Rainbow Thresholds

Alan Frieze
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA 15213

Trent G. Marbach
Department of Mathematics
Ryerson University
Toronto ON M5B 2K3

May 17, 2021

Abstract

We extend the recent breakthrough results relating expectation thresholds and actual thresholds to include rainbow versions.

1 Introduction

It has been observed for a long time that the threshold for the existence of various combinatorial objects in random graphs and hypergraphs occurs close to where the expected number of such objects tends to infinity. This informal observation has been given rigorous validation in two recent breakthrough papers. First of all, Frankston, Kahn, Narayanan and Park [11] showed that under fairly general circumstances, the threshold for the existence of combinatorial objects is within a factor $O(\log n)$ of the point where the expected number begins to take off. In a follow up paper, Kahn, Narayanan and Park [10] tightened their analysis for the case of the square of a Hamilton cycle and solved the existence problem up to a constant factor. A remarkable achievement, given the complexity of proofs of earlier weaker results.

There has been considerable research on random graphs where the edges have been randomly colored. Most notably several authors have considered the existence of rainbow colored combinatorial objects. A set of colored edges will be called rainbow if each edge has a different color. Improving on earlier results of Cooper and Frieze [2] and Frieze and Loh [7], Ferber and Krivelevich [4] showed that w.h.p. at the threshold for Hamiltonicity, randomly coloring the edges of $G_{n,p}$ with $n + o(n)$ colors yields a rainbow Hamilton cycle. Our aim in this short paper is to show that the proofs in [10] and [11] can easily be modified to incorporate rainbow questions. We begin by summarising the results of the papers [10] and [11].

A hypergraph $\mathcal{H}$ (thought of as a set of edges) is $r$-bounded if $e \in \mathcal{H}$ implies that $|e| \leq r$. The most important notion comes next. For a set $S \subseteq X = V(\mathcal{H})$ we let $\langle S \rangle = \{ T : S \subseteq T \subseteq X \}$ denote the subsets of $X$ that

*Research supported in part by NSF grant DMS1952285
contain \( S \). Let \( \langle \mathcal{H} \rangle = \bigcup_{S \in \mathcal{H}} \langle S \rangle \) be the collection of subsets of \( X \) that contain an edge of \( \mathcal{H} \). We say that \( \mathcal{H} \) is \( \kappa \)-spread if we have the following bound on the number of edges of \( \mathcal{H} \) that contain a particular set \( S \):

\[
|\mathcal{H} \cap \langle S \rangle| \leq \frac{|\mathcal{H}|}{\kappa^{|S|}}, \quad \forall S \subseteq X.
\]

Let \( X_m \) denote a random \( m \)-subset of \( X \) and \( X_p \) denote a subset of \( X \) where each \( x \in X \) is included independently in \( X_p \) with probability \( p \). The following theorem is from [11]:

**Theorem 1.** Let \( \mathcal{H} \) be an \( r \)-bounded, \( \kappa \)-spread hypergraph and let \( X = V(\mathcal{H}) \). There is an absolute constant \( K > 0 \) such that if

\[
m \geq \frac{(K \log r)|X|}{\kappa}
\]

then w.h.p. \( X_m \) contains an edge of \( \mathcal{H} \). Here w.h.p. assumes that \( r \to \infty \).

**Remark 1.** Let \( p = 1/\kappa \) and \( Z \) denote the number of edges of \( \mathcal{H} \) that are contained in \( X_p \). Then, assuming that \( \mathcal{H} \) is \( r \)-uniform, we have from [1] with \( S \in \mathcal{H} \) that \( |\mathcal{H}| \geq \kappa^r \) and then \( \mathbb{E}(Z) = |\mathcal{H}|p^r \geq 1 \). This gives the connection between spread and the expected value of \( Z_p \). Theorem [1] inflates \( p \) by a factor of order \( \log r \).

Suppose now that the elements of \( X \) are randomly colored from a set \( Q = [q] \). A set \( S \subseteq X \) is rainbow colored if no two elements of \( S \) have the same color. We modify the proof of Theorem [1] to prove

**Theorem 2.** Let \( \mathcal{H} \) be an \( r \)-bounded, \( \kappa \)-spread hypergraph and let \( X = V(\mathcal{H}) \) be randomly colored from \( Q = [q] \) where \( q \geq r \). Then there is an absolute constant \( K > 0 \) such that if

\[
m \geq \frac{(K \log r)|X|}{\kappa}
\]

then w.h.p. \( X_m \) contains a rainbow colored edge of \( \mathcal{H} \). Here w.h.p. assumes that \( r \to \infty \).

(We will assume that \( \kappa \gg \log r \). If \( \kappa = O(\log r) \) then \( m \geq |X| \) for sufficiently large \( K \).)

Taking \( X = \{\binom{[n]}{2}\} \) (the edges of \( K_n \)), this shows for example that if \( q = n \) and \( m = Kn \log n \) then w.h.p. a randomly edge colored copy of \( G_{n,m} \) contains a rainbow Hamilton cycle, see Bal and Frieze [1]. Here \( \mathcal{H} \) is the \( n \)-uniform hypergraph with \( (n - 1)!/2 \) edges, one for each Hamilton cycle of \( K_n \). Dudek, English and Frieze [3] studied rainbow Hamilton cycles in random hypergraphs. Theorem [2] strengthens Theorem 6 of that paper in that [3] requires \( m \geq \omega n \log n \) where \( \omega \to \infty \), whereas \( m = Kn \log n \) follows from Theorem [2]. Similarly, if \( q = n - 1 \) and \( m = Kn \log n \) and \( T \) is an \( n \)-vertex tree with bounded maximum degree then w.h.p. \( G_{n,m} \) contains a rainbow copy of \( T \). The uncolored version is due to Montgomery [12].

The paper [10] focusses exclusively on the square of Hamilton cycles. It removed a \( \log n \) factor to get within a constant factor of a sharp threshold. The proof is similar to that of Theorem [1] and it can be generalised to prove the following theorem. For \( S \in \mathcal{H} \), let

\[
f_{t,S} = |\mathcal{H}|^{-1} \left| \{ J \in \mathcal{H} : |J \cap S| = t \} \right|.
\]

**Theorem 3.** Suppose that \( \mathcal{H} \) is \( \kappa \)-spread and \( r \)-uniform. Assume that \( \kappa \gg 1, |X|/\kappa \gg r \) and that there exist constants \( \alpha, K_0 \) such that for all \( S \in \mathcal{H} \),

\[
f_{t,S} \leq \left( \frac{K_0}{\kappa} \right)^t \text{ for } 1 \leq t \leq \alpha r.
\]

Let \( X = V(\mathcal{H}) \) be randomly colored from \( Q = [q] \) where \( q \geq r \). Then

\[
\forall \varepsilon > 0, \exists C_\varepsilon \text{ such that } m \geq \frac{C_\varepsilon |X|}{\kappa} \text{ implies that } \mathbb{P}(X_m \text{ contains a rainbow copy of } e \in \mathcal{H}) \geq 1 - \varepsilon.
\]
It would seem likely that $1 - \varepsilon$ can be replaced by $1 - o(1)$ using Friedgut’s method \cite{5}, as was claimed in \cite{10}. We will not pursue this here though.

As examples of the use of Theorem 3 we can take the edges of $\mathcal{H}$ to be the collection of $k$th powers of Hamilton cycles of $K_n^k$, $k \geq 2$. In this case we can take $\kappa = O(n^{1/k})$. The spread condition was verified for $\kappa = 2$ in \cite{10} and the argument there generalises to $k \geq 3$, see an appendix.

## 2 Proof of Theorem 2

We closely follow \cite{11} changing the proof to account for the coloring. The idea is to choose a small random set $W$ and argue that w.h.p. there exists a rainbow edge $H \in \mathcal{H}$ such that $|H \setminus W|$ is significantly smaller than $|H|$. We then repeat the argument with respect to the hypergraph $\mathcal{H} \setminus W = \{ e \setminus W : e \in \mathcal{H} \}$. In this way, we build up a member of $\mathcal{H}$ piece by piece. After $O(\log r)$ iterations we can prove the existence of a final small piece by using a small modification to Janson’s inequality \cite{9}.

Let $\gamma$ be a moderately small constant (e.g. $\gamma = 0.1$ suffices), and let $C_0$ be a constant large enough to support the estimates that follow. Let $\mathcal{H}$ be an $r$-bounded, $\kappa$-spread hypergraph on a $Q$-colored set $X$ of size $N$, with $r, \kappa \geq C_0^2$. Let $X^* = X \times [q]$. We say that a set $S \subseteq X^*$ is rainbow if $c(x) \neq c(y)$ for $x, y \in S$. Then let

$$\mathcal{H}^* = \{ (x_i, c_i) \in X^*, 1 \leq i \leq \rho : \{ x_1, x_2, \ldots, x_\rho \} \in \mathcal{H}, c_i \neq c_j \text{ for } i \neq j \},$$

be the set of rainbow edges of $\mathcal{H}$. (Here $\rho \leq r$ always.)

For $x^* = (x, c) \in X^*$ we define $\xi(x^*) = x, c(x^*) = c$ and for $S \subseteq X^*$, we let $\xi(S) = \{ \xi(x), x \in S \}$. If $S \subseteq X^*$ is rainbow and $|S| = s$, then we have

$$|\mathcal{H}^* \cap \langle S \rangle| = (q - s) r_{r-s} |\mathcal{H} \cap \langle \xi(S) \rangle| \leq (q - s)^{r-s} \frac{|\mathcal{H}|}{\kappa^s} = \frac{(q - s)^{r-s} |\mathcal{H}^*|}{(q)^r} \leq \frac{e^s |\mathcal{H}^*|}{(q^k)^s}. \quad (7)$$

If $S$ is not rainbow or $|\xi(S)| \neq |S|$ then $\mathcal{H}^* \cap \langle S \rangle = \emptyset$ and so $\mathcal{H}^*$ is $q\kappa/e$ spread. It follows from Theorem 1 that w.h.p. $X^*_m$ contains an edge of $\mathcal{H}^*$ w.h.p. for $m \geq \left[ \frac{(K^{log r})^{X^*}}{q^k} \right] = \frac{K^{N log r}}{\kappa}$. The only fly in this ointment is that a random $m$-subset of $X^*$ may not correspond to a randomly colored subset of $X$. Let $\mathcal{B}$ denote the subsets of $X^*_m$ that contain a pair $(x, c_1), (x, c_2)$ i.e. where $x \in X$ has been given two colors. The expected number of such pairs in a random subset $X^*_m$ is $\approx \frac{N q^2}{2} \cdot \left( \frac{m}{q^k N} \right)^2 = O \left( \frac{N^{log^2 r}}{\kappa^2} \right)$. So we see immediately that Theorem 1 holds if $\frac{q^k}{N^{log^2 r}} \rightarrow \infty$, otherwise we have to work round the problem.

Let

$$\hat{E} = \{ W \subset X^* : W \notin \mathcal{B} \}.$$

Set $p = C/\kappa$ and $m = N p$ with $C_0 \leq C \leq C_0 / C_0$ (so $p \leq 1/C_0$), $r' = (1 - \gamma)r$. Let $X^*_m$ be chosen uniformly from $\hat{E}$. (We distinguish this from $X^*_m$, which is randomly chosen from $X \times [q]$.) We remark that choosing $W = X^*_m$ uniformly from $\hat{E}_m$ is the same as choosing $m$ randomly colored random edges i.e. we choose $W_1 = \{ x_1, x_2, \ldots, x_m \} \in (X^*_m)$ and then we choose $(c_1, c_2, \ldots, c_m)$ randomly from $[q]^m$.

Fix $\psi : \langle \mathcal{H}^* \rangle \rightarrow \mathcal{H}^*$ satisfying $\psi(Z) \subseteq Z$ for all $Z \in \langle \mathcal{H}^* \rangle$. I.e. given $Z \in \langle \mathcal{H}^* \rangle$, $\psi$ chooses a member of $\mathcal{H}^*$ contained in $Z$. It will be convenient to assume that $\psi$ selects a minimum size member of $\mathcal{H}^*$ that is contained in $Z$. For $W \subseteq X$ and $S \in \mathcal{H}^*$, set

$$\chi(S, W) = \psi(S \cup W) \setminus W; \quad \text{(Note that } |\chi(S, W)| \leq |S| \text{.)}$$

and say that the pair $(S, W)$ is bad if $|\chi(S, W)| > r'$ and good otherwise.
Lemma 4. For \( \mathcal{H} \) as above, and \( W \) chosen uniformly from \( \hat{E}_m = \left\{ W \in \hat{E} : |E| = m \right\} \),

\[ \mathbb{E}(\{ S \in \mathcal{H}^* : (S, W) \text{ is bad} \}) \leq |\mathcal{H}^*|C^{-r/3}. \]

**Proof.** Let \( W = \{(x_i, c_i) : i = 1, 2, \ldots, m\} \) be distributed as \( \hat{X}_m^* \). We let \( \mathcal{H}_s^* = \{ S \in \mathcal{H}^* : |S| = s \} \). It is enough to show, for \( s \in (r', r] \),

\[ \mathbb{E}(\{ S \in \mathcal{H}_s^* : (S, W) \text{ is bad} \}) \leq (\gamma r)^{-1}|\mathcal{H}^*|C^{-r/3}. \] (8)

Now

\[ \mathbb{E}(\{ S \in \mathcal{H}_s^* : (S, W) \text{ is bad} \}) = \sum_{S \in \mathcal{H}_s^*} \mathbb{P}( (S, W) \text{ is bad} ) \]

\[ = \frac{1}{N_{N_p}^2 q^{N_p}} \left| \left\{ S \in \mathcal{H}_s^*, W_1 \in \left( X_{N_p}, c \in [q^{N_p}] \right) : (S, (W_1, c)) \text{ is bad} \right\} \right|. \]

So, we instead concentrate on showing,

\[ \left| \left\{ S \in \mathcal{H}_s^*, W_1 \in \left( X_{N_p}, c \in [q^{N_p}] \right) : (S, (W_1, c)) \text{ is bad} \right\} \right| \leq (\gamma r)^{-1}\left( \frac{N_{N_p}}{N_p} \right) q^{N_p}|\mathcal{H}^*|C^{-r/3}. \] (9)

(Note that \( \gamma r = r - r' \) bounds the number of \( s \) for which the set in question can be nonempty, whence the factor \( (\gamma r)^{-1} \).)

For \( Z \supseteq S \in \mathcal{H}_s^* \), we say that \( (S, Z) \) is **pathological** if there is \( T \subseteq S \) with \( t := |T| > r' \) and

\[ |\{ S' \in \mathcal{H}_s^* : S' \in [T, Z] \}| > C^{r/2}|\mathcal{H}^*|(q\kappa)^{-t} \left( \frac{p}{q} \right)^{s-t}. \] (10)

From now on we will always take \( Z = W \cup S \) (with \( W = (W_1, c) \) as described at the beginning of the proof of this lemma). Note that in proving (9) we may assume \( s \leq N/2 \). This is justified as follows: we may assume that \( |\mathcal{H}_s^*| \) is at least the R.H.S. of (8). Otherwise equation (9) holds trivially. And then we have

\[ 1 \leq \left( \frac{e}{q \kappa} \right)^s |\mathcal{H}^*| \leq \left( \frac{e}{q \kappa} \right)^s \gamma r C^{r/3}|\mathcal{H}_s^*| \leq \left( \frac{e}{q \kappa} \right)^s \gamma r C^{r/3} 2^N q^s < 1, \]

if \( s > N/2 \) (since \( \kappa > C \)), contradiction.

We bound the non-pathological and pathological parts of (9) separately.

**Non-pathological contributions.** We first bound the number of pairs \( (S, W) \) in the left hand side of (9) with \( (S, Z) \) non-pathological, \( Z = S \cup W \).

**Step 1:** There are at most

\[ \sum_{i=0}^{s} \binom{N}{N_p+i} q^{N_p+i} \leq \binom{N+s}{N_p+s} q^{N_p+s} \leq \binom{N}{N_p} \frac{q^{N_p+s}}{p^s} \] (11)

choices for \( Z = W \cup S \).
Step 2: Given \( Z \), let \( S' = \psi(Z) \). Choose \( T := S \cap S' \), for which there are at most \( 2^{|S'|} \leq 2^r \) possibilities, and set \( t = |T| > r' \). (If \( t \leq r' \) then \((S, W)\) cannot be bad, as \( \chi(S, W) = S' \setminus W \subseteq T \).)

Step 3: Since we are only interested in non-pathological choices, and we choose \( S \setminus S' \) by choosing \( S \in [T, Z] \), the number of possibilities for \( S \) is now at most

\[
C^{r/2}|\mathcal{H}^*|(q\kappa)^{-t} \left( \frac{p}{q} \right)^{s-t}.
\]

Step 4: Complete the specification of \((S, W)\) by choosing \( W \cap S \), the number of possibilities for which is at most \( 2^s \).

Because \( s \leq r \) and \( t > r' = (1 - \gamma)r \), the number of nonpathological possibilities is at most

\[
\left( \frac{N}{N_p} \right) q^{Np+s} \cdot 2^s \cdot C^{r/2}|\mathcal{H}^*|(q\kappa)^{-t} \left( \frac{p}{q} \right)^{s-t} \cdot 2^s = 2^{r+s} \left( \frac{N}{N_p} \right) q^{Np}|\mathcal{H}^*|C^{r/2}(p\kappa)^{-t}
\]

\[
\leq \left( \frac{N}{N_p} \right) q^{Np}|\mathcal{H}^*|(4C^{1/2})^r C^{-t} < \left( \frac{N}{N_p} \right) q^{Np}|\mathcal{H}^*| [4C^{-(1/2 - \gamma)}]^r. \tag{12}
\]

Pathological contributions. We next bound the number of \((S, W)\) in the left hand side of (9) with \((S, Z)\) pathological, \( Z = S \cup W \).

Step 1: There are at most \(|\mathcal{H}^*|\) possibilities for \( S \).

Step 2: Choose \( T \subseteq S \) witnessing the pathology of \((S, Z)\) (i.e. for which (10) holds); there are at most \( 2^s \) possibilities for \( T \).

Step 3: Choose \( U \in [T, S] \) for which

\[
|\mathcal{H}^*_s \cap [U, (Z \setminus S) \cup U]| > 2^{-(s-t)}C^{r/2}|\mathcal{H}^*|(q\kappa)^{-t} \left( \frac{p}{q} \right)^{s-t}. \tag{13}
\]

(Here the left hand side counts the members of \( \mathcal{H}^*_s \) in \( Z \) whose intersection with \( S \) is precisely \( U \). Of course, the existence of \( U \) as in (13) follows from (10).) The number of possibilities for this choice is at most \( 2^{s-t} \).

Step 4: Choose \( Y = Z \setminus S \). Write \( \Phi \) for the R.H.S. of (13). Noting that \( Z \setminus S \) must belong to the set \( \bigcup_{i=0}^s \left( \binom{N}{N_p-i} \times [q]^{Np-i} \right) \), we consider, for \( Y \) drawn uniformly from this set,

\[
P(|\mathcal{H}^*_s \cap [U, Y \cup U]| > \Phi). \tag{14}
\]

We can then bound the number of choices for \( Z \setminus S \) by \( \left( \frac{N}{N_p} \right) q^{Np} \) times this probability. Set \( |U| = u \). We have

\[
|\mathcal{H}^*_s \cap (U)| \leq |\mathcal{H}^* \cap (U)| \leq |\mathcal{H}^*| \left( \frac{e}{q\kappa} \right)^u,
\]

while, for any \( S' \in \mathcal{H}^*_s \cap (U) \),

\[
P(Y \supseteq S' \setminus U) \leq \left( \frac{Np}{q(N - s)} \right)^{s-u}
\]
Lemma 5. For an $G$ where

$$\theta := \mathbb{E}(|\mathcal{H}_s^* \cap [U, Y \cup U]|) \leq |\mathcal{H}_s^*||\left(\frac{e}{q^r}\right)^u\left(\frac{Np}{q(N-s)}\right)^{s-u} \leq |\mathcal{H}_s^*||\left(\frac{e}{q^r}\right)^u\left(\frac{2p}{q}\right)^{s-u}$$

(since $N - s \geq N/2$). Markov’s Inequality then bounds the probability in (14) by $\theta/\Phi$, and this bounds the number of possibilities for $Z \setminus S$ by $\left(\frac{N}{N_p}\right)q^{Np}(\theta/\Phi)$.

Step 5: Complete the specification of $(S, W)$ by choosing $S \cap W$, which can be done in at most $2^s$ ways.

Combining we find that the number of pathological possibilities is at most

$$|\mathcal{H}_s^*| \cdot 2^s \cdot 2^{s-t} \cdot \left(\frac{N}{N_p}\right)q^{Np} \cdot \frac{|\mathcal{H}_s^*||\left(\frac{e}{q^r}\right)^u\left(\frac{2p}{q}\right)^{s-u}}{2^{-(s-t)}C^{r/2}(|\mathcal{H}_s^*|)(q^r)^{-t}\left(\frac{q}{p}\right)^{s-r}} \cdot 2^s =$$

$$\left(\frac{N}{N_p}\right)q^{Np}|\mathcal{H}_s^*| e^{u2^{5s-2t-u}k^{t-u}p^{t-u}C^{r/2}} = \left(\frac{N}{N_p}\right)q^{Np}|\mathcal{H}_s^*| e^{u2^{5s-2t-u}C^{t-u-r/2}} \leq \left(\frac{N}{N_p}\right)q^{Np}|\mathcal{H}_s^*| C^{-2r/5}. \quad (15)$$

Finally, the sum of the bounds in (12) and (15) is less than the $(\gamma r)^{-1}\left(\frac{N}{N_p}\right)q^{Np}|\mathcal{H}_s^*| C^{-r/3}$ of (9). \hfill \square

2.1 Small uniformities

Small set sizes are handled by a modification of Janson’s inequality \[9\]. For $Y = X \times [q]$ and $\alpha \in (0, 1)$ we define the random subset $Y_\alpha$ as follows: we start with the random set $X_\alpha$ and then each $x \in X_\alpha$ we give $x$ a random color $c(x) \in [q]$. The resulting set $Y_\alpha = \{(x, c(x)), x \in X_\alpha\}$.

Lemma 5. For an $r$-bounded, $\kappa$-spread $\mathcal{G}$ on $Y$, and $\alpha \in (0, 1)$,

$$\mathbb{P}(Y_\alpha \not\subset \langle \mathcal{G}^* \rangle) \leq \exp\left\{-2\sum_{t=1}^{r} \binom{r}{t} \left(\frac{e}{q^r}\right)^t\right\}, \quad (16)$$

where $\mathcal{G}^*$ is the set of rainbow edges of $\mathcal{G}$.

Proof. Denote the members of $\mathcal{G}^*$ by $S_i, i = 1, 2, \ldots, m$ and set $\zeta_i = 1_{\{Y_\alpha \supset S_i\}}$. We add $r - |S_i|$ new elements and colors $(x, c)$ to each $S_i$ to create an $r$-uniform hypergraph $\hat{\mathcal{G}}^*$ on a set of vertices $\hat{Y}$ together with an enlarged set of colors $[\hat{q}]$. Note that $\hat{Y}_\alpha \in \langle \hat{\mathcal{G}}^* \rangle$ implies $Y_\alpha \in \langle \mathcal{G}^* \rangle$. This addition of vertices and colors is for the purposes of the proof only and for the rest of the proof of this lemma, we let $\mathcal{G}^* = \hat{\mathcal{G}}^*$, assuming that $\mathcal{G}$ is $r$-uniform. Then

$$\mu := \sum_{i=1}^{m} \mathbb{E}(\zeta_i) = \left(\frac{\alpha}{q}\right)^r|\mathcal{G}^*| = \frac{\alpha^r(q)^r}{q^r}|\mathcal{G}|.$$
Let
\[
\Delta = \sum_{i,j: S_i \cap S_j \neq \emptyset} E(\zeta_i \zeta_j) = \sum_{G \in \mathcal{G}} \frac{\alpha^r(q)_r}{q^r} \sum_{t=1}^{r} \sum_{T \subseteq G} \frac{\alpha^{r-t}(q-t)_{r-t}}{q^{r-t}} \\
\leq \mu \max_{G \in \mathcal{G}} \left\{ \sum_{T \subseteq G} \frac{\alpha^{r-t}(q-t)_{r-t}}{q^{r-t}} \right\} \leq \mu \sum_{t=1}^{r} \left( \frac{e}{\alpha_K} \right)^t \frac{\alpha^r(q)_r}{q^r} \\
\leq \mu \sum_{t=1}^{r} \left( \frac{r}{t} \right) \left( \frac{e}{\alpha_K} \right)^t = \mu^2 \sum_{t=1}^{r} \left( \frac{r}{t} \right) \left( \frac{e}{\alpha_K} \right)^t.
\]

If applicable, Janson's Inequality would bound the probability in (16) by \(-\mu^2/2\Delta\). We claim however, that a straightforward modification of Janson's Inequality allows us to prove
\[
\mathbb{P}(Y_a \notin \langle \mathcal{G}^* \rangle) \leq \exp \left\{ -\frac{\mu^2}{2\Delta} \right\}.
\] (18)

We give details of our modification in Section 2.3

**Corollary 6.** Let \( \mathcal{G} \) be as in Lemma [5] let \( m = \alpha|Y| \) be an integer with \( \alpha \kappa \geq 2er \), and let \( W \) be distributed \( \hat{X}_m^* \) i.e. choose a random member of \( (X_m) \) and then give each element \( x \in W \) a random color \( c(x) \in [q] \). Then
\[
\mathbb{P}(W \notin \langle \mathcal{G}^* \rangle) \leq 2 \exp \left\{ -\frac{\alpha \kappa}{4er} \right\}.
\]

**Proof.** Note that
\[
\sum_{t=1}^{r} \left( \frac{r}{t} \right) \left( \frac{e}{\alpha_K} \right)^t = \left( 1 + \frac{e}{\alpha_K} \right)^r - 1 \leq \frac{2er}{\alpha_K}.
\]
Thus Lemma [5] gives

\[
\exp \left\{ -\frac{\alpha \kappa}{4er} \right\} \geq \mathbb{P}(Y_a \notin \langle \mathcal{G}^* \rangle) \geq \mathbb{P}(|Y_a| \leq m) \mathbb{P}(W \notin \langle \mathcal{G}^* \rangle) \geq \mathbb{P}(W \notin \langle \mathcal{G}^* \rangle)/2,
\]
where we use the fact that any binomial \( \xi \) with \( \mathbb{E}[\xi] \in \mathbb{Z} \) satisfies \( \mathbb{P}(\xi \leq \mathbb{E}[\xi]) \geq 1/2 \).

**2.2 Completing the proof**

It will be convenient to prove the theorem assuming \( \mathcal{H} \) is \((2\kappa)\)-spread. Let \( \gamma \) and \( C_0 \) be as in Section 2 and \( \mathcal{H} \) be as in the statement of Theorem [1] and recall that asymptotics refer to \( r \).

In what follows we will have a sequence \( \mathcal{H}^*_i \), with \( \mathcal{H}^*_0 = \mathcal{H}^* \) and
\[
\mathcal{H}^*_i \subseteq \{ \chi_i(S, W_i) : S \in \mathcal{H}^*_{i-1} \},
\]
where \( W_i \) and \( \chi_i \) will be defined below (with \( \chi_i \) a version of the \( \chi \) of Section 2).

Set \( C = C_0 \) and \( p = C/\kappa \), define \( \ell \) by \((1 - \gamma)\ell = \sqrt[\ell]{\log r}/r \), and set \( \rho = \log r/\kappa \). Then \( \ell \leq \gamma^{-1} \log r \)
and Theorem [1] will follow from the next assertion.
Claim 1. If \( W \) is a uniform \((lp + \rho)N\)-subset of \( \hat{E} \), then \( W \in \langle \mathcal{H}^* \rangle = \bigcup_{S \in \mathcal{H}^*} \langle S \rangle \) w.h.p.

Proof. Set \( \delta = 1/(2\ell) \). Let \( r_0 = r \) and \( r_i = (1 - \gamma)r_{i-1} = (1 - \gamma)^i r_0 \) for \( i \in [\ell] \). Let \( X_0 = \hat{E} \) and, for \( i = 1, \ldots, \ell \), let \( W_i \) be chosen uniformly from \( \binom{X_{i-1}}{Np} \) and set \( X_i = X_{i-1} \setminus W_i \). The sequence \( X_i, W_i, i = 1, 2, \ldots, \ell + 1 \) is defined below. Note the assumption \( \kappa \gg \log r \) ensures that

\[
|X_i| = N - \ell Np \geq N \left( 1 - \frac{C_0 \gamma^{-1} \log r}{\kappa} \right) \geq \frac{N}{2}.
\]

For \( S \in \mathcal{H}_{i-1}^* \) let \( \chi_i(S, W_i) = S' \setminus W_i \), where \( S' \) is a member of \( \mathcal{H}_{i-1}^* \) contained in \( W_i \cup S \). Say that \( S \) is good if \( |\chi_i(S, W_i)| \leq r_i \) (and bad otherwise), and set

\[
\mathcal{H}_i^* = \{ \chi_i(S, W_i) : S \in \mathcal{H}_{i-1}^* \text{ is good} \}.
\]

Thus \( \mathcal{H}_i^* \) is an \( r_i \)-bounded collection of subsets of \( X_i \). Finally, choose \( W_{\ell+1} \) uniformly from \( \binom{X_\ell}{Np} \). Then \( W = W_1 \cup \cdots \cup W_{\ell+1} \) is distributed as required for Claim 1. Note also that \( W \in \langle \mathcal{H}^* \rangle \) whenever \( W_{\ell+1} \in \langle \mathcal{H}_i^* \rangle \).

(More generally, \( W_1 \cup \cdots \cup W_i \cup Y \in \langle \mathcal{H}^* \rangle \) whenever \( Y \subseteq X_i \) lies in \( \langle \mathcal{H}_i^* \rangle \).)

So to prove the claim, we just need to show

\[
\mathbb{P}(W_{\ell+1} \in \langle \mathcal{H}_i^* \rangle) = 1 - o(1)
\]

(19)

where the \( \mathbb{P} \) refers to the entire sequence \( W_1 \ldots W_{\ell+1} \).

For \( i \in [\ell] \) call \( W_i \) successful if \( |\mathcal{H}_i^*| \geq (1 - \delta)|\mathcal{H}_{i-1}^*| \), call \( W_{\ell+1} \) successful if it lies in \( \langle \mathcal{H}_i^* \rangle \), and say a sequence of \( W_i \)'s is successful if each of its entries is. We show that

\[
\mathbb{P}(W_1 \ldots W_{\ell+1} \text{ is successful}) = 1 - \exp \left[ -\Omega(\sqrt{\log r}) \right].
\]

(20)

Now \( W_1 \ldots W_{i-1} \) is successful implies that \( |\mathcal{H}_i^*| > (1 - \delta)^i |\mathcal{H}^*| > |\mathcal{H}^*|/2 \). So for \( I \subseteq X_{i-1} \) we have

\[
|\mathcal{H}_{i-1}^* \cap \langle I \rangle| \leq |\mathcal{H}^* \cap \langle I \rangle| \leq \left( \frac{e}{2 \kappa q} \right)^{|I|} |\mathcal{H}^*| \leq \left( \frac{e}{q \kappa} \right)^{|I|} |\mathcal{H}_{i-1}^*|.
\]

We therefore have the spread condition \( \ref{eq:spread} \) for \( \mathcal{H}_{i-1}^* \). For \( i \in [\ell] \), according to Lemma \( \ref{lem:spread} \) (and Markov’s Inequality),

\[
\mathbb{P}(W_i \text{ is not successful} \mid W_1 \ldots W_{i-1} \text{ is successful}) < \delta^{-1} C^{-r_{i-1}/3},
\]

Thus

\[
\mathbb{P}(W_1 \ldots W_\ell \text{ is successful}) > 1 - \delta^{-1} \sum_{i=1}^{\ell} C^{-r_{i-1}/3} > 1 - \exp \left\{ -\sqrt{\log r}/4e \right\}
\]

(21)

(using \( r_\ell = \sqrt{\log r} \)).

Finally, if \( W_1 \ldots W_\ell \) is successful, then Corollary \( \ref{cor:bound} \) applied with \( G = \mathcal{H}_\ell \), \( Y = X_\ell \), \( \alpha = N \rho / |Y| \geq \rho \), \( r = r_\ell \), and \( W = W_{\ell+1} \) gives

\[
\mathbb{P}(W_{\ell+1} \notin \langle \mathcal{H}_\ell^* \rangle) \leq 2 \exp \left\{ -\sqrt{\log r}/4e \right\},
\]

(22)

and we have \( \ref{eq:claim} \) and the claim.

We used the following to obtain \( \ref{eq:claim} \):

\[
\frac{\alpha \kappa}{er_\ell} = \frac{N \rho \cdot \kappa}{e|X_\ell|r_\ell} \geq \frac{(N \log r_0 / \kappa) \cdot \kappa}{e|X_\ell|(1 - \gamma)^i r_0} = \frac{N \log r_0}{e|X_\ell|(\sqrt{\log r_0 / r_0}) r_0} \geq \frac{N \log r_0}{eN(\sqrt{\log r_0 / r_0}) r_0}.
\]

This completes the proof of Theorem \( \ref{thm:main} \) modulo the verification of inequality \( \ref{ineq:verification} \).
2.3 Proof of inequality \[18\]

We begin with a modification of Harris’s inequality \[8\].

2.3.1 A modification of Harris’s Inequality

We have a partition \(P_1, P_2, \ldots, P_n\) of the set \([n]\) and \(\Omega = \{x \in \{0, 1\}^n : \sum_{t \in P_j} x_t \leq 1, \text{for } j \in [N]\}\). Let \(f, g\) be real monotone increasing functions defined on \(\Omega\) in the sense that if \(x, y \in \Omega\) and \(x \leq y\) then \(f(x) \leq f(y)\), and similarly for \(g\). Turn \(\Omega\) into a probability space by setting \(\xi_j = 1, j \in [N]\) with probability \(p_j\) for \(j \in [N]\).

Then for each \(j \in [N]\) for which \(\xi_j = 1\), choose \(t\) uniformly at random from \(P_j\) and set \(x_t = 1\) and \(x_s = 0\) for \(s \in P_j \setminus \{t\}\). Harris’s Inequality comprises the following lemma when \(N = n\) and \(P_j = \{j\}, j \in [N]\).

Lemma 7. With \(f, g, \Omega\) as above, \(\mathbb{E}(fg) \geq \mathbb{E}(f)\mathbb{E}(g)\).

Proof. We just repeat the proof of Harris’s Inequality. We will prove the lemma by induction on \(N\). If \(N = 0\) then \(\mathbb{E}(f) = a, \mathbb{E}(g) = b\) and \(\mathbb{E}(fg) = ab\) for some constants \(a, b\).

So assume the truth for \(N - 1\). Suppose that \(\mathbb{E}(f | \xi_N = 0) = a\) and \(\mathbb{E}(g | \xi_N = 0) = b\) then

\[\mathbb{E}((f - a)(g - b)) - \mathbb{E}(f - a)\mathbb{E}(g - b) = \mathbb{E}(fg) - \mathbb{E}(f)\mathbb{E}(g).\]

By replacing \(f\) by \(f - a\) and \(g\) by \(g - b\) we may therefore assume that \(\mathbb{E}(f | \xi_N = 0) = \mathbb{E}(g | \xi_N = 0) = 0\). By monotonicity, we see that \(\mathbb{E}(f | \xi_N = 1), \mathbb{E}(g | \xi_N = 1) \geq 0\).

We observe that by the induction hypothesis that

\[
\begin{align*}
\mathbb{E}(fg | \xi_N = 0) &\geq \mathbb{E}(f | \xi_N = 0)\mathbb{E}(g | \xi_N = 0) = 0 \\
\mathbb{E}(fg | \xi_N = 1) &\geq \mathbb{E}(f | \xi_N = 1)\mathbb{E}(g | \xi_N = 1) \geq 0
\end{align*}
\]

Now, by the above inequalities,

\[
\mathbb{E}(fg) = \mathbb{E}(fg | \xi_N = 0)(1 - p_N) + \mathbb{E}(fg | \xi_N = 1)p_N
\]

\[
\geq \mathbb{E}(f | \xi_N = 1)\mathbb{E}(g | \xi_N = 1)p_N. \tag{23}
\]

Furthermore,

\[
\mathbb{E}(f)\mathbb{E}(g) = (\mathbb{E}(f | \xi_N = 0)(1 - p_N) + \mathbb{E}(f | \xi_N = 1)p_N) \times
\]

\[
(\mathbb{E}(g | \xi_N = 0)(1 - p_N) + \mathbb{E}(g | \xi_N = 1)p_N)
\]

\[
= \mathbb{E}(f | \xi_N = 1)\mathbb{E}(g | \xi_N = 1)p_N^2. \tag{24}
\]

The result follows by comparing (23) and (24) and using the fact that \(\mathbb{E}(f | \xi_N = 1), \mathbb{E}(g | \xi_N = 1) \geq 0\) and \(0 \leq p_N \leq 1\). \(\square\)

Of course, if \(f, g\) are both decreasing then we also have

\[
\mathbb{E}(fg) = \mathbb{E}((-f)(-g)) \geq \mathbb{E}(-f)\mathbb{E}(-g) = \mathbb{E}(f)\mathbb{E}(g). \tag{25}
\]
2.3.2 A modification of Janson’s Inequality

For \( x = (x_1, x_2, \ldots, x_n) \in \{0, 1\}^n \) define \( D(x) \subseteq [n] \), by \( i \in D(x) \) iff \( x_i = 1 \). On the other hand, given \( D \), define \( x(D) \) by \( D(x(D)) = D \). Fix a family of \( m \) subsets \( D_i \subseteq [n], i \in [m] \) where \( x(D_i) \in \Omega \) for \( i \in [m] \), where \( \Omega \) is as in Section 2.3.1. For each \( i \), let \( E_i = \{ j : D_i \cap P_j \neq \emptyset \} \).

We repeat the proof of Janson’s Inequality. Let \( R = D(x) \) where \( x \) is chosen randomly from \( \Omega \) as in Section 2.3.1. Let \( A_i \) be the event that \( D_i \) is a subset of \( R \). Moreover, let \( I_i \) be the indicator of the event \( A_i \). Note that, \( I_i \) and \( I_j \) are independent if \( E_i \cap E_j = \emptyset \) and if \( t \in E_i \cap E_j \) then \( D_i \cap P_t = D_j \cap P_t \) or \( I_i I_j = 0 \). One can easily see that the \( I_i \)'s are increasing. We let

\[
S_m = I_1 + I_2 + \cdots + I_m,
\]

and

\[
\mu = \mathbb{E}S_m = \sum_{i=1}^{m} \mathbb{E}(I_i).
\]

We write \( i \sim j \) if \( E_i \cap E_j \neq \emptyset \). Then, let

\[
\Delta = \sum_{\{i,j\} : i \sim j} \mathbb{E}(I_i I_j).
\] (26)

Let \( \phi(x) = (1 + x) \log(1 + x) - x \). Now, with \( S_m, \Delta, \phi \) given above one can establish the following upper bound on the lower tail of the distribution of \( S_m \).

**Lemma 8.** For any real \( t, 0 \leq t \leq \mu \),

\[
\mathbb{P}(S_m \leq \mu - t) \leq \exp \left\{ -\frac{\phi(-t/\mu)\mu^2}{\Delta} \right\} \leq \exp \left\{ -\frac{t^2}{2\Delta} \right\}.
\] (27)

**Proof.** Put \( \psi(\lambda) = \mathbb{E}(e^{-\lambda S_m}), \lambda \geq 0 \). By Markov’s inequality we have

\[
\mathbb{P}(S_m \leq \mu - t) \leq e^{\lambda(\mu - t)} \mathbb{E}(e^{-\lambda S_m}).
\]

Therefore,

\[
\log \mathbb{P}(S_m \leq \mu - t) \leq \log \psi(\lambda) + \lambda(\mu - t).
\] (28)

Now let us estimate \( \log \psi(\lambda) \) and minimise the right-hand-side of (28) with respect to \( \lambda \).

Note that

\[
-\psi'(\lambda) = \mathbb{E}(S_m e^{-\lambda S_m}) = \sum_{i=1}^{m} \mathbb{E}(I_i e^{-\lambda S_m}).
\] (29)

Now for every \( i \in [n] \), split \( S_m \) into \( Y_i \) and \( Z_i \), where

\[
Y_i = \sum_{j:j \sim i} I_j, \quad Z_i = \sum_{j:j \not\sim i} I_j, \quad S_m = Y_i + Z_i.
\]

We note that Lemma 7 remains true if we condition on \( I_i = 1 \) for some \( i \in [m] \). The conditioning basically just restricts \( \Omega \) to a smaller set. Then by Lemma 7 with \( f = -e^{-\lambda Y_i}, g = -e^{-\lambda Z_i} \) and \( R \) with \( I_i \) fixed at one, we get, setting \( p_i = \mathbb{E}(I_i) \),

\[
\mathbb{E}(I_i e^{-\lambda S_m}) = p_i \mathbb{E}(e^{-\lambda Y_i} e^{-\lambda Z_i} | I_i = 1) \geq p_i \mathbb{E}(e^{-\lambda Y_i} | I_i = 1) \mathbb{E}(e^{-\lambda Z_i} | I_i = 1).
\]
Since $Z_i$ and $I_i$ are independent we get
\[ \mathbb{E}(I_i e^{-\lambda S_m}) \geq p_i \mathbb{E}(e^{-\lambda Y_i} \mid I_i = 1) \mathbb{E}(e^{-\lambda Z_i}) \geq p_i \mathbb{E}(e^{-\lambda Y_i} \mid I_i = 1) \psi(\lambda). \] (30)

From [29] and [30], applying Jensen’s inequality (twice) and remembering that $\mu = \mathbb{E}S_m = \sum_{i=1}^{m} p_i$, we get
\[-(\log \psi(\lambda))' = -\frac{\psi'(\lambda)}{\psi(\lambda)} \geq \sum_{i=1}^{m} p_i e^{-\lambda Y_i} \mid I_i = 1 \]
\[ \geq \mu \sum_{i=1}^{m} \frac{p_i}{\mu} \exp \{-\mathbb{E}(\lambda Y_i \mid I_i = 1)\} \geq \mu \exp \left\{-\frac{1}{\mu} \sum_{i=1}^{m} p_i \mathbb{E}(\lambda Y_i \mid I_i = 1)\right\} \]
\[ = \mu \exp \left\{-\frac{\lambda}{\mu} \sum_{i=1}^{m} \mathbb{E}(Y_i I_i)\right\} = \mu e^{-\lambda \Delta/\mu}. \] (31)

So
\[-(\log \psi(\lambda))' \geq \mu e^{-\lambda \Delta/\mu} \] (32)
which implies that
\[-\log \psi(\lambda) \geq \int_{0}^{\lambda} \mu e^{-z/\mu} dz = \frac{\mu^2}{\Delta} (1 - e^{-\lambda \Delta/\mu}). \] (33)

Hence by [33] and [28]
\[ \log \mathbb{P}(S_m \leq \mu - t) \leq -\frac{\mu^2}{\Delta} (1 - e^{-\lambda \Delta/\mu}) + \lambda (\mu - t), \] (34)
which is minimized by choosing $\lambda = -\log(1 - \frac{\mu}{\Delta})/\mu$. It yields the first bound in [27], while the final bound in [27] follows from the fact that $\phi(x) \geq \frac{x^2}{2}$ for $x \leq 0$. \qed

To obtain (18) we partition $Y = X \times [q]$ into $P_i = X \times \{i\}$ for $i = 1, 2, \ldots, N$ and let the $D_i$ correspond to the edges of $G^*$. We then apply Lemma 8 with $t = \mu$.

### 3 Proof of Theorem 3

The strategy here is similar to that used to prove Theorem 2. The main difference is that we replace the $O(\log r)$ rounds by a single round where we obtain almost all of a rainbow edge of $\mathcal{H}$.

Let $N = |X|$ and $m = \frac{CN}{\kappa}$ and $k = r^{1/2}$. For $W \subseteq \hat{E}$, $|W| = m$ and $S \in \mathcal{H}^*$ we say that $(S, W)$ is bad if $|T \setminus W| > k$ for all $T \in \mathcal{H}^*$, $T \subseteq S \cup W$. Otherwise $(S, W)$ is good.

In the course of the proof, we make some claims that will be verified later.

Let $p = \frac{3m}{N}$ and $p_0 = \frac{2p}{5}$ and define $p_1$ by $(1 - p) = (1 - p_0)(1 - p_1)$ so that $X_{p_0}^* = X_{p_0}^* \cup X_{p_1}^*$, where $X_{p_0}^*$ is distributed as $Y_\alpha$ at the beginning of Section 2.1 with $Y = X$, $\alpha = p$. The size of $X_{p_0}^*$ is distributed as the binomial $\text{Bin}(N, 6m/5N)$ and so the Chernoff bounds imply that w.h.p. $|X_{p_0}^*| \geq m$. Let $W_0$ be distributed as $\tilde{X}_m^*$ i.e. as in the statement of Corollary 6. In what follows we generate $W = W_0 \cup W_1$ where $W_1$ is distributed as $|X_{p_1}^*|$. Let $W_0$ be a success if $|\{S \in \mathcal{H}^* : (S, W_0) \text{ is bad}\}| \leq |\mathcal{H}^*|/2$.

**Claim 2.** If $t \geq k$ then
\[ |\{ (S, W) : (S, W) \text{ is bad and } |S \cap W| = t \}| \leq 2C^{-k/3}|\mathcal{H}^*| \left(\frac{r}{t}\right)^{(N - r)} (m - t)^q. \] (35)
Now
\[ |\{W_0 : W_0 \text{ is a failure}\}| \times \frac{|\mathcal{H}^*|}{2} \leq \sum_{W \subseteq \hat{E}} |\{(S, W) : (S, W) \text{ is bad}\}|. \]

Claim 2 shows that
\[
\mathbb{P}(W_0 \text{ is not a success}) = \left| \left\{ W_0 : W_0 \text{ is not a success} \right\}\right| \leq 2 \sum_{W \subseteq \hat{E}} |\{(S, W) : (S, W) \text{ is bad}\}| \leq 4C^{-k/3}. 
\]
(We have used the Vandermonde identity \((\binom{N}{m}) = \sum_{t \geq 0} (\binom{N}{t} (\binom{N-r}{m-t})\).)

Suppose now that \(W_0 \) is a success and then let \(R = \{S \in \mathcal{H}^* : |S \setminus W_0| \leq k\} \) and for each \(S \in R \) let \(\eta(S)\) denote some \(k\)-subset of \(S\) that contains \(S \setminus W_0\). Let \(Z\) denote the number of sets \(S \in R\) such that \(\eta(S)\) is contained in \(W_1\). Then we have
\[
\mathbb{E}(Z) = |R| \frac{p_1^k}{q^k}. 
\]
(36)

Now
\[
\mathbb{V}(Z) \leq 2 \sum_{t=1}^{k} \sum_{A,B \in R} |A \cap B| \leq 2 |\mathcal{R}| \frac{p_1^k}{q^k} \sum_{t=1}^{k} \left( \frac{K_0}{\kappa} \right)^t \leq 2 |\mathcal{R}| \frac{p_1^k}{q^k} \sum_{t=1}^{k} \left( \frac{eK_0}{C} \right)^t \leq 3K_0 \mathbb{E}(Z)^2. 
\]
(37)

(We have used \(\kappa p_1 \geq \kappa m/N \geq C\) to get the third inequality.)

The Chebyshev inequality implies that
\[
\mathbb{P}(Z = 0) \leq \frac{\mathbb{V}(Z)}{\mathbb{E}(Z)^2} \leq \frac{3K_0}{C}. 
\]
Putting \(C_\varepsilon = 3K_0/\varepsilon\) verifies (5).

3.1 Proof of Claim 2

Fix \(t\). We bound the number of bad \((S, W)\)’s with \(|W \cap S| = t\). In which case, \(|W \setminus S| = m - t + r\). Call \(Y \in \binom{\hat{E}}{m-t+r}\) pathological if
\[
|\{S \subseteq Y : (S, Y \setminus S) \text{ is bad}\}| > C^{-k/3} q^{-t-r} |\mathcal{H}^*| \left( \binom{N-r}{m-t} \right) \left( \binom{N}{m-t+r} \right). 
\]
We say that \((S, W)\) is pathological, if \(Y = S \cup W\) is.

Non-pathological contributions: To specify such, we choose \(Y = S \cup W\), then \(S, W\). This gives at most
\[
\left(\binom{N}{m-t+r} q^{m-t+r} \times C^{-k/3} q^{-t-r} |\mathcal{H}^*| \left(\frac{N-r}{m-t} \right) \right) \left(\binom{N}{m-t} \times \binom{r}{t} \right) = C^{-k/3} |\mathcal{H}^*| \left(\frac{r}{t} \right) \left(\frac{N-r}{m-t} \right) q^{m-2t}.
\]
(38)

Pathological contributions:

Claim 3. For a fixed \(S \in \mathcal{H}^*\) and random \(W \setminus S\) from \(\binom{E \setminus S}{m-t}\) and \(t \geq k\) and large enough \(C\),
\[
\mathbb{E}(|\{\mathcal{H}^* \ni J \subseteq W \cup S : |J \cap S| = t\}|) \leq C^{-2k/3} q^{-t} |\mathcal{H}^*| \left(\frac{N-r}{m-t} \right) \left(\frac{N}{m-t+r} \right).
\]
(39)

Assume Claim 3 for now.

We choose \((S, W \cap S)\) in at most \(|\mathcal{H}^*| \left(\frac{r}{t} \right)\) ways.
(40)

Now \((S, W)\) bad implies that every \(\mathcal{H}^* \ni J \subseteq S \cup W\) satisfies \(|J \cap S| \geq |J \setminus W| \geq k\). Because \((S, W)\) is pathological,
\[
a_w = |\{\mathcal{H}^* \ni J \subseteq S \cup W : |J \cap S| = t\}| \geq C^{-k/3} q^{-t} |\mathcal{H}^*| \left(\frac{N-r}{m-t} \right) \left(\frac{N}{m-t+r} \right).
\]
Claim 3 implies that
\[
\frac{1}{\binom{N-r}{m-t} q^{m-t}} \sum_{W \setminus S \in \binom{E \setminus S}{m-t}} a_w \leq C^{-2k/3} q^{-t} |\mathcal{H}^*| \left(\frac{N-r}{m-t} \right) \left(\frac{N}{m-t+r} \right).
\]

Thus the number of choices for \(W \setminus S\) is at most
\[
\frac{C^{-2k/3} q^{m-t-r} |\mathcal{H}^*| \left(\frac{N-r}{m-t} \right) \left(\frac{N}{m-t+r} \right)^2}{\left(\frac{N}{m-t+r} \right)^2} = C^{-k/3} q^{m} \left(\frac{N-r}{m-t} \right).
\]
(41)

Claim 2 now follows from (38), (40) and (41).

\[\square\]

Proof of Claim 3 With \(f_{t,S}^* = f_{t,S} \left(\frac{q-t}{q} \right) \leq f_{t,S} \left(\frac{t}{q} \right)\) equal to the fraction of \(J\)'s in \(\mathcal{H}^*\) with \(|J \cap S| = t\), the left hand side of (39) is equal to
\[
f_{t,S}^* |\mathcal{H}^*| \mathbb{P}(J \subseteq W \cup S) = f_{t,S}^* |\mathcal{H}^*| \left(\frac{m-t}{N-r-t} \right) q^{r-t},
\]
for arbitrary \(J \in \mathcal{H}^*, \ |J \cap S| = t\).

It is therefore enough to show that
\[
f_{t,S}^* \left(\frac{m-t}{N-r-t} \right) \left(\frac{N-r}{m-t+r} \right) \left(\frac{N-r}{m-t} \right) \leq \frac{1}{C^{2k/3} q^t}.
\]
(42)
Observe that
\[
\frac{\binom{m-t}{r-t}}{\binom{N-r}{r-t}} \leq \left( \frac{m-t}{N-r} \right)^{r-t} \quad \text{and} \quad \frac{\binom{N}{m-t+r}}{\binom{N-r}{m-t}} \leq \left( \frac{N-r}{N-m} \right)^r.
\]
This implies that
\[
\frac{\binom{m-t}{r-t}}{\binom{N-r}{r-t}} \leq \left( \frac{N-r}{N-m} \right)^t \leq \left( \frac{N}{m} \right)^t = \left( \frac{\kappa}{C} \right)^t.
\]
(43)

Now for \( \alpha r \leq t \leq r \), the fact that \( H^* \) is \( e^{-1}q\kappa \) spread implies
\[
f^*_t S \leq \left( \frac{r}{t} \right) \left( \frac{e}{q\kappa} \right)^t \leq \left( \frac{e^2}{\alpha q\kappa} \right)^t.
\]
(44)

For \( k \leq t \leq \alpha r \) we have from (4) that
\[
f^*_t S \leq \left( \frac{eK_0}{q\kappa} \right)^t.
\]
(45)

Equation (42) (and Claim 3) now follow from (43), (44) and (45), assuming \( K_0 \geq e/\alpha \). □

Acknowledgement We thank Sam Spiro for pointing out an issue in a previous version.

References


A Powers of Hamiton cycles

We verify [4] for the hypergraph \( \mathcal{H} \) whose edges correspond to the \( k \)th power of a Hamilton cycle. As in [10] we split this into two propositions and modify their proof for \( k = 2 \).

Proposition 1. For \( J \subseteq \binom{[n]}{2} \), with \( t \leq n/3 \) edges, inducing \( c \) components,

\[
|\mathcal{H} \cap \langle J \rangle| \leq (2k)^{2t} \left( n - \left\lceil \frac{t + (2k-1)c}{k} \right\rceil - 1 \right)!.
\]

Proof. Let \( S_1, \ldots, S_c \) be the components of the subgraph induced by the edges \( S \) and let \( v = |V(S)| \) where \( V(A), E(A) \) is the set of vertices,edges used by a subgraph \( A \). The upper bound on \( t \) implies that no \( S_j \) can wrap around, and so \( |E(S_j)| \leq k|V(S_j)| - (2k - 1) \) for each \( j \) and so

\[
t \leq kv - (2k - 1)c.
\]

We designate a root vertex \( v_j \) for each \( S_j \) and order \( V(S_j) \) by some order \( \prec_j \) that begins with \( v_j \) and in which each \( v \neq v_j \) appears later than at least one of its neighbors. We may then bound \( |\mathcal{H} \cap \langle S \rangle| \) as follows. To specify a \( J \) containing \( S \), we first specify a cyclic permutation of \( \{v_1, \ldots, v_c\} \sqcup ([n] \setminus V(S)) \). By (46), the number of ways to do this (namely, \( (n - v + c - 1)! \)) is at most \( \left( n - \left\lceil \frac{t+(2k-1)c}{k} \right\rceil - 1 \right)! \). We then extend to a full cyclic ordering of \( [n] \) (thus determining \( J \)) by inserting, for \( j = 1, \ldots, c \), the vertices of \( V(S_j) \) \( \{v_j\} \) in the order \( \prec_j \). This allows at most \( 2k \) places to insert each vertex (since one of its neighbours has been inserted before it and the edge joining them must belong to \( J \)), so the number of possibilities here is less than \( (2k)^v \leq (2k)^{2t} \), and the proposition follows.

Proposition 2. For \( J \subseteq S \in \mathcal{H}, |J| = s \leq n/3 \), the number of subgraphs of \( J \) with \( t \) edges and \( c \) components is at most \( (4ke)^t \binom{k^t}{c} \).

Proof. To specify a subgraph \( J \) of \( S \) we proceed as follows. We first choose root vertices \( v_1, \ldots, v_c \) for the components, say \( S_1, \ldots, S_c \), of \( J \), the number of possibilities for this being at most \( \binom{k^s}{c} \). We then choose the sizes, say \( t_1, \ldots, t_c \), of \( S_1, \ldots, S_c \); here the number of possibilities is at most \( \binom{t}{c-1} \). Finally, we specify for each \( i \) a connected \( S_i \) of size \( t_i \) rooted at \( v_i \) in at most \( \prod_{i=1}^c (2ke)^{t_i} \) ways. Combining these estimates (with the crude \( \binom{t}{c-1} < 2^t \) yields the proposition.

It follows from these two propositions that if \( S \in \mathcal{H} \) and \( 1 \leq t \leq n/3 \) then

\[
f_{t,S} \leq \sum_{c=1}^{t} (2k)^{2t} \left( n - \left\lceil \frac{t + (2k-1)c}{k} \right\rceil - 1 \right)! \times (4ke)^t \binom{kt}{c} \times \frac{1}{(n-1)!}
\]

\[
\leq 2 \sum_{c=1}^{t} (16k^3e)^t \binom{kt}{c} \left( e \left\lceil \frac{t+(2k-1)c}{k} \right\rceil \right) \left( n - \left\lceil \frac{t+(2k-1)c}{k} \right\rceil \right) - 1 \right) \right)^n \left\lceil \frac{t+(2k-1)c}{k} \right\rceil - 1
\]

\[
\leq e^{O(t)} \sum_{c=1}^{t} n^{-\left\lceil \frac{t+(2k-1)c}{k} \right\rceil}
\]

\[
= O \left( \frac{O(1)}{n^{1/k}} \right)^t.
\]