PNAS

# On a two-truths phenomenon in spectral graph clustering

**Carey E. Priebe[a,b,c,1], Youngser Park[b], Joshua T. Vogelstein[b,d], John M. Conroy[e], Vince Lyzinski[c,f], Minh Tang[a], Avanti Athreya[a], Joshua Cape[a], and Eric Bridgeford[b,g]**

[a]Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218; [b]Center for Imaging Science, Johns Hopkins University, Baltimore, MD 21218; [c]Human Language Technology Center of Excellence, Johns Hopkins University, Baltimore, MD 21218; [d]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218; [e]Institute for Defense Analyses, Center for Computing Sciences, Bowie, MD 20715; [f]Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003; and [g]Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21218

Edited by Peter J. Bickel, University of California, Berkeley, CA, and approved February 8, 2019 (received for review August 21, 2018)

Carey E. Priebe

Johns Hopkins University

March 25, 2019

Heilbronn Institute for Mathematical Research

Home About Research Fellows Diversity News Events Opportunities Contact us

FOCUSED RESEARCH WORKSHOP AND CYBER-SECURITY DAY

University of Bristol

## Yogi Berra (purportedly):

*"In theory there is no difference between theory and practice. In practice, there is."*

(cf. *"That's all well and good in practice, but how does it work in theory?"*)

# On a two-truths phenomenon in spectral graph clustering

Carey E. Priebe[a,b,c,1], Youngser Park[b], Joshua T. Vogelstein[b,d], John M. Conroy[e], Vince Lyzinski[c,f], Minh Tang[a], Avanti Athreya[a], Joshua Cape[a], and Eric Bridgeford[b,g]

[a]Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218; [b]Center for Imaging Science, Johns Hopkins University, Baltimore, MD 21218; [c]Human Language Technology Center of Excellence, Johns Hopkins University, Baltimore, MD 21218; [d]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218; [e]Institute for Defense Analyses, Center for Computing Sciences, Bowie, MD 20715; [f]Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003; and [g]Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21218

Edited by Peter J. Bickel, University of California, Berkeley, CA, and approved February 8, 2019 (received for review August 21, 2018)

**Clustering is concerned with coherently grouping observations without any explicit concept of true groupings. Spectral graph clustering—clustering the vertices of a graph based on their spectral embedding—is commonly approached via *K*-means (or, more generally, Gaussian mixture model) clustering composed with either Laplacian spectral embedding (LSE) or adjacency spectral embedding (ASE). Recent theoretical results provide deeper understanding of the problem and solutions and lead us to a "two-truths" LSE vs. ASE spectral graph clustering phenomenon convincingly illustrated here via a diffusion MRI connectome dataset: The different embedding methods yield different clustering results, with LSE capturing left hemisphere/right hemisphere affinity structure and ASE capturing gray matter/white matter core–periphery structure.**

spectral embedding | spectral clustering | graph | network | connectome

The purpose of this paper is to cogently present a "two-truths" phenomenon in spectral graph clustering, to understand this phenomenon from a theoretical and methodological perspective, and to demonstrate the phenomenon in a real-data case consisting of multiple graphs each with multiple categorical vertex class labels.

A graph or network consists of a collection of vertices or nodes $V$ representing $n$ entities together with edges or links $E$ representing the observed subset of the $\binom{n}{2}$ possible pairwise relationships between these entities. Graph clustering, often associated with the concept of "community detection," is concerned with partitioning the vertices into coherent groups or clusters. By its very nature, such a partitioning must be based on connectivity patterns.

It is often the case that practitioners cluster the vertices of a graph—say, via $K$-means clustering composed with Laplacian spectral embedding—and pronounce the method as having performed either well or poorly based on whether the resulting clusters correspond well or poorly with some known or preconceived notion of "correct" clustering. Indeed, such a procedure may be used to compare two clustering methods and to pronounce that one works better (on the particular data under consideration). However, clustering is inherently ill-defined, as there may be multiple meaningful groupings, and two clustering methods that perform differently with respect to one notion of truth may in fact be identifying inherently different, but perhaps complementary, underlying structure. With respect to graph clustering, ref. 1 shows that there can be no algorithm that is optimal for all possible community detection tasks (Fig. 1).

We compare and contrast Laplacian and adjacency spectral embedding as the first step in spectral graph clustering and demonstrate that the two methods, and the two resulting clusterings, identify different—but both meaningful—graph structure. We trust that this simple, clear explication will contribute to an awareness that connectivity-based structure discovery via spectral graph clustering should consider both Laplacian and adjacency spectral embedding and the development of new methodologies based on this awareness.

## Spectral Graph Clustering

Given a simple graph $G=(V,E)$ on $n$ vertices, consider the associated $n\times n$ adjacency matrix $A$ in which $A_{ij}=0$ or 1 encoding whether vertices $i$ and $j$ in $V$ share an edge $(i,j)$ in $E$. For our simple undirected, unweighted, loopless case, $A$ is binary with $A_{ij}\in\{0,1\}$, symmetric with $A=A^\top$, and hollow with $diag(A)=\vec{0}$.

The first step of spectral graph clustering (2, 3) involves embedding the graph into Euclidean space via an eigendecomposition. We consider two options: Laplacian spectral embedding (LSE), wherein we decompose the normalized Laplacian of the adjacency matrix, and adjacency spectral embedding (ASE) given by a decomposition of the adjacency matrix itself. With target dimension $d$, either spectral embedding method produces $n$ points in $\Re^d$, denoted by the $n\times d$ matrix $X$. ASE employs the eigendecomposition to represent the adjacency matrix via $A=USU^\top$ and chooses the top $d$ eigenvalues by magnitude and their associated vectors to embed the graph via the scaled eigenvectors $U_d|S_d|^{1/2}$. Similarly, LSE embeds the graph via the top scaled eigenvectors of the normalized Laplacian $\mathcal{L}(A)=D^{-1/2}AD^{-1/2}$, where $D$ is the diagonal matrix of vertex degrees.

### Significance

**Spectral graph clustering—clustering the vertices of a graph based on their spectral embedding—is of significant current interest, finding applications throughout the sciences. But as with clustering in general, what a particular methodology identifies as "clusters" is defined (explicitly, or, more often, implicitly) by the clustering algorithm itself. We provide a clear and concise demonstration of a "two-truths" phenomenon for spectral graph clustering in which the first step—spectral embedding—is either Laplacian spectral embedding, wherein one decomposes the normalized Laplacian of the adjacency matrix, or adjacency spectral embedding given by a decomposition of the adjacency matrix itself. The two resulting clustering methods identify fundamentally different (true and meaningful) structure.**

Author contributions: C.E.P., J.T.V., J.M.C., and V.L. designed research; C.E.P., V.L., M.T., A.A., and J.C. performed research; C.E.P., Y.P., and E.B. analyzed data; and C.E.P. wrote the paper.

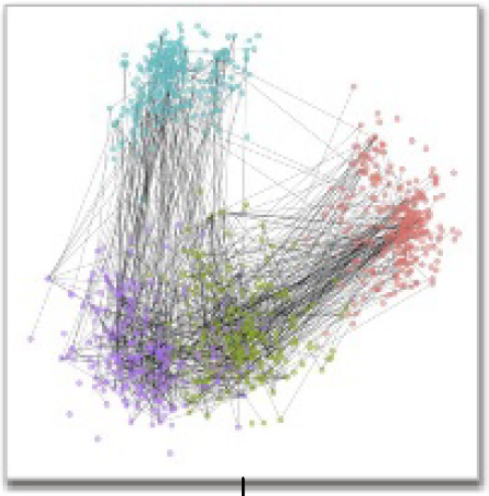
The authors declare no conflict of interest.

This article is a PNAS Direct Submission.



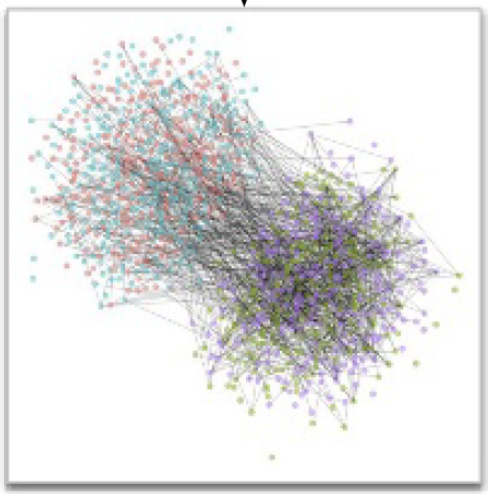
[1]To whom correspondence should be addressed. Email: cep@jhu.edu.

**L**aplacian
**S**pectral
**E**mbedding
(affinity)



**A**djacency
**S**pectral
**E**mbedding
(core-periphery)

**“Two Truths”**



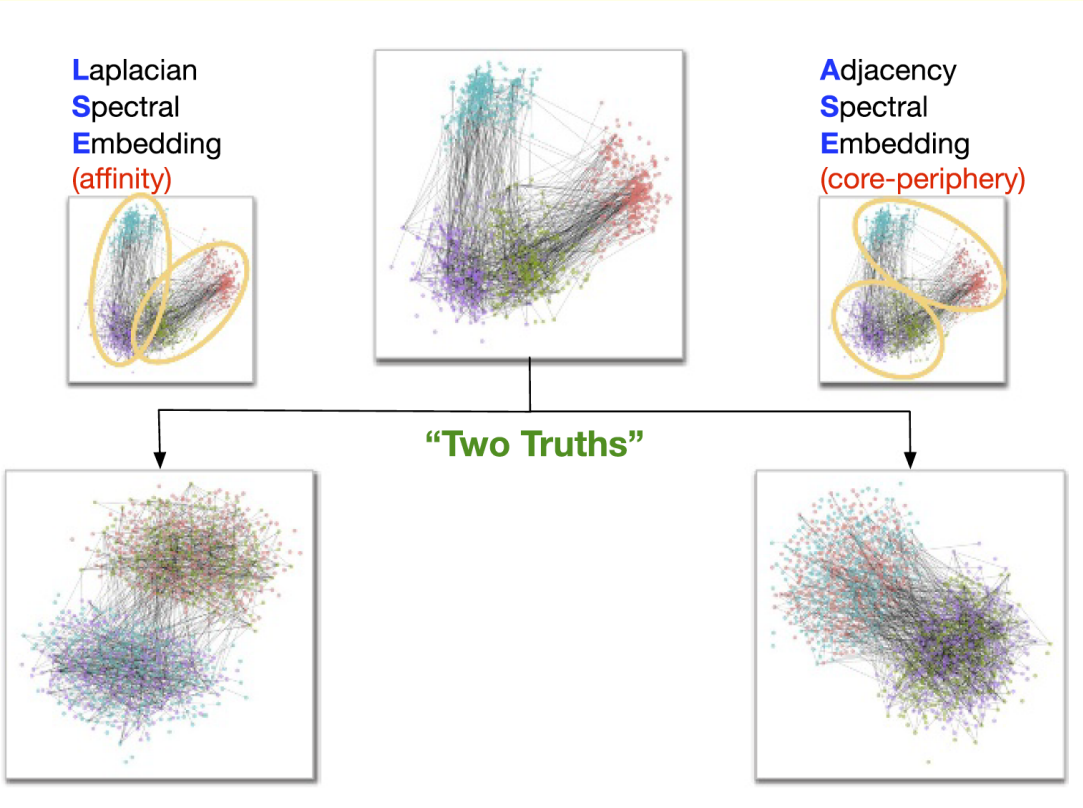



**Fig. 1.** A two-truths graph (connectome) depicting connectivity structure such that one grouping of the vertices yields affinity structure (e.g., left hemisphere/right hemisphere) and the other grouping yields core–periphery structure (e.g., gray matter/white matter). (*Top Center*) The graph with four vertex colors. (*Top Left* and *Top Right*) LSE groups one way and ASE groups another way. (*Bottom Left*) The LSE truth is two densely connected groups, with sparse interconnectivity between them (affinity structure). (*Bottom Right*) The ASE truth is one densely connected group, with sparse interconnectivity between it and the other group and sparse interconnectivity within the other group (core–periphery structure). This paper demonstrates the two-truths phenomenon illustrated here—that LSE and ASE find fundamentally different but equally meaningful network structure—via theory, simulation, and real data analysis.

We make significant conceptual use of the positive definite two-block SBM ($K = 2$), with

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

which henceforth we abbreviate as $B = [a, b; b, c]$. In this simple setting, two general/generic cases present themselves: affinity and core–periphery.

**Affinity: $a, c \gg b$.** An SBM with $B = [a, b; b, c]$ is said to exhibit affinity structure if each of the two blocks has a relatively high within-block connectivity probability compared with the between-block connectivity probability.

**Core-periphery: $a \gg b, c$.** An SBM with $B = [a, b; b, c]$ is said to exhibit core–periphery structure if one of the two blocks has a relatively high within-block connectivity probability compared with both the other block's within-block connectivity probability and the between-block connectivity probability.

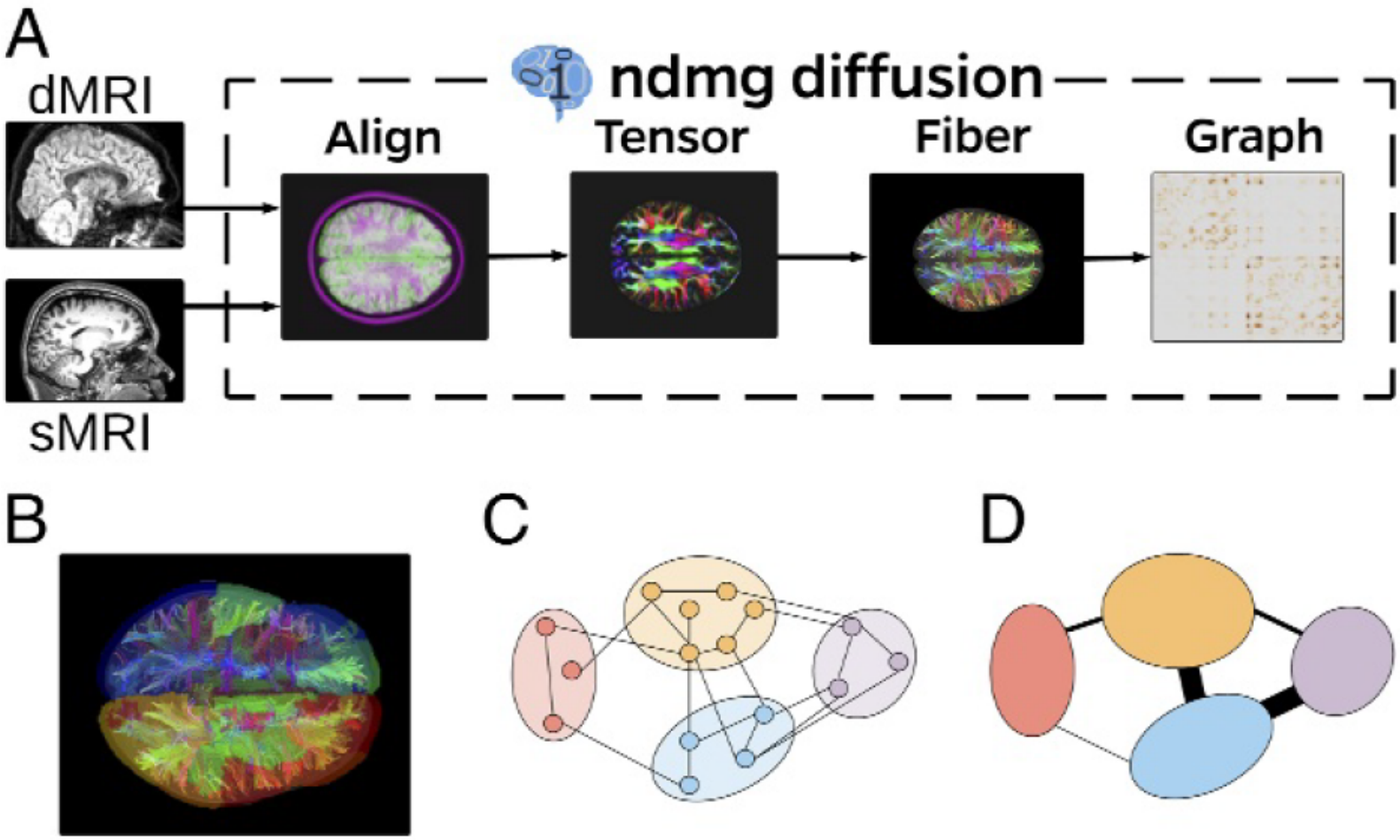



**Fig. 2.** Connectome data generation. (*A*) The pipeline. (*B*) Voxels and regions in tractography map. (*C*) Voxels and edges. (*D*) Contraction yields vertices and edges. The output is diffusion MRI graphs on ≈1 million vertices. Spatial vertex contraction yields graphs on ≈70,000 vertices from which we extract largest connected components of ≈40,000 vertices with {Left,Right} and {Gray,White} labels for each vertex. Fig. 1 depicts (a subsample from) one such graph.

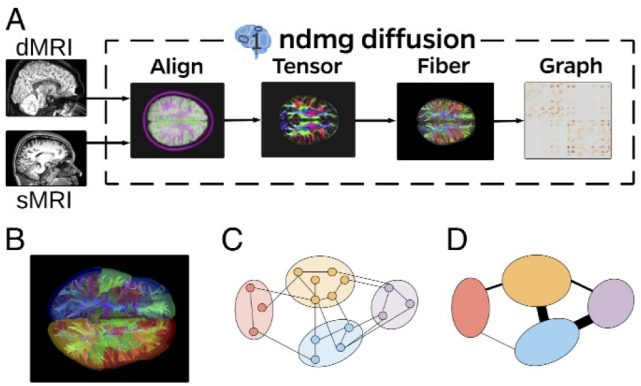



**Fig. 2.** Connectome data generation. (*A*) The pipeline. (*B*) Voxels and regions in tractography map. (*C*) Voxels and edges. (*D*) Contraction yields vertices and edges. The output is diffusion MRI graphs on ≈1 million vertices. Spatial vertex contraction yields graphs on ≈70,000 vertices from which we extract largest connected components of ≈40,000 vertices with {Left,Right} and {Gray,White} labels for each vertex. Fig. 1 depicts (a subsample from) one such graph.

## Connectome Data

We consider for illustration a diffusion MRI dataset consisting of 114 connectomes (57 subjects, two scans each) with 72,783 vertices each and both left/right/other hemispheric and gray/white/other tissue attributes for each vertex. Graphs were estimated using the NeuroData's MR Graphs pipeline (23), with vertices representing subregions defined via spatial proximity and edges defined by tensor-based fiber streamlines connecting these regions (Fig. 2).

The actual graphs we consider are the largest connected component (LCC) of the induced subgraph on the vertices labeled as both left or right and gray or white. This yields $m = 114$ connected graphs on $n \approx 40,000$ vertices. Additionally, for each graph every vertex has a {Left,Right} label and a {Gray,White} label, which we sometimes find convenient to consider as a single label in {LG,LW,RG,RW}.

**Sparsity.** The only notions of sparsity relevant here are linear algebraic: whether there are enough edges in the graph to support spectral embedding and whether there are few enough to allow for sparse matrix computations. We have a collection of observed connectomes and we want to cluster the vertices in these graphs, as opposed to in an unobserved sequence with the number of vertices tending to infinity. Our connectomes have, on average, $n \approx 40,000$ vertices and $e \approx 2,000,000$ edges, for an average degree $2e/n \approx 100$ and a graph density $e/\binom{n}{2} \approx 0.0025$.

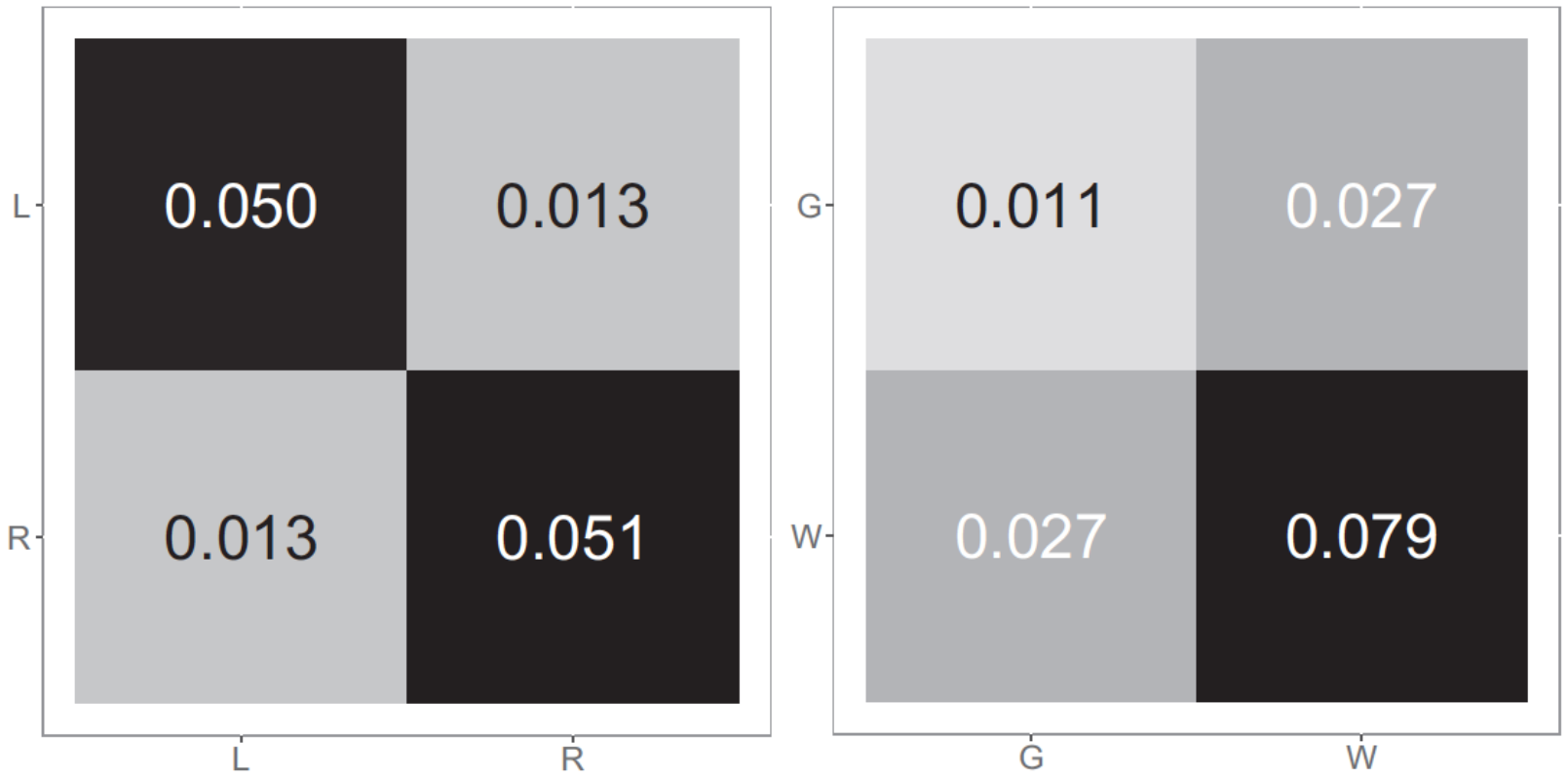
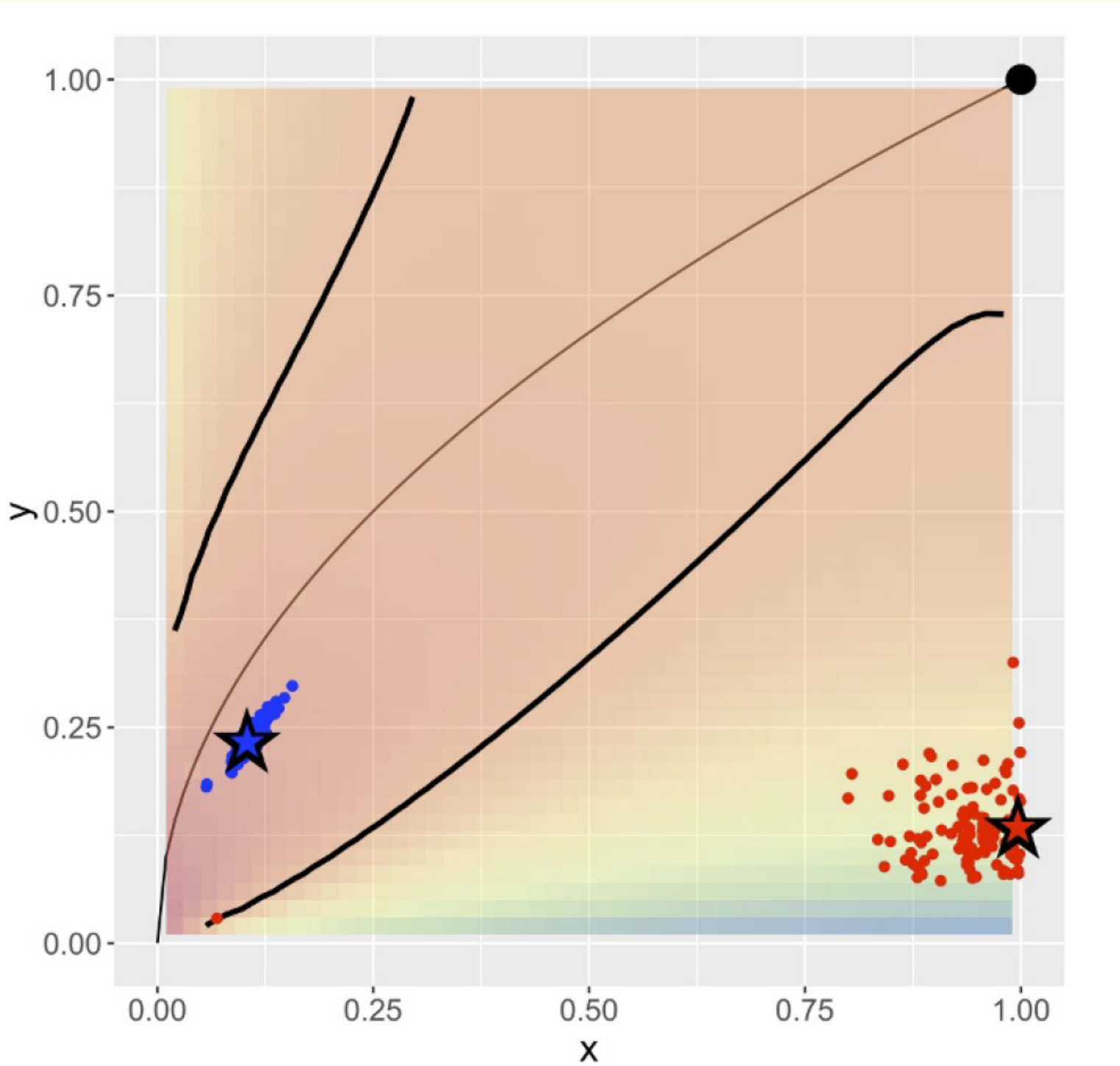



**Fig. 4.** Block connectivity probability matrices for the a priori projection of the composite connectome onto the two-block SBM for (*Left*) {Left, Right} and (*Right*) {Gray, White}. {Left, Right} exhibits affinity structure, with Chernoff ratio $<1$; {Gray, White} exhibits core–periphery structure, with Chernoff ratio $>1$.


1.00
0.75
0.50
0.25
0.00
y
0.00
0.25
0.50
0.75
1.00
x

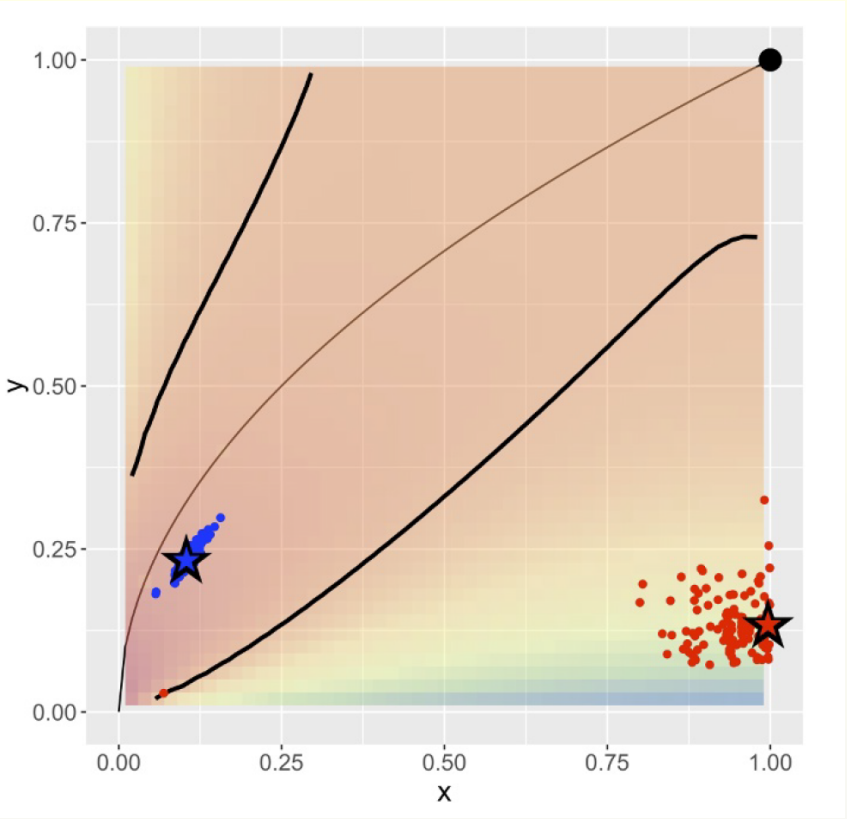
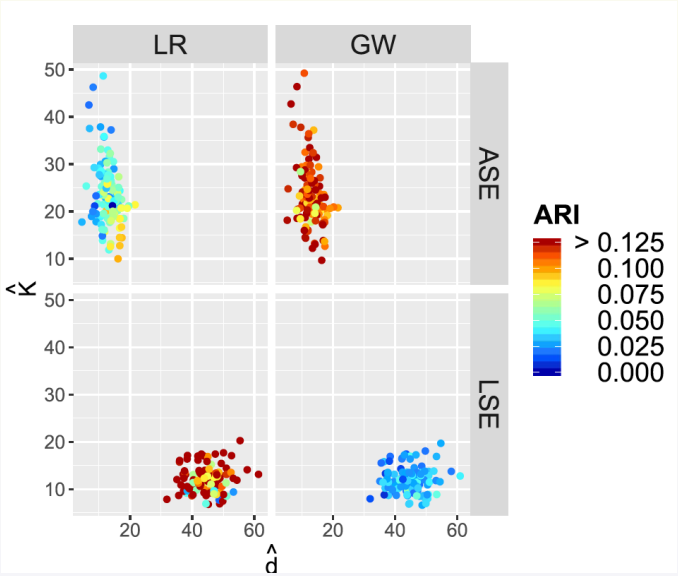
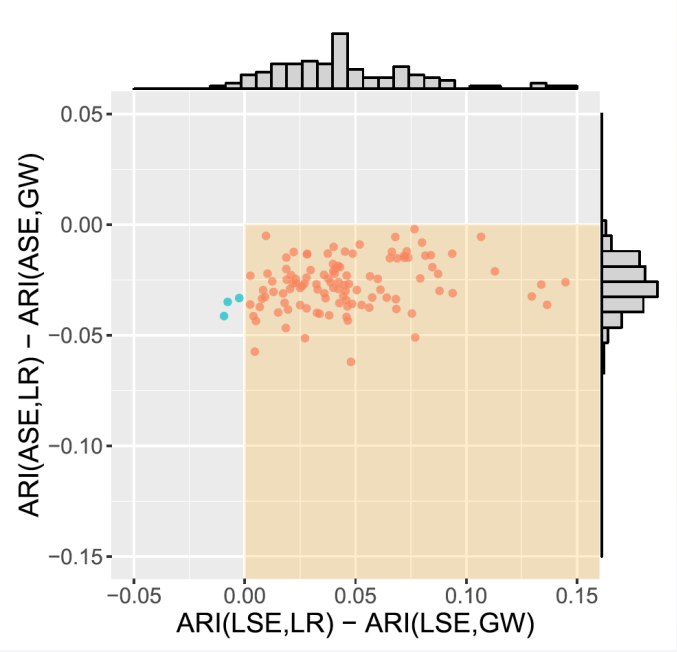



**Fig. 5.** For each of our 114 connectomes, we plot the a priori two-block SBM projections for {Left, Right} in red and {Gray, White} in blue. The coordinates are given by $x = \min(a, c)/\max(a, c)$ and $y = b/\max(a, c)$, where $B = [a, b; b, c]$ is the observed block connectivity probability matrix. The thin black curve $y = \sqrt{x}$ represents the rank 1 submodel separating positive definite (lower right) from indefinite (upper left). The background color shading is Chernoff ratio $\rho$, and the thick black curves are $\rho = 1$ separating the region where ASE is preferred (between the curves) from where LSE is preferred. The point (1, 1) represents Erdős–Rényi ($a = b = c$). The large stars are from the a priori composite connectome projections (Fig. 4). We see that the red {Left, Right} projections are in the affinity region where $\rho < 1$ and LSE is preferred while the blue {Gray, White} projections are in the core–periphery region where $\rho > 1$ and ASE is preferred. This analytical finding based on projections onto the SBM carries over to empirical spectral clustering results on the individual connectomes (Fig. 7).

# Two Truths:
# ASE $\Longrightarrow$ Gray/White ; LSE $\Longrightarrow$ Left/Right



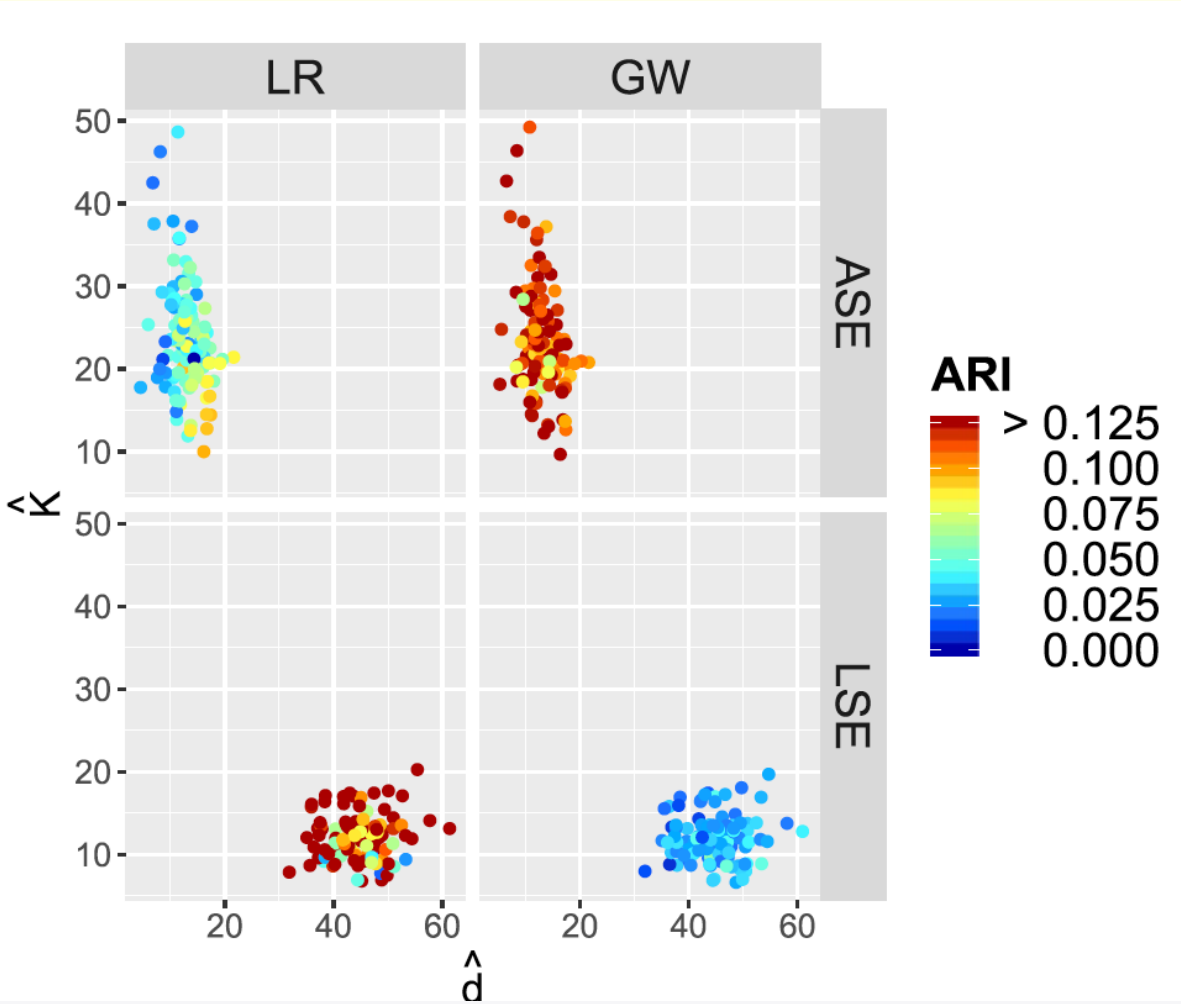



**Fig. 6.** Results of the $(\widehat{d}, \widehat{K})$ model selection for spectral graph clustering for each of our 114 connectomes. For LSE we see $\widehat{d} \in \{30, \ldots, 60\}$ and $\widehat{K} \in \{2, \ldots, 20\}$; for ASE we see $\widehat{d} \in \{2, \ldots, 20\}$ and $\widehat{K} \in \{10, \ldots, 50\}$. The color coding represents clustering performance in terms of ARI for each of LSE and ASE against each of the two truths {Left, Right} and {Gray, White} and shows that LSE clustering identifies {Left, Right} better than {Gray, White} and ASE identifies {Gray, White} better than {Left, Right}. Our two-truths phenomenon is conclusively demonstrated: LSE finds {Left, Right} (affinity) while ASE finds {Gray, White} (core–periphery).

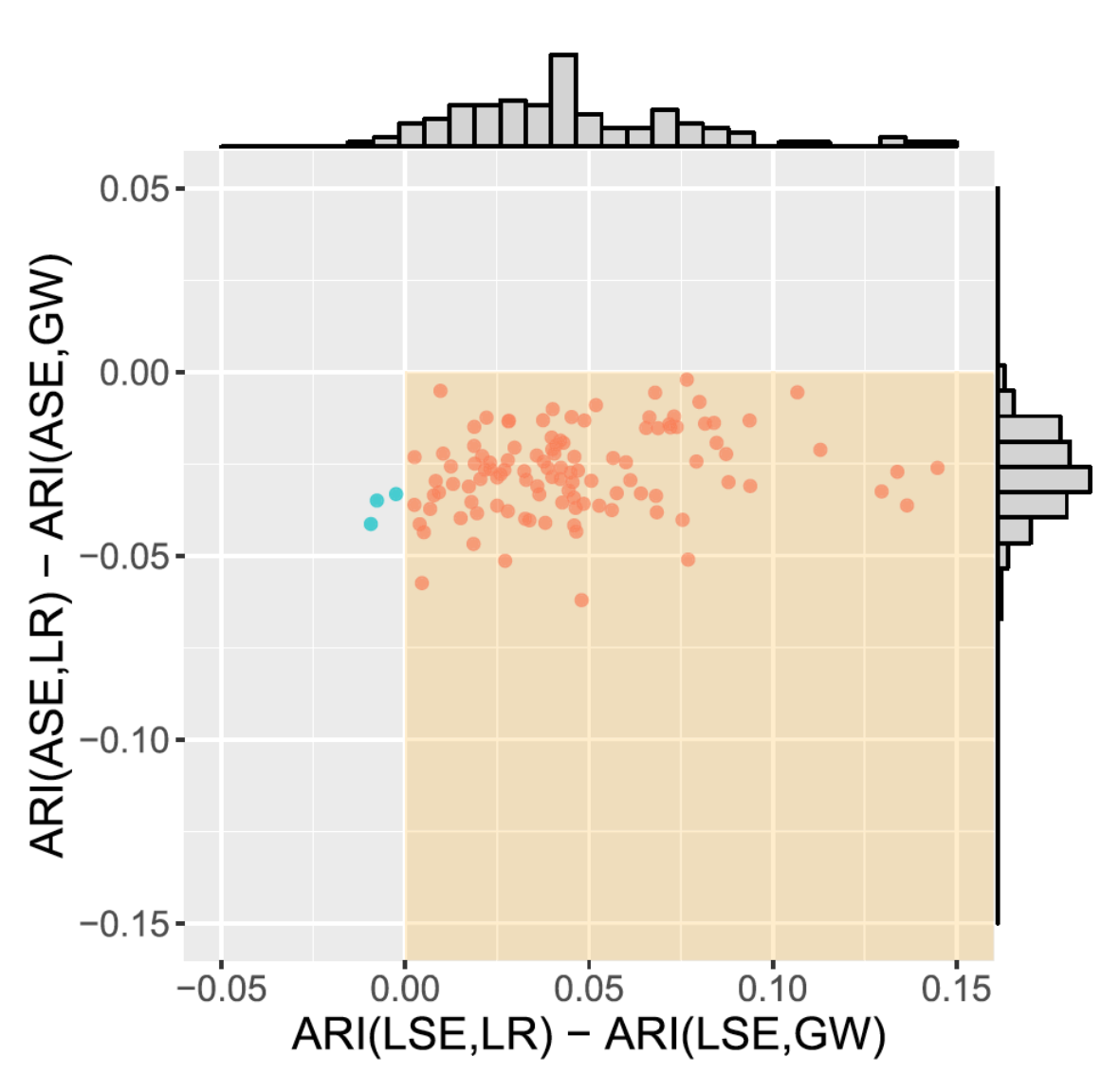



**Fig. 7.** Spectral graph clustering assessment via ARI. For each of our 114 connectomes, we plot the difference in ARI for the {Left, Right} truth against the difference in ARI for the {Gray, White} truth for the clusterings produced by each of LSE and ASE: $x$ = ARI(LSE,LR) − ARI(LSE,GW) vs. $y$ = ARI(ASE,LR) − ARI(ASE,GW). A point in the $(+, -)$ quadrant indicates that for that connectome the LSE clustering identified {Left, Right} better than {Gray, White} and ASE identified {Gray, White} better than {Left, Right}. Marginal histograms are provided. Our two-truths phenomenon is conclusively demonstrated: LSE identifies {Left, Right} (affinity) while ASE identifies {Gray, White} (core–periphery).

# On Spectral Graph Clustering

Minh Tang

**Limit theorems for eigenvectors of the normalized Laplacian for random graphs**

Minh Tang, Carey E. Priebe

*(Submitted on 28 Jul 2016)*

We prove a central limit theorem for the components of the eigenvectors corresponding to the $d$ largest eigenvalues of the normalized Laplacian matrix of a finite dimensional random dot product graph. As a corollary, we show that for stochastic blockmodel graphs, the rows of the spectral embedding of the normalized Laplacian converge to multivariate normals and furthermore the mean and the covariance matrix of each row are functions of the associated vertex's block membership. Together with prior results for the eigenvectors of the adjacency matrix, we then compare, via the Chernoff information between multivariate normal distributions, how the choice of embedding method impacts subsequent inference. We demonstrate that neither embedding method dominates with respect to the inference task of recovering the latent block assignments.

http://arxiv.org/abs/1607.08601

*Annals of Statistics*, 2018

# Spectral Clustering

*Spectral Clustering*
refers to a class of graph inference methodologies
in which the vertices of a graph $G$ are partitioned via

- some clustering algorithm

composed with

- some spectral embedding of $G$.

spectral embedding:

- Laplacian Spectral Embedding (LSE)
- Adjacency Spectral Embedding (ASE)

clustering:

- K-means
- Gaussian Mixture Modeling (GMM)

# Bickel & Sarkar, *AoS*, 2015

It was shown in B&S that for two-block stochastic blockmodels, for a large regime of parameters the normalized $\mathrm{LSE}$ reduces the within-block variance while preserving the between-block variance, as compared to that of the $\mathrm{ASE}$.

This suggests that for a large region of the parameter space for two-block stochastic blockmodels, the spectral embedding of the Laplacian is to be preferred over that of the adjacency matrix for subsequent inference.

However, the metric in B&S is intrinsically tied to the use of $\mathrm{K\text{-}means}$ as the clustering procedure, i.e., a smaller value of the metric for the $\mathrm{LSE}$ as compared to that for the $\mathrm{ASE}$ implies only that clustering the $\mathrm{LSE}$ using $\mathrm{K\text{-}means}$ is possibly better than clustering the $\mathrm{ASE}$ using $\mathrm{K\text{-}means}$.

# GMM ∘ ASE

Athreya et al., *Sankhya*, 2016
provides an ASE CLT
suggesting that the top $K$ eigenvectors from a $K$-SBM adjacency matrix behave approximately as a random sample from a mixture of $K$ Gaussians in $\mathbb{R}^K$.

Tang & P, *Annals of Statistics*, 2018
provides an LSE CLT
and demonstrates that the choice between ASE and LSE is a sticky wicket
as neither dominates the other for subsequent inference ...
and that K-means is inferior to GMM for spectral clustering.

## Definition (Adjacency Spectral Embedding)

Let $\mathbf{A}$ be a $n \times n$ adjacency matrix. Suppose the eigendecomposition of $\mathbf{A}$ is given by $\mathbf{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top$ where $|\lambda_1| \geqslant |\lambda_2| \geqslant \dots$ are the eigenvalues and $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$ are the corresponding orthonormal eigenvectors. Given a positive integer $d \leqslant n$, denote by $\mathbf{S}_\mathbf{A} = \operatorname{diag}(|\lambda_1|, \dots, |\lambda_d|)$ the diagonal matrix whose diagonal entries are the $|\lambda_1|, \dots, |\lambda_d|$, and denote by $\mathbf{U}_\mathbf{A}$ the $n \times d$ matrix whose columns are the corresponding eigenvectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_d$. The *adjacency spectral embedding* (ASE) of $\mathbf{A}$ into $\mathbb{R}^d$ is then the $n \times d$ matrix $\hat{\mathbf{X}} = \mathbf{U}_\mathbf{A} \mathbf{S}_\mathbf{A}^{1/2}$.

## Definition (Graph Laplacian)

For a given matrix $\mathbf{M}$ with non-negative entries, denote by $\mathcal{L}(\mathbf{M})$ the *normalized* Laplacian of $\mathbf{M}$ defined as

$$\mathcal{L}(\mathbf{M}) = (\mathrm{diag}(\mathbf{M1}))^{-1/2}\mathbf{M}(\mathrm{diag}(\mathbf{M1}))^{-1/2}$$

where, given $\boldsymbol{z} = (z_1, \ldots, z_n) \in \mathbb{R}^n$, $\mathrm{diag}(\boldsymbol{z})$ is the $n \times n$ diagonal matrix whose diagonal entries are the $z_i$'s.

Our definition of the normalized Laplacian is slightly different from that often found in the literature, wherein the normalized Laplacian is $\mathbf{I} - \mathcal{L}(\mathbf{M})$. For our purposes, namely the notion of the Laplacian spectral embedding via the eigenvalues and eigenvectors of the normalized Laplacian, these two definitions of the normalized Laplacian are equivalent. We shall henceforth refer to $\mathcal{L}(\mathbf{M})$ as the Laplacian of $\mathbf{M}$, in contrast to the *combinatorial* Laplacian $\mathrm{diag}(\mathbf{M1}) - \mathbf{M}$ of $\mathbf{M}$.

## Definition (Laplacian Spectral Embedding)

Let $\mathbf{A}$ be a $n \times n$ adjacency matrix. Let $\mathcal{L}(\mathbf{A})$ denote the normalized Laplacian of $\mathbf{A}$ and suppose the eigendecomposition of $\mathcal{L}(\mathbf{A})$ is given by $\mathcal{L}(\mathbf{A}) = \sum_{i=1}^{n} \tilde{\lambda}_i \tilde{\boldsymbol{u}}_i \tilde{\boldsymbol{u}}_i^\top$ where $|\tilde{\lambda}_1| \geqslant |\tilde{\lambda}_2| \geqslant \cdots \geqslant |\tilde{\lambda}_n| \geqslant 0$ are the eigenvalues and $\tilde{\boldsymbol{u}}_1, \tilde{\boldsymbol{u}}_2, \ldots, \tilde{\boldsymbol{u}}_n$ are the corresponding orthonormal eigenvectors. Then given a positive integer $d \leqslant n$, denote by $\tilde{\mathbf{S}}_{\mathbf{A}} = \operatorname{diag}(|\tilde{\lambda}_1|, \ldots, |\tilde{\lambda}_d|)$ the diagonal matrix whose diagonal entries are the $|\tilde{\lambda}_1|, \ldots, |\tilde{\lambda}_d|$ and denote by $\tilde{\mathbf{U}}_{\mathbf{A}}$ the $n \times d$ matrix whose columns are the eigenvectors $\tilde{\boldsymbol{u}}_1, \ldots, \tilde{\boldsymbol{u}}_d$. The *Laplacian spectral embedding* (LSE) of $\mathbf{A}$ into $\mathbb{R}^d$ is then the $n \times d$ matrix $\check{\mathbf{X}} = \tilde{\mathbf{U}}_{\mathbf{A}} \tilde{\mathbf{S}}_{\mathbf{A}}^{1/2}$.

## Definition (Random Dot Product Graph (RDPG))

Let $F$ be a distribution on a set $\mathcal{X} \subset \mathbb{R}^d$ satisfying $x^\top y \in [0,1]$ for all $x, y \in \mathcal{X}$. We say $(\mathbf{X}, \mathbf{A}) \sim \mathrm{RDPG}(F)$ with sparsity factor $\rho_n \leqslant 1$ if the following hold. Let $X_1, \ldots, X_n \sim^{iid} F$ be independent and identically distribtued random variables and define

$$\mathbf{X} = [X_1 \mid \cdots \mid X_n]^\top \in \mathbb{R}^{n \times d} \text{ and } \mathbf{P} = \rho_n \mathbf{X}\mathbf{X}^\top \in [0,1]^{n \times n}.$$

The $X_i$ are the *latent* positions for the random graph, i.e., we do not observe $\mathbf{X}$, rather we observe only the matrix $\mathbf{A}$. The matrix $\mathbf{A} \in \{0,1\}^{n \times n}$ is defined to be symmetric with all zeroes on the diagonal such that for all $i < j$, conditioned on $X_i, X_j$ the $A_{ij}$ are independent and

$$A_{ij} \sim \mathrm{Bernoulli}(\rho_n X_i^\top X_j);$$

that is,

$$\mathbb{P}[\mathbf{A} \mid \mathbf{X}] = \prod_{i<j} (\rho_n X_i^\top X_j)^{A_{ij}} (1 - \rho_n X_i^\top X_j)^{(1-A_{ij})}.$$

## Theorem (ASE LLN)

*Let* $(\mathbf{X}_n, \mathbf{A}_n) \sim \mathrm{RDPG}(F)$ *with sparsity factor* $\rho_n$. *Then there exists a* $d \times d$ *orthogonal matrix* $\mathbf{W}_n$ *and a* $n \times d$ *matrix* $\mathbf{R}_n$ *such that*

$$\hat{\mathbf{X}}_n \mathbf{W}_n - \rho_n^{1/2} \mathbf{X}_n = \rho_n^{-1/2} (\mathbf{A}_n - \mathbf{P}_n) \mathbf{X}_n (\mathbf{X}_n^\top \mathbf{X}_n)^{-1} + \mathbf{R}_n.$$

*Furthermore,* $\|\mathbf{R}_n\| = O_{\mathbb{P}}((n\rho_n)^{-1/2})$.
*Let* $\mu_F = \mathbb{E}[X_1]$ *and* $\Delta = \mathbb{E}[X_1 X_1^\top]$.

*If* $\rho_n = 1$ *for all* $n$, *then there exists a sequence of orthogonal matrices* $\mathbf{W}_n$ *such that*

$$\|\hat{\mathbf{X}}_n \mathbf{W}_n - \mathbf{X}_n\|_F^2 \xrightarrow{\text{a.s.}} \operatorname{tr}\, \Delta^{-1} \Big( \mathbb{E}[X_1 X_1^\top (X_1^\top \mu_F - X_1^\top \Delta X_1)] \Big) \Delta^{-1}.$$

*If, however,* $\rho_n \to 0$ *and* $n\rho_n = \omega(\log^4 n)$, *then*

$$\|\hat{\mathbf{X}}_n \mathbf{W}_n - \rho_n^{1/2} \mathbf{X}_n\|_F^2 \xrightarrow{\text{a.s.}} \operatorname{tr}\, \Delta^{-1} \Big( \mathbb{E}[X_1 X_1^\top (X_1^\top \mu_F)] \Big) \Delta^{-1}.$$

## Theorem (ASE CLT)

*Assume the setting and notation as above.*
*Denote by* $\hat{X}_i$ *the* $i$*-th row of* $\hat{\mathbf{X}}_n$*.*
*Let* $\Phi(z, \Sigma)$ *denote the cumulative distribution function for the multivariate normal, with mean zero and covariance matrix* $\Sigma$*, evaluated at* $z$*.*

## Theorem (ASE CLT ($\rho_n = 1$))

*If $\rho_n = 1$ for all $n$, then there exists a sequence of orthogonal matrices $\mathbf{W}_n$ such that for each fixed index $i$ and any $z \in \mathbb{R}^d$,*

$$\mathbb{P}\left\{\sqrt{n}(\mathbf{W}_n\hat{X}_i - X_i) \leqslant z\right\} \xrightarrow{\mathrm{d}} \int \Phi(z, \Sigma(x))dF(x)$$

*where*

$$\Sigma(x) = \Delta^{-1}\mathbb{E}[X_1X_1^\top(x^\top X_1 - x^\top X_1X_1^\top x)]\Delta^{-1}.$$

*That is, the sequence $\sqrt{n}(\mathbf{W}_n\hat{X}_i - X_i)$ converges in distribution to a mixture of multivariate normals. We denote this mixture by $\mathcal{N}(0, \tilde{\Sigma}(X_i))$.*

## Theorem (ASE CLT ($\rho_n \to 0$))

*If, however,* $\rho_n \to 0$ *and* $n\rho_n = \omega(\log^4 n)$ *then there exists a sequence of orthogonal matrices* $\mathbf{W}_n$ *such that*

$$\mathbb{P}\left\{\sqrt{n}(\mathbf{W}_n\hat{X}_i - \rho_n^{1/2}X_i) \leqslant z\right\} \xrightarrow{\mathrm{d}} \int \Phi(z, \Sigma_{o(1)}(x))dF(x)$$

*where* $\Sigma_{o(1)}(x) = \Delta^{-1}\mathbb{E}[X_1X_1^\top x^\top X_1]\Delta^{-1}$.

## Definition (SBM as RDPG)

Let

$$F = \sum_{k=1}^{K} \pi_k \delta_{\nu_k}, \quad \pi_1, \cdots, \pi_K > 0, \sum_k \pi_k = 1$$

be a mixture of $K$ point masses in $\mathbb{R}^d$ where $\delta_{\nu_k}$ is the Dirac delta measure at $\nu_k$.

## Corollary (ASE for SBM)

*If $\rho_n \equiv 1$, there exists a sequence of orthogonal matrices $\mathbf{W}_n$ such that for any fixed index $i$,*

$$\mathbb{P}\left\{\sqrt{n}(\mathbf{W}_n\hat{X}_i - X_i) \leqslant z \mid X_i = \nu_k\right\} \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Sigma_k)$$

*where $\Sigma_k = \Sigma(\nu_k)$.*

*If $\rho_n \to 0$ and $n\rho_n = \omega(\log^4(n))$ as $n \to \infty$, then the sequence of orthogonal matrices $\mathbf{W}_n$ satisfies*

$$\mathbb{P}\left\{\sqrt{n}(\mathbf{W}_n\hat{X}_i - \rho_n^{1/2}X_i) \leqslant z \mid X_i = \nu_k\right\} \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Sigma_{o(1),k})$$

*where $\Sigma_{o(1),k} = \Sigma_{o(1)}(\nu_k)$.*

We now provide analogues of the aforementioned ASE limit results for LSE.

## Theorem (LSE LLN)

*Let* $(\mathbf{A}_n, \mathbf{X}_n) \sim \mathrm{RDPG}(F)$ *for* $n \geqslant 1$ *be a sequence of random dot product graphs with sparsity factors* $(\rho_n)_{n\geqslant 1}$. *Denote by* $\mathbf{D}_n$ *and* $\mathbf{T}_n$ *the* $n \times n$ *diagonal matrices* $\mathrm{diag}(\mathbf{A}_n\mathbf{1})$ *and* $\mathrm{diag}(\rho_n\mathbf{X}_n\mathbf{X}_n^\top\mathbf{1})$, *respectively, i.e., the diagonal entries of* $\mathbf{D}_n$ *are the vertex degrees of* $\mathbf{A}_n$ *and the diagonal entries of* $\mathbf{T}_n$ *are the* expected *vertex degrees. Let* $\tilde{\mathbf{X}}_n = \rho_n^{1/2}\mathbf{T}_n^{-1/2}\mathbf{X}_n = \mathrm{diag}(\mathbf{X}_n\mathbf{X}_n^\top\mathbf{1})^{-1/2}\mathbf{X}_n$.

*Then for any* $n$, *there exists a* $d \times d$ *orthogonal matrix* $\mathbf{W}_n$ *and a* $n \times d$ *matrix* $\mathbf{R}_n$ *such that* $\zeta_n := (\check{\mathbf{X}}_n\mathbf{W}_n - \tilde{\mathbf{X}}_n)$ *satisfies*

$$\zeta_n = \mathbf{T}_n^{-1/2}(\mathbf{A}_n - \mathbf{P}_n)\mathbf{T}_n^{-1/2}\tilde{\mathbf{X}}_n(\tilde{\mathbf{X}}_n^\top\tilde{\mathbf{X}}_n)^{-1} + \tfrac{1}{2}(\mathbf{I} - \mathbf{D}_n\mathbf{T}_n^{-1})\tilde{\mathbf{X}}_n + \mathbf{R}_n. \tag{1}$$

*Furthermore,* $\|\mathbf{R}_n\|_F = O_{\mathbb{P}}((n\rho_n)^{-1})$, *i.e.,* $\|\mathbf{R}_n\|/\|\zeta_n\| \xrightarrow{\text{a.s.}} 0$ *as* $n \to \infty$.

## Theorem (LSE LLN)

*Define the following quantities*

$$\mu = \mathbb{E}[X_1]; \quad \tilde{\mu} = \mathbb{E}\Big[\frac{X_1}{X_1^\top \mu}\Big]; \quad \tilde{\Delta} = \mathbb{E}\Big[\frac{X_1 X_1^\top}{X_1^\top \mu}\Big]; \quad \textit{and} \tag{2}$$

$$g(X_1, X_2) = \Big(\frac{\tilde{\Delta}^{-1} X_1}{X_1^\top \mu} - \frac{X_2}{2X_2^\top \mu}\Big)\Big(\frac{\tilde{\Delta}^{-1} X_1}{X_1^\top \mu} - \frac{X_2}{2X_2^\top \mu}\Big)^\top. \tag{3}$$

## Theorem (LSE LLN)

*If $\rho_n \equiv 1$ then the sequence of orthogonal matrices $(\mathbf{W}_n)_{n\geqslant 1}$ satisfies*

$$n\|\check{\mathbf{X}}_n\mathbf{W}_n - \tilde{\mathbf{X}}_n\|_F^2 \overset{\text{a.s.}}{\to} \operatorname{tr}\,\mathbb{E}\left[g(X_1,X_2)\frac{X_1^\top X_2 - X_1^\top X_2 X_2^\top X_1}{X_2^\top \mu}\right] \tag{4}$$

*where the expectation in Eq. (4) is taken with respect to $X_1$ and $X_2$ being drawn i.i.d. according to $F$.*

*Equivalently, with $\Delta = \mathbb{E}[X_1 X_1^\top]$,*

$$n\|\check{\mathbf{X}}_n\mathbf{W}_n - \tilde{\mathbf{X}}_n\|_F^2 \xrightarrow{\text{a.s.}} \operatorname{tr}\,\mathbb{E}\left[\frac{\tilde{\Delta}^{-2}X_1X_1^\top(X_1^\top\tilde{\mu} - X_1^\top\tilde{\Delta}X_1)}{(X_1^\top\mu)^2} - \frac{3X_1X_1^\top}{4(X_1^\top\mu)^2}\right]$$
$$+ \operatorname{tr}\,\mathbb{E}\left[\frac{\tilde{\Delta}^{-1}X_1X_1^\top X_2X_2^\top(X_1^\top X_2)}{X_1^\top\mu(X_2^\top\mu)^2} - \frac{X_1X_1^\top(X_1^\top\Delta X_1)}{4(X_1^\top\mu)^3}\right]$$

## Theorem (LSE LLN)

*If* $\rho_n \to 0$ *and* $n\rho_n = \omega(\log^4 n)$ *then the sequence* $(\mathbf{W}_n)_{n\geqslant 1}$ *satisfies*

$$n\rho_n \|\check{\mathbf{X}}\mathbf{W}_n - \tilde{\mathbf{X}}_n\|_F^2 \xrightarrow{\text{a.s.}} \operatorname{tr} \mathbb{E}\Big[\frac{\tilde{\Delta}^{-2} X_1 X_1^\top (X_1^\top \tilde{\mu})}{(X_1^\top \mu)^2} - \frac{3 X_1 X_1^\top}{4(X_1^\top \mu)^2}\Big]. \quad (5)$$

## Theorem (LSE CLT)

*Assume the setting and notation as above.*
*Denote by* $\check{X}_i$ *and* $\tilde{X}_i$ *the* $i$*-th row of* $\check{\mathbf{X}}_n$ *and* $\tilde{\mathbf{X}}_n$*, respectively.*
*We note that* $\tilde{X}_i = \frac{X_i}{\sqrt{\sum_j X_i^\top X_j}}$.

## Theorem (LSE CLT)

*If $\rho_n \equiv 1$ then there exists a sequence of orthogonal matrices $\mathbf{W}_n$ such that for each fixed index $i$ and any $z \in \mathbb{R}^d$,*

$$\mathbb{P}\Big\{ n\big(\mathbf{W}_n \breve{X}_i - \frac{X_i}{\sqrt{\sum_j X_i^\top X_j}}\big) \leqslant z \Big\} \xrightarrow{\mathrm{d}} \int \Phi(z, \tilde{\Sigma}(x)) dF(x) \tag{6}$$

*where $\tilde{\Sigma}(x)$ is defined by*

$$\mathbb{E}\Big[ \Big( \frac{\tilde{\Delta}^{-1} X_1}{X_1^\top \mu} - \frac{x}{2x^\top \mu} \Big) \Big( \frac{X_1^\top \tilde{\Delta}^{-1}}{X_1^\top \mu} - \frac{x^\top}{2x^\top \mu} \Big) \frac{(x^\top X_1 - x^\top X_1 X_1^\top x)}{x^\top \mu} \Big]. \tag{7}$$

*That is, the sequence $n(\mathbf{W}_n \breve{X}_i - \tilde{X}_i)$ converges in distribution to a mixture of multivariate normals. We denote this mixture by $\mathcal{N}(0, \tilde{\Sigma}(X_i))$.*

## Theorem (LSE CLT)

*If* $\rho_n \to 0$ *and* $n\rho_n = \omega(\log^4 n)$ *then there exists a sequence of orthogonal matrices* $\mathbf{W}_n$ *such that*

$$\mathbb{P}\Big\{ n\rho_n^{1/2}\big(\mathbf{W}_n \breve{X}_i - \frac{X_i}{\sqrt{\sum_j X_i^\top X_j}}\big) \leqslant z \Big\} \xrightarrow{\mathrm{d}} \int \Phi(z, \tilde{\Sigma}_{o(1)}(x)) dF(x). \quad (8)$$

*where* $\tilde{\Sigma}_{o(1)}(x)$ *is defined by*

$$\tilde{\Sigma}_{o(1)}(x) = \mathbb{E}\Big[\Big(\frac{\tilde{\Delta}^{-1}X_1}{X_1^\top \mu} - \frac{x}{2x^\top \mu}\Big)\Big(\frac{X_1^\top \tilde{\Delta}^{-1}}{X_1^\top \mu} - \frac{x^\top}{2x^\top \mu}\Big)\frac{x^\top X_1}{x^\top \mu}\Big]. \quad (9)$$

## Corollary (LSE for SBM)

*Recall*

$$F = \sum_{k=1}^{K} \pi_k \delta_{\nu_k}, \quad \pi_1, \cdots, \pi_K > 0, \sum_k \pi_k = 1.$$

*If* $\rho_n \equiv 1$, *there exists a sequence of orthogonal matrices* $\mathbf{W}_n$ *such that for any fixed index* $i$,

$$\mathbb{P}\Big\{ n\big(\mathbf{W}_n \check{X}_i - \frac{\nu_k}{\sqrt{\sum_l n_l \nu_k^\top \nu_l}}\big) \leqslant z \mid X_i = \nu_k \Big\} \xrightarrow{\mathrm{d}} \mathcal{N}(0, \tilde{\Sigma}_k) \tag{10}$$

*where* $\tilde{\Sigma}_k = \tilde{\Sigma}(\nu_k)$ *is as defined in Eq.* (7).

*If instead* $\rho_n \to 0$ *and* $n\rho_n = \omega(\log^4(n))$ *as* $n \to \infty$ *then*

$$\mathbb{P}\Big\{ n\rho_n^{1/2}\big(\mathbf{W}_n \check{X}_i - \frac{\nu_k}{\sqrt{\sum_l n_l \nu_k^\top \nu_l}}\big) \leqslant z \mid X_i = \nu_k \Big\} \xrightarrow{\mathrm{d}} \mathcal{N}(0, \tilde{\Sigma}_{o(1),k}) \tag{11}$$

*where* $\tilde{\Sigma}_{o(1),k} = \tilde{\Sigma}_{o(1)}(\nu_k)$ *is as defined in Eq.* (9).

As a special case, let $\mathbf{A}$ be an Erdős-Rényi graph on $n$ vertices with edge probability $p^2$ – which corresponds to a random dot product graph where the latent positions are identically $p$.

Then for each fixed index $i$:

LSE yields

$$n\big(\breve{X}_i - \tfrac{1}{\sqrt{n}}\big) \xrightarrow{\mathrm{d}} \mathcal{N}\big(0, \tfrac{1-p^2}{4p^2}\big);$$

ASE yields

$$\sqrt{n}(\hat{X}_i - p) \xrightarrow{\mathrm{d}} \mathcal{N}(0, 1-p^2).$$

As another example, if $\mathbf{A}$ is a stochastic blockmodel graph with block probabilities matrix $\mathbf{B} = \left[\begin{smallmatrix} p^2 & pq \\ pq & q^2 \end{smallmatrix}\right]$ and block assignment probabilities $(\pi, 1-\pi)$ – which corresponds to a random dot product graph where the latent positions are either $p$ with probability $\pi$ or $q$ with probability $1-\pi$ – then letting $n_1$ and $n_2 = n - n_1$ denote the number of vertices of $\mathbf{A}$ with latent positions $p$ and $q$, we have that for each fixed $i$:

LSE yields

$$n\big(\check{X}_i - \tfrac{p}{\sqrt{n_1 p^2 + n_2 pq}}\big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \tfrac{\pi p(1-p^2)+(1-\pi)q(1-pq)}{4(\pi p+(1-\pi)q)^3}\Big) \text{ if } X_i = p,$$

$$n\big(\check{X}_i - \tfrac{q}{\sqrt{n_1 pq + n_2 q^2}}\big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \tfrac{\pi p(1-pq)+(1-\pi)q(1-q^2)}{4(\pi p+(1-\pi)q)^3}\Big) \text{ if } X_i = q;$$

ASE yields

$$\sqrt{n}(\hat{X}_i - p) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \tfrac{\pi p^4(1-p^2)+(1-\pi)pq^3(1-pq)}{(\pi p^2+(1-\pi)q^2)^2}\Big) \text{ if } X_i = p,$$

$$\sqrt{n}(\hat{X}_i - q) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \tfrac{\pi p^3 q(1-pq)+(1-\pi)q^4(1-q^2)}{(\pi p^2+(1-\pi)q^2)^2}\Big) \text{ if } X_i = q.$$

Section 3.1: sketch of (one *key* & *fun* part of) the proof

The LSE of RDPG $\mathbf{A}$ into $\mathbb{R}^d$ is the $n \times d$ matrix $\check{\mathbf{X}} = \tilde{\mathbf{U}}_{\mathbf{A}} \tilde{\mathbf{S}}_{\mathbf{A}}^{1/2}$.

Davis-Kahan implies $\tilde{\mathbf{U}}_{\mathbf{A}} \tilde{\mathbf{U}}_{\mathbf{A}}^{\top} = \tilde{\mathbf{U}}_{\mathbf{P}} \tilde{\mathbf{U}}_{\mathbf{P}}^{\top} + O_{\mathbb{P}}((n\rho_n)^{-1/2})$ and ...

Minh's Proposition B.2: There exists an orthogonal matrix $\mathbf{W}^*$ s.t.

$$\tilde{\mathbf{U}}_{\mathbf{P}}^{\top} \tilde{\mathbf{U}}_{\mathbf{A}} = \mathbf{W}^* + O_{\mathbb{P}}((n\rho_n)^{-1}).$$

Minh's Lemma B.3: Furthermore, $\mathbf{W}^*$ satisfies

$$\mathbf{W}^* \tilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} - \tilde{\mathbf{S}}_{\mathbf{P}}^{-1/2} \mathbf{W}^* = O_{\mathbb{P}}((n\rho_n)^{-1}).$$

By the Davis-Kahan theorem, the eigenspace spanned by the $d$ largest eigenvalues of $\mathcal{L}(\mathbf{A})$ is "close" to that spanned by the $d$ largest eigenvalues of $\mathcal{L}(\mathbf{P})$.
That is, $\tilde{\mathbf{U}}_\mathbf{A}\tilde{\mathbf{U}}_\mathbf{A}^\top = \tilde{\mathbf{U}}_\mathbf{P}\tilde{\mathbf{U}}_\mathbf{P}^\top + O_{\mathbb{P}}((n\rho_n)^{-1/2})$ and

$$\tilde{\mathbf{U}}_\mathbf{A}\tilde{\mathbf{S}}_\mathbf{A}^{1/2} - \tilde{\mathbf{U}}_\mathbf{P}\tilde{\mathbf{S}}_\mathbf{P}^{1/2}\tilde{\mathbf{U}}_\mathbf{P}^\top\tilde{\mathbf{U}}_\mathbf{A} = \mathcal{L}(\mathbf{A})\tilde{\mathbf{U}}_\mathbf{P}\tilde{\mathbf{U}}_\mathbf{P}^\top\tilde{\mathbf{U}}_\mathbf{A}\tilde{\mathbf{S}}_\mathbf{A}^{-1/2} - \mathcal{L}(\mathbf{P})\tilde{\mathbf{U}}_\mathbf{P}\tilde{\mathbf{S}}_\mathbf{P}^{-1/2}\tilde{\mathbf{U}}_\mathbf{P}^\top\tilde{\mathbf{U}}_\mathbf{A} + O_{\mathbb{P}}((n\rho_n)^{-1}).$$

Consider the terms $\tilde{\mathbf{S}}_\mathbf{P}^{-1/2}\tilde{\mathbf{U}}_\mathbf{P}^\top\tilde{\mathbf{U}}_\mathbf{A}$ and $\tilde{\mathbf{U}}_\mathbf{P}^\top\tilde{\mathbf{U}}_\mathbf{A}\tilde{\mathbf{S}}_\mathbf{A}^{-1/2}$.
Since $\tilde{\mathbf{U}}_\mathbf{P}$ and $\tilde{\mathbf{U}}_\mathbf{A}$ both have orthonormal columns, $\tilde{\mathbf{U}}_\mathbf{A}\tilde{\mathbf{U}}_\mathbf{A}^\top = \tilde{\mathbf{U}}_\mathbf{P}\tilde{\mathbf{U}}_\mathbf{P}^\top + O_{\mathbb{P}}((n\rho_n)^{-1/2})$ implies that there exists an orthogonal matrix $\mathbf{W}^*$ such that $\tilde{\mathbf{U}}_\mathbf{P}^\top\tilde{\mathbf{U}}_\mathbf{A} = \mathbf{W}^* + O_{\mathbb{P}}((n\rho_n)^{-1})$ (Proposition B.2).

Furthermore, $\mathbf{W}^*$ satisfies an important property, namely that $\mathbf{W}^*\tilde{\mathbf{S}}_\mathbf{A}^{-1/2} - \tilde{\mathbf{S}}_\mathbf{P}^{-1/2}\mathbf{W}^* = O_{\mathbb{P}}((n\rho_n)^{-1})$ (Lemma B.3).

We can thus juxtapose $\tilde{\mathbf{U}}_\mathbf{P}^\top\tilde{\mathbf{U}}_\mathbf{A}$ and $\tilde{\mathbf{S}}_\mathbf{A}^{-1/2}$ in the above expression and replace $\tilde{\mathbf{U}}_\mathbf{P}^\top\tilde{\mathbf{U}}_\mathbf{A}$ by the orthogonal matrix $\mathbf{W}^*$, thereby yielding

$$\tilde{\mathbf{U}}_\mathbf{A}\tilde{\mathbf{S}}_\mathbf{A}^{1/2} - \tilde{\mathbf{U}}_\mathbf{P}\tilde{\mathbf{S}}_\mathbf{P}^{1/2}\mathbf{W}^* = (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\tilde{\mathbf{U}}_\mathbf{P}\tilde{\mathbf{S}}_\mathbf{P}^{-1/2}\mathbf{W}^* + O_{\mathbb{P}}((n\rho_n)^{-1}).$$

By the Davis-Kahan theorem, the eigenspace spanned by the $d$ largest eigenvalues of $\mathcal{L}(\mathbf{A})$ is "close" to that spanned by the $d$ largest eigenvalues of $\mathcal{L}(\mathbf{P})$.
That is, $\tilde{\mathbf{U}}_{\mathbf{A}}\tilde{\mathbf{U}}_{\mathbf{A}}^{\top} = \tilde{\mathbf{U}}_{\mathbf{P}}\tilde{\mathbf{U}}_{\mathbf{P}}^{\top} + O_{\mathbb{P}}((n\rho_n)^{-1/2})$ and

$$\tilde{\mathbf{U}}_{\mathbf{A}}\tilde{\mathbf{S}}_{\mathbf{A}}^{1/2} - \tilde{\mathbf{U}}_{\mathbf{P}}\tilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\tilde{\mathbf{U}}_{\mathbf{P}}^{\top}\tilde{\mathbf{U}}_{\mathbf{A}} = \mathcal{L}(\mathbf{A})\tilde{\mathbf{U}}_{\mathbf{P}}\tilde{\mathbf{U}}_{\mathbf{P}}^{\top}\tilde{\mathbf{U}}_{\mathbf{A}}\tilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} - \mathcal{L}(\mathbf{P})\tilde{\mathbf{U}}_{\mathbf{P}}\tilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\tilde{\mathbf{U}}_{\mathbf{P}}^{\top}\tilde{\mathbf{U}}_{\mathbf{A}} + O_{\mathbb{P}}((n\rho_n)^{-1}).$$

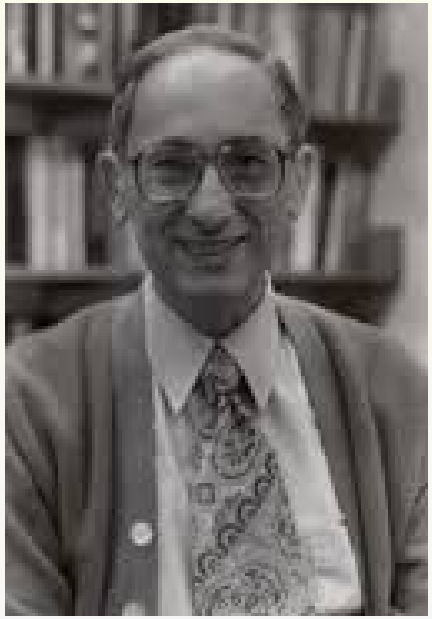
Consider the terms $\tilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\tilde{\mathbf{U}}_{\mathbf{P}}^{\top}\tilde{\mathbf{U}}_{\mathbf{A}}$ and $\tilde{\mathbf{U}}_{\mathbf{P}}^{\top}\tilde{\mathbf{U}}_{\mathbf{A}}\tilde{\mathbf{S}}_{\mathbf{A}}^{-1/2}$.
Since $\tilde{\mathbf{U}}_{\mathbf{P}}$ and $\tilde{\mathbf{U}}_{\mathbf{A}}$ both have orthonormal columns, $\tilde{\mathbf{U}}_{\mathbf{A}}\tilde{\mathbf{U}}_{\mathbf{A}}^{\top} = \tilde{\mathbf{U}}_{\mathbf{P}}\tilde{\mathbf{U}}_{\mathbf{P}}^{\top} + O_{\mathbb{P}}((n\rho_n)^{-1/2})$ implies that there exists an orthogonal matrix $\mathbf{W}^*$ such that $\tilde{\mathbf{U}}_{\mathbf{P}}^{\top}\tilde{\mathbf{U}}_{\mathbf{A}} = \mathbf{W}^* + O_{\mathbb{P}}((n\rho_n)^{-1})$ (Proposition B.2).

Furthermore, $\mathbf{W}^*$ satisfies an important property, namely that $\mathbf{W}^*\tilde{\mathbf{S}}_{\mathbf{A}}^{-1/2} - \tilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^* = O_{\mathbb{P}}((n\rho_n)^{-1})$ (Lemma B.3).

We can thus juxtapose $\tilde{\mathbf{U}}_{\mathbf{P}}^{\top}\tilde{\mathbf{U}}_{\mathbf{A}}$ and $\tilde{\mathbf{S}}_{\mathbf{A}}^{-1/2}$ in the above expression and replace $\tilde{\mathbf{U}}_{\mathbf{P}}^{\top}\tilde{\mathbf{U}}_{\mathbf{A}}$ by the orthogonal matrix $\mathbf{W}^*$, thereby yielding

$$\tilde{\mathbf{U}}_{\mathbf{A}}\tilde{\mathbf{S}}_{\mathbf{A}}^{1/2} - \tilde{\mathbf{U}}_{\mathbf{P}}\tilde{\mathbf{S}}_{\mathbf{P}}^{1/2}\mathbf{W}^* = (\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P}))\tilde{\mathbf{U}}_{\mathbf{P}}\tilde{\mathbf{S}}_{\mathbf{P}}^{-1/2}\mathbf{W}^* + O_{\mathbb{P}}((n\rho_n)^{-1}).$$

# Chernoff Information

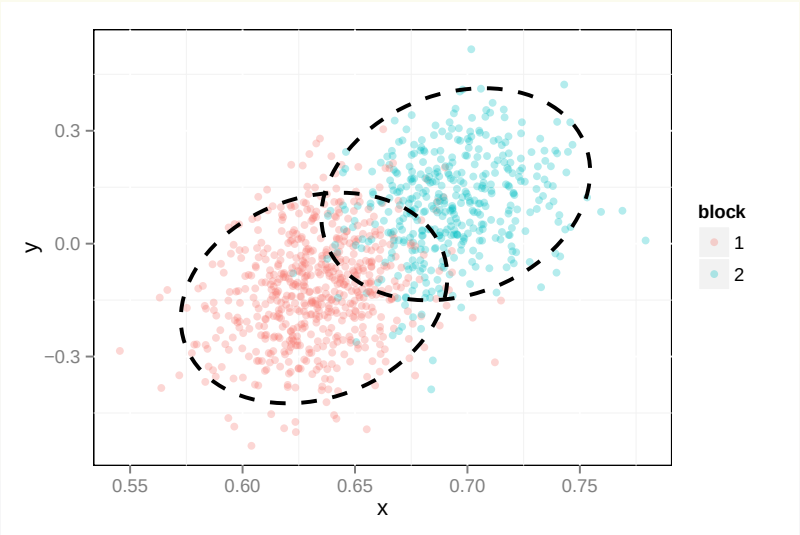


H. Chernoff, *Ann. Math. Stat.*, 1952 & 1956.

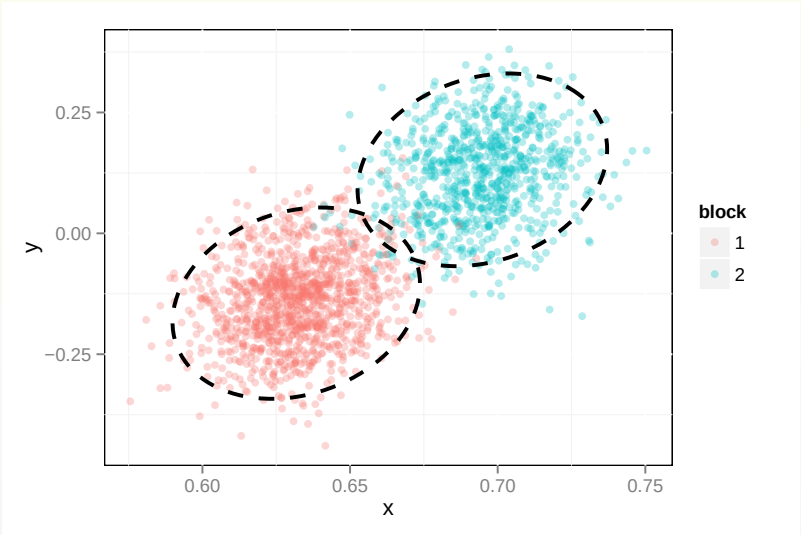
Consider

$$SBM\left( \quad \mathbf{B} = \begin{bmatrix} 0.42 & 0.42 \\ 0.42 & 0.5 \end{bmatrix} \quad , \quad \pi = [0.6, 0.4]^\top \quad \right)$$

# LSE(SBM)




$n = 1000$




$n = 2000$

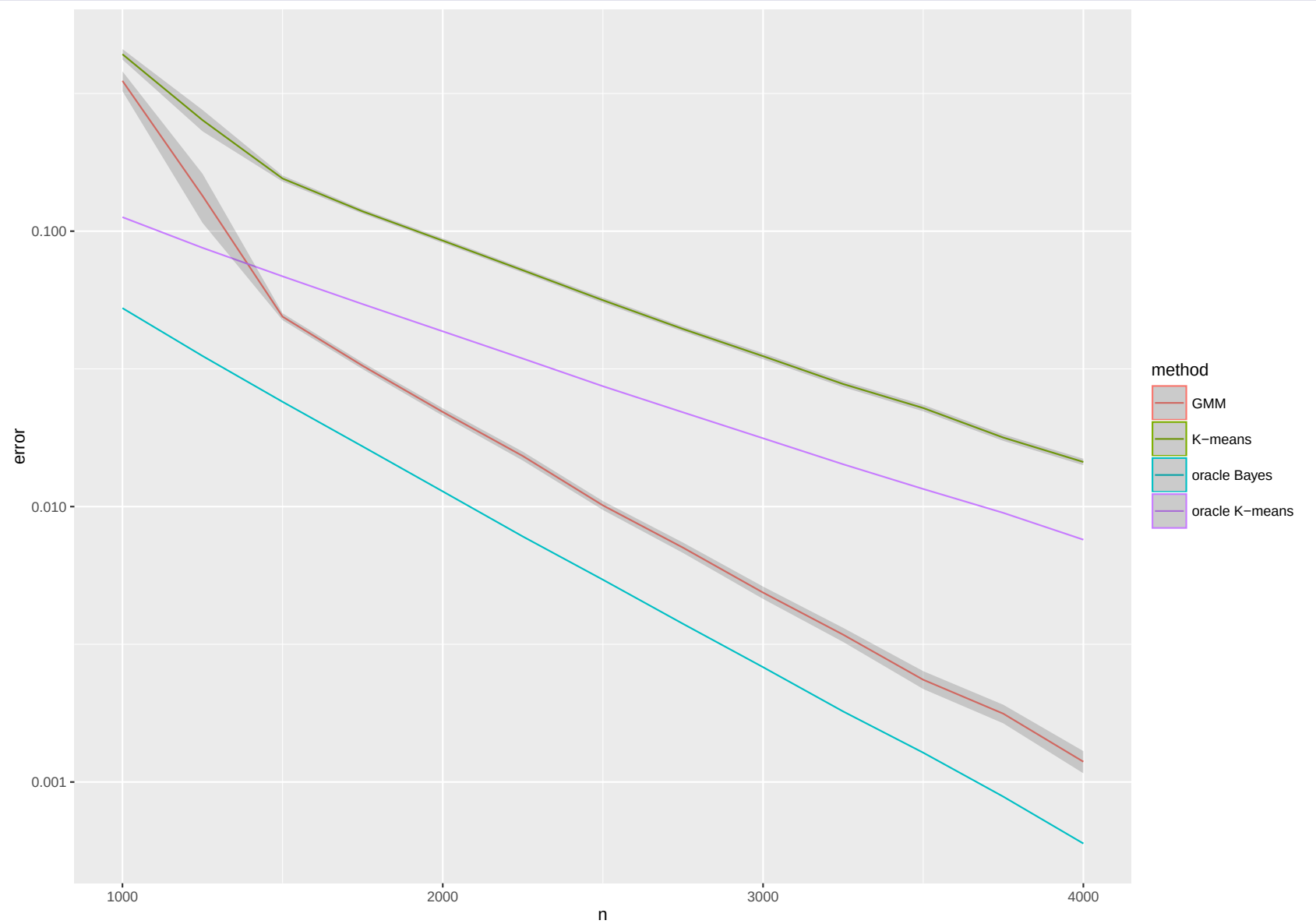



Figure 1: Clustering error rates (ordinate, on a $\log_{10}$ scale) vs. $n$ (abscissa) for $K$-means, oracle $K$-means, GMM, and oracle GMM.

# ASE vs LSE for subsequent inference

Section 4.1: within-block variances are insufficient

One metric for comparison is the notion of within-block variance for each block of the stochastic blockmodel.

We partially extend the results of B&S 2015 for two-block SBMs to $K$-block SBMs with positive semidefinite block probablity matrices.

However, while the collection of within-block variances is a meaningful surrogate for the performance of our subsequent inference task, we argue that it is not the "right" metric as it captures only the **trace** of the block-conditional covariance matrices.

That is to say, the use of the within-block variances as a surrogate measure is similar to the oracle $K$-means lower bound in the figure.

# ASE vs LSE for subsequent inference

Section 4.1: within-block variances are insufficient

A more appropriate surrogate is the collection of pairwise Chernoff informations between the block-conditional multivariate normals, which behave similarly to the oracle Bayes lower bound.

Roughly speaking, we want to compare, for a given SBM graph $\mathbf{A}$, the large-sample error rate of $\inf_T T \circ \mathrm{ASE}$ versus the large-sample error rate of $\inf_T T \circ \mathrm{LSE}$, where $T$ ranges over all possible transformations and clusterings procedure.

This comparison is facilitated by the ASE & LSE CLTs for SBMs.

Let $F_0$ and $F_1$ be two absolutely continuous multivariate distributions in $\Omega = \mathbb{R}^d$ with density functions $f_0$ and $f_1$, respectively.

Suppose that $Y_1, Y_2, \ldots, Y_m$ are independent and identically distributed random variables, with $Y_i$ distributed either $F_0$ or $F_1$.

We are interested in testing the simple null hypothesis $\mathbb{H}_0 \colon F = F_0$ against the simple alternative hypothesis $\mathbb{H}_1 \colon F = F_1$.

A test $T$ can be viewed as a sequence of mappings $T_m \colon \Omega^m \mapsto \{0, 1\}$ such that given $Y_1 = y_1, Y_2 = y_2, \ldots, Y_m = y_m$, the test rejects $\mathbb{H}_0$ in favor of $\mathbb{H}_1$ if $T_m(y_1, y_2, \ldots, y_m) = 1$; similarly, the test favors $\mathbb{H}_0$ if $T_m(y_1, y_2, \ldots, y_m) = 0$.

The Neyman-Pearson lemma states that, given $Y_1 = y_1, Y_2 = y_2, \ldots, Y_m = y_m$ and a threshold $\eta_m \in \mathbb{R}$, the likelihood ratio test which rejects $\mathbb{H}_0$ in favor of $\mathbb{H}_1$ whenever

$$\Big(\sum_{i=1}^{m} \log f_0(y_i) - \sum_{i=1}^{m} \log f_1(y_i)\Big) \leqslant \eta_m$$

is the most powerful test at significance level $\alpha_m = \alpha(\eta_m)$, i.e., the likelihood ratio test minimizes the type-II error $\beta_m$ subject to the contrainst that the type-I error is at most $\alpha_m$.

Assume that $\pi \in (0,1)$ is a prior probability that $\mathbb{H}_0$ is true. Then, for a given $\alpha_m^* \in (0,1)$, let $\beta_m^* = \beta_m^*(\alpha_m^*)$ be the type-II error associated with the likelihood ratio test when the type-I error is at most $\alpha_m^*$.

The quantity $\inf_{\alpha_m^* \in (0,1)} \pi\alpha_m^* + (1-\pi)\beta_m^*$ is then the Bayes risk in deciding between $\mathbb{H}_0$ and $\mathbb{H}_1$ given the $m$ independent random variables $Y_1, Y_2, \ldots, Y_m$.

A classical result of Chernoff (1952,1956) states that the Bayes risk is intrinsically linked to a quantity known as the *Chernoff information*. More specifically, let $C(F_0, F_1)$ be the quantity

$$\begin{aligned} C(F_0, F_1) &= -\log\left[\inf_{t\in(0,1)} \int_{\mathbb{R}^d} f_0^t(\boldsymbol{x}) f_1^{1-t}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}\right] \\ &= \sup_{t\in(0,1)} \left[-\log \int_{\mathbb{R}^d} f_0^t(\boldsymbol{x}) f_1^{1-t}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}\right]. \end{aligned} \tag{12}$$

Then we have

$$\lim_{m\to\infty} \frac{1}{m} \inf_{\alpha_m^* \in (0,1)} \log(\pi\alpha_m^* + (1-\pi)\beta_m^*) = -C(F_0, F_1). \qquad (13)$$

Thus $C(F_0, F_1)$, the Chernoff information between $F_0$ and $F_1$, is the *exponential* rate at which the Bayes error

$$\inf_{\alpha_m^* \in (0,1)} \pi\alpha_m^* + (1-\pi)\beta_m^*$$

decreases as $m \to \infty$.

Note that the Chernoff information is independent of $\pi$.

We also define, for a given $t \in (0,1)$ the Chernoff divergence $C_t(F_0, F_1)$ between $F_0$ and $F_1$ by

$$C_t(F_0, F_1) = -\log \int_{\mathbb{R}^d} f_0^t(\boldsymbol{x}) f_1^{1-t}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}.$$

The Chernoff divergence is an example of an $f$-divergence.

$C_{1/2}(F_0, F_1)$ is the Bhattacharyya distance between $F_0$ and $F_1$.

Any $f$-divergence satisfies the information processing lemma and is invariant with respect to invertible transformations.

Thus any $f$-divergence such as the Kullback-Liebler divergence can also be used to compare $\mathrm{ASE}$ & $\mathrm{LSE}$.

We choose the Chernoff information mainly because of its explicit relationship with the Bayes risk.

The result of Eq. (13) can be extended to $K+1 \geqslant 2$ hypotheses. Let $F_0, F_1, \ldots, F_K$ be distributions on $\mathbb{R}^d$ and suppose that $Y_1, Y_2, \ldots, Y_m$ are independent and identically distributed random variables with $Y_i$ distributed $F \in \{F_0, F_1, \ldots, F_K\}$. We are thus interested in determining the distribution of the $Y_i$ among the $K+1$ hypothesis $\mathbb{H}_0 \colon F = F_0, \ldots, \mathbb{H}_K \colon F = F_K$. Suppose also that hypothesis $\mathbb{H}_k$ has *a priori* probability $\pi_k$. Then for any decision rule $\delta$, the risk of $\delta$ is $r(\delta) = \sum_k \pi_k \sum_{l \neq k} \alpha_{lk}(\delta)$ where $\alpha_{lk}(\delta)$ is the probability of accepting hypothesis $\mathbb{H}_l$ when hypothesis $\mathbb{H}_k$ is true. Then we have

$$\inf_{\delta} \lim_{m \to \infty} \frac{r(\delta)}{m} = -\min_{k \neq l} C(F_k, F_l) \tag{14}$$

where the infimum is over all decision rules $\delta$.

That is, $r(\delta)$ decreases to 0 as $m \to \infty$ at a rate no faster than

$$\exp(-m \min_{k \neq l} C(F_k, F_l)).$$

For our purposes, we require the Chernoff information $C(F_0, F_1)$ when $F_0$ and $F_1$ are multivariate normals.

Suppose $F_0 = \mathcal{N}(\mu_0, \Sigma_0)$ and $F_1 = \mathcal{N}(\mu_1, \Sigma_1)$;
then, with $\Sigma_t = t\Sigma_0 + (1-t)\Sigma_1$, we have

$$C(F_0, F_1) = \sup_{t \in (0,1)} \Big( \frac{t(1-t)}{2} (\mu_1 - \mu_2)^\top \Sigma_t^{-1} (\mu_1 - \mu_2) + \frac{1}{2} \log \frac{|\Sigma_t|}{|\Sigma_0|^t |\Sigma_1|^{1-t}} \Big).$$

We now employ our $\mathrm{ASE}$ & $\mathrm{LSE}$ CLTs to compare the performance of the two spectral embedding methods for subsequent inference.

Our subsequent inference task is the recovery of block assignments.

We are interested in deriving the *large-sample optimal* error rate for recovering the underlying block assignments in stochastic blockmodel graphs after the spectral embedding step is carried out.

An appropriate measure for the large-sample optimal error rate for spectral clustering is in terms of the minimum of the pairwise Chernoff informations between the multivariate normal distributions as specified by the CLTs.

Let $\mathbf{B} \in [0,1]^{K \times K}$ and $\boldsymbol{\pi} \in \mathbb{R}^K$ be the matrix of block probabilities and the vector of block assignment probablities for a $K$-block stochastic blockmodel. Assume that $\mathbf{B}$ is positive semidefinite. Then given an $n$ vertex instantiation of the SBM graph with parameters $(\boldsymbol{\pi}, \mathbf{B})$, for sufficiently large $n$, the large-sample optimal error rate for recovering the block assignments ...

. . . when ASE is used as the initial embedding step can be characterized by the quantity $\rho_{\mathrm{A}} = \rho_{\mathrm{A}}(n)$ defined by

$$\rho_{\mathrm{A}} = \min_{k \neq l} \sup_{t \in (0,1)} \frac{1}{2} \log \frac{|\Sigma_{kl}(t)|}{|\Sigma_k|^t |\Sigma_l|^{1-t}} + \frac{nt(1-t)}{2} (\nu_k - \nu_l)^\top \Sigma_{kl}^{-1}(t) (\nu_k - \nu_l) \tag{15}$$

where $\Sigma_{kl}(t) = t\Sigma_k + (1-t)\Sigma_l$.

. . . when LSE is used as the initial embedding step can be characterized by the quantity $\rho_{\mathrm{L}} = \rho_{\mathrm{L}}(n)$ defined by

$$\rho_{\mathrm{L}} = \min_{k \neq l} \sup_{t \in (0,1)} \frac{1}{2} \log \frac{|\tilde{\Sigma}_{kl}(t)|}{|\tilde{\Sigma}_k|^t |\tilde{\Sigma}_l|^{1-t}} + \frac{nt(1-t)}{2} (\tilde{\nu}_k - \tilde{\nu}_l)^\top \tilde{\Sigma}_{kl}^{-1}(t) (\tilde{\nu}_k - \tilde{\nu}_l) \tag{16}$$

where $\tilde{\Sigma}_{kl}(t) = t\tilde{\Sigma}_k + (1-t)\tilde{\Sigma}_l$ and $\tilde{\nu}_k = \nu_k / (\sum_{k'} \pi_{k'} \nu_k^\top \nu_{k'})^{1/2}$.

Recall that as the Chernoff information increases,
the large-sample optimal error rate decreases.

For ease of comparison between $\rho_{\mathrm{A}}$ and $\rho_{\mathrm{L}}$, we have made the simplifying assumption that $n_k = n\pi_k$ in our expression for $\tilde{\nu}_k$ in Eq. (16).

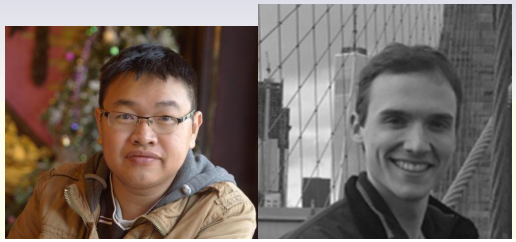
As an illustration, we first consider the collection of 2-block stochastic blockmodels where $\mathbf{B} = \begin{bmatrix} p^2 & pq \\ pq & q^2 \end{bmatrix}$ for $p, q \in (0,1)$ and $\boldsymbol{\pi} = (\pi_1, \pi_2)$ with $\pi_1 + \pi_2 = 1$.

Then for sufficiently large $n$ we have

$$\rho_A \approx \frac{n(p-q)^2(\pi_1 p^2 + \pi_2 q^2)^2}{2\left(\sqrt{\pi_1 p^4(1-p^2) + \pi_2 pq^3(1-pq)} + \sqrt{\pi_1 p^3 q(1-pq) + \pi_2 q^4(1-q^2)}\right)^2}$$

and

$$\rho_L \approx \frac{2n(\sqrt{p}-\sqrt{q})^2(\pi_1 p + \pi_2 q)^2}{\left(\sqrt{\pi_1 p(1-p^2) + \pi_2 q(1-pq)} + \sqrt{\pi_1 p(1-pq) + \pi_2 q(1-q^2)}\right)^2}.$$



# On spectral embedding performance and elucidating network structure in stochastic block model graphs

Joshua Cape, Minh Tang, Carey E. Priebe

*(Submitted on 14 Aug 2018)*

Statistical inference on graphs often proceeds via spectral methods involving low-dimensional embeddings of matrix-valued graph representations, such as the graph Laplacian or adjacency matrix. In this paper, we analyze the asymptotic information-theoretic relative performance of Laplacian spectral embedding and adjacency spectral embedding for block assignment recovery in stochastic block model graphs by way of Chernoff information. We investigate the relationship between spectral embedding performance and underlying network structure (e.g.~homogeneity, affinity, core-periphery, (un)balancedness) via a comprehensive treatment of the two-block stochastic block model and the class of $K$-block models exhibiting homogeneous balanced affinity structure. Our findings support the claim that, for a particular notion of sparsity, loosely speaking, "Laplacian spectral embedding favors relatively sparse graphs, whereas adjacency spectral embedding favors not-too-sparse graphs." We also provide evidence in support of the claim that "adjacency spectral embedding favors core-periphery network structure."

https://arxiv.org/abs/1808.04855

*Network Science*, forthcoming

# On spectral embedding performance and elucidating network structure in stochastic block model graphs

Joshua Cape and Minh Tang and Carey E. Priebe

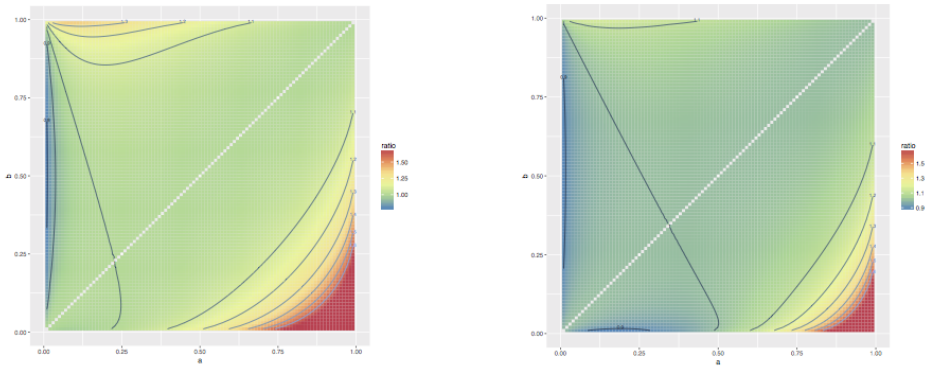
Department of Applied Mathematics and Statistics
The Johns Hopkins University, USA

January 27, 2018

**Abstract**

Statistical inference on graphs often proceeds via spectral methods involving low-dimensional embeddings of matrix-valued graph representations, such as the graph Laplacian or adjacency matrix. In this paper, we characterize the information-theoretic relative performance of Laplacian spectral embedding (LSE) and adjacency spectral embedding (ASE) for block assignment recovery in stochastic block model graphs via Chernoff information. We investigate the relationship between spectral embedding performance and underlying network structure (e.g. homogeneity, core-periphery, (un)balancedness) via a comprehensive treatment of the two-block stochastic block model and the class of $K$-block models exhibiting homogeneous balanced affinity structure. Our findings support the claim that, for a particular notion of sparsity, loosely speaking, "Laplacian spectral embedding favors relatively sparse graphs, whereas adjacency spectral embedding favors not-too-sparse graphs." We also provide evidence in support of the claim that "adjacency spectral embedding favors core-periphery network structure."

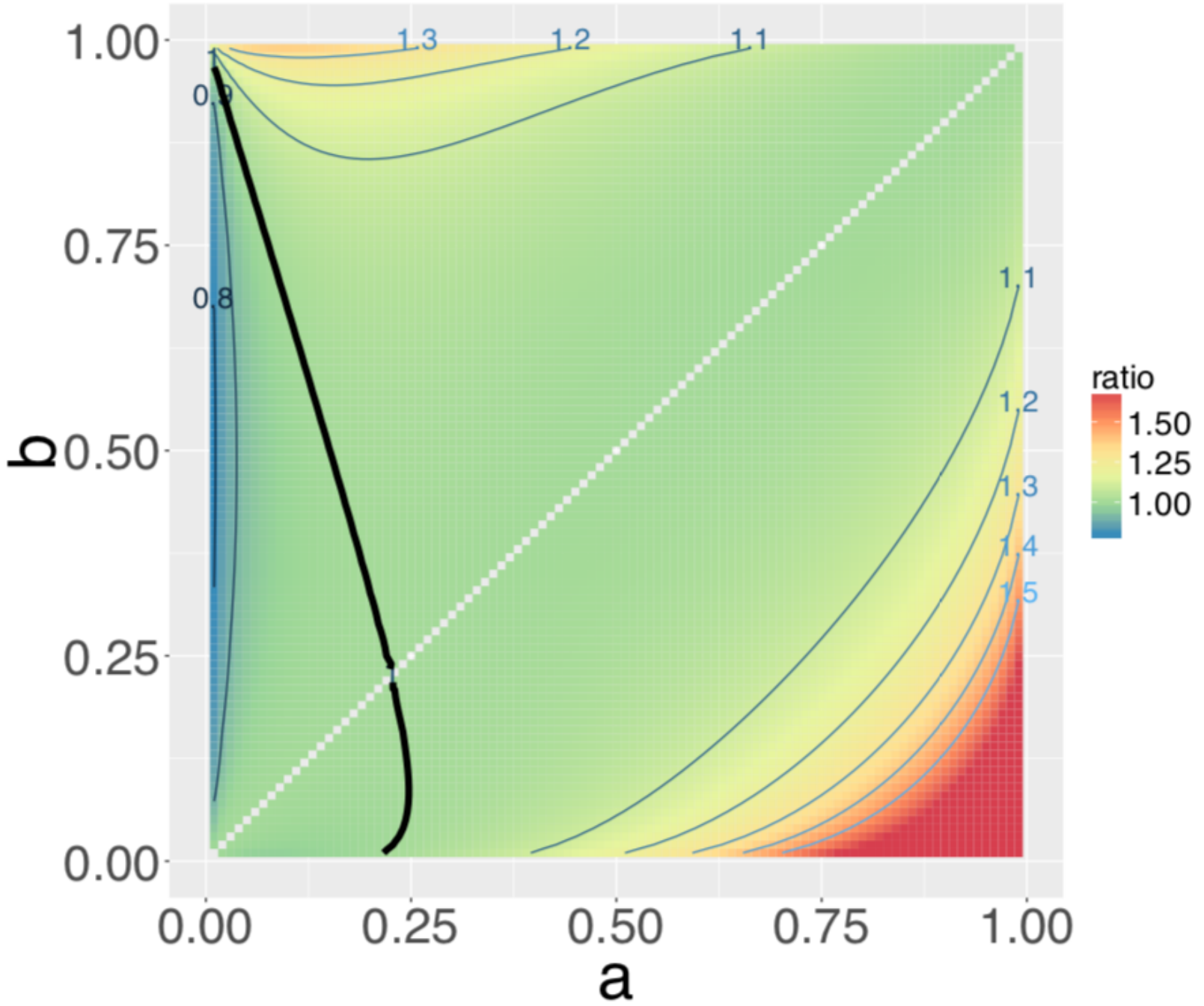
*Keywords:* Statistical network analysis; random graphs; stochastic block model; Laplacian spectral embedding; adjacency spectral embedding; Chernoff information; vertex clustering and classification

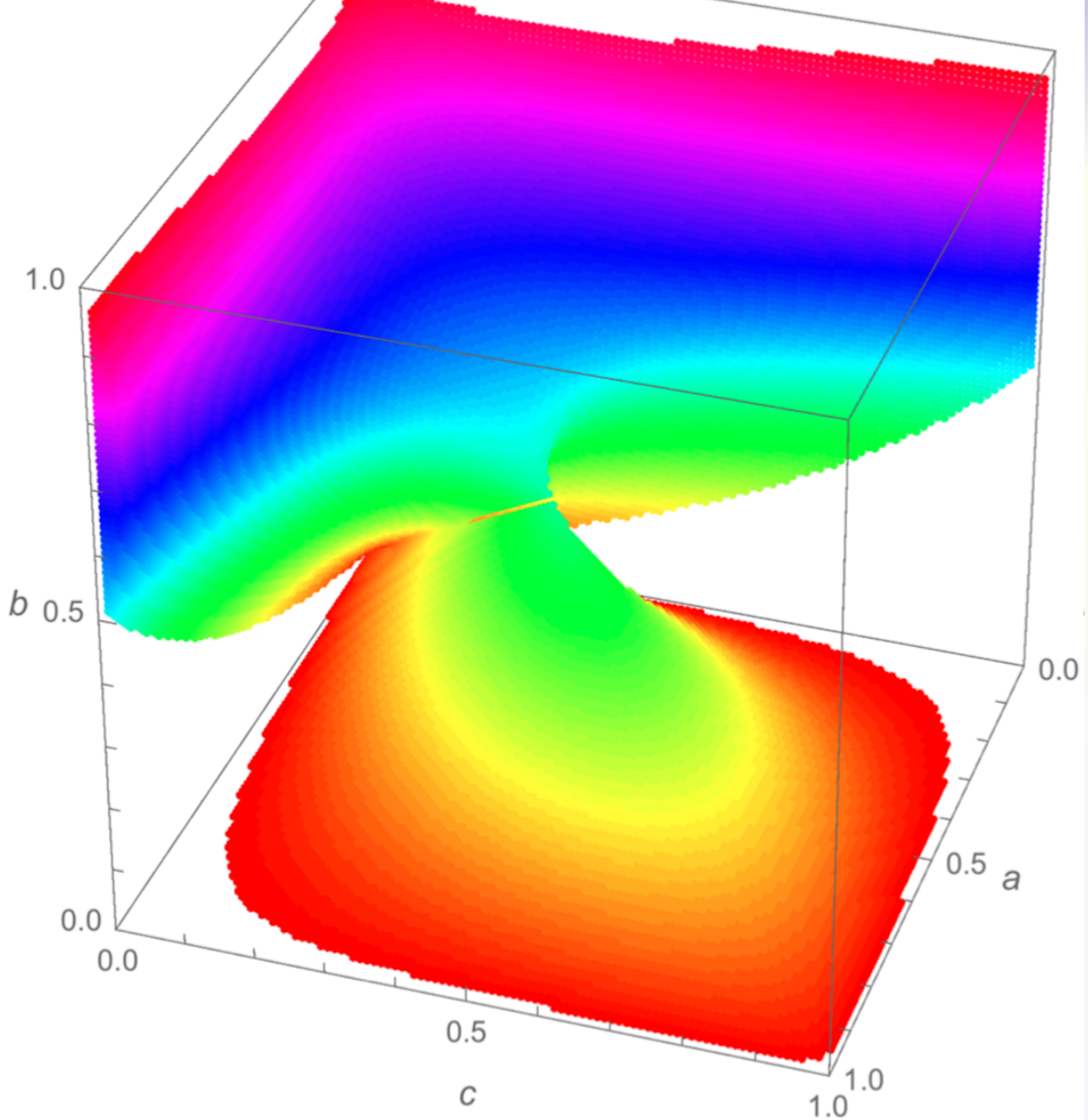


(a) The ratio $\rho^\star$ for $\mathbf{B} = \begin{bmatrix} a & b \\ b & b \end{bmatrix}$, $\boldsymbol{\pi} = (\frac{1}{2}, \frac{1}{2})$.

(b) The ratio $\rho^\star$ for $\mathbf{B} = \begin{bmatrix} a & b \\ b & b \end{bmatrix}$, $\boldsymbol{\pi} = (\frac{1}{4}, \frac{3}{4})$

Figure 1: Consider large $n$-vertex graphs from the $K$-block stochastic block model (SBM) with symmetric block edge probability matrix $\mathbf{B}$ and block probability vector $\boldsymbol{\pi}$ exhibiting block sizes $n_k = \pi_k n$ for each $k = 1, \ldots, K$. Using the concept of Chernoff information together with recent advances in random graph limit theory, we establish an information-theoretic summary statistic (ratio) $\rho^\star \equiv \rho^\star(\mathbf{B}, \boldsymbol{\pi})$ with the interpretation that the cases $\rho^\star > 1$, $\rho^\star < 1$, and $\rho^\star = 1$ correspond to comparative large-sample embedding performance summarized as ASE $>$ LSE, ASE $<$ LSE, and ASE $=$ LSE, respectively. For the collection of two-block SBMs exhibiting core-periphery structure with $\mathbf{B} \equiv \mathbf{B}(a, b)$ as specified in the above sub-captions, Figure 1(a) and Figure 1(b) show $\rho^\star$ evaluated over the parameter space $a, b \in (0, 1)$ in the balanced (block size) regime and an unbalanced regime, respectively. The empty diagonal depicts the Erdős-Rényi model singularity when $a = b$.


1.00
0.75
0.50
0.25
0.00
b
0.00
0.25
0.50
0.75
1.00
a
1.3
1.2
1.1
0.9
0.8
1.1
1.2
1.3
1.4
1.5
ratio
1.50
1.25
1.00




(a) $\rho^\star < 1$ for $\text{rank}(\boldsymbol{B}) = 2$ when $\boldsymbol{\pi} = (\frac{1}{2}, \frac{1}{2})$

# JMLR

## Journal of Machine Learning Research, 2018

### Statistical inference on random dot product graphs: a survey

Avanti Athreya, Donniell E. Fishkind, Keith Levin, Vince Lyzinski, Youngser Park, Yichen Qin, Daniel L. Sussman, Minh Tang, Joshua T. Vogelstein, Carey E. Priebe

*(Submitted on 16 Sep 2017)*

The random dot product graph (RDPG) is an independent-edge random graph that is analytically tractable and, simultaneously, either encompasses or can successfully approximate a wide range of random graphs, from relatively simple stochastic block models to complex latent position graphs. In this survey paper, we describe a comprehensive paradigm for statistical inference on random dot product graphs, a paradigm centered on spectral embeddings of adjacency and Laplacian matrices. We examine the analogues, in graph inference, of several canonical tenets of classical Euclidean inference: in particular, we summarize a body of existing results on the consistency and asymptotic normality of the adjacency and Laplacian spectral embeddings, and the role these spectral embeddings can play in the construction of single- and multi-sample hypothesis tests for graph data. We investigate several real-world applications, including community detection and classification in large social networks and the determination of functional and biologically relevant network properties from an exploratory data analysis of the Drosophila connectome. We outline requisite background and current open problems in spectral graph inference.

# Asymptotic Efficiency

## Asymptotically efficient estimators for stochastic blockmodels: the naive MLE, the rank-constrained MLE, and the spectral

Minh Tang, Joshua Cape, Carey E. Priebe

*(Submitted on 30 Oct 2017)*

We establish asymptotic normality results for estimation of the block probability matrix $\mathbf{B}$ in stochastic blockmodel graphs using spectral embedding when the average degrees grows at the rate of $\omega(\sqrt{n})$ in $n$, the number of vertices. As a corollary, we show that when $\mathbf{B}$ is of full-rank, estimates of $\mathbf{B}$ obtained from spectral embedding are asymptotically efficient. When $\mathbf{B}$ is singular the estimates obtained from spectral embedding can have smaller mean square error than those obtained from maximizing the log-likelihood under no rank assumption, and furthermore, can be almost as efficient as the true MLE that assume known $\mathrm{rk}(\mathbf{B})$. Our results indicate, in the context of stochastic blockmodel graphs, that spectral embedding is not just computationally tractable, but that the resulting estimates are also admissible, even when compared to the purportedly optimal but computationally intractable maximum likelihood estimation under no rank assumption.

# Generalized RDPG

Patrick Rubin-Delanchy, University of Bristol

## The generalised random dot product graph

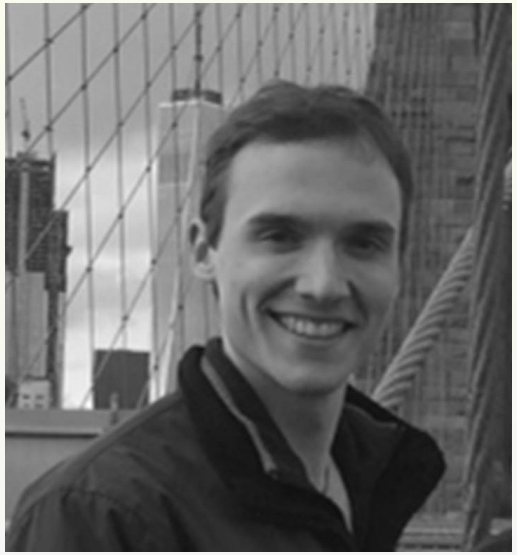
Patrick Rubin-Delanchy, Carey E. Priebe, Minh Tang

*(Submitted on 16 Sep 2017 (v1), last revised 21 Sep 2017 (this version, v2))*

This paper introduces a latent position network model, called the generalised random dot product graph, comprising as special cases the stochastic blockmodel, mixed membership stochastic blockmodel, and random dot product graph. In this model, nodes are represented as random vectors on $\mathbb{R}^d$, and the probability of an edge between nodes $i$ and $j$ is given by the bilinear form $X_i^T I_{p,q} X_j$, where $I_{p,q} = \mathrm{diag}(1, \ldots, 1, -1, \ldots, -1)$ with $p$ ones and $q$ minus ones, where $p + q = d$. As we show, this provides the only possible representation of nodes in $\mathbb{R}^d$ such that mixed membership is encoded as the corresponding convex combination of latent positions. The positions are identifiable only up to transformation in the indefinite orthogonal group $O(p, q)$, and we discuss some consequences for typical follow-on inference tasks, such as clustering and prediction.

# $2 \to \infty$

Our recent $2 \to \infty$ results precisely quantify spectral embedding estimation error for a broad class of random graph models while permitting heterogeneous, weakly dependent edge behavior.



Joshua Cape

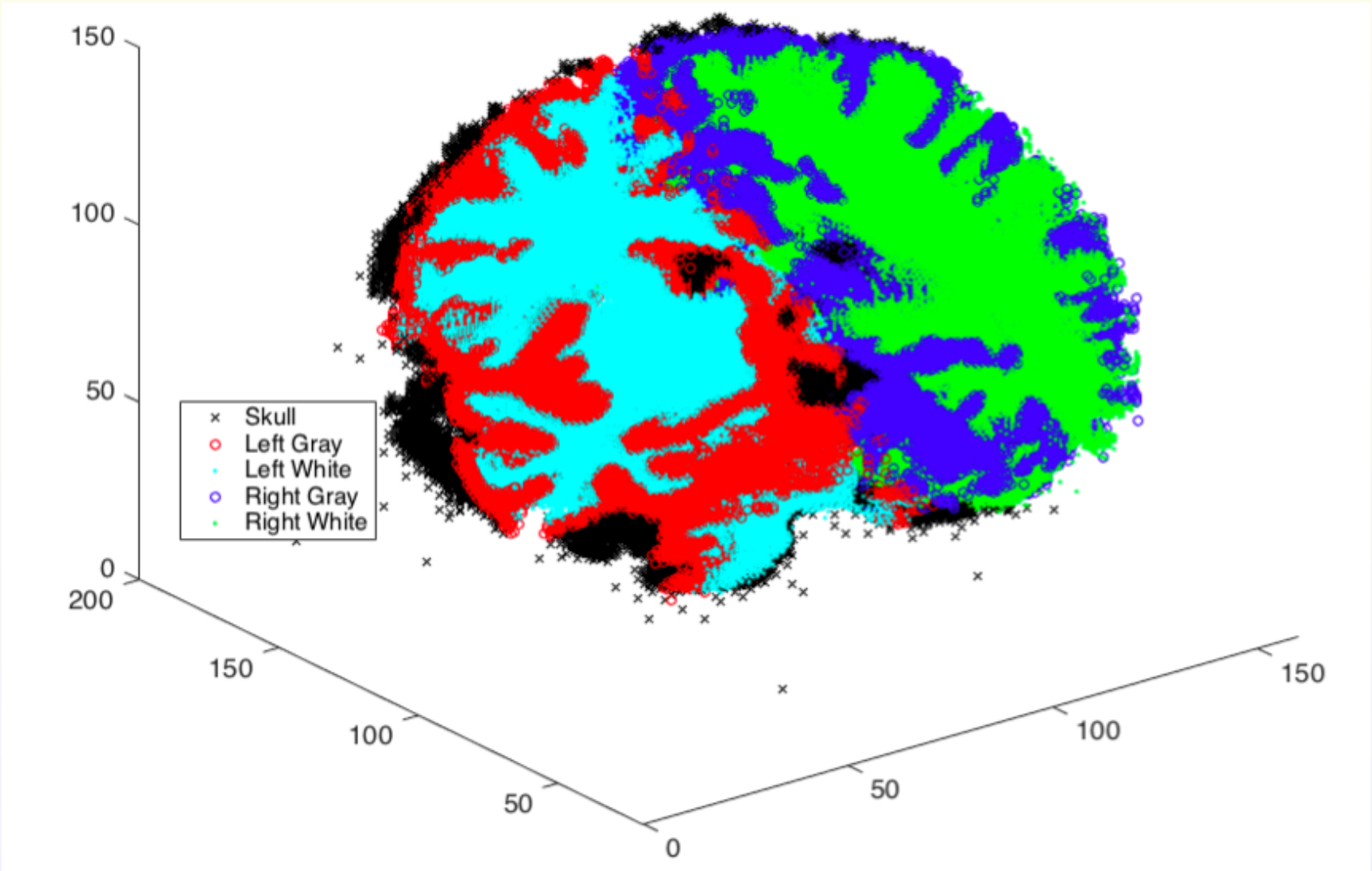
Joshua Cape, Minh Tang, CEP,
"The two-to-infinity norm and singular subspace geometry [...],"
http://arxiv.org/abs/1705.10735
*Annals of Statistics*, forthcoming

## Yogi Berra (purportedly):

*"In theory there is no difference between theory and practice. In practice, there is."*

(cf. *"That's all well and good in practice, but how does it work in theory?"*)

# Two Truths: Gray/White vs. Left/Right



# Our Connectomes I

Joshua Vogelstein and his team at Johns Hopkins University (special mention: Eric Bridgeford (JHU) & Greg Kiar (McGill))

have generated an exciting new connectome data set: multiresolution connectomes via a sequence of spatial vertex contractions with atlas annotation & tissue type.

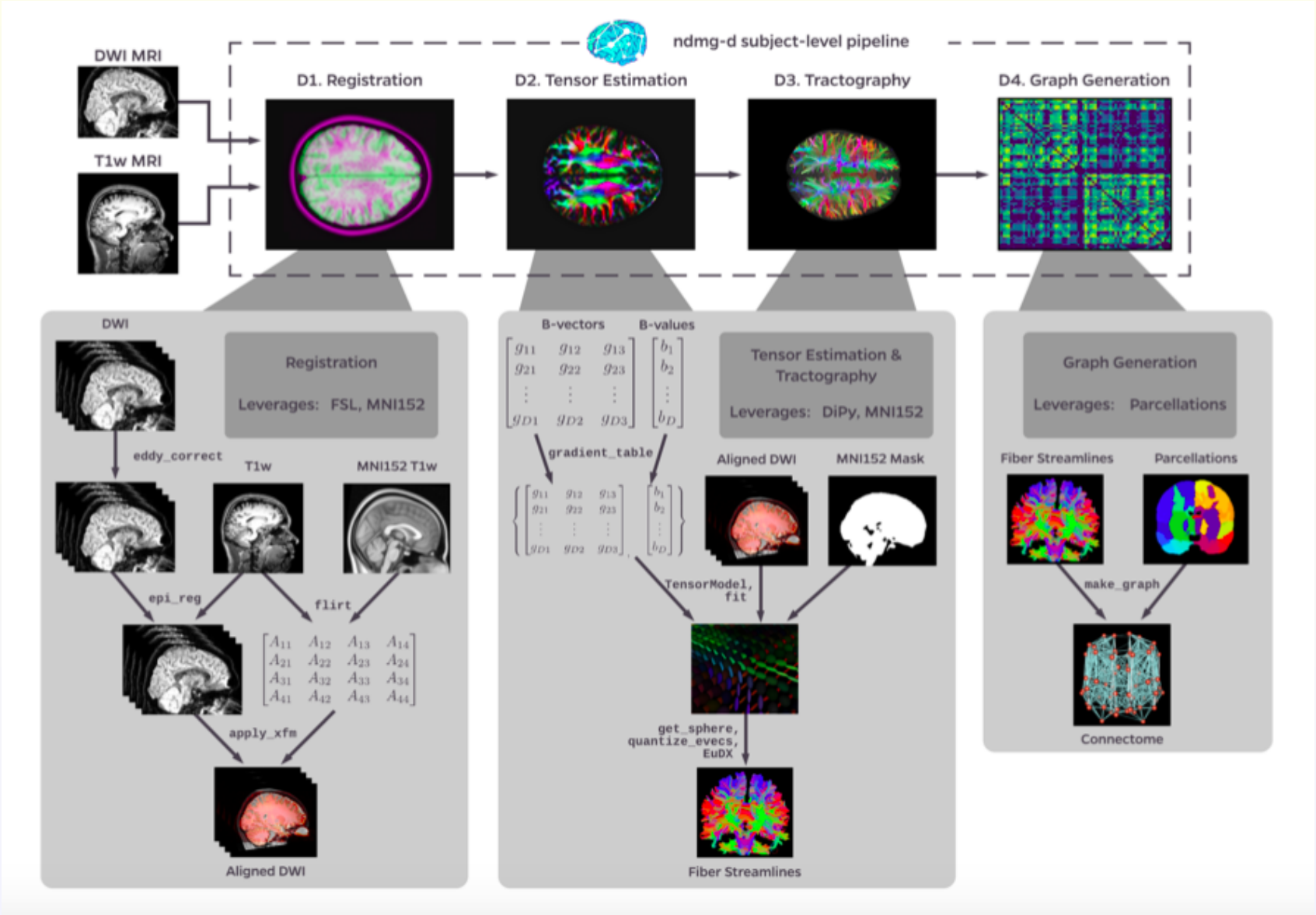
www.biorxiv.org/content/early/2018/03/20/188706

The subset we consider here includes 57 subjects, 2 scans each, dMRI with n $\approx$ 70K and Left/Right/x hemispheric & Gray/White/CSF/x tissue attributes for each vertex.

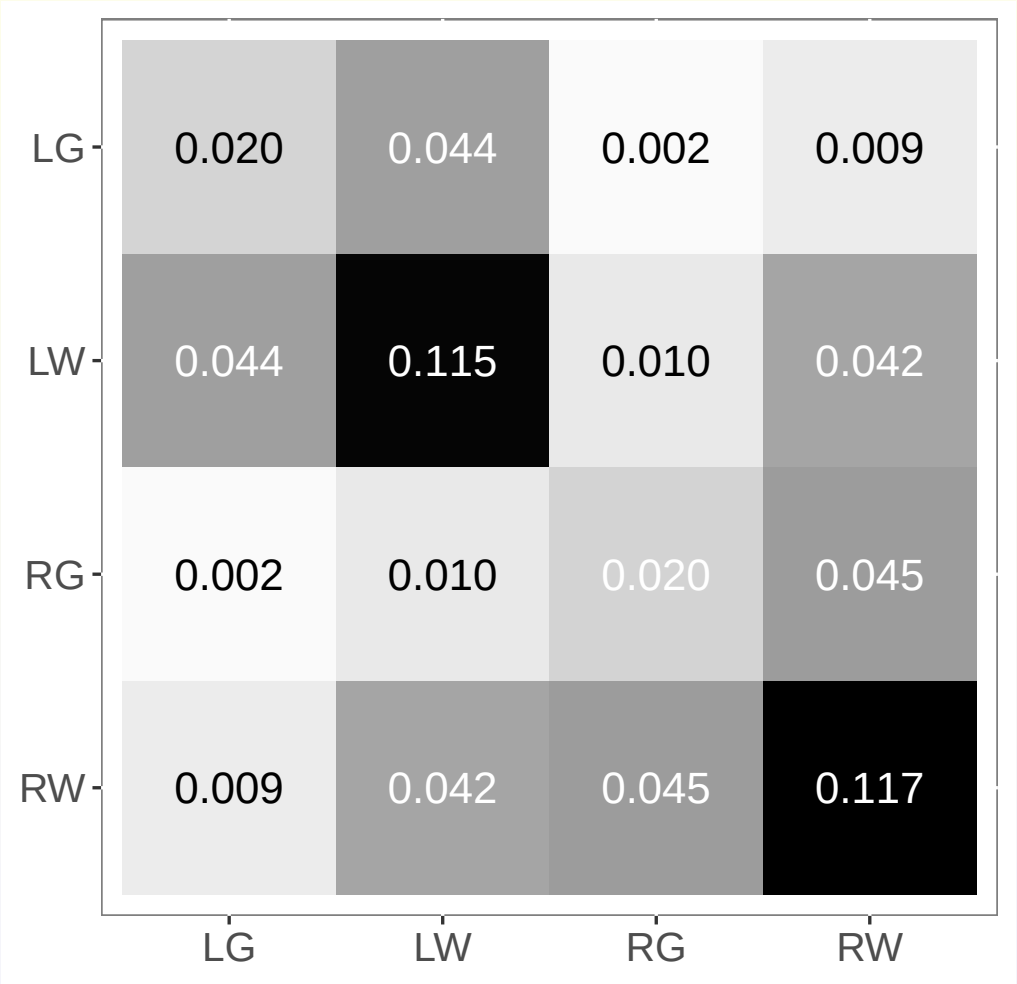
# Our Connectomes II

Two diffusion MRI (dMRI) and two structural MRI (sMRI) scans were done on an individual, collected over two sessions [59]. Graphs were estimated using the NDMG [59] pipeline. The dMRI scans were pre-processed for eddy currents using FSLs eddy-correct [3]. FSLs standard linear registration pipeline was used to register the sMRI and dMRI images to the MNI152 atlas [42, 54, 21, 33]. A tensor model was fit using DiPy [16] to obtain an estimated tensor at each voxel. A deterministic tractography algorithm was applied using DiPys EuDX [16, 15] to obtain a fiber streamline from each voxel. Graphs were formed by contracting fiber streamlines into sub-regions depending on spatial [35] proximity or neuro-anatomical [47, 9, 31, 26, 37, 19, 50, 43, 24] similarity.

# Our Connectomes III

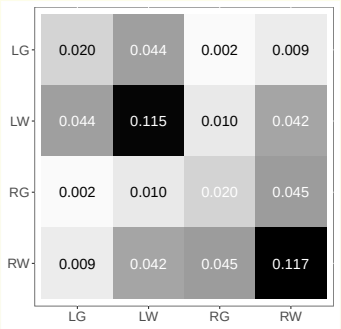


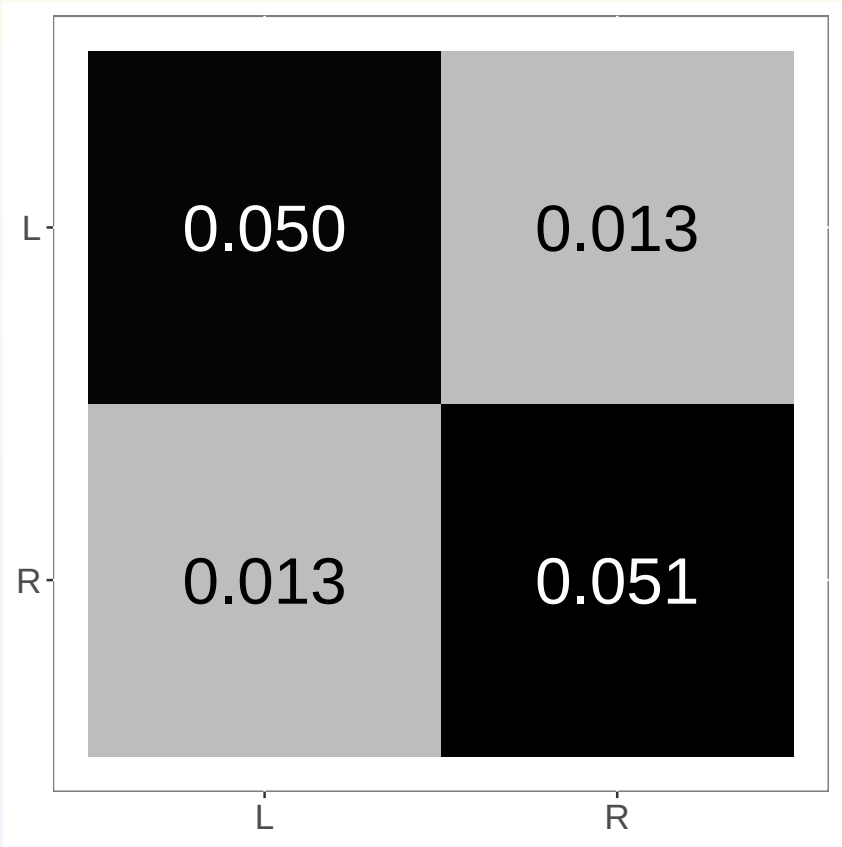
# Two Truths: Gray/White vs. Left/Right

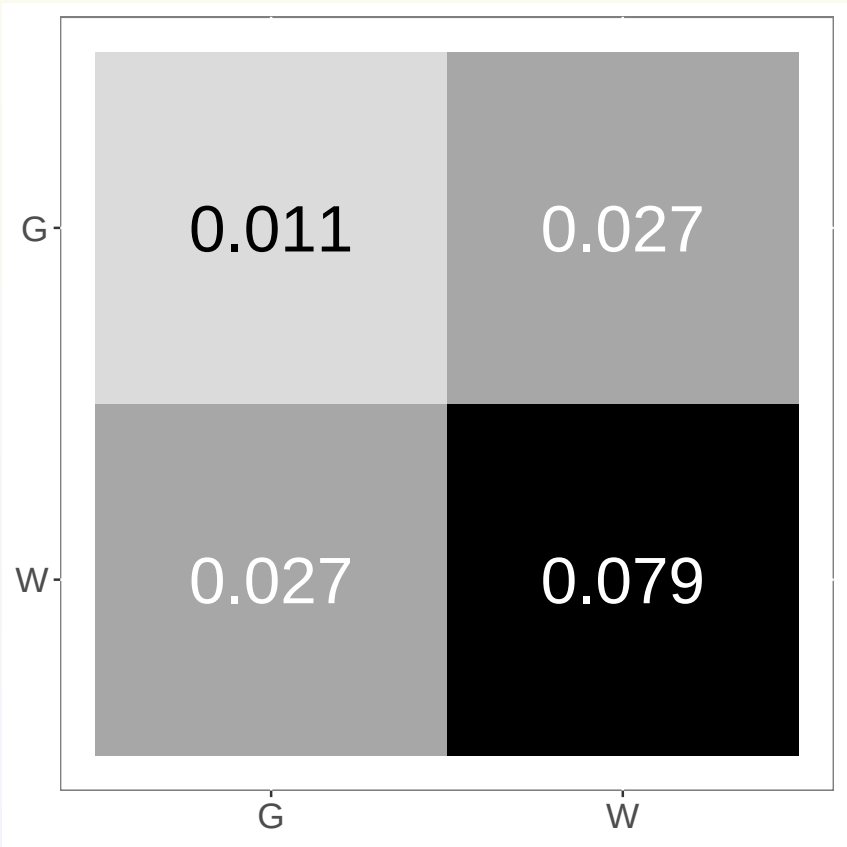



$$\pi = [n_{LG}, n_{LW}, n_{RG}, n_{RW}] = [0.279, 0.219, 0.282, 0.219]$$

# Two Truths: Gray/White vs. Left/Right




Left/Right $\approx$ Affinity

Gray/White $\approx$ Core-Periphery

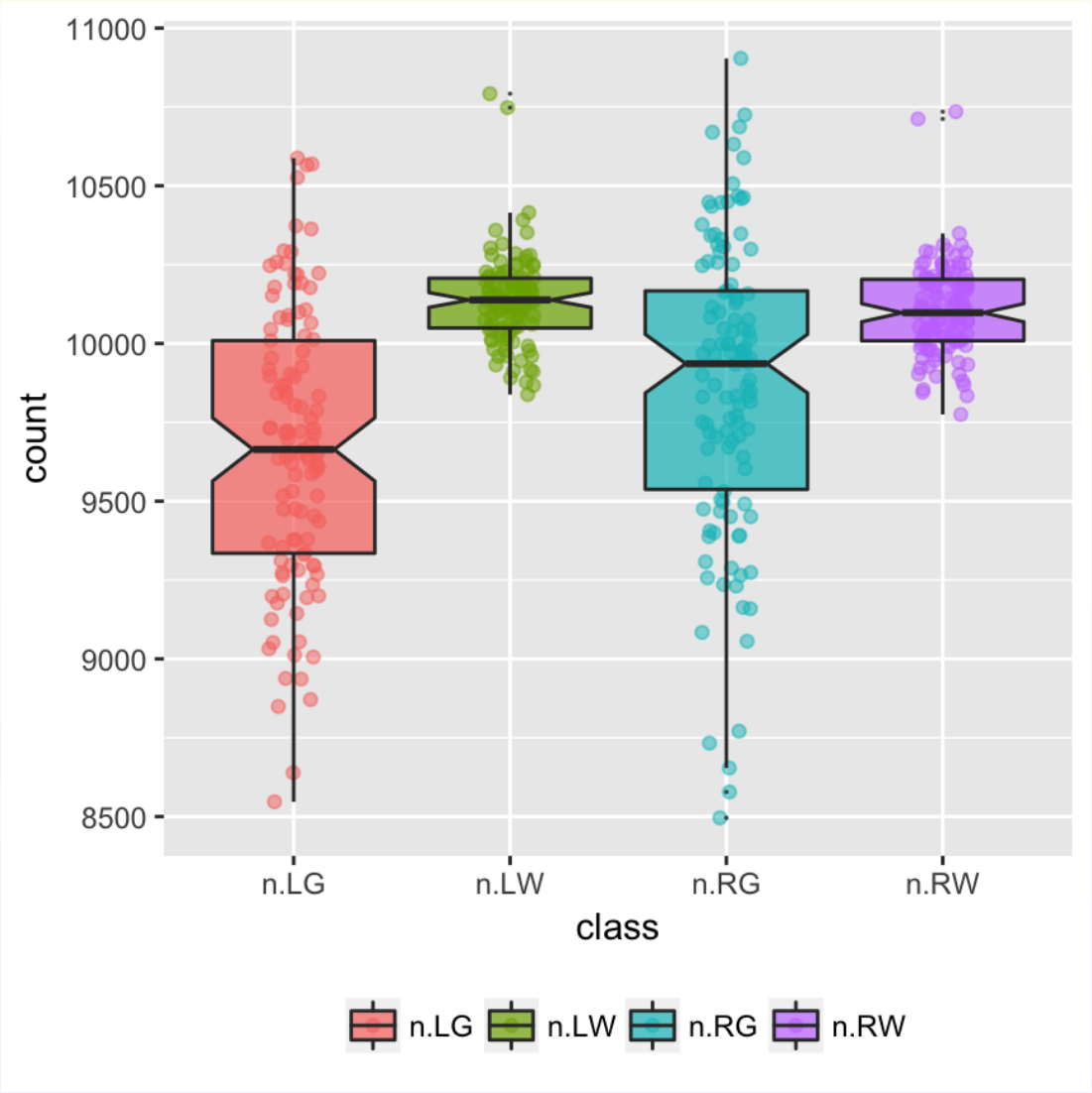
# Two Truths: ASE $\implies$ Gray/White ; LSE $\implies$ Left/Right (synthetic)

theory:
CLTs & Kullback-Leibler divergence shows that the
$(\widehat{d} = 2)$-dimensional embeddings of this $(K = 4)$-SBM,
when clustered via GMM into $\widehat{K} = 2$ clusters, will yield
{ {LG,LW} , {RG,RW} } for LSE
and
{ {LG,RG} , {LW,RW} } for ASE.

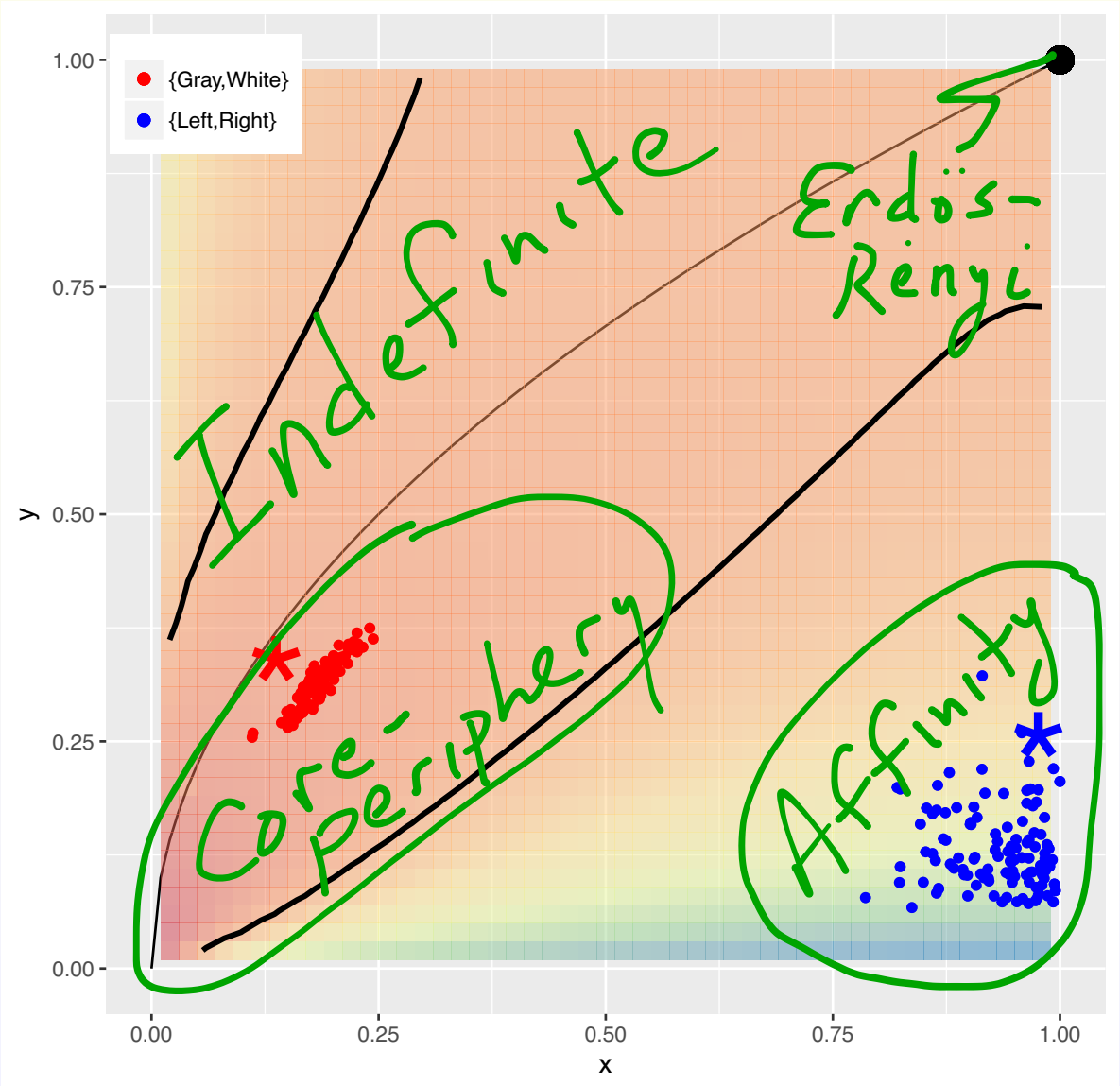
simulation:
$\widehat{d} = \widehat{K} = 2 \implies$
$P[ARI(GMM(LSE), LR) \approx 1] \approx 1]$
$P[ARI(GMM(LSE), GW) \approx 0] \approx 1]$
and
$P[ARI(GMM(ASE), LR) \approx 0] \approx 1]$
$P[ARI(GMM(ASE), GW) \approx 1] \approx 1]$

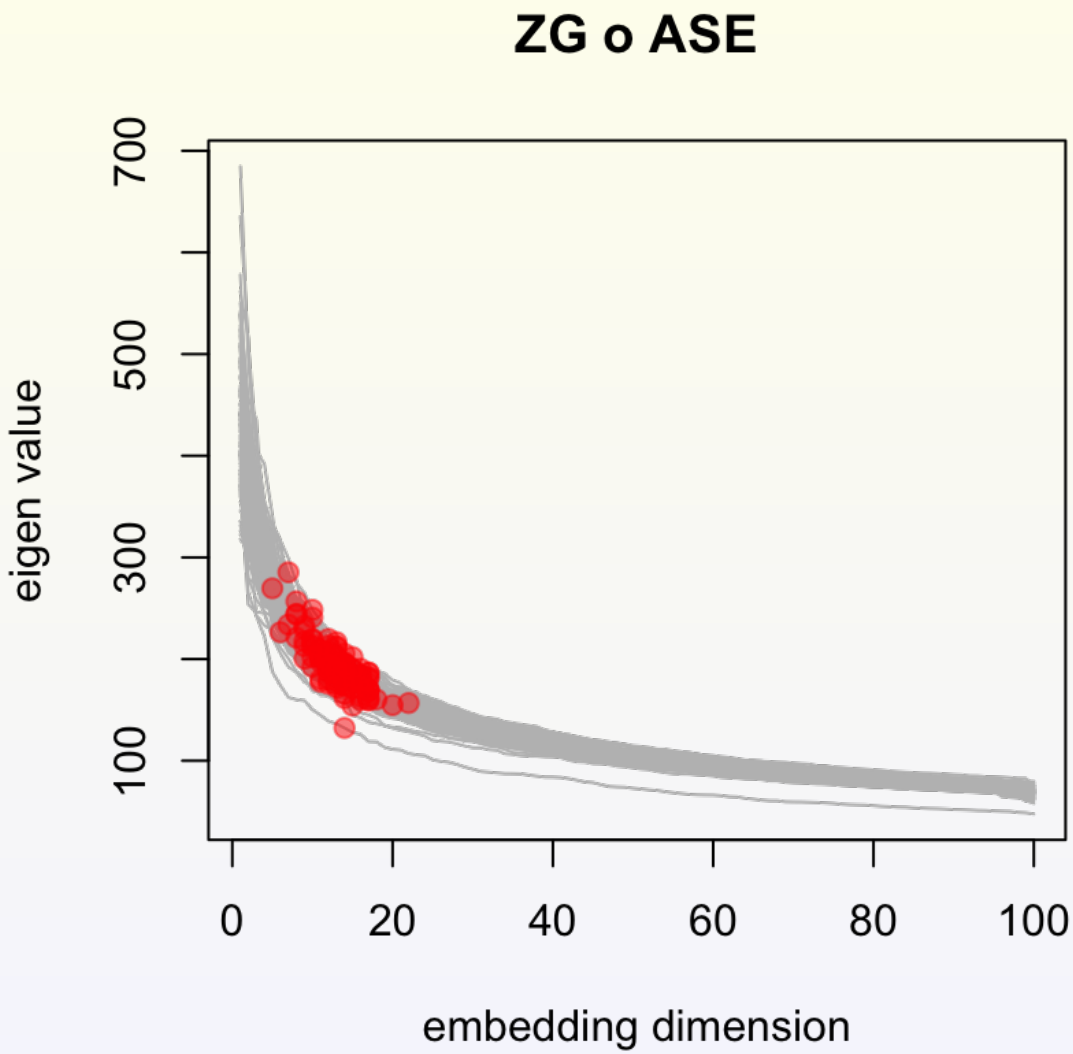
# back to our data ...

57 subjects, 2 scans each, dMRI with n $\approx$ 70K and Left/Right/x hemispheric & Gray/White/CSF/x tissue attributes for each vertex.

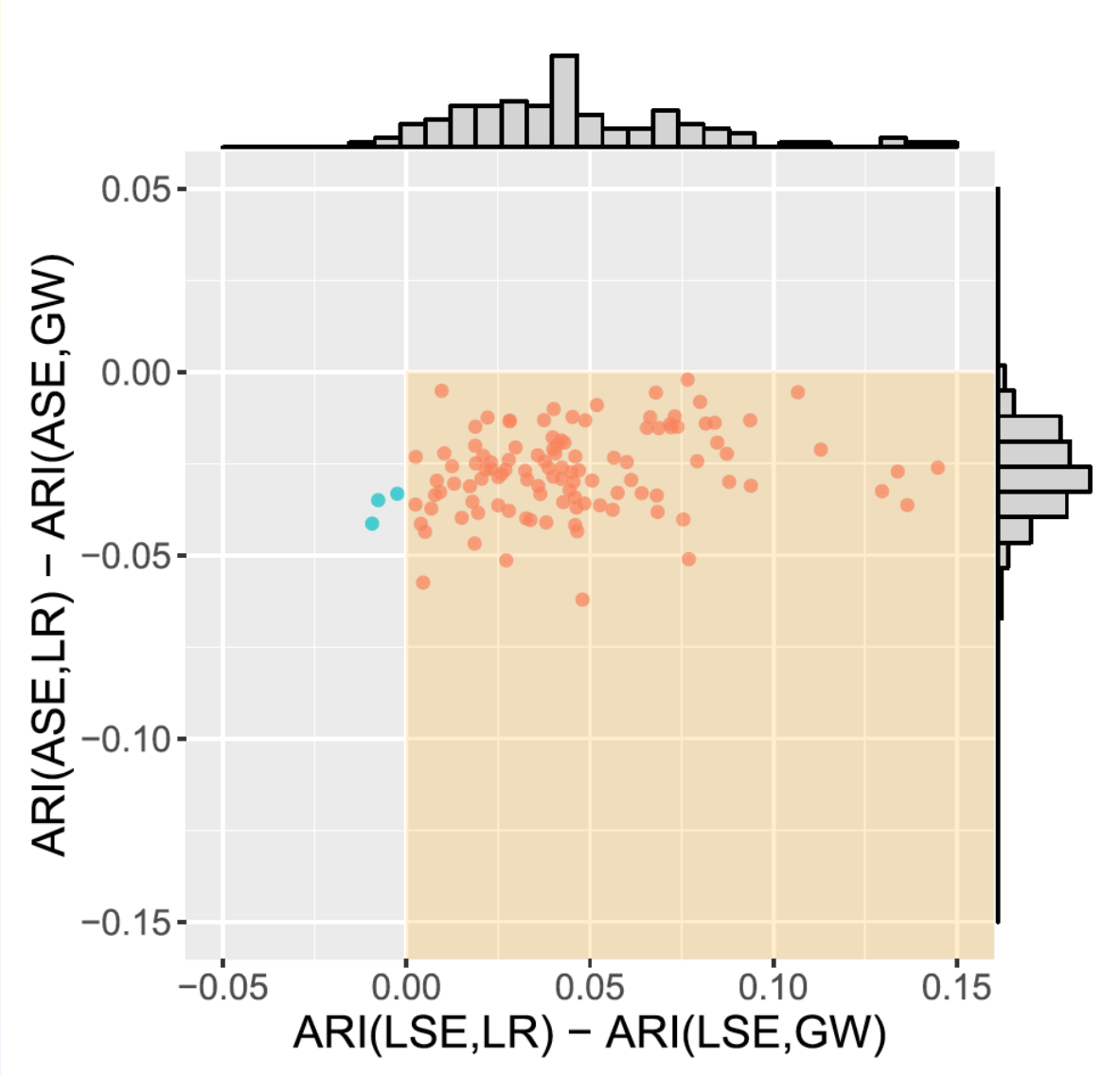


# Two Truths:
# G/W ≈ Core-Periphery ; L/R ≈ Affinity

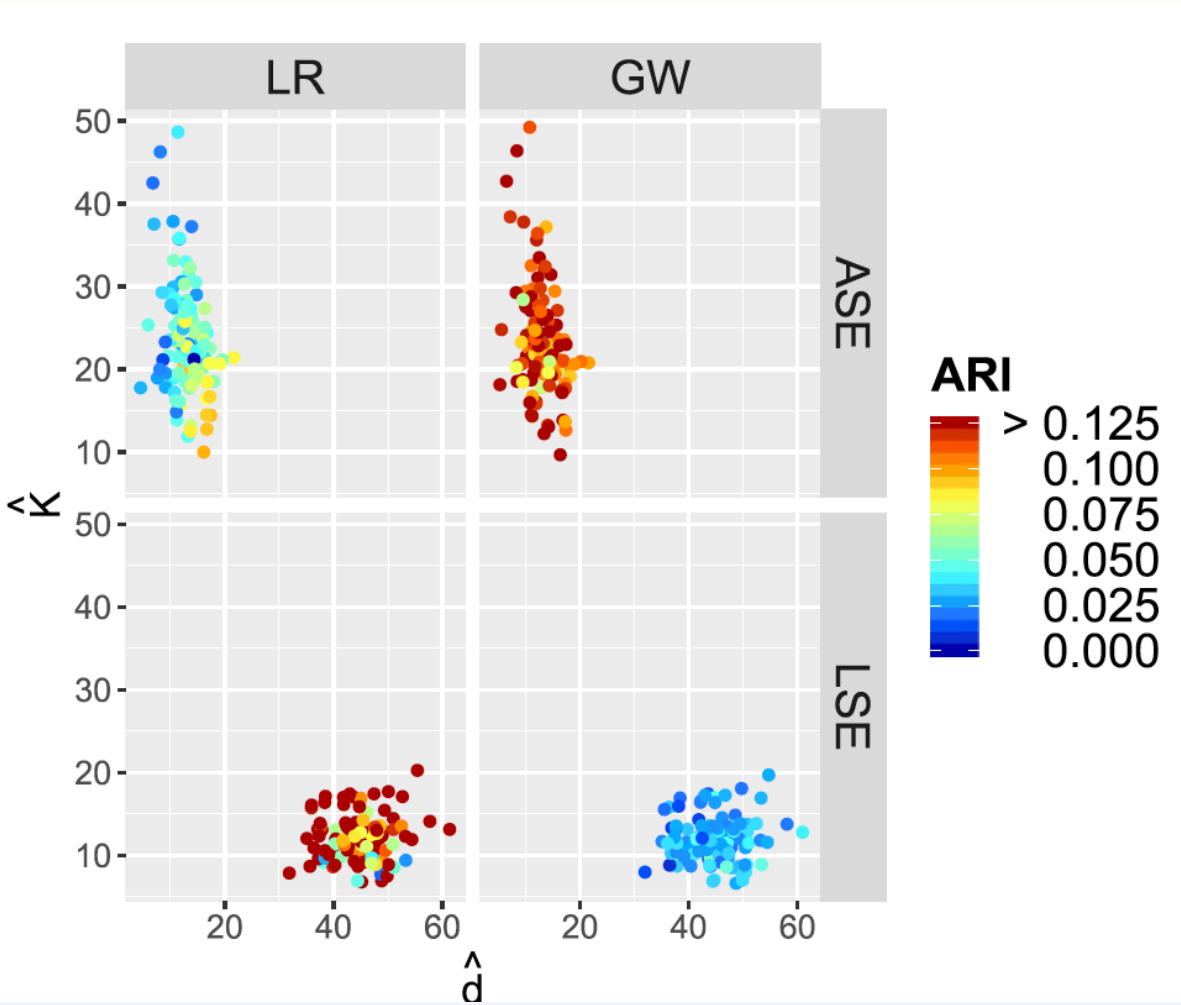



ZG o ASE
eigen value
700
500
300
100
0
20
40
60
80
100
embedding dimension

# Two Truths:
# ASE $\Longrightarrow$ Gray/White ; LSE $\Longrightarrow$ Left/Right



# Two Truths:
# ASE $\Longrightarrow$ Gray/White ; LSE $\Longrightarrow$ Left/Right



# Conclusions & Discussion

This illustrative example involves investigation of connectivity-based parcellations of the brain.

My great friend & mentor, JHU's very own Michael I Miller

once told me that the great Hopkins neuroscientist Vernon Mountcastle (the 'father of neuroscience')

once told him that **"to understand a neuron in cortex we must look to the neurons it is connected to."**

# Conclusions & Discussion

We end our Two Truths paper with the claim that our results show that the two methods ("LSE" & "ASE") capture different types of brain structure, suggesting that a comprehensive connectivity-based parcellation based on the powerful and popular mathematical approach of spectral clustering should combine the two competing approaches as these two spectral embedding approaches facilitate the identification of different and complementary connectivity-based clustering truths.

# Conclusions & Discussion

Neither $\mathrm{GMM} \circ \mathrm{ASE}$ nor $\mathrm{GMM} \circ \mathrm{LSE}$ dominates the other for subsequent inference . . .
and $\mathrm{K\text{-}means}$ is inferior to $\mathrm{GMM}$ for spectral clustering.

- Long-sought LSE CLT – in particular, LSE(SBM) $\sim$ GMM.
- LSE CLT, together with ASE CLT, allows Chernoff comparison.

- Two Truths: LSE likes Affinity ; ASE likes Core-Periphery.
- Two Truths: LSE likes Left-Right ; ASE likes Gray-White.

These results suggest that a connectivity-based parcellation based on spectral clustering should consider **both** LSE & ASE.

- regularized?
- $d' < d$? $d' > d$? $d_n \nearrow \infty$?
- omni?
- etc!

## Leopold Kronecker to Hermann von Helmholtz (1888):

*"The wealth of your practical experience*
*with sane and interesting problems*
*will give to mathematics*
*a new direction and a new impetus."*


Kronecker

Helmholtz