Regular subgraph counts in the Poisson regime

Mriganka Basu Roy Chowdhury

May 22, 2023

- Let $\mathcal{G}(n, p)$ be the Erdős–Rényi graph with n vertices and connection probability p, and let $G \sim \mathcal{G}(n, p)$ be a sample from this distribution.

- Let $\mathcal{G}(n, p)$ be the Erdős–Rényi graph with n vertices and connection probability p, and let $G \sim \mathcal{G}(n, p)$ be a sample from this distribution.
- H is a fixed small connected graph with q vertices, and let $Q_H(G)$ be the number of copies of H in G (up to isomorphism).

- Let $\mathcal{G}(n, p)$ be the Erdős–Rényi graph with n vertices and connection probability p, and let $G \sim \mathcal{G}(n, p)$ be a sample from this distribution.
- H is a fixed small connected graph with q vertices, and let $Q_H(G)$ be the number of copies of H in G (up to isomorphism).

Question

How does $Q_H(G)$ behave under various regimes of $p = p_n$?

- Let $\mathcal{G}(n, p)$ be the Erdős–Rényi graph with n vertices and connection probability p, and let $G \sim \mathcal{G}(n, p)$ be a sample from this distribution.
- H is a fixed small connected graph with q vertices, and let $Q_H(G)$ be the number of copies of H in G (up to isomorphism).

Question

How does $Q_H(G)$ behave under various regimes of $p = p_n$?

For example, if H is just an edge, it is clear that this is asymptotically normal (when p_n is large enough) since it is just a sum of i.i.d. random variables. But things are nontrivial when H is a general subgraph.

Subgraph counts

We know the asymptotic distributions of $Q_H(G)$ in many cases:

- [Ruc88] shows that under certain conditions, $Q_H(G)$ is *asymptotically normal*.

We know the asymptotic distributions of $Q_H(G)$ in many cases:

- [Ruc88] shows that under certain conditions, $Q_H(G)$ is *asymptotically normal*.
- However, this regime must necessarily exclude the case when $\mathbb{E}Q_H(G) = \Theta(1)$, where a natural guess for the asymptotic distribution is Poisson.

We know the asymptotic distributions of $Q_H(G)$ in many cases:

- [Ruc88] shows that under certain conditions, $Q_H(G)$ is *asymptotically normal*.
- However, this regime must necessarily exclude the case when $\mathbb{E}Q_H(G) = \Theta(1)$, where a natural guess for the asymptotic distribution is Poisson. Although this is not true in general, [Bol81] established this for the large class of *strictly balanced graphs*.

We know the asymptotic distributions of $Q_H(G)$ in many cases:

- [Ruc88] shows that under certain conditions, $Q_H(G)$ is *asymptotically normal*.
- However, this regime must necessarily exclude the case when $\mathbb{E}Q_H(G) = \Theta(1)$, where a natural guess for the asymptotic distribution is Poisson. Although this is not true in general, [Bol81] established this for the large class of *strictly balanced graphs*.

Definition (density, strictly balanced graphs)

Define the ***density*** *of a graph* H *to be* $d(H) = e(H)/v(H)$. *A graph* H *is* ***strictly balanced*** *if every proper subgraph* $H' \subsetneq H$ *satisfies* $d(H') < d(H)$. *Note that every connected regular graph is strictly balanced.*

We know the asymptotic distributions of $Q_H(G)$ in many cases:

- [Ruc88] shows that under certain conditions, $Q_H(G)$ is *asymptotically normal*.
- However, this regime must necessarily exclude the case when $\mathbb{E}Q_H(G) = \Theta(1)$, where a natural guess for the asymptotic distribution is Poisson. Although this is not true in general, [Bol81] established this for the large class of *strictly balanced graphs*.

Definition (density, strictly balanced graphs)

Define the ***density*** *of a graph H to be $d(H) = e(H)/v(H)$. A graph H is* ***strictly balanced*** *if every proper subgraph $H' \subsetneq H$ satisfies $d(H') < d(H)$. Note that every connected regular graph is strictly balanced.*

In fact, the criterion in [Ruc88] also depends on the related quantity

Since asymptotic distributions tell us little about the finite sample behavior, let us now ask the following:

Question

For a sequence $k_n \to \infty$, how does $\mathbb{P}(Q_H(G) \geq k_n)$ behave?

The regime which has received maximum attention is when $k_n = C \cdot \mathbb{E}[Q_H(G)]$ for $C > 1$.

Since asymptotic distributions tell us little about the finite sample behavior, let us now ask the following:

Question

For a sequence $k_n \to \infty$, how does $\mathbb{P}(Q_H(G) \geq k_n)$ behave?

The regime which has received maximum attention is when $k_n = C \cdot \mathbb{E}[Q_H(G)]$ for $C > 1$.

This problem turned out to be significantly harder than its predecessors.

- The seminal work of Chatterjee and Varadhan [CV11] achieved asymptotically sharp rates on this problem when p is a constant, using the theory of graphons (graph limits) and the Szemerédi regularity lemma, by reducing the problem to a natural "mean-field" variational problem.

Since asymptotic distributions tell us little about the finite sample behavior, let us now ask the following:

Question

For a sequence $k_n \to \infty$, how does $\mathbb{P}(Q_H(G) \geq k_n)$ behave?

The regime which has received maximum attention is when $k_n = C \cdot \mathbb{E}[Q_H(G)]$ for $C > 1$.

This problem turned out to be significantly harder than its predecessors.

- The seminal work of Chatterjee and Varadhan [CV11] achieved asymptotically sharp rates on this problem when p is a constant, using the theory of graphons (graph limits) and the Szemerédi regularity lemma, by reducing the problem to a natural "mean-field" variational problem.

- Due to poor quantitative estimates in the Szemerédi regularity lemma, [CV11] could only be extended to $p_n \gtrsim (\log n)^{-\varepsilon}$ for some small ε.

- Due to poor quantitative estimates in the Szemerédi regularity lemma, [CV11] could only be extended to $p_n \gtrsim (\log n)^{-\varepsilon}$ for some small ε.
- The arguments of Chatterjee and Dembo [CD16] developed new technology to deal with sparser cases, which for the case of the triangle extended previously known results down to $p_n \geq n^{-1/42} \log n$.

- Due to poor quantitative estimates in the Szemerédi regularity lemma, [CV11] could only be extended to $p_n \gtrsim (\log n)^{-\varepsilon}$ for some small ε.
- The arguments of Chatterjee and Dembo [CD16] developed new technology to deal with sparser cases, which for the case of the triangle extended previously known results down to $p_n \geq n^{-1/42} \log n$.
- This was however not satisfactory because one expects the mean-field variational problem to hold as long as $p_n \gg \frac{\log n}{n}$. After a sequence of follow-up works, finally in 2019, a breakthrough by Harel, Mousset and Samotij essentially solved the upper tail problem in the regime $p_n \gg 1/n$ using ideas inspired by classical moment method arguments of [JOR04].

Subgraph counts: $p_n = \Theta(1/n)$

- The fully sparse case of $p_n = \Theta(1/n)$, goes back to a famous question due to Aldous, who had asked about the structure of graphs conditioned on having many triangles.

Subgraph counts: $p_n = \Theta(1/n)$

- The fully sparse case of $p_n = \Theta(1/n)$, goes back to a famous question due to Aldous, who had asked about the structure of graphs conditioned on having many triangles. This, along with the related question of the upper tail probability, was answered by the works of Ganguly, Hiesmayr and Nam [GHN22] and Chakraborty, van der Hofstad and den Hollander [CHH21].

Subgraph counts: $p_n = \Theta(1/n)$

- The fully sparse case of $p_n = \Theta(1/n)$, goes back to a famous question due to Aldous, who had asked about the structure of graphs conditioned on having many triangles. This, along with the related question of the upper tail probability, was answered by the works of Ganguly, Hiesmayr and Nam [GHN22] and Chakraborty, van der Hofstad and den Hollander [CHH21].
- Note that $k_n = C \cdot \mathbb{E}[Q_H(G)]$ is no longer of interest because the expectation is $\Theta(1)$. Instead, k_n is chosen to be an arbitrary increasing function of n and upper tail behavior along these sequences was examined.

In [GHN22], a very interesting phenomenon was observed:

Subgraph counts: transition in tail behavior

In [GHN22], a very interesting phenomenon was observed:

Theorem (Ganguly, Hiesmayr, Nam '22, *simplified*)

If $k_n^{1/3} \log k_n < \Theta(\log n)$, in the upper tail $\{Q_{K_3}(G) \geq k_n\}$, we observe almost k_n disjoint triangles. If $k_n^{1/3} \log k_n > \Theta(\log n)$, we observe an almost clique containing almost all the excess triangles. In fact,

$$\mathbb{P}(Q_{K_3}(G) \geq k_n) = \Theta(\exp(-C \min(k_n \log k_n, k_n^{2/3} \log n)))$$

The first term is a Poisson tail, and the second one is due to the occurrence of a clique of the correct size.

Note: The result in [GHN22] is much sharper, with exact thresholds and exact exponent in the probability up to smaller order terms.

In [Cho23], we extend the above to all regular graphs.

In [Cho23], we extend the above to all regular graphs. Here H is a connected regular graph with degree Δ and q vertices. Note that the critical p_n to have $\Theta(1)$ copies of H in expectation is now $p_n = \Theta(n^{-2/\Delta})$.

In [Cho23], we extend the above to all regular graphs. Here H is a connected regular graph with degree Δ and q vertices. Note that the critical p_n to have $\Theta(1)$ copies of H in expectation is now $p_n = \Theta(n^{-2/\Delta})$.

Theorem (BRC, '23)

For any $k_n \geq 2$ and $p_n = n^{-2/\Delta}$, we have

$$C_1 \exp\left(-C_2 \min(k_n \log k_n, k_n^{2/q} \log n)\right) \leq \mathbb{P}(Q_H(G) \geq k_n) \leq C_3 \exp\left(-C_4 \min(k_n \log k_n, k_n^{2/q} \log n)\right)$$

for constants C_1, C_2, C_3, C_4 depending only on H.

Here C_i are not necessarily optimal.

It is possible to handle certain classes of H via direct combinatorial arguments (like when H is a cycle) which are essentially local. But due to the absence of a general structure among all regular graphs, a global approach is needed. A key tool we use in our analysis is the following result due to [Fin92]:

It is possible to handle certain classes of H via direct combinatorial arguments (like when H is a cycle) which are essentially local. But due to the absence of a general structure among all regular graphs, a global approach is needed. A key tool we use in our analysis is the following result due to [Fin92]:

Lemma (Finner's inequality)

Let $\mu_1, \mu_2, \ldots$ be probability measures on $\Omega_1, \Omega_2, \ldots$ respectively, and let $\Omega = \prod_{i=1}^{n} \Omega_i$, $\mu = \prod_{i=1}^{n} \mu_i$. Suppose $A_1, A_2, \ldots, A_m$ are nonempty subsets of $[n] = \{1, \ldots, n\}$, and for any set A, set $\mu_A = \prod_{i \in A} \mu_i$ and $\Omega_A = \prod_{i \in A} \Omega_i$. If $f_i \in L^{p_i}(\Omega_{A_i}, \mu_{A_i})$ (with $p_i \geq 1$) for each $i \in [m]$ such that

$$\sum_{i : x \in A_i} p_i^{-1} \leq 1, \quad \forall x \in [n] \tag{1}$$

then we have the inequality

$$\int \prod_{i=1}^{m} |f_i| d\mu \leq \prod_{i=1}^{m} \left(\int |f_i|^{p_i} d\mu_{A_i} \right)^{1/p_i}$$

Finner's inequality - II

- Finner's inequality generalizes Hölder's inequality. We can put $A_1 = A_2 = [n]$, and $p_1 = p$, $p_2 = q$ with $p^{-1} + q^{-1} = 1$, and check that all the conditions are satisfied.

Finner's inequality - II

- Finner's inequality generalizes Hölder's inequality. We can put $A_1 = A_2 = [n]$, and $p_1 = p$, $p_2 = q$ with $p^{-1} + q^{-1} = 1$, and check that all the conditions are satisfied.
- It also generalizes the celebrated Loomis-Whitney inequality. Consider d functions $g_i : \mathbb{R}^{d-1} \to \mathbb{R}$ such that $g_i \in L^{d-1}(\mathbb{R}^{d-1})$. We are interested in the function $f : \mathbb{R}^d \to \mathbb{R}$

$$f(x_1, \ldots, x_d) \stackrel{\text{def}}{=} \prod_{i=1}^{d} g_i(x_{-i}), \quad x_{-i} \stackrel{\text{def}}{=} (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$$

- Finner's inequality generalizes Hölder's inequality. We can put $A_1 = A_2 = [n]$, and $p_1 = p$, $p_2 = q$ with $p^{-1} + q^{-1} = 1$, and check that all the conditions are satisfied.
- It also generalizes the celebrated Loomis-Whitney inequality. Consider d functions $g_i : \mathbb{R}^{d-1} \to \mathbb{R}$ such that $g_i \in L^{d-1}(\mathbb{R}^{d-1})$. We are interested in the function $f : \mathbb{R}^d \to \mathbb{R}$

$$f(x_1, \ldots, x_d) \stackrel{\text{def}}{=} \prod_{i=1}^{d} g_i(x_{-i}), \quad x_{-i} \stackrel{\text{def}}{=} (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$$

Set $A_i = [n] - \{i\}$ and $p_i = d - 1$. Then this falls into the framework of Finner's inequality, yielding the Loomis-Whitney inequality:

$$\|f\|_{L^1(\mathbb{R}^d)} \leq \prod_{i=1}^{d} \|g_i\|_{L^{d-1}(\mathbb{R}^{d-1})}, \qquad \text{[LW49]}$$

A prototypical application for us is where say $\mathfrak{g}$ is a graphon, and H is a graph. We wish to find the homomorphism density of H in $\mathfrak{g}$. Then, we wish to bound

$$t_H(\mathfrak{g}) \stackrel{\text{def}}{=} \int_{[0,1]^{v(H)}} \prod_{(u,v)\in E(H)} \mathfrak{g}(t_u, t_v) \prod_{u\in V(H)} dt_u$$

A prototypical application for us is where say $\mathfrak{g}$ is a graphon, and H is a graph. We wish to find the homomorphism density of H in $\mathfrak{g}$. Then, we wish to bound

$$t_H(\mathfrak{g}) \stackrel{\text{def}}{=} \int_{[0,1]^{v(H)}} \prod_{(u,v)\in E(H)} \mathfrak{g}(t_u, t_v) \prod_{u\in V(H)} dt_u$$

This falls into the framework of the above lemma if

- $\Omega_v = [0,1], \mu_v = \text{Unif}[0,1], \quad \forall v \in V(H)$
- $A_{e=(u,v)} = \{u, v\}, \quad \forall e \in E(H)$.
- $f_{e=(u,v)}(s,t) = \mathfrak{g}(s,t), \quad \forall e \in E(H)$.

A prototypical application for us is where say $\mathfrak{g}$ is a graphon, and H is a graph. We wish to find the homomorphism density of H in $\mathfrak{g}$. Then, we wish to bound

$$t_H(\mathfrak{g}) \stackrel{\text{def}}{=} \int_{[0,1]^{v(H)}} \prod_{(u,v)\in E(H)} \mathfrak{g}(t_u, t_v) \prod_{u \in V(H)} dt_u$$

This falls into the framework of the above lemma if

- $\Omega_v = [0,1], \mu_v = \text{Unif}[0,1], \quad \forall v \in V(H)$
- $A_{e=(u,v)} = \{u, v\}, \quad \forall e \in E(H).$
- $f_{e=(u,v)}(s,t) = \mathfrak{g}(s,t), \quad \forall e \in E(H).$

If H is Δ-regular, one may choose $p_i = \Delta$ for all i.

A prototypical application for us is where say $\mathfrak{g}$ is a graphon, and H is a graph. We wish to find the homomorphism density of H in $\mathfrak{g}$. Then, we wish to bound

$$t_H(\mathfrak{g}) \stackrel{\text{def}}{=} \int_{[0,1]^{v(H)}} \prod_{(u,v)\in E(H)} \mathfrak{g}(t_u, t_v) \prod_{u \in V(H)} dt_u$$

This falls into the framework of the above lemma if

- $\Omega_v = [0,1], \mu_v = \text{Unif}[0,1], \quad \forall v \in V(H)$
- $A_{e=(u,v)} = \{u, v\}, \quad \forall e \in E(H).$
- $f_{e=(u,v)}(s,t) = \mathfrak{g}(s,t), \quad \forall e \in E(H).$

If H is Δ-regular, one may choose $p_i = \Delta$ for all i. For our purposes, we rely on this lemma to extract bounds on expected subgraph counts, expected subgraph counts containing given edges etc.

A straightforward but crucial consequence of this is the following lemma:

Lemma

Let G be a graph with E edges. Then,

$$Q_H(G) \leq C \cdot E^{q/2}$$

where $C = C(H)$ only depends on H.

A straightforward but crucial consequence of this is the following lemma:

Lemma

Let G be a graph with E edges. Then,

$$Q_H(G) \leq C \cdot E^{q/2}$$

where $C = C(H)$ only depends on H.

When H is a cycle, this may be done via direct spectral arguments, but those arguments do not extend to the regular case.

Main steps of the proof

The harder part of the proof is the upper bound on the probability, and requires several steps. The proof is essentially divided into two parts:

The harder part of the proof is the upper bound on the probability, and requires several steps. The proof is essentially divided into two parts:

- If $k_n \leq n^{\varepsilon}$ where $\varepsilon = \varepsilon(H)$ is a constant depending on H, the technique we use generalizes the ideas in [GHN22]. Crucially, in this part of the proof, we recover the transition at $k_n^{1-2/q} \log k_n = \Theta(\log n)$.

Main steps of the proof

The harder part of the proof is the upper bound on the probability, and requires several steps. The proof is essentially divided into two parts:

- If $k_n \leq n^{\varepsilon}$ where $\varepsilon = \varepsilon(H)$ is a constant depending on H, the technique we use generalizes the ideas in [GHN22]. Crucially, in this part of the proof, we recover the transition at $k_n^{1-2/q} \log k_n = \Theta(\log n)$.
- When $k_n \gg \operatorname{poly}\log n$, we use a modified version of the arguments due to [HMS22] and [CHH21].
- Together, these cases cover all possible values of k_n.

- When k_n is small i.e. $\leq n^{\varepsilon}$, following [GHN22] we divide the graph into subgraphs spanned by copies of H.

- When k_n is small i.e. $\leq n^{\varepsilon}$, following [GHN22] we divide the graph into subgraphs spanned by copies of H. Let S be such a spanned graph. The probability of it occurring in G is

$$\binom{n}{v(S)} n^{-\frac{2}{\Delta}\cdot e(S)} \leq n^{-\left(\frac{2e(S)}{\Delta} - v(S)\right)}$$

- When k_n is small i.e. $\leq n^{\varepsilon}$, following [GHN22] we divide the graph into subgraphs spanned by copies of H. Let S be such a spanned graph. The probability of it occurring in G is

$$\binom{n}{v(S)} n^{-\frac{2}{\Delta}\cdot e(S)} \leq n^{-\left(\frac{2e(S)}{\Delta} - v(S)\right)}$$

- Note that if $\Delta = 2$ like in a triangle, then this is $e(S) - v(S)$ which can be interpreted via the number of excess edges in S after choosing a spanning tree, which was crucially used in [GHN22].

- Unfortunately, the previous input is not available to us, and we instead prove a lemma showing that if S has at least $\ell \geq 2$ copies of H, then it must satisfy

$$\frac{2e(S)}{\Delta} - v(S) \geq Ce(S), \quad C = C(H).$$

  restoring control on the probability in terms of the number of copies of H in it, because $e(S) \gtrsim \ell^{2/q}$.

- Unfortunately, the previous input is not available to us, and we instead prove a lemma showing that if S has at least $\ell \geq 2$ copies of H, then it must satisfy

$$\frac{2e(S)}{\Delta} - v(S) \geq Ce(S), \quad C = C(H).$$

  restoring control on the probability in terms of the number of copies of H in it, because $e(S) \gtrsim \ell^{2/q}$.
- Now consider a particular configuration of assigning k_n copies of H to $s + m$ spanned components, where s of them contain a single copy of H, and the remaining m contain $\ell_1, \ldots, \ell_m$ copies of H (so that $s + \sum \ell_i \geq k_n$).

- Unfortunately, the previous input is not available to us, and we instead prove a lemma showing that if S has at least $\ell \geq 2$ copies of H, then it must satisfy

$$\frac{2e(S)}{\Delta} - v(S) \geq Ce(S), \quad C = C(H).$$

restoring control on the probability in terms of the number of copies of H in it, because $e(S) \gtrsim \ell^{2/q}$.

- Now consider a particular configuration of assigning k_n copies of H to $s + m$ spanned components, where s of them contain a single copy of H, and the remaining m contain $\ell_1, \ldots, \ell_m$ copies of H (so that $s + \sum \ell_i \geq k_n$). Then the probability that this particular configuration occurs is, by BK inequality, at most

- Due to the concavity of $\ell \mapsto \ell^{2/q}$, we see that it is not optimal to have multiple spanned components with ≥ 2 copies of H. This competes with the probability of s disjoint copies of H, resulting in the disjoint v.s. clique competition we see in the result.

- Due to the concavity of $\ell \mapsto \ell^{2/q}$, we see that it is not optimal to have multiple spanned components with ≥ 2 copies of H. This competes with the probability of s disjoint copies of H, resulting in the disjoint v.s. clique competition we see in the result.
- There is some work involved in controlling the entropy of these assignments, choices of spanned components and such, but I omit these details.

Main ideas: Many copies

- When $k_n \gg \operatorname{poly}\log n$, the idea is the following: look for planted graphs which increase the number of copies of H in G, and are *also present in G w.h.p.* if it does have k_n copies of H.

Main ideas: Many copies

- When $k_n \gg \operatorname{poly}\log n$, the idea is the following: look for planted graphs which increase the number of copies of H in G, and are *also present in G w.h.p.* if it does have k_n copies of H.
- These structures were defined by Harel, Mousset and Samotij [HMS22], and were called *seeds and cores*. It is known that every seed contains a core.

Main ideas: Many copies

- When $k_n \gg \operatorname{poly}\log n$, the idea is the following: look for planted graphs which increase the number of copies of H in G, and are *also present in* G *w.h.p.* if it does have k_n copies of H.
- These structures were defined by Harel, Mousset and Samotij [HMS22], and were called *seeds and cores*. It is known that every seed contains a core.
- The definitions are somewhat technical, but the takeaway is that these objects are modeled after cliques, but have a certain slack to them.

Main ideas: Many copies

- When $k_n \gg \text{poly}\log n$, the idea is the following: look for planted graphs which increase the number of copies of H in G, and are *also present in G w.h.p.* if it does have k_n copies of H.
- These structures were defined by Harel, Mousset and Samotij [HMS22], and were called *seeds and cores*. It is known that every seed contains a core.
- The definitions are somewhat technical, but the takeaway is that these objects are modeled after cliques, but have a certain slack to them.
- The following crucial theorem justifies their importance:

Lemma

$$\mathbb{P}(Q_H(G) \geq k_n) \leq (1+o(1)) \cdot \mathbb{P}(G_n \textit{ has a seed}) = (1+o(1)) \cdot \mathbb{P}(G_n \textit{ has a core})$$

- This is proved via computing very high moments of $X \stackrel{\text{def}}{=} Q_H(G)Z$ where $Z \stackrel{\text{def}}{=} 1_{G \text{ has no seed}}$ (in fact the moments are of order $\widetilde{\Omega}(k_n^{2/q})$).

Lemma

$\mathbb{P}(Q_H(G) \geq k_n) \leq (1 + o(1)) \cdot \mathbb{P}(G_n \textit{ has a seed}) = (1 + o(1)) \cdot \mathbb{P}(G_n \textit{ has a core})$

- This is proved via computing very high moments of $X \stackrel{\text{def}}{=} Q_H(G)Z$ where $Z \stackrel{\text{def}}{=} 1_{G \text{ has no seed}}$ (in fact the moments are of order $\widetilde{\Omega}(k_n^{2/q})$).
- Using the above, our task is reduced to analyzing cores. To perform this, we prove several structural results about cores (almost all of which require novel applications of Finner's inequality). Finally to finish the proof, we combine all these inputs via a multiscale decomposition argument to control the various sources of entropy, but I skip these details.

Thank You!

Reference I

[Bol81] Béla Bollobás. "Threshold functions for small subgraphs". In: *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 90. 2. Cambridge University Press. 1981, pp. 197–206.

[CD16] Sourav Chatterjee and Amir Dembo. "Nonlinear large deviations". In: *Advances in Mathematics* 299 (2016), pp. 396–450.

[CHH21] Suman Chakraborty, Remco van der Hofstad, and Frank den Hollander. "Sparse random graphs with many triangles". In: *arXiv preprint arXiv:2112.06526* (2021).

[Cho23] Mriganka Basu Roy Chowdhury. "Universality in prelimiting tail behavior for regular subgraph counts in the Poisson regime". In: *arXiv preprint arXiv:2304.01162* (2023).

[CV11] Sourav Chatterjee and SR Srinivasa Varadhan. "The large deviation principle for the Erdős-Rényi random graph". In: *European Journal of Combinatorics* 32.7 (2011), pp. 1000–1017.

[Fin92] Helmut Finner. "A generalization of Holder's inequality and some probability inequalities". In: *The Annals of probability* (1992), pp. 1893–1901.

Reference II

[GHN22] Shirshendu Ganguly, Ella Hiesmayr, and Kyeongsik Nam. "Upper tail behavior of the number of triangles in random graphs with constant average degree". In: *arXiv preprint arXiv:2202.06916* (2022).

[HMS22] Matan Harel, Frank Mousset, and Wojciech Samotij. "Upper tails via high moments and entropic stability". In: *Duke Mathematical Journal* 1.1 (2022), pp. 1–104.

[JOR04] Svante Janson, Krzysztof Oleszkiewicz, and Andrzej Ruciński. "Upper tails for subgraph counts in random graphs". In: *Israel Journal of Mathematics* 142 (2004), pp. 61–92.

[LW49] Lynn H Loomis and Hassler Whitney. "An inequality related to the isoperimetric inequality". In: (1949).

[Ruc88] Andrzej Ruciński. "When are small subgraphs of a random graph normally distributed?" In: *Probability Theory and Related Fields* 78.1 (1988), pp. 1–10.

Definition (Seeds and cores – Harel, Mousset, Samotij)

Let $S \subseteq K_n$ be a subset of edges. We call S a if

- $\mathbb{E}_S Q_H(G) \stackrel{def}{=} \mathbb{E}\left[Q_H(G) \mid S \subseteq E(G)\right] \geq$,

Definition (Seeds and cores – Harel, Mousset, Samotij)

Let $S \subseteq K_n$ *be a subset of edges. We call* S *a seed if*

- $\mathbb{E}_S Q_H(G) \stackrel{def}{=} \mathbb{E}\left[Q_H(G) \mid S \subseteq E(G)\right] \geq (1-\varepsilon_n)k_n,$
- $e(S) \leq \left[C_s \varepsilon_n^{-1} \log p_n\right] k_n^{2/q}$

Definition (Seeds and cores – Harel, Mousset, Samotij)

Let $S \subseteq K_n$ be a subset of edges. We call S a seed if

- $\mathbb{E}_S Q_H(G) \overset{def}{=} \mathbb{E}\left[Q_H(G) \mid S \subseteq E(G)\right] \geq (1-\varepsilon_n)k_n,$
- $e(S) \leq \left[C_s \varepsilon_n^{-1} \log p_n\right] k_n^{2/q}$

Due to the huge entropy of seeds, we define cores which also have the following condition

- $\mathbb{E}_S Q_H(G) - \mathbb{E}_{S-e} Q_H(G) \geq \left[\frac{\varepsilon_n^2}{C_s \log(1/p_n)}\right] k_n^{1-2/q}, \quad \forall e \in S.$

i.e., every edge contributes significantly. It is not hard to show that every seed has a core.

The model we are trying to emulate here are cliques, but with some (logarithmic) relaxation.

Definition (Seeds and cores – Harel, Mousset, Samotij)

Let $S \subseteq K_n$ be a subset of edges. We call S a core if

- $\mathbb{E}_S Q_H(G) \stackrel{def}{=} \mathbb{E}\left[Q_H(G) \mid S \subseteq E(G)\right] \geq (1 - 2\varepsilon_n)k_n,$
- $e(S) \leq \left[C_s \varepsilon_n^{-1} \log p_n\right] k_n^{2/q}$

Due to the huge entropy of seeds, we define cores which also have the following condition

- $\mathbb{E}_S Q_H(G) - \mathbb{E}_{S-e} Q_H(G) \geq \left[\frac{\varepsilon_n^2}{C_s \log(1/p_n)}\right] k_n^{1-2/q}, \quad \forall e \in S.$

i.e., every edge contributes significantly. It is not hard to show that every seed has a core.

The model we are trying to emulate here are cliques, but with some (logarithmic) relaxation.