UPPER TAILS FOR ARITHMETIC PROGRESSIONS REVISITED

MATAN HAREL, FRANK MOUSSET, AND WOJCIECH SAMOTIJ

Abstract. Let X be the number of k-term arithmetic progressions contained in the p-biased random subset of the first N positive integers. We give asymptotically sharp estimates on the logarithmic upper-tail probability $\log \mathbb{P}(X \geqslant \mathbb{E}[X] + t)$ for all $\Omega(N^{-2/k}) \leqslant p \ll 1$ and all $t \gg \sqrt{\mathrm{Var}(X)}$, excluding only a few boundary cases. In particular, we show that the space of parameters (p, t) is partitioned into three phenomenologically distinct regions, where the upper-tail probabilities either resemble those of Gaussian or Poisson random variables, or are naturally described by the probability of appearance of a small set that contains nearly all of the excess t progressions. We employ a variety of tools from probability theory, including classical tilting arguments and martingale concentration inequalities. However, the main technical innovation is a combinatorial result that establishes a stronger version of 'entropic stability' for sets with rich arithmetic structure.

1. Introduction

Let $k \geqslant 3$ and N be positive integers. We write AP_k for the set of k-term arithmetic progressions (k-APs for short) in the set $[\![N]\!] \coloneqq \{1, 2, \ldots, N\}$, that is, AP_k is the collection of k-element subsets of $[\![N]\!]$ of the form $\{a, a+b, a+2b, \ldots, a+(k-1)b\}$, where a and b are positive integers.[1] Given $p \in [0, 1]$, we may choose a random subset of $[\![N]\!]$ by including each number independently with probability p. We write $\mathbf{R}$ for the random set obtained in this way and let X be the number of elements of AP_k that are contained in $\mathbf{R}$.

The goal of this work is to calculate the asymptotic behaviour, as N tends to infinity and k is fixed, of the logarithmic upper-tail probability of X, in the sparse regime (that is, we assume throughout the paper that p vanishes as N grows). To be more precise, our goal is to compute the asymptotic rate of $\log \mathbb{P}(X \geqslant \mathbb{E}[X] + t)$ for all (well-behaved) sequences t. For notational convenience, we set $\mu \coloneqq \mathbb{E}[X]$ and $\sigma^2 \coloneqq \mathrm{Var}(X)$.

There are a few cases which are straightforward. First, if $\mu + t$ is greater than $|\mathrm{AP}_k|$, the maximal number of k-term arithmetic progressions that can possibly be contained in $\mathbf{R}$, then the event $\{X \geqslant \mu + t\}$ is empty, and the logarithmic upper-tail probability is negative infinity. If μ is bounded, then X is asymptotically Poisson (see, e.g., [6]), which answers the question when t is also bounded. Furthermore, a sequence of now-classical works from the 1980s (see, for example, [5, 24]) implies that X satisfies a Central Limit Theorem; i.e., whenever $\mu \to \infty$, then $(X - \mu)/\sigma$ converges weakly to a standard Gaussian

Date: September 12, 2024.

This research was supported by: the Israel Science Foundation grant 2110/22; the grant 2019679 from the United States–Israel Binational Science Foundation (BSF) and the United States National Science Foundation (NSF); and the ERC Consolidator Grant 101044123 (RandomHypGra).

[1] Clearly, AP_k depends on N. However, as the meaning of N will be same throughout the paper, we will often omit the explicit mention of N in the various notations. Similarly, we will omit mentioning the dependence on k whenever it seems safe to do so.

random variable, which answers the question in the case where t/σ is bounded. Unfortunately, when $t/\sigma \to \infty$, this result only tells us that $\mathbb{P}(X \geqslant \mu + t)$ vanishes and cannot be used directly to deduce any quantitative information on the rate of convergence. For values of p which vanish sufficiently slowly, it is possible to prove Berry–Esseen-like bounds on the rate of convergence via Stein's method (see, e.g., [23]); when they are available, such bounds can be leveraged to prove Gaussian behaviour if $t/\sigma \to \infty$ very slowly. Such techniques will not be sufficient to prove Gaussian bounds for a vast majority of the Gaussian regime that will be discussed in this paper.

The remaining regimes of the upper-tail problem can be divided into three cases: the case where t is much smaller than μ (but much larger than σ), known as the *moderate-deviation regime*; the case where t is commensurate with μ, known as the *large-deviation regime*; and the case where t is much greater than μ, which has received comparatively little attention, that we will term the *extreme-deviation regime*.

Historically, the large-deviation regime has been the most studied one. The main reason for this is that the upper bounds on the logarithmic upper-tail probability of X that can be proved via classical concentration inequalities do not match the known lower bounds, even up to constant factors; see [22] for a survey of such results. A breakthrough was achieved by the work of Chatterjee–Dembo [10], which established a large-deviation principle for a wide class of non-linear functions of independent random variables. Their result was later extended and generalised by Eldan [13], Augeri [2], and Austin [3]. Subsequently, Bhattacharya–Ganguly–Shao–Zhao [9] showed that these large-deviation principles apply in the context of k-APs and solved the associated variational problem to obtain asymptotically tight estimates for the logarithmic upper-tail probability, for a suboptimal range of the density parameter p. Around the same time, Warnke [25] developed a sophisticated moment-based approach in order to prove bounds on the logarithmic upper-tail probability that were correct only up to a multiplicative constant factor, but held in the entire large-deviation regime. Finally, the three authors [18] determined the asymptotic logarithmic upper-tail probability in the entire large-deviation regime using a combinatorial approach paired with a conditioned high-moment calculation.

The moderate-deviation regime has also garnered some recent attention. In particular, the aforementioned results of both Bhattacharya–Ganguly–Shao–Zhao [9] and Warnke [25] extend to portions of this regime. As before, the results of [9] determine the exact asymptotics whereas [25] computes only the order of magnitude. Both results hold under strong assumptions on the density p; moreover, [9] further requires the deviation t not to be too far from the expectation. The recent work of Griffiths, Koch, and Secco [17] determines exact asymptotics of the logarithmic upper-tail probability in a substantially larger, but still incomplete, portion of the moderate-deviation regime (see also [14], where a similar result is obtained in the setting of k-term arithmetic progressions modulo a prime).

Finally, although the extreme-deviation regime is not explicitly mentioned in most of the above works, many of the arguments can be extended to cases where t is much larger than μ, except when μ is only polylogarithmic in N.

1.1. **Main results.** Our main contribution is to determine the asymptotic rate of the logarithmic upper-tail probability $\log \mathbb{P}(X \geqslant \mu + t)$ for all values of p and t, with the exception of a few liminal cases and the regime $p = \Theta(1)$. To state the results, we require a few preliminaries. First, it is straightforward to

verify that, for some positive $C = C(k)$,

$$\mu = (1+o(1)) \cdot \frac{N^2 p^k}{2(k-1)} \qquad \text{and} \qquad \sigma^2 = (1+o(1)) \cdot \frac{N^2 p^k}{2(k-1)}(1 + CNp^{k-1}). \tag{1}$$

We also define the function

$$\mathrm{Po}(x) \coloneqq \int_0^x \log(1+y)\, dy = (1+x)\log(1+x) - x,$$

which naturally appears in the rate function of Poisson random variables. Finally, given a $U \subseteq [\![N]\!]$, we set $\mathbb{E}_U[X] \coloneqq \mathbb{E}[X \mid U \subseteq \mathbf{R}]$, and, for any $t \geqslant 0$, we define

$$\Psi(t) = \Psi_{N,p,k}(t) \coloneqq \min\{|U| : \mathbb{E}_U[X] \geqslant \mu + t\}, \tag{2}$$

with the convention that $\Psi(t) = \infty$ if the set being optimised over is empty.

Definition. We say that the sequence (p,t) is in:

- the *Gaussian regime* if

$$N^{-1/(k-1)} \ll p \ll 1, \quad t \gg \sigma, \quad \text{and} \quad \sqrt{t}\log(1/p) \gg t^2/\sigma^2;$$

- the *Poisson regime* if

$$\Omega(N^{-2/k}) \leqslant p \ll N^{-1/(k-1)}, \quad t \gg \sigma, \quad \text{and} \quad \sqrt{t}\log(1/p) \gg \mu \cdot \mathrm{Po}(t/\mu);$$

- the *localised regime* if either

$$N^{-1/(k-1)} < p \ll 1 \quad \text{and} \quad \sqrt{t}\log(1/p) \ll t^2/\sigma^2,$$

  or

$$\Omega(N^{-2/k}) \leqslant p \leqslant N^{-1/(k-1)} \quad \text{and} \quad \sqrt{t}\log(1/p) \ll \mu \cdot \mathrm{Po}(t/\mu).$$

The three regimes are depicted in Figure 1, together with a fourth regime where t/σ is bounded and the Central Limit Theorem applies.

Theorem 1.1. *Assume $k \geqslant 3$ and let X be the number of k-term arithmetic progressions contained in the random subset of $[\![N]\!]$ obtained by including each number independently with probability p.*

- *If (p,t) is in the Gaussian regime, then*

$$-\log \mathbb{P}(X \geqslant \mu + t) = (1+o(1)) \cdot \frac{t^2}{2\sigma^2}.$$

- *If (p,t) is in the Poisson regime, then*

$$-\log \mathbb{P}(X \geqslant \mu + t) = (1+o(1)) \cdot \mu \cdot \mathrm{Po}(t/\mu).$$

- *If (p,t) is in the localised regime, then*

$$-\log \mathbb{P}(X \geqslant \mu + t) = (1+o(1)) \cdot \Psi(t) \cdot \log(1/p);$$

  moreover, if $\mu + t \leqslant |\mathrm{AP}_k|$, then $\Psi(t) = (1+o(1)) \cdot \sqrt{2(k-1)t}$.

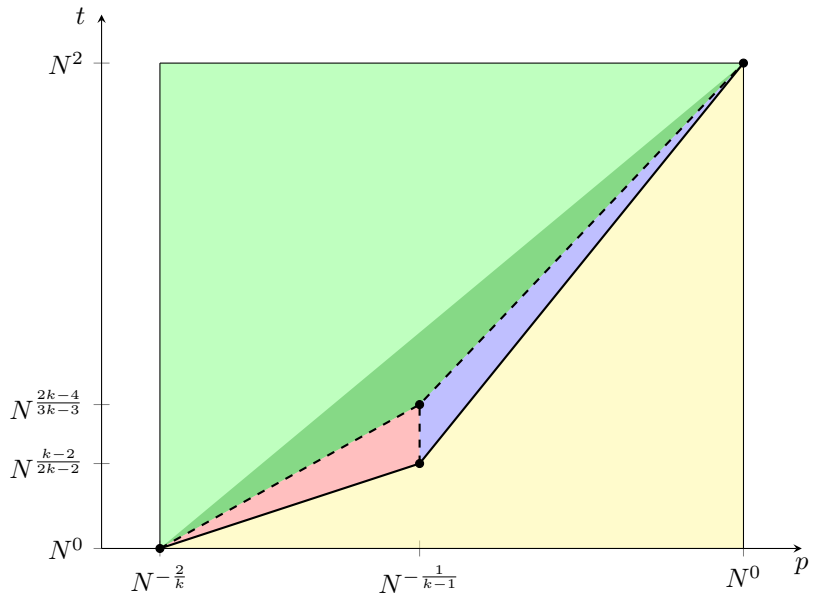

FIGURE 1. Phase diagram for the upper-tail problem for k-term arithmetic progressions, with logarithmic axes. The green region north west of the two oblique dashed lines represents the localised regime: the darker subregion is the moderate-deviation regime, the lighter one the extreme-deviation regime, and the boundary between the two is the large-deviation regime. The triangular blue region (east of the vertical dashed line segment) represents the Gaussian regime, and the triangular red region (west of the vertical dashed line segment) represents the Poisson regime. The yellow region south east of the two oblique solid lines is the region where the Central Limit Theorem holds. Theorem 1.1 gives no information on what happens on the dashed lines.

Heuristically, one may think of three different strategies to increase the number of k-term arithmetic progressions by t. First, we may add to the p-biased random set $\mathbf{R}$ a small but highly structured subset that contains the t excess arithmetic progressions: this leads to the localised regime. The other two strategies are more 'global', in the sense that the excess arithmetic progressions are spread out roughly evenly over $[\![N]\!]$. We can do this either by increasing the probability of the events $\{i \in \mathbf{R}\}$ in a roughly uniform fashion (this leads to the Gaussian regime), or by superimposing $\mathbf{R}$ and the union of t distinct, arithmetic progressions chosen uniformly at random (this leads to the Poisson regime). One can provide a convincing heuristic calculation that associates to each strategy the respective quantitative bound in Theorem 1.1; indeed, in each case, the rate function is precisely the Kullback–Leibler divergence of the random set obtained by applying the corresponding strategy from the original random set $\mathbf{R}$. (Having said that, turning these intuitions into rigorous arguments requires some work.) It is straightforward to check that the three regimes are the regions where the respective strategy is the 'cheapest', in the sense of leading to the smallest rate function. In light of this, the main contribution of Theorem 1.1 is to show that, away from the boundary between the regimes, one of these three strategies will always dominate the upper-tail event, up to lower order corrections.

Remark. Expanding $\mathrm{Po}(\cdot)$ in Taylor series, one can show that, when $t \ll \mu$,

$$\mu \cdot \mathrm{Po}(t/\mu) = (1 + o(1)) \cdot \frac{t^2}{2\mu}.$$

Under the additional assumption that $Np^{k-1} \ll 1$, and thus $\sigma^2 = (1 + o(1)) \cdot \mu$, the asymptotic rates of the Gaussian and the Poisson regimes coincide. Despite this, there are good reasons to consider the two regimes separately. First, for a narrow range of parameters (when μ is polylogarithmic in N), the Poisson regimes includes regions where t is commensurate or much larger than μ; when this occurs, the rate function in the Poisson regime is significantly smaller than $t^2/(2\sigma^2)$. Second, the two regimes are qualitatively very different, since, unlike in the Gaussian regime, the rate function in the Poisson regime no longer agrees with the naive mean-field prediction, as will be discussed in greater detail below. Last but not least, the different phenomenology in the two regimes requires vastly different approaches for bounding the tail probabilities from both above and below.

In the localised regime, Theorem 1.1 reduces the upper-tail question to the solution of a variational problem encoded by Ψ. Following [18], we consider the family of *t-seeds* – sets that increase the conditional expectation of X by at least t:

$$\mathcal{S}(t) = \mathcal{S}_{N,p,k}(t) \coloneqq \{U \subseteq [\![N]\!] : \mathbb{E}_U[X] \geqslant \mu + t\}. \tag{3}$$

As we will show in Section 4, the appearance of a $(1 + o(1))t$-seed implies the upper-tail event with a probability that is very high compared to the probability of appearance of the seed itself. Moreover, by picking a particular t-seed that realises the minimum in (2), one can deduce that the probability of appearance of a $(1 + o(1))t$-seed in $\mathbf{R}$ is bounded below by $p^{(1+o(1))\cdot\Psi(t)}$. From this, the lower bound of the localised regime in Theorem 1.1 follows immediately. The heart of the argument of [18] that resolved the large-deviation regime was showing that, for every fixed $\delta > 0$, the probability that $\mathbf{R}$ contains a $\delta\mu$-seed U satisfying $|U| = O\big(\Psi(\delta\mu)\log(1/p)\big)$ is $p^{(1+o(1))\Psi(\delta\mu)}$, which is the probability of the appearance of a smallest such seed. The following theorem, which is the main technical innovation of this paper, shows that the analogous statement about t-seeds remains true not only for all t but also for a much broader range of sizes of the seeds.

Theorem 1.2. *Assume $k \geqslant 3$ and let p, t, m be such that*

$$t \gg m \cdot \max\{1, Np^{k-1}\} \quad \text{and} \quad t \gg m^2 p^{k-2} \cdot N^{(k-2)(m/t)^{1/(k-1)}}. \tag{4}$$

Then

$$\log \mathbb{P}\big(U \subseteq \mathbf{R} \text{ for some } U \in \mathcal{S}_{N,p,k}(t) \text{ with } |U| \leqslant m\big) \leqslant (1 - o(1)) \cdot \Psi(t) \cdot \log p.$$

Remark. We claim that the lower-bound assumptions on t are natural. First, ignoring lower-order terms, every union of $t \leqslant m/k$ distinct k-APs forms a t-seed of size at most m, and so the probability of appearance of such a seed is at least $\mathbb{P}(X \geqslant t)$. However, since we expect that the planting of a smallest $(t - \mu)$-seed makes the event $\{X \geqslant t\}$ significantly more likely, it is plausible (and, up to lower-order corrections, true) that $\mathbb{P}(X \geqslant t) \geqslant p^{\Psi(t-\mu)}$, which is much larger than the upper bound in the theorem, at least when $t = O(\mu)$. A similar argument applies for all $t = O(m)$, so the assumption that t is much larger than m is really needed.

Second, observe that every set $U \subseteq [\![N]\!]$ with m elements satisfies

$$\mathbb{E}_U[X] - \mathbb{E}[X] \geqslant c_k \cdot \left(Nmp^{k-1} + m^2 p^{k-2}\right)$$

for some constant c_k that depends only on k; to see this, consider the k-APs intersecting U in either one or two elements. In particular, if $t \leqslant c_k Nmp^{k-1}$ or $t \leqslant c_k m^2 p^{k-2}$, then every m-element set is a t-seed. Consequently, at least for $m \leqslant Np$, the probability that $\mathbf{R}$ contains a t-seed with at most m elements is uniformly bounded from below, contradicting the vanishing upper bound stated by the theorem. Note that the above argument only justifies a lower bound of the form $t \geqslant Cm^2p^{k-2}$. The extra factor of $N^{(k-2)(m/t)^{1/(k-1)}}$ is needed for technical reasons; however, it is irrelevant once $t/m \gg (\log N)^{k-1}$.

One may well find it believable that Theorem 1.2 plays a direct role in the proof of the upper bound for the localised regime, where, following [18], we use a modified moment argument to show that the upper-tail event is dominated by the appearance of a 'small' seed. It is perhaps more surprising that it also plays a crucial role in proving the upper bound of the Poisson regime. In that context, it allows us to exclude certain inconvenient terms that arise when calculating the factorial moments of X; these terms correspond to small subsets with rich additive structure. In fact, the estimates of factorial moments of X that play the central role in our treatment of the Poisson regime extend to a portion of the localised regime. This proves crucial, as there is a small portion of the localised regime (which we term the *very sparse localised regime*) where the aforementioned argument based on estimating classical moments of X fails, but can be salvaged by factorial moment estimates. (This does not mean, however, that the very sparse localised regime is phenomenologically distinct from the rest of the localised regime; see Section 4.2 for further discussion.) In contrast, the upper bound for the Gaussian regime is proved by way of a truncated martingale concentration argument, generalising a classical inequality of Freedman [15]. The truncation scheme uses fairly straightforward moment estimates rather than the more powerful Theorem 1.2.

1.2. **The naive mean-field approximation.** One way to view Theorem 1.1 is in the context of the naive mean-field approximation. For a pair $\mathbb{P}$ and $\mathbb{Q}$ of measures on subsets of $[\![N]\!]$, with $\mathbb{Q}$ absolutely continuous with respect to $\mathbb{P}$, the *Kullback–Leibler divergence* of $\mathbb{Q}$ from $\mathbb{P}$ is defined by

$$D_{\mathrm{KL}}(\mathbb{Q} \,\|\, \mathbb{P}) \coloneqq \mathbb{E}_{\mathbb{Q}}\left[\log\left(\frac{d\mathbb{Q}}{d\mathbb{P}}(\mathbf{R})\right)\right] = \sum_{R \subseteq [\![N]\!]} \mathbb{Q}(\mathbf{R} = R) \log\left(\frac{\mathbb{Q}(\mathbf{R} = R)}{\mathbb{P}(\mathbf{R} = R)}\right), \tag{5}$$

where $\mathbb{E}_{\mathbb{Q}}$ is the expectation operator associated with the measure $\mathbb{Q}$. It is known (cf. Section 2) that the logarithmic probability of *any* event $\mathcal{A}$ can be obtained by optimising the Kullback–Leibler divergence over all measures that assign $\mathcal{A}$ probability one:

$$-\log \mathbb{P}(\mathcal{A}) = \inf_{\substack{\mathbb{Q} \ll \mathbb{P}, \\ \mathbb{Q}(\mathcal{A})=1}} D_{\mathrm{KL}}(\mathbb{Q} \,\|\, \mathbb{P}). \tag{6}$$

The usefulness of such a formulation is limited by the fact that measures that assign the upper-tail events probability one may be quite difficult to analyse. The idea of the naive mean-field approximation is to replace the complicated variational problem in (6) by a simpler one, where the infimum ranges only over product measures (the assumption $\mathbb{Q}(\mathcal{A}) = 1$ must then be relaxed somewhat). Roughly speaking, the naive mean-field approximation holds if minimising over this smaller set still achieves (6), up to lower

order corrections. More precisely, we say the naive mean-field approximation holds for a sequence of events $\mathcal{A}_N$, each defined on a measure space $(\Omega_N, \mathbb{P}_N)$, if

$$\inf_{\substack{\mathbb{Q}_N \ll \mathbb{P}_N, \\ \lim_{N\to\infty} \mathbb{Q}_N(\mathcal{A}_N)=1 \\ \mathbb{Q}_N \text{ is a product measure}}} D_{\mathrm{KL}}(\mathbb{Q}_N \,\|\, \mathbb{P}_N) = -(1+o(1)) \cdot \log \mathbb{P}_N(\mathcal{A}_N). \tag{7}$$

The aforementioned large-deviation principles proved in [2, 10, 13] establish a version of (7) when $\mathcal{A}_N$ are tail events for non-linear functions of independent random variables that satisfy certain complexity and smoothness properties. Bhattacharya–Ganguly–Shao–Zhao [9] showed that the number of arithmetic progressions in $\mathbf{R}$ has the requisite properties when the density p is sufficiently large. Furthermore, the same work solved the restricted variational problem of (7) in the case $\mathcal{A}_N = \{X \geqslant \mu + t\}$ for (nearly) all values of (p,t) with p vanishing and $t \gg \sigma$. Unsurprisingly, this solution matches the results of Theorem 1.1 in the entire Gaussian and localised regimes; *a posteriori*, Theorem 1.1 thus establishes that the naive mean-field approximation is valid in those regimes. In contrast, the naive mean-field approximation completely fails in the Poisson regime – the left-hand side of (7) is not even of the same order of magnitude as the right-hand side.

1.3. **Related works.** The study of large- and moderate-deviation regimes of the upper tail of random variables that arise from combinatorial settings has flowered in the last decade. Besides the aforementioned work of Chatterjee–Dembo [10], Eldan [13], Augeri [2], and Austin [3], which are concerned with rather general non-linear functions of independent random variables, there have been numerous works that focus on more specific cases. The most-studied family of examples are the random variables X_H that count copies of a given graph H in the binomial random graph $G_{n,p}$. Cook–Dembo [11] determined the asymptotics of the logarithmic upper-tail probability of X_H for all H and all p satisfying $n^{-c_H} \ll p \ll 1$ for some positive c_H that depends only on H. More specifically, they established that the naive mean-field approximation holds for X_H in the above range of densities. Later work of Cook–Dembo–Pham [12] extended these results to a wider range of densities p and generalised them to the case where H is a uniform hypergraph. The three authors [18] determined the asymptotics of the logarithmic upper-tail probability of X_H for all regular, non-bipartite H for essentially all densities p (also in the non-mean-field regime); their results were extended to regular, bipartite graphs by Basak–Basu [7]. In the the moderate-deviation regime, Goldschmidt–Griffths–Scott [16] proved asymptotic upper-tail estimates for arbitrary subgraphs for a certain restricted range of densities p and deviations t (using the notation of this paper). Recently, Alvarado–de Oliviera–Griffiths [1] successfully analysed a far greater (but still sub-optimal) portion of the moderate-deviation regime in the case where H is a triangle.

Finally, there has been some recent progress in the understanding of the typical deviations of the number of k-APs in random subsets of $\mathbb{Z}/(N\mathbb{Z})$, the cyclic group of order N. Berkowitz–Sah–Sawhney [8] showed that, at least when p is fixed, the standard notion of a local Central Limit Theorem fails for infinitely many N, in the sense that the probability that the number of k-APs equals a particular integer deviates significantly from the prediction one would get from the Gaussian limit.

1.4. **Organisation.** The paper is organised as follows: Section 2 includes an overview of the tilting argument, a classical method for producing lower bounds for rare events, as well as a proof of the

martingale concentration inequality used for the upper bound of the Gaussian regime. Section 3 is dedicated to proving Theorem 1.2. The remaining three sections (Sections 4 to 6) prove Theorem 1.1 for the localised, Gaussian, and Poisson regimes, respectively; the proof of the key estimate needed for the very sparse localised regime is postponed to Section 6, as it is uses methods developed for the Poisson regime. Finally, Appendix A proves a bound on the number of connected hypergraphs with small edge boundary that plays a key role in the analysis of the Poisson regime (and the very sparse localised regime), and may be of independent interest.

2. Probabilistic tools

2.1. **The tilting argument.** The tilting argument is a general method to bound the probability of an arbitrary event from below by constructing another measure that makes the event likely to occur and quantifying its 'distance' from the original measure. Suppose that $\mathbb{P}$ and $\mathbb{Q}$ are two measures on subsets of $[\![N]\!]$. If $\mathbb{Q} \ll \mathbb{P}$ (that is, if $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$), there is a unique (up to a set of measure zero) measurable function $d\mathbb{Q}/d\mathbb{P}$, called the Radon–Nikodym derivative, such that $\mathbb{Q}(\mathcal{A}) = \mathbb{E}\left[d\mathbb{Q}/d\mathbb{P} \cdot \mathbb{1}_{\mathcal{A}}\right]$ for every event $\mathcal{A}$. In this case, we define the *Kullback–Leibler divergence* of $\mathbb{Q}$ from $\mathbb{P}$ by

$$D_{\mathrm{KL}}(\mathbb{Q} \,\|\, \mathbb{P}) \coloneqq \mathbb{E}_{\mathbb{Q}}\left[\log \frac{d\mathbb{Q}}{d\mathbb{P}}(\mathbf{R})\right] = \sum_{R \subseteq [\![N]\!]} \mathbb{Q}(\mathbf{R} = R) \log\left(\frac{\mathbb{Q}(\mathbf{R} = R)}{\mathbb{P}(\mathbf{R} = R)}\right), \tag{8}$$

where we use the convention that $0 \log 0 = 0$. It is routine to verify that the Kullback–Leibler divergence between any two measures is nonnegative. We will also make use of the following easily verifiable additivity property of the Kullback–Leibler divergence.

Fact 2.1. *If $\mathbb{P}_1, \dots, \mathbb{P}_N$ and $\mathbb{Q}_1, \dots, \mathbb{Q}_N$ are probability measures with $\mathbb{Q}_i \ll \mathbb{P}_i$ for all $i \in [\![N]\!]$, then*

$$D_{\mathrm{KL}}\big(\mathbb{Q}_1 \times \cdots \times \mathbb{Q}_N \,\|\, \mathbb{P}_1 \times \cdots \times \mathbb{P}_N\big) = \sum_{i=1}^{N} D_{\mathrm{KL}}(\mathbb{Q}_i \,\|\, \mathbb{P}_i).$$

It is well known that one can use the notion of Kullback–Leibler divergence to produce a lower bound for the logarithmic probability of any event $\mathcal{A}$ under $\mathbb{P}$ by considering a measure $\mathbb{Q} \ll \mathbb{P}$ with $\mathbb{Q}(\mathcal{A}) = 1$.

Proposition 2.2. *Let $\mathcal{A}$ be an arbitrary event and let $\mathbb{P}$ and $\mathbb{Q}$ be two measures such that $\mathbb{Q}(\mathcal{A}) = 1$ and $\mathbb{Q} \ll \mathbb{P}$. Then*

$$\log \mathbb{P}(\mathcal{A}) \geqslant -D_{\mathrm{KL}}(\mathbb{Q} \,\|\, \mathbb{P}).$$

Proof. Since our assumptions imply that $\mathbb{P}(\mathcal{A}) > 0$, we may consider the conditioned measure $\mathbb{P}^* \coloneqq \mathbb{P}(\,\cdot \mid \mathcal{A})$. Denoting the indicator random variable of $\mathcal{A}$ by $\mathbb{1}_{\mathcal{A}}$, we observe that $d\mathbb{P}^*/d\mathbb{P} = \mathbb{1}_{\mathcal{A}}/\mathbb{P}(\mathcal{A})$ and that $\mathbb{Q} \ll \mathbb{P}^*$. Since the Kullback–Leibler divergence is always nonnegative,

$$-D_{\mathrm{KL}}(\mathbb{Q} \,\|\, \mathbb{P}) \leqslant D_{\mathrm{KL}}(\mathbb{Q} \,\|\, \mathbb{P}^*) - D_{\mathrm{KL}}(\mathbb{Q} \,\|\, \mathbb{P}) = \mathbb{E}_{\mathbb{Q}}\left[-\log \frac{d\mathbb{P}^*}{d\mathbb{P}}(\mathbf{R})\right],$$

where the final equality follows because, $\mathbb{Q}$-almost surely,

$$\frac{d\mathbb{Q}}{d\mathbb{P}^*} \cdot \left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)^{-1} = \left(\frac{d\mathbb{P}^*}{d\mathbb{P}}\right)^{-1}.$$

Since $\mathbb{Q}(\mathcal{A}) = 1$, we find that $d\mathbb{P}^*/d\mathbb{P} = 1/\mathbb{P}(\mathcal{A})$ holds $\mathbb{Q}$-almost surely. Therefore,

$$\mathbb{E}_{\mathbb{Q}}\left[-\log \frac{d\mathbb{P}^*}{d\mathbb{P}}(\mathbf{R})\right] = \log \mathbb{P}(\mathcal{A}),$$

which implies the desired inequality. □

In fact, the proof of Proposition 2.2 shows that $-\log \mathbb{P}(\mathcal{A})$ is precisely equal to the Kullback–Leibler divergence of $\mathbb{P}(\,\cdot \mid \mathcal{A})$ from $\mathbb{P}$. This allows us to restate the proposition as:

$$-\log \mathbb{P}(\mathcal{A}) = \inf_{\substack{\mathbb{Q} \ll \mathbb{P}, \\ \mathbb{Q}(\mathcal{A})=1}} D_{\mathrm{KL}}(\mathbb{Q} \,\|\, \mathbb{P}). \tag{9}$$

As mentioned before, (9) is a theoretically useful tool that is difficult to apply, since the set of measures that assign $\mathcal{A}$ probability one can be rather unwieldy. Below, we will derive two versions of this variational principle that are more immediately applicable. The first, Corollary 2.3, applies to arbitrary measures and will be used for lower bounds in the localised regime. The second, Proposition 2.4, applies only to measures that assign $\mathcal{A}$ probability one asymptotically; it will be used in the Gaussian and Poisson regimes.

Corollary 2.3. *For any event $\mathcal{A}$ and any measures $\mathbb{P}$ and $\mathbb{Q}$ such that $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{Q}(\mathcal{A}) > 0$,*

$$\log \mathbb{P}(\mathcal{A}) \geqslant \log \mathbb{Q}(\mathcal{A}) - \mathbb{E}_{\mathbb{Q}}\left[\log \frac{d\mathbb{Q}}{d\mathbb{P}}(\mathbf{R}) \mid \mathcal{A}\right] \tag{10}$$

Proof. We apply Proposition 2.2 to $\mathbb{Q}(\,\cdot \mid \mathcal{A})$. Since

$$\frac{d\mathbb{Q}(\,\cdot \mid \mathcal{A})}{d\mathbb{P}} = \frac{1}{\mathbb{Q}(\mathcal{A})} \cdot \frac{d\mathbb{Q}}{d\mathbb{P}}$$

holds $\mathbb{Q}(\,\cdot \mid \mathcal{A})$-almost surely, writing J in place of $\log(d\mathbb{Q}/d\mathbb{P})$, we find that

$$D_{\mathrm{KL}}(\mathbb{Q}(\,\cdot \mid \mathcal{A}) \,\|\, \mathbb{P}) = -\log \mathbb{Q}(\mathcal{A}) + \mathbb{E}_{\mathbb{Q}(\cdot \mid \mathcal{A})}[J(\mathbf{R})] = -\log \mathbb{Q}(\mathcal{A}) + \mathbb{E}_{\mathbb{Q}}[J(\mathbf{R}) \mid \mathcal{A}].$$

The claim now follows from Proposition 2.2. □

Proposition 2.4. *Let $\mathbb{P}_N$ and $\mathbb{Q}_N$ be two sequences of measures satisfying $\mathbb{Q}_N \ll \mathbb{P}_N$ for each N and suppose that $\{\mathcal{A}_N\}$ is a sequence of events such that $\limsup_{N\to\infty} \mathbb{P}_N(\mathcal{A}_N) < 1$ and $\lim_{N\to\infty} \mathbb{Q}_N(\mathcal{A}_N) = 1$. Then*

$$\liminf_{N\to\infty} \frac{D_{\mathrm{KL}}(\mathbb{Q}_N \,\|\, \mathbb{P}_N)}{-\log \mathbb{P}(\mathcal{A}_N)} \geqslant 1.$$

Proof. We first claim that

$$D_{\mathrm{KL}}(\mathbb{Q}_N \,\|\, \mathbb{P}_N) - \mathbb{Q}_N(\mathcal{A}_N) \cdot \log \frac{\mathbb{Q}_N(\mathcal{A}_N)}{\mathbb{P}_N(\mathcal{A}_N)} - \mathbb{Q}_N(\mathcal{A}_N^c) \cdot \log \frac{\mathbb{Q}(\mathcal{A}_N^c)}{\mathbb{P}(\mathcal{A}_N^c)} \geqslant 0. \tag{11}$$

Indeed, for every event $\mathcal{E}$ with $\mathbb{Q}_N(\mathcal{E}) > 0$ (and thus $\mathbb{P}_N(\mathcal{E}) > 0$), we have

$$\frac{d\mathbb{Q}_N(\,\cdot \mid \mathcal{E})}{d\mathbb{P}_N(\,\cdot \mid \mathcal{E})} = \frac{d\mathbb{Q}_N}{d\mathbb{P}_N} \cdot \frac{\mathbb{P}_N(\mathcal{E})}{\mathbb{Q}_N(\mathcal{E})}$$

and thus the left-hand side of the above inequality can be seen to equal

$$\mathbb{Q}_N(\mathcal{A}_N) \cdot D_{\mathrm{KL}}\big(\mathbb{Q}_N(\,\cdot \mid \mathcal{A}_N) \,\|\, \mathbb{P}_N(\,\cdot \mid \mathcal{A}_N)\big) + \mathbb{Q}_N(\mathcal{A}_N^c) \cdot D_{\mathrm{KL}}\big(\mathbb{Q}_N(\,\cdot \mid \mathcal{A}_N^c) \,\|\, \mathbb{P}_N(\,\cdot \mid \mathcal{A}_N^c)\big),$$

which is clearly nonnegative. Dividing (11) through by $-\log \mathbb{P}_N(\mathcal{A}_N)$ and rearranging the terms gives

$$\frac{D_{\mathrm{KL}}(\mathbb{Q}_N \,\|\, \mathbb{P}_N)}{-\log \mathbb{P}_N(\mathcal{A}_N)} \geqslant \mathbb{Q}_N(\mathcal{A}_N) - \frac{1}{\log \mathbb{P}_N(\mathcal{A}_N)} \cdot \left(\mathbb{Q}_N(\mathcal{A}_N) \cdot \log \mathbb{Q}_N(\mathcal{A}_N) + \mathbb{Q}_N(\mathcal{A}_N^c) \cdot \log \frac{\mathbb{Q}_N(\mathcal{A}_N^c)}{\mathbb{P}_N(\mathcal{A}_N^c)}\right).$$

Finally, our assumptions on the sequences $\mathbb{P}_N(\mathcal{A}_N)$ and $\mathbb{Q}_N(\mathcal{A}_N)$ imply that the first summand in the right-hand side of the above inequality tends to one whereas the second summand tends to zero. The desired inequality follows by taking the limit inferior of both sides. □

2.2. **A martingale concentration inequality.** The main tool for establishing the upper bound in the Gaussian regime is a martingale concentration inequality, which we formulate in the general context of hypergraphs. Let $\mathcal{H}$ be a hypergraph with vertex set $[\![N]\!]$ and let $\mathbf{R}$ denote the p-biased random subset of $[\![N]\!]$. Let X be the number of edges of in $\mathcal{H}[\mathbf{R}]$, the subhypergraph of $\mathcal{H}$ that is induced by $\mathbf{R}$, and denote the mean and the variance of X by μ and σ^2, respectively. If $\mathcal{H}$ comprises the k-term arithmetic progressions in $[\![N]\!]$, these notations coincide with the ones used in the rest of paper. Considering the upper-tail problem for arithmetic progressions in such an abstract setup of hypergraphs is not a new idea – both [17, 25] follow this route.

Our upper bound for the upper tail of X, which could be of independent interest, is a sum of a Gaussian-like tail bound and three upper-tail probabilities for various functions of the numbers of edges that the random set $\mathbf{R}$ induces in the link hypergraphs of the vertices of $\mathcal{H}$. For every $i \in [\![N]\!]$, we let

$$L_i \coloneqq \big|\big\{e \in \mathcal{H} : e \ni i \text{ and } e \setminus \{i\} \subseteq \mathbf{R}\big\}\big| \tag{12}$$

Proposition 2.5. *The following holds for all sufficiently small $\varepsilon > 0$. Suppose that $\mathcal{H}$ is a hypergraph with vertex set $[\![N]\!]$. Let $\mathbf{R}$ be the p-biased random subset of $[\![N]\!]$ and let $L_1, \dots, L_N$ be the random variables defined in* (12)*. Write $X \coloneqq e(\mathcal{H}[\mathbf{R}])$, $\mu \coloneqq \mathbb{E}[X]$, and $\sigma^2 \coloneqq \mathrm{Var}(X)$. Then for all $t \geqslant \varepsilon\sigma$, we have, letting $\lambda \coloneqq t/\sigma^2$,*

$$\mathbb{P}(X \geqslant \mu + t) \leqslant \exp\left(-\frac{(1-\varepsilon)t^2}{2\sigma^2}\right) + \frac{8N}{\varepsilon^3} \cdot \mathbb{P}\left(\sum_{i=1}^{N} L_i^2 > \left(1 + \frac{\varepsilon}{10}\right) \cdot \frac{\sigma^2}{p}\right) + \frac{8N}{\varepsilon^3} \cdot \mathbb{P}\left(\left|\left\{i : L_i > \frac{\varepsilon}{\lambda}\right\}\right| \geqslant \frac{\varepsilon\lambda^2\sigma^2}{20p^{1/2}}\right) + \mathbb{P}\left(\exists i \;\; L_i > \frac{\log(1/p)}{2\lambda}\right).$$

Proof. Let Y_i be the indicator random variable of the event $\{i \in \mathbf{R}\}$ and, for every $i \in \{0, \dots, N\}$, let $\mathcal{F}_i$ be the σ-algebra generated by $Y_1, \dots, Y_i$. The starting point for our considerations is the following identity, which holds for all $i \in [\![N]\!]$:

$$\mathbb{E}[X \mid \mathcal{F}_i] - \mathbb{E}[X \mid \mathcal{F}_{i-1}] = (Y_i - p) \cdot \mathbb{E}\left[L_i \mid \mathcal{F}_{i-1}\right]. \tag{13}$$

Instead of working with the Doob martingale $\big(\mathbb{E}[X \mid \mathcal{F}_i]\big)_{i=0}^N$ directly, we will consider a related martingale sequence whose differences are truncated versions of (13). More precisely, for each $i \in [\![N]\!]$, set

$$\hat{L}_i \coloneqq \min\left\{L_i, \log(1/p)/(2\lambda)\right\}$$

and define a martingale sequence $(M_i)_{i=0}^N$ by

$$M_0 \coloneqq \mathbb{E}[X] \qquad \text{and} \qquad M_i - M_{i-1} \coloneqq (Y_i - p) \cdot \mathbb{E}\big[\hat{L}_i \mid \mathcal{F}_{i-1}\big] \quad \text{for } i \in [\![N]\!].$$

Since the random variables X and M_N coincide on the event that $\hat{L}_i = L_i$ for all i, we find that

$$\mathbb{P}(X \geqslant \mu + t) \leqslant \mathbb{P}(M_N - M_0 \geqslant t) + \mathbb{P}\left(\exists i \;\; L_i > \frac{\log(1/p)}{2\lambda}\right). \tag{14}$$

In the remainder of the proof, we will estimate the first probability on the right-hand side of (14).

Define the function $\phi\colon \mathbb{R} \to \mathbb{R}$ by

$$\phi(0) \coloneqq \frac{1}{2} \qquad \text{and} \qquad \phi(x) \coloneqq \frac{e^x - x - 1}{x^2} \quad \text{if } x \neq 0$$

and observe that ϕ is positive and increasing. We also define, for each $i \in [\![N]\!]$,

$$W_i \coloneqq p \cdot \sum_{j=1}^{i} \mathbb{E}\big[\phi(\lambda \hat{L}_j) \cdot \hat{L}_j^2 \mid \mathcal{F}_{j-1}\big].$$

We first show that the upper-tail probability of M_N can be bounded from above by the sum of a Gaussian-like tail bound and the probability that W_N exceeds $\sigma^2/2$ by a macroscopic amount. Our proof is an adaptation of the argument used by Freedman [15] to prove a variance-dependent version of the Azuma–Hoeffding inequality; in contrast to [15], we do not assume an almost-sure bound on W_N.

Claim 2.6. *For any $\varepsilon > 0$,*

$$\mathbb{P}(M_N - M_0 \geqslant t) \leqslant \exp\left(-\frac{(1-\varepsilon)t^2}{2\sigma^2}\right) + \mathbb{P}\left(W_N > \frac{(1+\varepsilon)\sigma^2}{2}\right).$$

Proof. We will show that the sequence $Z_0, \dots, Z_N$, defined by

$$Z_i \coloneqq \exp\big(\lambda(M_i - M_0) - \lambda^2 W_i\big)$$

is a supermartingale. This fact will imply the assertion of the claim. Indeed, for every $w \geqslant 0$,

$$\begin{aligned}\mathbb{P}(M_N - M_0 \geqslant t) &= \mathbb{P}\left(Z_N \geqslant e^{\lambda t - \lambda^2 W_N}\right) \leqslant \mathbb{P}\left(Z_N \geqslant e^{\lambda t - \lambda^2 w}\right) + \mathbb{P}\left(W_N > w\right) \\ &\leqslant \mathbb{E}[Z_N] \cdot e^{\lambda^2 w - \lambda t} + \mathbb{P}\left(W_N > w\right),\end{aligned}$$

using Markov's inequality. If Z_i is in fact a supermartingale, then $\mathbb{E}[Z_N] \leqslant \mathbb{E}[Z_0] = 1$, and the assertion of the claim follows by letting $w \coloneqq (1+\varepsilon)\sigma^2/2$ in the above inequality (recall that $\lambda = t/\sigma^2$).

Since $e^{\lambda x} = 1 + \lambda x + \lambda^2 x^2 \cdot \phi(\lambda x)$, the definition of M_i yields

$$\begin{aligned}\mathbb{E}\left[\exp\big(\lambda(M_i - M_{i-1})\big) \mid \mathcal{F}_{i-1}\right] &= 1 + \lambda^2 \cdot \mathbb{E}\left[\phi\big(\lambda(M_i - M_{i-1})\big) \cdot (M_i - M_{i-1})^2 \mid \mathcal{F}_{i-1}\right] \\ &\leqslant 1 + \lambda^2 \cdot \mathbb{E}\left[\phi\left(\lambda \mathbb{E}\big[\hat{L}_i \mid \mathcal{F}_{i-1}\big]\right) \cdot (Y_i - p)^2 \cdot \mathbb{E}\big[\hat{L}_i \mid \mathcal{F}_{i-1}\big]^2 \mid \mathcal{F}_{i-1}\right] \\ &= 1 + \lambda^2 \cdot p(1-p) \cdot \phi\left(\lambda \mathbb{E}\big[\hat{L}_i \mid \mathcal{F}_{i-1}\big]\right) \cdot \mathbb{E}\big[\hat{L}_i \mid \mathcal{F}_{i-1}\big]^2,\end{aligned}$$

where the inequality holds as ϕ is increasing, $Y_i - p \leqslant 1$, and $\lambda \mathbb{E}\big[\hat{L}_i \mid \mathcal{F}_{i-1}\big] \geqslant 0$. Applying Jensen's inequality to the convex function $x \mapsto \phi(\lambda x) \cdot x^2 = \lambda^{-2} \cdot (e^{\lambda x} - \lambda x - 1)$ further gives

$$\begin{aligned}\mathbb{E}\left[\exp\big(\lambda(M_i - M_{i-1})\big) \mid \mathcal{F}_{i-1}\right] &\leqslant 1 + \lambda^2 \cdot p(1-p) \cdot \mathbb{E}\left[\phi\big(\lambda \hat{L}_i\big) \cdot \hat{L}_i^2 \mid \mathcal{F}_{i-1}\right] \\ &\leqslant \exp\left(\lambda^2 p \cdot \mathbb{E}\left[\phi\big(\lambda \hat{L}_i\big) \cdot \hat{L}_i^2 \mid \mathcal{F}_{i-1}\right]\right) \\ &= \exp\left(\lambda^2 \big(W_i - W_{i-1}\big)\right).\end{aligned}$$

Rearranging the above inequality gives $\mathbb{E}\left[Z_i \mid \mathcal{F}_{i-1}\right] \leqslant Z_{i-1}$, as claimed, which completes the proof. □

While the definition of W_i is convenient in the proof Claim 2.6, the sequentially conditioned random variables appearing in this definition make it difficult to work with this variable directly. Luckily, we may replace the upper-tail probability of W_N by a more facile upper-tail probability while incurring only a polynomial loss. Define

$$H_N \coloneqq p \cdot \sum_{i=1}^{N} \phi(\lambda \hat{L}_i) \cdot \hat{L}_i^2.$$

Claim 2.7. *For any $w \geqslant 0$, we have $\mathbb{E}\big[H_N \mid W_N > w\big] > w$.*

Proof. We begin by noting that $p \cdot \phi(\lambda \hat{L}_i) \cdot \hat{L}_i^2$ is an increasing function of $(Y_1, \dots, Y_N)$. Let $G_w \coloneqq \{W_N > w\}$ and let $\mathcal{G}_w$ be the σ-algebra generated by G_w. Harris's inequality [19] implies that, on G_w,

$$p \cdot \mathbb{E}\left[\phi\big(\lambda \hat{L}_i\big) \cdot \hat{L}_i^2 \mid \mathcal{F}_{i-1}, \mathcal{G}_w\right] \geqslant p \cdot \mathbb{E}\left[\phi\big(\lambda \hat{L}_i\big) \cdot \hat{L}_i^2 \mid \mathcal{F}_{i-1}\right].$$

In particular, we deduce that

$$\begin{aligned}\mathbb{E}[\mathbb{1}_{G_w} \cdot H_N] &= \mathbb{E}\left[\mathbb{1}_{G_w} \cdot p \cdot \sum_{i=1}^{N} \mathbb{E}\left[\phi\big(\lambda \hat{L}_i\big) \cdot \hat{L}_i^2 \mid \mathcal{F}_{i-1}, \mathcal{G}_w\right]\right] \\ &\geqslant \mathbb{E}\left[\mathbb{1}_{G_w} \cdot p \cdot \sum_{j=1}^{N} \mathbb{E}\left[\phi\big(\lambda \hat{L}_i\big) \cdot \hat{L}_i^2 \mid \mathcal{F}_{i-1}\right]\right] = \mathbb{E}\left[\mathbb{1}_{G_w} \cdot W_N\right] > w \cdot \mathbb{P}(G_w).\end{aligned}$$

Dividing through by the probability of G_w completes the proof. □

We now note for future reference that, for every $i \in [\![N]\!]$, since $\hat{L}_i \leqslant \log(1/p)/(2\lambda)$ by construction,

$$\phi\big(\lambda \hat{L}_i\big) \cdot \hat{L}_i^2 \leqslant \frac{e^{\lambda \hat{L}_i}}{\lambda^2} \leqslant \frac{1}{p^{1/2}\lambda^2}. \tag{15}$$

Claim 2.8. *For all $\varepsilon > 0$ and $t \geqslant \varepsilon\sigma$,*

$$\mathbb{P}\left(W_N > \frac{(1+\varepsilon)\sigma^2}{2}\right) < \frac{8N}{\varepsilon^3} \cdot \mathbb{P}\left(H_N > \frac{(1+3\varepsilon/4)\sigma^2}{2}\right).$$

Proof. Note that, by (15), we have $H_N \leqslant N/\lambda^2$ almost surely. In particular, this implies that

$$\mathbb{E}\left[H_N \mid W_N > \frac{(1+\varepsilon)\sigma^2}{2}\right] \leqslant \frac{N}{\lambda^2} \cdot \mathbb{P}\left(H_N > \frac{(1+3\varepsilon/4)\sigma^2}{2} \mid W_N > \frac{(1+\varepsilon)\sigma^2}{2}\right) + \frac{(1+3\varepsilon/4)\sigma^2}{2}.$$

On the other hand, by Claim 2.7,

$$\mathbb{E}\left[H_N \mid W_N > \frac{(1+\varepsilon)\sigma^2}{2}\right] > \frac{(1+\varepsilon)\sigma^2}{2}.$$

Combining these two inequalities, multiplying through by the probability that W_N exceeds $(1+\varepsilon)\sigma^2/2$, and recalling that $\lambda^2\sigma^2 = (t/\sigma)^2 \geqslant \varepsilon^2$ gives the assertion of the claim. □

Finally, we partition the upper tail of H_N. To this end, observe that when ε is sufficiently small, then for all i such that $\hat{L}_i \leqslant \varepsilon/\lambda$, we have $\phi(\lambda \hat{L}_i) \leqslant \phi(\varepsilon) \leqslant (1+\varepsilon/2)/2$. Using (15) for all remaining i, we obtain

$$H_N \leqslant \frac{(1+\varepsilon/2)}{2} \cdot p \cdot \sum_{i=1}^{N} \hat{L}_i^2 + \frac{p^{1/2}}{\lambda^2} \cdot \left|\left\{i : \hat{L}_i > \frac{\varepsilon}{\lambda}\right\}\right|.$$

Since $(1+\varepsilon/2)(1+\varepsilon/10) + 2\varepsilon/20 \leqslant 1 + 3\varepsilon/4$ for all sufficiently small $\varepsilon > 0$, we may conclude that

$$\mathbb{P}\left(H_N > \frac{(1+3\varepsilon/4)\sigma^2}{2}\right) \leqslant \mathbb{P}\left(\sum_{i=1}^{N} \hat{L}_i^2 \geqslant \left(1 + \frac{\varepsilon}{10}\right) \cdot \frac{\sigma^2}{p}\right) + \mathbb{P}\left(\left|\left\{i : \hat{L}_i > \frac{\varepsilon}{\lambda}\right\}\right| \geqslant \frac{\varepsilon\lambda^2\sigma^2}{20p^{1/2}}\right).$$

Combining (14), Claims 2.6 and 2.8, and the above estimate for the upper tail of H_N yields

$$\begin{aligned}\mathbb{P}(X \geqslant \mu + t) \leqslant \exp\left(-\frac{(1-\varepsilon)t^2}{2\sigma^2}\right) &+ \frac{8N}{\varepsilon^3} \cdot \mathbb{P}\left(\sum_{i=1}^{N} \hat{L}_i^2 > \left(1 + \frac{\varepsilon}{10}\right) \cdot \frac{\sigma^2}{p}\right) \\ &+ \frac{8N}{\varepsilon^3} \cdot \mathbb{P}\left(\left|\left\{i : \hat{L}_i > \frac{\varepsilon}{\lambda}\right\}\right| \geqslant \frac{\varepsilon\lambda^2\sigma^2}{20p^{1/2}}\right) + \mathbb{P}\left(\exists i \;\; L_i > \frac{\log(1/p)}{2\lambda}\right).\end{aligned}$$

Finally, since $\hat{L}_i \leqslant L_i$, we may replace the truncated variables in both probabilities above with the untruncated versions, thereby only increasing the right-hand side. □

3. The probability of small seeds: Proof of Theorem 1.2

In this section, we prove the main technical result of this paper, Theorem 1.2. It will be more convenient to state and prove an equivalent version of this result, where the function Ψ is replaced by its combinatorial analogue Ψ^*, which we now define. In order to do so, we first define, for all $U \subseteq [\![N]\!]$,

$$A_k(U) \coloneqq |\{B \in \mathrm{AP}_k : B \subseteq U\}|.$$

With this, for all $t \geqslant 0$, let

$$\Psi^*(t) = \Psi^*_{N,p,k}(t) \coloneqq \min\{|U| : U \subseteq [\![N]\!] \text{ and } A_k(U) \geqslant t\}, \tag{16}$$

cf. (2). As before we set $\Psi^*(t) = \infty$ when $t > |\mathrm{AP}_k|$. Since every k-AP contained in a set $U \subseteq [\![N]\!]$ contributes $1 - p^k$ to the difference $\mathbb{E}_U[X] - \mathbb{E}[X]$, we have $\Psi(t) \leqslant \Psi^*\big(t/(1-p^k)\big)$ for all $t \geqslant 0$. Furthermore, a straightforward computation shows that $A_k([\![m]\!]) = (1+o(1)) \cdot \frac{1}{k-1}\binom{m}{2}$ as $m \to \infty$, which implies that $\Psi^*(t) \leqslant (1+o(1)) \cdot \sqrt{2(k-1)t}$ whenever $1 \ll t \leqslant |\mathrm{AP}_k|$. Finally, we will show that $(1-o(1)) \cdot \sqrt{2(k-1)t}$ is a lower bound on $\Psi(t)$. Together, these facts will establish the following proposition.

Proposition 3.1. *Let $k \geqslant 3$ and assume that $\max\{1, N^2p^{2k-2}\} \ll t \leqslant |\mathrm{AP}_k| - \mu$ and $p \ll 1$. Then*

$$\Psi(t) = (1+o(1)) \cdot \Psi^*(t) = (1+o(1)) \cdot \sqrt{2(k-1)t}.$$

We remark that a version of Proposition 3.1 was proved by Bhattacharya–Ganguly–Shao–Zhao [9]. However, their version [9, Theorem 2.2] requires a stronger lower-bound assumption on t, which is in fact necessary for the continuous relaxation of Ψ which they consider (and which does not always coincide with the combinatorial notion of Ψ used in this work). Even though our proof of Proposition 3.1 essentially repeats the argument of [18, Proposition 4.3], we include it here for completeness.

We now state the aforementioned version of Theorem 1.2, with Ψ replaced by Ψ^*.

Proposition 3.2. *For every positive ε and every integer $k \geqslant 3$, there is some C such that the following holds. Let $N \in \mathbb{N}$ and $p \in (0, 1/2)$, and define $\mathcal{S}_{\mathrm{small}}(t, C)$ to be the set of all t-seeds $U \subseteq [\![N]\!]$ such that*

$$t \geqslant C|U| \cdot \max\{1, Np^{k-1}\} \qquad \text{and} \qquad t \geqslant C|U|^2 p^{k-2} \cdot N^{(k-2)(|U|/t)^{1/(k-1)}}. \tag{17}$$

Then, for every $t \geqslant 0$,

$$\log \mathbb{P}\big(U \subseteq \mathbf{R} \text{ for some } U \in \mathcal{S}_{\mathrm{small}}(t, C)\big) \leqslant (1-\varepsilon) \cdot \Psi^*\big((1-\varepsilon)t\big) \cdot \log p.$$

The remainder of this section is organised as follows. In Section 3.1, we prove Proposition 3.1 and present the short derivation of Theorem 1.2 from Proposition 3.2. The remaining two subsections are devoted to the proof of Proposition 3.2. The short Section 3.2 presents three auxiliary, technical results needed for the proof, which is presented in the much more substantial Section 3.3.

3.1. **Proof of Proposition 3.1 and derivation of Theorem 1.2.** Given a set $U \subseteq [\![N]\!]$ and an integer $k \geqslant 3$, it will be convenient to denote, for every $r \in [\![k]\!]$, the number of k-APs that intersect U in exactly r elements by $A_r^{(k)}(U)$; note that then $A_k(U) = A_k^{(k)}(U)$. If X denotes the number of k-APs in the p-biased random subset $\mathbf{R} \subseteq [\![N]\!]$, linearity of expectation allows us to write

$$\mathbb{E}_U[X] - \mathbb{E}[X] = \sum_{r=1}^{k} A_r^{(k)}(U) \cdot (p^{k-r} - p^k). \tag{18}$$

Since any two numbers lie in at most $\binom{k}{2}$ distinct k-APs (equivalently, the hypergraph AP_k of k-APs in $[\![N]\!]$ satisfies $\Delta_2(\mathrm{AP}_k) \leqslant \binom{k}{2}$), we may bound

$$A_1^{(k)}(U) \leqslant \binom{k}{2} N|U| \qquad \text{and} \qquad A_2^{(k)}(U) + \cdots + A_k^{(k)}(U) \leqslant \binom{k}{2}\binom{|U|}{2}. \tag{19}$$

Finally, the proof of Proposition 3.1 relies on the following combinatorial result that appears as [9, Theorem 2.4]. (We note that [9] considers a slightly different variational problem, since that work counts progressions with positive and negative common difference separately; this leads to a difference of $\sqrt{2}$ between the result quoted below and the one that appears in [9].)

Lemma 3.3. *For every $U \subseteq [\![N]\!]$ with m elements, $A_k(U) \leqslant A_k([\![m]\!]) = (1+o(1))\frac{m^2}{2(k-1)}$. In particular, if $1 \ll t \leqslant |\mathrm{AP}_k|$, then $\Psi^*(t) = (1+o(1))\sqrt{2(k-1)t}$.*

Proof of Proposition 3.1. We have already mentioned the bound $\Psi(t) \leqslant \Psi^*\big(t/(1-p^k)\big)$. The assumption $t \leqslant |\mathrm{AP}_k| - \mu = |\mathrm{AP}_k|(1-p^k)$ implies that $t/(1-p^k) \leqslant |\mathrm{AP}_k|$, so Lemma 3.3 gives

$$\Psi(t) \leqslant \Psi^*\big(t/(1-p^k)\big) = (1+o(1))\sqrt{2(k-1)t} = (1+o(1))\Psi^*(t),$$

where we used $p \ll 1$. In view of this, it remains to show that $\Psi(t) \geqslant (1-o(1)) \cdot \sqrt{2(k-1)t}$ as long as $t \gg \max\{1, N^2p^{2k-2}\}$. Fix $\varepsilon > 0$ and consider an arbitrary set $U \subseteq [\![N]\!]$ with $|U| \leqslant (1-\varepsilon)\sqrt{2(k-1)t}$. Using (18) and (19), we find that

$$\begin{aligned}\mathbb{E}_U[X] - \mathbb{E}[X] &\leqslant A_k(U) + \sum_{r=1}^{k-1} A_r^{(k)}(U) \cdot p^{k-r} \\ &\leqslant A_k(U) + p \cdot \binom{k}{2}\binom{|U|}{2} + \binom{k}{2} N|U|p^{k-1}.\end{aligned}$$

The final two terms on the right-hand side are $o(t)$; this follows from the upper bound on $|U|$ and the assumption $p \ll 1$ (for the second term) or the assumption $t \gg N^2p^{2k-2}$ (for the third term). Furthermore, Lemma 3.3 implies that $A_k(U) \leqslant (1-\varepsilon)t$. Thus, $\mathbb{E}_U[X] - \mathbb{E}[X] < t$ for every set U with at most $(1-\varepsilon)\sqrt{2(k-1)t}$ elements, as desired. □

Derivation of Theorem 1.2 from Proposition 3.2. Note that every t-seed with at most m elements, where t and m satisfy (4), belongs to $\mathcal{S}_{\mathrm{small}}(t, C)$ for every fixed $C > 0$ and all large enough N. It thus suffices to show that the existence of such a t-seed U implies that the two assumptions of Proposition 3.1 hold, so that we may replace $(1-\varepsilon)\Psi^*\big((1-\varepsilon)t\big)$ with $(1-o(1)) \cdot \Psi(t)$. To this end, suppose that $U \subseteq [\![N]\!]$ is a t-seed with at most m elements. Then, firstly, we know that $\mu + t \leqslant |\mathrm{AP}_k|$. Secondly, by (18) and (19),

$$t \leqslant \mathbb{E}_U[X] - \mathbb{E}[X] \leqslant \sum_{r=1}^{k} A_r^{(k)}(U) \cdot p^{k-r} \leqslant \binom{k}{2} Nm \cdot p^{k-1} + \binom{k}{2}\binom{m}{2}.$$

In particular, this means that $t \leqslant Km\left(Np^{k-1}+m\right)$ for some constant K that depends only on k. Since we have assumed that $t \gg mNp^{k-1}$, we conclude that $t \leqslant 2Km^2$ and thus $t \geqslant (t/m)^2/(2K) \gg N^2p^{2k-2}$. Finally, thanks to the assumption $t \gg m^2p^{k-2}N^{(k-2)(m/t)^{1/(k-1)}} \geqslant m^2p^{k-2}$, we conclude that $p \ll 1$. □

3.2. **Preliminaries for the proof of Proposition 3.2.** If U is a finite set and $f\colon \mathcal{P}(U) \to \mathbb{R}$ is a function on its power set, then the partial derivative of f with respect to $u \in U$ is the function $\partial_u f\colon \mathcal{P}(U) \to \mathbb{R}$ defined by $\partial_u f(U') \coloneqq f(U' \cup \{u\}) - f(U' \setminus \{u\})$ for every $U' \subseteq U$. The following simple lemma, which generalises [18, Lemma 3.8], is a key ingredient in the proof of Lemma 3.7 below. For the readers that are familiar with the general framework of [18], we remark that this lemma will allow us to extract a core from every seed.

Lemma 3.4. *If U is a finite set and $f\colon \mathcal{P}(U) \to \mathbb{R}$ is a function, then, for every $w \in \mathbb{R}^{|U|}$, there exists a subset $U^* \subseteq U$ such that*

(i) $f(U^*) \geqslant f(U) - \|w\|_1$ and

(ii) $\partial_u f(U^*) \geqslant w_{|U^*|}$ for all $u \in U^*$.

Proof. Let $U = U_0 \supsetneq U_1 \supsetneq \cdots \supsetneq U_k = U^*$ be a chain of maximal length such that $f(U_{i-1}) - f(U_i) < w_{|U_{i-1}|}$ for all $1 \leqslant i \leqslant k$. Then U^* satisfies (ii), since otherwise we could obtain a longer chain by setting $U_{k+1} = U_k \setminus \{u\}$ for an element $u \in U_k$ with $\partial_u f(U_k) < w_{|U_k|}$. To see that U^* also satisfies (i), note that

$$f(U) - f(U^*) = \sum_{i=1}^{k} \big(f(U_{i-1}) - f(U_i)\big) \leqslant \sum_{i=1}^{k} w_{|U_{i-1}|} \leqslant \|w\|_1,$$

because the cardinalities $|U_0|, \dots, |U_{k-1}|$ are distinct positive integers. □

The second ingredient is a version of Janson's inequality [20] for the hypergeometric distribution; it can be derived from the standard version of Janson's inequality and the fact that the mean of binomial distribution is also its median, provided that it is an integer (cf. the proof of [4, Lemma 3.1]).

Lemma 3.5. *Suppose that $(B_\alpha)_{\alpha\in A}$ is a family of subsets of a t-element set Ω. Let $s \in \{0, \dots, t\}$ and let*

$$\mu \coloneqq \sum_{\alpha\in A} \left(\frac{s}{t}\right)^{|B_\alpha|} \qquad \text{and} \qquad \Delta \coloneqq \sum_{\alpha\sim\beta} \left(\frac{s}{t}\right)^{|B_\alpha \cup B_\beta|},$$

where the second sum is over all ordered pairs $(\alpha, \beta) \in A^2$ such that $\alpha \neq \beta$ and $B_\alpha \cap B_\beta \neq \varnothing$. Let S be the uniformly chosen random s-element subset of Ω and let Z denote the number of $\alpha \in A$ such that $B_\alpha \subseteq S$. Then, for every $\varepsilon \in (0, 1]$,

$$\mathbb{P}\big(Z \leqslant (1-\varepsilon)\mu\big) \leqslant 2\exp\left(-\frac{\varepsilon^2}{2} \cdot \frac{\mu^2}{\mu+\Delta}\right).$$

In the proof of Proposition 3.2, we will encounter the function $\beta\colon (0, 1] \to \mathbb{R}_{>0}$ defined by

$$\beta(x) \coloneqq \frac{1}{(2-\log x)^2}. \tag{20}$$

The precise details of this definition have no deeper meaning; what we require is essentially a function on $(0, 1]$ that approaches 0 sufficiently slowly as $x \to 0$ while still having the property that $\int_0^1 x^{-1}\beta(x)\,\mathrm{d}x$ exists. Some relevant properties of β are collected in the following fact.

Fact 3.6. *The following statements hold:*

(i) β *is increasing;*

(ii) for $c > 0$*, the function* $x \mapsto x^{-c}\beta(x)$ *is decreasing on* $(0, e^{2-2/c}]$ *and increasing on* $[e^{2-2/c}, e^2)$*;*

(iii) $\int_0^1 x^{-1}\beta(x)\,\mathrm{d}x = 1/2$.

Proof. The first item is obvious. A direct computation shows that, for every $c > 0$,

$$\big(x^{-c}\beta(x)\big)' = \frac{c\log x - 2c + 2}{x^{c+1}\cdot(2-\log x)^3},$$

which is negative for $x \in (0, e^{2-2/c})$. For the last item, we have

$$\int_0^1 \frac{1}{x\cdot(2-\log x)^2}\,\mathrm{d}x = \left[\frac{1}{2-\log x}\right]_0^1 = 1/2. \qquad \square$$

3.3. **Proof of Proposition 3.2.** As earlier, we write $A_r^{(k)}(U)$ for the number of k-term arithmetic progressions in $[\![N]\!]$ that intersect U at precisely r elements. Throughout the proof, we will suppress the dependence of this quantity on k for notational convenience.

Definition. A set $U^* \subseteq [\![N]\!]$ is called a (t,ε,ξ)-*core*, for some $t, \varepsilon, \xi > 0$, if

(C1) $|U^*| \geqslant \Psi^*\big((1-\varepsilon)t\big)$ and

(C2) for some $r \in \{3,\dots,k\}$ and all $u \in U^*$,

$$\partial_u A_r(U^*) \geqslant \frac{\xi}{|U^*|}\cdot\max\left\{t, \left(\frac{t}{|U^*|^2}\right)^{\frac{1}{k-2}}\cdot\max_{K\subseteq U^*} A_{r-1}(K)\right\}.$$

We note that every interval in $[\![N]\!]$ of length $\lfloor\sqrt{2(k-1)t}\rfloor$ is a (t,ε,ξ)-core provided that ξ is smaller than some $\xi_0 = \xi_0(k) > 0$. Indeed, (C1) follows from Proposition 3.1 and (C2) (for $r = k$) is a straigthforward calculation (the left-hand side is linear in $\sqrt{t}$ whereas the right-hand side is linear in $\xi\sqrt{t}$).

Our first lemma states that every seed contains a core. This lemma, together with the definition of a core, lie at the the very heart of the proof of Proposition 3.2.

Lemma 3.7. *For every* $\varepsilon \in (0,1)$*, there exist positive* $\delta = \delta(\varepsilon,k)$ *and* $C = C(\varepsilon,k)$ *such that the following holds. Let* U *be a* t*-seed with* m *elements, where* $t \geqslant C\cdot\max\{m^2p^{k-2}, mNp^{k-1}, 1\}$*. Then* U *contains a subset* U^* *that is a* $\big(t,\varepsilon,\delta\beta(|U^*|/m)\big)$*-core, where* β *is defined as in* (20)*.*

Note that $\delta\beta(|U^*|/m) = \delta/\big(2-\log(|U^*|/m)\big)^2$ is not quite a constant – things would be a little simpler if it were – but it is good enough for the purposes of our next lemma, which bounds the number of cores of a given size.

Lemma 3.8. *For all* $\varepsilon, \delta, \eta \in (0,1/2)$*, there is a positive* $C = C(\delta,\varepsilon,\eta,k)$ *such that the following holds. Let* t *and* m *be positive integers with* $t \geqslant C\cdot\max\{m, mNp^{k-1}, m^2p^{k-2}N^{C(m/t)^{1/(k-1)}}\}$*. Denote by* $\mathcal{C}(s)$ *the set of* $\big(t,\varepsilon,\delta\beta(s/m)\big)$*-cores of size* s*. Then*

$$|\mathcal{C}(s)| \leqslant \begin{cases} (\eta/p)^s & \text{if } s \leqslant m, \\ (1/p)^{\eta s} & \text{if } s \leqslant \min\{m, \sqrt{4kt}\log(1/p)\}. \end{cases}$$

Before proving the two lemmas, let us show how they imply the statement of Proposition 3.2. Fix some $\varepsilon \in (0,1/2)$ and let $\delta = \delta(\varepsilon,k)$ be as in Lemma 3.7; we may clearly assume that $\delta \leqslant \xi_0$, where ξ_0 is the constant introduced below the definition of a core. Let $\eta \coloneqq \varepsilon/2 \leqslant 1/e$ and let m be such that $t \geqslant C \cdot \max\{m, mNp^{k-1}, m^2p^{k-2}N^{C(m/t)^{1/(k-1)}}\}$ for a sufficiently large $C = C(\delta,\varepsilon,\eta,k)$. Denote by $\mathcal{C}(s)$ the set of $\big(t,\varepsilon,\delta\beta(s/m)\big)$-cores of size s, as in the statement of Lemma 3.8, and let $s_0 \geqslant \Psi^*\big((1-\varepsilon)t\big)$ be the minimal value of s for which $\mathcal{C}(s)$ is nonempty. It is enough to show that

$$\mathbb{P}\big(U \subseteq \mathbf{R} \text{ for some } t\text{-seed } U \text{ with } |U| \leqslant m\big) \leqslant p^{(1-\varepsilon)s_0}.$$

By Lemma 3.7 and the union bound, the left-hand side of the above inequality is at most

$$\mathbb{P}\big(U^* \subseteq \mathbf{R} \text{ for some } (t,\varepsilon,\delta\beta(|U^*|/m))\text{-core } U^* \text{ with } |U^*| \leqslant m\big) \leqslant \sum_{s=s_0}^{m} |\mathcal{C}(s)| \cdot p^s.$$

Further, by Lemma 3.8,

$$\sum_{s=s_0}^{m} |\mathcal{C}(s)| \cdot p^s \leqslant \sum_{s=s_0}^{\sqrt{4kt}\log(1/p)} p^{(1-\eta)s} + \sum_{s=\sqrt{4kt}\log(1/p)}^{m} \eta^s \leqslant \frac{p^{(1-\varepsilon/2)s_0}}{1-p^{1-\varepsilon/2}} + \frac{p^{\sqrt{4kt}}}{1-1/e}.$$

As mentioned above, any interval of length $\lfloor\sqrt{2(k-1)t}\rfloor$ is a (t,ε,ξ)-core, whenever $\xi \leqslant \xi_0$. In particular, since $\delta\beta(s/m) \leqslant \delta\beta(1) \leqslant \xi_0$ for all $s \leqslant m$, we have $s_0 \leqslant \sqrt{2(k-1)t}$. Consequently, the right-hand side above is at most $4p^{(1-\varepsilon/2)s_0} \leqslant p^{(1-\varepsilon)s_0}$, as $p \leqslant 1/2$ and $s_0 \geqslant \Psi^*\big((1-\varepsilon)t\big) \geqslant \sqrt{t} \geqslant \sqrt{C}$, by Proposition 3.1.

The remaining part of this section is dedicated to proving Lemmas 3.7 and 3.8.

Proof of Lemma 3.7. Let $\eta = \eta(\varepsilon,k) > 0$ be sufficiently small and let $C = C(\varepsilon,\eta,k) > 0$ be sufficiently large. Define, for every $r \in \{2,\dots,k\}$, the function $a_r\colon \mathbb{R} \to \mathbb{R}$ given by

$$a_r(x) \coloneqq (1-\eta)\left(\frac{x^2}{\eta t}\right)^{\frac{k-r}{k-2}}.$$

We say that a subset $U' \subseteq U$ is *r-dense* if $A_r(U') \geqslant a_r(|U'|)\cdot t$. Observe that, as long as $\eta < 4/(k^2+4)$, no set $U' \subseteq U$ can be 2-dense. Indeed, by (19),

$$A_2(U') \leqslant \binom{k}{2}\binom{|U'|}{2} \leqslant \frac{k^2|U'|^2}{4} < \frac{(1-\eta)|U'|^2}{\eta} = a_2(|U'|)\cdot t.$$

Claim 3.9. *The t-seed U is r-dense for some $r \geqslant 3$.*

Proof. By the definition of a t-seed and (18), for every sequence $\lambda \in \mathbb{R}^k$ with $\lambda_1 + \dots + \lambda_k \leqslant 1$,

$$\sum_{r=1}^{k} \lambda_r t \leqslant t \leqslant \mathbb{E}_U[X] - \mathbb{E}[X] \leqslant \sum_{r=1}^{k} A_r(U)\cdot p^{k-r},$$

which implies that there is some $r \in [\![k]\!]$ such that $A_r(U)\cdot p^{k-r} \geqslant \lambda_r t$. With this in mind, we define λ_r as follows:

$$\lambda_r \coloneqq \begin{cases} (1-\eta)\eta^{k-1} & \text{if } r = 1, \\ a_r(|U|)\cdot p^{k-r} & \text{if } r \geqslant 2. \end{cases}$$

Our assumption $t \geqslant C|U|^2p^{k-2} \geqslant \eta^{1-k}|U|^2p^{k-2}$ (which holds for all large enough C) implies that $\lambda_r \leqslant (1-\eta)\eta^{k-r}$ for all $r \in [\![k]\!]$, so indeed

$$\lambda_1 + \dots + \lambda_k \leqslant (1-\eta)\sum_{r=1}^{k}\eta^{k-r} < (1-\eta)\sum_{r=0}^{\infty}\eta^r = 1.$$

Consequently, $A_r(U) \cdot p^{k-r} \geqslant \lambda_r t$ for some $r \in [\![k]\!]$. Note that we may rule out the case $r = 1$, thanks to the bound $A_1(U) \leqslant \binom{k}{2}|U|N$, see (19), and the assumption that $t \geqslant C|U|Np^{k-1}$ for a sufficiently large $C = C(\varepsilon, \eta, k)$. We may therefore conclude that $A_r(U) \geqslant a_r(|U|) \cdot t$, i.e., that U is r-dense, for some $r \in \{2, \dots, k\}$; since we have shown above that no subset of U can be 2-dense, we must have $r \geqslant 3$. □

The claim allows us to define $\tilde{U} \subseteq U$ as a smallest nonempty subset of U that is r-dense for some $r \geqslant 3$. In a slight abuse of notation, let r be the smallest index such that $\tilde{U}$ is r-dense. Then the minimality of $\tilde{U}$ and r implies that no subset of $\tilde{U}$ is $(r-1)$-dense.

Now let $w \in \mathbb{R}^{|\tilde{U}|}$ be defined by $w_\ell \coloneqq \eta A_r(\tilde{U}) \cdot \ell^{-1}\beta(\ell/|\tilde{U}|)$. Since $x^{-1}\beta(x)$ is decreasing on $(0,1]$ and $\int_0^1 x^{-1}\beta(x)\,\mathrm{d}x \leqslant 1$, we have

$$\|w\|_1 = \eta A_r(\tilde{U}) \cdot \frac{1}{|\tilde{U}|}\sum_{\ell=1}^{|\tilde{U}|}\frac{\beta(\ell/|\tilde{U}|)}{\ell/|\tilde{U}|} \leqslant \eta A_r(\tilde{U})\int_0^1 x^{-1}\beta(x)\,\mathrm{d}x \leqslant \eta A_r(\tilde{U}).$$

Therefore, we may apply Lemma 3.4 to obtain a subset $U^* \subseteq \tilde{U}$ such that

$$A_r(U^*) \geqslant A_r(\tilde{U}) - \|w\|_1 \geqslant (1-\eta)\cdot A_r(\tilde{U}) \geqslant (1-\eta)\cdot a_r(|\tilde{U}|)\cdot t \geqslant (1-\eta)\cdot a_r(|U^*|)\cdot t, \tag{21}$$

and, for every $u \in U^*$,

$$\partial_u A_r(U^*) \geqslant w_{|U^*|} = \eta A_r(|\tilde{U}|)\cdot\frac{\beta(|U^*|/|\tilde{U}|)}{|U^*|} \geqslant \frac{\eta\cdot\beta(|U^*|/|U|)}{|U^*|}\cdot a_r(|U^*|)\cdot t, \tag{22}$$

where the last inequality uses $A_r(\tilde{U}) \geqslant a_r(|\tilde{U}|)\cdot t \geqslant a_r(|U^*|)\cdot t$ and the fact that β is increasing.

Claim 3.10. *The set U^* is a $\big(t, \varepsilon, \eta^2(1-\eta)^2\beta(|U^*|/|U|)\big)$-core.*

Proof. We first prove that

$$A_k(U^*) \geqslant (1-2\eta)t \qquad \text{or} \qquad |U^*| \geqslant \sqrt{4kt}, \tag{23}$$

which implies (C1), in the first case by the definition of Ψ^* (provided that $\eta < \varepsilon/2$) and in the second case by Proposition 3.1. Suppose that $r \in \{3, \dots, k\}$ attains (21). First, if $r = k$, then (21) immediately gives

$$A_k(U^*) \geqslant (1-\eta)\cdot a_k(|U^*|)\cdot t = (1-\eta)^2 t \geqslant (1-2\eta)t,$$

as desired. Otherwise, if $3 \leqslant r < k$, then (21) gives

$$A_r(U^*) \geqslant (1-\eta)\cdot a_r(|U^*|)\cdot t = (1-\eta)^2\left(\frac{|U^*|^2}{\eta t}\right)^{\frac{k-r}{k-2}} t$$

which we combine with the bound $A_r(U^*) \leqslant \binom{k}{2}\binom{|U^*|}{2} \leqslant k^2|U^*|^2$, see (19), to obtain

$$|U^*|^{2r-4}k^{2k-4} \geqslant (1-\eta)^{2k-4}\eta^{r-k}t^{r-2}.$$

Since the exponent of η is negative, the required inequality $|U^*| \geqslant \sqrt{4kt}$ follows for sufficiently small η.

Finally, we prove that (C2) holds with $\xi \coloneqq \eta^2\beta(|U^*|/|U|)$. By (22), it is enough to show that

$$a_r(|U^*|) \geqslant \eta \qquad \text{and} \qquad a_r(|U^*|)\cdot t \geqslant \eta\left(\frac{t}{|U^*|^2}\right)^{\frac{1}{k-2}} \cdot \max_{K\subseteq U^*} A_{r-1}(K).$$

First, since $|U^*|^2 \geqslant \Psi^*\big((1-\varepsilon)t\big)^2 \geqslant t$, by (C1) and Proposition 3.1, we have

$$a_r(|U^*|) = (1-\eta)\left(\frac{|U^*|^2}{\eta t}\right)^{\frac{k-r}{k-2}} \geqslant \eta.$$

Second, the minimality of $\tilde{U}$ and r implies that every $K \subseteq U^* \subseteq \tilde{U}$ is not $(r-1)$-dense and therefore

$$A_{r-1}(K) < a_{r-1}(|K|)\cdot t \leqslant a_{r-1}(|U^*|)\cdot t = \left(\frac{|U^*|^2}{\eta t}\right)^{\frac{1}{k-2}} \cdot a_r(|U^*|)\cdot t \leqslant \frac{1}{\eta}\left(\frac{|U^*|^2}{t}\right)^{\frac{1}{k-2}} \cdot a_r(|U^*|)\cdot t.$$

This completes the proof of the claim. □

The assertion of the lemma now follows with $\delta \coloneqq \eta^2(1-\eta)^2$. □

Proof of Lemma 3.8. Fix some $r \in [\![k]\!]$, $U \subseteq [\![N]\!]$, and $u \in [\![N]\!]$. It will be convenient to introduce a quantity that is closely related to the discrete derivative $\partial_u A_r(U)$ but has a slightly simpler combinatorial interpretation. Define $A_r(U,u)$ as the number of k-APs in $[\![N]\!]$ that intersect $U \cup \{u\}$ in exactly r and $U \setminus \{u\}$ in exactly $r-1$ elements. We will first show that, for every $u \in U$,

$$\partial_u A_r(U) = A_r(U,u) - A_{r+1}(U,u) \leqslant A_r(U,u) = A_r(U \setminus \{u\}, u). \tag{24}$$

In order to see this, we count the k-APs that intersect both U and $U \setminus \{u\}$ in exactly r elements. We can express their number in two ways: first, by $A_r(U) - A_r(U,u)$ and second, by $A_r(U \setminus \{u\}) - A_{r+1}(U,u)$. This implies the first equality in (24); the remainder of (24) is straightforward.

Since we can clearly assume that $\mathcal{C}(s) \neq \varnothing$, property (C1) and Proposition 3.1 imply that $s \geqslant \Psi^*\big((1-\varepsilon)t\big) \geqslant \sqrt{t}$. For the sake of brevity, we let

$$\xi \coloneqq \delta \cdot \beta(s/m). \tag{25}$$

Suppose that $U^* \in \mathcal{C}(s)$. By (C2) and (24), there is some $r \geqslant 3$ such that

$$s \cdot A_r(U^*,u) \geqslant s \cdot \partial_u A_r(U^*) \geqslant \xi \cdot \max\left\{t, \left(\frac{t}{s^2}\right)^{\frac{1}{k-2}} \cdot \max_{K \subseteq U^*} A_{r-1}(K)\right\} \quad \text{for all } u \in U^*. \tag{26}$$

For every $r \geqslant 3$, let $\mathcal{C}_r(s)$ be the subset of $\mathcal{C}(s)$ containing those U^* that satisfy (26). Since $|\mathcal{C}(s)| \leqslant |\mathcal{C}_3(s)| + \cdots + |\mathcal{C}_k(s)|$, it is enough to bound each $|\mathcal{C}_r(s)|$ individually.

For the remainder of the proof, fix some $r \geqslant 3$. We will say that an element $u \in [\![N]\!]$ is *rich* with respect to a subset $K \subseteq [\![N]\!] \setminus \{u\}$ if

$$s \cdot A_r(K,u) \geqslant \frac{\xi}{2} \cdot \left(\frac{|K|}{s}\right)^{r-1}\left(\frac{t}{s^2}\right)^{\frac{1}{k-2}} \cdot A_{r-1}(K).$$

Let $\mathcal{R}(K)$ denote the set of all elements of $[\![N]\!] \setminus K$ that are rich with respect to K and observe that

$$\begin{aligned}\frac{\xi|\mathcal{R}(K)|}{2} \cdot \left(\frac{|K|}{s}\right)^{r-1}\left(\frac{t}{s^2}\right)^{\frac{1}{k-2}} \cdot A_{r-1}(K) &\leqslant \sum_{u \in \mathcal{R}(K)} s \cdot A_r(K,u) \\ &\leqslant \sum_{u \in [\![N]\!] \setminus K} s \cdot A_r(K,u) = (k-r+1)\cdot s \cdot A_{r-1}(K).\end{aligned}$$

This implies that, whenever K is nonempty,

$$|\mathcal{R}(K)| \leqslant \frac{2ks}{\xi} \cdot \left(\frac{s}{|K|}\right)^{r-1}\left(\frac{s^2}{t}\right)^{\frac{1}{k-2}}. \tag{27}$$

It is also easy to see that $|R(\varnothing)| = N$.

Let us briefly explain how the notion of rich elements can help us bound the number of cores $U^* \in \mathcal{C}_r(s)$. If we order the elements of U^* at random as $u_1, \dots, u_s$, then, on average, a $\left(\frac{d-1}{s}\right)^{r-1}$-fraction of the k-APs counted by $A_r(U^*, u_d)$ is also counted by $A_r(\{u_1, \dots, u_{d-1}\}, u_d)$. In particular, (26) suggests that, in a typical random ordering, u_d will be rich with respect to $\{u_1, \dots, u_{d-1}\}$ for most indices d. On the other hand, due to (27), the fact that u_d is rich means that it comes from a somewhat small set that depends only on $\{u_1, \dots, u_{d-1}\}$. This translates into an upper bound on the number of ways to choose the whole set U^* element-by-element.

We now turn to the implementation of this idea. Given an arbitrary sequence $u_1, \dots, u_s$ of distinct elements of $[\![N]\!]$, define the set of *poor* indices

$$\mathcal{P}(u_1, \dots, u_s) \coloneqq \big\{d \in [\![s]\!] : u_d \text{ is } \textit{not} \text{ rich with respect to } \{u_1, \dots, u_{d-1}\}\big\}.$$

First, we show that, for an average ordering $u_1, \dots, u_s$ of the elements of a core U^*, the set of poor indices is small. Second, we give an upper bound on the total number of s-element sequences for which the poor indices belong to a given set.

Claim 3.11. *Let $U^* \in \mathcal{C}_r(s)$ and let $u_1, \dots, u_s$ be a uniformly chosen random ordering of the elements of U^*. Then*

$$\mathbb{E}\big[|\mathcal{P}(u_1, \dots, u_s)|\big] \leqslant 2s \left(\frac{9k^3 s}{\xi t}\right)^{1/(k-1)}.$$

Proof. For every integer $d \in [\![s]\!]$ and all $u \in U^*$, define

$$\mu_d(u) \coloneqq \binom{d-1}{s-1}^{r-1} \cdot A_r(U^*, u) \geqslant \frac{\xi}{s} \cdot \binom{d-1}{s}^{r-1} \cdot \max\left\{t, \left(\frac{t}{s^2}\right)^{\frac{1}{k-2}} \cdot \max_{K \subseteq U^*} A_{r-1}(K)\right\}, \tag{28}$$

where the inequality follows from (26). Note that $d \in \mathcal{P}(u_1, \dots, u_s)$ implies

$$A_r\big(\{u_1, \dots, u_{d-1}\}, u_d\big) \leqslant \mu_d(u_d)/2.$$

Note also that $1 \notin \mathcal{P}(u_1, \dots, u_s)$. Consequently,

$$\mathbb{E}\big[|\mathcal{P}(u_1, \dots, u_s)|\big] \leqslant \sum_{d=2}^{s} \mathbb{P}\big(A_r\big(\{u_1, \dots, u_{d-1}\}, u_d\big) \leqslant \mu_d(u_d)/2\big). \tag{29}$$

Fix some $d \in [\![s]\!]$ and note that, conditioned on u_d, the set $\{u_1, \dots, u_{d-1}\}$ is a uniformly random $(d-1)$-element subset of $U^* \setminus \{u_d\}$. Thus, we may use Janson's inequality (Lemma 3.5) to get an upper bound on the probabilities in the above sum. For $u \in U^*$, let $\mathcal{J}_u$ be the multiset defined by

$$\mathcal{J}_u \coloneqq \big\{B \cap (U^* \setminus \{u\}) : B \in \mathrm{AP}_k,\, u \in B, \text{ and } |B \cap U^*| = r\big\},$$

where the multiplicity of each element is equal to the number of k-APs B containing u giving rise to the same $(r-1)$-element set $B \cap (U^* \setminus \{u\})$. Observe that, for every $u \in U^*$, we have $|\mathcal{J}_u| = A_r(U^*, u)$ and $|\{J \in \mathcal{J}_u : J \subseteq K\}| \leqslant A_r(K, u)$ for all $K \subseteq U^* \setminus \{u\}$. Gearing towards an application of Lemma 3.5, note first that

$$\sum_{J \in \mathcal{J}_u} \binom{d-1}{s-1}^{|J|} = |\mathcal{J}_u| \cdot \binom{d-1}{s-1}^{r-1} = \mu_d(u).$$

Further, writing $J \sim J'$ to mean that $J \neq J'$ and $J \cap J' \neq \varnothing$ (which also includes the case where J and J' are the same $(r-1)$-element subset of two different k-APs),

$$\Delta_d(u) \coloneqq \sum_{\substack{J,J' \in \mathcal{J}_u \\ J \sim J'}} \left(\frac{d-1}{s-1}\right)^{|J \cup J'|} \leqslant \mu_d(u) \cdot k^3,$$

where the inequality holds because, for every $J \in \mathcal{J}_u$, there are at most k^3 progressions of length k that contain u_d and some element of J. Lemma 3.5 implies that

$$\begin{aligned}\mathbb{P}\left(A_r\big(\{u_1,\dots,u_{d-1}\},u_d\big) \leqslant \mu_d(u_d)/2 \mid u_d\right) &\leqslant 2\exp\left(-\frac{\mu_d(u_d)^2}{8\big(\mu_d(u_d)+\Delta_d(u_d)\big)}\right)\\ &\leqslant 2\exp\left(-\frac{\mu_d(u_d)}{9k^3}\right).\end{aligned}$$

Observe moreover that (28) implies $\mu_d(u_d) \geqslant \frac{\xi}{s} \cdot \left(\frac{d-1}{s}\right)^{k-1} \cdot t$, so taking expectations of the above expression allows us to conclude that

$$\mathbb{P}\left(A_r\big(\{u_1,\dots,u_{d-1}\},u_d\big) \leqslant \mu_d(u_d)/2\right) \leqslant 2 \cdot \mathbb{E}\left[\exp\left(-\frac{\mu_d(u_d)}{9k^3}\right)\right] \leqslant 2 \cdot \exp\left(-\frac{(d-1)^{k-1}\xi t}{9k^3 s^k}\right).$$

Substituting this inequality into (29) and using $h\sum_{d\geqslant 1} f(d\cdot h) \leqslant \int_0^\infty f(x)\,\mathrm{d}x$ for the decreasing function $f(x) = \exp(-x^{k-1})$ and $h = \left(\frac{\xi t}{9k^3 s^k}\right)^{1/(k-1)}$, we obtain

$$\mathbb{E}[|\mathcal{P}(u_1,\dots,u_s)|] \leqslant 2\sum_{d=1}^{s-1} \exp\left(-\frac{\xi t}{9k^3 s} \cdot \frac{d^{k-1}}{s^{k-1}}\right) \leqslant 2s\left(\frac{9k^3 s}{\xi t}\right)^{1/(k-1)} \int_0^\infty \exp\left(-x^{k-1}\right)\,\mathrm{d}x.$$

Finally, since the gamma function is convex on $\mathbb{R}_{>0}$ and $\Gamma(1) = \Gamma(2) = 1$, we have

$$\int_0^\infty \exp(-x^{k-1})\,\mathrm{d}x = \frac{1}{k-1}\int_0^\infty y^{\frac{1}{k-1}-1}e^{-y}\,\mathrm{d}y = \frac{1}{k-1}\Gamma\left(\frac{1}{k-1}\right) = \Gamma\left(\frac{k}{k-1}\right) \leqslant 1,$$

which completes the proof of the claim. □

To state our second claim, it will be convenient to define, for any $P \subseteq [\![s]\!]$,

$$\mathcal{X}_P \coloneqq \big\{(u_1,\dots,u_s) \in [\![N]\!]^{\underline{s}} : \mathcal{P}(u_1,\dots,u_s) \subseteq P\big\}.$$

Claim 3.12. *For every $P \subseteq [\![s]\!]$,*

$$|\mathcal{X}_P| \leqslant (2ke^r)^s \cdot N^{|P|+1} \cdot \left(\frac{s^2}{t}\right)^{\frac{s}{k-2}} \cdot \xi^{-s} \cdot s!.$$

Proof. We can choose the elements of every sequence $(u_1,\dots,u_s) \in \mathcal{X}_P$ one-by-one as follows: Suppose that $u_1,\dots,u_{d-1}$ have already been chosen. If $d \in P$, then we have at most N choices for u_d. Otherwise, if $d \notin P$, then u_d must be chosen from the set $\mathcal{R}\big(\{u_1,\dots,u_{d-1}\}\big)$ of elements that are rich with respect the set of previously chosen elements. By (27), this set has at most $f(d-1)$ elements, where

$$f(x) \coloneqq \begin{cases} N & \text{if } x = 0\\ \frac{2ks}{\xi} \cdot \left(\frac{s}{x}\right)^{r-1}\left(\frac{s^2}{t}\right)^{\frac{1}{k-2}} & \text{otherwise.}\end{cases}$$

Since $s \geqslant \sqrt{t}$ implies that $f(d) \geqslant 1$ for all $d \in [\![s]\!]$,

$$|\mathcal{X}_P| \leqslant N^{|P|} \cdot \prod_{d\notin P} f(d-1) \leqslant N^{|P|+1} \cdot \prod_{d=1}^{s} f(d) = N^{|P|+1}\left(\frac{2ks^r}{\xi}\right)^s \cdot \left(\frac{s^2}{t}\right)^{\frac{s}{k-2}} \cdot \prod_{d=1}^{s}\frac{1}{d^{r-1}}.$$

Using the inequality $s! \geqslant (s/e)^s$, we moreover have

$$\prod_{d=1}^{s} \frac{1}{d^{r-1}} = (s!)^{1-r} \leqslant s! \cdot (e/s)^{rs},$$

which implies the claimed bound. □

We now use the two claims to prove Lemma 3.8, where we can assume that $s \leqslant m$. Let $\tau \coloneqq 4s\big(9k^3 s/(\xi t)\big)^{1/(k-1)}$. Claim 3.11 and Markov's inequality imply that, for every $U^* \in \mathcal{C}_r(s)$, at least $s!/2$ orderings of the elements of U^* belong to $\mathcal{X}_P$ for some $P \subseteq [\![s]\!]$ with at most τ elements. Therefore,

$$|\mathcal{C}_r(s)| \leqslant \frac{2}{s!} \cdot \sum_{\substack{P \subseteq [\![s]\!] \\ |P| \leqslant \tau}} |\mathcal{X}_P| \leqslant \frac{2^{s+1}}{s!} \cdot \max_{\substack{P \subseteq [\![s]\!] \\ |P| \leqslant \tau}} |\mathcal{X}_P|.$$

and thus, by Claim 3.12,

$$|\mathcal{C}_r(s)| \leqslant 2^{s+1} \cdot (2ke^r)^s \cdot N^{\tau+1} \cdot \left(\frac{s^2}{t}\right)^{\frac{s}{k-2}} \cdot \xi^{-s}.$$

Recall the definition of ξ given in (25). Using the fact that $x \mapsto x/\beta(x)$ is increasing on $(0,1]$, by Fact 3.6, we have $s/\beta(s/m) \leqslant m/\beta(1) = m/4$ and thus

$$\tau = 4s \left(\frac{9k^3 s}{\delta \beta(s/m) t}\right)^{1/(k-1)} \leqslant 4s \left(\frac{9k^3 m}{4\delta t}\right)^{1/(k-1)}.$$

Moreover, our assumptions imply that τ is sufficiently large, so we can assume that $\tau + 1 \leqslant 4^{1/(k-1)}\tau$. Set $h \coloneqq 10k^3 e^k/\delta$. Using (25) once more, we may write

$$\begin{aligned} |\mathcal{C}_r(s)|^{1/s} &\leqslant \frac{8ke^k}{\delta} \cdot N^{4\left(\frac{9k^3 m}{\delta t}\right)^{1/(k-1)}} \cdot \left(\frac{s^2}{t}\right)^{\frac{1}{k-2}} \cdot \frac{1}{\beta(s/m)} \\ &\leqslant h \cdot N^{h(m/t)^{1/(k-1)}} \cdot \left(\frac{s^2}{t}\right)^{\frac{1}{k-2}} \cdot \frac{1}{\beta(s/m)}. \end{aligned} \tag{30}$$

To prove the first assertion of the lemma, it suffices to show that $|\mathcal{C}_r(s)| \leqslant \frac{1}{k}(\eta/p)^s$. Since $(0,1] \ni x \mapsto x^{2/(k-2)}/\beta(x)$ achieves its maximum at some w_k that depends only on k (see Fact 3.6), we have

$$\frac{s^{2/(k-2)}}{\beta(s/m)} \leqslant m^{2/(k-2)} \cdot \frac{w_k^{2/(k-2)}}{\beta(w_k)}.$$

Using (30) and our assumption $t \geqslant Cm^2 p^{k-2} N^{C(m/t)^{1/(k-1)}}$, we therefore obtain

$$|\mathcal{C}_r(s)|^{1/s} \leqslant h \cdot N^{h(m/t)^{1/(k-1)}} \cdot \left(\frac{m^2}{t}\right)^{\frac{1}{k-2}} \cdot \frac{w_k^{2/(k-2)}}{\beta(w_k)} \leqslant \frac{\eta}{k^{1/s} p},$$

provided that C is sufficiently large.

Finally, for the second assertion of the lemma, assume that $s \leqslant \sqrt{4kt}\log(1/p)$. We then have

$$\left(\frac{s^2}{t}\right)^{\frac{1}{k-2}} \leqslant \big(4k \log(1/p)^2\big)^{\frac{1}{k-2}}$$

and, as $x \mapsto 1/\beta(x)$ is decreasing and $s^2 \geqslant t \geqslant m^2 p^{k-2}$,

$$\frac{1}{\beta(s/m)} \leqslant \frac{1}{\beta(p^{(k-2)/2})} = \left(\frac{k-2}{2}\log(1/p) + 2\right)^2.$$

To bound $N^{h(m/t)^{1/(k-1)}}$, we distinguish two cases. If $p < N^{-1/(2k-2)}$, then the assumption $t \geqslant Cm$ implies

$$N^{h(m/t)^{1/(k-1)}} \leqslant N^{hC^{-1/(k-1)}} \leqslant \left(\frac{1}{p}\right)^{\eta/2}$$

for all sufficiently large C. On the other hand, if $p \geqslant N^{-1/(2k-2)}$, then we have $t \geqslant CNmp^{k-1} \geqslant mN^{1/2}$. Moreover, the inequalities $CNmp^{k-1} \leqslant t \leqslant s^2 \leqslant Nm$ imply that $p \leqslant C^{-\frac{1}{k-1}}$, so if C is large enough, then

$$N^{h(m/t)^{1/(k-1)}} \leqslant N^{hN^{-1/(2k-2)}} \leqslant 2 \leqslant \left(\frac{1}{p}\right)^{\eta/2}$$

also in this case. Substituting the above inequalities into (30), we obtain

$$|\mathcal{C}_r(s)|^{1/s} \leqslant h \cdot \left(\frac{1}{p}\right)^{\eta/2} (4k\log(1/p)^2)^{\frac{1}{k-2}} \cdot \left(\frac{k-2}{2}\log(1/p) + 2\right)^2 \leqslant \frac{1}{k^{1/s}} \cdot \left(\frac{1}{p}\right)^{\eta}$$

provided C is large enough. This completes the proof of the lemma, and with it, Proposition 3.2. □

4. The localised regime

In this section, we prove Theorem 1.1 in the localised regime. We will assume throughout this section that (p,t) is in the localised regime, that is,

$$\begin{aligned} N^{-1/(k-1)} < p \ll 1 \quad &\text{and} \quad \sqrt{t}\log(1/p) \ll t^2/\sigma^2, \quad \text{or} \\ \Omega(N^{-2/k}) \leqslant p \leqslant N^{-1/(k-1)} \quad &\text{and} \quad \sqrt{t}\log(1/p) \ll \mu \cdot \mathrm{Po}(t/\mu). \end{aligned} \tag{31}$$

Since Theorem 1.1 holds vacuously when $\mu + t > |\mathrm{AP}_k|$, we can assume without loss of generality that $t \leqslant |\mathrm{AP}_k| \leqslant N^2$. Moreover, it is straightforward to show that these assumptions and $k \geqslant 3$ also imply that $t \gg \max\{1, N^2p^{2k-2}\}$; indeed, $\mu \cdot \mathrm{Po}(t/\mu) \leqslant t^2/\mu$ and $\sigma^2 = \Omega\big(N^2p^k + N^3p^{2k-1}\big)$.

We will prove the lower and the upper bounds on the upper-tail probability of X separately. The proof of the lower bound is a fairly straightforward application of Corollary 2.3. The proof of the upper bound involves a conditioned moment argument, adapted from [18], that crucially relies on Theorem 1.2 (or rather on its alternate version, Proposition 3.2). To complicate things further, this approach breaks down in a small sliver of the localised regime, where we have to resort to a much more delicate estimate of conditioned factorial moments of X.

4.1. Proof of the lower bound in the localised regime.

Since $t \gg \max\{1, N^2p^{2k-2}\}$ in the entire localised regime, the following proposition, together with Proposition 3.1, implies the required lower bound on the tail probability.

Proposition 4.1. *Suppose that the sequence (p,t) satisfies (31). Then, for every $\varepsilon > 0$ and integer $k \geqslant 3$, and all large enough N,*

$$\log \mathbb{P}\left(X \geqslant \mu + t\right) \geqslant -(1+\varepsilon) \cdot \Psi^*\big((1+\varepsilon)t\big) \cdot \log(1/p).$$

Proof. Fix an $\varepsilon > 0$ and let $U \subseteq [\![N]\!]$ be a smallest set satisfying $A_k(U) \geqslant (1+\varepsilon)t$, so that $|U| = \Psi^*\big((1+\varepsilon)t\big)$. Define $\mathbb{Q} \coloneqq \mathbb{P}(\cdot \mid U \subseteq \mathbf{R})$; more explicitly,

$$\mathbb{Q}(R) = \begin{cases} p^{|R|-|U|}(1-p)^{N-|R|} & \text{if } U \subseteq R, \\ 0 & \text{otherwise.} \end{cases}$$

We apply Corollary 2.3 to the event $\{X \geqslant \mu + t\}$, and conclude that

$$\log \mathbb{P}(X \geqslant \mu + t) \geqslant \log \mathbb{Q}(X \geqslant \mu + t) - \mathbb{E}_{\mathbb{Q}}\left[\log \frac{d\mathbb{Q}}{d\mathbb{P}}(\mathbf{R}) \mid X \geqslant \mu + t\right].$$

Since $\log(d\mathbb{Q}/d\mathbb{P})$ is identically equal to $|U|\log(1/p)$ on the support of $\mathbb{Q}$, we have

$$\log \mathbb{P}(X \geqslant \mu + t) \geqslant \log \mathbb{Q}(X \geqslant \mu + t) - |U|\log(1/p) = \log \mathbb{Q}(X \geqslant \mu + t) - \Psi^*\big((1+\varepsilon)t\big)\log(1/p).$$

As $\psi^*\big((1+\varepsilon)t\big) \geqslant \sqrt{t}$, by Proposition 3.1, in order to complete the proof, it suffices to show that

$$\log \mathbb{Q}(X \geqslant \mu + t) > -\varepsilon \cdot \sqrt{t}\log(1/p). \tag{32}$$

To this end, note that, by (18) and since $p \to 0$, we have

$$\mathbb{E}_{\mathbb{Q}}[X] = \mathbb{E}_U[X] \geqslant \mu + A_k(U) \cdot (1 - p^k) \geqslant \mu + (1 + \varepsilon/2)t$$

for all N large enough. Since $X \leqslant |\mathrm{AP}_k| \leqslant N^2$ always, we may conclude that

$$\mu + (1+\varepsilon/2)t \leqslant \mathbb{E}_{\mathbb{Q}}[X] \leqslant (\mu + t) + N^2 \cdot \mathbb{Q}(X \geqslant \mu + t),$$

which implies that

$$\log \mathbb{Q}(X \geqslant \mu + t) \geqslant \log\left(\frac{\varepsilon t}{2N^2}\right) \geqslant -2\log N.$$

Finally, we prove that $2\log N \leqslant \varepsilon\sqrt{t}\log(1/p)$ by distinguishing two cases. If $p \leqslant N^{-1/k}$, i.e., if $\log(1/p) \geqslant (\log N)/k$, this follows from $t \gg 1$. Otherwise, if $p > N^{-1/k}$, the assumption that (p,t) is in the localised regime implies that $\sqrt{t}\log(1/p) \gg \sqrt{t} \gg \sigma^{2/3}$, which easily implies the claim since $\sigma^2 = \Omega(N^2 p^k) = \Omega(N)$. □

4.2. **Proof of the upper bound in the localised regime.** In view of Proposition 3.2, to prove the upper bound in the localised regime, it suffices to show that the upper-tail event is dominated by the appearance of a small $(1-\varepsilon)t$-seed, that is, an element of $\mathcal{S}_{\mathrm{small}}((1-\varepsilon)t, C)$ for a suitably large $C = C(k,\varepsilon)$, where $\mathcal{S}_{\mathrm{small}}$ is defined as in the statement of Proposition 3.2.

Proposition 4.2. *Suppose that the sequence (p,t) satisfies (31). Then, for all $C, \varepsilon > 0$, every integer $k \geqslant 3$, and all sufficiently large N,*

$$\mathbb{P}\left(X \geqslant \mu + t\right) \leqslant (1+\varepsilon) \cdot \mathbb{P}\left(U \subseteq \mathbf{R} \text{ for some } U \in \mathcal{S}_{\mathrm{small}}\big((1-\varepsilon)t, C\big)\right).$$

In order to prove Proposition 4.2, we will bound the probability of the upper-tail event occurring *without* the appearance of a small $(1-\varepsilon)t$-seed and then compare that bound with the lower bound on the upper tail probability that we proved in the previous subsection. The following lemma will provide a suitable estimate on the probability of the upper-tail event occurring without the appearance of a small $(1-\varepsilon)t$-seed in all but a tiny sliver of the localised regime. Even though it is implicitly proved in [18], we include its proof (an adaptation of the elegant argument from [21]) here for completeness. Given $u, m > 0$, let $Z(u,m)$ be the indicator random variable of the event that $\mathbf{R}$ does not include a u-seed of size at most m.

Lemma 4.3. *For every $u \geqslant 0$ and every positive integer m,*

$$\mathbb{E}[X^m \cdot Z(u, km)] \leqslant (\mu + u)^m.$$

Proof. Recall that $\mathcal{S}(u)$ denotes the set of all u-seeds. For any $S \subseteq [\![N]\!]$, let Z_S be the indicator of the event that $\mathbf{R} \cap S$ does not contain any $U \in \mathcal{S}(u)$ with $|U| \leqslant km$ as a subset, so that $Z_{[\![N]\!]} = Z(u, km)$; note that $Z_S \leqslant Z_{S'}$ if $S' \subseteq S \subseteq [\![N]\!]$. For any $B \subseteq [\![N]\!]$, let Y_B be the indicator of the event that $B \subseteq \mathbf{R}$. Since we can write $X = \sum_{B \in \mathrm{AP}_k} Y_B$, it is straightforward to see that, for every nonnegative integer ℓ,

$$X^\ell \cdot Z = \sum_{B_1, \dots, B_\ell \in \mathrm{AP}_k} Y_{B_1 \cup \dots \cup B_\ell} \cdot Z \leqslant \sum_{B_1, \dots, B_\ell \in \mathrm{AP}_k} Y_{B_1 \cup \dots \cup B_\ell} \cdot Z_{B_1 \cup \dots \cup B_{\ell-1}} =: M_\ell.$$

The required inequality will clearly follow if we show that $\mathbb{E}[M_\ell] \leqslant (\mu + u)^\ell$ for all $\ell \in [\![m]\!]$. We prove this stronger estimate using induction on ℓ. The base case $\ell = 1$ holds vacuously, as $M_1 = X$. Suppose now that $\ell \geqslant 2$ and that $\mathbb{E}[M_{\ell-1}] \leqslant (\mu + u)^{\ell-1}$. Fix some $B_1, \dots, B_{\ell-1} \in \mathrm{AP}_k$, let $U \coloneqq B_1 \cup \dots \cup B_{\ell-1}$, and let $\mathcal{A}_U$ denote the event that $Y_U \cdot Z_U = 1$. Since $|U| \leqslant k\ell \leqslant km$, $\mathcal{A}_U$ holds if and only if $U \subseteq \mathbf{R}$ and $U \notin \mathcal{S}(u)$. Therefore, if $\mathcal{A}_U$ has nonzero probability, we have $\mathbb{E}[X \mid \mathcal{A}_U] = \mathbb{E}[X \mid Y_U = 1] \leqslant \mu + u$. This means that

$$\sum_{B_\ell \in \mathrm{AP}_k} \mathbb{E}[Y_{U \cup B_\ell} \cdot Z_U] = \mathbb{E}[Y_U \cdot Z_U] \cdot \mathbb{E}[X \mid \mathcal{A}_U] \leqslant \mathbb{E}[Y_U \cdot Z_{B_1 \cup \dots \cup B_{\ell-2}}] \cdot (\mu + u).$$

Summing this inequality over all $B_1, \dots, B_{\ell-1} \in \mathrm{AP}_k$ yields $\mathbb{E}[M_\ell] \leqslant \mathbb{E}[M_{\ell-1}] \cdot (\mu + u) \leqslant (\mu + u)^\ell$. □

In the following, let $m_{\max}(u, C)$ be the largest m that satisfies

$$u \geqslant Cm \cdot \max\{1, Np^{k-1}\} \qquad \text{and} \qquad u \geqslant Cm^2 p^{k-2} \cdot N^{(k-2)(m/u)^{1/(k-1)}}, \tag{33}$$

so that a u-seed U belongs to $\mathcal{S}_{\mathrm{small}}(u)$ precisely when $|U| \leqslant m_{\max}(u, C)$. Moreover, let $Z_u \coloneqq Z(u, m_{\max}(u, C))$ be the indicator random variable for the event that $\mathbf{R}$ does not contain a small u-seed.

As in [18], the above lemma permits us to derive an upper bound on the probability of the upper-tail event occurring without the appearance of a small $(1-\varepsilon)t$-seed by applying Markov's inequality to the variable $X^{m_{\max}((1-\varepsilon)t, C)/k} \cdot Z_{(1-\varepsilon)t}$. However, this bound is only useful when it is smaller than the available lower bound on the upper-tail. Unfortunately, there is a small portion of the localised regime where this is not the case. Specifically, this happens when $\mu \ll t \leqslant O\big(\big(\log(1/p)\big)^2\big)$; we shall informally say that this case falls into the *very sparse localised regime*. In this regime, instead of bounding the classical moments of X on the event that $\mathbf{R}$ does not contain a small $(1-\varepsilon)t$-seed, we will need to bound factorial moments of X. More precisely, given an integer r, let $X^{\underline{r}} \coloneqq X(X-1)\cdots(X-r+1)$ denote the r-th falling factorial of X. The following estimate is a special case of the more general Proposition 6.3, which is proved in Section 6.2.

Corollary 4.4. *For every $k \geqslant 3$, there exists a constant K such that the following holds for all $C, \varepsilon > 0$. Suppose that $\Omega(N^{-2/k}) \leqslant p \ll N^{-1/(k-1)}$ and that t is a positive integer satisfying $t \leqslant \big(\log(1/p)\big)^3$. Then, for all large enough N and all $u \leqslant t$,*

$$\mathbb{E}\left[X^{\underline{t}} \cdot Z_u\right] \leqslant \mu^t \cdot \exp\big((u + \varepsilon t/2)\log(1 + Kt/\mu)\big).$$

Proof of Proposition 4.2. Fix an $\varepsilon > 0$, let $u \coloneqq (1-\varepsilon)t$, and define

$$m \coloneqq \left\lceil \frac{2k \cdot (\mu + t) \cdot \log(1/p)}{\varepsilon\sqrt{t}} \right\rceil$$

We will first consider the case where $km \leqslant m_{\max}(u, C)$. Let $Z \coloneqq Z(u, km)$ and note that, as $km \leqslant m_{\max}(u, C)$, we have $Z \geqslant Z_u$. Since Z is an indicator random variable, Markov's inequality and Lemma 4.3 then give

$$\mathbb{P}\big(X \geqslant \mu + t\big) \leqslant \mathbb{P}\big(X \cdot Z \geqslant \mu + t\big) + \mathbb{P}(Z = 0) \leqslant \frac{\mathbb{E}[X^m \cdot Z]}{(\mu + t)^m} + \mathbb{P}(Z = 0) \leqslant \left(\frac{\mu + u}{\mu + t}\right)^m + \mathbb{P}(Z_u = 0).$$

Since

$$\mathbb{P}(Z_u = 0) = \mathbb{P}\left(U \subseteq \mathbf{R} \text{ for some } U \in \mathcal{S}_{\text{small}}\big((1-\varepsilon)t, C\big)\right),$$

the proposition will follow from

$$\left(\frac{\mu + u}{\mu + t}\right)^m \leqslant \frac{\varepsilon}{1+\varepsilon} \cdot \mathbb{P}\left(X \geqslant \mu + t\right).$$

To this end, observe that, by the definitions of u and m,

$$\left(\frac{\mu + u}{\mu + t}\right)^m = \left(1 - \frac{\varepsilon t}{\mu + t}\right)^m \leqslant \exp\left(-\frac{\varepsilon t m}{\mu + t}\right) \leqslant \exp\left(-2k\sqrt{t} \cdot \log(1/p)\right).$$

The assumptions of Proposition 3.1 hold throughout the localised regime, and so, thanks to Propositions 3.1 and 4.1, we may conclude that

$$\log \mathbb{P}\left(X \geqslant \mu + t\right) \geqslant -(1+\varepsilon) \cdot \Psi^*\big((1+\varepsilon)t\big) \cdot \log(1/p) \geqslant -(1+3\varepsilon) \cdot \sqrt{2(k-1)t} \cdot \log(1/p).$$

Finally, since $t \gg 1$ in the entire localised regime, we obtain

$$\mathbb{P}\left(X \geqslant \mu + t\right) \geqslant \frac{1+\varepsilon}{\varepsilon} \cdot \exp\left(-2k\sqrt{t} \cdot \log(1/p)\right) \geqslant \frac{1+\varepsilon}{\varepsilon} \cdot \left(\frac{\mu + u}{\mu + t}\right)^m$$

for all large enough N. This gives the assertion of the proposition in the case where $km \leqslant m_{\max}(u, C)$. The following claim states that this inequality holds unless $\mu \ll t = O\big(\big(\log(1/p)\big)^2\big)$.

Claim 4.5. *We have $km \leqslant m_{\max}(u, C)$ unless $\mu \ll t \leqslant C'\big(\log(1/p)\big)^2$ for some constant $C' = C'(k, C)$.*

Proof. Observe first that the inequality $km \leqslant m_{\max}(u, C)$ is equivalent to

$$u \geqslant Ckm \cdot \max\{1, Np^{k-1}\} \qquad \text{and} \qquad u \geqslant Ck^2m^2p^{k-2} \cdot N^{(k-2)(km/u)^{1/(k-1)}}. \tag{34}$$

Observe further that, if C is sufficiently large, then the first inequality in (34) implies that

$$N^{(k-2)(km/u)^{1/(k-1)}} \leqslant N^{(k-2)(C\max\{1, Np^{k-1}\})^{-1/(k-1)}} \leqslant p^{-1/4}$$

for all sufficiently large N, where the last inequality can be verified by considering separately the cases $Np^{k-1} \leqslant N^{\delta}$ and $Np^{k-1} > N^{\delta}$ for some small enough $\delta > 0$. In particular (34) follows from

$$u \geqslant Ckm \cdot \max\{1, Np^{k-1}\} \qquad \text{and} \qquad u \geqslant Ck^2m^2p^{k-9/4}. \tag{35}$$

Further, by the definition of m,

$$\frac{u}{m} \geqslant \frac{t}{2m} \geqslant \frac{\varepsilon}{6k\log(1/p)} \cdot \frac{t^{3/2}}{\mu + t} \qquad \text{and} \qquad \frac{u}{m^2} \geqslant \frac{t}{2m^2} \geqslant \left(\frac{\varepsilon}{3k\log(1/p)} \cdot \frac{t}{\mu + t}\right)^2$$

and, using also $p^{(4k-9)/8}\log(1/p) \ll p^{(k-2)/3}$ for $k \geqslant 3$, it is thus enough to show that

$$\frac{t^{3/2}}{\mu + t} \geqslant C' \max\{1, Np^{k-1}\}\log(1/p) \qquad \text{and} \qquad \frac{t}{\mu + t} \geqslant p^{(k-2)/3} \tag{36}$$

for some $C' = C'(C, k, \varepsilon)$, unless $\mu \ll t \leqslant 2C'\big(\log(1/p)\big)^2$.

Assume first that $Np^{k-1} > 1$ and thus $\sigma^2 = \Theta(N^3p^{2k-1})$ and $\sqrt{t}\log(1/p) \ll t^2/\sigma^2$, by the definition of the localised regime. In particular, we may assume that, for every $K > 0$ and all sufficiently large N, we have $t \geqslant t_K$, where

$$t_K \coloneqq \big(K\sigma^2\log(1/p)\big)^{2/3} = \Theta\big(K^{2/3}N^2p^{(4k-2)/3}\log(1/p)^{2/3}\big) \ll \mu;$$

the last inequality holds as $k \geqslant 3$ and $p \ll 1$. Since the function $(0,\infty) \ni x \mapsto \frac{x^{3/2}}{\mu+x}$ is increasing, we have

$$\frac{t^{3/2}}{\mu+t} \geqslant \frac{t_K^{3/2}}{\mu+t_K} \geqslant \frac{K\sigma^2\log(1/p)}{2\mu} \geqslant Kc_kNp^{k-1}\log(1/p),$$

for some $c_k > 0$. Choosing K appropriately, we thus obtain the first inequality in (36). Since the function $(0,\infty) \in x \mapsto \frac{x}{\mu+x}$ is also increasing,

$$\frac{t}{\mu+t} \geqslant \frac{t_K}{\mu+t_K} \geqslant \frac{\big(K\sigma^2\log(1/p)\big)^{2/3}}{2\mu} \geqslant \frac{K^{2/3}\sigma^{4/3}}{2\mu} \geqslant K^{2/3}c_k' \cdot \frac{N^2p^{(4k-2)/3}}{N^2p^k} \geqslant p^{(k-2)/3},$$

for some $c_k' > 0$, giving the second inequality in (36).

Assume now that $Np^{k-1} \leqslant 1$ and thus $\sqrt{t}\log(1/p) \ll \mu \cdot \mathrm{Po}(t/\mu)$. If $t = O(\mu)$, then the estimate $\mu \cdot \mathrm{Po}(t/\mu) \leqslant t^2/\mu$ immediately implies the first inequality in (36); further, since the same estimate implies that $t^{3/2} \gg \mu$, we obtain, using $Np^{k-1} \leqslant 1$, that

$$\frac{t}{\mu+t} \geqslant \frac{t}{O(\mu)} \gg \mu^{-1/3} \geqslant \big(N^2p^k\big)^{-1/3} \geqslant p^{(k-2)/3},$$

which gives the second inequality in (36). We may therefore assume that $t \gg \mu$. In this case,

$$\frac{t^{3/2}}{\mu+t} \geqslant \frac{\sqrt{t}}{2} \qquad \text{and} \qquad \frac{t}{\mu+t} \geqslant \frac{1}{2} \geqslant p^{(k-2)/3}.$$

In particular, both inequalities in (36) hold unless $\mu \ll t \leqslant \big(2C'\log(1/p)\big)^2$. □

Assume now that $\mu \ll t \leqslant \big(\log(1/p)\big)^3$. Since $X \geqslant \mu + t$ implies $X^{\underline{t}} \geqslant (\mu+t)^{\underline{t}}$, and since Z_u is an indicator random variable, we have

$$\mathbb{P}(X \geqslant \mu+t) \leqslant \mathbb{P}\big(X^{\underline{t}} \cdot Z_u \geqslant (\mu+t)^{\underline{t}}\big) + \mathbb{P}(Z_u = 0).$$

Since $X^{\underline{t}} \cdot Z_u \geqslant 0$, as X is integer-valued, we may apply Markov's inequality and Corollary 4.4 to conclude that

$$\mathbb{P}\big(X^{\underline{t}} \cdot Z_u \geqslant (\mu+t)^{\underline{t}}\big) \leqslant \frac{\mathbb{E}[X^{\underline{t}} \cdot Z_u]}{(\mu+t)^{\underline{t}}} \leqslant \frac{\mu^t}{(\mu+t)^{\underline{t}}} \cdot \exp\big((u+\varepsilon t/2)\log(1+Kt/\mu)\big).$$

Further,

$$\log\frac{\mu^t}{(\mu+t)^{\underline{t}}} \leqslant -\int_0^t \log\frac{\mu+x}{\mu}\,dx = -\mu\cdot\int_0^{t/\mu}\log(1+y)\,dy = -\mu\cdot\mathrm{Po}\left(\frac{t}{\mu}\right),$$

whereas the assumption $t \gg \mu$ gives

$$t\log\left(1+\frac{Kt}{\mu}\right) = (1+o(1))\cdot\left((\mu+t)\log\left(1+\frac{t}{\mu}\right)-t\right) = (1+o(1))\cdot\mu\cdot\mathrm{Po}\left(\frac{t}{\mu}\right).$$

Consequently,

$$\mathbb{P}(X \geqslant \mu+t) \leqslant \exp\big(-\varepsilon\mu/4\cdot\mathrm{Po}(t/\mu)\big) + \mathbb{P}(Z_u = 0).$$

Finally, since $\mu\cdot\mathrm{Po}(t/\mu) \gg \sqrt{t}\log(1/p)$ in the localised regime,

$$\exp\big(-\varepsilon\mu/4\cdot\mathrm{Po}(t/\mu)\big) \leqslant \exp\Big(-2k\sqrt{t}\cdot\log(1/p)\Big) \leqslant \frac{\varepsilon}{1+\varepsilon}\cdot\mathbb{P}\left(X \geqslant \mu+t\right),$$

which implies the assertion of the proposition. □

5. The Gaussian regime

In this section, we prove Theorem 1.1 in the Gaussian regime, i.e., for all sequences (p, t) satisfying

$$Np^{k-1} \gg 1, \quad t \gg \sigma, \quad \text{and} \quad \sqrt{t}\log(1/p) \gg t^2/\sigma^2.$$

The lower and the upper bounds on the upper-tail probability of X will be proved separately. In Section 5.1, we prove the lower bound by applying Proposition 2.4 to a product measure $\mathbb{Q}$ that is a small perturbation of the p-biased product measure $\mathbb{P}$. In Section 5.2, we prove the matching upper bound by applying our martingale concentration inequality (Proposition 2.5) to the hypergraph AP_k and bounding the three error terms using combinatorial arguments.

Throughout the section, it will be convenient to work with an expression that closely approximates the variance of X and involves only the numbers of k-APs that contain a given $i \in [\![N]\!]$. For any $i \in [\![N]\!]$, denote by $A_1(i)$ the number of k-term arithmetic progressions that contain i (recalling the notation $A_r^{(k)}(U)$ used in Section 3, we can see that this definition corresponds to $A_1^{(k)}(\{i\})$). Define

$$\mathcal{V} \coloneqq \sum_{i \in [\![N]\!]} A_1(i)^2 \cdot p^{2k-1}.$$

Lemma 5.1. *For every $\varepsilon > 0$, there exists a C such that, whevener $CN^{-1/(k-1)} \leqslant p \leqslant 1/C$,*

$$(1-\varepsilon)\cdot \mathcal{V} \leqslant \mathrm{Var}(X) \leqslant (1+\varepsilon)\cdot \mathcal{V}.$$

Proof. Since

$$\mathrm{Var}(X) - \mathcal{V} = \sum_{B,B' \in \mathrm{AP}_k} \left(p^{|B \cup B'|} - p^{2k} - |B \cap B'| \cdot p^{2k-1} \right),$$

we have

$$|\mathrm{Var}(X) - \mathcal{V}| \leqslant \left|\left\{(B,B') \in \mathrm{AP}_k^2 : |B \cap B'| = 1\right\}\right| \cdot p^{2k} + \left|\left\{(B,B') \in \mathrm{AP}_k^2 : |B \cap B'| \geqslant 2\right\}\right| \cdot p^k.$$

It is easy to see that the first term in the right-hand side is at most $p \cdot \mathcal{V}$. Moreover, as every pair of elements of $[\![N]\!]$ is contained in at most $\binom{k}{2}$ arithmetic progressions of length k, the second term is at most $|\mathrm{AP}_k| \cdot \binom{k}{2}^2 \cdot p^k$. Finally, since $|\mathrm{AP}_k| = \Theta(N^2)$ and $\mathcal{V} = \Theta(N^3 p^{2k-1})$, the assertion of the lemma now follows from the assumptions on p provided that C is sufficiently large as a function of ε. □

5.1. **Proof of the lower bound in the Gaussian regime.** We call a measure $\mathbb{Q}$ supported on subsets of $[\![N]\!]$ a *p-bounded product measure* if it is a product measure with $\mathbb{Q}(i \in \mathbf{R}) \in [p, 2p]$ for every $i \in [\![N]\!]$. We first show that the variance of X under an arbitrary p-bounded product measure is not much larger than the variance σ^2 taken with respect to the p-biased product measure $\mathbb{P}$:

Lemma 5.2. *Assume that $p \leqslant 1/2$ and that $\mathbb{Q}$ is a p-bounded product measure. Then, for all sufficiently large N,*

$$\mathrm{Var}_{\mathbb{Q}}(X) \leqslant 2^{2k} \cdot \sigma^2.$$

Proof. On the one hand,

$$\begin{aligned}\mathrm{Var}_{\mathbb{Q}}(X) &= \sum_{\substack{B,B'\in \mathrm{AP}_k\\ B\cap B'\neq\varnothing}} \big(\mathbb{Q}(B\cup B'\subseteq \mathbf{R}) - \mathbb{Q}(B\subseteq\mathbf{R})\cdot\mathbb{Q}(B'\subseteq\mathbf{R})\big)\\ &\leqslant \sum_{\substack{B,B'\in \mathrm{AP}_k\\ B\cap B'\neq\varnothing}} 2^{|B\cup B'|}\cdot\mathbb{P}(B\cup B'\subseteq\mathbf{R}) \leqslant 2^{2k-1}\cdot \sum_{\substack{B,B'\in \mathrm{AP}_k\\ B\cap B'\neq\varnothing}} \mathbb{P}(B\cup B'\subseteq\mathbf{R}).\end{aligned}$$

On the other hand,

$$\sigma^2 = \mathrm{Var}(X) = \sum_{\substack{B,B'\in \mathrm{AP}_k\\ B\cap B'\neq\varnothing}} \big(\mathbb{P}(B\cup B'\subseteq\mathbf{R}) - \mathbb{P}(B\subseteq\mathbf{R})\cdot\mathbb{P}(B'\subseteq\mathbf{R})\big) \geqslant \frac{1}{2}\sum_{\substack{B,B'\in \mathrm{AP}_k\\ B\cap B'\neq\varnothing}} \mathbb{P}(B\cup B'\subseteq\mathbf{R}),$$

where the inequality follows from $p\leqslant 1/2$. □

We restate the lower bound of Theorem 1.1 in the Gaussian regime in a non-asymptotic form and prove it by applying Proposition 2.4 to a well-chosen p-bounded product measure. Note that we state and prove the lower bound on the upper-tail probability of X in a larger region of the parameter space that includes also a part of the localised regime (where this Gaussian-like lower estimate is no longer optimal).

Proposition 5.3. *For every $\varepsilon>0$ and integer $k\geqslant 3$, there exists $C>0$ such that the following holds. If*

$$C\cdot N^{-1/(k-1)}\leqslant p\leqslant 1/C \qquad \text{and} \qquad C\cdot\sigma\leqslant t\leqslant \mu/C,$$

then

$$\log\mathbb{P}\left(X\geqslant \mu+t\right)\geqslant -(1+\varepsilon)\cdot\frac{t^2}{2\sigma^2}.$$

Proof. For each $i\in[\![N]\!]$, set

$$q_i \coloneqq \frac{(1+\varepsilon)tp^k}{\mathcal{V}}\cdot A_1(i)$$

and let $\mathbb{Q}$ be the product measure given by $\mathbb{Q}(i\in\mathbf{R}) = p+q_i$ for all i. Since $A_1(i)\leqslant kN$ for all i and $\sigma^2=\Theta(\mu\cdot Np^{k-1})$, Lemma 5.1 implies that

$$\max_{i\in[\![N]\!]}\frac{q_i}{p}\leqslant\frac{(1+\varepsilon)ktNp^{k-1}}{\mathcal{V}}\leqslant\frac{(1+\varepsilon)^2ktNp^{k-1}}{\sigma^2}\leqslant\frac{(1+\varepsilon)^2k\mu\cdot Np^{k-1}}{C\sigma^2}\leqslant\frac{c_k}{C}$$

for some c_k that depends only on k; in particular, $\mathbb{Q}$ is p-bounded, provided that C is large. Crucially,

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}[X]-\mu &= \sum_{B\in\mathrm{AP}_k}\big(\mathbb{Q}(B\subseteq\mathbf{R})-\mathbb{P}(B\subseteq\mathbf{R})\big) = \sum_{B\in\mathrm{AP}_k}\left(\prod_{i\in B}(p+q_i)-p^k\right)\\ &\geqslant \sum_{B\in\mathrm{AP}_k}\sum_{i\in B}q_ip^{k-1} = \sum_{i\in[\![N]\!]}A_1(i)q_ip^{k-1} = \frac{(1+\varepsilon)t}{\mathcal{V}}\cdot\sum_{i\in[\![N]\!]}A_1(i)^2p^{2k-1} = (1+\varepsilon)t\end{aligned}$$

and consequently, by Chebyshev's inequality and Lemma 5.2,

$$\mathbb{Q}(X<\mu+t)\leqslant\mathbb{Q}(X<\mathbb{E}_{\mathbb{Q}}[X]-\varepsilon t)\leqslant\frac{\mathrm{Var}_{\mathbb{Q}}(X)}{\varepsilon^2t^2}\leqslant\frac{2^{2k}\sigma^2}{\varepsilon^2t^2}\leqslant\frac{2^{2k}}{\varepsilon^2C^2}.$$

In particular, if C is sufficiently large as a function of ε, Proposition 2.4 implies that

$$\mathbb{P}(X\geqslant\mu+t)\geqslant-(1+\varepsilon)\cdot D_{\mathrm{KL}}(\mathbb{Q}\,\|\,\mathbb{P}).$$

Since both $\mathbb{P}$ and $\mathbb{Q}$ are products of Bernoulli distributions, Fact 2.1 implies that

$$D_{\mathrm{KL}}(\mathbb{Q}\,\|\,\mathbb{P}) = \sum_{i=1}^{N} D_{\mathrm{KL}}\big(\mathrm{Ber}(p+q_i)\,\|\,\mathrm{Ber}(p)\big) = \sum_{i=1}^{N} j_p(p+q_i),$$

where $j_p(x) \coloneqq x\log(x/p)+(1-x)\log((1-x)/(1-p))$. It is straightforward to verify that $j_p(p) = j_p'(p) = 0$ and that $j_p''(x) = 1/x(1-x)$. Finally, by expanding j_p in Taylor series with Lagrange remainder and using the assumption $p \leqslant 1/C$, we can conclude that, whenever C is sufficiently large as a function of ε,

$$D_{\mathrm{KL}}(\mathbb{Q}\,\|\,\mathbb{P}) = \sum_{i\in[\![N]\!]} j_p(p+q_i) \leqslant \sum_{i\in[\![N]\!]} \frac{q_i^2}{2p(1-p)} = \frac{(1+\varepsilon)^2t^2}{2(1-p)\mathcal{V}} \leqslant (1+\varepsilon)^4\cdot\frac{t^2}{2\sigma^2},$$

completing the proof. □

5.2. **Proof of the upper bound in the Gaussian regime.** Define, for every $i \in [\![N]\!]$,

$$\mathcal{B}_k'(i) \coloneqq \big\{B\setminus\{i\} : B \in \mathrm{AP}_k \text{ and } i\in B\big\},$$

and let $L_i \coloneqq |\{B\in\mathcal{B}_k'(i) : B\subseteq\mathbf{R}\}|$; in other words, $\mathcal{B}_k'(i)$ is the link of i in the hypergraph AP_k, cf. the definition of L_i given in (12). We will prove the claimed Gaussian-like upper bound for the upper tail of X by applying Proposition 2.5 and showing that, for all sequences (p,t) in the Gaussian regime, the three error terms involving upper-tail estimates on the respective functionals of L_i are dominated by the Gaussian term. In particular, it suffices to establish the following three estimates:

Proposition 5.4. *For all $\varepsilon>0$ and $k\geqslant 3$, there exists C such that the following holds. If*

$$C\cdot N^{-1/(k-1)} \leqslant p \leqslant 1/C \qquad \text{and} \qquad C\cdot\sigma \leqslant t \leqslant \frac{(\sigma^2\log(1/p))^{2/3}}{C},$$

then

$$\mathbb{P}\left(\exists i\ \ L_i > \frac{\sigma^2\log(1/p)}{2t}\right) < \exp\left(-\frac{t^2}{\sigma^2}\right), \tag{37}$$

$$\mathbb{P}\left(\left|\left\{i : L_i > \frac{\sigma^2}{t}\right\}\right| \geqslant \frac{\varepsilon t^2}{\sigma^2 p^{1/2}}\right) < N^{-1}\cdot\exp\left(-\frac{t^2}{\sigma^2}\right), \tag{38}$$

$$\mathbb{P}\left(\sum_{i=1}^{N} L_i^2 > (1+\varepsilon)\cdot\frac{\sigma^2}{p}\right) < N^{-1}\cdot\exp\left(-\frac{t^2}{\sigma^2}\right). \tag{39}$$

Since the asymptotic inequality $\sqrt{t}\log(1/p) \gg t^2/\sigma^2$ clearly implies that $t \leqslant \big(\sigma^2\log(1/p)\big)^{2/3}/C$ for all C and all sufficiently large N, Propositions 2.5 and 5.4 yield the claimed Gaussian-like upper bound on the upper-tail probability of X in the entire Gaussian regime.

The remainder of this section is dedicated to proving the three estimates of Proposition 5.4, sequentially. We will prove (37) (resp. (38)) by showing that the corresponding upper-tail events imply the appearance of a large collection of pairwise-disjoint elements of $\mathcal{B}_k'(i)$ (resp. 'nearly' pairwise-disjoint elements of $\bigcup\{\mathcal{B}_k'(i) : L_i > \sigma^2/t\}$) in $\mathbf{R}$ and then bounding the probability of the latter event using a straightforward first-moment calculation. The proof of (39) will be another adaptation of the elegant argument from [21], which played a key role in our proof of the upper bound in the localised regime. We begin with a helpful combinatorial fact that will help us find large collections of pairwise-disjoint elements of $\mathcal{B}_k'(i)$.

Fact 5.5. *For each $i \in [\![N]\!]$, every collection $\mathcal{B} \subseteq \mathcal{B}'_k(i)$ contains a subcollection $\mathcal{D} \subseteq \mathcal{B}$ with $|\mathcal{D}| \geqslant |\mathcal{B}|/k^3$ whose elements are pairwise disjoint.*

Proof. Since any two numbers are contained in at most $\binom{k}{2}$ k-APs, it follows that every $B \in \mathcal{B}'_k(i)$ intersects at most $(k-1)\binom{k}{2} \leqslant k^3$ sets from $\mathcal{B}'_k(i)$, including B itself. In particular, given an arbitrary collection $\mathcal{B} \subseteq \mathcal{B}'_k(i)$, we may construct a family $\mathcal{D} \subseteq \mathcal{B}$ of pairwise-disjoint sets greedily by adding sets $B \in \mathcal{B}$ one by one, each time removing the (at most k^3) sets that intersect B from $\mathcal{B}$. The resulting collection $\mathcal{D}$ satisfies $|\mathcal{D}| \geqslant |\mathcal{B}|/k^3$. □

Let us also note that the assumptions of Proposition 5.4 imply that

$$\frac{\sigma^2}{t} = \frac{(\sigma^2 \log(1/p))^{2/3} \cdot \sigma^{2/3}}{t \cdot (\log(1/p))^{2/3}} \geqslant \frac{C\sigma^{2/3}}{(\log(1/p))^{2/3}} \geqslant \frac{\sqrt{C}Np^{2k/3-1/3}}{(\log(1/p))^{2/3}} \geqslant \sqrt{C}Np^{k-1} \cdot p^{-1/4}, \tag{40}$$

whenever C is chosen to be sufficiently large as a function of k.

Proof of (37) *in Proposition 5.4.* Set $u \coloneqq \sigma^2 \log(1/p)/(2t)$. By Fact 5.5, the inequality $L_i \geqslant u$ implies that there is a collection $\mathcal{D} \subseteq \mathcal{B}'_k(i)$ of pairwise-disjoint subsets of $\mathbf{R}$ with cardinality at least $b \coloneqq \lceil u/k^3 \rceil$. Let Z_i denote the number of such collections, so that

$$\mathbb{P}(L_i \geqslant u) \leqslant \mathbb{P}(Z_i = 1) \leqslant \mathbb{E}[Z_i].$$

Since $|\mathcal{B}'_k(i)| \leqslant kN$ and the union of all sets in each collection counted by Z_i has $(k-1)b$ elements,

$$\mathbb{E}[Z_i] \leqslant \binom{kN}{b} \cdot p^{(k-1)b} \leqslant \left(\frac{ekNp^{k-1}}{b}\right)^b \leqslant \left(\frac{ek^4Np^{k-1}}{u}\right)^{u/k^3} \leqslant p^{u/(4k^3)},$$

where the penultimate inequality holds because $b \geqslant u/k^3 \geqslant kNp^{k-1}$ and, when $a > 0$, the function $x \mapsto (ea/x)^x$ is decreasing on $[a, \infty)$ and the ultimate inequality holds when C is sufficiently large as a function of k, by (40). By the union bound over all $i \in [\![N]\!]$,

$$\mathbb{P}\left(\exists i \;\; L_i \geqslant \frac{\sigma^2 \log(1/p)}{2t}\right) \leqslant Np^{u/(4k^3)} \leqslant N \cdot \exp\left(\frac{-C\sigma^2(\log(1/p))^2}{4k^3 t}\right) \leqslant \exp\left(\frac{-C\sigma^2(\log(1/p))^2}{5k^3 t}\right),$$

where the last inequality holds as $\sigma^2/t \geqslant N^{1/(4k-4)}$ by (40) and our lower-bound assumption on p. Finally, the upper-bound assumption on t implies that

$$\sigma^2(\log(1/p))^2/t \geqslant \sqrt{t}\log(1/p) \geqslant \frac{t^2}{\sigma^2},$$

which results in the claimed bound. □

Proof of (38) *in Proposition 5.4.* Set

$$u \coloneqq \frac{\sigma^2}{t} \qquad \text{and} \qquad I \coloneqq \{i \in [\![N]\!] : L_i > u\}.$$

We first claim that, when $|I| > (m-1)^2 \cdot u$ for some positive integer m, then there exist distinct $i_1, \dots, i_m \in [\![N]\!]$ and collections $\mathcal{D}_1 \subseteq \mathcal{B}'_k(i_1), \dots, \mathcal{D}_m \subseteq \mathcal{B}'_k(i_m)$, each comprising $b \coloneqq \lceil u/(4k^3) \rceil$ pairwise-disjoint subsets of $\mathbf{R}$, such that, for every $j \in [\![m]\!]$, letting

$$D_{<j} \coloneqq \bigcup_{j'<j} \bigcup \mathcal{D}_{j'},$$

either

(i) each $D \in \mathcal{D}_j$ is disjoint from $D_{<j}$ or

(ii) each $D \in \mathcal{D}_j$ intersects $D_{<j}$ in exactly one element.

To see this, suppose that $1 \leqslant j \leqslant m$ and that sequences $i_1, \dots, i_{j-1}$ and $\mathcal{D}_1, \dots, \mathcal{D}_{j-1}$ with the above properties have already been defined. Let $D_{<j}$ be the set defined above and define the sets

$$\mathcal{X} \coloneqq \big\{B \setminus \{i\} : i \in B \in \mathrm{AP}_k \text{ such that } |B \cap D_{<j}| \geqslant 2\big\}$$

and

$$I^{\mathcal{X}} \coloneqq \big\{i \in I : |\mathcal{B}'_k(i) \cap \mathcal{X}| \geqslant u/2\big\}.$$

The key observation is that, for every $i \in I \setminus I^{\mathcal{X}}$, the family $\mathcal{B}'_k(i)$ contains at least $u/2$ subsets of $\mathbf{R}$ that intersect $D_{<j}$ in at most one element; in particular, either at least $u/4$ of them are disjoint from $D_{<j}$ or at least $u/4$ of them intersect it in exactly one element. Consequently, Fact 5.5 allows us to find a suitable $\mathcal{D}_j$ for each $i_j \in I \setminus I^{\mathcal{X}}$. Thus, it suffices to show that $I \setminus I^{\mathcal{X}}$ is not empty. To this end, note that $|\mathcal{X}| \leqslant \binom{|D_{<j}|}{2} \cdot \binom{k}{2} \cdot k \leqslant (m-1)^2 b^2 k^3$ and that $|\mathcal{X}| \geqslant |I^{\mathcal{X}}| \cdot u/6$, as each $B \subseteq [\![N]\!]$ belongs to $\mathcal{B}'_k(i)$ for at most three different $i \in [\![N]\!]$ (at most two different $i \in [\![N]\!]$ if $k \geqslant 4$). Thus,

$$|I^{\mathcal{X}}| \leqslant \frac{6(m-1)^2 b^2 k^3}{u} \leqslant (m-1)^2 \cdot u < |I|.$$

Now, define

$$m \coloneqq \left\lceil \frac{\sqrt{\varepsilon} t^{3/2}}{\sigma^2 p^{1/4}} \right\rceil$$

and let Z denote the number of pairs of sequences $(i_1, \dots, i_m)$ and $(\mathcal{D}_1, \dots, \mathcal{D}_m)$ as above, so that

$$\mathbb{P}\left(\left|\left\{i : L_i > \frac{\sigma^2}{t}\right\}\right| \geqslant \frac{\varepsilon t^2}{\sigma^2 p^{1/2}}\right) \leqslant \mathbb{P}\big(|I| > (m-1)^2 \cdot u\big) \leqslant \mathbb{P}(Z \geqslant 1) \leqslant \mathbb{E}[Z].$$

In order to estimate the expectation of Z, for any $J \subseteq [\![m]\!]$, let Z_J denote the expected number of such pairs of sequences for which (ii) above holds for $j \in J$ and (i) above holds for $j \notin J$. There are at most N^m choices for $(i_1, \dots, i_m)$ and at most $\binom{kN}{b}$ further choices for each $\mathcal{D}_j$, as each $\mathcal{D}_j$ is a b-element subset of the set $\mathcal{B}'_k(i_j)$, which has size at most kN. If $j \in J$, however, the number of choices for $\mathcal{D}_j$ after $\mathcal{D}_1, \dots, \mathcal{D}_{j-1}$ have already been chosen is at most

$$\binom{|D_{<j}| + b - 1}{b} \cdot \binom{k}{2}^b \leqslant \binom{mkb}{b} \cdot k^{2b} \leqslant \big(emk^3\big)^b.$$

Therefore,

$$\begin{aligned}
\mathbb{E}[Z_J] &\leqslant N^m \cdot \left(\binom{kN}{b} \cdot p^{(k-1)b}\right)^{m-|J|} \cdot \Big(\big(emk^3\big)^b \cdot p^{(k-2)b}\Big)^{|J|} \\
&\leqslant N^m \cdot \left(\frac{ekNp^{k-1}}{b}\right)^{(m-|J|)b} \cdot \big(emk^3 p^{k-2}\big)^{|J|b} \\
&= N^m \cdot \left(\frac{ekNp^{k-1}}{b}\right)^{mb} \cdot \left(\frac{mk^2 b}{Np}\right)^{|J|b}.
\end{aligned}$$

Since

$$mk^2 b \leqslant mu = \frac{m\sigma^2}{t} \leqslant \frac{\sqrt{t}}{p^{1/4}} \leqslant \frac{\sigma^{2/3}\big(\log(1/p)\big)^{1/3}}{p^{1/4}} \leqslant Np \cdot p^{(2k-4)/3 - 1/4}\big(\log(1/p)\big)^{1/3} \leqslant Np,$$

we may conclude that

$$\mathbb{E}[Z] = \sum_{J \subseteq [\![m]\!]} \mathbb{E}[Z_J] \leqslant (2N)^m \cdot \left(\frac{ekNp^{k-1}}{b}\right)^{mb} \leqslant (2N)^m \cdot \left(\frac{4ek^4Np^{k-1}}{u}\right)^{mu/(4k^3)}$$
$$\leqslant (2N)^m \cdot p^{mu/(16k^3)},$$

where the last inequality follows from (40) and the penultimate inequality holds because $b \geqslant u/(4k^3)$ and, when $a > 0$, the function $x \mapsto (ea/x)^x$ is decreasing on $[a, \infty)$.

We conclude that

$$\mathbb{P}\left(\left|\left\{i : L_i > \frac{\sigma^2}{t}\right\}\right| \geqslant \frac{\varepsilon t^2}{\sigma^2 p^{1/2}}\right) \leqslant N^{-m} \cdot \left((2N^2)^{1/u} \cdot p^{1/(16k^3)}\right)^{mu} \leqslant N^{-m} \cdot \exp(-mu),$$

where the final inequality follows since $u \geqslant N^{1/(4k-4)}$, again by (40) and our lower-bound assumption on p. Finally, since $m \geqslant 1$, (38) follows as

$$mu \geqslant \frac{\sqrt{\varepsilon t}}{2p^{1/4}} \geqslant \sqrt{t}\log(1/p) \geqslant t^2/\sigma^2. \qquad \square$$

The following key lemma will allow us to deduce the claimed upper bound on the upper-tail probability of $L_1^2 + \cdots + L_N^2$. As we remarked above, its proof is another adaptation of the elegant argument of [21].

Lemma 5.6. *Let $\beta > 0$ and $k \geqslant 3$, and assume that $Np^{k-1} > 1$. Then there exists an $\alpha > 0$ depending only on β and k such that, for every nonnegative integer $\ell \leqslant \alpha Np^{(2k-2)/3}$, letting $V \coloneqq L_1^2 + \cdots + L_N^2$, we have*

$$\mathbb{E}[V^\ell] \leqslant (1+\beta)^\ell \cdot \mathbb{E}[V]^\ell.$$

Proof. We prove the claimed estimate by induction on ℓ. The inequality holds vacuously when $\ell \leqslant 1$, so we may assume that $\ell \geqslant 2$. Define the multiset[2]

$$\mathcal{V}_k \coloneqq \bigcup_{i=1}^N \big\{B \cup B' : B, B' \in \mathcal{B}'_k(i)\big\} = \big\{(B \cup B') \setminus \{i\} : B, B' \in \mathrm{AP}_k, i \in B \cap B'\big\}$$

so that V counts (with multiplicities) sets in $\mathcal{V}_k$ that are contained in $\mathbf{R}$. We have

$$\mathbb{E}[V^\ell] = \sum_{P_1,\dots,P_\ell \in \mathcal{V}_k} \mathbb{P}\big(P_1 \cup \cdots \cup P_\ell \subseteq \mathbf{R}\big)$$
$$= \sum_{P_1,\dots,P_{\ell-1} \in \mathcal{V}_k} \mathbb{P}\big(P_1 \cup \cdots \cup P_{\ell-1} \subseteq \mathbf{R}\big) \cdot \sum_{P_\ell \in \mathcal{V}_k} \mathbb{P}\big(P_\ell \subseteq \mathbf{R} \mid P_1 \cup \cdots \cup P_{\ell-1} \subseteq \mathbf{R}\big).$$

Note that the first sum above equals $\mathbb{E}[V^{\ell-1}]$ and the second sum is $\mathbb{E}_{P_1 \cup \cdots \cup P_{\ell-1}}[V]$. As each $P \in \mathcal{V}_k$ has at most $2k-2$ elements, we may use our inductive assumption to conclude that

$$\mathbb{E}[V^\ell] \leqslant (1+\beta)^{\ell-1} \cdot \mathbb{E}[V]^{\ell-1} \cdot \max\big\{\mathbb{E}_U[V] : |U| \leqslant 2k\ell\big\}.$$

In particular, it suffices to show that the maximum above is at most $(1+\beta) \cdot \mathbb{E}[V]$, provided that $\ell \leqslant \alpha Np^{(2k-2)/3}$ for some small $\alpha = \alpha(\beta, k)$.

Claim 5.7. *There is a constant C depending only on k such that, for every $U \subseteq [\![N]\!]$,*

$$\mathbb{E}_U[V] - \mathbb{E}[V] \leqslant C \cdot \left(|U|^3 + |U|^2 Np^{k-1} + |U|N^2p^{2k-3}\right).$$

[2]The map $\bigcup_{i=1}^N \left(\mathcal{B}'_k(i)\right)^2 \ni (B, B') \mapsto B \cup B' \in \mathcal{V}_k$ is generally not injective. For example, $(\{1,2,3\} \cup \{3,4,5\}) \setminus \{3\} = (\{1,3,5\} \cup \{2,3,4\}) \setminus \{3\}$.

Proof. Observe that

$$\mathbb{E}_U[V] - \mathbb{E}[V] = \sum_{P \in \mathcal{V}_k} \left(p^{|P \setminus U|} - p^{|P|}\right) \leqslant \sum_{\substack{P \in \mathcal{V}_k \\ P \cap U \neq \varnothing}} p^{|P \setminus U|}.$$

For every $s \geqslant 0$, let $v_s(U)$ denote the number of $P \in \mathcal{V}_k$ with $|P \setminus U| = s$. Since every pair of elements of $[\![N]\!]$ lies in at most $\binom{k}{2}$ arithmetic progressions of length k, one can check that, for some constant C that depends only on k,

$$v_s(U) \leqslant \begin{cases} C|U|^3 & \text{if } s \leqslant k-3, \\ C|U|^3 + C|U|N & \text{if } s \leqslant k-2, \\ C|U|^2 N & \text{if } s \leqslant 2k-4, \\ C|U|N^2 & \text{if } s \leqslant 2k-3. \end{cases}$$

(The odd term $C|U|N$ corresponds to 'degenerate' P of the form $(B \cup B) \setminus \{i\}$, where $B \in \mathrm{AP}_k$ and $i \in B$, that intersect U in exactly one element.) We may thus conclude that

$$\mathbb{E}_U[V] - \mathbb{E}[V] \leqslant \sum_{s=0}^{2k-3} v_s(U) \cdot p^s \leqslant kC|U|^3 + C|U|Np^{k-2} + kC|U|^2 Np^{k-1} + C|U|N^2 p^{2k-3}.$$

The claim follows, since $Np^{k-2} \leqslant N^2 p^{2k-3}$. □

Finally, the claim implies that, for every $U \subseteq [\![N]\!]$ with $|U| \leqslant 2k\alpha Np^{(2k-2)/3}$,

$$\mathbb{E}_U[V] - \mathbb{E}[V] \leqslant (2k)^3 \alpha \cdot N^3 p^{2k-2} \cdot C \cdot \left(1 + p^{(k-1)/3} + p^{(2k-5)/3}\right) \leqslant \beta \cdot \mathbb{E}[V],$$

where the last inequality holds for all sufficiently small α, as $\mathbb{E}[V] \geqslant c_k N^3 p^{2k-2}$ for some positive c_k that depends only on k. □

Proof of (39) *in Proposition 5.4.* Observe first that, for every $i \in [\![N]\!]$,

$$\begin{aligned} \mathbb{E}[L_i^2] &= \sum_{B, B' \in \mathcal{B}'_k(i)} p^{|B \cup B'|} \leqslant A_1(i)^2 \cdot p^{2k-2} + \sum_{\substack{B, B' \in \mathcal{B}'_k(i) \\ B \cap B' \neq \varnothing}} p^{k-1} \\ &\leqslant A_1(i)^2 \cdot p^{2k-2} + A_1(i) \cdot (k-1)\binom{k}{2} p^{k-1} \leqslant A_1(i)^2 \cdot p^{2k-2} \cdot \left(1 + \frac{k^4}{Np^{k-1}}\right), \end{aligned}$$

where the finaly inequality hods as $A_1(i) \geqslant N/k$ for every i. Since $Np^{k-1} \geqslant C$, summing the above estimate over all $i \in [\![N]\!]$, and recalling the definition of $\mathcal{V}$, we obtain

$$\mathbb{E}[V] = \sum_{i=1}^{N} \mathbb{E}[L_i^2] \leqslant \frac{(1+\eta) \cdot \mathcal{V}}{p}$$

for every constant $\eta > 0$ and all sufficiently large values of C. By Lemma 5.1, we may conclude that $\sigma^2/p \geqslant (1-2\eta)\mathbb{E}[V]$. Let $\ell := \lfloor \alpha Np^{(2k-2)/3} \rfloor$. Let $\beta > 0$ be such that $1 + \beta = (1+\varepsilon/2)^{1/2}$. Then by Lemma 5.6 and Markov's inequality,

$$\mathbb{P}\left(V > (1+\varepsilon) \cdot \frac{\sigma^2}{p}\right) \leqslant \mathbb{P}\left(V > (1+\varepsilon/2) \cdot \mathbb{E}[V]\right) \leqslant \frac{\mathbb{E}[V^\ell]}{(1+\varepsilon/2)^\ell \cdot \mathbb{E}[V]^\ell} \leqslant \left(\frac{1}{1+\varepsilon/2}\right)^{\alpha Np^{(2k-2)/3}/2}.$$

Finally, as $p \geqslant CN^{-1/(k-1)}$, we have

$$Np^{(2k-2)/3} \geqslant CN^{1/3} \geqslant C \log N$$

and, as $\sigma^2 \log(1/p) \geqslant (Ct)^{3/2}$ and $p \leqslant 1/C$,

$$Np^{(2k-2)/3} \geqslant (\sigma^2/p)^{1/3} \geqslant (\sigma^2 \log(1/p))^{4/3}/\sigma^2 \geqslant Ct^2/\sigma^2.$$

From these, the claim follows. □

6. The Poisson regime

In this section, we prove Theorem 1.1 in the Poisson regime, i.e., for all sequences (p, t) satisfying

$$\Omega(N^{-2/k}) \leqslant p \ll N^{-1/(k-1)}, \quad t \gg \sigma, \quad \text{and} \quad \sqrt{t}\log(1/p) \gg \mu \cdot \mathrm{Po}(t/\mu);$$

Note that these assumptions ensure that $\sigma = (1+o(1))Np^{k/2}$ is bounded away from zero, so, in particular, we have $t \gg 1$.

As usual, the lower and the upper bounds on the upper-tail probability of X will be proved separately. In Section 6.1, we prove the lower bound by applying Proposition 2.4 to a probability measure induced by the union of the random set $\mathbf{R}$ and (approximately) t random k-APs. In Section 6.2, we prove the matching upper bound by showing that the t-th factorial moment of X, restricted to the event that $\mathbf{R}$ does not contain a small εt-seed, is approximately μ^t, the t-th factorial moment of a genuine Poisson random variable with mean μ.

We recall the definition of the Poisson rate function $\mathrm{Po} \coloneqq [0,\infty) \to [0,\infty)$:

$$\mathrm{Po}(x) \coloneqq \int_0^x \log(1+y)\,dy = (1+x)\log(1+x) - x. \tag{41}$$

One property of Po that we will use several times throughout this section is that

$$x\log(1+x) \leqslant 2\mathrm{Po}(x), \tag{42}$$

which follows easily from the defintion of Po and the inequality $\log(1+y) \geqslant (y/x)\cdot\log(1+x)$, which is valid for all $y \in [0,x]$.

6.1. Proof of the lower bound in the Poisson regime.

In this section, we will prove the lower bound for the Poisson regime:

Proposition 6.1. *Suppose that the sequence (p,t) satisfies*

$$\Omega(N^{-2/k}) \leqslant p \ll N^{-1/(k-1)}, \quad t \gg \sigma, \quad \textit{and} \quad \sqrt{t}\log(1/p) \gg \mu \cdot \mathrm{Po}(t/\mu).$$

Then, for any $\varepsilon > 0$ and N large enough

$$\log \mathbb{P}(X > \mu + t) > -(1+\varepsilon)\cdot\mu\cdot\mathrm{Po}(t/\mu).$$

We prove this proposition by applying the tilting argument to a measure $\hat{\mathbb{P}}$ which is not close to any product measure. Instead, the random set $\mathbf{R}$ will be the union of $[\![N]\!]_p$ and approximately t random k-term arithmetic progressions. To be more precise, for any positive integer u, let $\mathrm{AP}_k^{\underline{u}}$ be the set containing all sequences of u distinct elements of AP_k. We define $\mathbb{P}_u$ to be the measure corresponding to an independent sample from $[\![N]\!]_p$ and a uniformly chosen element from $\mathrm{AP}_k^{\underline{u}}$; that is, for any $S \subseteq [\![N]\!]$ and $(\mathbf{A}_1,\dots,\mathbf{A}_u) \in \mathrm{AP}_k^{\underline{u}}$,

$$\mathbb{P}_u(S,(\mathbf{A}_1,\dots,\mathbf{A}_u)) = \frac{p^{|S|}(1-p)^{N-|S|}}{|\mathrm{AP}_k^{\underline{u}}|}.$$

We now let $\hat{\mathbb{P}}_u$ be the marginal of $\mathbb{P}_u$ onto the union $S \cup \mathbf{A}_1 \cup \cdots \cup \mathbf{A}_u$. A straightforward computation shows that, for any $R \subseteq [\![N]\!]$,

$$\hat{\mathbb{P}}_u(\mathbf{R} = R) = \sum_{\substack{(A_1,\dots,A_u)\in \mathrm{AP}_k^u \\ A_1\cup\cdots\cup A_u \subseteq R}} \frac{1}{|\mathrm{AP}_k^u|} \cdot p^{|R\setminus(A_1\cup\cdots\cup A_u)|} \cdot (1-p)^{N-|R|}. \tag{43}$$

As before, our argument has two parts. First, we will show that the upper-tail event $\{X \geqslant \mu + t\}$ is very likely to occur under the measure $\hat{\mathbb{P}}_u$, for an appropriately chosen $u \approx t$, and thus the logarithmic upper-tail probability can be bounded from below by $-(1+o(1)) \cdot D_{\mathrm{KL}}(\hat{\mathbb{P}}_u \,\|\, \mathbb{P})$. Second, we will show that $D_{\mathrm{KL}}(\hat{\mathbb{P}}_u \,\|\, \mathbb{P})$ is close to $\mu \cdot \mathrm{Po}(t/\mu)$, which will be a fairly simple consequence of the fact that $\hat{\mathbb{E}}_u[X] \leqslant \mu + t + o(t)$, where $\hat{\mathbb{E}}_u$ is the expectation operator associated with $\hat{\mathbb{P}}_u$.

Lemma 6.2. *For every $\varepsilon > 0$ and $k \geqslant 3$, there exists $C > 0$ such that the following holds for all $p \in (N^{-2/k}/C, N^{-1/(k-1)}/C)$ and all $t \in (C^k\sqrt{\mu}, N^{1-1/(k-1)}/C)$. Letting*

$$u \coloneqq \lceil t + C\sqrt{\mu} \rceil, \tag{44}$$

we have

$$\hat{\mathbb{P}}_u(X \geqslant \mu + t) \geqslant 1 - \varepsilon \qquad \text{and} \qquad \hat{\mathbb{E}}_u[X] \leqslant \mu + (1+\varepsilon)u.$$

Proof. Let Z count the progressions among $\mathbf{A}_1, \dots, \mathbf{A}_u$ that are contained in S and let Y count the progressions contained in $S \cup \mathbf{A}_1 \cup \cdots \cup \mathbf{A}_u$ that are neither fully contained in S nor belong to the set $\{\mathbf{A}_1, \dots, \mathbf{A}_u\}$. It is straightforward to see that

$$A_k(S) + u - Z \leqslant X \leqslant A_k(S) + u + Y. \tag{45}$$

Recall that, under $\hat{\mathbb{P}}$, the random set $\mathbf{R}$ is the union of S and the $\mathbf{A}_i$. The first inequality in (45) and the union bound yield

$$\hat{\mathbb{P}}_u(X \leqslant \mu + t) \leqslant \mathbb{P}_u(A_k(S) + u - Z \leqslant \mu + t) \leqslant \mathbb{P}_u\big(A_k(S) \leqslant \mu - (u-t)/2\big) + \mathbb{P}_u\big(Z \geqslant (u-t)/2\big).$$

Since the marginal of S under $\mathbb{P}_u$ is $\mathbb{P}$, the variance of $A_k(S)$ under $\mathbb{P}_u$ is equal to σ^2, the variance of X under $\mathbb{P}$. Thus, we may use Chebyshev's inequality to conclude that

$$\mathbb{P}_u\big(A_k(S) \leqslant \mu - (t-u)/2\big) \leqslant \frac{4\sigma^2}{(t-u)^2} \leqslant \frac{8\mu}{C^2\mu},$$

where the final bound follows from the estimate $\sigma^2 \leqslant (1+\varepsilon)\mu$, which, in turn, is a consequence of our bounds on p. Meanwhile, since $\mathbb{E}_u[Z] = up^k$, where $\mathbb{E}_u$ is the expectation operator associated with $\mathbb{P}_u$, a straightforward application of Markov's inequality yields

$$\mathbb{P}_u(Z \geqslant (u-t)/2) \leqslant \frac{2\mathbb{E}_u[Z]}{C\sqrt{\mu}} \leqslant \frac{2up^k}{C\sqrt{N^2p^k/(2k)}} \leqslant \frac{8ktp^{k/2}}{CN} \leqslant \frac{8k}{C},$$

where we use the inequalities $u \leqslant 2t \leqslant N$, which follow from the definition of u and the assumption on t. Choosing C large enough yields the first assertion of the lemma.

In light of the second inequality in (45) and the fact that $\mathbb{E}_u[A_k(S)] = \mathbb{E}[X] = \mu$, the second assertion of the lemma follows from the inequality $\mathbb{E}_u[Y] \leqslant \varepsilon u$, which we will establish in the remainder of the proof. To this end, note that Y can be written as a sum, over all $B \in \mathrm{AP}_k$, of the indicators of the event $\mathcal{E}_B$ that $B \subseteq S \cup \mathbf{A}_1 \cup \cdots \cup \mathbf{A}_u$, but B is neither fully contained in S nor is it equal to any of the

$\mathbf{A}_i$. Since $|\mathrm{AP}_k| \leqslant \binom{N}{2}$, it will be sufficient to prove that $\mathbb{P}_u(\mathcal{E}_B) \leqslant \varepsilon u/N^2$ for every $B \in \mathrm{AP}_k$. For the remainder of the proof, we will fix an arbitrary progression $B \in \mathrm{AP}_k$ and write $\mathcal{E}$ in place of $\mathcal{E}_B$.

For each $j \in [\![u]\!]$, let $\mathcal{B}_j$ and $\mathcal{D}_j$ be the events that $B \cap \mathbf{A}_j \neq \varnothing$ and $|B \cap \mathbf{A}_j| \geqslant 2$, respectively. The crucial observation is that, in order for $\mathcal{E}$ to occur, at least one of the following must occur:

- $\mathcal{D}_j$ occurs for some j and either $B \cap S \neq \varnothing$ or $\mathcal{B}_\ell$ occurs for some $\ell \neq j$; or
- there is an $r \in \{1, \dots, k\}$ such that $|B \cap S| = k - r$ and $\mathcal{B}_j$ occurs for r distinct indices j.

We shall denote the first event by $\mathcal{F}$ and the second by $\mathcal{T}$. As the above observation implies that $\mathbb{P}_u(\mathcal{E}) \leqslant \mathbb{P}_u(\mathcal{F}) + \mathbb{P}_u(\mathcal{T})$, it sufficies to bound the probabilities of these two events.

Recall that $|\mathrm{AP}_k| = (1 + o(1)) \cdot N^2/(2(k-1))$ and that there are at most kN progressions in AP_k that contain any given element of $[\![N]\!]$ and at most $\binom{k}{2}$ such progressions that contain any given pair of elements of $[\![N]\!]$. As $|B \cap S|$ follows a binomial distribution $\mathrm{Bin}(k,p)$ and is independent of all the events $\{\mathcal{B}_j\}_j$ and $\{\mathcal{D}_j\}_j$, we find that

$$\begin{aligned}\mathbb{P}_u(\mathcal{F}) &\leqslant \sum_{j=1}^{u} \mathbb{P}_u\big(\mathcal{D}_j \cap \{B \cap S \neq \varnothing\}\big) + \sum_{\substack{j,\ell \in [\![u]\!] \\ j \neq \ell}} \mathbb{P}_u\big(\mathcal{D}_j \cap \mathcal{B}_\ell\big) \\ &\leqslant u \cdot \frac{\binom{k}{2}^2}{|\mathrm{AP}_k|} \cdot kp + u^2 \cdot \frac{\binom{k}{2}^2 \cdot k^2 N}{|\mathrm{AP}_k|(|\mathrm{AP}_k| - 1)} \leqslant \frac{k^6 up}{N^2} + \frac{2k^8 u^2}{N^3},\end{aligned}$$

provided that C is sufficiently large (and thus N is sufficiently large). The inequalities $u \leqslant 2t \leqslant 2N/C$ and $p < 1/C$ for a large constant C imply that $\mathbb{P}_u(\mathcal{F}) \leqslant \frac{\varepsilon u}{2N^2}$. Similarly,

$$\begin{aligned}\mathbb{P}_u(\mathcal{T}) &\leqslant \sum_{r=1}^{k} \mathbb{P}_u(|B \cap S| = k - r) \cdot \mathbb{P}_u(\mathcal{B}_j \text{ occurs for } r \text{ distinct } j \in [\![u]\!]) \\ &\leqslant \sum_{r=1}^{k} (kp)^{k-r} \binom{u}{r} \frac{(k^2 N)^{\underline{r}}}{|\mathrm{AP}_k|^{\underline{r}}} \leqslant (kp)^k \cdot \sum_{r=1}^{k} \left(\frac{2k^2 u}{Np}\right)^r.\end{aligned}$$

In order to bound the right-hand side of the above inequality, we consider two cases. If $2k^2 u \leqslant Np$, then

$$\mathbb{P}_u(\mathcal{T}) \leqslant p^k \cdot k^{k+1} \cdot \frac{2k^2 u}{Np} = 2k^{k+3} N p^{k-1} \cdot \frac{u}{N^2} \leqslant \frac{\varepsilon u}{2N^2},$$

by our assumption that $p \leqslant N^{-1/(k-1)}/C$. Otherwise, if $2k^2 u > Np$, then

$$\mathbb{P}_u(\mathcal{T}) \leqslant p^k \cdot k^{k+1} \cdot \left(\frac{2k^2 u}{Np}\right)^k = \frac{2^k k^{3k+1} u^{k-1}}{N^{k-2}} \cdot \frac{u}{N^2} \leqslant \frac{\varepsilon u}{2N^2},$$

by our assumption that $u \leqslant 2t \leqslant 2N^{1-1/(k-1)}/C$. □

Proof of Proposition 6.1. By Proposition 2.4 and Lemma 6.2, it is sufficient to show that, for every fixed $\varepsilon > 0$,

$$D_{\mathrm{KL}}(\hat{\mathbb{P}}_u \,\|\, \mathbb{P}) \leqslant (1 + 5\varepsilon)\mu \cdot \mathrm{Po}(u/\mu) \qquad \text{and} \qquad \mathrm{Po}(u/\mu) \leqslant (1+\varepsilon) \cdot \mathrm{Po}(t/\mu), \tag{46}$$

where u is the integer defined in (44). To this end, recall (43) and note that, for every $R \subseteq [\![N]\!]$,

$$\frac{\hat{\mathbb{P}}_u(\mathbf{R} = R)}{\mathbb{P}(\mathbf{R} = R)} = \frac{1}{|\mathrm{AP}_k^{\underline{u}}|} \sum_{\substack{(A_1,\dots,A_u) \in \mathrm{AP}_k^{\underline{u}} \\ A_1 \cup \cdots \cup A_u \subseteq R}} p^{-|A_1 \cup \cdots \cup A_u|} \leqslant \frac{A_k(R)^{\underline{u}}}{|\mathrm{AP}_k^{\underline{u}}| \cdot p^{ku}}. \tag{47}$$

Consequently,

$$D_{\mathrm{KL}}(\hat{\mathbb{P}}_u \,\|\, \mathbb{P}) = \sum_{R \subseteq [\![N]\!]} \hat{\mathbb{P}}_u(\mathbf{R}=R) \log \frac{\hat{\mathbb{P}}_u(\mathbf{R}=R)}{\mathbb{P}(\mathbf{R}=R)} \leqslant \hat{\mathbb{E}}_u[\log X^{\underline{u}}] - \log\left(|\mathrm{AP}_k^{\underline{u}}| \cdot p^{ku}\right).$$

A straightforward calculation shows that

$$|\mathrm{AP}_k^{\underline{u}}| \cdot p^{ku} = \mu^u \cdot \frac{|\mathrm{AP}_k|^{\underline{u}}}{|\mathrm{AP}_k|^u} \geqslant \mu^u \cdot \left(1 - \frac{u}{|\mathrm{AP}_k|}\right)^u.$$

Since $|\mathrm{AP}_k| \gg \mu$, we can conclude that

$$-\log\left(|\mathrm{AP}_k^{\underline{u}}| \cdot p^{ku}\right) \leqslant -u\log\mu + \frac{\varepsilon}{2} u \log\left(1+\frac{u}{\mu}\right) \leqslant -u\log\mu + \varepsilon\mu \cdot \mathrm{Po}\left(\frac{u}{\mu}\right),$$

where the last inequality holds by (42).

Further, since the function $[u,\infty) \ni x \mapsto \log x^{\underline{u}}$ is concave, Jensen's inequality and Lemma 6.2 imply that

$$\begin{aligned}\hat{\mathbb{E}}_u[\log X^{\underline{u}}] &\leqslant \log\left(\hat{\mathbb{E}}_u[X]\right)^{\underline{u}} \leqslant \log\left(\mu + (1+\varepsilon)u\right)^{\underline{u}} = \sum_{i=1}^{u} \log(\mu + \varepsilon u + i) \\ &\leqslant \int_{\mu}^{\mu+u} \log(x + \varepsilon u + 1)\, dx = \int_{\mu}^{\mu+u} \log x\, dx + \int_{\mu}^{\mu+u} \log\left(1 + \frac{\varepsilon u + 1}{x}\right) dx.\end{aligned}$$

It follows from (41) that $\mu \cdot \mathrm{Po}(u/\mu) = \int_{\mu}^{\mu+u} \log x\, dx - u\log\mu$; so we conclude that

$$D_{\mathrm{KL}}(\hat{\mathbb{P}}_u \,\|\, \mathbb{P}) \leqslant (1+\varepsilon)\mu \cdot \mathrm{Po}\left(\frac{u}{\mu}\right) + \int_{\mu}^{\mu+u} \log\left(1 + \frac{\varepsilon u + 1}{x}\right) dx. \tag{48}$$

Further, as $\log(1+y) \leqslant y$ for every $y > -1$ and $u \geqslant t \geqslant 1/\varepsilon$,

$$\int_{\mu}^{\mu+u} \log\left(1 + \frac{\varepsilon u + 1}{x}\right) dx \leqslant \int_{\mu}^{\mu+u} \frac{2\varepsilon u}{x}\, dx = 2\varepsilon u \cdot \log\left(1 + \frac{u}{\mu}\right) \leqslant 4\varepsilon\mu \cdot \mathrm{Po}\left(\frac{u}{\mu}\right),$$

where the last inequality is again (42). Substituting this estimate into (48) yields the first inequality in (46). Finally,

$$\mathrm{Po}\left(\frac{u}{\mu}\right) - \mathrm{Po}\left(\frac{t}{\mu}\right) = \int_{t/\mu}^{u/\mu} \log(1+x)\, dx \leqslant \frac{u-t}{\mu} \cdot \log\left(1 + \frac{u}{\mu}\right) \leqslant 2\left(1 - \frac{t}{u}\right) \cdot \mathrm{Po}\left(\frac{u}{\mu}\right),$$

where the last inequality follows from (42). Our assumption that $t \gg \sigma = \Omega(\sqrt{\mu})$ implies that $t/u \to 1$, giving the second inequality in (46) and completing the proof. □

6.2. **Proof of the upper bound in the Poisson regime.** In this section, we prove an upper bound on the upper-tail probability in the Poisson regime. Our starting point here is the realisation that it would be enough to establish the following estimate for all $\varepsilon > 0$:

$$\mathbb{E}[X^{\underline{t}}] \leqslant \mu^t \cdot \exp\left(\varepsilon t \log(1 + t/\mu)\right). \tag{49}$$

Indeed, if (49) were true, then a simple application of Markov's inequality would yield

$$\mathbb{P}(X \geqslant \mu + t) = \mathbb{P}(X^{\underline{t}} \geqslant (\mu+t)^{\underline{t}}) \leqslant \frac{\mathbb{E}[X^{\underline{t}}]}{(\mu+t)^{\underline{t}}} \leqslant \frac{\mu^t}{(\mu+t)^{\underline{t}}} \cdot \exp\left(\varepsilon t \log(1+t/\mu)\right),$$

which gives the desired estimate, as $t\log(1+t/\mu) \leqslant 2\mu \cdot \mathrm{Po}(t/\mu)$, see (42), whereas

$$\log\left(\frac{\mu^t}{(\mu+t)^{\underline{t}}}\right) \leqslant -\int_0^t \log\left(\frac{\mu+x}{\mu}\right) dx = -\mu \cdot \int_0^{t/\mu} \log(1+y)\, dy = -\mu \cdot \mathrm{Po}\left(\frac{t}{\mu}\right).$$

Unfortunately, the estimate (49) is not correct in the entirety of the Poisson regime. To remedy this, we will define a family of indicator random variables $\{Z_u\}_{u\in\mathbb{R}}$ such that, for all sequences (p,t) falling into the Poisson regime, we have

$$\mathbb{E}[X^{\underline{t}}\cdot Z_u] \leqslant \mu^t\cdot\exp\big((u+\varepsilon t)\log(1+Kt/\mu)\big)$$

and

$$\mathbb{P}(Z_u=0)\leqslant\exp\big(-(1-\varepsilon)\Psi^*\big((1-\varepsilon)u\big)\log(1/p)\big)\,,$$

where $K=K(k)$ is some constant and Ψ^* was defined in (16).

We now set $u\coloneqq\varepsilon t$. Then $\Psi^*((1-\varepsilon)u)\geqslant\sqrt{(1-\varepsilon)u}$ holds by Proposition 3.1; the assumptions of that proposition apply since we assume $t\gg\sigma=\Omega(Np^{k/2})\geqslant\max\{1,N^2p^{2k-2}\}$. Therefore, for some positive $c=c(\varepsilon)$,

$$\begin{aligned}\mathbb{P}(X\geqslant\mu+t)&\leqslant\mathbb{P}(X\cdot Z_{\varepsilon t}\geqslant\mu+t)+\mathbb{P}(Z_{\varepsilon t}=0)\\&\leqslant\frac{\mathbb{E}[X^{\underline{t}}\cdot Z_{\varepsilon t}]}{(\mu+t)^{\underline{t}}}+\exp\big(-\Psi^*(\varepsilon t/2)\log(1/p)\big)\\&\leqslant\frac{\mu^t}{(\mu+t)^{\underline{t}}}\cdot\exp\big(2\varepsilon t\log(1+Kt/\mu)\big)+\exp\Big(-c\sqrt{t}\log(1/p)\Big)\\&\leqslant\exp\big(-\mu\cdot\mathrm{Po}(t/\mu)+2K\varepsilon t\log(1+t/\mu)\big)+\exp\Big(-c\sqrt{t}\log(1/p)\Big)\,.\end{aligned}$$

which gives the desired bound as $t\log(1+t/\mu)\leqslant 2\mu\cdot\mathrm{Po}(t/\mu)$, by (42), and $\sqrt{t}\log(1/p)\gg\mu\cdot\mathrm{Po}(t/\mu)$ across the Poisson regime.

We now define $Z_u=Z_u(C)$ to be the indicator of the event that $\mathbf{R}$ does not contain any set in $\mathcal{S}_{\mathrm{small}}(u,C)$, where $\mathcal{S}_{\mathrm{small}}$ is defined as in the statement of Proposition 3.2. The estimate

$$\mathbb{P}(Z_u=0)\leqslant\exp\big(-(1-\varepsilon)\Psi^*\big((1-\varepsilon)u\big)\log(1/p)\big)$$

is then immediate from Proposition 3.2, provided C is chosen to be sufficiently large.

It remains to prove the following proposition, which is the main business of this section.

Proposition 6.3. *For every $k\geqslant 3$, there exists a constant K such that the following holds for all $C,\varepsilon>0$. If $\Omega(N^{-2/k})\leqslant p\ll N^{-1/(k-1)}$ and t is a positive integer satisfying $t\leqslant\big(\mu\log(1/p)\big)^{2/3}+\big(\log(1/p)\big)^3$, then, for all sufficiently large N and all $u\leqslant t$,*

$$\mathbb{E}[X^{\underline{t}}\cdot Z_u(C)]\leqslant\mu^t\cdot\exp\big((u+\varepsilon t)\log(1+Kt/\mu)\big).$$

We remark that the upper-bound assumption on t in the proposition indeed holds in the entire Poisson regime, due to our assumption $\sqrt{t}\log(1/p)\gg\mu\mathrm{Po}(t/\mu)$. To see this, it helps to distinguish between the cases $t\leqslant\mu/2$ and $t>\mu/2$. In the first case, the inequality $\log(1+x)\geqslant x-x^2/2$ rather easily implies that $\mathrm{Po}(t/\mu)\geqslant t^2/4\mu^2$, and so our assumption results in $t\ll(\mu\log(1/p))^{2/3}$. In the second case, we can use (42) to get $\mu\mathrm{Po}(t/\mu)\geqslant\frac{1}{2}t(1+\log(t/\mu))>t/100$ and similarly obtain $t\ll(\log(1/p))^2$.

We now continue with the proof of Proposition 6.3. Since $X^{\underline{t}}$ is a sum of indicators of the events $\{A_1\cup\cdots\cup A_t\subseteq\mathbf{R}\}$ over all ordered sequences $(A_1,\dots,A_t)$ of t arithmetic progressions of length k, and

such an event precludes the event $\{Z_u = 1\}$ when the union $A_1 \cup \cdots \cup A_t$ contains a set from $\mathcal{S}_{\mathrm{small}}(u)$, we have

$$\mathbb{E}[X^{\underline{t}} \cdot Z_u] \leqslant \sum_{\substack{(A_1,\dots,A_t)\in \mathrm{AP}_k^t \\ \mathcal{P}(A_1\cup\cdots\cup A_t)\cap \mathcal{S}_{\mathrm{small}}(u)=\varnothing}} \mathbb{P}\big(A_1 \cup \cdots \cup A_t \subseteq \mathbf{R}\big). \tag{50}$$

In order to estimate the right-hand side of (50), we will analyse the overlap structure of progressions in each sequence in $\mathrm{AP}_k^{\underline{t}}$. It is convenient to introduce the notation

$$\psi(U) \coloneqq \mathbb{E}_U[X] - \mathbb{E}[X] = \sum_{r=1}^{k} A_r^{(k)}(U) \cdot (p^{k-r} - p^k),$$

where, as before, $A_r^{(k)}(U)$ is the number of k-APs in $[\![N]\!]$ that intersect U in precisely r elements. The following superadditivity property will turn out to be useful.

Lemma 6.4. *Suppose that $U_1, \dots, U_j$ are pairwise disjoint subsets of $[\![N]\!]$. Then*

$$\psi(U_1 \cup \cdots \cup U_j) \geqslant \psi(U_1) + \cdots + \psi(U_j).$$

Proof. It is enough to prove the claim for two disjoint sets U_1 and U_2. To this end, consider the conditional expectation $Y_1 \coloneqq \mathbb{E}[A_k(\mathbf{R} \cup U_1) - A_k(\mathbf{R}) \mid \mathbf{R} \setminus U_1]$ and note that Y_1 is an increasing random variable on a product probability space with coordinates $[\![N]\!] \setminus U_1$. So $\mathbb{E}[Y_1] \leqslant \mathbb{E}_{U_2}[Y_1]$ by Harris' inequality. To complete the proof, observe that $\mathbb{E}[Y_1] = \mathbb{E}_{U_1}[X] - \mathbb{E}[X] = \psi(U_1)$ and $\mathbb{E}_{U_2}[Y_1] = \mathbb{E}_{U_1\cup U_2}[X] - \mathbb{E}_{U_2}[X] = \psi(U_1 \cup U_2) - \psi(U_2)$. □

Definition. A *cluster* is a set $\mathfrak{C} \subseteq \mathrm{AP}_k$ that is connected when viewed as a hypergraph. We say that a cluster is:

- *small* if
$$|\mathfrak{C}| \leqslant 2\big(\log(1/p)\big)^3 \quad \text{and} \quad |\mathfrak{C}| \leqslant \frac{\varepsilon \log(1/p)}{\log\log(1/p)} \cdot |\bigcup \mathfrak{C}|;$$
- *L-bounded*, for a given positive real L, if
$$\psi\left(\bigcup \mathfrak{C}\right) \leqslant \frac{L\mu|\mathfrak{C}|}{Kt},$$
where $K > 0$ is a sufficiently large constant depending only on k.
- *heavy* if it is neither small nor 1-bounded.

For a sequence $\mathbf{A} = (A_1, \dots, A_t) \in \mathrm{AP}_k^t$, we will denote by $\mathfrak{B}(\mathbf{A})$ the collection of all heavy maximal[3] clusters in $\{A_1, \dots, A_t\}$ and let $U(\mathbf{A}) \coloneqq \bigcup\bigcup \mathfrak{B}(\mathbf{A})$ be the union of all progressions that belong to some (heavy, maximal) cluster in $\mathfrak{B}(\mathbf{A})$. The maximality ensures that the family $\{\bigcup \mathfrak{C}\}_{\mathfrak{C}\in\mathfrak{B}(\mathbf{A})}$ is a partition of $U(\mathbf{A})$ for every $\mathbf{A} \in \mathrm{AP}_k^t$, so Lemma 6.4 implies the following.

Corollary 6.5. *For every $\mathbf{A} \in \mathrm{AP}_k^t$, we have*

$$\psi(U(\mathbf{A})) \geqslant \sum_{\mathfrak{C}\in\mathfrak{B}(\mathbf{A})} \psi(\bigcup \mathfrak{C}).$$

Our next lemma establishes a key property of heavy maximal clusters.

[3]With respect to the subset relation. Such maximal clusters correspond to connnected components of the hypergraph spanned by $A_1, \dots, A_t$.

Lemma 6.6. *For every $\mathbf{A} \in \mathrm{AP}_k^t$, we have $U(\mathbf{A}) \in \mathcal{S}_{\mathrm{small}}(\psi(U(\mathbf{A})))$.*

Proof. For brevity, let us write $U \coloneqq U(\mathbf{A})$ and $\mathfrak{B} \coloneqq \mathfrak{B}(\mathbf{A})$. We can assume without loss of generality that U and $\mathfrak{B}$ are nonempty. Observe that U is a $\psi(U)$-seed by definition, so it only remains to show that it also satisfies the size constraint from the definition of $\mathcal{S}_{\mathrm{small}}(\psi(U))$, namely,

$$\psi(U) \geqslant C|U| \cdot \max\{1, Np^{k-1}, |U|p^{k-2} \cdot N^{(k-2)(|U|/\psi(U))^{1/(k-1)}}\}.$$

Due to our assumption $Np^{k-1} \ll 1$, which also implies the inequality $N^{(k-2)(|U|/\psi(U))^{1/(k-1)}} \leqslant N^{(k-2)C^{-1/(k-1)}} \leqslant p^{-1/4}$ when $|U| \leqslant \Psi(U)/C$ for a large enough C, it is in fact enough to show that

$$\psi(U) \geqslant C|U| \cdot \max\{1, |U|p^{k-9/4}\}.$$

Since U is the union of at most t progressions, we have $|U| \leqslant kt \leqslant k \cdot \big(\mu \log(1/p)\big)^{2/3} + k \cdot \big(\log(1/p)\big)^3$. Further,

$$\mu^{2/3} \leqslant (N^2p^k)^{2/3} = \big(Np^{k-1}\big)^{4/3} \cdot p^{(4-2k)/3} \leqslant p^{(4-2k)/3} \leqslant p^{7/3-k},$$

where the penultimate inequality follows from our assumption that $Np^{k-1} \leqslant 1$. Since p vanishes, we may conclude that $|U|p^{k-9/4} \leqslant 1$ for all sufficiently large N, and therefore it only remains to show that $\psi(U) \geqslant C|U|$.

To this end, note first that Corollary 6.5 implies that

$$\frac{\psi(U)}{|U|} \geqslant \frac{\sum_{\mathfrak{C} \in \mathfrak{B}} \psi(\bigcup \mathfrak{C})}{\sum_{\mathfrak{C} \in \mathfrak{B}} |\bigcup \mathfrak{C}|} \geqslant \min_{\mathfrak{C} \in \mathfrak{B}} \frac{\psi(\bigcup \mathfrak{C})}{|\bigcup \mathfrak{C}|}.$$

Let $\mathfrak{C} \in \mathfrak{B}$ be a cluster that realises this minimum. Suppose first that $|\mathfrak{C}| \leqslant 2\big(\log(1/p)\big)^3$. Since $\mathfrak{C}$ is heavy, we must have $|\mathfrak{C}|/|\bigcup \mathfrak{C}| > \frac{\varepsilon \log(1/p)}{\log\log(1/p)}$. Moreover, every progression in $\mathfrak{C}$ contributes $1 - p^k$ to $\psi(\bigcup \mathfrak{C})$, so we may conclude that

$$\frac{\psi(\bigcup \mathfrak{C})}{|\bigcup \mathfrak{C}|} \geqslant \frac{|\mathfrak{C}| \cdot (1 - p^k)}{|\bigcup \mathfrak{C}|} \geqslant \frac{\varepsilon \log(1/p)}{\log\log(1/p)} \cdot (1 - p^k) \geqslant C$$

for all large enough N. Suppose now that $|\mathfrak{C}| > 2\big(\log(1/p)\big)^3$. In this case, we have

$$2\log(1/p)^3 < |\mathfrak{C}| \leqslant t \leqslant \big(\mu \log(1/p)\big)^{2/3} + \big(\log(1/p)\big)^3,$$

which also implies that $\mu/t \geqslant \sqrt{t}/(3\log(1/p)) \geqslant \sqrt{\log(1/p)/3}$. Finally, since $|\bigcup \mathfrak{C}| \leqslant k|\mathfrak{C}|$ and $\mathfrak{C}$ is heavy, and thus not 1-bounded, we have

$$\frac{\psi(\bigcup \mathfrak{C})}{|\bigcup \mathfrak{C}|} \geqslant \frac{\psi(\bigcup \mathfrak{C})}{k|\mathfrak{C}|} \geqslant \frac{\mu}{Kkt} \geqslant \frac{\sqrt{\log(1/p)}}{\sqrt{3}Kk} \geqslant C$$

for all N large enough. □

In view of Lemma 6.6, we may conclude from (50) that

$$\mathbb{E}[X^t \cdot Z_u] \leqslant \sum_{\substack{\mathbf{A}=(A_1,\dots,A_t) \in \mathrm{AP}_k^t \\ \psi(U(\mathbf{A}))<u}} \mathbb{P}\big(A_1 \cup \dots \cup A_t \subseteq \mathbf{R}\big). \tag{51}$$

Indeed, if $A_1 \cup \dots \cup A_t$ does not contain any subset in $\mathcal{S}_{\mathrm{small}}(u)$, then, in particular, $U(\mathbf{A}) \notin \mathcal{S}_{\mathrm{small}}(u)$. Thus, either $U(\mathbf{A})$ is not a u-seed at all (but as it is a $\psi(U(\mathbf{A}))$-seed automatically, we then must have $u > \psi(U(\mathbf{A}))$) or it is a u-seed that is too large to be a member of $\mathcal{S}_{\mathrm{small}}(u)$: since $U(\mathbf{A}) \in \mathcal{S}_{\mathrm{small}}(\psi(U(\mathbf{A})))$ by the lemma, this implies $u > \psi(U(\mathbf{A}))$ as well.

In order to estimate the right-hand side of (51), we will first partition the set of all sequences $(A_1, \dots, A_t) \in \mathrm{AP}_k^t$ according to simple statistics of the maximal clusters in $\{A_1, \dots, A_t\}$. To do so, note first that it follows from Lemma 3.3 that $|\bigcup \mathfrak{C}| \geqslant |\mathfrak{C}|^{1/2}$ for every cluster $\mathfrak{C}$. In particular, any cluster $\mathfrak{C}$ satisfying

$$|\mathfrak{C}| \leqslant \left(\frac{\varepsilon \log(1/p)}{\log\log(1/p)} \right)^2 =: s_0$$

is automatically small (and thus cannot be heavy). Additionally, set $\xi := 1 + \varepsilon/15$, and define the *weight* of a heavy cluster $\mathfrak{C}$ to be the smallest integer ℓ such that $\mathfrak{C}$ is ξ^ℓ-bounded; note that the weight of a heavy cluster is necessarily positive, as heavy clusters are not 1-bounded.

Definition. Let $\mathcal{L}$ denote the family of all triples of integer sequences $(c_1, \dots, c_t)$, $(b_1, \dots, b_t)$, and $(\ell_{s,j} : s \in [\![t]\!], j \in [\![b_s]\!])$ satisfying:

(i) $0 \leqslant b_s \leqslant c_s$ and $b_s = 0$ when $s \leqslant s_0$;

(ii) $\ell_{s,1} \geqslant \cdots \geqslant \ell_{s,b_s} \geqslant 1$ for all s and $1 \leqslant j \leqslant b_s$; and

(iii) $c_1 + 2c_2 + \cdots + tc_t = t$.

Given a triple $(\mathbf{c}, \mathbf{b}, \mathbf{l}) \in \mathcal{L}$, denote by $\mathcal{C}(\mathbf{c}, \mathbf{b}, \mathbf{l})$ the family of all sequences $(A_1, \dots, A_t) \in \mathrm{AP}_k^t$ such that $\{A_1, \dots, A_t\}$ has, for every $s \in [\![t]\!]$, exactly c_s maximal clusters containing s progressions, out of which b_s are heavy clusters with weights $\ell_{s,1}, \dots, \ell_{s,b_s}$.

The following two key lemmas, whose proofs we postpone to the end of the section, will be used to bound from above the contribution of sequences from each collection $\mathcal{C}(\mathbf{c}, \mathbf{b}, \mathbf{l})$ to the right-hand side of (51). For every $s \geqslant 1$ and real L, let $\mathcal{D}_s$ denote the collection of all s-elements clusters that are small and let $\mathcal{D}_{s,L}$ denote the collection of s-element clusters that are not small, but L-bounded.

Lemma 6.7. *For every $s \geqslant 2$, we have*

$$D_s := \sum_{\mathfrak{C} \in \mathcal{D}_s} \mathbb{P}\left(\bigcup \mathfrak{C} \subseteq \mathbf{R} \right) \leqslant \frac{\mu^s}{t^s} \cdot \frac{\varepsilon t \log(1 + t/\mu)}{4^{s+1}}.$$

Lemma 6.8. *For every $s \geqslant 2$ and $L \geqslant 1$, we have*

$$D_{s,L} := \sum_{\mathfrak{C} \in \mathcal{D}_{s,L}} \mathbb{P}\left(\bigcup \mathfrak{C} \subseteq \mathbf{R} \right) \leqslant \frac{\mu^s}{t^s} \cdot L^s \cdot \frac{\varepsilon t \log(1 + t/\mu)}{4^{s+1}}.$$

Corollary 6.9. *For every $(\mathbf{c}, \mathbf{b}, \mathbf{l}) \in \mathcal{L}$, we have*

$$\Sigma(\mathbf{c}, \mathbf{b}, \mathbf{l}) := \sum_{(A_1, \dots, A_t) \in \mathcal{C}(\mathbf{c}, \mathbf{b}, \mathbf{l})} \mathbb{P}\big(A_1 \cup \cdots \cup A_t \subseteq \mathbf{R}\big) \leqslant \mu^t \cdot \prod_{s \geqslant 2} \left(\frac{1}{c_s!} \left(\frac{\varepsilon t \log(1 + t/\mu)}{2^s} \right)^{c_s} \cdot \prod_{j=1}^{b_s} \left(\xi^{\ell_{s,j}} \right)^s \right).$$

Proof. Fix an arbitrary sequence $(A_1, \dots, A_t) \in \mathcal{C}(\mathbf{c}, \mathbf{b}, \mathbf{l})$ and let $\mathfrak{C}_1, \dots, \mathfrak{C}_j$ be the maximal clusters in $\{A_1, \dots, A_t\}$. Since the sets $\bigcup \mathfrak{C}_1, \dots, \bigcup \mathfrak{C}_j$ partition $A_1 \cup \cdots \cup A_t$, we have

$$\mathbb{P}\big(A_1 \cup \cdots \cup A_t \subseteq \mathbf{R}\big) = \prod_{i=1}^{j} \mathbb{P}\left(\bigcup \mathfrak{C}_i \subseteq \mathbf{R} \right).$$

Further, for every collection $\{\mathfrak{C}_1, \dots, \mathfrak{C}_j\}$ of distinct clusters satisfying $|\mathfrak{C}_1| + \cdots + |\mathfrak{C}_j| = t$, there are exactly $t!$ sequences $(A_1, \dots, A_t) \in \mathrm{AP}_k^t$ with these clusters. Since non-heavy clusters of s progressions

belong to the set $\mathcal{D}_s \cup \mathcal{D}_{s,1}$ and heavy clusters with weight ℓ belong to the set $\mathcal{D}_{s,\xi^\ell}$, we have

$$\Sigma(\mathbf{c},\mathbf{b},\mathbf{l}) \leqslant \frac{1}{c_1!}\left(\sum_{A\in \mathrm{AP}_k}\mathbb{P}(A\subseteq \mathbf{R})\right)^{c_1}\cdot\prod_{s\geqslant 2}\left(\frac{\big(D_s+D_{s,1}\big)^{c_s-b_s}}{(c_s-b_s)!}\cdot\prod_{j=1}^{b_s}D_{s,\xi^{\ell_{s,j}}}\right)\cdot t!$$
$$\leqslant \mu^{c_1}\cdot\prod_{s\geqslant 2}\left(\frac{\big(D_s+D_{s,1}\big)^{c_s-b_s}\cdot c_s^{b_s}}{c_s!}\cdot\prod_{j=1}^{b_s}D_{s,\xi^{\ell_{s,j}}}\right)\cdot t^{t-c_1}.$$

By Lemmas 6.7 and 6.8, we have

$$D_s + D_{s,1} \leqslant \frac{\mu^s}{t^s}\cdot\frac{\varepsilon t\log(1+t/\mu)}{4^s}$$

and

$$D_{s,\xi^{\ell_{s,j}}} \leqslant \frac{\mu^s}{t^s}\cdot\xi^{s\ell_{s,j}}\cdot\frac{\varepsilon t\log(1+t/\mu)}{4^s}.$$

Using the fact that $t-c_1=\sum_{s\geqslant 2}sc_s$, we may conclude that

$$\Sigma(\mathbf{c},\mathbf{b},\mathbf{l}) \leqslant \mu^t\cdot\prod_{s\geqslant 2}\left(\frac{c_s^{b_s}}{c_s!}\cdot\left(\frac{\varepsilon t\log(1+t/\mu)}{4^s}\right)^{c_s}\cdot\prod_{j=1}^{b_s}\big(\xi^{\ell_{s,j}}\big)^s\right).$$

Finally, $b_s=0$ unless $s\geqslant s_0\gg\log t$. Since also $b_s\leqslant c_s\leqslant t$, it follows that $c_s^{b_s}\leqslant t^{c_s}\leqslant 2^{sc_s}$ for all s, which gives the desired inequality. □

The following lemma supplies a lower bound on $\psi(U(\mathbf{A}))$, valid for all $\mathbf{A}\in\mathcal{C}(\mathbf{c},\mathbf{b},\mathbf{l})$, that features the (logarithm of the) expression appearing in the upper bound on $\Sigma(\mathbf{c},\mathbf{b},\mathbf{l})$ given in Corollary 6.9.

Lemma 6.10. *If $\mathbf{A}\in\mathcal{C}(\mathbf{c},\mathbf{b},\mathbf{l})$ for some $(\mathbf{c},\mathbf{b},\mathbf{l})\in\mathcal{L}$, then*

$$\psi(U(\mathbf{A})) \geqslant \frac{1}{\xi^2\log(1+Kt/\mu)}\cdot\sum_{s=2}^{t}\sum_{j=1}^{b_s}s\ell_{s,j}\log\xi.$$

Proof. Fix some $(\mathbf{c},\mathbf{b},\mathbf{l})\in\mathcal{L}$ and let $\mathbf{A}$ be an arbitrary sequence in $\mathcal{C}(\mathbf{c},\mathbf{b},\mathbf{l})$. Since a heavy cluster with weight ℓ is not $\xi^{\ell-1}$-bounded, Corollary 6.5 gives (writing $\ell(\mathfrak{C})$ for the weight of a cluster $\mathfrak{C}$)

$$\psi(U(\mathbf{A})) \geqslant \sum_{\mathfrak{C}\in\mathfrak{B}(\mathbf{A})}\psi\Big(\bigcup\mathfrak{C}\Big) \geqslant \sum_{\mathfrak{C}\in\mathfrak{B}(\mathbf{A})}\frac{\xi^{\ell(\mathfrak{C})-1}\mu|\mathfrak{C}|}{Kt} = \sum_{s=2}^{t}\sum_{j=1}^{b_s}\frac{\xi^{\ell_{s,j}-1}\mu s}{Kt}.$$

Since the weight $\ell(\mathfrak{C})$ of every heavy cluster $\mathfrak{C}$ is positive and satisfies

$$|\mathfrak{C}|\cdot(1-p^k) \leqslant \psi\Big(\bigcup\mathfrak{C}\Big) \leqslant \frac{\xi^{\ell(\mathfrak{C})}\mu|\mathfrak{C}|}{Kt},$$

the asserted inequality will follow once we show that

$$\rho := \max\left\{\frac{Kt\ell\log\xi}{\xi^{\ell-1}\mu} : \ell\geqslant 1 \text{ and } \frac{\xi^\ell\mu}{Kt}\geqslant 1-p^k\right\} \leqslant \xi^2\log(1+Kt/\mu).$$

Since the function $\ell\mapsto\frac{Kt\ell\log\xi}{\xi^{\ell-1}\mu}$ is decreasing for $\ell\geqslant\log\xi$, we have $\rho\leqslant Kt(\log\xi)^2/(\xi^{\log\xi-1}\mu)$. Since we can assume without loss of generality that ε is rather small, and $\xi\leqslant 1+\varepsilon$, this implies, say, $\rho\leqslant Kt/(2\mu)$. In particular, if $Kt/\mu\leqslant 2$, then $\rho\leqslant\log(1+Kt/\mu)$. On the other hand, if $Kt/\mu>2$, then the smallest real solution ℓ_0 to the constraint $\xi^\ell\mu/(Kt)\geqslant 1-p^k$ satisfies $\ell_0=\log(Kt(1-p^k)/\mu)/(\log\xi)\geqslant 1\geqslant\log\xi$ and therefore the maximum in the definition of ρ is achieved at $\ell=\lceil\ell_0\rceil$. Thus,

$$\rho \leqslant \frac{\xi\log(Kt(1-p^k)/\mu)}{1-p^k} \leqslant \xi^2\log(1+Kt/\mu).$$ □

The upshot of Lemma 6.10 is that each sequence $\mathbf{A}$ that appears in the right-hand side of (51) belongs to the family $\mathcal{C}(\mathbf{c},\mathbf{b},\mathbf{l})$ for some triple $(\mathbf{c},\mathbf{b},\mathbf{l})$ from the set

$$\mathcal{L}_u \coloneqq \left\{ (\mathbf{c},\mathbf{b},\mathbf{l}) : \sum_{s\geqslant 2}\sum_{j=1}^{b_s} s\ell_{s,j}\log\xi \leqslant \xi^2 u\log(1+Kt/\mu) \right\}.$$

Therefore, by Corollary 6.9,

$$\begin{aligned}\frac{\mathbb{E}[X^{\underline{t}}\cdot Z_u]}{\mu^t} &\leqslant \sum_{(\mathbf{c},\mathbf{b},\mathbf{l})\in\mathcal{L}_u} \frac{\Sigma(\mathbf{c},\mathbf{b},\mathbf{l})}{\mu^t} \\ &\leqslant \underbrace{\sum_{(\mathbf{c},\mathbf{b},\mathbf{l})\in\mathcal{L}_u}\prod_{s\geqslant 2}\left(\frac{1}{c_s!}\left(\frac{\varepsilon t\log(1+t/\mu)}{2^s}\right)^{c_s}\right)}_{E}\cdot\exp\left(\xi^2 u\log(1+Kt/\mu)\right),\end{aligned}$$

where

$$\begin{aligned}E &\leqslant \prod_{s\geqslant 2}\left(\sum_{c=0}^{\infty}\frac{1}{c!}\left(\frac{\varepsilon t\log(1+t/\mu)}{2^s}\right)^c\right)\cdot\max_{\mathbf{c}}|\{(\mathbf{b},\mathbf{l}) : (\mathbf{c},\mathbf{b},\mathbf{l})\in\mathcal{L}_u\}| \\ &= \exp\left(\frac{\varepsilon t\log(1+t/\mu)}{2}\right)\cdot\max_{\mathbf{c}}|\{(\mathbf{b},\mathbf{l}) : (\mathbf{c},\mathbf{b},\mathbf{l})\in\mathcal{L}_u\}|.\end{aligned}$$

Finally, we bound the number of pairs $(\mathbf{b},\mathbf{l})$ such that $(\mathbf{c},\mathbf{b},\mathbf{l})\in\mathcal{L}_u$ for a given sequence $\mathbf{c}$. To this end, let π denote the (number-theoretic) partition function, for which we will only need the very crude bound $\pi(n)\leqslant e^n$. Note first that, for a given $\mathbf{L}=(L_1,\dots,L_t)$, the number of pairs $(\mathbf{b},\mathbf{l})$ such that $\ell_{s,1}+\dots+\ell_{s,b_s}=L_s$ for every s is at most $\prod_s\pi(L_s)\leqslant\exp\left(\sum_s L_s\right)$. Second, if $(\mathbf{c},\mathbf{b},\mathbf{l})\in\mathcal{L}_u$ for some such pair, then

$$\sum_{s\geqslant 1}L_s=\sum_{s\geqslant 1}\sum_{j=1}^{b_s}\ell_{s,j}=\sum_{s>s_0}\sum_{j=1}^{b_s}\ell_{s,j}\leqslant\frac{1}{s_0}\sum_{s>s_0}\sum_{j=1}^{b_s}s\ell_{s,j}\leqslant\frac{\xi^2 u\log(1+Kt/\mu)}{s_0\log\xi}\ll\frac{t\log(1+Kt/\mu)}{\log t},$$

where the last inequality follows from the assumption that $u\leqslant t$ and the fact that $s_0\gg\log(1/p)\geqslant\Omega(\log N)\geqslant\Omega(\log t)$. In particular, each L_s is at most t, and it follows that the number S of admissible sequences $\mathbf{L}$ satisfies

$$S\leqslant t^{\frac{\varepsilon t\log(1+Kt/\mu)}{8\log t}}=\exp\left(\frac{\varepsilon t\log(1+Kt/\mu)}{8}\right).$$

Consequently $E\leqslant\exp\big((2\varepsilon/3)\cdot t\log(1+Kt/\mu)\big)$. Since $\xi^2\leqslant 1+\varepsilon/6$, we may finally conclude that

$$\mathbb{E}[X^{\underline{t}}\cdot Z_u]\leqslant\mu^t\cdot\exp\big((u+\varepsilon t)\log(1+Kt/\mu)\big).$$

6.2.1. *Counting clusters.* Both Lemmas 6.7 and 6.8 will be derived from a more general upper bound on the number of clusters $\mathfrak{C}$ expressed in terms of the number of k-term arithmetic progressions that $\bigcup\mathfrak{C}$ is allowed to intersect. More precisely, for a vector $\mathbf{a}=(a_1,\dots,a_k)$, let let $\mathcal{C}_{m,s}(\mathbf{a})$ denote the set of all clusters $\mathfrak{C}$ with $|\mathfrak{C}|=s$, $|\bigcup\mathfrak{C}|=m$, and

$$A_{\geqslant i}\left(\bigcup\mathfrak{C}\right)\leqslant a_i \text{ for each } i\in[\![k]\!],$$

where we use the notation $A_{\geqslant i}(U)\coloneqq A_i^{(k)}(U)+\dots+A_k^{(k)}(U)$.

Proposition 6.11. *The following holds for every $k \geqslant 3$, $s \geqslant 2$, $p \in [0,1]$, $x \geqslant 0$, and vector $\mathbf{a} \in \mathbb{R}^k$. Letting $M \coloneqq \max_{i\in[\![k]\!]} a_i \cdot p^{k-i}$, we have*

$$\sum_{m\geqslant k+x} |\mathcal{C}_{m,s}(\mathbf{a})| \cdot p^m \leqslant \mu \cdot \left(\frac{e^2k^2M}{s}\right)^{s-1} \cdot \left(p \cdot \max_{i\in[\![k-1]\!]} \left(\frac{a_i}{M}\right)^{1/(k-i)}\right)^x \leqslant \mu \cdot \left(\frac{e^2k^2M}{s}\right)^{s-1}.$$

We postpone the proof of this proposition, showing first how to derive Lemmas 6.7 and 6.8.

Derivation of Lemma 6.7 from Proposition 6.11. We can assume that $s \leqslant 2(\log(1/p))^3$, as this is the maximum number of k-APs in a small cluster. We let $\mathbf{a} \in \mathbb{N}^k$ be the vector defined by

$$a_1 \coloneqq k^2sN \qquad \text{and} \qquad a_i \coloneqq k^4s^2 \text{ for } i \geqslant 2.$$

Since a union of s progressions contains at most ks elements and

$$|A_{\geqslant 1}(U)| \leqslant |U|kN \qquad \text{and} \qquad |A_{\geqslant i}(U)| \leqslant |U|^2k^2 \text{ for } i \geqslant 2,$$

every s-element cluster $\mathfrak{C}$ belongs to $\bigcup_m \mathcal{C}_{m,s}(\mathbf{a})$. Gearing up for an application of Proposition 6.11, we calculate

$$M \coloneqq \max_{i\in[\![k]\!]} a_i \cdot p^{k-i} = \max\left\{k^2sNp^{k-1}, k^4s^2\right\} = k^4s^2,$$

where we used the assumption $Np^{k-1} \ll 1$, and

$$\max_{i\in[\![k-1]\!]} \left(\frac{a_i}{M}\right)^{1/(k-i)} = \max\left\{\left(\frac{N}{k^2s}\right)^{1/(k-1)}, 1\right\} = \left(\frac{N}{k^2s}\right)^{1/(k-1)},$$

as $s \leqslant 2\big(\log(1/p)\big)^3 \leqslant N/k^2$. Let m_s denote the smallest cardinality of $\bigcup\mathfrak{C}$ for an s-element cluster $\mathfrak{C}$ that is small and observe that

$$m_s \geqslant \max\left\{k+1, \frac{\log\log(1/p)}{\varepsilon\log(1/p)} \cdot s\right\}, \tag{52}$$

by $s \geqslant 2$ and the definition of a small cluster. We may now deduce from Proposition 6.11 that

$$\begin{aligned} D_s \leqslant \sum_{m\geqslant m_s} |\mathcal{C}_{m,s}(\mathbf{a})| \cdot p^m &\leqslant \mu \cdot (k^6e^2s)^{s-1} \cdot \left(\frac{Np^{k-1}}{k^2s}\right)^{\frac{m_s-k}{k-1}} \\ &\leqslant \left(\frac{\mu}{t}\right)^s \cdot \frac{t}{4^{s+1}} \cdot \underbrace{16\left(\frac{4k^6e^2ts}{\mu}\right)^{s-1} \cdot \big(Np^{k-1}\big)^{m_s/k^2}}_{T}. \end{aligned}$$

Therefore, it suffices to show that $T \leqslant \varepsilon\log(1+t/\mu)$.

Assume first that $s \leqslant \mu/(4e^3k^6t)$. Since the function $s \mapsto (4k^6e^2ts/\mu)^{s-1}$ is decreasing on the interval $\big[2, \mu/(4e^3k^6t)\big]$, and since $Np^{k-1} \ll 1$ and $m_s \geqslant k+1$, we have

$$T \leqslant \frac{64k^6e^2t}{\mu} \cdot \big(Np^{k-1}\big)^{m_s/k^2} \leqslant \frac{t}{\mu} \cdot (Np^{k-1})^{1/k}.$$

Now, if $t \leqslant \mu$, then the claimed inequality follows from the inequality $t/\mu \leqslant 2\log(1+t/\mu)$. Otherwise, if $t \geqslant \mu$, then the assumption that $t \leqslant \big(\mu\log(1/p)\big)^{2/3} + \big(\log(1/p)\big)^3$ implies that $\mu \leqslant \big(2\log(1/p)\big)^3$ and thus $Np^{k-1} = O\big(\sqrt{\mu p^{k-2}}\big) = O\big(p^{1/3}\big)$. Since $t/\mu \leqslant O\big((\log(1/p))^3\big)$ also follows from the same assumption on t (as $\mu \geqslant \Omega(1)$), we may conclude that $T \ll 1$, whereas $\log(1+t/\mu) \geqslant \log 2$.

Finally, assume that $\mu/(4e^3k^6t) < s \leqslant 2\big(\log(1/p)\big)^3$. In this case, our assumption $t \leqslant \big(\mu\log(1/p)\big)^{2/3} + \big(\log(1/p)\big)^3$ implies that $\mu = O\big(\big(\log(1/p)\big)^{11}\big)$, and thus $Np^{k-1} = O(\sqrt{\mu p^{k-2}}) \leqslant p^{1/3}$. Since also $t/\mu \leqslant O\big((\log(1/p))^3\big)$, we obtain

$$T \leqslant \big(O\big((\log(1/p))^6\big)\big)^{s-1} \cdot p^{m_s/(3k^2)} \leqslant \exp\big(7s\log\log(1/p) - m_s/(3k^2)\cdot\log(1/p)\big).$$

Recalling from (52) that $s\log\log(1/p) \leqslant m_s\big(\varepsilon\log(1/p)\big)$, where we can assume without loss of generality that ε is somewhat small, we may conclude that

$$T \leqslant p^{m_s/(4k^2)} \leqslant p^{1/(4k)} \ll \frac{1}{2\mu} \leqslant \log(1+1/\mu) \leqslant \log(1+t/\mu),$$

as desired. □

Derivation of Lemma 6.8 from Proposition 6.11. Let $\mathbf{a} \in \mathbb{N}^k$ be the vector defined by

$$a_i \coloneqq \frac{L\mu s}{Kp^{k-i}(1-p)t}.$$

Since every s-element cluster $\mathfrak{C}$ that is L-bounded satisfies

$$A_{\geqslant i}(\bigcup\mathfrak{C})\cdot(p^{k-i}-p^k) \leqslant \psi(\bigcup\mathfrak{C}) \leqslant \frac{L\mu s}{Kt} \leqslant a_i\cdot(p^{k-i}-p^k)$$

for all $i \in [\![k]\!]$, it belongs to $\bigcup_m \mathcal{C}_{m,s}(\mathbf{a})$. We may thus deduce from Proposition 6.11 that

$$D_{s,L} \leqslant \sum_m |\mathcal{C}_{m,s}(\mathbf{a})|\cdot p^m \leqslant \mu\cdot\left(\frac{e^2k^2L\mu}{K(1-p)t}\right)^{s-1} = \frac{\mu^s}{t^s}\cdot\left(\frac{e^2k^2L}{K(1-p)}\right)^{s-1}\cdot t \leqslant \frac{\mu^s}{t^s}\cdot L^s\cdot\frac{e^{-s}\cdot t}{4^{s+1}},$$

when K is chosen to be sufficiently large as a function of k, as we assume $L \geqslant 1$.

Since $s \geqslant \big(\varepsilon\log(1/p)/\log\log(1/p)\big)^2 \gg \log(1/p)$ for every s-cluster that is not small, and as $Np^{k-1} \leqslant 1$ implies $N^2p^{2k} \ll 1$, we have

$$e^{-s} \leqslant p^k \ll \frac{1}{2\mu} \leqslant \log(1+1/\mu) \leqslant \log(1+t/\mu),$$

which implies the desired inequality. □

We finally move to proving Proposition 6.11. To do this, we will need to establish an upper bound on $\mathcal{C}_{m,s}(\mathbf{a})$ for all integers m and s and all vectors $\mathbf{a}$. Since there is nothing special here about the family (hypergraph) AP_k of arithmetic progressions that underlies the notion of a cluster, we will prove a more abstract statement that provides an analogous bound for the number of connected subhypegraphs with a given edge boundary, which could be of independent interest. The proof of this theorem can be found in Appendix A.

Let $\mathcal{H}$ be a hypergraph. Given a subset $W \subseteq V(\mathcal{H})$ and a positive integer i, we denote $\partial^{(i)}_{\mathcal{H}}(W) \coloneqq \{e \in \mathcal{H} : |e\cap W| = i\}$. Further, for a vector $\mathbf{a} = (a_1,\dots,a_k) \in \mathbb{R}^k$, we define $\mathcal{C}_{m,s}(\mathbf{a};\mathcal{H})$ to be the set of connected subhypergraphs $\mathcal{H}' \subseteq \mathcal{H}$ with m vertices and s edges that satisfy $|\partial^{(i)}_{\mathcal{H}}(V(\mathcal{H}'))| + \cdots + |\partial^{(k)}_{\mathcal{H}}(V(\mathcal{H}'))| \leqslant a_i$ for all $i \in [\![k]\!]$, so that $\mathcal{C}_{m,s}(\mathbf{a}) = \mathcal{C}_{m,s}(\mathbf{a};\mathrm{AP}_k)$.

Theorem 6.12. *Suppose that $\mathcal{H}$ is a k-uniform hypergaph. For all $m,s \in \mathbb{N}$ and $\mathbf{a} = (a_1,\dots,a_k) \in \mathbb{R}^k$,*

$$|\mathcal{C}_{m,s}(\mathbf{a};\mathcal{H})| \leqslant e(\mathcal{H})\cdot \sum_{\substack{s_1,\dots,s_k\geqslant 0\\ \sum_i (k-i)s_i = m-k\\ \sum_i s_i = s-1}} \prod_{i=1}^k \binom{a_i}{s_i}.$$

Proof of Proposition 6.11. We work in the hypergraph AP_k of k-APs in $[\![N]\!]$. Since every cluster of arithmetic progressions naturally corresponds to a connected subhypergraph of AP_k and since $\partial^{(i)}_{\mathrm{AP}_k}(U) = A_i(U)$ for every $U \subseteq [\![N]\!]$, one can see that Theorem 6.12 implies

$$|\mathcal{C}_{m,s}(\mathbf{a})| \leqslant |\mathcal{C}_{m,s}(\mathbf{a}; \mathrm{AP}_k)| \leqslant |\mathrm{AP}_k| \cdot \sum_{\substack{s_1,\dots,s_k \geqslant 0 \\ \sum_i (k-i)s_i = m-k \\ \sum_i s_i = s-1}} \prod_{i=1}^{k} \binom{a_i}{s_i}.$$

Fix some $x \geqslant 0$. Letting

$$\mathcal{S} \coloneqq \{(s_1,\dots,s_k) \in \mathbb{Z}^k_{\geqslant 0} : s_1 + \cdots + s_k = s-1 \text{ and } \sum_{i=1}^{k} (k-i)s_i \geqslant x\},$$

we then have

$$\begin{aligned}\sum_{m \geqslant k+x} |\mathcal{C}_{m,s}(\mathbf{a})| \cdot p^m &\leqslant |\mathrm{AP}_k| \cdot p^k \cdot \sum_{(s_1,\dots,s_k)\in\mathcal{S}} \prod_{i=1}^{k} \binom{a_i}{s_i} \cdot p^{(k-i)s_i} \\ &\leqslant \mu \cdot \sum_{(s_1,\dots,s_k)\in\mathcal{S}} \prod_{i=1}^{k} \left(\frac{e a_i p^{k-i}}{s_i}\right)^{s_i}\end{aligned}$$

where the convention $0^0 = 1$ is to be used when $s_i = 0$. Note that, for every $(s_1,\dots,s_k) \in \mathcal{S}$,

$$\prod_{i=1}^{k} \left(\frac{1}{s_i}\right)^{s_i} = \left(\frac{k}{s-1}\right)^{s-1} \cdot \exp\left(-\sum_{i=1}^{k} s_i \log\left(\frac{s_i}{(s-1)/k}\right)\right) \leqslant \left(\frac{k}{s-1}\right)^{s-1},$$

since the sum in the expression above is $s-1$ times the (nonnegative) Kullback–Leibler divergence of the random variable taking the value $i \in [\![k]\!]$ with probability $s_i/(s-1)$ from the uniform element of $[\![k]\!]$. This yields the bound

$$\sum_{m \geqslant k+x} |\mathcal{C}_{m,s}(\mathbf{a})| \cdot p^m \leqslant \mu \cdot \left(\frac{ek}{s-1}\right)^{s-1} \cdot \sum_{(s_1,\dots,s_k)\in\mathcal{S}} \prod_{i=1}^{k} \left(a_i p^{k-i}\right)^{s_i}.$$

By the definition of M, we have $a_i p^{k-i} \leqslant M$ for every $i \in [\![k]\!]$; so for every $(s_1,\dots,s_k) \in \mathcal{S}$, we have

$$\prod_{i=1}^{k} \left(a_i p^{k-i}\right)^{s_i} \leqslant M^{s-1} \cdot \prod_{i=1}^{k-1} \left(\frac{a_i p^{k-i}}{M}\right)^{s_i} \leqslant M^{s-1} \cdot \prod_{i=1}^{k-1} \left(\max_{j \in [\![k-1]\!]} \left(\frac{a_j p^{k-j}}{M}\right)^{1/(k-j)}\right)^{(k-i)s_i}.$$

Since the quantity in the iterated product is at most 1, we have

$$\sum_{m \geqslant k+x} |\mathcal{C}_{m,s}(\mathbf{a})| \cdot p^m \leqslant \mu \cdot \left(\frac{ekM}{s-1}\right)^{s-1} \cdot \left(p \cdot \max_{j \in [\![k-1]\!]} \left(\frac{a_j}{M}\right)^{1/(k-j)}\right)^{x} \cdot |\mathcal{S}|.$$

The claimed bound now follows after we observe that $|\mathcal{S}| \leqslant \binom{s-1+k-1}{s-1} \leqslant (ek)^{s-1}$. □

References

[1] J. D. Alvarado, L. G. de Oliveira, and S. Griffiths. Moderate deviations of triangle counts in sparse erd\h {o} sr\'enyi random graphs g(n,m) and g(n,p). *arXiv preprint arXiv:2305.04326*, 2023.

[2] F. Augeri. Nonlinear large deviation bounds with applications to Wigner matrices and sparse Erdős-Rényi graphs. *Ann. Probab.*, 48(5):2404–2448, 2020.

[3] T. Austin. The structure of low-complexity Gibbs measures on product spaces. *Ann. Probab.*, 47(6):4002–4023, 2019.

[4] J. Balogh, R. Morris, W. Samotij, and L. Warnke. The typical structure of sparse K_{r+1}-free graphs. *Trans. Amer. Math. Soc.*, 368(9):6439–6485, 2016.

[5] A. D. Barbour, M. Karoński, and A. Ruciński. A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B*, 47(2):125–145, 1989.

[6] Y. Barhoumi-Andréani, C. Koch, and H. Liu. Bivariate fluctuations for the number of arithmetic progressions in random sets. *Electronic J. Probab.*, 24:1–32, 2019.

[7] A. Basak and R. Basu. Upper tail large deviations of regular subgraph counts in erdős-rényi graphs in the full localized regime. *Communications on Pure and Applied Mathematics*, 76(1):3–72, 2023.

[8] R. Berkowitz, A. Sah, and M. Sawhney. Number of arithmetic progressions in dense random subsets of z/nz. *Israel Journal of Mathematics*, 244(2):589–620, 2021.

[9] B. B. Bhattacharya, S. Ganguly, X. Shao, and Y. Zhao. Upper tail large deviations for arithmetic progressions in a random set. *Int. Math. Res. Not. IMRN*, 2020(1):167–213, 2020.

[10] S. Chatterjee and A. Dembo. Nonlinear large deviations. *Adv. Math.*, 299:396–450, 2016.

[11] N. Cook and A. Dembo. Large deviations of subgraph counts for sparse erdős–rényi graphs. *Advances in Mathematics*, 373:107289, 2020.

[12] N. A. Cook, A. Dembo, and H. T. Pham. Regularity method and large deviation principles for the erdős–rényi hypergraph. *Duke Mathematical Journal*, 173(5):873–946, 2024.

[13] R. Eldan. Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *Geom. Funct. Anal.*, 28(6):1548–1596, 2018.

[14] G. Fiz Pontiveros, S. Griffiths, M. Secco, and O. Serra. Deviation probabilities for arithmetic progressions and other regular discrete structures. *Random Structures Algorithms*, 60(3):367–405, 2022.

[15] D. A. Freedman. On tail probabilities for martingales. *Ann. Probability*, 3:100–118, 1975.

[16] C. Goldschmidt, S. Griffiths, and A. Scott. Moderate deviations of subgraph counts in the erdős-rényi random graphs g(n,m) and g(n,p). *Transactions of the American Mathematical Society*, 373(8):5517–5585, 2020.

[17] S. Griffiths, C. Koch, and M. Secco. Deviation probabilities for arithmetic progressions and irregular discrete structures. *Electron. J. Probab.*, 28:Paper No. 172, 31, 2023.

[18] M. Harel, F. Mousset, and W. Samotij. Upper tails via high moments and entropic stability. *Duke Math. J.*, 171(10):2089–2192, 2022.

[19] T. E. Harris. A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.*, 56:13–20, 1960.

[20] S. Janson. Poisson approximation for large deviations. *Random Structures Algorithms*, 1(2):221–229, 1990.

[21] S. Janson, K. Oleszkiewicz, and A. Ruciński. Upper tails for subgraph counts in random graphs. *Israel J. Math.*, 142:61–92, 2004.

[22] S. Janson and A. Ruciński. The infamous upper tail. *Random Structures Algorithms*, 20(3):317–342, 2002. Probabilistic methods in combinatorial optimization.

[23] N. Ross. Fundamentals of Stein's method. *Probab. Surv.*, 8:210–293, 2011.

[24] A. Ruciński. When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields*, 78(1):1–10, 1988.

[25] L. Warnke. Upper tails for arithmetic progressions in random subsets. *Israel J. Math.*, 221(1):317–365, 2017.

Appendix A. Proof of the hypergraph cluster lemma

In this appendix, we prove Theorem 6.12, the bound for the number of connected subhypergraphs with a given edge boundary, which we used to control factorial moments of the number of k-term arithmetic progressions in the Poisson and the localised regimes. We restate the theorem here for reader's convenience. Recall that given a set W of vertices of a hypergraph $\mathcal{H}$ and a positive integer i, we denote $\partial_{\mathcal{H}}^{(i)}(W) \coloneqq \{e \in \mathcal{H} : |e \cap W| = i\}$. Further, for a vector $\mathbf{a} = (a_1, \dots, a_k) \in \mathbb{R}^k$, we defined $\mathcal{C}_{m,s}(\mathbf{a}; \mathcal{H})$ to be the set of connected subhypergraphs $\mathcal{H}' \subseteq \mathcal{H}$ with m vertices and s edges that satisfy $|\partial_{\mathcal{H}}^{(i)}(V(\mathcal{H}'))| + \cdots + |\partial_{\mathcal{H}}^{(k)}(V(\mathcal{H}'))| \leqslant a_i$ for all $i \in [\![k]\!]$.

Theorem 6.12. *Suppose that $\mathcal{H}$ is a k-uniform hypergaph. For all $m, s \in \mathbb{N}$ and $\mathbf{a} = (a_1, \dots, a_k) \in \mathbb{R}^k$,*

$$|\mathcal{C}_{m,s}(\mathbf{a}; \mathcal{H})| \leqslant e(\mathcal{H}) \cdot \sum_{\substack{s_1,\dots,s_k \geqslant 0 \\ \sum_i (k-i)s_i = m-k \\ \sum_i s_i = s-1}} \prod_{i=1}^{k} \binom{a_i}{s_i}.$$

Proof. We prove the claim by exhibiting an injective map from $\mathcal{C}_{n,m}(\mathbf{a}; \mathcal{H})$ into a set of combinatorial objects that are easier to count. To this end, let $I(\mathbf{a}) \coloneqq \{(i,j) : 1 \leqslant i \leqslant k \text{ and } 1 \leqslant j \leqslant a_i\}$ and let $\mathcal{T}_{m,s}(\mathbf{a})$ be the collection of all subsets $T \subseteq I(\mathbf{a})$ such that

$$|T| = s - 1 \qquad \text{and} \qquad \sum_{(i,j) \in T} (k - i) = m - k. \tag{53}$$

It may help to think of $I(\mathbf{a})$ as representing a grid or tableau consisting of k rows of respective lengths $a_1, \dots, a_k$ and of an element of $\mathcal{T}_{m,s}(\mathbf{a})$ as a certain way of marking some $s - 1$ cells of the tableau.

Further below, we will argue that there exists an injective map

$$\mathrm{emb} \colon \mathcal{C}_{m,s}(\mathbf{a}; \mathcal{H}) \to E(\mathcal{H}) \times \mathcal{T}_{m,s}(\mathbf{a}),$$

which clearly implies that

$$|\mathcal{C}_{m,s}(\mathbf{a}; \mathcal{H})| \leqslant e(\mathcal{H}) \cdot |\mathcal{T}_{m,s}(\mathbf{a})|. \tag{54}$$

From this, we may complete the proof by counting as follows. First, note that, for every $T \in \mathcal{T}_{m,s}(\mathbf{a})$, the 'row counts' $s_1, \dots, s_k$ defined by $s_i \coloneqq |\{(i', j) \in T : i' = i\}|$ satisfy $\sum_i (k - i)s_i = m - k$ and $\sum_i s_i = s - 1$. If $T \in \mathcal{T}_{m,s}(\mathbf{a})$, then T is fully specified by giving $(s_1, \dots, s_k)$ and then choosing, for each row $1 \leqslant i \leqslant k$, the s_i values $j \in [\![a_i]\!]$ such that $(i, j) \in T$. This gives the bound

$$|\mathcal{T}_{m,s}(\mathbf{a})| \leqslant \sum_{\substack{s_1,\dots,s_k \geqslant 0 \\ \sum_i (k-i)s_i = m-k \\ \sum_i s_i = s-1}} \prod_{i=1}^{k} \binom{a_i}{s_i},$$

which, together with (54), gives the assertion of the theorem.

Algorithm 1 takes an element $\mathcal{H}' \in \mathcal{C}_{n,m}(\mathbf{a}; \mathcal{H})$ as input and returns a pair (e, T) consisting of an edge $e \in E(\mathcal{H})$ and a subset $T \subseteq I(\mathbf{a})$. We will argue that one can define emb as the function computed by the algorithm. This requires showing the following:

(i) each step in the algorithm is well-defined and can indeed be carried out as described;

(ii) the pair (e, T) computed by the algorithm belongs to $E(\mathcal{H}) \times \mathcal{T}_{m,s}(\mathbf{a})$;

(iii) the map $\mathcal{H}' \mapsto (e, T)$ is injective.

Algorithm 1: The algorithm defining emb

Data: A hypergraph $\mathcal{H}' \in \mathcal{C}_{m,s}(\mathbf{a}; \mathcal{H})$, where $\mathcal{H}$ is a k-uniform hypergraph

Result: A pair (e, T) consisting of an edge $e \in \mathcal{H}$ and a subset $T \subseteq I(\mathbf{a})$

1 fix a total ordering $\preceq$ on the edges of $\mathcal{H}$;
2 **for** $i \in [\![k]\!]$ **do**
3 $\quad \sigma_1^{(i)} \leftarrow$ the empty sequence (of edges of $\mathcal{H}$);
4 **end**
5 $e_1 \leftarrow$ the $\preceq$-smallest edge of $\mathcal{H}'$;
6 $T_1 \leftarrow \varnothing \subseteq I(\mathbf{a})$;
7 **for** $\ell = 2, \dots, s$ **do**
8 $\quad$ **for** $i \in [\![k]\!]$ **do**
9 $\quad\quad \sigma_\ell^{(i)} \leftarrow \sigma_{\ell-1}^{(i)} \parallel (x_1, \dots, x_r)$, where $\parallel$ denotes concatenation of finite sequences and $x_1, \dots, x_r$ are the elements of $\partial_{\mathcal{H}}^{(i)}(e_1 \cup \dots \cup e_{\ell-1}) \setminus \sigma_{\ell-1}^{(i)}$ ordered according to $\preceq$;
10 $\quad$ **end**
11 $\quad (i, j) \leftarrow$ the lexicographically largest $(i, j) \in I(\mathbf{a})$ such that $(\sigma_\ell^{(i)})_j \in E(\mathcal{H}') \setminus \{e_1, \dots, e_{\ell-1}\}$;
12 $\quad e_\ell \leftarrow (\sigma_\ell^{(i)})_j$;
13 $\quad T_\ell \leftarrow T_{\ell-1} \cup \{(i, j)\}$;
14 **end**
15 **return** (e_1, T_m)

A first glance at the definition of the algorithm reveals that, in addition to computing (e, T), the algorithm defines sequences $(e_1, \dots, e_s)$, $(T_1, \dots, T_s)$, and $(\sigma_1^{(i)}, \dots, \sigma_s^{(i)})$ for every $i \in [\![k]\!]$. At this point, we can already convince ourselves, using straightforward induction, that each e_ℓ is a distinct edge of $\mathcal{H}'$, each T_ℓ is a subset of $I(\mathbf{a})$, and each $\sigma_\ell^{(i)}$ is a finite sequence of distinct edges of $\mathcal{H}$ (however, one and the same edge might belong to two different sequences $\sigma_\ell^{(i)}$ and $\sigma_\ell^{(i')}$).

As far as the well-definedness of the algorithm is concerned, the only critical step is on line 11 of the algorithm, where we need to show that at least one $(i, j) \in I(\mathbf{a})$ with $(\sigma_\ell^{(i)})_j \in E(\mathcal{H}') \setminus \{e_1, \dots, e_{\ell-1}\}$ exists. Since $\mathcal{H}'$ is connected and $\ell \leqslant s = e(\mathcal{H}')$, there exists at least one edge in $E(\mathcal{H}') \setminus \{e_1, \dots, e_{\ell-1}\}$ that intersects $e_1 \cup \dots \cup e_{\ell-1}$, in $i \in [\![k]\!]$ vertices. Since $\sigma_\ell^{(i)}$ contains all edges in $\partial_{\mathcal{H}}^{(i)}(e_1 \cup \dots \cup e_{\ell-1})$ by definition, there thus exists some j such that $(\sigma_\ell^{(i)})_j \in \mathcal{H}' \setminus \{e_1, \dots, e_{\ell-1}\}$. It now suffices to show that any such (i, j) belongs to $I(\mathbf{a})$. To this end, observe that every edge in $\sigma_\ell^{(i)}$ belongs to $\partial_{\mathcal{H}}^{(i)}(e_1 \cup \dots \cup e_{\ell-1}) \cup \dots \cup \partial_{\mathcal{H}}^{(k)}(e_1 \cup \dots \cup e_{\ell-1})$, since at some (possibly earlier) iteration it must have intersected the union of a prefix of $(e_1, \dots, e_{\ell-1})$ in i vertices. The assumption $\mathcal{H}' \in \mathcal{C}_{m,s}(\mathbf{a}; \mathcal{H})$ then implies that $\sigma_\ell^{(i)}$ has length at most a_i; so we have $(i, j) \in I(\mathbf{a})$. Therefore, a pair (i, j) as on line 11 of Algorithm 1 really exists, which shows that the algorithm is well-defined.

We now show that the pair (e, T) computed by Algorithm 1 belongs to $E(\mathcal{H}) \times \mathcal{T}_{m,s}(\mathbf{a})$. For this, it is sufficient to show that for every $1 \leqslant \ell \leqslant m$, we have

$$|T_\ell| = \ell - 1 \qquad \text{and} \qquad \sum_{(i,j)\in T_\ell} (k-i) = |e_1 \cup \cdots \cup e_\ell| - k.$$

Indeed, both statements are easily seen to hold for $\ell = 1$. Using induction on ℓ, the first statement now follows from the definition of T_ℓ on line 13 and the fact that $(\sigma_\ell^{(i)})_j = e_\ell \notin \{e_1, \ldots, e_{\ell-1}\}$, which implies that a different pair (i, j) is added on each iteration (since the sequences $\sigma_1^{(i)}, \ldots, \sigma_\ell^{(i)}$ are each a prefix of the next). For the second statement, we observe that the maximality of i on line 11, together with the fact that $\sigma_\ell^{(i)}$ contains all edges intersecting $e_1 \cup \cdots \cup e_{\ell-1}$ in exactly i vertices, implies that the edge e_ℓ intersects $e_1 \cup \cdots \cup e_{\ell-1}$ in *exactly* i vertices. Thus, adding e_ℓ to the list of edges adds precisely $k - i = \sum_{(i,j)\in T_\ell}(k-i) - \sum_{(i,j)\in T_{\ell-1}}(k-i)$ new vertices to their union.

Finally, we show that the function computed by the algorithm is injective. To this end, suppose that Algorithm 1 was run on two *different* hypegraphs $\mathcal{H}', \hat{\mathcal{H}}' \in \mathcal{C}_{m,s}(\mathbf{a}; \mathcal{H})$ and defined:

- sequences $(e_\ell)_\ell$, $(T_\ell)_\ell$, and $(\sigma_\ell^{(i)})_\ell$ for every $i \in [\![k]\!]$ while run on $\mathcal{H}'$;
- sequences $(\hat{e}_\ell)_\ell$, $(\hat{T}_\ell)_\ell$, and $(\hat{\sigma}_\ell^{(i)})_\ell$ for every $i \in [\![k]\!]$ while run on $\hat{\mathcal{H}}'$.

Since $e_1, \ldots, e_s$ and $\hat{e}_1, \ldots, \hat{e}_s$ are orderings of all edges of $\mathcal{H}'$ and $\hat{\mathcal{H}}'$, respectively, they must differ in at least one coordinate; let ℓ be the smallest such coordinate. If $\ell = 1$, then $e_1 \neq \hat{e}_1$ and the two outputs are clearly different. We will thus assume that $\ell > 1$. By the minimality of ℓ, we have $(e_1, \ldots, e_{\ell-1}) = (\hat{e}_1, \ldots, \hat{e}_{\ell-1})$. Observe that this implies that $(T_1, \ldots, T_{\ell-1}) = (\hat{T}_1, \ldots, \hat{T}_{\ell-1})$ and that $(\sigma_1^{(i)}, \ldots, \sigma_\ell^{(i)}) = (\hat{\sigma}_1^{(i)}, \ldots, \hat{\sigma}_\ell^{(i)})$ for every $i \in [\![k]\!]$.

Let (i, j) and $(\hat{i}, \hat{j})$ be the pairs chosen in line 11 of the ℓ-th iteration of the main for loop during the two respective executions of the algorithm. Since $(\sigma_\ell^{(i)})_j = e_\ell \neq \hat{e}_\ell = (\hat{\sigma}_\ell^{(\hat{i})})_{\hat{j}} = (\sigma_\ell^{(\hat{i})})_{\hat{j}}$, it must be that $(i, j) \neq (\hat{i}, \hat{j})$; without loss of generality, we may assume that (i, j) is lexicographically larger. This means that $e_\ell = (\sigma_\ell^{(i)})_j = (\hat{\sigma}_\ell^{(i)})_j \notin E(\hat{\mathcal{H}}') \setminus \{\hat{e}_1, \ldots, \hat{e}_{\ell-1}\}$, by maximality of $(\hat{i}, \hat{j})$. Finally, since $(\hat{\sigma}_{\ell'}^{(i)})_j = (\hat{\sigma}_\ell^{(i)})_j$ for every $\ell' \geqslant \ell$, we may conclude that (i, j) cannot be added to $\hat{T}_{\ell'}$ for any $\ell' \geqslant \ell$. In particular, as $(i, j) \notin T_{\ell-1} = \hat{T}_{\ell-1}$, we conclude that $(i, j) \in T_s \setminus \hat{T}_s$. This completes the proof. □

Department of Mathematics, Northeastern University, Boston, MA, USA
Email address: m.harel@northeastern.edu

Email address: moussetfrank@gmail.com

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel
Email address: samotij@tauex.tau.ac.il