Sparse pancyclic subgraphs of random graphs

Yahav Alon [*]   Michael Krivelevich [†]

November 29, 2024

Abstract

It is known that the complete graph K_n contains a pancyclic subgraph with $n + (1 + o(1)) \cdot \log_2 n$ edges, and that there is no pancyclic graph on n vertices with fewer than $n + \log_2(n-1) - 1$ edges. We show that, with high probability, $G(n,p)$ contains a pancyclic subgraph with $n + (1 + o(1)) \log_2 n$ edges for $p \geq p^*$, where $p^* = (1 + o(1)) \ln n/n$, right above the threshold for pancyclicity.

1 Introduction

Say that a graph G is pancyclic if G contains a cycle of every length between 3 and $|V(G)|$. See monograph [6] for generic information on pancyclic graphs. In his influential paper on pancyclic graphs, Bondy [2] asked what is the minimum number of edges in a pancyclic n-vertex graph. This can be rephrased as the minimum number of edges in a pancyclic subgraph of K_n, which motivates the following definition.

Definition 1. *Say that a pancyclic graph G on n vertices has pancyclicity excess k, and denote $Pex(G) = k$, if the minimum number of edges in a pancyclic subgraph of G is $n + k$.*

In other words, a pancyclic subgraph of G achieving the minimum number of edges is formed by a Hamilton cycle and $\text{Pex}(G)$ additional chords. In his paper, Bondy stated that, for every n,

$$\log_2(n-1) - 1 \leq \text{Pex}(K_n) \leq \log_2 n + \log^* n + O(1),$$

and did not provide a proof. Shi [10] later asserted the lower bound, by showing that an n-vertex graph with $n + k$ edges contains at most $2^{k+1} - 1$ distinct cycles, so every subgraph of K_n with fewer than $n + \log_2(n-1) - 1$ edges must have fewer than $2^{\log_2(n-1)} - 1 = n - 2$ cycles in total, regardless of their lengths. On the other hand, there are constructions for every n of an n-vertex pancyclic graph with $\log_2 n + \log^* n + O(1)$ chords (see e.g. [6], Chapter 4.5), so $\text{Pex}(K_n) \leq \log_2 n + \log^* n + O(1)$. What is the exact value of $\text{Pex}(K_n)$ within this range is still an open question.

In this paper, we study the typical behaviour of $\text{Pex}(G)$, for $G \sim G(n,p)$. Cooper and Frieze [4] showed that, for $p \in [0,1]$, the limiting probability of $G \sim G(n,p)$ being pancyclic is

$$\lim_{n\to\infty} \mathbb{P}(G(n,p) \text{ is pancyclic}) = \begin{cases} 1 & \text{if } np - \log n - \log\log n \to \infty; \\ e^{-e^{-c}} & \text{if } np - \log n - \log\log n \to c; \\ 0 & \text{if } np - \log n - \log\log n \to -\infty. \end{cases}$$

[*]School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, 6997801, Israel. Email: yahavalo@tauex.tau.ac.il.

[†]School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, 6997801, Israel. Email: krivelev@tauex.tau.ac.il. Partially supported by USA-Israel BSF grant 2018267.

Here and later, if the base of the logarithm is not stated then it is the natural base. The above expression is also the limiting probability of G being Hamiltonian, and the limiting probability of $\delta(G) \geq 2$. In particular, the three properties have the same threshold.

Clearly, $\text{Pex}(G) \geq \log_2(n-1) - 1$ for every pancyclic graph G on n vertices. On the other hand, Cooper [3] showed that if p is above the pancyclicity threshold, then with high probability $G \sim G(n,p)$ is a so called *1-pancyclic* graph, that is, it contains a Hamilton cycle H with the property that, for every $\ell \in [3, n-1]$, there is an edge $e \in E(G)$ such that $H \cup \{e\}$ contains a cycle of length ℓ and a cycle of length $n - \ell + 2$. Observe that if G is a 1-pancyclic n-vertex graph then $\text{Pex}(G) \leq \lceil \frac{n-3}{2} \rceil$. So Cooper's result implies that $\text{Pex}(G(n,p)) \leq \lceil \frac{n-3}{2} \rceil$ with high probability, for all p above the pancyclicity threshold.

Our result in this paper shows that, for $G \sim G(n,p)$, the pancyclicity excess of G is, typically, very close to the above stated general lower bound.

Theorem 1. *There is $p^* = p^*(n) = (1 + \varepsilon(n)) \cdot \frac{\log n}{n}$, where $\varepsilon(n) = O\left(\frac{1}{\log \log n}\right)$, such that, if $p \geq p^*$ and $G \sim G(n,p)$, then with high probability G is pancyclic with Pex$(G) = (1 + o(1)) \cdot \log_2 n$.*

It is worth noting that we did not attempt to optimize the error term $\varepsilon(n)$, opting rather for a cleaner proof. We therefore leave the question of whether $\text{Pex}(G(n,p))$ also typically satisfies $\text{Pex}(G(n,p)) = (1 + o(1)) \cdot \log_2 n$ for all p above the pancyclicity threshold as an open question.

Paper structure In Section 2 we introduce definitions and notation required for the rest of the paper, as well as auxiliary results to be used in our proof. In Section 3 we introduce a construction of a subgraph of a given n-vertex graph, which, if successful, produces a subgraph with $n + (1 + o(1)) \cdot \log_2 n$ edges. In Section 4 we show that, with high probability, the construction is possible in $G(n,p)$ for $p \geq p^*$, and in Section 5 we complete the proof of Theorem 1 by showing that the constructed subgraph is pancyclic.

2 Preliminaries

2.1 Definitions and notation

The following graph theoretic notation is used throughout the paper.

Let G be a graph and $U, W \subseteq V(G)$ vertex subsets. We denote by $E_G(U,W)$ the set of edges of G with vertex in U and one vertex in W, and $e_G(U,W) = |E_G(U,W)|$. We let $G[U]$ denote the subgraph induced by G on the vertex subset U, by $E_G(U)$ the set of edges in $G[U]$, and by $e_G(U)$ its size. We denote by $N_G(U)$ the (external) neighbourhood of U, that is, the set of vertices in $V(G) \setminus U$ adjacent to a vertex of U. The degree of a vertex $v \in V(G)$, denoted by $d_G(v)$, is the number of edges of G incident to v.

We let $\mathcal{L}(G)$ denote the set of cycle lengths found in G, that is, $\mathcal{L}(G)$ is the set of integers ℓ such that G contains a cycle of length ℓ.

While using the above notation we occasionally omit G if the identity of the specific graph is clear from the context.

We occasionally suppress the rounding signs to simplify the presentation.

Finally, we require the following definition.

Definition 2. *A graph G is called a (k, α)-expander if every subset $U \subseteq V(G)$ with $|U| \leq k$ satisfies $|N_G(U)| \geq \alpha \cdot |U|$.*

2.2 Auxiliary results

Theorem 2.1 (Cycle lengths in $G(n,p)$, a corollary of Łuczak [9]). *Let $p = p(n)$ be such that $np \to \infty$ and let $G \sim G(n,p)$. Then, with high probability, $[3, 0.99n] \subseteq \mathcal{L}(G)$.*

Theorem 2.2 (Tree embeddings in expanders, a corollary of [7] as given in [1]). *Let N, Δ be integers, and let G be a graph. Assume that there exists an integer k such that*

1. *For every $U \subseteq V(G)$ with $|U| \leq k$ we have $|N_G(U)| \geq \Delta \cdot |U| + 1$;*

2. *For every $U \subseteq V(G)$ with $k < |U| \leq 2k$ we have $|N_G(U)| \geq \Delta \cdot |U| + N$.*

Then, for every $v \in V(G)$ and every rooted tree T with at most N vertices and maximum degree at most Δ, the graph G contains a copy of T rooted in v.

Lemma 2.1 (Hamiltonicity and expansion of $G(n,p)$, see e.g. [8], Section 4). *Let $p = p(n)$ be such that $np - \log n - \log\log n \to \infty$, and let $G \sim G(n,p)$. Then, with high probability, there is a subset $S \subseteq V(G)$ of $\frac{n}{4}$ vertices, such that for every $s \in S$ there is a subset $T_s \subseteq V(G)$ of $\frac{n}{4}$ vertices, and for every $t \in T_s$ there is a Hamilton path between s and t.*

3 The constructed pancyclic subgraph

We emulate (an approximation of) the construction in [6].

Definition 3. *Let G be a graph and $H \subseteq G$ be a Hamilton cycle, and let $2 \leq \ell \leq n-2$. We say that an edge $e \in E(G)$ is an* ℓ-shortcut *with respect to H if (at least) one of the two intervals on H that connects the two endpoints of e has length $\ell + 1$.*

The motivation behind this definition is that by using H and an ℓ-shortcut we can find a cycle of length $n - \ell$ in G, by replacing an interval of length $\ell + 1$ with a single edge (the ℓ-shortcut). In the construction described in [6], one creates a sparse pancyclic graph by taking an n-cycle H and K shortcuts $e_0, e_1, ..., e_K$, where K is such that $\frac{1}{2}n \leq 2^{K+1} + K - 1 \leq n$ and e_i is a 2^i-shortcut. Additionally, these shortcuts are consecutive on the cycle, so that e_i, e_{i+1} and their corresponding intervals intersect in a vertex v_i. By taking intervals from the cycle H and a subset of shortcuts, one can now encode a cycle of every length between n and $n - 2^{K+1} + 1$. Next, by adding the edge between the first vertex of e_0 and the second vertex of e_K, all cycle lengths between $K + 2$ and $2^{K+1} + K$ can be encoded. This leaves out only a subset of cycle lengths contained in $[5, K+1]$, and adding these lengths to the set of cycle lengths in the graph can be done by inserting $O(\log^* n)$ additional edges. For the full details of the construction, we refer the reader to [6] Chapter 4.5.

We approximate this construction by finding a Hamilton cycle and shortcuts to encode an interval of $L = \Omega\left(\frac{n}{\sqrt{\log n}}\right)$ consecutive cycle lengths. Like in the deterministic version, we will utilize binary encoding of the cycle lengths, so that the number of required shortcuts is $(1 + o(1)) \log_2 n$. Additionally, we will require the shortcuts to reside on a short interval of the cycle (where in the deterministic version they intersected each other in a vertex). Next, by adding certain edges to the subgraph we can add an interval of L cycle lengths with each such added edge. If the said additional edges are chosen well (which we will show is possible to do with high probability), one can get a union of $O(\sqrt{\log n})$ of these intervals that covers all the lengths between some initial length $\ell^* = (1 + o(1)) \log_2 n$ and n.

To handle cycle lengths shorter than ℓ^* we will show that, with high probability, almost all of them (that is, all but $o(\log n)$ cycle lengths in $[3,\ell^*]$) can be encoded by $o(\log n)$ carefully chosen shortcuts, this time utilizing an encoding in base $b = \lceil \log\log n \rceil$. The remaining unencoded cycle lengths, which constitute a subset of $[3,\ell^*]$ of size $o(\log n)$, can now be added one-by-one by using at most $o(\log n)$ additional edges, with high probability.

Let

$$p_1 = p_5 = \frac{2\log\log n}{n},\ p_2 = p_3 = \frac{50\log n}{n\cdot\log\log n},\ p_4 = \frac{\log n + 10\sqrt{\log n}}{n},$$

and let

$$p^* = p^*(n) = 1 - \prod_{i=1}^{5}(1-p_i).$$

Letting $\varepsilon(n) := \frac{n}{\log n}\cdot p^* - 1$ we get that $\varepsilon(n) = O(\frac{1}{\log\log n})$, and since the property $\mathrm{Pex}(G) \le k$ is monotone increasing, it suffices to prove that $\mathrm{Pex}(G) \le (1+o(1))\log_2 n$ holds with high probability for $G(n,p^*) \sim \bigcup_{i=1}^{5} G(n,p_i)$. We note that we did not attempt to optimize the value of $\varepsilon(n)$ determined by $p_1,...,p_5$, aiming rather for simplicity.

Denote

$$\ell_i := 2^i + 1,$$

and

$$\beta = \beta(n) := \frac{2(\log\log n)^2}{\log n},\ d = d(n) = \lfloor \log_{(5\beta)^{-1}}(n/200)\rfloor.$$

Note that

$$d = \lfloor \log_{(5\beta)^{-1}}(n/200)\rfloor = (1+o(1))\cdot\frac{\log_2(n/200)}{-\log_2(5\beta)} = (1+o(1))\cdot\frac{\log_2 n}{\log_2\log n}.$$

For $1 \le i \le 5$ let $G_i \sim G(n,p_i)$. We divide the construction into five steps, where in the i'th step we sample G_i to try and produce a subgraph $H_i \subseteq \bigcup_{j=1}^{i} G_j$. If the construction is successful, the produced subgraph H_5 will be pancyclic with $|E(H_5)| = n + (1+o(1))\cdot\log_2 n$. The steps of our construction are as follows.

1. Let

   $$K_0 := \lfloor \log_2\left(\frac{\log n}{6\log\log\log n}\right)\rfloor,$$

   and

   $$b := \lceil \log\log n\rceil,\ t := \lceil \log_b \log n\rceil.$$

   Find a set of vertex disjoint cycles $C_0,...,C_{K_0},C_{\text{short}}$ in G_1 of respective lengths $\ell_0+1,\ell_1+1,...,\ell_{K_0}+1, t\cdot b+1$. The first K_0+1 cycles will later become the first K_0+1 shortcuts, and their corresponding intervals, where the edges of C_{short} will become the shortcuts required to handle short cycles. For every $0 \le i \le K_0$, choose an arbitrary edge $e_i \in C_i$ to serve as the shortcut. Denote $H_1 = C_{\text{short}} \cup \bigcup_{i=0}^{K_0} C_i$.

2. For every $0 \le i \le K_0$, find a path of length $d+2$ in G_2 between the second vertex of e_i and the first vertex of e_{i+1} (where for $i = K_0$ the path is between e_{K_0} and e_0), so that the K_0+1 paths are pairwise vertex disjoint from each other, and internally vertex disjoint from $V(H_1)$. Call the cycle formed by the edges $e_0,...,e_{k_0}$ and the newly found paths connecting pairs these edges C^*, and denote $\ell^* := e(C^*), H_2 := H_1 \cup C^*$. We have

   $$\ell^* = (1+o(1))\cdot K_0\cdot d = (1+o(1))\cdot\log_2 n.$$

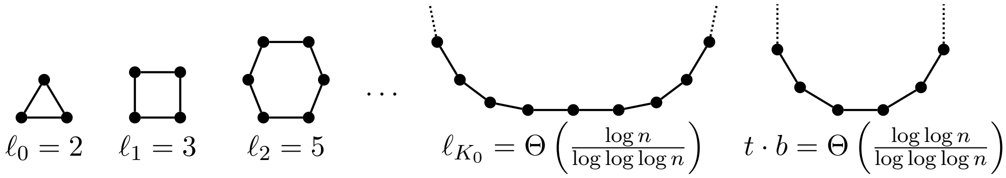

Figure 1: Step 1, with resulting graph H_1 depicted.

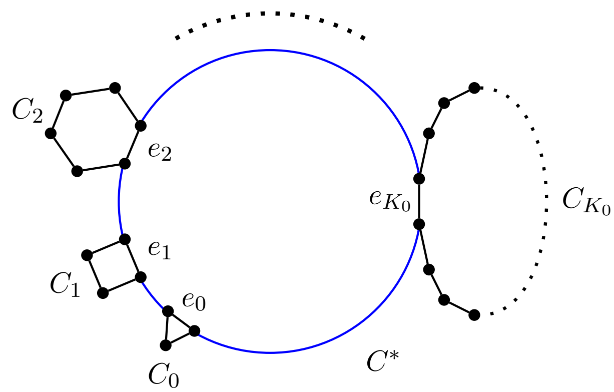

Figure 2: Step 2, with resulting graph H_2 depicted.

3. Let

$$K \coloneqq \lfloor \log_2 \left(\frac{n}{\sqrt{\log n}} \right) \rfloor,$$
$$L \coloneqq 2^{K+1} - 1,$$

so that $L+1 \in \left[\frac{n}{\sqrt{\log n}}, \frac{2n}{\sqrt{\log n}}\right]$. For $K_0 < i \le K$, construct the i'th shortcut by choosing an arbitrary (non-shortcut) edge e_i on C^*, and finding a path of length ℓ_i between its two vertices in G_3, such that these paths are internally vertex disjoint from each other and from $V(H_2)$. Letting C_i denote the cycle comprised of e_i and the ℓ_i-path in G_3 between its vertices, we get that the subgraph $C^* \cup \bigcup_{i=0}^{K} C_i$ contains all cycle lengths in the interval $[\ell^*, \ell^* + L]$. Choose an arbitrary edge $e^* \in C^* \setminus \{e_0, ..., e_K\}$.

Next, denote $E(C_{\text{short}}) = \{e_{\text{short}}\} \cup \{e_{i,j} \mid 0 \le i \le t-1, 0 \le j \le b-1\}$, with an arbitrary order. Find paths $\{P_{i,j} \mid 0 \le i \le t-1, 0 \le j \le b-1\}$ in G_3, where $P_{i,j}$ connects the endpoints of $e_{i,j}$, such that the paths are all internally vertex disjoint from each other, and from $V(H_2 \cup \bigcup_{i=K_0}^{K} C_i)$, and $P_{i,j}$ has length $d + 2 + j \cdot b^i$. Now the subgraph $C_{\text{short}} \cup \{P_{i,j} \mid 0 \le i \le t-1, 0 \le j \le b-1\}$ contains all cycle lengths in $\left[(d+b+1)\cdot t + 1, (d+b+1)\cdot t + b^t\right]$. Note that $b^t \ge \log n > \ell^*$, and that $(d+b+1)\cdot t = O\left(\frac{\log n}{\log\log\log n}\right)$.

Finally for this step, connect one vertex of e^* to one vertex of e_{short} by a path P^* of length $d+2$, internally disjoint from all previous construction, and denote $H_3 \coloneqq H_2 \cup P^* \cup \{P_{i,j}\}_{i,j} \cup \bigcup_{i=K_0}^{K} C_i$.

4. Construct a Hamilton cycle by connecting the vertex of e^* and the vertex of e_{short} that are not connected by P^* by a path P in G_4, whose internal vertices are exactly $V(G) \setminus V(H_3)$. Denote

$$C_H \coloneqq H_3 \cup P \setminus \left(\{e_0, ..., e_K, e^*, e_{\text{short}}\} \cup \{e_{i,j}\}_{i,j}\right).$$

Then the constructed $H_4 \coloneqq H_3 \cup P$ contains the Hamilton cycle C_H and $K + b \cdot t + 3$ additional edges, and all cycle lengths in $[(d+b+1)\cdot t + 1, \ell^* + L] \cup [n - L, n]$.

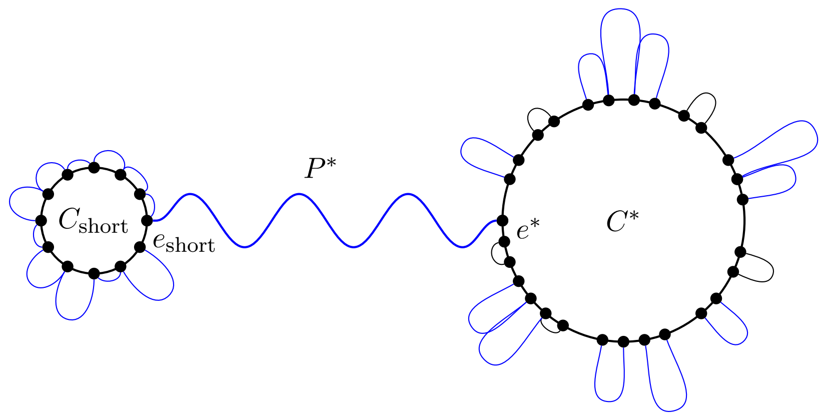

Figure 3: Step 3, with resulting graph H_3 depicted.

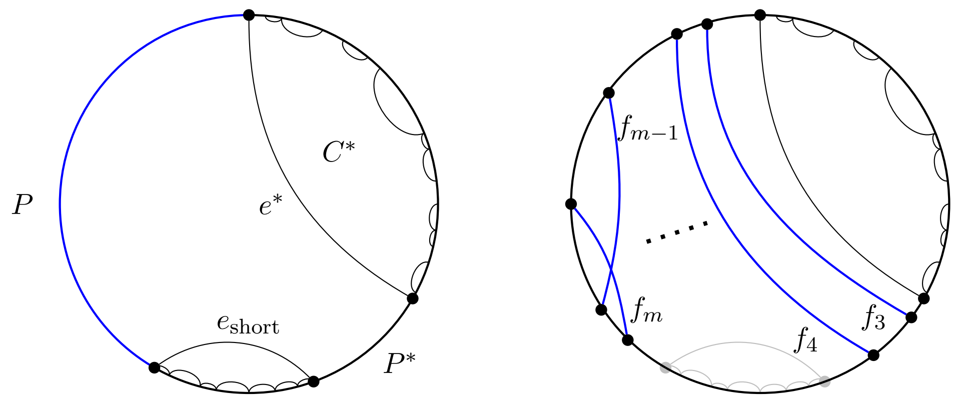

Figure 4: Steps 4 and 5, with resulting graph H_4 (left) and H_5 (right) depicted.

5. Let $m \coloneqq \lfloor n \cdot 2^{-K} \rfloor = o(\log n)$. For $3 \le i \le m$ find an ℓ_i^*-shortcut f_i in G_5, where ℓ_i^* is an integer such that $\left|\ell_i^* - i \cdot 2^K\right| \le n^{0.9}$, and such that the (ℓ_i^*+1)-path accompanying f_i contains $V\left(C^* \cup \bigcup_{i=0}^K C_i\right)$. We now have that $H_4 \cup \{f_i\}$ contains all cycle lengths in $[\ell_i^* + 2 - L, \ell_i^* + 2]$. Since $\ell_i^* \ge \ell_{i+1}^* - L$ for all i, and $\ell^* + L \ge \ell_3^* + 2 - L,\ \ell_m^* \ge n - L$, we get that $H_4 \cup \{f_3, ..., f_m\}$ contains all cycle lengths in $[(d+b+1) \cdot t + 1, n]$.

   Finally, add the remaining at most $(d+b+1) \cdot t = o(\log n)$ cycle lengths by finding in G_5 an edge g_ℓ that constitutes an $(\ell-2)$-shortcut with respect to C_H, for every $\ell \in [3, (d+b+1) \cdot t]$.

   This step adds at most $m + (d+b+1) \cdot t$ edges to the constructed subgraph $H_5 \coloneqq H_4 \cup \{f_3, ..., f_m\} \cup \{g_3, ..., g_{(d+b+1)t}\}$.

Observe that the resulting subgraph H_5 is a union of the Hamilton cycle C_H and an additional set of edges

$$\{e_0, ..., e_K, e^*, e_{\text{short}}\} \cup \{e_{i,j} \mid 0 \le i \le t-1, 0 \le j \le b-1\} \cup \{f_3, ..., f_m\} \cup \{g_3, ..., g_{(d+b+1)t}\}.$$

Therefore H_5 contains at most

$$n + K + b \cdot t + m + (d+b+1) \cdot t = n + (1 + o(1)) \cdot \log_2 n$$

edges. In Section 4 we prove that the construction of H_5 we described is possible with high probability in $G(n, p^*)$. In Section 5 we prove that H_5, if it exists as a subgraph of G, is indeed pancyclic.

4 Finding the subgraph in $G(n, p)$

We follow the steps described in Section 3, and show that, in each step, the desired substructure of the respective random graph G_i, $i = 1, 2, 3, 4, 5$, exists with high probability.

We will denote the subgraph output by the i'th step of the construction (if successful) by H_i.

Step 1

Recall the notation

$$K_0 := \lfloor \log_2 \left(\frac{\log n}{6 \log\log\log n} \right) \rfloor, \ \ell_i := 2^i + 1$$

and

$$b := \lceil \log\log n \rceil, \ t := \lceil \log_b \log n \rceil.$$

By Theorem 2.1, with high probability, $G_1 \sim G(n, p_1)$ (where $p_1 = \frac{2\log\log n}{n}$) contains a sequence of cycles $C_0, C_1, ..., C_{K_0}, C_{\text{short}}$ of respective lengths $\ell_0 + 1, ..., \ell_{K_0} + 1, b \cdot t + 1$.

The following lemma implies that these cycles are also typically vertex disjoint.

Lemma 4.1. *With high probability, no two cycles of length at most $\ell_{K_0} + 1$ in G_1 intersect each other.*

Proof. Using the union bound we can show that, with high probability, G_1 does not contain a subgraph with at most $2\ell_{K_0} + 1 \leq \frac{\log n}{2\log\log\log n}$ vertices and more edges than vertices, which implies the lemma. Indeed, the probability that such a subgraph exists is at most

$$\sum_{k=4}^{2\ell_{K_0}+1} \binom{n}{k} \cdot \binom{\binom{k}{2}}{k+1} \cdot {p_1}^{k+1} \leq \sum_{k=4}^{2\ell_{K_0}+1} (e^2 n p_1)^k \cdot k \cdot p_1 \leq \log^2 n \cdot (2e^2 \log\log n)^{\frac{\log n}{2\log\log\log n}} \cdot p_1 = o(1).$$

□

Step 2

Recall that $G_2 \sim G(n, p_2)$, where $p_2 = \frac{50 \log n}{n \cdot \log\log n}$. For each $0 \leq i \leq K_0$ let $\{s_i, t_i\} := e_i \in E(C_i)$ be an arbitrary edge of C_i.

Recall that $\beta := \frac{2(\log\log n)^2}{\log n}$ and $d = \lfloor \log_{(5\beta)^{-1}}(n/200) \rfloor = (1 + o(1)) \cdot \frac{\log n}{\log\log n}$.

Lemma 4.2. *With high probability G_2 contains paths $Q_0, ..., Q_{K_0}$ such that*

1. *Q_i is a path between t_i and s_{i+1} for $0 \leq i \leq K_0 - 1$, and Q_{K_0} is between t_{K_0} and s_0;*
2. *$Q_0, ..., Q_{K_0}$ all have length $d + 2$;*
3. *$Q_0, ..., Q_{K_0}$ are vertex disjoint, and are internally vertex disjoint from $V(H_1)$.*

Recall that $K = \lfloor \log_2 \left(\frac{n}{\sqrt{\log n}} \right) \rfloor$ and $L = 2^{K+1} - 1$. Before getting to the proof of Lemma 4.2, we show the following claim.

Claim 4.3. *With high probability, for every vertex subset $U \subseteq V(G)$ with $|U| \geq n - 2L$, there is a vertex subset $U^* \subseteq U$, with $|U^*| \geq (1 - \beta) \cdot n$, such that the induced subgraph $G_2[U^*]$ is a $(\beta n, 1/3\beta)$-expander.*

Proof. First, observe that, with high probability, for every $U, W \subseteq V(G)$ disjoint subsets with $|U| = |W| = \beta n$ there is an edge in G_2 between U and W. Indeed, the probability that there are such subsets with no edge between them is at most

$$\binom{n}{\beta n}^2 \cdot (1-p_2)^{\beta^2 n^2} \leq \left(\frac{e^2}{\beta^2} \cdot \exp(-\beta n p_2)\right)^{\beta n} \leq \left(\frac{\log^2 n}{\log\log n} \cdot \exp\left(-100 \log\log n\right)\right)^{\omega(1)} = o(1).$$

Now, assume that G_2 has the aforementioned property. We reiterate an argument from [5] and show that, in this case, for every such U there is $U^* \subseteq U$ with the desired properties.

For a given U, construct U^* as follows. Set $U_0 = U$. For $i \geq 0$, if $|U_i| \geq (1-\beta)n$ and there is $W_i \subseteq U_i$ with $|W_i| \leq \beta n$ and $|N_{G_2[U_i]}(W_i)| \leq \frac{1}{3\beta}|W_i|$, set $U_{i+1} = U_i \setminus W_i$. Otherwise, terminate the process with $U^* = U_i$.

Clearly, either the resulting $G_2[U^*]$ is a $(\beta n, 1/3\beta)$-expander, or $|U^*| < (1-\beta) \cdot n$. In fact, in the latter case, it must be that $(1-2\beta) \cdot n \leq |U^*| < (1-\beta) \cdot n$, since at most βn vertices are removed in every step of the process. Suppose that this is the case, and denote $W := U \setminus U^*$. Then $|N_{G_2}(W)| \leq \frac{1}{3\beta} \cdot |W| + |V(G) \setminus U| \leq \frac{2}{3}n$. We therefore have that W and $V(G) \setminus N_{G_2}(W)$ are subsets of size at least βn with no edges between them, a contradiction to our assumption. □

With Claim 4.3 at hand we are now able to prove Lemma 4.2 by appealing to Theorem 2.2.

Proof of Lemma 4.2. We expose edges of G_2 between $\{s_i, t_i\}_{i=0}^{K_0}$ and the rest of the vertices gradually, at each step making sure we only sample edges that have not been observed previously, so that their appearance in G_2 is independent in previous steps of the proof.

Assume that G_2 has the property in the assertion of Claim 4.3, and suppose that $Q_0, ..., Q_{i-1}$ have already been constructed. We attempt to construct Q_i.

Let $U := V(G) \setminus \left(V(H_1) \cup \bigcup_{j=0}^{i-1} V(Q_j)\right)$, so that

$$|U| \geq n - 2b \cdot t - 2^{K_0+1} - K_0 \cdot (d+2) \geq n - 2L,$$

and let $U^* \subseteq U$ be a subset of size at least $(1-\beta) \cdot n$ such that $G_2[U^*]$ is a $(\beta n, 1/3\beta)$-expander. Observe that $G_2[U^*]$ satisfies the conditions of Theorem 2.2 for $\Delta = \frac{1}{4\beta}, N = \frac{1}{50}n, k = \frac{1}{2}\beta n$.

Observe that, at this point, the edges of G_2 between $\{s_{i+1}, t_i\}$ (s_0 in the case $i = K_0$) and U^* have not been sampled yet. The probability that s_{i+1} does not have a neighbour in U^* is at most $(1-p_2)^{(1-\beta)n} = o({K_0}^{-1})$. Assume that there is such a neighbour, say u. By Theorem 2.2, $G_2[U^*]$ contains a complete $\frac{1}{5\beta}$-ary tree of depth d rooted in u. This tree has at least $\frac{\beta}{40} \cdot n$ leaves. The probability that none of these leaves is a neighbour of t_i in G_2 is at most

$$(1-p_2)^{\beta n/40} \leq \exp\left(-\frac{1}{40} \cdot \frac{50 \log n}{n \cdot \log\log n} \cdot \frac{2(\log\log n)^2}{\log n} \cdot n\right) = o({K_0}^{-1}).$$

Now, if indeed t_i has a neighbour among the tree's leaves, say w, the path from s_{i+1} to u, down the the tree to w, and from w to t_i is a path of length $d+2$ that intersects $V(C_{\text{short}}) \cup \bigcup_{j=0}^{K_0} V(C_j) \cup \bigcup_{j=0}^{i-1} V(Q_j)$ only in $\{s_{i+1}, t_i\}$.

Finally, for every i we showed that the probability that such a path Q_i does not exist is at most $o({K_0}^{-1})$, and therefore, by the union bound, a sequence $Q_0, ..., Q_{K_0}$ as required exists with high probability. □

Now $\left(\bigcup_{i=0}^{K_0} Q_i\right) \cup \{e_0, ..., e_{K_0}\}$ is a cycle, denote it by C^*. We have

$$\ell^* := |C^*| = (K_0 + 1) \cdot (d+3) = (1 + o(1)) \cdot \log_2 n.$$

Step 3

Recall that $G_3 \sim G(n, p_3)$, with $p_3 = \frac{50 \log n}{n \cdot \log\log n}$. Let $e_{K_0+1}, ..., e_K, e^*$ be distinct edges of $C^* \setminus \{e_0, ..., e_{K_0}\}$, such that e^* is vertex disjoint from $e_{K_0+1}, ..., e_K$, and denote $e_i = \{s_i, t_i\}$ such that s_i is the predecessor of t_i on C^* for all i, according to an arbitrary orientation of C^*.

As a preparation for a proof that the construction in Step 3 is possible with high probability, observe that G_3 and G_2 are drawn from the same distribution, and therefore Claim 4.3 also holds for G_3. That is, we have that, with high probability, for every $U \subseteq V(G)$ with $|U| \geq n - 2L$, there is $U^* \subseteq U$, with $|U^*| \geq (1-\beta) \cdot n$, such that $G_3[U^*]$ is a $(\beta n, 1/3\beta)$-expander. In the proofs of the following two lemmas, we assume that indeed G_3 has this property.

Lemma 4.4. *With high probability G_3 contains paths $Q_{K_0+1}, ..., Q_K$ such that*

1. *Q_i is a path between s_i and t_i for $K_0 + 1 \leq i \leq K$;*
2. *Q_i has length $\ell_i + 1$ for $K_0 + 1 \leq i \leq K$;*
3. *$Q_{K_0+1}, ..., Q_K$ are internally vertex disjoint from each other and from $V(H_2)$.*

Proof. Suppose that $Q_{K_0+1}, ..., Q_{i-1}$ were found, and attempt to construct Q_i.

Here, as in Lemma 4.2, we will appeal to Theorem 2.2.

Let $U \coloneqq V(G) \setminus \left(V(H_2) \cup \bigcup_{j=0}^{i-1} V(Q_j)\right)$ and observe that

$$|U| \geq n - 2^{K+1} - K_0 \cdot (d+2) \geq n - 2L.$$

Let $U^* \subseteq U$ be a subset with at least $(1-\beta) \cdot n$ vertices such that $G_3[U^*]$ is a $(\beta n, 1/3\beta)$-expander.

As in the proof of Lemma 4.2, $G_3[U^*]$ satisfies the conditions of Theorem 2.2 for the same parameters $\Delta = \frac{1}{4\beta}, N = \frac{1}{50}n, k = \frac{1}{2}\beta n$. Recall that $d = \lfloor \log_{(5\beta)^{-1}}(n/200) \rfloor$, and let T be the tree consisting of two complete $\frac{1}{5\beta}$-ary trees of depth d, whose roots are connected by a path of length $\ell_i - 2d - 1$ (which is positive for $i > K_0$). By Theorem 2.2, U^* contains a copy of T (rooted at an arbitrary vertex).

Let $L_{s_i}, L_{t_i} \subseteq U^*$ be the sets of leaves of the embedding of T that correspond to the first and the second subtrees of T that are connected by a path. By the definition of T we have that $|L_{s_i}| = |L_{t_i}| \geq \frac{\beta}{40}n$. Observe that s_i and t_i each belong to at most one other path among $Q_{K_0+1}, ..., Q_{i-1}$. For $v \in \{s_i, t_i\}$ do the following. If $v \notin V(Q_j)$ for $K_0 < j \leq i-1$, then choose an arbitrary subset of L_v of size $\frac{1}{2}|L_v|$ and connect v to one of the vertices in the subset by an edge from $E(G_3)$, if there is a neighbour of v in the subset. If $v \in V(Q_j)$ for some $K_0 < j \leq i-1$, connect v to a vertex of L_v by a previously unexposed edge from $E(G_3)$, if such an edge exists. In both cases, at least $\frac{1}{2}|L_v| \geq \frac{\beta}{80}n$ edges are considered. Therefore, the probability that there is no edge between v and (the subset of) L_v is at most

$$(1-p_3)^{\beta n/80} \leq \exp\left(-\frac{1}{80} \cdot \frac{50 \log n}{n \cdot \log\log n} \cdot \frac{2(\log\log n)^2}{\log n} \cdot n\right) = \exp\left(-\frac{5}{4}\log\log n\right) = o(K^{-1}).$$

In the case that an edge is found, denote it by e_v.

Now, e_{s_i}, e_{t_i} along with the path of length $\ell_i - 1$ in T between the two leaves connected to s_i and t_i constitute a path between s_i and t_i of length $\ell_i + 1$, which is internally contained in U, and therefore, by the definition of U, is internally vertex disjoint from $V(H_2), Q_{K_0}, ..., Q_{i-1}$. Call this path Q_i.

The probability that there is $K_0 + 1 \leq i \leq K$ for which we did not manage to find a path Q_i in this way is at most the probability that G_3 does not have the property in the assertion of Claim 4.3, or s_i or t_i did not have a leaf neighbour in the embedding of T for some i, both of which are of order $o(1)$. $\square$

For $K_0 + 1 \leq i \leq K$, denote by C_i the cycle $Q_i \cup \{e_i\}$.

Let $v_1, ..., v_{bt+1}$ be the vertices of C_{short} according to their order on the cycle, let $\sigma : \{0, 1, ..., t-1\} \times \{0, 1, ..., b-1\} \to [tb]$ be a bijection and denote $e_{i,j} = \{v_{\sigma(i,j)}, v_{\sigma(i,j)+1}\}$ and $e_{\text{short}} = \{v_1, v_{bt+1}\}$.

Lemma 4.5. *With high probability G_3 contains paths $\{P_{i,j} \mid 0 \leq i \leq t-1, 0 \leq j \leq b-1\}$ such that*

1. *$P_{i,j}$ is a path between $v_{\sigma(i,j)}$ and $v_{\sigma(i,j)+1}$, for all i and j;*
2. *$P_{i,j}$ has length $d + 2 + j \cdot b^i$, for all i and j;*
3. *$\{P_{i,j} \mid 0 \leq i \leq t-1, 0 \leq j \leq b-1\}$ are internally vertex disjoint from each other and from $V(H_2) \cup \bigcup_{i=K_0+1}^{K} V(C_i)$.*

Proof. The proof follows similar steps to the proofs of Lemma 4.2 and Lemma 4.4 by appealing to Theorem 2.2. Assume that $P_{\sigma^{-1}(1)}, ..., P_{\sigma^{-1}(k-1)}$ have already been found, and let $(i,j) = \sigma^{-1}(k)$. Let $U^* \subseteq U := V(G) \setminus \left(V(H_2) \cup \bigcup_{r=K_0+1}^{K} V(C_r) \cup \bigcup_{r=1}^{k-1} V(P_{\sigma^{-1}(r)})\right)$ be such that $|U^*| \geq (1-\beta) \cdot n$ and $G_3[U^*]$ is a $(\beta n, 1/3\beta)$-expander.

Let s_{k+1} be a neighbour of v_{k+1} from among the first $\frac{n}{\log \log n}$ vertices of U^*. The edges between v_{k+1} and U^* in G_3 have not been sampled yet, and the probability that no such neighbour exists is at most $(1-p_3)^{\frac{n}{\log \log n}} = o(1/tb)$.

As in the previous proofs, by Claim 4.3 and Theorem 2.2 we have that $G_3[U^*]$ contains a tree which consists of a complete $\frac{1}{5\beta}$-ary tree of depth d with a path of length $j \cdot b^i$ (this can possibly be 0, in which case the path is just a vertex) attached to its root, and such that the other end of the path is s_{k+1}. Let L_k be the set of leaves in the tree, so that $|L_k| \geq \frac{\beta}{40} n$. At most $\frac{n}{\log \log n}$ of the leaves were considered as neighbours of v_k in previous steps, and therefore the probability that v_k does not have a neighbour t_k from among the remaining leaves is at most $(1-p_3)^{\frac{\beta}{50} n} = o(1/tb)$. Now the path from v_{k+1}, through s_{k+1}, along the (jb^i)-path, down the tree to t_k and then to v_k, satisfies all the requirements to be $P_{i,j}$.

The probability that for some k one of the vertices s_{k+1}, t_k was not found is of order $o(1)$, and therefore, with high probability, this construction ends successfully. □

We remain with finding a path between e_{short} and e^*, which is done in the following lemma.

Lemma 4.6. *With high probability G_3 contains a path P^* of length $d+2$ between a vertex of e_{short} and a vertex of e^*, which is internally disjoint from $V(H_2) \cup \bigcup_{i=K_0+1}^{K} V(C_i) \cup \bigcup_{i,j} V(P_{i,j})$.*

Proof. As in previous constructions in this step, let $U = V(G) \setminus \left(V(H_2) \cup \bigcup_{i=K_0+1}^{K} V(C_i) \cup \bigcup_{i,j} V(P_{i,j})\right)$ and let U^* be a large subset spanning an expander. At least $\frac{1}{2} n$ vertices of U^* have not yet been considered as neighbours of v_{bt+1}, and with high probability at least one of them is, denote it by s. By Claim 4.3 and Theorem 2.2, $G_3[U^*]$ contains a complete $\frac{1}{5\beta}$-ary tree of depth d rooted in s. As none of the edges in G_3 of the vertices of e^* have been sampled yet, with high probability there is an edge between one of them and the tree's at least $\frac{\beta}{40} n$ leaves, which together with the path from the leaf to s and with $\{s, v_{bt+1}\}$ forms a path P^* satisfying the conditions. □

Step 4

Denote $X = V(G) \setminus V(H_3)$, and let $s_H \in e^*, t_H \in e_{\text{short}}$ be the vertices of e^*, e_{short} not already connected by P^*.

Claim 4.7. *With high probability there is a path P in G_4 between s_H and t_H, whose vertex set is $V(P) = X \cup \{s_H, t_H\}$.*

Proof. Consider the induced subgraph $G_4[X] \sim G(|X|, p_4)$. We have

$$\begin{aligned} |X| \cdot p_4 &\geq (n - 2L) \cdot \left(\frac{\log n + 10\sqrt{\log n}}{n} \right) \\ &\geq \log n \cdot \left(1 - \frac{4}{\sqrt{\log n}} \right) \cdot \left(1 + \frac{10}{\sqrt{\log n}} \right) \\ &= \log n + \omega(\log \log n) \\ &= \log |X| + \omega(\log \log |X|) \end{aligned}$$

Therefore, by Lemma 2.1, there is a set $S \subseteq X$ with $|S| = \frac{1}{4}|X| \geq \frac{1}{5}n$, and for every $s \in S$ there is a subset $T_s \subseteq X$ with $|T_s| \geq \frac{1}{4}|X| \geq \frac{1}{5}n$, such that there is a Hamilton path in $G_4[X]$ between s and t for every $t \in T_s$.

The set $E_{G_4}(s_H, X)$ has not yet been sampled. The probability that s_H has no neighbour in S is at most $(1 - p_4)^{n/5} = o(1)$. Assume that there is one, and denote it by s. Similarly, the probability that t_H has no neighbour in T_s is at most $(1 - p_4)^{n/5} = o(1)$, denote such a neighbour by t. Now the Hamilton path in $G_4[X]$ between s and t, along with the edges $\{s_H, s\}, \{t_H, t\}$, constitute a path P with $V(P) = X \cup \{s_H, t_H\}$, as desired. □

Denote the obtained Hamilton cycle $H_3 \cup P \setminus (\{e_0, ..., e_K, e^*, e_{\text{short}}\} \cup \{e_{i,j}\}_{i,j})$ by C_H.

Step 5

Let $m := \lfloor n \cdot 2^{-K} \rfloor$.

Lemma 4.8. *With high probability G_5 contains edges $f_3, ..., f_m$, such that the following hold for every i.*

1. *There is $\ell_i^* \in \left[i \cdot 2^K - n^{0.9}, i \cdot 2^K + n^{0.9}\right]$ such that f_i is an ℓ_i*-shortcut with respect to C_H;*
2. *The $(\ell_i^* + 1)$-path on C_H that connects the vertices of f_i contains $V\left(C^* \cup \bigcup_{i=0}^{K} C_i\right)$.*

Proof. For every $3 \leq i \leq m$ there is a set of at least $\frac{1}{5} \cdot n^{1.8}$ potential edges that satisfy the conditions. The probability that there is $3 \leq i \leq m$ for which none of these edges appears in G_5 is at most

$$m \cdot (1 - p_5)^{n^{1.8}/5} = o(1).$$

□

Lemma 4.9. *With high probability G_5 contains an $(\ell - 2)$-shortcut with respect to C_H for every $\ell \in [3, (d + b) \cdot t]$.*

Proof. For a given $\ell \in [3, (d + b) \cdot t]$, the probability that such an $(\ell - 2)$-shortcut does not exist is $(1 - p_5)^n = o(1/\log n)$, and by the union bound we obtain the lemma, as $(d + b)t = o(\log n)$. □

5 Proof of Theorem 1

The following lemma, referring to the subgraph $H_5 \subseteq G$ described in Section 3 and whose construction is shown to be possible with high probability in Section 4, completes the proof of Theorem 1.

Lemma 5.1. *The subgraph H_5 is pancyclic.*

Proof. Let $\ell \in [3, n]$. We show that H_5 contains a cycle of length ℓ. We divide the proof into cases based on a subinterval of $[3, n]$ that ℓ resides in. The subintervals are covering $[3, n]$ but not necessarily disjoint, so ℓ may be covered by more than one subinterval.

- If $\ell \in [3, (d+b+1) \cdot t]$, then g_ℓ is an $(\ell-2)$-shortcut with respect to C_H, so that g_ℓ and its accompanying $(\ell-1)$-path form an ℓ-cycle.

- If $\ell \in [(d+b+1) \cdot t + 1, (d+b+1) \cdot t + b^t]$, let $k = \ell - (d+b+1) \cdot t - 1$, so $0 \leq k \leq b^t - 1$ can be encoded in base b using t digits. Let $(k_{t-1}, k_{t-2}, ..., k_1, k_0)$ be its encoding, that is, $0 \leq k_i \leq b-1$ for all i, and $k = \sum_{i=0}^{t-1} k_i b^i$. Then
$$\{e_{\text{short}}\} \cup \bigcup_{j=k_i} P_{i,j} \cup \bigcup_{j \neq k_i} \{e_{i,j}\}$$
is a cycle of length ℓ in H_5. Indeed, it is a cycle, since it is the result of replacing a subset of the edges of C_{short} with internally disjoint paths, and its length is
$$1 + (b-1) \cdot t + \sum_{i=0}^{t-1} e(P_{i,k_1}) = 1 + (b-1) \cdot t + \sum_{i=0}^{t-1} \left(d + 2 + k_i b^i\right) = 1 + (d+b+1) \cdot t + k = \ell.$$

- If $\ell \in [\ell^*, \ell^* + L]$, let $k = \ell - \ell^*$, so $0 \leq k \leq 2^{K+1} - 1$ can be encoded by $K+1$ binary digits, say $k = \sum_{i=0}^{K} k_i 2^i$, where $k_i \in \{0, 1\}$. Then
$$\left(C^* \cup \bigcup_{i:k_i=1} C_i\right) \setminus \{e_i \mid k_i = 1\}$$
is a cycle in H_5 with length ℓ. It is a cycle because it is the result of replacing edges of C^* with internally disjoint paths. The length is indeed
$$\ell^* + \sum_{i=0}^{K} k_i \cdot (e(C_i) - 1) = \ell^* + \sum_{i=0}^{K} k_i 2^i = \ell^* + k = \ell.$$

- If $\ell \in [\ell_i^* + 2 - L, \ell_i^* + 2]$, where $3 \leq i \leq m$, then, similarly to the previous case, let $\ell_i^* + 2 - \ell = k = \sum_{i=0}^{K} k_i 2^i$, where $k_i \in \{0, 1\}$. Denote by C_i^* the cycle of length $\ell_i^* + 2$ comprised of f_i and its accompanying $(\ell_i^* + 1)$-path. Then
$$\left(C_i^* \setminus \bigcup_{i:k_i=1} C_i\right) \cup \{e_i \mid k_i = 1\}$$
is a cycle of length $\ell^* + 2 - k = \ell$.

- If $\ell \in [n-L, n]$ then for $n - \ell = k = \sum_{i=0}^{K} k_i 2^i,\ k_i \in \{0,1\}$, we get a cycle

$$\left(C_H \setminus \bigcup_{i:k_i=1} C_i\right) \cup \{e_i \mid k_i = 1\}$$

of length $n - k = \ell$.

Observe that

$$\begin{aligned}(d+b+1)\cdot t + b^t \geq b^{\log_b \log n} = \log n \geq \ell^* \ ;\\ \ell^* + L \geq L \geq 2^K + n^{0.9} + 3 \geq \ell_3^* + 2 - L \ ;\\ \ell_i^* + 2 \geq i \cdot 2^K - n^{0.9} + 2 \geq (i-1)\cdot 2^K + n^{0.9} + 3 \geq \ell_{i+1}^* + 2 - L \ ;\\ \ell_m^* + 2 \geq m \cdot 2^K - n^{0.9} \geq (n \cdot 2^{-K} - 1)\cdot 2^K - n^{0.9} \geq n - L \ ;\end{aligned}$$

and therefore the subintervals indeed cover $[3, n]$, and so H_5 is pancyclic. □

References

[1] J. Balogh, B. Csaba, M. Pei, and W. Samotij, *Large bounded degree trees in expanding graphs*, Electronic Journal of Combinatorics 17 (2010), Research Paper 6.

[2] J.A. Bondy, *Pancyclic graphs I*, Journal of Combinatorial Theory, Series B, 11 (1971), 80–84.

[3] C. Cooper, *1-Pancyclic Hamilton Cycles in Random Graphs*, Random Structures & Algorithms 3 (1992), 277–287.

[4] C. Cooper and A. Frieze, *Pancyclic random graphs*, In Proc. 3rd Annual Conference on Random Graphs, Poznań (1987), 29–39.

[5] L. Friedman and M. Krivelevich, *Cycle lengths in expanding graphs*, Combinatorica, 41 (2021), 53–74.

[6] J. C. George, A. Khodkar and W. D. Wallis, **Pancyclic and Bipancyclic Graphs**, Springer, 2016.

[7] P. E. Haxell, *Tree embeddings*, Journal of Graph Theory, 36 (2001), 121–130.

[8] M. Krivelevich, *Long paths and Hamiltonicity in random graphs.* In: Random Graphs, Geometry and Asymptotic Structure, N. Fountoulakis and D. Hefetz, Eds., London Math. Soc. Student Texts Vol. 84, 2016, Cambridge University Press, pp. 4–27.

[9] T. Łuczak, *Cycles in random graphs*, Discrete Mathematics, 98 (1991), 231–236.

[10] Y. Shi, *The number of cycles in a Hamilton graph*, Discrete Mathematics, 133 (1994), 249–257.