

Exchangeable Random Graphs

Econometric Methods for Networks,

GCEP, May 8th & 9th, 2017

Bryan S. Graham

University of California - Berkeley

Introduction

- First of two lectures on network nonparametrics
- Rest of today:
 - Aldous-Hoover representation
 - * Orbanz and Roy (2015)
 - Nearest neighborhood smoothing for edge probability estimation
 - * Zhang, Levina and Zhu (2015)

Introduction (continued)

- Tomorrow:
 - 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 $\mathcal{V} = \{1, \dots, N\}$,
- links/edges $\mathcal{E} = \{\{i, j\}, \{k, l\}, \dots\}$, and
- adjacency matrix $\mathbf{D} = [D_{ij}]$ 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

$$\mathbf{P} = [P_{ij}] = [\mathbb{E} [D_{ij} | U_1, \dots, U_N, \alpha]]$$

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 $\{1, \dots, N\}$.

- In many situations it is natural to assume that

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

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

– $\stackrel{d}{=}$ indicates equality of distribution.

- 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 $N \rightarrow \infty$?
- Dense graph implication:
 - if $[D_{ij}] \stackrel{d}{=} [D_{\pi(i)\pi(j)}]$ then $\rho = \Pr(D_{ij} = 1)$ is either bounded away from zero *or* zero.
 - exchangeable graphs are either dense or empty!

Exchangeable Sequences

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

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

- i.i.d. sequences are exchangeable...
- ... but non i.i.d. sequences can be too:

$$Z + Y_1, Z + Y_2, \dots$$

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

de Finetti Theorem

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

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

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 –

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

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 $i = 1, \dots, N$
- *Conditional* on α , Y_1, Y_2, \dots 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

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:

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

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*” (sequence function) $g_\alpha(u)$ is not identifiable...

– consider $g_\alpha(u)$ above with:

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

- ...but “moments” are identifiable:

$$- \frac{1}{N} \sum Y_i \xrightarrow{p} \mathbb{E}[g_\alpha(U) | \alpha] = \alpha$$

Aldous-Hoover

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

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

for $\{U_i\}_{i=1}^{\infty}$ and $\{V_{ij}\}_{i < j}$ 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

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

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

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

- $h_{\alpha}(U_i, U_j)$ is a *graphon*: short for **graph function**.

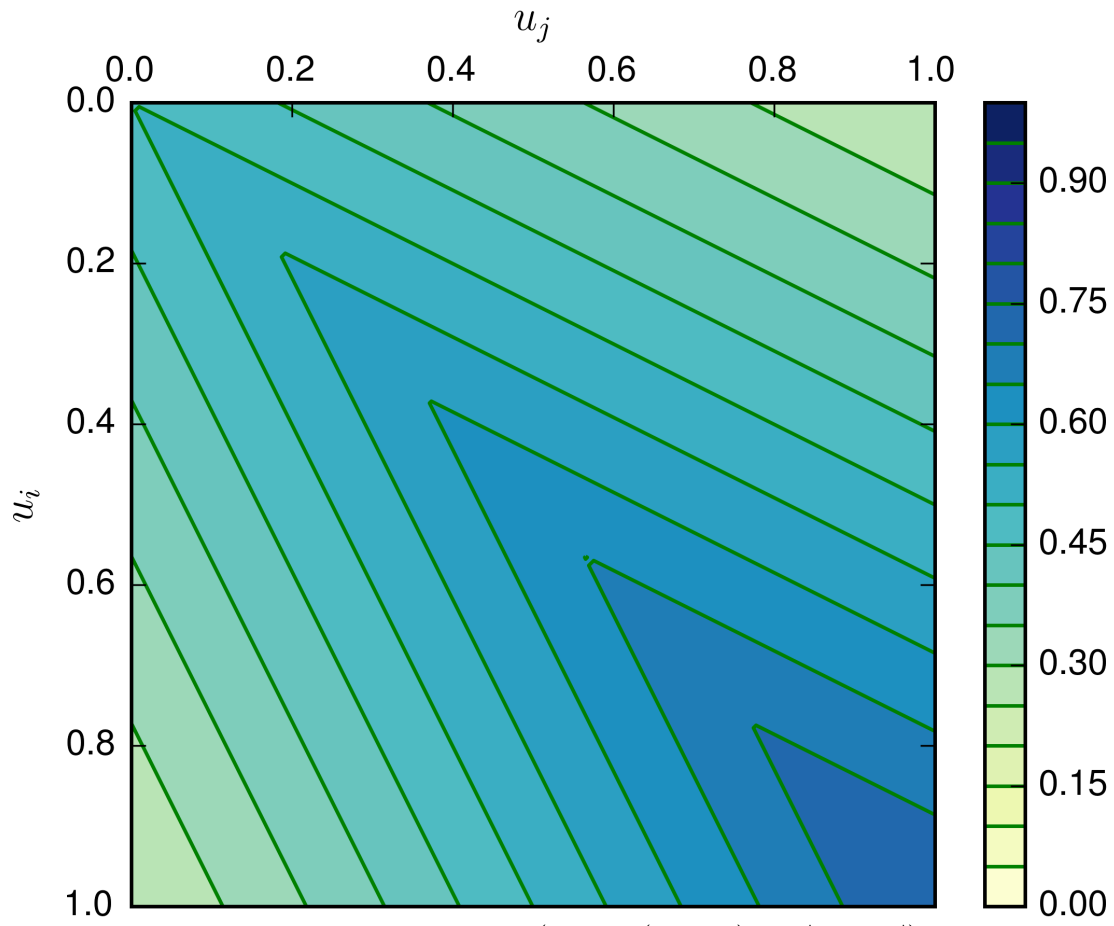
Aldous-Hoover (continued)

- The Aldous-Hoover representation theorem implies that, *under exchangeability*, we can proceed ‘as if’ links formed independently conditional on the agent-specific latent variables $\{U_i\}_{i=1}^{\infty}$ and α .
- A network generating process is:
 1. “Draw” α or choose a graphon;
 2. Draw $U_i \sim \mathcal{U}[0, 1]$ for agents $i = 1, \dots, N$;
 3. Construct \mathbf{D} , by sampling $D_{ij} \mid h_{\alpha}(\bullet, \bullet), U_i, U_j \sim \text{Bernoulli}\left(h_{\alpha}(U_i, U_j)\right)$ 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 large sample 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.

Graphon contour plot



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

Graphon

- The graphon $h_\alpha(u, v)$ is not identifiable...
 - consider the m.p.t. $\varphi(U) = 1 - U$ or $\varphi(U) = 2U \bmod 1$
 - $g_\alpha(U_i, U_j, V_{ij})$ and $g_\alpha(\varphi(U_i), \varphi(U_j), V_{ij})$ generate graphs with the same properties
- ...but link/edge probabilities *are* identifiable (under assumptions).
 - $p_{ij} = \mathbb{E} [D_{ij} | \mathbf{U}] = h_\alpha(U_i, U_j)$

Graphon (Bickel & Chen, 2009)

- For statistical analysis it is convenient to formulate the graphon somewhat differently.
- Consider the network DGP

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

and define

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

- 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 .
 - i.e., $\rho_{\alpha,N} \rightarrow 0$ as $N \rightarrow \infty$
 - in practice we ignore any dependence of $w_\alpha(u, v)$ on N
- The rate at which $\rho_{\alpha,N} \rightarrow 0$ controls the sparsity links
- If $\lambda_N = (N - 1) \rho_{\alpha,N} \rightarrow \lambda > 0$ as $N \rightarrow \infty$ the graph is *sparse*
 - other cases: $\lambda_N = O(N)$ (*dense*) or $\lambda_N = O(\ln N)$ (*semi-dense*)

Edge Probability Estimation

- Define the inner product

$$\langle f, g \rangle = \int f(u) g(u) du$$

with the associated norm

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

- Linking behavior of an agent of type u is summarized by the *graphon slice* $\rho w(u, \bullet)$.
- Measure “distance” between agent i , with $U_i = u$, and agent j , with $U_j = v$, by:

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

Network Neighbors

- $\mathbf{P} = \mathbb{E} [\mathbf{D} | \mathbf{U}]$ denotes the expected adjacency matrix.
- $\mathbf{P}_{i\bullet}$ denotes the i^{th} row of this matrix
- Distance between i and j is

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

- 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 $d_N(i, j) = 0$
 - *i* and *j* have identical (expected) adjacency (matrix) slices.
 - i.e., identical *ex ante* linking behavior
 - *realized* 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*:

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

where $q_i(h_N)$ 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.
- Estimate P_{ij} by the local average

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

- Unfortunately \mathbf{P} is not observed!

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

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

- Need estimates of $\langle \mathbf{P}_{i\bullet}, \mathbf{P}_{i\bullet} \rangle$, $\langle \mathbf{P}_{j\bullet}, \mathbf{P}_{j\bullet} \rangle$ and $\langle \mathbf{P}_{i\bullet}, \mathbf{P}_{j\bullet} \rangle$ 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) term

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

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:

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

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

- We get

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

Finding Neighbors (continued)

- Combining results we have that

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

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

- Zhang et al. (2015) estimate

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

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

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

- Estimated neighborhood* of agent i is then

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

Zhang et al. (2015) Estimate

- Estimate P_{ij} by

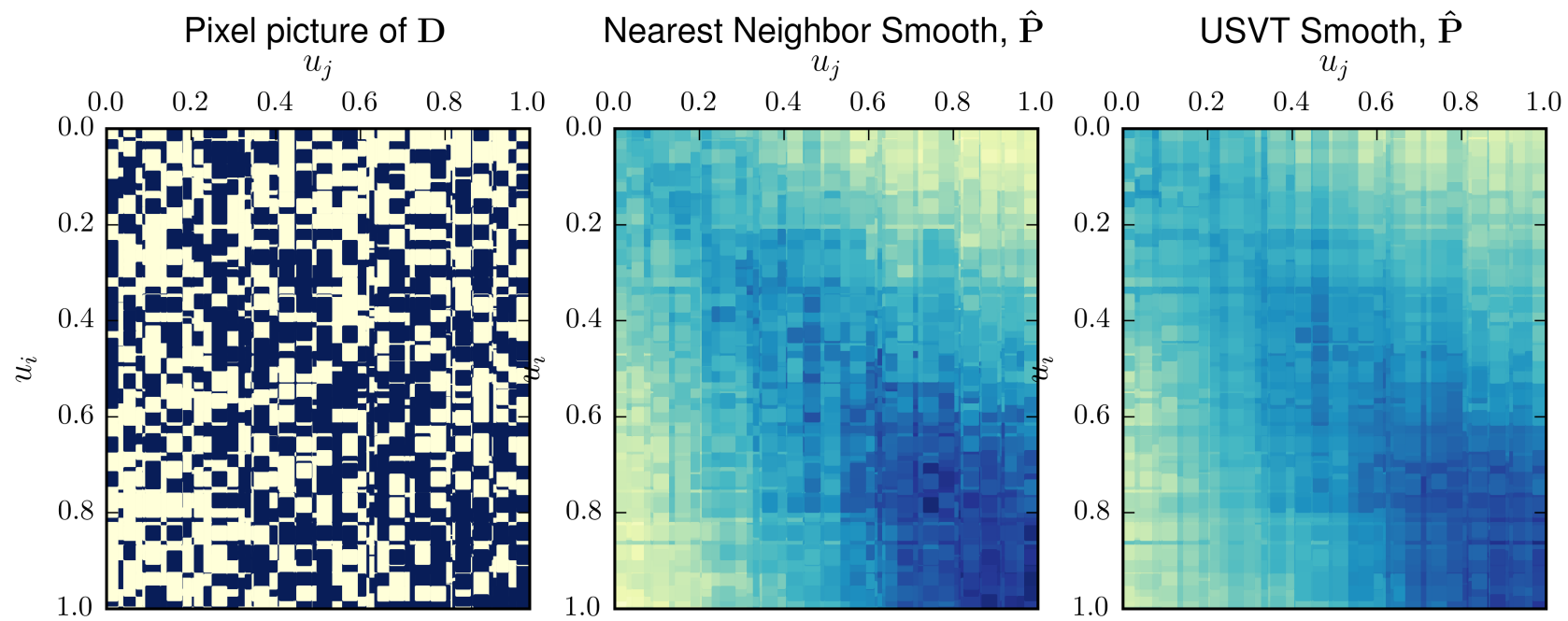
$$\hat{P}_{ij} = \frac{1}{2} \left(\frac{\sum_{k \in \hat{\mathcal{N}}_i} D_{kj}}{|\hat{\mathcal{N}}_i|} + \frac{\sum_{l \in \hat{\mathcal{N}}_j} D_{il}}{|\hat{\mathcal{N}}_j|} \right)$$

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

Alternative Distance Measure

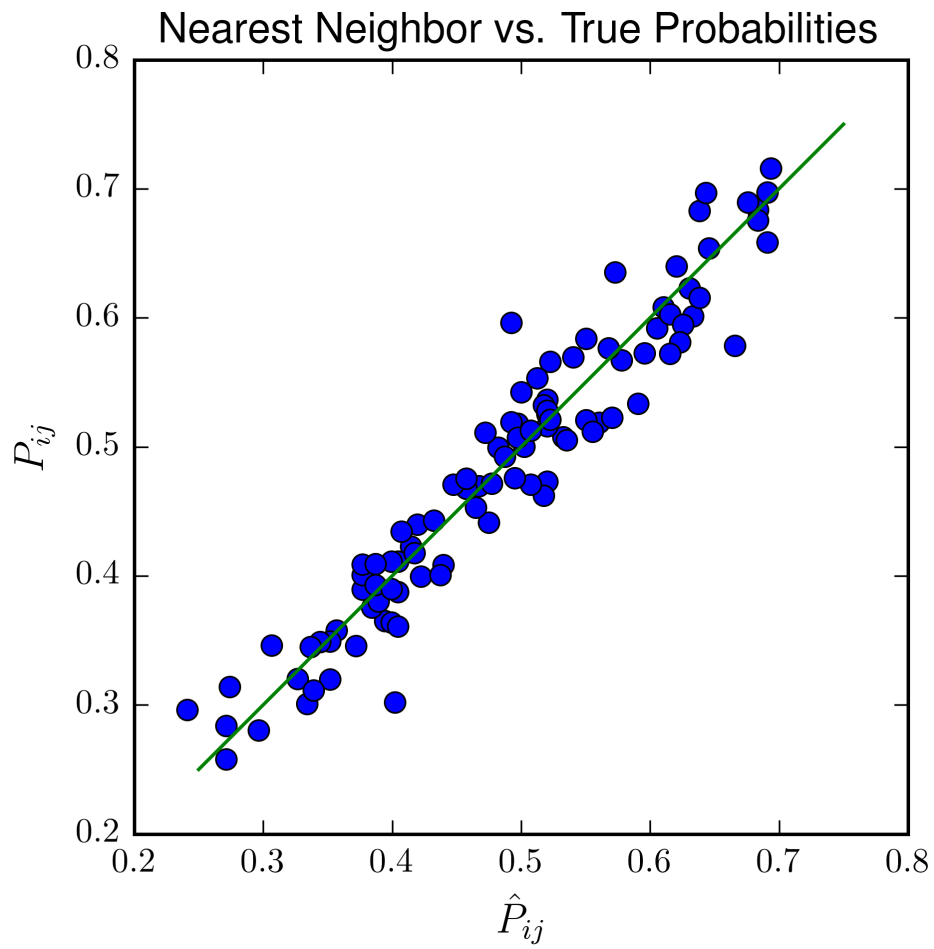
$$\begin{aligned}
 & \max_{l \neq i, j} |\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{l\bullet} \rangle| \\
 & \leq \sum_{l \neq i, j} |\langle \mathbf{P}_{i\bullet} - \mathbf{P}_{j\bullet}, \mathbf{P}_{l\bullet} \rangle| \cdot 1 \\
 \text{(HI)} & \leq \left[\sum_{l \neq i, j} (\langle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

We can use the ‘smoother’ $\hat{d}_N^*(i, j)$ to find nearest neighbors instead.



Goodness-of-Fit

($N = 2,000$, $h_N = 0.1$)



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 $\sum_j \hat{P}_{ij}$
 - should be sufficient to ‘see’ a block 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.