Overview of Hladký et al's (2021) Work on Inhomogeneous W-Random Graphs

Ernest Ng

Supervised by Anirban Chatterjee & Prof. Bhaswar Bhattacharya

May 12, 2022

Hladký et al. (2021) prove a limit theorem of the number of r-cliques in W-random graphs, which are an inhomogeneous variant of the Erdős-Rényi random graph. In this report, we discuss some of the relevant background in graphon theory required to understand Hladký et al's results, and provide a high-level overview of their main results.

Introduction to Graphons

Definition 1. *A **graphon** is a bounded, symmetric and measurable function*

$$W : [0,1]^2 \to [0,1], \quad W(x,y) = W(y,x)\ \forall\ x,y \in [0,1]$$

Let $\mathcal{W}_0$ denote the space of all graphons.

Intuitively, we may think of graphons as weighted symmetric graphs with uncountably many vertices, where the vertex set is $[0,1]$ and the weights are the values $W(x,y) = W(y,x)$.

Note that for any graph $G = (V,E)$, the associated empirical graphon $W^G \in \mathcal{W}_0$ if $w_e \in [0,1]$ for all $e \in E$.

Inhomogeneous Random Graphs

Given a graphon W, we generate the random graph $\mathbb{G}(n,W)$ as follows:

1. Sample independently n numbers $U_1,\dots,U_n \sim \text{Unif}(0,1)$. Call these numbers **types** (continuous analog of node colorings).

2. Identify each uniform random variable U_j with a node $j \in [1..n]$, i.e. assign each node a type.

3. Connect any two nodes i, j in $\mathbb{G}(n, W)$ with an edge (i, j) with probability $W(U_i, U_j)$

Call such a random graph a W-**random graph**.

From this construction, note that if the graphon W is constant, i.e. $W(x, y) \equiv p \in [0, 1]$, then $\mathbb{G}(n, W)$ is identical to the Erdös–Rényi random graph $\mathbb{G}(n, p)$.

Definition 2. *For a graphon W, the **degree function** $\deg_W : [0,1] \to [0,1]$ is defined as:*

$$\deg_W(x) = \int_0^1 W(x, y)\, dy$$

The degree function allows us to examine how the degree of a node varies as its type changes.

Recall that in an Erdos–Rényi random graph $\mathbb{G}(n, p)$, a node has expected degree $(n-1)\cdot p$. In $\mathbb{G}(n, W)$, if a node has type $x \in [0, 1]$, then its expected degree is $(n-1)\cdot \deg_W(x)$. Thus, we see that the degree function of a graphon generalizes the notion of the degree of a node for graphons.

Graph Homomorphisms and Homomorphism Density

Definition 3. *Let $F = (V', E')$ and $G = (V, E)$ be graphs. A **graph homomorphism** from F to G is a map*

$$\beta : V' \to V \quad \text{such that if } (i, j) \in E', \text{ then } (\beta(i), \beta(j)) \in E.$$

Write $F \to G$ if there exists a homomorphism from F to G.

*A graph homomorphism $F \to F$ that is bijective is called a **graph automorphism**.*

The intuition for graph homomorphisms is that it is a map $F \to G$ where the images of adjacent vertices remain adjacent. In particular, a homomorphism $K_r \to G$ indicates that G contains an r-clique.

Now, note that given any F and G, there may exist many possible homomorphisms $F \to G$. This motivates the notion of *homomorphism densities*.

Definition 4. *For a weighted graph $G = (V, E)$ on n nodes with adjacency matrix A, and a graph $F = (V', E')$ on k nodes, the **homomorphism density** of F in G is defined as:*

$$t(F, G) = \frac{1}{n^k} \sum_{\substack{\beta: V' \to V \\ \text{graph hom.}}} \left(\prod_{(i,j) \in E'} [A]_{\beta(i), \beta(j)} \right)$$

where $[A]_{\beta(i),\beta(j)}$ denotes the $(\beta(i), \beta(j))$-th entry of A.

Homomorphism densities are a relative measure of the number of ways in which F can be mapped into G in an adjacency-preserving manner. In particular, in the definition above, we weight each homomorphism $\beta : V' \to V$ by the product of edge weights in the image of β. (For an unweighted graph, we simply set all edge weights equal to 1.)

We may now define an analogous notion of homomorphism density for graphons.

Definition 5 (Equation 6, Hladký et al. 2021)**.** *For a graphon $W \in \mathcal{W}_0$ and a multigraph $H = (V, E)$ on n nodes, the **homomorphism density** of H in W is:*

$$t(H, W) = \int_{[0,1]^n} \prod_{(i,j)\in E} W(x_i, x_j) \prod_{i\in V} dx_i$$

For a clique K_r, the homomorphism density can be defined as:

$$t(K_r, W) = \mathbb{E} \prod_{(i,j)\in E} W(U_i, U_j)$$

Observe that this definition is similar to the definition of homomorphism density for weighted graphs, where $W(x_i, x_j)$ is the weight of the edge (i, j).

The quantity $X_n(H, W)$

Now, note that if H is a simple graph with k vertices, then the homomorphism density $t(H, W) \in [0, 1]$ is the probability that the W-random graph $\mathbb{G}(n, W)$ contains a subgraph that is isomorphic to H.

Let $X_n(H, W)$ denote the no. of subgraphs of $\mathbb{G}(n, W)$ that are isomorphic to H. To obtain the expectation of $X_n(H, W)$, we first take the probability $t(H, W)$ that a copy of H is in $\mathbb{G}(N, W)$. Then, we multiply this quantity by the no. of size-k subgraphs H of $\mathbb{G}(n, W)$. This quantity is given by $\frac{\frac{n!}{(n-k)!}}{\text{aut}(H)}$, where aut$(H)$ denotes the no. of graph automorphisms of H.

Note that $(n)_k := \frac{n!}{(n-k)!}$ is the no. of ways we can permute k out of n objects, and to avoid double-counting possible permutations of vertices within H, we need to divide by aut(H). Thus, we have that:

$$\mathbb{E}[X_n(H, W)] = \frac{(n)_k}{\text{aut}(H)} \cdot t(H, W)$$

Conditional homomorphism densities, K_r-regular graphons

Definition 6 (Equation 7, Hladký et al. 2021)**.** *For an integer $l \le k$, let J be an l-element subset of $[k] = \{1, 2, \ldots, k\}$.*
*Let H be a graph with vertex set $[k]$ where nodes in J are considered to be **marked**.*
Then, given a vector of values $\mathbf{x} = (x_j)_{j\in J} \in [0, 1]^l$, define the conditional homomor-

phism density $t_{\mathbf{x}}(H,W)$ as follows:

$$t_{\mathbf{x}}(H,W) = \mathbb{E}\left[\prod_{\{i,j\}\in E(H)} W(U_i,U_j) \;\middle|\; U_j = x_j : j \in J\right]$$

If H is a simple graph containing r nodes, then $t_{\mathbf{x}}(H,W)$ is the conditional probability that the W-random graph $\mathbb{G}(r,W) = H$ whenever node j is assigned type x_j (where $j \in J$).

Note that if $H = K_r$ is an r-clique, then $t_{\mathbf{x}}(K_r,W)$ depends only on the cardinality of J and not the elements of J (i.e. the marked nodes).

Recall that $t_{\mathbf{x}}(K_r,W)$ depends only on the no. of marked nodes in $\mathbb{G}(K_r,W)$. Then, let $K_r^{\bullet}$ and $K_r^{\bullet\bullet}$ denote K_r with one and two marked nodes respectively, with corresponding conditional homomorphism densities $t_x(K_r^{\bullet},W)$ and $t_{x,y}(K_r^{\bullet\bullet},W)$. This motivates the following definitions:

Definition 7. *A graphon W is K_r-**free** if $t(K_r,W) = 0$.*
*A graphon W is K_r-**complete** if $t(K_r,W) = 1$ almost everywhere.*

Definition 8. *A graphon W is K_r-**regular** if for almost every $x \in [0,1]$, we have:*

$$t_x(K_r^{\bullet},W) = t(K_r,W)$$

We first observe that for $r = 2$, we have $t_x(K_2^{\bullet},W) = t(K_2,W) = \int_0^1 W(x,y)\,dy = \deg_W(x)$ (by definition of the degree function of a graphon). This indicates that K_2-regularity coincides with the definition of regularity for a graphon.

Now, note that for $r \geq 3$, K_r-regularity indicates that in $\mathbb{G}(n,W)$, any node (regardless of its type) is expected to belong to the same no. of r-cliques.

Moreover, it can be shown that if a graphon W is *not* K_r-regular, then two copies of K_r in $\mathbb{G}(n,W)$ that share one vertex are positively correlated, resulting in greater variance in the no. of copies of K_r. This implies that if W is indeed K_r-regular, then any two copies of K_r that share exactly one vertex are uncorrelated. That is, the probability of one copy's existence is unrelated to the probability of the other copies' existence.

Spectrum of a graphon, the graphon $V_W^{(r)}$

Prior to stating the results obtained by Hladký et al., we first recall some preliminaries regarding the spectral property of graphons.

We first note that $L^2[0,1]$ is a real Hilbert space consisting of functions $f : [0,1] \to \mathbb{R}$ where $\int_0^1 |f(x)|^2\,dx < \infty$.

Now, for a graphon $W : [0,1]^2 \to [0,1]$, there is an associated integral linear operator $T_W : L^2[0,1] \to L^2[0,1]$, where $(T_W f)(x) = \int_0^1 W(x,y) f(y)\,dy$.

One can verify that the operator T_W is a Hilbert-Schmidt operator and that there are countably many non-zero real eigenvalues associated with W. Let $\mathrm{Spec}(W)$ denote the multiset of such eigenvalues. Now, it can be shown that:

$$\sum_{\lambda \in \mathrm{Spec}(W)} \lambda^2 = \int_{[0,1]^2} W(x,y)^2 \, dx \, dy \leq 1 \tag{1}$$

This fact is used in Hladký et al.'s proof when they discuss conditions pertaining to a normal limit distribution.

Moreover, it can be shown that if W is a regular graphon, i.e. $\deg_W(x) \equiv d$ for some constant d, then W has an eigenfunction $f \equiv 1$ with associated eigenvalue d. Then, let $\mathrm{Spec}^-(W)$ be the multiset of eigenvalues of W, where the multiplicity of the eigenvalue d is decreased by 1.

Now, to encode information regarding local clique densities in $\mathbb{G}(n,W)$, Hladký et al. construct an auxiliary graphon $V_W^{(r)}$. For a graphon W and $r \geq 2$, a graphon $V_W^{(r)} : [0,1]^2 \to [0,1]$ is defined where:

$$V_W^{(r)}(x,y) := t_{x,y}(K_r^{\bullet\bullet}, W)$$

That is, $V_W^{(r)}(x,y)$ is the homomorphism density of r-cliques K_r in $\mathbb{G}(n,W)$ that contain two nodes with types x and y. We note that $V_W^{(2)} = W$.

One can show that W is a K_r-regular graphon if and only if $V_W^{(r)}$ is a regular graphon. By computing the degree function of $V_W^{(r)}$ explicitly, we see that:

$$\begin{aligned}
\deg_{V_W^{(r)}}(x) &= \int_0^1 V_W^{(r)}(x,y) \, dy \\
&= \int_0^1 t_{x,y}(K_r^{\bullet\bullet}, W) \, dy \quad \text{(by definition of } V_W^{(r)}\text{)} \\
&= t_x(K_r^{\bullet}, W) \\
&= t(K_r, W) \quad \text{(by } K_r\text{-regularity of } W\text{)} \\
&= t_r
\end{aligned}$$

Since t_r is a constant, it follows that $V_W^{(r)}$ is a regular graphon with degree t_r. Then, $f \equiv 1$ is an eigenfunction of $V_W^{(r)}$ with eigenvalue t_r, and $\mathrm{Spec}^-(V_W^{(r)})$ is the eigenvalue spectrum of $V_W^{(r)}$ where the multiplicity of t_r is decreased by 1.

Statement of Hladký et al.'s results (Theorem 1.2)

Theorem (Theorem 1.2b, Hladký et al. (2021)). *Let W be a graphon. Fix $r \geq 2$ and let $t_r = t(K_r, W)$. Let $X_{n,r}$ denote the no. of r-cliques in $\mathbb{G}(n,W)$. Then:*

(a) If W is K_r-free or complete, then almost surely $X_{n,r} = 0$ or $X_{n,r} = \binom{n}{r}$ respectively.

(b) If W is not K_r-regular, then:

$$\frac{X_{n,r} - \binom{n}{r} t_r}{n^{r-1/2}} \xrightarrow{d} \hat{\sigma}_{r,W} \cdot Z$$

where $Z \sim N(0,1)$ and $\hat{\sigma}_{r,W} = \frac{1}{(r-1)!}\left(t(K_r \ominus K_r, W) - t_r^2\right)^{1/2} > 0$.

(c) Suppose W is a K_r-regular graphon that is neither K_r-free nor complete. Recall that X_n denotes the no. of r-cliques in $\mathbb{G}(n,W)$. Now, let $r \geq 2$ and set $t_r = t(K_r, W)$. Then:

$$\frac{X_n - \binom{n}{r} t_r}{n^{r-1}} \xrightarrow{d} \sigma_{r,W} \cdot Z + \frac{1}{2(r-2)!} \sum_{\lambda \in Spec^{-}(V_W^{(r)})} \lambda \cdot (Z_\lambda^2 - 1)$$

where Z and $(Z_\lambda)_{\lambda \in Spec^{-}(V_W^{(r)})}$ are independent standard normal.

The statement of part (b) is analogous to the following statement, which is akin to the statement of the Central Limit Theorem:

$$\frac{X_{n,r} - \mathbb{E}[X_{n,r}]}{\sqrt{\mathrm{Var}[X_{n,r}]}} \xrightarrow{d} Z$$

Now, recall that a chi-square distribution with k degrees of freedom of the distribution of the sum of squares of k independent standard normal random variables, i.e. if we have $Q = \sum_{i=1}^{k} Z_i^2$ where $Z_i \overset{i.i.d}{\sim} N(0,1)$, then $Q \sim \chi^2(k)$. Then, since $V_W^{(r)}$ has countably many eigenvalues λ and the independent standard normal variables Z_λ are indexed over $\lambda \in \mathrm{Spec}^{-}(V_W^{(r)})$, we may view the result of part (c) as a weighted analog of a chi-square distribution with countably infinite terms.

Hladký et al. show that the limit distribution in part (c) is normal if and only if $V_W^{(r)}$ is regular. To see the forward direction, note that if the limit distribution in part (c) is normal, then $\mathrm{Spec}^{-}(V_W^{(r)})$ is the empty set, i.e. $\mathrm{Spec}(V_W^{(r)}) = \{t_r\}$. Since W is assumed to be K_r-regular in part (c), we have that $V_W^{(r)}$ is regular with constant degree function $\deg_{V_W^{(r)}} \equiv t_r$.

Now, note that:

$$
\begin{aligned}
t_r^2 &= \left(\int_0^1 \deg_{V_W^{(r)}}(y)\, dy \right)^2 \\
&= \left(\int_{[0,1]^2} V_W^{(r)}(x,y)\, dx\, dy \right)^2 \quad \text{(by definition of the degree function)} \\
&\leq \int_{[0,1]^2} \left(V_W^{(r)}(x,y) \right)^2 dx\, dy \\
&\quad \text{(Applying Jensen's Inequality, since the quadratic function is convex)} \\
&= \sum_{\lambda \in \mathrm{Spec}(V_W{}^{(}r))} \lambda^2 \quad \text{(By equation (1))} \\
&= t_r^2
\end{aligned}
$$

Note that equality in the above inequality is attained if and only if $V_W^{(r)}$ is constant, i.e. $V_W^{(r)} \equiv t_r$.

The question regarding which graphons W lead to a constant graphon $V_W^{(r)}$ remains an open problem. Hladký et al. postulate that for $r \geq 3$, for $V_W^{(r)}$ to be constant, W must be a constant K_r-regular graphon. That is, among random graphs of the form $\mathbb{G}(n,W)$ where W is K_r-regular, only Erdös-Rényi random graphs $\mathbb{G}(n,p)$ for $p \in (0,1)$ that correspond to constant $W \equiv p$ have an asymptotically normal number of r-cliques.

Hladký et al. also discuss conditions where the normal term is absent in part (c) of the theorem above. Namely, this condition occcurs if and only if $W(x,y) = 1$ for almost every $(x,y) \in [0,1]^2$ for which $t_{x,y}(K_r^{\bullet\bullet}, W) > 0$. That is, the distribution in part (c) is normal-free when the graphon W attains a value of 1 for almost all (x,y) where the homomorphism density of an r-clique containing types x,y is non-zero.

Numerical Simulations

We perform some numerical simulations that verify Hladký et al.'s results.

We first consider the graphon $W(x,y) = xy$ and we consider the W-random graph $\mathbb{G}(n,W)$ for $n = 100$. To construct this graph, we sample types $U_1, \ldots, U_50$ independently from the uniform distribution on $[0,1]$ and connect nodes $i,j \in \{1,\ldots,50\}$ by an edge with probability $W(U_i, U_j)$. We set $r = 3$ to be the size of cliques whose we are interested in. Next, we record the no. of 3-cliques $X_{100,3}$ in the resultant random graph, and repeat this process for 1000 iterations.

In the histogram below, we plot the distribution of $\frac{X_{n,r} - \mathbb{E}[X_{n,r}]}{n^{r-1/2}} = \frac{X_{100,3} - \mathbb{E}[X_{100,3}]}{100^{3-1/2}}$, in accordance with the hypothesis of Theorem 1.2b.

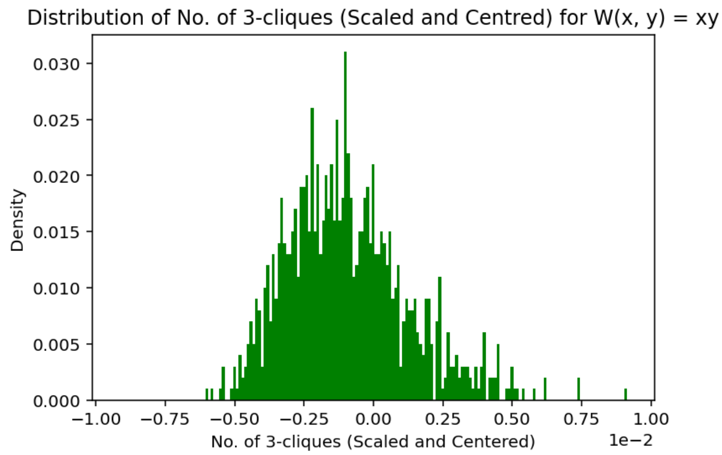

Figure: Distribution of $\frac{X_{100,3}-\mathbb{E}[X_{100,3}]}{100^{3-1/2}}$ for $\mathbb{G}(3,W)$ where $W(x,y)=xy$

We observe that the distribution resembles the shape of a scaled standard Gaussian, as stated in Theorem 1.2b.

We then consider an example of a K_3-regular graphon where the distribution in Theorem 1.2c is *not* Gaussian (discussed on pg.9-10 of Hladký et al.). We set $r=3$, and subdivide [0, 1] into 6 equally-sized subintervals. We place a copy of the complete 3-partite graphon on the first 3 subintervals, and another copy on the last 3 subintervals. We also connect the first and fourth subinterval with an arbitrary value. That is, we obtain a graphon W where:

$$W(x,y)=\begin{cases}1 & \text{if } x,y\in[0,1/2] \text{ where } x\neq y \text{ or } x,y\in[1/2,1] \text{ where } x\neq y\\ 0.5 & \text{if } x\in[0,1/6] \text{ and } y\in[3/6,4/6] \text{ or vice versa}\\ 0 & \text{otherwise}\end{cases}$$

(Here 0.5 was arbitrarily chosen as the value between the first and fourth subinterval)

An illustration of this graphon is included below:

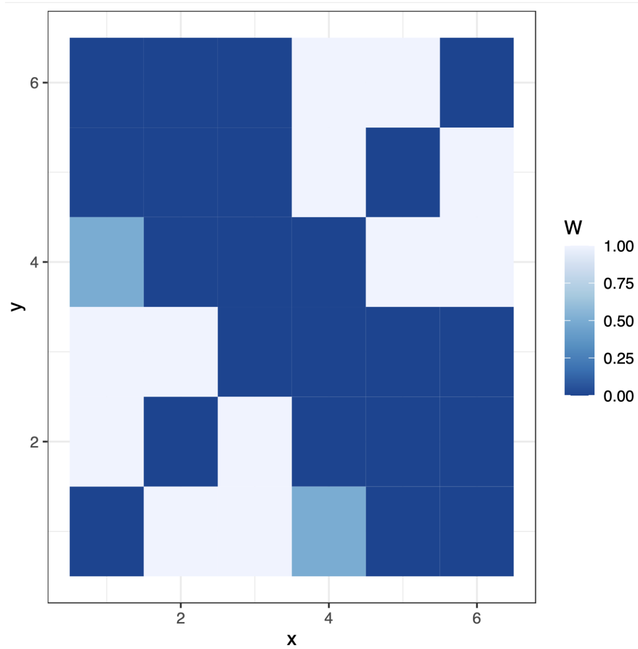

Figure: K_3-regular graphon discussed on pg. 9 of Hladký et al
(Image courtesy of Anirban Chatterjee)

According to Hladký et al, this graphon is K_3-regular but has $\sigma_{r,W} = 0$. Repeating the aforementioned simulation process with $n = 100$ for 1000 iterations, we note that the distribution of $X_{n,r}$ is non-Gaussian:

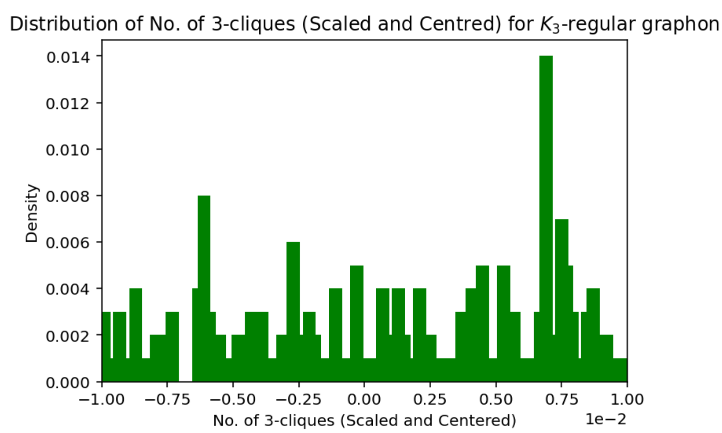

Figure: Distribution of $\frac{X_{100,3}-\mathbb{E}[X_{100,3}]}{100^{3-1/2}}$ for $\mathbb{G}(3, W)$ for the aforementioned K_3-regular graphon W

The Python code used to generate the above simulations is included below for reference:

```
# Import relevant Python packages
import numpy as np
import networkx as nx
from collections import defaultdict
import matplotlib.pyplot as plt

def simulate_graphon(n, r, num_iterations = 1000, W,
    graphon_name):
   """
   Plots the distribution of clique counts for a graphon

   Parameters:
   n (int): No. of nodes in the W-random graph
   r (int): Size of clique
   num_iterations (int): No. of iterations for simulation (by
       default 1000)
   W (function): Graphon function
   graphon_name (string): Name of graphon (for saving resulting
        plot)
   """
  # Instantiate a dictionary that maps the no. of r-cliques to
     their frequency
  clique_counts = defaultdict(int)

  for i in range(num_iterations):
   # Compute n uniform random variables U_1, ..., U_n
   U = np.random.uniform(low=0, high=1, size=n)

   # Create probability matrix where entry (i, j) = W(U_i, U_j)
   prob_matrix = np.array(
      [np.array([W(U[i], U[j]) for j in range(n)]) for i in
          range(n)]
   )

   # Populate lower diagonal entries of adjacency matrix
       according to edge probabilties
   adj = np.zeros((n, n))

   for i in range(n):
    for j in range(i):
     # Create an edge (i, j) with probability W(U_i, U_j)
     adj[i, j] = np.random.binomial(1, prob_matrix[i, j])
```

```
    # Make adjacency matrix symmetric
    adj = np.tril(adj) + np.tril(adj, 1).T

    G = nx.from_numpy_matrix(adj)

    # Compute no. of cliques of size r
    num_r_cliques = len([clique for clique in nx.find_cliques(G)
        if len(clique) == r])

    clique_counts[num_r_cliques] += 1

  # Compute frequencies
  clique_frequencies = np.array(list(clique_counts.values())) /
     num_iterations

  mean_clique_count = np.mean(list(clique_counts.keys()))

  transformed_clique_count = (list(clique_counts.keys()) -
     mean_clique_count) / (n ** (r - 1))

  transformed_clique_count = np.array(transformed_clique_count)

plt.bar(transformed_clique_count, clique_frequencies, color='g'
    , width=0.0005)
# Display x-axis labels using scientific notation
plt.ticklabel_format(style='sci', axis='x', scilimits=(0,0))
plt.title(f"Distribution of No. of {r}-cliques (Scaled and
    Centred) for {graphon_name}")
plt.xlabel(f"No. of {r}-cliques (Scaled and Centered)")

ax = plt.gca()

plt.xlim(-1*1e-2, 1*1e-2)
plt.ylabel("Density")

plt.savefig(f"{graphon_name}, n = {n}, r = {r}, {num_iterations
    } iterations.png", bbox_inches='tight', dpi=144)
```

Proof Idea of Theorem 1.2b

The proof of Theorem 1.2b uses a construction called *dependency graphs*. Given a collection of random variables $(Y_i : i \in I)$ for some index set I, we create a dependency graph $\mathcal{G}$ with vertex set I.

Now, for each vertex $i \in I$, let N_i denote the neighborhood of $i \in \mathcal{G}$. We construct $\mathcal{G}$ such that for all $i \in I$, the random variable Y_i is independent of $\{Y_j\}_{j \notin N_i}$.

Note that the dependency graph need not be unique for a given collection of random variables $(Y_i)_{i \in I}$.

Hladký et al. also use the following off-the-shelf bound for the Wasserstein distance between two random variables, which may be viewed as a distance function between probability distributions.

Theorem (Theorem 2.2, Hladký et al. 2021)**.** *Let $(Y_i : i \in I)$ be a finite collection of random variables where $\forall\ i \in I, \mathbb{E}[Y_i] = 0$ and $\mathbb{E}[Y_i^4] < \infty$. Let $\sigma^2 =$ $\mathrm{Var}\left[\sum_{i \in I} Y_i\right]$ and $Q = \sum_{i \in I} \frac{Y_i}{\sigma}$. Let $\mathcal{G}$ be a dependency graph for $(Y_i : i \in I)$, and let $D = \max_{i \in I} |N_i|$. Then, we have that:*

$$d_{Wass}(Q, Z) \leq \frac{D^2}{\sigma^3} \sum_i \mathbb{E}[|Y_i|^3] + \frac{\sqrt{28} D^{3/2}}{\sqrt{\pi} \sigma^2} \sqrt{\sum_i \mathbb{E}[Y_i^4]}$$

Hladký et al. apply the above bound to a collection of random variables $(Y_R : R \in \binom{[n]}{r})$. Here, $\binom{[n]}{r}$ is the set of all size-r subsets of $[n]$, so the random variables Y_R are indexed by size-r subsets $R \subseteq [n]$.

Let I_r be the indicator random variable for the event where R induces a clique in $\mathbb{G}(n, W)$. Then, let $Y_R := I_R - \mathbb{E}[I_R] = I_R - t_r$. Note that we have $\mathbb{E}[Y_R] = 0$ as required in the theorem above.

Then, we have $X_{n,r} - \binom{n}{r} t_r = \sum_{R \in \binom{[n]}{r}} I_R - t_r = \sum_{R \in \binom{[n]}{r}} Y_R$.

We then construct the dependency graph $\mathcal{G}$, where edges correspond to non-disjoint R_i, R_j, i.e. $R_i \cap R_j \neq \emptyset$. In $\mathcal{G}$, each neighbourhood N_R has the same size D, given by:

$$D = \sum_{l=1}^{r} \binom{r}{l} \binom{n-r}{r-l} = O(n^{r-1})$$

The summation on the left hand side sums over all possible ways to generate a neighborhood N_R of size r, by iterating over $l = \{1, \ldots, r\}$.

Then, the authors enumerate the no. of ordered pairs R_1, R_2 whose intersection has cardinality l for each $l \in [r]$, which they compute to be:

$$\binom{n}{l} \binom{n-l}{r-l} \binom{n-r}{r-l} = O(n^{2r-l})$$

Using this information, it can be shown that $\sigma_n^2 \approx \hat{\sigma}_{r,W}^2 \cdot n^{2r-1}$.

Then, the authors define $Q_n := \sum_{R \in \binom{[n]}{r}} \frac{Y_R}{\sigma_n}$. Applying Theorem 2.2 where we bound $\binom{n}{r} \leq n^r$ and examine powers of n, one can show that $d_{Wass}(Q_n, Z) = O(n^{-1/2}) \to 0$, i.e. $Q_n \xrightarrow{d} Z$. Next, applying Slutsky's Theorem, we have that:

$$\frac{\sum_{R \in \binom{[n]}{r}} Y_R}{n^{r-1/2}} = \frac{\sigma_n}{n^{r-1/2}} \cdot Q_n \xrightarrow{d} \hat{\sigma}_{r,W} Z$$

Since $\sum_{R \in \binom{[n]}{r}} Y_R = X_{n,r} - \binom{n}{r} t_r$, this completes the proof.

Proof Idea of Theorem 1.2c

The proof for Theorem 1.2c uses the method of moments to establish distributional convergence of $\frac{X_{n,r} - \binom{n}{r} t_r}{n^{r-1}}$.

Recalling that I_R be an indicator for the event that a size-r subset of vertices R induces a clique in $\mathbb{G}(n, W)$, we have that:

$$X_{n,r} - \binom{n}{r} t_r = \sum_{R \in \binom{[n]}{r}} I_R - t_r$$

Now, the authors analyse the structure of tuples $(R_1, \ldots, R_m)$ where each R_i is a subset of vertices of $\mathbb{G}(n, W)$. Defining $\Delta(R_1, \ldots, R_m) := \mathbb{E}\Big[\prod_{i=1}^m (I_{R_i} - t_r)\Big]$, we have that:

$$\mathbb{E}\left[\left(X_{n,r} - \binom{n}{r} t_r\right)^m\right] = \sum_{(R_1, \ldots, R_m) \,\in\, \binom{[n]}{r}^m} \Delta(R_1, \ldots, R_m)$$

The above equation indicates that for each m, we can analyze the m-th moments of $X_{n,r} - \binom{n}{r} t_r$ by examining the sum of the contributions $\Delta(R_1, \ldots R_m)$ of m-tuples $(R_1, \ldots, R_m)$.

The authors first define $\mathfrak{X}(n, r, m)$, which is a family of m-tuples where there exists some $i \in [m]$ such that $|R_i \cap \left(\cup_{j \neq i} R_j\right)| \leq 1$, we have that $\mathbb{E}\Big[\prod_{i=1}^m (I_{R_i} - t_r)\Big] = 0$. Recalling that $X_{n,r} - \binom{n}{r} t_r = \sum_{R \in \binom{[n]}{r}} I_R - t_r$, the above result shows that m-tuples in $\mathfrak{X}(n, r, m)$ do not contribute to the computation of the moments of $X_{n,r} - \binom{n}{r} t_r$. The authors thus consider m-tuples *not* in $\mathfrak{X}(n, r, m)$.

Letting $\mathsf{F}(n, r, m)$ denoting this complementary family of m-tuples, the authors show that hypergraphs $\mathcal{H}$ corresponding to such tuples have $(r-1)m$ nodes and are a union of vertex-disjoint loose cycles. Isomorphism classes of $\mathcal{H}$ can be encoded by a vector $\mathbf{k}$ whose i-th component is given by the no. of loose cycles of length i.

In Claim 4.3, the authors show that for each tuple in the aforementioned isomorphism classes, their contribution $\Delta(R_1, \ldots, R_m)$ is the same, and they obtain an explicit expression for the contribution. The authors proceed to count the no. of tuples in the isomorphism classes of $\mathcal{H}$ in Claim 4.4.

Using this information, the authors are able to express $\mathbb{E}[(X_{n,r} - \binom{n}{r}t_r)^m]$ as a formal power series $f(x)$ (Claim 4.5) and compute coefficients explicitly.

Then, the authors define a random variable Y equal to the right hand side of Theorem 1.2c, where:

$$Y := \sigma_{r,W} \cdot Z + \frac{1}{2(r-2)!} \sum_{\lambda \in \mathrm{Spec}^-(V_W^{(r)})} \lambda \cdot (Z_\lambda^2 - 1)$$

In Claim 4.6, the authors consider the moment generating function $M_Y(t) = \mathbb{E}[e^{tY}]$ and show that $M_y(x) = f(x)$ in a neighborhood of zero. Since the MGF of Y is finite and exists in this neighborhood, and recalling that the MGF of Y uniquely determines its distribution, we are thus able to conclude that $\frac{X_n - \binom{n}{r}t_r}{n^{r-1}} \xrightarrow{d} Y$ as desired.

Extensions and Concluding Remarks

Bhattacharya, Chatterjee & Janson (2022) extended Hladký et al's results for general subgraphs H in W-random graphs, and they introduce an analogous notion of H-regular graphons. Specifically, they found that if W is *not* H-regular, then the distribution of $X_n(H, W)$ is asymptotically Gaussian. Moreover, if W is H-regular, then the limiting distribution of $X_n(H, W)$ consists of a Gaussian term and a Chi-squared term.

Moreover, Kaur & Röllin (2021) provide a central limit theorem for *centred* subgraph counts in W-random graphs, and they demonstrate their distributional convergence to Gaussian distributions. They also developed test statistics for determining the presence of certain subgraphs, for example two edges sharing a common vertex.

There are also numerous open problems within this area. One such problem is to extend Hladký et al's results to sparse W-random graphs. Namely, to find the no. of cliques & other subgraphs in *sparse* $\mathbb{G}(n, p \cdot W)$ where $p = p(n) \to 0$ as $n \to \infty$. Another open area is to examine the variance of the no. of hyperedges of a certain size in W-random hypergraphs.

References

Bhaswar Chattacharya, Anirban Chatterjee, Svante Janson. "Fluctuations of Subgraph Counts in Graphon Based Random Graphs." Department of Statistics, University of Pennsylvania, 17 Jan. 2021, https://arxiv.org/abs/2104.07259

Gursharn Kaur, Adrian Röllin. "Higher-Order Fluctuations in Dense Random Graph Models." Department of Statistics & Applied Probability, National University of Singapore, 16 Jun. 2021, https://arxiv.org/abs/2006.15805v2

Jan Hladký, Christos Pelekis, Matas Šileikis. "A Limit Theorem for Small Cliques in Inhomogeneous Random Graphs." Journal of Graph Theory, vol. 97, no. 4, 2021, pp. 578–599, https://arxiv.org/abs/1903.10570