# Isomorphism for Random $k$-Uniform Hypergraphs

Debsoumya Chakraborti[*], Alan Frieze[†], Simi Haber[‡], Mihir Hasabnis[§]

February 16, 2020

**Abstract**

We study the isomorphism problem for random hypergraphs. We show that it is polynomially time solvable for the binomial random $k$-uniform hypergraph $H_{n,p;k}$, for a wide range of $p$. We also show that it is solvable w.h.p. for random $r$-regular, $k$-uniform hypergraphs $H_{n,r;k}, r = O(1)$.

## 1 Introduction

In this note we study the isomorphism problem for two models of random $k$-uniform hypergraphs, $k \geq 3$. A hypergraph is $k$-uniform if all of its edges

[*]Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA15213, USA

[†]Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA15213, USA, Research supported in part by NSF grant DMS1661063

[‡]Department of Mathematics, Bar-Ilan University. This research was partially supported by the BIU Center for Research in Applied Cryptography and Cyber Security in conjunction with the Israel National Directorate in the Prime Ministers office.

[§]Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA15213, USA

are of size $k$. The graph isomporphism problem for random graphs is well understood and in this note we extend some of the ideas to hypergraphs.

The first paper to study graph isomorphism in this context was that of Babai, Erdős and Selkow [12]. They considered the model $G_{n,p}$ where $p$ is a constant independent of $n$. They showed that w.h.p.[1] $G = G_{n,p}$ has a *canonical labelling* and that this labelling can be constructed in $O(n^2)$ time. In a canonical labelling we assign a unique label to each vertex of a graph such that labels are invariant under isomorphism. It follows that two graphs with the same vertex set are isomorphic, if and only if the labels coincide. (This includes the case where one graph has a unique labeling and the other does not. In which case the two graphs are not isomorphic.) The failure probability for their algorithm was bounded by $O(n^{-1/7})$. Karp [9], Lipton [11] and Babai and Kucera [3] reduced the failure probability to $O(c^n), c < 1$. These papers consider $p$ to be constant and the paper of Czajka and Pandurangan [13] allows $p = p(n) = o(1)$. We use the following result from [13]: the notation $A_n \gg B_n$ means that $A_n/B_n \to \infty$ as $n \to \infty$.

**Theorem 1.** *Suppose that $p \gg \frac{\log^4 n}{n \log\log n}$ and $p \leq \frac{1}{2}$. Then $G_{n,p}$ has a canonical labeling q.s.*[2]

Our first result concerns the random hypergraph $H_{n,p;k}$, the random $k$-uniform hypergraph on vertex set $[n]$ in which each of the possible edges $\binom{[n]}{k}$ occurs independently with probability $p$. We say that two $k$-uniform hypergraphs $H_1, H_2$ are isomorphic if there is a bijection $f : V(H_1) \to V(H_2)$ such that $\{x_1, x_2, \ldots, x_k\}$ is an edge of $H_1$ if and only if $\{f(x_1), f(x_2), \ldots, f(x_k)\}$ is an edge of $H_2$. We extend the notion of canonical labelling to hypergraphs.

**Theorem 2.** *Suppose that $k \geq 3$ and $p, 1-p \gg n^{-(k-2)} \log n$ then $H_{n,p;k}$ has a canonical labeling w.h.p.*

Bollobás [1] and Kucera [10] proved that random regular graphs have canonical labelings wh.p. We extend the argument of [1] to regular hypergraphs.

[1] A sequence of events $\mathcal{E}_n, n \geq 1$ occurs with high probability (w.h.p.) if $\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_n) = 1$.

[2] A sequence of events $\mathcal{E}_n, n \geq 1$ occurs quite surely (q.s.) if $\mathbb{P}(\mathcal{E}_n) = O(n^{-K})$ for any positive constant $K$.

A hypergraph is regular of degree $r$ if every vertex is in exactly $r$ edges. We denote a random $r$-regular, $k$-uniform hypergraph on vertex set $[n]$ by $H_{n,r;k}$.

**Theorem 3.** *$H_{n,r;k}$ has a canonical labeling w.h.p.*

## 2 Proof of Theorem 2

Given $H = H_{n,p;k}$ we let $H_i$ denote the $(k-1)$-uniform hypergraph with vertex set $[n] \setminus \{i\}$ and edges $\{e \in E(H) : \ i \in e\}$. Let $\mathcal{E}_k$ denote the event $\{\not\exists i, j : H_i \cong H_j\}$.

**Lemma 4.** *Suppose that $k \geq 3$ and $\omega \to \infty$ and $p, 1-p \geq \omega n^{-(k-2)} \log n$. Then $\mathcal{E}_k$ occurs q.s.*

*Proof.*

$$\begin{aligned}\mathbb{P}(\exists i, j : H_i \cong H_j) &\leq \binom{n}{2}(n-1)!(p^2+(1-p)^2)^{\binom{n-2}{k-1}} \\ &\leq \sqrt{2\pi} n^{5/2}\left(\frac{n}{e}\right)^n (p^2+(1-p)^2)^{\binom{n-2}{k-1}p} \\ &\leq n^{-\omega/k!}.\end{aligned}$$

**Explanation:** There are $(n-1)!$ possible isomorphisms and for every $(k-1)$-set of vertices $S$ that includes neither $i$ nor $j$, the probability for there to be an edge or non-edge in both $H_i$ and $H_j$ is given by the expression $p^2+(1-p)^2$. $\square$

Let $\mathcal{G}_k$ be the event that $H_{n,p;k}$ has a canonical labeling and that it can be constructed in $O(n^{2k})$ time. Now assume inductively that

$$\mathbb{P}(H_{n,p,k} \notin \mathcal{G}_k) \leq n^{-\omega/(k+1)!}. \tag{1}$$

The base case, $k = 2$, for (1) are given by the results of [13], [9] and [11]. Let $H_i', i = 1, 2, \ldots, n$ denote the $(k-1)$-uniform hypergraphs induced by the edges of $H'$ that contain $i$, (the *link* associated with vertex $i$). Let $\mathcal{B}_i$ be the event that $H_i \notin \mathcal{G}_{k-1}$. Then

$$\mathbb{P}(H_{n,p,k} \notin \mathcal{G}_k) \leq \mathbb{P}(\mathcal{E}_k) + \sum_{i=1}^{n} \mathbb{P}(\mathcal{B}_i). \tag{2}$$

Indeed, if none of the events in (2) occur then in time $O(n^2 \times n^{2(k-1)}) = O(n^{2k})$ we can by induction uniquely label each vertex via the canonical labeling of its link. After this we can confirm that $\mathcal{E}_k$ has occurred. This confirms the claimed time complexity. Given that $\mathcal{E}_k$ does not occur, this will determine the only possible isomorphism.

Going back to (2) we see by induction that

$$\mathbb{P}(H_{n,p,k} \notin \mathcal{G}_k) \leq n^{-\omega/k!} + n^2 \times (k-1)n^{2k-2}n^{-\omega/k!} \leq n^{-\omega/(k+1)!}.$$

This completes the proof of Theorem 2.

# 3 Proof of Theorem 3

We extend the analysis of Bollobás [1] to hypergraphs. We use the configureation model for hypergraphs, which is a simple generalisation of the model in Bollobás [2]. We let $W$ be a set of size $rn$ where $m = rn/k$ is an integer. Assume that it is partitioned into sets $W_1, W_2, \ldots, W_n$ of size $r$. We define $f : W \to [n]$ by $f(w) = i$ if $w \in W_i$. A configuration $F$ is a partition of $W$ into sets $F_1, F_2, \ldots, F_m$ of size $k$. Given $F$ we obtain the (multi)hypergraph $\gamma(F)$ where $F_i = \{w_1, w_2, \ldots, w_k\}$ gives rise to the edge $\{f(w_1), f(w_2), \ldots, f(w_k)\}$ for $i = 1, 2, \ldots, m$. It is known that if $\gamma(F)$ has a graph property w.h.p. then $H_{n,r;k}$ will also have this property w.h.p., see for example [4]. Let

$$\rho = (r-1)(k-1).$$

For a vertex $v$ we let $d_\ell(v)$ denote the number of vertices at hypergraph distance $\ell$ from $v$ in $H_{n,r;k}$. We show that if $\ell^* = \left\lceil \frac{3}{5} \log_\rho n \right\rceil$ then w.h.p. no two vertices have the same sequence $(d_\ell(v), \ell = 1, 2, \ldots, \ell^*)$. In the following $H = H_{n,r;k}$. For a set $S \subseteq [n]$, we let $e_H(S)$ denote the number of edges of $H$ that are contained in $S$.

**Lemma 5.**
*Let $\ell_0 = \left\lceil 100 \log_\rho \log n \right\rceil$. Then w.h.p., $e_H(S) \leq \frac{|S|}{k-1}$ for all $S \subseteq [n], |S| \leq 2\ell_0$.*

*Proof.* We have that

$$\mathbb{P}\left(\exists S : |S| \leq 2\ell_0, e_H(S) \geq \frac{|S|+1}{k-1}\right) \leq \sum_{s=4}^{2\ell_0} \binom{n}{s} \binom{sr}{\frac{s+1}{k-1}} \left(\frac{\binom{sr}{k-1}}{\binom{km-2k\ell_0}{k-1}}\right)^{\frac{s+1}{k-1}}$$

$$\leq \sum_{s=4}^{2\ell_0} \left(\frac{ne}{s}\right)^s (er(k-1))^{\frac{s+1}{k-1}} \left(\frac{rs}{rn-o(n)}\right)^{s+1}$$

$$\leq \frac{1}{n^{1-o(1)}} \sum_{s=4}^{2\ell_0} se^s \left(e(k-1)r\right)^{\frac{s+1}{k-1}} = o(1).$$

□

Let $\mathcal{E}$ denote the high probability event in Lemma 5. We will condition on the occurrence of $\mathcal{E}$.

Now for $v \in [n]$ let $S_\ell(v)$ denote the set of vertices at distance $\ell$ from $v$ and let $S_{\leq \ell}(v) = \bigcup_{j \leq \ell} S_j(v)$. We note that

$$|S_\ell(v)| \leq (k-1)r\rho^{\ell-1} \text{ for all } v \in [n], \ell \geq 1. \tag{3}$$

Furthermore, Lemma 5 implies that there exist $b_{r,k} < a_{r,k} < (k-1)r$ such that w.h.p., we have for all $v, w \in [n], 1 \leq \ell \leq \ell_0$,

$$|S_\ell(v)| \geq a_{r,k}\rho^{\ell-1}. \tag{4}$$

$$|S_\ell(v) \setminus S_\ell(w)| \geq b_{r,k}\rho^{\ell-1}. \tag{5}$$

This is because there can be at most one cycle in $S_{\leq \ell_0}(v)$ and the sizes of the relevant sets are reduced by having the cycle as close to $v, w$ as possible.

Now consider $\ell > \ell_0$. Consider doing breadth first search from $v$ or $v, w$ exposing the configuration pairing as we go. Let an edge be *dispensable* if exposing it contains two vertices already known to be in $S_{\leq \ell}$. Lemma 5 implies that w.h.p. there is at most one dispensable edge in $S_{\leq \ell_0}$.

**Lemma 6.** *With probability $1 - o(n^{-2})$, (i) at most 20 of the first $n^{\frac{2}{5}}$ exposed edges are dispensable and (ii) at most $n^{1/4}$ of the first $n^{\frac{3}{5}}$ exposed edges are dispensable.*

*Proof.* The probability that the $\sigma$th edge is dispensable is at most $\frac{(\sigma-1)(k-1)r}{rn-k\rho}$, independent of the history of the process. Hence,

$$\mathbb{P}(\exists\ 20 \text{ dispensable edges in the first } n^{2/5}) \leq \binom{n^{2/5}}{20}\left(\frac{(k-1)rn^{2/5}}{rn-o(n)}\right)^{20} = o(n^{-2}).$$

$$\mathbb{P}(\exists\ n^{1/4} \text{ dispensable edges in first } n^{3/5}) \leq \binom{n^{3/5}}{n^{1/4}}\left(\frac{(k-1)rn^{3/5}}{rn-o(n)}\right)^{n^{1/4}} = o(n^{-2}).$$

□

Now let $\ell_1 = \lceil \log_\rho n^{2/5} \rceil$ and $\ell_2 = \lceil \log_\rho n^{3/5} \rceil$. Then, we have that, conditional on $\mathcal{E}$, with probability $1 - o(n^{-2})$,

$$|S_\ell(v)| \geq (a_{r,k}\rho^{\ell_0-1} - 40)\rho^{\ell-\ell_0} : \ \ell_0 < \ell \leq \ell_1.$$
$$|S_\ell(v)| \geq (a_{r,k}\rho^{\ell_1-1} - 40\rho^{\ell_1-\ell_0} - 2n^{1/4})\rho^{\ell-\ell_1};\ \ell_1 < \ell \leq \ell_2.$$
$$|S_\ell(w) \setminus S_\ell(v)| \geq (b_{r,k}\rho^{\ell_0-1} - 40)\rho^{\ell-\ell_0} : \ \ell_0 < \ell \leq \ell_1.$$
$$|S_\ell(w) \setminus S_\ell(v)| \geq (b_{r,k}\rho^{\ell_1-1} - 40\rho^{\ell_1-\ell_0} - 2n^{1/4})\rho^{\ell-\ell_1};\ \ell_1 < \ell \leq \ell_2.$$

We deduce from this that if $\ell_3 = \lceil \log_{r-1} n^{4/7} \rceil$ and $\ell = \ell_3 + a, a = O(1)$ then with probability $1 - o(n^{-2})$,

$$|S_\ell(w)| \geq (a_{r,k} - o(1))\rho^{\ell-1} \approx a_{r,k}\rho^{a-1}n^{4/7}.$$
$$|S_\ell(w) \setminus S_\ell(v)| \geq (b_{r,k} - o(1))\rho^{\ell-1} \approx b_{r,k}\rho^{a-1}n^{4/7}.$$

Suppose now that we consider the execution of breadth first search up until we have exposed $S_k(v)$. Let $d_\ell(v)$ denote the number of vertices at distance $\ell$ from $v$. Then in order to have $d_\ell(v) = d_\ell(w)$, conditional on the history of the search, there has to be an exact outcome for $|S_\ell(w) \setminus S_\ell(v)|$. Now consider the pairings of the $W_x, x \in S_\ell(w) \setminus S_\ell(v)$. Now at most $n^{1/4}$ of these pairings are with vertices in $S_{\leq \ell}(v) \cup S_{\leq \ell}(w)$. Condition on these. There must now be $s = \Theta(n^{4/7})$ pairings between $W_x, x \in S_\ell(w) \setminus S_\ell(v)$ and $W_y, y \notin S_\ell(v) \cup S_\ell(w)$. Furthermore, to have $d_\ell(v) = d_\ell(w)$ these $s$ pairings must involve exactly $t$ of the sets $W_y, y \notin S_\ell(v) \cup S_\ell(w)$, where $t$ is determined *before* the choice of these $s$ pairings. The following lemma will easily show that w.h.p. $H$ has a canonical labeling defined by the values of $d_\ell(v), 1 \leq \ell \leq \ell^*$, $v \in [n]$.

**Lemma 7.** *Let $R = \bigcup_{i=1}^{\mu} R_i$ be a partitioning of an $r\mu$ set $R$ into $\mu$ subsets of size $r$. Suppose that $S$ is a random $s$-subset of $R$, where $\mu^{5/9} < s < \mu^{3/5}$. Let $X_S$ denote the number of sets $R_i$ intersected by $S$. Then*

$$\max_j \mathbb{P}(X_S = j) \leq \frac{c_0 \mu^{1/2}}{s},$$

*for some constant $c_0$.*

*Proof.* We may assume that $s \geq \mu^{1/2}$. The probability that $S$ has at least 3 elements in some set $R_i$ is at most

$$\frac{\mu\binom{r}{3}\binom{r\mu-3}{s-3}}{\binom{r\mu}{s}} \leq \frac{s^3}{6\mu^2} \leq \frac{\mu^{1/2}}{6s}.$$

But

$$\mathbb{P}(X_S = j) \leq \mathbb{P}\left(\max_i |S \cap R_i| \geq 3\right) + \mathbb{P}\left(X_S = j \text{ and } \max_i |S \cap R_i| \leq 2\right).$$

So the lemma will follow if we prove that for every $j$,

$$P_j = \mathbb{P}\left(X_S = j \text{ and } \max_i |S \cap R_i| \leq 2\right) \leq \frac{c_1 \mu^{1/2}}{s}, \tag{6}$$

for some constant $c_1$.

Clearly, $P_j = 0$ if $j < s/2$ and otherwise

$$P_j = \frac{\binom{\mu}{j}\binom{j}{s-j} r^{2j-s} \binom{r}{2}^{s-j}}{\binom{r\mu}{s}}. \tag{7}$$

Now for $s/2 \leq j < s$ we have

$$\frac{P_{j+1}}{P_j} = \frac{(\mu - j)(s - j)}{(2j + 2 - s)(2j + 1 - s)} \frac{2r}{r - 1}. \tag{8}$$

We note that if $s - j \geq \frac{10s^2}{\mu}$ then $\frac{P_{j+1}}{P_j} \geq \frac{10r}{3(r-1)} \geq 2$ and so the $j$ maximising $P_j$ is of the form $s - \frac{\alpha s^2}{\mu}$ where $\alpha \leq 10$. If we substitute $j = s - \frac{\alpha s^2}{\mu}$ into (8) then we see that

$$\frac{P_{j+1}}{P_j} \in \frac{2\alpha r}{r - 1}\left[1 \pm c_2 \frac{s}{\mu}\right]$$

for some absolute constant $c_2 > 0$.

It follows that if $j_0$ is the index maximising $P_j$ then

$$\left| j_0 - \left( s - \frac{(r-1)s^2}{2r\mu} \right) \right| \leq 1.$$

Furthermore, if $j_1 = j_0 - \frac{s}{\mu^{1/2}}$ then

$$\frac{P_{j+1}}{P_j} \leq 1 + c_3 \frac{\mu^{1/2}}{s} \text{ for } j_1 \leq j \leq j_0,$$

for some absolute constant $c_3 > 0$.

This implies that for all $j_1 \leq j \leq j_0$,

$$P_j \geq P_{j_0} \left( 1 + c_3 \frac{\mu^{1/2}}{s} \right)^{-(j_0 - j_1)} = \\ P_{j_0} \exp\left\{ -(j_0 - j_1) \left( c_3 \frac{\mu^{1/2}}{s} + O\left(\frac{\mu}{s^2}\right) \right) \right\} \geq P_{j_0} e^{-2c_3}.$$

It follows from this that

$$P_{j_0} \leq e^{2c_3} \min_{j \in [j_1, j_0]} P_j \leq \frac{e^{2c_3}}{j_0 - j_1} \sum_{j \in [j_1, j_0]} P_j \leq \frac{e^{2c_3} \mu^{1/2}}{s}.$$

□

We apply Lemma 7 with $\mu = n, s = \rho = \Theta(n^{4/7})$ to show that

$$\mathbb{P}(d_\ell(v) = d_\ell(w), \ell \in [\ell_3, \ell_3 + 14]) \leq \left( \frac{c_0 n^{1/2}}{n^{4/7}} \right)^{15} = o(n^{-2}).$$

This completes the proof of Theorem 3.

# References

[1] B. Bollobás, Distinguishing vertices of random graphs, *Annals of Discrete Mathematics* 13 (1982) 33-50.

[2] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled graphs, *European Journal on Combinatorics* 1 (1980) 311-316.

[3] L. Babai and L. Kucera, Canoncial labelling of graphs in linear average time, *20$^{th}$ Annual IEEE Symp. on Foundations of Comp. Sci. (Puerto Rico)* (1979), pp 39–46.

[4] C.Cooper, A.M. Frieze, M.Molloy and B.Reed, Perfect matchings in random$r$-regular, $s$-uniform hypergraphs, *Combinatorics, Probability and Computing* 5, 1-15.

[5] P. Erdős and J. Spencer, Probabilistic Methods in Combinatorics, *Akadémiai Kiadó Budapest* (1974).

[6] P. Erdős and R.J Wilson, On the Chromatic Index of almost all Graphs, *J. Comb. Theory - B* 23 (1977) pp 255–257.

[7] W. Feller, An Introduction to Probability Theory and it's Applications, *Vol. 1, 3$^{rd}$ ed., John Wiley, New York*, (1968).

[8] A. Frieze and M. Karoński, Introduction to Random Graphs, *Cambridge University Press* 2015.

[9] R.M. Karp, The fast approximation solution of hard combinatorial problems, *Proc.* 6$^{th}$ *South-Eastern Conf. Combinatorics, Graph Theory and Computing (Florida Atlantic U.*, (1975) pp 15–31.

[10] L. Kucera, Canonical labeling of regular graphs in linear average time, *Proceedings of the 28th Annual Symposium on Foundations of Computer Science* (1987) 271-279.

[11] R. J. Lipton, The beacon set approach to Graph Isomorphism, *Yale University*, (1978) preprint.

[12] L. Babai, P. Erdős and S. M. Selkow, Random Graph Isomorphism, *SIAM J. Computing* 9 (3) (1980) pp 628–635.

[13] T. Czajka and G. Pandurangan, Improved Random Graph Isomorphism, *J. Disc. Alg.* 6 (2008) pp 85–92.