Nonparametric Statistical Inference for Networks

Complex Networks Transition Workshop, SAMSI, June 2011

Sharmodeep Bhattacharyya and Peter Bickel

Statistics Dept. UC Berkeley

Nonparametric Asymptotic Model for Unlabeled Graphs

Given: P on ∞ graphs

Aldous/Hoover (1983)

$$\mathcal{L}(A_{ij} : i,j \geq 1\} \quad = \quad \mathcal{L}(A_{\pi_i,\pi_j} : i,j \geq 1),$$

for all permutations $\pi \iff$

$$\exists \quad g : [0,1]^4 \to \{0,1\} \text{ such that } A_{ij} \quad = \quad g(\alpha, \xi_i, \xi_j, \eta_{ij}),$$

where

α, ξ_i, η_{ij}, all i, $j \geq i$, i.i.d. $\mathcal{U}(0,1)$, $g(\alpha, u, v, w) = g(\alpha, v, u, w)$,

$\eta_{ij} = \eta_{ji}$.

Block Models (Holland, Laskey and Leinhardt 1983)

Probability model:

- Community label: $\mathbf{c} = (c_1, \cdots, c_n)$ i.i.d. multinomial $(\pi_1, \cdots, \pi_K) \equiv K$ "communities".
- Relation:

$$\mathbb{P}(A_{ij} = 1 | c_i = a, c_j = b) \quad = \quad P_{ab}.$$

- A_{ij} conditionally independent

$$\mathbb{P}(A_{ij} = 0) \quad = \quad 1 - \sum_{1 \leq a, b \leq K} \pi_a \pi_b P_{ab}.$$

- $K = 1$: E-R model.

Ergodic Models

$\mathcal{L}$ is an ergodic probability iff for g with $g(u,v,w) = g(v,u,w)$ $\forall(u,v,w)$,

$$A_{ij} = g(\xi_i, \xi_j, \eta_{ij}).$$

$\mathcal{L}$ is determined by

$$h(u,v) \equiv \mathbb{P}(A_{ij} = 1 | \xi_i = u, \xi_j = v),\ h(u,v) = h(v,u).$$

Notes:

1. K-block models and many other special cases
2. Model (also referred to as threshhold models) also suggested by Diaconis, Janson (2008)
3. More general models (Bollobás, Riordan & Janson (2007))

"Parametrization" of NP Model

- h is not uniquely defined.
- $h\big(\varphi(u), \varphi(v)\big)$, where φ is measure-preserving, gives same model.

  But, h_{CAN} = that $h(\cdot, \cdot)$ in equivalence class such that $P\left[A_{ij} = 1 | \xi_i = z\right] = \int_0^1 h_{\mathrm{CAN}}(z, v) dv \equiv \tau(z)$ with $\tau(\cdot)$ monotone increasing characterizes uniquely.
- ξ_i could be replaced by any continuous variables or vectors - but there is no natural unique representation.

Examples of models

i) Block models: on block of sizes π_a, π_b

$$h_{CAN}(u,v) = F_{ab}$$

ii) Power law: $$w(u,v) = a(u)a(v)$$

$$a(u) \sim (1-u)^{-\alpha} \text{ as } u \uparrow 1$$

iii) Dynamically defined model (preferential attachment):

$$w(u,v) = a(u)1(u \leq v) + a(v)1(u > v)$$

New vertex attaches to random old vertex and neighbors (not Hilbert-Schmidt)

$$a_{CAN}(u) = (1-u)^{-1} + \tau(u), \; a_{CAN}(u) = (1-u)^{-1} - \log(u(1-u))$$

Can One Fit Nonparametric Model?

- Even parametric models are difficult to fit. We have seen that even for simple parametric models such as block models, the efficient estimation of the parameters is not easy.
- But still many of the parametric models are not good enough representation of the naturally occurring graphs. The empirical and theoretical vulnerability of Exponential Random Graph Models have been pointed out by Chatterjee and Diaconis (2010) and Bhamidi et. al. (2008).

An Approach For Dense Models ($\lambda \to \infty$)

By Theorem 1(a), as $\lambda \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\tau(z_i) - \frac{D_i}{\bar{D}}\right)^2 = O\left(\frac{1}{\lambda}\right) \to 0 \qquad (1)$$

here, $\tau(z) = T(\mathbf{1})(z)$.

Let

$$\hat{W}_n(u,v) = \int_0^u \int_0^v \frac{1}{n\bar{D}} \sum_{i,j} A_{ij}\mathbf{1}(\hat{\xi}_i \le s, \hat{\xi}_j \le t)dsdt$$

where $\hat{\xi}_i \equiv \hat{F}(\frac{D_i}{\bar{D}})$ and $\hat{F}$ is the empirical df of $\{\frac{D_i}{\bar{D}} : 1 \le i \le n\}$. Let

$$W_n(u,v) = \int_0^u \int_0^v \frac{1}{n\bar{D}} \sum_{i,j} A_{ij}\mathbf{1}(\xi_i \le s, \xi_j \le t)dsdt.$$

Theorem 1

Suppose that the conditions of Theorem 1 hold.

a) If $w(\cdot,\cdot)$ is bounded, and F, the df of $\tau(\xi_1)$, is Lipschitz and strictly increasing, then uniformly in (u, v),

$$|\hat{W}_n(u,v) - W_n(u,v)| = O_P\left(\frac{(\log\lambda)^{3/2}}{\lambda^{1/2}}\right).$$

Theorem 1 (cont)

b) If $\rho \to 0$ and $\tau(\xi_1)$ takes on only a finite number of values $t_1, \cdots, t_K$, then uniformly in (u, v),

$$|\hat{W}_n(u, v) - W_n(u, v)| \quad = \quad O_P(\lambda^{-1/2})|.$$

Moreover, if $W(u, v) = \int_0^1 \int_0^1 w(s, t)(u - s)_+(v - t)_+ dsdt$, then uniformly in (u, v),

$$|W_n(u, v) - W(u, v)| \quad = \quad O_P(n^{-1/2})|.$$

Note:

$$\frac{\partial^4 W(u, v)}{(\partial u)^2 (\partial v)^2} \quad = \quad w(u, v). \tag{2}$$

An approach

a) Find smoothed empirical distribution function of $\frac{D_i}{\bar{D}}$,

$$\hat{F}(x) \equiv \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\frac{D_i}{\bar{D}} \leq x\right)$$

b) Divide $[0,1]$ into intervals $I_1, \ldots, I_M$, such that, $I_j = [\frac{j-1}{M}, \frac{j}{M})$,

$$\begin{aligned} \hat{w}(u,v) \; &\equiv \; \frac{1}{\bar{\bar{D}}}\sum_{a,b=1}^{M} \frac{1}{n^*}\mathbf{1}(u \in I_a)\mathbf{1}(v \in I_b) \\ &\times \left[\sum_{i,j=1}^{n} \mathbf{1}\left\{A_{ij} \; : \; \hat{F}\left(\frac{D_i}{\bar{D}}\right) \in I_a, \; \hat{F}\left(\frac{D_j}{\bar{D}}\right) \in I_b\right\}\right] \end{aligned}$$

where, $n^* = |I_a||I_b|$, if, $a \neq b$ and $n^* = (|I_a|(|I_a| - 1))/2$, if, $a = b$.

Example: Facebook Caltech Network

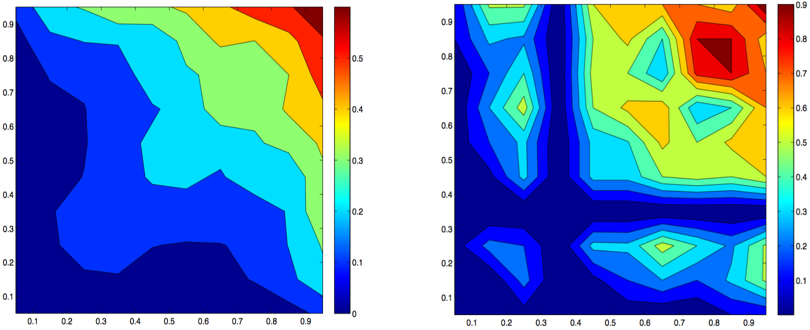

Figure : The LHS is estimate of h_{CAN} function for network of students of year 2008 and RHS is network of students of year 2008 residing in only 2 dorms. The proportions of classes in 2 distant modes are (0.3, 0.7) and (0.84, 0.16).

Why is the Result for Whole Network Uninstructive?

- $\xi \in U(0,1)$, w_{CAN} determine the probability uniquely but there are equivalent representation, which give very different results.
- $\xi \rightarrow$ degree suggest 'affinity', which is like 'linear' or first-order relation.
- We can now introduce higher-order relations, by making ξ a vector, that is, $(\xi) = (\xi^{(1)}, \xi^{(2)})$, where, $\xi^{(1)}, \xi^{(2)} \sim U(0,1)$, $\xi_1 \perp \xi_2$.
- One way of forming $\xi^{(1)}, \xi^{(2)}$ is: let the binary representation of ξ is $\xi = (\xi_1, \xi_2, \xi_3, \xi_4, \ldots)$. Now define, $\xi^{(1)} = (\xi_1, \xi_3, \ldots)$ and $\xi^{(2)} = (\xi_2, \xi_4, \ldots)$.
- We know that, if $\xi \sim U(0,1)$, then, $(\xi^{(1)}, \xi^{(2)}) \sim U(0,1)^2$. Also, $\xi \rightarrow (\xi^{(1)}, \xi^{(2)})$ is 1-1 onto.

Example: Facebook Caltech Network

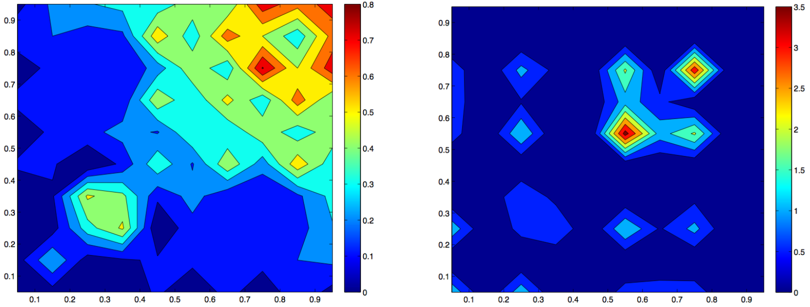

Figure : The LHS is estimate of h_{CAN} function for network of students of year 2008 residing in 3 dorms and RHS is sum of projections $\hat{h}_{CAN}(i, , i,)$ with two latent variables. The proportions of classes in 4 modes are (0.5, 0.13, 0.37), (0.67, 0.11, 0.22), (0.26, 0.66, 0.08), (0.32, 0.18, 0.5)

Another approach for Block Models

a) Cluster the normalized degrees $\frac{D_i}{\bar{D}}$ into K blocks. Estimate π_j from the normalized length of each interval in the cluster, for $j = 1, \ldots, K$.

b) Divide $[0, 1]$ into intervals $I_1, \ldots, I_k$, such that, $I_j = [\pi_{j-1}, \pi_j)$, with, $\pi_0 = 0$, $j = 1, \ldots, K$.

$$\hat{w}(u, v) \equiv \frac{1}{\bar{D}} \sum_{a,b=1}^{K} \frac{1}{n^*} \mathbf{1}(u \in I_a) \mathbf{1}(v \in I_b) \times \left[\sum_{i,j=1}^{n} \mathbf{1} \left\{ A_{ij} \; : \; \hat{F}\left(\frac{D_i}{\bar{D}}\right) \in I_a, \; \hat{F}\left(\frac{D_j}{\bar{D}}\right) \in I_b \right\} \right]$$

where, $n^* = |I_a||I_b|$, if, $a \neq b$ and $n^* = (|I_a|(|I_a| - 1))/2$, if, $a = b$.

Example: 3 block Model

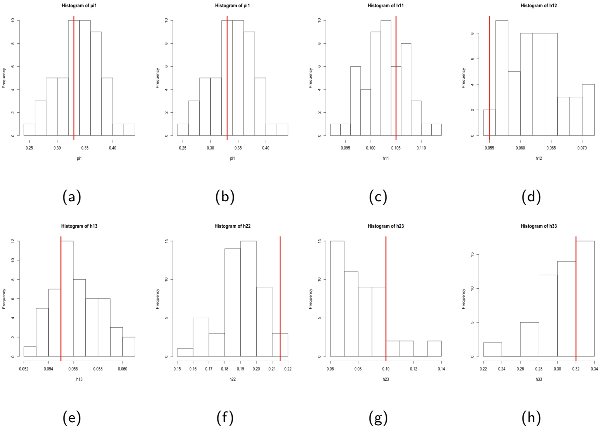

Figure : Histogram for estimated (a) π_1 (b) $(\pi_1 + \pi_2)$ (c) $h_{CAN}(1,1)$ (d) $h_{CAN}(1,2)$ (e) $h_{CAN}(1,3)$ (f) $h_{CAN}(2,2)$ (g) $h_{CAN}(2,3)$ (h) $h_{CAN}(3,3)$ with the original value indicated by red line.