

Universality of Cutoff for Riffle Shuffling

Mark Sellke

Jialu Shi

Jiamin Wang

Abstract

A Gilbert–Shannon–Reeds (GSR) shuffle is performed on a deck of N cards by cutting the top $n \sim \text{Bin}(N, 1/2)$ cards and interleaving the two resulting piles uniformly at random. The celebrated “Seven shuffles suffice” theorem of Bayer and Diaconis in [BD92] established cutoff for this Markov chain: to leading order, total variation mixing occurs after precisely $\frac{3}{2} \log_2 N$ shuffles. Later work of Lalley [Lal00] and the first author [Sel22] extended this result to asymmetric binomial cuts $n \sim \text{Bin}(N, p)$ for all $p \in (0, 1)$. These results relied heavily on the binomial condition and many natural chains were left open, including uniformly random cuts and exact bisections.

We establish cutoff for riffle shuffles with general pile size distribution. Namely, suppose the cut sizes $(n^{(t)})_{t \geq 1}$ are IID and the convergence in distribution $n^{(t)}/N \xrightarrow{d} \mu$ holds for some probability measure μ on the interval $[0, 1]$. Then the mixing time t_{mix} satisfies $t_{\text{mix}}/\log N \rightarrow \bar{C}_\mu$ for an explicit constant $\bar{C}_\mu$. The same result holds for any (deterministic or random) sequence of pile sizes with empirical distribution converging to μ on all macroscopic time intervals (of length $\Omega(\log N)$). It also extends to multi-partite shuffles where the deck is cut into more than 2 piles in each step. Qualitatively, we find that the “cold spot” phenomenon identified by Lalley characterizes the mixing time of riffle shuffling in great generality.

1 Introduction

Riffle shuffling is among the most common methods to randomize a deck of cards. The best-known probabilistic model is the Gilbert–Shannon–Reeds (GSR) shuffle [Gil55, Ree81], which proceeds in the following two steps for an N -card deck:

- (I) Remove the top n cards from the top of the deck, where $n \sim \text{Bin}(N, 1/2)$ is a binomial variable.
- (II) Combine the two resulting piles via a uniformly random interleaving among all $\binom{N}{n}$ possibilities.

The mixing time of the GSR shuffle was pinpointed in [BD92] following [Ald83, Example 4.17]: $(\frac{3}{2} \log_2 N) \pm O(1)$ shuffles are necessary and sufficient for mixing in total variation. The analysis relies on the study of *rising sequences*, which yield a sufficient statistic for mixing when (and only when) $n \sim \text{Bin}(N, 1/2)$. Later [Lal00, Sel22] extended this result to obtain cutoff for *asymmetric* binomial cuts, where $n \sim \text{Bin}(N, p)$ for $p \neq 1/2$ and the interleaving step (II) above is unchanged. Although rising sequences cannot be used, the binomial distribution of the cuts remains crucial in the analysis.

These results left open many natural questions. For example:

- Q1 Does cutoff occur at the same time whenever the deck is cut “about in half” at each step? For example one may consider exact bisections $n = N/2$, or larger fluctuations $n = N/2 \pm o(N)$.
- Q2 Are the existence and location of cutoff robust to a small fraction of “erroneous” cuts?
- Q3 Does the cutoff phenomenon persist when n/N does not concentrate, e.g. $n \sim \text{Unif}(\{0, 1, 2, \dots, N\})$?

We answer all of these questions, showing that the delicate first condition (I) on the GSR shuffle is wholly unnecessary for the analysis of cutoff. Namely, let μ be a probability distribution on the interval $[0, 1]$ satisfying the non-degeneracy condition $\mu(\{0, 1\}) < 1$. Suppose that the (deterministic or random) sequence of cut sizes $(n^{(1)}, n^{(2)}, \dots, n^{(K)})$ is such that the time-parametrized empirical distribution

$$\text{p-lim}_{N \rightarrow \infty} \frac{1}{K} \sum_{1 \leq t \leq K} \delta_{(n^{(t)}/N, t/K)}$$

of cut sizes converges to a product measure $\mu \times \text{Leb}([0, 1])$ as $N \rightarrow \infty$ in probability. Then we show cutoff occurs after $\bar{C}_\mu \log N$ shuffles where $\bar{C}_\mu$ is an explicit μ -dependent positive constant. Namely $K \geq$

$\overline{C}_\mu(1+\varepsilon) \log N$ shuffles suffice for total variation mixing while $K \leq \overline{C}_\mu(1-\varepsilon) \log N$ do not. This encompasses all of the settings highlighted above. As explained in more detail in the next subsection, our result extends also to the multi-partite riffle shuffles introduced in [DFP92] where the deck is cut into k piles in each step.

1.1 Main results

As most of our results hold conditional on the pile sizes, we first consider riffle shuffles with deterministic pile sizes. To define a shuffle of a deck N with pile sizes $(n_0, \dots, n_{k-1})$, where $\sum_{0 \leq i \leq k-1} n_i = N$, one splits the N cards into k piles by taking the top n_0 cards off the top to form the first pile, the next n_1 cards to form the second pile, and so on. One then interleaves the piles by choosing one of the $\binom{N}{n_0, \dots, n_{k-1}}$ possible interleavings uniformly at random. Equivalently at each time, if the current remaining pile sizes are $A_0, \dots, A_{k-1}$, we drop a card from one of these piles independently from previous shuffles, such that the next card is dropped from pile i with probability

$$\frac{A_i}{A_0 + A_1 + \dots + A_{k-1}}.$$

We call the above process a *riffle shuffle with pile sizes* $\vec{n} = (n_0, \dots, n_{k-1})$, which defines a Markov transition matrix on the N -permutation group S_N :

$$\mathbf{P}(\vec{n}) : \text{Prob}(S_N) \rightarrow \text{Prob}(S_N),$$

where we write $\text{Prob}(S_N)$ as the space of probability measures on S_N .

We study the following shuffling process. Fix $K = K(N)$ the number of total shuffles, satisfying that $\lim_{N \rightarrow \infty} K(N) = \infty$, k the number of piles, and non-negative integers $\vec{n}^{(t)} = (n_0^{(t)}, \dots, n_{k-1}^{(t)})$ for each $1 \leq t \leq K$, with $\sum_{i=0}^{k-1} n_i^{(t)} = N$ for each t . Then at time t , we perform a shuffle with pile sizes $\vec{n}^{(t)} = (n_0^{(t)}, \dots, n_{k-1}^{(t)})$.

To deal with general shuffling processes, fix a probability distribution μ on the simplex

$$\mathcal{D}_k = \{(p_0, \dots, p_{k-1}) : 0 \leq p_i \leq 1, p_0 + \dots + p_{k-1} = 1\}, \quad (1.1)$$

with vertex set

$$V_k = \{(p_0, \dots, p_{k-1}) \in \mathcal{D}_k : p_i = 1 \text{ for some } 0 \leq i \leq k-1\}. \quad (1.2)$$

In a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we say a shuffling process with pile sizes $(\mathbf{X}^{(t)})_{t \in \mathbb{Z}^+}$ is **randomized- μ -like** if the following convergence in probability holds in the space of probability measures on $\mathbb{R}^{k+1}$:

$$\text{p-lim}_{N \rightarrow \infty} \frac{1}{K(N)} \sum_{1 \leq t \leq K} \delta_{(\mathbf{X}^{(t)})/N, t/K} = \mu \times \text{Leb}([0, 1]). \quad (1.3)$$

Here δ_x denotes a point mass at $x \in \mathcal{D}_k$.

For any positive integer K , the shuffle process with pile sizes $\vec{n} = (\vec{n}^{(t)})_{t \in \mathbb{Z}^+}$ until time K defines a Markov transition matrix on S_N :

$$\mathbf{P}(\vec{n}, K) : \text{Prob}(S_N) \rightarrow \text{Prob}(S_N).$$

Equivalently, the transition matrix is defined by

$$\mathbf{P}(\vec{n}, K) = \mathbf{P}(\vec{n}^{(K)}) \mathbf{P}(\vec{n}^{(K-1)}) \dots \mathbf{P}(\vec{n}^{(1)}).$$

Moreover, for shuffling process with randomized pile sizes $(\mathbf{X}^{(t)})_{t \in \mathbb{Z}^+}$, let $\mathbf{P}(\mathbf{X}, K) = \mathbb{E}^{\vec{n} \sim \mathbf{X}} \mathbf{P}(\vec{n}, K)$ denote the distribution of the resulting K -times shuffle deck, averaged over the randomness of both the pile sizes and the interleavings. (We emphasize that both $\mathbf{X}, K$ depend implicitly on N here.)

The stationary measure for these transition matrices is the uniform distribution on S_N , which we denote by ν_{Unif} . Our interest will be in the rate of convergence to ν_{Unif} . We establish cutoff for all μ -like riffle shuffles, at a universal time depending only on μ . To state the result, we recall some constants defined similarly for **p**-shuffles in [Lal00, Sel22]. For each $\mathbf{p} \in \mathcal{D}_k$, with arbitrary tie-breaking, set $i_{\max} = \arg \max_{i \in \{0, 1, \dots, k-1\}} (p_i)$ and $p_{\max} = p_{i_{\max}}$. Define the functions

$$\phi_{\mathbf{p}}(x) = \sum_{i=0}^{k-1} p_i^x, \quad \psi_{\mathbf{p}}(x) = -\log \phi_{\mathbf{p}}(x), \quad \psi_{\mu}(x) = \mathbb{E}_{\mu} \psi_{\mathbf{p}}(x). \quad (1.4)$$

For each probability measure μ on $\mathcal{D}_k$ with $\mu(V_k) < 1$, define the positive constant θ_μ by the identity $\psi_\mu(\theta_\mu) = 2\psi_\mu(2)$. Observe that $\phi_{\mathbf{p}}$ is strictly monotone, hence so is $\psi_{\mathbf{p}}$ and thus also ψ_μ ; in particular θ_μ is uniquely determined. Noting that $\psi_{\mathbf{p}}(3) \leq 2\psi_{\mathbf{p}}(2) \leq \psi_{\mathbf{p}}(4)$ holds for all $\mathbf{p} \in \mathcal{D}_k$ (see [Sel22, Proof of Proposition 1.1]), we have

$$\theta_\mu \in [3, 4). \quad (1.5)$$

Then we define

$$C_\mu = \frac{3 + \theta_\mu}{4\psi_\mu(2)} = \frac{3 + \theta_\mu}{2\psi_\mu(\theta_\mu)}, \quad \tilde{C}_\mu = \frac{1}{\mathbb{E}_\mu \log(1/p_{\max})}, \quad \bar{C}_\mu = \max(\tilde{C}_\mu, C_\mu). \quad (1.6)$$

We now present our first main result: randomized- μ -like shuffles undergo (total variation) cutoff after $\bar{C}_\mu \log N$ steps. Note the case that μ is a delta mass at $\mathbf{p}$ already includes multinomial shuffles $\mu_N^{(t)} \sim \text{Mult}(N, \mathbf{p})$, recovering in particular the main result of [Sel22].

Theorem 1.1. *Recalling (1.1) and (1.2), let μ be a probability measure on $\mathcal{D}_k$ with $\mu(V_k) < 1$. Then any sequence of randomized μ -like shuffles undergoes total variation cutoff after $\bar{C}_\mu \log N$ steps. Equivalently, for any randomized μ -like shuffle with pile sizes $\mathbf{X} = (\mathbf{X}^{(t)})_{t \in \mathbb{Z}^+}$ and any $\varepsilon > 0$,*

$$\lim_{N \rightarrow \infty} d_{\text{TV}}(\mathbf{P}(\mathbf{X}, \lfloor (1 - \varepsilon)\bar{C}_\mu \log(N) \rfloor), \nu_{\text{Unif}}) = 1, \quad (1.7)$$

$$\lim_{N \rightarrow \infty} d_{\text{TV}}(\mathbf{P}(\mathbf{X}, \lfloor (1 + \varepsilon)\bar{C}_\mu \log(N) \rfloor), \nu_{\text{Unif}}) = 0. \quad (1.8)$$

Remark 1.2. *In the degenerate case $\mu(V_k) = 1$, we may naturally define $\bar{C}_\mu = \infty$. Indeed as explained in Remark 6.2, in this case one has $t_{\text{mix}} \geq \omega(\log N)$ also from our proof of (1.7), i.e. for any fixed $C > 0$:*

$$\lim_{N \rightarrow \infty} d_{\text{TV}}(\mathbf{P}(\mathbf{X}, C \log N), \nu_{\text{Unif}}) = 1. \quad (1.9)$$

Remark 1.3. *A simple example of a randomized- μ -like shuffle process is to take $(\mathbf{X}^{(t)})_{t \in \mathbb{Z}^+}$ a sequence of IID random vectors $\mathbf{X}^{(t)} = (\mathbf{X}_0^{(t)}, \dots, \mathbf{X}_{k-1}^{(t)}) \in (\mathbb{Z}^+)^k$, with $\mathbf{X}_0^{(t)} + \dots + \mathbf{X}_{k-1}^{(t)} = N$ and $\mathbf{X}^{(t)} \sim \mu$. However the condition (1.3) is much more general and only requires homogeneity on macroscopic time intervals. For instance, this includes periodic cases where e.g. $\mathbf{n}^{(2t)} = (N/3, 2N/3)$ and $\mathbf{n}^{(2t+1)} = (N/2, N/2)$.*

For IID shuffles, an equivalent interpretation of Theorem 2.1 (and Remark 1.2) is as follows. Fix k and consider any sequence $(\mu_N)_{N \geq 1}$ of probability measures on k -way cuts of a size N deck. Then for any distance d metrizing the compactification $[0, \infty]$ of $\mathbb{R}_+$ and any $\varepsilon \in (0, 1)$, we have with the obvious notations:

$$\lim_{N \rightarrow \infty} d(t_{\text{mix}}(\varepsilon; N, \mu_N) / \log N, \bar{C}_{\mu_N}) = \lim_{N \rightarrow \infty} d(t_{\text{mix}}(1 - \varepsilon; N, \mu_N) / \log N, \bar{C}_{\mu_N}) = 0.$$

(Here as usual $t_{\text{mix}}(\varepsilon)$ is the number of shuffles needed to reach total variation distance ε from stationarity.)

We finally provide a counterexample. As motivation, it is natural to expect some convexity properties for the rate of mixing, under averaging transition kernels. For example, if one dealer performs μ -like shuffles and a second dealer performs μ' -like shuffles, then choosing a random dealer in each step yields a $(\mu + \mu')/2$ -like shuffler. At least for reversible chains, the spectral gap is concave under such averaging, which suggests that mixing should also be sped up. Surprisingly, this intuition can fail for (total variation mixing of) riffle shuffles: there exist pairs of dealers who shuffle less efficiently when working together!

Proposition 1.4. *There exists $k \in \mathbb{N}$ and $\mathbf{p}, \check{\mathbf{p}} \in \mathcal{D}_k$ such that $\mu = (\delta_{\mathbf{p}} + \delta_{\check{\mathbf{p}}})/2$ satisfies*

$$\bar{C}_\mu > \max(\bar{C}_{\mathbf{p}}, \bar{C}_{\check{\mathbf{p}}}).$$

The proof (deferred to the Appendix) characterizes the closure of all possible ψ_μ up to rescaling, extends the definition of $\bar{C}_\mu$ to this space, and finds an explicit counterexample within it. The value k resulting from our proof is presumably quite large. We also note that $\tilde{C}_\mu$ is trivially convex, and C_μ is convex on $\{\mu \in \mathcal{D}_k : \theta_\mu = \theta_*\}$ for any fixed k, θ_* , both of which may make Proposition 1.4 more surprising.

1.2 Consequences and Discussion

As previously mentioned, the original GSR shuffle analysis [BD92] relies on the somewhat brittle notion of rising sequences. Our results are built instead on the cold spots picture introduced by Lalley [Lal00] and also used in [Sel22]. Thus a conceptual message of our work is that cold spots give a robust and sharp characterization of the mixing time for general riffle shuffles.

As discussed in [DG11, End of Chapter 8] and [Tod14, Pages 200, 465], for dividing a set of N items into parts of size $(n, N - n)$, the question of the “most natural” distribution of n has been debated for centuries and traces back to the origins of probability. Some may argue that the “most natural” way is to put each item to each part independently with probability $1/2$, which leads to the binomial distribution $\text{Bin}(N, 1/2)$. On the other hand, some may believe that the size should be distributed uniformly on $[0, N] \cap \mathbb{Z}$.

Our result is the first to handle the natural case of uniformly random cuts. It is straightforward that the corresponding shuffling process is randomized- μ -like for $\mu = \text{Unif}([0, 1])$. Using (1.6) and direct calculation, we find that $\bar{C}_\mu \approx 3.606$ in this case (yielding the approximation $\bar{C}_\mu \log(52) \approx 14.25$ for a real deck). Numerics for other natural choices of μ are given in Table 1.

Measure μ on $\mathcal{D}_k$	k	θ_μ	$\psi_\mu(2)$	C_μ	$\tilde{C}_\mu$	$\bar{C}_\mu$
$k = 2$ Beta(1, 1) (Uniform on $\mathcal{D}_2$)	2	3.197	0.430	3.606	3.256	3.606
$k = 2$ Beta(1/2, 1/2)	2	3.237	0.316	4.932	4.554	4.932
$k = 2$ Beta(2, 2)	2	3.149	0.525	2.926	2.562	2.926
$k = 2$ Beta(2, 16)	2	3.812	0.207	8.230	8.256	8.256
$k = 2$ Uniform on $\{(p, 1 - p) : p \in [1/4, 3/4]\}$	2	3.121	0.613	2.496	2.097	2.496
$k = 3$ Dirichlet(1, 1, 1) (Uniform on $\mathcal{D}_3$)	3	3.190	0.724	2.138	1.927	2.138
$k = 3$ Dirichlet(1/2, 1/2, 1/2)	3	3.232	0.551	2.830	2.607	2.830
$k = 3$ Dirichlet(2, 2, 2)	3	3.141	0.863	1.779	1.552	1.779
$k = 4$ Dirichlet(1, 1, 1, 1) (Uniform on $\mathcal{D}_4$)	4	3.183	0.948	1.631	1.465	1.631
Delta-mass on $(1/k, \dots, 1/k)$	k	3	$\log k$	$\frac{3}{2 \log k}$	$\frac{1}{\log k}$	$\frac{3}{2 \log k}$

Table 1: Numerical estimates of θ_μ , C_μ , $\tilde{C}_\mu$, and $\bar{C}_\mu$ for several Beta distributions and Dirichlet distributions. The last line $\mu = \delta_{(1/k, \dots, 1/k)}$ includes the symmetric multinomial k -shuffles analyzed in [BD92].

In fact, among all measures on $\mathcal{D}_k$, by definition of θ_μ and (1.5), we see that θ_μ reaches its minimum when μ is a δ -mass on $(1/k, \dots, 1/k)$. On the other hand, by Cauchy’s inequality we see that $\phi_{\mathbf{p}}(2)$ reaches its minimum, and thus $\psi_{\mathbf{p}}(2)$ reaches its maximum at $\mathbf{p} = (1/k, \dots, 1/k)$. Therefore, we conclude that C_μ reaches its minimum when μ is a δ -mass on $(1/k, \dots, 1/k)$. It is also straightforward that $\tilde{C}_\mu$ is minimized in this case. In other words, no distribution of cuts mixes asymptotically faster than GSR shuffles.

Other riffle shuffling chains have also been studied and would be very interesting to understand better. In light of our work, the main remaining challenge is to modify the uniform interleaving condition (II) of the GSR shuffle. An especially natural model is the “clumpy” shuffle where cards are more likely to drop from the same pile as the previous card. Polylogarithmic mixing for this chain was shown in [JM15]. Another model is the Thorp shuffle where one cuts the deck exactly in half and ensures the m -th cards from each pile end up adjacent to each other, yielding $2^{N/2}$ possible permutations from a single shuffle [Tho73, Mor08, Mor09, Mor13]. Cutoff is not known for any version of these chains. Many more results and problems on card shuffling can be found in [Dia03, DF23].

1.3 Proof Techniques and Challenges

The proof of Theorem 1.1 naturally splits into lower and upper bounds. We discuss these separately; the lower bound is easier but more clearly illustrates the underlying mechanism.

Lower Bound via Cold Spot Hypothesis Testing The mixing time lower bound requires pinpointing the fundamental obstruction to mixing. This was discovered in the multinomial case $\mu = \delta_{\mathbf{p}}$ by Lalley [Lal00], who identified a fractal set of *cold spots* within $[N]$ where the cards maintain an unusually strong correlation with their initial ordering. (More precisely, cold spots lead to the lower bound C_μ in (1.6), which is the more interesting obstruction.) To describe cold spots, we must identify each card with a sequence in $\{0, 1, \dots, k - 1\}^K$ of its piles in each shuffle. As described further in Subsection 2.2, the *inverse* permutation

of a riffle shuffle can be described by sorting the N such strings in lexicographic order, yielding a sequence of strings $s_1 \leq_{\text{lex}} s_2 \leq_{\text{lex}} \dots \leq_{\text{lex}} s_N$ in $[k]_0^K$. Equal adjacent strings $s_j = s_{j+1}$ then remain in order (by virtue of being in the same pile during all K shuffles), yielding extra ascents in the inverse permutation. The main contribution turns out to come from strings containing each digit j with frequency proportional to $p_j^{\theta_\mu}$ within their initial segment. The lengths of these initial segments are chosen to have approximately deterministic location, which are called cold spots. By counting ascents within the cold spots, [Lal00] constructed an explicit hypothesis test to detect under-shuffled decks and thus prove the sharp mixing time lower bound (for $\mathbf{p} \approx (1/k, \dots, 1/k)$ sufficiently close to uniform, a restriction removed in [Sel22] via truncation).

To extend this lower bound, we first approximate the measure μ by a discrete distribution (namely one uses $\sum_{0 \leq i \leq z} \mu(D_i) \delta_{\mathbf{p}^{(i)}}$ to approximate μ , where the D_i are as in (2.1)). Then we focus on strings containing each digit j among the “ $\mathbf{p}^{(i)}$ -like shuffles” with frequency proportional to $(p_j^{(i)})^{\theta_\mu}$ in an initial segment; cold spots are the (deterministic) typical locations of such strings (see (6.7)). If the pile sizes are random with fluctuations of order $O(N^{1/2+o(1)})$ as in the binomial case, then the same argument works with only minor changes. To handle larger sublinear fluctuations of pile sizes in the randomized case, we construct a many-to-one hypothesis test by exhausting all (suitably nicely behaved) sequences of pile sizes. For such a pile size sequence $(n_l^{(t)})_{1 \leq t \leq K, 0 \leq l \leq k-1}$, we test whether the deck appears more like it was shuffled according to such a sequence, or generated uniformly at random. By ensuring that each individual test has very small e^{-N^c} probability to incorrectly detect non-uniformity (see Lemma 6.6), we find that the resulting many-to-one test handles the random case, extending the lower bound to pile sizes with general distributions.

Upper Bound via Truncated Exponential Moments For the more difficult upper bound, it suffices to consider deterministic $\vec{n}$. Identifying each card with a string in $\{0, 1, \dots, k-1\}^K$ as described above, we consider the random subgraph of the N -vertex path graph by connecting $(j, j+1)$ when $s_j = s_{j+1}$. Via a truncated χ^2 argument, upper-bounding the total variation distance is reduced to upper-bounding the exponential moment of the number of shared edges between two IID such graphs by $1+o(1)$. (I.e. the number of shared edges should be stochastically dominated by a geometric random variable with mean $o(1)$.) This strategy was developed in [Sel22] for $\text{Mult}(N, \mathbf{p})$ shuffles, which relied on convenient properties of multinomial distributions in many places, including concentration, anti-concentration, and simple behavior of conditional laws. Even the case $\mu = \delta_{\mathbf{p}}$ requires significant new work, due to the much more complicated conditional laws that arise within the analysis and the presence of $o(\log N)$ non- $\mathbf{p}$ -like shuffles.

We first show the first moment is $N^{-\Omega(1)}$, assuming that the first shuffle is “good” in that it has at least 2 piles with $\Omega(N)$ cards. To achieve this, we divide the shuffle graph into carefully chosen blocks in Lemma 4.9 (corresponding to different possible prefix strings in $[k]_0^M$ for varying M) and argue that the edges are distributed “approximately uniformly” within each block. There are several key challenges.

First, understanding the fluctuation for the location of the strings with a given prefix (after conditioning) is an important part of the first moment proof. (See Section 4.1.) In [Sel22], the order of fluctuation for the location can be determined by initial consecutive 0-digits or 1-digits. In other words, by comparing the expected numbers of strings smaller and larger than this prefix, the fluctuation is given by the square root of the minimum of these two quantities. However, in our case, since we allow some shuffles to put $o(1)$ or $1-o(1)$ fraction of cards into certain piles, it is not immediately clear whether the fluctuation is determined by the left or the right endpoint. To adapt the proof, we carefully choose the information to reveal within the conditioning arguments to apply anti-concentration for hypergeometric variables (see Lemmas 4.5 and 4.6). By a delicate prefix decomposition, we construct the relevant hypergeometric variable $\text{Hyp}(n, m, n_1)$ to satisfy $\max(m, n_1)/n \leq 1 - \Omega(1/\log N)$. The fluctuation of this hypergeometric variable is of order the square root of its mean, up to this additional $\log N$ factor (which turns out to not affect the leading order mixing time).

Second, in [Sel22] each block is defined by all strings with a certain prefix, such that its location exhibits fluctuations of approximately the same order as its size, which can be done by recursively partitioning until a simple stopping criterion is met. Then one combines those blocks into classes according to the number of each digits, and proves that the expected number of edges is N^{-c} in each of a total of $O(\log N)$ classes. This becomes significantly more challenging in our case, especially because we allow some shuffles to put $o(1)$ fraction of cards into certain piles. For example, two blocks may behave very differently if one places an $N^{-1/2}$ fraction of cards in a certain pile, while the other places an $N^{-3/4}$ fraction. To handle this subtlety, we first define a set of prefixes such that the order of fluctuations is fixed once a string has a prefix in this set (see definition 4.3). After truncating very rare prefixes, there are at most $N^{c'}$ admissible prefixes for some sufficiently small c' . We then further decompose each prefix block into sub-blocks with fluctuations slightly

larger than its size (see lemma 4.9). We then group the sub-blocks according to both the counts of each digit and the precise pattern of “small pile” digit occurrences (definition 4.7), and union bound over the groups.

Having established the corresponding first moment estimate, we estimate the exponential moment by exploring a pair of IID shuffle graphs on $[N]$ one-by-one from each direction (both forward and backward). First by some shifting tricks and monotonicity of TV-distance under extra shuffles, we may assume some initial shuffles are “good” in that they have no extremely small piles (see lemma 3.2). The goal now is to argue that the first moment estimate applies conditionally to the unrevealed shared edges and conclude using a form of Portenko’s lemma. Since the pile sizes for the first shuffle are fixed, the shuffle graph naturally divides into k parts, corresponding to the k piles in the first shuffle. Therefore, instead of considering two exploration processes as in [Sel22] (forward and backward), we analyze $2k$ processes in Lemma 3.9, one in each direction within each part. Analyzing the exponential moment now reduces to analyzing the expected number of unrevealed edges, conditionally on the subgraph we have explored. To estimate this conditional quantity, we use a simple decomposition of the possible unrevealed strings into $O(\log N)$ disjoint prefix blocks, and then refine this decomposition further (see (5.5)) so that each prefix block is immediately followed by a “good” shuffle. The number of resulting blocks is still $N^{o(1)}$, so we can afford to apply the first moment bound to each separately and union bound (the $o(1)$ exponent can be arbitrarily slow depending on the convergence rate in (1.3)).

For “large” blocks, we apply the first-moment estimate in Proposition 4.2. Namely by carefully excluding specific events in each exploration step, we find that conditioning on the explored subgraph, the “sub-shuffle” in each block is still “ μ -like”. Thanks to concentration results in Lemma 4.4, the probability of each excluded event is at most e^{-N^c} . However, for “small” blocks, the concentration result does not apply, so that we may no longer use the first moment estimate. Instead, we bound the number of unrevealed edges directly by block size. The main challenge now comes from the more complicated conditional distributions of the unrevealed strings. Here we use an explicit combinatorial characterization of shuffling process introduced in Section 2.2, in which the conditional distribution is uniform on a specific set of matrices. Using this fact, we may count the number of those matrices in Section 5.4 and therefore give an estimate for the conditional distribution.

2 Preliminaries

In this section, we introduce notation and useful results. In Section 2.1, we reduce the original problem to studying a shuffling process with deterministic pile sizes. In Section 2.2, we define the shuffle graph and reformulate the problem into a more manageable form. Then, in Section 2.3, we establish some useful concentration inequalities. Finally, we conclude this section by detailing additional notation.

2.1 Reduction to deterministic pile sizes

To prove Theorem 1.1, we deal with shuffling processes with deterministic pile sizes. To this end, we define a class of deterministic shuffling processes that represent the “typical” behavior of a randomized- μ -like shuffling process. Recall the definition of $\mathcal{D}_k$ in (1.1) and recall that $V_k \subset \mathcal{D}_k$ is the vertex set of the domain $\mathcal{D}_k$. For convenience, we write $V_k = \{v_0, \dots, v_{k-1}\}$. For each $\chi > 0$, consider a finite decomposition

$$\mathcal{D}_k = D_0 \cup D_1 \cup \dots \cup D_z \tag{2.1}$$

of $\mathcal{D}_k$ into parts with disjoint interiors such that the following holds:

1. $z \leq O(\chi^{-2k})$;
2. $D_i = \{\mathbf{p} : d(\mathbf{p}, v_i) < \chi\}$ for $0 \leq i \leq k-1$;
3. D_i is a simplex for $k \leq i \leq z$;
4. For all $k \leq i \leq z$, it holds that $\text{diam}(D_i) < \chi^2$.

Here d and diam denote distance and diameter in $\mathcal{D}_k$ associated with the L^∞ norm on $\mathcal{D}_k \subseteq \mathbb{R}^k$. For convenience, we write $\tilde{V}_k = \cup_{0 \leq i \leq k-1} D_i$ which is a neighborhood of V_k . For a construction obeying these properties, note first that since $\mathcal{D}_k \setminus \tilde{V}_k$ is a convex polytope, it can be triangulated into finitely many interior-disjoint simplices on the same vertex set (see e.g. [LS17]). These simplices can then be further triangulated into subsimplices with arbitrarily small diameter using iterated barycentric subdivision.

We say a shuffle $\mathbf{n}^{(t)}$ is χ -bad if $\mathbf{n}^{(t)}/N \in \tilde{V}_k$; otherwise we call it χ -good. We say the shuffle process with pile sizes $(\tilde{\mathbf{n}}^{(t)})_{t \leq K}$ is (χ, ρ, φ) -**almost- μ -like** if there exists¹ $0 \leq t_* < \rho \log N$, such that for every $0 \leq i \leq \lceil K/\rho \log N \rceil$,

$$\left| \frac{1}{\rho \log N} |\{t_* + (i-1)\rho \log N < t \leq t_* + i\rho \log N : \tilde{\mathbf{n}}^{(t)}/N \in D_i\}| - \mu(D_i) \right| < \varphi. \quad (2.2)$$

The next theorem shows deterministic almost- μ -like shuffles undergo cutoff after $\bar{C}_\mu \log N$ steps. We state it separately from Theorem 1.1 because we will focus on the deterministic case first, and because dependence on the explicit parameters (χ, ρ, φ) will be tracked carefully in the proof.

Theorem 2.1. *For fixed μ such that $\mu(V_k) < 1$ and for fixed ε , there exists $\chi = \chi(\mu, \varepsilon), \rho = \rho(\mu, \varepsilon), \varphi = \varphi(\mu, \varepsilon) > 0$ such that the following holds. For any sequence of (χ, ρ, φ) -almost- μ -like shuffles with deterministic pile sizes $\mathbf{n}^{(t)} = (n_0^{(t)}, \dots, n_{k-1}^{(t)})$,*

$$\lim_{N \rightarrow \infty} d_{\text{TV}}(\mathbf{P}(\tilde{\mathbf{n}}, \lfloor (1 - \varepsilon)\bar{C}_\mu \log(N) \rfloor), \nu_{\text{Unif}}) = 1, \quad (2.3)$$

$$\lim_{N \rightarrow \infty} d_{\text{TV}}(\mathbf{P}(\tilde{\mathbf{n}}, \lfloor (1 + \varepsilon)\bar{C}_\mu \log(N) \rfloor), \nu_{\text{Unif}}) = 0, \quad (2.4)$$

where ν_{Unif} is the uniform measure on S_N and the limits are uniform in $\tilde{\mathbf{n}}$.

2.2 Shuffle graph

As in [Lal00, Sel22], it will be convenient to work with the inverse permutations of riffle shuffles, which admit a simpler description we recall now. Let $[N] = \{1, 2, \dots, N\}$ and $[k]_0 = \{0, 1, \dots, k-1\}$. For $M > 0$, we write $[k]_0^M$ as the set of strings with digits in $[k]_0$ and with length M . We then define a $\tilde{\mathbf{n}}$ -dependent distribution on length N sequences of length K strings

$$S_K = (s_1, \dots, s_N) \in ([k]_0^K)^N.$$

Namely, we consider all $N \times K$ matrices $S \in ([k]_0)^{N \times [K]}$ with entries in $[k]_0$ such that for all $t \in [K]$ and $l \in [k]_0$,

$$n_l^{(t)} = |\{i \in [N] : S[i, t] = l\}|$$

is exactly the number of digits l appearing in the t -th column. Here we write $S[i, t]$ for the entry of S in row i and column t . We call such matrices *shuffle matrices* associated with $\tilde{\mathbf{n}}$. Choosing such a matrix uniformly at random, each row of said matrix forms a string with length K , and we write $\bar{S}_K = \{\bar{s}_1, \dots, \bar{s}_N\}$ as the multi-set of these strings. Then S_K is obtained by sorting $\bar{S}_K$ into lexicographically increasing order

$$s_1 \leq_{\text{lex}} s_2 \leq_{\text{lex}} \dots \leq_{\text{lex}} s_N.$$

Here $x <_{\text{lex}} y$ indicates that x has a smaller digit at the first entry where x and y differ. We often omit the subscript K for simplicity, writing S for S_K . (We note that if $\tilde{\mathbf{n}}^{(t)}$ are IID multinomial as in [Lal00, Sel22], then the NK digits of $\bar{S}_K$ are IID which is a major simplification.)

Next we generate the *shuffle graph* $G = G(S)$. The vertex set is defined by

$$[N] = \{1, 2, \dots, N\},$$

and the edge set is defined by

$$\{(i, i+1) : s_i = s_{i+1}, i = 1, 2, \dots, N-1\}. \quad (2.5)$$

Thus $G(S)$ is always a subgraph of the path graph on N vertices. We call a connected component in G a G -component. For $\sigma \in S_N$ and a shuffle graph G , define σ^G by within each G -component, sorting the values $\sigma(i)$ into increasing order.

¹We point out the reason for defining such t_* rather than defining $t_* = 0$. In Sections 3.1 and 3.4, we will consider new shuffling processes which start at some time $\bar{t}$ between 1 and K . The above definition conveniently ensures that these new shuffling processes remain (χ, ρ, φ) -almost- μ -like.

Proposition 2.2. *Suppose that ν_{Unif} is the uniform distribution on $\mathcal{S}_N$, $\pi \sim \nu_{\text{Unif}}$, and G is the shuffle graph generated as above. Then we have*

$$d_{\text{TV}}(\pi^G, \pi) = d_{\text{TV}}(\Pi_{1 \leq i \leq K} \mathbf{P}(n_0^{(i)}, \dots, n_{k-1}^{(i)}), \nu_{\text{Unif}}).$$

Proof. For all $i = 1, 2, \dots, N$, let $\tilde{s}_i$ denote a $[k]_0$ -valued string of length K , such that $\tilde{s}_i[j] = l \in [k]_0$ if and only if card i belongs to pile l in the j -th shuffle. Consider the shuffle graph $G(\tilde{S})$ defined as in (2.5) with $\tilde{S} = (\tilde{s}_1, \dots, \tilde{s}_N)$. Now we sample $X_K \sim \Pi_{t \leq K} \mathbf{P}(n_0^{(i)}, \dots, n_{k-1}^{(i)})$. Direct calculation shows that, conditioned on $\tilde{S}$, the inverse permutation X_K^{-1} is distributed uniformly on the set

$$\{\sigma \in \mathcal{S}_N : \sigma = \sigma^{G(\tilde{S})}\}.$$

(See [Lal00, Lemma 3] for more discussion of this last assertion.)

It suffices to show that $S_K \stackrel{d}{=} \tilde{S}_K$. We prove this by induction on K . The case when $K = 1$ is obvious. Suppose that we have proved the claim for K . With the discussion in the above paragraph, to determine the inverse permutation of the deck, we first generate $\pi \in \mathcal{S}_N$ uniformly, then sort them in order in each $G(\tilde{S}_K)$ -component, and write j at the $(K+1)$ -th digits of s_i if

$$\pi^{G(\tilde{S}_K)}(i) \in \left[\sum_{k < j} n_k^{(K+1)} + 1, \sum_{k \leq j} n_k^{(K+1)} \right].$$

This process is equivalent to sampling a uniform vector with $n_j^{(K+1)}$ digits j (for all $j = 0, 1, \dots, k-1$), and sorting the vector in lexicographic order within each $G(\tilde{S}_K)$ -component. (Indeed generating a $N \times (K+1)$ matrix and then sorting in lexicographic order is equivalent to first generating a $N \times K$ matrix with a fixed number $n^{(i)}$ of j digits in column i and sorting it in lexicographic order, then generating column $K+1$ and sorting it within the rows with same first K digits.) Combined with the inductive hypothesis that $S_K \stackrel{d}{=} \tilde{S}_K$, we deduce that $S_{K+1} \stackrel{d}{=} \tilde{S}_{K+1}$. $\square$

2.3 Concentration results

In this subsection, we demonstrate concentration results for Y_x , the number of strings with prefix x in S_K . Since we work with fixed pile sizes most of the time, many relevant quantities will be hypergeometric random variables. Let $\text{Hyp}(n_1, n, m)$ be the distribution of the intersection size between two independent uniformly random subsets of $[n]$ with cardinalities n_1 and m . The following very useful lemma shows concentration for such distributions.

Lemma 2.3 ([HS05, Theorem 1]). *Let $X \sim \text{Hyp}(n_1, n, m)$, for all $x > 2$ we have*

$$\mathbb{P}(|X - \mathbb{E}X| \geq x) \leq \exp(-2\alpha_{n_1, n, m}(x^2 - 1)),$$

where

$$\alpha_{n_1, n, m} = \max \left(\frac{1}{n - n_1 + 1} + \frac{1}{n_1 + 1}, \frac{1}{n - m + 1} + \frac{1}{m + 1} \right).$$

We give stronger versions of the concentration for hypergeometric distribution.

Lemma 2.4. *Let $X \sim \text{Hyp}(n_1, n, m)$ with $\mathbb{E}X = \frac{n_1 m}{n} \geq 10$. There exists absolute constants C, c such that for all $\delta < 1/2$,*

$$\mathbb{P}(|X - \mathbb{E}X| \geq (\mathbb{E}X)^{1/2+\delta}) \leq C \exp(-c(\mathbb{E}X)^{2\delta}).$$

Proof. We follow the proof of [HS05, Theorem 1]. Write $p(k) = \mathbb{P}(X = k)$ and

$$q(k) \equiv p(k+1)/p(k) = \frac{(m-k)(n_1-k)}{(k+1)(n-n_1-m+k+1)}.$$

It is clear that $q(k)$ is decreasing in k . Denote $d(k) = \sum_{i \geq k} p(i)$, thus we have

$$d(k+1) = \sum_{i \geq k} p(i+1) = \sum_{i \geq k} q(i)p(i) \leq q(k) \sum_{i \geq k} p(i) = q(k)d(k).$$

It suffices to bound $q(k)$ to obtain a decay rate for $d(k)$. We extend q to be a function on $\mathbb{R}^+$ by defining

$$q(x) = \frac{(m-x)(n_1-x)}{(x+1)(n-n_1-m+x+1)}.$$

Taking the log derivative, we find

$$\frac{d}{dx} \log q(x) = -\frac{1}{m-x} - \frac{1}{n_1-x} - \frac{1}{x+1} - \frac{1}{n-n_1-m+x+1} < -\frac{1}{x}.$$

Writing $\kappa = (n_1m + m + n_1 - 1 - n)/(n+2)$, we have $q(\kappa) = 1$. For $x \in [\kappa, 2\kappa]$, we then have

$$\frac{d}{dx} \log q(x) \leq -\frac{1}{2\kappa} \text{ and } \log q(x) \leq -\frac{x-\kappa}{2\kappa},$$

which leads to the bound

$$\log d(k) \leq \log d(\lceil \kappa \rceil) + \sum_{i=\lceil \kappa \rceil}^{k-1} \log q(i) \leq -\sum_{i=\lceil \kappa \rceil}^{k-1} \frac{i-\kappa}{2\kappa} < -\frac{1}{2\kappa}((k-\kappa-1)^2 - 1).$$

Notice that $|\mathbb{E}X - \kappa| \leq 1$, we then obtain the upper bound

$$\mathbb{P}(X - \mathbb{E}X \geq (\mathbb{E}X)^{1/2+\delta}) \leq C \exp(-c(\mathbb{E}X)^{2\delta}).$$

By a similar argument we then obtain the lower bound, which then completes the proof. $\square$

Lemma 2.5. *Let $\delta \in (0, 1/2)$, $C > 0$, $X \sim \text{Hyp}(n_1, n, m)$ with $\mathbb{E}X = \frac{n_1m}{n} \leq CN^\delta$. There exists constants C, c depending on C such that*

$$\mathbb{P}(X \geq 2a) \leq C \exp(-ca) \text{ for all } a \geq 3CN^q. \quad (2.6)$$

Proof. By the calculation in the previous lemma, we see that $\log q(x) \leq -\log \frac{x}{\kappa}$. We have $q(a) \geq 2$ for $a \geq 2N^\delta$. Thus we have $\log d(a) \leq -(a - 2CN^\delta) \log 2 \geq ca$ for $a \geq 3CN^\delta$, which completes the proof. $\square$

Lemmas 2.4 and 2.5 give concentration for hypergeometric distributions in two different regimes. Lemma 2.4 states that if a hypergeometric variable has a macroscopic mean, it will concentrate around its mean with high probability. Lemma 2.5, on the other hand, states that a hypergeometric variable with a small mean is unlikely to take large values. With these two lemmas as inputs, we can deduce the following concentration results for the number of strings with fixed prefix.

Lemma 2.6. *For $K = O(\log N)$, let Y_x denote the number of strings with prefix x in S_K . Let T_q be the set of all strings x such that $\mathbb{E}Y_x \geq N^q$. Then there exist constants C, c such that for all $a > 0$,*

$$\mathbb{P}(|Y_x - \mathbb{E}Y_x| \leq CK(\mathbb{E}Y_x)^{1/2+a}) \geq 1 - C \exp(-c(\mathbb{E}Y_x)^{2a}). \quad (2.7)$$

Moreover,

$$\mathbb{P}(|Y_x - \mathbb{E}Y_x| \leq CK(\mathbb{E}Y_x)^{1/2+a}, \text{ for all } x \in T_q) \geq 1 - C \exp(-cN^{2qa}).$$

Proof. For a string $x = x[1] \cdots x[M]$, we define $x_t = x[1] \cdots x[t]$ for $1 \leq t \leq M$. We have immediately that condition on Y_{x_t} , $Y_{x_{t+1}} \sim \text{Hyp}(Y_{x_t}, N, n_{x_{t+1}}^{(t+1)})$. Suppose that $Y_{x_t} \geq \frac{1}{2} \mathbb{E}Y_{x_t}$, and we have

$$\mathbb{E}[Y_{x_{t+1}} | Y_{x_t}] = Y_{x_t} \cdot \frac{n_{x_{t+1}}^{(t+1)}}{N} \geq \frac{1}{2} \mathbb{E}Y_{x_{t+1}} \geq 10.$$

Therefore by Lemma 2.4, we obtain that for some positive absolute constant C, c ,

$$\begin{aligned} \mathbb{P}\left(|Y_{x_{t+1}} - \mathbb{E}[Y_{x_{t+1}} | Y_{x_t}]| \geq (\mathbb{E}[Y_{x_{t+1}} | Y_{x_t}])^{1/2+a} \mid Y_{x_t}\right) &\leq C \exp\left(-c(\mathbb{E}[Y_{x_{t+1}} | Y_{x_t}])^{2a}\right) \\ &\leq \exp\left(-c(\mathbb{E}Y_{x_{t+1}})^{2a}\right). \end{aligned}$$

Thus conditionally on Y_{x_t} , with probability at least $1 - \exp(-c(\mathbb{E}Y_{x_{t+1}})^{2a})$,

$$Y_{x_{t+1}} \in \left[Y_{x_t} \cdot \frac{n_{x[t+1]}^{(t+1)}}{N} (1 - c(\mathbb{E}Y_{x_{t+1}})^{-1/2+a}), Y_{x_t} \cdot \frac{n_{x[t+1]}^{(t+1)}}{N} (1 + c(\mathbb{E}Y_{x_{t+1}})^{-1/2+a}) \right].$$

By an induction on t , we deduce that with probability at least $1 - (t+1)\exp(-c(\mathbb{E}Y_{x_{t+1}})^a)$, the following holds:

$$Y_{x_{t+1}} \in \left[\mathbb{E}[Y_{x_{t+1}}] \cdot \prod_{i \leq t+1} (1 - c(\mathbb{E}Y_{x_i})^{-1/2+a}), \mathbb{E}[Y_{x_{t+1}}] \cdot \prod_{i \leq t+1} (1 + c(\mathbb{E}Y_{x_i})^{-1/2+a}) \right].$$

In particular, since $\mathbb{E}Y_{x_i} \geq \mathbb{E}Y_x \geq N^q$ and $t+1 \leq K = O(\log N)$, we have $Y_{x_{t+1}} \geq \frac{1}{2} \mathbb{E}Y_{x_{t+1}}$. Thus the induction is sound. The proof is then completed by a union bound over x . (Notice that the factor accounting for the number of possible x can be absorbed by adjusting c .) $\square$

Similarly, one may use Lemma 2.5 to obtain the following lemma.

Lemma 2.7. *For $K = O(\log N)$, let Y_x denote the number of strings with prefix x in S_K . If $\mathbb{E}Y_x \leq N^q$, then*

$$\mathbb{P}(Y_x \geq 3N^q) \leq \exp(-cN^q).$$

Proof. For a string $x = x[1] \cdots x[M]$ with $\mathbb{E}Y_x \leq N^q$. We define $x_t = x[1] \cdots x[t]$ for $1 \leq t \leq M$. Suppose t is the smallest integer such that $\mathbb{E}Y_{x_t} \leq N^q$. By (2.7), we have

$$\mathbb{P}(Y_{x_{t-1}} \geq 2\mathbb{E}Y_{x_{t-1}}) \leq \exp(-cN^q).$$

On the event $\{Y_{x_{t-1}} \leq 2\mathbb{E}Y_{x_{t-1}}\}$, we have $\mathbb{E}[Y_{x_t}|Y_{x_{t-1}}] \leq 2N^q$. We then obtain by Lemma 2.5 that

$$\mathbb{P}(Y_{x_t} \geq 3N^q) \leq \exp(-cN^q),$$

completing the proof since $Y_x \leq Y_{x_t}$. $\square$

2.4 Notation

We write $o_\chi(\cdot)$ to indicate quantities which are $o(\cdot)$ as χ tends to 0. We write e.g. $0 < \delta_1 \ll \delta_2 < 1$ if δ_1 is chosen sufficiently small depending on δ_2 , and $1 < m_1 \ll m_2 < \infty$ if m_2 is sufficiently large given m_1 . Throughout this paper, we work with several constants. First, we choose a large constant C satisfying

$$C \gg \frac{K}{\log N}, \quad (2.8)$$

i.e. C is sufficiently large depending on μ (with the implicit assumption that $\frac{K}{C_\mu \log N} \in [1/2, 2]$, say). Then we choose small constants in the order:

$$0 < \zeta \ll \varphi \ll \rho \ll \chi \ll \xi \ll \delta \ll \varepsilon. \quad (2.9)$$

In other words ε small depending on (μ, C) , then choose δ small depending on (μ, C, ε) , then choose ξ arbitrarily small depending on $(\mu, C, \varepsilon, \delta)$, and so on. The constants $\chi, \rho, \varphi, \varepsilon$ will correspond to the statement of Theorem 2.1. For fixed χ , we will fix a partition of $\mathcal{D}_k$ as in (2.1). We write

$$h_i = \mu(D_i) \text{ and choose } \mathbf{p}^{(i)} = (p_0^{(i)}, \dots, p_{k-1}^{(i)}) \in D_i^\circ, \text{ for all } 0 \leq i \leq z, \quad (2.10)$$

which will be treated as fixed values. (Here D_i° denotes the interior of D_i .) We also introduce the set

$$T = \{(i, l) : p_l^{(i)} > \chi\}. \quad (2.11)$$

The constant δ features crucially in the first moment analysis of Section 4 and the lower bound analysis of Section 6. The large constant C will appear in Definition 4.3, where we need to truncate the prefix set. Finally ξ and ζ will appear primarily in Lemma 5.6 near the end of the proof. All of these constants are fixed as $N \rightarrow \infty$, i.e. we also have $1/\log N \ll \zeta$. In particular they do not depend on the precise cut sizes $\vec{n}$.

Consider a shuffle process with pile sizes $\vec{n} = (\vec{n}^{(t)})_{t \in \mathbb{Z}^+}$, treated here as deterministic (by conditioning on $\vec{n}$, in the randomized case).

With arbitrary tie-breaking, we define $l_{\max}^{(t)} = \arg \max_{i \in [k]_0} n_i^{(t)}$. Since $\sum_{i \in [k]_0} n_i^{(t)} = N$, we have that

$$n_{l_{\max}^{(t)}}^{(t)} \geq \frac{N}{k}. \quad (2.12)$$

For a prefix $x \in [k]_0^M$, we write B_x as the set of strings in $[k]_0^K$ with prefix x . Moreover, we define

$$\iota(x) = |\{i \in [N] : s_i <_{\text{lex}} x\}| + 1, \quad \tau(x) = |\{i \in [N] : s_i <_{\text{lex}} x \text{ or } s_i \in B_x\}|. \quad (2.13)$$

These bookend the (random) discrete interval

$$\mathcal{I}(B_x) = \{i \in [N] : s_i \in B_x\} = \{\iota(x), \iota(x) + 1, \dots, \tau(x)\} \quad (2.14)$$

consisting of the indices of strings in B_x . Note that $Y_x = |\mathcal{I}(B_x)|$ (as defined in Lemma 2.6). Moreover, we define G_{B_x} to be the induced subgraph of G with vertex set $\mathcal{I}(B_x)$, which retains the edges $(i, i+1) \in E(G)$ such that $s_i = s_{i+1} \in B_x$. Denote its edge set by $E(G_{B_x})$. We also define t_x and λ_x in $[0, 1]$ to be the expected endpoints of $\mathcal{I}(B_x)$:

$$t_x = \frac{1}{N} \mathbb{E} \iota(x), \quad t_x + \lambda_x = \frac{1}{N} \mathbb{E} \tau(x). \quad (2.15)$$

By the definition of S_K , for any string $x = x[1] \cdots x[M]$, we have

$$\lambda_x = \prod_{t=1}^M \frac{n_{x[t]}^{(t)}}{N}. \quad (2.16)$$

For a prefix x , we let

$$J_x = [t_x, t_x + \lambda_x) \quad (2.17)$$

be the expected location of the strings with prefix x in S_K .

For convenience, we write $i^n = ii \cdots i$ for a length n string consisting solely of the digit i . For two strings $x = x[1] \cdots x[a]$ with length a and $y = y[1] \cdots y[b]$ with length b , we write $[xy]$ to denote their concatenation, i.e. the length $a+b$ string given by

$$[xy] = x[1] \cdots x[a]y[1] \cdots y[b].$$

Recall the definition of D_i in (2.1). We consider the following partition of the set $[K] \times [k]_0 = \mathbf{P} \cup \mathbf{Q} = \mathbf{P}(\vec{n}) \cup \mathbf{Q}(\vec{n})$, where

$$\mathbf{P}_i = \{(t, l) \in \mathbf{P} : \mathbf{n}^{(t)}/N \in D_i, p_l^{(i)} \geq \chi\}, \quad \mathbf{P} = \cup_{0 \leq i \leq z} \mathbf{P}_i \text{ and } \mathbf{Q} = ([K] \times [k]_0) \setminus \mathbf{P}. \quad (2.18)$$

It is elementary that $|\mathbf{P}| + |\mathbf{Q}| = Kk$, and

$$|\{l : (t, l) \in \mathbf{P}\}| + |\{l : (t, l) \in \mathbf{Q}\}| = k, \quad \forall t \in [K].$$

Given any k -tuple $(a_0, a_1, \dots, a_{k-1})$ of non-negative real numbers with sum $a_{\text{tot}} > 0$, let

$$H(a_0, a_1, \dots, a_{k-1}) = \frac{\sum_{i=0}^{k-1} a_i \log(\frac{a_i}{a_{\text{tot}}})}{a_{\text{tot}}}$$

be the entropy of the discrete distribution with weights $(\frac{a_i}{a_{\text{tot}}})_{0 \leq i \leq k-1}$. The following entropic approximation for multinomial coefficients will be useful.

Lemma 2.8 ([CS04, Lemma 2.2]). *For any fixed $A > 0$ and any non-negative real numbers $a_0, a_1, \dots, a_{k-1} \in [0, A]$, satisfying $a_i \log N \in \mathbb{Z}$,*

$$N^{a_{\text{tot}} H(a_0, \dots, a_{k-1}) - o_N(1)} \leq \binom{a_{\text{tot}} \log(N)}{a_0 \log(N), \dots, a_{k-1} \log(N)} \leq N^{a_{\text{tot}} H(a_0, \dots, a_{k-1})},$$

where the term $o_N(1)$ tends to 0 for any fixed A as $N \rightarrow \infty$, uniformly in $a_0, \dots, a_{k-1} \in [0, A]$.

Let $\mathbf{p}^t$ be the probability distribution given by $(\mathbf{p}^t)_i = \frac{p_i^t}{\phi_{\mathbf{p}}(t)}$. Define

$$I(\mathbf{p}, \mathbf{p}^t) \equiv D_{\text{KL}}(\mathbf{p}^t \parallel \mathbf{p}) + H(\mathbf{p}^t) = \sum_{i=0}^{k-1} (\mathbf{p}^t)_i \log(1/p_i) = \sum_{i=0}^{k-1} \frac{p_i^t \log(1/p_i)}{\phi_{\mathbf{p}}(t)} > 0.$$

We have

$$H(\mathbf{p}^t) = t \cdot I(\mathbf{p}, \mathbf{p}^t) - \psi_{\mathbf{p}}(t). \quad (2.19)$$

This quantity appears in calculations associated with the cold spot obstruction.

3 Upper Bound

In this section, we provide upper bounds on the mixing time. Throughout, we fix a deterministic sequence of pile sizes $(\bar{\mathbf{n}}^{(1)}, \bar{\mathbf{n}}^{(2)}, \dots)$ (except for Subsection 3.3 which addresses the random pile size case). We begin by proving (2.4), the case of deterministic almost $\mathbf{p}$ -like shuffles. In Section 3.1, we show that the initial shuffles can be assumed to behave well, which is needed later for technical reasons. In Section 3.2, we reduce the proof of (2.4) to verifying Lemma 3.9, which provides an estimate on the “truncated exponential moment” of the number of shared edges of two independent samples, and Lemma 3.12, which asserts that the shuffle graph is “sparse” with high probability. Finally, we conclude this section by proving that, with high probability, the random series of pile sizes $\mathbf{X}$ is (χ, ρ, φ) -almost- μ -like and deriving (1.8) from (2.4). In fact, in Sections 4 and 5, we will upper-bound the first moment and the exponential moment of the number of shared edges respectively, thereby establishing Lemmas 3.9 and 3.12.

3.1 Ensuring initial shuffles are good

For technical reasons in Section 5, we need to assume some condition on initial shuffles. The following lemmas explain why this assumption can be made without loss of generality. We first observe that ignoring an initial subsequence of shuffles only slows mixing.

Lemma 3.1. *For any sequence of pile sizes and any $1 \leq a \leq K$, we have:*

$$d_{\text{TV}}(\Pi_{1 \leq i \leq K} \mathbf{P}(n_0^{(i)}, \dots, n_{k-1}^{(i)}), \nu_{\text{Unif}}) \leq d_{\text{TV}}(\Pi_{a \leq i \leq K} \mathbf{P}(n_0^{(i)}, \dots, n_{k-1}^{(i)}), \nu_{\text{Unif}})$$

Proof. Let $\pi_{[1,a]} = \Pi_{1 \leq i \leq a-1} \mathbf{P}(n_0^{(i)}, \dots, n_{k-1}^{(i)})$ and $\pi_{[a,K]} = \Pi_{a \leq i \leq K} \mathbf{P}(n_0^{(i)}, \dots, n_{k-1}^{(i)})$. Then the claim follows from:

$$d_{\text{TV}}(\pi_{[1,a]} * \pi_{[a,K]}, \nu_{\text{Unif}}) = d_{\text{TV}}(\pi_{[1,a]} * \pi_{[a,K]}, \pi_{[1,a]} * \nu_{\text{Unif}}) \leq d_{\text{TV}}(\pi_{[a,K]}, \nu_{\text{Unif}}).$$

Here the first step holds because ν_{Unif} is the uniform distribution, while the second is a general property of convolutions. $\square$

Next, we show there must exist three χ -good shuffles within the first $\frac{\varepsilon}{2} \log N$ shuffles which are relatively close together. This condition guarantees that for any prefix x whose length $\mathbf{A}$ corresponds to the third χ -good shuffle, the expected size of $\mathcal{I}(B_x)$ —denoted by λ_x —admits a constant lower bound. Essentially speaking, this ensures that there remains sufficient “space” in $[k]_0^M$ for the future strings that appear in the exploration process. For related arguments and detailed estimates, we refer the reader to Section 5, in particular Section 5.4 and equation (5.43). Recall the definition of t_* in (2.2). Here and throughout the rest of the paper we set

$$\mathbf{A}(\mu) = 2(1 - \mu(\tilde{V}_k))^{-1}.$$

Lemma 3.2. *For any (χ, ρ, φ) -almost- μ -like shuffle process with pile sizes $(\bar{\mathbf{n}}^{(t)})_{t \leq K}$, there exists $t_* \leq t_0 < t_1 < t_2 \leq t_0 + \mathbf{A}(\mu) \leq t_* + \rho \log N$ such that $\bar{\mathbf{n}}^{(t_0)}$, $\bar{\mathbf{n}}^{(t_1)}$ and $\bar{\mathbf{n}}^{(t_2)}$ are χ -good.*

Proof. Suppose for contradiction that for all

$$t_* \leq t \leq t_* + \rho \log N - \mathbf{A}(\mu)$$

such that $\vec{n}^{(t)}$ is χ -good, there is at most one χ -good shuffle among $\vec{n}^{(t+i)}$ ($1 \leq i \leq A(\mu)$). This implies a total of at least

$$\rho \log N \cdot \frac{A(\mu) - 1}{1 + A(\mu)} > \rho \log N \cdot (\mu(\tilde{V}_k) + 2\varphi)$$

χ -bad shuffles in $[t_*, t_* + \rho \log N]$. However, this results in a contradiction, since for any (χ, ρ, φ) -almost- μ -like shuffle process, at most $\rho \log N \cdot (\mu(\tilde{V}_k) + \varphi)$ shuffles in $[t_*, t_* + \rho \log N]$ can be χ -bad. $\square$

By Lemma 3.2, we can find some $t_0 < 2\rho \log N \leq \frac{\varepsilon}{2} \log N$ such that $\vec{n}^{(t_0)}$ is χ -good, and $\vec{n}^{(t_0+i_1)}, \vec{n}^{(t_0+i_2)}$ are χ -good for some $i_1 < i_2 \leq A(\mu)$. Consider now the shuffle process with pile sizes

$$\vec{n}^{(t), \text{new}} = \vec{n}^{(t+t_0-1)}, \quad \text{for all } 1 \leq t \leq K^{\text{new}} \stackrel{\text{def.}}{=} K - t_0 + 1.$$

In this new shuffle process $(\vec{n}^{(t), \text{new}})_{t \leq K^{\text{new}}}$, the first shuffle is χ -good, and there are two χ -good shuffles in the interval $(1, A(\mu)]$. Moreover, the new shuffle remains (χ, ρ, φ) -almost- μ -like, and $K^{\text{new}} \geq (\bar{C}_\mu + \frac{\varepsilon}{2}) \log N$. We emphasize here that by introducing t_* , the (χ, ρ, φ) -almost- μ -like property is preserved after shifting. By Lemma 3.1, if we establish the desired upper bounds on the total variation distance for the new shuffle, we obtain the corresponding results for the original shuffle process. Therefore, without loss of generality, we may assume that the first shuffle, and two other shuffles before time $A(\mu)$ are χ -good.

3.2 Upper bound in the deterministic case

Recall that $\pi \sim \nu_{\text{Unif}}$ is uniformly distributed on the symmetric group S_N and $G = G(S_K)$ is the shuffle graph after K shuffles. We denote by $\mathbb{P}^\pi$ the probability measure under which $\pi \sim \nu_{\text{Unif}}$, and by $\mathbb{E}^\pi$ the corresponding expectation. We first describe the Radon–Nikodym derivative of π^G with respect to π , following [Sel22, Section 3]. Let v_i denote the number of vertices in the i -th component of G . For fixed $\sigma \in S_N$, recalling the definition of $(\cdot)^G$ from above Proposition 2.2, direct calculation shows that

$$\mathbb{P}^\pi(\pi^G = \sigma) = \frac{1_{\{\sigma^G = \sigma\}} \Pi_i(v_i!)}{N!}.$$

Therefore, the Radon–Nikodym derivative of π^G with respect to π is given by

$$f_{G, \sigma} \equiv \frac{\mathbb{P}^\pi(\pi^G = \sigma)}{\mathbb{P}^\pi(\pi = \sigma)} = \frac{1_{\{\sigma^G = \sigma\}}}{\mathbb{P}^\pi(\pi^G = \pi)}$$

We now consider σ as a random variable with law ν_{Unif} , and denote by $\mathbb{E}^\sigma$ the corresponding expectation. From the equation above, we can derive that the total variation distance to the uniform distribution after K shuffles is given by

$$d_{\text{TV}}(\Pi_{t \leq K} \mathbf{P}(n_0^{(t)}, \dots, n_{k-1}^{(t)}), \nu_{\text{Unif}}) = \frac{1}{2} \cdot \mathbb{E}^\sigma |\mathbb{E}^S[f_{G(S), \sigma} - 1]|.$$

Next, we apply a chi-squared upper bound for the total variation distance after excluding some exceptional rare events. Consider a partition $\mathcal{S} = \mathcal{S}^1 \cup \mathcal{S}^0$, where $\mathcal{S}$ denotes the space of all sequences generated by shuffle graphs of scale $N \times K$, and $\mathcal{S}^1$ consists of all “typical” sequences. (We will not fix such a partition explicitly until Section 5, but one should just think for now that $\mu_K(\mathcal{S}^0)$ is small and that $\mathcal{S}^1$ will have some nice properties to be determined as needed.) Recall that S_K distributes uniformly on $\mathcal{S}$. Take S'_K to be an independent copy of S_K and define for any shuffle graphs G, G' ,

$$f_{G, G'} \equiv \mathbb{E}^\sigma [f_{G, \sigma} f_{G', \sigma}].$$

By [Sel22, Equations (3.2) and (3.3)], we obtain

$$\mathbb{E}^\sigma |\mathbb{E}^S[f_{G(S_K), \sigma} - 1]| \leq \sqrt{\mathbb{E}^{S, S'} [(f_{G, G'} - 1) 1_{S_K, S'_K \in \mathcal{S}^1}] + 2\mu_K(\mathcal{S}^0)}. \quad (3.1)$$

We now upper-bound $f_{G, G'}$ using the number of edges shared by G and G' . Let $E(G, G') = E(G) \cap E(G')$. For two shuffle graphs G, G' , we define the new shuffle graph U to be their edge union and denote by $\mathcal{C}(U)$ the set of connected components of U . The next lemma shows how to upper-bound $f_{G, G'}$ based on $E(G, G')$ and the new shuffle graph U .

Lemma 3.3 ([Sel22, Lemma 3.1]). *Suppose the U -components have vertex-sizes $(u_1, \dots, u_c)$. Then*

$$f_{G,G'} \leq \prod_{\substack{1 \leq i \leq c, \\ E(U_i) \cap E(G,G') \neq \emptyset}} (u_i!). \quad (3.2)$$

We would like the right-hand side of (3.2) to be at most exponential in the number $E(G, G')$ of shared edges. This will be ensured by a truncation involving the following “local sparsity” condition (where we will take L large depending only on (μ, ε)).

Definition 3.4. *For $L \geq 10$ a positive integer, a shuffle graph G is **L -sparse** if within any discrete interval $\{i, i+1, \dots, i+L-1\} \subseteq [N]$ of L consecutive vertices, at most $L/3$ (of the possible $L-1$) edges are in $E(G)$.*

Lemma 3.5 ([Sel22, Lemma 3.3]). *Suppose G and G' are L -sparse shuffle graphs. Then*

$$f_{G,G'} \leq (L!)^{|E(G,G')|}.$$

To estimate the exponential moment, the idea will be to reduce to conditional first moment bounds via a form of Portenko’s lemma [Por75]. As in [Sel22], we explore (G, G') one vertex at a time and argue that at any intermediate stage, the expected number of unrevealed shared edges is $o(1)$. However as observed there, this argument may break down near the end of the exploration (in the most extreme case, all remaining Km digits could be $(k-1)$ when there are m cards left). To mitigate this issue we label shared edges of G and G' as either “forward” or “backward” edges, such that every edge satisfies at least 1 of the classifications (after truncating exceptional G and G'); this allows each exploration to stop before getting too close to the end. Since in our analysis the number $n_l^{(1)}$ of strings starting with $[l]$ is deterministic, we in fact define “forward” edges and “backward” edges separately for strings starting with each $l \in [k]_0$. Moreover, since the number of digits equal to 0 or $k-1$ may be unusually small in certain shuffles, unlike the setting in [Sel22], we introduce the notions of $I_{\text{left}}^{(t)}$ and $I_{\text{right}}^{(t)}$ to ensure that the exploration process does not proceed too close to the endpoints.

Definition 3.6. *For all $t \in [K]$, we let*

$$I_{\text{left}}^{(t)} = \inf \left\{ l \geq 0 : \frac{n_l^{(t)}}{N} \geq \chi \right\}, \quad I_{\text{right}}^{(t)} = \sup \left\{ l \leq k-1 : \frac{n_l^{(t)}}{N} \geq \chi \right\}$$

be the first and last piles with at least χN cards in the t -th shuffle. Note that $I_{\text{left}}^{(t)} < I_{\text{right}}^{(t)}$ if and only if t is a χ -good shuffle, and that if no pile sizes are small for shuffle t then $I_{\text{left}}^{(t)} = 0$ and $I_{\text{right}}^{(t)} = k-1$.

We can now introduce the “forward” and “backward” edges. For simplicity, we define

$$a_1 = 1, \quad a_{k+1} = \inf \{ t > a_k : \mathbf{n}^{(t)}/N \notin \tilde{V}_k \}. \quad (3.3)$$

Note that $I_{\text{left}}^{(t)} < I_{\text{right}}^{(t)}$ if and only if $t = a_k$ for some $k \geq 1$. Thus, by our assumption stated in Section 3.1, we have $a_3 \leq A(\mu)$.

Definition 3.7. *For all $p \in [k]_0$, we let*

$$S_{\text{forward}}^p = \{ x \in \cup_{a_3 \leq M \leq K} [k]_0^M : x[1] = p \text{ and } x[2]x[3] \cdots x[a_3] <_{\text{lex}} I_{\text{right}}^{(2)} I_{\text{right}}^{(3)} \cdots I_{\text{right}}^{(a_3)} \}.$$

Let $E_{\text{forward}}^p(G)$ consist of all edges $(i, i+1) \in E(G)$ for which $s_i = s_{i+1} \in S_{\text{forward}}^p$, and write

$$E_{\text{forward}}^p(G, G') = E_{\text{forward}}^p(G) \cap E_{\text{forward}}^p(G').$$

Similarly, for all $p \in [k]_0$, we let

$$S_{\text{backward}}^p = \{ x \in \cup_{a_3 \leq M \leq K} [k]_0^M : x[1] = p \text{ and } x[2]x[3] \cdots x[a_3] >_{\text{lex}} I_{\text{left}}^{(2)} I_{\text{left}}^{(3)} \cdots I_{\text{left}}^{(a_3)} \}.$$

Define $E_{\text{backward}}^p(G, G')$ in the same way but with S_{forward}^p replaced by S_{backward}^p .

Note that $E_{\text{forward}}^p(G) \subset [\iota([p]), \tau([p])]$ is contained within an interval that does not depend on the shuffle graph G . Therefore, $E_{\text{forward}}^p(G) \cap E_{\text{forward}}^{p'}(G') = \emptyset$ if $p \neq p'$. In the next truncation condition we require the 3 digits a_1, a_2, a_3 to have typical distribution, so that the forward and backward edges indeed cover all of $E(G, G')$. It is important for the truncation that this condition can be defined individually for G and G' separately rather than jointly.

For simplicity, we let

$$\begin{aligned}\lambda_{\text{left}}^p &= \frac{1}{N} \mathbb{E} \left| \left\{ i \in [\iota([p]), \tau([p])] : s_i[1]s_i[2] \cdots s_i[a_3] \leq_{\text{lex}} p \mathbf{I}_{\text{left}}^{(2)} \mathbf{I}_{\text{left}}^{(3)} \cdots \mathbf{I}_{\text{left}}^{(a_3)} \right\} \right|; \\ \lambda_{\text{right}}^p &= \frac{1}{N} \mathbb{E} \left| \left\{ i \in [\iota([p]), \tau([p])] : s_i[1]s_i[2] \cdots s_i[a_3] \geq_{\text{lex}} p \mathbf{I}_{\text{right}}^{(2)} \mathbf{I}_{\text{right}}^{(3)} \cdots \mathbf{I}_{\text{right}}^{(a_3)} \right\} \right|\end{aligned}\tag{3.4}$$

Definition 3.8. The sequence $S = (s_1, \dots, s_N) \in \mathcal{S}$ of strings is ϑ -regular if for all $p \in [k]_0$

$$\begin{aligned}\left| \left\{ i \in [\iota([p]), \tau([p])] : s_i[1]s_i[2] \cdots s_i[a_3] \leq_{\text{lex}} p \mathbf{I}_{\text{left}}^{(2)} \cdots \mathbf{I}_{\text{left}}^{(a_3)} \right\} \right| &\leq N \left(\lambda_{\text{left}}^p + \frac{\vartheta}{2} \frac{n_p^{(1)}}{N} \right); \\ \left| \left\{ i \in [\iota([p]), \tau([p])] : s_i[1]s_i[2] \cdots s_i[a_3] \geq_{\text{lex}} p \mathbf{I}_{\text{right}}^{(2)} \cdots \mathbf{I}_{\text{right}}^{(a_3)} \right\} \right| &\leq N \left(\lambda_{\text{right}}^p + \frac{\vartheta}{2} \frac{n_p^{(1)}}{N} \right).\end{aligned}\tag{3.5}$$

for some $\vartheta > 0$.

Lemma 3.9. If $S, S' \in \mathcal{S}$ are ϑ -regular for some $\vartheta > 0$ such that for all $p \in [k]_0$,

$$\lambda_{\text{left}}^p + \lambda_{\text{right}}^p < (1 - \vartheta) \frac{n_p^{(1)}}{N},$$

then

$$|E(G, G')| \leq \sum_{p=0}^{k-1} [|E_{\text{forward}}^p(G, G')| + |E_{\text{backward}}^p(G, G')|]$$

Proof. For simplicity, we let $E^p(G, G')$ denote the shared edges of G and G' in $[\iota([p]), \tau([p])]$. Regularity implies that $E_{\text{forward}}^p(G, G')$ contains all shared edges $(i, i+1) \in E^p(G, G')$ with

$$i \leq \tau([p]) - N \left(\lambda_{\text{right}}^p + \frac{\vartheta}{2} \frac{n_p^{(1)}}{N} \right),\tag{3.6}$$

and $E_{\text{backward}}^p(G, G')$ contains all shared edges $(i, i+1) \in E^p(G, G')$ with

$$i \geq \iota([p]) + N \left(\lambda_{\text{left}}^p + \frac{\vartheta}{2} \frac{n_p^{(1)}}{N} \right).\tag{3.7}$$

Thus, we have

$$\begin{aligned}& \left(\tau([p]) - N \left(\lambda_{\text{right}}^p + \frac{\vartheta}{2} \frac{n_p^{(1)}}{N} \right) \right) - \left(\iota([p]) + N \left(\lambda_{\text{left}}^p + \frac{\vartheta}{2} \frac{n_p^{(1)}}{N} \right) \right) \\ &= N \left((1 - \vartheta) \frac{n_p^{(1)}}{N} - \lambda_{\text{left}}^p - \lambda_{\text{right}}^p \right) \geq 0.\end{aligned}$$

Combining this with (3.6) and (3.7), we obtain

$$E^p(G, G') \subset E_{\text{forward}}^p(G, G') \cup E_{\text{backward}}^p(G, G').$$

The proof is then completed by observing that the following partition holds:

$$E(G, G') = \bigcup_{p \in [k]_0} E^p(G, G'). \quad \square$$

We now specify our choice of ϑ . Recall that $a_3 \leq A(\mu)$ as N varies, where $A(\mu)$ is a constant not depending on N . Let

$$\vartheta = \chi^{A(\mu)}.$$

Lemma 3.10. For all $\chi > 0$ and $p \in [k]_0$,

$$\lambda_{\text{left}}^p + \lambda_{\text{right}}^p < (1 - \vartheta) \frac{n_p^{(1)}}{N}.$$

Proof. By the definition of λ_{left}^p and λ_{right}^p in (3.4), we have

$$n_p^{(1)} - N(\lambda_{\text{left}}^p + \lambda_{\text{right}}^p) = \mathbb{E} \left| i \in [N] : p \mathbf{I}_{\text{left}}^{(2)} \cdots \mathbf{I}_{\text{left}}^{(a_3)} <_{\text{lex}} s_i[1] s_i[2] \cdots s_i[a_3] <_{\text{lex}} p \mathbf{I}_{\text{right}}^{(2)} \cdots \mathbf{I}_{\text{right}}^{(a_3)} \right|$$

Recall the definition of a_i in (3.3) and below, we have $\mathbf{I}_{\text{left}}^{(t)} < \mathbf{I}_{\text{right}}^{(t)}$ if and only if $t = a_k$ for some $k \geq 1$. Therefore, we have

$$n_p^{(1)} - N(\lambda_{\text{left}}^p + \lambda_{\text{right}}^p) \geq N \lambda_{p \mathbf{I}_{\text{left}}^{(2)} \cdots \mathbf{I}_{\text{left}}^{(a_3-1)} \mathbf{I}_{\text{right}}^{(a_3)}}. \quad (3.8)$$

Note that

$$\lambda_{p \mathbf{I}_{\text{left}}^{(2)} \cdots \mathbf{I}_{\text{left}}^{(a_3-1)} \mathbf{I}_{\text{right}}^{(a_3)}} = \frac{n_p^{(1)}}{N} \cdot \frac{n_{\mathbf{I}_{\text{right}}^{(a_3)}}^{(a_3)}}{N} \cdot \prod_{2 \leq t \leq a_3-1} \frac{n_{\mathbf{I}_{\text{left}}^{(t)}}^{(t)}}{N} \stackrel{(3.3)}{\geq} \frac{n_p^{(1)}}{N} \cdot \chi^{a_3-1} \geq \frac{n_p^{(1)}}{N} \cdot \chi^{\Lambda(\mu)}.$$

Combining this with (3.8) gives the desired result. $\square$

Lemma 3.11. For all $K \geq (\bar{C}_\mu + \varepsilon) \log N$, $t > 0$, and $p \in [k]_0$, there exists $\delta = \delta(\mu, \varepsilon) > 0$ and two sequence sets $\mathcal{S}_p^{\text{forward}}$ and $\mathcal{S}_p^{\text{backward}}$ such that the following holds.

(a) The sequence sets $\mathcal{S}_p^{\text{forward}}$ and $\mathcal{S}_p^{\text{backward}}$ are typical, i.e.

$$\mathbb{P}(S_K \in \mathcal{S}_p^{\text{forward}}) \geq 1 - \exp(-N^\delta), \quad \mathbb{P}(S_K \in \mathcal{S}_p^{\text{backward}}) \geq 1 - \exp(-N^\delta).$$

(b) The truncated exponential moment of number of shared edges is small: both of the inequalities

$$\begin{aligned} \mathbb{E}(e^{t|E_{\text{forward}}^p(G, G')|} \cdot 1_{S_K, S'_K \in \mathcal{S}_p^{\text{forward}}}) &\leq 1 + O(N^{-\delta}), \\ \mathbb{E}(e^{t|E_{\text{backward}}^p(G, G')|} \cdot 1_{S_K, S'_K \in \mathcal{S}_p^{\text{backward}}}) &\leq 1 + O(N^{-\delta}), \end{aligned} \quad (3.9)$$

hold for sufficiently large N .

(c) If the sequence $S_K \in \mathcal{S}^\dagger$, where

$$\mathcal{S}^\dagger = \bigcap_{p \in [k]_0} (\mathcal{S}_p^{\text{forward}} \cap \mathcal{S}_p^{\text{backward}}),$$

then S_K is ϑ -regular.

Lemma 3.12. For all $K \geq (\tilde{C}_\mu + \varepsilon) \log N$, there exists $L = L(\mu, \varepsilon)$ such that $G(S)$ is L -sparse with probability at least $1 - 1/N$, for N sufficiently large.

We now give the proof of (2.4) assuming Lemmas 3.11 and 3.12. We will complete the proof of Lemmas 3.11 and 3.12 in Section 5, after establishing a weaker first moment analog of (3.9) in Section 4.

Proof of (2.4). We take the integer L in Lemma 3.12. Then we define $\mathcal{S}^*$ to be the sequence set that consists of all L -sparse sequences, and we set $\mathcal{S}^1 = \mathcal{S}^\dagger \cap \mathcal{S}^*$, where $\mathcal{S}^\dagger$ is constructed in Lemma 3.11 when $\vartheta = \hat{\vartheta}$. Then we have

$$\mathbb{P}(\mathcal{S}^0) = O(1/N).$$

By (3.1), it suffices to upper-bound

$$\mathbb{E}^{S, S'} [(f_{G, G'} - 1) 1_{S_K, S'_K \in \mathcal{S}^1}]$$

by a quantity which tends to 0. Indeed we have

$$\begin{aligned}
& \mathbb{E}^{S, S'} [(f_{G, G'} - 1) 1_{S_K, S'_K \in \mathcal{S}^1}] \\
& \stackrel{\text{Lemma 3.5}}{\leq} \mathbb{E}^{S, S'} [((L!)^{|E(G, G')|} - 1) 1_{S_K, S'_K \in \mathcal{S}^1}] \\
& \stackrel{\text{Lemma 3.9}}{\leq} \mathbb{E}^{S, S'} [((L!)^{\sum_{p \in [k]_0} |E_{\text{forward}}^p(G, G')| + \sum_{p \in [k]_0} |E_{\text{backward}}^p(G, G')|} - 1) 1_{S_K, S'_K \in \mathcal{S}^1}] \\
& \leq \frac{1}{2k} \mathbb{E}^{S, S'} \left[\left(\sum_{p \in [k]_0} [((L!)^{2k|E_{\text{forward}}^p(G, G')|} - 1) + ((L!)^{2k|E_{\text{backward}}^p(G, G')|} - 1)] \right) 1_{S_K, S'_K \in \mathcal{S}^1} \right] \\
& \stackrel{(3.9)}{\leq} O(N^{-\delta}),
\end{aligned}$$

where we use (3.9) for $t = 2k \log(L!)$ in the last line. Moreover, since δ from (2.9) depends only on μ and ε , the upper bound is uniform in $\vec{n}$. $\square$

3.3 Upper bound for random pile sizes

In this subsection we extend the upper bound from deterministic to random pile sizes. Namely we will deduce (1.8) (the second part of Theorem 1.1) assuming the mixing time upper bound (2.4) for deterministic cuts. First we show that the shuffle process with pile sizes $(\mathbf{X}^{(t)})_{t \leq K}$ is (χ, ρ, φ) -almost- μ -like with high probability. We also introduce the event

$$\text{Like}(\chi, \rho, \varphi) = \left\{ (\mathbf{X}^{(t)})_{t \leq K} \text{ is } (\chi, \rho, \varphi)\text{-almost-}\mu\text{-like} \right\}.$$

Lemma 3.13. *Suppose that $(\mathbf{X}^{(t)})_{t \in \mathbb{Z}^+}$ is randomized- μ -like, we have*

$$\mathbb{P}(\text{Like}(\chi, \rho, \varphi)) = 1 - o(1).$$

Proof. Recall the definition of randomized- μ -like in (1.3). We conclude immediately by the definition of weak convergence and testing (1.3) with the indicator function $\mathbb{1}_{D_i \times (\frac{t^*}{\log N} + \rho(j-1), \frac{t^*}{\log N} + \rho j]}$. $\square$

We now obtain the mixing time upper bound (1.8) in Theorem 1.1:

Proof of (1.8). For any shuffle process with pile sizes $\vec{n}$, we set

$$\begin{aligned}
& \pi_{\vec{n}} \text{ is a random variable with law } \mathbf{P}(\vec{n}, \lfloor (1 + \varepsilon) \overline{C}_\mu \log(N) \rfloor); \\
& \sigma_{\vec{n}} \text{ is a random variable with law } \nu_{\text{Unif}}.
\end{aligned}$$

By definition of the total variation distance, we can construct a coupling $\mathbf{Q}_{\vec{n}}$ of these two random variables such that under $\mathbf{Q}_{\vec{n}}$,

$$\mathbf{Q}_{\vec{n}}(\pi_{\vec{n}} \neq \sigma_{\vec{n}}) \leq d_{\text{TV}}(\mathbf{P}(\vec{n}, \lfloor (1 + \varepsilon) \overline{C}_\mu \log(N) \rfloor), \nu_{\text{Unif}}) + \frac{1}{N}.$$

We now sample $(\pi_{\vec{n}}, \sigma_{\vec{n}})$ according to $\mathbf{Q}_{\vec{n}}$ independently, and sample the (random) pile sizes $\mathbf{X} = (\mathbf{X}^{(t)})_{t \in \mathbb{Z}^+}$ independent of $(\pi_{\vec{n}}, \sigma_{\vec{n}})$. We now have

$$\begin{aligned}
& \pi_{\mathbf{X}} \stackrel{\text{def.}}{=} \sum_{\vec{n}} \pi_{\vec{n}} \cdot 1_{\{\mathbf{X} = \vec{n}\}} \text{ is a random variable with law } \mathbf{P}(\mathbf{X}, \lfloor (1 + \varepsilon) \overline{C}_\mu \log(N) \rfloor); \\
& \sigma_{\mathbf{X}} \stackrel{\text{def.}}{=} \sum_{\vec{n}} \sigma_{\vec{n}} \cdot 1_{\{\mathbf{X} = \vec{n}\}} \text{ is a random variable with law } \nu_{\text{Unif}},
\end{aligned}$$

where the sum is over all possible series of pile sizes $\vec{n}$. Therefore, we have

$$\begin{aligned}
& \mathbb{P}(\pi_{\mathbf{X}} \neq \sigma_{\mathbf{X}}) \leq \mathbb{P}((\text{Like}(\chi, \rho, \varphi))^c) + \sum_{\vec{n} \text{ is } (\chi, \rho, \varphi)\text{-almost-}\mu\text{-like}} \mathbb{P}(\mathbf{X} = \vec{n}) \cdot \mathbf{Q}_{\vec{n}}(\pi_{\vec{n}} \neq \sigma_{\vec{n}}) \\
& \leq \mathbb{P}((\text{Like}(\chi, \rho, \varphi))^c) + \sup_{\vec{n} \text{ is } (\chi, \rho, \varphi)\text{-almost-}\mu\text{-like}} d_{\text{TV}}(\mathbf{P}(\vec{n}, \lfloor (1 + \varepsilon) \overline{C}_\mu \log(N) \rfloor), \nu_{\text{Unif}}) + \frac{1}{N}
\end{aligned}$$

for all $\chi, \rho, \varphi > 0$. We now apply Theorem 2.1 (only (2.4)) and Lemma 3.13 to obtain (1.8). $\square$

4 The first moment

In this subsection we upper-bound the first moment $\mathbb{E}[|E(G, G')|]$. This estimate will later be applied conditionally to a subset of the full shuffle graphs (see the last step of Lemma 5.6), which is still approximately described by μ . The next technical definition will ensure that our first moment estimates can be applied in such a conditional setting.

Definition 4.1. We say a shuffling process $\mathbf{n}$ is **weak- c - (χ, ρ, φ) -almost μ -like** if there exists another shuffling process $\hat{\mathbf{n}}$ which is (χ, ρ, φ) -almost μ -like such that

$$\left| \mathbf{n}^{(t)} - \hat{\mathbf{n}}^{(t)} \right| < N^{1-c}, \quad \forall t \in \mathbf{P}.$$

In this Subsection we will always assume that $\mathbf{n}$ is weak- c - (χ, ρ, φ) -almost μ -like, and in this subsection we define $\mathbf{P}_i$ with respect to $\hat{\mathbf{n}}$. Recalling (2.9), we emphasize that $N^{-c} \ll \varphi, \chi, \rho$. We will prove the following crucial bound for the first moment; note that the quantitative rate $O(N^{-\delta})$ is important since we will later sum over $O(N^\rho)$ applications of this bound.

Proposition 4.2. Fix $c \in (0, 1)$ and $\varepsilon > 0$. If $K \geq (\bar{C}_\mu + \varepsilon) \log N$, then there exists $\delta = \delta(\varepsilon, c)$ and $\rho = \rho(\delta, c), \varphi = \varphi(\delta, c)$ such that the following holds. For all weak- c - (χ, ρ, φ) -almost μ -like shuffle process $\mathbf{n}$, if the first shuffle is χ -good, we have

$$\mathbb{E}|E(G, G')| \leq O(N^{-\delta}). \quad (4.1)$$

Recall that we will choose parameters in the order (2.9) and a large constant C satisfying (2.8).

4.1 Fluctuation of the interval $\mathcal{I}(B_x)$

It is very useful to study the fluctuation of the interval $\mathcal{I}(B_x)$ (recall (2.14)). To this end, we introduce the following notations.

Definition 4.3. We define the following classes of prefixes:

$$\begin{aligned} \mathcal{A}_{\text{left}} &= \bigcup_{1 < M \leq K} \left\{ x \in [k]_0^M : x[M] > I_{\text{left}}^{(t)}, x[t] \leq I_{\text{left}}^{(t)}, \forall 1 < t < M \right\} \\ &\quad \bigcup \left\{ x \in [k]_0^K : x[t] \leq I_{\text{left}}^{(t)}, \forall 1 < t \leq K \right\}, \\ \mathcal{A}_{\text{right}} &= \bigcup_{1 < M < K} \left\{ x \in [k]_0^M : x[M] < I_{\text{right}}^{(t)}, x[t] \geq I_{\text{right}}^{(t)}, \forall 1 < t < M \right\} \\ &\quad \bigcup \left\{ x \in [k]_0^M : x[t] \geq I_{\text{right}}^{(t)}, \forall 1 < t \leq K \right\}. \end{aligned}$$

Moreover, we define

$$\begin{aligned} \tilde{\mathcal{A}}_{\text{left}} &= \left\{ x \in \mathcal{A}_{\text{left}} : |\{t : x[t] < I_{\text{left}}^{(t)}\}| < \frac{C}{\log(1/\chi)} \log N \right\}, \\ \tilde{\mathcal{A}}_{\text{right}} &= \left\{ x \in \mathcal{A}_{\text{right}} : |\{t : x[t] > I_{\text{right}}^{(t)}\}| < \frac{C}{\log(1/\chi)} \log N \right\}. \end{aligned}$$

We now define the quantity $c_F(x)$ which characterizes the order of fluctuations for location of the interval $\mathcal{I}(B_x)$. For any $y \in \mathcal{A}_{\text{left}}$ with length M_y and $z \in \mathcal{A}_{\text{right}}$ with length M_z , we define

$$A_{y,z} = \left\{ x \in \bigcup_{M \geq 1} [k]_0^M : x \text{ has prefixes } y \text{ and } z \right\}.$$

Of course $A_{y,z}$ is only non-trivial when y is a prefix of z or vice-versa. When $A_{y,z}$ is non-trivial, define

$$y \vee z = \begin{cases} y, & \text{if } M_y > M_z; \\ z, & \text{if } M_y \leq M_z. \end{cases}$$

Let M_x be the length of x . For $x \in A_{y,z}$, where $y \in \mathcal{A}_{\text{left}}$ has length M_y and $z \in \mathcal{A}_{\text{right}}$ has length M_z , we have the following decomposition:

$$\{s \in [k]_0^M : s < x\} = \bigcup_{1 \leq i \leq M_x} P_{x,i}, \quad (4.2)$$

where

$$P_{x,i} = \{s \in [k]_0^{M_x} : s[j] = x[j] \text{ for all } j < i \text{ and } s[i] < x[i]\}. \quad (4.3)$$

With these notations, we define $\text{Fluc}_{\text{left}}(x) = (\max_{2 \leq i \leq M_y} \mathbb{E}|P_{x,i}|)^{\frac{1}{2}}$, which represents the fluctuation seen from the left side. Similarly, we can define

$$\{s \in [k]_0^{M_x} : s > x\} = \bigcup_{1 \leq i \leq M} Q_{x,i}, \quad (4.4)$$

where

$$Q_{x,i} = \{s \in [k]_0^{M_x} : s[j] = x[j] \text{ for all } j < i \text{ and } s[i] > x[i]\},$$

and define $\text{Fluc}_{\text{right}}(x) = (\max_{2 \leq i \leq M_z} \mathbb{E}|Q_{x,i}|)^{\frac{1}{2}}$. With these notations, we are able to define

$$c_F(x) \equiv \min(\log_N \text{Fluc}_{\text{left}}(x), \log_N \text{Fluc}_{\text{right}}(x)).$$

It is useful to point out that

$$\mathbb{E}|\mathcal{I}(B_x)| \leq \frac{\min(\mathbb{E}|P_{x,M_y}|, \mathbb{E}|Q_{x,M_z}|)}{\chi} \leq N^{2c_F(x)+o(1)}. \quad (4.5)$$

For convenience, we summarize some concentration results here.

Lemma 4.4. *For any $y \in \mathcal{A}_{\text{left}}$, $z \in \mathcal{A}_{\text{right}}$ and $x \in A_{y,z}$, if $\mathbb{E}|\mathcal{I}(B_x)| < N^{c_F+\delta}$, then for any $a > 0$ independent of N :*

$$\mathbb{P}\left[|\iota(x) - Nt_x| \geq N^{\max(c_F, \delta) + \frac{a}{2} + \delta}\right] \leq e^{-\Omega(N^{a-o(1)})}, \quad (4.6)$$

$$\mathbb{P}\left[|\tau(x) - N(t_x + \lambda_x)| \geq N^{\max(c_F, \delta) + \frac{a}{2} + \delta}\right] \leq e^{-\Omega(N^{a-o(1)})}. \quad (4.7)$$

Proof. We first prove

$$\mathbb{P}\left[|\mathbb{E}|\mathcal{I}(B_x)| - \mathbb{E}|\mathcal{I}(B_x)|| \geq N^{\max(c_F, \delta) + \frac{a}{2} + \delta}\right] \leq e^{-\Omega(N^{a-o(1)})}, \quad (4.8)$$

When $\mathbb{E}|\mathcal{I}(B_x)| \geq N^\delta$, (4.8) holds immediately by (2.7) with $q = \log_N \mathbb{E}|\mathcal{I}(B_x)| > \delta$. Otherwise (4.8) holds by Lemma 2.7. For (4.6) and (4.7), suppose $c_F(x) = \log_N \text{Fluc}_{\text{left}}(x)$ without loss of generality. Recall that

$$\iota(x) = 1 + |\{i \in [N] : s_i <_{\text{lex}} x\}|, \quad \{s \in [k]_0^M : s <_{\text{lex}} x\} = \bigcup_{1 \leq i \leq M_x} P_{x,i},$$

where $P_{x,i}$ is defined in (4.3). Now we have

$$\{s \in [k]_0^{M_y} : s <_{\text{lex}} x\} = \{s : s[1] <_{\text{lex}} x[1]\} \cup \bigcup_{2 \leq i \leq M_x} P_{x,i}.$$

Thus, we obtain

$$\iota(x) = S_{x[1]-1} + \sum_{2 \leq i \leq M_x} |P_{x,i}|.$$

By definition of $\text{Fluc}_{\text{left}}(x)$, we have $\mathbb{E}|P_{x,i}| \leq (\text{Fluc}_{\text{left}}(x))^2$ for all $2 \leq i \leq M_y$. Moreover, for $M_y < i \leq M_x$, by (4.5) we have

$$\mathbb{E}|P_{x,i}| \leq \mathbb{E}|\mathcal{I}(B_y)| \leq N^{2c_F(y)+o(1)} = N^{2c_F(x)+o(1)}.$$

Therefore, (4.6) holds by Lemmas 2.6, 2.7 and a union bound. For (4.7), note that

$$\tau(x) = \iota(x) + |\mathcal{I}(B_x)|, \quad \mathbb{E}|\mathcal{I}(B_x)| = N\lambda_x,$$

(4.7) then holds by a combination of (4.8) and (4.6). $\square$

Now we are ready to lower bound the fluctuations of the location of $|\mathcal{I}(B_x)|$. The following lemma states that the endpoints of $\mathcal{I}(B_x)$ do not concentrate on a small set, even under conditioning. In [Sel22, Lemma 4.13], the corresponding step was implemented by simply conditioning on the multi-set

$$U_x = \{s \in S_K : s \text{ has prefix } x\}. \quad (4.9)$$

This determines the size and “internal structure” of $\mathcal{I}(B_x)$, but not its location. For multinomial cuts, given U_x , the left endpoint $\iota(x)$ of $\mathcal{I}(B_x)$ is conditionally binomial, so it is easy to analyze the homogenizing effect of this random shift. For deterministic $\vec{n}$, it appears quite complicated to resample the entirety of the strings outside U_x . Instead we will resample only carefully chosen restricted information.

Lemma 4.5. *For any $y \in \mathcal{A}_{\text{left}}, z \in \mathcal{A}_{\text{right}}$ and $x \in A_{y,z}$, there exists $\mathcal{S}'$ with $\mathbb{P}(\mathcal{S}') \geq 1 - \exp(-N^\delta)$ such that the following holds.*

$$\begin{aligned} \max_{j \in [N]} \mathbb{P}(\iota(x) = j, S_K \in \mathcal{S}' | U_x) &= O(N^{-c_F(x)+2\delta}), \\ \max_{j \in [N]} \mathbb{P}(\tau(x) = j, S_K \in \mathcal{S}' | U_x) &= O(N^{-c_F(x)+2\delta}). \end{aligned}$$

Before proving Lemma 4.5, it is useful to state the anti-concentration estimate for hypergeometric variables. See e.g. [Pit97, Eq. 25 and Eq. 38] for a proof. We note that our application of this estimate loses an ultimately irrelevant $\log N$ factor (from the depth of the prefix tree) as compared to the multinomial case, which is crucial to handle small pile sizes.

Lemma 4.6. *Let $X \sim \text{Hyp}(n_1, n, m)$ and let $\sigma^2 = \frac{n_1 m (n - n_1) (n - m)}{n^2 (n^2 - 1)}$ be the variance of X . There exists an absolute constant C such that for each $k \geq 0$,*

$$\mathbb{P}(X = k) \leq \frac{C}{\sigma}. \quad (4.10)$$

Proof of Lemma 4.5. When $c_F \leq 2\delta$, the statements hold immediately. Now we assume that $c_F > 2\delta$. Our proof strategy is to condition on a σ -field that includes U_x , and to use the remaining randomness to obtain a “shift” of U_x . In fact, we plan to use the randomness of just one shuffle.

For any $y \in \mathcal{A}_{\text{left}}$ with length M_y and $z \in \mathcal{A}_{\text{right}}$ with length M_z , by definition of $c_F(x)$, there exist $2 \leq u \leq M_y$ and $2 \leq v \leq M_z$, such that $\text{Fluc}_{\text{left}}(x) = (\mathbb{E}|P_{x,u}|)^{\frac{1}{2}}$ and $\text{Fluc}_{\text{right}}(x) = (\mathbb{E}|Q_{x,v}|)^{\frac{1}{2}}$. Without loss of generality, we suppose that $c_F(x) = \log_N \text{Fluc}_{\text{left}}(x)$. To simplify notations, we write $N' = n_{x[1]}^{(1)}$. Since the first shuffle is χ -good, we have $N' \leq (1 - \chi)N$. Recall the partitions (4.2) and (4.4), we have

$$N' = \sum_{i=2}^{M_x} \mathbb{E}|P_{x,i}| + \sum_{i=2}^{M_x} \mathbb{E}|Q_{x,i}| + \mathbb{E}|\mathcal{I}(B_x)| \leq 2M_x \mathbb{E}|Q_{x,v}| + \chi^{-1} \mathbb{E}|Q_{x,v}|.$$

Therefore, we have

$$\mathbb{E}|Q_{x,v}| \geq \Omega\left(\frac{N'}{\log N}\right).$$

We will divide the proof into three cases.

Case 1: $u > v$. Recall the definition of a shuffle matrix in Section 2.2. We first sample all columns except the v -th column. Let $R_1 \subseteq [N]$ be the set of $i \in [N]$ such that the i -th row $\bar{s}_i$ satisfies

$$\bar{s}[1 : v - 1] = x[1 : v - 1] \text{ and } \bar{s}_i[v + 1 : u] \geq x[v + 1 : u].$$

Since we have sampled all columns except the v -th column, R_1 is fixed. Furthermore, we sampled all digits in the v -th column and in rows which belong to R_1 . We write $\mathcal{F}_x$ for the σ -field associated with this conditioning. We see that U_x is $\mathcal{F}_x$ measurable. Now we define

$$\begin{aligned} a_1 &= |\{i : \bar{s}_i[1 : v - 1] < x[1 : v - 1]\}|; \\ a_2 &= |\{i : \bar{s}_i[1 : v - 1] = x[1 : v - 1] \text{ and } \bar{s}_i[v + 1 : u] < x[v + 1 : u]\}|; \\ a_3 &= |\{i \in R_1 : \bar{s}_i[v] \leq x[v]\}|; \\ a_4 &= |\{i : \bar{s}[1 : v - 1] = x[1 : v - 1], \bar{s}_i[v] < x[v] \text{ and } \bar{s}_i[v + 1 : u] \geq x[v + 1 : u]\}| \\ &\quad + |\{i : \bar{s}[1 : u] = x[1 : u] \text{ and } \bar{s}[u + 1 : K] < x[u + 1 : K]\}|. \end{aligned}$$

Notice that a_1, a_2 do not depend on the v -th shuffle, and a_3, a_4 depend only on rows in R_1 . Therefore, a_1, a_2, a_3, a_4 are all $\mathcal{F}_x$ measurable. Also define

$$Y_1 = |\{i : \bar{s}[1 : v-1] = x[1 : v-1], \bar{s}_i[v] \leq x[v], \text{ and } \bar{s}_i[v+1 : u] < x[v+1 : u]\}|$$

Now, conditioned on $\mathcal{F}_x$, we see that

$$\iota(x) = a_1 + a_4 + Y_1, \quad \tau(x) = a_1 + a_4 + Y_1 + |U_x|, \text{ and } Y_1 \sim \text{Hyp}(N - |R_1|, \sum_{l \leq x[v]} n_l^{(v)} - a_3, a_2).$$

Define the event

$$\begin{aligned} \mathcal{S}'_1 = & \left\{ \left| |R_1| - \mathbb{E}|R_1| \right| \leq \max((\mathbb{E}|R_1|)^{1/2+\delta}, N^\delta) \right\} \\ & \cap \left\{ |a_3 - \mathbb{E}a_3| \leq \max((\mathbb{E}a_3)^{1/2+\delta}, N^\delta) \right\} \cap \left\{ |a_2 - \mathbb{E}a_2| \leq (\mathbb{E}a_2)^{1/2+\delta} \right\} \end{aligned}$$

We have

$$\begin{aligned} \frac{\sum_{l \leq x[v]} n_l^{(v)}}{N} &= 1 - \frac{\sum_{l > x[v]} n_l^{(v)}}{N} \leq 1 - \frac{\mathbb{E}|Q_{x,v}|}{N} \leq 1 - \Omega\left(\frac{1}{\log N}\right); \\ \mathbb{E}a_3 &= \mathbb{E}|R_1| \cdot \frac{\sum_{l \leq x[v]} n_l^{(v)}}{N}; \\ a_2 + |R_1| &\leq N' \leq (1 - \chi)N; \\ \mathbb{E}a_2 \cdot \frac{\sum_{l \leq x[v]} n_l^{(v)}}{N} &= \mathbb{E}Y_1 \geq \mathbb{E}|P_{x,u}| = N^{2c_F(x)}. \end{aligned}$$

By (2.7) and (2.6), we have $\mathbb{P}(\mathcal{S}'_1) \geq 1 - \exp(-N^\delta)$ since $\mathbb{E}a_2 \geq N^{2c_F}$. Therefore on the event $\mathcal{S}'_1$,

$$\begin{aligned} a_2 &\leq c(\chi) \cdot (N - |R_1|); \\ \mathbb{E}(Y_1 | \mathcal{F}_x) &= a_2 \cdot \frac{\sum_{l \leq x[v]} n_l^{(v)} - a_3}{N - |R_1|} \\ &= a_2 \cdot \left(\frac{\sum_{l \leq x[v]} n_l^{(v)}}{N} + \frac{\sum_{l \leq x[v]} n_l^{(v)}}{N(N - |R_1|)} (|R_1| - \mathbb{E}|R_1|) + \frac{\mathbb{E}a_3 - a_3}{N - |R_1|} \right) \geq N^{2c_F(x) - \delta}. \end{aligned}$$

We then conclude by (4.10).

Case 2: $u < v$. Similarly to Case 1, we first sample all columns except the u -th column. Let $R_2 \subseteq [N]$ be the set of $i \in [N]$ such that the i -th row $\bar{s}_i$ satisfies

$$\bar{s}[1 : u-1] = x[1 : u-1] \text{ and } \bar{s}_i[u+1 : v] \leq x[u+1 : v].$$

Since we have sampled all columns except the u -th column, R_2 is fixed. Furthermore, we sampled all digits in the u -th column and in rows which belong to R_2 . We write $\mathcal{G}_x$ for the σ -field associated with this conditioning. We see that U_x is $\mathcal{G}_x$ measurable. Now we define

$$\begin{aligned} b_1 &= |\{i : \bar{s}_i[1 : u-1] < x[1 : u-1]\}|; \\ b_2 &= |\{i : \bar{s}_i[1 : u-1] = x[1 : u-1] \text{ and } \bar{s}_i[u+1 : v] > x[u+1 : v]\}|; \\ b_3 &= |\{i \in R_2 : \bar{s}_i[u] < x[u]\}|; \\ b_4 &= |\{i : \bar{s}_i[1 : u-1] = x[1 : u-1], \bar{s}_i[u] \leq x[u] \text{ and } \bar{s}_i[u+1 : v] < x[u+1 : v]\}| \\ &\quad + |\{i : \bar{s}_i[1 : u-1] = x[1 : u-1], \bar{s}_i[u] < x[u] \text{ and } \bar{s}_i[u+1 : v] = x[u+1 : v]\}| \\ &\quad + |\{i : \bar{s}_i[1 : v] = x[1 : v] \text{ and } \bar{s}_i[v+1 : K] < x[v+1 : K]\}| \end{aligned}$$

Notice that b_1, b_2 do not depend on the u -th shuffle, and b_3, b_4 depend only on rows in R_2 . Therefore, b_1, b_2, b_3, b_4 are all $\mathcal{G}_x$ measurable. Also define

$$Y_2 = |\{i : \bar{s}_i[1 : u-1] = x[1 : u-1], \bar{s}_i[u] < x[u] \text{ and } \bar{s}_i[u+1 : v] > x[u+1 : v]\}|.$$

Now, conditioned on $\mathcal{G}_x$, we see that

$$\iota(x) = b_1 + b_4 + Y_2, \tau(x) = b_1 + b_4 + Y_2 + |U_x|, \text{ and } Y_2 \sim \text{Hyp}(N - |R_2|, \sum_{l < x[u]} n_l^{(u)} - b_3, b_2).$$

Define the event

$$\begin{aligned} \mathcal{S}'_2 = & \left\{ \left| |R_2| - \mathbb{E}|R_2| \right| \leq \max((\mathbb{E}|R_2|)^{1/2+\delta}, N^\delta) \right\} \\ & \cap \left\{ |b_3 - \mathbb{E}b_3| \leq \max((\mathbb{E}b_3)^{1/2+\delta}, N^\delta) \right\} \cap \left\{ |b_2 - \mathbb{E}b_2| \leq (\mathbb{E}b_2)^{1/2+\delta} \right\}. \end{aligned}$$

By (2.7) and (2.6), we have $\mathbb{P}(\mathcal{S}'_2) \geq 1 - \exp(-N^\delta)$. Moreover, we have

$$\begin{aligned} \frac{\sum_{l < x[u]} n_l^{(u)}}{N} & \leq 1 - \frac{n_{x[u]}^{(u)}}{N} \leq 1 - \frac{\mathbb{E}|Q_{x,v}|}{N'} \leq 1 - \Omega\left(\frac{1}{\log N}\right); \\ \mathbb{E}b_3 & = \mathbb{E}|R_2| \cdot \frac{\sum_{l < x[u]} n_l^{(u)}}{N}; \\ b_2 + |R_2| & \leq N' \leq (1 - \chi)N; \\ \mathbb{E}b_2 \cdot \frac{\sum_{l < x[u]} n_l^{(u)}}{N} & \geq \mathbb{E}|Q_{x,v}| \cdot \frac{\sum_{l < x[u]} n_l^{(u)}}{N} \geq \Omega\left(\frac{N'}{\log N}\right) \cdot \frac{\sum_{l < x[u]} n_l^{(u)}}{N} = N' \cdot \frac{\sum_{l < x[u]} n_l^{(u)}}{N} \cdot \Omega\left(\frac{1}{\log N}\right) \\ & \geq \mathbb{E}|P_{x,u}| \cdot \Omega\left(\frac{1}{\log N}\right) \geq N^{2c_F(x) - \delta}. \end{aligned}$$

Therefore, on the event $\mathcal{S}'_2$, we have

$$\begin{aligned} b_2 & \leq c(\chi) \cdot (N - |R_2|); \\ \mathbb{E}(Y_2|\mathcal{G}_x) & = b_2 \cdot \frac{\sum_{l < x[u]} n_l^{(u)} - b_3}{N - |R_2|} \\ & = b_2 \cdot \left(\frac{\sum_{l < x[u]} n_l^{(u)}}{N} + \frac{\sum_{l < x[u]} n_l^{(u)}}{N(N - |R_2|)} (|R_2| - \mathbb{E}|R_2|) + \frac{\mathbb{E}b_3 - b_3}{N - |R_2|} \right) \geq N^{2c_F(x) - \delta}. \end{aligned}$$

We then conclude by (4.10).

Case 3: $u = v$. We sample all columns except the v -th column. Furthermore, we sample all the $x[v]$ -digit in the v -th column. We write $\mathcal{H}_x$ for the σ -field associated with this conditioning. Define

$$\begin{aligned} d_1 & = |\{i : \bar{s}_i[1 : v - 1] < x[1 : v - 1]\}|; \\ d_2 & = |\{i : \bar{s}_i[1 : v - 1] = x[1 : v - 1], \bar{s}_i[v] \neq x[v]\}|; \\ d_3 & = |\{i : \bar{s}_i[1 : v] = x[1 : v], \bar{s}_i[v + 1 : K] < x[v + 1 : K]\}|. \end{aligned}$$

Then d_1, d_2, d_3 are all $\mathcal{H}_x$ -measurable. Also define

$$Y_3 = |\{i : \bar{s}_i[1 : v - 1] = x[1 : v - 1] \text{ and } \bar{s}_i[v] < x[v]\}|.$$

Now, conditioned on $\mathcal{H}_x$, we see that

$$\iota(x) = d_1 + d_3 + Y_3, \quad \tau(x) = d_1 + d_3 + Y_3 + |U_x|, \text{ and } Y_3 \sim \text{Hyp}(N - n_{x[v]}^{(v)}, \sum_{l < x[u]} n_l^{(v)}, d_2).$$

Define the event $\mathcal{S}'_3 = \{|d_2 - \mathbb{E}d_2| \leq (\mathbb{E}d_2)^{1/2+\delta}\}$. Since $\mathbb{E}d_2 \geq \mathbb{E}|P_{x,u}| = N^{2c_F(x)} \geq N^{2\delta}$, we have $\mathbb{P}(\mathcal{S}'_3) \geq 1 - \exp(-N^\delta)$. Since $\mathbb{E}|P_{x,u}| \leq \mathbb{E}|Q_{x,v}|$, on the event $\mathcal{S}'_3$ we have

$$\begin{aligned} \sum_{l < x[u]} n_l^{(v)} & \leq \frac{1}{2}(N - n_{x[v]}^{(v)}); \\ d_2 & \leq \mathbb{E}d_2 + (\mathbb{E}d_2)^{1/2+\delta} \leq \frac{N'}{N} \cdot (N - n_{x[v]}^{(v)}) + \left(\frac{N'}{N} \cdot (N - n_{x[v]}^{(v)})\right)^{1/2+\delta} \leq (1 - c(\chi)) \cdot (N - n_{x[v]}^{(v)}). \\ \mathbb{E}(Y_3|\mathcal{H}_x) & = d_2 \cdot \frac{\sum_{l < x[u]} n_l^{(v)}}{N - n_{x[v]}^{(v)}} \geq c(\chi) \cdot \mathbb{E}d_2 \cdot \frac{\sum_{l < x[u]} n_l^{(v)}}{N} \geq c(\chi) \cdot \mathbb{E}|P_{x,u}| \geq N^{2c_F(x) - \delta}. \end{aligned}$$

We then conclude by (4.10). By taking $\mathcal{S}' = \mathcal{S}'_1 \cap \mathcal{S}'_2 \cap \mathcal{S}'_3$, we finish the proof. $\square$

4.2 Digit profiles

Definition 4.7 (digit profile). *Recall the definition of $\mathbf{Q}$ in (2.18) and definition of T in (2.11). For $y \in \mathcal{A}_{\text{left}}$ and $z \in \mathcal{A}_{\text{right}}$ and $x \in A_{y,z}$, the **digit profile** of x is*

$$\left(\{c_l^{(i)}(x)\}_{(i,l) \in T}, \{(t, x[t]) : (t, x[t]) \in \mathbf{Q}\} \right),$$

where $c_l^{(i)}(x) \in \mathbb{Z}/\log N$ is such that after the first $\max(M_y, M_z)$ digits, x contains $c_l^{(i)} \log N$ l -digits in $\mathbf{P}_i$.

Each digit profile will contribute a certain number of expected shared edges, of order a power of N . Our strategy is to prove that the number of possible digit profiles (via a truncation) is bounded by a small power of N , and control the maximal contribution from any single digit profile. Rescaling by $\log N$ is convenient because the optimal digit profiles will come from solving a continuous optimization problem over the probability simplex $\mathcal{D}_k$ which is independent of N .

Our definition of digit profile builds on that of [Sel22]. One change is that when $\vec{n}^{(1)}$ is deterministic, the sorting S_K only starts to behave randomly in its second digit. In particular, this causes all strings of the form $j00 \dots 0$ or $j(k-1)(k-1) \dots (k-1)$ to appear in highly predictable locations, so they have to be analyzed separately. In the multinomial case, this happens only for strings starting with many 0's or $(k-1)$'s, with no special exception for $y[1]$. Additionally in the multinomial case such “near boundary” prefixes suffice to completely determine the fluctuations, while our analysis must consider the much more combinatorially detailed information of pairs $(y, z) \in \mathcal{A}_{\text{left}} \times \mathcal{A}_{\text{right}}$.

We next define constants depending on the digit profile of $x \in [k]_0^M$. From now on, we always suppose by symmetry that $M_y(x) \geq M_z(x)$ (i.e. z is a prefix of y). Let

$$c_{\text{tot}}^{(i)}(x) = \sum_{l: (i,l) \in T} c_l^{(i)}(x)$$

be the number of digits in x that belong to $\mathbf{P}_i$ after y , and

$$c_{\text{tot}}(x) = \sum_{0 \leq i \leq z} c_{\text{tot}}^{(i)}(x)$$

be the number of digits in x that belong to $\mathbf{P}$ after y . Also define

$$\begin{aligned} c_L(x) &\equiv 1 - \sum_{t \leq M_y} \log_N(N/n_{x[t]}^{(t)}) - \sum_{(i,l) \in T} c_l^{(i)} \log \left(1/p_l^{(i)} \right) - \sum_{M_y < t \leq M_x : (t, x[t]) \in \mathbf{Q}} \log_N \left(N/n_{x[t]}^{(t)} \right), \\ c_E(x) &\equiv -\frac{1}{\log N} \sum_{t > M} \psi_{\mathbf{n}^{(t)}/N}(2). \end{aligned} \quad (4.11)$$

The former dictates the typical size of $\mathcal{I}(B_x)$, see (4.13) just below. Note that by (4.5) we have

$$c_F(x) \geq \frac{1}{2} c_L(x) - o(1). \quad (4.12)$$

Lemma 4.8. *Recall the definition of h_i and $\mathbf{p}^{(i)}$ in (2.10). We have*

$$c_L(x) = \log_N \mathbb{E}|\mathcal{I}(B_x)| + O(\delta), \quad (4.13)$$

$$c_E(x) = \left(\frac{M_x - K}{\log N} \right) \sum_{0 \leq i \leq z} h_i \psi_{\mathbf{p}^{(i)}}(2) + O(\delta), \quad (4.14)$$

$$c_{\text{tot}}^{(i)}(x) = \frac{1}{\log N} \left[h_i (M_x - M_y) - \sum_{M_y < t \leq M_x : (t, x[t]) \in \mathbf{Q}} 1_{\{\mathbf{n}^{(t)}/N \in D_i\}} \right] + O(\chi). \quad (4.15)$$

Proof. We first prove (4.13). Notice that when $\mathbf{n}^{(t)}/N \in D_i$ for some $k \leq i \leq z$, we have that $\|\mathbf{n}^{(t)}/N - \mathbf{p}^{(i)}\|_\infty < \chi^2$. In this case, when $p_{x[t]}^{(i)} > \chi$ we have

$$\left| \log \left(\frac{n_{x[t]}^{(t)}}{N} \right) - \log \left(p_{x[t]}^{(i)} \right) \right| < O(\chi). \quad (4.16)$$

When $\mathbf{n}^{(t)}/N \in D_i$ and $p_{x[t]}^{(i)} > \chi$ for some $0 \leq i \leq k-1$, we must have $p_{x[t]}^{(i)} > 1 - k\chi$, and (4.16) still holds. Therefore,

$$\begin{aligned}
& \log_N \mathbb{E}|\mathcal{I}(B_x)| - \log_N \mathbb{E}|\mathcal{I}(B_y)| = \frac{1}{\log N} \sum_{M_y < t \leq M_x} \log(n_{x[t]}^{(t)}/N) \\
&= \frac{1}{\log N} \sum_{M_y < t \leq M_x} \sum_{(i,l) \in T} 1_{\{x[t]=l, \mathbf{n}^{(t)} \in D_i\}} \log(n_{x[t]}^{(t)}/N) + \sum_{M_y < t \leq M_x: (t, x[t]) \in \mathcal{Q}} \log_N(N/n_{x[t]}^{(t)}) \\
&= \frac{1}{\log N} \sum_{(i,l) \in T} c_l^{(i)}(x) \log N \left(\log(p_l^{(i)}) + O(\chi) \right) + \sum_{M_y < t \leq M_x: (t, x[t]) \in \mathcal{Q}} \log_N(N/n_{x[t]}^{(t)}) \\
&= \sum_{(i,l) \in T} c_l^{(i)}(x) \log(p_l^{(i)}) + \sum_{M_y < t \leq M_x: (t, x[t]) \in \mathcal{Q}} \log_N(N/n_{x[t]}^{(t)}) + O(\chi),
\end{aligned}$$

Here we used the fact that $\sum_{(i,l) \in T} c_l^{(i)}(x) = c_{\text{tot}}(x)$ is bounded by $K/\log N$. (4.13) then holds by the fact

$$\log_N \mathbb{E}|\mathcal{I}(B_y)| = 1 - \sum_{t \leq M_y} \log_N(N/n_{x[t]}^{(t)}).$$

Now we prove (4.14). Again, note that if $\mathbf{n}^{(t)}/N \in D_i$ for some $0 \leq i \leq z$, then $\|\mathbf{n}^{(t)}/N - \mathbf{p}^{(i)}\|_\infty < \chi$, so

$$\left| \psi_{\mathbf{n}^{(t)}/N}(2) - \psi_{\mathbf{p}^{(i)}}(2) \right| < O(\chi).$$

Recall the definition of t_* in (2.2). Suppose that j_1 is the smallest integer such that $t_* + j_1 \rho \log N > M_x$, and j_2 is the largest integer such that $t_* + j_2 \rho \log N < K$. Thus we have $(j_2 - j_1)\rho < \frac{M_x - K}{\log N} < (j_2 - j_1)\rho + 2\rho$. In addition, we see that $\psi_{\mathbf{p}}(2)$ is bounded by $\log k$ for all $\mathbf{p} \in \mathcal{D}_k$. Therefore,

$$\begin{aligned}
c_E(x) &= -\frac{1}{\log N} \sum_{t_* + j_1 \rho \log N < t \leq t_* + j_2 \rho \log N} \psi_{\mathbf{n}^{(t)}/N}(2) + O(\rho) \\
&= -\frac{1}{\log N} \sum_{t_* + j_1 \rho \log N < t \leq t_* + j_2 \rho \log N} \sum_i 1_{\{\mathbf{n}^{(t)}/N \in D_i\}} \psi_{\mathbf{p}^{(i)}}(2) + O(\rho) + O(\chi) \\
&= -(j_2 - j_1)\rho \sum_{0 \leq i \leq z} h_i \psi_{\mathbf{p}^{(i)}}(2) + O(\rho) + O(\chi) + O(\varphi z) \\
&= \left(\frac{M_x - K}{\log N} \right) \sum_{0 \leq i \leq z} h_i \psi_{\mathbf{p}^{(i)}}(2) + O(\chi).
\end{aligned}$$

For (4.15), suppose that j_3 is the smallest integer such that $t_* + j_3 \rho \log N > M_y$, and j_4 is the largest integer such that $t_* + j_4 \rho \log N < M_x$. We have

$$\begin{aligned}
c_{\text{tot}}^{(i)}(x) + \frac{1}{\log N} \sum_{M_y < t \leq M_x: (t, x[t]) \in \mathcal{Q}} 1_{\{\mathbf{n}^{(t)}/N \in D_i\}} &= \frac{1}{\log N} \sum_{M_y < t \leq M_x} 1_{\{\mathbf{n}^{(t)}/N \in D_i\}} \\
&= \frac{1}{\log N} \sum_{t_* + j_3 \rho \log N < t \leq t_* + j_4 \rho \log N} 1_{\{\mathbf{n}^{(t)}/N \in D_i\}} + O(\rho) \\
&= h_i(j_4 - j_3)\rho + O(\rho) + O(\varphi) \\
&= \frac{h_i}{\log N} (M_x - M_y) + O(\rho) + O(\varphi),
\end{aligned}$$

which completes the proof. $\square$

4.3 Estimating the first moment

Now we control the probability that the edge $(i, i+1)$ belongs to $E(G, G')$. As in [Sel22], the idea is to find a prefix x such that $\mathcal{I}(B_x)$ typically contains $(i, i+1)$, but the location of $\mathcal{I}(B_x)$ has fluctuations of only

slightly smaller order than the typical size $|\mathcal{I}(B_x)|$ (recall the definition (2.14) of $\mathcal{I}(B_x)$). The fraction of edges near i can then be modelled as locally homogeneous, with local density depending on the digit profile of x . Note that x of course depends on i , and in fact will identify not just a single x but a small set of them. In preparation for this argument, we construct a tree-structured partition of S_K based on prefixes. First let $\mathcal{T}$ be a k -ary tree of depth K , which consists of all strings in $[k]_0^M$ for $0 \leq M \leq K$. Here the children of $s \in [k]_0^M$ are $[si]$ for $i = 0, 1, \dots, k-1$, and we let $\mathcal{F}x$ be the parent of any vertex x in $\mathcal{T}$. We then partition $\mathcal{T}$ as follows. For $y \in \mathcal{A}_{\text{left}}$ and $z \in \mathcal{A}_{\text{right}}$ such that $A_{y,z}$ is non-empty, suppose by symmetry that z is a prefix of y . Therefore, define

$$A_{y,z}^{\text{stable}} = \begin{cases} \{y\}, & \text{if } c_L(y) - c_F(y) < \delta; \\ \{x \in A_{y,z} : c_L(x) - c_F(x) < \delta \text{ and } c_L(\mathcal{F}x) - c_F(y) \geq \delta\}, & \text{if } c_L(y) - c_F(y) \geq \delta. \end{cases}$$

It is useful to point out that $c_L(x)$ is non-increasing down $\mathcal{T}$, and when $c_L(y) - c_F(y) \geq \delta$ and $x \in A_{y,z}^{\text{stable}}$,

$$c_F(y) = c_F(x) = c_F(\mathcal{F}x).$$

With the notation above, have the following partition lemma.

Lemma 4.9. *The following (resp. deterministic and random) partitions (i.e. disjoint unions) hold:*

$$[k]_0^K = \bigcup_{y \in \mathcal{A}_{\text{left}}, z \in \mathcal{A}_{\text{right}}} \bigcup_{x \in A_{y,z}^{\text{stable}}} B_x \quad \text{and} \quad [N] = \bigcup_{y \in \mathcal{A}_{\text{left}}, z \in \mathcal{A}_{\text{right}}} \bigcup_{x \in A_{y,z}^{\text{stable}}} \mathcal{I}(B_x). \quad (4.17)$$

Proof. By definition of $\mathcal{A}_{\text{left}}$ and $\mathcal{A}_{\text{right}}$, we have

$$[k]_0^K = \bigcup_{y \in \mathcal{A}_{\text{left}}, z \in \mathcal{A}_{\text{right}}} \bigcup_{x \in A_{y,z}} B_x.$$

It suffices to establish the partition

$$\bigcup_{x \in A_{y,z}} B_x = \bigcup_{x \in A_{y,z}^{\text{stable}}} B_x.$$

When $c_L(y) - c_F(y) < \delta$ the claim is obvious. When $c_L(y) - c_F(y) \geq \delta$, For $x' \in A_{y,z}$, the value $c_L(x') - c_F(x') = c_L(x') - c_F(y)$ is decreasing along any path in $\mathcal{T}$ from y to a leaf. When $x = y$ is the root, we have $c_L(x) - c_F(x) \geq \delta$. It suffices to check that $c_L(s) - c_F(s) < \delta$ for any leaf $s \in A_{y,z}$. In fact, by (4.12) and (4.13) we have

$$c_L(s) - c_F(s) \leq \frac{1}{2}c_L(s) + o(1) \leq \frac{1}{2} \log_N \mathbb{E}|\mathcal{I}(B_s)| - o(1).$$

Since $K \geq (\tilde{C}_\mu + \varepsilon) \log N$, we have

$$\begin{aligned} \log_N \mathbb{E}|\mathcal{I}(B_s)| &\leq 1 - \sum_{1 \leq t \leq K} \log_N(N/n_{l_{\max}^{(t)}}^{(t)}) = 1 - \frac{K}{\log N} \sum_{0 \leq i \leq z} h_i \log(1/p_{\max}^{(i)}) \pm O(\delta) \\ &= 1 - \frac{K \mathbb{E}_\mu \log(1/p_{\max}) \pm O(\delta)}{\log N} < 0, \end{aligned}$$

which concludes the proof. $\square$

Lemma 4.10. *For each index $i \in [N]$ and for each $y \in \mathcal{A}_{\text{left}}, z \in \mathcal{A}_{\text{right}}$, there exists a subset $A_{y,z,i}^{\text{stable}} \subseteq A_{y,z}^{\text{stable}}$ such that*

1. *The size of the subset is bounded: $|A_{y,z,i}^{\text{stable}}| \leq 2k$.*
2. *We have*

$$\mathbb{P}\left[i \in \bigcup_{y \in \mathcal{A}_{\text{left}}, z \in \mathcal{A}_{\text{right}}} \bigcup_{x \in A_{y,z,i}^{\text{stable}}} \mathcal{I}(B_x)\right] \geq 1 - e^{-\Omega(N^\delta)}.$$

Proof. For each $y \in \mathcal{A}_{\text{left}}, z \in \mathcal{A}_{\text{right}}$, consider the set

$$A_{y,z,i}^{\text{stable}} = \{x \in A_{y,z}^{\text{stable}}, \mathbb{P}(i \in \mathcal{I}(B_x)) \geq e^{-N^\delta}\}.$$

By (4.17), we have immediately that

$$\mathbb{P}\left[i \notin \bigcup_{y \in \mathcal{A}_{\text{left}}, z \in \mathcal{A}_{\text{right}}} \bigcup_{x \in A_{y,z,i}^{\text{stable}}} \mathcal{I}(B_x)\right] \leq e^{-\Omega(N^\delta)} \cdot |\mathcal{A}_{\text{left}}| \cdot |\mathcal{A}_{\text{right}}| \cdot \sup_{y \in \mathcal{A}} |A_{y,z}^{\text{stable}}|$$

Claim 2 follows directly from the fact that $|\mathcal{A}_{\text{left}}| \vee |\mathcal{A}_{\text{right}}| \vee |A_{y,z}^{\text{stable}}| \leq k^K = N^{O(1)}$. It suffices to prove that $|A_{y,z,i}^{\text{stable}}| \leq 2k$.

Without loss of generality, we assume that z is a prefix of y and $c_L(y) - c_F(y) \geq \delta$ (otherwise $|A_{y,z,i}^{\text{stable}}| \leq 1$). Consider the map (recall that $\mathcal{F}x$ is the parent of x)

$$\begin{aligned} \Phi : A_{y,z,i}^{\text{stable}} &\rightarrow \mathcal{T}; \\ x &\mapsto \mathcal{F}x, \end{aligned}$$

we claim that the range of Φ , denoted $\text{Range}(\Phi)$, satisfies $|\text{Range}(\Phi)| \leq 2$. Therefore, since for each $y \in \text{Range}(\Phi)$, we have $|\Phi^{-1}(y)| \leq k$, we conclude Claim 1.

We prove the claim on $\text{Range}(\Phi)$ by contradiction. Suppose that there are 3 strings y_1, y_2, y_3 in $\text{Range}(\Phi)$, such that $\mathbb{P}(i \in \mathcal{I}(B_{y_j})) \geq e^{-N^\delta}$ for each $j \in \{1, 2, 3\}$. By Lemma 4.9 we see J_{y_i} 's are disjoint. Therefore, we suppose without loss of generality that $J_{y_1}, J_{y_2}, J_{y_3}$ lies from left to right on the interval $[0, 1]$, and $\frac{i}{N}$ is not larger than the midpoint of J_{y_2} . Thus we have $i < (\iota_{y_2} + \tau_{y_2})/2$ and $\mathbb{P}(i \in \mathcal{I}(B_{y_3})) \leq \mathbb{P}(\iota(y_2) < i)$. Note that by the definition of $A_{y,z,i}^{\text{stable}}$, we have $c_L(\mathcal{F}x) - c_F(\mathcal{F}x) \geq \delta$, which by (4.12) implies that $c_F(\mathcal{F}x) > \delta - o(1)$ and $c_L(\mathcal{F}x) > 2\delta - o(1)$. By Lemma 4.4, we obtain that $\mathbb{P}(\iota(y_2) < i) \leq \mathbb{P}(|\iota(y_2) - \mathbb{E}\iota(y_2)| \geq N^{c_L(y_2)}/2) \leq e^{-\Omega(N^{2\delta})} \leq e^{-N^\delta}$, which leads to a contradiction. Thus $\text{Range}(\Phi)$ includes at most two elements as desired. $\square$

Lemma 4.11. *Recall the definition of $\tilde{\mathcal{A}}_{\text{left}}$ and $\tilde{\mathcal{A}}_{\text{right}}$ in Definition 4.3. With all assumptions in Proposition 4.2, we have*

$$|\tilde{\mathcal{A}}_{\text{left}}| \vee |\tilde{\mathcal{A}}_{\text{right}}| \leq N^\delta. \quad (4.18)$$

Proof. For fixed M with $1 < M \leq K$, we first count the number of strings x in $\tilde{\mathcal{A}}_{\text{left}} \cap [k]_0^M$. Notice that if we fix the set $\{t : 1 < t < K, x[t] < \mathbf{I}_{\text{left}}^{(t)}\}$, then there are at most k choices of x (because of the first digit). Therefore, we have

$$|\tilde{\mathcal{A}}_{\text{left}}| \leq \sum_{0 \leq i \leq \frac{C}{\log(1/\chi)} \log N} \binom{K}{i} \cdot k^{\frac{C}{\log(1/\chi)} \log N} \cdot k \leq N^\delta.$$

Here we used the fact that $\frac{C}{\log(1/\chi)} \ll \delta$. By symmetry the same inequality holds for $|\tilde{\mathcal{A}}_{\text{right}}|$. $\square$

Lemma 4.12. *With all assumptions in Proposition 4.2, we have for all $1 \leq i \leq N-1$,*

$$\mathbb{P} \left[(i, i+1) \in \bigcup_{\substack{(y,z) \in (\mathcal{A}_{\text{left}} \times \mathcal{A}_{\text{right}}); \\ (y,z) \notin (\tilde{\mathcal{A}}_{\text{left}} \times \tilde{\mathcal{A}}_{\text{right}})}} \bigcup_{x \in A_{y,z,i}^{\text{stable}}} E(G_{B_x}) \right] \leq N^{-C/2}.$$

Proof. For $(y, z) \in (\mathcal{A}_{\text{left}} \times \mathcal{A}_{\text{right}}) \setminus (\tilde{\mathcal{A}}_{\text{left}} \times \tilde{\mathcal{A}}_{\text{right}})$, suppose without loss of generality that z is a prefix of y . We have $y \in \mathcal{A}_{\text{left}} \setminus \tilde{\mathcal{A}}_{\text{left}}$, and

$$\begin{aligned} \mathbb{P} \left[(i, i+1) \in \bigcup_{x \in A_{y,z,i}^{\text{stable}}} G(B_x) \right] &\leq \mathbb{P}[(i, i+1) \in G(B_y)] \leq \mathbb{E}|\mathcal{I}(B_y)| \\ &\leq N \cdot \prod_{t: y[t] < \mathbf{I}_{\text{left}}^{(t)}} \frac{n_{y[t]}}{N} \leq N \cdot \chi^{\frac{C}{\log(1/\chi)} \log N} \leq N^{-C+1}. \end{aligned}$$

Therefore, we conclude since $|\mathcal{A}_{\text{left}} \times \mathcal{A}_{\text{right}}| \leq k^{2K} \leq N^{C/3}$, the former via a union bound. $\square$

Lemma 4.13. *With all assumptions in Proposition 4.2, we have*

$$\mathbb{E}[|E(G, G')|] \leq N^{-C/2} + N^{O(\delta)} \cdot \sum_{y \in \tilde{\mathcal{A}}_{\text{left}}, z \in \tilde{\mathcal{A}}_{\text{right}}} \sum_{x \in A_{y,z,i}^{\text{stable}}} \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2. \quad (4.19)$$

Proof. By definition of $E(G, G')$ we have

$$\mathbb{E}[|E(G, G')|] = \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G, G')] = \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G)]^2.$$

Recall the definition of $A_{y,z,i}^{\text{stable}}$ in Lemma 4.10, we have

$$\mathbb{P}[(i, i+1) \in E(G)] \leq e^{-\Omega(N^\delta)} + \mathbb{P}\left[(i, i+1) \in \bigcup_{y \in \tilde{\mathcal{A}}_{\text{left}}, z \in \tilde{\mathcal{A}}_{\text{right}}} \bigcup_{x \in A_{y,z,i}^{\text{stable}}} E(G_{B_x})\right].$$

By Lemma 4.12 and a union bound, we have

$$\begin{aligned} \mathbb{P}[(i, i+1) \in E(G)] &\leq N^{-C/2} + \mathbb{P}\left[(i, i+1) \in \bigcup_{y \in \tilde{\mathcal{A}}_{\text{left}}, z \in \tilde{\mathcal{A}}_{\text{right}}} \bigcup_{x \in A_{y,z,i}^{\text{stable}}} E(G_{B_x})\right] \\ &\leq N^{-C/2} + \sum_{y \in \tilde{\mathcal{A}}_{\text{left}}, z \in \tilde{\mathcal{A}}_{\text{right}}} \sum_{x \in A_{y,z,i}^{\text{stable}}} \mathbb{P}[(i, i+1) \in E(G_{B_x})]. \end{aligned}$$

Therefore, by (4.18) and the AM-GM inequality, we have

$$\begin{aligned} &\sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G)]^2 \\ &\leq \sum_{i=1}^{N-1} \left[N^{-C} + \sum_{y \in \tilde{\mathcal{A}}_{\text{left}}, z \in \tilde{\mathcal{A}}_{\text{right}}} \sum_{x \in A_{y,z,i}^{\text{stable}}} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 \right] \cdot N^{O(\delta)} \\ &\leq N^{-C/2} + N^{O(\delta)} \cdot \sum_{y \in \tilde{\mathcal{A}}_{\text{left}}, z \in \tilde{\mathcal{A}}_{\text{right}}} \sum_{x \in A_{y,z}^{\text{stable}}} \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2, \end{aligned}$$

which completes the proof. $\square$

Now it suffices to upper bound for each $(y, z) \in \tilde{\mathcal{A}}_{\text{left}} \times \tilde{\mathcal{A}}_{\text{right}}$ the sum

$$\sum_{x \in A_{y,z}^{\text{stable}}} \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2. \quad (4.20)$$

Lemma 4.14. *With all assumptions in Proposition 4.2, we have for any $x \in [k]_0^M$,*

$$\mathbb{E}(|E(G_{B_x})| \mid |\mathcal{I}(B_x)|) \leq |\mathcal{I}(B_x)|^2 N^{c_E(x) + o(1)}.$$

Proof. It is easy to see that for $x \in [k]_0^M$, and for a random pair of rows s, s' ,

$$\mathbb{P}[s = s' \mid s, s' \in B_x] = \prod_{M < t \leq K} \left(\sum_i \frac{n_i^{(t)}(n_i^{(t)} - 1)}{N(N-1)} \right) \leq N^{c_E(x) + o(1)}.$$

Summing over at most $|\mathcal{I}(B_x)|^2$ pairs of pre-sorted strings in B_x completes the proof. $\square$

Lemma 4.15. *With all assumptions in Proposition 4.2, we have for all $y \in \tilde{\mathcal{A}}_{\text{left}}, z \in \tilde{\mathcal{A}}_{\text{right}}$ and $x \in A_{y,z}^{\text{stable}}$, the following holds.*

(i) If $c_L(x) \geq 2\delta$, we have

$$\sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 \leq CN^{4c_L(x)-c_F(x)+2c_E(x)+O(\delta)} + e^{-\Omega(N^{\delta-o(1)})}. \quad (4.21)$$

(ii) If $c_L(x) < 2\delta$, $c_F \geq \delta - (\log N)^{-1/2}$ and $A_{y,z}^{\text{stable}} \neq \{y \vee z\}$, we have

$$\sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 \leq CN^{-c_F(x)+2c_E(x)+O(\delta)} + e^{-\Omega(N^{\delta-o(1)})}. \quad (4.22)$$

(iii) If $c_L(x) < 2\delta$ and $c_F(x) < \delta - (\log N)^{-1/2}$, then $x = y \vee z$ and $A_{y,z}^{\text{stable}} = \{y \vee z\}$.

(iv) If $A_{y,z}^{\text{stable}} = \{y \vee z\}$ and $c_L(y \vee z) < 2\delta$, we have $x = y \vee z$ and

$$\sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 \leq CN^{2c_L(x)+2c_E(x)+O(\delta)} + e^{-\Omega(N^{\delta-o(1)})}. \quad (4.23)$$

The idea behind Lemma 4.15 is that conditional on the internal structure of $G(B_x)$ (i.e. the multiset of strings with prefix x), the location of $\mathcal{I}(B_x)$ will be random with translational fluctuations of order $N^{c_F(x)}$. By definition of $A_{y,z}^{\text{stable}}$, this is of almost the same order as $|B_x|$ itself. Therefore one can expect to understand $\mathbb{P}[(i, i+1) \in E(G_{B_x})]$ based on the local edge density estimated in Lemma 4.14. This is significantly easier in the binomial case because the random translation is a binomial random variable with fairly simple parameters, but requires much more care in our setting. Thus to prove Lemma 4.15, we rely on Lemma 4.5, an anti-concentration result for conditional hypergeometric distributions.

Proof of Lemma 4.15. Recall the definition of U_x in (4.9) and we take the event $\mathcal{S}'$ from Lemma 4.5.

Claim (i). Since $x \in A_{y,z}^{\text{stable}}$, we have $c_L(x) - c_F(x) < \delta$. By Lemma 4.14 we have

$$\mathbb{E}(|E(G_{B_x})| \mid |\mathcal{I}(B_x)|) \leq |\mathcal{I}(B_x)|^2 N^{c_E(x)+o(1)}.$$

Furthermore, since we have by (4.12) and the assumption $c_L(x) \geq 2\delta$ that

$$c_F(x) \geq c_L(y(x))/2 - o(1) \geq c_L(x)/2 - o(1) \geq \delta - o(1),$$

Lemma 4.5 implies that $\max_{j \in [N]} \mathbb{P}(\iota(x) = j, S_K \in \mathcal{S}' | U_x) = O(N^{-c_F(x)+2\delta})$. Therefore,

$$\mathbb{P}((i, i+1) \in E(G_{B_x}), S_K \in \mathcal{S}' \mid |\mathcal{I}(B_x)|) \leq C|\mathcal{I}(B_x)|^2 N^{c_E(x)-c_F(x)+O(\delta)}. \quad (4.24)$$

For $i \in [Nt_x - N^{c_F(x)+2\delta}, N(t_x + \lambda_x) + N^{c_F(x)+2\delta}]$, Lemma 2.6 shows that

$$\mathbb{P}(|\mathcal{I}(B_x)| \geq 2N^{c_L(x)}) \leq e^{-\Omega(N^{\delta}-o(1))},$$

and thus

$$\mathbb{P}[(i, i+1) \in E(G_{B_x})] \leq CN^{2c_L(x)-c_F(x)+c_E(x)+O(\delta)} + e^{-\Omega(N^{\delta}-o(1))}.$$

While for $i \notin [Nt_x - N^{c_F(x)+2\delta}, N(t_x + \lambda_x) + N^{c_F(x)+2\delta}]$, since $c_L(x) - c_F(x) < \delta$, (4.6) and (4.7) imply

$$\mathbb{P}[(i, i+1) \in E(G_{B_x})] \leq e^{-\Omega(N^{\delta}-o(1))},$$

where we used $c_F(x) \geq \delta - o(1)$. We conclude that

$$\begin{aligned} & \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 \\ & \leq 2N^{c_F(x)+2\delta} \left(CN^{2c_L(x)-c_F(x)+c_E(x)+O(\delta)} + e^{-\Omega(N^{\delta}-o(1))} \right)^2 + Ne^{-\Omega(N^{\delta}-o(1))} \\ & \leq CN^{4c_L(x)-c_F(x)+2c_E(x)+O(\delta)} + e^{-\Omega(N^{\delta}-o(1))}. \end{aligned}$$

Claim (ii). By Lemma 2.7, we have that $\mathbb{P}(|\mathcal{I}(B_x)| \geq 2N^{2\delta}) < \exp(-\Omega(N^\delta))$. For any $x \in \mathcal{A}_{y,z}^{\text{stable}}$ with $c_L(x) < 2\delta$, $c_F(x) \geq \delta - (\log N)^{-1/2}$ and $i \in [Nt_x - N^{c_F(x)+2\delta}, N(t_x + \lambda_x) + N^{c_F(x)+2\delta}]$, by (4.24) we have

$$\mathbb{P}[(i, i+1) \in E(G_{B_x})] \leq CN^{-c_F(x)+c_E(x)+O(\delta)} + e^{-\Omega(N^{\delta-o(1)})}.$$

Meanwhile for $i \notin [Nt_x - N^{c_F(x)+2\delta}, N(t_x + \lambda_x) + N^{c_F(x)+2\delta}]$, since $c_L(x) - c_F(x) < \delta$, (4.6) and (4.7) imply

$$\mathbb{P}[(i, i+1) \in E(G_{B_x})] \leq e^{-\Omega(N^{\delta-o(1)})},$$

where we used $c_F(x) \geq \delta - (\log N)^{-\frac{1}{2}}$. We conclude that

$$\begin{aligned} & \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 \\ & \leq 2N^{c_F(x)+2\delta} \left(CN^{-c_F(x)+c_E(x)+O(\delta)} + e^{-\Omega(N^{\delta-o(1)})} \right)^2 + Ne^{-\Omega(N^{\delta-o(1)})} \\ & \leq CN^{-c_F(x)+2c_E(x)+O(\delta)} + e^{-\Omega(N^{\delta-o(1)})}. \end{aligned}$$

Claims (iii) and (iv). Without loss of generality we assume that z is a prefix of y . By (4.5), we have $c_L(y) \leq 2c_F(x) + O((\log N)^{-1})$. Combining this with $c_F(y) = c_F(x) \leq \delta - (\log N)^{-1/2}$, we obtain

$$c_L(y) - c_F(y) < \delta.$$

Thus, we obtain that $A_{y,z}^{\text{stable}} = \{y\}$ by the definition of $A_{y,z}^{\text{stable}}$ and conclude Claim (iii).

Lemma 4.5 implies that $\max_{j \in [N]} \mathbb{P}(\iota(x) = j, S_K \in \mathcal{S}'|U_x) = O(N^{-\max(c_F(x), \delta)+2\delta})$. Therefore, by Lemma 4.14 we have

$$\mathbb{P}((i, i+1) \in E(G_{B_x}), S_K \in \mathcal{S}' \mid |\mathcal{I}(B_x)|) \leq |\mathcal{I}(B_x)|^2 N^{-\max(c_F(x), \delta)+c_E(x)+O(\delta)}.$$

By Lemma 2.7 and the assumption that $c_L(x) < 2\delta$, we have $\mathbb{P}(|\mathcal{I}(B_x)| \geq N^{4\delta}) \leq e^{-\Omega(N^{\delta-o(1)})}$, and thus

$$\mathbb{E}|\mathcal{I}(B_x)|^2 \leq N^{4\delta} \mathbb{E}|\mathcal{I}(B_x)| + e^{-\Omega(N^\delta)} \leq N^{c_L(x)+O(\delta)} + e^{-\Omega(N^{\delta-o(1)})}.$$

Therefore,

$$\mathbb{P}((i, i+1) \in E(G_{B_x})) \leq N^{c_L(x)-\max(c_F(x), \delta)+c_E(x)+O(\delta)} + e^{-\Omega(N^{\delta-o(1)})}.$$

Now, for $i \in [Nt_x - N^{\max(c_F(x), \delta)+2\delta}, N(t_x + \lambda_x) + N^{\max(c_F(x), \delta)+2\delta}]$, we have

$$\mathbb{P}[(i, i+1) \in E(G_{B_x})] \leq N^{c_L(x)-\max(c_F(x), \delta)+c_E(x)+O(\delta)} + e^{-\Omega(N^{\delta-o(1)})},$$

While for $i \notin [Nt_x - N^{\max(c_F(x), \delta)+2\delta}, N(t_x + \lambda_x) + N^{\max(c_F(x), \delta)+2\delta}]$, (4.6) and (4.7) imply

$$\mathbb{P}[(i, i+1) \in E(G_{B_x})] \leq e^{-\Omega(N^{\delta-o(1)})}.$$

We conclude that

$$\begin{aligned} & \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 \\ & \leq \left(N^{c_L(x)+O(\delta)} + N^{\max(c_F(x), \delta)+O(\delta)} \right) \left(N^{c_L(x)-\max(c_F(x), \delta)+c_E(x)+O(\delta)} \right)^2 + Ne^{-\Omega(N^{\delta-o(1)})} \\ & \leq CN^{2c_L(x)+2c_E(x)+O(\delta)} + e^{-\Omega(N^{\delta-o(1)})}. \end{aligned} \quad \square$$

To apply Lemma 4.15, we need the following numerical lemma.

Lemma 4.16. *If $\delta, \chi, \varepsilon$ are as in (2.9), the following properties hold for each $y \in \mathcal{A}$.*

(a) *If $x \in A_{y,z}^{\text{stable}}$, then we have*

$$4c_L(x) - c_F(x) + 2c_E(x) + \sum_{0 \leq i \leq z} c_{\text{tot}}^{(i)} H(\{c_l^{(i)}\}_{(i,l) \in T}) \leq -\Omega(\varepsilon). \quad (4.25)$$

(b) If $x \in A_{y,z}^{\text{stable}}$ and $A_{y,z}^{\text{stable}} \neq \{y \vee z\}$, then we have

$$-c_F(x) + 2c_E(x) + \sum_{0 \leq i \leq z} c_{\text{tot}}^{(i)} H(\{c_l^{(i)}\}_{(i,l) \in T}) \leq -\Omega(\varepsilon). \quad (4.26)$$

(c) If $A_{y,z}^{\text{stable}} = \{y \vee z\}$, we have

$$c_E(y \vee z) + c_L(y \vee z) \leq -\Omega(\varepsilon). \quad (4.27)$$

Proof. We assume without loss of generality that z is a prefix of y . Recall M_y is the length of y .

Claim (a). To deal with the entropy term, we use the non-negativity of Kullback-Leibler divergence. For fixed i , define

$$\mathbf{c}^{(i)} = \left(\frac{c_l^{(i)}}{c_{\text{tot}}^{(i)}} \right)_{l:(i,l) \in T} \quad \text{and} \quad \mathbf{p}^{(i)} = \left(\frac{(p_l^{(i)})^{\theta_\mu}}{\sum_{l:(i,l) \in T} (p_l^{(i)})^{\theta_\mu}} \right)_{l:(i,l) \in T}. \quad (4.28)$$

We have:

$$\begin{aligned} H(\{c_l^{(i)}\}_{(i,l) \in T}) &= \sum_{l:(i,l) \in T} \frac{c_l^{(i)}}{c_{\text{tot}}^{(i)}} \log \left(\frac{\sum_{l:(i,l) \in T} (p_l^{(i)})^{\theta_\mu}}{(p_l^{(i)})^{\theta_\mu}} \right) - D_{\text{KL}}(\mathbf{c}^{(i)}, \mathbf{p}^{(i)}) \\ &\leq \sum_{l:(i,l) \in T} \frac{c_l^{(i)}}{c_{\text{tot}}^{(i)}} \log \left(\frac{\sum_{l:(i,l) \in T} (p_l^{(i)})^{\theta_\mu}}{(p_l^{(i)})^{\theta_\mu}} \right) \\ &= \log \left(\sum_{l:(i,l) \in T} (p_l^{(i)})^{\theta_\mu} \right) + \theta_\mu \sum_{l:(i,l) \in T} \frac{c_l^{(i)}}{c_{\text{tot}}^{(i)}} \log \left(\frac{1}{p_l^{(i)}} \right) \\ &= -\psi_{\mathbf{p}^{(i)}}(\theta_\mu) + \theta_\mu \sum_{l:(i,l) \in T} \frac{c_l^{(i)}}{c_{\text{tot}}^{(i)}} \log \left(\frac{1}{p_l^{(i)}} \right) \pm O(\chi). \end{aligned} \quad (4.29)$$

Combining (4.15) and (4.29), we have

$$\begin{aligned} \sum_{0 \leq i \leq z} c_{\text{tot}}^{(i)} H(\{c_l^{(i)}\}_{(i,l) \in T}) &\leq -c_{\text{tot}} \sum_{0 \leq i \leq z} h_i \psi_{\mathbf{p}^{(i)}}(\theta_\mu) + \theta_\mu \sum_{0 \leq i \leq z} \sum_{l:(i,l) \in T} c_l^{(i)} \log \left(\frac{1}{p_l^{(i)}} \right) + O(\delta) \\ &\leq -c_{\text{tot}} \psi_\mu(\theta_\mu) + \theta_\mu \sum_{0 \leq i \leq z} \sum_{l:(i,l) \in T} c_l^{(i)} \log \left(\frac{1}{p_l^{(i)}} \right) + O(\delta). \end{aligned}$$

Here we used $\rho, \chi, \varphi \ll \delta$. Recall that $\psi_\mu(\theta_\mu) = 2\psi_\mu(2)$, we have by (4.14),

$$\begin{aligned} 2c_E(x) + \sum_{0 \leq i \leq z} c_{\text{tot}}^{(i)} H(\{c_l^{(i)}\}_{(i,l) \in T}) \\ \leq 2\psi_\mu(2) \left(\frac{M_x - K}{\log N} - c_{\text{tot}} \right) + \theta_\mu \sum_{0 \leq i \leq z} \sum_{l:(i,l) \in T} c_l^{(i)} \log \left(\frac{1}{p_l^{(i)}} \right) + O(\delta). \end{aligned} \quad (4.30)$$

Meanwhile, we set

$$q \equiv \sum_{M_y < t \leq M_x: (t, x[t]) \in Q} \log_N \left(N/n_{x[t]}^{(t)} \right) \geq 0. \quad (4.31)$$

Then we have:

$$c_L(x) = c_L(y) - \sum_{0 \leq i \leq z} \sum_{l:(i,l) \in T} c_l^{(i)} \log \left(\frac{1}{p_l^{(i)}} \right) - q + O(\chi), \quad (4.32)$$

$$c_F(x) = c_F(y) \geq \frac{1}{2} c_L(y) - o(1) \quad (4.33)$$

$$c_L(x) - c_F(x) = c_L(y) - c_F(y) - \sum_{0 \leq i \leq z} \sum_{l:(i,l) \in T} c_l^{(i)} \log \left(\frac{1}{p_l^{(i)}} \right) - q + O(\chi) < \delta. \quad (4.34)$$

By (4.34), we obtain

$$\sum_{0 \leq i \leq z} \sum_{l: (i,l) \in T} c_l^{(i)} \log \left(\frac{1}{p_l^{(i)}} \right) \geq c_L(y) - c_F(y) - q - \delta - O(\chi). \quad (4.35)$$

Recall that by (1.5), we have $\theta_\mu \in [3, 4)$. Therefore, we have

$$\begin{aligned} & 4c_L(x) - c_F(x) + \theta_\mu \sum_{0 \leq i \leq z} \sum_{l: (i,l) \in T} c_l^{(i)} \log \left(\frac{1}{p_l^{(i)}} \right) \\ & \stackrel{(4.32), (4.33)}{=} 4c_L(y) - c_F(y) - 4q + (\theta_\mu - 4) \sum_{0 \leq i \leq z} \sum_{l: (i,l) \in T} c_l^{(i)} \log \left(\frac{1}{p_l^{(i)}} \right) \\ & \stackrel{(4.35), (1.5)}{\leq} -\theta_\mu q + \theta_\mu c_L(y) + (3 - \theta_\mu) c_F(y) + O(\delta) \\ & \stackrel{(1.5), (4.31), (4.33)}{\leq} \frac{\theta_\mu + 3}{2} c_L(y) + O(\delta). \end{aligned} \quad (4.36)$$

Here we used the fact that $\chi \ll \delta$. By the definition of M_y , we have

$$c_L(y) \leq 1 - \left(\frac{M_y}{\log N} - O(\rho) \right) \mathbb{E}_\mu \log(1/p_{\max}) + O(\delta). \quad (4.37)$$

Combining (4.30), (4.36) and (4.37), we obtain that

$$4c_L(x) - c_F(x) + 2c_E(x) + \sum_{0 \leq i \leq z} c_{\text{tot}}^{(i)} H(\{c_l^{(i)}\}_{(i,l) \in T}) \leq f(M_y) + O(\delta),$$

where

$$f(M_y) = \frac{\theta_\mu + 3}{2} \left[1 - \frac{M_y}{\log N} \cdot \mathbb{E}_\mu \log(1/p_{\max}) \right] + 2\psi_\mu(2) \cdot \frac{M_y - K}{\log N}.$$

Since $f(M_y)$ is a linear function for $M_y \in [0, K]$ and $\delta \ll \varepsilon$, it suffices to verify that $f(0) \leq -\Omega(\varepsilon)$ and $f(K) \leq -\Omega(\varepsilon)$. In fact, recall the definition of $C_\mu, \tilde{C}_\mu$ in (1.6), we have

$$\begin{aligned} f(0) &= \frac{\theta_\mu + 3}{2} - 2\psi_\mu(2) \cdot \frac{K}{\log N} = \frac{\theta_\mu + 3}{2} \left(1 - \frac{K}{C_\mu \log N} \right); \\ f(K) &= \frac{\theta_\mu + 3}{2} \left[1 - \frac{K}{\log N} \cdot \mathbb{E}_\mu \log(1/p_{\max}) \right] = \frac{\theta_\mu + 3}{2} \left(1 - \frac{K}{\tilde{C}_\mu \log N} \right). \end{aligned}$$

Therefore, we can obtain the desired result by noticing that $\frac{K}{\log N} \geq \bar{C}_\mu + \varepsilon$, where $\bar{C}_\mu = \max\{C_\mu, \tilde{C}_\mu\}$.

Claim (b). We continue using the notations in Claim (a) and find

$$\begin{aligned} & -c_F(x) + 2c_E(x) + \sum_{0 \leq i \leq z} c_{\text{tot}}^{(i)} H(\{c_l^{(i)}\}_{(i,l) \in T}) \\ & \leq -c_F(y) + 2\psi_\mu(2) \frac{M_y - K}{\log N} + \theta_\mu \sum_{0 \leq i \leq z} \sum_{l: (i,l) \in T} c_l^{(i)} \log \left(\frac{1}{p_l^{(i)}} \right). \end{aligned}$$

Recall the definition of $\mathcal{F}x$ above Lemma 4.9. Since we have assumed that $y \notin A_{y,z}^{\text{stable}}$, it holds that $c_F(\mathcal{F}x) = c_F(x)$ and $c_L(\mathcal{F}x) \leq c_L(y)$. We have

$$\begin{aligned} \delta &< c_L(\mathcal{F}x) - c_F(y) < c_L(y) - c_F(y) - q - \sum_{0 \leq i \leq z} \sum_{l: (i,l) \in T} c_l^{(i)} \log \left(\frac{1}{p_l^{(i)}} \right) + o(1) \\ &\stackrel{(4.33)}{\leq} \frac{1}{2} c_L(y) - \sum_{0 \leq i \leq z} \sum_{l: (i,l) \in T} c_l^{(i)} \log \left(\frac{1}{p_l^{(i)}} \right) + o(1). \end{aligned}$$

Thus we have

$$\begin{aligned}
& -c_F(y) + 2\psi_\mu(2) \frac{M_y - K}{\log N} + \theta_\mu \sum_{0 \leq i \leq z} \sum_{l: (i,l) \in T} c_l^{(i)} \log \left(\frac{1}{p_l^{(i)}} \right) \\
& \leq -\frac{1}{2} c_L(y) + 2\psi_\mu(2) \frac{M_y - K}{\log N} + \theta_\mu \sum_{0 \leq i \leq z} \sum_{l: (i,l) \in T} c_l^{(i)} \log \left(\frac{1}{p_l^{(i)}} \right) \\
& \leq \frac{1}{2} (\theta_\mu - 1) c_L(y) + 2\psi_\mu(2) \frac{M_y - K}{\log N} + O(\delta).
\end{aligned}$$

By (4.37) (whose proof does not use the hypothesis of Claim (a)), we have

$$\begin{aligned}
& -c_F(y) + 2\psi_\mu(2) \frac{M_y - K}{\log N} + \theta_\mu \sum_{0 \leq i \leq z} \sum_{l: (i,l) \in T} c_l^{(i)} \log \left(\frac{1}{p_l^{(i)}} \right) \\
& \leq \frac{\theta_\mu - 1}{2} \left[1 - \frac{M_y}{\log N} \cdot \log(1/p_{\max}) \right] + 2\psi_\mu(2) \frac{M_y - K}{\log N} + O(\delta).
\end{aligned}$$

By taking $M_y = 0$ and K in the right-hand side, we conclude the proof of Claim (b).

Claim (c). By (4.14) we have

$$c_E(y) = \left(\frac{M_y - K}{\log N} \right) \psi_\mu(2) + o_\chi(1).$$

By (4.37) we have

$$c_L(y) \leq 1 - \frac{1}{\log N} (M_y - 2\rho \log N) \mathbb{E}_\mu \log(1/p_{\max}) + O(\delta).$$

Therefore, we have

$$c_L(y) + c_E(y) \leq \left(\frac{M_y - K}{\log N} \right) \psi_\mu(2) + 1 - \frac{M_y}{\log N} \mathbb{E}_\mu \log(1/p_{\max}) + O(\delta)$$

By taking $M_y = 0$ and K in the right-hand side, we conclude the proof of Claim (c). $\square$

Proof of Proposition 4.2. By Lemma 4.13, it suffices to bound the sum (4.20). Let

$$\mathbf{q}(x) = |\{(t, x[t]) : (t, x[t]) \in \mathbf{Q}\}|.$$

When $\mathbf{q}(x) > \frac{C \log N}{\log(1/\chi)}$, we have immediately that

$$\sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 \leq N \cdot \mathbb{E} |\mathcal{I}(B_x)|^2 \leq N \cdot \chi^{\frac{2C \log N}{\log(1/\chi)}} \leq N^{-2C+1}.$$

Therefore,

$$\sum_{\mathbf{q}(x) > \frac{C}{\log(1/\chi)}} \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 \leq k^K \cdot N^{-2C+1} \leq N^{-C}. \quad (4.38)$$

It suffices to bound the sum (4.20) for those prefixes x with $\mathbf{q}(x) \leq \frac{C \log N}{\log(1/\chi)}$. We bound this term by grouping prefixes $x \in A_{y,z}^{\text{stable}}$ according to their digit profiles. For $y \in \tilde{\mathcal{A}}_{\text{left}}, z \in \tilde{\mathcal{A}}_{\text{right}}$, consider those strings $x \in A_{y,z}^{\text{stable}}$ with digit profile

$$\left(\{c_l^{(i)}(x)\}_{(i,l) \in T}, \{(t, x[t]) : (t, x[t]) \in \mathbf{Q}\} \right).$$

We call this group A for all $x \in A_{y,z}^{\text{stable}}$ with this digit profile. By Lemma 2.8, the corresponding number of strings is at most

$$\prod_{0 \leq i \leq z} \left(\frac{c_{\text{tot}}^{(i)} \log N}{\{c_l^{(i)} \log N\}_{(i,l) \in T}} \right) \leq N^{\sum_{0 \leq i \leq z} c_{\text{tot}}^{(i)} H(\{c_l^{(i)}\}_{l: (i,l) \in T}) + O(\delta)}.$$

For each group A , since the digit profile is fixed, we see that c_F and c_L are already fixed. Therefore, each group belongs to one of the following cases.

Case 1: $c_L \geq 2\delta$ for all $x \in A$. By (4.21) and (4.25), we have

$$\begin{aligned} & \sum_{x \in A} \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 \\ & \leq CN^{\sum_{0 \leq i \leq z} c_{\text{tot}}^{(i)} H(\{c_i^{(i)}\}_{(i,l) \in T}) + o(1)} \cdot N^{4c_L(x) - c_F(x) + 2c_E(x) + O(\delta)} + e^{-\Omega(N^{\delta-o(1)})} \\ & \leq CN^{-\Omega(\varepsilon)} + e^{-\Omega(N^{\delta-o(1)})}. \end{aligned}$$

Case 2. $c_L < 2\delta$ and $c_F \geq \delta - (\log N)^{-1/2}$ for all $x \in A$, and $A_{y,z}^{\text{stable}} \neq \{y \vee z\}$. By (4.22) and (4.26),

$$\begin{aligned} & \sum_{x \in A} \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 \\ & \leq CN^{\sum_{0 \leq i \leq z} c_{\text{tot}}^{(i)} H(\{c_i^{(i)}\}_{(i,l) \in T}) + o(1)} \cdot N^{-c_F(x) + 2c_E(x) + O(\chi) + 13\delta} + e^{-\Omega(N^{\delta-o(1)})} \\ & \leq CN^{-\Omega(\varepsilon)} + e^{-\Omega(N^{\delta-o(1)})}. \end{aligned}$$

Case 3. $c_L < 2\delta$ and either $c_F \geq \delta - (\log N)^{-1/2}$ and also $A_{y,z}^{\text{stable}} = \{y \vee z\}$, or else $c_F < \delta - (\log N)^{-1/2}$. By Lemma 4.15(iii), $c_L(y \vee z) < 2\delta$ and $A_{y,z}^{\text{stable}} = \{y \vee z\}$ (in either subcase). Therefore, by (4.23) and (4.27),

$$\sum_{x \in A} \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 \leq N^{2c_L(y \vee z) + 2c_E(y \vee z) + 4\delta} + e^{-\Omega(N^{\delta-o(1)})} \leq CN^{-\Omega(\varepsilon)} + e^{-\Omega(N^{\delta-o(1)})}.$$

Notice that there are at most

$$\log^{kz}(N) \cdot \sum_{\mathbf{q} \leq \frac{C \log N}{\log(1/\chi)}} \binom{K}{\mathbf{q}} k^{\mathbf{q}} = N^{O(\delta)}$$

total digit profiles with $\mathbf{q}(x) \leq \frac{C \log N}{\log(1/\chi)}$. Therefore, by (4.38) and combining the above three cases, we have

$$\begin{aligned} \sum_{x \in A_{y,z,i}^{\text{stable}}} \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 & \leq \sum_{\substack{x \in A_{y,z,i}^{\text{stable}} \\ \mathbf{q}(x) \leq \frac{C \log N}{\log(1/\chi)}}} \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 + N^{-C} \\ & \leq N^{O(\delta)} \cdot N^{-\Omega(\varepsilon)} \leq N^{-\Omega(\varepsilon)}. \end{aligned}$$

We conclude by (4.18) and (4.19) that

$$\begin{aligned} \mathbb{E}[|E(G, G')|] & \leq N^{-C/2} + N^{O(\delta)} \cdot \sum_{y \in \tilde{\mathcal{A}}_{\text{left}}, z \in \tilde{\mathcal{A}}_{\text{right}}} \sum_{x \in A_{y,z,i}^{\text{stable}}} \sum_{i=1}^{N-1} \mathbb{P}[(i, i+1) \in E(G_{B_x})]^2 \\ & \leq N^{-C/2} + N^{O(\delta)} \cdot \sum_{y \in \tilde{\mathcal{A}}_{\text{left}}, z \in \tilde{\mathcal{A}}_{\text{right}}} N^{-\Omega(\varepsilon)} \\ & \leq N^{-C/2} + |\tilde{\mathcal{A}}_{\text{left}}| \cdot |\tilde{\mathcal{A}}_{\text{right}}| \cdot N^{O(\delta)} \cdot N^{-\Omega(\varepsilon)} \leq O(N^{-\delta}). \end{aligned} \quad \square$$

5 The exponential moment

The main goal of this subsection is to prove Lemma 3.11, namely that $|E(G, G')|$ is typically $o(1)$ small even in the sense of exponential moments. Along the way we will also deduce Lemma 3.12. Recalling Subsection 3.1, we will assume throughout that for some $1 < t_1 < t_2 < A(\mu)$, the first shuffle, the t_1 -th shuffle and the t_2 -th shuffle are χ -good. This requirement appears in the discussion below (5.22), and in arguments of Lemmas 5.12, 5.13 and 5.8.

We now outline the organization of this section. Section 5.1 collects several preparatory results. In Section 5.2, we establish Lemmas 3.11 and 3.12 using an “exploration process” argument. The proof of a key technical lemma required for Lemma 3.11 is deferred to Sections 5.3 and 5.4.

5.1 Preparatory lemmas

The following simple lemma states that if an event $\tilde{A} \in \sigma(X, Y)$ occurs with high probability, we can construct another event $A \in \sigma(X)$ such that A itself occurs with high probability, and that if A holds, then $\tilde{A}$ occurs with high probability even after conditioning on X . We will use this lemma to exclude rare events in the exploration process; see Lemma 5.4.

Lemma 5.1. *Fix any probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and sub- σ -algebra $\mathcal{G} \subset \mathcal{F}$. For any $\delta > 0$ and event $A \in \mathcal{F}$ such that $\mathbb{P}(A) \geq 1 - p_0$, we have*

$$\mathbb{P}(\mathbb{P}(A|\mathcal{G}) \geq 1 - \sqrt{p_0}) \geq 1 - \sqrt{p_0} \quad (5.1)$$

Proof. Note that by the tower property of conditional expectations,

$$\mathbb{E}(\mathbb{P}(A^c|\mathcal{G})) \leq p_0.$$

Now (5.1) follows from the Markov inequality. $\square$

The next two easy lemmas from [Sel22] allow us to bound complicated expected edge intersections by simpler ones.

Lemma 5.2 ([Sel22, Lemma 5.2]). *Let A and B be independent random subsets of a finite set $\mathcal{A}$. Let A' and B' respectively be independent copies of A and B . Then*

$$\mathbb{E}[|A \cap B|] \leq \frac{\mathbb{E}[|A \cap A'|] + \mathbb{E}[|B \cap B'|]}{2}.$$

Lemma 5.3 ([Sel22, Lemma 5.3]). *Let A be a random subset of a finite set $\mathcal{A}$ and let $\mathcal{F}$ be a σ -algebra. Let A' be an independent copy of A and let $A_{\mathcal{F}}$ and $A'_{\mathcal{F}}$ be conditionally independent copies of A given $\mathcal{F}$. Then*

$$\mathbb{E}[|A \cap A'|] \leq \mathbb{E}[|A_{\mathcal{F}} \cap A'_{\mathcal{F}}|].$$

5.2 Exploration process

In this subsection, we will prove the crucial Lemma 3.11 via an exploration process (except for the technical Lemma 5.5 which will be deferred to the next subsection). We also deduce Lemma 3.12 at the end of the subsection. Recall the definition of τ_x and ι_x in (2.13) and $\mathcal{I}(B_x)$ in (2.14). Then, the interval $I(B_{[p]}) = \{\iota([p]), \dots, \tau([p])\}$ contains all strings with first digit p , and $\iota([p]), \tau([p])$ are determined by the cut sizes $n_p^{(t)}$ which we have fixed. For simplicity, we write $\iota(p)$ and $\tau(p)$ for $\iota([p])$ and $\tau([p])$, respectively. To estimate the exponential moment of $|E_{\text{forward}}^p(G, G')|$ on a typical event, we consider the exploration process on the interval $[\tau(p), \iota(p)]$ of strings beginning with p , and prove that $|E_{\text{forward}}^p(G, G')|$ can be dominated by a geometric random variable with mean $O(N^{-\delta})$ on this typical event. In the remainder of this subsection, we fix a digit p and always assume that $i \in \{\iota(p), \dots, \tau(p)\}$ and $K = O(\log N)$. For brevity, we write

$$\mathbf{a}_i = (s_{\iota(p)}, s_{\iota(p)+1}, \dots, s_i) \quad (5.2)$$

for the strings in $\mathcal{I}(B_{[p]})$ up to position i . Let G_i denote the induced subgraph of G on vertex set $\{\iota(p), \iota(p)+1, \dots, i\}$. We also let G' be an independent copy of shuffle graph G and define $s'_i, \mathbf{a}'_i$ and G'_i for G' similarly. We let $E_{\text{forward}}^p(G_i)$ consist of all edges $(i, i+1) \in E(G_i)$ for which $s_i = s_{i+1} \in \mathcal{S}_{\text{forward}}^p$, and let $E_{\text{forward}}^p(G_i, G'_i) = E_{\text{forward}}^p(G_i) \cap E_{\text{forward}}^p(G'_i)$; see also Definition 3.7 for reference.

We now explain how to upper-bound the exponential moment of $|E_{\text{forward}}^p(G, G')|$ on a typical event, using a variant of Portenko's lemma [Por75]. It requires truncating several atypical events, which ensure that the digits in our strings are evenly dispersed. Namely, we will consider a series of high-probability events $F_i \in \sigma(\mathbf{a}_i)$, under which we can upper-bound the expected number of unrevealed edges at each moment i . (The precise choice of F_i will be made only in (5.45) later.) With slight abuse of notation, we also define F'_i in the same way as F_i but with $\mathbf{a}_i$ replaced by $\mathbf{a}'_i$ in the following lemma. At the endpoints, we define

$$F_{\tau(p)-1} = \{s_{\tau(p)-1} \geq 1_{\text{ex}} [p \mathbf{I}_{\text{right}}^{(2)} \cdots \mathbf{I}_{\text{right}}^{(a_3)}]\}, \quad F'_{\tau(p)-1} = \{s'_{\tau(p)-1} \geq 1_{\text{ex}} [p \mathbf{I}_{\text{right}}^{(2)} \cdots \mathbf{I}_{\text{right}}^{(a_3)}]\},$$

where $\mathbf{I}_{\text{right}}^{(t)}$ and a_k are defined in Definition 3.6 and below. Also recall that $a_3 \leq A(\mu)$. We let

$$F^{(p)} = \bigcap_{i=\iota(p)}^{\tau(p)-1} (F_i \cap F'_i), \quad \forall p \in [k]_0.$$

Lemma 5.4. Consider events F_i, F'_i (for each $i \in \{\iota(p), \dots, \tau(p) - 2\}$), $F_{\tau(p)-1}, F'_{\tau(p)-1}$ and $F^{(p)}$ of the form above. Suppose that there exists $\gamma > 0$ such that for each $i \in [\iota(p), \tau(p) - 2]$, the conditional estimate

$$\mathbb{E}[|E_{\text{forward}}^p(G, G')| - |E_{\text{forward}}^p(G_i, G'_i)| \mid \mathbf{a}_i, \mathbf{a}'_i] \leq \gamma \quad (5.3)$$

holds on the event $F_i \cap F'_i$. Then

$$\mathbb{E}[e^{tE_{\text{forward}}^p(G, G')} \cdot 1_{F^{(p)}}] \leq 1 + 2e^t\gamma$$

for any N sufficiently large, $t > 0$ and $e^t\gamma < \frac{1}{10}$.

Proof. Define

$$\mathcal{F}_i = \sigma(\mathbf{a}_i, \mathbf{a}'_i, 1_{F^{(p)}}),$$

so that $(\mathcal{F}_i)_{i \in \mathcal{I}(B_x)}$ forms a filtration.

We now consider the random variable $X \equiv |E_{\text{forward}}^p(G, G')| \cdot 1_{F^{(p)}}$ and show that it can be dominated by an geometric variable Y with success probability γ . For $j > 0$, let

$$t_j = \inf\{i \in \{\iota(p), \iota(p) + 1, \dots, \tau(p)\} : |E_{\text{forward}}^p(G_i, G'_i)| \cdot 1_{F^{(p)}} \geq j\}$$

be a stopping time with respect to the filtration $(\mathcal{F}_i)_{i \in \{\iota(p), \iota(p)+1, \dots, \tau(p)\}}$, and then $s_{t_j} \in \mathcal{S}_{\text{forward}}^p$ on the event $F^{(p)}$. Therefore, on the event $\{t_j < \infty\}$, we have $F^{(p)}$ occurs and $t_j \leq \tau(p) - 2$. Hence,

$$\mathbb{P}(X > j \mid \mathcal{F}_{t_j}) \leq \mathbb{E}[|E_{\text{forward}}^p(G, G')| 1_{F^{(p)}} - |E_{\text{forward}}^p(G_i, G'_i)| 1_{F^{(p)}} \mid \mathcal{F}_{t_j}].$$

Moreover, (5.3) implies that for all $i \leq \tau(p) - 2$,

$$\mathbb{E}(|E_{\text{forward}}^p(G, G')| - |E_{\text{forward}}^p(G_i, G'_i)| \mid \mathbf{a}_i, \mathbf{a}'_i) \cdot 1_{F^{(p)}} \leq \mathbb{E}(|E_{\text{forward}}^p(G, G')| - |E_{\text{forward}}^p(G_i, G'_i)| \mid \mathbf{a}_i, \mathbf{a}'_i) \cdot 1_{F_i \cap F'_i} \leq \gamma,$$

thus we have (recalling the notation (5.2))

$$\begin{aligned} & \mathbb{E}[(|E_{\text{forward}}^p(G, G')| - |E_{\text{forward}}^p(G_i, G'_i)|) \cdot 1_{F^{(p)}} \mid \mathcal{F}_i] \\ &= \mathbb{E}[(|E_{\text{forward}}^p(G, G')| - |E_{\text{forward}}^p(G_i, G'_i)|) \mid \mathcal{F}_i] \cdot 1_{F^{(p)}} \\ &= \mathbb{E}[\mathbb{E}[|E_{\text{forward}}^p(G, G')| - |E_{\text{forward}}^p(G_i, G'_i)| \mid \mathbf{a}_i, \mathbf{a}'_i] \mid \mathcal{F}_i] \cdot 1_{F^{(p)}} \\ &= \mathbb{E}[\mathbb{E}[|E_{\text{forward}}^p(G, G')| - |E_{\text{forward}}^p(G_i, G'_i)| \mid \mathbf{a}_i, \mathbf{a}'_i] \cdot 1_{F^{(p)}} \mid \mathcal{F}_i] \\ &\leq \gamma \end{aligned}$$

Since we have $t_j \leq \tau(p) - 2$ on the event $\{t_j < \infty\}$, this implies that

$$\mathbb{P}(X > j \mid \mathcal{F}_{t_j}) \leq \gamma.$$

Note that $\{X \geq j\} = \{t_j < \infty\} \in \mathcal{F}_{t_j}$ is $\mathcal{F}_{t_j}$ -measurable, so we have

$$\mathbb{P}(X > j \mid X \geq j) \leq \gamma,$$

which implies that the random variable X is dominated by a geometric random variable Y with

$$\mathbb{P}(Y = j) = \gamma^j(1 - \gamma), \quad j \geq 0.$$

Recall that $X = |E_{\text{forward}}^p(G, G')| \cdot 1_{F^{(p)}}$. Using the assumption $e^t\gamma < \frac{1}{10}$, we find

$$\mathbb{E}[e^{t|E_{\text{forward}}^p(G, G')|} \cdot 1_{F^{(p)}}] \leq \mathbb{E}[e^{tX}] \leq \mathbb{E}[e^{tY}] \frac{1 - \gamma}{1 - e^t\gamma} \leq 1 + 2e^t\gamma,$$

which then completes the proof. $\square$

In the rest of this subsection, we focus on proving that

$$\mathbb{E}[|E_{\text{forward}}^p(G, G')| - |E_{\text{forward}}^p(G_i, G'_i)| \mid \mathbf{a}_i, \mathbf{a}'_i] \leq O(N^{-\zeta})$$

holds for some $\zeta > 0$ (as in (2.9)), after a suitable truncation. The idea will be to decompose the set of possible unrevealed strings into $N^{O(\rho)}$ disjoint blocks, and apply the first-moment estimate from the previous section to each one. It is easy to show (see e.g. [Sel22, Lemma 5.5]) that the lexicographic interval

$$I_{s_i, [p(k-1)(k-1)]} \equiv \{s \in [k]_0^K : s_i <_{\text{lex}} s <_{\text{lex}} [pI_{\text{right}}^{(2)} \cdots I_{\text{right}}^{(a_3)}]\}$$

can be written as a disjoint union of blocks

$$I_{s_i, [p(k-1)(k-1)]} = \bigcup_{\tilde{x} \in \widetilde{\text{Block}}_{\text{forward}}^p(s_i)} B_{\tilde{x}} \quad (5.4)$$

for some set $\widetilde{\text{Block}}_{\text{forward}}^p(s_i)$ containing at most $2kK = O(\log N)$ strings, such that each $x \in \widetilde{\text{Block}}_{\text{forward}}^p(s_i)$ has length at most K .

For prefixes $\tilde{x} \in \widetilde{\text{Block}}_{\text{forward}}^p$ with length $M = M(\tilde{x}) \leq K - 2\rho \log(N)$, there exists $j_0 \in \mathbb{Z}^+$ such that

$$M < j_0 \rho \log N \leq M + \rho \log N \leq (j_0 + 1) \rho \log N < K.$$

Since the shuffle process is (χ, ρ, φ) -almost- μ -like, there exists a χ -good shuffle $\hat{t} = \hat{t}(\tilde{x}) \in (j_0 \rho \log N, (j_0 + 1) \rho \log N]$, which implies $0 \leq \hat{t} - M < \rho \log N$ and $\hat{t} \leq K$. We may therefore further decompose $B_{\tilde{x}}$ into the following disjoint union of blocks,

$$B_{\tilde{x}} = \bigcup_{w \in [k]_0^{\hat{t}-M-1}} B_{[xw]}.$$

The number of blocks in the union above is at most $k^{\hat{t}-M-1} \leq k^{\rho \log N} = N^{O(\rho)}$.

Thus, the set of prefixes

$$\begin{aligned} \text{Block}_{\text{forward}}^p(s_i) &\equiv \{\tilde{x} \in \widetilde{\text{Block}}_{\text{forward}}^p(s_i) : M(\tilde{x}) > K - 2\rho \log(N)\} \cup \\ &\quad \{[xw] : \tilde{x} \in \widetilde{\text{Block}}_{\text{forward}}^p(s_i), M(\tilde{x}) \leq K - 2\rho \log(N), w \in [k]_0^{\hat{t}-M}\}. \end{aligned}$$

contains at most $O(\log N) \cdot N^{O(\rho)} = N^{O(\rho)}$ prefixes. Since the prefix $\tilde{x}$ corresponds to disjoint blocks $B_{\tilde{x}}$, the equation (5.4) implies that

$$I_{s_i, [p(k-1)(k-1)]} = \bigcup_{x \in \text{Block}_{\text{forward}}^p(s_i)} B_x \quad (5.5)$$

is a disjoint union of blocks. In addition, every prefix in $\text{Block}_{\text{forward}}^p(s_i)$ has length at most K , and any prefix of length $M \leq K - 2\rho \log N$ satisfies that $(M + 1)$ is a χ -good shuffle.

Recall the definition of S_{forward}^p in Definition 3.7. We estimate the number of unrevealed edges in each block B_x for $x \in \text{Blocks}(s_i)$, conditionally on $\mathbf{a}_i$ after truncation, according to the expected size $\lambda_x = \frac{1}{N} \mathbb{E}|\mathcal{I}(B_x)|$, where $\mathcal{I}(B_x) = \{j \in [N] : s_j \in B_x\}$.

When $\lambda_x \geq N^{-1+\xi}$, where $\xi > 0$ is a small constant as in (2.9), we will apply the upper bound from Proposition 4.2, established in the previous subsection, to control the number of unrevealed edges in B_x . (See (5.15) and below.) To use Proposition 4.2, it is necessary to ensure that, for every prefix x with $\lambda_x \geq N^{-1+\xi}$, the continuation of the shuffle process after x is conditionally “ μ -like” within the interval $\mathcal{I}(B_x)$. To guarantee this, we exclude certain rare events (included in the F_i below). Recall that M denotes the length of the prefix x . For brevity, denote by

$$n_l^{(t),x} = \sum_{j \in \mathcal{I}(B_x)} 1_{\{s_j[t]=l\}}, \quad (5.6)$$

the number of digits equal to l in the t -th shuffle among strings in B_x , for $t > M$ and $l \in [k]_0$. In Lemma 5.5 below, we show that both $|\mathcal{I}(B_x)|$ and $n_l^{(t),x}$ concentrate sharply for typical $\mathbf{a}_i$, which is sufficient to guarantee that the shuffle after prefix x remains conditionally “ μ -like.” Note that suitable concentration must hold even for quite small λ_x .

On the other hand, if $\lambda_x < N^{-1+\xi}$, we instead bound either the first moment of $|\mathcal{I}(B_x)|$ or the number of edges in B_x to obtain the desired estimate. In fact, for typical strings $\mathbf{a}_i$, we can establish upper bounds on both the first moment of $|\mathcal{I}(B_x)|$ and the number of edges in the block B_x , conditionally on $\mathbf{a}_i$. More precisely, there exists an event $F_i \in \sigma(\mathbf{a}_i)$ on which these quantities can be effectively bounded. The results are summarized in the following lemma. Recall that $E(G_{B_x})$ denotes the edges in G with vertices in B_x . Also recall the definition of $\mathbf{p}^{(i)}$ in (2.10).

Lemma 5.5. *For any $K \geq (\tilde{C}_\mu + \varepsilon) \log N$ and $i \in \{\iota(p), \dots, \tau(p) - 2\}$, and any $\xi > 0$, there exists an event $F_i = F_i(\xi) \in \sigma(\mathbf{a}_i)$ and $\zeta = \zeta(\mu, \xi)$, such that the following holds.*

(a) *We have*

$$\mathbb{P}(F_i) \geq 1 - \exp(-N^\zeta).$$

(b) *There exists $C(\mu, \chi) > 0$ such that if $s_i \in \mathcal{S}_{\text{forward}}^p$, we have that for all $x \in \mathcal{S}_{\text{forward}}^p$ and $x >_{\text{lex}} s_i$,*

$$\mathbb{E}[|\mathcal{I}(B_x)| | \mathbf{a}_i] \leq C(\mu, \chi) N(\lambda_x)^{1-o_\chi(1)} \cdot N^{O(\chi)}. \quad (5.7)$$

on the event F_i .

(c) *There exists $C(\mu, \chi) > 0$ such that if $s_i \in \mathcal{S}_{\text{forward}}^p$, we have*

$$\mathbb{E}[|\{j > i : s_j = s_i\}| | \mathbf{a}_i] \leq C(\mu, \chi) N(\lambda_{s_i})^{1-o_\chi(1)} \cdot N^{O(\chi)}. \quad (5.8)$$

on the event F_i .

(d) *There exists $C(\mu, \chi) > 0$ such that if $s_i \in \mathcal{S}_{\text{forward}}^p$, we have that for all $x \in \mathcal{S}_{\text{forward}}^p$ and $x >_{\text{lex}} s_i$,*

$$\mathbb{E}[|E(G_{B_x})| | \mathbf{a}_i] \leq C(\mu, \chi) N^{2 - \frac{K-M}{\log N} \sum_{1 \leq i \leq z} h_i \psi_{\mathbf{p}(i)}(2) + o_\chi(1)} \cdot (\lambda_x)^{2-o_\chi(1)} \quad (5.9)$$

on the event F_i , where M is the length of prefix x .

(e) *For any prefix x such that $\lambda_x \geq N^{-1+\xi}$, we have*

$$\mathbb{P}\left(|\mathcal{I}(B_x)| - N\lambda_x \leq (N\lambda_x)^{\frac{4}{5}} | \mathbf{a}_i\right) \geq 1 - \exp(-N^{\xi/2})$$

on the event F_i .

(f) *For any prefix x such that $\lambda_x \geq N^{-1+\xi}$, we have*

$$\mathbb{P}\left(\left|n_l^{(t),x} - N\lambda_x \cdot \frac{n_l^{(t)}}{N}\right| \leq (N\lambda_x)^{\frac{4}{5}} \text{ for all } (t, l) \in \mathbf{P} \text{ and } M < t \leq K | \mathbf{a}_i\right) \geq 1 - \exp(-N^{\xi/2})$$

on the event F_i , where M is the length of prefix x .

We postpone the proof of Lemma 5.5 to Section 5.3, after completing the proofs of Lemmas 3.11 and 3.12, as it is fairly involved and largely independent of the other arguments in this subsection.

Recall that G' is an independent copy of the shuffle graph G and s'_i , $\mathbf{a}'_i$ and F'_i are defined similarly for G' . We now prove an upper bound of the expected number of the “unrevealed” edges in $E_{\text{forward}}^p(G, G')$, conditionally on revealing the first i strings in G and G' , on the event that the first i strings are typical. Recall that G_i is the induced subgraph of G with vertex set $\iota(p), \iota(p) + 1, \dots, \tau(p)$ and for brevity define the “unrevealed” forward subgraphs

$$G_{u,1} = E_{\text{forward}}^p(G) \setminus E_{\text{forward}}^p(G_i), \quad G'_{u,1} = E_{\text{forward}}^p(G') \setminus E_{\text{forward}}^p(G'_i).$$

Note that we have suppressed i ; at this point, we have fixed i and aim to check the condition of Lemma 5.4 for that value of i .

Lemma 5.6. *For all $K \geq (\tilde{C}_\mu + \varepsilon) \log N$, $\chi, \rho, \varphi > 0$ as in (2.9) sufficiently small, (χ, ρ, φ) -almost- μ -like shuffle $(\vec{n}^{(t)})_{t \leq K}$ and $i \in \{\iota(p), \dots, \tau(p) - 2\}$, there exists $\zeta = \zeta(\mu, \varepsilon) > 0$ such that on the event $F_i \cap (F'_i)'$ constructed in Lemma 5.5, we have*

$$\mathbb{E}[|E(G_{u,1}, G'_{u,1})| | \mathbf{a}_i, \mathbf{a}'_i] \leq O(N^{-\zeta}).$$

Proof. Without loss of generality, suppose that s_i, s'_i is smaller than $[pI_{\text{right}}^{(2)} \cdots I_{\text{right}}^{(a_3)}]$. Define the graph $G_{u,2}$ (resp. $G'_{u,2}$) as a conditionally independent copy of $G_{u,1}$ (resp. $G'_{u,1}$) given

$$\tilde{\mathcal{F}}_i \stackrel{\text{def.}}{=} \sigma(\mathbf{a}_i, \mathbf{a}'_i, (\mathcal{I}(B_x))_{x \in \text{Block}_{\text{forward}}^p(s_i)}),$$

and $\tilde{G}_{u,1}$ (resp. $\tilde{G}'_{u,1}$) as a conditionally independent copy of $G_{u,1}$ (resp. $G'_{u,1}$) given $\mathbf{a}_i, \mathbf{a}'_i$. We now show that at any time i , the expected number of unrevealed common edges

$$\mathbb{E}[|E(G_{u,1}, G'_{u,1})| | \mathbf{a}_i, \mathbf{a}'_i] \leq \mathbb{E}[|E(G_{u,1}, G_{u,2})| | \mathbf{a}_i]. \quad (5.10)$$

Actually, we have

$$\begin{aligned} \mathbb{E}[|E(G_{u,1}, G'_{u,1})| | \mathbf{a}_i, \mathbf{a}'_i] &\stackrel{\text{Lemma 5.2}}{\leq} \frac{\mathbb{E}[|E(G_{u,1}, \tilde{G}_{u,1})| | \mathbf{a}_i, \mathbf{a}'_i] + \mathbb{E}[|E(G'_{u,1}, \tilde{G}'_{u,1})| | \mathbf{a}_i, \mathbf{a}'_i]}{2} \\ &\stackrel{\text{Lemma 5.3}}{\leq} \frac{\mathbb{E}[|E(G_{u,1}, G_{u,2})| | \tilde{\mathcal{F}}_i] + \mathbb{E}[|E(G'_{u,1}, G'_{u,2})| | \tilde{\mathcal{F}}_i]}{2} \end{aligned}$$

By symmetry, we obtain (5.10), so to complete the proof, it suffices to verify that

$$\mathbb{E}[|E(G_{u,1}, G_{u,2})| | \mathbf{a}_i] \leq O(N^{-\zeta}).$$

Recall the definition of $\text{Block}_{\text{forward}}^p(s_i)$ in (5.5). For any prefix $x \in \text{Block}_{\text{forward}}^p(s_i)$, let $E(G_{u,1}, G_{u,2}, x)$ be all shared edges in $E(G_{u,1}, G_{u,2})$ whose vertices in $G_{u,1}$ and in $G_{u,2}$ are all in B_x . Since $G_{u,1}$ and $G_{u,2}$ have the same $(|\mathcal{I}(B_x)|)_{x \in \text{Block}_{\text{forward}}^p(s_i)}$ and $\text{Block}_{\text{forward}}^p(s_i) \in \sigma(\mathbf{a}_i)$, we have the decomposition

$$\mathbb{E}[|E(G_{u,1}, G_{u,2})| | \mathbf{a}_i] = \mathbb{E}[|j \in \{i+1, \dots, \tau(p)\} : s_j = s_i| | \mathbf{a}_i] + \sum_{x \in \text{Block}_{\text{forward}}^p(s_i)} \mathbb{E}[|E(G_{u,1}, G_{u,2}, x)| | \mathbf{a}_i]. \quad (5.11)$$

We start with estimating the first term on the right-hand side of (5.11). Recall the definition of t_* in (2.2). Let j_0 be the biggest integer such that $t_* + j_0 \rho \log N \leq K$. We now have

$$\begin{aligned} \log_N \lambda_{s_i} &= \frac{1}{\log N} \sum_{1 \leq t \leq K} \log \frac{n_{s_i[t]}^{(t)}}{N} \\ &\leq \frac{1}{\log N} \sum_{t_* \leq t \leq t_* + j_0 \rho \log N} \sum_{1 \leq i \leq z} \log \frac{n_{s_i[t]}^{(t)}}{N} \cdot 1_{\{t \in P_i\}} \\ &\leq \frac{1}{\log N} \sum_{1 \leq i \leq z} \sum_{t_* \leq t \leq t_* + j_0 \rho \log N, t \in P_i} (\log p_{\max}^{(i)} + O(\chi)) \\ &\leq \frac{K}{\log N} \sum_{1 \leq i \leq z} h_i \log p_{\max}^{(i)} + O(\rho) + O(\chi) + O(\varphi) \\ &= -\frac{K}{\log N} \mathbb{E}_{\mu}(\log(1/p_{\max})) + o_{\chi}(1), \end{aligned} \quad (5.12)$$

where in the fourth line we have used that $K/(\log N) \leq 2\bar{C}_{\mu}$ is bounded without loss of generality, and (2.2), and in the last line we have used (2.9). Since

$$K \geq \left(\frac{1}{\mathbb{E}_{\mu}(\log(1/p_{\max}))} + \varepsilon \right) \log N,$$

we have $\lambda_{s_i} \leq N^{-1-\Omega(\varepsilon)}$ for all χ, ρ, φ sufficiently small. Therefore, by (5.8), we have

$$\mathbb{E}[|j \in \{i+1, \dots, \tau(p)\} : s_j = s_i| | \mathbf{a}_i] \leq C(\mu, \chi) N^{1-(1+\Omega(\varepsilon))(1-o_{\chi}(1))-O(\chi)}. \quad (5.13)$$

Since $o_{\chi}(1), \chi \ll \Omega(\varepsilon)$, by (5.13), there exists $\zeta > 0$ as in (2.9) such that

$$\mathbb{E}[|j \in \{i+1, \dots, \tau(p)\} : s_j = s_i| | \mathbf{a}_i] \leq O(N^{-\zeta}).$$

To estimate the second term on the right-hand side of (5.11), we will show that for all $x \in \text{Block}_{\text{forward}}^p(s_i)$,

$$\mathbb{E}[|E(G_{u,1}, G_{u,2}, x)| | \mathbf{a}_i] \leq O(N^{-\zeta}) \quad (5.14)$$

for some $\zeta > 0$. Since $\text{Block}_{\text{forward}}^p(s_i) = O(\log N)$, this suffices to conclude the proof. We now split into three cases according to the size of λ_x and write M for the number of digits in x in each case. The main one is Case 3 where we apply the first moment bound to the smaller system consisting of strings within B_x . As in (2.9), we let $\xi > 0$ be such that $\xi \ll \varepsilon$, and suppose that $\chi, \rho, \varphi \ll \xi$ and $\zeta \ll \chi, \rho, \varphi$.

Case 1. $\lambda_x \leq N^{-1-\xi}$. In this case, we have

$$\begin{aligned}
\mathbb{E}[|E(G_{u,1}, G_{u,2}, x)| | \mathbf{a}_i] &\leq \mathbb{E}[|\mathcal{I}(B_x)| | \mathbf{a}_i] \\
&\stackrel{(5.7)}{\leq} C(\mu, \chi, \zeta) N \lambda_x^{1-o_\chi(1)} N^{O(\chi)} \\
&\leq O(N^{1-(1+\xi)(1-o_\chi(1))-O(\chi)}) \\
&\leq O(N^{-\Theta(\xi)}) \\
&\leq O(N^{-\zeta})
\end{aligned}$$

Case 2. $N^{-1-\xi} \leq \lambda_x \leq N^{-1+\xi}$. Imitating the calculation in (5.12) gives

$$\log_N \lambda_x \leq -\frac{M}{\log N} \mathbb{E}_\mu(\log(1/p_{\max})) + o_\chi(1).$$

The condition $\lambda_x \geq N^{-1-\xi}$ now implies by rearranging that

$$\frac{M}{\log N} \leq \frac{1+\xi}{\mathbb{E}_\mu(\log(1/p_{\max}))} + o_\chi(1),$$

where we have used that $\mathbb{E}_\mu(\log(1/p_{\max})) = 1/\tilde{C}_\mu > 0$ depends only on μ in dividing (and that $\mu(V_k) < 1$). Since $o_\chi(1) \ll \varepsilon$ is sufficiently small (recall (2.9)), this implies $K - M \geq \Omega(\varepsilon) \log N$.

Trivially, the shared edges in $G_{u,1}$ and $G_{u,2}$ are in particular edges in $G_{u,1}$. Recall that $E(G_{B_x})$ is the set of edges in G with vertices in B_x . Therefore, we have

$$\begin{aligned}
&\mathbb{E}[|E(G_{u,1}, G_{u,2}, x)| | \mathbf{a}_i] \\
&\leq \mathbb{E}[|E((G_{u,1})_{B_x})| | \mathbf{a}_i] \\
&\stackrel{(5.9)}{\leq} C(\mu, \chi) N^{2-\frac{K-M}{\log N} \sum_{1 \leq i \leq z} h_i \psi_{\mathbf{p}(i)}(2) + o_\chi(1)} \cdot (\lambda_x)^{2-o_\chi(1)} \\
&= N^{2-\frac{K-M}{\log N} \mathbb{E}_\mu \psi_{\mathbf{p}}(2) + o_\chi(1)} (\lambda_x)^{2-o_\chi(1)}
\end{aligned}$$

Since $K - M \geq \Omega(\varepsilon) \log N$ and $\mathbb{E}_\mu(\psi_{\mathbf{p}}(2)) > 0$, we have

$$\mathbb{E}[|E(G_{u,1}, G_{u,2}, x)| | \mathbf{a}_i] \leq O(N^{2-2(1-\xi)(1-o_\chi(1))+o_\chi(1)-\Omega(\varepsilon)}) \leq O(N^{-\zeta}).$$

Here we used (2.9).

Case 3. $\lambda_x \geq N^{-1+\xi}$. Similarly to Case 2, we have $M \leq (1+o_\chi(1))\bar{C}_\mu \log(\lambda_x^{-1})$. If $|\mathcal{I}(B_x)| \in [\frac{1}{2}N\lambda_x, 2N\lambda_x]$, we have

$$\begin{aligned}
K - M &\geq (1+o_\chi(1))(\bar{C}_\mu + \varepsilon) \log(N\lambda_x) \\
&\geq \left(\bar{C}_\mu + \frac{\varepsilon}{2}\right) \log(2N\lambda_x) \\
&\geq \left(\bar{C}_\mu + \frac{\varepsilon}{2}\right) \log |\mathcal{I}(B_x)| \\
&\geq \left(\bar{C}_\mu + \frac{\varepsilon}{2}\right) \xi \log N.
\end{aligned} \tag{5.15}$$

Here we use $\chi, \rho, \varphi \ll \xi \ll \varepsilon$ in the third line and $|\mathcal{I}(B_x)| \geq \frac{1}{2}N\lambda_x \geq \frac{1}{2}N^\xi$.

Recall the definition of $n_l^{(t),x}$ in (5.6). We now apply a crucial self-reducibility argument. Conditioned on $\mathcal{I}(B_x)$ and $(n_l^{(T),x})_{T>M, l \in [k]_0}$, suppose we do a riffle shuffle with $|\mathcal{I}(B_x)|$ cards and pile sizes

$$(n_0^{(t+M),x}, n_1^{(t+M),x}, \dots, n_{k-1}^{(t+M),x})$$

for each $t \in [K - M]$. In this way we generate $|\mathcal{I}(B_x)|$ strings; if we append x to the beginning of each string, then the resulting $|\mathcal{I}(B_x)|$ strings of length K have the same distribution as the multi-set of strings starting with x . Thus we will apply the first moment result from Section 4 with $|\mathcal{I}(B_x)|$ cards. To do so, consider the event

$$\hat{F} = \left\{ \begin{aligned} &|\mathcal{I}(B_x)| \in [N\lambda_x/2, 2N\lambda_x]. \\ &\left| n_l^{(t+M),x} - |\mathcal{I}(B_x)| \frac{n_l^{(t+M)}}{N} \right| \leq |\mathcal{I}(B_x)|^{\frac{4}{5}} \text{ for all } (t+M, l) \in \mathbf{P}. \end{aligned} \right\}$$

Note that on the event $\widehat{F}$, we have $\log_N |\mathcal{I}(B_x)| = \xi + o(1)$ as $N \rightarrow \infty$. Therefore, by the second condition in the definition of $\widehat{F}$, for all $(t+M, l) \in \mathcal{P}$, the t -th shuffle in the new shuffle satisfies

$$\left| \frac{n_l^{(t+M),x}}{|\mathcal{I}(B_x)|} - \frac{n_l^{(t+M)}}{N} \right| \leq |\mathcal{I}(B_x)|^{-\frac{1}{5}}.$$

We now consider the shuffle process

$$(\widehat{n}_0^{(t),x}, \widehat{n}_1^{(t),x}, \dots, \widehat{n}_{k-1}^{(t),x}), \quad t \in [K-M],$$

such that

$$\left| \widehat{n}_l^{(t),x} - |\mathcal{I}(B_x)| \frac{n_l^{(t+M)}}{N} \right| \leq 2k, \quad \sum_{0 \leq l \leq k-1} \widehat{n}_l^{(t),x} = |\mathcal{I}(B_x)|,$$

$\frac{\widehat{n}_i^{(t),x}}{|\mathcal{I}(B_x)|}$ and $\frac{n_i^{(t+M)}}{N}$ are in the same D_i .

The existence of such a point $(\widehat{n}_l^{(t),x})_{0 \leq l \leq k-1}$ is guaranteed by the fact that D_i is a simplex. Hence, the new shuffle is weak- $\frac{1}{6}$ -($\chi, \frac{\rho \log N}{K-M}, \frac{\varphi \log N}{K-M}$)-almost μ -like. Note that (5.15) implies that $\frac{\log N}{K-M} \leq [(\overline{C}_\mu + \frac{\varepsilon}{2})\xi]^{-1}$. In addition, since $M \leq (1 + o_\chi(1))\overline{C}_\mu$ and $\rho, \chi \ll \varepsilon$, we can pick ρ and χ sufficiently small such that $M \leq K - 2\rho \log N$. Thus, by the discussion below (5.5), the first shuffle of the new shuffle process is χ -good. Therefore, we can now pick χ, ρ, φ sufficiently small and apply Proposition 4.2 (with $\frac{\rho \log N}{K-M}$ in place of ρ and $\frac{\varphi \log N}{K-M}$ in place of φ) to obtain

$$|E(G_{u,1}, G_{u,2}, x)| \stackrel{(4.1), (5.15)}{\leq} O(|\mathcal{I}(B_x)|^{-\delta}) \leq O(N^{-\zeta})$$

on the event $\widehat{F}$. In addition, by Lemma 5.5, on the event F_i , $\widehat{F}$ occurs with probability higher than $1 - \exp(-N^\xi)$. Therefore, we can use the trivial bound $|E(G_{u,1}, G_{u,2}, x)| \leq N$ on the event $(\widehat{F})^c$ to derive (5.14). This concludes the proof. $\square$

Combining Lemmas 5.6 and 5.4, we give the upper bound for the exponential moment, i.e. Lemma 3.11. Note that although we have focused on $E_{\text{forward}}^p(G, G')$, exactly analogous results for $E_{\text{backward}}^p(G, G')$ can be established by exploration in the opposite direction. Thus we only prove Lemma 3.11(a) and Lemma 3.11(b) in the cases that involve $\mathcal{S}_p^{\text{forward}}$, $E_{\text{forward}}^p(G, G')$, $\mathcal{S}_{\text{forward}}^p$, etc.

To establish Lemma 3.11(c), we need to do the following further truncation. Recall that ξ is a sufficiently small positive constant. For all x such that $\lambda_x \leq N^{-1+2\xi}/2$ and the length of x is smaller than K , we introduce the following event

$$J_x = \{|\mathcal{I}(B_x)| \leq 3N^{2\xi}\}, \quad J_\xi = \bigcap_{\substack{x \in \cup_{1 \leq M \leq K} [k]_0^M, \\ \lambda_x \leq N^{-1+2\xi}}} J_x.$$

By Lemma 2.7, we have that for all x with $\lambda_x \leq N^{-1+2\xi}$:

$$\mathbb{P}(J_x) \geq 1 - \exp(-cN^{2\xi}).$$

Since the intersection in the definition of J_ξ is over at most $N^{O(1)}$ events, each of which occurs with probability higher than $1 - \exp(-cN^{2\xi})$, the event $J_{i,\xi}$ occurs with probability higher than $1 - \exp(-N^\xi)$.

Similarly, we also introduce

$$H_\xi = \bigcap_{x: \lambda_x \geq N^{-1+\xi}} \{||\mathcal{I}(B_x)| - N\lambda_x| \leq (N\lambda_x)^{\frac{4}{5}}\}.$$

We have $\mathbb{P}(H_\xi) \geq 1 - \exp(-N^\xi)$.

Proof of Lemma 3.11. We take $\mathcal{S}_p^{\text{forward}}$ to consist of all sequences of strings S such that the event

$$\left(\bigcap_{i=\iota(p)}^{\tau(p)-2} F_i \right) \cap F_{\tau(p)-1} \cap J_\xi \cap H_\xi \quad (5.16)$$

occurs, where the event $F_{\tau(p)-1} = \{s_{\tau(p)-1} \geq [pI_{\text{right}}^{(2)} \cdots I_{\text{right}}^{(a_3)}]\}$ is defined in Lemma 5.4, and F_i ($i \in [\iota(p) + 1, \tau(p) - 2]$) is constructed in Lemma 5.5. It is elementary that

$$\begin{aligned} \{S_K \in \mathcal{S}_p^{\text{forward}}\} &= \left(\bigcap_{i=\iota(p)}^{\tau(p)-2} F_i \right) \cap F_{\tau(p)-1} \cap J_\xi \cap H_\xi, \\ \{S'_K \in \mathcal{S}_p^{\text{forward}}\} &= \left(\bigcap_{i=\iota(p)}^{\tau(p)-2} F'_i \right) \cap F'_{\tau(p)-1} \cap J'_\xi \cap H'_\xi. \end{aligned}$$

The inequality in Lemma 3.11(a) for $\mathcal{S}_p^{\text{forward}}$ follows from the union bound and the fact that $F_i, F_{\tau(p)-1}, J_\xi, H_\xi$ each occur with probability at least $1 - \exp(-N^\xi)$. The set $\mathcal{S}_p^{\text{backward}}$ is defined by replacing $(k-1)$ with 0 in the definition of $F_{\iota(p)+1}$ and F_i for $i \in [\iota(p) + 2, \tau(p)]$, and substituting the intersection of all these events for the event in (5.16). Lemma 3.11(a) in this case can be proved similarly.

To demonstrate Lemma 3.11(b), we apply Lemma 5.4 with:

$$\gamma = O(N^{-\zeta}),$$

and F_i as constructed in Lemma 5.5. Note that $e^t \gamma < \frac{1}{10}$ for sufficiently large N . By Lemma 5.6, (5.3) holds. Therefore, the first line of (3.9) holds for ζ in place of δ therein by the fact that

$$1_{\{S_K, S'_K \in \mathcal{S}_p^{\text{forward}}\}} \leq 1_{F^{(p)}},$$

where $F^{(p)}$ is defined above Lemma 5.4. The second line of (3.9) can be derived similarly.

We now turn to the proof of Lemma 3.11(c). We show the regularity first. Recall the definition of J_ξ and H_ξ above the proof of this lemma. We start with establishing (3.5). Actually, we have

$$\begin{aligned} &\left| i \in \{\iota(x), \iota(x) + 1, \dots, \tau(x)\} : s_i[1]s_i[2] \cdots s_i[a_3] \leq_{\text{lex}} [pI_{\text{left}}^{(2)} \cdots I_{\text{left}}^{(a_3)}] \right| \\ &= \sum_{2 \leq k \leq a_3} \sum_{0 \leq j < I_{\text{left}}^{(k)}} |\mathcal{I}(B_{[pI_{\text{left}}^{(2)} \cdots I_{\text{left}}^{(k-1)}]j})| + |\mathcal{I}(B_{[pI_{\text{left}}^{(2)} \cdots I_{\text{left}}^{(a_3)}]})|. \end{aligned} \quad (5.17)$$

For simplicity, we write

$$\mathcal{S}_{\text{left}}^p \equiv \{[pI_{\text{left}}^{(2)} \cdots I_{\text{left}}^{(k-1)}]j : 2 \leq k \leq a_3, 0 \leq j < I_{\text{left}}^{(k)}\} \cup \{[pI_{\text{left}}^{(2)} \cdots I_{\text{left}}^{(a_3)}]\},$$

so that the right-hand side of (5.17) can be written as $\sum_{x \in \mathcal{S}_{\text{left}}^p} |\mathcal{I}(B_x)|$. By the fact that $a_3 \leq A(\mu)$ is bounded, we have $|\mathcal{S}_{\text{left}}^p| = O(1)$.

For the strings in $\mathcal{S}_{\text{left}}^p$ such that $\lambda_x \geq N^{-1+\xi}$, the definition of the event H_ξ implies that

$$|\mathcal{I}(B_x)| \leq N\lambda_x + (N\lambda_x)^{\frac{4}{5}} \leq N\lambda_x(1 + N^{-\xi/6}) \quad (5.18)$$

on the event H_ξ . For the strings in $\mathcal{S}_{\text{left}}^p$ such that $\lambda_x < N^{-1+\xi}$, the definition of the event J_ξ implies

$$|\mathcal{I}(B_x)| \leq 3N^{2\xi}. \quad (5.19)$$

Combining (5.18), (5.19) with (5.17) gives

$$\left| i \in \{\iota(x), \iota(x) + 1, \dots, \tau(x)\} : s_i[1]s_i[2] \cdots s_i[a_3] \leq_{\text{lex}} [pI_{\text{left}}^{(2)} \cdots I_{\text{left}}^{(a_3)}] \right| \leq (1 + o(1))N \sum_{x \in \mathcal{S}_{\text{left}}^p} \lambda_x + |\mathcal{S}_{\text{left}}^p| \cdot 3N^{2\xi}$$

By (5.17) and (3.4), we have $\lambda_{\text{left}}^p = \sum_{x \in \mathcal{S}_{\text{left}}^p} \lambda_x$. Moreover, since $a_3 \leq A(\mu)$, we have $\lambda_{\text{left}}^p \geq \chi^{A(\mu)}$, a positive constant independent of N . Therefore, we further have

$$\left| i \in \{\iota(x), \iota(x) + 1, \dots, \tau(x)\} : s_i[1]s_i[2] \cdots s_i[a_3] \leq_{\text{lex}} [pI_{\text{left}}^{(2)} \cdots I_{\text{left}}^{(a_3)}] \right| \leq (1 + o(1))N\lambda_{\text{left}}^p,$$

which implies the first line of (3.5). The second line can be established similarly. Therefore, all sequences in $\mathcal{S}^\dagger$ are regular. $\square$

Proof of Lemma 3.12. Since $K \geq (\tilde{C}_\mu + \varepsilon) \log N$, we have $N \lambda_{s_i}^{1-o_\chi(1)} \leq O(N^{-\delta})$ for all $s_i \in \mathcal{S}_K$ and $\chi > 0$ sufficiently small (see (5.12) and (5.13)). By (5.8), on the event F_i (in Lemma 5.5 with sufficiently small $\delta > 0$), we have

$$\mathbb{P}[s_{i+1} = s_i | \mathbf{a}_i] \leq O(N^{-\delta}), \quad \forall s_i \in \mathcal{S}_{\text{forward}}^p. \quad (5.20)$$

Take $F = \cap_i F_i$ and $L > C_0/\delta$. For each discrete sub-interval

$$\{i, i+1, \dots, i+L-1\} \subseteq \{\iota(x), \iota(x)+1, \dots, \tau(x)\}$$

of L consecutive vertices, the probability for it to contain at least a ordinary edges in $\mathcal{S}_{\text{forward}}^p$ is at most

$$\mathbb{P}(F^c) + \sum_{i < j_1 < \dots < j_a \leq i+L-1} \mathbb{P}(s_{j_1} = s_{j_1-1}, \dots, s_{j_a} = s_{j_a-1}, F) \stackrel{(5.20)}{\leq} \mathbb{P}(F^c) + \binom{L}{a} (CN^{-\delta(1-o_\chi(1))})^a$$

Taking $a = \lfloor L/12 \rfloor$, we see that $\{i, i+1, \dots, i+L-1\}$ contains at most $L/12$ edges in $\mathcal{S}_{\text{forward}}^p$ with probability at least

$$1 - CN \exp(-N^\delta) - 2^L (CN^{-\delta(1-o_\chi(1))})^{L/12} \geq 1 - O\left(\frac{1}{N^2}\right).$$

By symmetry the same holds for edges formed by strings in $\mathcal{S}_{\text{backward}}^p$. We obtain that for each sub-interval $\{i, i+1, \dots, i+L-1\}$ in $[\iota([p]), S_p]$, it contains at most $L/6$ edges with probability $1 - O(\frac{1}{N^2})$. Taking a union bound, we see that the above happens for all i with probability at least $1 - O(\frac{1}{N})$. On this event, notice that each interval $\{i, i+1, \dots, i+L-1\}$ intersects with $\{\iota(x), \iota(x)+1, \dots, \tau(x)\}$ for at most two p 's, and thus contains at most $L/3$ edges. Thus we complete the proof. $\square$

5.3 Proof of Lemma 5.5

In this subsection, we prove a key lemma used in the proof of Lemma 5.5, namely Lemma 5.8, which, roughly speaking, estimates the probability that, conditioned on $\mathbf{a}_i$, m strings larger than s_i all start with a given prefix x . We will then use Lemma 5.8 to establish Lemma 5.5(b), (c) and (d) in the next subsection.

Recall $\mathbf{a}_i$ defined in (5.2). For convenience, we now introduce another way to characterize the distribution of the strings s_j ($j \in [N] \setminus \{\iota(p), \dots, i\}$) conditioned on $\mathbf{a}_i$. Suppose that there are $y_{p,i}$ strings in $\mathbf{a}_i$ equal to s_i , and let

$$n_l^{(t), \text{rem}} = n_l^{(t)} - \sum_{\iota(p) \leq j \leq i} 1_{\{s_j[t]=l\}} + y_{p,i} \cdot 1_{\{s_i[t]=l\}}$$

denote the number of remaining l -digits in the t -th shuffle after sampling all strings strictly smaller than s_i among $s_{\iota(p)}, \dots, s_i$. In particular, $i \in \{\iota(p), \iota(p)+1, \dots, \tau(p)\}$, and $n_p^{(1), \text{rem}} = \tau(p) - i + y_{p,i}$. For brevity, we set:

$$n^{\text{rem}} \equiv n_p^{(1), \text{rem}} = n_p^{(1)} - (i - \iota(p) + 1) + y_{p,i}.$$

We now provide another way to describe the distribution of the strings s_j ($j \in [N] \setminus \{\iota(p), \dots, i\}$) conditioned on $\mathbf{a}_i$. We also include a diagram to illustrate this representation in Figure 1.

Lemma 5.7. *Consider the set $\mathcal{M}$ of $(N - i + \iota(p) + y_{p,i} - 1) \times K$ matrices with entries in $[k]_0$ such that:*

1. *The first n^{rem} rows start with p .*
2. *The first n^{rem} rows are lexicographically greater than or equal to s_i .*
3. *Among the first n^{rem} rows, at least $y_{p,i}$ are equal to s_i .*
4. *For every $t \in [K]$ and $j \in [k]_0$, the number of j -digits in the t -th column is exactly $n_j^{(t), \text{rem}}$.*

We then select uniformly at random a matrix from $\widehat{\mathcal{M}}$. For each² $j \in [n^{\text{rem}}]$, denote by $\widehat{s}_j$ the j -th row of the selected matrix. If we sort $\{\widehat{s}_j\}_{j \in [n^{\text{rem}}]}$ into increasing lexicographic order, writing the result as $(\overline{s}'_j)_{j \in [n^{\text{rem}}]}$, then we have

$$(\overline{s}'_1, \dots, \overline{s}'_{n^{\text{rem}}}) \stackrel{\text{law.}}{=} (s_{i-y_{p,i}+1}, s_{i-y_{p,i}+2}, \dots, s_{\tau(p)}) \text{ conditioned on } \mathbf{a}_i,$$

which gives a natural coupling between the two sequences of strings with $\overline{s}'_j = s_{j+i-y_{p,i}}$ for each $j \in [n^{\text{rem}}]$, and thereby between the set $\{\widehat{s}_j\}_{j \in [n^{\text{rem}}]}$ and the sequence $(\overline{s}_{j+i-y_{p,i}})_{j \in [n^{\text{rem}}]}$. We denote this natural coupling by Q . In particular, the distribution of $|\mathcal{I}(B_x)|$ conditioned on $\mathbf{a}_i$ coincides with that of $|\{j \in [n^{\text{rem}}] : \widehat{s}_j \in B_x\}|$.

²Note the fourth condition above with $t = 1$ and $j = p$ implies that only the first n^{rem} rows begin with p ; recall $n^{\text{rem}} = n_p^{(1), \text{rem}}$.

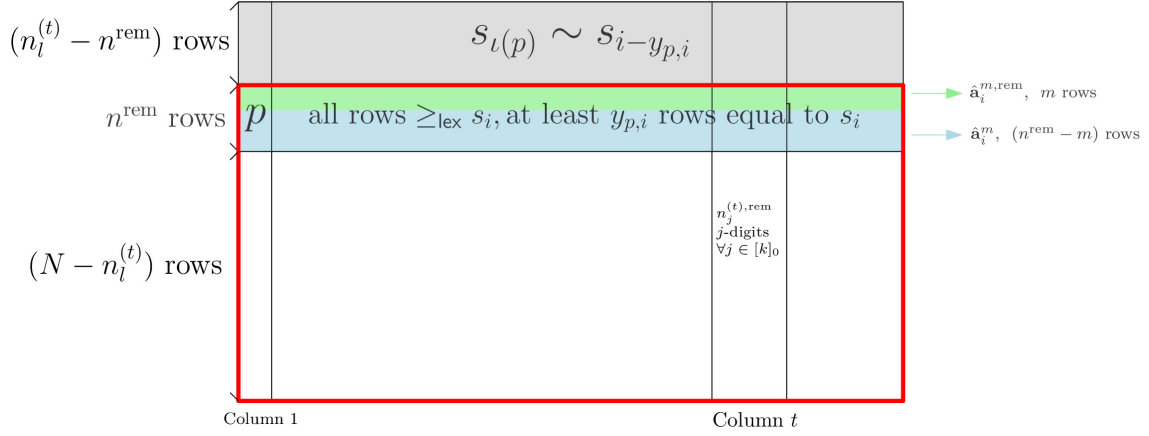


Figure 1: As defined in Lemma 5.7, a matrix in $\widehat{\mathcal{M}}$ is obtained by taking a realization of $\overline{S}_K$ and then moving all rows lexicographically smaller than s_i to the first $(i - y_{p,i} - \iota(p) + 1)$ positions, followed by placing all remaining rows that start with p in the next n^{rem} positions, while preserving their relative order. The portion enclosed by red bold lines corresponds exactly to a matrix in $\widehat{\mathcal{M}}$, and the four defining conditions of $\widehat{\mathcal{M}}$ are illustrated in the figure. The crucial Lemma 5.8, which estimates the probability that all rows in the green part start with a given prefix x conditioned on the blue part, is established by carefully counting the number of matrices whose green part satisfies certain prescribed properties.

Proof. Recall the definitions of S_K , $\overline{S}_K$, and $\bar{s}_j$ in Section 2.2. The matrices in $\widehat{\mathcal{M}}$ are precisely those obtained by removing the $(i - \iota(p) - y_{p,i} + 1)$ rows from $\overline{S}_K$ that correspond to $(s_j)_{j \in \{\iota(p), \iota(p)+1, \dots, i-y_{p,i}\}}$, and then moving all rows whose first digit is p to the first n^{rem} positions, while preserving their relative order. The four conditions defining $\widehat{\mathcal{M}}$ ensure that, when the rows $(s_j)_{j \in \{\iota(p), \iota(p)+1, \dots, i-y_{p,i}\}}$ are inserted back into any matrix in $\widehat{\mathcal{M}}$ as the first $(i - y_{p,i} - \iota(p) + 1)$ rows, and then the first $n_p^{(1)}$ rows are randomly interleaved with the remaining rows, the resulting matrix is a valid realization of $\overline{S}_K$, and the corresponding S_K satisfies that its rows from $\iota(p)$ to i are exactly $(s_j)_{j \in \{\iota(p), \iota(p)+1, \dots, i-y_{p,i}\}}$. The lemma then follows directly from the definition of s_j ; we omit further details. $\square$

For brevity, we define

$$\mu_{\mathbf{a}_i}(\cdot) = \mathbb{P}(\cdot | \mathbf{a}_i)$$

to be the conditional measure given $\mathbf{a}_i$, under which the distribution of $\hat{s}_j$ ($j = 1, 2, \dots, n^{\text{rem}} + y_{p,i}$) is defined in the proposition above. Unless otherwise specified, the rows and columns in this subsection refer to the rows and columns of the matrix defined in the previous paragraph.

For any $m \in \mathbb{Z}^+$, we write

$$\widehat{\mathbf{a}}_i^{m, \text{rem}} \stackrel{\text{def.}}{=} (\hat{s}_1, \dots, \hat{s}_m), \quad \widehat{\mathbf{a}}_i^m \stackrel{\text{def.}}{=} (\hat{s}_{m+1}, \hat{s}_{m+2}, \dots, \hat{s}_{n^{\text{rem}}}) \quad (5.21)$$

to be the first m and last $n^{\text{rem}} - m$ strings starting with p after $\mathbf{a}_i$; see Figure 1 for reference. Generating them from the distribution $\mu_{\mathbf{a}_i}(\cdot | \widehat{\mathbf{a}}_i)$, we condition on $\widehat{\mathbf{a}}_i^m$ and resample $\widehat{\mathbf{a}}_i^{m, \text{rem}}$ (or, more precisely, all rows $\hat{s}_j$ for $n^{\text{rem}} + y_{p,i} - m + 1 \leq j \leq N - i + \iota(p) + y_{p,i} - 1$) to study the distribution of the latter. We will fix m as a positive integer not depending on N so that $m \ll N$ when N is sufficiently large. This ensures that the total number of j -digits in the last $(N - n_p^{(1)} + m)$ rows of the t -th column can be approximated by $n_j^{(t)}(1 - \frac{n_p^{(1)}}{N})$ (recall that $n^{\text{rem}} = n_p^{(1), \text{rem}} = n_p^{(1)} - (i - \iota(p) + 1)$, so $N - i - n^{\text{rem}} + \iota(p) + m - 1 = N - n_p^{(1)} + m$); see also the proof of Lemma 5.8. Moreover, the condition $m \ll N$ also allows to estimate the “weight” of a sample of $\widehat{s}_{n^{\text{rem}} + y_{p,i}}$; see (5.32) and (5.34) for reference.

We now give the upper bound for the conditional probability, given $\widehat{\mathbf{a}}_i^m$, that all strings in $\widehat{\mathbf{a}}_i^{m, \text{rem}}$ start with a specific prefix x when $\widehat{\mathbf{a}}_i^m$ is “typical”. This constitutes the central step in the proof of Lemma 5.5.

We begin with defining the typical sequences of strings by explaining how to define the event F_i in Lemma 5.5. Let

$$\widehat{n}_i^{(t), \text{rem}} = n_i^{(t)} - \sum_{\iota(p) \leq j \leq \tau(p)} 1_{\{s_j[t] = l\}}.$$

For some $\delta_* \in (0, \frac{1}{2})$, we consider the following event (recall (2.18)):

$$\tilde{F}_i^* = \left\{ \left| \tilde{n}_l^{(t), \text{rem}} - n_l^{(t)} \cdot \left(1 - \frac{n_p^{(1)}}{N} \right) \right| \leq N^{\frac{1}{2} + \frac{\delta_*}{10}} \quad \text{for all } (t, l) \in \mathbf{P} \right\}. \quad (5.22)$$

Note that

$$\tilde{n}_l^{(t), \text{rem}} \sim \text{Hyp}(N - n_p^{(1)}, N, n_l^{(t)}).$$

Therefore, by Lemma 2.3, and the fact that the first shuffle is χ -good, we obtain that $\mathbb{P}(\tilde{F}_i^*) \geq 1 - \exp(-2N^{\delta_*/6})$. Furthermore, we can now define the following $\sigma(\mathbf{a}_i)$ -measurable event

$$F_i^* = \left\{ \mathbb{P}(\tilde{F}_i^* | \mathbf{a}_i) \geq 1 - \exp(-N^{\delta_*/6}) \right\}. \quad (5.23)$$

By Lemma 5.1, we have

$$\mathbb{P}(F_i^*) \geq 1 - \exp(-N^{\delta_*/6}), \quad (5.24)$$

which implies Lemma 5.5(a) by replacing η_* therein with $\delta_*/6$. We will see in the proof of the following lemma that, on the event $\tilde{F}_i^*$, the total number of j -digits in the last $(N - n_l^{(t)} + m)$ rows of the t -th column is $(1 + o(1))\tilde{n}_j^{(t)}(1 - \frac{n_p^{(1)}}{N})$ when $(t, j) \in \mathbf{P}$; see also (5.33) and its proof above and below (5.41).

Notably, we only require $\tilde{n}_l^{(t), \text{rem}}$ to concentrate well when $(t, l) \in \mathbf{P}$. This restriction is necessary because the desired concentration result for Y_x holds only when $\mathbb{E}Y_x \geq N^q$, which is necessary to ensure good concentration for $\tilde{n}_l^{(t), \text{rem}}$; see also Lemma 2.6. However for $(t, l) \in \mathbf{Q}$, $\mathbb{E}Y_x$ can be extremely small, making it impossible to achieve the desired concentration. Consequently, we can only expect concentration to hold for $\tilde{n}_l^{(t), \text{rem}}$.

The following lemma estimates the conditional probability, given $\hat{\mathbf{a}}_i^m$, that all strings in $\hat{\mathbf{a}}_i^{m, \text{rem}}$ start with a specific prefix x on the typical event F_i . Recall that $\lambda_x = \frac{1}{N} \mathbb{E}|\mathcal{I}(B_x)|$, see (2.15) for reference. Also recall that $y_{p,i}$ stands for the number of strings in $\mathbf{a}_i$ equal to s_i .

Lemma 5.8. *For all $K \geq (\tilde{C}_\mu + \varepsilon) \log N$, $\delta_* > 0$ and (χ, ρ, φ) -almost- μ -like shuffle $(\tilde{\mathbf{n}}^{(t)})_{t \leq K}$, there exists $\eta_*(\mu, \varepsilon, \delta_*) > 0$ such that the following holds for N sufficiently large. On the event F_i^* , if $s_i \in \mathbf{S}_{\text{forward}}^p$, then for each $0 < m < n^{\text{rem}}$ there exists an event $\bar{F}_i^m \in \sigma(\hat{\mathbf{a}}_i^m, \mathbf{a}_i)$ with*

$$\mu_{\mathbf{a}_i}(\bar{F}_i^m) \geq 1 - \exp(-N^{\eta_*}). \quad (5.25)$$

such that on the event $\bar{F}_i^m$, for all $x \in \mathbf{S}_{\text{forward}}^p$ and $x >_{\text{lex}} s_i$:

$$\mu_{\mathbf{a}_i}(\hat{s}_1, \dots, \hat{s}_m \text{ starts with } x | \hat{\mathbf{a}}_i^m) \leq C(\mu, m, \chi)(\lambda_x)^{m(1-o_\chi(1))} N^{O(m\chi)}.^3 \quad (5.26)$$

In addition, for all $\hat{\mathbf{a}}_i^m$ containing at least $y_{p,i}$ strings equal to s_i , on the event $\bar{F}_i^m$ we have

$$\mu_{\mathbf{a}_i}(\hat{s}_1 = \dots = \hat{s}_m = s_i | \hat{\mathbf{a}}_i^m) \leq C(\mu, m, \chi)(\lambda_{s_i})^{m(1-o_\chi(1))} N^{O(m\chi)}. \quad (5.27)$$

Since $\hat{\mathbf{a}}_i^m$ and $\hat{\mathbf{a}}_i^{m, \text{rem}}$ together contain at least $y_{p,i}$ copies of s_i (see the discussion at the beginning of this subsection), we require that $\hat{\mathbf{a}}_i^m$ alone contains at least $y_{p,i}$. This prevents any strings in $\hat{\mathbf{a}}_i^{m, \text{rem}}$ from being “forced” to equal s_i . We will use this lemma to establish Lemma 5.5(b), (c) and (d) in the next subsection.

The rigorous proof of Lemma 5.8 relies on carefully counting and estimating the number of matrices in which the m rows in $\hat{\mathbf{a}}_i^{m, \text{rem}}$ either start with x or equal to s_i , but the intuition is simpler. The condition $s_i \in \mathbf{S}_{\text{forward}}^p$ ensures that, roughly speaking, a constant proportion of strings in $[k]_0^M$ remains available for $\hat{s}_1, \dots, \hat{s}_m$ to choose from. Since λ_x denotes the proportion of strings in the $[k]_0^M$ that starts with x , so the probability that a single string starts with the prefix x is at most $C\lambda_x$ for some constant C . Because $m \ll N$, the m remaining strings $\hat{s}_1, \dots, \hat{s}_m$ are “nearly independent”, which gives the desired result.

We now turn to establishing Lemma 5.8. For simplicity, we write

$$\bar{N} = N - i - n^{\text{rem}} + \iota(p) + m - 1 = N - n_l^{(1)} + m.$$

³Here in light of (2.9), we write $O(m\chi)$ to indicate a positive sequence $(a_N)_{N \geq 0}$ satisfying $0 \leq a_N \leq C(\mu, \delta_*)m\chi$, where δ_* refers to the parameter appearing in (5.22). We follow this convention in the remainder of this section.

Moreover, recall $y_{p,i}$ denotes the number of strings in $\mathbf{a}_i$ equal to s_i , and $\widehat{\mathbf{a}}_i^m, \widehat{\mathbf{a}}_i^{m,\text{rem}}$ in (5.21). We introduce

$$\bar{n}_l^{(t),m,\text{rem}} = n_l^{(t)} - \sum_{\iota(p) \leq j \leq i - y_{p,i}} 1_{\{s_j[t]=l\}} - \sum_{j \in [n^{\text{rem}} - m]} 1_{\{\widehat{s}_{j+m}[t]=l\}}. \quad (5.28)$$

as the number of digits l not used by $\mathbf{a}_i$ and $\widehat{\mathbf{a}}_i^m$ in the t -th column.

Lemma 5.9. *Let M be the number of $\bar{N} \times K$ matrices with elements in $[k]_0 = \{0, 1, \dots, k-1\}$ such that*

- *the first digit of each the first m rows is p ;*
- *the number of l -digits among the first digits of the last $(\bar{N} - m)$ rows is exactly $n_l^{(1)}$ for all $l \in [k]_0$ and $l \neq p$;*
- *the number of l -digits in the t -th column is exactly $\bar{n}_l^{(t),m,\text{rem}}$ for all $2 \leq t \leq K$ and $l \in [k]_0$.*

We denote by $\mathcal{M}$ the set of these matrices. For all string sequences $\tilde{\mathbf{a}} = (\tilde{s}_1, \dots, \tilde{s}_m) \in (B_{[p]})^m$, the number of the matrices described above such that the first m lines are $\tilde{\mathbf{a}}$ is $M\nu(\tilde{\mathbf{a}})$, where

$$\nu(\tilde{\mathbf{a}}) = \prod_{t=2}^K \prod_{j=1}^m \frac{\bar{n}_{\tilde{s}_j[t]}^{(t),m,\text{rem}} - \sum_{i < j} 1_{\tilde{s}_i[t]=\tilde{s}_j[t]}}{\bar{N} - j + 1} = \prod_{t=2}^K \prod_{j=1}^m \frac{\bar{n}_{\tilde{s}_j[t]}^{(t),m,\text{rem}} - \sum_{i < j} 1_{\tilde{s}_i[t]=\tilde{s}_j[t]}}{N - n_l^{(1)} + j}.$$

Proof. If we fix the first m lines as $\tilde{\mathbf{a}}$, the number of such matrices is that of $(\bar{N} - m) \times K$ matrices such that the number of l -digits in the t -th column is exactly $(\bar{n}_l^{(t)} - \sum_{j \in [m]} 1_{\tilde{s}_j[t]=l})$ for all $t \in [K]$ and $j \in [k]_0$, i.e.

$$\prod_{t=1}^K \frac{(\bar{N} - m)!}{\prod_{l \in [k]_0} ((\bar{n}_l^{(t)} - \sum_{j \in [m]} 1_{\tilde{s}_j[t]=l})!)},$$

where we let $\bar{n}_p^{(1)} = 0$. Since we have $\sum_{j \in [m]} 1_{\tilde{s}_j[1]=p} = 0$ and

$$M = \frac{(\bar{N} - m)!}{\prod_{l \in [k]_0, l \neq p} (\bar{n}_l^{(1)}!)} \cdot \prod_{t=2}^K \frac{\bar{N}!}{\prod_{l \in [k]_0} (\bar{n}_l^{(t),m,\text{rem}}!)}, \quad (5.29)$$

$M\nu(\tilde{\mathbf{a}})$ is the number of the matrices whose first m lines are $\tilde{\mathbf{a}}$. $\square$

Based on this lemma, we can calculate the conditional probability $\mu_{\mathbf{a}_i}(\widehat{\mathbf{a}}_i^{\text{rem},m} = \tilde{\mathbf{a}} | \widehat{\mathbf{a}}_i^m)$ for all $\tilde{\mathbf{a}} = (\tilde{s}_1, \dots, \tilde{s}_m) \in (\mathcal{S}_{\geq \text{lex } s_i}^p)^m$, where $\mathcal{S}_{\geq \text{lex } s_i}^p$ is the set of the strings of length K that are larger than or equal to s and start with digit p . For simplicity, we introduce the following family of sequences of strings:

$$\mathcal{S}_{\geq \text{lex } s_i}^p(m, r) = \left\{ \tilde{\mathbf{a}} \in (\mathcal{S}_{\geq \text{lex } s_i}^p)^m \mid \text{There are at least } r \text{ strings in } \tilde{s}_1, \dots, \tilde{s}_m \text{ equal to } s_i. \right\} \quad (5.30)$$

Recall the definition of $\bar{n}_l^{(t),m,\text{rem}}$ in (5.28) and the fact that there are at least $y_{p,i}$ strings among $\widehat{s}_j$, $j \in [n^{\text{rem}}]$. Moreover, recall that $y_{p,i}$ is the number of strings equal to s_i in $\mathbf{a}_i$.

Lemma 5.10. *Suppose that there are y_1 strings equal to s_i among $\widehat{\mathbf{a}}_i$. Then for all $\tilde{\mathbf{a}} \in \mathcal{S}_{\geq \text{lex } s_i}^p(m, y_{p,i} - y_1)$, we have*

$$\mu_{\mathbf{a}_i}(\widehat{\mathbf{a}}_i^{\text{rem},m} = \tilde{\mathbf{a}} | \widehat{\mathbf{a}}_i^m) = \frac{\nu_{\mathbf{a}_i, \widehat{\mathbf{a}}_i^m}(\tilde{\mathbf{a}})}{\sum_{\tilde{\mathbf{a}} \in \mathcal{S}_{\geq \text{lex } s_i}^p(m, y_{p,i} - y_1)} \nu_{\mathbf{a}_i, \widehat{\mathbf{a}}_i^m}(\tilde{\mathbf{a}})}, \quad (5.31)$$

where

$$\nu_{\mathbf{a}_i, \widehat{\mathbf{a}}_i^m}(\tilde{\mathbf{a}}) = \prod_{t=1}^{K-1} \prod_{j=1}^m \frac{\bar{n}_{\tilde{s}_j[t]}^{(t),m,\text{rem}} - \sum_{i < j} 1_{\tilde{s}_i[t]=\tilde{s}_j[t]}}{N - n_l^{(1)} + j}. \quad (5.32)$$

For all $\tilde{\mathbf{a}} \in (\mathcal{S}_{\geq \text{lex } s_i}^p)^m \setminus \mathcal{S}_{\geq \text{lex } s_i}^p(m, y_{p,i} - y_1)$, we have $\mu_{\mathbf{a}_i}(\widehat{\mathbf{a}}_i^{\text{rem},m} = \tilde{\mathbf{a}} | \widehat{\mathbf{a}}_i^m) = 0$.

To prove this, we need the following characterization of $\mu_{\mathbf{a}_i}(\cdot | \widehat{\mathbf{a}}_i)$, which is an immediate result of Lemma 5.7; also see Figure 1. Recall that $\widehat{\mathbf{a}}_i^{\text{rem},m} = (\widehat{s}_1, \dots, \widehat{s}_m)$ and the distribution of $\widehat{s}_j$ ($j \in [n^{\text{rem}}]$).

Lemma 5.11. Consider all $\bar{N} \times K$ matrices with elements in $[k]_0$ such that:

1. The first m rows each start with p and are larger than or equal to s_i (in the lexicographical order);
2. The number of l -digits in the first column of each of the last $N - n_p^{(1)}$ rows is $n_l^{(1)}$ for all $l \in [k]_0 \setminus \{p\}$.
3. There are at least $(y_{p,i} - y_1)$ rows equal to s_i ;
4. The number of l -digits in the t -th column is exactly $\bar{n}_l^{(t),m,\text{rem}}$ for all $2 \leq t \leq K$ and $l \in [k]_0$.

Then for a uniformly random such matrix, the joint law of the first m rows is that of $\hat{\mathbf{a}}_i^{\text{rem},m}$ under $\mu_{\mathbf{a}_i}(\cdot | \hat{\mathbf{a}}_i)$.

Proof. The matrices considered here are precisely those obtained by removing the $(m+1)$ -th through n^{rem} -th rows from the matrices in $\widehat{\mathcal{M}}$, that is, the blue portion in Figure 1. The four requirements stated here correspond to the four conditions in Lemma 5.7. In particular, the third requirement follows because there are at least $y_{p,i}$ rows equal to s_i in total, but only y_1 rows equal to s_i among the removed rows (i.e., $\hat{\mathbf{a}}_i^m$), so the remaining rows must include at least $(y_{p,i} - y_1)$ rows equal to s_i . Hence, the desired result follows directly from the definition of $\mu_{\mathbf{a}_i}(\cdot | \hat{\mathbf{a}}_i^m)$ together with Lemma 5.7. We omit further details for brevity. $\square$

Proof of Lemma 5.10. The fourth condition in Lemma 5.11 implies that for all $\tilde{\mathbf{a}} \in (\mathcal{S}_{\geq_{\text{lex}} s_i}^p)^m \setminus \mathcal{S}_{\geq_{\text{lex}} s_i}^p(m, y_{p,i} - y_1)$, we have $\mu_{\mathbf{a}_i}(\hat{\mathbf{a}}_i^{\text{rem},m} = \tilde{\mathbf{a}} | \hat{\mathbf{a}}_i^m) = 0$. By Lemma 5.9, the number of matrices in Lemma 5.11 whose first m rows are $\tilde{\mathbf{a}}$ is $M \nu_{\mathbf{a}_i, \hat{\mathbf{a}}_i^m}(\tilde{\mathbf{a}})$, where M is a positive constant which does not depend on $\tilde{\mathbf{a}}$. Summing $M \nu_{\mathbf{a}_i, \hat{\mathbf{a}}_i^m}(\tilde{\mathbf{a}})$ over all $\tilde{\mathbf{a}} \in \mathcal{S}_{\geq_{\text{lex}} s_i}^p(m, y_{p,i} - y_1)$ gives the total number of the matrices described above. The desired result now follows from Lemma 5.11. $\square$

We now construct the event $\bar{F}_i^m$ in the statement of Lemma 5.8 and estimate the weight $\nu_{\mathbf{a}_i, \hat{\mathbf{a}}_i^m}(\tilde{\mathbf{a}})$. Recall the definition of $\bar{n}_l^{(t),m,\text{rem}}$ in (5.28). For some positive integer m and some $\delta_* \in (0, \frac{1}{2})$, we now define the event

$$\bar{F}_i^m = \left\{ \left| \bar{n}_l^{(t),m,\text{rem}} - n_l^{(t)} \cdot \left(1 - \frac{n_p^{(1)}}{N} \right) \right| \leq N^{\frac{1}{2} + \frac{\delta_*}{5}} \quad \text{for all } (t, l) \in \mathbf{P} \right\} \quad (5.33)$$

The event $\bar{F}_i^m$ is chosen so that we can establish good bounds for the weight $\nu_{\mathbf{a}_i, \hat{\mathbf{a}}_i^m}(\tilde{\mathbf{a}})$, as given in (5.32), in the case where $\hat{\mathbf{a}}_i^m$ contains at least $y_{p,i}$ strings equal to s_i . Recall the definition of λ_x in (2.15). The following lemma presents an upper bound for $\sum_{\tilde{\mathbf{a}} \in (B_x)^m} \nu_{\mathbf{a}_i, \hat{\mathbf{a}}_i^m}(\tilde{\mathbf{a}})$.

Lemma 5.12. Suppose that $\hat{\mathbf{a}}_i^m$ contains at least $y_{p,i}$ strings equal to s_i . Under the assumption of Lemma 5.8, and with $\bar{F}_i^m$ defined as in (5.33), for all $w \geq_{\text{lex}} s_i$ and $w \in \mathcal{S}_{\text{forward}}^p$, we have

$$\sum_{\tilde{\mathbf{a}} \in (B_w)^m} \nu_{\mathbf{a}_i, \hat{\mathbf{a}}_i^m}(\tilde{\mathbf{a}}) \leq C_2(\mu, m, \chi) (\lambda_w)^{m(1 - o_\chi(1))}. \quad (5.34)$$

for all $\tilde{\mathbf{a}} \in (\mathcal{S}_{\geq_{\text{lex}} s_i}^p)^m$.

Proof. We first establish that

$$\sum_{\tilde{\mathbf{a}} \in (B_w)^m} \nu_{\mathbf{a}_i, \hat{\mathbf{a}}_i^m}(\tilde{\mathbf{a}}) \leq \left(\prod_{t=2}^M \frac{\bar{n}_w^{(t),m,\text{rem}}}{N - n_p^{(1)}} \right)^m. \quad (5.35)$$

Recall that B_w is the collection of all the strings starting with the prefix w , and let the length of w be M_w . By Lemma 5.9, $M \sum_{\tilde{\mathbf{a}} \in (B_w)^m} \nu_{\mathbf{a}_i, \hat{\mathbf{a}}_i^m}(\tilde{\mathbf{a}})$ is exactly the number of $\bar{N} \times K$ matrices in $\mathcal{M}$ such that the first m lines start with w , so similarly to the proof of Lemma 5.9, we have

$$M \sum_{\tilde{\mathbf{a}} \in (B_w)^m} \nu_{\mathbf{a}_i, \hat{\mathbf{a}}_i^m}(\tilde{\mathbf{a}}) = \prod_{t=1}^{M_w} \frac{(\bar{N} - m)!}{\prod_{l \in [k]_0} ((\bar{n}_l^{(t)} - \sum_{j \in [m]} 1_{\tilde{x}[t]=l})!)} \cdot \prod_{t=M_w+1}^K \frac{\bar{N}!}{\prod_{l \in [k]_0} (\bar{n}_l^{(t),m,\text{rem}}!)}.$$

Therefore, by (5.29), we have

$$\sum_{\tilde{\mathbf{a}} \in (B_w)^m} \nu_{\mathbf{a}_i, \hat{\mathbf{a}}_i^m}(\tilde{\mathbf{a}}) = \prod_{t=2}^{M_w} \prod_{j=1}^m \frac{\bar{n}_{s_j[t]}^{(t),m,\text{rem}} - \sum_{i < j} 1_{\tilde{s}_i[t]=\tilde{s}_j[t]}}{N - n_p^{(1)} + j} \leq \left(\prod_{t=2}^M \frac{\bar{n}_w^{(t),m,\text{rem}}}{N - n_p^{(1)}} \right)^m,$$

which implies (5.35).

We now turn to estimating the product on the right side of (5.35). It suffices to estimate the product $\prod_{t=2}^{M_w} \frac{\bar{n}_{w[t]}^{(t),m,\text{rem}}}{N - n_p^{(1)}}$. Recall the definition of $\mathbf{P}$ and $\mathbf{Q}$ in (2.18).

By (5.33) and $K = O(\log N)$, there exists $C_1(\mu, m, \chi) > 0$ such that

$$\begin{aligned} \prod_{(t,w[t]) \in \mathbf{P}, 2 \leq t \leq M_w} \frac{\bar{n}_{w[t]}^{(t),m,\text{rem}}}{N - n_l^{(1)}} &\leq \prod_{(t,w[t]) \in \mathbf{P}, 2 \leq t \leq M_w} \left[\frac{n_{w[t]}^{(t)}}{N} \left(1 + C_1(\mu, m, \chi) N^{-\frac{1}{2} + \frac{\delta}{5}} \right) \right] \\ &\leq \left(1 + C_1(\mu, m, \chi) N^{-\frac{1}{2} + \frac{\delta}{4}} \right) \prod_{(t,w[t]) \in \mathbf{P}, 2 \leq t \leq M_w} \left(\frac{n_{w[t]}^{(t)}}{N} \right). \end{aligned} \quad (5.36)$$

Here we used the fact that $\frac{N - n_p^{(1)}}{N}, \frac{n_{w[t]}^{(t)}}{N} \in (\chi, 1)$ by definition for all $(t, w[t]) \in \mathbf{P}$ and N sufficiently large in the first line.

Moreover, by definition of $\bar{n}_l^{(t),m,\text{rem}}$ in (5.28), we have $\bar{n}_l^{(t),m,\text{rem}} \leq n_l^{(t)}$. Therefore, we have

$$\begin{aligned} \prod_{\substack{(t,w[t]) \in \mathbf{Q}, \\ 2 \leq t \leq M_w}} \frac{\bar{n}_{w[t]}^{(t),m,\text{rem}}}{N - n_p^{(1)}} &\leq \prod_{\substack{(t,w[t]) \in \mathbf{Q}, \\ 2 \leq t \leq M_w}} \left(\frac{n_{w[t]}^{(t)}}{N} \cdot \frac{N}{N - n_p^{(1)}} \right) \\ &\leq \prod_{\substack{(t,w[t]) \in \mathbf{Q}, \\ 2 \leq t \leq M_w}} \left(\frac{n_{w[t]}^{(t)}}{N} \right) \cdot (1 - 2\chi)^{-|\mathbf{Q} \cap \{(t,w[t]): 2 \leq t \leq M_w\}|}. \end{aligned} \quad (5.37)$$

Here we used the facts that $\frac{n_p^{(1)}}{N} \leq 1 - \chi$ (recall Subsection 3.1, and the first shuffle is χ -good). Therefore, by (2.16) we have

$$\lambda_w \leq \prod_{\substack{(t,w[t]) \in \mathbf{Q}, \\ 2 \leq t \leq M_w}} \left(\frac{n_{w[t]}^{(t)}}{N} \right) \leq \chi^{|\mathbf{Q} \cap \{(t,w[t]): 2 \leq t \leq M_w\}|}.$$

Thus, we can obtain

$$|\mathbf{Q} \cap \{(t, w[t]) : 2 \leq t \leq M_w\}| \leq \frac{\log \lambda_w}{\log \chi}.$$

Combining this with (5.37) gives

$$\prod_{\substack{(t,w[t]) \in \mathbf{Q}, \\ 2 \leq t \leq M_w}} \frac{\bar{n}_{w[t]}^{(t),m,\text{rem}}}{N - n_p^{(1)}} \leq \prod_{\substack{(t,w[t]) \in \mathbf{Q}, \\ 2 \leq t \leq M_w}} \left(\frac{n_{w[t]}^{(t)}}{N} \right) \cdot (\lambda_w)^{-\frac{\log(1-2\chi)}{\log \chi}}.$$

Combining the equation above with (2.16), (5.35) and (5.36), we can obtain

$$\begin{aligned} \sum_{\tilde{\mathbf{a}} \in (B_w)^m} \nu_{\mathbf{a}_i, \tilde{\mathbf{a}}_i^m}(\tilde{\mathbf{a}}) &\leq \left(\left(\prod_{t=2}^{M_w} \frac{n_{w[t]}^{(t)}}{N} \right)^m \cdot (\lambda_w)^{-m \frac{\log(1-2\chi)}{\log \chi}} \right) \cdot \left(1 + C_1(\mu, m, \chi) N^{-\frac{1}{2} + \frac{\delta}{4}} \right)^m \\ &\leq C_2(\mu, m, \chi) \lambda_w^{m(1-o_\chi(1))} \end{aligned} \quad (5.38)$$

where we use that m is a positive integer not depending on N in the last line. This completes the proof. $\square$

We now turn to establishing the lower bound for $\nu_{\mathbf{a}_i, \tilde{\mathbf{a}}_i^m}(\tilde{\mathbf{a}})$. When $(t, s[t]) \in \mathbf{Q}$, the number of digits can be abnormally small, preventing us from obtaining lower bounds for all $\tilde{\mathbf{a}} \in (\mathcal{S}_{\geq_{\text{lex}} s_i}^p)^m$; see also the discussion below (5.24). To address this issue, we introduce the following set of strings $\text{PM}(B_x)$ for each prefix x :

$$\text{PM}(B_x) = \{s \in B_x : (t, s[t]) \in \mathbf{P} \text{ for all } t \in [K]\}. \quad (5.39)$$

Thus $\text{PM}(B_x)$ consists of strings with prefix x all of whose digits are “ χ -good”. This restriction allows us to apply the definition of $\bar{F}_i^m$ to bound each term in the product appearing in (5.32).

Note that it is possible for $\text{PM}(B_x)$ to be empty for certain prefixes x , but we will choose the prefixes considered here carefully so that $\text{PM}(B_x) \neq \emptyset$. Moreover, restricting attention to strings in $\text{PM}(B_x)$ rather than $\mathcal{I}(B_x)$ typically does not discard a significant proportion of the strings $(s_j)_{j=i+1, \dots, \tau[p]}$ in each shuffle, provided χ is sufficiently small, since at least a proportion of $1 - (k-1)\chi$ of strings lie in $\mathbf{P}$ at each shuffle; see also (5.43) below.

We can provide the following lower bound for $\nu_{\mathbf{a}_i, \widehat{\mathbf{a}}_i^m}(\widetilde{\mathbf{a}})$ for all $\widetilde{\mathbf{a}} \in (\text{PM}(B_x))^m$.

Lemma 5.13. *Suppose that $\widehat{\mathbf{a}}_i^m$ contains at least y strings equal to s_i . Under the assumption of Lemma 5.8, and with $\overline{F}_i^m$ defined as in (5.33), there exists a constant $C_4(\mu, m, \chi) > 0$ not depending on the sequences of strings $\widehat{\mathbf{a}}_i^m$, $\mathbf{a}_i$ and $\widetilde{\mathbf{a}}$, such that*

$$\nu_{\mathbf{a}_i, \widehat{\mathbf{a}}_i^m}(\widetilde{\mathbf{a}}) \geq C_4(\mu, m, \chi) \prod_{j=1}^m \lambda_{\widetilde{s}_j} \quad (5.40)$$

for all $\widetilde{\mathbf{a}} \in (\text{PM}(B_x))^m$.

Proof. By (5.33), we obtain for all prefixes x , string $\widetilde{s} \in \text{PM}(B_x)$ and $j \in [m]$,

$$\frac{\overline{n}_{\widetilde{s}_j[t]}^{(t), m, \text{rem}} - \sum_{i < j} 1_{\widetilde{s}_i[t] = \widetilde{s}_j[t]}}{N - n_p^{(1)} + j} \geq \frac{n_{\widetilde{s}_j[t]}^{(t)}}{N} \left(1 + C_3(\mu, m, \chi) N^{-\frac{1}{2} + \frac{\delta}{6}}\right)$$

for some positive constant $C_3(\mu, m, \chi)$ not depending on the sequences of strings $\widehat{\mathbf{a}}_i^m$, $\mathbf{a}_i$ and $\widetilde{\mathbf{a}}$. This holds because, by definition, $\frac{N - n_p^{(1)}}{N}, \frac{n_{\widetilde{s}_j[t]}^{(t)}}{N} \in (\chi, 1)$ for all $j \in [m]$ and N sufficiently large in the first line. Imitating the calculation in (5.36) and (5.38) now gives the desired bound. $\square$

With Lemmas 5.12 and 5.13 in hand, we are now ready to establish Lemma 5.8.

Proof of Lemma 5.8. We first prove (5.25). Fix a positive integer m in the proof from now on. Recall the definition of $\overline{n}_l^{(t), m, \text{rem}}$ in (5.28), the definition of F_i in (5.23), the definition of $\overline{F}_i^m$ in (5.33) and the fact that there exists a natural coupling Q between $\widetilde{s}_j$ ($j = 1, 2, \dots, n^{\text{rem}}$) under $\mu_{\mathbf{a}_i}$ and $(s_{i+1}, \dots, s_{\tau(p)})$ conditioned on $\mathbf{a}_i$, see the part below Lemma 5.1 for reference. Therefore, we obtain that under Q ,

$$|\overline{n}_l^{(t), m, \text{rem}} - n_l^{(t), \text{rem}}| \leq m \quad (5.41)$$

almost surely. Hence, by (5.24) and Lemma 5.1, for all $\mathbf{a}_i \in F_i^*$, $\mu_{\mathbf{a}_i}(\overline{F}_i^m) \geq 1 - \exp(-N^{\delta^*/6})$. Taking $\eta_* = \frac{\delta_*}{6}$ concludes the proof of (5.25).

We now turn to the proof of (5.26). Let $\overline{\mathcal{F}}_i^m(\mathbf{a}_i)$ denote the collection of all sequences $\widetilde{\mathbf{a}} \in ([k]_0^K)^{n^{\text{rem}} - m}$ such that $\mathbf{a}_i$ and $\widetilde{\mathbf{a}}$ are two sequences such that $\overline{F}_i^m$ occurs. We now show that for all $\widehat{\mathbf{a}}_i^m \in \overline{\mathcal{F}}_i^m(\mathbf{a}_i)$, (5.26) holds.

If the number of strings equal to s_i in $\widehat{\mathbf{a}}_i^m$ is less than y , then $\widehat{\mathbf{a}}_i^{\text{rem}, m}$ contains at least one string equal to s_i . Since $x > s_i$, the left-hand side of (5.26) equals 0 and (5.26) holds.

In the case where $\widehat{\mathbf{a}}_i^m$ contains at least y strings equal to s_i , the event $\overline{F}_i^m$ is chosen so that we can derive a bound for $\nu_{\mathbf{a}_i, \widehat{\mathbf{a}}_i^m}(\widetilde{\mathbf{a}})$, as given in (5.32). Recall the definition of $\mathcal{S}_{\geq \text{lex } s_i}^p(m, r)$ in (5.30), and observe that $\mathcal{S}_{\geq \text{lex } s_i}^p(m, y_{p,i} - y_1) = (\mathcal{S}_{\geq \text{lex } s_i}^p)^m$ since $y_1 \geq y_{p,i}$, where y_1 represents the number of strings equal to s_i in $\widehat{\mathbf{a}}_i^m$.

Recall the definition of $\mathbf{I}_{\text{right}}^{(t)}$ in Definition 3.6, $\mathbf{S}_{\text{forward}}^p$ in Definition 3.7 and a_k in (3.3). Therefore, by the fact that the first shuffle is χ -good, in the nonempty set $\text{PM}(B_{[p\mathbf{I}_{\text{right}}^{(2)} \mathbf{I}_{\text{right}}^{(3)} \dots \mathbf{I}_{\text{right}}^{(a_3)}]})$, all strings start with $[p\mathbf{I}_{\text{right}}^{(2)} \mathbf{I}_{\text{right}}^{(3)} \dots \mathbf{I}_{\text{right}}^{(a_3)}]$. By (5.31), we have

$$\begin{aligned} \mu_{\mathbf{a}_i}(\widehat{s}_{n^{\text{rem}}-m+1}, \dots, \widehat{s}_{n^{\text{rem}}} \text{ starts with } x | \widehat{\mathbf{a}}_i^m) &= \sum_{\widetilde{\mathbf{a}} \in (B_x)^m} \mu_{\mathbf{a}_i}(\widehat{\mathbf{a}}_i^{\text{rem}, m} = \widetilde{\mathbf{a}} | \widehat{\mathbf{a}}_i^m) \\ &\stackrel{(5.31)}{\leq} \frac{\sum_{\widetilde{\mathbf{a}} \in (B_x)^m} \nu_{\mathbf{a}_i, \widehat{\mathbf{a}}_i^m}(\widetilde{\mathbf{a}})}{\sum_{\widetilde{\mathbf{a}} \in (\text{PM}(B_{[p\mathbf{I}_{\text{right}}^{(2)} \mathbf{I}_{\text{right}}^{(3)} \dots \mathbf{I}_{\text{right}}^{(a_3)}]})^m} \nu_{\mathbf{a}_i, \widehat{\mathbf{a}}_i^m}(\widetilde{\mathbf{a}})}. \end{aligned} \quad (5.42)$$

The numerator has been estimated in (5.34). For the denominator, we can now derive from (5.40):

$$\begin{aligned}
\sum_{\tilde{\mathbf{a}} \in (\text{PM}(B_{[pI_{\text{right}}^{(2)} I_{\text{right}}^{(3)} \dots I_{\text{right}}^{(a_3)}]}))^m} \nu_{\mathbf{a}_i, \tilde{\mathbf{a}}_i^m}(\tilde{\mathbf{a}}) &\geq \prod_{j=1}^m \left(\sum_{\tilde{s}_j \in \text{PM}(B_{[pI_{\text{right}}^{(2)} I_{\text{right}}^{(3)} \dots I_{\text{right}}^{(a_3)}]}))} \lambda_{\tilde{s}_j} \right) \\
&= \left(\frac{n_p^{(1)}}{N} \cdot \sum_{\tilde{s}_j \in \text{PM}(B_{[pI_{\text{right}}^{(2)} I_{\text{right}}^{(3)} \dots I_{\text{right}}^{(a_3)}]}))} \left(\prod_{2 \leq t \leq M} \frac{n_{\tilde{s}_j[t]}^{(t)}}{N} \right) \right)^m \\
&\stackrel{(5.39)}{=} \left(\frac{n_p^{(1)}}{N} \cdot \prod_{2 \leq t \leq a_3} \frac{n_{I_{\text{right}}^{(t)}}^{(t)}}{N} \cdot \prod_{a_3+1 \leq t \leq K} \left(\sum_{l: (t,l) \in P} \frac{n_l^{(t)}}{N} \right) \right)^m \\
&\stackrel{(2.18)}{\geq} \left(\frac{n_p^{(1)}}{N} \cdot \prod_{2 \leq t \leq a_3} \frac{n_{I_{\text{right}}^{(t)}}^{(t)}}{N} \cdot \prod_{a_3+1 \leq t \leq K} \left(1 - (k-1)\chi \right) \right)^m \\
&= N^{m \log(1-(k-1)\chi) \cdot \frac{K-a_3}{\log N}} (\lambda_{[pI_{\text{right}}^{(2)} I_{\text{right}}^{(3)} \dots I_{\text{right}}^{(a_3)}]})^m.
\end{aligned} \tag{5.43}$$

Since $(1, p)$ and $(a_i, I_{\text{right}}^{(a_i)})$ (for all $i \geq 1$) are not in $\mathcal{Q}$, we obtain

$$\lambda_{[pI_{\text{right}}^{(2)} I_{\text{right}}^{(3)} \dots I_{\text{right}}^{(a_3)}]} \geq \chi^3 (1 - (k-1)\chi)^{a_3}.$$

Combining (5.42), (5.34) and (5.43) now gives the desired result (5.26).

We now establish (5.27). Recall the definition of $\mathcal{S}_{\geq_{\text{lex}} s_i}^p(m, r)$ in (5.30). Since there are at least $y_{p,i}$ strings in $\tilde{\mathbf{a}}_i^m$ equal to s_i , $\mathcal{S}_{\geq_{\text{lex}} s_i}^p(m, y_{p,i} - y_1) = (\mathcal{S}_{\geq_{\text{lex}} s_i})^m$. Therefore, by (5.42), (5.43) and applying Lemma 5.12 with $w = s_i$, we can prove (5.26). This concludes the proof of Lemma 5.8. $\square$

5.4 Proof of Lemma 5.5

In this subsection, we will deduce Lemma 5.5. Parts (b),(c) and (d) essentially follow from Lemma 5.8. In order to establish Lemma 5.5(e),(f), we introduce the events

$$H_x = \left\{ \begin{aligned} &||\mathcal{I}(B_x)| - N\lambda_x| \leq (N\lambda_x)^{\frac{4}{5}}, \\ &\left| n_l^{(t),x} - N\lambda_x \cdot \frac{n_l^{(t)}}{N} \right| \leq (N\lambda_x)^{\frac{4}{5}} \text{ for all } (t, l) \in P \text{ and } M < t \leq K. \end{aligned} \right\}.$$

and

$$H_{i,x} = \{\mathbb{P}(H_x | \mathbf{a}_i) \geq 1 - \exp(-N^{\xi/2})\}.$$

Recall from Lemma 2.6 that T_ξ is the set of all strings x such that $\mathbb{E}|\mathcal{I}(B_x)| \geq N^\xi$. For simplicity, we write

$$H_{i,\xi} = \bigcap_{x \in T_\xi} H_{i,x}. \tag{5.44}$$

Define the $\sigma(\mathbf{a}_i)$ -measurable event

$$F_i = \bigcap_{j=S_{p-1}+1}^i (F_j^* \cap H_{j,\xi}), \tag{5.45}$$

for all $i \in \{\iota(p), \iota(p) + 1, \dots, \tau(p) - 2\}$.

Recall that $E(G_{B_x})$ denotes the edges in G with vertices in B_x . Also recall the definition of $\mathbf{p}^{(i)}$ in (2.10) and $\mathcal{I}(B_x)$ in (2.14). We now provide the proof of Lemma 5.5.

Proof of Lemma 5.5. We fix $\delta_* > 0$ so that we can construct the events F_i^* and $\overline{F}_i^1$ (resp. $\overline{F}_i^2$) as in Lemma 5.8 so that (5.26), (5.25) hold for $m = 1$ (resp. $m = 2$). In addition, we fix some $\xi > 0$ to construct the event $H_{i,\xi}$, and Lemma 5.5(e), (f) follows from the definition of $H_{i,x}$, $H_{i,\xi}$ and F_i . We then proceed to prove parts (a)–(d) of Lemma 5.5 sequentially.

Lemma 5.5(a). By Lemma 5.8, we have $\mathbb{P}(F_i^*) \geq 1 - \exp(-N^\eta)$. Moreover, by Lemma 2.3, we have

$$\mathbb{P}(H_x) \geq 1 - \exp(-2N^{\xi/2}).$$

Therefore, by Lemma 5.1, the event $H_{i,x}$ occurs with probability higher than $1 - \exp(-N^{\xi/2})$. By (5.44), since $|T_\xi| = N^{O(1)}$, the event $H_{i,\xi}$ occurs with probability higher than $1 - \exp(-\frac{1}{2}N^{\xi/2})$. Therefore, by (5.45) and the union bound, we have F_i occurs with probability at least $1 - \exp(-N^\zeta)$ for all ζ sufficiently small.

Lemma 5.5(b). By Lemma 5.7, the distribution of $|\mathcal{I}(B_x)|$ conditioned on $\mathbf{a}_i$ is the same as that of $|\{j \in [n^{\text{rem}}] : \hat{s}_j \in B_x\}|$ under $\mu_{\mathbf{a}_i}$; see Lemma 5.7 for the definition of $\hat{s}_j$. By symmetry, we have $\mu_{\mathbf{a}_i}(\hat{s}_j \in B_x) = \mu_{\mathbf{a}_i}(\hat{s}_1 \in B_x)$ for all $j \in [n^{\text{rem}}]$. Note that

$$\begin{aligned} \mu_{\mathbf{a}_i}(\hat{s}_{n^{\text{rem}}} \in B_x) &= \mathbb{E}_{\mu_{\mathbf{a}_i}}[\mu_{\mathbf{a}_i}(\hat{s}_j \in B_x | \hat{\mathbf{a}}_i^1; \bar{F}_i^1) + \mathbb{E}_{\mu_{\mathbf{a}_i}}[\mu_{\mathbf{a}_i}(\hat{s}_j \in B_x | \hat{\mathbf{a}}_i^1); (\bar{F}_i^1)^c] \\ &\stackrel{(5.26), (5.25)}{\leq} C(\mu, \chi)(\lambda_x)^{1-o_\chi(1)} \cdot N^{O(\chi)} + \exp(-N^{\eta*}), \end{aligned}$$

where $\hat{\mathbf{a}}_i^1 = (\hat{s}_1, \dots, \hat{s}_{n^{\text{rem}}-1})$. Therefore, we have

$$\begin{aligned} \mathbb{E}[|\mathcal{I}(B_x)| | \mathbf{a}_i] &= \sum_{j \in [n^{\text{rem}}]} \mu_{\mathbf{a}_i}(\hat{s}_j \in B_x) \\ &\leq N \left(C(\mu, \chi)(\lambda_x)^{1-o_\chi(1)} \cdot N^{O(\chi)} + \exp(-N^{\eta*}) \right) \\ &\leq C(\mu, \chi) N(\lambda_x)^{1-o_\chi(1)} \cdot N^{O(\chi)}. \end{aligned} \tag{5.46}$$

This concludes the proof of Lemma 5.5(b).

Lemma 5.5(c). We introduce the events

$$A(i, j, s_i) = \{\text{There are at least } y \text{ strings equal to } s_i \text{ among } \hat{s}_k \text{ for } k \in [n^{\text{rem}}] \setminus \{j\}\},$$

and

$$A(i, s_i) = \bigcap_{j \in [n^{\text{rem}}]} A(i, j, s_i).$$

Recall that y is the number of strings equal to s_i in $\mathbf{a}_i$, and by Lemma 5.7, there exists a natural coupling Q between $\hat{s}_j$ ($j = 1, 2, \dots, n^{\text{rem}}$) under $\mu_{\mathbf{a}_i}$ and $(s_{i-y_{p,i}+1}, \dots, s_{\tau(p)})$ conditioned on $\mathbf{a}_i$, and we have

$$|j > i : s_j = s_i| = |j \in [n^{\text{rem}}] : \hat{s}_j = s_i| - y.$$

Therefore, for all $j \in [n^{\text{rem}}]$, on the event $A(i, j, s_i)^c$, we have $|l > i : s_l = s_i| = 0$ and so

$$\begin{aligned} \mathbb{E}[|j > i : s_j = s_i| | \mathbf{a}_i] &= \mathbb{E}_{\mu_{\mathbf{a}_i}}[(|j \in [n^{\text{rem}}] : \hat{s}_j = s_i| - y) \cdot 1_{A(i, s_i)}] \\ &\leq \mathbb{E}_{\mu_{\mathbf{a}_i}}[|j \in [n^{\text{rem}}] : \hat{s}_j = s_i| \cdot 1_{A(i, s_i)}] \\ &= \sum_{j \in [n^{\text{rem}}]} \mu_{\mathbf{a}_i}(\hat{s}_j = s_i; A(i, s_i)) \\ &\leq \sum_{j \in [n^{\text{rem}}]} \mu_{\mathbf{a}_i}(\hat{s}_j = s_i; A(i, j, s_i)). \end{aligned}$$

By symmetry, $\mu_{\mathbf{a}_i}(\hat{s}_j = s_i; A(i, j, s_i)) = \mu_{\mathbf{a}_i}(\hat{s}_1 = s_i; A(i, n^{\text{rem}}, s_i))$. Therefore, we only need to upper-bound the latter. Note that $A(i, 1, s_i)$ is exactly the event that $\hat{\mathbf{a}}_i^m$ contains at least y strings equal to s_i , so by (5.27) and imitating the calculation in (5.46), we can obtain Lemma 5.5(c).

Lemma 5.5(d). Note that by Lemma 5.7, for all strings $\tilde{s}$ larger than s_i , the distribution of $|E(G, \tilde{s})|$ conditioned on $\mathbf{a}_i$ is the same as that of $|\{j, j' \in [n^{\text{rem}}] : \hat{s}_j = \hat{s}_{j'} = \tilde{s}, j \neq j'\}|$ under $\mu_{\mathbf{a}_i}$. The argument similar to the proof of Lemma 5.5(b) gives

$$\mu_{\mathbf{a}_i}(\hat{s}_j = \hat{s}_{j'} = \tilde{s}) = \mu_{\mathbf{a}_i}(\hat{s}_1 = \hat{s}_2 = \tilde{s}) \leq C(\mu, \chi)(\lambda_{\tilde{s}})^{2-o_\chi(1)} \cdot N^{O(\chi)} \tag{5.47}$$

for all $j \neq j'$.

Recall that the length of prefix x is M and note that

$$B_x = \{[xw] : w \in [k]_0^{K-M}\}.$$

Therefore, we have

$$\begin{aligned} \mathbb{E} [|E(G_{B_x})| | \mathbf{a}_i] &\leq \sum_{w \in [k]_0^{K-M}} \mathbb{E}_{\mu_{\mathbf{a}_i}} [|\{j, j' \in [n^{\text{rem}}] : \hat{s}_j = \hat{s}_{j'} = [xw], j \neq j'\}|] \\ &= \sum_{w \in [k]_0^{K-M}} \sum_{1 \leq j < j' \leq n^{\text{rem}}} \mu_{\mathbf{a}_i}(\hat{s}_j = \hat{s}_{j'} = [xw]) \\ &\stackrel{(5.47)}{\leq} C(\mu, \chi) N^{2+O(\chi)} \sum_{w \in [k]_0^{K-M}} (\lambda_{[xw]})^{2-o_\chi(1)}. \end{aligned} \tag{5.48}$$

In addition, using (2.15), the last sum is at most

$$\begin{aligned} \sum_{w \in [k]_0^{K-M}} (\lambda_{[xw]})^{2-o_\chi(1)} &= \sum_{w \in [k]_0^{K-M}} \prod_{t=1}^M \left(\frac{n_{x[t]}}{N} \right)^{2-o_\chi(1)} \prod_{t=M+1}^K \left(\frac{n_{w[t-M]}}{N} \right)^{2-o_\chi(1)} \\ &\leq (\lambda_x)^{2-o_\chi(1)} \sum_{w \in [k]_0^{K-M}} \prod_{M+1 \leq t \leq K} \left(\frac{n_{w[t-M]}}{N} \right)^{2-o_\chi(1)} \\ &= (\lambda_x)^{2-o_\chi(1)} \prod_{M+1 \leq t \leq K} \left(\sum_{l \in [k]_0} \left(\frac{n_l}{N} \right)^{2-o_\chi(1)} \right) \\ &= (\lambda_x)^{2-o_\chi(1)} N^{-\frac{1}{\log N} \sum_{t>M} \psi_{\mathbf{n}^{(t)}/N}(2-o_\chi(1))}. \end{aligned} \tag{5.49}$$

By the definition of $\psi_{\mathbf{p}}(x)$, for all $x \in (1, 2]$ and $\mathbf{p} \in \mathcal{D}_k$ (i.e. $\sum_{l \in [k]_0} p_l = 1$ and $0 \leq p_l \leq 1$ for all $l \in [k]_0$), we have

$$\left| \frac{d\psi_{\mathbf{p}}(x)}{dx} \right| = \frac{\sum_{l \in [k]_0} p_l^x \log(p_l^{-1}) \cdot 1_{\{p_l \neq 0\}}}{\sum_{l \in [k]_0} p_l^x} \stackrel{x \in (1, 2]}{\leq} \frac{\sum_{l \in [k]_0} p_l \log(p_l^{-1}) \cdot 1_{\{p_l \neq 0\}}}{\sum_{l \in [k]_0} p_l^2} \leq k^2 \cdot \sup_{p \in (0, 1]} [p \log(1/p)].$$

(Note $p \log(1/p)$ is bounded on $p \in (0, 1]$.) Since $\mathbf{n}^{(t)}/N \in \mathcal{D}_k$, we find that for all $\varsigma \in (0, \frac{1}{2})$:

$$|\psi_{\mathbf{n}^{(t)}/N}(2-\varsigma) - \psi_{\mathbf{n}^{(t)}/N}(2)| = \varsigma \cdot \left| \frac{d\psi_{\mathbf{p}}}{dx}(\bar{x}) \right| = O(\varsigma),$$

where $\bar{x}$ is a real number in $[2-\varsigma, 2]$. This implies that $\psi_{\mathbf{n}^{(t)}/N}(2-o_\chi(1)) = \psi_{\mathbf{n}^{(t)}/N}(2) \pm o_\chi(1)$. Plugging this inequality into (5.49), we obtain

$$\sum_{w \in [k]_0^{K-M}} (\lambda_{[xw]})^{2-o_\chi(1)} \leq (\lambda_x)^{2-o_\chi(1)} N^{-\frac{1}{\log N} \sum_{t>M} \psi_{\mathbf{n}^{(t)}/N}(2)+o_\chi(1)} \tag{5.50}$$

Recall from (4.11) that $c_E(x) = -\frac{1}{\log N} \sum_{t>M} \psi_{\mathbf{n}^{(t)}/N}(2)$. Therefore, combining (4.14), (5.48) and (5.50) gives (5.9). This concludes the proof of Lemma 5.5(d). $\square$

6 Lower Bound

The definition of $\overline{C}_\mu$ as a maximum in (1.6) corresponds to two distinct obstructions to mixing: long increasing contiguous subsequences in the inverse permutation and *cold spots*. Below we explain these obstructions separately in the following subsections, obtaining (1.7) which implies (2.3).

6.1 Increasing sequences

It is a rare event for a uniformly random permutation to have a long increasing contiguous subsequence of size $N^{\Omega(1)}$. Recall that $\tilde{C}_\mu = \frac{1}{\mathbb{E}_\mu \log(1/p_{\max})}$. The following proposition shows that if $K \leq (\tilde{C}_\mu - \varepsilon) \log N$, then an inverse K -shuffled permutation typically contains a long increasing sequence, and so the deck is not mixed. In particular, there will typically be $N^{\Omega(1)}$ strings consisting only of the most common digit in each shuffle. Due to the non-binomial cut sizes, the number of such strings is not binomial. However, a simple application of Lemma 2.6 gives the concentration.

Proposition 6.1. *If $K < (\tilde{C}_\mu - \varepsilon) \log N$, and $\mathbf{X}$ is randomized- μ -like, we have that*

$$\lim_{N \rightarrow \infty} d_{\text{TV}}(\mathbf{P}(\mathbf{X}, K), \nu_{\text{Unif}}) = 1.$$

Proof. We fix $\chi, \rho, \varphi \ll \varepsilon$ and condition that the shuffle process is (χ, ρ, φ) -almost- μ -like. Recall the definition of $l_{\max}$ above (2.12). Let x be the string with length K such that

$$x[t] = l_{\max}^{(t)}, \text{ for all } 1 \leq t \leq K.$$

Also recall that Y_x denotes the number of strings in S_K equal to x . The first moment of Y_x is at least:

$$\begin{aligned} \mathbb{E}Y_x &= N \cdot \prod_{t \leq K} \frac{n_i^{(t)}}{N} \geq N \cdot \prod_{0 \leq i \leq z} (p_{\max}^{(i)} - \chi)^{Kh_i - O(\varphi) - O(\rho)} \\ &\geq N^{1 - \frac{K}{\log N} \cdot \mathbb{E}_\mu \log(1/p_{\max}) - O(\delta)} \geq N^{\Omega(\varepsilon)}. \end{aligned} \quad (6.1)$$

Here we used the fact that $\chi, \rho, \varphi \ll \delta \ll \varepsilon$ and $K \leq (C_\mu - \varepsilon) \log N$. Thus by (2.7) and by taking $q = \Omega(\varepsilon)$ we obtain that

$$\mathbb{P}(Y_x \geq \frac{1}{2} \mathbb{E}Y_x) \geq 1 - \exp(-N^{\Omega(\varepsilon)}).$$

The above calculation and Lemma 3.13 indicate that with probability $1 - o(1)$, there exists an increasing sequence of length $N^{\Omega_\varepsilon(1)}$ after K shuffles. However, if $\pi \sim \nu_{\text{Unif}}$, we have

$$\nu_{\text{Unif}}(\pi \text{ has an increasing sequence of length } N^{\Omega_\varepsilon(1)}) \leq N \cdot \frac{1}{(N^{\Omega(\varepsilon)})!} = o(1),$$

which completes the proof. $\square$

Remark 6.2. *When $\mu(V_k) = 1$, we can define Y_x in the same way as in Proposition 6.1. The argument above indicates that for all $C > 0$, $K = C \log N$, and for sufficiently small $\chi = \chi(C), \rho = \rho(C), \varphi = \varphi(C)$, (6.1) can be modified into*

$$\mathbb{E}Y_x \geq N^{1+C \log(1-\chi)} \geq N^{\frac{1}{2}}.$$

Therefore we can conclude (1.9) using same arguments as in Proposition 6.1.

6.2 Cold spots

In this subsection, we take $K = \lfloor (C_\mu - \varepsilon) \log N \rfloor$ and further assume $K \geq (\tilde{C}_\mu + \varepsilon) \log N$. The following criterion for non-mixing follows exactly as in [Lal00].⁴

Lemma 6.3 ([Lal00, Proof of Proposition 2]). *For all shuffle process with pile size $\vec{n}$, suppose that there exists $H(\vec{n}) \subset [N]$ such that (with $E(H)$ the subgraph of the path graph induced by vertex set H):*

$$\begin{aligned} |H(\vec{n})| &\geq N^{1/2} \\ |\partial H(\vec{n})| &\leq 2|H|^{1/2} \\ \mathbb{P} \left[|E(G) \cap E(H(\vec{n}))| \geq |H(\vec{n})|^{\frac{1}{2}+\delta} \right] &\rightarrow 1. \end{aligned} \quad (6.2)$$

Then if $\nu \sim \mathbf{P}(\vec{n}, K)$ and $K \leq (\tilde{C}_\mu - \varepsilon) \log N$, we have

$$\nu \left(\#\{i \in H : \sigma(i) < \sigma(i+1)\} \leq |H(\vec{n})|/2 + |H(\vec{n})|^{1/2+\delta/2} \right) \rightarrow 0, \quad (6.3)$$

where the limit is uniform for all $\vec{n}$.

⁴In fact, the result in [Lal00, Proof of Proposition 2] is more general in that the explicit hypothesis $|H(\vec{n})| \geq N^{1/2}$ in Lemma 6.3 can be relaxed to $|H(\vec{n})| \rightarrow \infty$ as $N \rightarrow \infty$. We provide this explicit rate of divergence for clarity.

We will apply Lemma 6.3 for each (χ, ρ, φ) -almost- μ -like $\vec{n}$. We will prove in Lemma 6.4 that for each (χ, ρ, φ) -almost- μ -like $\vec{n}$, we can construct a subset $H(\vec{n})$ satisfying the conditions in Lemma 6.3. The proof of Lemma 6.4 will be postponed in Section 6.3.

Lemma 6.4. *For each (χ, ρ, φ) -almost- μ -like $\vec{n}$, there exists $H(\vec{n}) \subset [N]$ such that (6.2) holds.*

We then let $A_{\vec{n}}$ be the event that

$$|\{i \in H(\vec{n}) : \sigma(i) < \sigma(i+1)\}| > |H(\vec{n})|/2 + |H(\vec{n})|^{1/2+\delta/2}.$$

We also denote by $A_{\mu, \chi, \rho, \varphi}$ the event that $A_{\vec{n}}$ occurs for at least one (χ, ρ, φ) -almost- μ -like $\vec{n}$.

Lemma 6.5. *Suppose that $\mathbf{X}$ is randomized- μ -like, $\nu \sim \mathbf{P}(\mathbf{X}, K)$ and $K \leq (\tilde{C}_\mu - \varepsilon) \log N$. We have*

$$\nu(A_{\mu, \chi, \rho, \varphi}) \rightarrow 1.$$

Proof. By Lemma 6.3, we have that for all (χ, ρ, φ) -almost- μ -like pile sizes $\vec{n}$,

$$\nu(A_{\vec{n}} \text{ occurs} \mid \mathbf{X} = \vec{n}) \rightarrow 1 \text{ uniformly in } \vec{n}.$$

Thus we have

$$\begin{aligned} \nu(A_{\mu, \chi, \rho, \varphi}) &\geq \sum_{\vec{n} \text{ is } (\chi, \rho, \varphi)\text{-almost-}\mu\text{-like}} \left[\nu_{\text{Unif}}(A_{\mu, \chi, \rho, \varphi} \mid \mathbf{X} = \vec{n}) \nu_{\text{Unif}}(\mathbf{X} = \vec{n}) \right] \\ &\geq (1 - o(1)) \cdot \nu_{\text{Unif}}(\mathbf{X} \text{ is } (\chi, \rho, \varphi)\text{-almost-}\mu\text{-like}) = 1 - o(1). \end{aligned}$$

Here in the last line we used Lemma 3.13. $\square$

The event in (6.3) defines a hypothesis test to detect a $\mathbf{P}(\vec{n}, K)$ -shuffled deck. Next we show that for a uniformly random deck, one can run many such tests for different $\vec{n}$, and with high probability none of these tests will detect non-uniformity. We now view $H(\cdot)$ as a mapping from the set of all (χ, ρ, φ) -almost- μ -like pile sizes to subsets of $[N]$.

Lemma 6.6. *With ν_{Unif} the uniform measure on S_N and $A_{\mu, \chi, \rho, \varphi}$ the event defined in Lemma 6.5, we have*

$$\nu_{\text{Unif}}(A_{\mu, \chi, \rho, \varphi}) \rightarrow 0$$

for any mappings $H(\cdot)$ such that $|H(\vec{n})| \geq N^{1/2}$ and $|\partial H(\vec{n})| = O(|H(\vec{n})|^{1/2})$ for all (χ, ρ, φ) -almost- μ -like pile sizes $\vec{n}$.

Proof. Notice that for $j > i + 1$, we have that the events $\{\sigma(i) < \sigma(i+1)\}$ and $\{\sigma(j) < \sigma(j+1)\}$ are independent under ν_{Unif} , and both hold with probability $1/2$. Therefore, for fixed $\vec{n}$, if we define

$$\begin{aligned} E^{\text{odd}} &= \#\{i \text{ is odd} : (i, i+1) \in |E(G) \cap E(H(\vec{n}))|\}, \\ E^{\text{even}} &= \#\{i \text{ is even} : (i, i+1) \in |E(G) \cap E(H(\vec{n}))|\}, \end{aligned}$$

then E^{odd} and E^{even} are binomial. Thus by Hoeffding's inequality, we may obtain the concentration for $|E(G) \cap E(H(\vec{n}))|$:

$$\nu_{\text{Unif}}(|E(G) \cap E(H(\vec{n}))| \geq |H(\vec{n})|/2 + |H(\vec{n})|^{1/2+\delta/2}) = O(\exp(-|H(\vec{n})|^\delta)) = O(\exp(-N^{\delta/2})).$$

Here we used the fact that $|\partial H(\vec{n})| = O(|H(\vec{n})|^{1/2})$ and $|H(\vec{n})| \geq N^{1/2}$. Since the total choice of $\vec{n}$ is less than $N^{Kk} = \exp(O(\log N)^2)$, we then conclude by a union bound over $\vec{n}$. $\square$

Combining the above ingredients, we now deduce the desired mixing time lower bound for randomized μ -like shuffles.

Proof of (2.3) and (1.7). Recall the definition of $A_{\vec{n}}$ above Lemma 6.5. For (2.3), let $\nu \sim \mathbf{P}(\vec{n}, K)$ and ν_{Unif} be the uniform measure on S_N . We have that $\nu(A_{\vec{n}}) = 1 - o(1)$ by Lemma 6.3, and that $\nu_{\text{Unif}}(A_{\vec{n}}) = o(1)$ by Lemma 6.6. Thus

$$d_{\text{TV}}(\nu_{\text{Unif}}, \nu) \geq \nu(A_{\vec{n}}) - \nu_{\text{Unif}}(A_{\vec{n}}) = 1 - o(1).$$

For (1.7), let $\mathbf{X}$ be a randomized- μ -like shuffle process, $\nu \sim \mathbf{P}(\mathbf{X}, K)$ and ν_{Unif} be the uniform measure on S_N . Recall the definition of $A_{\mu, \chi, \rho, \varphi}$ in Lemma 6.5. We conclude by Lemmas 6.5 and 6.6 that

$$d_{\text{TV}}(\nu_{\text{Unif}}, \nu) \geq \nu(A_{\mu, \chi, \rho, \varphi}) - \nu_{\text{Unif}}(A_{\mu, \chi, \rho, \varphi}) = 1 - o(1). \quad \square$$

6.3 Proof of Lemma 6.4: construction of the cold spots

By Lemmas 6.5 and 6.6, the proof of Proposition 6.1 is now reduced to finding a deterministic set $H = H(\vec{n})$ that satisfies the conditions of Lemma 6.3. In order to define such a deterministic set H , we first define a set of “collision-likely” prefixes, and we will choose H to be the “expected positions” of strings with these prefixes. We will extend the definition of “cold spots” in [Sel22], since we need to tackle shuffle processes with χ -bad shuffles. Let

$$\alpha_{\text{tot}} \log N = \left\lfloor \frac{1 - \delta}{2 \sum_{0 \leq i \leq z} h_i I(\mathbf{p}^{(i)}, (\mathbf{p}^{(i)})^{\theta_\mu})} \log N \right\rfloor \quad \text{and} \quad \alpha_{\text{tot}}^{(i)} \log N = |\{t \leq \alpha_{\text{tot}} \log N : \mathbf{n}^{(t)}/N \in D_i\}|. \quad (6.4)$$

Recall the definition of T in (2.11) and definition of $\mathbf{p}^{(i)}$ in (4.28). For all $0 \leq i \leq z$, we choose positive integers $\alpha_l^{(i)} \log N$ for l such that $(i, l) \in T$ satisfying

$$\sum_{l: (i, l) \in T} \alpha_l^{(i)} = \alpha_{\text{tot}}^{(i)} \quad \text{and} \quad \left| \alpha_l^{(i)} \log(N) - \mathbf{p}_l^{(i)} \alpha_{\text{tot}}^{(i)} \log(N) \right| \leq 1.$$

We claim that $\alpha_{\text{tot}} \leq \frac{2}{3} C_\mu$. Indeed, by (2.19) we have

$$\alpha_{\text{tot}} + O(\delta) = \frac{1}{2 \sum_i h_i I(\mathbf{p}^{(i)}, (\mathbf{p}^{(i)})^{\theta_\mu})} < \frac{\theta_\mu}{2 \sum_i h_i \psi_{\mathbf{p}^{(i)}}(\theta_\mu)} \leq \frac{2\bar{C}_{\mathbf{p}}}{3},$$

where we used the fact that $\chi \ll 1$ and $\psi_\mu(\theta_\mu) = \sum_i h_i \psi_{\mathbf{p}^{(i)}}(\theta_\mu) + o_\chi(1)$. Therefore, we may let

$$\beta_{\text{tot}} \log(N) = K - \lfloor 2\rho \log N \rfloor - \alpha_{\text{tot}} \log(N) \geq \Omega(\log N),$$

$$\beta_{\text{tot}}^{(i)} \log N = |\{\alpha_{\text{tot}} \log N + \lfloor 2\rho \log N \rfloor < t \leq K : \mathbf{n}^{(t)}/N \in D_i\}|.$$

Similarly to $\mathbf{p}^{(i)}$, we define

$$\mathbf{q}^{(i)} = \left(\frac{(p_l^{(i)})^2}{\sum_{l: (i, l) \in T} (p_l^{(i)})^2} \right)_{l: (i, l) \in T}.$$

We choose positive integers $\beta_l^{(i)} \log(N)$ for all $(i, l) \in T$ with

$$\sum_{l=0}^{k-1} \sum_{0 \leq i \leq z} \beta_l^{(i)} = \beta_{\text{tot}} \quad \text{and} \quad \left| \beta_l^{(i)} \log(N) - \mathbf{q}_l^{(i)} \beta_{\text{tot}}^{(i)} \log(N) \right| \leq 1. \quad (6.5)$$

Note that $\alpha_{\text{tot}} \log N + \beta_{\text{tot}} \log N = K - \lfloor 2\rho \log N \rfloor$. Since the shuffle process is (χ, ρ, φ) -almost- μ -like, we have

$$\alpha_{\text{tot}}^{(i)} = h_i \alpha_{\text{tot}} \pm O(\rho) \pm O(\varphi), \quad \beta_{\text{tot}}^{(i)} = h_i \beta_{\text{tot}} \pm O(\rho) \pm O(\varphi).$$

We now consider G -edges coming from strings with $\alpha_l^{(i)} \log N$ digits l in $\mathbf{P}_i$ in the first $\alpha_{\text{tot}} \log N$ digits, and $\beta_l^{(i)} \log N$ digits l in $\mathbf{P}_i$ in the last $\beta_{\text{tot}} \log N$ digits for $i = 0, 1, \dots, k-1$. A straightforward calculation shows that

$$\sum_{(i, l) \in T} \alpha_l^{(i)} \log(p_l^{(i)}) \geq \frac{(1 - \delta)}{2 \sum_{0 \leq i \leq z} h_i I(\mathbf{p}^{(i)}, (\mathbf{p}^{(i)})^{\theta_\mu})} \cdot \sum_{(i, l) \in T} \frac{h_i (p_l^{(i)})^{\theta_\mu} \log(p_l^{(i)})}{\phi_{\mathbf{p}^{(i)}}(\theta_\mu)} \geq -\frac{1}{2} + \frac{\delta}{4}. \quad (6.6)$$

Here in the second inequality we used the fact that $\chi \ll \delta$ and

$$\sum_{l: (i, l) \in T} (p_l^{(i)})^{\theta_\mu} \log(p_l^{(i)}) = I(\mathbf{p}^{(i)}, (\mathbf{p}^{(i)})^{\theta_\mu}) \pm O(\chi).$$

Definition 6.7. The length $\alpha_{\text{tot}} \log(N)$ string $x \in [k]_0^M$ is a **collision-likely prefix** (we write $x \in \text{Pre}_{\text{CL}}$) if for all $(i, l) \in T$,

$$x \text{ contains } \alpha_l^{(i)} \log(N) \text{ of } l\text{-digits in positions } \{1 \leq t \leq \alpha_{\text{tot}} \log(N) : \mathbf{n}^{(t)}/N \in D_i\}.$$

Since $\vec{n}$ is (χ, ρ, φ) -almost- μ -like and $\alpha_{\text{tot}} \log N < K$, we can always find $g_1 < g_2$ in $[\alpha_{\text{tot}} \log N + 1, \alpha_{\text{tot}} \log N + 2\rho \log N]$ such that $\mathbf{n}^{(g_1)}, \mathbf{n}^{(g_2)}$ are χ -good. (Recall the definition of χ -good above (2.2).)

Definition 6.8. The string $s \in [k]_0^K$ is **collision-likely** (we write $s \in \text{CL}$) if s satisfies the following properties.

- The first $\alpha_{\text{tot}} \log(N)$ digits of s form a collision-likely prefix.
- $s[g_1] = \mathbf{I}_{\text{left}}^{(g_1)}$, $s[g_2] = \mathbf{I}_{\text{right}}^{(g_2)}$. (Recall (3.6).)
- $s[t] = l_{\text{max}}^{(t)}$ for all $t \in [\alpha_{\text{tot}} + 1, \dots, \alpha_{\text{tot}} + \lfloor 2\rho \log N \rfloor] \setminus \{g_1, g_2\}$.
- For all $(i, l) \in T$, we have

$$s \text{ contains } \beta_l^{(i)} \log(N) \text{ of } l\text{-digits in } \{\alpha_{\text{tot}} + \lfloor 2\rho \log N \rfloor + 1 \leq t \leq K : \mathbf{n}^{(t)}/N \in D_i\}.$$

We now give the definition of $H(\vec{n})$. Recalling (2.17), set

$$H(\vec{n}) \equiv \mathbb{Z} \cap \left(\bigcup_{x \in \text{Pre}_{\text{CL}}} NJ_x \right). \quad (6.7)$$

The set $H(\vec{n})$ now contains all expected locations of “collision-likely” prefixes for pile sizes $\vec{n}$.

Lemma 6.9. As $N \rightarrow \infty$, we have $|H(\vec{n})| \geq N^{1/2}$ and $|\partial H(\vec{n})| \leq 2|H(\vec{n})|^{1/2}$. More precisely,

$$|H(\vec{n})| = N^{1 + \sum_{(i,l) \in T} \alpha_l^{(i)} (\log(p_l^{(i)}) \pm O(\chi)) + \alpha_{\text{tot}} \sum_{i=0}^z h_i H(\{\alpha_l^{(i)}\}_{l:(i,l) \in T}) - o(\delta)} \geq N^{1/2}. \quad (6.8)$$

Proof. For each $x \in \text{Pre}_{\text{CL}}$, by (6.6), we have

$$\lambda_x = N^{\sum_{(i,l) \in T} \alpha_l^{(i)} (\log(p_l^{(i)}) \pm O(\chi))} = N^{\frac{-1+\delta}{2} \pm o(\delta)}, \quad (6.9)$$

where we have used $\sum_{(i,l) \in T} \alpha_l^{(i)} \leq \frac{K}{\log N}$ is bounded and $\chi \ll \delta$. Moreover, it is easy to see

$$\lfloor N\lambda_x \rfloor \leq \mathbb{Z} \cap NJ_x \leq \lceil N\lambda_x \rceil. \quad (6.10)$$

This implies that $|\mathbb{Z} \cap NJ_x| \geq N^{1/2+\delta/2}$ and thus $|H| \rightarrow \infty$ since $\text{Pre}_{\text{CL}} \neq \emptyset$. Moreover, since these intervals are mutually disjoint, we have $|\text{Pre}_{\text{CL}}| \leq N^{1/2-\delta/2}$, this means the number of connected components of H is smaller than the size of each component, hence $|\partial H| \leq 2|H|^{1/2}$. In addition, by Lemma 2.8 and the definition of $\alpha_l^{(i)}$, we have

$$\begin{aligned} |\text{Pre}_{\text{CL}}| &= \prod_{0 \leq i \leq z} \binom{\alpha_{\text{tot}}^{(i)} \log(N)}{\{\alpha_l^{(i)} \log N\}_{l:(i,l) \in T}} \\ &= N^{\sum_{i=0}^z \alpha_{\text{tot}}^{(i)} H(\{\alpha_l^{(i)}\}_{l:(i,l) \in T}) + o(1)} \\ &= N^{\alpha_{\text{tot}} \sum_{i=0}^z h_i H(\{\alpha_l^{(i)}\}_{l:(i,l) \in T}) \pm O(\chi)}. \end{aligned}$$

Here we used $\varphi, \rho \ll \chi \ll \delta$. Again, note that these intervals are mutually disjoint. Combining this with (6.9), (6.10) now gives the first equation of (6.8). $\square$

Lemma 6.10. With probability $1 - o(1)$, every $s_i \in \text{CL}$ obeys $i \in H$.

Proof. Consider the event

$$\Lambda = \left\{ \left| |\mathcal{I}(B_x)| - \mathbb{E}|\mathcal{I}(B_x)| \right| \leq N^{\frac{1}{2} + \frac{\delta}{20}}, \text{ for all } x \in [k]_0^M \right\}.$$

By Lemmas 2.6 and 2.7 and a union bound, we see $\mathbb{P}(\Lambda) = 1 - o(1)$. Note that

$$\iota(y) = \sum_{j=1}^K \sum_{0 \leq l < y[j]} |\mathcal{I}(B_{x[1:j-1]l})| \text{ and } \tau(y) = \sum_{j=1}^K \sum_{y[j] < l \leq k-1} |\mathcal{I}(B_{x[1:j-1]l})|.$$

On the event Λ we have that

$$|\iota(y) - Nt_y| \leq N^{\frac{1}{2} + \frac{\delta}{10}} \text{ and } |\tau(y) - N(t_y + \lambda_y)| \leq N^{\frac{1}{2} + \frac{\delta}{10}} \text{ for all } y \in [k]_0^M.$$

Now we prove that on the event Λ , $i \in H$ for all $i \in [N]$ such that $s_i \in \text{CL}$. For such i , we see that s_i starts with a prefix $x \in \text{Pre}_{\text{CL}}$. Recall the definition of g_1, g_2 above Definition 6.8. Let $u_1, u_2 \in [k]_0^{\alpha_{\text{tot}} \log N + \lfloor 2 \log N \rfloor}$ such that

$$\begin{aligned} u_j[t] &= s_i[t] \text{ for all } t \in [\alpha_{\text{tot}} \log N + \lfloor 2 \log N \rfloor] \setminus \{g_1, g_2\}, \\ u_1[g_1] &= \mathbf{I}_{\text{left}}^{(g_1)}, u_2[g_1] = \mathbf{I}_{\text{right}}^{(g_1)}, u_1[g_2] = \mathbf{I}_{\text{left}}^{(g_2)}, u_2[g_2] = \mathbf{I}_{\text{right}}^{(g_2)}. \end{aligned}$$

By Definition 6.8, we have $t_{s_i} \geq t_x + \lambda_{u_1}$ and $t_{s_i} + \lambda_{s_i} \leq t_x + \lambda_x - \lambda_{u_2}$. Therefore, we have that

$$i \geq \iota(s_i) \geq Nt_{s_i} - N^{\frac{1}{2} + \frac{\delta}{10}} \geq N(t_x + \lambda_{u_1}) - N^{\frac{1}{2} + \frac{\delta}{10}} \geq Nt_x. \quad (6.11)$$

Here we used that by (3.6) and (6.6),

$$N\lambda_{u_1} = N\lambda_x \cdot \prod_{t=\alpha_{\text{tot}} \log N}^{\alpha_{\text{tot}} \log N + \lfloor 2 \log N \rfloor} \frac{n_{u_1[t]}^{(t)}}{N} \geq N\lambda_x \cdot k^{-2\rho \log N} \cdot \chi^2 \geq N^{\frac{1}{2} + \frac{\delta}{10}}.$$

Similarly to (6.11), we have $i \leq N(t_x + \lambda_x)$, which indicates that $i \in H$. $\square$

Define

$$\gamma \equiv 2 + 2 \sum_{(i,l) \in T} (\alpha_l^{(i)} + \beta_l^{(i)}) \log(p_l^{(i)}) + \alpha_{\text{tot}} \sum_{i=0}^z h_i H(\{\alpha_l^{(i)}\}_{l:(i,l) \in T}) + \beta_{\text{tot}} \sum_{i=0}^z h_i H(\{\beta_l^{(i)}\}_{l:(i,l) \in T}).$$

Lemma 6.11. *We have*

$$\gamma \geq \frac{1}{2} \left(1 + \sum_{(i,l) \in T} \alpha_l^{(i)} \log(p_l^{(i)}) + \alpha_{\text{tot}} \sum_{i=0}^z h_i H(\{\alpha_l^{(i)}\}_{l:(i,l) \in T}) \right) + \Omega(\varepsilon). \quad (6.12)$$

Proof. It suffices to lower bound

$$\begin{aligned} & 3 + 3 \sum_{(i,l) \in T} \alpha_l^{(i)} \log(p_l^{(i)}) + 4 \sum_{(i,l) \in T} \beta_l^{(i)} \log(p_l^{(i)}) \\ & + \alpha_{\text{tot}} \sum_{i=0}^z h_i H(\{\alpha_l^{(i)}\}_{l:(i,l) \in T}) + 2\beta_{\text{tot}} \sum_{i=0}^z h_i H(\{\beta_l^{(i)}\}_{l:(i,l) \in T}). \end{aligned} \quad (6.13)$$

Recall the definition of $\beta_l^{(i)}$'s in (6.5). We write $\bar{\mathbf{p}}^{(i)} = (p_l^{(i)} / \sum_{l':(i,l') \in T} p_{l'}^{(i)})_{l:(i,l) \in T}$. We have

$$\begin{aligned} 2\beta_{\text{tot}} \sum_{i=0}^z h_i H(\{\beta_l^{(i)}\}_{l:(i,l) \in T}) &= 2\beta_{\text{tot}} \sum_{i=0}^z h_i \left(2I(\bar{\mathbf{p}}^{(i)}, \mathbf{q}^{(i)}) - \psi_{\bar{\mathbf{p}}^{(i)}}(2) \right) \\ &= -4\beta_{\text{tot}} \sum_{i=0}^z h_i \sum_{l=0}^{k-1} \frac{(p_l^{(i)})^2 \log(p_l^{(i)})}{\phi_{\bar{\mathbf{p}}^{(i)}}(2)} - 2\beta_{\text{tot}} \sum_{i=0}^z h_i \psi_{\bar{\mathbf{p}}^{(i)}}(2) \pm O(\chi) \\ &= -4 \sum_{(i,l) \in T} \beta_l^{(i)} \log(p_l^{(i)}) - 2\beta_{\text{tot}} \sum_{i=0}^z h_i \psi_{\bar{\mathbf{p}}^{(i)}}(2) \pm O(\chi) \\ &= -4 \sum_{(i,l) \in T} \beta_l^{(i)} \log(p_l^{(i)}) - 2\beta_{\text{tot}} \psi_{\mu}(2) \pm o_{\chi}(1). \end{aligned} \quad (6.14)$$

In addition, recall the definition of α_{tot} in (6.4), we have

$$\begin{aligned}
\alpha_{\text{tot}} \sum_{i=0}^z h_i H \left(\{\alpha_l^{(i)}\}_{l:(i,l) \in T} \right) &= \alpha_{\text{tot}} \sum_{i=0}^z h_i \left(\theta_\mu I(\bar{\mathbf{p}}^{(i)}, \mathbf{p}^{(i)}) - \psi_{\mathbf{p}^{(i)}}(\theta_\mu) \right) \pm O(\chi) \\
&= \frac{\theta_\mu}{2} - \alpha_{\text{tot}} \sum_{i=0}^z h_i \psi_{\mathbf{p}^{(i)}}(\theta_\mu) \pm O(\chi) \\
&= \frac{\theta_\mu}{2} - \alpha_{\text{tot}} \psi_\mu(\theta_\mu) + o_\chi(1) \pm O(\chi) \\
&= \frac{\theta_\mu}{2} - 2\alpha_{\text{tot}} \psi_\mu(2) \pm o_\chi(1).
\end{aligned} \tag{6.15}$$

Combining (6.6), (6.14) and (6.15), we see that (6.13) is lower-bounded by

$$3 + \frac{-3 + 3\delta}{2} + \frac{\theta_\mu}{2} - 2\alpha_{\text{tot}} \psi_\mu(2) - 2\beta_{\text{tot}} \psi_\mu(2) + O(\delta) = \frac{3 + \theta_\mu}{2} - \frac{2K}{\log N} \psi_\mu(2) + O(\delta) \geq \Omega(\varepsilon).$$

Here we used the fact that $\chi \ll \delta \ll \varepsilon$ and $K = \lfloor (C_\mu - \varepsilon) \log N \rfloor$. $\square$

Lemma 6.12. *For some $\delta = \delta(\mu, \varepsilon)$ the following holds: with probability $1 - o(1)$,*

$$|E(G(S_K)) \cap E(H)| \geq |H|^{1/2+\delta}.$$

Proof. For $s \in \text{CL}$, let A_s be the event that s occurs at least two times in S_K . By inclusion-exclusion,

$$\begin{aligned}
\mathbb{P}(A_s) &= \mathbb{P}\left(\bigcup_{i < j} \{\bar{s}_i = \bar{s}_j = s\}\right) \\
&\geq \sum_{i < j} \mathbb{P}(\bar{s}_i = \bar{s}_j = s) - \sum_{i < j, k < l, (i,j) \neq (k,l)} \mathbb{P}(\bar{s}_i = \bar{s}_j = \bar{s}_k = \bar{s}_l = s) \\
&\geq \sum_{i < j} \mathbb{P}(\bar{s}_i = \bar{s}_j = s) - 3 \sum_{i < j < k} \mathbb{P}(\bar{s}_i = \bar{s}_j = \bar{s}_k = s) - 3 \sum_{i < j < k < l} \mathbb{P}(\bar{s}_i = \bar{s}_j = \bar{s}_k = \bar{s}_l = s) \\
&= \binom{N}{2} \mathbb{P}(\bar{s}_i = \bar{s}_j = s) - 3 \binom{N}{3} \mathbb{P}(\bar{s}_i = \bar{s}_j = \bar{s}_k = s) - 3 \binom{N}{4} \mathbb{P}(\bar{s}_i = \bar{s}_j = \bar{s}_k = \bar{s}_l = s).
\end{aligned} \tag{6.16}$$

Direct calculation shows that for $1 \leq i < j \leq N$,

$$\mathbb{P}(\bar{s}_i = \bar{s}_j = s) = \prod_{t \leq K} \mathbb{P}(\bar{s}_i[t] = \bar{s}_j[t] = s[t]) = \prod_{t \leq K} \frac{n_{s[t]}^{(t)} (n_{s[t]}^{(t)} - 1)}{N(N-1)}.$$

Notice that for any $t \in \mathbf{P}_i$,

$$\frac{n_{s[t]}^{(t)} (n_{s[t]}^{(t)} - 1)}{N(N-1)} = \left(\frac{n_{s[t]}^{(t)}}{N} + O(1/N) \right)^2 = (p_{s[t]}^{(i)} + O(\chi))^2,$$

so we can immediately deduce from (6.17) that

$$\mathbb{P}(\bar{s}_i = \bar{s}_j = s) = \prod_{(i,l) \in T} (p_l^{(i)} + O(\chi))^{2(\alpha_l^{(i)} + \beta_l^{(i)}) \log N} = N^{2 \sum_{(i,l) \in T} (\alpha_l^{(i)} + \beta_l^{(i)}) \log p_l^{(i)} \pm O(\chi)}.$$

Similar calculation shows that for $i < j < k < l$,

$$\mathbb{P}(\bar{s}_i = \bar{s}_j = \bar{s}_k = s) \leq N^{3 \sum_{(i,l) \in T} (\alpha_l^{(i)} + \beta_l^{(i)}) \log p_l^{(i)} \pm O(\chi)},$$

$$\mathbb{P}(\bar{s}_i = \bar{s}_j = \bar{s}_k = \bar{s}_l = s) \leq N^{4 \sum_{(i,l) \in T} (\alpha_l^{(i)} + \beta_l^{(i)}) \log p_l^{(i)} \pm O(\chi)}.$$

Combining (6.18), (6.19) and (6.20), we have

$$\binom{N}{3} \mathbb{P}(\bar{s}_i = \bar{s}_j = \bar{s}_k = s) \leq N^{3+3 \sum_{(i,l) \in T} (\alpha_l^{(i)} + \beta_l^{(i)}) \log p_l^{(i)} + O(\chi)} = O(N^{-\delta/10}) \binom{N}{2} \mathbb{P}(\bar{s}_i = \bar{s}_j = s),$$

$$\binom{N}{4} \mathbb{P}(\bar{s}_i = \bar{s}_j = \bar{s}_k = \bar{s}_l = s) \leq N^{4+4 \sum_{(i,l) \in T} (\alpha_l^{(i)} + \beta_l^{(i)}) \log p_l^{(i)} + O(\chi)} = O(N^{-\delta/10}) \binom{N}{2} \mathbb{P}(\bar{s}_i = \bar{s}_j = s), \quad (6.22)$$

where we used the fact that

$$1 + \sum_{(i,l) \in T} (\alpha_l^{(i)} + \beta_l^{(i)}) \log p_l^{(i)} \leq 1 + \frac{K}{\log N} \cdot \sum_i h_i \log p_{\max}^{(i)} \leq -\delta$$

since $\chi \ll \delta$. Thus, combining (6.16), (6.21) and (6.22), we have

$$\mathbb{P}(A_s) = \binom{N}{2} \prod_{t \leq K} \left(\frac{n_{s[t]}^{(t)}}{N} + O(1/N) \right)^2 (1 - O(N^{-\delta/10})) = N^{2+2 \sum_{(i,l) \in T} (\alpha_l^{(i)} + \beta_l^{(i)}) \log p_l^{(i)} \pm O(\chi)}. \quad (6.23)$$

With $Z = \sum_{s \in \text{CL}} 1_{A_s}$, we have $|E(G(S_K)) \cap E(H)| \geq Z$. The above calculation and Lemma 2.8 yield

$$\mathbb{E}Z \geq N^{\gamma - O(\chi)}. \quad (6.24)$$

To calculate the second moment, first notice that for $s \neq s' \in \text{CL}$, we have

$$\begin{aligned} \mathbb{P}(A_s \cap A_{s'}) &\leq \sum_{i < j, \& k < l} \mathbb{P}(\bar{s}_i = \bar{s}_j = s, \bar{s}_k = \bar{s}_l = s') \\ &= \binom{N}{4} \binom{4}{2} \prod_{t \leq K} \mathbb{P}(\bar{s}_i[t] = \bar{s}_j[t] = s[t], \bar{s}_k[t] = \bar{s}_l[t] = s'[t]), \\ &= \binom{N}{4} \binom{4}{2} \prod_{t \leq K} \left(\frac{n_{s[t]}^{(t)}}{N} + O(1/N) \right)^2 \left(\frac{n_{s'[t]}^{(t)}}{N} + O(1/N) \right)^2, \end{aligned}$$

where we used the fact that

$$\mathbb{P}(\bar{s}_i[t] = \bar{s}_j[t] = s[t], \bar{s}_k[t] = \bar{s}_l[t] = s'[t]) = \begin{cases} \frac{n_{s[t]}^{(t)}(n_{s[t]}^{(t)} - 1)(n_{s[t]}^{(t)} - 2)(n_{s[t]}^{(t)} - 3)}{N(N-1)(N-2)(N-3)}, & \text{if } s[t] = s'[t], \\ \frac{n_{s[t]}^{(t)}(n_{s[t]}^{(t)} - 1)n_{s'[t]}^{(t)}(n_{s'[t]}^{(t)} - 1)}{N(N-1)(N-2)(N-3)}, & \text{if } s[t] \neq s'[t]. \end{cases}$$

Then, by the first equality in (6.23) we have that

$$\mathbb{P}(A_s) \mathbb{P}(A_{s'}) = (1 - O(N^{-\delta/10})) \binom{N}{2}^2 \prod_{t \leq K} \left[\left(\frac{n_{s[t]}^{(t)}}{N} + O(1/N) \right)^2 \left(\frac{n_{s'[t]}^{(t)}}{N} + O(1/N) \right)^2 \right].$$

Thus we obtain that

$$\mathbb{P}(A_s \cap A_{s'}) - \mathbb{P}(A_s) \mathbb{P}(A_{s'}) \leq O(N^{-\delta/10}) \cdot N^4 \prod_{t \leq K} \left[\left(\frac{n_{s[t]}^{(t)}}{N} \right)^2 \left(\frac{n_{s'[t]}^{(t)}}{N} \right)^2 \right] \leq O(N^{-\delta/10}) \mathbb{P}(A_s) \mathbb{P}(A_{s'}).$$

Thus we have

$$\text{Var}(Z) = \sum_{s, s' \in \text{CL}} [\mathbb{P}(A_s \cap A_{s'}) - \mathbb{P}(A_s) \mathbb{P}(A_{s'})] \leq \sum_{s, s' \in \text{CL}} O(N^{-\delta/10}) \mathbb{P}(A_s) \mathbb{P}(A_{s'}) = O(N^{-\delta/10}) (\mathbb{E}Z)^2. \quad (6.25)$$

Combining (6.24), (6.25) and $\chi, \eta \ll \delta \ll \varepsilon$, we obtain by the Markov inequality that

$$\mathbb{P}(|E(G(S_K)) \cap E(H)| \geq N^{\gamma - \delta}) \geq \mathbb{P}(Z \geq \frac{1}{2} \mathbb{E}Z) \geq 1 - \frac{4 \text{Var}(Z)}{(\mathbb{E}Z)^2} = 1 - o(1).$$

Then by (6.12), we know that $N^{\gamma - \delta} \geq |H|^{1/2 + \delta}$ (by taking δ sufficiently small depending only on ε), and thus we complete the proof. $\square$

A Non-Convexity of $\mu \mapsto \overline{C}_\mu$

Recall the definition of ψ_μ in and below (1.4). It is easy to see that $\psi_\mu(1) = 0$, and that $\psi_\mu(x)$ is concave and $\psi_\mu(x)/x$ is increasing (by the Hölder and power mean inequalities respectively). In fact these essentially characterize all possible functions ψ_μ up to rescaling. Moreover the smaller class $\cup_{k \geq 1} \{\psi_{\mathbf{p}}\}_{\mathbf{p} \in \mathcal{D}_k}$ suffices to approximate all such functions (this follows from Proposition A.1 below and convexity of Ψ).

Let Ψ be the set of functions $f : [1, 4] \rightarrow \mathbb{R}_+$ which are concave and satisfy $f(1) = 0$ and $f(2) > 0$ and have $f(x)/x$ increasing. We take θ_f to satisfy $f(\theta_f) = 2f(2)$; it follows easily from the definitions that $\theta_f \in [3, 4]$ for all $f \in \Psi$. We let

$$C_f = \frac{3 + \theta_f}{4f(2)}, \quad \tilde{C}_f = \frac{1}{f'(4)}, \quad \overline{C}_f = \max(C_f, \tilde{C}_f).$$

(Here $f'(4)$ is always well-defined as a derivative from the left since f is concave on domain $[1, 4]$.)

Proposition A.1. *Then for any $f \in \Psi$ and $\varepsilon > 0$ there exists $K_*(f, \varepsilon)$ and a sequence of probability distributions $\mathbf{p}^{(K)} \in \mathcal{D}_{k(K)}$ for $K \geq K_*$ such that $\psi_{\mathbf{p}^{(K)}}(x) \log K$ approximates f in the sense that:*

- (I) $\left| f(x) - \frac{\psi_{\mathbf{p}^{(K)}}(x)}{\log K} \right| \leq \varepsilon$ for all $x \in [2, 4]$.
- (II) $\left| \theta_f - \frac{\theta_{\mathbf{p}^{(K)}}}{\log K} \right| \leq \varepsilon$.
- (III) $\left| C_f - \frac{C_{\mathbf{p}^{(K)}}}{\log K} \right| \leq \varepsilon$.
- (IV) $\left| \tilde{C}_f - \frac{\tilde{C}_{\mathbf{p}^{(K)}}}{\log K} \right| \leq \varepsilon$.

(The last two points also then imply $\left| \overline{C}_f - \frac{\overline{C}_{\mathbf{p}^{(K)}}}{\log K} \right| \leq \varepsilon$.)

Proof. Without loss of generality we can slightly perturb f to be strictly concave (without changing any of the relevant quantities by more than $\varepsilon/2$). Such f can be approximated as a minimum of affine functions

$$f_\delta(x) = \min_{j \in [J]} [a_j x - b_j]$$

with $|f(x) - f_\delta(x)| \leq \delta$ on $x \in [2, 4]$. Explicitly, we can take $a_j x - b_j = (x - x_j)f'(x_j) + f(x_j)$ to be a tangent line approximation to f at x_j for a large discrete set of points $2 = x_1 \leq x_2 \leq \dots \leq x_J = 4$ with $|x_{j+1} - x_j| \leq \delta'$ with δ' small depending on f and δ . (Any choice of supergradient can be used for $f'(x_j)$ if x_j is not a point of differentiability). The condition that $f(x)/x$ is increasing implies that $a_j, b_j \geq 0$ for each j . Since we assumed f is strictly concave we also have $a_j > b_j$ for each j , i.e. each tangent line is strictly above the point $(1, 0) = (1, f(1))$.

We choose δ small depending on (f, ε) , and then K_* large depending on $(f, \varepsilon, \delta, \{x_1, \dots, x_J\})$. (Note that the notation of this appendix is unrelated to (2.9).) We take $\mathbf{p}$ to include $\lceil K^{b_j} \rceil$ coordinates with value K^{-a_j} , for each $1 \leq j \leq J$. Since $a_j > b_j$ these coordinates have total mass at most $2 \sum_{j=1}^J K^{b_j - a_j} < 1$ for large K_* . To make $\mathbf{p}$ a probability distribution, the remaining mass of $\mathbf{p}$ is placed on an extremely large number (say $K^{10 \cdot \max_j (|a_j| + |b_j|)}$) of coordinates with equal mass, which contribute negligibly to the value of $\phi_{\mathbf{p}}(x)$ for $x \in [1, 4]$. It is easy to see that with this choice, one has

$$\phi_{\mathbf{p}}(x) \in [K^{-f_\delta(x)}, (J+1)K^{-f_\delta(x)}], \quad \forall x \in [2, 4].$$

Since $\psi_{\mathbf{p}}(x) = -\log \phi_{\mathbf{p}}(x)$ we have

$$\psi_{\mathbf{p}}(x) \in [f_\delta(x) \log(K) - \log(J+1), f_\delta(x) \log(K)], \quad \forall x \in [2, 4].$$

This yields part (I) as K_* is large depending on J . Part (II) is immediate since f is strictly increasing, so θ_f is stable to small perturbations. Part (III) follows by combining the previous two (and adjusting ε). Finally part (IV) follows from concavity of f , which ensures that $f'(4)$ is the minimal value of $a_j = f'(x_j)$ that appears. $\square$

By taking explicit choices $f, \check{f} \in \Psi$, we now provide the counterexample claimed in Proposition 1.4.

Proof of Proposition 1.4. Given Proposition A.1, it suffices to find $f, \check{f} \in \Psi$ so that $\hat{f}(x) \equiv \frac{f(x) + \check{f}(x)}{2}$ satisfies

$$\overline{C}_{\hat{f}} > \max(\overline{C}_f, \overline{C}_{\check{f}}).$$

Indeed, then for K large the approximations $(\mathbf{p}^{(K)}, \check{\mathbf{p}}^{(K)})$ as in Proposition A.1 will furnish the desired counterexample.⁵ In fact we will have $\overline{C}_f = \overline{C}_{\check{f}}$ below, and parametrize the construction by a suitably small absolute constant η . We want $\theta_{\hat{f}}$ to be large given the values $\theta_f < \theta_{\check{f}}$ (e.g. larger than their average), and the idea below is simply to make f grow fast until θ_f and then grow slowly afterward.

We take f to be linear on each of the intervals $[1, 2]$ and $[2, 3.5 - \eta]$ and $[3.5 - \eta, 4]$, with values:

$$f(1) = 0, \quad f(2) = 6.5 - \eta, \quad f(3.5 - \eta) = 13 - 2\eta$$

and satisfy $f'(x) = 4$ on $x \in (3.5 - \eta, 4]$. Then f is concave because the derivatives on each subinterval are $6.5 - \eta > (6.5 - \eta) \cdot (2/3) > 4$, and also $f(x)/x$ is increasing (here it suffices to check the endpoints).

Meanwhile we take $\check{f}$ to be linear on $[1, 2]$ and $[2, 4]$ with values:

$$\check{f}(1) = 1, \quad \check{f}(2) = 6.5 + \eta, \quad \check{f}(3.5 + \eta) = 13 + 2\eta.$$

One can similarly check that $\check{f} \in \Psi$. It is immediate then that

$$\begin{aligned} 3 + \theta_f = 6.5 - \eta = f(2) &\implies C_f = 1/4; \\ 3 + \theta_{\check{f}} = 6.5 + \eta = \check{f}(2) &\implies C_{\check{f}} = 1/4. \end{aligned}$$

Meanwhile $f'(4) = 4$ so $\tilde{C}_f = 1/4$, and $\check{f}'(4) = \frac{6.5 + \eta}{1.5 + \eta} > 4$ so $\tilde{C}_{\check{f}} < 1/4$. Thus $\overline{C}_f = \overline{C}_{\check{f}} = 1/4$.

Finally for $\hat{f}$, we have $\hat{f}(2) = 6.5$. So to show $C_{\hat{f}} > 1/4$ it remains to verify that $\theta_{\hat{f}} > 3.5$, or equivalently that $\hat{f}(3.5) < 2\hat{f}(2) = 13$. Using piece-wise linearity it is easy to calculate that

$$\begin{aligned} f(3.5) &= (13 - 2\eta) + 4\eta = 13 + 2\eta, & \check{f}(3.5) &= 13 - \frac{7\eta}{3} \pm O(\eta^2) \\ \implies \hat{f}(3.5) &= 13 - \frac{\eta}{6} \pm O(\eta^2). \end{aligned}$$

Thus $\hat{f}(3.5) < 13$ indeed holds for a suitably small constant η , completing the construction. $\square$

References

- [Ald83] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Séminaire de Probabilités XVII 1981/82*, pages 243–297. Springer, 1983.
- [BD92] D. Bayer and P. Diaconis. Trailing the Dovetail Shuffle to its Lair. *Ann. Appl. Probab.*, 2(2):294–313, 1992.
- [CS04] I. Csiszár and P. Shields. *Information Theory and Statistics: A Tutorial*. Foundations and trends in communications and information theory. Now Publishers, 2004.
- [DF23] P. Diaconis and J. Fulman. *The mathematics of shuffling cards*. American Mathematical Society, 2023.
- [DFP92] P. Diaconis, J. A. Fill, and J. Pitman. Analysis of top to random shuffles. *Combinatorics, Probability & Computing*, 1:135–155, 1992.
- [DG11] P. Diaconis and R. Graham. *Magical mathematics: the mathematical ideas that animate great magic tricks*. Princeton University Press, 2011.
- [Dia03] P. Diaconis. Mathematical developments from the analysis of riffle shuffling. *Groups, combinatorics & geometry (Durham, 2001)*, pages 73–97, 2003.
- [Gil55] E. Gilbert. Theory of Shuffling. *Technical Memorandum*, 1955.
- [HS05] D. Hush and C. Scovel. Concentration of the hypergeometric distribution. *Statistics and Probability Letters*, 75(2):127–132, 2005.

⁵As stated the number of piles in $(\mathbf{p}^{(K)}, \check{\mathbf{p}}^{(K)})$ as constructed above will not be equal to each other, which is technically required in Proposition 1.4. However this is easily fixed: recall that both $(\mathbf{p}^{(K)}, \check{\mathbf{p}}^{(K)})$ include a very large number of equal-mass tiny coordinates which contribute negligibly to ψ . By increasing the number of such tiny coordinates in either $\mathbf{p}^{(K)}$ and $\check{\mathbf{p}}^{(K)}$ as needed, we can ensure they have an equal number of piles while retaining the conclusions of Proposition A.1.

- [JM15] J. Jonasson and B. Morris. Rapid mixing of dealer shuffles and clumpy shuffles. *Electronic Communications in Probability*, 20:1–11, 2015.
- [Lal00] S. P. Lalley. On the rate of mixing for p -shuffles. *Ann. Appl. Probab.*, 10(4):1302 – 1321, 2000.
- [LS17] C. W. Lee and F. Santos. Subdivisions and triangulations of polytopes. In *Handbook of discrete and computational geometry*, pages 415–447. Chapman and Hall/CRC, 2017.
- [Mor08] B. Morris. The Mixing Time of the Thorp Shuffle. *SIAM Journal on Computing*, 38(2):484–504, 2008.
- [Mor09] B. Morris. Improved mixing time bounds for the Thorp shuffle and L-reversal chain. *Ann. Probab.*, 37(2):453–477, 2009.
- [Mor13] B. Morris. Improved mixing time bounds for the Thorp shuffle. *Combinatorics, Probability and Computing*, 22(1):118–132, 2013.
- [Pit97] J. Pitman. Probabilistic bounds on the coefficients of polynomials with only real zeros. *Journal of Combinatorial Theory, Series A*, 77(2):279–303, 1997.
- [Por75] N. Portenko. Diffusion processes with unbounded drift coefficient. *Theory of Probability & Its Applications*, 20(1):27–37, 1975.
- [Ree81] J. Reeds. Unpublished Manuscript, 1981.
- [Sel22] M. Sellke. Cutoff for the asymmetric riffle shuffle. *Ann. Probab.*, 50(6):2244 – 2287, 2022.
- [Tho73] E. O. Thorp. Nonrandom shuffling with applications to the game of Faro. *Journal of the American Statistical Association*, 68(344):842–847, 1973.
- [Tod14] I. Todhunter. *A history of the mathematical theory of probability*. Cambridge University Press, 2014.