EXACT THRESHOLD AND LOGNORMAL LIMIT FOR NON-LINEAR HAMILTON CYCLES

BYRON CHIN

ABSTRACT. For positive integers $r > \ell \geq 1$, an ℓ-cycle in an r-uniform hypergraph is a cycle where each edge consists of r vertices and each pair of consecutive edges intersect in ℓ vertices. We show that for $\ell \geq 2$, a random r-uniform hypergraph contains a Hamilton ℓ-cycle with high probability whenever the expected number of such cycles tends to infinity. Moreover, for $\ell = 2$, we show that the normalized number of Hamilton 2-cycles converges to a lognormal distribution. This determines the exact threshold for the appearance of non-linear Hamilton cycles in random hypergraphs, confirming a conjecture of Narayanan and Schacht.

1. INTRODUCTION

The problem of determining the critical density at which a specific substructure will appear inside a random structure has been central to the development of probabilistic combinatorics. Starting with the case of fixed size subgraphs in random graphs, a fairly complete answer was determined by Bollobás using the second moment method [4]. The cases of spanning structures have been more difficult to handle, with a general partial answer provided by Riordan [21], also by a careful second moment calculation. A breakthrough work of Park and Pham [19] resolved the Kahn–Kalai conjecture, which gives a general purpose tool to locate these thresholds up to logarithmic factors. However, pinning down the sharp location of thresholds still remains a challenge.

Several substructures of interest have also been studied specifically. One of the most notable cases is that of perfect matchings. Early works of Erdős and Rényi [9] determined the threshold for the appearance of a perfect matching in a random graphs. However, the corresponding problem for hypergraphs, known as Shamir's problem, was notoriously difficult before it was resolved in a breakthrough work of Johansson, Kahn, and Vu [14].

We study the problem of determining when a random hypergraph contains another spanning structure of interest, the Hamilton cycle – that is a cycle that uses every vertex of the graph. This falls into a long line of work searching for spanning structures in random (hyper)graphs. Starting with the case of graphs, foundational works by Pósa [20], Komlós and Szemerédi [15], and Ajtai, Komlós, and Szemerédi [2] pinned down not only the sharp threshold, but also the more precise hitting time result equating the appearance of a Hamilton cycle with the time at which the minimum degree becomes two.

In the setting of hypergraphs, there is more than one way to form a Hamilton cycle. In particular, for any positive integers $r > \ell \geq 1$, we can consider the r-uniform ℓ-cycle, where each edge consists of r consecutive vertices and any two consecutive edges overlap in exactly ℓ vertices. The case when $\ell = 1$ is known as a linear, or loose, cycle whereas when $\ell = r - 1$ the cycle is called tight. Building on work of Dudek and Frieze [7, 8], Narayanan and Schacht [18] showed that the first moment threshold is sharp for the appearance of any non-linear Hamilton cycle. Their work is a combination of a careful second moment estimate with a general theorem of Friedgut [10] characterizing sharp thresholds.

DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY
E-mail address: `byronc@mit.edu`.

However, finding the sharp threshold in this setting leaves more to be desired, as any density a factor of $1 + \varepsilon$ above the first moment threshold produces *exponentially* many Hamilton cycles on average. This leaves the question of the appearance of Hamilton cycles when the expectation grows at a slower than exponential rate. Narayanan and Schacht conjectured that the first moment threshold is much sharper than they were able to show. In particular, a Hamilton cycle should appear as soon as the expected number tends to infinity, at any arbitrarily slow rate.

Our main result is to confirm this conjecture by carrying out a further refined second moment calculation. While for $\ell \geq 3$, a careful analysis reveals a sharp second moment, the special case of $\ell = 2$ requires using a technique known as small subgraph conditioning. In this case we show that the second moment method is sharp not for the number of Hamilton 2-cycles, but for a suitably modified random variable. A natural consequence of this method is the characterization that the number of Hamilton 2-cycles converges to an explicit lognormal distribution.

1.1. **Main definitions.** Let $G_r(n, p)$ be the r-uniform hypergraph where each subset of the vertices of size r is included as an edge independently with probability p, and let $C_{n,\ell}^{(r)}$ be the r-uniform ℓ-cycle on n vertices. For any hypergraph H, we will use $Z(H)$ to denote the random variable counting the number of copies of H in $G_r(n, p)$ and N_H to denote the number of copies of H in the complete hypergraph. In particular, $N_H = \frac{(n)_{v(H)}}{\mathrm{Aut}(H)}$ where $(n)_k = n(n-1) \cdots (n-k+1)$ is the falling factorial and $\mathrm{Aut}(H)$ is the number of automorphisms of H.

The first moment threshold for $C_{n,\ell}^{(r)}$ can be described as follows. Let $s = r - \ell$ and $1 \leq t \leq s$ be the unique integer satisfying $t \equiv r \pmod{s}$. Define $\lambda(r, \ell) = t!(s - t)!$, then the first moment threshold, where $\mathbb{E}\left[Z(C_{n,\ell}^{(r)}) \right] = 1$, is of the form

$$p^*(r, \ell) = (1 + o(1)) \frac{\lambda(r, \ell) e^{r-\ell}}{n^{r-\ell}}.$$

For the remainder of the paper, we will parameterize $p = cp^*(r, \ell)$. We will focus on the case where c is bounded, in particular c around 1, as this captures the setting where the behavior of the threshold is not known.

1.2. **Results.** With these definitions, we can more precisely describe the current knowledge on this problem. The state of the art was proven by Narayanan and Schacht, who showed that $p^*(r, \ell)$ is a sharp threshold for the appearance of $C_{n,\ell}^{(r)}$. However, note that the probability that a fixed copy of $C_{n,\ell}^{(r)}$ appears is $p^{\frac{n}{r-\ell}}$, so the expected number of copies of $C_{n,\ell}^{(r)}$ for $p > (1 + \varepsilon)p^*(r, \ell)$ is at least $(1 + \varepsilon)^{\frac{n}{r-\ell}}$, which is exponentially large. This motivates their following conjecture.

Conjecture 1.1 ([18, Conjecture 4.1]). *For all integers $r > \ell > 1$, as $n \to \infty$ with $(r - \ell) \mid n$, if $p = p(n)$ is such that $\mathbb{E}\left[Z(C_{n,\ell}^{(r)}) \right] \to \infty$ then*

$$\mathbb{P}\left(C_{n,\ell}^{(r)} \subset G_r(n, p) \right) \to 1.$$

Our main result is that for $\ell \geq 3$, the second moment matches the first moment squared, whereas for $\ell = 2$, the random variable $\frac{Z(C_{n,\ell}^{(r)})}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)}) \right]}$ has a log-normal limiting distribution whenever the expectation tends to infinity.

Theorem 1.2. *For all integers $r > \ell > 1$, as $n \to \infty$ with $(r - \ell) \mid n$, if $p = cp^*(r, \ell)$ is such that $\mathbb{E}\left[Z(C_{n,\ell}^{(r)}) \right] \to \infty$, then*
1. *For $\ell \geq 3$,*
$$\mathbb{E}\left[Z(C_{n,\ell}^{(r)})^2 \right] = (1 + o(1)) \mathbb{E}\left[Z(C_{n,\ell}^{(r)}) \right]^2.$$

2. For $\ell = 2$,

$$\frac{Z(C_{n,\ell}^{(r)})}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]} \to \text{Lognormal}\left(-\sum_{k=1}^{\infty} \frac{A_k c^{-k} e^{-k(r-\ell)}}{2(r-\ell)^2}, \sum_{k=1}^{\infty} \frac{A_k c^{-k} e^{-k(r-\ell)}}{(r-\ell)^2}\right).$$

Here, A_k are explicit constants whose definition we postpone to (1.1). We note that A_k remains bounded as $k \to \infty$, so the infinite series converges quickly to a constant. As a corollary, we confirm the above conjecture.

Corollary 1.3. *Conjecture 1.1 is true.*

Proof. For $\ell \geq 3$, the result follows from a standard application of Chebyshev's inequality. For $\ell = 2$, since the log-normal distribution has a density with respect to the positive real line, we have

$$\mathbb{P}\left(Z(C_{n,\ell}^{(r)}) = 0\right) \leq \mathbb{P}\left(\frac{Z(C_{n,\ell}^{(r)})}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]} = 0\right) \xrightarrow{n\to\infty} 0.$$

$\square$

1.3. **Proof overview.** The difficulty with determining the sharp threshold for Hamilton ℓ-cycles is that the standard second moment does not always work. Narayanan and Schacht [18] overcome this difficulty by estimating the second moment up to a constant factor, and applying a general purpose tool of Friedgut [10] to show that the threshold must be sharp. A key observation of this paper, which is implicit in their handling of the second moment, is that the additional constant factor in the second moment arises from the fluctuations of a fixed collection of small subgraphs. If one were able to isolate the contribution from these small subgraphs, the second moment estimate can be made sharp.

The tool to treat the contribution from these small subgraphs and handle situations where the second moment is off by a constant factor is known as small subgraph conditioning. The idea was introduced in a breakthrough work of Robinson and Wormald [22] to show that random regular graphs have Hamilton cycles with high probability. Janson [13] realized the applicability of the technique and presented a general method to establish contiguity between null and planted distributions. The small subgraph conditioning method went on to play an important role in the analysis of many statistical physics models [11, 5, 17, 16, 6]. On the other hand, applications to more combinatorial settings have been less frequent.

An important commonality between each of the above applications is a sparsity of the model, i.e. the number of edges is proportional to the number of vertices. Notable works of Banerjee [3] and Abbe, Li, and Sly [1] were able to implement a version of the small subgraph conditioning technique in dense models, namely the stochastic block model and symmetric perceptron respectively.

We rely on a similar application of small subgraph conditioning in the dense setting to explain the additional constant factor in the second moment. In particular, we analyze the random variable $X = \frac{Z(C_{n,\ell}^{(r)})}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]}\exp(-Y)$ where

$$Y = \sum_{k=1}^{\log n} t_k Y(P_k).$$

Here P_k is r-uniform ℓ-path with k edges, and $Y(H)$ is the normalized subgraph count of H in a (random) graph G defined by

$$Y(H) = \frac{1}{\sqrt{N_H}} \sum_H \prod_{e \in H} \frac{\mathbb{1}\{e \in G\} - p}{\sqrt{p(1-p)}}.$$

where the sum runs over all copies of H in the complete graph. Note that the normalization is chosen so that $\mathbb{E}\left[Y(H)\right] = 0$ and $\mathrm{Var}\left(Y(H)\right) = 1$.

The intuition for this choice of random variable is that each P_k can be viewed as a "short segment" of a Hamilton cycle. The appearance of each such short segment has a constant multiplicative effect on the expected number of Hamilton cycles. Thus, the total contribution heuristically takes an exponential form, and the correction e^{-Y} removes this contribution. With foresight, we choose the coefficients

$$t_k = \frac{\sqrt{A_k c^{-k} e^{-k(r-\ell)}}}{(r-\ell)} \text{ where } A_k = \frac{\mathrm{Aut}(P_k)}{(t!(s-t)!)^k}. \tag{1.1}$$

This is the precise quantitative version of the "constant multiplicative effect" alluded to above. Notice that for k a sufficiently large constant in terms of r and ℓ, A_k is constant (corresponding to the extra permutations arising from the two ends of the path). In particular, and importantly for our purposes, it is uniformly bounded by a function of r and ℓ.

1.4. **Organization.** In Section 2 we give a careful case analysis to get a sharp estimate on the second moment. This proves part 1 of Theorem 1.2. In Section 3 we quantify the contribution from the small subgraph conditioning modification of the random variable. Finally, Section 4 combines the two elements to prove part 2 of Theorem 1.2.

1.5. **Notation.** We use $\mathbb{P}^*$ and $\mathbb{E}^*$ to denote probability and expectation with respect to the measure $G_r(n,p)$ with a planted copy of $C_{n,\ell}^{(r)}$. For the second moment, we use $\mathbb{P}_t^{*2}$ and $\mathbb{E}_t^{*2}$ to denote probability and expectation with respect to the measure $G_r(n,p)$ with two planted copies of $C_{n,\ell}^{(r)}$, whose intersection consists of t edges. We use C_1 and C_2 to enumerate over all copies of $C_{n,\ell}^{(r)}$, and also C_1 and C_2 to denote two independent uniformly random copies of $C_{n,\ell}^{(r)}$ in the complete hypergraph. For ease of notation, sometimes we write C in place of $C_{n,\ell}^{(r)}$ in sub/superscripts.

2. SECOND MOMENT COMPUTATION

In this section, we evaluate the second moment of $Z(C_{n,\ell}^{(r)})$. We know from [18] that this is simply $\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]^2$ up to a constant factor. However, since we want to show a sharp second moment for X, we need to determine the exact constant factor. Thus, a more careful estimate of the overlap distribution is required. We begin by rewriting the second moment in terms of the equivalence classes of intersection subgraph:

$$\sum_{t=1}^{m} \frac{\mathbb{P}\left(|C_1 \cap C_2| = t\right)}{p^t} = \sum_{F \subset C_{n,\ell}^{(r)}} \frac{\mathbb{P}\left(C_1 \cap C_2 \cong F\right)}{p^{e(F)}} \leq \sum_{F \subset C_{n,\ell}^{(r)}} \frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}}.$$

We will handle the various subgraphs in a few cases:
- Lemma 2.1: if $v(F) \leq \log n$ we get the main contribution from the short segments.
- Lemma 2.2: if $\log n < v(F) < n$ the contribution is vanishing due to the discrepancy between $\left(\frac{n}{e}\right)^k$ and $(n)_k$.
- Lemma 2.3: if $v(F) = n$ and F is connected, then by comparison to $F = C_{n,\ell}^{(r)}$ the contribution is bounded by $\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]^{-1}$.
- Lemma 2.4: if $v(F) = n$ and F is disconnected, the contribution is vanishing due to the limited number of copies of F in $C_{n,\ell}^{(r)}$.

Each F in the sum can be decomposed into its connected components as follows. For each k, let there be n_k components isomorphic to P_k in F. Denote the remaining isomorphism classes of components by $F_1, \ldots, F_a$, with multiplicities $m_1, \ldots, m_a$. Then

$$\frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \leq \frac{(\frac{n}{r-\ell})^{2\sum_k n_k + \sum_j m_j}}{(n)_{v(F)}} \cdot \prod_k \frac{\operatorname{Aut}(P_k)^{n_k}}{n_k!} p^{-k n_k} \cdot \prod_{j=1}^a \frac{\operatorname{Aut}(F_j)^{m_j}}{m_j!} p^{-e(F_j) m_j}. \qquad (2.1)$$

We analyze this decomposition according to the cases listed above.

Lemma 2.1.

$$\sum_{F \subset C_{n,\ell}^{(r)}, v(F) \leq \log n} \frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \leq \exp\left((1 + o(1)) \sum_{k=1}^{\log n} \frac{A_k c^{-k} e^{-k(r-\ell)}}{(r-\ell)^2}\right).$$

Proof. Note that we have $(n)_k = (1 + o(1))n^k$ for $k = o(\sqrt{n})$. Thus,

$$\sum_{\substack{F \subset C_{n,\ell}^{(r)} \\ v(F) \leq \log n}} \frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \leq (1 + o(1)) \frac{(\frac{n}{r-\ell})^{2\sum_k n_k + \sum_j m_j}}{n^{v(F)}} \cdot \prod_k \frac{\operatorname{Aut}(P_k)^{n_k}}{n_k!} \cdot \prod_{j=1}^a \frac{\operatorname{Aut}(F_j)^{m_j}}{m_j!} \cdot p^{-e(F)}$$

$$\leq (1 + o(1)) \prod_{\substack{F \subset C_{n,\ell}^{(r)} \\ v(F) \leq \log n \\ F \text{ conn.}}} \sum_{k=1}^\infty \frac{1}{k!} \left(\frac{n^2}{(r-\ell)^2 n^{v(F)}} \operatorname{Aut}(F) p^{-e(F)}\right)^k$$

$$= (1 + o(1)) \prod_{\substack{F \subset C_{n,\ell}^{(r)} \\ v(F) \leq \log n \\ F \text{ conn.}}} \exp\left(\frac{n^2}{(r-\ell)^2 n^{v(F)}} \operatorname{Aut}(F) p^{-e(F)}\right)$$

$$= (1 + o(1)) \exp\left(\sum_{\substack{F \subset C_{n,\ell}^{(r)} \\ v(F) \leq \log n \\ F \text{ conn.}}} \frac{n^2}{(r-\ell)^2 n^{v(F)}} \operatorname{Aut}(F) p^{-e(F)}\right) \qquad (2.2)$$

We first handle the terms corresponding to $F = P_k$. For this we explicitly have $e(P_k) = k$, $v(P_k) = \ell + k(r - \ell)$, $\operatorname{Aut}(P_k) = A_k(t!(s-t)!)^k$ where A_k is the correction for the edges near the ends of the path. Importantly this is independent of the length of the path. Thus, we get

$$\frac{n^2}{(r-\ell)^2 n^{v(P_k)}} \operatorname{Aut}(P_k) p^{-e(P_k)} = \frac{n^2}{n^{\ell + (r-\ell)k}} \cdot \frac{A_k}{(r-\ell)^2} (t!(s-t)!)^k \left(\frac{n^{r-\ell}}{ct!(s-t)!e^{r-\ell}}\right)^k$$

$$\leq \frac{A_k}{(r-\ell)^2} c^{-k} e^{-k(r-\ell)} n^{2-\ell}.$$

Now consider a connected component F_j not isomorphic to any P_k. Suppose first that $v(F_j) < n$. Then $v(F_j) = v(P_{k_j})$ for some k_j. We compare the contribution of F_j with that of P_{k_j}. The key observation is that we have the bound

$$\operatorname{Aut}(F_j) \leq \operatorname{Aut}(P_{k_j}) \cdot (r!^2)^{k_j - e(F_j)}.$$

This is because for each edge removed from P_{k_j}, we free up the vertices within two edges to be permuted, and there are at most $r!$ ways to permute within each of these edges. Thus we can bound

$$\frac{n^2}{(r-\ell)^2 n^{v(F_j)}}\mathrm{Aut}(F_j)p^{-e(F_j)} \leq \frac{n^2}{n^{\ell+(r-\ell)k_j}} \cdot \frac{A_{k_j}}{(r-\ell)^2}(t!(s-t)!)^{k_j}(r!^2)^{k_j-e(F_j)}p^{k_j-e(F_j)}p^{-k_j}$$

$$= \frac{n^2}{n^{\ell+(r-\ell)k_j}} \cdot \frac{A_{k_j}}{(r-\ell)^2}(t!(s-t)!)^{k_j}(r!^2 p)^{k_j-e(F_j)}\left(\frac{n^{r-\ell}}{ct!(s-t)!e^{r-\ell}}\right)^{k_j}$$

$$\leq \frac{A_{k_j}}{(r-\ell)^2}c^{-k_j}e^{-k_j(r-\ell)}n^{2-\ell}\left(\frac{r!^2 ct!(s-t)!e^{r-\ell}}{n^{r-\ell}}\right)^{k_j-e(F_j)}$$

$$= \frac{A_{k_j}}{(r-\ell)^2}c^{-k_j}e^{-k_j(r-\ell)}n^{2-\ell}n^{-(r-\ell+o(1))(k_j-e(F_j))} \tag{2.3}$$

Notice here that $e(F_j) < k_j$ by assumption, so the contribution is smaller by a factor of at least n for every edge that is missing from P_{k_j}. In particular, we have the bound

$$\sum_{\substack{F \subset C_{n,\ell}^{(r)} \\ v(F)=v(P_{k_j}) \\ F \text{ conn.}}} \frac{n^2}{(r-\ell)^2 n^{v(F)}}\mathrm{Aut}(F)p^{-e(F)} \leq \frac{A_{k_j}}{(r-\ell)^2}c^{-k_j}e^{-k_j(r-\ell)}n^{2-\ell}\sum_{k=0}^{k_j}\binom{k_j}{k}n^{-(r-\ell+o(1))k}$$

$$= \frac{A_{k_j}}{(r-\ell)^2}c^{-k_j}e^{-k_j(r-\ell)}n^{2-\ell}\left(1+n^{-(r-\ell+o(1))}\right)^{k_j}$$

$$\leq (1+o(1))\frac{A_{k_j}}{(r-\ell)^2}c^{-k_j}e^{-k_j(r-\ell)}n^{2-\ell} \tag{2.4}$$

for $k_j \leq \log n$. Thus, all together we have

$$\sum_{F \subset C_{n,\ell}^{(r)},v(F)\leq \log n}\frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \leq \exp\left((1+o(1))n^{2-\ell}\sum_{k=1}^{\log n}\frac{A_k}{(r-\ell)^2}c^{-k}e^{-k(r-\ell)}\right).$$

$$\square$$

Lemma 2.2.
$$\sum_{\log n < v(F) \leq n-1}\frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \lesssim_{r,\ell} \frac{1}{n}.$$

Proof. For the remaining cases we require a more general purpose bound on the falling factorial in this regime. Stirling's approximation yields $(n)_k \geq n^k \exp\left(-n\left((1-\frac{k}{n})\log\left(1-\frac{k}{n}\right)+\frac{k}{n}\right)\right)$. Applying this estimate in (2.1), the right hand side factors according to the connected components.

$$\frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \leq \exp\left((n-v(F))\log\left(1-\frac{v(F)}{n}\right)+v(F)\right)\prod_k \frac{1}{n_k!}\left(\frac{n^2}{n^{v(P_k)}}\mathrm{Aut}(P_k)p^{-e(P_k)}\right)^{n_k}$$

$$\cdot \prod_{j=1}^a \frac{1}{m_j!}\left(\frac{n^2}{n^{v(F_j)}}\mathrm{Aut}(F_j)p^{-e(F_j)}\right)^{m_j}.$$

Note that we can express (2.3) as

$$\frac{n^2}{n^{v(F_j)}}\mathrm{Aut}(F_j)p^{-e(F_j)} \leq A_{k_j}c^{-k_j}e^{-v(F_j)}e^{\ell}n^{-(r-\ell+o(1))(k_j-e(F_j))}.$$

Substituting this into the above expression yields

$$\frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \leq \exp\left((n - v(F)) \log\left(1 - \frac{v(F)}{n}\right)\right) \prod_k \frac{1}{n_k!} \left(A_k c^{-k} e^\ell n^{o(1)}\right)^{n_k}$$

$$\cdot \prod_{j=1}^{a} \frac{1}{m_j!} \left(A_{k_j} c^{-k_j} e^\ell n^{-(r-\ell+o(1))(k_j - e(F_j))}\right)^{m_j}.$$

By the same calculation as (2.2), we have the estimate

$$\sum_{v(F)=t} \frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^t} \leq \exp\left((n - t) \log\left(1 - \frac{t}{n}\right)\right)$$

$$\exp\left(\sum_{F \subset C_{n,\ell}^{(r)}, v(F) \leq t, F \text{ conn.}} A_{k_j} c^{-k_j} e^\ell n^{-(r-\ell+o(1))(k_j - e(F))}\right)$$

The analog of (2.4) then gives

$$\sum_{v(F)=t} \frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \leq \exp\left((n - t) \log\left(1 - \frac{t}{n}\right)\right) \exp\left((1 + o(1)) \sum_{k=1}^{t} A_k c^{-k} e^\ell\right).$$

The second exponential term is uniformly bounded by a constant in r and ℓ. Moreover, for $2 \log n \leq t \leq n - 2$, the first exponential term is at most $\frac{4}{n^2}$, while for $\log n \leq t \leq 2 \log n$ or $t = n - 1$ it is at most $\frac{1}{n}$. All together,

$$\sum_{\log n < v(F) \leq n-1} \frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \lesssim_{r,\ell} \frac{1}{n}.$$

$\square$

Lemma 2.3.

$$\sum_{F \subset C_{n,\ell}^{(r)}, v(F)=n, F \text{ conn.}} \frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \lesssim_{r,\ell} \frac{1}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]}.$$

Proof. We now handle the terms that have a large number of vertices, and show that they contribute negligibly. We start with the case that F is a single connected spanning component. We let $m = \frac{n}{r-\ell}$ be the number of edges in $C_{n,\ell}^{(r)}$. By a similar calculation as above, when $v(F_j) = n$, we find

$$\frac{n^2}{(r-\ell)^2 n!} \mathrm{Aut}(F_j) p^{-e(F_j)} \leq \frac{n}{n!} (r!^2)^{m-e(F_j)} \mathrm{Aut}(C_{n,\ell}^{(r)}) p^{-m} p^{m-e(F_j)}$$

$$\leq \frac{n(r!^2)^{m-e(F_j)} p^{m-e(F_j)}}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]}.$$

In this case $F = F_j$ as there are no vertices left to form components. Thus, the total contribution of all spanning subgraphs F is bounded by

$$\sum_{\substack{F \subset C_{n,\ell}^{(r)} \\ v(F)=n \\ F \text{ conn.}}} \frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \leq \frac{1}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]} \left(1 + \sum_{k=1}^{m} \binom{m-1}{k-1} n(r!)^{2k} p^k\right)$$

$$= \frac{1}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]} \left(1 + \frac{c t!(s-t)!(r!)^2 e^{r-\ell}}{n^{r-\ell-1}} \sum_{k=1}^{m} \binom{m-1}{k-1} (r!)^{2k-2} p^{k-1}\right)$$

$$= \frac{1}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]} \left(1 + \frac{ct!(s-t)!(r!)^2 e^{r-\ell}}{n^{r-\ell-1}}(1 + (r!)^2 p)^{m-1}\right)$$

$$\leq \frac{1}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]} \left(1 + \frac{ct!(s-t)!(r!)^2 e^{r-\ell}}{n^{r-\ell-1}} e^{(r!)^2 p(m-1)}\right)$$

$$\lesssim_{r,\ell} \frac{1}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]}.$$

$\square$

Lemma 2.4.

$$\sum_{v(F)=n, F \; disconn.} \frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \lesssim_{r,\ell} \frac{1}{n^2}.$$

Proof. The only case that remains is when F consists of n vertices but multiple components. To handle this, we exploit the fact that in this case, the factorization (2.1) is off by a factor of at least n^2. Indeed, when choosing the locations of each component, since the subgraph covers all of the vertices, the last component is fixed and has no choices left. Thus, we gain a factor of n^2 over the above estimate, and so

$$\sum_{v(F)=n, F \; disconn.} \frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \lesssim_{r,\ell} \frac{1}{n^2}.$$

$\square$

Summing the inequalities yields the following (sharp) estimate on the second moment of $Z(C_{n,\ell}^{(r)})$. We will however need to use the refined cases above in the next section.

Proposition 2.5. *When* $\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right] \to \infty$,

$$\frac{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})^2\right]}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]^2} = \sum_{t=1}^{m} \frac{\mathbb{P}\left(|C_1 \cap C_2| = t\right)}{p^t} \leq \exp\left((1+o(1))n^{2-\ell} \sum_{k=1}^{\log n} \frac{A_k c^{-k} e^{-k(r-\ell)}}{(r-\ell)^2}\right).$$

Notice that when $\ell \geq 3$, the upper bound on the right hand side is $1 + o(1)$. This proves part 1 of Theorem 1.2.

3. Small subgraph conditioning

In this section, we set up the small subgraph conditioning and explain the improvement over the vanilla second moment. Since this improvement is only needed in the case $\ell = 2$, we will assume this for the remainder of the paper. The main claim is the following lemma.

Lemma 3.1.

$$\frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[X\right]^2} = \sum_{t=1}^{m} \frac{\mathbb{P}\left(|C_1 \cap C_2| = t\right)}{p^t} \cdot \frac{\mathbb{E}_t^{*2}[e^{-2Y}]}{\mathbb{E}^*[e^{-Y}]^2}.$$

Notice that compared to the vanilla second moment on $Z(C_{n,\ell}^{(r)})$, the right hand side has an additional term defined by the moment generating function of Y under a planted distribution. As will be shown later in Lemma 3.2, the additional term contributes a small (less than 1) constant factor for t small, and a large constant factor for t large. The key to the argument is that the small constant factor exactly cancels the extra constant factor in the vanilla second moment, and the large constant factor is negligible compared to the probability of large overlap.

Proof. Let $p(G) = \mathbb{P}\left(G_r(n, p) = G\right) = p^{e(G)}(1 - p)^{\binom{n}{r} - e(G)}$. We write

$$\mathbb{E}\left[Z(C_{n,\ell}^{(r)})e^{-Y}\right] = \sum_G p(G)Z(C_{n,\ell}^{(r)})e^{-Y(G)} = \sum_G \sum_{C_1} p(G)\mathbb{1}\{C_1 \subset G\}e^{-Y(G)}$$

where the first sum is over all graphs G. Note that

$$\mathbb{E}^*[e^{-Y}] = \frac{1}{N_C}\sum_{C_1}\frac{\sum_G \mathbb{1}\{C_1 \subset G\}p(G)e^{-Y(G)}}{\sum_G \mathbb{1}\{C_1 \subset G\}p(G)} = \frac{1}{\mathbb{E}\left[N_C(G)\right]}\sum_{C_1}\sum_G \mathbb{1}\{C_1 \subset G\}p(G)e^{-Y(G)}.$$

This implies

$$\mathbb{E}\left[Z(C_{n,\ell}^{(r)})e^{-Y}\right] = \mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]\mathbb{E}^*[e^{-Y}].$$

We perform a similar transformation for the second moment.

$$\mathbb{E}\left[Z^2 e^{-2Y}\right] = \sum_G \sum_{C_1,C_2} p(G)\mathbb{1}\{C_1 \cup C_2 \subset G\}e^{-2Y(G)}$$

$$= \sum_{k=0}^{e(C_{n,\ell}^{(r)})} \sum_{e(C_1 \cap C_2)=k} \sum_G p(G)\mathbb{1}\{C_1 \cup C_2 \subset G\}e^{-2Y(G)}.$$

This time, letting $N_C^{(2)}(k)$ be the number of pairs of copies of $C_{n,\ell}^{(r)}$ which intersect at k edges, we have

$$\mathbb{E}_k^{*2}[e^{-2Y}] = \frac{1}{N_C^{(2)}(k)}\sum_{e(C_1 \cap C_2)=k}\frac{\sum_G p(G)\mathbb{1}\{C_1 \cup C_2 \subset G\}e^{-2Y(G)}}{\sum_G p(G)\mathbb{1}\{C_1 \cup C_2 \subset G\}}$$

$$= \frac{1}{N_C^2\,\mathbb{P}\left(e(C_1 \cap C_2) = k\right)p^{2e(C)-k}}\sum_{e(C_1 \cap C_2)=k}\sum_G p(G)\mathbb{1}\{C_1 \cup C_2 \subset G\}e^{-2Y(G)}$$

$$= \frac{1}{\mathbb{E}\left[N_C(G)\right]^2\,\mathbb{P}\left(e(C_1 \cap C_2) = k\right)p^{-k}}\sum_{e(C_1 \cap C_2)=k}\sum_G p(G)\mathbb{1}\{C_1 \cup C_2 \subset G\}e^{-2Y(G)}$$

Substituting this in yields

$$\mathbb{E}\left[Z(C_{n,\ell}^{(r)})^2 e^{-2Y}\right] = \mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]^2 \sum_{k=0}^{e(C_{n,\ell}^{(r)})} \mathbb{P}\left(e(C_1 \cap C_2) = k\right)p^{-k}\mathbb{E}_k^{*2}[e^{-2Y}]$$

The conclusion follows by plugging in the first moment expression. $\qquad\square$

Following Lemma 3.1, to complete the "conditioning" portion of the method we need to evaluate the moment generating function of Y with respect to various planted distributions. The next lemma shows that the variable Y converges to a Gaussian random variable, and thus the moment generating functions will be easy to evaluate.

Lemma 3.2. *Under the null model* $\mathbb{P}$, *for any integer* k *as* $n \to \infty$,

$$(Y(P_1), Y(P_2), \ldots, Y(P_k)) \to \mathcal{N}(0, I_k).$$

Under the planted model $\mathbb{P}^*$, *for any integer* k *as* $n \to \infty$,

$$(Y(P_1), Y(P_2), \ldots, Y(P_k)) \to \mathcal{N}(\mu, I_k)$$

where $\mu_j = \frac{\sqrt{A_j c^{-j}e^{-j(r-\ell)}}}{(r-\ell)}$. *Under the double planted model* $\mathbb{P}_t^{*2}$, *for any integer* k *and* $t = o(n)$ *as* $n \to \infty$,

$$(Y(P_1), Y(P_2), \ldots, Y(P_k)) \to \mathcal{N}(2\mu, I_k).$$

Proof. The convergence to a Gaussian distribution under the null model follows from work of Janson using orthogonal decompositions [12, Theorem 3, Remark 5.1]. Note here that we have defined each $Y(P_k)$ in orthogonal decomposition form, and every component is connected. We now compute the mean and variances. Recall that for any j, we have

$$Y(P_j) = \frac{1}{\sqrt{N_{P_j}}} \sum_{P_j} \prod_{e \in P_j} Y_e$$

where $Y_e = \frac{\mathbb{1}\{e \in G\} - p}{\sqrt{p(1-p)}}$ and $\mathbb{E}[Y(P_j)] = 0$ and $\mathrm{Var}(Y(P_j)) = 1$. Notice that for any $j_1 \neq j_2$, and each copy of P_{j_1} and P_{j_2} we have $\mathbb{E}\left[\prod_{e_1 \in P_{j_1}} Y_{e_1} \prod_{e_2 \in P_{j_2}} Y_{e_2}\right] = 0$. This implies that $\mathbb{E}[Y(P_{j_1})Y(P_{j_2})] = 0$ so the statistics for distinct subgraphs are uncorrelated. Thus, since each variable converges to a standard Gaussian, their joint limit is a standard k-dimensional Gaussian.

We now study the behavior of the random variables under the planted model. The convergence to a Gaussian distribution is a little more subtle. We can view the planted model as sampling a random subgraph of $K_n \setminus C_n$. The copies of P_j that have planted edges now have coefficients $\sqrt{\frac{1-p}{p}}$ raised to the number of planted edges. Now the general orthogonal decomposition outlined by Janson no longer holds, as we distinguish equivalence classes only up to isomorphism of $K_n \setminus C_n$. Fortunately for us, we have written the definition of $Y(P_j)$ in terms of an orthogonal decomposition, and importantly the coefficients of all isomorphic subgraphs are the same – it is determined by the number of edges missing from P_j. Thus, the convergence follows once we check that the dominant contribution comes from connected subgraphs.

We compute the mean and covariance structure of the resulting Gaussian. From this computation, it will follows that the dominant contribution comes from the full P_j's with no planted edges. We can compute

$$\mathbb{E}^*[Y(P_j)] = \mathbb{E}^*\left[\frac{1}{\sqrt{N_{P_j}}} \sum_{P_j} \prod_{e \in P_j} Y_e\right] = \frac{1}{\sqrt{N_{P_j}}} \sum_{P_j} \prod_{e \in P_j} \mathbb{E}^*[Y_e] = \frac{n}{(r - \ell)} \cdot \sqrt{\frac{(1-p)^j}{p^j N_{P_j}}}$$

$$= (1 + o(1)) \frac{n}{r - \ell} \cdot \left(\frac{n^{(r-\ell)}}{ce^{(r-\ell)}\lambda}\right)^{j/2} \cdot \sqrt{\frac{\mathrm{Aut}(P_j)}{(n)_{v(P_j)}}} = (1 + o(1)) \frac{\sqrt{A_j c^{-j} e^{-j(r-\ell)}}}{(r - \ell)}.$$

The last equality follows since $\mathbb{E}^*[Y_e] = \sqrt{\frac{1-p}{p}}$ if e is a planted edge and $\mathbb{E}^*[Y_e] = 0$ if e is not a planted edge. The number of copies of P_j which consist only of planted edges is the number of copies of P_j in $C_{n,\ell}^{(r)}$, which is $\frac{n}{r-\ell}$.

$$\mathrm{Var}^*(Y(P_j)) = \mathbb{E}^*[Y(P_j)^2] - \mathbb{E}^*[Y(P_j)]^2$$

$$= \frac{1}{N_{P_j}} \sum_{P_1, P_2} \mathbb{E}^*\left[\prod_{e_1 \in P_1} Y_{e_1} \prod_{e_2 \in P_2} Y_{e_2}\right] - \mathbb{E}^*\left[\prod_{e_1 \in P_1} Y_{e_1}\right] \mathbb{E}^*\left[\prod_{e_2 \in P_2} Y_{e_2}\right]$$

$$= \frac{1}{N_{P_j}} \sum_{P_1, P_2} \left(\frac{1-p}{p}\right)^{e(P_1 \cap C^*)} \mathbb{1}\{P_1 \triangle P_2 \subset C^*, P_1 \cap P_2 \not\subset C^*\}.$$

For each pair where $P_1 = P_2$ and $P_1 \cap C^* = \emptyset$, we get a contribution of 1. This amounts to a total contribution of $1 - o(1)$. We show that all other terms are negligible. Let (P_1, P_2) be such that $e(P_1 \cap C^*) = i \neq 0$, $P_1 \triangle P_2 \subset C^*$ and $P_1 \cap P_2 \not\subset C^*$. This pair contributes $\left(\frac{1-p}{p}\right)^i = (1 + o(1))p^{-i}$. We now proceed to bound the number of such pairs. Let a denote the number of ℓ-path segments in $P_1 \cap C^*$. By stars and bars, we can choose the sizes of these segments in $\binom{i+a-1}{i}$ ways. We choose

the location of each segment along C^* in $\frac{n}{r-\ell}$ ways, leading to at most $\left(\frac{n}{r-\ell}\right)^a$ choices. There are $(r-\ell)(j-i)+\ell-\ell a$ vertices in P_1 that have yet to be fixed, leading to $n^{(r-\ell)(j-i)+\ell-\ell a}$ choices. Finally, given P_1, P_2 must live in the union $P_1 \cup C^*$. There are at most $ja2^a$ possibilities – j choices for which edge of P_2 first meets C^*, a choices for where P_2 meets C^*, and 2 choices for which structure to follow at each of the a branching points. All together, this is an upper bound of

$$\sum_{a=1}^{j} \binom{i+a-1}{i} \left(\frac{n}{r-\ell}\right)^a n^{(r-\ell)(j-i)+\ell-\ell a} \cdot ja2^a \lesssim_{r,\ell} n^{(r-\ell)(j-i)+1}.$$

Thus, the total contribution from such pairs is at most

$$\frac{1}{N_{P_j}} n^{(r-\ell)(j-i)+1} p^{-i} \lesssim_{r,\ell} n^{1-\ell}.$$

Summing from $i=1$ to j, we obtain the desired estimate that

$$\mathrm{Var}^*(Y(P_j)) = 1 + o(1).$$

For the covariance, we see that in computing $\mathbb{E}^*[Y(P_{j_1})Y(P_{j_2})]$ we no longer have the possibility that $P_1 = P_2$ and therefore there is no contribution of 1. The same argument shows that $\mathrm{Cov}^*(Y(P_{j_1}), Y(P_{j_2})) = o(1)$.

Similarly, under the double planted model $\mathbb{P}_t^{*2}$ the convergence to standard Gaussians follows by observing that all equivalence classes of isomorphic graphs have the same coefficients. By the calculation in the planted case, each of the two cycles contributes $\frac{\sqrt{A_j c^{-j} e^{-j(r-\ell)}}}{(r-\ell)}$ to the expectation. It remains to estimate the contributions involving both cycles. There are t edges at which the cycles intersect, and at most $j2^j$ paths emanating from each of these intersections. Thus, the total contribution from paths intersecting both cycles is bounded by

$$tj2^j \sqrt{\frac{(1-p)^j}{p^j N_{P_j}}} = o(1).$$

Thus all together the expectation is

$$\mathbb{E}_t^{*2}[Y(P_j)] = (1+o(1)) \frac{2\sqrt{A_j c^{-j} e^{-j(r-\ell)}}}{(r-\ell)}.$$

The covariance structure follows from a similar calculation to the planted model case. We again count the number of pairs P_1 and P_2 such that $e(P_1 \cap (C_1^* \cup C_2^*)) = i$, $P_1 \triangle P_2 \subset C_1^* \cup C_2^*$ and $P_1 \cap P_2 \not\subset C_1^* \cup C_2^*$. Once again let a denote the number of ℓ-path segments in $P_1 \cap (C_1^* \cup C_2^*)$. The number of ways to choose the sizes of these segments remains $\binom{i+a-1}{i}$. We can choose the locations of the segments in $\left(\frac{2n}{r-\ell}\right)^a$ ways, but these segments are no longer uniquely determined as they can oscillate between C_1^* and C_2^*. This can happen in at most 2^i ways. The number of ways to choose the remaining vertices of P_1 is still $n^{(r-\ell)(j-i)+\ell-\ell a}$. Finally, P_2 must live in the union $P_1 \cup C_1^* \cup C_2^*$. There are j choices for which edge of P_2 meets $C_1^* \cup C_2^*$, a choices for the meeting location, and 3 choices at each edge for which path to follow. Thus, there are at most $ja3^j$ choices for P_2. All together, we obtain an upper bound of

$$\sum_{a=1}^{j} \binom{i+a-1}{i} \left(\frac{2n}{r-\ell}\right)^a n^{(r-\ell)(j-i)+\ell-\ell a} ja3^j \lesssim_{r,\ell} n^{(r-\ell)(j-i)+1}$$

as before. Thus, the total contribution remains $n^{1-\ell}$, and we have shown that $\mathrm{Var}_t^{*2}(Y(P_j)) = 1 + o(1)$, and $\mathrm{Cov}_t^{*2}(Y(P_{j_1}), Y(P_{j_2})) = o(1)$. $\qquad\square$

Using the above central limit theorem, we evaluate the additional multiplicative terms from the right hand side of Lemma 3.1.

Lemma 3.3. *Let $Y_N = \sum_{k=1}^{N} t_k Y(P_k)$. For all $t = o(n)$ as $n \to \infty$,*

$$\frac{\mathbb{E}_t^{*2}[e^{-2Y_N}]}{\mathbb{E}^*[e^{-Y_N}]^2} = \exp\left(-(1+o(1)) \sum_{k=1}^{N} \frac{A_k c^{-k} e^{-k(r-\ell)}}{(r-\ell)^2}\right).$$

Proof. Using Lemma 3.2, we can compute

$$\mathbb{E}^*[e^{-Y_N}] = \exp\left(\sum_{k=1}^{N} -t_k\mu_k + \frac{t_k^2}{2}\right) = \exp\left(-(1+o(1)) \sum_{k=1}^{N} \frac{A_k c^{-k} e^{-k(r-\ell)}}{2(r-\ell)^2}\right)$$

$$\mathbb{E}_t^{*2}[e^{-2Y_N}] = \exp\left(\sum_{k=1}^{N} -2t_k(2\mu_k) + 2t_k^2\right) = \exp\left(-(1+o(1))2 \sum_{k=1}^{N} \frac{A_k c^{-k} e^{-k(r-\ell)}}{(r-\ell)^2}\right)$$

The convergence follows. $\square$

4. PROOF OF MAIN THEOREM

In this section, we combine the results of Sections 2 and 3 to prove the Theorem 1.2 in the case $\ell = 2$. To begin, we use the sharp second moment to deduce that X converges in probability to a constant.

Lemma 4.1.

$$\frac{Z(C_{n,\ell}^{(r)})}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right] e^Y} \xrightarrow{\text{p}} \exp\left(-\sum_{k=1}^{\infty} \frac{A_k c^{-k} e^{-k(r-\ell)}}{2(r-\ell)^2}\right).$$

Proof. Fix $\varepsilon > 0$. For any $\delta > 0$, choose $N = N_0(\varepsilon, \delta)$ such that $\exp\left(\sum_{k=N_0}^{\infty} \frac{A_k c^{-k} e^{-k(r-\ell)}}{(r-\ell)^2}\right) < 1 + \varepsilon^2 \delta$ and $e^{-N(r-\ell)} \le \delta(\log(1+\varepsilon))^2$.

Denote the desired limit by $L = \exp\left(-\sum_{k=1}^{\infty} \frac{A_k c^{-k} e^{-k(r-\ell)}}{2(r-\ell)^2}\right)$. We write

$$|X - L| \le |X - \widetilde{X}| + |\widetilde{X} - \widetilde{X}_N| + |\widetilde{X}_N - L|$$

where $\widetilde{X}$ be the truncation $\widetilde{X} = \frac{Z(C_{n,\ell}^{(r)})}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]} \exp\left(-\widetilde{Y}\right)$ with $\widetilde{Y} = \max\{\min\{Y, M\}, -M\}$ and $M = \min\{\log\log \mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right], \log\log n\}$. Recall that Y_N denotes the partial sum $\sum_{k=1}^{N} t_k Y(P_k)$, and define the truncation $\widetilde{\widetilde{X}}_N$ similarly.

We first control the difference between Y and Y_N. Notice that

$$\mathbb{E}\left[(Y - Y_N)^2\right] = \mathbb{E}\left[\left(\sum_{k=N}^{\log n} t_k Y(P_k)\right)^2\right] = \sum_{k=N}^{\log n} t_k^2 \mathbb{E}\left[Y(P_k)^2\right] = \sum_{k=N}^{\log n} \frac{A_k c^{-k} e^{-k(r-\ell)}}{(r-\ell)^2} \lesssim_{r,\ell} e^{-N(r-\ell)}.$$

The first difference is small since Y is close to Y_N, and Lemma 3.2 implies Y_N is normally distributed, and thus is very large in magnitude with small probability. Indeed for any $\varepsilon > 0$,

$$\mathbb{P}\left(|X - \widetilde{X}| > \varepsilon\right) \le \mathbb{P}\left(Y \ne \widetilde{Y}\right) = \mathbb{P}(|Y| > M) \le \mathbb{P}\left(|Y_N| > \frac{M}{2}\right) + \mathbb{P}\left(|Y - Y_N| > \frac{M}{2}\right) = o(1).$$

For the third difference, recall from (a variant of) Lemma 3.1 that we can write

$$\frac{\mathbb{E}[\widetilde{X}_N^2]}{\mathbb{E}[\widetilde{X}_N]^2} = \sum_{t=1}^{m} \frac{\mathbb{P}(|C_1 \cap C_2| = t)}{p^t} \cdot \frac{\mathbb{E}_t^{*2}[e^{-2\widetilde{Y}_N}]}{\mathbb{E}^*[e^{-\widetilde{Y}_N}]^2} = \sum_{F \subset C_{n,\ell}^{(r)}} \frac{\mathbb{E}[N_F(C_1 \cap C_2)]}{p^{e(F)}} \frac{\mathbb{E}_{e(F)}^{*2}[e^{-2\widetilde{Y}_N}]}{\mathbb{E}^*[e^{-\widetilde{Y}_N}]^2}.$$

We once again handle the various subgraphs in a few cases, similarly to the analysis in Section 2.

Case 1. If $v(F) \leq \log n$ then Lemmas 2.1 and 3.3 yield that

$$\sum_{F \subset C_{n,\ell}^{(r)}, v(F) \leq \log n} \frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \frac{\mathbb{E}_{e(F)}^{*2}[e^{-2\widetilde{Y}_N}]}{\mathbb{E}^*[e^{-\widetilde{Y}_N}]^2} \leq$$

$$\exp\left((1 + o(1)) \sum_{k=1}^{\log n} \frac{A_k}{(r-\ell)^2} c^{-k} e^{-k(r-\ell)}\right).$$

$$\exp\left(-(1 + o(1)) \sum_{k=1}^{N} \frac{A_k}{(r-\ell)^2} c^{-k} e^{-k(r-\ell)}\right) \leq 1 + \varepsilon^2 \delta + o(1).$$

Here we used that truncating Y_N to $\widetilde{Y}_N$ only changes the first and second moments by $o(1)$ factors.

Case 2. If $\log n < v(F) < n$ then Lemma 2.2 along with the truncation implies

$$\sum_{F \subset C_{n,\ell}^{(r)}, v(F) \leq \log n} \frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \frac{\mathbb{E}_{e(F)}^{*2}[e^{-2\widetilde{Y}_N}]}{\mathbb{E}^*[e^{-\widetilde{Y}_N}]^2} \lesssim_{r,\ell} \frac{e^{4M}}{n} \leq \frac{(\log n)^4}{n} = o(1).$$

Case 3. If $v(F) = n$ and F is connected, then Lemma 2.3 along with the truncation implies

$$\sum_{F \subset C_{n,\ell}^{(r)}, v(F)=n, F \text{ conn.}} \frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \frac{\mathbb{E}_{e(F)}^{*2}[e^{-2\widetilde{Y}_N}]}{\mathbb{E}^*[e^{-\widetilde{Y}_N}]^2} \lesssim_{r,\ell} \frac{e^{4M}}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]} \leq \frac{\left(\log \mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]\right)^4}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right]}.$$

Case 4. If $v(F) = n$ and F is disconnected then Lemma 2.4 along with the truncation implies

$$\sum_{F \subset C_{n,\ell}^{(r)}, v(F)=n, F \text{ disconn.}} \frac{\mathbb{E}\left[N_F(C_1 \cap C_2)\right]}{p^{e(F)}} \frac{\mathbb{E}_{e(F)}^{*2}[e^{-2\widetilde{Y}_N}]}{\mathbb{E}^*[e^{-\widetilde{Y}_N}]^2} \lesssim_{r,\ell} \frac{e^{4M}}{n^2} \leq \frac{(\log n)^4}{n^2} = o(1).$$

In particular, when $\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right] \to \infty$, we have that

$$\frac{\mathbb{E}[\widetilde{X}_N^2]}{\mathbb{E}[\widetilde{X}_N]^2} \leq 1 + \varepsilon^2 \delta + o(1).$$

By an application of Chebyshev's inequality we deduce

$$\mathbb{P}\left(\left|\widetilde{X}_N - \mathbb{E}[\widetilde{X}_N]\right| > \varepsilon \mathbb{E}[\widetilde{X}_N]\right) \leq \delta + o(1).$$

Note from the proof of Lemma 3.3 that

$$\mathbb{E}[\widetilde{X}_N] = \mathbb{E}\left[\frac{Z(C_{n,\ell}^{(r)})}{\mathbb{E}\left[Z(C_{n,\ell}^{(r)})\right] e^{\widetilde{Y}_N}}\right] = \mathbb{E}^*[e^{-\widetilde{Y}_N}] = (1 + o(1)) \exp\left(-\sum_{k=1}^{N} \frac{A_k c^{-k} e^{-k(r-\ell)}}{2(r-\ell)^2}\right)$$

which is bounded by a constant in r and ℓ. Since

$$\exp\left(-\sum_{k=1}^{N} \frac{A_k c^{-k} e^{-k(r-\ell)}}{2(r-\ell)^2}\right) - L =$$

$$\exp\left(-\sum_{k=1}^{N} \frac{A_k c^{-k} e^{-k(r-\ell)}}{2(r-\ell)^2}\right)\left(1 - \exp\left(\sum_{k=N}^{\infty} \frac{A_k c^{-k} e^{-k(r-\ell)}}{2(r-\ell)^2}\right)\right) \lesssim_{r,\ell} \varepsilon^2 \delta,$$

up to constant factors the third difference is greater than ε with probability at most δ.

For the middle difference, we estimate the probability

$$\mathbb{P}\left(\exp\left(|\widetilde{Y} - \widetilde{Y}_N|\right) > 1 + \varepsilon\right).$$

We decompose this event into three cases: either one of Y or Y_N is large, or the difference between Y and Y_N is large.

$$\mathbb{P}\left(\exp\left(|\widetilde{Y} - \widetilde{Y}_N|\right) > 1 + \varepsilon\right) \leq \mathbb{P}\left(Y \neq \widetilde{Y}\right) + \mathbb{P}\left(Y_N \neq \widetilde{Y}_N\right) + \mathbb{P}\left(|Y - Y_N| > \log(1 + \varepsilon)\right)$$

$$\lesssim_{r,\ell} \frac{e^{-N(r-\ell)}}{(\log(1 + \varepsilon))^2} + o(1) \leq \delta + o(1).$$

This implies that

$$\mathbb{P}\left(|\widetilde{X} - \widetilde{X}_N| > \varepsilon \widetilde{X}_N\right) \leq \delta + o(1).$$

Recall from above that with probability at least $1 - \delta$, $\widetilde{X}_N$ is close to $\mathbb{E}[\widetilde{X}_N]$ which is bounded by a constant. This yields

$$\mathbb{P}\left(|\widetilde{X} - \widetilde{X}_N| > \varepsilon \mathbb{E}[\widetilde{X}_N]\right) \leq 2\delta + o(1)$$

which is the desired bound on the middle difference.

Combining the estimates on all three terms, and since $\delta > 0$ was chosen arbitrarily, the claimed convergence follows. □

Proof of Theorem 1.2. By Lemma 3.2

$$Y_N \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \sum_{k=1}^{N} \frac{A_k c^{-k} e^{-k(r-\ell)}}{(r - \ell)^2}\right).$$

We also have from the proof of Lemma 4.1 that $\mathbb{E}\left[(Y - Y_N)^2\right] \lesssim_{r,\ell} e^{-N(r-\ell)}$, i.e. Y_N converges to Y in L^2. Thus, sending $N \to \infty$ we deduce that

$$Y \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \sum_{k=1}^{\infty} \frac{A_k c^{-k} e^{-k(r-\ell)}}{(r - \ell)^2}\right).$$

Thus, we know that

$$e^Y \xrightarrow{\mathrm{d}} \mathrm{Lognormal}\left(0, \sum_{k=1}^{\infty} \frac{A_k c^{-k} e^{-k(r-\ell)}}{(r - \ell)^2}\right).$$

The theorem follows from applying Lemma 4.1 and noting that multiplying by an exponential factor shifts the mean. □

References

[1] Emmanuel Abbe, Shuangping Li, and Allan Sly, *Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron*, 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science—FOCS 2021, IEEE Computer Soc., Los Alamitos, CA, [2022] ©2022, pp. 327–338. 3

[2] M. Ajtai, J. Komlós, and E. Szemerédi, *First occurrence of Hamilton cycles in random graphs*, Cycles in graphs (Burnaby, B.C., 1982), North-Holland Math. Stud., vol. 115, North-Holland, Amsterdam, 1985, pp. 173–178. 1

[3] Debapratim Banerjee, *Contiguity and non-reconstruction results for planted partition models: the dense case*, Electron. J. Probab. **23** (2018), Paper No. 18, 28. 3

[4] Béla Bollobás, *Threshold functions for small subgraphs*, Math. Proc. Cambridge Philos. Soc. **90** (1981), 197–206. 1

[5] Amin Coja-Oghlan, Charilaos Efthymiou, and Samuel Hetterich, *On the chromatic number of random regular graphs*, J. Combin. Theory Ser. B **116** (2016), 367–439. 3

[6] Amin Coja-Oghlan, Charilaos Efthymiou, Nor Jaafari, Mihyun Kang, and Tobias Kapetanopoulos, *Charting the replica symmetric phase*, Comm. Math. Phys. **359** (2018), 603–698. 3

[7] Andrzej Dudek and Alan Frieze, *Loose Hamilton cycles in random uniform hypergraphs*, Electron. J. Combin. **18** (2011), Paper 48, 14. 1

[8] Andrzej Dudek and Alan Frieze, *Tight Hamilton cycles in random uniform hypergraphs*, Random Structures Algorithms **42** (2013), 374–385. 1

[9] P. Erdős and A. Rényi, *On the existence of a factor of degree one of a connected random graph*, Acta Math. Acad. Sci. Hungar. **17** (1966), 359–368. 1

[10] Ehud Friedgut, *Sharp thresholds of graph properties, and the k-sat problem*, J. Amer. Math. Soc. **12** (1999), 1017–1054, With an appendix by Jean Bourgain. 1, 3

[11] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda, *Inapproximability of the partition function for the anti-ferromagnetic Ising and hard-core models*, Combin. Probab. Comput. **25** (2016), 500–559. 3

[12] Svante Janson, *Orthogonal decompositions and functional limit theorems for random graph statistics*, Mem. Amer. Math. Soc. **111** (1994), vi+78. 10

[13] Svante Janson, *Random regular graphs: asymptotic distributions and contiguity*, Combin. Probab. Comput. **4** (1995), 369–405. 3

[14] Anders Johansson, Jeff Kahn, and Van Vu, *Factors in random graphs*, Random Structures Algorithms **33** (2008), 1–28. 1

[15] János Komlós and Endre Szemerédi, *Limit distribution for the existence of Hamiltonian cycles in a random graph*, Discrete Math. **43** (1983), 55–63. 1

[16] Elchanan Mossel, Joe Neeman, and Allan Sly, *Reconstruction and estimation in the planted partition model*, Probab. Theory Related Fields **162** (2015), 431–461. 3

[17] Elchanan Mossel, Dror Weitz, and Nicholas Wormald, *On the hardness of sampling independent sets beyond the tree threshold*, Probab. Theory Related Fields **143** (2009), 401–439. 3

[18] Bhargav Narayanan and Mathias Schacht, *Sharp thresholds for nonlinear Hamiltonian cycles in hypergraphs*, Random Structures Algorithms **57** (2020), 244–255. 1, 2, 3, 4

[19] Jinyoung Park and Huy Tuan Pham, *A proof of the Kahn-Kalai conjecture*, J. Amer. Math. Soc. **37** (2024), 235–243. 1

[20] L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Math. **14** (1976), 359–364. 1

[21] Oliver Riordan, *Spanning subgraphs of random graphs*, Combin. Probab. Comput. **9** (2000), 125–148. 1

[22] R. W. Robinson and N. C. Wormald, *Almost all regular graphs are Hamiltonian*, Random Structures Algorithms **5** (1994), 363–374. 3