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

Alexander Modell

University of Bristol

September 13 2022

▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ 三三 - のへぐ

• 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.



▲□▶▲□▶▲□▶▲□▶ □ のQの

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

Node degrees



ヘロト ヘ回ト ヘヨト ヘヨト

æ

20 Ó Cluster degree 1 ----2 10 ..... ÓÓÓÓ 0000000000 Aragog Mary Riddle Cedric Diggory Frank Longbottom Igor Karkaroff Cho Chang Lavender Brown Fluffy Dolores Janes Umbridge Ron Weasley Hermione Granger Bartemius "Barty" Crouch Sr. Bill Weasley Charlie Weasley Narcissa Malfoy Arthur Weasley Neville Longbottom Alastor "Mad-Eye" Moody Nymphadora Toriks Ginny Weasley Tom Riddle Sr. Dudley Dursley Gregory Goyle Cornelius Fudge Minerva McGonagall Quirinus Quirrell Draco Malfoy Bartemius "Barty" Crouch Jr. Lucius Malfoy Sirius Black Albus Dumbledore Bellatrix Lestrange Argus Filch Peter Pettigrew Lord Voldemort Fleur Delacour Remus Lupir Dobb Lily Potter Rubeus Hagric Harry Potter Rita Skeet Vernon Dursle Severus Snap Regulus Arcturus Blac Molly Weask James Potte Fred Weask Vincent Crabb George Weash Petunia Dursh

Node degrees

#### Spectral clustering algorithm

1. Form a matrix representation of the graph, e.g.

• A  
adjacency matrix,  
• 
$$L_{un} = D - A$$
  
unnormalised Laplacian matrix,  
•  $L_{sym} = D^{-1/2}AD^{-1/2}$   
symmetric Laplacian matrix,  
•  $L_{rw} = D^{-1}A$   
random walk Laplacian matrix.

 Compute it's d largest-in-magnitude eigenvalues s<sub>1</sub>,..., s<sub>d</sub> and eigenvectors u<sub>1</sub>,..., u<sub>d</sub>.

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

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

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

## Stochastic block model (Holland et al., 1983)

Given

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

the adjacency matrix A follows a stochastic block model if

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

・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・

for i < j.

## Stochastic block model (Holland et al., 1983)

Given

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

the adjacency matrix A follows a stochastic block model if

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

for i < j.



◆□▶ ◆□▶ ◆臣▶ ◆臣▶ ─臣 ─のへで



◆□▶ ◆□▶ ◆臣▶ ◆臣▶ ─臣 ─の�?



◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへで

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  $\mathbf{B} \in [0,1]^{K imes K}$ ,

the adjacency matrix 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  $\mathbf{B} \in [0, 1]^{K imes K}$ ,

the adjacency matrix A follows a degree-corrected stochastic block model if

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





◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへ⊙



◆□▶ ◆□▶ ◆臣▶ ◆臣▶ ─臣 ─の�?





▲ロト ▲御 ト ▲臣 ト ▲臣 ト → 臣 → の々ぐ



Random Walk Laplacian spectral embedding



Random Walk Laplacian spectral embedding

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ □臣 ○のへ⊙









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

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

for i < j.

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

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

▲□▶ ▲□▶ ▲□▶ ▲□▶ ▲□ ● ● ●

for i < j.

Remark

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

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

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

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

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

▲□▶ ▲□▶ ▲□▶ ▲□▶ ▲□ ● ● ●

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

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

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

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

(日) (四) (日) (日) (日)



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

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

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 **Q** such that with high probability...

under adjacency spectral embedding

$$\max_{i} \left\| \mathbf{Q} \hat{X}_{i} - X_{i} \right\| 
ightarrow 0,$$

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

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

where  $\Sigma(X_i)$  has rank d-1.



イロト イボト イヨト イヨト 三日

## Projective plane representations of the latent positions

#### Lemma

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

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

then

$$ilde{X}_i := P_{\mathcal{H}}(X_i) = rac{X_i}{t_i}$$



## Central limit theorem



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

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

◆□▶ ◆□▶ ◆ 臣▶ ◆ 臣▶ ○ 臣 ○ の Q @

## Central limit theorem



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

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

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

#### 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.

・ロト ・ 目 ・ ・ ヨト ・ ヨト ・ シック