Annals of Mathematics **188** (2018), 49–143
https://doi.org/10.4007/annals.2018.188.1.2

The sharp threshold for making squares

By Paul Balister, Béla Bollobás, and Robert Morris

Abstract

Consider a random sequence of N integers, each chosen uniformly and independently from the set $\{1, \ldots, x\}$. Motivated by applications to factorization algorithms such as Dixon's algorithm, the quadratic sieve, and the number field sieve, Pomerance in 1994 posed the following problem: how large should N be so that, with high probability, this sequence contains a subsequence, the product of whose elements is a perfect square? Pomerance determined asymptotically the logarithm of the threshold for this event and conjectured that it in fact exhibits a *sharp threshold* in N. More recently, Croot, Granville, Pemantle and Tetali determined the threshold up to a factor of $4/\pi + o(1)$ as $x \to \infty$ and made a conjecture regarding the location of the sharp threshold.

In this paper we prove both of these conjectures by determining the sharp threshold for making squares. Our proof combines techniques from combinatorics, probability and analytic number theory; in particular, we use the so-called method of self-correcting martingales in order to control the size of the 2-core of the random hypergraph that encodes the prime factors of our random numbers. Our method also gives a new (and completely different) proof of the upper bound in the main theorem of Croot, Granville, Pemantle and Tetali.

Contents

1. Introduction 50
2. An outline of the proof 55
3. Number-theoretic facts 61
4. Probabilistic facts and preliminary results 71
5. Approximation by an independent hypergraph model 79

Keywords: integer factorization, perfect square, random graph process

AMS Classification: Primary: 11Y05; Secondary: 60C05.

The first two authors were partially supported by NSF grant no. DMS-1600742. The second author was also supported by MULTIPLEX grant no. 317532. The third author was partially supported by CNPq (Proc. 303275/2013-8), by FAPERJ (Proc. 201.598/2014), and by ERC Starting Grant 680275 MALIG.

© 2018 Department of Mathematics, Princeton University.

6. The exploration process 94
7. Tracking the process when most columns are empty 113
8. Tracking the process in the critical range 122
9. The proof of Theorems 2.2 and 2.6 131
10. The proof of Theorem 1.1 137
References 140

1. Introduction

Many of the fastest known algorithms for factoring large integers rely on finding subsequences of randomly generated sequences of integers whose product is a perfect square. Examples include Dixon's algorithm [15], the quadratic sieve [28], and the number field sieve (see, e.g., [25]); an excellent elementary introduction to the area is given by Pomerance [31]. In each of these algorithms one generates a sequence of congruences of the form

$$a_i \equiv b_i^2 \pmod{n}, \qquad i = 1, 2, \ldots$$

and then one aims to find subsets of the a_i whose product is a perfect square, say $\prod_{i\in I} a_i = X^2$, so then one has $X^2 \equiv Y^2 \pmod{n}$ with $Y = \prod_{i\in I} b_i$. If one is lucky, then $X \not\equiv \pm Y \pmod{n}$, in which case one can generate non-trivial factors of n as $\gcd(X \pm Y, n)$.

A useful heuristic, suggested by Schroeppel in the 1970s (see [31]), is to imagine that the numbers a_i are chosen independently and uniformly at random from the set $\{1, \ldots, x\}$ for some suitably chosen integer x. Motivated by this idea, Pomerance [29] posed in 1994 the problem of determining the *threshold* for the event that such a collection of random numbers contains a subset whose product is a square. To be precise, given $x \in \mathbb{N}$, let us define a probability space $\Omega(x)$ by choosing $a_1, a_2, \ldots$ independently and uniformly at random from $\{1, \ldots, x\}$, and a random variable $T(x)$ by setting

$$T(x) := \min\Big\{N \in \mathbb{N} : \prod_{i\in I} a_i \text{ is a perfect square for some } I \subseteq \big\{1, \ldots, N\big\},\ I \neq \emptyset\Big\}.$$

Pomerance [30] proved that for all $\varepsilon > 0$,

$$\exp\Big(\big(1-\varepsilon\big)\sqrt{2\log x \log\log x}\Big) \leqslant T(x) \leqslant \exp\Big(\big(1+\varepsilon\big)\sqrt{2\log x \log\log x}\Big) \tag{1}$$

with high probability,[1] and he conjectured that $T(x)$ in fact exhibits a *sharp threshold*, i.e., that there exists a function $f(x)$ such that $(1-\varepsilon)f(x) \leqslant T(x) \leqslant$

[1] We use the term *with high probability* to mean with probability tending to 1 as $x \to \infty$.

$(1+\varepsilon)f(x)$ with high probability for all $\varepsilon > 0$. Croot, Granville, Pemantle and Tetali [13] significantly improved these bounds (see (3), below) and stated a conjecture as to the location of the threshold, i.e., the value of the function $f(x)$. In this paper we shall prove these two conjectures.

In order to state the theorem and conjecture of Croot, Granville, Pemantle and Tetali, we will need to recall some standard notation. Let $\pi(y)$ denote the number of primes less than or equal to y, let $\Psi(x,y)$ denote the number of *y-smooth* integers in $\{1,\dots,x\}$, that is, the number of integers with no prime factor strictly greater than y, and define

$$J(x) = \min_{2\leqslant y\leqslant x} \frac{\pi(y)x}{\Psi(x,y)}. \tag{2}$$

It can be shown (see Section 3) that the minimum in (2) occurs at

$$y_0 = y_0(x) = \exp\Big(\big(1+o(1)\big)\sqrt{\tfrac{1}{2}\log x \log\log x}\Big)$$

and that

$$J(x) = y_0^{2+o(1)} = \exp\Big(\big(1+o(1)\big)\sqrt{2\log x\log\log x}\Big).$$

An asymptotic formula for $J(x)$ was obtained by McNew [26]. We remark that a relatively straightforward argument due to Schroeppel (see [30]) shows that, for all $\varepsilon > 0$,

$$T(x) \leqslant \big(1+\varepsilon\big)J(x)$$

with high probability, which implies the upper bound in (1). Indeed, if $N \geqslant (1+\varepsilon)J(x)$, then with high probability at least $\pi(y_0)+1$ of the numbers $a_1,\dots,a_N$ will be y_0-smooth, since each a_i is y_0-smooth with probability $\Psi(x,y_0)/x = \pi(y_0)/J(x)$. Now, by simple linear algebra, it follows that the vectors encoding the primes that divide a_i an odd number of times are linearly dependent over $\mathbb{F}_2$, and hence there exists a subset whose product is a square, as required.

Pomerance's conjecture remained wide open for over ten years, until a fundamental breakthrough was obtained by Croot, Granville, Pemantle and Tetali [13], who used a combination of techniques from number theory, probability theory and combinatorics to dramatically improve both the upper bound of Schroeppel and the lower bound of Pomerance [30], determining the location of the threshold to within a factor of $4/\pi$. To be precise, they proved that

$$\frac{\pi}{4}\big(e^{-\gamma}-\varepsilon\big)J(x) \leqslant T(x) \leqslant \big(e^{-\gamma}+\varepsilon\big)J(x) \tag{3}$$

with high probability, where $\gamma \approx 0.5772$ is the Euler–Mascheroni constant. Recall that $e^{-\gamma}$ is (amongst other things) the limit as $y\to\infty$ of the ratio of the density of integers with no prime divisor smaller than y, to the proportion of elements of $\{1,\dots,y\}$ that are prime.

Croot, Granville, Pemantle and Tetali [13] conjectured that the upper bound in (3) is sharp. Our main theorem confirms their conjecture.

THEOREM 1.1. *For all $\varepsilon > 0$, we have with high probability*

$$\big(e^{-\gamma} - \varepsilon\big)J(x) \leqslant T(x) \leqslant \big(e^{-\gamma} + \varepsilon\big)J(x).$$

As a simple corollary, we also deduce the following asymptotic expression for the expected value of $T(x)$.

COROLLARY 1.2. $\mathbb{E}\big[T(x)\big] = \big(e^{-\gamma} + o(1)\big)J(x)$ *as* $x \to \infty$.

Since the upper bound in Theorem 1.1 was proved in [13], we are only required to prove the lower bound. However, we will also obtain a new proof of the upper bound, quite different from that given in [13], as a simple consequence of our method; see Section 10. We would like to thank Jonathan Lee for pointing out to us a particularly simple and natural way of deducing this from our proof.

Another significant advantage of our proof, which is outlined in Section 2, is that it gives detailed structural information about the typical properties of the set of numbers that are left over after sieving and "singleton removal" (see, e.g., [22]). We plan to study this structure in a more general setting, and in greater detail, in a follow-up paper together with Lee [2].

Croot, Granville, Pemantle and Tetali [13] proved their lower bound in (3) via the first moment method, by counting the expected number of non-empty subsets $I \subseteq \{1, \ldots, N\}$ such that $\prod_{i \in I} a_i$ is a square. Unfortunately, there exists a constant $c > 0$ such that this expected number blows up when $N \geqslant (e^{-\gamma} - c)J(x)$, which implies that a sharp lower bound cannot be obtained by this method. (See the comments after the proof of Theorem 1.1 in Section 10.)

Instead, we shall use the method of self-correcting martingales[2] to follow a random process that removes numbers from the set[3] $\{a_1, \ldots, a_N\}$ as soon as we can guarantee that they are not contained in a subset whose product is a square. This is in one sense very simple: a number a_i can be discarded if there exists a prime for which a_i is the only remaining number that it divides an odd number of times. However, this apparent simplicity is deceiving, and

[2]This technique is based on the so-called "differential equations method" (see, e.g., [23], [35]), and it involves the use of martingales to control a collection of interacting random variables that exhibit "self-correction" in a certain natural sense (see Sections 7 and 8). It was introduced in [4], [7], [33] and has more recently been further developed in [5], [6], [16]; our approach is, in particular, based on that used in [16].

[3]This is, strictly speaking, a multi-set, since the numbers a_i are chosen independently with replacement. However, since we are very unlikely to choose the same number twice (indeed, if we do so we immediately have a square), we shall ignore this possibility for the sake of this discussion.

the technical challenges involved in tracking the process are substantial. For example, we will need to reveal the random numbers $\{a_1, \ldots, a_N\}$ gradually (roughly speaking, prime by prime, in decreasing order), and the amount of information we are allowed to reveal at each step is rather delicate. Moreover, the removal of a number can trigger an avalanche, causing many other numbers to be removed in the same step. Fortunately, however, self-correction (which is partly a result of these avalanches) will allow us to show that the process remains subcritical (in a certain natural sense), which will in turn allow us to control the upper tail of the size of the avalanches; see Section 6. In order to do so, we will need good control over the dependence between the prime factors of the numbers $\{a_1, \ldots, a_N\}$ conditioned on the information we have observed so far. This is achieved in Section 5, where we obtain strong bounds on the ratio between the (conditional) probability of certain "basic" events, and the corresponding probabilities in a simpler independent model. These bounds require some number-theoretic estimates (stated in Section 3), most of which follow easily from the fundamental work of Hildebrand and Tenenbaum [18] on smooth numbers.

Using the method described above, we shall be able to show that with high probability the number of "active" numbers (i.e., elements of $\{a_1, \ldots, a_N\}$ that we have not yet discarded) tracks a deterministic function (see Theorem 2.2, below) until there are very few numbers remaining (roughly $e^{-C\sqrt{\log y_0}} y_0$ for some large constant C), at which point we can apply the first moment calculation from [13]. The heuristic reason for the appearance of the formula in Theorem 2.2 is that the number of y-smooth numbers is concentrated (e.g., by Chernoff's inequality) for all reasonably large values of y, since the a_i are chosen independently, and is equal to the number of isolated vertices in a certain natural (random) hypergraph (see Definition 2.3). We will control the average degree of this hypergraph (see Theorem 2.6) and show (using Theorem 5.1) that its edges are chosen *almost* independently so, in particular, its degree distribution is close to Poisson. Equating these two estimates for the number of isolated vertices gives (7). The Euler–Mascheroni constant γ appears in our proof at this point, since when we reveal the zth smallest prime, the (typical) average degree of the hypergraph is close to $\mathrm{Ein}(m(z)/z)$, where $m(z)$ is the number of active numbers at this point, and Ein is the exponential integral

$$\mathrm{Ein}(w) := \int_0^w \frac{1 - e^{-t}}{t}\, dt.$$

In order to prove the upper bound in Theorem 1.1, we observe (Lemma 10.2) that the ratio of the number of active numbers and active primes (that is, primes that could still appear in some square) approaches 1 when $z = \pi(y_0)$ and $N/J(x)$ approaches $e^{-\gamma}$. Thus, by adding just a few extra y_0-smooth

numbers, we can apply the linear algebra approach of Schroeppel to obtain a subset whose product is a square, as required.

The rest of the paper is organized as follows. In Section 2 we give a detailed outline of the proof, state our main auxiliary results, and define precisely the random process mentioned above. In Section 3 we deduce the number-theoretic estimates we need from known results on smooth numbers, and in Section 4 we recall some basic results from probability theory, define some useful events, and use the results of Section 3 to prove various useful properties of these events. In Section 5 we shall again apply the results of Section 3, this time to control the dependence between the prime divisors of our random numbers in the probability space obtained by conditioning on the information revealed in the random process so far, and then in Section 6 we apply the main theorem of Section 5 to control the size of avalanches in the process. In Section 7 we use these results and the method of self-correcting martingales to control the process for large primes, and in Section 8 we do the same in the critical range $z = e^{O(\sqrt{\log y_0})} y_0$. Finally, in Sections 9 and 10, we will deduce the main auxiliary theorems stated in Section 2, as well as Theorem 1.1 and Corollary 1.2.

1.1. *Notation and basic definitions.* Let us conclude this introduction by collecting together for convenience some of the basic definitions and notation that we shall use throughout the paper. We shall denote the primes by $q_1, q_2, \ldots$, so q_z is the zth prime. (We use q here to avoid overusing the letter p, which will often denote a probability.) We shall write $[n] = \{1, \ldots, n\}$ for the first n positive integers and $[m, n]$ for the set $\{m, \ldots, n\}$. We shall also use the notation $a \in b \pm c$ to mean that

$$b - c \leqslant a \leqslant b + c.$$

In this paper, a *hypergraph* $\mathcal{H}$ will consist of a set $V(\mathcal{H})$ of *vertices* and a multi-set $E(\mathcal{H})$ of *hyperedges* (which we will usually refer to simply as *edges*). A hyperedge is just a subset of $V(\mathcal{H})$, and a k-edge is a hyperedge of size k. Note that we allow multiple copies of the same edge; all edge counts are taken with multiplicity. A hypergraph $\mathcal{H}' = (V', E')$ is a sub-hypergraph of $\mathcal{H} = (V, E)$ if $V' \subseteq V(\mathcal{H})$ and $E' \subseteq E(\mathcal{H})$ (so that each $e \in E'$ is a subset of V'). The *degree* of a vertex $v \in V(\mathcal{H})$ in $\mathcal{H}$ is the number of hyperedges containing it, counted with multiplicity. An *isolated vertex* is a vertex of degree 0. An *even* hypergraph is one in which all vertices have even degree.

The 2-*core* of a hypergraph $\mathcal{H}$ is the hypergraph obtained by repeatedly removing any vertex of degree at most 1, along with the corresponding edge when the degree is exactly one. Clearly, the 2-core is the union of all sub-hypergraphs of minimum vertex degree at least 2.

Finally, let us recall the standard Landau notation, which we shall use throughout the paper. Given functions $f(x)$ and $g(x)$, we write $f(x) = O(g(x))$

if $|f(x)| \leqslant C|g(x)|$ for some constant C and all sufficiently large x; and $f(x) = \Theta(g(x))$ if $f(x) = O(g(x))$, $g(x) = O(f(x))$, and $f(x)/g(x)$ is positive for all sufficiently large x. We write $f(x) = o(g(x))$ if $f(x)/g(x) \to 0$ as $x \to \infty$, and $f(x) = \omega(g(x))$ if $g(x) = o(f(x))$. Unless stated otherwise, all limits are as $x \to \infty$, where $\{1, \dots, x\}$ is the set from which the random numbers a_i are chosen. We shall avoid the notation $\Omega(f(x))$, $\ll$, and $\gg$, as these may mean different things to different mathematical communities.

2. **An outline of the proof**

In this section we shall precisely define the random process that we shall use to prove Theorem 1.1, and state our key results about this process, Theorems 2.2 and 2.6. Throughout the proof we fix a constant $\eta > 0$ and a sufficiently large integer x.[4] We set $N = \eta J(x)$ and define an N-tuple $(a_1, \dots, a_N)$ by choosing N elements of $[x]$ independently and uniformly at random (with replacement), and we form an $N \times \pi(x)$ 0-1 matrix A by setting $A_{ij} = 1$ if and only if the jth prime q_j occurs an odd number of times in the prime factorization of a_i. Thus, to find a subset $I \subseteq [N]$ such that $\prod_{i \in I} a_i$ is a square, it is enough to find a set of rows of A such that all column sums within these rows are even.

Note that the rows of A are chosen independently, but the columns are not. For example, the condition $a_i \leqslant x$ puts a limit on the number of times a 1 can occur in row i of A. More precisely, let $\widetilde{\Psi}(x, y)$ be the number of integers in $[x]$ all of whose prime factors that are strictly greater than y occur to an even power. Thus

$$\widetilde{\Psi}(x, y) = \sum_{t \in P(y)} \Psi\big(x/t^2, y\big), \tag{4}$$

where $P(y)$ is the set of all $t \geqslant 1$ that have no prime factor less than or equal to y. Define

$$p_j(x) := \frac{\widetilde{\Psi}(x, q_j) - \widetilde{\Psi}(x, q_{j-1})}{\widetilde{\Psi}(x, q_j)} \tag{5}$$

for each $j \in [\pi(x)]$, and observe that $p_j(x)$ is equal to the conditional probability that $A_{ij} = 1$ if $A_{ij'} = 0$ for every $j' > j$. Indeed, more generally we have

$$\mathbb{P}\Big(A_{ij} = 1 \,\Big|\, (A_{ij'})_{j'=j+1}^{\pi(x)}\Big) = p_j\Bigg(x \prod_{j' > j,\, A_{ij'} = 1} \frac{1}{q_{j'}}\Bigg). \tag{6}$$

[4] In the definitions below, we shall suppress the dependence on x and η.

Typically, $p_j(x)$ will be only slowly varying with x, and so the entries in a row of A will depend only mildly on one another. Nevertheless, this dependency is a major technicality that we shall need to overcome.

We can also think of the matrix A as a hypergraph whose vertices are the primes and whose edges correspond to the set of primes dividing a_i an odd number of times. We shall often wish to ignore small primes here, so a precise definition is as follows.

Definition 2.1. For each $z \in [0, \pi(x)]$, define $\mathcal{H}_A(z)$ to be the hypergraph with vertex set $V(\mathcal{H}_A(z)) = [z+1, \pi(x)]$ and hyperedge set $E(\mathcal{H}_A(z)) = \{e'_i : i \in [N]\}$, where

$$e'_i := \big\{j \in [z+1, \pi(x)] : A_{ij} = 1\big\}.$$

In particular, when $z = \pi(x)$, all of the edges of $\mathcal{H}_A(z)$ are empty.

Croot, Granville, Pemantle and Tetali [13] proved the upper bound in Theorem 1.1 by counting the number of *acyclic* (also called *Berge-acyclic*) even sub-hypergraphs of this hypergraph. An acyclic hypergraph is one in which there does not exist, for any $k \geqslant 2$, a cycle of k distinct hyperedges $e_0, e_1, \ldots, e_k = e_0$ and distinct vertices $v_1, \ldots, v_k$ with each $v_i \in e_{i-1} \cap e_i$, $i = 1, \ldots, k$. They showed that if $\eta > e^{-\gamma}$ then, for a suitable z, $\mathcal{H}_A(z)$ contains more than z edge-disjoint acyclic even sub-hypergraphs[5] with high probability. This guarantees more than z disjoint sets of rows of A, each of which has a sum in the subspace (taken over $\mathbb{F}_2$) of vectors that are supported on the first z columns. As any set of more than z vectors in this z-dimensional subspace is linearly dependent, this guarantees a linear relation between the rows of A. As noted in the introduction, the authors of [13] used the first moment method to prove their lower bound, counting the expected number of even sub-hypergraphs of $\mathcal{H}_A(0)$ or, equivalently, the number of sets of rows of A that sum to zero over $\mathbb{F}_2$. However, as mentioned in the introduction, this method does not yield a sharp lower bound as this expected number blows up before the threshold for the existence of a single such set given by Theorem 1.1.

2.1. *The* 2*-core of* $\mathcal{H}_A(z)$. Instead of counting the even sub-hypergraphs of $\mathcal{H}_A(0)$, we shall study the 2-core $\mathcal{C}_A(z)$ of the hypergraph $\mathcal{H}_A(z)$ for each $z_- \leqslant z \leqslant \pi(x)$, where $z_- = e^{-\Theta(\sqrt{\log y_0})} y_0$ is defined below; see (12). Since all even sub-hypergraphs (after removing isolated vertices) are sub-hypergraphs of the 2-core, it is enough to restrict attention to $\mathcal{C}_A(z)$. As noted in the introduction, this is equivalent to iteratively removing any a_i for which there

[5]Note that an empty hyperedge is an acyclic even sub-hypergraph corresponding to a q_z-smooth integer a_i; however, in order to find sufficiently many relations for all $\eta > e^{-\gamma}$, the authors of [13] needed to consider even sub-hypergraphs with an arbitrarily large (but bounded) number of hyperedges.

exists a prime $q > q_z$ that occurs to an odd power in a_i, but to an even power in all other remaining a_j. We shall show that if $\eta < e^{-\gamma}$, then the 2-core $\mathcal{C}_A(z_-)$ of $\mathcal{H}_A(z_-)$ is (with high probability) small by tracking the size of $\mathcal{C}_A(z)$ throughout the range $z \in [z_-, \pi(x)]$. In particular, we shall show that $\mathcal{C}_A(z_-)$ has fewer than z_- edges with high probability. As a consequence, any linear relation between the rows of A must involve fewer than z_- rows. This however is ruled out by a result in [13] that shows (via a first moment calculation) that any linear relation between the rows of A must involve many rows. We remark that this approach was partly inspired by the work of Pittel and Sorkin [27] in a closely related setting, where again a direct first moment calculation fails to find the correct threshold for the appearance of linear relations in the rows of a random matrix, but succeeds once restricted to the 2-core.

The following theorem tracks the size of the 2-core of $\mathcal{H}_A(z)$ for all $z \in [z_-, \pi(x)]$ and is the key technical statement we will need in order to prove Theorem 1.1. Let $M(z)$ be the set of rows of A corresponding to the set $E(\mathcal{C}_A(z)) = \{e_i' : i \in M(z)\}$ of hyperedges of the 2-core, and let $m(z) = |M(z)| = |E(\mathcal{C}_A(z))|$.

THEOREM 2.2. *If $\eta < e^{-\gamma}$ and $\varepsilon_0 > 0$, then with high probability,*

$$\frac{m(z)}{z} \exp\Big(- \mathrm{Ein}\Big(\frac{m(z)}{z}\Big)\Big) \in (1 \pm \varepsilon_0)\eta J(x) \frac{\Psi(x, q_z)}{xz} \tag{7}$$

for every $z \in [z_-, \pi(x)]$, where z_- is defined in (12) *below.*

Recall that the exponential integral $\mathrm{Ein}(w)$ is an entire function and is related to the incomplete gamma function via the relation

$$\mathrm{Ein}(w) := \int_0^w \frac{1 - e^{-t}}{t}\, dt = \Gamma(0, w) + \log w + \gamma. \tag{8}$$

Since $\Gamma(0, w) = \int_w^\infty e^{-t} \frac{dt}{t}$ decreases to 0 as $w \to \infty$, we see that $we^{-\mathrm{Ein}(w)}$ is a strictly increasing function of w that converges to $e^{-\gamma}$ as $w \to \infty$. Thus we can define $\alpha(\eta) \in [0, \infty)$ uniquely by the equation

$$\alpha(\eta) e^{-\mathrm{Ein}(\alpha(\eta))} = \eta \tag{9}$$

for any $\eta \in [0, e^{-\gamma})$. Note that $\frac{d}{dw} we^{-\mathrm{Ein}(w)} = e^{-w - \mathrm{Ein}(w)}$ and hence

$$\alpha'(\eta) = e^{\alpha(\eta) + \mathrm{Ein}(\alpha(\eta))}$$

is an increasing function of η. Thus $\alpha(\eta)$ is a convex function that strictly increases from 0 to ∞ as η increases from 0 to $e^{-\gamma}$.

Let us assume from now on that $0 < \eta < e^{-\gamma}$, so that $\alpha(\eta) \in (0, \infty)$ is well defined by (9). We shall fix sufficiently small positive constants ε_0, ε_1 and δ satisfying the following inequalities:

$$0 < \varepsilon_0 < e^{-\gamma} - \eta, \qquad 0 < \varepsilon_1 < \frac{\varepsilon_0}{16} e^{-C_0} \text{ and } 0 < \delta < \varepsilon_1 e^{-3/\varepsilon_1}, \tag{10}$$

where

$$C_0 := \alpha\big((1+\varepsilon_0)\eta\big). \tag{11}$$

Note that the upper bound on ε_0 implies that $(1+\varepsilon_0)\eta < e^{-\gamma}$ and hence $C_0 < \infty$. For convenience we shall also assume that $1/\varepsilon_1$ is an integer.

The constant ε_0 appears in Theorem 2.2 and determines the accuracy with which we track $m(z)$, while the constant ε_1 will appear (via Definition 2.5) in Theorem 2.6 below and will determine the accuracy to which we track various other parameters of the process. The constant δ plays a different role: it determines the "critical" range $[z_-, z_+]$, above which we shall have to use a slightly different approach, and below which we will lose control of the process. To be precise, set

$$z_- := \min\big\{z : \Lambda(z) \geqslant \delta\big\} \quad \text{and} \quad z_+ := \max\big\{z : \Lambda(z) \geqslant \delta\big\}, \tag{12}$$

where

$$\Lambda(z) = \Lambda_x(z) := J(x)\frac{\Psi(x,q_z)}{xz} \tag{13}$$

for each $z \in [\pi(x)]$. As in the introduction, let $\pi(y)x/\Psi(x,y)$ be minimized at $y = y_0$, and define $z_0 = \pi(y_0)$. Observe that we can take y_0 to be prime, since $\pi(y)$ and $\Psi(x,y)$ only change at prime y. It follows from (2) that $\Lambda(z) \leqslant 1$ and that $\Lambda(z_0) = 1$. Theorem 2.2 and (9) together imply that

$$m(z) \in \alpha\big((1 \pm \varepsilon_0)\eta\Lambda(z)\big)z \tag{14}$$

so, in particular, $m(z) \leqslant C_0 z$ for all $z \in [z_-, \pi(x)]$ (see Observation 4.7 below) and, moreover, $m(z) \approx \alpha(\eta)z$ when $z \approx z_0$. We will show later (see (37)) that $z_\pm = e^{\pm\Theta(\sqrt{\log z_0})}z_0$.

2.2. *The hypergraph* $\mathcal{S}_A(z)$. As mentioned in the introduction, the equation (7) comes from counting the number of isolated vertices in a certain hypergraph in two different ways. This hypergraph is not $\mathcal{C}_A(z)$, but (in a certain sense) its "dual," defined as follows.

Definition 2.3. For each $k \geqslant 2$ and $z \in [0, \pi(x)]$, let $S_k(z)$ denote the collection of vertices of degree k in $\mathcal{C}_A(z)$,

$$S_k(z) := \Big\{j \in [z+1, \pi(x)] : \Big|\big\{i \in M(z) : A_{ij} = 1\big\}\Big| = k\Big\}.$$

In other words, $S_k(z)$ is the set of all columns of A after column z that have column sum k when restricted to the set of rows $M(z)$. Note that for $z = \pi(x)$, we have $S_k(z) = \emptyset$. Also, these column sums are zero for $j \notin V(\mathcal{C}_A(z))$, so $S_k(z) \subseteq V(\mathcal{C}_A(z))$. Set $s_k(z) := |S_k(z)|$, and define $S(z) := \bigcup_{k\geqslant 2} S_k(z)$.

We shall think of $S_k(z)$ as labeling the k-edges $e_j := \{i \in M(z) : A_{ij} = 1\}$ of a hypergraph $\mathcal{S}_A(z)$ with vertex set $M(z)$ and edge set $\{e_j : j \in S(z)\}$. Note that we are now thinking of the rows (corresponding to numbers a_i) of A as

being the vertices and the columns (primes) as hyperedges, where each prime q corresponds to the set of $i \in M(z)$ such that q divides a_i an odd number of times.

Later we shall show that $p_j(x) = (1 + o(1))/j$ when j is in the critical range $z_- \leqslant j \leqslant z_+$. Thus heuristically one would expect that

$$\mathbb{P}\big(j \in S_k(z)\big) \approx \mathbb{P}\big(\mathrm{Bin}(m(z), 1/j) = k\big)$$

for $k \geqslant 2$, where $\mathrm{Bin}(n, p)$ denotes a binomial random variable with n trials and success probability p. Indeed, if a column has at least two 1s in active rows (i.e., rows in $M(z)$), then this column has no effect on the construction of the 2-core. Also, one would expect that the events $\big\{j \in S_k(z) : j \in [z+1, \pi(x)]\big\}$ are "approximately independent." This leads one (after a short calculation) to predict that $s_k(z)$ tracks the following function.

Definition 2.4. For each $k \geqslant 2$ and every $z \in [\pi(x)]$, set

$$\hat{s}_k(z) := \frac{m(z)}{k(k-1)} e^{-m(z)/z} \sum_{\ell=k-1}^{\infty} \frac{1}{\ell!} \Big(\frac{m(z)}{z}\Big)^{\ell}. \tag{15}$$

Note that $\hat{s}_k(z)$ is a decreasing function of k and that

$$\hat{s}_2(z) = \frac{m(z)}{2}\big(1 - e^{-m(z)/z}\big). \tag{16}$$

We next define a function that we shall use to bound the error in $s_k(z)$.

Definition 2.5. For each $z \in [z_-, z_+]$ and each $k \geqslant 2$, define

$$\varepsilon(k, z) := \frac{\varepsilon_1^k \cdot k!}{\Lambda(z)}.$$

The function $\varepsilon(k, z)$ decreases exponentially fast in k while k is relatively small, and then increases super-exponentially fast when k is large. We will need the former property in order to obtain the self-correction (see below) that will play a crucial role in our proof, and the latter property in order to show that the bound (17) holds when k is reasonably large.[6] We shall prove the following theorem.

THEOREM 2.6. *Suppose $\eta < e^{-\gamma}$. Then, with high probability,*

$$s_k(z) \in \big(1 \pm \varepsilon(k, z)\big)\hat{s}_k(z) \tag{17}$$

holds for every $k \geqslant 2$ and every $z \in [z_-, z_+]$.

Note that $\varepsilon(k, z) > 1$ for all sufficiently large k, so to prove that (17) holds for these values of k, it will suffice to show that $s_k(z) = 0$. We shall in fact

[6]More precisely, it will be important that $\varepsilon(k, z)\hat{s}_k(z)$ decreases only exponentially fast in k; see Observation 4.16 and its applications in Section 8.

show that, with high probability, $s_k(z) = 0$ for all $k \geqslant 5 \log z_0 / \log\log z_0$ and for all $z \in [z_-, \pi(x)]$ (see Lemma 4.13 below).

As mentioned in the introduction, we shall prove Theorem 2.6 using the method of self-correcting martingales (see, e.g., [16, §3]). Roughly speaking, we shall show (see Lemmas 7.5 and 8.1) that if nothing has yet gone wrong, then the (expected) drift in the *relative* error of $s_k(z)$ depends mainly on the current error, and (unless it is already quite small) tends to push the error towards zero. We emphasize that the function $\varepsilon(k, z)$ was chosen with exactly these lemmas in mind; in particular, it will be important that $\varepsilon(k, z)$ decreases rapidly for small values of k, since we shall use this fact to bound the influence of the error in $s_{k+1}(z)$ on the drift of $s_k(z)$. Combining these lemmas with bounds on the probability of a large jump in the relative error (see Lemmas 7.6 and 8.2), it will be relatively easy to deduce sufficiently strong bounds on the probability that $s_k(z)$ is the first variable to "go astray."

Theorem 2.2 will be proved simultaneously with Theorem 2.6, but we will not show that $m(z)$ is self-correcting; instead we shall show that $m(z)$ is unlikely to go off track before any of the $s_k(z)$. More precisely, we shall use a martingale approach to show that $m(z)$ does not drift off course too quickly, together with an occasional application of Lemma 9.3 to put it back on track. Since the probability of failure in Lemma 9.3 is relatively large, we can only apply it a small number of times; however, this will be sufficient to prove Theorem 2.2 over the "critical" interval $[z_-, z_+]$, while larger values of z are easier to deal with.

2.3. *The random process.* Let us finish this section by defining the random process we shall use to reveal the 2-core $\mathcal{C}_A(z)$. In each step we reveal just enough information to proceed; in particular, and crucially, we shall not reveal the exact locations of the 1s in a column until it has only a single non-zero element in an *active* row, that is, a row of $M(z)$.

Algorithm 2.7. We start with $z := \pi(x)$, $M(z) := [N]$ and $S_k(z) := \emptyset$ for each $k \geqslant 2$. Now repeat the following steps until $z = 0$:

1. Set $M := M(z)$, $S_k := S_k(z)$ for each $k \geqslant 2$ and $S_1 := \emptyset$.
2. Reveal the (random) quantity $d(z) := \left|\left\{i \in M(z) : A_{iz} = 1\right\}\right|$, that is, the number of active non-zero entries of column z.
3. If $d(z) = d > 0$, set $S_d := S_d \cup \{z\}$.
4. While $S_1 \neq \emptyset$ do the following:
   (a) Pick the smallest[7] $z' \in S_1$, observe which row i is such that $i \in M$, $A_{iz'} = 1$.

[7] An arbitrary $z' \in S_1$ would do here, but we shall later wish to ensure that the order in which rows are removed by the algorithm is uniquely specified.

(b) Set $M := M \setminus \{i\}$, $S_1 := S_1 \setminus \{z'\}$.
(c) For each $k \geqslant 1$ and each $j \in S_k$, reveal whether column j has a 1 in row i; if it does, remove j from S_k and (if $k > 1$) add it to S_{k-1}.

5. Set $M(z-1) := M$ and $S_k(z-1) := S_k$ for each $k \geqslant 2$.
6. Set $z := z - 1$; if $z > 0$, then return to Step 1, otherwise stop.

It is easy to see that this algorithm tracks the 2-core $\mathcal{C}_A(z)$ as z decreases from $\pi(x)$ to 0. Define a filtration $\mathcal{F}_{\pi(x)} \subseteq \mathcal{F}_{\pi(x)-1} \subseteq \cdots$ by taking $\mathcal{F}_y$ to be the information observed at the moment the index z is set equal to y. More precisely, $\mathcal{F}_y$ reveals which rows and columns of A correspond to the edges and vertices of the 2-core $\mathcal{C}_A(y)$, as well as the degrees (column sums) of all the vertices in the 2-core. The only other information revealed by $\mathcal{F}_y$ concerns rows of A outside of $M(y)$ (as a result of earlier steps of the algorithm), which will be irrelevant for our purposes.

Define the σ-algebra $\mathcal{F}_y^+$ to include $\mathcal{F}_y$ and also the information observed at Step 2 when $z = y$, namely, the value of $d(y)$. Thus $\mathcal{F}_y^+$ specifies the column sums of A of all columns in $[y, \pi(x)]$, summing only over rows in $M(y)$. The matrix A conditioned on $\mathcal{F}_y^+$ and restricted to $M(y) \times [y, \pi(x)]$ can be constructed with the correct probability distribution by taking a uniform distribution on all choices of the multi-set of numbers $\{a_i : i \in M(y)\}$ whose column sums are compatible with $\mathcal{F}_y^+$. Indeed, any such multi-set, combined with the original a_i for all $i \notin M(y)$, would result in the algorithm constructing the same 2-core, and all such choices of the a_i are equally likely. Understanding this probability space will be the main aim of the next three sections and will be key to the proof of Theorem 1.1.

3. **Number-theoretic facts**

In order to understand the distribution of the numbers $\{a_i : i \in M(z)\}$ conditioned on the information observed in $\mathcal{F}_z$ or $\mathcal{F}_z^+$, we shall need some detailed information about the smooth number counting function $\Psi(x, y)$ and its close relative $\widetilde{\Psi}(x, y)$. The first result in this direction was obtained by Dickman [14] in 1930, who proved that if u is fixed, then

$$\lim_{x\to\infty} \frac{\Psi(x, x^{1/u})}{x} = \rho(u),$$

where ρ is the (unique) continuous solution to the delay differential equation

$$u\rho'(u) + \rho(u-1) = 0 \tag{18}$$

for $u > 1$, with the boundary condition $\rho(u) = 1$ for all $0 \leqslant u \leqslant 1$. This function is now known as the Dickman–de Bruijn function, and it is asymptotically of the form

$$\rho(u) = u^{-(1+o(1))u} \tag{19}$$

as $u \to \infty$. Further important breakthroughs were made in 1938, by Rankin [32], and in 1951, by de Bruijn [9], who determined $\Psi(x,y)$ asymptotically when $y \geqslant \exp\big((\log x)^{5/8+\varepsilon}\big)$ for some $\varepsilon > 0$. Upper and lower bounds in a much wider range were later proved by de Bruijn [10] and by Canfield, Erdős and Pomerance [11], respectively. We will use the following asymptotic result, due to Hildebrand [17].

THEOREM 3.1 (Hildebrand, 1986). *Let* $\exp\big((\log\log x)^2\big) \leqslant y \leqslant x$, *and set* $u = \frac{\log x}{\log y}$. *Then*

$$\frac{\Psi(x,y)}{x} = \rho(u)\bigg(1 + O\bigg(\frac{\log(u+1)}{\log y}\bigg)\bigg)$$

uniformly in x *and* y.

We remark that the main result of [17] is even more general than Theorem 3.1, but the version above is a little simpler to state, and more than sufficient for our purposes. Indeed, it follows from Theorem 3.1 and (19) (see, for example, [13, §2.1]) that[8]

$$J(x) = y_0^{2+o(1)} = \exp\Big(\big(1+o(1)\big)\sqrt{2\log x \log\log x}\Big) \tag{20}$$

as $x \to \infty$, as claimed in the introduction. It also follows that

$$\Psi(x, y_0^\beta) = x y_0^{-1/\beta + o(1)} \tag{21}$$

for every $\beta = \beta(x)$ bounded away from 0, which implies the following crude estimate for $\Lambda(z)$.

COROLLARY 3.2. *Let* $\beta = \beta(x)$ *be bounded away from* 0, *and set* $z = z_0^\beta$, *where* $z_0 = \pi(y_0)$. *Then*

$$\Lambda(z) = z_0^{2-\beta-1/\beta+o(1)}$$

as $x \to \infty$.

Note that Corollary 3.2 also follows from [13, Lemma 2.1] and implies that

$$z_\pm = z_0^{1+o(1)}. \tag{22}$$

Let us define $u_0 := \frac{\log x}{\log y_0}$, so that $y_0 = x^{1/u_0}$. For future reference, we note here the following immediate and useful consequence of (20):

$$\log z_0 = (1+o(1))u_0 \log u_0. \tag{23}$$

In order to obtain more detailed information about the function $\Lambda(z)$, we shall need some fundamental results of Hildebrand and Tenenbaum [18, Th. 3],

[8] In fact, to prove (20) one only needs the results of de Bruijn [10] and Canfield, Erdős and Pomerance [11], which imply that $\Psi(x,y) = x u^{-(1+o(1))u}$ as $u \to \infty$ for all $y \geqslant (\log x)^{1+\varepsilon}$.

which control the "local" dependence of $\Psi(x,y)$ on the variable x. (We remark that the idea of using these to understand the matrix A was one of the key innovations of [13].) Instead of quoting these results directly, we shall prove a form (Theorem 3.5) that will be more convenient for our purposes. We will need the following two results on the Dickman–de Bruijn function $\rho(u)$.

THEOREM 3.3 (Hildebrand [17, proof of Lemma 1]). *The function $\rho(u)$ is log-concave for $u \geqslant 1$. Equivalently, $\frac{\rho(u-1)}{u\rho(u)}$ is increasing in u.*

Define the function $\xi = \xi(u)$ for $u > 1$ to be the unique positive solution of the equation

$$e^{\xi(u)} = 1 + u\xi(u). \tag{24}$$

THEOREM 3.4 (Hildebrand and Tenenbaum [19, eq. (2.1)′]). *For $u > 1$, we have*

$$\rho(u) = \left(1 + \frac{O(1)}{u}\right)\sqrt{\frac{\xi'(u)}{2\pi}}\exp\left(\gamma - \int_1^u \xi(t)\,dt\right).$$

We now state our key estimate on the rate of change of the function $\rho(u)$. This essentially follows from [19, Cor. 2.4], but since the precise version we need is not an immediate consequence of the results stated in [19], we shall give the proof for completeness.

THEOREM 3.5. *Let $u > 1$ and $a, b \geqslant 0$ satisfy $a + b \leqslant u$. Then*

$$\frac{\rho(u-a-b)}{\rho(u-a)} \leqslant \exp\left(b\xi(u) + \frac{O(1)}{u}\right). \tag{25}$$

If, in addition, $a + b \leqslant u/2$, then

$$\frac{\rho(u-a-b)}{\rho(u-a)} = \exp\left(b\xi(u) + \frac{O(b^2+ab+1)}{u}\right). \tag{26}$$

We shall use the following simple facts about the function $\xi(u)$, which we collect here for convenience:

$$\xi(u) = \big(1 + o(1)\big)\log u, \qquad \xi'(u) = \frac{\Theta(1)}{u} \qquad \text{and} \qquad \xi''(u) = -\frac{\Theta(1)}{u^2} \tag{27}$$

for all $u > 1$. We remark that here the $o(1)$ is as $u \to \infty$.

Proof. Suppose first that $a = 0$ and $u - b > 1$, and apply Theorem 3.4 to $u - b$ and u to obtain

$$\log\frac{\rho(u-b)}{\rho(u)} = \frac{1}{2}\log\frac{\xi'(u-b)}{\xi'(u)} + \int_{u-b}^u \xi(t)\,dt + \frac{O(1)}{u-b}.$$

We shall bound each of the terms on the right in turn. First, observe that

$$\log\frac{\xi'(u-b)}{\xi'(u)} = -\int_{u-b}^u \frac{\xi''(t)}{\xi'(t)}\,dt = O(1)\int_{u-b}^u \frac{dt}{t} = O\Big(\log\frac{u}{u-b}\Big),$$

where the first step follows by differentiating $\log \xi'(t)$, and the second follows from (27). Next, note that integration by parts gives

$$b\xi(u) - \int_{u-b}^{u} \xi(t)\,dt = \int_{u-b}^{u} \xi'(t)(t-u+b)\,dt$$
$$= \Theta(1)\int_{u-b}^{u} \frac{t-u+b}{t}\,dt = \frac{\Theta(b^2)}{u},$$

where the second step follows from (27), and the last equality holds for $u-b \geqslant \varepsilon u$ as $\int_{u-b}^{u}(t-u-b)\,dt = b^2/2$ and $t = \Theta(u)$ for $t \in [\varepsilon u, u]$. On the other hand, the last integral increases to b as $u-b \to 0^+$, so it also holds for $1 < u-b \leqslant \varepsilon u$.

Combining the three equations above, we obtain

$$\log \frac{\rho(u-b)}{\rho(u)} = b\xi(u) + O\Big(\log \frac{u}{u-b}\Big) - \frac{\Theta(b^2)}{u} + \frac{O(1)}{u-b}.$$

Now for $0 \leqslant b \leqslant u/2$, we have $\log \frac{u}{u-b} = O(b/u)$ and $1/(u-b) = O(1/u)$, so

$$\frac{\rho(u-b)}{\rho(u)} = \exp\left(b\xi(u) + \frac{O(b+1) - \Theta(b^2)}{u}\right). \tag{28}$$

This clearly also holds for $0 \leqslant u-b \leqslant 1$ as then $u \leqslant 2$ is bounded. Thus (26) holds for $a = 0$.

Now suppose $1 < u-b \leqslant u/2$. Then $b^2/u = \Theta(u)$, while $\log \frac{u}{u-b} \leqslant \log u$ and $1/(u-b) = O(1)$. Thus we obtain

$$\frac{\rho(u-b)}{\rho(u)} \leqslant \exp\left(b\xi(u) + \frac{O(1)}{u}\right). \tag{29}$$

However, (29) also follows from (28) when $u-b \geqslant u/2$ as the $O(b+1) - \Theta(b^2)$ term is bounded above by a constant for all $b \geqslant 0$. Finally, note that $\xi(u) \geqslant 0$ and $\rho(u-b) = 1$ for every $0 \leqslant u-b \leqslant 1$, so it follows that (29) in fact holds for all $0 \leqslant b \leqslant u$. In particular, (25) holds for $a = 0$.

In order to deduce (25) and (26) in the case when $a > 0$, we substitute $u-a$ for u in (28) and (29). In the former case, this gives

$$\frac{\rho(u-a-b)}{\rho(u-a)} = \exp\left(b\xi(u-a) + \frac{O(b+1) - \Theta(b^2)}{u-a}\right)$$

for all $a+b \leqslant u/2$, which implies (26) since $\xi(u) = \xi(u-a) + O(a/u)$, by (27), and $u-a = \Theta(u)$. Similarly, substituting $u-a$ for u in (29) gives

$$\frac{\rho(u-a-b)}{\rho(u-a)} \leqslant \exp\left(b\xi(u-a) + \frac{O(1)}{u-a}\right) \leqslant \exp\left(b\xi(u) + \frac{O(1)}{u-a}\right)$$

since ξ is an increasing function. This is enough to prove (25) when $u-a \geqslant \varepsilon u$, so let us assume instead that $u-a \leqslant \varepsilon u$. Now, observe that

$$-\frac{\rho'(u)}{\rho(u)} = \frac{\rho(u-1)}{u\rho(u)} \leqslant \frac{e^{\xi(u)+O(1/u)}}{u} = \xi(u) + o(1)$$

as $u \to \infty$, where the first step follows by the definition (18) of ρ, the second step follows by (29) applied with $b = 1$, and the third step follows by the definition (24) of ξ and (27). It follows that $-\rho'(u)/\rho(u) \leqslant \xi(Cu)$ for some absolute constant $C > 1$ (since $\xi'(u) = \Theta(1/u)$), and hence

$$\log \frac{\rho(u-a-b)}{\rho(u-a)} = -\int_{u-a-b}^{u-a} \frac{\rho'(t)}{\rho(t)}\, dt \leqslant b\xi\big(C(u-a)\big) \leqslant b\xi(u)$$

if ε was chosen sufficiently small, since ξ is increasing. Thus (25) also holds in this case, and the proof is complete. □

The probabilities in our model are given in terms of the modified smooth number counting function $\widetilde{\Psi}(x,y)$. We now show that there is little difference between $\Psi(x,y)$ and $\widetilde{\Psi}(x,y)$.

Lemma 3.6. *If* $y \geqslant \exp\big((\log\log x)^2\big)$, *then*

$$\Psi(x,y) \leqslant \widetilde{\Psi}(x,y) \leqslant \big(1 + y^{-1+o(1)}\big)\Psi(x,y)$$

uniformly in y *as* $x \to \infty$.

Proof. The lower bound holds trivially, by definition; we shall prove the upper bound. We may assume $y \leqslant x$, as otherwise $\Psi(x,y) = \widetilde{\Psi}(x,y) = x$. Recall from (4) that

$$\widetilde{\Psi}(x,y) = \sum_{t \in P(y)} \Psi(x/t^2, y),$$

where $P(y)$ is the set of positive integers whose prime factors are all strictly larger than y. Applying Theorems 3.1 and 3.5 with $a = 0$, $b = 2\frac{\log t}{\log y}$, we obtain

$$\begin{aligned}\frac{t^2\Psi(x/t^2,y)}{\Psi(x,y)} &\leqslant \left(1 + \frac{O(\log(u+1))}{\log y}\right)\frac{\rho(u-b)}{\rho(u)} \\ &\leqslant \exp\left(b\xi(u) + \frac{O(1)}{u} + \frac{O(\log(u+1))}{\log y}\right)\end{aligned}$$

for all $t \in [y, \sqrt{x}]$, where $u = \frac{\log x}{\log y} \geqslant u - b = \frac{\log(x/t^2)}{\log y} \geqslant 0$. Note that in the case when $x/t^2 < y$, we apply Theorem 3.1 only to $\Psi(x,y)$, as in this case $t^2\Psi(x/t^2,y) \leqslant x$ and $\rho(u-b) = 1$. Now $\xi(u) = O(\log u) = o(\log y)$ for $y \geqslant \exp\big((\log\log x)^2\big)$. Hence $t^2\Psi(x/t^2,y)/\Psi(x,y) = t^{o(1)}$, and so

$$\begin{aligned}\widetilde{\Psi}(x,y) - \Psi(x,y) &= \sum_{1<t\in P(y)} \Psi(x/t^2,y) \\ &\leqslant \Psi(x,y)\sum_{t\geqslant y} t^{-2+o(1)} \leqslant y^{-1+o(1)}\Psi(x,y),\end{aligned}$$

as claimed. □

We can now state our estimates for the rate of change of the functions $\Psi(x,y)$ and $\widetilde{\Psi}(x,y)$. The form of the statement below is designed to facilitate its application in Section 5.

THEOREM 3.7. *Let* $z \in [z_-, \pi(x)]$, *set* $y = q_z$ *and* $u = \frac{\log x}{\log y}$, *and assume* $0 \leqslant a \leqslant u$ *and* $b \geqslant 1$. *Then*

$$\log \frac{\Psi(xy^{-(a+b)}, y)}{\Psi(xy^{-a}, y)} \leqslant b\big(\xi(u) - \log y\big) + \frac{O(1)}{u}. \tag{30}$$

Moreover, if $a + b \leqslant u/2$, *then*

$$\log \frac{\Psi(xy^{-(a+b)}, y)}{\Psi(xy^{-a}, y)} = b\big(\xi(u) - \log y\big) + \frac{O(b^2 + ab + 1)}{u}. \tag{31}$$

Also, both statements hold with $\widetilde{\Psi}$ *in place of* Ψ.

Note that we interpret $\log 0 = -\infty$ in the case when $a + b > u$ (and so $\Psi(xy^{-(a+b)}, y) = 0$).

Proof. The bound $y \geqslant \exp((\log\log x)^2)$ follows from (20), since $z \geqslant z_-$, so $y \geqslant y_0^{1+o(1)}$, by (22). Thus by Theorem 3.1, we have

$$\frac{\Psi(xy^{-(a+b)}, y)}{\Psi(xy^{-a}, y)} = y^{-b} \cdot \frac{\rho(u-a-b)}{\rho(u-a)}\left(1 + O\left(\frac{\log(u+1)}{\log y}\right)\right)$$

if $a + b \leqslant u - 1$ and $b \geqslant 0$, since these inequalities imply that $y \leqslant xy^{-(a+b)} \leqslant xy^{-a}$. Both (30) and (31) now follow by Theorem 3.5, since $\log(u+1)/\log y = O(1/u)$, by (23), and the corresponding bounds with Ψ replaced by $\widetilde{\Psi}$ follow using Lemma 3.6.

The case when $a + b > u - 1$ and $a + b \leqslant u/2$ is ruled out by the assumption[9] that $b \geqslant 1$, so it only remains to prove that (30) holds when $u - 1 \leqslant a + b \leqslant u + b$ and $b \geqslant 1$. If $a \leqslant u - 1$, then we can apply (30) with b replaced by $b' := u - 1 - a$ (noting that we in fact proved it for all $a + b \leqslant u - 1$ and $b \geqslant 0$) to obtain

$$\log \frac{\Psi(xy^{-(a+b')}, y)}{\Psi(xy^{-a}, y)} = \log \frac{y}{\Psi(xy^{-a}, y)} \leqslant b'\big(\xi(u) - \log y\big) + \frac{O(1)}{u}.$$

Noting that $\Psi(xy^{-(a+b)}, y) \leqslant xy^{-(a+b)} = y^{1+b'-b}$, it follows that

$$\log \frac{\Psi(xy^{-(a+b)}, y)}{\Psi(xy^{-a}, y)} \leqslant \log \frac{y^{1+b'-b}}{\Psi(xy^{-a}, y)} \leqslant b'\xi(u) - b\log y + \frac{O(1)}{u},$$

[9]Note that we need some condition on the parameters in the statement of the theorem in order to rule out the case when, say, $y^b \approx 2$, but $1 \leqslant xy^{-(a+b)} < xy^{-a} < 2$, and $u \to \infty$.

which implies (30) since $b' \leqslant b$. The proof with Ψ replaced by $\widetilde{\Psi}$ is identical. Finally, if $a > u-1$ and $b \geqslant 1$, then $xy^{-(a+b)} < 1$, and so the claimed bound holds trivially. □

We are now ready to bound the conditional probability of $A_{ij} = 1$ that is given by (6).

Corollary 3.8. *If $z \in [z_-, z_+]$ and $a = o(u_0)$, then*

$$p_z\big(xq_z^{-a}\big) = \frac{1+o(1)}{z}.$$

Moreover, for any $z \in [z_-, \pi(x)]$ and $a \geqslant 1$, we have

$$p_z\big(xq_z^{-a}\big) \leqslant \big(1+o(1)\big)p_z(x) \leqslant \frac{1+o(1)}{z}.$$

Furthermore, $p_z(x) = z^{-1+o(1)}$ for every $z \in [z_-, \pi(x)]$.

We shall use the following observation in the proof of Corollary 3.8.

Observation 3.9. *If $z = z_0^{1+o(1)}$ and $x = q_z^u$, then $e^{\xi(u)} = (1+o(1))q_z/z$.*

Proof. Note that $z = z_0^{1+o(1)}$ implies that $u = (1+o(1))u_0$, and hence

$$\begin{aligned} e^{\xi(u)} &= 1 + u\xi(u) = (1+o(1))u\log u = (1+o(1))u_0\log u_0 \\ &= (1+o(1))\log z_0 = (1+o(1))\log q_z = (1+o(1))q_z/z, \end{aligned}$$

by (23), (24) and the prime number theorem, as claimed. □

Proof of Corollary 3.8. Set $x' = xq_z^{-a}$, and observe that

$$\widetilde{\Psi}(x', q_z) - \widetilde{\Psi}(x', q_{z-1}) = \widetilde{\Psi}(x'q_z^{-1}, q_{z-1}) = \widetilde{\Psi}(x'q_z^{-1}, q_z) - \widetilde{\Psi}(x'q_z^{-2}, q_{z-1})$$

and $0 \leqslant \widetilde{\Psi}(x'q_z^{-b}, q_{z-1}) \leqslant \widetilde{\Psi}(x'q_z^{-b}, q_z)$ for $b \in \{1,2\}$. Therefore, recalling the definition (5) of $p_z(x)$,

$$\frac{\widetilde{\Psi}(x'q_z^{-1}, q_z) - \widetilde{\Psi}(x'q_z^{-2}, q_z)}{\widetilde{\Psi}(x', q_z)} \leqslant p_z(x') \leqslant \frac{\widetilde{\Psi}(x'q_z^{-1}, q_z)}{\widetilde{\Psi}(x', q_z)}. \tag{32}$$

Let $x = q_z^u$, and note that if $z \in [z_-, z_+]$, then $u = (1+o(1))u_0$, since $z_\pm = z_0^{1+o(1)}$, by (22). Thus, if $a = o(u_0)$ then, by Theorem 3.7, we have

$$\log \frac{\widetilde{\Psi}(x'q_z^{-b}, q_z)}{\widetilde{\Psi}(x', q_z)} = b\big(\xi(u) - \log q_z\big) + o(1)$$

for $b \in \{1,2\}$. Moreover, $e^{\xi(u)} = (1+o(1))q_z/z$ by Observation 3.9, since $z_\pm = z_0^{1+o(1)}$. It follows that

$$\frac{\widetilde{\Psi}(x'q_z^{-b}, q_z)}{\widetilde{\Psi}(x', q_z)} = \left(\frac{1+o(1)}{z}\right)^b,$$

which together with (32) implies that $p_z(x') = (1+o(1))/z$ when $a = o(u_0)$, as claimed.

The bounds $p_z(x) \leqslant (1+o(1))/z$ and $p_z(x) = z^{-1+o(1)}$ follow by a similar argument. Indeed, by Theorem 3.7 (applied with $a = 0$), we have

$$\log \frac{\widetilde{\Psi}(xq_z^{-b}, q_z)}{\widetilde{\Psi}(x, q_z)} \leqslant b\big(\xi(u) - \log q_z\big) + \frac{O(1)}{u} \tag{33}$$

for $b \in \{1, 2\}$, and moreover a matching lower bound holds when $u \geqslant 2b$. Note also that, since $1 \leqslant u \leqslant (1+o(1))u_0$, we have

$$e^{\xi(u)+O(1/u)} \leqslant \big(1+o(1)\big)e^{\xi(u_0)} = \big(1+o(1)\big)\frac{q_{z_0}}{z_0} \leqslant \big(1+o(1)\big)\frac{q_z}{z}$$

by (27), Observation 3.9 and the prime number theorem. Thus by (33),

$$\frac{\widetilde{\Psi}(xq_z^{-b}, q_z)}{\widetilde{\Psi}(x, q_z)} \leqslant \left(\frac{1+o(1)}{z}\right)^b, \tag{34}$$

and (32) then implies $p_z(x) \leqslant (1+o(1))/z$. Moreover, the lower bound $p_z(x) \geqslant e^{O(1)}/q_z$ holds for $u \geqslant 2$ by (32), using the matching lower bound in (33) when $b = 1$. For $1 \leqslant u < 2$, $\widetilde{\Psi}(x/q_z, q_z) = \lfloor x/q_z \rfloor = \Theta(x/q_z)$ and $\widetilde{\Psi}(x, q_z) \leqslant x$, so again $p_z(x) \geqslant c/q_z$ for some $c > 0$. In particular, $p_z(x) \geqslant z^{-1+o(1)}$.

Finally, the bound $p_z(xq_z^{-a}) \leqslant (1+o(1))p_z(x)$ is also similar, though since it does not always hold when $a < 1$, we shall need to be a little careful. Note first that if $u < a+1$, then the claimed bounds hold trivially, since $xq_z^{-a} < q_z$, and hence $p_z(xq_z^{-a}) = 0$. We may therefore assume that $u \geqslant a+1$, which implies that $xq_z^{-1} \geqslant q_z$ and $x' \geqslant q_z$. We claim that

$$p_z(x') \leqslant \frac{\widetilde{\Psi}(x'q_z^{-1}, q_z)}{\widetilde{\Psi}(x', q_z)} \leqslant \big(1+o(1)\big)\frac{\widetilde{\Psi}(xq_z^{-1}, q_z)}{\widetilde{\Psi}(x, q_z)} \leqslant \big(1+o(1)\big)p_z(x). \tag{35}$$

Indeed, the first inequality follows by (32), and the second follows since Theorem 3.3 implies that $\frac{\rho(u-1)}{u\rho(u)}$, and hence $\frac{\rho(u-1)}{\rho(u)}$, is an increasing function of u, so if $u \geqslant a+2$, then

$$\begin{aligned}\frac{\widetilde{\Psi}(x'q_z^{-1}, q_z)}{\widetilde{\Psi}(x', q_z)} &\leqslant \big(1+o(1)\big)q_z^{-1} \cdot \frac{\rho(u-a-1)}{\rho(u-a)} \\ &\leqslant \big(1+o(1)\big)q_z^{-1} \cdot \frac{\rho(u-1)}{\rho(u)} = \big(1+o(1)\big)\frac{\widetilde{\Psi}(xq_z^{-1}, q_z)}{\widetilde{\Psi}(x, q_z)}\end{aligned}$$

by Theorem 3.1 and Lemma 3.6. On the other hand, if $a+1 \leqslant u \leqslant a+2$, then $x'q_z^{-1} < q_z$, so in this case we may replace the bound on $\widetilde{\Psi}(x'q_z^{-1}, q_z)$ given by Theorem 3.1 by the equality $\widetilde{\Psi}(x'q_z^{-1}, q_z) = x'q_z^{-1} = \rho(u-a-1)x'q_z^{-1}$, since by definition, $\rho(u) = 1$ for all $0 \leqslant u \leqslant 1$.

To deduce (35) from (32), it therefore only remains to observe that, by (34) and two applications of Theorem 3.7,

$$\frac{\widetilde{\Psi}(xq_z^{-2}, q_z)}{\widetilde{\Psi}(x, q_z)} \leqslant O(1) \cdot \left(\frac{\widetilde{\Psi}(xq_z^{-1}, q_z)}{\widetilde{\Psi}(x, q_z)}\right)^2 = o(1) \cdot \frac{\widetilde{\Psi}(xq_z^{-1}, q_z)}{\widetilde{\Psi}(x, q_z)},$$

where we again used the fact that $u \geqslant a+1 \geqslant 2$ when applying Theorem 3.7 with $b=1$. This proves (35), and hence completes the proof of the corollary. □

We next deduce some more refined estimates concerning the function $\Lambda(z)$.

LEMMA 3.10. *If* $z = z_0^{1+o(1)}$, *then*

$$\frac{\Lambda(z-1)}{\Lambda(z)} = 1 + \frac{o(1)}{z}.$$

Proof. Noting that $\Psi(x,q_z) = \Psi(x,q_{z-1}) + \Psi(xq_z^{-1},q_z)$, and recalling the definition (13) of $\Lambda(z)$, we have

$$\frac{\Lambda(z)-\Lambda(z-1)}{\Lambda(z)} = \frac{(z-1)\Psi(x,q_z) - z\Psi(x,q_{z-1})}{(z-1)\Psi(x,q_z)} = \frac{z\Psi(xq_z^{-1},q_z) - \Psi(x,q_z)}{(z-1)\Psi(x,q_z)}.$$

Now, applying Theorem 3.7 with $y=q_z$, $a=0$ and $b=1$, we obtain

$$\frac{\Psi(xq_z^{-1},q_z)}{\Psi(x,q_z)} = \exp\Big(\xi(u) - \log q_z + o(1)\Big),$$

where $u = \frac{\log x}{\log q_z} = (1+o(1))u_0 \to \infty$. Hence

$$\frac{\Lambda(z)-\Lambda(z-1)}{\Lambda(z)} = \frac{1}{z-1}\left(\frac{ze^{\xi(u)}}{q_z}\Big(1+o(1)\Big) - 1\right) = \frac{o(1)}{z}, \tag{36}$$

by Observation 3.9. □

LEMMA 3.11. *Write* $z = z_0 \exp\Big(c\sqrt{\log z_0}\Big)$ *for some* $c = c(x)$. *Then*

$$\Lambda(z) = e^{-c^2} + o(1).$$

Proof. We may assume without loss of generality that $z = z_0^{1+o(1)}$, and hence $c = o(\sqrt{\log z_0})$, as otherwise the result follows from Corollary 3.2 with $\Lambda(z) = o(1)$. Set $y = q_z$ and $u = \frac{\log x}{\log y}$. Since $\Lambda(z) = J(x)\Psi(x,q_z)/xz$ and $\Lambda(z_0) = 1$, it follows by Theorem 3.1 that

$$\Lambda(z) = \frac{\Lambda(z)}{\Lambda(z_0)} = \frac{z_0}{z}\cdot\frac{\Psi(x,y)}{\Psi(x,y_0)} = \Big(1+o(1)\Big)\frac{z_0}{z}\cdot\frac{\rho(u)}{\rho(u_0)}.$$

Set $b = u_0 - u$, and note that $b = o(u_0)$, which by Theorem 3.5 implies that

$$\frac{\rho(u)}{\rho(u_0)} = \exp\left(b\xi(u_0) + \frac{O(b^2+1)}{u_0}\right).$$

Thus, defining κ_0 by $u_0\xi(u_0) = (1+\kappa_0)\log y_0$, we obtain

$$\Lambda(z) = \Big(1+o(1)\Big)\frac{z_0}{z}\cdot y_0^{b(1+\kappa_0)/u_0 + o(b^2/u_0^2)},$$

where we have used (23) (in particular, the fact that $u_0 = o(\log y_0)$) to replace the error factor $e^{O(b^2+1)/u_0}$ by $(1+o(1))y_0^{o(b^2/u_0^2)}$.

Now, by the prime number theorem we have

$$\frac{y_0}{y} = \big(1+o(1)\big)\frac{z_0 \log y_0}{z \log y} = \big(1+o(1)\big)\frac{z_0}{z},$$

and by the definitions of u, u_0 and b, and the fact that $b = o(u_0)$, we have

$$\frac{y_0}{y} = x^{1/u_0 - 1/(u_0-b)} = y_0^{-b/(u_0-b)} = y_0^{-b/u_0-(1+o(1))b^2/u_0^2}.$$

Thus, we obtain

$$\begin{aligned}\Lambda(z) &= \big(1+o(1)\big)y_0^{b\kappa_0/u_0-(1+o(1))b^2/u_0^2}\\ &= \exp\left(\frac{b}{u_0}\kappa_0 \log y_0 - \big(1+o(1)\big)\frac{b^2}{u_0^2}\log y_0 + o(1)\right),\end{aligned}$$

and hence, noting that

$$c\sqrt{\log z_0} = \log\frac{z}{z_0} = \log\frac{y}{y_0} + o(1) = \big(1+o(1)\big)\frac{b}{u_0}\log y_0 + o(1),$$

and that $\kappa_0 = o(1)$, by (23) and (27), it follows that

$$\Lambda(z) = \exp\Big(\big(1+o(1)\big)\kappa_0 c\sqrt{\log y_0} - \big(1+o(1)\big)c^2 + o(1)\Big).$$

But $\Lambda(z)$ is maximized at $c = 0$, so this implies that $\kappa_0\sqrt{\log y_0} = o(1)$, which in turn implies that $\Lambda(z) = e^{-c^2} + o(1)$, as required. □

Note that as an immediate corollary of Lemma 3.11, we have

$$z_\pm = z_0 \exp\Big(\pm\big(\sqrt{\log(1/\delta)} + o(1)\big)\sqrt{\log z_0}\Big) \tag{37}$$

as well as

$$\Lambda(z) \geqslant \delta + o(1) \qquad \text{for all } z \in [z_-, z_+] \tag{38}$$

and

$$\Lambda(z_\pm) = \delta + o(1). \tag{39}$$

Finally, let us make a trivial observation.

Observation 3.12. $\sum_{j \geqslant z_-} A_{ij} \leqslant 2u_0$ *for every* $i \in [N]$.

Proof. If some number a_i is divisible by k distinct primes, each larger than z_-, then $z_-^k \leqslant x$. Since $z_- = z_0^{1+o(1)} = y_0^{1+o(1)} = x^{(1+o(1))/u_0}$, this implies that $k \leqslant (1+o(1))u_0$. □

4. **Probabilistic facts and preliminary results**

In this section we shall recall some standard probabilistic tools, define some events that will be important in later sections, and prove some basic facts about these events. Let us begin by defining the events that encode our induction hypothesis. Recall that $\eta < e^{-\gamma}$ was fixed in Section 2 and that ε_0 and $\varepsilon(k,z)$ were defined in (10) and Definition 2.5.

Definition 4.1. For each $z \in [z_-, \pi(x)]$, let $\mathcal{M}(z)$ denote the event that

$$\frac{m(z)}{z} \exp\left(-\operatorname{Ein}\left(\frac{m(z)}{z}\right)\right) \in (1 \pm \varepsilon_0)\eta\Lambda(z) \tag{40}$$

holds, and let $\mathcal{M}^*(z)$ denote the event that $\mathcal{M}(w)$ holds for every $w \in [z, \pi(x)]$. For each $z \in [z_-, z_+]$ and $k \geqslant 2$, let $\mathcal{T}_k(z)$ denote the event that

$$s_k(z) \in \big(1 \pm \varepsilon(k,z)\big)\hat{s}_k(z).$$

Note that Theorem 2.2 states that the event $\mathcal{M}^*(z_-)$ holds with high probability, and Theorem 2.6 states that with high probability the event $\mathcal{T}_k(z)$ holds for every $k \geqslant 2$ and every $z \in [z_-, z_+]$. We shall also need the following slightly more technical events.

Definition 4.2. For each $z \in [z_-, \pi(x)]$, let $\mathcal{Q}(z)$ denote the event that

- $\mathcal{M}(z)$ holds;
- $S(z) = \bigcup_{k \geqslant 2} S_k(z)$ contains no element w with $w > z_0^5$;
- $s_k(z) = 0$ for all $k \geqslant 4u_0$.

Let $\mathcal{K}(z)$ denote the event that $\mathcal{Q}(z)$ holds and also

$$\sum_{k \geqslant 2} 2^k s_k(z) \leqslant 2e^{C_0} m(z), \tag{41}$$

where C_0 was defined in (11).

We shall prove the following two lemmas, which allow us to deduce the technical events (which we shall need in the sections below) from our induction hypothesis.

Lemma 4.3. *With high probability, $\mathcal{Q}(z) \cup \mathcal{M}^*(z)^c$ holds for every $z \in [z_-, \pi(x)]$.*

In other words, the lemma above says that the probability that there exists $z \in [z_-, \pi(x)]$ such that $\mathcal{M}^*(z)$ holds but $\mathcal{Q}(z)$ does not is $o(1)$ as $x \to \infty$.

Lemma 4.4. *Let $z \in [z_-, z_+]$. If $\mathcal{Q}(z)$ holds and $\mathcal{T}_k(z)$ holds for all $k \geqslant 2$, then $\mathcal{K}(z)$ holds.*

Before proceeding to the proofs of Lemmas 4.3 and 4.4, let us give a simple but important application of the event $\mathcal{K}(z)$. Recall from Algorithm 2.7 that

$d(z)$ denotes the number of rows of $M(z)$ that contain a 1 in column z. The following lemma shows that the distribution of $d(z)$ is close to that of a Poisson random variable with mean $m(z)/z$.

LEMMA 4.5. *Let $z \in [z_-, z_+]$. If $\mathcal{K}(z)$ holds, then*

$$\mathbb{P}\big(d(z) = k \mid \mathcal{F}_z\big) = \frac{\big(1+o(1)\big)^k}{k!} e^{-m(z)/z} \left(\frac{m(z)}{z}\right)^k + \frac{O(1)}{z} \tag{42}$$

for every $k \geqslant 0$.

In the proof of Lemma 4.5 we shall use the following bound on the sum of independent Bernoulli random variables due to Le Cam [24].

LEMMA 4.6 (Le Cam, 1960). *Let $X_1, \dots, X_n$ be independent Bernoulli random variables, and let $X := \sum_{i=1}^n X_i$. Then*

$$\sum_{k \geqslant 0} \left| \mathbb{P}(X = k) - \frac{e^{-\mu}\mu^k}{k!} \right| \leqslant 2 \sum_{i=1}^n p_i^2,$$

where $p_i := \mathbb{P}(X_i = 1)$ and $\mu := \mathbb{E}[X] = \sum_{i=1}^n p_i$.

We shall also use the following fact on numerous occasions throughout the paper.

OBSERVATION 4.7. *If $\mathcal{M}(z)$ holds, then $m(z)/z \leqslant C_0 \Lambda(z) \leqslant C_0$.*

Proof. Recall from (8) and (9) that $\alpha(w)$ and $we^{-\mathrm{Ein}(w)}$ are strictly increasing functions, that $\alpha(w)$ is convex, and that $\alpha(w)e^{-\mathrm{Ein}(\alpha(w))} = w$. It follows that, if $\mathcal{M}(z)$ holds, then

$$\frac{m(z)}{z} \leqslant \alpha\big((1+\varepsilon_0)\eta\Lambda(z)\big) \leqslant \alpha\big((1+\varepsilon_0)\eta\big)\Lambda(z) = C_0\Lambda(z) \leqslant C_0,$$

as claimed, since $C_0 = \alpha\big((1+\varepsilon_0)\eta\big)$ and $\Lambda(z) \leqslant 1$. □

Proof of Lemma 4.5. Recall that

$$p_z = p_z(x) = \frac{\widetilde{\Psi}(x, q_z) - \widetilde{\Psi}(x, q_{z-1})}{\widetilde{\Psi}(x, q_z)}$$

denotes the probability that a uniformly chosen random number in $[x]$ is divisible by q_z to an odd power, conditioned on the event that all larger prime factors occur to an even power. We shall prove that the lemma holds even when, instead of conditioning on $\mathcal{F}_z$, we in fact condition on the entire matrix to the right of column z. In this case, $d(z)$ is a sum of independent Bernoulli random variables $\{X_i : i \in M(z)\}$ with $\mathbb{P}(X_i = 1) = p_z(xq_z^{-\alpha_i})$, where

$$q_z^{\alpha_i} = \prod_{w > z,\, A_{iw} = 1} q_w$$

is the product of the primes q_w greater than q_z that divide a_i an odd number of times.

However, the event $\mathcal{K}(z)$ implies that

$$\sum_{i\in M(z)} \alpha_i \leqslant \sum_{i\in M(z)} \sum_{w>z,\, A_{iw}=1} 6 = \sum_{k\geqslant 2} 6ks_k(z) = O(m(z)),$$

where the first step follows since $S(z)$ contains no prime q_w with $q_w \geqslant q_z^6$ (since $q_z^6 \geqslant z_-^6 \geqslant q_{z_0^5} = z_0^{5+o(1)}$), and the last follows from (41). By the pigeonhole principle, it follows that $\alpha_i = o(u_0)$ for all but a $o(1)$-proportion of the $i \in M(z)$. Now by Corollary 3.8, we have $\mathbb{P}(X_i = 1) = (1 + o(1))/z$ whenever $\alpha_i = o(u_0)$, and $\mathbb{P}(X_i = 1) \leqslant (1 + o(1))/z$ for every $i \in M(z)$. Thus

$$\sum_{i\in M(z)} \mathbb{P}(X_i = 1) = \big(1 + o(1)\big)\frac{m(z)}{z} \quad \text{and} \quad \sum_{i\in M(z)} \mathbb{P}(X_i = 1)^2 = \big(1 + o(1)\big)\frac{m(z)}{z^2}$$

whenever $\mathcal{K}(z)$ holds. As $m(z) = O(z)$ by Observation 4.7 (since $\mathcal{K}(z)$ implies $\mathcal{M}(z)$), it follows by Lemma 4.6 that

$$\sum_{k\geqslant 0} \left| \mathbb{P}\big(d(z) = k \mid \mathcal{F}_z\big) - \frac{e^{-\mu}\mu^k}{k!} \right| = \frac{O(1)}{z},$$

where $\mu := \mathbb{E}[d(z)] = (1 + o(1))m(z)/z$, as required. □

For the next result, it will be useful to have the following simple estimate on the function $\alpha(w)$ defined in (9).

Observation 4.8. *For $0 \leqslant w \leqslant 0.2$, we have*

$$w \leqslant \alpha(w) \leqslant w + 2w^2. \tag{43}$$

Proof. The lower bound follows from (9) as $we^{-\mathrm{Ein}(w)} \leqslant w$, and hence $\alpha(w) \geqslant w$. The upper bound also follows from (9) as $\mathrm{Ein}(w) \leqslant w$, and

$$\begin{aligned}(w + 2w^2)e^{-\mathrm{Ein}(w+2w^2)} &\geqslant (w + 2w^2)e^{-(w+2w^2)} \geqslant (w + 2w^2)\big(1 - (w + 2w^2)\big)\\ &= w + w^2(1 - 4w - 4w^2) \geqslant w,\end{aligned}$$

so $\alpha(w) \leqslant w + 2w^2$ for $0 \leqslant w \leqslant 0.2$. □

In Section 7 we shall also need the following weaker bounds on the distribution of $d(z)$ in the range $[z_+, \pi(x)]$.

Lemma 4.9. *Let $z \in [z_+, \pi(x)]$. If $\mathcal{K}(z)$ holds, then*

$$\mathbb{P}\big(d(z) \geqslant 1 \mid \mathcal{F}_z\big) \leqslant \big(1 + o(1)\big)m(z)p_z(x)$$

and, moreover,

$$\mathbb{P}\big(d(z) = k \mid \mathcal{F}_z\big) \leqslant \frac{(2\delta\eta)^{k-1}}{k!}\, \mathbb{P}\big(d(z) = 1 \mid \mathcal{F}_z\big)$$

for every $k \geqslant 2$.

Proof. Let the independent Bernoulli random variables $\{X_i : i \in M(z)\}$ be defined as in the proof of Lemma 4.5, and observe that, since $\alpha_i \in \{0\} \cup [1, \infty)$ for every $i \in M(z)$, we have

$$\mathbb{P}\big(X_i = 1\big) \leqslant \big(1 + o(1)\big)p_z(x) \leqslant \frac{1 + o(1)}{z}$$

for every $i \in M(z)$, by Corollary 3.8. The first claimed inequality now follows by the union bound. For the second bound, we note that in general if $p_i := \mathbb{P}(X_i = 1)$, then

$$\begin{aligned}
\frac{k\,\mathbb{P}\big(\sum_i X_i = k\big)}{\mathbb{P}\big(\sum_i X_i = 1\big)} &= \frac{\sum_{i_0 \in [m(z)]} \sum_{i_0 \in S \subseteq [m(z)],\, |S|=k} \prod_{i \in S} p_i \prod_{i \notin S}(1 - p_i)}{\sum_{i_0 \in [m(z)]} p_{i_0} \prod_{i \neq i_0}(1 - p_i)} \\
&\leqslant \max_{i_0} \sum_{S \subseteq [m(z)] \setminus \{i_0\},\, |S|=k-1} \prod_{i \in S} \frac{p_i}{1 - p_i} \\
&\leqslant \binom{m(z) - 1}{k - 1} \max_i \left(\frac{p_i}{1 - p_i}\right)^{k-1} \\
&\leqslant \frac{m(z)^{k-1}}{(k-1)!} \max_i \left(\frac{p_i}{1 - p_i}\right)^{k-1}.
\end{aligned}$$

Thus

$$\frac{\mathbb{P}\big(d(z) = k \mid \mathcal{F}_z\big)}{\mathbb{P}\big(d(z) = 1 \mid \mathcal{F}_z\big)} \leqslant \frac{1}{k!}\left((1 + o(1))\frac{m(z)}{z}\right)^{k-1}.$$

The result follows as $\mathcal{M}(z)$ and $z \in [z_+, \pi(x)]$ imply that

$$(1 + o(1))\frac{m(z)}{z} \leqslant (1 + o(1))\alpha\big((1 + \varepsilon_0)\eta\Lambda(z)\big) \leqslant \alpha(1.6\delta\eta) \leqslant 2\delta\eta$$

by (43) as $\Lambda(z) \leqslant \delta + o(1)$ and $\varepsilon_0 < e^{-\gamma} < 0.6$. □

4.1. *The proof of Lemma* 4.3. The first step is to control $m_0(z)$, the number of isolated vertices in the hypergraph $\mathcal{S}_A(z)$. As noted above, this simple fact lies at the heart of our proof and will be used several times in later sections.

LEMMA 4.10. *For each* $z \in [z_-, \pi(x)]$,

$$\mathbb{P}\Big(m_0(z) \notin \big(1 \pm \varepsilon_1\big)\eta\Lambda(z)z\Big) \leqslant \frac{1}{x^2}.$$

In particular, with high probability, $m_0(z) \in (1 \pm \varepsilon_1)\eta\Lambda(z)z$ *for every* $z \in [z_-, \pi(x)]$.

Note that since $m(z) \geqslant m_0(z)$, this implies, in particular, that with high probability we have $m(z) \geqslant (1 - \varepsilon_1)\eta\Lambda(z)z$ for every $z \in [z_-, \pi(x)]$. We remark (and also note for future reference) that the event $\mathcal{M}(z)$ implies deterministically that

$$m(z) \geqslant \alpha\big((1-\varepsilon_0)\eta\Lambda(z)\big)z \geqslant (1-\varepsilon_0)\eta\Lambda(z)z \geqslant (1-\varepsilon_0)\eta\Lambda(z_-)z_- \geqslant z_0^{1+o(1)} \tag{44}$$

for every $z \in [z_-, \pi(x)]$, by (14), (43), (22), and the fact that $\Lambda(z)z = J(x)\Psi(x, q_z)/x$ is increasing in z.

Lemma 4.10 is a straightforward consequence of the following special case of the well-known inequality of Chernoff [12].

LEMMA 4.11 (Chernoff's inequality). *Let $\varepsilon > 0$, let $X_1, \dots, X_n$ be independent Bernoulli random variables, and let $X := \sum_{i=1}^n X_i$. Then for any $\varepsilon > 0$,*

$$\mathbb{P}\big(|X - \mu| \geqslant \varepsilon\mu\big) \leqslant 2e^{-\varepsilon^2\mu/(2+\varepsilon)},$$

where $\mu := \mathbb{E}[X]$.

Proof of Lemma 4.10. The number of isolated vertices in $\mathcal{S}_A(z)$ is precisely the number of rows of A with no non-zero entry to the right of column z (all of which will lie in $M(z)$). This is also the same as the number of integers a_i, $i \in [N]$ such that every prime $q > q_z$ divides a_i an even number of times. Let μ denote the expected number of a_i with this property, and observe that

$$\mu = \eta J(x) \cdot \frac{\widetilde{\Psi}(x, q_z)}{x} = \big(1 + o(1)\big)\eta J(x) \cdot \frac{\Psi(x, q_z)}{x} = \big(1 + o(1)\big)\eta\Lambda(z)z$$

by Lemma 3.6 and the definition (13) of $\Lambda(z)$. Since the numbers a_i are independent, it follows by Lemma 4.11 that

$$\mathbb{P}\big(m_0(z) \notin \big(1 \pm \varepsilon_1\big)\eta\Lambda(z)z\big) \leqslant \mathbb{P}\big(m_0(z) \notin \big(1 \pm \varepsilon_1/2\big)\mu\big) \leqslant 2e^{-\varepsilon_1^2\mu/(8+2\varepsilon_1)}.$$

Now simply note that μ is increasing in z, so (approximating very crudely) we have $\mu \geqslant (1 + o(1))\eta\Lambda(z_-)z_- \geqslant \eta\delta z_-/2 \geqslant (\log x)^2$, and hence the right-hand side of the above inequality is at most $1/x^2$. The last part follows by the union bound over $z \in [z_-, \pi(x)]$. □

We shall next use the lower bound on $m(z)$ given by Lemma 4.10 to prove the following upper bounds on the random variables $s_k(z)$.

LEMMA 4.12. *With high probability, the following all hold*:

(a) $s_k(z) = 0$ *for every* $k \geqslant 2$ *and every* $z \geqslant z_0^5$.
(b) $s_2(z) + s_3(z) \leqslant z_0^{-1/2} m(z)$ *for every* $z \geqslant z_0^3$.
(c) $s_k(z) = 0$ *for every* $k \geqslant 4$ *and every* $z \geqslant z_0^3$.

Proof. Note that a prime q divides a uniformly chosen random element of $[x]$ with probability $\lfloor x/q \rfloor / x \leqslant 1/q$. Recall that A has N rows, and that $N = \eta J(x) = z_0^{2+o(1)}$, by (20). It follows that the expected number of primes $q \geqslant w$ that divide at least k of the integers a_i, $i \in [N]$, is at most

$$\sum_{q \geqslant w} \frac{N^k}{q^k} \leqslant \frac{z_0^{(2+o(1))k}}{w^{k-1}}. \tag{45}$$

Now if $s_k(z) > 0$, then there must be a prime $q > q_z$ that divides at least k of the a_i. Hence applying (45) with $k \geqslant 2$ and $w = q_{z_0^5} \geqslant z_0^5$ gives part (a), and applying it with $k \geqslant 4$ and $w = q_{z_0^3} \geqslant z_0^3$ gives part (c). To prove (b), observe first that with high probability

$$m(z) \geqslant m_0(z) \geqslant (1-\varepsilon_1)\eta\Lambda(z)z \geqslant (1-\varepsilon_1)\eta J(x) \cdot \frac{\Psi(x, q_{z_0^3})}{x} = z_0^{5/3+o(1)}$$

for every $z \geqslant z_0^3$. Indeed, the first inequality is trivial, the second follows by Lemma 4.10, the third since $\Psi(x,y)$ is increasing in y, and the fourth by (20) and (21). Hence, applying (45) with $k = 2$ and $w = q_{z_0^3} \geqslant z_0^3$, it follows that the expected number of primes that can contribute to the value of $s_2(z) + s_3(z)$ for any $z \geqslant z_0^3$ is at most $z_0^{1+o(1)} \leqslant z_0^{-2/3+o(1)} m(z)$. Part (b) then follows by Markov's inequality. □

We similarly obtain the following bound on $s_k(z)$ for large k. Note that $u_0 = (1+o(1)) \log z_0 / \log\log z_0$ by (23).

Lemma 4.13. *With high probability, for every $z \in [z_-, \pi(x)]$, either $\mathcal{M}^*(z)$ fails to hold, or $s_k(z) = 0$ for every $k \geqslant 4u_0$.*

Proof. Suppose that $\mathcal{M}^*(z)$ holds, and recall that this implies $m(w) \leqslant C_0 w$ for all $w \geqslant z$, by Observation 4.7. Any element $w \in \bigcup_{k \geqslant 4u_0} S_k(z)$ must have been "born" with $d(w) \geqslant 4u_0$. But, by Lemma 4.12, with high probability no such w exists in $[z_0^3 + 1, \pi(x)]$ as it would contribute to $s_k(w-1)$ for some $k \geqslant 4u_0$. Thus it is enough to show that with high probability no w exists in $[z_-, z_0^3]$ with $d(w) \geqslant 4u_0$ and $m(w) \leqslant C_0 w$.

The probability that $A_{iw} = 1$, conditioned on all entries of A to the right of column w, is always at most $(1+o(1))/w$, by Corollary 3.8, and is conditionally independent for each i. It follows that

$$\mathbb{P}\big(d(w) \geqslant k \mid \mathcal{F}_w\big) \leqslant \binom{m(w)}{k}\left(\frac{1+o(1)}{w}\right)^k \leqslant \frac{(2C_0)^k}{k!} \tag{46}$$

when $m(w) \leqslant C_0 w$. Thus the expected number of w in $[z_-, z_0^3]$ with $d(w) \geqslant 4u_0$ and $m(w) \leqslant C_0 w$ is at most

$$\frac{(2C_0)^{4u_0}}{(4u_0)!} \cdot z_0^3 \leqslant \exp\big(-4u_0 \log u_0 + O(u_0) + 3\log z_0\big) = o(1)$$

by Stirling's formula and the fact that $\log z_0 = (1+o(1)) u_0 \log u_0$, by (23). Hence, with high probability, no such w exists. □

Lemma 4.3 follows immediately from Lemmas 4.12 and 4.13. Note that if $w \in S_k(z)$, then $w \in S_{k'}(w-1)$ for some $k' \geqslant k$, so if $s_k(w) = 0$ for all $k \geqslant 2$, $w \geqslant z_0^5$, then $S_k(z)$ cannot contain any element $w > z_0^5$.

4.2. *The proof of Lemma* 4.4. We shall next prove that the event $\mathcal{Q}(z) \cap \bigcap_{k\geqslant 2} \mathcal{T}_k(z)$ implies (deterministically) that the event $\mathcal{K}(z)$ holds. To do so, we need to prove that $\sum_{k\geqslant 2} 2^k s_k(z) \leqslant 2e^{C_0} m(z)$. We begin with a simple but useful observation.

Observation 4.14. $(k+1)\hat{s}_{k+1}(z) \leqslant \dfrac{m(z)}{z} \cdot \hat{s}_k(z)$.

Proof. This follows easily from Definition 2.4. Indeed, writing $\lambda := m(z)/z$, we have

$$(k+1)\hat{s}_{k+1}(z) = \frac{m(z)}{k} e^{-\lambda} \sum_{\ell=k-1}^{\infty} \frac{\lambda^{\ell+1}}{(\ell+1)!} \leqslant \lambda \cdot \frac{m(z)}{k(k-1)} e^{-\lambda} \sum_{\ell=k-1}^{\infty} \frac{\lambda^{\ell}}{\ell!} = \lambda \cdot \hat{s}_k(z),$$

as claimed. □

We shall first prove the following bound on the sum over k of $2^k \varepsilon(k,z)\hat{s}_k(z)$. We shall need this bound again in Sections 9 and 10.

Lemma 4.15. *Let* $z \in [z_-, z_+]$. *If* $\mathcal{M}(z)$ *holds, then*

$$\sum_{k=2}^{\infty} 2^k \varepsilon(k,z)\hat{s}_k(z) \leqslant \varepsilon_1 m(z).$$

Proof. Note that

$$\sum_{k\geqslant 2} 2^k \varepsilon(k,z)\hat{s}_k(z) = \sum_{k\geqslant 2} \frac{2^k \varepsilon_1^k}{\Lambda(z)} k!\hat{s}_k(z) \leqslant \sum_{k\geqslant 2} \frac{2^k \varepsilon_1^k}{\Lambda(z)} \left(\frac{m(z)}{z}\right)^{k-2} 2\hat{s}_2(z) \tag{47}$$

by Definition 2.5 and Observation 4.14. Now

$$2\hat{s}_2(z) = m(z)\big(1 - e^{-m(z)/z}\big) \leqslant \frac{m(z)^2}{z} \leqslant C_0 \Lambda(z) m(z),$$

and $m(z) \leqslant C_0 z$, by (16) and Observation 4.7. Noting from (10) that $\varepsilon_1 C_0 < \varepsilon_1 e^{C_0} \leqslant 1/16$, it follows that the right-hand side of (47) is at most

$$\sum_{k\geqslant 2} 2^k \varepsilon_1^k C_0^{k-1} m(z) = \frac{4\varepsilon_1 C_0}{1 - 2\varepsilon_1 C_0} \varepsilon_1 m(z) \leqslant \varepsilon_1 m(z)$$

as required. □

Proof of Lemma 4.4. Since $\mathcal{T}_k(z)$ holds for all $k \geqslant 2$, we have

$$\sum_{k\geqslant 2} 2^k s_k(z) \leqslant \sum_{k=2}^{\infty} 2^k \hat{s}_k(z) + \sum_{k=2}^{\infty} 2^k \varepsilon(k,z)\hat{s}_k(z).$$

The second term is at most $\varepsilon_1 m(z)$, by Lemma 4.15, and since $\mathcal{Q}(z)$ implies that the event $\mathcal{M}(z)$ holds. To bound the first term, observe that, writing

$\lambda = m(z)/z$, we have

(48)
$$\sum_{k\geqslant 2} 2^k \hat{s}_k(z) = m(z)e^{-\lambda} \sum_{k\geqslant 2} \frac{2^k}{k(k-1)} \sum_{\ell=k-1}^{\infty} \frac{\lambda^\ell}{\ell!} = m(z)e^{-\lambda} \sum_{\ell=1}^{\infty} \frac{\lambda^\ell}{\ell!} \sum_{k=2}^{\ell+1} \frac{2^k}{k(k-1)}$$
$$\leqslant m(z)e^{-\lambda} \sum_{\ell=1}^{\infty} \frac{\lambda^\ell}{\ell!} \cdot 2^\ell \leqslant m(z)e^{-\lambda}e^{2\lambda} = e^{m(z)/z} m(z) \leqslant e^{C_0} m(z),$$

as required, where in the final step we used Observation 4.7. □

Finally, let us make a simple observation that will play an important role in Section 8.

Observation 4.16. *If $\mathcal{M}(z)$ holds, then $\varepsilon(k,z)\hat{s}_k(z) = z_0^{1+o(1)}$ uniformly for every $z \in [z_-, z_+]$ and $2 \leqslant k \leqslant 4u_0$.*

Proof. Let $z \in [z_-, z_+]$, and observe that, since $\mathcal{M}(z)$ implies $m(z)/z = \Theta(1)$, we have

$$\varepsilon(k,z)\hat{s}_k(z) = \frac{\varepsilon_1^k \cdot k!}{\Lambda(z)} \cdot \frac{m(z)}{k(k-1)} e^{-m(z)/z} \sum_{\ell=k-1}^{\infty} \frac{1}{\ell!} \left(\frac{m(z)}{z}\right)^\ell = e^{O(k)} z. \tag{49}$$

Since $u_0 = o(\log z_0)$, by (23), and $z = z_0^{1+o(1)}$ for every $z \in [z_-, z_+]$, by (22), it follows that $\varepsilon(k,z)\hat{s}_k(z) = z_0^{1+o(1)}$ uniformly for every $z \in [z_-, z_+]$ and $k \leqslant 4u_0$, as claimed. □

4.3. *Martingales.* We finish this section by recalling some standard results about martingales that we shall use in later sections. Recall that a *supermartingale* with respect to a filtration $(\mathcal{F}_t)_{t\geqslant 0}$ of σ-algebras is a sequence of random variables $(X_t)_{t\geqslant 0}$ such that the following hold for each $t \geqslant 0$: X_t is $\mathcal{F}_t$-measurable, $\mathbb{E}[|X_t|] < \infty$, and $\mathbb{E}\big[X_{t+1} \mid \mathcal{F}_t\big] \leqslant X_t$.

The following inequality was proved by Azuma [1] and Hoeffding [20] (see, e.g., [8]).

Lemma 4.17 (The Azuma–Hoeffding inequality). *Let $(X_t)_{t=0}^{\ell}$ be a supermartingale with respect to a filtration $(\mathcal{F}_t)_{t=0}^{\ell}$, and assume $\mathbb{P}(|X_t - X_{t-1}| > c_t) = 0$ for $t = 1, \ldots, \ell$. Then*

$$\mathbb{P}\big(X_\ell - X_0 \geqslant a\big) \leqslant \exp\left(\frac{-a^2}{2\sum_{t=1}^{\ell} c_t^2}\right)$$

for every $a > 0$.

Recall that a *stopping time* T with respect to the filtration $(\mathcal{F}_t)_{t\geqslant 0}$ is a non-negative integer valued random variable such that the event $\{T \leqslant t\}$ is $\mathcal{F}_t$-measurable. A stopping time T is called *bounded* if there exists a deterministic

$C > 0$ such that $\mathbb{P}(T \leqslant C) = 1$. We shall require the following well-known theorem (see, for example, [34]).

LEMMA 4.18 (Optional Stopping Theorem). *Suppose $(X_t)_{t \geqslant 0}$ is a supermartingale and T is a bounded stopping time. Then $\mathbb{E}[X_T] \leqslant \mathbb{E}[X_0]$.*

5. Approximation by an independent hypergraph model

In order to control the evolution of the variables $m(z)$ and $s_k(z)$ as z decreases, we shall need to understand the structure of the 2-core $\mathcal{C}_A(z)$ of $\mathcal{H}_A(z)$, conditioned on $\mathcal{F}_z^+$. More precisely, we shall need to prove good approximations for the probability that certain substructures occur in A, conditioned on the column sums (over the rows $M(z)$) of the columns $[z, \pi(x)]$. We shall show that, up to relatively small error, these probabilities are the same as they would be if the columns were independent.

Note that without any conditioning the rows are independent, and the entries in the rows are almost independent, meaning that we can estimate the probability of given hypergraph structures using (6). The problem is that we wish to condition on an event $\mathcal{E} \in \mathcal{F}_z^+$ of the form

$$\mathcal{E} := \left\{ M(z) = M \text{ and } \sum_{i \in M(z)} A_{ij} = d_j, \text{ for } j \geqslant z \right\}, \tag{50}$$

which specifies $M(z)$ and all the degrees of vertices in $\mathcal{C}_A(z)$ as well as $d(z)$. Thus both rows and columns are now dependent. If the entries in each row of A were independent, then we could just forget the probabilities $p_j(x)$ and model the matrix as placing d_j 1s in column j uniformly at random. Unfortunately this is not the case, so we need to be a bit more careful.

The substructures that we shall need to consider involve a (typically small) subset $I \subseteq M(z)$ of rows of A, and we shall need to estimate the probability that the entries in these rows are of a given form. The most general version of this requirement is that the submatrix obtained by just considering the set I of rows and a (usually larger) set $C \subseteq [z, \pi(x)]$ of columns forms a specific $I \times C$ matrix R. This corresponds to specifying exactly which vertices of C lie in a fixed set I of edges in $\mathcal{C}_A(z)$.

In order to state the main result of this section (Theorem 5.1, below), we shall need some notation. Given any matrix B and subsets I and C of the rows and columns of B respectively, define $B[I \times C]$ to be the submatrix of B given by the set I of rows and the set C of columns. Given $z \in [z_-, \pi(x)]$ and an event $\mathcal{E} \in \mathcal{F}_z^+$ of the form (50), which determines the set $M(z) = M$ and the sequence $(d_j)_{j \geqslant z}$, let $\tilde{A}_{\mathcal{E}}$ denote the random $M \times [z, \pi(x)]$ matrix obtained by choosing (for each j) column j uniformly among the $\binom{|M|}{d_j}$ possible choices of column with column sum d_j, independently for each column. We shall

write $m := |M|$ and $A_M := A\big[M \times [z, \pi(x)]\big]$ for the corresponding submatrix of A, even when $\mathcal{E}$ does not hold. Of course if $\mathcal{E}$ holds, then $m(z) = m$ and $M(z) = M$.

Assume $I \subseteq M$ and $C \subseteq [z, \pi(x)]$. If R is an $I \times C$ 0-1 matrix, define

$$|R|_1 = \sum_{i \in I} \sum_{j \in C} R_{ij} \qquad \text{and} \qquad |R|_2 = \sum_{i \in I} \bigg(\sum_{j \in C} R_{ij} \bigg)^2.$$

We shall prove the following theorem, which allows us to control the (conditional) probability of each "basic event" of the form $\{A[I \times C] = R\}$ in terms of the probability of the corresponding event $\{\tilde{A}_{\mathcal{E}}[I \times C] = R\}$ in the independent model $\tilde{A}_{\mathcal{E}}$.

THEOREM 5.1. *Let $z \in [z_-, \pi(x)]$, and let $\mathcal{E} \in \mathcal{F}_z^+$ be an event of the form* (50) *such that $\mathcal{K}(z)$ holds and $d(z) \leqslant 4u_0$. Let $I \subseteq M$, $C \subseteq [z, \pi(x)]$ with $|I| = e^{O(u_0)}$, and let R be an $I \times C$ 0-1 matrix with $|R|_1 = O(|M|/u_0)$. Then*

$$\mathbb{P}\big(A[I \times C] = R \mid \mathcal{E}\big) \leqslant \exp\bigg(\frac{O\big(|I| + |R|_1\big)}{u_0}\bigg) \mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times C] = R\big). \tag{51}$$

Moreover, if every row sum of R is at most $u_0/150$, then

$$\mathbb{P}\big(A[I \times C] = R \mid \mathcal{E}\big) = \exp\bigg(\frac{O\big(|I| + |R|_2\big)}{u_0}\bigg) \mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times C] = R\big). \tag{52}$$

For each $I \subseteq M$ and $C \subseteq [z, \pi(x)]$, let us define $\mathcal{R}_{\mathcal{E}}(I, C)$ to be the collection of $I \times C$ 0-1 matrices R whose jth column sum does not exceed d_j for each $j \in C$. Note that matrices that do not have this property are inconsistent with the event $\mathcal{E}$, and so the theorem holds trivially for such matrices, since all the probabilities are then zero.

If $\mathcal{K}(z)$ holds, then $d_j = 0$ for $j > z_0^5$, so if $R \in \mathcal{R}_{\mathcal{E}}(I, C)$, then the probabilities in (51) and (52) are unaffected if we replace C by $C \cap [z, z_0^5]$. Thus we may assume without loss of generality that $C \subseteq [z, z_0^5]$. Similarly, we may assume without loss of generality that $z \leqslant z_0^5$. Thus, for convenience, let us fix for the rest of the section an integer $z \in [z_-, z_0^5]$ and an event $\mathcal{E} \in \mathcal{F}_z^+$ of the form (50) such that $\mathcal{K}(z)$ holds and $d(z) \leqslant 4u_0$. We note that the conditions $\mathcal{K}(z)$ and $d(z) \leqslant 4u_0$ also imply that

$$\sum_{j=z}^{\pi(x)} d_j = d(z) + \sum_{k \geqslant 2} k s_k(z) = O(m(z)), \tag{53}$$

and that $m(z) \geqslant z_0^{1+o(1)}$, by (44).

The idea of the proof is to randomly permute some of the entries of A in $M \times \{j\}$ for each $j \in C$ so as to obtain a random submatrix in $M \times C$ that

is zero in $I \times C$ and to show that in most cases the expected probability of obtaining this permuted matrix is roughly the same as obtaining the original. This will not always be the case, but such exceptional cases occur rarely, and the probabilities in these cases are smaller than normal, so they have little effect on the probability of seeing a particular pattern in the entries $I \times C$. This allows us to reduce the general case to the case $R = 0$. Moreover, by summing over R we can reduce the trivial "true" event to the case $R = 0$, thus indirectly estimating the probability when $R = 0$. Lemmas 5.6 and 5.8 below will prove (51) and (52) respectively, subject to the result holding for $R = 0$. Lemma 5.9 will then deal with the case when $R = 0$. One reason for splitting the result into three lemmas is that the proof of the lower bound in Lemma 5.8 actually relies heavily on the upper bound from Lemma 5.6, and the proof of Lemma 5.9 also relies heavily on Lemma 5.8.

Given an $M \times [z, \pi(x)]$ matrix B, we say B is *consistent with* $\mathcal{E}$ if the column sums $\sum_{i \in M} B_{ij}$ are equal to d_j for all $j \geqslant z$. Let $\mathcal{B}$ be the set of all $M \times [z, \pi(x)]$ 0-1 matrices that are consistent with $\mathcal{E}$, and for $R \in \mathcal{R}_{\mathcal{E}}(I, C)$, define

$$\mathcal{B}_R := \{B \in \mathcal{B} : B[I \times C] = R\} \text{ and } \mathcal{B}_0 := \{B \in \mathcal{B} : B[I \times C] = 0\}.$$

We will use the following simple observation several times in the proof below.

Observation 5.2. $\mathbb{P}(A_M = B \mid \mathcal{E}) = \mathbb{P}(A_M = B \mid A_M \in \mathcal{B})$ *for every* $B \in \mathcal{B}$, *and hence*

$$\mathbb{P}\big(A[I \times C] = R \mid \mathcal{E}\big) = \mathbb{P}\big(A_M \in \mathcal{B}_R \mid A_M \in \mathcal{B}\big)$$

for every $R \in \mathcal{R}_{\mathcal{E}}(I, C)$.

Proof. Note that the event $\mathcal{E}$ is equal to $\{M(z) = M\} \cap \{A_M \in \mathcal{B}\}$. We claim that, conditional on the event $A_M \in \mathcal{B}$, the event $M(z) = M$ depends only on rows outside of M and so is independent of the choice of $A_M \in \mathcal{B}$. Indeed, since every d_j (for $j > z$) is either zero or at least 2, none of the rows of M will be deleted by the algorithm that constructs the 2-core $\mathcal{C}_A(z)$. (To see this, consider the first such row to be deleted.) Thus the event $M(z) = M$ holds (conditional on $A_M \in \mathcal{B}$) if and only if all other rows are removed by step z of the algorithm, which depends on the sequence $(d_{z+1}, \ldots, d_{\pi(x)})$ and on the rows outside M, but not on the choice of $A_M \in \mathcal{B}$, as claimed. This proves the first statement, and the second follows immediately since $\{A[I \times C] = R\} \cap \mathcal{B} = \mathcal{B}_R$. □

Given $B \in \mathcal{B}$, define a random matrix $\phi_{I,C}(B)$ as follows. For each $j \in C$, remove all 1s in the submatrix $B[I \times \{j\}]$ and place them on a uniformly chosen random subset of the zero entries in $B[(M \setminus I) \times \{j\}]$, choosing the random subsets independently for each $j \in C$. The result is a matrix $\phi_{I,C}(B)$

with the same column sums, and hence $\phi_{I,C}(B)$ is still consistent with $\mathcal{E}$, but $\phi_{I,C}(B)[I \times C] = 0$, so $\phi_{I,C}(B) \in \mathcal{B}_0$. The choices made in the construction of $\phi_{I,C}(B)$ will always be assumed to be independent of any random choice of B, or the matrix A. The following observation is then immediate.

Observation 5.3. *For any fixed $R \in \mathcal{R}_{\mathcal{E}}(I, C)$, if B is chosen uniformly at random from $\mathcal{B}_R$, then the distribution of $\phi_{I,C}(B)$ is uniform on $\mathcal{B}_0$.*

Proof. Indeed, the distribution is invariant under any permutation of the rows $M \setminus I$. □

Now, for any 0-1 matrix B and each row i of B, let

$$t_i(B) = \prod_{j \in [z, \pi(x)]} q_j^{B_{ij}},$$

and define the *weight* of the ith row, $w_i(B)$, by

$$t_i(B) = q_{z-1}^{w_i(B)}.$$

For completeness, define $w_i(B) = 0$ if i is not a row of B.

Observation 5.4. *If $B_{ij} = 0$ for all $j \notin [z, z_0^5]$, then $\sum_j B_{ij} \leqslant w_i(B) \leqslant 6 \sum_j B_{ij}$.*

Proof. All primes dividing $t_i(B)$ are in the range $[q_z, q_{z_0^5}]$, and $q_{z_0^5} = z_0^{5+o(1)} \leqslant z_-^6 \leqslant q_{z-1}^6$. □

Assume $B \in \mathcal{B}_R$. Let B^- be the matrix obtained from B by setting all entries in $I \times C$ to 0. Note that B^- is not in general consistent with $\mathcal{E}$. Write

$$\delta_i := w_i(B) - w_i(B^-), \qquad \text{and} \qquad \delta_i^\phi := w_i(\phi_{I,C}(B)) - w_i(B^-).$$

Note that $\delta_i = w_i(R)$ depends only on R and that $\delta_i = 0$ for $i \notin I$ while $\delta_i^\phi = 0$ for $i \in I$. Moreover, by Observation 5.4, we have $\delta_i \in \{0\} \cup [1, \infty)$ for every $i \in M$, and δ_i (respectively δ_i^ϕ) is, up to a constant factor, equal to the number of 1s removed from (respectively added to) row i by the map $\phi_{I,C}$. Define

$$\Delta := \sum_{i \in I} \delta_i = \sum_{i \in M \setminus I} \delta_i^\phi,$$

and note that $\Delta = \Theta(|R|_1)$ and Δ depends only on R. Also observe that, by (53) and Observation 5.4,

$$\sum_{i \in M} w_i(B) = O(|M|) \tag{54}$$

for any $B \in \mathcal{B}_R$. We shall also write

$$\alpha_R(B) := \sum_{i \in I} \delta_i w_i(B^-). \tag{55}$$

Our first challenge will be to prove the following lemma.

LEMMA 5.5. *Let $z \in [z_-, z_0^5]$, and let $\mathcal{E} \in \mathcal{F}_z^+$ be an event of the form* (50) *such that $\mathcal{K}(z)$ holds and $d(z) \leqslant 4u_0$. Let $I \subseteq M$ and $C \subseteq [z, z_0^5]$ be such that $|I| = e^{O(u_0)}$, and let $R \in \mathcal{R}_{\mathcal{E}}(I, C)$. If $|R|_1 = O(|M|/u_0)$, then for all $B \in \mathcal{B}_R$,*

$$\mathbb{P}\big(A_M = B\big) \leqslant \exp\left(\frac{O\big(|R|_1\big)}{u_0}\right)\mathbb{P}\big(A_M = \phi_{I,C}(B)\big). \tag{56}$$

Moreover, if no row $i \in I$ of B contains more than $u/12$ 1s, where $q_{z-1} = x^{1/u}$, then

$$\mathbb{P}\big(A_M = B\big) = \exp\left(\frac{O\big(\alpha_R(B) + |R|_2\big)}{u_0}\right)\mathbb{P}\big(A_M = \phi_{I,C}(B)\big). \tag{57}$$

Note that the probabilities here are *unconditional* and are both over the choice of A and the (uniform and independent) choice of $\phi_{I,C}(B)$.

Proof. For any $B \in \mathcal{B}_R$, there are exactly $\widetilde{\Psi}(x/t_i(B), q_{z-1})$ choices of integer $a_i \in [x]$ such that $A\big[\{i\} \times [z, \pi(x)]\big] = B\big[\{i\} \times [z, \pi(x)]\big]$. Thus, by counting the number of choices for a_i, $i \in M$, we have

$$\frac{\mathbb{P}(A_M = B)}{\mathbb{P}(A_M = B^-)} = \prod_{i \in M} \frac{\widetilde{\Psi}\big(x/t_i(B), q_{z-1}\big)}{\widetilde{\Psi}\big(x/t_i(B^-), q_{z-1}\big)} = \prod_{i\colon \delta_i > 0} \frac{\widetilde{\Psi}\big(xq_{z-1}^{-w_i(B^-)-\delta_i}, q_{z-1}\big)}{\widetilde{\Psi}\big(xq_{z-1}^{-w_i(B^-)}, q_{z-1}\big)}. \tag{58}$$

Thus, by Theorem 3.7, and recalling that $\delta_i \in \{0\} \cup [1, \infty)$ for every $i \in M$, we have

$$\begin{aligned} \log \frac{\mathbb{P}(A_M = B)}{\mathbb{P}(A_M = B^-)} &\leqslant \sum_{i\colon \delta_i > 0} \left(\delta_i\big(\xi(u) - \log q_{z-1}\big) + \frac{O(1)}{u}\right) \\ &= \Delta\big(\xi(u) - \log q_{z-1}\big) + \frac{O(|R|_1)}{u_0}. \end{aligned} \tag{59}$$

The error bound in the last line follows as there are clearly at most $|R|_1$ rows where B and B^- differ, and since $u = \Theta(u_0)$, which holds because $z \in [z_-, z_0^5]$.

Similarly, conditioned on the choice of $\phi_{I,C}(B)$,

$$\begin{aligned} \log \frac{\mathbb{P}(A_M = \phi_{I,C}(B) \mid \phi_{I,C}(B))}{\mathbb{P}(A_M = B^-)} &\leqslant \sum_{i\colon \delta_i^\phi > 0} \left(\delta_i^\phi\big(\xi(u) - \log q_{z-1}\big) + \frac{O(1)}{u}\right) \\ &= \Delta\big(\xi(u) - \log q_{z-1}\big) + \frac{O(|R|_1)}{u_0}. \end{aligned} \tag{60}$$

Again, the error bound follows as there are at most $|R|_1$ rows where $\phi_{I,C}(B)$ and B^- differ.

We now aim to deduce corresponding lower bounds from Theorem 3.7. To do so, note first that if no row $i \in I$ of B contains more than $u/12$ 1s, then, by Observation 5.4,

$$w_i(B^-) + \delta_i = w_i(B) \leqslant \frac{u}{2}.$$

Hence, by (58) and Theorem 3.7, we have

(61)

$$\begin{aligned}\log \frac{\mathbb{P}(A_M = B)}{\mathbb{P}(A_M = B^-)} &= \log \prod_{i \in M} \frac{\widetilde{\Psi}\big(x, q_{z-1}^{-w_i(B^-)-\delta_i}, q_{z-1}\big)}{\widetilde{\Psi}\big(x, q_{z-1}^{-w_i(B^-)}, q_{z-1}\big)} \\ &= \sum_{i\colon \delta_i > 0} \left(\delta_i(\xi(u) - \log q_{z-1}) + \frac{O\big(\delta_i^2 + \delta_i w_i(B^-) + 1\big)}{u_0} \right) \\ &= \Delta(\xi(u) - \log q_{z-1}) + \frac{O(\alpha_R(B) + |R|_2)}{u_0}.\end{aligned}$$

Note that the last equality follows by Observation 5.4, since

$$\sum_{i\colon \delta_i > 0} (\delta_i^2 + \delta_i w_i(B^-) + 1) \leqslant 6^2 |R|_2 + \alpha_R(B) + |R|_1$$

and $|R|_1 \leqslant |R|_2$.

Proving a similar bound for $\phi_{I,C}(B)$ (and without assuming that the rows of B have few 1s) is a little more complicated, since $w_i(B^-) + \delta_i^\phi \leqslant u/2$ does not necessarily hold for all $\phi_{I,C}(B)$. To get around this problem, we use the following event (which depends on the choice of $\phi_{I,C}(B)$):

$$\mathcal{G}_B := \Big\{ w_i(B^-) \leqslant u/3 \text{ and } \delta_i^\phi \leqslant 6 \text{ for each } i \in M \text{ such that } \delta_i^\phi > 0 \Big\}.$$

The first step is to show that this event occurs for most choices of $\phi_{I,C}(B)$.

Claim 1: For every $B \in \mathcal{B}_R$, we have

$$\mathbb{P}(\mathcal{G}_B) = \exp\left(-\frac{O(|R|_1)}{u_0} \right).$$

Proof of Claim 1. Let $M' = \{i \in M : w_i(B^-) \leqslant u/3\}$, and note that if $\phi_{I,C}(B)$ places 1s only in rows of M', and places no more than a single 1 in each row, then $\mathcal{G}_B$ holds. Indeed, only those rows i where a 1 is inserted have $\delta_i^\phi > 0$, and if only a single 1 is inserted in row i, then by Observation 5.4 we have $\delta_i^\phi \leqslant 6$.

We can construct $\phi_{I,C}(B)$ with the correct distribution by processing each one of the $|R|_1$ 1 entries of $R = B[I \times C]$ in turn, removing it from B and then adding a 1 to a uniformly chosen non-zero entry of the same column of B, outside of the rows I. The number of possible choices in this process is clearly at most $|M|^{|R|_1}$. If instead we consider only those choices where 1s are placed only in the rows $M' \setminus I$, and no two 1s are placed in the same row, the

number of choices will still be at least $(|M'| - |I| - 4u_0 - |R|_1)^{|R|_1}$. Indeed, at each step we have $|M' \setminus I|$ choices of row, but we must avoid the at most $4u_0$ original 1s of B in that column and the at most $|R|_1$ rows where we have already placed a 1. Since the choice of $\phi_{I,C}(B)$ is uniform, we have

$$\mathbb{P}(\mathcal{G}_B) \geqslant \left(\frac{|M'| - |I| - 4u_0 - |R|_1}{|M|}\right)^{|R|_1} \geqslant \left(1 - \frac{O(1)}{u_0}\right)^{|R|_1} = e^{-O(|R|_1/u_0)},$$

since $|I| + 4u_0 + |R|_1 = O(|M|/u_0)$, by our assumptions and using (23) and (44), and since

$$|M \setminus M'| \leqslant \frac{3}{u} \sum_{i \in M \setminus M'} w_i(B^-) \leqslant \frac{3}{u} \sum_{i \in M} w_i(B) = \frac{O(|M|)}{u_0}$$

by (54) and since $u = \Theta(u_0)$. □

We are now ready to prove our final lower bound; the lemma follows easily from the following claim, together with (59), (60) and (61).

Claim 2: For every $B \in \mathcal{B}_R$, we have

$$\log \frac{\mathbb{P}\big(A_M = \phi_{I,C}(B)\big)}{\mathbb{P}\big(A_M = B^-\big)} \geqslant \Delta\big(\xi(u) - \log q_{z-1}\big) + \frac{O(|R|_1)}{u_0}.$$

Proof of Claim 2. Note first that

$$\frac{\mathbb{P}\big(A_M = \phi_{I,C}(B)\big)}{\mathbb{P}\big(A_M = B^-\big)} \geqslant \mathbb{P}\big(\mathcal{G}_B\big)\, \mathbb{E}\left[\frac{\mathbb{P}\big(A_M = \phi_{I,C}(B) \mid \phi_{I,C}(B)\big)}{\mathbb{P}\big(A_M = B^-\big)} \;\middle|\; \mathcal{G}_B\right],$$

and observe that if $\mathcal{G}_B$ holds, then $w_i(B^-) + \delta_i^\phi \leqslant u/3 + 6 \leqslant u/2$. Hence, by Theorem 3.7, it follows (cf. (61)) that

$$\begin{aligned}
&\log \frac{\mathbb{P}\big(A_M = \phi_{I,C}(B) \mid \phi_{I,C}(B)\big)}{\mathbb{P}\big(A_M = B^-\big)} \\
&\quad = \log \prod_{i \in M} \frac{\widetilde{\Psi}\big(x, q_{z-1}^{-w_i(B^-) - \delta_i^\phi}, q_{z-1}\big)}{\widetilde{\Psi}\big(x, q_{z-1}^{-w_i(B^-)}, q_{z-1}\big)} \\
&\quad = \sum_{i:\, \delta_i^\phi > 0} \left(\delta_i^\phi(\xi(u) - \log q_{z-1}) + \frac{O(w_i(B^-) + 1)}{u_0}\right) \\
&\quad = \Delta\big(\xi(u) - \log q_{z-1}\big) + \frac{O(\alpha_\phi + |R|_1)}{u_0}
\end{aligned}$$

whenever $\mathcal{G}_B$ holds, where

$$\alpha_\phi := \sum_{\delta_i^\phi > 0} w_i(B^-).$$

Thus, by the convexity of the exponential function,

$$\begin{aligned}(62)\qquad &\frac{\mathbb{P}\big(A_M = \phi_{I,C}(B)\big)}{\mathbb{P}\big(A_M = B^-\big)} \\ &\geqslant \mathbb{P}\big(\mathcal{G}_B\big)\mathbb{E}\bigg[\exp\bigg(\Delta(\xi(u) - \log q_{z-1}) + \frac{O(\alpha_\phi + |R|_1)}{u_0}\bigg) \;\bigg|\; \mathcal{G}_B\bigg] \\ &\geqslant \mathbb{P}\big(\mathcal{G}_B\big)\exp\bigg(\mathbb{E}\bigg[\Delta(\xi(u) - \log q_{z-1}) + \frac{O(\alpha_\phi + |R|_1)}{u_0} \;\bigg|\; \mathcal{G}_B\bigg]\bigg).\end{aligned}$$

We claim that

$$\mathbb{E}\big[\alpha_\phi \mid \mathcal{G}_B\big] = O(|R|_1). \tag{63}$$

Indeed, as in the proof of Claim 1, the probability that a given 1 entry in R is moved to row i is at most $\big(|M'| - |I| - 4u_0 - |R|_1\big)^{-1} = O(1/|M|)$ for each $i \in M$. Thus the probability that $\delta_i^\phi > 0$ (i.e., that row i receives some 1) is at most $O(|R|_1/|M|)$, and so

$$\mathbb{E}\big[\alpha_\phi \mid \mathcal{G}\big] = \frac{O(|R|_1)}{|M|}\sum_{i\in M} w_i(B^-) = O(|R|_1),$$

by (54), as claimed.

Hence, combining (62) and (63), and using Claim 1, we obtain

$$\frac{\mathbb{P}\big(A_M = \phi_{I,C}(B)\big)}{\mathbb{P}\big(A_M = B^-\big)} \geqslant \exp\bigg(\Delta\big(\xi(u) - \log q_{z-1}\big) + \frac{O(|R|_1)}{u_0}\bigg),$$

as required. □

To complete the proof, simply observe that combining Claim 2 with (59) gives (56), and that combining Claim 2 with (60) and (61) gives (57). □

We can now easily deduce the following lemma.

Lemma 5.6. *Let $z \in [z_-, z_0^5]$, and let $\mathcal{E} \in \mathcal{F}_z^+$ be an event of the form* (50) *such that $\mathcal{K}(z)$ holds and $d(z) \leqslant 4u_0$. Let $I \subseteq M$ and $C \subseteq [z, z_0^5]$ be such that $|I| = e^{O(u_0)}$, and let $R \in \mathcal{R}_{\mathcal{E}}(I, C)$. If $|R|_1 = O(|M|/u_0)$, then*

$$\frac{\mathbb{P}\big(A[I\times C] = R \mid \mathcal{E}\big)}{\mathbb{P}\big(A[I\times C] = 0 \mid \mathcal{E}\big)} \leqslant \exp\bigg(\frac{O(|R|_1)}{u_0}\bigg)\frac{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I\times C] = R\big)}{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I\times C] = 0\big)}. \tag{64}$$

Proof. Write $\mathbb{E}_{B\in\mathcal{B}_R}$ for the expectation over a uniform random choice of $B \in \mathcal{B}_R$, and similarly for $\mathbb{E}_{B\in\mathcal{B}_0}$. Note that, by Observation 5.2,

$$\mathbb{P}\big(A[I\times C] = R \mid \mathcal{E}\big) = \mathbb{P}\big(A_M \in \mathcal{B}_R \mid A_M \in \mathcal{B}\big) = \frac{|\mathcal{B}_R|\,\mathbb{E}_{B\in\mathcal{B}_R}\mathbb{P}\big(A_M = B\big)}{\mathbb{P}\big(A_M \in \mathcal{B}\big)},$$

and that by Lemma 5.5 and Observation 5.3,

$$\begin{aligned}\mathbb{E}_{B\in\mathcal{B}_R}\mathbb{P}\big(A_M = B\big) &\leqslant \exp\left(\frac{O\big(|R|_1\big)}{u_0}\right)\mathbb{E}_{B\in\mathcal{B}_R}\mathbb{P}\big(A_M = \phi_{I,C}(B)\big)\\ &= \exp\left(\frac{O\big(|R|_1\big)}{u_0}\right)\mathbb{E}_{B\in\mathcal{B}_0}\mathbb{P}\big(A_M = B\big).\end{aligned}$$

Thus, using Observation 5.2 again, it follows that

$$\mathbb{P}\big(A[I\times C] = R \mid \mathcal{E}\big) \leqslant \exp\left(\frac{O\big(|R|_1\big)}{u_0}\right)\frac{|\mathcal{B}_R|}{|\mathcal{B}_0|}\mathbb{P}\big(A[I\times C] = 0 \mid \mathcal{E}\big).$$

Now, since $A_{\mathcal{E}}$ is distributed uniformly on $\mathcal{B}$, we have

$$\frac{\mathbb{P}\big(A_{\mathcal{E}}[I\times C] = R\big)}{\mathbb{P}\big(A_{\mathcal{E}}[I\times C] = 0\big)} = \frac{|\mathcal{B}_R|}{|\mathcal{B}_0|},$$

and so the lemma follows. □

Our next task, which will be somewhat harder, is to prove an almost-matching lower bound when the row sums of R are not too large. In order to do so we will use the following simple observation, which will also be useful in Section 6.

Observation 5.7. *Let $X_1,\dots,X_n$ be independent Bernoulli random variables, and let $X = \sum_{i=1}^n X_i$. Then for any $\lambda > 0$,*

$$\mathbb{E}\big[e^{\lambda X}\big] \leqslant \exp\big((e^\lambda - 1)\mathbb{E}[X]\big).$$

Proof. We have, for each Bernoulli random variable X_i,

$$\mathbb{E}\big[e^{\lambda X_i}\big] = 1 + (e^\lambda - 1)\mathbb{P}(X_i = 1) \leqslant \exp\big((e^\lambda - 1)\mathbb{E}[X_i]\big).$$

Thus by independence of the X_i,

$$\mathbb{E}\big[e^{\lambda X}\big] = \prod_{i=1}^n \mathbb{E}\big[e^{\lambda X_i}\big] \leqslant \prod_{i=1}^n \exp\big((e^\lambda - 1)\mathbb{E}[X_i]\big) = \exp\big((e^\lambda - 1)\mathbb{E}[X]\big),$$

as required. □

The following lemma provides us with the lower bound on $\mathbb{P}\big(A[I\times C] = R \mid \mathcal{E}\big)$ that we require in order to prove the second statement in Theorem 5.1.

Lemma 5.8. *Under the same assumptions as Lemma 5.6, but with the extra condition that no row sum of R exceeds $u/24$, where $q_{z-1} = x^{1/u}$, we have*

$$\frac{\mathbb{P}\big(A[I\times C] = R \mid \mathcal{E}\big)}{\mathbb{P}\big(A[I\times C] = 0 \mid \mathcal{E}\big)} = \exp\left(\frac{O\big(|R|_2\big)}{u_0}\right)\frac{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I\times C] = R\big)}{\mathbb{P}(\tilde{A}_{\mathcal{E}}[I\times C] = 0\big)}. \tag{65}$$

To prove Lemma 5.8 we would like to repeat the calculation in the proof of Lemma 5.6, except using the second statement in Lemma 5.5. However, there is a problem: we have no control over the entries of B^- in the rows of I. Indeed, if B^- contains too many 1s in some row $i \in I$ that is non-zero in R, then we will be unable to find any non-trivial lower bound on $\mathbb{P}(A_M = B)$. Fortunately however, very few matrices B have this property, so we can simply use the trivial lower bound (i.e., zero) for those matrices. The main challenge is to show that we also do not have a large contribution to $\mathbb{P}(A[I \times C] = 0 \mid \mathcal{E})$ in this case.

Proof of Lemma 5.8. We are only required to prove the lower bound, since the upper bound follows from Lemma 5.6. Define

$$\mathcal{B}' := \Big\{B \in \mathcal{B} : B^- \text{ has no more than } u/24 \text{ 1s in row } i \text{ for all } i \in I\Big\},$$

and set $\mathcal{B}'_R = \mathcal{B}' \cap \mathcal{B}_R$ and $\mathcal{B}'_0 = \mathcal{B}' \cap \mathcal{B}_0$. Note that if $B \in \mathcal{B}'_R$, then, since no row of R contains more than $u/24$ 1s, we have at most $u/12$ 1s in each row $i \in I$ of B. Thus we can apply Lemma 5.5 to deduce that

$$\mathbb{P}\big(A_M = B\big) = \exp\left(\frac{O(\alpha_R(B) + |R|_2)}{u_0}\right)\mathbb{P}\big(A_M = \phi_{I,C}(B)\big) \tag{66}$$

for every $B \in \mathcal{B}'_R$. Note that $\alpha_R(B) = \alpha_R(\phi_{I,C}(B))$, since δ_i only depends on R, and $w_i(B^-) = w_i(\phi_{I,C}(B)^-)$ for every $i \in I$, and observe that, as in Observation 5.3, if B is chosen uniformly from $\mathcal{B}'_R$, then $\phi_{I,C}(B)$ is uniform on $\mathcal{B}'_0$, since the extra condition that $B \in \mathcal{B}'$ does not depend on the columns C of the matrix B. Thus, taking the expectation of (66) over a uniform choice of $B \in \mathcal{B}'_R$, we obtain

$$\begin{aligned}
&\frac{\mathbb{P}\big(A_M \in \mathcal{B}'_R\big)}{|\mathcal{B}'_R|} \\
&\quad = \mathbb{E}_{B\in\mathcal{B}'_R}\left[\exp\left(\frac{O\big(\alpha_R(\phi_{I,C}(B)) + |R|_2\big)}{u_0}\right)\mathbb{P}\big(A_M = \phi_{I,C}(B)\big)\right] \\
&\quad = \mathbb{E}_{B\in\mathcal{B}'_0}\left[\exp\left(\frac{O\big(\alpha_R(B) + |R|_2\big)}{u_0}\right)\mathbb{P}\big(A_M = B\big)\right] \\
&\quad = \frac{1}{|\mathcal{B}'_0|}\mathbb{E}\left[\exp\left(\frac{O\big(\alpha_R(A_M) + |R|_2\big)}{u_0}\right) \;\middle|\; A_M \in \mathcal{B}'_0\right]\mathbb{P}\big(A_M \in \mathcal{B}'_0\big).
\end{aligned} \tag{67}$$

The following claim will allow us to bound the right-hand side of (67).

Claim 1: $\mathbb{E}\big[\alpha_R(A_M) \mid A_M \in \mathcal{B}'_0\big] = O(|R|_1)$.

Proof of Claim 1. Since δ_i depends only on the fixed matrix R, the definition (55) of α_R and Observation 5.4 imply that

$$\begin{aligned}\mathbb{E}\big[\alpha_R(A_M) \mid A_M \in \mathcal{B}_0'\big] &= \sum_{i\in I} \delta_i\, \mathbb{E}\big[w_i(A_M^-) \mid A_M \in \mathcal{B}_0'\big] \\ &\leqslant 6\sum_{i\in I}\delta_i \sum_{j\in[z,\pi(x)]\setminus C} \mathbb{P}\big(A_{ij} = 1 \mid A_M \in \mathcal{B}_0'\big).\end{aligned} \tag{68}$$

We claim that

$$\mathbb{P}\big(A_{ij} = 1 \mid A_M \in \mathcal{B}_0'\big) \leqslant e^{O(1/u_0)}\frac{d_j}{|M|} \tag{69}$$

for every $i \in I$ and $j \in [z, z_0^5] \setminus C$. Indeed, if B is chosen uniformly at random from the set $\mathcal{C}_1 := \{B \in \mathcal{B}_0' : B_{ij} = 1\}$, then $\phi_{\{i\},\{j\}}(B)$ is uniform on $\mathcal{C}_0 := \{B \in \mathcal{B}_0' : B_{ij} = 0\}$. Lemma 5.5 then implies (cf. (67)) that

$$\begin{aligned}\mathbb{P}\big(A_M \in \mathcal{C}_1\big) &= |\mathcal{C}_1| \cdot \mathbb{E}_{B\in\mathcal{C}_1}\mathbb{P}\big(A_M = B\big) \\ &\leqslant |\mathcal{C}_1| e^{O(1/u_0)}\mathbb{E}_{B\in\mathcal{C}_1}\mathbb{P}\big(A_M = \phi_{I,C}(B)\big) \\ &= |\mathcal{C}_1| e^{O(1/u_0)}\mathbb{E}_{B\in\mathcal{C}_0}\mathbb{P}\big(A_M = B\big) = \frac{|\mathcal{C}_1|}{|\mathcal{C}_0|}e^{O(1/u_0)}\mathbb{P}\big(A_M \in \mathcal{C}_0\big).\end{aligned}$$

Now $|\mathcal{C}_1|/|\mathcal{C}_0| = d_j/(|M| - d_j)$, so

$$\frac{\mathbb{P}\big(A_{ij} = 1 \mid A_M \in \mathcal{B}_0'\big)}{\mathbb{P}\big(A_{ij} = 0 \mid A_M \in \mathcal{B}_0'\big)} = \frac{\mathbb{P}\big(A_M \in \mathcal{C}_1\big)}{\mathbb{P}\big(A_M \in \mathcal{C}_0\big)} \leqslant e^{O(1/u_0)}\frac{d_j}{|M| - d_j},$$

which implies (69), because $d_j \leqslant 4u_0$ (since $\mathcal{K}(z)$ holds) and $u_0^2 = o(|M|)$, by (44).

Now, combining (68) with (69) gives

$$\mathbb{E}\big[\alpha_R(A_M) \mid A_M \in \mathcal{B}_0'\big] \leqslant O(1)\sum_{i\in I}\delta_i \sum_{j=z}^{\pi(x)} \frac{d_j}{|M|} = O(|R|_1),$$

where the final step follows by (53), and since $\sum_{i\in I}\delta_i = O(|R|_1)$. □

By Claim 1 and the convexity of the exponential function, we obtain

$$\mathbb{E}\left[\exp\left(\frac{O\big(\alpha_R(A_M)\big)}{u_0}\right) \Big| A_M \in \mathcal{B}_0'\right] \geqslant \exp\left(\frac{O(|R|_1)}{u_0}\right).$$

Now, combining this with (67) and noting that $|R|_1 \leqslant |R|_2$ gives

$$\mathbb{P}\big(A_M \in \mathcal{B}_R\big) \geqslant \mathbb{P}\big(A_M \in \mathcal{B}_R'\big) \geqslant e^{O(|R|_2/u_0)}\frac{|\mathcal{B}_R'|}{|\mathcal{B}_0'|}\mathbb{P}\big(A_M \in \mathcal{B}_0'\big). \tag{70}$$

Note also that

$$\frac{|\mathcal{B}_R'|}{|\mathcal{B}_0'|} = \frac{|\mathcal{B}_R|}{|\mathcal{B}_0|} = \frac{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I\times C] = R\big)}{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I\times C] = 0\big)}, \tag{71}$$

since the condition that $B \in \mathcal{B}$ lies in $\mathcal{B}'$ does not affect the columns in C. The following claim will therefore be sufficient to complete the proof of the lemma.

Claim 2: $\mathbb{P}\big(A_M \notin \mathcal{B}'_0 \mid A_M \in \mathcal{B}_0\big) = O(e^{-u})$.

Proof of Claim 2. Recall that if $B \in \mathcal{B}_0 \setminus \mathcal{B}'_0$, then $B\big[\{i\} \times ([z, z_0^5] \setminus C)\big]$ contains at least $u/24$ 1s for some $i \in I$. Our strategy will be to use Lemma 5.6 to bound the probability that this property is satisfied by A_M in terms of the probability that it is satisfied by $\tilde{A}_{\mathcal{E}}$, and then use the independence of the columns of $\tilde{A}_{\mathcal{E}}$ to deduce the desired bound.

In order to apply Lemma 5.6, we will first have to cover the event $A_M \in \mathcal{B}_0 \setminus \mathcal{B}'_0$ with a suitable collection of events. To do so, let $\mathcal{C}$ be the collection of subsets $C' \subseteq [z, z_0^5] \setminus C$ of size exactly $\lceil u/24 \rceil$, and let $\mathcal{R}(C')$ be the set of $R' \in \mathcal{R}(I, C')$ with some row $i \in I$ of R' consisting entirely of 1s. Now if $A_M \in \mathcal{B}_0 \setminus \mathcal{B}'_0$, then there must exist some $C' \in \mathcal{C}$ and $R' \in \mathcal{R}(C')$ such that $A[I \times C'] = R'$. Thus, by the union bound, we have

$$\mathbb{P}\big(A_M \notin \mathcal{B}'_0 \mid A_M \in \mathcal{B}_0\big) \leqslant \sum_{C' \in \mathcal{C}} \sum_{R' \in \mathcal{R}(C')} \mathbb{P}\big(A[I \times C'] = R' \mid A_M \in \mathcal{B}_0\big). \tag{72}$$

We claim that

$$\mathbb{P}\big(A[I \times C'] = R' \mid A_M \in \mathcal{B}_0\big) = \frac{\mathbb{P}\big(A[I \times C'] = R',\, A[I \times C] = 0 \mid \mathcal{E}\big)}{\mathbb{P}\big(A[I \times C] = 0 \mid \mathcal{E}\big)} \tag{73}$$

for every $C' \in \mathcal{C}$ and $R' \in \mathcal{R}(C')$. Indeed, by Observation 5.2 both sides are equal to

$$\frac{\mathbb{P}\big(A[I \times C'] = R', A[I \times C] = 0 \mid A_M \in \mathcal{B}\big)}{\mathbb{P}\big(A[I \times C] = 0 \mid A_M \in \mathcal{B}\big)}.$$

We are now ready to use Lemma 5.6 to show that

$$\begin{aligned} &\frac{\mathbb{P}\big(A[I \times C'] = R',\, A[I \times C] = 0 \mid \mathcal{E}\big)}{\mathbb{P}\big(A[I \times (C' \cup C)] = 0 \mid \mathcal{E}\big)} \\ &\leqslant e^{O(u_0)} \frac{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times C'] = R',\, \tilde{A}_{\mathcal{E}}[I \times C] = 0\big)}{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times (C' \cup C)] = 0\big)}. \end{aligned} \tag{74}$$

Indeed, this follows by applying Lemma 5.6 with R an $I \times (C \cup C')$ matrix that is R' on $I \times C'$ and zero on $I \times C$, noting that $|R'|_1 \leqslant 4u_0|C'| \leqslant u_0^2$.

Note that the right-hand side of (73) is at most the left-hand side of (74), since we have added the condition that $A[I \times C'] = 0$ in the denominator. Observe also that

$$\frac{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times C'] = R',\, \tilde{A}_{\mathcal{E}}[I \times C] = 0\big)}{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times (C' \cup C)] = 0\big)} = \frac{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times C'] = R'\big)}{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times C'] = 0\big)}, \tag{75}$$

since the columns of $\tilde{A}_{\mathcal{E}}$ are independent, and also that

(76)
$$\begin{aligned}\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I\times C'] = 0\big) &= \prod_{j\in C'}\binom{|M|-|I|}{d_j}\binom{|M|}{d_j}^{-1} = \prod_{j\in C'}\prod_{t=0}^{d_j-1}\frac{|M|-|I|-t}{|M|-t}\\ &= \prod_{j\in C'}\exp\bigg(-\frac{O(|I|d_j)}{|M|}\bigg)\\ &\geqslant \exp\bigg(-\frac{O(|I|u_0^2)}{|M|}\bigg) = 1-o(1),\end{aligned}$$

since $\sum_{j\in C'} d_j \leqslant 4u_0|C'| \leqslant u_0^2$, and $|I| = e^{O(u_0)} = o(|M|/u_0^2)$, by (23) and (44).

Combining (72), (73), (74), (75) and (76), we obtain

$$\begin{aligned}\mathbb{P}\big(A_M \notin \mathcal{B}_0' \mid A_M \in \mathcal{B}_0\big) &\leqslant e^{O(u_0)}\sum_{C'\in\mathcal{C}}\sum_{R'\in\mathcal{R}(C')}\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I\times C'] = R'\big)\\ &\leqslant e^{O(u_0)}\sum_{C'\in\mathcal{C}}\sum_{i\in I}\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[\{i\}\times C'] = 1\big),\end{aligned} \tag{77}$$

where 1 indicates the all 1s vector, since the events $\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I\times C'] = R'\big)$ are disjoint.

Now, let $X = \sum_{j=z}^{z_0^5} X_j$, where $X_z, \dots, X_{z_0^5}$ are independent Bernoulli random variables with $\mathbb{P}(X_j = 1) = d_j/|M|$. We claim that

$$\sum_{C'\in\mathcal{C}}\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[\{i\}\times C'] = 1\big) = e^{O(u_0)}\mathbb{P}\big(X = \lceil u/24\rceil\big) \tag{78}$$

for every $i \in I$. Indeed, simply note that

$$\begin{aligned}\mathbb{P}\big(X = \lceil u/24\rceil\big) &= \sum_{C'\in\mathcal{C}}\prod_{j\in C'}\frac{d_j}{|M|}\prod_{j\in[z,z_0^5]\setminus C'}\bigg(1-\frac{d_j}{|M|}\bigg)\\ &= e^{O(u_0)}\sum_{C'\in\mathcal{C}}\prod_{j\in C'}\frac{d_j}{|M|} = e^{O(u_0)}\sum_{C'\in\mathcal{C}}\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[\{i\}\times C'] = 1\big)\end{aligned}$$

for each $i \in I$, since the columns of $\tilde{A}_{\mathcal{E}}$ are independent, and using (53).

Recalling that $|I| = e^{O(u_0)}$ and $u = \Theta(u_0)$, it follows from (77) and (78) that

$$\mathbb{P}\big(A_M \notin \mathcal{B}_0' \mid A_M \in \mathcal{B}_0\big) \leqslant e^{\lambda u}\cdot\mathbb{P}\big(X = \lceil u/24\rceil\big)$$

for some constant $\lambda > 0$. Noting that

$$\mathbb{E}[X] = \sum_{j=z}^{z_0^5}\frac{d_j}{|M|} = O(1)$$

by (53), and that $e^{24(\lambda+1)X} \geqslant e^{(\lambda+1)u}$ for $X = \lceil u/24 \rceil$, it follows by Observation 5.7 that

$$\begin{aligned}\mathbb{P}\big(A_M \notin \mathcal{B}_0' \mid A_M \in \mathcal{B}_0\big) &\leqslant \mathbb{E}\big[e^{24(\lambda+1)X}\big]e^{-u} \\ &\leqslant \exp\Big(\big(e^{24(\lambda+1)} - 1\big)\mathbb{E}[X]\Big)e^{-u} = O\big(e^{-u}\big),\end{aligned}$$

as required. □

To complete the proof, we simply combine Claim 2 with (70) and (71). This gives

$$\mathbb{P}\big(A_M \in \mathcal{B}_R\big) \geqslant e^{O(|R|_2/u_0)} \frac{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times C] = R\big)}{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times C] = 0\big)} \mathbb{P}\big(A_M \in \mathcal{B}_0\big).$$

But, by Observation 5.2, we have

$$\frac{\mathbb{P}\big(A[I \times C] = R \mid \mathcal{E}\big)}{\mathbb{P}\big(A[I \times C] = 0 \mid \mathcal{E}\big)} = \frac{\mathbb{P}\big(A_M \in \mathcal{B}_R \mid A_M \in \mathcal{B}\big)}{\mathbb{P}\big(A_M \in \mathcal{B}_0 \mid A_M \in \mathcal{B}\big)} = \frac{\mathbb{P}\big(A_M \in \mathcal{B}_R\big)}{\mathbb{P}\big(A_M \in \mathcal{B}_0\big)},$$

and so the required lower bound follows. □

To prove Theorem 5.1, it just remains to estimate $\mathbb{P}(A[I \times C] = 0 \mid \mathcal{E})$. To do so we prove the following lemma, which follows from $|C|$ applications of Lemma 5.8.

Lemma 5.9. *Under the same assumptions as in Lemma* 5.6,

$$\mathbb{P}\big(A[I \times C] = 0 \mid \mathcal{E}\big) = \exp\big(O(|I|/u_0)\big)\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times C] = 0\big). \tag{79}$$

Proof. Enumerate the elements of C as $\{j_1, \ldots, j_t\}$, and write

$$C_i = \{j_i, j_{i+1}, \ldots, j_t\}.$$

Let $p_i := \mathbb{P}(A[I \times C_i] = 0 \mid \mathcal{E})$ and $\tilde{p}_i := \mathbb{P}(\tilde{A}_{\mathcal{E}}[I \times C_i] = 0)$, and observe that

$$\log \mathbb{P}(A[I \times C] = 0 \mid \mathcal{E}) = -\sum_{i=1}^{t} \log \frac{p_{i+1}}{p_i} = -\sum_{i=1}^{t} \log\left(1 + \frac{p_{i+1} - p_i}{p_i}\right), \tag{80}$$

and similarly for $\tilde{A}_{\mathcal{E}}$ and $\tilde{p}_i$, since $p_{t+1} = \tilde{p}_{t+1} = 1$. We will use Lemma 5.8 to show that

$$\frac{p_{i+1} - p_i}{p_i} = e^{O(d_{j_i}/u_0)} \cdot \frac{\tilde{p}_{i+1} - \tilde{p}_i}{\tilde{p}_i} \tag{81}$$

for each $i \in [t]$. To prove this, define a family

$$\mathcal{R}(i) := \Big\{R \in \mathcal{R}_{\mathcal{E}}(I, C_i) : R[I \times \{j_i\}] \neq 0 \text{ and } R[I \times C_{i+1}] = 0\Big\},$$

and observe that we can write

$$p_{i+1} - p_i = \sum_{R \in \mathcal{R}(i)} \mathbb{P}\big(A[I \times C_i] = R \mid \mathcal{E}\big),$$

and similarly for $\tilde{p}_{i+1} - \tilde{p}_i$. Note that each row sum of each $R \in \mathcal{R}(i)$ is at most 1, and so $|R|_2 = |R|_1 \leqslant d_{j_i} \leqslant 4u_0 = o(|M|/u_0)$. Hence, by applying Lemma 5.8 (with $C = C_i$), we obtain

$$\frac{\mathbb{P}\big(A[I \times C_i] = R \mid \mathcal{E}\big)}{p_i} = e^{O(d_{j_i}/u_0)} \frac{\mathbb{P}(\tilde{A}_{\mathcal{E}}[I \times C_i] = R\big)}{\tilde{p}_i}$$

for each $R \in \mathcal{R}(i)$. It follows that

$$\frac{p_{i+1} - p_i}{p_i} = e^{O(d_{j_i}/u_0)} \sum_{R \in \mathcal{R}(i)} \frac{\mathbb{P}(\tilde{A}_{\mathcal{E}}[I \times C_i] = R\big)}{\tilde{p}_i} = e^{O(d_{j_i}/u_0)} \cdot \frac{\tilde{p}_{i+1} - \tilde{p}_i}{\tilde{p}_i},$$

as claimed.

Next, observe that $\tilde{p}_i/\tilde{p}_{i+1} = \mathbb{P}(\tilde{A}_{\mathcal{E}}[I \times \{j_i\}] = 0\big)$, since the columns of $\tilde{A}_{\mathcal{E}}$ are independent, and so, recalling that $d_{j_i}|I| = e^{O(u_0)} = o(|M|)$, by (23) and (44), we have

$$\frac{\tilde{p}_{i+1} - \tilde{p}_i}{\tilde{p}_i} = \binom{|M|}{d_{j_i}} \binom{|M| - |I|}{d_{j_i}}^{-1} - 1 = \frac{O(d_{j_i}|I|)}{|M|} = o(1). \tag{82}$$

Now $\log(1 + e^a b) = \log(1 + b + O(ab)) = \log(1 + b) + O(ab)$ for all $a = O(1)$ and $b = o(1)$. Applying this with $b = (\tilde{p}_{i+1} - \tilde{p}_i)/\tilde{p}_i$, and using (81) and (82), gives

$$\begin{aligned}
\log\left(1 + \frac{p_{i+1} - p_i}{p_i}\right) &= \log\left(1 + e^{O(d_{j_i}/u)} \cdot \frac{\tilde{p}_{i+1} - \tilde{p}_i}{\tilde{p}_i}\right) \\
&= \log\left(1 + \frac{\tilde{p}_{i+1} - \tilde{p}_i}{\tilde{p}_i}\right) + O\left(\frac{d_{j_i}}{u_0} \cdot \frac{\tilde{p}_{i+1} - \tilde{p}_i}{\tilde{p}_i}\right) \\
&= \log \frac{\tilde{p}_{i+1}}{\tilde{p}_i} + O\left(\frac{d_{j_i}^2 |I|}{u_0 |M|}\right).
\end{aligned}$$

Finally, recall that $\sum_{i=1}^t d_{j_i}^2 \leqslant d(z)^2 + \sum_{k \geqslant 2} k^2 s_k(z) = O(|M|)$, since $\mathcal{K}(z)$ holds and $d(z) \leqslant 4u_0$. Thus, using (80), we obtain

$$\begin{aligned}
\log \mathbb{P}(A[I \times C] = 0 \mid \mathcal{E}) &= -\sum_{i=1}^t \log \frac{\tilde{p}_{i+1}}{\tilde{p}_i} + O\left(\frac{|I|}{u_0|M|} \sum_{i=1}^t d_{j_i}^2\right) \\
&= \log \mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times C] = 0\big) + \frac{O(|I|)}{u_0},
\end{aligned}$$

as required. □

Proof of Theorem 5.1. The reduction to the case $z \in [z_-, z_0^5]$, $C \subseteq [z, z_0^5]$ and $R \in \mathcal{R}_{\mathcal{E}}(I, C)$ was given after the statement of the theorem, so we may assume these conditions hold. Multiplying (79) by (64) gives (51). Similarly, multiplying (79) by (65), and noting that $u > u_0/6$ for all $z \in [z_-, z_0^5]$, gives (52). □

6. The exploration process

When there is exactly one active non-zero entry in column z (i.e., when $d(z) = 1$), a chain reaction is set off that reduces the number of active rows and can significantly alter the hypergraph $\mathcal{S}_A(z)$. In order to control the evolution of the variables $s_k(z)$ (and hence $m(z)$) we shall need a very precise understanding of this *deterministic* process. In this section we shall use techniques from the theory of branching processes to control the expected change of various key parameters of the hypergraph $\mathcal{S}_A(z)$, where we average over the possible matrices A that are consistent with the information observed so far in the filtration; see Algorithm 2.7. Importantly, we shall also obtain strong bounds on the probability of large deviations.

We begin by defining the various parameters that we shall need to control.

Definition 6.1. For each $z \in [\pi(x)]$, and each $k \geqslant 2$, define the following random variables:

(i) $D(z) := m(z) - m(z-1)$, the number of rows removed in step z.
(ii) $\Delta_k(z) := |S_k(z) \setminus S_k(z-1)|$, the number of edges of size k that contain a vertex removed in step z.
(iii) $R_1(z) := d(z) + \sum_{k \geqslant 2} k(s_k(z) - s_k(z-1))$, the number of 1s removed from the matrix in step z (including those in column z) if $d(z) = 1$.
(iv) $\Delta'(z)$, the number of edges of size at least three that have at least two vertices removed in step z.

The following theorem will play a key role in the proof of Theorem 2.6.

THEOREM 6.2. *Suppose that* $z \in [z_-, \pi(x)]$, $\mathcal{K}(z)$ *holds, and* $2s_2(z) \leqslant (1-\varepsilon_1)m(z)$. *Then*

(a) $\mathbb{E}\big[D(z) \mid \mathcal{F}_z,\, d(z) = 1\big] = \left(1 - \dfrac{2s_2(z)}{m(z)} + o(1)\right)^{-1}$;
(b) $\mathbb{E}\big[\Delta_k(z) \mid \mathcal{F}_z,\, d(z) = 1\big] = \left(1 - \dfrac{2s_2(z)}{m(z)} + o(1)\right)^{-1} \dfrac{ks_k(z)}{m(z)} + O\big(m(z)^{-1/3}\big)$ *for all* $k \geqslant 2$;
(c) $\mathbb{E}\big[D(z)^2 \mid \mathcal{F}_z,\, d(z) = 1\big] = O(1)$;
(d) $\mathbb{E}\big[\Delta'(z) \mid \mathcal{F}_z,\, d(z) = 1\big] = O\big(m(z)^{-1/3}\big)$,

where the bounds implicit in the $o(\cdot)$ *and* $O(\cdot)$ *notation are uniform in* k *and* z. *Moreover*,

$$\mathbb{P}\big(R_1(z) \geqslant u_0^2 \mid \mathcal{F}_z,\, d(z) = 1\big) \leqslant z_0^{-20}. \tag{83}$$

The idea is to prove a corresponding theorem in the simpler (independent) model $\tilde{A}_{\mathcal{E}}$ and then use Theorem 5.1 to deduce the statement in $\mathcal{S}_A(z)$. We now recall this model and define some additional notation.

Definition 6.3. Fix $z \in [z_-, \pi(x)]$ and an event $\mathcal{E} \in \mathcal{F}_z^+$ of the form (50) such that $\mathcal{K}(z)$ holds and $d(z) = 1$, and define $\tilde{\mathcal{S}}_{\mathcal{E}}$ to be the random hypergraph with vertex set M and edge set

$$E(\tilde{\mathcal{S}}_{\mathcal{E}}) = \big\{\tilde{e}_j : j \in [z+1, \pi(x)],\, d_j > 0\big\},$$

where $\tilde{e}_j$ is a random subset of M chosen uniformly over all $\binom{|M|}{d_j}$ choices of subsets of M of size d_j, independently for each j. For $k \geqslant 2$, let

$$\tilde{S}_k = \big\{j \in [z+1, \pi(x)] : d_j = k\big\}$$

denote the set of columns corresponding to the k-edges in $\tilde{\mathcal{S}}_{\mathcal{E}}$, and let $\tilde{S} = \bigcup_{k \geqslant 2} \tilde{S}_k$. We also define $\tilde{e}_z = \{v\}$, where v is chosen uniformly at random from M, independently of $\tilde{e}_j$, $j > z$.

We define a deterministic process on the random hypergraph $\tilde{\mathcal{S}}_{\mathcal{E}}$ by "infecting" vertex v at time zero and then at each subsequent step infecting any vertex that is the last non-infected vertex in an edge of $\tilde{S}$. To be precise, we set $\tilde{D}_0 = \{v\}$ and, for each $t \geqslant 1$,

$$\tilde{D}_t := \tilde{D}_{t-1} \cup \big\{w \in M : \text{there exists } j \in \tilde{S} \text{ with } w \in \tilde{e}_j \subseteq \tilde{D}_{t-1} \cup \{w\}\big\}.$$

We remark that such processes are usually referred to as "bootstrap percolation" and have been extensively studied in both deterministic and random settings; see, e.g., [3, 21].

We now define $\tilde{D}$, $\tilde{R}_1$, $\tilde{\Delta}_k$, and $\tilde{\Delta}'$ to be the quantities corresponding to $D(z)$, $R_1(z)$, $\Delta_k(z)$, and $\Delta'(z)$ respectively. That is,

(i) $\tilde{D} := |\tilde{D}_\infty|$, where $\tilde{D}_\infty := \bigcup_{t \geqslant 0} \tilde{D}_t$.
(ii) $\tilde{\Delta}_k := \Big|\big\{j \in \tilde{S}_k : \tilde{e}_j \cap \tilde{D}_\infty \neq \emptyset\big\}\Big|$ for each $k \geqslant 2$.
(iii) $\tilde{R}_1 := 1 + \sum_{j \in \tilde{S}} |\tilde{e}_j \cap \tilde{D}_\infty|$. (Note that the extra 1 is for the single 1 in column z.)
(iv) $\tilde{\Delta}' := \Big|\big\{j \in \bigcup_{k \geqslant 3} \tilde{S}_k : |\tilde{e}_j \cap \tilde{D}_\infty| \geqslant 2\big\}\Big|$.

For $k \geqslant 3$, we also define $\tilde{\Delta}_k^{(1)}$ by

$$\tilde{\Delta}_k^{(1)} := \Big|\big\{j \in \tilde{S}_k : |\tilde{e}_j \cap \tilde{D}_\infty| = 1\big\}\Big|.$$

Note that for $k \geqslant 3$, $\tilde{\Delta}_k^{(1)} \leqslant \tilde{\Delta}_k \leqslant \tilde{\Delta}_k^{(1)} + \tilde{\Delta}'$.

The hypergraph $\tilde{\mathcal{S}}_{\mathcal{E}}$ is the hypergraph corresponding to $\mathcal{S}_A(z)$, except that we use the matrix $\tilde{A}_{\mathcal{E}}$ from Section 5 in place of A. In particular, if $\mathcal{E}$ holds, then $\tilde{S}_k = S_k(z)$ and $M = M(z)$. We remark that, for emphasis, we shall put tildes over all random variables that are functions of the random hypergraph $\tilde{\mathcal{S}}_{\mathcal{E}}$ and write $s_k = |\tilde{S}_k|$ and $m = |M|$ so that if $\mathcal{E}$ holds, then $m = m(z)$ and $s_k = s_k(z)$. We label the hyperedges of $\tilde{\mathcal{S}}_{\mathcal{E}}$ by the column indices $j \in [z+1, \pi(x)]$; the realizations of these edges as subsets of vertices are then given by the random variables $\tilde{e}_j = \{i \in M : (\tilde{A}_{\mathcal{E}})_{ij} = 1\}$. However,

as we analyse the progress of the "avalanche" we shall reveal information about these random subsets only when necessary. Although technically not part of the hypergraph $\tilde{\mathcal{S}}_{\mathcal{E}}$, we also define $\tilde{e}_z$ as the random 1-edge corresponding to column z. We remark also that, since we will always assume that $\mathcal{K}(z)$ (and hence $\mathcal{M}(z)$) holds, it follows from (44) that $m \geqslant z_0^{1+o(1)}$.

Our first main task will be to prove the following bound on the probability (in the independent random hypergraph model) of large deviations of $\tilde{R}_1$.

LEMMA 6.4. *Suppose that $\mathcal{E}$ is such that $\mathcal{K}(z)$ holds, $d(z) = 1$, and $2s_2 \leqslant (1-\varepsilon_1)m$. Then there exists a constant $\lambda > 0$, depending only on ε_1, such that*

$$\mathbb{P}\big(\tilde{R}_1 \geqslant t\big) = \frac{O(s_2)}{m} e^{-\lambda t} \tag{84}$$

and

$$\mathbb{E}\Big[\tilde{\Delta}_k^{(1)} \mathbb{1}_{\{t \leqslant \tilde{R}_1 < m^{1/2}\}}\Big] = \frac{O(ks_k)}{m} e^{-\lambda t} \tag{85}$$

for all $2 \leqslant t \leqslant m^{1/2}$, uniformly in t and $k \geqslant 3$.

We shall next precisely define the process via which we reveal the set of vertices removed in the independent model. For each integer $t \geqslant 0$, let $\tilde{A}_t$ and $\tilde{V}_t$ (the *active* and *visited* vertices respectively) be subsets of M given by the following algorithm. For technical reasons, we shall need an upper bound on the number of vertices we visit during the process.

Algorithm 6.5. We start with $t := 1$, $\tilde{A}_0 = \tilde{V}_0 := \{v\}$ and $\tilde{E}_0 := \{z\}$, where v is the vertex corresponding to the unique 1 in column z. Define $\tilde{\jmath}(v) = z$ and repeat the following steps until $|\tilde{A}_t| = 0$:

1. Pick $u \in \tilde{A}_{t-1}$ with the smallest value of $\tilde{\jmath}(u)$, and list the elements of $M \setminus \tilde{V}_{t-1}$ (in increasing order, say) as $w_1, \dots, w_r$. Set $\tilde{A}^{(0)} := \tilde{A}_{t-1}$, $\tilde{V}^{(0)} := \tilde{V}_{t-1}$, $\tilde{E}^{(0)} := \tilde{E}_{t-1}$ and $\ell := 0$.
2. While $|\tilde{V}^{(\ell)}| < m^{1/2}$ and $\ell < r$, repeat the following steps:
   (a) Set $\ell := \ell + 1$.
   (b) Let $\tilde{\jmath}(w_\ell)$ be the smallest $j \in \tilde{S}$ with $\{u, w_\ell\} \subseteq \tilde{e}_j \subseteq \tilde{V}_{t-1} \cup \{w_\ell\}$, if such a j exists. Set $\tilde{V}^{(\ell)} := \tilde{V}^{(\ell-1)} \cup \{w_\ell\}$, $\tilde{A}^{(\ell)} := \tilde{A}^{(\ell-1)} \cup \{w_\ell\}$, and $\tilde{E}^{(\ell)} := \tilde{E}^{(\ell-1)} \cup \{\tilde{\jmath}(w_\ell)\}$. If no such j exists, then set $\tilde{V}^{(\ell)} := \tilde{V}^{(\ell-1)}$, $\tilde{A}^{(\ell)} := \tilde{A}^{(\ell-1)}$, and $\tilde{E}^{(\ell)} := \tilde{E}^{(\ell-1)}$.
3. Set $\tilde{A}_t := \tilde{A}^{(\ell)} \setminus \{u\}$, $\tilde{V}_t := \tilde{V}^{(\ell)}$, and $\tilde{E}_t := \tilde{E}^{(\ell)}$.
4. If $|\tilde{A}_t| = 0$, then set $\tilde{V}_\infty := \tilde{V}_t$ and $\tilde{E}_\infty := \tilde{E}_t$; otherwise set $t := t + 1$ and return to Step 1.

Let us begin by making a couple of simple but key observations about this algorithm.

Observation 6.6. *$|\tilde{A}_t| = |\tilde{V}_t| - t$ for every $0 \leqslant t \leqslant |\tilde{V}_\infty|$, and $|\tilde{V}_\infty| = \min\left\{\tilde{D}, \lceil m^{1/2} \rceil\right\}$.*

Proof. The equation $|\tilde{A}_t| = |\tilde{V}_t| - t$ follows since we add the same elements to $\tilde{A}_t$ and $\tilde{V}_t$ but remove one element from $\tilde{A}_t$ at each time step. To see that $|\tilde{V}_\infty| \leqslant \tilde{D}$, observe that in fact we have $\tilde{V}_t \subseteq \tilde{D}_t$ for all $t \geqslant 0$ by induction, as we only add vertices to $\tilde{V}_t$ that are added to $\tilde{D}_t$. Moreover, if $|\tilde{V}_t| < \lceil m^{1/2} \rceil$ for every $t \geqslant 0$, then the algorithm discovers all vertices that are included in $\tilde{D}_\infty$, so in that case $\tilde{V}_\infty = \tilde{D}_\infty$, and so $|\tilde{V}_\infty| = \tilde{D}$. On the other hand, if $|\tilde{V}_t| = \lceil m^{1/2} \rceil$ for some t, then we visit no new vertices after that point, and therefore $\tilde{V}_{t'} = \tilde{V}_t$ for all $t < t' \leqslant |\tilde{V}_\infty|$, and hence $|\tilde{V}_\infty| = \lceil m^{1/2} \rceil$, as claimed. □

We think of $(|\tilde{A}_t|)_{t \geqslant 0}$ as a random walk and of $|\tilde{V}_\infty|$ as the hitting time of 0. However, the steps of this random walk might be large and are not independent. Therefore, in order to control the walk we shall need to break the steps up into smaller pieces. Let us define random variables $\tilde{X}_{t,w} \in \{0,1\}$ for each $1 \leqslant t \leqslant |\tilde{V}_\infty|$ and $w \in M \setminus \tilde{V}_{t-1}$ by setting

$$\tilde{X}_{t,w} = 1 \quad \Leftrightarrow \quad w \in \tilde{V}_t \setminus \tilde{V}_{t-1}.$$

Abusing notation slightly, let us define a filtration $\mathcal{F}_z^+ = \tilde{\mathcal{F}}_0 \subseteq \tilde{\mathcal{F}}_1 \subseteq \cdots \subseteq \tilde{\mathcal{F}}_{|\tilde{V}_\infty|} = \tilde{\mathcal{F}}_\infty$ by defining $\tilde{\mathcal{F}}_t$ to be the information observed (about the independent model) at the moment $\tilde{V}_t$ is defined.[10] For each $1 \leqslant t \leqslant |\tilde{V}_\infty|$, let us define a further filtration

$$\tilde{\mathcal{F}}_{t-1} = \tilde{\mathcal{F}}_{t,<w_1} \subseteq \tilde{\mathcal{F}}_{t,<w_2} \subseteq \cdots \subseteq \tilde{\mathcal{F}}_{t,<w_r} \subseteq \tilde{\mathcal{F}}_t$$

by defining $\tilde{\mathcal{F}}_{t,<w}$ to be the information observed just before we begin Step 2(b) in the round of Algorithm 6.5 in which we discover whether or not $w \in \tilde{V}_t \setminus \tilde{V}_{t-1}$.

Remark 6.7. Note that in Step 2(b) of the algorithm we only need to observe whether or not the hyperedge $\tilde{e}_j$ satisfies $\{u, w_\ell\} \subseteq \tilde{e}_j \subseteq \tilde{V}_{t-1} \cup \{w_\ell\}$ in turn[11] for each j until we find one that does, or we have exhausted all $j \in \tilde{S}$. Moreover, if we do find such an edge, then we do not test this condition for larger j. We emphasize that this is the only (new) information contained in $\tilde{\mathcal{F}}_{t,<w_{\ell+1}}$, and therefore, for edges $\tilde{e}_j \in E(\tilde{\mathcal{S}}_{\mathcal{E}})$ that are not used in the process, we only have "negative" information (i.e., information of the form "the hyperedge $\tilde{e}_j$ does not satisfy $\{u, w_\ell\} \subseteq \tilde{e}_j \subseteq \tilde{V}_{t-1} \cup \{w_\ell\}$"). This fact will play an important role in the proof; see Lemmas 6.8, 6.12 and 6.13, below.

[10] Note that after we have visited $m^{1/2}$ vertices, the algorithm does not observe any further new information, and so the σ-algebras of the filtration are all the same from that point on.

[11] Since $\tilde{\jmath}(w_\ell)$ is the smallest $j \in \tilde{S}$ with this property, we consider the elements of $\tilde{S}$ in increasing order.

In order to bound $\tilde{R}_1$ and the other variables introduced in Definition 6.3, we shall need to consider both edges $\tilde{e}_j$ with $j \in \tilde{E}_\infty$ (i.e., edges of $\tilde{\mathcal{S}}_{\mathcal{E}}$ that are used in the algorithm), and edges $\tilde{e}_j$, $j \in \tilde{S} \setminus \tilde{E}_\infty$, that nonetheless intersect $\tilde{V}_\infty$. Before embarking on the proof of Lemma 6.4, we shall use Remark 6.7 to control the distributions of the number of both types of edges.

We begin with the edges $\tilde{e}_j$, $j \in \tilde{E}_\infty$. Let us define a random variable $\tilde{X}^{(k)}_{t,w} \in \{0,1\}$ for each $k \geqslant 2$, $1 \leqslant t \leqslant |\tilde{V}_\infty|$ and $w \in M \setminus \tilde{V}_{t-1}$ by setting

$$\tilde{X}^{(k)}_{t,w} = 1 \iff w \in \tilde{V}_t \setminus \tilde{V}_{t-1} \quad \text{and} \quad \tilde{\jmath}(w) \in \tilde{S}_k.$$

Note that $\tilde{X}_{t,w} = \sum_{k \geqslant 2} \tilde{X}^{(k)}_{t,w}$ for every t and w. Define $\tilde{\mathcal{X}}_{t,w}$ to be the event that $\tilde{V}^{(\ell)} < m^{1/2}$ just before we test vertex w in time step t. Thus $\tilde{X}_{t,w} = 1$ is only possible if $\tilde{\mathcal{X}}_{t,w}$ holds.

We write $\mathbb{1}_{\mathcal{A}}$ to denote the indicator function of an event $\mathcal{A}$.

Lemma 6.8. *Suppose that $\mathcal{E}$ is such that $\mathcal{K}(z)$ holds, $d(z) = 1$, and $2s_2 \leqslant (1-\varepsilon_1)m$. Then*

$$\mathbb{E}\big[\tilde{X}^{(2)}_{t,w} \mid \tilde{\mathcal{F}}_{t,<w}\big] = \left(\frac{2s_2}{m^2} + O\big(m^{-3/2}\big)\right) \mathbb{1}_{\tilde{\mathcal{X}}_{t,w}}$$

and if $k \geqslant 3$, then

$$\mathbb{E}\big[\tilde{X}^{(k)}_{t,w} \mid \tilde{\mathcal{F}}_{t,<w}\big] \leqslant \frac{2k^2 s_k}{m^{(k+2)/2}} \mathbb{1}_{\tilde{\mathcal{X}}_{t,w}}$$

for every $t \geqslant 1$ and $w \in M \setminus \tilde{V}_{t-1}$. As a consequence,

$$\mathbb{E}\big[\tilde{X}_{t,w} \mid \tilde{\mathcal{F}}_{t,<w}\big] = \left(\frac{2s_2}{m^2} + O\big(m^{-3/2}\big)\right) \mathbb{1}_{\tilde{\mathcal{X}}_{t,w}}.$$

Moreover, the constants implicit in the $O(\cdot)$ notation are uniform in z, t and w.

Proof. Let $\tilde{V}^{(\ell)}$ be the set of visited vertices just before we ask whether or not $w \in \tilde{V}_t \setminus \tilde{V}_{t-1}$. If $|\tilde{V}^{(\ell)}| \geqslant m^{1/2}$, then $\tilde{X}_{t,w} = \mathbb{1}_{\tilde{\mathcal{X}}_{t,w}} = 0$, so we may assume that $|\tilde{V}^{(\ell)}| < m^{1/2}$. As $\mathcal{K}(z)$ holds, we may also assume $k \leqslant 4u_0$, as otherwise $\tilde{S}_k = \emptyset$ and hence $\tilde{X}^{(k)}_{t,w} = 0$. By (44), we may also assume that $m \geqslant z_0^{1+o(1)}$, so $u_0 = o(\log m)$. We shall prove the first two statements by bounding (for each $k \geqslant 2$) the probability that $w \in \tilde{V}_t \setminus \tilde{V}_{t-1}$ and $\tilde{\jmath}(w) \in \tilde{S}_k$ by the expected number of edges $j \in \tilde{S}_k \setminus \tilde{E}^{(\ell)}$ of size k with $\{u,w\} \subseteq \tilde{e}_j \subseteq \tilde{V}_{t-1} \cup \{w\}$.

Indeed, by Remark 6.7, conditioned on $\tilde{\mathcal{F}}_{t,<w}$, the edges $\tilde{e}_j$, $j \in \tilde{S}_k \setminus \tilde{E}^{(\ell)}$, are each chosen independently and uniformly from the collection of sets that fail to satisfy the test in Step 2(b) in any prior time step where the edge $\tilde{e}_j$ was actually tested. This (crucially) includes all k-subsets of M that contain at least two vertices of $M \setminus \tilde{V}_{t-1}$. Since $|\tilde{V}_{t-1}| < m^{1/2}$, the number of such sets

is therefore

$$\binom{m}{k} - O(m)\binom{|\tilde{V}_{t-1}|}{k-1} = \big(1+o(1)\big)\binom{m}{k} = \big(1+o(1)\big)\frac{m^k}{k!},$$

where we have used the fact that $2 \leqslant k \leqslant 4u_0 = o(m^{1/2})$, by (44), and the general fact that $\binom{n}{r} = (1 - O(r^2/n))n^r/r!$. It follows that, for each $k \geqslant 2$ and each $j \in \tilde{S}_k \setminus \tilde{E}^{(\ell)}$, the edge $\tilde{e}_j$ satisfies $\{u,w\} \subseteq \tilde{e}_j \subseteq \tilde{V}_{t-1} \cup \{w\}$ with (conditional) probability

$$\big(1+o(1)\big)\frac{k!}{m^k} \cdot \binom{|\tilde{V}_{t-1}|}{k-2} \leqslant \big(1+o(1)\big)\frac{k(k-1)|\tilde{V}_{t-1}|^{k-2}}{m^k} \tag{86}$$

so, in particular, for each $k \geqslant 3$, we have

$$\mathbb{E}\big[\tilde{X}^{(k)}_{t,w} \mid \tilde{\mathcal{F}}_{t,<w}\big] \leqslant (1+o(1))\frac{k(k-1)|\tilde{V}_{t-1}|^{k-2}}{m^k} \cdot s_k \leqslant \frac{2k^2 s_k}{m^{(k+2)/2}}$$

since $|\tilde{V}_{t-1}| < m^{1/2}$, as claimed. When $k = 2$, on the other hand, we can replace (86) by

$$\left(\binom{m}{2} - O\big(m \cdot |\tilde{V}_{t-1}|\big)\right)^{-1} = \frac{2}{m^2} + O\big(m^{-5/2}\big),$$

since $|\tilde{V}_{t-1}| < m^{1/2}$, and that at most $|\tilde{V}_{t-1}|$ edges have already been used in the process (since every time a new edge is used, we visit a new vertex). Hence the probability that some 2-edge satisfies $\tilde{e}_j = \{u,w\}$ is

$$\left(\frac{2}{m^2} + O\big(m^{-5/2}\big)\right)\big(s_2 + O\big(m^{1/2}\big)\big) = \frac{2s_2}{m^2} + O\big(m^{-3/2}\big),$$

as claimed, since $s_2 = O(m)$.

For the last part, we note that $\tilde{X}_{t,w} = \sum_{k\geqslant 2} \tilde{X}^{(k)}_{t,w}$ and so, assuming $\tilde{\mathcal{X}}_{t,w}$ holds,

$$\mathbb{E}\big[\tilde{X}_{t,w} - \tilde{X}^{(2)}_{t,w} \mid \tilde{\mathcal{F}}_{t,<w}\big] \leqslant \sum_{k\geqslant 3} \frac{2k^2 s_k}{m^{(k+2)/2}} = O\big(m^{-3/2}\big)$$

as $\mathcal{K}(z)$ holds, so $\sum_{k\geqslant 2} s_k = O(m)$. Uniformity in z, t and w follows as all the $o(\ )$ terms are in fact bounded by functions of m, and since $\mathcal{K}(z)$ holds we have $m \geqslant z_0^{1+o(1)}$, by (44). □

Next, define

$$\tilde{f}(t,w) := \begin{cases} k & \text{if } \tilde{X}_{t,w} = 1 \text{ and } \tilde{\jmath}(w) \in \tilde{S}_k, \\ 0 & \text{if } \tilde{X}_{t,w} = 0. \end{cases}$$

Using Lemma 6.8, we can easily deduce the following bounds, which will be needed in the proof of Lemma 6.4, below.

LEMMA 6.9. *Suppose* $\mathcal{E}$ *is such that* $\mathcal{K}(z)$ *holds,* $d(z) = 1$, *and* $2s_2 \leqslant (1-\varepsilon_1)m$. *Then for any* $0 < \lambda \leqslant \varepsilon_1$, *we have*

$$\mathbb{E}\big[e^{\lambda \tilde{X}_{t,w}} \mid \tilde{\mathcal{F}}_{t,<w}\big] \leqslant \exp\left(\frac{\lambda - \lambda^2/2}{m}\right)$$

and

$$\mathbb{E}\big[e^{\lambda \tilde{f}(t,w)} \mid \tilde{\mathcal{F}}_{t,<w}\big] \leqslant \exp\left(\frac{2\lambda}{m}\right)$$

for every $t \geqslant 1$ *and* $w \in M \setminus \tilde{V}_{t-1}$.

Proof. Since $2s_2 \leqslant (1-\varepsilon_1)m$, it follows by Observation 5.7 and Lemma 6.8 that

$$\begin{aligned}\mathbb{E}\big[e^{\lambda \tilde{X}_{t,w}} \mid \tilde{\mathcal{F}}_{t,<w}\big] &\leqslant \exp\left(\big(e^{\lambda}-1\big)\left(\frac{2s_2}{m^2} + O\big(m^{-3/2}\big)\right)\right)\\ &\leqslant \exp\left(\frac{\big(e^{\lambda}-1\big)(1-\varepsilon_1+o(1))}{m}\right) \leqslant \exp\left(\frac{\lambda-\lambda^2/2}{m}\right),\end{aligned}$$

since $0 < \lambda \leqslant \varepsilon_1 < 1$ and so $(e^{\lambda}-1)(1-\varepsilon_1) \leqslant (\lambda + \frac{\lambda^2}{2} + \cdots)(1-\lambda) < \lambda - \frac{\lambda^2}{2}$. Similarly, recalling that $m \geqslant z_0^{1+o(1)}$ and that $\sum_{k\geqslant 2} 2^k s_k = O(m)$, since $\mathcal{K}(z)$ holds, we have

$$\begin{aligned}\mathbb{E}\big[e^{\lambda \tilde{f}(t,w)} \mid \tilde{\mathcal{F}}_{t,<w}\big] &= 1 + \sum_{k=2}^{4u_0} \big(e^{k\lambda}-1\big)\mathbb{P}\big(\tilde{X}_{t,w}^{(k)} = 1 \mid \tilde{\mathcal{F}}_{t,<w}\big)\\ &\leqslant 1 + \big(e^{2\lambda}-1\big)\left(\frac{2s_2}{m^2} + O\big(m^{-3/2}\big)\right) + \sum_{k\geqslant 3} \frac{2k^2(e^{k\lambda}-1)}{m^{(k+2)/2}} s_k\\ &\leqslant 1 + \big(e^{2\lambda}-1\big)\frac{1-\varepsilon_1}{m} + O\big(\lambda m^{-3/2}\big)\\ &\leqslant 1 + \frac{2\lambda}{m} \leqslant \exp\left(\frac{2\lambda}{m}\right),\end{aligned}$$

since $0 < \lambda \leqslant \varepsilon_1 < 1$ and so $(e^{2\lambda}-1)(1-\varepsilon_1) \leqslant (2\lambda + \frac{4\lambda^2}{2} + \frac{8\lambda^3}{6} + \cdots)(1-\lambda) < 2\lambda$. □

In the proof of (85) we shall use the inequality

$$\mathbb{E}\big[\Delta_k^{(1)} \mathbb{1}_{\{t \leqslant \tilde{R}_1 < m^{1/2}\}}\big] \leqslant e^{-\lambda t} \cdot \mathbb{E}\big[\Delta_k^{(1)} e^{\lambda \tilde{R}_1} \mathbb{1}_{\{2 \leqslant \tilde{R}_1 < m^{1/2}\}}\big], \tag{87}$$

so it will be important that we have some control over the distributions of $\Delta_k^{(1)}$ and $\tilde{R}_1$ conditioned on the "positive" information that $\tilde{R}_1 > 1$. The next observation provides us with this control.

OBSERVATION 6.10. *The random variable* $|\tilde{V}_1| - 1$ *is stochastically dominated by the binomial random variable* $\mathrm{Bin}(s_2, 2/m)$. *In particular,* $\mathbb{E}\big[e^{|\tilde{V}_1|} \mid \tilde{R}_1 > 1\big] = O(1)$.

Proof. Only 2-edges can be included in $\tilde{E}_1$ as $j \in \tilde{E}_1$ implies $|\tilde{e}_j \setminus \{v\}| = 1$. Each 2-edge is included precisely when it contains v and is not identical to a previously encountered 2-edge, and this occurs with probability at most $2/m$. Thus $|\tilde{V}_1|$ is stochastically dominated by a $1 + \mathrm{Bin}(s_2, 2/m)$ random variable. The condition that $\tilde{R}_1 > 1$ is precisely the condition that the exploration process does not immediately die out, so it is equivalent to the condition that $|\tilde{V}_1| > 1$. Now conditioning on $\tilde{R}_1 > 1$ is equivalent to conditioning on at least one 2-edge containing v. But conditioned on that, $|\tilde{V}_1| - 2$ is stochastically bounded by a $\mathrm{Bin}(s_2, 2/m)$ random variable, as there are at most s_2 remaining edges to test, and each adds 1 to $|\tilde{V}_1|$ with probability at most $2/m$. By Observation 5.7, $\mathbb{E}[e^{\mathrm{Bin}(s_2,2/m)}] \leqslant \exp((e-1)2s_2/m) = O(1)$, so the second result follows. □

We are ready to bound the contribution to $\tilde{R}_1$ of the edges used in Algorithm 6.5. Let

$$\tilde{W}_\infty := \sum_{w \in \tilde{V}_\infty} |\tilde{e}_{\tilde{\jmath}(w)}| = \sum_{j \in \tilde{E}_\infty} d_j$$

denote the sum of the sizes of these edges; as these edges all lie inside $\tilde{V}_\infty$, this is precisely their contribution to $\tilde{R}_1$. The following lemma controls the size of $\tilde{W}_\infty$.

Lemma 6.11. *There exists some $\lambda > 0$, depending only on ε_1, such that*

$$\mathbb{E}\Big[e^{\lambda \tilde{W}_\infty} \mid \tilde{R}_1 > 1\Big] = O(1).$$

Proof. Note that the condition $\tilde{R}_1 > 1$ is equivalent to $\tilde{W}_\infty > 0$ and is $\tilde{\mathcal{F}}_1$-measurable, as one discovers whether or not $\tilde{R}_1 > 1$ in the first round of Algorithm 6.5. We will first need to control the large deviations of $|\tilde{V}_\infty|$.

Claim 1: For every $t \geqslant 1$,

$$\mathbb{P}\Big(|\tilde{V}_\infty| \geqslant t \mid \tilde{R}_1 > 1\Big) = O\Big(e^{-\varepsilon_1^2 t/2}\Big).$$

Proof of Claim 1. Since $|\tilde{V}_t| = |\tilde{A}_t| + t \geqslant t$ for $t \leqslant |\tilde{V}_\infty|$, we have

(88)
$$\mathbb{P}\Big(|\tilde{V}_\infty| \geqslant t \mid \tilde{R}_1 > 1\Big) = \mathbb{P}\Big(|\tilde{V}_t| \geqslant t \mid \tilde{R}_1 > 1\Big) \leqslant e^{-\varepsilon_1 t}\, \mathbb{E}\Big[e^{\varepsilon_1 |\tilde{V}_t|} \mid \tilde{R}_1 > 1\Big].$$

We shall bound $\mathbb{E}\Big[e^{\varepsilon_1 |\tilde{V}_t|} \mid \tilde{R}_1 > 1\Big]$ using Lemma 6.9 and the law of iterated expectations. Indeed, setting $\tilde{X}_t = \sum_w \tilde{X}_{t,w}$, so that $\tilde{X}_t = |\tilde{V}_t \setminus \tilde{V}_{t-1}|$, we have

$$\begin{aligned} \mathbb{E}\Big[e^{\varepsilon_1 |\tilde{V}_t|} \mid \tilde{R}_1 > 1\Big] &= \mathbb{E}\Big[e^{\varepsilon_1 (1 + \tilde{X}_1 + \cdots + \tilde{X}_t)} \mid \tilde{R}_1 > 1\Big] \\ &= \mathbb{E}\Big[e^{\varepsilon_1 |\tilde{V}_{t-1}|} \cdot \mathbb{E}\Big[e^{\varepsilon_1 \tilde{X}_t} \mid \tilde{\mathcal{F}}_{t-1}\Big] \,\Big|\, \tilde{R}_1 > 1\Big] \end{aligned} \tag{89}$$

and similarly,

$$\begin{aligned}(90)\quad \mathbb{E}\Big[e^{\varepsilon_1 \tilde{X}_t} \mid \tilde{\mathcal{F}}_{t-1}\Big] &= \mathbb{E}\Big[e^{\varepsilon_1(\tilde{X}_{t,w_1}+\cdots+\tilde{X}_{t,w_r})} \mid \tilde{\mathcal{F}}_{t-1}\Big] \\ &= \mathbb{E}\Big[e^{\varepsilon_1(\tilde{X}_{t,w_1}+\cdots+\tilde{X}_{t,w_{r-1}})} \cdot \mathbb{E}\Big[e^{\varepsilon_1 \tilde{X}_{t,w_r}} \mid \tilde{\mathcal{F}}_{t,<w_r}\Big] \mid \tilde{\mathcal{F}}_{t-1}\Big].\end{aligned}$$

Now, applying Lemma 6.9 with $\lambda = \varepsilon_1$, we have

$$\mathbb{E}\Big[e^{\varepsilon_1 \tilde{X}_{t,w}} \mid \tilde{\mathcal{F}}_{t,<w}\Big] \leqslant \exp\left(\frac{\varepsilon_1 - \varepsilon_1^2/2}{m}\right) \tag{91}$$

for every $t \geqslant 1$ and $w \in M \setminus \tilde{V}_{t-1}$. Combining this with (90), and iterating the procedure, we obtain

$$\begin{aligned}\mathbb{E}\Big[e^{\varepsilon_1 \tilde{X}_t} \mid \tilde{\mathcal{F}}_{t-1}\Big] &\leqslant \exp\left(\frac{\varepsilon_1 - \varepsilon_1^2/2}{m}\right)\mathbb{E}\Big[e^{\varepsilon_1(\tilde{X}_{t,w_1}+\cdots+\tilde{X}_{t,w_{r-1}})} \mid \tilde{\mathcal{F}}_{t-1}\Big] \\ &\leqslant \cdots \leqslant \exp\left(\frac{(\varepsilon_1 - \varepsilon_1^2/2)r}{m}\right) \leqslant \exp\big(\varepsilon_1 - \varepsilon_1^2/2\big),\end{aligned}$$

since $r = |M \setminus \tilde{V}_{t-1}| \leqslant m$. Hence, using (89), and iterating again, we have

$$\begin{aligned}\mathbb{E}\Big[e^{\varepsilon_1|\tilde{V}_t|} \mid \tilde{R}_1 > 1\Big] &\leqslant e^{\varepsilon_1 - \varepsilon_1^2/2}\mathbb{E}\Big[e^{\varepsilon_1|\tilde{V}_{t-1}|} \mid \tilde{R}_1 > 1\Big] \\ &\leqslant \cdots \leqslant e^{(\varepsilon_1 - \varepsilon_1^2/2)(t-1)}\mathbb{E}\Big[e^{\varepsilon_1|\tilde{V}_1|} \mid \tilde{R}_1 > 1\Big].\end{aligned}$$

Thus by Observation 6.10,

$$\mathbb{E}\Big[e^{\varepsilon_1|\tilde{V}_t|} \mid \tilde{R}_1 > 1\Big] = O\Big(e^{(\varepsilon_1 - \varepsilon_1^2/2)t}\Big).$$

Finally, it follows from (88) that

$$\mathbb{P}\Big(|\tilde{V}_\infty| \geqslant t \mid \tilde{R}_1 > 1\Big) \leqslant e^{-\varepsilon_1 t} \cdot \mathbb{E}\Big[e^{\varepsilon_1|\tilde{V}_t|} \mid \tilde{R}_1 > 1\Big] = O\Big(e^{-\varepsilon_1^2 t/2}\Big),$$

as claimed. □

We shall next use a similar argument to control

$$\tilde{W}_t := \sum_{w \in \tilde{V}_t} |\tilde{e}_{\tilde{\jmath}(w)}| = \sum_{j \in \tilde{E}_t} d_j,$$

the sum of the sizes of the edges used in the first t iterations of Algorithm 6.5.

Claim 2: If $0 < \lambda \leqslant \varepsilon_1$, then

$$e^{\lambda \tilde{W}_t - 2\lambda t}$$

is a super-martingale with respect to the filtration $(\mathcal{F}_t)_{t \geqslant 0}$.

Proof of Claim 2. The proof is similar to that of Claim 1, except the bound (91) is replaced by

$$\mathbb{E}\Big[e^{\lambda \tilde{f}(t,w)} \mid \tilde{\mathcal{F}}_{t,<w}\Big] \leqslant \exp\left(\frac{2\lambda}{m}\right),$$

which also follows from Lemma 6.9. Indeed,

$$\begin{aligned}\mathbb{E}\Big[e^{\lambda(\tilde{W}_t-\tilde{W}_{t-1})} \mid \tilde{\mathcal{F}}_{t-1}\Big] &= \mathbb{E}\Big[e^{\lambda(\tilde{f}(t,w_1)+\cdots+\tilde{f}(t,w_{r-1}))} \\ &\qquad \cdot \mathbb{E}\Big[e^{\lambda\tilde{f}(t,w_r)} \mid \tilde{\mathcal{F}}_{t,<w_r}\Big] \mid \tilde{\mathcal{F}}_{t-1}\Big] \\ &\leqslant \exp\left(\frac{2\lambda}{m}\right)\mathbb{E}\Big[e^{\lambda(\tilde{f}(t,w_1)+\cdots+\tilde{f}(t,w_{r-1}))} \mid \tilde{\mathcal{F}}_{t-1}\Big] \\ &\leqslant \cdots \leqslant \exp\left(\frac{2\lambda r}{m}\right) \leqslant e^{2\lambda},\end{aligned}$$

since $r \leqslant m$. Since $\tilde{W}_{t-1}$ is $\tilde{\mathcal{F}}_{t-1}$-measurable, it follows immediately that

$$\mathbb{E}\Big[e^{\lambda\tilde{W}_t-2\lambda t} \mid \tilde{\mathcal{F}}_{t-1}\Big] \leqslant e^{\lambda\tilde{W}_{t-1}-2\lambda(t-1)},$$

as required. □

We are now ready to bound the expectation of $e^{\lambda\tilde{W}_\infty}$. Observe first that

$$\mathbb{E}\Big[e^{\lambda\tilde{W}_\infty/2} \mid \tilde{R}_1 > 1\Big] \leqslant \frac{1}{2}\Big(\mathbb{E}\Big[e^{\lambda\tilde{W}_\infty-2\lambda|\tilde{V}_\infty|} \mid \tilde{R}_1 > 1\Big] + \mathbb{E}\Big[e^{2\lambda|\tilde{V}_\infty|} \mid \tilde{R}_1 > 1\Big]\Big),$$

by the convexity of e^x. Now, if $\lambda < \varepsilon_1^2/4$, then

$$\mathbb{E}\Big[e^{2\lambda|\tilde{V}_\infty|} \mid \tilde{R}_1 > 1\Big] \leqslant \sum_{t=0}^{\infty} e^{2\lambda t}\mathbb{P}(|\tilde{V}_\infty| \geqslant t \mid \tilde{R}_1 > 1) = O(1)$$

by Claim 1. Moreover, since the event $\tilde{R}_1 > 1$ is $\tilde{\mathcal{F}}_1$-measurable, it follows from Claim 2 by the optional stopping theorem that

$$\mathbb{E}\Big[e^{\lambda\tilde{W}_\infty-2\lambda|\tilde{V}_\infty|} \mid \tilde{R}_1 > 1\Big] \leqslant \mathbb{E}\Big[e^{\lambda\tilde{W}_1} \mid \tilde{R}_1 > 1\Big] = O(1)$$

for every $\lambda < 1/2$. Indeed, since $\tilde{W}_1 = 2(|\tilde{V}_1| - 1)$ is twice the number of 2-edges used in the first round of Algorithm 6.5, the last equality follows by Observation 6.10. Thus

$$\mathbb{E}\Big[e^{\lambda\tilde{W}_\infty} \mid \tilde{R}_1 > 1\Big] = O(1)$$

for every $\lambda < \varepsilon_1^2/8$, as required. □

We shall next use Remark 6.7 to control the distribution of the number of remaining edges that nonetheless intersect $\tilde{V}_\infty$. For each $k \geqslant 2$, define

$$\tilde{R}(k) := \Big|\Big\{j \in \tilde{S}_k \setminus \tilde{E}_\infty : \tilde{e}_j \cap \tilde{V}_\infty \neq \emptyset\Big\}\Big|$$

to be the number of edges of $\tilde{\mathcal{S}}_{\mathcal{E}}$ of size k that have at least one vertex removed but are not used in Algorithm 6.5. For each $k \geqslant 2$, define binomial random variables

$$Z(k) \sim \mathrm{Bin}\left(s_k, \frac{k|\tilde{V}_\infty|}{m-k}\right).$$

LEMMA 6.12. *The random variables $\tilde{R}(k)$ are conditionally independent given $\tilde{\mathcal{F}}_\infty$ and, conditioned on $\tilde{\mathcal{F}}_\infty$, are stochastically dominated by $Z(k)$ for each $k \geqslant 2$.*

Proof. Since $\mathcal{K}(z)$ holds, we have $s_k = 0$ for all $k > 4u_0$, so we may assume that $k \leqslant 4u_0$. Run the algorithm to reveal $\tilde{V}_\infty$ and $\tilde{E}_\infty$. By Remark 6.7, we have not revealed any edge of $\tilde{S}_k \setminus \tilde{E}_\infty$, though we have gained some "negative" information about the events $\{\tilde{e}_j \cap \tilde{V}_\infty \neq \emptyset\}$. To be precise, conditioned on the information we have observed during the process (i.e., $\tilde{\mathcal{F}}_\infty$), the edges $\tilde{e}_j$, $j \in \tilde{S}_k \setminus \tilde{E}_\infty$, are each chosen independently and uniformly from the collection of sets of size k that would not have resulted in j being picked as some $\tilde{\jmath}(w)$. The crucial observation in this case is that this collection is $\tilde{\mathcal{F}}_\infty$-measurable and includes all k-subsets of $M \setminus \tilde{V}_\infty$; cf. the proof of Lemma 6.8. Thus, the (conditional) probability that $\tilde{e}_j$ meets $\tilde{V}_\infty$ is at most the probability that a uniformly chosen k-set of M meets $\tilde{V}_\infty$.

The events $\{\tilde{e}_j \cap \tilde{V}_\infty \neq \emptyset\}$, $j \in \tilde{S}_k \setminus \tilde{E}_\infty$, are therefore conditionally independent given $\tilde{\mathcal{F}}_\infty$, and each has (conditional) probability at most

$$\frac{|\tilde{V}_\infty|\binom{m}{k-1}}{\binom{m}{k}} = \frac{k|\tilde{V}_\infty|}{m-k}.$$

As $|\tilde{S}_k \setminus \tilde{E}_\infty| \leqslant s_k$, $\tilde{R}(k)$ is stochastically dominated by $Z(k)$. The $\tilde{R}(k)$ are conditionally independent given $\tilde{\mathcal{F}}_\infty$ as they depend on disjoint sets $\tilde{S}_k \setminus \tilde{E}_\infty$ of random edges that are themselves conditionally independent given $\tilde{\mathcal{F}}_\infty$. $\square$

We are finally ready to prove the key lemma of this section.

Proof of Lemma 6.4. Recall that

$$\tilde{W}_\infty := \sum_{w \in \tilde{V}_\infty} |\tilde{e}_{\tilde{\jmath}(w)}| = \sum_{j \in \tilde{E}_\infty} d_j$$

is the sum of the sizes of the edges used in Algorithm 6.5. Observe that

$$\min\left\{\tilde{R}_1, m^{1/2}\right\} \leqslant \tilde{W}_\infty + \sum_{k=2}^{\infty} k \cdot \tilde{R}(k),$$

since each edge counted by $\tilde{R}(k)$ can contribute at most k to $\tilde{R}_1$, so this holds when $|\tilde{V}_\infty| < m^{1/2}$, and $\tilde{W}_\infty \geqslant |\tilde{V}_\infty| \geqslant m^{1/2}$ otherwise. Recall that $\tilde{W}_\infty$ is $\tilde{\mathcal{F}}_\infty$-measurable and that $\mathcal{K}(z)$ implies $\tilde{R}(k) = 0$ for all $k \geqslant 4u_0$. Given $\tilde{\mathcal{F}}_\infty$, the random variables $\tilde{R}(k)$ are conditionally independent, so Lemma 6.12 implies that

$$\begin{aligned}(92)\quad \mathbb{E}\Big[e^{\lambda \min\{\tilde{R}_1, m^{1/2}\}} \mid \tilde{R}_1 > 1\Big] &\leqslant \mathbb{E}\Big[\mathbb{E}\Big[e^{\lambda \tilde{W}_\infty + \sum_{k \geqslant 2} \lambda k \tilde{R}(k)} \mid \tilde{\mathcal{F}}_\infty\Big] \,\Big|\, \tilde{R}_1 > 1\Big] \\ &= \mathbb{E}\Big[e^{\lambda \tilde{W}_\infty} \prod_{k=2}^{4u_0} \mathbb{E}\Big[e^{\lambda k \tilde{R}(k)} \mid \tilde{\mathcal{F}}_\infty\Big] \,\Big|\, \tilde{R}_1 > 1\Big].\end{aligned}$$

Moreover, it follows from Lemma 6.12 and Observation 5.7 that

$$\mathbb{E}\big[e^{\lambda k\tilde{R}(k)} \mid \tilde{\mathcal{F}}_\infty\big] \leqslant \mathbb{E}\big[e^{\lambda k Z(k)} \mid \tilde{\mathcal{F}}_\infty\big] \leqslant \exp\bigg(\big(e^{\lambda k}-1\big)\frac{k|\tilde{V}_\infty|}{m-k}\cdot s_k\bigg),$$

since $Z(k)$ is a sum of s_k independent Bernoulli random variables, each of expectation $k|\tilde{V}_\infty|/(m-k)$. Since $e^w-1=(1-e^{-w})e^w \leqslant we^w$ for all $w \geqslant 0$, and $\sum_{k\geqslant 2} 2^k s_k = O(m)$ (since $\mathcal{K}(z)$ holds), it follows that

$$\sum_{k=2}^{4u_0}\big(e^{\lambda k}-1\big)\frac{k|\tilde{V}_\infty|}{m-k}\cdot s_k \leqslant \frac{|\tilde{V}_\infty|}{m-4u_0}\sum_{k=2}^{4u_0}\lambda k^2 e^{\lambda k}s_k = O(\lambda|\tilde{V}_\infty|)$$

for $\lambda \leqslant \varepsilon_1 < \log 2$. Thus, recalling that $\tilde{W}_\infty \geqslant |\tilde{V}_\infty|$, we have

$$\begin{aligned}\mathbb{E}\big[e^{\lambda\min\{\tilde{R}_1,m^{1/2}\}} \mid \tilde{R}_1 > 1\big] &\leqslant \mathbb{E}\big[e^{\lambda\tilde{W}_\infty+O(\lambda|\tilde{V}_\infty|)} \mid \tilde{R}_1 > 1\big]\\ &= \mathbb{E}\big[e^{O(\lambda)\tilde{W}_\infty} \mid \tilde{R}_1 > 1\big]\end{aligned} \tag{93}$$

for all $0<\lambda\leqslant\varepsilon_1$. Hence, by Lemma 6.11, it follows that

$$\mathbb{E}\big[e^{\lambda\min\{\tilde{R}_1,m^{1/2}\}} \mid \tilde{R}_1 > 1\big] \leqslant \mathbb{E}\big[e^{O(\lambda)\tilde{W}_\infty} \mid \tilde{R}_1 > 1\big] = O(1)$$

for sufficiently small $\lambda > 0$. Now, by Observation 6.10, $\mathbb{P}(\tilde{R}_1 > 1) = \mathbb{P}(|\tilde{V}_1| > 1) \leqslant 2s_2/m$. Thus

$$\mathbb{P}(\tilde{R}_1 \geqslant t) \leqslant e^{-\lambda t}\,\mathbb{E}\big[e^{\lambda\min\{\tilde{R}_1,m^{1/2}\}} \mid \tilde{R}_1 > 1\big]\,\mathbb{P}\big(\tilde{R}_1 > 1\big) = \frac{O(s_2)}{m}e^{-\lambda t}$$

for all sufficiently small $\lambda > 0$ and all $2 \leqslant t \leqslant m^{1/2}$, as required.

For the second part, observe first that, as noted in (87), we have

$$\begin{aligned}\mathbb{E}\big[\Delta_k^{(1)}\mathbb{1}_{\{t\leqslant\tilde{R}_1<m^{1/2}\}}\big] &\leqslant e^{-\lambda t}\cdot\mathbb{E}\big[\Delta_k^{(1)}e^{\lambda\tilde{R}_1}\mathbb{1}_{\{2\leqslant\tilde{R}_1<m^{1/2}\}}\big]\\ &\leqslant e^{-\lambda t}\cdot\mathbb{E}\big[\Delta_k^{(1)}e^{\lambda\tilde{R}_1}\mathbb{1}_{\{\tilde{R}_1<m^{1/2}\}} \mid \tilde{R}_1 > 1\big].\end{aligned}$$

Recall that $\tilde{\Delta}_k^{(1)} = |\{j\in\tilde{S}_k : |\tilde{e}_j\cap\tilde{D}_\infty| = 1\}|$, and note that therefore

$$\mathbb{E}\Big[\Delta_k^{(1)}e^{\lambda\tilde{R}_1}\mathbb{1}_{\{\tilde{R}_1<m^{1/2}\}} \mid \tilde{R}_1 > 1\Big] \leqslant \sum_{j\in\tilde{S}_k}\mathbb{E}\Big[\mathbb{1}_{\{|\tilde{e}_j\cap\tilde{V}_\infty|=1\}}e^{\lambda\min\{\tilde{R}_1,m^{1/2}\}}\,\Big|\,\tilde{R}_1 > 1\Big],$$

since $\tilde{V}_\infty = \tilde{D}_\infty$ when $\tilde{R}_1 < m^{1/2}$. Now, repeating the argument of (92)–(93), we obtain

$$\mathbb{E}\Big[\mathbb{1}_{\{|\tilde{e}_j\cap\tilde{V}_\infty|=1\}}e^{\lambda\min\{\tilde{R}_1,m^{1/2}\}}\,\Big|\,\tilde{R}_1 > 1\Big] \leqslant \mathbb{E}\Big[\mathbb{1}_{\{|\tilde{e}_j\cap\tilde{V}_\infty|=1\}}e^{\lambda+O(\lambda)\tilde{W}_\infty}\,\Big|\,\tilde{R}_1 > 1\Big]$$

for every $j\in\tilde{S}_k$ and $0<\lambda\leqslant\varepsilon_1$, since $|\tilde{e}_j\cap\tilde{V}_\infty| = 1$ and $k\geqslant 3$ imply that $j\notin\tilde{E}_\infty$ (and that $\tilde{e}_j$ contributes exactly one to $\tilde{R}_1$), and the edges not used in

the algorithm are conditionally independent given $\tilde{\mathcal{F}}_\infty$. Combining the above inequalities, we obtain

$$\begin{aligned}\mathbb{E}\big[\Delta_k^{(1)}\mathbb{1}_{\{t\leqslant \tilde{R}_1<m^{1/2}\}}\big] &\leqslant e^{-\lambda t}\sum_{j\in\tilde{S}_k}\mathbb{E}\Big[\mathbb{1}_{\{|\tilde{e}_j\cap\tilde{V}_\infty|=1\}}e^{\lambda+O(\lambda)\tilde{W}_\infty}\,\Big|\,\tilde{R}_1>1\Big]\\ &\leqslant e^{-\lambda t}\sum_{j\in\tilde{S}_k}\mathbb{E}\left[\frac{k|\tilde{V}_\infty|}{m-k}e^{\lambda+O(\lambda)\tilde{W}_\infty}\;\Big|\;\tilde{R}_1>1\right];\end{aligned}$$

cf. the proof of Lemma 6.12. Since $|\tilde{S}_k|=s_k$ and $|\tilde{V}_\infty|\leqslant\tilde{W}_\infty=O(e^{\lambda\tilde{W}_\infty})$, it follows by Lemma 6.11 that

$$\mathbb{E}\big[\Delta_k^{(1)}\mathbb{1}_{\{t\leqslant\tilde{R}_1<m^{1/2}\}}\big]\leqslant\frac{O(ks_k)}{m}e^{-\lambda t}\cdot\mathbb{E}\big[e^{O(\lambda)\tilde{W}_\infty}\mid\tilde{R}_1>1\big]=\frac{O(ks_k)}{m}e^{-\lambda t}$$

for all sufficiently small $\lambda>0$, as required. □

Having done the hard part, it is now relatively straightforward to deduce Theorem 6.2, using Theorem 5.1. The next step is to use Lemmas 6.4, 6.8 and 6.12 to deduce the following estimates for the other quantities introduced in Definition 6.3.

LEMMA 6.13. *Suppose that $\mathcal{E}$ is such that $\mathcal{K}(z)$ holds, $d(z)=1$, and $2s_2\leqslant(1-\varepsilon_1)m$. Then, in the independent random hypergraph model,*

(a) $\mathbb{E}\big[\tilde{D}\big]=\left(1-\dfrac{2s_2}{m}+o(1)\right)^{-1}$,

(b) $\mathbb{E}\big[\tilde{\Delta}_k\big]=\left(1-\dfrac{2s_2}{m}+o(1)\right)^{-1}\dfrac{ks_k}{m}+O\big(m^{-1/2}\big)$ *for all* $k\geqslant 2$,

(c) $\mathbb{E}\big[\tilde{D}^2\big]=O(1)$,

(d) $\mathbb{E}\big[\tilde{\Delta}'\big]=O(m^{-1/2})$,

where the bounds implicit in the $o(\cdot)$ and $O(\cdot)$ notation are uniform in k and z.

Proof. Parts (a) and (c) are easier: we shall prove them first. Part (c) follows immediately from Lemma 6.4, since $\tilde{D}\leqslant\tilde{R}_1$, so

$$\mathbb{E}\big[\tilde{D}^2\big]\leqslant\mathbb{E}\big[\tilde{R}_1^2\big]\leqslant\sum_{t=1}^{\infty}t^2\cdot\mathbb{P}(\tilde{R}_1\geqslant t)=O(1),$$

as required.

To prove (a), recall that $\tilde{X}_t=|\tilde{V}_t\setminus\tilde{V}_{t-1}|=\sum_w\tilde{X}_{t,w}$ and that $\tilde{\mathcal{X}}_{t,w}$ is the event that $\tilde{V}^{(\ell)}<m^{1/2}$ just before we test whether or not $\tilde{X}_{t,w}=1$. Also, by Lemma 6.8, we have

$$\mathbb{E}\big[\tilde{X}_{t,w}\mid\tilde{\mathcal{F}}_{t,<w}\big]=\left(\frac{2s_2}{m^2}+O\big(m^{-3/2}\big)\right)\mathbb{1}_{\tilde{\mathcal{X}}_{t,w}}\tag{94}$$

for every $t \in \mathbb{N}$ and $w \in M \setminus \tilde{V}_{t-1}$. Let us define, for each $t \geqslant 1$, an $\tilde{\mathcal{F}}_{t-1}$-measurable event

$$\tilde{\mathcal{X}}_t := \left\{ \mathbb{P}\Big(|\tilde{V}_t| \geqslant m^{1/2} \mid \tilde{\mathcal{F}}_{t-1}\Big) \leqslant \frac{1}{m^2} \right\},$$

and observe that if $\tilde{\mathcal{X}}_t$ holds, then $\mathbb{E}\big[\mathbb{1}_{\tilde{\mathcal{X}}_{t,w}^c} \mid \tilde{\mathcal{F}}_{t-1}\big] \leqslant 1/m^2$ for each $w \in M \setminus \tilde{V}_{t-1}$, which in turn implies that

$$\begin{aligned} \mathbb{E}\big[\tilde{X}_t \mid \tilde{\mathcal{F}}_{t-1}\big] &= \sum_{w \in M \setminus \tilde{V}_{t-1}} \mathbb{E}\big[\tilde{X}_{t,w} \mid \tilde{\mathcal{F}}_{t-1}\big] \\ &= \frac{2s_2 + o(m)}{m^2} \cdot |M \setminus \tilde{V}_{t-1}| = \frac{2s_2}{m} + o(1), \end{aligned} \tag{95}$$

by (94) and since $|\tilde{V}_{t-1}| = o(m)$. Observe also that, by Lemma 6.4, we have

$$\mathbb{P}\big(\tilde{\mathcal{X}}_t^c\big) \leqslant m^2 \mathbb{P}\big(|\tilde{V}_\infty| \geqslant m^{1/2}\big) \leqslant m^2 \mathbb{P}\big(\tilde{R}_1 \geqslant m^{1/2}\big) = O\big(m^2 e^{-\lambda m^{1/2}}\big), \tag{96}$$

since $\tilde{D} \leqslant \tilde{R}_1$ and $|\tilde{V}_\infty| = \min\big\{\tilde{D}, \lceil m^{1/2} \rceil\big\}$, by Observation 6.6.

We now define sequences $\tilde{Y}_t^+$ and $\tilde{Y}_t^-$ as follows. Let $0 < \varepsilon < \varepsilon_1$ be an arbitrarily small constant, set $\tilde{Y}_0^+ = 1$ and $\tilde{Y}_0^- = -1$, and for each $1 \leqslant t \leqslant |\tilde{V}_\infty|$, define

$$\tilde{Y}_t^+ := \begin{cases} |\tilde{A}_t| + \left(1 - \dfrac{2s_2}{m} - \varepsilon\right) t & \text{if } \tilde{\mathcal{X}}_t \text{ holds,} \\ \tilde{Y}_{t-1}^+ & \text{otherwise.} \end{cases}$$

Similarly, define $\tilde{Y}_t^- := -|\tilde{A}_t| - \big(1 - \frac{2s_2}{m} + \varepsilon\big)t$ if $\tilde{\mathcal{X}}_t$ holds, and $\tilde{Y}_t^- := \tilde{Y}_{t-1}^-$ otherwise. We claim that $\tilde{Y}_t^+$ and $\tilde{Y}_t^-$ are both super-martingales with respect to the filtration $(\tilde{\mathcal{F}}_t)_{t \geqslant 0}$. Indeed, for each $1 \leqslant t \leqslant |\tilde{V}_\infty|$, if $\tilde{\mathcal{X}}_t$ holds, then

$$\mathbb{E}\big[\tilde{Y}_t^+ - \tilde{Y}_{t-1}^+ \mid \tilde{\mathcal{F}}_{t-1}\big] = \mathbb{E}\big[|\tilde{A}_t| - |\tilde{A}_{t-1}| \mid \tilde{\mathcal{F}}_{t-1}\big] + \left(1 - \frac{2s_2}{m} - \varepsilon\right) \leqslant 0$$

by (95), since $|\tilde{A}_t| - |\tilde{A}_{t-1}| = \tilde{X}_t - 1$, and similarly for $\tilde{Y}_t^-$. Therefore, by the optional stopping theorem applied to $\tilde{Y}_t^+$, and writing $\tilde{Y}_\infty^+ := \tilde{Y}_{|\tilde{V}_\infty|}^+$, it follows that

$$\mathbb{E}\big[\tilde{Y}_\infty^+\big] \leqslant \tilde{Y}_0^+ = 1.$$

Now, recalling that $\tilde{A}_{|\tilde{V}_\infty|} = 0$, it follows that

$$\tilde{Y}_\infty^+ = \left(1 - \frac{2s_2}{m} - \varepsilon\right) |\tilde{V}_\infty|$$

if $\tilde{\mathcal{X}}_t$ holds for every $1 \leqslant t \leqslant |\tilde{V}_\infty|$, and in general $|\tilde{Y}_\infty^+| \leqslant m^{1/2}$, by Observation 6.6. Recall also that, by Observation 6.6, we have $\tilde{D} = |\tilde{V}_\infty|$ if

$|\tilde{V}_\infty| < m^{1/2}$, and otherwise $\tilde{D} \leqslant m$. Hence, by (96), it follows that

$$\mathbb{E}\big[\tilde{Y}_\infty^+\big] \geqslant \bigg(1 - \frac{2s_2}{m} - 2\varepsilon\bigg)\mathbb{E}\big[\tilde{D}\big].$$

Applying the same argument to $\tilde{Y}_t^-$, we obtain $1 \leqslant \big(1 - \frac{2s_2}{m} + 2\varepsilon\big)\mathbb{E}\big[\tilde{D}\big]$. Since $\varepsilon > 0$ was arbitrary, and $2s_2 \leqslant (1-\varepsilon_1)m$, it follows that

$$\mathbb{E}\big[\tilde{D}\big] = \bigg(1 - \frac{2s_2}{m} + o(1)\bigg)^{-1},$$

as required.

To prove parts (b) and (d), we shall need to separately consider edges of $\tilde{S}_k \cap \tilde{E}_\infty$ (i.e., edges that are used in the process) and edges of $\tilde{S}_k \setminus \tilde{E}_\infty$ that intersect $\tilde{V}_\infty$. We will show that $\tilde{E}_\infty$ is unlikely to contain any edges of size at least three but is expected to have the required number of edges of size 2.

The proof of part (b) in the case $k = 2$ is very similar to that of part (a), so we shall be somewhat brief with the details. Let $\lambda > 0$ be a sufficiently large constant, set $\tilde{Z}_0^+ = 0$, and for each $1 \leqslant t \leqslant |\tilde{V}_\infty|$, define

$$\tilde{Z}_t^+ := \begin{cases} |\tilde{S}_2 \cap \tilde{E}_t| - \Big(\dfrac{2s_2}{m} + \dfrac{\lambda}{m^{1/2}}\Big) \cdot t & \text{if } \tilde{\mathcal{X}}_t \text{ holds,} \\ \tilde{Z}_{t-1}^+ & \text{otherwise.} \end{cases}$$

Similarly, define $\tilde{Z}_t^- := -|\tilde{S}_2 \cap \tilde{E}_t| + \big(\frac{2s_2}{m} - \lambda m^{-1/2}\big)t$ if $\tilde{\mathcal{X}}_t$ holds, and $\tilde{Z}_t^- := \tilde{Z}_{t-1}^-$ otherwise. Recall that $\tilde{X}_{t,w}^{(k)} = 1$ if and only if $w \in \tilde{V}_t \setminus \tilde{V}_{t-1}$ and $\tilde{\jmath}(w) \in \tilde{S}_k$, and that

$$\mathbb{E}\big[\tilde{X}_{t,w}^{(2)} \mid \tilde{\mathcal{F}}_{t,<w}\big] = \bigg(\frac{2s_2}{m^2} + O\big(m^{-3/2}\big)\bigg)\, \mathbb{1}_{\tilde{\mathcal{X}}_{t,w}}$$

for every $t \in \mathbb{N}$ and $w \in M \setminus \tilde{V}_{t-1}$, by Lemma 6.8. It follows that

$$\mathbb{E}\big[\tilde{Z}_t^+ - \tilde{Z}_{t-1}^+ \mid \tilde{\mathcal{F}}_{t-1}\big] = \mathbb{E}\big[|\tilde{S}_2 \cap \tilde{E}_t| - |\tilde{S}_2 \cap \tilde{E}_{t-1}| \mid \tilde{\mathcal{F}}_{t-1}\big] - \bigg(\frac{2s_2}{m} + \frac{\lambda}{m^{1/2}}\bigg) \leqslant 0$$

if $\tilde{\mathcal{X}}_t$ holds, since $s_2 = O(m)$. Thus $\tilde{Z}_t^+$ is a super-martingale with respect to the filtration $(\tilde{\mathcal{F}}_t)_{t \geqslant 0}$, and hence, by the optional stopping theorem,

$$\mathbb{E}\big[\tilde{Z}_\infty^+\big] \leqslant \tilde{Z}_0^+ = 0,$$

where $\tilde{Z}_\infty^+ := \tilde{Z}_{|\tilde{V}_\infty|}^+$. Now,

$$\tilde{Z}_\infty^+ = |\tilde{S}_2 \cap \tilde{E}_\infty| - \bigg(\frac{2s_2}{m} + \frac{\lambda}{m^{1/2}}\bigg)|\tilde{V}_\infty| \tag{97}$$

if $\tilde{\mathcal{X}}_t$ holds for every $1 \leqslant t \leqslant |\tilde{V}_\infty|$, and $|\tilde{Z}_\infty^+| \leqslant s_2 \leqslant m$ otherwise. Thus, by (96), it follows that

$$\mathbb{E}\big[|\tilde{S}_2 \cap \tilde{E}_\infty|\big] \leqslant \bigg(\frac{2s_2}{m} + \frac{\lambda}{m^{1/2}}\bigg)\mathbb{E}\big[|\tilde{V}_\infty|\big] + \frac{1}{m^2}. \tag{98}$$

Now, using part (a) and recalling that $\varepsilon > 0$ was arbitrary and $2 \leqslant 2s_2 \leqslant (1-\varepsilon_1)m$ (since if $s_2 = 0$ then we trivially have $\mathbb{E}[\tilde{\Delta}_2] = 0$), it follows that

$$\mathbb{E}\big[|\tilde{S}_2 \cap \tilde{E}_\infty|\big] \leqslant \frac{2s_2}{m}\bigg(1 - \frac{2s_2}{m} + o(1)\bigg)^{-1} + O\big(m^{-1/2}\big).$$

Repeating the argument for $\tilde{Z}_t^-$, we obtain a corresponding lower bound, and hence

$$\mathbb{E}\big[|\tilde{S}_2 \cap \tilde{E}_\infty|\big] = \frac{2s_2}{m}\bigg(1 - \frac{2s_2}{m} + o(1)\bigg)^{-1} + O\big(m^{-1/2}\big). \tag{99}$$

When $k \geqslant 3$, we only need a weaker bound, and as a consequence the argument is simpler. Recall first that

$$\mathbb{E}\big[\tilde{X}_{t,w}^{(k)} \mid \tilde{\mathcal{F}}_{t,<w}\big] \leqslant \frac{2k^2 s_k}{m^{(k+2)/2}} \mathbb{1}_{\tilde{\mathcal{X}}_{t,w}}$$

for every $t \in \mathbb{N}$, $k \geqslant 3$ and $w \in M \setminus V_{t-1}$, by Lemma 6.8. Note also that

$$|\tilde{S}_k \cap \tilde{E}_t| = |\tilde{S}_k \cap \tilde{E}_{t-1}| + \sum_{w \in M \setminus \tilde{V}_{t-1}} \tilde{X}_{t,w}^{(k)},$$

and therefore $\tilde{U}_t^{(k)} := |\tilde{S}_k \cap \tilde{E}_t| - 2k^2 s_k m^{-k/2} t$ is a super-martingale, since

$$\mathbb{E}\big[\tilde{U}_t^{(k)} - \tilde{U}_{t-1}^{(k)} \mid \tilde{\mathcal{F}}_{t-1}\big] = \mathbb{E}\big[|\tilde{S}_k \cap \tilde{E}_t| - |\tilde{S}_k \cap \tilde{E}_{t-1}| \mid \tilde{\mathcal{F}}_{t-1}\big] - \frac{2k^2 s_k}{m^{k/2}} \leqslant 0$$

for every $t \geqslant 1$. Hence, by the optional stopping theorem, we have

$$\mathbb{E}\big[|\tilde{S}_k \cap \tilde{E}_\infty|\big] \leqslant \frac{2k^2 s_k}{m^{k/2}} \cdot \mathbb{E}\big[|\tilde{V}_\infty|\big]. \tag{100}$$

Since $\mathbb{E}\big[|\tilde{V}_\infty|\big] = O(1)$ (by Lemma 6.4, or by part (a)), it follows, using the event $\mathcal{K}(z)$, that

$$\sum_{k=3}^{\infty} \mathbb{E}\big[|\tilde{S}_k \cap \tilde{E}_\infty|\big] = O\big(m^{-1/2}\big). \tag{101}$$

We shall next deal with edges that are not used in the process, but nevertheless intersect $\tilde{V}_\infty$ in at least two vertices. To be precise, we shall show that the expected size of

$$\tilde{R}_2(k) := \Big|\big\{j \in \tilde{S}_k \setminus \tilde{E}_\infty : |\tilde{e}_j \cap \tilde{V}_\infty| \geqslant 2\big\}\Big|$$

is small for each $2 \leqslant k \leqslant 4u_0$. Indeed, recall (see Remark 6.7 and the proofs of Lemmas 6.8 and 6.12) that conditional on the information we have observed during the process, each edge $\tilde{e}_j$, $j \in \tilde{S}_k \setminus \tilde{E}_\infty$, is chosen uniformly from a collection of sets of size k that depends on j but always includes all k-subsets

that contain at least two elements of $M \setminus \tilde{V}_\infty$. The (conditional) probability of the event $\{|\tilde{e}_j \cap \tilde{V}_\infty| \geqslant 2\}$ is therefore at most

$$\binom{|\tilde{V}_\infty|}{2}\binom{m}{k-2}\binom{m-|\tilde{V}_\infty|}{k}^{-1} \leqslant \frac{k^2 \cdot |\tilde{V}_\infty|^2}{m^2} \tag{102}$$

since $|\tilde{V}_\infty| \leqslant m^{1/2}$, so $k \cdot |\tilde{V}_\infty| = o(m)$. Note also that if $k \geqslant 3$, then the (conditional) probability of the event $\{|\tilde{e}_j \cap \tilde{V}_\infty| = 1\}$ is at least

$$|\tilde{V}_\infty|\binom{m-|\tilde{V}_\infty|}{k-1}\binom{m}{k}^{-1} = \big(1+o(1)\big)\frac{k \cdot |\tilde{V}_\infty|}{m}.$$

Thus the expected number of edges of size k that intersect $\tilde{V}_\infty$ is at least

$$\big(1+o(1)\big)\frac{ks_k}{m}\mathbb{E}\big[|\tilde{V}_\infty|\big], \tag{103}$$

which together with part (a) proves the lower bound of part (b) when $k \geqslant 3$.

Now, using part (c) and $\mathcal{K}(z)$ to bound $\mathbb{E}\big[|\tilde{V}_\infty|^2\big] = O(1)$ and $s_k = O(m)$, respectively, it follows from (102) that

$$\mathbb{E}[\tilde{R}_2(k)] \leqslant \frac{k^2 s_k}{m^2} \cdot \mathbb{E}\big[|\tilde{V}_\infty|^2\big] = O\bigg(\frac{k^2 s_k}{m^2}\bigg). \tag{104}$$

Note that if $|\tilde{V}_\infty| < m^{1/2}$, then every edge of size 2 is (by definition) either contained in or disjoint from $\tilde{D}_\infty$, and that otherwise $\tilde{\Delta}_2 = O(m)$, by $\mathcal{K}(z)$. Hence, combining (104) with (99), and using (96), we obtain the case $k = 2$ of part (b). Moreover, combining (104) with (101), and recalling that $\sum_{k \geqslant 2} k^2 s_k = O(m)$, by $\mathcal{K}(z)$, it follows that

$$\mathbb{E}\big[\tilde{\Delta}'\big] = O\big(m^{-1}\big) + O\big(m^{-1/2}\big) = O\big(m^{-1/2}\big),$$

which proves (d).

It only remains to prove the upper bound in part (b) when $k \geqslant 3$. By (100), it is sufficient to show that the expected number of edges of $\tilde{S}_k \setminus \tilde{E}_\infty$ that intersect $\tilde{V}_\infty$ in at least one vertex is at most

$$\bigg(1 - \frac{2s_2}{m} + o(1)\bigg)^{-1}\frac{ks_k}{m}$$

for each $k \geqslant 3$. This follows by Lemma 6.12 and part (a), since

$$\mathbb{E}\big[Z(k)\big] = \mathbb{E}\Big[\mathbb{E}\big[Z(k) \mid \tilde{\mathcal{F}}_\infty\big]\Big] = \big(1+o(1)\big)\frac{ks_k}{m}\mathbb{E}\big[|\tilde{V}_\infty|\big]$$

for every $3 \leqslant k \leqslant 4u_0$. This completes the proof of part (b), and therefore of the lemma. □

We are finally ready to deduce Theorem 6.2. The key observation, which allows us to apply Theorem 5.1, is that each of the variables $D(z)$, $R_1(z)$, $\Delta_k(z)$ and $\Delta'(z)$ depends only on $\mathcal{F}_z^+$ and the submatrix $R(I) := A\big[I \times [z, \pi(x)]\big]$,

where I is the set of rows that are removed in step z. Note also that we can deduce from $\mathcal{F}_z^+$ and $R(I)$ whether or not I is the set of removed rows. Indeed, in Algorithm 2.7 we only need to look at the entries of a row once we know we shall remove it. The event $\{D(z) = t\}$ (for example) is therefore a disjoint union of events of the form $\big\{A\big[I \times [z, \pi(x)]\big] = R\big\}$ with R having exactly t rows.

It follows that we would be able to deduce Theorem 6.2 from Lemmas 6.4 and 6.13 if we could restrict to events $\mathcal{E} \in \mathcal{F}_z^+$ and $\big\{A\big[I \times [z, \pi(x)]\big] = R\big\}$ for which

$$\mathbb{P}\big(A\big[I \times [z, \pi(x)]\big] = R \mid \mathcal{E}\big) = \big(1 + o(1)\big)\mathbb{P}\big(\tilde{A}_{\mathcal{E}}\big[I \times [z, \pi(x)]\big] = R\big).$$

Theorem 5.1 provides us with such a bound as long as $|R|_1$ is not too large, since $|I| \leqslant |R|_1 \leqslant |R|_2 \leqslant (|R|_1)^2$ for the matrices we shall be dealing with. Moreover, observe that if I is the set of removed rows, then $|R(I)|_1 = R_1(z)$, so we shall be able to use Theorem 5.1 and Lemma 6.4 to bound from above the probability that $|R|_1$ is large.

Proof of Theorem 6.2. We partition the probability space into three pieces, using the events

$$\mathcal{D}_1 = \big\{R_1(z) < u_0^{1/3}\big\}, \quad \mathcal{D}_2 = \big\{u_0^{1/3} \leqslant R_1(z) < u_0^2\big\}, \quad \mathcal{D}_3 = \big\{R_1(z) \geqslant u_0^2\big\}.$$

Let us fix $z \in [z_-, \pi(x)]$ and an event $\mathcal{E} \in \mathcal{F}_z^+$ of the form (50) such that $\mathcal{K}(z)$ holds, $2s_2(z) \leqslant (1 - \varepsilon_1)m(z)$ and $d(z) = 1$, and say that an $I \times [z, \pi(x)]$ matrix R is 1-*acceptable* (with respect to $\mathcal{E}$) if it is consistent with $\mathcal{D}_1 \cap \mathcal{E}$ and if the event $\big\{A\big[I \times [z, \pi(x)]\big] = R\big\} \cap \mathcal{E}$ implies that I is the set of removed rows in step z.[12] Note that $\mathcal{D}_1$ is the disjoint union of the events $\big\{A\big[I \times [z, \pi(x)]\big] = R\big\}$ over the family $\mathcal{U}_1$ of all 1-acceptable matrices R. We claim that, for every 1-acceptable matrix $R \in \mathcal{U}_1$, we have

$$\mathbb{P}\big(A\big[I \times [z, \pi(x)]\big] = R \mid \mathcal{E}\big) = \big(1 + o(1)\big)\mathbb{P}\big(\tilde{A}_{\mathcal{E}}\big[I \times [z, \pi(x)]\big] = R\big). \tag{105}$$

Indeed, this follows by applying Theorem 5.1 since

$$\frac{|I| + |R|_2}{u_0} = o(1),$$

which holds because every row of R is non-empty, so $|I| \leqslant |R|_2 \leqslant (|R|_1)^2 < u_0^{2/3}$, and also every row sum of R is at most $|R|_1 = o(u_0)$.

We shall next prove that $\mathbb{P}(\mathcal{D}_3 \mid \mathcal{F}_z, d(z) = 1) \leqslant z_0^{-20}$. To do so, let us say that an $I \times [z, \pi(x)]$ matrix R is 3-*acceptable* (with respect to $\mathcal{E}$) if it is consistent with $\mathcal{E}$ and if I is the set of rows removed in Algorithm 2.7 at

[12] To be precise, an $I \times [z, \pi(x)]$ matrix R with this latter property is said to be consistent with $\mathcal{D}_1 \cap \mathcal{E}$ if and only if the column sums of R satisfy $\sum_{i \in I} R_{ij} \leqslant d_j$ for each $j \in [z, \pi(x)]$, and $|R|_1 < u_0^{1/3}$.

the first point at which $\mathcal{D}_3$ is guaranteed to hold, i.e., the first point at which $|R|_1 \geqslant u_0^2$. Thus $\mathcal{D}_3$ is the disjoint union of the events $\big\{A\big[I \times [z, \pi(x)]\big] = R\big\}$ over the family $\mathcal{U}_3$ of all 3-acceptable matrices R. Now, by Theorem 5.1, we have

$$\frac{\mathbb{P}\big(A\big[I \times [z, \pi(x)]\big] = R \mid \mathcal{E}\big)}{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times [z, \pi(x)]\big] = R\big)} \leqslant \exp\left(\frac{O\big(|I| + |R|_1\big)}{u_0}\right) = e^{O(u_0)}$$

for any 3-acceptable matrix R, since $|I| \leqslant |R|_1 = O(u_0^2)$. Indeed, we have $|I| \leqslant |R|_1$ (as before) since R has no empty rows, and $|R|_1 = O(u_0^2)$ since I is minimal, and since each row of R contains at most $2u_0$ non-zero entries (by Observation 3.12), so each row can increase $|R|_1$ by at most $O(u_0)$.

Now, by Lemma 6.4, we have

$$\sum_{R \in \mathcal{U}_3} \mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times [z, \pi(x)]\big] = R\big) = \mathbb{P}\big(\tilde{R}_1 \geqslant u_0^2\big) = O\big(e^{-\lambda u_0^2}\big),$$

and hence, noting that $\log z_0 = o(u_0^2)$, by (23), we obtain

$$\mathbb{P}\big(\mathcal{D}_3 \mid \mathcal{F}_z,\, d(z) = 1\big) \leqslant \exp\big(O(u_0) - \lambda u_0^2\big) \leqslant z_0^{-20},$$

as claimed, which proves (83). Note that this also implies that

$$\mathbb{E}\big[R_1(z)^2 \mathbb{1}_{\mathcal{D}_3} \mid \mathcal{F}_z,\, d(z) = 1\big] \leqslant z_0^{-9}, \tag{106}$$

since the event $\mathcal{K}(z)$ implies that $R_1(z)^2 \leqslant (4u_0 z_0^5)^2 \leqslant z_0^{11}$.

Finally, let $\mathcal{U}_2$ denote the family of (2-*acceptable* with respect to $\mathcal{E}$) $I \times [z, \pi(x)]$ matrices R that are consistent with $\mathcal{D}_2 \cap \mathcal{E}$ and are such that the event $\big\{A\big[I \times [z, \pi(x)]\big] = R\big\} \cap \mathcal{E}$ implies that I is the set of removed rows in step z. By Theorem 5.1, we have

$$\frac{\mathbb{P}\big(A\big[I \times [z, \pi(x)]\big] = R \mid \mathcal{E}\big)}{\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times [z, \pi(x)]\big] = R\big)} \leqslant \exp\left(\frac{O(|R|_1)}{u_0}\right)$$

for any 2-acceptable matrix R, since $|I| \leqslant |R|_1$, and hence

$$\begin{aligned}\mathbb{P}\big(\mathcal{D}_2 \mid \mathcal{F}_z,\, d(z) = 1\big) &\leqslant \sum_{R \in \mathcal{U}_2} \exp\left(\frac{O(|R|_1)}{u_0}\right) \mathbb{P}\big(\tilde{A}_{\mathcal{E}}[I \times [z, \pi(x)]\big] = R\big) \\ &\leqslant \sum_{t = u_0^{1/3}}^{u_0^2} e^{O(t/u_0)} \mathbb{P}\big(\tilde{R}_1 \geqslant t\big) \leqslant \exp\left(-\frac{\lambda u_0^{1/3}}{2}\right),\end{aligned}$$

by Lemma 6.4. Since (by definition) $R_1(z) \leqslant u_0^2$ if $\mathcal{D}_2$ holds, it follows that

$$\mathbb{E}\big[R_1(z)^2 \cdot \mathbb{1}_{\mathcal{D}_2} \mid \mathcal{F}_z,\, d(z) = 1\big] \leqslant u_0^4 \exp\left(-\frac{\lambda u_0^{1/3}}{2}\right) = o(1). \tag{107}$$

Combining (107) with (105), (106) and Lemma 6.13, this completes the proof of parts (a) and (c) of Theorem 6.2. Moreover, for part (d), we have

$$\begin{aligned}\mathbb{E}\big[\Delta'(z)\big] &= \mathbb{E}\big[\Delta'(z)\mathbb{1}_{\mathcal{D}_1}\big] + \mathbb{E}\big[\Delta'(z)\mathbb{1}_{\mathcal{D}_2}\big] + \mathbb{E}\big[\Delta'(z)\mathbb{1}_{\mathcal{D}_3}\big] \\ &\leqslant (1+o(1))\mathbb{E}\big[\tilde{\Delta}'\big] + e^{O(u_0)}\mathbb{E}\big[\tilde{\Delta}'\big] + z_0^{-9} = O\big(m(z)^{-1/3}\big),\end{aligned}$$

by Lemma 6.13, since $u_0 = o(\log m(z))$ and $m(z) \leqslant N = z_0^{2+o(1)}$.

For part (b), we need to take more care in the case that $s_k(z)$ is small. For $k \geqslant 3$, we have

$$\mathbb{E}\big[\Delta_k(z)\mathbb{1}_{\mathcal{D}_2}\big] \leqslant \sum_{t=u_0^{1/3}}^{u_0^2} \mathbb{E}\big[\Delta_k^{(1)}(z)\mathbb{1}_{\{R_1(z)=t\}}\big] + \mathbb{E}\big[\Delta'(z)\big],$$

and applying Theorem 5.1 to each 2-acceptable matrix R with $|R|_1 = t$ gives

$$\mathbb{E}\big[\Delta_k^{(1)}(z)\mathbb{1}_{\{R_1(z)=t\}}\big] \leqslant \mathbb{E}\big[\tilde{\Delta}_k^{(1)} e^{O(t/u_0)}\mathbb{1}_{\{\tilde{R}_1=t\}}\big] \leqslant \frac{O(ks_k(z))}{m(z)} e^{-\lambda t/2}$$

for each $u_0^{1/3} \leqslant t \leqslant u_0^2$, where the second inequality follows by Lemma 6.4. The claimed bound now follows by (105), (106) and Lemma 6.13, and using part (d) to bound $\mathbb{E}\big[\Delta'(z)\big]$. For $k = 2$, we note that $\Delta_2(z) \leqslant R_1(z)$, so by Theorem 5.1 and Lemma 6.4,

$$\mathbb{E}\big[\Delta_2(z)\mathbb{1}_{\{R_1(z)=t\}}\big] \leqslant \mathbb{E}\big[\tilde{R}_1 e^{O(t/u_0)}\mathbb{1}_{\{\tilde{R}_1=t\}}\big] \leqslant \frac{O(s_2(z))}{m(z)} e^{-\lambda t/2}.$$

Summing over $t \in [u_0^{1/3}, u_0^2]$ gives $\mathbb{E}[\Delta_2(z)\mathbb{1}_{\mathcal{D}_2}] = o(s_2(z)/m(z))$, so we are done as before. □

7. Tracking the process when most columns are empty

In this section we shall track $s_k(z)$ and $m(z)$ above the "critical" range $[z_-, z_+]$ by showing that for $z \geqslant z_+$, there are few edges in $\mathcal{S}_A(z)$ and as a consequence that $m(z) \approx \eta\Lambda(z)z$. For $z \geqslant z_0^3$, these results follow almost immediately from Lemma 4.12, so for most of this section we shall be interested in the range $z \in [z_+, z_0^3]$.

Set $\tilde{\delta} := (1-2\varepsilon_1)\eta\delta$ and, recalling Definition 2.5, define

$$\sigma_k := \frac{\varepsilon(k, z_+)}{k!} \cdot \frac{\tilde{\delta}^{k-1}}{2k} = \frac{\varepsilon_1^k}{\Lambda(z_+)} \cdot \frac{\tilde{\delta}^{k-1}}{2k}. \tag{108}$$

We remark that the fact that σ_k decreases only exponentially fast (as a function of k) will play an important role in the proofs of Lemmas 7.5 and 7.6, below. Note that Lemma 4.10 implies that $m(z_+) \geqslant m_0(z_+) \geqslant \tilde{\delta} z_+$ with high probability, since $\Lambda(z_+) \geqslant \delta$.

Recall that, by Definition 4.1, $\mathcal{M}^*(z) = \bigcap_{w \geqslant z} \mathcal{M}(z)$ implies that $m(w)$ is well controlled for all $w \geqslant z$, and $\mathcal{T}_k(z)$ implies that $s_k(z)$ is well controlled. Define $\mathcal{D}(z)$ to be the event that $m(z) - m(z-1) \leqslant u_0^2$, and set

$$\mathcal{D}^*(z) := \bigcap_{w=z+1}^{\pi(x)} \mathcal{D}(w). \tag{109}$$

Note that $\mathcal{D}^*(z)$ does *not* include the event $\mathcal{D}(z)$. We shall prove the following upper bound on $s_k(z)$ when $z \geqslant z_+$.

Proposition 7.1. *With high probability, $\mathcal{M}^*(z_+)$ and $\mathcal{D}^*(z_+)$ hold and, moreover,*

$$s_k(z) \leqslant \sigma_k m(z) \tag{110}$$

for every $k \geqslant 2$ and every $z \in [z_+, \pi(x)]$.

We remark that (110) is extremely weak at the beginning of the process but becomes progressively stronger as time goes on (i.e., as z decreases). We shall begin by showing that at $z = z_+$ it implies the events $\mathcal{T}_k(z_+)$.

Corollary 7.2. *If $s_k(z_+) \leqslant \sigma_k m(z_+)$ and $m(z_+) \geqslant \tilde{\delta} z_+$, then*

$$s_k(z_+) \in \left(1 \pm \frac{\varepsilon(k, z_+)}{2}\right) \hat{s}_k(z_+).$$

In particular, with high probability $\mathcal{T}_k(z_+)$ holds for every $k \geqslant 2$.

Note that our bound on $s_k(z_+)$ is slightly stronger than necessary for $\mathcal{T}_k(z_+)$ here. This is because we shall require the stronger bound in Section 9. To prove Corollary 7.2, we first note the following observation, which will also be used later in the proof of Lemma 7.9.

Observation 7.3. *For every $k \geqslant 2$, we have*

$$\frac{\varepsilon_1^{k+1} k!}{3^k} \geqslant \Lambda(z_+),$$

and hence $\varepsilon(k, z_+) \geqslant 3^k/\varepsilon_1$.

Proof. Note that the expression $\varepsilon_1^{k+1} k!/3^k$ is minimized by taking $k = 3/\varepsilon_1$ (which we have assumed is an integer) and that $k! \geqslant 2(k/e)^k$ for all $k \geqslant 2$. Thus

$$\frac{\varepsilon_1^{k+1} k!}{3^k} \geqslant 2\varepsilon_1 \left(\frac{3}{e\varepsilon_1}\right)^{3/\varepsilon_1} \left(\frac{\varepsilon_1}{3}\right)^{3/\varepsilon_1} = 2\varepsilon_1 e^{-3/\varepsilon_1} \geqslant 2\delta$$

by (10). The first result follows since $\Lambda(z_+) = \delta + o(1)$, by (39). It now follows immediately from Definition 2.5 that

$$\varepsilon(k, z_+) = \frac{\varepsilon_1^k k!}{\Lambda(z_+)} \geqslant \frac{3^k}{\varepsilon_1},$$

as required. □

Proof of Corollary 7.2, *assuming Proposition* 7.1. The lower bound on $s_k(z_+)$ holds trivially as $\varepsilon(k, z_+) \geqslant 2$ by Observation 7.3. To prove the upper bound, observe that

$$\frac{s_k(z_+)}{\varepsilon(k, z_+)} \leqslant \frac{\sigma_k m(z_+)}{\varepsilon(k, z_+)} = \frac{\tilde{\delta}^{k-1} m(z_+)}{k! \cdot 2k} \leqslant \frac{m(z_+)}{2(k-1)k!}\left(\frac{m(z_+)}{z_+}\right)^{k-1}, \tag{111}$$

by (108) and the assumption that $m(z_+) \geqslant \tilde{\delta} z_+$. Thus, by considering just the first term in the sum,

$$s_k(z_+) \leqslant \frac{\varepsilon(k, z_+) m(z_+)}{2k(k-1)} \sum_{\ell=k-1}^{\infty} \frac{1}{\ell!}\left(\frac{m(z_+)}{z_+}\right)^{\ell} = \frac{\varepsilon(k, z_+)}{2} \cdot \hat{s}_k(z_+),$$

as required.

The last part follows as by Proposition 7.1, $s_k(z_+) \leqslant \sigma_k m(z_+)$, and by Lemma 4.10, $m(z_+) \geqslant \tilde{\delta} z_+$, with high probability. □

Note that $m(z) - m(z-1) = D(z) \leqslant R_1(z)$ when $d(z) = 1$, and $m(z) - m(z-1) = 0$ otherwise. So by Theorem 6.2, if $\mathcal{K}(z)$ holds and $2s_2(z) \leqslant (1-\varepsilon_1)m(z)$, then

$$\mathbb{P}\big(\mathcal{D}(z)^c \mid \mathcal{F}_z\big) \leqslant \mathbb{P}\big(\mathcal{D}(z)^c \mid \mathcal{F}_z,\, d(z) = 1\big) \leqslant z_0^{-20}. \tag{112}$$

Let

$$s_k^*(z) := \begin{cases} \dfrac{s_k(z)}{\sigma_k m(z)} & \text{if } \mathcal{D}^*(z) \text{ holds,} \\ s_k^*(z+1) & \text{otherwise} \end{cases}$$

for each $k \geqslant 2$ and $z \in [z_+, \pi(x)]$, and define

$$\mathcal{L}^*(z) := \mathcal{D}^*(z) \cap \bigcap_{w \in [z, \pi(x)]} \Big(\mathcal{Q}(w) \cap \bigcap_{k \geqslant 2} \big\{s_k^*(w) \leqslant 1\big\}\Big) \tag{113}$$

for each $z \geqslant z_+$. We shall in fact show that the event $\mathcal{L}^*(z_+)$ holds with high probability, which will be sufficient to prove Proposition 7.1. Let us quickly note, for future reference, that the event $\mathcal{L}^*(z)$ implies that the conditions of Theorem 6.2 are satisfied.

Lemma 7.4. *Let $z \in [z_+, \pi(x)]$. If $\mathcal{D}^*(z)$ holds and $s_k^*(z) \leqslant 1$ for every $k \geqslant 2$, then*

$$\sum_{k \geqslant 2} 2^k s_k(z) \leqslant \varepsilon_1 m(z). \tag{114}$$

In particular, if $\mathcal{L}^*(z)$ holds, then $2s_2(z) \leqslant \varepsilon_1 m(z)$ and $\mathcal{K}(z)$ holds.

Proof. Note first that if $\mathcal{D}^*(z)$ holds and $s_k^*(z) \leqslant 1$ for every $k \geqslant 2$, then

$$\sum_{k \geqslant 2} 2^k s_k(z) \leqslant \sum_{k \geqslant 2} 2^k \sigma_k m(z) = \sum_{k \geqslant 2} \frac{(2\varepsilon_1)^k \tilde{\delta}^{k-1}}{2k\Lambda(z_+)} m(z) \leqslant \varepsilon_1 m(z),$$

since $\tilde{\delta} \leqslant \delta \leqslant \Lambda(z_+) \leqslant 1$ and $\varepsilon_1 \leqslant 1/16$, which proves (114). It follows that $\mathcal{L}^*(z)$ also implies (114), which in turn implies that $2s_2(z) \leqslant \varepsilon_1 m(z)$ and that (41) holds. Since $\mathcal{L}^*(z)$ also implies $\mathcal{Q}(z)$, this is sufficient to show that $\mathcal{K}(z)$ holds, as claimed. □

In order to apply the method of self-correcting martingales, we shall need (for each $k \geqslant 2$) an estimate of the expected change in $s_k^*(z)$ as well as a bound on the largest jump in $s_k^*(z)$. To be precise, we shall prove the following two lemmas.

Lemma 7.5. *Let $z \in [z_+, z_0^3]$. If $\mathcal{L}^*(z)$ holds, then*

$$\mathbb{E}\big[s_k^*(z-1) - s_k^*(z) \mid \mathcal{F}_z\big] \leqslant (k-1)p_z(x)\big(-s_k^*(z) + 16\varepsilon_1\big)$$

for every $k \geqslant 2$.

Lemma 7.6. *Let $z = z_0^\beta \in [z_+, z_0^3]$. If $\mathcal{L}^*(z)$ holds, then*

$$|s_k^*(z-1) - s_k^*(z)| \leqslant z_0^{-2+1/\beta+\varepsilon_1}$$

for every $k \geqslant 2$.

We shall first deduce Lemma 7.6, which is relatively straightforward, and then Lemma 7.5, the proof of which will require a little more work. The first step is to recall the following simple facts about $m(z)$.

Observation 7.7. *Let $z \in [z_+, \pi(x)]$. If $\mathcal{M}(z)$ holds, then $z_0^{1+o(1)} \leqslant m(z) \leqslant z_0^{2+o(1)}$. More precisely, if $z = z_0^\beta$, then $m(z) = z_0^{2-1/\beta+o(1)}$.*

Proof. Recall first that if $\mathcal{M}(z)$ holds, then $m(z) = \Theta(\Lambda(z)z)$, by Observation 4.7 and (44). Moreover, $\Lambda(z) = z_0^{2-\beta-1/\beta+o(1)}$ for $z = z_0^\beta$, by Corollary 3.2, so $m(z) = z_0^{2-1/\beta+o(1)}$, as claimed. Since $\beta \geqslant 1$ for every $z_+ \geqslant z_0$, the bounds $z_0^{1+o(1)} \leqslant m(z) \leqslant z_0^{2+o(1)}$ follow. □

Proof of Lemma 7.6. Recall first that $\mathcal{L}^*(z)$ implies $\mathcal{Q}(z)$, which implies that $s_k(z) = 0$ if $k > 4u_0$ (see Definition 4.2), so we may assume that $k \leqslant 4u_0$. Also, by definition of $s_k^*(z)$, we may assume $\mathcal{D}(z)$ holds. Recall that $u_0 = o(\log z_0)$, by (23), and note that therefore

$$\sigma_k = \frac{\varepsilon_1^k}{\Lambda(z_+)} \cdot \frac{\tilde{\delta}^{k-1}}{2k} = e^{O(k)} = z_0^{o(1)} = m(z)^{o(1)}. \tag{115}$$

By Observation 3.12, each row of A contains at most $2u_0$ non-zero entries to the right of z, and hence $|s_k(z-1) - s_k(z)| \leqslant 2u_0|m(z-1) - m(z)|$ if $d(z) = 1$. On the other hand, $|s_k(z-1) - s_k(z)| \leqslant 1$ and $m(z-1) = m(z)$ if $d(z) \neq 1$. Thus $\mathcal{D}(z)$ implies

$$|s_k(z-1) - s_k(z)| \leqslant 2u_0^3. \tag{116}$$

Finally, note that, by the definition of $s_k^*(z)$, we have

$$\begin{aligned}\Big|\sigma_k\Big(s_k^*(z-1)-s_k^*(z)\Big)\Big| &= \left|\frac{s_k(z-1)}{m(z-1)}-\frac{s_k(z)}{m(z)}\right| \\ &\leqslant \frac{|s_k(z-1)-s_k(z)|}{m(z-1)}+\frac{s_k(z)|m(z)-m(z-1)|}{m(z)m(z-1)} \\ &\leqslant \frac{2u_0^3}{m(z-1)}+\frac{s_k(z)u_0^2}{m(z)m(z-1)} \\ &\leqslant \frac{3u_0^3}{m(z)-u_0^2} = m(z)^{-1+o(1)},\end{aligned}$$

since $s_k(z) \leqslant \varepsilon_1 m(z)$, by $\mathcal{L}^*(z)$ and Lemma 7.4, and since $u_0 \leqslant \log z_0$ and $m(z) \geqslant z_0^{1+o(1)}$. The lemma now follows from (115) and Observation 7.7 as $\mathcal{L}^*(z)$ implies $\mathcal{M}(z)$. □

We shall now prove Lemma 7.5. The first step is to note the following immediate consequence of Theorem 6.2 and Lemma 7.4. The key observation is that since $s_2(z)$ is small, if $d(z) = 1$, then we are likely to only remove a single row.

Lemma 7.8. *Let $z \in [z_+, \pi(x)]$. If $\mathcal{L}^*(z)$ holds, then*

$$\mathbb{E}\Big[\Big(m(z)-m(z-1)\Big)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z,\, d(z)=1\Big] \in 1 \pm 2\varepsilon_1.$$

Proof. By Lemma 7.4, $2s_2(z) \leqslant \varepsilon_1 m(z)$. Thus by Theorem 6.2, and recalling that $D(z) = m(z) - m(z-1)$, it follows that

$$\begin{aligned}\mathbb{E}\Big[(m(z)-m(z-1))\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z,\, d(z)=1\Big] &= \left(1-\frac{2s_2(z)}{m(z)}+o(1)\right)^{-1} \\ &\quad + O(m(z)z_0^{-20}) \in 1 \pm 2\varepsilon_1\end{aligned}$$

by (112), as claimed. □

Next, we shall calculate the expected change in $s_k(z)$.

Lemma 7.9. *Let $z \in [z_+, \pi(x)]$. If $\mathcal{L}^*(z)$ holds, then*

(117)
$$\mathbb{E}\Big[\Big(s_k(z-1)-s_k(z)\Big)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z\Big] \leqslant k\sigma_k\Big(-s_k^*(z)+5\varepsilon_1\Big)\mathbb{P}\Big(d(z)=1 \mid \mathcal{F}_z\Big)$$

for every $k \geqslant 2$.

Recall from Definition 6.1 that $\Delta_k(z) = |S_k(z) \setminus S_k(z-1)|$ denotes the number of edges of size k that contain a vertex removed in step z, and $\Delta'(z)$ denotes the number of edges of size at least 3 that have at least two vertices removed in step z. We shall use the following simple observation to prove Lemma 7.9, and again in Section 8.

Observation 7.10. *For every $z \in [z_-, \pi(x)]$ and $k \geqslant 2$, we have*

$$s_k(z-1) - s_k(z) \in \mathbb{1}_{\{d(z)=k\}} + \big(-\Delta_k(z) + \Delta_{k+1}(z) \pm \Delta'(z)\big)\mathbb{1}_{\{d(z)=1\}}.$$

Proof. Recall from Algorithm 2.7 that if $d(z) \notin \{1,k\}$, then $S_k(z-1) = S_k(z)$, while if $d(z) = k$, then $S_k(z-1) = S_k(z) \cup \{z\}$. If $d(z) = 1$, then exactly $\Delta_k(z)$ edges are removed from $S_k(z)$, and $\Delta_{k+1}(z) \pm \Delta'(z)$ edges are added to $S_k(z)$, as required. Note that the $\Delta'(z)$ bounds both the number of $(k+1)$-edges that lose more than one vertex, as well as edges of size at least $k+2$ that lose enough vertices to become k-edges. □

Proof of Lemma 7.9. Note first that, since $\mathcal{L}^*(z)$ holds, we have $s_k^*(z) \leqslant 1$ for each $k \geqslant 2$, $2s_2(z) \leqslant \varepsilon_1 m(z)$, and $\mathcal{K}(z)$ holds (by Lemma 7.4), and $s_k(z) = \sigma_k m(z) s_k^*(z)$ (since $\mathcal{L}^*(z)$ implies $\mathcal{D}^*(z)$). Thus, by Theorem 6.2,

$$\begin{aligned}\mathbb{E}\big[\Delta_k(z) \mid \mathcal{F}_z,\, d(z) = 1\big] &= \left(1 - \frac{2s_2(z)}{m(z)} + o(1)\right)^{-1} \frac{ks_k(z)}{m(z)} + O(m(z)^{-1/3}) \\ &\in \big(1 \pm 2\varepsilon_1\big) k\sigma_k s_k^*(z) + O(m(z)^{-1/3})\end{aligned}$$

for each $k \geqslant 2$, and also

$$\mathbb{E}\big[\Delta'(z) \mid \mathcal{F}_z,\, d(z) = 1\big] = O(m(z)^{-1/3}).$$

Note also that (108) implies

$$(k+1)\sigma_{k+1} = \varepsilon_1 \tilde{\delta} \cdot k\sigma_k.$$

By Observation 7.10, and using (112) and (115), it follows that

$$\begin{aligned}&\mathbb{E}\big[\big(s_k(z-1) - s_k(z)\big)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z,\, d(z) = 1\big] \\ &\quad\leqslant -(1-2\varepsilon_1)k\sigma_k s_k^*(z) + (1+2\varepsilon_1)(k+1)\sigma_{k+1}s_{k+1}^*(z) + O\big(m(z)^{-1/3}\big) \\ &\quad\leqslant k\sigma_k\big(-s_k^*(z) + 2\varepsilon_1 + (1+2\varepsilon_1)\varepsilon_1\tilde{\delta} + o(1)\big) \\ &\quad\leqslant k\sigma_k\big(-s_k^*(z) + 3\varepsilon_1\big).\end{aligned}$$

Finally, by Lemma 4.9 and Observation 7.3, we have

$$\frac{\mathbb{P}\big(d(z) = k \mid \mathcal{F}_z\big)}{\mathbb{P}\big(d(z) = 1 \mid \mathcal{F}_z\big)} \leqslant \frac{(2\delta\eta)^{k-1}}{k!} \leqslant \frac{(3\tilde{\delta})^{k-1}}{k!} \leqslant \frac{\varepsilon_1^{k+1}\tilde{\delta}^{k-1}}{\Lambda(z_+)} = 2\varepsilon_1 k\sigma_k$$

for every $z \in [z_+, \pi(x)]$ and every $k \geqslant 2$. Hence

$$\mathbb{E}\big[(s_k(z-1) - s_k(z))\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z\big] \leqslant k\sigma_k\big(-s_k^*(z) + 5\varepsilon_1\big)\mathbb{P}\big(d(z) = 1 \mid \mathcal{F}_z\big),$$

as required. □

Lemma 7.5 will now follow as a straightforward consequence of Lemmas 7.8 and 7.9.

Proof of Lemma 7.5. As in the proof of Lemma 7.6, we may assume that $k \leqslant 4u_0$, and hence that $\sigma_k = z_0^{o(1)}$, by (115). Observe that

$$\begin{aligned}(118)\quad \sigma_k\big(s_k^*(z-1) - s_k^*(z)\big) &= \left(\frac{s_k(z-1)}{m(z-1)} - \frac{s_k(z)}{m(z)}\right)\mathbb{1}_{\mathcal{D}(z)} \\ &= \frac{m(z)}{m(z-1)}\left(\frac{s_k(z-1)-s_k(z)}{m(z)} + \frac{s_k(z)\big(m(z)-m(z-1)\big)}{m(z)^2}\right)\mathbb{1}_{\mathcal{D}(z)}.\end{aligned}$$

Since $\mathcal{L}^*(z)$ implies $\mathcal{K}(z)$, by Lemma 7.4, it follows from Lemmas 4.9 and 7.8 that

$$\mathbb{E}\big[\big(m(z)-m(z-1)\big)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z\big] \leqslant (1+3\varepsilon_1)m(z)p_z(x).$$

Since $\mathcal{L}^*(z)$ implies that $s_k(z) = \sigma_k m(z) s_k^*(z)$ and $s_k^*(z) \leqslant 1$, it follows that

$$(119)\quad \mathbb{E}\big[s_k(z)\big(m(z)-m(z-1)\big)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z\big] \leqslant \big(s_k^*(z)+3\varepsilon_1\big)\sigma_k m(z)^2 p_z(x).$$

Similarly, by Lemmas 4.9 and 7.9, we have

$$(120)\quad \mathbb{E}\big[\big(s_k(z-1)-s_k(z)\big)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z\big] \leqslant \big(-ks_k^*(z)+6\varepsilon_1 k\big)\sigma_k m(z)p_z(x).$$

It remains to bound

$$(121)\quad \left(\frac{m(z)-m(z-1)}{m(z-1)}\right)\left(\frac{s_k(z-1)-s_k(z)}{m(z)} + \frac{s_k(z)\big(m(z)-m(z-1)\big)}{m(z)^2}\right)$$

under the assumption that $\mathcal{D}(z)$ holds. To do so, recall that $s_k(z) \leqslant \varepsilon_1 m(z)$, by Lemma 7.4, that $m(z) = z_0^{2-1/\beta+o(1)}$, by Observation 7.7, where $z = z_0^\beta$, and that $1 \leqslant \beta \leqslant 3$, since $z \in [z_+, z_0^3]$. It follows from (116) that (121) is at most

$$\left(\frac{u_0^2}{m(z-1)}\right)\left(\frac{2u_0^3}{m(z)} + \frac{s_k(z)u_0^2}{m(z)^2}\right) \leqslant \frac{3u_0^5}{m(z)(m(z)-u_0^2)} \leqslant z_0^{-4+2/\beta+\varepsilon_1}.$$

Now $p_z(x) = z^{-1+o(1)} = z_0^{-\beta+o(1)}$, by Corollary 3.8, and $\sigma_k = z_0^{o(1)}$, by (115). Thus

$$z_0^{-4+2/\beta+\varepsilon_1} \leqslant z_0^{-\beta-\varepsilon_1} \leqslant \varepsilon_1\sigma_k p_z(x),$$

where we have used the fact that $\beta + 2/\beta \leqslant \frac{11}{3} < 4 - 2\varepsilon_1$ for $1 \leqslant \beta \leqslant 3$. Hence, using (118), (119) and (120), we have

$$\begin{aligned}\mathbb{E}\big[s_k^*(z-1) - s_k^*(z) \mid \mathcal{F}_z\big] &\leqslant \big(-(k-1)s_k^*(z) + 6\varepsilon_1 k + 4\varepsilon_1\big)p_z(x) \\ &\leqslant (k-1)p_z(x)\big(-s_k^*(z) + 16\varepsilon_1\big)\end{aligned}$$

for every $k \geqslant 2$, as required. □

We can now deduce the main result of the section.

Proof of Proposition 7.1. We shall show that $\mathcal{L}^*(z_+)$ holds with high probability, which implies both the events $\mathcal{M}^*(z_+)$ and $\mathcal{D}^*(z_+)$, and the inequality (110). The idea is (roughly speaking) to bound, for each $a \geqslant z_+$, the probability that a is maximal such that $\mathcal{L}^*(a)$ does not hold. The main step is the proof of the following claim.

Claim: For each $z_+ \leqslant a \leqslant z_0^3$ and $k \geqslant 2$, we have

$$\mathbb{P}\Big(\mathcal{L}^*(a+1) \cap \big\{s_k^*(a) > 1\big\} \cap \big\{s_k^*(z_0^3) \leqslant 3/4\big\}\Big) \leqslant z_0^{-20}.$$

Proof of claim. For each $z_+ \leqslant a < b \leqslant z_0^3$ and $k \geqslant 2$, let us define $\mathcal{U}_k(a,b)$ to be the event that the following all occur:

(a) $s_k^*(a) > 1$.
(b) $s_k^*(z) > 3/4$ for every $a < z < b$.
(c) $s_k^*(b) \leqslant 3/4$.
(d) $\mathcal{L}^*(a+1)$ holds.

Note that if $\mathcal{L}^*(a+1)$ holds, $s_k^*(a) > 1$ and $s_k^*(z_0^3) \leqslant 3/4$, then the event $\mathcal{U}_k(a,b)$ occurs for some (unique) b, $a < b \leqslant z_0^3$. By the union bound, it will therefore suffice to prove that

$$\mathbb{P}\big(\mathcal{U}_k(a,b)\big) \leqslant z_0^{-23}$$

for every $z_+ \leqslant a < b \leqslant z_0^3$.

By Lemma 7.6 we may assume $s_k^*(b) \geqslant 3/4 - \varepsilon_1$, as otherwise $s_k^*(b-1) \leqslant 3/4$ and so $\mathcal{U}_k(a,b)$ is impossible. For each $t \in \{0, \dots, b-a\}$, define

$$X_t := \begin{cases} s_k^*(b-t) - s_k^*(b) & \text{if } X_{t-1} \geqslant 0 \text{ or } t = 0, \\ X_{t-1} & \text{otherwise.} \end{cases}$$

We claim that X_t is a super-martingale with respect to the filtration $(\mathcal{F}_{b-t})_{t=0}^{b-a}$. Indeed, if $X_t < 0$, then $X_{t+1} = X_t$, and if $X_t \geqslant 0$, then

$$\mathbb{E}\big[X_{t+1} - X_t \mid \mathcal{F}_{b-t}\big] \leqslant (k-1)p_{b-t}(x)\big(-s_k^*(b-t) + 16\varepsilon_1\big) \leqslant 0$$

by Lemma 7.5, since $X_t \geqslant 0$ and (10) imply $s_k^*(b-t) \geqslant s_k^*(b) \geqslant 3/4 - \varepsilon_1 \geqslant 16\varepsilon_1$. Write $c_t = z_0^{-2+1/\beta+\varepsilon_1}$, where $b - t = z_0^\beta$. Then

$$|X_{t+1} - X_t| \leqslant c_t,$$

by Lemma 7.6. Also,

$$\sum_{t=0}^{b-a-1} c_t^2 \leqslant \sum_{z=z_+}^{z_0^3} z_0^{-4+2/\beta+2\varepsilon_1} \leqslant \sum_{z=z_+}^{z_0^3} z_0^{-4/3+2\varepsilon_1} z^{-2/3} \leqslant z_0^{-1/3+3\varepsilon_1},$$

where $\beta = \beta(z)$ is defined by $z = z_0^\beta$, and the second inequality holds as $2/\beta \leqslant (8-2\beta)/3$ for $1 \leqslant \beta \leqslant 3$. But $\mathcal{U}_k(a,b)$ implies that $X_{b-a} > 1/4$, so by the Azuma–Hoeffding inequality, we obtain

$$\mathbb{P}\big(\mathcal{U}_k(a,b)\big) \leqslant \mathbb{P}\big(X_{b-a} > 1/4\big) \leqslant \exp\big(-z_0^{1/3-4\varepsilon_1}\big) \leqslant z_0^{-23},$$

as claimed. □

To complete the proof of the proposition, we will show that with high probability there does not exist $z \in [z_+, \pi(x)]$ such that $\mathcal{L}^*(z+1)$ holds but $\mathcal{L}^*(z)$ does not hold. (Here $\mathcal{L}^*(\pi(x)+1)$ holds vacuously.) Note first that, by Lemma 7.4 and (112), with high probability there does not exist $z \in [z_+, \pi(x)]$ such that $\mathcal{L}^*(z+1)$ holds but $\mathcal{D}(z+1)$ does not; since $\mathcal{D}^*(z+1)$ and $\mathcal{D}(z+1)$ imply $\mathcal{D}^*(z)$, we may assume that $\mathcal{D}^*(z)$ holds.

Now consider the condition $s_k^*(z) \leqslant 1$. Lemma 4.12 implies that with high probability we have $s_2(z) + s_3(z) \leqslant z_0^{-1/2} m(z)$ for every $z \geqslant z_0^3$, and $s_k(z) = 0$ for every $k \geqslant 4$ and every $z \geqslant z_0^3$. Since $\sigma_k = \Theta(1)$ for $k \in \{2,3\}$, it follows that with high probability (110) holds and, moreover, $s_k^*(z) \leqslant 3/4$ for all $z \geqslant z_0^3$ and every $k \geqslant 2$. But if $s_k^*(z_0^3) \leqslant 3/4$, then the claim implies that with high probability there does not exist $z \in [z_+, z_0^3]$ and $k \geqslant 2$ such that $\mathcal{L}^*(z+1)$ and $s_k^*(z) > 1$.

We next consider the event $\mathcal{M}(z)$. Note first that, by Lemma 7.4, and using the assumptions that $\mathcal{D}^*(z)$ holds and that $s_k^*(z) \leqslant 1$ for every $k \geqslant 2$, we have

$$m_0(z) \leqslant m(z) \leqslant m_0(z) + \sum_{k \geqslant 2} k s_k(z) \leqslant m_0(z) + \varepsilon_1 m(z), \tag{122}$$

since the number of non-isolated vertices is at most the sum of the degrees. Recall that with high probability we have $m_0(z) \in (1 \pm \varepsilon_1)\eta\Lambda(z)z$ for every $z \in [z_-, \pi(x)]$, by Lemma 4.10. Assuming this holds, it follows from (122) that

$$(1-\varepsilon_1)\eta\Lambda(z) \leqslant \frac{m(z)}{z} \leqslant \frac{1+\varepsilon_1}{1-\varepsilon_1}\eta\Lambda(z).$$

Recalling from (12) and (39) that $\Lambda(z) \leqslant \delta + o(1)$ when $z \geqslant z_+$, and that

$$w(1-w) \leqslant we^{-w} \leqslant we^{-\mathrm{Ein}(w)} \leqslant w$$

for $0 \leqslant w \leqslant 1$, since $0 \leqslant \mathrm{Ein}(w) \leqslant w$, by (8), we obtain

$$(1-2\delta\eta)(1-\varepsilon_1)\eta\Lambda(z) \leqslant \frac{m(z)}{z} e^{-\mathrm{Ein}(m(z)/z)} \leqslant \frac{1+\varepsilon_1}{1-\varepsilon_1}\eta\Lambda(z).$$

This implies that

$$\frac{m(z)}{z} e^{-\mathrm{Ein}(m(z)/z)} \in \big(1 \pm 3\varepsilon_1\big)\eta\Lambda(z), \tag{123}$$

which implies that $\mathcal{M}(z)$ holds. As $\mathcal{L}^*(z+1)$ implies $\mathcal{M}^*(z+1)$, this implies $\mathcal{M}^*(z)$ holds. However, by Lemma 4.3, with high probability there does not exist $z \in [z_+, \pi(x)]$ such that $\mathcal{M}^*(z)$ holds but $\mathcal{Q}(z)$ does not. It follows that with high probability, there does not exist $z \geqslant z_+$ such that $\mathcal{L}^*(z+1)$ but $\mathcal{L}^*(z)$ fails to hold, and the proof is complete. $\square$

8. Tracking the process in the critical range

In the next two sections we shall track $s_k(z)$ and $m(z)$ in the "critical" range $[z_-, z_+]$. The main aim of this section is to prove two lemmas corresponding to Lemmas 7.5 and 7.6 from the previous section. We shall need these lemmas in Section 9 in order to track $s_k(z)$ using the method of self-correcting martingales.

The first step is to define the event we shall use to prove our key lemmas. Recall that $\mathcal{D}(z)$ denotes the event that $m(z) - m(z-1) \leqslant u_0^2$ and $\mathcal{D}^*(z)$ denotes the event that $\mathcal{D}(w)$ holds for all $w > z$. Now , for each $z \in [z_-, z_+]$ and each $k \geqslant 2$, define

$$s_k^*(z) := \begin{cases} \dfrac{s_k(z) - \hat{s}_k(z)}{\varepsilon(k,z)\hat{s}_k(z)} & \text{if } \mathcal{D}^*(z) \text{ holds,} \\ s_k^*(z+1) & \text{otherwise.} \end{cases} \tag{124}$$

Recall that $\mathcal{T}_k(z)$ denotes the event that $s_k(z) \in (1 \pm \varepsilon(k,z))\hat{s}_k(z)$, and define

$$\mathcal{T}^*(z) := \mathcal{D}^*(z) \cap \bigcap_{w=z}^{z_+} \left(\mathcal{Q}(w) \cap \bigcap_{k=2}^{4u_0} \mathcal{T}_k(w) \right)$$

for each $z \in [z_-, z_+]$. Note that $\mathcal{Q}(z)$ implies $\mathcal{T}_k(z)$ for every $k \geqslant 4u_0$ since for such k, we have $\varepsilon(k,z) > 1$ (since $u_0 = \omega(1)$), and $s_k(z) = 0$ (see Definition 4.2).

In this section we shall prove the following two lemmas.

Lemma 8.1. *Let $z \in [z_-, z_+]$. If $\mathcal{T}^*(z)$ holds, then*

$$\mathbb{E}\big[s_k^*(z-1) - s_k^*(z) \mid \mathcal{F}_z\big] \in -\frac{k-1}{z}\left(s_k^*(z) \pm \frac{1}{2}\right)$$

for every $2 \leqslant k \leqslant 4u_0$.

Lemma 8.2. *Let $z \in [z_-, z_+]$. If $\mathcal{T}^*(z)$ holds, then*

$$|s_k^*(z-1) - s_k^*(z)| \leqslant z_0^{-1+\varepsilon_1}$$

for every $2 \leqslant k \leqslant 4u_0$.

The proofs of these two lemmas are, in outline, similar to those of Lemmas 7.5 and 7.6, but the calculation is more delicate in the critical range, and as a consequence the details are somewhat more complicated. We begin by noting that $\mathcal{T}^*(z)$ implies the conditions of Theorem 6.2 are satisfied and so, in particular, that (112) holds.

Lemma 8.3. *Let $z \in [z_-, z_+]$. If $\mathcal{T}^*(z)$ holds, then $2s_2(z) \leqslant (1-\varepsilon_1)m(z)$ and $\mathcal{K}(z)$ holds.*

Proof. Since $\mathcal{T}^*(z)$ implies that $\mathcal{Q}(z)$ holds and that $\mathcal{T}_k(z)$ holds for every $k \geqslant 2$, it follows from Lemma 4.4 that $\mathcal{K}(z)$ holds. To prove the bound on $s_2(z)$, note that $\mathcal{T}_2(z)$ implies that

$$2s_2(z) \leqslant 2\big(1+\varepsilon(2,z)\big)\hat{s}_2(z) = \left(1+\frac{2\varepsilon_1^2}{\Lambda(z)}\right)m(z)\big(1-e^{-m(z)/z}\big), \tag{125}$$

by (16) and Definition 2.5. Since

$$1-e^{-m(z)/z} \leqslant \frac{m(z)}{z} \leqslant C_0\Lambda(z),$$

by Observation 4.7, it follows that for $\Lambda(z) \leqslant 2\varepsilon_1$,

$$2s_2(z) \leqslant \big(C_0\Lambda(z)+2\varepsilon_1^2C_0\big)m(z) \leqslant 2\varepsilon_1C_0(1+\varepsilon_1)m(z) \leqslant \frac{m(z)}{2}$$

since $\varepsilon_1C_0 \leqslant \varepsilon_1e^{C_0} < 1/16$ by (10). On the other hand, if $2\varepsilon_1 \leqslant \Lambda(z) \leqslant 1$, then

$$2s_2(z) \leqslant (1+\varepsilon_1)\big(1-e^{-C_0}\big)m(z) \leqslant (1+\varepsilon_1)(1-2\varepsilon_1)m(z) \leqslant (1-\varepsilon_1)m(z),$$

as required, as $e^{-C_0} \geqslant 2\varepsilon_1$. □

8.1. *The expected change in $m(z)$.* The next step is to use Theorem 6.2 to bound the expected number of vertices removed in each step. Recall that if $\mathcal{T}^*(z)$ holds, then the average degree in the graph $S_2(z)$ is close to $2\hat{s}_2(z)/m(z) = 1-e^{-m(z)/z}$, and so the expected size of $D(z) = m(z)-m(z-1)$ should be about $e^{m(z)/z}\cdot\mathbb{P}(d(z)=1) \approx m(z)/z$, by Lemma 4.5. The following lemma makes this precise.

Lemma 8.4. *Let $z \in [z_-, z_+]$. If $\mathcal{T}^*(z)$ holds, then*

$$\mathbb{E}\big[m(z)-m(z-1) \mid \mathcal{F}_z\big] = \big(1+\gamma(z)+o(1)\big)\frac{m(z)}{z}$$

and

$$\mathbb{E}\big[\big(m(z)-m(z-1)\big)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z\big] = \big(1+\gamma(z)+o(1)\big)\frac{m(z)}{z},$$

where $\gamma(z)$ is defined by

$$\gamma(z) := \frac{\varepsilon(2,z)s_2^*(z)\big(e^{m(z)/z}-1\big)}{1-\varepsilon(2,z)s_2^*(z)\big(e^{m(z)/z}-1\big)}. \tag{126}$$

Observation 8.5. *Let $z \in [z_-, z_+]$. If $\mathcal{T}^*(z)$ holds, then $|\gamma(z)| \leqslant \varepsilon_1$.*

Proof. Since $m(z)/z \leqslant C_0\Lambda(z)$, by Observation 4.7, it follows that

$$\varepsilon(2,z)\big(e^{m(z)/z}-1\big) \leqslant \frac{2\varepsilon_1^2}{\Lambda(z)}\big(e^{C_0\Lambda(z)}-1\big).$$

Now $(e^x-1)/x$ is an increasing function of x and $\Lambda(z) \leqslant 1$, so

$$\varepsilon(2,z)\big(e^{m(z)/z}-1\big) \leqslant 2\varepsilon_1^2(e^{C_0}-1) \leqslant \varepsilon_1/8$$

by (10). The result follows from the definition (126) of $\gamma(z)$ and the fact that $\mathcal{T}^*(z)$ implies $\mathcal{T}_2(z)$, so $|s_2^*(z)| \leqslant 1$. □

In the proof below, and also several times later in the section, we will need the fact that $m(z)=\Theta(z)$ uniformly in $z \in [z_-, z_+]$, which follows from $\mathcal{M}(z)$ (and hence from $\mathcal{T}^*(z)$) by Observation 4.7, and by (38) and (44).

Proof of Lemma 8.4. Note first that $\mathcal{T}^*(z)$ implies that $\mathcal{K}(z)$ and $\mathcal{M}(z)$ hold, by Lemma 8.3 and Definition 4.2. It follows that $\mathbb{P}(d(z)=1 \mid \mathcal{F}_z) = (1+o(1))\frac{m(z)}{z}e^{-m(z)/z} = \Theta(1)$, by Lemma 4.5 and since $m(z)=\Theta(z)$. Recall also that $D(z)=m(z)-m(z-1)=0$ if $d(z)\neq 1$. By Theorem 6.2, Lemma 8.3, and (112), it follows that

$$\begin{aligned}
&\mathbb{E}\big[D(z)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z\big] \\
&\qquad = \Big(\mathbb{E}\big[D(z) \mid \mathcal{F}_z,\, d(z)=1\big] - O\big(m(z)z_0^{-20}\big)\Big)\mathbb{P}\big(d(z)=1 \mid \mathcal{F}_z\big) \\
&\qquad = \left(1-\frac{2s_2(z)}{m(z)}+o(1)\right)^{-1}\frac{m(z)}{z}e^{-m(z)/z}.
\end{aligned}$$

Now, $\mathcal{T}^*(z)$ implies $\mathcal{D}^*(z)$, so

$$\frac{2s_2(z)}{m(z)} = \big(1+\varepsilon(2,z)s_2^*(z)\big)\frac{2\hat{s}_2(z)}{m(z)} = \big(1+\varepsilon(2,z)s_2^*(z)\big)\big(1-e^{-m(z)/z}\big),$$

by (16) and (124). Thus we obtain

$$\begin{aligned}
&\mathbb{E}\big[D(z)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z\big] \\
&\qquad = \Big(e^{m(z)/z} - \big(1+\varepsilon(2,z)s_2^*(z)\big)\big(e^{m(z)/z}-1\big)+o(1)\Big)^{-1}\frac{m(z)}{z} \\
&\qquad = \Big(1-\varepsilon(2,z)s_2^*(z)\big(e^{m(z)/z}-1\big)+o(1)\Big)^{-1}\frac{m(z)}{z} \\
&\qquad = \big(1+\gamma(z)+o(1)\big)\frac{m(z)}{z},
\end{aligned}$$

and similarly for $\mathbb{E}\big[D(z) \mid \mathcal{F}_z\big]$, as required. □

8.2. *The expected change in* $\hat{s}_k(z)$. The next step is to bound the expected change of $\hat{s}_k(z)$. We shall use Lemma 8.4 to bound the first moment of $D(z)$, and Theorem 6.2 to bound its second moment. To simplify the statement, let us define

$$g_k(z) := \frac{\hat{s}_k(z)}{z} + \frac{e^{-m(z)/z}}{k!}\left(\frac{m(z)}{z}\right)^k \tag{127}$$

for each $k \geqslant 2$. Note that $g_k(z) = O(1)$ if the event $\mathcal{T}^*(z)$ holds, since $\hat{s}_k(z) \leqslant m(z)$ and $m(z) = O(z)$, by Observation 4.7. We shall prove the following lemma.

LEMMA 8.6. *Let $z \in [z_-, z_+]$. If $\mathcal{T}^*(z)$ holds, then*

$$\mathbb{E}\Big[\big(\hat{s}_k(z-1) - \hat{s}_k(z)\big)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z\Big] = -\frac{\hat{s}_k(z)}{z} - \big(\gamma(z) + o(1)\big)g_k(z) \tag{128}$$

for every $2 \leqslant k \leqslant 4u_0$.

Note that, since $g_k(z) = O(1)$, the error term is $o(1)$. However, it will be important in the proof of Lemma 8.1 that $g_k(z)$ is significantly smaller than this when $k \to \infty$. The first step in the proof of Lemma 8.6 is to obtain deterministic bounds on $\hat{s}_k(z-1) - \hat{s}_k(z)$, which follow via some easy algebra. We give the details for completeness.

LEMMA 8.7. *Let $z \in [z_-, z_+]$. If $\mathcal{T}^*(z)$ and $\mathcal{D}(z)$ hold, then*

$$\hat{s}_k(z) - \hat{s}_k(z-1) = \frac{\hat{s}_k(z)}{z} + \frac{zg_k(z)}{m(z)}\left(D(z) - \frac{m(z)}{z} + \frac{O\big(k(D(z)^2+1)\big)}{z}\right) \tag{129}$$

for $2 \leqslant k \leqslant 4u_0$.

Proof. For each $k \geqslant 2$ and $w \geqslant 0$, set

$$f_k(w) := \frac{we^{-w}}{k(k-1)} \sum_{\ell=k-1}^{\infty} \frac{w^\ell}{\ell!}$$

so that $\hat{s}_k(z)/z = f_k(m(z)/z)$. Observe that

$$f_k'(w) = \frac{(1-w)e^{-w}}{k(k-1)} \sum_{\ell=k-1}^{\infty} \frac{w^\ell}{\ell!} + \frac{we^{-w}}{k(k-1)} \sum_{\ell=k-2}^{\infty} \frac{w^\ell}{\ell!} = \frac{1}{w}\left(f_k(w) + \frac{w^k e^{-w}}{k!}\right)$$

so that $f_k'(m(z)/z) = zg_k(z)/m(z)$. Observe also that

$$f_k''(w) = \frac{1}{w}\frac{d}{dw}\big(wf_k'(w) - f_k(w)\big) = \frac{1}{w}\frac{d}{dw}\frac{w^k e^{-w}}{k!} = \frac{w^{k-2}e^{-w}}{(k-1)!} - \frac{w^{k-1}e^{-w}}{k!}. \tag{130}$$

Thus, if w is bounded away from 0, $f_k''(w) = O(kf_k'(w))$. Moreover, if $|w' - w| = O(u_0^2/z) = o(1/k)$, then we still have $f_k''(w'') = O(kf_k'(w))$ for all $w'' \in [w, w']$ as neither term in (130) changes by more than a constant factor. Thus, by Taylor's Theorem,

$$f_k(w) - f_k(w') = f_k'(w)\Big(w - w' + O\big(k(w-w')^2\big)\Big). \tag{131}$$

Writing $w = m(z)/z$ and $w' = m(z-1)/(z-1)$ we have

$$w - w' = \frac{m(z)}{z} - \frac{m(z) - D(z)}{z-1} = \frac{1}{z-1}\left(D(z) - \frac{m(z)}{z}\right).$$

If $\mathcal{T}^*(z)$ holds, then w is indeed bounded away from zero as $\mathcal{T}^*(z)$ implies $\mathcal{Q}(z)$, which implies $\mathcal{M}(z)$. Also $\mathcal{D}(z)$ implies $D(z) \leqslant u_0^2$, so $|w' - w| = O(u_0^2/z)$. Thus substituting these values of w and w' into (131) gives

$$\frac{\hat{s}_k(z)}{z} - \frac{\hat{s}_k(z-1)}{z-1} = \frac{z g_k(z)}{(z-1)m(z)}\left(D(z) - \frac{m(z)}{z} + \frac{O(k(D(z)^2+1))}{z}\right).$$

Multiplying by $z-1$ then gives (129). □

We note here, for future reference, the following identity, which follows immediately from the definition (15) of $\hat{s}_k(z)$.

Observation 8.8. *For each $k \geqslant 2$ and every $z \in [\pi(x)]$,*

$$(k-1)\hat{s}_k(z) - (k+1)\hat{s}_{k+1}(z) = \frac{e^{-m(z)/z}}{k!}\frac{m(z)^k}{z^{k-1}}.$$

Note that as an immediate corollary of this observation, it follows that

$$g_k(z)z = k\hat{s}_k(z) - (k+1)\hat{s}_{k+1}(z) \leqslant k\hat{s}_k(z). \tag{132}$$

Moreover, from Lemma 8.7 and (132) it follows that if the events $\mathcal{T}^*(z)$ and $\mathcal{D}(z)$ hold for some $z \in [z_-, z_+]$, then

$$\frac{\hat{s}_k(z-1)}{\hat{s}_k(z)} = 1 + O\big(z^{-1+o(1)}\big), \tag{133}$$

since $\mathcal{D}(z)$ implies $D(z) \leqslant u_0^2$ and $\mathcal{M}(z)$ implies $m(z) = \Theta(z)$, as noted above.

To deduce Lemma 8.6 from Lemma 8.7, we just need to take expectations of both sides and apply Theorem 6.2 and Lemma 8.4.

Proof of Lemma 8.6. Recall that

$$\mathbb{E}\big[D(z)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z\big] = \big(1 + \gamma(z) + o(1)\big)\frac{m(z)}{z},$$

by Lemma 8.4, and that $\mathbb{E}[D(z)^2 \mid \mathcal{F}_z] = O(1)$ by Theorem 6.2 and Lemma 8.3. It therefore follows from Lemma 8.7 and (112) that

$$\begin{aligned}\mathbb{E}\big[\big(\hat{s}_k(z) - \hat{s}_k(z-1)\big)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z\big] &= \frac{\hat{s}_k(z)}{z} \\ &\quad + \left(\gamma(z) + o(1) + \frac{O(k)}{m(z)}\right)g_k(z) + O\big(z_0^{-20}\big),\end{aligned}$$

since $\hat{s}_k(z)/z \leqslant g_k(z) = O(1)$. The result follows since $k \leqslant 4u_0 = o(m(z))$. □

8.3. *The expected change in $s_k(z)$.* We now arrive at the main calculation: that of the expected change in the number of edges of size k.

LEMMA 8.9. *Let $z \in [z_-, z_+]$. If $\mathcal{T}^*(z)$ holds, then*

$$\mathbb{E}\Big[\big(s_k(z-1) - s_k(z)\big)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z\Big] \in -\frac{\hat{s}_k(z)}{z} - \gamma(z)g_k(z) - \frac{k \cdot \varepsilon(k,z)\hat{s}_k(z)}{z}\bigg(s_k^*(z) \pm \frac{1}{6}\bigg)$$

for every $2 \leqslant k \leqslant 4u_0$.

We shall first prove the following deterministic lemma.

LEMMA 8.10. *Let $z \in [z_-, z_+]$. If $\mathcal{T}^*(z)$ holds, then*

$$ks_k(z) - (k+1)s_{k+1}(z) \in g_k(z)z + k \cdot \varepsilon(k,z)\hat{s}_k(z)\bigg(s_k^*(z) \pm \frac{1}{8}\bigg)$$

for every $k \geqslant 2$.

Proof. Observe first that since the event $\mathcal{T}^*(z)$ holds, we have

$$\begin{aligned} ks_k(z) &- (k+1)s_{k+1}(z) \\ &\in k\big(1 + \varepsilon(k,z)s_k^*(z)\big)\hat{s}_k(z) - (k+1)\big(1 \pm \varepsilon(k+1,z)\big)\hat{s}_{k+1}(z) \\ &= g_k(z)z + k \cdot \varepsilon(k,z)s_k^*(z)\hat{s}_k(z) \pm (k+1)\varepsilon(k+1,z)\hat{s}_{k+1}(z), \end{aligned}$$

by (132). But by Definition 2.5 and Observations 4.7 and 4.14, we have

(134)
$$(k+1)\varepsilon(k+1,z)\hat{s}_{k+1}(z) \leqslant \frac{m(z)}{z}\varepsilon(k+1,z)\hat{s}_k(z) \leqslant \varepsilon_1 C_0(k+1) \cdot \varepsilon(k,z)\hat{s}_k(z),$$

so since $\varepsilon_1 C_0(k+1) \leqslant 2k\varepsilon_1 e^{C_0} < k/8$, by (10), the claimed bounds follow. □

We can now use Theorem 6.2 to bound the expected change in $s_k(z)$.

Proof of Lemma 8.9. Let $z \in [z_-, z_+]$, and suppose that $\mathcal{T}^*(z)$ holds. Recall that

$$s_k(z-1) - s_k(z) \in \mathbb{1}_{\{d(z)=k\}} + \big(-\Delta_k(z) + \Delta_{k+1}(z) \pm \Delta'(z)\big)\mathbb{1}_{\{d(z)=1\}} \tag{135}$$

for each $k \geqslant 2$, by Observation 7.10. Also,

$$\begin{aligned} \mathbb{E}\big[\Delta_k(z)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z,\, d(z) = 1\big] &= \frac{ks_k(z)}{m(z)}\Big(\mathbb{E}\big[D(z) \mid \mathcal{F}_z,\, d(z) = 1\big] + o(1)\Big) + O\big(m(z)^{-1/3}\big) \end{aligned} \tag{136}$$

for each $k \geqslant 2$, and

$$\mathbb{E}\big[\Delta'(z) \mid \mathcal{F}_z,\, d(z) = 1\big] = O\big(m(z)^{-1/3}\big), \tag{137}$$

by Theorem 6.2 and Lemma 8.3, where we have used the fact that $\Delta_k(z) \leqslant s_k(z)$ deterministically and (112) to bound the error when $\mathcal{D}(z)$ fails. Moreover,

$$\mathbb{E}\big[D(z) \mid \mathcal{F}_z,\, d(z) = 1\big]\,\mathbb{P}\big(d(z) = 1 \mid \mathcal{F}_z\big) = \big(1 + \gamma(z) + o(1)\big)\frac{m(z)}{z}, \tag{138}$$

by Lemma 8.4. We claim that

$$\mathbb{E}\Big[\big(s_k(z-1)-s_k(z)\big)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z,\, d(z)=1\Big]\,\mathbb{P}\big(d(z)=1 \mid \mathcal{F}_z\big) = \frac{1+\gamma(z)}{z}\big((k+1)s_{k+1}(z)-ks_k(z)\big)+\frac{o\big(k\varepsilon(k,z)\hat{s}_k(z)\big)}{z}. \tag{139}$$

Indeed, in order to deduce this from (135)–(138), we just need to show that the various error terms are all $o\big(k\varepsilon(k,z)\hat{s}_k(z)/z\big)$. To see this, note first

$$\varepsilon(k,z) \geqslant \varepsilon_1^k k! \geqslant \varepsilon_1^{1/\varepsilon_1}(1/\varepsilon_1)! \tag{140}$$

is bounded below by a positive constant, since $\Lambda(z)\leqslant 1$ for every $z\in[z_-,z_+]$, and recall that $\varepsilon(k,z)\hat{s}_k(z)=z_0^{1+o(1)}$, by Observation 4.16 and our assumption that $k\leqslant 4u_0$, that $\varepsilon(k+1,z)\hat{s}_{k+1}(z)=O(\varepsilon(k,z)\hat{s}_k(z))$ by (134), that $|s_k^*(z)|+|s_{k+1}^*(z)|=O(1)$, by $\mathcal{T}^*(z)$, and that $m(z)=\Theta(z)$. The error terms are therefore at most

$$\frac{o\big(ks_k(z)+(k+1)s_{k+1}(z)\big)}{z}+O\big(m(z)^{-1/3}\big)=\frac{o\big(k\varepsilon(k,z)\hat{s}_k(z)\big)}{z},$$

as claimed, so we have proved (139).

To deal with the case when $d(z)>1$, observe also that, by Lemmas 4.5 and 8.3, we have

$$\mathbb{P}\big(d(z)=k \mid \mathcal{F}_z\big)=\big(1+o(1)\big)^k\frac{e^{-m(z)/z}}{k!}\left(\frac{m(z)}{z}\right)^k+\frac{O(1)}{z}.$$

We claim that in fact

$$\mathbb{P}\big(d(z)=k \mid \mathcal{F}_z\big)=\frac{e^{-m(z)/z}}{k!}\left(\frac{m(z)}{z}\right)^k+\frac{o\big(\varepsilon(k,z)\hat{s}_k(z)\big)}{z}. \tag{141}$$

To see this, note that if $k=O(1)$, then $(1+o(1))^k=1+o(1)$ and $\frac{e^{-m(z)/z}}{k!}\big(\frac{m(z)}{z}\big)^k = O(\hat{s}_k(z)/z)$, by (15). On the other hand, if $k=\omega(1)$ then $\mathbb{P}(d(z)=k \mid \mathcal{F}_z)\leqslant (2C_0)^k/k!+O(1/z)$ decreases super-exponentially with k, while $\varepsilon(k,z)\hat{s}_k(z)=e^{O(k)}z$, by (49) and our assumption that $k\leqslant 4u_0$.

Now, noting that $d(z)=k>1$ implies $\mathcal{D}(z)$, it follows from (135), (139) and (141) that

$$\mathbb{E}\Big[\big(s_k(z-1)-s_k(z)\big)\mathbb{1}_{\mathcal{D}(z)} \mid \mathcal{F}_z\Big]=\frac{1+\gamma(z)}{z}\big((k+1)s_{k+1}(z)-ks_k(z)\big) + \frac{e^{-m(z)/z}}{k!}\frac{m(z)^k}{z^k}+\frac{o\big(k\varepsilon(k,z)\hat{s}_k(z)\big)}{z}.$$

By Lemma 8.10, this is contained in

$$\frac{e^{-m(z)/z}}{k!}\frac{m(z)^k}{z^k}-\frac{1+\gamma(z)}{z}\left(g_k(z)z+k\cdot\varepsilon(k,z)\hat{s}_k(z)\left(s_k^*(z)\pm\frac{1}{7}\right)\right),$$

which is equal to

$$-\frac{\hat{s}_k(z)}{z} - \gamma(z)g_k(z) - \frac{1+\gamma(z)}{z}\bigg(k \cdot \varepsilon(k,z)\hat{s}_k(z)\bigg(s_k^*(z) \pm \frac{1}{7}\bigg)\bigg).$$

Since $|\gamma(z)| \leqslant \varepsilon_1$ and $|s_k^*(z)| \leqslant 1$, by Observation 8.5 and $\mathcal{T}^*(z)$, the lemma follows. □

8.4. *The proof of Lemmas* 8.1 *and* 8.2. We are finally ready to prove the two main lemmas of the section. We will prove Lemma 8.2 first, since we shall need (a weak form of) it in the proof of Lemma 8.1.

Proof of Lemma 8.2. If $\mathcal{D}(z)$ fails to hold, then $s_k^*(z-1) = s_k^*(z)$ by definition, so we may assume that $|m(z) - m(z-1)| \leqslant u_0^2$, and hence $|s_k(z-1) - s_k(z)| \leqslant 2u_0^3$, by (116). We claim first that

$$\begin{aligned} s_k^*(z-1) - s_k^*(z) &= \frac{s_k(z-1) - \hat{s}_k(z-1)}{\varepsilon(k,z-1)\hat{s}_k(z-1)} - \frac{s_k(z) - \hat{s}_k(z)}{\varepsilon(k,z)\hat{s}_k(z)} \\ &= \frac{s_k(z-1) - \hat{s}_k(z-1) - s_k(z) + \hat{s}_k(z)}{\varepsilon(k,z)\hat{s}_k(z)} + O\bigg(\frac{s_k^*(z-1)}{z^{1+o(1)}}\bigg). \end{aligned}$$

Indeed, to see this simply recall that, by Definition 2.5, Lemma 3.10 and (133), the events $\mathcal{T}^*(z)$ and $\mathcal{D}(z)$ imply that

$$\frac{\varepsilon(k,z-1)\hat{s}_k(z-1)}{\varepsilon(k,z)\hat{s}_k(z)} = \frac{\Lambda(z)}{\Lambda(z-1)} \cdot \frac{\hat{s}_k(z-1)}{\hat{s}_k(z)} = 1 + O\big(z^{-1+o(1)}\big).$$

Now, note that $\hat{s}_k(z-1) - \hat{s}_k(z) = O(u_0^2)$, by Lemma 8.7, since $m(z) = \Theta(z)$, $\hat{s}_k(z)/z \leqslant g_k(z) = O(1)$ and $u_0 = z_0^{o(1)}$. Thus

$$|s_k^*(z-1) - s_k^*(z)| = \frac{O(u_0^3)}{\varepsilon(k,z)\hat{s}_k(z)} + O\big(z^{-1+o(1)}\big)$$

as $\mathcal{T}^*(z)$ implies $|s_k^*(z)| \leqslant 1$. Now $z = z_0^{1+o(1)}$ for every $z \in [z_-, z_+]$, by Lemma 3.11, and $\varepsilon(k,z)\hat{s}_k(z) = z_0^{1+o(1)}$ for all $k \leqslant 4u_0$, by Observation 4.16. Hence

$$|s_k^*(z-1) - s_k^*(z)| \leqslant z_0^{-1+o(1)},$$

as required. □

To finish the section, we shall deduce Lemma 8.1 from Lemmas 8.2, 8.6 and 8.9.

Proof of Lemma 8.1. Let $z \in [z_-, z_+]$, and suppose that $\mathcal{T}^*(z)$ holds. We shall break

$$s_k^*(z-1) - s_k^*(z) = \bigg(\frac{s_k(z-1) - \hat{s}_k(z-1)}{\varepsilon(k,z-1)\hat{s}_k(z-1)} - \frac{s_k(z) - \hat{s}_k(z)}{\varepsilon(k,z)\hat{s}_k(z)}\bigg)\mathbb{1}_{\mathcal{D}(z)}$$

into two pieces (see (142) and (143), below) and bound the expected size of each of them in turn. We note that, by Lemmas 8.6 and 8.9, we have

$$\mathbb{E}\Big[\big(s_k(z-1)-\hat{s}_k(z-1)-s_k(z)+\hat{s}_k(z)\big)\mathbb{1}_{\mathcal{D}(z)}\,\Big|\,\mathcal{F}_z\Big] \in -\frac{k\varepsilon(k,z)\hat{s}_k(z)}{z}\bigg(s_k^*(z)\pm\frac{1}{6}\bigg)+o(g_k(z)).$$

But $g_k(z)\leqslant k\hat{s}_k(z)/z=O(k\varepsilon(k,z)\hat{s}_k(z)/z)$ by (132) and the fact that $\varepsilon(k,z)$ is bounded away from 0. Thus

$$\mathbb{E}\bigg[\frac{s_k(z-1)-\hat{s}_k(z-1)-s_k(z)+\hat{s}_k(z)}{\varepsilon(k,z)\hat{s}_k(z)}\mathbb{1}_{\mathcal{D}(z)}\,\Big|\,\mathcal{F}_z\bigg] \in -\frac{k}{z}\bigg(s_k^*(z)\pm\frac{1}{5}\bigg). \tag{142}$$

We claim that

$$\mathbb{E}\bigg[\bigg(\frac{\varepsilon(k,z)\hat{s}_k(z)-\varepsilon(k,z-1)\hat{s}_k(z-1)}{\varepsilon(k,z)\hat{s}_k(z)}\bigg)s_k^*(z-1)\mathbb{1}_{\mathcal{D}(z)}\,\Big|\,\mathcal{F}_z\bigg] \in \frac{s_k^*(z)\pm 3\varepsilon_1 k}{z}. \tag{143}$$

To prove (143), observe first that if $\mathcal{D}(z)$ holds, then

$$\varepsilon(k,z)\hat{s}_k(z)-\varepsilon(k,z-1)\hat{s}_k(z-1)=\varepsilon(k,z)\big(\hat{s}_k(z)-\hat{s}_k(z-1)\big)+o\bigg(\frac{\varepsilon(k,z)\hat{s}_k(z)}{z}\bigg),$$

since $\varepsilon(k,z-1)/\varepsilon(k,z)=1+o(1/z)$, by Definition 2.5 and Lemma 3.10, and $\hat{s}_k(z-1)=\Theta\big(\hat{s}_k(z)\big)$ if $\mathcal{T}^*(z)$ and $\mathcal{D}(z)$ hold, by (133). Therefore, by Lemma 8.6,

$$\begin{aligned}\mathbb{E}\Big[\big(\varepsilon(k,z)\hat{s}_k(z)-\varepsilon(k,z-1)\hat{s}_k(z-1)\big)\mathbb{1}_{\mathcal{D}(z)}\mid\mathcal{F}_z\Big]&=\frac{\varepsilon(k,z)\hat{s}_k(z)}{z}\bigg(1+\frac{(\gamma(z)+o(1))g_k(z)z}{\hat{s}_k(z)}+o(1)\bigg)\\&\in\big(1\pm 2\varepsilon_1 k\big)\frac{\varepsilon(k,z)\hat{s}_k(z)}{z},\end{aligned}$$

since $g_k(z)\leqslant k\hat{s}_k(z)/z$, by (132), and $|\gamma(z)|\leqslant\varepsilon_1$, by Observation 8.5. Since $|s_k^*(z-1)-s_k^*(z)|\leqslant\varepsilon_1$ by Lemma 8.2, it follows that

$$\mathbb{E}\bigg[\bigg(\frac{\varepsilon(k,z)\hat{s}_k(z)-\varepsilon(k,z-1)\hat{s}_k(z-1)}{\varepsilon(k,z)\hat{s}_k(z)}\bigg)s_k^*(z-1)\mathbb{1}_{\mathcal{D}(z)}\,\Big|\,\mathcal{F}_z\bigg] \in \frac{(1\pm 2\varepsilon_1 k)(s_k^*(z)\pm\varepsilon_1)}{z}\in\frac{s_k^*(z)\pm 3\varepsilon_1 k}{z},$$

as claimed. As noted above, adding (142) and (143) we obtain

$$\begin{aligned}\mathbb{E}\big[s_k^*(z-1)-s_k^*(z)\mid\mathcal{F}_z\big]&\in-\frac{k}{z}\bigg(s_k^*(z)\pm\frac{1}{5}\bigg)+\frac{s_k^*(z)\pm 3\varepsilon_1 k}{z}\\&\subseteq-\frac{k-1}{z}\bigg(s_k^*(z)\pm\frac{1}{2}\bigg)\end{aligned}$$

for every $k\geqslant 2$, which completes the proof of the lemma. □

9. The proof of Theorems 2.2 and 2.6

In this section we shall use the method of self-correcting martingales to deduce Theorems 2.2 and 2.6 from Lemmas 8.1 and 8.2. Recall that Theorem 2.2 states that the event $\mathcal{M}^*(z_-)$ holds with high probability, and Theorem 2.6 states that with high probability the event $\mathcal{T}_k(z)$ holds for every $k \geqslant 2$ and every $z \in [z_-, z_+]$. Since we already know that $\mathcal{M}^*(z_+)$ holds with high probability, by Proposition 7.1, it will suffice to prove that the event $\mathcal{T}^*(z_-)$ holds with high probability, since $\mathcal{T}^*(z_-)$ implies that $\mathcal{M}(z)$ holds for every $z \in [z_-, z_+]$, and that $\mathcal{T}_k(z)$ holds for every $k \geqslant 2$ and every $z \in [z_-, z_+]$.

As in Section 7, the rough idea is to bound the probability, for each $a \in [z_-, z_+]$, that a is maximal such that $\mathcal{T}^*(a)$ fails to hold. Let us begin by proving the base case.

Lemma 9.1. *$\mathcal{T}^*(z_+)$ holds with high probability.*

Proof. Recall that by Lemma 4.3, $\mathcal{Q}(z_+) \cup \mathcal{M}^*(z_+)^c$ holds with high probability and that by Proposition 7.1 and Corollary 7.2, $\mathcal{D}^*(z_+) \cap \mathcal{M}^*(z_+) \cap \bigcap_{k=2}^{4u_0} \mathcal{T}_k(z_+)$ holds with high probability. Thus

$$\mathcal{T}^*(z_+) = \mathcal{D}^*(z_+) \cap \mathcal{Q}(z_+) \cap \bigcap_{k=2}^{4u_0} \mathcal{T}_k(z_+)$$

holds with high probability, as claimed. □

By Lemma 9.1, we can assume that $z_- \leqslant a < z_+$ and that $\mathcal{T}^*(a+1)$ holds. For each pair $z_- \leqslant a < b \leqslant z_+$ and each $k \geqslant 2$, let $\mathcal{W}_k(a,b)$ denote the event that the following all occur:

(a) $s_k^*(a) > 1$.
(b) $s_k^*(z) > 3/4$ for every $a < z < b$.
(c) $s_k^*(b) \leqslant 3/4$.
(d) $\mathcal{T}^*(a+1)$ holds.

Note that if $\mathcal{T}^*(a+1) \cap \left\{s_k^*(z_+) \leqslant 3/4\right\}$ holds and $s_k^*(a) > 1$, then the event $\mathcal{W}_k(a,b)$ holds for some $a < b \leqslant z_+$. We shall prove the following lemma.

Lemma 9.2. *If $z_- \leqslant a < b \leqslant z_+$, then*

$$\mathbb{P}\big(\mathcal{W}_k(a,b)\big) \leqslant z_0^{-20}$$

for every $2 \leqslant k \leqslant 4u_0$.

Proof. We may assume $s_k^*(b) \geqslant 3/4 - \varepsilon_1$, as otherwise $s_k^*(b-1) < 3/4$ by Lemma 8.2, and so $\mathcal{W}_k(a,b)$ fails to hold. Set $\ell = b - a$, and for each $0 \leqslant t \leqslant \ell$, define

$$X_t := \begin{cases} s_k^*(b-t) - s_k^*(b) & \text{if } X_{t-1} \geqslant 0 \text{ or } t = 0, \\ X_{t-1} & \text{otherwise.} \end{cases}$$

We claim that X_t is a super-martingale with respect to the filtration $(\mathcal{F}_{b-t})_{t=0}^{\ell}$. Indeed, since $\mathcal{T}^*(a+1)$ holds, we have either $X_t < 0$, in which case $X_{t+1} = X_t$, or $X_t \geqslant 0$, in which case

$$\mathbb{E}\big[X_{t+1} - X_t \mid \mathcal{F}_{b-t}\big] \leqslant \frac{k-1}{b-t}\bigg(- s_k^*(b-t) + \frac{1}{2}\bigg)$$

by Lemma 8.1. But if $X_t \geqslant 0$, then $s_k^*(b-t) \geqslant s_k^*(b) \geqslant 3/4 - \varepsilon_1 > 1/2$, so in all cases $\mathbb{E}[X_{t+1} - X_t \mid \mathcal{F}_{b-t}] \leqslant 0$. Recalling that $\ell \leqslant z_+ = z_0^{1+o(1)}$, it follows by the Azuma–Hoeffding inequality that

$$\mathbb{P}\big(\mathcal{W}_k(a,b)\big) \leqslant \exp\big(- z_0^{1-3\varepsilon_1}\big) \leqslant z_0^{-20},$$

as claimed. □

It is easy to see that one can deal with the case $s_k^*(a) < -1$ in exactly the same way, so we leave the details to the reader. Using the union bound over all choices of b, it follows that

$$\mathbb{P}\Big(\big\{|s_k^*(a)| > 1\big\} \cap \mathcal{T}^*(a+1) \cap \big\{s_k^*(z_+) \leqslant 3/4\big\}\Big) \leqslant z_0^{-18}$$

for every $a \in [z_-, z_+]$ and $2 \leqslant k \leqslant 4u_0$. Moreover, it follows from Corollary 7.2 that, with high probability, $s_k^*(z_+) \leqslant 1/2$ for every $k \geqslant 2$.

It therefore only remains to bound the probability that $\mathcal{D}(z+1)^c \cup \mathcal{Q}(z)^c \cup \mathcal{T}^*(z+1)$ holds for some $z \in [z_-, z_+]$. By (112) and Lemmas 4.3 and 8.3, it will therefore suffice to show that $m(z)$ is unlikely to be the first variable to go off track. Unfortunately, unlike with $s_k(z)$, $m(z)$ is not self-correcting, and so a simple martingale approach will not work. Instead we shall prove this using a two stage approach: we shall show, using super-martingales, that $m(z)$ can only drift off track slowly, but every so often (and well before it drifts so far as to cause $\mathcal{M}(z)$ to fail) we shall use the following lemma to put it firmly back on track. The following lemma, which follows from Lemma 4.10 and Theorem 5.1, ensures that $m(z)$ is far closer to its target value than required by $\mathcal{M}(z)$, but unfortunately it has a relatively large failure probability. Thus we cannot use it for very many values of z.

Lemma 9.3. *For every* $z \in [z_-, z_+]$,

$$\mathbb{P}\bigg(\mathcal{K}(z) \cap \bigg\{m(z)\exp\bigg(- \sum_{k\geqslant 2} \frac{ks_k(z)}{m(z)}\bigg) \notin (1 \pm 3\varepsilon_1)\eta\Lambda(z)z\bigg\}\bigg) = \frac{O(1)}{u_0}.$$

Proof. Recall that, by Lemma 4.10, the number of isolated vertices $m_0(z)$ in the hypergraph $\mathcal{S}_A(z)$ satisfies

$$m_0(z) \in \big(1 \pm \varepsilon_1\big)\eta\Lambda(z)z \tag{144}$$

with probability at least $1 - 1/x^2$. We shall use Theorem 5.1 to give another way of approximating $m_0(z)$, and together these estimate will imply the lemma.

Recall that a vertex i in $\mathcal{S}_A(z)$ is isolated if the submatrix $A\big[\{i\} \times [z+1, \pi(x)]\big]$ is the all zero vector. We bound the number $m_0(z)$ of isolated vertices using the second moment method, using Theorem 5.1 to estimate both the mean and variance of $m_0(z)$. First, let $\mathcal{E} \in \mathcal{F}_z^+$ be an event of the form (50) for which $\mathcal{K}(z)$ holds and $d(z) \leqslant 4u_0$, and recall from Definition 6.3 that the random matrix $\tilde{A}_{\mathcal{E}}$ is obtained by choosing each column uniformly at random from all $\binom{m(z)}{d_j}$ possible choices, independently in each column. Thus for $j > z$,

$$\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[\{i\} \times \{j\}] = 0\big) = 1 - \frac{d_j}{m(z)} = \exp\bigg(-\frac{d_j}{m(z)} + \frac{O(d_j^2)}{m(z)^2}\bigg).$$

As the columns of $\tilde{A}_{\mathcal{E}}$ are independent,

$$\mathbb{P}\big(\tilde{A}_{\mathcal{E}}\big[\{i\} \times [z+1, \pi(x)]\big] = 0\big) = \exp\bigg(-\frac{1}{m(z)}\sum_{j>z} d_j + \frac{O(1)}{m(z)^2}\sum_{j>z} d_j^2\bigg).$$

Now $\sum_{j>z} d_j = \sum_{k\geqslant 2} k s_k(z)$ and $\sum_{j>z} d_j^2 = \sum_{k\geqslant 2} k^2 s_k(z) = O(m(z))$ by condition $\mathcal{K}(z)$. Thus

$$\mathbb{P}\big(\tilde{A}_{\mathcal{E}}\big[\{i\} \times [z+1, \pi(x)]\big] = 0\big) = \exp\bigg(-\sum_{k\geqslant 2}\frac{k s_k(z)}{m(z)} + \frac{O(1)}{m(z)}\bigg).$$

Applying Theorem 5.1 with $I = \{i\}$, $C = [z+1, \pi(x)]$, and $R = 0$, gives

$$\begin{aligned}\mathbb{P}\big[i \text{ is isolated in } \mathcal{S}_A(z) \mid \mathcal{E}\big] &= \mathbb{P}\big(A\big[\{i\} \times [z+1, \pi(x)]\big] = 0 \mid \mathcal{E}\big)\\ &= \exp\bigg(-\sum_{k\geqslant 2}\frac{k s_k(z)}{m(z)} + \frac{O(1)}{u_0}\bigg).\end{aligned}$$

Thus, summing over $i \in M(z)$,

$$\mu := \mathbb{E}\big[m_0(z) \mid \mathcal{E}\big] = m(z)\exp\bigg(-\sum_{k\geqslant 2}\frac{k s_k(z)}{m(z)} + \frac{O(1)}{u_0}\bigg). \tag{145}$$

For the second moment of $m_0(z)$, we consider the probability that two distinct vertices i_1, i_2 are both isolated. In the independent model, we have

$$\begin{aligned}\mathbb{P}\big(\tilde{A}_{\mathcal{E}}[\{i_1, i_2\} \times \{j\}] = 0\big) &= \bigg(1 - \frac{d_j}{m(z)}\bigg)\bigg(1 - \frac{d_j}{m(z)-1}\bigg)\\ &= \exp\bigg(-\frac{2d_j}{m(z)} + \frac{O(d_j^2)}{m(z)^2}\bigg).\end{aligned}$$

As the columns of $\tilde{A}_{\mathcal{E}}$ are independent, a similar argument to the above yields

$$\mathbb{P}\big(\tilde{A}_{\mathcal{E}}\big[\{i_1, i_2\} \times [z+1, \pi(x)]\big] = 0\big) = \exp\bigg(-\sum_{k\geqslant 2}\frac{2k s_k(z)}{m(z)} + \frac{O(1)}{m(z)}\bigg).$$

Applying Theorem 5.1 with $I = \{i_1, i_2\}$, $C = [z+1, \pi(x)]$, and $R = 0$, gives

$$\mathbb{P}\Big(A\big[\{i_1, i_2\} \times [z+1, \pi(x)]\big] = 0 \mid \mathcal{E}\Big) = \exp\bigg(-\sum_{k \geqslant 2} \frac{2ks_k(z)}{m(z)} + \frac{O(1)}{u_0}\bigg).$$

Summing over all ordered pairs (i_1, i_2), we obtain

(146)

$$\begin{aligned}\mathbb{E}\big[m_0(z)(m_0(z)-1) \mid \mathcal{E}\big] &= m(z)(m(z)-1)\exp\bigg(-\sum_{k \geqslant 2} \frac{2ks_k(z)}{m(z)} + \frac{O(1)}{u_0}\bigg) \\ &= \bigg(1 + \frac{O(1)}{u_0}\bigg)\mu^2.\end{aligned}$$

Combining (145) and (146), we have

$$\mathrm{Var}\Big(m_0(z) \mid \mathcal{E}\Big) = \mathbb{E}\big[m_0(z)(m_0(z)-1) \mid \mathcal{E}\big] + \mu - \mu^2 = \mu + O(1/u_0)\mu^2.$$

Now $\sum ks_k(z) = O(m(z))$ by $\mathcal{K}(z)$ and so $\mu = \Theta(m(z))$. Thus $\mathrm{Var}\Big(m_0(z) \mid \mathcal{E}\Big) = O(\mu^2/u_0)$ and hence, by Chebychev's inequality,

$$\mathbb{P}\Big(m_0(z) \notin (1 \pm \varepsilon_1)\mu \mid \mathcal{E}\Big) \leqslant \frac{\mathrm{Var}\Big(m_0(z) \mid \mathcal{E}\Big)}{\varepsilon_1^2 \mu^2} = \frac{O(1)}{u_0}.$$

Since the event $\mathcal{E}$ was chosen arbitrarily amongst those consistent with $\mathcal{K}(z)$ and satisfying $d(z) \leqslant 4u_0$, and using (46) to bound the probability that $d(z) > 4u_0$, it follows that

$$\mathbb{P}\Big(m_0(z) \notin (1 \pm \varepsilon_1)\mu \mid \mathcal{K}(z)\Big) = \frac{O(1)}{u_0}.$$

Combining this with (144) and (145) completes the proof of the lemma. □

We shall also need the following simple identity.

Lemma 9.4. *For every* $z \in [\pi(x)]$,

$$\sum_{k \geqslant 2} k\hat{s}_k(z) = m(z)\mathrm{Ein}\bigg(\frac{m(z)}{z}\bigg).$$

Proof. Define

$$f(w) := \sum_{k \geqslant 2} \frac{1}{k-1} \sum_{\ell=k-1}^{\infty} \frac{e^{-w} w^\ell}{\ell!},$$

and note that

$$\sum_{k \geqslant 2} k\hat{s}_k(z) = \sum_{k \geqslant 2} \frac{m(z)}{k-1} e^{-m(z)/z} \sum_{\ell=k-1}^{\infty} \frac{1}{\ell!}\bigg(\frac{m(z)}{z}\bigg)^\ell = m(z) f\big(m(z)/z\big).$$

Now, we have $f(0) = 0$ and

$$\begin{aligned} f'(w) &= \sum_{k\geqslant 2} \frac{e^{-w}}{k-1}\bigg(\sum_{\ell=k-1}^{\infty} \frac{w^{\ell-1}}{(\ell-1)!} - \sum_{\ell=k-1}^{\infty} \frac{w^{\ell}}{\ell!}\bigg) \\ &= e^{-w}\sum_{k\geqslant 2} \frac{w^{k-2}}{(k-1)!} = e^{-w}\bigg(\frac{e^{w}-1}{w}\bigg) = \frac{1-e^{-w}}{w}. \end{aligned}$$

It follows that $f(w) = \displaystyle\int_0^w \frac{1-e^{-t}}{t}\,dt = \mathrm{Ein}(w)$, as required. □

We can now complete the proof of our main auxiliary results.

Proof of Theorems 2.2 *and* 2.6. As noted above, it will suffice to prove that the event $\mathcal{T}^*(z_-)$ holds with high probability. Recall that $\mathcal{T}^*(z_+)$ holds with high probability, by Lemma 9.1, and let $a \in [z_-, z_+]$ be maximal such that $\mathcal{T}^*(a)$ fails to hold. As $\mathcal{T}^*(a+1)$ holds, one of the events $\mathcal{D}^*(a)$, $\mathcal{Q}(a)$, or $\mathcal{T}_k(a)$ for some $2 \leqslant k \leqslant 4u_0$, must fail. By Lemma 8.3 and (112), with high probability there is no a such that $\mathcal{T}^*(a+1)$ holds but $\mathcal{D}(a+1)$ fails, so we may assume that $\mathcal{D}(a+1)$, and hence $\mathcal{D}^*(a)$, holds. By Lemma 9.2, with high probability there is no a such that $\mathcal{T}^*(a+1)$ and $\mathcal{D}^*(a)$ hold, but $\mathcal{T}_k(a)$ fails for some $k \geqslant 2$. By Lemma 4.3, with high probability there is no a such that $\mathcal{M}^*(a)$ holds but $\mathcal{Q}(a)$ fails. As $\mathcal{T}^*(a+1)$ implies $\mathcal{M}^*(a+1)$, we deduce that with high probability there is no a such that $\mathcal{T}^*(a+1)$ and $\mathcal{M}(a)$ hold but $\mathcal{T}^*(a)$ fails. It will therefore suffice to bound the probability that $\mathcal{M}(a)$ fails to hold, assuming that $\mathcal{D}^*(a)$ holds, and that $\mathcal{T}_k(a)$ holds for every $2 \leqslant k \leqslant 4u_0$.

To do so, let us choose a set $W = \{w_0, w_1, \dots, w_\ell\}$, where $z_- = w_0 < w_1 < \cdots < w_\ell = z_+$, such that $\ell = O(\log(z_+/z_-))$ and $w_i \leqslant 2w_{i-1}$ for each $i \in [\ell]$. Since $\log(z_+/z_-) = \Theta(\sqrt{\log z_0}) = o(u_0)$, by Lemma 3.11 and (23), it follows from Lemma 9.3 that with high probability either

$$m(w)\exp\bigg(-\sum_{k\geqslant 2}\frac{ks_k(w)}{m(w)}\bigg) \in (1 \pm 3\varepsilon_1)\eta\Lambda(w)w, \tag{147}$$

or $\mathcal{K}(w)$ fails to hold for every $w \in W$. Since by Lemma 4.4, $\mathcal{T}^*(z)$ implies $\mathcal{K}(z)$, it will suffice to bound the probability that (147) holds for $w = w_i$, say, and

$$\mathcal{M}(a)^c \cap \mathcal{D}^*(a) \cap \mathcal{T}^*(a+1) \cap \bigcap_{k=2}^{4u_0} \mathcal{T}_k(a) \tag{148}$$

holds for some $w_{i-1} \leqslant a < w_i$. We shall show that this has probability at most z_0^{-15}.

To bound the probability of the event (148), we shall use a martingale approach to control $m(z)$ in the interval $a \leqslant z \leqslant w$. Define $X_0 := 0$, and for

$0 \leqslant t := w - z < w - a$,

$$X_{t+1} := X_t + \left(\frac{m(w-t-1)}{w-t-1} - \frac{m(w-t)}{w-t}\right)\mathbb{1}_{\mathcal{D}(w-t)\cap\mathcal{T}^*(w-t)} - 5\varepsilon_1\frac{m(w)}{w^2}.$$

We claim that X_t is a super-martingale with respect to the filtration $(\mathcal{F}_{w-t})_{t=0}^{w-a}$. Indeed, $\mathcal{T}^*(w-t)$ is $\mathcal{F}_{w-t}$-measurable and clearly $X_{t+1} \leqslant X_t$ when $\mathcal{T}^*(w-t)$ fails. Assuming $\mathcal{T}^*(w-t) = \mathcal{T}^*(z)$ holds, we have

$$\mathbb{E}\big[m(z) - m(z-1) \mid \mathcal{F}_z\big] = \big(1 + \gamma(z) + o(1)\big)\frac{m(z)}{z}$$

by Lemma 8.4. Recalling that $\mathcal{T}^*(z)$ implies that $|\gamma(z)| \leqslant \varepsilon_1$, by Observation 8.5, it follows (using (112)) that

$$\begin{aligned}\mathbb{E}\big[X_{t+1} - X_t \mid \mathcal{F}_z\big] &= \mathbb{E}\left[\frac{m(z-1)-m(z)}{z-1} + \frac{m(z)}{z(z-1)} - \frac{5\varepsilon_1 m(w)}{w^2} \;\middle|\; \mathcal{F}_z\right] \\ &\quad + O\big(z_0^{-20}\big) \\ &= -\big(\gamma(z) + o(1)\big)\frac{m(z)}{z(z-1)} - \frac{5\varepsilon_1 m(w)}{w^2} \leqslant 0\end{aligned}$$

for every $a < z \leqslant w$, as claimed, since $w \leqslant 2a$ and $m(z) \leqslant m(w)$.

Now, $|X_{t+1} - X_t| \leqslant z_0^{-1+\varepsilon_1}$ for every $a < z \leqslant w$, since $m(z) = O(z)$ (by Observation 4.7) and $z = z_0^{1+o(1)}$ (by (22)), and since $\mathcal{D}(z)$ implies that $|m(z) - m(z-1)| \leqslant u_0^2 = z_0^{o(1)}$. Thus, noting that $w - a \leqslant z_+ = z_0^{1+o(1)}$, by the Azuma–Hoeffding inequality we obtain

$$\mathbb{P}\Big(\big\{X_{w-a} > z_0^{-\varepsilon_1}\big\} \cap \mathcal{T}^*(a+1)\Big) \leqslant \exp\big(-z_0^{1-5\varepsilon_1}\big) \leqslant z_0^{-20}.$$

Observe that if $\mathcal{D}^*(a) \cap \mathcal{T}^*(a+1)$ holds, then $X_{w-a} \leqslant z_0^{-\varepsilon_1}$ implies that

$$\frac{m(a)}{a} \leqslant \big(1 + 5\varepsilon_1\big)\frac{m(w)}{w} + z_0^{-\varepsilon_1}. \tag{149}$$

Now, to finish the proof, we shall show that if $\mathcal{T}^*(w)$ and (147) hold, then

$$\frac{m(w)}{w} \leqslant \alpha\big((1 + 5\varepsilon_1)\eta\Lambda(w)\big). \tag{150}$$

To prove this, observe first that if $\mathcal{T}^*(w)$ holds, then

$$\sum_{k\geqslant 2} k s_k(w) \leqslant \sum_{k\geqslant 2}\big(1 + \varepsilon(k,z)\big)k\hat{s}_k(w) \leqslant \left(\mathrm{Ein}\left(\frac{m(w)}{w}\right) + \varepsilon_1\right)m(w),$$

by Lemmas 4.15 and 9.4. Thus (147) implies that

$$\frac{m(w)}{w}\exp\left(-\mathrm{Ein}\left(\frac{m(w)}{w}\right)\right) \leqslant e^{\varepsilon_1}\big(1 + 3\varepsilon_1\big)\eta\Lambda(w),$$

which implies (150), since α is increasing and $e^{\varepsilon_1}(1+3\varepsilon_1) \leqslant 1+5\varepsilon_1$. Combining (149) and (150), and recalling that $\Lambda(a) = (1+o(1))\Lambda(w) \geqslant \delta + o(1)$ by Lemma 3.10 and (38), and that $\alpha(t)/t$ is increasing, by (9), it follows that,

$$\frac{m(a)}{a} \leqslant \alpha\big((1+11\varepsilon_1)\eta\Lambda(a)\big).$$

By the definition of α (and again using the fact that α is increasing), this implies that

$$\frac{m(a)}{a}\exp\bigg(-\operatorname{Ein}\bigg(\frac{m(a)}{a}\bigg)\bigg) \leqslant \big(1+11\varepsilon_1\big)\eta\Lambda(a).$$

A corresponding lower bound can be proved similarly, and thus

$$\mathbb{P}\bigg(\mathcal{M}(a)^c \cap \mathcal{D}^*(a) \cap \mathcal{T}^*(a+1) \cap \bigcap_{k=2}^{4u_0} \mathcal{T}_k(a)\bigg) \leqslant z_0^{-15},$$

as claimed. As explained above, this implies that $\mathcal{T}^*(z_-)$ holds with high probability, and hence this completes the proof of Theorems 2.2 and 2.6. □

10. **The proof of Theorem 1.1**

Once we have Theorem 2.2, it is straightforward to deduce Theorem 1.1. Indeed, the deduction of the lower bound follows from the results of [13], the extra ingredient provided by Theorem 2.2 being that any linear relation can only involve at most $m(z_-) \approx \eta\Lambda(z_-)z_-$ rows. We shall use the following result, which was proved in [13].

Proposition 10.1 (Croot, Granville, Pemantle and Tetali). *There exists $c > 0$ such that if $N \leqslant e^{-\gamma}J(x)$, then with high probability there does not exist a set $I \subseteq [N]$ with*

$$0 < |I| \leqslant z_0 \exp\Big(-c\sqrt{\log z_0}\Big)$$

such that $\prod_{i\in I} a_i$ is a square.

We remark that Proposition 10.1 follows from the proof of [13, Th. 1.3]; see [13, §3.5], for any constant $c > \sqrt{2-\log 2}$. However, we shall only use the fact that $c = O(1)$.

As noted in the introduction, we shall also give a new proof of the upper bound, which was originally proved in [13]. For the upper bound, we shall need the following simple identity.

Lemma 10.2. *For every $z \in [\pi(x)]$,*

$$\sum_{k\geqslant 2} \hat{s}_k(z) = m(z) - z\big(1 - e^{-m(z)/z}\big).$$

Proof. Using the fact that $\frac{1}{k(k-1)} = \frac{1}{k-1} - \frac{1}{k}$, we have

$$\begin{aligned}
\sum_{k\geqslant 2} \hat{s}_k(z) &= m(z)e^{-m(z)/z} \sum_{k\geqslant 2} \frac{1}{k(k-1)} \sum_{\ell=k-1}^{\infty} \frac{1}{\ell!}\Big(\frac{m(z)}{z}\Big)^{\ell} \\
&= m(z)e^{-m(z)/z} \sum_{\ell=1}^{\infty} \frac{1}{\ell!}\Big(\frac{m(z)}{z}\Big)^{\ell} \sum_{k=2}^{\ell+1} \Big(\frac{1}{k-1} - \frac{1}{k}\Big) \\
&= m(z)e^{-m(z)/z} \sum_{\ell=1}^{\infty} \frac{1}{\ell!}\Big(\frac{m(z)}{z}\Big)^{\ell} \Big(1 - \frac{1}{\ell+1}\Big) \\
&= m(z)e^{-m(z)/z} \sum_{\ell=1}^{\infty} \frac{1}{\ell!}\Big(\frac{m(z)}{z}\Big)^{\ell} - ze^{-m(z)/z} \sum_{\ell=1}^{\infty} \frac{1}{(\ell+1)!}\Big(\frac{m(z)}{z}\Big)^{\ell+1} \\
&= m(z)e^{-m(z)/z}\big(e^{m(z)/z} - 1\big) - ze^{-m(z)/z}\big(e^{m(z)/z} - 1 - m(z)/z\big) \\
&= m(z) - z\big(1 - e^{-m(z)/z}\big),
\end{aligned}$$

as claimed. □

Proof of Theorem 1.1. To prove the lower bound, it is enough to show that for $\eta < e^{-\gamma}$, there is with high probability no linear relation between the rows of A. Any such relation would correspond to an even sub-hypergraph of $\mathcal{H}_A(z_-)$ without isolated vertices and so must lie in the 2-core $\mathcal{C}_A(z_-)$. In particular, the number of rows involved is at most $m(z_-)$, which satisfies

$$m(z_-) \leqslant \alpha\big((1+\varepsilon_0)\Lambda(z_-)\eta\big)z_- \leqslant \alpha\big(2\delta e^{-\gamma}\big)z_- \leqslant z_- \tag{151}$$

with high probability, by Theorem 2.2. We may therefore assume that there is no linear relation involving more than z_- rows. However, by Lemma 3.11 we have

$$z_- = z_0 \exp\Big(-\big(1+o(1)\big)\sqrt{\log(1/\delta)\log z_0}\Big),$$

and so by Proposition 10.1 there is with high probability no linear relation involving at most z_- rows if δ is taken sufficiently small. Hence there is with high probability no linear relation between the rows of A, as required.

To prove the upper bound, assume that we have $\eta' J(x)$ numbers with $\eta' = e^{-\gamma} + \nu$, $\nu > 0$. Pick $\eta < e^{-\gamma}$, and construct the 2-core $\mathcal{C}_A(z_0)$ starting with just the first $N = \eta J(x)$ numbers. Observe that, by Theorems 2.2 and 2.6, and Lemmas 4.15 and 10.2, the number of columns of A that either have a non-zero entry in one of the rows of $M(z_0)$, or are to the left of z_0, is with high probability at most

$$\begin{aligned}
z_0 + \sum_{k\geqslant 2} s_k(z_0) &\leqslant z_0 + \sum_{k\geqslant 2} \big(1+\varepsilon(k,z_0)\big)\hat{s}_k(z_0) \\
&\leqslant (1+\varepsilon_1)m(z_0) + z_0 e^{-m(z_0)/z_0}.
\end{aligned} \tag{152}$$

On the other hand, we have at least $\nu J(x)$ remaining unused numbers, and among these there are, with high probability, at least

$$\frac{\nu J(x)}{2} \cdot \frac{\Psi(x, y_0)}{x} = \frac{\nu z_0}{2}$$

y_0-smooth numbers. Thus we have a total of at least $m(z_0) + \nu z_0/2$ rows of A, all of whose non-zero entries lie in a set of columns of size at most $(1+\varepsilon_1)m(z_0) + z_0 e^{-m(z_0)/z_0}$. Hence, if

$$\frac{\nu}{2} > \varepsilon_1 m(z_0)/z_0 + e^{-m(z_0)/z_0},$$

then we obtain a linear relation between the rows. Now, recall that $m(z_0)/z_0 \geqslant \alpha((1-\varepsilon_0)\eta)$ with high probability, by Theorem 2.2, and that $\alpha(w) \to \infty$ as $w \to e^{-\gamma}$. Hence, by choosing η sufficiently close to $e^{-\gamma}$, and ε_0 sufficiently small, we can make $m(z_0)/z_0$ arbitrarily large. In particular we can force $e^{-m(z_0)/z_0} < \nu/4$. Since $m(z_0)/z_0 \leqslant C_0$ with high probability, with $C_0 = C_0(\eta)$ fixed, the result follows by taking ε_1 sufficiently small. $\square$

The proof of the upper bound in Theorem 1.1 can be modified to show that the expected number of linear relations between the rows of A blows up at some $\eta_0 J(x)$, with $\eta_0 < e^{-\gamma}$, thus demonstrating that a straightforward application of the first moment method cannot give a sharp lower bound on $T(x)$. To see this, let $\eta < e^{-\gamma}$ and consider $N = \eta J(z)$ integers a_i. The number $m_0(z_0)$ of y_0-smooth numbers is binomially distributed with mean ηz_0, but can be much higher. Indeed,

$$\mathbb{P}\big(m_0(z_0) = 2\eta z_0\big) \approx \frac{(\eta z_0)^{2\eta z_0}}{(2\eta z_0)!} e^{-\eta z_0} \approx (e/4)^{(1+o(1))\eta z_0}.$$

However, the remaining numbers are still uniformly distributed over non-smooth numbers, and smooth numbers have no effect on the algorithm determining the 2-core. Thus if we remove about ηz_0 smooth numbers, the distribution of the 2-core $\mathcal{C}_A(z_0)$ of the remaining numbers has approximately the same distribution as if we had started with $N - \eta z_0 = (1+o(1))N$ numbers initially. Thus, with probability $(e/4)^{(1+o(1))\eta z_0}$, we have a submatrix of A with $m(z_0) + \eta z_0$ rows and $(1+\varepsilon_1)m(z_0) + z_0 e^{-m(z_0)/z_0}$ non-zero columns. Taking η sufficiently close to $e^{-\gamma}$ and $\varepsilon_0, \varepsilon_1$ sufficiently small, we obtain a submatrix of A with $(\eta - \varepsilon)z_0$ more rows than non-zero columns. This results in at least $2^{(\eta-\varepsilon)z_0} - 1$ non-trivial linear relations between the rows. As this occurs with probability $(e/4)^{(1+o(1))\eta z_0}$ and $e/4 > 1/2$, the expected number of linear relations grows exponentially with z_0, even though we are below the threshold.

Finally we give a proof of Corollary 1.2.

Proof of Corollary 1.2. Since finding a square product among $\{a_1, \dots, a_t\}$ is independent of finding one among $\{a_{t+1}, \dots, a_{2t}\}$ we have that $\mathbb{P}\big(T(x) \geqslant 2t\big) \leqslant \mathbb{P}\big(T(x) \geqslant t\big)^2$, and more generally $\mathbb{P}\big(T(x) \geqslant kt\big) \leqslant \mathbb{P}\big(T(x) \geqslant t\big)^k$ for every $k \in \mathbb{N}$.

Setting $t = \big(e^{-\gamma} + \varepsilon\big)J(x)$ and $\theta = \mathbb{P}\big(T(x) \geqslant t\big)$, we have

$$\begin{aligned}\mathbb{E}[T(x)] &\leqslant t\,\mathbb{P}\big(T(x) \in [0,t)\big) + 2t\,\mathbb{P}\big(T(x) \in [t,2t)\big) \\ &\quad + 3t\,\mathbb{P}\big(T(x) \in [2t,3t)\big) + \cdots \\ &= t + t\,\mathbb{P}\big(T(x) \geqslant t\big) + t\,\mathbb{P}\big(T(x) \geqslant 2t\big) + \cdots \\ &\leqslant \big(1 + \theta + \theta^2 + \cdots\big)t = \frac{e^{-\gamma} + \varepsilon}{1 - \theta} \cdot J(x).\end{aligned}$$

Now, observe that, by Theorem 1.1, $\theta \to 0$ as $x \to \infty$ for any $\varepsilon > 0$. Since $\varepsilon > 0$ is arbitrary, it follows that $\mathbb{E}[T(x)] \leqslant \big(e^{-\gamma} + o(1)\big)J(x)$. On the other hand, taking $t = \big(e^{-\gamma} - \varepsilon\big)J(x)$, we have

$$\mathbb{E}[T(x)] \geqslant t\,\mathbb{P}\big(T(x) \geqslant t\big) = \big(1 + o(1)\big)t = \big(e^{-\gamma} - \varepsilon - o(1)\big)J(x)$$

as $x \to \infty$, so, since $\varepsilon > 0$ was arbitrary, $\mathbb{E}[T(x)] \geqslant \big(e^{-\gamma} + o(1)\big)J(x)$, as required. □

Acknowledgements. This research was begun while the authors were visiting IMT, Lucca, and it was partly carried out while the first and third authors were Visiting Fellow Commoners of Trinity College, Cambridge. We would like to thank both institutions for providing a wonderful working environment.

References

[1] K. Azuma, Weighted sums of certain dependent random variables, *Tôhoku Math. J.* (2) **19** (1967), 357–367. MR 0221571. Zbl 0178.21103. https://doi.org/10.2748/tmj/1178243286.

[2] P. Balister, P. Bollobás, J. Lee, and R. Morris, On the 2-core of a random set of numbers, in preparation.

[3] J. Balogh, B. Bollobás, H. Duminil-Copin, and R. Morris, The sharp threshold for bootstrap percolation in all dimensions, *Trans. Amer. Math. Soc.* **364** no. 5 (2012), 2667–2701. MR 2888224. Zbl 1238.60108. https://doi.org/10.1090/S0002-9947-2011-05552-2.

[4] T. Bohman, A. Frieze, and E. Lubetzky, A note on the random greedy triangle-packing algorithm, *J. Comb.* **1** no. 3-4 (2010), 477–488. MR 2799220. Zbl 1244.05213. https://doi.org/10.4310/JOC.2010.v1.n4.a5.

[5] T. Bohman, A. Frieze, and E. Lubetzky, Random triangle removal, *Adv. Math.* **280** (2015), 379–438. MR 3350225. Zbl 1312.05123. https://doi.org/10.1016/j.aim.2015.04.015.

[6] T. Bohman and P. Keevash, Dynamic concentration of the triangle-free process, in *The Seventh European Conference on Combinatorics*, *Graph Theory and Applications*, *CRM Series* **16**, Ed. Norm., Pisa, 2013, pp. 489–495. MR 3185850. Zbl 1291.05183. https://doi.org/10.1007/978-88-7642-475-5_78.

[7] T. Bohman and M. Picollelli, SIR epidemics on random graphs with a fixed degree sequence, *Random Structures Algorithms* **41** no. 2 (2012), 179–214. MR 2956054. Zbl 06099390. https://doi.org/10.1002/rsa.20401.

[8] B. Bollobás, *Random Graphs*, second ed., *Cambridge Stud. Adv. Math.* **73**, Cambridge University Press, Cambridge, 2001. MR 1864966. Zbl 0979.05003. https://doi.org/10.1017/CBO9780511814068.

[9] N. G. de Bruijn, On the number of positive integers $\leq x$ and free of prime factors $> y$, *Nederl. Acad. Wetensch. Proc. Ser. A.* **54** (1951), 50–60. MR 0046375. Zbl 0042.04204. https://doi.org/10.1016/S1385-7258(51)50008-2.

[10] N. G. de Bruijn, On the number of positive integers $\leq x$ and free prime factors $> y$. II, *Nederl. Akad. Wetensch. Proc. Ser. A.* **69** (1966), 239–247. MR 0205945. Zbl 0139.27203. https://doi.org/10.1016/S1385-7258(66)50029-4.

[11] E. R. Canfield, P. Erdős, and C. Pomerance, On a problem of Oppenheim concerning "factorisatio numerorum", *J. Number Theory* **17** no. 1 (1983), 1–28. MR 0712964. Zbl 0513.10043. https://doi.org/10.1016/0022-314X(83)90002-1.

[12] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Statistics* **23** (1952), 493–507. MR 0057518. Zbl 0048.11804. https://doi.org/10.1214/aoms/1177729330.

[13] E. Croot, A. Granville, R. Pemantle, and P. Tetali, On sharp transitions in making squares, *Ann. of Math.* (2) **175** no. 3 (2012), 1507–1550. MR 2912710. Zbl 1321.11122. https://doi.org/10.4007/annals.2012.175.3.10.

[14] K. Dickman, On the frequency of numbers containing prime factors of a certain relative magnitude, *Ark. Mat. Astron. Fys.* **22** (1930), 1–14. Zbl 56.0178.04.

[15] J. D. Dixon, Asymptotically fast factorization of integers, *Math. Comp.* **36** no. 153 (1981), 255–260. MR 0595059. Zbl 0452.10010. https://doi.org/10.2307/2007743.

[16] G. Fiz Pontiveros, S. Griffiths, and R. Morris, The triangle-free process and the Ramsey numbers $R(3,k)$, *Mem. Amer. Math. Soc.*, to appear. arXiv 1302.6279.

[17] A. Hildebrand, On the number of positive integers $\leq x$ and free of prime factors $> y$, *J. Number Theory* **22** no. 3 (1986), 289–307. MR 0831874. Zbl 0575.10038. https://doi.org/10.1016/0022-314X(86)90013-2.

[18] A. Hildebrand and G. Tenenbaum, On integers free of large prime factors, *Trans. Amer. Math. Soc.* **296** no. 1 (1986), 265–290. MR 0837811. Zbl 0601.10028. https://doi.org/10.2307/2000573.

[19] A. Hildebrand and G. Tenenbaum, Integers without large prime factors, *J. Théor. Nombres Bordeaux* **5** no. 2 (1993), 411–484. MR 1265913. Zbl 0797.11070. https://doi.org/10.5802/jtnb.101.

[20] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58** (1963), 13–30. MR 0144363. Zbl 0127.10602. https://doi.org/10.1080/01621459.1963.10500830.

[21] S. Janson, T. Łuczak, T. Turova, and T. Vallier, Bootstrap percolation on the random graph $G_{n,p}$, *Ann. Appl. Probab.* **22** no. 5 (2012), 1989–2047. MR 3025687. Zbl 1254.05182. https://doi.org/10.1214/11-AAP822.

[22] T. Kleinjung, K. Aoki, J. Franke, A. K. Lenstra, E. Thomé, J. W. Bos, P. Gaudry, A. Kruppa, P. L. Montgomery, D. A. Osvik, H. te Riele, A. Timofeev, and P. Zimmermann, Factorization of a 768-bit RSA modulus, in *Advances in Cryptology—CRYPTO* 2010, *Lecture Notes in Comput. Sci.* **6223**, Springer, Berlin, 2010, pp. 333–350. MR 2725602. Zbl 1196.11167. https://doi.org/10.1007/978-3-642-14623-7_18.

[23] T. G. Kurtz, Solutions of ordinary differential equations as limits of pure jump Markov processes, *J. Appl. Probability* **7** no. 1 (1970), 49–58. MR 0254917. Zbl 0191.47301. https://doi.org/10.2307/3212147.

[24] L. Le Cam, An approximation theorem for the Poisson binomial distribution, *Pacific J. Math.* **10** (1960), 1181–1197. MR 0142174. Zbl 0118.33601. https://doi.org/10.2140/pjm.1960.10.1181.

[25] H. W. Lenstra, Jr., The number field sieve: an annotated bibliography, in *The Development of the Number Field Sieve*, *Lecture Notes in Math.* **1554**, Springer, Berlin, 1993, pp. 1–3. MR 1321217. Zbl 0806.11065f. https://doi.org/10.1007/BFb0091535.

[26] N. McNew, The most frequent values of the largest prime divisor function, *Exp. Math.* **26** no. 2 (2017), 210–224. MR 3623870. Zbl 06752333. https://doi.org/10.1080/10586458.2016.1155188.

[27] B. Pittel and G. B. Sorkin, The satisfiability threshold for k-XORSAT, *Combin. Probab. Comput.* **25** no. 2 (2016), 236–268. MR 3455676. Zbl 1372.68193. https://doi.org/10.1017/S0963548315000097.

[28] C. Pomerance, Analysis and comparison of some integer factoring algorithms, in *Computational Methods in Number Theory*, *Part I*, *Math. Centre Tracts* **154**, Math. Centrum, Amsterdam, 1982, pp. 89–139. MR 0700260. Zbl 0508.10004.

[29] C. Pomerance, The role of smooth numbers in number-theoretic algorithms, in *Proceedings of the International Congress of Mathematicians*, *Vol.* 1, 2 (Zürich, 1994), Birkhäuser, Basel, 1995, pp. 411–422. MR 1403941. Zbl 0854.11047.

[30] C. Pomerance, Multiplicative independence for random integers, in *Analytic Number Theory*, *Vol.* 2 (Allerton Park, IL, 1995), *Progr. Math.* **139**, Birkhäuser Boston, Boston, MA, 1996, pp. 703–711. MR 1409387. Zbl 0865.11085.

[31] C. Pomerance, A tale of two sieves, *Notices Amer. Math. Soc.* **43** no. 12 (1996), 1473–1485. MR 1416721. Zbl 1042.11529.

[32] R. A. Rankin, The difference between consecutive prime numbers, *J. London Math. Soc.* **11** no. 4 (1936), 242–245. MR 1574971. https://doi.org/10.1112/jlms/s1-13.4.242.

[33] A. Telcs, N. Wormald, and S. Zhou, Hamiltonicity of random graphs produced by 2-processes, *Random Structures Algorithms* **31** no. 4 (2007), 450–481. MR 2362639. Zbl 1129.05026. https://doi.org/10.1002/rsa.20133.

[34] D. Williams, *Probability with Martingales*, *Cambridge Math. Textbooks*, Cambridge University Press, Cambridge, 1991. MR 1155402. Zbl 0722.60001. https://doi.org/10.1017/CBO9780511813658.

[35] N. Wormald, The differential equation method for random graph processes and greedy algorithms, in *Lectures on Approximation and Randomized Algorithms*, PWN, Warsaw, 1999, Karonski, M. and Prömel, H. J., eds., pp. 73–155. Zbl 0943.05073.

(Received: August 15, 2016)

Department of Mathematical Sciences, University of Memphis, Memphis, TN, USA
E-mail: pbalistr@memphis.edu

Department of Pure Mathematics and Mathematical Statistics, Cambridge, United Kingdom and Department of Mathematical Sciences, University of Memphis, Memphis, TN, USA and London Institute for Mathematical Sciences, London, United Kingdom
E-mail: b.bollobas@dpmms.cam.ac.uk

IMPA, Rio de Janeiro, Brazil
E-mail: rob@impa.br