Random Union-Closed Families

Paul Balister[*] Béla Bollobás[*†‡]

June 21, 2013

Abstract

We shall show that the union-closed families conjecture holds for a random union-closed family with high probability. This strengthens a recent result of Bruhn and Schaudt.

1 Introduction

A family $\mathcal{A}$ of subsets of a set N is said to be an *up-set* if the conditions $A \in \mathcal{A}$ and $A \subset B \subset N$ imply that $B \in \mathcal{A}$, and is *union-closed* (UC) if the union of any two members of $\mathcal{A}$ is again in $\mathcal{A}$. Also, we call $\mathcal{A}$ *globally large* (GL) if the average size of a member of $\mathcal{A}$ is at least $|N|/2$, and *locally large* (LL) if some element $x \in N$ is in at least half of the sets in $\mathcal{A}$. Clearly, every globally large family is also locally large.

In what follows, we shall consider only *non-trivial* families, i.e., families containing at least one non-empty set, although occasionally we may not explicitly say so. Note that every non-trivial up-set $\mathcal{A}$ is globally large and locally large in the much stronger sense that *every* $x \in N$ is contained in at least half of the sets in $\mathcal{A}$. Indeed, for $\mathcal{A}_- = \{A \in \mathcal{A} : x \notin A\}$ and $\mathcal{A}_+ = \{A \in \mathcal{A} : x \in A\}$, the map $\mathcal{A}_- \to \mathcal{A}_+$ given by $A \mapsto A \cup \{x\}$ is

[*]Department of Mathematical Sciences, University of Memphis, Memphis, TN 38152 USA.

[†]Department of Pure Mathematics and Mathematical Statistics, Wilberforce Road, Cambridge CB3 0WB, UK.

[‡]The work of the authors was supported in part by NSF grant DMS 1301614, and that of the second author by MULTIPLEX grant 317532.

an injection. The *union-closed conjecture* states that every non-trivial UC-family is locally large. This conjecture has been a folklore conjecture since the late 1960s, and was made well-known by Frankl, who rediscovered it in the late 1970s (see [7, 8]). The aim of this note is to show that with high probability a random UC-family is globally large, so the union-closed conjecture holds for 'almost all' UC families.

For this statement to have any meaning, we have to decide how we define a random UC-family. Perhaps the most natural way goes as follows. Select a family $\mathcal{B}$ of random subsets of N, and take for $\mathcal{A}$ the collection $\mathcal{U}(\mathcal{B})$ of sets of the form $\bigcup_{B\in\mathcal{C}} B$, $\mathcal{C} \subset \mathcal{B}$; we call $\mathcal{A} = \mathcal{U}(\mathcal{B})$ the UC-family with *basis* $\mathcal{B}$, or the UC-family *generated by* $\mathcal{B}$. Having said this, we have to decide how we choose our random basis $\mathcal{B}$. Here is a simple way; we shall discuss other possibilities later. For $0 < p < 1$, let $B \subset N$ be obtained by choosing the elements of N with probability p, independently of each other. (In particular, $|B|$ has binomial distribution with parameters $n = |N|$ and p.) We call B a *p-random subset* of N. Then let $\mathcal{B} = \mathcal{B}(n, m; p)$ be a sequence of m independent p-random subsets of V: $\mathcal{B} = \{B_1, \dots, B_m\}$. Note that the elements of $\mathcal{B}$ need not be distinct. Finally,

$$\mathcal{A}(n, m; p) = \mathcal{U}(\mathcal{B})$$

is our random UC-family.

Recall that there is a trivial way of identifying a sequence $\mathcal{B} = (B_i)_{i=1}^m$ of subsets of N with a bipartite graph $G(\mathcal{B})$ with bipartition $(N, \mathcal{B})$: simply join $x \in N$ to $B_i \in \mathcal{B}$ by an edge if $x \in B_i$. Conversely, a bipartite graph G with bipartition (N, M) is identified with the multi-family $\mathcal{B}(G)$ on N consisting of all the neighbourhoods $\Gamma(y) \subset N$ of the vertices $y \in M$. (Clearly, $\mathcal{B}(G)$ is a family and not a multi-family of subsets of N if and only if no two vertices of M are *twins*, i.e. any two of them have different neighbourhoods.)

With this identification, the random bipartite graph $G\big(\mathcal{B}(n, m; p)\big)$ is precisely the random n by m bipartite graph $G_{N,M,p}$ defined by taking bipartite classes N and M with $|N| = n$, $|M| = m$, and including edges xy, $x \in N$, $y \in M$, independently with probability p. This gives us the following natural definition of a *random UC-family with parameters n, m and p*:

$$\mathcal{A}(n, m; p) = \mathcal{U}\big(\mathcal{B}(G(n, m; p))\big).$$

This identification of $\mathcal{B}$ with $G(\mathcal{B})$ also gives us a simple description of the elements of $\mathcal{A} = \mathcal{U}(\mathcal{B}) = \mathcal{U}(G)$ in terms of the bipartite graph G. Note that a set $A \cup \mathcal{C}$ of vertices of G, with $A \subset N$ and $\mathcal{C} \subset \mathcal{B}$, is an independent set in $G(\mathcal{B})$ if no vertex $x \in A$ is in any of the sets in $\mathcal{C}$, i.e., if A and $\bigcup_{B \in \mathcal{C}} B$ are disjoint sets. Hence, $A \cup \mathcal{C}$ is a *maximal independent set* of vertices if $A = N \setminus \bigcup_{B \in \mathcal{C}} B$ and $\mathcal{C} = \overline{\mathcal{C}} = \{B' \in \mathcal{B} : B' \subset \bigcup_{B \in \mathcal{C}} B\}$. But what are the sets of the form $\bigcup_{B \in \mathcal{C}} B$ $(= \bigcup_{B \in \overline{\mathcal{C}}} B)$? The elements of $\mathcal{A}$. Thus $\mathcal{A}$ consists of the complements of $I \cap N$ in N, where I is a maximal independent set of vertices of G. Equivalently, $\mathcal{A}$ consists of the subsets of N of the form $N \setminus I$, where I is a maximal independent set of vertices of G. This was first observed by Bruhn, Charbit and Telle [2]; see also ElZahar [6].

As we have seen, if G is a bipartite graph with bipartition (N, M) then for every set $A \subset N$ there is at most one maximal independent set of vertices of G intersecting N in A; if there *is* one, we call A *good*, so that the elements of $\mathcal{A}$ are precisely the complements of the good sets in N. Hence $\mathcal{U}(G)$ is globally large if the average size of the good sets is at most $|N|/2$, and it is locally large if some vertex $x \in N$ is in at most half of the good sets.

Our main aim of this paper is to show that if $\max\{n, m\} \to \infty$ and $p \in (0, 1)$ is fixed, then with high probability (whp) $G = G_{N,M,p}$ is such that $\mathcal{U}(G)$ is locally large, i.e., the probability that $\mathcal{U}(G)$ is locally large tends to 1. Indeed, whp G is globally large after the removal of isolated vertices. An approximate version of this was proved by Bruhn and Schaudt [3]. Earlier, the average size of a set in a union-closed family had been studied by Reimer [9], Czédli [4], Czédli, Maróti and Schmidt [5], and Balla, Bollobás and Eccles [1].

Having pointed out the connection between random union-closed families and random bipartite graphs, from now on we shall work exclusively with random bipartite graphs $G_{N,M,p}$ and prove our result in the following form.

Theorem 1. *For fixed* $p \in (0, 1)$ *and* $\max\{|N|, |M|\} \to \infty$*, whp there exists a vertex in* N *which lies in at most half of all maximal independent subsets of* $G_{N,M,p}$*. Indeed, after removing isolated vertices from* N*, the average of* $|I \cap N|$ *over all maximal independent sets* I *is at most* $|N|/2$*.*

2 Proofs

We start by restating the following observations made in the introduction.

Lemma 2. *If* $A \subset N$ *then there is at most one maximal independent set* I *of* G *with* $I \cap N = A$*. Moreover, such a maximal independent set exists iff* *there is no* $x \in N \setminus A$ *with* $\Gamma(x) \subset \Gamma(A)$.

Proof. If $I \cap N = A$ and I is an independent set then $\Gamma(A) \cap I = \emptyset$. On the other hand, by maximality of I, any element of M that is not in $\Gamma(A)$ must lie in I as we can safely add such points to I while keeping the set I independent. Hence $I \cap M = M \setminus \Gamma(A)$ and $I = A \cup (M \setminus \Gamma(A))$ is uniquely determined by A. This set is a maximal independent set precisely when no $x \in N \setminus A$ can be added to I retaining independence, i.e., when there is no $x \in N \setminus A$ with $\Gamma(x) \subset \Gamma(A)$. $\square$

We call a subset $A \subset N$ *good* if it is the N-part of a maximal independent set, i.e. there is a (unique) maximal independent set I (in $G_{N,M,p}$) such that $A = I \cap N$ and *bad* otherwise.

Lemma 3. *Fix a set* $A \subset N$ *of size* a. *Then the probability that* A *is good is*

$$\sum_{s=1}^{m} \binom{m}{s} q^{as}(1-q^a)^{m-s}(1-q^s)^{n-a},$$

where $q = 1 - p$. *Also, the probability that* A *is bad is bounded above by* $n(1-pq^a)^m$.

Proof. Fix a set $S \subset M$ of size s. The probability that $\Gamma(A) = M \setminus S$ is exactly $(q^a)^s(1-q^a)^{m-s}$. Conditioning on the edges from A, the probability that the neighbourhood of every $x \in N \setminus A$ meets S is $(1-q^s)^{n-a}$. Multiplying these probabilities and summing over all choices of S gives the expression for the probability that A is good.

We use a different approach to bound the probability that A is bad. Fix A and $x \in N \setminus A$. The event that $\Gamma(x) \subset \Gamma(A)$ is just the intersection over $y \in M$ of the event that it is not the case that $y \in \Gamma(x)$ and $y \notin \Gamma(A)$. For each y the probability of this event is just $1 - pq^a$. As these events are independent for different y, the probability that $\Gamma(x) \subset \Gamma(A)$ is $(1-pq^a)^m$. As there are $n - a \leq n$ choices for x, the union bound gives that the probability that A is bad is at most $n(1-pq^a)^m$. (Note that the events that $\Gamma(x) \subset \Gamma(A)$ are not independent for different values of x as there is a strong dependence via the size of $\Gamma(A)$.) $\square$

As noted in the introduction, to show that there is a vertex of N that is in at most half of all maximal independent sets, it is clearly enough to show that the average size of a good set is at most $n/2$, or equivalently, the average size of a bad set is least $n/2$.

Define for $0 \le t \le n/2$,

$$g_t = \sum_{A \text{ good}, |A| \ge n-t} (2|A| - n),$$
$$b_t = \sum_{A \text{ bad}, |A| \le t} (n - 2|A|),$$
$$c_t = \sum_{|A| \ge n-t} (2|A| - n) = \sum_{|A| \le t} (n - 2|A|) = \sum_{i=0}^{t} (n - 2i)\binom{n}{i}$$

Note that $g_t, b_t \le c_t$.

Lemma 4.
(a) $c_t = (n-t)\binom{n}{t}$.
(b) $c_{n/2 - o(\sqrt{n})} \sim c_{n/2} \sim \sqrt{n/2\pi}\, 2^n$.
(c) c_t *is log-concave for* $0 \le t \le n/2$.

Recall that a positive sequence a_t is *log-concave* if $\log a_t$ is a concave function of t, or equivalently, $a_t^2 \ge a_{t-1}a_{t+1}$ for all t.

Proof. For (a),

$$c_t = \sum_{i=0}^{t} (n-i)\binom{n}{i} - \sum_{i=0}^{t} i\binom{n}{i} = \sum_{i=0}^{t} \frac{n!}{(n-i-1)!i!} - \sum_{i=1}^{t} \frac{n!}{(n-i)!(i-1)!}$$
$$= \sum_{i=0}^{t} \frac{n!}{(n-i-1)!i!} - \sum_{i=0}^{t-1} \frac{n!}{(n-i-1)!i!} = \frac{n!}{(n-t-1)!t!} = (n-t)\binom{n}{t}.$$

For (b), note that as $\binom{n}{\lfloor n/2 \rfloor} \sim \sqrt{2/\pi n}\, 2^n$, $c_{n/2} = \lceil n/2 \rceil \binom{n}{\lfloor n/2 \rfloor} \sim \sqrt{n/2\pi}\, 2^n$. Finally, $c_{n/2} \ge c_{n/2-t} \ge c_{n/2} - O(t^2)\binom{n}{\lfloor n/2 \rfloor} = (1 - o(1))c_{n/2}$ when $t = o(\sqrt{n})$. For (c), note first that if a_t is a log-concave sequence then so is $s_t = \sum_{i=0}^{t} a_i$. Indeed, $s_t^2 - s_{t+1}s_{t-1} = s_t^2 - (s_t - a_t)(s_t + a_{t+1}) = s_t(a_t - a_{t+1}) + a_t a_{t+1}$. This is non-negative when $a_t \ge a_{t+1}$, so suppose $a_t = \gamma a_{t+1}$ with $\gamma < 1$. Then $a_{t-i} \le \gamma^i a_t$, and so $s_t \le a_t/(1-\gamma)$. Then $s_t(a_t - a_{t+1}) + a_t a_{t+1} \ge a_t^2(1 - 1/\gamma)/(1-\gamma) - a_t^2/\gamma = 0$. The result follows as $a_t = (n-2t)\binom{n}{t}$ is log-concave:

$$\begin{aligned}a_{t+1}a_{t-1} &= (n-2t+2)(n-2t-2)\binom{n}{t-1}\binom{n}{t+1}\\ &= ((n-2t)^2-4)\binom{n}{t}^2\frac{t}{n-t+1}\,\frac{n-t}{t+1}\\ &< (n-2t)^2\binom{n}{t}^2 = a_t^2.\end{aligned}$$

□

Lemma 5. *Assume* $0 \le t \le n/2$. *If* $b_{n/2} + g_t \le c_t$ *or if* $b_t + g_{n/2} \le c_t$, *then the average size of a good set is at most* $n/2$.

Proof. In the first case

$$\sum_{A \text{ bad}} (2|A|-n) = (c_t - g_t) - b_{n/2} + \sum_{\substack{A \text{ bad}\\ n/2<|A|<n-t}} (2|A|-n) \ge c_t - g_t - b_{n/2} \ge 0,$$

so the average size of a bad set is at least $n/2$. In the second case

$$\sum_{A \text{ good}} (2|A|-n) = g_{n/2} + (b_t - c_t) + \sum_{\substack{A \text{ good}\\ t<|A|<n/2}} (2|A|-n) \le g_{n/2} + b_t - c_t \le 0,$$

so the average size of a good set is at most $n/2$. □

Lemma 6. *Suppose* t *satisfies* $0 \le t \le n/2$ *and* $mpq^t \ge 2\log n$. *Then*

$$\mathbb{E}(b_t) \le c_t/n.$$

Proof. If A is a set of size $a \le t$ then, by Lemma 3, the probability that A is bad is at most $n(1-pq^a)^m \le ne^{-mpq^a} \le 1/n$ as $mpq^a \ge mpq^t \ge 2\log n$.The result follows by linearity of expectation. □

We can now easily deal with the case when m is very large compared to n.

Lemma 7. *If* $\max\{m,n\} \to \infty$ *in such a way that* $mpq^{n/2} \ge n$, *then whp* $b_{n/2} = 0$. *In particular, the average size of a good set is at most* $n/2$ *whp.*

Proof. The expected number of bad sets A of size at most $n/2$ is at most 2^n times $n(1-pq^{n/2})^m$ by Lemma 3. Now

$$2^n n(1-pq^{n/2})^m \le \exp\{n\log 2 + \log n - mpq^{n/2}\},$$

which tends to 0 under the conditions of the lemma. (If $n \to \infty$ then this is clear as $mpq^{n/2} \ge n$. For bounded n we must have $m \to \infty$, for which it is also clear.) Hence whp all sets $A \subset N$ of size at most $n/2$ are good, i.e., $b_{n/2} = 0$. The last statement follows from Lemma 5 as $b_{n/2} + g_t = g_t \le c_t$. □

Lemma 8. *Assume k is an integer such that $m \le q^{-k}$ and t is such that $0 \le t \le n/2$. Then*

$$\mathbb{E}(g_t) \le en2^k + ec_t/\binom{n-k}{t}.$$

Proof. By Lemma 3,

$$\begin{aligned}\mathbb{E}(g_t) &= \sum_{a \ge n-t} (2a-n)\binom{n}{a} \sum_{s=1}^{m} \binom{m}{s} q^{as}(1-q^a)^{m-s}(1-q^s)^{n-a} \\ &\le \sum_{s=1}^{m} \binom{m}{s} \sum_{a\ge n-t} (2a-n)\binom{n}{a} q^{as}(1-q^s)^{n-a}.\end{aligned}$$

If $q^s \ge 1/2$ then

$$\sum_{a\ge n-t} (2a-n)\binom{n}{a} q^{as}(1-q^s)^{n-a} \le n \sum_{a=0}^{n} \binom{n}{a} q^{as}(1-q^s)^{n-a} = n.$$

Now $\binom{m}{s} n \le nm^s/s! \le nq^{-sk}/s! \le n2^k/s!$. Summing over s gives at most $en2^k$. If $q^s < 1/2$ then

$$\sum_{a\ge n-t} (2a-n)\binom{n}{a} q^{as}(1-q^s)^{n-a} \le c_t q^{(n-t)s}(1-q^s)^t.$$

Now

$$\binom{m}{s} c_t q^{(n-t)s}(1-q^s)^t \le c_t s!^{-1}(q^s)^{n-t-k}(1-q^s)^t.$$

But $\binom{n-k}{t}(q^s)^{n-t-k}(1-q^s)^t \le (q^s + (1-q^s))^{n-k} = 1$, so

$$\binom{m}{s} c_t q^{(n-t)s}(1-q^s)^t \le c_t s!^{-1}/\binom{n-k}{t}.$$

Summing over s gives the second term. □

Lemma 9. *Suppose $mpq \ge 2\log n$, $mpq^{n/2} < n$, and $p \in (0,1)$ is fixed. Then the average size of a good set is at most $n/2$ whp as $n \to \infty$ (uniformly in m).*

Proof. Let t be maximal such that $mpq^t \ge 2\log n$. Then $1 \le t \le n/2 + O(\log n)$. Indeed, $mpq^{n/2+C\log n} < nq^{C\log n} \le 1 < 2\log n$ for a sufficiently large constant C depending on q. Let k be minimal such that $m \le q^{-k}$. Then $0 \le k - t \le \log_{1/q}((2\log n)/p) + 2 = o(\log n)$. Also, by Lemma 8, $\mathbb{E}(g_s) \le en2^k + ec_s/\binom{n-k}{s}$ for $0 \le s \le n/2$.

Assume first that $k \geq n/2$. Then $k = n/2 + O(\log n)$. Let $s = n - k - 1$. Then

$$\mathbb{E}(g_{n/2}) \leq (c_{n/2} - c_s) + \mathbb{E}(g_s) \leq (c_{n/2} - c_s) + en2^k + ec_s/(n-k).$$

But $s = n/2 - O(\log n)$, so $(c_{n/2} - c_s) = o(c_{n/2})$ by Lemma 4. Also $k \leq (1-\varepsilon)n$, so $en2^k = o(c_{n/2})$ and as $n \to \infty$, $ec_s/(n-k) = o(c_{n/2})$. Write $t' = \min\{t, \lfloor n/2 \rfloor\}$. Then $t' = n/2 - O(\log n)$, so $c_{t'} \sim c_{n/2}$. Thus $\mathbb{E}(g_{n/2}) = o(c_{t'})$. However, $\mathbb{E}(b_{t'}) = o(c_{t'})$ by Lemma 6, so by Markov's inequality $b_{t'} + g_{n/2} \leq c_{t'}$ whp. The result now follows from Lemma 5. Note that the bounds in the $o()$-notation depend only on n, so the probability that the average size of a good set is less than $n/2$ tends to zero uniformly in m.

Now assume $k < n/2$. Then

$$\mathbb{E}(g_{n/2}) \leq en2^k + ec_{n/2}/\tbinom{n-k}{\lfloor n/2 \rfloor}.$$

The difference $k - t$ is (up to a constant) a function of n. Fixing $k - t = \alpha$ (which is essentially the same as fixing n) and letting m vary, we note that both c_t and $\binom{n-k}{\lfloor n/2 \rfloor}$ are log-concave as functions of t. Thus if $ec_{n/2}/\binom{n-k}{\lfloor n/2 \rfloor} \leq \varepsilon c_t$ for $t = t_0, t_1$ then this holds throughout the range $t_0 \leq t \leq t_1$. Similarly, if $en2^k \leq \varepsilon c_t$ for $t = t_0, t_1$, then this holds throughout the range $t_0 \leq t \leq t_1$. Thus if $\mathbb{E}(g_{n/2}) \leq \varepsilon c_t$ for $t = t_0, t_1$ then $\mathbb{E}(g_{n/2}) \leq 2\varepsilon c_t$ for all $t \in [t_0, t_1]$.

Fix any sufficiently small $\varepsilon > 0$. Take $t = t_1 = \lceil n/2 \rceil - 1 - \alpha$ so that $k = \lceil n/2 \rceil - 1$. Then $\mathbb{E}(g_{n/2}) \leq en2^{n/2} + ec_{n/2}/(n/2) \leq \varepsilon c_t$ for sufficiently large n. If $t = t_0 = 1$ then $c_t = n(n-1)$ and $k = o(\log n)$, so $ec_{n/2}/\binom{n-k}{\lfloor n/2 \rfloor} = \Theta(n2^n/2^{n-k}) = \Theta(n2^k) = o(n^2)$. Thus for n sufficiently large, $\mathbb{E}(g_{n/2}) \leq \varepsilon c_t$ for $t = t_0 = 1$. Thus $\mathbb{E}(g_{n/2}) = o(c_t)$ for $t \geq 1$ and $k < n/2$ as $n \to \infty$, uniformly in m. Once again $\mathbb{E}(b_t) \leq c_t/n = o(c_t)$ uniformly in m, and so by Markov, $g_{n/2} + b_t \leq c_t$ whp as $n \to \infty$, uniformly in m. The result now follows from Lemma 5. □

To deal with the case when mpq is only slightly less than $2\log n$, we refine the argument above to more carefully bound the contribution from small good sets.

Lemma 10. *Suppose $mpq < 2\log n$ and $m = \omega(\log\log n)$. Then the average size of a good set is at most $n/2$ whp as $n \to \infty$.*

Proof. Let

$$\tilde{g} = \sum_{A \text{ good}, |A| < n/2} (n - 2|A|).$$

Then it is enough to show that $\tilde{g} \geq g_{n/2}$. Now $m \leq q^{-k}$ where $k = O(\log \log n)$, so $\mathbb{E}(g_{n/2}) = O(n2^k) \leq n(\log n)^C$ for some $C > 0$. For each $x \in N$, consider the set $[x] = \{y \in N : \Gamma(y) \subset \Gamma(x)\}$. Clearly $[x]$ is a good set. Hence

$$\tilde{g} \geq \tfrac{n}{2}\big|\{[x] : |[x]| < n/4\}\big|.$$

Now $[x] = [y]$ iff $\Gamma(x) = \Gamma(y)$, which occurs with probability at most $\max\{p, q\}^m = \gamma^m$, uniformly in the size of $\Gamma(x)$. Let $r = \min\{n, \gamma^{-m}/2\}$ and choose r elements $x_1, \dots, x_r$ from N. For fixed i, the probability that $[x_i]$ is equal to some other $[x_j]$ is at most $r\gamma^m \leq 1/2$. If $|M \setminus \Gamma(x_i)| = s$ then the size of $[x_i]$ is given as $1 + \mathrm{Bin}(n-1, q^s)$, where $\mathrm{Bin}(n-1, q^s)$ is a binomial random variable with parameters $n-1$ and q^s. Thus as $n \to \infty$, the Chernoff bound implies $|[x_i]| < n/4$ whp when $q^s < 1/8$, say. Another application of Chernoff shows that as $m \to \infty$, $s > qm/2$ whp, and in particular $q^s < 1/8$ whp. Hence $\mathbb{P}(|[x_i]| < n/4) \to 1$. Thus $[x_i]$ contributes to $\tilde{g}$ with probability at least 1/3, say. Hence $\mathbb{E}(\tilde{g}) \geq rn/6$. If $r/6 \gg (\log n)^C$, then by Markov, $\tilde{g} \geq g_{n/2}$ whp. Clearly this holds if $r = n \to \infty$. If $r = \gamma^{-m}/2$ then this holds when $m = \omega(\log \log n)$. □

We are now left to deal with the case when m is very small, including the case when m is constant. In this case it is not enough to just consider the average sizes of the good sets. Indeed, there may be many isolated vertices in N which will be included in all good sets, making these sets larger.

To determine whether a point is in more or less than a half of all independent sets, it is enough, and will be more convenient, to consider both a set $A \subset N$ and its complement $N \setminus A$ and consider only the cases where one is good and the other is bad.

Lemma 11. *Suppose $A \subset N$ is bad, but $N \setminus A$ is good. Then there exists a $C \subset M$ such that $A = \Gamma(C)$. Conversely, if $A = \Gamma(C) \neq N$ and $\Gamma(A) = M$ then A is bad and $N \setminus A$ is good.*

Proof. Assume A is bad and $N \setminus A$ is good. Let $C = M \setminus \Gamma(N \setminus A)$. The fact that $N \setminus A$ is good implies that for each $x \in A$, $\Gamma(x) \not\subset \Gamma(N \setminus A)$. Hence there must be a vertex in C adjacent to x. Hence $\Gamma(C) \supseteq A$. But by definition $\Gamma(C) \cap (N \setminus A) = \emptyset$, so in fact $\Gamma(C) = A$.

For the converse, assume $A = \Gamma(C) \neq N$ and in addition assume $\Gamma(A) = M$. Pick $x \in N \setminus A$. Clearly $\Gamma(x) \subset M = \Gamma(A)$, so A is bad. On the other hand, if $x \in A$, then $\emptyset \neq \Gamma(x) \subset C$. But $\Gamma(N \setminus A) \cap C = \emptyset$, so $\Gamma(x) \not\subset \Gamma(N \setminus A)$. Hence $N \setminus A$ is good. □

Corollary 12. *If* $G = G_{N,M,p}$ *is such that every two vertices in* M *has a common neighbour in* N*, and if the average size of a set* $A \subset N$ *that can be written in the form* $A = \Gamma(C)$ *is at least* $|\Gamma(M)|/2$*, then the conclusion of Theorem 1 holds.*

Proof. Every maximal independent set contains all isolated vertices, so it is enough to prove the result for the subgraph of G obtained by removing all isolated vertices from N. Hence we may assume $\Gamma(M) = N$. If every pair of vertices in M have a common neighbour, then $\Gamma(\Gamma(C)) = M$ for all $\emptyset \neq C \subset M$. Thus by Lemma 11, the sets A with A bad and $N \setminus A$ good are precisely the sets of the form $A = \Gamma(C)$, $A \neq \emptyset, N$. If the average size of these sets is at most $n/2$, then the same must be true for the set of all good sets. (Note that $\emptyset = \Gamma(\emptyset)$ and $N = \Gamma(M)$ are both good.) $\square$

Corollary 13. *If* $\max\{m, n\} \to \infty$ *in such a way that* $n \min\{pq^{m-1}, p^2/2\} - \log m \to \infty$ *then the conclusion of Theorem 1 holds.*

Proof. Let $x \in M$. We say x has a *private neighbour* in N if there exists a $y \in N$ such that $y \in \Gamma(x)$ but $y \notin \Gamma(x')$ for any $x' \in M$, $x' \neq x$. The probability that x has no private neighbour is $(1 - pq^{m-1})^n \leq e^{-npq^{m-1}}$. The probability that there is some element of M which has no private neighbour is then at most $me^{-npq^{m-1}}$. By assumption, this tends to zero, so we may assume every $x \in M$ has a private neighbour. In this case all sets $\Gamma(C)$, $C \subset M$ are distinct. Also, if $x, y \in M$ then the probability that $\Gamma(x) \cap \Gamma(y) = \emptyset$ is $(1 - p^2)^n$. Thus the probability that there exist $x, y \in M$ with no common neighbour in N is at most $m^2(1 - p^2)^n \leq m^2 e^{-np^2}$. But the assumptions imply this tends to zero, so we may assume every pair of vertices in M have a common neighbour. It is clear that the average size of $\Gamma(C)$, $C \subset M$ is at least $|\Gamma(M)|/2$. Indeed, $|\Gamma(C)| + |\Gamma(M \setminus C)| \geq |\Gamma(M)|$ for all C. Hence Corollary 12 implies that the conclusion of Theorem 1 holds. $\square$

Proof of Theorem 1. The case when $mpq^{n/2} \geq n$ is dealt with by Lemma 7. For $mpq \geq 2 \log n$ but $mpq^{n/2} < n$ we can apply Lemma 9. For $mpq < 2 \log n$ but $m = \omega(\log \log n)$ we can apply Lemma 10. Finally, if $m = O(\log \log n)$ then $n \min\{pq^{m-1}, p^2/2\} - \log m \to \infty$ and we can apply Corollary 13. $\square$

3 Further Models

Taking a random bipartite graph is only one of the many natural ways of choosing a random base: let us mention some more. We may take for $\mathcal{B}$ a random r-uniform hypergraph on N with m edges. This is essentially the same as taking $p = r/n$ in the model we have considered.

Another, even more regular model is defined by taking a quadruple (n, m, r, s) with $nr = ms$, and taking for G a random (r, s)-regular bipartite graph with bipartition (N, M). This is again close to $G(n, m; p)$ with $p = s/n$. Although our results are bound to hold for these models, a fair amount of work would be needed to prove them.

A considerably more general model is obtained by taking a random bipartite graph with varying probabilities. Thus, we may take an edge ij with a probability depending on both i and j. In a less extreme case, the probability p_{ij} of an edge ij, $1 \leq i \leq n$, $1 \leq j \leq m$, is just p_i. In an even less general case, which is still much more general than the one we have studied, we take integers $m_i \geq 1$ with $\sum_{i=1}^{k} m_i = m$, and probabilities $0 < p_1, \ldots, p_k < 1$, and define a random m by n bipartite graph with probability p_i of an edge leaving m_i of the vertices in the second class. The union-closed conjecture for such a general base is likely to be very close to the full union-closed conjecture.

References

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

[2] H. Bruhn, P. Charbit and J.A. Telle, The graph formulation of the union-closed sets conjecture, `arXiv.1212.4175`, 11pp.

[3] H. Bruhn, and O. Schaudt, The union-closed sets conjecture almost holds for almost all random bipartite graphs, `arXiv:1302.7141`, 26pp.

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

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

[6] M.H. El-Zahar, A graph-theoretic version of the union-closed sets conjecture, *J. Graph Theory* **26** (1997), 155–163.

[7] P. Frankl, Problem, in *Graphs and Order* (I. Rival, ed.), NATO ASI Series **147**, Springer, 1985.

[8] Frankl, P., Extremal set systems, in *Handbook of Combinatorics, Vol 2* (R.L. Graham, M. Grötschel and L. Lovász, eds), MIT Press, Cambridge, MA, USA, 1995, pp. 1293–1329. (See page 1296.)

[9] D. Reimer, An average set size theorem, *Combin. Probab. Comput.* **12** (2003), 89–93.