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.

October 2, 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 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 $p_k$ 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

Let $L_{c,n}$ denote the length of the longest cycle in the random graph $G_{n,c/n}$. Erdős [8] 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 [22] who proved that if $c > 4\log 2$ then $f(c) = 1 - O(c^{-1})$. See also Suen [21]. 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 $\ell(c) \geq 1 - c^{24}e^{-c/2}$. This was then improved by Bollobás, Fenner and Frieze [5] to $\ell(c) \geq 1 - 6^c e^{-c}$ and then by Frieze [13] to $\ell(c) \geq 1 - (1 + \varepsilon_c)(1 + c)e^{-c}$ where $\varepsilon_c \to 0$ as $c \to \infty$. This last result is optimal up to the value of $\varepsilon_c$, 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
The fundamental 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.

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 in \( G_{n,c/n} \) differs from this by \( O(\log n) \) w.h.p. This will be for two reasons. The first reason is that we will establish a Hamiltonian subgraph of \( C_2 \) that contains the longest path in \( C_2 \) and the second 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 [13] 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 \( S_{\ell} \), \( \ell \geq 0 \). We construct \( S_{\ell+1} \) from \( S_\ell \) via one of two cases: let

\[
c_0 = 10.
\]

**Construction of \( \Gamma_L \)**

**Case a:** If there is \( v \in S_\ell \) that has at least one but fewer than \( c_0 \) neighbors \( W \) in \( C_2 \setminus S_\ell \), then we add \( W \) to \( S_\ell \) to make \( S_{\ell+1} \).

**Case b:** If there is a vertex \( v \in C_2 \setminus S_\ell \) that has at most \( c_0 - 1 \) neighbors in \( C_2 \setminus S_\ell \) then we define \( S_{\ell+1} \) to be \( S_\ell \) plus \( v \) plus the neighbors of \( v \) in \( C_2 \setminus S_\ell \).

Note that we allow \( d < c_0 \) here and so low degree vertices are always added to some \( S_\ell \).

\( S_L \) 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 \( \Gamma_L \) induced by \( S_L \) is a forest plus a few small components. Each tree in \( \Gamma_L \) 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 \( T \) denote the set of trees in \( \Gamma_L \). For a tree \( T \in T \) let \( \mathcal{P}_T \) be the set of vertex disjoint path packings of \( T \) where we allow only paths whose start- and end- vertex
are incident to $C_2 \setminus V(T)$. Here we allow paths of length 0, so that a single vertex with neighbors in $C_2 \setminus V(T)$ 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 $Q(T) \in \mathcal{P}$ 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 $c > 1$ is a sufficiently large constant. Then w.h.p.

$$L_{c,n} \approx |V(C_2)| - \sum_{T \in \mathcal{T}} \phi(T).$$

(1)

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. [16], Lemma 2.16),

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

(2)

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

(3)

Equation (4.5) of Erdős and Rényi [9] tells us that

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

(4)

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

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

(5)

We therefore have the following improvement to the estimate in [13].

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

$$L_{c,n} \approx (1 - (c + 1)e^{-c} - c^2 e^{-2c} + O(c^3 e^{-3c})) n.$$  

(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 $n_\epsilon$ such that for $n \geq n_\epsilon$,

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

(7)

(b) 

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

We will prove Theorem 1.3 in Section 3.
1.1 Structure of $\Gamma_L$:

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 $n_0 = n/10e^3$ contains less than $3|S|/2$ edges in $G_{n,p}$.

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

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

Now consider the construction of $S_L$. Let $A$ be the set of the vertices with degree at most $D = 59$ that belong to $S_L$ and $S'_0 = A \cup (N(A) \cap S_L)$. Then $S'_0 \subseteq S_L$. If we start with $S_0 = S'_0$ instead of $S_0 = \emptyset$ and run the process for constructing $\Gamma_L$ then we will achieve the same $S_L$ as in the given version of the process. In this alternating version consider marking red every vertex that invokes Case a or Case b. If the algorithm runs for $k$ additional step then $S_k$ is of size $|S'_0| + k$ and it contains $k$ red vertices. Observe that all the neighbors of any such red vertex lie in $S_k$. In addition no vertex in $A$ will invoke Case a or Case b, hence every red vertex has degree at least $D + 1$ and $S_k$ spans at least $k(D + 1)/2$ edges.

Now w.h.p. there are at most $n_D = \frac{2e^D e^{-e}}{D^e} n$ vertices of degree at most $D$ in $G_{n,p}$, (see for example Theorem 3.3 of [16]) and so $|S'_0| \leq (D + 1)n_D$. Now suppose that the process runs for at least $7n_D$ rounds. Then $S_{7n_D}$ is well defined. From the above $e(S_{7n_D}) \geq (D + 1)7n_D/2 = 210n_D$. Moreover at step $1 \leq i \leq 7n_D$ at most $c_0$ vertices are added to $S_i$ thus $|S_{7n_D}| \leq |S'_0| + c_0 \cdot 7n_D \leq (D + 1)n_d + 7c_0n_D = 130n_D$. Thus

$$\frac{e(S_{7n_D})}{|S_{7n_D}|} \geq \frac{3}{2}.$$

But, $|S_{7n_D}| \leq 67c_0n_D < n_0$, contradicting Lemma 1.4. So, we can assert that w.h.p.

$$|V(\Gamma_L)| \leq 130n_D \leq ne^{-c/2}. \quad (8)$$

We note the following properties of $S_L$. Let

$$V_2 = \{ v \in S_L : v \text{ has at least one neighbor in } V_1 \} \text{ and } V_1 = C_2 \setminus S_L.$$

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 $c_0$ neighbors in $V_1$. 

4
Given the definition of $V_2$, for $T \in T$ we can express $v_0(T)$ as

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

We will now show that each component $K$ of $\Gamma_L$ satisfies

$$|v_0(K)| \geq \frac{|V(K)|}{c_0}. \tag{9}$$

We will prove that for $0 \leq i \leq L$ and each component $K$ spanned by $S_i$,

$$|v_{0,i}(K)| \geq \frac{|V(K)|}{c_0}. \tag{10}$$

Here $v_{0,i}(K)$ is taken to be the number of vertices in $V(K)$ with no neighbors in $C_2 \setminus K$. Taking $i = L$ in (10) yields (9). We proceed by an induction on $i$.

$S_0 = \emptyset$ and so for $i = 0$, (10) is satisfied by every component spanned by $S_0$. Suppose that at step $i = \ell$, (10) is satisfied by every component spanned by $S_\ell$. At step $\ell + 1$, if Case a is invoked, $v \in K$ and $K'$ is the new component, then $|K'| \leq |K| + c_0$ and $v_0(K)$ increases by at least one and so (9) continues to hold, because

$$v_0(K') \geq v_0(K) + 1 \geq (|K| + c_0)/c_0 \geq |K'|/c_0.$$

Adding $v$ in Case b could merge components $K_1, K_2, \ldots, K_r$ into one component $K'$ while adding at most $c_0$ vertices. Hence $c_0 + \sum_{i=1}^r |K_i| \geq |K'|$ and so

$$v_0(K') \geq 1 + v_0(K) \geq 1 + \frac{1}{c_0} \sum_{i=1}^r |K_i| \geq 1 + \frac{|K'| - c_0}{c_0} = \frac{|K'|}{c_0}.$$

and so (10) 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 $k \leq ne^{-c/2}$ that satisfy this condition is at most

$$\binom{n}{k} \frac{k^{k-2} (\frac{c}{n})^{k-1} (\frac{k}{k/c_0}) (1-p)^{k(n-k)/c_0}}{k} \leq \binom{ne}{k} \frac{k^{k-2} (\frac{c}{n})^{k-1} 2^k e^{-ck/2c_0}}{k} \leq \frac{n}{ck^2} (2ce^{1-c/2c_0})^k = o(n^{-2}), \tag{11}$$

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

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

$$\sum_{k=3}^{\log n} \binom{n}{k} \frac{k^{k+1} (\frac{c}{n})^k (\frac{k}{k/2c_0}) (1-p)^{k(n-k)/2c_0}}{k} \leq \sum_{k=3}^{\log n} \binom{ne}{k} \frac{k^{k+1} (\frac{c}{n})^k (e^{-ck/3c_0})}{k}.$$
\[
\leq \sum_{k=3}^{\log n} k \left(2ce^{1-c/3c_0}\right)^k = O(1).
\]

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

**Notation 2:** For \(T \in \mathcal{T}\), let \(M_T\) be the matching on \(V_2\) obtained by replacing each path of \(Q(T)\) of length at least 1 by an edge and let \(M^* = \bigcup_{T \in \mathcal{T}} M_T\). We let \(\Gamma_1^*\) be the subgraph of \(G\) induced by \(V_1\). We also let \(\Gamma_2^* = \bigcup_{T \in \mathcal{T}} M_T\). 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 \(\Gamma_L\), 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 \(V_1 \cup \bigcup_{T \in \mathcal{T}} V(Q(T))\) (see Notation 1). The length of \(H\) is equal to the RHS of (1). Equivalently, we show that

w.h.p. there is a Hamilton cycle \(H^*\) in \(\Gamma^*\) that contains all the edges of \(M^*\). \hfill (12)

2.1 Proof of (5)

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 (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 endpoint. (If \(T\) is a star of degree 3 for example, it can be covered by a path of length 2 that covers the central vertex and a path of a vertex 0. Here we are using that every vertex in \(V(T) \setminus V_2 \subset C_2\) must have degree at least 2, hence every vertex of \(T\) of degree 1 belongs to \(V_2\) and is incident to \(C_2 \setminus V(T)\).) Therefore, in \(G_{n,p}\),

\[
E \left(\sum_{T \in \mathcal{T}} \phi(T)\right) \leq \sum_{k \geq 7} k \cdot \binom{n}{k} k^{k-2} p^{k-1} (1 - p)^{(n-k)\max\{3,k/c_0\}}
\]

\[
\leq \sum_{k \geq 7} \left(\frac{ne}{k}\right)^k k^{k-1} \left(\frac{c}{n}\right)^{k-1} \exp \{-c \max\{3,k/c_0\}\}
\]

\[
= O(c^6 e^{-3c}) n, \hfill (13)
\]

At the first line we used that every tree that contributes to \(E \left(\sum_{T \in \mathcal{T}} \phi(T)\right)\) must satisfy \(v_0(T) > 2\). In addition (9) states that \(v_0(T) \geq |T|/c_0\). We obtain (5) from (13).
2.2 Structure of $\Gamma_1^*$

Suppose now that $|V_1| = N$ and that $V_1$ contains $M$ edges. The construction of $\Gamma_L$ does not involve the edges inside $V_1$, but we do know that $\Gamma_1^*$ has minimum degree at least $c_0$. The distribution of $\Gamma_1^*$ will be that of $G_{V_1, M}$ subject to this degree condition, viz. the random graph $G_{V_1, M}^d$ which is sampled uniformly from the set $G_{V_1, M}^d$, the set of graphs with vertex set $V_1$, $M$ edges and minimum degree at least $c_0$. This is because, we can replace $\Gamma_1^*$ by any graph in $G_{V_1, M}^d$ without changing $\Gamma_L$. By the same token, we also know that each $v \in V_2$ has at least $c_0$ random neighbors in $V_1$. We have that

$$N \geq n(1 - 2e^{-c/2})$$

and

$$M \leq \left(1 \pm \varepsilon_1\right) \frac{cN^2}{2},$$

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 $G_{n,p}$,

$$\Pr\left(\exists S : |S| = N, e(S) \not\in \left(1 \pm \varepsilon_1\right) \binom{N}{2} \right) \leq 2 \binom{n}{N} \exp\left\{-\frac{\varepsilon_1^2 N(N - 1)p}{3}\right\} = o(1).$$

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

We will use the edge coloring argument of Fenner and Frieze [11] to verify (12). 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 $\Gamma_b, \Gamma_g$ 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 $c_0$ 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 $c_0$ 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 $d_g$ of $\Gamma_g$, every graph $H$ with vertex set $V_1$ and degree sequence $d_g$ is equally likely to be $\Gamma_g$.

We can justify **C3** as follows: Amending $G$ by replacing $\Gamma_g$ 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 extensions to \(G_{n,p}\) and as such is equally likely.

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

### 2.4 Expansion of \(\Gamma_b^*\)

We wish to estimate the probability that small sets have relatively few neighbors in the graph \(\Gamma_b^*\). For \(S \subseteq V^*\) we let

\[
N_b(S) = \{w \in V_1 \setminus S : \exists v \in S \text{ with } \{v, w\} \in E(\Gamma_b^*)\} = \{w \in V_1 \setminus S : \exists v \in S \text{ with } \{v, w\} \in E(\Gamma_b)\}
\]

The second equality follows from the fact that \(M^*\) is spanned by \(V_2\) only. We have slightly abused notation here since \(N_b(S)\) is implicitly defined in both \(G\) and \(\Gamma^*\).

**Lemma 2.1.** W.h.p. there does not exist \(S \subset V^*\) of size \(|S| \leq n/4\) such that \(|N_b(S)| \leq 2|S|\).

**Proof.** Assume that the above fails for some set \(S\).

**Case 1:** \(|S| \leq n/(100c^3)\).

In this case \(S \cup N_b(S)\) has cardinality at most \(s + t \leq 3s\) and contains at least \(c_0s/2 > 3(s+t)/2\) edges, contradicting Lemma 1.4.

**Case 2:** \(n/(100c^3) < |S| \leq n/4\).

The particular values for the sets \(V_1, V_2\) condition \(G_{n,p}\). 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 \(T = N_b(S)\) and \(W = N_G(S) \setminus N_b(S)\), that is \(T\) and \(W\) is the neighborhood of \(S\) 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) \(t \leq 2s\).

(ii) \(w \leq n_0 = ne^{-c/2}\), where \(n_0\) is from (8).

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

(iv) \(S \cup N_b(S)\) spans at least \(c_0s/2 \geq s + t\) edges.
Thus,

\[
\Pr(\mathcal{E}_S \mid s, t, w) \\
\leq \binom{n}{s} \binom{n}{t} \binom{n}{w} \left( \frac{s+t}{2} w \right) \left( 1 - \frac{p}{2000} \right)^{s(n-s-t-w)} \\
\leq \left( \frac{en}{s} \right)^s \left( \frac{en}{t} \right)^t \left( \frac{en}{w} \right)^w \left( \frac{e(s+t)}{2t} \right)^{s+t} \left( \frac{c}{n} \right)^{s+t+w} \exp \left\{ - \frac{p}{2000} \left( \frac{sn}{5} \right) \right\} \\
\leq (ec)^{2(s+t)} \left( \frac{s+t}{2s} \right)^s \left( \frac{s+t}{2t} \right)^t \left( \frac{ec}{w} \right)^w \exp \left\{ - \frac{cs}{10^5} \right\} \\
\leq (ec)^{6s} \exp \left\{ s \cdot \frac{t-s}{2s} \right\} \exp \left\{ t \cdot \frac{s-t}{2t} \right\} \left( \frac{ec}{n_0} \right)^{n_0} \exp \left\{ - \frac{cs}{10^5} \right\} \\
\leq (ec)^{6s} (ce^{1-c/3})^{se^{-c/3}} \exp \left\{ - \frac{cs}{10^5} \right\} = \left( (ec)^6 (ce^{1-c/3})^{e^{-c/3}} e^{-c/10^5} \right)^s.
\]

At the 5th line we used that \( w \leq n_0 \leq 100c^3 e^{-c/2} s \leq e^{-c/3} s \). Hence

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

\[
2.5 \text{ The Degrees of the Green Subgraph}
\]

**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 \( 2c_0 n \) 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 \( \epsilon \). The probability that a vertex has degree less than \( c/4 \) is bounded by \( 2^{e^{-c\epsilon/c/4}} < 1/1000 \). 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 \( [(n/100 - n/1000)c/4]/2 \) edges, hence \( [(n/100 - n/1000)c/4]/2 - (1 + \epsilon)cn/4000 - 2c_0 n \geq c/50 \cdot n/100 \) 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

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

\[\square\]
2.6 Posá Rotations

We say that a path/cycle $P$ in $\Gamma^*$ is compatible if for every $\{v, w\} \in M^*$ either $P$ contains the edge $\{v, w\}$ or $V(P) \cap \{v, w\} = \emptyset$. We are thus going to show that w.h.p. $\Gamma^*$ contains a compatible hamilton cycle. Suppose that $\Gamma^*$ and hence $\Gamma_b^*$ is not Hamiltonian and that $P = (v_1, v_2, \ldots, v_s)$ is a longest compatible path in both $\Gamma^*$ and $\Gamma_b^*$. If $\{v_s, v_i\} \in E(\Gamma^*)$ 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 $v_i$ as the fixed endpoint. Observe that since $P$ is compatible and $\{v_i, v_{i+1}\} \notin M^*$ (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 $v_i$ as the fixed endpoint. Then, for $v \in END_b^*(P, v_1)$ we let $END_b^*(P, v)$ 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 $|N_b(\text{END}^*(P, v_1))| \leq 2|\text{END}^b_b(P, v_1)|$. Indeed, assume otherwise. Then there exist vertices $v_i, u$ such that $u \in END_b^*(P, v_1)$, $v_i \in N_b(u)$, $v_{i-1}, v_{i+1} \notin END_b^*(P, v_1)$ and the edge $\{u, v_i\}$ can be used by an acceptable rotation with $v_i$ 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 $v_{i-1}$. Now $v_i \notin V(M^*)$ and so the rotation will be acceptable and hence $v_{i-1} \in END_b^*(P, v_1)$ resulting in a contradiction. So, from Lemma 2.1 we see that w.h.p. $|END_b^*(P, v)| \geq n/4$ for all $v \in END_b^*(P, v_1)$.

We let

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

2.7 Coloring argument

We use a modification of a double counting argument that was first used in [11]. The specific version is from [12]. Given a two-colored $\Gamma^*$, 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 $\xi_v$ blue if necessary. There are at most $\Pi = \prod_{v \in V_1} d(v)$ choices for $\xi = (\xi_v, v \in V_1)$.

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

We now let $a(\xi, \Gamma_b^*) = 1$ if the following hold:

H1 $\Gamma_b^*$ is not Hamiltonian.

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

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

10
We observe first that if $\Gamma^*$ is not Hamiltonian and $H2$ holds then there exists $\xi$ such that $a(\xi, \Gamma^*_g) = 1$. Indeed, let $P = (v_1, v_2, \ldots, v_r)$ be a longest path in $\Gamma^*$. Then we simply let $\xi_{v_i}$ be the edge $\{v_i, v_{i+1}\}$ for $1 \leq i < r$. It follows that if $\Phi$ denotes the number of choices for $\Gamma^*_g$ and $\pi_H$ is the probability that $\Gamma^*$ is not Hamiltonian, then

$$
\pi_H \leq \frac{\sum_{\xi, \Gamma^*_g} a(\xi, \Gamma^*_g)}{\Phi} + o(1),
$$

(15)

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

On the other hand $\Gamma^*_g$ is a random graph over all the graphs with degree sequence $D^*_g$. Hence

$$
\sum_{\xi, \Gamma^*_g} a(\xi, \Gamma^*_g) \leq \Phi \Pi \max_{\Gamma^*_b} \pi_b,
$$

(16)

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

$$
\max_{\Gamma^*_b} \pi_b \leq O(1) \times \prod_{v \in END^*_b(P)} \left(1 - \frac{\sum_{w \in END^*_b(P, v)} d_{\Gamma^*_g}(w)}{2M}\right)^{\frac{1}{2}}
$$

(17)

$$
\leq O(1) \times \exp \left\{ -\frac{\sum_{v \in END^*_b(P)} d_{\Gamma^*_g}(v) \sum_{w \in END^*_b(P, v)} d_{\Gamma^*_g}(w)}{4M} \right\}.
$$

(18)

The extra $\frac{1}{2}$ factor at the exponent accounts for the cases where $w, u \in END^*_b(P)$, $w \in END^*_b(P, v)$ and $v \in END^*_b(P, w)$. 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 $END^*_b(P, v) \cup \{v\}$ is of size at least $n/4$ and so has total degree at least $cn/250$. Thus from (18), it follows that

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

The Arithmetic-Geometric-mean inequality implies that

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

11
It then follows that for sufficiently large $c$

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

and completes the proof of (12).

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 (multi)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 $G_1, G_2 \in \mathcal{G}_{d_g}$ then $\Pr(\gamma(F) = G_1 | \gamma(F) \text{ is simple}) = \Pr(\gamma(F) = G_2 | \gamma(F) \text{ is simple}).$

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

These are well established properties of the configuration model, see for example Chapter 11 of [16]. Note that P2 uses the fact that w.h.p. $G_{d_g}^{\geq 3}$ $(\text{and hence } \Gamma_g^*)$ has an exponential tail, as shown for example in [14]. 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 $1/\Pr(\gamma(F) \text{ is simple})$ 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,

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

(19)

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

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

Note that for large $c$, we have

$$k_1 \leq \frac{2}{c} \log \frac{1}{\epsilon}.$$  

(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 \( K_{c_0,c_0} \) 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 \( \Gamma_L \) on \( G_v \) and let \( C_{2,v}, \Gamma_v, V_{1,v}, V_{2,v}, S_{L,v}, v_0,v(T) \) 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 \( f(v) = f(T)/v_0,v(T) \). 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 \( \Gamma_L \), \( v \) 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 (11) 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 \( |V(C)|/c_0 \) 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

\[
\frac{\log^2 n}{n} \sum_{k=k_1-1}^{\log n} \sum_{j=k}^{n} c_0 k \left( \frac{e}{c_0 k} \right) ^{c_0 k} 2^{c_0 k} (c_0 k)^{c_0 k - 2} e^{-c k / 4} < \frac{\epsilon n}{3}.
\]

A vertex \( v \in [n] \) is good if the \( i \)th 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 \( \approx \epsilon / 3k_1 \) and so the expected number of bad vertices is bounded by \( \epsilon n / 2 \). Thus

\[
\mathbb{E} \left( \sum_{v \in V} \phi(v) - \sum_{v \in V} f(v) \right) \leq \mathbb{E} \left( \sum_{v \in V} \phi(v) - \sum_{v \in V} f(v) \right) + \mathbb{E} \left( \sum_{v \in V} f(v) \right) \\
\leq \mathbb{E} \left( \sum_{v : t(v) = 0} \phi(v) - f(v) \right) + \mathbb{E} \left( \sum_{v \in V} 1 \right) \\
\leq \mathbb{E} \left( \sum_{v : t(v) = 0} 1 \right) + \frac{\epsilon n}{3}.
\]
Let $\mathcal{H}_\varepsilon$ be the set of BFS neighborhoods that are good i.e. whose $i$th levels are of size at most $3c^i k_1/\varepsilon$ 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 $\text{int}(H)$ be the set of vertices incident to the first $k_1 - 1$ neighborhoods of $o_H$ and let $\text{Aut}(H, o_H)$ 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 $M = |E(C_2)|, N = |C_2|,$

$$
\mathbb{E}\left( \sum_{v \text{ is good}} f(v) \mid M, N \right) = \sum_v \sum_{k \geq 1} \sum_{(H, o_H) \in \mathcal{H}_\varepsilon} \rho_{H, o_H} f(o_H) \\
= o(n) + \sum_v \sum_{k \geq 1} \sum_{(H, o_H) \in \mathcal{H}_\varepsilon} \rho_{H, o_H} f(o_H), \quad (21)
$$

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

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

where $f_k$ is defined in (25) below and $\lambda$ 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

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

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 (36)),

**Lemma 3.1.**

$$
\Pr(|L_{c,n} - \mathbb{E}(L_{c,n})| \geq \varepsilon n + n^{3/4}) = O(e^{-\Omega(n^{1/5})}).
$$

**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

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

The inequality follows from a version of Azuma’s inequality (see (36)), and the lemma follows from taking a union bound over
\[
\exp \left\{ O \left( \frac{c^{k_1} k_1(\epsilon)}{\epsilon} \right) \right\} = \exp \left\{ O \left( \frac{c^{2 \log \frac{1}{c} + 2 \log \frac{1}{c}}{\epsilon} \right) \right\} \\
= \exp \left\{ O \left( \frac{(1/\epsilon)^{2 \log c/c} \log \frac{1}{c}}{\epsilon} \right) \right\} = \exp \left\{ O((1/\epsilon)^{2+2 \log c/c}) \right\}
\]

graphs \(H\). Note also that the \(o(n)\) term in (21) is bounded by the same \(e^{O((1/\epsilon)^{2+2 \log c/c})}\) 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 \(\epsilon \to 0\) and from the Borel-Cantelli lemma.

### 3.1 A Model of \(C_2\)

It is known that given \(M, N\) that, up to relabeling vertices, \(C_2\) is distributed as \(G_{N,M}^{d \geq 2}\). The random graph \(G_{N,M}^{d \geq 2}\) is chosen uniformly from \(G_{N,M}^{d \geq 2}\) 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 \(G_{N,M}^{d \geq 2}\). We use a variation on the pseudo-graph model of Bollobás and Frieze [6] and Chvátal [7]. Given a sequence \(x = (x_1, x_2, \ldots, x_{2M}) \in [n]^{2M}\) of \(2M\) integers between \(1\) and \(N\) we can define a (multi)-graph \(G_x = G_x(N, M)\) with vertex set \([N]\) and edge set \(\{(x_2i-1, x_2i) : 1 \leq i \leq M\}\). The degree \(d_x(v)\) of \(v \in [N]\) is given by

\[
d_x(v) = |\{j \in [2M] : x_j = v\}|.
\]

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

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

\[
[N]_{d \geq 2}^{2M} = \{x \in [N]^{2M} : d_x(j) \geq 2 \text{ for } j \in [N]\}.
\]

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

Let

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

for $k \geq 0$.

\textbf{Lemma 3.2.} Let $x$ 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 $P$, where

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

Here $\lambda$ satisfies

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

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

\textit{Proof.} This can be derived as in Lemma 4 of [2].

It follows from (14) and (26) and the fact that $f_1(\lambda)/f_2(\lambda) \to 1$ as $c \to \infty$ that for large $c$,

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

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

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

Furthermore,

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

and

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

This is an example of a local central limit theorem. See for example, (5) of [2] or (3) of [14]. 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

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

Let $\nu_x(s)$ denote the number of vertices of degree $s$ in $G_x$. 

16
Lemma 3.3. Suppose that $\log N = O((N\lambda)^{1/2})$. Let $x$ be chosen randomly from $[N]^{2M}_{\delta \geq 2}$. Then as in equation (7) of [2], we have that with probability $1 - o(N^{-10})$,

$$\left| \nu_x(j) - \frac{N\lambda^j}{j!f_2(\lambda)} \right| \leq \left( 1 + \frac{N\lambda^j}{j!f_2(\lambda)} \right)^{1/2} \log^2 N, \quad 2 \leq j \leq \log N. \quad (31)$$

$$\nu_x(j) = 0, \quad j \geq \log N. \quad (32)$$

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

$$\Pr(G_x \text{ is simple}) = \Omega(1). \quad (33)$$

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 $\rho_H$ is the probability that $G(N_{k_1}(v)) = H$ in $G_x$ then where $\Phi(2m) = \frac{(2m)!}{m^{2m}}$, we have

$$\rho_H \approx \left( \frac{N}{k-1} \right)^{(k-1)!} \sum_{D=2k}^{\infty} \prod_{D=2k}^{\infty} \prod_{i=1}^{k} \frac{\lambda^{d_i}}{d_i!f_2(\lambda)} \frac{d_i!}{(d_i-z_i)!} \frac{\Phi(2M-2k+2)}{\Phi(2M)} \quad (34)$$

$$= \left( \frac{N}{k-1} \right)^{(k-1)!} \sum_{D=2k}^{\infty} \prod_{D=2k}^{\infty} \prod_{i=1}^{k} \frac{\lambda^{d_i-z_i}}{d_i!z_i!f_2(\lambda)} \frac{\Phi(2M-2k+2)}{\Phi(2M)} \quad (35)$$

$$= \left( \frac{N}{k-1} \right)^{(k-1)!} \sum_{D=2k}^{\infty} \frac{(k\lambda)^{D-2(k-1)}}{(D-2(k-1))!} \quad (36)$$

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

**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 $H$ occurring in $[k]$. We then use the fact that $k$ is small to argue that w.h.p. $H$ is induced.

**Explanation for (35):** We use the identity

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

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

(36)

(Specifically we can use Lemma 11 of Frieze and Pittel [18] or Section 3.2 of McDiarmid [20].) 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 [15]. It would be more interesting to do the analysis for small $c > 1$. Applying the coupling of McDiarmid [19] 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 [17] did for the earlier result of [13].

References


