Statistics Seminars – Alma Mater Studiorum, Università di Bologna

Model selection and latent substructure inference in spectral graph clustering

Imperial College London

Francesco Sanna Passino

Department of Mathematics, Imperial College London

f.sannapassino@imperial.ac.uk, fraspass@outlook.com

20th May, 2020

Joint work with:

- **Professor Nick Heard**
  Department of Mathematics, Imperial College London
- **Dr Patrick Rubin-Delanchy**
  School of Mathematics & Statistics, University of Bristol

Acknowledgements:

- **Dr Joshua Neil, Dr Melissa Turcotte**
  Microsoft 365 Defender, Microsoft Corporation (Redmond, WA)

More details about this work:

Sanna Passino, F. and N. A. Heard (2020). “Bayesian estimation of the latent dimension and communities in stochastic blockmodels”. In: *Statistics and Computing* 30.5, pp. 1291–1307.

Sanna Passino, F. and N. A. Heard (2021). “Latent structure blockmodels for Bayesian spectral graph clustering”. In: *arXiv e-prints (forthcoming).*

Sanna Passino, F., N. A. Heard, and P. Rubin-Delanchy (2020). “Spectral clustering on spherical coordinates under the degree-corrected stochastic blockmodel”. In: *arXiv e-prints.* arXiv: 2011.04558 [stat.ML].

Graphs

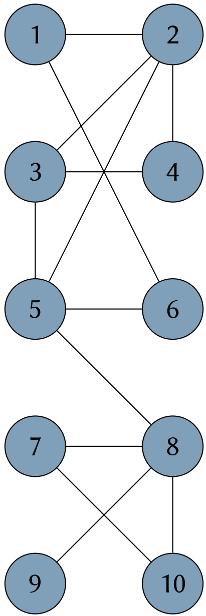
- **Graph** $\mathbb{G} = (V, E)$ where:
  - V is the **node set**, $n = |V|$,
  - $E \subseteq V \times V$ is the **edge set**, containing dyads $(i,j),\ i,j \in V$.
- An edge is drawn if a node $i \in V$ connects to $j \in V$, written $(i,j) \in E$.
  - If the graph is **undirected**, then $(i,j) \in E \Leftrightarrow (j,i) \in E$.
  - For **directed** graphs, $(i,j) \in E \not\Leftrightarrow (j,i) \in E$.
  - For **bipartite** graphs $(i,j) \in E \Leftrightarrow i \in V_1, j \in V_2$, with $V_1 \cap V_2 = \varnothing, V_1 \cup V_2 = V$.
- 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.
- For bipartite graphs, a **rectangular** adjacency matrix $\mathbf{A} \in \{0,1\}^{|V_1| \times |V_2|}$ is preferred.

A toy example

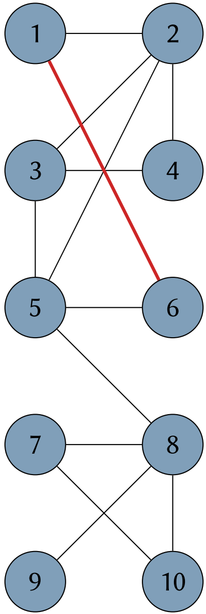
$\mathbb{G} =$



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

A toy example

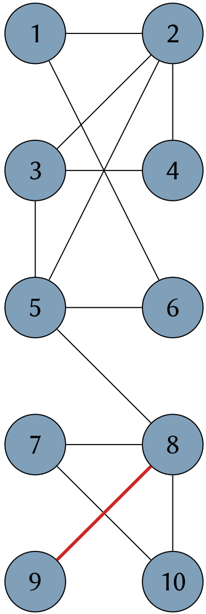
$\mathbb{G} =$



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

A toy example

$\mathbb{G} =$



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

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 $\mathcal{X}$.
- The edges are generated *independently* using a **kernel function** $\kappa : \mathcal{X} \times \mathcal{X} \to [0,1]$:

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

- The latent positions are represented as a $(n \times d)$-dimensional matrix $\mathbf{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n]^\intercal$.
- 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 $\mathcal{X}$ is chosen such that $0 \leq \boldsymbol{x}^\intercal \boldsymbol{x}' \leq 1\ \forall\, \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$:

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

- In RDPGs, the latent dimension has a nice interpretation: $d = \text{rank}\{\mathbb{E}(\mathbf{A})\} = \text{rank}(\mathbf{X}\mathbf{X}^\intercal)$.

RDPG and ASE

Definition (Random dot product graph – RDPG, Young and Scheinerman, 2007)

For an integer d, let F be a probability measure supported on $\mathcal{X} \subset \mathbb{R}^d$, where $\mathcal{X}$ is a d-dimensional inner product distribution, such that $\boldsymbol{x}^\intercal\boldsymbol{x}' \in [0,1]\ \forall\ \boldsymbol{x},\boldsymbol{x}' \in \mathcal{X}$. Furthermore, let $\mathbf{A} \in \{0,1\}^{n\times n}$ be a symmetric binary matrix and $\mathbf{X} = (\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)^\intercal \in \mathcal{X}^n$. Then $(\mathbf{A},\mathbf{X}) \sim \text{RDPG}_d(F^n)$ if $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n \overset{iid}{\sim} F$ and for $i<j$, independently,

$$\mathbb{P}(A_{ij}=1 \mid \boldsymbol{x}_i,\boldsymbol{x}_j) = \boldsymbol{x}_i^\intercal\boldsymbol{x}_j.$$

Definition (ASE – Adjacency spectral embedding)

For a given integer $d \in \{1,\dots,n\}$ and a symmetric adjacency matrix $\mathbf{A} \in \{0,1\}^{n\times n}$, the d-dimensional adjacency spectral embedding (ASE) $\hat{\mathbf{X}} = [\hat{\boldsymbol{x}}_1,\dots,\hat{\boldsymbol{x}}_n]^\intercal$ of $\mathbf{A}$ is

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

where $\boldsymbol{\Lambda}$ is a $d\times d$ diagonal matrix containing the absolute values of the d largest eigenvalues in magnitude, and $\boldsymbol{\Gamma}$ is a $n\times d$ matrix containing the corresponding eigenvectors.

A simple example: a Hardy-Weinberg graph

- Each node is given a latent score $\phi_i \in [0,1],\ i=1,\dots,n$.
- The latent positions $\boldsymbol{x}_i \in \mathbb{R}^3$ are uniquely determined from ϕ_i:

$$\boldsymbol{x}_i = (\phi_i^2, 2\phi_i(1-\phi_i), (1-\phi_i)^2).$$

- Graphs are simulated for $n \in \{100, 1000, 5000\}$ and $\phi_i \sim \text{Unif}(0,1)$.
- ASE is calculated for $d=3$ from the adjacency matrices.
- The true latent positions are coloured in **black**, whereas their estimates are in **blue**.

(a) $n=100$ (b) $n=1000$ (c) $n=5000$

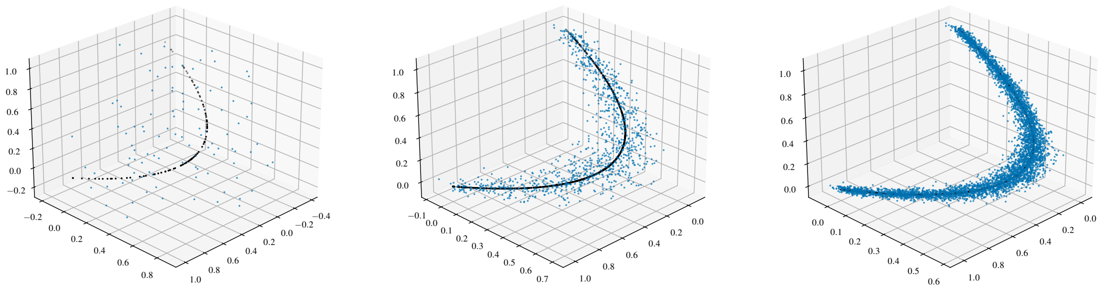


Figure 1. 3-dimensional ASE from a simulated Hardy-Weinberg graph with $\phi_i \sim \text{Unif}(0,1)$ for $n \in \{100, 1000, 5000\}$.

Central limit theorem for ASE

Theorem (ASE central limit theorem)

Let $(\mathbf{A}^{(n)}, \mathbf{X}^{(n)}) \sim$ *RDPG*$_d(F^n), n = 1, 2, \dots,$ *be a sequence of adjacency matrices and corresponding latent positions, and let* $\hat{\mathbf{X}}^{(n)}$ *be the* d*-dimensional ASE of* $\mathbf{A}^{(n)}$*. For an integer* $m > 0$*, and for the sequences of points* $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathcal{X}$ *and* $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m \in \mathbb{R}^d$*, there exists a sequence of orthogonal matrices* $\mathbf{Q}_1, \mathbf{Q}_2, \dots \in \mathbb{O}(d)$ *such that for* $n \to \infty$*:*

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

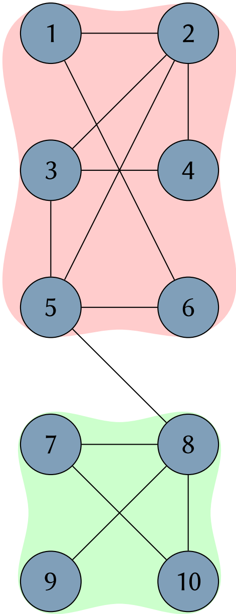
where $\Phi\{\cdot\}$ *is the CDF of a* d*-dimensional normal distribution, and* $\boldsymbol{\Sigma}(\cdot)$ *is a covariance matrix which depends on the true value of the latent position.*

- References: Athreya et al., 2016; Rubin-Delanchy et al., 2017; Athreya et al., 2018.
- The theorem has *crucial* relevance in practice. Approximately, for n large:

$$\hat{\boldsymbol{x}}_i \approx \mathbb{N}\{\mathbf{Q}_n^\intercal \boldsymbol{x}_i, n^{-1}\mathbf{Q}_n^\intercal \boldsymbol{\Sigma}(\boldsymbol{x}_i)\mathbf{Q}_n\}.$$

Graph clustering / Community detection

$\mathbb{G} =$



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

RDPGs and spectral clustering

- **Spectral clustering** (Ng, Jordan, and Weiss, 2001; von Luxburg, 2007) is one of the most popular methods for community detection (Fortunato, 2010).

Algorithm: Spectral clustering

Input: adjacency matrix $\mathbf{A}$, dimension d, and number of communities K.

1 from $\mathbf{A}$, compute ASE $\hat{\mathbf{X}} = [\hat{\boldsymbol{x}}_1, \dots, \hat{\boldsymbol{x}}_n]^\intercal$ (von Luxburg, 2007) or its row-normalised version $\tilde{\mathbf{X}} = [\tilde{\boldsymbol{x}}_1, \dots, \tilde{\boldsymbol{x}}_n]^\intercal$ (Ng, Jordan, and Weiss, 2001) into $\mathbb{R}^d$,

2 fit a clustering model (e.g. GMM, k-means, hierarchical clustering) with K components on the d-dimensional embedding space.

Result: node memberships $z_1, \dots, z_n$.

- The theory holds on the assumption that d and K are **known**.
  - In practice the two parameters are estimated **sequentially**. This is **sub-optimal**.
    - 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.
- Different embeddings imply **different modelling choices** under a RDPG perspective.
  - $\mathbf{X}$ + GMM $=$ stochastic blockmodel (SBM; Holland, Laskey, and Leinhardt, 1983),
  - $\tilde{\mathbf{X}}$ + GMM $\approx$ degree-corrected stochastic blockmodel (DCSBM; Karrer and Newman, 2011),
  - SBMs and DCSBMs assume fairly simple community structure under the RDPG: what if the communities have **complex latent substructure**?

SBMs and DCSBMs

- The **stochastic blockmodel** (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 $\boldsymbol{\psi} = (\psi_1,\dots,\psi_K)$, from the $K-1$ probability simplex.
- The probability of a link depends on the **community allocations** z_i and z_j of the nodes:

$$\mathbb{P}(A_{ij}=1) = B_{z_i z_j}.$$

- Real-world networks often present **within-community degree heterogeneity**. In this case, **degree-corrected stochastic blockmodels** (Karrer and Newman, 2011) are more appropriate. Each node is given a degree-correction parameter $\rho_i \in (0,1)$ such that:

$$\mathbb{P}(A_{ij}=1) = \rho_i \rho_j B_{z_i z_j}.$$

SBMs and DCSBMs as special cases of RDPGs

- SBMs and DCSBMs can be interpreted as a **special cases** of RDPGs.
- For simplicity, initially assume that $\mathbf{B}$ is *positive semi-definite.*
- Let $B_{kh} = \boldsymbol{\mu}_k^\intercal \boldsymbol{\mu}_h$ for some $\boldsymbol{\mu}_k, \boldsymbol{\mu}_h \in \mathcal{X}$.
- If the nodes in community k are assigned the latent position $\boldsymbol{\mu}_k$, then, for the SBM:

$$\mathbb{P}(A_{ij}=1) = B_{z_i z_j} = \boldsymbol{\mu}_{z_i}^\intercal \boldsymbol{\mu}_{z_j}.$$

- Extension to *any* $\mathbf{B}$: generalised RDPG (GRDPG, Rubin-Delanchy et al., 2017).
- For the DCSBM, it is assumed that $\boldsymbol{x}_i = \rho_i \boldsymbol{\mu}_{z_i}$, which gives:

$$\mathbb{P}(A_{ij}=1) = \rho_i \rho_j B_{z_i z_j} = \rho_i \rho_j \boldsymbol{\mu}_{z_i}^\intercal \boldsymbol{\mu}_{z_j}.$$

- **Inference** on SBMs and DCSBMs as (G)RDPGs:
  - Latent dimension d,
  - Number of communities K,
  - Community allocations $\boldsymbol{z} = (z_1, \dots, z_n)$,
  - Nuisance parameters: latent positions $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K$, degree-correction parameters $\rho_1, \dots, \rho_n$.
- **This talk discusses a novel framework for joint estimation of d and K.**

ASE of SBMs and DCSBMs

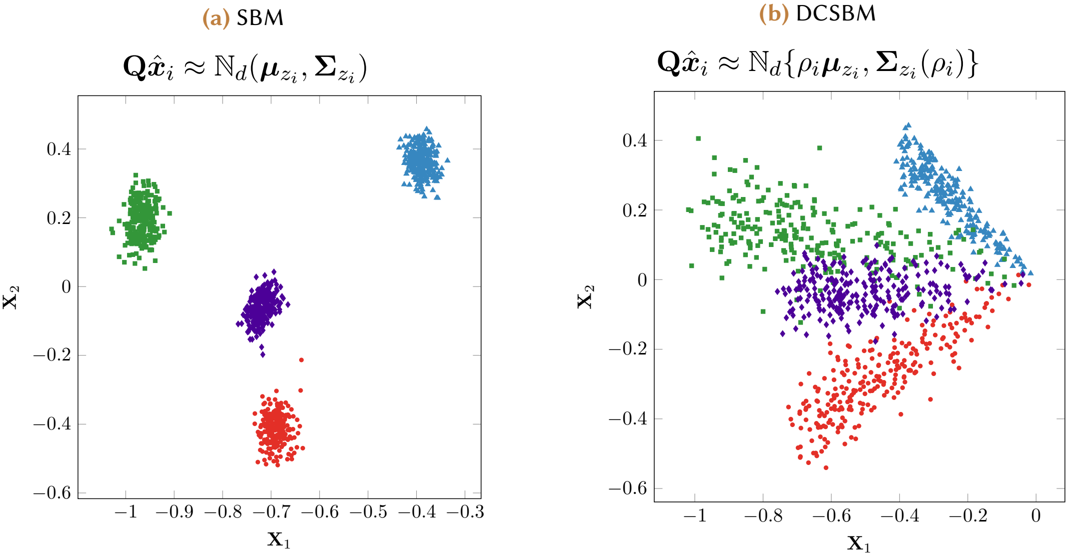


Figure 2. Scatterplot of the 2-dimensional ASE for a simulated SBM with $d = K = 4$, $\mathbf{B} \sim$ Uniform$(0,1)^{K\times K}$, and 100 nodes per community, and corresponding DCSBM corrected with $\rho_i \sim$ Beta$(2,1)$.

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 "extended" embedding $\hat{\mathbf{X}} = [\hat{\boldsymbol{x}}_1, \dots, \hat{\boldsymbol{x}}_n]^\intercal \in \mathbb{R}^{n\times m},\ \boldsymbol{x}_i \in \mathbb{R}^m$ for some m.
  - *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 SBM network embeddings

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

$$
\begin{aligned}
\hat{\boldsymbol{x}}_i | d, z_i, \boldsymbol{\mu}_{z_i}, \boldsymbol{\Sigma}_{z_i}, \boldsymbol{\sigma}^2_{z_i} &\sim \mathbb{N}_m\left(\begin{bmatrix}\boldsymbol{\mu}_{z_i} \\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{z_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma}^2_{z_i}\mathbf{I}_{m-d}\end{bmatrix}\right),\ i = 1,\dots,n, \\
(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)|d &\overset{iid}{\sim} \mathrm{NIW}_d(\mathbf{0}, \kappa_0, \nu_0 + d - 1, \boldsymbol{\Delta}_d),\ k = 1,\dots,K, \\
\sigma^2_{kj} &\overset{iid}{\sim} \text{Inv-}\chi^2(\lambda_0, \sigma_0^2),\ j = d+1,\dots,m, \\
d|\boldsymbol{z} &\sim \mathrm{Uniform}\{1,\dots,K_\varnothing\}, \\
z_i|\boldsymbol{\psi} &\overset{iid}{\sim} \mathrm{Discrete}(\boldsymbol{\psi}),\ i = 1,\dots,n,\ \boldsymbol{\psi} \in \mathcal{S}_{K-1}, \\
\boldsymbol{\psi}|K &\sim \mathrm{Dirichlet}\left(\frac{\alpha}{K},\dots,\frac{\alpha}{K}\right), \\
K &\sim \mathrm{Geometric}(\omega).
\end{aligned}
$$

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

- Alternative: $d \sim \mathrm{Geometric}(\delta)$.
- Yang et al., 2021, independently and simultaneously proposed a similar frequentist model.

Empirical model validation

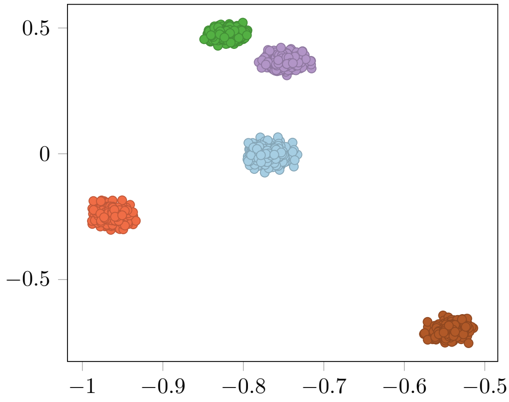


Figure 3. Scatterplot of the columns $\hat{\mathbf{X}}_1$ and $\hat{\mathbf{X}}_2$ of the ASE.

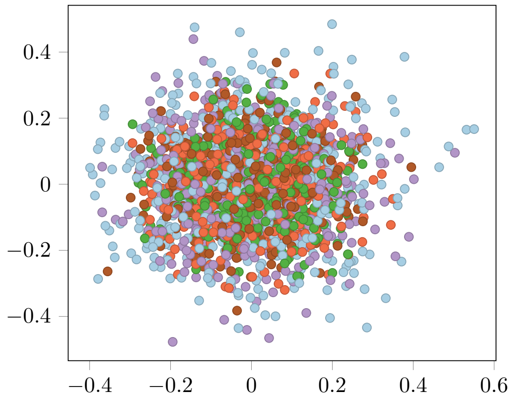


Figure 4. Scatterplot of the columns $\hat{\mathbf{X}}_3$ and $\hat{\mathbf{X}}_4$ of the ASE.

- Simulated GRDPG-SBM with $n = 2500$, $d = 2$, $K = 5$.
- Nodes allocated to communities with probability $\psi_k = \mathbb{P}(z_i = k) = 1/K$.

Empirical model validation

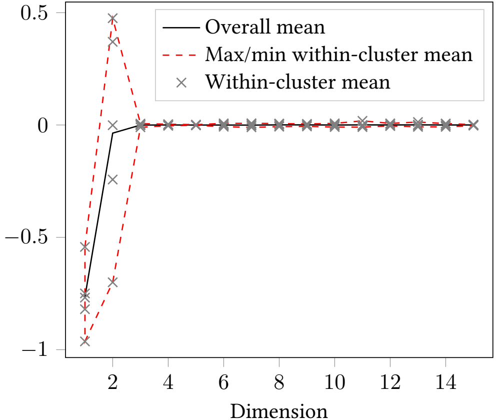


Figure 5. Within-cluster and overall means of $\hat{\mathbf{X}}_{:15}$.

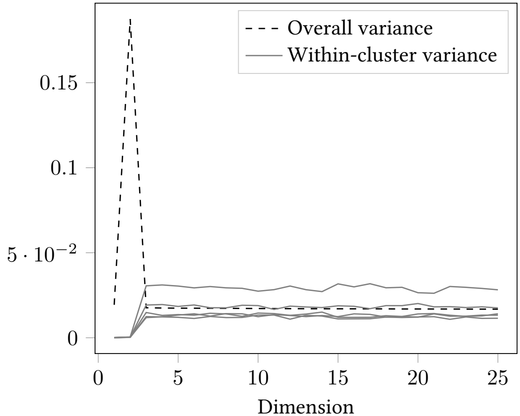


Figure 6. Within-cluster variance of $\hat{\mathbf{X}}_{:25}$.

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

Empirical model validation

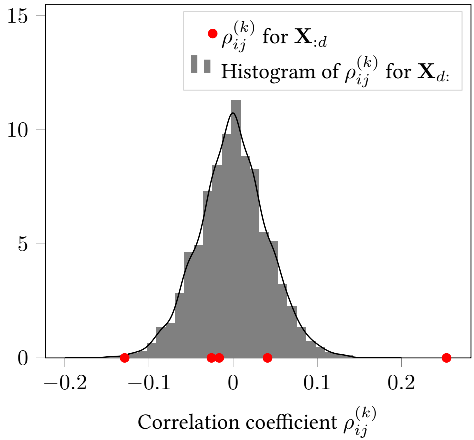


Figure 7. Within-cluster correlation coefficients of $\hat{\mathbf{X}}_{:30}$.

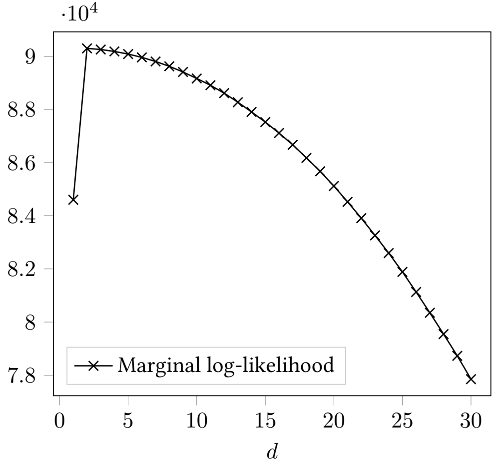


Figure 8. Marginal likelihood as a function of d.

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

Inference

- **Integrate out nuisance parameters** $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$, σ^2_{jk} and $\boldsymbol{\psi}$ $\rightarrow$ inference on d, K and $\boldsymbol{z}$.
- Inference via MCMC: **collapsed Metropolis-within-Gibbs sampler** $\rightarrow$ 4 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.
- **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** $\rightarrow$ might be worth exploring constrained and unconstrained models.
- 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 $\mathbf{A} = \hat{\mathbf{U}}\hat{\mathbf{D}}\hat{\mathbf{V}}^\intercal + \hat{\mathbf{U}}_\perp\hat{\mathbf{D}}_\perp\hat{\mathbf{V}}_\perp^\intercal$ 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.
- $\hat{\mathbf{X}}$ and $\hat{\mathbf{X}}'$ could also be analysed *separately*.

ICL NetFlow data

- Bipartite graph of HTTP (port 80) and HTTPS (port 443) connections from machines hosted in computer labs at ICL.
- 439×60635 nodes, 717912 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.
- $K = 4$.

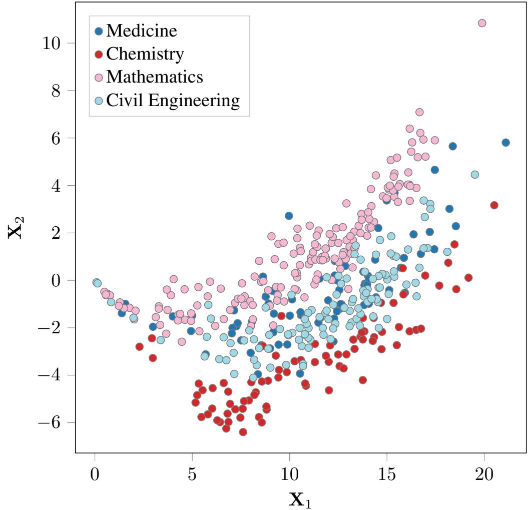


Figure 9. Scatterplot of $\hat{\mathbf{X}}_{:2}$, coloured by department.

ICL NetFlow: embeddings

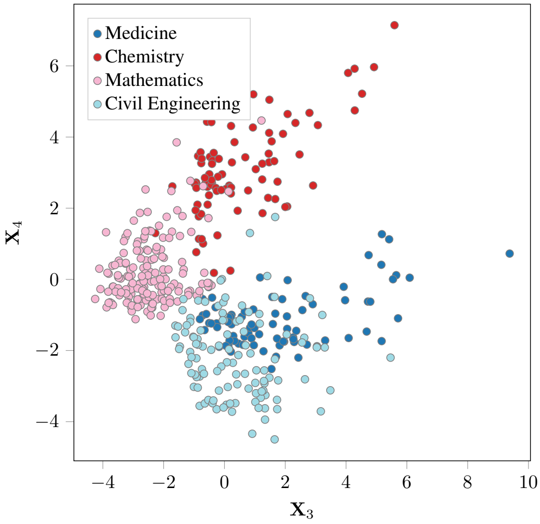


Figure 10. Scatterplot of $\hat{\mathbf{X}}_3$ and $\hat{\mathbf{X}}_4$, coloured by department.

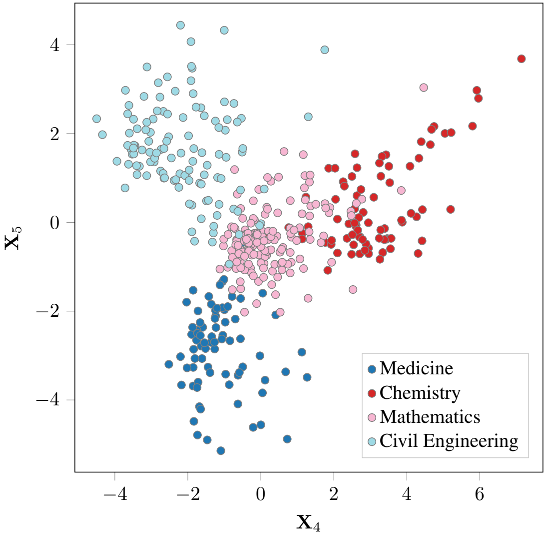


Figure 11. Scatterplot of $\hat{\mathbf{X}}_4$ and $\hat{\mathbf{X}}_5$, coloured by department.

ICL NetFlow: number of clusters

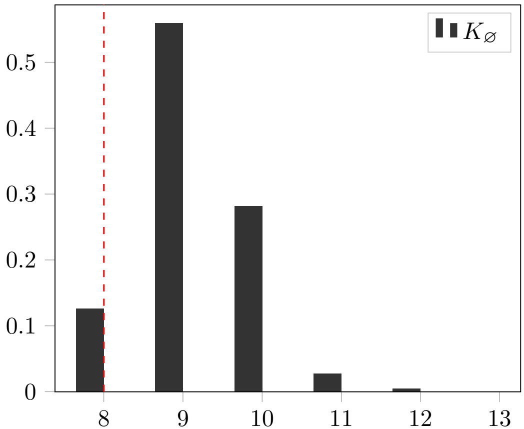


Figure 12. Posterior histogram of $K_\varnothing$, **constrained** model, MAP for d in **red**.

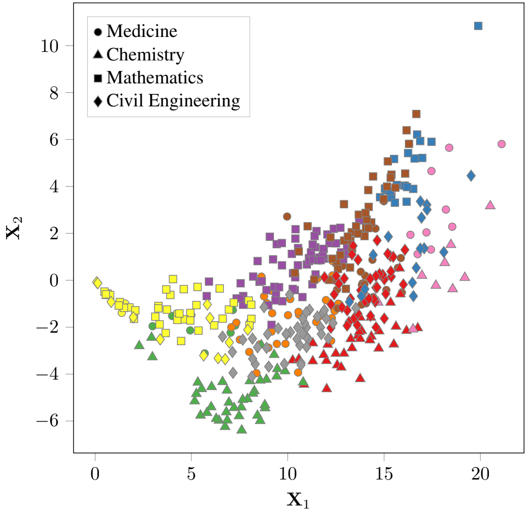


Figure 13. Scatterplot of $\hat{\mathbf{X}}_1$ and $\hat{\mathbf{X}}_2$, labelled by estimated clustering ($K = 9$) and department.

ICL NetFlow: effect of out-degree

- The ASE is strongly correlated with out-degree ⇒ **DCSBM** might be more appropriate.

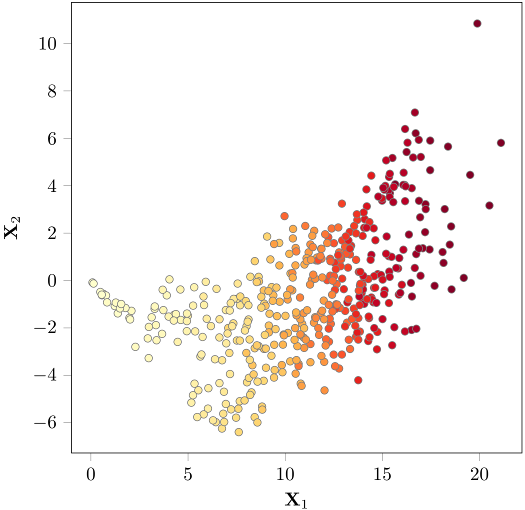


Figure 14. Scatterplot of $\hat{\mathbf{X}}_1$ and $\hat{\mathbf{X}}_2$, coloured by out-degree.

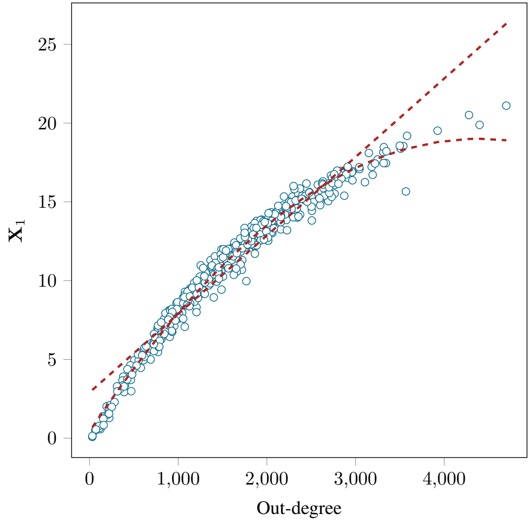


Figure 15. Scatterplot of $\hat{\mathbf{X}}_1$ versus out-degree of the node.

ICL NetFlow: SBM or DCSBM?

- The DCSBM seems to be a better model for the ICL NetFlow data.
- Further evidence: comparison between the observed out-degree distribution and simulated out-degree distributions from SBMs and DCSBMs.

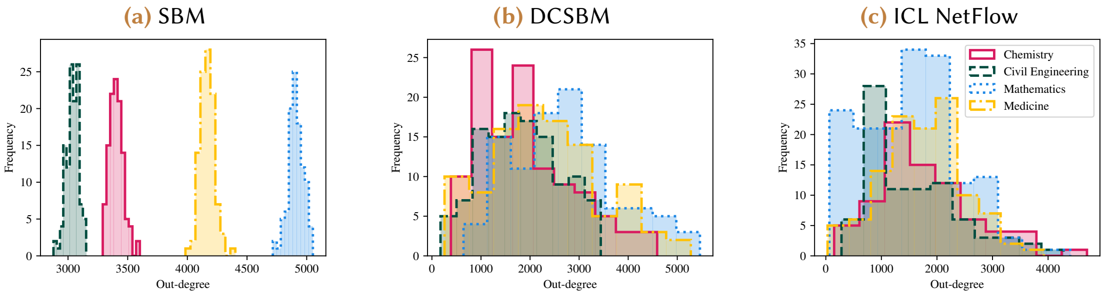


Figure 16. Histogram of within-community degree distributions from three bipartite networks with size 439×60635, obtained from (a) a simulation of a SBM, (b) a simulation of a DCSBM, and (c) the ICL NetFlow network.

A synthetic example

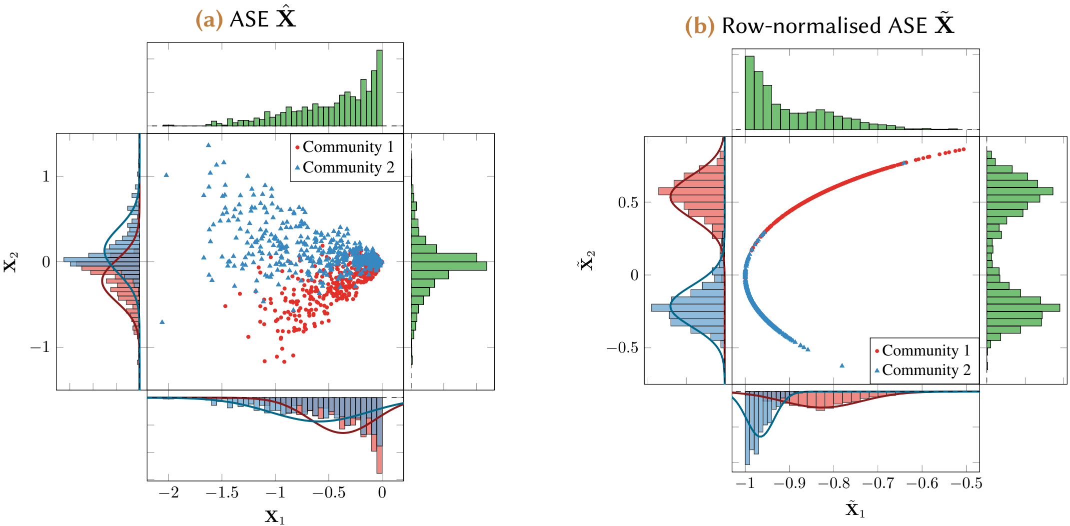


Figure 17. Scatterplot of the 2-dimensional **ASE** and row-normalised ASE for a simulated DCSBM with $d = K = 2$, $B_{11} = 0.1, B_{12} = B_{21} = 0.05$ and $B_{22} = 0.15$, and 500 nodes per community, corrected with $\rho_i \sim \text{Beta}(2, 1)$.

A model for DCSBM embeddings

- Proposed solution: parametric model on the **spherical coordinates** of the embedding.
- Consider a m-dimensional vector $\boldsymbol{x} \in \mathbb{R}^m$. The m Cartesian coordinates $\boldsymbol{x} = (x_1, \dots, x_m)$ can be converted in $m-1$ spherical coordinates $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{m-1})$ on the unit m-sphere using a mapping $f_m : \mathbb{R}^m \to [0, 2\pi)^{m-1}$ such that $f_m : \boldsymbol{x} \mapsto \boldsymbol{\theta}$, where:

$$\theta_1 = \begin{cases} \arccos(x_2/\|\boldsymbol{x}_{:2}\|) & x_1 \geq 0, \\ 2\pi - \arccos(x_2/\|\boldsymbol{x}_{:2}\|) & x_1 < 0, \end{cases}$$

$$\theta_j = 2\arccos(x_{j+1}/\|\boldsymbol{x}_{:j+1}\|),\ j = 2, \dots, m-1.$$

- From the $(m+1)$-dimensional adjacency embedding $\hat{\mathbf{X}} \in \mathbb{R}^{n\times(m+1)}$, define its transformation $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n]^\top \in [0, 2\pi)^{n\times m}$, such that $\boldsymbol{\theta}_i = f_{m+1}(\hat{\boldsymbol{x}}_i),\ i = 1, \dots, n$.

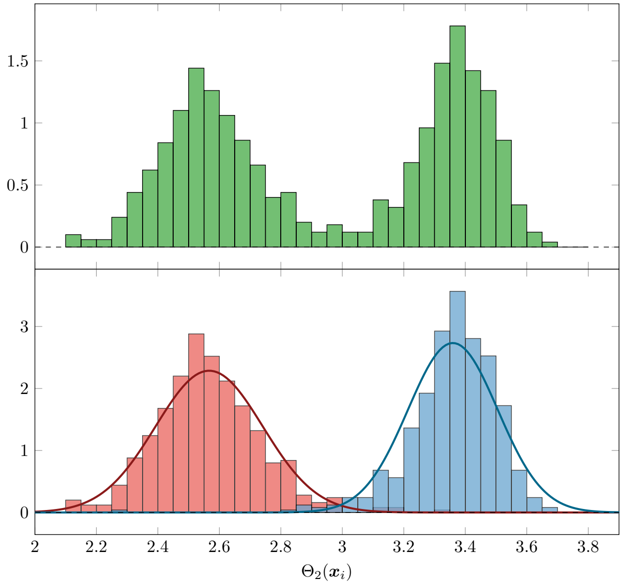


Figure 18. Scatterplot of the **transformed ASE** $\boldsymbol{\Theta}$ for the simulated DCSBM in Figure 17.

“Gaussianisation” of the ASE

A model on spherical coordinates for DCSBM spectral embeddings

- Let $\mathbf{\Theta}_{:d}$ and $\boldsymbol{\theta}_{i,:d}$ denote respectively the first d columns of the matrix and d elements of the vector, and $\mathbf{\Theta}_{d:}$ and $\boldsymbol{\theta}_{i,d:}$ the remaining $m-d$ components.
- For a given pair (d, K), the transformed ASE $\mathbf{\Theta}$ is assumed to have the distribution:

$$\boldsymbol{\theta}_i | d, z_i, \boldsymbol{\vartheta}_{z_i}, \mathbf{\Sigma}_{z_i}, \boldsymbol{\sigma}^2_{z_i} \sim \mathbb{N}_m \left(\begin{bmatrix} \boldsymbol{\vartheta}_{z_i} \\ \pi \mathbf{1}_{m-d} \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_{z_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma}^2_{z_i} \mathbf{I}_{m-d} \end{bmatrix} \right),$$

  where $\boldsymbol{\vartheta}_{z_i} \in [0, 2\pi)^d$ represents a community-specific mean angle, $\mathbf{1}_m$ is a m-dimensional vector of ones, $\mathbf{\Sigma}_{z_i}$ is a $d \times d$ full covariance matrix, and $\boldsymbol{\sigma}^2_k = (\sigma^2_{k,d+1}, \dots, \sigma^2_{k,m})$ is a vector of positive variances.
- The model specification is again completed using a hierarchical prior structure.
- The pair (d, K) could also be chosen using BIC, for m **fixed** (Yang et al., 2021).
- The conjecture for the likelihood mirrors the SBM model for Cartesian coordinates.

Empirical model validation

- $N = 1000$ simulations of a GRDPG-DCSBM with $n = 1500$, $d = K = 3$;
- $\mathbf{B} \sim \text{Uniform}(0,1)^{K \times K}$ fixed across all N simulations, communities of equal size;
- $\rho_i \sim \text{Beta}(2,1)$.

(a) Boxplots of $\boldsymbol{\vartheta}_{k,:2}$ (b) Boxplots of $\boldsymbol{\vartheta}_{k,2:}$

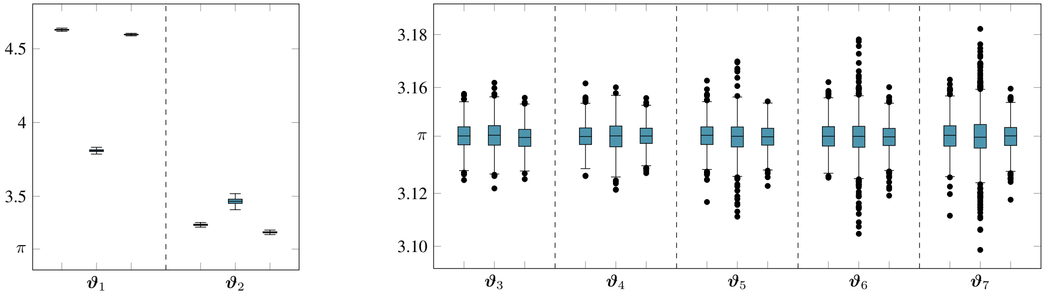


Figure 19. Boxplots for $N = 1{,}000$ simulations of a DCSBM with $n = 1{,}500$ nodes, $K = 3$, equal number of nodes allocated to each group, and $\mathbf{B} \sim \text{Uniform}(0,1)^{K \times K}$, corrected by $\rho_i \sim \text{Beta}(2,1)$.

Empirical model validation

- $N = 1000$ simulations of a GRDPG-DCSBM with $n = 1500$, $d = K = 3$;
- $\mathbf{B} \sim \text{Uniform}(0,1)^{K \times K}$ fixed across all N simulations, communities of equal size;
- $\rho_i \sim \text{Beta}(2,1)$.

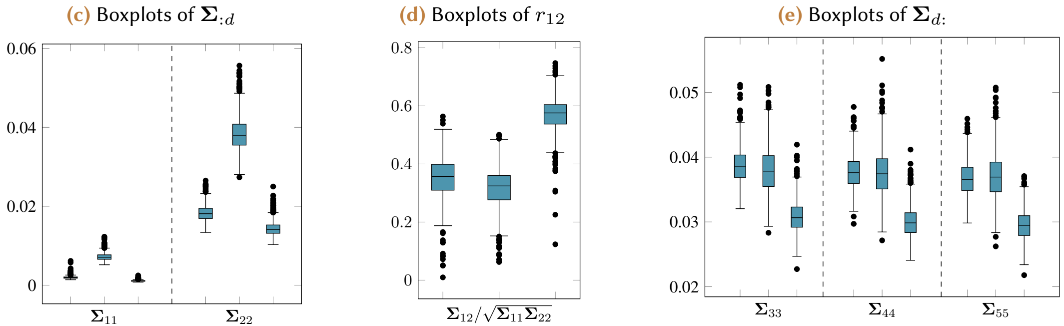
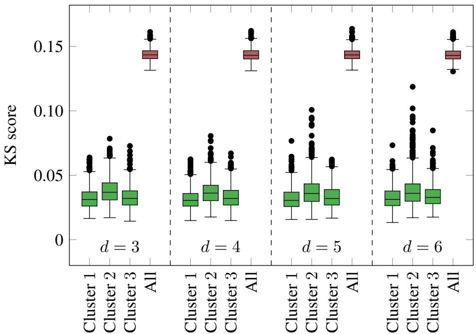


Figure 6. Boxplots for $N = 1{,}000$ simulations of a DCSBM with $n = 1{,}500$ nodes, $K = 3$, equal number of nodes allocated to each group, and $\mathbf{B} \sim \text{Uniform}(0,1)^{K \times K}$, corrected by $\rho_i \sim \text{Beta}(2,1)$.

Empirical model validation

- $N = 1000$ simulations of a GRDPG-DCSBM with $n = 1500$, $d = K = 3$;
- $\mathbf{B} \sim \text{Uniform}(0,1)^{K\times K}$ fixed across all N simulations, communities of equal size;
- $\rho_i \sim \text{Beta}(2,1)$.

(f) Kolmogorov-Smirnov scores for Gaussian fit in $\boldsymbol{\Theta}_d$:



(g) Boxplots of $r_{k\ell}$ for the redundant components

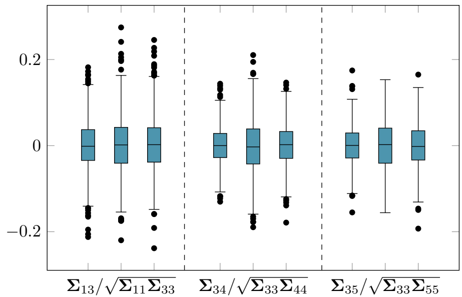


Figure 6. Boxplots for $N = 1{,}000$ simulations of a DCSBM with $n = 1{,}500$ nodes, $K = 3$, equal number of nodes allocated to each group, and $\mathbf{B} \sim \text{Uniform}(0,1)^{K\times K}$, corrected by $\rho_i \sim \text{Beta}(2,1)$.

ICL NetFlow: row-normalised and transformed embeddings

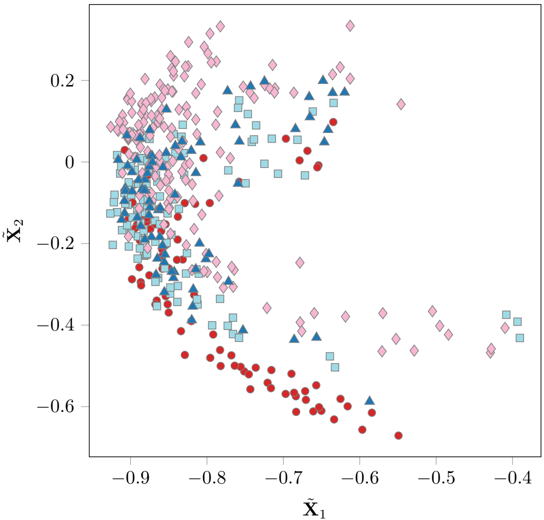


Figure 7. Scatterplot of $\tilde{\mathbf{X}}_{:2}$ for $m = 30$.

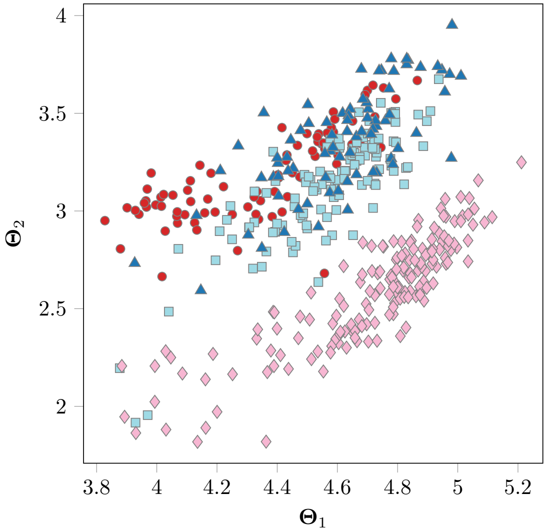


Figure 8. Scatterplot of $\mathbf{\Theta}_{:2}$ for $m = 30$.

ICL NetFlow: parameter estimates and community detection

	$m=30$			$m=50$		
	$\hat{\mathbf{X}}$	$\tilde{\mathbf{X}}$	$\boldsymbol{\Theta}$	$\hat{\mathbf{X}}$	$\tilde{\mathbf{X}}$	$\boldsymbol{\Theta}$
Estimated (d, K)	$(28, 5)$	$(8, 7)$	$(15, 4)$	$(29, 4)$	$(8, 7)$	$(15, 4)$
Adjusted Rand Index (ARI)	0.441	0.736	0.938	0.359	0.743	0.938

Table 1. Estimates of (d, K) and ARIs for the embeddings $\hat{\mathbf{X}}$, $\tilde{\mathbf{X}}$ and $\boldsymbol{\Theta}$ for $m \in \{30, 50\}$.

- Estimates from $\hat{\mathbf{X}}$ and $\tilde{\mathbf{X}}$ are obtained using the model for the SBM (Sanna Passino and Heard, 2020; Yang et al., 2021).
- Estimates from $\boldsymbol{\Theta}$ are obtained using the model for the DCSBM (Sanna Passino, Heard, and Rubin-Delanchy, 2020).
- Using $\boldsymbol{\Theta}$, the correct value of K is estimated (corresponding to the number of departments).
- Using $\boldsymbol{\Theta}$, only 9 **nodes** are misclassified.
- The constraint of unit row-norm on $\tilde{\mathbf{X}}$ causes issues in the estimation of K.
- Estimates appear to be stable for different values of m.

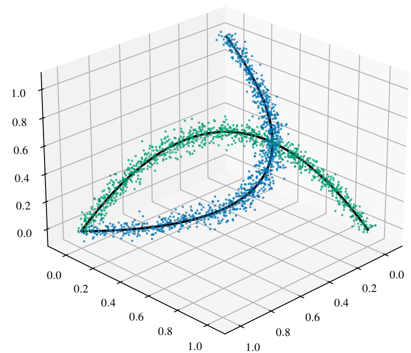
Beyond SBMs and DCSBMs: latent structure blockmodels (LSBMs)

- The SBM and DCSBM correspond to very simple community-specific latent structure under the RDPG.
  - SBM: each cluster corresponds to a latent *point*.
  - DCSBM: each cluster corresponds to a latent *ray*.
- More generally: each community might be associated with a different **one-dimensional structural support submanifold** $\mathcal{S}_k,\ k=1,\dots,K$.
- Parametrically, latent positions can be expressed as:

$$\boldsymbol{x}_i = \boldsymbol{f}(\phi_i, z_i).$$

- The function $\boldsymbol{f}=(f_1,\dots,f_d):\mathbb{R}\times\{1,\dots,K\}\to\mathbb{R}^d$ maps the latent draw ϕ_i to the corresponding node latent position on the community-specific submanifold corresponding to the community allocation z_i.
- Proposal: **latent structure blockmodel (LSBM)**.

Hardy-Weinberg LSBM, $K=2$



$$\boldsymbol{f}(\phi_i,1) = (\phi_i^2, 2\phi_i(1-\phi_i), (1-\phi_i)^2),$$
$$\boldsymbol{f}(\phi_i,2) = (2\phi_i(1-\phi_i), (1-\phi_i)^2, \phi_i^2).$$

LSBMs: some examples

- SBMs and DCSBMs are **special cases of LSBMs**. ICL NetFlow: **quadratic LSBM**?
- From the ASE-CLT:

$$\mathbf{Q}\hat{\boldsymbol{x}}_i \approx \mathbb{N}_d\{\boldsymbol{f}(\phi_i, z_i), \boldsymbol{\Sigma}(\phi_i, z_i)\},$$

  for some orthogonal matrix $\mathbf{Q}$ and covariance matrix function $\boldsymbol{\Sigma} : \mathbb{R}\times\{1,\dots,K\} \to \mathbb{R}^{d\times d}$.
- More examples and details: Sanna Passino and Heard, 2021 (forthcoming on *arXiv*).

(a) SBM $\boldsymbol{f}(\phi_i, z_i) = \boldsymbol{\mu}_{z_i}$ | **(b)** DCSBM $\boldsymbol{f}(\phi_i, z_i) = \phi_i\boldsymbol{\mu}_{z_i}$ | **(c)** Quadratic LSBM $\boldsymbol{f}(\phi_i, z_i) = \boldsymbol{\alpha}_{z_i}\phi_i^2 + \boldsymbol{\beta}_{z_i}\phi_i$

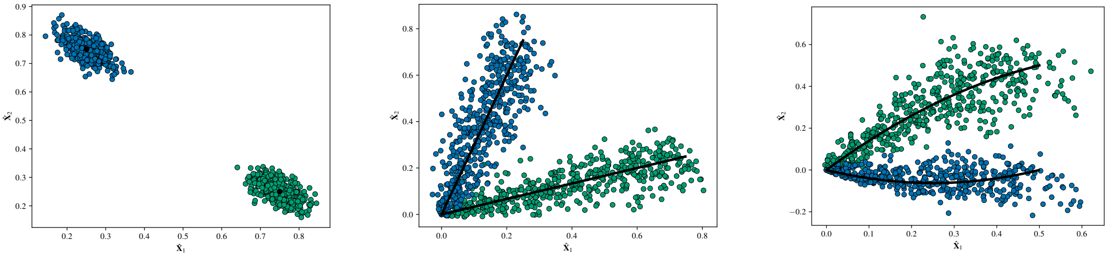


Figure 9. Scatterplots of the 2-dimensional ASE of simulated graphs with $n = 1000$ and $K = 2$, arising from different LSBMs, and true underlying latent curves (in black).

Bayesian modelling of LSBMs

- Inferential task: recover $\boldsymbol{z} = (z_1, \dots, z_n)$ given a realisation of the adjacency matrix $\mathbf{A}$.
- Problem: $\boldsymbol{f}(\cdot)$ is **unknown** $\rightarrow$ a prior on functions is needed.
- Most commonly used prior on unknown functions: **Gaussian process**.
  - $f \sim \mathrm{GP}(\nu, \xi)$, if for any $\boldsymbol{x} = (x_1, \dots, x_n)$, $f(\boldsymbol{x}) \sim \mathbb{N}_n\{\nu(\boldsymbol{x}), \boldsymbol{\Xi}(\boldsymbol{x}, \boldsymbol{x})\}$, where $\boldsymbol{\Xi}(\boldsymbol{x}, \boldsymbol{x})$ is a $n \times n$ matrix such that $[\boldsymbol{\Xi}(\boldsymbol{x}, \boldsymbol{x})]_{k\ell} = \xi(x_k, x_\ell)$ for a positive semi-definite kernel function ξ.
- Hierarchical Bayesian model:

$$
\begin{aligned}
\hat{\boldsymbol{x}}_i | z_i, \phi_i, \boldsymbol{f}, \boldsymbol{\sigma}^2_{z_i} &\sim \prod_{j=1}^{d} \mathbb{N}\left\{\hat{x}_{i,j} \mid f_j(\phi_i, z_i), \sigma^2_{z_i,j}\right\},\ i = 1, \dots, n, \\
f_j(\cdot, k) | \sigma^2_{k,j} &\sim \mathrm{GP}(0, \xi_{k,j}),\ k = 1, \dots, K,\ j = 1, \dots, d, \\
\sigma^2_{k,j} &\sim \text{Inv-Gamma}(a_0, b_0),\ k = 1, \dots, K,\ j = 1, \dots, d.
\end{aligned}
$$

- Simplification: $\boldsymbol{\Sigma}(\phi_i, z_i) = \boldsymbol{\sigma}^2_{z_i} \mathbf{I}_{d \times d} \rightarrow$ approximately “functional” k-means.
- The model specification is completed by the following priors:

$$
\begin{aligned}
& z_i \sim \text{Discrete}(\boldsymbol{\psi}),\ \boldsymbol{\psi} = (\psi_1, \dots, \psi_K),\ i = 1, \dots, n, \\
& \boldsymbol{\psi} \sim \text{Dirichlet}(\alpha/K, \dots, \alpha/K), \\
& \phi_i \sim \mathbb{N}(\mu_\phi, \sigma^2_\phi),\ i = 1, \dots, n.
\end{aligned}
$$

A special case: inner product kernels

- **Inner product kernels** $\Rightarrow$ **linear models** (linear & polynomial regression, splines...).
- Essentially a **Bayesian linear regression** model with suitably chosen **basis functions** with **conjugate normal-inverse-gamma priors** on the parameters.
- Closed-form marginals are available $\rightarrow$ MCMC inference reduces to (ϕ_i, z_i).
- According to the model choice, **identifiability issues** might arise. For example, for the DCSBM:

$$\phi_i \boldsymbol{\mu}_{z_i} = (\phi_i/\kappa)(\kappa\boldsymbol{\mu}_{z_i}), \kappa \in \mathbb{R}.$$

- On the ICL NetFlow data, it might be suitable to use a **quadratic LSBM** $\rightarrow$ the curves $\mathcal{S}_1, \dots, \mathcal{S}_4$ are parabolas passing through the origin.

ICL NetFlow: quadratic LSBM

- Consider an inner product kernel such that:

$$\boldsymbol{f}(\phi_i, z_i) = \boldsymbol{\alpha}_{z_i}\phi_i^2 + \boldsymbol{\beta}_{z_i}\phi_i,\ \boldsymbol{\alpha}_{z_i}, \boldsymbol{\beta}_{z_i} \in \mathbb{R}^d.$$

- Adjusted Rand Index $> 0.94 \to 8$ misclassified nodes, slightly better than DCSBM.

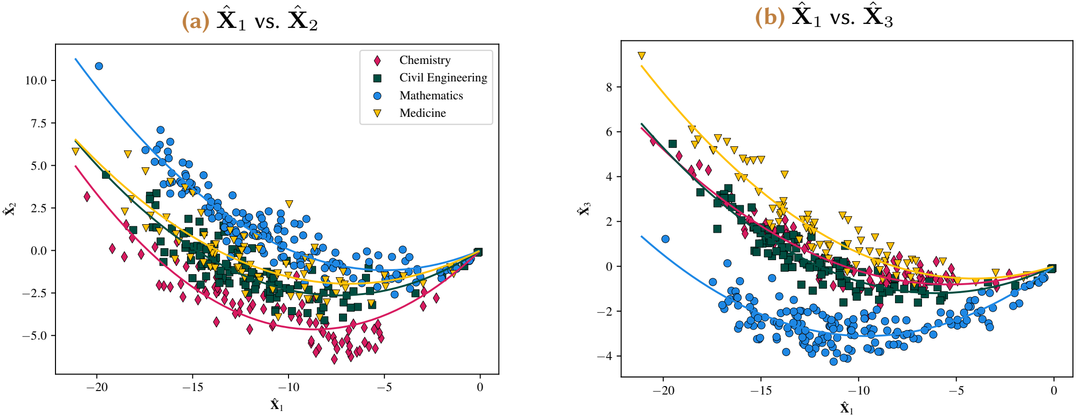


Figure 10. Scatterplots of $\{\hat{\mathbf{X}}_2, \hat{\mathbf{X}}_3, \hat{\mathbf{X}}_4, \hat{\mathbf{X}}_5\}$ vs. $\hat{\mathbf{X}}_1$, coloured by department, and estimated best fitting quadratic curves after clustering.

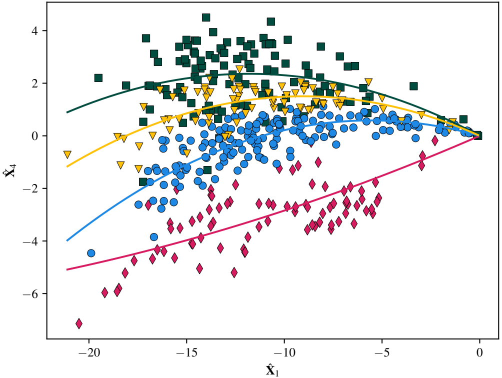
ICL NetFlow: quadratic LSBM

- Consider an inner product kernel such that:

$$\boldsymbol{f}(\phi_i, z_i) = \boldsymbol{\alpha}_{z_i}\phi_i^2 + \boldsymbol{\beta}_{z_i}\phi_i,\ \boldsymbol{\alpha}_{z_i}, \boldsymbol{\beta}_{z_i} \in \mathbb{R}^d.$$

- Adjusted Rand Index $> 0.94 \rightarrow 8$ misclassified nodes, slightly better than DCSBM.

(c) $\hat{\mathbf{X}}_1$ vs. $\hat{\mathbf{X}}_4$



(d) $\hat{\mathbf{X}}_1$ vs. $\hat{\mathbf{X}}_5$

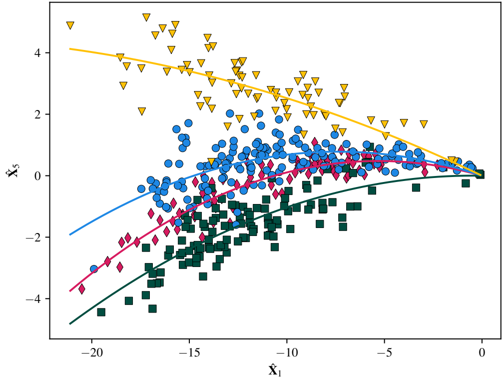


Figure 11. Scatterplots of $\{\hat{\mathbf{X}}_2, \hat{\mathbf{X}}_3, \hat{\mathbf{X}}_4, \hat{\mathbf{X}}_5\}$ vs. $\hat{\mathbf{X}}_1$, coloured by department, and estimated best fitting quadratic curves after clustering.

ICL NetFlow: LSBMs with splines

- Consider a cubic truncated power basis with three equally spaced knots $\kappa_\ell,\ \ell = 1, 2, 3$:

$$\tilde{f}_{j,1}(\phi) = \phi,\ \tilde{f}_{j,2}(\phi) = \phi^2,\ \tilde{f}_{j,3}(\phi) = \phi^3,\ \tilde{f}_{j,3+\ell}(\phi) = (\phi - \kappa_\ell)^3_+,\ \ell = 1, 2, 3,$$

where $(\cdot)_+ = \max\{0, \cdot\}$. This gives:

$$f_j(\phi_i, z_i) = \sum_{h=1}^{6} \beta_{j,h,z_i} \tilde{f}_{j,h}(\phi_i).$$

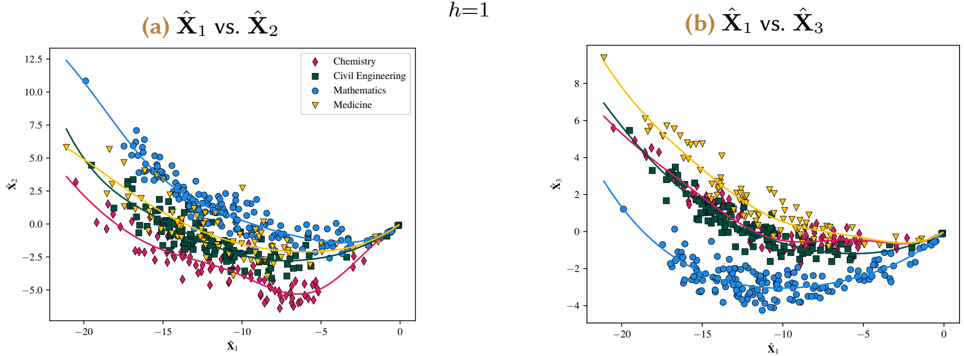
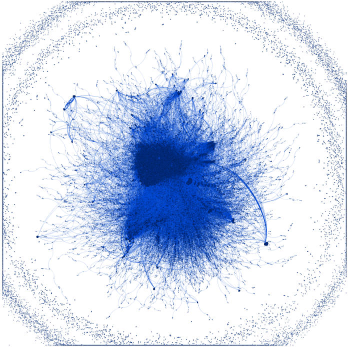


Figure 12. Scatterplots of $\{\hat{\mathbf{X}}_2, \hat{\mathbf{X}}_3, \hat{\mathbf{X}}_4, \hat{\mathbf{X}}_5\}$ vs. $\hat{\mathbf{X}}_1$, coloured by department, and estimated best curves after clustering.

Conclusion / Summary of contributions

- **Model selection** under the SBM and DCSBM:
  - Simultaneous selection of d and K under the GRDPG,
  - Allow for initial misspecification of the arbitrarily large parameter m, then refine estimate d,
  - SBM: Gaussian mixture model (with constraints),
  - DCSBM: Gaussian mixture model on spherical coordinates (with constraints),
  - Easy to extend to directed and bipartite graphs.
- **Latent substructure inference** in GRDPG:
  - **Latent structure blockmodels** admitting community-specific structural support submanifolds,
  - Flexible **Gaussian process priors** for Bayesian inference on unknown latent functions,
  - The SBM and DCSBM are special cases of the LSBM.
- **What's next**: simultaneous model selection of d and K in LSBMs, automatic selection of the complexity of the latent functions.



References I

- Athreya, A. et al. (2016). "A Limit Theorem for Scaled Eigenvectors of Random Dot Product Graphs". In: *Sankhya A* 78.1, pp. 1–18.
- 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.
- Fortunato, S. (2010). "Community detection in graphs". In: *Physics Reports* 486.3, pp. 75–174.
- 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).
- Ng, A. Y., M. I. Jordan, and Y. Weiss (2001). "On Spectral Clustering: Analysis and an Algorithm". In: *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic.* Vancouver, British Columbia, Canada, pp. 849–856.
- 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.

References II

- Sanna Passino, F. and N. A. Heard (2020). "Bayesian estimation of the latent dimension and communities in stochastic blockmodels". In: *Statistics and Computing* 30.5, pp. 1291–1307.
- Sanna Passino, F. and N. A. Heard (2021). "Latent structure blockmodels for Bayesian spectral graph clustering". In: *arXiv e-prints (forthcoming).*
- Sanna Passino, F., N. A. Heard, and P. Rubin-Delanchy (2020). "Spectral clustering on spherical coordinates under the degree-corrected stochastic blockmodel". In: *arXiv e-prints.* arXiv: 2011.04558 [stat.ML].
- von Luxburg, U. (2007). "A tutorial on spectral clustering". In: *Statistics and Computing* 1.4, pp. 395–416.
- Yang, C. et al. (2021). "Simultaneous dimensionality and complexity model selection for spectral graph clustering". In: *Journal of Computational and Graphical Statistics* (to appear).
- 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.