The hitting time of clique factors

(Extended abstract)

Annika Heckel[*]   Marc Kaufmann[†]   Noela Müller[‡]   Matija Pasch[§]

Abstract

In [13], Kahn gave the strongest possible, affirmative, answer to Shamir's problem, which had been open since the late 1970s: Let $r \geqslant 3$ and let n be divisible by r. Then, in the random r-uniform hypergraph process on n vertices, as soon as the last isolated vertex disappears, a perfect matching emerges. In the present work, we prove the analogue of this result for clique factors in the random graph process: At the time that the last vertex joins a copy of the complete graph K_r, the random graph process contains a K_r-factor. Our proof draws on a novel sequence of couplings which embeds the random hypergraph process into the cliques of the random graph process. An analogous result is proved for clique factors in the s-uniform hypergraph process $(s \geqslant 3)$.

1 Introduction

When can we cover the vertices of a graph with disjoint isomorphic copies of a small subgraph? The study of this question goes back at least to 1891, when Julius Petersen, in his *Theorie der regulären graphs* [16], provided sufficient conditions for a graph to contain a perfect matching, that is, a cover of the vertices with pairwise disjoint edges. Let $H_r(n, \pi)$ be the random r-uniform hypergraph on the vertex set $V = [n]$ where each of the $N_r = \binom{n}{r}$ possible hyperedges of size r is present independently with probability π. The binomial random graph in this notation is then $G(n, p) = H_2(n, p)$. In 1979, Shamir asked the following natural question, as reported by Erdős [6]:

Question 1.1. How large does $\pi = \pi(n)$ need to be for $H_r(n, \pi)$ to contain a perfect matching whp[1], that is, a collection of n/r vertex-disjoint hyperedges[2]?

A closely related question, posed by Ruciński [18] and Alon and Yuster [1], is:

Question 1.2. For which $p = p(n)$ does the random graph $G(n, p)$ contain a K_r-factor whp?

That is, for which $p(n)$ does $G(n, p)$ contain a collection of n/r vertex-disjoint copies of K_r? In the following, we will also call a copy of K_r an r-clique. For $r = 2$, the two questions are the same — and thanks to Erdős and Rényi [5], we have known since 1966 that there is a sharp threshold[3] for the existence of a perfect matching at $p_0 = \frac{\log n}{n}$. The lower bound for this is immediate: At $p = (1 - \varepsilon)p_0$, some vertices in the graph are still isolated, so there cannot be a perfect matching. The upper bound relies on Tutte's Theorem, for which there is no known hypergraph analogue.

On the other hand, for the case $r \geqslant 3$, these questions remained some of the most prominent open problems in random (hyper-)graph theory. Initial results on perfect matchings in random r-uniform hypergraphs were obtained by Schmidt and Shamir [19] - guaranteeing a perfect matching for hypergraphs with expected degree $\omega(\sqrt{n})$, with improvements by Frieze and Janson [7] to $\omega(n^{\frac{1}{3}})$ and further to $\omega(n^{1/(5+2/(r-1))})$ by Kim [14]. For clique factors, even determining the special case of triangle factors proved hard, despite partial results by Alon and Yuster [1], Ruciński [18] and Krivelevich [15]. Finally, both questions were jointly resolved up to constant factors by Johansson, Kahn and Vu in their seminal paper [11]. It had long been assumed that, as in the case $r = 2$, the main obstacle in finding a perfect matching in $H_r(n, \pi)$ were *isolated vertices*, that is, vertices not contained in any hyperedge. In the clique factor setting, the obstacle corresponding to isolated vertices are vertices not contained in any r-clique. Let

$$\pi_0 = \pi_0(r) = \frac{\log n}{\binom{n-1}{r-1}} \quad \text{and} \quad p_0 = p_0(r) = \pi_0^{1/\binom{r}{2}};$$

---

[1] We say that a sequence of events $(E_n)_{n \geqslant 1}$ holds with high probability (whp) if $P(E_n) \to 1$ as $n \to \infty$.

[2] Here and in the following, we implicitly assume $n \in r\mathbb{Z}_+$ whenever necessary.

[3] Recall that a sequence $p^* = p^*(n)$ is called a *sharp threshold* for a graph property $\mathcal{P}$, if for all fixed $\epsilon > 0$ we have $G(n, p) \notin \mathcal{P}$ whp if $p(n) < (1 - \epsilon)p^*(n)$, and $G(n, p) \in \mathcal{P}$ whp if $p(n) > (1 + \epsilon)p^*(n)$. For a (weak) threshold, the conditions become $p = o(p^*)$ and $p^* = o(p)$, respectively.

then π_0 and p_0 are known to be sharp thresholds for the properties 'minimum degree at least 1' in $H_r(n, \pi)$ and 'every vertex is covered by an r-clique' in $G(n, p)$, respectively [4, 10]. Johansson, Kahn and Vu [11] showed that π_0 and p_0 are indeed (weak) thresholds for the existence of a perfect matching in $H_r(n, \pi)$ and for the existence of an r-clique factor in $G(n, p)$, respectively.

Recently, Kahn [12] proved that π_0 is in fact a *sharp* threshold for the existence of a perfect matching in $H_r(n, \pi)$. Indeed, he was able to confirm the conjecture that isolated vertices are essentially the only obstacle, and thereby answer Shamir's question, in the strongest possible sense:

Let $\mathbf{h}_1, \ldots, \mathbf{h}_{N_r}$ be a uniformly random order of the hyperedges in $\binom{V}{r}$, then the *random r-uniform hypergraph process* $(H_t^r)_{t=0}^{N_r}$ is given by $H_t^r = \{\mathbf{h}_1, \ldots, \mathbf{h}_t\}$. Let

$$T_H = \min\{t : H_t^r \text{ has no isolated vertices}\}$$

be the hyperedge cover hitting time, i.e., the time t where the last isolated vertex 'disappears' by being included in a hyperedge. In the graph case $r = 2$, Bollobás and Thomason [2] proved in 1985 that this hitting time whp coincides with the hitting time for a perfect matching. Kahn [13] showed that this is indeed also the case when $r \geqslant 3$:

Theorem 1.3 ([13]). *Let $r \geqslant 3$ and $n \in r\mathbb{Z}_+$, then whp $H_{T_H}^r$ has a perfect matching.*

Can we get a similarly strong answer to the clique factor question? For $r = 3$, the question whether a triangle factor exists in the random graph process as soon as every vertex is covered by a triangle was attributed to Erdős and Spencer in [3, §5.4]. This question seems much harder than its Shamir counterpart because, unlike hyperedges in the random hypergraph, cliques do not appear independently of each other. However, for sharp thresholds it has indeed been possible to *reduce* the clique factor problem to the perfect matching problem, using the following coupling result of Riordan (for $r \geqslant 4$) and the first author (for $r = 3$):

Theorem 1.4 ([8, 17]). *Let $r \geqslant 3$. There are constants $\varepsilon(r), \delta(r) > 0$ such that, for any $p = p(n) \leqslant n^{-2/r+\varepsilon}$, letting $\pi = p^{\binom{r}{2}}(1-n^{-\delta})$, we may couple the random graph $G = G(n, p)$ with the random r-uniform hypergraph $H = H_r(n, \pi)$ so that, whp, for every hyperedge in H there is a copy of K_r in G on the same vertex set.*[4]

Together with Kahn's sharp threshold result [12], the following corollary is immediate.

Corollary 1.5. *There is a sharp threshold for the existence of a K_r-factor at p_0.*

In the same spirit, we wish to transfer Kahn's hitting time theorem, Theorem 1.3, directly to the random graph process setting, showing its clique factor analogue. Such a derivation of the factor result from its Shamir counterpart was believed to be out of reach — Kahn remarks in [12] that 'there seems little chance of anything analogous' for

---

[4]In [8, 17], Theorem 1.4 was given with an unspecified $o(1)$-term in place of $n^{-\delta}$; the formulation above is Remark 4 in [17] and in the case $r = 3$, an unnumbered remark near the end of [8].

Theorem 1.3, and in [13] that the connection between the factor version and the Shamir version of the result 'seems unlikely to extend to' Theorem 1.3. One important reason for this is that the original coupling provides merely a one-way bound. While it guarantees a copy of K_r in $G = G(n,p)$ on the same vertex set for every h in $H = H_r(n, \pi)$, we cannot, as observed by Riordan [17], expect to find a corresponding hyperedge of H for every K_r in G, since there we will find roughly $n^{2r-2}p^{2\binom{r}{2}-1}$ pairs of K_r sharing two vertices, which is much larger than the expected number $n^{2r-2}\pi^2$ of pairs of hyperedges of H sharing two vertices. A second obstacle is that whenever we do have such a pair of overlapping hyperedges in H, the corresponding cliques in G will not appear independently of each other in the associated random graph process — for example the shared edge could be the last to appear, and then those cliques emerge simultaneously in the random graph process. And indeed, extra cliques and pairs of overlapping cliques do pose a challenge, but they will not appear 'near' those candidate vertices which may be among the last vertices to be covered by cliques.

Now, let $(G_t)_{t=0}^{N_2}$ be the random graph process, which is the random r-uniform hypergraph process for $r = 2$. Denote the hitting time of an r-clique cover by

$$T_G = \min\{t : \text{ every vertex in } G_t \text{ is contained in at least one } r\text{-clique}\}.$$

Then, to apply Kahn's hitting time result to the clique factor setting, we need to find a copy of $H_{T_H}^r$ within the cliques of G_{T_G}. That this can be achieved is our main result:

Theorem 1.6. *Let $r \geqslant 3$. We may couple the random graph process $(G_t)_{t=0}^{N_2}$ with the random r-uniform hypergraph process $(H_t^r)_{t=0}^{N_r}$ so that, whp, for every hyperedge in $H_{T_H}^r$ there is a clique in G_{T_G} on the same vertex set. In particular, whp G_{T_G} contains a K_r-factor.*

What is more, a simplification of the proof of Theorem 1.6 yields a corresponding result for $K_r^{(s)}$-factors. For this, let $r > s \geqslant 3$ and $K_r^{(s)}$ denote the complete s-uniform hypergraph on r vertices. Let $(G_t)_{t=1}^{N_s} = (H_t^s)_{t=1}^{N_s}$ and denote the hitting time of a $K_r^{(s)}$-cover by T_G. Then:

Theorem 1.7. *Let $r > s \geqslant 3$. We may couple the stopped random r-uniform hypergraph process H_{T_H} and the stopped random s-uniform hypergraph process G_{T_G} so that, whp, for every hyperedge in H_{T_H} there is copy of $K_r^{(s)}$ in G_{T_G} on the same vertex set. In particular, whp G_{T_G} has a $K_r^{(s)}$-factor.*

2   Preliminaries

In the remainder, we fix $r \geqslant 3$ and suppress the dependence on r writing H_t instead of H_t^r, etc. Let $M = N_r = \binom{n}{r}$ and $N = N_2 = \binom{n}{2}$. By an r-*uniform hypergraph* H on the vertex set $V = [n]$, we mean a subset of $\binom{V}{r}$, the set of all r-subsets of V. That is, we will use H as a set (of sets of vertices of size r) for convenient notation. For a hypergraph H on

the vertex set $V = [n]$ and $v \in [n]$, we use $d(v)$ to denote the degree of v in H. In a graph G, an r-*clique* is a clique on r vertices. We denote by $\text{cl}(G)$ the set of vertex sets from $\binom{V}{r}$ which span r-cliques in G (so $\text{cl}(G)$ is an r-uniform hypergraph in the aforementioned sense). Throughout the paper, we fix an arbitrary function $g(n)$ satisfying

$$g(n) = o\big(\log n / \log \log n\big) \quad \text{and} \quad g(n) = \omega(1). \tag{1}$$

2.1 The standard coupling and the critical window

It will be useful to work with the following standard device which gives a convenient coupling of the random hypergraphs $H(n, \pi)$ for all $\pi \in [0, 1]$ and the random hypergraph process.

Definition 2.1 (Standard coupling). For every $h \in \binom{V}{r}$, let U_h be an independent random variable, uniform from $[0, 1]$. Let

$$H_\pi = (V, \{h : U_h \leqslant \pi\}).$$

Then $H_\pi \sim H(n, \pi)$. Almost surely all values U_h, $h \in \binom{V}{r}$, are distinct, yielding an instance of the random hypergraph process $(H_t)_{t=0}^M$, as we can add the hyperedges in ascending order of U_h.

We will operate within the following *critical window*: Define π_- and π_+ by setting

$$\pi_\pm = \frac{\log n \pm g(n)}{\binom{n-1}{r-1}}, \tag{2}$$

where $g(n)$ is the function which was fixed globally in (1), and, using δ, ε from Theorem 1.4, let

$$p_\pm = \big(\pi_\pm / (1 - n^{-\delta})\big)^{1/\binom{r}{2}}. \tag{3}$$

For n large enough we have $p_+ \leqslant n^{-2/r+\epsilon}$, so Theorem 1.4 applies with $p = p_+$ and $\pi = \pi_+$.

It is well-known that (π_-, π_+) is the 'critical window' for the disappearance of the last isolated vertex in a random r-uniform hypergraph (see [4, Lemma 5.1(a)]), and so by Theorem 1.3 for the appearance of a perfect matching. So if we couple as in Definition 2.1, then whp we have

$$G_{p_-} \subset G_{T_G} \subset G_{p_+} \quad \text{and} \quad H_{\pi_-} \subset H_{T_H} \subset H_{\pi_+}. \tag{4}$$

2.2 Proof overview

Define p_+, π_+ as in equations (2) and (3). Our starting point is the coupling of $G \sim G(n, p_+)$ and $H \sim H(n, \pi_+)$ given by Theorem 1.4. We review this coupling in §3. In §4, the heart of the proof, we take the coupled $G \sim G(n, p_+)$ and $H \sim H(n, \pi_+)$ and proceed by carefully coupling uniform orders of the edges of G and hyperedges of H. Since p_+ and π_+ are

the upper ends of the respective critical windows (see §2.1), whp this couples (copies of) the stopped graph process G_{T_G} and the stopped hypergraph process H_{T_H}. This coupling *almost* does what we want: for all hyperedges $h \in H_{T_H}$, except those in a small exceptional set $\mathcal{E}$, there is an r-clique in G_{T_G} on the same vertex set. Moreover, we show that whp all $h \in \mathcal{E}$ have a partner hyperedge which appears between time T_H and time $T_H + \lfloor g(n)n \rfloor$. To prove Theorem 1.6, we are left to show that we can get rid of the hyperedges in $\mathcal{E}$ and still have an instance of the stopped random hypergraph process. To this end, $\mathcal{E}$ can be whp embedded into a binomial random subset $\mathcal{R} \subset H_{T_H}$ where each hyperedge $h \in H_{T_H}$ is included independently with a small probability. We proceed to show that if we remove the hyperedges in $\mathcal{R}$ from the hypergraph process up to time T_H, whp this essentially does not change the hitting time T_H, and in particular whp $H_{T_H} \setminus \mathcal{R}$ is still an instance of the stopped random hypergraph process. Chaining the couplings together then proves Theorem 1.6. The necessary modifications in the proof of Theorem 1.7 are detailed in [9].

3   Coupling of $G(n, p_+)$ and $H(n, \pi_+)$

In §3.1 we briefly review Riordan's coupling from Theorem 1.4 for $r \geqslant 4$.[5] We let $\pi = \pi_+$ from equation (2) and $p = p_+$ from equation (3).

3.1   The coupling algorithm for $r \geqslant 4$

Order the $M = \binom{n}{r}$ potential hyperedges in some arbitrary way as $h_1, \ldots, h_M$, and for $1 \leqslant j \leqslant M$, let A_j be the event that there is an r-clique in $G \sim G(n, p)$ on the vertex set of h_j. We construct the coupling of $G \sim G(n, p)$ and $H \sim H(n, \pi)$ step by step; in step j revealing whether or not $h_j \in H$, as well as some information about A_j.

Coupling algorithm: For each j from 1 to M:

- Calculate π_j, the conditional probability of A_j given all the information revealed so far.

- If $\pi_j \geqslant \pi$, toss a coin which lands heads with probability π/π_j, independently of everything else. If the coin lands heads, then test whether A_j holds (which it does with probability exactly π_j). Include the hyperedge h_j in H if and only if the coin lands heads and A_j holds. (Note that the probability of including h_j is exactly $\pi/\pi_j \cdot \pi_j = \pi$.)

- If $\pi_j < \pi$, then toss a coin which lands heads with probability π (independently of everything else), and declare h_j present in H if and only if the coin lands heads. If this happens for any j, we say that the coupling has failed.

After steps $j = 1, \ldots, M$, we have decided all hyperedges of H, and revealed information on the events $A_1, \ldots, A_M$ of G. Now choose G conditional on the revealed information on the events A_j.

---

[5]For the modifications in the case $r = 3$ we refer the reader to [8].

4 Process coupling

Building upon Theorem 1.4, we couple the random graph process with the random hypergraph process. Roughly speaking, we may couple the random graph process and the random hypergraph process so that there is *almost* a copy of H_{T_H} within the r-cliques of G_{T_G}: for all hyperedges in H_{T_H} except those in a set $\mathcal{E}$ (the *exceptional* hyperedges), there is an r-clique in G_{T_G} on the same vertex set. Moreover, the hyperedges in $\mathcal{E}$ all gain a partner hyperedge shortly after time T_H.

Proposition 4.1. *We may couple the random graph process $(G_t)_{t=0}^N$ and the random hypergraph process $(H_t)_{t=0}^M$ so that whp the following holds. There is a set of hyperedges $\mathcal{E} \subset H_{T_H}$ so that*

a) $H_{T_H} \setminus \mathcal{E} \subset \mathrm{cl}(G_{T_G})$, *and*

b) for every $h_1 \in \mathcal{E}$ there is a $h_2 \in H_{T_H + \lfloor g(n)n \rfloor} \setminus H_{T_H}$ so that $|h_1 \cap h_2| = 2$.

Now whp, we can embed the set $\mathcal{E}$ of 'exceptional' hyperedges from Proposition 4.1 into a random set $\mathcal{R}$ which includes every $h \in H_{T_H}$ independently with a small probability.

Proposition 4.2. *We may couple the random r-uniform hypergraph process $(H_t)_{t=0}^M$ and a set $\mathcal{R} \subset \binom{[n]}{r}$ of hyperedges so that both of the following properties hold.*

a) We have $\mathcal{R} \subseteq H_{T_H}$, and (given only H_{T_H}) each hyperedge $h \in H_{T_H}$ is included in $\mathcal{R}$ independently with probability $\pi_{\mathcal{R}} = \frac{10r^4 g(n)}{n}$.

b) Let $\mathcal{F} \subset H_{T_H}$ be the set of hyperedges in H_{T_H} with a partner hyperedge in $H_{T_H + \lfloor g(n)n \rfloor} \setminus H_{T_H}$. Then, whp, $\mathcal{F} \subset \mathcal{R}$.

As the final puzzle piece, we find that after removing every hyperedge from H_{T_H} independently with a small probability, whp we still have an instance of the stopped random hypergraph process.

Proposition 4.3. *Let H_{T_H} be the stopped random hypergraph process, and let $\mathcal{R} \subset H_{T_H}$ be a subset of hyperedges where we include every $h \in H_{T_H}$ independently with probability $\pi_{\mathcal{R}} = \frac{10r^4 g(n)}{n}$. We may couple H_{T_H} and $\mathcal{R}$ with another instance $H'_{T'_H}$ of the stopped random hypergraph process so that, whp, $H_{T_H} \setminus \mathcal{R} = H'_{T'_H}$.*

Combining Propositions 4.1, 4.2 and 4.3 yields a chain of couplings that whp embeds the stopped hypergraph process into the cliques of the stopped graph process, completing the proof:

$$ H'_{T'_H} \stackrel{\mathrm{whp}}{=} H_{T_H} \setminus \mathcal{R} \stackrel{\mathrm{whp}}{\subseteq} H_{T_H} \setminus \mathcal{F} \stackrel{\mathrm{whp}}{\subseteq} \mathrm{cl}(G_{T_H}). $$

References

[1] Noga Alon and Raphael Yuster. Threshold functions for H-factors. *Combin. Probab. Comput.*, 2(2):137–144, 1993.

[2] Béla Bollobás and Andrew Thomason. Random graphs of small order. In *Random graphs '83 (Poznań, 1983)*, volume 118 of *North-Holland Math. Stud.*, pages 47–97. North-Holland, Amsterdam, 1985.

[3] Fan Chung and Ron Graham. *Erdős on graphs: His legacy of unsolved problems*. AK Peters/CRC Press, 1998.

[4] Pat Devlin and Jeff Kahn. Perfect fractional matchings in k-out hypergraphs. *Electron. J. Combin.*, 24(3):Paper No. 3.60, 12, 2017.

[5] Pál Erdős and Alfréd Rényi. On the existence of a factor of degree one of a connected random graph. *Acta Math. Acad. Sci. Hungar.*, 17:359–368, 1966.

[6] Pál Erdős. On the combinatorial problems which I would most like to see solved. *Combinatorica*, 1(1):25–42, 1981.

[7] Alan Frieze and Svante Janson. Perfect matchings in random s-uniform hypergraphs. *Random Structures & Algorithms*, 7(1):41–57, 1995.

[8] Annika Heckel. Random triangles in random graphs. *Random Structures & Algorithms*, 59(4):616–621, 2021.

[9] Annika Heckel, Marc Kaufmann, Noela Müller, and Matija Pasch. The hitting time of clique factors. *arXiv:2302.08340*, 2023.

[10] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[11] Anders Johansson, Jeff Kahn, and Van Vu. Factors in random graphs. *Random Structures Algorithms*, 33(1):1–28, 2008.

[12] Jeff Kahn. Asymptotics for Shamir's problem. *arXiv:1909.06834*, 2019.

[13] Jeff Kahn. Hitting times for Shamir's problem. *Trans. Amer. Math. Soc.*, 375(1):627–668, 2022.

[14] Jeong Han Kim. Perfect matchings in random uniform hypergraphs. *Random Struct. Algorithms*, 23(2):111–132, 2003.

[15] Michael Krivelevich. Triangle factors in random graphs. *Combinatorics, Probability and Computing*, 6:337 – 347, 1997.

[16] Julius Petersen. Die theorie der regulären graphs. *Acta Math.*, 15:193–220, 1891.

[17] Oliver Riordan. Random cliques in random graphs and sharp thresholds for F-factors. *Random Structures & Algorithms*, 61(4):619–637, 2022.

[18] Andrzej Ruciński. Matching and covering the vertices of a random graph by copies of a given graph. *Discrete Math.*, 105(1-3):185–197, 1992.

[19] Jeanette Schmidt and Eli Shamir. A threshold for perfect matchings in random d-pure hypergraphs. *Discrete Mathematics*, 45(2):287–295, 1983.