A counterexample to the very sparse greedy conjecture for random hitting sets

Peter Graham
https://orcid.org/0009-0007-8493-1505

November 3, 2026

Zenodo archive (Lean development in `lean`): https://doi.org/10.5281/zenodo.23915337

hexagon:2611.00859v1

Abstract

Let A be a random $m \times n$ zero-one matrix with independent Bernoulli(p) entries, whose rows are to be hit by a set of columns. Arpino, Dmitriev and Grometto proved that with high probability the linear programming (LP) relaxation, the fractional hitting set, is within a constant factor of the optimum when $mp \lesssim \log n$, and the greedy algorithm is within a constant factor of the optimum when $p \geq n^{-\delta}$ with $\delta < 1/2$. They conjectured that in the very sparse regime $mp \to 0$ the ratio of the greedy value to the LP value tends to one. We show that this fails. For $m = k$, $n = k^{10}$ and $p = 1/(10k^4)$, with the explicit constant $c = 1/(196608 \cdot 10^9)$, the greedy algorithm that breaks ties by the leftmost column, as in their analysis, uses at least $(1 + c/8)$ times the LP value with probability at least $c/4$ for every $k \geq 10^{20}$; these parameters satisfy their Assumption 2. The same holds with probability at least $c/8$ when ties are broken uniformly at random. So the ratio does not converge to one in probability. The obstruction is a path of three triples whose middle triple greedy takes first; exchanging it for its two ends saves half a column per path. The proofs are elementary counting arguments, checked in Lean from the definitions.

This work has not yet been reviewed by experts. The proofs, the Lean formalization and the text were produced by AI models (from OpenAI and Anthropic), working under the author's direction.

Keywords: hitting set, set cover, greedy algorithm, linear programming relaxation, random hypergraph, formal proof.
MSC 2020: 05C80, 68W40, 90C05, 60C05.

1 Introduction

1.1 Main results

Let A be an $m \times n$ matrix with entries in $\{0, 1\}$. Its rows are the sets to be hit and its columns the elements; write $A_j = \{i : A_{ij} = 1\}$ for the set of rows that column j hits. The matrix is *feasible* if every row is hit by some column. The greedy algorithm keeps a set U of uncovered rows, initially all of them. While $U \neq \emptyset$, it chooses a column j maximizing the *gain* $|A_j \cap U|$, removes A_j from U, and counts the choice. We write $G(A)$ for the count when ties are broken by the smallest column index. On a feasible matrix it stops with a set of $G(A)$ columns hitting every row. The LP value is

$$L(A) = \min\Big\{\sum_{j=1}^{n} x_j : x \in [0,1]^n, \sum_{j:\, i \in A_j} x_j \geq 1 \text{ for every row } i\Big\}.$$

Always $L(A) \le G(A)$ on feasible matrices.

Arpino, Dmitriev and Grometto [1] study a random matrix whose entries are independent Bernoulli(p) variables, with $m = m(n)$ and $p = p(n)$. Their Assumption 2 [1, p. 30:3] requires $c_0 n^{c_0} \le m \le C_0 n^{C_0}$ for some $c_0, C_0 > 0$, and $n^{-\delta} \le p \le 1/2$ for some $\delta \in (0,1)$, for all large n. Their Conjecture 12 [1, Section 6.2, p. 30:13], labeled "Very sparse", reads: "For $mp \ll 1$, $\mathrm{val}_{\mathrm{Gr}}/\mathrm{val}_{\mathrm{LP}} \to 1$." Here $\mathrm{val}_{\mathrm{Gr}}$ is the output of their Algorithm 1 [1, p. 30:7], the greedy algorithm above, and $\mathrm{val}_{\mathrm{LP}}$ is the LP value [1, equation (1.2)]. In the proof of their Theorem 10 [1, p. 30:11] ties are broken by "selecting the left-most column in the associated matrix". The conjecture first appeared, as Conjecture 1 in Section 7.2, in the preprint [2] of May 9, 2023, and Dmitriev's thesis restates it as open [6, Conjecture 6.1(1)].

Throughout, $c = 1/(196608 \cdot 10^9)$, and for an integer $k \ge 1$ we take

$$m = k, \qquad n = k^{10}, \qquad p = \frac{1}{10k^4}. \tag{1}$$

Theorem 1. *For every integer $k \ge 10^{20}$, with parameters (1),*

$$\Pr\Big[A \text{ is feasible and } G(A) \ge \big(1 + \tfrac{c}{8}\big) L(A)\Big] \ge \frac{c}{4}.$$

For the second theorem, let $G_{\mathrm{u}}(A)$ be the number of choices of the greedy algorithm when, at each step and given everything before it, the column is chosen uniformly at random among the columns of maximum positive gain. Columns with equal sets A_j count as different columns.

Theorem 2. *For every integer $k \ge 10^{20}$, with parameters (1),*

$$\Pr\Big[A \text{ is feasible and } G_{\mathrm{u}}(A) \ge \big(1 + \tfrac{c}{8}\big) L(A)\Big] \ge \frac{c}{8},$$

the probability being over the matrix and the choices together.

The proof in Section 4 gives the bound $c/4$ of Theorem 1; we state the weaker bound $c/8$, which is the one in the Lean statement (Section 5).

Corollary 3. *Let $m(n) = \lfloor n^{1/10} \rfloor$ and $p(n) = n^{-2/5}/10$. This sequence satisfies Assumption 2 of [1], $m(n)p(n) \to 0$, and $m(k^{10}) = k$, $p(k^{10}) = 1/(10k^4)$. Assign arbitrary real values to the ratios G/L and G_{u}/L on infeasible matrices. Then neither ratio converges to 1 in probability as $n \to \infty$.*

Convergence in probability is implied by almost sure convergence and by convergence in mean, so the corollary refutes Conjecture 12 under each of these readings, for the leftmost rule that [1] use and for uniform tie-breaking. The constants are far from optimal; we made no attempt to improve them. We do not know whether the ratio tends to one for some other tie-breaking rule, or for the best of all greedy runs, along this sequence.

1.2 Prior work

Lovász [12] and Johnson [9] analyzed the greedy algorithm for covering. Lovász showed that "the ratio of optimal integral and fractional covers of a hypergraph does not exceed $1 + \log d$" [12, p. 383], with d the maximum degree, as recalled in [1, p. 30:2]. Slavík bounds the greedy value by $\ln m - \ln\ln m + 0.78$ times the optimum when m elements are to be covered [15, Theorem 1] and shows that this is nearly tight [15, Theorem 2]. These are worst-case bounds. For the random model, Corollary 9(ii) of [1] states that if $mp \lesssim \log n$ then $\Pr(\mathrm{IPGAP} \sim 1) \ge 1 - n^{-D}$ for any $D > 0$ and n large, where $\mathrm{IPGAP} = \mathrm{val}_{\mathrm{IP}}/\mathrm{val}_{\mathrm{LP}}$ and $\sim$ means equality up to multiplicative constants;

their Theorem 10 states that under Assumption 2 with $\delta < 1/2$, $\Pr(\mathrm{val}_{\mathrm{Gr}} \sim \mathrm{val}_{\mathrm{IP}}) \geq 1 - \exp(-\sqrt{n})$ for n large. Since $p(n) = n^{-2/5}/10 \geq n^{-\delta}$ for large n whenever $\delta > 2/5$, these results apply to the sequence of Corollary 3: greedy is within a constant factor of the LP value there with high probability. Theorems 1 and 2 show that the constant cannot tend to one.

Telelis and Zissimopoulos [16, Theorem 4, p. 174] show, for the same independent model with fixed p, that a simple $O(nm)$ algorithm approximates random set cover "within a term of $o(\log m)$ from the optimum almost everywhere". Fernandez de la Vega, Paschos and Saad [8] study a greedy algorithm for hitting m random k-element subsets of an n-element set, with k fixed and $m, n \to \infty$ in fixed proportion; their Theorem 1 [8, p. 131] states that its approximation ratio "ranges between 1.094 and 1.257" for $3 \leq k \leq 7$ and a range of values of the proportion. See also the later article of Blot, Fernandez de la Vega, Paschos and Saad [4]. Mézard and Tarzia [14, Sections V.C–V.D and Fig. 5] ran highest-degree greedy on random hypergraphs in which every variable lies in $L = 4$ hyperedges and every hyperedge has $K = 6$ variables, with 12288 variables and 8192 hyperedges, and report greedy covers of density $\rho_{\mathrm{gr}} \simeq 0.212$, against $\rho_{\mathrm{cov}} \simeq 0.178$ for the minimal cover and $1/K$; on such regular instances the LP value is $1/K$ times the number of variables.

The isolated bad paths below are paths of three triples, consecutive triples sharing one row, in a sparse random hypergraph. Greedy matching on sparse random graphs, the setting of Karp and Sipser [10] and of Dyer, Frieze and Pittel [7], is related in spirit, but no matching limit theorem is used here.

We could not read four sources in full: the journal version of Slavík's paper [15], whose abstract and references are those of the conference version cited above; pages 132–138 of [8]; Chebolu's thesis [5], which [3, Section 1] describes as extending the analysis of the random greedy matching process [7] to sparse random hypergraphs; and a Zenodo deposit, *Spectral gaps and algorithmic phase transitions in sparse hypergraphs* (DOI 10.5281/zenodo.17834080), removed before it could be read, whose surviving abstract concerns coloring, satisfiability and community detection.

1.3 Method

On the event that every column hits at most three rows and every pair of rows is exactly the set of some column, greedy takes a family of disjoint triples and then pairs, so its count is determined by the number K of triples it takes (Lemma 4). The obstruction is an *isolated bad path*: three triples $\{a, b, c\}, \{c, d, e\}, \{e, f, g\}$ on seven rows such that no other column with three rows meets these rows; the definition in Section 2 adds that each triple is the set of exactly one column and that the middle column has the smaller index. Leftmost greedy takes the middle triple, after which the two ends are useless to it; but the two ends cover six of the seven rows, and the row d joins the pairs (Lemma 5). The paths are disjoint, so exchanging every such middle at once saves half a column per path, up to one rounding.

Counting expected occurrences (a first-moment count) gives order k such paths in expectation (Lemma 6). Their number is at most m on every matrix, so this lower bound on the expectation gives a fixed fraction of m paths with fixed probability.

For uniform ties, the triples greedy takes have the law of a scan of the triples in a uniformly random order, which is leftmost greedy on a matrix with randomly permuted columns; the distribution of A does not change under this permutation (Section 4).

The theorems and the corollary are proved in Lean 4 [11] with Mathlib [13], from definitions of the random matrix, the greedy algorithms and the LP value (Section 5).

2 Greedy on the good event

Let H be the event that $|A_j| \le 3$ for every column j and that for every two rows $a \ne b$ some column has $A_j = \{a, b\}$. If $m \ge 2$, then H implies that A is feasible. Gains never increase as U shrinks.

Lemma 4. *Assume H and $m \ge 2$, and break ties by any rule. Greedy first chooses K columns of gain 3, whose sets are pairwise disjoint, then columns of gain 2 while at least two rows are uncovered, then at most one column of gain 1. Hence $G = \lceil (m - K)/2 \rceil$, with G the number of choices.*

Proof. Gains are at most 3 and do not increase. While two rows a, b remain uncovered, the column with set $\{a, b\}$ has gain 2, so a choice of gain 1 happens only when one row remains. A column of gain 3 hits three uncovered rows, so its set is disjoint from those of earlier choices. If there are T choices of gain 2 and $S \in \{0, 1\}$ of gain 1, then $m = 3K + 2T + S$, so $S \equiv m - 3K \pmod 2$ and $K + T + S = \lceil (m - K)/2 \rceil$. □

An *isolated bad path* consists of seven distinct rows $a, \dots, g$ and three columns with sets $P = \{a, b, c\}$, $M = \{c, d, e\}$, $Q = \{e, f, g\}$ such that each of these three sets is the set of exactly one column, the column of M has a smaller index than those of P and Q, and every other column with at least three rows avoids $\{a, \dots, g\}$. Columns with at most two rows may meet these rows. Let $B(A)$ be the number of columns that are the middle M of an isolated bad path.

Lemma 5. *Assume every column has at most three rows, and let K be the number of gain-3 choices of leftmost greedy.*

(a) Every middle column of an isolated bad path is a gain-3 choice of leftmost greedy, and every other gain-3 choice avoids the seven rows of that path.

(b) Isolated bad paths with distinct middles have disjoint row sets, and $3(K + B) \le m$.

(c) If moreover H holds and $m \ge 2$, then $L \le \lceil (m - K - B)/2 \rceil$ and $G - L \ge (B - 1)/2$.

Proof. (a) Let $R = \{a, \dots, g\}$. Until greedy first chooses a column meeting R, the middle column has gain 3, so greedy does not stop before such a step. At that step the chosen column has gain 3, the maximum possible, so it has three rows, meets R, and by isolation is one of the three path columns. All three have gain 3 and the middle has the smallest index, so greedy chooses the middle. Afterwards the ends have gain at most 2, and every other column with three rows avoids R.

(b) Let M' be a second middle, with ends P', Q'. By (a), M' avoids R. If P' met R, isolation would make P' one of the columns of the first path, so $P' \subseteq R$; but P' meets M', which avoids R. So P' and likewise Q' avoid R. Now replace each of the B middles among the K gain-3 choices by its two ends. The ends of one path are disjoint, different paths are disjoint, and the remaining gain-3 choices avoid every path by (a), so the resulting $K + B$ triples are pairwise disjoint and $3(K + B) \le m$.

(c) Cover the $m - 3(K + B)$ rows outside these triples in pairs by columns with exactly those two rows, and a possible last row by any column hitting it. The middle-only rows d are among them. This gives a cover with $K + B + \lceil (m - 3K - 3B)/2 \rceil = \lceil (m - K - B)/2 \rceil$ columns, and its indicator vector is feasible for the LP. With Lemma 4, $G - L \ge (m - K)/2 - (m - K - B + 1)/2 = (B - 1)/2$. □

Since one column per row gives $L \le m$, part (c) yields: if H holds, $m \ge 2$, $\alpha \ge 0$, $B \ge \alpha m$ and $\alpha m \ge 2$, then

$$G - L \ge \frac{\alpha m - 1}{2} \ge \frac{\alpha m}{4} \ge \frac{\alpha L}{4}. \tag{2}$$

3 Counting paths

Lemma 6. *With parameters* (1) *and* $k \geq 14$, $\mathbb{E}\, B \geq ck$.

Proof. Call an ordered 7-tuple $t = (a, \ldots, g)$ of distinct rows *occurring* if $\{a, b, c\}, \{c, d, e\}, \{e, f, g\}$ are the sets of an isolated bad path. Each middle column is represented by exactly 8 occurring tuples: its ends are determined by isolation, and the representations are given by the choice of which end is P and the orders of the two private rows of each end. Hence $\sum_t \mathbf{1}[t \text{ occurs}] = 8B$ on every matrix. By Lemma 9 in the appendix, each of the $k(k-1)\cdots(k-6) \geq k^7/128$ tuples occurs with probability at least $1/(192 \cdot 10^9 k^6)$. By linearity, $\mathbb{E}\, B \geq k/(8 \cdot 128 \cdot 192 \cdot 10^9) = ck$. □

Lemma 7. *On every matrix,* $B \leq m$. *With parameters* (1), $\Pr[\text{some } |A_j| \geq 4] \leq 1/(10^4 k^2)$, *and for* $k \geq 2$, $\Pr[\text{some pair of rows is the set of no column}] \leq 160000/k^2$.

Proof. Replace every column with at least four rows by an empty column in the same position. Every isolated bad path survives, since its three columns have three rows and an empty column meets nothing; so B does not decrease, and Lemma 5(b) for the new matrix gives $B \leq m/3$, in particular $B \leq m$, which is all we use. The two probability bounds are union bounds, computed in Appendix A. □

Proof of Theorem 1. Since $B \leq m = k$ by Lemma 7, we have $B \leq ck/2 + k\,\mathbf{1}[B \geq ck/2]$ pointwise. Taking expectations and using Lemma 6, $ck \leq \mathbb{E}\, B \leq ck/2 + k \Pr[B \geq ck/2]$, so $\Pr[B \geq ck/2] \geq c/2$. By Lemma 7, H fails with probability at most $(10^{-4} + 160000)/k^2 \leq c/4$ for $k \geq 10^{20}$. So H and $B \geq ck/2$ hold together with probability at least $c/4$. On this event A is feasible, and (2) with $\alpha = c/2$, valid because $ck/2 \geq 2$, gives $G - L \geq cL/8$. □

4 Uniform tie-breaking

For a permutation π of the columns, let A^π be the matrix with columns $A^\pi_j = A_{\pi(j)}$. The law of A, the event H, feasibility and L are unchanged when A is replaced by A^π.

Lemma 8. *Let* $(T_\ell)_{\ell \in \Lambda}$ *be a family of* 3*-element sets indexed by a finite set* Λ *of labels; different labels may carry the same set. Consider two processes: (X) repeatedly choose, uniformly at random, a label whose set is disjoint from the sets of all labels chosen so far, until none is left; (Y) list* Λ *in a uniformly random order and accept each label whose set is disjoint from the sets of the labels already accepted. The sets of chosen and of accepted labels have the same law.*

Proof. Induct on $|\Lambda|$. In (Y), the first accepted label is the first in the order, uniform on Λ, as in (X). Given that it is r, the remaining order is uniform on $\Lambda \setminus \{r\}$, labels whose sets meet T_r (other labels carrying T_r included) are never accepted, and the order it induces on the set Λ' of labels whose sets are disjoint from T_r is uniform. So both processes continue as the same processes on Λ'. □

Proof of Theorem 2. Assume H and $m \geq 2$. While some column has gain 3, the columns of maximum gain are the columns with three rows disjoint from all earlier choices, so the gain-3 phase of uniform greedy is process (X) with the columns with three rows as labels. Leftmost greedy on A^π is process (Y) with the order given by π, which is uniform when π is. By Lemmas 4 and 8, for every predicate F on integers, the conditional probability given A that $F(G_{\mathrm{u}}(A))$ holds is the

average of $\mathbf{1}[F(G(A^\pi))]$ over all permutations π. Take $F(g)$ to be "$(1+c/8)L(A) \le g$". Averaging over A, using that the law of A, H and L are invariant and substituting A for A^π,

$$\Pr\big[H \text{ and } G_{\mathrm{u}} \ge (1+\tfrac{c}{8})L\big] = \Pr\big[H \text{ and } G \ge (1+\tfrac{c}{8})L\big] \ge \frac{c}{4},$$

where the inequality is the proof of Theorem 1. Since H implies feasibility, the theorem follows, with $c/4 \ge c/8$. □

Proof of Corollary 3. For $n^{1/10} \ge 10$, $\frac{1}{20}n^{1/20} \le \frac{1}{2}n^{1/10} < m(n) \le n$ and $n^{-1/2} \le p(n) \le 1/2$, so Assumption 2 holds with $c_0 = 1/20$, $C_0 = 1$, $\delta = 1/2$; and $m(n)p(n) \le n^{-3/10}/10 \to 0$. On the event of Theorem 1 the matrix is feasible, $L \ge 1$ and $G/L - 1 \ge c/8$, whatever values are assigned on infeasible matrices. Along $n = k^{10}$ this event has probability at least $c/4$, so $\Pr[|G/L-1| \ge c/8]$ does not tend to 0. The same argument with Theorem 2 applies to G_{u}/L. □

5 Formal verification

All of the above is proved in Lean 4 [11] (version 4.19.0) with Mathlib [13], and the build checks that the main theorems use only the axioms `propext`, `Classical.choice` and `Quot.sound`. A matrix is a function giving the set of rows of each column. The probability of an event is the finite sum, over all matrices, of the product over columns of $\prod_i(p \text{ or } 1-p)$. Greedy is a recursive function scanning columns by index; L is the real infimum of the objective over the LP's feasible set, the actual LP value rather than a bound. Uniform ties are a recursive kernel that averages over the columns of maximum positive gain at each step. Theorem 1 is

```
theorem GreedyLPHittingSet.leftmost_explicit : ∀ k ≥ Uniform.threshold,
    (Uniform.c : ℝ) / 4 ≤ matrixProb k (k^10) (parameterP k)
      (fun A => feasible A ∧ (1 + (Uniform.epsilon : ℝ)) * L A ≤ (G A : ℝ))
```

where the build also checks that `Uniform.threshold` is 10^{20}, `Uniform.c` is c, `Uniform.epsilon` is $c/8$ and `parameterP k` is $1/(10k^4)$. The other statements are:

- Theorem 2: `ADG.Uniform.uniform_probability_endpoint`, the same statement with the joint probability of matrix and uniform choices and the bound $c/8$.
- Corollary 3, leftmost ties: `ADG.full_frozen_endpoint`; uniform ties: `ADG.Uniform.full_uniform_endpoint`. Each states Assumption 2 of [1] as printed, the identities $m(k^{10}) = k$ and $p(k^{10}) = 1/(10k^4)$, $m(n)p(n) \to 0$, and, for every assignment of values on infeasible matrices, the failure of convergence in probability to one.

Nothing beyond these definitions and the Lean kernel is assumed; the comparison of the definitions with [1] is the one made in Section 1. Appendix B lists the modules.

To reproduce, run `lake exe cache get` and `lake build` in the `lean` folder accompanying this paper; with the Mathlib cache this takes a few minutes.

Appendix A Probability estimates

Write $\mu(s) = p^{|s|}(1-p)^{m-|s|}$ for the probability that a column has set s; these masses sum to 1. Under (1), $kp \le 1/10$.

Lemma 9. *For $k \ge 14$ and a fixed tuple t of seven distinct rows,* $\Pr[t \text{ occurs}] \ge 1/(192 \cdot 10^9 k^6)$.

Proof. Let $q = p^3(1-p)^{k-3}$ be the mass of each of the three path sets, and F the total mass of the sets s with $|s| \geq 3$ and $s \cap \{a, \dots, g\} \neq \emptyset$; this includes the three path sets. The three path columns occupy three distinct positions with the middle first, $2\binom{n}{3} = n(n-1)(n-2)/3$ choices, and these events are disjoint because each path set occurs exactly once. So

$$\Pr[t \text{ occurs}] = \frac{n(n-1)(n-2)}{3}\, q^3(1-F)^{n-3}.$$

A set counted in F contains one of the seven rows and two others, so $F \leq 7\binom{k-1}{2}p^3$ and $nF \leq 7/1000$. Bernoulli's inequality gives $q \geq p^3(1-kp) \geq p^3/2$ and $(1-F)^{n-3} \geq 1 - nF \geq 1/2$, and $n(n-1)(n-2) \geq n^3/4$ since $n \geq 4$. Hence the probability is at least $n^3p^9/192 = 1/(192 \cdot 10^9 k^6)$. □

Large columns. A column with at least four rows contains one of the $\binom{k}{4}$ four-sets, each with probability p^4, so $\Pr[\text{some } |A_j| \geq 4] \leq n\binom{k}{4}p^4 \leq nk^4p^4 = 1/(10^4k^2)$.

Missing pairs. A column has set exactly $\{a, b\}$ with probability $r = p^2(1-p)^{k-2} \geq p^2/2$. Its absence from all n columns has probability $(1-r)^n$. The binomial term $\binom{n}{2}r^2(1-r)^{n-2}$ is at most 1, so $(1-r)^n \leq 1/(\binom{n}{2}r^2) \leq 16/(n^2p^4)$, using $\binom{n}{2} \geq n^2/4$. Over at most k^2 pairs, the probability is at most $16k^2/(n^2p^4) = 160000/k^2$.

Threshold. For $k \geq 10^{20}$ we have $ck/2 \geq 2$, since $4 \cdot 196608 \cdot 10^9 \leq 10^{20}$, and $(10^{-4} + 160000)/k^2 \leq c/4$, since $4 \cdot 196608 \cdot 10^9(10^{-4} + 160000) \leq 10^{40}$.

Appendix B The Lean modules

`Endpoint` and `SourceEndpoint` contain the definitions and statements; `LP` the basic facts on L, among them that L is at most the size of any cover; `PathDefinitions` the definition of isolated bad paths and of B; `Probability`, `PathProbability` and `QuantitativeProbability` the finite probability calculations; `Greedy` Lemma 4; `GreedyPaths`, `DeterministicExchange` and `DeterministicGap` Lemma 5 and (2); `PathCounting` and `ExpectationConstruction` Lemma 6; `BadMiddleBound` and `ExpectationTail` Lemma 7 and the tail bound; `FullConstruction`, `SourceTransport` and `SourceConsequence` Theorem 1 and Corollary 3. For uniform ties, `UniformEndpoint` defines the kernel and statements; `UniformKernel` shows that the kernel gives probability laws; `ConflictCoupling` and `PermutationScan` prove Lemma 8 for an arbitrary directed conflict relation; `UniformPhase`, `OrderedPhase` and `UniformCount` identify the gain-3 phases; `UniformInvariance`, `UniformTransport` and `UniformGoodLaw` give the permutation invariance; `UniformConstruction` proves Theorem 2. `Statement` restates the theorems with their definitions unfolded, `Examples` checks small matrices (a single path, where greedy and the exchange both use three columns, and two paths, where greedy uses six and the exchange five), and `Axioms` fails the build if a main theorem uses another axiom.

References

[1] G. Arpino, D. Dmitriev and N. Grometto, *Greedy heuristics and linear relaxations for the random hitting set problem*, in: Approximation, Randomization, and Combinatorial Optimization (APPROX/RANDOM 2024), LIPIcs 317, Paper 30, 2024. doi:10.4230/LIPIcs.APPROX/RANDOM.2024.30.

[2] G. Arpino, D. Dmitriev and N. Grometto, *Greedy heuristics and linear relaxations for the random hitting set problem*, arXiv:2305.05565v1, 2023.

[3] D. Bal and P. Bennett, *The matching process and independent process in random regular graphs and hypergraphs*, Electron. J. Combin. 30(1) (2023), P1.11. doi:10.37236/10698.

[4] J. Blot, W. Fernandez de la Vega, V. Th. Paschos and R. Saad, *Average case analysis of greedy algorithms for optimisation problems on set systems*, Theoret. Comput. Sci. 147 (1995), 267–298. doi:10.1016/0304-3975(95)00242-O.

[5] P. Chebolu, *Topics in random graphs*, PhD thesis, Carnegie Mellon University, 2008.

[6] D. Dmitriev, *High-dimensional optimization problems in average-case and worst-case settings*, PhD thesis, ETH Zürich, 2025. doi:10.3929/ethz-b-000743377.

[7] M. Dyer, A. Frieze and B. Pittel, *The average performance of the greedy matching algorithm*, Ann. Appl. Probab. 3 (1993), 526–552. doi:10.1214/aoap/1177005436.

[8] W. Fernandez de la Vega, V. Th. Paschos and R. Saad, *Average case analysis of a greedy algorithm for the minimum hitting set problem*, in: LATIN '92, Lecture Notes in Comput. Sci. 583, Springer, 1992, 130–138. doi:10.1007/BFb0023824.

[9] D. S. Johnson, *Approximation algorithms for combinatorial problems*, J. Comput. System Sci. 9 (1974), 256–278. doi:10.1016/S0022-0000(74)80044-9.

[10] R. M. Karp and M. Sipser, *Maximum matching in sparse random graphs*, in: 22nd Annual Symposium on Foundations of Computer Science (1981), 364–375. doi:10.1109/SFCS.1981.21.

[11] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: Automated Deduction – CADE 28, Lecture Notes in Comput. Sci. 12699, Springer, 2021, 625–635. doi:10.1007/978-3-030-79876-5_37.

[12] L. Lovász, *On the ratio of optimal integral and fractional covers*, Discrete Math. 13 (1975), 383–390. doi:10.1016/0012-365X(75)90058-8.

[13] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), 367–381. doi:10.1145/3372885.3373824.

[14] M. Mézard and M. Tarzia, *Statistical mechanics of the hitting set problem*, Phys. Rev. E 76 (2007), 041124. doi:10.1103/PhysRevE.76.041124; arXiv:0707.0189v2.

[15] P. Slavík, *A tight analysis of the greedy algorithm for set cover*, in: Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC 1996), 435–441. doi:10.1145/237814.237991. Journal version: J. Algorithms 25 (1997), 237–254, doi:10.1006/jagm.1997.0887.

[16] O. A. Telelis and V. Zissimopoulos, *Absolute $o(\log m)$ error in approximating random set covering: an average case analysis*, Inform. Process. Lett. 94 (2005), 171–177. doi:10.1016/j.ipl.2005.02.009.