EXCHANGEABLE RANDOM GRAPHS, GRAPH LIMITS, AND GRAPHONS

Jiasen Yang
March 31, 2016

Department of Statistics
Purdue University
This talk is NOT about:

- Deep learning, AlphaGo, [insert hype here] ...
- What I have been working on recently (network sampling, statistical relational learning)

This talk is about:

- A fundamental theorem in probability theory
- A recent development in graph limit theory
- And how they are connected to nonparametric models of network data
Graph Data

Exchangeable Random Graphs

Graph Limits and Graphons

Graphon Estimation
Graph Data
A graph \( G = (V, E) \), where

- \( V = \{1, 2, \cdots, n\} \) is the set of vertices;
- \( E \subseteq V \times V \) is the set of edges,

can be represented by an **adjacency matrix** \( A \), where

\[
A_{ij} = \begin{cases} 
1, & \text{if } (i,j) \in E; \\
0, & \text{otherwise.}
\end{cases}
\]

Assume \( G \) is simple and undirected \( \Rightarrow \) \( A \) is binary, symmetric, and \( A_{ii} = 0, \forall i \).
They are the same network!
They are the same network!
Exchangeable Random Graphs
Definition

An exchangeable sequence is an infinite sequence $X_1, X_2, \cdots$ of random variables whose joint distribution satisfies

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2, \cdots) = \mathbb{P}(X_{\pi(1)} \in A_1, X_{\pi(2)} \in A_2, \cdots)$$

for every permutation $\pi$ of $\mathbb{N} := \{1, 2, \cdots\}$ and collection $A_1, A_2, \cdots$ of (measurable) sets.

In words, order of observations does not matter.

$i.i.d. \Rightarrow \text{exchangeable}; \text{exchangeable} \not\Rightarrow i.i.d.$

---

Example (Polyá’s Urn scheme)
Consider an urn with $b$ black balls and $w$ white balls. Draw a ball at random and note its color. Replace the ball together with $a$ balls of the same color. Repeat the procedure \textit{ad infinitum}. Let $X_i = \mathbb{I}\{\text{the } i\text{-th draw yields a black ball}\}$.

The sequence $X_1, X_2, \cdots$ is exchangeable but not \textit{i.i.d.}/Markov.

E.g., $P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1)$
\[
= \frac{b}{b + w} \frac{b + a}{b + w + a} \frac{w + a}{b + w + 2a} \frac{b + 2a}{b + w + 3a} \\
= \frac{b}{b + w} \frac{w + a}{b + w + 2a} \frac{b + a}{b + w + a} \frac{b + 2a}{b + w + 3a} \\
= P(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1).
\]
Theorem (de Finetti, 1931)

Let \( X_1, X_2, \cdots \) be an infinite sequence of random variables with values in a space \( \mathcal{X} \). The sequence \( X_1, X_2, \cdots \) is exchangeable if and only if there is a random probability measure \( \Theta \) on \( \mathcal{X} \)—i.e., a random variable with values in the set \( \mathcal{M}(\mathcal{X}) \) of probability distributions on \( \mathcal{X} \)—such that the \( X_i \) are conditionally i.i.d. given \( \Theta \) and

\[
P(X_1 \in A_1, X_2 \in A_2, \cdots ) = \int_{\mathcal{M}(\mathcal{X})} \prod_{i=1}^{\infty} \theta(A_i) \nu(d\theta)
\]

where \( \nu \) is the distribution of \( \Theta \).

Adj. matrix of undirected graph = binary & symmetric array

⇒ de Finetti exchangeability too strong!

⇒ Need generalization of exchangeability to arrays.
Adj. matrix of undirected graph = binary & symmetric array
⇒ de Finetti exchangeability too strong!
⇒ Need generalization of exchangeability to arrays.

\[
\sigma = \pi \otimes \pi \\
\sigma = \pi_1 \otimes \pi_2
\]
Exchangeable Arrays

Definition
A random array \((X_{ij})_{i,j \in \mathbb{N}}\) is called separately exchangeable if

\[
(X_{ij}) \overset{d}{=} (X_{\pi(i)\pi'(j)})
\]

holds for every pair of permutations \(\pi, \pi'\) of \(\mathbb{N}\).

Definition
A random array \((X_{ij})_{i,j \in \mathbb{N}}\) is called jointly exchangeable if

\[
(X_{ij}) \overset{d}{=} (X_{\pi(i)\pi(j)})
\]

holds for every permutation \(\pi\) of \(\mathbb{N}\).

Aldous-Hoover Theorem

Theorem (Aldous, 1981-Hoover, 1979)

A random array \((X_{ij})_{i,j \in \mathbb{N}}\) is

- **jointly exchangeable** iff. \(\exists\) a random measurable function \(F : [0, 1]^3 \rightarrow \mathcal{X}\) s.t.

\[
(X_{ij}) \overset{d}{=} (F(U_i, U_j, U_{\{i,j\}}))
\]

- **separately exchangeable** iff. \(\exists\) a random measurable function \(F : [0, 1]^3 \rightarrow \mathcal{X}\) s.t.

\[
(X_{ij}) \overset{d}{=} (F(U_i, U'_j, U_{ij}))
\]

where \(U_i, U_{\{i,j\}}, U'_j, U_{ij} \overset{i.i.d.}{\sim} \text{Uniform}[0, 1].\)

Definition

An undirected random graph $G = (\mathbb{N}, E)$ with an infinite (countable) vertex set $\mathbb{N}$ and a random edge set $E$ is called an exchangeable random graph if its random adjacency matrix $A = (A_{ij})$ is a jointly exchangeable array.

Thus, $G$ is exchangeable if its distribution is invariant under relabeling of the vertices.

Adjacency matrix $A = (A_{ij})$ symmetric & binary:

$$A_{ij} \overset{d}{=} F(U_i, U_j, U_{\{i,j\}}) \overset{d}{=} \mathbb{I}\{U_{\{i,j\}} < W(U_i, U_j)\}$$
Adjacency matrix $A = (A_{ij})$ symmetric & binary:

$$A_{ij} \overset{d}{=} F(U_i, U_j, U_{\{i,j\}}) \overset{d}{=} \mathbb{I}\{U_{\{i,j\}} < W(U_i, U_j)\}$$

Sample $W : [0, 1]^2 \to [0, 1]$ measurable & symmetric (graphon);

Sample $U_1, U_2, \ldots \overset{i.i.d.}{\sim} \text{Uniform}[0, 1]$;

Sample $A_{ij} \sim \text{Bernoulli}(W(U_i, U_j))$ for $i < j$. 

---

Graph Limits and Graphons
Rado graph: \( A_{ij} = A_{ji} \overset{i.i.d.}{\sim} \text{Bernoulli}(1/2), \ i \neq j \in \mathbb{N} \).
Graph Limits

Rado graph: $A_{ij} = A_{ji} \overset{i.i.d.}{\sim} \text{Bernoulli}(1/2), \ i \neq j \in \mathbb{N}$.

Limit graphon (graph function):

$$f : [0,1]^2 \to [0,1], (x,y) \mapsto \frac{1}{2}.$$
Growing uniform attachment: Let $G_1 = \circ$. For $n \geq 2$, construct $G_n$ from $G_{n-1}$ by adding one new vertex, then, drawing an edge between each pair of non-adjacent vertices with probability $\frac{1}{n}$.

Growing uniform attachment: Let $G_1 = \bullet$. For $n \geq 2$, construct $G_n$ from $G_{n-1}$ by adding one new vertex, then, drawing an edge between each pair of non-adjacent vertices with probability $\frac{1}{n}$.

Limit graphon:

$$f : [0, 1]^2 \rightarrow [0, 1], (x, y) \mapsto 1 - \max(x, y).$$
Definition

A labeled graphon is a symmetric, Lebesgue-measurable a.e. function $W : [0, 1]^2 \rightarrow [0, 1]$. A labeled graphon determines the equivalence class of graphons

$$[W] = \{W^\phi : (x, y) \mapsto W(\phi(x), \phi(y))\},$$

where $\phi$ is an invertible, measure-preserving transformation of $[0, 1]$. Such equivalence classes are called unlabeled graphons.
**Definition**

The cut distance between two labeled graphons $W$ and $U$ is

$$\delta_{\Box}(W, U) = \inf_{\phi, \psi \text{ m.p.t. of } [0,1]} \sup_{S, T \subseteq [0,1]} \int_{S \times T} |W^\phi(x, y) - U^\psi(x, y)|.$$  

**Theorem**

*Every graphon is the $\delta_{\Box}$-limit of a sequence of finite graphs.*

**Theorem (Lovász-Szegedy)**

*Let $\mathcal{G}$ be the space of unlabeled graphons (modulo weak isometry). The metric space $(\mathcal{G}, \delta_{\Box})$ is compact.*

---

Graphon Estimation
Adjacency matrix $A = (A_{ij})$ symmetric & binary:

$$A_{ij} \overset{d}{=} F(U_i, U_j, U_{\{i,j\}}) \overset{d}{=} \mathbb{I}\{U_{\{i,j\}} < W(U_i, U_j)\}$$

Sample $W : [0, 1]^2 \to [0, 1]$ measurable & symmetric (graphon);
Sample $U_1, U_2, \ldots \overset{i.i.d.}{\sim} \text{Uniform}[0, 1]$;
Sample $A_{ij} \sim \text{Bernoulli}(W(U_i, U_j))$ for $i < j$. 

---

**Problem:** Estimate graphon $W$ from a set of observed networks.

**Nonparametric regression:** estimate $W$ from $\{(U_i, U_j), A_{ij}\}_{i,j \in \mathbb{N}}$.

**Challenge:** design points $\{(U_i, U_j)\}$ are latent.

**Proposed estimators:**

- Stochastic blockmodel approximation [Airoldi et.al., 2013], [Gao et.al., 2015]
- Histogram estimator (sorting-and-smoothing) [Chan and Airoldi, 2014]
- Gaussian process model [Lloyd et.al., 2012], [Orbanz and Roy, 2015]
APPLICATIONS

- Community detection
- Link prediction:
  \[ P((i, j) \in E) = P(A_{ij} = 1) = W(U_i, U_j) \]
- Network comparison & hypothesis testing

Real-world graphs are sparse (finite # of edges per vertex).

- Power-law degree distribution;
- Small-world phenomena.
Real-world graphs are **sparse** (finite # of edges per vertex).

- Power-law degree distribution;
- Small-world phenomena.

**Theorem (Misspecification)**

*If a random graph is exchangeable, it is either dense or empty.*

**Proof.**

For a random graph $G_n$ with $n$ vertices, the expected proportion of present edges is

$$p := \int_{[0,1]^2} W(x, y) \, dx \, dy.$$ 

If $p = 0$, $G_n$ is empty; if $p > 0$, $G_n$ has $p \cdot \binom{n}{2} = \Theta(n^2)$ edges in expectation.

The limit object of a convergent sequence of *sparse* graphs in the cut metric is always the *empty* graphon.

Representation theorems for *sparse* random graphs?

