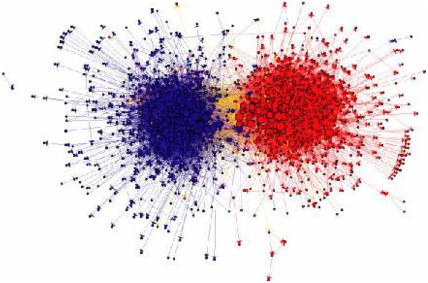
Inference in large random graphs

Laurent Massoulié

Inria

January 2021

Community Detection: cluster nodes $i \in V$ of graph $G = (V, E)$, into subsets $V_1, \ldots, V_K$ of "statistically similar nodes"



Applications:
-recommendation (graph between users and products);
-biology (graph of chemical interactions between proteins);
-...

$\to$ Focus on $G = (V, E)$ drawn from the **Stochastic Block Model** $\mathcal{G}(n, \alpha, P)$:

- $\alpha = \{\alpha_1, \ldots, \alpha_K\}$ probability distribution on $[K]$
- $P \in [0,1]^{K\times K}$: Symmetric matrix
- $\{\sigma_i\}_{i\in[n]}$: i.i.d., $\sim \alpha$
- Conditionally on $\sigma_{[n]} := \{\sigma_1, \ldots, \sigma_n\}$, edge $(i,j) \in E$ with probability P_{σ_i,σ_j}

$\rightarrow$ Focus on $G = (V, E)$ drawn from the **Stochastic Block Model** $\mathcal{G}(n, \alpha, P)$:

- $\alpha = \{\alpha_1, \ldots, \alpha_K\}$ probability distribution on $[K]$
- $P \in [0, 1]^{K \times K}$: Symmetric matrix
- $\{\sigma_i\}_{i \in [n]}$: i.i.d., $\sim \alpha$
- Conditionally on $\sigma_{[n]} := \{\sigma_1, \ldots, \sigma_n\}$, edge $(i, j) \in E$ with probability P_{σ_i, σ_j}

$\rightarrow$ Generative model that generalizes the Erdős-Rényi random graph $\mathcal{G}(n, p)$

$\rightarrow$ Community Detection: inference of node attributes σ_i from observation $G = (V, E)$

$\rightarrow$ Spectral methods and phase transitions on feasibility of community detection

Graphical models, a generalization of the Ising model on $\{-1,1\}^n$:
$\mathbb{P}(X_{[n]} = x_{[n]}) \propto e^{-h\sum_{i\in[n]} x_i + \sigma \sum_{i,j:i\sim j} x_i x_j}$

Goal: Infer unobserved variables X_i of nodes i of graph G

Graphical models, a generalization of the Ising model on $\{-1,1\}^n$:
$\mathbb{P}(X_{[n]} = x_{[n]}) \propto e^{-h\sum_{i\in[n]} x_i + \sigma \sum_{i,j:i\sim j} x_i x_j}$

Goal: Infer unobserved variables X_i of nodes i of graph G

$\rightarrow$ Belief Propagation Algorithm

$\rightarrow$ Emphasis on **Tree Reconstruction** problem: infer characteristics X_a where a: ancestor in genealogical tree, from characteristics of its descendants

$\rightarrow$ Phase transitions on feasibility of inference, and links to community detection

Hypothesis tests: does observed graph G have some structure (ex: is drawn from Stochastic Block Model with $K > 1$ blocks) or is it "totally random", i.e. an Erdős-Rényi graph?

Special Case: the Planted Clique Problem

$\rightarrow$ Highlight existence of two kinds of phase transitions:

- **Informational**, i.e. is there enough information present in the observation
- **Computational**, i.e. can the information present in the observation be extracted in polynomial time

Outline for today

- Background results from Linear Algebra
- Bounds on spectral radius of random matrices
- Spectral methods for community detection in Stochastic Block Model

Singular Value Decomposition (SVD)

SVD of matrix $X \in \mathbb{C}^{n\times p}$:

$X = U\Lambda V^* = \sum_{i=1}^{n\wedge p} \sigma_i u_i v_i^*$, where:

$U = (u_1, \ldots, u_n) \in \mathbb{C}^{n\times n}$, $U^*U = I_n$, (u_i: i-th left singular vector)

$V = (v_1, \ldots, v_p) \in \mathbb{C}^{p\times p}$, $V^*V = I_p$, (v_i: i-th right singular vector)

$$\Lambda \in \mathbb{R}^{n\times p},\ \Lambda = \begin{pmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \ddots & & \vdots \\ 0 & 0 & \sigma_{n\wedge p} & 0 \end{pmatrix},$$

$\sigma_1 \geq \cdots \geq \sigma_{n\wedge p} \geq 0$, ($\sigma_i$: i-th singular value)

Principal Component Analysis (Karl Pearson "On Lines and Planes of Closest Fit to Systems of Points in Space", 1901)

Definition

For matrix $X \in \mathbb{C}^{n\times p}$ with SVD $X = \sum_{i=1}^{n\wedge p} \sigma_i u_i v_i^*$,

Operator norm: $\|X\|_{op} := \sup_{u\in\mathbb{C}^p} \frac{\|Xu\|}{\|u\|} = \sigma_1$;

Frobenius norm: $\|X\|_F := \sqrt{\sum_{i,j} X_{ij}^2} = \sqrt{\sum_i \sigma_i^2}$.

For $r < n \wedge p$, let $X_r := \sum_{i=1}^{r} \sigma_i u_i v_i^*$. Then:

X_r best rank-r approximation of X both for $\|\cdot\|_F$ and $\|\cdot\|_{op}$, with

$$\inf_{Y:\ rk(Y)=r} \|X - Y\|_{op} = \|X - X_r\|_{op} = \sigma_{r+1},$$

$$\inf_{Y:\ rk(Y)=r} \|X - Y\|_F = \|X - X_r\|_F = \sqrt{\sum_{i>r} \sigma_i^2}.$$

Perturbation results: Eigenvalues

Definition

For $M \in \mathbb{C}^{n \times n}$,
spectral radius $\rho(M) := \sup\{|\lambda|,\ \lambda \in \mathrm{Spectrum}(M)\}$.

For Hermitian $M \in \mathbb{C}^{n \times n}$, order its (real) eigenvalues as $\lambda_1(M) \geq \cdots \geq \lambda_n(M)$.

Then $\rho(M) = \sigma_1(M) = \max(|\lambda_1(M)|, |\lambda_n(M)|)$.

Perturbation results: Eigenvalues

Definition

For $M \in \mathbb{C}^{n\times n}$,
spectral radius $\rho(M) := \sup\{|\lambda|,\ \lambda \in \mathrm{Spectrum}(M)\}$.

For Hermitian $M \in \mathbb{C}^{n\times n}$, order its (real) eigenvalues as $\lambda_1(M) \geq \cdots \geq \lambda_n(M)$.

Then $\rho(M) = \sigma_1(M) = \max(|\lambda_1(M)|, |\lambda_n(M)|)$.

Lemma

(Weyl's inequality) *For Hermitian* $H,\ W$ *in* $\mathbb{C}^{n\times n}$,
for all $i \in [n]$, $|\lambda_i(H) - \lambda_i(\overline{H+W})| \leq \rho(W)$

Proof: by Courant-Fisher theorem,

$$\lambda_i(H+W) = \sup_{\dim(E)=i} \inf_{x\in E, \|x\|=1} x^T(H+W)x$$

Apply to $E = \mathrm{Vect}\{x_1(H), \ldots, x_i(H)\}$ to obtain

$$\begin{aligned} \lambda_i(H+W) &\geq \inf_{x\in E,\|x\|=1} x^T(H+W)x \\ &\geq \inf_{x\in E,\|x\|=1} x^T Hx + \inf_{x\in E,\|x\|=1} x^T Wx \\ &\geq \lambda_i(H) - \rho(W). \end{aligned}$$

Similarly, $\lambda_i(H) \geq \lambda_i(H+W) - \rho(-W)$ hence the result

Cauchy's interlacing theorem

Theorem

For Hermitian $A \in \mathbb{C}^{n\times n}$, $P \in \mathbb{C}^{n\times n}$, $m < n$ *such that* $P^*P = I_m$, *then* $B := P^*AP$ *verifies*

$\lambda_i(A) \geq \lambda_i(B) \geq \lambda_{n-m+i}(A)$

Cauchy's interlacing theorem

Theorem

For Hermitian $A \in \mathbb{C}^{n\times n}$, $P \in \mathbb{C}^{n\times n}$, $m < n$ *such that* $P^*P = I_m$, *then* $B := P^*AP$ *verifies*
$\lambda_i(A) \geq \lambda_i(B) \geq \lambda_{n-m+i}(A)$

Proof: Let $x_1(B), \ldots, x_n(B)$ eigenvector basis of B,
$V_i = \mathrm{Vect}(x_1(B), \ldots, x_i(B))$, $W_i = P(V_i)$.

Courant-Fisher:
$\lambda_i(B) = \inf_{u \in V_i, \|u\|=1} u^*P^*APu = \inf_{v \in W_i, \|v\|=1} v^*Av \leq \lambda_i(A)$.

Similarly:
$\lambda_{m-i+1}(-A) \geq \lambda_{m-i+1}(-B)$; equivalently:
$-\lambda_{n-m+i}(A) \geq -\lambda_i(B)$.

Perturbation of eigenvectors: the Davis-Kahane "sin Θ" theorem

Theorem

For symmetric, real $H, \hat{H}, W = \hat{H} - H \in \mathbb{R}^{n \times n}$ *any* $1 \leq r \leq s \leq n$, $d = s - r + 1$, *if*

$$\delta := \min(\lambda_{r-1}(H) - \lambda_r(H), \lambda_s(H) - \lambda_{s+1}(H)) > 0,$$

then for matrices $V = (v_r, \ldots, v_s)$, $\hat{V} = (\hat{v}_r, \hat{v}_s)$ *of orthonormal eigenvectors of* H *resp. of* $\hat{H}$, *there exists* $O \in \mathbb{R}^{d \times d}$ *such that* $O^\top O = I_d$ *and*

$$\|\hat{V}O - V\|_{op} \leq \|\hat{V}O - V\|_F \leq \frac{2^{3/2} \min(d^{1/2}\|W\|_{op}, \|W\|_F)}{\delta}.$$

Perturbation of eigenvectors: a more elementary result

Lemma

For fixed i let $\Delta := \inf_{j:\lambda_j(H)\neq\lambda_i(H)} |\lambda_j(H) - \lambda_i(H)|$. Assume $\rho(W) < \Delta$. Then for any normed eigenvector $\hat{x}_i$ of $\hat{H} = H + W$ associated with $\lambda_i(\hat{H})$ there exists x_i normed eigenvector of H associated with $\lambda_i(H)$ such that $\langle x_i, \hat{x}_i\rangle \geq \sqrt{1 - \left(\frac{\rho(W)}{\Delta-\rho(W)}\right)^2}$

Perturbation of eigenvectors: a more elementary result

Lemma

For fixed i let $\Delta := \inf_{j:\lambda_j(H)\neq\lambda_i(H)} |\lambda_j(H) - \lambda_i(H)|$. Assume $\rho(W) < \Delta$. Then for any normed eigenvector $\hat{x}_i$ of $\hat{H} = H + W$ associated with $\lambda_i(\hat{H})$ there exists x_i normed eigenvector of H associated with $\lambda_i(H)$ such that $\langle x_i, \hat{x}_i\rangle \geq \sqrt{1 - \left(\frac{\rho(W)}{\Delta - \rho(W)}\right)^2}$

Proof: Decompose $\hat{x}_i = \sum_j \theta_j x_j$ to get

$\hat{H}\hat{x}_i = \lambda_i(\hat{H})\hat{x}_i = \sum_j \theta_j \lambda_j(H) x_j + W\hat{x}_i$

hence $W\hat{x}_i = \sum_j (\lambda_i(\hat{H}) - \lambda_j(H))\theta_j x_j$

By Weyl's inequality, $|\lambda_i(\hat{H}) - \lambda_j(H)| \geq \Delta - \rho(W)$ if $\lambda_i(H) \neq \lambda_j(H)$. Thus

$\rho(W) \geq (\Delta - \rho(W))\sqrt{1 - \sum_{k:\lambda_k(H)=\lambda_i(H)} |\theta_k|^2}$

$\Rightarrow \sum_{k:\lambda_k(H)=\lambda_i(H)} |\theta_k|^2 \geq 1 - \left(\frac{\rho(W)}{\Delta - \rho(W)}\right)^2$

Perturbation arguments recap

For $\hat{H} = H + W$, eigenvalue $\lambda_i(\hat{H})$ and associated eigenvector $\hat{x}_i$ of $\hat{H}$, with $\Delta = \inf_{\lambda_j(H) \neq \lambda_i(H)} |\lambda_j(H) - \lambda_i(H)|$:

$|\lambda_i(H) - \lambda_i(\hat{H})| \leq \rho(W)$

$\exists x_i \leftrightarrow \lambda_i(H)$ such that $\langle x_i, \hat{x}_i \rangle = 1 - O\left(\frac{\rho(W)^2}{\Delta^2}\right)$.

Spectral properties of graphs

Non-oriented graph $G = (V, E)$, $V = [n]$, $E = \{\text{edges } (i, j)\}$

Adjacency matrix $A(G)$: $A_{ij} = \mathbb{I}_{(i,j)\in E} = \mathbb{I}_{i\sim j}$, $i, j \in V = [n]$
$\to$ Symmetric for non-oriented graph

Laplacian matrix $L(G)$: $L_{ij} = \begin{cases} -A_{ij} & \text{if } i \neq j, \\ d_i := \sum_{k\neq i} A_{ik} & \text{if } j = i. \end{cases}$
d_i: **degree** of node i.

$x^T L x = \sum_{i<j} A_{ij}(x_i - x_j)^2$ so that:
$0 \preceq L$, where $\preceq$: semi-definite order on symmetric matrices, and
$\lambda_n(L) = 0$ with associated eigenvector $\{1/\sqrt{n}\}_{i\in[n]}$

L: infinitesimal generator of continuous-time random walk on G,
with transition rates A_{ij} from i to $j \neq i$

Isoperimetric constant of graph G:
$I(G) := \min\{\frac{|E(S,\overline{S})|}{|S|}, S \subset V, 0 < |S| \leq \frac{n}{2}\}$
Let $\Delta(G) := \max_{i \in V=[n]} d_i(G)$, largest node degree.

Lemma
(Cheeger's inequality). $I(G) \leq \sqrt{2\Delta(G)\lambda_{n-1}(L(G))}$.

Easier result:

Lemma
$I(G) \geq \frac{\lambda_{n-1}(L(G))}{2}$

Isoperimetric constant of graph G:

$I(G) := \min\{\frac{|E(S,\overline{S})|}{|S|}, S \subset V, 0 < |S| \leq \frac{n}{2}\}$

Let $\Delta(G) := \max_{i \in V=[n]} d_i(G)$, largest node degree.

Lemma

(Cheeger's inequality). $I(G) \leq \sqrt{2\Delta(G)\lambda_{n-1}(L(G))}$.

Easier result:

Lemma

$I(G) \geq \frac{\lambda_{n-1}(L(G))}{2}$

Proof: Courant-Fisher: $\lambda_{n-1} = \inf\left\{\frac{x^\top Lx}{\|x\|^2}, x : \langle x, e\rangle = 0\right\}$

For x : $x_i = \mathbb{I}_{i \in S} - \frac{|S|}{n}$, yields

$\lambda_{n-1} \leq \frac{|E(S,\overline{S})|}{|S|(1-|S|/n)}$

Isoperimetric constant of graph G:

$I(G) := \min\{\frac{|E(S,\overline{S})|}{|S|}, S \subset V, 0 < |S| \leq \frac{n}{2}\}$

Let $\Delta(G) := \max_{i \in V=[n]} d_i(G)$, largest node degree.

Lemma

(Cheeger's inequality). $I(G) \leq \sqrt{2\Delta(G)\lambda_{n-1}(L(G))}$.

Easier result:

Lemma

$I(G) \geq \frac{\lambda_{n-1}(L(G))}{2}$

Proof: Courant-Fisher: $\lambda_{n-1} = \inf\left\{\frac{x^\top Lx}{\|x\|^2}, x : \langle x, e\rangle = 0\right\}$

For $x :\ x_i = \mathbb{I}_{i\in S} - \frac{|S|}{n}$, yields

$\lambda_{n-1} \leq \frac{|E(S,\overline{S})|}{|S|(1-|S|/n)}$

Corollary

Graph connected iff $\lambda_{n-1}(L) > 0$

Alon-Boppana theorem

Definition

d-regular graph G is **Ramanujan** if $\lambda_2(A) \leq 2\sqrt{d-1}$

Note: $L = dI_n - A$, hence $\lambda_{n-1}(L) = d - \lambda_2(A)$

Alon-Boppana theorem

Definition
d-regular graph G is **Ramanujan** if $\lambda_2(A) \leq 2\sqrt{d-1}$

Note: $L = dI_n - A$, hence $\lambda_{n-1}(L) = d - \lambda_2(A)$

Theorem
For d-regular G with diameter $\geq 2r+1$,
$\lambda_2(A) \geq 2\sqrt{d-1}\cos\left(\frac{\pi}{r+2}\right) = 2\sqrt{d-1}(1 - O(r^{-2}))$.

Hence Ramanujan graph G has maximal spectral gap $\lambda_{n-1}(L)$

Proof

Take $i, j \in V : d_G(i,j) \geq 2r+1$.
G': subgraph of G induced by nodes of $\mathcal{B}(i,r)$, $\mathcal{B}(j,r)$
$A'_G = P^* A_G P$, where $P \in \mathbb{R}^{n\times m}$ such that $P^*P = I_m$

Cauchy's interlacing theorem
$\Rightarrow \lambda_2(A_G) \geq \lambda_2(A_{G'}) = \min\{\lambda_1(A_{\mathcal{B}(i,r)}), \lambda_1(A_{\mathcal{B}(j,r)})\}$

Nb of closed walks of length $2q$ started at i in $\mathcal{B}(i,r)$: $w_{2q}(\mathcal{B}(i,r))$.

$w_{2q}(\mathcal{B}(i,r)) \geq w_{2q}(\mathcal{T}_{d-1,r}) = (d-1)^q w_{2q}(\mathcal{P}_r)$

$\lim_{q\to\infty} [w_{2q}(\mathcal{B}(i,r))]^{1/2q} = \lambda_1(A_{\mathcal{B}(i,r)}) \geq \cdots$
$\cdots \geq \sqrt{d-1}\lambda_1(\mathcal{P}_r) = \sqrt{d-1} \times 2\cos\frac{\pi}{r+2}$

Bounding spectral radius $\rho(W)$ of random matrices W: the Trace method

(method due to [Füredi-Komlos'81]; here a sub-optimal version avoiding sharp combinatorics)

Lemma

Let $W \in \mathbb{R}^{n\times n}$: symmetric matrix with entries independent up to symmetry, such that $|W_{ij}| \leq 1$, $\mathbb{E}(W_{ij}) = 0$, and $\mathbb{E}(W_{ij}^2) \leq O(d/n)$ for some $d \geq 1$. Then for any fixed $\epsilon > 0$, with high probability, $\rho(W) \leq O(\sqrt{d}n^{\epsilon})$.

Corollary

If we assume $d \geq n^{\delta}$ for some $\delta \in]0,1[$, then with high probability $\rho(W) = o(d)$.

Proof: Take $\epsilon < \delta/2$ to obtain $\rho(W) = O(n^{\delta/2+\epsilon}) = o(n^{\delta})$.

Proof

For fixed $k \in \mathbb{N}$, write $\rho^{2k} \leq \sum_{i \in [n]} \lambda_i(W)^{2k} = \mathrm{Trace}(W^{2k})$

Thus $\mathbb{P}(\rho \geq x) \leq x^{-2k} \mathbb{E}(\rho^{2k}) \leq x^{-2k} \mathbb{E}\mathrm{Trace}(W^{2k})$

Combinatorial expression of trace:

$$\mathrm{Trace}(W^{2k}) = \sum_{i_0^{2k} \in [n]^{2k+1} : i_0 = i_{2k}} \prod_{j=1}^{2k} W_{i_{j-1} i_j}$$

Recall W_{ij}: centered and independent

$\rightarrow$ Only paths contributing non-zero expectation: traverse each edge at least twice

$$\Rightarrow \mathbb{E}\mathrm{Trace}(W^{2k}) \leq \sum_{e=1}^{k} \sum_{v=1}^{e+1} C(e,v) n^v O((d/n)^e)$$

Yields $\mathbb{E}\mathrm{Trace}(W^{2k}) = O(nd^k)$.

For $x = \sqrt{d} n^{\epsilon}$, yields $\mathbb{P}(\rho \geq x) \leq O(n^{1-2k\epsilon})$

Result follows by taking $k > 1/(2\epsilon)$

Method 2: Bernstein's inequality for matrices

Theorem

(Tropp'15) Let $X_1, \dots, X_m$ *be independent Hermitian random matrices such that:*
$\mathbb{E}(X_k) = 0, \quad \|X_k\|_{op} \le L$ *almost surely*, $k \in [m]$.

Let $Y = \sum_{k\in[m]} X_k$, *and* $v(Y) := \|\mathbb{E}(Y^2)\|_{op} = \|\sum_{k\in[m]} \mathbb{E}X_k^2\|_{op}$.

Then for all $t > 0$, $\mathbb{P}(\lambda_1(Y) \ge t) \le n \exp\left(\frac{-t^2}{2(v(Y)+Lt/3)}\right)$.

This implies for all $t > 0$: $\mathbb{P}(\|Y\|_{op} \ge t) \le 2n \exp\left(\frac{-t^2}{2(v(Y)+Lt/3)}\right)$.

Corollary

For W *s.t.* W_{ij} *independent up to symmetry*, $\mathbb{E}W_{ij} = 0$, $|W_{ij}| \le 1$,
$\mathbb{E}W_{ij}^2 = O(d/n)$:
If $d \gg \log(n)$ *then with high probability* $\rho(W) = o(d)$.

Stronger bounds on $\rho = \rho(W)$

Theorem (Feige and Ofek, 2005)

Let $A \in \mathbb{R}^{n\times n}$*: symmetric matrix with entries independent up to symmetry,* $A_{ij} \in [0,1]$*, and such that* $\mathbb{E}(A_{ij}) \leq d/n$*, where* $d \leq n^{1/5}$*.*
Then for some (universal) constant $\kappa > 0$*,*
with high probability $\rho(A - \mathbb{E}(A)) \leq \kappa\sqrt{\max(d, \log(n))}$*.*

Corollary

For $d \gg \sqrt{\max(d, \log(n))}$*, i.e.* $d \gg \sqrt{\log(n)}$*, with high probability* $\rho(A - \mathbb{E}A) = o(d)$

Proof of Bernstein matrix inequality

Lemma

For independent Hermitian matrices X_k, $k \in [m]$, *and* $Y = \sum_{k\in[m]} X_k$*:*

$$\mathbb{E}\mathrm{Tr} e^{\theta Y} \leq \mathrm{Tr}\exp\left(\sum_{k\in[m]} \ln \mathbb{E} e^{\theta X_k}\right)$$

Lemma

For Hermitian X *such that* $\mathbb{E}(X) = 0$ *and* $\|X\| \leq L$ *almost surely, then:*

$$\forall\theta \in (0, 3/L), \begin{cases} \mathbb{E} e^{\theta X} \preceq \exp\left(\frac{\theta^2/2}{1-\theta L/3}\mathbb{E}X^2\right), \\ \ln \mathbb{E} e^{\theta X} \preceq \frac{\theta^2/2}{1-\theta L/3}\mathbb{E}X^2, \end{cases}$$

Lemma

For Hermitian A, B, *if* $A \preceq B$, *then* $\forall i \in [n], \lambda_i(A) \leq \lambda_i(B)$. *Hence for all non-decreasing* $f : \mathbb{R} \to \mathbb{R}$, $\mathrm{Tr}\, f(A) \leq \mathrm{Tr}\, f(B)$.

Application: Community Detection in the Stochastic Block Model

$\mathcal{G}(n, \{\alpha_i\}_{i\in[K]}, P)$, where $\alpha_i > 0, \sum_{i\in[K]} \alpha_i = 1$, $P \in [0,1]^{K\times K}$: multi-type version of the Erdős-Rényi random graph

- n vertices partitioned into K communities
- Type (community) of node $i : \sigma_i \in [K]$, $\sigma_i :$ i.i.d., $\sim \alpha$
- Conditionally on $\sigma_{[n]}$, independently for each pair $i, j \in [n]$: edge (i, j) present with probability $P_{\sigma(i),\sigma(j)}$.

Strong signal regime:: fixed $K, \alpha, B \in \mathbb{R}_+^{K\times K}$; $P = (d/n)B$, with $\lim_{n\to+\infty} d = +\infty$

Spectral embedding

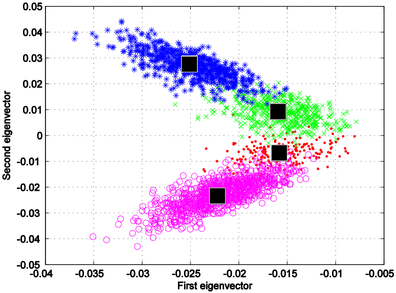
- Extract top two (more generally top R) eigenvalues λ_1, λ_2 of graph's adjacency matrix $A \in \mathbb{R}^{n \times n}$ (ordered by absolute value: $|\lambda_1| \geq |\lambda_2| \geq \cdots$)
- Let $x_1, x_2 \in \mathbb{R}^n$: corresponding normalized eigenvectors
- Embed vertex $k \in [n]$ into $\mathbb{R}^2$ by letting $z_k := \sqrt{n}(x_1(k), x_2(k))$

Spectral embedding

- Extract top two (more generally top R) eigenvalues λ_1, λ_2 of graph's adjacency matrix $A \in \mathbb{R}^{n \times n}$ (ordered by absolute value: $|\lambda_1| \geq |\lambda_2| \geq \cdots$)
- Let $x_1, x_2 \in \mathbb{R}^n$: corresponding normalized eigenvectors
- Embed vertex $k \in [n]$ into $\mathbb{R}^2$ by letting $z_k := \sqrt{n}(x_1(k), x_2(k))$

$\rightarrow$ based on PCA dimensionality reduction of A to dimension R

Example: spectral embedding for SBM



A case with $K = 4$ communities

Spectral embedding seems to reflect community structure

$\rightarrow$ Why / when do spectral methods work?

Theorem

Assume communities are **distinguishable**, *i.e. for each* $k \neq \ell \in [K]$, *there exists* $m \in [K]$ *such that* $B_{km} \neq B_{\ell m}$.
Assume $\sqrt{\ln(n)} \ll d \ll n^{\delta}$ *for some fixed* $\delta \in]0,1[$. *Let* R: *rank of matrix* B. *Then with high probability:*
(i) the spectrum of A *consists of* R *eigenvalues of order* $\Theta(d)$ *and* $n-R$ *eigenvalues of order* $o(d)$.
(ii) R*-dimensional spectral embedding reveals underlying communities: except for vanishing fraction of nodes* $i \in [n]$,

$$||z_i - z_j|| = \begin{cases} o(1) & \text{if } \sigma(i) = \sigma(j), \\ \Omega(1) & \text{if } \sigma(i) \neq \sigma(j) \end{cases}$$

Theorem

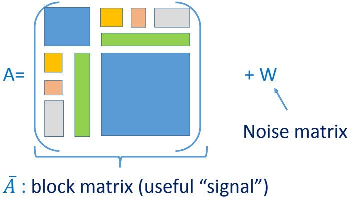
Assume communities are **distinguishable**, *i.e. for each* $k \neq \ell \in [K]$, *there exists* $m \in [K]$ *such that* $B_{km} \neq B_{\ell m}$.
Assume $\sqrt{\ln(n)} \ll d \ll n^{\delta}$ *for some fixed* $\delta \in]0,1[$. *Let* R*: rank of matrix* B. *Then with high probability:*
(i) the spectrum of A *consists of* R *eigenvalues of order* $\Theta(d)$ *and* $n - R$ *eigenvalues of order* $o\,(d)$.
(ii) R*-dimensional spectral embedding reveals underlying communities: except for vanishing fraction of nodes* $i \in [n]$,

$$||z_i - z_j|| = \begin{cases} o(1) & \text{if } \sigma(i) = \sigma(j), \\ \Omega(1) & \text{if } \sigma(i) \neq \sigma(j) \end{cases}$$

Corollary

Under these conditions any sensible clustering scheme (eg K-means properly initialized) correctly classifies all but vanishing fraction of nodes.

Proof strategy



- Write adjacency matrix as $A = \overline{A} + W$ with $\overline{A}_{ij} = \frac{d}{n} B_{\sigma(i),\sigma(j)}$
- R leading eigen-elements of $\overline{A}$ capture community structure
- Control perturbation of eigen-elements of a symmetric matrix $\overline{A}$ by addition of symmetric matrix W in terms of **spectral radius** $\rho(W)$ of noise matrix
- Prove bound on $\rho(W)$ for random noise matrix W

Eigenstructure of $\overline{A}$

Block structure of $\overline{A} \Rightarrow \overline{A}x$ constant on each block $\Rightarrow$ eigenvectors associated to non-zero eigenvalue are block-constant.

For $t \in \mathbb{R}^K$ define $x := \phi(t) = (t_{\sigma(i)})_{i\in[n]} \in \mathbb{R}^n$.

Then $\overline{A}\phi(t) = d\phi(Mt)$, where $M_{uv} := B_{uv}\alpha_v$.

Lemma

Spectrum of $\overline{A}$:
R eigen-pairs $(\lambda_u = d\mu_u, \overline{x}_u = \phi(t_u))$ where (μ_u, t_u): eigen-pairs of M with $\mu_u \neq 0$;
0: eigenvalue with multiplicity $n - R$

Eigenstructure of $\overline{A}$ (continued)

Lemma

Under distinguishability hypothesis there exists $\epsilon > 0$ *function of* B, α *such that for any choice of normalized leading eigenvectors* $\overline{x}_1, \ldots, \overline{x}_R$, $\overline{z}_i = \sqrt{n}(\overline{x}_1(i), \ldots, \overline{x}_R(i))^T$ *verify*

$$\sigma(i) \neq \sigma(j) \Rightarrow ||\overline{z}_i - \overline{z}_j|| \geq \epsilon > 0$$

Proof: Let $t_u \in \mathbb{R}^K$ be such that $\sqrt{n}\overline{x}_u = \phi(t_u)$, and $\sqrt{\alpha} = \text{Diag}(\sqrt{\alpha_u})$.

Then: $\{\sqrt{\alpha} t_u\}_{u \in [R]}$: orthonormal family by orthonormality of the $\overline{x}_u$.

t_u eigenvectors of matrix $M = B\alpha$, hence $\sqrt{\alpha} t_u$: orthonormal family of eigenvectors of matrix $\sqrt{\alpha} B \sqrt{\alpha}$.

Thus $\sqrt{\alpha} B \sqrt{\alpha} = \sum_{u \in [R]} \mu_u (\sqrt{\alpha} t_u)(\sqrt{\alpha} t_u)^T$.

Equivalently: $B = \sum_{u \in [R]} \mu_u t_u t_u^T$.

Hence minimum of $||\overline{z}_i - \overline{z}_j||$ over $\sigma(i) \neq \sigma(j)$ strictly positive, for otherwise B has two identical rows, i.e. distinguishability fails.

Proof

- Matrix $\overline{A}$ of rank R, spectral gaps $|\overline{\lambda}_i - \overline{\lambda}_j| = \Omega(d)$, R-dimensional spectral embedding with $\overline{x}_1, \ldots, \overline{x}_R$ separates clusters $V_k = \{i \in [n] : \sigma_i = k\}$
- Assuming $\rho = \rho(A - \overline{A}) \ll d$, Weyl's inequality: R eigenvalues λ_i close to $\overline{\lambda}_i = \Omega(d)$, others of order $\rho \ll d$
- Associated eigenvectors x_i such that $\langle x_i, \overline{x}_i \rangle = 1 - O((\rho/d)^2)$

Then $\sum_{i\in[n]} \|z_i - \overline{z}_i\|^2 = n \sum_{u\in[R]} \|x_u - \overline{x}_u\|^2 = n\theta$ with $\theta = O((\rho/d)^2) = o(1)$
Hence (Tchebitchev inequality):
$|\{i : \|z_i - \overline{z}_i\| \geq \theta^{1/3}\}| \leq n\theta^{1/3} = o(n)$
Yields desired conclusion: except for vanishing fraction $\theta^{1/3}$ of nodes, spectral representatives z_i $\theta^{1/3}$-close of corresponding $\overline{z}_i$, themselves clustered according to community structure
Feige-Ofek: $d \gg \sqrt{\log n} \Rightarrow \rho(A - \overline{A}) \ll d$