{m,d}|$. If d is a constant or diverges sufficiently slowly, then the RW on $(H_{m,d})_k^{\pm}$ exhibits cutoff whp. Further, there is a density-1 set $\mathbb{A} \subseteq \mathbb{N}$ so that if $m \in \mathbb{A}$ and $\log d \ll \log\log m$ then the RW on $(H_{m,d})_k^{\pm}$ exhibits cutoff whp.*

Remark D. We use the same techniques in §2.1 to find the limit profile when the underlying group G is Abelian. (There the regime is $1 \ll k \ll \log|G| = \log|G^{\mathrm{ab}}|$.) The density-1 set $\mathbb{A}$ comes from a number-theoretic result. For $m \in \mathbb{N}$, if $\operatorname{div} m$ is the number of divisors of m, then 'typically' $\operatorname{div} m$ is order $\log m$; see [40, §18]. We choose $\mathbb{A} := \{m \in \mathbb{N} \mid \operatorname{div} m \leq (\log m)^2\}$. △

1.3.2 Mixing Time Comparison for Nilpotent Groups

The previous results established cutoff. The next results are of a slightly different flavour. They consider *nilpotent* groups: these are groups G whose *lower central series*, ie 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.

We establish a conjecture of Wilson [77] in the nilpotent set-up; see [77, Conjecture 7].

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

As noted in Remark A.1, for Abelian groups this follows from our cutoff result and the abstract entropic definition of the cutoff time $t_*(k, G)$ for Abelian G. The extension to nilpotent groups is then established by Theorem F below, which is of independent interest and quite significantly stronger than Wilson's conjecture.

Theorem F. *Let G be a nilpotent group. Set $\overline{G} := \oplus_1^L (G_{\ell-1}/G_\ell)$ where $(G_\ell)_{\ell\geq 0}$ is the lower central series of G and $L := \min\{\ell \geq 0 \mid G_\ell = \{\mathsf{id}\}\}$. Suppose that $1 \ll \log k \ll \log|G|$ and $k - d(\overline{G}) \gg 1$. Then $t_{\mathrm{mix}}(G_k)/t_{\mathrm{mix}}(\overline{G}_k) \leq 1 + o(1)$ whp.*

Remark F.1. Wilson's conjecture requires $k - \log_2|G| \gg 1$ and compares $t_{\mathrm{mix}}(G_k)$ with $t_*(k, \mathbb{Z}_2^d)$. We have $d(\overline{G}) \leq \max\{\ell \in \mathbb{N} \mid p^\ell \text{ divides } |G| \text{ for some prime } p\} \leq \log_2|G|$; often $d(\overline{G})$ is much smaller than $\log_2|G|$. (In fact, in some precise sense of choosing an Abelian group H uniformly, typically $d(H) \ll \log_2|H|$.) Further, $t_{\mathrm{mix}}(G_k)$ may be significantly smaller than $t_*(k, \mathbb{Z}_2^d)$.

The bounds on $t_*(k, \overline{G})$, for Abelian $\overline{G}$, described in Theorem A complement the upper bound $t_{\mathrm{mix}}(G_k) \leq t_{\mathrm{mix}}(\overline{G}_k)$ to give explicit bounds on $t_{\mathrm{mix}}(G_k)$ which hold whp. △

Remark F.2. In the course of proving Theorem F, we prove an *exact* relation between the L_2 mixing time for the RWs on G_k and $\overline{G}_k$, namely that the expected L_2 distance for the RW on G_k at time t is at most that for the RW on $\overline{G}_k$ at time t. We actually prove a more refined version of this which allows us to compare the expected L_2 distances given that $W(t)$ lies in some 'typical set'. We use such a modified L_2 calculation in the proof of Theorem A to upper bound the TV mixing time. From these two considerations combined, Theorem F then follows. △

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

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

Conjecture F. *Let G be a group. Let $(G_\ell)_{\ell\geq 0}$ be its lower central series and $L := \min\{\ell \geq 0 \mid G_\ell = \{\mathsf{id}\}\}$. Let the prime decomposition of $|G_L|$ be $|G_L| = \prod_1^r p_j$. Set $\overline{G} := (\oplus_1^L (G_{\ell-1}/G_\ell)) \oplus (\oplus_1^r \mathbb{Z}_{p_j})$. Suppose that $1 \ll \log k \ll \log|G|$ and $k - d(\overline{G}) \gg 1$. Then $t_{\mathrm{mix}}(G_k)/t_{\mathrm{mix}}(\overline{G}_k) \leq 1 + o(1)$ whp.*

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

1.3.3 Geometry of Random Cayley Graphs

We move onto considering distances in the random Cayley graphs. We start by analysing typical distance and diameter for general Abelian groups. Then we analyse these for the non-Abelian Heisenberg group, and contrast the results. Finally we consider the spectral gap for Abelian groups.

1.3.3.1 Typical Distance and Diameter for Abelian Groups

Informally, we show that the mass (in terms of number of vertices) concentrates at a thin 'slice', or 'shell', consisting of vertices at a distance $M \pm o(M)$ from the origin, with M explicit.

For an Abelian group G, write $d(G)$ for the minimal size of a generating subset of G and

$$m_*(G) := \max\big\{\min_{j\in[d]} m_j \;\big|\; \oplus_1^d \mathbb{Z}_{m_j} \text{ is a decomposition of } G\big\}.$$

Our first statement is on typical distance for Abelian groups. More refined statements are given in Theorems 4.1.2, 4.2.2 and 4.3.2; see also Hypotheses F to H.

Theorem G (Typical Distance). *Let G be an Abelian group.*

Consider $1 \ll k \ll \log|G|$; suppose that $k - d(G) \asymp k$ and $d(G) \ll \log|G|/\log\log k$. Write $\mathfrak{D}^+ := |G|^{1/k}/(2e)$ and $\mathfrak{D}^- := |G|^{1/k}/e$. For all $\beta \in (0,1)$, we have $\mathcal{D}^{\pm}_{G_k}(\beta)/\mathfrak{D}^{\pm} \to^{\mathbb{P}} 1$.

Consider $k \asymp \lambda \log|G|$ with $\lambda \in (0,\infty)$; suppose that $d(G) \leq \frac{1}{2}\log|G|/\log\log|G|$ and $m_*(G) \gg 1$. There is a constant $\alpha^{\pm}_\lambda \in (0,\infty)$ so that, for all $\beta \in (0,1)$, we have $\mathcal{D}^{\pm}_{G_k}(\beta)/(\alpha^{\pm}_\lambda k) \to^{\mathbb{P}} 1$.

Consider $k \gg \log|G|$ with $\log k \ll \log|G|$; write $\rho := \log k/\log\log|G|$ so that $k = (\log|G|)^\rho$. For all $\beta \in (0,1)$, we have $\mathcal{D}^{\pm}_{G_k}(\beta)/\big(\frac{\rho}{\rho-1}\log_k|G|\big) \to^{\mathbb{P}} 1$.

In all three cases, the implicit lower bound holds deterministically and for all Abelian groups.

Remark G. We establish the concentration of typical distance via three distinct approaches, in §4.1, §4.2 and §4.3. Conceptually, all involve sizes of lattice balls and drawing elements uniformly from balls. A precise statement for each approach is given, as is an outline of the proof. In summary, Theorem G is a direct consequence of Theorems 4.1.2, 4.2.2 and 4.3.2; see also Hypotheses F to H.

It is interesting how we prove this theorem. It is common in mixing time proofs to use geometric properties of the graph, such as expansion or distance properties. We do the opposite: we use mixing techniques to prove this geometric result. This is in the same spirit as [52]; see §1.5.4. △

We consider the diameter when $k \gtrsim \log|G|$. Our first result is a concentration statement akin to Theorem H. A more refined statement is given in Theorem 4.4.1.

Theorem H (Diameter). *Let G be an Abelian group.*

Consider $k \approx \lambda \log|G|$ with $\lambda \in (0,\infty)$; suppose that $d(G) \le \frac{1}{2}\log|G|/\log\log|G|$ and $m_*(G) \gg 1$. Let $\alpha_\lambda^\pm \in (0,\infty)$ be the constant from Theorem G. We have $\operatorname{diam} G_k^\pm/(\alpha_\lambda^\pm k) \to^{\mathbb{P}} 1$.

Consider $k \gg \log|G|$ with $\log k \ll \log|G|$; write $\rho := \log k/\log\log|G|$ so that $k = (\log|G|)^\rho$. We have $\operatorname{diam} G_k^\pm/(\frac{\rho}{\rho-1}\log_k|G|) \to^{\mathbb{P}} 1$. The implicit upper bound here holds for arbitrary groups.

In both cases, the implicit lower bound holds deterministically and for all Abelian groups.

Remark H. Note that for any graph H one has $\mathcal{D}_H(\frac{1}{2}) \le \operatorname{diam} H \le 2\,\mathcal{D}_H(\frac{1}{2}) + 1$. (Note that $(x_1,...,x_\ell)$ is a path in $G(z)$ if and only if $(x_\ell,...,x_1)$ is a path in $G(z^{-1})$.) So the typical distance and diameter are always equivalent up to constants. Theorem H gives conditions under which they are asymptotically equivalent whp for random Cayley graphs.

Combining Theorem G with Theorem A shows that $t_{\text{mix}}(G_k) \asymp (\operatorname{diam} G_k)^2/k$ whp when $k - d(G) \asymp k$. One can also consider non-Abelian groups; see Theorem J. △

Our next diameter result shows, in a well-defined sense, that, amongst all groups, when $k - \log_2|G| \asymp k$ with $\log k \ll \log|G|$, the group $\mathbb{Z}_2^d$ gives rise to the largest typical diameter.

We define the candidate radius which we show is an upper bound for $\operatorname{diam} G_k$ whp.

Definition I. *Write $\mathfrak{R}(k,n)$ for the minimal $R \in \mathbb{N}$ with $\binom{k}{R} \ge n$.*

We now state the second diameter result. A more refined statement is given in Theorem 4.4.3.

Theorem I. *Let G be an arbitrary group. Suppose that $k - \log_2|G| \asymp k$ and $1 \ll \log k \ll \log|G|$. Then $\operatorname{diam} G_k \le \mathfrak{R}(k,|G|)$ up to sot whp; further, if $H := \mathbb{Z}_2^d$, then $\operatorname{diam} H_k \approx \mathfrak{R}(k,|H|)$ whp.*

This gives a quantitative sense in which $\mathbb{Z}_2^d$ is the group giving rise to the largest diameter.

Corollary I. *For all diverging d and n with $n \le 2^d$ and all groups G of size n, if $k - \log_2 n \asymp k$ and $\log k \ll \log n$, then $\operatorname{diam} G_k \le \operatorname{diam} H_k$ where $H := \mathbb{Z}_2^d$ up to sot whp over Z.*

1.3.3.2 Typical Distance and Diameter for Heisenberg Groups

Our next result concerns typical distance in the random Cayley graph.

Definition J. *For a group G, $k \in \mathbb{N}$ and $\beta \in (0,1)$, define the β-typical distance $\mathcal{D}_{G_k}(\beta)$ via*

$$\mathcal{B}_{G_k}^\pm(R) := \{x \in G \mid \operatorname{dist}_{G_k^\pm}(\mathsf{id},x) \le R\} \quad \text{and} \quad \mathcal{D}_{G_k}^\pm(\beta) := \min\{R \ge 0 \mid |\mathcal{B}_{G_k}^\pm(R)| \ge \beta|G|\},$$

with the $\pm$-superscript indicating definitions for both the directed and undirected cases.

Informally, we show that the mass (in terms of number of vertices) concentrates at a thin 'slice', or 'shell', consisting of vertices at a distance $M \pm o(M)$ from the origin, with M explicit.

Investigating this typical distance when k diverges with $|G|$ was suggested to us by Benjamini [10]. Previous work concentrated on fixed k, ie independent of $|G|$; see §1.5.4.

A more refined statement than the one below is given in Theorem 3.2.1.

Theorem J (Typical Distance). *Let p be prime and $d \ge 3$. Let $H := H_{p,d}$ and $A := H_{p,d}^{\text{ab}}$. Write*

$$M_k^+ := k|A|/e, \quad M_k^- := k|A|^{1/k}/(2e) \quad \text{and} \quad M_k^* := \tfrac{\rho}{\rho-1}\log_k|A|$$

where $\rho := \log k/\log\log|A|$, ie $k = (\log|A|)^\rho$. Assume that the following conditions hold:

- *if $1 \ll k \ll \log|A|$, then either d is fixed or $d \ll \max\{\log k, k^{1/2}/|A|^{1/(4k)}\}$ and $k \le \frac{3}{2}\log_d|A|$;*
- *if $k \gtrsim \log|A|$, then $\log k \ll \log|H|$ and $\log d \ll \log\log|H|$.*

For all $\lambda \in (0,\infty)$, there exists a constant $\alpha_\lambda^\pm \in (0,\infty)$ so that, for all constants $\beta \in (0,1)$, the following convergences in probability hold:

$$\begin{aligned} \mathcal{D}^\pm_{H_k}(\beta)/M_k^\pm &\to^{\mathbb{P}} 1 \quad \text{if} \quad 1 \ll k \ll \log|A|; \\ \mathcal{D}^\pm_{H_k}(\beta)/(\alpha_\lambda^\pm k) &\to^{\mathbb{P}} 1 \quad \text{if} \quad k \asymp \lambda \log|A|; \\ \mathcal{D}^\pm_{H_k}(\beta)/\max\{M_k^*,\, \log_k|H|\} &\to^{\mathbb{P}} 1 \quad \text{if} \quad k \gg \log|A|. \end{aligned}$$

Alternatively, the typical distance concentration value can be given by the maximum of $\log_k|H|$ and the minimal radius of a k-dimensional lattice ball of volume at least $|A|$. Note that

$$\max\{M_k^*,\, \log_k|H|\} = \max\{\tfrac{\rho}{\rho-1},\, \tfrac{1}{2}d\}\log|A|.$$

Remark J.1. By a classical result, to generate a nilpotent group it is enough that the maps of the generators under $g \mapsto gG^{\mathrm{com}} : G \mapsto G^{\mathrm{ab}}$ generate the Abelianisation G^{ab}; this follows from the fact that for nilpotent groups $G^{\mathrm{com}} \leq \Phi(G)$, the Frattini subgroup of non-generators of G.

We prove a quantitative version of this result, where the typical distance in G is very close to the typical distance in G^{ab} for the Cayley graph with generating multiset $[Z_1G^{\mathrm{com}}, ..., Z_kG^{\mathrm{com}}]$.

See El-Baz and Pagano [6] for a recent different result in the same spirit. △

Remark J.2. In Remark C.2, we interpret the cutoff time for the RW in the following way: if the walk has run for long enough so that the projection to the Abelianisation is mixed *and* almost every vertex can be reached with reasonable probability, then the walk is mixed on the full group. Theorem J says that the typical distance and mixing time agree when $k \gg \log|H^{\mathrm{ab}}_{p,d}|$; this gives a sense of rigour to the above interpretation. △

In the regime $k \gtrsim \log|H^{\mathrm{ab}}_{p,d}|$, when $d \asymp 1$, we can extend the typical distance argument to determine the *diameter*, ie the maximal distance between pairs of vertices. In this regime, the two are the same, up to sot, whp. For a graph H, denote by $\operatorname{diam} H$ its diameter.

Theorem K (Diameter)**.** *Let p be prime and $d \geq 3$ fixed. Suppose that $1 \ll \log k \ll \log|H_{p,d}|$. For all $\lambda \in (0,\infty)$, with $\alpha_\lambda^\pm$ the constant from Theorem J, the following convergences hold:*

$$\begin{aligned} (\operatorname{diam} H_k)/(\alpha_\lambda^\pm k) &\to^{\mathbb{P}} 1 \quad \text{if} \quad k \asymp \lambda \log|H^{\mathrm{ab}}_{p,d}|; \\ (\operatorname{diam} H_k)/\max\{M_k^*,\, \log_k|H^{\mathrm{ab}}_{p,d}|\} &\to^{\mathbb{P}} 1 \quad \text{if} \quad k \gg \log|H^{\mathrm{ab}}_{p,d}|. \end{aligned}$$

Remark K. Theorems C, J and K combined give $t_{\mathrm{mix}}(H_k) \asymp (\operatorname{diam} H_k)^2/k$ whp. △

Interesting is the way we prove Theorem J, and by extension Theorem K. It is quite common in mixing proofs to use geometric properties of the graph, such as expansion or distance properties. We, in essence, do the opposite: we adapt the mixing proof to this geometric set-up. (We give a proof-outline in §3.2.2.) This is in the same spirit as [52]; see §1.5.4.

Remark. When $k \gtrsim \log|H^{\mathrm{ab}}_{p,d}|$, for typical distance, and by extension diameter, we can remove the primarily assumption on p. This involves using ideas in the proof of Theorem D.2, ie the extension to non-prime p for cutoff with $k \gtrsim \log|H^{\mathrm{ab}}_{p,d}|$. We do not go into detail here. △

1.3.3.3 Spectral Gap for Abelian Groups

Our final result concerns the spectral gap, and relaxation time, of the random Cayley graph.

Definition L. *Consider a reversible Markov chain with (real) eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq -1$ of its transition matrix. The usual, respectively absolute, spectral gap is defined as*

$$\gamma := \min_{i\neq 1}\{1-\lambda_i\} = 1-\lambda_2, \quad \textit{respectively} \quad \gamma_* := \min_{i\neq 1}\{1-|\lambda_i|\} = 1-\max\{|\lambda_2|,|\lambda_n|\};$$

the usual, respectively absolute, relaxation time is defined as $t_{\mathrm{rel}} := 1/\gamma$, respectively , $t^*_{\mathrm{rel}} := 1/\gamma_*$.

By the spectral gap or relaxation time of a graph*, we mean that of the SRW on the graph.*

It is classical that under reversibility in continuous-time the spectral gap asymptotically determines the exponential rate of convergence to equilibrium, whereas in discrete-time it is determined by the absolute spectral gap; see Remark 4.5.7.

Our result finds the correct order of the relaxation time. We do not require $k \to \infty$ as $|G| \to \infty$. The same statement and proof hold for both the usual and absolute relaxation times. Recall that we write $d(G)$ for the minimal number of generators required to generate the group.

A more refined statement than the one given below is given in Theorem 4.5.1 in §4.5.

Theorem L. *There exists a positive constant c so that, for all Abelian groups G, all k and all multisets of generators z of size k, we have*

$$t^*_{\mathrm{rel}}\big(G^-(z)\big) \geq t_{\mathrm{rel}}\big(G^-(z)\big) \geq c|G|^{2/k}.$$

For all $\delta > 0$, there exists a constant $C_\delta > 0$ so that, for all Abelian groups G, if $k \geq (2+\delta)d(G)$, then

$$\mathbb{P}\big(t^*_{\mathrm{rel}}(G^-_k) \leq C_\delta|G|^{2/k}\big) \geq 1 - C_\delta 2^{-k/C_\delta}.$$

Further, for all $\varepsilon \in (0,1)$, there exists a density-$(1-\varepsilon)$ subset $\mathbb{A} \subseteq \mathbb{N}$ so that if $|G| \in \mathbb{A}$ then the condition $k \geq (2+\delta)d(G)$ can be relaxed to $k \geq (1+\delta)d(G)$; the constants now also depend on ε.

The method of proof for this result is rather different to our previous results, and also somewhat different to those used by others to study the spectral gap of random Cayley graphs; see §1.5.5.

1.3.4 Additional Cutoff and Typical Distance Results

To close, we state some additional results on cutoff and typical distance. These tend to be some slightly more refined results than those given above, but with additional conditions on the group. Eg, in Theorem M we determine the limit profile, not just the existence of cutoff, in the regime $k \asymp \log|G|$, which we could not do in Theorem A. However, while previously we considered all groups, now there are conditions on the group.

1.3.4.1 Cutoff: Limit Profile for Random Walks on Abelian Groups

For $t \geq 0$, write μ_t for the law of $W(t)$. Define $Q(t) := -\log \mu_t(W(t))$.

Definition M. *For all $k, n \in \mathbb{N}$ and all $\alpha \in \mathbb{R}$, define $t_\alpha := t_\alpha(k,n)$ so that*

$$\mathbb{E}\big(Q(t_\alpha)\big) = \big(\log n + \alpha\sqrt{vk}\big) \quad \textit{where} \quad v := \mathbb{V}\mathrm{ar}\big(Q(t_0)\big)/k.$$

We call t_0 the entropic time and the $\{t_\alpha\}_{\alpha\in\mathbb{R}}$ cutoff times.

An asymptotic evaluation of these times is given in Proposition 5.1.2.

For an Abelian group G, write $d(G)$ for the minimal size of a generating subset of G and

$$m_*(G) := \max\big\{\min_{j\in[d]} m_j \;\big|\; \oplus_1^d \mathbb{Z}_{m_j} \text{ is a decomposition of } G\big\}.$$

A more refined statement is given in Theorem 5.1.4; see also Hypothesis I.

Theorem M. *Let G be an Abelian group. Let $\lambda \in (0,\infty)$ and suppose that $k \asymp \lambda \log|G|$. Suppose that $d(G) \leq \frac{1}{35}\log|G|/\log\log|G|$ and $m_*(G) > (\log k)^2$.*

Then, whp, the RW on G_k exhibits cutoff at $t_0 := t_0(k,|G|)$; moreover, it has Gaussian profile given by $\{t_\alpha := t_\alpha(k,|G|)\}_{\alpha\in\mathbb{R}}$, namely, writing $\Psi : \mathbb{R} \to [0,1]$ for the standard Gaussian tail,

$$d_{G_k}(t_\alpha) \to^{\mathbb{P}} \Psi(\alpha) \quad \textit{(in probability)} \quad \textit{for all } \alpha \in \mathbb{R}.$$

(The randomness is over the uniform choice of generators $Z = [Z_1, \ldots, Z_k]$.) Further,

$$t_0 \asymp k \quad \textit{and} \quad |t_\alpha - t_0| \asymp \alpha t_0/\sqrt{k} \quad \textit{for all} \quad \alpha \in \mathbb{R}.$$

Remark M. We can write the cutoff statement in terms of the mixing time, rather than the TV distance: writing $t_{\text{mix}}(\varepsilon)$ for the ε-mixing time, for all $\varepsilon \in (0,1)$, we have

$$\big(t_{\text{mix}}(\varepsilon) - t_0\big)/w \to^{\mathbb{P}} \Psi^{-1}(\varepsilon),$$

where t_0 is the mixing time and w is the cutoff window defined, via $\{t_\alpha - t_0\}_{\alpha\in\mathbb{R}}$. For a more explicit formula, using asymptotic evaluation of the cutoff times, see Remark 5.1.5. △

1.3.4.2 Cutoff: A Detailed Investigation of $\mathbb{Z}_p^d$

For our next theorem, we specialise to the case $G := \mathbb{Z}_p^d$ with p prime. This specialisation allows us to derive some very refined results. In particular, before we could not allow d to be close to k; here we consider any $k \geq d$. Now every element of G has order p; as such, need only consider the auxiliary W mod p. We redefine the entropic times to take this into account.

Definition N. *Define $t_0 := t_0(k,p,d)$ to be the time at which the entropy of the RW on $\mathbb{Z}_p^k$ is $\log(p^d)$.*
An asymptotic evaluation of this time is given in Proposition 5.2.2.

A more refined statement is given in Theorem 5.2.4. In particular, by defining t_α appropriately, we can also also consider the cutoff window; see Definition 5.2.1 and Theorem 5.2.2.

Theorem M. *Let $G := \mathbb{Z}_p^d$ with p prime. Assume that $1 \ll k \lesssim d \log p$.*
Suppose that $(k-d)p \gg 1$. Then, whp, the RW on G_k exhibits cutoff at t_0.
Also, if $0 \leq k - d \lesssim 1$, then, conditional that the uniformly chosen multisubset $[Z_1, ..., Z_k]$ generates the group, there is cutoff whp at time $\frac{1}{2} d \log d/(1 - \cos(2\pi/p))$.

1.3.4.3 Cutoff: No Cutoff when k Is Constant

Lastly for cutoff, it is natural to ask what happens when k is constant. This regime has already been analysed by Diaconis and Saloff-Coste [27]. We give an exposition of their results, using the language which we have developed. We emphasise that this is a result of Diaconis and Saloff-Coste.
A more refined statement is given in Corollary 5.4.5.

Theorem N (cf [27, Corollary 5.3])**.** *Let G be a finite, nilpotent group of bounded step. Suppose that $k \asymp 1$. Then the RW on $G^-(Z)$ does not exhibit cutoff for any choice of Z with $|Z| = k$.*

1.3.4.4 Cutoff: Extending Arguments from Heisenberg to Other Nilpotent Groups

In Theorem C, we studied cutoff for random walks on Heisenberg groups. In the introduction there, specifically in Remark C.3, we claimed that some of our analysis extends from Heisenberg groups to more general nilpotent groups. We discuss this claim further in §5.3.

1.3.4.5 Typical Distance: Generalised Graph Distance for Abelian Groups

Locally, when $\log k \ll \log |G|$, typical degree-k Cayley graphs of an Abelian group look like $\mathbb{Z}^k$. In a lattice, graph distance corresponds to L_1 distance; this can be extended to an L_q distance, for $q \in [1,\infty]$. Analogously, we can extend the usual L_1 graph distances to an L_q-type, for $q \in [1,\infty]$.
Consider a collection $z = [z_1, ..., z_k]$ of generators and distances in the Cayley graph $G(z)$. For a path ρ in $G(z)$, for each $i \in [k]$, write $\rho_{i,+}$ for the number of times z_i is used, $\rho_{i,-}$ for the number of times z_i^{-1} is used (if in the undirected case otherwise $\rho_{i,-} := 0$) and $\rho_i := \rho_{i,+} - \rho_{i,-}$. The path connects the identity with $\rho \cdot z$. Then the L_1 length of ρ is $\|\rho\|_1 := \sum_1^k (\rho_{i,+} + \rho_{i,-})$.
For any $q \in [1,\infty)$, define the L_q *graph distance* of ρ by $\|\rho\|_q^q := \sum_1^k (\rho_{i,+}^q + \rho_{i,-}^q)$. For the L_∞-*graph distance*, define $\|\rho\|_\infty := \max_i (\rho_{i,+} + \rho_{i,-})$. (The usual graph distance is given by $q = 1$.)
For Abelian groups, clearly for any $q \in [1,\infty)$ an L_q *geodesic*, ie a path of minimal length, will only use either z_i or z_i^{-1}, not both (since the terms in the product can be reordered), ie $\rho_{i,+}\rho_{i,-} = 0$ for all i. Thus $\|\rho\|_q^q = \sum_1^k |\rho_i|^q$. Similarly, any L_∞-geodesic ρ can be adjusted into a new path ρ' with $\|\rho\|_\infty = \|\rho'\|_\infty$ and $\rho'_{i,+}\rho'_{i,-} = 0$ for all i.

We define the L_q *typical distance* $\mathcal{D}_{G(z),q}(\cdot)$ analogously to $\mathcal{D}_{G(z)}(\cdot)$, ie the $q=1$ case.

For an Abelian group G, recall that $d(G)$ is the minimal size of a generating subset of G and

$$m_*(G) = \max\big\{\min_{j\in[d]} m_j \;\big|\; \oplus_1^d \mathbb{Z}_{m_j} \text{ is a decomposition of } G\big\}.$$

Finally we set up a little more notation. Make the following definitions for $q \in [1,\infty]$:

$$C_q^- := 2\,\Gamma(1/q+1)(qe)^{1/q}, \quad C_q^+ := \tfrac{1}{2}C_q^- \quad \text{and} \quad \mathfrak{D}_q^{\pm}(k,n) := k^{1/q}n^{1/k}/C_q^{\pm},$$

where the case $q=\infty$ is to be interpreted as the limit $q\to\infty$; eg, $C_\infty^- = 2$ and $\mathfrak{D}_\infty^+(k,n) = n^{1/k}$.

A more refined statement is given in Theorem 5.5.1. Write $k^{1/\infty} := 1$.

Theorem O. *Let G be an Abelian group and $q \in [1,\infty]$. Abbreviate $n := |G|$, $d := d(G)$ and $m_* := m_*(G)$. Suppose that $1 \ll k \ll \log|G|$. Draw $Z_1,..,Z_k \sim^{\text{iid}} \text{Unif}(G)$. Suppose that $m_*(G) \gg k^{1/q}n^{1/k}$ and if $q \in (1,\infty)$ then additionally require $k \le \log n/\log\log n$. Suppose that $\limsup d/k < 1$ for undirected graphs and $\limsup d/k < \frac{1}{2}$ for directed graphs.*

Then, whp, the L_q typical distance on G_k concentrates at $\mathfrak{D}_q^{\pm}$, namely

$$\mathcal{D}^{\pm}_{G(Z),q}(\beta)/\mathfrak{D}_q^{\pm} \to^{\mathbb{P}} 1 \quad \textit{(in probability)} \quad \textit{for all } \beta\in(0,1).$$

(The randomness is over the uniform choice of generators $Z=[Z_1,...,Z_k]$.)

1.4 Entropic Method and Cutoff for 'Generic' Markov Chains

We noted at the start of §1.3 that we use an *entropic method*. We briefly described how it worked there. Here we give a brief history before giving a detailed description of the our application.

1.4.1 A Brief History

We now put our results into a broader context. As mentioned above, a common 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, whp, give rise to a sequence of Markov chains which exhibits cutoff. A few notable examples include random birth and death chains [31, 72], the simple or non-backtracking RW on various models of sparse random graphs, including random regular graphs [53], random graphs with given degrees [8, 9, 11, 12], the giant component of the Erdős–Rényi random graph [11] (where the authors consider mixing from a 'typical' starting point) and a large family of sparse Markov chains [12], as well as RWs on a certain generalisation of Ramanujan graphs [13] and random lifts [13, 20].

A recurring idea in the aforementioned works establishing the cutoff phenomenon for certain families of random instances 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 sot) 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 [11, 12, 13, 20]. For 'most' regimes of k, this is the case for us too; further, for the non-Abelian groups, we use a similar idea. As our auxiliary random process, we use a SRW, respectively DRW, in the undirected, respectively directed, case.

With the exception of the very recent [43], 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); for non-Abelian groups, the local behaviour of the graph is more complex. Admittedly, this has less of an impact on the walk since each vertex is of diverging degree.

1.4.2 An Application to Random Cayley Graphs

We now describe in a little more detail the entropic method applied to the set-up of (random) Cayley graphs. We do not give an abstract definition of entropic times here, but rather in each chapter we define them in the way appropriate for that application.

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 $i = 1, ..., k$, 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, increment $W_i \to W_i - 1$ (rather than $W_i \to W_i + 1$). In this case, $W(\cdot)$ is a SRW on $\mathbb{Z}^k$.

If the underlying group is Abelian, then the order in which the generators are applied is irrelevant and generator-inverse pairs cancel; hence we can write $S(t) = \sum_{i=1}^k W_i(t)Z_i = W(t) \cdot Z$. For non-Abelian groups, this simple projection does not hold; just looking at $W(t)$ loses information.

Recall that the invariant distribution is uniform, regardless of the group. 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: using the equivalence $-\log \mu \geq \log |G|$ if and only if $\mu \leq 1/|G|$, 'typically' $W(t)$ takes values to which it assigns probability smaller than $1/|G|$; informally, this means that $W(t)$ is 'well spread out'. If we could immediately deduce that $S(t)$ typically takes values to which it assigns probability approximately $1/|G|$, we would be basically done. However, one could 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, ie well mixed. In contrast, if the entropy is much smaller than $\log |G|$, then $W(t)$ is not well spread out: it is highly likely to live on a set of size $o(1/|G|)$. The same must then be true for $S(t)$; hence it is not mixed.

For a non-Abelian group, as noted above, just looking at $W(t)$ loses information. We decompose $H_{p,d}$ into its Abelianisation $H_{p,d}^{\text{ab}}$ and commutator $H_{p,d}^{\text{com}}$. The above heuristics for Abelian groups suggest that the walk projected to the Abelianisation $H_{p,d}^{\text{ab}}$ should be mixed at the time at which W has entropy $\log |H_{p,d}^{\text{ab}}|$. We then need to check that the walk on the commutator $H_{p,d}^{\text{com}}$ is mixed at this time. For Heisenberg groups, $H_{p,d}^{\text{ab}}$ corresponds to the super-diagonal. Diaconis and Hough [26] showed that coordinates mix faster the farther they are from the diagonal. It is thus natural then to expect the commutator to mix faster than the Abelianisation, at least for d not too large. We need to make sure that all elements of the group can be reached with reasonable probability, and so need to run for at least $\log_k |H_{p,d}|$. This suggests $\max\{t_0(k, |H_{p,d}^{\text{ab}}|), \log_k |H_{p,d}|\}$ as the mixing time.

To study typical distance, we define a related auxiliary variable, A, corresponding to the number of times each generator is used: A is uniformly distributed on a k-dimensional lattice ball of a certain radius. We apply the chosen generators in a uniformly random order. We do not apply an entropic method here, per se, but the underlying principles of the proof are extremely similar.

1.5 Historic Overview

In this section, we give a fairly comprehensive account of previous work on mixing and cutoff for random walk on random Cayley graphs, and compare our results with existing ones. The occurrence of cutoff in particular has received a great deal of attention over the years.

1.5.1 Motivation: Random Cayley Graphs and Cutoff for Random Walks

In their seminal paper, Aldous and Diaconis [1] considered random walks on *random* Cayley graphs. Diaconis [25] gave the following (paraphrased) motivation.

> *Erdős, when considering classes of mathematical objects, often combinatorial or graph theoretic, would often ask, "What does a typical object in this class 'look like'?" If an object is chosen uniformly at random, are there natural properties which hold whp?*
>
> *It is then natural to ask, "What does a typical random walk on a group 'look like'?"*

This lead him, with Aldous, to consider the set of all Cayley graphs of a given group G with a given number k of generators. Drawing such a Cayley graph uniformly at random corresponds precisely to our G_k, ie choosing generators $Z_1, ..., Z_k \sim^{\text{iid}} \text{Unif}(G)$.

In their pioneering work on the cutoff phenomenon [1], Aldous and Diaconis had the remarkable insight to conjecture that when $k \gg \log|G|$ the RW on G_k exhibits cutoff whp, regardless of which underlying group is used. They also suggested a candidate mixing time in terms of k and $|G|$.

It was shown in the '90s that this conjecture is true for Abelian groups; further, the upper bound on mixing was valid for all groups. Moreover, extending the conjecture to $1 \ll k \lesssim \log|G|$, in Theorem A here we establish cutoff for all Abelian groups. In Theorem C here, using Heisenberg matrix groups, we give the first example of cutoff for a non-Abelian group (when k does not grow super-polynomially with $|G|$); we consider any $1 \ll \log k \ll \log|G|$. Contrary to Abelian groups, the mixing time cannot always be written as a function only of k and $|G|$, even for some $k \gg \log|G|$.

1.5.2 Universal Cutoff: The Aldous–Diaconis Conjecture

Aldous and Diaconis [1] stated their conjecture for $k \gg \log|G|$. An upper bound, valid for arbitrary groups, was established by Dou and Hildebrand [34, Theorem 1] and later Roichman [69, Theorems 1 and 2], who simplified their argument. A matching lower bound, valid only for Abelian groups, was given by Hildebrand [44, Theorem 3]. Combined, this established the Aldous–Diaconis conjecture for Abelian groups. Moreover, the cutoff time was known explicitly:

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

(To have $k \gg \log|G|$, one needs $\rho - 1 \gg 1/\log\log|G|$.) See also Dou [33] and Hildebrand [45].

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

In Theorem C, using the Heisenberg group $H_{p,d}$, we disprove the conjecture: taking $k \coloneqq \lfloor\log|H_{p,d}|\rfloor^2$ and $d \coloneqq 3$, there is cutoff at $\frac{2}{3}T(k,|H_{p,d}|)$. In fact, $T(k,n)$ does not even capture the correct order: letting $d \to \infty$ sufficiently slowly, k can be chosen so that $k \gg \log|H_{p,d}|$ and there is cutoff at a time smaller order than $T(k,|H_{p,d}|)$.

To extend consideration to $1 \ll k \lesssim \log|G|$, one naturally needs some conditions. For example, if $k < d$ and $G \coloneqq \mathbb{Z}_2^d$, then the group is not generated, and so no mixing can occur. There has been some investigation into the regime $1 \ll k \lesssim \log|G|$, but with much less success. Hildebrand [44, Theorem 4] showed that the mixing time must be super-polylogarithmic, unlike for $k \gg \log|G|$. Wilson [77, 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 (in 2017) when Hough [46, 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.

1.5.3 Random Walks on the Heisenberg Group

Random walks on the Heisenberg group have been the focus of a great deal of attention; focus has primarily been on 3×3 matrices. (See in particular [16, §2.1] and [66, §1.1], upon which we have based the description below.) The probabilistic study of random walks on $H_{p,d}$ was initiated by Zack [79]; she interpreted the walk in terms of random number generation; focus has been on $d = 3$. Using a specific generating set of size 4, the correct order of mixing was established by Diaconis and Saloff-Coste [27, 29, 30] using geometric theoretical tools. Further proofs were given by Diaconis [24], Stong [73, 74, 75] and Bump et al [16].

Moving even further from the realm of Abelian groups, consider p fixed and general $d \geq 3$. One can consider a simple walk on $H_{p,d}$: a row is chosen uniformly and added to or subtracted from the row above. Ellenberg [36] studied the diameter of the associated Cayley graph, with d growing, subsequently improving this in Ellenberg and Tymoczko [37]. Stong [75] gave mixing bounds via analysis of eigenvalues. Coppersmith and Pak [21, 63] look directly at mixing. This has then been further studied, improved upon and generalised by Peres and Sly [66], Nestoridi [59] and Nestoridi and Sly [60]; Nestoridi and Sly [60] are the first to optimise bounds for p and d simultaneously.

In a recent impressive work, Diaconis and Hough [26] introduced a new method for proving a CLT for random walks on nilpotent groups. They illustrate the method on $H_{p,d}$, obtaining some

extremely precise results on the rate at which individual coordinates mix as a function of their distance from the diagonal. They show that the greater the distance from the diagonal is, the faster the mixing time of the coordinate is. In the same spirit, we show that, in many cases, the bottleneck for mixing is the super-diagonal coordinates, while in the rest of the cases, the cutoff time is given by the diameter-based lower bound $\log_k |H_{p,d}|$.

Related work includes analysis of the spectrum of a random walk on the Heisenberg group by Béguin, Valette and Zuk [7]. There has also been work on a random walk on a 3×3 Heisenberg group with entries in $\mathbb{R}$. See, for example, Breuillard [14, 15].

1.5.3.1 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 [62]. There he studies universal mixing bounds (ie 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 [77] (and by us above).

A related universal bound in which $\mathbb{Z}_2^d$ is the worst case is given by Pak [64]. Let $\varphi_k(G) := \mathbb{P}(G_k \text{ is connected})$, ie the probability that the group G is generated by k uniformly chosen generators. Then Pak [64, Lecture 1, Theorem 6] proves that if $|G| \leq 2^d$ then $\varphi_k(G) \geq \varphi(\mathbb{Z}_2^d)$ for all k.

1.5.4 Typical Distance and Diameter

As well as determining cutoff for these random Cayley graph, we study a geometric property of a diameter flavour; recall the concept of *typical distance* from §1.2.3. Previous work (detailed below) had concentrated on the case where the number of generators k is a *fixed* number, ie one which does not increase as the size n of the group increases. In contrast, our results are in the situation where $k \to \infty$ as $n \to \infty$; this line of enquiry was suggested to use by Benjamini [10].

Amir and Gurel-Gurevich [4] studied the diameter of the random Cayley graph of cyclic groups of prime order. They prove (for fixed k) that the diameter is order $|G|^{1/k}$; see [4, Theorems 1 and 2]. They conjecture that the diameter divided by $|G|^{1/k}$ converges in distribution to some non-trivial random variable as $|G| \to \infty$; see [4, Conjecture 3].

Marklof and Strömbergsson [55] consider, as a consequence of a quite general framework, the diameter of the random Cayley graph of $\mathbb{Z}_n$ with respect to a fixed number k of random generators, for a random n, without any primality assumption. They derive distributional limits for the diameter, the average distance (defined with respect to various L_p metrics) and the girth. They determine limit distributions for each of these, and in some cases derive explicit formulas.

Shapira and Zuck [71] build on the framework of Marklof and Strömbergsson [55], again only for fixed k; they are able to consider non-random n, as well as Abelian groups of arbitrary (fixed) rank, instead of only cyclic groups. In particular, they verify the conjecture of Amir and Gurel-Gurevich [4, Conjecture 3]; they additionally work with average distance and girth.

Lubetzky and Peres [52] derive an analogous typical distance result for n-vertex, d-regular Ramanujan graphs: whp all by $o(n)$ of the vertices lie at a distance $\log_{d-1} n \pm \mathcal{O}(\log \log n)$; they establish this by proving cutoff for the non-backtracking random walk at time $\log_{d-1} n$.

Related work on the diameter of random Cayley graphs, including concentration of certain measures, can be found in [50, 70].

The Aldous–Diaconis conjecture for mixing can be transferred naturally to typical distance: the mass should concentrate at a distance M, where M can be written as a function only of k and n; ie there is concentration of mass at a distance independent of the algebraic structure of the group.

1.5.5 Spectral Gap

Hough [46, Theorem 1.1] showed that, for any prime p, the relaxation time of the RW on any Cayley graph of $\mathbb{Z}_p$ with respect to an arbitrary set of k generators is order at least $|\mathbb{Z}_p|^{2/k} = p^{2/k}$, provided that $k \leq \log p / \log \log p$. Using a different approach, we extend Hough's result, removing the restrictions on p and k and considering general Abelian groups; see Theorem L.

This extends, in the Abelian set-up, a celebrated result of Alon and Roichman [3, Corollary 1], which asserts that, for any finite group G, the random Cayley graph with at least $C_\varepsilon \log |G|$ random

generators is whp an ε-expander, provided C_ε is a 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 the spectral gap of the RW on the graph being bounded below by ε.) There has been a considerable line of work building upon this general result of Alon and Roichman. (Pak [61, 62] proves a similar result.) Their proof was simplified and extended, independently, by Loh and Schulman [51] and Landau and Russell [47]; 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 [17]. Both [17, 47] use some Chernoff-type bounds on operator valued random variables.

Christofides and Markström [18] improve these further by using matrix martingales and proving a Hoefdding-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, whp, 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.

Alon and Roichman [3, Theorem 2] also specifically consider Abelian groups. There they do a calculation directly in terms of the eigenvalues, rather than using a probabilistic tail bound.

There are some fairly standard ways in which one can get bounds on the (usual) spectral gap. The first is to look at the mixing time. It is standard that, for $c > 0$ and $\varepsilon \in (0, 1/n^c]$, we have

$$t_{\text{mix}}(\varepsilon) \asymp t_{\text{rel}} \log(1/\varepsilon),$$

where n is the size of the state space of the Markov chain and c is a constant; see, eg, [49, Theorems 12.5 and 20.6]. Thus, if one can bound the mixing time at level $1/n^c$ then one can bound the relaxation time. This method is used by Alon and Roichman [3], as well as by Pak [61].

Another method is to obtain a tail estimate on the value of a random eigenvalue; one can then use the union bound to say that all (non-unitary) eigenvalues are at most some fixed value, which in turn lower bounds the spectral gap (ie upper bounds the relaxation time).

All these references consider the regime $k \asymp \log |G|$; our results also apply when $1 \ll k \ll \log |G|$. From a technical perspective, in order to obtain failure probability via a large deviation bound for a random eigenvector of $\mathcal{O}(1/|G|)$, one needs $k \gtrsim \log |G|$. The purpose of this is to carry out a union bound over the $|G|$ eigenvalues; see, eg, [18]. Likewise, arguments that bound the $1/|G|^c$ mixing time, for some constant c, in terms of some generator getting picked once (cf [69]) cannot work unless $k \gtrsim \log |G|$. As such, to consider $1 \ll k \ll \log |G|$, a different approach is needed. We still use a union bound, but instead of asking for an error probability $\mathcal{O}(1/|G|)$ for each eigenvalue, we group together eigenvalues according to a certain gcd and bound the error for each group.

1.6 Additional Remarks

1.6.1 Typical Cayley Graphs

Given a group G and an integer k, we are drawing the generators $[Z_1, ..., Z_k]$ independently and uniformly at random. Thus $G(Z)$ is in fact a uniform Cayley graph. So when we say that our results hold "whp over Z", we could equivalently say that the result holds "for almost all degree-k Cayley graphs of G". Not only this, but since our asymptotic evaluation does not depend on the particular choice of Z, this shows that the statistics in question depends very weakly on the particular choice of generators for almost all choices of generators. This is a strong sense of 'universality'.

1.6.2 Simple Cayley Graphs

The Cayley graph is simple if and only if no generator is picked twice, ie $Z_i \neq Z_j$ for all $i \neq j$ and no generator is the identity; in the undirected case, additional no generator may be the inverse of another, ie $Z_i \neq Z_j^{-1}$ for all i and j. Since $k/\sqrt{|G|} \to 0$ as $|G| \to \infty$, the probability of this event tends to 1 as $|G| \to \infty$. Hence our "whp over Z" results all also hold when the generators are chosen uniformly at random from G but conditional on giving rise a simple Cayley graph.

1.6.3 Asymptotic Results and Notation

Our results are asymptotic as the size of the group diverges. As such, we implicitly consider a sequence $(G_N)_{N\in\mathbb{N}}$ of groups; we also assume that k (and hence $Z = [Z_1, ..., Z_k]$) is indexed by N. For simplicity of notation, we tend to drop the sequence notation, eg writing G or k.

For functions f and g, write $f \eqsim g$ if $f(N)/g(N) \to 1$ as $N \to \infty$; also write $f \ll g$, or $g \gg f$, if $f(N)/g(N) \to 0$ as $N \to \infty$. Write $f \lesssim g$, or $g \gtrsim f$, if there exists a constant C so that $f(N) \leq Cg(N)$ for all N; also write $f \asymp g$ if $g \lesssim f \lesssim g$. Also write $f = \mathcal{O}(g)$ if $f \lesssim g$, and $f = o(g)$ if $f \ll g$. Throughout the paper, unless otherwise explicitly mentioned all limits will be as the size of the group diverges; so if a term is $o(1)$, then it tends to 0 as the group gets larger.

Throughout the paper, we frequently consider undirected and directed graphs, or simple and directed RWs, simultaneously. We use a $+$-sub/superscript to indicated directed and $-$ to indicated undirected. We use $\pm$ to indicate that a statement holds for both; when such an identifier is omitted, it means the same. Eg, G_k is a Cayley graph which can be either directed or undirected.

When dealing with a sequences $(k_N)_{N\in\mathbb{N}}$ and $(G_N)_{N\in\mathbb{N}}$, we abbreviate

$$d^{\pm}_{G_k,N}(t) \coloneqq \left\|\mathbb{P}_{G^{\pm}_N([Z_1,\ldots,Z_{k_N}])}\big(S(t)\in\cdot\big) - \pi_{G_N}\right\|_{\mathrm{TV}} \quad \text{where} \quad Z_1, ..., Z_{k_N} \sim^{\mathrm{iid}} \mathrm{Unif}(G_N).$$

We write $\mathcal{D}^{\pm}_{G_k,N}(\beta)$ for the typical distance similarly. For simplicity, we tend to drop the sequence notation, and sometimes the $\pm$-superscript, eg writing $d^{\pm}_{G_k}(t)$, $\mathcal{D}_{G_k}(\beta)$ or G.

1.6.4 Acknowledgements

- We thank Allan Sly for suggesting the underlying entropy-based approach taken in this paper.
- For multiple insightful discussions on mixing for the Heisenberg group, as well as for more general nilpotent groups, we thank Péter Varjú.
- For discussions on typical distance, we thank Itai Benjamini.
- For general discussions, consultation and advice, we thank Evita Nestoridi and Persi Diaconis.

2 Cutoff for Almost All Random Walks on Abelian Groups

Abstract for Chapter 2

We establish cutoff whp for *all* Abelian groups in the regime $1 \ll k \lesssim \log|G|$ with $k - d(G) \gg 1$; the latter condition is almost optimal for generating the group whp. (Cutoff for all Abelian groups in the regime $k \gg \log|G|$ had already been established.)

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 depends, up to subleading order terms, only on k and $|G|$ when $d(G) \ll \log|G|$ and $k - d(G) \asymp k \gg 1$. This is not so when $d(G) \asymp \log|G| \asymp k$ or even for some $k \ll \log|G|$ if $1 \ll k - d(G) \ll k$.

Table of Contents for Chapter 2

2.1 TV 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 2.1.4; see also Hypothesis A and Remark 2.1.5.

The outline of this section is as follows:

- §2.1.1 defines *entropic times* and states a CLT;
- §2.1.2 sketches arguments to evaluate these entropic times;
- §2.1.3 states precisely the main theorem of the section;
- §2.1.4 outlines the argument;
- §2.1.5 is devoted to the lower bound;
- §2.1.6 is devoted to the upper bound.

2.1.1 Entropic Times: Definition and Concentration

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

$$Q_i(t) := -\log \nu_{t/k}\big(W_i(t)\big), \quad \text{and set} \quad Q(t) := -\log \mu_t\big(W(t)\big) = \textstyle\sum_1^k Q_i(t).$$

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

Definition 2.1.1 (Entropic and Times). *For all $k, n \in \mathbb{N}$ and all $\alpha \in \mathbb{R}$, define $t_\alpha := t_\alpha(k,n)$ so that*

$$\mathbb{E}\big(Q_1(t_\alpha)\big) = \big(\log n + \alpha\sqrt{vk}\big)/k \quad \textit{and} \quad s_\alpha := t_\alpha/k, \quad \textit{where} \quad v := \mathbb{V}\mathrm{ar}\big(Q_1(t_0)\big),$$

assuming that $\log n + \alpha\sqrt{vk} \geq 0$. We call t_0 the entropic time and the $\{t_\alpha\}_{\alpha\in\mathbb{R}}$ cutoff times.

Direct calculation with the Poisson distribution and SRW on $\mathbb{Z}$ gives the following relations. These calculations are sketched below in §2.1.2; rigorous arguments are given in §6.1.

Proposition 2.1.2 (Entropic and Cutoff Times; Proposition 6.1.2). *Assume that $1 \ll k \ll \log n$. For all $\alpha \in \mathbb{R}$, we have $t_\alpha \approx t_0$ and furthermore*

$$t_0 \approx k \cdot n^{2/k}/(2\pi e) \quad \textit{and} \quad (t_\alpha - t_0)/t_0 \approx \alpha\sqrt{2/k}.$$

Since $Q = \sum_1^k Q_i$ is a sum of k iid random variables, $Q(t_0)$ concentrates around $\log N$. One can show that if the time is multiplied by a factor $1+\xi$ for any constant $\xi > 0$ then the entropy increases by a significant amount; similarly, if $\xi < 0$ then the entropy decreases by a significant amount. Further, the change is by an additive term of larger order than the standard deviation $\sqrt{\mathbb{V}\mathrm{ar}(Q(t_0))}$. Thus $Q((1+\xi)t_0)$ concentrates around this new value.

The following proposition quantifies this change in entropy and this concentration; see §6.1.

Proposition 2.1.3 (CLT; Proposition 6.1.3). *Assume that $1 \ll k \ll \log n$. For all $\alpha \in \mathbb{R}$, we have*

$$\mathbb{P}\big(Q(t_\alpha) \leq \log n \pm \omega\big) \to \Psi(\alpha) \quad \textit{for} \quad \omega := \mathbb{V}\mathrm{ar}\big(Q(t_0)\big)^{1/4} = (vk)^{1/4}.$$

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

2.1.2 Entropic Times: Sketch Evaluation

In this subsection, we sketch details towards a proof of Proposition 2.1.2. The full, rigorous details can be found in Proposition 6.1.2, where all of the approximations below are justified.

Recall that t_0 is the time t at which the entropy of $W_1(t)$, which is a rate-$1/k$ process, is $\log n/k$. We need to find the variance $\mathbb{V}\mathrm{ar}(Q_1(s_0 k))$, as this is used in the definition of t_α, given in Definition 2.1.1. In the sketch below, we replace $\mathbb{V}\mathrm{ar}(Q_1(t_0))$ by an approximation.

For $s \geq 0$, denote $X_s := W_1(sk)$ for $s \geq 0$ and the entropy of X_s as $H(s)$. The target entropy $\log n/k \gg 1$, and so $s_0 \gg 1$. For $s \gg 1$, we find that 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 of X_s, which we denoted $H(s)$, by the entropy of a $N(0, s)$ random variable, which we denoted $\overline{H}(s)$. Direct calculation with the normal distribution shows that

$$\overline{H}(s) = \tfrac{1}{2}\log(2\pi e s) \quad \text{and hence} \quad \overline{H}'(s) = 1/(2s).$$

Define $\overline{s}_\alpha$ as the entropic times for the approximation:

$$\overline{H}(\overline{s}_\alpha) = \big(\log n + \alpha\sqrt{vk}\big)/k \quad \text{where} \quad v := \mathbb{V}\mathrm{ar}\big(\overline{Q}_1(\overline{s}_0 k)\big),$$

where $\overline{Q}_1(sk)$ is the analogue of $Q_1(sk)$, except with $W_1(sk)$ replaced by $N(0, s)$. Hence $\overline{s}_0 = n^{2/k}/(2\pi e)$. Direct calculation with the normal distribution, one finds

$$\mathbb{V}\mathrm{ar}\big(\overline{Q}_1(sk)\big) = \tfrac{1}{2}.$$

As mentioned above, for this sketch, to ease the calculation of t_α in Definition 2.1.1, we replace $\mathbb{V}\mathrm{ar}(Q_1(t_0))$ by its approximation $\frac{1}{2}$, and assume the above normal distribution approximation.

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

$$s_\alpha - s_0 = \textstyle\int_0^\alpha \frac{ds_a}{da}\, da.$$

Again, we replace s_α with $\overline{s}_\alpha$. By definition, $\overline{s}_\alpha$ satisfies

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

Using the expressions for $d\overline{s}_a/da$ and $\overline{H}'(s)$ above, we find that

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

since $\overline{s}_a$ only varies by sot over $a \in [0, \alpha]$. The argument is analogous for $\alpha < 0$.

We have now shown the desired result for $\overline{s}_\alpha$, ie when approximating $W_1(sk)$ by $N(\mathbb{E}(X_s), s)$. It will turn out that this approximation is sufficiently good for the results to pass over to the original case, ie to apply to s_0 and $t_0 = s_0 k$. This is made rigorous with a local CLT.

2.1.3 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:*

$$\lim_{N\to\infty}|G_N| = \infty, \quad \lim_{N\to\infty}\big(k_N - d(G_N)\big) = \infty \quad \text{and}$$
$$\frac{k_N - d_N(G_N) - 1}{k_N} \geq 5\frac{k_N}{\log|G_N|} + 2\frac{d_N(G_N)\log\log k_N}{\log|G_N|} \text{ for all } N \in \mathbb{N}.$$

In Remark 2.1.5 below, we give some sufficient conditions of Hypothesis A to hold. Throughout the proofs, we drop the subscript-N from the notation, eg writing k or n, considering sequences implicitly. Recall that we abbreviate the TV distance from uniformity at time t as

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

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

Theorem 2.1.4. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ a sequence of finite, Abelian groups; for each $N \in \mathbb{N}$, define $Z_{(N)} := [Z_1, \dots, Z_{k_N}]$ by drawing $Z_1, \dots, Z_{k_N} \sim^{\mathrm{iid}} \mathrm{Unif}(G_N)$.*

Suppose that the sequence $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis A. Let $\alpha \in \mathbb{R}$. Then

$$d^{\pm}_{G_k,N}\big(t_\alpha(k_N, |G_N|)\big) \to^{\mathbb{P}} \Psi(\alpha) \quad \textit{(in probability)} \quad \text{as } N \to \infty.$$

That is, for all $\alpha \in \mathbb{R}$, whp, t_α is, up to sot, the mixing time $t_{\mathrm{mix}}(\Psi^{-1}(\alpha))$. Moreover, the implicit lower bound holds deterministically, ie for all choices of generators.

Remark. Using Proposition 2.1.2, we can write the cutoff statement in the form

$$\big(t_{\mathrm{mix}}(\varepsilon) - t_0\big)/w \to^{\mathbb{P}} \Psi^{-1}(\varepsilon) \quad \text{for all} \quad \varepsilon \in (0,1),$$

where $t_0 \approx k|G|^{2/k}/(2\pi e)$ is the mixing time and $w \approx \sqrt{k}|G|^{2/k}/(\sqrt{2}\pi e)$ the window. △

Remark 2.1.5. Write $n := |G|$. Note that the final condition of Hypothesis A implies that $k \le \frac{1}{5}\log n$; so we are in the regime $1 \ll k \lesssim \log n$. Any of the following conditions imply Hypothesis A:

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

△

Remark. The CLT, Proposition 2.1.3, will give the dominating term in the TV distance:
- on the event $\{Q(t_\alpha) \le \log n - \omega\}$, we lower bound the TV distance by $1 - o(1)$;
- on the event $\{Q(t_\alpha) \ge \log n + \omega\}$, we upper bound the expected TV distance by $o(1)$.

Combined with the CLT, we deduce that the $d_{G_k}(t_\alpha) \to \Psi(\alpha)$ in probability. △

Remark. Observe that Hypothesis A does not cover the regime $k \gtrsim \log|G|$. Under certain conditions on the group we can apply a variation on the argument given below to obtain a limit profile result for any k with $1 \ll \log k \ll \log|G|$. We do not carry out the analysis here; see §5.1. △

2.1.4 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; rather, this section merely introduces ideas. Recall the definitions from the previous sections.

In all cases we show that cutoff occurs around the entropic time. As $Q(t)$ is a sum of iid random variables, we expected it to be concentrated around its mean. Loosely speaking, we show that the shape of the cutoff, ie 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 2.1.3), are determined by $\mathbb{V}\mathrm{ar}(Q(t))$, for t 'close' to 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 (§2.1.4), we write 0 for the identity element of the Abelian group G. 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

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

But clearly $|E| \le ne^{-\omega}$. Choosing the time t slightly smaller than the entropic time t_0 and $\omega \gg 1$ suitably, the event $\{Q(t) \le \log n - \omega\}$ will hold whp. Thus, whp, $S(t)$ is restricted to a set of size $o(n)$. It hence 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 $Z = [Z_1, ..., Z_k]$, that

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

Observe that the probability on the right-hand side is independent of Z. Thus we are naturally interested in the fluctuations of $Q(t)$ for t close to t_0. Using the CLT application above, ie Proposition 2.1.3 with $\omega := \mathbb{V}\mathrm{ar}(Q(t_0))^{1/4}$, we deduce the lower bound in Theorem 2.1.4.

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 uniform randomness 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 un- 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 Cauchy–Schwarz) says that

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

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

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

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

$$\mathbb{P}\big(V(t_0) \cdot Z = 0 \mid Z\big) \geq \mathbb{P}\big(V(t_0) = (0, ..., 0) \in \mathbb{Z}^k\big) = \mathbb{P}\big(X(2t_0/k) = 0\big)^k \gg 1/n.$$

(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

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

note that here we are (implicitly) averaging over Z. (The set $\mathcal{W} \subseteq \mathbb{Z}^k$ will be chosen later; the idea is that W 'typically' lies in it.) We now perform the same L_2 calculation, but for $\overline{\mathbb{P}}$ rather than $\mathbb{P}$:

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

see Lemma 2.1.6. 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* ones say that

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

where $\omega := (vk)^{1/4}$ as above; the *local* ones will come later. For a precise statement of the typicality requirements, see Definition 2.1.7. These have the property that $\mathbb{P}(W \notin \mathcal{W}) = \Psi(\alpha) + o(1) \asymp 1$ when $t = t_\alpha$; see Proposition 2.1.8. This has the advantage that now

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

since $-\log x \geq \log n + \omega$ if and only if $x \leq n^{-1}e^{-\omega}$.

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 $v \cdot Z \sim \mathrm{Unif}(\mathfrak{g}_v G)$ where $\mathfrak{g}_v := \gcd(v_1, ..., v_k, n)$; see Lemma 2.1.11. (For an Abelian group G and $\gamma \in \mathbb{N}$, define $\gamma G := \big\{\sum_1^\gamma g \mid g \in G\big\}$; eg, $\gamma\mathbb{N} = \{\gamma, 2\gamma, ...\}$.) Then

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

(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 $\gamma \in [2r_*]$ for the gcd. It is here where the two approaches (§2.1 and §2.2) diverge.

In the first approach (§2.1) we use a rather direct approach. First, it is elementary that

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

see Lemma 2.1.12. 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 order $1/n^{1/k} \asymp 1/\sqrt{t_\alpha/k}$. This leads to

$$\overline{\mathbb{P}}(\mathfrak{g}_V = \gamma) \lesssim \big(2/n^{1/k} + 1/\gamma\big)^k;$$

see Lemma 2.1.14. Provided at least one of $d(G)$ or k is not too close to $\log n$, we are able to use this inequality to control the expectation, showing $\overline{\mathbb{E}}(\mathfrak{g}_V^{d(G)}\mathbf{1}(V \neq 0)) = 1 + o(1)$; see Corollary 2.1.15.

Combining these two analyses, we deduce that

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

The modified L_2 calculation then says that the TV distance is roughly $\Psi(\alpha)$ plus a term $o_{\mathbb{P}}(1)$, ie tending to 0 in probability. This establishes a matching limiting upper bound of $\Psi(\alpha)$ in probability.

The second approach (§2.2) analyses the term $\overline{\mathbb{P}}(\mathfrak{g}_V = \gamma)$ and uses it to kill $|G/\gamma G|$ directly in

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

We outline in more detail the adaptation in §2.2.5, including where Approach #1 breaks down.

This concludes the outline; we now move onto the formal proofs.

2.1.5 Lower Bound on Mixing

In this subsection, we prove the lower bound on mixing, which holds for every choice of Z.

Proof of Lower Bound in Theorem 2.1.4. For this proof, assume that Z is given and suppress it.

We convert the CLT, Proposition 2.1.3, from a concentration statement about Q into one about W: for all $\alpha \in \mathbb{R}$, by the CLT, we have

$$\mathbb{P}\big(\mathcal{E}_\alpha\big) \eqsim \Psi(\alpha) \quad \text{where} \quad \mathcal{E}_\alpha \coloneqq \big\{\mu\big(W(t_\alpha)\big) \ge n^{-1}e^{\omega}\big\} = \big\{Q(t_\alpha) \le \log n - \omega\big\};$$

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

$$E_\alpha \coloneqq \big\{x \in G \,\big|\, \exists\, w \in \mathbb{Z}^d \text{ st } \mu_{t_\alpha}(w) \ge n^{-1}e^{\omega} \text{ and } x = w \cdot Z\big\}.$$

Since we use W to generate S, we have $\mathbb{P}(S(t_\alpha) \in E_\alpha \mid \mathcal{E}_\alpha) = 1$. Every element $x \in E_\alpha$ can be realised as $x = w_x \cdot Z$ for some $w_x \in \mathbb{Z}^k$ with $\mu_{t_\alpha}(w_x) \ge n^{-1}e^{\omega}$. Hence, for all $x \in E_\alpha$, we have

$$\mathbb{P}\big(S(t_\alpha) = x\big) \ge \mathbb{P}\big(W(t_\alpha) = w_x\big) = \mu_{t_\alpha}(w_x) \ge n^{-1}e^{\omega}.$$

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

$$1 \ge \textstyle\sum_{x \in E_\alpha} \mathbb{P}\big(S(t_\alpha) = x\big) \ge |E_\alpha| \cdot n^{-1}e^{\omega}, \quad \text{and hence} \quad |E_\alpha|/n \le e^{-\omega} = o(1).$$

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

$$\big\|\mathbb{P}\big(S(t_\alpha) \in \cdot \mid Z\big) - \pi_G\big\|_{\mathrm{TV}} \ge \mathbb{P}\big(S(t_\alpha) \in E_\alpha\big) - \pi_G(E_\alpha) \ge \mathbb{P}(\mathcal{E}_\alpha) - \tfrac{1}{n}|E_\alpha| \ge \Psi(\alpha) - o(1). \qquad \square$$

Remark. Using a variant of this argument, in §3.1.5 we prove an analogous lower bound for general groups: where $t_0(k, |G|)$ was the lower bound above (for Abelian groups), we establish a lower bound of $t_0(k, |G/[G,G]|)$ for any group. (If a group is Abelian, then $[G,G]$ is trivial.) In many cases, this is a significant improvement over previous best-known bound of $\log_{k-1}|G|$. $\triangle$

2.1.6 Upper Bound on Mixing

It is often easier to consider L_2 distances than L_1: roughly, squares are easier to deal with than absolute values. TV has the significant advantage, though, of being uniformly bounded (by 1); as such, we can condition on high probability events, and upper bound the by 1 when this event fails.

We use a 'modified L_2 calculation': first conditioning that W is 'typical'; then using a standard L_2 calculation on the conditioned law. Let W' be an independent copy of W; then $S' \coloneqq W' \cdot Z$ is an independent copy of S.

Lemma 2.1.6. *For all $t \ge 0$ and all $\mathcal{W} \subseteq \mathbb{Z}^k$, the following inequalities hold:*

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

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

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

$$4\left\|\mathbb{P}_{G_k}\left(S\in\cdot\mid W\in\mathcal{W}\right)-\pi_G\right\|_{\mathrm{TV}}^2\le n\sum_x\left(\mathbb{P}_{G_k}\left(S=x\mid W\in\mathcal{W}\right)-\tfrac{1}{n}\right)^2$$
$$=n\sum_x\mathbb{P}_{G_k}\left(S=x\mid W\in\mathcal{W}\right)^2-1=n\sum_x\mathbb{P}_{G_k}\left(S=S'=x\mid W,W'\in\mathcal{W}\right)-1,$$

as $S=W\cdot Z$ and $S'=W'\cdot Z$. The claim follows by taking expectations. □

We now make the specific choice of the 'typical' set $\mathcal{W}$; we make a different choice for each $\alpha\in\mathbb{R}$. Write Ψ for the standard Gaussian tail. The collection $\{\mathcal{W}_\alpha\}_{\alpha\in\mathbb{R}}$ of sets will satisfy $\mathbb{P}(W(t_\alpha)\notin\mathcal{W}_\alpha)\approx\Psi(\alpha)$, using the CLT (Proposition 2.1.3). We show that the L_2 distance is $o(1)$; see Proposition 2.1.9. Applying Lemma 2.1.6, we find that $d_{G_k}(t_\alpha)\le\Psi(\alpha)+o(1)$ whp over Z. This matches the lower bound from §2.1.5.

By considering all $\alpha\in\mathbb{R}$, we are able to 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, $\mathbb{P}(W(t_\alpha)\notin\mathcal{W}_\alpha)\approx\Psi(\alpha)\approx 0$, which explains the use of the word 'typically' in describing $\mathcal{W}_\alpha$.

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

Definition 2.1.7. *For all $\alpha\in\mathbb{R}$, define the local and global typicality conditions, respectively:*

$$\mathcal{W}_{\alpha,\mathrm{loc}}:=\left\{w\in\mathbb{Z}^k\;\middle|\;\left|w_i-\mathbb{E}\big(W_1(t_\alpha)\big)\right|\le r_*\;\forall\,i=1,\dots,k\right\}\quad\text{where}\quad r_*:=\tfrac{1}{2}n^{1/k}(\log k)^2;$$
$$\mathcal{W}_{\alpha,\mathrm{glo}}:=\left\{w\in\mathbb{Z}^k\;\middle|\;\mathbb{P}\big(W(t_\alpha)=w\big)\le n^{-1}e^{-\omega}\right\}.$$

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 proposition determines the probability that $W(t_\alpha)$ lies in $\mathcal{W}_\alpha$, ie of typicality.

Proposition 2.1.8. *For each $\alpha\in\mathbb{R}$, we have*

$$\mathbb{P}\big(W(t_\alpha)\notin\mathcal{W}_\alpha\big)\to\Psi(\alpha).$$

Proof. By our CLT, Proposition 2.1.3, the probability that the global conditions hold converges to $1-\Psi(\alpha)$. Proposition 2.1.2 and Definitions 6.3.1 and 6.3.2 and Proposition 6.3.3 together say that the probability that a single coordinate fails the local condition is at most $k^{-3/2}$. By the union bound, the probability that local typicality fails to hold is then at most $k^{-1/2}=o(1)$. □

Herein, we fix $\alpha\in\mathbb{R}$ and frequently suppress the t_α from the notation, eg writing $W_\cdot$ for $W_\cdot(t_\alpha)$ or $\mathcal{W}$ for $\mathcal{W}_\alpha$. Let $V:=W-W'$, so $\{W\cdot Z=W'\cdot Z\}=\{V\cdot Z=0\}$. Write

$$D:=D(t_\alpha):=n\,\mathbb{P}\big(V(t_\alpha)\cdot Z=0\mid\mathsf{typ}_\alpha\big)-1\quad\text{where}\quad\mathsf{typ}:=\mathsf{typ}_\alpha:=\big\{W(t_\alpha),W'(t_\alpha)\in\mathcal{W}_\alpha)\big\}.$$

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

$$\frac{k-d}{k}-4\frac{d\log\log k}{\log n}\ge 10\frac{k}{\log n}\quad\text{and}\quad k-d\gg 1.$$

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

Proposition 2.1.9. *Suppose that (d,n,k) jointly satisfy Hypothesis A. (Recall that, implicitly, (d,n,k) is a sequence of integers.) Write $\mathfrak{g}:=\gcd(V_1,\dots,V_k,n)$. Then, for all $\alpha\in\mathbb{R}$, we have*

$$0\le D(t_\alpha)=\textstyle\sum_{\gamma\in\mathbb{N}}\mathbb{P}(\mathfrak{g}=\gamma\mid\mathsf{typ})\cdot|G|/|\gamma G|-1=o(1).$$

Given this proposition, we can prove the upper bound in the main theorem, Theorem 2.1.4.

Proof of Upper Bound in Theorem 2.1.4 Given Proposition 2.1.9. Hypothesis A are precisely the conditions required for Proposition 2.1.9. Apply the modified L_2 calculation, Lemma 2.1.6 and Definition 2.1.7, and use Propositions 2.1.8 and 2.1.9 to control the two resulting terms. Combined, these say that $d_{G_k}(t_\alpha)\le\Psi(\alpha)+o(1)$ whp over Z. □

It remains to prove Proposition 2.1.9, ie 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 probability; its proof is deferred to the end of the subsection.

Lemma 2.1.10. *We have*

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

Now, linear combinations of independent uniform random variables in an Abelian group are themselves uniform on their support. Hence the distribution of $v \cdot Z$ is uniform on $\gcd(v_1, ..., v_k, n)G$; this is proved carefully in Lemma 6.6.1. (Recall that $\gamma G := \{\gamma g \mid g \in G\}$ for $\gamma \in \mathbb{N}$.)

Lemma 2.1.11. *For all $v \in \mathbb{Z}^k$, we have*

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

We thus need to control $|\gamma G|$, since Lemma 2.1.11 implies that

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

Lemma 2.1.12. *For all Abelian groups H and all $\gamma \in \mathbb{N}$, we have*

$$|H|/|\gamma H| \le \gamma^{d(H)}.$$

Proof. Decompose H as $\oplus_1^d \mathbb{Z}_{m_j}$ with $d = d(H)$ and some $m_1, ..., m_d \in \mathbb{N}$. Then γH can be decomposed as $\oplus_1^d \gcd(\gamma, m_j)\mathbb{Z}_{m_j}$. Hence $|\gamma H| = \prod_1^d (m_j/\gcd(\gamma, m_j)) \ge \prod_1^d (m_j/\gamma) = |H|/\gamma^d$. □

These lemmas combine to produce a simple, but key, corollary. Recall that $\mathfrak{g} = \gcd(V_1, ..., V_k, n)$.

Corollary 2.1.13. *We have*

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

Proof. The conditioning does not affect Z. The corollary follows from Lemmas 2.1.11 and 2.1.12. □

In order to control this gcd, we need to determine the probability that an individual coordinate is a multiple of a given number. We evaluate the RW around the entropic time t_0. The proof of the following auxiliary lemma is deferred to the end of the subsection. This, along with Corollary 2.1.13, are the key elements to the proof of Proposition 2.1.9.

Lemma 2.1.14. *For all $\gamma \in \mathbb{N}$, we have*

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

From this, using the conditions of Hypothesis A, we can deduce that $\mathbb{E}(\mathfrak{g}^d\,\mathbf{1}(V \neq 0) \mid \mathsf{typ}) = 1 + o(1)$. We refer to this as a "corollary", since its proof is purely technical, not relying on any properties of the RW or the generators, just algebraic manipulation. Its proof is briefly deferred.

Corollary 2.1.15. *Given Hypothesis A, we have $\mathbb{E}(\mathfrak{g}^d\,\mathbf{1}(V \neq 0) \mid \mathsf{typ}) = 1 + o(1)$.*

Proposition 2.1.9 now follows immediately from Lemma 2.1.10 and Corollaries 2.1.13 and 2.1.15.

Proof of Proposition 2.1.9. By Lemma 2.1.10 and Corollaries 2.1.13 and 2.1.15, we have

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

□

We now give the deferred proof of Corollary 2.1.15.

Proof of Corollary 2.1.15. By local typicality, $\mathfrak{g} \le 2r_* = n^{1/k}(\log k)^2$ when $V \neq 0$. Hence

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

For $\gamma \ge 2$, we use Lemma 2.1.14. Let $\delta \in (0,1)$. For $2 \le \gamma \le \delta n^{1/k}$, we use the bound

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

For $\gamma \ge \delta n^{1/k}$, we use the slightly crude bound

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

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

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

This is $o(1)$, by the conditions of Hypothesis A, as we now outline. Write $\eta := (k-d-1)/k \in (0,1)$. We wish to choose δ as large as possible so that the first term is $o(1)$; set $\delta := \frac{1}{4}\eta$. With this definition, it is not difficult to see that the assumption $\eta \ge 4k/\log n$, which follows immediately from Hypothesis A, is sufficient to make the middle term small. Finally, the inequality in Hypothesis A is designed precisely so that the final term is $o(1)$, noting that $\eta k \ge k - d - 1 \ge \frac{1}{2}(k-d)$. □

Remark 2.1.16. From our analysis, it follows that if $k - d = M \ge 2$ is fixed (ie not diverging), then the order of the mixing time is still given by t_0. However, our argument does not give cutoff in this case. For many groups we expect there to be cutoff, but not always. In fact, for certain groups, eg $\mathbb{Z}_2^d$, it is not even the case that the group is generated whp. △

It remains to prove the auxiliary lemmas, namely Lemmas 2.1.10 and 2.1.14.

Proof of Lemma 2.1.10. By direct calculation, since W and W' are independent copies,

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

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

$$n\,\mathbb{P}\big(V = 0 \mid \mathsf{typ}\big) \le n \textstyle\sum_{w\in\mathcal{W}} \mathbb{P}\big(W = w\big)^2/\mathbb{P}(\mathsf{typ}) \le e^{-\omega}/\mathbb{P}(\mathsf{typ}).$$ □

Proof of Lemma 2.1.14. Let $X = (X_s)_{s\ge 0}$ be a rate-1 SRW on $\mathbb{Z}$. To calculate the expectation, we use that $V = W - W'$ has the distribution of a SRW run at twice the speed; in particular, $V_i(t) \sim X_{2t/k}$, and that coordinates of V are independent. (This holds for both the un- and directed cases.) Clearly the distribution of X is symmetric about 0.

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

$$|V_1| \sim \mathrm{Unif}\{1,\dots,Y\} \quad \text{conditional on} \quad V_1 \neq 0,$$

where Y has some distribution. Hence we have

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

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

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

noting that $\mathbb{P}(\mathsf{typ}) \asymp 1$. Using Proposition 2.1.2 to argue that $\mathbb{P}(V_1 = 0) \le 2/n^{1/k}$, we deduce that

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

2.2 TV Cutoff: Approach #2

Recall that the cutoff statement for arbitrary Abelian groups, 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'. In this section we use a new approach to deal with the case that k is 'not too small'. The main result of the section is Theorem 2.2.6; see also Hypothesis B and Remark 2.2.7.

The outline of this section is roughly the same as that of the previous one:

- §2.2.1 discusses the new, refined entropic methodology;
- §2.2.2 defines the new *entropic times*;
- §2.2.3 states bounds on the growth rate of the entropy and concentration;
- §2.2.4 states precisely the main theorem of the section;
- §2.2.5 outlines the differences between this argument and the previous approach;
- §2.2.6 is devoted to the lower bound;
- §2.2.7 is devoted to the upper bound.

2.2.1 Entropic Times: New Methodology and Definition

The underlying principles of the method used in this section (§2.2) are the same as those of the previous (§2.1). We adjust the method slightly to deal with the cases not covered in §2.1.

We first discuss where the previous approach broke down, and how we might fix it. The primary issue was when d was very large. Eg consider $\mathbb{Z}_2^d$. Since all elements are of order 2, instead of looking at W, a RW on $\mathbb{Z}$, we could equally have looked at W taken modulo 2. The entropy of W_1 mod 2 may be significantly smaller than that of W_1 at the original entropic time t_0.

We saw that $V \cdot Z \sim \mathrm{Unif}(\gamma G)$ when $\gcd(V_1, ..., V_k, n) = \gamma$. (This assumes that the group G is Abelian.) This motivates defining t_γ to be the time at which the entropy of W_1 mod γ is $\log|G/\gamma G|$. The proposed upper bound is then given by $t_* := \max_{\gamma\in\mathbb{N}} t_\gamma$.

While this method will be able to handle arbitrary Abelian groups, we only get an abstract definition of the cutoff time, which is not easily calculable for many groups.

As in the previous sections, not only are we interested in the entropy at this proposed mixing time t_*, but we also desire quantitative information about the rate of change of entropy at this time, and the variance of the 'random entropy', denoted Q.

2.2.2 Entropic Times: Definition and Concentration

In this section, we *redefine* entropic times. There is some overlap with notation from before, but all entropic definitions from §2.1.1 should be forgotten; all terms will be defined below.

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 $\mathbb{Z}$, counting the uses of generators, as in the previous sections. (This can be either a SRW on $\mathbb{Z}$ or DRW on $\mathbb{Z}_+$.) As before, $S(t) = W(t)\cdot Z$. For $\gamma \in \mathbb{N}$, define W_γ via $W_{\gamma,i}(t) := W_i(t) \bmod \gamma$; write $W_\infty := W$. Then W_γ is a RW on $\mathbb{Z}_\gamma^k$; so $W_{\gamma,i} := (W_{\gamma,i}(t))_{t\geq 0}$ forms an iid sequence (over $i \in [k]$) of rate-$1/k$ RWs on $\mathbb{Z}_\gamma$.

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

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

then, $Q_{\gamma,i}$ forms an iid sequence over $i \in [k]$, and

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

So $h_\gamma(t)$ and $H_\gamma(s)$ are the entropies of $W(t)$ and $W_1(sk)$, respectively. Note that $h_\gamma(t) = kH_\gamma(t/k)$ and that $h_\gamma : [0,\infty) \to [0, \log(\gamma^k))$ is a strictly increasing bijection.

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

Definition 2.2.1. *For $N < \gamma^k$, define the entropic time*

$$t_0(\gamma, N) := h_\gamma^{-1}(\log N) \quad \textit{and} \quad s_0(\gamma, N) := t_0(\gamma, N)/k = H_\gamma^{-1}(\log N/k).$$

We are interested primarily in $N := |G/\gamma G|$. For an Abelian group G, define

$$t_*(k, G) := \max_{\gamma \wr |G|} t_0(\gamma, |G/\gamma G|).$$

Our next result determines the asymptotics of t_*. The first part is for $k - d(G) \asymp k$: it shows that here the mixing time is the same order as that from Approach #1, ie $kn^{2/k}$. Combining the two approaches, this means that all Abelian groups have mixing time order $kn^{2/k}$ when $1 \ll k \lesssim \log n$ and $k - d(G) \asymp k$. The second part allows $k - d$ to diverge arbitrarily slowly: in this case the mixing time can be as large as order $kn^{2/k} \log k$. The final part evaluates t_* up to sot when $d(G) \ll \log |G|$ and $k - d(G) \asymp k$. The proofs are given in §6.2.3.2.

Proposition 2.2.2a (Proposition 6.2.17)**.** *Suppose that $1 \ll k \lesssim \log |G|$. The following hold:*

$$\begin{aligned} &\text{if} \quad k - d(G) \asymp k, \quad \text{then} \quad t_*(k, G) \asymp k|G|^{2/k}; \\ &\text{if} \quad k - d(G) > 1, \quad \text{then} \quad t_*(k, G) \lesssim k|G|^{2/k} \log k. \end{aligned}$$

Proposition 2.2.2b (Proposition 6.2.18)**.** *Suppose that $d(G) \ll \log |G|$ and $k - d(G) \asymp k \gg 1$. Then $t_*(k, G) \asymp t_0(\infty, |G|)$. (Note that $t_0(\infty, |G|) = t_0(k, |G|)$ from Definition 2.1.1.)*

Heuristics Behind Proofs. For the RW on $\mathbb{Z}_\gamma$, until time γ^2 the walk looks roughly the same as if it were on $\mathbb{Z}$. In particular, the entropy growth rates are comparable. From this, we are able to see that s_* is the same order as the entropic time s_0 from §2.1 when $k - d \asymp k$.

For $k - d \gg 1$, by Lemma 2.1.12, we have $s_0(\gamma, |G/\gamma G|) \le s_0(\gamma, \gamma^d) = s_0(\gamma, |\mathbb{Z}_\gamma^d|) = R_\gamma^{-1}(\zeta_\gamma)$. So the worst case is studying relative entropy for the RW on $\mathbb{Z}_\gamma^d$. In §5.2 we analyse in detail RWs on random Cayley graphs of $\mathbb{Z}_p^d$. In particular, we analyse this entropic time for $1 \ll k - d \ll k$.

The same heuristics hold for the regime $1 \ll k \ll \log |G|$, except that now one checks that the optimal γ satisfies $\gamma \gg 1$ and $s_0(\gamma, \log |G/\gamma G|) \ll \gamma^2$. In this case, the RW on $\mathbb{Z}_\gamma$ is almost indistinguishable from that on $\mathbb{Z}$. Hence the entropic times are asymptotically equivalent.

For $k \asymp \log |G|$, the mixing time is order $k \asymp \log |G|$. As such one expects each generator to be picked an order 1 number of times. One can then separate into large γ and small γ; upper bound $|G/\gamma G| \le |G|$ in the former case and $|G/\gamma G| \le \gamma^{d(G)}$ in the latter. The optimal γ must be small. □

In §2.2.6, we show that $t_0(\gamma, |G/\gamma G|)$ is a lower bound on mixing for all γ, for all Z. Throughout this section, we work under the assumption that $k \lesssim \log n$. (Recall from §1.5.2 that cutoff had already been established for all Abelian groups when $k \gg \log n$.) As a result of this, taking $\gamma := n$, we see that the mixing time is at least order k. There hence exists a $\varsigma > 0$ so that the mixing time is at least $2\varsigma k$. (This is true for all Z, not just whp over Z.)

The following definitions are made purely for technical convenience.

Definition 2.2.3. *For $s \ge 0$ and $\gamma \ge 2$, define the (adjusted) entropic time and relative entropy via*

$$s_\gamma := s_0(\gamma, |G/\gamma G|) \vee \varsigma, \quad t_\gamma := s_\gamma k \quad \text{and} \quad R_\gamma(s) := \log \gamma - H_\gamma(s).$$

We have $\max_{\gamma \in \mathbb{N}} t_\gamma = \max_{\gamma \in \mathbb{N}} t_0(\gamma, |G/\gamma G|)$.

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 $\mathrm{Unif}(\mathbb{Z}_\gamma)$.

2.2.3 Entropic Times: Entropy Growth Rate and Concentration

In the previous approach, we had a CLT for the random variable Q. Here we do not give such precise results; this means that while we show cutoff, we do not find the profile. (Even if we knew such refined information, it would be difficult to calculate $\max_{\gamma \in \mathbb{N}} t_\gamma$, as this is highly dependent on the group.) Instead, we determine the rate of change of the entropy around the entropic time, and determine concentration estimates on the 'random entropy', ie the Q_γ random variable, at a time shortly after the entropic time.

The first lemma controls the rate of change of the entropy near the entropic time; see §6.2.

Lemma 2.2.4 (Lemma 6.2.20). *There exists a continuous function $\bar{c} : (0,1) \to (0,1)$ so that, for all $\gamma \geq 2$, all $\xi \in (-1,1) \setminus \{0\}$ and all $s \geq \varsigma$, we have*

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

Given that we know how much the entropy, ie the expectation of Q_γ, changes, we now want a concentration result, giving upper and lower tail estimates. The upper tail is used for the lower bound on mixing: it says that Q_γ is *at most* some value whp. Similarly, the lower tail is used for the upper bound on mixing. These are given in Proposition 2.2.5, which is proved in §6.2.

Recall that $t_* = \max_{\gamma \in \mathbb{N}} t_\gamma$ and $d = d(G)$. For $\gamma \in \mathbb{N}$, write $\zeta_\gamma \coloneqq \frac{1}{k}(k - d(G)) \log \gamma$.

Proposition 2.2.5 (Proposition 6.2.21). *Assume that $k > d$. There exists a continuous function $c : (0,1) \to (0,1)$ so that, for all $\gamma \geq 2$ and all $\varepsilon \in (0,1)$, the following hold:*

$$\mathbb{P}\big(Q_\gamma\big(t_*(1+\varepsilon)\big) \leq \log|G/\gamma G| + c_\varepsilon(\zeta_\gamma \wedge 1)k\big) \leq \exp\big(-c_\varepsilon(\zeta_\gamma \wedge 1)k\big);$$
$$\mathbb{P}\big(Q_\gamma\big(t(1-\varepsilon)\big) \geq \log|G/\gamma G| - c_\varepsilon(\zeta_\gamma \wedge 1)k\big) = o(1) \quad \text{for all} \quad t \leq t_\gamma.$$

The proof of this proposition is given in §6.2. We give a brief outline here. Recall that $Q_\gamma(t) = \sum_1^k Q_{\gamma,i}(t)$ is a sum of iid terms, each of which has mean $H_\gamma(t/k)/k$. Applying the entropy growth rate lemma, ie Lemma 2.2.4, we see, for any $\xi \in (-1,1) \setminus \{0\}$, that the change in entropy between times s and $(1+\xi)s$ is order $R_\gamma(s) \wedge 1$ (with implicit constant depending on $|\xi|$). Taking $s \coloneqq s_0(\gamma, |G/\gamma G|)$, recalling that $|G/\gamma G| \leq \gamma^{d(G)}$ by Lemma 2.1.12, gives

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

(We are interested in the times s_γ, not $s_0(\gamma, |G/\gamma G|)$; this is a minor technical complication.)

Regarding the concentration, the non-quantitative part is then an application of Chebyshev's inequality, once one has shown that the variance $\mathbb{V}\mathrm{ar}(Q_{\gamma,1}(sk))$ is uniformly bounded over $s \geq \varsigma$; the quantitative part requires a (one-sided) large deviations estimate.

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

$$\limsup_N k_N/\log|G_N| < \infty, \quad \liminf_N\big(k_N - d(G_N)\big) = \infty \quad \text{and} \quad \liminf_N k_N/\log|\mathcal{H}_N| = \infty,$$
$$\text{where} \quad \mathcal{H}_N \coloneqq \big\{\gamma G_N \;\big|\; \gamma \wr |G_N| \text{ and } \gamma \in [2, n_{*,N}]\big\} \quad \text{and} \quad n_{*,N} \coloneqq \lfloor |G_N|^{1/k_N}(\log k_N)^2 \rfloor.$$

In Remark 2.2.7 below, we give a sufficient condition for Hypothesis B to hold. Throughout the proofs, we drop the subscript-N from the notation, eg writing k or n, considering sequences implicitly. Recall that we abbreviate the TV distance from uniformity at time t as

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

We now state the main theorem of this section. Recall that $t_* = \max_{\gamma\in\mathbb{N}} t_0(\gamma, |G/\gamma G|)$.

Theorem 2.2.6. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ a sequence of finite, Abelian groups; for each $N \in \mathbb{N}$, define $Z_{(N)} \coloneqq [Z_1,\dots,Z_{k_N}]$ by drawing $Z_1,\dots,Z_{k_N} \sim^{\mathrm{iid}} \mathrm{Unif}(G_N)$.*

Suppose that the sequence $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis B. Let $c \in (-1,1) \setminus \{0\}$. Then

$$d^{\pm}_{G_k,N}\big((1+c)t^{\pm}_*(k_N, G_N)\big) \to^{\mathbb{P}} \mathbf{1}(c<0) \quad \textit{(in probability)} \quad \text{as } N \to \infty.$$

That is, whp, there is cutoff at $\max_\gamma t^{\pm}_0(\gamma, |G/\gamma G|)$. Moreover, the implicit lower bound holds deterministically, ie for all choices of generators.

Remark 2.2.7. If $k \gg \sqrt{\log n}$, then $k \gg \log|\mathcal{H}|$, since $|\mathcal{H}| \leq n_* \leq n^{1/k}(\log k)^2$. △

2.2.5 Outline of Proof

The general outline of this approach is the same as that of the previous; see §2.1.4 for an outline of the previous approach. The previous 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 can only decrease the TV distance. The idea, then, is that where before we looked at a RW on $\mathbb{Z}^k$ and waited until it has entropy $\log|G|$, instead we look at a RW on $\mathbb{Z}^k_\gamma$ and wait until it has entropy $\log|G/\gamma G|$; see Definition 2.2.1. We then take a worst-case over $\gamma\in\mathbb{N}$.

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

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

see Propositions 2.1.9 and 2.2.13. In §2.1.6, we upper bounded $|G|/|\gamma G|\le\gamma^{d(G)}$. In certain situations, this is too crude. Instead, observe that if $\mathfrak{g}=\gamma$ then $V\equiv 0 \bmod\gamma$. But $W_\gamma:=W \bmod\gamma$ and $W'_\gamma:=W' \bmod\gamma$ are simply RWs on $\mathbb{Z}^k_\gamma$. Just as we used entropy and typicality to get

$$\mathbb{P}\big(W=W'\mid\mathsf{typ}\big)\ll 1/|G|$$

in Lemma 2.1.10, here we adjust the entropic time (and typicality) so that

$$\mathbb{P}\big(W_\gamma=W'_\gamma\mid\mathsf{typ}\big)\ll|\gamma G|/|G|;$$

see Definitions 2.2.1 and 2.2.8 and the proof of Proposition 2.2.13.

2.2.6 Lower Bound on Mixing

In this subsection, we state and prove the lower bound, matching the upper bound of Theorem 2.2.6; it holds not only for all groups G but also for all choices of Z, not just whp.

The idea is to quotient out by γG, and show that the walk on this quotient is not mixed at time $(1-\varepsilon)t_0(\gamma,|G/\gamma G|)$, and hence the original walk is not mixed on G either. We use the same idea as in §2.1.5 to show that, for each γ, the walk is not mixed on $G/\gamma G$ at time $(1-\varepsilon)t_0(\gamma,|G/\gamma,G|)$.

In §2.2.6 we used a CLT to control the entropic variables. Here we use the entropy growth rate and variance bounds, detailed in Proposition 2.2.5.

Proof of Lower Bound in Theorem 2.2.6. For this proof, assume that Z is given and suppress it.

We first convert the statement from one about Q_γ to one about W_γ. Let $\varepsilon\in(0,1)$ and set $t:=(1-\varepsilon)t_0(\gamma,|G/\gamma G|)$. Write $\zeta_\gamma:=R_\gamma(s_0(\gamma,|G/\gamma G|))$. From Proposition 2.2.5, we obtain

$$\mathbb{P}(\mathcal{E})=1-o(1)\quad\text{where}\quad\mathcal{E}:=\big\{\mu_{\gamma,t}\big(W_\gamma(t)\big)\ge\delta_{\gamma,\varepsilon}^{-1}/|G/\gamma G|\big\}\quad\text{and}\quad\delta_{\gamma,\varepsilon}:=\exp\big(-c_\varepsilon(\zeta_\gamma\wedge 1)k\big).$$

From Lemma 2.1.12, we have $|G/\gamma G|\le\gamma^{d(G)}$. Thus

$$\zeta_\gamma=R_\gamma\big(s_0(\gamma,\log|G/\gamma G|)\big)=\log\gamma-\log|G/\gamma G|/k\ge\tfrac{1}{k}\big(k-d(G)\big)\log\gamma;$$

also, $k-d(G)\gg 1$. Thus $\delta_{\gamma,\varepsilon}=o(1)$ uniformly in γ. Consider the set

$$A:=\big\{x\in G/\gamma G\;\big|\;\exists\, w\in\mathbb{Z}^k_\gamma\text{ st }\mu_{\gamma,t}(w)\ge\delta_{\gamma,\varepsilon}^{-1}/|G/\gamma G|\text{ and }x=(w\cdot Z)\gamma G\big\}.$$

Define S_γ to be the projection of S to $G/\gamma G$. Since we use W to generate S_γ, we have $\mathbb{P}(S_\gamma(t)\in A\mid\mathcal{E})=1$. Every element $x\in A$ can be realised as $x=w_x\cdot Z$ for some $w_x\in\mathbb{Z}^k_\gamma$ with $\mu_{\gamma,t}(w_x)\ge\delta_{\gamma,\varepsilon}^{-1}/|G/\gamma G|$. Hence, for all $x\in A$, we have

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

recalling that S_γ lives in the quotient $G/\gamma G$. Summing over $x\in A$, we deduce that

$$1\ge\textstyle\sum_{x\in A}\mathbb{P}\big(S_\gamma(t)=x\big)\ge|A|\cdot\delta_{\gamma,\varepsilon}^{-1}/|G/\gamma G|,\quad\text{and hence}\quad|A|/|G/\gamma G|\le\delta_{\gamma,\varepsilon}=o(1).$$

Projecting onto $G/\gamma G$ (which can only decrease the TV distance), we see that

$$\big\|\mathbb{P}_{G_k}\big(S(t)\in\cdot\big)-\pi_G\big\|_{\mathrm{TV}}\ge\mathbb{P}\big(S_\gamma(t)\in A\big)-\pi_{G/\gamma G}(A)\ge\mathbb{P}(\mathcal{E})-|A|/|G/\gamma G|=1-o(1).$$

Finally, recall that $\max_{\gamma\in\mathbb{N}}t_\gamma=\max_{\gamma\in\mathbb{N}}t_0(\gamma,|G/\gamma G|)$. This completes the proof. □

2.2.7 Upper Bound on Mixing

To upper bound the mixing time, we use a 'modified L_2 calculation', as in the previous approach. This involves first conditioning that W has some 'typical' properties, laid out in the following definition, and then performing a standard TV–L_2 upper bound on the conditioned law.

Abbreviate $t_{*,\varepsilon} := t_*(1+\varepsilon)$. Recall that $d = d(G)$ and $\zeta_\gamma = \frac{1}{k}(k-d)\log\gamma$; set $\hat{\zeta}_\gamma := \zeta_\gamma \wedge 1$.

Definition 2.2.8. *Let $\varepsilon > 0$; recall the constant $c_\varepsilon > 0$ from Proposition 2.2.5. The following depend on ε; we suppress this in the notation. Define global typical sets for $\gamma \in \mathbb{N}$ by*

$$\mathcal{W}_{\gamma,\mathrm{glo}} := \big\{w \in \mathbb{Z}_\gamma^k \;\big|\; \mathbb{P}(W_\gamma(t_{*,\varepsilon}) = w) \le \delta_{\gamma,\varepsilon}/|G/\gamma G|\big\} \quad \text{where} \quad \delta_\gamma := \delta_{\gamma,\varepsilon} := e^{-c_\varepsilon \hat{\zeta}_\gamma k}.$$

Also define $\delta_\infty := \delta_{\infty,\varepsilon} := e^{-c_\varepsilon k}$. Define the local typicality set by

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

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

$$\mathsf{typ} := \big\{W(t_{*,\varepsilon}), W'(t_{*,\varepsilon}) \in \mathcal{W}_{\mathrm{loc}}\big\} \cap \big(\cap_{\gamma\in\Gamma}\big\{W_\gamma(t_{*,\varepsilon}), W'_\gamma(t_{*,\varepsilon}) \in \mathcal{W}_{\gamma,\mathrm{glo}}\big\}\big),$$

where Γ is a subset of $[2,n]$ to be defined below in Definition 2.2.11.

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

Lemma 2.2.9. *For all $\Gamma \subseteq \mathbb{N}\setminus\{1\}$, we have $\sum_{\gamma\in\Gamma}\delta_\gamma \le \delta_{\infty,\varepsilon}|\Gamma| + o(1)$.*

Proof. Since $\min\Gamma \ge 2$ and $k - d \gg 1$, we have

$$\textstyle\sum_{\gamma\in\Gamma}\delta_\gamma \le \sum_{\gamma\in\Gamma}(e^{-c_\varepsilon k} + e^{-c_\varepsilon\zeta_\gamma k}) = e^{-c_\varepsilon k}|\Gamma| + \sum_{\gamma\in\Gamma}\gamma^{-c_\varepsilon(k-d)} = \delta_\infty|\Gamma| + o(1). \qquad \square$$

Proposition 2.2.10. *For all $\varepsilon > 0$ and any subset $\Gamma \subseteq \mathbb{N}\setminus\{1\}$, we have*

$$\mathbb{P}(\mathsf{typ}) \ge 1 - 2\delta_{\infty,\varepsilon}|\Gamma| - o(1).$$

Proof. Suppress the time-dependence from the notation, eg writing W for $W(t_{*,\varepsilon})$.

Consider global typicality. First, observe that

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

Hence, recalling that $\delta_\gamma = \delta_{\gamma,\varepsilon} = \exp(-c_\varepsilon\hat{\zeta}_\gamma k)$, by Proposition 2.2.5, we have

$$\mathbb{P}\big(\mu_\gamma(W_\gamma) \le \delta_\gamma/|G/\gamma G|\big) \le \delta_\gamma, \quad \text{and hence} \quad \mathbb{P}\big(\cap_{\gamma\in\Gamma}\big\{W_\gamma \in \mathcal{W}_{\gamma,\mathrm{glo}}\big\}\big) \ge 1 - \textstyle\sum_{\gamma\in\Gamma}\delta_\gamma,$$

by the union bound. Recall that $\zeta_\gamma = \frac{1}{k}(k-d)\log\gamma$. Applying Lemma 2.2.9, we deduce that

$$\mathbb{P}\big(\cap_{\gamma\in\Gamma}\big\{W_\gamma \in \mathcal{W}_{\gamma,\mathrm{glo}}\big\}\big) \ge 1 - \delta_{\infty,\varepsilon}|\Gamma| - o(1) \quad \text{where} \quad \delta_{\infty,\varepsilon} = e^{-c_\varepsilon k}.$$

Now consider local typicality. Proposition 2.2.2a says that $t/k \le |G|^{2/k}\log k$. Then Definitions 6.3.1 and 6.3.2 and Proposition 6.3.3 together give

$$\mathbb{P}\big(\cap_i\big\{|W_i - \mathbb{E}(W_i)| \le r_*\big\}\big) = 1 - o(1), \quad \text{and hence} \quad \mathbb{P}\big(W \in \mathcal{W}_{\mathrm{loc}}\big) = 1 - o(1).$$

The claim follows by combining local and global typicality and applying the union bound. $\square$

We now choose the set Γ, to make sense of typicality. Recall that "$H \le G$" means that H is a subgroup of G and that we write $\alpha \wr \beta$, for $\alpha, \beta \in \mathbb{N}$, if α divides β.

Definition 2.2.11. *Abbreviate $n_* := (n-1) \wedge \lfloor 2r_*\rfloor$. Define $\Delta := \{\gamma \in [2, n_*] \mid \gamma \wr n\}$. Write $\mathcal{H}$ for the set of all proper subgroups H of G which can be represented as $H = \gamma G$ for some $\gamma \in \Delta$:*

$$\mathcal{H} := \big\{H \mid H = \gamma G \ne G \text{ for some } \gamma \wr n \text{ with } 2 \le \gamma \le n_*\big\}.$$

Given $H \in \mathcal{H}$, write $\Gamma_H := \{\gamma \in \Delta \mid H = \gamma G\}$ and denote by γ_H the minimal $\gamma \wr n$ so that $H = \gamma G$, ie $\gamma_H := \inf\Gamma_H$. 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 the subsection, will also be needed.

Lemma 2.2.12. *Given $H \in \mathcal{H}$, for all $\gamma \in \Gamma_H$, we have $\gamma_H \wr \gamma$.*

As shown below, we can combine our results to control the L_2 distance conditioned on typicality. In analogy with §2.1.6 and Proposition 2.1.9, write

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

Proposition 2.2.13. *Write $\mathfrak{g} := \gcd(V_1, ..., V_k, n)$. Then, for all $\varepsilon \in (0,1)$, we have*

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

(The conditions of Hypothesis B imply immediately that this last term is $o(1)$.)

From Propositions 2.2.10 and 2.2.13, it is straightforward to deduce the upper bound on mixing.

Proof of Upper Bound in Theorem 2.2.6. We use a modified L_2 calculation.

- Condition that W satisfies typicality; see Definition 2.2.8 and Proposition 2.2.10.
- Perform the standard TV–L_2 upper bound on the law of S conditioned that W is typical.
- Upper bound the resulting L_2 distance by $(\delta_{\infty,\varepsilon}|\mathcal{H}| + o(1))/\mathbb{P}(\mathsf{typ})$; see Proposition 2.2.13.
- This gives an upper bound on the expected TV distance of $(\delta_{\infty,\varepsilon}|\mathcal{H}| + o(1))/\mathbb{P}(\mathsf{typ}) + \mathbb{P}(\mathsf{typ}^c)$.
- From the definition of Γ, it is clear that $|\Gamma| \le |\mathcal{H}| + 1$. Since $\delta_{\infty,\varepsilon} = e^{-c_\varepsilon k} = o(1)$, with c_ε an arbitrary constant, the assumed condition $k \gg \log|\mathcal{H}|$ gives a final bound of $o(1)$ on the expected TV distance, recalling that $\mathbb{P}(\mathsf{typ}) = 1 - o(1)$ by Proposition 2.2.10.
- By Markov's inequality, this means that the TV distance is $o(1)$ whp over Z.

These calculations are all performed at time $t_{*,\varepsilon} = (1+\varepsilon)\max_\gamma t_\gamma$. This completes the proof. □

We now prove Proposition 2.2.13. To ease exposition, while all terms are evaluated at time $t_{*,\varepsilon} = (1+\varepsilon)\max_{\gamma\in\mathbb{N}} t_\gamma$, we suppress this from the notation.

Proof of Proposition 2.2.13. Write $V := W - W'$ and $\mathfrak{g} := \gcd(V_{\infty,1}, ..., V_{\infty,k}, n)$. If $\mathfrak{g} = \gamma$, which must have $\gamma \wr n$ as the gcd is with $n = |G|$, then $V \cdot Z \sim \mathrm{Unif}(\gamma G)$ by Lemma 2.1.11. Then

$$\big\|\mathbb{P}\big(S \in \cdot \mid \mathsf{typ}\big) - \pi_G\big\|_2^2 = n\,\mathbb{P}\big(V\cdot Z = 0 \mid \mathsf{typ}\big) - 1 = |G| \textstyle\sum_{\gamma\wr n} \mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ})/|\gamma G| - 1.$$

We consider various cases. For γ such that $\gamma G = G$, we have $|\gamma G| = |G|$ and upper bound

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

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

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

cf Lemma 2.1.10. If $V_\infty \neq 0$, then, given (local) typicality, $\mathfrak{g} \le n_* = (n-1) \wedge \lfloor 2r_* \rfloor$.

So it remains to study $\gamma \in \Delta$. As a consequence of Lemma 2.2.12, for any $H \in \mathcal{H}$, we have

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

(Recall that $V_\gamma \in \mathbb{Z}_\gamma^k$ for each γ.) This is key: it allows us to collapse the consideration of all $\gamma \in \Gamma_H$ down to the single element γ_H. Indeed, using the above we have

$$\begin{aligned}\textstyle\sum_{\gamma\in\Gamma_H} \mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ})/|\gamma G| &= \mathbb{P}\big(\textstyle\cup_{\gamma\in\Gamma_H}\{\mathfrak{g} = \gamma\} \mid \mathsf{typ}\big)/|H| \\ &\le \mathbb{P}\big(V_\gamma = 0 \text{ for some } \gamma \in \Gamma_H\big)/|H| \le \mathbb{P}\big(V_{\gamma_H} = 0 \mid \mathsf{typ}\big)/|H| \le (\delta_{\gamma_H}/|G|)/\mathbb{P}(\mathsf{typ}),\end{aligned}$$

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

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

(Note that every γ gives rise to a unique H such that $\gamma G = H$ and, by definition, $\mathcal{H}$ is the set of all H which can be obtained as γG for some γ; hence this decomposition neither overcounts nor undercounts $\gamma \in \Delta$.) Combining all these and using Lemma 2.2.9, we deduce the proposition:

$$0 \le n\,\mathbb{P}\big(V \cdot Z = 0 \mid \mathsf{typ}\big) - 1 = |G| \sum_{\gamma \wr n} \mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ})/|\gamma G| - 1 \le \big(\delta_\infty |\mathcal{H}| + o(1)\big)/\mathbb{P}(\mathsf{typ}). \quad \square$$

It remains to give the deferred proof of the divisibility lemma, namely Lemma 2.2.12.

Proof of Lemma 2.2.12. Consider any decomposition of G as $\oplus_1^r \mathbb{Z}_{m_j}$; this does not require $r = d(G)$. Fix some $\beta \in \Gamma_H$. Since $\alpha G = \beta G$ if and only if $\gcd(\alpha, m_j) = \gcd(\beta, m_j)$ for all j, we may decompose H as $\oplus_1^r h_j \mathbb{Z}_{m_j}$ where $h_j := \gcd(\beta, m_j)$ for all j. Set $\gamma_* := \operatorname{lcm}(h_1, ..., h_r)$. We show that $\gamma_* G = H$ and that $\gamma_* \wr \alpha$ for all $\alpha \in \Gamma_H$; this proves the lemma.

Fix $j \in [r]$. Now, $h_j \wr \gamma_* = \operatorname{lcm}(h_1, ..., h_r)$ 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 $\gamma_* = \operatorname{lcm}(h_1, ..., h_r) \wr \beta$ 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 $\operatorname{lcm}(h_1, .., h_r) \wr \alpha$, ie $\gamma_* \wr \alpha$. $\square$

2.3 TV Cutoff: Combining Approaches #1 and #2

In this section we combine the analysis from the previous two approaches to study the regime

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

We use the more refined notion of the entropic times; see §2.2.2.

2.3.1 Precise Statements and Results

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

$$\liminf_N k_N/\sqrt{\log|G_N|/\log\log\log|G_N|} > 0, \quad \limsup_N k_N/\sqrt{\log|G_N|} < \infty,$$
$$\liminf_N \big(k_N - d(G_N)\big) = \infty \quad \textit{and} \quad \limsup_N \big(k_N - d(G_N)\big)/k_N = 0.$$

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

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

We now state the main theorem of this section. Recall that $t_* = \max_{\gamma\in\mathbb{N}} t_0(\gamma, |G/\gamma G|)$.

Theorem 2.3.1. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ a sequence of finite, Abelian groups; for each $N \in \mathbb{N}$, define $Z_{(N)} := [Z_1, ..., Z_{k_N}]$ by drawing $Z_1, ..., Z_{k_N} \sim^{\mathrm{iid}} \mathrm{Unif}(G_N)$.*

Suppose that the sequence $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis C. Let $c \in (-1, 1) \setminus \{0\}$. Then

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

That is, whp, there is cutoff at $\max_\gamma t^{\pm}_0(\gamma, |G/\gamma G|)$. Moreover, the implicit lower bound holds deterministically, ie for all choices of generators.

Remark 2.3.2. In short, the conditions of Hypothesis C say that

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

The regime of smaller k is covered by Approach #1 and of larger k by Approach #2. $\triangle$

Remark. Recall that the lower bound from §2.2 is valid whenever $1 \ll k \lesssim \log|G|$ and $k - d(G) \gg 1$. It thus suffices to consider only the upper bound. $\triangle$

2.3.2 Outline of Proof

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

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

see Propositions 2.1.9 and 2.2.13. In §2.1.6 we used $|G/\gamma G| \le \gamma^{d(G)}$. In §2.2.7 we used

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

The upper bound $|G/\gamma G| \le \gamma^{d(G)}$ is fairly crude. Roughly the idea here is to show, for this interim regime of k around $\sqrt{\log |G|}$, that for all but $e^{o(k)}$ of the γ we can improve it; for the remaining γ, we use the second approach. (Before we considered $|\mathcal{H}|$ different γ, and so required $|\mathcal{H}| = e^{o(k)}$.)

2.3.3 Upper Bound on Mixing

Let G be an Abelian group; set $n := |G|$. One can find a decomposition $\oplus_1^d \mathbb{Z}_{m_j}$ of G such that $d = d(G)$, the minimal size of a generating set, and $m_i \wr m_j$ for all $i \le j$. (This can be proved by induction. Alternatively, write G as a direct sum of p-groups then merge the p-groups appropriately.) For the remainder of this section we fix such a decomposition.

We use the more refined concept of typicality from Approach #2. Let $\varepsilon > 0$ and let $t := t_{*,\varepsilon} := (1+\varepsilon)t_*(k, G)$. We frequently suppress the t and ε dependence in the notation. Let $c := c_\varepsilon > 0$ be the constant from Proposition 2.2.5. Recall some notation:

$$\zeta_\gamma := \tfrac{1}{k}(k-d)\log\gamma, \quad \hat\zeta_\gamma := \zeta_\gamma \wedge 1 \quad \text{and} \quad \delta_\gamma := e^{-c\hat\zeta_\gamma k}.$$

Note that $k - d \gg 1$ and $k \lesssim \sqrt{\log n}$; thus $\hat\zeta_n = 1$; set $\hat\zeta_\infty := 1$. Recall that W is a RW on $\mathbb{Z}$ and we define W_γ by $W \bmod \gamma$; set $W_\infty := W$. We repeat the definition of typicality for convenience.

Definition 2.3.3 (Definition 2.2.8)**.** *Define typical sets for $\gamma \in \mathbb{N}_\infty$ by the following:*

$$\mathcal{W}_{\gamma,\mathrm{glo}} := \big\{w \in \mathbb{Z}_\gamma^k \;\big|\; \mathbb{P}(W_\gamma(t_{*,\varepsilon}) = w) \le \delta_{\gamma,\varepsilon}/|G/\gamma G|\big\} \quad \textit{where} \quad \delta_\gamma := \delta_{\gamma,\varepsilon} := e^{-c_\varepsilon \hat\zeta_\gamma k};$$
$$\mathcal{W}_{\mathrm{loc}} := \big\{w \in \mathbb{Z}^k \;\big|\; |w_i - \mathbb{E}(W_i(t_{*,\varepsilon}))| \le r_* \;\forall\, i \in [k]\big\} \quad \textit{where} \quad r_* := \tfrac{1}{2}n^{1/k}(\log k)^2.$$

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

$$M := \exp\big(\sqrt{\log n/\log\log n}\big); \quad \textit{set} \quad \Gamma := \big\{r \cdot m \;\big|\; r \in [\sqrt{k}],\, m \wr m_L,\, rm \wr n\big\} \setminus \{1\}.$$

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

$$\mathsf{typ} := \big\{W(t_{*,\varepsilon}), W'(t_{*,\varepsilon}) \in \mathcal{W}_{\mathrm{loc}}\big\} \cap \big(\cap_{\gamma\in\Gamma}\big\{W_\gamma(t_{*,\varepsilon}), W'_\gamma(t_{*,\varepsilon}) \in \mathcal{W}_{\gamma,\mathrm{glo}}\big\}\big).$$

Lemma 2.3.4. *We have $\log|\Gamma| \ll k$. In particular, $\delta_\infty|\Gamma| = o(1)$.*

Proof. We have $|\Gamma| \le \sqrt{k}\,\operatorname{div} m_L$ where $\operatorname{div} m$ is the number of divisors of $m \in \mathbb{N}$. We have

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

by [40, §18.1]. Also $\log\sqrt{k} \asymp \log k \ll k$. Thus $\log|\Gamma| \ll k$. Recall that $\log(1/\delta_\infty) \asymp k$. □

We use a union bound over $\gamma \in \Gamma$, which we then bound via Lemma 2.3.4.

Lemma 2.3.5 (Lemma 2.2.9)**.** *We have $\sum_{\gamma\in\Gamma}\delta_\gamma = o(1)$.*

Proposition 2.3.6 (Proposition 2.2.10)**.** *For all $\varepsilon > 0$, we have*

$$\mathbb{P}(\mathsf{typ}) = 1 - o(1).$$

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

Proposition 2.3.7. *Let $\varepsilon > 0$ be fixed and set $t := (1+\varepsilon)t_*(k,G)$. Then*

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

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

Lemma 2.3.8. *We have $0 \le d - L \le \sqrt{\log n/\log\log n} \ll k$. In particular, $L \approx d \approx k$.*

Proof. Since $n = m_1 \cdots m_L$ and $m_1 \le \cdots \le m_L$, if $L < d$ then $M^{d-L} \le m_{L+1}^{d-L} \le n$. Rearranging gives the inequalities. Recall that $k \gtrsim \sqrt{\log n/\log\log\log n}$ and $k \approx d$. This completes the proof. □

We prove Proposition 2.3.7 by separating the sum over γ into two according to Γ.

Proof of Proposition 2.3.7. Observe that $|G/\gamma G|\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \le 1$ when $\gamma = 1$. Also, $\mathfrak{g} \wr n$. Thus

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

where $\overline{\Gamma} := \{\gamma \in [2,n] \mid \gamma \wr n\} \setminus \Gamma$. We analyse these sums with Approach #1 and #2, respectively. Namely we show below that both sums are $o(1)$, when $t := (1+\varepsilon)t_*(k,G)$ with $\varepsilon > 0$ a constant. □

Analysis via Approach #1. Suppose that $\gamma \in \overline{\Gamma}$, ie $\gamma \notin \Gamma \cup \{1\}$. We improve the inequality $|G/\gamma G| \le \gamma^d$ via the following argument. For each $j \in [L]$, we may write

$$\gamma = r_j \cdot \gcd(\gamma, m_j) \quad \text{and} \quad m_j = r_j' \cdot \gcd(\gamma, m_j) \quad \text{where} \quad \gcd(r_j, r_j') = 1.$$

By definition of Γ, if $\gamma = \tilde{r}\cdot m$ for some $m \wr m_j$, then $\tilde{r} \ge \sqrt{k}$. 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 $|G/\gamma G| = \prod_1^d \gcd(\gamma, m_j) \le \gamma^d/k^{L/2}$.

Exactly the same analysis as in the proof of Corollary 2.1.15 then leads us to

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

where δ is any value in $(0,1)$. As in Corollary 2.1.15, the first two terms are $o(1)$ if $k - d \gg 1$ and $k \ll \log n$. For the third term, in Corollary 2.1.15 we needed the complicated condition from Hypothesis A. Now, however, observe that $4^k(\log k)^{2(d+1)}/k^{L/2} \ll 1$ as $L \approx k \approx d$; thus the final term is also $o(1)$. We thus deduce that the sum over $\gamma \in \overline{\Gamma}$ is $o(1)$. □

Analysis via Approach #2. The typicality conditions set out in Definition 2.3.3 imply that

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

cf Lemma 2.1.10. Combining this with Lemma 2.3.5, we deduce that the sum over $\gamma \in \Gamma$ is $o(1)$:

$$\textstyle\sum_{\gamma\in\Gamma} |G/\gamma G|\,\mathbb{P}\big(\mathfrak{g} = \gamma \mid \mathsf{typ}\big) \le \sum_{\gamma\in\Gamma} \delta_\gamma = o(1).$$ □

2.4 Separation Cutoff

In this section we prove Theorem B, namely that there is cutoff in the separation metric for $k \gtrsim \log|G|$. Recall that the separation distance is defined by

$$s(t) := \max_{x,y}\{1 - P_t(x,y)/\pi(y)\},$$

where $P_\cdot$ is the heat kernel (ie transition probabilities) and π the invariant distribution. We write $s^{\pm}_{G_k,N}$ when considering sequences $(k_N, G_N)_{N\in\mathbb{N}}$, analogously to $d^{\pm}_{G_k,N}$.

We now state the main theorem. We require $k \gtrsim \log|G|$, $\log k \ll \log|G|$ and $k - d(G) \asymp k$.

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

$$\liminf_N k_N/\log|G_N| > 0, \quad \limsup_N \log k_N/\log|G_N| < \infty \quad \text{and} \quad \liminf_N \big(k_N - d(G_N)\big)/k_N > 0.$$

Theorem 2.4.1. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ a sequence of finite, Abelian groups; for each $N\in\mathbb{N}$, define $Z_{(N)} := [Z_1, ..., Z_{k_N}]$ by drawing $Z_1, ..., Z_{k_N} \sim^{\text{iid}} \text{Unif}(G_N)$.*

Suppose that the sequence $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis D. Let $c\in(-1,1)\setminus\{0\}$. Then

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

That is, there is cutoff in the separation metric at $t_*(k,G)$ whp. Moreover, the implicit lower bound holds deterministically, ie for all choices of generators.

Remark 2.4.2. *While we only state and prove the result for $k \gtrsim \log|G|$ with $k-d(G)\asymp k$, the argument can be extended to larger regimes in a couple of ways:*

- *$k \ll \log|G|$ with $k-d(G)\asymp k$, provided $\log|G|/k$ diverges sufficiently slowly;*
- *$k\gtrsim\log|G|$ with $1\ll k-d(G)\ll k$ provided $k-d(G)$ diverges sufficiently rapidly.*

These regimes require a little more care; we do not explore the details here.

The proof uses the TV mixing time as a building block.

Proof of Theorem 2.4.1. Since TV is a lower bound on separation (see, eg, [49, Lemma 6.16]), the lower bound follows from the TV result. References for the TV result are as follows. See Theorem 2.2.6, specifically §2.2.6 for the lower bound on mixing, for the regime $k\asymp\log|G|$. For $k\gg\log|G|$, TV cutoff had already been established; see §1.5.2.

We turn to the upper bound. For $y,z\in G$ and $t\geq 0$, write $P_t(y,z) := \mathbb{P}_y(S(t)=z)$ for the transition probability from y to z in time t. Write $n := |G|$. We want to show, for fixed $\xi>0$, that

$$\min_{x\in G} P^{\pm}_t(0,x) \geq \tfrac{1}{n}\big(1-o(1)\big) \quad \text{when} \quad t\geq(1+2\xi)t^{\pm}_*(k,G).$$

Let $\varepsilon>0$ with $\varepsilon=o(1)$ to be specified later. Let $A := [Z_1,...,Z_{(1-\varepsilon)k}]$ be the first $(1-\varepsilon)k$ generators and $B := [Z_{(1-\varepsilon)k+1},...,Z_k]$ be the remaining εk. Since G is Abelian, we may write $P_t = P_{t,A}P_{t,B}$ where in $P_{t,A}$, respectively $P_{t,B}$, we pick each generator of A, respectively B, at rate $1/k$ independently. (In words, we first apply the generators from A and then those from B.)

Let $t' := (1+\xi)t_*((1-\varepsilon)k,G)$; we can then choose $\delta=o(1)$ so that t' is larger than the $\frac{1}{2}\delta^2$-TV mixing time for the rate-1 RW on $G(A)$ for a typical choice of A. To relate this to the rate-1 RW on $G(Z)$, rescale time by $k/|A| = 1/(1-\varepsilon)$: set $t := t'/(1-\varepsilon)$. Since $\varepsilon=o(1)$ and $k-d(G)\asymp k\gtrsim\log|G|$, we have $t\leq(1+2\xi)t_*(k,|G|)$ by Proposition 2.2.2a. (Note that $k-d(G)\asymp k$ implies that $(1-\varepsilon)k-d(G)\asymp k$ when $\varepsilon=o(1)$.) It thus suffices to show that

$$\min_x P_t(0,x) \geq \tfrac{1}{n}\big(1-o(1)\big).$$

Now condition on a typical realisation of A, namely write $\mathcal{A} := \{A \mid t_{\text{mix}}(\frac{1}{2}\delta^2; G(A)) \leq t'\}$ and condition on $A=a$ for a fixed $a\in\mathcal{A}$. We have $\mathbb{P}(A\in\mathcal{A})=1-o(1)$. Given $A=a\in A$, the set

$$D := \big\{z\in G \;\big|\; P_{t,a}(0,z)\geq\tfrac{1}{n}(1-\delta)\big\} \quad \text{satisfies} \quad |D|\geq|G|-\delta|G| = |G|\cdot\big(1-o(1)\big).$$

For the undirected case (ie the RW on G^-_k), by reversibility, conditional on A, we have

$$P^-_t(0,x)\geq P^-_{t,B}(x,D)\cdot\tfrac{1}{n}(1-\delta).$$

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

$$P^+_t(0,x)\geq Q^+_{t,B}(x,D)\cdot\tfrac{1}{n}(1-\delta)$$

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 $x\in G$.

This is a RW on $G^{\pm}(B^{-1})$ run for time t. By considering just the final step of this RW, it suffices to prove the following statement: we can choose ε and η with $\varepsilon,\eta=o(1)$ so that, for all (deterministic) sets $D\subseteq G$ with $|G\setminus D|\leq\delta|G|$ and all $x\in G$ uniformly, we have

$$\mathbb{P}\big(Q_B(x,D)\leq 1-\eta\big)=o(1/|G|) \quad \text{where} \quad Q_B(y,z) := |B_{\pm}|^{-1}\textstyle\sum_{b\in B_{\pm}}\mathbf{1}(y+b^{-1}=z)$$

for $y, z \in G$ where $B_+ := B$ and $B_- := B \cup B^{-1}$ (as multisets). Indeed, this failure probability allows us to perform a union bound to say (conditional on $A = a$) that

$$\mathbb{P}\big(\min_x Q_{t,B}(x, D) \le 1 - 2\eta \;\big|\; A = a\big) = o(1),$$

where the randomness is over the generators B, provided η decays sufficiently slowly (taking into account the uniform $o(1)$ probability that the walk makes no steps). For $A \in \mathcal{A}$ we have the desired lower bound on $\min_x P_t(0, x)$. Finally we average over A and use $\mathbb{P}(A \in \mathcal{A}) = 1 - o(1)$ to show that $\min_x P_t(0, x) \ge \frac{1}{n}(1 - o(1))$ whp. It remains to prove the claim given above.

Fix an arbitrary $x \in G$. We desire at least a proportion $1-\eta$ of the generators in B to connect x to D. The generators are chosen independently, and each connect with probability $|D|/|G| \ge 1-\delta$. Since there are εk generators, it thus suffices to choose η with $\eta = o(1)$ so that

$$\mathbb{P}\big(\mathrm{Bin}(\varepsilon k, 1 - \delta) \le \varepsilon k(1 - \eta)\big) = o(1).$$

Direct calculation, using standard inequalities, gives

$$\mathbb{P}\big(\mathrm{Bin}(\varepsilon k, 1 - \delta) \le \varepsilon k(1 - \eta)\big) = \mathbb{P}\big(\mathrm{Bin}(\varepsilon k, \delta) \ge \eta\varepsilon k\big) \le \tbinom{\varepsilon k}{\eta\varepsilon k}\delta^{\eta\varepsilon k} \le (\delta e/\eta)^{\eta\varepsilon k}.$$

Given $\delta = o(1)$, choose ε and η to be decaying sufficiently slowly so that so that $(\delta e/\eta)^{\eta\varepsilon} = o(1)$. Since $k \gtrsim \log n$, with this choice of ε and η, we have

$$\mathbb{P}\big(Q_B(x, D) \le 1 - \eta\big) = \mathbb{P}\big(\mathrm{Bin}(\varepsilon k, \delta) \ge \eta\varepsilon k\big) \le (\delta e/\eta)^{\eta\varepsilon k} = o(1/|G|).$$

This bound is independent of x, and hence holds for all $x \in G$ uniformly, completing the proof. □

2.5 Mixing Time Comparison for Nilpotent Groups

In this section we compare the mixing time of a general nilpotent group G with a 'corresponding' Abelian group $\overline{G}$: we show that $t_{\mathrm{mix}}(G_k)/t_{\mathrm{mix}}(\overline{G}_k) \le 1 + o(1)$ whp.

2.5.1 Precise Statement

We compare the mixing time for G with that for $\overline{G}$. Specifically, we prove Theorem F, which we recall here for the reader's convenience.

Theorem 2.5.1. *Let G be a nilpotent group. Set $\overline{G} := \oplus_1^L (G_{\ell-1}/G_\ell)$ 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 $k - d(\overline{G}) \gg 1$. Let $\varepsilon > 0$ and let $t \ge (1+\varepsilon)t_*(k, \overline{G})$. Then $d_{G_k}(t) = o(1)$ whp.*

Remark. An upper bound valid for all groups has already been established in the regime $k \gg \log|G|$ at $T(k, |G|) \asymp t_*(k, \overline{G})$; recall Remark A.1. Thus we need only consider $1 \ll k \lesssim \log|G|$. △

2.5.2 Outline of Proof

Let L be the minimal integer such that G_L is the trivial group. 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 representatives for $Q_\ell = G_{\ell-1}/G_\ell$.

In order 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 2.5.2. Then $Z_{i,\ell}G_\ell \sim \mathrm{Unif}(Q_\ell)$ independently for each i and ℓ; see Corollary 2.5.3.

Suppose that M steps are taken; let $\sigma : [M] \to [k]$ 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 choice σ.

Since these are RWs on Abelian groups, the ordering in σ will 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

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

where $\mathfrak{g} := \gcd(W_1 - W'_1, ..., W_k - W'_k, n)$; see Proposition 2.5.5 and Corollary 2.5.8. Via analysing $|\overline{G}/\mathfrak{g}\overline{G}|$, we showed in §2.1–§2.3 that the RW on $\overline{G}_k$ is mixed whp shortly after $t_*(k, \overline{G})$; see specifically Lemma 2.1.11. From this and the inequality above, we are able to deduce that the RW on G_k is mixed whp shortly after the same time.

2.5.3 Reduction to Abelian-Type Calculations

Let L be the minimal integer such that G_L is the trivial group. 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 representatives for $Q_\ell = G_{\ell-1}/G_\ell$, ie a set R_ℓ with $|R_\ell| = |Q_\ell|$ and $\{rG_\ell\}_{r\in R_\ell} = G_{\ell-1}/G_\ell = Q_\ell$.

We want to sample the uniform generators by using uniform random variables on each of the quotients. In this way, projecting to one of the quotients, we get a RW on this quotient. The following two proofs are deferred to Lemma 6.6.3 and Corollary 6.6.4, respectively.

Lemma 2.5.2. *For each $\ell \in [L]$, let $Y_\ell \sim \mathrm{Unif}(R_\ell)$ independently. Then $Y := Y_1 \cdots Y_L \sim \mathrm{Unif}(G)$.*

Corollary 2.5.3. *For each $(i, \ell) \in [k] \times [L]$, sample $Z_{i,\ell} \sim \mathrm{Unif}(R_\ell)$ independently and set $Z_i := Z_{i,1} \cdots Z_{i,L}$. Then $Z_1, ..., Z_L \sim^{\mathrm{iid}} \mathrm{Unif}(G)$. Further, $Z_{i,\ell}G_\ell \sim \mathrm{Unif}(Q_\ell)$ independently for each (i, ℓ).*

For the remainder of the section, assume that Z is drawn in this way. The next main result (Proposition 2.5.5) is the key element of the proof of Theorem 2.5.1. Informally, it reduces the problem to a collection of Abelian calculations, the like of which were handled when we established cutoff when the underlying group was Abelian. We first need a preliminary 'worst-case' lemma.

As is standard, we write 0 for the identity of an Abelian group.

Lemma 2.5.4. *Let H be an Abelian group. Let $Z_1, ..., Z_k \sim^{\mathrm{iid}} \mathrm{Unif}(H)$. Let $v \in \mathbb{Z}^k$. Then*

$$\max_{h\in H} \mathbb{P}\big(v \cdot Z = h\big) = \mathbb{P}\big(v \cdot Z = 0\big).$$

Proof. 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

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

We now prove the decomposition theorem. It crucially uses the nilpotency of the group.

Proposition 2.5.5. *Let $M, M' \in \mathbb{N}$. Let $\sigma : [M] \to [k]$ and $\sigma' : [M'] \to [k]$. Let $\eta \in \{\pm 1\}^M$ and $\eta' \in \{\pm 1\}^{M'}$. For $\ell \in [L]$, set*

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

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

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

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

$$\mathbb{P}\big(S = S'\big) = \textstyle\prod_1^L \mathbb{P}\big(\mathcal{E}_\ell \mid \mathcal{E}_{\ell-1}\big).$$

For all $g \in G$ and $h \in G_{\ell-1}$, we have $[g, h] \in G_\ell$ and $hg = gh[h^{-1}, g^{-1}] = gh[g, h]^{-1}$. We can hence write $S'S^{-1}$ in the following way:

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

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 is independent of N_ℓ.

We claim the following:

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

To prove this, first make the following 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

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

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

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

Now, $G_{\ell-1}/G_\ell$ 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 2.5.4. Applying said lemma, we deduce that

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

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

2.5.4 Evaluation of Abelian-Type Calculations

When establishing cutoff for RWs on Abelian groups, we had to bound a very similar expression to those in the product of Proposition 2.5.5. In particular, since the Q_ℓ are Abelian groups, it does not matter in which order the generators are applied. So instead of considering the exact sequence $\sigma : [M] \to [k]$, it suffices to consider W where $W_i := \sum_{m=1}^{M} \mathbf{1}(\sigma(m) = i)$ for each $i \in [k]$.

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

$$\mathfrak{g}_{(w,w')} := \gcd\big(w_1 - w'_1,\ w_2 - w'_2,\ \ldots,\ w_k - w'_k,\ |G|\big).$$

We use this to evaluate the right-hand side of Proposition 2.5.5, culminating in Corollary 2.5.8.

Lemma 2.5.6. *Let $\ell \in [L]$. For all $w, w' \in \mathbb{Z}^k$, we have*

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

Proof. Corollary 2.5.3 says that each $Z_{i,\ell}G_\ell$ is an independent $\mathrm{Unif}(Q_\ell)$. Lemma 6.6.1 in the supplementary material 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 leads us to a bound on $\mathbb{P}_{(w,w')}(S = S')$ in terms of a product of $|Q_\ell|/|\gamma Q_\ell|$ over $\ell \in [L]$, for some γ which is a suitable gcd. The following lemma controls this product.

Lemma 2.5.7. *For all $\gamma \in \mathbb{N}$, we have $\prod_{\ell=1}^{L} |\gamma Q_\ell| = |\gamma \overline{G}|$.*

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 $\overline{G}$ 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 2.5.5 and Lemmas 2.5.6 and 2.5.7 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 2.5.8. *For all $w, w' \in \mathbb{Z}^k$, we have*

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

Proof. Note that $|Q_\ell|$ divides $|G|$, and so $\gcd(v_1, ..., v_k, |Q_\ell|) \le \gcd(v_1, ..., v_k, |G|)$ for all $v \in \mathbb{Z}^k$. Also, for any Abelian subgroup H of G, if $\alpha \wr |H|$ and $\alpha \wr \beta$, then $\alpha H \le \beta H$. Combined with Proposition 2.5.5 and Lemma 2.5.6, this proves the inequality. The first equality follows immediately from Lemma 2.5.7. The second equality follows from Lagrange's theorem. □

Observe that the right-hand side of this corollary depends only on the Abelian group $\overline{G}$. By applying the results used for Abelian groups, we can prove Theorem 2.5.1; we explain this now. Here, as there, we use a modified L_2 calculation; see Lemma 2.1.6.

Lemma 2.5.9 (Lemma 2.1.6). *For all $t \ge 0$ and all $\mathcal{W} \subseteq \mathbb{Z}^k$, the following inequalities hold:*

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

Proof of Theorem 2.5.1. Let $\mathcal{W} \subseteq \mathbb{Z}^k$ be arbitrary for the moment. Set

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

Abbreviate $\mathfrak{g} := \mathfrak{g}_{(W,W')}$. Applying now Corollary 2.5.8, we obtain

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

This latter expression is purely a statistics of the Abelian group $\overline{G}$. 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 §2.1.6 we upper bounded $|\overline{G}/\gamma\overline{G}| \le \gamma^{d(\overline{G})}$; we then used unimodality to show that $\mathbb{P}(\gamma \wr W_i \mid W_i \neq 0) \le 1/\gamma$, and convert this into $\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \le (1/\gamma + \mathbb{P}(W_1 = 0 \mid \mathsf{typ}))^k$.
- In §2.2.7 we analysed (W, W') taken modulo γ, for each γ; we then used entropic considerations to bound $\mathbb{P}(\mathfrak{g} = \gamma \mid \mathsf{typ}) \ll |\overline{G}/\gamma\overline{G}|$ in a quantitative sense.
- In §2.3.3 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 $(k, \overline{G})$; see Hypotheses A to C. At least one of these is satisfied if $1 \ll k \lesssim \log|G|$ and $k - d(\overline{G}) \gg 1$; see Remarks 2.1.5, 2.2.7 and 2.3.2.

We need to choose the set $\mathcal{W}$; see Definitions 2.1.7 and 2.2.8 for the respective definitions. (In those definitions, replace G with $\overline{G}$.) See Propositions 2.1.9, 2.2.13 and 2.3.7 specifically for the results bounding this sum. The conclusion of these results is that

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

Combined with the modified L_2 calculation of Lemma 2.5.9 this completes the proof. □

2.6 Concluding Remarks and Open Questions

§2.6.1 We discuss some statistics in the regime where k is a fixed constant.

§2.6.2 We give a very short proof, which is a small variant on Roichman's argument [69, Theorem 2], establishing an upper bound on mixing, for arbitrary groups and any $k \gg \log|G|$.

§2.6.3 To conclude, we discuss some questions which remain open and gives some conjectures.

Throughout this section, we only sketch details.

2.6.1 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 [27], using their concept of *moderate growth*. 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 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 $t_{\mathrm{rel}} := t_{\mathrm{rel}}(G(Z))$ for the relaxation time (ie inverse of the spectral gap) and $t_{\mathrm{mix}} := t_{\mathrm{mix}}(\varepsilon; G(Z))$ for the (TV) ε-mixing time, for $\varepsilon \in (0,1)$. When considering sequences $(G_N(Z_{(N)}))_{N\in\mathbb{N}}$, add an N-sub/superscript.

We phrase the result of Diaconis and Saloff-Coste [27] in our language.

Theorem 2.6.1 (cf [27, 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 $t^N_{\mathrm{mix}}/k_N \lesssim \Delta_N^2 \lesssim t^N_{\mathrm{rel}} \lesssim t^N_{\mathrm{mix}}$ 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 [27], including the definition of moderate growth, leading to this conclusion in §5.4.

2.6.2 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 [69, Theorem 2], but it applies in both the undirected and directed cases. (Roichman [69, Theorem 1] deals with the directed case, but requires additional matrix algebra machinery.)

The proof proceeds as follows. Assume that $k \gg \log|G|$ and $\log k \ll \log|G|$; let $\varepsilon > 0$ and let $t := (1+\varepsilon)\log|G|/\log(k/\log|G|)$. Note that $1 \ll k \ll t$. Choose some $\omega \gg 1$, diverging arbitrarily slowly; set $t_\pm := \lfloor t(1 \pm \omega/\sqrt{t})\rfloor$ and $L := \omega\lfloor t^2/k\rfloor$. Whp the number of generators picked at most once is at least $k - L$; whp of these the number picked exactly once lies in $[t_-, t_+]$. 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, eg, Lemma 2.1.6) meaning that we need to control

$$|G|\,\mathbb{P}\big(S = S' \mid W = W',\, \mathsf{typ}\big) - 1.$$

Let $\mathcal{E}$ be the event that some generator is used once in W and not at all in W' or vice versa, ie

$$\mathcal{E} := \bigcup_{i\in[k]} \big(\{|W_i| = 1,\ |W'_i| = 0\} \cup \{|W'_i| = 1,\ |W_i| = 0\}\big).$$

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 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

$$\mathbb{P}\big(\mathcal{E} \mid \mathsf{typ}\big) \le \min_{a\in[t_-,t_+], b\le L} 1/\tbinom{k-b}{a-b} = 1/\tbinom{k-L}{t_- - L}.$$

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 whp after t steps.

2.6.3 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 [65] proposed a simple spectral criterion for a sequence $(X^N)_{N\in\mathbb{N}}$ of reversible Markov chains, known as the *product condition*:

$$\text{cutoff} \quad \text{is equivalent to} \quad t_{\text{mix}}(X^N)\gamma_{\text{gap}}(X^N) \to \infty \text{ as } N \to \infty.$$

It is well-known that the product condition is a necessary condition for cutoff; see, eg, [49, Proposition 18.4]. It is relatively easy to artificially create counter-examples, but these are not 'natural'; see, eg, [49, §18] where constructions due to Aldous and due to Pak are described. The product condition is widely believed to be sufficient for "most" 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 $\operatorname{step} G$), ie 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}, respectively γ^N_{gap}, for the mixing time, respectively spectral gap, of the SRW on $G_N(Z_{(N)})$.*
>
> *Suppose that $\limsup_{N\to\infty} \operatorname{step} G_N < \infty$ and that the product condition holds, ie $t^N_{\text{mix}}\gamma^N_{gap} \to \infty$ 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 [32] and, more generally, for RWs on trees by Basu, Hermon and Peres [5]. 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 [52].

2: An Explicit Choice of Generators

We have shown that if one chooses the generators Z uniformly, then one obtains cutoff whp, at a time which does not depend on Z. 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 $t_*(k,G)$ from Definition 2.2.1. A question raised to us by Diaconis [25] is to find *explicit* sets of generators for which cutoff occurs; see also [23, Chapter 4G, Question 2].

> **Open Problem 2.** *Let G be an Abelian group and $1 \ll k \lesssim \log|G|$. Find an explicit choice of generators Z so that the RW on $G(Z)$ exhibits cutoff. Further, find generators so that the cutoff time is $t_*(k,G)$.*

For the cyclic group $\mathbb{Z}_p$ with p prime, Hough [46, Theorem 1.11] shows that the choice $Z := [0, \pm 1, \pm 2, ..., \pm 2^{\lceil \log_2 p \rceil - 1}]$, 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.

3 Cutoff and Geometry for Random Walks on Heisenberg Groups

Abstract for Chapter 3

We establish cutoff whp for the RW on $H := H_{p,d}$ with p prime and $d \geq 3$, provided d does not diverge too quickly, for all $1 \ll \log k \ll \log |H|$. When k does not grow super-polylogarithmically in the size of the group, this is the first example of cutoff for the RW on the random Cayley graph of a non-Abelian group. When $k \gtrsim \log |H^{\mathrm{ab}}|$, where $H^{\mathrm{ab}} \cong \mathbb{Z}_p^{d-1}$ is the Abelianisation, we can remove the primality assumption on p.

The cutoff time is described (abstractly) in terms of the entropy of RW on $\mathbb{Z}^k$. Unlike for Abelian groups, this cutoff time does not depend, up to subleading order terms, only on k and $|H| = p^{d(d-1)/2}$; rather, one needs to know $|H^{\mathrm{ab}}|$ too. In fact, for $k \leq (\log |H^{\mathrm{ab}}|)^{1+2/(d-2)}$, the cutoff time is asymptotically equivalent to the cutoff time for the RW on the random Cayley graph of the Abelian group H^{ab}.

We also study typical distance for the random Cayley graph of H. We show that all but $o(|H|)$ of the elements of H lie at a graph distance $M \pm o(M)$ from the identity whp, where M is the minimal radius of a ball in $\mathbb{Z}^k$ of cardinality at least $|H^{\mathrm{ab}}| = p^{d-1}$.

When $k \gtrsim \log |H|$ and $d \asymp 1$, we show that the diameter of the random Cayley graph is asymptotically equivalent to the typical distance value M whp.

Table of Contents for Chapter 3

3.1 Cutoff for Random Walk

In this section, we consider mixing for the random walk on the random directed Cayley graph of the Heisenberg group $H_{p,d}$. We take p prime and "$\equiv$" means "equivalent modulo p".

Recall that we denote by $H_{p,d}^{\mathrm{com}} = [H_{p,d}, H_{p,d}]$ the *commutator* and $H_{p,d}^{\mathrm{ab}} = H_{p,d}/[H_{p,d}, H_{p,d}]$ the *Abelianisation* (noting that the latter is an Abelian group). It is straightforward to see that $H_{p,d}^{\mathrm{ab}} \cong \mathbb{Z}_p^{d-1}$ (corresponding to the super-diagonal terms). Set $n := |H_{p,d}| = p^{d(d-1)/2}$.

Informally, we show that there is competition between mixing of the Abelianisation and of the commutator. Which part governs the mixing depends on the regime of k: for $k \ll (\log n)^{1+2/(d-2)}$, it is the Abelianisation, meaning that the overall mixing time is the same as that for $\mathbb{Z}_p^{d-1}$; for $k \gg (\log n)^{1+2/(d-2)}$, it is the non-Abelian part, and the overall mixing time is given by the standard diameter-based lower bound of $\log_k |H_{p,d}|$; see Definition 3.1.4 and Theorem 3.1.6.

Throughout this section, we use the following notation:

$$k = \big(\log |H_{p,d}^{\mathrm{ab}}|\big)^{\rho}, \quad \tfrac{1}{2}d = \big(\log |H_{p,d}|\big)^{\nu} \text{ and } n = |H_{p,d}| = p^{d(d-1)/2};$$

the choice of ν is so that $\log |H_{p,d}| = (\log |H_{p,d}^{\mathrm{ab}}|)^{1+\nu}$, so also $k = (\log |H_{p,d}|)^{\rho/(1+\nu)}$.

3.1.1 Entropic Times: Definition and Concentration

In this section, we define the notion of *entropic times*. For $t \geq 0$, write μ_t, respectively ν_t, for the law of $W(t)$, respectively $W_1(t)$; so $\mu_t = \nu_t^{\otimes k}$. Also, for each $i = 1, ..., k$, define

$$Q_i(t) := -\log \nu_{t/k}\big(W_i(t)\big), \quad \text{and set} \quad Q(t) := -\log \mu_t\big(W(t)\big) = \textstyle\sum_1^k Q_i(t).$$

Definition 3.1.1. *For $k, N \in \mathbb{N}$, define the entropic time $t_0(k, N)$ so that $\mathbb{E}(Q_1(t_0(k, N))) = \log N/k$. We apply this with $N := p^{d-1} = n^{2/d}$; abbreviate $t_0 := t_0(k, p^{d-1}) = t_0(k, n^{2/d})$.*

Direct calculation, with the SRW and Poisson laws gives the following relations. We sketch the argument in §3.1.2; full, rigorous details are given in Proposition 6.1.2 and §6.1.5. Recall that the $+$-superscript corresponds to the DRW and the $-$-superscript to the SRW.

Proposition 3.1.2. *Assume that $1 \ll \log k \ll \log N$. Write $\kappa := k/\log N$. For all $\lambda \in (0, \infty)$, the following relations hold, for some continuous, decreasing bijection $f^{\pm} : (0, \infty) \to (0, \infty)$:*

$$t_0^{\pm}(k, N) \asymp k \cdot N^{2/k}/(2\pi e) \quad \textit{when} \quad k \ll \log N; \tag{3.1.1a}$$

$$t_0^{\pm}(k, N) \asymp k \cdot f^{\pm}(\lambda) \quad \textit{when} \quad k \asymp \lambda \log N; \tag{3.1.1b}$$

$$t_0^{\pm}(k, N) \asymp k \cdot 1/(\kappa \log \kappa) \quad \textit{when} \quad k \gg \log N. \tag{3.1.1c}$$

By a standard argument considering appropriate subsequences, to cover the general case $k \asymp \log N$, it suffices to assume that $k/\log N$ actually converges, say to $\lambda \in (0, \infty)$.

Since $Q = \sum_1^k Q_i$ is a sum of k iid random variables, $Q(t_0)$ concentrates around $\log N$. One can show that if the time is multiplied by a factor $1 + \xi$ for any constant $\xi > 0$ then the entropy increases by a significant amount; similarly, if $\xi < 0$ then the entropy decreases by a significant amount. Further, the change is by an additive term of larger order than the standard deviation $\sqrt{\mathbb{V}\mathrm{ar}(Q(t_0))}$. Thus $Q((1 + \xi)t_0)$ concentrates around this new value.

The following proposition quantifies this change in entropy and this concentration. For rigorous details, see Definition 6.1.1 and Propositions 6.1.2 and 6.1.3 in the supplementary material.

Proposition 3.1.3. *Assume that k satisfies $1 \ll \log k \ll \log N$. Then $\mathbb{V}\mathrm{ar}(Q(t_0)) \gg 1$, and further, for $\xi \in \mathbb{R} \setminus \{0\}$, writing $v := \mathbb{V}\mathrm{ar}(Q_1(t_0))$ and $\omega := \mathbb{V}\mathrm{ar}(Q(t_0))^{1/4} = (vk)^{1/4}$, we have*

$$\mathbb{P}\big(Q((1 + \xi)t_0) \geq \log N \pm \omega\big) \to \mathbf{1}(\xi > 0). \tag{3.1.2}$$

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

3.1.2 Entropic Times: Sketch Evaluation

In this subsection, we sketch details towards a proof of Proposition 3.1.2. The full, rigorous details can be found in Proposition 6.1.2 and §6.1.5, where all the approximations below are justified.

(3.1.1a). When $k \ll \log N$, the target entropy for the rate-$1/k$ RW W_1 on $\mathbb{Z}$ is $\log N/k \gg 1$. Hence $t_0(k,N)/k \gg 1$. When a rate-1 RW on $\mathbb{Z}$ is run for time $s \gg 1$, it approximates a normal distribution with variance s; the SRW has mean 0 and the DRW mean s. Direct calculation shows that the entropy of such a normal distribution is precisely $\frac{1}{2}\log(2\pi e s)$. Assuming that we can approximate the entropy of the RW by that of the normal distribution sufficiently well (which is precisely what we show in Proposition 6.1.9), the claim now follows.

(3.1.1b). When $k \asymp \log N$, the target entropy is $\log N/k \asymp 1$. So we consider a rate-1 RW on $\mathbb{Z}$ run for time order 1. The claim now follows with $f(\lambda) \coloneqq H^{-1}(1/\lambda)$, where $H(s)$ is the entropy of the rate-1 RW on $\mathbb{Z}$ run for time s. See Proposition 6.1.12 for a more formal treatment.

(3.1.1c). When $k \gg \log N$, the target entropy is $\log N/k \ll 1$. The rate-1 RW on $\mathbb{Z}$ run for time $s \ll 1$ is approximated by a Bernoulli distribution with success probability $1-e^{-s} \approx s$ (along with a uniformly chosen sign for the SRW). Such a (possibly signed) Bernoulli distribution has entropy $s\log(1/s)+\mathcal{O}(s)$. Again assuming that this approximation can be suitably justified (which is precisely what we show in Proposition 6.1.13), the claim now follows.

3.1.3 Precise Statement and Remarks

In this section, we state the more refined version of Theorem C.

Definition 3.1.4. *Define* $t_{\mathrm{diam}}(k,n) \coloneqq \log_k n$. *Define*

$$t_*^{\pm}(k,p,d) \coloneqq \max\bigl\{t_0^{\pm}(k,|H_{p,d}^{\mathrm{ab}}|),\ t_{\mathrm{diam}}(k,|H_{p,d}|)\bigr\}.$$

Abbreviate $t_{\mathrm{diam}} \coloneqq t_{\mathrm{diam}}(k,p^{d(d-1)/2})$ *and* $t_0^{\pm} \coloneqq t_0(k,p,d)$.

The following proposition determines t_* up to a $1\pm o(1)$ factor; it follows easily from Proposition 3.1.2 and Definition 3.1.4, using $N \coloneqq |H_{p,d}^{\mathrm{ab}}| = |H_{p,d}|^{2/d}$.

Proposition 3.1.5. *We have the following approximation to* t_**:*

$$t_* \eqsim \begin{cases} k\cdot\frac{1}{2\pi e}|H_{p,d}^{\mathrm{ab}}|^{2/k} & \text{when} \quad 1 \ll k \ll \log|H_{p,d}^{\mathrm{ab}}|; & \text{(3.1.3a)}\\ k\cdot f(\lambda) & \text{when} \quad k \eqsim \lambda\log|H_{p,d}^{\mathrm{ab}}|; & \text{(3.1.3b)}\\ \frac{\rho}{\rho-1}\frac{2}{d}\log_k|H_{p,d}| & \text{when} \quad \log|H_{p,d}^{\mathrm{ab}}| \ll k \le (\log|H_{p,d}^{\mathrm{ab}}|)^{1+2/(d-2)}; & \text{(3.1.3c)}\\ \log_k|H_{p,d}| & \text{when} \quad (\log|H_{p,d}^{\mathrm{ab}}|)^{1+2/(d-2)} \le k,\ \log k \ll \log|H_{p,d}|; & \text{(3.1.3d)} \end{cases}$$

here f *is the function from Proposition 3.1.2. (The third regime is empty if* $(\log|H_{p,d}^{\mathrm{ab}}|)^{1/d} \asymp 1$, *ie* $d \gtrsim \log\log p$*; in this case, the lower bound in the fourth regime becomes* $k \gg \log|H_{p,d}^{\mathrm{ab}}|$*.)*

There are some simple conditions that the parameters must satisfy for our proof to be valid. The conditions will be assumed throughout the remainder of the section, often not explicitly stated.

Hypothesis E. *The triple* (k,p,d) *satisfies Hypothesis E if the following conditions hold:*

- $1 \ll \log k \ll \log|H_{p,d}|$;
- *if* $k \ll \log|H_{p,d}^{\mathrm{ab}}|$, *then* $d^3 \ll k$ *and* $k \le \frac{3}{2}\log|H_{p,d}^{\mathrm{ab}}|/\log d$ *(eg* $k \le d\log p/\log d$*);*
- *if* $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$, *then* $\log d \ll \log\log p$ *(equivalently* $\log d \ll \log\log|H_{p,d}|$*).*

(Recall that implicitly we consider sequences $(k_N,p_N,d_N)_{N\in\mathbb{N}}$*.)*

Remark. Hypothesis E holds when $d \asymp 1$ and $1 \ll \log k \ll \log|H_{p,d}^{\mathrm{ab}}|$. As noted in the introduction, there is no cutoff for k outside this regime. Thus our conditions are optimal when $d \asymp 1$. △

We now state the main result of this section; it is in essence a restatement of Theorem C.

Theorem 3.1.6 (Cutoff). *Let (k,p,d) be integers with p prime and $d \geq 3$, satisfying Hypothesis E. Then the random walk on $H_k^{\pm}$ exhibits cutoff at time $t_*^{\pm}(k,p,d)$, given in Definition 3.1.4, whp over Z. Moreover, the implicit lower bound on mixing holds deterministically for all Z.*

Remark. For ease of presentation, consider for the moment d independent of n. Define

$$T(\rho, N) \coloneqq \tfrac{\rho}{\rho-1} \log_k N = \tfrac{\rho}{\rho-1} t_{\mathrm{diam}}(k, N).$$

Simple algebraic manipulations give $T(\rho,N) \approx t_0^{\pm}((\log N)^{\rho}, N)$ when $\rho > 1$ is bounded away from 1. This is the universal mixing time upper bound for a group of size N from §1.5.2 when $k = (\log N)^{\rho}$ with $\rho > 1$. (It is tight for Abelian groups.) Recall that the Abelianisation has size $|H_{p,d}^{\mathrm{ab}}| = p^{d-1}$.

Consider $k = (\log |H_{p,d}^{\mathrm{ab}}|)^{\rho}$ with $\rho > 1$. Hence the walk projected to the Abelianisation has cutoff at $t_0(k, |H_{p,d}^{\mathrm{ab}}|)$. Our proof shows that the random walk on the whole group exhibits cutoff, with time given by the maximum of this and the diameter lower bound, $t_{\mathrm{diam}}(k, |H_{p,d}|)$.

This heuristic is only valid when $\rho > 1$. From Theorem 2.1.4, one sees that the walk projected to the Abelianisation has cutoff at $t_0(k, |H_{p,d}^{\mathrm{ab}}|)$ for $\rho \leq 1$ too; in this regime, $t_0(k, |H_{p,d}^{\mathrm{ab}}|) \gg t_{\mathrm{diam}}(k, |H_{p,d}|)$. We show that the mixing time is upper bounded by $t_0(k, |H_{p,d}^{\mathrm{ab}}|)$ when $\rho \leq 1$. △

The fact that the mixing time is a maximum of two quantities suggests some sort of 'competition' between the Abelianisation and the rest of the group; this leads to a 'phase transition' in the mixing time, which has an interesting consequence for the Aldous–Diaconis conjecture.

Remark. Consider for the moment $\rho \coloneqq 1 + \frac{1}{d}$. If $d \to \infty$ sufficiently slowly, then

$$\log |H_{p,d}^{\mathrm{ab}}| \ll (\log |H_{p,d}^{\mathrm{ab}}|)^{1+1/d} \ll (\log |H_{p,d}^{\mathrm{ab}}|)^{1+2/(d-2)}.$$

According to the Aldous–Diaconis conjecture, there should be cutoff at $T(\rho, n)$; however, Proposition 3.1.5 shows that the mixing time t_* satisfies $t_* \approx T(\rho, |H_{p,d}^{\mathrm{ab}}|) = \frac{2}{d} T(\rho, |H_{p,d}|)$ in this regime (provided d does not grow too quickly). Hence the Aldous–Diaconis conjecture is off by a factor of $\frac{2}{d}$, and so does not even capture the correct order of the mixing (since we allow $d \to \infty$).

Recall that the conjecture has been verified for Abelian groups, in the entire $k \gg \log n$ regime. These Heisenberg groups give a counter-example once one allows non-Abelian groups. △

Recall that cutoff is already established (for all groups) when k grows super-polylogarithmically in n, ie $\log k \gg \log\log |H_{p,d}|$. Below assume that $\log k \lesssim \log\log |H_{p,d}|$, ie $k = (\log |H_{p,d}|)^{\mathcal{O}(1)}$.

3.1.4 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; rather, this section merely introduces ideas. Recall the definitions from the previous sections.

For ease of notation, we suppress the p and d dependence from $H_{p,d}$, writing just H. Similarly we write $H^{\mathrm{ab}} \coloneqq H_{p,d}^{\mathrm{ab}}$ for the Abelianisation and $H^{\mathrm{com}} \coloneqq H_{p,d}^{\mathrm{com}}$ for the commutator.

We start by discussing the lower bound. In §2.1.5 we consider an analogous entropic lower bound but where the underlying group is Abelian. To apply this, we simply project the walk from H to H^{ab}, which is an Abelian group. Projection cannot increase the TV distance.

If Q is sufficiently small, then W, and hence also S, is restricted to a small set. Indeed, $Q \leq \log |H^{\mathrm{ab}}| - \omega$ if and only if $\mu(W) \geq |H^{\mathrm{ab}}|^{-1} e^{\omega}$, and thus if this is the case then $W \in \{w \mid \mu(w) \geq |H^{\mathrm{ab}}|^{-1} e^{\omega}\}$. Write S^{ab} for S projected to the Abelianisation H^{ab}. Since H^{ab} is an Abelian group, $S^{\mathrm{ab}}(t)$ depends only on $W(t)$ (not additionally any $W(t')$ for $t' < t$). It is thus also the case that

$$S^{\mathrm{ab}}(t) \in E \coloneqq \{a \in H^{\mathrm{ab}} \mid \mathbb{P}\big(S^{\mathrm{ab}}(t) = a\big) \geq |H^{\mathrm{ab}}|^{-1} e^{\omega}\}.$$

But clearly $|E| \leq e^{-\omega} |H^{\mathrm{ab}}|$. Choosing the time t slightly smaller than the entropic time t_0 and $\omega \gg 1$ suitably, the event $\{Q(t) \leq \log |H^{\mathrm{ab}}| - \omega\}$ will hold whp. Thus, whp, $S^{\mathrm{ab}}(t)$ is restricted to a set of size $o(|H^{\mathrm{ab}}|)$. It hence cannot be mixed. This heuristic applies for any choice of generators.

Precisely, we show for any ω with $1 \ll \omega \ll \log|H^{\mathrm{ab}}|$, all t and all $Z = [Z_1, ..., Z_k]$, that

$$d_{G(Z)}(t) \geq \mathbb{P}\big(Q(t) \leq \log|H^{\mathrm{ab}}| - \omega\big) - e^{-\omega}.$$

Observe that the probability on the right-hand side is independent of Z. Thus we are naturally interested in the fluctuations of $Q(t)$ for t close to t_0. Using the concentration of Q, ie Proposition 3.1.3 with $\xi < 0$ and $\omega := \mathbb{V}\mathrm{ar}(Q(t_0))^{1/4}$, we deduce the lower bound in Theorem 3.1.6.

We now turn to discussing the upper bound. We use a *modified L_2 calculation*; see Lemma 3.1.8. If S and S' are independent copies, with auxiliary W and W', and $\mathcal{W} \subseteq \mathbb{Z}^k$ is some set, then

$$\mathbb{E}\big(\big\|\mathbb{P}\big(S(t) \in \cdot\big) - \pi_H\big\|_{\mathrm{TV}}\big) \leq \tfrac{1}{2}\sqrt{|H|\,\mathbb{P}\big(S(t) = S'(t) \mid W(t), W'(t) \in \mathcal{W}\big) - 1} + \mathbb{P}\big(W(t) \notin \mathcal{W}\big).$$

We choose $\mathcal{W}$ so that $\mathbb{P}(W(t) \notin \mathcal{W}) = o(1)$, and think of $\mathcal{W}$ as the set of 'typical $W(t)$'. By imposing some mild typicality conditions, we show that $S'(t)S(t)^{-1}$ is uniformly distributed on H when $W(t) \neq W'(t)$ with both $W(t)$ and $W'(t)$ typical. It thus remains to show that

$$|H|\,\mathbb{P}\big(S(t) = S'(t),\ W(t) = W'(t) \,\big|\, W(t), W'(t) \in \mathcal{W}\big) - 1 = o(1/|G|).$$

This is where we analyse the Abelianisation and commutator separately. Drop the t from the notation. Also write $\overline{\mathbb{P}}(\cdot) := \mathbb{P}(\cdot \mid W(t), W'(t) \in \mathcal{W})$.

For the regime in which the mixing time is the entropic time $t_0(k, |H^{\mathrm{ab}}|)$, for the Abelianisation, we add an entropic condition to typicality: if $w \in \mathcal{W}$, then $\mu(w) \leq |H^{\mathrm{ab}}|^{-1}e^{-\omega}$. By definition of the entropic time, like in the lower bound, we have

$$\overline{\mathbb{P}}\big(W = W'\big) \leq e^{-\omega}|H^{\mathrm{ab}}|^{-1} \ll |H^{\mathrm{ab}}|^{-1};$$

see Lemma 3.1.13. If $W = W'$, then $S'S^{-1} \in H^{\mathrm{com}}$. Given $W = W'$, we desire $S'S^{-1}$ to be approximately uniformly distributed on H^{com}, say with modal probability order $|H^{\mathrm{com}}|^{-1}$, as then

$$\overline{\mathbb{P}}\big(S = S',\ W = W'\big) = \overline{\mathbb{P}}\big(S = S' \,\big|\, W = W'\big) \cdot \overline{\mathbb{P}}\big(W = W'\big) \ll |H^{\mathrm{ab}}|^{-1} \cdot |H^{\mathrm{com}}|^{-1} = |H|^{-1},$$

as desired. When the mixing time is $\log_k|H|$, then we perform an analogous analysis, but this time we calculate the entropy shortly after $\log_k|H|$, which is larger than $t_0(k, |H^{\mathrm{ab}}|)$; see Lemmas 3.1.9 and 3.1.13. We can then relax the "approximate uniformity" appropriately.

We now briefly explain how to establish this "approximate uniformity" of $S'S^{-1}$ on H^{com} given $W = W'$. We explain the method for $d = 3$; general d imposes some additional technical hurdles.

We define a combinatorial event $\mathcal{E}$ in terms of W and W', which depends on the order in which the generators are chosen not just on the final counts $W(t)$ and $W'(t)$. We can describe this event in terms of a random walk on a free nilpotent group. Let $\widetilde{H}_k$ be a free nilpotent group of step 2 with k generators. Let $\widetilde{S}$ be the RW on $\widetilde{H}_k$; assign to it auxiliary process W. Let $(\widetilde{S}', W')$ be an independent copy of $(\widetilde{S}, W)$. Let $w \in \mathbb{Z}^k$. Given $W(t) = w = W'(t)$, the combinatorial event $\mathcal{E}$ is exactly the event $\{\widetilde{S}'(t)\widetilde{S}(t)^{-1} \in [\widetilde{H}_k, \widetilde{H}_k]\} \setminus \{\widetilde{S}'(t)\widetilde{S}(t)^{-1} \in [\widetilde{H}_k, [\widetilde{H}_k, \widetilde{H}_k]] \setminus \{\mathsf{id}\}\}$. It is interesting that this condition turns out to be the relevant condition for arguing that $S'(t)S(t)^{-1}$ is roughly a uniformly distributed commutator $H^{\mathrm{com}} = [H, H]$. We plan to investigate this further in the context of general step 2 nilpotent groups in future work.

It remains to control the probability of $\mathcal{E}$; we again use typicality conditions for this. This is the only place in which the method differs according to the regime of k; see Lemma 3.1.15.

When the mixing time is the entropic time, this 'error probability' will be smaller than $1/|H^{\mathrm{com}}|$, meaning that $\overline{\mathbb{P}}(S = S',\ W = W') = o(1/|H|)$, as described above. When the mixing time is $\log_k|H|$, the error is larger, but combined with $\mathbb{P}(W = W')$ gives $o(1/|H|)$; see Lemma 3.1.18.

3.1.5 Lower Bound

The lower bound is relatively straightforward to prove: we project onto the Abelianisation, then use the lower bound for Abelian groups from Chapter 2, specifically §2.1.5.

For ease of notation, we suppress the p and d dependence from $H_{p,d}$, writing just H. Similarly we write $A := H^{\mathrm{ab}}_{p,d}$ for the Abelianisation. Also write $n := |H_{p,d}| = p^{d(d-1)/2}$.

Proof of Lower Bound in Theorem 3.1.6. We assume that Z is given, and suppress it.

For any $\varepsilon > 0$, a lower bound is given by $(1-\varepsilon)\log_k n$: in m steps the support of the random walk is (at most) k^m, and hence the walk cannot be mixed in this many steps; cf [52, Fact 2.1].

Write $\Pi : H \to A$ for the canonical projection. Write $N := |A| = p^{d-1}$. Let $\varepsilon > 0$ and let $t := (1-\varepsilon)t_0(k,N)$. Write

$$\mathcal{E} := \big\{\mu\big(W(t)\big) \geq N^{-1}e^{\omega}\big\} = \big\{Q(t) \leq \log N - \omega\big\},$$

with μ, Q and $\omega \gg 1$ from §3.1.1. By Proposition 3.1.3, we have $\mathbb{P}(\mathcal{E}) = 1 - o(1)$.

For $w \in \mathbb{Z}_+^k$ and $z_1, ..., z_k \in H$, write $z^w := z_1^{w_1} \cdots z_k^{w_k}$. Recall that reordering the terms corresponds to multiplication by a particular element of the commutator. Consider the set

$$E := \big\{x \in A \;\big|\; \exists\, w \in \mathbb{Z}_+^k \text{ st } \mu_t(w) \geq N^{-1}e^{\omega} \text{ and } x = \Pi(Z^w)\big\} \subseteq A.$$

Since we use W to generate S, we have $\mathbb{P}(\Pi(S(t)) \in E \mid \mathcal{E}) = 1$. Every element $x \in E$ satisfies $x = \Pi(w_x \cdot Z)$ for some $w_x \in \mathbb{Z}_+^k$ with $\mu_t(w_x) \geq N^{-1}e^{\omega}$. Hence, for all $x \in E$, we have

$$\mathbb{P}\big(\Pi(S(t)) = x\big) \geq \mathbb{P}\big(W(t) = w_x\big) = \mu_t(w_x) \geq N^{-1}e^{\omega}.$$

Taking the sum over all $x \in E \subseteq A$, we deduce that

$$1 \geq \textstyle\sum_{x\in E} \mathbb{P}\big(\Pi(S(t)) = x\big) \geq |E| \cdot N^{-1}e^{\omega}, \quad \text{and hence} \quad |E|/N \leq e^{-\omega} = o(1).$$

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

$$\big\|\mathbb{P}\big(S(t) \in \cdot \mid Z\big) - \pi_G\big\|_{\mathrm{TV}} \geq \mathbb{P}\big(S(t) \in \Pi^{-1}(E)\big) - \pi_G\big(\Pi^{-1}(E)\big) \geq \mathbb{P}(\mathcal{E}) - \tfrac{1}{N}|E| \geq 1 - o(1). \quad \square$$

Remark 3.1.7. For the entropic lower bound, all that we used was the size of the Abelianisation. The same argument shows, for all finite groups G, all $k \gg 1$ and all multisubsets Z of G of size k, that $\max\{t_0(k, |G^{\mathrm{ab}}|), \log_k |G|\}$ is a lower bound on the mixing time.

The lower bound here can be used to determine the profile of the convergence to equilibrium; this is done in §3.1.9.2, and in §2.1. Another lower bound is proved in §2.2.6; this cannot be used to determine the profile. For many groups these lower bounds will be equivalent, but for some the latter captures the correct mixing time while the former is a constant factor too small. △

3.1.6 Upper Bound Preliminaries

For ease of notation, we suppress the p and d dependence from $H_{p,d}$, writing just H. Similarly we write $A := H_{p,d}^{\mathrm{ab}}$ for the Abelianisation. Also write $n := |H_{p,d}| = p^{d(d-1)/2}$.

We first prove the upper bound for $d = 3$. The majority of the ideas are exposed in this case, while the technical details involved in the general d case somewhat obscure the ideas. Note that the conditions on d from Hypothesis E are always satisfied when $d = 3$ (or, in fact, any fixed d). Similarly, we first analyse the DRW (ie directed graphs); in §3.1.9.1 we then describe the (simple) adaptations to the proof required for the SRW.

Before doing so, we need some preliminary results (for both $d = 3$ and general d). First, we need a concept of 'typicality' for the auxiliary random variable $W(t)$: later in the proof, we define a set $\mathcal{W} \subseteq \mathbb{Z}_+^k$ (dropping the t-dependence from the notation) with the property

$$\mathcal{W} \subseteq \big\{w \in \mathbb{Z}_+^k \;\big|\; \mu_t(w) \leq e^{-h},\ \max_i w_i < \tfrac{1}{2}p\big\}, \tag{3.1.4}$$

where h is roughly the entropy of $W(t)$; see Definition 3.1.11 for the precise definition of h, and also Lemma 3.1.9 for the relation to the entropy. This set will satisfy $\mathbb{P}(W \in \mathcal{W}) = 1 - o(1)$, hence the name 'typical'; see Lemma 3.1.12.

It is often easier to work with L_2, rather than TV, distance, since it has a nice explicit representation; on the other hand, with TV one can condition on 'typical' events. We combine the two with a 'modified L_2 calculation': let S and S' be independent copies (given Z), and let W and W' be their associated auxiliary random variables; write $\mathsf{typ} := \{W, W' \in \mathcal{W}\}$.

Lemma 3.1.8. *Assume that $\mathbb{P}(W((1+\varepsilon)t_*) \notin \mathcal{W}) = o(1)$ for all constants $\varepsilon > 0$. Then the upper bound in Theorem 3.1.6 is established by showing, for all constants $\varepsilon > 0$, that*

$$t \mapsto D(t) := |H|\,\mathbb{P}\big(S(t) = S'(t) \mid \mathsf{typ}\big) - 1 \quad \textit{satisfies} \quad D\big((1+\varepsilon)t_*\big) = o(1).$$

Proof. Using the triangle inequality and then Cauchy-Schwarz inequality, we obtain the following:

$$d_{H_k}(t) = \big\|\mathbb{P}_{H_k}\big(S(t) \in \cdot \mid Z\big) - \pi_H\big\|_{\mathrm{TV}} \le \big\|\mathbb{P}_{H_k}\big(S(t) \in \cdot \mid W(t) \in \mathcal{W}\big) - \pi_H\big\|_{\mathrm{TV}} + \mathbb{P}\big(W(t) \notin \mathcal{W}\big);$$
$$\mathbb{E}\big(2\big\|\mathbb{P}_{H_k}\big(S(t) \in \cdot \mid W(t) \in \mathcal{W}\big) - \pi_H\big\|_{\mathrm{TV}}\big)^2 \le |H|\,\mathbb{P}\big(S(t) = S'(t) \mid \mathsf{typ}\big) - 1 = D(t).$$

Combining these with the assumption $\mathbb{P}(W \notin \mathcal{W}) = o(1)$, and Markov's inequality, gives

$$\big\|\mathbb{P}_{H_k}\big(S(t) \in \cdot\big) - \pi_H\big\|_{\mathrm{TV}} = o(1) \quad \text{whp over } Z. \qquad \square$$

To upper bound $D := D(t)$, we separate into cases according to whether or not $W = W'$:

$$\begin{aligned}\mathbb{P}\big(S = S' \mid \mathsf{typ}\big) &= \mathbb{P}\big(S = S' \mid W = W',\, \mathsf{typ}\big)\mathbb{P}\big(W = W' \mid \mathsf{typ}\big)\\ &\quad + \mathbb{P}\big(S = S' \mid W \neq W',\, \mathsf{typ}\big)\mathbb{P}\big(W \neq W' \mid \mathsf{typ}\big).\end{aligned}$$

Were the underlying group Abelian, $W = W'$ would imply $S = S'$. This is not the case for non-Abelian groups; in fact estimating $\mathbb{P}(S = S' \mid W = W', \mathsf{typ})$ is the main part of the proof.

First we control $\mathbb{P}(W = W' \mid \mathsf{typ})$. To do this, we must estimate the entropy shortly after the proposed mixing time. Recall that $W(\cdot)$ is a RW on $\mathbb{Z}_+^k$, that μ_t is the law of $W(t)$ and that $Q(t) = -\log \mu_t(W(t))$. Denote by

$$h(t) = \mathbb{E}\big(Q(t)\big), \quad \text{the entropy of } W(t).$$

Recall that $t_{\mathrm{diam}} = \log_k |H|$, $\frac{1}{2}d = (\log|A|)^\nu$, $k = (\log|A|)^\rho = (\log|H|)^{\rho/(1+\nu)}$ and $|H| = |A|^{d/2}$.

Lemma 3.1.9. *Let $\xi > 0$. Then, for any $\omega \ll \min\{k, \log|A|\}$, the following lower bounds hold.*

- *For $t \ge (1+\xi)t_0(k, |A|)$, we have $h(t) = \mathbb{E}(Q(t)) \ge \log|A| + 2\omega$.*
- *For $t \ge (1+\xi)t_{\mathrm{diam}}$, if $\rho \ge 1 + 2/(d-2)$, then $h(t) = \mathbb{E}(Q(t)) \ge (1 - \frac{1}{\rho})\log|H| + 2\omega$.*

To prove this lemma, we use the following result, which will be used independently later.

Lemma 3.1.10. *Let t_0 and $t_{2\omega}$ be the entropic times for entropy $\log|A|$ and $\log|A| + 2\omega$, respectively. Then we have $t_{2\omega} \asymp t_0$ if $\omega \ll \min\{k, \log|A|\}$.*

We defer the proof of Lemma 3.1.10 to Chapter 6; see Lemma 6.1.8. We now prove Lemma 3.1.9.

Proof of Lemma 3.1.9. Consider first time $t_0(k, |A|)$. We have $h(t) \ge \log|A|$ by definition of the entropic time. The $+2\omega$ additive term then follows immediately from Lemma 3.1.10.

Consider now the time t_{diam}. Recall from §3.1.2 that the entropy of the rate-1 RW on $\mathbb{Z}$ at time $s \ll 1$ satisfies $H(s) \asymp s\log(1/s)$. Take $s := t_{\mathrm{diam}}/k$. Direct calculation gives $s \ll 1$. Thus

$$\begin{aligned}h(t_{\mathrm{diam}})/\log n &\asymp t_{\mathrm{diam}}\log(k/t_{\mathrm{diam}})/\log n = \tfrac{1}{\log k}\log\big(\tfrac{k\log k}{\log n}\big)\\ &= \tfrac{1}{\rho\log\log|A|}\big((\rho - 1 - \nu)\log\log|A| + \log\log((\log|A|)^\rho)\big) \ge 1 - \tfrac{1+\nu}{\rho}.\end{aligned}$$

For $\xi \in (0,1)$ fixed, $h((1+\xi)t_{\mathrm{diam}}) \asymp (1+\xi)h(t_{\mathrm{diam}})$. The conditions on d gives $1 - 1/\rho \ge 2/d \gg \nu$. Hence the claim is true for all $\xi > 0$ (fixed) provided $k \ll \frac{2}{d}\log n$, ie $k \ll \log|A|$. $\square$

Motivated by this lemma, recalling that $t_* = \max\{t_0, t_{\mathrm{diam}}\}$, we make the following definition.

Definition 3.1.11. *Define h_0 as follows:*

$$h_0 := \begin{cases}\log|A| & \text{when} \quad k \le (\log|A|)^{1+2/(d-2)};\\ (1 - \frac{1}{\rho})\log|H| & \text{when} \quad k \ge (\log|A|)^{1+2/(d-2)}.\end{cases}$$

Fix some ω such that $1 \ll \omega \ll \min\{k, \log|A|\}$, and set $h := h_0 + \omega$.

Not only does the entropy satisfy this lower bound, but the Q random variable, which is defined, in §3.1.1, so that $\mathbb{E}(Q(t)) = h(t)$, concentrates, giving the following result.

Lemma 3.1.12. *Assume that $\omega \ll \min\{k, \log|A|\}$. Let $\varepsilon > 0$ and $t \geq (1+3\varepsilon)\max\{t_0, t_{\mathrm{diam}}\}$. Then*

$$\mathbb{P}\big(Q(t) \geq h\big) = \mathbb{P}\big(\mu_t\big(W(t)\big) \leq e^{-h}\big) = 1 - o(1).$$

Proof. Rearrange the inequality $\mu \leq e^{-h}$ into $Q := -\log\mu \geq h$, use Lemma 3.1.9, the definition of h and h_0 from Definition 3.1.11 and apply the concentration result Proposition 3.1.3. □

We now control $\mathbb{P}(W = W' \mid \mathsf{typ})$. This is where the typicality condition in (3.1.4) comes in.

Lemma 3.1.13. *Recall h as defined in Definition 3.1.11. We have*

$$\mathbb{P}\big(W = W' \mid \mathsf{typ}\big) \leq e^{-h}/\mathbb{P}(\mathsf{typ}).$$

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

$$\mathbb{P}\big(W = W',\, \mathsf{typ}\big) = \mathbb{P}\big(W = W',\, W \in \mathcal{W}\big) = \textstyle\sum_{w\in\mathcal{W}} \mathbb{P}\big(W = w\big)\mathbb{P}\big(W' = w\big) \leq e^{-h},$$

using the fact that $\sum_{w\in\mathcal{W}} \mathbb{P}\big(W = w\big) \leq 1$ and $\mathbb{P}(W' = w) \leq e^{-h}$ for all $w \in \mathcal{W}$. □

Consideration of $\mathbb{P}(S = S' \mid W \neq W', \mathsf{typ})$ is the topic of the next two subsections (§3.1.7 for $d = 3$ and §3.1.8 for general d). The main ingredient is the following lemma, which follows immediately from the fact that p is prime. (Recall that $[m] = \{1, ..., m\}$ for $m \in \mathbb{N}$.)

Lemma 3.1.14. *Let $X_1, ..., X_\ell \sim^{\mathrm{iid}} \mathrm{Unif}(\mathbb{Z}_p)$ and $a_1, ..., a_\ell \in [p-1]$. Then $\sum_1^\ell a_i X_i \sim \mathrm{Unif}(\mathbb{Z}_p)$.*

We can extend this to general $p \in \mathbb{N}$: we have $\sum_1^\ell a_i X_i \sim \mathrm{Unif}(\mathfrak{g}\mathbb{Z}_p)$ where $\mathfrak{g} := \gcd(a_1, ..., a_\ell, p)$ and $\mathfrak{g}\mathbb{Z}_p = \{g, 2g, ..., p\}$; see Lemma 2.1.11. (These statements are in $\mathbb{Z}_p$, ie modulo p.)

3.1.7 Upper Bound for 3×3 Heisenberg Matrices

As in §3.1.6, we use the abbreviations $H := H_{p,d}$, $A := H_{p,d}^{\mathrm{ab}}$ and $n := |H_{p,d}| = p^{d(d-1)/2}$. Here (§3.1.7), we study only $d = 3$. In the 3×3 case, we only have three terms to deal with. Abbreviate

$$\text{a matrix } M \in H_{p,3} \quad \text{by} \quad (M_{1,2}, M_{2,3}, M_{3,3}).$$

For matrices $M_1, M_2, ... \in H_{p,3}$, writing $M_j := (a_j, b_j, c_j)$ for each j, we have

$$\begin{gathered}\textstyle\prod_1^t M_s = \Big(\sum_1^t a_s,\, \sum_1^t b_s,\, \sum_1^t c_s + f\big((a_s)_1^t,\, (b_s)_1^t\big)\Big) \\ \text{where} \quad f\big((a_j)_1^t,\, (b_j)_1^t\big) := \textstyle\sum_{s=1}^t b_s \sum_{r=1}^{s-1} a_r,\end{gathered} \tag{3.1.5}$$

Note that the first two terms are 'Abelian' (and correspond to the Abelianisation): we can reorder the product $M_1 \cdots M_t$ in any way we desire, and the first two terms are unchanged; also, so is the first part of the third term, but the polynomial f is not.

We have k generators $Z = [Z_1, ..., Z_k]$; write $Z_i := (A_i, B_i, C_i)$ for each i. Recall that W is a DRW on $\mathbb{Z}_+^k$. Suppose that $N := N(t)$ steps are taken. Write $(\alpha_1, \beta_1, \gamma_1)$, ..., $(\alpha_N, \beta_N, \gamma_N)$ for the steps taken by S. Write G_m for the generator index chosen at step $m \in [N]$, ie $G_m = i$ if $(\alpha_m, \beta_m, \gamma_m) = (A_i, B_i, C_i)$. Write $\boldsymbol{\alpha} := (\alpha_m)_1^N$ and $\boldsymbol{\beta} := (\beta_m)_1^N$. Let S' be an independent copy of S, and make similar definitions. From (3.1.5), we have

$$S(t) = \big(\textstyle\sum_1^k A_i W_i(t),\, \sum_1^k B_i W_i(t),\, \sum_1^k C_i W_i(t) + f(\boldsymbol{\alpha}, \boldsymbol{\beta})\big). \tag{3.1.6}$$

Recall that we write "≡" to mean "equivalent modulo p".

Proof of Theorem 3.1.6 (when $d = 3$). First, we claim that

$$\mathbb{P}\big(S = S' \mid W \neq W', \, \mathsf{typ}\big) = 1/n = 1/p^3. \tag{3.1.7}$$

Indeed, for any $v \in \mathbb{Z}_p^k \setminus \{0\}$, by Lemma 3.1.14, each of $\sum_1^k A_i v_i$, $\sum_1^k B_i v_i$ and $\sum_1^k C_i v_i$ is an independent $\mathrm{Unif}(\mathbb{Z}_p)$; also, $f(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is independent of $\sum_1^k C_i W_i(t)$. Note also that, by typicality, $|W_i - W_i'| < p$ for all i, and so $W_i \equiv W_i' \bmod p$ if and only if $W_i = W_i'$. Hence conditioning on W and W' and then using (3.1.6) establishes the claim. Next, recall from Lemma 3.1.13 that

$$\mathbb{P}\big(W = W' \mid \mathsf{typ}\big) \leq e^{-h}/\mathbb{P}\big(\mathsf{typ}\big) = e^{-\omega} e^{-h_0}/\mathbb{P}\big(\mathsf{typ}\big). \tag{3.1.8}$$

It remains to consider the case that $W(t) = W'(t) = w$, for some $w \in \mathcal{W}$. In particular, S and S' take the same number of steps: $N = N'$. Note, by (3.1.6), that $S - S' = (0, 0, f(\boldsymbol{\alpha}, \boldsymbol{\beta}) - f(\boldsymbol{\alpha}', \boldsymbol{\beta}'))$.

Expanding the definition of f in (3.1.5), we may write

$$f(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \textstyle\sum_{i,j=1}^k C_{i,j} A_i B_j, \tag{3.1.9}$$

for appropriate $\{C_{i,j}\}_{i,j=1}^k$; specifically, for $i, j \in [k]$, we have

$$C_{i,j} \coloneqq \textstyle\sum_{\ell=1}^N \mathbf{1}\big(G_\ell = j\big) \sum_{m=1}^{\ell-1} \mathbf{1}\big(G_m = i\big); \quad \text{write} \quad \boldsymbol{C} \coloneqq (C_{i,j} \mid i, j \in [k]). \tag{3.1.10}$$

Define $C_{i,j}'$ and $\boldsymbol{C}'$ analogously with respect to W'. The body of the proof will be controlling the probability that $\boldsymbol{C} \equiv \boldsymbol{C}'$ conditional on $W(t) = W'(t) = w$, for some typical $w \in \mathcal{W}$. Write

$$\mathcal{E} \coloneqq \big\{\boldsymbol{C} \equiv \boldsymbol{C}'\big\} = \big\{C_{i,j} \equiv C_{i,j}' \; \forall \, i, j \in [k]\big\}. \tag{3.1.11}$$

We have $\mathbb{P}(S = S' \mid W = W' = w, \, \mathcal{E}) = 1$. We now argue that

$$\mathbb{P}\big(S = S' \mid W = W' = w, \, \mathcal{E}^c\big) \leq 2/p. \tag{3.1.12}$$

Write $D_{i,j} \coloneqq C_{i,j} - C_{i,j}'$. On the event $\mathcal{E}^c$, there exist $i', j' \in [k]$ with $D_{i',j'} \not\equiv 0$. Then

$$f(\boldsymbol{\alpha}, \boldsymbol{\beta}) - f(\boldsymbol{\alpha}', \boldsymbol{\beta}') = A_{i'}\big(D_{i',j'} B_{j'} + \textstyle\sum_{j \neq j'} D_{i,j} B_j\big) + \sum_{i \neq i'} A_i \sum_j D_{i,j} B_j. \tag{3.1.13}$$

We can write this final expression (with the natural association) as

$$U\big(V + X\big) + Y. \tag{3.1.14}$$

Since $D_{i',j'} \not\equiv 0$ (by choice of i' and j') and p is prime, $U, V \sim^{\mathrm{iid}} \mathrm{Unif}(\mathbb{Z}_p)$. Moreover, U is jointly independent of X and Y and V is independent of X (but not of Y); hence $V + X \sim \mathrm{Unif}(\mathbb{Z}_p)$, independent of U, and so $U(V+X) \sim \mathrm{Unif}(\mathbb{Z}_p)$ and is independent of Y on the event $\{V + X \not\equiv 0\}$. (These independence statements are all conditional on $W = W' = w$.) Thus

$$\mathbb{P}\big(U(V + X) + Y \equiv 0\big) \leq \max_u \mathbb{P}\big(U \equiv u\big) + \mathbb{P}\big(V + X \equiv 0\big) = 2/p. \tag{3.1.15}$$

This establishes (3.1.12).

Combining these results, recalling that $\mathcal{E} = \{\boldsymbol{C} \equiv \boldsymbol{C}'\}$, writing

$$q(t) \coloneqq \max_{w \in \mathcal{W}} \mathbb{P}\big(\mathcal{E} \mid W = W' = w\big),$$

recalling that w is an arbitrary (fixed) element of $\mathcal{W}$, we find that

$$\mathbb{P}\big(S = S' \mid W = W' = w, \, \mathsf{typ}\big) \leq 2/p + q(t). \tag{3.1.16}$$

Once we average over $w \in \mathcal{W}$, recalling (3.1.8), we obtain

$$\mathbb{P}\big(S = S', \, W = W' \mid \mathsf{typ}\big) \leq 2e^{-h}\big(1/p + q(t)\big)/\mathbb{P}\big(\mathsf{typ}\big). \tag{3.1.17}$$

It remains to make an appropriate definition of typicality, ie of $\mathcal{W}$: we require that it satisfies (3.1.4), that $\mathbb{P}(W \in \mathcal{W}) = 1 - o(1)$, and hence $\mathbb{P}(\mathsf{typ}) = 1 - o(1)$, and that $e^{-h}(2/p + q(t)) = o(1/n)$. This is done in Lemma 3.1.15 below; it is the main technical part of the proof.

Once this is done, combining (3.1.7, 3.1.17) gives

$$n\,\mathbb{P}\big(S = S' \mid \mathsf{typ}\big) - 1 = o(1).$$

The upper bound in Theorem 3.1.6 then follows from Lemma 3.1.8, modulo Lemma 3.1.15. □

It remains to appropriately upper bound $q(t)$ so that the right-hand side of (3.1.16) is $o(e^h/n)$.

Lemma 3.1.15. *Suppose that $1 \ll \log k \ll \log n$. There exists a $\mathcal{W} \subseteq \mathbb{Z}_+^k$, satisfying (3.1.4), so that*

$$\mathbb{P}\big(W \in \mathcal{W}\big) = 1 - o(1) \quad \text{and} \quad ne^{-h}\big(1/p + q(t)\big) = o(1).$$

For this proof, let $\varepsilon > 0$, and assume that it is as small as required (but independent of n). Recall that here $d = 3$, so $\log|A| \asymp \log n$ and $1 + \frac{2}{d-2} = 3$. Hence there are three main regimes:

$$k \ll \log|A|, \quad \log|A| \lesssim k \le (\log|A|)^3 \quad \text{and} \quad k \ge (\log|A|)^3.$$

Proof of Lemma 3.1.15 when $k \ge (\log|A|)^3$. We have $t \ge (1+3\varepsilon)t_{\mathrm{diam}} = (1+3\varepsilon)\log_k|G|$. Recall from Definition 3.1.11 that, in this regime, we take $h_0 := (1-\frac{1}{\rho})\log n$. Hence $e^{-h_0} = n^{-1+1/\rho}$.

Since $t \ll k$, almost all the generators are picked at most once whp. For $w \in \mathbb{Z}_+^k$, define

$$\mathcal{J}(w) := \big\{i \in [k] \mid w_i = 1\big\} \quad \text{and} \quad J(w) := |\mathcal{J}(w)|.$$

Using this, we make precise our definition of typicality:

$$\mathcal{W} := \big\{w \in \mathbb{Z}_+^k \;\big|\; \mu_t(w) \le e^{-h},\ |J(w) - te^{-t/k}| \le \tfrac{1}{2}\varepsilon te^{-t/k},\ \max_i w_i < \tfrac{1}{2}p\big\},$$

satisfying (3.1.4). Using the conditions of Hypothesis E, we have $\log_k n \ll \sqrt{p}$ and $t_0 \lesssim k \ll \sqrt{p}$. Thus the condition $\{\max_i w_i < \frac{1}{2}p\}$ holds with probability $1 - o(1)$. By Binomial concentration and Lemma 3.1.12, we then have $\mathbb{P}(W \in \mathcal{W}) = 1 - o(1)$.

Let $w \in \mathcal{W}$. We now argue that

$$\mathbb{P}\big(\mathcal{E} \;\big|\; W = W' = w,\ |\mathcal{J}(w)| = J\big) \le 1/J!. \tag{3.1.18}$$

This holds since, conditional on $W = W' = w$, *different* (relative) orderings, between S and S', of the coordinates chosen once, ie in $\mathcal{J}(w)$, must result in some pair (i,j) such that $C_{i,j} = 1$ and $C'_{i,j} = 0$. There are $J!$ different orderings.

Applying (3.1.18), using the condition $|J(w) - te^{-t/k}| \le \frac{1}{2}\varepsilon te^{-t/k}$ for $w \in \mathcal{W}$, gives

$$q(t) \le 1/\big((1-\varepsilon)t\big)! \tag{3.1.19}$$

Note that $k = (\log|H|)^{\rho/(1+\nu)}$, and so $t_{\mathrm{diam}} = \log_k n = \frac{1+\nu}{\rho}\log n/\log\log n$. Using $t \ge (1+3\varepsilon)t_{\mathrm{diam}}$ in (3.1.19), direct calculation with Stirling's approximation gives

$$q(t) \le \big((1+\varepsilon)t_{\mathrm{diam}}/e\big)^{-(1+\varepsilon)t_{\mathrm{diam}}} \le n^{-1/\rho}. \tag{3.1.20}$$

Recalling that $e^{-h} = e^{-\omega}n^{-1+1/\rho}$, the proof is completed in the regime $k \ge (\log|A|)^3$:

$$ne^{-h}\big(1/p + q(t)\big) \le 2n \cdot e^{-\omega}n^{-1+(1+\nu)/\rho} \cdot n^{-(1+\nu)/\rho} = 2e^{-\omega} \ll 1. \qquad \square$$

Proof of Lemma 3.1.15 when $\log|A| \lesssim k \le (\log|A|)^3$. We have $t \ge (1+3\varepsilon)t_0$. Recall from Definition 3.1.11 that, in this regime, we take $h_0 := \log|A|$. Hence $e^{-h_0} = |A|^{-1}$.

Since $d \asymp 1$, we have $1 \ll t \lesssim k$. We use the same definition of typicality here as for $k \ge (\log|A|)^3$. Since $1 \ll t \lesssim k$, we have $\mathbb{P}(W \in \mathcal{W}) = 1 - o(1)$.

Since $t \ge (1+3\varepsilon)t_0$, direct calculation using (3.1.1c, 3.1.19) and Stirling's approximation gives

$$q\big((1+3\varepsilon)t_0\big) \le \big((1+\varepsilon)t_0/e\big)^{-(1+\varepsilon)t_0} \le |A|^{1/(\rho-1)} \quad \text{when} \quad k \gg \log|A|. \tag{3.1.21a}$$

For $k \asymp \lambda\log|A|$, with $\lambda \in (0,\infty)$, we have $t_0 \asymp f(\lambda)k \asymp \lambda f(\lambda)\log|A|$ by (3.1.3b), and thus

$$\mathbb{E}\big(|\mathcal{J}|\big) \asymp \lambda f(\lambda)e^{-f(\lambda)}\log|A|.$$

Applying (3.1.18), using the condition $|J(w) - te^{-t/k}| \le \frac{1}{2}\varepsilon te^{-t/k}$ for $w \in \mathcal{W}$, gives

$$q(t) \le 1/\big((1-\varepsilon)\lambda f(\lambda)e^{-f(\lambda)}\log|A|\big)!;$$

applying Stirling's approximation, it is easy to see that this decays super-polynomially, ie

$$\log\big(1/q\big((1+3\varepsilon)t_0\big)\big) \gg \log|A| \quad \text{when} \quad k \asymp \lambda \log|A|, \tag{3.1.21b}$$

provided ε is sufficiently small. Hence

$$q(t) \le |A|^{1/(\rho-1)} \quad \text{when} \quad \log n \lesssim k \le (\log|A|)^3. \tag{3.1.22}$$

Recall that we want to compare $q(t)$ with $1/p = |A|/|H|$. Recall that $|H| = |A|^{2/d}$. Some simple algebra then shows that $q(t) \le 1/p$ when $\rho \ge 1 + \frac{2}{d-2} = 3$. Recalling that $e^{-h} = e^{-\omega}|A|^{-1}$, the proof is completed in the regime $\log|A| \lesssim k \le (\log|A|)^3$:

$$ne^{-h}\big(2/p + q(t)\big) \le |H| \cdot e^{-\omega}|A|^{-1} \cdot 3|A|/|H| = 3e^{-\omega} \ll 1. \qquad \square$$

Proof of Lemma 3.1.15 when $1 \ll k \ll \log|A|$. We have $t \ge (1+3\varepsilon)t_0$. Recall from Definition 3.1.11 that, in this regime, we take $h_0 := \log|A|$. Hence $e^{-h_0} = |A|^{-1}$.

Since $d \asymp 1$, we have $t_0 \asymp k|A|^{2/k} = kp^{4/k} \gg k$. Hence the same generator is picked lots of times, and so we need a new approach for calculating $q(t)$. The expected number of times a generator is picked is $s := t/k \gg 1$. As part of our typicality requirements, we ask that 'most' pairs $(2i, 2i-1)$, with $i \in \{1, ..., \lfloor k/2 \rfloor\}$, are picked between ηs and $\eta^{-1}s$ times, for a small positive constant η, to be chosen later; for the moment, let $\eta \in (0,1)$. For $w \in \mathbb{Z}_+^k$, write

$$\mathcal{C}(w) := \big\{ i \in \{1, ..., \lfloor k/2 \rfloor\} \;\big|\; \eta s \le \min\{w_{2i}, w_{2i-1}\} \le \max\{w_{2i}, w_{2i-1}\} \le \eta^{-1}s \big\}.$$

Then, for η sufficiently small (but still a constant), we have

$$\mathbb{P}\big(|\mathcal{C}(W)| \ge \tfrac{2}{5}k\big) = 1 - o(1). \tag{3.1.23}$$

(We could replace $\frac{2}{5}$ by any constant less than $\frac{1}{2}$, at the cost only of making η a smaller constant.) We use this to make precise our definition of typicality for this regime:

$$\mathcal{W} := \big\{ w \in \mathbb{Z}_+^k \;\big|\; \mu_t(w) \le e^{-h},\ |\mathcal{C}(w)| \ge \tfrac{2}{5}k,\ \max_i w_i < \tfrac{1}{2}p \big\}.$$

Then, like before and additionally using (3.1.23), we have $\mathbb{P}(W \in \mathcal{W}) = 1 - o(1)$. If $i \in \mathcal{C}(w)$, then $\max\{C_{2i,2i-1}, C'_{2i,2i-1}\} \le w_i^2 \lesssim s^2 \ll p$ as $d \asymp 1$, so $\{C_{2i,2i-1} \equiv C'_{2i,2i-1}\} = \{C_{2i,2i-1} = C'_{2i,2i-1}\}$.

We claim that it is sufficient to fix an arbitrary $w \in \mathcal{W}$ and prove the bound

$$\max_i q_i \le p^{-3/k} \quad \text{where} \quad q_i := \max_x \mathbb{P}\big(C_{2i,2i-1} = x \;\big|\; W = w\big)\mathbf{1}\big(i \in \mathcal{C}(w)\big). \tag{3.1.24}$$

To see this, first make the simple observation that, for any $\mathcal{I} \subseteq \{1, ..., \lfloor k/2 \rfloor\}$, we have

$$\big\{(C_{i,j})_{i,j\in[k]} = (C'_{i,j})_{i,j\in[k]}\big\} \subseteq \big\{(C_{2i,2i-1})_{i\in\mathcal{I}} = (C'_{2i,2i-1})_{i\in\mathcal{I}}\big\}.$$

Given $W = W' = w$, the event $\{C_{i,j} = C'_{i,j}\}$ is determined by the relative order in which the generators i and j are chosen. Hence, since the pairs $(2i, 2i-1)$ are disjoint, the events $\{C_{2i,2i-1} = C'_{2i,2i-1}\}$ are independent for different i, conditional on $W = W' = w$. Take $\mathcal{I} := \mathcal{C}(w)$, which has size at least $\frac{2}{5}k$. By the aforementioned independence, given (3.1.24), we have

$$\mathbb{P}\big(\boldsymbol{C} = \boldsymbol{C}' \;\big|\; W = W' = w\big) \le (\max_i q_i)^{2k/5} \le p^{(-3/k)(2k/5)} = p^{-6/5} \ll 1/p,$$

and hence $q(t) \ll 1/p$. The proof is then completed as in the regime $\log|A| \lesssim k \le (\log|A|)^3$.

It remains to prove (3.1.24). For simplicity of notation, we assume that $1 \in \mathcal{C}(w)$ and set $i := 1$. Let $r := w_1 + w_2$, and write our random word as $S = Z_{G_1} \cdots Z_{G_N}$; here $N = \sum_1^k w_i$ is the number of steps taken and G_ℓ is the generator index chosen in the ℓ-th step. Let $J_1 < \cdots < J_r$ be the (random) indices with $G_{J_\ell} \in \{1,2\}$. Now define the vector $I \in \{1,2\}^r$ by $I_\ell := G_{J_\ell}$. Thus I encodes the relative order between the different occurrences (with multiplicities) of the generators labelled by $\{1,2\}$ in the word S. By typicality, $2\eta s \le r \le 2\eta^{-1}s$.

Let I' be the random vector obtained from I by picking a 2 uniformly at random and omitting it from I. (Eg, if $I = (2,2,1,1,2,1)$ and we pick the last 2, then $I' = (2,2,1,1,1)$.) Importantly, we

are omitting elements of the *relative* order of appearances of Z_1 and Z_2, not the *absolute* locations of the corresponding generators.

By the definition of $C_{1,2}$, given in (3.1.10), given $W = w$, the value of $C_{1,2}$ is a function only of the relative locations I. Hence, given I' also, it is a function only of the location of the omitted 2. It is constant on the set of locations which give rise to the same I: two different placements of the omitted 2 give rise to the same I if and only if they both lie in the same (possibly empty) interval of consecutive 2s. (Eg, if $I' = (2, 2, 1)$, then there are three locations in which we can insert a 2 to get $I = (2, 2, 2, 1)$, namely the first, second and third positions, and only one to get $I = (2, 2, 1, 2)$, namely the fourth position; the first three give rise to $C_{1,2} = 0$ and the fourth to $C_{1,2} = 1$.)

Hence, writing $L(I')$ for the longest interval of 2s in I', we have

$$\max_x \mathbb{P}\big(C_{1,2} = x \;\big|\; W = w,\ I'\big) \le \big(L(I') + 1\big)/r.$$

By Claim 3.1.16 below, with $m = 2$, we find that $L(I')/(C \log r) \preceq \mathrm{Geom}(\frac{1}{2})$ given $W = w$ for a sufficiently large constant C, and so $\mathbb{E}(L(I') \mid W = w) \lesssim \log r$. Hence

$$\max_x \mathbb{P}\big(C_{1,2} = x \;\big|\; W = w\big) \le \mathbb{E}\big(L(I') + 1 \;\big|\; W = w\big)/r \lesssim (\log r)/r.$$

Since $2\eta s \le r \le 2\eta^{-1}s$, as $w \in \mathcal{W}$, and η is a (small) constant, this last expression is $o(1/s^{3/4})$. (In fact, it is $\mathcal{O}(\log s)/s$.) Recalling that $s \asymp p^{4/k}$ establishes (3.1.24). This completes the proof. □

It remains to state and prove the claim regarding $\mathbb{E}(L(I'))$. We actually state and prove a slightly more general claim, that we are then able to use in the analysis of the $d \times d$ matrices.

Claim 3.1.16. *Let $m \in \mathbb{N}$ and $\eta \in (0, 1)$. Let $\{w_1, ..., w_m\}$ be arbitrary positive integers satisfying $w_i/w_j \in [\eta^2, \eta^{-2}]$ for all $i, j \in [m]$. For each $k \in \{1, ..., m\}$, let there be w_k balls of colour k; write $r \coloneqq \sum_{k=1}^m w_k$ for the total number of balls. Choose a uniform permutation of the balls on positions $\{1, ..., r\}$. For each $k \in \{1, ..., m\}$, let L_k be the longest interval* without *any balls of colour k. Then, for each k, we have the stochastic domination*

$$L_k/(\eta^{-2} m \log r) \preceq \mathrm{Geom}(\tfrac{1}{2}).$$

Proof. Without loss of generality, take $k \coloneqq 1$ and write $L \coloneqq L_1$. By assumption, $w_i/w_j \in [\eta^2, \eta^{-2}]$ for all i and j, and $\eta \in (0, 1)$ is a constant. Hence $r \le m\eta^{-2}w_1$, and so $w_1 \ge \eta^2 r/m$. Let $\ell \in \mathbb{N}$ to be chosen shortly; write $[1, r] \subseteq [1, \ell] \cup [2, \ell + 1] \cup \cdots \cup [r, r + \ell - 1]$. By direct calculation,

$$\begin{aligned}
\mathbb{P}\big(L > \ell\big) &\le r\mathbb{P}\big(\text{no 1 in the interval } [1, \ell]\big) \\
&= r \cdot \big(1 - \tfrac{w_1 - 1}{r-1}\big)\big(1 - \tfrac{w_1-1}{r-2}\big) \cdots \big(1 - \tfrac{w_1-1}{r-\ell}\big) \\
&\le r\big(1 - \tfrac{w_1-1}{r-1}\big)^{\ell} \le r\exp\big(-\ell w_1/r\big) \le r\exp\big(-\ell\eta^2/m\big),
\end{aligned}$$

where for the penultimate inequality we used the fact that $\frac{w_1-1}{r-1} \le \frac{w_1}{r}$, which holds since $w_1 \le r$. Choosing $\ell \coloneqq (k + 1)\eta^{-2} m \log r$ gives

$$\mathbb{P}\big(L > (k + 1)\eta^{-2} m \log r\big) \le r\exp\big(-(k + 1)\log r\big) = r^{-k}.$$

Thus we may stochastically dominate

$$L/(\eta^{-2} m \log r) \preceq \mathrm{Geom}(1 - 1/r) \preceq \mathrm{Geom}(\tfrac{1}{2}).$$ □

3.1.8 Upper Bound for $d \times d$ Heisenberg Matrices

The high-level ideas of the proof will be similar to the $d = 3$ case, but there are a number of subtleties which need to be navigated. Analogously to Lemma 3.1.15, there will be a certain probability that requires bounding, and the argument for bounding this will differ depending on k; the specific reference will be Lemmas 3.1.18 and 3.1.10, and comes at the end of the section.

We also use the same preliminaries (see §3.1.6), and in particular consider

$$D(t) = |H_{p,d}|\,\mathbb{P}\big(S = S' \mid \mathsf{typ}\big) - 1.$$

The analogues of (3.1.5, 3.1.6) are different for general d than for $d = 3$: they have the same basic structure, but with the addition of 'higher order' terms (given by $g_{a,b}$ in the lemma below). The following lemma is for both the DRW and SRW; take $\sigma_\ell := 1$ for all ℓ to reduce to the DRW.

Lemma 3.1.17. *Let $Z_1, ..., Z_k \in H_{p,d}$. Let $\gamma \in [k]^L$ and $\sigma \in \{\pm 1\}^L$. For $i, j \in [k]$, set*

$$C_{i,j}(\gamma, \sigma) := \sum_{\ell=0}^{L} \sum_{m=0}^{\ell-1} \sigma_m \sigma_\ell \mathbf{1}(\gamma_m = i,\ \gamma_\ell = j) + \mathbf{1}(i = j) \sum_{\ell=0}^{L} \mathbf{1}(\gamma_\ell = i,\ \sigma_\ell = -1).$$

Set $M := Z_{\gamma_1}^{\sigma_1} \cdots Z_{\gamma_L}^{\sigma_L}$. Then, for all $a \in [d]$, we have

$$M(a, a) = 1 \quad \text{and} \quad M(a, a+1) = \sum_{\ell=1}^{L} \sigma_{\gamma_\ell} Z_{\gamma_\ell}(a, a+1),$$

and, for all $a, b \in [d]$ with $b \geq a + 2$, we have

$$M(a, b) = \sum_{\ell \in [L]} Z_{\gamma_\ell}(a, b) + \sum_{i,j \in [k]} C_{i,j}(\gamma, \sigma) Z_i(a, a+1) Z_j(a+1, b) + g_{a,b}(\gamma, \sigma; Z_1, ..., Z_k), \tag{3.1.25}$$

where $g_{a,b}(\gamma, \sigma; Z_1, ..., Z_k)$ is a polynomial in $(Z_i(x, y) : i \in [k],\ x \in [d-1],\ y > x)$. Further, in this polynomial, each monomial contains the term $Z_i(a, a+1)$ either 0 times or exactly once and no monomial contains a term of the form $Z_i(a, a+1)Z_j(a+1, b)$ for $i, j \in [k]$.

We give a sketch of the argument here; the rigorous details are deferred to Lemma 6.6.2.

Proof Sketch of Lemma 3.1.17. The fundamental idea is to write a matrix $M_\ell \in H_{p,d}$ as $I + N_\ell$ where N_ℓ is *strictly* upper triangular. For $M_1, ..., M_L \in H_{p,d}$ written like this, one then has

$$M_1 \cdots M_L = \prod_{\ell=1}^{L} (1 + N_\ell) = I + \sum_{\ell=1}^{L} N_\ell + \sum_{m_1 < m_2} N_{m_1} N_{m_2} + \sum_{\ell=3}^{L} \sum_{m_1 < \cdots < m_\ell} \prod_{r=1}^{\ell} N_{m_r},$$

where the indices m_r run over $[L]$. Further, for (a, b) with $b - a \geq 2$, one can write

$$\begin{aligned}\big(N_{m_1} N_{m_2}\big)(a, b) &= \sum_{c \in [1,d]} N_{m_1}(a, c) N_{m_2}(c, b) = \sum_{c \in [a+1, b-1]} M_{m_1}(a, c) M_{m_2}(c, b) \\ &= M_{m_1}(a, a+1) M_2(a+1, b) + \sum_{c=a+2}^{b-1} M_{m_1}(a, c) M_2(c, b).\end{aligned}$$

We consider this latter sum along with all products of degree at least 3 as 'higher order' terms. Writing $\sum_{m_1 < m_2}$ as $\sum_{m_2=1}^{L} \sum_{m_1=1}^{m_2-1}$, the formula for $C_{i,j}$ follows for the DRW (ie $\sigma_\ell := 1$ for all ℓ).

The SRW analysis is similar. Since N_ℓ is strictly upper triangular, $N_\ell^d = 0$. Thus

$$M_\ell^{-1} = (1 + N_\ell)^{-1} = I - N_\ell + N_\ell^2 + \sum_{t=3}^{d} (-1)^t N_\ell^t.$$

Separating out 'higher order' terms similarly, we deduce the formula for the SRW. □

As previously, we use the abbreviations $H := H_{p,d}$, $A := H_{p,d}^{\mathrm{ab}}$ and $n := |H_{p,d}| = p^{d(d-1)/2}$.

Proof of Theorem 3.1.6 (general d). When $W(t) \neq W'(t)$, the same argument as for $d = 3$, using Lemma 3.1.14, applies, replacing (3.1.6) by (3.1.25):

$$\mathbb{P}\big(S = S' \mid W \neq W',\ \mathsf{typ}\big) = 1/n = 1/p^{d(d-1)/2} = p^{-(d-1)(d-2)/2} \cdot p^{-(d-1)}. \tag{3.1.26}$$

This is the analogue of (3.1.7). Next, recall from Lemma 3.1.13 that

$$\mathbb{P}\big(W = W' \mid \mathsf{typ}\big) \leq e^{-h} / \mathbb{P}\big(\mathsf{typ}\big) = e^{-\omega} e^{-h_0} / \mathbb{P}\big(\mathsf{typ}\big). \tag{3.1.27}$$

Now suppose that $W(t) = W'(t) = w$, where w is some fixed element of $\mathcal{W}$ (yet to be defined fully). Then the 'Abelian' parts of S and S', corresponding to the first term in the right-hand side of (3.1.25), cancel (as was the case when $d = 3$). Write $\boldsymbol{C} := (C_{i,j})$ and $\boldsymbol{C}' := (C'_{i,j})$ for the $C(\gamma)$ in Lemma 3.1.17 generated by S and S', respectively. Write $\mathcal{E} := \{\boldsymbol{C} = \boldsymbol{C}'\}$. On $\mathcal{E}$, the middle terms of (3.1.25) cancel, leaving only the higher-order terms; upper bound $\mathbb{P}(S = S' \mid W = W',\ \mathcal{E}) \leq 1$.

Now suppose that $\mathcal{E}$ does not hold; choose, and fix, (i', j') so that $C_{i',j'} \not\equiv C'_{i',j'}$. By the condition (3.1.4) which $\mathcal{W}$ must satisfy and the definition of $C_{i,j}$, this implies that $C_{i',j'} \neq C'_{i',j'}$. Analogously to (3.1.13, 3.1.14), where d was equal to 3, letting

$$U_{a,b} := Z_{i'}(a, a+1) \quad \text{and} \quad V_{a,b} := (C_{i',j'} - C'_{i',j'}) Z_{j'}(a+1, b), \tag{3.1.28}$$

we can, for some random variables $X_{a,b}$ and $Y_{a,b}$, write

$$\textstyle\sum_{i,j}^{k}(C_{i,j}-C'_{i,j})Z_i(a,a+1)Z_j(a+1,b) \quad \text{naturally as} \quad U_{a,b}(V_{a,b}+X_{a,b})+Y_{a,b}.$$

For the moment, fix (a,b). Analogously to the $d=3$ case, ie (3.1.12–3.1.15), the following hold: $U_{a,b}, V_{a,b} \sim \mathrm{Unif}(\mathbb{Z}_p)$; $U_{a,b}$ is independent of $(V_{a,b}, X_{a,b}, Y_{a,b})$; $V_{a,b}$ is independent of $X_{a,b}$ (but not of $Y_{a,b}$). Thus $U_{a,b}(V_{a,b}+X_{a,b}) \sim \mathrm{Unif}(\mathbb{Z}_p)$ is independent of Y on the event $V_{a,b}+X_{a,b} \not\equiv 0$. Hence

$$\max_r \mathbb{P}\big(U_{a,b}(V_{a,b}+X_{a,b})+Y_{a,b} \equiv r\big) \le \max_u \mathbb{P}\big(U_{a,b} \equiv u\big) + \mathbb{P}\big(V_{a,b}+X_{a,b} \equiv 0\big) \le 2/p; \quad (3.1.29)$$

Now compare $S_{a,b}$ and $S'_{a,b}$. Since $W = W'$, the 'Abelian' part cancels; we are left with the $U_{a,b}(V_{a,b}+X_{a,b})+Y_{a,b}$ part and the higher-order terms, given by the $g_{a,b}$ polynomials in (3.1.25). These two parts are independent, by the conditions of Lemma 3.1.17. Hence (3.1.29) implies that

$$\mathbb{P}\big(S_{a,b} = S'_{a,b} \,\big|\, W = W' = w,\ \mathcal{E}^c\big) \le 2/p. \quad (3.1.30)$$

Now, the random variables $\{X_{a,b}, Y_{a,b}\}_{a,b}$ are not independent. Also, $U_{a,b} = Z_{i'}(a,a+1)$ does not depend on b, and so $\{U_{a,b}, V_{a,b} \mid b \ge a+2\}_{a,b}$ are not independent either. However, if we fix b then $\{U_{a,b}, V_{a,b} \mid b \ge a+2\}_a$ is a collection of independent variables. We exploit this.

Partition the $[k]$ generators into $d-2$ sets $(P_3, ..., P_d)$. For each (fixed) $b \in \{3, ..., d\}$, we use generators only from P_b; this will give independence when we consider all b. (Note that for $b \in \{1,2\}$ there are no terms above the super-diagonal.) Then for the (a,b)-th coordinate we try to get $C_{i',j'} \not\equiv C'_{i',j'}$ for some (i',j') with $i', j' \in P_b$. Now, for each b, using this pair (i',j') in the definition (3.1.28) of $U_{a,b}$ and $V_{a,b}$, the random variables $\{U_{a,b}, V_{a,b} \mid b \ge a+2\}_a$ are independent, since they depend on a disjoint set of generators.

For each $b \in \{3, ..., d\}$, write

$$\boldsymbol{C}_b := (C_{i,j})_{i,j \in P_b}, \quad \boldsymbol{C}'_b := (C'_{i,j})_{i,j \in P_b} \quad \text{and} \quad \mathcal{E}_b := \{\boldsymbol{C}_b = \boldsymbol{C}'_b\}. \quad (3.1.31)$$

We wish to get an analogue of (3.1.16), for general d. Write $\overline{\mathbb{P}}_w(\cdot) := \mathbb{P}(\cdot \mid W = W' = w)$ for $w \in \mathcal{W}$, and $S_{:,b} := (S_{a,b} \mid a = 1, ..., b-2)$ for the b-th column strictly above the super-diagonal; also, henceforth, in $\sum_3^d$ and $\prod_3^d$, the implicit index is always b. Then

$$\overline{\mathbb{P}}_w\big(S = S'\big) = \textstyle\prod_3^d \overline{\mathbb{P}}_w\big(S_{:,b} = S'_{:,b} \,\big|\, S_{:,b'} = S'_{:,b'}\ \forall\, b' = 3, ..., b-1\big)$$

Using (3.1.30), and noting that $S_{:,b}$ has $b-2$ entries, we obtain

$$\overline{\mathbb{P}}_w\big(S_{:,b} = S'_{:,b} \,\big|\, S_{:,b'} = S'_{:,b'}\ \forall\, b' = 3, ..., b-1\big) \le (2/p)^{b-2} + \overline{\mathbb{P}}_w\big(\mathcal{E}_b\big);$$

this uses the aforementioned independence between columns, guaranteed by the partitioning of the generators. Combining these two equations, we obtain

$$\overline{\mathbb{P}}_w\big(S = S'\big) \le 2^{d^2/2} \textstyle\prod_3^d \big(1/p^{b-2} + q_b(t)\big) \quad \text{where} \quad q_b(t) := \max_{w \in \mathcal{W}} \prod_3^d \overline{\mathbb{P}}_w(\mathcal{E}_b). \quad (3.1.32)$$

It remains to make an appropriate definition of typicality, ie of $\mathcal{W}$, and choose the partition $(P_3, ..., P_d)$ appropriately. For reasons explained later, we end up choosing P_b so that $R_b := |P_b|/k = (b-2)/\binom{d-1}{2}$, omitting floor/ceiling signs. (Note that $\sum_3^d R_b = 1$, as required.) We justify the omission of floor/ceiling signs by the fact that $|P_b| \asymp (b-2)kd^{-2} \gg 1$ (as $d^2 \ll k$).

This is all done in Lemma 3.1.18, which gives the following bound:

$$n\,\overline{\mathbb{P}}_w\big(S = S'\big) \equiv n\,\mathbb{P}\big(S = S' \mid W = W' = w\big) \le e^{h_0} 2^{d^2}. \quad (3.1.33)$$

Combined with (3.1.26, 3.1.27) this implies that

$$n\,\mathbb{P}\big(S = S' \mid \mathsf{typ}\big) - 1 \le 2 \cdot e^{-\omega} 2^{d^2}, \quad (3.1.34)$$

where we shall choose typ so that $\mathbb{P}(\mathsf{typ}) = 1 - o(1)$. If we can show that we can choose $\omega \gg d^2$, then the upper bound in Theorem 3.1.6 then follows from Lemma 3.1.8, modulo Lemma 3.1.18.

It remains to prove that we can choose $\omega \gg d^2$. Lemma 3.1.10 says that we can choose any $\omega \ll \min\{k, \log|A|\}$. Hypothesis E implies that $d^2 \ll \min\{k, \log|A|\}$ is satisfied, as required:

$$d^3 \ll k \text{ when } k \ll \log|A| \quad \text{and} \quad d \ll \log\log p \text{ when } k \gtrsim \log|A|. \qquad \square$$

It remains to appropriately bound $q_b(t)$, defined in (3.1.32).

Lemma 3.1.18. *Let $\varepsilon > 0$ and set $t := (1+3\varepsilon)t_*$. Assume the conditions of Hypothesis E. Then there exists a $\mathcal{W} \subseteq \mathbb{Z}_+^k$, satisfying (3.1.4), so that*

$$\mathbb{P}\big(W \in \mathcal{W}\big) = 1 - o(1) \quad \text{and} \quad |H| e^{-h_0} \textstyle\prod_3^d \big(1/p^{b-2} + q_b(t)\big) \le 2^{d^2/2}.$$

Recall the condition on $\mathcal{W}$ given by (3.1.4). Since $t \ge (1+3\varepsilon)t_0$, by Lemma 3.1.12, this condition is satisfied with probability $1-o(1)$. Hence we need only check that any additional constraints are also satisfied with probability $1-o(1)$. Recall that we use the notation

$$k = \big(\log|A|\big)^\rho \quad \text{and} \quad \tfrac{1}{2}d = \big(\log|H|\big)^\nu, \quad \text{so} \quad k = \big(\log|H|\big)^{\rho/(1+\nu)}, \quad \text{and} \quad n = |H| = p^{d(d-1)/2}.$$

Proof of Lemma 3.1.18 for $k \gtrsim \log|A|$. Let $\varepsilon > 0$ and set $t := (1+3\varepsilon)t_*$; write $s := t/k$.

Typicality. As when $d = 3$, when $k \gg \log|A|$ almost all the generators are picked at most once; when $k \asymp \log|A|$, a constant proportion are. As part of our typicality requirement (typ), we ask that at least $(1-\varepsilon)te^{-t/k}$ generators are picked exactly once—ie at least $(1-\varepsilon)$ times the expected number. Given this, we can then choose our partition so that, for each $b \in \{3,\dots,d\}$, writing $R_b := |P_b|/k$, at least $(1-\varepsilon)te^{-t/k}R_b$ generators from P_b are picked exactly once.

We can hence use the same definition of typicality, for $k \gtrsim \log|A|$, as when $d = 3$:

$$\mathcal{W} := \big\{w \in \mathbb{Z}_+ \;\big|\; \mu_t(w) \le e^{-h},\ |J(w) - se^{-s}k| \le \tfrac{1}{2}\varepsilon se^{-s}k,\ \max_i w_i < \sqrt{p}\big\}, \tag{3.1.35}$$

satisfying (3.1.4), recalling that $J(w) = \sum_1^k \mathbf{1}(w_i = 1)$. As previously, $\mathbb{P}(W \in \mathcal{W}) = 1 - o(1)$.

Analogously to (3.1.19), when $k \gg \log|A|$, we have

$$q_b(t) \le 1/\big((1-2\varepsilon)tR_b\big)!, \tag{3.1.36}$$

where we have absorbed the $e^{-t/k} = 1 - o(1)$ term into the $(1-2\varepsilon)$; we consider $k \asymp \log|A|$ later.

Regime $k \ge (\log|A|)^{1+2/(d-2)}$. We have $t \ge (1+3\varepsilon)t_{\mathrm{diam}} = (1+3\varepsilon)\log_k|A|$. Direct calculation, analogous to (3.1.20), using (3.1.3d) and (3.1.36) and Stirling's approximation gives

$$q_b(t) \le 1/\big((1-2\varepsilon)\cdot(1+3\varepsilon)t_{\mathrm{diam}}\cdot R_b\big)! \le n^{-R_b 1/\rho}.$$

In (3.1.20), we upper bounded $q(t) \le n^{-1/\rho}$, and this term was dominant in the sum $1/p + n^{-1/\rho}$. Here, we compare $q_b(t) \le n^{-R_b/\rho}$ and $1/p^{b-2}$. It is thus natural to choose $R_b \propto b-2$, ie $R_b := (b-2)/\binom{d-1}{2}$, for $b \in \{3,\dots,d\}$. Observe that

$$1/p^{b-2} \le n^{-R_b/\rho} \quad \text{if and only if} \quad \rho(b-2)/\tbinom{d}{2} \ge R_b = (b-2)/\tbinom{d-1}{2};$$

hence we need $\rho \ge \binom{d}{2}/\binom{d-1}{2} = 1 + \frac{2}{d-2}$, which is precisely the regime which we are considering.

Combining the upper bounds just developed, we deduce that

$$\textstyle\prod_3^d\big(1/p^{b-2} + q_b(t)\big) \le 2^d \prod_3^d q_b(t) \le 2^d n^{-1/\rho},$$

since $\sum_3^d R_b = 1$. Recalling that $h_0 = (1-\frac{1}{\rho})\log n$ in this regime, we deduce the desired bound.

Regime $\log|A| \lesssim k \ll (\log|A|)^{1+2/(d-2)}$. We have $t \ge (1+3\varepsilon)t_0$. Recall from Definition 3.1.11 that, in this regime, we take $h_0 := \log|A|$. Hence $e^{-h_0} = |A|^{-1}$. We subdivide the regime.

Consider first $k \gg \log|A|$. Direct calculation, analogous to (3.1.21a), using (3.1.3c) and (3.1.36) and Stirling's approximation, gives

$$q_b(t) \le 1/\big((1-2\varepsilon)\cdot(1+3\varepsilon)t_0\cdot R_b\big)! \le \exp\big(-\tfrac{2}{d}\tfrac{1}{\rho-1}R_b\log n\big); \tag{3.1.37}$$

again, this crucially uses the fact that $d = (\log n)^{o(1)}$ and $\varepsilon > 0$ is a constant.

In (3.1.21a), we upper bounded $q(t) \le |A|^{1/(\rho-1)}$, and this term was subdominant in the sum $1/p + |A|^{1/(\rho-1)}$. Here, we compare $q_b(t) \le |A|^{R_b/(\rho-1)}$ with $1/p^{b-2}$. Again, it is thus natural to choose $R_b \propto b-2$, ie $R_b := (b-2)/\binom{d-1}{2}$, for $b \in \{3,\dots,d\}$. Observe that

$$1/p^{b-2} \ge \exp\bigl(-\tfrac{2}{d}\tfrac{1}{\rho-1}R_b \log n\bigr) \quad \text{if and only if} \quad (\rho-1)(b-2)/(d-1) \le R_b = (b-2)/\tbinom{d-1}{2};$$

hence we need $\rho \le 1 + \frac{2}{d-2}$, which is precisely the regime that we are considering.

Combining the upper bounds just developed, we deduce that

$$\textstyle\prod_3^d \bigl(1/p^{b-2} + q_b(t)\bigr) \le 2^d \prod_3^d 1/p^{b-2} \le 2^d p^{-\binom{d-1}{2}} = 2^d |A|/|H|.$$

Recalling that $h_0 = \log|A|$ in this regime, we deduce the desired bound.

Consider now $k \asymp \log|A|$. Suppose that $k \approx \lambda \log|A|$ with $\lambda \in (0,\infty)$. Direct calculation, analogous to (3.1.21b), using (3.1.3b) and (3.1.36) and Stirling's approximation gives

$$\log\bigl(1/q_b(t)\bigr) \asymp d^{-1} \log\log|A| \cdot (b-2)\log p \gg (b-2)\log p, \tag{3.1.38}$$

using the conditions on d. The proof is then completed in exactly the same way as above. □

Proof of Lemma 3.1.18 for $k \ll \log|A|$. Set $t := (1+3\varepsilon)t_* \ge (1+3\varepsilon)t_0$. Recall from Definition 3.1.11 that, in this regime, we take $h_0 := \log|A|$. Hence $e^{-h_0} = |A|^{-1}$. Then $s := t/k \asymp |A|^{2/k} \gg 1$, by Proposition 3.1.2 and the assumption $k \ll \log|A|$.

As noted in the $d=3$ case, neither the actual value of t nor the fact that W and W' are independent DRWs is of much consequence. Even the particular form of s is not important: it can be changed, subject to changing the conditions on d appropriately.

In the case $d=3$, we looked at (adjacent) pairs of indices $(2i, 2i-1)$. For general d, this is insufficient; instead, we look at m-tuples, where m is a (growing) function of d.

In this regime, the same generator is picked lots of times, with expectation $s = t/k \gg 1$. For the moment, let $\eta \in (0,1)$. For $w \in \mathbb{Z}_+^k$, write

$$\mathcal{C}(w) := \bigl\{ i \in [k] \;\big|\; \eta s \le w_i \le \eta^{-1} s \bigr\}. \tag{3.1.39a}$$

Then, for η sufficiently small (but still a constant), we have

$$\mathbb{P}\bigl(|\mathcal{C}(W)|/k \ge \tfrac{4}{5}\bigr) = 1 - o(1). \tag{3.1.39b}$$

(We could replace $\frac{4}{5}$ by any constant less than 1.) This will form part of our typicality requirements:

$$\mathcal{W} := \bigl\{ w \in \mathbb{Z}_+^k \;\big|\; \mu_t(w) \le e^{-h},\ |\mathcal{C}(w)| \ge \tfrac{4}{5}k,\ \max_i w_i < p \bigr\}. \tag{3.1.40}$$

Note that this definition satisfies (3.1.4). For $i,j \in \mathcal{C}(w)$, as when $d=3$, we have $\{C_{i,j} \equiv C'_{i,j}\} = \{C_{i,j} = C'_{i,j}\}$, since $\max\{C_{i,j}, C'_{i,j}\} \le w_i w_j \lesssim s^2 \asymp p^{2(d-1)/k} \ll p$.

Now recall the partition $(P_3,\dots,P_d)$ of k, and the definition $R_b = |P_k|/k = (b-2)/\binom{d-1}{2}$. Let m be an integer (allowed to depend on other parameters) with $m \ll \min_b |P_b| = k/\binom{d-1}{2} \asymp k/d^2$. By exchangeability of the generators, for each $b \in \{3,\dots,d\}$ assume that the first $\frac{4}{5}|P_b|$ entries i of P_b satisfy $\eta s \le w_i \le \eta^{-1} s$.

Our aim is to show that the mode of the vector $\boldsymbol{C}_m := (C_{i,j})_{i,j\in[m]}$, conditional on $W = w$, which we denote μ_m, is bounded by $s^{-f(m)}$, for some (suitable) super-linearly growing function f, recalling that $s \asymp p^{2(d-1)/k}$. We prove this in Claim 3.1.19 below, and in fact show that we can take $f(m) \asymp m^2$; for now, assume that claim.

Partition $\{1,\dots,\frac{4}{5}|P_b|\}$ into $N := \lfloor \frac{4}{5}|P_b|/m \rfloor \ge \frac{3}{4}|P_b|/m$ sequential intervals of length m, say $I_{1,b},\dots,I_{N,b}$. This allows us to decompose

$$\mathcal{E}_b = \bigl\{C_{i,j} = C'_{i,j}\ \forall\, i,j \in P_b\bigr\} \subseteq \cap_{\ell=1}^N \bigl\{C_{i,j} = C'_{i,j}\ \forall\, i,j \in I_{\ell,b}\bigr\}.$$

Moreover, the events in the intersection are independent. We upper bound each using the mode:

$$q_b(t) \le \mu_m^N \le s^{-f(m)N} \le p^{-(d-1)k^{-1}\cdot f(m)\cdot(3/4)|P_b|/m} \le p^{-dR_b\cdot\frac{1}{2}f(m)/m}, \tag{3.1.41}$$

as $s \asymp p^{2(d-1)/k}$, and so in particular $s \geq p^{(d-1)/k}$ (recall that we had said that the exact value of s would be unimportant); we have $dR_b = (b-2)d/\binom{d-1}{2} \geq 2(b-2)/d$, and so this becomes

$$q_b(t) \leq p^{-(b-2)\cdot d^{-1} f(m)/m}. \tag{3.1.42}$$

Since $f(m)/m \asymp m$, we can choose a constant C large enough so that $m \coloneqq Cd$ satisfies $f(m)/m \geq d$, and hence $q_b(t) \leq 1/p^{b-2}$.

We still need $m \ll k/\binom{d-1}{2} \asymp k/d^2$; since $m \asymp d$, this is equivalent to requiring $d^3 \ll k$. Finally, to apply Claim 3.1.19, we need $\log m \leq \frac{4}{3}\log|A|/k$. But $m \asymp d$ and $k \ll \log|A|$, so this is implied by the condition $\log d \leq \frac{3}{2}\log|A|/k$ from Hypothesis E.

This establishes the desired bound, as it did for $\log|A| \lesssim k \leq (\log|A|)^{1+2/(d-2)}$. □

Claim 3.1.19. *In the notation and under the assumptions of the above proof, there exists an absolute positive constant c so that, assuming that $\log m \leq \frac{4}{3}\log|A|/k$ (so that $m \ll s$), we have*

$$\mu_m \leq s^{-cm^2}.$$

Proof. First, note that $\log m \leq \frac{4}{3}\log|A|/k$ implies that $m \leq |A|^{(4/3)/k} \ll |A|^{2/k} \asymp s$.

For this proof, we use the following notation: for $w \in \mathcal{W}$, write $\mathbb{P}_w(\cdot) \coloneqq \mathbb{P}(\cdot \mid W = W' = w)$ and $\mathbb{E}_w(\cdot)$ similarly; often we consider events that depend on W but not on W', in which case we ignore the conditioning on W' (noting that W and W' are independent). Recall that we write $G_\ell \in [k]$ for the *index* of the generators chosen in the ℓ-th step.

Further, we abuse notation and terminology slightly by always assuming that "pairs (i,j)" have $i \neq j$, and write $[m]^2 = \{(i,j) \mid i \neq j\}$, so $|[m]^2| = m(m-1)$.

Take an arbitrary ordering of all $m(m-1)$ distinct pairs $(i,j) \in [m]^2$; write $K \coloneqq m(m-1)$ and the κ-th term (of the ordering) as y_κ, for $\kappa \in [K]$.

Let $\boldsymbol{x}_m = (x_{i,j})_{i,j\in[m]^2}$, with $x_{i,j} \in \mathbb{N}_0$ for all $(i,j) \in [m]^2$, be arbitrary. We are interested in $\mathbb{P}(\boldsymbol{C}_m = \boldsymbol{x}_m \mid W = w)$; cf (3.1.24). We do this by sequentially estimating the conditional probabilities that $C_{i,j} = x_{i,j}$. For each $\kappa \in [K]$, let $\chi_\kappa \coloneqq \mathbf{1}(C_{y_\kappa} = x_{y_\kappa})$, recalling that $(y_1, ..., y_K)$ is the chosen ordering of all pairs $(i,j) \in [m]^2$. Then we want to bound $\mathbb{E}_w(\chi_1 \cdots \chi_K)$.

To do this, we use the following general bound, which is an immediate consequence of the tower property for conditional expectation: for random variables $V_1, ..., V_K$, we have

$$\begin{aligned}\mathbb{E}\big(f_1(V_1,...,V_K)\cdots f_K(V_1,...,V_K)\big) &\leq \max_{v_1,...,v_K} \mathbb{E}\big(f_1(V_1,v_2,...,v_K)\cdots f_K(v_1,...,v_{K-1},V_K)\big)\\ &= \max_{v_1,...,v_K} \mathbb{E}\Big(\textstyle\prod_{\kappa=1}^K \mathbb{E}\big(f_\kappa(v_1,...,v_{\kappa-1},V_\kappa,v_{\kappa+1},...,v_K \mid V_1,...,V_{\kappa-1}\big)\Big).\end{aligned}$$

The following argument is analogous to that given in the $d = 3$ case; see the end of the proof of Lemma 3.1.15. Let $r \coloneqq \sum_1^m w_i$, and write our random word as $S = Z_{G_1} \cdots Z_{G_N}$; here $N = \sum_1^k w_i$ is the number of steps taken and G_ℓ is the generator index chosen in the ℓ-th step. Let $J_1 < \cdots < J_r$ be the (random) indices with $G_{J_\ell} \in [m]$. Now define the vector $I \in \{1, ..., m\}^r$ by $I_\ell \coloneqq G_{J_\ell}$. Thus I encodes the relative order between the different occurrences (with multiplicities) of the generators labelled by $[m]$ in the word S. By typicality, $mn\eta s \leq r \leq mn\eta^{-1}s$

Sequentially, and without replacements, for each pair (i,j), or index κ, choose a uniformly random element of $\{\ell \mid G_\ell = j\}$; call this $U_{i,j}$, or U_κ. (This can be done since $|\{\ell \mid G_\ell = j\}| \gtrsim s \gg m$ by assumption.) For each pair (i,j), let $V_{i,j}$ be the location in I of the random element $U_{i,j}$. Now define the vector I' from I by omitting the $K = m(m-1)$ locations $\{V_{i,j}\}_{i,j\in[m]^2}$; so $I' \in \{1, ..., m\}^{r-K}$. Importantly, we are omitting elements of the *relative* order, not the *absolute* order.

By definition of $C_{i,j}$, given in (3.1.10), given $W = w$, the value of $C_{i,j}$ is a function only of the relative locations I. Hence, given I' also, it is a function only of the location of the omitted j. It is constant on the set of locations which give rise to the same relative order between choices of i and j: two different placements of the omitted j give rise to the same relative order between choices of i and j if and only if they both lie in the same (possibly empty) interval in which there are no i-s; this is analogous to the case $d = 3$. (Recall that for the pair (i,j) we omitted the location of a j.)

Recall that the random variables $\{V_\kappa\}$ are drawn uniformly at random without replacement from $[r]$; hence the distribution of V_κ given $V_1, ..., V_{\kappa-1}$ is uniform on $[r] \setminus \{V_1, ..., V_{\kappa-1}\}$. Hence,

writing $L_i := L_i(I')$ for the longest interval in I' without an i in it, we obtain

$$\mathbb{E}_w\big(\chi_\kappa(v_1,...,v_{\kappa-1},V_\kappa,v_{\kappa+1},...,v_K)\;\big|\;V_1,...,V_{\kappa-1},\;I'\big)\le\big(L_\kappa+m(m-1)+1\big)/(r-\kappa+1).$$

Hence, applying the above formula and noting that $m(m-1)+1\le m^2$, we obtain

$$\mathbb{E}_w\big(\chi_1\cdots\chi_K\;\big|\;I'\big)\le\frac{\prod_1^m(L_i+m^2)^{m-1}}{r(r-1)\cdots(r-m(m-1)+1)}.$$

We now average over I'. To bound this expectation, we use the generalisation of Hölder's inequality to the product of m variables: for non-negative random variables $X_1,...,X_m$, we have

$$\mathbb{E}\big(X_1\cdots X_m\big)\le\big(\mathbb{E}\big(X_1^m\big)\cdots\mathbb{E}\big(X_m^m\big)\big)^{1/m}.$$

In our application of this, we take $X_i=(L_i+m^2)^{m-1}$. This gives

$$\mathbb{E}_w\big(\textstyle\prod_1^m(L_i+m^2)^{m-1}\big)\le\max_i\mathbb{E}_w\big((L_i+m^2)^{m(m-1)}\big).$$

Since $r\ge m\eta s$ and $s\gg m$, we have $m^2\le\frac12 r$, and so the denominator is at least $2^{-m^2}r^{-m(m-1)}$; recall also that $r\ge m\eta s$. In summary, we have proved that

$$\mathbb{E}_w\big(\chi_1\cdots\chi_K\big)\le 2^m(2\eta^{-1}m^{-1})^{m(m-1)}s^{-m(m-1)}\max_i\mathbb{E}_w\big((L_i+m^2)^{m^2}\big).\tag{3.1.43}$$

Recall also that $s\asymp p^{2(d-1)/k}$. It remains to bound this latter expectation.

To do this, recall Claim 3.1.16, which, for given w, states that

$$L_i/(\eta^{-2}m\log r)\preceq\mathrm{Geom}(\tfrac12).$$

Using the the inequality $(a+b)^\ell\le 2^{\ell-1}(a^\ell+b^\ell)$, valid for $a,b\ge0$ and $\ell\in\mathbb{N}$, we obtain

$$\big(L_i+m^2\big)^{m^2}\le\big(2\eta^{-2}m\log r\big)^{m^2}\big(L_i/(\eta^{-2}m\log r\big)^{m^2}+(2m^2)^{m^2}.$$

If $X\sim\mathrm{Geom}(\frac12)$, then one can show, for $\ell\ge3$, that $\mathbb{E}(X^\ell)\le\ell^\ell$. (This follows by comparison with the exponential-$(\log 2)$ distribution.) We apply this with $X:=L_i/(\eta^{-2}m\log r)$ and $\ell:=m^2$:

$$\mathbb{E}_w\big((L_i+m^2)^{m^2}\big)\le\big(2\eta^{-2}m\log r\big)^{m^2}\cdot(m^2)^{m^2}+(4m^2)^{m^2}\le 2\big(2\eta^{-2}m^3\log r\big)^{m^2}.$$

Plugging this back into (3.1.43), we obtain

$$-\log\mathbb{E}_w\big(\chi_1\cdots\chi_K\big)\asymp m^2\log s+m^2\log m+m^2\log r.$$

Recalling that $r\asymp ms$ and $\log s\asymp\log|A|/k$, we obtain

$$-\log\mathbb{E}_w\big(\chi_1\cdots\chi_K\big)\asymp m^2\log s+\log m\asymp m^2\big(\log|A|/k+\log m\big)\asymp m^2\log|A|/k,$$

using the condition $\log m\le\frac43\log|A|/k$. Recall that we desire

$$\mathbb{E}_w\big(\chi_1\cdots\chi_K\big)\le s^{-f(m)}\le|A|^{-f(m)/k},$$

as $s\approx|A|^{2/k}/(2\pi e)\ge|A|$, where f is some super-linearly growing function; this holds if

$$-\log\mathbb{E}\big(\chi_1\cdots\chi_K\;\big|\;W=w\big)\ge f(m)\log|A|/k.$$

We hence see that this is satisfied for some f with $f(m)\asymp m^2$. □

3.1.9 Extensions

In this subsection we describe some extensions to the argument.

§3.1.9.1 We consider undirected Cayley graphs.
§3.1.9.2 We describe the limit profile in the regime $1\ll k\ll\log|H_{p,d}^{\mathrm{ab}}|$.
§3.1.9.3 We relax that condition that p (in $H_{p,d}$) is prime when $k\gtrsim\log|H_{p,d}|$.

3.1.9.1 Undirected Cayley Graphs

Throughout the paper we have always been assuming that the graph is directed. Here we describe the required adaptations to consider *undirected* Cayley graphs, rather than *directed*. We still use an auxiliary process W to generate the walk; it is now a SRW, rather than a DRW, on $\mathbb{Z}^k$.

Recall that in the directed case the mixing time was the maximum of the time t_0^+ at which the entropy of W, ie a DRW on $\mathbb{Z}^k$, reaches $\log|H_{p,d}^{\mathrm{ab}}|$ and the diameter-based bound of $\log_k|H_{p,d}|$; see Definition C. The undirected case is completely analogous: the mixing time is the maximum of the time t_0^- at which the entropy of W, now a SRW on $\mathbb{Z}^k$, reaches $\log|H_{p,d}^{\mathrm{ab}}|$ and the diameter-based lower bound of $\log_k|H_{p,d}|$. (The directed graphs are k-regular, while the undirected graphs are $2k$-regular; but $\log(2k) \approx \log k$, and so $\log_{2k}|H_{p,d}| \approx \log_k|H_{p,d}|$, when $k \gg 1$.)

We still use exactly the same outline, namely we use a modified L_2 calculation and 'typicality'.

Adaptation from DRW to SRW. The primary difference comes from our expression for $S(t)$ given the generators chosen: for the DRW there were no inverses. However, Lemma 3.1.17 had basically the same form whether the inverses were included or not; the 'remainder polynomial' g is different in the two cases, as is the expression for $C_{i,i}$, but neither of these forms were needed for our proof. Recall that, in fact, $C_{i,j}$ has the same form for the DRW and SRW for $i \neq j$. The SRW generalises the DRW in some sense: if no inverses are applied, then the formulas for the SRW are exactly those for the DRW. (In the terminology of Lemma 3.1.17 where the $C_{i,j}$ are defined, the DRW corresponds to the 'signs' $\sigma_\ell := 1$ for all ℓ.) Recall also that the entropic results of §3.1.1 are stated for both the DRW and the SRW; further, the statements in these two cases are analogous.

The remaining adaptations are relatively simple. This is one of the strengths of our entropic method. Importantly, note that $V := W - W'$ is a rate-2 SRW on $\mathbb{Z}^k$ in both the undirected and the directed cases, where W' is an independent copy of the auxiliary process W. The definition of $C_{i,j}$ for $i \neq j$ for the SRW is almost the same as for the DRW, except that now when an inverse is chosen 1 is subtracted from the count, rather than added.

For the DRW, conditioning on $(W(t), W'(t))$ told us exactly how many times each generator had been applied. Now if a generator–inverse pair is applied at some point, this is not seen by $(W(t), W'(t))$. Instead we define $W_{i,+}(t)$ to be the number of times that generator Z_i is applied (normally) and $W_{i,-}(t)$ to be the number of times that the inverse Z_i is applied; then $W_i(t) = W_{i,+}(t) - W_{i,-}(t)$. Write $W_+(t) := (W_{i,+}(t))_{i\in[k]}$ and define $W_-(t)$, $W'_+(t)$ and $W'_-(t)$ similarly.

Using this construction, when considering $\mathbb{P}(S = S' \mid W(t) = W'(t), \mathsf{typ})$, rather than conditioning on $W(t) = w = W'(t)$, for some typical $w \in \mathcal{W}$, we condition on

$$\big(W_+(t), W_-(t)\big) = (w_+, w_-) \quad \text{and} \quad \big(W'_+(t), W'_-(t)\big) = (w'_+, w'_-) \quad \text{with} \quad w_+ - w_- = w'_+ - w'_-.$$

We need these to be 'typical' in the appropriate sense. Recall that throughout the proof there have been two types of typicality: first, the entropic part, requiring $\mu_t(w) \le e^{-h}$; second, some technical requirements; call these sets $\mathcal{W}_1$ and $\mathcal{W}_2$, respectively, so $\mathcal{W} = \mathcal{W}_1 \cap \mathcal{W}_2$. The main part of the technical requirements asks for $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$ that approximately the correct number of generators precisely once, while for $1 \ll k \ll \log|H_{p,d}^{\mathrm{ab}}|$ it asks that most generators are picked within some constant factor of the expected number of times. It is the entropic part which was the key part; see also (3.1.4). For the entropy, we are interested in the law of $W(t)$. As such, we require $w := w_+ - w_-$ and $w' := w'_+ - w'_-$ to be in $\mathcal{W}_1$. For the technical requirements, we ask that $w_+, w_-, w'_+, w'_- \in \mathcal{W}_2$.

The only other place where the proof needs to be adapted is in Claim 3.1.19, namely the analysis of $\mathbb{P}(S(t) = S'(t) \mid W(t) = W'(t), \mathsf{typ})$ in the regime $1 \ll k \ll \log|H_{p,d}^{\mathrm{ab}}|$. There we considered the relative order of the choices of generators the multiset $[Z_i, Z_j]$; we redacted one Z_j from this partial order and then considered the longest interval without Z_i. Here we consider the relative order of the choices of generators $[Z_i, Z_i^{-1}, Z_j, Z_j^{-1}]$; we redact one Z_j or Z_j^{-1} and then consider the longest interval with neither Z_i nor Z_i^{-1}. This analysis is completely analogous. □

3.1.9.2 Cutoff Window

We determine the limit profile in the regime $k \ll \log|H_{p,d}^{\mathrm{ab}}|$. Here the mixing time is $t_0(k, |H_{p,d}^{\mathrm{ab}}|)$.

Theorem D.1. *Let p be prime and $d \geq 3$ a fixed constant. Suppose that $1 \ll k \ll \log |H_{p,d}^{\mathrm{ab}}| \asymp \log p$. There exist times $(t_\alpha)_{\alpha \in \mathbb{R}}$ satisfying*

$$t_0 \approx k \cdot \tfrac{1}{2\pi e} |H_{p,d}^{\mathrm{ab}}|^{2/k}, \quad t_\alpha - t_0 \approx \alpha \sqrt{2} t_0 / \sqrt{k} = o(t_0) \quad \textit{and} \quad d^{\pm}_{(H_{p,d})_k}(t_\alpha) \approx \Psi(\alpha) \textit{ whp.}$$

Proof. Recall that t_0 is determined by the entropy of W, which was the expectation of the random variable Q; see Definition 3.1.1. For $\alpha \in \mathbb{R}$, we now define $t_\alpha(k, N)$ according to the variations of Q:

$$\mathbb{E}\big(Q_1(t_\alpha(k,N))\big) = \big(\log N + \alpha \sqrt{vk}\big)/k \quad \text{where} \quad v \coloneqq \mathbb{V}\mathrm{ar}\big(Q_1(t_0(k,N))\big).$$

Analogously to before, we consider the entropic time $t_\alpha \coloneqq t_\alpha(k, |H_{p,d}^{\mathrm{ab}}|)$. We show in the supplementary material, in §6.1, that, for all $\alpha \in \mathbb{R}$, if $1 \ll \omega \ll \sqrt{vk}$ and $1 \ll k \ll \log |H_{p,d}^{\mathrm{ab}}|$, then

$$t_0 \approx k |H_{p,d}^{\mathrm{ab}}|^{2/k}, \quad t_\alpha - t_0 \approx \alpha \sqrt{2} t_0 / \sqrt{k} = o(t_0) \quad \text{and} \quad \mathbb{P}\big(Q(t_\alpha) \leq \log |H_{p,d}^{\mathrm{ab}}| \pm \omega\big) \approx \Psi(\alpha),$$

where Ψ is the standard Gaussian tail; see Propositions 6.1.2 and 6.1.3.

Recall from the analysis of the total variation distance *given typicality*, from §3.1.8, in the regime $k \ll |H_{p,d}^{\mathrm{ab}}|$, that the particular value of t_0 is unimportant—changing it by a constant would not affect the proof or the result. The above says that $t_\alpha \approx t_0$ for all $\alpha \in \mathbb{R}$. Hence this contribution is $o(1)$ when t_0 is replaced by t_α, regardless of $\alpha \in \mathbb{R}$. All that changes is Lemma 3.1.12: now

$$\mathbb{P}\big(\mu_{t_\alpha}\big(W(t_\alpha)\big) \leq e^{-h}\big) = \mathbb{P}\big(Q(t_\alpha) \geq \log |H_{p,d}^{\mathrm{ab}}| \pm \omega\big) \approx 1 - \Psi(\alpha).$$

Hence exactly the same argument gives $d_{(H_{p,d})_k}(t_\alpha) \approx \Psi(\alpha)$ whp over Z. □

3.1.9.3 Lifting the Primality Condition for p when $k \gtrsim \log |H_{p,d}^{\mathrm{ab}}|$

Throughout the paper we have always been assuming that p is prime. Here we describe how to remove this assumption under the condition that $k \gtrsim \log |H_{p,d}^{\mathrm{ab}}|$ and d is constant; we also define a density-1 set $\mathbb{A} \subseteq \mathbb{N}$ so that if $|G| \in \mathbb{A}$ then we can allow d to diverge. To emphasise the lack of primality, we consider $H_{m,d}$ with $m, d \in \mathbb{N}$ (ie replace the letter p by the letter m).

Theorem D.2. *Let $m, d \in \mathbb{N}$ with $d \geq 3$. Suppose that $k \gtrsim \log |H_{m,d}^{\mathrm{ab}}|$. If d is a constant or diverges sufficiently slowly, then the RW on $(H_{m,d})_k^{\pm}$ exhibits cutoff whp. Further, there is a density-1 set $\mathbb{A} \subseteq \mathbb{N}$ so that if $m \in \mathbb{A}$ and $\log d \ll \log \log m$ then the RW on $(H_{m,d})_k^{\pm}$ exhibits cutoff whp.*

Lower Bound. First note that the lower bound holds easily: all it required was that the walk projected to the Abelianisation was not mixed. The lower bound for mixing on an Abelian group is valid for any Abelian group and any choice of generators. □

We now turn to the upper bound. We use the same definition of typicality as when we studied $m = p$ prime with a directed graph. Let W' be an independent copy of W and define S' via W'; write $V \coloneqq W - W'$. We are interested in the probability that $S = S'$ when W and W' are typical.

Analysis when $W(t) \neq W'(t)$. When $m = p$ is prime, for any $v \in \mathbb{Z}_p^k \setminus \{0\}$, each of $\sum_i A_i v_i$, $\sum_i B_i v_i$ and $\sum_i C_i v_i$ was an independent $\mathrm{Unif}(\mathbb{Z}_p)$ random variable, since $\{A_i, B_i, C_i\}_1^k$ is an independent collection of $\mathrm{Unif}(\mathbb{Z}_p)$-s. As described above, in the (a,b)-th coordinate there is a sum $\sum_i Z_i(a,b) W_i(t)$, independent of the other terms in the coordinate. Hence, as in (3.1.26), we have

$$\mathbb{P}\big(S = S' \mid W \neq W', \, \mathsf{typ}\big) = 1/|H_{m,d}| = 1/m^{d(d-1)/2}.$$

This is not so when m is composite. Instead, they are uniform on the subset $\mathfrak{g}_v \mathbb{Z}_m$, where $\mathfrak{g}_v \coloneqq \gcd(v_1, \ldots, v_k, m)$; this is proved in Lemma 6.6.1. Since in the (a,b)-th coordinate there is a sum $\sum_i Z_i(a,b) W_i(t)$, independent of the other terms in the coordinate, noting that there are $\frac{1}{2} d(d-1)$ non-trivial coordinates in the matrix, exactly the same arguments show that

$$\mathbb{P}\big(S = S' \mid W = w, \, W' = w'\big) \leq (\mathfrak{g}_{w - w'}/m)^{d(d-1)/2} \quad \text{for any} \quad w, w' \in \mathbb{Z}^k.$$

We then need to take expectation of (W, W') with $W \neq W'$. We now argue that

$$\mathbb{E}\big(\mathfrak{g}_{W-W'}^{d(d-1)/2} \mid W = W',\, \mathsf{typ}\big) = 1 + o(1).$$

An analysis of such a gcd-expectation is carried out in §5.1.5, where Abelian groups are considered; the application here is for the Abelian group $\mathbb{Z}_m^{d(d-1)/2}$, so n there should be replaced with $m^{d(d-1)/2}$. The gcd is taken with respect to n, not m; this only increases the right-hand side.

We now take expectation over (W, W') with $W \neq W'$. In §5.1.5, there are conditions on (k, d, n); see Hypothesis I. Since $\log d \ll \log\log m$, it is straightforward to see that $m \gg \log\log|H_{m,d}|$. Recall that we can assume that $\log k \lesssim \log\log|H_{m,d}|$, otherwise cutoff is already established at time $\log_k|H_{m,d}|$. Thus the conditions are satisfied.

The conclusion of §5.1.5—namely that the expectation of the power of the gcd is $1 + o(1)$; see specifically Proposition 5.1.10—thus follows. In summary,

$$\mathbb{P}\big(S = S' \mid W \neq W',\, \mathsf{typ}\big) \leq \big(1 + o(1)\big)/n. \tag{3.1.44}$$

Analysis when $W(t) = W'(t)$. In the regime $k \gtrsim \log|H_{m,d}^{\mathrm{ab}}|$, part of the typicality conditions were that a large number of generators are picked at most once; see (3.1.35). When calculating the probability that $C_{i,j} = C'_{i,j}$ for all $i, j \in [k]$, we restricted ourselves to just looking at those (i, j) with $C_{i,j}, C'_{i,j} \in \{0, 1\}$; see (3.1.18, 3.1.19). (Now we have a sign for $C_{i,j}$, so look at (i, j) with $C_{i,j}, C'_{i,j} \in \{0, \pm 1\}$.) As such, we may replace the event

$$\mathcal{E} := \{\boldsymbol{C} = \boldsymbol{C}'\} \quad \text{where} \quad \boldsymbol{C} := (C_{i,j}) \quad \text{and} \quad \boldsymbol{C}' := (C'_{i,j})$$

from (3.1.11) with the event

$$\mathcal{E}' := \big\{\exists\,(i,j) \text{ st } |D_{i,j}| = 1\big\} \quad \text{where} \quad D_{i,j} := C_{i,j} - C'_{i,j} \quad \text{for all} \quad i, j \in [k].$$

Herein assume that there exists (i, j) with $|C_{i,j}| = 1$. By Lemma 3.1.17, we need to control AB for $A, B \sim^{\mathrm{iid}} \mathrm{Unif}(\mathbb{Z}_m)$, representing the $Z_{\gamma_i}(a, a+1)$ and $Z_{\gamma_j}(a+1, b)$, respectively. (When $m = p$ was prime, it did not matter whether or not $|C_{i,j}| = 1$, just that $C_{i,j} \not\equiv 0 \bmod p$.) Then the law of AB conditional on $\gcd(B, m) =: \mathfrak{g}$ is uniform on $\mathfrak{g}\mathbb{Z}_m$. Hence

$$\max_{x \in \mathbb{Z}_m} \mathbb{P}\big(AB \equiv x \bmod m\big) = \max_{x \in \mathbb{Z}_m} \mathbb{E}\big(\mathbb{P}\big(AB \equiv x \bmod m \mid \mathfrak{g}\big)\big) \leq \mathbb{E}\big(\mathfrak{g}\big)/m.$$

It remains to bound this expectation. For $\alpha, \beta \in \mathbb{N}$, write $\alpha \wr \beta$ to indicate that α divides β. We have

$$\mathbb{E}(\mathfrak{g}) = \textstyle\sum_{r \in [m]} r\, \mathbb{P}(\mathfrak{g} = r) \leq \sum_{r \in [m]} r\, \mathbb{P}(r \wr B) \mathbf{1}(r \wr m) = \sum_{r \in [m]} \mathbf{1}(r \wr m) =: \operatorname{div} m,$$

ie the number of divisors of m. From this we deduce the analogue of (3.1.30):

$$\mathbb{P}\big(S_{a,b} = S'_{a,b} \,\big|\, W = W',\, \exists\,(i,j) \text{ st } |D_{i,j}| = 1\big) \leq (\operatorname{div} m)/m,$$

where (a, b) is an element of the matrix above the super-diagonal, ie $a \leq b - 2$.

Consider first $d = 3$. Here there is only one such element (a, b), namely $(1, 3)$. Thus

$$\mathbb{P}\big(S = S' \,\big|\, W = W',\, \exists\,(i,j) \text{ st } |D_{i,j}| = 1\big) \leq (\operatorname{div} m)/m,$$

which is the direct analogue of (3.1.12). When $d = 3$, we trivially bound $\mathbb{P}(S = S' \mid W = W', \mathcal{E}') \leq 1$. This leads us to the analogue of (3.1.17), still for $d = 3$:

$$\begin{gathered}\mathbb{P}\big(S = S',\, W = W' \mid \mathsf{typ}\big) \leq (\operatorname{div} m)\, e^{-h}\big(1/m + q\big)/\mathbb{P}(\mathsf{typ}) \\ \text{where} \quad q := \max_{w \in \mathcal{W}} \mathbb{P}\big(\mathcal{E}' \mid W = W' = w\big)\end{gathered}$$

and h is the entropy at the time t. Lemma 3.1.15 controls this last upper bound. An easy inspection of the proof shows that when $k \gtrsim \log|H_{m,d}^{\mathrm{ab}}|$, the $o(1)$ term in Lemma 3.1.15 is $\mathcal{O}(e^{-\omega})$. Hence

$$|H_{m,d}|\, \mathbb{P}\big(S = S',\, W = W' \mid \mathsf{typ}\big) \lesssim e^{-\omega} \cdot \operatorname{div} m.$$

We may take any $1 \ll \omega \ll \min\{k, \log|H^{\mathrm{ab}}_{m,d}|\} \asymp \log|H_{m,d}|$ (as $d = 3$ here); see Lemmas 3.1.9 and 3.1.12 and Definition 3.1.11 for entropic calculations. It is well-known that $\operatorname{div} r \le r^{\mathcal{O}(1/\log\log r)}$ for all $r \in \mathbb{N}$; see, eg, [40, §18.1]. So here $\operatorname{div} m \le \operatorname{div}|H_{m,d}| = |H_{m,d}|^{o(1)}$, as $m \wr |H_{m,d}|$. Choosing ω sufficiently close to $\log|H_{m,d}|$, we obtain $e^{-\omega}\operatorname{div} m = o(1)$. In summary,

$$|H_{m,d}|\,\mathbb{P}\big(S = S',\, W = W' \mid \mathsf{typ}\big) = o(1). \tag{3.1.45a}$$

Consider now general d. We use the partitioning argument to deduce an analogue of (3.1.33):

$$|H_{m,d}|\,\mathbb{P}\big(S = S',\, W = W' \mid \mathsf{typ}\big) \lesssim e^{-\omega}\cdot(\operatorname{div} m)^{d^2}.$$

Using the bound $\operatorname{div} r \le r^{\mathcal{O}(1/\log\log r)}$ for all $r \in \mathbb{N}$, if $d = o(\log\log|H_{m,d}|)$, ie $d \ll \log\log m$, then the same upper bound as for $d = 3$ holds for the number of divisors, namely $(\operatorname{div} m)^{d^2} = n^{o(1)}$. Hence, with exactly the same justification as for (3.1.45a), we deduce for such d that

$$|H_{m,d}|\,\mathbb{P}\big(S = S',\, W = W' \mid \mathsf{typ}\big) = o(1). \tag{3.1.45b}$$

Conclusion for Fixed or Slowly Diverging d. Combining (3.1.44, 3.1.45), we deduce that

$$|H_{m,d}|\,\mathbb{P}\big(S = S' \mid \mathsf{typ}\big) - 1 = o(1).$$

Combined with the modified L_2 calculation Lemma 3.1.8, we deduce the theorems for such d. □

Adaptations for Diverging d. Now consider the part of the theorem where d is allowed to diverge with $\log d \ll \log\log m$. In [40, §18.2] it is shown that if $U \sim \mathrm{Unif}([m])$ then $\mathbb{E}(\operatorname{div} U)/(m\log m) \to 1$ as $m \to \infty$. In particular, $\mathbb{P}(\operatorname{div} U \le (\log m)^2) \to 1$ as $m \to \infty$. Define

$$\mathbb{A} := \big\{m \in \mathbb{N} \;\big|\; \operatorname{div} m \le (\log m)^2\big\}.$$

Then $\mathbb{A}$ has density 1 in $\mathbb{N}$. Assume that $m \in \mathbb{A}$. For such m, we have

$$(\operatorname{div} m)^L \le (\log m)^{2\cdot d^2/2} = (\log m)^{d^2} = e^{d^2\log\log m}.$$

As above, we require $\omega \ll \min\{k, \log|H^{\mathrm{ab}}_{m,d}|\} \asymp \log|H^{\mathrm{ab}}_{m,d}| \asymp d\log m$; we desire $\omega \gg d^2\log\log m$. Thus $d^2 \ll \log m/\log\log m$ suffices. But we already required $\log d \ll \log\log m$ in order to apply Lemma 3.1.18 (see Hypothesis E), and this implies that $d^2 = (\log m)^{o(1)} \ll \log m/\log\log m$. □

3.2 Typical Distance and Diameter

This section focuses on distances from a fixed point in the directed random Cayley graph of a Heisenberg matrix group $H_{p,d}$, with p prime and $d \ge 3$. Recall the definition of *typical distance*: when $G := H_{p,d}$ and there are k generators, for $R \ge 0$ and $\beta \in (0,1)$, write

$$\mathcal{B}_k(R) := \big\{x \in G \;\big|\; \mathrm{dist}_{G_k}(\mathsf{id}, x) \le R\big\} \quad\text{and}\quad \mathcal{D}_k(\beta) := \min\big\{R \ge 0 \;\big|\; |\mathcal{B}_k(R)| \ge \beta|G|\big\},$$

emphasising explicitly the dependence on d for the latter statistic.

3.2.1 Precise Statement and Remarks

In this section, we state the more refined version of Theorem J. Again, there are some simple conditions that the parameters must satisfy.

We now state the main result of this section; it is in essence a restatement of Theorem J.

Theorem 3.2.1a (Typical Distance for $1 \ll k \ll \log|H^{\mathrm{ab}}_{p,d}|$)**.** *Let $k, d, p \in \mathbb{N}$. Write*

$$M_k^+ := k|H^{\mathrm{ab}}_{p,d}|^{1/k}/e \quad\text{and}\quad M_k^- := k|H^{\mathrm{ab}}_{p,d}|^{1/k}/(2e); \quad\text{recall that}\quad |H^{\mathrm{ab}}_{p,d}| = p^{d-1}.$$

Suppose that (k, d, p) jointly satisfy the following conditions:

$$1 \ll k \ll \log|H_{p,d}^{\mathrm{ab}}|, \quad k \le \tfrac{1}{2}\log|H_{p,d}^{\mathrm{ab}}|/\log d \quad \text{and} \quad d \ll \max\{\log k,\ k^{1/2}/|H_{p,d}^{\mathrm{ab}}|^{1/(4k)}\}.$$

For all constants $\beta \in (0,1)$, we have

$$\mathcal{D}^{\pm}_{(H_{p,d})_k}(\beta)/M_k^{\pm} \to^{\mathbb{P}} 1 \quad \textit{(in probability)} \quad \text{as } N \to \infty.$$

Moreover, the implicit lower bound holds deterministically, ie for all choices of generators, and for all Heisenberg groups, requiring only $1 \ll k \ll \log|H_{p,d}^{\mathrm{ab}}|$.

Theorem 3.2.1b (Typical Distance for $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$)**.** *Let (k,p,d) be integers with p prime and $d \ge 3$. Suppose that $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$ and $\log k \ll \log|H_{p,d}|$; suppose also that $\log d \ll \log\log p$. Write*

$$M_k := \tfrac{\rho}{\rho-1}\log_k|H_{p,d}^{\mathrm{ab}}| \quad \textit{where} \quad \rho := \log k/\log\log|H_{p,d}^{\mathrm{ab}}|, \quad \textit{ie} \quad k = \big(\log|H_{p,d}^{\mathrm{ab}}|\big)^{\rho}.$$

For all $\lambda \in (0,\infty)$, there exists a constant $\alpha_\lambda^{\pm} \in (0,\infty)$ so that, for all constants $\beta \in (0,1)$, the following convergences in probability hold:

$$\mathcal{D}^{\pm}_{(H_{p,d})_k}(\beta)/\big(\alpha_\lambda^{\pm}k\big) \to^{\mathbb{P}} 1 \quad \textit{if} \quad k \asymp \lambda\log|H_{p,d}^{\mathrm{ab}}|;$$
$$\mathcal{D}^{\pm}_{(H_{p,d})_k}(\beta)/\max\big\{M_k,\ \log_k|H_{p,d}|\big\} \to^{\mathbb{P}} 1 \quad \textit{if} \quad k \gg \log|H_{p,d}^{\mathrm{ab}}|.$$

Moreover, the implicit lower bound holds deterministically, ie for all choices of generators, and for all Heisenberg groups, requiring only $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$ and $\log k \ll \log|H_{p,d}|$. Note that

$$\max\big\{M_k,\ \log_k|H_{p,d}|\big\} = \max\big\{\tfrac{\rho}{\rho-1},\ \tfrac{1}{2}d\big\}\log|H_{p,d}^{\mathrm{ab}}|.$$

We also state the result on the diameter; it is a restatement of Theorem K.

Theorem 3.2.2 (Diameter for $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$)**.** *Let (k,p,d) be integers with p prime and $d \ge 3$. Suppose that $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$ and $\log k \ll \log|H_{p,d}|$; suppose also that $\log d \ll \log\log p$. Write*

$$M_k := \tfrac{\rho}{\rho-1}\log_k|H_{p,d}^{\mathrm{ab}}| \quad \textit{where} \quad \rho := \log k/\log\log|H_{p,d}^{\mathrm{ab}}|, \quad \textit{ie} \quad k = \big(\log|H_{p,d}^{\mathrm{ab}}|\big)^{\rho}.$$

For all $\lambda \in (0,\infty)$, let $\alpha_\lambda^{\pm}$ be the constant from Theorem 3.2.1b.

For all $\lambda \in (0,\infty)$, the following convergences in probability hold:

$$\big(\mathrm{diam}\,(H_{p,d})_k\big)/\big(\alpha_\lambda^{\pm}k\big) \to^{\mathbb{P}} 1 \quad \textit{if} \quad k \asymp \lambda\log|H_{p,d}^{\mathrm{ab}}|;$$
$$\big(\mathrm{diam}\,(H_{p,d})_k\big)/\max\big\{M_k,\ \log_k|H_{p,d}|\big\} \to^{\mathbb{P}} 1 \quad \textit{if} \quad k \gg \log|H_{p,d}^{\mathrm{ab}}|.$$

Moreover, the implicit lower bound holds deterministically, ie for all choices of generators, and for all Heisenberg groups, requiring only $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$ and $\log k \ll \log|H_{p,d}|$. Note that

$$\max\big\{M_k,\ \log_k|H_{p,d}|\big\} = \max\big\{\tfrac{\rho}{\rho-1},\ \tfrac{1}{2}d\big\}\log_k|H_{p,d}^{\mathrm{ab}}|.$$

As a proxy for the size of the (L_1) balls in the (directed) Cayley graph with k generators, denoted $\mathcal{B}_k(\cdot)$, we use the size of discrete, directed L_1 balls in dimension k, denoted $B_k(\cdot)$: for $R \ge 0$, define $B_k(R) := \{x \in \mathbb{Z}_+^k \mid \mathrm{dist}_{\mathbb{Z}_+^k}(0,x) \le R\}$. This is done in Lemma 3.2.4 below.

Were the underlying group Abelian, we would have the easy inequality $|\mathcal{B}_k(R)| \le |B_k(R)|$. For the Heisenberg group $H_{p,d}$, we develop a similar inequality; roughly, we use the inequality for the Abelianisation and upper bound the number of elements which can be seen by the other vertices by the maximum amount, ie $|H_{p,d}|/|H_{p,d}^{\mathrm{ab}}|$. In §4.1, we studied typical distance for general Abelian groups, using the same (overall) method; there, the radius R of the balls in question was chosen so that $|B_k(R)| \approx |G|$. Here our candidate radius M_k^* satisfies $|B_k(M_k^*)| \gg |H_{p,d}^{\mathrm{ab}}|$.

Definition 3.2.3. *Set $\omega := \max\{(\log k)^2,\ k/|H_{p,d}^{\mathrm{ab}}|^{1/(2k)}\} = \max\{(\log k)^2,\ k/p^{(d-1)/(2k)}\}$. Choose M_k^* to be the minimal integer satisfying $|B_k(M_k^*)| \ge e^{\omega}|H_{p,d}^{\mathrm{ab}}|$.*

For the sake or presentation, we first analyse in §3.2.2–§3.2.4 typical distance for directed graphs with $1 \ll k \ll \log|H_{p,d}^{\mathrm{ab}}|$. We then describe in §3.2.5 the requisite extensions to the argument to handle undirected graphs and $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$ for both typical distance and diameter.

3.2.2 Outline of Proof

As remarked after the summarised statement (in §1.3.3.2), when considering the mixing time on a graph, geometric properties of the graph are often derived and used. In a reversal of this, we use knowledge about the mixing properties of the random walk to derive a geometric result; the style of proof is similar enough that we even quote lemmas from the mixing section.

The main difference between the proofs is the following: previously, $W(\cdot)$ was a DRW on $\mathbb{Z}_+^k$; we replace this $W(t)$ by A which is uniformly distributed on a $\mathbb{Z}_+^k$-ball of radius R, for R defined later; this A tells us how many times each generator is used; we apply the sequence of generators, with multiplicities, in an order chosen uniformly at random; call the resulting element S.

We choose M so that this ball has size slightly larger than $|H_{p,d}^{\mathrm{ab}}|$—recall that this size was used for the entropic time $t_0(k, |H_{p,d}^{\mathrm{ab}}|)$ in the mixing. For a constant $\xi > 0$, if $R := M(1-\xi)$, then we use a counting argument to show that the ball cannot cover more than a proportion $o(1)$ of the vertices of the graph; hence this gives a *deterministic* lower bound, valid for all Z. For a constant $\xi > 0$, if $R := M(1+\xi)$, then we show that not only does the ball cover (almost) all the graph, but the random variable S is well-mixed whp, in the sense that it is very close to the uniform distribution. From this we deduce that, for a proportion $1-o(1)$ of the vertices, there is a non-zero probability that S is at that vertex, and hence a path to it must exist; furthermore, by choice of A, the path must have length at most $R = M(1+\xi)$. To prove this, we even use an analogous L_2 calculation to that used for the mixing, namely Lemma 3.1.8.

3.2.3 Size of Ball Estimates and Lower Bound

Lemma 3.2.4. *For all $R \geq 0$, we have*

$$\big|B_k(R)\big| = \tbinom{\lfloor R \rfloor + k}{k}.$$

Proof. Assume that $R \in \mathbb{N}$. It is a standard combinatorial identity that

$$\big|B_k(R)\big| = \big|\{\alpha \in \mathbb{Z}_+^k \mid \textstyle\sum_1^k \alpha_i \leq R\}\big| = \tbinom{R+k}{k}. \qquad \square$$

Recall that $M_k = M_k^+ = k|H_{p,d}^{\mathrm{ab}}|^{1/k}/e$. The next lemma shows that the difference between M_k^* and M_k^* is only in sot, and so can be absorbed into the error terms.

Lemma 3.2.5. *For $k \ll \log |H_{p,d}^{\mathrm{ab}}|$, for all constants $\xi \in (0,1)$, we have*

$$M_k^* \leq \big\lceil M_k(1+\xi) \big\rceil \quad \text{and} \quad \big|B_k\big(M_k(1-\xi)\big)\big| \ll |H_{p,d}^{\mathrm{ab}}|.$$

Proof. *Upper bound.* Set $M := e^{\xi} k|H_{p,d}^{\mathrm{ab}}|^{1/k}/e$. By Stirling's approximation, we have

$$\tbinom{M+k}{k} \geq M^k/k! \gtrsim k^{-1/2}(eM/k)^k = k^{-1/2} e^{\xi k} |H_{p,d}^{\mathrm{ab}}|.$$

Since $\omega \ll k$, we have $k^{-1/2}e^{\xi k} \gg e^{\omega}$ and $\binom{M+k}{k} \geq e^{\omega}|H_{p,d}^{\mathrm{ab}}|$.

Lower bound. Set $M := e^{-\xi} k|H_{p,d}^{\mathrm{ab}}|^{1/k}/e$. Using the inequality $\binom{N}{k} \leq (eN/k)^k$, we have

$$\tbinom{M+k}{k} \leq \big(e(M+k)/k\big)^k \leq (eM/k)^k \exp\big(k^2/M\big) \leq e^{-\xi k}|H_{p,d}^{\mathrm{ab}}| \exp\big(e^{1+\xi} k/|H_{p,d}^{\mathrm{ab}}|^{1/k}\big).$$

Since $k \ll \log |H_{p,d}^{\mathrm{ab}}|$, we have $k/|H_{p,d}^{\mathrm{ab}}|^{1/k} \ll \xi k$ and $\binom{M+k}{k} \ll |H_{p,d}^{\mathrm{ab}}|$. $\square$

From these, it is straightforward to deduce the lower bound (for all Z) in Theorem 3.2.1a.

Proof of Lower Bound in Theorem 3.2.1a. Were the underlying group Abelian, we would be able to upper bound $|\mathcal{B}_k(M)| \leq |B_k(M)|$. However, this does not hold for general groups.

Recall that the Abelianisation of $H_{p,d}$ corresponds to 'modding out all but the super-diagonal':

$$H_{p,d}^{\mathrm{ab}} = H_{p,d}/[H_{p,d}, H_{p,d}] \cong \mathbb{Z}_p^{d-1} \quad \text{and} \quad |H_{p,d}^{\mathrm{com}}| = |[H_{p,d}, H_{p,d}]| = p^{(d-1)(d-2)/2}.$$

for a given number of steps, the number of different elements that can be seen is at most $|H_{p,d}^{\mathrm{com}}|$ times the number that can be seen in the Abelianisation $H_{p,d}^{\mathrm{ab}}$; that is,

$$\big|\mathcal{B}_k(M)\big| \le L \cdot |H_{p,d}^{\mathrm{com}}| \quad \text{where} \quad L := \big|\big\{gH_{p,d}^{\mathrm{com}} \;\big|\; g \in \mathcal{B}_k(M)\big\}\big|.$$

Since $L \le |B_k(M)|$, we have $|\mathcal{B}_k(M)| \le |B_k(M)||H_{p,d}^{\mathrm{com}}|$. We choose M so that $|B_k(M)| \approx |H_{p,d}^{\mathrm{ab}}|$. More precisely, for any constant $\xi > 0$, by Lemma 3.2.5, we have

$$|B_k\big(M_k^*(1-\xi)\big)| \ll |H_{p,d}^{\mathrm{ab}}| \cdot |H_{p,d}^{\mathrm{com}}| = |H_{p,d}|.$$

Thus, for any constant $\beta \in (0,1)$, we have $\mathcal{D}_k(\beta) \ge M_k^*(1-\xi)$, asymptotically. □

Remark 3.2.6. This proof generalises further. Instead of looking at just Heisenberg groups, we can take *any* group G. We then obtain a lower bound analogously, but where now M_k^* is defined so that $|B_k(M_k^*)| \ge e^{\omega}|G^{\mathrm{ab}}|$, for some suitable $\omega \gg 1$. (For $H_{p,d}$, this is $e^{\omega}p^{d-1}$.) △

Remark 3.2.7. The statements are for *directed* lattice balls, in $\mathbb{Z}_+^k$. Changing to *undirected* lattice balls, in $\mathbb{Z}^k$, increases the size by a factor at most 2^k. Since $k \ll \log|H_{p,d}^{\mathrm{ab}}|$ and we are looking at sizes of the order $|H_{p,d}^{\mathrm{ab}}|$, analogous statements can easily be proved for directed balls. △

3.2.4 Mixing-Type Results and Upper Bound

As stated in the outline (§3.2.2), we replace the auxiliary W with $A \sim \mathrm{Unif}(B_k(M_k^*))$, and then apply the generators in a uniformly chosen order. More precisely, we have the following algorithm.

Definition 3.2.8. *Define S via the following random algorithm.*

- *First draw $A \sim \mathrm{Unif}(B_k(M_k^*))$; this tells us how many times we use each of the k generators. Define the vector g by $g_1 = \cdots = g_{A_1} = 1$, $g_{A_1+1} = \cdots = g_{A_1+A_2} = 2$ and so on.*
- *To decide in which order we apply the generators, label the steps $1, \dots, N$, so $N := \sum_1^k A_i$, and then draw a uniform permutation σ on $[N] = \{1, \dots, N\}$; this will tell us in which order we the generators: $S := Z_{g_{\sigma(1)}} \cdots Z_{g_{\sigma(N)}}$.*

In words, we choose how many times each generator is going to be used by A, and then apply them in a uniformly chosen order. In particular, we can define $(C_{i,j})_{i,j\in[k]}$ as before; see Lemma 3.1.17.

We now present our 'mixing-type' result, showing that S is close to uniform.

Proposition 3.2.9. *Assume that the conditions of Theorem 3.2.1a hold. Then*

$$\mathbb{E}\big(\big\|\mathbb{P}_{G_k}\big(S \in \cdot\big) - \pi_G\big\|_{\mathrm{TV}}\big) = o(1).$$

Proof. For notational ease, write $M := M_k^*$. Let S', A' and σ' be independent copies of S, A and σ, respectively. For a set $\mathcal{A}$ (to be defined), the modified L_2 calculation used in Lemma 3.1.8 gives

$$\mathbb{E}\big(\big\|\mathbb{P}_{G_k}\big(S = \cdot \mid Z\big) - \pi_G\big\|_{\mathrm{TV}}\big) \le n\,\mathbb{P}\big(S = S' \mid \mathsf{typ}\big) - 1 + \mathbb{P}\big(A \notin \mathcal{A}\big),$$

where $\mathsf{typ} := \{A, A' \in \mathcal{A}\}$. Write $n := |H_{p,d}| = p^{d(d-1)/2}$. Similarly to the mixing case, we separate according to whether or not $A = A'$. If $A = A'$, then we do an analysis similar to that of $W = W'$ from §3.1. Using Lemma 3.1.14 in an analogous way as was used to obtain (3.1.26), we obtain

$$\mathbb{P}\big(S = S' \mid A \ne A',\, \mathsf{typ}\big) = 1/n = 1/|H_{p,d}| = 1/p^{d(d-1)/2}. \tag{3.2.1}$$

recall that the coefficients in Lemma 3.1.14 (corresponding to the entries of $A - A'$ here) are deterministic, and hence (3.2.1) holds *regardless* of the choice of $\mathcal{A}$. This also uses the fact that $M \ll p$, and so $|A_i - A_i'| \ll p$ for all i; this follows from manipulating the conditions of Theorem 3.2.1a and using $M \asymp k|H_{p,d}^{\mathrm{ab}}|^{1/k}$. Using the definition of $M = M_k^*$, it is easy to calculate

$$\mathbb{P}\big(A = A' \mid \mathsf{typ}\big) \le \big|B_k(M)\big|^{-1}/\mathbb{P}\big(\mathsf{typ}\big) \le |H_{p,d}^{\mathrm{ab}}|^{-1}e^{-\omega}/\mathbb{P}\big(\mathsf{typ}\big); \tag{3.2.2}$$

this replaces the entropic calculation (3.1.27). Combining (3.2.1, 3.2.2) establishes

$$n\,\mathbb{P}\big(S = S' \mid \mathsf{typ}\big) - 1 \le n|H_{p,d}^{\mathrm{ab}}|^{-1}\,\mathbb{P}\big(S = S' \mid A = A',\, \mathsf{typ}\big)/\mathbb{P}\big(\mathsf{typ}\big)$$

As stated above, the analysis of $\mathbb{P}(S = S' \mid A = A',\, \mathsf{typ})$ is analogous to the $W = W'$ case from §3.1. There we stated that it was not important that W was a DRW, and that we would apply the same proof here (§3.2) for a "different W"—the A just defined is this "different W". Recall that we separated the generators using the partition $\{P_3, ..., P_d\}$; we do the same here. Define $\mathcal{E}_b$ as in (3.1.31):

$$\boldsymbol{C}_b \coloneqq (C_{i,j})_{i,j\in P_b}, \quad \boldsymbol{C}'_b \coloneqq (C'_{i,j})_{i,j\in P_b} \quad \text{and} \quad \mathcal{E}_b \coloneqq \{\boldsymbol{C}_b = \boldsymbol{C}'_b\}.$$

So far, we have in essence been following *Proof of Theorem 3.1.6 Given Lemmas 3.1.18 and 3.1.10* from the start of §3.1.8, but with $W(t)$ replaced by A and $\mathcal{W}$ replaced by $\mathcal{A}$; it is not until Lemma 3.1.18, which upper bounds the analogue of $\mathbb{P}(S = S' \mid A = A',\, \mathsf{typ})$, that the choice of typicality is made. Hence (3.1.32) still holds here: write $\overline{\mathbb{P}}_a(\cdot) \coloneqq \mathbb{P}(\cdot \mid A = A' = a,\, \mathsf{typ})$; we have

$$\overline{\mathbb{P}}_a\big(S = S'\big) \le 2^{d^2/2}\prod_3^d\big(1/p^{d-2} + q_b\big) \quad \text{where} \quad q_b \coloneqq \max_{a\in\mathcal{A}}\prod_3^d \overline{\mathbb{P}}_a(\mathcal{E}_b). \tag{3.2.3}$$

In the mixing context, Lemma 3.1.18 upper bounded this probability by $2^{d^2}n^{-1}e^{h_0}$, where h_0 was the entropy; for $k \ll \log|H_{p,d}^{\mathrm{ab}}|$, we chose $h_0 = \log|H_{p,d}^{\mathrm{ab}}|$, so this upper bound became $2^{d^2/2}n^{-1}|H_{p,d}^{\mathrm{ab}}|$. Combined with the entropic calculation (3.1.27), of which (3.2.2) is the analogue, established (3.1.34): $n\,\mathbb{P}(S = S' \mid \mathsf{typ}) - 1 \le 2\cdot e^{-\omega}2^{d^2}$. Conditions on d ensured that we could choose ω to make this $o(1)$.

We claim that we can copy the proof of Lemma 3.1.18 to show that $q_b \le 1/p^{b-2}$. The proof of this claim is deferred to the end of the subsection (§3.2.4). From this claim, we deduce that

$$\overline{\mathbb{P}}_a\big(S = S'\big) \le 2^{d^2}\prod_3^d 1/p^{b-2} = 2^{d^2}p^{-(d-1)(d-2)/2} = 2^{d^2}|H_{p,d}^{\mathrm{ab}}|/n. \tag{3.2.4}$$

Combining (3.2.3, 3.2.4), we obtain

$$n\,\mathbb{P}\big(S = S' \mid \mathsf{typ}\big) - 1 \le e^{-\omega}2^{d^2}/\mathbb{P}\big(\mathsf{typ}\big),$$

where we shall choose typ so that $\mathbb{P}(\mathsf{typ}) = 1 - o(1)$; this is analogous to (3.1.34). We now check that our conditions on d allow us to choose ω so that $\omega \gg d^2$: recall from Definition 3.2.3 that $\omega = \max\{(\log k)^2, k/|H_{p,d}^{\mathrm{ab}}|^{1/(2k)}\}$; the condition $d^2 \ll \omega$ is included in the conditions of Theorem 3.2.1a.

It remains to prove our claim that we can copy of the proof of Lemma 3.1.18 to prove that $q_b \le 1/p^{b-2}$. In said proof, we were particularly interested in the (expected) number of times that an individual generator was picked; this was $s \coloneqq t/k$, and, in the regime $k \ll \log|H_{p,d}^{\mathrm{ab}}|$, satisfied $s \asymp |H_{p,d}^{\mathrm{ab}}|^{2/k}$. At the start of *Proof of Lemma 3.1.18* for $k \ll \log|H_{p,d}^{\mathrm{ab}}|$, we emphasised that the proof did not rely heavily on the distribution of W, nor did it need $s \asymp p^{2(d-1)/k}$; we apply the same arguments with W replaced by A, and in this case the expected number of times that an individual generator is picked, which we still denote s, satisfies $s \asymp |H_{p,d}^{\mathrm{ab}}|^{1/k}$ since $A \sim \mathrm{Unif}(B_k(M_k^*))$ and, by Lemma 3.2.5, $M_k^* \asymp k|H_{p,d}^{\mathrm{ab}}|^{1/k}$. We elaborate further on how to adapt the proof to this context.

Let $\eta \in (0,1)$ be a (small) constant. For $a \in \mathbb{Z}_+^k$, writing

$$\mathcal{C}(a) \coloneqq \big\{i \in [k] \;\big|\; \eta s \le a_i \le \eta^{-1}s\big\}, \quad \text{we have} \quad \mathbb{P}\big(|\mathcal{C}(A)|/k \ge \tfrac{4}{5}\big) = 1 - o(1),$$

if η is sufficiently small; this is analogous to (3.1.39). We use this to define typicality, analogously to (3.1.40) except recalling that we no longer require the entropic part:

$$\mathcal{A} \coloneqq \big\{a \in \mathbb{Z}_+^k \;\big|\; |\mathcal{C}(a)| \ge \tfrac{4}{5}k,\ \max_i a_i < p\big\}.$$

We use exactly the same decomposition of generators; we look at m-tuples, and require $m \ll k/d^2$. We take $m \asymp d$, and so need $d^3 \ll k$; this is implied by the conditions of Theorem 3.2.1a. Fixing some $a \in \mathcal{A}$, consider the mode μ_m of the vector $\boldsymbol{C}_m \coloneqq (C_{i,j})_{i,j\in[m]}$, conditional on $A = a$; here $C_{i,j}$ is defined as in (3.1.10), but now with S defined using A instead of W. Since $\log s = \log|H_{p,d}^{\mathrm{ab}}|/k \asymp d\log p/k$, as it did in §3.1.8, we may apply Claim 3.1.19, provided $m \ll |H_{p,d}^{\mathrm{ab}}|^{1/k}$. This condition is in Theorem 3.2.1a. Applying Lemma 3.1.18 gives

$$q_b \le p^{-(b-2)\cdot d^{-1}f(m)/m} \quad \text{for some} \quad f(m) \asymp m^2,$$

exactly as in (3.1.42); setting $m \coloneqq Cd$ for a sufficiently large constant C gives $q_b \le 1/p^{b-2}$. □

3.2.5 Extensions

In this subsection we describe two extensions to the argument.

§3.2.5.1 We consider typical distance $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$.

§3.2.5.2 We consider diameter for $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$.

§3.2.5.3 We consider undirected Cayley graphs for $1 \ll k \lesssim \log|H_{p,d}^{\mathrm{ab}}|$.

3.2.5.1 Extending Typical Distance to $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$

In this subsection we extend the argument to $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$. We consider first $k \asymp \log|H_{p,d}^{\mathrm{ab}}|$.

Typical Distance for $k \asymp \log|H_{p,d}^{\mathrm{ab}}|$. Key to the typical distance analysis was adapting the mixing time analysis from §3.1.8 to the case where the auxiliary process W, which is a RW, is replaced with A, which is uniform on a certain ball. To define the element S of G, we use the generators chosen by A (ie Z_i A_i times), applied in a uniformly chosen order; see Definition 3.2.8. We then adapted the mixing analysis of §3.1 to show, for $k \ll d\log p$, that S is well-mixed.

The method for $k \asymp \log|H_{p,d}^{\mathrm{ab}}|$ is exactly the same, except that now we use mixing analysis for $k \asymp \log|H_{p,d}^{\mathrm{ab}}|$. We define $\mathcal{A}$ analogously to $\mathcal{W}$, given in (3.1.35), except without the entropic part:

$$\mathcal{A} := \big\{a \in \mathbb{Z}_+^k \;\big|\; |J(a) - Re^{-R/k}| \le \tfrac{1}{2}\varepsilon R e^{-R/k},\ \max_i a_i < \sqrt{p}\big\} \quad \text{where} \quad J(a) := \textstyle\sum_1^k \mathbf{1}(a_i = 1).$$

Write $\mathsf{typ} := \{A, A' \in \mathcal{A}\}$. It then suffices to show that

$$\mathbb{E}\big(\big\|\mathbb{P}_{G_k}\big(S \in \cdot \mid A \in \mathcal{A}\big) - \pi_G\big\|_2^2\big) = n\,\mathbb{P}\big(S = S' \mid A = A',\ \mathsf{typ}\big) - 1 = o(1).$$

(Indeed, if the L_2 norm is $o(1)$ in expectation, then it is $o(1)$ whp, and in particular the support of S must be a proportion $1 - o(1)$ of the vertices of G whp.) As in (3.2.3), we have

$$\overline{\mathbb{P}}_a\big(S = S'\big) \le 2^{d^2/2} \textstyle\prod_3^d\big(1/p^{d-2} + q_b\big) \quad \text{where} \quad q_b := \max_{a\in\mathcal{A}} \prod_3^d \overline{\mathbb{P}}_a(\mathcal{E}_b),$$

recalling that we wrote $\overline{\mathbb{P}}_a(\cdot) := \mathbb{P}(\cdot \mid A = A' = a, \mathsf{typ})$. Now, however, to bound $\overline{\mathbb{P}}_a(\mathcal{E}_b)$, we use the method from Lemma 3.1.18 with $k \asymp \log|H_{p,d}^{\mathrm{ab}}|$, rather than $1 \ll k \ll \log|H_{p,d}^{\mathrm{ab}}|$. There, key was to consider only generators which are chosen either once or not at all—we needed such generators to have the same relative order in S as in S'. For more details, see (3.1.18, 3.1.19, 3.1.36, 3.1.38). The remainder of the mixing-type proof follows analogously to the regime $1 \ll k \ll \log|H_{p,d}^{\mathrm{ab}}|$.

It remains to describe the adaptations in calculating the minimal radius of a ball of cardinality at least $|H_{p,d}^{\mathrm{ab}}| = p^{d-1}$. When $k \asymp \log|H_{p,d}^{\mathrm{ab}}|$, it is not difficult to see that we need a radius order k; so typically each of the k coordinates gets displaced by an order 1 amount. We wish to choose $R \asymp k \asymp \log|H_{p,d}^{\mathrm{ab}}|$ so that $\binom{R+k}{k} \approx p^{d-1}$. A more refined approximation to $\binom{R+k}{k}$ is needed, with $R \asymp k$: an application of Stirling's approximation gives

$$\log\tbinom{R+k}{k} = (R+k)h\big(k/(R+k)\big)\cdot\big(1 + o(1)\big),$$

where $h : [0,1] \to [0,1] : q \mapsto -q\log q - (1-q)\log(1-q)$ is the entropy (in nats) of $\mathrm{Bern}(q)$. Writing $R = \alpha k$, if $k \approx \lambda \log|H_{p,d}^{\mathrm{ab}}|$, then, motivated by the above display, we choose α to satisfy

$$(\alpha+1)h\big(1/(\alpha+1)\big) = \lambda$$

in the directed case. In the undirected case, the same analysis holds, but the lattice balls have slightly different sizes; cf §3.2.5.3. The desired radius R still satisfies $R \asymp k$, but now with a different implicit constant; cf $\alpha^{\pm}$ in Theorem 4.2.2. Further details can be found in §4.2.3 and §6.5. (In §4.2, the groups G are Abelian, so $|G^{\mathrm{ab}}| = n$.) □

The analysis for $k \gg \log|H_{p,d}^{\mathrm{ab}}|$ is almost identical.

Typical Distance for $k \gg \log|H_{p,d}^{\mathrm{ab}}|$. The mixing time of the RW gives an upper bound on the typical distance. As noted in Remark J.2, the mixing time agrees with the desired upper bound on the typical distance. This completes the proof of the upper bound.

We turn to the lower bound. The lower bound of $\log_k|H_{p,d}|$ is trivial. For the other part of the maximum, we project to the Abelianisation $H_{p,d}^{\mathrm{ab}}$, as in §3.1.5. As in Lemma 3.2.4, we have $|B_k(R)| \le 2^R\binom{R+k}{R}$. It is easy to check that this is $o(|H_{p,d}^{\mathrm{ab}}|)$ when $R = (1-\xi)\cdot\frac{\rho}{\rho-1}\log_k|H_{p,d}^{\mathrm{ab}}|$ and $k = (\log|H_{p,d}^{\mathrm{ab}}|)^\rho$. This is done carefully in §4.3. (There the groups are Abelian, so $G = G^{\mathrm{ab}}$.) □

3.2.5.2 Diameter for $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$

In this subsection, we outline the diameter argument. Here $d \asymp 1$, so $\log|H_{p,d}| \asymp \log|H_{p,d}^{\mathrm{ab}}|$. In §4.2 we consider typical distance for Abelian groups in the regime $k \asymp \log|G|$. (For Abelian groups, $G = G^{\mathrm{ab}}$.) In §4.4 we show carefully how to adapt the argument to the diameter; here we give a sketch of the argument there, adapted slightly to Heisenberg groups.

Proof of Theorem 3.2.2. Clearly typical distance is a lower bound on the diameter.

We first assume that $k \approx \lambda\log|H_{p,d}^{\mathrm{ab}}|$ with $\lambda \in (0,\infty)$. Our aim is to show that the diameter is, up to sot, the same as the typical distance; by Theorem J, this is αk for some constant $\alpha := \alpha_\lambda^{\pm}$. We thus let $\xi > 0$ and set $R := \alpha k(1+\xi)$; we show that $\operatorname{diam}(H_{p,d})_k \le R+1$ whp.

Split the generators into two sets: $A := [Z_1,...,Z_{(1-\varepsilon)k}]$ and $B := [Z_{(1-\varepsilon)k+1},...,Z_k]$, with $\varepsilon = o(1)$ to be determined. Roughly, one first uses the generators in A to connect the identity to the set $S \subseteq H_{p,d}$ of elements of G which can be reached by paths of length at most R. By Theorem J, we have $|H_{p,d}\setminus S| \ll |H_{p,d}|$ whp; assume this herein. Given an arbitrary $g \in G$, if $g \notin S$ then one uses the remaining generators from B to connect g to S directly, and by extension to the identity. More precisely, we try to connect $g \in G\setminus S$ to S via $hz = g$ for some $z = Z_i$ for some $i > (1-\varepsilon)k$ and $h \in S$. The probability that this fails for a given g for all such Z_i and all $h \in S$ is at most $(|H_{p,d}\setminus S|/|H_{p,d}|)^{\varepsilon k}$. Since $|H_{p,d}\setminus S| \ll |H_{p,d}|$ and $k \asymp \log|H_{p,d}^{\mathrm{ab}}| \asymp \log|H_{p,d}|$, we can choose $\varepsilon \to 0$ sufficiently slowly so that this latter probability is $o(1/|H_{p,d}|)$. By the union bound, the probability that this fails for some such g is $o(1)$. When this does not fail, $\operatorname{diam}(H_{p,d})_k \le R+1$.

Finally, when $k \asymp \log|H_{p,d}^{\mathrm{ab}}|$ replacing k with $(1-\varepsilon)k$ changes the typical distance by a factor $1+o_{\varepsilon\to0}(1)$. This completes the proof for this regime.

The same argument holds for $k \gg \log|H_{p,d}^{\mathrm{ab}}|$, using the typical distance result of that regime. □

3.2.5.3 Undirected Cayley Graphs

Here we describe the required adaptations to the proof to allow undirected graphs. The only major difference is that the size of the discrete lattice balls changes: previously we considered $a \in \mathbb{Z}_+^k$ with $\sum_1^k a_i \le R$, while now we consider $a \in \mathbb{Z}^k$ with $\sum_1^k|a_i| \le R$. Indeed, besides estimates on the sizes of balls, the only tool required was an adaptation of the mixing proof. Since this proof works for both the directed and undirected cases, the same holds true here.

When $1 \ll k \ll \log|H_{p,d}^{\mathrm{ab}}|$, the desired radius R satisfies $R \gg k$. Comparing Lemma 6.5.3a with Lemma 3.2.4 shows that the radius of the desired ball changes by approximately a factor 2. This is shown carefully in §4.1.3; there the groups G are Abelian, so $|G^{\mathrm{ab}}| = |G|$.

When $k \asymp \log|H_{p,d}^{\mathrm{ab}}|$, the desired radius R satisfies $R \asymp k$. In general there is not an easy closed form for the implicit constant in either the directed or undirected cases; cf Theorem 4.2.2.

3.3 Concluding Remarks and Open Questions

§3.3.1 We discuss some statistics in the regime where k is a fixed constant.

§3.3.2 We discuss very briefly how our methods can be extended to more general nilpotent groups.

§3.3.3 To conclude, we discuss some questions which remain open and gives some conjectures.

3.3.1 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 [27], using their concept of *moderate growth*. There is no cutoff.

Recall that a group G is called *nilpotent of step at most* L 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 $t_{\mathrm{rel}} := t_{\mathrm{rel}}(G(Z))$ for the relaxation time (ie inverse of the spectral gap) and $t_{\mathrm{mix}} := t_{\mathrm{mix}}(\varepsilon; G(Z))$ for the (TV) ε-mixing time, for $\varepsilon \in (0,1)$. When considering sequences $(G_N(Z_{(N)}))_{N\in\mathbb{N}}$, add an N-sub/superscript.

We phrase the result of Diaconis and Saloff-Coste [27] in our language.

Theorem 3.3.1 (cf [27, 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 $(t^N_{\mathrm{mix}})_{N\in\mathbb{N}}$ does not exhibit the cutoff phenomenon; in particular, $t^N_{\mathrm{mix}}/k_N \lesssim \Delta_N^2 \lesssim t^N_{\mathrm{rel}} \lesssim t^N_{\mathrm{mix}}$ as $N \to \infty$.*

We give a very brief exposition of the results of Diaconis and Saloff-Coste [27], including the definition of moderate growth, leading to this conclusion in §5.4.

3.3.2 Extending Our Arguments from Heisenberg to Other Nilpotent Groups

Our analysis has focussed on Heisenberg matrix groups; these are a canonical class of nilpotent groups. In the introduction, in Remark C.3, we claimed that some of our analysis extends from Heisenberg groups to more general nilpotent groups. This extension is based primarily on observations made by Péter Varjú during discussions of our work with him.

Recall that we wrote S for the location of the walk and W for its auxiliary variable; let W' be an independent copy of W, and define S' correspondingly. Recall the definition of $C_{i,j}$ from (3.1.10). The difference $C_{i,j} - C'_{i,j}$ has a natural group theoretic interpretation. Indeed, for a step-2 nilpotent group, one can write $S'S^{-1}$ in the canonical form $\prod_1^k Z_i^{W_i - W'_i} Z_i^{V_i} \cdot \prod_{i<j} [Z_i, Z_j]^{C_{i,j} - C'_{i,j}}$; for general groups, $S'S^{-1}$ can be written like this up to multiplication on the right by a term in $[[G,G],G]$. When $W = W'$, we are left just with the product of commutator terms in the expression for $S'S^{-1}$.

For a p-group G, when $W \neq W' \bmod p$, one can show that $S'S^{-1}$ is uniformly distributed on G. (In the current article, we used the specific structure of $H_{p,d}$ to reach this conclusion.)

We give a fairly detailed discussion, elaborating on the above points, in §5.3.

3.3.3 Open Questions and Conjectures

We close the paper with some questions which are left open.

1: Sufficient Conditions for Cutoff for Nilpotent and General Groups

We have established cutoff for a family of non-Abelian groups. The group is guaranteed to be generated whp if $k \gg \log |G|$.

> **Conjecture 1.** *For all groups G, for $k \gg \log |G|$ with $\log k \ll \log |G|$, the random walk on G_k exhibits cutoff whp.*

It is natural to ask at which time this cutoff occurs.

> **Open Problem 1.** *Find an expression for the cutoff time in Conjecture 1.*
>
> *Find conditions under which this time can be in terms of a few statistics of the group, eg the size of the Abelianisation, the number of low dimensional irreducible representations or the size of the largest Abelian subgroup.*

In Theorem A in Chapter 2 we establish cutoff for all Abelian groups in the regime $1 \ll k \lesssim \log|G|$. There are some necessary conditions to generate the group whp. Using the techniques from there, one can look to extend Open Problem 1 to the regime $1 \ll k \lesssim \log|G^{\mathrm{ab}}|$.

Since $t_0(k, |G^{\mathrm{ab}}|)$ is a lower bound on the mixing time, clearly the size of the Abelianisation plays a role. Relatedly, the size of the largest Abelian subgroup appears to play a role. Indeed, consider the dihedral group D_{2n} of order $2n$. This has an Abelian subgroup congruent to $\mathbb{Z}_n$ (corresponding to rotations). Some basic calculations suggest that the mixing time likely should be the same as that of $\mathbb{Z}_n$. This is perhaps related to irreducible representations (*irreps*): D_{2n} has at most 4 irreps of dimension 1 (and hence $|D_{2n}^{\mathrm{ab}}| \leq 4$), and all the remaining irreps are of dimension 2. Instead of just considering $|G^{\mathrm{ab}}|$, which is the number of irreps of dimension 1, more generally the number of low dimensional irreps (in some precise sense) is likely to affect the mixing time.

As a starting point, one should perhaps study nilpotent groups. Finally, we mention work by Gowers [39], on *quasirandom groups*. He looks for groups whose (non-trivial) irreps all have high dimension; such groups he describes as being 'very far from Abelian'. The farther a group is from Abelian, the faster we expect its mixing time to be. Perhaps a similar criterion would be useful for this question of comparing the mixing of the Abelianisation with that of the full group.

2: Cutoff for Heisenberg Group $H_{p,d}$ with p Small (eg $p = 2$)

Our conditions on d appear to be more than artefact of the proof than actual necessities. It is natural to expect there to be cutoff even for larger d. Here W should now be a DRW on $\mathbb{Z}_p^k$ (ie taken mod p), rather than $\mathbb{Z}^k$; see below, or §2.2. As such, the entropic times are now defined with respect to the RW on $\mathbb{Z}_p^k$.

> **Conjecture 2.** *Let p be prime and $d \geq 3$. Suppose that $1 \ll \log k \ll \log|H_{p,d}|$. Then the RW on $(H_{p,d})_k$ exhibits cutoff whp over Z, with behaviour similar to Theorem 3.1.6.*

One place in which we loose information is in moving from 3×3 matrices to $d \times d$ matrices. Indeed, compare (3.1.6) with (3.1.25): in both cases, we used the 'Abelian' terms (ie monomials of degree 1) and the first 'non-Abelian' terms (ie monomials of size 2); for general d, there are $d-1$ terms—coordinates at distance ℓ from the diagonal contain monomials of size at most ℓ. (This corresponds to the fact that $H_{p,3}$ is step-2 nilpotent, while in general $H_{p,d}$ is step-$(d-1)$ nilpotent; the further from the diagonal a coordinate is, the 'more non-Abelian' it is.) Considering only the first two terms meant that the analysis was more analogous to the 3×3 case, however we threw away a lot of information. In order to study the case where d is very large (compared with p), one surely needs to analyse these higher order terms.

In §5.2, we study in detail cutoff for the Abelian group $\mathbb{Z}_p^d$, in particular allowing (prime) p to be fixed; this extends Wilson's consideration of $\mathbb{Z}_2^d$ (ie $p = 2$) in [77]. One key difference is that instead of letting W be a RW on $\mathbb{Z}^k$, we take each coordinate modulo p; this leads from entropy to relative entropy considerations. The same adaptation should be made here. In the current article, p has been large enough so that almost all coordinates of W never reach p, and hence the distinction between a RW on $\mathbb{Z}^k$ and $\mathbb{Z}_p^k$ is negligible (cf §2.1 vs §2.2); this will not be the case for small p.

3: Cutoff for Heisenberg Group $H_{p,d}$ with p Not Prime

Throughout this paper, we primarily considered $H_{p,d}$ with p prime. In Theorem D.2 we relieved this condition, but only in the regime $k \gtrsim \log|H_{p,d}^{\mathrm{ab}}|$. We conjecture that the analogous behaviour holds for p not prime in the regime $1 \ll k \ll \log|H_{p,d}^{\mathrm{ab}}|$. This is work in progress.

> **Conjecture 3.** *Subject to potentially stronger conditions, analogous results hold when p is not prime with the same cutoff time.*

When we studied Abelian groups in Chapter 2, we did not assume that the analogue of p was prime; we did the gcd analysis. It is not unreasonable to imagine that similar techniques applied there (see §2.1.6/§2.2.7) may well be applicable here too.

4: Spectral Gap for Heisenberg Group $H_{p,d}$ with $k \gtrsim \log |H_{p,d}^{\mathrm{ab}}|$

We studied typical distance for $k \lesssim \log |H_{p,d}^{\mathrm{ab}}|$. In the boundary regime $k \asymp \log |H_{p,d}^{\mathrm{ab}}|$, the typical distance is order k. We study typical distance for Abelian groups in Theorem G, obtaining analogous results—for an Abelian group, $G = G^{\mathrm{ab}}$. The regime $k \asymp \log |G|$ is the point at which the Cayley graph of an Abelian group becomes an expander; see Theorem L. It is natural to conjecture that the analogue holds for Heisenberg groups. It is known that the G_k is an expander whp for *any* group is when, eg, $k \geq 2 \log_2 |G|$; this is the celebrated Alon–Roichman theorem [3].

Conjecture 4. *If $k \gtrsim \log |H_{p,d}^{\mathrm{ab}}|$ with $k - d \asymp k$, then $(H_{p,d})_k$ is an expander whp.*

If this were proved for some diverging d, then it would provide the first example of a group with the property that its k-uniform Cayley graph is an expander whp for some k with $k \ll \log |G|$.

5: Diameter for Heisenberg Group $H_{p,d}$ for Diverging k

We have shown concentration of typical distance, but never considered the diameter. It is trivial that the typical distance is a lower bound on the diameter, and that twice the typical distance is an upper bound. Further, in the regime $k \asymp \log |H_{p,d}^{\mathrm{ab}}|$, we argued that the diameter and typical distance are asymptotically equivalent; see §3.2.5.1. Can more be determined?

Conjecture 5. *For $G = H_{p,d}$ and $Z_1, ..., Z_k \sim^{\mathrm{iid}} \mathrm{Unif}(H_{p,d})$, write Δ_Z for the diameter of the Cayley graph with generators Z. Assume that k diverges, sufficiently rapidly in terms of d. Then the law of Δ_Z concentrates.*

6: Diameter for Heisenberg Group $H_{p,d}$ for Fixed k

Instead of requiring $k \gg 1$, as in the previous question, for fixed d we can ask that k is (at least) a sufficiently large constant (depending on d). Theorem 3.3.1 and considering just the Abelianisation shows that the correct order for the diameter if $|H_{p,d}^{\mathrm{ab}}|^{1/k}$, and suggest that in the limit as k grows it is order $k|H_{p,d}^{\mathrm{ab}}|^{1/k}$. Shapira and Zuck [71] establish, for Abelian groups, convergence in distribution of the normalised diameter to some non-trivial random variable. (Cf Amir and Gurel-Gurevich [4].) We conjecture that the same holds for Heisenberg groups.

Conjecture 6. *For $G = H_{p,d}$ and $Z_1, ..., Z_k \sim^{\mathrm{iid}} \mathrm{Unif}(H_{p,d})$, write Δ_Z for the diameter of the Cayley graph with generators Z. For all $k \in \mathbb{N}$, under suitable conditions on d, there exists a non-trivial random variable Δ_k so that $\Delta_Z/(k|H_{p,d}^{\mathrm{ab}}|^{1/k}) \to^{\mathrm{dist}} \Delta_k$ (in distribution) as $p \to \infty$. Further, the sequence $(\Delta_k)_{k \in \mathbb{N}}$ of random variables is tight.*

Subsequent to the original posting of this conjecture, El-Baz and Pagano [6, Theorem 1.2] established the convergence in distribution for each fixed k, without any primality assumption on p. They derive this as a consequence of a general inequality, showing that the diameter of a Cayley graph of a nilpotent group is governed by the diameter of its Abelianisation.

Questions for Typical Distance

Questions for typical distance can be asked analogous to those detailed in Questions 2 and 3.

Replacing the Heisenberg Group with a Nilpotent Group

Questions 4–6 for the Heisenberg group can all be extended by replacing $G = H_{p,d}$ with a general nilpotent group G; where one sees the Abelianisation $H_{p,d}^{\mathrm{ab}}$, this should be replaced with G^{ab}.

4 Geometry of Random Cayley Graphs of Abelian Groups

Abstract for Chapter 4

We show that the distance from the identity for all but $o(|G|)$ of the elements of G lies in the interval $[M - o(M), M + o(M)]$. In the regime $k \gtrsim \log|G|$, we show that the diameter of the graph is asymptotically equivalent to the typical distance value M whp. In the spirit of the Aldous–Diaconis conjecture, this M depends only on k and $|G|$.

Additionally, when $k - \log_2|G| \asymp k$, we show, in a quantitative sense, that the group $\mathbb{Z}_2^d$ gives rise to the largest diameter amongst all (not just Abelian) groups whp.

We prove that the spectral gap of the graph is order $|G|^{-2/k}$ when $k - d(G) \asymp k$ whp. This extends, for Abelian groups, a celebrated result of Alon and Roichman [3].

Table of Contents for Chapter 4

4.1 Typical Distance: $1 \ll k \ll \log|G|$

This section focusses on concentration of distances from the identity in the random Cayley graph of an Abelian group when $1 \ll k \ll \log|G|$. (Subsequent sections deal with $k \gtrsim \log|G|$.) The main result of the section is Theorem 4.1.2; see also Hypothesis F.

The outline of this section is as follows:
- §4.1.1 states precisely the main theorem of the section;
- §4.1.2 outlines the argument;
- §4.1.3 gives some crucial estimates on the size of lattice balls;
- §4.1.4 is devoted to the lower bound;
- §4.1.5 is devoted to the upper bound.

4.1.1 Precise Statement and Remarks

To start the section, we recall the typical distance statistic.

Definition 4.1.1. *Let H be a graph and fix a vertex $0 \in H$. For $r \in \mathbb{N}$, write $\mathcal{B}_H(r)$ for the r-ball in the graph H, ie $\mathcal{B}_H(r) := \{h \in H \mid d_H(0,h) \le r\}$, where d_H is the graph distance in H. Define*

$$\mathcal{D}_H(\beta) := \min\big\{r \ge 0 \,\big|\, |\mathcal{B}_H(r)| \ge \beta n\big\} \quad \text{for} \quad \beta \in (0,1).$$

When considering sequences $(k_N, G_N)_{N\in\mathbb{N}}$ of integers and Abelian groups, abbreviate

$$\mathcal{D}_N(\beta) := \mathcal{D}_{G_N([Z_1,\dots,Z_{k_N}])}(\beta) \quad \text{where} \quad Z_1,\dots,Z_{k_N} \sim^{\text{iid}} \text{Unif}(G_N).$$

Finally, considering such sequences, we define the candidate radius for the typical distance:

$$\mathfrak{D}_N^+ := k_N|G_N|^{1/k_N}/(2e) \quad \text{and} \quad \mathfrak{D}_N^- := k_N|G_N|^{1/k_N}/e \quad \text{for each} \quad N \in \mathbb{N}.$$

As always, if we write $\mathfrak{D}_N$, then this is either $\mathfrak{D}_N^+$ or $\mathfrak{D}_N^-$ according to context.

We show that, whp over the graph (ie choice of Z), this statistics concentrates. The result will be valid for all Abelian groups, under some conditions on k in terms of G. Further, the value at which the typical distance concentrates, written as $\mathfrak{D}^\pm$ below, depends only on k and $|G|$. This is in agreement with the spirit of the Aldous–Diaconis conjecture.

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

$$\liminf_{N\to\infty} |G_N| = \infty, \quad \limsup_{N\to\infty} k_N/\log|G_N| = 0, \quad \liminf_{N\to\infty}(k_N - d(G_N)) = \infty$$
$$\text{and} \quad \frac{k_N - d(G_N) - 1}{k_N} \ge 5\frac{k_N}{\log|G_N|} + 2\frac{d(G_N)\log\log k_N}{\log|G_N|} \quad \text{for all } N \in \mathbb{N}.$$

So we are only studying $1 \ll k \ll \log|G|$ here. In Remark 4.1.3 below, we give some sufficient conditions for Theorem 4.1.2 to hold. Throughout the proofs, we drop the subscript-N from the notation, eg writing k or $n = |G|$, considering sequences implicitly. Write $\mathcal{D}_k(\beta)$ for the β-typical distance of the random Cayley graph G_k.

We now state the main theorem of this section.

Theorem 4.1.2. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ a sequence of finite, Abelian groups; for each $N \in \mathbb{N}$, define $Z_{(N)} := [Z_1,\dots,Z_{k_N}]$ by drawing $Z_1,\dots,Z_{k_N} \sim^{\text{iid}} \text{Unif}(G_N)$.*

Suppose that $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis F. Then, for all $\beta \in (0,1)$, we have

$$\mathcal{D}_N^\pm(\beta)/\mathfrak{D}_N^\pm \to^{\mathbb{P}} 1 \quad \text{(in probability)} \quad \text{as } N \to \infty.$$

Moreover, the implicit lower bound holds deterministically, ie for all choices of generators, and for all Abelian groups, ie Hypothesis F need not be satisfied—we just need $\limsup_N k_N/\log|G_N| = 0$.

Remark 4.1.3. Write $n \coloneqq |G|$. Any of the following conditions imply Hypothesis F:

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

△

4.1.2 Outline of Proof

As remarked after the summarised statement (in §1.3.3.1), when considering the mixing of SRW on a graph, geometric properties of the graph are often derived and used. In a reversal of this, we use knowledge about the mixing properties of the random variable to derive a geometric result. We explain this in a little more detail now.

For the lower bound, for any Cayley graph $\mathcal{G}$ of an Abelian group of degree k, (trivially) we have $|\mathcal{B}_{\mathcal{G}}(R)| \leq |B_k(R)|$, where $B_k(R)$ is the k-dimensional lattice ball of radius R. If $|B_k(R)| \ll n$, then immediately $|\mathcal{B}_{\mathcal{G}}(R)| \ll n$, and so $\mathcal{D}_{\mathcal{G}}(\beta) \geq R$ for all $\beta \in (0,1)$, asymptotically in n.

Consider first the upper bound. We fix some target radius kL and draw $W_1, ..., W_k \sim^{\text{iid}} \text{Geom}(1/L)$ in the directed case. For the undirected case, we add to each W_i a uniform sign. It is well-known that the law of $W \coloneqq (W_1, ..., W_k)$ given $\|W\|_1 = R$ is uniform on the L_1 sphere of radius R. Since the $\|W\|_1 = \sum_1^k |W_i|$ is an iid sum, it concentrates around its mean, ie kL. So this is roughly like drawing uniformly from a sphere of radius kL, except that we have the added benefit that the coordinates $W_1, ..., W_k$ are independent.

We can then interpret W_i as the number of times which generator i is used in getting from the identity to $W \cdot Z$. We show that $W \cdot Z$ is well-mixed whp over Z when kL is slightly larger than the target radius. Now, if the law of $W \cdot Z$ is mixed in TV and $\|W\|_1 \leq kL(1+\delta)$ whp, then the law of $W \cdot Z$ conditional on $\|W\|_1 \leq kL(1+\delta)$ is also mixed in TV. Thus, using the concentration of $\|W\|_1$, we deduce that a proportion $1 - o(1)$ of vertices $x \in G$ can be written as $x = w \cdot Z$ for some w with $\|w\|_1 \approx kL$; this gives a path of length at most kL from the identity to x.

We show this mixing estimate via a (modified) L_2 argument. The most important part is to bound the probability that two independent copies of W are equal; this must be $o(1/n)$. Since $\|W\|_1$ concentrates and W is uniform on the sphere of this radius, we need to choose L so that the sphere of radius kL has volume slightly more than n. In high dimensions—here we consider balls in $k \gg 1$ dimensions—(discrete) spheres and balls are of asymptotically the same volume. Thus the desired radius coincides with that of the lower bound up to sot.

In an ideal world, we would directly sample W uniformly from a ball of radius kL. However, the lack of independence between the coordinate causes difficulties, in particular in Lemma 4.1.13 below. We thus use this vector of geometrics as a proxy for the uniform distribution, but with the key property that the coordinates are independent.

4.1.3 Estimates on Sizes of Balls in $\mathbb{Z}^k$

We desire an M so that $|B_k^{\pm}(M)| \approx n$, where $B_k^{\pm}(M)$ is the lattice ball of radius M, ie

$$B_k^-(M) \coloneqq \big\{w \in \mathbb{Z}^k \;\big|\; \|w\|_1 \leq M\big\} \quad \text{and} \quad B_k^+(M) \coloneqq \big\{w \in \mathbb{Z}_+^k \;\big|\; \|w\|_1 \leq M\big\}.$$

Definition 4.1.4. *Set* $\omega \coloneqq \max\{(\log k)^2,\, k/n^{1/(2k)}\}$. *Note that* $1 \ll \omega \ll k$. *Define*

$$\mathcal{R}_0^{\pm} \coloneqq \inf\big\{R \in \mathbb{N} \;\big|\; |B_k(R)| \geq ne^{\omega}\big\}.$$

The following lemma controls the size of balls. Its proof is given in §6.5; see in particular Lemmas 6.5.2a and 6.5.3a where the index q corresponds to a type of L_q lattice balls; take $q \coloneqq 1$ to recover the usual L_1 lattice balls here. Recall $\mathfrak{D}_k$ from Definition 4.1.1.

Lemma 4.1.5. *Assume that* $1 \ll k \ll \log n$. *For all* $\xi \in (0,1)$, *we have*

$$|\mathcal{R}_0 - \mathfrak{D}_k|/\mathfrak{D}_k \ll 1 \quad \textit{and} \quad \big|B_k\big(\mathfrak{D}_k(1-\xi)\big)\big| \ll n.$$

4.1.4 Lower Bound on Typical Distance

From the results in §4.1.3, it is straightforward to deduce the lower bound in Theorem 4.1.2.

Proof of Lower Bound in Theorem 4.1.2. Let $\xi \in (0,1)$ and set $R \coloneqq \mathcal{R}_k(1-\xi)$. Since the underlying group is Abelian, applying Lemma 4.1.5, we have $|\mathcal{B}_k(R)| \le |B_k(R)| \ll n$. Hence, for all $\beta \in (0,1)$ and all Z, we have $\mathcal{D}_k(\beta) \ge R = \mathcal{R}_k(1-\xi)$, asymptotically in n. □

4.1.5 Upper Bound on Typical Distance

The argument given here is in a similar vein to that of §2.1.6; there we analysed the mixing time of the random walk on the (random) Cayley graph. Let $\varepsilon > 0$ and set $L \coloneqq (1+3\varepsilon)\mathcal{R}_0/k$.

Draw $W = (W_i)_1^k \sim \mathrm{Geom}(1/L)^{\otimes k}$; later, we condition on $\|W\|_1 \le Lk$. Here the geometric random variable has support $\{1,2,\dots\}$. Define $\chi \coloneqq (\chi_i)_1^k$ as follows: in the undirected case, $\chi_i \sim^{\mathrm{iid}} \mathrm{Unif}(\{\pm 1\})$; in the directed case, $\chi_i \coloneqq 1$ for all i. Set $S \coloneqq (\chi W)\cdot Z$ where $(\chi W) = (\chi_i W_i)_1^k$. Define W' and χ' as independent copies of W and χ, respectively; set $S' \coloneqq (\chi' W')\cdot Z$.

In §2.1.6, a key ingredient was conditioning that the auxiliary variable W was 'typical' in a precise sense. There we were interested in the law of S, ie the random walk; the introduction of typicality was a tool to study this, and establishing mixing bounds for the random walk. Here, somewhat in reverse, we can choose which random variable we study.

Definition 4.1.6. *Abbreviate* $L_- \coloneqq \lceil L(1-\log k/\sqrt{k})\rceil$. *Define*

$$\mathcal{W} \coloneqq \big\{w \in \mathbb{Z}_+^k \;\big|\; L\big(1-\log k/\sqrt{k}\big)+1 \le \|w\|_1/k \le L,\ \max_i w_i \le 3L\log k\big\}.$$

When W *and* W' *are independent copies, write* $\mathsf{typ} \coloneqq \{W, W' \in \mathcal{W}\}$.

Lemma 4.1.7 (Typicality). *We have* $\mathbb{P}(W \in \mathcal{W}) \asymp 1$ *and hence* $\mathbb{P}(\mathsf{typ}) \asymp 1$.

Proof. We consider the three parts of typicality separately:
- the lower bound on $\|W\|_1$ holds with probability $1-o(1)$ by Chebyshev's inequality;
- the upper bound on $\|W\|_1$ holds with probability bounded away from 0 by Berry–Esseen;
- the upper bound on $\max_i W_i$ holds with probability $1-o(1)$ by the union bound. □

We control the L_2 distance between S conditional on $W \in \mathcal{W}$ and the uniform distribution.

Proposition 4.1.8. *Suppose that* (d,n,k) *jointly satisfy Hypothesis F. Then*

$$\mathbb{E}\big(\big\|\mathbb{P}_{G_k}\big(S \in \cdot \mid W \in \mathcal{W}\big) - \mathrm{Unif}(G)\big\|_2^2\big) = o(1),$$

where we recall that $\mathbb{P}_{G_k}(\cdot)$ *is the* random *law corresponding to the random Cayley graph* G_k.

We now have all the ingredients to prove the upper bound on typical distance.

Proof of Upper Bound in Theorem 4.1.2. Let $\overline{W}$ have the law of W conditional on $W \in \mathcal{W}$. By Proposition 4.1.8, the L_2 distance between $\overline{S} \coloneqq \overline{W}\cdot Z$ and $\mathrm{Unif}(G)$ is $o(1)$ whp over Z. Thus the support $\mathcal{S}$ of $\overline{S}$ is a proportion $1-o(1)$ of the vertices whp. In particular, there is a path of length at most Lk from id to all vertices in $\mathcal{S}$ whp, as $\|\overline{W}\|_1 \le Lk$ by definition of typicality. Hence $\mathcal{D}_k(\beta) \le Lk = (1+3\varepsilon)\mathcal{R}_0$ whp. Applying Lemma 4.1.5 then gives $(\mathcal{D}_k(\beta)-\mathfrak{D}_k)/\mathfrak{D}_k \le 4\varepsilon$ whp. □

The remainder of this subsection is devoted to proving Proposition 4.1.8. We have

$$\mathbb{E}\big(\big\|\mathbb{P}_{G_k}\big(S \in \cdot \mid W \in \mathcal{W}\big) - \mathrm{Unif}(G)\big\|_2^2\big) = n\,\mathbb{P}\big(S = S' \mid \mathsf{typ}\big) - 1.$$

First we control the probability that $\chi W = \chi' W'$, since in this case we necessarily have $S = S'$.

Lemma 4.1.9. *We have* $n\,\mathbb{P}(\chi W = \chi' W' \mid \mathsf{typ}) = o(1)$.

Proof. Recall that $L_- := \lceil L(1-\log k/\sqrt{k})\rceil$. Consider the directed case first, ie $\chi = 1 = \chi'$. Then

$$\begin{aligned}
\mathbb{P}\big(W = W',\, \mathsf{typ}\big) &\le \textstyle\sum_{w:\|w\|_1 \ge kL_-} \mathbb{P}\big(W = w = W'\big) \\
&= \textstyle\sum_{w:\|w\|_1 \ge kL_-} \mathbb{P}\big(W' = w\big) \prod_1^k \mathbb{P}\big(W_i = w_i\big) \\
&= \textstyle\sum_{w:\|w\|_1 \ge kL_-} \mathbb{P}\big(W' = w\big) \prod_1^k L^{-1}(1-L^{-1})^{w_i - 1} \\
&= \textstyle\sum_{w:\|w\|_1 \ge kL_-} \mathbb{P}\big(W' = w\big) \cdot L^{-k}(1-L^{-1})^{\|w\|_1 - k} \\
&\le L^{-k}(1-L^{-1})^{kL_-} = \big(L^{-1}(1-L^{-1})^{\lceil L(1-\sqrt{\log k/k})\rceil}\big)^k \\
&\le (eL)^{-k} \exp\big(\sqrt{k\log k}\big) \le n^{-1} e^{-3\varepsilon k/2},
\end{aligned}$$

with the final inequality using $L^+ \ge (1+2\varepsilon)n^{1/k}/e$, using Lemma 4.1.5. In the undirected case, we also need $\chi = \chi'$, which happens with probability 2^{-k}, and is independent of (W, W'). Hence the same inequality holds with the event $\{W = W'\}$ replaced by $\{\chi W = \chi' W'\}$, recalling that $L^- = \frac{1}{2}L^+$. Finally, $\mathbb{P}(\mathsf{typ}) \asymp 1$, and so Bayes's rule combined with the above calculation gives

$$\mathbb{P}\big(\chi W = \chi' W' \mid \mathsf{typ}\big) \le n^{-1} e^{-\varepsilon k} \ll 1/n. \qquad \square$$

The following lemma describing the distribution of $v \cdot Z$ for a given $v \in \mathbb{Z}^k$ is crucial.

Lemma 4.1.10. *For all $v \in \mathbb{Z}^k$ with $\gcd(v_1, ..., v_k, n) = \gamma$, we have*

$$v \cdot Z \sim \mathrm{Unif}(\gamma G).$$

We thus now need to control $|\gamma G|$.

Lemma 4.1.11. *For all Abelian groups H and all $\gamma \in \mathbb{N}$, we have*

$$|H|/|\gamma H| \le \gamma^{d(H)}.$$

These two lemmas were used in §2.1.6; see Lemmas 2.1.11 and 2.1.12 for proofs. Define

$$V := \chi W - \chi' W' \quad \text{and} \quad \mathfrak{g} := \gcd(V_1, ..., V_k, n).$$

Corollary 4.1.12. *We have*

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

Proof. The conditioning does not affect Z. The result follows from Lemmas 4.1.10 and 4.1.11. $\square$

Lemma 4.1.13. *Given Hypothesis F, we have $\mathbb{E}(\mathfrak{g}^d \mathbf{1}(V \ne 0) \mid \mathsf{typ}) = 1 + o(1)$.*

Proof. Each coordinate of V is unimodal and symmetric about 0.

$$\mathbb{P}\big(V_1 \in \gamma\mathbb{Z} \mid V_1 \ne 0\big) \le 1/\gamma,$$

as in Lemma 2.1.14. The probability of $V_1 = 0$ is roughly $1/(2L) \asymp n^{-1/k}$; in particular, it is at most $3n^{-1/k}$. The coordinates are independent. Since $\mathbb{P}(\mathsf{typ}) \asymp 1$, we thus have

$$\mathbb{P}\big(\mathfrak{g} = \gamma \mid \mathsf{typ}\big) \lesssim \big(1/\gamma + 3/n^{1/k}\big)^k.$$

By typicality, $\mathfrak{g} \le 6L\log k \le 3n^{1/k}\log k$. Hence, summing over γ, we obtain

$$\mathbb{E}\big(\mathfrak{g}^d \mathbf{1}(V \ne 0)\,|\big) \lesssim \textstyle\sum_{\gamma=1}^{3n^{1/k}\log k} \gamma^d \big(1/\gamma + 3/n^{1/k}\big)^k.$$

We handle almost exactly the same sum in Corollary 2.1.15. Hypothesis F here are designed precisely to control this sum; they are identical to Hypothesis A. There the $3/n^{1/k}$ part is replaced with $2/n^{1/k}$, but exactly the same arguments apply showing that the sum is $1 + o(1)$. $\square$

Proposition 4.1.8 now follows immediately from Lemmas 4.1.9 and 4.1.13 and Corollary 4.1.12.

Proof of Proposition 4.1.8. By Lemmas 4.1.9 and 4.1.13 and Corollary 4.1.12, we have

$$\begin{aligned}\mathbb{P}\big(S = S' \mid \mathsf{typ}\big) &\le \mathbb{P}\big(V = 0 \mid \mathsf{typ}\big) + \mathbb{P}\big(V \cdot Z = 0,\, V \neq 0 \mid \mathsf{typ}\big) \\ &\le \mathbb{P}\big(V = 0 \mid \mathsf{typ}\big) + \mathbb{E}\big(\mathfrak{g}^d\, \mathbf{1}(V \neq 0) \mid \mathsf{typ}\big) = 1 + o(1). \qquad \square\end{aligned}$$

4.2 Typical Distance: $k \asymp \log|G|$

This section focusses on concentration of distances from the identity in the random Cayley graph of an Abelian group when $k \asymp \log|G|$. (The previous section dealt with $1 \ll k \ll \log|G|$ and the next deal with $k \gg \log|G|$.) The main result of the section is Theorem 4.2.2; see also Hypothesis G.

The outline of this section is as follows:

- §4.2.1 states precisely the main theorem of the section;
- §4.2.2 outlines the argument;
- §4.2.3 gives some crucial estimates on the size of lattice balls;
- §4.2.4 is devoted to the lower bound;
- §4.2.5 is devoted to the upper bound under additional constraints;
- §4.2.6 describes how to relax these additional constraints;
- §4.2.7 describes an extension for L_1-type graph distances to L_q-type.

4.2.1 Precise Statement and Remarks

To start the section, we recall the typical distance statistic.

Definition 4.2.1. *Let H be a graph and fix a vertex $0 \in H$. For $r \in \mathbb{N}$, write $\mathcal{B}_H(r)$ for the r-ball in the graph H, ie $\mathcal{B}_H(r) := \{h \in H \mid d_H(0,h) \le r\}$, where d_H is the graph distance in H. Define*

$$\mathcal{D}_H(\beta) := \min\big\{r \ge 0 \;\big|\; |\mathcal{B}_H(r)| \ge \beta n\big\} \quad \text{for} \quad \beta \in (0,1).$$

When considering sequences $(k_N, G_N)_{N\in\mathbb{N}}$ of integers and Abelian groups, abbreviate

$$\mathcal{D}_N(\beta) := \mathcal{D}_{G_N([Z_1,\dots,Z_{k_N}])}(\beta) \quad \text{where} \quad Z_1,\dots,Z_{k_N} \sim^{\text{iid}} \mathrm{Unif}(G_N).$$

As always, if we write $\mathfrak{D}_N$, then this is either $\mathfrak{D}_N^+$ or $\mathfrak{D}_N^-$ according to context.

We show that, whp over the graph (ie choice of Z), this statistics concentrates. Here we consider $k \approx \lambda \log|G|$ for any $\lambda \in (0,\infty)$. The result holds for a large class of Abelian groups. Further, for these groups, the typical distance concentrates at $\alpha_\lambda k$ where $\alpha_\lambda \in (0,\infty)$ is a constant; so this depends only on k and $|G|$. This is in agreement with the spirit of the Aldous–Diaconis conjecture.

Recall that any Abelian group can be decomposed as $\oplus_1^d \mathbb{Z}_{m_j}$ for some $d, m_1, \dots, m_d \in \mathbb{N}$. For an Abelian group G, we define the *dimension* and *minimal side-length*, respectively, as follows:

$$\begin{aligned} d(G) &:= \min\big\{d \in \mathbb{N} \;\big|\; \oplus_1^d \mathbb{Z}_{m_j} \text{ is a decomposition of } G\big\}; \\ m_*(G) &:= \max\big\{\min_{j\in[d]} m_j \;\big|\; \oplus_1^d \mathbb{Z}_{m_j} \text{ is a decomposition of } G\big\}. \end{aligned}$$

It can be shown that there is a decomposition which is optimal for both these statistics: there exist $d, m_1, \dots, m_d \in \mathbb{N}$ so that $\oplus_1^d \mathbb{Z}_{m_j}$ is a decomposition of G with $d = d(G)$ and $\min_{j\in[d]} m_j = m_*(G)$. From now on, we assume that we are always using such an optimal decomposition.

There are some conditions which the Abelian groups must satisfy.

Hypothesis G. *The sequence $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis G if*

$$\begin{gathered} \lim_N k_N = \infty, \quad \lim_N k_N / \log|G_N| \in (0,\infty), \quad \liminf_N m_*(G_N) = \infty \\ \text{and} \quad d(G_N) \le \tfrac{1}{2} \log|G_N| / \log\log|G_N| \text{ for all } N. \end{gathered}$$

We are now ready to state the main theorem of this section.

Theorem 4.2.2. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ a sequence of finite, Abelian groups; for each $N \in \mathbb{N}$, define $Z_{(N)} \coloneqq [Z_1, ..., Z_{k_N}]$ by drawing $Z_1, ..., Z_{k_N} \sim^{\text{iid}} \text{Unif}(G_N)$.*

Suppose that $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis G. Let $\lambda \coloneqq \limsup_N k_N/\log|G_N|$. Then there exists a constant $\alpha_\lambda^\pm$ so that, for all $\beta \in (0,1)$, we have

$$\mathcal{D}_N^\pm(\beta)/(\alpha_\lambda^\pm k_N) \to^{\mathbb{P}} 1 \quad \textit{(in probability)} \quad \textit{as } N \to \infty.$$

Moreover, the implicit lower bound holds deterministically, ie for all choices of generators, and for all Abelian groups, ie Hypothesis G need not be satisfied—we just need $\lim_N k_N/\log|G_N| \in (0,\infty)$.

For ease of presentation, in the proof we drop the N-subscripts.

Remark 4.2.3. In §4.2.7 we extend the usual L_1-type graph distances to L_q-type. An analogous concentration of typical distance is given. See Hypotheses G′ and Theorem 4.2.11. △

4.2.2 Outline of Proof

The outline here is very similar to that from before; see §4.1.2. In particular, the lower bound is exactly the same idea. For the upper bound, we were trying to bound the expectation of a d-th power of a gcd. Issues arose when k became too large while $k-d$ is fairly small; see the proof of Lemma 4.1.13. This arose from the fact that we used the estimate

$$\mathbb{P}\big(V_1 \in \gamma\mathbb{Z}\big) \le \mathbb{P}\big(V_1 \in \gamma\mathbb{Z} \mid V_1 \neq 0\big) + \mathbb{P}\big(V_1 = 0\big) \le 1/\gamma + 3/n^{1/k}.$$

Once this was raised to the power k, the second term became an issue. We alleviate this by defining

$$\mathcal{I} \coloneqq \big\{i \in [k] \mid V_i \neq 0\big\}$$

and studying $\mathbb{P}(V_i \in \gamma\mathbb{Z} \mid i \in \mathcal{I})$; the problematic term $3/n^{1/k}$ then does not exist. If $G = \oplus_1^d \mathbb{Z}_{m_j}$, then we are actually interested in $V_i \bmod m_j$ for each j. Recall that $m_* = \min_j m_j$. 'Typically', one has $|V_i| \le m_*$. We assume m_* is sufficiently large so that $\max_i |V_i| < m_*$ whp. Thus looking at $V_i = 0$ or $V_i \equiv 0 \bmod m_j$ is no different.

For large $|I|$, the gcd analysis goes through similarly to before. When $|I|$ is small, eg smaller than d, it is more difficult to control; in this case, we use a fairly naive bound on the gcd, but control carefully the probability of realising such an $\mathcal{I}$. The case $\mathcal{I} = \emptyset$ corresponding to $V = 0$, which is handled by taking the ball to be of large enough volume.

Previously we used a vector of geometrics as a proxy for a uniform distribution on a ball. Here we are able to let W be uniform on a ball. The coordinates are no longer independent, which makes the gcd analysis is slightly complicated. However, since we only consider i with $V_i \neq 0$, this can be handled; see Lemma 4.2.9. This uniformity simplifies some other calculations somewhat.

4.2.3 Estimates on Sizes of Balls in $\mathbb{Z}^k$

We wish to determine the size of balls $B_k(R)$ when $k \asymp \log n$. In particular, we are interested in the growth when the volume is around n.

Definition 4.2.4. *Define $M_*^\pm(k,N)$ to be the minimal integer M satisfying $|B_k^\pm(M)| \ge N$.*

Lemma 4.2.5. *For all $\lambda > 0$, there exists a function $\omega \gg 1$ and a constant $\alpha^\pm$ so that, for all $\varepsilon \in (0,1)$, if $k \asymp \lambda \log n$, then $M_*^\pm \coloneqq M_*^\pm(k, ne^\omega)$ satisfies*

$$M_*^\pm \asymp \alpha^\pm k \asymp \alpha^\pm \lambda \log n \quad \textit{and} \quad \big|B_k^\pm\big(\alpha^\pm k(1-\varepsilon)\big)\big| \ll n.$$

(There, the $\pm$-superscript indicates that α takes different values in the un/directed cases.)

This will follow easily from the following auxiliary lemma controlling the size of lattice balls.

Lemma 4.2.6. *There exists a strictly increasing, continuous function $c^{\pm} : (0,\infty) \to (0,\infty)$ so that, for all $a \in (0,\infty)$, we have*

$$\left|B_k^{\pm}(ak)\right| = \exp\big(k\big(c^{\pm}(a) + o(1)\big)\big).$$

Proof. The directed case follows immediately from Stirling's approximation and the fact that

$$\left|B_k^{+}(ak)\right| = \left|\left\{b \in \mathbb{Z}_+^k \;\middle|\; \textstyle\sum_1^k b_i \le ak\right\}\right| = \binom{\lfloor ak \rfloor + k}{k} = \binom{\lfloor (a+1)k \rfloor}{k}$$

Consider now the undirected case. Omit all floor/ceiling signs. By considering the number i of coordinates which equal 0, we obtain

$$\left|B_k^{-}(ak)\right| = \textstyle\sum_{i=0}^{k} A_i \quad \text{where} \quad A_i := A_i(k,a) := \binom{k}{i} 2^{k-i} \binom{k-i+ak}{ak}.$$

Choose $i_* := i_*(k,a)$ that maximises A_i. Then $A_{i_*} \le |B_k^{-}(ak)| \le (k+1)A_{i^*}$. Observe that

$$\frac{A_{i+1}}{A_i} = \frac{(k-i)^2}{2(i+1)(k(1+a)-i)},$$

and hence one can determine i_* as a function of k and a, conclude that $i_*(a,k)/k$ converges as $k \to \infty$ and thus determine $c^{+}(a)$ in terms of the last limit. We omit the details. Knowing this limit allows us to plug this into the definition of A_i and use Stirling's approximation to get

$$A_{i_*} = \exp\big(k\big(c^{-}(a) + o(1)\big)\big),$$

for some strictly increasing function $c^{-} : (0,\infty) \to (0,\infty)$. Since $k+1 = e^{o(k)}$, the claim follows. □

From this lemma, Lemma 4.2.5 follows easily.

Proof of Lemma 4.2.5. Set $\alpha := c^{-1}(1/\lambda)$. The upper bound is an immediate consequence of the continuity of c. The lower bound follows from the exponential growth rate. □

4.2.4 Lower Bound on Typical Distance

From the results in §4.2.3, it is straightforward to deduce the lower bound in Theorem 4.2.2.

Proof of Lower Bound in Theorem 4.2.2. Let $\xi \in (0,1)$ and set $R := \alpha_\lambda^{\pm} k(1-\xi)$. Since the underlying group is Abelian, applying Lemma 4.2.5, we have $|\mathcal{B}_k^{\pm}(R)| \le |B_k^{\pm}(R)| \ll n$. Hence, for all $\beta \in (0,1)$ and all Z, we have $\mathcal{D}_k^{\pm}(\beta) \ge R = \alpha_\lambda^{\pm} k(1-\xi)$, asymptotically in n. □

4.2.5 Upper Bound on Typical Distance

Define $M_*^{\pm}$, ω and $\alpha^{\pm}$ as in Definition 4.2.4 and Lemma 4.2.5. In this subsection we draw $W^{\pm} \sim \mathrm{Unif}(B_k^{\pm}(M_*^{\pm}))$, ie uniform on a ball of radius $M_*^{\pm}$. We show that $W^{\pm} \cdot Z$ is well-mixed on G, and hence its support contains almost all the vertices.

Proposition 4.2.7. *Suppose that (k,n,d,m) satisfy Hypothesis G. Then*

$$\mathbb{E}\Big(\big\|\mathbb{P}_{G_k}\big(W^{\pm} \cdot Z \in \cdot\big) - \pi_G\big\|_2^2\Big) = o(1),$$

Given this proposition, the upper bound in Theorem 4.2.2 follows easily.

Proof of Upper Bound in Theorem 4.2.2 Given Proposition 4.2.7. If $\|\mathbb{P}_{G_k}(W^{\pm} \cdot Z \in \cdot) - \pi_G\|_2 \le \varepsilon$, then the support $\mathcal{S}$ of $W^{\pm} \cdot Z$ satisfies $\pi_G(\mathcal{S}^c) \le \varepsilon$. Combined with Lemma 4.2.5 and Proposition 4.2.7, the upper bound in Theorem 4.2.2 follows. □

The remainder of this subsection is devoted to proving Proposition 4.2.7. We tend to drop the $\pm$-superscript from the notation, only writing $+$ or $-$ if there is ambiguity. Let $W, W' \sim^{\text{iid}} \mathrm{Unif}(B_k(M_*))$ and let $V \coloneqq W - W'$. The standard L_2 calculation gives

$$\mathbb{E}\big(\big\|\mathbb{P}_{G_k}\big(W \cdot Z \in \cdot\big) - \pi_G\big\|_2^2\big) = \mathbb{E}\big(n\,\mathbb{P}\big(V \cdot Z = 0 \mid Z\big) - 1\big) = n\,\mathbb{P}\big(V \cdot Z = 0\big) - 1.$$

First, it is immediate that $\mathbb{P}(V = 0) = \mathbb{P}(W = W') = |B_k(M_*)|^{-1} \le n^{-1}e^{-\omega} \ll n^{-1}$. Now consider $V \neq 0$. As in §4.1.5, it is key to analyse certain gcds. Specifically, set

$$\mathfrak{g}_j \coloneqq \gcd\big(V_1, ..., V_k, m_j\big) \quad \text{for each} \quad j \in [d]; \quad \text{set} \quad \mathfrak{g} \coloneqq \gcd\big(V_1, ..., V_k, n\big).$$

The following lemma is an immediate application of (the analogous) Lemma 4.1.10.

Lemma 4.2.8. *Conditional on V, we have $V \cdot Z \sim \mathrm{Unif}(\oplus_1^d \mathfrak{g}_j\mathbb{Z}_{m_j})$.*

Since the minimal side-length m_* satisfies $m_* \gg k \asymp M_*$, we have $\max_{i\in[k]} |V_i| < \max_{j\in[d]} m_j$. An immediate corollary of this is that

$$\mathcal{I} \coloneqq \big\{i \in [k] \;\big|\; V_i \not\equiv 0 \bmod m_j \;\forall\, j \in [d]\big\} = \big\{i \in [k] \;\big|\; W_i \neq W_i'\big\}.$$

To analyse the expected gcd, we breakdown according to the value of $\mathcal{I}$.

Lemma 4.2.9. *There exists a constant C so that, for all $I \subseteq [k]$ with $I \neq \emptyset$, we have*

$$n\,\mathbb{P}\big(V \cdot Z = 0 \mid \mathcal{I} = I\big) \le \mathbb{E}\big(\mathfrak{g}^d \mid \mathcal{I} = I\big) \le \begin{cases} C2^d(2M_*)^{d-|I|+2} & \text{when} \quad |I| \le d+1, \\ 1 + 5\cdot(\frac{3}{2})^{2d-|I|} & \text{when} \quad |I| \ge d+2. \end{cases}$$

Lemma 4.2.10. *For all $I \subseteq [k]$ with $|I| \ll k$, we have $\mathbb{P}(\mathcal{I} = I) \le e^{-\omega}n^{-1+o(1)}$. If $I = \emptyset$, then the $o(1)$ term may be taken to be 0.*

Given these two lemmas, we have all the ingredients required to prove Proposition 4.2.7, from which we deduced the main theorem (Theorem 4.2.2). We defer the proofs of Lemmas 4.2.9 and 4.2.10 until after the proof of Proposition 4.2.7, which we give now.

Proof of Proposition 4.2.7. Here $k \asymp \lambda \log n$, $M \coloneqq M_* \asymp \alpha k \asymp \alpha\lambda \log n$ and $d \le \frac{1}{2}\log n/\log\log n$. As noted previously, the standard L_2 calculation gives

$$\begin{aligned}\mathbb{E}\big(\big\|\mathbb{P}_{G_k}\big(W \cdot Z \in \cdot\big) - \pi_G\big\|_2^2\big) &= \mathbb{E}\big(n\,\mathbb{P}\big(V \cdot Z = 0 \mid Z\big) - 1\big) \\ &= n\,\mathbb{P}\big(V \cdot Z = 0\big) - 1 = n\textstyle\sum_{I\subseteq[k]} \mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) - 1.\end{aligned}$$

Consider $I = \emptyset$. Then $V \cdot Z = 0$ (for all Z). By Lemma 4.2.10, we have $\mathbb{P}(\mathcal{I} = \emptyset) \le n^{-1}e^{-\omega}$. Thus

$$n\,\mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = \emptyset\big) \le e^{-\omega} = o(1).$$

Consider $I \subseteq [k]$ with $1 \le |I| \le d+1$. There are at most $(d+1)\binom{k}{d+1} \le k^{d+2}$ such sets I. Since $\log k = \log\log n + \log\lambda + o(1)$, we have $k^{d+2} \le n^{2/3}$. Applying Lemmas 4.2.9 and 4.2.10 gives

$$n\,\mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) \le C2^d(3\alpha\lambda\log n)^{d+2-|I|} \cdot n^{-1+o(1)} \le k^{-d-2}n^{-1/4},$$

noting that $2^d = n^{o(1)}$. We now sum over all I with $1 \le |I| \le d+1$:

$$n\textstyle\sum_{1\le|I|\le d+1} \mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) \le n^{-1/4} = o(1).$$

Consider $I \subseteq [k]$ with $d+2 \le |I| \le L \coloneqq \frac{2}{3}\log n/\log\log n$; then $L - 2d \gg 1$. Similarly to above, there are at most $L\binom{k}{L} \le k^{L+1}$ such sets I. Applying Lemmas 4.2.9 and 4.2.10 gives

$$n\,\mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) \le n^{-1+o(1)} \le k^{-L-1}n^{-1/4}.$$

We now sum over all I with $d+2 \le |I| \le L$:

$$n \sum_{d+2 \le |I| \le L} \mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) \le n^{-1/4} = o(1).$$

Finally consider $I \subseteq [k]$ with $|I| \ge L$. Sum over these using Lemma 4.2.9:

$$n \sum_{L \le |I| \le k} \mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) \le 1 + 5 \cdot (\tfrac{3}{2})^{2d-L} = 1 + o(1).$$

Combining these four parts into a single sum, we deduce the result. □

It remains to prove the auxiliary Lemmas 4.2.9 and 4.2.10.

Proof of Lemma 4.2.9. The first inequality is an immediate consequence of Lemma 4.2.8.

Note that $\mathfrak{g} \le 2M_*$ since $\max_i |V_i| \le 2M_*$. For $\alpha, \beta \in \mathbb{Z}$, write $\alpha \wr \beta$ if α divides β. Thus

$$\mathbb{E}\big(\mathfrak{g}^d \mid \mathcal{I} = I\big) \le \sum_{\gamma=1}^{2M} \gamma^d \, \mathbb{P}\big(\gamma \wr V_i \,\forall\, i \in I \mid \mathcal{I} = I\big)$$

For a set $I \subseteq [k]$, write $W_I \coloneqq (W_i)_{i \in I}$ and $W_{\setminus I} \coloneqq W_{[k] \setminus I}$. Consider conditioning on $\mathcal{I} = I$. Let $W_{\setminus I}$ and $W'_{\setminus I}$ be given; since $\mathcal{I} = I$, we have $W_{\setminus I} = W'_{\setminus I}$. Let U have the distribution of W_I given $W_{\setminus I}$ and define U' analogously. Write $D_i \coloneqq D_i(\gamma) \coloneqq \{\gamma \wr (U_i - U'_i)\}$. Then

$$\mathbb{P}\big(\gamma \wr V_i \,\forall\, i \in I \mid \mathcal{I} = I,\, \|W_{\setminus I}\|_1\big) = \mathbb{P}\big(D_i \,\forall\, i \in I\big).$$

By exchangeability, it suffices to consider the case $I = \{1, ..., \ell\}$. We then have

$$\mathbb{P}\big(D_i \,\forall\, i \in I\big) = \mathbb{P}\big(D_\ell\big)\mathbb{P}\big(D_{\ell-1} \mid D_\ell\big) \cdots \mathbb{P}\big(D_1 \mid D_2, ..., D_\ell\big) = \prod_{i=1}^{\ell} \mathbb{P}\big(D_i \mid D_{i+1}, ..., D_\ell\big).$$

For $i \in [k]$, define $M_i \coloneqq M_* - \|W_{\setminus\{1,...,i\}}\|_1$ and M'_i analogously. Let $i \in [\ell - 1]$. Let $(u_{i+1}, ..., u_\ell)$ and $(u'_{i+1}, ..., u'_\ell)$ be two vectors in the support of $(U_{i+1}, ..., U_\ell)$. Then,

$$\begin{gathered}\text{conditional on} \quad (U_{i+1}, ..., U_\ell) = (u_{i+1}, ..., u_\ell) \quad \text{and} \quad (U'_{i+1}, ..., U'_\ell) = (u'_{i+1}, ..., u'_\ell) \\ \text{we have} \quad (U_1, ..., U_i) \sim \mathrm{Unif}\big(B_i(R)\big) \quad \text{and} \quad (U'_1, ..., U'_i) \sim \mathrm{Unif}\big(B_i(R')\big) \quad \text{for some} \quad R, R' \in \mathbb{R}.\end{gathered}$$

(Recall that the subscript in B_k denotes the dimension of the ball.)

In the case of undirected balls, the law of $U_i - U'_i$ given this conditioning is symmetric and unimodal on $\mathbb{Z} \setminus \{0\}$; see [68, Theorem 2.2]. It follows, as in the proof of Lemma 4.1.13, that

$$\mathbb{P}\big(D_i^- \mid D_{i+1}^-, ..., D_\ell^-\big) \le 1/\gamma.$$

Further, this holds not just conditional on $D_{i+1}^- \cap \cdots \cap D_\ell^-$, but conditional on any choice of $(U_{i+1}, ..., U_\ell)$ and $(U'_{i+1}, ..., U'_\ell)$ which satisfy $D_{i+1}^- \cap \cdots \cap D_\ell^-$. By the same reasoning, $\mathbb{P}(D_\ell^-) \le 1/\gamma$. Hence, for undirected balls,

$$\mathbb{P}\big(D_i^- \,\forall\, i \in I\big) = \mathbb{P}\big(\gamma \wr V_i^- \,\forall\, i \in I \mid \mathcal{I} = I\big) \le \gamma^{-|I|}.$$

(The $-$-superscript emphasises that this is for undirected balls.)

We now turn our attention to directed balls. In this case, U_i and U'_i are both unimodal, but with potentially different modes, if $R \ne R'$. Instead of direct computation, we compare with the undirected case. Specifically, if U_i and U'_i have the same sign in the undirected case, then $|V_i| = |U_i - U'_i|$ has the same law as in the directed case. The choice of sign is independent of everything else; the two have the same sign with probability $\frac{1}{2}$. Hence, by conditioning on the specific values of $(U_{i+1}, ..., U_\ell)$ and $(U'_{i+1}, ..., U'_\ell)$, we obtain

$$1/\gamma \ge \mathbb{P}\big(D_i^- \mid D_{i+1}^-, ..., D_\ell^-\big) \ge \tfrac{1}{2}\mathbb{P}\big(D_i^+ \mid D_{i+1}^+, ..., D_\ell^+\big).$$

For $\gamma = 2$, note that the probabilities are actually the same: this is because $x - y$ is even if and only if $|x| - |y|$ is even, since x and $-x$ have the same parity.

From this we deduce, for both the undirected and directed cases, that

$$\mathbb{E}\big(\mathfrak{g}^d \mid \mathcal{I} = I\big) \le 1 + 2^{d-|I|} + \sum_{\gamma=3}^{2M} \gamma^d (2/\gamma)^{|I|} = 1 + 2^{d-|I|} + 2^d \sum_{\gamma=3}^{2M} (\gamma/2)^{d-|I|}.$$

A case-by-case analysis, according to $d - |I|$, completes the proof. □

Proof of Lemma 4.2.10. Recall from Definition 4.2.4 that $|B_k(M_*)| \geq ne^{\omega}$. Thus

$$\mathbb{P}\big(\mathcal{I} = \emptyset\big) = \mathbb{P}\big(W = W'\big) = \big|B_k(M_*)\big|^{-1} \leq n^{-1}e^{-\omega}.$$

Using the law of W_I given $W_{\setminus I}$ determined in the previous proof, we have

$$\mathbb{P}\big(W_{\setminus I} = W'_{\setminus I}\big) = \frac{\mathbb{P}(W = W')}{\mathbb{P}(W = W' \mid W_{\setminus I} = W'_{\setminus I})} = \frac{|B_k(M_*)|^{-1}}{\mathbb{E}(|B_{|I|}(M_* - \|W_{\setminus I}\|_1)|^{-1})} \leq \frac{|B_{|I|}(M_*)|}{|B_k(M_*)|}.$$

It is a standard balls-in-bins combinatorial identity that

$$\big|B^+_\ell(R)\big| = \big|\big\{b \in \mathbb{Z}^\ell_+ \;\big|\; \textstyle\sum_1^\ell b_i \leq R\big\}\big| = \binom{\lfloor R \rfloor + \ell}{\ell}.$$

For the undirected case, we can choose a sign $\pm b_i$. Hence we see that

$$\big|B^+_\ell(R)\big| \leq \big|B^-_\ell(R)\big| = \big|\big\{b \in \mathbb{Z}^\ell \;\big|\; \textstyle\sum_1^\ell |b_i| \leq R\big\}\big| \leq 2^\ell \binom{\lfloor R \rfloor + \ell}{\ell}.$$

Abbreviate $M := M_*$ and $\ell := |I|$. It suffices to consider I with $\ell \leq ck$, for an arbitrarily small positive constant c. From Lemma 4.2.5, we have $M \leq 2\alpha k$. So

$$\big|B^\pm_\ell(M)\big| \leq 2^\ell \binom{\lfloor M \rfloor + \ell}{\ell} \leq \big(2e(2\alpha k/\ell + 1)\big)^\ell \leq (8e\alpha k/\ell)^\ell,$$

with the last inequality requiring $2\alpha k/\ell \geq 1$, which holds if c is sufficiently small, as $\ell \leq ck$. Now, for c sufficiently small, the map $\ell \mapsto (8e\alpha k/\ell)^\ell$ is increasing on $[1, ck]$. Hence

$$\big|B^\pm_\ell(M)\big| \leq (8e\alpha k/\ell)^\ell \leq (8e\alpha/c)^{ck} \leq (8e\alpha/c)^{2c\lambda \log n}.$$

By taking c sufficiently small, we can upper bound this by an arbitrarily small power of n. □

4.2.6 Relaxing Condition on Minimal Side-Length $m_*(G)$

For the upper bound, we have been assuming that the minimal side length $m_*(G)$ satisfies $m_*(G) \gg \log|G|$. (Recall that the lower bound had no conditions on $m_*(G)$.) We now describe how to relax this condition to $m_*(G) \gg 1$. We could go even further, with statements like "only a small number of j in $G = \oplus_1^d \mathbb{Z}_{m_j}$ have $m_j \asymp 1$". Since we have no reason to believe our other conditions are optimal, we settle for the simpler $m_*(G) \gg 1$.

In this proof we consider both L_1 and L_∞ balls. To distinguish these we use a superscript:

- $B^1_\ell(R)$ will be the L_1 ball in ℓ dimensions of radius R;
- $B^\infty_\ell(R)$ will be the L_∞ ball in ℓ dimensions of radius R.

For a set $I \subseteq [k]$ we write $W_I := (W_i)_{i \in I}$ and $W_{\setminus I} := (W_i)_{i \notin I}$.

We describe the adaptations for *undirected* graphs. The adaptations for *directed* graphs are completely analogous: simply replace appearances of $\mathbb{Z}^k$ with $\mathbb{Z}^k_+$ and $|W_i|$ with W_i.

Outline of Proof. The idea behind the proof is intuitive. Since $R \asymp k$, by symmetry we have $\mathbb{E}(|W_i|) \leq R/k \asymp 1$ for all i. Thus 'almost all' the coordinates should be smaller than any diverging function ('good'). Further, the contribution to the radius $\|W\|_1$ due to these 'bad' coordinates should be small, ie $o(k)$. Roughly this allows us to replace k with $\overline{k} = k(1 - o(1))$ and R with $\overline{R} = R(1 - o(1))$. Choosing $R := \alpha_{k/\log n} k \cdot (1 + 2\varepsilon)$ for $\varepsilon > 0$ then gives

$$\overline{R} \geq \alpha_{\overline{k}/\log n}\overline{k} \cdot (1 + \varepsilon) \quad \text{and hence} \quad |B^1_{\overline{k}}(\overline{R})| \gg n.$$

This was the key element in the proof previously, and the remainder of the proof is as before. □

We now proceed formally and rigorously.

Relaxing Minimal Side-Length Condition. Let $\varepsilon > 0$ and $\lambda \coloneqq \lim k/\log n$. Set $R \coloneqq \alpha_\lambda k(1+2\varepsilon)$ and draw $W \sim \mathrm{Unif}(B_{k,1}(R))$. Let ν satisfy $1 \ll \nu \ll m_*(G)$. For $w \in \mathbb{Z}^k$, define

$$\mathcal{J}(w) \coloneqq \big\{ i \in [k] \;\big|\; |w_i| \le \nu \big\}.$$

These are the 'good' coordinates. By Markov's inequality, clearly $|[k] \setminus \mathcal{J}(W)| \lesssim 1/\nu = o(1)$ whp.

As always, we look at two independent realisations W and W'. We then wish to look at coordinates $i \in [k]$ which are 'good' for both W and W', ie in $\overline{\mathcal{J}} \coloneqq \mathcal{J}(W) \cap \mathcal{J}(W')$. We need to make sure that the contribution to the radius from the (abnormally large) 'bad' coordinates is not too large. For $\delta > 0$ and $w \in \mathbb{Z}^k$, write $\mathcal{L}_\delta(w)$ for the collection of the $\lceil 2\delta k \rceil$-largest (in absolute value) coordinates of w. We then define typicality in the following way: for $\delta, \delta' > 0$, set

$$\mathcal{W} \coloneqq \big\{ w \in \mathbb{Z}^k \;\big|\; \|w\|_1 \le R,\ \big|[k] \setminus \mathcal{J}(w)\big| \le \delta k,\ \|w_{\mathcal{L}_\delta(w)}\| \le \delta' k \big\}.$$

In particular now, if $w, w' \in \mathcal{W}$, then $\|w_{\mathcal{J}(w)\cap\mathcal{J}(w')}\|_1 \ge k - 2\delta' k$. It is not difficult to see that we can choose $\delta, \delta' = o(1)$ with $\mathbb{P}(W \in \mathcal{W}) = 1 - o(1)$; we give justification at the end of the proof.

Consider now $W, W' \sim^{\mathrm{iid}} \mathrm{Unif}(B_{k,1}(R))$. We have the following conditional law:

$$W_{\overline{J}}, W'_{\overline{J}} \sim^{\mathrm{iid}} \mathrm{Unif}\big(B^1_{\overline{k}}(\overline{R}) \cap B^\infty_{\overline{k}}(\nu)\big) \quad \text{conditional on} \quad W_{\setminus J} = w_{\setminus J} = W'_{\setminus J} \quad \text{and} \quad \overline{\mathcal{J}} = \overline{J}$$
$$\text{where} \quad \overline{\mathcal{J}} \coloneqq \mathcal{J}(W) \cap \mathcal{J}(W'), \quad \overline{k} \coloneqq |\overline{\mathcal{J}}| \quad \text{and} \quad \overline{R} \coloneqq R - \|w_{\setminus \overline{J}}\|.$$

Write $\mathsf{typ} \coloneqq \{W, W' \in \mathcal{W}\}$. On the event typ, given $\overline{\mathcal{J}} = \overline{J}$ and $(W_{\overline{J}}, W'_{\overline{J}})$, we have

$$\overline{k} \ge k(1-\delta) = k\big(1 - o(1)\big) \quad \text{and} \quad \overline{R} \ge R(1-\delta') = R\big(1 - o(1)\big).$$

In particular, we may choose $\eta > 0$ sufficiently small but constant (depending on ε) so that

$$\overline{R} \ge \alpha_{\lambda(1-\eta)} k(1-\eta)(1+\varepsilon) \quad \text{and} \quad \overline{k} \ge k(1-\eta), \quad \text{and hence} \quad |B^1_{\overline{k}}(\overline{R})| \gg n.$$

Since typicality holds with probability $1 - o(1)$, we have

$$\big|B^1_{\overline{k}}(\overline{R}) \cap B^\infty_{\overline{k}}(\nu)\big| \gg n.$$

The remainder of the proof follows similarly to before. Formally, we define $\overline{W}$ and $\overline{W}'$ as follows:

$$\overline{W}_i \coloneqq W_i \quad \text{and} \quad \overline{W}'_i \coloneqq W'_i \quad \text{for} \quad i \in \overline{\mathcal{J}};$$
$$\overline{W}_i \coloneqq 0 \quad \text{and} \quad \overline{W}'_i \coloneqq 0 \quad \text{for} \quad i \notin \overline{\mathcal{J}}.$$

Since this is a projection, $\{W_I = W'_I\} \subseteq \{\overline{W}_I = \overline{W}_I\}$ for any $I \subseteq [k]$. Now instead of decomposing according to the value (or size) of $\mathcal{I} \coloneqq \{i \in [k] \mid W_i \ne 0\}$, we use the set $\overline{\mathcal{I}} \coloneqq \mathcal{I} \cap \overline{\mathcal{J}}$. The fact that $\big|B^1_{\overline{k}}(\overline{R}) \cap B^\infty_{\overline{k}}(\nu)\big| \gg n$ allows all the previous estimates for $\mathcal{I}$ to follow through for $\overline{\mathcal{I}}$ here.

The last change to mention is the gcd calculations of Lemma 4.2.9. The only property of the distribution of (W, W') required was that each coordinate (while not independent) is unimodal and symmetric about 0, even conditional on $W_I = W'_I$ and $W'_I = w'_I$ for some $I \subseteq [k]$ and $w_I, w'_I \in \mathbb{Z}^{|I|}$. For $(\overline{W}, \overline{W}')$, this property still holds. Hence the identical argument applies here too.

It remains to argue that $\mathbb{P}(W \in \mathcal{W}) = 1 - o(1)$ for some $\delta, \delta' = o(1)$. First, as noted above, $\mathbb{P}(|[k] \setminus \mathcal{J}(W)| > \delta k) = o(1)$ by Markov's inequality and the fact that $\mathbb{E}(|W_1|) \asymp 1$. The fact that $\mathbb{P}(\|W_{\mathcal{L}_\delta(W)}\| > \delta k) = o(1)$ follows by a union bound over all $\binom{k}{\lceil 2\delta k \rceil}$ possible values of the set $\mathcal{L}_\delta(W)$ and applying Bernstein's inequality; take $\delta' \coloneqq C\delta \log(1/\delta)$ for a sufficiently large constant C. □

Remark. We believe that the typical distance should concentrate if $k \asymp \log|G|$ and $k - d \gg 1$ without any condition on like that on $m_*(G)$. However, without any such condition, we do have reason to believe that the *value* at which this concentration happens should depend on more than just k and $|G|$—the algebraic structure of G should be important. This exact phenomenon occurs when studying the mixing time of the random walk on the Cayley graph. See Theorem A, in particular contrasting the cases $k \asymp \log|G|$ and $1 \ll k \ll \log|G|$. △

4.2.7 Typical Distances for L_q-Type Graph Distances

Graphs distances in Cayley graphs have some special properties. Consider a collection $z = [z_1, ..., z_k]$ of generators and distances in the Cayley graph $G(z)$. For a path ρ in $G(z)$, for each $i \in [k]$, write $\rho_{i,+}$ for the number of times z_i is used, $\rho_{i,-}$ for the number of times z_i^{-1} is used (if in the undirected case, otherwise $\rho_{i,-} := 0$) and $\rho_i := \rho_{i,+} - \rho_{i,-}$. The path connects the identity with $\rho \cdot z$. Then the length, in the usual graph distance, of ρ is $\|\rho\|_1 := \sum_1^k (\rho_{i,+} + \rho_{i,-})$.

For any $q \in [1, \infty)$, define the L_q *graph distance* of ρ by $\|\rho\|_q^q := \sum_1^k (\rho_{i,+}^q + \rho_{i,-}^q)$. For the L_∞-*graph distance*, define $\|\rho\|_\infty := \max_i (\rho_{i,+} + \rho_{i,-})$. (The usual graph distance is given by $q = 1$.)

For Abelian groups, clearly for any $q \in [1, \infty)$ an L_q *geodesic*, ie a path of minimal length, will only use either z_i or z_i^{-1}, not both (since the terms in the product can be reordered), ie $\rho_{i,+}\rho_{i,-} = 0$ for all i. Thus $\|\rho\|_q^q = \sum_1^k |\rho_i|^q$. Similarly, any L_∞-geodesic ρ can be adjusted into a new path ρ' with $\|\rho\|_\infty = \|\rho'\|_\infty$ and $\rho'_{i,+}\rho'_{i,-} = 0$ for all i.

We define the L_q *typical distance* $\mathcal{D}_{G(z),q}(\cdot)$ analogously to $\mathcal{D}_{G(z)}(\cdot)$, ie the $q = 1$ case. When the k generators are chosen uniformly at random, we write $\mathcal{D}^{\pm}_{k,q}(\cdot)$, with the $\pm$-superscript indicating whether or not the Cayley graph is directed.

Hypothesis G′. *The sequence $(k_N, G_N)_{N\in\mathbb{N}}$ and $q \in [1, \infty]$ jointly satisfy Hypotheses G′ if the following conditions hold (defining $k^{1/\infty} := 1$ for $k \in \mathbb{N}$):*

$$\lim_N k_N = \infty, \quad \lim_N k_N / \log|G_N| = 0 \quad \text{and} \quad \lim_N k_N^{1/q} |G_N|^{1/k_N} / m_*(G_N) = 0;$$
$$\text{if} \quad q \in (1, \infty) \quad \text{then additionally} \quad k_N \le \log|G_N| / \log\log|G_N| \text{ for all } N \in \mathbb{N};$$
$$\limsup_N d_N / k_N < \begin{cases} 1 & \text{for undirected graphs,} \\ \frac{1}{2} & \text{for directed graphs.} \end{cases}$$

Finally we set up a little more notation. Make the following definitions:

$$C_q^- := 2\,\Gamma(1/q + 1)(qe)^{1/q}, \quad C_q^+ := \tfrac{1}{2} C_q^- \quad \text{and} \quad \mathfrak{D}_q^{\pm}(k, n) := k^{1/q} n^{1/k} / C_q^{\pm},$$

where the case $q = \infty$ is to be interpreted as the limit $q \to \infty$; eg, $C_\infty^- = 2$ and $\mathfrak{D}_\infty^+(k, n) = n^{1/k}$. When these are sequences $(k_N, |G_N|)_{N\in\mathbb{N}}$, for $N \in \mathbb{N}$ and $q \in [1, \infty]$, write $\mathfrak{D}^{\pm}_{N,q} := \mathfrak{D}^{\pm}_q(k_N, |G_N|)$.

Similarly, for a sequence $(G_N)_{N\in\mathbb{N}}$ of finite groups with corresponding multisubsets $(Z_{(N)})_{N\in\mathbb{N}}$ of sizes $(k_N)_{N\in\mathbb{N}}$, for $N \in \mathbb{N}$, $\beta \in [0, 1]$ and $q \in [1, \infty]$, define $\mathcal{D}^{\pm}_{N,q} := \mathcal{D}_{G_N^{\pm}(Z_{(N)})}(\beta)$.

Using an extension of the methodology from this section (§4.2), including analysis of L_q lattice balls, we can prove the following theorem. We have already considered $q = 1$ and $k \asymp \log|G|$.

Theorem 4.2.11. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ a sequence of finite, Abelian groups; for each $N \in \mathbb{N}$, define $Z_{(N)} := [Z_1, ..., Z_{k_N}]$ by drawing $Z_1, ..., Z_{k_N} \sim^{\text{iid}} \text{Unif}(G_N)$.*

Suppose that $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypotheses G′. Then, for all $\beta \in (0, 1)$, we have

$$\mathcal{D}^{\pm}_{N,q}(\beta) / \mathfrak{D}^{\pm}_{N,q} \to^{\mathbb{P}} 1 \quad \text{(in probability)} \quad \text{as } N \to \infty.$$

Moreover, the implicit lower bound holds for all choices of generators and for all Abelian groups, only requiring the conditions in Hypotheses G′ which depend only on $(k_N, |G_N|)_{N\in\mathbb{N}}$ and q.

The arguments used to prove this theorem really are analogous to those used in this section (§4.2). The only real difference is that we have to look at lattice balls under an L_q and in dimension $1 \ll k \ll \log n$, rather than L_1 and $k \asymp \log n$. Other than this, the remainder of the analysis, in particular the reduction to a gcd and the consideration of the set $\mathcal{I}$ of non-zero coordinates of W, is exactly the same. (Now W is uniform on an L_q ball of appropriate radius.) We do not give the proof here; it is deferred to §5.5.

4.3 Typical Distance: $k \gg \log|G|$

This section focusses on concentration of distances from the identity in the random Cayley graph of an Abelian group when $k \gg \log|G|$. (The previous sections dealt with $1 \ll k \lesssim \log|G|$.) The main result of the section is Theorem 4.3.2; see also Hypothesis H.

The outline of this section is as follows:
- §4.3.1 states precisely the main theorem of the section;
- §4.3.2 outlines the argument;
- §4.3.3 gives some crucial estimates on the size of lattice balls;
- §4.3.4 is devoted to the lower bound;
- §4.3.5 is devoted to the upper bound.

4.3.1 Precise Statement and Remarks

To start the section, we recall the typical distance statistic.

Definition 4.3.1. *Let H be a graph and fix a vertex $0 \in H$. For $r \in \mathbb{N}$, write $\mathcal{B}_H(r)$ for the r-ball in the graph H, ie $\mathcal{B}_H(r) := \{h \in H \mid d_H(0,h) \le r\}$, where d_H is the graph distance in H. Define*

$$\mathcal{D}_H(\beta) := \min\big\{r \ge 0 \;\big|\; |\mathcal{B}_H(r)| \ge \beta n\big\} \quad \text{for} \quad \beta \in (0,1).$$

When considering sequences $(k_N, G_N)_{N\in\mathbb{N}}$ of integers and Abelian groups, abbreviate

$$\mathcal{D}_N(\beta) := \mathcal{D}_{G_N([Z_1,\dots,Z_{k_N}])}(\beta) \quad \text{where} \quad Z_1,\dots,Z_{k_N} \sim^{\text{iid}} \text{Unif}(G_N).$$

Finally, considering such sequences, we define the candidate radius for the typical distance:

$$\overline{\mathfrak{D}}_N := \tfrac{\rho_N}{\rho_N - 1} \log|G_N| / \log k_N \quad \text{where} \quad \rho_N := \log k_N / \log\log|G_N| \quad \text{for each} \quad N \in \mathbb{N}.$$

To leading order, the typical distance will be the same for the undirected graphs as for the directed.

We show that, whp over the graph (ie choice of Z), this statistics concentrates. Here we consider $k \gg \log|G|$. The result holds for all Abelian groups; in fact, the implicit upper bound is valid for all groups. Further, the typical distance concentrates at a distances which depends only on k and $|G|$. This is in agreement with the spirit of the Aldous–Diaconis conjecture.

Hypothesis H. *The sequence $(k_N, n_N)_{N\in\mathbb{N}}$ satisfies Hypothesis H if*

$$\liminf_N k_N / \log n_N = \infty \text{ and } \liminf_N \log k_N / \log n_N = 0.$$

Theorem 4.3.2. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ a sequence of finite, Abelian groups; for each $N \in \mathbb{N}$, define $Z_{(N)} := [Z_1,\dots,Z_{k_N}]$ by drawing $Z_1,\dots,Z_{k_N} \sim^{\text{iid}} \text{Unif}(G_N)$.*

Suppose that $(k_N, |G_N|)_{N\in\mathbb{N}}$ satisfies Hypothesis H. Then, for all $\beta \in (0,1)$, we have

$$\mathcal{D}_N^{\pm}(\beta) / \overline{\mathfrak{D}}_N \to^{\mathbb{P}} 1 \quad \text{(in probability)} \quad \text{as } N \to \infty.$$

Moreover, the implicit lower bound holds deterministically, ie for all choices of generators, and the implicit upper bound holds for all groups, not just Abelian groups.

As always, for ease of presentation, in the proof we drop the N-subscripts.

4.3.2 Outline of Proof

When $k \gg \log|G|$, one can see that the typical distances statistic $\mathcal{D}$ must satisfy $\mathcal{D} \ll k$. By symmetry, the expected number of times a generator is used when drawing from a ball $B_k(R)$ is $o(1)$. The number of ways that precisely R can be chosen is $\binom{k}{r}$. Choose R with $\binom{k}{R} \approx |G|$.

4.3.3 Estimates on Sizes of Balls in $\mathbb{Z}^k$

We consider balls and spheres in the L_1 and L_∞ senses: write $B_{k,1}(\cdot)$, respectively $S_{k,1}(\cdot)$, for the L_1 ball, respectively sphere, in $\mathbb{Z}^k$; write $B_{k,\infty}(1)$ for the L_∞ unit ball in $\mathbb{Z}^k$.

Lemma 4.3.3. *For all $R \geq 0$, we have*

$$|B^{\pm}_{k,1}(R)| \leq 2^R \binom{\lfloor R \rfloor + k}{\lfloor R \rfloor} \quad \text{and} \quad \left|S^{\pm}_{k,1}(R) \cap B^{\pm}_{k,\infty}(1)\right| \geq \binom{k}{\lfloor R \rfloor}.$$

Furthermore, if $R \ll k$, then

$$2^R \binom{\lfloor R \rfloor + k}{\lfloor R \rfloor} = \exp\big(R \log(k/R) \cdot (1 + o(1))\big) = \binom{k}{\lfloor R \rfloor}$$

In particular, if $k = (\log n)^\rho$ and $\varepsilon > 0$ is a constant, then

$$\left|S^{\pm}_{k,1}\left(\tfrac{\rho}{\rho - 1} \log_k n\right) \cap B_{k,\infty}(1)\right| \gg n.$$

Proof. In the first display, the upper bound is proved in Lemma 6.5.2a; the lower bound is the usual formula for the number of subsets of $[k]$ of size R. The second display is a simple application of Stirling's approximation and asymptotics of the binary entropy function. The final display follows by combining the previous two and performing a simple calculation. □

4.3.4 Lower Bound on Typical Distance

From the results in §4.3.3, it is straightforward to deduce the lower bound in Theorem 4.3.2.

Proof of Lower Bound in Theorem 4.3.2. Let $\xi \in (0,1)$ and set $R := \overline{\mathfrak{D}}(1 - \xi)$. Since the underlying group is Abelian, applying Lemma 4.3.3, a simple calculation gives

$$|\mathcal{B}_k(R)| \leq |B_{k,1}(R)| \leq \exp\big(\mathfrak{D} \log(k/\mathfrak{D}) \cdot (1 - \tfrac{1}{2}\xi)\big) \ll n.$$

Hence, for all $\beta \in (0,1)$ and all Z, we have $\mathcal{D}_k(\beta) \geq R = \mathfrak{D}(1 - \xi)$, asymptotically in n. □

4.3.5 Upper Bound on Typical Distance

Lemma 4.3.3 gives a quantitative sense in which $|B_{k,1}(R)| \approx \left|S_{k,1}(R) \cap B_{k,\infty}(1)\right| \geq \binom{k}{\lfloor R \rfloor}$; informally, this means that we do not really lose any volume by restricting to the sphere and requiring that each generator is used at most once. We show the upper bound for arbitrary groups.

Proof of Upper Bound in Theorem 4.3.2. Let $\xi > 0$ and set $R := \overline{\mathfrak{D}}(1 + \xi)$. Draw $W, W' \sim^{\text{iid}} \mathrm{Unif}(S_{k,1}(R) \cap B_{k,\infty}(1))$. Define $S := Z_1^{W_1} \cdots Z_k^{W_k}$ and S' similarly. We show that S is well-mixed whp (this time in the L_2 sense) to deduce the upper bound. Then, by the standard L_2 calculation,

$$\mathbb{E}\big(\|\mathbb{P}_{G_k}(S \in \cdot) - \pi_G\|_2^2\big) = n\,\mathbb{P}\big(S' = S'\big) - 1.$$

If $W \neq W'$, then there exists an $i \in [k]$ so that $W_i = 1$ and $W'_i = 0$ or vice versa. By the uniformity and independence of the generators, $S'S^{-1} \sim \mathrm{Unif}(G)$ for *all* (not just Abelian) groups. Thus

$$n\,\mathbb{P}\big(S = S'\big) - 1 \leq n\,\mathbb{P}\big(W = W'\big) = n\left|S_{k,1}(R) \cap B_{k,\infty}(1)\right|^{-1} \ll 1,$$

using Lemma 4.3.3 for the final relation. This completes the proof. □

Remark. We remark that this upper bound, ie on typical distance with $k \gg \log|G|$, can be easily deduced from mixing results proved in the '90s. Specifically, it was shown by Dou and Hildebrand [34, Theorem 1] that the mixing time for the usual random walk is upper bounded by $\frac{\rho}{\rho - 1} \log_k |G|$ for any group; Roichman [69, Theorems 1 and 2] subsequently gave a simpler proof, using an argument not that dissimilar from our proof above. The lower bound does not follow from mixing results, though.

There are a few reasons for including the proof above. Foremost is that we use the same argument in §4.4.2 to obtain universal bounds for $k \geq (1 + \delta)\log_2|G|$ (with $\delta > 0$ a constant), not just $k \gg \log|G|$. Additionally, we need to do most of the work for the lower bound anyway, and it demonstrates how easily our method adapts to this new regime. △

4.4 Diameter

In this section we consider the diameter of the random Cayley graph. Our analysis is separated into two distinct sections.

§4.4.1 We show that the diameter concentrates for $k \gtrsim \log|G|$, and that the value at which it concentrates is the same as for typical distance.

§4.4.2 We show, for $k \geq (1+\delta)\log_2|G|$, with $\delta > 0$ constant, that the group giving rise to the largest diameter (amongst all groups) is $\mathbb{Z}_2^d$.

4.4.1 Concentration for $k \gtrsim \log|G|$

Recall that in Theorem 4.2.2 we showed, in the regime $k \asymp \log n$ and under some assumptions, that, up to sot, the typical distance concentrates at αk, for some constant α. The next theorem shows, in the same set-up, that the diameter does the same. The argument is a relatively straightforward adaptation of the typical distance argument. Recall Hypothesis G.

Theorem 4.4.1. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ a sequence of finite, Abelian groups; for each $N \in \mathbb{N}$, define $Z_{(N)} := [Z_1, ..., Z_{k_N}]$ by drawing $Z_1, ..., Z_{k_N} \sim^{\text{iid}} \text{Unif}(G_N)$.*

Suppose that $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies either Hypotheses G or H. For $\lambda \in (0,\infty)$, let $\alpha_\lambda^\pm \in (0,\infty)$ be the constant from Theorem 4.2.2; for each $N \in \mathbb{N}$, write $\rho_N := \log k_N / \log\log|G_N|$, so that $k_N = (\log|G_N|)^{\rho_N}$. Then the following convergences in probability hold:

$$\begin{aligned} \operatorname{diam} G_N(Z_{(N)}) / \big(\alpha_\lambda^\pm k_N\big) &\to^{\mathbb{P}} 1 \quad \textit{when} \quad \lim_N k_N/\log|G_N| = \lambda; \\ \operatorname{diam} G_N(Z_{(N)}) / \big(\tfrac{\rho_N}{\rho_N - 1}\log_{k_N}|G_N|\big) &\to^{\mathbb{P}} 1 \quad \textit{when} \quad \lim_N k_N/\log|G_N| = \infty. \end{aligned}$$

Moreover, the implicit lower bound on the diameter holds deterministically, ie for all choices of generators, and for all Abelian groups, and, when $k \gg \log|G|$, the implicit upper bound holds for all groups, not just Abelian groups.

Remark 4.4.2. *While we only state and prove the result for $k \gtrsim \log|G|$, the argument can be extended to allow $k \ll \log|G|$, provided $\log|G|/k$ diverges sufficiently slowly. This require a little more care; we do not explore the details here.*

As always, we drop the N-subscripts in the proof, eg writing $\operatorname{diam} G_k$ or $|G|$.

Proof of Theorem 4.4.1. Clearly $\operatorname{diam} G_k = \mathcal{D}_k(1) \geq \mathcal{D}_k(\beta)$ for all $\beta \in [0,1]$. Hence typical distance is trivially a lower bound on the diameter. It remains to consider the upper bound.

Assume first Hypotheses G, so $k \asymp \log|G|$. Let $\varepsilon \ll 1$, vanishing slowly. Define $\alpha := \alpha_\lambda^\pm$ as in Theorem 4.2.2. Let $A := [Z_1, ..., Z_{(1-\varepsilon)k}]$ be the first $(1-\varepsilon)k$ generators and $B := [Z_{(1-\varepsilon)k+1}, ..., Z_k]$ be the remaining εk. By transitivity, it suffices to consider distances from the identity. The idea is to take L steps using A and then one more using B, where L is the minimal radius of a ball in $\mathbb{Z}_\pm^{|A|}$ of volume at least ne^ω, for some slowly diverging ω. Write $M := \alpha k$. By Lemma 4.2.5, we have $L/M \asymp 1-\varepsilon \asymp 1$. (This does not hold for $1 \ll k \ll \log|G|$ by Lemma 4.1.5.) The key point is that when $k \asymp \log|G|$ replacing k with $(1-\varepsilon)k$ changes the typical distance by a factor $1 + o_{\varepsilon\to 0}(1)$.

By Theorem 4.2.2, whp, A is *typical* in the sense that the proportion of elements of the group which can be reached via a word of length at most L, using only the generators from A, is $1-e^{-\nu}$, for some $\nu \gg 1$, independent of ε.

Condition on A, and that it is typical; write $\overline{\mathbb{P}}$ for the probability measure induced by this conditioning. Denote by H the set of elements which can be reached in the above sense. (This is the vertex set of the ball of radius L in $G(A)$.) Fix $x \in G$. Note that if $b \sim \text{Unif}(G)$, then

$$\overline{\mathbb{P}}\big(x \in b + H\big) = 1 - e^{-\nu} \quad \text{where} \quad b + H := \{b + h \mid h \in H\}.$$

Furthermore, if $b, b' \sim \text{Unif}(G)$ are independent then the events $\{x \in b + H\}$ and $\{x \in b' + H\}$ are $\overline{\mathbb{P}}$-independent; this is because we have conditioned on A, and so H is a deterministic set.

Using the εk generators from B, informally we get εk Bernoulli trials to get to x using $b+H$ for $b \in B$, and each trial has success probability $1-o(1)$. Formally, write R for the set of elements reachable from the identity via a word of length at most $L+1$ (ie the 'range'); let b' be an arbitrary element of B, so $b' \sim \mathrm{Unif}(G)$. (Recall that the conditioning makes H non-random.) Then

$$\overline{\mathbb{P}}\big(x \notin R\big) \le \overline{\mathbb{P}}\big(x \notin B+H\big) = \overline{\mathbb{P}}\big(x \notin b+H \,\forall\, b \in B\big) = \overline{\mathbb{P}}\big(x \notin b'+H\big)^{|B|} = e^{-\nu\varepsilon k}.$$

Since $\nu \to \infty$, we may choose $\varepsilon \to 0$ so that $\nu\varepsilon \to \infty$. Then, since $k \asymp \log n$, we have

$$\overline{\mathbb{P}}\big(R \neq G\big) = \overline{\mathbb{P}}\big(\exists\, x \in G \text{ st } x \notin R\big) \le n\,\overline{\mathbb{P}}\big(x \notin R\big) \le ne^{-\nu\varepsilon k} = o(1).$$

Averaging over typical A establishes an upper bound of $\operatorname{diam} G_k \le L+1$ whp, and $L \le M(1+\varepsilon)$.

Finally consider Hypotheses H, so $k \gg \log|G|$. Exactly the same argument holds here, using the typical distance to first get to almost all the elements and then one more step. Recall from Theorem 4.3.2 that the upper bound is valid for arbitrary groups. □

4.4.2 Universal Bounds for $k \ge (1+\delta)\log_2|G|$

In this subsection we show that the group $\mathbb{Z}_2^d$ gives rise to the random Cayley graph with the largest diameter when $k \ge (1+\delta)\log_2|G|$ whp, up to sot.

Recall that $\mathfrak{R}(k,n)$ is the minimal $R \in \mathbb{N}$ with $\binom{k}{R} \ge n$.

Theorem 4.4.3. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ a sequence of finite groups; for each $N \in \mathbb{N}$, define $Z_{(N)} := [Z_1, \dots, Z_{k_N}]$ by drawing $Z_1, \dots, Z_{k_N} \sim^{\mathrm{iid}} \mathrm{Unif}(G_N)$.*

Suppose that $\liminf_N (k_N - \log_2|G_N|)/k_N > 0$ and $\limsup_N \log k_N / \log|G_N| = 0$. Then

$$\limsup_N \operatorname{diam} G_N(Z_{(N)})/\mathfrak{R}(k_N,|G_N|) \le 1 \quad \textit{in probability.}$$

As noted in the introduction, the proof of this statement is that it will follow from our previous typical distance and diameter considerations with relatively little extra work.

Proof. From Lemma 4.3.3 and Theorem 4.4.1, when $k \gg \log|G|$, the diameter concentrates at $\mathfrak{R}(k,|G|)$ when the underlying group is Abelian, and this is an upper bound for all groups.

Thus it remains to consider k with $k - \log_2|G| \asymp k$ and $k \asymp \log|G|$. All that was required for the upper bound on typical distance when $k \gg \log|G|$ was that $\mathbb{P}(W = W') \ll 1/|G|$ where $W, W' \sim^{\mathrm{iid}} \mathrm{Unif}(S_{k,1}(D) \cap B_{k,\infty}(1))$ with $D := \overline{\mathfrak{D}}(1+\xi)$, where $\overline{\mathfrak{D}}$ was the candidate typical distance radius and $\xi > 0$ was a constant. We show that the analogous statement holds here.

Let $\xi > 0$ be fixed and set $R := \mathfrak{R}(k,|G|)(1+\xi)$. Before proceeding, let us determine some estimates on $\mathfrak{R}$. Let $h : (0,1) \to (0,1) : p \mapsto -p\log p - (1-p)\log(1-p)$ denote the binary entropy function (in nats). It is standard that Stirling's approximation, like in Lemma 4.3.3, gives

$$\tbinom{k}{r} = \exp\big(k\,h(r/k)\cdot\big(1+o(1)\big)\big).$$

Thus if $k - \log_2|G| \asymp k$, then we see that $\mathfrak{R}(k,|G|) \asymp k$. Further, the fact that the derivative of h is continuous and strictly positive on $(0,\frac{1}{2})$ gives $\binom{k}{R} \gg |G|$; hence $\mathbb{P}(W = W') \ll 1/|G|$.

This shows that the typical distance $\mathcal{D}_k(\beta) \le \mathfrak{R}(k,|G|)$ whp up to sot for all constants $\beta \in (0,1)$. This is then converted from a statement about typical distance to one about the diameter via the same method as used previously (in §4.4.1), noting that $\mathfrak{R}(k,|G|) \asymp k$. □

4.5 Spectral Gap

In this section, we calculate the spectral gap; see Theorem L. We first prove it for $k \ge 3d(G)$. In §4.5.4, we explain how to extend to $k \ge (2+\delta)d(G)$ and then to $k \ge (1+\delta)d(G)$ for a density-$(1-\varepsilon)$ subset of values for $|G|$. The lower bound holds deterministically, without any conditions.

4.5.1 Precise Statement

For an Abelian group G, we write $d(G)$ for the minimal size of a generating set. It is convenient to phrase the statement in terms of the *relaxation time*, which is the inverse of the spectral gap.

Theorem 4.5.1 (Spectral Gap). *First, there exists an absolute constant $c > 0$ so that, for all Abelian groups G and all (multi)sets Z of generators of size k, we have*

$$t_{\mathrm{rel}}\big(G(Z)\big) \geq c|G|^{2/k}. \tag{4.5.1a}$$

Second, for all $\delta > 0$, there exist constants $c_\delta, C_\delta > 0$ so that, for all Abelian groups G, if $k \geq (2+\delta)d$ and $Z_1, ..., Z_k \sim^{\mathrm{iid}} \mathrm{Unif}(G)$, then

$$\mathbb{P}\big(t_{\mathrm{rel}}(G_k) \leq C_\delta |G|^{2/k}\big) \geq 1 - C_\delta 2^{-k/c_\delta}. \tag{4.5.1b}$$

Furthermore, for all $\varepsilon \in (0,1)$, there exists a subset $\mathbb{A} \subseteq \mathbb{N}$ of density at least $1-\varepsilon$ so that if $|G| \in \mathbb{A}$ then then condition $k \geq (2+\delta)d(G)$ can be relaxed to $k \geq (1+\delta)d(G)$ and (4.5.1b) still holds; the constant C_δ now also depends on ε, ie becomes $C_{\delta,\varepsilon}$, but c_δ need not be adjusted.

We prove this for the non-absolute spectral gap, ie $\min_{\lambda \neq 1}\{1-\lambda\}$, where the minimum is over eigenvalues; the same proof also works for the absolute spectral gap, ie $\min_{\lambda \neq 1}\{1-|\lambda|\}$.

4.5.2 Lower Bound on Relaxation Time

In this subsection, we establish the lower bound in Theorem 4.5.1.

Proof of Lower Bound in Theorem 4.5.1. Write $n \coloneqq |G|$. We may assume that $k \leq \log_3(\frac{1}{2}n)$, as otherwise (4.5.1a) indeed holds for some $c > 0$. Let $L \coloneqq \lfloor \frac{1}{2}((\frac{1}{2}n)^{1/k} - 1)\rfloor$. By our assumption on k, we have $L \geq 1$. Consider the set

$$A \coloneqq \big\{w \cdot Z \mid w \in \mathbb{Z}^k \text{ and } |w_i| \leq L \,\forall\, i = 1, ..., k\big\} \subseteq G. \tag{4.5.2}$$

Clearly $|A| \leq (2L+1)^k \leq \frac{1}{2}n$. Let $t \geq 0$, and let $(Y_s)_{s \geq 0}$ be a continuous-time rate-1 SRW on $\mathbb{Z}$. Writing $\tau_{A^c} \coloneqq \inf\{s \geq 0 \mid S_s \notin A\}$ for the exit time of A by the SRW S, observe that

$$\mathbb{P}_0\big(\tau_{A^c} > t\big) \leq \mathbb{P}_0\big(\max_{s \in [0,t/k]} |Y_s| \leq L\big)^k, \tag{4.5.3}$$

where $0 \in A$ is the identity of the group. It follows from Lemma 4.5.3 below that

$$\mathbb{P}_0\big(\max_{s \in [0,t/k]} |Y_s| \leq L\big) \geq \exp\big(-\tfrac{1}{8}\pi^2 (t/k)/(L+1)^2\big).$$

Substituting this into (4.5.3) we get

$$\mathbb{P}_0\big(\tau_{A^c} > t\big) \geq \exp\big(-\tfrac{1}{8}t\pi^2/(L+1)^2\big). \tag{4.5.4}$$

The minimal Dirichlet eigenvalue of a set A is defined to be the minimal eigenvalue of minus the generator of the walk killed upon exiting A; we denote it by λ_A. For connected A, we show in Lemma 4.5.4 below that, for all $a \in A$, we have

$$-\tfrac{1}{t}\log \mathbb{P}_a\big(\tau_{A^c} > t\big) \to \lambda_A \quad \text{as } t \to \infty.$$

From this and (4.5.4), it then follows that $\lambda_A \leq \lambda$ where

$$\lambda \coloneqq \tfrac{1}{8}\pi^2/(L+1)^2 \leq \pi^2/\big((\tfrac{1}{2}n)^{1/k} + 1\big)^2.$$

Since $|A| \leq \frac{1}{2}n$, applying [2, Corollary 3.34], we get

$$t_{\mathrm{rel}} \geq (1 - \tfrac{1}{n}|A|)/\lambda \geq 1/(2\lambda).$$

This concludes the proof of the lower bound in Theorem 4.5.1, namely (4.5.1a). □

4.5.3 Upper Bound on Relaxation Time

In this subsection, we establish the upper bound in Theorem 4.5.1, namely (4.5.1b). For ease of presentation, we assume first that $k \geq 3d(G)$. In §4.5.4, we explain how to relax this condition, to prove the complete theorem.

Proof of Upper Bound in Theorem 4.5.1. Decompose G as $\oplus_1^d \mathbb{Z}_{m_j}$. An orthogonal basis of eigenvectors for P, the transition matrix of the corresponding discrete-time walk, is given by

$$(f_x \mid x \in G) \quad \text{where} \quad f_x(y) := \cos\bigl(2\pi \textstyle\sum_{i=1}^d x_i y_i / m_i\bigr),$$

with corresponding eigenvalues given by

$$\begin{gathered}(\lambda_x \mid x \in G) \quad \text{where} \quad \lambda_x = \tfrac{1}{k} \textstyle\sum_{i=1}^k \cos\bigl(2\pi(\bar{x} \cdot Z_i)\bigr), \\ \text{where} \quad \bar{x}_j = x_j/m_j \text{ for all } j = 1, \ldots, d \quad \text{and} \quad \bar{x} \cdot Z_i = \textstyle\sum_{j=1}^d x_j Z_i^j / m_j\end{gathered}$$

is the standard inner-product on $\mathbb{R}^d$, where Z_i^j is the j-th coordinate of the i-th generator Z_i; here we identify $\bar{x}$ and Z_j with elements of $\mathbb{R}^d$ in a natural manner.

Observe that $\lambda_0 = 1$. Our goal is to bound $\min_{x \in G \setminus \{0\}}\{1 - \lambda_x\}$ from below. For $\alpha \in \mathbb{R}$, let $\{\alpha\}$ be the unique number in $(-\frac{1}{2}, \frac{1}{2}]$ so that $\alpha - \{\alpha\} \in \mathbb{Z}$. It follows from Lemma 4.5.5 below that

$$1 - \lambda_x \geq \tfrac{2\pi^2}{3k} \textstyle\sum_{i=1}^k \{\bar{x} \cdot Z_i\}^2. \tag{4.5.5}$$

For each $x \in G$, we make the following definitions:

$$\begin{aligned} g_j &:= g_j(x) := \gcd(x_j, m_j) && \text{for each } j \geq 1; \\ s_* &:= s_*(x) := \max\bigl\{m_j/g_j \mid j \in \{1, \ldots, d\}\bigr\}; \\ A(s) &:= \bigl\{x \in G \mid s_*(x) = s\bigr\} && \text{for each } s \geq 1; \\ \phi(j) &:= \bigl|\bigl\{j' \in \{1, \ldots, j\} \mid \gcd(j, j') = 1\bigr\}\bigr| && \text{for each } j \geq 1. \end{aligned}$$

From this, we claim that we are able to deduce, for $s \geq 2$, that

$$|A(s)| \leq \bigl(\textstyle\sum_{j=1}^s \phi(j)\bigr)^d \leq \bigl(1 + \textstyle\sum_{j=2}^s (j-1)\bigr)^d \leq \bigl(\tfrac{1}{2}s^2\bigr)^d. \tag{4.5.6}$$

Indeed, $\phi(j) \leq j - 1$ for $j \geq 2$, and observe that

$$\text{if } r \text{ divides } m, \quad \text{then} \quad \bigl|\bigl\{a \in \{1, \ldots, m\} \bigm| \gcd(a, m) = r\bigr\}\bigr| = \phi(m/r);$$

hence, summing over the set of possible values for m_j/g_j, which by definition of $A(s)$ is $\{1, \ldots, s\}$, we have $|A(s)|^{1/d} \leq \sum_{j=1}^s \phi(j)$. We are then able to deduce the upper bound, ie (4.5.1b), from Proposition 4.5.2, which we state precisely below. Indeed, first write

$$p(s) := \max_{x : s_*(x) = s} \mathbb{P}\bigl(1 - \lambda_x \leq c_1 n^{-2/k}\bigr).$$

By (4.5.5) along with Proposition 4.5.2 and Lemma 4.5.5 (stated below), for $c' := c_1 \cdot \frac{3}{2\pi^2}$, we have

$$\begin{aligned} \sum_{x \in G \setminus \{0\}} \mathbb{P}\bigl(1 - \lambda_x \leq c_1' n^{-2/k}\bigr) &\leq n \max_{s > C_2 n^{1/k}} p(s) + \sum_{2 \leq s \leq C_2 n^{1/k}} |A(s)|\, p(s) \\ &\leq 2^{-k} + 2^{-d} \textstyle\sum_{s \geq 2} s^{2d} (2s)^{-9k/10} \lesssim 2^{-k}, \end{aligned}$$

where we have used $k \geq 3d$ and the fact that $s_*(x) > 1$ for all $x \neq 0$.

Modulo the proofs of the quoted results, ie Proposition 4.5.2 and Lemmas 4.5.3 to 4.5.5, this concludes the proof of the upper bound in Theorem 4.5.1, namely (4.5.1b). □

It remains to state and prove the quoted results, ie Proposition 4.5.2 and Lemmas 4.5.3 to 4.5.5.

Proposition 4.5.2. *There exist absolute constants $c_1 \in (0,1)$ and C_2 such that*

$$\mathbb{P}\big(\tfrac{1}{k}\textstyle\sum_{i=1}^k \{\bar{x}\cdot Z_i\}^2 \le c_1 n^{-2/k}\big) \le \begin{cases} s_*(x)^{-9k/10} & \text{when} \quad s_*(x) \le C_2 n^{1/k}, \qquad (4.5.7a)\\ 2^{-k}/n & \text{when} \quad s_*(x) > C_2 n^{1/k}. \qquad (4.5.7b)\end{cases}$$

Proof. Fix $x \in G$. First consider the case that $s := s_*(x) > C_2 n^{1/k}$, ie (4.5.7b). Let $j := j(x)$ be a coordinate satisfying $s = m_j/g_j$. Denote $m := m_{j(x)}$ and $g := g_{j(x)}$. Observe that $x_j Z_i^j \sim^{\text{iid}} \text{Unif}\{g, 2g, ..., m\}$ for each i. Hence, for each i, we have

$$U_i := \bar{x}_j Z_i^j \sim \text{Unif}\{1/s, 2/s, ..., 1\}. \tag{4.5.8}$$

By averaging over $(a_i)_{i=1}^k$, where $a_i := \{\sum_{\ell\in\{1,...,d\}\setminus\{j\}} x_\ell Z_i^\ell/m_\ell\}$, recalling that $\{\alpha\}$ is the unique number in $(-\frac{1}{2}, \frac{1}{2}]$ so that $\alpha - \{\alpha\} \in \mathbb{Z}$, it suffices to show that

$$\max_{b_1,...,b_k\in[-1/2,1/2]} \mathbb{P}\big(\tfrac{1}{k}\textstyle\sum_{i=1}^k \{U_i + b_i\}^2 \le c_1 n^{-2/k}\big) \le 2^{-k}/n. \tag{4.5.9}$$

Replacing c_1 with $4c_1$ we may assume that $b_i \in \frac{1}{s}\mathbb{Z}$. Indeed, if

$$|b_i - \ell/s| \le 1/(2s), \quad \text{ie} \quad |b_i - \ell/s| = \min\big\{|b_i - \alpha| \mid \alpha \in \tfrac{1}{s}\mathbb{Z}\big\},$$

then $\{U_i + \ell/s\}^2 \le 4\{U_i + b_i\}^2$. Hence

$$\text{if} \quad \tfrac{1}{k}\textstyle\sum_{j=1}^k \{U_i + b_i\}^2 \le c_1 n^{-2/k} \quad \text{then} \quad \tfrac{1}{k}\textstyle\sum_{j=1}^k \{U_i + \ell/s\}^2 \le 4c_1 n^{-2/k}.$$

In this case, $\{U_i + b_i\}$ has the same law as $\{U_i\}$. It thus suffices to prove (4.5.9) for $b_1 = \cdots = b_k = 0$.

We now split $[0, \frac{1}{2}]$ into $M := \lceil 4n^{1/k}\rceil$ consecutive intervals of equal length $J_1, ..., J_M$, where $J_1 := [0, \frac{1}{2M}]$ and $J_\ell := (\frac{\ell-1}{2M}, \frac{\ell}{2M}]$ for $\ell > 1$. Let $Y_i := \ell - 1$ if $|\{U_i\}| \in J_\ell$. Clearly, $\frac{1}{4}Y_i/M^2 \le \frac{1}{4}Y_i^2/M^2 \le \{U_i\}^2$. It thus suffices to show that

$$\mathbb{P}\big(\tfrac{1}{k}\textstyle\sum_{i=1}^k Y_i \le \frac{1}{10}\big) \le 2^{-k}/n.$$

This last claim follows by a simple counting argument: there are M^k total assignments of the Y_i-s, but at most $L(k) := \binom{\lceil 11k/10\rceil}{k-1} \le 2^k$ assignments satisfy $\frac{1}{k}\sum_{i=1}^k Y_i \le \frac{1}{10}$, since $L(k)/M^k \le 2^{-k}n^{-1}$.

We now prove the case $s := s_*(x) \le C_2 n^{1/k}$, ie (4.5.7a). By the same reasoning as for (4.5.9), it suffices to show that

$$\max_{b_1,...,b_k\in[-1/2,1/2]} \mathbb{P}\big(\tfrac{1}{k}\textstyle\sum_{i=1}^k \{U_i + b_i\}^2 \le c_1 n^{-2/k}\big) \le s^{-9k/10}. \tag{4.5.10}$$

Regardless of b_i, there is at most one $a := a(b_i) \in \{1/s, 2/s, ..., 1\}$ such that $\{a + b_i\}^2 < (2s)^{-2}$, and hence by (4.5.8), for all i, we have

$$\mathbb{P}\big(\{U_i + b_i\}^2 < (2s)^{-2}\big) \le 1/s.$$

If there is no such value $a(b_i)$, then set $a(b_i) := -1$.

If $\{U_i + b_i\}^2 \ge (2s)^{-2}$ for at least $q := k \cdot 4c_1 s^2 n^{-2/k}$ of the i-s, ie if

$$\big|\big\{i \in \{1, ..., k\} \mid U_i \ne a(b_i)\big\}\big| \ge q,$$

then $\frac{1}{k}\sum_{i=1}^k \{U_i + b_i\}^2 \ge c_1 n^{-2/k}$, as desired. As $s \le C_2 n^{1/k}$, by taking c_1 sufficiently small in terms of C_2, we can make q/k sufficiently small so that the following holds:

$$\mathbb{P}\big(\big|\big\{i \in \{1, ..., k\} \mid U_i \ne a(b_i)\big\}\big| < q\big) \lesssim \tbinom{k}{q} s^{q-k} \lesssim s^{-9k/10}. \qquad \square$$

We now state the auxiliary lemmas referenced above, ie Lemmas 4.5.3 to 4.5.5. These are technical results; their proofs are given in §6.4.

Lemma 4.5.3. *Let $\ell \in \mathbb{N}$ and $\tau := \inf\{s \ge 0 \mid |Y_s| = \ell\}$, where $(Y_s)_{s\ge 0}$ is a continuous-time rate-1 SRW on $\mathbb{Z}$. Let $\theta := \frac{1}{2}\pi/\ell$ and $\lambda := 1 - \cos\theta$. Then, for all $s \ge 0$, we have*

$$\mathbb{P}_0(\tau > s) \ge e^{-\lambda s} \ge \exp\big(-\tfrac{1}{8}s(\pi/\ell)^2\big).$$

For a transition matrix P and a set A, let λ_A be the *minimal Dirichlet eigenvalue*, defined to be the minimal eigenvalue of minus the generator of the chain killed upon exiting A, ie of

$$I_A - P_A \quad \text{where} \quad (I_A - P_A)(x,y) := \mathbf{1}\big(x,y \in A\big)\big(\mathbf{1}(x=y) - P(x,y)\big).$$

Also, for a set A, write τ_{A^c} for the (first) exit time of this set by the chain.

Lemma 4.5.4. *Consider a rate-1, continuous-time, reversible Markov chain with transition matrix P. Let A be a connected set, and let λ_A and τ_{A^c} be as above. Then, for all $a \in A$, we have*

$$-\tfrac{1}{t}\log \mathbb{P}_a\big(\tau_{A^c} > t\big) \to \lambda_A \quad \text{as } t \to \infty.$$

Lemma 4.5.5. *For $\theta \in [-\frac{1}{2}, \frac{1}{2}]$, we have*

$$2(\pi\theta)^2 \geq 1 - \cos(2\pi\theta) \geq \tfrac{2}{3}(\pi\theta)^2.$$

4.5.4 Relaxing the Conditions on k

In this subsection, we explain how to relax the conditions on k. First we can relax from $k \geq 3d$ to $k \geq (2+\delta)d$, with $\delta > 0$, valid for every group size $n = |G|$. (The constants now depend on δ.)

We now give conditions under which this can be relaxed to $k \geq (1+\delta)d$. If $G = \mathbb{Z}_p^d$ for a prime p, then one can relax this further to $k \geq (1+\delta)d$, and even allow δ to tend to 0, provided p diverges. (In this case, the term 2^{-k} has to be replaced by another term which tends to zero at a slower rate as $k \to \infty$.) This follows from the fact that now we only need to consider (4.5.6) above with $s := p$ and we can replace (4.5.6) with $|A(p)| = p^d - 1$. So the condition $k \geq (1+\delta)d$ is sufficient when $G = \mathbb{Z}_p^d$ with p prime.

We now show that if $|G|$ is 'typical' (in a precise sense), then the same condition is sufficient. In the proof above, in (4.5.6), we used the crude bound

$$|A(s)| \leq \big(\textstyle\sum_{i\in[s]} \phi(i)\big)^d \leq \big(\frac{1}{2}s^2\big)^d.$$

Instead, recalling that we write $i \wr n$ to mean that i divides n, we can use the improved bound

$$|A(s)| \leq \big(\textstyle\sum_{i\in[s]} i\,\mathbf{1}(i \wr n)\big)^d.$$

In Lemma 6.6.5, we show that, for all $\varepsilon > 0$, there exists a constant C'_ε and a density-$(1-\varepsilon)$ set $\mathbb{B}_\varepsilon \subseteq \mathbb{N}$ such that, for all $n \in \mathbb{B}_\varepsilon$ and all $2 \leq s \leq n$, we have

$$\textstyle\sum_{i\in[s]} i\,\mathbf{1}(i \wr n) \leq C'_\varepsilon s(\log s)^2.$$

Using this to derive an improved bound on $|A(s)|$, and adjusting some of the constants in the proof in an appropriate manner, an inspection of the proof reveals that, for all $n \in \mathbb{B}_\varepsilon$ and all $\delta > 0$, there exists a positive constant $C_{\varepsilon,\delta}$ so that, for all Abelian groups of size n, if $k \geq (1+\delta)d$, then

$$\mathbb{P}\big(t_{\text{rel}}(G_k) \geq C_{\varepsilon,\delta} n^{2/k}\big) \leq e^{-k/C_{\varepsilon,\delta}}.$$

4.5.5 Remarks and Extensions

Now that we have completed the proof (modulo the deferred lemmas), we make two remarks.

Remark 4.5.6. Our proof gives an explicit form for c in (4.5.1a). If $k \ll \log n$, then we get

$$t_{\text{rel}} \geq 2\pi^{-2}|G|^{2/k} \cdot \big(1 + o(1)\big).$$

Indeed, in this case, in the definition of the set A in (4.5.2), we can take $L := \lfloor \frac{1}{2}(\varepsilon n)^{1/k} \rfloor$ for any $\varepsilon > 0$, making $|A|/|G|$ arbitrary small. △

Remark 4.5.7. It is classical that

$$\tfrac{1}{2}e^{-\gamma t} \le \max_{x\in V} \left\|P_t(x,\cdot) - \pi\right\|_{\mathrm{TV}} \le \tfrac{1}{2}\big(\min_{y\in V}\pi(y)\big)^{-1/2} e^{-\gamma t},$$

where $P_t := e^{-t(I-P)}$ is the heat-kernel of the corresponding continuous-time chain; see, for example, [49, Theorems 12.4, 12.5 and 20.6]. To complement this, we note that the same holds with P^t in the role of P_t if we replace $e^{-\gamma t}$ with $(1-\gamma_*)^t$, where γ_* is the absolute spectral gap.

The argument in the proof of (4.5.1b) can be used to show, for a positive constant C, that

$$\mathbb{P}\big(1/\gamma_* \le C|G|^{2/k}\big) \ge 1 - C2^{-k},$$

where γ_* is the absolute spectral gap of the transition matrix of the SRW, when $k \ge 3d(G)$. △

4.6 Open Questions and Conjectures

We close the paper with some questions which are left open.

1: Typical Distance for All Abelian Groups

In our typical distance theorem, there were some conditions on the group. We allowed any group with $d(G) \ll \log|G|/\log\log k$ if $1 \ll k \ll \log|G|$, but once $d(G)$ became larger than this or k became order $\log|G|$, we had to impose conditions. We conjecture that these are artefacts of the proof.

> **Conjecture 1.** *Let G be an Abelian group. Suppose that $k - d(G) \gg 1$ and $1 \ll k \lesssim \log|G|$. Then the typical distance statistic concentrates. Further, if $k \ll \log|G|$ and $k - d(G) \asymp k$, then it concentrates at a value which depends only on k and $|G|$.*

The claim when $1 \ll k \ll \log|G|$ and $k - d(G) \asymp k$ is a natural extension of Theorem 4.1.2. Further, if $k \ll \sqrt{\log|G|/\log\log\log|G|}$, then $k - d(G) \gg 1$ is sufficient, by Hypothesis F. Once we relax to $k - d(G) \gg 1$, for larger k, we still expect concentration of typical distance for all Abelian groups, but now the value will likely depend on the specific group. Compare this with the occurrence of cutoff for the random walk on the random Cayley graph established in Chapter 2.

2: Diameter for Abelian Groups for Diverging k

We have shown concentration of typical distance, but never considered the diameter. It is trivial that the typical distance is a lower bound on the diameter, and that twice the typical distance is an upper bound. Can more be determined? Recall that $d(G)$ is the minimal size of a generating set.

> **Conjecture 2.** *For an Abelian group G and $Z_1, \dots, Z_k \sim^{\mathrm{iid}} \mathrm{Unif}(G)$, write Δ_Z for the diameter of the Cayley graph with generators Z. Assume that k diverges, sufficiently rapidly in terms of $d(G)$. Then the law of Δ_Z concentrates. Further, if $k - d(G) \asymp k$ and $k \ll \log|G|$, then it concentrates at a value $\Delta_{k,|G|}$ which depends only on k and $|G|$.*

3: Isoperimetry for Random Cayley Graphs

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

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

More generally, the isoperimetric constant is defined for Markov chains; see [49, §7.2]. For a given stochastic matrix P, it is easy to see that the original chain P, the time-reversal P^* and the additive symmetrisation $\frac{1}{2}(P + P^*)$ all have the same isoperimetric profile. Thus the isoperimetric constant for a directed Cayley graphs is the same as that for the undirected version.

The following conjecture assets that the Cheeger constant is, up to a constant factor, the same as that of the standard Cayley graph of $\mathbb{Z}_L^k$ where L is such that $n \asymp L^k$.

Conjecture 3. *There exists a constant c so that, for all $\varepsilon \in (0,1)$, there exist constants n_ε and M_ε so that, for every finite group G of size at least n_ε, when $k \geq M_\varepsilon$, we have*

$$\mathbb{P}\big(\Phi_*(G_k) \leq c|G|^{-1/k}\big) \leq \varepsilon,$$

where $\Phi_*(G_k)$ is the Cheeger constant of a random Cayley graph with k generators.

By [54, Theorem 6.29], which regards expansion of general Cayley graphs, along with out upper bound on typical distance (and hence on diameter), we can prove this conjecture up to a factor k.

By the well-known discrete analogue of Cheeger's inequality, discovered independently by multiple authors—see, for example, [49, Theorem 13.10]—we have $\frac{1}{2}\gamma \leq \Phi_* \leq \sqrt{2\gamma}$. Determining the correct order of Φ_* in our model remains an open problem. We conjecture that the correct order of Φ_* is given by $\sqrt{\gamma}$, ie order $|G|^{-1/k}$, using Theorem L for the order of the spectral gap.

The celebrated Alon–Roichman theorem states that the Cayley graph of any finite group G is a $(1-\varepsilon)$-expander (ie $\Phi_* \geq 1-\varepsilon$) whp when $k \geq C_\varepsilon \log|G|$, for some constant C_ε; the best known upper bound on C_ε is $\mathcal{O}(1/\varepsilon^2)$. Naor [58, Theorem 1.2] refines this for Abelian groups: he showed that one can in fact bound $|\Phi(S)-1| \leq \varepsilon\sqrt{\log|S|/\log|G|}$ for all S with $1 \leq |S| \leq \frac{1}{2}|V|$ simultaneously, when $k/\log n \geq C/\varepsilon^2$, for a constant C.

5 Additional Cutoff and Typical Distance Results for Abelian Groups

Abstract for Chapter 5

The results of this chapter supplement those of the previous chapters. In general, the results proved here are more refined, but require additional conditions on the underlying group. The three main results are the following.

For $k \asymp \log|G|$, under suitable conditions on the group, we determine the limit profile (not just the existence of cutoff) for the RW on the random Cayley graph. (This was found in Chapter 2, but only for $1 \ll k \ll \log|G|$.)

We study in greater detail the special case of $G \coloneqq \mathbb{Z}_p^d$ for a prime p. In particular, if $p \gg 1$ then we allow $k - d \asymp 1$, even $k = d + 1$ or $k = d$. We establish cutoff and determine bounds on the window. There are two regimes of behaviours according to the parameter $\zeta \coloneqq \frac{1}{k}(k-d)\log p$, namely $\zeta \ll 1$ and $\zeta \gtrsim 1$.

For Abelian groups, we extend the concept of graph distance from an L_1 to an L_q sense, analogously to L_q distances in (multi-dimensional) lattices. Under suitable conditions on the group, we establish typical distance results akin to those of Chapter 4.

Table of Contents for Chapter 5

5.1 Cutoff: Limit Profile for Random Walks on Abelian Groups

In §2.1, we established the shape of the cutoff profile for *arbitrary* Abelian groups G, with some conditions on the number k of generators in terms of G. For an Abelian group G, we wrote $d(G)$ for the minimal size of a generating set. For our results, the following conditions were sufficient:

$$\begin{aligned}
&\text{to consider any } k - d(G) \gg 1, && \text{we needed } k \ll \sqrt{\log|G|/\log\log\log|G|};\\
&\text{to consider any } k - d(G) \asymp k, && \text{we needed } k \ll \log|G|/\log\log\log|G|;\\
&\text{to consider any } k \ll \log|G|, && \text{we needed } d \ll \log|G|/\log\log\log|G|;
\end{aligned}$$

see Hypothesis A. In particular, we could never consider general $k \gtrsim \log|G|$. Recall that cutoff had already been established for arbitrary Abelian groups when $k \gg \log|G|$, but the window, never mind the profile, was not known. In this section, we outline how to alleviate the conditions on k, at the cost of some conditions on the group.

5.1.1 Entropic Times: Methodology, Definition and Concentration

We use an 'entropic method', as mentioned in §1.3.4.1; cf [11, 12, 13, 20]. 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 $i = 1, ..., k$, 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, increment $W_i \to W_i - 1$ (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

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

Recall that the invariant distribution is uniform on G, giving mass $1/n$ to each vertex. The proposed mixing time is then the time at which the auxiliary process W obtains entropy $\log n$. This time can be calculated fairly precisely in many situations; see Proposition 5.1.2.

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

$$Q_i(t) := -\log \nu_{t/k}\big(W_i(t)\big), \quad \text{and set} \quad Q(t) := -\log \mu_t\big(W(t)\big) = \textstyle\sum_1^k Q_i(t).$$

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

Definition 5.1.1 (Entropic and Times). *For all $k, n \in \mathbb{N}$ and all $\alpha \in \mathbb{R}$, define $t_\alpha := t_\alpha(k,n)$ so that*

$$\mathbb{E}\big(Q_1(t_\alpha)\big) = \big(\log n + \alpha\sqrt{vk}\big)/k \quad \textit{and} \quad s_\alpha := t_\alpha/k, \quad \textit{where} \quad v := \mathbb{V}\mathrm{ar}\big(Q_1(t_0)\big),$$

assuming that $\log n + \alpha\sqrt{vk} \geq 0$. We call t_0 the entropic time and the $\{t_\alpha\}_{\alpha\in\mathbb{R}}$ cutoff times.

Direct calculation with the Poisson distribution and SRW on $\mathbb{Z}$ gives the following relations. A sketch is given below; the rigorous details are given in §6.1.

Proposition 5.1.2 (Entropic and Cutoff Times, Proposition 6.1.2). *Assume that $1 \ll \log k \ll \log n$. Write $\kappa := k/\log n$. For all $\alpha \in \mathbb{R}$, we have $t_\alpha \eqsim t_0$ and furthermore, for some functions f and g and all $\lambda > 0$, the following relations hold:*

$$\textit{if} \quad k \ll \log n, \quad \textit{then} \quad t_\alpha \eqsim k \cdot n^{2/k}/(2\pi e) \quad \textit{and} \quad (t_\alpha - t_0)/t_0 \eqsim \alpha\sqrt{2}/\sqrt{k}; \tag{5.1.1a}$$

$$\textit{if} \quad k \eqsim \lambda \log n, \quad \textit{then} \quad t_\alpha \eqsim k \cdot f(\lambda) \quad \textit{and} \quad (t_\alpha - t_0)/t_0 \eqsim \alpha g(\lambda)/\sqrt{k}; \tag{5.1.1b}$$

$$\textit{if} \quad k \gg \log n, \quad \textit{then} \quad t_\alpha \eqsim k \cdot 1/(\kappa\log\kappa) \quad \textit{and} \quad (t_\alpha - t_0)/t_0 \eqsim \alpha\sqrt{\kappa\log\kappa}/\sqrt{k}. \tag{5.1.1c}$$

Moreover, $f, g : (0,\infty)$ are continuous bijections, whose value differs between SRW and DRW.

Sketch of Proof. In §2.1.2, we sketched the argument for $k \ll \log n$. For $k \asymp \log n$, the target entropy is order 1, and so all the random variables are bounded in probability, away from both 0 and ∞. For $k \gg \log n$, we have $t_0 \ll k$, so approximate the RW by a Bernoulli distribution (with a uniformly chosen sign for the SRW); in §2.1.2, for $k \ll \log n$, we had $t_0 \gg k$ and so approximated by a normal distribution. With this adaptation, the sketch from §2.1.2 passes over. □

Since the W_i, and hence the Q_i, are iid, Q is a sum of k iid random variables. Also, it turns out that $\mathbb{V}\mathrm{ar}(Q(t)) \eqsim \mathbb{V}\mathrm{ar}(Q(t_0)) \gg 1$ when $t \eqsim t_0$; see Corollary 6.1.7. It then stands to reason that a CLT holds for $Q = \sum_1^k Q_i$; this is indeed the case. The following propositions, which will be of great importance, is proved in §6.1.

Proposition 5.1.3 (CLT, Proposition 6.1.3)**.** *Assume that $1 \ll \log k \ll \log n$. For all $\alpha \in \mathbb{R}$, we have*

$$\mathbb{P}\big(Q(t_\alpha) \le \log n \pm \omega\big) \to \Psi(\alpha) \quad \text{for} \quad \omega := \mathbb{V}\mathrm{ar}\big(Q(t_0)\big)^{1/4} = (vk)^{1/4}.$$

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

5.1.2 Precise Statement and Remarks

In this section we give the more refined version of Theorem M. Recall that, for an Abelian group G, we write $d(G)$ for the minimal size of a generating subset of G and

$$m_*(G) := \max\big\{\min_{j\in[d]} m_j \;\big|\; \oplus_1^d \mathbb{Z}_{m_j} \text{ is a decomposition of } G\big\}.$$

Hypothesis I. *An Abelian group G and integer k jointly satisfy Hypothesis I if*

$$k \ge \tfrac{1}{2}\log|G|/\log\log|G|, \quad m_*(G) > |G|^{1/k}(\log k)^2 \quad \text{and} \quad d(G) \le \tfrac{1}{30}\log|G|/\log k.$$

Recall that we write $d^{\pm}_{G_k,N}(t)$ for the TV distance from uniform at time t for the walk on $G_k^{\pm}$ and Ψ for the standard Gaussian tail. Throughout the proofs, we drop the subscript-N from the notation, considering sequences implicitly. We now state the main theorem of this section.

Theorem 5.1.4. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ a sequence of finite, Abelian groups; for each $N \in \mathbb{N}$, define $Z_{(N)} := [Z_1, ..., Z_{k_N}]$ by drawing $Z_1, ..., Z_{k_N} \sim^{\mathrm{iid}} \mathrm{Unif}(G_N)$.*

Suppose that the sequence $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis I. Then, for all $\alpha \in \mathbb{R}$, we have

$$d^{\pm}_{G_k,N}\big(t_\alpha(k_N, |G_N|)\big) \to^{\mathbb{P}} \Psi(\alpha) \quad \textit{(in probability)} \quad \text{as } N \to \infty.$$

That is, for all $\alpha \in \mathbb{R}$, whp, t_α is, up to sot, the mixing time $t_{\mathrm{mix}}(\Psi^{-1}(\alpha))$. Moreover, the implicit lower bound holds deterministically, ie for all choices of generators.

Remark 5.1.5. We can write the cutoff statement, emphasising the N-dependence, in the form

$$\big(t^{Z,N}_{\mathrm{mix}}(\varepsilon) - t_{0,N}\big)/w_N \to^{\mathbb{P}} \Psi^{-1}(\varepsilon) \quad \text{for all} \quad \varepsilon \in (0,1),$$

where $(t_{0,N})_{N\in\mathbb{N}}$ is the mixing time and $(w_N)_{N\in\mathbb{N}}$ is the window, defined by Proposition 5.1.2: for all $\lambda \in (0,\infty)$ and all $\varepsilon \in (0,1)$, combining Proposition 5.1.2 and Theorem 5.1.4, we have

$$\frac{t^Z_{\mathrm{mix}}(\varepsilon) - f(\lambda)k}{g(\lambda)\sqrt{k}} \to \Psi^{-1}(\varepsilon) \quad \text{for all} \quad \varepsilon \in (0,1) \quad \text{when} \quad k \eqsim \lambda \log|G|. \qquad \triangle$$

Remark. The CLT, Proposition 5.1.3, gives the dominating term in the TV distance Theorem 5.1.4:
- on the event $\{Q(t_\alpha) \le \log n - \omega\}$, we lower bound the TV distance by $1 - o(1)$;
- on the event $\{Q(t_\alpha) \ge \log n + \omega\}$, we upper bound the expected TV distance by $o(1)$.

Combined with the CLT, we deduce that the $d_Z(t_\alpha) \to \Psi(\alpha)$ in probability. $\triangle$

Remark. Observe that Hypothesis I requires $k \ge \frac{1}{2}\log n/\log\log n$ and $d \le \frac{1}{30}\log n/\log k$; in particular, this implies that $k/d \ge 10$. This method does in fact apply for all k with $1 \ll \log k \ll \log n$. We do not give details, though. This is because of results proved in §2.1. Write $d := d(G)$.

- If $1 \ll k \ll \log n / \log\log\log n$ and $k - d \asymp k$, then the argument of §2.1 establishes the same cutoff profile, but with no conditions on the Abelian group.
- If $1 \ll k \ll \log n$, then the conditions $k - d \asymp k$ and $d \ll \log n / \log\log\log n$ are sufficient.

For $\frac{1}{2}\log n / \log\log n \leq k \ll \log n$, these are weaker than Hypothesis I. However, $k \asymp \log n$ is never allowed in §2.1. (It is considered in §2.2, and there the group is arbitrary, but the profile is not found, only the occurrence of cutoff is shown.)

As such, the method of this article is only really of interest for $k \gtrsim \log n$, where cutoff is known but the profile is not: for $k \gg \log_2 n \geq d(G)$, this had been established prior to our work (see §1.5.2); for $k \gg \sqrt{\log n}$ and $k - d(G) \gg 1$, this is established by our work (see §2.2. △

Remark 5.1.6. The regime $k \asymp \log n$ is of particular interest. It can be thought of as a 'critical regime': if $k \ll \log n$, then $t_{\text{mix}} \gg k$; if $k \asymp \log n$, then $t_{\text{mix}} \asymp k$; if $k \gg \log n$, then $t_{\text{mix}} \ll k$.

Further, in this regime, the analysis of §2.2 disproved the Abelian Aldous–Diaconis conjecture: eg, the mixing times for $\mathbb{Z}_2^{2r}$ and $\mathbb{Z}_4^r$ are different if $k \asymp \log n$, but not if $k \gg 2\log_2 n$; note that $d(\mathbb{Z}_2^{2r}) = 2r = \log_2 |\mathbb{Z}_2^{2r}|$ and $d(\mathbb{Z}_4^r) = r = \frac{1}{2}\log_2 |\mathbb{Z}_4^r|$, so we need $k \geq d \asymp \log n$.

The Abelian Aldous–Diaconis conjecture was already known to hold when $k \gg \log |G|$. In §2.1, we were able to give sufficient conditions for it to hold when $k \ll \log n$; see, eg, the bullets in the previous remark or Remark 2.1.5. In this article, we close the gap, giving sufficient conditions for the conjecture to hold when $k \asymp \log n$. △

5.1.3 Outline of Proof

The outline here is very similar to that from the main article. For a detailed outline, see §2.1.4 there; here we outline the difference. Note that the lower bound in §2.1.5 was valid for all groups; we repeat it here for convenience.

For the upper bound, we were trying to bound the expectation of a d-th power of a gcd. Issues arose when k became too large while $k - d$ is fairly small; see the proof of Corollary 2.1.15. This arose from the fact that we used the following estimate from Lemma 2.1.14:

$$\mathbb{P}\big(V_1 \in \gamma\mathbb{Z}\big) \leq \mathbb{P}\big(V_1 \in \gamma\mathbb{Z} \mid V_1 \neq 0\big) + \mathbb{P}\big(V_1 = 0\big) \leq 1/\gamma + 2/n^{1/k}.$$

Once this was raised to the power k, the second term became an issue. We alleviate this by defining

$$\mathcal{I} \coloneqq \big\{i \in [k] \mid V_i \neq 0\big\}$$

and studying $\mathbb{P}(V_i \in \gamma\mathbb{Z} \mid i \in \mathcal{I})$; the problematic term $2/n^{1/k}$ then does not exist. If $G = \oplus_1^d \mathbb{Z}_{m_j}$, then we are actually interested in $V_i \bmod m_j$ for each j. Recall that $m_* = \min_j m_j$. 'Typically', one has $|V_i| \leq m_*$. We assume m_* is sufficiently large so that $\max_i |V_i| < m_*$ whp. Thus looking at $V_i = 0$ or $V_i \equiv 0 \bmod m_j$ is no different.

For large $|I|$, the gcd analysis goes through similarly to before. When $|I|$ is small, eg smaller than d, it is more difficult to control; in this case, we use a fairly naive bound on the gcd, but control carefully the probability of realising such an $\mathcal{I}$. The case $\mathcal{I} = \emptyset$ corresponding to $V = 0$, is handled using the concentration around the entropic time in exactly the same way as before.

5.1.4 Lower Bound on Mixing

In this subsection, we prove the lower bound on mixing, which holds for every choice of Z.

In §2.1, we only considered $1 \ll k \ll \log n$. As such, we only stated the entropic results for this regime. Above, in Propositions 5.1.2 and 5.1.3, we stated analogous results for the full regime $1 \ll \log k \ll \log n$. In the lower bound given in §2.1.5, valid for arbitrary groups, there were no conditions on k beyond those required for the entropic concentration, namely Proposition 2.1.3. As such, the identical proof passes over to the full regime $1 \ll \log k \ll \log n$ unchanged.

5.1.5 Upper Bound on Mixing

We use a modified L_2 calculation, as in Lemma 5.2.6 and Definition 5.2.7 in §5.2.5 above. There we only bounded the order of the cutoff window; now we desire the profile. We use definitions

analogous to Lemma 2.1.6 and Definition 2.1.7 in §2.1.6, where the profile is studied. Herein, we often suppress the time and α-subscripts, eg writing W for $W(t_\alpha)$ or $W(t)$, depending on context.

Let W' be an independent copy of W; then $S' := W' \cdot Z$ is an independent copy of S. We recall the modified L_2 calculation; the following lemma is the same as Lemma 5.2.6.

Lemma 5.1.7. *For all $t \geq 0$ and all $\mathcal{W} \subseteq \mathbb{Z}^k$, the following inequalities hold:*

$$\big\|\mathbb{P}_Z\big(S(t) \in \cdot\big) - \pi_G\big\|_{\mathrm{TV}} \leq \big\|\mathbb{P}_Z\big(S(t) \in \cdot \mid W(t) \in \mathcal{W}\big) - \pi_G\big\|_{\mathrm{TV}} + \mathbb{P}\big(W(t) \notin \mathcal{W}\big); \tag{5.1.2a}$$

$$4\,\mathbb{E}\big(\big\|\mathbb{P}_Z\big(S(t) \in \cdot \mid W(t) \in \mathcal{W}\big) - \pi_G\big\|_{\mathrm{TV}}^2\big) \leq \mathbb{P}\big(S(t) = S'(t) \mid W(t), W'(t) \in \mathcal{W}\big) - 1. \tag{5.1.2b}$$

We now make the specific choice of the 'typical' set $\mathcal{W}$; we make a different choice for each $\alpha \in \mathbb{R}$. Write Ψ for the standard Gaussian tail. The collection $\{\mathcal{W}_\alpha\}_{\alpha\in\mathbb{R}}$ of sets will satisfy

$$\mathbb{P}\big(W(t_\alpha) \notin \mathcal{W}_\alpha\big) \approx \Psi(\alpha),$$

using the CLT (Proposition 5.1.3). We show that the expression in (5.1.2b) is $o(1)$. Then applying (5.1.2a) gives $d_Z(t_\alpha) \leq \Psi(\alpha) + o(1)$ whp over Z. This matches the lower bound in §5.1.4.

By considering all $\alpha \in \mathbb{R}$, we are able to 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, $\mathbb{P}(W(t_\alpha) \notin \mathcal{W}_\alpha) \approx \Psi(\alpha) \approx 0$, which explains the use of the word 'typically' in describing $\mathcal{W}_\alpha$.

In order to define precisely the set $\mathcal{W}_\alpha$ here, we first define two parameters, r_α and p_α.

Definition 5.1.8a. *For all $\alpha \in \mathbb{R}$, define $r_\alpha(k,n)$ and $p_\alpha(k,n)$ as follows:*

$$r_\alpha(k,n) := \min\big\{r \in \mathbb{Z}_+ \;\big|\; \mathbb{P}\big(\big|W_1(t_\alpha) - \mathbb{E}\big(W_1(t_\alpha)\big)\big| > r\big) \leq 1/k^{3/2}\big\};$$
$$p_\alpha(k,n) := \min\big\{\mathbb{P}\big(W_1(t_\alpha) - \mathbb{E}\big(W_1(t_\alpha)\big) = j\big) \;\big|\; |j| \leq r_\alpha(k,n)\big\}.$$

Also define $r_*(k,n) := \frac{1}{2}n^{1/k}(\log k)^2$ and $p_*(k,n) := n^{-1/k}k^{-2}$.

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

Definition 5.1.8b. *For all $\alpha \in \mathbb{R}$, define the local and global typicality conditions, respectively:*

$$\mathcal{W}_{\alpha,\ell} := \big\{w \in \mathbb{Z}^k \;\big|\; \big|w_i - \mathbb{E}\big(W_1(t_\alpha)\big)\big| \leq r_\alpha \;\forall\, i = 1, \ldots, k\big\};$$
$$\mathcal{W}_{\alpha,g} := \big\{w \in \mathbb{Z}^k \;\big|\; \mathbb{P}\big(W(t_\alpha) = w\big) \leq n^{-1}e^{-\omega}\big\}.$$

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

The following proposition determines the probability that $W(t_\alpha)$ lies in $\mathcal{W}_\alpha$, ie of typicality.

Proposition 5.1.9. *For each $\alpha \in \mathbb{R}$, we have*

$$\mathbb{P}\big(W(t_\alpha) \notin \mathcal{W}_\alpha\big) \to \Psi(\alpha).$$

Proof. By our CLT, Proposition 5.1.3, the probability that the global conditions hold converges to $1 - \Psi(\alpha)$. By Definition 5.1.8a, the probability that a single coordinate fails the local condition is at most $k^{-3/2}$. By the union bound, the probability that local typicality fails to hold is then at most $k^{-1/2} = o(1)$. The claim follows. □

Herein, we fix $\alpha \in \mathbb{R}$ and frequently suppress the t_α from the notation, eg writing $W_\cdot$ for $W_\cdot(t_\alpha)$ or $\mathcal{W}$ for $\mathcal{W}_\alpha$. Let $V := W - W'$, so $\{W \cdot Z = W' \cdot Z\} = \{V \cdot Z = 0\}$. Write

$$D := D_\alpha := n\,\mathbb{P}\big(V(t_\alpha) \cdot Z = 0 \mid \mathsf{typ}_\alpha\big) - 1 \quad \text{where} \quad \mathsf{typ} := \mathsf{typ}_\alpha := \big\{W(t_\alpha), W'(t_\alpha) \in \mathcal{W}_\alpha)\big\}.$$

It remains to show that $D_\alpha = o(1)$ for all $\alpha \in \mathbb{R}$. Recall the conditions of Hypothesis I:

$$\min_j m_j > |G|^{1/k}(\log k)^2, \quad k \geq \tfrac{1}{2}\log|G|/\log\log\log|G| \quad \text{and} \quad d \leq \tfrac{1}{30}\log|G|/\log k.$$

Proposition 5.1.10. *Suppose that (k, G) jointly satisfy Hypothesis I. (Recall that, implicitly, (k, G) is a sequence of Abelian groups and integers.) Then, for all $\alpha \in \mathbb{R}$, we have $D_\alpha = o(1)$.*

Given this proposition, we can prove the upper bound in the main theorem, Theorem 5.1.4.

Proof of Upper Bound in Theorem 5.1.4 Given Proposition 5.1.10. Hypothesis I imply that conditions required for Proposition 5.1.10. Apply the modified L_2 calculation, Lemma 5.1.7 and Definition 5.1.8b, and use Propositions 5.1.9 and 5.1.10 to control the two resulting terms. Combined, these says that $d_Z(t_\alpha) \le \Psi(\alpha) + o(1)$ whp over Z. □

It remains to prove Proposition 5.1.10, ie to bound the modified L_2 distance. The remainder of the section is dedicated to this goal. Up to here, the proof has been very similar to that given in §5.2.5 or in §2.2.7; here it diverges somewhat.

Write $[k] := \{1, ..., k\}$. For $v \in \mathbb{Z}^k$, write

$$\mathcal{I}(v) := \big\{ i \in [k] \;\big|\; v_i \not\equiv 0 \bmod m_j \text{ for all } j = 1, ..., d \big\}.$$

We always consider V conditioned on typicality. By local typicality, $|V_i| \le 2r_*$. By Hypothesis I, $r_* = n^{1/k}(\log k)^2 < m_j$ for all i and j. Thus, conditioned on local typicality,

$$\mathcal{I}(V) = \big\{ i \in [k] \;\big|\; V_i \neq 0 \big\}; \quad \text{abbreviate} \quad \mathcal{I} := \mathcal{I}(V).$$

Thus we may write $D := D_\alpha$ as

$$D + 1 = n \textstyle\sum_{I \subseteq [k]} \mathbb{P}\big(V \cdot Z \equiv 0,\, \mathcal{I} = I \mid \mathsf{typ}\big).$$

We now split the sum into 'large $\mathcal{I}$', 'small $\mathcal{I}$' and 'empty $\mathcal{I}$'. In the sums below, we always have $I \subseteq [k]$. Let L be a number greater than 1, allowed to depend on n. We then have

$$\begin{aligned} D + 1 \le\; & n \textstyle\sum_{1 \le |I| < L} \mathbb{P}\big(V \cdot Z \equiv 0 \mid \mathcal{I} = I,\, \mathsf{typ}\big)\mathbb{P}\big(\mathcal{I} = I \mid \mathsf{typ}\big) \\ & + n \textstyle\sum_{|I| \ge L} \mathbb{P}\big(V \cdot Z \equiv 0 \mid \mathcal{I} = I,\, \mathsf{typ}\big)\mathbb{P}\big(\mathcal{I} = I \mid \mathsf{typ}\big) + n\,\mathbb{P}\big(\mathcal{I} = \emptyset \mid \mathsf{typ}\big), \end{aligned} \tag{5.1.3}$$

noting that if $\mathcal{I} = \emptyset$ then $V = 0 \in \mathbb{Z}^k$ (as a vector), and hence $V \cdot Z = 0$.

We first bound the third term on the right-hand side of (5.1.3), ie consider $\mathcal{I} = \emptyset$. The global typicality condition is designed precisely so that the following lemma holds.

Lemma 5.1.11. *We have*

$$n\,\mathbb{P}\big(\mathcal{I} = \emptyset \mid \mathsf{typ}\big) \le e^{-\omega}/\mathbb{P}\big(\mathsf{typ}\big). \tag{5.1.4}$$

Proof. We have $\{\mathcal{I} = \emptyset\} = \{W = W'\}$ given typicality. So the claim follows as in Lemma 5.2.9. □

We now turn our attention to $\mathcal{I} \neq \emptyset$, where we must also analyse $\mathbb{P}(V \cdot Z \equiv 0 \mid \mathcal{I} = I,\, \mathsf{typ})$. For $r_1, ..., r_\ell \in \mathbb{Z} \setminus \{0\}$, we use the convention $\gcd(r_1, ..., r_\ell, 0) := \gcd(|r_1|, ..., |r_\ell|)$. Define

$$\mathfrak{g}_j := \gcd\big(V_1, ..., V_k, m_j\big) \text{ for } j = 1, ..., d, \quad \text{and also define} \quad \mathfrak{g} := \gcd\big(V_1, ..., V_k, n\big).$$

We end up needing to separate the concepts of *local* and *global* typicality: define

$$\mathsf{typ}_\ell := \big\{W, W' \in \mathcal{W}_\ell\big\} \quad \text{and} \quad \mathsf{typ}_g := \big\{W, W' \in \mathcal{W}_g\big\}, \quad \text{so} \quad \mathsf{typ} = \mathsf{typ}_\ell \cap \mathsf{typ}_g.$$

We now state a simple lemma, describing the law of $V \cdot Z$ given $\mathcal{I} \neq \emptyset$. The same lemma is used in §2.1.6; its proof is given in Lemma 6.6.1.

Lemma 5.1.12 (Lemma 6.6.1). *For any $v \in \mathbb{Z}^k$, writing $\mathfrak{g}_j(v) := \gcd(v_1, ..., v_k, m_j)$, we have*

$$v \cdot Z \sim \mathrm{Unif}\big(\textstyle\prod_1^d \mathfrak{g}_j(v)\mathbb{Z}_{m_j}\big); \quad \textit{note that} \quad \mathfrak{g}_j(v) = \big\{\mathfrak{g}_j(v), 2\mathfrak{g}_j(v), ..., m_j\big\}.$$

To control this gcd, we determine the probability an individual coordinate is a multiple of a given number in the following auxiliary lemma; it is taken from §2.1.6. Write $\alpha \wr \beta$ if α divides β.

Lemma 5.1.13. *For all non-empty $I \subseteq [k]$ with $\{\mathcal{I} = I\} \cap \mathsf{typ} \neq \emptyset$ and all $\gamma \in \mathbb{N}$, we have*

$$\mathbb{P}\big(\gamma \wr V_i \,\forall\, i \in I \mid \mathcal{I} = I,\, \mathsf{typ}_\ell\big) \leq \gamma^{-|I|}.$$

Proof. The coordinates are independent and *local* typicality merely conditions each coordinate to lie in a certain interval centred at 0. The claim now follows immediately from Lemma 2.1.14. □

Note that $\mathfrak{g}_j \leq \mathfrak{g}$ since m_j divides n, for all $j = 1, ..., d$. From the lemma we now deduce that

$$n\,\mathbb{P}\big(V \cdot Z \equiv 0 \mid \mathcal{I} = I,\, \mathsf{typ}\big) = n\,\mathbb{E}\big(\textstyle\prod_1^d \mathfrak{g}_j/m_j \mid \mathcal{I} = I,\, \mathsf{typ}\big) \leq \mathbb{E}\big(\mathfrak{g}^d \mid \mathcal{I} = I,\, \mathsf{typ}\big), \tag{5.1.5}$$

as $\prod_1^d m_j = n$ and since, by local typicality, we have $|V_i| < m_j$ for all i and j and observing that the conditioning affects V, but not Z. We now bound the expectation of this gcd.

In order to do this, in one situation we consider a 'worst-case' for W. For this, we need to know bounds on r_α. Also given are bounds on p_α, which will be used in Lemma 5.1.17 below.

Proposition 5.1.14 (Proposition 6.3.3)**.** *For all $\alpha \in \mathbb{R}$, we have*

$$r_\alpha(k,n) \geq r_*(k,n) \quad \text{and} \quad p_\alpha(k,n) \geq p_*(k,n). \tag{5.1.6}$$

Proof. This follows from standard large deviation theory. Its proof can be found in §6.3. □

Remark. The exponent 2 in $(\log k)^2$ is not optimal, but is chosen for convenience of proof and to enable us to deal with all regimes of k simultaneously. △

Lemma 5.1.15. *There exists a constant C so that, for all $I \neq \emptyset$ with $\{\mathcal{I} = I\} \cap \mathsf{typ} \neq \emptyset$, we have*

$$\mathbb{E}\big(\mathfrak{g}^d \mid \mathcal{I} = I,\, \mathsf{typ}\big) \leq \begin{cases} C(2r_*)^{d-|I|+2}/\mathbb{P}\big(\mathsf{typ}_g \mid \mathcal{I} = I,\, \mathsf{typ}_\ell\big) & \text{when} \quad |I| \leq d+1, \qquad \text{(5.1.7a)} \\ 1 + 3 \cdot 2^{d-|I|}/\mathbb{P}\big(\mathsf{typ}_g \mid \mathcal{I} = I,\, \mathsf{typ}_\ell\big) & \text{when} \quad |I| \geq d+2. \qquad \text{(5.1.7b)} \end{cases}$$

Furthermore, recalling the definition of r_* from Definition 5.1.8a, we also have

$$\mathbb{E}\big(\mathfrak{g}^d \mid \mathcal{I} = I,\, \mathsf{typ}\big) \leq (2r_*)^d = n^{d/k}(\log k)^{2d}. \tag{5.1.8}$$

An easy corollary of this says that the contribution to (5.1.3) by 'large $\mathcal{I}$' is $1 + o(1)$.

Corollary 5.1.16. *For any L with $L \geq d+2$, we have*

$$n \textstyle\sum_{|I| \geq L} \mathbb{P}\big(V \cdot Z \equiv 0,\, \mathcal{I} = I \mid \mathsf{typ}\big) \leq 1 + 3 \cdot 2^{d-L}/\mathbb{P}\big(\mathsf{typ}\big).$$

Proof. This proof is a direct calculation. By (5.1.7b), using Bayes's rule, specifically the fact that $\mathbb{P}(B \mid C)/\mathbb{P}(C \mid B) = \mathbb{P}(B)/\mathbb{P}(C)$ for non-null events B and C, for $L \geq d+2$ we deduce that

$$\begin{aligned} n \textstyle\sum_{|I| \geq L} \mathbb{P}\big(V \cdot Z \equiv 0,\, \mathcal{I} = I \mid \mathsf{typ}\big) &= n \textstyle\sum_{|I| \geq L} \mathbb{P}\big(V \cdot Z \equiv 0 \mid \mathcal{I} = I,\, \mathsf{typ}\big)\mathbb{P}\big(\mathcal{I} = I \mid \mathsf{typ}\big) \\ &\leq \textstyle\sum_{|I| \geq L} \big(\mathbb{P}\big(\mathcal{I} = I \mid \mathsf{typ}\big) + 3 \cdot 2^{d-|I|}\,\mathbb{P}\big(\mathcal{I} = I\big)/\mathbb{P}\big(\mathsf{typ}\big)\big) \\ &\leq \mathbb{P}\big(|\mathcal{I}| \geq L \mid \mathsf{typ}\big) + 3 \cdot 2^{d-L}\mathbb{P}\big(|\mathcal{I}| \geq L\big)/\mathbb{P}\big(\mathsf{typ}\big) \leq 1 + 3 \cdot 2^{d-L}/\mathbb{P}\big(\mathsf{typ}\big). \end{aligned}$$

□

Proof of Lemma 5.1.15. The definition of r_* from (5.1.8) along with (5.1.5) immediately imply the final claim (5.1.8). Write $\overline{\mathbb{P}}$ and $\overline{\mathbb{E}}$ to denote probability and expectation, respectively, conditioned on $\mathcal{I} = I$ and typ_ℓ (ie *local* typicality). As for (5.1.5), we obtain

$$n\,\mathbb{P}\big(V \cdot Z \equiv 0 \mid \mathcal{I} = I,\, \mathsf{typ}\big) \leq 1 + \mathbb{E}\big(\mathfrak{g}^d - 1 \mid \mathcal{I} = I,\, \mathsf{typ}\big) \leq 1 + \overline{\mathbb{E}}\big(\mathfrak{g}^d - 1\big)/\overline{\mathbb{P}}\big(\mathsf{typ}_g\big).$$

Hence, to prove (5.1.7a, 5.1.7b), we need to bound $\overline{\mathbb{E}}(g^d)$. To do this, note that

$$\overline{\mathbb{E}}\big(\mathfrak{g}^d\big) = \textstyle\sum_{\gamma=1}^{2r_*} \gamma^d\,\overline{\mathbb{P}}\big(\mathfrak{g} = \gamma\big) \leq \sum_{\gamma=1}^{2r_*} \gamma^d\,\overline{\mathbb{P}}\big(\gamma \wr V_i \,\forall\, i \in I\big).$$

Applying Lemma 5.1.13, we obtain

$$\overline{\mathbb{E}}(\mathfrak{g}^d) \le \textstyle\sum_{\gamma=1}^{2r} \gamma^{d-|I|}.$$

To bound this sum, we now consider separate cases, according to the value of $d-|I|$. In particular, we can summarise all these cases in the following way:

$$\overline{\mathbb{E}}(g^d) \le \begin{cases} 1+3\cdot 2^{d-|I|} & \text{when} \quad |I|-d \ge 2, \\ C(2r_*)^{d-|I|+2} & \text{when} \quad |I|-d \le 1, \end{cases}$$

where C is the implicit constant in the previous equation. We thus deduce (5.1.7a, 5.1.7b). □

We now consider the probability of a given realisation of $\mathcal{I}$. Recall that $t \coloneqq t_\alpha$ still.

Lemma 5.1.17. *We have*

$$\mathbb{P}\big(\mathcal{I}=I,\ \mathsf{typ}\big) \le n^{-1}e^{-\omega}/p_*^{|I|} = e^{-\omega}n^{-1+|I|/k}k^{2|I|}. \tag{5.1.9}$$

Proof. Requiring $\mathcal{I}=I$ places restrictions on the coordinates in I^c, but not on the coordinates of I other than that they are non-zero; we ignore the latter to get an upper bound (see below).

For a vector $w \in \mathbb{Z}^k$, write

$$\mathcal{W}_I(w) \coloneqq \big\{w' \in \mathbb{Z}^k \mid \mathcal{I}(w-w') = I\big\}.$$

Then, using the independence of W and W', we have

$$\mathbb{P}\big(\mathcal{I}=I,\ W\in\mathcal{W}\big) = \textstyle\sum_{w\in\mathcal{W}} \mathbb{P}\big(W=w\big)\mathbb{P}\big(W'\in\mathcal{W}_I(w)\big).$$

Hence, using the independence of the coordinates of W', given $w\in\mathcal{W}$ we have

$$\mathbb{P}\big(W'\in\mathcal{W}_I(w)\big) = \mathbb{P}\big(W'=w\big)\cdot\prod_{i\in I}\frac{\mathbb{P}(W_i'\ne w_i)}{\mathbb{P}(W_i'=w_i)} \le \mathbb{P}\big(W'=w\big)\cdot\prod_{i\in I}\frac{1}{\mathbb{P}(W_i'=w_i)}.$$

An immediate consequence of the definitions of r and p, in Definition 5.1.8a, is that,

$$\text{for all } \alpha\in\mathbb{R}, \quad \text{if} \quad \big|w_1-\mathbb{E}\big(W_1(t_\alpha)\big)\big| \le r_\alpha(k,n) \quad \text{then} \quad \mathbb{P}\big(W_1(t_\alpha)=w_1\big) \ge p_\alpha(k,n).$$

By Proposition 5.1.14, we have $p_\alpha \ge p_*$. Hence, for $w\in\mathcal{W}$, we then obtain

$$\mathbb{P}\big(W'\in\mathcal{W}_I(w)\big) \le \mathbb{P}\big(W'=w\big)/p_*^{|I|} \le n^{-1}e^{-\omega}/p_*^{|I|}.$$

From this and the sum above, (5.1.9) follows by summing over all $w\in\mathcal{W}$:

$$\mathbb{P}\big(\mathcal{I}=I,\ \mathsf{typ}\big) \le \mathbb{P}\big(\mathcal{I}=I,\ W\in\mathcal{W}\big) \le n^{-1}e^{-\omega}p_*^{-|I|}\textstyle\sum_{w\in\mathcal{W}}\mathbb{P}\big(W=w\big) \le n^{-1}e^{-\omega}p_*^{-|I|};$$

finally we substitute the definition $p_* = n^{-1/k}k^{-2}$ from Definition 5.1.8a. □

We have now done all the hard work in proving Proposition 5.1.10, from which we deduced Theorem 5.1.4. It remains to go through the details of how to combine the previous results; there are no more interesting ideas to prove the proposition, but the details are quite technical.

Proof of Proposition 5.1.10. Assume that (k,G) satisfy Hypothesis I. Set $L \coloneqq \frac{1}{15}\log n/\log k$; this satisfies $d \le \frac{1}{2}L$. Also, $\log k \ll \log n$, so $L \gg 1$ and hence also $L-d \gg 1$.

Consider first $I\subseteq[k]$ with $1\le|I|\le L$. We have

$$\begin{aligned} n\,\mathbb{P}\big(V\cdot Z\equiv 0,\ \mathcal{I}=I,\ \mathsf{typ}\big) &= n\,\mathbb{P}\big(V\cdot Z\equiv 0 \mid \mathcal{I}=I,\ \mathsf{typ}\big)\,\mathbb{P}\big(\mathcal{I}=I,\ \mathsf{typ}\big) \\ &\overset{(5.1.8,\,5.1.9)}{\le} (2r_*)^d\cdot n^{-1}e^{-\omega}p_*^{-|I|} \\ &\overset{(5.1.6)}{\le} n^{d/k}(\log k)^{2d}\cdot n^{-1}e^{-\omega}\cdot n^{|I|/k}k^{2|I|} \\ &= e^{-\omega}n^{-1+(d+|I|)/k}k^{2|I|}(\log k)^{2d}. \end{aligned}$$

We now sum over the I with $1 \le |I| \le L$:

$$n \sum_{1 \le |I| < L} \mathbb{P}\big(V \cdot Z \equiv 0,\, \mathcal{I} = I \mid \mathsf{typ}\big) \le L k^L n^{-1+(d+L)/k} k^{2L+d-1},$$

since $\binom{k}{\ell} \le k^\ell \le k^L$ for $\ell \le L$. We now use the fact that $d + |I| \le \frac{3}{2}L = \frac{1}{10} \log n / \log k$ and $k \ge \frac{1}{2} \log n / \log\log n$ to deduce that $(d + |I|)/k \le \frac{2}{5}$. Also, since $d \le L$, we have

$$k^{3|I|+d} \le e^{4L \log k} = e^{4 \log n/10} = n^{2/5},$$

by definition of L. Hence

$$n \sum_{1 \le |I| \le L} \mathbb{P}\big(V \cdot Z \equiv 0,\, \mathcal{I} = I,\, \mathsf{typ}\big) \le n^{-1+2/5+2/5} = n^{-1/5}. \tag{5.1.10}$$

Finally we consider $I \subseteq [k]$ with $L \le |I| \le k$. By Corollary 5.1.16, we have

$$n \sum_{L \le |I| \le k} \mathbb{P}\big(V \cdot Z \equiv 0,\, \mathcal{I} = I \mid \mathsf{typ}\big) \le 1 + 3 \cdot 2^{d-L} / \mathbb{P}\big(\mathsf{typ}\big). \tag{5.1.11}$$

Plugging (5.1.4, 5.1.10, 5.1.11) into (5.1.3), recalling that $L - d \gg 1$, we obtain

$$D = n \sum_I \mathbb{P}\big(V \cdot Z \equiv 0,\, \mathcal{I} = I \mid \mathsf{typ}\big) - 1 = o(1)/\mathbb{P}\big(\mathsf{typ}\big) = o(1). \qquad \square$$

5.2 Cutoff: A Detailed Investigation of $\mathbb{Z}_p^d$

In this section we perform a detailed analysis of the behaviour of the mixing time for the random walk on the uniform random Cayley graph of degree k of $\mathbb{Z}_p^d$. In Chapter 2 we established cutoff in this set-up, under the assumption that $k - d \gg 1$. If $G = \mathbb{Z}_p^d$ for p prime, then

$$\big\{\gamma G \mid \gamma \wr n\big\} = \big\{\gamma G \mid \gamma \wr p\big\} = \{G\} \cup \{pG\} = \{G,\, \{\mathsf{id}\}\};$$

that is, the only options are the group itself and the trivial group, corresponding to $\gamma = 1$ and $\gamma = p$, respectively. Thus, applying Theorem 2.2.6, we deduce that there is cutoff at the entropic time $t_0(p, |G|) = t_0(p, p^d)$, ie the time at which the entropy of the RW on $\mathbb{Z}_p^k$ becomes $\log |G| = d \log p$.

In this exposition, we consider some cases not covered in Chapter 2. In particular, we allow $k - d$ to be a fixed constant, not diverging. When this is the case and p diverges, it can be shown that choosing k elements $Z_1, ..., Z_k \sim^{\text{iid}} \text{Unif}(\mathbb{Z}_p^d)$ generates the group whp. On the other hand, if p is also fixed, then this is not the case; we establish cutoff conditional on generating the group.

To ease notation, we drop completely any p-s, and often drop the p; to be explicit, we state in the next subsection precisely what notation we are going to use.

5.2.1 Entropic Times: Methodology, Definition and Concentration

We use an 'entropic method'; for further details, see Chapter 2. To make this as self-contained as possible, we now explain the specific application in a little more depth.

We define an auxiliary random process $(W(t))_{t \ge 0}$, recording how many times, mod p, each generator has been used: for $t \ge 0$, for each generator $i = 1, ..., k$, 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 $\mathbb{Z}_p^k$. For the undirected case, recall that we either apply a generator or its inverse; when we apply the inverse of generator i, increment $W_i \to W_i - 1$ (rather than $W_i \to W_i + 1$). In this case, $W(\cdot)$ is a SRW (rather than DRW) on $\mathbb{Z}_p^k$. Note that every element of $G = \mathbb{Z}_p^d$ has order p, since p is prime. Hence it suffices to look at the walk W mod p, ie on $\mathbb{Z}_p^k$, rather than 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 invariant distribution is uniform on G, giving mass $1/n$ to each vertex. The proposed mixing time is then the time at which the auxiliary process W obtains entropy $\log n$. This time will be calculated fairly precisely in many situations; see Proposition 5.2.2.

Write μ_t, respectively ν_s, for the law of $W(t)$, respectively $W_1(sk)$; so $\mu_t = \nu_{t/k}^{\otimes k}$. Define

$$Q(t) := -\log \mu_t\big(W_i(t)\big) \quad \text{and} \quad Q_i(t) := -\log \nu_{t/k}\big(W_i(t)\big);$$

then, Q_i forms an iid sequence over $i \in [k]$, and

$$Q(t) = \textstyle\sum_{i=1}^k Q_i(t), \quad h(t) := \mathbb{E}\big(Q(t)\big) \quad \text{and} \quad H(s) := \mathbb{E}\big(Q_1(sk)\big).$$

So $h(t)$ and $H(s)$ are the entropies of $W(t)$ and $W_1(sk)$, respectively. Note that $h(t) = kH(t/k)$ and that $h : [0,\infty) \to [0, \log(p^k))$ is a strictly increasing bijection.

While all our results can be phrased in terms of (Shannon) entropy, from a technical point of view it will be convenient to define the *relative entropy*:

$$R(s) := \log p - H(s).$$

The maximal entropy of a random variable on $\mathbb{Z}_p$ is $\log p$, obtained uniquely by the uniform distribution. Since the RW converges to the uniform distribution, $R(s) \to 0$ as $s \to \infty$. Of great importance will be the parameter

$$\zeta := \tfrac{1}{k}(k-d)\log p = \log p - \tfrac{1}{k}\log n \quad \text{where} \quad n := |\mathbb{Z}_p^d| = p^d.$$

Definition 5.2.1. *Define* $\zeta := \frac{1}{k}(k-d)\log p = \log p - \frac{1}{k}\log n$, *and, for* $\alpha \in \mathbb{R}$, *define*

$$\zeta_\alpha := \zeta\big(1 - 2\alpha/\sqrt{\zeta k(\zeta \vee 1)}\big), \quad s_\alpha := H^{-1}(\zeta_\alpha) \quad \text{and} \quad t_\alpha := s_\alpha k.$$

5.2.2 Entropic Times: Evaluation and Concentration

In this subsection, we estimate the entropic times in different regimes, and give a concentration result. The proofs are given in §6.2.4; precise references are given at the appropriate times.

The first proposition estimates the entropic times t_0 and the difference $t_\alpha - t_0$; the second gives concentration of the Q random variable around these times.

Proposition 5.2.2a (Proposition 6.2.25a)**.** *Suppose that* $1 \ll k \lesssim d\log p$. *The following hold:*

$$\begin{aligned} &\text{if} \quad \zeta \ll 1, \quad \text{then} \quad t_0/k = s_0 \eqsim \tfrac{1}{2}\log(1/\zeta)/\big(1-\cos(2\pi/p)\big);\\ &\text{if} \quad \zeta \gtrsim 1, \quad \text{then} \quad t_0/k = s_0 \asymp p^2 e^{-2\zeta} = (p^d)^{2/k};\end{aligned}$$

further, if in fact $1 \ll k \ll d\log p$, *then*

$$\text{if} \quad \zeta \gg 1, \quad \text{then} \quad t_0/k = s_0 \eqsim p^2 e^{-2\zeta}/(2\pi e) = (p^d)^{2/k}/(2\pi e).$$

Note that $1 - \cos(2\pi/p) \eqsim_{p\to\infty} 2\pi^2/p^2 = 2\pi^2 p^{-2d/k} e^{2\zeta}$.

Proposition 5.2.2b (Proposition 6.2.25b)**.** *Suppose that* $1 \ll k \lesssim d\log p$ *and* $(k-d)p \gg 1$, *ie* $\zeta \gg 1/k$. *Then, for all* $\alpha \in \mathbb{R}$, *we have* $t_\alpha \eqsim t_0$ *and furthermore the following hold:*

$$\begin{aligned} &\text{if} \quad \zeta \lesssim 1, \quad \text{then} \quad (t_\alpha - t_0)/t_0 \lesssim 1/\big(\sqrt{\zeta k}\log((1/\zeta)\vee e)\big) = o(1);\\ &\text{if} \quad \zeta \gg 1, \quad \text{then} \quad (t_\alpha - t_0)/t_0 \lesssim 1/\sqrt{k} = o(1) \quad \text{for the SRW.}\end{aligned}$$

Proposition 5.2.3 (Concentration, Proposition 6.2.27)**.** *For* $\alpha \in \mathbb{R}$, *define*

$$Q_\alpha^+ := \{Q(t_\alpha) \ge d\log p + \alpha\sqrt{k(\zeta\wedge 1)}\} \quad \text{and} \quad Q_\alpha^- := \{Q(t_{-\alpha}) \le d\log p - \alpha\sqrt{k(\zeta\wedge 1)}\};$$

For all $\alpha \in (0,\infty)$ *with* $|\zeta_\alpha - \zeta_0| \le \frac{1}{2}\zeta_0$, *we have* $\mathbb{P}((Q_\alpha^\pm)^c) \lesssim \alpha^{-2}$.

5.2.3 Precise Statement and Remarks

Recall that $d_Z(t)$ is the TV distance from uniform after time t with realisation Z of generators.

Theorem 5.2.4 (Cutoff). *Let G be a finite, Abelian group admitting a decomposition $G \coloneqq \mathbb{Z}_p^d$ with p prime. Assume that $1 \ll k \lesssim d \log p$. Define the entropic times $\{t_\alpha\}_{\alpha\in\mathbb{R}}$ as in Definition 5.2.1. The entropic times are asymptotically evaluated in Proposition 5.2.2.*

Suppose that $(k-d)p \gg 1$, ie $\zeta k \gg 1$. Then the RW on G_k exhibits cutoff whp at t_0. More precisely, choose a sequence $(\beta_N)_{N\in\mathbb{N}} \subseteq \mathbb{R}_+^{\mathbb{N}}$ with $\beta_N \to \infty$ (arbitrarily slowly) and let $c \in \{\pm 1\}$. Then

$$d^{\pm}_{G_k,N}(t_{c\beta_N}) \to^{\mathbb{P}} \mathbf{1}(c=-1) \quad \textit{(in probability)} \quad \textit{as } N \to \infty.$$

Moreover, the implicit lower bound holds deterministically, ie for all choices of generators.

Also, if $0 \le (k-d)p \lesssim 1$, then, conditional that the uniformly chosen multisubset $[Z_1, ..., Z_k]$ generates the group, there is cutoff whp at time $\frac{1}{2} d \log d/(1-\cos(2\pi/p))$.

Remark 5.2.5. The outline of the proof is the same for all ζ with $\zeta k \gg 1$; we assume this initially.

We also consider the case where $0 \le k-d = \mathcal{O}(1)$ conditional on generating the group. This uses a standard argument for the case $k = d$, and then compares the walk which uses $Z = [Z_1, ..., Z_k]$ with another walk using a subset Z' of size d which generates the group.

We explain how to do this at the end of the section in §5.2.6. △

Remark. Prior to our work, cutoff had already been established for any Abelian group when $k \gg \log n$, with an explicit mixing time; see §1.5.2. Although our technique can be adapted to allow $k \gtrsim \log n$ (when $n = |\mathbb{Z}_p^d|$, this is equivalent to $k \gtrsim d \log p$), we do not give details here. △

5.2.4 Lower Bound on Mixing for $\mathbb{Z}_p^d$

In this subsection, we prove the lower bound on mixing, which holds for every choice of Z. (This argument is almost identical to the one which we give in §2.1.5.)

Proof of Lower Bound. For this proof, we assume that Z is given, and suppress it.

The concentration result Proposition 5.2.3 gives $\mathbb{P}(Q_\alpha^-) \to 1$ as $\alpha \to \infty$. Consider the set

$$A_\alpha \coloneqq \big\{x \in G \;\big|\; \exists\, w \in \mathbb{Z}^d \text{ st } \mu_{t_\alpha}(w) \ge n^{-1}e^{\omega} \text{ and } x = w \cdot Z\big\}.$$

Since we use W to generate S, we have $\mathbb{P}(S(t_\alpha) \in A_\alpha \mid \mathcal{E}_\alpha) = 1$. Every element $x \in A_\alpha$ can be realised as $x = w_x \cdot Z$ for some $w_x \in \mathbb{Z}^k$ with $\mu_{t_\alpha}(w_x) \ge n^{-1}e^{\omega}$. Hence, for all $x \in A_\alpha$, we have

$$\mathbb{P}\big(S(t_\alpha) = x\big) \ge \mathbb{P}\big(W(t_\alpha) = w_x\big) = \mu_{t_\alpha}(w_x) \ge n^{-1}e^{\omega}.$$

From this we deduce that

$$1 \ge \textstyle\sum_{x\in A_\alpha} \mathbb{P}\big(S(t_\alpha) = x\big) \ge |A_\alpha| \cdot n^{-1}e^{\omega}, \quad \text{and hence} \quad |A_\alpha|/n \le e^{-\omega} = o(1).$$

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

$$\big\|\mathbb{P}\big(S(t_\alpha) \in \cdot \mid Z\big) - \pi_G\big\|_{\mathrm{TV}} \ge \mathbb{P}\big(S(t_\alpha) \in A_\alpha\big) - \pi_G(A_\alpha) \ge \mathbb{P}\big(Q_\alpha^-\big) - \tfrac{1}{n}|A_\alpha| = 1 - o(1). \qquad \square$$

5.2.5 Upper Bound on Mixing for $\mathbb{Z}_p^d$ for $(k-d)p \gg 1$

This subsection is devoted to the upper bound.

Outline of Proof. Consider α with $\alpha \to \infty$, but arbitrarily slowly. We show that the TV distance from uniform is $o(1)$ on the event Q_α^+ (whp over Z) and that $\mathbb{P}(Q_\alpha^+) = 1 - o(1)$, using similar techniques to those in §2.1.6. Theorem 5.2.4 follows from this and Propositions 5.2.2 and 5.2.3. △

We now make this outline precise and rigorous. Herein, we frequently suppress the time and α-subscripts, eg writing W for $W(t_\alpha)$ or $W(t)$, depending on context.

Key is a 'modified L_2 calculation'; cf Lemma 2.1.6. In short, one condition that W is 'typical' (in some precise sense), and then applies the standard TV–L_2 calculation on the conditioned law.

Let W' be an independent copy of W; then $S' \coloneqq W' \cdot Z$ is an independent copy of S.

Lemma 5.2.6. *For all $t \geq 0$ and all $\mathcal{W} \subseteq \mathbb{Z}_p^k$, the following inequalities hold:*

$$\big\|\mathbb{P}\big(S(t) \in \cdot \mid Z\big) - \pi_G\big\|_{\mathrm{TV}} \leq \big\|\mathbb{P}_Z\big(S(t) \in \cdot \mid W(t) \in \mathcal{W}\big) - \pi_G\big\|_{\mathrm{TV}} + \mathbb{P}\big(W(t) \notin \mathcal{W}\big); \tag{5.2.1a}$$

$$4\,\mathbb{E}\big(\big\|\mathbb{P}_Z\big(S(t) \in \cdot \mid W(t) \in \mathcal{W}\big) - \pi_G\big\|_{\mathrm{TV}}^2\big) \leq n\,\mathbb{P}\big(S(t) = S'(t) \mid W(t), W'(t) \in \mathcal{W}\big) - 1. \tag{5.2.1b}$$

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

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

as $S = W \cdot Z$, $S' = W' \cdot Z$ and $V = W - W'$. The claim follows from Jensen's inequality. □

We now make the specific choice of the 'typical' set $\mathcal{W}$; we make a different choice for each $\alpha \in \mathbb{R}$. Cf Definition 2.1.7. The collection $\{\mathcal{W}_\alpha\}_{\alpha \in \mathbb{R}}$ will satisfy

$$\mathbb{P}\big(W(t_\alpha) \notin \mathcal{W}_\alpha\big) \approx 0 \quad \text{for large } \alpha,$$

using the concentration result Proposition 5.2.3. We show that the expression (5.2.1b) is $o(1)$. Then applying (5.2.1a) gives $d_Z(t_\alpha) \approx 0$ whp over Z, for large α.

Definition 5.2.7. *For all $\alpha \in \mathbb{R}$, define $\omega_\alpha \coloneqq \alpha\sqrt{k(\zeta \wedge 1)} \gg 1$,*

$$\mathcal{W}_\alpha \coloneqq \big\{w \in \mathbb{Z}_p^k \mid \mathbb{P}\big(W(t_\alpha) = w\big) \leq n^{-1} e^{-\omega_\alpha}\big\}, \quad \textit{and} \quad \mathsf{typ}_\alpha \coloneqq \big\{W(t_\alpha), W'(t_\alpha) \in \mathcal{W}_\alpha\big\}.$$

The following proposition determines the probability that $W(t_\alpha)$ lies in $\mathcal{W}_\alpha$, ie of typicality.

Lemma 5.2.8. *For all $\alpha \in (0, \infty)$ with $|\zeta_\alpha - \zeta_0| \leq \frac{1}{2}\zeta_0$, we have*

$$\mathbb{P}\big(W(t_\alpha) \notin \mathcal{W}_\alpha\big) \lesssim \alpha^{-2}.$$

Proof. The lemma follows immediately from Proposition 5.2.3, since $\{W(t_\alpha) \in \mathcal{W}_\alpha\} = Q_\alpha^+$. □

Herein, inside proofs we often drop the time dependence and α-subscripts from the notation, eg writing W for $W(t_\alpha)$ and $\mathcal{W}$ for $\mathcal{W}_\alpha$ or typ for typ_α. The 'typical set' $\mathcal{W}$ is designed precisely so that the following lemma holds.

Lemma 5.2.9. *For all $\alpha \in (0, \infty)$, we have*

$$\mathbb{P}\big(W(t_\alpha) = W'(t_\alpha) \mid \mathsf{typ}_\alpha\big) \leq n^{-1} e^{-\omega_\alpha} / \mathbb{P}\big(\mathsf{typ}_\alpha\big) \ll n^{-1}.$$

Proof. By direct calculation, using independence of W and W', we have

$$\mathbb{P}\big(W = W',\, \mathsf{typ}\big) = \mathbb{P}\big(W = W',\, W \in \mathcal{W}\big) = \textstyle\sum_{w \in \mathcal{W}} \mathbb{P}\big(W = w\big)^2 \leq n^{-1} e^{-\omega},$$

with the final inequality using global typicality. The result follows by Bayes's rule. □

When $W = W'$, we necessarily have $S = S'$ (since the group is Abelian). Now consider when $W \neq W'$. The following lemma is a special case of Lemma 2.1.11.

Lemma 5.2.10 (Lemma 2.1.11)**.** *For any $v \in \mathbb{Z}_p^k \setminus \{0\}$, we have $v \cdot Z \sim \mathrm{Unif}(G)$.*

Corollary 5.2.11. *For all $\alpha \in \mathbb{R}$, we have*

$$\mathbb{P}\big(S(t_\alpha) = S'(t_\alpha),\ W(t_\alpha) \neq W'(t_\alpha) \mid \mathsf{typ}_\alpha\big) \leq \tfrac{1}{n}.$$

Proof. Condition on $W = w$ and $W' = w'$ with $w \neq w'$ and $w, w' \in \mathcal{W}$. This conditioning is independent of Z. Hence $S - S' = (w - w') \cdot Z \sim \mathrm{Unif}(G)$ by Lemma 5.2.10. The claim follows. □

Proof of Upper Bound in Theorem 5.2.4 Given Propositions 5.2.2 and 5.2.3. We are assuming that $(k-d)p \gg 1$, ie $\zeta k \gg 1$. This means that $|\zeta_\alpha - \zeta_0| \leq \frac{1}{2}\zeta_0$ for all $\alpha \in \mathbb{R}$. Choose α with $\alpha \to \infty$, arbitrarily slowly. We need to show that $d_Z(t_\alpha) = o(1)$ whp over Z.

Apply Lemma 5.2.9 and Corollary 5.2.11 to deduce that $\mathbb{P}(S = S' \mid \mathsf{typ}) = o(1)$. Apply the modified L_2 calculation of Lemma 5.2.6 using Definition 5.2.7 for the definition of typicality. Bound the 'error term' using Lemma 5.2.8. This gives $d_Z(t_\alpha) = o(1)$ whp. □

5.2.6 Removing the Condition $(k-d)p \gg 1$

In this subsection, we explain how to remove the condition $(k-d)p \gg 1$ in Theorem 5.2.4 conditional on generating the group, as referenced in Remark 5.2.5. There are two cases to consider: $k = d$ with p arbitrary (allowed to diverge) and $0 < k - d = \mathcal{O}(1)$ with p a fixed prime.

Case $k = d$. Here we do not need to assume that p is prime. Note also that $d = k \gg 1$. The occurrence of cutoff in this set-up is not difficult. As we could not find a proof in the literature—note that p is not assumed to be fixed—we give the details.

A key observation is that if Z' is a set of size d that generates $\mathbb{Z}_p^d$, then the Cayley graph with respect to it is isomorphic to the Cayley graph with respect to the standard basis $\{e_1, ..., e_d\}$. Namely, it is the d-fold Cartesian product chain of the p-cycle with itself.

For the lower bound, we combine the method of distinguishing statistics and Wilson's method [78]. Let $f_2(y) := \cos(2\pi y/p)$ and $\lambda_p := \cos(2\pi/p)$. Then λ_p is the second largest eigenvalue of the transition matrix of SRW on $\mathbb{Z}_p$, and the corresponding eigenvector is f_2. Then $f(x_1, ..., x_d) := \frac{1}{d}\sum_{i=1}^d f_2(x_i)$ is an eigenvector of the transition matrix of SRW on $\mathbb{Z}_p^d$ with eigenvalue

$$\Lambda_{p,d} := (d-1)/d + \lambda_p/d = 1 - (1-\lambda_p)/d.$$

We use initial state $(0, ..., 0)$. To apply the method of distinguishing statistics, we need to bound both the expectation and the variance of f, both under the uniform and the RW distributions (at time t); write these as π and μ_t, respectively. Under the uniform distribution, since the coordinates are independent and $|f_2(z)| \leq 1$ for all $z \in \mathbb{Z}_p$, we have $\mathrm{Var}_\pi(f) \leq \frac{1}{d}$; similarly, we have $\mathrm{Var}_{\mu_t}(f) \leq \frac{1}{d}$. Also, since f is an eigenvector, we have $E_\pi(f) = 0$ and $E_{\mu_t}(f) = e^{-(1-\Lambda_{p,d})t}$. Applying the method of distinguishing statistics, eg as stated in [49, Proposition 7.12], for all $\varepsilon \in (0,1)$, whenever $e^{-(1-\Lambda_{p,d})t} \geq C_\varepsilon/\sqrt{d}$, for a sufficient large constant C_ε, we have $t \leq t_{\mathrm{mix}}(1-\varepsilon)$.

Rearranging $e^{-(1-\Lambda_{p,d})t} \geq C_\varepsilon/\sqrt{d}$ and recalling the definitions of $\Lambda_{p,d}$ and λ_p, we obtain

$$t \leq \frac{\frac{1}{2}\log d - \log C_\varepsilon}{1-\Lambda_{p,d}} = \frac{\frac{1}{2}d\log d - d\log C_\varepsilon}{1-\lambda_p} = \frac{\frac{1}{2}d\log d}{1-\lambda_p}\cdot\big(1-o(1)\big).$$

We now prove a matching upper bound on the mixing time. Let P_t be the time-t transition probabilities of the walk on $\mathbb{Z}_p^d$. We identify this walk with the aforementioned d-fold Cartesian product of the p-cycle with itself. Using independence of the coordinates, we have

$$p^d\, P_{2t}(x,x) - 1 = \big(p\, Q_{2t/d}(0,0)\big)^d - 1,$$

where Q_s is the time-s transition kernel for a rate-1 SRW on $\mathbb{Z}_p$. This is the L_∞ distance at time $2t$, and hence the square of the L_2 distance at time t. Let $\varepsilon > 0$ be a constant. Using the fact that

$$(1 + \tfrac{1}{2}\varepsilon^2/d)^d \leq \varepsilon^2$$

when ε is sufficiently small, we have

$$p^d\, P_{2t}(x,x) - 1 \leq \varepsilon \quad \text{when} \quad p\, Q_{2t/d}(0,0) - 1 \leq \tfrac{1}{2}\varepsilon^2/d.$$

Using the eigenvalue representation,

$$\text{if} \quad 2t/d \geq (1+\varepsilon)\log(2d/\varepsilon^2)/(1-\lambda_p), \quad \text{then} \quad p\,Q_{2t/d}(0,0) - 1 \leq \tfrac{1}{2}\varepsilon^2/d,$$

and hence $p^d P_{2t}(x,x) - 1 \leq \varepsilon^2$. Thus

$$\text{if} \quad t \geq \frac{\frac{1}{2}(1+\varepsilon)d\log d + \frac{1}{2}(1+\varepsilon)d\log(2/\varepsilon^2)}{1-\lambda_p} = \frac{\frac{1}{2}d\log d}{1-\lambda_p}\cdot\big(1+o(1)\big) \quad \text{then} \quad d_2(t) \leq \varepsilon,$$

provided ε is sufficiently small. This upper bound matches our lower bound. □

Case $0 < k - d = \mathcal{O}(1)$. When $p \gg 1$, we already established cutoff, and so the group is generated, whp. Thus we may assume that both p and $k-d$ are fixed, but $k \geq d \gg 1$.

The following statement is key: if $Z_1, ..., Z_k$ generate $\mathbb{Z}_p^d$ for p prime, then there exists a set $S \subseteq [k]$ such that $|S| = d$ and $\{Z_\ell\}_{\ell\in S}$ generate $\mathbb{Z}_p^d$. This is immediate by viewing $\mathbb{Z}_p^d$ as a vector space over the field $\mathbb{F}_p$ and noting that a set generates $\mathbb{Z}_p^d$ if and only if it spans it.

We begin by obtaining an upper bound. Choose a subset of generators of size d which generate the group. We consider the walk on the Cayley graph corresponding to this subset of generators. In the natural realisation of this walk, each coordinate is updated at rate $1/d$; we want it to be updated at rate $1/k$. If this walk is mixed, then since the walk on G_k is obtained by a random independent shift of that walk, the walk on G_k is also mixed. Hence the previous entropic upper bound on the mixing time from the case $k = d$ is still valid, after multiplication by $k/d = 1 + \mathcal{O}(1/d) = 1 + o(1)$, due to replacing the rate $1/d$ by $1/k$.

We now turn to the lower bound. Set $\zeta := \frac{1}{k}(k-d)\log p \asymp 1/k \ll 1$. Since $k > d$, we can apply our argument from §5.2.4. To bound the variance, we can no longer assume that $|\zeta_\alpha - \zeta_0| \leq \frac{1}{2}\zeta_0 = \frac{1}{2}\zeta$, since $\zeta k \asymp 1$. (Recall the definition of ζ_α from Definition 5.2.1.) This means that there is an extra factor of $\zeta_\alpha/\zeta_0 = 1 - 2\alpha/\sqrt{\zeta k} \asymp 1$ multiplying the variance. For the lower bound, we need only consider $\alpha < 0$ with $|\alpha|$ large. Multiplying the variance by a constant only changes the window by a constant—it does not affect the occurrence of cutoff. Hence the entropic time lower bound is still valid. Also, as $k/d = 1 + o(1)$, multiplying it by k/d does not affect the leading order.

Finally we must asymptotically evaluate this entropic time when $0 < (k-d)p = \mathcal{O}(1)$. Up to sot, by Proposition 5.2.2 it equals the desired time:

$$\tfrac{1}{2}k\log\big(k(k-d)^{-1}/\log p\big)/\big(1-\cos(2\pi/p)\big) \asymp \tfrac{1}{2}d\log d/\big(1-\cos(2\pi/p)\big), \qquad \square$$

5.3 Cutoff: From Heisenberg to General Nilpotent Groups

Most of the following discussion is based on observations made by Péter Varjú during discussions of our work with him. A group is *nilpotent of step at most ℓ* if all iterated commutators of order at least $\ell+1$ vanish necessarily. For example, step-1 is Abelian; step-2 has $[[g_1,g_2],g_3] = \mathsf{id}$ for all g_1, g_2 and g_3, ie the commutator subgroup is central. Our analysis has focussed on Heisenberg matrix groups; these are a canonical class of nilpotent groups—$H_{p,d}$ is step-$(d-2)$ nilpotent. However, some of our analysis does extend somewhat to more general nilpotent, as we now explain.

Recall that we wrote S for the location of the walk and W for its auxiliary variable; let W' be an independent copy of W, and define S' correspondingly. As previously, we work in the directed regime; so in the word S there are no inverses. Recall the definition of $C_{i,j}$ from (3.1.10):

$$C_{i,j} := \textstyle\sum_{\ell=1}^N \mathbf{1}\big(G_\ell = j\big)\sum_{m=1}^{\ell-1}\mathbf{1}\big(G_m = i\big) \quad \text{and} \quad C_{i,i} := 0 \quad \text{for all} \quad i,j \in [k],$$

where there are N steps and G_m is the index of the generator chosen in step m.

Lemma 5.3.1. *Up to multiplication by an element of $[G, G^{\mathrm{com}}]$, we can express S as*

$$S = \big(\textstyle\prod_1^k Z_i^{W_i}\big)\cdot\big(\prod_{i<j}[Z_i^{-1}, Z_j^{-1}]^{-C_{i,j}}\big)$$

If G is step-2 nilpotent then $[G, G^{\mathrm{com}}] = \{\mathsf{id}\}$ is the trivial group.

(The second product is unordered, since we are working up to an element of $[G, G^{\mathrm{com}}]$, and so we may assume that commutators commute with any element of G; the first is ordered $i = 1, ..., k$.)

Sketch of Proof. Writing a rigorous proof of this lemma is technical, and can obscure what is going on; we use an example to demonstrate how to prove the lemma. In essence, we wish to move all the Z_1-s to the left, then all the Z_2-s to the left-but-one and so on. To reverse the order terms, we use the fact that $hg = ghh^{-1}g^{-1}gh = gh[h^{-1}, g^{-1}]$ and $[h^{-1}, g^{-1}] = [g^{-1}, h^{-1}]^{-1}$. For example,

$$ghhg = gh \cdot gh[g^{-1}, h^{-1}]^{-1} = g \cdot gh[g^{-1}, h^{-1}]^{-1} \cdot h[g^{-1}, h^{-1}]^{-1} = g^2h^2[g^{-1}, h^{-1}]^{-2}.$$

To move Z_i past Z_j, with $i < j$, for each occurrence of Z_i we need to count the number of times that Z_j appears before it in the word; this is precisely (the definition of) $C_{i,j}$. □

Expressing $S^{-1}S'$ as a similar product, it is straightforward to see what we get when $W = W'$. (We actually only need $W_i \equiv W_i'$ mod ord Z_i for each i, but $W = W'$ is generally easier to analyse.)

Corollary 5.3.2. *If $W = W'$, then $S^{-1}S' \in [G, G]/[G, [G, G]]$. The converse holds in the free group, ie when considering $Z_1, ..., Z_k$ as formal variables (ie with no relations between them).*

If $W = W'$, then, up to multiplication by an element of $[G, G^{\mathrm{com}}]$, we can express $S^{-1}S'$ as

$$S^{-1}S' = \textstyle\prod_{i<j}[Z_i^{-1}, Z_j^{-1}]^{D_{i,j}} \quad \text{where} \quad D_{i,j} \coloneqq C_{i,j} - C'_{i,j}; \quad \text{write} \quad D \coloneqq (D_{i,j})_{i,j}.$$

In particular, if $C_{i,j} = C'_{i,j}$ for all i and j (which implies that $W_i = W_i'$ for all i by taking $i = j$), then $S^{-1}S' \in [G, G^{\mathrm{com}}]$; if the group is step-2 nilpotent, then $[G, G^{\mathrm{com}}] = \{\mathsf{id}\}$, and hence $S = S'$.

Consider now step-2 nilpotent groups, of which $H_{p,3}$ is an example. We are interested in analysing $\mathbb{P}(S = S' \mid \mathsf{typ})$; typicality will primarily involve entropic considerations. For ease of presentation, here we drop typ from the notation. As in §3.1, we separate this probability as

$$\mathbb{P}\big(S = S'\big) \le \mathbb{P}\big(S = S' \mid W = W'\big)\mathbb{P}\big(W = W'\big) + \mathbb{P}\big(S = S' \mid W \ne W'\big).$$

Typicality (entropy) bounds $\mathbb{P}(W = W') \ll 1/|G^{\mathrm{ab}}| = |G^{\mathrm{com}}|/|G|$, as for Heisenberg groups.

Assume that $t \ll k$, and that every generator is picked at most once—eg, this is the case if $k \gg \log n$. The assumption means that some generator is picked once in S and never in S' (or vice versa); this will allow us to deduce that $S^{-1}S' \sim \mathrm{Unif}(G)$, and hence $\mathbb{P}(S = S' \mid W \ne W') = 1/n$.

Since $S = S'$ when $D_{i,j} = 0$ for all i and j, we have

$$\mathbb{P}\big(S = S' \mid W = W'\big) \le \mathbb{P}\big(S = S' \mid W = W',\, D \ne 0\big) + \mathbb{P}\big(D = 0\big).$$

When the nilpotent group is of higher step, the bound $\mathbb{P}(S = S' \mid D = 0) \le 1$ may be too crude. We analysed $\mathbb{P}(D = 0)$ in §3.1, obtaining $\mathbb{P}(D = 0) \approx 1/t!$. We desire this to be close to $1/|G^{\mathrm{com}}|$.

We wish to get $\mathbb{P}(S = S' \mid W = W',\, D \ne 0)$ close to $1/|G^{\mathrm{com}}|$. To do this, write

$$S^{-1}S' = \textstyle\prod_{i<j:D_{i,j}\ne 0}\big[Z_i^{-1}, Z_{j'}^{-1}\big]^{D_{i,j}}.$$

While these commutators are neither uniformly random nor independent, we aim to have suitably many $D_{i,j} \ne 0$ so that the commutator product is sufficiently close to uniform (on G^{com}).

If "close" can mean "up to a sufficiently small factor", then combining all these bounds gives $\mathbb{P}(S = S',\, W = W') \ll 1/n$. The modified L_2 distance is then given by $n\,\mathbb{P}(S = S') - 1 = o(1)$.

We can apply the method for nilpotent groups of greater step, by quotienting out $[G, G^{\mathrm{com}}]$. However, as the step increases the bounds become more crude: we could have $\mathbb{P}(S = S') \ll \mathbb{P}(S^{-1}S' \in [G, G^{\mathrm{com}}])$, which would be bad for this method; this is in essence what we did for $H_{p,d}$. The analysis also applies to non-nilpotent groups, for which such issues can be even worse.

5.4 Cutoff: No Cutoff When k Is Constant

Throughout the paper we have always been assuming that $k \to \infty$ as $n \to \infty$. It is natural to ask what happens when k does not diverge. This case has actually already been covered by Diaconis and Saloff-Coste [27], using their concept of *moderate growth*. Here we give a short exposition of their results leading to the conclusion that, for nilpotent groups of bounded step, there is no cutoff—for any choice of generating set, not only when one draws the Cayley graph uniformly.

Recall that a group G is called *nilpotent of step at most L* if its lower central series terminates in the trivial group after at most L steps: $G_0 \coloneqq G$ and $G_\ell \coloneqq [G_{\ell-1}, G]$ for $\ell \in \mathbb{N}$ with $G_L = \{\mathsf{id}\}$.

Definition 5.4.1 ([27]). *Let G be a finite group. Let Z be a symmetric generating subset; that is, $\{z_1 \cdots z_r \mid r \in \mathbb{N}_0, z_1, ..., z_r \in Z\} = G$ and if $z \in Z$ then $z^{-1} \in Z$ also. For $R \in \mathbb{N}_0$, let $\mathcal{B}(R)$ denote the R-ball around the identity in $\mathcal{G}$. Write $\Delta := \inf\{R \in \mathbb{N}_0 \mid |\mathcal{B}(R)| = |G|\}$ for the diameter of $G(Z)$. We say that $G(Z)$ is of (A, d)-moderate growth if $|\mathcal{B}(R)| \geq A^{-1}|G|(R/\Delta)^a$ for all $R \in \mathbb{N}_0$.*

The main abstract result of Diaconis and Saloff-Coste [27] considers simple random walks on general Cayley graphs of moderate growth; see [27, Theorem 3.1] for a slight extension, considering more general random walks, which fundamentally gives the same conclusion.

Theorem 5.4.2 ([27, Theorem 1.2]). *Let $G(Z)$ be a Cayley graph of (A, d)-moderate growth; write $\Delta := \operatorname{diam} G(Z)$. For $t \in \mathbb{N}_0$, let $d_{\mathrm{TV}}(t)$ denote the TV distance between the law of the lazy SRW run for t steps and the uniform distribution. Let $c > 0$. Then the following hold:*

$$d_{\mathrm{TV}}\big(2(1+c)|Z|\Delta^2\big) \leq Be^{-c} \quad \text{where} \quad B := A^{1/2}2^{d(d+3)/4};$$
$$d_{\mathrm{TV}}\big(c\Delta^2/(2^{4d+1}A^2)\big) \geq \tfrac{1}{2}e^{-c}.$$

Further, the corresponding relaxation time t_{rel} satisfies $t_{\mathrm{rel}} \geq \Delta^2/(4^{2d+1}A^2)$.

The claim on the spectral is not included in the statements of Diaconis and Saloff-Coste [27]. However, the lower bound is proved precisely via the standard eigenvalue analysis; [27, (3.2)] gives the required inequality (in the notation there β_1 is the largest non-trivial eigenvalue).

Diaconis and Saloff-Coste then make the following observation, formalised below.

Corollary 5.4.3. *Let $A, d > 0$. Let $(G_N(Z_{(N)}))_{N\in\mathbb{N}}$ be a sequence of finite, undirected Cayley graphs of (A, d)-moderate growth and with $\sup_N |Z_{(N)}| < \infty$. Then the corresponding sequences of lazy simple random walks does not exhibit the cutoff phenomenon; in fact,*

$$t_{\mathrm{mix}}\big(G_N(Z_{(N)})\big)/k_N \lesssim \big(\operatorname{diam} G_N(Z_{(N)})\big)^2 \lesssim t_{\mathrm{rel}}\big(G_N(Z_{(N)})\big) \lesssim t_{\mathrm{mix}}\big(G_N(Z_{(N)})\big) \quad \text{as } N \to \infty.$$

Diaconis and Saloff-Coste apply this to nilpotent groups of bounded step.

Theorem 5.4.4 ([27, Lemma 5.1 and Theorem 5.2]). *Let G be a nilpotent group of step L. Let Z be a symmetric set of generators for G. Then $G(Z)$ is of $(A, \log_2 A)$-moderate growth for some $A := A(|Z|, L)$, depending only on the number of generators $|Z|$ and the step L.*

As a corollary of this, if the number of generators is bounded and the underlying group is nilpotent of bounded step, then the corresponding simple random walks do not exhibit cutoff.

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 $t_{\mathrm{rel}} := t_{\mathrm{rel}}(G(Z))$ for the relaxation time (ie inverse of the spectral gap) and $t_{\mathrm{mix}} := t_{\mathrm{mix}}(\varepsilon; G(Z))$ for the (TV) ε-mixing time, for $\varepsilon \in (0, 1)$. When considering sequences $(G_N(Z_{(N)}))_{N\in\mathbb{N}}$, add an N-sub/superscript.

Corollary 5.4.5 (cf [27, 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 $(t^N_{\mathrm{mix}})_{N\in\mathbb{N}}$ does not exhibit the cutoff phenomenon; in particular, $t^N_{\mathrm{mix}}/k_N \lesssim \Delta_N^2 \lesssim t^N_{\mathrm{rel}} \lesssim t^N_{\mathrm{mix}}$ as $N \to \infty$.*

5.5 Typical Distance: Generalised Graph Distance

This section focuses on distances from a fixed point in the uniform random Cayley graph of degree k of an Abelian group G. The analysis is very similar to that of §4.2 where the same statistic was studied; here we are more general. In particular, there we only considered $k \asymp \log|G|$. Here we adapt that analysis to consider $1 \ll k \ll \log|G|$; we also extend the concept of graph distance from an L_1-type concept to an L_q-type, for general $q \in [1, \infty]$.

5.5.1 Definition of L_q Typical Distance

Graphs distances in Cayley graphs have some special properties. Consider a collection $z = [z_1, ..., z_k]$ of generators and distances in the Cayley graph $G(z)$. For a path ρ in $G(z)$, for each $i \in [k]$, write $\rho_{i,+}$ for the number of times z_i is used, $\rho_{i,-}$ for the number of times z_i^{-1} is used (if in the undirected case otherwise $\rho_{i,-} \coloneqq 0$) and $\rho_i \coloneqq \rho_{i,+} - \rho_{i,-}$. The path connects the identity with $\rho \cdot z$. Then the length, in the usual graph distance, of ρ is $\|\rho\|_1 \coloneqq \sum_1^k (\rho_{i,+} + \rho_{i,-})$.

For any $q \in [1, \infty)$, define the L_q *graph distance* of ρ by $\|\rho\|_q^q \coloneqq \sum_1^k (\rho_{i,+}^q + \rho_{i,-}^q)$. For the L_∞-*graph distance*, define $\|\rho\|_\infty \coloneqq \max_i(\rho_{i,+} + \rho_{i,-})$. (The usual graph distance is given by $q = 1$.)

For Abelian groups, clearly for any $q \in [1, \infty)$ an L_q *geodesic*, ie a path of minimal length, will only use either z_i or z_i^{-1}, not both (since the terms in the product can be reordered), ie $\rho_{i,+}\rho_{i,-} = 0$ for all i. Thus $\|\rho\|_q^q = \sum_1^k |\rho_i|^q$. Similarly, any L_∞-geodesic ρ can be adjusted into a new path ρ' with $\|\rho\|_\infty = \|\rho'\|_\infty$ and $\rho'_{i,+}\rho'_{i,-} = 0$ for all i.

We define the L_q *typical distance* $\mathcal{D}_{G(z),q}(\cdot)$ analogously to $\mathcal{D}_{G(z)}(\cdot)$, ie the $q = 1$ case. When the k generators are chosen uniformly at random, we write $\mathcal{D}^{\pm}_{G_k,q}(\cdot)$, with the $\pm$-superscript indicating whether or not the Cayley graph is directed.

5.5.2 Precise Statement

For an Abelian group, we define the *dimension* and *minimal side-length*, respectively, as follows:

$$d(G) \coloneqq \min\{d \in \mathbb{N} \mid \oplus_1^d \mathbb{Z}_{m_j} \text{ is a decomposition of } G\};$$
$$m_* \coloneqq \max\{\min_{j=1,...,d} m_j \mid \oplus_1^d \mathbb{Z}_{m_j} \text{ is a decomposition of } G\}.$$

It can be shown that there exists an optimal decomposition $\{m_j\}_1^d$ for m_* with $d = d(G)$. Our main constraints will be $\limsup d/k < 1$ and $k^{1/q}n^{1/k}/m^* \ll 1$.

Hypothesis J. *The sequence $(k_N, G_N)_{N\in\mathbb{N}}$ and $q \in [1, \infty]$ jointly satisfy Hypothesis J if the following conditions hold (defining $k^{1/\infty} \coloneqq 1$ for $k \in \mathbb{N}$):*

$$\lim_N k_N = \infty, \quad \lim_N k_N/\log|G_N| = 0 \quad \text{and} \quad \lim_N k_N^{1/q}|G_N|^{1/k_N}/m_*(G_N) = 0;$$
$$\text{if} \quad q \in (1,\infty) \quad \text{then additionally} \quad k_N \le \log|G_N|/\log\log|G_N| \text{ for all } N \in \mathbb{N};$$
$$\limsup_N d_N/k_N < \begin{cases} 1 & \text{for undirected graphs,} \\ \frac{1}{2} & \text{for directed graphs.} \end{cases}$$

Finally we set up a little more notation. Make the following definitions:

$$C_q^- \coloneqq 2\,\Gamma(1/q+1)(qe)^{1/q}, \quad C_q^+ \coloneqq \tfrac{1}{2}C_q^- \quad \text{and} \quad \mathfrak{D}_q^{\pm}(k,n) \coloneqq k^{1/q}n^{1/k}/C_q^{\pm},$$

where the case $q = \infty$ is to be interpreted as the limit $q \to \infty$; eg, $C_\infty^- = 2$ and $\mathfrak{D}_\infty^+(k,n) = n^{1/k}$. When these are sequences $(k_N, G_N)_{N\in\mathbb{N}}$, for $N \in \mathbb{N}$ and $q \in [1,\infty]$, write $\mathfrak{D}_{N,q}^{\pm} \coloneqq \mathfrak{D}_q^{\pm}(k_N, |G_N|)$.

Similarly, for a sequence $(G_N)_{N\in\mathbb{N}}$ of finite groups with corresponding multisubsets $(Z_{(N)})_{N\in\mathbb{N}}$ of sizes $(k_N)_{N\in\mathbb{N}}$, for $N \in \mathbb{N}$, $\beta \in [0,1]$ and $q \in [1,\infty]$, define $\mathcal{D}_{N,q}^{\pm} \coloneqq \mathcal{D}_{G_N^{\pm}(Z_{(N)})}(\beta)$.

Theorem 5.5.1. *Let $(k_N)_{N\in\mathbb{N}}$ be a sequence of positive integers and $(G_N)_{N\in\mathbb{N}}$ a sequence of finite, Abelian groups; for each $N \in \mathbb{N}$, define $Z_{(N)} \coloneqq [Z_1, ..., Z_{k_N}]$ by drawing $Z_1, ..., Z_{k_N} \sim^{\text{iid}} \text{Unif}(G_N)$.*

Suppose that the sequence $(k_N, G_N)_{N\in\mathbb{N}}$ satisfies Hypothesis J. Then, for all $\beta \in (0,1)$, we have

$$\mathcal{D}_{N,q}^{\pm}(\beta)/\mathfrak{D}_{N,q}^{\pm} \to^{\mathbb{P}} 1 \quad \textit{(in probability)} \quad \text{as } N \to \infty.$$

Moreover, the implicit lower bound holds for all choices of generators and for all Abelian groups, only requiring the conditions in Hypothesis J which depend only on $(k_N, |G_N|)_{N\in\mathbb{N}}$ and q.

Remark. We initially prove this theorem for undirected Cayley graphs. In §5.5.6, we explain how to adapt the proof from the undirected case to the directed case. Doing this, rather than making every statement apply for both the un- and directed cases, significantly increases the readability. In particular, when we speak of $\mathbb{Z}$ we are referring to the set of all integers, positive and negative. △

Remark. We use the same methodology as §4.2. An outline of the proof is given in §4.2.2. △

5.5.3 Size of Ball Estimates and Lower Bound

In the lemmas below, used to prove this theorem, instead of writing one lemma with multiple parts, we split into separate lemmas according to q and k, eg $q \in (1,\infty)$ or $k \asymp \log n$; these parts are indexed with letters, eg Lemmas 5.5.2a, 5.5.2b and 5.5.2c.

We wish to determine the size of the L_q balls in $\mathbb{R}^k$. This is done by Lemmas 5.5.2 and 5.5.4; the statements are given below, with proofs are deferred to the supplementary material, §6.5.

For $q \in [1,\infty)$, write $V_{k,q}(R)$ for the (Lebesgue) volume of the L_q ball of radius R in $\mathbb{R}^k$, ie

$$V_{k,q}(R) \coloneqq \mathrm{vol}\big\{x \in \mathbb{R}^k \;\big|\; \|x\|_q \le R\big\};$$

also write $V_{k,q} \coloneqq V_{k,q}(1)$ and note that $V_{k,q}(R) = R^k V_{k,q}$. It is known (see [76]) that

$$V_{\ell,q} = 2^\ell \Gamma(1/q+1)^\ell / \Gamma(\ell/q+1). \tag{5.5.1}$$

We can use this, along with Lemma 5.5.2b below, to well-approximate $|B_{k,q}(R)|$ when $q \notin \{1,\infty\}$; for $q=1$ we directly bound $|B_{k,1}(\cdot)|$, while for $q=\infty$ we have an exact expression.

Lemma 5.5.2a. *For $q=1$ and all $R \ge 0$, we have*

$$2^k \tbinom{\lfloor R \rfloor}{k} \mathbf{1}(R \ge k) \le \big|B_{k,1}(R)\big| \le 2^k \tbinom{\lfloor R \rfloor + k}{k}. \tag{5.5.2a}$$

Lemma 5.5.2b. *For $q \in (1,\infty)$ and all $R \ge k^{1+1/q}$, we have*

$$\big|B_{k,q}(R)\big| = V_{k,q}(R)\,\big(1 + \mathcal{O}\big(k^{1+1/q}/R\big)\big). \tag{5.5.2b}$$

Lemma 5.5.2c. *For $q=\infty$ and all $R \ge 0$, we have*

$$\big|B_{k,\infty}(R)\big| = \big(2\lfloor R \rfloor + 1\big)^k. \tag{5.5.2c}$$

We use this lemma to find an M so that $|B_{k,q}(M)| \approx n$.

Definition 5.5.3. *Set $\omega \coloneqq \max\{(\log k)^2, k/n^{1/(2k)}\}$, and choose $M_{k,q}$ to be the minimal integer satisfying $|B_{k,q}(M_{k,q})| \ge n e^\omega$. Note that ω satisfies $1 \ll \omega \ll k$ if $k \ll \log n$.*

Recall that $\mathfrak{M}_{k,q} = k^{1/q} n^{1/k}/C_q$, and that $C_q = 2\,\Gamma(1/q+1)(qe)^{1/q}$. The next lemma shows that the difference between M and $\mathfrak{M}$ is only by sot. Also, let K be a constant, assumed to be as large as required, and let $\xi \coloneqq 1 - e^{-K\omega/k}$.

Lemma 5.5.4a. *For $k \ll \log n$ and $q=1$, we have*

$$M_{k,1} \le \big\lceil \mathfrak{M}_{k,1}(1+\xi) \big\rceil \quad \text{and} \quad \big|B_{k,1}\big(\mathfrak{M}_{k,1}(1-\xi)\big)\big| \ll n. \tag{5.5.3a}$$

Lemma 5.5.4b. *For $k \le \log n/\log\log n$ and all $q \in [1,\infty)$, we have*

$$M_{k,q} \le \big\lfloor \mathfrak{M}_{k,q}(1+\xi) \big\rfloor \quad \text{and} \quad \big|B_{k,q}\big(\mathfrak{M}_{k,q}(1-\xi)\big)\big| \ll n. \tag{5.5.3b}$$

Lemma 5.5.4c. *For $q=\infty$, we have*

$$M_{k,\infty} = \big\lceil \tfrac{1}{2} n^{1/k} e^{\omega/k} - \tfrac{1}{2} \big\rceil \quad \text{and} \quad \big|B_{k,\infty}\big(\mathfrak{M}_{k,\infty}(1-\xi)\big)\big| \ll n. \tag{5.5.3c}$$

Moreover, if $k \ll \log n$ then $M_{k,\infty} \asymp \mathfrak{M}_{k,\infty}$.

5.5.4 Lower Bound on Typical Distance

From this lemma, it is straightforward to deduce the lower bound in Theorem 5.5.1.

Proof of Lower Bound in Theorem 5.5.1. Observe that $|\mathcal{B}_{k,q}(M)| \le |B_{k,q}(M)|$. By Lemma 5.5.4, $|B_{k,q}(M)| = o(n)$ when $M \coloneqq M_{k,q}(1-\xi)$ when $k \ll \log n$. Thus $\mathcal{D}_{G_{k,q}}(\beta) \ge M$ for all β and Z. □

5.5.5 Upper Bound on Typical Distance

The outline of this subsection follows closely that of §4.2.5.

Proposition 5.5.5. *Let $q \in [1,\infty]$. Suppose that $k \ll \log n$. If $q \in (1,\infty)$, then further restrict to $k \le \log n / \log\log n$. Suppose also that $\limsup_n d/k < 1$. Then $\mathbb{E}(\|\mathbb{P}(W \cdot Z = \cdot \mid Z) - \pi_G\|_2^2) = o(1)$.*

Once we prove these propositions, we have all we need to prove Theorem 5.5.1.

Proof of Theorem 5.5.1 Given Lemma 5.5.4 and Proposition 5.5.5. Write $\mathcal{S}$ for the support of $A \cdot Z$. If $\|\mathbb{P}(A \cdot Z = \cdot \mid Z) - \pi_G\|_2 \le \varepsilon$, then $\pi_G(\mathcal{S}^c) \le \varepsilon$. Combining this with Lemma 5.5.4 and Proposition 5.5.5, we deduce the upper bound in Theorem 5.5.1. □

Remark. Proposition 5.5.5 actually holds even if $\eta := 1 - d/k \downarrow 0$, provided it does so sufficiently slowly and $k/\log n$ is sufficiently small. It turns out that $k/\log n \ll \eta$ and $\eta \gg 1/\sqrt{k}$ is sufficient; this allows k very close to both d and $\log n$. △

Let $W, W' \sim^{\text{iid}} \mathrm{Unif}(B_{k,q}(M))$, and let $V := W - W'$. Then we have

$$\mathbb{E}\big(\big\|\mathbb{P}_{G_k}\big(W \cdot Z \in \cdot\big) - \pi_G\big\|_2^2\big) = \mathbb{E}\big(n\,\mathbb{P}\big(V \cdot Z = 0 \mid Z\big) - 1\big) = n\,\mathbb{P}\big(V \cdot Z = 0\big) - 1.$$

First, it is immediate to see that $\mathbb{P}(W = W') = |B_{k,q}(M)|^{-1} \le n^{-1}e^{-\omega}$. Analogously in §5.1, the side-lengths $\{m_j\}_1^d$ satisfy $\min_j m_j > 2M$. Then we have

$$\mathcal{I} := \big\{i \in [k] \;\big|\; V_i \not\equiv 0 \bmod m_j \;\forall\, j = 1,\dots,d\big\} = \big\{i \in [k] \;\big|\; W_i \neq W_i'\big\}.$$

Lemma 5.5.6a. *For all k and all q, we have*

$$\mathbb{P}\big(\mathcal{I} = \emptyset\big) \le n^{-1}e^{-\omega}. \tag{5.5.4a}$$

Lemma 5.5.6b. *Suppose that $k \ll \log n$ and $q \in [1,\infty)$. If $q \in (1,\infty)$, then restrict further to $k \le \log n / \log\log n$. Then, for all $I \subseteq [k]$, we have*

$$\mathbb{P}\big(\mathcal{I} = I\big) \le e^{k(1/(eq)+\xi_q)} n^{-1+|I|/k} \tag{5.5.4b}$$

where $\xi_q := K_q\omega/k \ll 1$, for some constant K_q.

Lemma 5.5.6c. *For $q = \infty$, for all $I \subseteq [k]$, we have*

$$\mathbb{P}\big(\mathcal{I} = I\big) \le e^{-\omega(1-|I|/k)} n^{-1+|I|/k} \le n^{-1+|I|/k} \tag{5.5.4c}$$

While we have been stating results for undirected graphs, Lemma 5.5.6 holds in the directed case too. Contrastingly, the following lemma distinguishes between the directed and undirected graphs at one point. A proof of the lemma can be found in Lemma 4.2.9 in the main article. (There, while we studied both undirected and directed graphs, it was sufficient to use the worst-case bound for both; there we need the slightly more refined statement. The identical proof works.) Define

$$\mathfrak{g} := \gcd\big(V_1,\dots,V_k,n\big).$$

Lemma 5.5.7. *For all $I \subseteq [k]$, we have*

$$n\,\mathbb{P}\big(V \cdot Z = 0 \mid \mathcal{I} = I\big) \le \mathbb{E}\big(\mathfrak{g}^d \mid \mathcal{I} = I\big). \tag{5.5.5}$$

Further, there exists a constant C so that, for all $I \subseteq [k]$, we have

$$\mathbb{E}\big(\mathfrak{g}^d \mid \mathcal{I} = I\big) \le \begin{cases} C2^d(2M)^{d-|I|+2} & \text{when} \quad |I| \le d+1; & \text{(5.5.6a)} \\ 1 + 3\cdot 2^{d-|I|} & \text{when} \quad |I| \ge d+2 \quad \text{for undirected grahs,} & \text{(5.5.6b)} \\ 1 + 5\cdot (\frac{3}{2})^{2d-|I|} & \text{when} \quad |I| \ge d+2 \quad \text{for directed graphs.} & \text{(5.5.6c)} \end{cases}$$

The idea behind (5.5.5) is that linear combinations of independent uniform random variables are uniform on their support. Writing $G = \oplus_1^d \mathbb{Z}_{m_j}$, we obtain $V \cdot Z \sim \mathrm{Unif}(\oplus_1^d \mathfrak{g}_j \mathbb{Z}_{m_j})$ where $\mathfrak{g}_j := \gcd(V_1, ..., V_k, m_j) \le \mathfrak{g}$. For a rigorous argument, see Lemma 4.2.8 in the main article.

When $|I|$ is large, if $\mathfrak{g} > 1$ then we are asking that a large number of coordinates have a common divisor; naturally this decays exponentially in $|I|$. Using this decay, we can sum over all "large I".

Remark 5.5.8. We firmly believe that the stronger (5.5.6c) should hold for both the undirected and directed graphs (ie (5.5.6b) is unnecessary). It is merely a technical hurdle which is holding us back from proving this. When $q = \infty$, the coordinates of V are independent; in this case, we can prove that (5.5.6b) holds for both the undirected and directed graphs. As a result, we can relax $\limsup d/k < \frac{1}{2}$ to $\limsup d/k < 1$ for $q = \infty$. △

Corollary 5.5.9. *For any L with $L \ge d + 2$, we have*

$$n \sum_{|I| \ge L} \mathbb{P}\big(V \cdot Z \equiv 0,\, \mathcal{I} = I\big) \le \begin{cases} 1 + 3 \cdot 2^{d-L} & \text{for undirected graphs} \quad (5.5.7a) \\ 1 + 5 \cdot (\frac{3}{2})^{2d-L} & \text{for directed graphs} \quad (5.5.7b) \end{cases}$$

Proof. This proof is a direct calculation. By (5.5.5, 5.5.6b), using Bayes's rule, specifically the fact that $\mathbb{P}(B \mid C)/\mathbb{P}(C \mid B) = \mathbb{P}(B)/\mathbb{P}(C)$ for non-null events B and C, for $L \ge d + 2$ we deduce that

$$\begin{aligned} n \sum_{|I| \ge L} \mathbb{P}\big(V \cdot Z \equiv 0,\, \mathcal{I} = I \mid \mathsf{typ}\big) &= n \sum_{|I| \ge L} \mathbb{P}\big(V \cdot Z \equiv 0 \mid \mathcal{I} = I,\, \mathsf{typ}\big) \mathbb{P}\big(\mathcal{I} = I \mid \mathsf{typ}\big) \\ &\le \sum_{|I| \ge L} \big(\mathbb{P}\big(\mathcal{I} = I \mid \mathsf{typ}\big) + 3 \cdot 2^{d-|I|}\, \mathbb{P}\big(\mathcal{I} = I\big)/\mathbb{P}\big(\mathsf{typ}\big)\big) \\ &\le \mathbb{P}\big(|\mathcal{I}| \ge L \mid \mathsf{typ}\big) + 3 \cdot 2^{d-L} \mathbb{P}\big(|\mathcal{I}| \ge L\big)/\mathbb{P}\big(\mathsf{typ}\big) \le 1 + 3 \cdot 2^{d-L}/\mathbb{P}\big(\mathsf{typ}\big) \end{aligned}$$

for undirected graphs. The case of directed graphs follows analogously. □

We first prove the results on $\mathbb{P}(\mathcal{I} = I)$. For a set $I \subseteq [k]$ and $W \in \mathbb{Z}^k$, write $W_I = (W_i)_{i \in I}$ and $W_{\setminus I} = W_{I^c}$. Recall that if $\mathcal{C} \subseteq \mathcal{C}'$ and $U \sim \mathrm{Unif}(\mathcal{C}')$, then $(U \mid U \in \mathcal{C}) \sim \mathrm{Unif}(\mathcal{C})$. Hence we have

$$\mathbb{P}\big(W_{\setminus I} = W'_{\setminus I}\big) = \frac{\mathbb{P}(W = W')}{\mathbb{P}(W_I = W'_I \mid W_{\setminus I} = W'_{\setminus I})} = \frac{|B_{k,q}(M)|^{-1}}{\mathbb{E}(|B_{|I|,q}(M - \|A_{\setminus I}\|_1)|^{-1})} \le \frac{|B_{|I|,q}(M)|}{|B_{k,q}(M)|}. \tag{5.5.8}$$

Write $\ell := |I|$. Recall that, by choice of M, we have $|B_{k,q}(M)| \ge ne^{\omega}$, and so

$$\mathbb{P}\big(W_{\setminus I} = W'_{\setminus I}\big) \le n^{-1} e^{-\omega} \big|B_{\ell,q}(M)\big|.$$

Proof of Lemma 5.5.6a. Recall the choice of $M_{k,q}$, from Definition 5.5.3. Then (5.5.4a) follows:

$$\mathbb{P}\big(\mathcal{I} = \emptyset\big) = \mathbb{P}\big(W = W'\big) = \big|B_{k,q}(M_{k,q})\big|^{-1} \le n^{-1} e^{-\omega}. \qquad \square$$

Proof of Lemma 5.5.6b. Consider first $q = 1$. From Lemma 5.5.4a, recall that $M_1 \le (2e)^{-1} k n^{1/k} e^{\xi}$ with $\xi \asymp \omega/k$. Using Lemma 5.5.2a, for $\ell \le k$, we have

$$\big|B_{\ell,1}(M_1)\big| \le 2^{\ell} \tbinom{M_1 + \ell}{\ell} \le \big(2e(M_1/\ell + 1)\big)^{\ell} \le e^{\xi\ell} (k/\ell)^{\ell} n^{\ell/k} \le e^{k(1/e + \xi)} n^{\ell/k},$$

using the fact that $\binom{N}{\ell} \le (eN/\ell)^{\ell}$, that $\ell \mapsto (k/\ell)^{\ell}$ is maximised by $\ell = k/e$ and that $1 + x \le e^x$. The proof is completed by noting that $\{\mathcal{I} = I\} \subseteq \{A_{\setminus I} = A'_{\setminus I}\}$, and applying (5.5.8).

Now consider $q \in (1, \infty)$. Justified by Lemma 5.5.2b and Lemma 5.5.4b, which shows that $M_{k,q} \gg k^{1+1/q}$ for all q, we replace this discrete ball by the continuous ball, and lose only a factor $1 + o(1)$; for readability, we do not carry this factor in subsequent formulae.

Using Stirling's formula and the upper for $M_{k,q}$ from Lemma 5.5.4b gives

$$V_{\ell,q}(M_{k,q}) \le V_{\ell,q} \cdot \big((1 + \xi) k^{1/q} n^{1/k} / C_q\big)^{\ell} \le q^{1/2} e^{K_q \omega} (k/\ell)^{\ell/q} n^{\ell/k}.$$

From this, similarly to in Lemma 5.5.6a, using (5.5.8), we deduce that

$$\mathbb{P}\big(W_{\setminus I} = W'_{\setminus I}\big) \le q^{1/2} e^{K_q \omega} (k/\ell)^{\ell/q} n^{-1+\ell/k} \le e^{k(1/(eq) + \xi)} n^{-1+\ell/q},$$

where $\xi := K_q \omega/k \ll 1$, using again the fact that $\binom{N}{\ell} \le (eN/\ell)^{\ell}$ and that $\ell \mapsto (k/\ell)^{\ell}$ is maximised by $\ell = k/e$ The proof is completed by noting that $\{\mathcal{I} = I\} \subseteq \{W_{\setminus I} = W'_{\setminus I}\}$. □

Proof of Lemma 5.5.6c. The coordinates of W satisfy $W_i \sim^{\text{iid}} \text{Unif}(\{0, \pm 1, ..., \pm M_\infty\})$, for $i = 1, ..., k$. Write $\ell := |I|$. Hence, by (5.5.8) and (5.5.4a), we have

$$\mathbb{P}\big(W_{\setminus I} = W'_{\setminus I}\big) \le \big|B_{\ell,\infty}(M_\infty)\big| / \big|B_{k,\infty}(M_\infty)\big| = (2M_\infty + 1)^{\ell - k}.$$

By (5.5.3c), we have $2M_\infty + 1 \ge n^{1/k} e^{\omega/k}$. Hence

$$\mathbb{P}\big(\mathcal{I} = I\big) \le \mathbb{P}\big(A_{\setminus I} = A'_{\setminus I}\big) \le e^{\omega(-1+\ell/k)} n^{-1+\ell/k}. \qquad \square$$

We have now done all the hard work in proving Proposition 5.5.5, from which we deduced Theorem 5.5.1. It remains to go through the details of how to combine the previous results; there are no more interesting ideas to prove the propositions, but the details are quite technical.

Similarly to the mixing proof, we use an L_2 calculation:

$$\mathbb{E}\big(\big\|\mathbb{P}\big(W \cdot Z = \cdot \mid Z\big) - \pi_G\big\|_2^2\big) = n \textstyle\sum_{I \subseteq [k]} \mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) - 1. \tag{5.5.9}$$

Proof of Proposition 5.5.5 (when $q < \infty$). Recall that here $k \ll \log n$. For undirected graphs, $\limsup d/k < 1$; for directed, $\limsup d/k < \frac{1}{2}$. Set $\eta^- := 1 - \limsup d/k > 0$ and $\eta^+ := \frac{1}{2} - \limsup d/k$; set $L^- := d + \frac{1}{4}\eta^-(k-d)$ and $L^+ := 2d + \frac{1}{4}\eta^+(k-2d)$. Use L^- for undirected graphs and L^+ for directed. Then

$$\limsup L^\pm/k \le \tfrac{1}{4}\eta^\pm + (1 - \tfrac{1}{4}\eta^\pm)(1 - \eta^\pm) \le 1 - \tfrac{2}{3}\eta^\pm < 1;$$

also, $L^- - d \gg 1$ and $L^+ - 2d \gg 1$. Suppress the $\pm$-superscript: write $L := L^\pm$. By Lemma 5.5.4, recalling that $C_q = 2\,\Gamma(1/q+1)\,(qe)^{1/q}$, for some $\varepsilon_q = \mathcal{O}(\omega/k) = o(1)$, we can write

$$M = (1 + \varepsilon_q) k^{1/q} n^{1/k} / C_q.$$

It can be shown that $C_q \ge 2$ for all $q \in [1, \infty]$, and hence

$$2M \le e^{\varepsilon_q} k^{1/q} n^{1/k}. \tag{5.5.10}$$

Recall that when we consider $q = 1$, we only require $k \ll \log n$; when we consider $q \in (1, \infty)$, we require further that $k \le \log n / \log\log n$. Note that if $\mathcal{I} = \emptyset$ then $B = 0$, and so $B \cdot Z = 0$. Hence

$$n\,\mathbb{P}\big(B \cdot Z = 0 \mid \mathcal{I} = \emptyset\big) = n\,\mathbb{P}\big(\mathcal{I} = \emptyset\big) \le e^{-\omega},$$

by the choice of the radius $M_{k,q}$.

Consider $I \subseteq [k]$ with $1 \le \ell = |I| \le d + 1$. There are at most 2^k such sets I. Recall ξ_q given in Lemma 5.5.6b, and that $\xi_q = \mathcal{O}(\omega/k) = o(1)$. Applying (5.5.4b, 5.5.5, 5.5.6a, 5.5.10), we obtain

$$n\,\mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) \le C 2^d e^{k\varepsilon_q} e^{k\xi_q} k^{(d+2-\ell)/q} n^{(d+2-\ell)/k} \cdot e^{k/(eq)} n^{-1+\ell/k};$$

algebraic manipulations using the fact that $\limsup d/k < 1$ and $2^d = n^{o(1)}$ then give

$$n\,\mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) = 2^{-k} o(1). \tag{5.5.11}$$

Consider $I \subseteq [k]$ with $d + 2 \le \ell = |I| \le L = d$. Applying (5.5.4b, 5.5.5, 5.5.6b, 5.5.6c) gives

$$n\,\mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) \le 15 \cdot e^{k(1/(eq) + \xi_q)} n^{-1+\ell/k};$$

similar algebraic manipulations to those used when $1 \le |I| \le d + 1$ give

$$n\,\mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) = 2^{-k} o(1). \tag{5.5.12}$$

We now sum over all I with $1 \le |I| \le L$, using (5.5.11, 5.5.12):

$$n \textstyle\sum_{1 \le |I| \le L} \mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) = o(1). \tag{5.5.13}$$

Finally we consider $I \subseteq [k]$ with $L \le |I| \le k$. By Corollary 5.5.9, we have

$$n \sum_{L \le |I| \le k} \mathbb{P}\big(V \cdot Z \equiv 0,\, \mathcal{I} = I\big) \le \begin{cases} 1 + 3 \cdot 2^{d-L} = 1 + o(1) & \text{for undirected graphs,} \quad (5.5.14a) \\ 1 + 5 \cdot (\frac{3}{2})^{2d-L} = 1 + o(1) & \text{for directed graphs.} \quad (5.5.14b) \end{cases}$$

This last result actually holds for all $q \in [1, \infty]$ and all $1 \ll k \ll \log n$.

The proof is completed by combining (5.5.13, 5.5.14) with (5.5.9). $\square$

Proof of Proposition 5.5.5 (when $q = \infty$). Recall that $k \ll \log n$. As discussed in Remark 5.5.8, here (5.5.6c) holds for both the undirected and directed balls (ie (5.5.6b) is unnecessary); we only assume that $\limsup d/k < 1$ in either case. Set $\eta := 1 - \limsup d/k > 0$.

By (5.5.3c), we have $2M_{k,\infty} \le n^{1/k} e^{\omega/k} + 1$. Consider $I \subseteq [k]$ with $1 \le \ell = |I| \le d+1$. There are at most 2^k such sets I. Applying (5.5.3c, 5.5.4c, 5.5.5, 5.5.6a), we obtain

$$n\,\mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) \le C 2^d n^{(d-\ell+2)/k} e^{\omega(d+2-\ell)/k} (1 + e^{-\omega/k}/n^{1/k})^{d+2-\ell} \cdot e^{-\omega(1-\ell/k)} n^{-1+\ell/k};$$

algebraic manipulations using the fact that $\limsup d/k < 1$ and $2^d = n^{o(1)}$ then give

$$n\,\mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) = 2^{-k} o(1). \tag{5.5.15}$$

For $I \subseteq [k]$ with $d+2 \le \ell = |I| \le (1-\eta)k =: L$, applying (5.5.4c, 5.5.5, 5.5.6b) gives

$$n\,\mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) \le 15\, n^{-1+\ell/k} \le 2^{-k} n^{-1+L/k+o(1)}, \tag{5.5.16}$$

since $k \ll \log n$. We now sum over the I with $1 \le |I| \le L = (1-\eta)k$, using (5.5.15, 5.5.16):

$$n \textstyle\sum_{1 \le |I| \le L} \mathbb{P}\big(V \cdot Z = 0,\, \mathcal{I} = I\big) \le 2^k \cdot 2^{-k} n^{-1+L/k+o(1)} \le n^{-\eta+o(1)} = o(1). \tag{5.5.17}$$

Finally we consider $I \subseteq [k]$ with $L \le |I| \le k$. As above, by Corollary 5.5.9, we have

$$n \sum_{L \le |I| \le k} \mathbb{P}\big(V \cdot Z \equiv 0,\, \mathcal{I} = I\big) \le \begin{cases} 1 + 3 \cdot 2^{d-L} = 1 + o(1) & \text{for undirected graphs,} \quad (5.5.18a) \\ 1 + 5 \cdot (\frac{3}{2})^{2d-L} = 1 + o(1) & \text{for directed graphs.} \quad (5.5.18b) \end{cases}$$

The proof is completed by combining (5.5.17, 5.5.18) with (5.5.9). □

5.5.6 Adapting Proof to Directed Cayley Graphs

Where the random variable A was uniform on a certain undirected lattice ball, it is now uniform on a directed ball (of a different radius). Other than this, the only adaptation that needs be made is in determining the sizes of the discrete lattice balls: now instead of being a subset of $\mathbb{Z}^k$, for some k, they are restricted to the first quadrant, ie to $\mathbb{Z}_+^k$. Assuming that their radius is large enough, this simply reduces their size by a factor (roughly) 2^k.

Since all the sizes in question scale like R^k when the ball-radius is R, when $k \ll \log n$ (and so $R \gg 1$), the desired radius for the directed ball is twice that of the undirected ball. When $k \asymp \log n$ (and we consider the L_1 ball), the directed ball has size $\binom{R+k}{k}$, so we are still interested in $R \asymp k \asymp \log n$, just the constant is different for directed compared with directed.

Finally, for directed graphs, we have a slightly weakened bound on the expected gcd, ie $\mathbb{E}(\mathfrak{g}^d \mid \mathcal{I} = I)$; see Lemma 5.5.7. We addressed this in the proof of Proposition 5.5.5 at the time.

6 Supplementary Material

Abstract for Chapter 6

This document contains supplementary material for the previous chapters.

We prove very refined results about simple random walks on the integers and on the cycle. (See §6.1 and §6.2.) We are primarily interested in these random walks' entropy at certain times, and how this entropy changes when the time changes slightly. Additionally, we prove some large deviation and exit time estimates. (See §6.3 and §6.4.)

We prove some results on the size of discrete lattice balls, and how this size changes when the radius changes slightly, in a general L_q norm ($q \in [1, \infty]$). (See §6.5.)

We also prove some technical results deferred from the previous chapters. (See §6.6.)

We hope that some of the results, particularly the simple random walk estimates, will be useful in their own right for other researchers.

Table of Contents for Chapter 6

6.0 Notation and Terminology

While notation will often be recalled later, we list here the majority of what we use below.

- The *simple random walk*, abbreviated *SRW*, on $\mathbb{Z}^k$ (or $\mathbb{Z}_\gamma^k$) is the rate-1 RW which, in each step, chooses a coordinate uniformly at random and adds/subtracts 1 from this value (mod γ) with equal probability. The *directed random walk*, abbreviated *DRW*, on $\mathbb{Z}^k$ (or $\mathbb{Z}_\gamma^k$) is the same except that it only ever adds 1 (mod γ).

  When we wish to specify the DRW, we add a $+$-superscript; for the SRW, we add a $-$-superscript. When we do not wish to specify to which walk we are referring (as the statement applies for both), we simply speak of the *random walk*, abbreviated *RW*; if we wish to emphasise that the statement applies for both, then we sometimes add a $\pm$-superscript.

- Rate-1 RW on $\mathbb{Z}^k$:
  - $W = \big(W_i(t) \mid i \in [k],\, t \geq 0\big)$ is a rate-1 RW on $\mathbb{Z}^k$;
  - $\mu_t(\cdot) \coloneqq \mathcal{L}\big(W(t)\big)$, the *law*;
  - $Q(t) \coloneqq -\log \mu_t\big(W(t)\big)$, the *random (Shannon) entropy*;
  - $h(t) \coloneqq \mathbb{E}\big(Q(t)\big)$, the *(Shannon) entropy*.
- Rate-1 RW on $\mathbb{Z}$:
  - $W_i = (W_1(t) \mid i \in [k],\, t \geq 0)$ is a rate-$1/k$ RW on $\mathbb{Z}$;
  - $X = (X(s) = X_s \mid s \geq 0)$ defined by $X_s \coloneqq W_1(sk)$ is a rate-1 RW on $\mathbb{Z}$;
  - $\nu_s(\cdot) \coloneqq \mathcal{L}\big(W_1(sk)\big) = \mathcal{L}\big(X(s)\big)$, the *law*;
  - $Q_i(t) \coloneqq -\log \nu_{t/k}\big(W_i(t)\big)$, the *random (Shannon) entropy*;
  - $H(s) \coloneqq \mathbb{E}\big(Q_1(sk)\big) = \mathbb{E}\big(-\log \nu_s\big(X(s)\big)\big)$, the *(Shannon) entropy*.
- Rate-1 RW on $\mathbb{Z}_\gamma^k$:
  - $W_\gamma = \big(W_{\gamma,i}(t) \mid t \geq 0\big)$ defined by $W_{\gamma,i}(t) \coloneqq W_i(t) \bmod \gamma$ is a rate-1 RW on $\mathbb{Z}_\gamma$;
  - $\mu_{\gamma,t}(\cdot) \coloneqq \mathcal{L}\big(W_\gamma(t)\big)$, the *law*;
  - $Q_\gamma(t) \coloneqq -\log \mu_{\gamma,t}\big(W_\gamma(t)\big)$, the *random (Shannon) entropy*;
  - $h_\gamma(t) \coloneqq \mathbb{E}\big(Q_\gamma(t)\big)$, the *(Shannon) entropy*;
  - $r_\gamma(t) \coloneqq \log(\gamma^k) - h_\gamma(t)$, the *relative entropy* wrt $\mathrm{Unif}(\mathbb{Z}_\gamma^k)$.
- Rate-1 RW on $\mathbb{Z}_\gamma$:
  - $W_{\gamma,i} = \big(W_{\gamma,i}(t) \mid t \geq 0\big)$ is a rate-$1/k$ RW on $\mathbb{Z}_\gamma$;
  - $X_\gamma = \big(X_\gamma(s) \mid s \geq 0\big)$ defined by $X_\gamma(s) \coloneqq W_{\gamma,1}(sk)$ is a rate-1 RW on $\mathbb{Z}_\gamma$;
  - $\nu_{\gamma,s}(\cdot) \coloneqq \mathcal{L}\big(W_{\gamma,1}(sk)\big) = \mathcal{L}\big(X_\gamma(s)\big)$, the *law*;
  - $Q_{\gamma,i}(t) \coloneqq -\log \nu_{t/k}\big(W_{\gamma,i}(t)\big)$, the *random (Shannon) entropy*;
  - $H_\gamma(s) \coloneqq \mathbb{E}\big(Q_{\gamma,1}(sk)\big) = \mathbb{E}\big(-\log \nu_{\gamma,s}\big(X_\gamma(s)\big)\big)$, the *(Shannon) entropy*;
  - $R_\gamma(s) \coloneqq \log\gamma - H_\gamma(s)$, the *relative entropy* wrt $\mathrm{Unif}(\mathbb{Z}_\gamma)$.
- Consider "mod ∞" to mean no modulation; eg, $W = W_\infty$ or $H = H_\infty$.
- For $\gamma \in \mathbb{N}$ and $p \in [1, \infty]$, write

$$d_{p,\gamma}(s) \coloneqq \big\|\mathbb{P}\big(X_\gamma(s) \in \cdot\big) - \mathrm{Unif}(\mathbb{Z}_\gamma)\big\|_{p,\gamma} = \big(\textstyle\sum_{x \in \mathbb{Z}_\gamma} \frac{1}{\gamma}\big|n\mathbb{P}\big(X_\gamma(s) = x\big) - 1\big|^p\big)^{1/p};$$

  also write $d_{\mathrm{TV},\gamma}(s) \coloneqq \frac{1}{2} d_{\gamma,1}(s)$ for the total variation distance, which can be represented as

$$d_{\mathrm{TV},\gamma}(s) = \max_{A \subseteq \mathbb{Z}_\gamma} \big|\mathbb{P}\big(X_\gamma(s) \in A\big) - \tfrac{1}{\gamma}|A|\big|.$$

  For $\gamma = \infty$, we usually drop the γ-subscript.

6.1 Shannon Entropy Estimates and Central Limit Theorem

This part of the appendix (§6.1) is devoted to properties of the entropic time t_0 and cutoff window $t_\alpha - t_0$; this is done through analysis of a CLT for Q (Proposition 6.1.3) and variance of Q_1 at the entropic time, $\mathbb{V}\mathrm{ar}(Q_1(t_0))$. Accordingly, here we mainly derive properties of the SRW on $\mathbb{Z}$ evaluated at t/k or of Poisson(t/k), for t around the entropic time.

6.1.1 Key Definitions and Results for Shannon Entropy

We now define precisely the notion of *entropic times*. Let $W = (W(t))_{t\geq 0}$ be a RW on $\mathbb{Z}^k$. Write μ_t, respectively ν_s, for the law of $W(t)$, respectively $W_1(sk)$; so $\mu_t = \nu_{t/k}^{\otimes k}$. Define

$$Q_i(t) \coloneqq -\log \nu_{t/k}\big(W_i(t)\big), \quad \text{and set} \quad Q(t) \coloneqq -\log \mu_t\big(W(t)\big) = \textstyle\sum_1^k Q_i(t).$$

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

Definition 6.1.1 (Entropic and Times). *For all $k, n \in \mathbb{N}$ and all $\alpha \in \mathbb{R}$, define $t_\alpha \coloneqq t_\alpha(k,n)$ so that*

$$\mathbb{E}\big(Q_1(t_\alpha)\big) = \big(\log n + \alpha\sqrt{vk}\big)/k \quad \text{and} \quad s_\alpha \coloneqq t_\alpha/k, \quad \text{where} \quad v \coloneqq \mathbb{V}\mathrm{ar}\big(Q_1(t_0)\big).$$

We call t_0 the entropic time and the $\{t_\alpha\}_{\alpha\in\mathbb{R}}$ cutoff times.

The following proposition gives a detailed approximate evaluation of these entropic times.

Proposition 6.1.2 (Entropic and Cutoff Times). *Assume that $1 \ll \log k \ll \log n$. Write $\kappa \coloneqq k/\log n$. For all $\alpha \in \mathbb{R}$ and $\lambda > 0$, the following relations hold, for some functions f and g: we have $t_\alpha \approx t_0$;*

$$\text{for } k \ll \log n, \quad \text{we have} \quad t_0 \approx k \cdot n^{2/k}/(2\pi e) \quad \text{and} \quad t_\alpha - t_0 \approx \sqrt{2}\cdot \alpha t_0/\sqrt{k}; \tag{6.1.1a}$$

$$\text{for } k \approx \lambda \log n, \quad \text{we have} \quad t_0 \approx k \cdot f(\lambda) \quad \text{and} \quad t_\alpha - t_0 \approx g(\lambda)\cdot \alpha t_0/\sqrt{k}; \tag{6.1.1b}$$

$$\text{for } k \gg \log n, \quad \text{we have} \quad t_0 \approx k \cdot 1/(\kappa\log\kappa) \quad \text{and} \quad t_\alpha - t_0 \approx \sqrt{\kappa\log\kappa}\cdot \alpha t_0/\sqrt{k}. \tag{6.1.1c}$$

Moreover, $f, g : (0,\infty) \to (0,\infty)$ are continuous functions, whose value differs between the un- and directed cases. In particular, for all $\alpha \in \mathbb{R}$, in all cases, we have $t_\alpha \approx t_0$.

Observe that $Q(t) = \sum_1^k Q_i(t)$ is a sum of iid random variables.

Proposition 6.1.3 (CLT). *Assume that $1 \ll k \ll \log n$. For all $\alpha \in \mathbb{R}$, we have*

$$\mathbb{P}\big(Q(t_\alpha) \leq \log n \pm \omega\big) \to \Psi(\alpha) \quad \text{for} \quad \omega \coloneqq \mathbb{V}\mathrm{ar}\big(Q(t_0)\big)^{1/4} = (vk)^{1/4}.$$

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

6.1.2 Local CLT for RW on $\mathbb{Z}$

We repeatedly use a local CLT for Poisson and simple random walk distributions. We state it here precisely; the particular version is given in [48, Theorem 2.5.6].

Theorem 6.1.4 (Local CLT, [48, Theorem 2.5.6]). *Let $\varsigma > 0$ and let $s \in (\varsigma,\infty)$; the implicit constants in the $\mathcal{O}$-notation notation depend on ς. Let $X = (X_s)_{s\geq 0}$ be either a rate-1 SRW or rate-1 DRW on $\mathbb{Z}$. For all $x \in \mathbb{R}$ with $x - \mathbb{E}(X_s) \in \mathbb{Z}$ and $|x| \leq \frac{1}{2}s$, we have*

$$\mathbb{P}\big(X_s - \mathbb{E}(X_s) = x\big) = \frac{1}{\sqrt{2\pi s}}\exp\left(-\frac{x^2}{2s}\right)\exp\left(\mathcal{O}\left(\frac{1}{\sqrt{s}} + \frac{|x|^3}{s^2}\right)\right).$$

In particular, if $|x| \leq s^{7/12}$, then

$$\mathbb{P}\big(X_s - \mathbb{E}(X_s) = x\big) = \frac{1}{\sqrt{2\pi s}}\exp\left(-\frac{x^2}{2s}\right)\exp\big(\mathcal{O}\big(s^{-1/4}\big)\big). \tag{6.1.2}$$

Proof. The result for the SRW is given in [48, Theorem 2.5.6]. For the DRW, observe that $X_s \sim \text{Poisson}(s)$ and use Stirling's approximation. □

6.1.3 Derivation of CLT for Q

We first justify our CLT application in Proposition 6.1.3. The distribution of $Q_i(t_\alpha)$ depends on k (and n), and so we cannot apply the standard CLT. Instead, we apply a CLT for 'triangular arrays'; specifically, we now state a special case of the Lindeberg–Feller theorem.

Theorem 6.1.5 (CLT for Triangular Arrays; cf [35, Theorem 3.4.5]). *For each $k \in \mathbb{N}$, let $\{Y_{i,k}\}_{i=1}^k$ be an iid sequence of centralised, normalised random variables, and suppose that $\mathbb{E}(Y_{1,k}^4) \ll k$. Then*

$$\textstyle\sum_{i=1}^k Y_{i,k}/\sqrt{k} \to^d N(0,1) \quad \text{as } k \to \infty,$$

where $N(0,1)$ is a standard normal.

This version can be deduced easily from the version given in Durrett [35, Theorem 3.4.5]. Indeed, apply [35, Theorem 3.4.5] to the iid triangular array defined by $X_{i,k} := Y_{i,k}/\sqrt{k}$. Note that

$$\begin{aligned}\big(\textstyle\sum_1^k \mathbb{E}\big(X_{i,k}^2 \mathbf{1}(|X_{i,k}| \geq \varepsilon)\big)\big)^2 &= \mathbb{E}\big(Y_{1,k}^2 \mathbf{1}(|Y_{1,k}| \geq \varepsilon\sqrt{k})\big)^2 \\ &\leq \mathbb{E}\big(Y_{1,k}^4\big)\mathbb{P}\big(|Y_{1,k}| \geq \varepsilon\sqrt{k}\big) \leq \mathbb{E}\big(Y_{1,k}^4\big)/(\varepsilon^2 k) \to 0.\end{aligned}$$

Using this CLT for triangular arrays, we can deduce a CLT for Q.

Proof of Proposition 6.1.3. For our application of Theorem 6.1.5, for each $\alpha \in \mathbb{R}$, we take

$$Y_{i,k} := Y_{i,k}(\alpha) := \frac{Q_i(t_\alpha) - \mathbb{E}(Q_i(t_\alpha))}{\sqrt{\mathbb{V}\text{ar}(Q_i(t_0))}}. \tag{6.1.3}$$

Observe that $\mathbb{E}(Y_{i,k}) = 0$ and $\mathbb{V}\text{ar}(Y_{i,k}) = \mathbb{E}(Y_{i,k}^2) = 1$. Assuming that $\mathbb{E}(Y_{i,k}^4) \ll k$, we deduce the following result: for any sequence $(\alpha_n)_{n\geq 1}$ which converges to α, we deduce that

$$\mathbb{P}\big(Q(t) - \mathbb{E}\big(Q(t)\big) \geq \alpha_n \sqrt{\mathbb{V}\text{ar}(Q(t))}\big) \to \Psi(\alpha). \tag{6.1.4}$$

(We are also using Slutsky's theorem to allow α_n to depend on n, and, of course, the fact that $k \to \infty$ as $n \to \infty$.) We also further rely on the following claim:

$$\text{if} \quad t \asymp t_0, \quad \text{then} \quad \mathbb{V}\text{ar}\big(Q_1(t)\big) \asymp \mathbb{V}\text{ar}\big(Q_1(t_0)\big); \quad \text{also} \quad \mathbb{V}\text{ar}\big(Q(t_0)\big) \gg 1. \tag{6.1.5}$$

We prove these two statements in this claim (independently of the proof of the CLT) in Corollary 6.1.7 in §6.1.4. Now recall (6.1.1), which says that $t_\alpha \asymp t_0$ for all $\alpha \in \mathbb{R}$. Taking

$$\alpha_n := -\alpha\sqrt{\mathbb{V}\text{ar}(Q(t_0))/\mathbb{V}\text{ar}(Q(t_\alpha))} \pm \omega/\sqrt{\mathbb{V}\text{ar}(Q(t_\alpha))} \quad \text{with} \quad \omega := \mathbb{V}\text{ar}\big(Q(t_0)\big)^{1/4} \gg 1,$$

applying (6.1.4, 6.1.5) along with the above recollections we obtain the desired result:

$$\mathbb{P}\big(Q(t_\alpha) \leq \log n \pm \omega\big) \to \Psi(\alpha).$$

It remains to verify that $\mathbb{E}(Y_{i,k}^4) \ll k$. Roughly, $|W_1(t)|$ is 'well-approximated' by the following:

$$\begin{array}{ll} \big|N\big(\mathbb{E}(W_1(t)),\, t/k\big)\big| & \text{when} \quad t/k \gg 1, \quad \text{ie} \quad k \ll \log n; \\ \text{Bernoulli}(t/k) & \text{when} \quad t/k \ll 1, \quad \text{ie} \quad k \gg \log n. \end{array}$$

In the interim regime $k \asymp \log n$, we have that W_1 behaves like an 'order 1' random variable, in the sense that its mean and variance are bounded away from both 0 and ∞. It will actually turn out that the normal approximation is sufficient in the $k \asymp \log n$ regime also. Below, we abbreviate $Q_1(t_\alpha)$ by Q_1, $W_1(t_\alpha)$ by W_1 and t_α by t.

Write $s := t/k$. We consider separately the cases $s \gtrsim 1$ and $s \ll 1$. When $s \gtrsim 1$, we have $t \gtrsim k \gg 1$; when considering $s \ll 1$, however, we shall only consider t with $1 \ll t \ll k$. We shall be interested in $t := t_\alpha \approx t_0$, and Proposition 6.1.2 says that $t_0 \gg 1$ in all regimes; hence we need only consider $t \gg 1$. Let $\delta > 0$ be some (arbitrarily) small number.

Consider first $s = t/k$ with $s \geq \delta$. In this regime, we approximate $W_1(t)$ by a $N(\mathbb{E}(W_1), s)$ distribution, where $s = t/k$. Let $Z \sim N(\mathbb{E}(W_1), s)$, and write f for its density function:

$$f(x) := (2\pi s)^{-1/2} \exp\bigl(-\tfrac{1}{2s}(x - \mathbb{E}(W_1))^2\bigr) \quad \text{for} \quad x \in \mathbb{R}. \tag{6.1.6}$$

Let R_1 be a real valued random variable defined so that

$$R_1 = -\log f(x) \quad \text{when} \quad W_1 = x. \tag{6.1.7}$$

Also write $G := W_1 + U$, where $U \sim \mathrm{Unif}[-\frac{1}{2}, \frac{1}{2})$ is independent of W_1; then G has density function

$$g(x) := \mathbb{P}\bigl(W_1 = [x]\bigr) \quad \text{for} \quad x \in \mathbb{R}, \tag{6.1.8}$$

where $[x] \in \mathbb{Z}$ is $x \in \mathbb{R}$ rounded to the nearest integer (rounding up when $x \in \mathbb{Z} + \frac{1}{2}$). We have

$$(a-b)^4 \leq 3^4\bigl((a-a')^4 + (a'-b')^4 + (b'-b)^4\bigr) \quad \text{for all} \quad a, a', b, b' \in \mathbb{R}.$$

Applying this inequality with $a = Q_1$, $a' = R_1$, $b = \mathbb{E}(Q_1)$ and $b' = \mathbb{E}(R_1)$, we obtain

$$\begin{aligned} 3^{-4}\mathbb{E}\bigl((Q_1 - \mathbb{E}(Q_1))^4\bigr) &\leq \mathbb{E}\bigl((Q_1 - R_1)^4\bigr) + \mathbb{E}\bigl((R_1 - \mathbb{E}(R_1))^4\bigr) + \mathbb{E}\bigl(R_1 - Q_1\bigr)^4 \\ &\leq \mathbb{E}\bigl((R_1 - \mathbb{E}(R_1))^4\bigr) + 2\,\mathbb{E}\bigl((Q_1 - R_1)^4\bigr), \end{aligned} \tag{6.1.9}$$

with the second inequality following from Jensen (or Cauchy–Schwarz twice). We study these terms separately. Approximately, the local CLT will say that the second term is small; up to an error term which we control with the local CLT, the first term we can calculate directly using properties of the normal distribution.

We consider first the first term of (6.1.9). In terms of an integral, it is given by

$$\mathbb{E}\bigl((R_1 - \mathbb{E}(R_1))^4\bigr) = \int_{\mathbb{R}} g(x)\bigl(-\log f(x) - \mathbb{E}(R_1)\bigr)^4\,dx.$$

The local CLT suggests that we can approximately replace the $g(x)$ factor by $f(x)$, at least for a large range of x. So let us first study

$$\int_{\mathbb{R}} f(x)\bigl(-\log f(x) - \mathbb{E}(R_1)\bigr)^4\,dx = \int_{\mathbb{R}} f(x + \mathbb{E}(W_1))\bigl(-\log f(x + \mathbb{E}(W_1)) - \mathbb{E}(R_1)\bigr)^4\,dx.$$

A direct calculation reveals, remarkably, that the last expression is independent of the mean of W_1—this is a property special to the family of normal distributions. Expanding the fourth power and using moments of $N(0,1)$, one finds that this equals $\frac{15}{4}$; the exact numerical value is unimportant.

Now, by the local CLT (6.1.2), we have

$$\begin{aligned} \int_{-s^{7/12}}^{s^{7/12}} g(x)\bigl(-\log f(x) - \mathbb{E}(R_1)\bigr)^4\,dx &= \bigl(1 + \mathcal{O}(s^{-1/4})\bigr) \int_{-s^{7/12}}^{s^{7/12}} f(x)\bigl(-\log f(x) - \mathbb{E}(R_1)\bigr)^4\,dx \\ &\leq \bigl(1 + \mathcal{O}(s^{-1/4})\bigr) \cdot \tfrac{15}{4}. \end{aligned}$$

Using bounds on the tail of the SRW and Poisson distribution, as given in Propositions 6.3.4 and 6.3.5, it is straightforward to see, in both the undirected and directed cases, that

$$\int_{\mathbb{R}\setminus[-s^{7/12}, s^{7/12}]} f(x)\bigl(-\log f(x) - \mathbb{E}(R_1)\bigr)^4\,dx = o\bigl(s^{-10}\bigr). \tag{6.1.10}$$

(In fact, it is easy to see that it is $\mathcal{O}(\exp(-cs^{1/6}))$ for some sufficiently small constant c.) Hence

$$\mathbb{E}\bigl((R_1 - \mathbb{E}(R_1))^4\bigr) = \tfrac{15}{4}\bigl(1 + \mathcal{O}(s^{-1/4})\bigr) = \tfrac{15}{4}\bigl(1 + o(1)\bigr). \tag{6.1.11}$$

We now turn to the second term of (6.1.9). In terms of an integral, it is given by

$$\mathbb{E}\bigl((Q_1 - R_1)^4\bigr) = \mathbb{E}\bigl(\bigl(\log f(W_1) - \log g(W_1)\bigr)^4\bigr) = \int_{\mathbb{R}} g(x) \log\bigl(f(x)/g(x)\bigr)^4\,dx.$$

Again by the local CLT (6.1.2), we have

$$\int_{-s^{7/12}}^{s^{7/12}} g(x)\log\big(f(x)/g(x)\big)^4\,dx = \mathcal{O}\big(s^{-1/4}\big)\int_{-s^{7/12}}^{s^{7/12}} g(x)\,dx \le \mathcal{O}\big(s^{-1/4}\big),$$

and a similar application of the tail bounds in Propositions 6.3.4 and 6.3.5 shows that

$$\int_{\mathbb{R}\setminus[-s^{7/12},s^{7/12}]} g(x)\log\big(f(x)/g(x)\big)^4\,dx = o\big(s^{-10}\big) = \mathcal{O}\big(s^{-1/4}\big). \tag{6.1.12}$$

Hence, combining (6.1.11, 6.1.12) into (6.1.9), we obtain

$$\mathbb{E}\big((Q_1-\mathbb{E}(Q_1))^4\big) \le \tfrac{15}{4}\cdot 3^4 + o(1) \le 1000.$$

Now consider $\mathbb{V}\mathrm{ar}(Q_1)$. The arguments used for Proposition 6.1.6 (in §6.1.4) below show that

$$\mathbb{E}\big(\big(Q_1(sk)-\mathbb{E}\big(Q_1(sk)\big)\big)^4\big) \lesssim 1 \quad \text{when} \quad s \gtrsim 1.$$

Since $s_0 = t_0/k \gtrsim 1$ in the regime $k \lesssim \log n$, we deduce that $\mathbb{E}(Y_{1,k}^4) \lesssim 1 \ll k$.

Consider now $s = t/k$ with $s \le \delta$ but $t \gg 1$. In this regime, we approximate the number of steps taken by Bernoulli(t/k). Indeed, we have

$$\mathbb{E}\big(W_1=0\big) = 1-s+\mathcal{O}(bs^2) \quad \text{and} \quad \mathbb{E}\big(|W_1|=1\big) = s+\mathcal{O}\big(s^2\big).$$

We also use the fact that, for both the undirected and directed cases, for $x \ge 0$ we have

$$\mathbb{P}\big(W_1=x\big) \ge \mathbb{P}\big(\mathrm{Poisson}(s)=x\big)\cdot 2^{-x} = 2^{-x}e^{-s}s^x/x! \ge (s^2/x)^x; \tag{6.1.13}$$

from this one deduces that $-\log\mathbb{P}(W_1=x) \le x\log(x/s^2) = x(x+2\log(1/s))$. We use this to show that the terms with $|x| \ge 2$ contribute subleading order to the expectation

$$\mathbb{E}\big(Q_1\big) = \textstyle\sum_x \mathbb{P}\big(W_1=x\big)\log 1/\mathbb{P}\big(W_1=x\big) = s\log(1/s)+\mathcal{O}(s).$$

Similarly, we can use (6.1.13) to ignore the terms with $|x| \ge 2$ in

$$\begin{aligned}\mathbb{E}\big(\big|Q_1-\mathbb{E}(Q_1)\big|^r\big) &= \textstyle\sum_x \mathbb{P}\big(W_1=x\big)\big|-\log\mathbb{P}\big(W_1=x\big)-s\log(1/s)+\mathcal{O}(s)\big|^r\\ &= s\log(1/s)^r\big(1+\mathcal{O}(s)\big),\end{aligned} \tag{6.1.14}$$

for any fixed $r \in \mathbb{N}$ with $r \ge 2$, say $r \in \{2,3,4\}$.

In particular, this says that $\mathbb{V}\mathrm{ar}(Q_1) \eqsim s\log(1/s)^2$, and so

$$\mathbb{E}\big(Y_{i,k}^4\big) \eqsim \big(s\log(1/s)^4\big)\big/\big(s\log(1/s)^2\big)^2 = 1/s = k/t \ll k,$$

with the final relation holding since while $s \ll 1$ we do have $t \gg 1$. □

We now have all that we need to get on and calculate the entropic time t_0 in the three regimes of k. However, in order to find the cutoff times t_α, we need to know what the variance of the terms in the sum $Q(t)$, ie $\mathbb{V}\mathrm{ar}(Q_1(t))$, is for $t \eqsim t_0$.

6.1.4 Variance of $Q_1(t)$

Recall that, for all $t \ge 0$, we have

$$Q(t) = -\log\mu(t) = -\textstyle\sum_{i=1}^k \log\nu_t\big(W_i(t)\big) = \sum_{i=1}^k Q_i(t),$$

and that the $Q_i(t)$-s are iid (for fixed t). We now determine what its variance is at the entropic time t_0, and how the variance changes around this time. Note that $\mathbb{V}\mathrm{ar}(Q(t)) = k\mathbb{V}\mathrm{ar}(Q_1(t))$.

Proposition 6.1.6. *In both the undirected and the directed case,*

$$\mathbb{V}\mathrm{ar}\big(Q_1(sk)\big) \eqsim \begin{cases} 1/2 & \text{as} \quad s\to\infty, & \text{(6.1.15a)}\\ s\log(1/s)^2 & \text{as} \quad s\to 0; & \text{(6.1.15b)}\end{cases}$$

furthermore, the map $s \mapsto \mathbb{V}\mathrm{ar}(Q_1(sk)) : [0,\infty) \to \mathbb{R}_+$ is continuous.

From this, it is easy to calculate the variance at the entropic time t_0. Note that knowledge of the variance *is not* required to calculate t_0. The reader should recall Proposition 6.1.2. (Knowledge of the variance is required in calculating t_α with $\alpha \neq 0$, but not with $\alpha = 0$.)

Corollary 6.1.7. *For all regimes of k, in both the undirected and directed case,*

$$\text{if} \quad t \approx t_0, \quad \text{then} \quad \mathbb{V}\mathrm{ar}\big(Q_1(t)\big) \approx \mathbb{V}\mathrm{ar}\big(Q_1(t_0)\big) \gg 1/k. \tag{6.1.16}$$

Moreover, for all $\lambda > 0$, we have

$$\mathbb{V}\mathrm{ar}\big(Q_1(t_0)\big) \approx \begin{cases} 1/2 & \text{when} \quad k \ll \log n, & \text{(6.1.17a)} \\ v(\lambda) & \text{when} \quad k \approx \lambda \log n, & \text{(6.1.17b)} \\ \log n \log(k/\log n)/k & \text{when} \quad k \gg \log n, & \text{(6.1.17c)} \end{cases}$$

where $v : (0,\infty) \to (0,\infty) : \lambda \mapsto \mathbb{V}\mathrm{ar}(Q_1(f(\lambda)k))$ is a continuous function whose value differs between the undirected and directed cases.

Proof of Corollary 6.1.7. The first claim is immediate from Proposition 6.1.6. The claim for $k \ll \log n$ also uses (6.1.1). For $k \gg \log n$, there is a small amount of work to do. Set $s_0 := t_0/k$, and so

$$s_0 = \frac{t_0}{k} \approx \frac{\log n/k}{\log(k/\log n)} = \frac{1}{\kappa \log \kappa} \quad \text{where} \quad \kappa := \frac{k}{\log n} \gg 1.$$

We then also have

$$\log(1/s_0) = \log\log\kappa + \log\kappa + o(1) \approx \log\kappa,$$

and hence

$$s_0 \log(1/s_0)^2 \approx (\log\kappa)^2/(\kappa\log\kappa) = \log\kappa/\kappa = \log n \log(k/\log n)/k.$$

Note that while this has $\mathbb{V}\mathrm{ar}(Q_1(t_0)) \ll 1$, it does have $\mathbb{V}\mathrm{ar}(Q(t_0)) = k\mathbb{V}\mathrm{ar}(Q_1(t_0)) \gg 1$.

Finally consider $k \approx \lambda \log n$. Each coordinate runs at rate $1/k$, so for all $s \in \mathbb{R}_+$ the map $s \mapsto \mathbb{V}\mathrm{ar}(Q_1(sk))$ is continuous. Hence given $C > 0$ there exists an M so that

$$1/M \leq \mathbb{V}\mathrm{ar}\big(Q_1(sk)\big) \leq M \quad \text{for all} \quad s \text{ with } 1/C \leq s \leq C.$$

By (6.1.1b), we have $s = t_0/k \to f(\lambda)$. Hence $\mathbb{V}\mathrm{ar}(Q_1(t_0)) \to v$ for some constant $v \in (0,\infty)$ depending only on λ. This v is not the same in the directed and undirected cases. □

Proof of Proposition 6.1.6. Consider first $s \to \infty$. This proof is similar to the $s \gtrsim 1$ case, in justifying the CLT application. In particular, if

$$g(x) := \mathbb{P}\big(W_1(sk) = [x]\big) \quad \text{and} \quad f(x) := (2\pi s)^{-1/2} \exp\big(-\tfrac{1}{2s}(x - \mathbb{E}\big(W_1(sk)\big))^2\big),$$

then the local CLT (6.1.2) says, for $s \gtrsim 1$, that

$$g(x) = f(x)\,\big(1 + \mathcal{O}(s^{-1/4})\big) \quad \text{for} \quad x \in \mathbb{R} \quad \text{with} \quad \big|x - \mathbb{E}\big(W_1(sk)\big)\big| \leq s^{7/12}.$$

Under the assumption that $W_1(sk)$ is actually distributed as $N(0,s)$, direct calculation as in the previous section shows that the variance is then $\frac{1}{2}$. Considering the same approximations as before, namely splitting the integration range into $|x - \mathbb{E}(W_1)| \leq s^{7/12}$ and $|x - \mathbb{E}(W_1)| > s^{7/12}$, and using the local CLT to argue that $\log(g(x)/f(x)) = \mathcal{O}(s^{-1/4})$ for x in the first range, we obtain

$$\mathbb{V}\mathrm{ar}\big(Q_1(sk)\big) = \tfrac{1}{2} + \mathcal{O}\big(s^{-1/4}\log s\big) \approx \tfrac{1}{2} \quad \text{when} \quad s \gg 1.$$

Consider next $s \to 0$. In the CLT justification in the case $s \gtrsim 1$, we showed that

$$\mathbb{E}\big(\big|Q_1(sk) - \mathbb{E}(Q_1(sk))\big|^r\big) = s\log(1/s)^r + \mathcal{O}\big(s^2\log(1/s)^r\big), \tag{6.1.14}$$

and in particular deduced that $\mathbb{V}\mathrm{ar}(Q_1(sk)) \approx s\log(1/s)^2$. This applies for $s \ll 1$ also.

The continuity of $s \mapsto \mathbb{V}\mathrm{ar}(Q_1(sk))$ follows from the dominated convergence theorem. □

6.1.5 Calculating the Entropic and Cutoff Times

In this section we calculate the entropic time t_0, and the cutoff times t_α. Recall that

$$h(t) = \mathbb{E}\big(Q(t)\big) \quad \text{and} \quad H(s) \coloneqq \mathbb{E}\big(Q_1(sk)\big);$$

note that $H(s)$ is the entropy of $W_1(sk)$, which forms a rate-1 RW on $\mathbb{Z}$. The primary purpose of this section is to prove Proposition 6.1.2, which the reader should recall. To prove this, we derive asymptotic expressions for the entropy of rate-1 RW on $\mathbb{Z}$. As a consequence of these, one can see by how much the entropy changes when t_0 is replaced by $(1+\xi)t_0$ for a (small) constant $\xi \in \mathbb{R}$.

Lemma 6.1.8. *Let t_0 and $t_{\pm 2\omega}$ be the times at which the entropy of rate-1 RW on $\mathbb{Z}^k$ obtains entropy $\log N$ and $\log N \pm 2\omega$, respectively. Then $t_{\pm 2\omega} \eqsim t_0$ if $\omega \ll \min\{k, \log N\}$.*

We give this proof straight away, quoting results which are proved in the upcoming subsections.

Proof of Lemma 6.1.8. We prove the claim for $t_{2\omega}$; the analysis for $t_{-2\omega}$ is identical.

For $s \geq 0$, write $H(s)$ for the entropy rate-1 RW on $\mathbb{Z}$ evaluated at time s. In Propositions 6.1.9 and 6.1.13 below, we establish the following relations:

$$H(s) = \begin{cases} \frac{1}{2}\log(2\pi e s) + \mathcal{O}(s^{-1/4}) & \text{when} \quad s \gg 1; \qquad (6.1.18a) \\ s\log(1/s) + \mathcal{O}(s) & \text{when} \quad s \ll 1; \qquad (6.1.18b) \end{cases}$$

when $s \asymp 1$, we use continuity of $\lambda \mapsto H(1/\lambda)$ (cf Proposition 6.1.12 below). Let $\delta > 0$; assume that $\delta \downarrow 0$, but more slowly than the error terms in (6.1.18). Recall that the entropic time $t_0 = s_0 k$ is defined so that $H(s_0) = \log N/k$. We need to choose δ and ω so that

$$H\big(s_0(1+\delta)\big) \geq \log N/k + 2\omega/k \quad \text{with} \quad \delta \ll 1 \text{ and } \omega \gg 1.$$

Note that the statement is monotone in ω: if it holds for some ω, then it holds for any $0 \leq \omega' \leq \omega$, since then $t_0 \leq t_{2\omega'} \leq t_{2\omega}$. Hence we may assume lower bounds on ω, if desired.

Regime $k \ll \log N$. By (3.1.3a), we have $s_0 \gg 1$. By (6.1.18a), we have

$$\begin{aligned} H\big(s_0(1+\delta)\big) &= \tfrac{1}{2}\log(2\pi e s_0) + \tfrac{1}{2}\log(1+\delta) + \mathcal{O}\big(s_0^{-1/4}\big) \\ &= \log N/k + \tfrac{1}{2}\delta + \mathcal{O}\big(\min\big\{\delta^2, s_0^{-1/4}\big\}\big). \end{aligned}$$

We take $\delta \coloneqq 5\omega/k$, and so need $s_0^{-1/4} \ll \omega/k \ll 1$. Hence $\omega \ll k$ suffices.

Regime $k \gg \log N$. By (3.1.3c, 3.1.3d), we have $s_0 \ll 1$. By (6.1.18b), we have

$$\begin{aligned} H\big(s_0(1+\delta)\big) &= (1+\delta)\cdot s_0\log(1/s_0) - s_0(1+\delta)\log(1+\delta) + \mathcal{O}(s_0) \\ &= \log N/k + \delta\log N/k + \mathcal{O}(s_0). \end{aligned}$$

We take $\delta \coloneqq 2\omega/\log N$, and so need $s_0 \ll \omega/\log N \ll 1$. Hence $\omega \ll \log N$ suffices.

Regime $k \asymp \log N$. By continuity and the strict increasing property of the entropy, all we require is that $\log N/k + 2\omega/k = (1+o(1))\log N/k$, and hence only require $\omega \ll \log N \asymp k$. □

6.1.5.1 Regime $k \ll \log n$

We first consider the regime $k \ll \log n$, which corresponds to $s_0 = t_0/k \gg 1$.

Proposition 6.1.9. *For $s \gtrsim 1$, the entropy H of a rate-1 SRW or DRW on $\mathbb{Z}$ satisfies*

$$H(s) = \tfrac{1}{2}\log(2\pi e s) + \mathcal{O}\big(s^{-1/4}\big). \qquad (6.1.19)$$

Proof. We consider both the directed and undirected cases together. Write $t \coloneqq sk$. Define f, R_1 and g as in (6.1.6, 6.1.7, 6.1.8), respectively. By (6.1.12), we have

$$\big|\mathbb{E}(Q_1) - \mathbb{E}(R_1)\big| \leq \mathbb{E}\big((Q_1 - R_1)^4\big)^{1/4} = o\big(s^{-5/2}\big) = \mathcal{O}\big(s^{-1/4}\big) \quad \text{when} \quad s \gtrsim 1.$$

Direct calculation with its pdf shows that the entropy of $N(0,s)$ is precisely $\frac{1}{2}\log(2\pi es)$. Using this along with a similar calculation as used for (6.1.11) gives

$$\mathbb{E}\big(R_1\big) = \big(1+\mathcal{O}(s^{-1/4})\big)\cdot\log(2\pi es).$$

Hence we obtain our desired expression, namely (6.1.19). □

We now calculate the derivative of this entropy.

Proposition 6.1.10. *For $s \gtrsim 1$, the entropy H of a rate-1 SRW or DRW on $\mathbb{Z}$ satisfies*

$$H'(s) = (2s)^{-1}\big(1+\mathcal{O}\big(s^{-10}\big)\big). \tag{6.1.20}$$

By the chain rule, for $t \gtrsim k$, the entropy h of rate-1 SRW or DRW on $\mathbb{Z}^k$ then satisfies

$$h'(t) = H'(t/k) = (2t/k)^{-1}\big(1+\mathcal{O}\big((t/k)^{-10}\big)\big).$$

Proof. Write $t \coloneqq sk$. Define f, R_1 and g as in (6.1.6, 6.1.7, 6.1.8) respectively. We have

$$H(s) = -\textstyle\sum_{x\in\mathbb{Z}} \mathbb{P}\big(X_s = x\big)\log\mathbb{P}\big(X_s = x\big).$$

Differentiating this with respect to t we obtain

$$H'(s) = -\textstyle\sum_{x\in\mathbb{Z}} \frac{d}{ds}\mathbb{P}\big(X_s = x\big)\big(\log\mathbb{P}\big(X_s = x\big)+1\big) = -\sum_{x\in\mathbb{Z}} \frac{d}{ds}\mathbb{P}\big(X_s = x\big)\cdot\log\mathbb{P}\big(X_s = x\big).$$

Consider first the SRW. Using the Kolmogorov backward equations for the SRW, we obtain

$$\tfrac{d}{ds}\mathbb{P}\big(X_s = x\big) = \tfrac{1}{2}\mathbb{P}\big(X_s = x+1\big) + \tfrac{1}{2}\mathbb{P}\big(X_s = x-1\big) - \mathbb{P}\big(X_s = x\big).$$

Recall that $\nu_s(x) \coloneqq \mathbb{P}(X_s = x)$; write $g_s(x) \coloneqq \nu_s([x])$. Since $\sum_{x\in\mathbb{Z}}\nu_s(x) = 1$, we obtain

$$\begin{aligned}
H'(s) &= \textstyle\sum_{x\in\mathbb{Z}}\big(\nu_s(x) - \frac{1}{2}(\nu_s(x+1)+\nu_s(x-1))\big)\log\nu_s(x)\\
&= \textstyle\int_{\mathbb{R}}\big(g_s(x) - \frac{1}{2}(g_s(x+1)+g_s(x-1))\big)\log g(x).\\
&= \textstyle\int_{\mathbb{R}}\big(g_s(x) - \frac{1}{2}(g_s(x+1)+g_s(x-1))\big)\log f_s(x)\,dx &\text{(6.1.21a)}\\
&\quad + \textstyle\int_{\mathbb{R}}\big(g_s(x) - \frac{1}{2}(g_s(x+1)+g_s(x-1))\big)\log\big(g_s(x)/f_s(x)\big)\,dx, &\text{(6.1.21b)}
\end{aligned}$$

where $f_s(x) \coloneqq (2\pi s)^{-1/2}\exp(-x^2/(2s))$. The same arguments as used for (6.1.10) show that the integral in (6.1.21b) is $o(s^{-10})$. Now consider the integral in (6.1.21a). Using a simple shift,

$$\textstyle\int_{\mathbb{R}} g_s(x+1)\log f_s(x)\,dx = \int_{\mathbb{R}} g_s(x)\log f_s(x)\,dx - \int_{\mathbb{R}} g_s(x)\log\big(f_s(x-1)/f_s(x)\big)\,dx,$$

and we consider $\int_{\mathbb{R}} g_s(x-1)\log f_s(x)\,dx$ similarly; hence we have

$$\begin{aligned}
&\textstyle\int_{\mathbb{R}}\big(g_s(x) - \frac{1}{2}(g_s(x+1)+g_s(x-1))\big)\log f_s(x)\,dx\\
&\qquad = \tfrac{1}{2}\textstyle\int_{\mathbb{R}} g_s(x)\big(\log\big(f_s(x-1)/f_s(x)\big) + \log\big(f_s(x+1)/f_s(x)\big)\big)\,dx\\
&\qquad = \tfrac{1}{2}\textstyle\int_{\mathbb{R}} g_s(x)\log\big(f_s(x-1)f_s(x+1)/f_s(x)^2\big)\,dx.
\end{aligned}$$

Since $f_s(x) = (2\pi s)^{-1/2}\exp(-x^2/(2s))$, this log is precisely $1/s$ (independent of x). Since it is a distribution, g_s integrates to 1, so the integral equals $1/(2s)$. This proves the SRW case.

Now consider the DRW. Here the backward Kolmogorov equations read

$$\tfrac{d}{ds}\mathbb{P}\big(X_s = x\big) = \mathbb{P}\big(X_s = x-1\big) - \mathbb{P}\big(X_s = x\big) \quad \text{for} \quad x\in\mathbb{N}$$

and $\frac{d}{ds}\mathbb{P}(X_s = 0) = -\mathbb{P}(X_s = 0) = -e^{-s}$. Hence, as above, we have

$$\begin{aligned}
H'(s) - se^{-s} &= \textstyle\sum_{x\in\mathbb{N}}\big(\nu_s(x) - \nu_s(x-1)\big)\log\nu_s(x)\\
&= \textstyle\int_{1/2}^{\infty}\big(g_s(x) - g_s(x-1)\big)\log f_s(x)\,dx &\text{(6.1.22a)}\\
&\quad + \textstyle\int_{1/2}^{\infty}\big(g_s(x) - g_s(x-1)\big)\log\big(g_s(x)/f_s(x)\big)\,dx. &\text{(6.1.22b)}
\end{aligned}$$

As for (6.1.21b) above, the same arguments as used for (6.1.10) show that the integral in (6.1.22b) is $o(s^{-10})$. Note also that $se^{-s} = o(s^{-10})$ as $s \to \infty$. Now consider the integral in (6.1.22a). Using a simple shift as before, we have

$$\begin{aligned}\int_{1/2}^{\infty}\big(g_s(x) - g_s(x-1)\big)\log f_s(x)\,dx &= -\int_{1/2}^{\infty} g_s(x)\log\big(f_s(x+1)/f_s(x)\big)\,dx \\ &= \int_{1/2}^{\infty} g_s(x)\big((x-s)/s + 1/(2s)\big)\,dx = 1/(2s),\end{aligned}$$

recalling that here $f_s(x) = (2\pi s)^{-1/2}\exp(-(x-s)^2/(2s))$, $\mathbb{E}(X_s) = s$ and g_s integrates to 1. In the same way as for the SRW, this proves the DRW case. □

We wish to find the times $s_\alpha = t_\alpha/k$ defined so that, recalling (6.1.17a),

$$H(s_\alpha) = \big(\log n + \alpha\sqrt{vk}\big)/k \quad \text{where} \quad v := \mathbb{V}\mathrm{ar}\big(Q_1(t_0)\big) \approx \tfrac{1}{2}.$$

Proposition 6.1.11. *For $k \ll \log n$, we have*

$$s_0 = t_0/k \approx n^{2/k}/(2\pi e), \tag{6.1.1a}$$

and, for each $\alpha \in \mathbb{R}$, we have $s_\alpha \approx s_0$, and furthermore

$$(s_\alpha - s_0)/s_0 = (t_\alpha - t_0)/t_0 \approx \alpha\sqrt{2/k} = o(1). \tag{6.1.1a}$$

Proof. We consider the directed and undirected cases simultaneously. By directly manipulating (6.1.19), we see that if $H(s_0) = \log n/k$ then

$$s_0 = n^{2/k}/(2\pi e)\cdot\big(1 + \mathcal{O}\big(s_0^{-1/4}\big)\big) \approx n^{2/k}/(2\pi e),$$

noting that $k \ll \log n$ and so $n^{2/k} \gg 1$. This proves the first part of (6.1.1a).

We now turn to finding t_α. Fix $\alpha \in \mathbb{R}$. Note that H is increasing and $\alpha\sqrt{v/k} = o(1)$. So from the form of $H(s)$ given in (6.1.19) we see that, for all $\varepsilon > 0$, we have $(1-\varepsilon)s_0 \le s_\alpha \le (1+\varepsilon)s_0$ for n sufficiently large (depending on α); hence $s_\alpha \approx s_0$ for all $\alpha \in \mathbb{R}$.

By definition of s_α, we have

$$H(s_\alpha) - H(s_0) = \alpha\sqrt{v/k}, \quad \text{and hence} \quad \tfrac{ds_\alpha}{d\alpha}H'(s_\alpha) = \sqrt{v/k}.$$

Hence we have

$$s_\alpha - s_0 = \int_0^\alpha \tfrac{ds_a}{da}\,da = \sqrt{v/k}\int_0^\alpha 1/H'(s_a)\,da.$$

But, by Proposition 6.1.10, we may write $H'(s) = (2s)^{-1}(1+o(1))$ with $o(1)$ term uniform over $s \in [\frac{1}{2}s_0, 2s_0]$, which is an interval containing the cutoff window. Hence, recalling from (6.1.17a) that $v \approx \frac{1}{2}$ in this regime, the second part of (6.1.1a) follows:

$$s_\alpha - s_0 \approx 2\alpha s_0\sqrt{1/2}/\sqrt{k} \approx \alpha s_0\sqrt{2/k}.$$ □

6.1.5.2 Regime $k \asymp \log n$

We next consider the regime $k \approx \lambda\log n$ with $\lambda \in (0,\infty)$, which corresponds to $s_0 = t_0/k \asymp 1$.

Proposition 6.1.12. *There exists a decreasing, continuous bijection $f : (0,\infty) \to (0,\infty)$, whose value differs between the undirected and directed cases, so that, for all $\lambda > 0$, for $k \approx \lambda\log n$, we have*

$$s_0 = t_0/k \approx f(\lambda) \quad \text{where} \quad f(\lambda) := H^{-1}(1/\lambda), \tag{6.1.1b}$$

and, for each $\alpha \in \mathbb{R}$, we have $s_\alpha \approx s_0$, and furthermore

$$\begin{aligned}(s_\alpha - s_0)/s_0 = (t_\alpha - t_0)/t_0 \approx \alpha g(\lambda)/\sqrt{k} = o(1) \\ \text{where} \quad g(\lambda) := \sqrt{\mathbb{V}\mathrm{ar}(Q_1(f(\lambda)\,k))}/\big(f(\lambda)H'(f(\lambda))\big).\end{aligned} \tag{6.1.1b}$$

(Note that, for $s \in \mathbb{R}_+$, the law of $Q_1(sk)$ is independent of n and k, so g is a continuous function.)

Proof. Since $\log n/k \approx 1/\lambda$, we must choose s_0 so that $H(s_0) \approx 1/\lambda \in \mathbb{R}$. Since H is strictly increasing and continuous, we thus deduce that $t_0/k = s_0 \approx H^{-1}(1/\lambda) =: f(\lambda)$. So f is a decreasing, continuous bijection from $(0,\infty)$ to itself. This proves the first part of (6.1.1b).

We wish to find times s_α defined so that, recalling (6.1.17b),

$$H(s_\alpha) = \big(\log n + \alpha\sqrt{vk}\big)/k \quad \text{where} \quad v := \mathbb{V}\mathrm{ar}\big(Q_1(t_0)\big) \approx \mathbb{V}\mathrm{ar}\big(Q_1(f(\lambda)k)\big),$$

which is a constant whose value differs between the undirected and directed cases.

We now turn to finding s_α. Fix $\alpha \in \mathbb{R}$. Note that G is increasing and $\alpha\sqrt{v/k} = o(1)$. So from the continuity of H we see that, for all $\varepsilon > 0$, we have $(1-\varepsilon)s_0 \le s_\alpha \le (1+\varepsilon)s_0$ for n sufficiently large (depending on α); hence $s_\alpha \approx s_0$ for all $\alpha \in \mathbb{R}$.

Using the same arguments as in the previous derivative proof, we have

$$s_\alpha - s_0 = \textstyle\int_0^\alpha \frac{ds_a}{da}\, da = \sqrt{v/k}\int_0^\alpha 1/H'(s_a)\, da.$$

Using continuity of H' along with $s_\alpha \approx s_0 \approx f(\lambda)$, the second part of (6.1.1b) follows:

$$\begin{aligned} s_\alpha - s_0 &\approx \alpha\sqrt{v_*/k}/H'(s_0) \approx \alpha\sqrt{v_*/k}/H'(f(\lambda)) \\ &\approx \alpha s_0\big(\sqrt{\mathbb{V}\mathrm{ar}(Q_1(f(\lambda)\,k)}/\big(f(\lambda)H'(f(\lambda))\big)\big)/\sqrt{k}. \end{aligned}$$

□

6.1.5.3 Regime $k \gg \log n$

Finally we consider the regime $k \gg \log n$, which corresponds to $s_0 = t_0/k \ll 1$ but $t_0 \gg 1$. We have to handle the directed and undirected cases slightly differently here. The entropic time t_0 and cutoff times t_α will be the same (up to sot), but the technical details of the proofs will differ ever so slightly.

Proposition 6.1.13. *For $s \ll 1$, the entropy H of a rate-1 SRW or DRW on $\mathbb{Z}$ satisfies*

$$H(s) = s\log(1/s) + \mathcal{O}(s). \tag{6.1.23}$$

Proof. This follows immediately from (6.1.14) given in the justification of the CLT when $s \ll 1$. □

Proposition 6.1.14. *For $s \ll 1$, the entropy H of a rate-1 SRW or DRW on $\mathbb{Z}$ satisfies*

$$H'(s) = \log(1/s) + \mathcal{O}(1). \tag{6.1.24}$$

(For SRW, this $\mathcal{O}(1)$ is $\log 2 + \mathcal{O}(s)$; for DRW, it is $\mathcal{O}(s\log(1/s))$.)

Proof. We proceed as in the previous derivative proof, ie the proof of Proposition 6.1.10.

Consider first the undirected case. Using the Kolmogorov backward equations, we obtain

$$H'(s) = \textstyle\sum_{x\in\mathbb{Z}}\big(\nu_s(x) - \frac{1}{2}(\nu_s(x+1) + \nu_s(x-1))\big)\log\nu_s(x).$$

Recall that we have

$$\mathbb{P}\big(X_s = 0\big) = 1 - s + \mathcal{O}\big(s^2\big) \quad \text{and} \quad \mathbb{P}\big(X_s = x\big) = \tfrac{1}{2}s + \mathcal{O}\big(s^2\big) \text{ for } x \in \{\pm 1\},$$

and hence $\mathbb{P}(X_s = x) = \mathcal{O}(s^2)$ for $x \notin \{0, \pm 1\}$. Also, as previously, in the above sum we may ignore the x with $x \notin \{0,\pm1\}$ to give an error $\mathcal{O}(s\log(1/s))$. (Note that it is not $\mathcal{O}(s^2\log(1/s))$, since the x-th term of the sum contains $\nu_s(x+1)$ and $\nu_s(x-1)$.) Direct calculation then gives

$$H'(s) = \log(1/s) + \log 2 + \mathcal{O}(s) = \log(1/s) + \mathcal{O}(1).$$

This proves the undirected case.

Now consider the directed case. Here, $X_s \sim \text{Poisson}(s)$, and so $\mathbb{P}(X_s = x) = e^{-s}x^s/x!$. Then $W_1(t) \sim X_s$. Direct differentiation shows that

$$\tfrac{d}{ds}\mathbb{P}\big(X_s = x\big) = \mathbb{P}\big(X_s = x-1\big) - \mathbb{P}\big(X_s = x\big) = e^{-s}s^{x-1}(x-s)/x! \quad \text{for} \quad x \in \mathbb{N},$$

and $\frac{d}{ds}\mathbb{P}(X_s = 0) = -\mathbb{P}(X_s = 0) = -e^{-s}$, as in the previous derivative proof. As there, we have

$$H'(s) = -\sum_{x\in\mathbb{Z}_+} \tfrac{d}{ds}\mathbb{P}\big(X_s = x\big)\big(\log \mathbb{P}\big(X_s = x\big) + 1\big).$$

As previously, we may ignore the terms with $x \notin \{0, \pm 1\}$, giving an error $\mathcal{O}(s\log(1/s))$. Plugging in the derivative, we obtain

$$\begin{aligned} H'(s) &= -e^{-s}\log\big(e^{-s}\big) - e^{-s}(1-s)\log\big(se^{-s}\big) + \mathcal{O}\big(s\log(1/s)\big) \\ &= s\big(1 - s + \mathcal{O}(s^2)\big) - (1-s)\big(1 - s + \mathcal{O}(s^2)\big)\big(\log s - s\big) + \mathcal{O}\big(s\log(1/s)\big) \\ &= \log(1/s) + \mathcal{O}\big(s\log(1/s)\big) = \log(1/s) + \mathcal{O}(1). \end{aligned}$$

This proves the directed case. □

We wish to find the times $s_\alpha = t_\alpha/k$ defined so that, recalling (6.1.17c),

$$H(s_\alpha) = \big(\log n + \alpha\sqrt{vk}\big)/k \quad \text{where} \quad v := \mathbb{V}\mathrm{ar}\big(Q_1(t_0)\big) \eqsim (\log n/k)\log(k/\log n).$$

Proposition 6.1.15. *For $k \gg \log n$, we have*

$$s_0 = t_0/k \eqsim k^{-1}\log n/\log(k/\log n), \tag{6.1.1c}$$

and, for each $\alpha \in \mathbb{R}$, we have $s_\alpha \eqsim s_0$, and furthermore

$$(s_\alpha - s_0)/s_0 = (t_\alpha - t_0)/t_0 \eqsim \alpha\sqrt{\log(k/\log n)/\log n} = o(1). \tag{6.1.1c}$$

Proof. We consider the directed and undirected cases simultaneously. By directly manipulating (6.1.23), we see that if $H(s_0) = \log n/k$ then

$$s_0\log(1/s_0) \eqsim \log n/k. \quad \text{and hence} \quad \log(1/s_0) \eqsim \log(k/\log n),$$

with the final relation holding since $k \gg \log n$ and so $\log(k/\log n) \gg 1$; this implies that

$$s_0 = t_0/k \eqsim k^{-1}\log n/\log(k/\log n).$$

We now turn to finding s_α. Fix $\alpha \in \mathbb{R}$. From the form (6.1.23) of H, observe that

$$H\big(s_0(1\pm\varepsilon)\big) = (1\pm\varepsilon)H(s_0) + \mathcal{O}(s_0) = (1\pm\varepsilon)H(s_0)\cdot\big(1 + o(1)\big),$$

noting that $s_0 \ll 1$ and so $H(s_0) \eqsim s_0\log(1/s_0) \gg s_0$. Note also that

$$\sqrt{vk} \eqsim \sqrt{\log n\log(k/\log n)} \ll \log n,$$

since $\log k \ll \log n$. Hence $H(s_\alpha) = h(t_0)\cdot(1 + o(1))$. Hence, for all $\varepsilon > 0$, we have $(1-\varepsilon)s_0 \le s_\alpha \le (1+\varepsilon)s_0$ for n sufficiently large (depending on α); hence $s_\alpha \eqsim s_0$ for all $\alpha \in \mathbb{R}$.

As in the previous derivative proofs, we have

$$s_\alpha - s_0 = \textstyle\int_0^\alpha \frac{ds_a}{da}\,da = \sqrt{v/k}\int_0^\alpha 1/H'(s_a)\,da.$$

But, by Proposition 6.1.14, we may write $H'(s) = \log(1/s)(1 + o(1))$ with $o(1)$ term uniform over $t \in [\frac{1}{2}s_0, 2s_0]$, which is an interval containing the cutoff window. Hence, recalling the expressions for v from (6.1.15b) and s_0 from above, the second part of (6.1.1c) follows:

$$s_\alpha - s_0 \eqsim \alpha\sqrt{s_0/k} = \alpha s_0/\sqrt{s_0 k} \eqsim \alpha s_0/\sqrt{\log n/\log(k/\log n)}.$$

Note that $\log k \ll \log n$, and so $\log n/\log(k/\log n) \gg 1$. So we do indeed have $|s_\alpha - s_0| = o(s_0)$. □

Remark. In the directed case, we can actually find an explicit closed-form solution for the entropy:

$$H(s) = s\big(\log(1/s) + 1 + e^{-s}\textstyle\sum_{\ell=2}^\infty s^{\ell-1}\log(\ell!)/\ell!\big).$$

From this explicit expression, one can derive an approximation to the entropy when $s \gg 1$; see [38]. An analogous result for $s \ll 1$ is easy to obtain. For $s \asymp 1$, no simple closed form is known. △

6.2 Relative Entropy Estimates, Growth and Concentration

Let $X^+_\gamma := (X^+_\gamma(s))_{s\geq 0}$ be a DRW on $\mathbb{Z}_\gamma$ and $X^-_\gamma := (X^-_\gamma(s))_{s\geq 0}$ be a SRW on $\mathbb{Z}_\gamma$. Throughout this section, we use $+$-superscript to indicate DRW, eg $X^+_\gamma(s)$, and $-$-superscript to indicate SRW, eg $X^-_\gamma(s)$; when the result holds for both the SRW and DRW, we use either $\pm$-superscript or none at all, eg $X^\pm_\gamma(s)$ or $X_\gamma(s)$. Write $\nu_{\gamma,s}(\cdot)$ for the law of $X_\gamma(s)$, adding $+/-$-superscripts as appropriate.

The aim of this section is to derive some estimates on relative entropy. In the first two subsections (§6.2.1 and §6.2.2), the results will be for general times s. In the final subsection (§6.2.3), we are interested in the behaviour of the relative entropy around the so-called *entropic times*; see §6.2.3 for the definition, namely Definition 6.2.8.

First, we prove some general estimates on the relative entropy for RW on $\mathbb{Z}_\gamma$. We then specialise to $s \gtrsim \gamma^2$ (in §6.2.1) and then to $s \ll \gamma^2$ (in §6.2.2).

Lemma 6.2.1. *There exists an absolute constant $c > 0$ so that, for all $\gamma \geq 2$ and all $s \geq c$, we have*

$$R_\gamma(s) \geq c \log(\gamma/\sqrt{s}). \tag{6.2.1}$$

Moreover, for all $p \geq 2$ and $s \geq 0$, we have

$$\tfrac{1}{2} e^{-2\gamma_2 s} \leq 2\, d_{\mathrm{TV},\gamma}(s)^2 \leq R_\gamma(s) \leq d_{\infty,\gamma}(2s) \leq \textstyle\sum_{\ell=2}^{\gamma} e^{-2\gamma_\ell s} \tag{6.2.2}$$

where $\gamma_\ell := 1 - \cos(2\pi(\ell-1)/\gamma)$ for $\ell \in [\gamma]$. In particular, the following hold:

$$\begin{aligned} R_\gamma(s) &\ll 1 \quad \text{if and only if} \quad s \gg \gamma^2; \\ R_\gamma(s) &\asymp 1 \quad \text{if and only if} \quad s \asymp \gamma^2; \\ R_\gamma(s) &\gg 1 \quad \text{if and only if} \quad s \ll \gamma^2. \end{aligned}$$

Proof. The first claim is an immediate consequence of [42, Proposition 4.1]; it applies for both SRW and DRW. In particular, in the notation of [42, Proposition 4.1], the set A is chosen to be an interval of width $2\sqrt{s}$ around the mode of the RW location.

For the lower bound, recall Pinsker's inequality, which says that

$$R_\gamma(s) \geq 2\, d_{\mathrm{TV},\gamma}(s)^2.$$

Recall the standard fact that, for any eigenvalue ψ of the transition matrix and $s \geq 0$, we have

$$d_{\mathrm{TV},\gamma}(s) \geq \tfrac{1}{2} e^{-s\Re(1-\psi)};$$

see [49, (12.15)] for the discrete-time analogue. Write $q := e^{-2\pi i/\gamma}$, where i is the imaginary unit (not an index). The eigenvalues of the transition matrix for DRW, respectively SRW, are given by

$$\big(\lambda^+_\ell := q^{\ell-1} \mid \ell \in [\gamma]\big), \quad \text{respectively} \quad \big(\lambda^-_\ell := \Re(q^{\ell-1}) = \Re(\lambda^+_\ell) \mid \ell \in [\gamma]\big).$$

Apply this with $\psi := \lambda^\pm_2$. As $\lambda^-_2 = \Re(\lambda^+_2)$, this proves the lower bound for both SRW and DRW.

We turn to the upper bounds. By Jensen's inequality, for two measures μ and π, we have

$$\begin{aligned} D(\mu \,\|\, \pi) &= \textstyle\int \mu(x) \log\big(\mu(x)/\pi(x)\big)\, dx \leq \log\big(\int \mu(x)^2/\pi(x)\, dx\big) \\ &= \log\big(1 + \textstyle\int \pi(x) |\mu(x)/\pi(x) - 1|^2\, dx\big) = \log\big(1 + \|\mu - \pi\|^2_{L^2(\pi)}\big). \end{aligned}$$

Applying this and using the inequality $\log(1+x) \leq x$ for $x > -1$, we obtain

$$R_\gamma(s) \leq \log\big(1 + d_{2,\gamma}(s)^2\big) \leq d_{2,\gamma}(s)^2.$$

For reversible chains, it is well-known that $d_{2,\gamma}(s)^2 = d_{\infty,\gamma}(2s)$. For the DRW, we have $d_{2,\gamma}(s)^2 \leq d_{\infty,\gamma}(2s)$. Indeed, by symmetry, the L_2 mixing profile for the DRW and its time reversal are identical. The claim then follows from L_2–L_∞ mixing time relations in [57, Appendix].

For the SRW, by transitivity, and since we are working in continuous time,

$$d^-_{\infty,\gamma}(2s) = \gamma\, \nu^-_{\gamma,2s}(0) - 1 = \mathrm{trace}(P^-_{\gamma,2s}) - 1 = \textstyle\sum_{\ell=2}^{\gamma} e^{-2\gamma_\ell s},$$

where $P^-_{\gamma,\cdot}$ is the transition kernel for rate-1 SRW on $\mathbb{Z}_\gamma$. (See [2, Lemma 3.20, (3.60)] for justification of the first equality.) This establishes the upper bound for SRW.

For the DRW, we use the spectral decomposition. Let $(f^+_\ell \mid \ell \in [\gamma])$ be the orthonormal eigenbasis corresponding to $(\lambda^+_\ell \mid \ell \in [\gamma])$. We have $f^+_\ell(x) \coloneqq \exp(-2\pi i(\ell-1)x/\gamma)$ for $x \in \mathbb{Z}_\gamma$. By the spectral decomposition, for all $s \geq 0$ and all $x, y \in \mathbb{Z}_\gamma$, we have

$$P^+_{\gamma,2s}(x,y) - \tfrac{1}{\gamma} = \tfrac{1}{\gamma}\textstyle\sum_{\ell=2}^{\gamma} f^+_\ell(x)\overline{f^+_\ell(y)}\exp\big(-2s(1-\lambda^+_\ell)\big) \leq \tfrac{1}{\gamma}\sum_{\ell=2}^{\gamma}\exp\big(-2s\big(1-\Re(\lambda^+_\ell)\big)\big),$$

where $P^+_{\gamma,\cdot}$ is the transition kernel for rate-1 DRW on $\mathbb{Z}_\gamma$ and we have used the fact that $|f^+_\ell(z)| = 1$ for all $z \in \mathbb{Z}_\gamma$. As $\Re(\lambda^+_\ell) = \lambda^-_\ell$ for all $\ell \in [\gamma]$, this establishes the upper bound for DRW. □

6.2.1 Estimates for $s \gtrsim \gamma^2$

This subsection is devoted to analysing the regime $s \gtrsim \gamma^2$. (Recall that centred RW is diffusive, and so γ^2 is the order of the mixing and maximal hitting time of the RW.)

Lemma 6.2.1 has the following simple, but extremely useful, corollary.

Corollary 6.2.2. *For all $\gamma \geq 2$, if $s \gtrsim \gamma^2$, then*

$$d_{\mathrm{TV},\gamma}(s)^2 \asymp R_\gamma(s) \asymp d_{\infty,\gamma}(2s) \asymp d_{\infty,\gamma}(s)^2 \asymp e^{-2\gamma_2 s}.$$

Proof. Note that $\gamma_2 = \gamma_m$. Hence from (6.2.2) we deduce that

$$\tfrac{1}{2}e^{-2\gamma_2 s} \leq R_\gamma(s) \leq e^{-2\gamma_2 s}\big(2 + \textstyle\sum_{\ell=3}^{\gamma-1} e^{-2(\gamma_\ell-\gamma_2)s}\big).$$

Since $s \gtrsim \gamma^2$ and $\gamma_\ell - \gamma_2 \gtrsim \min\{\ell, \gamma-\ell\}^2/\gamma^2$, the sum above is $\mathcal{O}(1)$. □

Lemma 6.2.3. *For all $c > 0$, there exists a constant $\sigma \in (0,\infty)$ so that, for all $s \geq c\gamma^2$, we have*

$$1/\big(1+\sigma\sqrt{R_\gamma(s)}\big) \leq \gamma \min_{x\in\mathbb{Z}_\gamma} \nu_{\gamma,s}(x) \leq \gamma \max_{x\in\mathbb{Z}_\gamma} \nu_{\gamma,s}(x) \leq 1 + \sigma\sqrt{R_\gamma(s)}.$$

Proof. By Lemma 6.2.1 and Corollary 6.2.2, there exists a constant $\sigma_+ \in (0,\infty)$ so that

$$d_{\infty,\gamma}(s) \leq \sigma_+\sqrt{R_\gamma(s)}, \quad \text{and hence} \quad p\max_x \nu_{\gamma,s}(x) \leq 1 + \sigma_+\sqrt{R_\gamma(s)}.$$

If $R_\gamma(s) \leq (2\sigma_+)^{-2}$, then $1 - \sigma_+\sqrt{R_\gamma(s)} \geq 1/(1+2\sigma_+\sqrt{R_\gamma(s)})$; the claim follows with $\sigma \coloneqq 2\sigma_+$.

It remains to prove the lower bound under the assumption that $R_\gamma(s) \geq (2\sigma_+)^{-2}$. It then suffices to show that $\min_x \nu_{\gamma,s}(x) \gtrsim 1/\gamma$. This follows from a relatively simple application of the local CLT (see Theorem 6.1.4), for either SRW or DRW, noting that $s \gtrsim L^2$. □

Corollary 6.2.4. *For all $c > 0$, there exists a constant $\sigma > 0$ so that, for all $s \geq c\gamma^2$, we have*

$$\mathbb{V}\mathrm{ar}\big(Q_{\gamma,1}(sk)\big) \leq \sigma^2 R_\gamma(s).$$

Proof. Abbreviate $\rho_x \coloneqq \nu_{\gamma,s}(x)\cdot\gamma$ for $x \in \mathbb{Z}_\gamma$. Since $Q_{\gamma,1}(sk)$ takes the value $-\log\nu_{\gamma,s}(x)$ with probability $\nu_{\gamma,s}(x)$ for each $x \in \mathbb{Z}_\gamma$, if we define the random variable Y to take the value $\log\rho_x$ with probability $\nu_{\gamma,s}(x)$ for each $x \in \mathbb{Z}_\gamma$, then

$$\mathbb{V}\mathrm{ar}\big(Q_{\gamma,1}(sk)\big) = \mathbb{V}\mathrm{ar}(Y) \leq \mathbb{E}\big(Y^2\big).$$

Applying Lemma 6.2.3, we deduce the corollary:

$$\mathbb{E}\big(Y^2\big) = \textstyle\sum_x \nu_{\gamma,s}(x)|\log\rho_x|^2 \leq \max_x|\log\rho_x|^2 \leq \log\big(1+\sigma R_L(s)^{1/2}\big)^2 \leq \sigma^2 R_L(s).$$ □

6.2.2 Estimates for $s \ll \gamma^2$

This subsection is devoted to analysing the regime $s \ll \gamma^2$; however, we only consider $s \geq \varsigma$, for some absolute constant ς. Many of the constants below will depend on the choice of ς; however, since ς should be thought of as fixed throughout this whole section, we do not restate this dependence.

Proposition 6.2.5. *Uniformly in all $\gamma \in \mathbb{N}$, we have*

$$\max_{s\in[r,c\gamma^2]} \left|H_\gamma(s) - \tfrac{1}{2}\log(2\pi e s)\right| = o_{r\to\infty}(1) + o_{c\to 0}(1).$$

Equivalently, uniformly in all $\gamma \in \mathbb{N}$, we have

$$\max_{s\in[r,c\gamma^2]} \left|R_\gamma(s) - \tfrac{1}{2}\log(\gamma^2/s) - \tfrac{1}{2}\log(2\pi e)\right| = o_{r\to\infty}(1) + o_{c\to 0}(1).$$

Proof. We can uniquely write $X_\infty(s) = \tilde{X}_\gamma(s) + \gamma M_\gamma(s) + m_s$ with the following definitions:

- $\tilde{X}_\gamma(\cdot)$ is the RW on $[-\frac{1}{2}\gamma, \frac{1}{2}\gamma]$ centred to have mean 0;
- $M_\gamma(s)$ indicates in which interval of width L the RW on $\mathbb{Z}$ lives;
- m_s is the mode of $X_\infty(s)$.

As $X_\infty(s)$ determines $(X_\gamma(s),\ M_\gamma(s))$ and vice versa, by standard properties of entropy, we have

$$\mathrm{Ent}\big(X_\gamma(s)\big) \leq \mathrm{Ent}\big(X_\gamma(s),\ M_\gamma(s)\big) = \mathrm{Ent}\big(X_\infty(s)\big) \leq \mathrm{Ent}\big(X_\gamma(s)\big) + \mathrm{Ent}\big(M_\gamma(s)\big).$$

The upper bound on $H_\gamma(s)$ now follows immediately from Proposition 6.1.9.

We now turn to the lower bound. Using large deviations estimates for the SRW and the Poisson distribution from Propositions 6.3.4 and 6.3.5, it is routine to show that

$$-\log \mathrm{Ent}\big(M_\gamma(s)\big) \asymp \gamma^2/s \geq 1/c. \qquad \square$$

The above proof actually quantifies the errors, in the way described below.

Corollary 6.2.6. *There exists a constant c so that, for all $\gamma \in \mathbb{N}$ and all $\varsigma \leq s \leq c\gamma^2$, we have*

$$0 \leq H_\infty(s) - H_\gamma(s) \leq e^{-c\gamma^2/s}/c.$$

Lemma 6.2.7. *There exist positive constants c and C so that, for all $\gamma \in \mathbb{N}$, if $\varsigma \leq s \leq c\gamma^2$, then*

$$\mathbb{V}\mathrm{ar}\big(Q_{\gamma,1}(sk)\big) \leq C.$$

Proof. We may assume that γ is larger than any constant which we desire, otherwise all random variables are order 1 and so the statement holds easily. For $\delta \in (0, \frac{1}{2})$, consider the set

$$A_\delta := \big\{x \in \mathbb{Z}_\gamma \;\big|\; \nu_{\gamma,s}(x) \geq \delta/\sqrt{s}\big\}; \quad \text{write} \quad B_\delta := \mathbb{Z}_\gamma \setminus A_\delta.$$

By Proposition 6.2.5 and the local CLT (Theorem 6.1.4), since $\varsigma \leq s \leq c\gamma^2$, we have

$$\mathbb{E}\big(Q_{\gamma,1}(sk)\big) = H_\gamma(s) = \tfrac{1}{2}\log s + \mathcal{O}(1) \quad \text{and} \quad \max_x \nu_{\gamma,s}(x) = \nu_{\gamma,s}(0) \gtrsim 1/\sqrt{s}.$$

From this and the definition of A_δ, we deduce the following relations:

$$\alpha := \textstyle\sum_{x\in A_\delta} \nu_{\gamma,s}(x)\big(\log(1/\nu_{\gamma,s}(x)) - \mathbb{E}\big(Q_{\gamma,1}(sk)\big)\big)^2 \lesssim 1/\delta;$$
$$\beta := \textstyle\sum_{x\in B_\delta} \nu_{\gamma,s}(x)\big(\log(1/\nu_{\gamma,s}(x)) - \mathbb{E}\big(Q_{\gamma,1}(sk)\big)\big)^2 \lesssim 1 + \sum_{x\in B_\delta} \nu_{\gamma,s}(x)\log\big(\sqrt{s}\nu_{\gamma,s}(x)\big)^2.$$

To analyse the sum over $x \in B_\delta$, note that $\delta \leq \frac{1}{2}$. Under this assumption,

$$\nu_{\gamma,s}(x) \leq s^{-1/2}e^{-\sqrt{u}} \quad \text{iff} \quad \log\big(\sqrt{s}\nu_{\gamma,s}(x)\big) \leq -\sqrt{u} \quad \text{iff} \quad \log\big(\sqrt{s}\nu_{\gamma,s}(x)\big)^2 \geq u.$$

From this, using a simple change of variables, taking $\delta := \exp(-\sqrt{10}) \in (0, \frac{1}{2})$, we find that

$$\begin{aligned}
\textstyle\sum_{x\in B_\delta} \nu_{\gamma,s}(x)\log\big(\sqrt{s}\nu_{\gamma,s}(x)\big)^2 &= \textstyle\int_0^\infty \mathbb{P}\big(\log\big(\sqrt{s}\nu_{\gamma,s}(X_s)\big)^2 > u \mid X_s \in B_\delta\big)\mathbb{P}\big(X_s \in B_\delta\big)du \\
&= \textstyle\int_0^\infty \mathbb{P}\big(\nu_{\gamma,s}(X_s) \leq s^{-1/2}\min\{e^{-\sqrt{u}}, \delta\}\big)du \\
&\leq \textstyle\int_{10}^\infty \nu_{\infty,s}\big(x \in \mathbb{Z} \mid \nu_{\infty,s}(x) \leq s^{-1/2}e^{-\sqrt{u}}\big)du + 10.
\end{aligned}$$

It is easy to verify that the last integral is bounded from above, uniformly in $s \geq \varsigma$.

The result now follows, since $\mathbb{V}\mathrm{ar}(Q_{\gamma,1}(sk)) = \alpha + \beta$, and $\delta = \exp(-\sqrt{10})$. $\square$

6.2.3 Variations Around the Entropic Time: General Abelian Groups

For rate-1 RW on $\mathbb{Z}_\gamma^k$, the entropy function is denoted $h_\gamma(\cdot)$. For rate-1 RW on $\mathbb{Z}_\gamma$, the Shannon, respectively relative, entropy function is denoted $H_\gamma(\cdot)$, respectively $R_\gamma(\cdot)$; recall that $h_\gamma(\cdot) = \log\gamma - R_\gamma(\cdot)$. The Shannon entropy functions are strictly increasing bijections with

$$h_\gamma : [0,\infty) \to \big[0, \log(\gamma^k)\big) = [0, k\log\gamma) \quad \text{and} \quad H_\gamma : [0,\infty) \to [0, \log\gamma).$$

6.2.3.1 Entropic Time Definitions and Preliminaries

We are primarily interested in a target entropy of $N \coloneqq \log|G/\gamma G|$, where G is an arbitrary Abelian group G. For certain γ, the time $H_\gamma^{-1}(\log|G/\gamma G|/k)$ may well be $o(1)$, but since $k \lesssim \log n$ the maximum over γ is at least order 1. In the definition below, we take a maximum with ς.

Definition 6.2.8. *For $\gamma, N \in \mathbb{N}$, the entropic time is defined by*

$$s_0(\gamma, N) \coloneqq H_\gamma^{-1}\big((\log N)/k\big) \quad \textit{and} \quad t_0(\gamma, N) \coloneqq s_0(\gamma, N)k = h_\gamma^{-1}(\log N).$$

For the special case $N \coloneqq |G/\gamma G|$, write $t_\gamma \coloneqq s_\gamma k$ and $t_* \coloneqq s_* k$ where

$$s_\gamma \coloneqq s_0(\gamma, |G/\gamma G|) \vee \varsigma \quad \textit{and} \quad s_* \coloneqq \max_{\gamma\in\mathbb{N}} s_\gamma.$$

For an Abelian group G, write $d(G)$ for the minimal size of a generating subset of G. Abbreviate

$$\zeta_\gamma \coloneqq \tfrac{1}{k}\big(k - d(G)\big)\log\gamma.$$

Lemma 6.2.9. *For all $\gamma \in \mathbb{N}$, we have $|\gamma G| \geq \gamma^{-d(G)}|G|$, and in particular $|G/\gamma G| \leq \gamma^{d(G)}$.*

Proof. Decompose G as $\oplus_1^d \mathbb{Z}_{m_j}$. Then γG can be decomposed as $\oplus_1^d \mathbb{Z}_{m_j/\gcd(\gamma,m_j)}$. Thus $|\gamma G| = \prod_1^d m_j/\gcd(\gamma, m_j) \geq \prod_1^d m_j/\gamma = |G|/\gamma^d$. The second part follows from Lagrange's theorem. □

Corollary 6.2.10. *For all $\gamma \geq 2$, we have*

$$R_\gamma\big(s_0(\gamma, |G/\gamma G|)\big) = \log\gamma - (\log|G/\gamma G|)/k \geq \tfrac{1}{k}\big(k - d(G)\big)\log\gamma = \zeta_\gamma.$$

We first determine the asymptotic behaviour of s_*. Afterwards, we determine the rate of growth of the entropy around the entropy time. For both investigations, the following is useful.

Use the usual functional inner product: $\langle f, g\rangle_\pi \coloneqq \sum_z f(z)g(z)$.

Definition 6.2.11. *For all transition matrices P and all functions f and g, define the Dirichlet form*

$$\mathcal{E}_P(f,g) \coloneqq \big\langle f, (1-P)g\big\rangle_\pi = \textstyle\sum_{x,y} f(x)\big(g(x) - g(y)\big)P(x,y)\pi(x).$$

For a transition matrix P, write P^* for its time reversal; then $P^\times \coloneqq \frac{1}{2}(P + P^*)$ is its additive symmetrisation. Observe that, for all functions f and g, we have the following:

$$\mathcal{E}_P(f,f) = \tfrac{1}{2}\textstyle\sum_{x,y}\big(f(x) - f(y)\big)^2 P(x,y)\pi(x) \quad \text{and} \quad \mathcal{E}_P(f,f) = \mathcal{E}_{P^*}(f,f) = \mathcal{E}_{P^\times}(f,f);$$

if P is reversible, then also $\mathcal{E}(f,g) = \mathcal{E}(g,f)$. We now define logarithmic-Sobolev constants.

Recall that we write $\mathrm{Ent}_\pi(g) \coloneqq E_\pi(g\log(g/E_\pi(g))) = D(g/E_\pi(g)\,\|\,\pi)E_\pi(g)$ for a function g.

Definition 6.2.12. *Define the usual, respectively modified, log-Sobolev constants by*

$$c_{\mathrm{LS},P} \coloneqq \inf_{f:f\neq 0} \frac{\mathcal{E}_P(f,f)}{\mathrm{Ent}_\pi(f^2)} \quad \textit{and} \quad c_{\mathrm{MLS},P} \coloneqq \inf_{f:f>0} \frac{\mathcal{E}_P(f,\log f)}{\mathrm{Ent}_\pi(f)}.$$

Observe that $c_{\mathrm{LS},P} = c_{\mathrm{LS},P^*} = c_{\mathrm{LS},P^\times}$, ie this is the same for the reversal and the symmetrisation.

Lemma 6.2.13. *For every irreducible transition matrix P, we have $2c_{\mathrm{LS},P} \leq c_{\mathrm{MLS},P}$.*

Proof. Using the inequality $\log c \geq 1-1/c$ for $c > 0$, it is straightforward to show that $\mathcal{E}(f, \log f) \geq 2\,\mathcal{E}(\sqrt{f}, \sqrt{f})$ for $f > 0$; see, eg, [57, Lemma 2.8]. From this and the definitions, the claim follows. □

Simple direct calculations establish the following lemma; see, eg, [57, Lemma 2.4].

Lemma 6.2.14 ([57, Lemma 2.4])**.** *Let Ω be a state space and let $s \geq 0$. Let P be an irreducible transition matrix with invariant distribution π and write $P_s := e^{s(P-I)}$ for its heat kernel. Let μ be a distribution on Ω and write $\mu_s := \mu P_s$, ie the law of the chain started from μ and run for time s. For $x \in \Omega$, write $h_s(x) := \mu_s(x)/\pi(x)$, ie the density with respect to π. Then*

$$\tfrac{d}{ds} D(\mu_s \,\|\, \pi) = \tfrac{d}{ds} \operatorname{Ent}_\pi(h_s) = -\mathcal{E}(h_s, \log h_s) \leq -c_{\mathrm{MLS},P} \operatorname{Ent}_\pi(h_s) = -c_{\mathrm{MLS},P} D(\mu_s \,\|\, \pi).$$

Corollary 6.2.15. *In the set-up of Lemma 6.2.14, we have*

$$D(\mu_s \,\|\, \pi) \leq D(\mu \,\|\, \pi) e^{-c_{\mathrm{MLS},P} s} \leq D(\mu \,\|\, \pi) e^{-2c_{\mathrm{LS},P} s}.$$

Proof. This follows immediately from Lemmas 6.2.13 and 6.2.14 and Gronwall's lemma. □

It remains to estimate the log-Sobolev constant for the random walks on $\mathbb{Z}_\gamma$—recall that this is the same for the SRW and DRW, as the SRW is the additive symmetrisation of the DRW.

Lemma 6.2.16. *For all $\gamma \in \mathbb{N}$, the log-Sobolev constants of the RW on $\mathbb{Z}_\gamma$ satisfy*

$$c_{\mathrm{MLS},\gamma} \geq 2c_{\mathrm{LS},\gamma} \gtrsim 1/\gamma^2.$$

Proof. In [22, Corollary 3.11], it is shown that the L_2 mixing time is bounded below by $1/(2c_{\mathrm{LS},P})$. For the SRW on $\mathbb{Z}_L$, the L_2 mixing time is well-known to be order γ^2. The claim follows. □

6.2.3.2 Asymptotic Evaluation of Entropic Time

The precise definitions of $s_0(\gamma, N)$ and $t_0(\gamma, N)$ Here we asymptotically evaluate the entropic time. We first determine its order in general; second we evaluated it up to smaller order terms when $1 \ll k \ll \log|G|$ and $k - d(G) \asymp k$.

Proposition 6.2.17a. *Let $d, n \in \mathbb{N}$. Suppose that $1 \ll k \lesssim \log n$ and $k - d \asymp k$. Then, with implicit constant uniform over all Abelian groups G with $|G| = n$ and $d(G) = d$, we have*

$$\max_{\gamma \in \mathbb{N}} t_0(\gamma, |G/\gamma G|) \asymp k|G|^{2/k}.$$

Proposition 6.2.17b. *Let $d, n \in \mathbb{N}$. Suppose that $1 \ll k \lesssim \log n$ and $k > d$. Then, with implicit constant uniform over all Abelian groups G with $|G| = n$ and $d(G) = d$, we have*

$$\max_{\gamma \in \mathbb{N}} t_0(\gamma, |G/\gamma G|) \lesssim k|G|^{2/k} \log k.$$

When $d \ll \log n$ and $k - d(G) \asymp k$, the entropic time $t_*(k, G)$ is asymptotically equivalent to $t_0(\infty, |G|)$, ie the time at which the entropy of the RW on $\mathbb{Z}^k$ reaches $\log|G|$. This entropic time $t_0(\infty, |G|)$ is evaluated, in different regimes, in Proposition 6.1.2—there it is denoted $t_0(k, |G|)$.

Proposition 6.2.18. *Let $d, n \in \mathbb{N}$. Suppose that $d \ll \log n$ and $k - d \asymp k$. Then, with implicit constant uniform over all Abelian groups G with $|G| = n$ and $d(G) = d$, we have*

$$\max_{\gamma \in \mathbb{N}} t_0(\gamma, |G/\gamma G|) \asymp t_0(\infty, |G|).$$

Proposition 6.2.19. *Suppose that $k \gg \log|G|$. Write $\rho := \log k / \log\log|G|$. Then*

$$\max_{\gamma \in \mathbb{N}} t_0(\gamma, |G/\gamma G|) \asymp \tfrac{\rho}{\rho - 1} \log_k |G|.$$

Proof of Proposition 6.2.17a. For the lower bound, take $\gamma := \infty$ and use Proposition 6.1.2

$$\tilde{s}_* := \max_{\gamma\in\mathbb{N}} s_0(\gamma, |G/\gamma G|) \geq s_0(\infty, |G|) \asymp |G|^{2/k}.$$

In particular, this says that the maximising γ satisfies $s_0(\gamma, |G/\gamma G|) \gtrsim |G|^{2/k} \gtrsim 1$.

We now turn to the upper bound. Assume that $k - d \geq \varepsilon k$ with $\varepsilon \in (0, 1)$. By Lemma 6.2.9, the entropy of a single-coordinate at $t := t_0(\gamma, |G/\gamma G|)$, which we denote x, satisfies

$$x = \tfrac{1}{k} \log |G/\gamma G| \leq \big(\tfrac{1}{k} \log n\big) \wedge \big(\tfrac{1}{k} \log(\gamma^d)\big) \leq \big(\tfrac{1}{k} \log n\big) \wedge \big((1-\varepsilon) \log \gamma\big).$$

The relative entropy, which we denote ξ, thus satisfies $\xi = \log \gamma - x \geq \varepsilon \log \gamma \gtrsim 1$. By Lemma 6.2.1 and Proposition 6.2.5, there exists a constant C so that if $\xi \geq C$ then $t/k \leq Ce^{2x} \leq Cn^{2/k}$. On the other hand, if $\xi \leq C$, then necessarily $\xi \asymp 1$ and $\varepsilon \log \gamma \leq C$, ie $\gamma \leq e^{C/\varepsilon}$. Lemma 6.2.1 then implies that $t/k \asymp \gamma^2 \leq e^{2C/\varepsilon} \leq e^{2C/\varepsilon} n^{2/k}$, with implicit constant uniform over the compact interval $[\varepsilon \log 2, C]$ in which ξ lies. Hence, in either case, we have $t \lesssim C' k n^{2/k}$. □

Proof of Proposition 6.2.17b. As in the proof of Proposition 6.2.17a, the relative entropy of a single coordinate at time $t := t_0(\gamma, |G/\gamma G|)$, which we denote ξ, satisfies

$$\xi = \log \gamma - \tfrac{1}{k} \log |G/\gamma G| \geq \log \gamma - \big(\tfrac{1}{k} \log n\big) \wedge \big(\tfrac{1}{k} \log(\gamma^d)\big) = \big(\tfrac{1}{k}(k-d) \log \gamma\big) \vee \big(\log \gamma - \tfrac{1}{k} \log n\big).$$

Consider first the case that $\xi = \frac{1}{k}(k-d) \log \gamma =: \zeta$, ie $\gamma \leq n^{1/d}$. This is precisely the entropic time studied in §6.2.4 below, specifically Definition 6.2.24 and Proposition 6.2.25a with $p := \gamma$ and $\alpha := 0$ (in the notation there). These references show that $t/k \lesssim (\gamma^d)^{2/k} |\log(\zeta \wedge \frac{1}{2})|$. Since here we are considering $L \leq n^{1/d}$, the proof is completed in this case.

Now suppose that $\xi = \log \gamma - \frac{1}{k} \log n$, ie $L \geq n^{1/d}$. Then $\xi \geq k^{-2}(k-d) \log n \geq 1/k^2$. By Lemma 6.2.1 and Proposition 6.2.5, there exists a constant C so that if $\log \gamma - \frac{1}{k} \log n \geq C$ then $t/k \leq Cn^{2/k}$ as in the proof of Proposition 6.2.17a. On the other hand, if $\log \gamma - \frac{1}{k} \log n \leq C$, then necessarily $\xi \lesssim 1$ and $n^{1/d} \leq \gamma \lesssim n^{1/k}$. Corollary 6.2.2 implies that $R_\gamma(C' \gamma^2 \log k) \leq 1/k^2 \leq \xi$ for a large enough constant C'. Since $C' \gamma^2 \log k \lesssim n^{2/k} \log k$ the proof is completed in this case. □

Proof of Proposition 6.2.18. For the lower bound, take $\gamma := \infty$ and use Proposition 6.1.2:

$$\tilde{s}_* := \max_{\gamma\in\mathbb{N}} s_0(\gamma, |G/\gamma G|) \geq s_0(\infty, n).$$

When $k \ll \log n$, we have $s_0(\infty, n) \approx \frac{1}{2\pi e} n^{2/k} \gg 1$ while $s_0(\infty, n) \asymp 1$ when $k \asymp \log n$.

We now turn to the upper bound. Choose γ to be an optimiser, ie with $\tilde{s}_* = s_0(\gamma, |G/\gamma G|)$. By Lemma 6.2.9, we have $\tilde{s}_* \leq s_0(\gamma, \gamma^d) =: \tilde{s}_\gamma$. We consider first the case that $1 \ll k \ll \log n$ with $k - d \asymp k$; so $\tilde{s}_* \gg 1$. If $k \gtrsim \log(\gamma^d)$, then $\tilde{s}_\gamma \asymp 1$. But $\tilde{s}_\gamma \gg 1$, so we must have $1 \ll k \ll \log(\gamma^d)$. But $k \geq d$, so we must have $\gamma \gg 1$. Hence $\zeta := \frac{1}{k}(k-d) \log \gamma \gg 1$. By Lemma 6.2.1 and Corollary 6.2.10, we thus have $R_\gamma(\tilde{s}_\gamma) \geq \zeta \gg 1$ and $\tilde{s}_\gamma \ll \gamma^2$. By Proposition 6.2.5,

$$H_\gamma(\tilde{s}_\gamma) = \tfrac{1}{2} \log(2\pi e \tilde{s}_\gamma) + o(1).$$

The target entropy is $\frac{1}{k} \log |G/\gamma G| \leq \frac{1}{k} \log n$. Thus $\tilde{s}_\gamma \approx \frac{1}{2\pi e} n^{2/k}$, completing the upper bound.

Consider now $k \asymp \log n$ with $d \ll \log n$. Choose $m \gg 1$ such that $d \log m \ll \log n \asymp k$. Consider first $\gamma \leq m$. Since $\log(m^d)/k = d \log m/k \ll 1$, we have $\tilde{s}_\gamma \leq s_0(\gamma, m^d) \ll 1$. But $\tilde{s}_* \asymp 1$, so we must have $\gamma \geq m \gg 1$. Since $\tilde{s}_\gamma = \tilde{s}_* \asymp 1$, for $\gamma \geq m$, we have $\tilde{s}_\gamma \ll \gamma^2$. Thus we may apply Corollary 6.2.6 to deduce that $0 \leq H_\infty(\tilde{s}_*) - H_\gamma(\tilde{s}_*) = o(1)$. Further, $H_\infty(\tilde{s}_*) \asymp 1$. We thus deduce that the entropic times for the RW on $\mathbb{Z}^k$ and $\mathbb{Z}_\gamma^k$ are asymptotically equivalent. □

Proof of Proposition 6.2.19. We start with the lower bound. Clearly $t_0(\gamma, |G/\gamma G|) \geq t_0(\infty, |G|)$. By (6.1.1b) and some simple algebraic manipulations, we have $t_0(\infty, |G|) \approx \frac{\rho}{\rho - 1} \log_k |G|$.

We turn to the upper bound. Clearly $t_0(\gamma, |G/\gamma G|) \leq t_0(2, |G|)$ for all $\gamma \in \mathbb{N}$. In the regime $k \gg \log |G|$, in §6.1, to prove (6.1.1b), we approximated the rate-1 RW run for time $s \ll 1$ on $\mathbb{Z}$ by one on $\mathbb{Z}_2$. Thus the arguments for (6.1.1b) imply the upper bound here. □

6.2.3.3 Rate of Change of Entropy Around the Entropic Time

We now move onto determining the rate of growth of the entropy. The following lemma is valid for any $s \geq \varsigma$, but we are particularly interested in applying it at an entropic time s_γ. (This is one place in which we need the bound $s \geq \varsigma$, and so need to deal with s_γ, rather than $s_0(\gamma, |G/\gamma G|)$.)

Lemma 6.2.20. *There exists a continuous function $\tilde{c} : (0,1) \to (0,1)$ so that, for all $\gamma \geq 2$, all $\xi \in (-1,1) \setminus \{0\}$ and all $s \geq \varsigma$, we have*

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

Proof. If $s \asymp 1$, then the claim is immediate, noting that $R_\gamma(s) \asymp 1$. Now assume that $s \gg 1$.

Consider first the case where s/γ^2 is small; in particular, γ is large. By Proposition 6.2.5, there exists constants $\gamma_0 \in \mathbb{N}$ and $\alpha, c \in (0,\infty)$ so that, for all $\gamma \geq \gamma_0$ and all $s \in [\varsigma, 2\alpha\gamma^2]$, the difference in entropy is $\frac{1}{2}\log(1+\xi) + o(1)$. The claim thus follows in this case.

Now suppose that $s \geq \alpha\gamma^2$. By Corollary 6.2.15, for all $\gamma \geq 2$ and all $s \geq \alpha\gamma^2$, we have

$$H_\gamma\big(s(1+\varepsilon)\big) - H_\gamma(s) = R_\gamma(s) - R_\gamma\big(s(1+\varepsilon)\big) \geq (1 - e^{-2c_{\mathrm{LS},\gamma} s\varepsilon}) R_\gamma(s) \geq \delta_\varepsilon R_\gamma(s),$$

where $\delta_\varepsilon \coloneqq \liminf_\gamma \{1 - e^{-2\alpha\varepsilon c_{\mathrm{LS},\gamma}\gamma^2}\} \in (0,1)$ by Lemma 6.2.16. This completes the proof. □

Abbreviate $\rho_\gamma \coloneqq R_\gamma(s_\gamma)$. By Corollary 6.2.10, if $s_\gamma = s_0(\gamma, |G/\gamma G|)$, then $\rho_\gamma \geq \zeta_\gamma$.

Proposition 6.2.21. *Suppose that $k - d(G) \gg 1$ There exists a continuous function $c : (0,1) \to (0,1)$ so that, for all $\gamma \geq 2$ with $\gamma \wr |G|$ and all $\varepsilon \in (0,1)$, the following hold:*

$$\mathbb{P}\big(Q_\gamma\big(t_*(1+\varepsilon)\big) \leq \log|G/\gamma G| + c_\varepsilon(\zeta_\gamma \wedge 1)k\big) \leq \exp\big(-c_\varepsilon(\zeta_\gamma \wedge 1)k\big);$$
$$\mathbb{P}\big(Q_\gamma\big(t(1-\varepsilon)\big) \geq \log|G/\gamma G| - c_\varepsilon(\zeta_\gamma \wedge 1)k\big) = o(1) \quad \textit{for all} \quad t \leq t_0(\gamma, |G/\gamma G|).$$

The outline is of the proof is relatively straightforward. Replace s_γ with $s_0(\gamma, |G/\gamma G|)$. Consider $k - d \asymp k$, so that $\zeta_\gamma \asymp \zeta_\gamma \wedge 1 \asymp 1$. Both parts use the entropy growth rate lemma, Lemma 6.2.20. The non-quantitative part is then an application of Chebyshev's inequality, once one has shown that the variance $\mathbb{V}\mathrm{ar}(Q_{\gamma,1}(sk))$ is uniformly bounded over $s \geq \varsigma k$. The quantitative part requires a (one-sided) large deviations estimate given below in Theorem 6.2.22. We are not exactly sure who proved this originally; the earliest reference we found is in a survey by McDiarmid [56, Theorem 2.7]; we use the version given in the very nice survey paper by Chung and Lu [19, Theorem 3.4].

Theorem 6.2.22. *Let $(\xi_i)_{i=1}^k$ be a sequence of iid, mean-0 random variables with $\xi_1 \geq -M$ (deterministically), for some M. Set $\sigma^2 \coloneqq \mathbb{V}\mathrm{ar}(\xi_1) = \mathbb{E}(\xi_1^2)$. For all $x > 0$, we have*

$$\mathbb{P}\big(\textstyle\sum_1^k \xi_i \leq -x\big) \leq B(x, M, k\sigma^2) \quad \textit{where} \quad B(x, M, v^2) \coloneqq \exp\big(-\tfrac{1}{2}x^2/(v^2 + xM/3)\big).$$

Recall the definition of the random variable Q and the entropies h and H. Define

$$\xi_i \coloneqq Q_{\gamma,i}(t) - H_\gamma(t/k); \quad \text{in particular recall that} \quad \mathbb{E}\big(Q_{\gamma,i}(t)\big) = H_\gamma(t/k).$$

To apply the large deviations estimate to $\sum_1^k \xi_i$, we wish to find an $M \in \mathbb{R}$ so that $\xi_1 \geq -M$ deterministically. We also need to bound the variance. The following auxiliary lemmas do these.

Lemma 6.2.23a. *There exists an absolute constant M so that, for all $\gamma \geq 2$ and $s \geq \varsigma$, we have*

$$Q_{\gamma,1}(sk) - \mathbb{E}\big(Q_{\gamma,1}(sk)\big) \geq -M\big(\sqrt{R_\gamma(s)} \wedge 1\big).$$

Lemma 6.2.23b. *There exist an absolute constant σ^2 so that, for all $\gamma \geq 2$ and $s \geq \varsigma$, we have*

$$\mathbb{V}\mathrm{ar}\big(Q_{\gamma,1}(sk)\big) \leq \sigma^2\big(R_\gamma(s) \wedge 1\big).$$

We combine these lemmas to get our own large deviations estimate on $Q_\gamma(\cdot)$.

Proof of Proposition 6.2.21. Let $\varepsilon \in (0,1)$. We are interested at looking at time t_*, which satisfies

$$\max_{\gamma\in\mathbb{N}} t_\gamma = t_* = \max_{\gamma\in\mathbb{N}} t_0(\gamma, |G/\gamma G|).$$

Let $\gamma \in \mathbb{N}$ and set $t_0 \coloneqq t_0(\gamma, |G/\gamma G|)$ and $s_0 \coloneqq t_0/k$. Abbreviate $r_\gamma \coloneqq R_\gamma(s_0)$ and $\hat{r}_\gamma \coloneqq r_\gamma \wedge 1$. By Corollary 6.2.10, we have $r_\gamma \geq \zeta_\gamma$. By Lemma 6.2.20, there exists a constant $\tilde{c}_\varepsilon > 0$ so that

$$h_\gamma\big(t_0(1+\varepsilon)\big) - h_\gamma(t_0) \geq 2\tilde{c}_\varepsilon \hat{r}_\gamma k \geq 2\tilde{c}_\varepsilon \hat{\zeta}_\gamma k,$$

where $\zeta_\gamma \coloneqq \zeta_\gamma \wedge 1$. Recall that $\mathbb{E}(Q_\gamma(t')) = h(t')$ for all $t' \geq 0$. For each $i \in [k]$, set

$$\xi_i \coloneqq Q_{\gamma,i}\big(t_0(1+\varepsilon)\big) - \mathbb{E}\big(Q_{\gamma,i}\big(t_0(1+\varepsilon)\big)\big).$$

Altogether, these relations imply that

$$\big\{Q_\gamma\big(t_0(1+\varepsilon)\big) \leq \log|G/\gamma G| + \tilde{c}_\varepsilon\hat{\zeta}_\gamma k\big\} \subseteq \big\{\textstyle\sum_1^k \xi_i \leq -\tilde{c}_\varepsilon\hat{\zeta}_\gamma k\big\}.$$

Applying the large deviations estimate Theorem 6.2.22, with parameters controlled by Lemma 6.2.23, setting $c_\varepsilon \coloneqq \frac{1}{2}\tilde{c}_\varepsilon^2/(\sigma^2 + \frac{1}{3}\tilde{c}_\varepsilon M)$, a little algebra shows that

$$\mathbb{P}\big(Q_\gamma\big(t_0(1+\varepsilon)\big) \leq \log|G/\gamma G| + \tilde{c}_\varepsilon\hat{\zeta}_\gamma k\big) \leq \exp\big(-c_\varepsilon\hat{\zeta}_\gamma k\big).$$

We want to apply this for $\max_{\gamma\in\mathbb{N}} t_0(\gamma, |G/\gamma G|)$ instead of each individual $t_0(\gamma, |G/\gamma G|)$. This follows from the above analysis, due to the fact that $t' \mapsto Q_\gamma(t')$ is stochastically increasing.

For the lower bound, observe that the growth rate lemma Lemma 6.2.20 has the same form for time $1-\varepsilon$ as for $1+\varepsilon$. As for the upper bound, it suffices to prove the result for $t \coloneqq t_0(\gamma, |G/\gamma G|)$. We apply Chebyshev's inequality, with the variance controlled by Lemma 6.2.23b. The standard deviation is order $(\hat{\zeta}_\gamma k)^{1/2}$ and the displacement order $\hat{\zeta}_\gamma k$. So to deduce the result from Chebyshev's inequality, we need $\hat{\zeta}_\gamma k \gg 1$. This is immediate: $\hat{\zeta}_\gamma k = (k-d)\log\gamma \gg 1$ as $k - d(G) \gg 1$. □

It remains to prove Lemma 6.2.23, which has two parts.

Proof of Lemma 6.2.23a. Recall that $R_\gamma(s) \lesssim 1$ if $s \gtrsim \gamma^2$ and $R_\gamma(s) \gtrsim 1$ if $s \lesssim \gamma^2$. We have

$$\mathbb{E}\big(Q_{\gamma,1}(sk)\big) - Q_{\gamma,1}(sk) \leq \log\gamma + \log\big(\max_{x\in\mathbb{Z}_\gamma} \nu_{\gamma,s}(x)\big) = \log\big(1 + d_{\infty,\gamma}(s)\big) \leq d_{\infty,\gamma}(s).$$

By Lemma 6.2.1, writing $\gamma_\ell \coloneqq 1 - \cos(2\pi(\ell-1)/\gamma)$ for $\ell \in [\gamma]$, for both SRW and DRW, we have

$$d_{\infty,\gamma}(s) \leq \textstyle\sum_{\ell=2}^{\gamma} e^{-\gamma_\ell s} = 2e^{-\gamma_2 s}\big(1 + \frac{1}{2}\sum_{\ell=3}^{\gamma-1} e^{-(\gamma_\ell - \gamma_2)s}\big).$$

Since $\gamma_\ell - \gamma_2 \asymp \min\{\ell, \gamma-\ell\}^2/\gamma^2$, there exists a constant β so that, if $s \geq \beta\gamma^2$, then

$$\mathbb{E}\big(Q_{\gamma,1}(sk)\big) - Q_{\gamma,1}(sk) \leq d_{\infty,\gamma}(s) \leq 5\sqrt{R_\gamma(s)}.$$

On the other hand, if $\varsigma \leq s \leq \beta\gamma^2$, then we upper bound

$$\mathbb{E}\big(Q_{\gamma,1}(sk)\big) = H_\gamma(s) \leq H_\infty(s) = \tfrac{1}{2}\log(2\pi e s) + \mathcal{O}\big(s^{-1/4}\big),$$

with the last relation following from Proposition 6.1.9. By the local CLT (see, eg, [48, Theorem 2.5.6] or Theorem 6.1.4), the mode has probability order $1/\sqrt{s}$ in this regime. Hence

$$\mathbb{E}\big(Q_{\gamma,1}(sk)\big) - Q_{\gamma,1}(sk) \leq \mathcal{O}(1).$$ □

Proof of Lemma 6.2.23b. This is an immediate consequence of Corollary 6.2.4 and Lemma 6.2.7. □

Proof of Lemma 6.2.23b. Recall that $R_\gamma(s) \lesssim 1$ if $s \gtrsim \gamma^2$ and $R_\gamma(s) \gtrsim 1$ if $s \lesssim \gamma^2$.

By Lemma 6.2.7, there exist positive constants β and C so that, if $\varsigma \leq s \leq \beta\gamma^2$, then

$$\mathbb{V}\mathrm{ar}\big(Q_{\gamma,1}(sk)\big) \leq C.$$

On the other hand, if $s \geq \beta\gamma^2$, then by Corollary 6.2.4 we have

$$\mathbb{V}\mathrm{ar}\big(Q_{\gamma,1}(sk)\big) \lesssim R_\gamma(s).$$ □

6.2.4 Variations Around the Entropic Time: The Special Case of $\mathbb{Z}_p^d$

In this section we specialise to the group $\mathbb{Z}_p^d$; these entropic results do not require p to be prime. (Note that $d(\mathbb{Z}_p^d) = d$.) Here we not only establish cutoff, but also get a bound on the order of the cutoff window when $(k-d)p \gg 1$. This section has two main propositions.

Use the notation from the previous sections, but drop the γ-subscripts: we only consider RWs on $\mathbb{Z}_p$ or $\mathbb{Z}_p^k$. This is because the maximiser γ is clearly given by $\gamma = p$.

We first define precisely the entropic times under consideration here.

Definition 6.2.24. *Recall that $\zeta = \frac{1}{k}(k-d)\log p$. For $\alpha \in \mathbb{R}$, define*

$$t_\alpha := h^{-1}\big(d\log p + 2\alpha\sqrt{k(\zeta \wedge 1)}\big).$$

Equivalently, $t_\alpha := s_\alpha k$ where

$$\zeta_\alpha := \zeta - 2\alpha\sqrt{(\zeta\wedge 1)/k} = \zeta\big(1 - 2\alpha/\sqrt{\zeta k(\zeta \vee 1)}\big) \quad \text{and} \quad s_\alpha := R^{-1}(\zeta_\alpha).$$

We call t_0 the entropic time and $\{t_\alpha\}_{\alpha\in\mathbb{R}}$ cutoff times. Note that $\zeta_0 = \zeta$.

The next proposition estimates these entropic times.

Proposition 6.2.25a (Entropic Times). *Suppose that $1 \ll k \lesssim d\log p$. The following hold:*

$$\begin{aligned} &\text{if} \quad \zeta \ll 1, \quad \text{then} \quad t_0/k = s_0 \eqsim \tfrac{1}{2}\log(1/\zeta)/\big(1-\cos(2\pi/p)\big);\\ &\text{if} \quad \zeta \gtrsim 1, \quad \text{then} \quad t_0/k = s_0 \asymp p^2 e^{-2\zeta} = (p^d)^{2/k};\end{aligned}$$

further, if in fact $1 \ll k \ll d\log p$, then

$$\text{if} \quad \zeta \gg 1, \quad \text{then} \quad t_0/k = s_0 \eqsim p^2 e^{-2\zeta}/(2\pi e) = (p^d)^{2/k}/(2\pi e).$$

Note that $1-\cos(2\pi/p) \eqsim_{p\to\infty} 2\pi^2/p^2 = 2\pi^2 p^{-2d/k} e^{2\zeta}$.

Proposition 6.2.25b (Cutoff Times). *Suppose that $1 \ll k \lesssim d\log p$ and $(k-d)p \gg 1$, ie $\zeta \gg 1/k$. Then, for all $\alpha \in \mathbb{R}$, we have $t_\alpha \eqsim t_0$ and furthermore the following hold:*

$$\begin{aligned} &\text{if} \quad \zeta \lesssim 1, \quad \text{then} \quad (t_\alpha - t_0)/t_0 \lesssim 1/\big(\sqrt{\zeta k}\log((1/\zeta)\vee e)\big);\\ &\text{if} \quad \zeta \gg 1, \quad \text{then} \quad (t_\alpha - t_0)/t_0 \lesssim 1/\sqrt{k} \quad \text{for the SRW.}\end{aligned}$$

Remark 6.2.26. We strongly believe that the last result also holds for the DRW; see Remark 6.2.29 for justification of this belief. In short, $\zeta \gg 1$ implies that $s_0 \ll p^2$, and so the RW on $\mathbb{Z}_p$ should look almost the same as the RW on $\mathbb{Z}$, once recentred to have mean 0. In particular, the growth of the entropy (as a function of time) should be similar. △

The next result is a concentration result. For $\alpha \in \mathbb{R}$, define

$$Q_\alpha^+ := \{Q(t_\alpha) \ge \log n + \alpha\sqrt{k(\zeta\wedge 1)}\} \quad \text{and} \quad Q_\alpha^- := \{Q(t_{-\alpha}) \le \log n - \alpha\sqrt{k(\zeta\wedge 1)}\};$$

Proposition 6.2.27 (Concentration). *For all $\alpha \in (0,\infty)$ with $|\zeta_\alpha - \zeta_0| \le \frac{1}{2}\zeta_0$, we have*

$$\mathbb{P}\big((Q_\alpha^\pm)^c\big) \lesssim \alpha^{-2}.$$

This proposition is an easy consequence of relative entropy results proved earlier in §6.2.

Proof of Proposition 6.2.27. Using the definition of ζ_α and ζ, we have

$$h(t_\alpha) = \log n + 2\alpha\sqrt{k}(\zeta\wedge 1);$$

recall that $n = d\log p$. Note that $\mathbb{V}\mathrm{ar}(Q) = k\mathbb{V}\mathrm{ar}(Q_1)$. By Chebyshev's inequality, we have

$$\mathbb{P}\big(\big|Q(t_\alpha) - \big(\log n + 2\alpha\sqrt{k(\zeta\wedge 1)}\big)\big| \geq |\alpha|\sqrt{k(\zeta\wedge 1)}\big) \leq \alpha^{-2}\,\mathbb{V}\mathrm{ar}\big(Q_1(t_\alpha)\big)/(\zeta\wedge 1).$$

Consider $\zeta \lesssim 1$. Lemma 6.2.1 implies that $s_\alpha \gtrsim p^2$. Then, by Corollary 6.2.4, we have $\mathbb{V}\mathrm{ar}(Q(t_\alpha)) \lesssim \zeta_\alpha \asymp \zeta$. From this and the definition of $Q_\alpha^\pm$, the proposition follows. When $\zeta \gg 1$, Lemma 6.2.1 gives $s_\alpha \ll p^2$. The argument proceeds as before, replacing Corollary 6.2.4 with Lemma 6.2.7. □

We separate the proof of Proposition 6.2.25 into multiple parts.

Proof of 6.2.25a. Apply Lemma 6.2.1 and Corollary 6.2.2 together: for $\zeta \ll 1$, they imply that $s_0 \eqsim \frac{1}{2}\gamma_2^{-1}\log(1/\zeta)$; for $\zeta \asymp 1$, they imply that $s_0 \asymp \gamma_2^{-1}$. Also, note that $\gamma_2 = 1-\cos(2\pi/p) \eqsim_{p\to\infty} 2\pi^2/p^2$. For $\zeta \gg 1$, Proposition 6.2.5 implies that $s_0 \asymp n^{2/k} = p^{2d/k}$. Further, if $k \ll d\log p = \log n$, then $s_0 \gg 1$, and so in fact Proposition 6.2.5 implies that $s_0 \eqsim n^{2/k}/(2\pi e)$. □

We first show cutoff, namely $t_\alpha \eqsim t_0$ for all $\alpha \in \mathbb{R}$. This just uses the entropy growth rate.

Proof of 6.2.25b: Cutoff. Since $\zeta k \gg 1$, we have

$$R(s_\alpha) = \zeta_\alpha = \zeta\big(1+o(1)\big) = R(s_0)\big(1+o(1)\big).$$

But by Lemma 6.2.20, replacing s_0 by $s_0(1+\xi)$ changes the entropy by at least order $\zeta\wedge 1$. Case analysis gives $s_0(1-\varepsilon) \leq s_\alpha \leq s_0(1+\varepsilon)$ asymptotically for all $\alpha \in \mathbb{R}$ for all $\varepsilon \in (0,1)$. □

We next bound the window. Note that $(k-d)p \gg 1$ implies $\zeta k(\zeta\vee 1) \gg 1$.

Proof of 6.2.25b: Window when $\zeta \lesssim 1$. Recall Corollary 6.2.15 and Lemma 6.2.16:

$$R(u+v) \leq R(v)e^{-2c_{\mathrm{LS},p}u} \quad \text{for all} \quad u,v \geq 0 \quad \text{and} \quad c_{\mathrm{LS},p} \gtrsim 1/p^2.$$

Consider first $\alpha > 0$. Applying this with $v \coloneqq s_0$ and $u \coloneqq s_\alpha - s_0$ gives

$$\zeta_\alpha = R(s_\alpha) \leq e^{-c_{\mathrm{LS},p}(s_\alpha - s_0)}R(s_0) = e^{-c_{\mathrm{LS},p}(s_\alpha-s_0)}\zeta_0.$$

We hence deduce that $s_\alpha - s_0 \leq c_{\mathrm{LS},p}^{-1}\log(\zeta_0/\zeta_\alpha)$, ie

$$s_\alpha - s_0 \leq -c_{\mathrm{LS},p}^{-1}\log\big(1-\alpha/\sqrt{\zeta k(\zeta\vee 1)}\big) \asymp \alpha p^2/\sqrt{\zeta k(\zeta\vee 1)} \asymp \alpha p^2/\sqrt{\zeta k}.$$

For $\alpha < 0$, use $v \coloneqq s_\alpha$ and $u \coloneqq s_0 - s_\alpha$ to deduce the analogous result. Hence

$$|s_\alpha - s_0| \asymp |\alpha|p^2/\sqrt{\zeta k} \asymp |\alpha|s_0/\big(\sqrt{\zeta k}\log((1/\zeta)\vee e)\big).$$ □

To analyse the window, we need to use the derivative of the entropy.

Lemma 6.2.28. *There exist positive constants c and c' so that, for all $\gamma \in \mathbb{N}$, if $\varsigma \leq s \leq c\gamma^2$, then*

$$\tfrac{d}{ds}H_\gamma^-(s) = -\tfrac{d}{ds}R_\gamma^-(s) \geq c'/s.$$

Remark 6.2.29. This intuition for this claim is simple. For $s \ll \gamma^2$, the RW on $\mathbb{Z}_\gamma$ and $\mathbb{Z}$ look almost the same. This is quantified by Corollary 6.2.6. We showed the the RW on $\mathbb{Z}$ that the derivative satisfies $H'_\infty(s) \eqsim 1/(2s)$. We thus expect the RW on $\mathbb{Z}_\gamma$ to exhibit the same property when $s \ll \gamma^2$. However, due to a technical hurdle, we have only been able to show this for the SRW.

The claim is somewhat analogous to the usual log-Sobolev inequality (see Corollary 6.2.15):

$$R_\gamma(u+v) \leq R_\gamma(v)e^{-c_{\mathrm{LS},\gamma}v} \quad \text{for all} \quad u,v \geq 0.$$ △

Proof of 6.2.25b: Window when $\zeta \gg 1$ for SRW. In Proposition 6.1.11 we performed an analogous calculation. Exactly as there, using Lemma 6.2.28, we deduce that $|s_\alpha - s_0|/s_0 \lesssim |\alpha|/\sqrt{k}$. □

Finally, we prove Lemma 6.2.28.

Proof of Lemma 6.2.28. We may assume that s, and hence γ, is larger than any constant which we desire, otherwise all terms on the left-hand side are order 1 and so the statement holds easily.

There is a constant C sufficiently large so that, writing $K := C\sqrt{s}$, we have

$$\nu^-_{\gamma,s}([-K,K]) := \textstyle\sum_{x\in[-K,K]} \nu^-_{\gamma,s}(x) \geq \frac{8}{9}.$$

Moreover, provided c is sufficiently small, we can choose C so that $K \leq \gamma/10$. Define ξ^- by

$$\xi^-_x := \nu^-_{\gamma,s}(x)\mathbf{1}(x \in [-K,K])/\nu^-_{\gamma,s}([-K,K]).$$

Let $\mathcal{U}$ denote the uniform distribution on $\mathbb{Z}_{2K+1}$. For $g : \mathbb{Z}_{2K+1} \to (0,\infty)$, define

$$\mathrm{Ent}_{\mathcal{U}}(g) := E_{\mathcal{U}}\big(g \log\big(g/E_{\mathcal{U}}(g)\big)\big).$$

By definition of the *modified log-Sobolev* constant (of RW on $\mathbb{Z}_{2K+1}$), denoted $c_{\mathrm{MLS},2K+1}$, we have

$$\textstyle\sum_{x\in\mathbb{Z}_{2K+1}} \frac{1}{2}\mathcal{U}(x)\big(g(x)-g(x+1)\big)\log\big(g(x)/g(x+1)\big) \geq c_{\mathrm{MLS},2K+1}\,\mathrm{Ent}_{\mathcal{U}}(g); \tag{6.2.3}$$

see Definition 6.2.12 below. Further, it is well-known that, $c_{\mathrm{MLS},2K+1} \gtrsim c_{\mathrm{LS},2K+1} \gtrsim 1/K^2$; see Lemma 6.2.16 below. This holds for both the SRW and the DRW.

Using the backward Kolmogorov equations, an elementary calculation for the SRW gives

$$\begin{aligned} -\tfrac{d}{ds}R^-_\gamma(s) &= \textstyle\sum_{x\in\mathbb{Z}_\gamma} \frac{d}{ds}\nu^-_{\gamma,s}(x)\cdot\log\nu^-_{\gamma,s}(x) \\ &= \textstyle\sum_{x\in\mathbb{Z}_\gamma} \frac{1}{2}\big(\nu^-_{\gamma,s}(x)-\nu^-_{\gamma,s}(x+1)\big)\log\big(\nu^-_{\gamma,s}(x)/\nu^-_{\gamma,s}(x+1)\big). \end{aligned} \tag{6.2.4}$$

This actually holds for the DRW too; we explain this at the end. Note that $(a-b)\log(a/b) \geq 0$ whenever $a,b>0$. Hence all terms above are non-negative, and so, recalling the definition of ξ^- above, we have

$$-\tfrac{d}{ds}R^-_\gamma(s) \geq \nu^-_{\gamma,s}([-K,K])\textstyle\sum_{x\in[-K,K]} \frac{1}{2}\big(\xi^-_x-\xi^-_{x+1}\big)\log\big(\xi^-_x/\xi^-_{x+1}\big),$$

where we identify $\mathbb{Z}_{2K+1}$ with $[-K,K]\cap\mathbb{Z}$ and $K+1 \equiv -K$, and used the fact that $\xi^-_K = \xi^-_{-K}$.

Combining this with (6.2.3), noting that $\nu_{\gamma,s}([-K,K]) \geq \frac{8}{9}$, we see that

$$\begin{aligned} -\tfrac{d}{ds}R^-_\gamma(s) &\geq \nu_{\gamma,s}([-K,K])\textstyle\sum_{x\in\mathbb{Z}_{2K+1}} \frac{1}{2}\big(\xi^-_x-\xi^-_{x+1}\big)\log\big(\frac{\xi^-_x}{\xi^-_{x+1}}\big) \\ &= \nu_{\gamma,s}([-K,K])\textstyle\sum_{x\in\mathbb{Z}_{2K+1}} \mathcal{U}(x)\frac{1}{2}\big((2K+1)\xi^-_x-(2K+1)\xi^-_{x+1}\big)\log\big(\frac{(2K+1)\xi^-_x}{(2K+1)\xi^-_{x+1}}\big) \\ &\gtrsim K^{-2}\,\mathrm{Ent}_{\mathcal{U}}\big((2K+1)\xi^-\big) = K^{-2}D(\xi^-\,\|\,\mathcal{U}), \end{aligned}$$

where the final expression is the the relative entropy of ξ^- with respect to $\mathcal{U}$ on $\mathbb{Z}_{2K+1}$.

We argue that $D(\xi^-\,\|\,\mathcal{U}) \gtrsim 1$; the lemma then follows, since $K \asymp \sqrt{s}$. By standard exit time estimates for SRW on a cycle, $\nu^-_{\gamma,s}([-K,K]) \asymp 1$ since $K \asymp \sqrt{s}$. Since $K \leq \gamma/10$, the support of ξ, namely $[-K,K] \subseteq \mathbb{Z}_\gamma$, contains fewer than half the vertices of $\mathbb{Z}_\gamma$. Hence by Pinsker's inequality and then the triangle inequality, we have $\frac{1}{2}D(\xi^-\,\|\,\mathcal{U})^2 \geq \|\xi^- - \mathcal{U}\|_{\mathrm{TV}} \geq \frac{1}{2}$. □

6.3 Large Deviation Estimates for Random Walk on $\mathbb{Z}$

The aim of this section is to prove a large deviations result for the RW on $\mathbb{Z}$. First we must define the times at which we wish to evaluate the RW. Roughly, we look at the time at which the entropy of the RW on $\mathbb{Z}$ is $\log n/k$. This corresponds to roughly the time in the following definition.

Definition 6.3.1. *Abbreviate $\kappa := k/\log n$. Let $s_0 := s_0(k,n)$ be any time satisfying*

$$s_0 \lesssim n^{2/k}\log k \text{ when } k \lesssim \log n \quad \text{and} \quad s_0 \eqsim 1/(\kappa\log\kappa) \text{ when } k \gg \log n.$$

When we say "RW", we mean either a SRW or a DRW.

Definition 6.3.2. *Let $X = (X_s)_{s\geq 0}$ be a rate-1 RW on $\mathbb{Z}$. Define $r(k,n)$ and $p(k,n)$ as follows:*

$$r(k,n) \coloneqq \min\{r \in \mathbb{Z} \mid \mathbb{P}\big(|X_{s_0} - \mathbb{E}(X_{s_0})| > r\big) \leq 1/k^{3/2}\};$$
$$p(k,n) \coloneqq \min\{\mathbb{P}\big(X_{s_0} - \mathbb{E}(X_{s_0}) = j\big) \mid |j| \leq r(k,n)\}.$$

Also define $r_*(k,n) \coloneqq \frac{1}{2}n^{1/k}(\log k)^2$ and $p_*(k,n) \coloneqq n^{-1/k}k^{-2}$.

Proposition 6.3.3. *We have $r(k,n) \geq r_*(k,n)$ and $p(k,n) \geq p_*(k,n)$.*

This proposition will follow from standard large deviation theory, but the details are non-trivial. The exponent 2 in $(\log k)^2$ is not optimal, but is chosen for convenience of proof and to enable us to deal with all regimes of k simultaneously.

The following propositions provide asymptotic estimates for tails of the Poisson distribution and for continuous-time SRW on $\mathbb{Z}$, as well as for the ratio between the 'tail' and 'point' probabilities. We note that in the regime $r \in [\sqrt{s}, s^{2/3}]$ stronger assertions can be made via the local CLT (6.1.2).

Below, for $a, b \in \mathbb{R}$, we write $a \vee b \coloneqq \max\{a,b\}$ and $a \wedge b \coloneqq \min\{a,b\}$.

Proposition 6.3.4 (Poisson Bounds)**.** *For $s \in (0,\infty)$, let $X_s \sim \text{Poisson}(s)$. Then, uniformly in $s \in (0,\infty)$ and in r with $r \geq \sqrt{s}$ and $s + r \in \mathbb{Z}$, we have the following relations:*

$$-\log \mathbb{P}\big(X_s \geq s + r\big) \asymp r\big((r/s) \wedge 1\big)\log\big((r/s) \vee e\big); \tag{6.3.1a}$$
$$\mathbb{P}\big(X_s \geq s + r\big)/\mathbb{P}\big(X_s = s + r\big) \asymp (s/r) \vee 1. \tag{6.3.2a}$$

Moreover, uniformly in $s \in (0,\infty)$ and in $r \in [\sqrt{s}, s]$ with $s - r \in \mathbb{Z}$ we have the following relations:

$$-\log \mathbb{P}\big(X_s \leq s - r\big) \asymp r\big((r/s) \wedge 1\big)\log\big((r/s) \vee e\big); \tag{6.3.1b}$$
$$\mathbb{P}\big(X_s \leq s - r\big)/\mathbb{P}\big(X_s = s - r\big) \asymp (s/r) \vee 1. \tag{6.3.2b}$$

Proposition 6.3.5 (SRW Bounds)**.** *Let $X = (X_s)_{s\geq 0}$ be a rate-1 SRW on $\mathbb{Z}$ started at 0. Then, uniformly in $s \in (0,\infty)$ and in r with $r \geq \sqrt{s}$ and $r \in \mathbb{Z}$, we have the following relations:*

$$-\log \mathbb{P}\big(X_s \geq r\big) \asymp r\big((r/s) \wedge 1\big)\log\big((r/s) \vee e\big); \tag{6.3.3}$$
$$\mathbb{P}\big(X_s \geq r\big)/\mathbb{P}\big(X_s = r\big) \asymp (s/r) \vee 1. \tag{6.3.4}$$

From these, we can deduce the proof of Proposition 6.3.3.

Proof of Proposition 6.3.3. Recall that $\kappa = k/\log n$ and that the time s being considered satisfies

$$s \lesssim n^{2/k}\log k \text{ when } k \lesssim \log n \quad \text{and} \quad s \eqsim 1/(\kappa \log \kappa) \text{ when } k \gg \log n.$$

Consider the SRW. Equations (6.3.1–6.3.4) are all "$f \asymp g$"-type statements; let $c > 0$ be a universal constant such that c is a lower and $C \coloneqq 1/c$ an upper bound for these relations.

For r, it is enough to find an $\tilde{r}$ so that

$$-\log \mathbb{P}\big(X_s \geq \tilde{r}\big) \geq 2\log k.$$

For p, since we only consider j with $|j| \leq r$, and r is defined as a minimum, we have $\mathbb{P}(X_s \geq |j|) \geq k^{-3/2}$ for all such j. We split into two regimes, namely $s \geq 2C\log k$ and $s < 2C\log k$.

First suppose that $s \geq 2C\log k$. Set $\tilde{r} \coloneqq \sqrt{2Cs\log k}$. Then $\tilde{r} \leq s$, and so, by (6.3.3), we have

$$-\log \mathbb{P}\big(X_s \geq \tilde{r}\big) \geq c\tilde{r}\big((\tilde{r}/s) \wedge 1\big)\log\big((\tilde{r}/s) \vee e\big) = c\tilde{r}^2/s \geq 2\log k.$$

For p_*, since $\tilde{r} \leq s$, by (6.3.4), we have

$$\mathbb{P}\big(X_s = j\big) \gtrsim (s/r)\mathbb{P}\big(X_s \geq j\big) \gtrsim (\log k)^{1/2}n^{-1/k} \cdot k^{-3/2} \gg n^{-1/k}k^{-2}.$$

Suppose now that $s < 2C\log k$. Set $\tilde{r} \coloneqq 2C\log k$. Then $\tilde{r} \geq s$, and so, by (6.3.3), we have

$$-\log \mathbb{P}\big(X_s \geq \tilde{r}\big) \geq c\tilde{r}\big((\tilde{r}/s) \wedge 1\big)\log\big((\tilde{r}/s) \vee e\big) \geq c\tilde{r} = 2\log k.$$

For p_*, since $\tilde{r} \geq s$, by (6.3.4), we have

$$\mathbb{P}\big(X_s = j\big) \gtrsim \mathbb{P}\big(X_s \geq j\big) \geq k^{-3/2} \gg k^{-2} \geq n^{-1/k} k^{-2}.$$

Observe that, in either regime, we have $\tilde{r} \leq r_*$, with r_* defined in Definition 6.3.2. This completes the proof of Proposition 6.3.3 in the undirected case.

The DRW case, using Poisson bounds, is in essence the same, due to the similarity of Propositions 6.3.4 and 6.3.5. It is slightly messier to write down, since one must take care that $s+r \geq 0$. □

Proof of Proposition 6.3.4 (Poisson)**.** For $s \leq 10$, all that is needed is the observation that

$$\mathbb{P}\big(X_s \geq r\big) \asymp \mathbb{P}\big(X_s = r\big) \asymp s^r/r! \asymp \big(es/r\big)^r/\sqrt{r}.$$

We now consider the case $s \geq 1$. First we state that, for all $r \geq 0$, we have

$$\max\big\{\mathbb{P}\big(X_s \geq s+r\big), \mathbb{P}\big(X_s \leq s-r\big)\big\} \leq \exp\big(-\tfrac{1}{2}r^2/(s+r/3)\big); \tag{6.3.5}$$

this follows from Bernstein's inequality, by taking an appropriate limit.

A direct calculation involving Stirling's approximation shows, uniformly in s and in r with $r \geq \frac{1}{2}s$ and $s+r \in \mathbb{Z}$, respectively $\frac{1}{2}s \leq r \leq s$, the following relations:

$$\mathbb{P}\big(X_s \geq s+r\big) \asymp \mathbb{P}\big(X_s = s+r\big) \asymp \frac{e^r\big(s/(s+r)\big)^{s+r}}{\sqrt{2\pi(s+r)}},$$
$$\mathbb{P}\big(X_s \leq s-r\big) \asymp \mathbb{P}\big(X_s = s-r\big) \asymp \frac{e^r\big(s/(s-r)\big)^{s-r}}{\sqrt{2\pi(s-r)}};$$

from these, one can verify (6.3.2a, 6.3.2b) for such r.

We can obtain lower bounds on $\mathbb{P}(X_s \geq s+r)$ and $\mathbb{P}(X_s \leq s-r)$ for $r \leq \frac{1}{2}s$, from which, together with (6.3.5), we can verify (6.3.2a, 6.3.2b) for such r:

$$\mathbb{P}\big(X_s = s+r\big)\sqrt{2\pi(s+r)} \asymp e^r\Big(\frac{s}{s+r}\Big)^{s+r} \asymp \exp\bigg(-\frac{r^2}{2(s+r)} - \mathcal{O}\Big(\frac{r^3}{(s+r)^2}\Big)\bigg),$$
$$\mathbb{P}\big(X_s = s-r\big)\sqrt{2\pi(s-r)} \asymp e^{-r}\Big(\frac{s}{s-r}\Big)^{s-r} \asymp \exp\bigg(-\frac{r^2}{2(s-r)} - \mathcal{O}\Big(\frac{r^3}{(s-r)^2}\Big)\bigg);$$

these are found using Stirling's approximation, and both hold uniformly for $r \leq \frac{1}{2}s$.

We now prove (6.3.1a); the proof of (6.3.1b) is similar and is omitted. We consider $s \geq 10$, having already considered $s \leq 10$ initially. Observe that $r \mapsto \mathbb{P}(X_s = s \pm r)$ is decreasing on $r \geq 0$ with $s \pm r \in \mathbb{Z}$. Using the formula for $\mathbb{P}(\text{Poisson}(\lambda) = k)$, we have

$$\frac{\mathbb{P}(X_s = s+r)}{\mathbb{P}(X_s = s+r+1)} = \frac{s+r+1}{s}.$$

If $r \geq \frac{1}{4}s$, then this ratio is at least $11/9$, when $s \geq 10$, from which one can readily see that (6.3.1a) holds. Now suppose that $r \in [\sqrt{s}, \frac{1}{4}s]$. To conclude the proof, we show that there exist universal constants $c_1, c_2 \in (0,1)$ so that, for such r, we have

$$c_1\,\mathbb{P}\big(X_s = s+r\big) \leq \mathbb{P}\big(X_s = s+r+\lceil s/(2r)\rceil\big) \leq c_2\,\mathbb{P}\big(X_s = s+r\big). \tag{6.3.6}$$

This, together with the decreasing statement above, can easily be seen to imply (6.3.1a). We now prove (6.3.6). If $\sqrt{s} \leq r \leq \frac{1}{4}s$, then

$$\frac{\mathbb{P}(X_s = s+r)}{\mathbb{P}(X_s = r+r+j)} = \prod_{i=1}^{j} \frac{s+r+i}{s} = \prod_{i=1}^{j}\big(1 + (r+i)/s\big)$$
$$\leq \exp\big(\textstyle\sum_{i=1}^{j}(r+i)/s\big) = \exp\big(\frac{1}{2}j(j+2r+1)/s\big).$$

If in addition $j \leq \frac{1}{2}s/r$, then the last estimate is tight up to a constant factor. Indeed, in this case we have $\exp(\frac{1}{2}j(j+2r+1)/s) \leq e^3$. Conversely, using the fact that $1+\theta \geq \exp(\theta - 2\theta^2)$ for $\theta \in [0, \frac{1}{2}]$, we find some universal constant $c_0 > 1$ so that $\exp(\frac{1}{2}j(j+2r+1)/s) \geq c_0$. □

Proof of Proposition 6.3.5 (SRW). Fix an $s \in (0,\infty)$; without loss of generality, assume $r \geq 0$. Recall that X has the same law as $Y_N := \sum_1^N \xi_i$, where $(\xi_i)_{i\in\mathbb{N}}$ is an iid sequence of random variables with $\mathbb{P}(\xi_1 = +1) = \frac{1}{2} = \mathbb{P}(\xi_1 = -1)$ and $N \sim \text{Poisson}(s)$, independent of $(\xi_i)_{i\in\mathbb{N}}$. Then $(Y_k := \sum_1^k \xi_i)_{k\in\mathbb{Z}_+}$ is a discrete-time SRW on $\mathbb{Z}$ started at the origin.

We first prove (6.3.3). Observe that $\mathbb{E}(e^{\lambda\xi_1}) = \frac{1}{2}e^{\lambda} + \frac{1}{2}e^{-\lambda} \leq e^{\lambda^2/2}$, and so $\mathbb{E}(e^{\lambda Y_k}) \leq e^{\lambda^2 k/2}$, and hence $\mathbb{P}(Y_k \geq r) \leq \exp(-r^2/(2k))$, by taking $\lambda := r/k$. Further, an elementary calculation involving Stirling's approximation shows, uniformly over r with $\sqrt{k\log k} < r \leq k$, that

$$-\log \mathbb{P}\big(Y_k \geq r\big) \leq -\log \mathbb{P}\big(Y_k \in \{r, r+1\}\big) \asymp r^2/k;$$

for $\sqrt{k} \leq r \leq \sqrt{k\log k}$ one can use the local CLT (see Theorem 6.1.4) to verify that

$$-\log \mathbb{P}\big(Y_k \geq r\big) \asymp r^2/k.$$

For $r \leq \sqrt{2}s$, we average over N and use the above bounds on Y_k. In particular, we have

$$\mathbb{E}\big(e^{\lambda X_s}\big) \leq \textstyle\sum_{r=0}^{\infty} \mathbb{P}\big(N = k\big)\, e^{\lambda^2 k/2} = \mathbb{E}\big(e^{\lambda^2 N/2}\big) = \exp\big(s(e^{\lambda^2/2} - 1)\big) \leq \exp\big(s(\lambda^2/2 + (\lambda^2/2)^2)\big),$$

with the final inequality holding when $\lambda^2 \leq 2$, applying the inequality $e^{\theta} - 1 \leq \theta + \theta^2$ valid for $\theta \in [-1, 1]$. We now set $\lambda := r/s$ and use Chernoff to deduce that

$$\mathbb{P}\big(X_s \geq r\big) \leq \exp\big(-\tfrac{1}{2}(r^2/s)(1 - \tfrac{1}{2}(r/s))\big) \leq \exp\big(-\tfrac{1}{8}(r^2/s)\big).$$

For $r \geq \sqrt{2}s$, we use the inequalities

$$\mathbb{P}\big(X_s \geq r\big) \leq \mathbb{P}\big(\text{Poisson}(s) \geq r\big) \quad \text{and} \quad \mathbb{P}\big(X_s \geq r\big) \geq \mathbb{P}\big(N = 2r\big)\, \mathbb{P}\big(Y_{2r} \geq r\big).$$

This case is completed by applying (6.3.1, 6.3.2), ie Proposition 6.3.4.

We now prove (6.3.4). For $r \geq \frac{1}{2}s$, it follows from the fact that $r \mapsto \mathbb{P}(X_s = r)$ is decreasing and

$$\sup_{s,\, r \text{ st } r \geq s/2} \mathbb{P}\big(X_s = r+2\big)/\mathbb{P}\big(X_s = r\big) < 1,$$

which can be verified via a direct calculation involving averaging over N and applying Stirling's approximation; we omit the details. For $r \leq \frac{1}{2}s$, it suffices to prove the following corresponding result for $(Y_k)_{k\in\mathbb{Z}_+}$: uniformly in $k > 0$ and $r \in [\sqrt{k}, \frac{1}{2}k]$ with $r \in \mathbb{Z}$, we have

$$\frac{\mathbb{P}(Y_{2k} \geq 2r)}{\mathbb{P}(Y_{2k} = 2r)} \asymp \frac{k}{r} \asymp \frac{\mathbb{P}(Y_{2r+1} \geq 2r+1)}{\mathbb{P}(Y_{2r+1} = 2r+1)}; \tag{6.3.7}$$

from this, the original claim follows by averaging over N. Using Stirling's approximation, it is not hard to verify for $r \in [\sqrt{k}, \frac{1}{2}k]$ that there exist universal constants $c_1, c_2 \in (0,1)$ such that the following hold:

$$\begin{aligned} c_1\, \mathbb{P}\big(Y_{2k} = 2r\big) \leq \mathbb{P}\big(Y_{2k} = 2(r + \lceil k/r \rceil)\big) \leq c_2\, \mathbb{P}\big(Y_{2k} = 2r\big); \\ c_1\, \mathbb{P}\big(Y_{2k+1} = 2r+1\big) \leq \mathbb{P}\big(Y_{2k+1} = 2(r + \lceil k/r \rceil) + 1\big) \leq c_2\, \mathbb{P}\big(Y_{2k+1} = 2r+1\big). \end{aligned}$$

This, together with the fact that both $r \mapsto \mathbb{P}(Y_{2k} = 2r)$ and $r \mapsto \mathbb{P}(Y_{2k+1} = 2r+1)$ are decreasing on $[0, k]$, is easily seen to imply (6.3.7). □

6.4 Simple Random Walk Exit Times Estimates

In this section, we prove some estimates on exit times for SRW on the integers. These results were used in the spectral gap analysis of Chapter 4. The following auxiliary lemma is needed.

Lemma 6.4.1. *For $\varphi \in [-\frac{1}{2}, \frac{1}{2}]$, we have*

$$2(\pi\varphi)^2 \geq 1 - \cos(2\pi\varphi) \geq \tfrac{2}{3}(\pi\varphi)^2.$$

Proof of Lemma 6.4.1. Let $\varphi \in [-\frac{1}{2}, \frac{1}{2}]$. Then, using the fact that $\log(1-x) \geq -x-x^2$ for $|x| < 1$, that $\sum_1^\infty 1/i^2 = \frac{1}{6}\pi^2$, that $\sum_1^\infty 1/i^4 = \frac{1}{90}\pi^4$ and that $\varphi \in [-\frac{1}{2}, \frac{1}{2}]$, we can calculate directly:

$$1 \geq \frac{1-\cos(2\pi\varphi)}{2(\pi\varphi)^2} = \Big(\frac{\sin(\pi\varphi)}{\pi\varphi}\Big)^2 = \prod_{\ell=1}^{\infty}\Big(1-\frac{\varphi^2}{\ell^2}\Big)^2 \geq \exp\Big(-2\sum_{\ell=1}^{\infty}\Big(\frac{\varphi^2}{\ell^2}+\frac{\varphi^4}{\ell^4}\Big)\Big) \geq 0.383 \geq \tfrac{1}{3}. \quad \square$$

Lemma 6.4.2. *Let $\ell \in \mathbb{N}$ and $\tau := \inf\{s \geq 0 \mid |Y_s| = \ell\}$, where $(Y_s)_{s\geq 0}$ is a continuous-time rate-1 SRW on $\mathbb{Z}$. Let $\theta := \frac{1}{2}\pi/\ell$ and $\lambda := 1-\cos\theta$. Then, for all $s \geq 0$, we have*

$$\mathbb{P}_0\big(\tau > s\big) \geq e^{-\lambda s} \geq \exp\big(-\tfrac{1}{8}s(\pi/\ell)^2\big).$$

Proof of Lemma 6.4.2. The second inequality follows from Lemma 6.4.1.

For the first inequality, we first note that

$$\mu : x \mapsto \cos(\theta x)\big/\textstyle\sum_{j=-\ell}^{\ell}\cos(\theta j) : \{-\ell, ..., \ell\} \to [0,1]$$

is a distribution satisfying $\mu(\pm\ell) = 0$ and

$$(\mu\widehat{P})(x) = \mu(x)\,\cos\theta \quad \text{for} \quad x \in J = \{-\ell+1, ..., \ell-1\},$$

where $\widehat{P}$ is the transition matrix of discrete-time SRW on $\{-\ell, ..., \ell\}$ with absorption at the boundary. Indeed, using $\mu(\pm\ell) = 0$ we have $(\mu\widehat{P})(x) = \frac{1}{2}(\mu(x+1)+\mu(x-1)) = \mu(x)\cos(\pi/(2\ell))$, where we have used $\cos(a+b)+\cos(a-b) = 2\cos a\cos b$. If follows that starting from initial distribution μ we have $\mu\widehat{P}^i(J) = (1-\lambda)^i$, where $\widehat{P}^i$ is the matrix $\widehat{P}$ raised to the power i, and so $\mu\widehat{P}^i(J)$ is the probability of not getting absorbed at the boundary by the i-th step when the initial distribution is μ. It follows that

$$\mathbb{P}_\mu\big(\tau > s\big) = \textstyle\sum_{i=0}^{\infty}\mu\widehat{P}^i(J)\,\mathbb{P}\big(\text{Poisson}(t) = i\big) = e^{-\lambda s}.$$

By considering the continuous-time chain with jump-matrix $\widehat{P}$, we obtain, for all $s \geq 0$, that

$$\mathbb{P}_0\big(\tau > s\big) = \max_{j\in J}\mathbb{P}_j\big(\tau > s\big),$$

as can be seen by a simple coupling argument; cf [49, Example 5.1]. This concludes the proof. $\square$

Definition 6.4.3. *For a transition matrix P and a set A, let λ_A be the minimal Dirichlet eigenvalue, namely the minimal eigenvalue of minus the generator of the chain killed upon exiting A, ie of*

$$I_A - P_A \quad \text{where} \quad (I_A - P_A)(x,y) := \mathbf{1}(x,y \in A)\big(\mathbf{1}(x=y) - P(x,y)\big).$$

Also, for a set A, write τ_{A^c} for the (first) exit time of this set by the chain.

Lemma 6.4.4. *Consider a rate-1, continuous-time, reversible Markov chain with transition matrix P. Let A be a connected set, and let λ_A and τ_{A^c} be as in Definition 6.4.3. For all $a \in A$, we have*

$$-\tfrac{1}{t}\log\mathbb{P}_a\big(\tau_{A^c} > t\big) \to \lambda_A \quad \text{as } t \to \infty.$$

Proof of Lemma 6.4.4. For connected A, by the Perron–Frobenius theorem, the quasi-stationary distribution of A, which we denote by $\mu = (\mu_a)_{a\in A}$, is positive everywhere on A. (See [2, §3.6.5] for the definition of quasi-stationarity.) Since $\mathbb{P}_\mu(\tau_{A^c} > t) = \sum_{a\in A}\mu_a\,\mathbb{P}_a\big(\tau_{A^c} > t\big)$, we have

$$\mathbb{P}_a\big(\tau_{A^c} > t\big) \leq \mu_a^{-1}\,\mathbb{P}_\mu\big(\tau_{A^c} > t\big) = \mu_a^{-1}\exp(-\lambda_A t),$$

since the exit time starting from the quasi-stationary distribution is exponential with rate λ_A, as shown in the equation proceeding (3.83) in [2]. This proves the upper bound, taking $t \to \infty$.

For the other direction, we claim that there exists a constant c, independent of a and t, so that

$$\min_{a\in A}\mathbb{P}_a\big(\tau_{A^c} > t\big) \geq c\,\max_{a\in A}\mathbb{P}_a\big(\tau_{A^c} > t+1\big).$$

Indeed, let a' be an element of A attaining the maximum at time $t+1$. Using the connectedness of A, for any other $a \in A$ there exists a path from a' to a consisting of states belonging to A. The probability that the walk traverses this path, and does so in time less than 1, is at least c, for some c independent of t. From this we deduce that

$$\min_{a\in A} \mathbb{P}_a\big(\tau_{A^c} > t\big) \geq c\,\mathbb{P}_\mu\big(\tau_{A^c} > t+1\big) = c\,\exp\big(-\lambda_A(t+1)\big).$$

This proves the lower bound, taking $t \to \infty$, and hence proves the lemma. □

6.5 Size of Discrete Lattice Ball Estimates

We wish to determine the size of the L_q balls in $\mathbb{Z}^k$. In particular, we desire $R_{k,q}$ so that

$$|B_{k,q}(R_{k,p})| \approx n \quad \text{where} \quad B_{k,q}(R) := \big\{a \in \mathbb{Z}^k \;\big|\; \textstyle\sum_1^k |a_i|^q \leq R^q\big\}.$$

This is done by Lemmas 6.5.2 and 6.5.3. First we need a definition and preliminary lemma.

Definition 6.5.1. *Set $\omega := \max\{(\log k)^2, k/n^{1/(2k)}\}$, and choose $R_{k,q}$ to be the minimal integer satisfying $|B_{k,q}(R_{k,q})| \geq ne^{\omega}$. Note that ω satisfies $1 \ll \omega \ll k$ if $k \ll \log n$.*

For $q \in [1,\infty)$, write $V_{k,q}(R)$ for the (Lebesgue) volume of the L_q ball of radius R in $\mathbb{R}^k$, ie

$$V_{k,q}(R) := \mathrm{vol}\big\{x \in \mathbb{R}^k \;\big|\; \|x\|_q \leq R\big\};$$

also write $V_{k,q} := V_{k,q}(1)$ and note that $V_{k,q}(R) = R^k V_{k,q}$. It is known (see [76]) that

$$V_{\ell,q} = 2^\ell \Gamma(1/q+1)^\ell / \Gamma(\ell/q+1). \tag{6.5.1}$$

We can use this, along with Lemma 6.5.2b below, to well-approximate $|B_{k,q}(R)|$ when $q \notin \{1,\infty\}$; for $q=1$ we directly bound $|B_{k,1}(\cdot)|$, while for $q=\infty$ we have an exact expression.

Lemma 6.5.2a. *For $q=1$ and all $R \geq 0$, we have*

$$2^{k\wedge R}\tbinom{\lfloor R\rfloor}{k}\mathbf{1}(R \geq k) \leq \big|B_{k,1}(R)\big| \leq 2^{k\wedge R}\tbinom{\lfloor R\rfloor + k}{k}. \tag{6.5.2a}$$

Lemma 6.5.2b. *For $q \in (1,\infty)$ and all $R \geq k^{1+1/q}$, we have*

$$\big|B_{k,q}(R)\big| = V_{k,q}(R)\,\big(1 + \mathcal{O}\big(k^{1+1/q}/R\big)\big). \tag{6.5.2b}$$

Lemma 6.5.2c. *For $q=\infty$ and all $R \geq 0$, we have*

$$\big|B_{k,\infty}(R)\big| = \big(2\lfloor R\rfloor + 1\big)^k. \tag{6.5.2c}$$

Proof of Lemma 6.5.2a. Assume $R \in \mathbb{N}$. Observe that

$$\big|B_{k,1}(R)\big| = \big|\big\{a \in \mathbb{Z}^k \mid \textstyle\sum_{i=1}^k |a_i| \leq R\big\}\big|.$$

Moreover, it is a standard combinatorial identity that

$$\big|\big\{\alpha \in \mathbb{Z}_+^k \mid \textstyle\sum_{i=1}^k \alpha_i \leq R\big\}\big| = \tbinom{R+k}{k}.$$

The upper and lower bounds will follow easily from this view point, setting $\alpha_i := |a_i|$.

For the upper bound, note that $\alpha_i = |\pm a_i|$, and so given the value of α_i, there are two choices for a_i if $\alpha_i > 0$, otherwise there is only one (since $0 = -0$). Hence

$$\big|\big\{a \in \mathbb{Z}^k \mid \textstyle\sum_{i=1}^k |a_i| \leq R\big\}\big| \leq 2^{k\wedge R}\big|\big\{a \in \mathbb{Z}_+^k \mid \textstyle\sum_{i=1}^k \alpha_i \leq R\big\}\big| = 2^{k\wedge R}\tbinom{R+k}{k},$$

noting that there are at most $k \wedge R$ non-zero coordinates for which a sign can be chosen.

For the lower bound, we get the factor of $2^{k\wedge R}$ by only considering $a \in \mathbb{Z}^k$ with $|a_i| > 0$ for all i, and then setting $\beta_i := \alpha_i - 1$. Concretely, for $R \geq k$, we have

$$\begin{aligned}\big|\{a \in \mathbb{Z}^k \mid \textstyle\sum_{i=1}^k |a_i| \leq R\}\big| &\geq \big|\{a \in \mathbb{Z}^k \mid \textstyle\sum_{i=1}^k |a_i| \leq R,\ a_i \neq 0\,\forall\, i = 1, ..., k\}\big| \\ &= 2^{k\wedge R}\big|\{\alpha \in \mathbb{Z}^k \mid \textstyle\sum_{i=1}^k \alpha_i \leq R,\ \alpha_i > 0\,\forall\, i = 1, ..., k\}\big| \\ &= 2^{k\wedge R}\big|\{\beta \in \mathbb{Z}^k \mid \textstyle\sum_{i=1}^k \beta_i \leq R - k,\ \beta_i \geq 0\,\forall\, i = 1, ..., k\}\big| = 2^{k\wedge R}\tbinom{R}{k}.\end{aligned}$$

□

Proof of Lemma 6.5.2b. For any R, writing diam_q for the L_q diameter (in $\mathbb{R}^k$), we have

$$B_q^k\big(R - \operatorname{diam}_q[-\tfrac{1}{2}, \tfrac{1}{2})^k\big) \subseteq B_q^k(R) \subseteq B_q^k\big(R + \operatorname{diam}_q[-\tfrac{1}{2}, \tfrac{1}{2})^k\big).$$

Note that $\operatorname{diam}_q[-\frac{1}{2}, \frac{1}{2})^k = k^{1/q}$. Hence, for R with $R \geq k^{1+1/q}$, we have

$$\big|B_q^k(R)\big| = \big(1 + \mathcal{O}(k^{1/q}/R)\big)^k = 1 + \mathcal{O}\big(k^{1+1/q}/R\big).$$

Cf [46, Lemma 2.5], where the case $q = 2$ is considered; there, convolutions are employed. □

Proof of Lemma 6.5.2c. In the L_∞ norm, the coordinates are independent. The claim follows. □

We use this lemma to find an $R_{k,p}$ from Definition 6.5.1, which is the minimal integer satisfying $|B_{k,q}(R_{k,p})| \geq ne^\omega$. Recall that $\mathfrak{M}_{k,q} = k^{1/q}n^{1/k}/C_q$, and that $C_q = 2\,\Gamma(1/q + 1)(qe)^{1/q}$. The next lemma shows that the difference between M and $\mathfrak{M}$ is only by sot. Also, let K be a constant, assumed to be as large as required, and let $\xi := 1 - e^{-K\omega/k}$ when $k \ll \log n$. (As such, we can always replace $1 \pm \xi$ by $e^{\pm\xi}$.)

Lemma 6.5.3a. *For $k \ll \log n$ and $q = 1$, we have*

$$R_{k,1} \leq \big\lceil \mathfrak{M}_{k,1}(1 + \xi)\big\rceil \quad \text{and} \quad \big|B_{k,1}\big(\mathfrak{M}_{k,1}(1 - \xi)\big)\big| \ll n. \tag{6.5.3a}$$

Lemma 6.5.3b. *For $k \leq \log n/\log\log n$ and all $q \in [1, \infty)$, we have*

$$R_{k,q} \leq \big\lfloor \mathfrak{M}_{k,q}(1 + \xi)\big\rfloor \quad \text{and} \quad \big|B_{k,q}\big(\mathfrak{M}_{k,q}(1 - \xi)\big)\big| \ll n. \tag{6.5.3b}$$

Lemma 6.5.3c. *For $q = \infty$, we have*

$$R_{k,\infty} = \big\lceil \tfrac{1}{2}n^{1/k}e^{\omega/k} - \tfrac{1}{2}\big\rceil \quad \text{and} \quad \big|B_{k,\infty}\big(\mathfrak{M}_{k,\infty}(1 - \xi)\big)\big| \ll n. \tag{6.5.3c}$$

Moreover, if $k \ll \log n$ then $R_{k,\infty} \asymp \mathfrak{M}_{k,\infty}$.

Lemma 6.5.3d. *For all $\lambda > 0$, for $k \asymp \lambda \log n$, there exists a function $\omega \gg 1$ and a constant α so that, for all $\varepsilon \in (0, 1)$, the minimal integer M_1 satisfying $|B_{k,1}(M_1)| \geq ne^\omega$ satisfies*

$$R_{k,1} \asymp \alpha k \asymp \alpha\lambda \log n \quad \text{and} \quad \big|B_{k,1}\big(\alpha k(1 - \varepsilon)\big)\big| \ll n. \tag{6.5.3d}$$

In fact, the result holds for any $1 \ll \omega \ll k$.

Proof of Lemma 6.5.3a. *Upper Bound.* Write $M := \lceil e^\xi kn^{1/k}/(2e)\rceil$. Note that $k \ll \log n$, and so $n^{1/k} \gg 1$, and so $M \gg k$. Then, by (6.5.2a) and Stirling's formula, we have

$$\begin{aligned}\big|B_{k,1}(M)\big| \geq 2^k\tbinom{M}{k} \geq 2^k(M - k)^k/k! &\gtrsim k^{-1/2}(1 - k/M)^k(2eM/k)^k \\ &\geq k^{-1/2}\exp\big(-k(2k/M + \xi)\big) \cdot n.\end{aligned}$$

Take $\xi := 2\omega/k$: then $k/M \asymp n^{-1/k} \ll n^{-1/(2k)} \leq \xi$ and $e^{-\xi k} \gg k^{1/2}$. Hence $|B_{k,1}(M)| \geq ne^\omega$.

Lower Bound. Set $M := kn^{1/k}e^{-K\omega/k}/(2e)$. Using $\binom{N}{k} \leq (eN/k)^k$ and (6.5.2a), we have

$$\big|B_{k,1}(M)\big| \leq \big(2e(M/k + 1)\big)^k \leq ne^{-K\omega}\exp\big(6k/n^{1/k}\big) \ll n,$$

using $1 + x \leq e^x$ with $x = k/M$, $\binom{N}{r} \leq (eN/r)^r$ and $\omega \geq k/n^{1/(2k)} \gg k/n^{1/k}$ as $k \ll \log n$. □

Proof of Lemma 6.5.3b. *Upper Bound.* From the formula (6.5.1), we see that

$$R_{k,q} \coloneqq n^{1/k} e^{2\omega/k} / V_{k,q}^{1/k} = \tfrac{1}{2} n^{1/k} e^{2\omega/k} \Gamma(k/q+1)^{1/k} / \Gamma(1/q+1)$$

satisfies $V_q^k(R_{k,q}) = ne^{2\omega}$. Using Stirling's formula, and the fact that $k \gg 1$, we then deduce that

$$R_{k,q} \le n^{1/k} k^{1/q} e^{\xi} / C_q.$$

Observe that $k^{1+1/q}/R_{k,q} \asymp k/n^{1/k} \ll k/n^{1/(2k)} \le \omega$. Applying Lemma 6.5.2b with $R \coloneqq R_{k,q}$, which is valid since $k \le \log n / \log\log n$, implying $n^{1/k} \gg k$ and hence $R_{k,q} \gg k^{1+1/q}$, gives

$$\big|B_q^k(R_{k,q})\big| / V_q^k(R_{k,q}) = 1 + \mathcal{O}\big(k^{1+1/q}/R_{k,q}\big) = \exp\big(o(\omega)\big).$$

Noting that $V_q^k(R_{k,q}) = ne^{2\omega}$, we hence deduce that $|B_q^k(R_{k,q})| \ge ne^{\omega}$.

Lower Bound. Set $M \coloneqq k^{1/q} n^{1/k} e^{-K\omega/k} / C_q$. Then, by (6.5.1) and Stirling, we have

$$V_q^k(M) = C_k M^k = ne^{-K\omega} k^{k/q} \big/ \big(\Gamma(k/q+1)(qe)^{k/q}\big) \ll n.$$

Note that $M \gg k^{1+1/q}$ since $k \le \log n / \log\log n$, and hence $|B_q^k(M)| \ll n$ by Lemma 6.5.2b. □

Proof of Lemma 6.5.3c. *Upper Bound.* This is immediate from (6.5.2c) and the relation $n^{1/k} \gg 1$.

Lower Bound. Recall Lemma 6.5.2c. Observe that

$$(2M+1)^k \le ne^{-\nu} \quad \text{if and only if} \quad k\log(2M) + k\log\big(1 + 1/(2M)\big) \le \log n - \nu.$$

Let us set $M \coloneqq \frac{1}{2} n^{1/k} e^{-K\omega/k}$, for a constant K. Then

$$(2M+1)^k \le ne^{-\nu} \quad \text{if and only if} \quad \log n - K\omega + k\log\big(1 + 1/(2M)\big) \le \log n - \nu.$$

Recall that $\omega \ge k/n^{1/(2k)} \gg k/n^{1/k} \asymp k/M$. Hence, for any constant K, we have

$$\big|B_\infty^k(M)\big| \le (2M+1)^k \ll n,$$

by choosing $\nu \gg 1$ but with $\nu = o(\omega)$. Also, $k \ll \log n$, so $\lfloor M \rfloor \gg 1$. □

Proof of Lemma 6.5.3d. We first prove that there exists a strictly increasing function $c : (0,\infty) \to (0,\infty)$ so that, for all $a > 0$, omitting here and below all ceiling signs, we have

$$\big|B_{k,1}(ak)\big| = \exp\big(k\big(c(a) + o(1)\big)\big).$$

By considering the number i of coordinates which equal 0, we have $|B_{k,1}(ak)| = \sum_{i=0}^{k} A_i$, where

$$A_i \coloneqq A_i(k,a) \coloneqq \tbinom{k}{i} 2^{k-i} \tbinom{k-i+ak}{ak}.$$

Choose $i_* \coloneqq i_*(k,a)$ that maximises A_i. Then $A_{i_*} \le |B_{k,1}(ak)| \le (k+1)A_{i^*}$. Observe that

$$\frac{A_{i+1}}{A_i} = \frac{(k-i)^2}{2(i+1)(k(1+a)-i)},$$

and hence one can determine i_* as a function of k and a, conclude that $i_*(a,k)/k$ converges as $k \to \infty$ and thus determine $c(a)$ (in terms of the last limit). We omit the details. Knowing this limit allows us to plug this into the definition of A_i and use Stirling's approximation to get

$$A_{i_*} = \exp\big(k\big(c(a) + o(1)\big)\big),$$

for some strictly increasing function $c : (0,\infty) \to (0,\infty)$. Since $k+1 = e^{o(k)}$, the claim follows.

Upper Bound. Since $k \approx \lambda \log n$, we have $M_1/k \to c^{-1}(1/\lambda)$ as $n \to \infty$; set $\alpha \coloneqq c^{-1}(1/\lambda)$.

Lower Bound. It follows from the exponential increase in the size of the L_1 ball that $|B_{k,1}((1-\varepsilon)\alpha k)| = o(n)$ for all $\varepsilon > 0$, where $M_1 \approx \alpha k$ and $\alpha = c^{-1}(1/\lambda)$. □

6.6 Some Further Deferred Proofs

6.6.1 Uniformity of Linear Combination of Uniform Random Variables

Lemma 6.6.1. *Let G be Abelian, $k \in \mathbb{N}$ and $v \in (\mathbb{Z} \setminus \{0\})^k$. Draw $Z_1, ..., Z_k \sim^{\text{iid}} \text{Unif}(G)$. Then*

$$v \cdot Z = \sum_1^k v_i Z_i \sim \text{Unif}(\mathfrak{g}G) \quad \textit{where} \quad \mathfrak{g} \coloneqq \gcd(v_1, ..., v_k, |G|).$$

Proof. Decompose G as $\oplus_1^d \mathbb{Z}_{m_j}$. Write $\mathfrak{g}_j \coloneqq \gcd(v_1, ..., v_k, m_j)$ for each $j \in [d]$. Then, for each $i \in [k]$, we may write $Z_i = (\zeta_{i,1}, ..., \zeta_{i,d})$ with $\zeta_{i,j} \sim \text{Unif}(\mathbb{Z}_{m_j})$ with all the $\zeta_{i,j}$ independent. Then

$$(v \cdot Z)_j = \sum_{i=1}^k v_i \zeta_{i,j},$$

where $(v \cdot Z)_j$ is the j-th component of $v \cdot Z \in \mathbb{Z}^d$, and in particular $((v \cdot Z)_j)_{j=1}^d$ are independent. Assuming the $d = 1$ case, the above then shows that $(v \cdot Z)_j \sim \text{Unif}(\mathfrak{g}_j \mathbb{Z}_{m_j/\mathfrak{g}_j})$ for each j. Hence it suffices to prove the $d = 1$ case.

We now prove the $d = 1$ case. Since any $i \in [k]$ with $v_i \equiv 0 \bmod m$ does not contribute to the sum, by passing to a subsequence, we may assume that $v_i \not\equiv 0 \bmod n$ for all $i \in [k]$.

We use induction on $|\mathcal{I}|$. Let $U \sim \text{Unif}\{1, ..., n\}$ and set $R \coloneqq mU$ where $m \in \{1, ..., n\}$. Define

$$\mathfrak{g} \coloneqq \gcd(m, n) \text{ and } r \coloneqq m/\mathfrak{g} \quad \text{so that} \quad R = mU = \mathfrak{g} \cdot (rU).$$

We then have $\gcd(r, n) = 1$, and so $rU \sim \text{Unif}\{1, ..., n\}$: indeed, for any $x \in \{1, ..., n\}$, we have

$$\mathbb{P}\big(rU = x\big) = \mathbb{P}\big(U = xr^{-1}\big) = \tfrac{1}{n} \quad \text{where} \quad r^{-1} \text{ is the inverse of } r \bmod n.$$

Thus we have $R = \mathfrak{g} \cdot (rU) \sim \text{Unif}\{\mathfrak{g}, 2\mathfrak{g}, ..., n\}$, since $\mathfrak{g} \wr n$. This proves the base case $|\mathcal{I}| = 1$.

Now consider independent $X, Y \sim \text{Unif}\{1, ..., n\}$ and set $R \coloneqq aX + bY$. By pulling out a constant as above, we may assume that $a, b \wr n$. Write $c \coloneqq \gcd(a, b, n)$. Then there exist $r, s \in \{1, ..., n\}$ with

$$ar + bs \equiv c \bmod n, \quad \text{and hence} \quad a(mr) + b(ms) \equiv cm \bmod n \text{ for any } m \in \{1, ..., n\}.$$

Thus $\{c, 2c, ..., n\} \subseteq \text{supp}(R)$. By writing $R \coloneqq c(ac^{-1}X + bc^{-1}Y)$, with c^{-1} the inverse $\bmod n$, we see that in fact $\text{supp}(R) = \{c, 2c, ..., n\}$. It remains to show that R is uniform on its support.

Pulling out the factor c, it is enough to consider $\gcd(a, b, n) = 1$. For $m \in \{0, 1, ..., n-1\}$, set

$$\Omega_m \coloneqq \big\{(x, y) \in [n]^2 \mid ax + by \equiv m \bmod n\big\}.$$

We show that $|\Omega_m|$ is the same for all m, and hence deduce that R is uniform on $\{c, 2c, ..., n\}$. Indeed, for every m there exists a pair $(x_m, y_m) \in [n]^2$ so that $ax_m + by_m = m$. If also $(x, y) \in \Omega_m$, then letting $x' = x - x_m$ and $y' = y - y_m$, we see that $(x', y') \in \Omega_0$. This proves the case $|\mathcal{I}| = 2$.

Now suppose that $X_1, ..., X_L \sim^{\text{iid}} \text{Unif}\{1, ..., n\}$ and $a_1, ..., a_L \in \{1, ..., n-1\}$. By the hypothesis,

$$\sum_{\ell=1}^{L-1} a_\ell X_\ell \sim c_0 U \quad \text{where} \quad U \sim \text{Unif}\{1, ..., n\} \quad \text{and} \quad c_0 \coloneqq \gcd(a_1, ..., a_{L-1}, n).$$

Now, X_L is independent of this sum, and so the previous case applies to say that

$$\sum_{\ell=1}^{L} a_\ell X_\ell \sim cU \quad \text{where} \quad U \sim \text{Unif}\{1, ..., n\} \quad \text{and} \quad c \coloneqq \gcd(c_0, a_L, n) = \gcd(a_1, ..., a_L, n).$$

This completes the induction, and hence proves the claim. □

6.6.2 Decomposition for Product of Upper Triangular Matrices

Write $H_{p,d}$ for the set of $d \times d$ uni-upper triangular matrices with entries in $\mathbb{Z}_p$.

Lemma 6.6.2. *Let $Z_1, ..., Z_k \in H_{p,d}$. Let $\gamma \in [k]^L$ and $\sigma \in \{\pm 1\}^L$. For $i, j \in [k]$, set*

$$C_{i,j}(\gamma, \sigma) \coloneqq \sum_{\ell=0}^{L} \sum_{m=0}^{\ell-1} \sigma_m \sigma_\ell \mathbf{1}(\gamma_m = i,\, \gamma_\ell = j) + \mathbf{1}(i = j) \sum_{\ell=0}^{L} \mathbf{1}(\gamma_\ell = i,\, \sigma_\ell = -1).$$

Set $M \coloneqq Z_{\gamma_1}^{\sigma_1} \cdots Z_{\gamma_L}^{\sigma_L}$. *Then, for all* $a \in [d]$, *we have*

$$M(a,a) = 1 \quad \text{and} \quad M(a,a+1) = \textstyle\sum_{\ell=1}^{L} \sigma_{\gamma_\ell} Z_{\gamma_\ell}(a,a+1),$$

and, for all $a, b \in [d]$ *with* $b \geq a+2$, *we have*

$$M(a,b) = \textstyle\sum_{\ell \in [L]} Z_{\gamma_\ell}(a,b) + \sum_{i,j \in [k]} C_{i,j}(\gamma,\sigma) Z_i(a,a+1) Z_j(a+1,b) + g_{a,b}(\gamma,\sigma; Z_1, ..., Z_k),$$

where $g_{a,b}(\gamma,\sigma; Z_1, ..., Z_k)$ *is a polynomial in* $(Z_i(x,y) : i \in [k],\ x \in [d-1],\ y > x)$. *Further, in this polynomial, each monomial contains the term* $Z_i(a,a+1)$ *either 0 times or exactly once and no monomial contains a term of the form* $Z_i(a,a+1)Z_j(a+1,b)$ *for* $i, j \in [k]$.

Proof. Given $M_\ell \in H_{p,d}$, we can write $M_\ell = I + N_\ell$ with N_ℓ strictly upper triangular. Consider now $M_1, ..., M_L \in H_{p,d}$; write $M \coloneqq M_1 \cdots M_L$. From the above expression for $M_\ell^{\sigma_\ell}$ and the fact that N_ℓ is strictly upper triangular, the claimed expression for $M(a,a+1)$ is immediate—specifically, $(\prod_{r=1}^{\ell} N_{m_r})(a,a+1) = 0$ for all $m_1, ..., m_\ell \in [L]$ and $a \in [L-1]$ when $\ell \geq 2$.

Herein we consider the terms above the super-diagonal, ie (a,b) with $b \geq a+2$. Observe that

$$\textstyle\prod_{\ell=1}^{L}(I + N_\ell) = \sum_{\ell=0}^{L} \sum_{m_1 < \cdots < m_\ell} \prod_{r=1}^{\ell} N_{m_r}$$

where the indices $m_1, ..., m_\ell$ run over all of $[L]$. Then, for (a,b) with $b \geq a+2$, we have

$$\begin{aligned}\big(\textstyle\prod_{r=1}^{\ell} N_{m_r}\big)(a,b) &= \textstyle\sum_{c_0,\ldots,c_\ell \in [d]} \mathbf{1}(c_0 = a, c_\ell = b) \prod_{r=1}^{\ell} N_{m_r}(c_{r-1}, c_r) \\ &= \textstyle\sum_{a = c_0 < c_1 < \cdots < c_{\ell-1} < c_\ell = b} \prod_{r=1}^{\ell} M_{m_r}(c_{r-1}, c_r),\end{aligned}$$

using the strict upper triangular property of the N_ℓ. Similarly, for (a,b) with $b \geq a+2$, we have

$$\big(N_{m_1} N_{m_2}\big)(a,b) = M_{m_1}(a,a+1) M_2(a+1,b) + \textstyle\sum_{c=a+2}^{b-1} M_{m_1}(a,c) M_2(c,b).$$

Next observe that $N_\ell^d = 0$ as N_ℓ is strictly upper triangular. Hence, for any $\sigma \in \{\pm 1\}$, we have

$$M_\ell^{\sigma_\ell} = I + \sigma_\ell N_\ell + N_\ell^2 \mathbf{1}(\sigma_\ell = -1) + \textstyle\sum_{t=3}^{d} (-1)^t N_\ell^t \mathbf{1}(\sigma_\ell = -1).$$

Recall that $M = M_1 \cdots M_L$. Then, for (a,b) with $b \geq a+2$, we may write

$$\begin{aligned}M(a,b) = \textstyle\sum_m M(a,b) + \sum_{m_1 < m_2} M_{m_1}(a,a+1) M_{m_2}(a+1,b) + R(a,b) \\ + \textstyle\sum_m M_m(a,a+1) M_m(a+1,b) \mathbf{1}(\sigma_m = -1)\end{aligned}$$

where $R(a,b)$ is a 'remainder' polynomial, containing the matrix products of degree 2 and higher *except for* those of the form $M_{m_1}(a,a+1)M_{m_2}(a+1,b)$. Indeed, since the sequence $(c_0, ..., c_\ell)$ is strictly increasing, each monomial in $R(a,b)$ contains the term $M_m(a,a+1)$ for $m \in [L]$ either 0 times or exactly once and, since $\ell \geq 3$ and $c_\ell = b$, no monomial in $R(a,b)$ contains a term of the form $M_{m_1}(a,a+1)M_{m_2}(a+1,b)$ for $m_1, m_2 \in [L]$.

Suppose now that $M_\ell = Z_{\gamma_\ell}$ for some $\gamma \coloneqq (\gamma_\ell)_1^L \in [k]^L$. By the above analysis, the 'first order' term $\sum_{\ell=1}^{L} Z_{\gamma_\ell}(a,b)$ has the desired form and the 'remainder' term has the desired property. Thus it only remains to check that the 'second order' term has the desired form. Writing $\sum_{m_1 < m_2}$ as $\sum_{m_2=1}^{L} \sum_{m_1=1}^{m_2-1}$, the $i \neq j$ case follows from some simple algebra; cf (3.1.5, 3.1.9, 3.1.10) for the $d = 3$ case. The analysis of $C_{i,i}$ is similar (and depends on whether or not inverses are allowed). □

6.6.3 Uniform Random Variables in Nilpotent Groups

Lemma 6.6.3. *For each* $\ell \in [L]$, *let* $Y_\ell \sim^{\text{iid}} \text{Unif}(R_\ell)$. *Then* $Y \coloneqq Y_1 \cdots Y_L \sim \text{Unif}(G)$.

Proof. Let $r_0 \in G$ and consider the event $\{Y = r_0\}$.

If $r_0 = Y_1 \cdots Y_L$, then $r_1 \coloneqq Y_1^{-1} r_0 = Y_2 \cdots Y_L$. Clearly the right-hand side is in G_1, and so the left-hand side must be too. Hence $r_0 \equiv Y_1 \bmod G_1$, ie $\pi_1(r_0) = Y_1$. Since $Y_1 \sim \text{Unif}(R_1)$, the

probability of this is $1/|R_1| = 1/|G_0/G_1|$. Similarly, $r_2 \coloneqq Y_2^{-1} r_1 = Y_3 \cdots Y_L$, and we deduce that $r_2 \equiv Y_2 \bmod G_2$, the probability of which is $1/|R_2| = 1/|G_1/G_2|$.

Iterating this argument, recalling that the Y_ℓ are independent, we deduce that

$$\mathbb{P}\big(Y = r_0\big) = \prod_1^L 1/|G_{\ell-1}/G_\ell| = \prod_1^L |G_\ell|/|G_{\ell-1}| = |G_L|/|G_0| = 1/|G|.$$

Since $r_0 \in G$ was arbitrary, we deduce that $Y \sim \mathrm{Unif}(G)$. □

This gives the following corollary.

Corollary 6.6.4. *For each $i \in [k]$ and $\ell \in [L]$, sample $Z_{i,\ell} \sim \mathrm{Unif}(R_\ell)$ independently; set $Z_i \coloneqq Z_{i,1} \cdots Z_{i,L}$. Then $Z_1, ..., Z_L \sim^{\mathrm{iid}} \mathrm{Unif}(G)$. Further, $Z_{i,\ell} G_\ell \sim \mathrm{Unif}(Q_\ell)$ independently for each (i, ℓ).*

For the remainder of the section, assume that Z is drawn in this way.

Proof. All the independence claims are immediate. The first claim is immediate from Lemma 6.6.3.

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 $XG_\ell \sim \mathrm{Unif}(Q_\ell)$. □

6.6.4 A Bound on the Number of Divisors of an Integer

In this section, we prove the following number-theoretic result.

Lemma 6.6.5. *For all $\varepsilon > 0$, there exists a density-$(1-\varepsilon)$ set $\mathbb{A}_\varepsilon \subseteq \mathbb{N}$ such that, for all $n \in \mathbb{A}_\varepsilon$, all $m \geq 2$ and all $\lambda > 0$, we have*

$$\textstyle\sum_{i \in [m]} i\, \mathbf{1}(i \wr n) \leq 40(\lambda\varepsilon)^{-1} m (\log m)^2.$$

Proof. Choose $N \in \mathbb{N}$ (large) and sample $n \sim \mathrm{Unif}(\{1, ..., N\})$; let $\varepsilon, \lambda \in (0, 1)$. We prove that

$$\mathbb{P}\big(\cap_{m \in [1,n]} \big\{ \textstyle\sum_{i \in [m]} i \mathbf{1}(i \wr n) \leq 20(\varepsilon\lambda)^{-1} m (\log m)^{1+\lambda} \big\}\big) \geq 1 - \varepsilon.$$

This implies the lemma. We have $\mathbb{P}(i \wr n) \leq 1/i$ for each $i \in [N]$. For $i \in [\lfloor \log_2 N \rfloor]$, defining

$$N_i \coloneqq \textstyle\sum_{j \in [2^{i-1}, 2^i - 1]} \mathbf{1}(j \wr n), \quad \text{we have} \quad \mathbb{E}(N_i) \leq \sum_{j=2^{i-1}}^{2^i-1} 1/j \leq 1.$$

By Markov's inequality, for any $\lambda, C > 0$, we then have

$$\mathbb{P}\big(N_i \geq C i^{1+\lambda}\big) \leq 1/(C i^{1+\lambda}).$$

Using the union bound, this gives

$$\mathbb{P}(\mathcal{E}) \geq 1 - 2\lambda^{-1}/C \quad \text{where} \quad \mathcal{E} \coloneqq \cap_{i \geq 1} \{N_i \leq C i^{1+\lambda}\}.$$

Set $r \coloneqq \lceil \log_2 m \rceil$; then $m \leq 2^r$. On the event $\mathcal{E}$, we then have

$$\textstyle\sum_{i \in [m]} i \mathbf{1}(i \wr n) \leq \sum_{i=1}^r 2^i N_i \leq C \sum_{i=1}^r 2^i i^{1+\lambda} \leq C r^{1+\lambda} 2^{r+1} \leq 20 C m (\log m)^{1+\lambda}.$$

Now let $\varepsilon \in (0, 1)$ and set $C \coloneqq 2/(\varepsilon\lambda)$, so then $\mathbb{P}(\mathcal{E}) \geq 1 - \varepsilon$. The result follows. □

Exactly the same argument can be used to show the following result.

Lemma 6.6.6. *For all $\varepsilon > 0$, there exists a density-$(1-\varepsilon)$ set $\mathbb{A}_\varepsilon \subseteq \mathbb{N}$ such that, for all $n \in \mathbb{A}_\varepsilon$, all $m \geq 2$ and all $\lambda > 0$, we have*

$$\textstyle\sum_{i \in [m]} \mathbf{1}(i \wr n) \leq 10(\lambda\varepsilon)^{-1} (\log m)^{2+\lambda}.$$

Bibliography

[1] D. Aldous and P. Diaconis (1985). Shuffling Cards and Stopping Times. *Technical Report 231, Department of Statistics, Stanford University*. *Available online*

[2] D. Aldous and J. A. Fill (2002). *Reversible Markov Chains and Random Walks on Graphs*. Unfinished Monograph *Available at* `stat.berkeley.edu/~aldous/RWG/book.html`

[3] N. Alon and Y. Roichman (1994). Random Cayley Graphs and Expanders. *Random Structures Algorithms*. **5**.2 (271–284) MR1262979 DOI

[4] G. Amir and O. Gurel-Gurevich (2010). The Diameter of a Random Cayley Graph of $\mathbb{Z}_q$. *Groups Complex. Cryptol.* **2**.1 (59–65) MR2672553 DOI

[5] R. Basu, J. Hermon and Y. Peres (2017). Characterization of Cutoff for Reversible Markov Chains. *Ann. Probab.* **45**.3 (1448–1487) MR3650406 DOI

[6] D. El-Baz and C. Pagano (2020). Diameters of Random Cayley Graphs of Finite Nilpotent Groups. *arXiv:2002.08870*

[7] C. Béguin, A. Valette and A. Zuk (1997). On the Spectrum of a Random Walk on the Discrete Heisenberg Group and the Norm of Harper's Operator. *J. Geom. Phys.* **21**.4 (337–356) MR 1436310 DOI

[8] A. Ben-Hamou, E. Lubetzky and Y. Peres (2018). Comparing Mixing Times on Sparse Random Graphs. *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, SIAM, Philadelphia, PA (1734–1740) MR3775901 DOI

[9] A. Ben-Hamou and J. Salez (2017). Cutoff for Nonbacktracking Random Walks on Sparse Random Graphs. *Ann. Probab.* **45**.3 (1752–1770) MR3650414 DOI

[10] I. Benjamini (2018). Private Communication.

[11] N. Berestycki, E. Lubetzky, Y. Peres and A. Sly (2018). Random Walks on the Random Graph. *Ann. Probab.* **46**.1 (456–490) MR3758735 DOI

[12] C. Bordenave, P. Caputo and J. Salez (2019). Cutoff at the "Entropic Time" for Sparse Markov Chains. *Probab. Theory Related Fields*. **173**.1-2 (261–292) MR3916108 DOI

[13] C. Bordenave and H. Lacoin (2018). Cutoff at the Entropic Time for Random Walks on Covered Expander Graphs. *arXiv:1812.06769*

[14] E. F. Breuillard (2004). Equidistribution of Random Walks on Nilpotent Lie Groups and Homogeneous Spaces. Thesis, ProQuest LLC, Ann Arbor, MI MR2705764

[15] E. F. Breuillard (2005). Local Limit Theorems and Equidistribution of Random Walks on the Heisenberg Group. *Geom. Funct. Anal.* **15**.1 (35–82) MR2140628 DOI

[16] D. Bump, P. Diaconis, A. Hicks, L. Miclo and H. Widom (2017). An Exercise(?) in Fourier Analysis on the Heisenberg Group. *Ann. Fac. Sci. Toulouse Math. (6)*. **26**.2 (263–288) MR 3640891 DOI

[17] S. Chen, C. Moore and A. Russell (2013). Small-Bias Sets for Nonabelian Groups: Derandomizations of the Alon–Roichman Theorem. *Approximation, Randomization, and Combinatorial Optimization*, Lecture Notes in Comput. Sci. Springer, Heidelberg **8096** (436–451) MR3126546 DOI

[18] D. Christofides and K. Markström (2008). Expansion Properties of Random Cayley Graphs and Vertex Transitive Graphs via Matrix Martingales. *Random Structures Algorithms*. **32**.1 (88–100) MR2371053 DOI

[19] F. Chung and L. Lu (2006). Concentration Inequalities and Martingale Inequalities: A Survey. *Internet Mathematics*. **3**.1 (79–127) MR2283885

[20] G. Conchon–Kerjan (2019). Cutoff for Random Lifts of Weighted Graphs. *arXiv:1908.02898*

[21] D. Coppersmith and I. Pak (2000). Random Walk on Upper Triangular Matrices Mixes Rapidly. *Probab. Theory Related Fields*. **117**.3 (407–417) MR1774070 DOI

[22] P. Diaconis and L. Saloff-Coste (1996). Logarithmic Sobolev Inequalities for Finite Markov Chains. *Ann. Appl. Probab.* **6**.3 (695–750) MR1410112 DOI

[23] P. Diaconis (1988). *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, CA **11** MR964069

[24] P. Diaconis (2010). Threads Through Group Theory. *Character Theory of Finite Groups*, Contemp. Math. Amer. Math. Soc., Providence, RI **524** (33–47) MR2731916 DOI

[25] P. Diaconis (2019). Private Communication.

[26] P. Diaconis and R. Hough (2015). Random Walk on Unipotent Matrix Groups. *arXiv:1512.06304*

[27] P. Diaconis and L. Saloff-Coste (1994). Moderate Growth and Random Walk on Finite Groups. *Geom. Funct. Anal.* **4**.1 (1–36) MR1254308 DOI

[28] P. Diaconis and L. Saloff-Coste (1995). An Application of Harnack Inequalities to Random Walk on Nilpotent Quotients. *Proceedings of the Conference in Honor of Jean-Pierre Kahane (Orsay, 1993)*, Journal of Fourier Analysis and Applications, (189–207) MR1364885

[29] P. Diaconis and L. Saloff-Coste (1995). An Application of Harnack Inequalities to Random Walk on Nilpotent Quotients. *Proceedings of the Conference in Honor of Jean-Pierre Kahane (Orsay, 1993)*, J. Fourier Anal. Appl. (189–207) MR1364885

[30] P. Diaconis and L. Saloff-Coste (1996). Nash Inequalities for Finite Markov Chains. *J. Theoret. Probab.* **9**.2 (459–510) MR1385408 DOI

[31] P. Diaconis and P. M. Wood (2013). Random Doubly Stochastic Tridiagonal Matrices. *Random Structures Algorithms*. **42**.4 (403–437) MR3068032 DOI

[32] J. Ding, E. Lubetzky and Y. Peres (2010). Total Variation Cutoff in Birth-and-Death Chains. *Probab. Theory Related Fields*. **146**.1-2 (61–85) MR2550359 DOI

[33] C. Dou (1992). Studies of Random Walks on Groups and Random Graphs. Thesis, Massachusetts Institute of Technology MR2716375

[34] C. Dou and M. Hildebrand (1996). Enumeration and Random Random Walks on Finite Groups. *Ann. Probab.* **24**.2 (987–1000) MR1404540 DOI

[35] R. Durrett (2010). *Probability: Theory and Examples*. Fourth ed., Cambridge University Press, Cambridge **31** MR2722836 DOI

[36] J. Ellenberg (1993). A Sharp Diameter Bound for Upper Triangular Matrices. Thesis, Harvard University

[37] J. S. Ellenberg and J. Tymoczko (2010). A Sharp Diameter Bound for Unipotent Groups of Classical Type Over $\mathbb{Z}/p\mathbb{Z}$. *Forum Math.* **22**.2 (327–347) MR2607568 DOI

[38] R. J. Evans, J. Boersma, N. M. Blachman and A. A. Jagers (1988). The Entropy of a Poisson Distribution: Problem 87-6. *SIAM Review*. **30**.2 (314–317) JSTOR2030812 DOI

[39] W. T. Gowers (2008). Quasirandom Groups. *Combin. Probab. Comput.* **17**.3 (363–387) MR2410393 DOI

[40] G. H. Hardy and E. M. Wright (2008). *An Introduction to the Theory of Numbers*. Sixth ed., Oxford University Press, Oxford MR2445243

[41] J. Hermon, H. Lacoin and Y. Peres (2016). Total Variation and Separation Cutoffs Are Not Equivalent and Neither One Implies the Other. *Electronic Journal of Probability*. **21** (Paper No. 44, 36 pp.) MR3530321 DOI

[42] J. Hermon and Y. Peres (2018). A Characterization of L_2 Mixing and Hypercontractivity via Hitting Times and Maximal Inequalities. *Probab. Theory Related Fields*. **170**.3-4 (769–800) MR 3773799 DOI

[43] J. Hermon, A. Sly and P. Sousi (2020). Universality of Cutoff for Graphs With an Added Random Matching. *arXiv:2008.08564*

[44] M. Hildebrand (1994). Random Walks Supported on Random Points of $\mathbb{Z}/n\mathbb{Z}$. *Probab. Theory Related Fields*. **100**.2 (191–203) MR1296428 DOI

[45] M. Hildebrand (2005). A Survey of Results on Random Random Walks on Finite Groups. *Probab. Surv.* **2** (33–63) MR2121795 DOI

[46] R. Hough (2017). Mixing and Cut-Off in Cycle Walks. *Electron. J. Probab.* **22** (Paper No. 90, 49 pp.) MR3718718 DOI

[47] Z. Landau and A. Russell (2004). Random Cayley Graphs Are Expanders: A Simple Proof of the Alon–Roichman Theorem. *Electron. J. Combin.* **11**.1 (Research Paper 62, 6 pp.) MR2097328 DOI

[48] G. F. Lawler and V. Limic (2010). *Random Walk: A Modern Introduction*. Cambridge University Press, Cambridge **123** MR2677157 DOI

[49] D. A. Levin, Y. Peres and E. L. Wilmer (2017). *Markov Chains and Mixing Times*. Second ed., American Mathematical Society, Providence, RI, USA MR3726904 DOI

[50] M. E. Lladser, P. Potočnik, J. Širáň and M. C. Wilson (2012). Random Cayley Digraphs of Diameter 2 and Given Degree. *Discrete Math. Theor. Comput. Sci.* **14**.2 (83–90) MR2992954

[51] P.-S. Loh and L. J. Schulman (2004). Improved Expansion of Random Cayley Graphs. *Discrete Math. Theor. Comput. Sci.* **6**.2 (523–528) MR2180056

[52] E. Lubetzky and Y. Peres (2016). Cutoff on All Ramanujan Graphs. *Geom. Funct. Anal.* **26**.4 (1190–1216) MR3558308 DOI

[53] E. Lubetzky and A. Sly (2010). Cutoff Phenomena for Random Walks on Random Regular Graphs. *Duke Math. J.* **153**.3 (475–510) MR2667423 DOI

[54] R. Lyons and Y. Peres (2016). *Probability on Trees and Networks*. Cambridge University Press, New York **42** MR3616205 DOI

[55] J. Marklof and A. Strömbergsson (2013). Diameters of Random Circulant Graphs. *Combinatorica*. **33**.4 (429–466) MR3133777 DOI

[56] C. McDiarmid (1998). Concentration. *Probabilistic Methods for Algorithmic Discrete Mathematics*, Algorithms and Combinatorics, Springer, Berlin, Heidelberg **16** (195–248) MR1678578 DOI

[57] R. Montenegro and P. Tetali (2006). Mathematical Aspects of Mixing Times in Markov Chains. *Found. Trends Theor. Comput. Sci.* **1**.3 (x+121) MR2341319 DOI

[58] A. Naor (2012). On the Banach-Space-Valued Azuma Inequality and Small-Set Isoperimetry of Alon–Roichman Graphs. *Combin. Probab. Comput.* **21**.4 (623–634) MR2942733 DOI

[59] E. Nestoridi (2019). Super-Character Theory and Comparison Arguments for a Random Walk on the Upper Triangular Matrices. *J. Algebra*. **521** (97–113) MR3884694 DOI

[60] E. Nestoridi and A. Sly (2020). The Random Walk on Upper Triangular Matrices over $\mathbb{Z}/m\mathbb{Z}$. *In preparation*

[61] I. Pak (1999). Random Cayley Graphs with $O(\log |G|)$ Generators Are Expanders. *Algorithms—ESA '99 (Prague)*, Lecture Notes in Comput. Sci. Springer, Berlin **1643** (521–526) MR1729149 DOI

[62] I. Pak (1999). Random Walks on Finite Groups with Few Random Generators. *Electron. J. Probab.* **4** (Paper No. 1, 11 pp.) MR1663526 DOI

[63] I. Pak (2000). Two Random Walks on Upper Triangular Matrices. *J. Theoret. Probab.* **13**.4 (1083–1100) MR1820503 DOI

[64] I. Pak (2001). Combinatorics, Probability, and Computations on Groups Lecture Notes. *Available at* www.math.ucla.edu/~pak/courses/pg.html

[65] Y. Peres (2004). American Institute of Mathematics Research Workshop "Sharp Thresholds for Mixing Times" (Palo Alto). *Summary available at* www.aimath.org/WWN/mixingtimes

[66] Y. Peres and A. Sly (2013). Mixing of the Upper Triangular Matrix Walk. *Probab. Theory Related Fields*. **156**.3-4 (581–591) MR3078280 DOI

[67] C. Pomerance (2001). The Expected Number of Random Elements to Generate a Finite Abelian Group. *Period. Math. Hungar.* **43**.1-2 (191–198) MR1830576 DOI

[68] S. Purkayastha (1998). Simple Proofs of Two Results on Convolutions of Unimodal Distributions. *Statistics & Probability Letters*. **39**.2 (97–100) MR1652520 DOI

[69] Y. Roichman (1996). On Random Random Walks. *Ann. Probab.* **24**.2 (1001–1011) MR1404541 DOI

[70] N. T. Sardari (2019). Diameter of Ramanujan Graphs and Random Cayley Graphs. *Combinatorica*. **39**.2 (427–446) MR3962908 DOI

[71] U. Shapira and R. Zuck (2019). Asymptotic Metric Behavior of Random Cayley Graphs of Finite Abelian Groups. *Combinatorica*. **39**.5 (1133–1148) MR4039604 DOI

[72] A. Smith (2017). The Cutoff Phenomenon for Random Birth and Death Chains. *Random Structures Algorithms*. **50**.2 (287–321) MR3607126 DOI

[73] R. Stong (1995). Eigenvalues of Random Walks on Groups. *Ann. Probab.* **23**.4 (1961–1981) MR1379176

[74] R. Stong (1995). Eigenvalues of the Natural Random Walk on the Burnside Group $B(3, n)$. *Ann. Probab.* **23**.4 (1950–1960) MR1379175

[75] R. Stong (1995). Random Walks on the Groups of Upper Triangular Matrices. *Ann. Probab.* **23**.4 (1939–1949) MR1379174

[76] X. Wang (2005). Volumes of Generalized Unit Balls. *Mathematics Magazine*. **78**.5 (390–395) JSTOR30044198 DOI

[77] D. B. Wilson (1997). Random Random Walks on $\mathbb{Z}_2^d$. *Probab. Theory Related Fields*. **108**.4 (441–457) MR1465637 DOI

[78] D. B. Wilson (2004). Mixing Times of Lozenge Tiling and Card Shuffling Markov Chains. *Ann. Appl. Probab.* **14**.1 (274–325) MR2023023 DOI

[79] M. Zack (1990). Measuring Randomness and Evaluating Random Number Generators Using the Finite Heisenberg Group. *Limit Theorems in Probability and Statistics (Pécs, 1989)*, Colloq. Math. Soc. János Bolyai, North-Holland, Amsterdam **57** (537–544) MR1116809