Statistics Webinars – Collegio Carlo Alberto, Torino

Bayesian estimation of the latent dimension and communities in stochastic blockmodels

Imperial College London

Francesco Sanna Passino, Nick Heard
Department of Mathematics, Imperial College London
francesco.sanna-passino16@imperial.ac.uk

8th May 2020

Undirected graphs

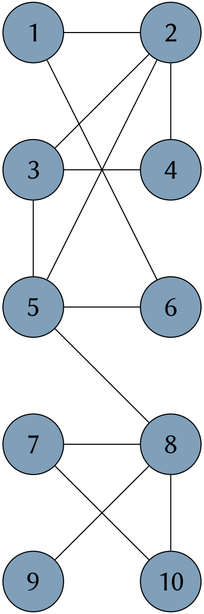
- **Undirected graph** $\mathbb{G} = (V, E)$ where:
  - V is the **node set**, $n = |V|$,
  - $E \subseteq V \times V$ is the **edge set**, containing dyads $(i, j),\ i, j \in V$.
- An edge is drawn if a node $i \in V$ connects to $j \in V$, written $(i, j) \in E$.
- From $\mathbb{G}$, an **adjacency matrix** $\mathbf{A} = \{A_{ij}\}$, of dimension $n \times n$, can be obtained:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 1 & \cdots & 1 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 1 & \cdots & 1 & 0 \end{pmatrix} \qquad A_{ij} = \begin{cases} 1 & \text{if } (i, j) \in E \\ 0 & \text{otherwise} \end{cases}$$

- Commonly, self-edges are not allowed, implying that $\mathbf{A}$ is a **hollow** matrix.

A toy example

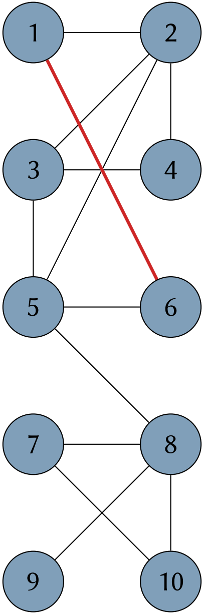
$\mathbb{G} =$



$$\Rightarrow \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

A toy example

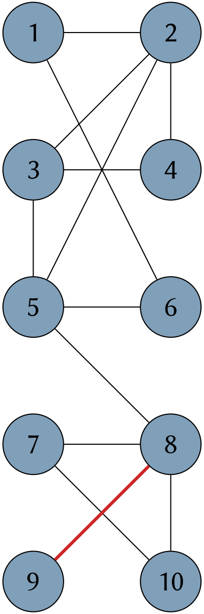
$\mathbb{G} =$



$$\Rightarrow \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & {\color{red}1} & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ {\color{red}1} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

A toy example

$\mathbb{G} =$



$$\Rightarrow \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & \color{red}{1} & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \color{red}{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Statistical models for undirected graphs

- Consider an **undirected graph** with **symmetric adjacency matrix** $\mathbf{A} \in \{0,1\}^{n\times n}$.
- **Latent feature models** (Hoff, Raftery, and Handcock, 2002): each node is assigned a latent position $\boldsymbol{x}_i$ in a d-dimensional latent space $\mathbb{X}$.
- The edges are generated *independently* using a **kernel function** $\psi : \mathbb{X} \times \mathbb{X} \to [0,1]$:

$$\mathbb{P}(A_{ij}=1) = \psi(\boldsymbol{x}_i, \boldsymbol{x}_j),\ i<j,\ A_{ij}=A_{ji}.$$

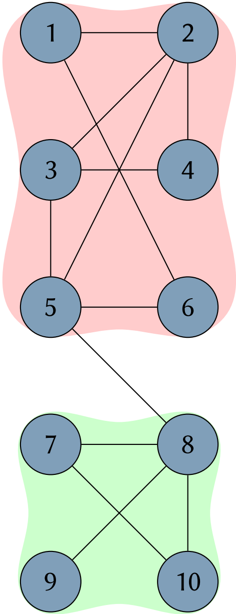
- The latent positions are represented as a $(n\times d)$-dimensional matrix $\mathbf{X} = [\boldsymbol{x}_1,\dots,\boldsymbol{x}_n]^\top$.
- In **random dot product graphs** (RDPG) (Young and Scheinerman, 2007; Athreya et al., 2018), the kernel is the **inner product** of the latent positions, and $\mathbb{X}$ is chosen such that $0 \leq \boldsymbol{x}^\top \boldsymbol{y} \leq 1\ \forall\ \boldsymbol{x},\boldsymbol{y} \in \mathbb{X}$:

$$\mathbb{P}(A_{ij}=1) = \boldsymbol{x}_i^\top \boldsymbol{x}_j,\ i<j,\ A_{ij}=A_{ji}.$$

- In RDPGs: $d = \mathrm{rank}\{\mathbb{E}(\mathbf{A})\} = \mathrm{rank}(\mathbf{X}\mathbf{X}^\top)$.

A toy example: community detection

$$\mathbb{G} =$$



$$\Rightarrow \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Clustering nodes in undirected graphs

- The **stochastic blockmodel** (SBM) (Holland, Laskey, and Leinhardt, 1983) is the classical model for community detection in graphs.
- Assume K communities, and a matrix $\mathbf{B} \in [0,1]^{K \times K}$ of within-community probabilities.
- Each node is assigned a community $z_i \in \{1, \dots, K\}$ with probability $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$, from the $K-1$ probability simplex.
- The probability of a link depends on the **community allocations** z_i and z_j of the nodes:

$$\mathbb{P}(A_{ij} = 1) = B_{z_i z_j},\ i < j,\ A_{ij} = A_{ji}.$$

- The likelihood for an observed symmetric adjacency matrix $\mathbf{A}$ is:

$$L(\mathbf{A}|\boldsymbol{z}, \mathbf{B}) = \prod_{1 \leq i < j \leq n} B_{z_i z_j}^{A_{ij}} (1 - B_{z_i z_j})^{1 - A_{ij}}.$$

The SBM as a special case of RDPG

- The stochastic blockmodel can be interpreted as a **special case** of a RDPG.
- For simplicity, initially assume that $\mathbf{B}$ is *positive semi-definite*.
- Assume that $B_{kh} = \boldsymbol{\mu}_k^\top \boldsymbol{\mu}_h$ for some $\boldsymbol{\mu}_k, \boldsymbol{\mu}_h \in \mathbb{X}$.
- If all the nodes in community k are assigned the latent position $\boldsymbol{\mu}_k$, then:

$$\mathbb{P}(A_{ij} = 1) = B_{z_i z_j} = \boldsymbol{\mu}_{z_i}^\top \boldsymbol{\mu}_{z_j},\ i < j,\ A_{ij} = A_{ji}.$$

- In this framework: $d = \text{rank}\{\mathbb{E}(\mathbf{A})\} = \text{rank}(\mathbf{X}\mathbf{X}^\top) = \text{rank}(\mathbf{B}) \leq K$.
- Extension to *any* $\mathbf{B}$: generalised RDPG (GRDPG, Rubin-Delanchy et al., 2017).
- **Inference** on SBMs as (G)RDPGs:
  - Latent dimension d,
  - Number of communities K,
  - Community allocations $\boldsymbol{z} = (z_1, \dots, z_n)$,
  - Latent positions $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K$.

Beyond RDPGs: the generalised random dot product graph

Definition (Generalised random dot product graph, GRDPG, Rubin-Delanchy et al., 2017)

Let d_+, d_- be non-negative integers such that $d = d_+ + d_-$. Let $\mathbb{X} \subseteq \mathbb{R}^d$ such that $\forall\, \boldsymbol{x}, \boldsymbol{x}' \in \mathbb{X}$, $0 \leq \boldsymbol{x}^\top \mathbf{I}(d_+, d_-)\boldsymbol{x}' \leq 1$, where

$$\mathbf{I}(p,q) = \mathrm{diag}(\underbrace{1,\dots,1}_{p}, \underbrace{-1,\dots,-1}_{q}).$$

Let $\mathcal{F}$ be a probability measure on $\mathbb{X}$, $\mathbf{A} \in \{0,1\}^{n\times n}$ be a symmetric matrix and $\mathbf{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)^\top \in \mathbb{X}^n$. Then $(\mathbf{A}, \mathbf{X}) \sim \mathrm{GRDPG}_{d_+,d_-}(\mathcal{F})$ if $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \stackrel{iid}{\sim} \mathcal{F}$ and for $i < j$, independently

$$\mathbb{P}(A_{ij} = 1) = \boldsymbol{x}_i^\top \mathbf{I}(d_+, d_-)\boldsymbol{x}_j.$$

- To represent the K-community SBM as a GRDPG, $\mathcal{F}$ can be chosen to have mass concentrated at $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K \in \mathbb{R}^d$ such that $\boldsymbol{\mu}_i^\top \mathbf{I}(d_+, d_-)\boldsymbol{\mu}_j = B_{ij}\ \forall\, i,j \in \{1, \dots, K\}$.

Network embeddings

Definition (Adjacency spectral embedding, ASE)

For $d \in \{1, \dots, n\}$, consider the spectral decomposition

$$\mathbf{A} = \hat{\mathbf{\Gamma}}\hat{\mathbf{\Lambda}}\hat{\mathbf{\Gamma}}^\top + \hat{\mathbf{\Gamma}}_\perp\hat{\mathbf{\Lambda}}_\perp\hat{\mathbf{\Gamma}}_\perp^\top,$$

where $\hat{\mathbf{\Lambda}}$ is a $d \times d$ diagonal matrix containing the top d eigenvalues in magnitude, in decreasing order, $\hat{\mathbf{\Gamma}}$ is a $n \times d$ matrix containing the corresponding orthonormal eigenvectors, and the matrices $\hat{\mathbf{\Lambda}}_\perp$ and $\hat{\mathbf{\Gamma}}_\perp$ contain the remaining $n-d$ eigenvalues and eigenvectors. The adjacency spectral embedding $\hat{\mathbf{X}} = [\hat{\boldsymbol{x}}_1, \dots, \hat{\boldsymbol{x}}_n]^\top$ of $\mathbf{A}$ in $\mathbb{R}^d$ is

$$\hat{\mathbf{X}} = \hat{\mathbf{\Gamma}}|\hat{\mathbf{\Lambda}}|^{1/2} \in \mathbb{R}^{n\times d},$$

where the operator $|\cdot|$ applied to a matrix returns the absolute value of its entries.

- $\hat{\mathbf{X}}\mathbf{I}(d_+, d_-)\hat{\mathbf{X}}^\top$ represents an estimate of $\mathbb{E}(\mathbf{A}) = \mathbf{X}\mathbf{I}(d_+, d_-)\mathbf{X}^\top \rightarrow$ **link prediction**.

Network embeddings

Definition (Laplacian spectral embedding, LSE)

For $d \in \{1, \dots, n\}$, consider the (modified) normalised Laplacian matrix

$$\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2},\ \mathbf{D} = \operatorname{diag}\left(\sum_{j=1}^{n} A_{ij}\right),$$

and its spectral decomposition

$$\mathbf{L} = \tilde{\mathbf{\Gamma}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Gamma}}^\top + \tilde{\mathbf{\Gamma}}_\perp\tilde{\mathbf{\Lambda}}_\perp\tilde{\mathbf{\Gamma}}_\perp^\top.$$

The Laplacian spectral embedding $\tilde{\mathbf{X}} = [\tilde{\boldsymbol{x}}_1, \dots, \tilde{\boldsymbol{x}}_n]^\top$ of $\mathbf{A}$ in $\mathbb{R}^d$ is

$$\tilde{\mathbf{X}} = \tilde{\mathbf{\Gamma}}|\tilde{\mathbf{\Lambda}}|^{1/2}.$$

- The modified Laplacian $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ (Rohe, Chatterjee, and Yu, 2011) is preferred to the version $\mathbf{I}_n - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ since its eigenvalues lie in $(-1, 1)$, providing a convenient interpretation for disassortative networks (Rubin-Delanchy, Adams, and Heard, 2016).

Limit theorems for ASE (Rubin-Delanchy et al., 2017)

- Let $\boldsymbol{\xi}$ be a random vector such that $\boldsymbol{\xi} \sim F$, where F is supported on $\mathbb{X}$ and $\boldsymbol{\xi}$ has full rank second order moment matrix $\boldsymbol{\Delta} = \mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}^\top) \in \mathbb{R}^{d\times d}$, for d **fixed**, **constant** and **known**.
- Introduce a *sparsity factor* ρ_n, requiring $\rho_n = 1$ or $\rho_n \to 0$.
- The latent positions $\boldsymbol{x}_1^{(n)} = \rho_n^{1/2}\boldsymbol{\xi}_1^{(n)}, \dots, \boldsymbol{x}_n^{(n)} = \rho_n^{1/2}\boldsymbol{\xi}_n^{(n)}$ at each step are assumed to be independent replicates of the random vector $\rho_n^{1/2}\boldsymbol{\xi}$.
- Consequently, $\mathcal{F}$ is assumed to factorise into a product F_ρ^n of n identical marginal distributions that are equal to F up to scaling.

Theorem (ASE two-to-infinity norm bound)

Consider $(\mathbf{A}^{(n)}, \mathbf{X}^{(n)}) \sim \mathrm{GRDPG}_{d_+,d_-}(F_\rho^n)$. *There exists a universal constant* $\varepsilon > 0$ *such that, provided that* $n\rho_n = \omega\{(\log n)^{4\varepsilon}\}$, *there exists* $\mathbf{Q}_n \in \mathbb{O}(d_+, d_-)$ *such that*

$$\left\|\mathbf{Q}_n\hat{\boldsymbol{x}}_i^{(n)} - \boldsymbol{x}_i^{(n)}\right\|_{2\to\infty} = \max_i \left\|\mathbf{Q}_n\hat{\boldsymbol{x}}_i^{(n)} - \boldsymbol{x}_i^{(n)}\right\| = O_{\mathbb{P}}\left\{\frac{(\log n)^\varepsilon}{n^{1/2}}\right\}.$$

$X = O_{\mathbb{P}}\{f(n)\}$ if, for any $\varepsilon > 0$, $\exists\, n_\varepsilon \in \mathbb{N}, C_\varepsilon > 0$, s.t. $\mathbb{P}\{|X| \leq C_\varepsilon f(n)\} \geq 1 - n^{-\varepsilon}\ \forall\, n \geq n_\varepsilon$.

Limit theorems for ASE (Rubin-Delanchy et al., 2017)

Theorem (ASE central limit theorem)

Consider the sequence of graphs $(\mathbf{A}^{(n)}, \mathbf{X}^{(n)}) \sim \mathrm{GRDPG}_{d_+,d_-}(F_\rho^n)$, *such that* $n\rho_n = \omega\{(\log n)^{4\varepsilon}\}$ *for the universal constant* $\varepsilon > 0$. *For any integer* $m > 0$, *choose points* $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{X}$ *in the support of* F, *and points* $\boldsymbol{q}_1, \dots, \boldsymbol{q}_m \in \mathbb{R}^d$. *Then there exists a sequence of random matrices* $\mathbf{Q}_n \in \mathbb{O}(d_+, d_-)$ *such that*

$$\mathbb{P}\left\{\bigcap_{i=1}^m n^{1/2}\left(\mathbf{Q}_n\hat{\boldsymbol{x}}_i^{(n)} - \boldsymbol{x}_i^{(n)}\right) \leq \boldsymbol{q}_i \,\middle|\, \boldsymbol{\xi}_1^{(n)} = \boldsymbol{x}_1, \dots, \boldsymbol{\xi}_m^{(n)} = \boldsymbol{x}_m\right\} \longrightarrow \prod_{i=1}^m \Phi\left\{\boldsymbol{q}_i, \boldsymbol{\Sigma}(\boldsymbol{x}_i)\right\},$$

where $\Phi\{\boldsymbol{q}, \boldsymbol{\Sigma}\}$ *is the cumulative distribution function of a multivariate normal distribution with mean* $\mathbf{0}$ *and covariance* $\boldsymbol{\Sigma}$, *evaluated at* $\boldsymbol{q}$, *and*

$$\boldsymbol{\Sigma}(\boldsymbol{x}) = \begin{cases} \mathbf{I}(d_+,d_-)\boldsymbol{\Delta}^{-1}\mathbb{E}[\{\boldsymbol{x}^\top\mathbf{I}(d_+,d_-)\boldsymbol{\xi}\}\{1-\boldsymbol{x}^\top\mathbf{I}(d_+,d_-)\boldsymbol{\xi}\}\boldsymbol{\xi}\boldsymbol{\xi}^\top]\boldsymbol{\Delta}^{-1}\mathbf{I}(d_+,d_-) & \text{if } \rho_n = 1 \\ \mathbf{I}(d_+,d_-)\boldsymbol{\Delta}^{-1}\mathbb{E}[\{\boldsymbol{x}^\top\mathbf{I}(d_+,d_-)\boldsymbol{\xi}\}\boldsymbol{\xi}\boldsymbol{\xi}^\top]\boldsymbol{\Delta}^{-1}\mathbf{I}(d_+,d_-) & \text{if } \rho_n \to 0 \end{cases}.$$

- The theorem has *crucial* relevance in practice.

Practical utility of the limit theorems

- ***Uniqueness up to indefinite orthogonal transformations***
  For any matrix $\mathbf{Q} \in \mathbb{O}(d_+, d_-)$, the indefinite orthogonal group with signature (d_+, d_-),
  $$(\mathbf{Q}\boldsymbol{\mu}_{z_i})^\top \mathbf{I}(d_+, d_-)(\mathbf{Q}\boldsymbol{\mu}_{z_j}) = \boldsymbol{\mu}_{z_i}^\top \mathbf{I}(d_+, d_-)\boldsymbol{\mu}_{z_j}.$$
- **If d is known**, conditioning on K, the ASE CLT implies that **Gaussian mixture modelling** gives a **consistent estimate** of the locations $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K$ in SBMs.
- Intuitively, the algorithm approximately holds because:
  $$\hat{\boldsymbol{x}}_i \approx \mathbb{N}\{\mathbf{Q}_n\boldsymbol{\mu}_{z_i}, n^{-1}\mathbf{Q}_n\boldsymbol{\Sigma}(\boldsymbol{\mu}_{z_i})\mathbf{Q}_n^\top\},\ n \to \infty,\ i = 1, \dots, m.$$
- Importantly, K-means, with Euclidean distance, which has been traditionally extensively used in spectral clustering, is **suboptimal** and **unsound** for identifiability reasons.
- Similar asymptotic results are also available for the **Laplacian spectral embedding**.

Spectral estimation of the stochastic blockmodel

- Based the asymptotic properties derived in Rubin-Delanchy et al., 2017, the following algorithm should be used for consistent estimation of the latent positions in stochastic blockmodels, **when d and K are known**:

Algorithm: Spectral estimation of the stochastic blockmodel (**spectral clustering**)

Input: adjacency matrix $\mathbf{A}$ (or the Laplacian matrix $\mathbf{L}$), dimension d, and number of communities $K \geq d$.

1 compute spectral embedding $\hat{\mathbf{X}} = [\hat{\boldsymbol{x}}_1, \dots, \hat{\boldsymbol{x}}_n]^\top$ or $\tilde{\mathbf{X}} = [\tilde{\boldsymbol{x}}_1, \dots, \tilde{\boldsymbol{x}}_n]^\top$ into $\mathbb{R}^d$,

2 fit a Gaussian mixture model with K components,

Result: return cluster centres $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K \in \mathbb{R}^d$ and node memberships $z_1, \dots, z_n$.

- What about d and K? In practice the two parameters are estimated **sequentially**.
  - The latent dimension d is chosen according to the scree-plot criterion (Jolliffe, 2002), or the universal singular value thresholding method (Zhu and Ghodsi, 2006).
  - The number of communities K is *usually* chosen using information criteria, conditional on d.
- **This talk discusses a novel framework for joint estimation of d and K.**

Estimation of d: *"overshooting"*

- Main issues for estimation of d and K:
  - Sequential approach is **sub-optimal**: the estimate of K depends on choice of d.
  - Theoretical results only hold for d **fixed and known**.
  - Distributional assumptions when d is misspecified are **not available**.
  - What is the **distribution of the last $m-d$ columns of the embedding**, for $m > d$?
- How to deal with uncertainty in the estimate of d? *"Overshooting"*.
  - Obtain embeddings $\mathbf{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n]^\top \in \mathbb{R}^{n \times m}$, $\boldsymbol{x}_i \in \mathbb{R}^m$ (ASE or LSE) for some m.
  - *Here* $\mathbf{X}$ *represents an estimate of the latent positions (ASE or LSE), dropping "hats" and "tildes".*
  - *Ideally*, m must be $d \leq m \leq n$, so it can be given an **arbitrarily large value**.
  - The parameter m is always assumed to be fixed and obtained from a preprocessing step.
  - Choosing an appropriate value of m is arguably **much easier** than choosing the correct d.
  - Under the estimation framework that will be proposed, the correct d can be recovered for any choice of m, as long as $d \leq m$.

A Bayesian model for network embeddings

- Choose integer $m \leq n$ and obtain embedding $\mathbf{X} \in \mathbb{R}^{n \times m} \to m$ arbitrarily large.
- Bayesian model for simultaneous estimation of d and $K \to$ allow for $d = \mathrm{rank}(\mathbf{B}) \leq K$.

$$\boldsymbol{x}_i | d, z_i, \boldsymbol{\mu}_{z_i}, \boldsymbol{\Sigma}_{z_i}, \boldsymbol{\sigma}^2_{z_i} \sim \mathbb{N}_m \left(\begin{bmatrix} \boldsymbol{\mu}_{z_i} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{z_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma}^2_{z_i} \mathbf{I}_{m-d} \end{bmatrix} \right), \ i = 1, \dots, n,$$

$$(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) | d \overset{iid}{\sim} \mathrm{NIW}_d(\mathbf{0}, \kappa_0, \nu_0 + d - 1, \boldsymbol{\Delta}_d), \ k = 1, \dots, K,$$

$$\sigma^2_{kj} \overset{iid}{\sim} \mathrm{Inv}\text{-}\chi^2(\lambda_0, \sigma^2_0), \ j = d+1, \dots, m,$$

$$d | \boldsymbol{z} \sim \mathrm{Uniform}\{1, \dots, K_\varnothing\},$$

$$z_i | \boldsymbol{\theta} \overset{iid}{\sim} \mathrm{Discrete}(\boldsymbol{\theta}), \ i = 1, \dots, n, \ \boldsymbol{\theta} \in \mathcal{S}_{K-1},$$

$$\boldsymbol{\theta} | K \sim \mathrm{Dirichlet}\left(\frac{\alpha}{K}, \dots, \frac{\alpha}{K} \right),$$

$$K \sim \mathrm{Geometric}(\omega).$$

where $K_\varnothing$ is the number of non-empty communities.

- Alternative: $d \sim \mathrm{Geometric}(\delta)$.
- Yang et al., 2019, independently and simultaneously proposed a similar frequentist model.

Empirical model validation

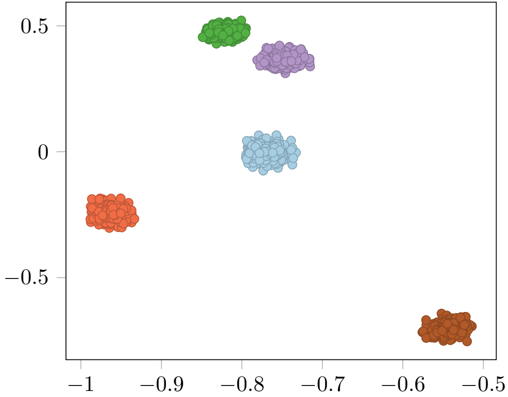


Figure 1. Scatterplot of the columns $\mathbf{X}_1$ and $\mathbf{X}_2$ of the ASE.

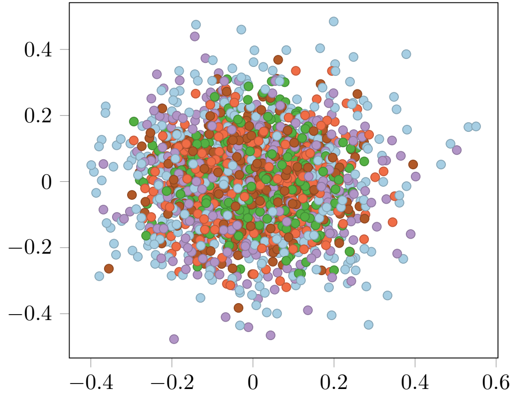


Figure 2. Scatterplot of the columns $\mathbf{X}_3$ and $\mathbf{X}_4$ of the ASE.

- Simulated GRDPG-SBM with $n = 2{,}500$, $d = 2$, $K = 5$.
- Nodes allocated to communities with probability $\theta_k = \mathbb{P}(z_i = k) = 1/K$.

Empirical model validation

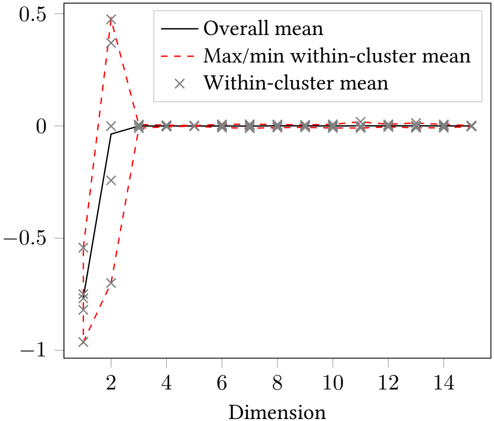


Figure 3. Within-cluster and overall means of $\mathbf{X}_{:15}$.

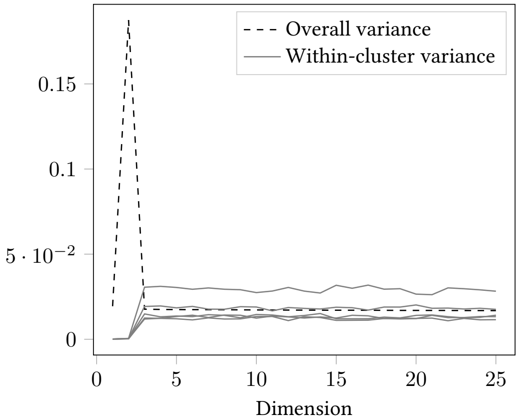


Figure 4. Within-cluster variance of $\mathbf{X}_{:25}$.

- Means are approximately $\mathbf{0}$ for columns with index $> d$.
- Different cluster-specific variances even for columns with index $> d$.

Empirical model validation

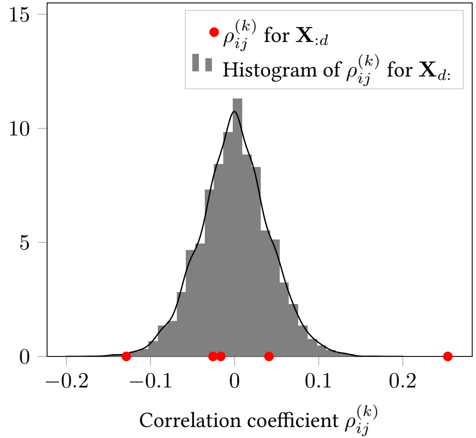


Figure 5. Within-cluster correlation coefficients of $\mathbf{X}_{:30}$.

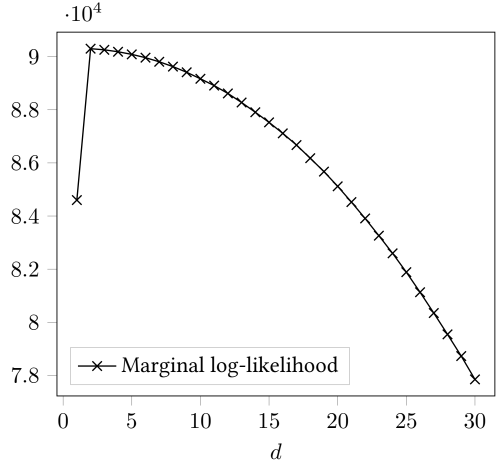


Figure 6. Marginal likelihood as a function of d.

- Reasonable to assume correlation $\rho_{ij}^{(k)} = 0$ for $i, j > d$.
- Marginal likelihood has maximum at the true value of d.

Curse of dimensionality

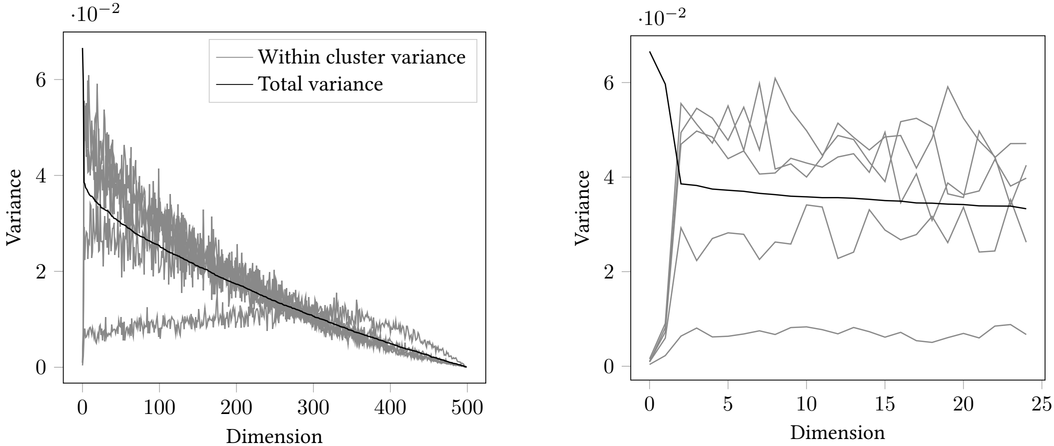


Figure 7. Within-block variance and total variance for the adjacency embedding obtained from a simulated SBM with $d = 2$, $K = 5$, $n = 500$, and well separated means $\boldsymbol{\mu}_1 = [0.7, 0.4]$, $\boldsymbol{\mu}_2 = [0.1, 0.1]$, $\boldsymbol{\mu}_3 = [0.4, 0.8]$, $\boldsymbol{\mu}_4 = [-0.1, 0.5]$ and $\boldsymbol{\mu}_5 = [0.3, 0.5]$, and $\boldsymbol{\theta} = (0.2, 0.2, 0.2, 0.2, 0.2)$.

- For some k and k': $\sigma^2_{kj} \approx \sigma^2_{k'j}$ for $j \gg d$ and $k \neq k'$.

Second order clustering

- **Bayesian model parsimony**: K underestimated for $d \ll m$.
- Possible solution: **second order clustering** $\boldsymbol{v} = (v_1, \dots, v_K)$ with $v_k \in \{1, \dots, H\}$.
- If $v_k = v_{k'}$, then $\sigma^2_{kj} = \sigma^2_{k'j}$ for $j > d$:

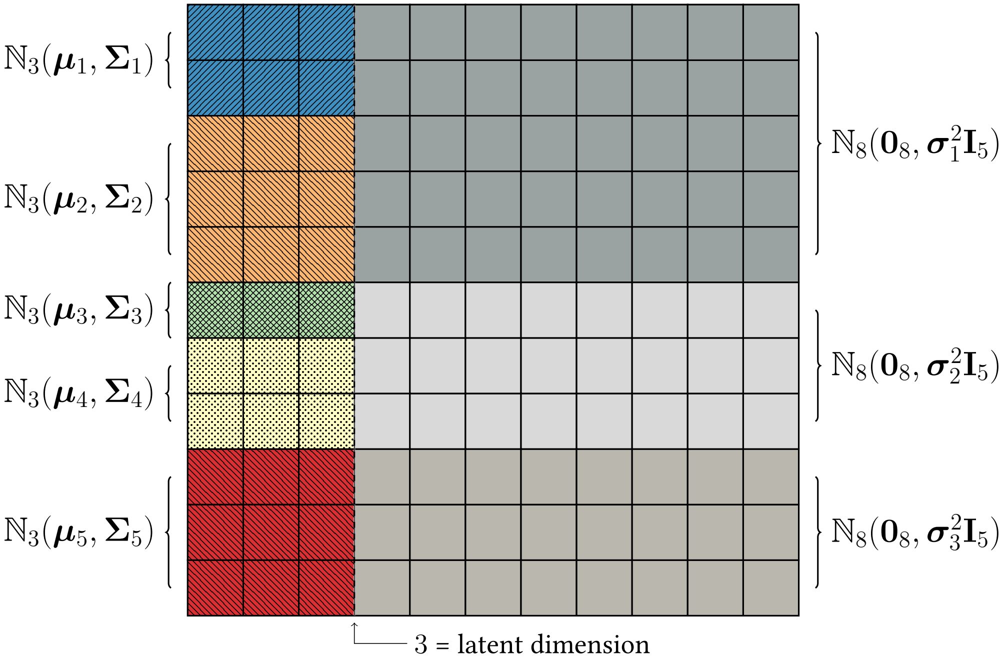
$$\boldsymbol{x}_i | d, z_i, v_{z_i}, \boldsymbol{\mu}_{z_i}, \boldsymbol{\Sigma}_{z_i}, \boldsymbol{\sigma}^2_{v_{z_i}} \sim \mathbb{N}_m \left(\begin{bmatrix} \boldsymbol{\mu}_{z_i} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{z_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma}^2_{v_{z_i}} \mathbf{I}_{m-d} \end{bmatrix} \right), \; i = 1, \dots, n,$$

$$v_k | K, H \sim \mathrm{Discrete}(\boldsymbol{\phi}), \; k = 1, \dots, K,$$

$$\boldsymbol{\phi} | H \sim \mathrm{Dirichlet} \left(\frac{\beta}{H}, \dots, \frac{\beta}{H} \right),$$

$$H | K \sim \mathrm{Uniform}\{1, \dots, K\}.$$

- The parameter v_k defines **clusters of clusters**.
- Empirical results show that the model is able to handle $d \ll m$.
- If $H = 1$, the model is a special case of Raftery and Dean, 2006 $\to$ ***ordinal* variable selection in clustering**.


$\mathbb{N}_3(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$
$\mathbb{N}_3(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$
$\mathbb{N}_3(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$
$\mathbb{N}_3(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_4)$
$\mathbb{N}_3(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_5)$
$\mathbb{N}_8(\mathbf{0}_8, \boldsymbol{\sigma}_1^2\mathbf{I}_5)$
$\mathbb{N}_8(\mathbf{0}_8, \boldsymbol{\sigma}_2^2\mathbf{I}_5)$
$\mathbb{N}_8(\mathbf{0}_8, \boldsymbol{\sigma}_3^2\mathbf{I}_5)$
3 = latent dimension

Inference

- **Integrate out nuisance parameters** $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$, σ^2_{jk} and $\boldsymbol{\theta}$ → inference on d, K, H and $\boldsymbol{z}$.
- Inference via MCMC: **collapsed Metropolis-within-Gibbs sampler** → 7 moves.
  - Propose a **change in the community allocations** $\boldsymbol{z}$,
  - Propose to **split (or merge) two communities**,
  - Propose to **create (or remove) an empty community**,
  - Propose a **change in the latent dimension** d,
  - Propose a change in the second order community allocations $\boldsymbol{v}$,
  - Propose to split (or merge) two second-order communities,
  - Propose to create (or remove) an empty second-order community.
- **Initialisation**: K-means clustering, choose K from scree-plot + uninformative priors (with zero means and variances comparable in scale with the observed data).
- Posterior for d is usually similar to a **point mass** → might be worth exploring constrained and unconstrained model.
- The latent dimension d could also be treated as a nuisance parameter and **marginalised out** (often not computationally feasible).

Extension to directed and bipartite graphs

- Consider a **directed graph** with adjacency matrix $\mathbf{A} \in \{0,1\}^{n\times n}$.
- The d-dimensional "directed" adjacency embedding (DASE) of $\mathbf{A}$ in $\mathbb{R}^{2d}$, is defined as:

$$\hat{\mathbf{U}}\hat{\mathbf{D}}^{1/2} \oplus \hat{\mathbf{V}}\hat{\mathbf{D}}^{1/2} = \begin{bmatrix}\hat{\mathbf{U}}\hat{\mathbf{D}}^{1/2} & \hat{\mathbf{V}}\hat{\mathbf{D}}^{1/2}\end{bmatrix} = \begin{bmatrix}\hat{\mathbf{X}} & \hat{\mathbf{X}}'\end{bmatrix},$$

  where $\mathbf{A} = \hat{\mathbf{U}}\hat{\mathbf{D}}\hat{\mathbf{V}}^\top + \hat{\mathbf{U}}_\perp\hat{\mathbf{D}}_\perp\hat{\mathbf{V}}_\perp^\top$ is the **SVD decomposition** of $\mathbf{A}$, where $\hat{\mathbf{D}} \in \mathbb{R}_+^{d\times d}$ is a diagonal matrix containing the top d singular values in decreasing order, and $\hat{\mathbf{U}} \in \mathbb{R}^{n\times d}$ and $\hat{\mathbf{V}} \in \mathbb{R}^{n\times d}$ contain the corresponding left and right singular vectors.
- Extended model:

$$\boldsymbol{x}_i|d,K,z_i \sim \mathbb{N}_{2m}\left(\begin{bmatrix}\boldsymbol{\mu}_{z_i}\\ \mathbf{0}\\ \boldsymbol{\mu}'_{z_i}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{z_i} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\sigma}^2_{z_i}\mathbf{I}_{m-d} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}'_{z_i} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma}^{2\prime}_{z_i}\mathbf{I}_{m-d}\end{bmatrix}\right).$$

- **Co-clustering**: different clusters for sources and receivers $\to$ bipartite graphs.

Empirical model validation

- Simulate bipartite 250×300 graph with $K = 5$ and $K' = 3$ obtained from $\mathbf{B} \in [0,1]^{K \times K'}$ with $B_{k\ell} \sim \text{Beta}(1.2, 1.2)$, $\boldsymbol{\theta} = (1/K, \ldots, 1/K)$, $\boldsymbol{\theta}' = (1/K', ..., 1/K')$, and $d = 2$.

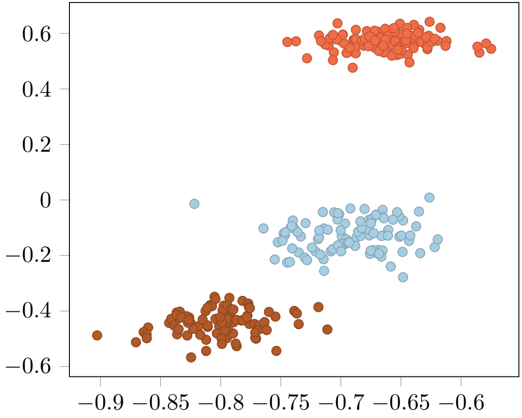


Figure 8. Scatterplot of the first two columns of $\hat{\mathbf{X}}'$.

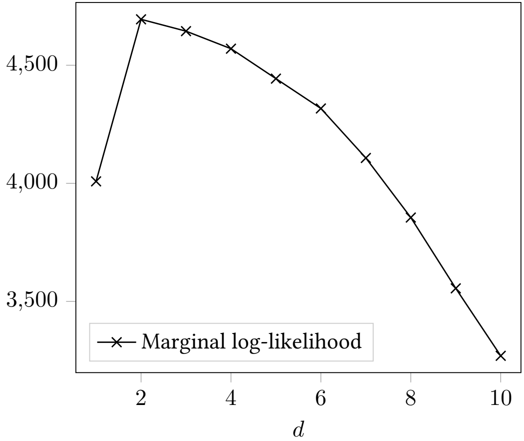


Figure 9. Marginal likelihood as a function of d.

Empirical model validation

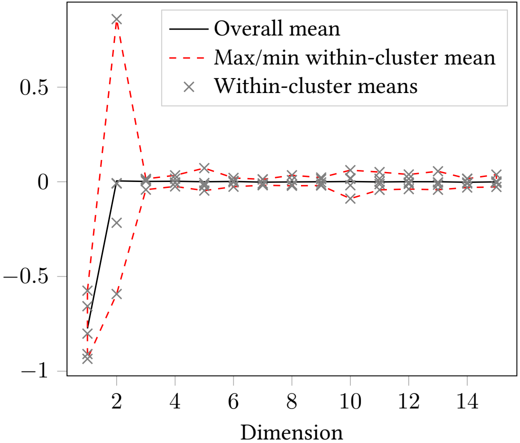


Figure 10. Within-cluster means of $\hat{\mathbf{X}}$.

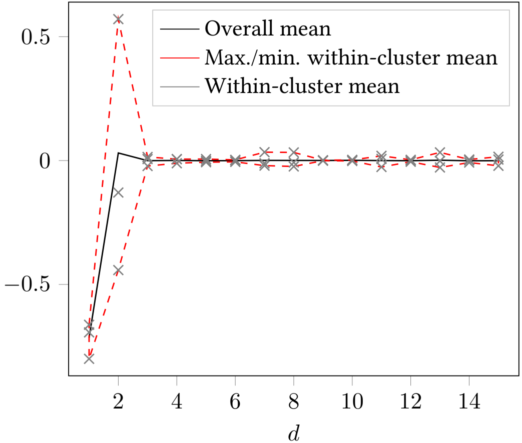


Figure 11. Within-cluster means of $\hat{\mathbf{X}}'$.

- Means are approximately $\mathbf{0}$ for columns with index $> d$, even for a relatively small graph.

Empirical model validation

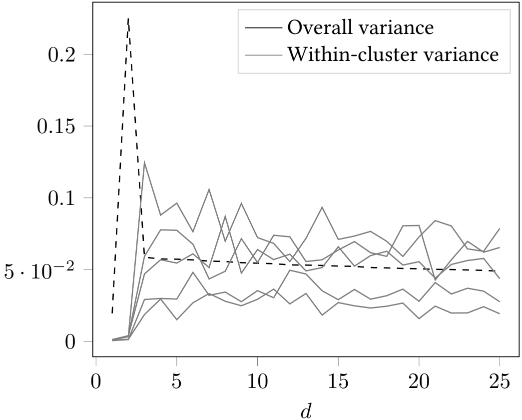


Figure 12. Within-cluster variances of $\hat{\mathbf{X}}$.

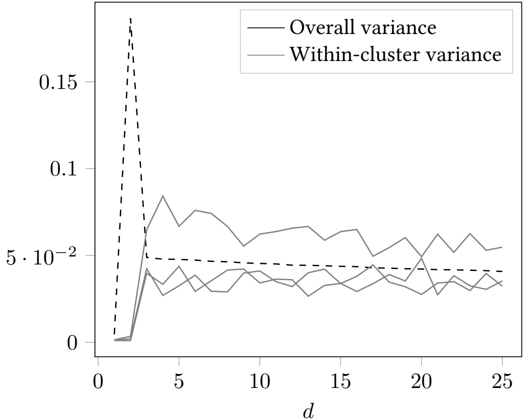


Figure 13. Within-cluster variances of $\hat{\mathbf{X}}'$.

- Different cluster-specific variances even for columns with index $> d$.
- Some evidence of second-order clustering.

Simulated data: parameter estimation

(d, K)	Model	$m = 25$		
		$\bar{d}$	$\bar{K}_\varnothing$	$\bar{H}_\varnothing$
$(2, 2)$	constrained, ASE	2.00	2.00	1.99
	unconstrained, ASE	2.00	2.00	1.99
	constrained, LSE	2.01	2.03	1.99
	unconstrained, LSE	2.02	2.02	1.99
$(2, 5)$	constrained, ASE	2.00	5.05	1.77
	unconstrained, ASE	2.00	5.07	1.80
	constrained, LSE	2.05	5.10	3.11
	unconstrained, LSE	2.07	5.11	3.10
$(6, 7)$	constrained, ASE	6.00	7.04	2.10
	unconstrained, ASE	6.00	7.05	2.20
	constrained, LSE	6.00	7.10	2.47
	unconstrained, LSE	6.00	7.07	2.39

Table 1. Results of the inferential procedure for undirected SBMs simulated using different (d, K) pairs, $n = 1{,}000$.

Simulated data: parameter estimation

(d, K)	Model	$m = 25$		
		$\bar{d}$	$\bar{K}_\varnothing$	$\bar{H}_\varnothing$
$(9, 9)$	constrained, ASE	8.97	9.01	2.08
	unconstrained, ASE	9.00	9.01	1.98
	constrained, LSE	9.00	9.02	2.12
	unconstrained, LSE	9.00	9.04	2.11
$(9, 12)$	constrained, ASE	9.00	12.02	1.96
	unconstrained, ASE	9.00	12.01	1.90
	constrained, LSE	9.00	12.03	2.60
	unconstrained, LSE	9.00	12.02	2.53
$(10, 15)$	constrained, ASE	10.00	14.78	1.25
	unconstrained, ASE	10.00	14.11	1.27
	constrained, LSE	10.00	14.81	1.81
	unconstrained, LSE	10.00	15.01	1.87

Table 2. Results of the inferential procedure for undirected SBMs simulated using different (d, K) pairs, $n = 1{,}000$.

Simulated data: effect of second-order clustering

(d,K)	m	H random				$H=K$		
		$\hat{d}$	$\hat{K}_\varnothing$	$\bar{H}_\varnothing$	ARI	$\hat{d}$	$\hat{K}_\varnothing$	ARI
$(3,5)$	15	3	5	1.669	1.000	3	5	1.000
	50	3	5	1.577	1.000	3	4	0.768
	150	3	5	1.467	1.000	3	4	0.768
	500	3	5	1.006	1.000	3	4	0.768
$(9,12)$	15	9	12	1.979	1.000	9	12	1.000
	50	9	12	1.912	1.000	9	12	1.000
	150	9	12	1.875	1.000	9	11	0.942
	500	9	12	1.388	1.000	9	5	0.517

Table 3. Results for the MCMC sampler on simulated undirected SBMs for different values of m, with and without second order clustering, $n=1{,}000$, assuming the unconstrained model.

Santander cycles data

- Aka *Boris bikes*.
- Santander cycles $\rightarrow$ bike sharing system in central London.
- £2 for access for 24 hours, first 30 minutes of each ride are free. Limited speed.
- Data freely available at `https://cycling.data.tfl.gov.uk/`, powered by TfL.
- One week of data: 5 – 11 September, 2018.
- $|V| = 783$ nodes/stations, $|E| = 69{,}153$ (excluding self-loops).

- Undirected graph:

$$A_{ij} = \begin{cases} 1 & \text{if at least one journey between stations } i \text{ and } j \text{ is completed,} \\ 0 & \text{otherwise.} \end{cases}$$

Santander cycles data: number of clusters

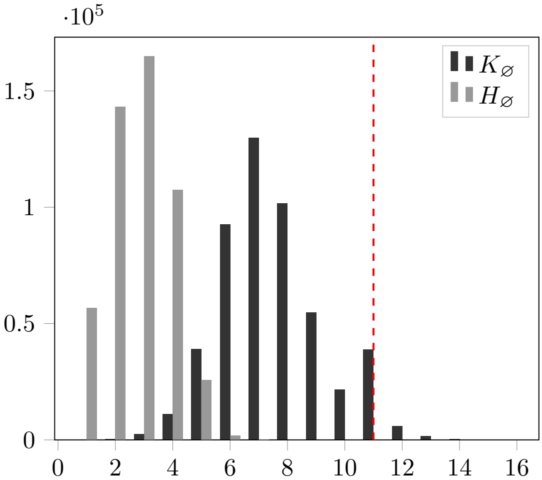


Figure 14. **Adjacency** embedding – Posterior histogram of $K_\varnothing$ and $H_\varnothing$, **unconstrained** model, MAP for d in **red**.

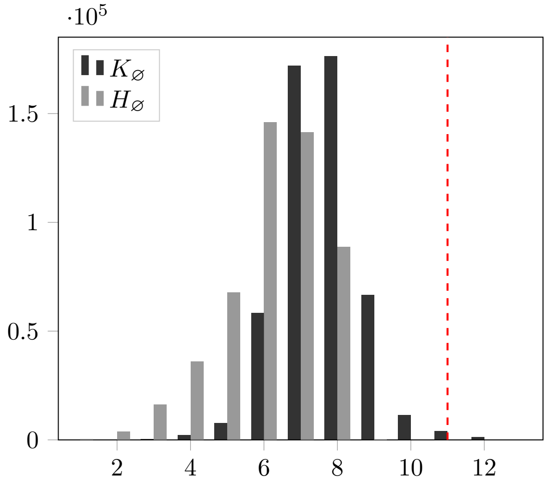


Figure 15. **Laplacian** embedding – Posterior histogram of $K_\varnothing$ and $H_\varnothing$, **unconstrained** model, MAP for d in **red**.

Santander cycles data: scree-plots

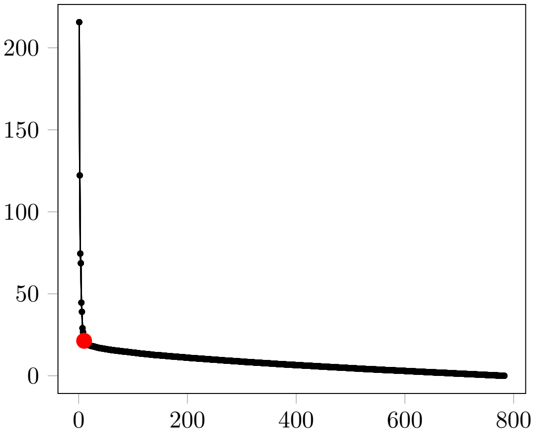


Figure 16. Magnitude of eigenvalues of the **adjacency** matrix.

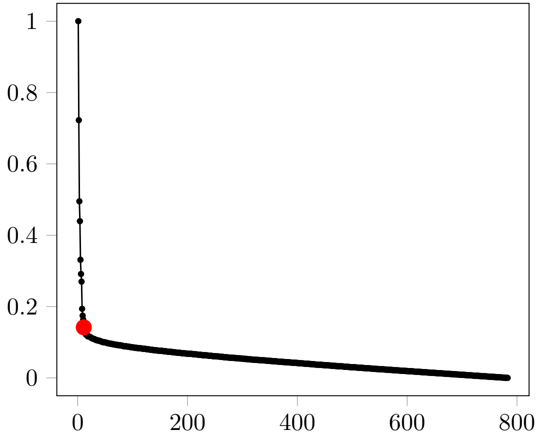


Figure 17. Magnitude of eigenvalues of the **Laplacian** matrix.

- Choice of d is consistent with the *elbow* of the scree-plot.

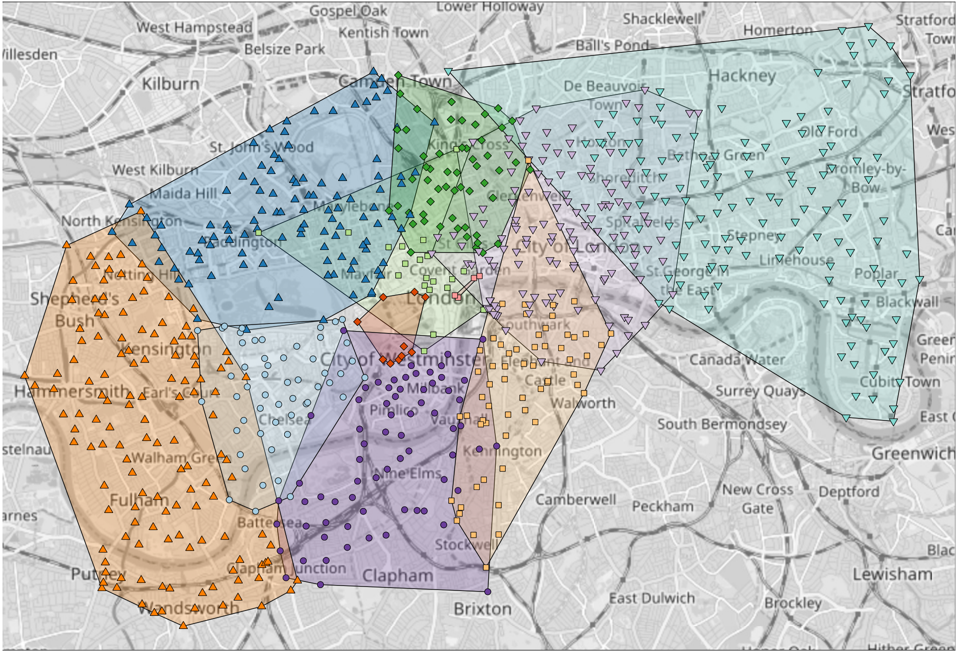


Figure 18. Adjacency embedding – Estimated communities for $K = 11$.

Enron e-mail network

- Corpus of e-mails sent by the employees of Enron corporation.
- Data freely available at `https://www.cs.cmu.edu/~enron/`.
- Version of dataset: May 7, 2015.
- $|V| = 184$ nodes/employees, $|E| = 3{,}010$.
- Extensively analysed in Priebe et al., 2005.

- Directed graph:

$$A_{ij} = \begin{cases} 1 & \text{if employee } i \text{ sends at least one e-mail to employee } j, \\ 0 & \text{otherwise.} \end{cases}$$

Image: Paul Rand, `https://commons.wikimedia.org/wiki/File:Logo_de_Enron.svg`.

Enron e-mail network: number of clusters

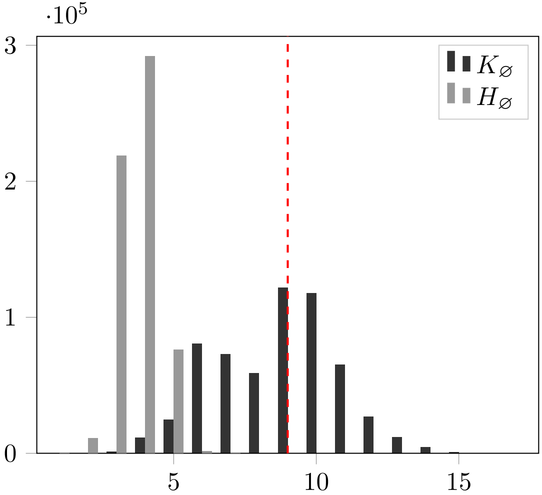


Figure 19. **ASE** – Posterior histogram of $K_\varnothing$ and $H_\varnothing$, **unconstrained** model, MAP for d in **red**.

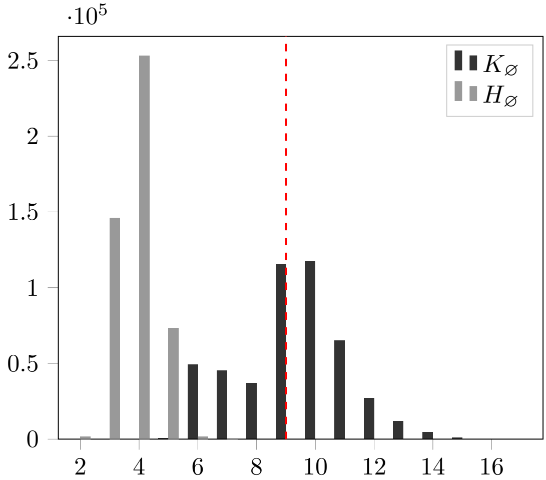


Figure 20. **ASE** – Posterior histogram of $K_\varnothing$ and $H_\varnothing$, **constrained** model, MAP for d in **red**.

Enron e-mail network: number of clusters

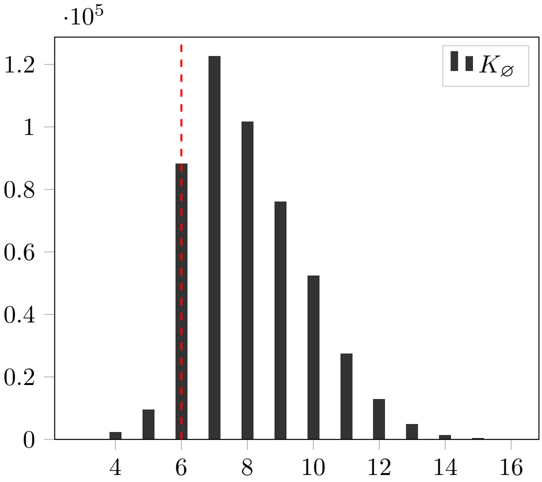


Figure 21. **ASE** – Posterior histogram of $K_\varnothing$, **unconstrained** model **without second order clustering**, MAP for d in **red**.

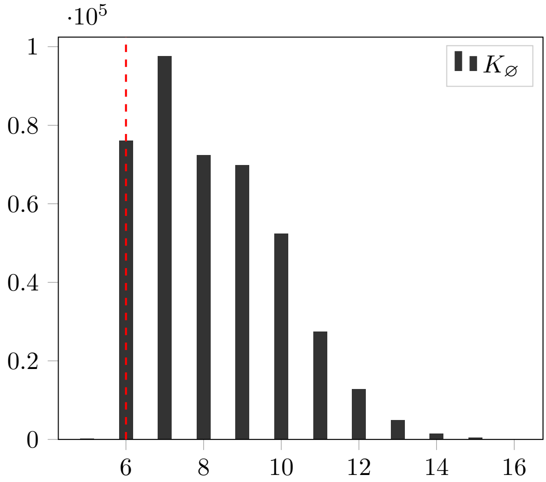


Figure 22. **ASE** – Posterior histogram of $K_\varnothing$, **constrained** model **without second order clustering**, MAP for d in **red**.

Enron e-mail network: scree-plot

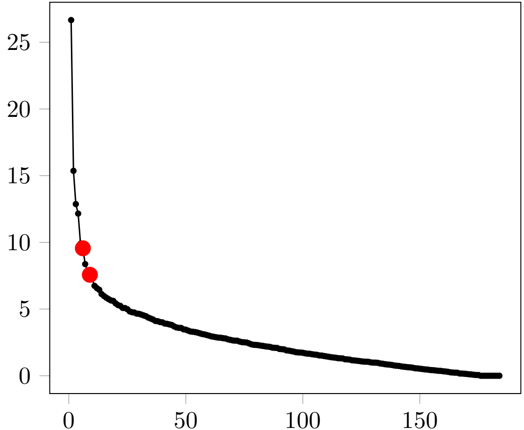


Figure 23. Singular values of the **adjacency** matrix.

- Choice of d is consistent with the *elbow* of the scree-plot.

Imperial College netflow data

- Bipartite graph of HTTP (port 80) and HTTPS (port 443) connections from machines hosted in computer labs at ICL.
- $439 \times 60{,}635$ nodes, $717{,}912$ links.
- Observation period: 1–31 January 2020.
- Periodic activity filtered according to opening hours of the buildings.
- Departments can be used as labels.
  - Chemistry,
  - Civil & Environmental Engineering,
  - Mathematics,
  - School of Medicine.

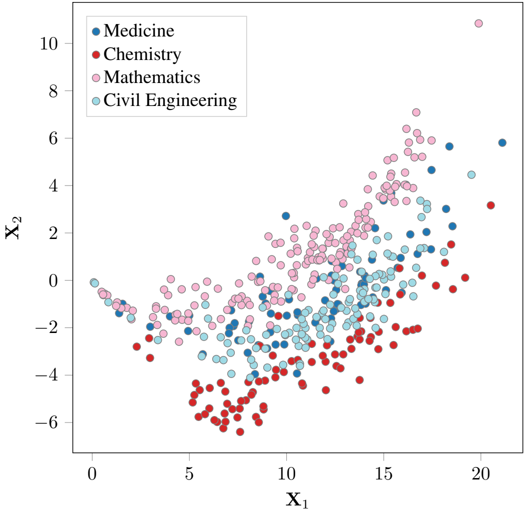


Figure 24. Scatterplot of $\mathbf{X}_{:2}$, coloured by department.

ICL WEB: EMBEDDINGS

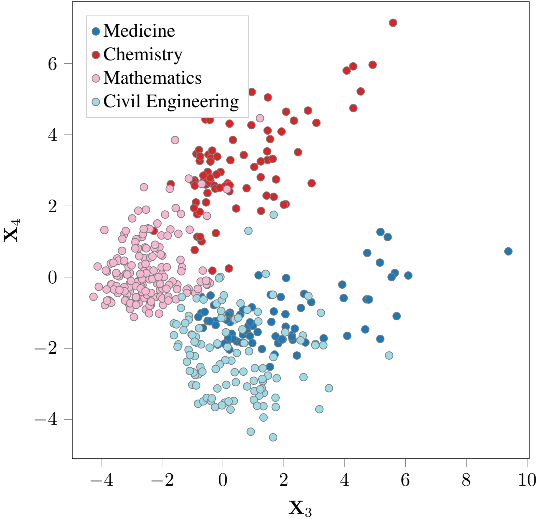


Figure 25. Scatterplot of $\mathbf{X}_3$ and $\mathbf{X}_4$, coloured by department.

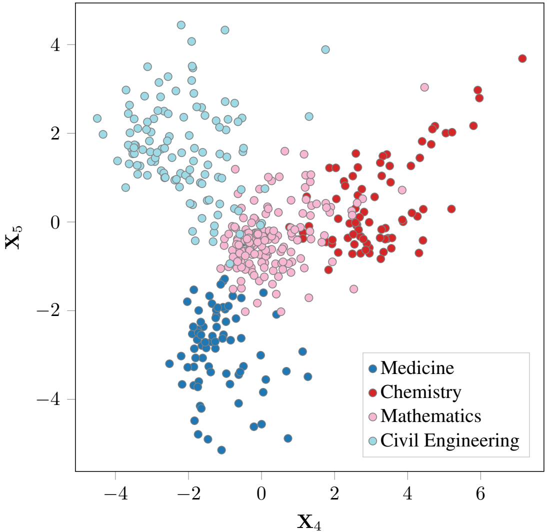


Figure 26. Scatterplot of $\mathbf{X}_4$ and $\mathbf{X}_5$, coloured by department.

ICL web: effect of out-degree

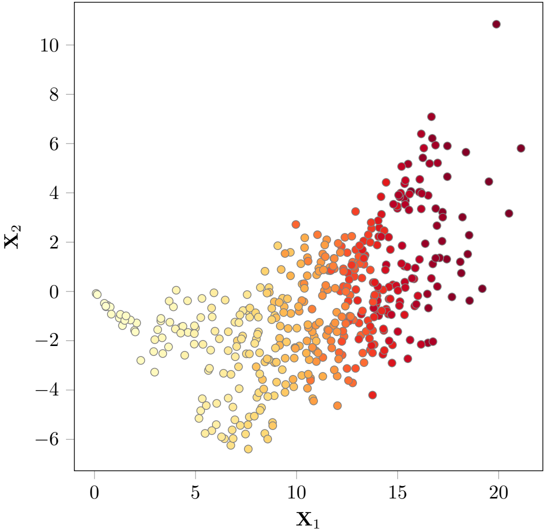


Figure 27. Scatterplot of $\mathbf{X}_1$ and $\mathbf{X}_2$, coloured by out-degree percentile.

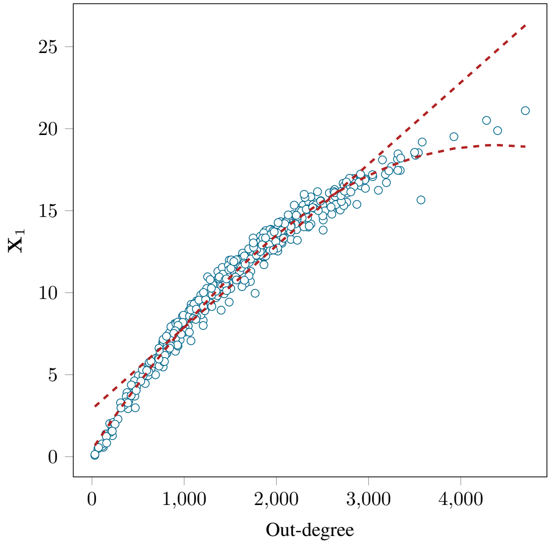


Figure 28. Scatterplot of $\mathbf{X}_1$ versus out-degree of the node.

ICL WEB: NUMBER OF CLUSTERS

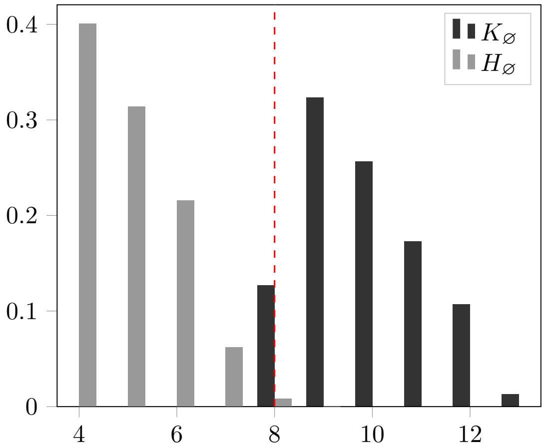


Figure 29. Posterior histogram of $K_\varnothing$ and $H_\varnothing$, **constrained** model **with second order clustering**, MAP for d in **red**.

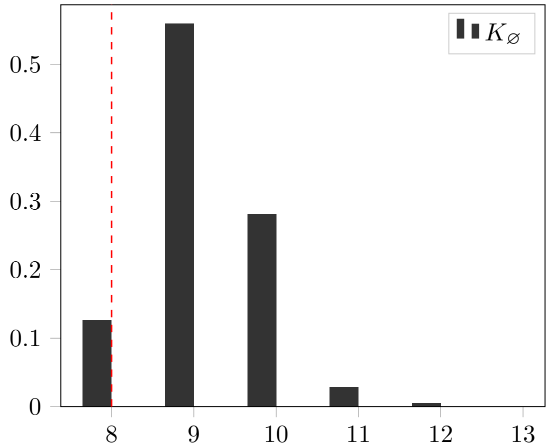


Figure 30. Posterior histogram of $K_\varnothing$, **constrained** model **without second order clustering**, MAP for d in **red**.

ICL WEB: NUMBER OF CLUSTERS

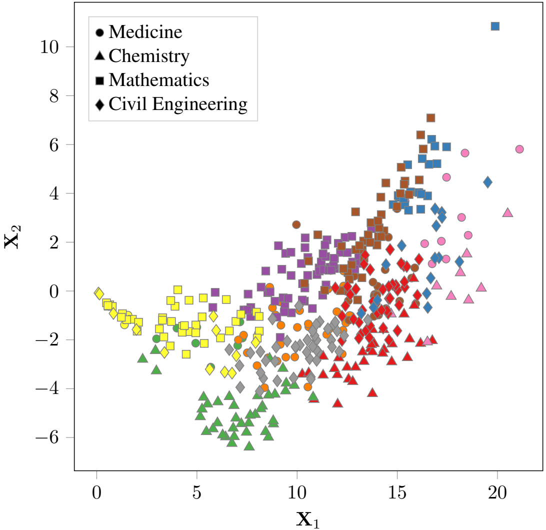


Figure 31. Scatterplot of $\mathbf{X}_1$ and $\mathbf{X}_2$, labelled by estimated clustering ($K = 9$) and department.

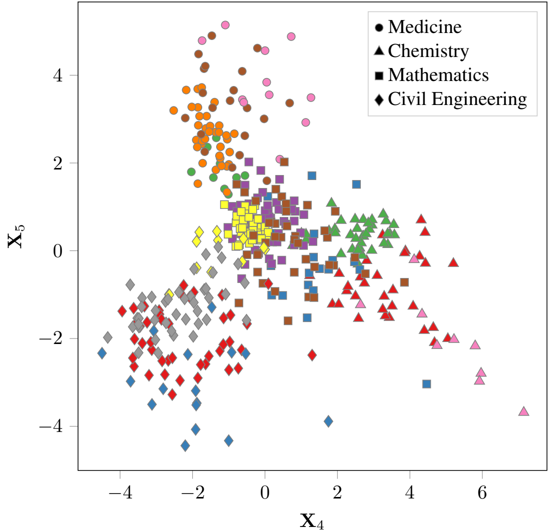


Figure 32. Scatterplot of $\mathbf{X}_4$ and $\mathbf{X}_5$, labelled by estimated clustering ($K = 9$) and department.

Beyond SBMs: the degree-corrected SBM (DCSBM)

- Problem: SBMs do not account for **within-community degree heterogeneity**.
- Solution: **degree-corrected** stochastic blockmodel (DCSBM, Karrer and Newman, 2011).
- Assign a correction $\rho_i \in (0,1)$ to each node.
- Model adjacency matrix as:

$$\mathbb{P}(A_{ij}=1) = \rho_i \rho_j \boldsymbol{\mu}_{z_i}^\top \boldsymbol{\mu}_{z_j}, i<j, A_{ij}=A_{ji}.$$

- Theory predicts that DCSBM embeddings have K rays from the origin.
- How to do spectral clustering in this setting? **More on this (hopefully) coming soon!**

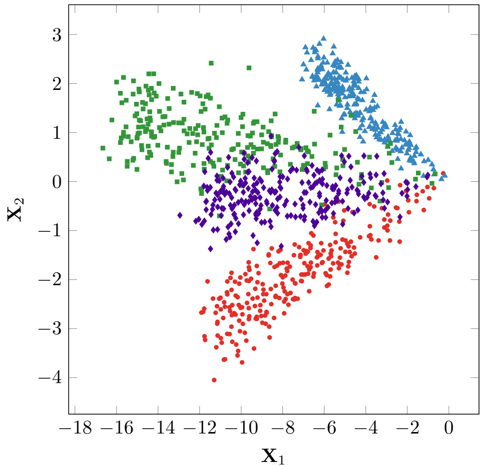
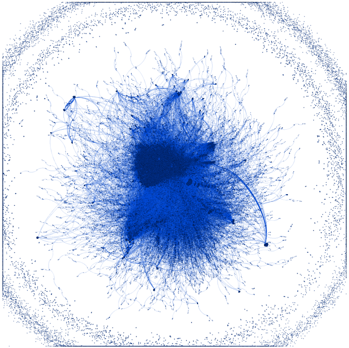


Figure 33. Scatterplot of the initial 2 dimensions of the ASE for a simulated DCSBM with $n = 1{,}000$, $K = 4$, and degree corrections $\rho_i \sim \mathrm{Beta}(2,1)$.

Conclusion

- Community detection and stochastic blockmodels:
  - Bayesian model for simultaneous selection of K and d in generalised random dot product graphs,
  - Allow for initial misspecification of the arbitrarily large parameter m, then refine estimate d,
  - Gaussian mixture model (with constraints) based on spectral embedding,
  - Easy to extend to directed and bipartite graphs.
- More details:
  Sanna Passino and Heard, 2019 – arXiv: 1904.05333.
- What's next: simultaneous model selection of d and K in spectral clustering under the DCSBM.



References I

- Athreya, A. et al. (2018). "Statistical Inference on Random Dot Product Graphs: a Survey". In: *Journal of Machine Learning Research* 18.226, pp. 1–92.
- Cape, J., M. Tang, and C. E. Priebe (2018). "On spectral embedding performance and elucidating network structure in stochastic block model graphs". In: *arXiv e-prints*, arXiv:1808.04855, arXiv:1808.04855. arXiv: `1808.04855 [math.ST]`.
- Hoff, P. D, A. E. Raftery, and M. S. Handcock (2002). "Latent space approaches to social network analysis". In: *Journal of the American Statistical Association* 97.460, pp. 1090–1098.
- Holland, P. W., K. B. Laskey, and S. Leinhardt (1983). "Stochastic blockmodels: First steps". In: *Social Networks* 5.2, pp. 109 –137.
- Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Series in Statistics. Springer.
- Karrer, B. and M. E. J. Newman (2011). "Stochastic blockmodels and community structure in networks". In: *Phys. Rev. E* 83 (1).
- Priebe, C. E. et al. (2005). "Scan Statistics on Enron Graphs". In: *Computational & Mathematical Organization Theory* 11.3, pp. 229–247.
- Raftery, A. E. and N. Dean (2006). "Variable Selection for Model-Based Clustering". In: *Journal of the American Statistical Association* 101.473, pp. 168–178.

References II

- Rohe, K., S. Chatterjee, and B. Yu (2011). "Spectral clustering and the high-dimensional stochastic blockmodel". In: *Annals of Statistics* 39.4, pp. 1878–1915.
- Rubin-Delanchy, P., N.M. Adams, and N.A. Heard (2016). "Disassortativity of computer networks". In: *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, pp. 243–247.
- Rubin-Delanchy, P. et al. (2017). "A statistical interpretation of spectral embedding: the generalised random dot product graph". In: *ArXiv e-prints*. arXiv: 1709.05506.
- Sanna Passino, F. and N. A. Heard (2019). "Bayesian estimation of the latent dimension and communities in stochastic blockmodels". In: *arXiv e-prints*. arXiv: 1904.05333.
- Yang, C. et al. (2019). "Simultaneous dimensionality and complexity model selection for spectral graph clustering". In: *arXiv e-prints*. arXiv: 1904.02926.
- Young, S. J. and E. R. Scheinerman (2007). "Random Dot Product Graph Models for Social Networks". In: *Algorithms and Models for the Web-Graph*. Ed. by A. Bonato and F. R. K. Chung. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 138–149.
- Zhu, M. and A. Ghodsi (2006). "Automatic dimensionality selection from the scree plot via the use of profile likelihood". In: *Computational Statistics & Data Analysis* 51.2, pp. 918 –930.

Identifiability of the GRDPG

- The GRDPG has **two sources of non-identifiability** (Cape, Tang, and Priebe, 2018).

1. ***Identifiability: uniqueness up to indefinite orthogonal transformations***
   For any matrix $\mathbf{Q} \in \mathbb{O}(d_+, d_-)$, the indefinite orthogonal group with signature (d_+, d_-),

$$(\mathbf{Q}\boldsymbol{\mu}_{z_i})^\top \mathbf{I}(d_+, d_-)(\mathbf{Q}\boldsymbol{\mu}_{z_j}) = \boldsymbol{\mu}_{z_i}^\top \mathbf{I}(d_+, d_-)\boldsymbol{\mu}_{z_j},$$

   which implies that the likelihood is invariant to any such transformation.

2. ***Uniqueness up to artificial dimension blow-up***
   For $(\mathbf{A}, \mathbf{X}) \sim \mathrm{GRDPG}_{d_+,d_-}(\mathcal{F})$, there exists $\mathcal{F}^\star$ on $\mathbb{R}^{d^\star}$, with $d^\star > d$, such that

$$(\mathbf{A}, \mathbf{X}) \overset{d}{=} (\mathbf{A}^\star, \mathbf{X}^\star) \text{ with } (\mathbf{A}^\star, \mathbf{X}^\star) \sim \mathrm{GRDPG}_{d_+^\star,d_-^\star}(\mathcal{F}^\star).$$

   In the SBM setting, this essentially means that **any** matrix $\mathbf{B} \in [0,1]^{K\times K}$ with rank d can be obtained as an inner product between latent positions on **arbitrarily large** dimensions.

ASE and SBMs: an example

- Simulate a 2-block stochastic blockmodel using the within-community probability matrix

$$\mathbf{B} = \begin{bmatrix} 0.02 & 0.03 \\ 0.03 & 0.01 \end{bmatrix}.$$

- Eigenvalues: $\lambda_1 \approx 0.045$ and $\lambda_2 \approx -0.015 \Rightarrow$ GRDPG ($\mathbf{B}$ is indefinite).
- Simulate the community allocations under two settings:
  - $\boldsymbol{\theta} = (0.5, 0.5)$ (balanced communities),
  - $\boldsymbol{\theta} = (0.9, 0.1)$ (unbalanced communities).
- Simulate two adjacency matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ under both settings, for $n = 4{,}000$.
- Take ASE of $\mathbf{A}_1$ and $\mathbf{A}_2$ in $\mathbb{R}^2$, say $\hat{\mathbf{X}}_1$ and $\hat{\mathbf{X}}_2$.

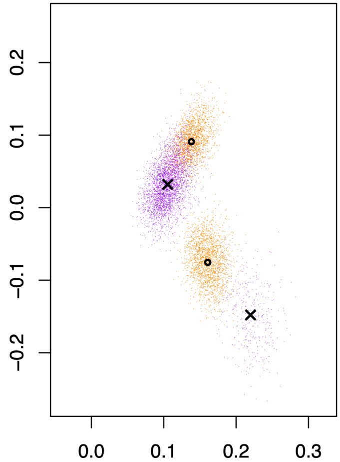


Figure 34. ASEs of simulated 2-block SBMs with same $\mathbf{B}$, but two different $\boldsymbol{\theta}$. Illustration from Rubin-Delanchy et al., 2017.

ASE and SBMs: an example of the role of $\mathbf{Q}_n$

- In the simulation, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are **known**.
- The **purple** point cloud $\hat{\mathbf{X}}_2$ is reconfigured, and aligned to the **orange** point cloud $\hat{\mathbf{X}}_1$, using two (indefinite) orthogonal transformations estimated from the two ASEs.
- The two representations of the **purple** point cloud are **equivalent**.
- In the CLT, $\mathbf{Q}_n$ is *unidentifiable*, but it *materially affects (Euclidean) distances* between points.
- The picture confirms that GMMs are preferable over K-means.

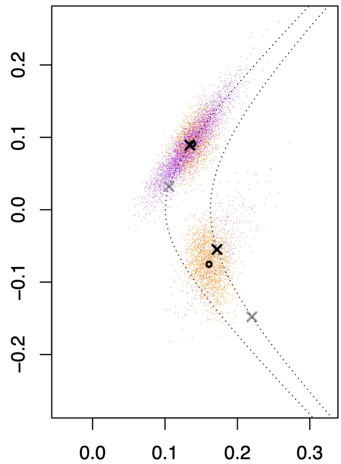


Figure 35. Transformed ASEs of simulated 2-block SBMs with same $\mathbf{B}$, but two different $\boldsymbol{\theta}$. Illustration from Rubin-Delanchy et al., 2017.