The Minimization of Random Hypergraphs

Thomas Bläsius

Hasso Plattner Institute, University of Potsdam, Germany

Tobias Friedrich

Hasso Plattner Institute, University of Potsdam, Germany

Martin Schirneck[1]

Hasso Plattner Institute, University of Potsdam, Germany

Abstract

We investigate the maximum-entropy model $\mathcal{B}_{n,m,p}$ for random n-vertex, m-edge multi-hypergraphs with expected edge size pn. We show that the expected size of the minimization $\min(\mathcal{B}_{n,m,p})$, i.e., the number of inclusion-wise minimal edges of $\mathcal{B}_{n,m,p}$, undergoes a phase transition with respect to m. If m is at most $1/(1-p)^{(1-p)n}$, then $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|]$ is of order $\Theta(m)$, while for $m \geq 1/(1-p)^{(1-p+\varepsilon)n}$ for any $\varepsilon > 0$, it is $\Theta(2^{(\mathrm{H}(\alpha)+(1-\alpha)\log_2 p)\, n}/\sqrt{n})$. Here, H denotes the binary entropy function and $\alpha = -(\log_{1-p} m)/n$. The result implies that the maximum expected number of minimal edges over all m is $\Theta((1+p)^n/\sqrt{n})$. Our structural findings have algorithmic implications for minimizing an input hypergraph. This has applications in the profiling of relational databases as well as for the Orthogonal Vectors problem studied in fine-grained complexity. We make several technical contributions that are of independent interest in probability. First, we improve the Chernoff–Hoeffding theorem on the tail of the binomial distribution. In detail, we show that for a binomial variable $Y \sim \mathrm{Bin}(n,p)$ and any $0 < x < p$, it holds that $\mathrm{P}[Y \leq xn] = \Theta(2^{-\mathrm{D}(x\,\|\,p)\,n}/\sqrt{n})$, where D is the binary Kullback–Leibler divergence between Bernoulli distributions. We give explicit upper and lower bounds on the constants hidden in the big-O notation that hold for all n. Secondly, we establish the fact that the probability of a set of cardinality i being minimal after m i.i.d. maximum-entropy trials exhibits a sharp threshold behavior at $i^* = n + \log_{1-p} m$.

2012 ACM Subject Classification Mathematics of computing → Information theory; Mathematics of computing → Hypergraphs; Theory of computation → Random network models; Mathematics of computing → Random graphs

Keywords and phrases Chernoff–Hoeffding theorem, maximum entropy, maximization, minimization, phase transition, random hypergraphs

Digital Object Identifier 10.4230/LIPIcs.ESA.2020.21

Related Version A full version of the paper is available at `https://arxiv.org/abs/1910.00308`.

Acknowledgements The authors thank Benjamin Doerr, Timo Kötzing, and Martin Krejca for the fruitful discussions on the Chernoff–Hoeffding theorem, including valuable pointers to the literature.

1 Introduction

A plethora of work has been dedicated to the analysis of random graphs. Random hypergraphs, however, received much less attention. For many types of data, hypergraphs provide a much more natural model. This is especially true if the data has a hierarchical structure or reflects interactions between groups of entities. In non-uniform hypergraphs, where edges can have different numbers of vertices, a phenomenon occurs that is unknown to graphs: an edge may be contained in another, with multiple edges even forming chains of inclusion. We are often only interested in the endpoints of those chains, namely, the collections of inclusion-wise minimal or maximal edges. This is the *minimization* or *maximization* of the hypergraph.

[1] Corresponding author: `martin.schirneck@hpi.de`.

28th Annual European Symposium on Algorithms (ESA 2020).
Editors: Fabrizio Grandoni, Grzegorz Herman, and Peter Sanders; Article No. 21; pp. 21:1–21:15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

We investigate the maximum-entropy model $\mathcal{B}_{n,m,p}$ for random multi-hypergraphs with n vertices and m edges and expected edge size pn for some constant sampling probability p. In other words, out of all probability distributions on hypergraphs with expected edge size pn, $\mathcal{B}_{n,m,p}$ is the one of maximum entropy.[2] This is equivalent to sampling m independent edges by adding any vertex $v \in [n]$ independently with probability p (see Section 2 for details). We are interested in the expected size of the minimization/maximization of $\mathcal{B}_{n,m,p}$, that is, the expected number of minimal/maximal edges. Most of our results are phrased in terms of the minimization, but replacing the probability p with $1-p$ immediately transfers them to the maximization. We show that the size of the minimization undergoes a phase transition with respect to m with the point of transition at $m = 1/(1-p)^{(1-p)n}$. While the number of edges is still small, a constant fraction of them is minimal and the minimization grows linearly in the total sample sizes. For m beyond the transition, we can instead characterize the size of the minimization in terms of the entropy function of $\log_{1-p} m$, see Theorem 1.2 for a precise statement. This characterization shows that the minimality ratio goes down dramatically when m increases. It also allows us to prove that the maximum expected number of minimal edges over all m is of order $\Theta((1+p)^n/\sqrt{n})$. These results draw from another, more hidden, threshold behavior. The probability of a set to be minimal in the hypergraph $\mathcal{B}_{n,m,p}$ depends only on its cardinality i and we show that this probability falls sharply from almost 1 to almost 0 at $i^* = n + \log_{1-p} m$.

The main tool in our analysis is the Chernoff–Hoeffding theorem bounding the tail of the binomial distribution via the Kullback–Leibler divergence from information theory. However, the existing inequalities are not sharp enough to derive tight statements on the expected size of the minimization. So far, there is a gap of order $\sqrt{n}$ between the best-known upper and lower estimates. In this work, we improve these bounds such that they match up to constant factors. We give an explicit interval for the constants involved that holds for all positive integers n making the result useful also in a non-asymptotic setting.

Our structural findings have algorithmic implications for the computation of the minimization $\min(\mathcal{H})$ from an input hypergraph $\mathcal{H}$. We discuss two examples in the context of fine-grained complexity as well as data profiling. There are reasons to believe that there exists no minimization algorithm running in time $m^{2-\varepsilon} \cdot \mathrm{poly}(n)$ for any $\varepsilon > 0$ on m-edge, n-vertex hypergraphs. The reason is as follows: The Sperner Family problem is to decide whether $\mathcal{H}$ comprises two edges such that one is contained in the other, i.e., whether $|\min(\mathcal{H})| < |\mathcal{H}|$. It is equivalent under subquadratic reductions to the more prominent Orthogonal Vectors problem [14, 25]. Hence, a truly subquadratic algorithm would falsify the Orthogonal Vectors Conjecture[3] and in turn the Strong Exponential Time Hypothesis [52]. Partitioning the edges by the number of vertices and processing them in order of increasing cardinality gives an algorithm running in $\mathrm{O}(mn\,|\min(\mathcal{H})| + mn)$, which is $\mathrm{O}(m^2 n)$ in the worst case. However, when looking at the average-case complexity for $\mathcal{B}_{n,m,p}$, we get a run time of $\mathrm{O}(mn\,\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] + mn)$. Our main result therefore shows that the algorithm is subquadratic for all m beyond the phase transition, and even linear for $m \geq 1/(1-p)^n$.

There is also a connection to the profiling of relational databases. Data scientists regularly need to compile and output a comprehensive list of metadata, like unique column combinations, functional dependencies, or, more general, denial constraints, cp. [1]. These multi-column dependencies can all be described as the minimal hitting sets of certain hypergraphs created from comparing pairs of rows in the database and recording the sets

[2] The notation $\mathcal{B}_{n,m,p}$ is mnemonic of the binomial distribution emerging in the sampling process.
[3] Precisely, we mean the Orthogonal Vectors Conjecture for moderate dimensions, see [25].

of attributes in which they differ [24, 11, 10, 40]. Computing these difference sets one by one generates an incoming stream of seemingly random subsets. Filtering the inclusion-wise minimal ones from this stream does not affect the solution, but reduces the number of sets to store and the complexity of the resulting hitting set instance. Minimizing the input is therefore a standard preprocessing technique in data profiling. It has been observed in practice that the minimal sets make up only a small fraction of the whole input [45]. Usually there are fewer minimal difference sets than rows in the database, let alone pairs thereof [11]. The upper bounds given in the Theorems 1 and 2 provide a way to explain this phenomenon. We show that only a few edges can be expected to be minimal, their number may even shrink as the database grows, provided that the number of rows is large enough compared to the number of columns. The respective lower bounds can further be seen as the smallest amount of data any dependency enumeration algorithm needs to hold in memory.

Related Work. Erdős–Rényi graphs $\mathcal{G}_{n,m}$ [23] and Gilbert graphs $\mathcal{G}_{n,p}$ [27] are arguably the most-discussed random graph models in the literature. We refer the reader to the monograph by Bollobás [12] for an overview. A majority of the work on these models concentrates on various phase transitions with respect to the number of edges m or the sample probability p, respectively. This intensive treatment is fueled by the appealing property that Erdős–Rényi graphs are "maximally random" in that they do not assume anything but the number of vertices and edges. More formally, among all probability distributions on graphs with n vertices and m edges, $\mathcal{G}_{n,m}$ is the unique distribution of maximum entropy. The same holds for $\mathcal{G}_{n,p}$ under the constraint that the expected number of edges is $p\binom{n}{2}$, see [2].

The intuition of being maximally random is captured by the Shannon entropy, which is the central concept in information theory [18, 50]. A discrete stochastic system that can be described by the probability distribution $(p_i)_i$ has a (binary) *entropy* of $H((p_i)_i) = -\sum_i p_i \log_2 p_i$. The *self-information* of a single state with probability p is $-\log_2 p$, the entropy is thus the expected information of the whole system. It is a measure of surprisal or how "spread out" the distribution is. Originally stemming from thermodynamics [39], the versatility of this definition is key to the successful application of information theory to fields as diverse as cryptography [15], machine learning [28], quantum computing [44], and of course network analysis [43], to name only a few topics close to computer science. The *principle of maximum entropy* states that out of an ensemble of probability distributions that all describe the observed phenomena equally well, the one of maximum entropy is to be preferred in order to minimize any outside bias. The principle is usually attributed to Jaynes [37, 32, 33]. In the context of random graphs, it is mainly used to define so-called *null models* [53]. One fixes certain graph statistics to mimic those of an observed network and then chooses the maximum-entropy distribution that meets these constraints. By comparing the original network with a "typical graph" drawn from the null model, one can infer whether other observed properties are correlated with the constraints. This method was made rigorous by Park and Newman [46] building on earlier work in general statistics. Prescribing the exact or expected number of edges leads to the $\mathcal{G}_{n,m}$ or $\mathcal{G}_{n,p}$ distributions, respectively. The configuration model fixes the whole degree sequence [13], and in the soft configuration model the degrees hold at least in expectation [8, 26].

Many early attempts to transfer the concept of null models to hypergraphs were only indirect in that they studied hypergraphs via their clique-expansion [42] or as bipartite graphs [48]. This is unsatisfactory since these projections alter relevant observables, like node degrees or the number of triangles. Only recently, Chodrow generalized the configuration model directly to multi-hypergraphs [16]. There also seems to be not much literature on hypergraph models that can be cast into the maximum-entropy framework without being intentionally designed as such. A notable early exception is the work by Schmidt-Pruzan and Shamir [49]. They fixed the exact/expected edge sequence such that the largest edge has cardinality $\mathrm{O}(\log n)$ and showed a "double jump" phase transition in the size of the largest connected component. Most of the recent literature on random hypergraphs concentrates on the k-uniform model where every edge has exactly k vertices [34, 5, 6] or, equivalently, on random binary matrices with k 1s per column [17]. In our model, we do not prescribe the exact cardinalities of the edges and neither do we bound their maximum size, instead we only require that the *expected* edge size is pn.

Probably closest to our work is a string of articles by Demetrovics et al. [20] as well as Katona [35, 36]. They investigated random databases and connected the Rényi entropy of *order 2* of the logarithmic number of rows with the probability that certain unique column combinations or functional dependencies hold. In contrast, we connect the Shannon entropy of the logarithmic number of *pairs* of rows with the expected number of minimal difference sets. Unique column combinations and functional dependencies are dual to the difference sets of record pairs, one are the minimal hitting sets of the other [1, 9]. Also, the Shannon entropy is the same as the Rényi entropy of *order 1* [18]. In this sense, we complement the result by Demetrovics et al. by showing that the duality also pertains to the order of entropy.

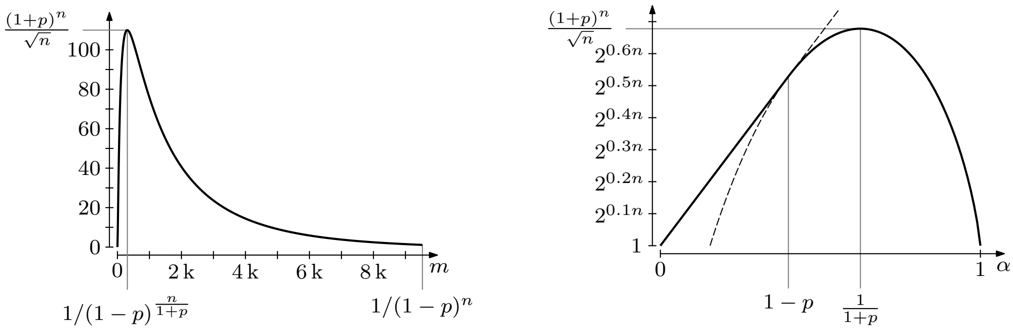
The analysis of random (hyper-)graphs naturally uses tools from combinatorics and probability theory. Conversely, it has always helped to advance the fields by sharpening those tools [7, 12, 31]. In this work, we improve the bounds of the Chernoff–Hoeffding theorem [30] on the tail of the binomial distribution. We use an observation by Klar [38] on the relation between the distribution function and the probability mass function. There were some refined inequalities known before. By Cramér's theorem [19], Chernoff–Hoeffding is asymptotically tight up to subexponential factors. The gap was subsequently reduced to order $\mathrm{O}(\sqrt{n})$, cp. [3], we close it down to a constant. There also exist some comparatively tight bounds based on the normal limit of the binomial distribution, contributions by Prokhorov [47] and Slud [51] founded major lines of research. However, we avoid this approach since the normal approximation cannot be expressed in terms of elementary functions. Also, it tends to place unnecessary restrictions on the success probability p when deriving non-asymptotic results.

Outline. Next, we introduce the hypergraph model and state our results in full detail. We review some notation in Section 3. Section 4 is dedicated to the Chernoff–Hoeffding theorem. Section 5 adds further technical contributions, including the sharp threshold of minimal sets at a certain cardinality. The main theorem is proven in Section 6. Section 7 discusses the phase transition and concludes the work.

2 Model and Main Theorem

Fix a probability p and positive integers n and m. The random multi-hypergraph $\mathcal{B}_{n,m,p}$ is defined by independently sampling m (not necessarily distinct) subsets of $[n]$. Each set is generated by including a vertex $v \in [n]$ with probability p independently of all other choices.

We quickly argue that this is indeed the maximum-entropy model. Besides the size of the universe n and the number of edges m, the only other constraint is the expected edge size pn. The independence bound on the entropy reads as follows: Let X_1 to X_m be random variables with joint distribution $\mathrm{P}_{X_1,\ldots,X_m}$ and marginal distributions P_{X_j}. Then, their entropies observe the inequality $H(\mathrm{P}_{X_1,\ldots,X_m}) \leq \sum_{j=1}^{m} H(\mathrm{P}_{X_j})$, equality holds if and only if the X_j are independent, see [18]. This suggests that we should choose the edges independently if we



(a) Our bound as a function of m for $n = 10$ and $p = 0.6$ in the information-theoretic regime. The vertical line at $m = 1/(1-p)^{\frac{n}{1+p}}$ indicates the maximum (Theorem 2). For $m > 1/(1-p)^n$, the size goes to 1. The linear bound for $m \le 1/(1-p)^{(1-p)n}$ is not shown as it is too close to 0.

(b) Our bound as a function of α for $p = 0.6$ (the plot is independent of n). The vertical line at $\alpha = 1/(1+p)$ indicates the maximum (Theorem 2). For $\alpha \le 1-p$, the linear bound holds, for larger α, we get the information-theoretic bound. They are continued as dashed lines into the other regime.

Figure 1 Illustration of Theorem 1 showing the expected size of the minimization of a random hypergraph depending on the number of edges m (a) and on α (b). As α grows logarithmically in m, (b) shows the same plot as (a) but with both axes being logarithmic.

want to maximize the entropy and the same is true for the vertices inside an edge. Finally, the fact that setting the sampling probability to be equal for all vertices indeed maximizes the entropy under a given mean set size was proven by Harremoës [29].

We are interested in the expected number of inclusion-wise minimal sets in $\mathcal{B}_{n,m,p}$, denoted by $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|]$. We describe the asymptotic behavior of this expectation with respect to n. In more detail, we view $m = m(n)$ as a function of n assuming integer values and bound the univariate asymptotics of $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|]$ in n for different choices of m. The probability p is considered to be a constant throughout. We show that the size of the minimization can be described precisely in terms of p and the Shannon entropy of the logarithm of m.

We let $\mathrm{H}(x) = H((x, 1-x))$ denote the binary entropy function and define the quantity $\alpha = \log_{\frac{1}{(1-p)^n}} m = -(\log_{1-p} m)/n$. The quantity α is well-defined for all $0 < p < 1$ and n, $m \ge 1$. It is always non-negative and asymptotically of order $\Theta((\log m)/n)$. If p and n are fixed, choosing a value for α determines m since we can rewrite m as $1/(1-p)^{\alpha n}$.

▶ **Theorem 1.** *Let p be a probability, and n, m be two positive integers. If $p = 0$ or $p = 1$, then $|\min(\mathcal{B}_{n,m,p})| = 1$ holds deterministically. For $0 < p < 1$, the following statements hold.*

1. *If $m \le 1/(1-p)^{(1-p)n}$, then $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] = \Theta(m)$.*
2. *For any two $\varepsilon, \varepsilon' > 0$ and all m such that $1/(1-p)^{(1-p+\varepsilon)n} \le m \le 1/(1-p)^{(1-\varepsilon')n}$, i.e., all α such that $1-p+\varepsilon \le \alpha \le 1-\varepsilon'$, we have*

$$\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] = \Theta\Big(2^{(\mathrm{H}(\alpha)+(1-\alpha)\log_2 p)\,n} \Big/ \sqrt{n}\Big) = \Theta\left(\left(\frac{p^{1-\alpha}}{(1-\alpha)^{1-\alpha}\,\alpha^{\alpha}}\right)^{n} \Big/ \sqrt{n}\right);$$

3. *If $m = 1/(1-p)^{n+\omega(\log n)}$, then $1 \le \mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] = 1 + \mathrm{o}(1)$.*

The bounds in the distinct cases are very different in nature. They are visualized in Figure 1 showing the expectation both as a function of the number of trials m and of α. To distinguish the behavior also in writing, we use the term *linear regime* if m is between 1 and $1/(1-p)^{(1-p)n}$, corresponding to $0 \le \alpha \le 1-p$, likewise, we refer to m being between $1/(1-p)^{(1-p)n}$ and $1/(1-p)^n$, i.e., $1-p \le \alpha \le 1$, as the *information-theoretic regime.*

All asymptotic estimates in Theorem 1 are at least tight up to constants, the third statement is even tight up to lower-order terms. The constants hidden in the big-O-notation are universal in the sense that they do not depend on m or n, and also not on α describing the relation between the former two. However, they may depend on the probability p and, in case of Statement 2, on the particular choices for ε and ε'. We note that the bounds for the information-theoretic regime have two gaps at $m = 1/(1-p)^{(1-p)n}$ and $m = 1/(1-p)^n$. These gaps can be made arbitrarily small: Let $c = 1/(1-p)$, then Statement 2 holds if $m \le (c-\gamma)^n$ for any constant $\gamma > 0$ and Statement 3 takes over at $m \ge (c + \delta(n))^n$, where $\delta(n)$ is a function converging to 0 as n increases.

From the main theorem, we derive bounds on the maximum expectation over all m.

▶ **Theorem 2.** *If $p = 0$ or $p = 1$, then $\max_{m\ge1} |\min(\mathcal{B}_{n,m,p})| = 1$. For $0 < p < 1$, we have $\max_{m\ge1} \mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] = \Theta((1+p)^n/\sqrt{n})$, attained at $m = 1/(1-p)^{\frac{n}{1+p}}$.*

3 Preliminaries and Notation

Multi-Hypergraphs. A *hypergraph* on $[n] = \{1,\dots,n\}$ is a set of subsets $\mathcal{H} \subseteq \mathcal{P}([n])$, called the *(hyper-)edges.* If $\mathcal{H}$ is a multiset instead, we have a *multi-hypergraph.* We do not allow multiple copies of the same vertex in one edge. The *minimization* of a hypergraph $\mathcal{H}$ is the collection of its inclusion-wise minimal edges, $\min(\mathcal{H}) = \{E \in \mathcal{H} \mid \forall E' \in \mathcal{H} \colon E' \subseteq E \Rightarrow E' = E\}$. We extend this notion to multi-hypergraphs by requiring that whenever a minimal edge has multiple copies, only one of them is included in the minimization. This way $\min(\mathcal{H})$ is always a mere hypergraph (a set). For a multi-hypergraph $\mathcal{H}$, we use $|\mathcal{H}|$ to denote the total number of edges counting multiplicities, and $\|\mathcal{H}\|$ for the number of *distinct* edges, i.e., the cardinality of the support of $\mathcal{H}$. Evidently, we have $|\min(\mathcal{H})| \le \|\mathcal{H}\| \le |\mathcal{H}|$.

Information Theory. We intend the expressions $0 \cdot \log_a 0$ and $0 \cdot \log_a(\frac{0}{0})$ to all mean 0 for any positive real base $a > 0$. Note that this convention also implies $0^0 = a^{0 \log_a 0} = 1$ and $(\frac{0}{0})^0 = 1$. We use $\operatorname{ld} x$ for the binary (base-2) logarithm of x. The (binary) *entropy function* H is defined for all probabilities x as $\mathrm{H}(x) = -x \operatorname{ld} x - (1-x) \operatorname{ld}(1-x)$. It describes the Shannon entropy or, equivalently, the Rényi entropy of order 1, of the Bernoulli distribution with parameter x. In the notation of the previous sections, $\mathrm{H}(x) = H((x, 1-x))$. Evidently, the entropy function is symmetric around 1/2 with $\mathrm{H}(x) = \mathrm{H}(1-x)$. On the open unit interval, H is positive and differentiable with derivative $\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{H}(x) = \operatorname{ld}(\frac{1-x}{x})$. This is the negative (binary) *logit function*, also dubbed *log-odds* in statistics. H is strictly concave and has its maximum at 1/2 with value $\mathrm{H}(1/2) = 1$. The *perplexity* of x is $2^{\mathrm{H}(x)} = 1/(x^x(1-x)^{1-x})$. We utilize it to estimate binomial coefficients. The bounds are well-known in the literature [18].

▶ **Lemma 3.** *Let n be a positive integer and $0 < x < 1$ such that xn is an integer, then $2^{\mathrm{H}(x)n}/\sqrt{8nx(1-x)} \le \binom{n}{xn} \le 2^{\mathrm{H}(x)n}/\sqrt{\pi nx(1-x)}$.*

Let $(p_i)_i$ and $(q_i)_i$ be two distributions on the same state space such that $(p_i)_i$ is absolutely continuous with respect to $(q_i)_i$, i.e., $q_i = 0$ implies $p_i = 0$ for all i. The (binary) *Kullback–Leibler divergence*[4] from $(q_i)_i$ to $(p_i)_i$ is given by $D((p_i)_i \,\|\, (q_i)_i) = -\sum_i p_i \operatorname{ld}(\frac{q_i}{p_i})$. It is the expected information loss when assuming that the distribution is $(q_i)_i$ while the system actually follows $(p_i)_i$. The divergence is a premetric in that it is non-negative and 0 iff the distributions are the same. However, it is neither symmetric nor does it observe the triangle

[4] The divergence is sometimes also called *relative entropy*, we avoid this term due to ambiguities, cf. [18].

inequality. In this work, we only need the divergence between Bernoulli distributions. For any two probabilities x, y, the divergence between two Bernoulli distributions with respective parameters x and y is $\mathrm{D}(x \,\|\, y) = D((x, 1-x) \,\|\, (y, 1-y)) = -x \operatorname{ld}\left(\frac{y}{x}\right) - (1-x) \operatorname{ld}\left(\frac{1-y}{1-x}\right)$. The function $\mathrm{D}(x \,\|\, y)$ is convex in both x and y, attains its minimum 0 for $x = y$, and observes $\mathrm{D}(x \,\|\, y) = \mathrm{D}(1-x \,\|\, 1-y)$. We often use the derived quantity $2^{-\mathrm{D}(x \,\|\, y)} = \left(\frac{y}{x}\right)^x \left(\frac{1-y}{1-x}\right)^{1-x}$.

Polynomials of Probabilities.

▶ **Lemma 4.** *Let n be a non-negative integer and x a probability, then it holds that $e^{-nx}\left(1 - nx^2\right) \le (1-x)^n \le e^{-nx}$.*

▶ **Lemma 5** (Lemma 10 in [4])**.** *Let n be a non-negative integer and x a probability, then $nx/(1+nx) \le 1-(1-x)^n \le nx$.*

▶ **Lemma 6.** *Consider a random experiment with outcomes A, B, and C, where $\mathrm{P}[B] > 0$. In a series of m i.i.d. trials, let A_j denote the event that the outcome of the j-th trial is A, same with B. Then, we have $\mathrm{P}[\forall j \le m\colon \neg A_j \mid \exists k \le m\colon B_k] \le \mathrm{P}[\forall j \le m\colon \neg A_j \mid B_m]$.*

4 The Chernoff–Hoeffding Theorem

In this section, we tighten the Chernoff–Hoeffding theorem bounding the tail of the binomial distribution. The result will later help us with the random hypergraphs, but more importantly it provides a powerful tool of general interest in probability theory. Fix a positive integer n and probabilities x and p. Recall that the Kullback–Leibler divergence between the respective Bernoulli distributions is $\mathrm{D}(x \,\|\, p) = -x \operatorname{ld}\left(\frac{p}{x}\right) - (1-x) \operatorname{ld}\left(\frac{1-p}{1-x}\right)$. The Chernoff–Hoeffding theorem [30, 22] employs the divergence to bound the probability that a binomially distributed random variable $Y \sim \mathrm{Bin}(n,p)$ deviates from its expected value $\mathrm{E}[Y] = pn$. If $x \le p$, then $\mathrm{P}[Y \le xn] \le 2^{-\mathrm{D}(x \,\|\, p)\, n} = \left(\frac{p}{x}\right)^{xn} \left(\frac{1-p}{1-x}\right)^{(1-x)n}$. Similarly, if $p \le x$, we have $\mathrm{P}[Y \ge xn] \le 2^{-\mathrm{D}(x \,\|\, p)\, n}$. Several weaker but more practical inequalities have been inferred from this, summarized as *Chernoff bounds* [41, 21]. We sharpen these inequalities by a $\sqrt{n}$-factor for all but the extreme values of x. While the upper bound of Chernoff–Hoeffding holds for all probabilities x, there are some lower bounds known for $\mathrm{P}[Y \le xn]$ if the product xn is an integer, c.f. the textbook by Ash [3, Lemma 4.7.2]. We use a proposition by Klar [38] to improve the upper bound such that it matches the lower one up to constants. We then extend both bounds to the general case of arbitrary products xn.

▶ **Theorem 7.** *Let n be a positive integer, x and p two probabilities with $0 < p < 1$, and $Y \sim \mathrm{Bin}(n,p)$ a binomial random variable.*

1. *If $1/n \le x < p$, then* $\frac{(1-p)\sqrt{x}}{2e\sqrt{2\,(1-x)}} \cdot \frac{2^{-\mathrm{D}(x \,\|\, p)\, n}}{\sqrt{n}} \le \mathrm{P}[Y \le xn] \le \frac{\sqrt{1-x}}{(p-x)\sqrt{\pi\, x}} \cdot \frac{2^{-\mathrm{D}(x \,\|\, p)\, n}}{\sqrt{n}}$.
2. *If $p < x \le 1 - 1/n$, then* $\frac{p\sqrt{1-x}}{2e\sqrt{2\, x}} \cdot \frac{2^{-\mathrm{D}(x \,\|\, p)\, n}}{\sqrt{n}} \le \mathrm{P}[Y \ge xn] \le \frac{\sqrt{x}}{(x-p)\sqrt{\pi\,(1-x)}} \cdot \frac{2^{-\mathrm{D}(x \,\|\, p)\, n}}{\sqrt{n}}$.

Proof sketch. The second statement of the theorem is implied by the first one by applying it to the complementary variable $\overline{Y} \sim \mathrm{Bin}(n, 1-p)$. Let $x \le p$ be a probability. We mainly confine ourselves here to the case that the product xn is integral and show that then we get $\frac{1}{\sqrt{8\,x(1-x)}} \cdot \frac{2^{-\mathrm{D}(x \,\|\, p)\, n}}{\sqrt{n}} \le \mathrm{P}[Y \le xn] \le \frac{p\sqrt{1-x}}{(p-x)\sqrt{\pi\, x}} \cdot \frac{2^{-\mathrm{D}(x \,\|\, p)\, n}}{\sqrt{n}}$.

Lemma 3 provides the following error bounds for the probability mass function of Y: $1/\sqrt{8nx(1-x)} \leq \mathrm{P}[Y = xn]/(2^{\mathrm{H}(x)n} \cdot p^{xn}(1-p)^{(1-x)n}) \leq 1/\sqrt{\pi n x(1-x)}$. We further have $2^{\mathrm{H}(x)n} \cdot p^{xn}(1-p)^{(1-x)n} = 2^{-\,\mathrm{D}(x \,\|\, p)\, n}$. This proves the first part that $\mathrm{P}[Y \leq xn] \geq \mathrm{P}[Y = xn] \geq 2^{-\,\mathrm{D}(x \,\|\, p)\, n}/\sqrt{8\,n\,x(1-x)}$ holds.

A result by Klar [38, Proposition 1 (c)] states that the ratio $\mathrm{P}[Y \leq xn]/\,\mathrm{P}[Y = xn]$ is at most $f_{n,xn}(p) = p\,(1 - \frac{xn}{n+1})/(p - \frac{xn}{n+1})$. The partial discrete derivative of $f_{n,xn}$ with respect to n, that is, $\Delta_n(f_{n,xn})(p) = f_{n+1,x(n+1)}(p) - f_{n,xn}(p)$ can be shown to be positive whenever $x < p$. Thus $f_{n,xn}(p)$ converges from below to $p(1-x)/(p-x)$ as n increases. Combined with the error bounds this is $\mathrm{P}[Y \leq xn] \leq \frac{p\,(1-x)}{p-x} \cdot \frac{1}{\sqrt{\pi\,n\,x(1-x)}} \cdot 2^{-\,\mathrm{D}(x \,\|\, p)\, n} = \frac{p\,\sqrt{1-x}}{(p-x)\,\sqrt{\pi\,x}} \cdot \frac{2^{-\mathrm{D}(x \,\|\, p)\, n}}{\sqrt{n}}$.

Transferring the improvements also to non-integral products xn is not straightforward. A careful analysis of the monotonicity of the entropy function H as well as that of the divergence D reveals that this transition weakens the upper bound only by a additional factor of $1/p$ and the lower bound by $x\,(1-p)/e$, independently of n. ◀

We showed that for all values x strictly between 0 and p, the Chernoff–Hoeffding theorem can be asymptotically improved to $\mathrm{P}[Y \leq xn] = \Theta(2^{-\mathrm{D}(x \,\|\, p)\, n}/\sqrt{n})$. However, the constants hidden in the big-O notation diverge at the boundaries. This caveat cannot be healed, there is no way to extend the improvement also to $x = 0$ or p. Simply put, the original formulation of the theorem is tight. First, pn is not only the mean but also the median of the binomial distribution, whence $\mathrm{P}[Y \leq pn] \geq 1/2$ is constant and *not* of order $\mathrm{O}(2^{-\mathrm{D}(p \,\|\, p)\, n}/\sqrt{n}) = \mathrm{O}(1/\sqrt{n})$. Secondly, the initial bound $\mathrm{P}[Y \leq 0] = (1-p)^n = 2^{-\mathrm{D}(0 \,\|\, p)\, n}$ even is exact.

5 Distinct Sets and Minimality

We now return to the main topic of this work, which is determining the expected size of the minimization $\min(\mathcal{B}_{n,m,p})$ of the maximum-entropy multi-hypergraph $\mathcal{B}_{n,m,p}$. The sampling probabilities $p = 0$ or $p = 1$ are trivial, we thus assume $0 < p < 1$ in this work unless explicitly stated otherwise. Every subset of $[n]$ then has a non-vanishing chance to be sampled. Such a set is minimal for $\mathcal{B}_{n,m,p}$ iff it is generated in one of the trials and no proper subset ever occurs. Both of these aspects influence the chance of minimality, but their impact varies depending on the cardinality of the set in question. The number of vertices per edge is heavily concentrated around pn and the more vertices there are in an edge, the less likely it is minimal. Intuitively, almost no sets with very low cardinalities are sampled, but if so, they are often included in $\min(\mathcal{B}_{n,m,p})$. There are plenty of edges with a medium number of vertices and there is a good chance they are minimal. Finally, sets of very high cardinality rarely occur and usually they are then dominated by smaller ones. This disparity is exacerbated by a large number of trials. Boosting m increases the probability that also sets of cardinality a bit further away from pn are sampled, at the same time the process generates more duplicates of sets that occurred before. More importantly though, the likelihood of a larger set being minimal is even smaller with many trials. Eventually, the last effect outweighs all others, creating a situation in which the only minimal edge is empty.

We start making this intuition rigorous by giving preliminary bounds on the number of minimal edges as a first step towards the proof of Theorem 1. The results are binomial sums of polynomials of probabilities, depending on which factors we choose, we get an upper or a lower bound. The estimates are already tight up to constants but are rather unwieldy. They will serve as the basis for our further analysis. Let $\mathcal{D}_{n,p}$ denote the maximum-entropy distribution on the power set $\mathcal{P}([n])$ provided that $\mathrm{E}_{X \sim \mathcal{D}_{n,p}}[|X|] = pn$, meaning each vertex is included independently with probability p.

▶ **Lemma 8.** *Let $0 < p < 1$ be a probability, n, m positive integers, and let $X_j \sim \mathcal{D}_{n,p}$ denote the outcome of the j-th independent trial. For any integer i with $0 \le i \le n$, define $s_{n,p}(i,m) = \mathrm{P}[\exists j \le m \colon X_j = [i]]$ and $w_{n,p}(i,m) = \mathrm{P}[\forall j \le m \colon \neg(X_j \subsetneq [i])]$ to be the respective probabilities*[5] *that some trial produces the set $[i]$ and no trial produces a proper subset of $[i]$. Then, we have $s_{n,p}(i,m) = 1 - (1 - p^i(1-p)^{n-i})^m$ and $w_{n,p}(i,m) = \Big(1 - (1-p)^{n-i}(1-p^i)\Big)^m$. Furthermore, the following statements hold.*

1. $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] \ge \sum_{i=0}^{n} \binom{n}{i} s_{n,p}(i,m) \cdot w_{n,p}(i,m)$.
2. $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] \le \sum_{i=0}^{n} \binom{n}{i} s_{n,p}(i,m) \cdot w_{n,p}(i,m-1)$.
3. $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] \le 1 + \frac{1}{p} \sum_{i=0}^{n} \binom{n}{i} s_{n,p}(i,m) \cdot w_{n,p}(i,m)$.

Proof sketch. The formula for $w_{n,p}(i,m) = \mathrm{P}[\forall j \le m \colon \neg(X_j \subsetneq [i])]$ can be seen as follows. The random set $X_j \sim \mathcal{D}_{n,p}$ is a subset of $[i]$ iff it does not contain an element of $[n]\backslash[i]$, which happens with probability $(1-p)^{n-i}$. Conditioned on being any subset, X_j is a *proper* subset if it is missing at least one element of $[i]$, having conditional probability $1 - p^i$.

Regarding the main statements, a set $S \subseteq [n]$ is in $\min(\mathcal{B}_{n,m,p})$ iff it is sampled in one of the m trials and no proper subset is sampled. The probability for both events depends only on the cardinality $|S|$: $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] = \sum_{S \subseteq [n]} \mathrm{P}[\exists k \le m \colon X_k = S \wedge \forall j \le m \colon \neg(X_j \subsetneq S)] = \sum_{i=0}^{n} \binom{n}{i} \cdot \mathrm{P}[\exists k \le m \colon X_k = [i]] \cdot \mathrm{P}[\forall j \le m \colon \neg(X_j \subsetneq [i]) \mid \exists k \le m \colon X_k = [i]]$. Generating any *other* set than $[i]$ in a single trial has probability $1 - p^i(1-p)^{n-i}$, over the independent trials we thus get $s_{n,p}(i,m) = \mathrm{P}[\exists j \le m \colon X_j = [i]] = 1 - (1 - p^i(1-p)^{n-i})^m$

The last factor $\mathrm{P}[\forall j \le m \colon \neg(X_j \subsetneq [i]) \mid \exists k \le m \colon X_k = [i]]$ in each term describes the likelihood that the set $[i]$ is minimal, conditioned on it being sampled at all. Conditioning on at least one trial producing $[i]$ itself only increases the chances of never sampling a proper subset, which gives Statement 1. To prove Statement 2, we apply Lemma 6. Statement 3 follows from the ratio between $w_{n,p}(i,m)$ and $w_{n,p}(i,m-1)$ being the probability that a non-subset of $[i]$ is sampled in a single trial. ◀

The part that all three bounds of Lemma 8 have in common describes the expected number of *distinct* sets in $\mathcal{B}_{n,m,p}$. Recall that we use $\|\mathcal{H}\|$ to denote the number of distinct sets of some multi-hypergraph $\mathcal{H}$. That means, we have $\mathrm{E}[\|\mathcal{B}_{n,m,p}\|] = \sum_{i=0}^{n} \binom{n}{i} s_{n,p}(i,m)$. We weighted the terms of this sum by $w_{n,p}(i,m)$ or $w_{n,p}(i,m{-}1)$, respectively. In the following, we analyze the two parts separately, starting with the *weighting factors* $w_{n,p}$. They are of interest beyond their application to random multi-hypergraphs. Consider m trials according to the maximum-entropy distribution $\mathcal{D}_{n,p}$ on subsets of $[n]$ with expected set size pn. The quantity $w_{n,p}(i,m)$ is, by definition, the probability that any fixed subset of cardinality i survives as minimal after all trials. Equivalently, $1 - w_{n,p}(i,m)$ is the probability of any proper subset being sampled. We prove next that the weighting factors are in fact threshold functions falling abruptly from almost 1 to almost 0 as i increases from 0 to n, the position of the transition depends on n, m, and p. Recall that α abbreviates $-(\log_{1-p} m)/n$. Lemma 9 below establishes a sharp threshold at $i^* = n + \log_{1-p} m = (1-\alpha)n$. Note that i^* is always at most n since $\log_{1-p} m$ is non-positive. The definition ensures the equality $m = 1/(1-p)^{n-i^*} = 1/(1-p)^{\alpha n}$. For increasing m, the threshold gets smaller relative to n. Once m grows beyond $1/(1-p)^n$, i.e., $\alpha > 1$, the quantity i^* can no longer be interpreted as a cardinality as it becomes negative. Later, in Lemma 12, we will see that m being this large is in fact irrelevant for the minimization.

[5] The notation $s_{n,p}$ refers the set being *sampled*; these probabilities are then *weighted* by the factors $w_{n,p}$.

▶ **Lemma 9.** *Let $0 < p < 1$ be a probability, and n, m positive integers, then $w_{n,p}(0, m) = 1$, and $w_{n,p}(n, m) = p^{nm}$. Now let $i = i(n)$ with $0 < i < n$ be a function taking integer values.*

1. *We have* $\exp(-m\,(1-p)^{n-i}) \cdot (1 - m\,(1-p)^{2(n-i)}) \le w_{n,p}(i, m) \le \exp(-m\,(1-p)^{n-i+1})$.

In particular, the following statements hold.[6]

2. *If $i = n + \log_{1-p} m + \omega(1)$, then $\lim_{n\to\infty} w_{n,p}(i, m) = 0$.*
3. *If $i = n + \log_{1-p} m - \omega(1)$, then $\lim_{n\to\infty} w_{n,p}(i, m) = 1$.*
4. *If $i = n + \log_{1-p} m \pm \Theta(1)$, then $w_{n,p}(i, m) = \Theta(1)$.*

Proof. Suppose $0 < i < n$, we estimate $w_{n,p}(i, m)$ using mainly Lemma 4. This yields $w_{n,p}(i, m) = (1-(1-p)^{n-i}(1-p^i))^m \le (1-(1-p)^{n-i}(1-p))^m \le \exp\Big(-m\,(1-p)^{n-i}\cdot(1-p)\Big)$. Since $1 - p$ is constant, the limit behavior is entirely determined by the product $m\,(1-p)^{n-i}$. If $i = n + \log_{1-p} m + \omega(1)$, then $m\,(1-p)^{n-i} = m\,(1-p)^{n-n-(\log_{1-p} m)-\omega(1)} = (1-p)^{-\omega(1)}$ diverges and thus the weighting factor $w_{n,p}(i, m)$ converges to 0. Conversely, from $1 - p^i \le 1$ we get that $w_{n,p}(i, m) \ge (1 - (1-p)^{n-i})^m \ge \exp\Big(-m\,(1-p)^{n-i}\Big) \cdot (1 - m\,(1-p)^{2(n-i)})$. If $i = n + \log_{1-p} m - \omega(1)$, both $m\,(1-p)^{n-i} = (1-p)^{\omega(1)}$ and $m\,(1-p)^{2(n-i)} = (1-p)^{\omega(1)}/m$ tend to 0, implying $\lim_{n\to\infty} w_{n,p}(i, m) = 1$.

Finally, if the cardinality i is around the threshold $i^* = n + \log_{1-p} m$, the limit may not exist. We show that $w_{n,p}(i, m)$ is still bounded away from 0. Suppose $i = n + \log_{1-p} m \pm \Theta(1)$; in particular, the difference $i^* - i$ is bounded for all n. If m is constant w.r.t. n, so is $w_{n,p}(i, m) \ge (1 - (1-p)^{n-i})^m \ge p^m$. Here, we used the assumption $i < n$. Finally, if m diverges, then $n - i = \log_{1-p} m \mp \Theta(1) = \omega(1)$ diverges with it. Together with the fact that $m\,(1-p)^{n-i} = (1-p)^{i^*-i}$ holds by the definition of i^*, we get that $w_{n,p}(i, m)$ is bounded since $w_{n,p}(i, m) \ge \exp\Big(-(1-p)^{i^*-i}\Big) \cdot (1 - (1-p)^{(i^*-i)+(n-i)}) = \Omega(1)$. ◀

After we have shown the existence of a sharp threshold for the weighting factors, we now treat the number of distinct sets $\|\mathcal{B}_{n,m,p}\|$ in the multi-hypergraph. This is a natural upper bound for the size of the minimization. In turn, a trivial cap for the number of distinct sets is the total number of sets $|\mathcal{B}_{n,m,p}| = m$. When starting the sampling, many different sets are generated and $\|\mathcal{B}_{n,m,p}\|$ is indeed close to m. As the number of trials increases though, duplicates occur in the sample and the two quantities grow apart.

To discuss this in more detail, we introduce some notation. For a pair of integers ℓ, u with $0 \le \ell \le u \le n$, let $\|\mathcal{B}_{n,m,p}(\ell, u)\|$ denote the number of distinct sampled sets whose cardinality is between ℓ and u, including. This is also at most as large as the total number of samples in that range. It thus makes sense to expect an upper bound in terms of the binomial distribution. We confirm this below and further prove that there is also a lower bound of the same flavor.

▶ **Lemma 10.** *Let $0 < p < 1$ be a probability, n, m positive integers, and $Y \sim \mathrm{Bin}(n, p)$ a binomially distributed random variable with parameters n and p. Let ℓ, u be integers such that $0 \le \ell \le u \le n$ and define $\mathbf{p} = \max_{\ell \le i \le u}\{p^i(1-p)^{n-i}\}$. Then, $\mathbf{p}$ is equal to $p^\ell(1-p)^{n-\ell}$ if $p \le 1/2$; otherwise, we have $\mathbf{p} = p^u(1-p)^{n-u}$. Further, the expected number of distinct sets in $\mathcal{B}_{n,m,p}$ with cardinality between ℓ and u observes $\frac{m}{1+m\mathbf{p}} \cdot \mathrm{P}[\ell \le Y \le u] \le \mathrm{E}[\,\|\mathcal{B}_{n,m,p}(\ell, u)\|\,] \le m \cdot \mathrm{P}[\ell \le Y \le u]$.*

---

[6] We understand $\omega(1)$ as the class of all non-negative unbounded functions of n. In particular, the classes $n + \log_{1-p} m + \omega(1)$ and $n + \log_{1-p} m - \omega(1)$ are disjoint.

6 Proof of the Main Theorem

We prove the main results on the expected size of the minimization of the hypergraph $\mathcal{B}_{n,m,p}$ in this section with the help of the tools above. The key observation is that the minimization is dominated by the sets with cardinalities around the threshold $i^* = n + \log_{1-p} m$.

6.1 The Lower Bound

We prove the main results, Theorem 1, with the help of the tools above. The key observation is that the minimization is dominated by the sets with cardinalities around the threshold $i^* = n + \log_{1-p} m$ of the weighting factors. We will see that the distinct edges make up a constant fraction of $\mathcal{B}_{n,m,p}$ as long as m is at most $1/(1-p)^{(1-p)n}$. In turn, a constant fraction of those distinct edges are indeed minimal. However, the linear growth of $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|]$ cannot be maintained for a larger sample size. We prove that once m is so large that the threshold i^* is below pn, the ratio of minimal edges decreases significantly. The minimization then enters a regime governed by the entropy of $\alpha = -(\log_{1-p} m)/n$.

The next lemma shows both lower bounds of Theorem 1 together. The information-theoretic one is slightly more general than what was stated in Theorem 1.2 in that it pertains to all m between $1/(1-p)^{(1-p)n}$ and $1/(1-p)^{(1-\varepsilon')n}$. Let H denote the entropy function.

▶ **Lemma 11** (Theorem 1.1 and the lower bound of Theorem 1.2). *Let $0 < p < 1$. If $m \le 1/(1-p)^{(1-p)n}$, then $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] = \Theta(m)$. For any $\varepsilon' > 0$ and m such that $1/(1-p)^{(1-p)n} \le m \le 1/(1-p)^{(1-\varepsilon')n}$, corresponding to $1-p \le \alpha \le 1-\varepsilon'$, we have $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] = \Omega\big(2^{(\mathrm{H}(\alpha)+(1-\alpha)\,\mathrm{ld}\,p)\,n}/\sqrt{n}\big)$.*

Proof sketch. The sought expectation is at least as large as the number of distinct sets up to some cardinality i that are minimal after m trials for arbitrary values of i. As an ansatz, we choose this to be the threshold $i^* = n + \log_{1-p} m$. Lemmas 8 and 10 together then imply that $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] \ge (m/(1+m\mathbf{p})) \cdot \mathrm{P}[Y \le i^*] \cdot w_{n,p}(i^*, m)$. It can be shown via Lemma 10 that the denominator $1 + m\mathbf{p}$ is at most 2 as long as $m \le 1/(1-p)^n$. Lemma 9.1 shows that there exists a universal constant $\delta > 0$ (again for all $m \le 1/(1-p)^n$) such that $w(i^*, m) \ge \delta$.

The bounds in the two regimes differ in the way the product $m \cdot \mathrm{P}[Y \le i^*]$ is estimated. If $m \le 1/(1-p)^{(1-p)n}$, then $i^* \ge pn$ is at least as large as the median of Y, whence $m \cdot \mathrm{P}[Y \le i^*] \ge m/2$. This gives the lower bound in the linear regime. In the information-theoretic regime, we use the rewrite $m = 1/(1-p)^{\alpha n}$. Suppose first that there are constants $\varepsilon, \varepsilon' > 0$ such that $1 - p + \varepsilon \le \alpha \le 1 - \varepsilon'$ holds. These are exactly the prerequisites of Theorem 1.2. We apply the improved lower bound of the Chernoff–Hoeffding theorem, Theorem 7.1. Let D denote the Kullback–Leibler divergence. There exists a positive constant $C > 0$–independent of n and m but possibly dependent on p, ε, and ε'–such that

$$m \cdot \mathrm{P}[Y \le i^*] = m \cdot \mathrm{P}[Y \le (1-\alpha)n] \ge m \cdot C\, \frac{2^{-\mathrm{D}(1-\alpha\,\|\,p)\,n}}{\sqrt{n}}$$
$$= \frac{1}{(1-p)^{\alpha n}} \cdot \frac{C}{\sqrt{n}} \left(\frac{p}{1-\alpha}\right)^{(1-\alpha)n} \left(\frac{1-p}{\alpha}\right)^{\alpha n} = \frac{C}{\sqrt{n}} \cdot \left(\frac{p^{1-\alpha}}{(1-\alpha)^{1-\alpha}\,\alpha^{\alpha}}\right)^{n}.$$

The latter expression equals $C \cdot 2^{(\mathrm{H}(\alpha)+(1-\alpha)\,\mathrm{ld}\,p)\,n}/\sqrt{n}$. Finally, if $m(1-p)^{(1-p)n}$ converges to 1 from above, i.e., $\alpha \searrow 1-p$, then the result follows from a direct application of Lemma 3. ◀

6.2 The Upper Bound

The upper bound draws from the same core observations as the lower one: the threshold position of the weighting factors $w_{n,p}$ and the proportion of distinct sets in the sample. First, we show that once m is more than a polynomial factor larger than $1/(1-p)^n$, the minimization essentially consists of a single edge, the empty set. Lemma 13 then proves our claim that the information-theoretic lower bound is tight beyond the phase transition.

▶ **Lemma 12** (Theorem 1.3). *If* $m = 1/(1-p)^{n+\omega(\log n)}$, *then* $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] = 1 + \mathrm{o}(1)$.

▶ **Lemma 13** (Upper bound of Theorem 1.2). *Let* $0 < p < 1$ *be a probability,* $\varepsilon, \varepsilon' > 0$ *positive reals, and* n*,* m *positive integers such that* m *is between* $1/(1-p)^{(1-p+\varepsilon)n}$ *and* $1/(1-p)^{(1-\varepsilon')n}$*, i.e.,* $1-p+\varepsilon \le \alpha \le 1-\varepsilon'$*. Then, we have* $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] = \mathrm{O}\big(2^{(\mathrm{H}(\alpha)+(1-\alpha)\,\mathrm{ld}\,p)\,n}/\sqrt{n}\big)$.

Proof sketch. We get $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|] \le \sum_{i=0}^{n} \binom{n}{i}\big(1-(1-p^i(1-p)^{n-i})^m\big) \cdot w_{n,p}(i, m-1)$ from Lemma 8.2. The idea of this proof is to split the sum at the threshold $i^* = (1-\alpha)n$ and handle the two parts separately. Let $Y \sim \mathrm{Bin}(n,p)$ be a binomial variable. Lemma 10 shows for the first part that $\sum_{i=0}^{i^*} \binom{n}{i}(1-(1-p^i(1-p)^{n-i})^m) \cdot w_{n,p}(i, m-1) \le m \cdot \mathrm{P}[Y \le i^*]$. The new Chernoff–Hoeffding theorem (Theorem 7.1) gives a constant $C' = C'(p, \varepsilon, \varepsilon')$ with

$$m \cdot \mathrm{P}[Y \le i^*] = m \cdot \mathrm{P}[Y \le (1-\alpha)n] \le m \cdot C' \, \frac{2^{-\mathrm{D}(1-\alpha\,\|\,p)\,n}}{\sqrt{n}} = \mathrm{O}\bigg(\frac{2^{(\mathrm{H}(\alpha)+(1-\alpha)\,\mathrm{ld}\,p)\,n}}{\sqrt{n}}\bigg).$$

The lemma follows from the second part of the sum being at most a constant factor larger than the first one. This is shown using the assumption $\alpha \le 1-\varepsilon'$ and the weighting factors $w_{n,p}(i,m)$ going doubly exponentially to 0 if i crosses the threshold i^*, see Lemma 9.1. ◀

7 Conclusion

We examined the expected number of minimal edges of the maximum-entropy multi-hypergraph model with expected edge size pn. We discovered a phase transition with respect to the total number of edges at $m = 1/(1-p)^{(1-p)n}$. Now that we have tight upper and lower bounds in place, we can discuss the transition in full detail. For small m, $\mathrm{E}[|\min(\mathcal{B}_{n,m,p})|]$ is linear in m. Beyond that point, the minimization instead follows $2^{(\mathrm{H}(\alpha)+(1-\alpha)\,\mathrm{ld}\,p)\,n}/\sqrt{n}$ with $\alpha = -(\log_{1-p} m)/n$. In the information-theoretic regime the size of the minimization is decoupled from the number of edges. It continues to grow initially, but now sublinearly in m and only until $m = 1/(1-p)^{\frac{n}{1+p}}$. From there on, the size of the minimization decays rapidly although the total number of trials increases. Once m exceeds $1/(1-p)^n$, the minimization collapses under the sheer likelihood of sampling the empty set.

We gain additional insights by contrasting the results in the two regimes (ignoring constant factors here). The ratio between the two bounds at $m = 1/(1-p)^{\alpha n}$ for any $0 \le \alpha \le 1$ is $\big((2^{(\mathrm{H}(\alpha)\,+\,(1-\alpha)\;\mathrm{ld}\,p)\,n})/\sqrt{n}\big)/m = (2^{-\mathrm{D}(1-\alpha\,\|\,p)\,n})/\sqrt{n}$. This is exponentially small in n when α lies strictly between 0 and $1-p$. Therefore, the information-theoretic *lower* bound of Theorem 1.2 also pertains to the linear regime, but is unnecessarily loose there. If the number of trials m is close to $1/(1-p)^{(1-p)n}$, the two bounds coincide, up to a factor of $\sqrt{n}$, since the divergence vanishes at $1-\alpha = p$. This overlap is indicated in Figure 1b by dashed lines. Finally, for α beyond $1-p$ the relative share of minimal edges becomes exponentially small.

The Chernoff–Hoeffding theorem played an integral role in verifying these results. We tightened the tail bounds on the binomial distribution and provided explicit upper and lower bounds on the constants involved. We are convinced that this sharpened tool can help researchers in all of probability beyond the scope of this paper.

There is more work needed for the upper and lower bounds in the information-theoretic regime. Currently, α has to be bounded away from from $1-p$ and 1 for the bounds to be tight. For $\alpha \searrow 1-p$ the lower bound goes to $\Omega(2^{(\mathrm{H}(1-p)+p\,\mathrm{ld}\,p)\,n}/\sqrt{n}) = \Omega(m/\sqrt{n})$. Here, we profit from the hidden constant not depending on α. In actuality though, the minimization at $m = 1/(1-p)^{(1-p)n}$ has size $\Theta(m)$, so the share of minimal edges in the sample moves from order $1/\sqrt{n}$ to a constant. The speed of this shift depends on how fast $\alpha = 1-p+\mathrm{o}(1)$ converges. The situation for $\alpha \nearrow 1$ is different as in this parameter range there is a huge disparity between the number of minimal edges $|\min(\mathcal{B}_{n,m,p})|$ and the number of distinct edges $||\mathcal{B}_{n,m,p}||$. Thus, the expected size of the minimization is not completely captured by the binomial distribution and additional tools are needed for tight estimates. An immediate extension of our work would therefore be to pinpoint the exact behavior of the minimization at the two transitions points. Another interesting question in light of the original motivation of random databases is to allow different sample probabilities per vertex as well as dependencies between the elements. To fit the maximum-entropy setting, this would require the model to incorporate additional constraints.

References

1 Ziawasch Abedjan, Lukasz Golab, Felix Naumann, and Thorsten Papenbrock. *Data Profiling*. Synthesis Lectures on Data Management. Morgan & Claypool Publishers, San Rafael, CA, USA, 2018. doi:10.2200/S00878ED1V01Y201810DTM052.

2 Kartik Anand and Ginestra Bianconi. Entropy Measures for Networks: Toward an Information Theory of Complex Topologies. *Physical Review E*, 80:045102, 2009. doi:10.1103/PhysRevE.80.045102.

3 Robert B. Ash. *Information Theory*. Dover Books on Mathematics. Dover Publications, Mineola, NY, USA, 1990. Reprint of the Interscience Publishers 1965 edition.

4 Golnaz Badkobeh, Per Kristian Lehre, and Dirk Sudholt. Black-box Complexity of Parallel Search with Distributed Populations. In *Proceedings of the 2015 Conference on Foundations of Genetic Algorithms (FOGA)*, pages 3–15, 2015. doi:10.1145/2725494.2725504.

5 Michael Behrisch, Amin Coja-Oghlan, and Mihyun Kang. The Order of the Giant Component of Random Hypergraphs. *Random Structures and Algorithms*, 36:149–184, 2010. doi:10.1002/rsa.v36:2.

6 Michael Behrisch, Amin Coja-Oghlan, and Mihyun Kang. Local Limit Theorems for the Giant Component of Random Hypergraphs. *Combinatorics, Probability and Computing*, 23:331—366, 2014. doi:10.1017/S0963548314000017.

7 Claude Berge. *Hypergraphs - Combinatorics of Finite Sets*, volume 45 of *North-Holland Mathematical Library*. North-Holland, Amsterdam, Netherlands, 1989.

8 Ginestra Bianconi. The Entropy of Randomized Network Ensembles. *Europhysics Letters*, 81:28005, 2007. doi:10.1209/0295-5075/81/28005.

9 Thomas Bläsius, Tobias Friedrich, and Martin Schirneck. The Parameterized Complexity of Dependency Detection in Relational Databases. In *Proceedings of the 11th International Symposium on Parameterized and Exact Computation (IPEC)*, pages 6:1–6:13, 2016. doi:10.4230/LIPIcs.IPEC.2016.6.

10 Tobias Bleifuß, Sebastian Kruse, and Felix Naumann. Efficient Denial Constraint Discovery with Hydra. *Proceedings of the VLDB Endowment*, 11:311–323, 2017. doi:10.14778/3157794.3157800.

11 Thomas Bläsius, Tobias Friedrich, Julius Lischeid, Kitty Meeks, and Martin Schirneck. Efficiently Enumerating Hitting Sets of Hypergraphs Arising in Data Profiling. In *Proceedings of the 21st Meeting on Algorithm Engineering and Experiments (ALENEX)*, pages 130–143, 2019. doi:10.1137/1.9781611975499.11.

12 Béla Bollobás. *Random Graphs*. Studies in Advanced Mathematics. Cambridge University Press, Cambridge, UK, 2 edition, 2001. doi:10.1017/CBO9780511814068.

13 Béla Bollobás. A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs. *European Journal of Combinatorics*, 1:311–316, 1980. doi:10.1016/S0195-6698(80)80030-8.

14 Michele Borassi, Pierluigi Crescenzi, and Michel Habib. Into the Square: On the Complexity of Some Quadratic-time Solvable Problems. *Electronic Notes in Theoretical Computer Science*, 322:51–67, 2016. doi:10.1016/j.entcs.2016.03.005.

15 Aiden A. Bruen and Mario A. Forcinito. *Cryptography, Information Theory, and Error-Correction: A Handbook for the 21st Century*. Wiley-Interscience, New York, NY, USA, 2004. doi:10.1002/9781118033296.

16 Philip Samuel Chodrow. Configuration Models of Random Hypergraphs. *ArXiv e-prints*, 2019. arXiv:1902.09302.

17 Colin Cooper, Alan M. Frieze, and Wesley Pegden. On the Rank of a Random Binary Matrix. *Electronic Journal of Combinatorics*, 26:P4.12, 2019. doi:10.37236/8092.

18 Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience, New York, NY, USA, 2nd edition, 2006.

19 Harald Cramér. Sur un nouveau théorème-limite de la théorie des probabilités. In *Actualités scientifiques et industrielles*, volume 763, pages 5–23, 1938. Colloque consacré à la théorie des probabilités. (On a New Limit Theorem in Probability.) In French.

20 J. Demetrovics, Gyula O. H. Katona, D. Miklos, O. Seleznjev, and B. Thalheim. Asymptotic Properties of Keys and Functional Dependencies in Random Databases. *Theoretical Computer Science*, 190:151–166, 1998. doi:10.1016/S0304-3975(97)00089-3.

21 Benajmin Doerr. Probabilistic Tools for the Analysis of Randomized Optimization Heuristics. In Benjamin Doerr and Frank Neumann, editors, *Theory of Evolutionary Computation*, chapter 1. Springer International, Basel, Switzerland, 2020. doi:10.1007/978-3-030-29414-4.

22 Devdatt Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, New York, NY, USA, 2009.

23 Paul Erdős and Alfréd Rényi. On Random Graphs I. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

24 Vincent Froese, René van Bevern, Rolf Niedermeier, and Manuel Sorge. Exploiting Hidden Structure in Selecting Dimensions That Distinguish Vectors. *Journal of Computer and System Sciences*, 82:521–535, 2016. doi:10.1016/j.jcss.2015.11.011.

25 Jiawei Gao, Russell Impagliazzo, Antonina Kolokolova, and Ryan Williams. Completeness for First-Order Properties on Sparse Structures With Algorithmic Applications. *Transactions on Algorithms*, 15:23:1–23:35, 2018. doi:10.1145/3196275.

26 Diego Garlaschelli and Maria I. Loffredo. Maximum Likelihood: Extracting Unbiased Information from Complex Networks. *Physical Review E*, 78:015101, 2008. doi:10.1103/PhysRevE.78.015101.

27 Edgar N. Gilbert. Random Graphs. *Annals of Mathematical Statistics*, 30:1141–1144, 1959. doi:10.1214/aoms/1177706098.

28 Yves Grandvalet and Yoshua Bengio. Entropy Regularization. In Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien, editors, *Semi-Supervised Learning*, chapter 9. MIT Press, Cambridge, MA, USA, 2006. doi:10.7551/mitpress/9780262033589.001.0001.

29 Peter Harremoës. Binomial and Poisson Distributions as Maximum Entropy Distributions. *Transactions on Information Theory*, 47:2039–2041, 2001. doi:10.1109/18.930936.

30 Wassily Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

31 Remco van der Hofstad. *Random Graphs and Complex Networks*, volume 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, UK, 2016. doi:10.1017/9781316779422.

32 Edwin Thompson Jaynes. Information Theory and Statistical Mechanics. *Phyical Review Series II*, 106:620–630, 1957. doi:10.1103/PhysRev.106.620.

33 Edwin Thompson Jaynes. Information Theory and Statistical Mechanics II. *Phyical Review Series II*, 108:171–190, 1957. doi:10.1103/PhysRev.108.171.

34 Michał Karoński and Tomasz Łuczak. The Phase Transition in a Random Hypergraph. *Journal of Computational and Applied Mathematics*, 142:125–135, 2002. doi:10.1016/S0377-0427(01)00464-2.

35 Gyula O. H. Katona. Random Databases with Correlated Data. In Antje Düsterhöft, Meike Klettke, and Klaus-Dieter Schewe, editors, *Conceptual Modelling and Its Theoretical Foundations: Essays Dedicated to Bernhard Thalheim on the Occasion of His 60th Birthday*, pages 29–35. Springer, Berlin and Heidelberg, Germany, 2012. Festschrift. doi:10.1007/978-3-642-28279-9_4.

36 Gyula O. H. Katona. Testing Functional Connection Between Two Random Variables. In Albert N. Shiryaev, S. R. S. Varadhan, and Ernst L. Presman, editors, *Prokhorov and Contemporary Probability Theory*, pages 335–348. Springer, Berlin and Heidelberg, Germany, 2013. Festschrift. doi:10.1007/978-3-642-33549-5_20.

37 Hiremagalur Krishnaswamy Kesavan. Jaynes' Maximum Entropy Principle. In Christodoulos A. Floudas and Panos M. Pardalos, editors, *Encyclopedia of Optimization*, pages 1779–1782. Springer, Boston, MA, USA, 2009. doi:10.1007/978-0-387-74759-0_312.

38 Bernhard Klar. Bounds on Tail Probabilities of Discrete Distributions. *Probability in the Engineering and Informational Sciences*, 14:161–171, 2000.

39 Elliott H. Lieb and Jakob Yngvason. The Physics and Mathematics of the Second Law of Thermodynamics. *Physics Reports*, 310:1–96, 1999. doi:10.1016/S0370-1573(98)00082-9.

40 Ester Livshits, Alireza Heidari, Ihab F. Ilyas, and Benny Kimelfeld. Approximate Denial Constraints. *CoRR*, abs/2005.08540, 2020. Arxiv preprint. To appear in PVLDB 13. arXiv:2005.08540.

41 Michael Mitzenmacher and Eli Upfal. *Probability and Computing*. Cambridge University Press, New York, NY, USA, 2nd edition, 2017.

42 Mark E. J. Newman. Scientific Collaboration Networks. I. Network Construction and Fundamental Results. *Physical Review E*, 64:016131, 2001. doi:10.1103/PhysRevE.64.016131.

43 Mark E. J. Newman. *Networks: An Introduction*. Oxford University Press, New York, NY, USA, 2010. doi:10.1093/acprof:oso/9780199206650.001.0001.

44 Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge, UK, 2010. doi:10.1017/CBO9780511976667.

45 Thorsten Papenbrock, Jens Ehrlich, Jannik Marten, Tommy Neubert, Jan-Peer Rudolph, Martin Schönberg, Jakob Zwiener, and Felix Naumann. Functional Dependency Discovery: An Experimental Evaluation of Seven Algorithms. *Proceedings of the VLDB Endowment*, 8:1082–1093, 2015. doi:10.14778/2794367.2794377.

46 Juyong Park and Mark E. J. Newman. The Statistical Mechanics of Networks. *Physical Review E*, 70:066117, 2004. doi:10.1103/PhysRevE.70.066117.

47 Yuri Vasilyevich Prokhorov. Asymptotic Behavior of the Binomial Distribution. *Uspekhi Matematicheskikh Nauk*, 8:135–142, 1953. In Russian.

48 Fabio Saracco, Riccardo Di Clemente, Andrea Gabrielli, and Tiziano Squartini. Randomizing Bipartite Networks: The Case of the World Trade Web. *Scientific Reports*, 5:10595, 2015. doi:10.1038/srep10595.

49 Jeanette Schmidt-Pruzan and Eli Shamir. Component Structure in the Evolution of Random Hypergraphs. *Combinatorica*, 5:81–94, 1985. doi:10.1007/BF02579445.

50 Claude Elwood Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:379–423, 1948. doi:10.1002/j.1538-7305.1948.tb01338.x.

51 Eric V. Slud. Distribution Inequalities for the Binomial Law. *The Annals of Probability*, 5:404–412, 1977. URL: https://projecteuclid.org/euclid.aop/1176995801.

52 Ryan Williams. A New Algorithm for Optimal 2-Constraint Satisfaction and Its Implications. *Theoretical Computer Science*, 348:357–365, 2005. doi:10.1016/j.tcs.2005.09.023.

53 Katharina Anna Zweig. *Network Analysis Literacy: A Practical Approach to the Analysis of Networks*. Lecture Notes in Social Networks. Springer, Vienna, Austria, 2014. doi:10.1007/978-3-7091-0741-6.