

Pancyclicity of Hamiltonian graphs

Nemanja Draganić*

David Munhá Correia*

Benny Sudakov*

Abstract

An n -vertex graph is *Hamiltonian* if it contains a cycle that covers all of its vertices, and it is *pancyclic* if it contains cycles of all lengths from 3 up to n . In 1972, Erdős conjectured that every Hamiltonian graph with independence number at most k and at least $n = \Omega(k^2)$ vertices is pancyclic. In this paper we prove this old conjecture in a strong form by showing that if such a graph has $n = (2 + o(1))k^2$ vertices, it is already pancyclic, and this bound is asymptotically best possible.

1 Introduction

Hamiltonicity is one of the most central notions in graph theory, and it has been extensively studied by numerous researchers. The problem of deciding Hamiltonicity of a graph is NP-complete and therefore, a central theme in Combinatorics is to derive sufficient conditions for this property. The most classical one is Dirac's theorem [12] which dates back to 1952 and states that every n -vertex graph with minimum degree at least $n/2$ contains a Hamilton cycle. Since then, many other interesting results about various aspects of Hamiltonicity have been obtained, see e.g. [1, 8, 9, 10, 17, 23, 24, 25, 30], and the surveys [19, 26].

A related notion to Hamiltonicity is that of pancyclicity. An n -vertex graph is said to be *pancyclic* if it contains all cycles of length from 3 up to n . Trivially, pancyclicity is a stronger property than Hamiltonicity, and one might ask how much stronger it really is. In 1973, Bondy [3] stated his celebrated meta-conjecture, indicating that the first property should be only slightly stronger than the latter. Indeed, he proposed that any non-trivial condition which implies that a graph is Hamiltonian should also imply that it is pancyclic (up to a certain collection of simple exceptional graphs). As an example, Bondy [4] himself first showed that every n -vertex graph with minimum degree at least $n/2$ is either pancyclic or isomorphic to the complete bipartite graph $K_{n/2, n/2}$, thus strengthening Dirac's theorem. His meta-conjecture sparked a lot of research which in turn has led to various appealing results and methods. For example, Bauer and Schmeichel [2], relying on previous results of Schmeichel and Hakimi [31], showed that the sufficient conditions for Hamiltonicity given by Bondy [5], Chvátal [7] and Fan [16] all imply pancyclicity, up to a certain small family of exceptional graphs. Furthermore, much like with Dirac's theorem, the classical result of Chvátal and Erdős [7] that a graph with connectivity number $\kappa(G)$ at least as large as its independence number $\alpha(G)$ is Hamiltonian, has also been addressed. Namely, in 1990, Jackson and Ordaz [20], conjectured that if $\kappa(G) > \alpha(G)$, then G must be pancyclic. An approximate form of this conjecture was proven by Keevash and Sudakov [22], who showed that $\kappa(G) \geq 600\alpha(G)$ is sufficient. The authors [13] recently showed an asymptotically optimal result, proving that $\kappa(G) \geq (1 + o(1))\alpha(G)$ is enough for pancyclicity. Finally, building on our work, Letzter [28] proved the Jackson-Ordaz conjecture.

Bondy's meta-conjecture is about conditions for Hamiltonicity which imply pancyclicity. A natural and closely related question was first studied by Erdős in the 1970s. Let G be a Hamiltonian graph; under which assumptions can we guarantee that G is also pancyclic or more generally, that it has many cycle lengths? One example of such a problem was suggested by Jacobson and Lehel at the 1999 conference

*Department of Mathematics, ETH, Zürich, Switzerland. Research supported in part by SNSF grant 200021_196965.
Emails: {nemanja.draganic,david.munhacanascorreia, benjamin.sudakov}@math.ethz.ch.

“Paul Erdős and His Mathematics”. They asked for the minimal number of cycle lengths in a k -regular n -vertex Hamiltonian graph. They conjectured (see Verstraëte [32] for a stronger conjecture) that already when $k \geq 3$, there are $\Omega(n)$ many lengths. Improving on the previously best known lower bound of $\Omega(\sqrt{n})$ by Milans et al. [29], recently Bucić, Gishboliner and Sudakov [6] showed that any Hamiltonian graph with minimum degree at least 3 has $n^{1-o(1)}$ different cycle lengths.

As we already mentioned above, the earliest question of this flavor was studied by Erdős about 50 years ago. In 1972 he posed the following problem in [14]. Given an n -vertex Hamiltonian graph with independence number $\alpha(G) \leq k$, how large does n have to be in terms of k in order to guarantee that G is pancyclic? Erdős [14] proved that it is enough to have $n = \Omega(k^4)$ and conjectured that already $n = \Omega(k^2)$ should be enough. A simple construction shows that this is best possible. Let C_1, \dots, C_k be disjoint cliques of size $2k - 2$, and let each C_i have two distinguished vertices a_i and b_i . Let G be the graph obtained by connecting a_i and b_{i+1} by an edge for each i (taking addition modulo k). Notice that this graph has $n = 2k^2 - 2k$ vertices, is Hamiltonian and its independence number is k . On the other hand, it is easy to check that it does not contain a cycle of length $2k - 1$, and thus it is not pancyclic. Indeed, observe that every cycle must be either a subgraph of one of the cliques C_i , or contain all the vertices a_i, b_i for each i . The first type of cycles all have length at most $2k - 2$ and the latter have length at least $2k$.

In the last 50 years, there have been several improvements upon Erdős’s initial result. Firstly, Keevash and Sudakov [22] proved that $n = \Omega(k^3)$ vertices are enough to guarantee pancyclicity. Then, Lee and Sudakov [27] improved this to $n = \Omega(k^{7/3})$, and more recently Dankovics [11] showed that $n = \Omega(k^{11/5})$ vertices suffice. In this paper we completely resolve the conjecture of Erdős in the following strong form.

Theorem 1.1. *Every Hamiltonian graph G with $\alpha(G) \leq k$ and at least $2k^2 + o(k^2)$ vertices is pancyclic.*

As shown by the previous construction, our result is tight up to the $o(k^2)$ error term. The rest of this paper is organized as follows. In Section 2, we state some well-known tools, and introduce some useful definitions. There we also prove the two key propositions that are used in the proof of Theorem 1.1, which is given in Section 3. Finally, in Section 4 we make some concluding remarks and mention some open questions; we also show how to get a short proof of the conjecture of Erdős, that is, a proof of Theorem 1.1 with a sufficiently large constant C instead of the precise factor of 2.

2 Preliminaries

2.1 Notation and definitions

We mostly use standard graph theoretic notation. Let G be a finite graph. Denote by $V(G)$ its vertex set, and let $S_1, S_2 \subseteq V(G)$. We denote by $G[S_1]$ the subgraph of G induced by S_1 , and by $E[S_1, S_2]$ the set of edges with one endpoint in S_1 and the other in S_2 . Let H be a subgraph of G . We denote by $G[H]$ the graph $G[V(H)]$. A path $P = (x_0, x_1, \dots, x_\ell)$ of length ℓ is a graph on vertex set $\{x_0, x_1, \dots, x_\ell\}$ with an edge between x_{i-1} and x_i for all $i \in [\ell]$. We say that x_0 and x_ℓ are the endpoints of P , and we call P an x_0x_ℓ -path. If the vertices of the graph G come with a given ordering, then we say that a path $P = (x_0, x_1, \dots, x_\ell)$ contained in G is increasing if $x_0 < x_1 < \dots < x_\ell$. We denote by $\alpha(G)$ the independence number of G . Given a digraph D , its independence number $\alpha(D)$ is defined as the independence number of the underlying graph.

Given sets $A_1, A_2 \subset \mathbb{N}$, we denote by $A_1 + A_2$ the set of integers c such that $c = a_1 + a_2$ for some $a_1 \in A_1$ and $a_2 \in A_2$. Throughout the paper we omit floor and ceiling signs for clarity of presentation, whenever it does not impact the argument.

Definition 2.1. Let a, b, p be positive real numbers. Given a graph G , and two vertices x and y , we say that the pair xy is p -dense in the interval $[a, b]$ if for every subinterval $[a', b']$ with $b' - a' \geq p$ there is an integer $\ell \in [a', b']$ and an xy -path in G of length ℓ .

2.2 Standard tools

Here we state and prove some standard facts which we use in our proof. We start with the following well-known result about directed graphs of Gallai and Milgram [18]. A *path cover* in a directed graph is a partition of its vertex set into directed paths, and its size is the number of such paths.

Lemma 2.2 ([18]). *Every directed graph D has a path cover of size at most $\alpha(D)$.*

We also use the celebrated Ramsey's theorem.

Theorem 2.3. *For every two positive integers k, t , there exists a large enough integer n , such that for any k -coloring of the edges of K_n , there is a monochromatic copy of K_t in K_n .*

The next lemma shows that we can partition a large proportion of the vertex set of a graph into sets with small radius, such that there are no edges between the parts.

Lemma 2.4. *Let G be an n -vertex graph and let $0 < \gamma < \frac{1}{2}$. Then, there exists a collection of vertices v_1, v_2, \dots, v_r and disjoint sets $U_1, U_2, \dots, U_r \subseteq V(G)$ such that the following hold.*

1. $v_j \in U_j$ for all j , and $\left| \bigcup_j U_j \right| \geq (1 - \gamma)n$.
2. Every vertex $u \in U_j$ has $\text{dist}(v_j, u) \leq \log_{1+\gamma} n$.
3. There is no edge between any two sets $U_j, U_{j'}$ with $j \neq j'$.

Proof. We find the required sets and vertices with the following algorithm. We start with an arbitrary vertex $v_1 \in V(G)$ and consider the breadth-first-search tree rooted at v_1 , that is, consider the sets $V_0^{(1)}, V_1^{(1)}, V_2^{(1)}, \dots \subseteq V(G)$ defined as $V_i^{(1)} := \{u \in G : \text{dist}(v_1, u) = i\}$. Now, define $i_1 \geq 0$ to be the minimal i such that $|V_{i+1}^{(1)}| \leq \gamma |V_0^{(1)} \cup \dots \cup V_i^{(1)}|$ and let $U_1 := V_0^{(1)} \cup \dots \cup V_{i_1}^{(1)}$. We now continue the process and do the same on the graph $G' := G \setminus (U_1 \cup V_{i_1+1}^{(1)})$. More precisely, take an arbitrary vertex $v_2 \in G'$ and consider again the breadth-first-search tree in G' rooted at v_2 . Like before, this gives an i_2 and sets $V_0^{(2)}, \dots, V_{i_2+1}^{(2)}, U_2$. We then repeat this on the graph $G'' := G' \setminus (U_2 \cup V_{i_2+1}^{(2)})$ and continue doing this until we have no vertices left. By construction, the desired properties hold. Indeed, the only vertices not contained in $\bigcup_j U_j$ are in some $V_{i_j+1}^{(j)}$, hence there are at most γn of them. For the second property, observe that each U_j is of size at least $(1 + \gamma)^{i_j}$, so $i_j \leq \log_{1+\gamma} n$. The third condition holds by construction, since we deleted all the neighbors of U_i before defining U_{i+1} . \square

Finally, we state a trivial observation used throughout the proof of Theorem 1.1. It will be used to state that appropriate combinations of internally vertex-disjoint paths of different lengths result in the construction of cycles of many different lengths.

Observation 2.5. *Let G be a graph whose vertex set contains t disjoint sets S_1, \dots, S_t and another set of t vertices v_1, \dots, v_t outside of $\bigcup_{i=1}^t S_i$. For each $i \in [t]$, let $A_i \subset \mathbb{N}$ and suppose that for every i the induced subgraph $G[v_i \cup S_i \cup v_{i+1}]$ is such that it contains a $v_i v_{i+1}$ -path of length ℓ for each $\ell \in A_i$ (with $v_{t+1} = v_1$). Then for every $\ell \in A_1 + \dots + A_t$, the graph G contains a cycle of length ℓ .*

2.3 Finding consecutive path lengths

Next we show that in a graph with small independence number, we can find two vertices between which there exist paths of almost all ‘possible’ lengths. We believe that this result is of independent interest and pose a problem related to it after its proof.

Proposition 2.6. *Let G be an n -vertex graph with $\alpha(G) = k$ and let $0 < \gamma < 1/2$. Then there exist two vertices $u, v \in V(G)$ such that for every integer $\ell \in [\log_{1+\gamma} n, (1 - \gamma)\frac{n}{k}]$ there is a uv -path of length ℓ .*

Proof. First, we apply Lemma 2.4 to G to get vertices v_i and sets U_i for all $i \leq r$. Fix the graph $H := G[U_1 \cup \dots \cup U_r]$ and let us orient the edges of H in the following manner. For an edge xy in H with $x, y \in U_j$ (recall property 3 of the sets U_j) orient it as $x \rightarrow y$ if $\text{dist}(v_j, x) < \text{dist}(v_j, y)$ and as $y \rightarrow x$ if $\text{dist}(v_j, x) > \text{dist}(v_j, y)$. In the case that $\text{dist}(v_j, x) = \text{dist}(v_j, y)$, orient the edge arbitrarily. Now, since $\alpha(H) \leq \alpha(G) = k$ and $|H| \geq (1 - \gamma)n$, by Lemma 2.2 there must exist a directed path $\vec{P} = x_1 \rightarrow x_2 \rightarrow \dots \rightarrow x_m$ in H of length $m := \frac{|H|}{\alpha(H)} \geq (1 - \gamma)\frac{n}{k}$. Let j be such that $\vec{P} \subseteq U_j$, and for each $i \in [m]$ denote $d_i := \text{dist}(v_j, x_i)$ and note that $d_i \leq \log_{1+\gamma} n$.

Now we show that for all $\ell \in [d_m, m + d_1] \supseteq [\log_{1+\gamma} n, m]$ there is path in G of length ℓ between v_j and x_m . For each $i \in [m]$ look at the $v_j x_m$ -path P_i obtained by concatenating the shortest $v_j x_i$ -path with the path $x_{i+1} x_{i+2} \dots x_m$. Since by definition, we have that $d_i \leq d_\ell$ for $i < \ell$, these two paths are vertex disjoint and their union is indeed a path as well. Moreover, for each i , we have that $|P_i| - 1 \leq |P_{i+1}| \leq |P_i|$ since $x_i \rightarrow x_{i+1}$ is an edge and thus, $d_i \leq d_{i+1} \leq d_i + 1$. Because $|P_1| = d_1 + m$ and $|P_m| = d_m$, each path length in $[d_m, m + d_1]$ is then attained by at least one of the constructed paths. Since $d_m \leq \log_{1+\gamma} n$ and $m \geq (1 - \gamma)\frac{n}{k}$, this finishes the proof. \square

Before moving on to the next section, it is worth noting that the above proposition is asymptotically tight. Indeed, an n -vertex graph G with $\alpha(G) = k$ does not need to contain a path of length larger than $\frac{n}{k}$, as we can see from a disjoint union of cliques of size n/k . Since the proof of this proposition uses a result about directed paths, we further ask if the following directed variant of it might be true as well.

Problem 2.7. Does Proposition 2.6 generalize to directed graphs? If G is a directed graph, how large an interval $I \subseteq [0, \frac{n}{k}]$ can we guarantee for which there are vertices u, v with a directed uv -path of length ℓ for all $\ell \in I$?

2.4 Path shortening

In this section we show that if a graph has small independence number and contains a long path P , then we can find a slightly shorter path P' with the same endpoints, which satisfies certain additional properties. In a graph with independence number k , a path can clearly be shortened by considering $2k + 1$ consecutive vertices on the path, and observing that there must be an edge (not contained in the path) between those vertices. This is the statement of the next simple lemma.

Lemma 2.8. *Let G be a graph with independence number k , and P a path in G with endpoints x, y such that $|P| > 2k$. Then, there is an xy -path P' with $V(P') \subseteq V(P)$ such that $|P| - 2k \leq |P'| < |P|$.*

The following proposition is one of the central results of this paper. It shows that we can shorten a path only by a little while also preserving a pre-specified set of vertices in the newly obtained path.

Proposition 2.9. *Let G be an n -vertex graph with independence number k , and let P be a path in G with endpoints x, y . Further, let $c \in \mathbb{N}$ and $U \subseteq V(P)$, where $c \left(\frac{|P| - (4c-1)|U|}{2k} - 1 \right) > k$. Then there is an xy -path P' with the following properties.*

1. $U \subseteq V(P') \subseteq V(P)$.
2. $|P| - (4c - 3) \leq |P'| < |P|$.

Before we give a proof of Proposition 2.9, we introduce some useful concepts. The key idea to prove the proposition is to find a certain structure in our graph which can be used to shorten a path in a graph with low independence number. The properties of this structure are captured by the notion of a *special edge set*, defined below.

Definition 2.10. Given a graph H with ordered vertex set $(1, 2, \dots, n)$, we say that a sequence of vertices (v_1, \dots, v_ℓ) is *special* if the following hold (see Figure 1).

- $v_{i+1} > v_i$ for all $i \in [\ell - 1]$.
- $(v_i, v_{i+1} + 1)$ is an edge in H for all $i \in [\ell - 1]$. We call the set formed by those edges a *special edge set*.

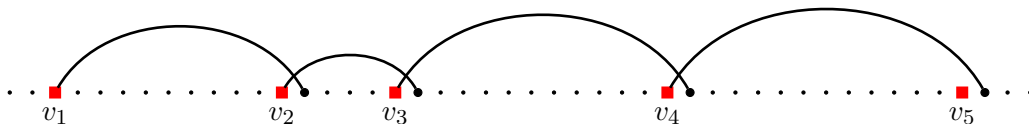


Figure 1: An illustration of a special edge set. The special vertices are colored red.

We can now find special vertex sequences in graphs in the following manner.

Lemma 2.11. *Let H be a graph on the vertex set $[n]$, let $U' \subseteq [n]$ and suppose H has independence number k . Then, there exist a special vertex sequence with at least $\frac{n - |U'|}{2k}$ vertices in $[n] \setminus U'$.*

Proof. Let H' be the graph obtained from H by removing all edges of the form $\{i, i + 1\}$. Since the removed edges form a graph with chromatic number at most 2, we must have that $\alpha(H') \leq 2\alpha(G) \leq 2k$. Let us also direct the edges of H' so that an edge ij is oriented from i to j if $i < j$. Applying Lemma 2.2 to this directed graph, we obtain a partition of $[n]$ into at most $2k$ directed paths. Denote the digraph formed by the union of those paths as F . An important property of F that we are going to use is that all outdegrees and indegrees of vertices in F are at most one.

Having obtained this path cover, we want to find another decomposition but now of the edges of F , and into a small number of special edge sets. Simply take \mathcal{M} to be a smallest collection of edge disjoint special edge sets which decompose the edges of F (this exists as one such collection is the set of all single edges in F). If $(v_1, v_2, \dots, v_\ell)$ is the special sequence corresponding to a special edge set M in \mathcal{M} , then note that v_ℓ must be a vertex of out-degree 0 in F . Indeed, if this is not the case, then there exists $u > v_\ell$ such that $v_\ell \rightarrow u$ is an edge of F . Let $M' \in \mathcal{M}$ be the special edge set which contains it and note that then v_ℓ is the first vertex of this special edge set, since otherwise it would contain the edge $\{v_{\ell-1}, v_\ell + 1\} \in M$ and contradict the edge-disjointness of the special edge sets in \mathcal{M} . But now we can 'concatenate' M and M' to form a larger special edge set $M \cup M'$, which contradicts the minimality of \mathcal{M} .

Notice that the special edge sets in \mathcal{M} have disjoint corresponding special vertex sequences. Indeed, since these special edge sets are non-empty and edge-disjoint, the only possibility for a common vertex in two different special vertex sequences would be if some vertex v is the first vertex of one sequence and the last vertex of another one. In turn, as shown in the previous paragraph, this would contradict the maximality of \mathcal{M} . Now, a consequence of having disjoint corresponding special vertex sequences in \mathcal{M}

and the previous paragraph is that each one of these has a unique vertex of out-degree 0. Moreover, let $S \subseteq V(H)$ denote the set of vertices which do not belong to any special vertex sequence of \mathcal{M} and notice that all $v \in S$ also have out-degree 0 in F . In turn, since F is a decomposition of $[n]$ with at most $2k$ paths, there are at most $2k$ vertices of outdegree zero and thus $|\mathcal{M}| + |S| \leq 2k$. Hence, there exists one special vertex sequence in $\mathcal{M} \cup S$ (allowing a single vertex to be a special vertex sequence) with at least $\frac{n-|U'|}{|\mathcal{M}|+|S|} \geq \frac{n-|U'|}{2k}$ vertices not in U' , which finishes the proof. \square

We can now prove our proposition.

Proof of Proposition 2.9. Let $(1, 2, \dots, |P|)$ be an ordering of the vertices of P , with the endpoints $x = 1$ and $y = |P|$. Let U_c be the set of vertices v in P such that there is a $u \in U$ with $|v - u| \leq 2c - 1$. By applying Lemma 2.11 to $G[V(P)]$ and the set U_c , we obtain a special vertex sequence (v_1, \dots, v_ℓ) corresponding to a special edge set M , with at least $\frac{|P|-|U_c|}{2k}$ vertices outside of U_c . First, notice that for each $v_{i+1} \notin U_c$ we may assume that $v_{i+1} - v_i \geq 2c$. Indeed, if that was not the case, then we obtain the desired path P' from P by replacing the interval $[v_i, v_{i+1} + 1]$ by the edge $(v_i, v_{i+1} + 1)$, noting that there are no vertices from U inside of the removed interval, since otherwise v_{i+1} would be in U_c , a contradiction.

Now we define, for each special vertex $v = v_i$ which is not in $U_c \cup \{v_1\}$, the c -element set $S_v := \{v - 1, v - 3, \dots, v - (2c - 1)\}$ contained in the $(2c - 1)$ -element interval $I_v := [v - (2c - 1), v - 1]$, which is disjoint to U . Since $v_i - v_{i-1} \geq 2c$, all of these sets are disjoint and further, disjoint to U . Now, the union S of those sets is of size at least $\left(\frac{|P|-|U_c|}{2k} - 1\right)c$. Therefore, since $|U_c| \leq (4c - 1)|U|$, we get that $|S| \geq k + 1$ by our assumption on c . Hence, there exists an edge e in G spanned by S , since the independence number of G is k . We may assume that this edge does not lie inside some S_v , as otherwise we can again get the desired path P' by using e instead of the interval which it bridges, avoiding at least one, but at most $2c - 1$ vertices which are not in U . Hence, $e = ab$ with $a < b$ is between two distinct sets S_{v_i} and S_{v_j} . Now we can find the required path P' as shown in Figure 2, avoiding at most $4c - 3$ vertices. More precisely, we obtain the required path P' as the union of the following paths:

- The part of the path P which connects x to a , plus the edge (a, b) .
- The increasing path P_2 obtained by the following iterative procedure. First, initialize P_2 to be the edge $(v_i, v_i + 1)$. Repeat the following. Let r be the last vertex of P_2 . If r is a special vertex $r = v_t$, then add the edge $(r, v_{t+1} + 1)$ to P_2 , and update $r = v_{t+1} + 1$. If r is not a special vertex, add the edge $(r, r + 1)$ to P_2 , and update $r = r + 1$. We stop when either $r = b$ or $r = v_j + 1$.
- The increasing path P_3 obtained by the following iterative procedure. First we initialize P_3 to be the edge $(v_i, v_{i+1} + 1)$. Repeat the following (exactly as for the previous path). Let r be the last vertex of P_3 . If r is a special vertex $r = v_t$, then add the edge $(r, v_{t+1} + 1)$ to P_3 , and update $r = v_{t+1} + 1$. If r is not a special vertex, add the edge $(r, r + 1)$ to P_3 , and update $r = r + 1$. We stop when either $r = b$ or $r = v_j + 1$.
- The part of the path P which connects $v_j + 1$ to y .

It is easy to see that this path contains only vertices of P and that it does not contain the vertex v_j . Furthermore, the only other vertices from P which the new path P' avoids are the vertices in I_{v_i} which are strictly larger than a , and the vertices in I_{v_j} which are strictly larger than b . So in total, the new path P' avoids at most $1 + (|I_{v_i}| - 1) + (|I_{v_j}| - 1) = 4c - 3$ vertices of P , which completes the proof. \square

so adding the edge xy to it creates a cycle of length $n - 1$. We remove the remaining vertex and repeat. This gives all cycle lengths in $[2k^2 + 2k, n]$.

Now we turn to the cycle lengths in $[\frac{1000}{\varepsilon^2}k, 2k^2 + 2k]$. For this we need the following lemma.

Lemma 3.3. *Let G be a Hamiltonian graph on n vertices with independence number k . Then, there is a partition of the vertices of G into a cycle C and a set S of size $|S| = \frac{\varepsilon n}{20}$, such that there is a matching $M \subseteq E[C, S]$ which covers S .*

Proof. To show this claim, we apply Proposition 2.9 iteratively $\varepsilon n/20$ times as follows. We always have $c = 1$, and in the beginning we set $U_M = S = \emptyset$, and we set M to be an empty matching and C a Hamilton cycle in G . During the procedure C and S are always disjoint and partition the vertices of G , M is a matching in $E[C, S]$ which covers S , and U_M is the set of endpoints of M in C .

In the first step we apply Proposition 2.9 with the mentioned values of $c = 1$ and $U = \emptyset$, to get a cycle C' of length $n - 1$, and a vertex v which is not on C' . Let v' be a neighbor of v in the cycle C . Now we set $C = C'$, $S = \{v\}$, $M = \{vv'\}$ and $U_M = \{v'\}$.

In the i -th step of the procedure, we let U denote the set of vertices which are either in U_M , or adjacent to a vertex in U_M on the current cycle C . We apply Proposition 2.9 with $c = 1$ and U to the graph $G[C]$, to get a cycle C' of length $|C| - 1$ and a vertex $v \in C \setminus C'$ which is not in U . Again, we denote by v' the neighbor of v in C , and we set $M := M \cup \{vv'\}$, $S := S \cup \{v\}$, $U_M = U_M \cup \{v'\}$ and $C = C'$. Notice that since we only perform $\varepsilon n/20$ steps, at each point we have that $|U_M| \leq \varepsilon n/20$, and that $|C| \geq n - \varepsilon n/20$. Together with the fact that $|U| \leq 3|U_M|$, this gives that $\frac{|C| - 3|U|}{2k} - 1 > k$, so we can always successfully apply Proposition 2.9 with $c = 1$. The resulting matching is then of size $\frac{\varepsilon n}{20}$, and evidently satisfies the given requirements. □

We are ready to show how to get the cycle lengths in $[\frac{1000}{\varepsilon^2}k, 2k^2 + 2k]$. We first apply the above lemma to get a cycle C of length $n - \frac{\varepsilon n}{20} \geq 2k^2 + \frac{2\varepsilon k^2}{3}$ and outside of it a set S of size $|S| = \frac{\varepsilon n}{20}$, together with a matching M between them which covers S . Split the cycle C into $4/\varepsilon$ intervals of (almost) equal size. By pigeonholing, at least one of those intervals contains at least $\frac{\varepsilon^2 n}{80}$ endpoints of M . Let S' be the subset of S of vertices corresponding to those endpoints in M . We apply Lemma 2.6 to the graph $G[S']$ with say $\gamma = 1/100$, and conclude that there are two vertices x' and y' in $G[S']$, between which there exists a path of length ℓ in $G[S']$, for every $\ell \in [\frac{\varepsilon^2 k}{100}, \frac{\varepsilon^2 k}{50}]$.

Let x and y be the vertices in C corresponding to x' and y' in M . Let P be the longer path in C which connects x and y , so that by the choice of S' we have $|P| \geq |C| - \frac{\varepsilon n}{4} > 2k^2 + 2k$. Now we use Proposition 2.9 to show that in $G[P]$ the pair (x, y) is $\frac{\varepsilon^2 k}{100}$ -dense in $[\frac{900k}{\varepsilon^2}, 2k^2 + 2k]$. Denote $P_0 := P$, and we obtain the path P_i from path P_{i-1} by applying Proposition 2.9 with $c = \frac{\varepsilon^2 k}{400}$ and $U = \emptyset$. We do this until $|P_i| < \frac{900k}{\varepsilon^2}$ and then we stop our procedure. Notice we could do each step of the procedure as we always had $c \left(\frac{|P_i|}{2k} - 1 \right) > k$. Hence we obtain a sequence of xy -paths of decreasing lengths, where $|P_i| \geq |P_{i-1}| - 4c + 3 \geq |P_{i-1}| - \frac{\varepsilon^2 k}{100}$. Since $|P_0| > 2k^2 + 2k$ and the last path is of length at most $\frac{900k}{\varepsilon^2}$, we indeed get that in $G[P]$ the pair (x, y) is $\frac{\varepsilon^2 k}{100}$ -dense in $[\frac{900k}{\varepsilon^2}, 2k^2 + 2k]$.

Now, applying Observation 2.5 to the graph $G[P \cup S']$ with $v_1 = x$ and $v_2 = y$, gives all cycle lengths in $[\frac{1000k}{\varepsilon^2}, 2k^2 + 2k]$.

3.3 Middle range : from $(2 + \varepsilon)k$ to $\frac{1000}{\varepsilon^2}k$

By Lemma 3.3, there exists a partition of the vertices of G into a cycle C and a set S such that $|S| = \varepsilon n/20$, along with a matching $M \subseteq E[C, S]$ which covers S . Denote the vertices along the cycle C with $(1, \dots, N)$. For each vertex $x \in S$, we denote by $m(x)$ the vertex in C matched to x in M . We now remove from S

the at most $\frac{1000k}{\varepsilon^2}$ vertices x which have $m(x) \in \{1, 2, \dots, \frac{1000k}{\varepsilon^2}\}$. Hence, there are at least $\varepsilon n/22$ vertices remaining in S .

We let $B_1 := S$ and apply Proposition 2.6 with $\gamma = 1/100$ to the graph $G[B_1]$, to find a pair of vertices x_1, y_1 such that for all $\ell \in [\frac{\varepsilon k}{100}, \frac{\varepsilon k}{50}]$, there is an $x_1 y_1$ path of length ℓ in $G[S]$; we set $B_2 := B_1 - \{x_1, y_1\}$. We repeat this $t = \varepsilon k^2/40$ times, i.e., we apply Proposition 2.6 to B_i to obtain vertices x_i, y_i such that for all $\ell \in [\frac{\varepsilon k}{100}, \frac{\varepsilon k}{50}]$, there is an $x_i y_i$ -path of length ℓ in $G[B_i] \subseteq G[S]$, and we set $B_{i+1} := B_i - \{x_i, y_i\}$. Note that each B_i is of size at least $|S| - 2t \geq \frac{\varepsilon n}{22} - 2t \geq \frac{\varepsilon k^2}{11} - \frac{\varepsilon k^2}{20} \geq \frac{\varepsilon k^2}{30}$, so we always can successfully apply Proposition 2.6.

Before we make our first crucial observation, we introduce some notation. First, with possible re-labeling, let us suppose that for each i , we have that $m(x_i) < m(y_i)$. Now, for each i , let P_i denote the subpath $(m(x_i) - \frac{1000k}{\varepsilon^2}, m(x_i) - \frac{1000k}{\varepsilon^2} - 1, \dots, m(x_i) - 1, m(x_i))$ of length $\frac{1000k}{\varepsilon^2}$. Notice that, since $m(x_i) > \frac{1000k}{\varepsilon^2}$, none of those paths contains vertex 1.

Lemma 3.4. *If there is an i such that $G[P_i]$ does not contain an increasing path of length $\varepsilon^3 k$ as an induced subgraph with $m(x_i)$ as an endpoint¹, then G contains all cycle lengths in $[(2 + \varepsilon)k, \frac{1000k}{\varepsilon^2}]$.*

Proof. Assume for sake of contradiction that $G[P_i]$ does not contain such a path. Then, the following holds.

Claim. *In the graph $G[P_i]$, the endpoints of P_i are $\varepsilon^3 k$ -dense in $[0, \frac{1000k}{\varepsilon^2}]$.*

Proof of Claim. Consider the following procedure. We begin with P_i ; by assumption, there exists a chordal edge among the last $\varepsilon^3 k$ vertices of P_i ending with $m(x_i)$. Otherwise, these vertices would induce an increasing path in $G[P_i]$. Thus we obtain a path P'_i by adding this edge to P_i instead of the interval between the endpoints of this edge. We repeat this procedure, each time finding a chordal edge in the newly obtained path, and we do this until our path has length at most $\varepsilon^3 k$. Hence, we obtain a sequence of paths such that two consecutive paths lengths are at most $\varepsilon^3 k$ apart, while the last path has length at most $\varepsilon^3 k$. Since the endpoints always remain the same, this implies the statement of the claim. \square

Consider now the path P contained in the cycle C , and which is spanned by the vertices in the interval $[m(y_i), N] \cup [1, m(x_i) - \frac{1000k}{\varepsilon^2}]$. By Lemma 2.8, there exists a path P' with the following properties: $V(P') \subseteq V(P)$, the endpoints of P' are the same as those of P , and $|P'| \leq 2k$.

Then, in order to finish, recall that x_i, y_i are such that for all $\ell \in [\frac{\varepsilon k}{100}, \frac{\varepsilon k}{50}]$, there exists an $x_i y_i$ -path of length ℓ in $G[S]$. Further, by the above claim we got that in the graph $G[P_i]$, the endpoints of P_i are $\varepsilon^3 k$ -dense in $[0, \frac{1000k}{\varepsilon^2}]$. Hence we can use Observation 2.5 on the graph G with $v_1 = m(x_i), v_2 = m(x_i) - \frac{1000k}{\varepsilon^2}$ and $v_3 = m(y_i)$, where S_1 are the internal vertices of P_i , S_2 the internal vertices of P and $S_3 = S$ to obtain all cycle lengths in $[(2 + \varepsilon)k, \frac{1000k}{\varepsilon^2}]$. \square

We have $t = \varepsilon k^2/40$ paths P_i of length $1000k/\varepsilon^2$ such that each of them corresponds to an interval of vertices in C and therefore intersects at most $2000k/\varepsilon^2$ other such paths. Thus, we can choose a collection of $r = \frac{t}{2000k/\varepsilon^2 + 1} \geq \sqrt{k}$ of those paths which are all disjoint. With possible renaming, w.l.o.g. we may assume that those paths are P_1, \dots, P_r . Using Lemma 3.4, we may also assume that there are induced increasing subpaths Q_1, \dots, Q_r of $G[P_1], \dots, G[P_r]$ with endpoints $m(x_1), \dots, m(x_r)$ respectively and of length $\varepsilon^3 k$.

Let us now define an auxiliary colored complete graph H on $[r]$ in the following manner. For each $i \in [r]$, partition Q_i into three consecutive subpaths Q_i^3, Q_i^2, Q_i^1 of size $|Q_i|/3 = \varepsilon^3 k/3$, with Q_i^1 containing $m(x_i)$. Now, for $i, j \in [r]$, we color the edge ij in H *red* if in G both $E[Q_i^1, Q_j^1]$ and $E[Q_i^3, Q_j^3]$ are non-empty. We color it *blue* if $E[Q_i^1, Q_j^1] = \emptyset$, and in the remaining case, we color it *green*.

¹Recall that a path P is an induced subgraph of $G[P_i]$ if its vertices belong to $V(P_i)$ and except for the edges of P , $G[P]$ does not contain any other edges.

Claim. *There are no blue or green cliques in H of size larger than $6/\varepsilon^3$.*

Proof. Suppose there exists a blue clique $\{i_1, \dots, i_\ell\}$ in H . Since each $Q_{i_j}^1$ is an induced path, its odd vertices form an independent set of size $|Q_{i_j}^1|/2$. Moreover, by assumption, there are no edges between two $Q_{i_j}^1$'s and therefore the set $\bigcup_{1 \leq j \leq \ell} V(Q_{i_j}^1)$ must contain an independent set of size at least

$$\sum_{1 \leq j \leq \ell} \frac{|Q_{i_j}^1|}{2} \geq \ell \cdot (\varepsilon^3 k / 6).$$

Since $\alpha(G) \leq k$, this implies that $\ell \leq 6/\varepsilon^3$. An analogous argument deals with green cliques. \square

Given the above claim and Theorem 2.3, and since H has $r > \sqrt{k}$ vertices where we chose k large enough in terms of ε , we have that there exists a red clique in H of size at least $4\varepsilon^{-7}$.

Denote by I the vertices/indices contained in this clique, so that for all $i, j \in I$ we have that there is an edge e_{ij}^1 between Q_i^1 and Q_j^1 and an edge e_{ij}^3 between Q_i^3 and Q_j^3 . For simplicity of notation, w.l.o.g. we may assume that the indices in I are $\{1, 2, 3, \dots, |I|\}$ according to the ordering of the vertices $m(x_i)$ for $i \in I$ - that is, we now have $m(x_1) < m(x_2) < \dots < m(x_{|I|})$. Denote by z the endpoint of Q_1 which is not $m(x_1)$. In order to complete the proof, we will need the following lemma.

Lemma 3.5. *The path Q , defined by the interval $[z, m(x_{|I|})]$ is such that its endpoints are $3\varepsilon^3 k$ -dense in $[0, \frac{k}{\varepsilon^4}]$ in the graph $G[Q]$.*

Proof. Denote by R_1 the path consisting only of vertex z . We now recursively define for each $i < |I|$ a path R_i whose one endpoint is z and the other endpoint z_i lies in either Q_i^1 or Q_i^3 (see Figure 3 for an illustration). First, let $z_1 = z$. Suppose z_i is in Q_i^a for some $a \in \{1, 3\}$, and let $b \in \{1, 3\} \setminus \{a\}$; let R_{i+1} be the path obtained from R_i by concatenating to it the path contained in Q_i which starts at z_i , goes through Q_i^b and touches the edge $e_{i,i+1}^b$, and also add the edge $e_{i,i+1}^b$ itself.

Now we also define paths R'_i for each $i \leq |I| - 1$, obtained from R_i as follows. If $z_i \in Q_i^a$, again we let $b \neq a$ and $b \in \{1, 3\}$. Let R'_i be the path obtained by concatenating with R_i the path starting at z_i , going through Q_i until the edge $e_{i,|I|}^b$, then also adding this edge $e_{i,|I|}^b$ itself, together with the path in $Q_{|I|}$ which connects the endpoint of this edge with $m(x_{|I|})$.

Note that the length of two paths R'_i and R'_{i+1} differs by at most $|Q_i| + |Q_{i+1}| + |Q_{|I|} \leq 3\varepsilon^3 k$, since the only vertices which belong to exactly one of these two paths are contained in $G[Q_i \cup Q_{i+1} \cup Q_{|I|}]$. Furthermore, the length of the first path R'_1 obtained by our procedure is at most $|Q_1| + |Q_{|I|} \leq 2\varepsilon^3 k$ and the length of the last path $R'_{|I|-1}$ is at least $(|I| - 1)\varepsilon^3 k / 3 \geq \frac{k}{\varepsilon^4}$, since it contains all paths Q_i^2 for all $i \leq |I| - 1$. This implies that the path Q , defined as the path between $z = z_1$ and $m(x_{|I|})$ is such that its endpoints are $3\varepsilon^3 k$ -dense in $[0, \frac{k}{\varepsilon^4}]$ in the graph $G[Q]$. \square

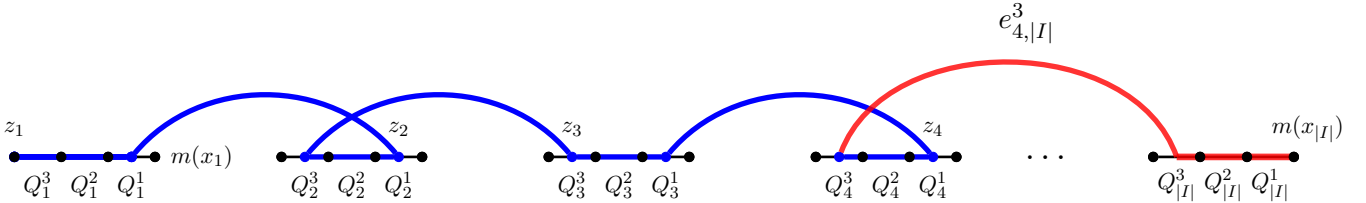


Figure 3: The thick blue path represents path R_4 , and adding to it the red path creates R'_4 .

Let Q^* be the path in C spanned by the interval $[1, z] \cup [m(y_{|I|}), N]$. By applying Lemma 2.8, we get a path Q' of length at most $2k$ in $G[Q^*]$ with the same endpoints as Q^* . Recalling that the pair of vertices $x_{|I|}, y_{|I|}$ is connected by paths of all lengths in $[\frac{\varepsilon k}{100}, \frac{\varepsilon k}{50}]$ in the subgraph $G[S]$, and Lemma 3.5 above, we are done by Observation 2.5. Indeed, we apply it to G with $v_1 = z$, $v_2 = m(y_{|I|})$ and $v_3 = m(x_{|I|})$, while S_1 is the set of internal vertices of Q^* , $S_2 = S$ and S_3 are the internal vertices of Q to get all cycle lengths in the interval $[(2 + \varepsilon)k, \frac{1000}{\varepsilon^2}k]$. \square

4 Concluding remarks

In this paper we proved that every Hamiltonian graph on $n \geq 2k^2 + o(k^2)$ vertices with independence number k is pancyclic, which is tight up to the $o(k^2)$ error term. Furthermore, our methods allow us to give a short proof of Erdős's conjecture that $n = \Omega(k^2)$ vertices are enough for G to be pancyclic. For this, we first note that while getting a bound of $n = \Omega(k^3)$, Keevash and Sudakov [22] implicitly proved the following result.

Lemma 4.1 ([22]). *There exists a large constant C such that every Hamiltonian graph on $n \geq Ck^2$ vertices with independence number k , contains all cycle lengths in $[3, n/C]$.*

This reduces Erdős's conjecture to the following problem.

(*) Does there exist $C' > 0$ such that every Hamiltonian graph on $n \geq C'k^2$ vertices with independence number k , contains a cycle of length $n - 1$?

Indeed, suppose that the above is true for some large constant C' and let G be a Hamiltonian graph on n vertices with independence number k . Then, combining this with the above lemma of Keevash and Sudakov, it follows that if $n \geq CC'k^2$ then G is pancyclic, thus proving Erdős's conjecture. Indeed, note that by the lemma above, G contains all cycle lengths up to n/C and one can see that it contains all cycle lengths from n/C to n by iteratively applying the assumption that whenever $n' \geq C'k^2$, there is a cycle of length $n' - 1$. The previous results [22], [27] and [11] are all improvements towards question (*) above. As discussed in the beginning of Section 3.2, applying Proposition 2.9 with $c = 1$ and $U = \emptyset$ solves this problem in the following stronger form.

Theorem 4.2. *Every Hamiltonian graph on $n > 2k^2 + 2k$ vertices with independence number k , contains a cycle of length $n - 1$.*

Let us note that although Theorem 4.2 shows the existence of a cycle of length $n - 1$ already with $n > 2k^2 + 2k$, this is not sufficient to prove that $n > 2k^2 + \varepsilon k^2$ implies pancyclicity. At this threshold, Lemma 4.1 does not apply, so one needs a different argument to find the cycle lengths in the interval $[3, 2k^2 + 2k]$. It turns out that in this setting, the cycle lengths which are hardest to find are those around $2k$, that is, precisely the cycle lengths which are missed by the lower bound construction given in the introduction. Finding them is the most technical part of our proof, given in Section 3.3.

A very interesting open question is to understand the best bound on the number of vertices n in (*) which guarantees the cycle of length $n - 1$. Here the answer might be linear in n as the following question asked in [22] suggests.

Problem 4.3. Does there exist a constant C such that every Hamiltonian graph with independence number k and $n \geq Ck$ vertices contains a cycle of length $n - 1$?

Acknowledgements. We thank the referee for their careful reading of the paper and for useful comments.

References

- [1] M. Ajtai, J. Komlós, and E. Szeraerédi. First occurrence of Hamilton cycles in random graphs. In *Cycles in graphs* (Burnaby, B.C., 1982), Vol. 115, North-Holland Mathematical Studies, North-Holland, Amsterdam, 115:173–178, 1985.
- [2] D. Bauer and E. Schmeichel. Hamiltonian degree conditions which imply a graph is pancyclic. *Journal of Combinatorial Theory, Series B*, 48(1):111–116, 1990.
- [3] J. A. Bondy. Pancyclic graphs: recent results, infinite and finite sets. In *Colloq. Math. Soc. János Bolyai*, Keszthely, pp. 181–187, 1973.
- [4] J. A. Bondy. Pancyclic graphs I. *Journal of Combinatorial Theory, Series B*, 11(1):80–84, 1971.
- [5] J. A. Bondy. *Longest paths and cycles in graphs of high degree*. Department of Combinatorics and Optimization, University of Waterloo, 1980.
- [6] M. Bucić, L. Gishboliner, and B. Sudakov. Cycles of many lengths in Hamiltonian graphs. *Forum of Mathematics, Sigma*, 10, E70, 2022.
- [7] V. Chvátal. On Hamilton’s ideals. *Journal of Combinatorial Theory, Series B*, 12(2):163–168, 1972.
- [8] V. Chvátal and P. Erdős. A note on Hamiltonian circuits. *Discrete Mathematics*, 2(2):111–113, 1972.
- [9] B. Csaba, D. Kühn, A. Lo, D. Osthus, and A. Treglown. Proof of the 1-factorization and Hamilton decomposition conjectures. *Memoirs of the American Mathematical Society*, 244, monograph 1154, 170 pages, 2016.
- [10] B. Cuckler and J. Kahn. Hamiltonian cycles in Dirac graphs. *Combinatorica*, 29(3):299–326, 2009.
- [11] A. Dankovics. Low independence number and Hamiltonicity implies pancyclicity. *Journal of Graph Theory*, 95(2):181–191, 2020.
- [12] G. A. Dirac. Some theorems on abstract graphs. *Proceedings of the London Mathematical Society*, 3(1):69–81, 1952.
- [13] N. Draganić, D. Munhá Correia, and B. Sudakov. Chvátal-Erdős condition for pancyclicity. *arXiv preprint arXiv:2301.10190*, 2023.
- [14] P. Erdős. Some problems in graph theory. In *Hypergraph Seminar*, pp. 187–190. Springer, 1972.
- [15] P. Erdős, R. J. Faudree, C. C. Rousseau, and R. H. Schelp. On cycle-complete graph Ramsey numbers. *Journal of Graph Theory*, 2(1):53–64, 1978.
- [16] G. H. Fan. New sufficient conditions for cycles in graphs. *Journal of Combinatorial Theory, Series B*, 37(3):221–227, 1984.
- [17] A. Ferber, E. Long, and B. Sudakov. Counting Hamilton decompositions of oriented graphs. *International Mathematics Research Notices*, 2018(22):6908–6933, 2018.
- [18] T. Gallai and A.N. Milgram. Verallgemeinerung eines graphentheoretischen Satzes von Rédei. *Acta Sc. Math*, 21:181–186, 1960.
- [19] R. J. Gould. Recent advances on the Hamiltonian problem: Survey III. *Graphs and Combinatorics*, 30(1):1–46, 2014.
- [20] B. Jackson and O. Ordaz. Chvátal-Erdős conditions for paths and cycles in graphs and digraphs. A survey. *Discrete mathematics*, 84(3):241–254, 1990.
- [21] P. Keevash, E. Long, and J. Skokan. Cycle-complete Ramsey numbers. *International Mathematics Research Notices*, 2021(1):275–300, 2021.
- [22] P. Keevash and B. Sudakov. Pancyclicity of Hamiltonian and highly connected graphs. *Journal of Combinatorial Theory, Series B*, 100(5):456–467, 2010.
- [23] M. Krivelevich. The critical bias for the Hamiltonicity game is $(1 + o(1))n/\ln n$. *Journal of the American Mathematical Society*, 24(1):125–131, 2011.
- [24] M. Krivelevich, C. Lee, and B. Sudakov. Robust Hamiltonicity of Dirac graphs. *Transactions of the American Mathematical Society*, 366(6):3095–3130, 2014.
- [25] D. Kühn and D. Osthus. Hamilton decompositions of regular expanders: a proof of Kelly’s conjecture for large tournaments. *Advances in Mathematics*, 237:62–146, 2013.
- [26] D. Kühn and D. Osthus. Hamilton cycles in graphs and hypergraphs: an extremal perspective. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. IV*, pp. 381–406. Kyung Moon Sa, Seoul, 2014.

- [27] C. Lee and B. Sudakov. Hamiltonicity, independence number, and pancyclicity. *European Journal of Combinatorics*, 33(4):449–457, 2012.
- [28] S. Letzter. Pancyclicity of highly connected graphs. *arXiv preprint arXiv:2306.12579*, 2023.
- [29] K. G. Milans, F. Pfender, D. Rautenbach, F. Regen, and D. B. West. Cycle spectra of Hamiltonian graphs. *Journal of Combinatorial Theory, Series B*, 102(4):869–874, 2012.
- [30] L. Pósa, Hamiltonian circuits in random graphs. *Discrete Mathematics*, 14(4):359–364, 1976.
- [31] E. F. Schmeichel and S. L. Hakimi. A cycle structure theorem for Hamiltonian graphs. *Journal of Combinatorial Theory, Series B*, 45(1):99–107, 1988.
- [32] J. Verstraëte. Extremal problems for cycles in graphs. In *Recent trends in combinatorics*, pp. 83–116. Springer, 2016.