Two-colorability of random non-uniform hypergraphs

(Extended abstract)

Jakub Kozik [*][†]   Grzegorz Ryn [‡]   Toullec Streicher Ronan [§]

Abstract

Two-colorability of random hypergraphs has been extensively studied in the literature for over half a century. Most of the efforts were devoted to studying uniform hypergraphs (i.e. with all the edges of the same size). Random models played significant role in this research. Our main motivation was to find out how some of the well-known results on random sparse uniform hypergraphs can be translated to the non-uniform case. We have examined both algorithmic and non-constructive bounds for 2-colorability thresholds for certain natural models of random nonuniform hypergraphs. In both of these problems, we have managed to find sufficient conditions for 2-colorability that aggregate information about different sizes of edges. Additionally, in the case of algorithmic lower bounds, we discovered an interesting behaviour when the discrepancy of the sizes of edges can be used to improve the natural generalization of the uniform bounds.

1   Introduction

Hypergraph $\mathcal{H} = (V, E)$ is a pair consisting of a set of *vertices* V and a family of sets of vertices $E \subset \mathcal{P}(V)$, called *(hyper-)edges*. Hypergraph is *k-uniform* if all its edges are of size k. Such hypergraphs are also called *k-graphs*. Just like hypergraphs can be considered as generalizations of ordinary graphs, their coloring problems can be seen as straightforward analogs of graph coloring problems. For a hypergraph $\mathcal{H} = (V, E)$, vertex coloring $c : V \to \mathbb{N}$ is *proper* if every edge contains at least two vertices x, y such that $c(x) \neq c(y)$ (in other words no edge is *monochromatic*). Two colors are enough to make this notion interesting – as observed by Lovász [1] the computational problem of recognizing 2-colorable hypergraphs is NP-complete.

An interesting perspective on hypergraph 2-coloring comes from considering it as a constraint satisfaction problem. Apparent similarity to SAT allows for a translation of a number of

---

[*]Theoretical Computer Science, Jagiellonian University, Kraków, Poland. E-mail: `jakub.kozik@uj.edu.pl`.
[‡]Theoretical Computer Science, Jagiellonian University, Kraków, Poland. E-mail: `grzegorz.ryn@student.uj.edu.pl`.

[§]École normale supérieure de Lyon, Lyon, France. E-mail: `ronan.toullec_streicher@ens-lyon.fr`.

results to the framework of hypergraph coloring. Significant difference reveals itself when one tries to construct a small no-instance. In [2] Erdős and Hajnal asked about the minimum number of edges in a k-graph which is not 2-colorable. Shortly after that Erdős came up with asymptotic lower and upper bounds. In [3], he showed that for large enough k, any k-uniform hypergraph with at most $(1 - \varepsilon)2^k \ln 2$ edges is 2-colorable. A year later, in [4] he showed with a probabilistic argument that there exist a non-2-colorable hypergraph with $k^2 \cdot 2^{k+1}$ edges. This is one of the textbook examples of the situation when randomization yields a construction which is significantly smaller than any known deterministic examples. Constraint satisfaction perspective also inspired parallel line of research where 2-colorability is studied for random hypergraphs with fixed size of edges but the number of vertices tending to infinity. Erdős' upper bound literally translate to this framework yielding Theorem 1.1 below. See Alon and Spencer [5] for the adapted proof. By a direct analogy to the well-known random graph model $G(n, p)$, we denote by $H_k(n, p)$ the probabilistic space of random k-graphs on the set V of n vertices, given by including each k-subset of V to the set of edges independently with probability p.

Theorem 1.1 (uniform upper bound [5]).
For $p(n) \geq \frac{\ln 2}{2} \cdot 2^k \cdot n$, random hypergraph $H_k(n, p(n))$ is a.a.s. not 2-colorable.

The evolution of lower bounds has been more convoluted. In [5] Alon and Spencer presented an algorithm that almost surely produces a proper 2-coloring for $H_k(n, p)$ whenever the expected number of edges in that hypergraph does not exceed $c \cdot (2^k/k^2) \cdot n$. The next significant progress came in 2001 with Achlioptas, Kim, Krivelevich, Tetali [6] publishing an algorithm that almost surely 2-colors random k-graphs $H_k(n, p)$ with at most $c \cdot (2^k/k) \cdot n$ edges on average. Finally lower bound for 2-colorability has been improved in 2002 by Achlioptas and Moore [7] to almost asymptotically match the upper one. This argument however is not constructive.

Theorem 1.2 (uniform 2-colorability lower bound [7]).
Let $\varepsilon > 0$ and $k \geq k_0(\varepsilon)$. Then for $p(n)$ such that the expected number of edges in a random hypergraph is $\left(\frac{\ln 2}{2} \cdot \left(2^k - 1\right) - \frac{1+\varepsilon}{2}\right) \cdot n$, random hypergraph $H_k(n, p(n))$ is a.a.s. 2-colorable.

Both, the algorithm from [6] and the nonconstructive lower bound from [7] are going to be our reference points for the developments on non-uniform hypergraphs.

It's worth noting that [6] provides the best currently known algorithmic lower bound for 2-colorability. Moreover, there are reasons to believe that no *significant* improvement is possible. In 2008 Achlioptas and Coja-Oghlan [8] published a paper in which they analyze how the space (which is essentially a subset of the Hamming hypercube) of proper 2-colorings changes as probability of an edge increases. They have showed that for random k-uniform hypergraphs with the expected number of edges between $(1+\varepsilon_k)\frac{2^{k-1}}{k} \ln k \cdot n$ and $(1-\varepsilon_k)2^{k-1} \ln 2 \cdot n$ the space of the solutions is *shattered*. This means that the space of proper 2-colorings becomes disconnected into *many* small components that are *very far* apart of each other wrt the Hamming distance. In this and a number of related problems the occurrence of this phenomenon coincides with the point where the best known algorithms break. Moreover this kind of behaviour has been formally proven to be a barrier for certain classes of algorithms [9].

1.1 Non-uniform hypergraphs

A hypergraph is called a k^+-graph if all its edges are of size at least k. Extending the described results to the case of k^+-graphs is often a challenging endeavor. The basic question of Erdős

and Hajnal (about the minimum number of edges in a non-2-colorable k-graph) has been re-formulated for non-uniform hypergraphs by Erdős and Lovász in [10] as follows: what is the minimum number $f(k)$ for which any k^+-graph with the expected number of monochromatic edges[1] not exceeding $f(k)$ is 2-colorable. Erdős and Lovász conjectured that such defined function is unbounded. It has been confirmed by Beck in 1978 [11], who provided a lower bound of the order $\log^*(k)$. The best currently known lower bound for $f(k)$ from [12] is of the order $\log(k)$. Note that the corresponding bound for k-graphs from [13] is roughly $\sqrt{k/\log(k)}$. While it seems harder to color non-uniform hypergraphs, no one so far managed to exploit non-uniformity in the construction of non-2-colorable hypergraphs. In fact, the best upper bound for $f(k)$ is still $\Theta(k^2)$ which results from the Erdős' upper bound for the uniform case.

Note that $f(k)$ can be equivalently defined as the sum of weights of edges, where an edge of size k gets weight 2^{-k+1}. That gives us a clue about how to translate the statements of the theorems to the non-uniform world.

2 Main results

We are going to work with the following model of a random non-uniform hypergraphs. For a function $\mathcal{M} : \mathbb{N} \to [0,1]$ let random hypergraph $H(n; \mathcal{M})$ have set of vertices $V = \{1, \ldots, n\}$ and let the (random) set of edges be constructed by including every subset of $S \subset V$ independently with probability $\mathcal{M}(|S|)$. The support of $\mathcal{M}$, denoted by $supp(\mathcal{M})$, is the set of numbers k for which $\mathcal{M}(k) \neq 0$. By an analogy to k being fixed in the case of uniform hypergraphs, and in order to avoid discussing certain technical issues, we consider only functions $\mathcal{M}$ with finite support. This is implicitly assummed in the statements of the theorems. We are mainly interested in asymptotic results, when the sizes of edges are bounded and the number of vertices tends to infinity. Then, the assymptotic statements about random hypergraphs have to involve working with the sequence of random hypergraphs. For the problems of our interest, the edge probabilities in such sequences are tending to 0 with n. By the common convention of the area we do not explicitly represent the dependence of the probabilities in $\mathcal{M}$ on n. The main theorems are formulated in a way that implicitely defines the probabilities for every n by specifying the expected number of edges of given size.

2.1 Algorithmic 2-coloring of non-uniform hypergraphs

We start by presenting the algorithm of Achlioptas, Kim, Krivelevich, Tetali from [6] (see listing below). For convenience we assume that the number of vertices n is even. Given some partial coloring[2] an edge of size k is called an *l-tail* if $(k - l)$ of its vertices are colored and all of them are assigned the same color.
The main result of [6] can be now restated as follows: for p such that $p \cdot \binom{n}{k} \leq \lambda \cdot \frac{2^k}{k} \cdot n$, AKKT algorithm a.a.s. properly colors $H_k(n, p)$[3].

Observe that this algorithm essentially does not care about any edge till it becomes a 3-tail. Moreover, when an edge becomes a 3-tail, it does not matter how many of its vertices are already colored (i.e. what is the size of that edge). Hence, it is natural to consider its behaviour

---

[1]in a uniformly random 2-coloring of the vertices
[2]throughout this paper the colorings will be done with RED and BLUE.
[3]For $k \geq 10$, $\lambda \leq \frac{1}{50}$ is enough. For $k \geq 40$, it can be improved to $\lambda \leq \frac{1}{10}$.

Algorithm 1 AKKT algorithm

1: **do** $n/2$ times
2:    **if** there exists some 3-tail or 2-tail e **then**
3:        pick any two uncolored vertices $x, y \in e$ and color x to RED and y to BLUE.
4:    **else**
5:        pick any two uncolored vertices x, y and color x to RED and y to BLUE.
6:    **end if**
7: **end do**
---

for non-uniform hypergraphs.

An important property of this procedure is that every considered partial coloring is equitable (i.e. the number of RED vertices equals the number of BLUE ones). Moreover, the coloring decisions depend only on the currently observed 3-tails and 2-tails. In particular the potential edges with a larger number of not yet colored vertices are irrelevant at that point and do not even have to be inspected for being an actual edges of the underlying hypergraph. As a result, for every $i \in \{0, \ldots, n/2 - 1\}$, the distribution of new 3-tails produced by coloring selected vertices in the i-th step does not depend on the previously chosen vertices. Then it is easy to conclude that the new 3-tails are in fact distributed as the edges of a random 3-graph on the remaining vertices with specific probability of including an edge at the i-th step $p(i)$. (In case of k-graphs the value of $p(i)$ depends only on the number of remaining vertices and the original p.) We can therefore abstract from the underlying (random) instance and focus on the probability that the algorithm is successful as a function of the sequence of probabilities $(p(i))_{i=0,\ldots,n/2-1}$. Most of the arguments from the original analysis of this procedure in [6] are valid in such extended framework.

In the uniform case, the algorithm succeeds on graphs with $\lambda \cdot (2^k/k) \cdot n$ edges on average. That suggests assigning weights $(2^k/k)$ to edges of size k. Let $\lambda_{k_1}, \lambda_{k_2}, \ldots, \lambda_{k_l}$ be a sequence of positive real numbers and $H(n; \mathcal{M})$ be the random hypergraph model with $supp(\mathcal{M}) = \{k_1, \ldots, k_l\}$, such that the expected number of hyperedges of arity k_i is $\lambda_{k_i} \cdot (2^{k_i}/k_i) \cdot n$. The results of [6] imply that if $\lambda_{k_i} \geq 1$ for some i, then the edges of arity k_i are enough to make the algorithm fail. Natural generalization of the result for k-graphs, would be to show that whenever the average weighted sum of edges is bounded by $\lambda \cdot n$, the procedure is likely to succeed (for the same value of λ as used in [6]). Indeed, once the (random) instance has been abstracted to the sequence of probabilities $(p(i))_{i=0,\ldots,n/2-1}$, the tools of [6] can be used in a straightforward manner to deduce such result – if

$$\sum_{k \in supp(\mathcal{M})} \lambda_k < \lambda, \tag{1}$$

then the procedure a.a.s. succeeds. Interestingly, this natural condition in some cases is a serious overkill. E.g. when we consider only two sizes of edges and they differ significantly (e.g. edge sizes k and k' satisfy $k' = k^4$), there are no steps for which both λ_k and $\lambda_{k'}$ contribute significantly to $p(i)$. In such a case the success condition would be closer to the maximum of λ_k and $\lambda_{k'}$ being bounded by λ. Taking into account this kind of behaviour we have shown the following general condition for the random hypergraph to be a.a.s. 2-colorable by AKKT procedure.

Theorem 2.1 (AKKT nonuniform sufficient condition).
If in every phase of AKKT algorithm run on the random hypergraph, the expected number of new 3-tails is bounded by $1 - \varepsilon$ for $\varepsilon > 0$, then a.a.s. the algorithm succeeds in finding a proper 2-coloring.

This condition, although not directly related to the specific parametrization (i.e., the expected number of hyperedges of each size), can be used to derive the key properties used in [6] to show that AKKT succeeds. That enable easy adaptation to the non-uniform setting. We present below a more technical version of this condition, by expressing the expected number of new 3-tails using the parametrization of our random hypergraph model. Fourth derivative and exponent 3 in the formula are the consequence of focusing on 3-tails.

Theorem 2.2 (AKKT nonuniform aggregating condition).
Let $H(n; \mathcal{M})$ be a random hypergraph and let $(\lambda_k)_{k \in supp(\mathcal{M})}$ be such that the expected number of edges of each size k in $H(n; \mathcal{M})$ equals $\lambda_k \cdot \frac{2^k}{k} \cdot n$. Let

$$f(x) := \sum_{k \in supp(\mathcal{M})} \frac{\lambda_k}{k} \cdot x^k$$

Then, for any $\varepsilon > 0$ and $\min(supp(\mathcal{M}))$ large enough, the following condition is sufficient for the AKKT procedure to produce a proper 2-coloring almost surely

$$\max_{x \in [0,1]} \left(\frac{16}{3}(1-x)^3 f^{(4)}(x) \right) < 1 - \varepsilon,$$

where $f^{(4)}$ is the fourth derivative of f.

It is not immediately clear how to compare the technical condition above with the natural generalization (1). In order to illustrate the difference we constructed a family of the sequences of edge sizes and corresponding lambdas s.t. their sum can be arbitrarily large.

Theorem 2.3 (Unbounded sum of lambdas).
For every $c > 0$, there exists a random hypergraph $H(n; \mathcal{M})$ with corresponding sequence $(\lambda_k)_{k \in supp(\mathcal{M})}$ such that the expected number of edges of each size k in $H(n; \mathcal{M})$ equals $\lambda_k \cdot \frac{2^k}{k} \cdot n$, for which

$$\sum_{k \in supp(\mathcal{M})} \lambda_k \geq c$$

and AKKT procedure succeeds a.a.s.

As we can see, 2-colorable non-uniform random hypergraphs can have much more[4] edges on the average than the uniform ones as long as their sizes are sufficiently spread.

2.2 2-coloring threshold for non-uniform hypergraphs

It is already known, that the 2-colorability threshold for k-graphs is around the values of $p^*(n)$ for which the expected number of edges is $\lambda_k^* \cdot 2^k \cdot n$, for some λ_k^* that is close to $\ln 2/2$

---

[4]in the sense of the sum of lambdas

but whose precise value has not been determined yet[5]. For $\lambda_k \geq \ln 2/2$ the corresponding random hypergraph is a.a.s. not-2-colorable, and by the results of [7] and later improvements [14, 15], for slightly smaller λ_k it is eventually very likely to be 2-colorable. That justifies parametrization of the average number of edges as $\lambda_k \cdot 2^k \cdot n$ and focusing on how the values of λ_k affect 2-colorability. In a nonuniform case, for a sequence $\lambda_{k_1}, \lambda_{k_2}, \ldots, \lambda_{k_l}$ of positive real numbers, we are going to work with corresponding random hypergraph $H(n; \mathcal{M})$ in which the expected number of edges of size k_i is $\lambda_{k_i} \cdot 2^{k_i} \cdot n$. Within this setting, the proof of the upper bound for k-graphs generalizes in a straightforward way.

Theorem 2.4 (non-uniform upper bound)**.**
Let $H(n; \mathcal{M})$ be a random hypergraph model with corresponding sequence $(\lambda_k)_{k \in supp(\mathcal{M})}$ for which, for every $k \in supp(\mathcal{M})$, the expected number of edges of size k equals $\lambda_k \cdot 2^k \cdot n$. Suppose that $\sum_{k \in supp(\mathcal{M})} \lambda_k > \frac{\ln 2}{2}$, then

$$\lim_{n \to \infty} \mathcal{P}(H(n; \mathcal{M}) \text{ is 2-colorable}) = 0.$$

We immediately observe that, unlike in the case of the algorithmic lower bound, the sizes of edges are irrelevant provided they are compensated by probability. In other words it is really the sum of λ_k (not the maximum) that makes the (random) hypergraph not 2-colorable. The non-uniform structure which we have exploited with AKKT appears to be completely insignificant around the threshold. As explained below, it is not only a feature of this particular argument, since we observe an analogous behavior in the analysis of nonconstructive lower bounds.

The first lower bound for the uniform case of the order $2^k \cdot n$ was obtained in [7] by the second moment method. Later is has been observed (see e.g. [14]) that this proof can be significantly simplified by focusing on equitable colorings. Note that once again, the equitable colorings announce their prominent role. Second moment argument is used to prove that, for certain range of the values of p, the probability that corresponding random k-graph is 2-colorable is bounded away from 0. Then, by the general properties of the thresholds of monotonic properties developed by Friedgut [16], it can be concluded that for every slightly smaller value of p corresponding random hypergraph is a.a.s. 2-colorable. The proof that 2-colorability of hypergraphs exhibits a sharp (non-uniform) threshold can be found in [17].

With a bit of technical effort we managed to show that non-uniform hypergraphs are 2-colorable with positive probability just before the non-2-colorability bound of Theorem 2.4. Our proof is also based on the second moment method, and just like [14] focuses on equitable 2-colorings. The obtained bound for the sum of λ_k is weaker than the known corresponding bounds for the uniform case. However, the ideas that allowed for the improvements of the uniform case are likely to be adapted to the non-uniform framework. The technical cost of doing so seems to be considerable.

Theorem 2.5 (non-uniform 2-colorability threshold)**.**
Let $\varepsilon > 0$ and $H(n; \mathcal{M})$ be a random hypergraph model with large enough $\min(supp(\mathcal{M}))$ and corresponding sequence λ_k for $k \in supp(\mathcal{M})$ such that the expected number of edges of size k equals $\lambda_k \cdot 2^k \cdot n$.

---

[5]To be precise it is also possible that the threshold is nonuniform, i.e. $\lambda_k^* = \lambda_k^*(n)$ and it does not converge with $n \to \infty$.

Suppose that $\sum_{k \in supp(\mathcal{M})} \frac{\lambda_k}{1-2^{1-k}} < \frac{\ln 2}{2}(1 - \varepsilon)$, then

$$\liminf_{n \to \infty} \mathcal{P}(H(n; \mathcal{M}) \text{ is 2-colorable}) > 0.$$

The thing stopping us from finishing the proof and showing that such hypergraphs are 2-colorable almost surely is the lack of tools for handling sharp thresholds in the non-uniform case. We explored more recent generalizations of these methods from [18] by Hatami and Molloy, and tried to embed our non-uniform model in their (much more general than the original) random CSP model. We did not succeed and eventually convinced ourselves that stronger tools are necessary. We are currently working on the multi-dimensional analogue of the Friedgut's results from [16].

3   Conclusion

A number of results concerning the limits of 2-colorability of large (and sparse) random k-graphs can be carried out to non-uniform framework. While the methods for non-constructive upper and lower bounds turns out to be rather oblivious to the distribution of edge sizes, specific configurations can be exploited in the proofs of the algorithmic (but weaker) lower bounds. It is particularly interesting to study the behavior of such coloring procedures from the point of view of algorithmic barriers (as defined in [8]).

It is expected that 2-colorability for non-uniform models exhibit (possibly non-uniform) sharp-threshold behavior. That would allow for strengthening our non-constructive lower bound from Theorem 2.5 and show that for slightly smaller sums of lambdas, the limit is not merely positive but equals 1. Although nonuniformity do not seem to introduce any essentially new phenomena when the sharp-thresholds are considered, the common tools of this area need to be sharpened by a multi-dimensional analysis before they can be applied to nonuniform hypergraphs.

[1] L. Lovász. Coverings and colorings of hypergraphs. *Proc. 4th Southeastern Conference of Combinatorics, Graph Theory, and Computing*, pages 3–12, 1973.

[2] Paul Erdős and András Hajnal. On a property of families of sets. *Acta Mathematica Academiae Scientiarum Hungaricae*, 12:87–123, 1961.

[3] Paul Erdős. On a combinatorial problem. *Nordisk Matematisk Tidskrift*, 11:5–10, 40, 1963.

[4] Paul Erdős. On a combinatorial problem. II. *Acta Mathematica Academiae Scientiarum Hungaricae*, 15:445–447, 1964.

[5] Noga Alon and Joel H. Spencer. A note on coloring random k-sets.

[6] Dimitris Achlioptas, Jeong Han Kim, Michael Krivelevich, and Prasad Tetali. Two-coloring random hypergraphs. *Random Structures Algorithms*, 20(2):249–259, 2002.

[7] Dimitris Achlioptas and Cristopher Moore. On the 2-colorability of random hypergraphs. In *Randomization and approximation techniques in computer science*, volume 2483 of *Lecture Notes in Comput. Sci.*, pages 78–90. Springer, Berlin, 2002.

[8] Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '08, pages 793–802, Washington, DC, USA, 2008. IEEE Computer Society.

[9] David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41):e2108492118, 2021.

[10] Paul Erdős and László Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II*, volume 10 of *Colloquia Mathematica Societatis János Bolyai*, pages 609–627. North-Holland, Amsterdam, 1975.

[11] József Beck. On 3-chromatic hypergraphs. *Discrete Mathematics*, 24(2):127–137, 1978.

[12] Lech Duraj, Grzegorz Gutowski, and Jakub Kozik. A Note on Two-Colorability of Nonuniform Hypergraphs. In Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella, editors, *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*, volume 107 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 46:1–46:13, Dagstuhl, Germany, 2018. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[13] Jaikumar Radhakrishnan and Aravind Srinivasan. Improved bounds and algorithms for hypergraph 2-coloring. *Random Structures & Algorithms*, 16(1):4–32, 2000.

[14] Martin Dyer, Alan Frieze, and Catherine Greenhill. On the chromatic number of a random hypergraph. *J. Combin. Theory Ser. B*, 113:68–122, 2015.

[15] Peter Ayre, Amin Coja-Oghlan, and Catherine Greenhill. Hypergraph coloring up to condensation. *Random Structures & Algorithms*, 54(4):615–652, 2019.

[16] Ehud Friedgut and Jean Bourgain. Sharp thresholds of graph properties, and the k-sat problem. *Journal of the American Mathematical Society*, 12(4):1017–1054, 1999.

[17] Ehud Friedgut. Hunting for sharp thresholds. *Random Struct. Algorithms*, 26(1–2):37–51, January 2005.

[18] Michael Molloy Hamed Hatami. Sharp thresholds for constraint satisfaction problems and homomorphisms. *Random Structures & Algorithms*, 33(3):310–332, 2008.