The threshold for the square of a Hamilton cycle

Jeff Kahn, Bhargav Narayanan, and Jinyoung Park

Abstract. Resolving a conjecture of Kühn and Osthus from 2012, we show that $p = 1/\sqrt{n}$ is the threshold for the random graph $G_{n,p}$ to contain the square of a Hamilton cycle.

1. Introduction

Understanding thresholds for various properties of interest has been central to the study of random graphs since its initiation by Erdős and Rényi [4], and thresholds for containment of (copies of) specific graphs in the random graph has been the subject of some of the most powerful work in the area; see [3, 11, 10], both for a broad overview, and for threshold basics.

Hamilton cycles in random graphs in particular are the subject of an extensive literature, with, to begin, the question of when they appear posed in [4] and answered in [19, 14, 2, 1]; see [9] for a thorough account. Here, we consider a related question first raised by Kühn and Osthus [15]: *when does the square of a Hamilton cycle appear in the random graph*? We remind the reader that the k-th *power* of a graph G is the graph on $V(G)$ with two vertices joined if and only if their distance in G is at most k.

For this discussion, we write $\mathcal{H}_n^k$ for the k-th power of an n-vertex cycle i.e., a Hamilton cycle of K_n. The expected number of copies of $\mathcal{H}_n^k$ in the binomial random graph $G_{n,p}$ is

$$((n-1)!/2)p^{kn},$$

implying that the threshold for appearance of $\mathcal{H}_n^k$ in $G_{n,p}$ (henceforth, simply the 'threshold for $\mathcal{H}_n^k$') is at least $n^{-1/k}$; here, we follow a standard abuse in using *'the' threshold* for an order of magnitude rather than a specific value. For $k = 1$, it was famously shown by Pósa [19] that the threshold for a Hamilton cycle is $\log n/n$ — this is driven not by expectation considerations, but by the need to avoid isolated vertices — while for $k \geq 3$, it follows from a general result of Riordan [20], based on the second-moment method, that the threshold for $\mathcal{H}_n^k$ is $n^{-1/k}$. The case $k = 2$ has proved more stubborn: here, there is no obvious analogue of isolated vertices pushing the threshold above $n^{-1/2}$, but, unlike for $k \geq 3$, the second-moment method yields

Date: 14 September, 2020.

2010 *Mathematics Subject Classification.* Primary 05C80; Secondary 05C38.

only weak upper bounds. Kühn and Osthus [15] conjectured that $n^{-1/2}$ is correct, and showed that the threshold is at most $n^{-1/2+o(1)}$, an upper bound subsequently improved to $(\log n)^4 n^{-1/2}$ by Nenadov and Škorić [18], to $(\log n)^3 n^{-1/2}$ by Fischer, Škorić, Steger and Trujić [5], and to $(\log n)^2 n^{-1/2}$ in unpublished work of Montgomery [16]. Here we resolve this question, proving the conjecture of [15].

Theorem 1.1. *There is a universal $K > 0$ such that for $p \geq K/\sqrt{n}$,*

$$\mathbb{P}(G_{n,p} \text{ contains the square of a Hamilton cycle}) \to 1$$

as $n \to \infty$.

While the aforementioned attempts are all rooted in the notion of 'absorption' introduced in [21], the proof of Theorem 1.1 takes a different approach, based on the recent resolution, by Frankston and the present authors [6], of Talagrand's relaxation [22] of the 'expectation threshold' conjecture of [12]. We say that a hypergraph $\mathcal{H}$ on a finite vertex set V is q-*spread* if

$$|\mathcal{H} \cap \langle I \rangle| \leq q^{|I|}|\mathcal{G}|$$

for each $I \subset V$, where $\langle I \rangle$ is the increasing family generated by I; in this language, the main result of [6] says that there is a universal $C > 0$ such that if a hypergraph $\mathcal{H}$ with edges of size at most ℓ is q-spread, then a $(Cq \log \ell)$-random subset of V is likely to contain some edge of $\mathcal{H}$.

Applied to the hypergraph $\mathcal{G}$ consisting of all copies of $\mathcal{H}_n^2$ — which is q-spread with $q \sim \sqrt{e/n}$ — the result of [6] says that the threshold for $\mathcal{H}_n^2$ is at most $\log n/\sqrt{n}$. A key point in our proof of Theorem 1.1, which eliminates the offending $\log n$, is the observation that large 'local spreads' $(|\mathcal{G} \cap \langle I \rangle|/|\mathcal{G}|)^{1/|I|}$ are relatively rare, a typical value being more like $1/n$ than $1/\sqrt{n}$.

Formally, we prove the following result which on the surface appears weaker than Theorem 1.1.

Theorem 1.2. *For each $\varepsilon > 0$ there is a K such that for $p \geq K/\sqrt{n}$,*

$$\mathbb{P}(G_{n,p} \text{ contains the square of a Hamilton cycle}) \geq 1 - \varepsilon$$

for sufficiently large $n \in \mathbb{N}$.

However, it is easily seen that Theorem 1.2 implies Theorem 1.1; this just requires applying the machinery of Friedgut [7, 8] to say that the property of containing $\mathcal{H}_n^2$ has a sharp threshold. We omit this by now routine step (and the relevant definitions), and refer the reader to [17] for a similar argument.

This paper is organised as follows: the proof of Theorem 1.2 is given in Section 3, with some basic calculations supporting the argument provided in Section 2, and a few final remarks follow in Section 4.

2. Preliminaries

We will use $\mathcal{M}$ for $E(K_n)$ and from now on write $\mathcal{H}$ for $\mathcal{H}_n^2$. As above, $\mathcal{G}$ is the $(2n)$-uniform hypergraph on vertex set $\mathcal{M}$ consisting of all copies of $\mathcal{H}$ in K_n. Thus $|\mathcal{G}| = (n-1)!/2$, and it is not hard to see that $\mathcal{G}$ is q-spread with

$$q = (2/(n-1)!)^{1/(2n)} \sim \sqrt{e/n},$$

which we recall means that

$$|\mathcal{G} \cap \langle I \rangle| \leq q^{|I|}|\mathcal{G}| \tag{1}$$

for all $I \subset \mathcal{M}$.

The next two observations implement the basic idea mentioned above, that large values of $|\mathcal{G} \cap \langle I \rangle|/|\mathcal{G}|$ are relatively rare.

Proposition 2.1. *For an $I \subset \mathcal{M}$ with $\ell \leq n/3$ edges and c components,*

$$|\mathcal{G} \cap \langle I \rangle| \leq (16)^\ell \left(n - \left\lceil \frac{\ell + c}{2} \right\rceil - 1\right)!$$

Proof. Let $I_1, \ldots, I_c$ be the components of I and $v = |V(I)|$, where for $H \subset \mathcal{M}$, we write $V(H)$ for the set of vertices spanned by H. The upper bound on ℓ implies that no I_j can 'wrap around', so we have $|E(I_j)| \leq 2|V(I_j)| - 3$ for each j and

$$\ell \leq 2v - 3c. \tag{2}$$

We first designate a root vertex v_j for each I_j and order $V(I_j)$ by some $\prec_j$ that begins with v_j and in which each $v \neq v_j$ appears later than at least one of its neighbors. We may then bound $|\mathcal{G} \cap \langle I \rangle|$ as follows.

To specify a J ($\in \mathcal{G}$) containing I, we first specify a cyclic permutation of $\{v_1, \ldots, v_c\} \cup (V(K_n) \setminus V(I))$. By (2), the number of ways to do this, namely, $(n - v + c - 1)!$, is at most

$$\left(n - \left\lceil \frac{\ell + c}{2} \right\rceil - 1\right)!.$$

We then extend to a full cyclic ordering of $V(K_n)$ (thus determining J) by inserting, for $j = 1, \ldots, c$, the vertices of $V(I_j) \setminus \{v_j\}$ in the order $\prec_j$. This allows at most four places to insert each vertex (since one of its neighbours has been inserted before it and the edge joining them must belong to J), so the number of possibilities here is less than $4^v \leq (16)^\ell$, and the proposition follows. □

Proposition 2.2. *For an $F \subset \mathcal{H}$ of size h, the number of subgraphs of F with ℓ edges and c components is at most*

$$(8e)^\ell \binom{2h}{c}.$$

Proof. We need the following standard bound, which follows from the fact (see [13], for example) that the infinite Δ-branching rooted tree contains precisely

$$\frac{\binom{\Delta v}{v}}{(\Delta-1)v+1} \leq (e\Delta)^{v-1}$$

rooted subtrees with v vertices.

Lemma 2.3. *For a graph G of maximum degree Δ, the number of connected, h-edge subgraphs of G containing a given vertex is less than $(e\Delta)^h$.* □

To specify a subgraph J of F as in the proposition, we proceed as follows. We first choose root vertices $v_1, \ldots, v_c$ for the components, say $J_1, \ldots, J_c$, of J, the number of possibilities for this being at most $\binom{2h}{c}$. We then choose the sizes, say $\ell_1, \ldots, \ell_c$, of $J_1, \ldots, J_c$; here the number of possibilities is at most the number of positive integer solutions of $\ell_1 + \cdots + \ell_c = \ell$, which is $\binom{\ell-1}{c-1}$. Finally, we specify for each i, a connected J_i of size ℓ_i rooted at v_i, which according to Lemma 2.3 can be done in at most $\prod(4e)^{\ell_i} = (4e)^\ell$ ways. Combining these estimates with the crude bound of $\binom{\ell-1}{c-1} < 2^\ell$ yields the bound in the proposition. □

3. Proof of the main result

Recall that $\mathcal{M} = E(K_n)$ and $\mathcal{G}$ is the hypergraph of copies of $\mathcal{H} = \mathcal{H}_n^2$ in K_n, and set $m = |\mathcal{M}| = \binom{n}{2}$. For $S \in \mathcal{G}$ and $X \subset \mathcal{M}$, an (S, X)-*fragment* is a set of the form $J \setminus X$ with $J \in \mathcal{G}$ contained in $S \cup X$.

Our main point, Lemma 3.1 below, says that for a suitably large w, most pairs (S, W) with $S \in \mathcal{G}$ and $W \in \binom{\mathcal{M}}{w}$ admit small fragments.

Set $k = 4\sqrt{n}$ and for $S \in \mathcal{G}$ and $X \subset \mathcal{M}$, call the pair (S, X) *good* if some (S, X)-fragment has size at most k, and *bad* otherwise. In what follows, we will always assume $S, J \in \mathcal{G}$ and $W \in \binom{\mathcal{M}}{w}$, where w will be $Cn^{3/2}$ for some large constant C.

Lemma 3.1. *There is a fixed C_0 such that for all $C \geq C_0$ and $n \in \mathbb{N}$, with $w = Cn^{3/2}$,*

$$|\{(S, W) : (S, W) \text{ is bad}\}| \leq 2C^{-k/3}|\mathcal{G}|\binom{m}{w}. \tag{3}$$

Proof. We may of course assume n is large, since values below any fixed n_0 can be handled trivially by adjusting C_0. It is enough to show

$$|\{(S, W) : (S, W) \text{ is bad}, |W \cap S| = t\}| \leq 2C^{-k/3}|\mathcal{G}|\binom{2n}{t}\binom{m-2n}{w-t} \tag{4}$$

for $t \in \{0, \ldots, 2n\}$, since summing over t then gives (3).

Now aiming for (4), we fix t, set $w' = w - t$, and bound the number of bad (S, W)'s with $|W \cap S| = t$ (so $|W \setminus S| = w'$ and $|W \cup S| = w' + 2n$).

Call $Z \in \binom{\mathcal{M}}{w'+2n}$ *pathological* if

$$|\{S \subset Z : (S, Z \setminus S) \text{ is bad}\}| > C^{-k/3}|\mathcal{G}|\binom{m-2n}{w'}\Big/\binom{m}{w'+2n}$$
$$= C^{-k/3}|\mathcal{G}|\binom{w'+2n}{2n}\Big/\binom{m}{2n},$$

and, when $|S \cup X| = w' + 2n$, say (S, X) is pathological if $S \cup X$ is. We bound the nonpathological and pathological parts of (4) separately.

Nonpathological contributions. We claim that the number of nonpathological (S, W)'s in (4) is less than

$$C^{-k/3}|\mathcal{G}|\binom{2n}{t}\binom{m-2n}{w'}. \tag{5}$$

To see this, we specify (S, W) by specifying first $Z = S \cup W$, then S, and then W. The number of possibilities for Z is at most

$$\binom{m}{w'+2n},$$

while, since (S, W) being bad implies that $(S, Z \setminus S)$ is bad (and Z is nonpathological), the number of possibilities for S given Z is at most

$$C^{-k/3}|\mathcal{G}|\binom{m-2n}{w'}\Big/\binom{m}{w'+2n}.$$

Of course the number of possibilities for W given Z and S is at most $\binom{2n}{t}$, and we have (5).

Pathological contributions. The main point here is the following estimate.

Claim 3.2. *For a given $S \in \mathcal{G}$, Y chosen uniformly from $\binom{\mathcal{M}\setminus S}{w'}$, and large enough C,*

$$\mathbb{E}\left[|\{J \in \mathcal{G} : J \subset Y \cup S \text{ and } |J \cap S| \geq k\}|\right] \leq C^{-2k/3}|\mathcal{G}|\binom{w'+2n}{2n}\Big/\binom{m}{2n}. \tag{6}$$

This is proved below, but assuming for the moment it is true, we show that the number of pathological (S, W)'s in (4) is, for C as in the claim, less than

$$C^{-k/3}|\mathcal{G}|\binom{2n}{t}\binom{m-2n}{w'}. \tag{7}$$

To see this we think of choosing $(S, W \cap S)$ — which can be done in at most $|\mathcal{G}|\binom{2n}{t}$ ways — and then $W \setminus S$. For the latter, notice that (S, W) being bad means that *every* $J \subset S \cup W$ has $|J \cap S|$ $(\geq |J \setminus W|) \geq k$. Hence, since (S, W) is pathological, we have

$$|\{J \subset S \cup (W \setminus S) : |J \cap S| \geq k\}| \geq C^{-k/3}|\mathcal{G}|\binom{w'+2n}{2n}\Big/\binom{m}{2n}.$$

But then Claim 3.2, along with Markov's inequality, says the number of possibilities for $W \setminus S$ is at most

$$C^{-k/3}\binom{m-2n}{w'}.$$

Thus, we have (7), and combining this with (5) completes the proof of Lemma 3.1. □

Proof of Claim 3.2. With f_i the fraction of J's (in $\mathcal{G}$) with $|J \cap S| = i$, the left-hand side of (6) is

$$\sum_{i\ge k} |\mathcal{G}| f_i \binom{w'}{2n-i} \Big/ \binom{m-2n}{2n-i},$$

so it is enough to show

$$f_i \left(\binom{w'}{2n-i} \Big/ \binom{m-2n}{2n-i} \right) \left(\binom{w'+2n}{2n} \Big/ \binom{m}{2n} \right)^{-1} = e^{O(i)} C^{-i}, \tag{8}$$

where — here and below — implied constants do not depend on C. The terms other than f_i on left-hand side of (8) reduce to

$$\frac{(w')_{2n-i}}{(w'+2n)_{2n-i}} \cdot \frac{(m)_{2n-i}}{(m-2n)_{2n-i}} \cdot \frac{(m-2n+i)_i}{(w'+i)_i}; = e^{O(i)} C^{-i} n^{i/2}.$$

we omit the routine calculation establishing this, just noting that $\sqrt{n} = O(i)$ since $i \ge k$. Hence, for (8) we just need

$$f_i \le e^{O(i)} n^{-i/2}. \tag{9}$$

For $n/3 \le i \le 2n$, this follows from the fact that $\mathcal{G}$ is q-spread with $q \sim \sqrt{e/n}$, see (1), which gives

$$f_i \le \binom{2n}{i} q^i = e^{O(i)} n^{-i/2}.$$

For $k \le i \le n/3$, Propositions 2.1 and 2.2 with Stirling's formula give

$$f_i \le |\mathcal{G}|^{-1} (128e)^i \sum_{c=1}^{i} \binom{4n}{c} \left(n - \left\lceil \frac{i+c}{2} \right\rceil -1 \right)! = e^{O(i)} n^{-i/2} \sum_{c=1}^{i} (\sqrt{n}/c)^c = e^{O(i)} n^{-i/2},$$

where at the end we use the fact that $(a/x)^x \le e^{a/e}$ and that $i \ge k$. □

We are now ready to prove Theorem 1.2.

Proof of Theorem 1.2. For a given $\varepsilon > 0$ as in the statement of the theorem, we prove the result for $K = 3C_0 + C$, with C_0 as in Lemma 3.1 and C a suitable function of ε (essentially $1/\varepsilon$). Let $p_0 = 3C_0/\sqrt{n}$, $p_1 = C/\sqrt{n}$ and $p = p_0 + p_1 - p_0p_1 < K/\sqrt{n}$. We generate $G_{n,p}$ in two rounds, as $W_0 \cup W_1$, where W_0 and W_1 are independent with W_0 and W_1 distributed as G_{n,p_0} and G_{n,p_1} respectively, and W_1 chosen after W_0 (at which point we are really just interested in $W_1 \setminus W_0$).

Call W_0 *successful* if

$$|\{S : (S, W_0) \text{ is bad}\}| \leq |\mathcal{G}|/2.$$

We first observe that W_0 is (very) likely to be successful: standard concentration estimates give (say)

$$\mathbb{P}(|W_0| < C_0 n^{3/2}) = \exp\left(-n^{3/2}\right),$$

and Lemma 3.1 gives

$$\mathbb{P}(W_0 \text{ unsuccessful} \mid |W_0| \geq C_0 n^{3/2}) < 4C_0^{-k/3};$$

in particular W_0 is successful with probability $1 - o(1)$.

Suppose now that W_0 is successful. For each S with (S, W_0) being good, let $\chi(S, W_0)$ be some k-element subset of S containing an (S, W_0)-fragment, and let $\mathcal{R}$ be the k-uniform *multihypergraph*

$$\{\chi(S, W_0) : (S, W_0) \text{ is good}\}. \tag{10}$$

To finish, we use the second moment method to show that W_1 is reasonably likely to contain a member of $\mathcal{R}$. Setting

$$X = |\{A \in \mathcal{R} : A \subset W_1\}|,$$

we have

$$\mu = \mathbb{E}[X] = |\mathcal{R}| p_1^k$$

and

$$\mathrm{Var}(X) \leq p_1^{2k} \sum \left\{ p_1^{-|A \cap B|} : A, B \in \mathcal{R}, A \cap B \neq \emptyset \right\}. \tag{11}$$

For $R \in \mathcal{R}$ and $1 \leq i \leq k$, as in the proof of Claim 3.2, Propositions 2.1 and 2.2 with Stirling's formula give

$$\begin{aligned} |\{A \in \mathcal{R} : |A \cap R| = i\}| &\leq \sum_{I \subset R, |I| = i} |\mathcal{R} \cap \langle I \rangle| \leq \sum_{I \subset R, |I| = i} |\mathcal{G} \cap \langle I \rangle| \\ &= e^{O(i)} \sum_{1 \leq c \leq i} \binom{2k}{c} \left(n - \left\lceil \frac{i+c}{2} \right\rceil - 1\right)! \\ &= e^{O(i)} n^{-i/2} |\mathcal{G}|. \end{aligned}$$

Recall that W_0 was assumed to be successful, so this means that $|\mathcal{R}| \geq |\mathcal{G}|/2$, and hence, the sum in (11) is at most

$$2|\mathcal{R}|^2 p_1^{2k} \sum_{i=1}^{k} e^{O(i)} p_1^{-i} n^{-i/2} = O(\mu^2/C)$$

for large enough C where, again, the implied constant does not depend on C. Finally, Chebyshev's inequality gives

$$\mathbb{P}(X = 0) \leq \mathrm{Var}(X)/\mu^2 = O(1/C),$$

which is at most ε provided $C = \Omega(1/\varepsilon)$ and we are done. $\square$

4. Conclusion

Our work still leaves open a natural question, namely that of locating the sharp threshold for $G_{n,p}$ to contain a square of the Hamilton cycle. Though there seems little hope of proving such a statement along the present lines, it is natural to guess that the above expectation considerations drive the threshold more precisely.

Conjecture 4.1. *For fixed $\varepsilon > 0$ and $p > (1+\varepsilon)\sqrt{e/n}$,*

$$\mathbb{P}(G_{n,p} \text{ contains the square of a Hamilton cycle}) \to 1$$

as $n \to \infty$.

Acknowledgments

The first and second authors were supported by NSF grants DMS1954035 and DMS-1800521 respectively. The third author's work was supported directly by NSF grant DMS-1926686 and indirectly by NSF grant CCF-1900460.

References

1. M. Ajtai, J. Komlós, and E. Szemerédi, *First occurrence of Hamilton cycles in random graphs*, Cycles in graphs (Burnaby, B.C., 1982), vol. 115, North-Holland, Amsterdam, 1985, pp. 173–178. 1
2. B. Bollobás, *The evolution of sparse graphs*, Graph theory and combinatorics (Cambridge, 1983), Academic Press, London, 1984, pp. 35–57. 1
3. ______, *Random graphs*, second ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001. 1
4. P. Erdős and A. Rényi, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **5** (1960), 17–61. 1
5. M. Fischer, N. Škorić, A. Steger, and M. Trujić, *Triangle resilience of the square of a Hamilton cycle in random graphs*, Preprint, arXiv:1809.07534. 2
6. K. Frankston, J. Kahn, B. Narayanan, and J. Park, *Thresholds versus fractional expectation-thresholds*, Preprint, arXiv:1910.13433. 2
7. E. Friedgut, *Sharp thresholds of graph properties, and the k-sat problem*, J. Amer. Math. Soc. **12** (1999), 1017–1054, With an appendix by Jean Bourgain. 2
8. ______, *Hunting for sharp thresholds*, Random Structures Algorithms **26** (2005), 37–51. 2
9. A. Frieze, *Hamilton cycles in random graphs: a bibliography*, Online Survey, arXiv:1901.07139. 1

10. A. Frieze and M. Karoński, *Introduction to random graphs*, Cambridge University Press, Cambridge, 2016. 1
11. S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000. 1
12. J. Kahn and G. Kalai, *Thresholds and expectation thresholds*, Combin. Probab. Comput. **16** (2007), 495–502. 2
13. D. E. Knuth, *The art of computer programming. Vol.* 1, Addison-Wesley, Reading, MA, 1997, Fundamental algorithms, Third edition. 4
14. J. Komlós and E. Szemerédi, *Limit distribution for the existence of Hamiltonian cycles in a random graph*, Discrete Math. **43** (1983), 55–63. 1
15. D. Kühn and D. Osthus, *On Pósa's conjecture for random graphs*, SIAM J. Discrete Math. **26** (2012), 1440–1457. 1, 2
16. R. Montgomery, Personal communication. 2
17. B. Narayanan and M. Schacht, *Sharp thresholds for nonlinear Hamiltonian cycles in hypergraphs*, Random Structures Algorithms **57** (2020), 244–255. 2
18. R. Nenadov and N. Škorić, *Powers of Hamilton cycles in random graphs and tight Hamilton cycles in random hypergraphs*, Random Structures Algorithms **54** (2019), 187–208. 2
19. L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Math. **14** (1976), 359–364. 1
20. O. Riordan, *Spanning subgraphs of random graphs*, Combin. Probab. Comput. **9** (2000), 125–148. 1
21. V. Rödl, A. Ruciński, and E. Szemerédi, *A Dirac-type theorem for* 3*-uniform hypergraphs*, Combin. Probab. Comput. **15** (2006), 229–251. 2
22. M. Talagrand, *Are many small sets explicitly small?*, Proceedings of the 2010 ACM International Symposium on Theory of Computing, 2010, pp. 13–35. 2

Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA
Email address: jkahn@math.rutgers.edu

Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA
Email address: narayanan@math.rutgers.edu

School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA
Email address: jpark@math.ias.edu