The union-closed sets conjecture almost holds for almost all random bipartite graphs

Henning Bruhn and Oliver Schaudt

Abstract

Frankl's union-closed sets conjecture states that in every finite union-closed family of sets, not all empty, there is an element in the ground set contained in at least half of the sets. The conjecture has an equivalent formulation in terms of graphs: In every bipartite graph with least one edge, both colour classes contain a vertex belonging to at most half of the maximal stable sets.

We prove that, for every fixed edge-probability, almost every random bipartite graph almost satisfies Frankl's conjecture.

1 Introduction

One of the most basic conjectures in extremal set theory is Frankl's conjecture on union-closed set families. A family $\mathcal{F}$ of sets is *union-closed* if $X \cup Y \in \mathcal{F}$ for all $X, Y \in \mathcal{F}$.

Union-closed sets conjecture. *Let $\mathcal{F} \neq \{\emptyset\}$ be a finite union-closed family of sets. Then there is a $x \in \bigcup \mathcal{F}$ that lies in at least half of the members of $\mathcal{F}$.*

While Frankl [11] dates the conjecture to 1979, it apparently did not appear in print before 1985, when it was mentioned as an open problem in Rival [19]. Despite being widely known, there is only little substantial progress on the conjecture.

The conjecture has two equivalent formulations, one in terms of lattices and one in terms of graphs. For the latter, let us say that a vertex set S in a graph is *stable* if no two of its vertices are adjacent, and that it is *maximally stable* if, in addition, every vertex outside S has a neighbour in S.

Conjecture 1 (Bruhn, Charbit, Schaudt and Telle [4])**.** *Let G be a bipartite graph with at least one edge. Then each of the two bipartition classes contains a vertex belonging to at most half of the maximal stable sets.*

We prove a slight weakening of Conjecture 1 for random bipartite graphs. For $\delta > 0$, we say that a bipartite graph *satisfies the union-closed sets conjecture up to δ* if each of its two bipartition classes has a vertex for which the number of maximal stable sets containing it is at most $\frac{1}{2} + \delta$ times the total number of maximal stable sets. A *random bipartite graph* is a graph on bipartition classes of cardinalities m and n, where any two vertices from different classes are independently joined by an edge with probability p. We say that *almost every* random bipartite graph has property P if for every $\varepsilon > 0$ there is an N such that, whenever $m+n \geq N$, the probability that a random bipartite graph on $m+n$ vertices has P is at least $1-\varepsilon$.

We prove:

Theorem 2. *Let $p \in (0,1)$ be a fixed edge-probability. For every $\delta > 0$, almost every random bipartite graph satisfies the union-closed sets conjecture up to δ.*

While Frankl's conjecture has attracted quite a lot of interest, a proof seems still out of reach. For a survey of the literature on this conjecture we refer to [5].

Some of the earliest results verified the conjecture for few sets or few elements in the ground set, that is, when $n = |\mathcal{F}|$ or $m = |\bigcup_{X\in\mathcal{F}} X|$ are small. The current best results show that the conjecture holds for $m \leq 11$, which is due to Bošnjak and Marković [3], and for $n \leq 46$, proved by Lo Faro [10] and independently Roberts and Simpson [20]. The conjecture is also known to be true when n is large compared to m, that is $n \geq 2^m - \frac{1}{2}\sqrt{2^m}$ (Nishimura and Takahashi [15]). The latter result was improved upon by Czédli [6], who shows that $n \geq 2^m - \sqrt{2^m}$ is enough. Recently, Balla, Bollobás and Eccles [1] pushed this to $n \geq \lceil \frac{1}{3}2^{m+1} \rceil$.

The lattice formulation of the conjecture was apparently known from very early on, as it is already mentioned in Rival [19]. Poonen [16] investigated several variants and gave proofs for geometric as well as distributive lattices. Reinhold [18] extended this, with a very concise argument, to lower semimodular lattices. Finally, the conjecture holds as well for large semimodular lattices and for planar semimodular lattices (Czédli and Schmidt [8]).

The third view, in terms of graphs, on the union-closed sets conjecture is more recent. So far, the graph formulation is only verified for chordal-bipartite graphs, subcubic bipartite graphs, bipartite series-parallel graphs and for bipartitioned circular interval graphs (Bruhn, Charbit, Schaudt and Telle [4]).

One of the main techniques that is used for the set formulation of Frankl's conjecture as well as for the lattice formulation, is averaging: The average frequency of an element is computed, and if that average is at least half of the size of the family, it is concluded that the conjecture holds for the family. Averaging is also our main tool. We discuss averaging and its limits in Section 3.

2 Basic tools and definitions

In our graph-theoretic notation we usually follow Diestel [9], while we refer to Bollobás [2] for more details on random graphs.

All our graphs are finite and simple. We always consider a bipartite graph G to have a fixed bipartition, which we denote by $(L(G), R(G))$. When discussing the bipartition classes, we will often refer to $L(G)$ as the *left side* and to $R(G)$ as the *right side* of the graph.

Throughout the paper we consider a fixed edge probability p with $0 < p < 1$; and we will always put $q = 1-p$. A *random bipartite graph* G is a bipartite graph where every pair $u \in L(G)$ and $v \in R(G)$ is joined by an edge independently with probability p. We denote by $\mathcal{B}(m, n; p)$ the probability space whose elements are the random bipartite graphs G with $|L(G)| = m$ and $|R(G)| = n$. We will always tacitly assume that $m \geq 1$ and $n \geq 1$. Indeed, if one of the sides of the random bipartite graph is empty, then the graph has no edge and is therefore trivial with respect to Conjecture 1.

Markov's inequality states that for a non-negative random variable X and any $\alpha > 0$,

$$\Pr[X \geq \alpha] \leq \frac{\mathrm{E}[X]}{\alpha}. \tag{1}$$

Chebyshev's inequality is as follows. Let X be a random variable with finite variance $\sigma^2 = \mathrm{E}[X^2] - \mathrm{E}[X]^2$. Then, for every real $\lambda > 0$,

$$\Pr[|X - \mathrm{E}[X]| \geq \lambda] \leq \frac{\sigma^2}{\lambda^2}. \tag{2}$$

3 Discussion of averaging

Most of the partial results on Frankl's conjecture are based on one of two techniques: *Local configurations* and *averaging*. By a local configuration we mean a subfamily of the union-closed family $\mathcal{F}$, that guarantees that one element of the ground set lies in at least half of the members of $\mathcal{F}$. For instance, one of the earliest results is the observation of Sarvate and Renaud [21] that the element of a singleton will always belong to at least half of the sets. More local configurations have later been found by Poonen [16], Vaughan [23], Morris [14] and others.

The second technique consists in taking the average of the number of member sets containing a given element, where the average ranges over the set $U = \bigcup \mathcal{F}$ of all elements. If that average is at least $\frac{1}{2}|U|$ then clearly $\mathcal{F}$ will satisfy the conjecture. Averaging was used successfully by Balla, Bollobás and Eccles [1] when $n \geq \lceil \frac{1}{3} 2^{m+1} \rceil$. Reimer [17] showed that the average is always at least $\log_2(|U|)$.

Averaging will not always work. It is easy to construct union-closed families in which the average is too low. Czédli, Maróti and Schmidt [7] even found such families of size $|\mathcal{F}| = \lfloor 2^{|U|+1}/3 \rfloor$. Nevertheless, we will see that, in the graph formulation, averaging will almost always allow us to conclude that the union-closed sets conjecture is satisfied (up to any $\delta > 0$).

To describe the averaging technique for bipartite graphs, let us write $\mathcal{A}(G)$ for the set of maximal stable sets of a bipartite graph G. The graph formulation of the union-closed sets conjecture, Conjecture 1, is satisfied if G contains an *sparse* vertex in both bipartition classes, that is, a vertex that lies in at most half of the maximal stable sets. We note first that exchanging the sides turns a random bipartite graph $G \in \mathcal{B}(m, n; p)$ into a member of $\mathcal{B}(n, m; p)$, which means that it will suffice to show the existence of a sparse vertex in $L(G)$. All the discussion and proofs that follows will focus on the left side $L(G)$.

That a vertex v is sparse means that $|\mathcal{A}_v(G)|$, the number of maximal stable sets containing v, is at most $\frac{1}{2}|\mathcal{A}(G)|$. Thus, if for the average

$$\sum_{v\in L(G)} \frac{|\mathcal{A}_v(G)|}{|\mathcal{A}(G)|} \leq \frac{1}{2}|L(G)|$$

then $L(G)$ will contain a sparse vertex. Double-counting shows that the above average is equal to

$$\text{left-avg}(G) := \sum_{A\in\mathcal{A}(G)} \frac{|A\cap L(G)|}{|\mathcal{A}(G)|},$$

and thus our aim is to show that when $m+n$ is very large, it follows with high probability that $\text{left-avg}(G) \leq \frac{m}{2}$ for any $G \in \mathcal{B}(m,n;p)$.

Unfortunately, we will not reach this aim. While we will show for large parts of the parameter space (m,n) that the average is, with high probability, small enough, we will also see that when n is roughly $q^{-\frac{m}{2}}$ or larger the average becomes very close to $\frac{m}{2}$, so close that our tools are not sharp enough to separate the average from slightly above $\frac{m}{2}$. Therefore, we provide for a bit more space by settling on bounding the average away from $(\frac{1}{2}+\delta)m$ for any positive δ, which then only allows us to deduce the existence of a vertex $v \in L(G)$ that is *almost sparse*, in the sense that v lies in at most $(\frac{1}{2}+\delta)|\mathcal{A}(G)|$ maximal stable sets.

Much of the previous discussion is subsumed in the following lemma.

Lemma 3. *Let G be a bipartite graph, and let $\delta \geq 0$. If*

$$\text{left-avg}(G) \leq \left(\tfrac{1}{2}+\delta\right)|L(G)|$$

then there exists a vertex in $L(G)$ that lies in at most $\left(\frac{1}{2}+\delta\right)|\mathcal{A}(G)|$ maximal stable sets.

Proof. Double counting yields $\sum_{y\in L(G)} |\mathcal{A}_y(G)| = \sum_{A\in\mathcal{A}(G)} |A \cap L(G)|$, from which we deduce that $\sum_{y\in L(G)} |\mathcal{A}_y(G)| \leq |L(G)| \cdot \left(\frac{1}{2}+\delta\right)|\mathcal{A}(G)|$. Thus there is a $y \in L(G)$ with $|\mathcal{A}_y(G)| \leq \left(\frac{1}{2}+\delta\right)|\mathcal{A}(G)|$. $\square$

Most of the effort in this article will be spent on proving the following result, which is the heart of our main result, Theorem 2:

Theorem 4. *For all $\delta > 0$ and all $\varepsilon > 0$ there is an integer N so that for $G \in \mathcal{B}(m,n;p)$*

$$\Pr\left[\text{left-avg}(G) \leq \left(\tfrac{1}{2}+\delta\right)m\right] \geq 1-\varepsilon$$

for all m,n with $m+n \geq N$ and $n \geq \max\{20, (\lceil 3\log_{1/q}(2)\rceil+2)^2\}+1$.

In order to show how Theorem 2 follows from Theorem 4, we need to deal with the special case when one side is of constant size while the other becomes ever larger. Indeed, in this case averaging might fail—for a trivial reason. If we fix a constant right side $R(G)$, while $L(G)$ becomes ever larger, then $L(G)$ will contain many isolated vertices. Since the isolated vertices lie in every maximal stable set they may push up left-avg(G) to above $\frac{m}{2}$.

However, isolated vertices are never a threat to Frankl's conjecture: A bipartite graph satisfies the union-closed sets conjecture if and only if it satisfies the conjecture with all isolated vertices deleted. More generally, it turns out that the special case of a constant right side is easily taken care of:

Lemma 5. *Let c be a positive integer, and let $\varepsilon > 0$. Then there is an N so that for $G \in \mathcal{B}(m, n; p)$*

$$\Pr\left[L(G) \text{ contains a sparse vertex}\right] \geq 1 - \varepsilon,$$

for all m, n with $m \geq N$ and $n \leq c$.

Proof. Let G be any bipartite graph, and suppose there is a vertex $v \in L(G)$ that is adjacent with every vertex in $R(G)$. Then, the only maximal stable set that contains v is $L(G)$. Since the fact that v is incident with an edge implies that G has at least two maximal stable sets, v is sparse.

We now calculate the probability that there is such a vertex. The probability that $R(G) = N(v)$ for a fixed vertex $v \in L(G)$ is $p^n \geq p^c$ if $n \leq c$. Thus the probability that no such vertex exists in $L(G)$ is at most $(1 - p^c)^m$, which tends to 0 as $m \to \infty$. □

Proof of Theorem 2. Let $\delta > 0$ be given. By symmetry, it is enough to show that the left side $L(G)$ of almost every random bipartite graph G in $\mathcal{B}(m, n; p)$ contains a vertex that lies in at most $(\frac{1}{2} + \delta)|\mathcal{A}|$ maximal stable sets. For this, consider an $\varepsilon > 0$, and let N be the maximum of the N given by Theorem 4 and Lemma 5 with $c = \max\{20, (\lceil 3 \log_{1/q}(2) \rceil + 2)^2\}$. Consider a pair m, n of positive integers with $m + n \geq N$. If $n \leq \max\{20, (\lceil 3 \log_{1/q}(2) \rceil + 2)^2\}$ then Lemma 5 yields a sparse vertex in $L(G)$ with probability at least $1 - \varepsilon$. If, on the other hand, $n \geq \max\{20, (\lceil 3 \log_{1/q}(2) \rceil + 2)^2\} + 1$, Theorem 4 becomes applicable, which is to say that with probability at least $1 - \varepsilon$ we have left-avg$(G) \leq \left(\frac{1}{2} + \delta\right) m$. Now, Lemma 3 yields the desired vertex in $L(G)$. □

We close this section with the obvious but useful observation that if there are many more maximal stable sets with small left side than with large left side, then the average over the left sides is small, too. We will use this lemma repeatedly.

Lemma 6. *Let $\nu > 0$ and $\delta \geq 0$, and let G be a bipartite graph with $|L(G)| = m$. Let $\mathcal{L}$ be the maximal stable sets A of G with $|A \cap L(G)| \geq (\frac{1}{2} + \delta)m$, and let $\mathcal{S}$ be those maximal stable sets B with $|B \cap L(G)| \leq (1 - \nu)\frac{m}{2}$. If $|\mathcal{S}| \geq \frac{1}{\nu}|\mathcal{L}|$ then*

$$\text{left-avg}(G) \leq \left(\tfrac{1}{2} + \delta\right) m.$$

Proof. Let $\mathcal{M} = \mathcal{A}(G) \setminus (\mathcal{L} \cup \mathcal{S})$, that is, $\mathcal{M}$ is the set of those maximal stable sets A with $(1-\nu)\frac{m}{2} < |A \cap L(G)| < (\frac{1}{2}+\delta)m$. Then

$$\sum_{A\in\mathcal{A}(G)} \frac{|A\cap L(G)|}{\left(\frac{1}{2}+\delta\right)m} \leq \sum_{A\in\mathcal{L}} \frac{m}{\left(\frac{1}{2}+\delta\right)m} + \sum_{B\in\mathcal{M}} \frac{(\frac{1}{2}+\delta)m}{\left(\frac{1}{2}+\delta\right)m} + \sum_{C\in\mathcal{S}} \frac{(1-\nu)\frac{m}{2}}{\left(\frac{1}{2}+\delta\right)m}$$
$$\leq 2|\mathcal{L}| + |\mathcal{M}| + (1-\nu)|\mathcal{S}| = |\mathcal{A}(G)| - (\nu|\mathcal{S}| - |\mathcal{L}|).$$

Thus, from $|\mathcal{S}| \geq \frac{1}{\nu}|\mathcal{L}|$ it follows that $\sum_{A\in\mathcal{A}(G)} \frac{|A\cap L(G)|}{\left(\frac{1}{2}+\delta\right)m} \leq |\mathcal{A}(G)|$, which is equivalent to the inequality of the lemma. $\square$

4 Proof of Theorem 4

In order to prove Theorem 4, we distinguish several cases, depending on the relative sizes, m and n, of the two sides of the random bipartite graph $G \in \mathcal{B}(m,n;p)$. In each of the cases we need a different method.

The general strategy follows Lemma 6: We bound the number of maximal stable sets with large left side, usually counted by a random variable $\mathcal{L}_G$, and at the same time we show that there are many maximal stable sets with a small left side; those we count with $\mathcal{S}_G$.

Up to $n < q^{-\frac{m}{2}}$ we are able to use the same bound for the number $\mathcal{L}_G$ of maximal stable sets whose left sides are of size at least $\frac{m}{2}$: We prove that with high probability $\mathcal{L}_G$ is bounded by a polynomial in n. For right sides that are much larger than the left side, i.e. $m \gg n$, we even extend such a bound to maximal stable sets with left side $\geq \frac{m}{3}$.

For the maximal stable sets with small left side, counted by $\mathcal{S}_G$, we need to consider some cases. When the left side of the graph is much larger than the right side, namely $m \geq q^{-\sqrt[5]{n}}$, we find with high probability a large induced matching in G. This in turn implies that the total number of maximal stable sets is high, and thus clearly also the number of those with small left side.

When the sides of the graph do not differ too much in size, meaning $m \leq q^{-\sqrt[5]{n}}$ and $n \leq q^{-\sqrt[5]{m}}$, the variance of the number of maximal stable sets with small left side is moderate enough to apply Chebychev's inequality. Since the expectation of $\mathcal{S}_G$ is high, we again can use Lemma 6 to deduce Theorem 4.

However, when the left side of the graph becomes much larger than the right side, we cannot control the variance of $\mathcal{S}_G$ anymore. Instead, for $q^{-\sqrt[5]{m}} \leq n \leq q^{-\frac{m}{16}}$, we cut the right side into many pieces each of large size and apply Hoeffding's inequality to each of the pieces together with the left side. The inequality ensures that we find on at least one of the pieces a large number of maximal stable sets of small left side. Surpassing $n \geq q^{-\frac{m}{16}}$, we have to refine our estimations but we can still use this strategy up to slightly below $n = q^{-\frac{m}{2}}$.

In the interval $q^{-\frac{m}{2}} \leq n \leq q^{-m^3}$, we encounter a serious obstacle. There, we have to cope with an average that is very close to $\frac{m}{2}$. It is precisely for this reason that, overall, we only prove that left-avg$(G) \leq \left(\frac{1}{2}+\delta\right)m$ instead of left-avg$(G) \leq \frac{m}{2}$. To keep below the slightly higher average, we only need to bound the number of maximal stable sets with left side $> (\frac{1}{2} + \delta)m$. This number we will almost trivally bound by $2^{\lambda m}$, with some $\lambda < 1$. On the other hand, we will see that the number $\mathcal{S}_G$ of maximal stable sets of small left side is $2^{\lambda' m}$ with a λ' as close to 1 as we want.

In the remaining case, we are dealing with an enormous right side: $n \geq q^{-m^3}$. Then, it is easy to see that with high probability there is an induced matching that covers all of the left side, which implies that every subset of $L(G)$ is the left side of a maximal stable set. This immediately gives us left-avg$(G) = \frac{m}{2}$.

4.1 The case $m \geq q^{-\sqrt[5]{n}}$

In this section we treat the graphs whose left side is much larger than the right side. From an easy argument it follows that, with high probability, any large enough random graph $G \in \mathcal{B}(m, n; p)$ contains an induced matching of size $s \log_2(n)$, for any constant s. This directly implies that the total number of maximal stable sets is large. At the same time, we shall bound the number of maximal stable sets of large left side, which then shows that there are many of small left side.

However, if we take *large* left side to mean at least $\frac{m}{2}$ then it might be that most of those of *small* left side have a left side whose size is very close to $\frac{m}{2}$. Such a left side does not help much to drop the average. Therefore, we will consider a more generous notion of a large left side and bound the number of maximal stable sets that have a left side of $\geq \frac{m}{3}$; then small will mean $< \frac{m}{3}$.

For a random graph $G \in \mathcal{B}(m, n; p)$, let stab$(\geq \ell; \geq r)$ denote the number of stable sets of that have at least ℓ vertices in $L(G)$ and at least r vertices in $R(G)$.

Lemma 7. *Let $\ell^* \leq m$ and $r^* \leq n$ so that $nq^{\ell^*} \leq \frac{1}{2}$. Then for $G \in \mathcal{B}(m, n; p)$*

$$\mathrm{E}[stab(\geq \ell^*; \geq r^*)] \leq 2^{m+1} \left(nq^{\ell^*}\right)^{r^*}.$$

Proof. The expectation is given by

$$\begin{aligned}
\mathrm{E}[\mathrm{stab}(\geq \ell^*; \geq r^*)] &= \sum_{\ell=\ell^*}^{m} \sum_{r=r^*}^{n} \binom{m}{\ell}\binom{n}{r} q^{\ell r} \\
&\leq \left(\sum_{\ell=\ell^*}^{m} \binom{m}{\ell}\right) \left(\sum_{r=r^*}^{n} \binom{n}{r} q^{\ell^* r}\right) \\
&\leq 2^m \left(\sum_{r=r^*}^{n} \left(nq^{\ell^*}\right)^r\right) \leq 2^m \left(\sum_{r=r^*}^{\infty} \left(nq^{\ell^*}\right)^r\right).
\end{aligned}$$

The error estimation for the geometric series yields $\sum_{i=k}^{\infty} z^i \leq 2|z|^k$ for any $|z| \leq \frac{1}{2}$. Applying this for $z = nq^{\ell^*}$, we obtain the claimed bound of the lemma. □

In the following, we denote by $\mathcal{L}'_G$ the number of maximal stable sets S of a bipartite graph G with $|S \cap L(G)| \geq \frac{m}{3}$.

Lemma 8. *Let $r^* = \lceil 3 \log_{1/q}(2) \rceil + 1$. Then for any $\varepsilon > 0$ there is an N so that for $G \in \mathcal{B}(m, n; p)$*

$$\Pr[\mathcal{L}'_G \leq n^{r^*}] \geq 1 - \varepsilon$$

for all m, n with $m \geq q^{-\sqrt[5]{n}}$ and $m + n \geq N$.

Proof. Throughout the proof we assume that $m \geq q^{-\sqrt[5]{n}}$.

Setting $\ell^* = \frac{m}{3}$, we get from Lemma 7 that

$$\begin{aligned}\mathrm{E}[\mathrm{stab}(\geq \tfrac{m}{3}; \geq r^*)] &\leq 2^{m+1} \left(nq^{\frac{m}{3}}\right)^{r^*} \\ &= 2n^{r^*} \cdot q^{m(\frac{r^*}{3} - \log_{1/q}(2))} \leq 2n^{r^*} \cdot q^{\frac{m}{3}},\end{aligned}$$

by choice of r^*.

Choose N so that $2n^{r^*} \cdot q^{\frac{m}{3}} \leq \varepsilon$ for all m, n with $m + n \geq N$. Then, from Markov's inequality it follows that

$$\Pr[\mathrm{stab}(\geq \tfrac{m}{3}; \geq r^*) > 0] \leq \varepsilon. \tag{3}$$

The set of maximal stable sets S of G whose left side $S \cap L(G)$ has size at least $\frac{m}{3}$ is divided into those S with $|S \cap R(G)| \geq r^*$ and those whose right sides have $< r^*$ vertices; let the number of the latter ones be t. Since there are at most n^{r^*} subsets of $R(G)$ with at most r^* vertices, $t \leq n^{r^*}$. Hence,

$$\begin{aligned}\mathcal{L}'_G &\leq \mathrm{stab}(\geq \tfrac{m}{3}; \geq r^*) + t \\ &\leq \mathrm{stab}(\geq \tfrac{m}{3}; \geq r^*) + n^{r^*}.\end{aligned}$$

From (3), we deduce $\Pr[\mathcal{L}'_G > n^{r^*}] \leq \varepsilon$. $\square$

Lemma 9. *Let s be a positive integer, and let $\varepsilon > 0$. Then there is an N so that for $G \in \mathcal{B}(m, n; p)$*

$$\Pr\left[G \text{ has an induced matching of size } \geq s \log_2(n)\right] \geq 1 - \varepsilon$$

for all m, n with $m + n \geq N$, $n \geq s \log_2(n)$ and $m \geq q^{-\sqrt[5]{n}}$.

Proof. Let us assume throughout the proof that $n \geq s \log_2(n)$ and $m \geq q^{-\sqrt[5]{n}}$.

Put $k := \lceil s \log_2(n) \rceil$, and choose $\lfloor m/k \rfloor$ pairwise disjoint subsets $L_1, \ldots, L_{\lfloor m/k \rfloor}$ of size k of $L(G)$. Since $n \geq s \log_2(n)$, $n \geq k$ and so we may choose a set $R' \subseteq R(G)$ with $|R'| = k$. For $i = 1, \ldots, \lfloor m/k \rfloor$ let M_i be the random indicator variable for an induced matching of size k on $L_i \cup R'$. It is straightforward that

$$\Pr\left[M_i = 1\right] = k! p^k q^{k^2 - k} \geq p^k q^{k^2}.$$

Since the M_i are independent,

$$\Pr\left[\sum_{i=1}^{\lfloor m/k \rfloor} M_i = 0\right] \leq \left(1 - p^k q^{k^2}\right)^{\lfloor m/k \rfloor} \leq e^{-p^k q^{k^2} \lfloor m/k \rfloor},$$

using the standard inequality $1-x \leq e^x$ for all $x < 1$. Now for large $m+n$ the dominating term in $p^k q^{k^2} \lfloor m/k \rfloor$ is $q^{k^2} m$, since $m \geq q^{-\sqrt[5]{n}}$, which becomes arbitrarily large for large $m+n$ as $k = \lceil s \log_2(n) \rceil$ and $n \leq \left(\log_{1/q}(m)\right)^5$. Thus, there is an N so that $\Pr\left[\sum_{i=1}^{\lfloor m/k \rfloor} M_i = 0\right] \leq \varepsilon$ for all m,n with $m+n \geq N$. $\square$

We have now bounded the number $\mathcal{L}'_G$ of maximal stable sets of large left side, while the previous lemma will let us to conclude that the number of those with small left side is large. Together this allows us prove the first case of Theorem 4:

Lemma 10. *For every $\varepsilon > 0$ there exists an N so that for $G \in \mathcal{B}(m,n;p)$*

$$\Pr\left[\text{left-avg}(G) \leq \tfrac{m}{2}\right] \geq 1-\varepsilon,$$

for all m,n with $m+n \geq N$, $n \geq \max\{20, (\lceil 3\log_{1/q}(2)\rceil + 2)^2\}$ and $m \geq q^{-\sqrt[5]{n}}$.

Proof. Set $r^* = \lceil 3\log_{1/q}(2)\rceil + 1$. Choose N to be the maximum of the N obtained from Lemma 8 for $\frac{\varepsilon}{2}$ and the one from Lemma 9 applied to $s = r^*+1 = \lceil 3\log_{1/q}(2)\rceil + 2$ and $\frac{\varepsilon}{2}$.

Now, consider m,n with $m+n \geq N$, $n \geq \max\{20, (\lceil 3\log_{1/q}(2)\rceil + 2)^2\}$ and $m \geq q^{-\sqrt[5]{n}}$. We note that $n \geq \max\{20, (\lceil 3\log_{1/q}(2)\rceil + 2)^2\}$ implies that $n \geq s\log_2(n)$. By choice of s and N, we obtain from Lemma 9 that the probability that $G \in \mathcal{B}(m,n;p)$ does not contain an induced matching of size at least $s\log_2(n)$ is at most $\frac{\varepsilon}{2}$. On the other hand, the probability that the number $\mathcal{L}'_G$ of maximal stable sets A with $|A \cap L(G)| \geq \frac{m}{3}$ surpasses n^{r^*} is as well $\leq \frac{\varepsilon}{2}$. Thus, the probability that none of these two events occur is at least $1-\varepsilon$. We claim that in this case left-avg$(G) \leq \frac{m}{2}$.

So, assume that G contains an induced matching of cardinality $\geq s\log_2(n)$ and that $\mathcal{L}'_G \leq n^{r^*}$. There are at least $2^{s\log_2(n)}$ maximal stable sets on the subgraph restricted to the matching edges. Since each extends to a distinct maximal stable set of G, the number of maximal stable sets of G is at least $2^{s\log_2(n)} = n^s = n^{r^*+1}$. On the other hand, from $\mathcal{L}'_G \leq n^{r^*}$ it follows that at least $n^{r^*}(n-1)$ of the maximal stable sets have a left side of size at most $\frac{m}{3}$. As $n-1 \geq 3$, we may apply Lemma 6 with $\delta = 0$ in order to see that left-avg$(G) \leq \frac{m}{2}$. $\square$

The key observation in the argument above is that the number of maximal stable sets with a large left side is bounded by a polynomial in n, the size of the right-hand side. We will continue to exploit this, in a slightly strengthened version, below. The second part of the argument here is to note that there is always a relatively large induced matching, from which we deduce that the total number of maximal stable sets is not too small. Then we also have a large number of maximal stable sets with small left side, so that we are guaranteed a small average.

This strategy fails once m becomes smaller than n. Assume $m < n$ and, for simplicity, $p = q = \frac{1}{2}$. Below we will in that case bound the number of maximal stable sets with large left side by about $2n^2$. Thus, for our strategy to work, we should better find an induced matching of size at least $2\log_2(n)$. An easy calculation, however, shows that the expected number of induced matchings of size $2\log_2(n)$ is below one.

4.2 The case $n \le q^{-\sqrt[5]{m}}$ and $m \le q^{-\sqrt[5]{n}}$

From now on we will denote by $\mathcal{L}_G$ the number of maximal stable sets S with $|S \cap L(G)| \ge \frac{m}{2}$ of a random bipartite graph $G \in \mathcal{B}(m, n; p)$. We first bound $\mathcal{L}_G$ by a polynomial in n, a bound that will be useful up to slightly below $n = 2^{\frac{m}{2}}$.

Lemma 11. *For every $\alpha < \frac{1}{2}$ and every $\varepsilon > 0$ there exists an N so that for $G \in \mathcal{B}(m, n; p)$*

$$\Pr[\mathcal{L}_G \le 2n^{\log_q(1/4)}] \ge 1 - \varepsilon$$

when $m + n \ge N$ and $n \le q^{-\alpha m}$.

Proof. Let $\alpha < \frac{1}{2}$ be given and assume $n \le q^{-\alpha m}$.

We determine first the probability that a random bipartite graph contains many stable sets (not necessarily maximal) with left side $\ge \frac{m}{2}$ and right side $\ge \lfloor \log_q(1/4) \rfloor + 1$.

For this, note that $\alpha < \frac{1}{2}$ implies $nq^{\frac{m}{2}} \le q^{m(\frac{1}{2}-\alpha)} \le \frac{1}{2}$ for large m. Moreover, it follows that

$$\nu := (1/2 - \alpha)(\lfloor \log_q(1/4) \rfloor + 1 - \log_q(1/4)) > 0.$$

Thus, applying Lemma 7 yields

$$\begin{aligned}
&\mathrm{E}[\mathrm{stab}(\ge \tfrac{m}{2}; \ge \lfloor \log_q(1/4) \rfloor + 1)] \\
&\quad \le 2^{m+1} \left(nq^{\frac{m}{2}}\right)^{\lfloor \log_q(1/4) \rfloor + 1} \\
&\quad = 2n^{\log_q(1/4)} 2^m n^{\lfloor \log_q(1/4) \rfloor + 1 - \log_q(1/4)} q^{\frac{\lfloor \log_q(1/4) \rfloor + 1}{2} m} \\
&\quad \le 2n^{\log_q(1/4)} q^{-\log_q(1/2)m - \alpha(\lfloor \log_q(1/4) \rfloor + 1 - \log_q(1/4))m + \frac{\lfloor \log_q(1/4) \rfloor + 1}{2} m} \\
&\quad = 2n^{\log_q(1/4)} q^{m(1/2-\alpha)(\lfloor \log_q(1/4) \rfloor + 1 - \log_q(1/4))} \\
&\quad \le 2n^{\log_q(1/4)} q^{\nu m}
\end{aligned}$$

for sufficiently large m. With Markov's inequality we deduce

$$\Pr[\mathrm{stab}(\ge \tfrac{m}{2}; \ge \lfloor \log_q(1/4) \rfloor + 1) > n^{\log_q(1/4)}] \le \frac{2n^{\log_q(1/4)} q^{\nu m}}{n^{\log_q(1/4)}} = 2q^{\nu m},$$

which tends to 0 as $m \to \infty$. Since $n \le q^{-\alpha m}$ implies that also m must be large for large $m + n$, we may find an N so that $\Pr[\mathrm{stab}(\ge \frac{m}{2}; \ge \lfloor \log_q(1/4) \rfloor + 1) > n^{\log_q(1/4)}] \le \varepsilon$, for all for all integers m, n with $m + n \ge N$.

Considering such m and n, we turn now to the number of maximal stable sets $\mathcal{L}_G$ with left side $\ge \frac{m}{2}$. As in Lemma 8, we argue that the maximal stable sets counted by $\mathcal{L}_G$ split into those whose right side have at least $\lfloor \log_q(1/4) \rfloor + 1$ vertices and those with at most $\lfloor \log_q(1/4) \rfloor$ vertices in $R(G)$. Of the latter ones, there are at most $n^{\log_q(1/4)}$ many sets. By choice of N, the probability that we have more than $n^{\log_q(1/4)}$ of the former is bounded by ε. □

Let us quickly calculate the probability that a given set of vertices is a maximal stable set.

Lemma 12. *For a random bipartite graph $G \in \mathcal{B}(m, n; p)$, let S be a subset of $V(G)$. If $|S \cap L| = \ell$ and $|S \cap R| = r$ then*

$$\Pr[S \in \mathcal{A}] = q^{\ell r} \left(1 - q^r\right)^{m-\ell} \left(1 - q^{\ell}\right)^{n-r}.$$

Proof. The factor $q^{\ell r}$ is the probability that there is no edge from $S \cap L(G)$ to $S \cap R(G)$, that is, S is a stable set. The factor $(1 - q^r)^{m-\ell}$ is the probability that every of the $m - \ell$ many vertices in $L(G) \setminus S$ has a neighbour in $S \cap R(G)$, and $\left(1 - q^{\ell}\right)^{n-r}$ is the probability that every of the $n - r$ many vertices in $R(G) \setminus S$ has a neighbour in $S \cap L(G)$. The latter two conditions ensure that S is a maximal stable set. □

Next, we calculate the expectation and the variance of the number of maximal stable sets of small left side. Since they outnumber the other maximal stable sets by far, we concentrate on those maximal stable sets with a left side equal to $\approx \log_{1/q}(n)$ and a right side equal to $\approx \log_{1/q}(m)$. This choice is somewhat forced by the maximality requirement for maximal stable sets: For logarithmic sized left sides the maximality requirement eliminates only a constant proportion of the possible maximal stable sets. With smaller sides, on the other hand, we lose more sets due to the stability condition, that is, the expectation becomes much smaller.

For $G \in \mathcal{B}(m, n; p)$ we denote by $\mathcal{S}_G$ the number of maximal stable sets S of G with $|S \cap L(G)| = \lfloor \log_{1/q}(n) \rfloor$ and $|S \cap R(G)| = \lfloor \log_{1/q}(m) \rfloor$.

Lemma 13. *Let $c = e^{-(2/q+1)}$. There are $m_0, n_0 \in \mathbb{N}$ such that for $G \in \mathcal{B}(m, n; p)$*

$$\mathrm{E}[\mathcal{S}_G] \geq c \binom{m}{\lfloor \log_{1/q}(n) \rfloor} \lfloor \log_{1/q}(m) \rfloor^{-\lfloor \log_{1/q}(m) \rfloor}$$

for all $m \geq m_0$, $n \geq n_0$ with $m \geq \log_{1/q}(n)$ and $n \geq \log_{1/q}(m)$.

Proof. Assume that m, n are integers with $m \geq \log_{1/q}(n)$ and $n \geq \log_{1/q}(m)$. We first note that

$$\begin{aligned}
\left(1 - q^{\lfloor \log_{1/q}(m) \rfloor}\right)^{m - \lfloor \log_{1/q}(n) \rfloor} &\geq \left(1 - q^{\lfloor \log_{1/q}(m) \rfloor}\right)^{m} \\
&\geq \left(1 - q^{\log_{1/q}(m) - 1}\right)^{m} \\
&= \left(1 - \frac{1}{qm}\right)^{m}.
\end{aligned}$$

Since $\lim_{m\to\infty}\left(1-\frac{1}{qm}\right)^m = e^{-1/q}$, there is $m_0 \in \mathbb{N}$ such that

$$\left(1-\frac{1}{qm}\right)^m \geq e^{-\left(\frac{1}{q}+\frac{1}{2}\right)}, \tag{4}$$

for all $m \geq m_0$. With the same arguments, we see that there is an $n_0 \in \mathbb{N}$, so that

$$\left(1-q^{\lfloor \log_{1/q}(n)\rfloor}\right)^{n-\lfloor \log_{1/q}(m)\rfloor} \geq e^{-\left(\frac{1}{q}+\frac{1}{2}\right)}.$$

for all $n \geq n_0$.

Now consider a random bipartite graph $G \in \mathcal{B}(m,n;p)$ with $m \geq m_0$ and $n \geq n_0$, and let S be any vertex subset with $|S \cap L(G)| = \lfloor \log_{1/q}(n)\rfloor =: a$ and $|S \cap R(G)| = \lfloor \log_{1/q}(m)\rfloor =: b$. By Lemma 12, the probability that S is a maximal stable set of G amounts to

$$\Pr[S \text{ is maximally stable}] = q^{ab}\left(1-q^b\right)^{m-a}\left(1-q^a\right)^{n-b}.$$

The first term is at least equal to $n^{-\lfloor \log_{1/q}(m)\rfloor}$, while, by (4), the remaining terms together are at least equal to $c = e^{-(2/q+1)}$. This yields

$$\Pr[S \text{ is maximally stable}] \geq cn^{-\lfloor \log_{1/q}(m)\rfloor}.$$

Thus,

$$\begin{aligned}
\mathrm{E}[\mathcal{S}_G] &\geq \binom{m}{\lfloor \log_{1/q}(n)\rfloor}\binom{n}{\lfloor \log_{1/q}(m)\rfloor} cn^{-\lfloor \log_{1/q}(m)\rfloor} \\
&\geq \binom{m}{\lfloor \log_{1/q}(n)\rfloor}\left(\frac{n}{\lfloor \log_{1/q}(m)\rfloor}\right)^{\lfloor \log_{1/q}(m)\rfloor} cn^{-\lfloor \log_{1/q}(m)\rfloor} \\
&= c\binom{m}{\lfloor \log_{1/q}(n)\rfloor}\lfloor \log_{1/q}(m)\rfloor^{-\lfloor \log_{1/q}(m)\rfloor}.
\end{aligned}$$

□

We use Chebyshev's inequality to show that, with high probability, $\mathcal{S}_G$ does not differ much from the expected value.

Lemma 14. *For every $\varepsilon > 0$ there is an N so that for $G \in \mathcal{B}(m,n;p)$*

$$\Pr\left[\mathcal{S}_G > \tfrac{1}{2}\mathrm{E}[\mathcal{S}_G]\right] \geq 1-\varepsilon,$$

for all m,n with $m+n \geq N$, $m \leq q^{-\sqrt[5]{n}}$ and $n \leq q^{-\sqrt[5]{m}}$.

Proof. Since the statement of Lemma 14 is symmetric in m and n, we may assume throughout the proof that $m \leq n$. Moreover we assume that $m \leq q^{-\sqrt[5]{n}}$ and $n \leq q^{-\sqrt[5]{m}}$.

Let $a := \lfloor \log_{1/q}(n) \rfloor$ and $b := \lfloor \log_{1/q}(m) \rfloor$. Chebyshev's inequality (2) gives us

$$\Pr\left[\mathcal{S}_G \le \tfrac{1}{2}\mathrm{E}[\mathcal{S}_G]\right] \le \frac{4\sigma^2}{\mathrm{E}[\mathcal{S}_G]^2},$$

where $\sigma^2 = \mathrm{E}[\mathcal{S}_G^2] - \mathrm{E}[\mathcal{S}_G]^2$ is the variance of the random variable $\mathcal{S}_G$. We have

$$\mathrm{E}[\mathcal{S}_G^2] = \sum_{i=0}^{a} \sum_{j=0}^{b} A_{i,j},$$

where $A_{i,j}$ denotes the expected number of pairs (S,T) of maximal stable sets of G with $|S \cap L(G)| = a = |T \cap L(G)|$, $|S \cap R(G)| = b = |T \cap R(G)|$, $|S \cap T \cap L(G)| = i$, and $|S \cap T \cap R(G)| = j$. By Lemma 12,

$$\mathrm{E}[\mathcal{S}_G]^2 = \binom{m}{a}^2 \binom{n}{b}^2 q^{2ab}(1-q^a)^{2(n-b)}(1-q^b)^{2(n-a)}. \tag{5}$$

We will first show that there is an N_1 so that

$$\frac{A_{0,0} - \mathrm{E}[\mathcal{S}_G]^2}{\mathrm{E}[\mathcal{S}_G]^2} \le \frac{\varepsilon}{8}, \tag{6}$$

for all m, n with $m + n \ge N_1$.

To prove this, observe that

$$A_{0,0} \le \binom{m}{a}\binom{m-a}{a}\binom{n}{b}\binom{n-b}{b} q^{2ab}(1-q^a)^{2(n-2b)}(1-q^b)^{2(m-2a)}.$$

Indeed, while the binomial coefficients count the number of possibilities to choose the disjoint sets S and T, the factor q^{2ab} is the probability that S and T are stable sets. Furthermore, the probability that every vertex in $R(G) \setminus (S \cup T)$ has a neighbour in S and a neighbour in T is equal to $(1-q^a)^{2(n-2b)}$; the factor $(1-q^b)^{2(m-2a)}$ expresses the analogous probability for $L(G)$.

The above estimation for $A_{0,0}$ together with (5) yields $A_{0,0}/\mathrm{E}[\mathcal{S}_G]^2 \le (1-q^a)^{-2b}(1-q^b)^{-2a}$, and consequently

$$\frac{A_{0,0} - \mathrm{E}[\mathcal{S}_G]^2}{\mathrm{E}[\mathcal{S}_G]^2} \le (1-q^a)^{-2b}(1-q^b)^{-2a} - 1.$$

Next, note that if n is large enough so that $\frac{1}{nq} \le \frac{1}{2}$ then

$$\begin{aligned}(1-q^a)^{2b} &\ge (1-q^{\log_{1/q}(n)-1})^{2\log_{1/q}(m)} \ge \left(1 - \frac{1}{nq}\right)^{2\sqrt[5]{n}} \\ &\ge \left(e^{-\frac{2}{nq}}\right)^{2\sqrt[5]{n}} \to e^0 = 1 \text{ as } n \to \infty,\end{aligned}$$

where we have used that $1 - x \ge e^{-2x}$ for all $0 \le x \le 1/2$. The analogous estimation holds for $(1-q^b)^{-2a}$. Thus, if m and n are large enough, then

$$(1-q^a)^{-2b}(1-q^b)^{-2a} - 1 \le \frac{\varepsilon}{8}.$$

Since it follows from $m \le q^{-\sqrt[5]{n}}$ and $n \le q^{-\sqrt[5]{m}}$ that both of m and n have to be large if $m+n$ is large, we may therefore choose N_1 so that (6) holds.

We will now investigate $A_{i,j}/\mathrm{E}[\mathcal{S}_G]^2$ when $i+j \ge 1$. For this, let

$$B_{i,j} = \binom{m}{i}\binom{m-i}{a-i}\binom{m-a}{a-i}\binom{n}{j}\binom{n-j}{b-j}\binom{n-b}{b-j}q^{2ab-ij}. \tag{7}$$

Note that $B_{i,j}$ equals the expected number of pairs (S,T) of stable sets (not necessarily maximal) with $|S \cap L(G)| = a = |T \cap L(G)|$, $|S \cap R(G)| = b = |T \cap R(G)|$, $|S \cap T \cap L(G)| = i$, and $|S \cap T \cap R(G)| = j$. Hence, $A_{i,j} \le B_{i,j}$ for all $0 \le i \le a$ and $0 \le j \le b$.

For $r, s \in \mathbb{N}$ with $r \ge s$, let $(r)_s$ denote the s-th falling factorial of r, i.e., $(r)_s = r(r-1)\cdots(r-s+1)$. For the binomial coefficients appearing in $B_{i,j}/\mathrm{E}[\mathcal{S}_G]^2$ that involve m, we deduce

$$\frac{\binom{m}{i}\binom{m-i}{a-i}\binom{m-a}{a-i}}{\binom{m}{a}^2} = \frac{(m-a)_{a-i}\,(a)_i^2}{(m)_a\,i!} \le \frac{(a)_i^2}{(m)_i\,i!} \le \frac{(a)_i^2}{m^i} \le \frac{a^{2i}}{m^i},$$

for all integers i with $0 \le i \le a$. With the analogous estimation for the binomial coefficients involving n, we obtain

$$\frac{\binom{m}{i}\binom{m-i}{a-i}\binom{m-a}{a-i}\binom{n}{j}\binom{n-j}{b-j}\binom{n-b}{b-j}}{\binom{m}{a}^2\binom{n}{b}^2} \le \frac{a^{2i}b^{2j}}{m^i n^j}, \tag{8}$$

for all integers i, j with $0 \le i \le a$ and $0 \le j \le b$.

An easy calculation (very similar to (4)) shows that there is a constant c such that

$$c \ge 4(1-q^a)^{-2(n-b)}(1-q^b)^{-2(m-a)} \tag{9}$$

for all m.

Recalling the explicit expression (5) for $\mathrm{E}[\mathcal{S}_G]^2$, and then applying first (8) and then (9) we deduce

$$\begin{aligned}
\frac{(a+1)(b+1)B_{i,j}}{\mathrm{E}[\mathcal{S}_G]^2} &= \frac{(a+1)(b+1)\binom{m}{i}\binom{m-i}{a-i}\binom{m-a}{a-i}\binom{n}{j}\binom{n-j}{b-j}\binom{n-b}{b-j}q^{2ab-ij}}{\binom{m}{a}^2\binom{n}{b}^2 q^{2ab}(1-q^a)^{2(n-b)}(1-q^b)^{2(m-a)}} \\
&\overset{(8)}{\le} \frac{(a+1)(b+1)a^{2i}b^{2j}}{m^i n^j q^{ij}(1-q^a)^{2(n-b)}(1-q^b)^{2(m-a)}} \\
&\le \frac{4a^{2i+1}b^{2j+1}}{m^i n^j q^{ij}(1-q^a)^{2(n-b)}(1-q^b)^{2(m-a)}} \\
&\overset{(9)}{\le} \frac{ca^{2i+1}b^{2j+1}}{m^i n^j q^{ij}}
\end{aligned}$$

for all i, j with $i+j \ge 1$, and where we assume in the third step that m and n are large enough so that $a = \lfloor \log_{1/q}(n) \rfloor \ge 1$ and $b = \lfloor \log_{1/q}(m) \rfloor \ge 1$. (Again, this is possible since $m \leq q^{-\sqrt[5]{n}}$ and $n \leq q^{-\sqrt[5]{m}}$ implies that both of m and n have to be large if $m+n$ is large.)

In order to continue with the estimation we consider the term $m^i n^j q^{ij}$. For $0 \leq i \leq a$ and $0 \leq j \leq b$, we see that $m^i q^{ij} \geq (mq^b)^i \geq (mq^{\log_{1/q}(m)})^i = (\frac{m}{m})^i = 1$. In a similar way, we obtain $n^j q^{ij} \geq 1$. Now, if $i \geq j$ then $m^i n^j q^{ij} = m^i(n^j q^{ij}) \geq m^i$. If, on the other hand, $i < j$ then $n^j m^i q^{ij} \geq n^j \geq m^j$, since $m \leq n$. Thus

$$m^i n^j q^{ij} \geq m^{\max(i,j)}, \text{ for all } m, n \text{ with } m \leq n. \tag{10}$$

Using (10), we obtain

$$\begin{aligned}
\frac{(a+1)(b+1)B_{i,j}}{\mathrm{E}[\mathcal{S}_G]^2} &\leq \frac{ca^{2i+1}b^{2j+1}}{m^i n^j q^{ij}} \leq \frac{ca^{2i+1}b^{2j+1}}{m^{\max(i,j)}} \\
&\leq \frac{c\log_{1/q}(n)^{2i+1}\log_{1/q}(m)^{2j+1}}{m^{\max(i,j)}} \\
&\leq \frac{c(\sqrt[5]{m})^{2i+1}\log_{1/q}(m)^{2j+1}}{m^{\max(i,j)}} \\
&\leq \frac{cm^{\frac{2}{5}i+\frac{1}{5}}\log_{1/q}(m)^{2j+1}}{m^{\max(i,j)}},
\end{aligned}$$

since $n \leq q^{-\sqrt[5]{m}}$ and $m \leq n$. Now, since $i + j \geq 1$, the last term tends to 0 for $m \to \infty$. Therefore, there is an N_2 independent of i, j such that for all $m + n \geq N_2$

$$\frac{(a+1)(b+1)B_{i,j}}{\mathrm{E}[\mathcal{S}_G]^2} \leq \frac{\varepsilon}{8},$$

whenever $i + j \geq 1$.

Thus, for $N = \max(N_1, N_2)$ and m, n with $m + n \geq N$, we get with (6)

$$\begin{aligned}
\Pr[\mathcal{S}_G \leq \tfrac{1}{2}\mathrm{E}[\mathcal{S}_G]] &\leq \frac{4\sigma^2}{\mathrm{E}[\mathcal{S}_G]^2} \\
&\leq 4\frac{A_{0,0} - \mathrm{E}[\mathcal{S}_G]^2}{\mathrm{E}[\mathcal{S}_G]^2} \\
&+ \frac{4(a+1)(b+1)\max\{B_{i,j} : 0 \leq i \leq a, 0 \leq j \leq b, i+j \geq 1\}}{\mathrm{E}[\mathcal{S}_G]^2} \\
&\leq 4\frac{\varepsilon}{8} + 4\frac{\varepsilon}{8} = \varepsilon.
\end{aligned}$$

□

Let us quickly explain why the method of Lemma 14 ceases to work when $n \geq q^{-\sqrt{m}}$. In the proof we aim for $\Pr[\mathcal{S}_G < \frac{1}{2}\mathrm{E}[\mathcal{S}_G]]$ to tend to 0 with growing $m + n$. We achieve that by forcing the right hand side of (8) to vanish for large $m + n$, and all i, j with $i + j \geq 1$. In particular, when $i = 1$ and $j = 0$, we need $\frac{a^2}{m} = \lfloor \log_{1/q}(n) \rfloor^2 / m$ to vanish with growing m, which in turn requires that $n < q^{-\sqrt{m}}$.

To finish this case, we observe that, with high probability, Lemma 11 bounds the number $\mathcal{L}_G$ of maximal stable sets with large left side with a polynomial in n, while we will see below that, again with high probability, Lemmas 13 and 14 translate into superpolynomially many maximal stable sets with small left side.

Lemma 15. *For every $\varepsilon > 0$ there is an N so that for every $G \in \mathcal{B}(m, n; p)$*

$$\Pr[\text{left-avg}(G) \leq \tfrac{m}{2}] \geq 1 - \varepsilon$$

for all m, n with $m + n \geq N$, $m \leq q^{-\sqrt[5]{n}}$ and $n \leq q^{-\sqrt[5]{m}}$.

Proof. Choose N_1 large enough so that it is at least as large as the N in Lemma 13 with $\frac{\varepsilon}{4}$ and as the N in Lemma 14, as well with $\frac{\varepsilon}{4}$, and so that $\lfloor \log_{1/q}(n) \rfloor \geq \frac{1}{2}\log_{1/q}(n)$ for any m, n with $m + n \geq N_1$ and $m \leq q^{-\sqrt[5]{n}}$.

In the remainder of the proof, we consider integers m, n with $m + n \geq N_1$, $m \leq q^{-\sqrt[5]{n}}$ and $n \leq q^{-\sqrt[5]{m}}$. By Lemmas 13 and 14, there is a constant $c > 0$, independent of m and n, so that the probability that

$$\begin{aligned}
\mathcal{S}_G &\leq c\left(\frac{m}{\lfloor \log_{1/q}(n) \rfloor}\right)^{\lfloor \log_{1/q}(n) \rfloor} \lfloor \log_{1/q}(m) \rfloor^{-\lfloor \log_{1/q}(m) \rfloor} \\
&\leq c\binom{m}{\lfloor \log_{1/q}(n) \rfloor} \lfloor \log_{1/q}(m) \rfloor^{-\lfloor \log_{1/q}(m) \rfloor},
\end{aligned}$$

is at most $\frac{\varepsilon}{2}$.

Now, using that $\log_{1/q}(m) \leq \sqrt[5]{n}$ and $\log_{1/q}(n) \leq \sqrt[5]{m}$ we obtain

$$\begin{aligned}
&c\left(\frac{m}{\lfloor \log_{1/q}(n) \rfloor}\right)^{\lfloor \log_{1/q}(n) \rfloor} \lfloor \log_{1/q}(m) \rfloor^{-\lfloor \log_{1/q}(m) \rfloor} \\
&\geq cm^{\frac{1}{2}\log_{1/q}(n)} \cdot \log_{1/q}(n)^{-\log_{1/q}(n)} \cdot \log_{1/q}(m)^{-\log_{1/q}(m)} \\
&\geq cn^{\frac{1}{2}\log_{1/q}(m)} \cdot n^{-\log_{1/q}(\log_{1/q}(n))} \cdot \left(\sqrt[5]{n}\right)^{-\log_{1/q}(m)} \\
&\geq cn^{\frac{1}{2}\log_{1/q}(m)} \cdot n^{-\log_{1/q}(\sqrt[5]{m})} \cdot (n)^{-\frac{1}{5}\log_{1/q}(m)} = cn^{\frac{1}{10}\log_{1/q}(m)}.
\end{aligned}$$

It follows that

$$\Pr\left[\mathcal{S}_G \leq cn^{\frac{1}{10}\log_{1/q}(m)}\right] \leq \frac{\varepsilon}{2}. \tag{11}$$

On the other hand, we obtain from Lemma 11 that there is a N_2 so that

$$\Pr\left[\mathcal{L}_G > 2n^{\log_q(1/4)}\right] \leq \frac{\varepsilon}{2}, \tag{12}$$

whenever $m + n \geq N_2$.

Recall that $\mathcal{S}_G$ is a lower bound on the number of maximal stable sets S with $|S \cap L(G)| = \lfloor \log_{1/q}(n) \rfloor$. Since $m \leq q^{-\sqrt[5]{n}}$ and $n \leq q^{-\sqrt[5]{m}}$ implies that both of m and n have to be large if $m + n$ is large, we may choose N_3 large enough so that $\sqrt[5]{m} \le \frac{m}{4}$ and $cn^{\frac{1}{10}\log_{1/q}(m)} \ge 4n^{\log_q(1/4)}$ whenever $m+n \ge N_3$. We claim that

$$\textit{if } \mathcal{S}_G \ge cn^{\frac{1}{10}\log_{1/q}(m)} \textit{ and } \mathcal{L}_G \le 2n^{\log_q(1/4)} \textit{ then } \text{left-avg}(G) \le \tfrac{m}{2}. \tag{13}$$

Indeed, since $\lfloor \log_{1/q}(n) \rfloor \le \sqrt[5]{m} \le \frac{m}{4}$ this is a direct consequence of Lemma 6 with $\nu = \frac{1}{2}$ and $\delta = 0$.

Finally, the lemma follows from (11), (12) and (13) if N is chosen to be at least $\max(N_1, N_2, N_3)$. $\square$

4.3 The case $q^{-\sqrt[5]{m}} \le n \le q^{-\frac{m}{16}}$

When the right side of the random bipartite graph becomes much larger than the left side, we cannot control the variance of $\mathcal{S}_G$ anymore, and indeed Chebyshev's inequality cannot even give a positive probability, however small, that the graph contains many maximal stable sets of small left side. We therefore use another standard tool, Hoeffding's inequality, which yields a tiny but non-zero probability. We then leverage this tiny probability to a high probability by cutting up the right side of the graph into a large number of large pieces to each of which we apply Hoeffding's theorem:

Theorem 16 (Hoeffding [12]). *For $i = 1, \dots, s$, let $X_i : \Omega_i \to [0, \rho]$ be independent random variables, and let $X = \sum_{i=1}^{s} X_i$. Then*

$$\Pr[X \ge \mathrm{E}[X] + \lambda] \le e^{-\frac{2\lambda^2}{s\rho^2}}$$

and

$$\Pr[X \le \mathrm{E}[X] - \lambda] \le e^{-\frac{2\lambda^2}{s\rho^2}}$$

for every λ with $\lambda > 0$.

We will use Theorem 16 in the simpler case, when there is only one random variable, that is, s will be equal to 1.

By $\mathcal{S}'_G$ we denote the number of maximal stable sets S of G with $|S \cap L(G)| = \lfloor \log_{1/q}(\lfloor n/m^{\log_{1/q}(m)} \rfloor) \rfloor$. Note that, in contrast to $\mathcal{S}_G$, we put no restriction on $|S \cap R(G)|$.

Lemma 17. *For integers m, n let $a' := \lfloor \log_{1/q}(\lfloor n/m^{\log_{1/q}(m)} \rfloor) \rfloor$ and $b := \lfloor \log_{1/q}(m) \rfloor$. Then there exists a constant $c > 0$ so that for every $\varepsilon > 0$ there is an N such that for $G \in \mathcal{B}(m, n; p)$*

$$\Pr\left[\mathcal{S}'_G \ge c\binom{m}{a'} b^{-b}\right] \ge 1 - \varepsilon,$$

for all m, n with $m + n \ge N$ and $m^{2\log_{1/q}(m)} \le n \le q^{-m}$.

Proof. In the following we assume that $m^{2\log_{1/q}(m)} \le n \le q^{-m}$. Let $k = m^{\log_{1/q}(m)}$, and let $R_1, R_2, \dots, R_{\lfloor k \rfloor}$ be disjoint subsets of $R(G)$ of size $\lfloor n/k \rfloor$

each. Moreover, let $G_i = G[L(G) \cup R_i]$ for all $1 \leq i \leq \lfloor k \rfloor$. Note that every G_i may be considered as a random bipartite graph in $\mathcal{B}(m, \lfloor n/k \rfloor; p)$.

Let $a' = \lfloor \log_{1/q}(\lfloor n/k \rfloor) \rfloor$ and $b = \lfloor \log_{1/q}(m) \rfloor$. Note that, since $m^{2\log_{1/q}(m)} \leq n \leq q^{-m}$, also $m \geq a'$ and $n \geq b$. Recall that, for $1 \leq i \leq k$, $\mathcal{S}_{G_i}$ is the number of maximal stable sets of G_i with $|\mathcal{S}_{G_i} \cap L(G)| = a'$ and $|\mathcal{S}_{G_i} \cap R(G)| = b$.

Applying Lemma 13 to G_i, we get

$$\mathrm{E}[\mathcal{S}_{G_i}] \geq c\binom{m}{a'} b^{-b}, \tag{14}$$

where $c > 0$ is some constant.

Clearly the value of $\mathcal{S}_{G_i}$ does never exceed $\binom{m}{a'}$. Thus, by Theorem 16,

$$\Pr[\mathcal{S}_{G_i} \leq \tfrac{1}{2}\mathrm{E}[\mathcal{S}_{G_i}]] \leq e^{-\frac{1}{2}\mathbf{E}[\mathcal{S}_{G_i}]^2\binom{m}{a'}^{-2}} \overset{(14)}{\leq} e^{-\frac{1}{2}c^2 b^{-2b}}.$$

The edge sets of the G_i are pairwise disjoint, and thus the random variables $\mathcal{S}_{G_i}$ are independent:

$$\Pr[\mathcal{S}_{G_i} \leq \tfrac{1}{2}\mathbf{E}[\mathcal{S}_{G_i}] \text{ for all } 1 \leq i \leq \lfloor k \rfloor] \leq e^{-\lfloor k \rfloor \frac{1}{2}c^2 b^{-2b}}.$$

Since $\lfloor k \rfloor = \lfloor m^{\log_{1/q}(m)} \rfloor$, it dominates b^{-2b}. Hence, $\lim_{m\to\infty} \lfloor k \rfloor \frac{1}{2}c^2 b^{-2b} = \infty$. Thus, there is an N such that $e^{-\lfloor k \rfloor \frac{1}{2}c^2 b^{-2b}} \leq \varepsilon$ whenever $m + n \geq N$, as $m^{2\log_{1/q}(m)} \leq n \leq q^{-m}$ implies that m grows with $m + n$. Hence, assuming $m + n \geq N$,

$$\Pr[\mathcal{S}_{G_i} \leq \tfrac{1}{2}\mathbf{E}[\mathcal{S}_{G_i}] \text{ for all } 1 \leq i \leq \lfloor k \rfloor] \leq e^{-\lfloor k \rfloor \frac{1}{2}c^2 b^{-2b}} \leq \varepsilon, \tag{15}$$

Thus, with probability $1 - \varepsilon$, there is an i for which $\mathcal{S}_{G_i} \geq c\binom{m}{a'}b^{-b}$. Note that every maximal stable set S of G_i can be extended to a maximal stable set S' of G such that $S' \cap V(G_i) = S$. This extension is injective and, moreover, $|S' \cap L(G)| = |S \cap L(G)| = a'$. Hence, every maximal stable set counted by $\mathcal{S}_{G_i}$ is also counted by $\mathcal{S}'_G$, i.e., $\mathcal{S}'_G \geq \mathcal{S}_{G_i}$. This completes the proof. □

Observe that, in order to apply (15), we need k to dominate b^{2b}, where $b = \lfloor \log_{1/q}(m) \rfloor$. Hence, k and thus also n should be of the order at least $m^{2\log_{1/q}(m)\log_{1/q}(\log_{1/q}(m))}$. This means that we could not use this method before, when m and n had about the same size.

Lemma 18. *For every $\varepsilon > 0$ there is an N so that for $G \in \mathcal{B}(m, n; p)$*

$$\Pr\left[\text{left-avg}(G) \leq \tfrac{m}{2}\right] \geq 1 - \varepsilon,$$

for all m, n with $m + n \geq N$ and $q^{-\sqrt[5]{m}} \leq n \leq q^{-\frac{m}{16}}$.

Proof. In this proof consider integers m, n with $q^{-\sqrt[5]{m}} \leq n \leq q^{-\frac{m}{16}}$, and let $a' := \lfloor \log_{1/q}(\lfloor n/m^{\log_{1/q}(m)} \rfloor) \rfloor$. Note that $a' \geq \log_{1/q}(n) - \log_{1/q}(m^{\log_{1/q}(m)}) - 2$.

Then

$$\begin{aligned}
\binom{m}{a'} &\geq \left(\frac{m}{a'}\right)^{a'} \geq \left(\frac{m}{\log_{1/q}(n)}\right)^{a'} \\
&\geq \left(\frac{m}{\log_{1/q}(n)}\right)^{\log_{1/q}(n)-\log_{1/q}(m^{\log_{1/q}(m)})-2} \\
&\geq m^{\log_{1/q}(n)-(\log_{1/q}(m)^2+2)} \cdot \Big(\log_{1/q}(n)\Big)^{-\log_{1/q}(n)} \\
&= n^{\log_{1/q}(m)} \cdot m^{-(\log_{1/q}(m)^2+2)} \cdot n^{-\log_{1/q}(\log_{1/q}(n))} \\
&\geq n^{\log_{1/q}(m)} \cdot m^{-(\log_{1/q}(m)^2+2)} \cdot n^{-\log_{1/q}(m/16)} \\
&= n^{\log_{1/q}(16)} \cdot m^{-(\log_{1/q}(m)^2+2)} = n^{\log_{1/q}(8)} \cdot n^{\log_{1/q}(2)} \cdot m^{-(\log_{1/q}(m)^2+2)}
\end{aligned}$$

In Lemma 17 the binomial coefficient $\binom{m}{a'}$ is divided by $\lfloor \log_{1/q}(m) \rfloor^{\lfloor \log_{1/q}(m) \rfloor}$. So, let us compare this factor times $m^{\log_{1/q}(m)^2+2}$ against $n^{\log_{1/q}(2)}$. When m is large enough, which we may assume since $m^{2\log_{1/q}(m)} \leq n \leq q^{-m}$ implies that m grows with $m+n$, we get that $(\log_{1/q}(m))^2 \geq 2 + \log_{1/q}(\log_{1/q}(m))$. Thus

$$\begin{aligned}
m^{\log_{1/q}(m)^2+2} \cdot \lfloor \log_{1/q}(m) \rfloor^{\lfloor \log_{1/q}(m) \rfloor} &\leq m^{\log_{1/q}(m)^2+2} \cdot (\log_{1/q}(m))^{\log_{1/q}(m)} \\
&= m^{\log_{1/q}(m)^2+2+\log_{1/q}(\log_{1/q}(m))} \\
&\leq m^{2\log_{1/q}(m)^2}.
\end{aligned}$$

Using $q^{-\sqrt[5]{m}} \leq n$, we get

$$\begin{aligned}
m^{2\log_{1/q}(m)^2} &\leq \Big((\log_{1/q}(n))^5\Big)^{2(\log_{1/q}(\log_{1/q}(n)^5))^2} \\
&= \Big(\log_{1/q}(n)\Big)^{250(\log_{1/q}(\log_{1/q}(n))^2} \\
&= q^{-250(\log_{1/q}(\log_{1/q}(n))^3}.
\end{aligned}$$

Since $\log_{1/q}(2) > 0$ and $n^{\log_{1/q}(2)} = q^{-\log_{1/q}(2)\cdot\log_{1/q}(n)}$, we see that $n^{\log_{1/q}(2)} > m^{2\log_{1/q}(m)^2}$ for large enough m and n.

In conjunction with Lemma 17 this yields that there is a constant $c > 0$ and an N_1 so that

$$\Pr[\mathcal{S}'_G \geq cn^{\log_{1/q}(8)}] \geq 1 - \frac{\varepsilon}{2}, \tag{16}$$

whenever $m + n \geq N_1$.

On the other hand, Lemma 11 yields an N_2 so that

$$\Pr[\mathcal{L}_G > 2n^{\log_q(1/4)}] \leq \frac{\varepsilon}{2},$$

when $m + n \geq N_2$.

Now we choose an $N \geq \max(N_1, N_2)$ so that $2n^{\log_q(1/4)} = 2n^{\log_{1/q}(4)}$ is much smaller than $cn^{\log_{1/q}(8)}$, by a factor of 2, say, when $m+n \geq N$. (This is possible as $m^{2\log_{1/q}(m)} \leq n \leq q^{-m}$ implies that m is large when $m+n$ is large.) Thus $2\mathcal{S}'_G \geq \mathcal{L}_G$ with a probability of $\geq 1-\varepsilon$, and Lemma 6 (with $\delta = 0$) completes the proof. Indeed, note that $\mathcal{S}'_G$ counts the number of maximal stable sets whose left sides have size $\lfloor \log_{1/q}(\lfloor n/m^{\log_{1/q}(m)} \rfloor) \rfloor \leq \log_{1/q}(n) \leq \frac{m}{16}$. □

Note that the estimation leading to (16) ceases to work when $n > q^{-\frac{m}{4}}$. We need therefore a finer estimation, which is what we do in the next section.

4.4 The case $q^{-\frac{m}{16}} \leq n < q^{-\frac{m}{2}}$

For $\kappa \in (0,1)$ the *binary entropy* is defined as

$$H(\kappa) = \kappa \log_2(\tfrac{1}{\kappa}) + (1-\kappa)\log_2(\tfrac{1}{1-\kappa}).$$

Observe that the binary entropy $H(\kappa)$ is always strictly smaller than 1, except for $\kappa = \frac{1}{2}$. Moreover, H is monotonously increasing in the interval $\left[0, \frac{1}{2}\right]$. For further details see [13].

We will use the following bound on the binomial coefficient, which can be found for instance in Mitzenmacher and Upfal [13, Lemma 9.2].

Lemma 19. *For all $m, k \in \mathbb{N}$ with $0 < k < m$,*

$$\binom{m}{k} \geq \frac{1}{m+1} \cdot 2^{H(k/m)\cdot m}.$$

Lemma 20. *For integers m, n let $\lambda := \log_{1/q}(n)/m$. For every $\varepsilon, \varphi > 0$ there is an N such that for $G \in \mathcal{B}(m,n;p)$,*

$$\Pr[\mathcal{S}'_G \geq 2^{(1-\varphi)\cdot H(\lambda)\cdot m}] \geq 1-\varepsilon,$$

for all m, n with $m+n \geq N$ and $q^{-\frac{m}{16}} \leq n \leq q^{-\frac{m}{2}}$.

Proof. Throughout the proof assume $q^{-\frac{m}{16}} \leq n \leq q^{-\frac{m}{2}}$.

Let $\varphi > 0$. First we choose a δ with $0 < \delta < 1$ which satisfies

$$H((1-\delta)\kappa) \geq \left(1 - \frac{\varphi}{2}\right) \cdot H(\kappa) \tag{17}$$

for all $\kappa \in \left[\frac{1}{16}, \frac{1}{2}\right]$. This is possible since H is uniformly continuous in $\left[\frac{1}{16}, \frac{1}{2}\right]$ and $\min\{H(\kappa) : \kappa \in \left[\frac{1}{16}, \frac{1}{2}\right]\} > 0$.

Let $a' := \lfloor \log_{1/q}(\lfloor n/m^{\log_{1/q}(m)} \rfloor) \rfloor$. Let N_1 be such that when $m+n \geq N_1$

$$\lceil (1-\delta)\log_{1/q}(n) \rceil \leq a' \leq \left\lfloor \tfrac{m}{2} \right\rfloor. \tag{18}$$

The choice of N_1 is possible since $\delta > 0$ and $\log_{1/q}(n) \leq \frac{m}{2}$. In the following, we restrict our attention to these m, n with $m+n \geq N_1$. From (18) it follows that

$$\binom{m}{a'} \geq \binom{m}{\lceil (1-\delta)\log_{1/q}(n) \rceil}. \tag{19}$$

Lemma 19 gives

$$\binom{m}{\lceil (1-\delta)\log_{1/q}(n)\rceil} \geq \frac{1}{m+1} \cdot 2^{H(\lceil (1-\delta)\log_{1/q}(n)\rceil m^{-1})\cdot m}. \tag{20}$$

Since $H(\kappa)$ is monotonically increasing for $\kappa \in \left[0, \frac{1}{2}\right]$, it follows that $H(\lceil(1-\delta)\log_{1/q}(n)\rceil m^{-1}) \geq H((1-\delta)\lambda)$, where we recall that $\lambda = \log_{1/q}(n)/m$. As $\frac{1}{16} \leq \lambda \leq \frac{1}{2}$, we get

$$2^{H(\lceil (1-\delta)\log_{1/q}(n)\rceil m^{-1})\cdot m} \geq 2^{H((1-\delta)\lambda)\cdot m} \overset{(17)}{\geq} 2^{\left(1-\frac{\varphi}{2}\right)H(\lambda)\cdot m}. \tag{21}$$

Lemma 17 gives an N_2 and a constant $c > 0$ such that for $b = \lfloor \log_{1/q}(m) \rfloor$

$$\Pr\left[\mathcal{S}'_G \geq c\binom{m}{a'} b^{-b}\right] \geq 1 - \varepsilon,$$

when $m + n \geq N_2$. Since $q^{-m/16} \leq n \leq q^{-m/2}$, there is an N_3 such that $m + n \geq N_3$ implies

$$\frac{c}{m+1} \cdot b^{-b} \cdot 2^{\frac{\varphi}{2}\cdot H(\lambda)\cdot m} \geq 1,$$

where we use that $\lambda \geq \frac{1}{16}$. For such m and n,

$$\frac{c}{m+1} \cdot b^{-b} \cdot 2^{(1-\frac{\varphi}{2})\cdot H(\lambda)\cdot m} \geq 2^{(1-\varphi)\cdot H(\lambda)\cdot m}. \tag{22}$$

Now, taking $N = \max(N_1, N_2, N_3)$, the inequalities (19), (20), (21) and (22) finish the proof. □

Lemma 21. *For every α with $\frac{1}{16} \leq \alpha < \frac{1}{2}$ and every $\varepsilon > 0$ there is an N so that for $G \in \mathcal{B}(m, n; p)$*

$$\Pr\left[\text{left-avg}(G) \leq \tfrac{m}{2}\right] \geq 1 - \varepsilon,$$

for all m, n with $m + n \geq N$ and $q^{-\frac{m}{16}} \leq n \leq q^{-\alpha m}$.

Proof. Let $\frac{1}{16} \leq \alpha < \frac{1}{2}$ be given, and assume $q^{-\frac{m}{16}} \leq n \leq q^{-\alpha m}$.

Note that $2\kappa < H(\kappa)$ for all $\frac{1}{16} \leq \kappa \leq \alpha$. Since H is continuous on the compactum $[\frac{1}{16}, \alpha]$, we may choose $\varphi > 0$ such that for all $\frac{1}{16} \leq \kappa \leq \alpha$, $2\kappa < (1-\varphi)H(\kappa)$. Moreover, we can put $\gamma := \min\{(1-\varphi)H(\kappa) - 2\kappa : \kappa \in \left[\frac{1}{16}, \alpha\right]\}$ and have $\gamma > 0$.

By Lemma 20, there is an N_1 such that $m + n \geq N_1$ yields

$$\Pr[\mathcal{S}'_G \geq 2^{(1-\varphi)\cdot H(\lambda)\cdot m}] \geq 1 - \tfrac{\varepsilon}{2},$$

where $\lambda = \log_{1/q}(n)/m$.

By Lemma 11, there is an N_2 such that $\Pr[\mathcal{L}_G > 2^{2\log_{1/q}(n)+1}] \leq \varepsilon/2$ when $m+n \geq N_2$. Let $N_3 = \max(N_1, N_2)$ and assume $m+n \geq N_3$. With probability $1-\varepsilon$,

$$\mathcal{S}'_G/\mathcal{L}_G \geq 2^{(1-\varphi)\cdot H(\lambda)\cdot m} \cdot 2^{-2\log_{1/q}(n)-1} \geq 2^{\gamma m - 1}.$$

Thus, there is an $N \geq N_3$ such that $\mathcal{S}'_G/\mathcal{L}_G \geq (1-2\alpha)^{-1}$, whenever $m+n \geq N$. Lemma 6 (with $\delta = 0$ and $\nu = 1 - 2\alpha$) completes the proof. □

4.5 The case $q^{-\frac{m}{2}} \leq n \leq q^{-m^3}$

Once the size n of the right side reaches $q^{-\frac{m}{2}}$, the expected average left side of a maximal stable set becomes very close to $\frac{m}{2}$, so close in fact that the methods developed so far begin to fail: We cannot any longer show that the average is at most $\frac{m}{2}$.

The two main obstacles we face are: Firstly, when n approaches $q^{-m/2}$ the upper bound on $\mathcal{L}_G$, Lemma 11, becomes useless as it reaches 2^m. Secondly, the maximality requirement for maximal stable sets of small left side becomes harder to satisfy. Recall that we focus on left sides of size $\approx \log_{1/q}(n)$ because with this size the maximality requirement eliminates only a constant proportion of the possible small maximal stable sets. However, when n surpasses $q^{-\frac{m}{2}}$, the maximal stable sets with left side $\log_{1/q}(n)$ can no longer be considered as *small*, since $\log_{1/q}(n) > \frac{m}{2}$.

Therefore we lower our goals and aim instead for an average of at most $(\frac{1}{2}+\delta)m$, for any given $\delta > 0$. Then the large maximal stable sets, those with left side $> (\frac{1}{2}+\delta)m$, suddenly make up a significantly smaller proportion of the power set.

The key to that observation lies in the following basic lemma, a version of which can be found in, for instance, van Lint [22, Theorem 1.4.5].

Lemma 22. *For all $\frac{1}{2} < \gamma < 1$ it holds that*

$$\sum_{i=\lceil \gamma m \rceil}^{m} \binom{m}{i} \leq 2^{H(1-\gamma)m}.$$

Moreover, $H(1-\gamma) < 1$.

Lemma 23. *For every $\delta > 0$ and $\varepsilon > 0$ there is an N and an $\alpha < \frac{1}{2}$ so that for $G \in \mathcal{B}(m, n; p)$*

$$\Pr\left[\text{left-avg}(G) \leq (\tfrac{1}{2}+\delta)m\right] \geq 1-\varepsilon,$$

for all m, n with $m+n \geq N$ and $q^{-\alpha m} \leq n \leq q^{-m^3}$.

Proof. For $G \in \mathcal{B}(m, n; p)$, let us denote by $\mathcal{L}^\delta_G$ the number of maximal stable sets S with $|L(G) \cap S| \geq \left(\frac{1}{2}+\delta\right)m$.

Note that

$$\mathcal{L}^\delta_G \leq \sum_{i=\lceil (1/2+\delta)m \rceil}^{m} \binom{m}{i} \leq 2^{H(1/2-\delta)m}, \tag{23}$$

where we used Lemma 22 for the second inequality.

We show now that, with high probability, there are many more maximal stable sets with small left side in a random graph $G \in \mathcal{B}(m, n; p)$, if $q^{-\alpha m} \leq n \leq q^{-m^3}$ and $m+n \geq N$, for an N and an $\alpha < \frac{1}{2}$ that we will determine below. In order to do so, note first that we may assume δ to be small enough so that $\alpha' := \frac{1}{2} - \frac{\delta}{3} \geq \frac{1}{16}$. Next, fix $n'(m) = n' := \lceil q^{-\alpha' m} \rceil$ and delete arbitrary $n - n'$

vertices from $R(G)$. The resulting graph G' may be viewed as a random graph in $\mathcal{B}(m, n'; p)$, and we will see that with probability $\geq 1-\varepsilon$ it contains many maximal stable sets with small left side. More precisely, we will prove that

$$\mathcal{S}'_{G'} \geq \tfrac{3}{2\delta}\mathcal{L}^{\delta}_{G} \tag{24}$$

with probability at least $1-\varepsilon$. Note that the maximal stable sets counted by $\mathcal{S}'_{G'}$ have a left side of size

$$a' = \lfloor \log_{1/q}(\lfloor n'/m^{\log_{1/q}(m)} \rfloor) \rfloor \leq \alpha' m = \left(1 - \tfrac{2}{3}\delta\right)\tfrac{m}{2}.$$

Since every maximal stable set of G' extends to a maximal stable set of G with the same left side, we may then use Lemma 6 with $\nu = \frac{2}{3}\delta$ to conclude that left-avg$(G) \leq \left(\frac{1}{2}+\delta\right) m$.

Let us now see how we need to choose N and α in order to guarantee (24), which is all we need to finish the proof. For α, we could take α' if it were not for the fact that we round up $q^{-\alpha' m}$ to get n' (which turns out to be useful below). So we simply choose α to be somewhat larger than α': Let $\alpha = \frac{1}{2} - \frac{\delta}{4} > \alpha'$ and choose N_1 large enough so that $q^{-\alpha m} \geq n' = \lceil q^{-\alpha' m} \rceil$ for all m, n with $m + n \geq N_1$ and $n \leq q^{-m^3}$. (This is possible as $n \leq q^{-m^3}$ implies that m has to be large as well if $m+n$ is large.) Throughout the rest of the proof we will always assume that m, n are integers with $q^{-\alpha m} \leq n \leq q^{-m^3}$.

Next, as H is monotonously increasing in the interval $\left[0, \frac{1}{2}\right]$, it follows that

$$\varphi := 1 - \frac{H\left(\frac{1}{2} - \frac{\delta}{2}\right)}{H\left(\alpha'\right)} = 1 - \frac{H\left(\frac{1}{2} - \frac{\delta}{2}\right)}{H\left(\frac{1}{2} - \frac{\delta}{3}\right)} < 1.$$

Applying Lemma 20 with ε and φ yields an integer N'. Choose $N_2 \geq N_1$ large enough so that $m + n \geq N_2$ implies $m + n' \geq N'$. (Again, this is possible as m has to be large if $m+n$ is large.) Then, as $\alpha' \geq \frac{1}{16}$, which in turn leads to $n' \geq q^{-\frac{m}{16}}$, we obtain for G' that

$$\Pr[\mathcal{S}'_{G'} \geq 2^{(1-\varphi)\cdot H(\log_{1/q}(n')/m)\cdot m}] \geq 1 - \varepsilon,$$

for all m with $m + n \geq N_2$. Note that for m, n with $m+n \geq N_2$

$$H(\log_{1/q}(n')/m) = H(\log_{1/q}(\lceil q^{-\alpha' m}\rceil)/m) \geq H(\log_{1/q}(q^{-\alpha' m})/m) = H(\alpha')$$

and thus

$$2^{(1-\varphi)\cdot H(\log_{1/q}(n')/m)\cdot m} \geq 2^{(1-\varphi)\cdot H(\alpha')\cdot m} = 2^{H(1/2-\delta/2)\cdot m}.$$

Put $\mu := H(1/2 - \delta/2) - H(1/2 - \delta)$ and note that $\mu > 0$. Inequality (23) gives that with probability $1-\varepsilon$,

$$\mathcal{S}'_{G'}/\mathcal{L}^{\delta}_{G} \geq 2^{(H(1/2-\delta/2) - H(1/2-\delta))\cdot m} = 2^{\mu m}.$$

Finally, choosing $N \geq N_2$ large enough so that $2^{\mu m} \geq \frac{3}{2}\delta$ for all m, n with $m + n \geq N$ ensures (24). Again, this is possible as m grows with $m+n$. □

For the proof technique to work, we need G' to be a large graph. Otherwise, Lemma 20 cannot guarantee a high probability. In particular, m has to grow with $m+n$, which is why we assumed $n \leq q^{-m^3}$.

4.6 The case $q^{-m^3} \leq n$ and proof of Theorem 4

If the right side of the random bipartite graph G is huge in comparision to the left side, that is, if $q^{-m^3} \leq n$, then almost surely $L(G)$ may be inductively matching into the right side. As a consequence, the set of left sides of maximal stable sets is equal to the power set of $L(G)$, and thus left-avg$(G) = \frac{m}{2}$.

Lemma 24. *For every $\varepsilon > 0$ there is an N so that for $G \in \mathcal{B}(m, n; p)$*

$$\Pr\left[\text{left-avg}(G) \leq \tfrac{m}{2}\right] \geq 1 - \varepsilon,$$

for all m, n with $m + n \geq N$ and $q^{-m^3} \leq n$.

Proof. Consider positive integers m, n with $n \geq q^{-m^3}$. We will give an N such that for all such m, n with $m + n \geq N$ there are, with probability $1 - \varepsilon$, exactly 2^m maximal stable sets in G. Then, every subset of $L(G)$ is the left side of a maximal stable set, which implies left-avg$(G) = \frac{m}{2}$.

We proceed in a similar way as in the proof of Lemma 9 and therefore skip some of the details. Let $R_1, \dots, R_{\lfloor n/m \rfloor}$ be pairwise disjoint subsets of $R(G)$, of size m each. For $i = 1, \dots, \lfloor n/m \rfloor$ let M_i be the random indicator variable for an induced matching of size m on $L(G) \cup R_i$. Since $n \geq q^{-m^3}$, $n \geq m$ and so $\Pr[M_i = 1] \geq p^m q^{m^2}$. Thus

$$\Pr\left[\sum_{i=1}^{\lfloor n/m \rfloor} M_i = 0\right] \leq \left(1 - p^m q^{m^2}\right)^{\lfloor n/m \rfloor} \leq e^{-p^m q^{m^2} \lfloor n/m \rfloor},$$

where we use that $1 + x \leq e^x$ for all $x \in \mathbb{R}$. Since $n \geq q^{-m^3}$, the term $p^m q^{m^2} \lfloor n/m \rfloor$ becomes arbitrarily large for large $m + n$. Thus, there is an N so that $\Pr\left[\sum_{i=1}^{\lfloor m/k \rfloor} M_i = 0\right] \leq \varepsilon$ for all m, n with $m + n \geq N$ and $n \geq q^{-m^3}$.

Now assume that there is an induced matching on $L(G) \cup R_i$ of size m for some $1 \leq i \leq \lfloor n/m \rfloor$. Then are 2^m many maximal stable sets of the graph $G[L(G) \cup R_i]$ and each can be extended to a maximal stable set of G without changing its left side. This completes the proof. □

Having exhausted all of the parameter space (m, n), we may finally prove our main theorem.

Proof of Theorem 4. For given $\varepsilon > 0$ and $\delta > 0$ choose N_1 and $\alpha < \frac{1}{2}$ as in Lemma 23. Then let N be at least as large as N_1 and the N in Lemmas 10, 15, 18, 21 (with α as chosen) and 24. Then the theorem follows. □

Acknowledgements

We thank Carola Doerr for inspiring discussions and help with some of the probabilistic arguments. The second author is supported by a post-doc grant of the Fondation Sciences Mathématiques de Paris.

References

[1] I. Balla, B. Bollobas, and T. Eccles, *Union-closed families of sets*, J. Combin. Theory (Series A) **120** (2013), 531–544.

[2] B. Bollobás, *Random graphs* (2nd edition), Cambridge University Press, 2001.

[3] I. Bošnjak and P. Marković, *The 11-element case of Frankl's conjecture*, Europ. J. Combin. **15** (2008), R88.

[4] H. Bruhn, P. Charbit, O. Schaudt, and J.A. Telle, *The graph formulation of the union-closed sets conjecture*, Preprint, 2012.

[5] H. Bruhn and O. Schaudt, *The journey of the union-closed sets conjecture*, in preparation.

[6] G. Czédli, *On averaging Frankl's conjecture for large union-closed sets*, J. Combin. Theory (Series A) **116** (2009), 724–729.

[7] G. Czédli, M. Maróti, and E.T. Schmidt, *On the scope of averaging for Frankl's conjecture*, Order **26** (2009), 31–48.

[8] G. Czédli and E.T. Schmidt, *Frankl's conjecture for large semimodular and planar semimodular lattices*, Acta Univ. Palacki. Olomuc., Fac. rer. nat., Mathematica **47** (2008), 47–53.

[9] R. Diestel, *Graph theory* (4th edition), Springer-Verlag, 2010.

[10] G. Lo Faro, *Union-closed sets conjecture: Improved bounds*, J. Combin. Math. Combin. Comput. **16** (1994), 97–102.

[11] P. Frankl, *Handbook of combinatorics (vol. 2)*, MIT Press, Cambridge, MA, USA, 1995, pp. 1293–1329.

[12] W. Hoeffding, *Probability inequalities for sums of bounded variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30.

[13] M. Mitzenmacher and E. Upfal, *Probability and computing: Randomized algorithms and probabilistic analysis*, Cambridge University Press, 2005.

[14] R. Morris, *FC-families, and improved bounds for Frankl's conjecture*, Europ. J. Combin. **27** (2006), 269–282.

[15] T. Nishimura and S. Takahashi, *Around Frankl conjecture*, Sci. Rep. Yokohama Nat. Univ. Sect. I Math. Phys. Chem. **43** (1996), 15–23.

[16] B. Poonen, *Union-closed families*, J. Combin. Theory (Series A) **59** (1992), 253–268.

[17] D. Reimer, *An average set size theorem*, Comb., Probab. Comput. (2003), 89–93.

[18] J. Reinhold, *Frankl's conjecture is true for lower semimodular lattices*, Graphs Comb. **16** (2000), no. 1, 115–116.

[19] I. Rival (ed.), *Graphs and order*, NATO ASI Series, vol. 147, Springer Netherlands, 1985.

[20] I. Roberts and J. Simpson, *A note on the union-closed sets conjecture*, Australas. J. Combin. **47** (2010), 265–267.

[21] D.G. Sarvate and J.-C. Renaud, *Improved bounds for the union-closed sets conjecture*, Ars Combin. **29** (1990), 181–185.

[22] J.H. van Lint, *Introduction to coding theory* (3rd edition), Springer, 1999.

[23] T.P. Vaughan, *Families implying the Frankl conjecture*, Europ. J. Combin. **23** (2002), 851–860.

Version 18 Feb 2013

Henning Bruhn `<bruhn@math.jussieu.fr>`
Oliver Schaudt `<schaudt@math.jussieu.fr>`

Combinatoire et Optimisation
Université Pierre et Marie Curie
4 place Jussieu
75252 Paris cedex 05
France