The Minimization of Random Hypergraphs

Thomas Bläsius
Hasso Plattner Institute
University of Potsdam
Potsdam, Germany

Tobias Friedrich
Hasso Plattner Institute
University of Potsdam
Potsdam, Germany

Martin Schirneck
Hasso Plattner Institute
University of Potsdam
Potsdam, Germany

Abstract
We investigate the maximum-entropy model $B_{n,m,p}$ for random $n$-vertex, $m$-edge multi-hypergraphs with expected edge size $pn$. We show that the expected number of inclusion-wise minimal edges of $B_{n,m,p}$ undergoes a phase transition with respect to $m$. If $m \leq 1/(1-p)^{(1-p)n}$, the expectation is of order $O(m)$, while for $m \geq 1/(1-p)^{(1-p)n}$, it is $O(2^H(\alpha + (1-\alpha)\log p)n)$. Here, $H$ denotes the binary entropy function and $\alpha = -(\log_{1-p} m)/n$. This implies that the maximum expected number of minimal edges over all $m$ is $O((1 + p)n)$. All asymptotics are with respect to $n$, for all upper bounds we have (almost) matching lower bounds.

As a technical contribution, 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 $i^* = n + \log_{1-p} m$, which is of independent interest.

Our structural findings have algorithmic implications, for example, for computing the minimal hitting sets of a hypergraph as well as the profiling of relational databases.

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 Maximum entropy, maximization, minimization, multi-hypergraph, phase transition, random graphs.

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 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: one edge may be contained in another, even forming chains of inclusion. Often, we are only interested in the endpoints of those chains, namely, the collections of inclusion-wise minimal or maximal edges, respectively. This is the minimization or maximization of the hypergraph.

We investigate the maximum-entropy model $B_{n,m,p}$ for random multi-hypergraphs with $n$ vertices and $m$ edges. The sole constraint is that the expected cardinality of an edge is

---

The notation $B_{n,m,p}$ is mnemonic of the binomial distribution emerging in the sampling process.
The Minimization of Random Hypergraphs

We pinpoint the transition at when looking at the average-case complexity, we get a run time of \( O((1 + p)^n) \). These results draw from another, more hidden, threshold behavior. We show that the probability of a set to be minimal for \( B_{n, m, p} \) depends only on its cardinality \( i \) and falls sharply from almost 1 to almost 0 at \( i^* = n + \log_{1-p} m \).

Our structural findings have some algorithmic implications for the computation of the minimization \( \min(H) \). This is a standard preprocessing routine, e.g., when solving Hitting Set problems, but it is also an interesting computational problem in itself. There are reasons to believe that no \( m^{2-\varepsilon}\cdot\text{poly}(n) \)-algorithm exists for any \( \varepsilon > 0 \). Namely, such a procedure would falsify the orthogonal vectors conjecture\(^2\) and in turn the exponential time hypothesis.

Partitioning the edges by the number of vertices and processing them in order of increasing cardinality gives a run time of \( O(mn\lfloor \min(H) \rfloor) \), which is \( O(m^2n) \) in the worst case. However, when looking at the average-case complexity, we get a run time of \( O(mn E[\min(B_{n, m, p})]) \).

Our results thus show that the algorithm is subquadratic for all \( m \) beyond the phase transition, i.e., \( m > 1/(1 - p)^{(1-p)n} \), and is even linear in the input size for \( m \) larger than \( 1/(1 - p)^n \).

There is also a connection to the profiling of relational databases. Data scientist regularly need to compile a comprehensive list of all minimal unique column combinations or functional dependencies of a database. These dependencies are the hitting sets of the hypergraph of difference sets, that is, the collections of columns in which any pair of rows of the database disagree. When computing the difference sets one by one, this generates an incoming stream of seemingly random subsets. Filtering the inclusion-wise minimal ones from this stream does not affect the solutions but greatly reduces the number of sets to store as well as the complexity of the resulting instance. One can thus see the size of the minimization as the smallest amount of data any dependency enumeration algorithm needs to hold in memory.

Related Work. Erdős–Rényi graphs \( G_{n,m} \)\(^{17}\) and Gilbert graphs \( G_{n,p} \)\(^{20}\) are arguably the most discussed random graph models in the literature. We refer the reader to the monograph by Bollobás \(^8\) for an overview. A majority of the work on these models concentrates on various phase transitions with respect to the number of edges \( m \) and the sample probability \( p \), respectively. This intensive treatment is fueled by the appealing property that Erdős–Rényi graphs are “maximally random” and does not assume anything but the number of vertices and edges. More formally, among all probability distributions on graphs with \( n \) vertices and \( m \) edges, \( G_{n, m} \) is the unique distribution of maximum entropy. The same holds for \( G_{n, p} \) under the constraint that the expected number of edges is \( p(n) \), see \(^2\).

Shannon entropy is the central concept in information theory \(^{14, 38}\). If a stochastic system is described by the distribution \( \{p_i\}_i \), its entropy is \( H(\{p_i\}) = -\sum_i p_i \log p_i \). Since the self-information of a single state with probability \( p \) is \( \log(1/p) \), entropy is the expected information of a system. It is a measure of surprisal or how “spread out” the distribution is. Originally stemming from thermodynamics \(^{30}\), the versatility of this definition is the key

---

\(^2\) Precisely, we mean the orthogonal vectors conjecture for moderate dimensions. The reduction from the OV-complete Sperner Family problem is immediate, cp. \(^{10, 15}\).
to the successful application of information theory to fields as diverse as cryptography [11], machine learning [21], quantum computing [34], and of course network analysis [33], to name only a few topics close to computer science.

The principle of maximum entropy states that out of an ensemble of probability distribution that all describe the phenomena in question equally well, the one of maximum entropy is to be preferred in order to minimize bias. The principle is usually attributed to Jaynes [24, 25, 29]. In the context of random graphs, it is mainly used to define null models [39]. 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 caused by the constraints. This method was made rigorous by Park and Newman [35] building on earlier work in general statistics. Prescribing the exact, respectively expected, number of edges leads to the $G_{n,m}$ and $G_{n,p}$ distributions, the exact degree sequence is fixed in the configuration model [37], and in the soft configuration model the degrees at least hold in expectation [6, 19].

Many early attempts to transfer the concept of null models to hypergraphs were only indirect in that they studied hypergraphs via their clique-expansion [32] or as bipartite graphs [36]. This is unsatisfactory since these projections alter relevant observables, like node degrees or the number of triangles. Just recently, Chodrow generalized the configuration model directly to multi-hypergraphs [12]. Also, the literature on hypergraph models that happen to be maximum-entropy without being designed as such is limited. A notable early exception is the work by Schmidt-Pruzan and Shamir [37]. They fixed the exact/expected edge sequence and showed a “double jump” phase transition in the size of the largest connected component, provided that the largest edge has cardinality $O(\log n)$. Most of the literature, however, concentrates on $k$-uniform hypergraphs where every edge has exactly $k$ vertices [4, 5, 26] or, equivalently, on random binary matrices with $k$ 1s per row [13]. 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. [15] as well as Katona [27, 28]. They investigated random databases and connected the Rényi entropy of order 2 of the logarithm of the number of rows with the probability that a certain unique column combination or functional dependency holds. These dependencies are the hitting sets of the difference sets and vice versa [1, 7]. Furthermore, it is known that the Shannon entropy equals the Rényi entropy of order 1 [14]. In this sense, we complement their result by connecting the order-1 entropy of the logarithm of the number of pairs of rows with the expected number of minimal difference sets.

Outline. Next, we introduce our hypergraph model and state the results in full detail. In Section 3, we review some notation and general concepts. Section 4 contains our technical contributions. This includes the sharp threshold behavior of the probability that a set of a certain size is minimal. The main theorem and the maximum number of minimal edges are then proven in Section 5. Section 6 concludes the paper.

2 Model and Main Theorem

Fix a probability $0 \leq p \leq 1$ and positive integers $n, m$. We define the random multi-hypergraph $B_{n,m,p}$ by independently sampling $m$ (not necessarily distinct) subsets of $[n]$. Each set is generated by including any vertex $v \in [n]$ with probability $p$ independently of all others.

We quickly argue that this is the maximum-entropy model. The three constraints are
The Minimization of Random Hypergraphs

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

(b) Our bound as a function of \( \alpha \) for \( p = 0.6 \) (the plot is independent of \( n \)). The vertical line at \( \alpha = \frac{1}{1 + p} \) indicates the maximum (Theorem 2). For \( \alpha \leq 1 - p \), the linear bound holds; for larger \( \alpha \), 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 \( m \) and on \( \alpha \). As \( \alpha \) grows logarithmically in \( m \), shows the same plot as (a) but with both axes being logarithmic.

The independence bound on entropy reads as follows [13]. Let \( X_1 \) to \( X_m \) be random variables with joint distribution \( P_{X_1,\ldots,X_m} \) and marginal distributions \( P_{X_j} \). Then, their entropies satisfy

\[
H(P_{X_1,\ldots,X_m}) \leq \sum_{j=1}^m H(P_{X_j}),
\]

with equality if and only if the \( X_j \) are independent. This suggests that we should choose the edges independently if we want to maximize the entropy. The same holds for the vertices inside an edge. Finally, the fact that setting the sampling probabilities per vertex to be all equal indeed maximizes the entropy under a given mean set size was proven by Harremoës [22].

We are interested in the expected number of inclusion-wise minimal sets in \( B_{n,m,p} \), denoted by \( E[|\min(B_{n,m,p})|] \). We describe the asymptotic behavior of this expectation with respect to \( n \). In more detail, we view \( m = n(n) \) as a function of \( n \) taking integer values and bound the univariate asymptotics of \( E[|\min(B_{n,m,p})|] \) in \( n \) for any choice of \( m \). The probability \( p \), however, is considered to be a constant.

To state our result in full detail, we let \( H(x) = H(|x,1-x|) \) denote the binary entropy function. Further, we define

\[
\alpha = \log \frac{1}{1-p} m = \frac{\log_{1-p} m}{n}.
\]

The quantity \( \alpha \) is a non-negative function of \( p \), \( n \), and \( m \); it exists for all \( 0 < p < 1 \) and \( n, m \geq 1 \). Asymptotically in \( n \), it is of order \( \Theta((\log m)/n) \). If \( p \) and \( n \) are fixed, choosing a value for \( \alpha \) determines \( m \) since we can rewrite \( m \) as \( \frac{1}{(1-p)^n} \).

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

1. For all \( m \leq \frac{1}{1-p} \), we have \( E[|\min(B_{n,m,p})|] = \Theta(m) \).
2. For any \( \varepsilon > 0 \) and all \( m \) such that \( 1/(1-p)^{(1-p)n} \leq m \leq \frac{1}{1-p} \),

\[
E[|\min(B_{n,m,p})|] = O \left( \frac{2^{\log(\alpha + (1-\alpha)\log p)}}{\alpha^{\alpha}} \right) = O \left( \left( \frac{1-\alpha}{\alpha^{\alpha-1}} p^{1-\alpha} \right)^n \right);
\]
The maximum is attained for the information-theoretic regime, where
\[ \delta \]
the exponential upper and lower bounds in the information-theoretic regime are at least
\[ H(x) = -x \log_a x - (1 - x) \log_a (1 - x). \]

On the open unit interval, \( H(x) \) 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, we have \( H(x) = H_y(x) \). Evidently, the entropy function is symmetric with \( H(x) = H(1 - x) \).

On the open unit interval, \( H(x) \) is positive and differentiable with derivative
\[ \frac{dH(x)}{dx} = \log \left( \frac{1 - x}{x} \right). \]
H is strictly concave and has a single maximum at position $x^* = 1/2$ with value $H(x^*) = 1$.

We utilize the entropy function to estimate binomial coefficients. The bounds are well-known in the literature and can be found, e.g., in the textbook by Cover and Thomas [14].

Lemma 3. Let $n$ be a positive integer and $0 < x < 1$, then

$$\frac{2H(x)^n}{\sqrt{8n}x(1-x)} \leq \binom{n}{xn} \leq \frac{2H(x)^n}{\sqrt{\pi}nx(1-x)}.$$

We utilize the entropy function to estimate binomial coefficients. The bounds are well-known in the literature and can be found, e.g., in the textbook by Cover and Thomas [14].

Lemma 3. Let $n$ be a positive integer and $0 < x < 1$, then

$$\frac{2H(x)^n}{\sqrt{8n}x(1-x)} \leq \binom{n}{xn} \leq \frac{2H(x)^n}{\sqrt{\pi}nx(1-x)}.$$

Let $\{p_i\}$ and $\{q_i\}$ be two distributions on the same state space such that $q_i = 0$ implies $p_i = 0$ for any $i$. The (binary) Kullback–Leibler divergence\(^3\) between the two is given by

$$D(\{p_i\} \parallel \{q_i\}) = - \sum_i p_i \log \left( \frac{q_i}{p_i} \right).$$

It is the expected information loss when assuming that the distribution is $\{q_i\}$ while the system in fact follows $\{p_i\}$. 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 inequality. We mainly use the following derived function. For any two reals $x$ and $y$ in the unit interval, the divergence between two Bernoulli distributions with parameters $x$ and $y$, respectively, is

$$D(x \parallel y) = -x \log \left( \frac{y}{x} \right) - (1 - x) \log \left( \frac{1 - y}{1 - x} \right).$$

It is not hard to show that $D(x \parallel y)^n = D(\text{Bin}(n,x) \parallel \text{Bin}(n,y))$ is the divergence between binomial distributions. The Chernoff–Hoeffding theorem uses this to bound the probability that a binomial random variable deviates additively from its mean.

Lemma 4 (Chernoff–Hoeffding theorem [16, 23]). Let $n$ be a non-negative integer, $p$ a probability, and $\varepsilon$ a real number such that $0 \leq \varepsilon \leq p$. Further, let $Y \sim \text{Bin}(n,p)$ be a binomially distributed random variable with parameters $n$ and $p$. Then,

$$P[Y \leq (p - \varepsilon)n] \leq 2^{-D(p-\varepsilon \parallel p)^n} = \left( \left( \frac{p}{p - \varepsilon} \right)^{p-\varepsilon} \left( \frac{1 - p}{1 - p + \varepsilon} \right)^{1-p+\varepsilon} \right)^n.$$

Polynomials of Probabilities. We regularly need to estimate expressions of the form $(1 - x)^n$ where $x$ is a probability. The first inequality we use for this task is taken from the textbook by Motwani and Raghavan [31].

Lemma 5. Let $n$ be a positive integer and $x$ a real number such that $|x| \leq n$, then

$$e^x \left( 1 - \frac{x^2}{n} \right) \leq \left( 1 + \frac{x}{n} \right)^n.$$

We reach rather tight bounds on $(1 - x)^n$ by substituting $x$ for $-nx$ above, and combining this with the simple fact that $(1 + x) \leq e^x$ holds for all $x$.

Corollary 6. Let $n$ be a positive integer and $0 \leq x \leq 1$, then

$$e^{-nx} \left( 1 - nx^2 \right) \leq (1 - x)^n \leq e^{-nx}.$$ 

The next set of inequalities was given by Badkobeh, Lehre, and Sudholt [3].

\(^3\) The divergence is sometimes also called relative entropy, we avoid this term due to ambiguities, cf. [14].
Lemma 7 (Lemma 10 in [3]). Let \( n \) be a non-negative integer and \( 0 \leq x \leq 1 \), then
\[
\frac{nx}{1 + nx} \leq 1 - (1 - x)^n \leq nx.
\]

Finally, we prepare the following lemma for later use.

Lemma 8. Consider a random experiment with outcomes \( A, B \), and \((\neg A \land \neg B)\), where \( P[B] > 0 \). 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 \( P[\forall j \leq m: \neg A_j \land \exists k \leq m: B_k] \leq P[\forall j \leq m: \neg A_j \land B_m] \).

Proof. First, we prove that the claim is equivalent to
\[
P[\forall j \leq m: \neg A_j \land \exists k < m: B_k] \leq P[\forall j \leq m: \neg A_j \land B_m].
\]
To this end, observe that for any four reals \( x, y, z, w \) such that \( y, w \) are all non-zero, we have \( \frac{x}{y} < \frac{z}{w} \) if and only if \( \frac{x}{y} + \frac{z}{w} < \frac{x}{y} \), which can be seen by elementary means.

The event \([\exists k \leq m: B_k]\) can be partitioned into \([B_m]\) and \([\neg B_m \land \exists k < m: B_k]\). Thus,
\[
P[\forall j \leq m: \neg A_j \land \exists k \leq m: B_k] = \frac{P[\forall j \leq m: \neg A_j \land \exists k \leq m: B_k]}{P[\exists k \leq m: B_k]} = \frac{P[\forall j \leq m: \neg A_j \land \exists k \leq m: B_k]}{P[\exists k \leq m: B_k]}.
\]

Applying the observation to the real numbers \( x = P[\forall j \leq m: \neg A_j \land \exists k < m: B_k] \), \( y = P[\neg B_m \land \exists k < m: B_k] \), \( z = P[\forall j \leq m: \neg A_j \land B_m] \) and \( w = P[B_m] \) gives the equivalence.

Now we prove the actual lemma by induction over the \( m \) trials. The case \( m = 1 \) is trivial since there both sides of the claimed inequality simplify to \( P[\neg A_1 \land B_1] \). In the following, suppose \( P[\forall j < m: \neg A_j \land \exists k < m: B_k] \leq P[\forall j < m: \neg A_j \land B_{m-1}] \) holds. It is sufficient to conclude \( P[\forall j \leq m: \neg A_j \land \exists k < m: B_k] \leq P[\forall j \leq m: \neg A_j \land B_m] \) from that. As the trials are independent, we get
\[
P[\forall j \leq m: \neg A_j \land \exists k < m: B_k] = \frac{P[\forall j \leq m: \neg A_j \land \exists k < m: B_k]}{P[\neg B_m \land \exists k < m: B_k]}
\]
\[
= \frac{P[\neg A_m \land \neg B_m] \cdot P[\forall j < m: \neg A_j \land \exists k < m: B_k]}{P[\neg B_m] \cdot P[\exists k < m: B_k]}
\]
\[
= P[\neg A_m \land \neg B_m] \cdot P[\forall j < m: \neg A_j \land \exists k < m: B_k].
\]

By induction, the latter is at most \( P[\neg A_m \land \neg B_m] \cdot P[\forall j < m: \neg A_j \land B_{m-1}] \). The probabilities of the outcomes do not change over the trials; also, event \( B_m \) implies \( \neg A_m \). Therefore,
\[
P[\neg A_m \land \neg B_m] \cdot P[\forall j < m: \neg A_j \land B_{m-1}] = \frac{1 - P[A_m] - P[B_m] \cdot (1 - P[A_m])^{m-2}}{1 - P[B_m]} \cdot \frac{P[B_{m-1}]}{P[B_{m-1}]^{m-1}}
\]
\[
= 1 - \frac{P[A_m]}{1 - P[B_m]} \cdot (1 - P[A_m])^{m-2} \leq \left(1 - P[A_m]\right)^{m-1} = P[\forall j \leq m: \neg A_j \land B_m]
\]

4 Distinct Sets and Minimality

As a first step towards the proof of Theorem 1, we give preliminary bounds on the expected number of minimal edges in \( B_{n,m,p} \). The bounds have the form of binomial sums of products of probabilities, depending on which factors we choose we get an upper or a lower bound. They are tight up to a constant factor and will serve as the basis for the further analysis.
The Minimization of Random Hypergraphs

The cases in which the probability to include a vertex is either \( p = 0 \) or \( p = 1 \) are trivial. \( \mathcal{B}_{n,m,p} \) then deterministically consists of \( m \) copies of the empty set or the whole universe \( [n] \), respectively. Either way, the minimization \( \min(\mathcal{B}_{n,m,p}) \) contains only a single edge. We therefore assume \( 0 < p < 1 \) in the remainder of this work unless explicitly stated otherwise. Every subset of \( [n] \) then has a non-vanishing chance to occur. Such a set is minimal for \( \mathcal{B}_{n,m,p} \) if and only if it is sampled in one of the trials and no proper subset is ever generated.

Both aspects influence the minimality but their impact varies depending on the cardinality of the set in question. The number of vertices per edge is heavily concentrated around \( pm \), and edges with more vertices are less likely to be minimal. Consequently, almost no sets with very low cardinality are generated, but if such a set is sampled, it is often minimal. Many edges with a medium number of vertices are sampled and there is a good chance they are included in \( \min(\mathcal{B}_{n,m,p}) \). Very high cardinalities rarely occur, and even if so, those sets are usually 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. We will see that for large \( m \) the last effect outweighs all others, creating situations in which only very small sets have a chance to be minimal.

Let \( \Delta_{n,p} \) denote the maximum-entropy distribution on the power set \( \mathcal{P}([n]) \) under the constraint that \( E_X \sim \Delta_{n,p}. |X| = pn \). That is to say, every vertex is included independently and identically distributed with probability \( p \).

Lemma 9. Let \( 0 < p < 1 \) be a probability, \( n, m \) positive integers, and let \( X_j \sim \Delta_{n,p} \) denote the outcome of the \( j \)-th trial. For any integer \( i \) such that \( 0 \leq i \leq n \), define

\[
 w_{n,p}(i, m) = P[\forall j \leq m: \neg(X_j \subseteq [i])].
\]

Then, we have \( w_{n,p}(i, m) = \left(1 - (1 - p)^{n-i}(1 - p')\right)^m \). It further holds that

1. \( E[|\min(\mathcal{B}_{n,m,p})|] \geq \sum_{i=0}^{n} \binom{n}{i} (1 - (1 - p')(1 - p)^{n-i})^m \cdot w_{n,p}(i, m) \);
2. \( E[|\min(\mathcal{B}_{n,m,p})|] \leq \sum_{i=0}^{n} \binom{n}{i} (1 - (1 - p')(1 - p)^{n-i})^m \cdot w_{n,p}(i, m - 1) \);
3. \( E[|\min(\mathcal{B}_{n,m,p})|] \leq 1 + \frac{1}{m} \sum_{i=0}^{n} \binom{n}{i} (1 - (1 - p')(1 - p)^{n-i})^m \cdot w_{n,p}(i, m) \).

Proof. First, we show that indeed we have

\[
 w_{n,p}(i, m) = \left(1 - (1 - p)^{n-i}(1 - p')\right)^m.
\]

The random set \( X_j \sim \Delta_{n,p} \) is a subset of \([i]\) if it does not contain an element of \([n]\setminus[i] \), 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 \(1 - p'\). Therefore, the probability \( P[\forall j \leq m: \neg(X_j \subseteq [i])] \) is as desired.

We now turn to 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 \(|S|\) as all sets with the same number of elements are equally likely.

\[
 E[|\min(\mathcal{B}_{n,m,p})|] = \sum_{S \subseteq [n]} P[\exists k \leq m: X_k = S \land \forall j \leq m: \neg(X_j \subseteq S)]
 = \sum_{i=0}^{n} \binom{n}{i} P[\exists k \leq m: X_k = [i]] \cdot P[\forall j \leq m: \neg(X_j \subseteq [i]) | \exists k \leq m: X_k = [i]].
\]
Generating any other set than \([i]\) in a single trial has probability \(1 - p^i(1 - p)^{n-i}\). This probability does not change over the independent trials, giving an expected value of

\[
\sum_{i=0}^{n} \binom{n}{i} \left(1 - (1 - p^i(1 - p)^{n-i})^m\right) \cdot \mathbb{P}[\forall j \leq m: \neg(X_j \subseteq [i]) \mid \exists k \leq m: X_k = [i]].
\]

The last factor describes the likelihood that a set with \(i\) elements is minimal, conditioned on it being sampled at all. The stated bounds differ only in the way this factor is estimated. We claim that it is at least as large as \(\mathbb{P}[\forall j \leq m: \neg(X_j \subseteq [i])]\) (i.e., without the condition) while at the same time being at most \(\mathbb{P}[\forall j < m: \neg(X_j \subseteq [i])]\) (with one fewer trial). The first inequality is obvious because conditioning on at least one trial producing the set \([i]\) itself only increases the chances of never sampling a proper subset. For the second one, we apply Lemma 8 to the events \(A_j = [X_j \subseteq [i]]\) and \(B_j = [X_j = [i]]\), showing that

\[
\mathbb{P}[\forall j \leq m: \neg(X_j \subseteq [i]) \mid \exists k \leq m: X_k = [i]] \leq \mathbb{P}[\forall j \leq m: \neg(X_j \subseteq [i]) \mid X_m = [i]].
\]

The proof of the claim is completed by observing that \(\mathbb{P}[\forall j \leq m: \neg(X_j \subseteq [i]) \mid X_m = [i]]\) is the same as \(\mathbb{P}[\forall j < m: \neg(X_j \subseteq [i])]\), which holds due to the independence of the trials. This proves the Statements 1 and 2.

The relative difference between \(w_{n,p}(i, m-1)\) and \(w_{n,p}(i, m)\) is \(1/(1 - (1 - p)^{n-i}(1 - p^i))\), a quantity independent of \(m\) and non-decreasing in the cardinality \(i\). Thus, for \(i < n\), the ratio is upper bounded by \(1/(1 - (1 - p)(1 - p^{n-i})) \leq 1/p\). If \(i = n\), the difference is super-constant, namely, \(1/p^n\). Notwithstanding, Statement 3 follows from the fact that the contribution of the last term to the whole sum is at most 1.

Recall that we use \(|\mathcal{H}|\) to denote the number of distinct sets in a multi-hypergraph. The part that all three bounds above have in common describes the expected number of distinct sets in \(\mathcal{B}_{n,m,p}\). That is to say that

\[
\mathbb{E}[|\mathcal{B}_{n,m,p}|] = \sum_{i=0}^{n} \binom{n}{i} \left(1 - (1 - p^i(1 - p)^{n-i})^m\right).
\]

To reach the bounds the terms of sum are weighted by \(w_{n,p}(i, m)\) and \(w_{n,p}(i, m-1)\), respectively. We analyze the two parts separately, starting with the \(w_{n,p}\).

These factors are of interest beyond their application to random multi-hypergraphs. For the maximum-entropy distribution \(\Delta_{n,p}\) on subsets of \([n]\) with expected set size \(pn\), the weighting factor \(w_{n,p}(i, m)\) is, by definition, the probability that any fixed subset of cardinality \(i\) is minimal after \(m\) i.i.d. trials according to \(\Delta_{n,p}\). Equivalently, \(1 - w_{n,p}(i, m)\) is the probability of any proper subset being sampled.

It is easy to see that \(w_{n,p}(i, m) = (1 - (1 - p)^{n-i} + p^i(1 - p)^{n-i})^m\) is non-increasing in both \(i\) and \(m\). 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 \(\alpha\) abbreviates \(-(\log_{1-p} m)/n\). In full detail, Lemma 10 below establishes a sharp threshold behavior at

\[
i^* = n + \log_{1-p} m = (1 - \alpha)n.
\]

Note that \(i^*\) is always at most \(n\) since \(\log_{1-p} m \leq 0\). Moreover, 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 since it becomes negative. Later, in Lemma 15 we will see that \(m\) being this large is in fact irrelevant for the analysis of \(\mathbb{E}[\min(|\mathcal{B}_{n,m,p}|)]\).
Lemma 10. Let $0 < p < 1$ be a probability, and $n$, $m$ positive integers. Let $i = i(n)$ be a function of $n$ taking integer values between 0 and $n$, including.

1. We have $w_{n,p}(0,m) = 1$, and $w_{n,p}(n,m) = p^m$.

Suppose $0 < i < n$ for the remainder.

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 $w_{n,p}(i,m) = \Omega(1)$.

Proof. The corner cases in Statement 1 are elementary. Suppose $0 < i < n$. We mainly use Corollary 6 to estimate $w_{n,p}(i,m)$. For the upper bound, we have

$$w_{n,p}(i,m) = (1 - (1 - p)^{n-i})^m \leq (1 - (1 - p)^{n-i}(1-p))^m \leq \exp\left(-m (1-p)^{n-i} \cdot (1-p)\right).$$

Since $1 - p$ is constant, the limit behavior is determined entirely 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 $w_{n,p}(i,m)$ tends to 0.

Conversely, as $1 - p$ is at most 1, we get

$$w_{n,p}(i,m) \geq (1 - (1 - p)^{n-i})^m \geq \exp\left(-m (1-p)^{n-i}ight) \cdot (1-m(1-p)^{(2n-i)}).$$

If $i = n + \log_{1-p} m - \omega(1)$, both $m (1-p)^{n-i} = (1-p)^{\omega(1)}$ and $m (1-p)^{2(n-i)} = (1-p)^{\omega(1)}/m$ converge to 0. These two facts together imply $\lim_{n \to \infty} w_{n,p}(i,m) = 1$.

Finally, if the cardinality $i$ is around the threshold, the limit may not exist. Let the constant $C$ be such that $|i^* - i| \leq C$; in particular, it holds that $-\log_{1-p} m - C \leq n - i \leq -\log_{1-p} m + C$. We use Lemma 7 to cover this case,

$$w_{n,p}(i,m) \geq (1 - (1 - p)^{n-i})^m \geq \frac{m (1-p)^{n-i}}{1 + m(1-p)^{n-i}} \geq \frac{(1-p)^C}{1 + (1-p)^C} = \Omega(1).$$

We restate the precise bounds of Statement 2 and 4 for later use.

Corollary 11. If $0 < p < 1$ and $0 < i < n$, then $w_{n,p}(i,m) \leq \exp\left(-m (1-p)^{n+i+1}\right)$. If additionally there is a $C \geq 0$ such that $|i^* - i| \leq C$, then $w_{n,p}(i,m) \geq (1-p)^C/(1+(1-p)^C)$.

After showing the existence of a threshold for the weighting factors, we turn to the number of distinct sets in $\mathcal{B}_{n,m,p}$. This is a natural upper bound for the size of the minimization, $|\min(\mathcal{B}_{n,m,p})| \leq |\mathcal{B}_{n,m,p}| \leq |\mathcal{B}_{n,m,p}| = m$. When starting the sampling, many different sets are generated and $|\mathcal{B}_{n,m,p}|$ is close to $m$. As the number of trials increases though, duplicates occur in the sample and the two quantities grow apart.

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

Lemma 12. Let $0 < p < 1$ be a probability, $n$, $m$ positive integers, and $Y \sim \text{Bin}(n,p)$ a binomially distributed random variable with parameters $n$ and $p$. Further, let $\ell$, $u$ be integers
such that $0 \leq \ell \leq u \leq n$ and define $p = \max_{i \leq u} \{ p^i(1-p)^{n-i} \}$. Then, we have

$$p = \begin{cases} 
p^\ell(1-p)^{n-\ell}, & \text{if } p < 1/2; \\
1/2^n, & \text{if } p = 1/2; \\
p^u(1-p)^{n-u}, & \text{otherwise.} \end{cases}$$

and the expected number of distinct sets in $B_{n,m,p}$ with cardinality between $\ell$ and $u$ is

$$\frac{m}{1+mp} \cdot P[\ell \leq Y \leq u] \leq E[\|B_{n,m,p}(\ell,u)\|] \leq m \cdot P[\ell \leq Y \leq u].$$

**Proof.** The closed form for $p$ can be seen from the equality $p^i(1-p)^{n-i} = (p/(1-p))^i(1-p)^n$ and the fact that the odds $p/(1-p)$ are smaller than 1 iff $p \leq 1/2$. Lemma 9 implies for the number of distinct sets in $B_{n,m,p}(\ell,u)$ that

$$E[\|B_{n,m,p}(\ell,u)\|] = \sum_{i=\ell}^{u} \binom{n}{i} (1-(1-p^i(1-p)^{n-i})^m) \leq m \cdot \sum_{i=\ell}^{u} \binom{n}{i} p^i(1-p)^{n-i}.$$

Conversely, we have

$$E[\|B_{n,m,p}(\ell,u)\|] \geq \sum_{i=\ell}^{u} \frac{m p^i(1-p)^{n-i}}{1+mp} \geq \frac{m}{1+mp} \cdot \sum_{i=\ell}^{u} \binom{n}{i} p^i(1-p)^{n-i}.$$

The sum in both bounds equals the probability that $Y \sim \text{Bin}(n,p)$ is between $\ell$ and $u$. ◀

# 5 Proof of the Main Theorem

In this section, we prove the main result with the tools above. A key observation is that the minimization is dominated by the sets with cardinalities around the threshold of the weighting factors.

## 5.1 The Lower Bound

For a small number of trials, many different sets are created. Namely, we show that the distinct edges make up at least a constant fraction of $B_{n,m,p}$ as long as $m$ is at most $1/(1-p)^{(1-p)n}$. In turn, since the cardinalities of the sets are concentrated around $pn$ a constant fraction of them are indeed minimal. For a larger number of trials this no longer holds true. We show that once $m$ is so large that the threshold $i^* = n + \log_{1-p} m$ falls below $pn$, the number of minimal edges also decreases significantly.

**Lemma 13 (Theorem 11).** For all $m \leq 1/(1-p)^{(1-p)n}$, we have $E[\|\text{min}(B_{n,m,p})\|] = \Theta(m)$.

**Proof.** The upper bound is trivial as $\text{min}(B_{n,m,p})$ has at most $m$ edges. To get a lower bound on $E[\|\text{min}(B_{n,m,p})\|]$, we truncate the sum given in Lemma 9 at index $i = pn$ and use that the weighting factors $w_{n,p}(i,m)$ are non-increasing in $i$. Let $Y \sim \text{Bin}(n,p)$ be a random variable and set $p = \max_{0 \leq i \leq pn} \{ p^i(1-p)^{n-i} \}$. Lemma 12 yields

$$E[\|\text{min}(B_{n,m,p})\|] \geq \sum_{i=0}^{pn} \binom{n}{i} \left( 1 - (1-p^i(1-p)^{n-i}) \right)^m \cdot w_{n,p}(i,m) \geq \sum_{i=0}^{pn} \binom{n}{i} \left( 1 - (1-p^{pn}(1-p)^{n-pn}) \right)^m \cdot w_{n,p}(pn,m) \geq E[\|B_{n,m,p}(0,pn)\|] \cdot w_{n,p}(pn,m) \geq m \cdot \frac{P[Y \leq pn]}{1+mp} \cdot w_{n,p}(pn,m).$$
We are done if the last two coefficients are bounded below by positive constants. The upper limit of the summation is the median of the binomial distribution Bin(n,p), giving 
\[ P[Y \leq pn] \geq 1/2. \]
Next we bound the product \( mp \) from above by
\[
mp \leq (1 - p)^{(p-1)n} p = \begin{cases} 
(1 - p)^{pn}, & \text{if } p < 1/2; \\
1/\sqrt{2n}, & \text{if } p = 1/2; \\
p^{\alpha n}, & \text{otherwise.}
\end{cases}
\]
Either way, \( mp \) is at most 1. (In fact, since \( p = 0 \) and \( p = 1 \) are excluded, \( mp \) goes to 0.) The threshold of the weighting factors is at \( i^* = n + \log_{1-p} m \geq n + \log_{1-p}((1 - p)^{(p-1)n}) = pn \).
Hence, \( pn \) is below the threshold \( i^* \) and applying Statement 3 or 4 of Lemma 10 shows that there exists a constant \( \delta > 0 \) such that \( w_n(pn, m) \geq \delta \). Combining the three bounds finally proves
\[ E[\min(B_{n,m,p})] \geq m \cdot \delta/4 = \Omega(m). \]

For large sample sizes, the linear growth of the number of minimal sets cannot be maintained. Instead, \( \min(B_{n,m,p}) \) enters a regime governed by the entropy of \( \alpha = -(\log_{1-p} m)/n \).
We first show the lower bound, which is a bit simpler. It holds for all \( \alpha \) in the unit interval. However, we will see that it is only meaningful in in the information-theoretic regime, i.e., for \( k \geq 1/(1 - p)^{(1-p)n} \). Again, \( H \) is the entropy function.

**Lemma 14 (Theorem 13).** For all \( m \leq 1/(1 - p)^n \), i.e., \( 0 \leq \alpha \leq 1 \), it holds that
\[ E[\min(B_{n,m,p})] = \Omega \left( 2^{(H(\alpha) + (1-\alpha)\log p)n}/\sqrt{n} \right). \]

**Proof.** The sought expectation is at least as large as the number of distinct sets of cardinality \( i \) that are minimal after \( m \) trials, for arbitrary values of \( i \). As an ansatz, we chose the cardinality to be directly at the threshold \( i^* = n + \log_{1-p} m \).
Let again \( Y \sim \text{Bin}(n,p) \) be a random variable. Lemmas 9 and 12 together imply that
\[
E[\min(B_{n,m,p})] \geq \frac{\min(B_{n,m,p}(i^*, i^*))}{m} \cdot w_n(i^*, m)
\]
\[ \geq \frac{1}{1 + mp^*(1 - p)^{n - \alpha p}} \cdot P[Y = i^*] \cdot w_n(i^*, m). \]
Corollary 11 gives that the weighting factor at the threshold \( i^* \) is bounded from below by a constant, namely \( 1/2 \), uniformly for all \( m \). We now apply the writeouts \( m = 1/(1 - p)^{\alpha n} \) and \( i^* = (1 - \alpha)n \). For \( \alpha = 0 \) or \( \alpha = 1 \), the claimed bound degenerates to \( \Omega(p^n/\sqrt{n}) \) or \( \Omega(1/\sqrt{n}) \), respectively. We can thus assume \( 0 < \alpha < 1 \) and arrive at
\[
E[\min(B_{n,m,p})] \geq \frac{1}{2} \cdot \frac{(1 - p)^{-\alpha n}}{1 + (1 - p)^{-\alpha n} p^{(1-\alpha)n}(1 - p)^{\alpha n}} \cdot P[Y = (1 - \alpha)n]
\]
\[ = \frac{1}{2} \cdot \frac{(1 - p)^{-\alpha n}}{1 + (1 - p)^{-\alpha n} p^{(1-\alpha)n}(1 - p)^{\alpha n}} \cdot \left( (1 - \alpha)n \right)^{p^{(1-\alpha)n}(1 - p)^{\alpha n}}
\]
\[ = \frac{1}{2} \cdot \frac{1}{1 + p^{(1-\alpha)n}} \left( (1 - \alpha)n \right)^{p^{(1-\alpha)n}}. \]

The fraction \( 1/(1 + p^{(1-\alpha)n}) \) is never smaller than \( 1/2 \). Hence, the lower bound is asymptotically dominated by \( \left( (1 - \alpha)n \right)^{p^{(1-\alpha)n}} \). Lemma 3 provides information-theoretic estimates of the binomial coefficient. Note that \( H(1 - \alpha) \) and \( H(\alpha) \) are equal.
\[
\left( \frac{n}{(1 - \alpha)n} \right)^{p^{(1-\alpha)n}} \geq \frac{2^{H(\alpha)n}}{\sqrt{8n(1 - \alpha)\alpha}} \cdot \frac{1}{\sqrt{8n(1 - \alpha)\alpha}} \cdot \frac{1}{\sqrt{8n(1 - \alpha)\alpha}} \cdot \frac{1}{\sqrt{8n(1 - \alpha)\alpha}} \cdot 2^{(H(\alpha) + (1-\alpha)\log p)n}. \]
The proof is completed by the observation that \( 1/\sqrt{8(1 - \alpha)\alpha} \geq 1/\sqrt{2} \) for \( 0 < \alpha < 1 \).
We have seen bounds on the expected size of the minimization for two different ranges of \( m \). The tight one given in Lemma 13 holds only if \( m \leq 1/(1 - p)^{(1-\epsilon)n} \). There, the threshold \( i^* \) is not smaller than the expected edge size \( pn \). The analysis in that case includes the full sample size. The information-theoretic bound (Lemma 14) focuses on the edges with cardinality at the threshold. It holds for all \( m \leq 1/(1 - p)^n \), but is rather slack for \( m \) much below \( 1/(1 - p)^{(1-\epsilon)n} \). If the number of trials is close to \( 1/(1 - p)^{(1-\epsilon)n} \) the two bounds coincide (up to polynomial factors). For large \( m \) the information-theoretic lower bound is even decreasing. One could suspect that this is only an artifact of the particular techniques we used in its proof. We show next that this is not the case by giving a corresponding upper bound, implying a phase transition of the behavior of \( E[\min(B_{n,m,p})] \) around \( m = 1/(1 - p)^{(p-1)n} \).

5.2 The Upper Bound

The upper bound draws from the same core observations as the previous section, namely, the position of the threshold of the weighting factors as well as the ratio of distinct sets in the sample. First, we show that once \( m \) is more than a polynomial larger than \( 1/(1 - p)^n \), the minimization essentially consists of a single edge, namely, the empty set. We then prove in Lemma 15 our claim that for intermediate values of \( m \) between the phase transition and \( 1/(1 - p)^n \), \( E[\min(B_{n,m,p})] \) follows closely the lower bound shown in Lemma 14.

For the next result, note that \( \min(B_{n,m,p}) \) always contains at least a single edge.

\begin{lemma}[Theorem 14] If \( m = 1/(1 - p)^{n+\omega(\log n)} \), then \( E[\min(B_{n,m,p})] = 1 + o(1) \).
\end{lemma}

\textbf{Proof.} Suppose \( m = 1/(1 - p)^{n+f(n)} \) for some function \( f = \omega(\log n) \). As soon as the empty set is sampled in one of the \( m \) trials, the minimization of \( B_{n,m,p} \) contains only a single set; otherwise, we fall back to the trivial estimate \( \min(B_{n,m,p}) \leq m \). Let \( A \) abbreviate the event \( \{0 \in B_{n,m,p}\} \). The law of total expectation together with Corollary 6 thus implies that

\[
E[\min(B_{n,m,p})] = E[\min(B_{n,m,p}) | A] \cdot P[A] + E[\min(B_{n,m,p}) | \neg A] \cdot P[\neg A] \\
\leq P[A] + m \cdot (1 - (1 - p)^n) m \leq 1 + \exp(\ln m - m(1 - p)^n) \\
= 1 + \exp(\ln m - (1 - p)^{-f(n)}) .
\]

By the assumption on \( m \), the logarithm \( \ln m \) is of order \( O(n + f(n)) \). As \( 1/(1 - p) \) is strictly larger than 1 and \( f = \omega(\log n) \), \( \ln m \) is negligible compared to \( 1/(1 - p)^f(n) \). Therefore, the exponential expression converges to 0.

We prove next the last remaining statement of the main theorem, which is the upper bound in the information-theoretic regime.

\begin{lemma}[Theorem 12] Fix an \( \epsilon > 0 \) and suppose \( m \) is between \( 1/(1 - p)^{(1-\epsilon)n} \) and \( 1/(1 - p)^{(1-\epsilon)\alpha n} \), that is, \( 1 - p \leq \alpha \leq 1 - \epsilon \). Then, we have

\[
E[\min(B_{n,m,p})] = O\left(2^{(H(\alpha)+(1-\alpha)\log p)n}\right).
\]

\end{lemma}

\textbf{Proof.} Lemma 13 states \( E[\min(B_{n,m,p})] \leq \sum_{i=0}^{\alpha n}\binom{n}{i}(1 - (1 - p)(1 - p)^{n-i})^m w_{n,p}(i, m-1). \) The main idea of this proof is to split this sum at the threshold \( i^* \) and handle the two parts separately. A constant fraction of the distinct sets with cardinality below the threshold are minimal, cp. Lemma 10. Let \( Y \sim \text{Bin}(n, p) \) be a binomial variable. Lemma 12 implies that the first part of the sum is bounded above by

\[
\sum_{i=0}^{i^*}\binom{n}{i}(1 - (1 - p)(1 - p)^{n-i})^m w_{n,p}(i, m-1) \leq E[\|B_{n,m,p}(0, i^*)\|] \leq m \cdot P[Y \leq i^*].
\]
Recall that we denote the Kullback–Leibler divergence between Bernoulli random variables by $D$, the entropy function by $H$, and that we let $\alpha$ abbreviate $-(\log_{1-p} m)/n$. From the assumption $m \geq (1-p)^{(1-\alpha)n}$, we get $i^* = (1-\alpha)n \leq pn = E[Y]$. The Chernoff–Hoeffding theorem (Lemma 4) thus yields

$$m \cdot P[Y \leq (1-\alpha)n] \leq m \cdot 2^{-D(1-\alpha \| p)} = \frac{1}{(1-p)^{\alpha m}} \cdot \left( \frac{p}{1-\alpha} \right)^{1-\alpha} \left( \frac{1-p}{\alpha} \right)^{\alpha n}$$

which already has the same form as the claimed bound.

We now turn to the sets with cardinalities beyond $i^*$. We claim that the whole second part of the sum is at most a constant factor larger than $\binom{n}{i^*}^*$. Let $\ell \leq n^* - i^*$ be a positive integer. By Lemma 7, the relative difference of the term of the sum at position $i^* + \ell$ is

$$\frac{\binom{n}{i^* + \ell} (1 - (1-p)^{i^* + \ell}(1-p)^{n-i^* - \ell}m)}{\binom{n}{i^*}^* m(1-p)^{n-i^* - \ell} p^{i^* + \ell} \cdot w_{n,p}(i^* + \ell, m-1)}$$

$$\leq (\frac{\binom{n}{i^* + \ell}}{\binom{n}{i^*}}) \cdot \frac{(1-p)^{i^* + \ell} \cdot w_{n,p}(i^* + \ell, m-1)}{\binom{n}{i^*}^*} = (\frac{n}{i^*})^\ell \cdot \frac{1-p}{1-p} = \frac{1-p}{1-p} w_{n,p}(i^* + \ell, m-1).$$

In the following, we estimate the first and last factor in this product using that $m$ is at most $1/(1-p)^{(1-\alpha)n}$, i.e., that $\alpha \leq 1 - \varepsilon$ is bounded away from 1.

$$\frac{\binom{n}{i^* + \ell}}{\binom{n}{i^*}} = \prod_{j=1}^{\ell} \frac{n - i^* - j}{i^* + j} \leq \left( \frac{n - i^*}{i^*} \right)^\ell = \left( \frac{\alpha}{1-\alpha} \right)^\ell \leq \left( \frac{1}{\varepsilon} - 1 \right)^\ell.$$ 

We apply Corollary 11 to the weighting factor and get

$$w_{n,p}(i^* + \ell, m-1) \leq \frac{1}{p} \exp(-m(1-p)^{n-i^* - \ell+1}) = \frac{1}{p} \exp(-(1-p)^{-\ell + 1}).$$

Set the constants $a = (\frac{1}{\varepsilon}) - 1 \cdot p/(1-p)$ and $b = 1/(1-p)$. So far, we have established a bound of $s(\ell) = a^\ell/(p \cdot \exp(b^{\ell-1}))$ on the ratio between the term at $i^* + \ell$ and $\binom{n}{i^*}^*$. Observe that $s$ lacks any dependence on $n$, $m$, or $\alpha$. However, we claimed that $\sum_{\ell=1}^{\infty} s(\ell)$ is a constant. To complete the proof of the claim, we show the even stronger assertion that the series $s(\ell)$ is summable. To this end, we prove that there exists an $\ell_0$ such that for all $\ell \geq \ell_0$, $s(\ell) \leq 2^{-\ell}$ holds. Consider the sequence $t(\ell) = s(\ell) \cdot 2^\ell$. Both $a > 0$ and $b > 1$ hold and therefore $\ln(t(\ell)) = \ell \ln(2a) - \ln p - b^{\ell-1}$ diverges to $-\infty$ as $\ell$ increases. This in turn implies $t(\ell) \to 0$, in particular, there is an $\ell_0$ such that $t(\ell) \leq 1$ for all $\ell \geq \ell_0$.

$$\sum_{\ell=1}^{\infty} s(\ell) = \sum_{\ell=1}^{\infty} \frac{a^\ell}{p \cdot \exp(b^{\ell-1})} \leq \sum_{\ell=1}^{\ell_0} \frac{a^\ell}{p \cdot \exp(b^{\ell-1})} + \sum_{\ell=\ell_0+1}^{\infty} \frac{1}{2^\ell} \leq \sum_{\ell=1}^{\ell_0} \frac{a^\ell}{b \cdot \exp(b^{\ell-1})} + 2 = O(1).$$

Let $C = \sum_{\ell=1}^{\infty} s(\ell)$ denote the constant factor. With the help of Lemma 3 and the
symmetry of the entropy function \( H \), we finally arrive at an estimate for the second part.

\[
\sum_{i=i^*+1}^{n} \binom{n}{i} (1 - (1 - p^\prime (1 - p)^{n-i})^m) \cdot w_{n,p}(i,m) \leq C \binom{n}{i^*} p^{i^*} = C \binom{n}{(1 - \alpha)n} p^{(1 - \alpha)n} \\
\leq C \frac{2^{H(\alpha)n}}{\sqrt{\pi n(1 - \alpha)}} p^{(1 - \alpha)n} = C \frac{1}{\sqrt{n}} \cdot 2^{(H(\alpha) + (1 - \alpha) \log p)n}.
\]

Since \( \alpha \) is bounded away from both 0 and 1, the coefficient is a constant not larger than \( C/\sqrt{\pi \varepsilon(1 - p)} \). In summary, we have established an \( O(2^{(H(\alpha) + (1 - \alpha) \log p)n}) \) bound on both parts of the sum, which completes the proof. \hfill \□

After putting the corresponding upper bound in place, we can discuss the phase transition in full generality. For \( m \leq 1/(1 - p)n \), the expected size of the minimization is linear in \( m \). However, the trivial upper bound \( |\min(B_{n,m,p})| \leq m \) holds also for higher \( m \). Conversely, if the total number of edges is above this while still being smaller than \( 1/(1 - p)n \), the minimization follows \( 2^{H(\alpha)\log(1 - \alpha) + (1 - \alpha) \log p)n} \), where \( \alpha = -\log_{1-p} m/n \). That means, in the information-theoretic regime the size of the minimization continues to grow at first but now sublinearly w.r.t. \( m \). After peaking at its maximum, it is even falling as the number of trials further increases. Still, the lower bound shown in Lemma 14 is also valid for small \( m \), although not tight. This overlap is indicated in Figure 12 by dashed lines.

The differences of the bounds stem from their respective focus. The linear upper bound is for the whole sample, where the dominant edge size is \( np \). The information-theoretic one instead employs the number of edges at the threshold as an estimate for \( |\min(B_{n,m,p})| \). That is to say, the dominant edge size is assumed to be \( i^* = (1 - \alpha)n \). By contrasting the sample size and the expected number of edges with \( i^* \) vertices, we can quantify their multiplicative difference in terms of the Kullback–Leibler divergence \( D(x \parallel y) = -x \log \left( \frac{x}{y} \right) - (1 - x) \log \left( \frac{1 - y}{1 - x} \right) \).

For \( m = 1/(1 - p)\alpha n \), this difference is

\[
\frac{m}{2^{H(\alpha) + (1 - \alpha) \log p}n} = \left( \frac{\alpha}{1 - p} \right)^{\alpha} \left( \frac{1 - \alpha}{p} \right)^{1 - \alpha} \right)^n = 2^{D(1 - \alpha \parallel p)n},
\]

as we have seen before in the proof of Lemma 16. Recall that \( D(1 - \alpha \parallel p)n \) is the divergence between two order-\( n \) binomial distributions with parameters \( 1 - \alpha \) and \( p \). Above equality is not fully surprising since the divergence marks, by definition, the information loss when assuming that the dominant edge size is \( (1 - \alpha)n \), while in reality it is \( np \).

Nevertheless, this observation has some interesting consequences. \( D(1 - \alpha \parallel p) \) as a function of \( \alpha \) is convex with the sole minimum at \( \alpha = 1 - p \). The multiplicative gap is exponential in \( n \) for small \( m \), namely, for \( m = 1, (\alpha = 0) \), it grows up to \( 1/p^n \). But as \( m \) approaches \( 1/(1 - p)(1 - p)^n \), i.e., \( 1 - \alpha \) close to \( p \), the gap vanishes. For larger \( m \), the difference increases again. The threshold \( i^* \) is below \( pm \) if and only if \( m > 1/(1 - p)(1 - p)^n \).

The corresponding upper and lower bounds for the information-theoretic regime show that assuming a dominant edge size of \( i^* \) in fact gives the better estimate there.

### 5.3 The Maximum Number of Minimal Sets

The size of the minimization grows linearly with \( m \), around \( 1/(1 - p)(1 - p)^n \) this trend slows down, but \( |\min(B_{n,m,p})| \) continues to increase. For even larger \( m \), the threshold \( i^* \) gets close to 0, and the amount of minimal sets goes down with it. Finally, when the number of trials crosses \( 1/(1 - p)n \), the minimization collapses under the sheer likelihood that the empty set
is sampled. This suggests that there is a sweet spot for which the size of the minimization is maximum. We apply the main theorem to calculate the maximum of $E[|\min(B_{n,m,p})|]$ over any number of trials. In more detail, we want to show a bound of $(1+p)^n$.

First, observe that $(1-p)^{p-1} < 1+p$ holds for all probabilities $p > 0$. This can be seen, for example, from the fact that $(1-p)^{p-1}$ is strictly concave on the open unit interval and $1+p$ is its tangent at position $p=0$. By the first statement of Theorem 1, $m$ in the linear regime are too small to lead to a proof of the bound. Conversely, $2^{H(1-\varepsilon)+\varepsilon ld p}$ as a function of $\alpha$ converges to 1 from above as $\alpha \to 1$, regardless of $p$. We can thus choose an $\varepsilon > 0$ small enough such that $2^{H(1-\varepsilon)+\varepsilon ld p} < 1+p$, doing so may only increase the leading constant of the bound. The sought maximum then occurs for an $m \leq 1/(1-p)^{(1-\varepsilon)n}$. That is to say, we only need to look at the information-theoretic regime.

\textbf{Lemma 17 (Theorem 2).} For any $0 \leq p < 1$, it holds that

1. $\max_{m \geq 1} E[|\min(B_{n,m,p})|] = O((1+p)^n)$ and
2. $\max_{m \geq 1} E[|\min(B_{n,m,p})|] = \Omega((1+p)^n/\sqrt{n})$.

The maximum is attained for $m = 1/(1-p)^{\frac{1}{\sqrt{n}}}$. 

\textbf{Proof.} There is nothing to show for $p = 0$. If $p \neq 0$, we use the rewrite $m = 1/(1-p)^{\alpha n}$. Fix an $\varepsilon$ with $0 < \varepsilon < p$ such that $H(1-\varepsilon) + \varepsilon ld p < ld(1+p)$. Theorem 1 shows that there exist positive constants $C_1, C_2 > 0$, possibly depending on $p$ and $\varepsilon$, such that

$$\frac{C_1}{\sqrt{n}} \cdot 2^{(H(\alpha)+(1-\alpha) ld p)n} \leq E[|\min(B_{n,m,p})|] \leq C_2 \cdot 2^{(H(\alpha)+(1-\alpha) ld p)n}.$$

for all $\alpha$ such that $1-p \leq \alpha \leq 1-\varepsilon$ and all $n$ sufficiently large.

Hence, we only need to determine the extremum of the exponential part. Let $g(\alpha,p) = H(\alpha)+(1-\alpha) \cdot ld p = -\alpha \cdot ld \alpha - (1-\alpha) \cdot ld(1-\alpha) + (1-\alpha) \cdot ld p$ be the exponent of $2^n$ above. As a sum of positive concave functions, $g$ is positive and concave as well. Its partial derivative

$$\frac{\partial g(\alpha,p)}{\partial \alpha} = \frac{1}{\alpha} - \frac{1}{1+p} \cdot ld \left(\frac{1}{1+p}\right) - \frac{p}{1+p} \cdot ld \left(\frac{p}{1+p}\right) + \frac{p}{1+p} \cdot ld p = ld(1+p).$$

has a single zero in the open interval $(1-p, 1-\varepsilon)$ at $\alpha^* = 1/(1+p)$, resulting in an exponent

$$g(\alpha^*,p) = -\frac{1}{1+p} \cdot ld \left(\frac{1}{1+p}\right) - \frac{p}{1+p} \cdot ld \left(\frac{p}{1+p}\right) + \frac{p}{1+p} \cdot ld p = ld(1+p).$$

This corresponds to an upper and lower bound on the maximum expected size of $\min(B_{n,m,p})$ of $C_2 (1+p)^n$ and $C_1 (1+p)^n/\sqrt{n}$, respectively. The number of trials for which the maximum is attained is $m = 1/(1-p)^{\frac{1}{\sqrt{n}}} = 1/(1-p)^{\frac{1}{\sqrt{n}}}$. \hfill ▶

\section{Conclusion}

We investigated the minimization of random hypergraphs. If the number of edges $m$ is at most $1/(1-p)^{(1-p)n}$, the size of the minimization is linear. When increasing the number of edges beyond that point, the minimization continues to grow sublinearly until $m$ passes $1/(1-p)^{\frac{1}{\sqrt{n}}}$. From there on, the size of the minimization drops quickly. Increasing $m$ significantly above $1/(1-p)n$ leads to the degenerated minimization consisting of only the empty set. An immediate extension of our work is to close the $\sqrt{n}$-gap between the information-theoretic bounds. In fact, we conjecture the lower bound to be tight. The only
place in the proofs where we lose more than a constant factor is when upper bounding the number of minimal sets with cardinality below the threshold in Lemma 16. Another interesting question in light of the application to random databases would be to incorporate different sample probabilities per vertex and dependencies between the elements. To fit the maximum-entropy setting, this would have to be modeled as additional constraints.
References

1. Z. Abedjan, L. Golab, and F. Naumann. Profiling Relational Data: A Survey. *The VLDB Journal*, 24:557–581, 2015.
2. K. Anand and G. Bianconi. Entropy Measures for Networks: Toward an Information Theory of Complex Topologies. *Physical Review E*, 80:045102, 2009.
3. G. Badkobeh, P. K. Lehre, and D. Sudholt. Black-box Complexity of Parallel Search with Distributed Populations. In *Proceedings of the 2015 Conference on Foundations of Genetic Algorithms (FOGA)*, pp. 3–15, 2015.
4. M. Behrisch, A. Coja-Oghlan, and M. Kang. The order of the giant component of random hypergraphs. *Random Structures and Algorithms*, 36:149–184, 2010.
5. M. Behrisch, A. Coja-Oghlan, and M. Kang. Local Limit Theorems for the Giant Component of Random Hypergraphs. *Combinatorics, Probability and Computing*, 23:331–366, 2014.
6. G. Bianconi. The Entropy of Randomized Network Ensembles. *Europhysics Letters*, 81:28005, 2007.
7. T. Bläsius, T. Friedrich, and M. Schirneck. The Parameterized Complexity of Dependency Detection in Relational Databases. In *Proceedings of the 11th International Symposium on Parameterized and Exact Computation (IPEC)*, pp. 6:1–6:13, 2016.
8. B. Bollobás. *Random Graphs*. Studies in Advanced Mathematics. Cambridge University Press, Cambridge, UK, 2nd edition, 2001.
9. B. Bollobás. A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs. *European Journal of Combinatorics*, 1:311 – 316, 1980.
10. M. Borassi, P. Crescenzi, and M. Habib. Into the Square: On the Complexity of Some Quadratic-time Solvable Problems. *Electronic Notes in Theoretical Computer Science*, 322:51–67, 2016.
11. A. A. Bruen and M. A. Forcinito. *Cryptography, Information Theory, and Error-Correction: A Handbook for the 21st Century*. Wiley-Interscience, New York, NY, USA, 2004.
12. P. S. Chodrow. Configuration Models of Random Hypergraphs. *ArXiv e-prints*, 2019.
13. C. Cooper, A. M. Frieze, and W. Pegden. On the Rank of a Random Binary Matrix. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 946–955, 2019.
14. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience, New York, NY, USA, 2nd edition, 2006.
15. J. Demetrovics, G. O. H. Katona, D. Miklós, O. Seleznev, and B. Thalheim. Asymptotic Properties of Keys and Functional Dependencies in Random Databases. *Theoretical Computer Science*, 190:151–166, 1998.
16. D. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, New York, NY, USA, 2009.
17. P. Erdős and A. Rényi. On Random Graphs I. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.
18. J. Gao, R. Impagliazzo, A. Kolokolova, and R. Williams. Completeness for First-order Properties on Sparse Structures with Algorithmic Applications. *Transactions on Algorithms*, 15:23:1–23:35, 2018.
19. D. Garlaschelli and M. I. Loffredo. Maximum Likelihood: Extracting Unbiased Information from Complex Networks. *Physical Review E*, 78:015101, 2008.
20. E. N. Gilbert. Random Graphs. *Annals of Mathematical Statistics*, 30:1141–1144, 1959.
21 Y. Grandvalet and Y. Bengio. Entropy Regularization. In O. Chapelle, B. Schölkopf, and A. Zien, editors, *Semi-Supervised Learning*, chapter 9. MIT Press, Cambridge, MA, USA, 2006.

22 P. Harremoës. Binomial and Poisson Distributions as Maximum Entropy Distributions. *IEEE Transactions on Information Theory*, 47:2039–2041, 2001.

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

24 E. T. Jaynes. Information Theory and Statistical Mechanics. *Physical Review Series II*, 106:620–630, 1957.

25 E. T. Jaynes. Information Theory and Statistical Mechanics. II. *Physical Review Series II*, 108:171–190, 1957.

26 M. Karoński and T. Łuczak. The Phase Transition in a Random Hypergraph. *Journal of Computational and Applied Mathematics*, 142:125–135, 2002.

27 G. O. H. Katona. Testing Functional Connection Between Two Random Variables. In A. N. Shiryaev, S. R. S. Varadhan, and E. L. Presman, editors, *Prokhorov and Contemporary Probability Theory*, pp. 335–348. Springer, Berlin and Heidelberg, Germany, 2013.

28 G. O. H. Katona. Random databases with correlated data. In A. Düsterhöft, M. Klettke, and K.-D. Schewe, editors, *Conceptual Modelling and Its Theoretical Foundations: Essays Dedicated to Bernhard Thalheim on the Occasion of His 60th Birthday*, pp. 29–35. Springer, Berlin and Heidelberg, Germany, 2012.

29 H. K. Kesavan. Jaynes’ Maximum Entropy Principle. In C. A. Floudas and P. M. Pardalos, editors, *Encyclopedia of Optimization*, pp. 1779–1782. Springer, Boston, MA, USA, 2009.

30 E. H. Lieb and J. Yngvason. The Physics and Mathematics of the Second Law of Thermodynamics. *Physics Reports*, 310:1–96, 1999.

31 R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, Cambridge, UK, 1995.

32 M. E. J. Newman. Scientific Collaboration Networks. I. Network Construction and Fundamental Results. *Physical Review E*, 64:016131, 2001.

33 M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, New York, NY, USA, 2010.

34 M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge, UK, 2010.

35 J. Park and M. E. J. Newman. The Statistical Mechanics of Networks. *Physical Review E*, 70:066117, 2004.

36 F. Saracco, R. Di Clemente, A. Gabrielli, and T. Squartini. Randomizing Bipartite Networks: The Case of the World Trade Web. *Scientific Reports*, 5:10595, 2015.

37 J. Schmidt-Pruzan and E. Shamir. Component Structure in the Evolution of Random Hypergraphs. *Combinatorica*, 5:81–94, 1985.

38 C. E. Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:379–423, 1948.

39 K. A. Zweig. *Network Analysis Literacy: A Practical Approach to the Analysis of Networks*. Lecture notes in social networks. Springer, Vienna, Austria, 2014.