A large hole in pseudo-random graphs

Sahar Diskin * Michael Krivelevich * Itay Markbreit * Maksim Zhukovskii †

Abstract

We show that there exist constants $\delta_1, \delta_2 > 0$ such that if G is an (n, d, λ)-graph with $\lambda/d \le \delta_1$, then G contains an induced cycle of length at least $\delta_2 n/d$. We further demonstrate that, up to a constant factor, this is best possible. Utilising our techniques, we derive that the number of non-isomorphic induced subgraphs of such G is at least exponential in $n \log d/d$, and further demonstrate that this is tight up to a constant factor in the exponent.

1 Introduction and main results

The study of extremal problems about induced subgraphs is a popular theme in combinatorics. Of particular interest is finding the lengths of a longest induced path and of a longest induced cycle (also called a *hole*) in a given graph G. Two key examples, that have been studied in the past, are when G is the d-dimensional hypercube, wherein finding the length of a longest induced path is known as the 'Snake-in-the-Box' problem [1, 19, 29], and when G is the binomial random graph $G(n, p)$. Let us note here that the length of a longest induced path is at most twice the size of a largest independent set in G.

The study of induced cycles in $G(n, p)$ dates back to the late '80s. Frieze and Jackson [18] showed that for each sufficiently large constant d, **whp**[1] the random graph $G(n, d/n)$ contains an induced cycle of length at least $c(d)\, n$ for some constant $c(d) > 0$. Łuczak [26] and, independently, Suen [28] later improved this, proving that whenever $d > 1$, **whp** $G(n, d/n)$ contains an induced cycle of length at least $(1 + o(1))\frac{\ln d}{d}\, n$. A simple first-moment argument implies that **whp** the length of a longest induced cycle for large d in $G(n, d)$ is at most $(1 + o_d(1))\frac{2 \ln d}{d}\, n$. This upper bound was shown to be asymptotically tight by Draganić, Glock, and Krivelevich [10]. Dutta and Subramanian [12] further established a two-point concentration result on the length of a longest induced path in $G(n, p)$ for $p \ge \log^2 n/\sqrt{n}$.

It appears natural to extend the study of extremal problems for induced paths and cycles for random to *pseudo-random* graphs. The latter can be informally described as graphs whose edge distribution resembles that of a truly random graph $G(n, p)$ of a similar density. Formally, given a d-regular graph G on n vertices, denote the eigenvalues of its adjacency matrix by $d = \lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$. Letting $\lambda \coloneqq \max\{|\lambda_2|, |\lambda_n|\}$, we then say that G is an (n, d, λ)-graph. The expander mixing lemma, due to Alon and Chung [4], relates the *spectral ratio* $\frac{\lambda}{d}$ to the edge-distribution of these graphs (see Lemma 2.2). We refer the reader to [20, 22, 23] for comprehensive surveys on the subject of pseudo-random graphs and expanders.

As noted before, it is thus quite natural to ask whether one can find long induced cycles, similar to the case of the binomial random graph $G(n, d/n)$, in (n, d, λ)-graphs (naturally under some assumption on the aforementioned spectral ratio). A first result in this direction was recently obtained by Draganić and Keevash [11], who showed the following: any (n, d, λ)-graph G with $\lambda < d^{3/4}/100$ and $d < n/10$ contains an induced path of length $\frac{n}{64d}$; their paper did not address the problem of long induced cycles.

*School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. Emails: sahardiskin@mail.tau.ac.il, krivelev@tauex.tau.ac.il, markbreit@mail.tau.ac.il.

†School of Computer Science, University of Sheffield, UK. Email: m.zhukovskii@sheffield.ac.uk.

[1]With high probability, that is, with probability tending to one as n tends to infinity

Our first main result improves upon the result of Draganić and Keevash [11] by significantly relaxing the assumption on the spectral ratio, as well as showing the existence of a long induced *cycle*, instead of a path.

Theorem 1. *There exist constants $\delta_1, \delta_2, \delta_3 > 0$ such that the following holds. For any integers n, d and for any (n, d, λ)-graph G such that $d \leq \delta_3 n$ and $\lambda/d \leq \delta_1$, G contains an induced cycle of length at least $\delta_2 n/d$.*

A few comments are in place. We note that for δ_1 sufficiently small, one can take $\delta_2 = 1/150$; in fact, as we will soon see, we can find a longer induced *path*, of length at least $\frac{n}{48d}$, thus improving also the constant factor in the result of [11]. Further, since a random d-regular graph on n vertices $G_{n,d}$ is typically an (n, d, λ)-graph with $\lambda \leq 2\sqrt{d}$ (see, e.g., [16]), we obtain that for large enough d, **whp** $G_{n,d}$ contains an induced cycle of length $\Omega(n/d)$. This improves the bound of Frieze and Jackson [18] from 1985, who showed that **whp** $G_{n,d}$ contains an induced cycle of length $\Omega(n/d^2)$ (see also [13]), and makes progress towards resolving [17, Problem 74], which asks to determine the typical length of a longest induced cycle in $G_{n,d}$. Finally, let us note that our proof yields a randomised algorithm, which finds in a linear in n time an induced cycle of length at least $\delta_2 n/d$ **whp**.

It turns out, perhaps somewhat surprisingly, that insights and results from the setting of *site percolation* prove very efficient, leading to a rather simple (in hindsight) proof of Theorem 1. Intuitively, finding an induced cycle is 'easier' when the graph is sparse, and in this case, choosing a random set of vertices typically yields a sparser graph, wherein every induced structure is also an induced structure in the host graph. Formally, given a host graph $G = (V, E)$, form a random subset V_p by retaining every $v \in V$ independently with probability p. The p-site-percolated[2] subgraph G_p is then $G_p := G[V_p]$. We will derive Theorem 1 from our next result.

Theorem 2. *Let $\epsilon > 0$ be a sufficiently small constant. Let $n, d := d(n) \in \mathbb{N}$ be such that $d = o(n)$. Let $p = \frac{1+\epsilon}{d}$. Then, there exists a constant $\delta := \delta(\epsilon) > 0$ such that the following holds. Let G be an (n, d, λ)-graph with $\frac{\lambda}{d} \leq \delta$. Then,* ***whp*** *$G_p$ contains an induced path of length at least $\frac{\epsilon^2 n}{3d}$.*

Note that Theorem 2 naturally implies that *deterministically* the whole graph G contains a long induced path. In fact, with a little more effort, one can typically find an induced cycle of length $\Omega(n/d)$ in G_p, see details in Remark 3.1. Further, as we mentioned after Theorem 1, it is not hard to verify that one can take any $\epsilon \leq 1/4$, and thus obtain an induced path of length $\frac{n}{48d}$.

Since the edge-distribution of an (n, d, λ)-graph G (when $\lambda/d \leq \delta$ for some sufficiently small constant δ) resembles that of the binomial random graph $G(n, d/n)$, one might expect to find an induced path (or even cycle) of length $\Theta(\frac{\ln d}{d})n$ in G. Perhaps somewhat surprisingly, as our next result shows, this is not the case.

Theorem 3. *For every constant $\delta > 0$, there exists a constant $C := C(\delta) > 0$ such that the following holds. For every sufficiently large $d \in \mathbb{N}$, there exist infinitely many n for which there exists an (n, d, λ)-graph G with $\frac{\lambda}{d} \leq \delta$, whose longest induced path is of length at most $\frac{Cn}{d}$.*

In particular, Theorem 3 shows that Theorem 1 is tight up to a multiplicative constant. Let us note that this also marks a key difference between a longest path in an (n, d, λ)-graph, which is of length linear in n [24], and a longest induced path, which we now see might be of length at most linear in n/d.

[2]In many papers, this notation is reserved for the *bond*-percolated random subgraph. Throughout this paper, we reserve G_p for the p-*site*-percolated random subgraph.

Given an (n, d, λ)-graph G with λ/d being a (small) constant, by Theorems 1 and 3 we know that a largest hole of G is of size $\Omega(n/d)$, and this estimate is tight. It would be interesting to relate the spectral ratio λ/d (where we allow the ratio to depend on d) to the size of a largest hole of an (n, d, λ)-graph G. In particular, one may wonder whether a largest hole is of size $\Omega(n \log(d/\lambda)/d)$ — indeed, when $\lambda \ll d$, the independence number of an (n, d, λ)-graph G satisfies $\alpha(G) = \Omega(n \log(d/\lambda)/d)$ (see, for example, [23, Proposition 4.6]).

Application Theorem 2 and its proof bear interesting consequences in terms of counting non-isomorphic induced subgraphs. Let $\mu(G)$ be the number of non-isomorphic induced subgraphs in G. Erdős and Rényi conjectured that for every constant $c_1 > 0$, there exists a constant $c_2 > 0$ such that if G has no subset S of $c_1 \log n$ vertices on which $G[S]$ is either the complete graph or the empty graph (that is, G is c_1-Ramsey), then $\mu(G) \geq \exp\{c_2 n\}$. Several results in this direction were obtained [3, 5]; in particular, in 1976 Müller [27] showed that **whp** $\mu(G(n, 1/2)) = 2^{(1-o(1))n}$. The conjecture was finally confirmed by Shelah in 1998 [28].

In a recent work [25], the second and fourth authors determined the asymptotics of $\mu(G(n, p))$ for (almost) the entire range of $G(n, p)$. They further showed that **whp** the number of non-isomorphic induced subgraphs in a random d-regular graph $G_{n,d}$ is exponential in n as well, and the base of the exponent grows to 2 with growing d. Utilising Theorem 2, we are able to show that for pseudo-random graphs, $\mu(G)$ is exponential in $(n \log d)/d$.

Theorem 4. *There exist constants $\delta_1, \delta_2, \delta_3 > 0$ such that the following holds. For any integers n, d and for any (n, d, λ)-graph G such that $d \leq \delta_3 n$ and $\lambda/d \leq \delta_1$, $\mu(G) \geq \exp\left(\frac{\delta_2 n \log d}{d}\right)$.*

We further show that the above is tight up to the constant in the exponent.

Theorem 5. *For every constant $\delta > 0$, there exists a constant $C > 0$ such that the following holds. For every sufficiently large $d \in \mathbb{N}$, there exist infinitely many n for which there exists an (n, d, λ)-graph G with $\lambda/d \leq \delta$ satisfying that $\mu(G) \leq \exp\left\{\frac{Cn \log d}{d}\right\}$.*

Structure of the paper In Section 2 we present notation used throughout the paper, describe an adaptation of the Depth-First-Search algorithm which we will employ, and collect several lemmas to be utilised in the proof. Then, in Section 3 we prove Theorem 2 and derive Theorem 1 from it. In Section 4 we prove Theorem 4. Finally, in Section 5 we give constructions proving Theorems 3 and 5.

2 Preliminaries

Given a graph $G = (V, E)$ and sets $A, B \subseteq V(G)$, we denote by $e(A, B)$ the number of edges with one endpoint in A and the other endpoint in B. We further abbreviate $e(A) \coloneqq e(A, A)/2 = e(G[A])$. We denote by $N(A)$ the external neighbourhood of A, that is,

$$N(A) = \{v \in V \setminus A \colon \exists u \in A, uv \in E\}.$$

Given $v \in V$ and $A \subseteq V$, we denote by $d(v, A)$ the number of neighbours of v in A. When $A = V$, we write $d(v) \coloneqq d(v, V)$ for the degree of v in G. Recall that given $G = (V, E)$, we form V_p by retaining every $v \in V$ independently and with probability p; we then abbreviate $G_p = G[V_p]$. Throughout the paper, we systematically ignore rounding signs as long as it does not affect the validity of our arguments.

2.1 Auxiliary Lemmas

We will make use of the following fairly standard Chernoff-type probability bound (see, for example, Appendix A in [6]).

Lemma 2.1. *Let $X \sim Bin(n,p)$. Then, for any $0 \le t \le np$,*

$$\mathbb{P}\left(|X-\mathbb{E}[X]| \ge t\right) \le 2\exp\left\{-\frac{t^2}{3np}\right\}.$$

Throughout the rest of the section, we set $\epsilon > 0$ to be a sufficiently small constant. We assume that G is an (n,d,λ)-graph (in fact, we consider a sequence of pairs $(d_k, n_k) \in \mathbb{N}^2$ and a sequence of (n_k, d_k, λ_k)-graphs $(G_k)_{k\in\mathbb{N}}$ satisfying $d_k = o(n_k)$). We write $\delta = \frac{\lambda}{d}$ for the spectral ratio. We further set $p = \frac{1+\epsilon}{d}$, and let V_p and G_p be as defined above.

Let us first collect several lemmas that will be useful for us throughout the paper. The first is the aforementioned expander mixing lemma, due to Alon and Chung [4].

Lemma 2.2. *For any pair of subsets $A, B \subseteq V(G)$,*

$$\left|e(A,B) - \frac{d}{n}|A||B|\right| \le \lambda\sqrt{|A||B|}.$$

2.2 A modified Depth First Search algorithm

Let us present a variant of the Depth First Search (DFS) algorithm, which we will utilise to show a typical existence of a long induced path when proving Theorem 2. The variant combines ideas of the algorithm presented in [10] together with the one presented in [21].

The algorithm is fed a graph $G = (V, E)$ with an ordering σ on its vertices, and a sequence $(X_v)_{v\in V}$ of i.i.d. Bernoulli(p) random variables (with $0 \le p \le 1$). We maintain five sets of vertices: T, the set of vertices yet to be processed; $U \subseteq V_p$, the set of active vertices, kept in a stack (the last vertex to enter U is the first to leave); S_1 and S_2, the sets of processed vertices (which fell into V_p); and W, the set of processed vertices which fell outside of V_p. We initialise $U, S_1, S_2, W = \varnothing$ and $T = V$. The algorithm terminates once $U \cup T = \varnothing$. As we will see, throughout the execution, U will span an induced path in G_p.

Each step of the algorithm corresponds to exposing a random variable X_v, and proceeds as follows.

1. If U is empty, we consider the first vertex v in T according to σ.
   (a) If $X_v = 1$, we move v from T to U.
   (b) Otherwise (that is, if $X_v = 0$), we move v from T to W.
2. If $U \neq \varnothing$, let u be the vertex on the top of the stack U.
   (a) If u has no neighbours in T, move u from U into S_1, and return to 1 and proceed.
   (b) Otherwise, let v be the first vertex in T according to σ such that $uv \in E$.
      i. If $X_v = 0$, we move v from T to W.
      ii. If $X_v = 1$ and v has a neighbour in $U \setminus \{u\}$, we move v from T to S_2.
      iii. Otherwise (that is, if $X_v = 1$ and v has no neighbours in $U \setminus \{u\}$), we move v from T to U.

We will make use of the following simple observations about the above algorithm. First, at every step of the algorithm, $G_p[U]$ spans an induced path. Indeed, if a vertex v gets added to U at Case

2b, then v augments the existing path spanned by U by attaching itself to its last vertex u. Also, by Case 2(b)iii the only neighbour of v along the path is u, thus the new path is induced as well. Further, at every step, there are no edges between S_1 and T, and thus $N_G(S_1) \subseteq S_2 \cup U \cup W$. Moreover, for every integer $0 \le k \le n$, after k steps $|S_1 \cup S_2 \cup U \cup W| = k$.

Finally, observe that at any step, the connected component C of G_p currently explored (that is, the one containing vertices in U) stays connected when restricted to $S_1 \cup U$, and every vertex of C in S_2 sends at least two edges to $S_1 \cup U$. Therefore, $|S_2| \le |E(G_p)| - |V(G_p)|$.

2.3 Upper bound on the excess of G_p

In order to find a quantitative bound for the size of $|S_2|$ from the DFS algorithm in the previous subsection, we use several results on site percolation on pseudo-random graphs to derive an upper bound for $|E(G_p)|$ and lower bound for $|V(G_p)|$. The first result relates the spectral ratio of G to the vertex-expansion (in G) of sets which lie in G_p.

Lemma 2.3 (Lemma 2.4 in [9], see also [21]). *Let $\alpha := \alpha(\epsilon) \in (0, \epsilon^8)$ be a constant. Suppose that $\delta \le \alpha^{2/\alpha}$. Then, **whp** V_p does not contain a set S with $|S| = m$, $\frac{\alpha n}{d} \le m \le \frac{n}{3d}$, such that $|N_G(S)| < (1-\alpha)\left(dm - \frac{d^2m^2}{2n}\right)$.*

Let us now recall the notion of *excess*. For a connected graph H, we define the excess of H as $\mathrm{exc}(H) := |E(H)| - |V(H)| + 1$. If H has more than one connected component, then we set $\mathrm{exc}(H)$ to be the sum of the excesses of each of its components. Recall that S_2 from the DFS algorithm in the previous subsection satisfies $|S_2| \le |E(G_p)| - |V(G_p)| \le \mathrm{exc}(G_p)$. Thus, to control $|S_2|$, it will be of use for us to estimate the typical excess of G_p. We require some additional results on site percolation on pseudo-random graphs.

Given G_p, we denote by L_1 the largest component of G_p. Let us further define x to be the unique solution in $(0, 1)$ of

$$x = (1+\epsilon)(1 - \exp\{-x\}). \tag{1}$$

We note that $x = 2\epsilon - \frac{2\epsilon^2}{3} + O(\epsilon^3)$ (see [9, Equation (4)]). The following theorem estimates the typical order of L_1.

Theorem 2.4 (Theorem 2 of [9]). *Let $\alpha := \alpha(\epsilon) \in (0, \epsilon^8)$ be a constant. Suppose that $\delta \le \alpha^{2/\alpha}$. Then, **whp**,*

$$\left| |V(L_1)| - \frac{xn}{d} \right| \le \frac{7\alpha n}{d},$$

where x is as in (1).

The next result estimates the typical number of edges in L_1.

Theorem 2.5 (Theorem 4 of [9]). *Let $\alpha := \alpha(\epsilon) \in (0, \epsilon^8)$ be a constant. Suppose that $\delta \le \alpha^{2/\alpha}$. Then, **whp**,*

$$\left| e(L_1) - \frac{((1+\epsilon)^2 - (1+\epsilon-x)^2)n}{2d} \right| \le \frac{8\alpha^{1/4} n}{d},$$

where x is as in (1).

Finally, the following result estimates the typical number of edges in G_p which are neither in L_1 nor in isolated trees.

Theorem 2.6 (Lemma 6.4 of [9])**.** *Let $\alpha := \alpha(\epsilon) \in (0, \epsilon^8)$ be a constant. Suppose that $\delta \le \alpha^{2/\alpha}$. Then,* ***whp****, the number of edges in G_p which are in components that are neither the giant component nor isolated trees is at most $\frac{7\alpha^{1/4}n}{d}$.*

With these three results at hand, we can now estimate the typical excess of G_p.

Lemma 2.7. *Suppose that $\delta \le \epsilon^{24/\epsilon^{12}}$. Then,* ***whp****, $\text{exc}(G_p) = O(\epsilon^3)n/d$.*

Proof. Note that under the above assumption, we can take α from Theorems 2.4, 2.5, and 2.6 to be $\alpha = \epsilon^{12}$. Then, by Theorems 2.4 and 2.5,

$$\begin{aligned}\text{exc}(G_p[L_1]) &\le \frac{((1+\epsilon)^2 - (1+\epsilon-x)^2)n}{2d} + \frac{8\epsilon^{12/4}n}{d} - \frac{xn}{d} + \frac{7\epsilon^{12}n}{d} + 1\\ &\le \frac{\left(\epsilon x - x^2/2\right)n}{d} + \frac{O(\epsilon^3)n}{d} = \frac{O(\epsilon^3)n}{d},\end{aligned}$$

where the last equality follows from $\epsilon x - x^2/2 = O(\epsilon^3)$. Since isolated trees have no excess, by the above together with Theorem 2.6, we conclude that **whp**

$$\text{exc}(G_p) \le \text{exc}(G_p[L_1]) + \frac{7\epsilon^{12/4}n}{d} = \frac{O(\epsilon^3)n}{d},$$

as required. □

3 Existence of induced paths and cycles

We begin by finding an induced path in the percolated graph G_p. To do so, we apply the modified DFS algorithm presented in Section 2 and obtain a lower bound on the size of the set of active vertices, which spans an induced path. The proof relies on Lemma 2.7 and Lemma 2.3.

Proof of Theorem 2. Let $\delta = \epsilon^{24/\epsilon^{12}}$. Run the DFS algorithm described in Section 2 on G. We claim that after ϵn steps in the algorithm, **whp**, $|U| \ge \frac{\epsilon^2 n}{3d}$, and thus, as U forms an induced path in G_p, the statement follows.

Indeed, after ϵn steps, $|U \cup S_1 \cup S_2| \sim Bin(\epsilon n, p)$. Thus, by Lemma 2.1, with probability at least $1 - 2\exp\left\{-\frac{(n/d)^{4/3}}{4\epsilon n/d}\right\}$ we have that $\left||U_1 \cup S_1 \cup S_2| - \frac{(1+\epsilon)\epsilon n}{d}\right| \le (n/d)^{2/3}$. Further, by Lemma 2.7 together with our assumption on δ, we have that **whp** $|S_2| = O(\epsilon^3)n/d$. Assume towards contradiction that $|U| < \frac{\epsilon^2 n}{3d}$. Then, **whp**,

$$|S_1| \ge \frac{(1+\epsilon)\epsilon n}{d} - n^{2/3} - O(\epsilon^3)\frac{n}{d} - \frac{\epsilon^2 n}{3d} \ge \frac{(\epsilon + 3\epsilon^2/5)n}{d}.$$

On the other hand, **whp**,

$$|S_1| \le |U \cup S_1 \cup S_2| \le 4\epsilon n/(3d) < n/(3d),$$

and thus by Lemma 2.3 applied with $\alpha = \epsilon^{12}$ (which is possible by our assumption on δ), we have that **whp**

$$|N_G(S_1)| \ge (1 - \epsilon^{12})\left(\epsilon + 3\epsilon^2/5 - \frac{1}{2}(\epsilon + 3\epsilon^2/5)^2\right)n > \epsilon n,$$

a contradiction, since $N_G(S_1) \subseteq S_2 \cup U \cup W$ and $|S_2 \cup U \cup W| \le |S_1 \cup S_2 \cup U \cup W| \le \epsilon n$. □

With this result at hand, we are ready to prove Theorem 1. We take the induced path obtained in Theorem 2 and, using Lemma 2.2, "close" (a substantial part of) it into a cycle. By carrying this out carefully, we ensure that the resulting cycle is induced and is long enough.

Proof of Theorem 1. Let $\epsilon, \delta(\epsilon) > 0$ be as in the statement of Theorem 2. We may assume that $\delta_1 \leq \delta(\epsilon)$. Let $p = \frac{1+\epsilon}{d}$. Then, by Theorem 2, **whp** G_p contains an induced path of length at least $\frac{\epsilon^2 n}{3d}$. Thus, deterministically, G contains an induced path P on $k = \frac{\epsilon^2 n}{3d}$ vertices (note that for the latter to hold, we merely needed the first statement to hold with positive probability).

Let $P = \{v_1, \ldots, v_k\}$. Let $P_1 = \{v_1, \ldots, v_{k/3}\}$ be the first $\frac{\epsilon^2 n}{9d}$ vertices of P, and let $P_2 = \{v_{2k/3+1}, \ldots, v_k\}$ be the last $\frac{\epsilon^2 n}{9d}$ vertices of P. Let $P' = \{v_{k/2-k/20+1}, \ldots, v_{k/2+k/20}\}$ be the set of $\frac{\epsilon^2 n}{30d}$ vertices at the middle of P. By Lemma 2.3, $|N_G(P_1)|, |N_G(P_2)| \geq \frac{\epsilon^2 n}{10}$. Since G is d-regular, $|N_G(P')| \leq d|P'| = \frac{\epsilon^2 n}{30}$. Set $N_1 = N_G(P_1) \setminus (P \cup N_G(P'))$ and similarly $N_2 = N_G(P_2) \setminus (P \cup N_G(P'))$. Note that $|N_1|, |N_2| \geq \frac{\epsilon^2 n}{30}$. In particular, by Lemma 2.2, $e(N_1, N_2) \geq \epsilon^5 dn > 0$.

Let $u_1 \in N_1$, $u_2 \in N_2$ be such that $u_1 u_2 \in E(G)$. Consider $L = \{v_1, \ldots, v_{k/2-k/20}\}$ and $R = \{v_{k/2+k/20+1}, \ldots, v_k\}$. If u_1 has a neighbour in both L and R, we let w_1 be the neighbour of u_1 from L closest to P', that is, to $v_{k/2-k/20}$. Similarly, let w_2 be the neighbour of u_1 from R closest to P'. Recall that u_1 has no neighbours in P', hence u_1, together with the subpath of P starting at w_1 and ending at w_2 forms an induced cycle in G whose length is at least $|P'| = \frac{\epsilon^2 n}{30d}$. This completes the proof, with $\delta_2 = \epsilon^2/30$. A similar argument applies when u_2 has a neighbour in both L and R.

Now, since $u_1 \in N_1 \subseteq L$ and $u_2 \in N_2 \subseteq R$, it remains to consider the case where u_1 has a neighbour in L but not in R and u_2 has a neighbour in R but not in L. Let w_1 as the neighbour of u_1 from L closest to P', and similarly, define w_2 as the neighbour of u_2 from R closest to P'. $u_1 u_2$ together with the subpath of P starting at w_1 and ending at w_2 forms an induced cycle in G whose length is at least $|P'| = \frac{\epsilon^2 n}{30d}$ which again completes the proof, with $\delta_2 = \epsilon^2/30$. $\square$

Remark 3.1. We note that, with a bit more effort, one can show the typical existence of an induced cycle of length $\Omega(\epsilon^2 n/d)$ in G_p. Let us give a sketch of the proof. We may employ a sprinkling argument, setting $p_2 = \frac{\epsilon^3}{d}$, and p_1 to be such that $(1-p_1)(1-p_2) = 1-p$. We then have that G_p has the same distribution as $G[V_{p_1} \cup V_{p_2}]$, and $p_1 \geq \frac{1+\epsilon-\epsilon^3}{d}$. By Theorem 2, **whp** there is an induced path $G[V_{p_1}]$ of length at least $\Omega(\epsilon^2 n/d)$. Similar to the above proof, one can consider N_1, N_2 the neighbourhoods in G of some sufficiently long prefix and suffix of the path, which have **whp** at least $\epsilon^5 dn$ edges between them. We can then consider sequentially every vertex in N_1, and whether it falls into V_{p_2}, noting that a typical vertex in N_1 will have $\Omega(d)$ neighbours in N_2. Upon reaching a sufficiently large subset $W \subseteq N_1$ in V_{p_2} whose neighbourhood (in G) in N_2 is of order $d|W|$, we may percolate N_2 with probability p_2, and **whp** obtain an edge in $G[V_{p_2}]$ between N_1 and N_2, and then complete the proof as before.

4 Non-isomorphic induced subgraphs

The proof of Theorem 4 will utilise Theorem 2, together with Lemma 2.3 to construct a long induced path P and a set W' of size $\Theta(d|P|) = \Theta(n)$, such that each vertex in W' has a unique neighbour in P, and $N_P(W')$ avoids the endpoints of P. their neighbours, and has pairwise distances at least 4 along P. We then show that the induced subgraph on $V(P) \cup W'$ contains the required number of non-isomorphic induced subgraphs.

Proof of Theorem 4. Let $\epsilon, \delta(\epsilon) > 0$ be as in the statement of Theorem 2. We may assume that $\delta_1 \le \delta(\epsilon)$, and in particular, $\delta_1 \le \epsilon^{8/\epsilon^4}$. Let $p = \frac{1+\epsilon}{d}$. Then, by Theorem 2, **whp** G_p contains an induced path P of length exactly $\frac{\epsilon^2 n}{3d}$. Furthermore, by Lemma 2.3 and by our assumption on δ_1, **whp** $|N_G(P)| \ge (1-\epsilon^4)\left(\frac{\epsilon^2 n}{3} - \frac{\epsilon^4 n}{18}\right) \ge \frac{\epsilon^2 n}{3} - \epsilon^4 n$. Thus, there exists (deterministically) an induced path $P = \{v_1, v_2, \dots, v_k\}$ in G on $k = \frac{\epsilon^2 n}{3d}$ vertices, such that $|N_G(P)| \ge \frac{\epsilon^2 n}{3} - \epsilon^4 n$.

Now, since G is d-regular, we have that $\sum_{v \in N_G(P)} d(v, P) \le \sum_{v \in P} d(v) = \epsilon^2 n/3$. Suppose towards contradiction that there are more than $2\epsilon^4 n$ vertices in $N_G(P)$ each having at least two neighbours in P. Then,

$$\sum_{v \in N_G(P)} d(v, P) \ge |N_G(P)| + 2\epsilon^4 n > \epsilon^2 n/3,$$

which is a contradiction. Hence, there exists a set $U \subseteq N_G(P)$ of size at least $\frac{\epsilon^2 n}{3} - 3\epsilon^4 n$, such that every $u \in U$ has exactly one neighbour in P. In particular, this implies that there are at least $\frac{\epsilon^2 n}{10d}$ vertices in P which have at least $\frac{d}{10}$ neighbours in U. Let us denote the set of these vertices in P by W. Let us then choose a subset $W' \subseteq W$, such that $v_1, v_2, v_{k-1}, v_k \notin W'$, and the distance in P between any $u, u' \in W'$ is at least 4. Note that we can choose such W' with $|W'| \ge |W|/5 \ge \frac{\epsilon^2 n}{50d}$. Crucially, observe that every $u \in N_G(W') \cap U := N_{W'}$ has a unique neighbour $v \in W'$ on P.

Now, let $H := G[V(P) \cup N_{W'}]$. Let $A \ne A'$ be subsets of $N_{W'}$, such that $G[V(P) \cup A] \cong G[V(P) \cup A']$. Let $\psi : V(P) \cup A \to V(P) \cup A'$ be an isomorphism between these two graphs. We claim that $\psi \mid_{V(P)}$ is either the trivial automorphism of P, or the non-trivial involution automorphism of P. First, note that the only two vertices in H which are of degree one and have a neighbour of degree two are v_1 and v_k. Indeed, $v_2, v_{k-1} \notin W'$ and thus have degree two, and any other vertex of degree one is in $N_{W'}$, and thus has a neighbour on $V(P)$ of degree at least three. It thus suffices to show that ψ sends vertices of degree two in $V(P)$ to (possibly other) vertices of degree two in the same set $V(P)$.

Suppose towards contradiction that there exists $u \in V(P)$ such that $d_H(u) = 2 = d_{G[V(P) \cup A]}(u)$ and $\psi(u) \in A'$. Since every vertex in $A' \subseteq N_{W'}$ has a unique neighbour in P and since $\psi(u)$ has degree two in $G[V(P) \cup A']$, we have that $\psi(u)$ has a neighbour in A', which we denote by v. Let x_1 be the unique neighbour of $\psi(u)$ in P, and let x_2 be the unique neighbour of v in P, noting that both x_1 and x_2 have degree at least three in $G[V(P) \cup A']$. We then have that $\psi(u)$ belongs to the path $x_1 \psi(u) v x_2$ in $G[V(P) \cup A'] \subseteq H$. However, by construction of H, since $d_H(u) = 2$, the two closest vertices to u which have degree at least 3 in $G[V(P) \cup A]$ (and therefore, in H) lie on P, and thus are at distance at least four from each other — contradiction.

Therefore, any automorphism ψ has that $\psi \mid_{V(P)}$ is one of the two involution automorphisms. Hence, for a fixed automorphism of P, any two subsets $A, A' \in N_{W'}$ having $|A' \cap N(v) \cap N_{W'}| \ne |A' \cap N(v) \cap N_{W'}|$ for some $v \in W'$ cannot span (together with P) isomorphic subgraphs. We thus conclude that, for every tuple $(0 \le s_v \le d/10)_{v \in W'}$, there exists at most one *other* tuple $(0 \le s'_v \le d/10)_{v \in W'}$ satisfying the following: there exist two subsets $A, A' \subset N_{W'}$ such that $|A \cap N(v) \cap N_{W'}| = s_v$ and $|A' \cap N(v) \cap N_{W'}| = s'_v$ for every $v \in W'$ and $G[V(P) \cup A], G[V(P) \cup A']$ are isomorphic. Therefore,

$$\mu(G) \ge \frac{1}{2}(d/10)^{|W'|} \ge \frac{1}{2}\left(\frac{d}{10}\right)^{\frac{\epsilon^2 n}{50d}} \ge \exp\left(\frac{\delta_2 n \log d}{d}\right),$$

for small enough δ_2, as required. □

5 Constructions

Both the proof of Theorem 3 and of Theorem 5 will follow from quite similar constructions. Let us first collect some notation and results about the spectra of graphs. We refer the reader to [7, 8] for a comprehensive study of the spectra of graphs and graph operations.

Lemma 5.1 (see, e.g., [7]). *The eigenvalues of the complete graph on n vertices are $n-1$ of multiplicity 1, and -1 of multiplicity $n-1$. The eigenvalue of the empty graph on n vertices is zero with multiplicity n.*

We will make use of the following graph operation. Let $r > 0$ be an integer. Let G and H be two graphs, such that H is r-regular. The *lexicographic product* (denoted by $\mathrm{lex}(G, H)$) of G and H has vertex set $V(G) \times V(H)$ and for every $u, v \in V(G)$ and $x, y \in V(H)$, (u, x) is adjacent to (v, y) if and only if either $uv \in E(G)$ or $u = v \wedge xy \in E(H)$.

One can think of such an operation as taking a graph G and replacing each vertex of G by a copy of H; in particular, when H is the empty graph, this is simply the blow-up operation. The lexicographic product satisfies the following property (see, e.g., [7]).

Lemma 5.2. *Suppose $\lambda_1, \dots, \lambda_n$ are the eigenvalues of a graph G on n vertices and $\mu_1 = r, \dots, \mu_m$ are the eigenvalues of an r-regular graph H on m vertices. Then, the eigenvalues of the adjacency matrix of the lexicographic product $\mathrm{lex}(G, H)$ are $\lambda_i m + r$ of multiplicity 1 for $1 \le i \le n$ and μ_j of multiplicity n for $2 \le j \le m$.*

Let us begin with the proof of Theorem 5, whose construction is slightly simpler.

Proof of Theorem 5. Let $\delta > 0$ be a constant. Let $d \in \mathbb{N}$ be sufficiently large, and let $n \in \mathbb{N}$ be such that d, n satisfy the parity assumptions which are implicit below (in particular, n is divisible by d). Let $d_0 \coloneqq d_0(\delta)$ be the smallest integer satisfying $d_0 \ge 3$ and $\sqrt{d_0} \ge \frac{3}{\delta}$.

Assume first, for the sake of clarity of presentation, that d is divisible by d_0. Let $n_0 \coloneqq n \cdot d_0/d$, and let H_1 be an (n_0, d_0, λ_0)-graph with $\lambda_0 \le 3\sqrt{d_0}$ — indeed, such a graph exists, since a random d_0-regular graph on n_0 vertices typically satisfies this (see, e.g., [23]). We note that we assume here that n_0 is sufficiently large with respect to d_0, and in turn, n is sufficiently large with respect to d. Let H_2 be the empty graph on d/d_0 vertices. Set $G = \mathrm{lex}(H_1, H_2)$. By Lemmas 5.1 and 5.2 we have that G is an (n, d, λ)-graph with $\lambda = \lambda_0 \cdot \frac{d}{d_0}$. Hence, by our choice of d_0,

$$\frac{\lambda}{d} = \frac{\lambda_0}{d_0} \le \frac{3}{\sqrt{d_0}} \le \delta,$$

and thus satisfies the requirement of Theorem 5.

Let $V(H_1) = \{v_1, \dots, v_{n_0}\}$. Let $V_1, \dots, V_{n_0}$ be subsets of $V(G)$, where V_i is the blow-up of v_i. Note that, by construction, given two graphs $G_1, G_2 \subseteq G$, if for every $i \in [n_0]$, $|V(G_1) \cap V_i| = |V(G_2) \cap V_i|$, then $G_1 \cong G_2$. We thus conclude that the number of non-isomorphic induced subgraphs of G is bounded from above by the number of different tuples $(0 \le s_i \le d/d_0)_{i=1}^{n_0}$. Hence,

$$\mu(G) \le \left(\frac{d}{d_0} + 1\right)^{n_0} = \exp\left\{\frac{nd_0}{d}\log(d/d_0 + 1)\right\} \le \exp\left\{\frac{Cn\log d}{d}\right\},$$

as required.

In the general case, when d is not divisible by d_0, let us write $d = qd_0 + r$ where $q \in \mathbb{N}$ and $1 \le r \le d_0 - 1$. Suppose first that qr is even. We then set H_2 to be a graph on q vertices, formed by taking a disjoint union of $\lfloor \frac{q}{r+1} \rfloor - 1$ cliques of size $r+1$, and on the remaining $q - (r+1)\left(\lfloor \frac{q}{r+1} \rfloor - 1\right)$

vertices we draw an arbitrary r-regular graph (indeed, such a graph exists since qr and $r(r+1)$ are both even). If qr is odd, then both q and r are odd, and thus $q-1$ is even. Let us write $d = (q-1)d_0 + (d_0 + r)$. We then set H_2 to be a graph on $q-1$ vertices, formed by taking a disjoint union $\lfloor \frac{q-1}{d_0+r+1} \rfloor - 1$ cliques of size d_0+r+1, and on the remaining $(q-1)-(d_0+r+1)\left(\lfloor \frac{q-1}{d_0+r+1} \rfloor - 1\right)$ vertices we draw an arbitrary (d_0+r)-regular graph (indeed, such a graph exists since $q-1$ and $(d_0+r)(d_0+r+1)$ are both even). The rest of the proof, for both cases (both choices of H_2), is quite similar to the case where d is divisible by d_0. $\square$

The construction for the proof of Theorem 3 is very similar, where instead of replacing every vertex with an independent set, we will replace every vertex with a copy of the complete graph.

Proof of Theorem 3. Let $\delta > 0$ be a constant. Let $d \in \mathbb{N}$ be sufficiently large, and let $n \in \mathbb{N}$ be such that d, n satisfy the parity assumptions which are implicit below (in particular, n is divisible by d). Let d_0 be the smallest integer satisfying $d_0 \geq 3$ and $\sqrt{d_0} \geq \frac{4}{\delta}$.

Assume first, for the sake of clarity of presentation, that $d+1$ is divisible by d_0+1. Let $k \coloneqq \frac{d+1}{d_0+1}$, and let $n_0 = n/k$. Let H be an (n_0, d_0, λ_0)-graph satisfying that $\lambda_0 \leq 3\sqrt{d_0}$ (indeed, a random d_0-regular graph on n_0 vertices typically satisfies this). In particular, here too we assume that n_0 is sufficiently large with respect to d_0, and in turn, n is sufficiently large with respect to d. Set $G = \mathrm{lex}(H, K_k)$. Noting that $kd_0 + k - 1 = d$, we then have that G is a d-regular graph on $kn_0 = n$ vertices. Furthermore, by Lemmas 5.1 and 5.2 we have that the second largest eigenvalue of G, λ, satisfies that

$$\frac{\lambda}{d} = \frac{k\lambda_0 + k - 1}{kd_0 + k - 1} \leq \frac{4}{\sqrt{d_0}} \leq \delta,$$

and thus G satisfies the assumptions of Theorem 3.

By the construction of G, we have that a largest independent set of G is of size at most $n_0 = (d_0+1)n/d$. Therefore, a longest induced path in G is of length at most $2(d_0+1)n/d$. Setting $C \coloneqq C(\delta) = 2(d_0+1)$ completes the proof.

In the general case, where $d+1$ is not divisible by d_0+1, let r be the residue of $d+1$ modulo d_0+1. Let $k \coloneqq \frac{d+1-r}{d_0+1}$, and let H_2 be a copy of K_k with an r-regular graph removed. We then take $G = \mathrm{lex}(H, H_2)$. The rest of the proof is quite similar to the case where $d+1$ is divisible by d_0+1. $\square$

6 Concluding Remarks

In this paper we studied two quantities: the size of a largest hole, and the number of non-isomorphic induced subgraphs of (n, d, λ)-graphs under the standard mild assumption that the spectral ratio λ/d is bounded by a small constant. Concretely, we have shown a tight bound on the size of a largest hole in (n, d, λ)-graphs, which is of order n/d. We have also shown that the number of non-isomorphic induced subgraphs of (n, d, λ)-graphs is of order at least $\exp(\Theta(n \log d/d))$ and that this bound is tight up to a multiplicative factor in the exponent.

It is plausible to expect similar results to hold for other definitions of pseudo-random graphs (see [23, Section 2] for a comprehensive survey on the subject). It is also natural to ask how to generalise our result on induced paths to arbitrary induced bounded degree trees. This question has been studied in random graphs, see for example [2, 14, 15].

Moreover, as discussed earlier in this paper, it would be interesting to relate the spectral ratio λ/d (where we allow the ratio to depend on d) to the size of a largest hole of an (n, d, λ)-graph G.

Supported by the fact that when $\lambda \ll d$, the independence number of an (n, d, λ)-graph G satisfies $\alpha(G) = \Omega(n \log(d/\lambda)/d)$ (see [23, Proposition 4.6] for example), we tend to believe the following:

Conjecture 6.1. *There exists a constant $\delta > 0$ such that the following holds. For any integers n, d and for any (n, d, λ)-graph G such that $\lambda/d \leq \delta$, G contains an induced cycle of length $\Omega(n \log(d/\lambda)/d)$.*

Acknowledgements. Part of this work was done while the fourth author was visiting Tel Aviv University, and he would like to thank the university for its hospitality. The second author was supported in part by NSF-BSF grant 2023688.

References

[1] H. L. Abbott and M. Katchalski. On the snake in the box problem. *J. Comb. Theory, Ser. B*, 45(1):13–24, 1988.

[2] M. Akhmejanova, V. Kozhevnikov, and M. Zhukovskii. Maximum induced trees and forests of bounded degree in random graphs. Preprint, arXiv:2408.15215 [math.CO], 2024.

[3] N. Alon and B. Bollobás. Graphs with a small number of distinct induced subgraphs. *Discrete Math.*, 75(1-3):23–30, 1989. Graph theory and combinatorics (Cambridge, 1988).

[4] N. Alon and F. R. K. Chung. Explicit construction of linear sized tolerant networks. *Discrete Math.*, 72(1-3):15–19, 1988.

[5] N. Alon and A. Hajnal. Ramsey graphs contain many distinct induced subgraphs. *Graphs Combin.*, 7(1):1–6, 1991.

[6] N. Alon and J. H. Spencer. *The probabilistic method.* Wiley-Intersci. Ser. Discrete Math. Optim. Hoboken, NJ: John Wiley & Sons, 4th edition edition, 2016.

[7] S. Barik, D. Kalita, S. Pati, and G. Sahoo. Spectra of graphs resulting from various graph operations and products: a survey. *Spec. Matrices*, 6:323–342, 2018.

[8] D. M. Cvetković, M. Doob, and H. Sachs. *Spectra of graphs. Theory and application*, volume 87 of *Pure Appl. Math., Academic Press.* New York, NY; Academic Press, 1980.

[9] S. Diskin and M. Krivelevich. Site percolation on pseudo-random graphs. *Random Struct. Algorithms*, 63(2):406–441, 2023.

[10] N. Draganić, S. Glock, and M. Krivelevich. The largest hole in sparse random graphs. *Random Struct. Algorithms*, 61(4):666–677, 2022.

[11] N. Draganić and P. Keevash. Long induced paths in expanders. *Comb. Probab. Comput.*, 34(2):276–282, 2025.

[12] K. Dutta and C. R. Subramanian. On induced paths, holes, and trees in random graphs. *SIAM J. Discrete Math.*, 37(1):279–303, 2023.

[13] N. Enriquez, G. Faraud, L. Ménard, and N. Noiry. Long induced paths in a configuration model. Preprint, arXiv:2106.11130 [math.PR] (2021), 2021.

[14] P. Erdős and Z. Palka. Trees in random graphs. *Discrete Math.*, 46:145–150, 1983.

[15] W. Fernandez de la Vega. The largest induced tree in a sparse random graph. *Random Struct. Algorithms*, 9(1-2):93–97, 1996.

[16] J. Friedman. A proof of Alon's second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.*, 195(910):viii+100, 2008.

[17] A. Frieze. Hamilton Cycles in Random Graphs: a bibliography. Preprint, arXiv:1901.07139 [math.CO] (2019), 2019.

[18] A. M. Frieze and B. Jackson. Large holes in sparse random graphs. *Combinatorica*, 7:265–274, 1987.

[19] I. Hartarsky. Maximal bootstrap percolation time on the hypercube via generalised snake-in-the-box. *Electron. J. Comb.*, 25(3):research paper p3.10, 12, 2018.

[20] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Am. Math. Soc., New Ser.*, 43(4):439–561, 2006.

[21] M. Krivelevich. The phase transition in site percolation on pseudo-random graphs. *Electron. J. Comb.*, 23(1):research paper p1.12, 11, 2016.

[22] M. Krivelevich. Expanders – how to find them, and what to find in them. In *Surveys in combinatorics 2019. Papers based on the 27th British combinatorial conference, University of Birmingham, Birmingham, UK, July 29 – August 2, 2019*, pages 115–142. Cambridge: Cambridge University Press, 2019.

[23] M. Krivelevich and B. Sudakov. Pseudo-random graphs. In *More sets, graphs and numbers. A salute to Vera Sós and András Hajnal*, pages 199–262. Berlin: Springer. Budapest: János Bolyai Mathematical Society, 2006.

[24] M. Krivelevich and B. Sudakov. The phase transition in random graphs: A simple proof. *Random Structures Algorithms*, 43(2):131–138, 2013.

[25] M. Krivelevich and M. Zhukovskii. Non-isomorphic subgraphs in random graphs. Preprint, arXiv:2505.14623 [math.CO] (2025), 2025.

[26] T. Łuczak. The size of the largest hole in a random graph. *Discrete Math.*, 112(1-3):151–163, 1993.

[27] V. Müller. Probabilistic reconstruction from subgraphs. *Commentat. Math. Univ. Carol.*, 17:709–719, 1976.

[28] W. C. S. Suen. On large induced trees and long induced paths in sparse random graphs. *J. Comb. Theory, Ser. B*, 56(2):250–262, 1992.

[29] G. Zémor. An upper bound on the size of the snake-in-the-box. *Combinatorica*, 17(2):287–298, 1997.