Exchangeable Random Graphs

Econometric Methods for Social Spillovers and Networks

University of St. Gallen, September 28th to October 6th, 2020

Bryan S. Graham

University of California - Berkeley

Introduction

- First of two lectures on network nonparametrics

- **Lecture 1**:
  - Aldous-Hoover representation
    - Orbanz and Roy (2015)
  - Nearest neighborhood smoothing for edge probability estimation
    - Zhang, Levina and Zhu (2015)

Introduction (continued)

- **Lecture 2**:
  - graph limits (e.g., Lovász, 2012)
  - estimation of network moments
    * Holland and Leinhardt (1976)
    * Bickel, Chen and Levina (2011)
    * Bhattacharya and Bickel (2015)

Setup

Let $G(\mathcal{V}, \mathcal{E})$ be a finite undirected random graph with

- agents/vertices $\mathcal{V} = \{1, \ldots, N\}$,

- links/edges $\mathcal{E} = \{\{i, j\}, \{k, l\}, \ldots\}$, and

- adjacency matrix $\mathbf{D} = \left[D_{ij}\right]$ with

$$D_{ij} = \begin{cases} 1 & \text{if } \{i, j\} \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

Setup (continued)

- The *expected* adjacency matrix equals

$$\mathbf{P} = \left[P_{ij}\right] = \left[\mathbb{E}\left[D_{ij}\middle| U_1, \ldots, U_N, \alpha\right]\right]$$

  for $i < j$.

- Here $\{U_i\}_{i=1}^N$ and α are *latent* random variables introduced (and explained below).

- Form of $\mathbf{P}$ might indicate community structure (e.g., block model)...

- ...or guide other aspects of model formulation

Exchangeable Networks

- Let π be a permutation of the index set $\{1, \ldots, N\}$.

- In many situations it is natural to assume that

$$\left[D_{ij}\right] \stackrel{d}{=} \left[D_{\pi(i)\pi(j)}\right] \tag{1}$$

  for every permutation π and $i < j$, $j = 1, \ldots, N$.

- Condition (1) $\Rightarrow$ our beliefs about the probability of a link between two agents does not depend on their labels.

- Networks with this property are *jointly exchangeable*.

Exchangeable Networks (continued)

- Does exchangeability have any modeling implications?

- Does $\mathbf{D}$ converge to a *graph limit* as $N \to \infty$?

- Dense graph implication:

  – if $\left[D_{ij}\right] \stackrel{d}{=} \left[D_{\pi(i)\pi(j)}\right]$ then $\rho = \Pr\left(D_{ij} = 1\right)$ is either bounded away from zero *or* zero.

  – (infinitely) exchangeable graphs are either dense or empty!

Exchangeable Sequences

The sequence $Y_1, Y_2, \ldots$ is said to be **infinitely exchangeable** if, for every $N \geq 2$ and permutation π,

$$(Y_1, Y_2, \ldots, Y_N) \stackrel{d}{=} \left(Y_{\pi(1)}, Y_{\pi(2)}, \ldots, Y_{\pi(N)}\right).$$

i.i.d. sequences are exchangeable...

... but non i.i.d.sequences can be too:

$$Z + Y_1, Z + Y_2, \ldots$$

for Z some non-trivial random variable, drawn independently of the i.i.d. sequence $Y_1, Y_2, \ldots$

de Finetti Theorem

de Finetti (1931): the sequence of binary random variable $Y_1, Y_2, \ldots$ is infinitely exchangeable if, and only if,

$$\Pr(Y_1 = y_1, Y_2 = y_2, \ldots, Y_N = y_N) = \int_0^1 \alpha^{t_N} (1-\alpha)^{N-t_N} \, \mathrm{d}\Pi(\alpha)$$

for $t_N = \sum_{i=1}^N y_i$, all $N \geq 2$, and Π some measure on $\alpha \in [0,1]$.

For any infinitely exchangeable sequence we have that – conditional on the (latent) random variable α–

$$\Pr(Y_1 = y_1, Y_2 = y_2, \ldots, Y_N = y_N | \alpha) = F_\alpha(y_1) F_\alpha(y_2) \times \cdots \times F_\alpha(y_N)$$

for $F_\alpha(y) = \alpha^y (1-\alpha)^{1-y}$ if $y \in \{0,1\}$ and zero otherwise.

de Finetti Theorem (continued)

Representation result: any exchangeable binary sequence can be modeled 'as if' the DGP were:

1. Draw $\alpha \sim \Pi$

2. Draw $Y_i \sim F_\alpha$ for $i = 1, \ldots, N$

Conditional on α, $Y_1, Y_2, \ldots$ is an i.i.d. sequence, where each of its members have the same *random* distribution function $F_\alpha(y)$.

See Orbanz and Roy (2015) for non-technical survey of de Finetti type results (cf., Diaconis and Janson, 2008).

Alternative Formulation

The right-continuous inverse of $F_\alpha(u)$ (i.e., quantile function) is

$$g_\alpha(u) \stackrel{\text{def}}{=} F_\alpha^{-1}(u) = \begin{cases} 0 & \text{if } 0 < u < 1 - \alpha \\ 1 & \text{if } 1 - \alpha \leq u < 1 \end{cases}.$$

This gives:

$$(Y_1, Y_2, \ldots) \stackrel{d}{=} (g_\alpha(U_1), g_\alpha(U_2), \ldots)$$

for $\{U_i\}_{i=1}^{\infty}$ a sequence of independent $\mathcal{U}[0,1]$ random variables.

We further have that

$$\mathbb{E}[Y_i | U_i = u, \alpha] = g_\alpha(u).$$

Alternative Formulation

The "*sequon*" (**sequ**ence functi**on**) $g_\alpha(u)$ is not identifiable...

Consider $g_\alpha(u)$ as above with:

$$g^*_\alpha(u) = \begin{cases} 0 & \text{if } 0 < u < \frac{1-\alpha}{2} \\ 1 & \text{if } \frac{1-\alpha}{2} \le u < \frac{1}{2} \\ 0 & \text{if } \frac{1}{2} \le u < \frac{2-\alpha}{2} \\ 1 & \text{if } \frac{2-\alpha}{2} \le u < 1 \end{cases}.$$

...but "moments" are identifiable:

$\frac{1}{N}\sum Y_i \xrightarrow{p} \mathbb{E}\left[g_\alpha(U)|\,\alpha\right] = \alpha$ and $\frac{1}{N}\sum (Y_i - \alpha)^2 \xrightarrow{p} \alpha(1-\alpha)$

Aldous-Hoover

Aldous (1981) and Hoover (1979) (essentially) showed that a random graph is (infinitely) jointly exchangeable if, and only if, it admits the representation

$$\left[D_{ij}\right] \overset{d}{=} \left[g_{\alpha}\left(U_i, U_j, V_{ij}\right)\right]$$

for $\{U_i\}_{i=1}^{\infty}$ and $\left\{V_{ij}\right\}_{i<j}$ sequences of independent $\mathcal{U}[0,1]$ random variables.

Here α is a mixing parameter as in de Finetti (1931).

$g_{\alpha}(\cdot,\cdot,\cdot)$ is a *random function.*

Aldous-Hoover (continued)

Averaging over V_{ij} yields

$$\begin{aligned} h_\alpha\left(u_i, u_j\right) &= \mathbb{E}\left[D_{ij} \middle| U_i = u_i, U_j = u_j, \alpha\right] \\ &= \mathbb{E}\left[g_\alpha\left(u_i, u_j, V_{ij}\right) \middle| \alpha\right] \\ &= \int_0^1 g_\alpha\left(u_i, u_j, v\right) \mathrm{d}v, \end{aligned}$$

from which we get the more convenient representation, for $i < j$,

$$\boxed{\left[D_{ij}\right] \stackrel{d}{=} \left[\mathbf{1}\left(V_{ij} \leq h_\alpha\left(U_i, U_j\right)\right)\right]}.$$

$h_\alpha\left(U_i, U_j\right)$ is a *graphon*: short for **graph** functi**on**.

Aldous-Hoover (continued)

The Aldous-Hoover representation theorem implies that, *under (infinite) exchangeability*, we can proceed ‘as if’ links formed independently conditional on the agent-specific latent variables $\{U_i\}_{i=1}^{\infty}$ and α.

Aldous-Hoover (continued)

A network generating process is:

1. "Draw" α or choose a graphon;

2. Draw $U_i \sim \mathcal{U}\,[0,1]$ for agents $i = 1, \ldots, N$;

3. Construct $\mathbf{D}$, by sampling

$$D_{ij}\Big|\, h_\alpha(\bullet,\bullet), U_i, U_j \sim \text{Bernoulli}\left(h_\alpha\left(U_i, U_j\right)\right)$$

   for every dyad $\{i, j\}$ with $i < j$.

Aldous-Hoover (continued)

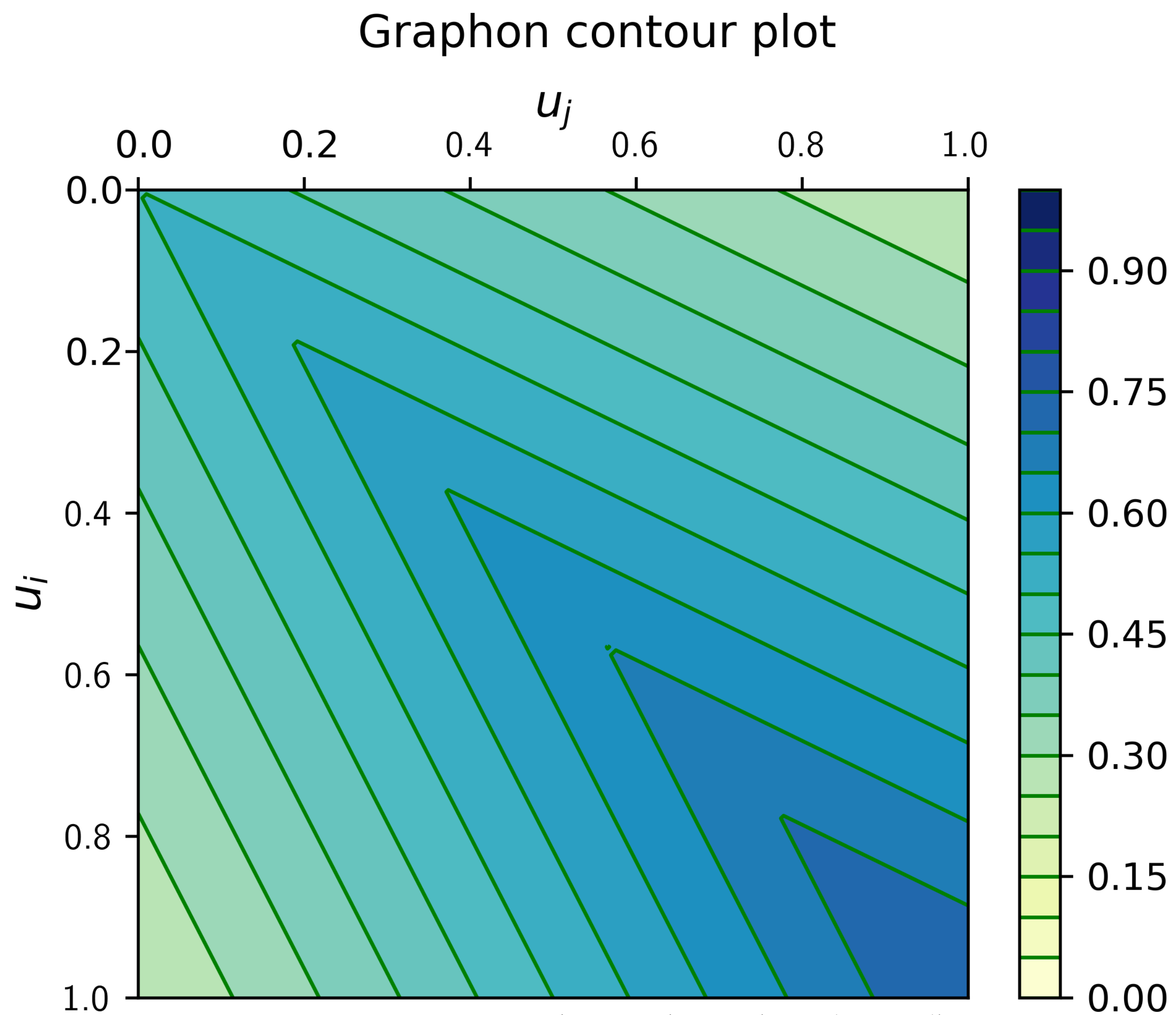
Any exchangeable random graph may be modeled as a mixture of conditionally independent edge formation processes.

Conditional independence structure useful for developing large network theory.

Representation result! Actual network generating process may not coincide with representation (cf., reduced form).

Graphon associated with a specific economic model may be very complicated.

(note: finite vs. infinite exchangeability).



Note: $h\left(u_i, u_j\right) = \frac{\exp(\alpha_0+\alpha_1(u_i+u_j)+\alpha_2|u_i-u_j|)}{1+\exp(\alpha_0+\alpha_1(u_i+u_j)+\alpha_2|u_i-u_j|)}$

Graphon

The graphon $h_\alpha(u,v)$ is not identifiable...

- consider the m.p.t. $\varphi(U) = 1 - U$ or $\varphi(U) = 2U \bmod 1$
- $g_\alpha\left(U_i, U_j, V_{ij}\right)$ and $g_\alpha\left(\varphi(U_i), \varphi\left(U_j\right), V_{ij}\right)$ generate graphs with the same properties

...but link/edge probabilities *are* identifiable (under assumptions).

- $P_{ij} = \mathbb{E}\left[D_{ij} \middle| \mathbf{U}\right] = h_\alpha\left(U_i, U_j\right)$

...as are network moments (next lecture).

Graphon (Bickel & Chen, 2009)

For statistical analysis it is convenient to formulate the graphon somewhat differently.

Consider the network DGP

$$\Pr\left(D_{ij}=1\middle|\, U_i,U_j,\alpha\right)=h_\alpha\left(U_i,U_j\right)$$

and define

$$\rho_\alpha=\int_0^1\int_0^1 h_\alpha(u,v)\,\mathrm{d}u\mathrm{d}v$$
$$w_\alpha(u,v)=f_{U_i,U_j|D_{ij},\alpha}\left(u,v|\,D_{ij}=1,\alpha\right).$$

Since $f_{U_i,U_j|\alpha}(u,v|\,\alpha)=1$ on $[0,1]^2$ we get the formulation

$$\boxed{h_\alpha(u,v)=\rho_\alpha w_\alpha(u,v)}$$

Graphon (Bickel & Chen, 2009)

The Bickel and Chen (2009) formulation is useful for sequences of network NGPs where ρ_α, the network density, is indexed by N.

We allow $\rho_{\alpha,N} \to 0$ as $N \to \infty$.

In practice we ignore any dependence of $w_\alpha(u,v)$ on N.

The rate at which $\rho_{\alpha,N} \to 0$ controls the sparsity links.

If $\lambda_N = (N-1)\,\rho_{\alpha,N} \to \lambda > 0$ as $N \to \infty$ the graph is *sparse*

- other cases: $\lambda_N = O(N)$ (*dense*) or $\lambda_N = O(\ln N)$ (*semi-dense*).

Edge Probability Estimation

Can we recovered link probabilities from $\mathbf{D}$?

Graphon is non-identifiable, but its conditional mean given $\mathbf{U} = (U_1, \ldots, U_N)'$ is identifiable.

Knowledge of

$$\mathbf{P} = \left[P_{ij}\right] = \left[\mathbb{E}\left[D_{ij}\middle| U_1, \ldots, U_N, \alpha\right]\right]$$

can aid in structural model formulation (cf., agent beliefs).

Edge Probability Estimation

Define the inner product

$$\langle f, g \rangle = \int f(u)\, g(u)\, \mathrm{d}u$$

with the associated norm

$$\|f\| = \langle f, f \rangle^{1/2} = \left[\int f(u)^2\, \mathrm{d}u \right]^{1/2}.$$

Linking behavior of an agent of type u is summarized by the *graphon slice* $\rho w(u, \bullet)$.

Edge Probability Estimation

Measure "distance" between agent i, with $U_i = u$, and agent j, with $U_j = v$, by:

$$\begin{aligned} d(u,v) &= \|\rho w(u,\cdot) - \rho w(v,\cdot)\|_2 \\ &= \rho \left[\int [w(u,t) - w(v,t)]^2 \,\mathrm{d}t \right]^{1/2} \end{aligned} \tag{2}$$

Network Neighbors

$\mathbf{P} = \mathbb{E}\left[\mathbf{D} \middle| \mathbf{U}\right]$ denotes the expected adjacency matrix.

$\mathbf{P}_{i\bullet}$ denotes the i^{th} row of this matrix

Distance between i and j is

$$\begin{aligned} d_N(i,j) &= \left\| \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet} \right\|_2 \\ &\stackrel{def}{\equiv} \left[\frac{1}{N-2} \sum_{k \neq i,j} \left(P_{ik} - P_{jk} \right)^2 \right]^{1/2} \end{aligned} \tag{3}$$

i^{th} and j^{th} elements of both $\mathbf{P}_{i\bullet}$ and $\mathbf{P}_{j\bullet}$ are removed prior to calculating $d(i,j)$.

Nearest Network Neighbors

j *is an exact neighbor of* i if $d_N(i,j) = 0$.

Exact neighbors i and j have identical (expected) adjacency (matrix) slices.

Ex ante their linking behavior is identical.

Their *realized* (ex post) links may differ.

Nearest Neighbor Averaging

In a finite network it may be that agent i has no exact neighbors, but we can still find a set of *nearest neighbors*:

$$\mathcal{N}_i = \left\{ j \, : \, \left\| \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet} \right\|_2 \leq q_i \left(h_N \right) \right\} \tag{4}$$

where $q_i \left(h_N \right)$ is the h_N^{th} sample quantile of $\left\{ \left\| \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet} \right\|_2 \right\}_{j=1, j \neq i}^{N}$.

If $N = 1,000$ and $h_N = 0.05$, then we would take the 50 nearest neighbors.

Nearest Neighbor Averaging

Estimate P_{ij} by the local average

$$\boxed{\widehat{P}_{ij}^{\text{oracle}} = \frac{1}{2}\left(\frac{\sum_{k\in\mathcal{N}_i} D_{kj}}{|\mathcal{N}_i|} + \frac{\sum_{l\in\mathcal{N}_j} D_{il}}{\left|\mathcal{N}_j\right|}\right)}. \tag{5}$$

Unfortunately $\mathbf{P}$ is not observed! Hence "oracle"...

Generalizes to the directed case easily.

Finding Network Neighbors

Can we construct a measure of distance between two agents based on the (observed) adjacency matrix alone?

Zhang et al. (2015) observe that

$$\begin{aligned} d_N^2(i,j) &= \left\|\mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}\right\|_2^2 \\ &= \left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet} \right\rangle \\ &= \left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{i\bullet} \right\rangle - \left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{j\bullet} \right\rangle \\ &\leq \left|\left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{i\bullet} \right\rangle\right| + \left|\left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{j\bullet} \right\rangle\right| \end{aligned}$$

Need estimates of $\langle \mathbf{P}_{i\bullet}, \mathbf{P}_{i\bullet} \rangle$, $\left\langle \mathbf{P}_{j\bullet}, \mathbf{P}_{j\bullet} \right\rangle$ and $\left\langle \mathbf{P}_{i\bullet}, \mathbf{P}_{j\bullet} \right\rangle$ to form estimate of $d_N(i,j)$.

Finding Neighbors (continued)

$\langle \mathbf{P}_{i\bullet}, \mathbf{P}_{i\bullet} \rangle = \frac{1}{N-1} \sum_{i \neq j} P_{ij}^2$ is hard to estimate...

...apparently requires estimate of P_{ij} (which is our target!)

However the (limit of the) cross-product term

$$\left\langle \mathbf{P}_{i\bullet}, \mathbf{P}_{j\bullet} \right\rangle = \frac{1}{N-2} \sum_{k \neq i,j} P_{ik} P_{jk}$$

is not hard to estimate since (iterated expectations)

$$\mathbb{E}\left[\frac{1}{N-2} \sum_{k} D_{ik} D_{jk} \right] = \mathbb{E}\left[\frac{1}{N-2} \sum_{k \neq i,j} P_{ik} P_{jk} \right].$$

Recall edges form independently conditional on $\mathbf{U}$.

Finding Neighbors (continued)

Assume that $w(u,v)$ is Lipschitz continuous:

$$\rho\left\|w(u,\cdot)-w(v,\cdot)\right\|_2 \leq C\left\|u-v\right\|_2.$$

With N large we can find an agent $k \neq i,j$ such that $|U_i - U_k| \leq \epsilon_N$ for $\epsilon_N = o(1)$.

We get

$$\begin{aligned}
\left|\left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{i\bullet} \right\rangle\right| &= \left|\left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{k\bullet} \right\rangle\right. \\
&\quad + \left.\left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{i\bullet} - \mathbf{P}_{k\bullet} \right\rangle\right| \\
\text{(TI)} &\leq \left|\left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{k\bullet} \right\rangle\right| \\
&\quad + \left|\left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{i\bullet} - \mathbf{P}_{k\bullet} \right\rangle\right| \\
\text{(CS)} &\leq \left|\left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{k\bullet} \right\rangle\right| \\
&\quad + \left\|\mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}\right\|_2 \|\mathbf{P}_{i\bullet} - \mathbf{P}_{k\bullet}\|_2 \\
&\leq \left|\left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{k\bullet} \right\rangle\right| + C_{i,j}\epsilon_N
\end{aligned}$$

with $C_{i,j} = C \left\|\mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}\right\|_2$.

Finding Neighbors (continued)

Combining results we have that

$$d^2(i,j) \le 2 \max_{l \neq i,j} \left| \left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{l\bullet} \right\rangle \right| + 2C_{i,j}\epsilon_N$$

...if $2 \max_{l \neq i,j} \left| \left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{l\bullet} \right\rangle \right| \approx 0$, then $d^2(i,j) \approx 0$ if N is large (upper bound).

Zhang et al. (2015) Estimate

Zhang et al. (2015) suggest estimating

$$2 \max_{l \neq i,j} \left| \left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{l\bullet} \right\rangle \right|$$

by $\widehat{\tilde{d}}^2(i,j)$ equal to

$$2 \max_{l \neq i,j} \left| \frac{1}{N-2} \left[\sum_{k \neq i,j} D_{ik} D_{lk} - \sum_{k \neq i,j} D_{jk} D_{lk} \right] \right|$$

Estimated neighborhood of agent i is then

$$\widehat{\mathcal{N}}_i = \left\{ j \, : \, \widehat{\tilde{d}}^2(i,j) \leq q_i(h_N) \right\}.$$

Zhang et al. (2015) Estimate (continued)

Estimate P_{ij} by

$$\widehat{P}_{ij} = \frac{1}{2}\left(\frac{\sum_{k\in\widehat{\mathcal{N}}_i} D_{kj}}{\left|\widehat{\mathcal{N}}_i\right|} + \frac{\sum_{l\in\widehat{\mathcal{N}}_j} D_{il}}{\left|\widehat{\mathcal{N}}_j\right|}\right)$$

Consistency requires that $h_N = C\sqrt{\frac{\ln N}{N}}$ for some C.

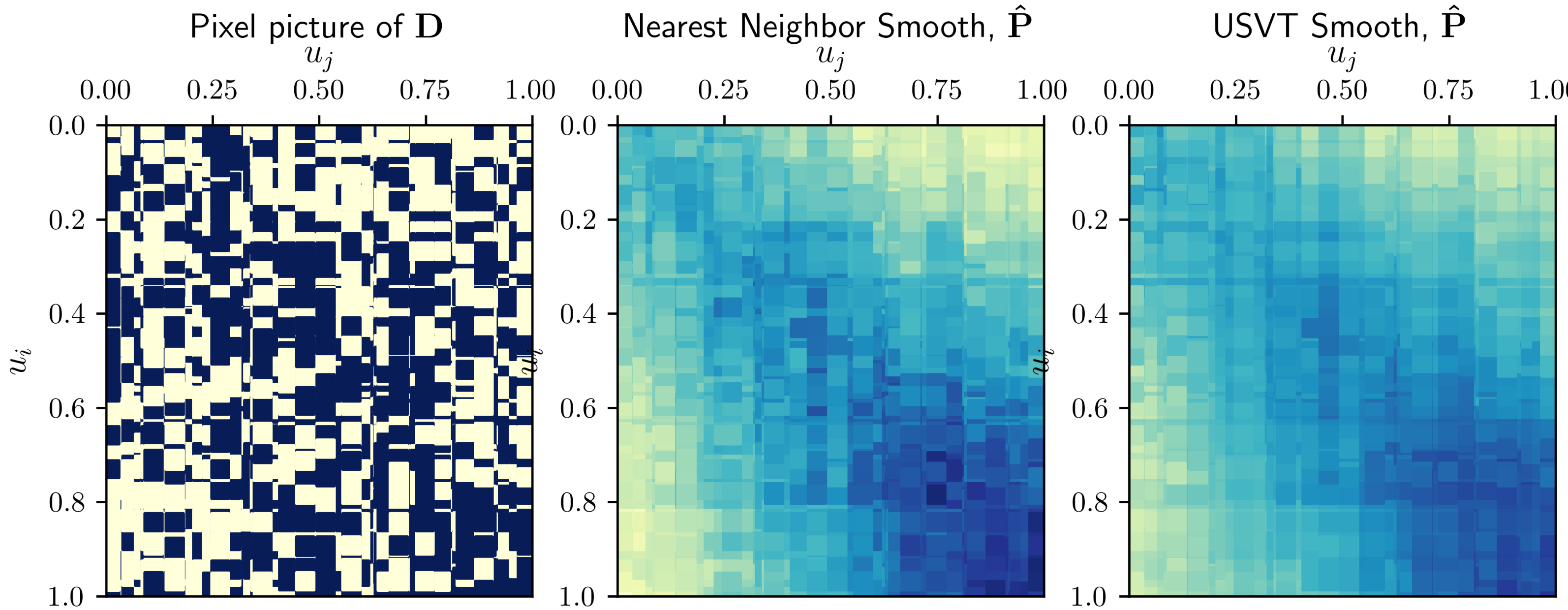
Zhang et al. (2015) suggest that $C = 0.1$ works well in practice.

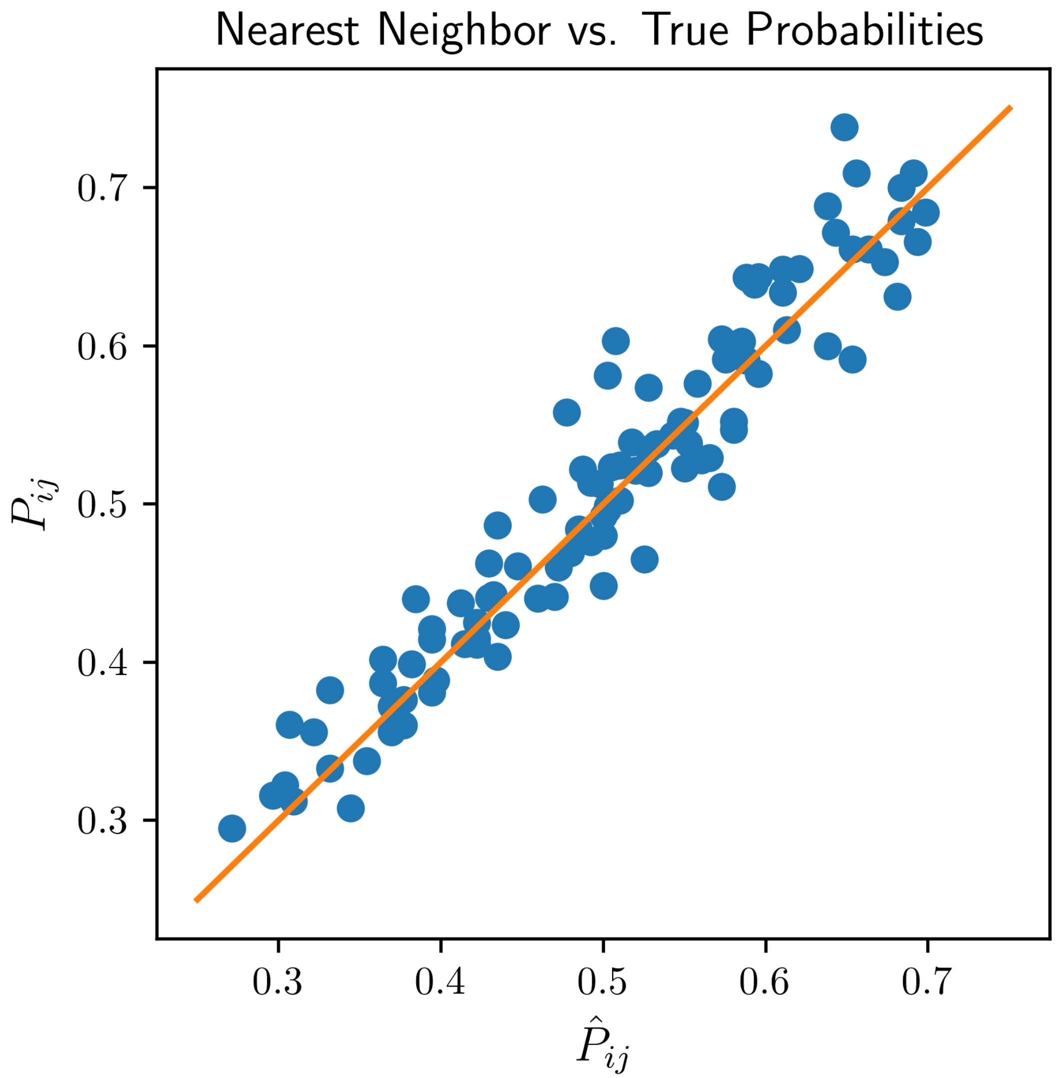
- $K_N = \lfloor Nh_N \rfloor = \left\lfloor 0.1\,(N \ln N)^{1/2} \right\rfloor$ or $K_{1000} \approx 8$ and $K_{2000} \approx 12$.

Alternative Distance Measure

We can use a $\widehat{d}_N^*(i,j)$ to find nearest neighbors instead. We have $\max\limits_{l\neq i,j} \left|\left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{l\bullet}\right\rangle\right| \leq \sum_{l\neq i,j} \left|\left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{l\bullet}\right\rangle\right| \cdot 1$ and hence:

$$
\begin{aligned}
\text{(HI)} &\leq \left[\sum_{l\neq i,j} \left(\left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{l\bullet}\right\rangle\right)^2\right]^{1/2} \cdot \left[\sum_{k\neq i,j} 1^2\right]^{1/2} \\
&= \left[(N-2) \sum_{k\neq i,j} \left(\left\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{l\bullet}\right\rangle\right)^2\right]^{1/2} \\
&= \left[\frac{1}{N-2} \sum_{l\neq i,j} \left(\sum_{k\neq i,l} P_{ik}P_{kl} - \sum_{k\neq j,l} P_{jk}P_{kl}\right)^2\right]^{1/2} \\
&= d_N^*(i,j)
\end{aligned}
$$


Pixel picture of $\mathbf{D}$
Nearest Neighbor Smooth, $\hat{\mathbf{P}}$
USVT Smooth, $\hat{\mathbf{P}}$
u_j
0.00
0.25
0.50
0.75
1.00
u_i
0.0
0.2
0.4
0.6
0.8
1.0


Nearest Neighbor vs. True Probabilities
0.7
0.6
0.5
0.4
0.3
P_{ij}
0.3
0.4
0.5
0.6
0.7
$\hat{P}_{ij}$

Practicalities

In example it is natural to order i by their realized values of U_i.

This information is not available in real world examples.

In practice, we can order agents by degree or its smoothed estimate $\sum_j \widehat{P}_{ij}$.

This should be sufficient to 'see' a block/community structure (for example) in many cases.

Edge Probability Estimation: Why do we care?

Adjacency matrix smoothing can be a useful for exploratory data analysis.

Can be used to discover community structure and other high-level network features.

Can also be used as a preliminary step in the context of economic model estimation.

Application to Nyakatoke Network