A scaling limit for the length of the longest cycle in a sparse random graph

Michael Anastos*and Alan Frieze†

January 8, 2021

Abstract

We discuss the length $L_{c,n}$ of the longest cycle in a sparse random graph $G_{n,p}$, $p = c/n$, c constant. We show that for large c there exists a function $f(c)$ such that $L_{c,n}/n \to f(c)$ a.s. The function $f(c) = 1 - \sum_{k=1}^{\infty} p_k(c)e^{-kc}$ where $p_k(c)$ is a polynomial in c. We are only able to explicitly give the values p_1, p_2, although we could in principle compute any p_k. We see immediately that the length of the longest path is also asymptotic to $f(c)n$ w.h.p.

1 Introduction

There are several basic questions that can be asked in the context of a class of graphs. E.g. what is the chromatic number? Is the graph Hamiltonian? Another such basic question is the following: how long is the longest cycle? In this paper we study this question in relation to the sparse random graph $G_{n,p}, p = c/n$ for a constant $c > 0$. Thus, let $L_{c,n}$ denote the length of the longest cycle in the random graph $G_{n,c/n}$. Erdős [10] conjectured that if $c > 1$ then w.h.p. $L_{c,n} \geq \ell(c)n$ where $\ell(c) > 0$ is independent of n. This was proved by Ajtai, Komlós and Szemerédi [1] and in a slightly weaker form by de la Vega [26] who proved that if $c > 4 \log 2$ then $f(c) = 1 - O(c^{-1})$. See also Suen [25]. Although this answered Erdős's question it only gives us a lower bound for the length of the longest cycle. Bollobás [4] realised that for large c one could find a large path/cycle w.h.p. by concentrating on a large subgraph with large minimum degree and demonstrating Hamiltonicity. In this way he showed that $\ell(c) \geq 1 - c^{24}e^{-c/2}$. This was then improved by Bollobás, Fenner and Frieze [7] to $\ell(c) \geq 1 - c^6 e^{-c}$ and then by Frieze [15] to $\ell(c) \geq 1 - (1 + \varepsilon_c)(1 + c)e^{-c}$ where $\varepsilon_c \to 0$ as

*Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany, email:**manastos@zedat.fu-berlin.de**

†Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA, U.S.A. email:**alan@random.math.cmu.edu**; the author is supported in part by NSF Grant DMS1363136

$c \to \infty$. This last result is optimal up to the value of ε_c, as there are w.h.p. $(1+c)e^{-c}n+o(n)$ vertices of degree 0 or 1.

The basic open question to this point, is at to whether or not there exists a function $f(c)$ such that w.h.p. the $L_{c,n} = (1+\varepsilon_n)f(c)n$ where $\varepsilon_n \to 0$ as $n \to 0$. And what is $f(c)$. In this paper we establish the existence of $f(c)$ for large c and give a method of computing it to arbitrary accuracy. We note that this is one case of a fundamental extremal random variable where the existence of a scaling limit has not previously been shown to exist and does not appear to be susceptible to the interpolation method as in Bayati, Gamarnik and Tetali [3].

Let $p = c/n$ and let $G = G_{n,p}$. We will assume throughout that c is sufficiently large. To approximate the length of the longest path we construct a cycle C and then argue that w.h.p. its length is equal to $L_{c,n} - O(\log n)$. It is well known, see for example Chapter 2 of [18] that w.h.p. G consists of a unique linear size *giant* component C_1 plus a collection of smaller components of size bounded by $O(\log n)$. So to look for a long cycle, we must look inside C_1. Now, no vertex of degree one or less can be in a cycle and so we remove such vertices from consideration. This may create more vertices of degree one and so we continue until we have a subgraph with minimum degree at least two. This will be C_2, it is the 2-core of the giant component C_1 and consists of all the vertices in C_1 that are in at least one cycle.

C_2 has minimum degree at least two, but it is unlikely to be Hamiltonian. One reason is because there are a large number of triples of degree two vertices that share a common neighbor. Given this, we first identify $C_{3,ext}$, a large subgraph of C_2 of minimum degree 3. $C_{3,ext}$ can be proven to be Hamiltonian, a fact that we use as a starting point. To construct an even longer cycle we consider how paths in $C_2 \setminus C_{3,ext}$ can be inserted into a Hamilton cycle in $C_{3,ext}$. Indeed, in Section 3, we show that given a fixed set of vertex disjoint paths whose endpoints are adjacent to $C_{3,ext}$ and cover a set of vertices V_{paths} we can find a cycle that spans $V(C_{3,ext}) \cup V_{paths}$. By considering a suitable set of paths such that V_{paths} is (almost) maximized we find a long cycle in C_2. The length of the longest path in $G_{n,c/n}$ differs from the length of this cycle $O(\log n)$ w.h.p. The reason for the latter statement is that $L_{c,n} - (|V_{paths}| + |C_{3,ext}|)$ will be bounded by the size of the first and last component in $G_{n,p} \setminus C_{3,ext}$ that a longest path traverses plus the number of vertices found in the non-tree components of $C_2 \setminus C_{3,ext}$. The latter two quantities, as seen by Lemmas 2.6 and 2.7 sum up to $O(\log n)$ w.h.p.

Notation 1.1. *Let $C_{3,ext}$ be the maximal subgraph of C_2 such that (i) every vertex in $C_{3,ext}$ has at least 3 neighbors in $C_{3,ext}$ and (ii) every vertex in $C_2 \setminus C_{3,ext}$ that is adjacent to a vertex in $C_{3,ext}$ has at least 3 neighbors in $C_{3,ext}$. Note that if S_1, S_2 are two sets satisfying (i) and (ii) then $S_1 \cup S_2$ also satisfies (i), (ii) and so $C_{3,ext}$ is well-defined.*

We let Γ be the induced subgraph of C_2 spanned $V(C_2) \setminus V(C_{3,ext})$.

In Section 2, we study the structure of Γ by considering a peeling process that constructs $C_{3,ext}$ as in the papers [4], [7] and [15].

Notation 1.2. *Let $\mathcal{T}$ denote the set of trees in Γ. For a tree $T \in \mathcal{T}$ let $\mathcal{P}_T$ be the set of path packings of T where we allow only paths whose start- and end- vertex have neighbors in $C_{3,ext}$.*

Here by a path packing we mean a set of vertex disjoint paths in which we also allow paths of length 0. So a single vertex with neighbors in $C_{3,ext}$ counts as a path. For $P \in \mathcal{P}_T$ let $n(T, P)$ be the number of vertices in T that are not covered by P. Let $\phi(T) = \min_{P \in \mathcal{P}_T} n(T, P)$ and $\mathcal{Q}(T) \in \mathcal{P}_T$ denote a set of paths that leaves $\phi(T)$ vertices of T uncovered i.e. satisfies $n(T, \mathcal{Q}(T)) = \phi(T)$. Finally we let $\mathcal{Q}(\mathcal{T}) = \cup_{T \in \mathcal{T}} \mathcal{Q}(T)$.

Observe that any cycle in C_2 fails to span at least $\sum_{T \in \mathcal{T}} \phi(T)$ vertices in the tree components of Γ. Hence it spans at most $|V(C_2)| - \sum_{T \in \mathcal{T}} \phi(T)$ vertices in C_2. By finding a cycle in C_2 that spans exactly this many vertices we prove,

Theorem 1.3. *Let $p = c/n$ where $c > 1$ is a sufficiently large constant. Then w.h.p.*

$$-1 \leq L_{c,n} - \left[|V(C_2)| - \sum_{T \in \mathcal{T}} \phi(T) \right] \leq 3 \log n. \tag{1}$$

Notation 1.4. *If $A = A(n), B = B(n)$ then we write $A \approx B$ if $A = (1 + o(1))B$ as $n \to \infty$.*

The size of C_2 is well-known. Let x be the unique solution of $xe^{-x} = ce^{-c}$ in $(0, 1)$. Then w.h.p. (see e.g. [18], Lemma 2.16),

$$|C_2| \approx (1 - x) \left(1 - \frac{x}{c}\right) n.$$
$$|E(C_2)| \approx \left(1 - \frac{x}{c}\right)^2 \frac{c}{2} n.$$

Equation (4.5) of Erdős and Rényi [11] tells us that

$$x = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (ce^{-c})^k = ce^{-c} + c^2 e^{-2c} + 3c^3 e^{-3c}/2 + O(c^4 e^{-4c}).$$

Hence,

$$|C_2| = (1 - (c + 1)e^{-c} - c^2 e^{-2c} - c^2 (c + 1) e^{-3c}/2 + O(c^4 e^{-4c}))n. \tag{2}$$

We will argue in Section 4 that w.h.p., as c grows, that

$$\sum_{T \in \mathcal{T}} \phi(T) = \frac{c^6 e^{-3c}}{36} + O(c^6 e^{-4c})n. \tag{3}$$

We therefore have the following improvement to the estimate in [15].

Corollary 1.5. *W.h.p., as c grows, we have that*

$$L_{c,n} = \left(1 - (c + 1)e^{-c} - c^2 e^{-2c} - c^2 (c + 1) e^{-3c}/2 - c^6 e^{-3c}/36 + O(c^6 e^{-4c})\right) n. \tag{4}$$

Note the term $(c+1)e^{-c}$ which accounts for vertices of degree 0 or 1. In principle we can compute more terms than what is given in (4). We claim next that there exists some function $f(c)$ such that the sum in (1) is concentrated around $f(c)n$ w.h.p.

Theorem 1.6. *Let $p = c/n$ where $c > 1$ is a sufficiently large constant.*

(a) There exists a function $f(c)$ such that for any $\epsilon > 0$, there exists n_ε such that for $n \geq n_\varepsilon$,

$$\left| \frac{\mathbf{E}[L_{c,n}]}{n} - f(c) \right| \leq \epsilon. \tag{5}$$

(b)

$$\frac{L_{c,n}}{n} \to f(c) \ a.s. \tag{6}$$

Beginning with Theorem 1.3 we will prove Theorem 1.6 in Section 5. The proof of Theorem 1.3 is given in Section 3. In Section 2 we study the components of Γ.

2 Structure of Γ

To construct $C_{3,ext}$ we consider a peeling process that sequentially removes vertices from C_2 as described below. We let $S_0 = \emptyset, S_1, S_2, \ldots, S_L \subseteq C_2$ be the sequence of vertex sets that have been removed by the steps/iterations of the process. Thus L is the number of iterations of the process and $C_{3,ext}$ is shown in Lemma 2.1 to be the graph spanned by $V(C_2) \setminus S_L$.

Algorithm Γ-Construction
Let $S_0 = \emptyset$. Suppose now that we have constructed S_ℓ, $\ell \geq 0$. We construct $S_{\ell+1}$ from S_ℓ via one of two cases:

Case a: If there is $v \in S_\ell$ that has exactly one or two neighbors W in $C_2 \setminus S_\ell$, then we add W to S_ℓ to make $S_{\ell+1}$.
Case b: If there is a vertex $v \in C_2 \setminus S_\ell$ that has at most two neighbors in $C_2 \setminus S_\ell$ then we define $S_{\ell+1}$ to be S_ℓ plus v plus the neighbors of v in $C_2 \setminus S_\ell$.

If none of the two above cases apply we let the current vertex set be S_L and we terminate the algorithm.

Lemma 2.1. *Let S_L be the set of vertices output by the above algorithm. Then, $C_{3,ext}$ and Γ are the graphs spanned by $V(C_2) \setminus S_L$ and S_L respectively.*

Proof. First observe that since the algorithm terminates after L steps we see that there does not exist $v \in V(C_2) \setminus S_L$ such that either (i) v has fewer than 3 neighbors in $V(C_2) \setminus S_L$ or (ii) v is adjacent to a vertex $V(C_2)$ that has fewer than 3 neighbors in $V(C_2) \setminus S_L$. Since $V(C_{3,ext})$ spans the maximal such subgraph we have that $V(C_2) \setminus S_L \subseteq V(C_{3,ext})$.

Now assume that $C_2 \setminus S_L \neq V(C_{3,ext})$ and let w be the first vertex in $V(C_{3,ext})$ that was removed from $C_{3,ext}$ and let i be the corresponding iteration i.e. $w \notin S_i$ but $w \in S_{i+1}$. Then either (i) w invoked Case b or (ii) a neighbor of w invoked Case a of the above algorithm. For (i) we have $C_{3,ext} \subset C_2 \setminus S_i$ implies $N(w) \cap C_{3,ext} \subset N(w) \cap (C_2 \setminus S_i)$. Hence w has at least 3 neighbors in $C_2 \setminus S_i$ and at step i it did not invoke Case b. For (ii) let $u \in N(w) \cap S_i$. Then $N(u) \cap C_{3,ext} \subset N(u) \cap (C_2 \setminus S_i)$ and so u has at least 3 neighbors in $C_2 \setminus S_i$ and so u did not invoke Case a. Hence we have a contradiction and $V(C_{3,ext}) = V(C_2 \setminus S_L)$ and $V(\Gamma) = S_L$. □

Lemma 2.2. *S_L does not depend on the order of adding vertices.*

Proof. The proof of Lemma 2.1 can be adapted to prove this. We assume there are two possibilities S, S' for S_L and let w be the first vertex of S' not in S. The argument of Lemma 2.1 can then be repeated. □

In Lemma 2.4 we bound the size of $V(\Gamma) = S_L$. For its proof we need the following lemma on the density of small sets.

Lemma 2.3. *W.h.p., every set $S \subseteq [n]$ of size at most $n_0 = n/10c^3$ contains less than $3|S|/2$ edges in $G_{n,p}$.*

Proof. The expected number of sets invalidating the claim can be bounded by

$$\sum_{s=4}^{n_0} \binom{n}{s} \binom{\binom{s}{2}}{3s/2} \left(\frac{c}{n}\right)^{3s/2} \leq \sum_{s=4}^{n_0} \left(\frac{ne}{s} \cdot \left(\frac{se}{3}\right)^{3/2} \cdot \left(\frac{c}{n}\right)^{3/2}\right)^s = \sum_{s=4}^{n_0} \left(\frac{e^{5/2}c^{3/2}s^{1/2}}{3^{3/2}n^{1/2}}\right)^s = o(1).$$

□

Lemma 2.4. *Let $p = c/n$ where $c > 1$ is a sufficiently large constant. Then w.h.p.*

$$|V(\Gamma)| \leq ne^{-c/2}. \tag{7}$$

Proof. Consider the construction of S_L. Let A be the set of the vertices in C_2 with degree less than $D = 100$ and let $S_0' = (A \cup N(A)) \cap S_L \subseteq S_L$. If we start with $S_0 = S_0'$ and run the process for constructing Γ then we will produce the same S_L as if we had started with $S_0 = \emptyset$, see Lemma 2.2. Now w.h.p. there are at most $n_D = \frac{2c^De^{-c}}{D!}n$ vertices of degree at most D in $G_{n,p}$, (see for example Theorem 3.3 of [18]) and so $|S_0'| \leq Dn_D$.

Now suppose that the process runs for another k rounds and let v_i be the vertex that invokes either Case a or Case b at the ith iteration of the Construction of Γ. Then $v_1, v_2, \ldots, v_k$ are all distinct, none of them belongs to A and the sets $N(v_1), N(v_2), \ldots, N(v_k)$ belong to S_L. Because $v_i \notin A$ we have $|N(v_i)| \geq D$ for $i \in [k]$. In addition at the ith iteration at most three *new* vertices are added to S_i. Thus S_k has a least $(\sum_{i\in[k]} |N(v_i)|)/2 \geq kD/2$ edges and at most $|S_0'| + 3k \leq Dn_D + 3k$ vertices.

If k reaches $4n_D$ then,
$$\frac{e(S_k)}{|S_k|} \geq \frac{4Dn_D}{2} \cdot \frac{1}{(D+12)n_D} > \frac{3}{2}.$$
As $Dn_D + 3 \times 4n_D \leq n/10c^3$, from Lemma 2.3, we can assert that w.h.p. the process runs for less than $4n_D$ rounds and,
$$|V(\Gamma)| \leq (D+12)n_D \leq ne^{-c/2}.$$
$\square$

We note the following properties of $S_L = V(\Gamma)$. Let
$$V_1 = V(C_2) \setminus S_L \text{ and } V_2 = \{v \in S_L : \ v \text{ has at least one neighbor in } V_1\}.$$
Then,

G1 Each vertex $v \in S_L \setminus V_2$ has no neighbors in V_1.

G2 Each $v \in V_1 \cup V_2$ has at least 3 neighbors in V_1.

Given the definition of V_2, for a component K of Γ we define $v_0(K)$ as
$$v_0(K) = V(K) \setminus V_2.$$
Hence $v_0(K)$ consists of the vertices in $V(K)$ with no neighbors in V_1. We prove the following lemma.

Lemma 2.5. *W.h.p. each component K of Γ satisfies*
$$|v_0(K)| \geq \frac{|V(K)|}{3}. \tag{8}$$

Proof. We will prove that for $0 \leq i \leq L$ and each component K spanned by S_i,
$$|v_{0,i}(K)| \geq \frac{|V(K)|}{3}. \tag{9}$$
Here $v_{0,i}(K)$ is taken to be the number of vertices in $V(K)$ with no neigbors in $C_2 \setminus K$. Taking $i = L$ in (9) yields (8). We proceed by an induction on i.

$S_0 = \emptyset$ and so for $i = 0$, (9) is satisfied by every component spanned by S_0. Suppose that at step $i = \ell$, (9) is satisfied by every component spanned by S_ℓ.

At step $\ell+1$, assume that v invokes either Case a or Case b. In both cases $S_{\ell+1} = S_\ell \cup \big(\{v\} \cup N(v)\big)$. The addition of the new vertices into S_ℓ could merge components $K_1, K_2, \ldots, K_r$ into one component K' while adding at most 3 vertices. Hence $3 + \sum_{j\in[r]} |K_i| \geq |K'|$. In addition every vertex that contributed to $\upsilon_{0,\ell}(K_j)$, $j = 1, 2, \ldots, r$ now contributes towards $\upsilon_{0,\ell+1}(K')$. Also v has neighbors outside S_ℓ but no neighbors outside $S_{\ell+1}$. The inductive hypothesis implies that $\upsilon_{0,\ell}(K_j) \geq |K_j|/3$ for $j \in [r]$. Thus,

$$\upsilon_{0,\ell+1}(K') \geq 1 + \sum_{j\in[r]} \upsilon_{0,\ell}(K_j) \geq 1 + \frac{1}{3}\sum_{j\in[r]} |K_j| \geq 1 + \frac{|K'| - 3}{3} = \frac{|K'|}{3}.$$

And so (9) continues to hold for all the components spanned by $S_{\ell+1}$. □

We show next that w.h.p., only a small component K can satisfy (8).

Lemma 2.6. *Let $p = c/n$ where $c > 1$ is a sufficiently large constant. Then w.h.p. the tree components of $G_{n,p} \setminus C_{3,ext}$, hence of Γ, are bounded in size by $\log n$.*

Proof. Let K be a tree component of Γ and K' the component of $G_{n,p} \setminus C_{3,ext}$ that contains $K \subset C_2$. Then $K' \setminus K \subset G_{n,p} \setminus C_2$ and $K \subset C_2$ imply that $K' \setminus K$ consists of trees (or small unicyclic components) that are connected to C_2 via a single vertex that belongs to K and hence these trees are not adjacent to $V(C_{3,ext})$. Thus (9) implies that K' contains at least $|K|/3 + |K' \setminus K| \geq |K'|/3$ vertices that are not adjacent to $V(G_{n,p}) \setminus K$.

Thus the probability a tree component of $G_{n,p} \setminus C_{3,ext}$, hence of Γ, contains more than $\log n$ vertices is bounded by

$$\begin{aligned}\sum_{k\geq \log n} \binom{n}{k} k^{k-2}\left(\frac{c}{n}\right)^{k-1}\binom{k}{k/3}\left(1-\frac{c}{n}\right)^{k(n-k)/3} &\leq \sum_{k\geq \log n}\left(\frac{ne}{k}\right)^k k^{k-2}\left(\frac{c}{n}\right)^{k-1} 2^k e^{-ck/6} \quad (10)\\ &\leq \sum_{k\geq\log n}\frac{n}{ck^2}\left(2ce^{1-c/6}\right)^k = o(1).\end{aligned}$$

Explanation for (10): We first choose K' in $\binom{n}{k}$ ways, then choose a spanning tree of K' in k^{k-2} ways and then choose a subset K_1 of size $k/3$ in $\binom{k}{k/3}$ ways. K_1 consists of the vertices in $V(K')$ with no neighbor outside $V(K')$. □

So, we can assume that all tree components are of size at most $\log n$.

Lemma 2.7. *Let $p = c/n$ where $c > 1$ is a sufficiently large constant. Then w.h.p. the non-tree components in either $G_{n,p} \setminus C_{3,ext}$ or Γ, span at most $\log n$ vertices.*

Proof. Every non-tree of component of $V(G_{n,p}) \setminus C_{3,ext}$ contains a cycle. It is either disjoint from the giant component C_1 or it intersects C_2 and contains a non-tree component of Γ. Thus we can bound both quantities in question by the expected number of vertices of $V(G_{n,p}) \setminus C_{3,ext}$ on components that are not trees. Similarly to Lemma 2.6 we have that the latest is bounded by

$$\sum_{k\geq 3} k\binom{n}{k}k^{k-2}\binom{k}{2}\left(\frac{c}{n}\right)^k\binom{k}{k/3}\left(1-\frac{c}{n}\right)^{k(N_2-k)/3} \leq \sum_{k\geq 3} k\left(2ce^{1-c/6}\right)^k = O(1). \tag{11}$$

The $k^{k-2}\binom{k}{2}$ in the above expression bounds the number of spanning unicyclic graphs on k vertices that can be decomposed into a spanning tree and an edge.

Markov's inequality implies that w.h.p. such components span at most $\log n$ vertices. □

3 Proof of Theorem 1.3

Notation 3.1. *For $T \in \mathcal{T}$, let $\mathbb{M}_T$ be the matching on V_2 obtained by replacing each path of $\mathcal{Q}(T)$ of length at least 1 by an edge joining its endpoints. The internal vertices of such paths are removed. We let $\mathbb{M}^* = \bigcup_{T\in\mathcal{T}} \mathbb{M}_T$. Let $I(T)$ denote the internal vertices of the paths $\mathcal{Q}(\mathcal{T})$ and $I^* = \bigcup_{T\in\mathcal{T}} I(T)$ and $V_2^* = V_2 \setminus I^*$. We let Γ_1^* be the subgraph of G induced by V_1. We also let Γ_2^* be the bipartite graph with vertex partition V_1, V_2^* and all edges $\{e \in E(G) : e \in V_1 \times V_2^*\}$. Finally let $\Gamma^* = \Gamma_1^* \cup \Gamma_2^* \cup \mathbb{M}^*$ and $V^* = V_1 \cup V_2^* = V(\Gamma^*)$.*

Theorem 3.2. *W.h.p. there is a Hamilton cycle H^* in Γ^* that contains all the edges of $\mathbb{M}^*$*

This section is devoted to the proof of Theorem 3.2. We begin by giving an outline of the proof and then we show how Theorem 1.3 follows. Following this, we prove Theorem 3.2.

Outline of proof To prove Theorem 3.2 we begin by partitioning Γ^* into 2 subgraphs, the blue and the green subgraphs denoted by Γ_b^* and Γ_g^* respectively. The blue graph will have "nice" expansion properties while the green graph will be distributed uniformly among a set of graphs $\mathcal{G}$. Then, in Section 3.6 we use a modification of a double counting argument that was first used in [13] to bound the number of graphs $G \in \mathcal{G}$ such that $G_b^* \cup G$ is not Hamiltonian. The specific version is from [14]. Given the decomposition of Γ^* into Γ_b^* and Γ_g^* if Γ^* is not Hamiltonian then one may further decompose the edges of the green graph Γ_g^* into two subgraphs, the yellow and red subgraphs denoted by Γ_y^* and Γ_r^* respectively, such that (i) the yellow edges form a set of paths and (ii) a longest path in Γ^* is spanned by the blue and yellow edges. Then we argue, using Pósa rotations, that there is a large set of edges E' none of which belongs to $E(\Gamma_b^*) \cup E(\Gamma_y^*)$ such for every $e \in E'$ the subgraph spanned $\{e\} \cup E(\Gamma_b^*) \cup E(\Gamma_y^*)$ either spans a path longer than the one spanned by $\Gamma_b^* \cup \Gamma_y^*$ hence by Γ^* or it is Hamiltonian. Pósa rotations (introduced in Section 3.5), define a procedure that starts with a longest path in a graph and produces many pairs of vertices that are the endpoints of longest paths. Hence, $E' \cap E(\Gamma_r^*) = \emptyset$ which will imply that for each possible set of yellow edges there are only a small number of sets of red edges such that $\Gamma_b^* \cup \Gamma_y^* \cup \Gamma_r^* = \Gamma^*$ is not Hamiltonian.

We finish this subsection by proving Theorem 1.3.

Proof of Theorem 1.3: Let H^* be the Hamilton cycle given in Theorem 3.2. Replacing the edges in $\mathbb{M}^*$ with the corresponding paths in $\mathcal{Q}(\mathcal{T})$ gives a cycle in $G_{n,p}$ of size $|V(C_2)| - \sum_{T\in\mathcal{T}} \phi(T)$. Hence, $L_{c,n} \geq |V(C_2)| - \sum_{T\in\mathcal{T}} \phi(T)$.

On the other hand let $P_{longest}$ be a longest path in $G_{n,p}$ and $P_1, P_2, \ldots, P_a$ be its subpaths that are spanned by $G_{n,p} \setminus C_{3,ext}$ in the order that they appear. Then the endpoints of $P_2, P_3, \ldots, P_{a-1}$ are adjacent to V_1 and therefore $P_2, P_3, \ldots, P_{a-1}$ do not cover at least $\sum_{T\in\mathcal{T}} \phi(T)$ vertices that are spanned by the tree components of $C_2 \setminus C_{3,ext}$ (see notation 1.2). Each of P_1, P_a may traverse vertices in a single component of $G_{n,p} \setminus C_{3,ext}$. Thus $|P_{longest}|$ is bounded by above by $|C_2| - \sum_{T\in\mathcal{T}} \phi(T)$ plus twice the size of the maximum component of $G_{n,p} \setminus C_{3,ext}$ plus the number of vertices in Γ that do not belong to a tree component of Γ. Lemmas 2.6 and 2.7 imply that the last two quantities sum to at most $3 \log n$. $\square$

3.1 Structure of Γ_1^*

Suppose now that $|V_1| = N$ and that V_1 contains M edges. The construction of Γ does not involve the edges inside V_1, but we do know that that Γ_1^* has minimum degree at least 3. The distribution of Γ_1^* will be that of $G_{V_1,M}$ subject to this degree condition, viz. the random graph $G_{V_1,M}^{\delta\geq 3}$ which is sampled uniformly from the set $\mathcal{G}_{V_1,M}^{\delta\geq 3}$, the set of graphs with vertex set V_1, M edges and minimum degree at least 3. This is because, we can replace Γ_1^* by any graph in $G_{V_1,M}^{\delta\geq 3}$ without changing Γ. By the same token, we also know that each $v \in V_2^*$ has at least 3 random neighbors in V_1. We have that

$$N \geq n(1 - 2e^{-c/2}) \text{ and } M \in \frac{(1 \pm \varepsilon_1)cN}{2}, \tag{12}$$

where $\varepsilon_1 = c^{-1/3}$. The bound on N follows from (2) and (7) and the bound on M follows from the fact that in $G_{n,p}$,

$$\mathbf{Pr}\left(\exists S : |S| = N, e(S) \notin (1 \pm \varepsilon_1)\binom{N}{2}p\right) \leq 2\binom{n}{N} \exp\left\{-\frac{\varepsilon_1^2 N(N-1)p}{3}\right\} = o(1).$$

The inequality follows from the Chernoff bound for the Binomial distribution.

3.2 Partitioning/Coloring $G = G_{n,p}$ and Γ^*

In this section we describe how to color/partition the edges of both $G = G_{n,p}$ and Γ^*. We first color most of the edges of G light blue, dark blue or green. This will induce a partial coloring of $E(\Gamma^*)$ which we then extend to a complete coloring of $E(\Gamma^*)$. We denote the resultant blue and green subgraphs in G by Γ_b, Γ_g respectively (an edge is blue if it is either dark or light blue). We later show that the blue graph has expansion properties while the green graph has suitable randomness.

Notation 3.3. *For a graph G and vertex sets $A, B \subseteq V(G)$ we write*

$$A : B = \{\{a, b\} \in E(G) : a \in A, b \in B\}.$$

Every vertex $v \in V_1$ independently chooses $\min\left\{\deg_{V_1}(v), 100\right\}$ neighbors in V_1 and we color the chosen edges light blue. Then we color every edge in $V_2^* : V_1$ light blue. Thereafter we independently color (re-color) every edge of G dark blue with probability $1/2000$. This coloring is done independently of the structure of Γ^*. Finally we color green all the uncolored edges that are contained in V_1. (Some of the edges of G will remain uncolored and play no significant role in the proof.)

The above coloring satisfies the following properties:

(C1) Every vertex in $V_1 \cup V_2^*$ is joined to at least 3 vertices in V_1 by a blue edge.

(C2) In G, every dark blue edge appears independently with probability $\frac{p}{2000}$.

(C3) Given the degree sequence $\mathbf{d}_g$ of Γ_g, every graph H with vertex set V_1 and degree sequence $\mathbf{d}_g$ is equally likely to be Γ_g.

We can justify **C3** as follows: Amending G by replacing Γ_g by any other graph Γ_g' with vertex set V_1 and the same degree sequence and executing our construction of S_L will result in the same set S_L and sets V_1, V_2^*. So, each possible Γ_g' has the same set of extensions to $G_{n,p}$ and as such is equally likely.

Now given $\Gamma_b, \Gamma_g \subset G$ we color the edges in Γ^* as follows. Every edge in Γ^* that exists in G inherits its color from the coloring in G. Every edge in $\mathbb{M}^* \subseteq E(\Gamma^*)$ is colored light blue. We let Γ_b^*, Γ_g^* be the blue and the green subgraphs of Γ^*. Observe that $\Gamma_g^* = \Gamma_g$, hence Γ_g^* satisfies property $(C3)$ as well.

3.3 Expansion of Γ_b^*

We wish to estimate the probability that small sets have relatively few neighbors in the graph Γ_b^*. For $S \subseteq V^* = V_1 \cup V_2^*$ we let

$$\begin{aligned} N_b(S) &= \{w \in V_1 \setminus S : \exists v \in S \text{ with } \{v, w\} \in E(\Gamma_b^*)\} \\ &= \{w \in V_1 \setminus S : \exists v \in S \text{ with } \{v, w\} \in E(\Gamma_b)\}. \end{aligned}$$

We have slightly abused notation here since $N_b(S)$ is implicitly defined in both G and Γ^* in the same way.

It is shown in [6] and also in [16] that if S is the set of endpoints of longest paths created by Pósa rotations (see Section 3.5) then $S \cup N(S)$ is connected and contains at least two distinct cycles hence, at least $|S| + |N(S)| + 1$ edges. Hence the condition (iii) in the following lemma.

Lemma 3.4. *W.h.p. there does not exist $S \subset V^*$ of size $|S| \le n/4$ and (i) $|N_b(S)| \le 2|S|$, (ii) $S \cup N_b(S)$ is connected in $G_{n,p}$ and (iii) $S \cup N_b(S)$ spans at least $|S| + |N_b(S)| + 1$ edges in $G_{n,p}$.*

Proof. Assume that the above fails for some set S.
Case 1: $|S| \le n_1 = n/(100c^3)$**.**
Let $t = |N_b(S)|$. We will suppose first that S contains at least $s/10$ vertices of degree at least 100. In this case $S \cup N_S$ has cardinality at most $s + t \le 3s$ and contains at least $5s > 3(s+t)/2$ edges, contradicting Lemma 2.3.

On the other hand, if there are at least $9s/10$ vertices in S of degree at most 99 then there are at least $3(s+t)/10$ vertices of degree at most 99 in a connected subgraph of size $s_0 \le s+t \le 3n_1$. In addition that subgraph spans at least $s+t+1$ edges. But the probability of this occurring in $G_{n,p}$ is at most

$$\sum_{k=1}^{3n_1} \binom{n}{k} k^{k-2} \binom{\binom{k}{2}}{2} p^{k+1} \binom{k}{3k/10} \left(\sum_{\ell=1}^{99} \binom{n-k}{\ell} p^\ell (1-p)^{n-k-\ell} \right)^{3k/10}$$
$$\le \sum_{k=1}^{3n_1} \left(\frac{ne}{k}\right)^k k^{k+2} \left(\frac{c}{n}\right)^{k+1} 2^k e^{-3kc/20} \le \sum_{k=1}^{3n_1} \frac{ck^2}{n} \cdot \left(2ce^{1-3c/20}\right)^k = o(1).$$

This completes the proof for Case 1.

Case 2: $n_1 < |S| \le n/4$**.**
The choice of the sets V_1, V_2^* conditions $G_{n,p}$. To get around this, we describe a larger event $\mathcal{E}_S$ in $G = G_{n,p}$ that (a) occurs as a consequence of there being a set S with small expansion and (b) only occurs with probability $o(1)$. This event involves an arbitrary choice for V_1, V_2^*.

Let $T = N_b(S)$ and $W = N_G(S) \setminus N_b(S)$, that is T and W are the neighborhoods of S in G inside and outside of V_1 respectively. Then the following event $\mathcal{E}_S$ must hold. There exist S, T, W such that, where $s = |S|, t = |T|$ and $w = |W|$,

(i) $t \le 2s$.

(ii) $w \le n_0 = ne^{-3c/5}$, where n_0 is a bound on $|V(\Gamma)| + |V(G \setminus C_2)|$ (see (2) and (7)).

(iii) No vertex in S is connected to a vertex in $V \setminus (S \cup T \cup W)$ by a dark blue edge.

(iv) $S \cup N_S$ spans at least $s+t$ edges (at least $s+t+1$ in fact).

Thus,

$$\mathbf{Pr}(\mathcal{E}_S \mid s,t,w)$$
$$\le \binom{n}{s}\binom{n}{t}\binom{n}{w}\binom{\binom{s+t}{2}}{s+t} s^w p^{s+t+w} \left(1 - \frac{p}{2000}\right)^{s(n-s-t-w)}$$

$$\leq \left(\frac{en}{s}\right)^s \left(\frac{en}{t}\right)^t \left(\frac{en}{w}\right)^w \left(\frac{e(s+t)}{2}\right)^{s+t} s^w \left(\frac{c}{n}\right)^{s+t+w} \exp\left\{-\frac{p}{2000}\left(\frac{sn}{5}\right)\right\}$$
$$\leq (ec)^{2(s+t)} \left(\frac{s+t}{2s}\right)^s \left(\frac{s+t}{2t}\right)^t \left(\frac{ecs}{w}\right)^w \exp\left\{-\frac{cs}{10^5}\right\}$$
$$\leq (ec)^{6s} \exp\left\{s \cdot \frac{t-s}{2s}\right\} \exp\left\{t \cdot \frac{s-t}{2t}\right\} \left(\frac{ecs}{n_0}\right)^{n_0} \exp\left\{-\frac{cs}{10^5}\right\}$$
$$\leq (ec)^{6s} (ce^{1-c/3})^{se^{-c/3}} \exp\left\{-\frac{cs}{10^5}\right\} = \left((ec)^6 (ce^{1-c/3})^{e^{-c/3}} e^{-c/10^5}\right)^s.$$

At the 5th line we used $\frac{s+t}{2s} = 1 + \frac{t-s}{2s} \leq \exp\left\{\frac{t-s}{2s}\right\}$ and $w \leq n_0 \leq 100c^3 e^{-c/2} s \leq e^{-c/3} s$. Hence

$$\mathbf{Pr}(\exists S : \mathcal{E}_S) \leq n \sum_{s=n/(100c^3)}^{n/4} \sum_{t=0}^{2s} \left((ec)^6 (ce^{1-c/3})^{e^{-c/3}} e^{-c/10^5}\right)^s = o(1).$$

□

3.4 The Degrees of the Green Subgraph

Lemma 3.5. *W.h.p. at least* $99n/100$ *vertices in* V_1 *have green degree at least* $c/50$*. In addition every set* $S \subset V_1$ *of size at least* $n/4$ *has total green degree at least* $cn/250$*.*

Proof. At most $100n$ edges are colored light blue and thereafter the Chernoff bounds imply that w.h.p. at most $(1+\epsilon)cn/4000$ edges are colored dark blue, for some arbitrarily small positive ε. The degree of a fixed vertex in $G_{n,p}$ is asymptotically Poisson with mean c (see [18], Chapter 3). So, the probability that a vertex has degree less than $c/4$ in $G_{n,p}$ is bounded by $\frac{2e^{-c}\lambda^{c/4}}{c/4!} < 1/1000$. Azuma's inequality or the Chebyshev inequality can be employed to show that w.h.p. there are at most $n/1000$ vertices of degree less than $c/4$ in $G_{n,p}$. Therefore every set of $n/100$ vertices is incident with at least $[(n/100 - n/1000)c/4]/2$ edges. And hence with at least $[(n/100 - n/1000)c/4]/2 - (1+\epsilon)cn/4000 - 100n \geq c/50 \cdot n/100$ green edges. Thus in every set of vertices of size at least $n/100$ there exists a vertex that is incident to $c/50$ green edges, proving the first part of our Lemma.

It follows that w.h.p. every set of size $n/4$ has total green degree at least

$$\left(\frac{n}{4} - \frac{n}{100}\right) \times \frac{c}{50} > \frac{cn}{250}.$$

□

3.5 Pósa Rotations

Pósa Rotations [24] are a standard tool in the analysis of Hamilton cycles in random graphs, see for example [18], Chapter 6.2. It is a procedure that starts with a longest path and outputs many pairs of vertices that are the endpoints of longest paths. Here we marginally modify the standard argument.

We say that a path/cycle P in Γ^* is *compatible* if for every $\{v,w\} \in \mathbb{M}^*$ either P contains the edge $\{v,w\}$ or $V(P) \cap \{v,w\} = \emptyset$. Our aim therefore is to show that w.h.p. Γ^* contains a compatible Hamilton cycle. Suppose that Γ^* is not Hamiltonian and that $P = (v_1, v_2, \ldots, v_s)$ is a longest compatible path in some graph Γ'_b, $\Gamma^*_b \subseteq \Gamma'_b \subseteq \Gamma^*$. If $\{v_s, v_i\} \in E(\Gamma^*_b)$ and $v_i \in V_1$ then the path $P' = (v_1, v_2, \ldots, v_i, v_s, v_{s-1}, \ldots, v_{i+1})$ is said to be obtained from P by an *acceptable* rotation with v_1 as the fixed endpoint. We also call v_i the *pivot vertex*, the edges $\{v_s, v_i\}, \{v_i, v_{i+1}\}$ the *pivot edges* and the edge $\{v_s, v_i\}$ the inserting edge. Observe that even though we are searching for the longest path in Γ'_b we only allow the insertion of edges from Γ^*_b. In addition, since P is compatible and $\{v_i, v_{i+1}\} \notin \mathbb{M}^*$ (since $v_i \in V_1$) then P' is also compatible.

Let $END'_b(P, v_1)$ be the set of vertices that are endpoints of paths that are obtainable from P by a sequence of acceptable rotations with v_1 as the fixed endpoint. Then, for $v \in END'_b(P, v_1)$ we let $END'_b(P_v, v)$ be defined similarly. Here P_v is a path with endpoints v_1, v obtainable from P by a sequence of acceptable rotations.

Pósa's lemma states that $|N_b(END'_b(P, v_1))| < 2|END'_b(P, v_1)|$ in the case where $\mathbb{M}^* = \emptyset$ (see for example Lemma 6.6.of [18]). Arguing as in the proof of Pósa's lemma we see that

$$|N_b(END'_b(P, v_1))| < 2|END'_b(P, v_1)|. \tag{13}$$

Indeed, assume otherwise. Then there exist vertices $v_i, u \in V(P)$ such that $u \in END'_b(P, v_1)$, $v_i \in N_b(u) \subseteq V_1$, $v_{i-1}, v_{i+1} \notin END'_b(P, v_1)$. $v_i \in V_1$ implies that neither of $\{v_{i-1}, v_i\}, \{v_i, v_{i+1}\}$ belongs to $\mathbb{M}^*$ and the edge $\{u, v_i\}$ can be used by an acceptable rotation with v_1 as the fixed endpoint that "rotates out" u. Any such rotation will create a path with either v_{i-1} or v_{i+1} as a new endpoint, say v_{i-1}. Hence $v_{i-1} \in END'_b(P, v_1)$ resulting in a contradiction.

Lemma 3.6. *Let Γ'_b be any graph satisfying $\Gamma^*_b \subseteq \Gamma'_b \subseteq \Gamma^*$. W.h.p. for every path P of maximal length in Γ'_b and an endpoint v of P we have that $|END'_b(P_v, v)| \geq n/4$.*

Proof. We will show that $S = END'_b(P_v, v)$ satisfies (i), (ii), (iii) of Lemma 3.4. For this let $R = R(P_v, v)$ be the set of pivot points and $E_R = E_R(P)$ be the set of pivot edges. It is shown in [6] (Lemma 5) and also in [16] (Lemma 2.1) that if S is the set of endpoints created by Pósa rotations then E_R spans a connected subgraph on $S \cup R$ that consists of at least $|S| + |R \setminus S| + 1$ edges.

The key observation is that if v is the pivot vertex of an acceptable rotation then, by definition, the associated pivot edges do not belong to $\mathbb{M}^*$. Consequently every edge in E_R belongs to $E(\Gamma) \setminus \mathbb{M}^* \subseteq E(G_{n,p})$. This would not have necessarily been true if E_R contained an edge of $\mathbb{M}^*$. Finally, $N_b(S) \setminus R \subset V_1$ and therefore $(N_b(S) \setminus R) : S$ spans at least $|N_b(S) \setminus R|$ edges in $E(\Gamma) \setminus \mathbb{M}^* \subseteq E(G_{n,p})$. Hence $N_b(S) \cup S$ is connected in $G_{n,p}$ and spans at least $(|S| + |R \setminus S| + 1) + |N_b(S) \setminus R| = |S| + |N_b(S)| + 1$ edges. This verifies conditions (ii) and (iii) of Lemma 3.4. Finally (13) implies condition (i). □

From Lemma 3.6 we see that w.h.p. $|END'_b(P_v, v)| \geq n/4$ for all $v \in END'_b(P, v_1)$. We let

$$END'_b(P) = END'_b(P, v_1) \cup \bigcup_{v \in END'_b(P,v_1)} END'_b(P_v, v).$$

3.6 Coloring argument

We use a modification of a double counting argument that was first used in [13]. The specific version is from [14]. Given a two edge-colored Γ^*, we choose for each $v \in V_1$, an incident edge $\xi_v = \{v, \eta_v\}$ where $\eta_v \in V_1 \cup V_2^*$. We color ξ_v yellow if it is not already colored blue. We then color the rest of the green edges red. We denote the yellow and red subgraphs of Γ_g^* by Γ_y^* and Γ_r^* respectively. There are at most $\Pi = \prod_{v \in V_1} d(v)$ choices for $\boldsymbol{\xi} = (\xi_v, v \in V_1)$.

Let $\mathcal{G}(\mathbf{d}_g)$ be the set of graphs with degree sequence $\mathbf{d}_g$ and $\Phi = |\mathcal{G}(\mathbf{d}_g)|$. For a fixed set of yellow edges, defined by $\boldsymbol{\xi}$, we let $\mathbf{d}_g^{\boldsymbol{\xi}}$ be the degree sequence of the red graph and $\mathcal{G}(\mathbf{d}_g^{\boldsymbol{\xi}})$ be the set of graphs with degree sequence $\mathbf{d}_g^{\boldsymbol{\xi}}$. Thus given $\mathbf{d}_g$ and conditional on $\boldsymbol{\xi}$, Γ_r^* is a random member of $\mathcal{G}(\mathbf{d}_g^{\boldsymbol{\xi}})$. In addition, since every red graph can be extended to a green graph via the addition of the yellow edges, we have that $\Phi_{\boldsymbol{\xi}} \leq \Phi$ where $\Phi_{\boldsymbol{\xi}}$ denotes $|\mathcal{G}(\mathbf{d}_g^{\boldsymbol{\xi}})|$.

For a graph Γ, $\Gamma = \Gamma^*$ or $\Gamma_b^* \cup \Gamma_y^*$ we let $\ell(\Gamma)$ denote the length of the longest compatible path in Γ.

We now reveal Γ_b^*. For given $\boldsymbol{\xi}$ and $\Gamma_r^* \in \mathcal{G}(\mathbf{d}_g^{\boldsymbol{\xi}})$ we let $a(\boldsymbol{\xi}, \Gamma_r^*) = 1$ if H1,H2,H3 below hold, and equal to 0 otherwise:

H1 : Γ^* is not Hamiltonian.

H2 : $\ell(\Gamma_b^* \cup \Gamma_y^*) = \ell(\Gamma^*)$.

H3 : With $\Gamma'_b = \Gamma_b^* \cup \Gamma_y^*$, for every path P of maximal length in Γ'_b and an endpoint v of P we have that $|END'_b(P_v, v)| \geq n/4$.

Let $\pi_{\bar{H}}$ be the probability that Γ^* is not Hamiltonian.

Lemma 3.7.

$$\pi_{\bar{H}} \leq \frac{\displaystyle\sum_{\boldsymbol{\xi}} \sum_{\Gamma_r^* \in \mathcal{G}(\mathbf{d}_g^{\boldsymbol{\xi}})} a(\boldsymbol{\xi}, \Gamma_r^*)}{\Phi} + o(1). \tag{14}$$

Proof. The $o(1)$ term accounts for the probability that H3 fails which is related to the already revealed Γ_b^* and by Lemma 3.6 is o(1). If H3 is satisfied and Γ^* is not Hamiltonian then Γ_g^* belongs to

$$\mathcal{G}_{\bar{H}} = \{\Gamma' \in \mathcal{G}(\mathbf{d}_g) : \Gamma_b^* \cup \Gamma' \text{ is not Hamiltonian}\}.$$

If Γ_g^* belongs to $\mathcal{G}_{\bar{H}}$ then there exists $\boldsymbol{\xi}$ such that $a(\boldsymbol{\xi},\Gamma_r^*)=1$. Indeed, let $P=(v_1,v_2,\ldots,v_r)$ be a longest path in Γ^*. Then we simply let ξ_{v_i} be the edge $\{v_i,v_{i+1}\}$ for $1\leq i<r$. Since Γ_g^* is a random member of $\mathcal{G}(\mathbf{d}_g)$, it follows that

$$\pi_{\bar{H}}\leq\frac{|\mathcal{G}_{\bar{H}}|}{\Phi}+o(1)\leq\frac{\sum_{\boldsymbol{\xi}}\sum_{\Gamma_r^*\in\mathcal{G}(\mathbf{d}_g^{\boldsymbol{\xi}})}a(\boldsymbol{\xi},\Gamma_r^*)}{\Phi}+o(1).$$

□

For fixed $\boldsymbol{\xi}$ we let $P_{\boldsymbol{\xi}}$ be a fixed longest path in $\Gamma_b^*\cup\Gamma_y^*$ and $\pi_{\boldsymbol{\xi}}$ be the probability that a random element of $\mathcal{G}(\mathbf{d}_g^{\boldsymbol{\xi}})$ does not include a pair $\{x,y\}$ where $y\in END_b'(P_{\boldsymbol{\xi}},x)$. It follows that

$$\sum_{\boldsymbol{\xi}}\sum_{\Gamma_r^*\in\mathcal{G}(\mathbf{d}_g^{\boldsymbol{\xi}})}a(\boldsymbol{\xi},\Gamma_r^*)\leq\sum_{\boldsymbol{\xi}}\Phi_{\boldsymbol{\xi}}\pi_{\boldsymbol{\xi}}\leq\Phi\Pi\max_{\boldsymbol{\xi}}\pi_{\boldsymbol{\xi}}. \tag{15}$$

Lemma 3.8.

$$\max_{\boldsymbol{\xi}}\pi_{\boldsymbol{\xi}}\leq e^{-cn/10^6}.$$

Proof. This is an exercise in the use of the configuration model of Bollobás [5]. Let $W=[2M_g]$ where M_g is the number of green edges and let $W_1,W_2,\ldots,W_N$ be a partition of W where $|W_v|=d_{\Gamma_g^*}(v),v\in V_1$. The elements of W will be referred to as *configuration points* or just as points. A *configuration* F is a partition of W into M_g pairs. Next define $\psi:W\to[N]$ by $x\in W_{\psi(x)}$. Given F, we let $\gamma(F)$ denote the (muti)graph with vertex set V_1 and an edge $\{\psi(x),\psi(y)\}$ for all $\{x,y\}\in F$. We say that $\gamma(F)$ is simple if it has no loops or multiple edges. Suppose that we choose F at random. The properties of F that we need are

P1 If $G_1,G_2\in\mathcal{G}_{\mathbf{d}_g}$ then $\mathbf{Pr}(\gamma(F)=G_1\mid\gamma(F)\text{ is simple})=\mathbf{Pr}(\gamma(F)=G_2\mid\gamma(F)\text{ is simple})$.

P2 $\mathbf{Pr}(\gamma(F)\text{ is simple})=\Omega(1)$.

These are well established properties of the configuration model, see for example Chapter 11 of [18]. Note that **P2** uses the fact that w.h.p. $G_{V_1,M}^{\delta\geq3}$ (and hence Γ_g^*) has an exponential tail, as shown for example in [17].

Given all this, in the context of the configuration model, we have the following simple consequence of a random pairing of W.

$$\max_{\boldsymbol{\xi}}\pi_{\boldsymbol{\xi}}\leq\max_{\boldsymbol{\xi}}O(1)\times\prod_{v\in END_b'(P_{\boldsymbol{\xi}})}\left(1-\frac{d_{\Gamma_g^*}(v)\sum_{w\in END_b'(P_v,v)}d_{\Gamma_g^*}(w)}{2M}\right)^{\frac{1}{2}} \tag{16}$$

$$\leq \max_{\boldsymbol{\xi}} O(1) \times \exp\left\{-\frac{\sum_{v \in END'_b(P_{\boldsymbol{\xi}})} d_{\Gamma_g^*}(v) \sum_{w \in END'_b(P_v,v)} d_{\Gamma_g^*}(w)}{4M}\right\}. \tag{17}$$

The $O(1)$ factor is $1/\,\mathbf{Pr}(\gamma(F)$ is simple) and bounds the effect of the conditioning. We take the square root to account for the possibility that $w \in END'_b(P_v, v)$ and $v \in END'_b(P_w, w)$.

Lemma 3.5 implies that at least $n/4 - n/100$ out of the at least $n/4$ vertices in $END'_b(P)$ have $d_{\Gamma_g^*}(v) \geq c/50$. Also, for such v the set $END'_b(P_v, v) \cup \{v\}$ is of size at least $n/4$ and so has total degree at least $cn/250$. Thus from (17), it follows that

$$\max_{\boldsymbol{\xi}} \pi_{\boldsymbol{\xi}} \leq O(1) \times \exp\left\{-\frac{\frac{c}{50} \cdot \left(\frac{n}{4} - \frac{n}{100}\right) \cdot \frac{cn}{250}}{4M}\right\} \leq e^{-cn/10^6}.$$

□

The Arithmetic-Geometric-mean inequality implies that

$$\Pi \leq \prod_{v \in V_1} d(v) \leq \left(\frac{\sum_{v \in V} d(v)}{N}\right)^N \leq (2c)^n. \tag{18}$$

It then follows from Lemmas 3.7, 3.8 and from (18) that for sufficiently large c

$$\pi_{\overline{H}} \leq (2c)^n \cdot e^{-cn/10^6} + o(1) = o(1),$$

and this completes the proof of Theorem 3.2.

4 Proof of (3)

We are not able at this time to give a simple estimate of $\sum_{T \in \mathcal{T}} \phi(T)$ as a function of c. We will have to make do with (3). On the other hand, $\sum_{T \in \mathcal{T}} \phi(T)$ can be approximated to within arbitrary accuracy, using the argument in Section 5.

We work in $G_{n,p}$. Observe that a tree T is spanned by C_2 and satisfies $\phi(T) > 0$ only if (i) it has a vertex with at least 3 neighbors in $V(\Gamma) \setminus V_2$ each having degree at least 2 in T and (ii) all the vertices of T of degree 1 belong to V_2. Here we are using that no vertex in $V(T) \cap V_2$ contributes to $\phi(T)$ as it can be considered as an individual path of length 0.

The smallest such tree is T' the tree on seven vertices that consists of three paths of length two with a common endpoint. In addition every tree T satisfying (i) and (ii) and intersects $V(\Gamma) \setminus V_2$ in exactly 3 vertices has T' as a subtree. Since $\phi(T') = 1$ we have in $G_{n,p}$, as in the proof of Lemma 10,

$$\mathbf{E}\left(\sum_{T \in \mathcal{T}} \phi(T)\right) = n\binom{n-1}{3}\binom{n-4}{3} p^6 (1-p)^{3(n-7)}$$

$$
\begin{aligned}
&+ O\Bigg(\sum_{k\geq 7} k\cdot \binom{n}{k} k^{k-2}p^{k-1}(1-p)^{(n-k)\max\{4,k/3\}}\Bigg)\\
&\approx \frac{c^6e^{-3}n}{36} + \sum_{k\geq 7}\left(\frac{ne}{k}\right)^k k^{k-1}\left(\frac{c}{n}\right)^{k-1}\exp\left\{-c(1-k/n)\max\left\{4,\frac{k}{3}\right\}\right\}\\
&= \frac{c^6e^{-3}n}{36} + O(c^6e^{-4c})n. \qquad (19)
\end{aligned}
$$

In the first line we used that every tree that contributes to $\mathbf{E}\left(\sum_{T\in\mathcal{T}}\phi(T)\right)$ either satisfies $\upsilon_0(T)=3$ and spans a copy of T' or satisfies both $\upsilon_0(T)\geq 4$ and (8) i.e. $\upsilon_0(T)\geq |T|/3$. We obtain (3) from (19).

5 Proof of Theorem 1.6

For $v\in C_2$ we let $\phi(v)=\phi(T)/|\upsilon_0(T)|$ if $v\in \upsilon_0(T)$ for some $T\in\mathcal{T}$ and $\phi(v)=0$ otherwise. (Recall that $\upsilon_0(T)=V(T)\setminus V_2$.) Thus

$$\sum_{T\in\mathcal{T}}\phi(T)=\sum_{v\in C_2}\phi(v).$$

Hence (1) can be rewritten as,

$$L_{c,n}\approx |C_2|-\sum_{v\in C_2}\phi(v). \qquad (20)$$

To prove Theorem 1.6 we show that there for every $\epsilon>0$ there exists a set of vertices S_ϵ of size $|S_\epsilon|\geq (1-\epsilon)|C_2|$ such that for every $v\in S_\epsilon$ we can evaluate correctly $\phi(v)$ via a procedure described later on. This evaluation will be based on the first $k=k(\epsilon)$, neighborhoods of v. Hence the distribution of $\sum_{v\in C_2}\phi(v)$ can be tied to the distribution of the first k neighborhoods of a random vertex which we then relate to the expected number of appearances of small subgraphs in C_2.

Let $\epsilon>0$. Let $k_1=k_1(\epsilon,c)$ be the smallest positive integer such that

$$\sum_{k=k_1-1}^{\infty}(e^3 2^3 c e^{-c/4})^k<\frac{\epsilon}{3}.$$

Note that for large c, we have

$$k_1\leq \frac{2}{c}\log\frac{1}{\varepsilon}. \qquad (21)$$

Notation 5.1. *For $v\in C_2$ let $N_k(v)$ (and $N_{\leq k}(v)$ respectively) be the set of vertices in $V(C_2)$ that are in distance exactly k (at most k respectively) from v in C_2.*

For $v \in C_2$ let G_v be the graph that is formed as follows: Starting with the graph spanned by $N_{\leq k}(w)$ for every vertex $w \in N_k(v)$ we introduce $K_{3,3}^w$, a copy of $K_{3,3}$, and we join w to each vertex of the same part of the bipartition of $K_{3,3}^w$. We consider the algorithm for the construction of Γ on G_v and let $C_{2,v}, \Gamma_v, V_{1,v}, V_{2,v}, S_{L,v}, \upsilon_{0,v}(T)$ be the corresponding sets/quantities.

For a tree $T \in S_{L,v}$ let $f(T)$ be equal to $|T|$ minus the maximum number of vertices that can be covered by a set of vertex disjoint paths with endpoints in $V_{2,v}$ (we allow paths of length 0). For $v \in C_{2,v}$, if v belongs to some tree $T \in S_{L,v}$ set $f(v) = f(T)/\upsilon_{0,v}(T)$, otherwise set $f(v) = 0$.

For $v \in C_2$ let $t(v) = 1$ if $v \in V_1$ or if $v \in S_L$ and in Γ, v lies in a component with at most $k_1 - 2$ vertices that are not connected to V_1 in G. Set $t(v) = 0$ otherwise. Observe that if $t(v) = 1$ then $\phi(v) = f(v)$. Otherwise $|\phi(v) - f(v)| \leq 1$.

Lemma 5.2. *The expected number of vertices v satisfying $t(v) = 0$ is bounded by $\frac{\epsilon n}{3}$.*

Proof. By repeating the arguments used to prove (10) and (8) it follows that if $t(v) = 0$ then v lies on a component C of size at most $\log n$. In addition at least $\max\{|V(C)|/3, k_1 - 1\}$ vertices in $V(C)$ are not adjacent to any C_2-vertex outside $V(C)$. So,

$$\begin{aligned}
\mathbf{E}(|\{v : t(v) = 0\}|) &\leq \sum_{k=k_1-1}^{\log^2 n} \sum_{j=k}^{3k} \binom{n}{j} \binom{j}{k} j^{j-2} p^{j-1} (1-p)^{k(n-j)} \\
&\leq n \sum_{k=k_1-1}^{\log^2 n} 3k \left(\frac{e}{3k}\right)^{3k} 2^{3k} (3k)^{3k-2} c^{k-1} e^{-ck/4} \\
&\leq n \sum_{k=k_1-1}^{\infty} (e^3 2^3 c e^{-c/4})^k < \frac{\epsilon n}{3}.
\end{aligned}$$

□

Notation 5.3. *A vertex $v \in [n]$ is* ε-good *if $N_i(v) \leq 3c^i k_1/\epsilon$ for every $i \leq k_1$ and it is ε* bad *otherwise.*

Lemma 5.4.

$$\mathbf{E}\left(\left|\sum_{v \in V} \phi(v) - \sum_{v \text{ is } \varepsilon\text{-good}} f(v)\right|\right) \leq \epsilon n.$$

Proof. Because the expected size of the i^{th} neighborhood of every vertex in G is $\approx c^i$ we have by the Markov inequality that v is ε-bad with probability at most $\approx \varepsilon/3k_1$ and so the expected number of ε-bad vertices is bounded by $\varepsilon n/2$. Thus,

$$\mathbf{E}\left(\left|\sum_{v \in V} \phi(v) - \sum_{v \text{ is } \varepsilon\text{-good}} f(v)\right|\right) \leq \mathbf{E}\left(\left|\sum_{v \in V} \phi(v) - \sum_{v \in V} f(v)\right|\right) + \mathbf{E}\left(\left|\sum_{v \text{ is } \varepsilon\text{-bad}} f(v)\right|\right)$$

$$\le \mathbf{E}\left(\left|\sum_{v:t(v)=0} |\phi(v) - f(v)|\right|\right) + \mathbf{E}\left(\sum_{v \text{ is } \varepsilon\text{-bad}} 1\right)$$

$$\le \mathbf{E}\left(\sum_{v:t(v)=0} 1\right) + \frac{\epsilon n}{2}$$

$$\le \frac{\epsilon n}{3} + \frac{\epsilon n}{2} < \epsilon n.$$

□

Let $\mathcal{H}_\varepsilon$ be the set of pairs (H, o_H) where H is a graph, o_H is a distinguished vertex of H, that is considered to be the root, every vertex in $V(H)$ is at distance at most k_1 from o_H and all the neighborhoods of o_H are ε-good. For $v \in C_2$ let $G(N_{k_1}(v))$ be the subgraph induced by the ${k_1}^{th}$ neighborhood of v in C_2. For $(H, o_H) \in \mathcal{H}_\varepsilon$ let $Aut(H, o_H)$ be the number of automorphisms of H that fix o_H. Note that each ε-good vertex v is associated with a pair $(H, o_H) \in \mathcal{H}_\varepsilon$ from which we can compute $f(v)$, since $f(v) = f(o_H)$. Let

$$f_\varepsilon(c) = \sum_{k \ge 1} \sum_{\substack{(H,o_H) \in \mathcal{H}_\varepsilon \\ H \text{ is a tree}}} \frac{f(o_H)}{Aut(H, o_H)} \left(\frac{N_2}{2M_2}\right)^{k-1} \lambda^{2k-2} \frac{f_2(k\lambda)}{f_2(\lambda)^k}. \tag{22}$$

Lemma 5.5. *Let $M_2 = |E(C_2)|$ and $N_2 = |C_2|$. Then*

$$\mathbf{E}\left(\sum_{v \text{ is } \varepsilon\text{-good}} f(v) \middle| M_2, N_2\right) = o(n) + f_\varepsilon(c)n.$$

Proof.

$$\mathbf{E}\left(\sum_{v \text{ is } \varepsilon\text{-good}} f(v) \middle| M_2, N_2\right) = \sum_v \sum_{k \ge 1} \sum_{\substack{(H,o_H) \in \mathcal{H}_\varepsilon \\ (G(N_{k_1}(v)),v)=(H,o_H) \\ |V(H)|=k}} \rho_{H,o_H} f(o_H)$$

$$= o(n) + \sum_v \sum_{k \ge 1} \sum_{\substack{(H,o_H) \in \mathcal{H}_\varepsilon \\ H \text{ is a tree} \\ (G(N_{k_1}(v)),v)=(H,o_H)}} \rho_{H,o_H} f(o_H), \tag{23}$$

where ρ_{H,σ_H} is the probability $(G(N_{k_1}(v)), v) = (H, o_H)$ in C_2. The $o(n)$ term in (23) is an upper bound on the number of vertices v such that $N_{\le k}(v)$ spans a cycle in G, hence in C_2.

We show in Section 5.1 that

$$\rho_{H,o_H} \approx \frac{1}{Aut(H, o_H)} \left(\frac{N_2}{2M_2}\right)^{k-1} \lambda^{2k-2} \frac{e^{k\lambda}}{f_2(\lambda)^k}, \tag{24}$$

where f_k is defined in (28) below and λ satisfies (29) below. □

Proof of part (a) of Theorem 1.6: $f_\varepsilon(c)$ is monotone increasing as $\varepsilon \to 0$. This is simply because $\mathcal{H}_\varepsilon$ grows. Furthermore, $f_\varepsilon(c) \leq 1$ and so the limit $f(c) = \lim_{\varepsilon\to 0} f_\varepsilon(c)$ exists.

Let $\epsilon' > 0$. Take ϵ sufficiently small such that $\max\{|f_\varepsilon(c) - f(c)|, \varepsilon\} \leq \varepsilon'/3$. Theorem 1.3 and Lemmas 5.4 and 5.5 imply that for sufficiently large n,

$$\left|\frac{\mathbf{E}[L_{c,n}]}{n} - f(c)\right| \leq \left|\mathbf{E}\left(\left|\frac{\sum_{v\in V}\phi(v) - \sum_{v \text{ is } \varepsilon\text{-good}} f(v)}{n}\right|\right)\right| + |f_\varepsilon(c) - f(c)| + o(1) \leq \varepsilon.$$

□

Proof of part (b) of Theorem 1.6: For a graph G let $C_2(G)$ be the 2-core of its largest component. We let $\mathcal{G}$ be the set of graphs on n vertices and with at most n^2p edges such that for $G \in \mathcal{G}$ the following holds:

(i) the largest component in $G \setminus C_2(G)$ is of size at most $\log n$.

(ii) at most $\log n$ vertices lie in a non-tree component in $G \setminus C_2(G)$.

(iii) the length of the largest path in G satisfies (1).

Theorem 1.3 and Lemmas 2.6 and 2.7 imply that $\mathbf{Pr}(G_{n,p} \notin \mathcal{G}) = o(1)$. Hence

$$\mathbf{E}(L_{c,n}) = \mathbf{E}(L_{c,n}|G_{n,p} \in \mathcal{G}) + o(n). \tag{25}$$

We now implement an edge exposure martingale to reveal $G_{n,p}$, conditioned that it belongs to $\mathcal{G}$ and $|E(G_{n,p}| = m$: let $e_1, e_2, \ldots, e_{2m}$ be chosen randomly from $\binom{n}{2}^m$.

Now let $e_1, e_2, \ldots, e_m$ and $e'_1, e'_2, \ldots, e'_m$ be two edge sequences that differ in a single edge say $e_i \neq e'_i$ such that the corresponding graphs G and G' belong to $\mathcal{G}$. Then, G, G' differ in at most 4 components (the ones containing a vertex in $e_i \cup e'_i$) and therefore conditions (i)-(iii) imply that the length of the longest paths in G,G' differ by at most $1 + 3\log n + 8\log n$. The 1 and $3\log n$ originate from (1), a $4\log n$ term accounts for the difference in the size of the 2-cores and a $4\log n$ term for the difference in at most 4 components outside the 2-cores. Azuma's inequality (see Lemma 11 of Frieze and Pittel [20] or Section 3.2 of McDiarmid [22]) implies that

$$\mathbf{Pr}\left[\left|L_{c,n} - E[L_{c,n}|G_{n,p} \in \mathcal{G}]\right|\Big| G_{n,p} \in \mathcal{G} \geq n^{0.8}\right] \leq e^{-0.5n}. \tag{26}$$

(25), (26) and part (a) of Theorem 1.6 imply that for $\varepsilon > 0$ and sufficiently large n,

$$\mathbf{Pr}\left[\left|\frac{L_{c,n}}{n} - f(c)\right| \geq \epsilon\right] \leq e^{-0.5n}. \tag{27}$$

(6) follows from (27) and the Borel-Cantelli lemma. □

5.1 A Model of C_2

It is known that given M_2, N_2 that, up to relabeling vetices, C_2 is distributed as $G^{\delta\geq 2}_{N_2,M_2}$ (see for example the first section of [20]). The random graph $G^{\delta\geq 2}_{N_2,M_2}$ is chosen uniformly from $\mathcal{G}^{\delta\geq 2}_{N_2,M_2}$ which is the set of graphs with vertex set $[N_2]$, M_2 edges and minimum degree at least two. From now, we replace M_2, N_2 by M, N respectively.

5.1.1 Random Sequence Model

We must now take some time to explain the model we use for $G^{\delta\geq 2}_{N,M}$. We use a variation on the pseudo-graph model of Bollobás and Frieze [8] and Chvátal [9]. Given a sequence $\mathbf{x} = (x_1, x_2, \ldots, x_{2M}) \in [n]^{2M}$ of $2M$ integers between 1 and N we can define a (multi)-graph $G_{\mathbf{x}} = G_{\mathbf{x}}(N, M)$ with vertex set $[N]$ and edge set $\{(x_{2i-1}, x_{2i}) : 1 \leq i \leq M\}$. The degree $d_{\mathbf{x}}(v)$ of $v \in [N]$ is given by

$$d_{\mathbf{x}}(v) = |\{j \in [2M] : x_j = v\}|.$$

If $\mathbf{x}$ is chosen randomly from $[N]^{2M}$ then $G_{\mathbf{x}}$ is close in distribution to $G_{N,M}$. Indeed, conditional on being simple, $G_{\mathbf{x}}$ is distributed as $G_{N,M}$. To see this, note that if $G_{\mathbf{x}}$ is simple then it has vertex set $[N]$ and M edges. Also, there are $M!2^M$ distinct equally likely values of $\mathbf{x}$ which yield the same graph.

Our situation is complicated by there being a lower bound of 2 on the minimum degree. So we let

$$[N]^{2M}_{\delta\geq 2} = \{\mathbf{x} \in [N]^{2M} : d_{\mathbf{x}}(j) \geq 2 \text{ for } j \in [N]\}.$$

Let $G_{\mathbf{x}}$ be the multi-graph $G_{\mathbf{x}}$ for $\mathbf{x}$ chosen uniformly from $[N]^{2M}_{\delta\geq 2}$. It is clear then that conditional on being simple, $G_{\mathbf{x}}$ has the same distribution as $G^{\delta\geq 2}_{N,M}$. It is important therefore to estimate the probability that this graph is simple. For this and other reasons, we need to have an understanding of the degree sequence $d_{\mathbf{x}}$ when $\mathbf{x}$ is drawn uniformly from $[N]^{2M}_{\delta\geq 2}$. Let

$$f_k(\lambda) = e^\lambda - \sum_{i=0}^{k-1} \frac{\lambda^i}{i!} \tag{28}$$

for $k \geq 0$.

Lemma 5.6. *Let $\mathbf{x}$ be chosen randomly from $[N]^{2M}_{\delta\geq 2}$. Let $Z_j, j = 1, 2, \ldots, N$ be independent copies of a* truncated Poisson *random variable $\mathcal{P}$, where*

$$\mathbf{Pr}(\mathcal{P} = t) = \frac{\lambda^t}{t! f_2(\lambda)}, \qquad t \geq 2.$$

Here λ satisfies

$$\frac{\lambda f_1(\lambda)}{f_2(\lambda)} = \frac{2M}{N}. \tag{29}$$

Then $\{d_{\mathbf{x}}(j)\}_{j\in[N]}$ is distributed as $\{Z_j\}_{j\in[N]}$ conditional on $Z = \sum_{j\in[n]} Z_j = 2M$.

Proof. This can be derived as in Lemma 4 of [2]. □

It follows from (12) and (29) and the fact that $f_1(\lambda)/f_2(\lambda) \to 1$ as $c \to \infty$ that for large c,

$$\lambda = c\left(1 + O(ce^{-c})\right). \tag{30}$$

We note that the variance σ^2 of $\mathcal{P}$ is given by

$$\sigma^2 = \frac{\lambda(e^\lambda - 1)^2 - \lambda^3 e^\lambda}{f_2^2(\lambda)}.$$

Furthermore,

$$\mathbf{Pr}\left(\sum_{j=1}^{N} Z_j = 2M\right) = \frac{1}{\sigma\sqrt{2\pi N}}(1 + O(N^{-1}\sigma^{-2})) \tag{31}$$

and

$$\mathbf{Pr}\left(\sum_{j=2}^{N} Z_j = 2M - d\right) = \frac{1}{\sigma\sqrt{2\pi N}}\left(1 + O((d^2+1)N^{-1}\sigma^{-2})\right). \tag{32}$$

This is an example of a local central limit theorem. See for example, (5) of [2] or (3) of [17]. It follows by repeated application of (31) and (32) that if $k = O(1)$ and $d_1^2 + \cdots + d_k^2 = o(N)$ then

$$\mathbf{Pr}\left(Z_i = d_i, i = 1, 2, \ldots, k \mid \sum_{j=1}^{N} Z_j = 2M\right) \approx \prod_{i=1}^{k} \frac{\lambda^{d_i}}{d_i! f_2(\lambda)}. \tag{33}$$

Let $\nu_{\mathbf{x}}(s)$ denote the number of vertices of degree s in $G_{\mathbf{x}}$.

Lemma 5.7. *Suppose that* $\log N = O((N\lambda)^{1/2})$. *Let* $\mathbf{x}$ *be chosen randomly from* $[N]^{2M}_{\delta\geq 2}$. *Then as in equation (7) of [2], we have that with probability* $1 - o(N^{-10})$,

$$\left|\nu_{\mathbf{x}}(j) - \frac{N\lambda^j}{j! f_2(\lambda)}\right| \leq \left(1 + \left(\frac{N\lambda^j}{j! f_2(\lambda)}\right)^{1/2}\right)\log^2 N,\ 2 \leq j \leq \log N. \tag{34}$$

$$\nu_{\mathbf{x}}(j) = 0, \quad j \geq \log N. \tag{35}$$

We can now show $G_{\mathbf{x}}$, $\mathbf{x} \in [n]^{2m}_{\delta\geq 2}$ is a good model for $G^{\delta\geq 2}_{n,m}$. For this we only need to show now that

$$\mathbf{Pr}(G_{\mathbf{x}} \text{ is simple}) = \Omega(1). \tag{36}$$

Again, this follows as in [2].

Given a tree H with k vertices of degrees $z_1, z_2, \ldots, z_k$ and a fixed vertex v we see that if ρ_H is the probability that $G(N_{k_1}(v)) = H$ in $G_{\mathbf{x}}$ then we have

$$\rho_{H,o_H} \approx \binom{N}{k-1}\frac{(k-1)!}{Aut(H, o_H)} \times$$

$$
\sum_{D=2k-2}^{\infty} \sum_{\substack{d_1 \geq z_1, \ldots, d_k \geq z_k \\ d_1+\cdots+d_k=D}} \prod_{i=1}^{k} \frac{\lambda^{d_i}}{d_i! f_2(\lambda)} \cdot \binom{M}{k-1} 2^{k-1}(k-1)! \cdot \prod_{i=1}^{k} \frac{d_i!}{(d_i-z_i)!} \frac{1}{(2M)^{2k-2}} \tag{37}
$$

$$
\approx \left(\frac{N}{2M}\right)^{k-1} \frac{\lambda^{2k-2}}{Aut(H,o_H) f_2(\lambda)^k} \sum_{D=2k-2}^{\infty} \sum_{\substack{d_1 \geq z_1, \ldots, d_k \geq z_k \\ d_1+\cdots+d_k=D}} \prod_{i=1}^{k} \frac{\lambda^{d_i - z_i}}{(d_i - z_i)!}
$$

$$
= \left(\frac{N}{2M}\right)^{k-1} \frac{\lambda^{2k-2}}{Aut(H,o_H) f_2(\lambda)^k} \sum_{D=2k-2}^{\infty} \frac{(k\lambda)^{D-2(k-1)}}{(D-2(k-1))!} \tag{38}
$$

$$
\approx \frac{1}{Aut(H,o_H)} \left(\frac{N}{2M}\right)^{k-1} \lambda^{2k-2} \frac{e^{k\lambda}}{f_2(\lambda)^k}.
$$

Explanation for (37)**:** We use (33) to obtain the probability that the degrees of $[k]$ are $d_1, \ldots, d_k$. This explains the product $\prod_{i=1}^{k} \frac{\lambda^{d_i}}{d_i! f_2(\lambda)}$. Implicit here is that $d_i = O(\log n)$, from (35). The contribution to the degree sum D for $D \geq 2k \log n$ can therefore be shown to be negligible. We use the fact that k is small to argue that w.h.p. H is induced. We choose the vertices, other than v in $\binom{N}{k-1}$ ways and then $\frac{(k-1)!}{Aut(H,o_H)}$ counts the number of copies of H in K_k. We then choose the place in the sequence to put these edges in $\binom{M}{k-1} 2^{k-1}(k-1)!$ ways. Finally note that the probability the z_i occurrences of the ith vertex are as claimed is asymptotically equal to $\frac{d_i(d_i-1)\cdots(d_i-z_i+1)}{(2M)^{z_i}}$ and this explains the factor $\prod_{i=1}^{k} \frac{d_i!}{(d_i-z_i)!} \frac{1}{(2M)^{2k-2}}$.

Explanation for (38)**:** We use the identity

$$
\sum_{\substack{d_1, \ldots, d_k \\ d_1+\cdots+d_k=D}} \frac{D!}{d_1! \cdots d_k!} = k^D.
$$

6 Summary and open problems

We have derived an expression for the length of the longest path in $G_{n,p}$ that holds for large c w.h.p. It would be interesting to have a more algebraic expression. Also, we could no doubt make this proof algorithmic, by using the arguments of Frieze and Haber [17]. It would be more interesting to do the analysis for small $c > 1$. Applying the coupling of McDiarmid [21] we see that the random digraph $D_{n,p}, p = c/n$ contains a path at least as long as that given by the R.H.S. of (4). It should be possible to improve this, just as Krivelevich, Lubetzky and Sudakov [19] did for the earlier result of [15].

References

[1] M. Ajtai, J. Komlós and E. Szemerédi, The longest path in a random graph, *Combinatorica* 1 (1981) 1-12.

[2] J. Aronson, A.M. Frieze and B.G. Pittel, Maximum matchings in sparse random graphs: Karp-Sipser re-visited, *Random Structures and Algorithms* 12 (1998) 111-178.

[3] M. Bayati, D. Gamarnik and P. Tetali, Combinatorial approach to the interpolation method and scaling limits in sparse random graphs, *The Annals of Probability* 41 (2013) 4080-4115.

[4] B. Bollobás, Long paths in sparse random graphs, *Combinatorica* 2 (1982) 223-228.

[5] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labeled regular graphs, *European Journal on Combinatorics* 1 (1980) 311-316.

[6] B. Bollobás, C. Cooper, T.I.Fenner and A.M.Frieze, On Hamilton cycles in sparse random graphs with minimum degree at least k, *Journal of Graph Theory* 34 (2000) 42-59.

[7] B.Bollobás, T.I.Fenner and A.M.Frieze, Long cycles in sparse random graphs, *Graph theory and combinatorics, Proceedings of Cambridge Combinatorial Conference in honour of Paul Erdos* (1984) 59-64.

[8] B.Bollobás and A.M.Frieze, On matchings and Hamiltonian cycles in random graphs, *Annals of Discrete Mathematics* 28 (1985) 23-46.

[9] V. Chvátal, Almost all graphs with $1.44n$ edges are 3-colourable, *Random Structures and Algorithms* 2 (1991) 11-28.

[10] P.Erdős, Problems and results on finite and infinite graphs , in *Proceedings Symposium Prague* 1974.

[11] P.Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* 5 (1960) 17-61.

[12] B. Bollobás, C. Cooper, T. Fenner and A.M. Frieze, On Hamilton cycles in sparse random graphs with minimum degree at least k, *Journal of Graph Theory* 34 (2000) 42-59.

[13] T.I.Fenner and A.M.Frieze, On the existence of Hamiltonian cycles in a class of random graphs, *Discrete Mathematics* 45 (1983) 301-305.

[14] T.I.Fenner and A.M.Frieze, Hamiltonian cycles in random regular graphs, *Journal of Combinatorial Theory, Series B* 40 (1984) 103-112.

[15] A.M. Frieze, On large matchings and cycles in sparse random graphs, *Discrete Mathematics* 59 (1986) 243-256.

[16] A.M.Frieze and B. Pittel. On a sparse random graph with minimum degree three: Likely Posa's sets are large, *Journal of Combinatorics* 4 (2013) 123-156.

[17] A.M. Frieze and S. Haber, An almost linear time algorithm for finding Hamilton cycles in sparse random graphs with minimum degree at least three, *Random Structures and Algorithms* 47 (2015) 73-98.

[18] A.M. Frieze and M. Karoński, Introduction to Random Graphs, Cambridge University Press. 2015.

[19] M. Krivelevich, E. Lubetzky and B. Sudakov, Longest cycles in sparse random digraphs, *Random Structures and Algorithms* 43 (2013) 1-15.

[20] A.M. Frieze and B. Pittel, Perfect matchings in random graphs with prescribed minimal degree, *Trends in Mathematics*, Birkhauser Verlag, Basel (2004) 95-132.

[21] C. McDiarmid, Clutter percolation and random graphs, *Mathematical Programming* 13 (1980) 17-25.

[22] C. McDiarmid, Concentration, *Probabilistic Methods for Algorithmic Discrete Mathematics* (M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, B. Reed eds.), Springer, Berlin (1998) 1-46.

[23] B. Pittel, On tree census and the giant component in sparse random graphs, *Random Structures and Algorithms* 1 (1990), 311-342.

[24] L. Pósa, Hamiltonian circuits in random graphs, *Discrete Mathematics* 14 (1976) 359-364.

[25] S. Suen, On large induced trees and long induced paths in sparse random graphs, *Journal of Combinatorial Theory B* 56 (1992) 250-262.

[26] W. de la Vega, Long paths in random graphs, *Studia Scient. Math. Hungar.* 14 (1979) 335-340.