Enumeration of linear hypergraphs with given size and its applications

Fang Tian
Department of Applied Mathematics
Shanghai University of Finance and Economics

December 21, 2022

Introduction

- Let r and ℓ be given fixed integers such that $2 \leqslant \ell \leqslant r-1$. A hypergraph H on vertex set $[n]$ is an *r-uniform hypergraph* (*r-graph* for short) if each edge is a set of r vertices.
- An *r-graph* is said to be *linear* if every pair of distinct edges intersect in at most one vertex.
- Linear hypergraphs are the subject of much study, and one reason is that they are a natural generalization of simple graphs.
- An r-graph is called a partial Steiner (n, r, ℓ)-system, if every subset of size ℓ (*ℓ-set* for short) lies in at most one edge of H.
- In particular, partial Steiner $(n, r, 2)$-systems are linear hypergraphs.
- Let $\mathcal{H}_r(n, m)$ denote the set of r-graphs on the vertex set $[n]$ with m edges, and let $\mathcal{L}_r(n, m)$ denote the set of all linear hypergraphs in $\mathcal{H}_r(n, m)$.

- The *uniform hypergraph process* $\mathbb{H}_r(n, m)$ is a Markov process with time running through the set $\{0, 1, \cdots, \binom{n}{r}\}$. It is the typical random graph process $\mathbb{G}(n, m)$ introduced by Erdős and Rényi when $r = 2$.
- Similarly, the *partial Steiner* (n, r, ℓ)*-system process* begins with no edges on vertex set $[n]$ at time 0, all r-sets arrive one by one according to a uniformly chosen random permutation, and each one is added if and only if it does not overlap any of the previously added edges in ℓ or more vertices.
- In particular, it is the *linear hypergraph process* when $\ell = 2$.
- Let $\mathbb{L}_r^{\ell}(n, m)$ with $2 \leqslant \ell \leqslant r - 1$ denote the m-th stage of the uniform partial Steiner (n, r, ℓ)-system process, and $\mathbb{L}_r^2(n, m)$ is also denoted as $\mathbb{L}_r(n, m)$.

Introduction

- The hitting time of connectivity is a classic problem which has been extensively studied in the theory of random graph processes.
- Bollobás and Thomason in 1985 proved that, with probability approaching to 1 when $n \to \infty$ (*w.h.p.* for short), $m = \frac{n}{2} \log n$ is a sharp threshold of connectivity for $\mathbb{G}(n, m)$ and the very edge which links the last isolated vertex with another vertex makes the graph connected.
- Poole in 2015 proved the analogous result for $\mathbb{H}_r(n, m)$ when $r \geqslant 3$ is a fixed integer, which means that $m = \frac{n}{r} \log n$ is the hitting time of connectivity for $\mathbb{H}_r(n, m)$.

Introduction

- When working with random graphs (or random hypergraphs) with a given number of edges, Bollobas and Thomason (and Poole, respectively) could instead work in the binomial random graph $\mathbb{G}(n,p)$(and $\mathbb{H}_r(n,p)$,respectively).
- The proofs are due to the fact that the m-th stage $\mathbb{H}_r(n,m)$ can be identified with the uniform random hypergraph from $\mathcal{H}_r(n,m)$, and behaves in a similar fashion when m equals or is close to the expected number of edges of $\mathbb{H}_r(n,p)$.
- This approach does not work for linear hypergraphs, as choosing edges independently is very unlikely to result in a linear hypergraph.

Introduction

- It might be surmised that the threshold of connectivity for $\mathbb{L}_r^\ell(n,m)$ is smaller than the one for $\mathbb{H}_r(n,m)$ because of its constraint on r-graphs.
- Let $\tau_c = \min\{m : \mathbb{L}_r^\ell(n,m) \text{ is connected}\}$.
- Let $\tau_o = \min\{m : \mathbb{L}_r^\ell(n,m) \text{ has no isolated vertices}\}$.
- These two properties are certainly monotone increasing properties, so τ_c and τ_o are well-defined in $\mathbb{L}_r^\ell(n,m)$.
- For any fixed integers r and ℓ with $2 \leqslant \ell \leqslant r-1$, we will show that $\mathbb{L}_r^\ell(n,m)$ has the same threshold function of connectivity with $\mathbb{H}_r(n,m)$.
- And $\mathbb{L}_r^\ell(n,m)$ also becomes connected exactly at the moment when the last isolated vertex disappears.

Some Results

Theorem

For any fixed integers r and ℓ with $2 \leqslant \ell \leqslant r-1$, w.h.p., $m = \frac{n}{r}\log n$ is a sharp threshold of connectivity for $\mathbb{L}_r^\ell(n,m)$ and $\tau_c = \tau_o$ for $\mathbb{L}_r^\ell(n,m)$.

We also have a corollary about the distribution of the number of isolated vertices in $\mathbb{L}_r^\ell(n,m)$ when $m = \frac{n}{r}\left(\log n + c_n\right)$ and $c_n \to c \in \mathbb{R}$.

Corollary

For any fixed integers r and ℓ with $2 \leqslant \ell \leqslant r-1$, let $m = \frac{n}{r}\left(\log n + c_n\right)$ with $c_n \to c \in \mathbb{R}$. The number of isolated vertices in $\mathbb{L}_r^\ell(n,m)$ tends in distribution to the Poisson distribution with mean $\exp[-c]$.

We will rely on the enumeration results.

Our Approach

- We ever obtained the asymptotic enumeration formula for $\mathcal{L}_r(n,m)$ when $m = o(r^{-3}n^{\frac{3}{2}})$. In fact, we can apply exactly the same approach to obtain an asymptotic formula for $|\mathcal{L}_r^{\ell}(n,m)|$ when $3 \leqslant \ell \leqslant r-1$ and $m = o(n^{\frac{\ell+1}{2}})$.
- It turns out that the proof is a little easier when $\ell \geqslant 3$, as only one type of cluster needs to be considered, compared with four clusters in the case $\ell = 2$.
- Hence, the asymptotic expression when $\ell \geqslant 3$ is simpler than the corresponding expression when $\ell = 2$, so the statements cannot be combined.

Our Approach

Theorem

Let $r = r(n) \geqslant 3$ and $m = m(n)$ be integers with $m = o(r^{-3} n^{\frac{3}{2}})$. Then, as $n \to \infty$,

$$|\mathcal{L}_r(n,m)| = $$
$$= \frac{N^m}{m!} \exp\left[-\frac{[r]_2^2 [m]_2}{4n^2} - \frac{[r]_2^3 (3r^2 - 15r + 20) m^3}{24 n^4} + O\left(\frac{r^6 m^2}{n^3} \right) \right].$$

Theorem

For fixed integers r and ℓ such that $3 \leqslant \ell \leqslant r-1$, let $m = m(n)$ be an integer with $m = o(n^{\frac{\ell+1}{2}})$. Then, as $n \to \infty$,

$$|\mathcal{L}_r^\ell(n,m)| = \frac{N^m}{m!} \exp\left[-\frac{[r]_\ell^2 [m]_2}{2\ell!\, n^\ell} + O\left(\frac{m^2}{n^{\ell+1}} \right) \right].$$

Theorem

Let $r = r(n) \geq 3$ *and let* $pN = m_0$ *with* $m_0 = o(r^{-3} n^{\frac{3}{2}})$. *Then, as* $n \to \infty$,

$$\mathbb{P}[H_r(n,p) \in \mathcal{L}_r(n)] = \begin{cases} \exp\left[-\frac{[r]_2^2 m_0^2}{4n^2} + O\left(\frac{r^6 m_0^2}{n^3}\right)\right], & \text{if } m_0 = O(r^{-2}n); \\ \exp\left[-\frac{[r]_2^2 m_0^2}{4n^2} + \frac{[r]_2^3 (3r-5) m_0^3}{6n^4} + O\left(\frac{\log^3(r^{-2}n)}{\sqrt{m_0}} + \frac{r^6 m_0^2}{n^3}\right)\right], & \text{if } r^{-2}n \leq m_0 = o(r^{-3}n^{\frac{3}{2}}). \end{cases}$$

Our Approach

We obtain the probability that H contains a given hypergraph as a subhypergraph.

Theorem

Let $r = r(n) \geq 3$, $m = m(n)$ and $x = x(n)$ be integers with $m = o(r^{-3} n^{\frac{3}{2}})$ and $x = o\big(\frac{n^3}{r^6 m^2}\big)$. Let $X = X(n)$ be a given r-graph in $\mathcal{L}_r^\ell(n, x)$ and $H \in \mathcal{L}_r^\ell(n, m)$ be chosen uniformly at random. Then, as $n \to \infty$,

$$\mathbb{P}[X \subseteq H] = \frac{[m]_x}{N_0^x} \exp\left[\frac{[r]_\ell^2 x^2}{2\ell!\, n^\ell} + O\left(\frac{x}{n^\ell} + \frac{m^2 x}{n^{\ell+1}}\right)\right].$$

Proof of Main Results

Let

$$m_L = \frac{n}{r}(\log n - \omega(n)) \quad \text{and} \quad m_R = \frac{n}{r}(\log n + \omega(n)).$$

Lemma

Let H be chosen from $\mathcal{L}_r^\ell(n, m)$ uniformly at random. W.h.p. there are at most $2 \log n$ isolated vertices in H when $m = m_L$, while w.h.p. there are no isolated vertices in H when $m = m_R$. Thus, $\tau_o \in [m_L, m_R]$.

Lemma

If H is chosen uniformly at random from $\mathcal{L}_r^\ell(n, m_L)$, then w.h.p. H has at most $2 \log n$ isolated vertices and all remaining vertices are in a giant component.

Proof of Main Results

Let H be chosen uniformly at random from $\mathcal{L}_r^\ell(n, m_L)$. Assume that H consists of a connected component and at most $2\log n$ isolated vertices. Let V_1 denote the collection of these isolated vertices in H. We add $m_R - m_L$ random edges to H, which are denoted by $e_1, \cdots, e_{m_R - m_L}$ in sequence. If $\tau_o < \tau_c$ then at least one edge e_j for $1 \leqslant j \leqslant m_R - m_L$ must be added which contains only isolated vertices.

If $H_{m_R - m_L}$ is chosen uniformly at random from $\mathcal{L}_r^\ell(n, m_R)$, then we have

$$\begin{aligned}
\mathbb{P}[\tau_o < \tau_c] &\leqslant o(1) + (m_R - m_L)\binom{2\log n}{r}\frac{m_R}{N_0}\exp\left[O\left(\frac{1}{n^\ell} + \frac{m_R^2}{n^{\ell+1}}\right)\right] \\
&= o(1) + O\left(\frac{n^2(\log n)^{r+1}\log\log n}{N_0}\right) \\
&= o(1).
\end{aligned}$$

We have *w.h.p.* $\mathcal{L}_r^\ell(n, m_R)$ is connected.

For $n \geq 3$, let $r = r(n) \geq 3$, $m = o(r^{-3}n^{3/2})$ and $t = t(n) = \min\{m, o(\frac{n^3}{r^6 m^2})\}$. The expected number of hypertrees with t edges in an r-uniform linear hypergraph with m edges is

$$\mathbb{E}(|T|) = \frac{(rt - t + 1)^{t-2} r^t [m]_t}{n^{t-1} t!} \exp\Big[\frac{[r]_2^2 t^2}{4n^2} - \frac{(r-1)^2 [t]_2}{2n} + O\Big(\frac{r^4 t}{n^2} + \frac{r^6 m^2 t}{n^3}\Big)\Big].$$

The expected number of matchings with t edges .... the expected number of loose cycles with t edges....

Further Problems

We show the process $\mathbb{L}_r^{\ell}(n,m)$ has the same threshold of connectivity with $\mathbb{H}_r(n,m)$. What about other extremal properties of the partial Steiner (n,r,ℓ)-systems process? For any fixed integer $g \geqslant 4$, some researchers applied a natural constrained random process to typically produce a partial Steiner $(n,3,2)$-system with $(1/6 - o(1))n^2$ edges and girth larger than g. The process iteratively adds random 3-set subject to the constraint that the girth remains larger than g. In future work, we will consider the final size of the partial Steiner (n,r,ℓ)-system process with some constraints on the girth.

Thank You