Network Representation Using Graph Root Distributions

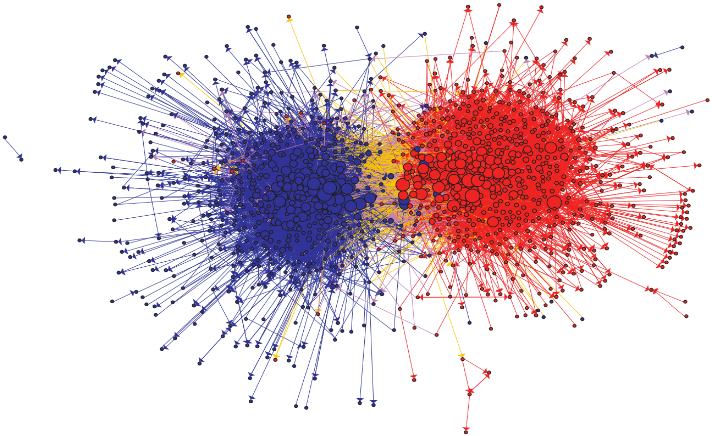
Jing Lei
Department of Statistics and Data Science
Carnegie Mellon University

2018.04

Network Data

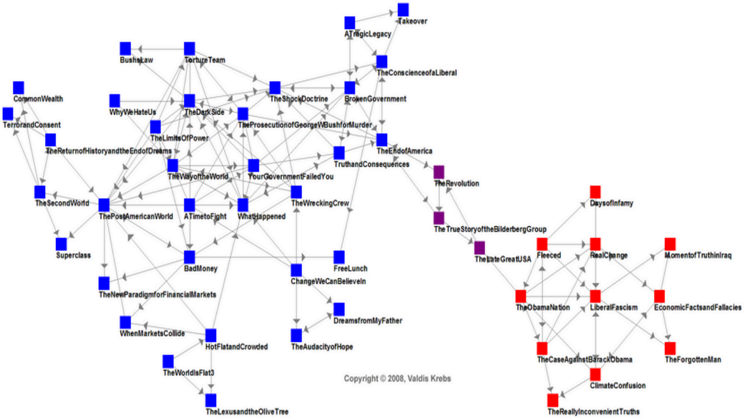
- Network data record interactions (edges) between individuals (nodes).
- From WIKIPEDIA: "... a complex network is a graph (network) with non-trivial topological features ..."
- Examples of "non-trivial topological features"
  - heavy-tail degree distribution (a.k.a "scale-free", "power law")
  - large clustering coefficient (transitivity)
  - community structure: the nodes can be grouped into subsets with dense internal connection.
  - ...

Example: Links Between Political Blogs



[Adamic & Glance ’05] The political blogosphere and the 2004 US election: divided they blog

Example: Co-purchase of Political Books



[V. Krebs ’04] Co-purchased political books on Amazon.

Exchangeable Random Graphs

- Symmetric binary array $\mathbf{A} = (A_{ij} : 1 \leq i < j < \infty)$, $A_{ij} \in \{0, 1\}$.
- Joint exchangeability:

$$(A_{ij} : 1 \leq i < j < \infty) \overset{d}{=} (A_{\sigma(i),\sigma(j)} : 1 \leq i < j < \infty)$$

for all permutation σ

$$\sigma(i) = \begin{cases} i & \text{if } i \notin \{i_0, j_0\}, \\ j_0 & \text{if } i = i_0, \\ i_0 & \text{if } i = j_0 . \end{cases}$$

- Idea: nodes are subjects, so the order does not matter.

A two-way de Finetti Theorem

de Finetti for two-way array (Hoover '79, Aldous '81, Kallenberg '89): All such random graphs must be generated as

$$s_i \overset{iid}{\sim} \text{Unif}(0,1), \quad i \geq 1\,.$$

$$(A_{ij}|s) \overset{indep.}{\sim} \text{Bernoulli}(W(s_i, s_j))\,,\ 1 \leq i < j\,.$$

where $W : [0,1]^2 \mapsto [0,1]$, measurable and symmetric, is called a graphon (graph function).

Popular Special Cases

- The stochastic block model (SBM, Holland *et al* '83): W is block-wise constant.
- The degree corrected block model (DCBM, Karrer & Newman '11): W is block-wise rank-one.
- Random dot product graph (RDPG, Tang *et al*, '13; Rubin-Delanchy *et al*, '17): W is positive semidefinite and low-rank.
- Random geometric graphs (Penrose '03).

Inference Problems

- Estimation
  - Community recovery: find block structure of W in SBM and DCBM.
  - Nonparametric estimation: estimate W from observed $\mathbf{A}_n = (A_{ij} : 1 \leq i,j \leq n)$.

Inference Problems

- Estimation
  - Community recovery: find block structure of W in SBM and DCBM.
  - Nonparametric estimation: estimate W from observed $\mathbf{A}_n = (A_{ij} : 1 \leq i,j \leq n)$.
- Identifiability of W:
  - Let $h : [0,1] \mapsto [0,1]$ be measure-preserving:

    $$\mu(h^{-1}(B)) = \mu(B)\,,\ \forall \text{ measurable } B,$$

    where μ is Lebesgue Measure.
  - $W(\cdot,\cdot)$ and $W(h(\cdot),h(\cdot))$ lead to the same distribution of $\mathbf{A}_n$.

Identifiability of Graphons

- W_1 and W_2 lead to the same distribution of $\mathbf{A}$ if and only if there exist measure-preserving h_1, h_2 such that

$$W_1(h_1(\cdot), h_1(\cdot)) = W_2(h_2(\cdot), h_2(\cdot)), \quad \text{a.e.}$$

- Cut-distance:

$$\delta_\square(W_1, W_2) = \inf_{h_1, h_2} \sup_{S,T} \left| \int_{S \times T} [W_1(h_1(s), h_1(t)) - W_2(h_2(s), h_2(t))] \right|$$

- When $\delta_\square(W_1, W_2) = 0$, write $W_1 \overset{w.i.}{=} W_2$ (*weakly isomorphic*), which defines an equivalence relationship on $\mathscr{W}_0 := \{W : [0,1]^2 \mapsto [0,1], \text{ symmetric}\}$.

Identifiability of Graphons

- In general, we can only hope to recover W up to a measure-preserving change-of-variable transform.
- Existing methods assume smoothness to specify a particular member in the equivalence class (Wolfe & Olhede '13, Airoldi *et al* '13, Gao *et al* '15, Klopp *et al* '17).

The latent space approach

- Sample $\xi_1, \ldots, \xi_n$ independently from a distribution F on $\mathbb{R}^d$.
- Connect nodes i, j by $f(\xi_i, \xi_j)$, for some simple function f, such as inner products and distances [Hoff et al '02, Hoff 07, Tang et al 13].
- The node embedding carries rich, interpretable structures about the network.

The latent space approach

- Sample $\xi_1, \ldots, \xi_n$ independently from a distribution F on $\mathbb{R}^d$.
- Connect nodes i, j by $f(\xi_i, \xi_j)$, for some simple function f, such as inner products and distances [Hoff et al '02, Hoff 07, Tang et al 13].
- The node embedding carries rich, interpretable structures about the network.
- Question: Can we use latent space models with simple f to study exchangeable random graphs, with better identifiability?

The latent space approach

- Sample $\xi_1, \ldots, \xi_n$ independently from a distribution F on $\mathbb{R}^d$.
- Connect nodes i, j by $f(\xi_i, \xi_j)$, for some simple function f, such as inner products and distances [Hoff et al '02, Hoff 07, Tang et al 13].
- The node embedding carries rich, interpretable structures about the network.
- Question: Can we use latent space models with simple f to study exchangeable random graphs, with better identifiability?
- Yes. Use graph root distributions on a separable Kreĭn space.

Graph Root Distributions on a Kreĭn Space

Definition: Kreĭn Space

A Kreĭn space $\mathscr{K} = \mathscr{H}_+ \ominus \mathscr{H}_-$ is the direct sum of two Hilbert spaces $\mathscr{H}_+$, $\mathscr{H}_-$, with inner product (for $x, x' \in \mathscr{H}_+$, $y, y' \in \mathscr{H}_-$)

$$\langle (x;y),(x';y')\rangle_{\mathscr{K}} = \langle x,x'\rangle_{\mathscr{H}_+} - \langle y,y'\rangle_{\mathscr{H}_-} .$$

$\mathscr{K}$ is isomorphic to a Hilbert space $\mathscr{H} = \mathscr{H}_+ \oplus \mathscr{H}_-$ with norm $\|\cdot\|_{\mathscr{K}}$: for $z = (x;y) \in \mathscr{K}$

$$\|z\|_{\mathscr{K}} = \|(x,y)\| = \left(\|x\|^2_{\mathscr{H}_+} + \|y\|^2_{\mathscr{H}_-}\right)^{1/2} .$$

Graph Root Distributions on a Kreĭn Space

Definition: Graph Root Distribution (GRD)

A graph root distribution is a probability distribution F on $\mathscr{K}$ such that for $Z, Z' \overset{iid}{\sim} F$,

$$\mathbb{P}\left(\langle Z, Z' \rangle_{\mathscr{K}} \in [0,1]\right) = 1\,.$$

From GRD's to Exchangeable Random Graphs

Given a GRD F on $\mathscr{K}$, one can generate exchangeable random graphs as follows.

1. Generate $(Z_i : i \geq 1) \overset{iid}{\sim} F$.
2. Generate A_{ij} independently from Bernoulli$(\langle Z_i, Z_j \rangle_{\mathscr{K}})$.

Related work

- Eigenmodel [Hoff '07]
- Random dot-product graph [Tang et al '13; Rubin-Delanchy et al '17]

Interpretation of GRD

- Edge probability $= \langle Z_i, Z_j \rangle_{\mathscr{K}} = \langle X_i, X_j \rangle - \langle Y_i, Y_j \rangle$
- Nodes i, j are more likely to connect if
  - $\|X_i\|, \|X_j\|$ are large (active nodes)
  - $\langle X_i / \|X_i\|, X_j / \|X_j\| \rangle$ is large (good match)
- Analogous interpretations for negative components Y_i, Y_j.

Questions to be answered about GRD's

- Existence: What kind of exchangeable random graphs can be generated by GRD's?
- Uniqueness/Identifiability: When do two GRD's lead to the same distribution of exchangeable random graphs?
- What is the relationship between GRD and graphon?
- What is an interesting topology in the space of GRD's?
- How to estimate the generating GRD from an observed network?

Existence of GRD Representation

- View a graphon W as the kernel of an integral operator on $L^2([0,1])$. By compactness, W admits spectral decomposition

$$W(s,s') = \sum_{j=1}^{\infty} \lambda_j \phi_j(s)\phi_j(s') - \sum_{j=1}^{\infty} \gamma_j \psi_j(s)\psi_j(s')$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$, $\gamma_1 \geq \gamma_2 \geq \ldots > 0$.

Existence of GRD Representation

Theorem

If W is trace-class (i.e. $\sum_{j\geq 1}(\lambda_j + \gamma_j) < \infty$), then there exists a GRD F on a separable Kreĭn space $\mathscr{K}$ such that W and F lead to the same exchangeable random graph distribution.

Idea of Proof

- Recall that

$$W(s,s') = \sum_{j=1}^{\infty} \lambda_j \phi_j(s)\phi_j(s') - \sum_{j=1}^{\infty} \gamma_j \psi_j(s)\psi_j(s')$$

- $Z(s) = (X(s); Y(s)) : [0,1] \mapsto \mathscr{K}$ with

$$X(s) = (\lambda_j^{1/2}\phi_j(s) : j \geq 1), \;\; Y(s) = (\gamma_j^{1/2}\psi_j(s) : j \geq 1).$$

- summability of $\lambda_j, \gamma_j \Rightarrow \|X\|, \|Y\| < \infty$ a.s. $\Rightarrow Z$ is well-defined.
- F is the measure induced by Z with $s \sim U(0,1)$.
- summability of $\lambda_j, \gamma_j \Rightarrow \; W(\cdot,\cdot) = \langle Z(\cdot), Z(\cdot)\rangle_{\mathscr{K}}$ a.e.

Idea of Proof

- Recall that

$$W(s,s') = \sum_{j=1}^{\infty} \lambda_j \phi_j(s)\phi_j(s') - \sum_{j=1}^{\infty} \gamma_j \psi_j(s)\psi_j(s')$$

- $Z(s) = (X(s); Y(s)) : [0,1] \mapsto \mathscr{K}$ with

$$X(s) = (\lambda_j^{1/2}\phi_j(s) : j \geq 1), \;\; Y(s) = (\gamma_j^{1/2}\psi_j(s) : j \geq 1).$$

- summability of $\lambda_j, \gamma_j \Rightarrow \|X\|, \|Y\| < \infty$ a.s. $\Rightarrow Z$ is well-defined.
- F is the measure induced by Z with $s \sim U(0,1)$.
- summability of $\lambda_j, \gamma_j \Rightarrow \; W(\cdot,\cdot) = \langle Z(\cdot), Z(\cdot)\rangle_{\mathscr{K}}$ a.e.
- Z can be viewed as the square root of W.

Existence of GRD Representation

Special cases:

- Continuity: $W = W_+ - W_-$ with continuous and positive semidefinite W_+, W_- (Mercer's theorem).
- Smoothness: W is smooth.

Identifiability of GRD's

- When do two GRD's lead to the same sampling distribution of exchangeable random graphs?
- Concatenation. Let $Z = (X;Y) \sim F$, and R an arbitrary random variable. Let $\tilde{Z} = (\tilde{X};\tilde{Y})$ with $\tilde{X} = (X,R)$, $\tilde{Y} = (Y,R)$.
- Inner product preserving transforms. $H : \mathscr{K} \mapsto \mathscr{K}$ such that $\langle z, z' \rangle_{\mathscr{K}} = \langle Hz, Hz' \rangle_{\mathscr{K}}$. Let $Z \sim F$ and $\tilde{Z} = HZ$.
  - Direct sum of orthogonal transforms. Let Q_+, Q_- be orthogonal transforms on $\mathscr{H}_+$, $\mathscr{H}_-$. Let $Z = (X;Y) \sim F$, and $\tilde{Z} = (Q_+X;Q_-Y)$.

Identifiability of GRD's

- When do two GRD's lead to the same sampling distribution of exchangeable random graphs?
- Concatenation. Let $Z = (X;Y) \sim F$, and R an arbitrary random variable. Let $\tilde{Z} = (\tilde{X};\tilde{Y})$ with $\tilde{X} = (X,R)$, $\tilde{Y} = (Y,R)$.
- Inner product preserving transforms. $H : \mathscr{K} \mapsto \mathscr{K}$ such that $\langle z,z' \rangle_{\mathscr{K}} = \langle Hz,Hz' \rangle_{\mathscr{K}}$. Let $Z \sim F$ and $\tilde{Z} = HZ$.
  - Direct sum of orthogonal transforms. Let Q_+, Q_- be orthogonal transforms on $\mathscr{H}_+$, $\mathscr{H}_-$. Let $Z = (X;Y) \sim F$, and $\tilde{Z} = (Q_+X;Q_-Y)$.
  - Hyperbolic rotations.

Hyperbolic Rotations: An Example

- Let $\mathscr{H}_+ = \mathscr{H}_- = \mathbb{R}$, $z = (x;y) \in \mathbb{R}^2$.
- An example of hyperbolic rotation is, for $\theta \in \mathbb{R}^1$,

$$
\begin{aligned}
H[(x,y)^T] &= \left(\frac{e^\theta + e^{-\theta}}{2} x + \frac{e^\theta - e^{-\theta}}{2} y,\ \frac{e^\theta - e^{-\theta}}{2} x + \frac{e^\theta + e^{-\theta}}{2} y \right)^T \\
&= \begin{bmatrix} \cosh(\theta) & \sinh(\theta) \\ \sinh(\theta) & \cosh(\theta) \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= H_\theta \begin{pmatrix} x \\ y \end{pmatrix}.
\end{aligned}
$$

- $H_\theta^T \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} H_\theta = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

Identifiability of GRD's: Where is the Hope?

Key observation:

- Both concatenation and hyperbolic rotation necessarily mix up the positive and negative components, so they can be precluded by disentangling the positive and negative components.

Identifiability of GRD's

Let Q_+, Q_- be orthogonal transforms on $\mathscr{H}_+$, $\mathscr{H}_-$. Define $Q = (Q_+ \oplus Q_-) : \mathscr{K} \mapsto \mathscr{K}$ as $Q(x;y) = (Q_+x; Q_-y)$.

Theorem

Two square-integrable GRD's F_1, F_2 with uncorrelated positive and negative components lead to the same sampling distribution of exchangeable random graphs if and only if there exists $Q = Q_+ \oplus Q_-$ such that $Z_1 \sim F_1 \Leftrightarrow Z_2 = QZ_1 \sim F_2$ (denoted as $F_1 \overset{o.t.}{=} F_2$).

Proof Sketch of Identifiability

- For $i = 1, 2$, let $Z_i(\cdot) : [0, 1] \mapsto \mathscr{K}$ be an *inverse transform sampling* (ITS) mapping such that $s \sim U(0, 1) \Rightarrow Z_i(s) \sim F_i$.
- Let $W_i(\cdot, \cdot) = \langle Z_i(\cdot), Z_i(\cdot) \rangle_{\mathscr{K}}$.
- By assumption, $W_1 \overset{w.i.}{=} W_2$.
- Can choose appropriate orthogonal transforms so that $W_i(\cdot, \cdot) = \langle Z_i(\cdot), Z_i(\cdot) \rangle_{\mathscr{K}}$ indeed corresponds to the spectral decomposition of W_i.
- Apply Kallenberg's representation theorem of exchangeable random arrays using spectral decompositions.

Existence and Identifiability of GRD: Summary

Corollay

There exists a one-to-one correspondence between trace-class graphons (up to "$\stackrel{w.i.}{=}$") and square-integrable GRD's with uncorrelated positive and negative components (up to "$\stackrel{o.t.}{=}$").

Canonical GRD. Given a square integrable GRD, one can always make the positive and negative components uncorrelated and choose a canonical pair of orthogonal transforms so that the covariance is diagonalized.

Distances between GRD equivalence classes

- Given two GRD's F_1 and F_2, each representing their own equivalence class, how do we measure the difference between them?
- For graphons, the cut-distance is linked to the large-sample subgraph densities.

$$\delta_{\square}(W_1, W_2) = \inf_{h_1, h_2} \sup_{S,T} \left| \int_{S\times T} [W_1(h_1(s), h_1(t)) - W_2(h_2(s), h_2(t))] \right|$$

- Taking inf over h_1 and h_2 is to find a common ITS for two distributions, which is essentially coupling.

Wasserstein Distance

- Let F_1, F_2 be two distributions on $\mathscr{K}$, the Wasserstein distance between F_1, F_2 is

$$d_{\mathrm{w}}(F_1, F_2) = \inf_{\nu \in \mathscr{V}(F_1, F_2)} \mathbb{E}_{(Z_1, Z_2) \sim \nu} \|Z_1 - Z_2\|,$$

where $\mathscr{V}(F_1, F_2)$ is the set of all distributions ν on $\mathscr{K} \times \mathscr{K}$ with marginals being F_1, F_2.

Orthogonal Wasserstein Distance

Definition: Orthogonal Wasserstein Distance

The *orthogonal Wasserstein distance* between two square-integrable GRD's F_1, F_2 is

$$d_{\text{ow}}(F_1, F_2) = \inf_{Q_+, Q_-} \inf_{\nu \in \mathscr{V}(F_1, F_2)} \mathbb{E}_{(Z_1, Z_2) \sim \nu} \|Z_1 - (Q_+ \oplus Q_-) Z_2\|,$$

where the first inf is taken over all pairs of orthogonal transforms on $\mathscr{H}_+$, $\mathscr{H}_-$.

Remark: OWD measures the distance between two equivalence classes of GRD's.

Stronger Topology: $d_{\text{ow}}(\cdot,\cdot) \succeq \delta_\square(\cdot,\cdot)$

Theorem

Let F, F_1, F_2, ..., F_N, ..., be square-integrable GRD's and W, W_1, W_2, ..., W_N, ..., the corresponding graphons. Then

$$\delta_\square(W_1, W_2) \leq (\mathbb{E}_{F_1}\|Z\| + \mathbb{E}_{F_2}\|Z\|)\, d_{\text{ow}}(F_1, F_2)\,.$$

Moreover, if $d_{\text{ow}}(F_N, F) \to 0$, then $\delta_\square(W_N, W) \to 0$.

Estimating the GRD

- Data: $\mathbf{A}_n = (A_{ij} : 1 \leq i,j \leq n)$, symmetric, $A_{ii} = 0$.
- Model: $A_{ij} \overset{ind.}{\sim}$ Bernoulli$(\langle Z_i, Z_j \rangle_{\mathscr{K}})$, $1 \leq i < j \leq n$, where $(Z_i : 1 \leq i \leq n) \overset{iid}{\sim} F$, a square-integrable GRD on a $\mathscr{K}$.
- Let $\mathscr{H}_+ = \mathscr{H}_- = \{x \in \mathbb{R}^{\mathbb{N}} : \sum_{j \geq 1} x_j^2 < \infty\}$.
- To identify, let $Z = (X; Y) \sim F$ have diagonal covariance:

$$\mathbb{E}XX^T = \mathrm{diag}(\lambda_1, \lambda_2, \ldots)\,, \;\; \mathbb{E}YY^T = \mathrm{diag}(\gamma_1, \gamma_2, \ldots)\,, \;\; \mathbb{E}XY^T = 0\,.$$

Truncated Weighted Signed Spectral Embedding

- Let $\mathbf{A}_n = \sum_j \hat{\lambda}_j \hat{a}_j \hat{a}_j^T - \sum_j \hat{\gamma}_j \hat{b}_j \hat{b}_j^T$ be the eigen decomposition of $\mathbf{A}_n$, with positive eigenvalues $\hat{\lambda}_j$ and (absolute) negative eigenvalues $\hat{\gamma}_j$ ranked in decreasing order.
- Let p_1, p_2 be positive integers to be specified later.
- For $1 \leq i \leq n$ let $\hat{Z}_i = (\hat{X}_i; \hat{Y}_i)$ with
  $\hat{X}_i = (\hat{\lambda}_1^{1/2} \hat{a}_{1i} \ , \ \ldots \ , \ \hat{\lambda}_{p_1}^{1/2} \hat{a}_{p_1 i} \ , \ 0 \ , \ldots)$
  $\hat{Y}_i = (\hat{\gamma}_1^{1/2} \hat{b}_{1i} \ , \ \ldots \ , \ \hat{\gamma}_{p_2}^{1/2} \hat{b}_{p_2 i} \ , \ 0 \ , \ldots)$
- $\hat{F}$ is the distribution that puts $1/n$ mass at each $\hat{Z}_i$.

Regularity Conditions

- Eigen decay and eigen gap: for some $1 < \alpha \leq \beta$ and all $j \geq 1$

$$\lambda_j, \gamma_j \asymp j^{-\alpha}, \quad \min(\lambda_j - \lambda_{j+1},\ \gamma_j - \gamma_{j+1}) \gtrsim j^{-\beta}$$

- Fourth moment bounded: $\mathbb{E}_{Z \sim F} \|Z\|^4 < \infty$.
- These are standard assumptions in functional data analysis, where truncated PCA is used to recover sample curves in $L^2([0,1])$.

Estimation Error Bound

Let $\tilde{F}$ be the distribution that puts $1/n$ mass at $\tilde{Z}_i = (\tilde{X}_i, \tilde{Y}_i)$, with $\tilde{X}_i = (X_{i1}, \ldots, X_{ip}, 0, \ldots)$, $\tilde{Y}_i = (Y_{i1}, \ldots, Y_{ip}, 0, \ldots)$

Theorem

Under the regularity conditions, if $p_1 = p_2 = p = o(n^{1/(2\beta+\alpha)})$ then

$$d_{\mathrm{w}}(\hat{F}, \tilde{F}) = O_P(p^{-(\alpha-1)/2})$$

and

$$d_{\mathrm{w}}(\hat{F}, F) = O_P(p^{-(\alpha-1)/2} + n^{-1/p}).$$

Proof Sketch

- Treat positive and negative components separately.
- $X_i = (X_{ij} : j \geq 1)$
  $\hat{X}_i = (\hat{\lambda}_1^{1/2}\hat{a}_{1i} \, , \, \dots \, , \, \hat{\lambda}_p^{1/2}\hat{a}_{pi} \, , \, 0 \, , \, \dots)$
  $\tilde{X}_i = (X_{i1} \, , \, \dots \, , \, X_{ip} \, , \, 0 \, , \, \dots)$
- $\mathbf{G}_n = (\langle (X_i; Y_i), (X_j; Y_j) \rangle_{\mathscr{K}} : \ 1 \leq i,j \leq n) = \mathbb{E}(\mathbf{A}_n | Z_1, \dots, Z_n)$.
- $\mathbf{G}_{n,X} = (\langle X_i, X_j \rangle : \ 1 \leq i,j \leq n)$, $\mathbf{G}_n = \mathbf{G}_{n,X} - \mathbf{G}_{n,Y}$
- $\tilde{\mathbf{G}}_{n,X} = (\langle \tilde{X}_i, \tilde{X}_j \rangle : 1 \leq i,j \leq n)$
- Spectral perturbation: $\mathbf{A}_n \approx \mathbf{G}_n$
  $\Rightarrow \hat{X}_i \approx$ T.W. spectral embedding of $\mathbf{G}_n$.
- Uncorrelatedness+eigen-decay: $\mathbf{G}_n \approx \mathbf{G}_{n,X} \approx \tilde{\mathbf{G}}_{n,X}$ in leading subspace $\Rightarrow \ \hat{X}_i \approx \tilde{X}_i \ \Rightarrow d_{\mathrm{w}}(\hat{F}, \tilde{F}) \approx 0$.

Sparse Graphs

- GRD's, as graphons, can only generate dense graphs with total number of edges proportional to n^2 .
- Given graphon W and node sample size n, one can consider sparse sampling with a sparsity rate $\rho_n = o(1)$ (see e.g. [Bickel & Chen '09]):

$$A_{ij} \sim \text{Bernoulli}(\rho_n W(s_i, s_j)) .$$

- In GRD representation, the sparse sampling is equivalent to generating the network using $\rho_n^{1/2} F$ (scaling down F by a factor of $\rho_n^{1/2}$).

Sparse Graphs

Theorem

Assume $\mathbf{A}_n$ is generated by a GRD F with sparsity parameter ρ_n. Under the regularity conditions, if $\beta \geq 3\alpha/2$ and $p = o(n^{1/(2\beta+\alpha)} \wedge (n\rho_n)^{1/(2\beta)})$ then

$$d_{\mathrm{w}}(\rho_n^{-1/2}\hat{F}, F) = O_P\left(\frac{p^{\beta-(\alpha-1)/2}}{(n\rho_n)^{1/2}} + p^{-(\alpha-1)/2} + n^{-1/p}\right).$$

Other values of β and p are allowed, but complicated to present.

How to Choose p_1, p_2?

- The truncated empirical eigen decomposition resembles methods in functional principal components analysis, where one can choose the number of PC's by fraction of variance explained.
- However, network data are different
  - Low-rank: the number of significant eigen components is usually small;
  - High noise: network data are observed with entry-wise independent noise.
- Singular value thresholding [Chatterjee '14]: use eigenvalues larger than $\sqrt{n}$.

Examples

- $B \in [0,1]^{K \times K}$, $B = B^T$.
- Stochastic block models: mixture of point mass

$$\mathbb{E}(A_{ij}) = B_{g_i,g_j}\,, \quad g_i \in \{1,\ldots,K\}\,.$$

- Degree corrected block models: mixture of 1-D subspaces

$$\mathbb{E}(A_{ij}) = \psi_i \psi_j B_{g_i,g_j}\,, \quad g_i \in \{1,\ldots,K\}\,, \quad \psi_i \in \mathbb{R}^+\,.$$

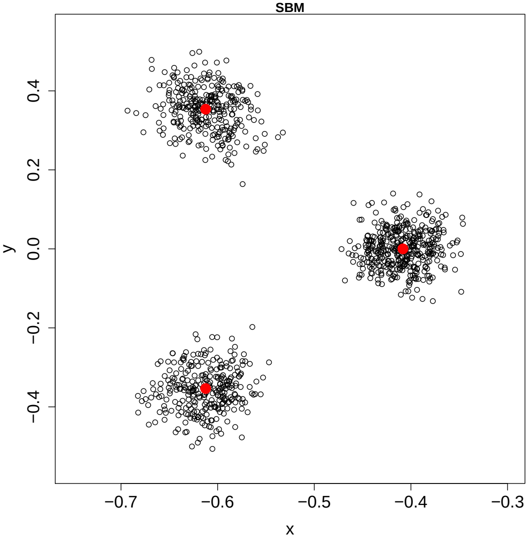
- Mixed membership block models: convex polytope

$$\mathbb{E}(A_{ij}) = a_i^T B a_j\,, \quad a_i \in (K-1) \text{ dim. simplex}\,.$$

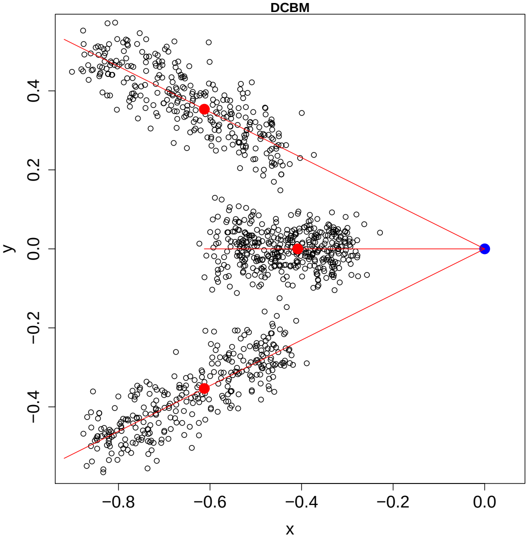
Simulation, $K = 3$

- $K = 3$, $g_i \sim \text{Multinomial}(0.3, 0.3, 0.4)$, $n = 1000$.
- $B = \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/6 \end{bmatrix}$
- Three communities but $\text{rank}(B) = 2$, with one positive and one negative component.
- DCBM node activeness: $\psi_i \sim U(0.7, 1.4)$.
- MMBM node mixture: $a_i \sim \text{Dir}(0.5, 0.5, 0.5)$.

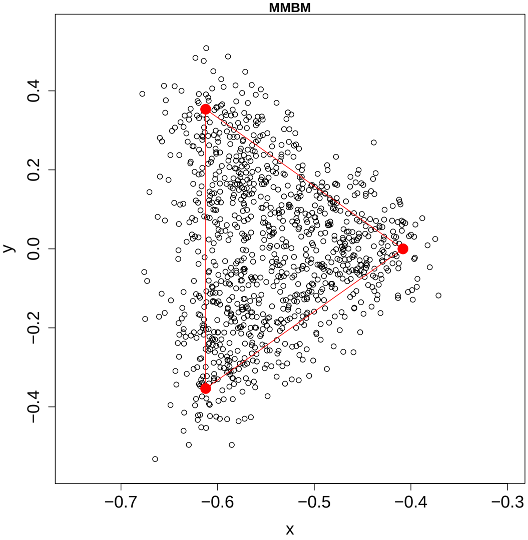
SBM: Point Mass Mixture



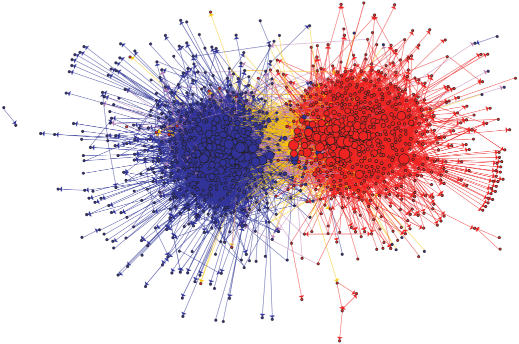
DCBM: Subspace Mixture



MMBM: Convex Polytope

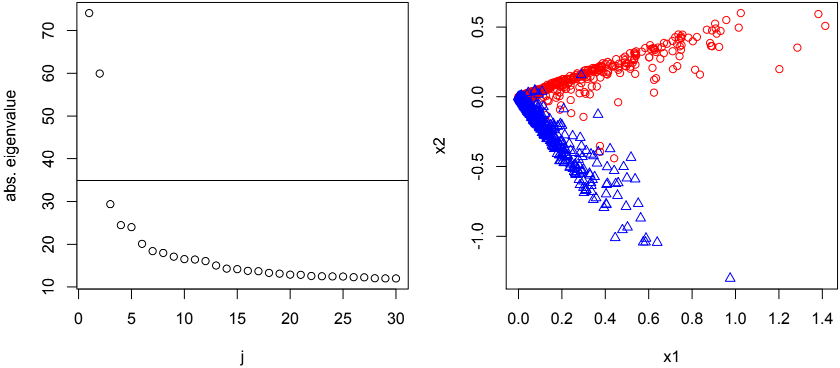


Data example: U.S. political blogs

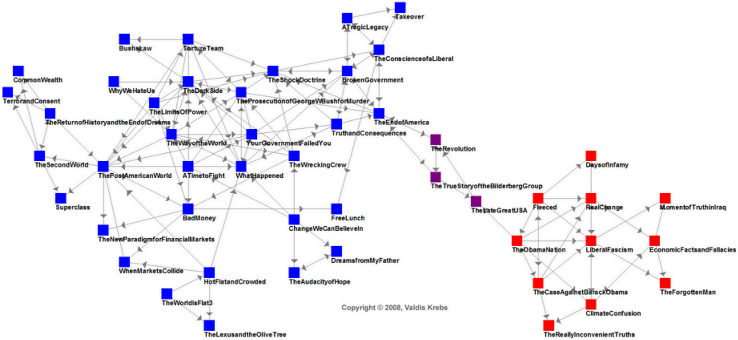


- [Adamic & Glance '05] Snapshots of weblogs shortly before 2004 U.S. Presidential Election. Nodes: weblogs; edges: hyperlinks.
- Fitted well by a DCBM with two clusters.

Political Blogs Data

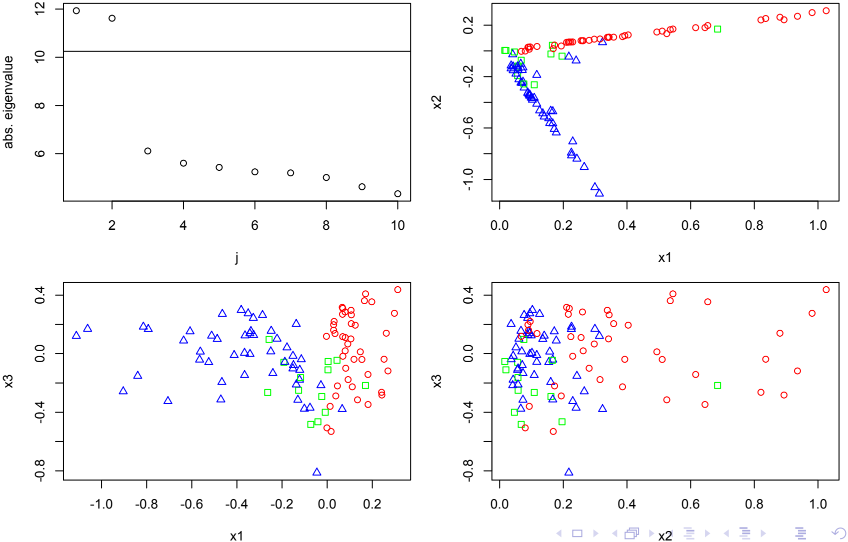


Political Books Data



Co-purchase of political books on Amazon (V. Krebs '04)

Political Books Data



Next

- GRD with logistic link: $A_{ij} \sim \text{Bernoulli}\left(\frac{1}{1+e^{-\langle Z_i, Z_j \rangle_{\mathscr{K}}}}\right)$
- Two sample testing: are $\mathbf{A}_1$ and $\mathbf{A}_2$ from the same GRD?
- Bi-partite graph: $\mathscr{A}$ is asymmetric. e.g. gene-cell matrix.
- Multiple networks: modeling and predicting the movement of latent node embeddings.

Thank You! Questions?

1. Lei, J. “Network Representation Using Graph Root Distributions”, `arXiv:1802.09684`
2. Code: easy to write but also available upon request
3. Slides: `www.stat.cmu.edu/~jinglei/talk.shtml`