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

Michael Anastos and Alan Frieze*
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA15213
U.S.A.

July 8, 2019

Abstract

We discuss the length of the longest cycle in a sparse random graph $G_{n,p}, p = c/n$. c constant. We show that for large c there is a function $f(c)$ such that $L_n(c)/n \to f(c)$ a.s. The function $f(c) = 1 - \sum_{k=1}^{\infty} p_k(c)e^{-kc}$ where p_k is a polynomial in k. 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

Erdős conjectured that if $c > 1$ then w.h.p. $G_{n,c/n}$ contains a path of length $f(c)n$ where $f(c) > 0$. This was proved by Ajtai, Komlós and Szemerédi [1] and in a slightly weaker form by de la Vega [21] who proved that if $c > 4 \log 2$ then $f(c) = 1 - O(c^{-1})$. See also Suen [20]. Bollobás [3] 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 $f(c) \geq 1 - c^{24}e^{-c/2}$. This was then improved by Bollobás, Fenner and Frieze [5] to $f(c) \geq 1 - c^6 e^{-c}$ and then by Frieze [12] to $f(c) \geq 1 - (1 + \varepsilon_c)(1 + c)e^{-c}$ where $\varepsilon_c \to 0$ as $c \to \infty$. This last result is optimal up to the value of ε_c, as there are w.h.p. $\approx (1 + c)e^{-c}n$ vertices of degree 0 or 1.

*Research supported in part by NSF Grant DMS1363136

Let $p = c/n$ and let $G = G_{n,p}$. We will assume throughout that c is sufficiently large. Let C_2 denote the 2-core of G. By this we mean that part of the giant component consisting of vertices that are in at least one cycle. The longest cycle in G is contained in C_2 and the length of the longest path differs from this by $O(\log n)$ w.h.p. The reason for this is that w.h.p. the giant component of G consists of C_2 plus a forest of trees with maximum diameter $O(\log n)$.

As in the papers, [3], [5] and [12] we consider a process that builds a large Hamiltonian subgraph. We construct a sequence of sets $S_0 = \emptyset, S_1, S_2, \ldots, S_L \subseteq C_2$ and their induced subgraphs $\Gamma_0, \Gamma_1, \Gamma_2, \ldots, \Gamma_L$. Suppose now that we have constructed S_ℓ, $\ell \geq 0$. We construct $S_{\ell+1}$ from S_ℓ via one of two cases:

Construction of Γ
Case a: If there is $v \in S_\ell$ that has at least one but fewer than 3 neighbors W outside S_ℓ, then we add W to S_ℓ to make $S_{\ell+1}$.
Case b: If there is a vertex $v \in C_2 \setminus S_\ell$ of degree d in G that has more than $d-3$ neighbors in S_ℓ then we define $S_{\ell+1}$ to be S_ℓ plus v plus the neighbors of v that are currently not in S_ℓ.

Note that we allow $d < 3$ here and so low degree vertices are always added to some S_ℓ.

S_L is the set we end up with when there are no more vertices to add. We note that S_L is well-defined and does not depend on the order of adding vertices. Indeed, suppose we have two distinct outcomes $O_1 = v_1, v_2, \ldots, v_r$ and $O_2 = w_1, w_2., \ldots, w_s$. Assume without loss of generality that there exists i which is the smallest index such that $w_i \notin O_1$. Then, $X = \{w_1, w_2, \ldots, w_{i-1}\} \subseteq Y = \{v_1, v_2, \ldots, v_r\}$. If w_i was added in Step a as $v \in X$ then $v \in Y$, contradiction. If w_i is a neighbor of $v \in X$ then v qualifies for Step a at the end of O_1, again a contradiction. Suppose then that w_i is added in Step b. If $w_i = v$ then it would be added to O_1 because we would have added $S_1 \cup X$ and maybe more, a contradiction. If w_i is the neighbor of v then it would also be added after O_1 for the same reason, giving the final contradiction. It follows that $\{w_1, w_2., \ldots, w_s\} \subseteq \{v_1, v_2, \ldots, v_r\}$ and vice-versa, by the same reasoning.

We will argue below in Section 1.1 that w.h.p. the graph Γ_L induced by S_L is a forest plus a few small components. Each tree in Γ_L will w.h.p. have at most $\log n$ vertices. For a tree component T let $v_0(T)$ denote the the set of vertices of T that have no neighbors outside S_L.

Notation 1: Let $\mathcal{T}$ denote the set of trees in Γ_L. For a tree $T \in \mathcal{T}$ let $\mathcal{P}_T$ be the set of vertex disjoint path packings of T where every endpoint of a path in P has a neighbor outside T. Here we allow paths of length 0, so that a single vertex with neighbors outside T 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 $\mathcal{Q}(T) \in \mathcal{P}$ denote a set of paths that leaves $\phi(T)$ vertices of T uncovered i.e. satisfies $n(T, Q(T)) = \phi(T)$.

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

We will prove

Theorem 1.1. *Let $p = c/n$ where $c > 0$ is a sufficiently large constant. Then w.h.p.*

$$L_n \approx |V(C_2)| - \sum_{T \in \mathcal{T}} \phi(T). \tag{1}$$

The size of C_2 is well-known. Let $x < 1$ be the unique solution to $xe^{-x} = ce^{-c}$. Then w.h.p. (see e.g. [15], Lemma 2.16),

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

$$|E(C_2)| \approx \left(1 - \frac{x}{c}\right)^2 \frac{c}{2} n. \tag{3}$$

Equation (4.5) of Erdős and Rényi [8] tells us that

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

We will argue below that w.h.p., as c grows, that

$$\sum_{T \in \mathcal{T}} \phi(T) = O(c^6 e^{-3c}) n. \tag{5}$$

We therefore have the following improvement to the estimate in [12].

Corollary 1.2. *W.h.p., as c grows, that*

$$L_n \approx \left(1 - (c+1)e^{-c} - c^2 e^{-2c} + O(c^6 e^{-3c})\right) n. \tag{6}$$

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 (6). We claim next that there exists some function $f(c)$ such that the sum in (1) is concentrated around $f(c)n$. In other words, the sum in (1) has the form $\approx f(c)n$ w.h.p.

Theorem 1.3. *(a) There exists a function $f(c)$ such that for any $\epsilon > 0$, there exists n_ε such that for $n \geq n_\varepsilon$,*

$$\left| \frac{\mathbf{E}[L_n]}{n} - f(c) \right| \leq \epsilon. \tag{7}$$

(b)

$$\frac{L_n}{n} \to f(c) \ a.s.$$

We will prove Theorem 1.3 in Section 3.

1.1 Structure of Γ_L:

We first bound the size of S_L. We need the following lemma on the density of small sets.

Lemma 1.4. *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

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

□

Now consider the construction of S_L. Let S_0 consist of the vertices with degree at most $D = 18$ that appear in the sequence. If we start with this S_0 and run the process then we will achieve the same S_L as in the given version of the process. Now w.h.p. there are at most $n_D = \frac{2c^D e^{-c}}{D!} n$ vertices of degree at most D in $G_{n,p}$, (see for example Theorem 3.3 of [15]) and so $|S_1| \leq 3n_D$. Now suppose that the process continues for another k rounds. Then S_{k+1} has at least $kD/2$ edges and at most $3n_D + 3k$ vertices. This is because each round adds a vertex v and at the end of the round i the neighbors of v are in S_i. Suppose k reaches n_D before the process stops. Then $e(S_{k+1})/|S_{k+1}| \geq Dn_D/(12n_D) = 3/2$. But, $6n_D < n_0$, contradicting Lemma 1.4. So, we can assert that w.h.p.

$$|V(\Gamma_L)| \leq 6n_D \leq ne^{-c/2}. \tag{8}$$

We note the following properties of S_L. Let

$$V_2 = \{v \in V(\Gamma) \subseteq S_L : \ v \text{ has at least one neighbor in } V_1\} \text{ and } V_1 = C_2 \setminus S_L$$

and for $T \in \mathcal{T}$ we let

$$v_0(T) = V(T) \setminus V_2.$$

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

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

We will now show that each component K of Γ satisfies

$$|v_0(K)| \geq \frac{|V(K)|}{3}. \tag{9}$$

$S_0 = \emptyset$ and so (9) is satisfied by every component spanned by S_0. Suppose that at step ℓ (9) is satisfied by every component spanned by S_ℓ. At step $\ell + 1$, if Case a is invoked, $v \in K$

and K' is the new component, then $|K'| \le |K| + 3$ and $v_0(K)$ increases by at least one and so (9) continues to hold, because

$$v_0(K') \ge v_0(K) + 1 \ge (|K| + 3)/3 \ge |K'|/3.$$

Adding v in Case b could merge components $K_1, K_2, \dots, K_r$ into one component K' while adding at most 3 vertices. Hence $3 + \sum_{i=1}^r |K_i| \ge |K'|$ and so

$$v_0(K') \ge 1 + v_0(K) \ge 1 + \frac{1}{3}\sum_{i=1}^{r} |K_i| \ge 1 + \frac{|K'| - 3}{3} = \frac{|K'|}{3}.$$

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

We next show that w.h.p., only a small component can satisfy (9). The expected number of components of size $k \le ne^{-c/2}$ that satisfy this condition is at most

$$\begin{aligned}\binom{n}{k} k^{k-2}\left(\frac{c}{n}\right)^{k-1}\binom{k}{k/3}(1-p)^{k(n-k)/3} &\le \left(\frac{ne}{k}\right)^k k^{k-2}\left(\frac{c}{n}\right)^{k-1} 2^k e^{-ck/6} \\ &\le \frac{n}{ck^2}\left(2ce^{1-c/6}\right)^k = o(n^{-2}), \end{aligned} \tag{10}$$

if c is large and $k \ge \log n$.

So, we can assume that all components are of size at most $\log n$. Then the expected number of vertices on components that are not trees is bounded by

$$\begin{aligned}\sum_{k=3}^{\log n}\binom{n}{k} k^{k+1}\left(\frac{c}{n}\right)^{k}\binom{k}{k/6}(1-p)^{k(n-k)/6} &\le \sum_{k=3}^{\log n}\left(\frac{ne}{k}\right)^k k^{k+1}\left(\frac{c}{n}\right)^{k} (e^{-ck/9}) \\ &\le \sum_{k=3}^{\log n} k\left(2ce^{1-c/9}\right)^k = O(1).\end{aligned}$$

Markov's inequality implies that whp such components span at most $\log n = o(n)$ vertices.

Notation 2: For $T \in \mathcal{T}$, let M_T be the matching obtained by replacing each path of $\mathcal{Q}(T)$ by an edge and let $M^* = \bigcup_{T\in\mathcal{T}} M_T$. 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 $\Gamma^* = \Gamma_1^* \cup \Gamma_2^* \cup M^*$ and $V^* = V_1 \cup V_2 = V(\Gamma^*)$.

2 Proof of Theorem 1.1

The RHS of (1), modulo the $o(n)$ number of vertices that are spanned by non tree components in Γ_L, is clearly an upper bound on the largest cycle in C_2. Any cycle must omit at least $\phi(T)$ vertices from each $T \in \mathcal{T}$. On the other hand, as we show, w.h.p. there is cycle H that spans $V_1 \cup \bigcup_{T\in\mathcal{T}} V(\mathcal{Q}(T))$ (see Notation 1). The length of H is equal to the RHS of (1). Equivalently, we show that

$$\text{w.h.p. there is a Hamilton cycle } H^* \text{ in } \Gamma^* \text{ that contains all the edges of } M^*. \tag{11}$$

2.1 Proof of (5)

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

We work in $G_{n,p}$. Observe that T must have a vertex of degree three in order that $\phi(T)>0$. The smallest such tree has seven vertices and consists of three paths of length two with a common vertex. Therefore, in $G_{n,p}$,

$$\begin{aligned}\mathbf{E}\left(\sum_{T\in\mathcal{T}}\phi(T)\right) &\leq_{\mathcal{O}} \sum_{k\geq 7}\binom{n}{k}p^{k-1}(1-p)^{(n-k)\max\{3,k/3\}} \\ &\leq_{\mathcal{O}} \sum_{k\geq 7}\left(\frac{ne}{k}\right)^{k}\left(\frac{c}{n}\right)^{k-1}\exp\left\{-c\max\left\{3,k/3\right\}\right\} \\ &= O(c^6e^{-3c})n, \end{aligned} \tag{12}$$

We obtain (5) from (12).

2.2 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_L. 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

$$N\geq n(1-2e^{-c/2}) \text{ and } M\in\frac{(1\pm\varepsilon_1)cN}{2}, \tag{13}$$

where $\varepsilon_1=c^{-1/3}$. The bound on N follows from (2) and (8) and the bound on M follows from the fact that in $G_{n,p}$,

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

2.3 Partitioning/Coloring $G=G_{n,p}$

We will use the edge coloring argument of Fenner and Frieze [10] to verify (11). In this section we describe how to color edges.

We color most of the edges of G light blue, dark blue or green. We denote the resultant blue and green subgraphs 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.

Every vertex $v \in V_1$ independently chooses 3 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. 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) 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 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 G' has the same set of extenstions to $G_{n,p}$ and as such is equally likely.

2.4 Expansion of Γ_b^*

We wish to estimate the probability that small sets have relatively few neighbors in the graph Γ_b^*. For $S \subseteq V^*$ we let $N_b(S) = \{w \in V^* \setminus S : \exists v \in S, \{v, w\} \in E(\Gamma_b^*)\}$.

It is known that for a graph with minimum degree at least three that a set of endpoints S obtained by rotations, that

$$S \cup N(S) \text{ contains at least } |S| + |N(S)| + 1 \text{ edges with an endpoint in } S, \tag{14}$$

see for example Lemma 5 of [9].

Lemma 2.1. *W.h.p. there does not exist $S \subset V^*$ of size $|S| \leq n/4$ such that $|N_b(S)| \leq 2|S|$ and $S : (N_b(S) \cap V_1)$ contains at least $|S| + |N_b(S) \cap V_1| + 1$ edges in Γ_b^*.*

Proof. Assume that the above fails for some set S. The particular values for the sets V_1, V_2 conditions $G_{n,p}$. To get round 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) and only occurs with probability $o(1)$. This event involves an arbitrary choice for V_1, V_2 etc.

Let $T = V^* \cap N_b(S)$ and $W = N_b(S) \setminus V^* \subseteq S_L$, that is T and W is the neighborhood of S inside and outside of V^* 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) $s + t \leq 2s$.

(ii) $S \cup T$ spans $s + t + 1$ edges in G and every vertex in W is connected to a vertex in S by a dark blue edge.

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

Thus,

$$\begin{aligned}
&\mathbf{Pr}(\mathcal{E}_S \mid s, t, w) \\
&\leq \binom{n}{s}\binom{n}{t}\binom{n}{w}\binom{\binom{s+t}{2}}{s+t+1} s^w p^{s+t+w+1}\left(1 - \frac{p}{2000}\right)^{s(n-s-t-w)} \\
&\leq \left(\frac{en}{s}\right)^s \left(\frac{en}{t}\right)^t \left(\frac{en}{w}\right)^w \left(\frac{e(s+t)}{2}\right)^{s+t+1} s^w \left(\frac{c}{n}\right)^{s+t+w+1} \exp\left\{-\frac{p}{2000}\left(\frac{sn}{4}\right)\right\} \\
&\leq (ec)^{2(s+t)} \left(\frac{s+t}{2s}\right)^s \left(\frac{s+t}{2t}\right)^t \left(\frac{ecs}{w}\right)^w \left(\frac{ec(s+t)}{2n}\right) \exp\left\{-\frac{cs}{10^5}\right\} \\
&\leq (ec)^{6s} \exp\left\{s \cdot \frac{t-s}{2s}\right\} \exp\left\{t \cdot \frac{s-t}{2t}\right\} \left(\frac{ecs}{2s}\right)^{2s} \left(\frac{3ecs}{2n}\right) \exp\left\{-\frac{cs}{10^5}\right\} \\
&= (ec)^{6s} \left(\frac{ec}{2}\right)^{2s} \left(\frac{3ecs}{2n}\right) \exp\left\{-\frac{cs}{10^5}\right\} = \left((ec)^6 \left(\frac{ec}{2}\right)^2 e^{-\frac{c}{10^5}}\right)^s \cdot \left(\frac{3ecs}{2n}\right).
\end{aligned}$$

At the 5th line we used that $t + w \leq 2s$, thus $t, w \leq 2s$. Hence

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

□

2.5 The Degrees of the Green Subgraph

Lemma 2.2. *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 $6n$ 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 probability that a vertex has degree less than $c/4$ is bounded by $2\frac{e^{-c}\lambda^{c/4}}{c/4!} < 1/1000$. 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. Therefore every set of $n/100$ vertices spans at least $[(n/100 - n/1000)c/4]/2 > (1+\epsilon)cn/4000 + 6n + c/50 \cdot n/100$ edges. Thus in every set of vertices of size at least $n/100$ there is 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

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

□

2.6 Posá Rotations

We say that a path/cycle P in Γ^* is *compatible* if for every $\{v,w\} \in M^*$ and $V(P) \cap \{v,w\} \neq \emptyset$ implies that P contains the edge $\{v,w\}$. We are thus going to show that w.h.p. Γ^* contains a compatible hamilton cycle.

Suppose that Γ^* and hence Γ_b^* is not Hamiltonian and that $P = (v_1, v_2, \ldots, v_s)$ is a longest compatible path in both Γ^* and Γ_b^*. If $\{v_s, v_i\} \in E_b^* \setminus M^*$ then the path $(v_1, v_2, \ldots, v_i, v_s, v_{s-1}, \ldots, v_{i+1})$ is said to be obtained from P by an *acceptable* rotation with v_1 as the fixed endpoint. Let $END_b^*(P, v_1)$ be the set of endpoints of paths 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 acceptable rotations.

Arguing as in the proof of Posá's lemma we see that $|N_b(END^*(P, v_1))| \leq 2|END_b^*(P, v_1)|$. So, from Lemma 2.1 we see that w.h.p. $|END_b^*(P_v, v)| \geq N_0$ for all $v \in END_b^*(P, v_1)$.

We let

$$END_b^*(P) = END_b^*(P, v_1) \cup \bigcup_{v \in END^*(P, v_1)} END_b^*(P_v, v).$$

2.7 Coloring argument

We use a modification of a double counting argument that was first used in [10]. The specific version is from [11]. Given a two-colored Γ^*, we choose for each $v \in V_1$, an additional incident edge $\xi_v = \{v, \eta_v\}$ where $\eta_v \in V_1 \cup V_2$. We re-color ξ_v blue if necessary. There are at most $\Pi = \prod_{v \in V_1} d(v)$ choices for $\boldsymbol{\xi} = (\xi_v, v \in V_1)$.

For a graph Γ, $\Gamma = \Gamma^*$ or Γ_b^*, we let $\ell(\Gamma)$ denote the length of the longest compatible path in Γ. We indicate that Γ has a compatible Hamilton cycle by $\ell(\Gamma) = N$.

We now let $a(\boldsymbol{\xi}, \Gamma_g^*) = 1$ if

H1 Γ_b^* is not Hamiltonian.

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

H3 $|N_b(S)| \geq 2|S|$ for all $S \subseteq V(\Gamma^*), |S| \leq n/4$.

We observe first that if Γ^* is not Hamiltonian and H2 holds then there exists $\boldsymbol{\xi}$ such that $a(\boldsymbol{\xi}, \Gamma_g^*) = 1$. 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 $1 \leq i < r$. It follows that if Φ denotes the number of choices for Γ_g^* and $\pi_{\bar{H}}$ is the probability that Γ^* is not Hamiltonian, then

$$\pi_{\bar{H}} \leq \frac{\sum_{\boldsymbol{\xi}, \Gamma_g^*} a(\boldsymbol{\xi}, \Gamma_g^*)}{\Phi} + o(1), \tag{15}$$

where the $o(1)$ term accounts for failure of the high probability events that we have identified so far.

On the other hand Γ_g^* is a random graph over all the graphs with degree sequence D_g^*. Hence

$$\sum_{\boldsymbol{\xi}, G_g^*} a(\boldsymbol{\xi}, \Gamma_g^*) \leq \Phi \Pi \max_{\Gamma_g^*} \pi_g, \tag{16}$$

where π_g is defined as follows: let P be some longest path in Γ_b^*. Then π_g is the probability that a random realization of Γ_g^* does not include a pair $\{x, y\}$ where $y \in END_b^*(P, x)$. We will argue below that

$$\max_{E_b} \pi_g \leq O(1) \times \prod_{v \in END_b^*(P)} \left(1 - \frac{d_{\Gamma_g^*}(v) \sum_{w \in END_b^*(P_v, v)} d_{\Gamma_g^*}(w)}{2 \times 2M} \right) \tag{17}$$

$$\leq O(1) \times \exp \left\{ - \frac{\sum_{v \in END_b^*(P)} d_{\Gamma_g^*}(v) \sum_{w \in END_b^*(P_v, v)} d_{\Gamma_g^*}(w)}{4M} \right\}. \tag{18}$$

The extra factor $2\times$ accounts for the cases where $w \in END_b^*(P_v, v)$ and $v \in END_b^*(P_w, w)$. Lemma 2.2 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 (18), it follows that

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

The Arithmetic-Geometric-mean inequality implies that

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

It then follows that

$$\pi_H \leq \frac{e^{-cn/10^5}}{(2c)^n} + o(1) = o(1),$$

and completes the proof of (11).

Proof of (17)**:** This is an exercise in the use of the configuration model of Bollobás [4]. 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 [15]. Note that **P2** uses the fact that w.h.p. $G_{V_1,M}^{\delta\geq 3}$ (and hence Γ_g^*) has an exponential tail, as shown for example in [13]. But, given all this, in the context of the configuration model, (17) is a simple consequence of a random pairing of W. The $O(1)$ factor is $1/\,\mathbf{Pr}(\gamma(F) \text{ is simple})$ and bounds the effect of the conditioning.

3 Proof of Theorem 1.3

For $v \in C_2$ we let $\phi(v) = \phi(T)/|v_0(T)|$ if $v \in v_0(T)$ for some $T \in \mathcal{T}$ and $\phi(v) = 0$ otherwise. Thus

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

Hence (1) can be rewritten as,

$$L_n \approx |C_2| - \sum_{v \in C_2} \phi(v). \tag{19}$$

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

$$k_1 \leq \frac{2}{c} \log \frac{1}{\varepsilon}. \tag{20}$$

For $v \in C_2$ let G_v be the graph consisting of (i) the vertices of G that are within distance k_1 from v and (ii) a copy of $K_{3,3}$ where every vertex in the k_1 neighborhood of v is adjacent to each vertex of the same one part of the bipartition. 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 $v \in C_2$, if v belongs to some tree $T \in S_{L,v}$ set $f(v) = f(T)/\upsilon_{0,v}(T)$. Else 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 $|\phi(v) - f(v)| \leq 1$.

By repeating the arguments used to prove (10) and (9) it follows that if $t(v) = 0$ then v lies on a component C of size at most $\log n$. In addition at least $|V(C)|/3$ vertices in $V(C)$ are not adjacent to any vertex outside $V(C)$. Thus the expected number of vertices v satisfying $t(v) = 0$ is bounded by

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

A vertex $v \in [n]$ is *good* if the ith level of its BFS neighborhood has size at most $3c^i k_1/\epsilon$ for every $i \leq k_1$ and it is *bad* otherwise. Because the expected size of the i^{th} neighborhood is $\approx c^i$ we have by the Markov inequality that v is bad with probability at most $\approx \varepsilon/3k_1$ and so the expected number of bad vertices is bounded by $\varepsilon n/2$. Thus

$$\begin{aligned}
\mathbf{E}\left(\left|\sum_{v \in V} \phi(v) - \sum_{v \text{ is good}} f(v)\right|\right) &\leq \mathbf{E}\left(\left|\sum_{v \in V} \phi(v) - \sum_{v \in V} f(v)\right|\right) + \mathbf{E}\left(\left|\sum_{v \text{ is bad}} f(v)\right|\right) \\
&\leq \mathbf{E}\left(\left|\sum_{v: t(v)=0} |\phi(v) - f(v)|\right|\right) + \mathbf{E}\left(\sum_{v \text{ is bad}} 1\right) \\
&\leq \mathbf{E}\left(\sum_{v: t(v)=0} 1\right) + \frac{\epsilon n}{3}
\end{aligned}$$

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

Let $\mathcal{H}_\varepsilon$ be the set of BFS neighborhoods that are good i.e. whose ith levels are of size at most $3c^i k_1/\epsilon$ for every $i \leq k_1$. Every element of $\mathcal{H}_\varepsilon$ corresponds to a pair (H, o_H) where H is a graph and o is a distinguished vertex of H, that is considered to be the root. Also for $v \in C_2$ let $G(N_{k_1}(v))$ be the subgraph induced by the ${k_1}^{th}$ neighborhood of v. For $(H, o_H) \in \mathcal{H}_\varepsilon$ let $int(H)$ be the set of vertices incident to the first $k_1 - 1$ neighborhoods of o_H and let $Aut(H, o_H)$ be the number of automorphisms of H that fix o_H. Note that each good vertex v is associated with a pair $(H, o_H) \in \mathcal{H}_\varepsilon$ from which we can compute $f(v)$, since $f(v) = f(o_H)$. Thus, if now $M = |E(C_2)|, N = |C_2|$,

$$\begin{aligned}
\mathbf{E}\left(\sum_{v \text{ is good}} f(v) \middle| M, N\right) &= \sum_v \sum_{k\geq 1} \sum_{\substack{(H,o_H)\in\mathcal{H}_\varepsilon \\ (G(N_{k_1}(v)),v)=(H,o_H) \\ |V(H)|=k}} \rho_{H,o_H} f(o_H) \\
&= o(n) + \sum_v \sum_{k\geq 1} \sum_{\substack{(H,o_H)\in\mathcal{H}_\varepsilon \\ H \text{ is a tree} \\ (G(N_{k_1}(v)),v)=(H,o_H)}} \rho_{H,o_H} f(o_H), \qquad (21)
\end{aligned}$$

where ρ_{H,σ_H} is the probability $(G(N_{k_1}(v)), v) = (H, o_H)$ in C_2. We show in Section 3.1 that

$$\rho_{H,o_H} \approx \frac{1}{Aut(H, o_H)} \left(\frac{N}{2M}\right)^{k-1} \lambda^{2k-2} \frac{f_2(k\lambda)}{f_2(\lambda)^k}, \qquad (22)$$

where f_k is defined in (25) below and λ satisfies (26) below.

Finally observe that with the exception of the $o(1)$ term, all the terms in (21) are independent of n. We let

$$f_\varepsilon(c) = \sum_{k\geq 1} \sum_{\substack{(H,o_H)\in\mathcal{H}_\varepsilon \\ H \text{ is a tree}}} \frac{f(o_H)}{Aut(H, o_H)} \left(\frac{N}{2M}\right)^{k-1} \lambda^{2k-2} \frac{f_2(k\lambda)}{f_2(\lambda)^k}. \qquad (23)$$

Then for a fixed c, we see that $f_\varepsilon(c)$ 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. This verifies part (a) of Theorem 1.3. For part (b), we prove, (see (38)),

Lemma 3.1.
$$\mathbf{Pr}(|L_n - \mathbf{E}(L_n)| \geq \varepsilon n + n^{3/4}) = O(e^{-\Omega(n^{1/5})}).$$

Proof. To prove this we show that if $\nu(H)$ is the number of copies of H in C_2 then $H \in \mathcal{H}_\varepsilon$ implies that

$$\mathbf{Pr}(|\nu(H) - \mathbf{E}(\nu(H))| \geq n^{3/5}) = O(e^{-\Omega(n^{1/5})}). \qquad (24)$$

The inequality follows from a version of Azuma's inequality (see (38)), and the lemma follows from taking a union bound over

$$\exp\left\{O\left(\frac{c^{k_1(\epsilon)}k_1(\epsilon)}{\epsilon}\right)\right\} = \exp\left\{O\left(\frac{c^{\frac{2\log\frac{1}{\varepsilon}}{c}}\frac{2\log\frac{1}{\varepsilon}}{c}}{\varepsilon}\right)\right\}$$
$$= \exp\left\{O\left(\frac{(1/\varepsilon)^{2\log c/c}\log\frac{1}{\varepsilon}}{c\varepsilon}\right)\right\} = \exp\left\{O((1/\varepsilon)^{2+2\log c/c})\right\}$$

graphs H. Note also that the $o(n)$ term in (21) is bounded by the same $e^{O((1/\varepsilon)^{2+2\log c/c})}$ term times the number of cycles of length at most $2k_1$ in G. The probability that this exceeds $n^{1/2}$ is certainly at most the RHS of (24). We will give details of our use of the Azuma inequality in Section 3.1. □

Part (b) of Theorem 1.3 follows by letting $\varepsilon \to 0$ and from the Borel-Cantelli lemma.

3.1 A Model of C_2

It is known that given M, N that, up to relabeling vetices, C_2 is distributed as $G_{N,M}^{\delta\geq 2}$. The random graph $G_{N,M}^{\delta\geq 2}$ is chosen uniformly from $\mathcal{G}_{N,M}^{\delta\geq 2}$ which is the set of graphs with vertex set $[N]$, M edges and minimum degree at least two.

3.1.1 Random Sequence Model

We must now take some time to explain the model we use for $G_{N,M}^{\delta\geq 2}$. We use a variation on the pseudo-graph model of Bollobás and Frieze [6] and Chvátal [7]. 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

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

Let $G_{\mathbf{x}}$ be the multi-graph $G_{\mathbf{x}}$ for $\mathbf{x}$ chosen uniformly from $[N]_{\delta\geq 2}^{2M}$. It is clear then that conditional on being simple, $G_{\mathbf{x}}$ has the same distribution as $G_{N,M}^{\delta\geq 2}$. 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

$$f_k(\lambda) = e^{\lambda} - \sum_{i=0}^{k-1} \frac{\lambda^i}{i!} \tag{25}$$

for $k \geq 0$.

Lemma 3.2. *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*

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

Here λ satisfies

$$\frac{\lambda f_1(\lambda)}{f_2(\lambda)} = \frac{2M}{N}. \tag{26}$$

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 (13) and (26) and the fact that $f_1(\lambda)/f_2(\lambda) \to 1$ as $c \to \infty$ that for large c,

$$\lambda = c\left(1 + O(ce^{-c})\right). \tag{27}$$

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,

$$\mathbf{Pr}\left(\sum_{j=1}^{N} Z_j = 2M\right) = \frac{1}{\sigma\sqrt{2\pi N}}(1 + O(N^{-1}\sigma^{-2})) \tag{28}$$

and

$$\mathbf{Pr}\left(\sum_{j=2}^{N} Z_j = 2M - d\right) = \frac{1}{\sigma\sqrt{2\pi N}}\left(1 + O((d^2+1)N^{-1}\sigma^{-2})\right). \tag{29}$$

This is an example of a local central limit theorem. See for example, (5) of [2] or (3) of [13]. It follows by repeated application of (28) and (29) that if $k = O(1)$ and $d_1^2 + \cdots + d_k^2 = o(N)$ then

$$\mathbf{Pr}\left(Z_i = d_i, i = 1, 2, \ldots, k \mid \sum_{j=1}^{N} Z_j = 2M\right) \approx \prod_{i=1}^{k} \frac{\lambda^{d_i}}{d_i! f_2(\lambda)}. \tag{30}$$

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

Lemma 3.3. *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 $1 - o(N^{-10})$,*

$$\left|\nu_{\mathbf{x}}(j) - \frac{N\lambda^j}{j!f_2(\lambda)}\right| \leq \left(1 + \left(\frac{N\lambda^j}{j!f(\lambda)}\right)^{1/2}\right)\log^2 N,\ 2 \leq j \leq \log N. \tag{31}$$

$$\nu_{\mathbf{x}}(j) = 0, \quad j \geq \log N. \tag{32}$$

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

$$\mathbf{Pr}(G_{\mathbf{x}} \text{ is simple}) = \Omega(1). \tag{33}$$

Again, this follows as in [2].

Given a tree H with k vertices of degrees $z_1, z_2, ..., z_k$ 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 where $\Phi(2m) = \frac{(2m)!}{m!2^m}$, we have

$$\rho_H \approx \binom{N}{k-1}\frac{(k-1)!}{Aut(H,o_H)}\sum_{D=2k}^{\infty}\sum_{\substack{d_1,\ldots,d_k\\ d_1+\cdots+d_k=D}}\prod_{i=1}^{k}\frac{\lambda^{d_i}}{d_i!f_2(\lambda)}\cdot\prod_{i=1}^{k}\frac{d_i!}{(d_i-z_i)!}\frac{\Phi(2M-2k+2)}{\Phi(2M)} \tag{34}$$

$$= \binom{N}{k-1}\frac{(k-1)!}{Aut(H,o_H)}\lambda^{2k-2}\sum_{D=2k}^{\infty}\sum_{\substack{d_1,\ldots,d_k\\ d_1+\cdots+d_k=D}}\prod_{i=1}^{k}\frac{\lambda^{d_i-z_i}}{(d_i-z_i)!f_2(\lambda)}\frac{\Phi(2M-2k+2)}{\Phi(2M)} \tag{35}$$

$$= \binom{N}{k-1}\frac{(k-1)!}{Aut(H,o_H)}\lambda^{2k-2}\frac{\Phi(2M-2k+2)}{\Phi(2M)f_2(\lambda)^k}\sum_{D=2k}^{\infty}\frac{(k\lambda)^{D-2(k-1)}}{(D-2(k-1)!)} \tag{36}$$

$$\approx \frac{1}{Aut(H,o_H)}\left(\frac{N}{2M}\right)^{k-1}\lambda^{2k-2}\frac{f_2(k\lambda)}{f_2(\lambda)^k}. \tag{37}$$

Explanation for (34)**:** We use (30) to obtain the probability that the degrees of $[k]$ are $d_1, \ldots, d_k$. Implicit here is that $d_i = O(\log n)$, from (32). The contribution to the sum of $D \geq 2k\log n$ can therefore be shown to be negligible. Having fixed $d_1, \ldots, d_k$ we can condition on $d_{k+1}, \ldots, d_N$ and then we essentially are dealing with the configuration model. In which case $\Phi(2M)$ is the total number of pairings of all points and $\Phi(2M-k)$ is the number of pairings, given we have H occuring in $[k]$. We then use the fact that k is small to argue that w.h.p. H is induced.

Explanation for (36)**:** We use the identity

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

It only remains to verify (24). It follows from the above that $\mathbf{E}(\nu(H) \mid M, N) = \Omega(N)$. We first condition on a degree sequence $\mathbf{x}$ satisfying (31). We then work in the associated configuration model. We can generate a configuration F as a permutation of the multi-set

$\{d_i \times i : i \in [N]\}$. Interchanging two elements in a permutation can only change $\nu(H)$ by $O(1)$. We can therefore apply Azuma's inequality to show that

$$\mathbf{Pr}(|\nu(H) - \mathbf{E}(\nu(H))| \geq n^{3/5}) = O(e^{-\Omega(n^{1/5})}). \tag{38}$$

(Specifically we can use Lemma 11 of Frieze and Pittel [17] or Section 3.2 of McDiarmid [19].) This verifies (24).

4 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 [14]. It would be more interesting to do the analysis for small $c > 1$. Applying the coupling of McDiarmid [18] 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 (6). It should be possible to improve this, just as Krivelevich, Lubetzky and Sudakov [16] did for the earlier result of [12].

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

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

[5] 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.

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

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

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

[9] 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.

[10] 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.

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

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

[13] A.M. Frieze, 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.

[14] 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.

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

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

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

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

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

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

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