Network community detection under degree heterogeneity: spectral clustering with the random walk Laplacian

Alexander Modell

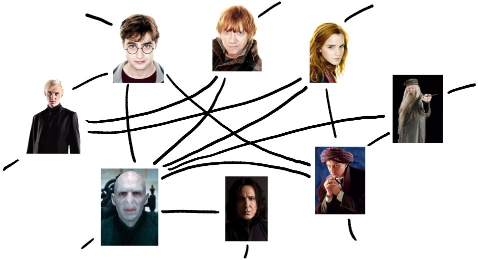
University of Bristol

September 13 2022

- Joint work with Patrick Rubin-Delanchy.

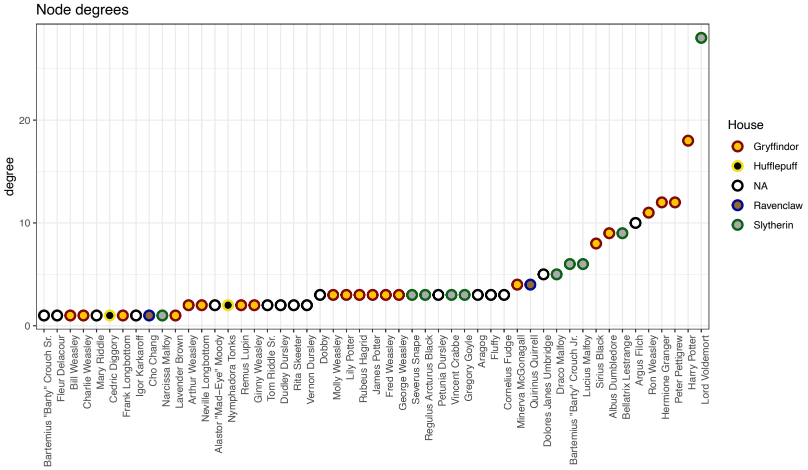
Running example

As a running example, we consider a network of 111 enmities between 51 characters from the Harry Potter book series.

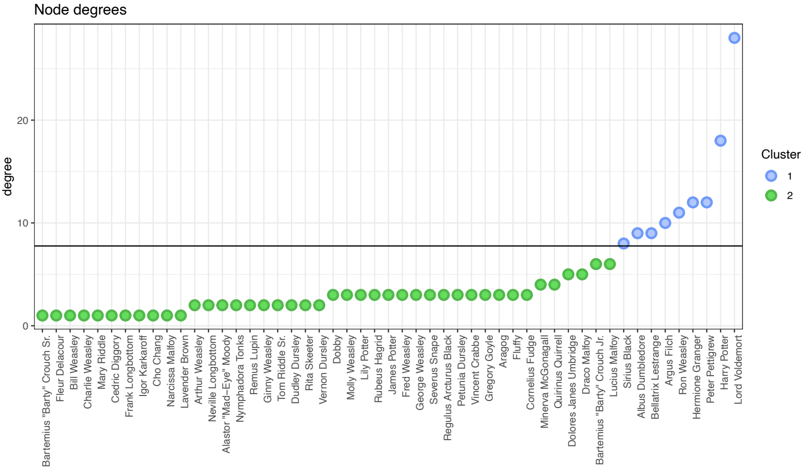


Data available from `github.com/efekarakus/potter-network`.

Harry Potter enmity graph



Harry Potter enmity graph



Spectral clustering algorithm

1. Form a matrix representation of the graph, e.g.
   - $\mathbf{A}$
     adjacency matrix,
   - $\mathbf{L}_{\text{un}} = \mathbf{D} - \mathbf{A}$
     unnormalised Laplacian matrix,
   - $\mathbf{L}_{\text{sym}} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$
     symmetric Laplacian matrix,
   - $\mathbf{L}_{\text{rw}} = \mathbf{D}^{-1}\mathbf{A}$
     random walk Laplacian matrix.

2. Compute it's d largest-in-magnitude eigenvalues $s_1, \ldots, s_d$ and eigenvectors $u_1, \ldots, u_d$.

3. Let $\hat{\mathbf{X}} = (\hat{X}_1, \ldots, \hat{X}_n)^\top := \mathbf{U}|\mathbf{S}|^{1/2}$ where
   - $\mathbf{U} := (u_1, \ldots, u_d)$,
   - $\mathbf{S} := \text{diag}(s_1, \ldots, s_d)$.

4. Cluster $\hat{X}_1, \ldots, \hat{X}_n$.

Stochastic block model (Holland et al., 1983)

Given

- a symmetric probability matrix $\mathbf{B} \in [0,1]^{K \times K}$,
- communities $z_1, \ldots, z_n \in \{1, \ldots, K\}$,

the adjacency matrix $\mathbf{A}$ follows a *stochastic block model* if

$$a_{ij} \overset{\text{ind.}}{\sim} \text{Bernoulli}(b_{z_i, z_j})$$

for $i < j$.

Stochastic block model (Holland et al., 1983)

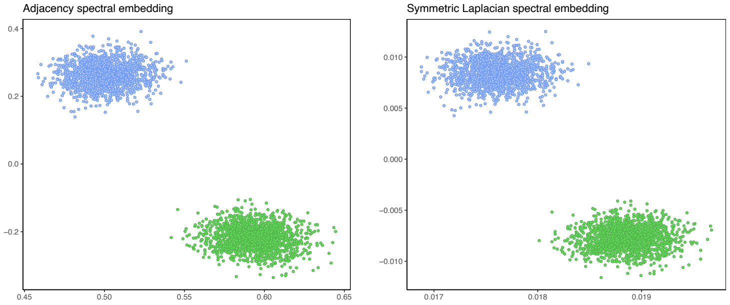
Given

- a symmetric probability matrix $\mathbf{B} \in [0,1]^{K \times K}$,
- communities $z_1, \ldots, z_n \in \{1, \ldots, K\}$,

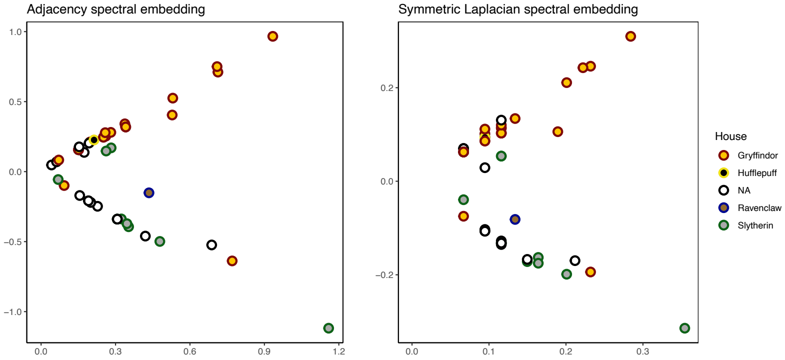
the adjacency matrix $\mathbf{A}$ follows a *stochastic block model* if

$$a_{ij} \overset{\text{ind.}}{\sim} \text{Bernoulli}(b_{z_i, z_j})$$

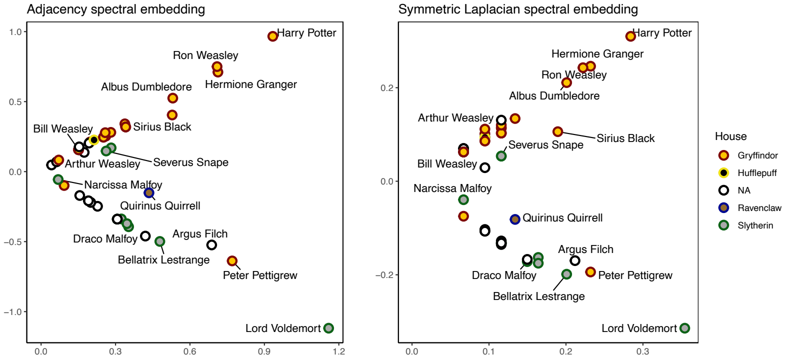
for $i < j$.



Harry Potter enmity graph



Harry Potter enmity graph



Degree-corrected stochastic block model (Karrer and Newman, 2011)

Given

- weights $w_1, \ldots, w_n \in [0, 1]$,
- communities $z_1, \ldots, z_n \in \{1, \ldots, K\}$,
- a symmetric probability matrix $\mathbf{B} \in [0, 1]^{K \times K}$,

the adjacency matrix $\mathbf{A}$ follows a *degree-corrected stochastic block model* if

$$a_{ij} \stackrel{\text{ind.}}{\sim} \text{Bernoulli}(w_i w_j b_{z_i, z_j})$$

for $i < j$.

Degree-corrected stochastic block model (Karrer and Newman, 2011)

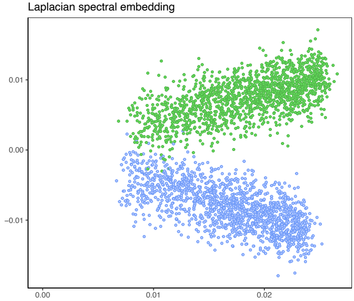
Given

- weights $w_1, \ldots, w_n \in [0, 1]$,
- communities $z_1, \ldots, z_n \in \{1, \ldots, K\}$,
- a symmetric probability matrix $\mathbf{B} \in [0, 1]^{K \times K}$,

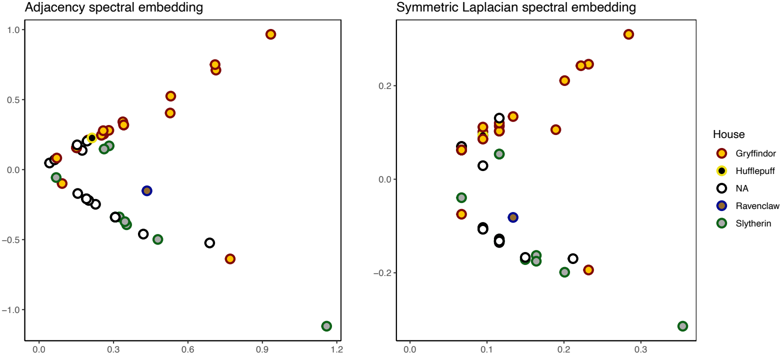
the adjacency matrix $\mathbf{A}$ follows a *degree-corrected stochastic block model* if

$$a_{ij} \overset{\text{ind.}}{\sim} \text{Bernoulli}(w_i w_j b_{z_i, z_j})$$

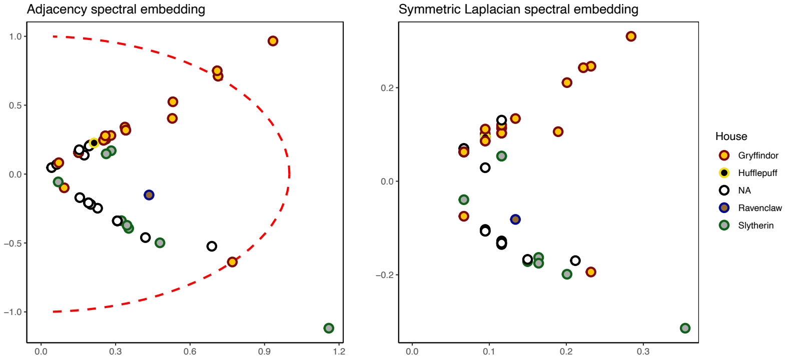
for $i < j$.



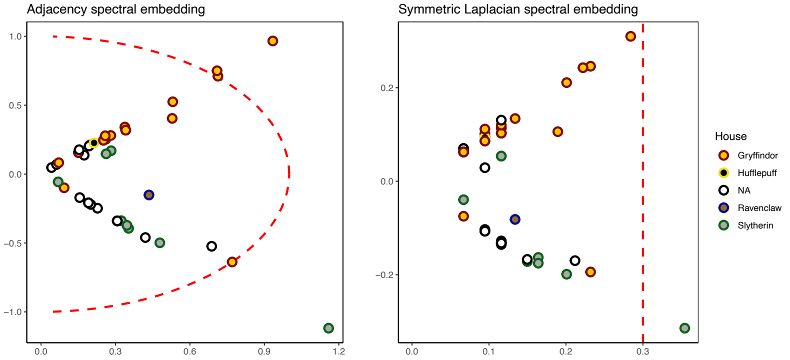
Harry Potter enmity graph



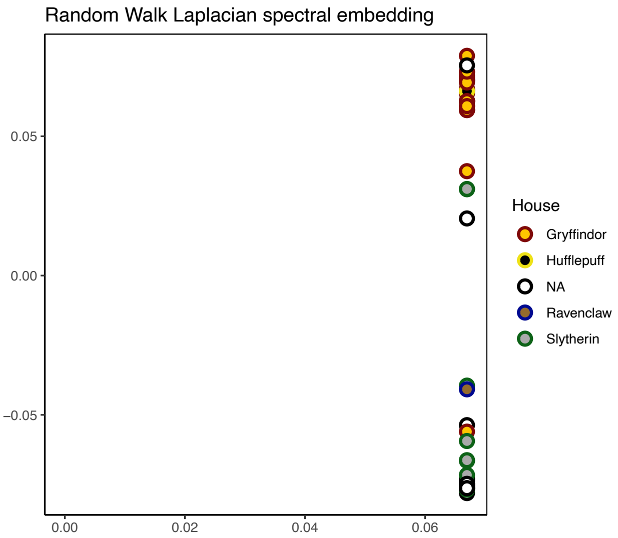
Harry Potter enmity graph



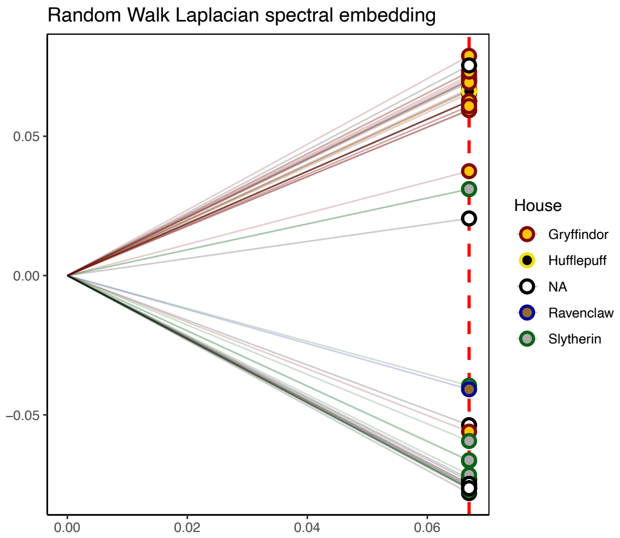
Harry Potter enmity graph



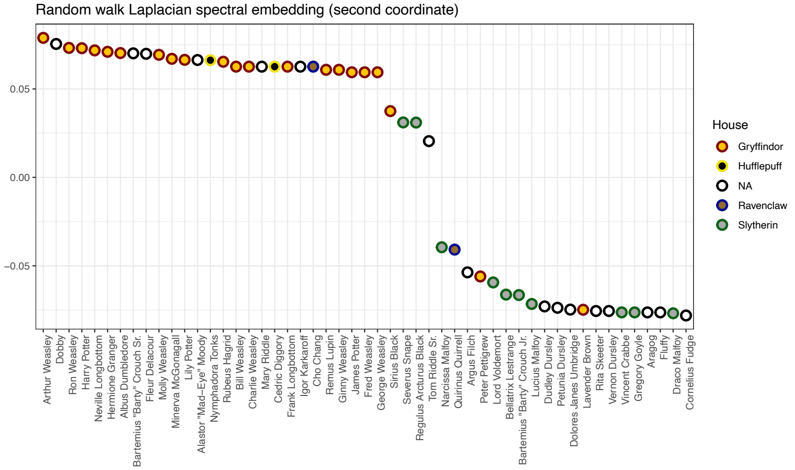
Harry Potter enmity graph



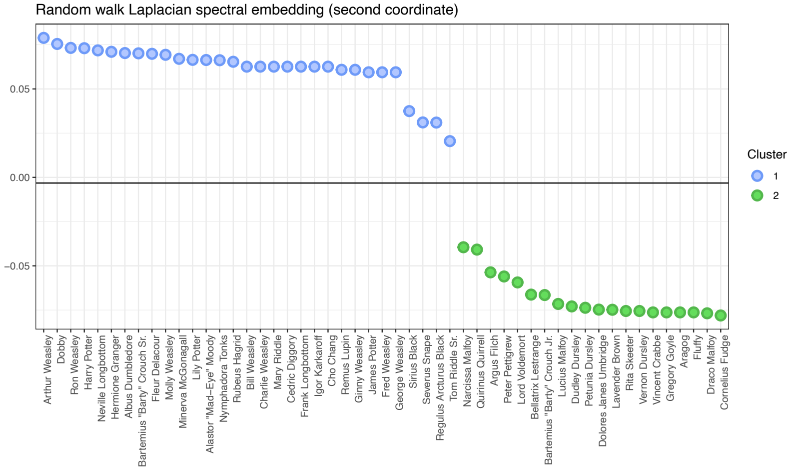
Harry Potter enmity graph



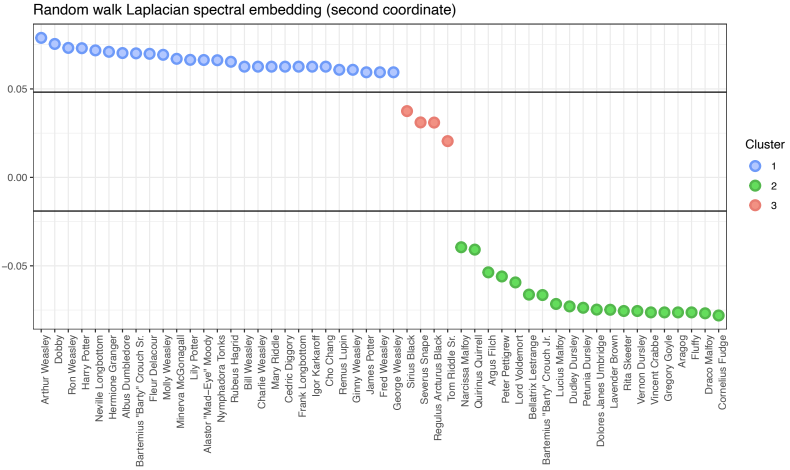
Harry Potter enmity graph



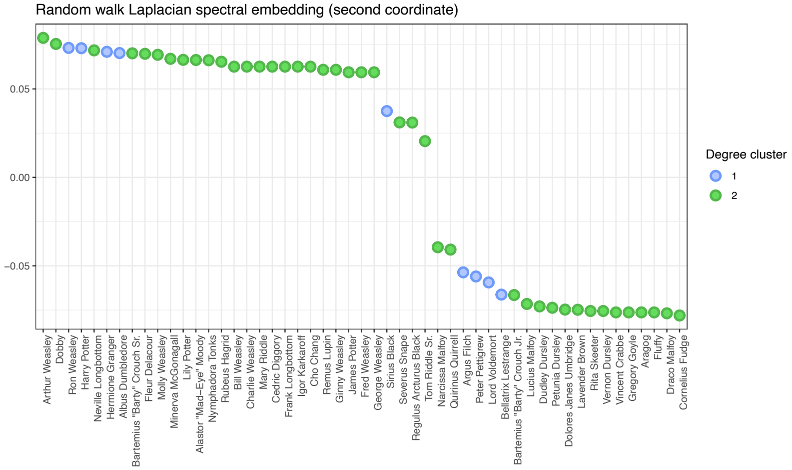
Harry Potter enmity graph



Harry Potter enmity graph



Harry Potter enmity graph



Random dot product graph (Young and Scheinerman, 2007)

Given $X_1, \ldots, X_n \in \mathbb{R}^d$, the adjacency matrix $\mathbf{A}$ follows a *random dot product graph* if

$$a_{ij} \overset{\text{ind.}}{\sim} \text{Bernoulli}(\langle X_i, X_j \rangle)$$

for $i < j$.

Random dot product graph (Young and Scheinerman, 2007)

Given $X_1, \ldots, X_n \in \mathbb{R}^d$, the adjacency matrix $\mathbf{A}$ follows a *random dot product graph* if

$$a_{ij} \overset{\text{ind.}}{\sim} \text{Bernoulli}(\langle X_i, X_j \rangle)$$

for $i < j$.

Remark

$X_1, \ldots, X_n$ are identified only up to a common orthogonal transformation.

Random dot product graph (Young and Scheinerman, 2007)

Given $X_1, \ldots, X_n \in \mathbb{R}^d$, the adjacency matrix $\mathbf{A}$ follows a *random dot product graph* if

$$a_{ij} \overset{\text{ind.}}{\sim} \text{Bernoulli}(\langle X_i, X_j \rangle)$$

for $i < j$.

Remark

$X_1, \ldots, X_n$ are identified only up to a common orthogonal transformation.

Example (Degree-corrected stochastic block model)

Let $v_1, \ldots, v_K \in \mathbb{R}^d$ be such that $b_{k\ell} = \langle v_k, v_\ell \rangle$, then

$$X_i = w_i v_{z_i}, \qquad i = 1, \ldots, n.$$

Random dot product graph (Young and Scheinerman, 2007)

Given $X_1, \ldots, X_n \in \mathbb{R}^d$, the adjacency matrix $\mathbf{A}$ follows a *random dot product graph* if

$$a_{ij} \overset{\text{ind.}}{\sim} \text{Bernoulli}(\langle X_i, X_j \rangle)$$

for $i < j$.

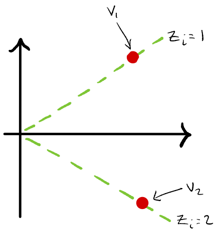
Remark

$X_1, \ldots, X_n$ are identified only up to a common orthogonal transformation.

Example (Degree-corrected stochastic block model)

Let $v_1, \ldots, v_K \in \mathbb{R}^d$ be such that $b_{k\ell} = \langle v_k, v_\ell \rangle$, then

$$X_i = w_i v_{z_i}, \qquad i = 1, \ldots, n.$$



Random dot product graph (Young and Scheinerman, 2007)

Given $X_1, \ldots, X_n \in \mathbb{R}^d$, the adjacency matrix $\mathbf{A}$ follows a *random dot product graph* if

$$a_{ij} \overset{\text{ind.}}{\sim} \text{Bernoulli}(\langle X_i, X_j \rangle)$$

for $i < j$.

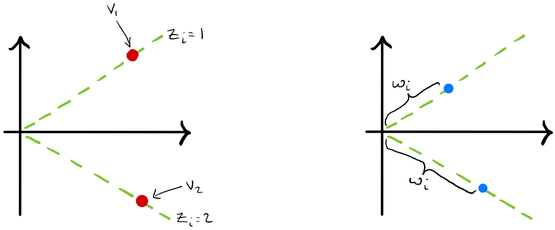
Remark

$X_1, \ldots, X_n$ are identified only up to a common orthogonal transformation.

Example (Degree-corrected stochastic block model)

Let $v_1, \ldots, v_K \in \mathbb{R}^d$ be such that $b_{k\ell} = \langle v_k, v_\ell \rangle$, then

$$X_i = w_i v_{z_i}, \qquad i = 1, \ldots, n.$$



Spectral embedding as estimating a random dot product graph

Theorem (Rubin-Delanchy et al. (2017))

There exists an orthogonal matrix $\mathbf{Q}$ *such that with high probability...*

- *under adjacency spectral embedding*

$$\max_i \left\| \mathbf{Q}\hat{X}_i - X_i \right\| \to 0,$$

- *under symmetric Laplacian spectral embedding*

$$\max_i \left\| \mathbf{Q}\hat{X}_i - \frac{X_i}{\sqrt{t_i}} \right\| \to 0,$$

where $t_i = \sum_j \langle X_i, X_j \rangle$.

Theorem (Athreya et al. (2013); Tang and Priebe (2016))

The error distributions are asymptotically Gaussian.

New results

Theorem (Modell and Rubin-Delanchy (2021))

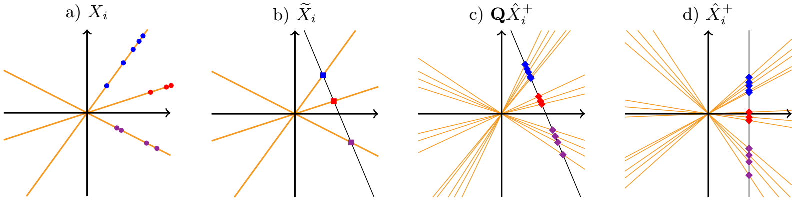
There exists an orthogonal matrix $\mathbf{Q}$ *such that under random walk Laplacian spectral embedding, with high probability,*

$$\max_i \left\| \mathbf{Q}\hat{X}_i - \tilde{X}_i \right\| \to 0,$$

where $\tilde{X}_i$ *is a projective plane representation of* X_i*, and*

$$n^{3/2}\left(\mathbf{Q}\hat{X}_i - \tilde{X}_i\right) \xrightarrow{d} \mathcal{N}(0, \mathbf{\Sigma}(X_i))$$

where $\mathbf{\Sigma}(X_i)$ *has rank* $d-1$.



Projective plane representations of the latent positions

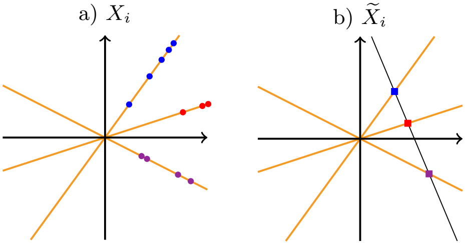
Lemma

Let $P_{\mathcal{H}}$ denote the projection onto the hyperplane

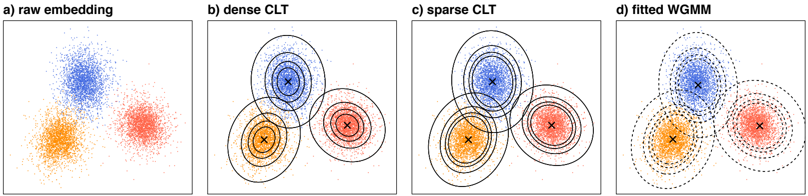
$$\mathcal{H} = \{x \in \mathbb{R}^d : (\sum_j X_j)^\top x = 1\},$$

then

$$\tilde{X}_i := P_{\mathcal{H}}(X_i) = \frac{X_i}{t_i}.$$



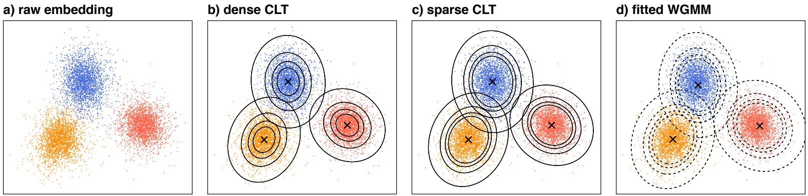
Central limit theorem



Demonstration of the central limit theorem for a degree-corrected stochastic block model with $n = 8000$ nodes and parameters

$$\mathbf{B} = \begin{pmatrix} 0.08 & 0.06 & 0.06 \\ 0.06 & 0.10 & 0.06 \\ 0.06 & 0.06 & 0.12 \end{pmatrix}, \quad w_1, \ldots, w_n \overset{\text{i.i.d.}}{\sim} \text{uniform}(0.25, 1), \quad \pi = (1/3, 1/3, 1/3).$$

Central limit theorem



Demonstration of the central limit theorem for a degree-corrected stochastic block model with $n = 8000$ nodes and parameters

$$\mathbf{B} = \begin{pmatrix} 0.08 & 0.06 & 0.06 \\ 0.06 & 0.10 & 0.06 \\ 0.06 & 0.06 & 0.12 \end{pmatrix}, \quad w_1, \ldots, w_n \overset{\text{i.i.d.}}{\sim} \text{uniform}(0.25, 1), \quad \pi = (1/3, 1/3, 1/3).$$

Remark

Under a sparse degree-corrected stochastic block model, $\mathbf{\Sigma}(X_i) \propto w_i^{-1}$.

References

Athreya, A., Lyzinski, V., Marchette, D. J., Priebe, C. E., Sussman, D. L., and Tang, M. (2013). A central limit theorem for scaled eigenvectors of random dot product graphs. *arXiv preprint arXiv:1305.7388*.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137.

Karrer, B. and Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical review E*, 83(1):016107.

Modell, A. and Rubin-Delanchy, P. (2021). Spectral clustering under degree heterogeneity: a case for the random walk laplacian. *arXiv preprint arXiv:2105.00987*.

Rubin-Delanchy, P., Cape, J., Tang, M., and Priebe, C. E. (2017). A statistical interpretation of spectral embedding: the generalised random dot product graph. *arXiv preprint arXiv:1709.05506*.

Tang, M. and Priebe, C. E. (2016). Limit theorems for eigenvectors of the normalized laplacian for random graphs. *arXiv preprint arXiv:1607.08601*.

Young, S. J. and Scheinerman, E. R. (2007). Random dot product graph models for social networks. In *International Workshop on Algorithms and Models for the Web-Graph*, pages 138–149. Springer.