Hamilton Cycles in Random Subgraphs of Pseudo-Random Graphs

Alan Frieze* Michael Krivelevich†

September 20, 2001

Abstract

Given an r-regular graph G on n vertices with a Hamilton cycle, order its edges randomly and insert them one by one according to the chosen order, starting from the empty graph. We prove that if the eigenvalue of the adjacency matrix of G with the second largest absolute value satisfies $\lambda = o(r^{5/2}/(n^{3/2}(\log n)^{3/2}))$, then for almost all orderings of the edges of G at the very moment τ^* when all degrees of the obtained random subgraph H_{τ^*} of G become at least two, H_{τ^*} has a Hamilton cycle. As a consequence we derive the value of the threshold for the appearance of a Hamilton cycle in a random subgraph of a pseudo-random graph G, satisfying the above stated condition.

Key-words Pseudo-random Graphs, Hamilton Cycles, Random Graphs.

1 Introduction

Pseudo-random graphs (sometimes also called quasi-random graphs) can be informally defined as graphs whose edge distribution resembles closely that of truly random graphs on the same number of vertices and with the same edge density. Pseudo-random graphs, their constructions and properties have been a subject of intensive study for the last fifteen years (see [15], [16], [8], [14], [3], to mention just a few).

*Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA15213, U.S.A. Supported in part by NSF grant CCR-9818411.

†Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: krivelev@math.tau.ac.il, Research supported in part by USA-Israel BSF Grant 99-0013 and by a Bergmann Memorial Grant.

For the purposes of this paper, a pseudo-random graph is an r-regular graph $G = (V, E)$ with vertex set $V = [n] = \{1, \ldots, n\}$, all of whose eigenvalues but the first one are significantly smaller than r in their absolute values. More formally, let $A = A(G)$ be the adjacency matrix of G. This is an n-by-n matrix such that $A_{ij} = 1$ if $(i, j) \in E(G)$ and $A_{ij} = 0$ otherwise. Then A is a real symmetric matrix with non-negative values of its entries. Let $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_n$ be the eigenvalues of A, usually called also the eigenvalues of G. It follows from the Perron-Frobenius theorem that $\lambda_1 = r$ and $|\lambda_i| \leq r$ for all $2 \leq i \leq n$. We thus denote $\lambda = \lambda(G) = \max_{2 \leq i \leq n} |\lambda_i|$. The reader is referred to a monograph of Chung [7] for further information on spectral graph theory.

It is known (see, e.g. [2]) that the greater is the so-called spectral gap (i.e. the difference between r and λ) the more tightly the distribution of the edges of G approaches that of the random graph $G(n, r/n)$. We will cite relevant quantitative results later in the text (see (6), (7)), for now we just state informally that a spectral gap ensures pseudo-randomness. Thus in the rest of the paper we will stick to the view of a pseudo-random graph as an r-regular graph on n vertices with $\lambda \ll r$.

In this paper we study certain properties of a random subgraph of a pseudo-random graph. Given a graph $G = (V, E)$ and an edge probability $0 \leq p = p(n) \leq 1$, the *random subgraph* G_p is formed by choosing each edge of G independently and with probability p. The most studied random graph is the so called binomial random graph $G(n, p)$, formed by choosing the edges of the complete graph on n labeled vertices independently with probability p. Here rather than studying random subgraphs of one particular graph, we investigate the properties of random subgraphs of graphs from a wide class of regular pseudo-random graphs. As we will see, all such subgraphs viewed as probability spaces share certain common features.

Another related notion is that of a *graph process*. Given a graph $G = (V, E)$, choose a permutation $\sigma = (e_1, e_2, \ldots)$ of the edges of E uniformly at random and then define an increasing sequence of subgraphs (G_m) of G, where $G_m = (V, E_m)$ and $E_m = \{e_1, \ldots, e_m\}$. This sequence is called a random graph process. Choosing a random graph process (G_m) and then "stopping" it at time i is easily seen to produce a subgraph of G with i edges, chosen uniformly at random from all such subgraphs. We will thus denote by G_i the probability space of the subgraphs of G with i edges and with the uniform measure. Random subgraphs and graph processes are intimately related, and studying one of these two probabilistic objects usually provides immediate consequences for its counterpart. Our paper is not exceptional in this aspect, we will draw conclusions about random subgraphs based on studying random graph processes.

Given a graph process (G_m) and a graph property $\mathcal{A}$ possessed by G, the *hitting time* $\tau_{\mathcal{A}}$ is the minimal m, $0 \leq m \leq |E(G)|$, for which the subgraph G_m has $\mathcal{A}$.

As customary when studying random graphs, asymptotic conventions and notations apply. In particular, we assume where necessary the number of vertices n of the base graph G to be as large as needed. Also, we say that a graph property $\mathcal{A}$ holds *with high probability*, or

whp for brevity, in G_p if the probability that G_p has $\mathcal{A}$ tends to 1 as n tends to infinity. A recent monograph [11] provides a necessary background and reflects the state of affairs in the theory of random graphs.

The subject of this paper is Hamilton cycles in random subgraphs of pseudo-random graphs. Observe that in order for a random graph G_m to contain a Hamilton cycle all vertices should have degree at least two in G_m. Thus the corresponding graph process problem can be formulated in general as follows: given a graph G with a Hamilton cycle, is it true that for almost all graph processes (G_m) the first Hamilton cycle appears exactly at the moment when all vertex degrees become at least two? This problem has been solved in the affirmative for the case $G = K_n$, i.e. for the model $G(n,p)$, by Bollobás [5], based on a breakthrough technique developed by Posa in [13]. The ingenious rotation-extension technique of Posa plays a central role in our arguments as well.

Our main result can be formulated as follows. Let $G = ([n], E)$ be an r-regular graph with a Hamilton cycle. Consider a hitting time problem. Let $e_1, e_2, \ldots, e_N, N = rn/2$ be a random ordering of the edges of G. For $0 \le m \le |E|$ let $E_m = \{e_1, e_2, \ldots, e_m\}$ and let $G_m = ([n], E_m)$. Now consider two *hitting times.*

$$\begin{aligned} \tau_2 = \tau_2(G) &= \min\{m : \ \delta(G_m) \ge 2\}. \\ \tau_H = \tau_H(G) &= \min\{m : \ G_m \text{ is Hamiltonian}\}. \end{aligned}$$

Let λ denote the second largest by absolute value eigenvalue of the adjacency matrix of G.

Theorem 1 *If* $\lambda = o\left(\frac{r^{5/2}}{n^{3/2}(\log n)^{3/2}}\right)$ *then* **whp** $\tau_2(G) = \tau_H(G)$.

Remark 1 *Since* $\lambda = \Omega(r^{1/2})$ *(see, e.g., [12]), the condition of the theorem implies that*

$$r \gg n^{3/4}(\log n)^3.$$

Throughout the paper we omit systematically all floor and ceiling signs for the sake of clarity of presentation. All logarithms are natural.

2 Proof of Theorem 1

Theorem 1 follows immediately from the Lemmas 1 and 2 below. We start with some notation. Let $H = ([n], F)$ be a graph on vertex set $[n]$. Let d_H denote degree in H. Let

$$\text{LARGE} = \text{LARGE}(H) = \{v \in [n] : \ d_H(v) \ge \frac{\log n}{10}\}$$

and SMALL = SMALL(H) = $[n] \setminus$ LARGE. Vertices $v \in$ SMALL will be called *small vertices*.

For $S \subseteq [n]$ we let $N_H(S) = \{w \in [n] : \exists v \in S$ such that $(v,w) \in F\}$.

Consider the following list of properties:

P1 $\delta(H) \geq 2$.

P2 SMALL contains no edges.

P3 No $v \in V$ is within distance 2 of more than one small vertex.

P4 $S \subseteq$ LARGE, $|S| \leq n\frac{\log\log n}{\log n}$ implies that $|N_H(S)| \geq |S|\frac{\log n}{10\log\log n}$.

P5 $A, B \subseteq V$, $A \cap B = \emptyset$, $|A|, |B| \geq 20n\frac{\log\log n}{\log n}$ implies that H contains at least $|A||B|\frac{\log n}{2n}$ edges joining A and B.

P6 $A, B \subseteq V$, $A \cap B = \emptyset$, $|A| \leq |B| \leq 4|A|$ and $|B| \leq 200n\frac{\log\log n}{\log n}$ implies that there are at most $2400|A|\log\log n$ edges joining A and B.

P7 If $|A| \leq 30n\frac{\log\log n}{\log n}$ then A contains at most $100|A|\log\log n$ edges.

Lemma 1 *Let $H = ([n], F)$ satisfy P1–P7 above. Then H is Hamiltonian.*

$\square$

Lemma 2 *G_{τ_2} satisfies P1–P7* **whp**.

$\square$

3 Proof of Lemma 1

We assume throughout this section that P1–P7 hold. We first prove

Lemma 3 *H is connected.*

Proof If H is not connected then from P5 it has a component C of size at most $20n\frac{\log\log n}{\log n}$. But then P3 and P4 imply $C \subseteq$ SMALL. P1 and P2 give a contradiction. $\square$

3.1 Construction of an initial long path

We use rotations and extensions in H to find a maximal path with large rotation endpoint sets, see for example [6], [9]. Let $P_0 = (v_1, v_2, \ldots, v_l)$ be a path of maximum length in H. If $1 \leq i < l$ and $\{v_l, v_i\}$ is an edge of H then $P' = (v_1 v_2 \ldots v_i v_l v_{l-1} \ldots v_{i+1})$ is also of maximum length. It is called a *rotation* of P_0 with *fixed endpoint* v_1 and *pivot* v_i. Edge (v_i, v_{i+1}) is called the *broken* edge of the rotation. We can then, in general, rotate P' to get more maximum length paths.

For $t \geq 0$ let $S_t = \{v \in \text{LARGE}: \ v \neq v_1$, is the endpoint of a path obtainable from P_0 by t rotations with fixed endpoint v_1 and all broken edges in $P_0\}$.

It follows from P1, P3 and P4 that $S_1 \neq \emptyset$. It then follows that if $|S_t| \leq n^{\frac{\log\log n}{\log n}}$ then $|S_{t+1}| \geq |S_t| \frac{\log n}{21 \log\log n}$. We prove this by induction. It is clearly true for $t = 0$ and

$$\begin{aligned} |S_{t+1}| &\geq \frac{1}{2}|N_H(S_t)| - (1 + |S_1| + |S_2| + \cdots + |S_t|) \qquad (1) \\ &\geq |S_t| \frac{\log n}{20 \log\log n} - (1 + |S_1| + |S_2| + \cdots + |S_t|) \\ &\geq |S_t| \frac{\log n}{21 \log\log n}. \end{aligned}$$

Explanation of (1) Let

$$T = \{i \geq 2: \ v_i \in N_H(S_t), v_{i-1}, v_{i+1} \notin \bigcup_{\tau=0}^{t} S_\tau\}. \qquad (2)$$

Note that if $v_1 \in N_H(S_t)$ then we know already that H is Hamiltonian. It is connected and there is a cycle containing a longest path. So we assume that $v_1 \notin N_H(S_t)$

Suppose now that $x = v_i \in T$ and $y \in S_t$ are neighbours. t rotations starting with P_0 produce a path Q with y as an endpoint. Now, $v_{i-1}, v_{i+1} \notin \bigcup_{\tau=0}^{t} S_\tau$ and so a rotation with pivot v_i will place one of v_{i-1}, v_{i+1} in S_{t+1}. Suppose for example that it is v_{i-1}. The only other vertex other than v_i that can place v_{i-1} in S_{t+1} in this way is v_{i-2}. Thus we obtain at least half the RHS of (2) in this way as members of S_{t+1}.

Thus there exists $t_0 \leq (1 + o(1)) \log n / \log\log n$ such that $|S_{t_0}| \geq cn$, $c = 1/21$. Let $B(v_1) = S_{t_0}$ and $A_0 = B(v_1) \cup \{v_1\}$. Similarly, for each $v \in B(v_1)$ we can construct a set $B(v)$, $|B(v)| \geq cn$, of endpoints of maximum length paths with endpoint v. Note that $l \geq cn$ as every vertex of S_{t_0} lies on P_0.

In summary, for each $a \in A_0$, $b \in B(a)$ there is a maximum length path $P(a, b)$ joining a and b and this path is obtainable from P_0 by at most $(2 + o(1)) \log n / \log\log n$ rotations.

3.2 Closure of the maximal path

This section follows closely both the notation and the proof methodology used in [1].

Given path P_0 and a set of vertices S of P_0, we say $s \in S$ is an *interior* point of S if both neighbours of s on P_0 are also in S. The set of all interior points of S will be denoted by $int(S)$.

Lemma 4 *Given a set S of vertices with $|int(S)| \geq 120n\frac{\log\log n}{\log n}$, there is a subset $S' \subseteq S$ such that, for all $s' \in S'$ there are at least $k = \frac{\log n}{5n}|int(S)|$ edges between s' and $int(S')$. Moreover, $|int(S')| \geq |int(S)|/2$.*

Proof We use the proof given in [1]. If there is $s_1 \in S$ such that the number of edges from s_1 to $int(S)$ is less than k we delete s_1, and define $S_1 = S \setminus \{s_1\}$. If possible we repeat this procedure for S_1, to define $S_2 = S_1 \setminus \{s_2\}$ (etc). If this continued for $r = \frac{1}{6}|int(S)|$ steps, we would have a set S_r and a set $R = \{s_1, s_2, \ldots, s_r\}$, with

$$|int(S_r)| \geq |int(S)| - 3r \geq \frac{|int(S)|}{2}.$$

However, there are fewer than

$$k|R| = \frac{\log n}{5n}|int(S)|\;|R| \leq \frac{2\log n}{5n}|int(S_r)|\;|R|,$$

edges from R to $int(S_r)$, which contradicts P5. □

In Section 3.1 we proved the existence of maximum length paths $P(a,b),\ b \in B(a),\ a \in A_0$ where $|A_0|,\ |B(a)| \geq cn$. Thus there are at least c^2n^2 distinct endpoint pairs (a,b) and for each such pair there is a path $P(a,b)$ derived from at most $\rho = (2+o(1))\log n/\log\log n$ rotations starting with some fixed maximal path P_0.

We consider P_0 to be directed and divided into 2ρ segments $I_1, I_2, \ldots, I_{2\rho}$ of length at least $\lfloor |P_0|/2\rho \rfloor$, where $|P_0| \geq cn$. As each $P(a,b)$ is obtained from P_0 by at most ρ rotations, the number of segments of P_0 which occur on this path, although perhaps reversed, is at least ρ. We say that such a segment is *unbroken*. These segments have an absolute orientation given by P_0, and another, relative to this by $P(a,b)$, which we regard as directed from a to b. Let t be a fixed natural number. We consider sequences $\sigma = I_{i_1}, ..., I_{i_t}$ of unbroken segments of P_0, which occur in this order on $P(a,b)$, where we consider that σ also specifies the relative orientation of each segment. We call such a sequence σ a *t-sequence*, and say $P(a,b)$ *contains* σ.

For given σ, we consider the set $L = L(\sigma)$ of ordered pairs (a,b), $a \in A_0,\ b \in B(a)$ which contain the sequence σ.

The total number of such sequences of length t is $(2\rho)_t 2^t$. Any path $P(a,b)$ contains at least $\rho \geq \log n/\log\log n$ unbroken segments, and thus at least $\binom{\rho}{t}$ t-sequences. The average, over

t-sequences, of the number of pairs containing a given t-sequence is therefore at least

$$c^2n^2\frac{\binom{\rho}{t}}{(2\rho)_t2^t} \geq \alpha n^2,$$

where $\alpha = c^2/(4t)^t$. Thus there is a t-sequence σ_0 and a set $L = L(\sigma_0)$, $|L| \geq \alpha n^2$ of pairs (a,b) such that for each $(a,b) \in L$ the path $P(a,b)$ contains σ_0. Let $\hat{A} = \{a :\ L$ contains at least $\alpha n/2$ pairs with a as first element$\}$. Then $|\hat{A}| \geq \alpha n/2$. For each $a \in \hat{A}$ let $\hat{B}(a) = \{b :\ (a,b) \in L\}$.

Let $t = 480000/c$ and let C_1 denote the union of the first $t/2$ segments of σ_0, in the fixed order and with the fixed relative orientation in which they occur along *any* of the paths $P(a,b)$, $(a,b) \in L$. Let C_2 denote the union of the second $t/2$ segments of σ_0. C_1 and C_2 both contain at least $\frac{t}{2}\ cn\ \frac{\log\log n}{4\log n}(1-o(1))$ interior points which from Lemma 4 gives sets C_1', C_2' with at least

$$\frac{tc(1-o(1))}{16}n\frac{\log\log n}{\log n} \geq 30000n\frac{\log\log n}{\log n} \tag{3}$$

interior points and with all vertices $v \in C_i'$ having at least $\frac{\log n}{5n}|C_i'|$ neighbors in the corresponding sets $int(C_i')$.

It follows from P5 that there exists $\hat{a} \in \hat{A}$ such that H contains an edge from $\hat{a}$ to $int(C_1')$. Similarly, H contains an edge joining some $\hat{b} \in \hat{B}(\hat{a})$ to $int(C_2')$. Let x be some vertex separating C_1' and C_2' along $\hat{P} = P(\hat{a},\hat{b})$. We now consider the two half paths P_1, P_2 obtained by splitting $\hat{P}$ at x. We consider rotations of P_i, $i = 1,2$ with x as a fixed endpoint. We show that in both cases the finally constructed endpoint sets V_1, V_2 are large enough so that P5 guarantees an edge $e \in E(H)$ from V_1 to V_2. Since in a connected graph every non-Hamiltonian cycle can be augmented to a path with a larger number of vertices, we deduce that H is Hamiltonian as the path e closes is of maximum length.

Consider P_1. Let $T_i = \{v \in C_1' :\ v \neq x$ is the endpoint of a path obtainable from P_1 by i rotations with fixed endpoint x, all pivots in $int(C_1')$ and all broken edges in $P_1\}$. We claim we can choose sets $U_i \subseteq T_i$, $i = 1,2,\ldots$ such that $|U_1| = 1$ and $|U_{i+1}| = 2|U_i|$, as long as $|U_i| \leq 15n\frac{\log\log n}{\log n}$. Thus there is an i^* such that $|U_{i^*}| \geq 15n\frac{\log\log n}{\log n}$ and we are done. Note that $T_1 \neq \emptyset$ because $\hat{a}$ has an H-neighbour in $int(C_1')$. Note also that if we make a rotation with pivot in $int(C_1')$ and broken edge in P_1 then the new endpoint created is in C_1'.

Let y be a vertex of U_i. Then by Lemma 4 and (3) there are at least $6000\log\log n$ edges between y and $int(C_1')$. Thus the number of edges from U_i to $int(C_1')$ is at least $3000|U_i|\log\log n$. As $|\bigcup_{j=1}^{i} U_i| < 2|U_i|$ at most $200|U_i|\log\log n$ of these edges are contained in $\bigcup_{j=1}^{i} U_j$ (from P7), and so by P6 we have $|T_{i+1}| \geq \frac{1}{2}|N_H(U_i) \cap int(C_1')| > 2|U_i|$ and we can select a subset of size exactly $2|U_i|$. □

4 Proof of Lemma 2

For $0 \leq p \leq 1$ let $G_p = ([n], E_p)$ denote the random subgraph of G where each edge of E is independently included in E_p with probability p.

Let $\mathcal{A}$ be a property of graphs. Such a property is monotone increasing if $G \in \mathcal{A}$ implies that $G' \in \mathcal{A}$ for all G' which contain G as a subgraph. A similar definition holds for monotone decreasing properties. We state the following easily verified results: Let $p = m/N$.

$$\mathbf{Pr}(G_m \in \mathcal{A}) \leq 3m^{1/2}\mathbf{Pr}(G_p \in \mathcal{A}). \tag{4}$$

If $\mathcal{A}$ is monotone then

$$G_m \in \mathcal{A} \textbf{ whp} \text{ iff } G_p \in \mathcal{A} \textbf{ whp}. \tag{5}$$

We also need the following results related to edge density in G. They are taken from Alon and Spencer [4], Chapter 9.

Recall that G is an r-regular graph with vertex set $V(G) = [n]$. Let $K, L \subseteq [n]$, $|K| = k, |L| = \ell$ be disjoint. Let $e_G(K)$ denote the number of G-edges which are contained in K and let $e_G(K, L)$ denote the number of G-edges joining K and L. Then

$$\left| e_G(K) - \frac{k^2}{2}\frac{r}{n} \right| \leq \frac{1}{2}\lambda k \tag{6}$$

$$\left| e_G(K, L) - k\ell\frac{r}{n} \right| \leq \lambda\sqrt{k\ell} \tag{7}$$

Thus if λ is small, the values of $e_G(K), e_G(K, L)$ are close to what one would expect in the standard random graph model $G_{n,r/n}$.

Let

$$m_i = \frac{1}{2}n(\log n + \log\log n + (2i-3)\log\log\log n),\ i = 1, 2,$$

and where $N = rn/2$ let

$$p_i = \frac{2m_i}{N} = \frac{1}{r}(\log n + \log\log n + (2i-3)\log\log\log n),\ i = 1, 2.$$

Lemma 5

$$m_1 \leq \tau_2 \leq m_2 \textbf{ whp}.$$

Proof Having minimum degree at least two is a monotone property and so we can prove the lemma by verifying that

$$\delta(G_{p_1}) < 2 \textbf{ whp} \text{ and } \delta(G_{p_2}) \geq 2 \textbf{ whp},$$

and then applying (5).

Let Z_i denote the number of vertices of degree 0 or 1 in $G_{p_i}, i = 1,2$. Then

$$\begin{aligned}\mathbf{E}(Z_i) &= n((1-p_i)^r + rp_i(1-p_i)^{r-1}) \\ &= (1+o(1))(\log\log n)^{3-2i}.\end{aligned}$$

Thus $\mathbf{E}(Z_2) = o(1)$ and $\delta(G_{p_2}) \geq 2$ **whp**.

Furthermore, $\mathbf{E}(Z_1) \approx \log\log n$ and

$$\begin{aligned}\mathbf{E}(Z_1(Z_1-1)) = 2N(p_1(1-p_1)^{2r-2} + (1-p_1)((1-p_1)^{r-1} + (r-1)p_1(1-p_1)^{r-2})^2 \\ + (n(n-1)-2N)((1-p_i)^r + rp_i(1-p_i)^{r-1})^2 \\ \approx (\log\log n)^2\end{aligned}$$

and so

$$\mathbf{Pr}(Z_1 \neq 0) \geq \frac{\mathbf{E}(Z_1)^2}{\mathbf{E}(Z_1^2)} \to 1.$$

□

We now go through the list of properties P1–P7 and confirm them one by one. Clearly G_{τ_2} satisfies P1 and so we start with P2.

P2 and P3:
We will prove that **whp** there do not exist $v, w \in \text{SMALL}(G_{m_1})$ such that v, w are joined by a path of length 4 or less in G_{m_2}. This implies that P2 and P3 both hold in G_{τ_2}. Now if Z denotes the number of such pairs v, w then

$$\begin{aligned}\mathbf{E}(Z) \leq n\sum_{t=1}^{4} r^t \sum_{k=0}^{(\log n)/5} \frac{\binom{2r-2}{k}\binom{N-2r+2-t}{m_1-k-t}}{\binom{N}{m_1}} \\ \leq (1+o(1))n\sum_{t=1}^{4} r^t \sum_{k=0}^{(\log n)/5} \left(\frac{2re}{k}\right)^k \left(\frac{m_1}{N}\right)^{k+t} \left(1-\frac{2r}{N}\right)^{m_1} \\ \leq 2n^{-1}(10e)^{(\log n)/5}\sum_{t=1}^{4}(\log n)^t \\ = o(1).\end{aligned}$$

P4:
Suppose that $K \subseteq [n]$ with $|K| = k$. We show first that **whp**

$$k \leq \frac{n}{2(\log n)^{3/2}} \text{ implies } e_{m_2}(K) \leq 3k \tag{8}$$

for every such subset K, where $e_{m_2}(K)$ is the number of G_{m_2} edges contained in K.

It follows from (6) that

$$e_G(K) \leq \frac{1}{2}k^2\frac{r}{N} + \frac{1}{2}\lambda k. \tag{9}$$

Case 1:

$$k \leq \frac{\lambda n}{r} \leq \frac{r^{5/2}}{\omega n^{3/2}(\log n)^{3/2}} \cdot \frac{n}{r} = \frac{r^{3/2}}{\omega n^{1/2}(\log n)^{3/2}}$$

where $\omega \to \infty$.

Let $e_{p_2}(K)$ denote the number of G_{p_2} edges contained in K.

$$\begin{aligned}
\mathbf{Pr}(\exists K:\ |K| \leq \lambda n/r,\ e_{p_2}(K) > 3k) &\leq \sum_{k=8}^{\lambda n/r} \binom{n}{k}\binom{\binom{k}{2}}{3k} p_2^{3k} \\
&\leq \sum_{k=8}^{\lambda n/r} \left(\frac{ne}{k}\left(\frac{k\log n}{2r}\right)^3\right)^k \\
&\leq \sum_{k=8}^{\lambda n/r} \left(\frac{n}{2}\frac{k^2(\log n)^3}{r^3}\right)^k \\
&\leq \sum_{k=8}^{\lambda n/r} (2\omega^2)^{-k} \\
&= o(1).
\end{aligned}$$

The existence of K containing $\geq 3|K|$ edges is monotone and so we see from (5) that **whp** $e_{m_2}(K) \leq 3|K|$ for $|K| \leq \lambda n/r$.

Case 2: $k > \lambda n/r$.

It follows from (9) that $e_G(K) \leq k^2 r/n$. So

$$\begin{aligned}
&\mathbf{Pr}\left(\exists K:\ \lambda n/r < |K| \leq \frac{n}{2(\log n)^{3/2}},\ e_{p_2}(K) > 3k\right) \\
&\leq \sum_{k=\lambda n/r}^{n/(2(\log n)^{3/2})} \binom{n}{k}\binom{k^2 r/n}{3k} p_2^{3k} \\
&\leq \sum_{k=\lambda n/r}^{n/(2(\log n)^{3/2})} \left(\frac{ne}{k}\left(\frac{kre}{3n}\cdot\frac{(1+o(1))\log n}{r}\right)^3\right)^k \\
&\leq \sum_{k=\lambda n/r}^{n/(2(\log n)^{3/2})} \left((1+o(1))\frac{e^4}{27}\cdot\frac{k^2(\log n)^3}{n^2}\right)^k \\
&= o(1).
\end{aligned}$$

Applying (5) we see that we have now verified (8).

It follows that **whp**

$$K \subseteq \text{LARGE},\ |K| \leq \frac{n}{(\log n)^{5/2}} \text{ implies } |N_{G_{m_2}}(K)| \geq \frac{\log n}{40}|K|. \tag{10}$$

Indeed, suppose there exists K for which (10) does not hold and let $L = N_{G_{m_2}}(K)$. Then $|K \cup L| \leq (\frac{\log n}{40} + 1)|K| \leq \frac{n}{39(\log n)^{3/2}}$ and it contains at least $\frac{\log n}{10}|K| - e_{m_2}(K)$ edges, contradicting (8).

To finish the verification of P4, consider the event $\mathcal{B}$ that G_{m_1} contains a set of vertices K, $\frac{n}{(\log n)^{5/2}} \leq |K| = k \leq n\frac{\log\log n}{\log n}$ where $|N_{G_{m_1}}(K)| \leq \frac{\log n}{10 \log\log n}|K|$. P4 follows from the non-occurrence of this event. Let $L = N_{G_{m_1}}(K)$, $\ell = |L|$ and $R = [n] \setminus K \cup L$. Now (7) and our assumption on the value of λ imply that

$$e_G(K,R) = (1-o(1))k(n-k-l)\frac{r}{n} \geq (1-o(1))9kr/10.$$

We calculate in G_{p_1} and translate to G_{m_1} via (7). First consider the case $\ell \leq k$. The probability that such K, L exist with no $K - R$ edges in G_{p_1} is then at most

$$\sum_{k=n/(\log n)^{5/2}}^{n\log\log n/(\log n)} \binom{n}{k}^2 (1-p_1)^{(1-o(1))k(n-2k)r/n}$$
$$\leq \sum_{k=n/(\log n)^{5/2}}^{n\log\log n/(\log n)} \left(\frac{ne}{k}\right)^{2k} \exp\{-(1-o(1))k\log n\} = o(1).$$

We can now use monotonicity and (5) to rule out this event in G_{m_1}.

Now consider the case $\ell > k$. Now (7) implies that $e_G(K,L) = (1+o(1))k\ell\frac{r}{n}$. Hence there is a set L_1 of at least $\ell/2$ vertices of L, each having at most $2(1+o(1))k\frac{r}{n}$ K-neighbours in G. Each vertex of L has a G_{p_1} neighbour in K. So the probability that such K, L_1 exist in G_{p_1} is at most

$$\sum_{k=n/(\log n)^{5/2}}^{n\log\log n/(\log n)} \binom{n}{k} \sum_{\ell=k}^{k\log n/(10\log\log n)} \binom{n}{\ell} 2^\ell \left((1+o(1))\frac{2kr}{n} \cdot p_1\right)^{\ell/2} (1-p_1)^{(1-o(1))9kr/10}$$
$$\leq \sum_{k=n/(\log n)^{5/2}}^{n\log\log n/(\log n)} \left(\frac{ne}{k}\right)^k n^{-8k/9} \sum_{\ell=k}^{k\log n/(10\log\log n)} \left(\frac{2ne}{\ell} \cdot \sqrt{(1+o(1))\frac{2k\log n}{n}}\right)^\ell$$
$$\leq \sum_{k=n/(\log n)^{5/2}}^{n\log\log n/(\log n)} \left(\frac{ne}{k}\right)^k n^{-8k/9} \cdot n \cdot \left((1+o(1))\frac{20en\log\log n}{k\log n}\sqrt{\frac{2k\log n}{n}}\right)^{k\log n/(10\log\log n)}$$
$$\leq \sum_{k=n/(\log n)^{5/2}}^{n\log\log n/(\log n)} (\log n)^{5k/2} \cdot n^{-8k/9} \cdot n \cdot n^{k/7}$$
$$= o(n^{-2}).$$

So applying (5) we see that **whp** $\mathcal{B}$ does not happen in G_{m_1} and this completes the verification of P4.

P5:
It follows from (7) that $e_G(A,B) \geq (1-o(1))ab\frac{r}{n}$ where $a=|A|, b=|B|$. Then by Chernoff bounds,

$$\mathbf{Pr}(\exists A,B:\ e_{p_1}(A,B) < ab\frac{\log n}{2n}) \leq \sum_{a,b \geq 200n\frac{\log\log n}{\log n}} \binom{n}{a}\binom{n}{b}\exp\{-\frac{1}{10}ab\frac{\log n}{n}\}$$

$$\leq \sum_{a,b \geq 200n\frac{\log\log n}{\log n}} \left(\frac{ne}{a}\exp\left\{-\frac{b\log n}{20n}\right\}\right)^a \left(\frac{ne}{b}\exp\left\{-\frac{a\log n}{20n}\right\}\right)^b$$

$$= o(1).$$

Since we are discussing a monotone property, we deduce its occurence in G_{m_1} **whp** and hence in G_{τ_2}.

P6:
We can assume w.l.o.g. that $|B| = 4|A|$. If $|A| \leq \frac{n}{(\log n)^2}$ then we can use (8) because if P6 fails then $A \cup N(A)$ contains $5|A| \leq \frac{n}{(\log n)^{3/2}}$ vertices and at least $2400|A|\log\log n$ edges. So we can assume that $|A| > \frac{n}{(\log n)^2}$. It follows from (7) that $e_G(A,B) \leq (1+o(1))ab\frac{r}{n}$. Then by Chernoff bounds,

$$\begin{aligned}
&\mathbf{Pr}(\exists A,B: e_{p_2}(A,B) \geq 2400|A|\log\log n)\\
\leq\ & \sum_{a \leq 200n\frac{\log\log n}{\log n}} \binom{n}{a}\binom{n}{4a}\binom{\frac{(1+o(1))4a^2r}{n}}{2400a\log\log n} p_2^{2400a\log\log n}\\
\leq\ & \sum_{a \leq 200n\frac{\log\log n}{\log n}} \binom{n}{a}\binom{n}{4a}\left(\frac{(1+o(1))4ea^2rp_2}{2400na\log\log n}\right)^{2400a\log\log n}\\
\leq\ & \sum_{a \leq 200n\frac{\log\log n}{\log n}} \left(\frac{ne}{a}\cdot\left(\frac{ne}{4a}\right)^4\cdot\left(\frac{(1+o(1))4ea\log n}{2400n\log\log n}\right)^{2400\log\log n}\right)^a\\
\leq\ & \sum_{a \leq 200n\frac{\log\log n}{\log n}} \left(\frac{ne}{a}\cdot\left(\frac{ne}{4a}\right)^4\cdot\left(\frac{e(1+o(1))}{3}\right)^{2400\log\log n}\right)^a\\
\leq\ & n\cdot\left(e(\log n)^{10}\left(\frac{e(1+o(1))}{3}\right)^{2400\log\log n}\right)^{\frac{n}{\log^2 n}}\\
=\ & o(1)\quad .
\end{aligned}$$

Since we are discussing a monotone property, we deduce its occurrence in G_{m_2} **whp** and hence in G_{τ_2}.

P7:
For $|A| = a \leq \frac{n}{(\log n)^{3/2}}$ we can appeal to (8). For $a > \frac{n}{(\log n)^{3/2}}$ we see from (7) that

$e_G(A) \le (1+o(1))\frac{a^2 r}{2n}$. So

$$\mathbf{Pr}\left(\exists A: \frac{n}{(\log n)^{3/2}} < a \le 30n\frac{\log\log n}{\log n}, e_{p_2}(A) \ge 100a\log\log n\right)$$

$$\le \sum_{a=\frac{n}{(\log n)^{3/2}}}^{30n\frac{\log\log n}{\log n}} \binom{n}{a}\binom{(1+o(1))a^2r/n}{100a\log\log n} p_2^{100a\log\log n}$$

$$\le \sum_{a=\frac{n}{(\log n)^{3/2}}}^{30n\frac{\log\log n}{\log n}} \left(\frac{ne}{a}\left(\frac{ae(1+o(1))\log n}{100n\log\log n}\right)^{100\log\log n}\right)^a$$

$$\le \sum_{a=\frac{n}{(\log n)^{3/2}}}^{30n\frac{\log\log n}{\log n}} (e(\log n)^{3/2}(.8)^{100\log\log n})^a$$

$$= o(1).$$

Since we are discussing a monotone property, we deduce its occurrence in G_{m_2} **whp** and hence in G_{τ_2}.

5 Concluding remarks

- Our main result, Theorem 1, and its proof imply readily the following result on the threshold probability for Hamilton cycles in the random subgraph G_p:

  Corollary 1 *Let G be an r-regular subgraph on n vertices all of whose eigenvalues, but the first one, are at most λ in their absolute values. Assume that $\lambda = o\left(\frac{r^{5/2}}{n^{3/2}(\log n)^{3/2}}\right)$. Form a random subgraph G_p of G by choosing each edge of G independently with probability p. Then for any function $\omega(n)$ tending to infinity arbitrarily slowly:*

  1. *if $p(n) = \frac{1}{r}(\log n + \log\log n - \omega(n))$, then* **whp** *$G_p$ is not Hamiltonian;*
  2. *if $p(n) = \frac{1}{r}(\log n + \log\log n + \omega(n))$, then* **whp** *$G_p$ is Hamiltonian.*

- Our result enables to estimate from below the number of Hamilton cycles in pseudo-random graphs as follows:

  Corollary 2 *Let G satisfy the conditions of Theorem 1. Then G contains at least $\left(\frac{r}{(1+o(1))\log n}\right)^n$ Hamilton cycles.*

Proof Denote by $HC(G)$ the number of Hamilton cycles in G. Consider the random subgraph G_p with $p = p(n) = (\log n + 2\log\log n)/r$. Denote by X the random variable counting the number of Hamilton cycles in $G(p)$. As by our main Theorem and the preceding corollary G_p has **whp** a Hamilton cycle, we get $\mathbf{E}(X) \geq 1 - o(1)$. On the other hand, the probability a given Hamilton cycle of G appears in G_p is exactly p^n. Therefore the linearity of expectation implies $\mathbf{E}(X) = HC(G) \cdot p^n$. Combining the above two estimates we get:

$$HC(G) \geq \frac{1 - o(1))}{p^n} = \left(\frac{r}{(1 + o(1))\log n}\right)^n .$$

□

Note that the number of Hamilton cycles in any r-regular graph on n vertices obviously does not exceed r^n. Thus for graphs satisfying the conditions of Theorem 1 the above corollary provides an asymptotically tight estimate on the exponent of the number of Hamilton cycles. The above result improves upon an estimate of Thomason ([15], Corollary 2.9) for the number of Hamilton cycles in pseudo-random graphs. We remark that for the case of pseudo-random graphs of linear degrees a recent result of the first author [10] gives an even better lower bound for the number of Hamilton cycles.

- We do not believe that the restriction on λ imposed in the formulation of Theorem 1 is optimal. It would be interesting to figure out what is the weakest possible requirement on the spectral gap which still guarantees that the hitting time for Hamiltonicity still coincides **whp** with that of having all degrees at least two. We conjecture that at least for the case of linear degrees the weakest possible condition $\lambda = o(r)$ should ensure the above stated property.

- Our paper can be viewed as the first step in studying random subgraphs of pseudo-random graphs. Questions of a similar kind can be asked with respect to other properties of pseudo-random graphs, like independence and chromatic numbers, existence of perfect matchings, factors and many others. Their study should combine existing techniques for the binomial random graphs $G(n, p)$ with known results on the edge distribution of pseudo-random graphs. We plan to return to this subject in the future.

References

[1] M. Ajtai, J. Komlós and E. Szemerédi. *The first occurrence of Hamilton cycles in random graphs.*, Annals of Discrete Mathematics 27 (1985), 173–178.

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

[3] N. Alon, M. Krivelevich and B. Sudakov, List coloring of random and pseudo-random graphs, *Combinatorica* **19** (1999), 453–472.

[4] N. Alon and J. Spencer, **The probabilistic method**, Wiley, New York 1992.

[5] B. Bollobás, *The evolution of sparse graphs*, In: Graph Theory and Combinatorics, Proc. Cambridge Combin. Conf. in honour of Paul Erdős (B. Bollobás, ed.), Academic Press, 1984, pp. 35–57.

[6] B.Bollobás, T.I.Fenner and A.M.Frieze, *An algorithm for finding Hamilton paths and cycles in random graphs*, Combinatorica 7 (1987), 327–341.

[7] F. R. K. Chung, **Spectral graph theory**, Regional Conference Series in Mathematics 92, Amer. Math. Soc., Providence, 1997.

[8] F. R. K. Chung, R. L. Graham and R. M. Wilson, *Quasi-random graphs*, Combinatorica 9 (1989), 345–362.

[9] J. Komlós and E. Szemerédi, *Limit distributions for the existence of Hamilton circuits in a random graph*, Discrete Mathematics 43 (1983), 55–63.

[10] A. M. Frieze, *On the number of perfect matchings and Hamilton cycles in ϵ-regular non-bipartite graphs*, Electronic J. Combinatorics Vol. 7 (2000), publ. R57.

[11] S. Janson, T. Łuczak and A. Ruciński, **Random graphs**, Wiley, New York, 2000.

[12] A. Nilli, *On the second eigenvalue of a graph*, Discrete Math. 91 (1991), 207–210.

[13] L. Posa, *Hamiltonian circuits in random graphs*, Discrete Math. 14 (1976), 359–364.

[14] M. Simonovits and V. T. Sós, *Hereditary extended properties, quasi-random graphs and not necessarily induced subgraphs*, Combinatorica 17 (1997), 577–596.

[15] A. Thomason, *Pseudorandom graphs*, In: Random graphs'85 (Poznań 85), North-Holland Math. Stud. 144, North Holland, Amsterdam, 1987, pp. 307–331.

[16] A. Thomason, *Random graphs, strongly regular graphs and pseudorandom graphs*, In: Surveys in Combinatorics 1987, London Math. Soc. Lecture Note Ser. 123, Cambridge Univ. Press, Cambridge, 1987, pp. 173–195,