Hypergraph covering and small gaps between primes

Kevin Ford

University of Illinois at Urbana-Champaign

May, 2018

Prime gaps and hypergraph covering

Goal: Cover $[0, y]$ with residue classes $a_p \mod p$, $p \leqslant x$.

- Q - a random set of primes in $[1, y]$, size $\gg \frac{y}{\log^2 y}$. (Stage 2)
- $\mathcal{P}$ - the set of primes in $(x/4, x/2]$
- a_p - some residue class modulo p, for each $p \in P$.
- $e_p = Q \cap (a_p \mod p)$; primes in Q that are $\equiv a_p \mod p$

Goal: choose the a_p so that the sets e_p cover most of Q.

Hypergraph language: Q is the vertex set, e_p are hyperedges

Sieve input: whp, for most $p \in \mathcal{P}$, a_p exist so that $e_p \gg \log_2 y$.

Big question. Can one choose the a_p so that the e_p are both large *and* cover Q efficiently (allowing a larger choice for y)?

Hypergraph covering

$H = (V, E)$ - a hypergraph
V - finite set of vertices
E - collection of nonempty subsets of V (hyperedges)

A covering of H is a subset of E that covers all of V
A packing (or matching) is a subset of disjoint elements of E
A perfect matching (packing) is both a matching and a covering.

Problems: under what general conditions on H does their exist

- a perfect matching
- a $(1-\varepsilon)$-near perfect matching (a matching that covers all but at most $\varepsilon|V|$ vertices
- a $(1+\varepsilon)$-efficient covering (a covering where at most $\varepsilon|V|$ vertices are covered twice)

Pippenger-Spencer

Pippenger-Spencer Theorem (1989). building on earlier work of Pippenger (unpublished) and Frankl-Rödl.

Three basic conditions on H:

1. (l-uniformity) $|e| = l$ for all $e \in E$, l fixed;
2. (regularity) $\forall\ v, w \in V$, deg$(v) \sim$ deg(w);
3. (small codegrees) $\forall\ v, w \in V$, $v \neq w$, codeg$(v, w) = o($deg$(v))$, where codeg$(v, w) = |\{e \in E : v, w, \in e\}|$.

Here $o(1)$ means as $|V| \to \infty$, where we assume that the typical vertex degree is also $\to \infty$.

Conclusion: There is a $(1 - o(1))$-near perfect matching of H.

An inefficient method of covering/matching, I

The naive method of choosing edges randomly and independently is very inefficient for producing (near) matchings/coverings. Why?

1. After relative few choices one encounters overlaps
2. After many choices the overlapped parts begin to dominate the non-overlapped parts
3. Even after a great number of choices, there is still a lot left uncovered

An inefficient method of covering/matching, II

Analysis of the random, uniform choice method:

Assumptions on H: $|E| = l$; $\deg(v) \sim d\ \ \forall v \in V$ (regularity)

A (near) perfect matching/covering will use about l/d edges

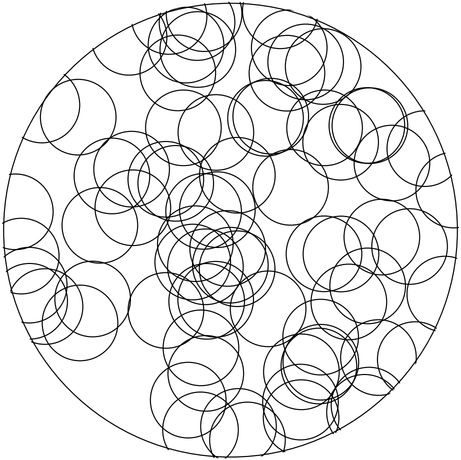
Suppose we have chosen $J = \lambda l/d$ edges $e_1, \ldots, e_J$, $\lambda > 0$ fixed. For any vertex v,

$$\mathcal{P}\left(v \notin \bigcup_{j=1}^{J} e_j\right) = \prod_{j=1}^{J}(1 - \mathcal{P}(v \in e_j)) \approx \left(1 - \frac{d}{l}\right)^J \sim e^{-\lambda}.$$

Therefore, we expect about $e^{-\lambda}|V|$ uncovered vertices. That is, no matter how large we take λ, there is a lot left uncovered **and** what is covered is highly overlapped.

An inefficient method of covering/matching, III

Big circle = V; small circles = hyperedges

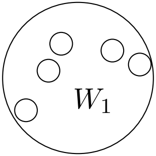


$\lambda = 0.1389 \lambda = 0.2778 \lambda = 0.4167 \lambda = 0.5556 \lambda = 0.6944 \lambda = 0.8333 \lambda = 0.9722 \lambda = 1.1111 \lambda = 1.2500 \lambda = 1.3889 \lambda = 1.5278 \lambda = 1.6667 \lambda = 1.8056 \lambda = 1.9444 \lambda = 2.0833$

Better method: The Rödl Nibble

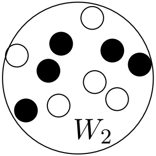
(Nibble # 1). Choose a small number, n_1 ($n_1 = o(l/d)$, say), of edges independently at random: $e_1, \ldots, e_{n_1}$. With high probability, they are disjoint. Let

$$W_1 = V \backslash \{e_1 \cup \cdots \cup e_{n_1}\}$$



(Nibble # 2). Choose a small number, n_2, of edges at random, $e_{n_1+1}, \ldots, e_{n_1+n_2}$, **but only from those edges** $\subset W_1$. Let

$$W_2 = W_1 \backslash \{e_{n_1+1} \cup \cdots \cup e_{n_1+n_2}\}$$



Continue for k nibbles.

Relaxing the hypotheses

Pippenger-Spencer: WHP (with high probability), get a $(1 - o(1))$-near perfect matching, assuming uniformity, regularity, small codegrees.

For our prime gap application, our hypergraph is much more irregular:

- The hyperedges have greatly varying sizes (but none are too big);
- The vertices have greatly varying degrees (but none are too large);

Our hypotheses

1. $|e| \leqslant r$ for all $e \in E$; (r need not be fixed)
2. $\forall\, v \in V$, $\deg(v) \leqslant d$;
3. $\forall\, v, w \in V$, $v \neq w$, $\mathrm{codeg}(v, w) \leqslant \delta \deg(v)$ for some small δ.

Rödle nibble under relaxed hypotheses

Hyp: $|e| \leqslant r$; $\deg(v) \leqslant d$; $\mathrm{codeg}(v,w) \leqslant \delta \deg(v))$; $|E| = l$.

(Nibble # 1). Choose random edges $e_1, \ldots, e_{n_1}$. WHP, they are disjoint. Denote $W_1 = V \backslash \{e_1 \cup \cdots \cup e_{n_1}\}$. For all $v \in V$,

$$\mathbb{P}(v \in W_1) = \prod_{i=1}^{n_1} (1 - \mathbb{P}(v \in e_i))$$

$$\sim \exp\left(- \sum_{i=1}^{n_1} \mathbb{P}(v \in e_i) \right) = \exp\left(- \frac{n_1 \deg(v)}{l} \right) =: P_1(v).$$

Note: $\deg(v)$ may be highly variable, hence so is $P_1(v)$. However, we have a universal lower bound on $P_1(v)$ from the upper bound on $\deg(v)$. $\mathbb{P}(v \in W_1) \sim \exp(-\frac{n_1 \deg(v)}{l}) =: P_1(v)$. Hence

$$\mathbb{E}|W_1| = \sum_{v \in V} P_1(v).$$

(Nibble # 2). Choose random edges $e_{n_1+1}, \ldots, e_{n_1+n_2} \subset W_1$, but **not with identical distribution.** Choose $e_i = e$ with probability

Main Theorem

$\exists C_0$ s.t. for $D, r \geqslant 1$, $0 < \kappa < \frac{1}{2}$, $m \geqslant 0$, n_i arbitrary,

$$0 < \delta \leqslant \left(\frac{\kappa^{2rm+1}}{C_0 \exp\{D(2rm+1)\}} \right)^{10^{m+2}},$$

and the hypergraph satisfies

1. $|e| \leqslant r$ for all $e \in E$;
2. $\deg(v) \leqslant \frac{\delta l}{\sqrt{\min(n_i)}}$ for all $v \in V$;
3. $\text{codeg}(v, w) \leqslant \frac{\delta l}{\min(n_i)}$ for $v \neq w$;
4. $\frac{n_i \deg(v)}{l P_i(v)} \leqslant D$ for $1 \leqslant i \leqslant m$; $P_m(v) \geqslant \kappa$ $(v \in V)$;

Then there are edges $e_1, \ldots, e_N \in E$, $N \leqslant n_1 + \cdots + n_m$, so that

$$|V \backslash (e_1 \cup \cdots \cup e_N)| \ll \sum_{v \in V} P_m(v)$$

Near perfect coverings

Corollary. Let $H = (V, E)$ be a hypergraph satisfying

1. $|e| = O(1)$ for all $e \in E$;
2. $d \leqslant \deg(v) \leqslant O(d)$, with $d = o(|E|)$, $d \to \infty$ as $|V| \to \infty$;
3. $\operatorname{codeg}(v, w) = o(d)$ for distinct $v, w \in V$;

Then there are $e_1, \ldots, e_N \in E$ with $N \leqslant (1 + o(1))\frac{|E|}{d}$ and

$$|e_1 \cup \cdots \cup e_N| = (1 + o(1))|V|.$$

If most vertex degrees are close to d, this is an efficient near-covering.