Asymptotic behavior of the differences between latent positions and their estimates in RDPG models

Konstantinos Pantazis

STAT705 Presentation

Dec 2nd 2019

Table of Contents

Table of Contents

Random Dot Product Graph

Definition (Random Dot Product Graph(RDPG)[1])

Let $X = [X_1^T \cdots X_n^T]^T \in \mathbb{R}^{n \times d}$, where the rows X_i of X are independent random variables following a distribution F(F is a distribution on a set $\mathscr{X} \subset \mathbb{R}^d$ satisfying $\langle x, x' \rangle \in [0,1]$ for all $x, x' \in \mathscr{X}$. We say $(X, A) \sim$RDPG(F, n) if the matrix $A \in \{0,1\}^{n \times n}$ is defined to be a symmetric, hollow matrix such that for all $i < j$, conditioned on X_i, X_j,

$$A_{i,j} \sim^{ind} Bern(X_i^T X_j).$$

Remark

We say that $A \in \mathbb{R}^{n \times n}$ is the adjacency matrix of a random dot product graph with latent positions given by the rows of X, $X_i \in \mathbb{R}^d$.

Definition (Stochastic Block Model[4])

The stochastic block model(SBM) is a generative model for random graphs. This model tends to produce graphs containing communities, subsets characterized by being connected with one another with particular edge densities. The stochastic block model takes the following parameters:

- The number n of vertices;
- A partition of the vertex set $\{1, 2, \cdots, n\}$ into disjoint subsets $C_1, C_2, \cdots, C_K$ called communities;
- A symmetric $K \times K$ matrix P of edge probabilities.

The edge set is then sampled at random as follows: any two vertices $u \in C_i$ and $v \in C_j$ are connected by an edge with probability P_{ij}.

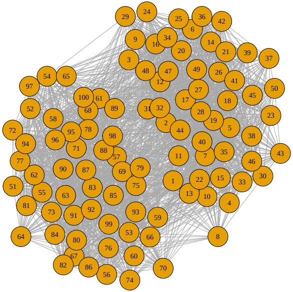
Remark

Stochastic Blockmodels(SBM) *are a special case of RDPG models, where* $d = K$ *represent the number of communities.*

Example SBM

Consider a random graph generated using SBM, with $n = 100$ vertices, $K = 2$ communities and probability matrix

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}.$$



Adjacency Spectral Embedding [3]

Definition

Let $A \in \mathbb{R}^{n\times n}$ be the adjacency matrix of an undirected d-dimensional RDPG. The d-**dimensional adjacency spectral embedding** (ASE) of A is a spectral decomposition of A based on its top d eigenvalues, obtained by $ASE(A,d) = U_A S_A^{1/2}$, where $S_A \in \mathbb{R}^{d\times d}$ is a diagonal matrix whose entries are the top eigenvalues of A(in nondecreasing order) and $U_A \in \mathbb{R}^{n\times d}$ is the matrix whose columns are orthonormal eigenvectors corresponding to the eigenvalues in S_A.

Remark

The rows of the matrix $U_A S_A^{1/2}$ are the estimated latent positions of the true latent positions X_i.

Joint RDPG [2]

Definition

We say that random graphs $A^{(1)}, A^{(2)} \cdots, A^{(m)}$ are distributed as a *joint random dot product graph* (JRDPG) if $(A^{(k)}, X) \sim \text{RDPG}(F, n)$ for each $k = 1, 2, \cdots, m$. That is, the $A^{(k)}$ are conditionally independent given X, with edges independently distributed as $A^{(k)}_{i,j} \sim Bernoulli((XX^T)_{ij})$ for all $1 \leq i < j \leq n$ and all $k \in [m]$.

Omnibus Matrix

Let $A^{(1)}, A^{(2)}, \cdots, A^{(m)} \in \mathbb{R}^{n \times n}$ be adjacency matrices of a collection of m undirected graphs. We define the mn-by-mn omnibus matrix of $A^{(1)}, A^{(2)}, \cdots, A^{(m)}$ by

$$M = \begin{bmatrix} A^{(1)} & \frac{A^{(1)}+A^{(2)}}{2} & \frac{A^{(1)}+A^{(3)}}{2} & \cdots & \frac{A^{(1)}+A^{(m)}}{2} \\ \frac{A^{(2)}+A^{(1)}}{2} & A^{(2)} & \frac{A^{(2)}+A^{(3)}}{2} & \cdots & \frac{A^{(2)}+A^{(m)}}{2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{A^{(m)}+A^{(1)}}{2} & \frac{A^{(m)}+A^{(2)}}{2} & \frac{A^{(m)}+A^{(3)}}{2} & \cdots & A^{(m)} \end{bmatrix},$$

and the d-dimensional *omnibus embedding* of $A^{(1)}, A^{(2)}, \cdots, A^{(m)}$ is the adjacency spectral embedding of M:

$$OMNI(A^{(1)}, A^{(2)}, \cdots, A^{(m)}, d) = ASE(M, d).$$

Table of Contents

Main Theorem

Theorem

Let $(A^{(1)}, A^{(2)}, \cdots, A^{(m)}, X) \sim JRDPG(F, n, m)$ *and let* M *denote the omnibus matrix. Let* $Z = [X^T, X^T, \cdots, X^T]^T \in \mathbb{R}^{mn \times d}$ *and let* $\hat{Z} = ASE(M, d)$ *its estimate(of the mn latent positions collected in the matrix Z). Let* $h = n(s-1)+i$ *for* $i \in [n], s \in [m]$*, so that* $\hat{Z}_h$ *denotes the estimated latent position of the i-th vertex in the s-th graph* $A^{(s)}$*. Let* $\Phi(x, \Sigma)$ *denote the cdf of a (multivariate) Gaussian with mean zero and covariance matrix* Σ*, evaluated at* $x \in \mathbb{R}^d$*. There exists a sequence of orthogonal d-by-d matrices* $(W_n)_{n=1}^{\infty}$ *such that for all* $x \in \mathbb{R}^d$*,*

$$\lim_{n\to\infty} \left\{ \sqrt{n}(\hat{Z}W_n - Z)_h \le x \right\} \to \int_{supp\ F} \Phi(x, \Sigma(y)) dF(y),$$

where $\Sigma(y) = (m+3)\Delta^{-1}\tilde{\Sigma}(y)\Delta^{-1}/4m$*,* $\Delta = [X_1 X_1^T] \in \mathbb{R}^{d\times d}$ *and* $\tilde{\Sigma}(y) = [(y^T X_1 - (y^T X_1)^2) X_1 X_1^T]$*.*

Main Theorem cont'd

Remark

The above theorem shows that the scaled differences between the estimated and the true latent positions asymptotically follow a mixture of multivariable normals with mean 0 *and covariance matrix* Σ.

Extension

Now, let the off-diagonal entries of the omnibus matrix M to be $\frac{c_i A_i + c_j A_j}{c_i + c_j}$, where $c_i, c_j \in (0,1)$.

$$M^{(new)} = \begin{bmatrix} A^{(1)} & \frac{c_1 A^{(1)} + c_2 A^{(2)}}{c_1 + c_2} & \frac{c_1 A^{(1)} + c_3 A^{(3)}}{c_1 + c_3} & \dots & \frac{c_1 A^{(1)} + c_m A^{(m)}}{c_1 + c_m} \\ \frac{c_2 A^{(2)} + c_1 A^{(1)}}{c_2 + c_1} & A^{(2)} & \frac{c_2 A^{(2)} + c_3 A^{(3)}}{c_2 + c_3} & \dots & \frac{c_2 A^{(2)} + c_m A^{(m)}}{c_2 + c_m} \\ \frac{c_3 A^{(3)} + c_1 A^{(1)}}{c_3 + c_1} & \frac{c_3 A^{(3)} + c_2 A^{(2)}}{c_3 + c_2} & A^{(3)} & \dots & \frac{c_3 A^{(3)} + c_m A^{(m)}}{c_3 + c_m} \\ \vdots & \vdots & \ddots & \vdots & \\ \frac{c_m A^{(m)} + c_1 A^{(1)}}{c_m + c_1} & \frac{c_m A^{(m)} + c_2 A^{(2)}}{c_m + c_2} & \frac{c_m A^{(m)} + c_3 A^{(3)}}{c_m + c_3} & \dots & A^{(m)} \end{bmatrix}$$

GOAL

Our goal is to prove a similar asymptotic result as in the previous theorem with the more general matrix $M^{(new)}$, and moreover, instead of taking the differences between the estimated and the true latent positions, we look after the differences between estimated latent positions with each other, assuming both have the same true latent position.

Conjecture

Let $(A^{(1)}, A^{(2)}, \cdots, A^{(m)}, X) \sim \text{JRDPG}(F, n, m)$ and let $M^{(new)}$ denote the omnibus matrix above. Let $h_1 = n(s_1 - 1) + i, h_2 = n(s_2 - 1) + i$ for $i \in [n], s_1, s_2 \in [m]$, so that $\widehat{M}_{h_1}^{(new)}, \widehat{M}_{h_2}^{(new)}$ denote the estimated latent positions of the i-th vertex in the s_1, s_2-th graphs $A^{(s_1)}, A^{(s_2)}$ respectively. Let $\Phi(x, \Sigma)$ denote the cdf of a (multivariate) Gaussian with mean zero and covariance matrix Σ, evaluated at $x \in \mathbb{R}^d$. Then,

$$\lim_{n\to\infty} \left\{ \sqrt{n} (\widehat{M}_{h_1}^{(new)} - \widehat{M}_{h_2}^{(new)}) \le x \right\} \to \int_{\text{supp F}} \Phi(x, \Sigma(y)) dF(y),$$

where $\Sigma(y) = \frac{2}{m^2} \Delta^{-1} \widetilde{\Sigma}(y) \Delta^{-1}$, $\Delta = [X_1 X_1^T] \in \mathbb{R}^{d \times d}$ and

$$\widetilde{\Sigma}(y) = \frac{2}{m^2}[X_j X_j^T (y^T X_j - (y^T X_j)^2)]\Big[1+$$
$$\sum_{k=3}^{m} \frac{c_1}{c_1+c_k} + \frac{c_2}{c_2+c_k} + \frac{c_1^2}{(c_1+c_k)^2} + \frac{c_2^2}{(c_2+c_k)^2} - \frac{c_1 c_2 (c_1 c_2 + c_1 c_k + c_2 c_k + c_k^2)}{(c_1+c_k)^2 (c_2+c_k)^2} -$$
$$\sum_{k<l, k\geq 3} \frac{c_1^2 c_2 (c_2 + c_l + c_k) + c_1 c_2^2 (c_1 + c_l + c_k) + c_k c_l (c_1^2 + c_2^2)}{(c_1+c_k)(c_2+c_k)(c_1+c_l)(c_2+c_l)}\Big].$$

Remark

The choice of c_i's has to be such that the covariance matrix Σ has positive entries.

Table of Contents

Description

For the simulations, for each $N \in \{300, 310, \cdots, 800\}$ number of vertices, we generate $m = 3$ random graphs according to the SBM model with $K = 2$ communities, with their corresponding adjacency matrices

- $A^{(1)}, A^{(2)}, A^{(3)} \sim SBM(N, 2, P)$
- Moreover, the probability block matrix $P \in \mathbb{R}^{2\times 2}$ is given by

$$P = \begin{bmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{bmatrix}$$

Cont'd

The first experiment investigates whether the asymptotic normality of the row differences hold for different coefficients $C_i = (c_1, c_2, c_3)$.

Specifically, we take the differences of the embedded omnibus matrix $\widehat{M}^{(new)}$ between the 1st,(N+1)st and (2N+1)st rows.

- 12 corresponds to $\widehat{M}_1^{(new)} - \widehat{M}_{N+1}^{(new)}$
- 13 corresponds to $\widehat{M}_1^{(new)} - \widehat{M}_{2N+1}^{(new)}$
- 23 corresponds to $\widehat{M}_{N+1}^{(new)} - \widehat{M}_{2N+1}^{(new)}$

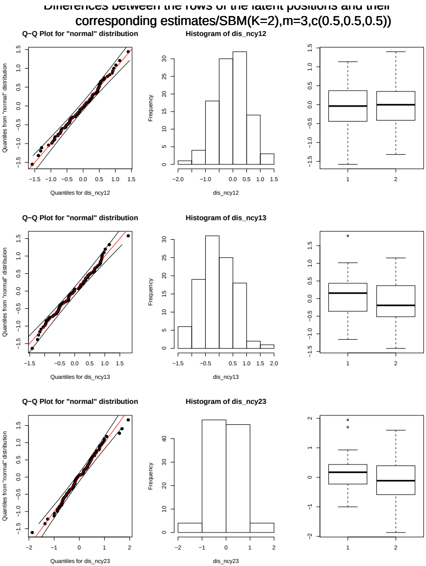


Figure 1: $C_1 = (0.5, 0.5, 0.5)$, corresponds to the initial omnibus matrix. Shapiro-Wilkinson test for 12: 0.94, for 13: 0.67, for 23: 0.54

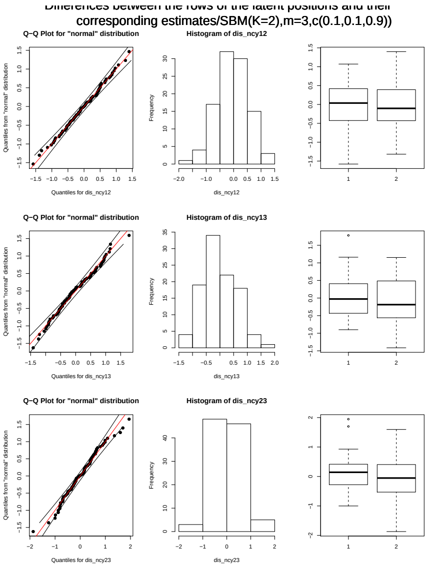


Figure 2: $C_2 = (0.1, 0.1, 0.9)$
Shapiro-Wilkinson test for 12: 0.99, for 13:0.45, for 23: 0.28

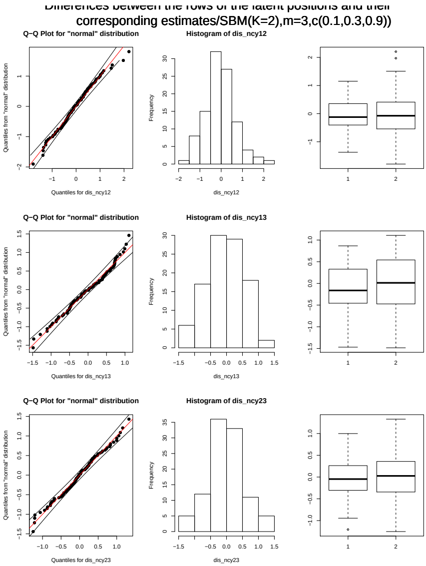


Figure 3: $C_3 = (0.1, 0.3, 0.9)$
Shapiro-Wilkinson test for 12: 0.44, for 13:0.28, for 23: 0.54

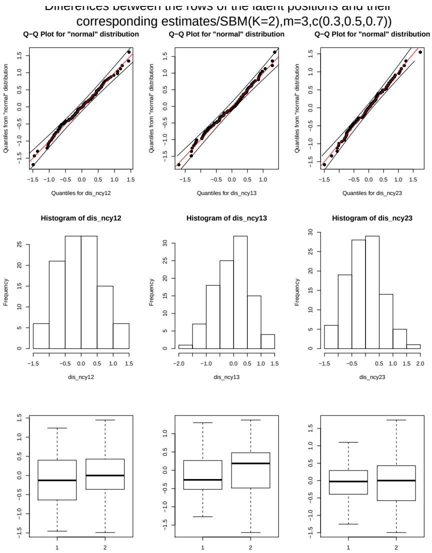


Figure 4: $C_4 = (0.3, 0.5, 0.7)$

Shapiro-Wilkinson test for 12: 0.816, for 13:0.64, for 23: 0.97

Not asymptotic normality example

Choosing the probability block matrix P to be:

$$P = \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{bmatrix},$$

the row differences do not converge to a mixture of multivariate normals. That is because an eigengap assumption is violated, i.e., the smallest eigenvalue of P is too small.

Differences between the rows of the latent positions and their corresponding estimates/SBM(K=2),m=3,c(0.5,0.5,0.5))

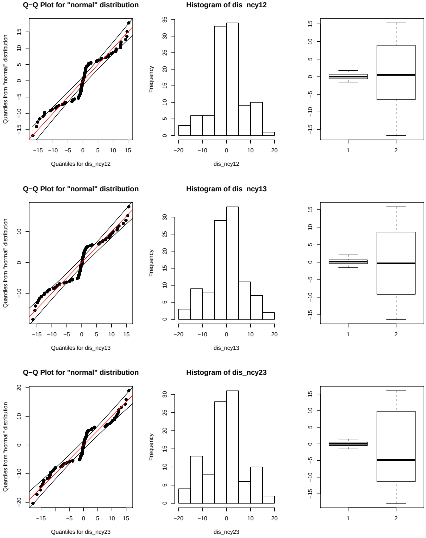


Figure 5: $C_5 = (0.5, 0.5, 0.5)$
Shapiro-Wilkinson test for 12,13,23 $\approx 10e-6 \ll 0.05$.

Simulations Cont'd

The second set of figures illustrates the estimated latent positions in $\mathbb{R}^2$ for different C_i's, and the ellipse corresponds to the theoretical covariance matrix Σ.

We seek to determine how well the estimated latent positions from the simulations fit to the theoritical results.

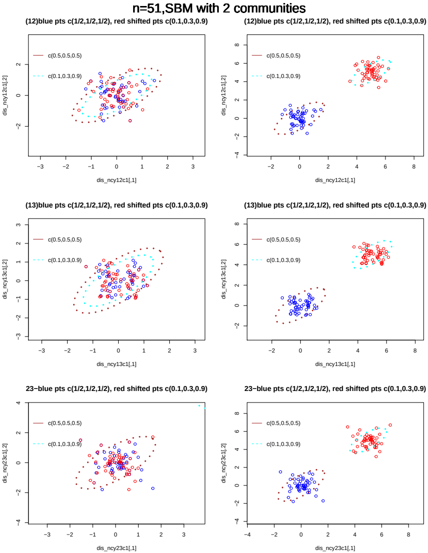


Figure 6: We compare the residuals of the estimated latent positions of $\widehat{M}^{(1)}, \widehat{M}^{(2)}$ computed using the coefficients $C_1 = (0.5, 0.5, 0.5), C_2 = (0.1, 0.3, 0.9)$ respectively.

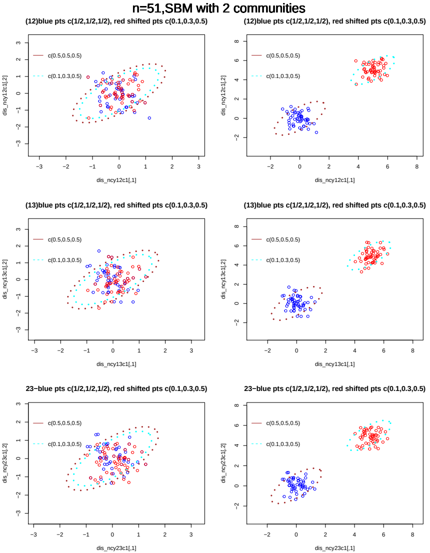
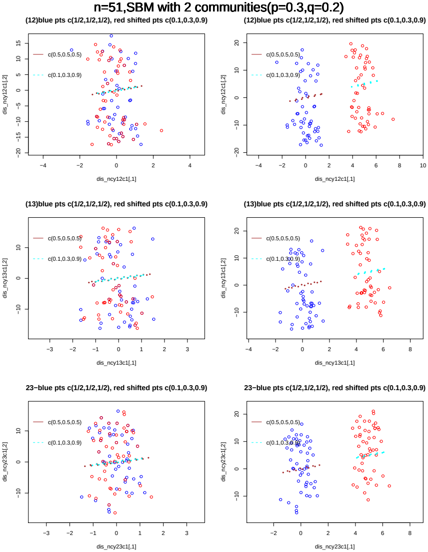


Figure 7: We compare the residuals of the estimated latent positions of $\widehat{M}^{(1)}, \widehat{M}^{(2)}$ computed using the coefficients $C_1 = (0.5, 0.5, 0.5), C_2 = (0.1, 0.3, 0.5)$ respectively.

Not asymptotic normality example



A. Athreya, C. E. Priebe, M. Tang, V. Lyzinski, D. J. Marchette, and D. L. Sussman.
A limit theorem for scaled eigenvectors of random dot product graphs.
Sankhya A, pages 1–18, 2013.

Keith Levin, Avanti Athreya, Minh Tang, Vince Lyzinski, Youngser Park, and Carey E. Priebe.
A central limit theorem for an omnibus embedding of multiple random graphs and implications for multiscale network inference, 2017.

D. L. Sussman, M. Tang, D. E. Fishkind, and C. E. Priebe.
A consistent adjacency spectral embedding for stochastic blockmodel graphs.
Journal of the American Statistical Association, 107(499):1119–1128, 2012.

Wikipedia contributors.
Stochastic block model — Wikipedia, the free encyclopedia, 2019.
[Online; accessed 30-November-2019].