Bayesian nonparametric models of sparse and exchangeable random graphs

François Caron   Emily B. Fox

STA4513 Presentation, October 2014
Presenter: Victor Veitch

Big Picture

The Problem

Exchangeable random arrays are either empty or dense and thus not appropriate for most real applications.

Big Picture Question

How can we salvage a useful notion of exchangeability for graphs?

Big Picture

The Problem

Exchangeable random arrays are either empty or dense and thus not appropriate for most real applications.

Big Picture Question

How can we salvage a useful notion of exchangeability for graphs?

Main Ideas

Point Process $\leftrightarrow$ Random Graph

Set up a correspondence between random graphs and random discrete measures (point processes)

Symmetry

- Natural notion of exchangeability of point processes
- Use associated representation theorem to study random graphs

Point Processes and Graphs

Completely Random Measures

- $W = \sum_{i=1}^{\infty} w_i \delta_{\theta_i}$ random measure
- w_i sociability parameter
- θ_i embedding of node i in $\mathbb{R}^+$

Point Process

- $Z = \sum_i \sum_j z_{ij} \delta_{(\theta_i, \theta_j)}$
- $z_{ij} = 1$ if there is a link between θ_i, θ_j
- $z_{ij} = f(w_i, w_j)$

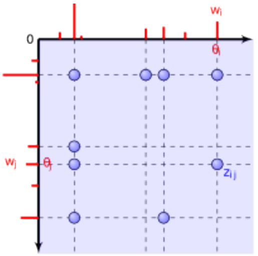

Figure : Edge between θ_i and θ_j represented by points at (θ_i, θ_j) and (θ_j, θ_i)

Point Processes and Directed Multigraphs

Directed Multigraph

- $D = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} n_{ij} \delta_{(\theta_i, \theta_j)}$
- n_{ij} number of directed edges from θ_i to θ_j

Given $W \sim \text{CRM}(\rho, \lambda)$

- $D | W \sim PP(W \times W)$ on $\mathbb{R}_+^2$
- informally, n_{ij} are generated as $\text{Poisson}(w_i w_j)$

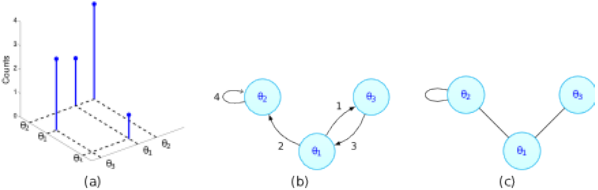

Figure : (Restricted) atomic measure D to directed multigraph to corresponding undirected graph

Undirected Graphs

Hierarchical Model

$$W = \sum_{i=1}^{\infty} w_i \delta_{\theta_i} \qquad W \sim \mathrm{CRM}(\rho, \lambda)$$

$$D = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} n_{ij} \delta_{(\theta_i, \theta_j)} \qquad D | W \sim \mathrm{PP}(W \times W)$$

$$Z = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \min(n_{ij} + n_{ji}, 1) \, \delta_{(\theta_i, \theta_j)}$$

Observation

The distribution of the random graph is determined by the distribution of W

Completely Random Measures

W a CRM if

For any countable collection $A_1, A_2 \ldots$ of measurable sets

- random variables $W(A_1), W(A_2), \ldots$ are independent
- $W(\cup_j A_j) = \sum_j W(A_j)$
- the distribution of $W([t,s])$ depends only on $t - s$

Important Facts

- $W = \sum_{i=1}^{\infty} w_i \delta_{\theta_i}$ almost surely
- For any measurable A the Laplace transform may be written as

$$\mathscr{L}[W(A)] = \exp\left(-\int_{\mathbb{R}_+ \times A} [1 - \exp(-tw)]\, \rho(dw)\, \lambda(d\theta)\right)$$

Jump Part

Jump Part ρ

- CRM W is characterized by measure ρ on $\mathbb{R}^+$ such that $\int_0^\infty (1 - e^{-w}) \rho(dw) < \infty$
- $\int_0^\infty \rho(dw) = \infty \iff$ number of jumps is infinite in any interval $[0, T]$
- Infinite number of jumps $\leftrightarrow$ infinite number of nodes (almost all with degree 0)

Representation

This (possibly non-parametric) model for random graphs has a representation in terms of ρ

Finite Size Samples

Truncation

- Aldous-Hoover construction: truncate at finite number of nodes n
- Random measure construction: truncate $\mathbb{R}_+$ to interval $[0, \alpha]$
  - Define $W^*_\alpha = W([0, \alpha])$
  - Number of directed edges: $D^*_\alpha | W^*_\alpha \sim \text{Poisson}\left(W^{*2}_\alpha\right)$

Exchangeability of Random Measures

Point process Z on $\mathbb{R}_+^2$

- π, σ permutations of $\mathbb{N}$
- $A_i = [h(i-1), hi]\ h > 0$
- Z is exchangeable if and only if $Z(A_i \times A_j) \overset{d}{=} Z\left(A_{\pi(i)} \times A_{\sigma(j)}\right)$

Point process defining the graph construction is exchangeable

- $W(A_i) \overset{d}{=} W\left(A_{\sigma(i)}\right)$
- $D(A_i \times A_j) \sim \text{Poisson}\left(W(A_i) \times W(A_j)\right)$

Kallenberg Representation Theorem

Construction of CRM:

- (θ_i, ϑ_i) a unit rate Poisson process on $\mathbb{R}_+ \times \mathbb{R}_+$
- $L(x) = \int_x^\infty \rho(dw)$
- $w_i \equiv L^{-1}(\vartheta_i)$ then $W = \sum w_i \delta_{\theta_i}$ is CRM with $\rho(dw)\, d\theta$

Kallenberg Representation:

- $Z = \sum_{i,j} f\left(\vartheta_i, \vartheta_j, \zeta_{\{i,j\}}\right) \delta_{\theta_i, \theta_j}$ (transformed Poisson Processes)
- $f\left(\vartheta_i, \vartheta_j, \zeta_{\{i,j\}}\right) = \begin{cases} 1 & \zeta_{\{i,j\}} \leq M(\vartheta_i, \vartheta_j) \\ 0 & \text{ow} \end{cases}$
- $M(\vartheta_i, \vartheta_j) = \begin{cases} 1 - \exp\left(-2L^{-1}(\vartheta_i)\, L^{-1}(\vartheta_j)\right) & \vartheta_i \neq \vartheta_j \\ 1 - \exp\left(-L^{-1}(\vartheta_i)^2\right) & \vartheta_i = \vartheta_j \end{cases}$

Kallenberg Representation Continued

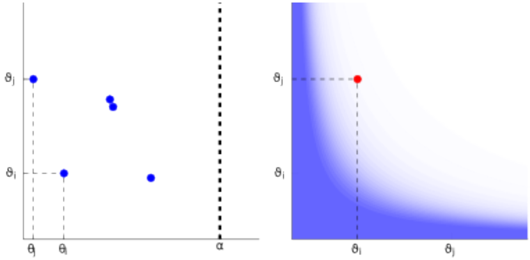

Figure : Model construction based on Kallenberg representation. (left) Unit rate Poisson process. (right) graphical representation of L^{-1}

Example 1: Poisson Process

Representations

- $\rho(dw) = \delta_{w_0}(dw)$ so $\int_0^\infty \rho(dw) < \infty$
- $L(x) = \begin{cases} 1 & x < w_0 \\ 0 & \text{ow} \end{cases}$

Construction for fixed α

- sample $n \sim \text{Poisson}(\alpha)$, number of nodes
- set $z_{ij} = z_{ji} = 1$ with probability $1 - \exp\left(-2w_0^2\right)$
- this is Erdős-Renyi conditional on n

Example 2: Compound Poisson Process

Representations

- $\rho(dw) = h(w)\,dw$ with $\int_0^\infty h(w)\,dw = 1$
- $L(x) = 1 - H(x)$

Construction for fixed α

- sample $n \sim \text{Poisson}(\alpha)$, number of nodes
- set $z_{ij} = z_{ji} = 1$ with probability $M(U_i, U_j)$, U_i uniform
- $M(U_i, U_j) = 1 - \exp\left(-2H^{-1}(U_i)\,H^{-1}(U_j)\right)$
- Aldous-Hoover representation, conditional on n

Example 3: Generalized Gamma Process

Representations

- $\rho(dw) = \frac{1}{\Gamma(1-\sigma)} w^{-1-\sigma} \exp(-\tau w)\, dw, \;\; \sigma \in [0,1)\, \tau \geq 0$
- $L(x) = \begin{cases} \frac{\tau^{\sigma}\Gamma(-\sigma,\tau x)}{\Gamma(1-\sigma)} & \tau > 0 \\ \frac{x^{-\sigma}}{\Gamma(1-\sigma)\sigma} & \tau = 0 \end{cases}$

Features

- CRM has infinite number of jumps in any interval
- Exact sampling is possible via urn process
- The network growth is not dense

Example 3: Generalized Gamma Process Continued

Power Law

Let $N_{\alpha,j}$ number of nodes in directed graph D with j outgoing or ingoing edges, then

$$\frac{N_{\alpha,j}}{N_{\alpha,1}} \to \frac{\sigma\Gamma(j-\sigma)}{\Gamma(1-\sigma)\Gamma(j+1)}, \ \alpha \to \infty$$
$$\sim \frac{\sigma}{\Gamma(1-\sigma)} j^{-1-\sigma}, \ j \to \infty$$

Sparsity

Let E_α the number of edges in the undirected graph. Then for $0 < \varepsilon < \sigma$

$$E_\alpha = O\left(N_\alpha^{2-\sigma+\varepsilon}\right)$$

almost surely as $\alpha \to \infty$

Discussion

Main points

- Correspondence between point processes and random graphs
- exchangeability of random measures of $\mathbb{R}_+^2$ gives tractable representation
- there are random graphs in this family that are asymptotically sparse