JSM19 – Novel Approaches for Analyzing Dynamic Networks

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

July 30, 2019

DENVER, COLORADO JSM2019 JULY 27–AUGUST 1

Stochastic blockmodels as random dot product graphs

- Consider an **undirected graph** with **symmetric adjacency matrix** $\mathbf{A} \in \{0,1\}^{n \times n}$.
- In **random dot product graphs**, the probability of a link between two nodes is expressed as the inner product between two latent positions $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{F}$, $0 \leq \boldsymbol{x}^\top \boldsymbol{y} \leq 1 \,\forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{F}$:

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

- The **stochastic blockmodel** is the classical model for community detection in graphs. Given a matrix $\mathbf{B} \in [0,1]^{K \times K}$ of within-community probabilities, the probability of a link depends on the **community allocations** z_i and $z_j \in \{1, \dots, K\}$ of the two nodes:

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

- The stochastic blockmodel can be interpreted as a **special case** of a random dot product graph. If $B_{kh} = \boldsymbol{\mu}_k^\top \boldsymbol{\mu}_h$ with $\boldsymbol{\mu}_k, \boldsymbol{\mu}_h \in \mathcal{F}$, and all the nodes in community k are assigned the latent position $\boldsymbol{\mu}_k$, then:

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

Network embeddings

- Consider an undirected graph with symmetric adjacency matrix $\mathbf{A} \in \{0,1\}^{n \times n}$, and modified Laplacian $\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, $\mathbf{D} = \mathrm{diag}(\sum_{i=1}^n A_{ij})$.
- The **adjacency embedding** of $\mathbf{A}$ in $\mathbb{R}^d$ is:

$$\hat{\mathbf{X}} = [\hat{\boldsymbol{x}}_1, \dots, \hat{\boldsymbol{x}}_n]^\top = \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Lambda}}^{1/2} \in \mathbb{R}^{n \times d},$$

  where $\hat{\boldsymbol{\Lambda}}$ is a $d \times d$ diagonal matrix containing the top d largest eigenvalues of $\mathbf{A}$, and $\hat{\boldsymbol{\Gamma}}$ is a $n \times d$ matrix containing the corresponding orthonormal eigenvectors.
- The **Laplacian embedding** of $\mathbf{A}$ in $\mathbb{R}^d$ is:

$$\tilde{\mathbf{X}} = [\tilde{\boldsymbol{x}}_1, \dots, \tilde{\boldsymbol{x}}_n]^\top = \tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{\Lambda}}^{1/2} \in \mathbb{R}^{n \times d},$$

  where $\tilde{\boldsymbol{\Lambda}}$ is a $d \times d$ diagonal matrix containing the top d largest eigenvalues of $\mathbf{L}$, and $\tilde{\boldsymbol{\Gamma}}$ is a $n \times d$ matrix containing the corresponding orthonormal eigenvectors.

Spectral estimation of the stochastic blockmodel

- Based on asymptotic properties, Rubin-Delanchy et al., 2017, propose the following algorithm for consistent estimation of the latent positions in stochastic blockmodels:

Algorithm 1: 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 $z_1, \dots, z_n$.

- In practice: d and K are estimated **sequentially**. Issues:
  - 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**.
- **This talk discusses a novel framework for joint estimation of d and K.**

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 $d = \mathrm{rank}(\mathbf{B}) \leq K$.

$$\begin{aligned}
\boldsymbol{x}_i | d, z_i, \boldsymbol{\mu}_{z_i}, \boldsymbol{\Sigma}_{z_i}, \boldsymbol{\sigma}^2_{z_i} &\overset{d}{\sim} \mathbb{N}_m\left(\begin{bmatrix}\boldsymbol{\mu}_{z_i} \\ \mathbf{0}_{m-d}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{z_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma}^2_{z_i}\mathbf{I}_{m-d}\end{bmatrix}\right),\ i = 1,\dots,n, \\
(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)|d &\overset{iid}{\sim} \mathrm{NIW}_d(\mathbf{0}, \kappa_0, \nu_0 + d - 1, \boldsymbol{\Delta}_d),\ k = 1,\dots,K, \\
\sigma^2_{kj} &\overset{iid}{\sim} \mathrm{Inv}\text{-}\chi^2(\lambda_0, \sigma^2_0),\ j = d+1,\dots,m, \\
d|\boldsymbol{z} &\overset{d}{\sim} \mathrm{Uniform}\{1,\dots,K_\varnothing\}, \\
z_i|\boldsymbol{\theta} &\overset{iid}{\sim} \mathrm{Multinoulli}(\boldsymbol{\theta}),\ i = 1,\dots,n,\ \boldsymbol{\theta} \in \mathcal{S}_{K-1}, \\
\boldsymbol{\theta}|K &\overset{d}{\sim} \mathrm{Dirichlet}\left(\frac{\alpha}{K},\dots,\frac{\alpha}{K}\right), \\
K &\overset{d}{\sim} \mathrm{Geometric}(\omega).
\end{aligned}$$

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

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

DENVER, COLORADO JSM2019 JULY 27–AUGUST 1

Empirical model validation

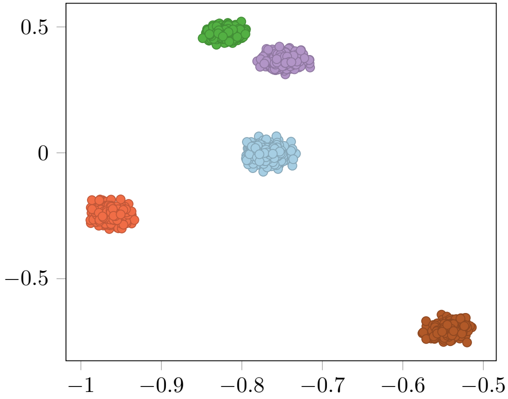


Figure 1. Scatterplot of the simulated $\mathbf{X}_1$ and $\mathbf{X}_2$ – i.e. $\mathbf{X}_{:d}$

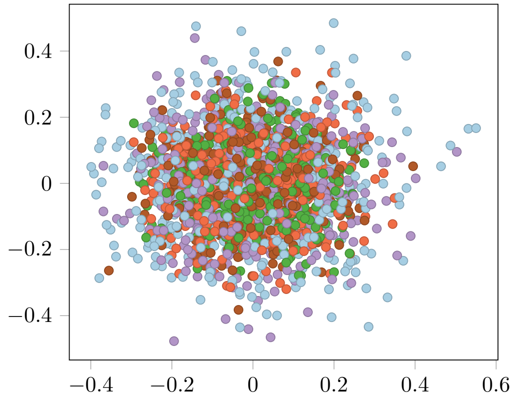


Figure 2. Scatterplot of the simulated $\mathbf{X}_3$ and $\mathbf{X}_4$

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

Empirical model validation

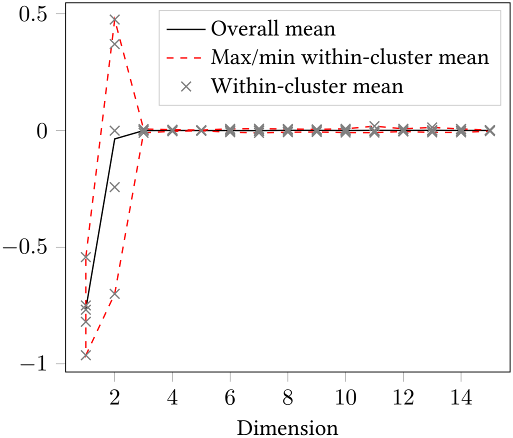


Figure 3. Within-cluster and overall means of $\mathbf{X}_{:15}$

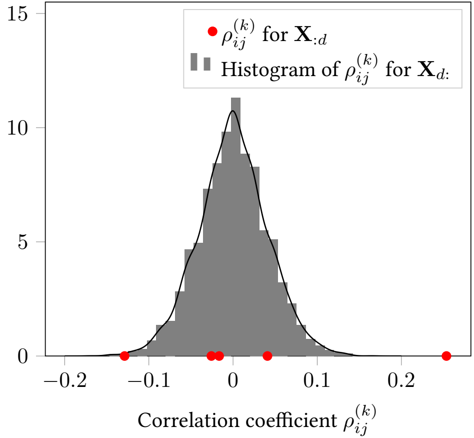


Figure 4. Within-cluster correlation coefficients of $\mathbf{X}_{:30}$

- Means are approximately $\mathbf{0}$ for columns with index $> d$.
- Reasonable to assume correlation $\rho_{ij}^{(k)} = 0$ for $i, j > d$.

Curse of dimensionality

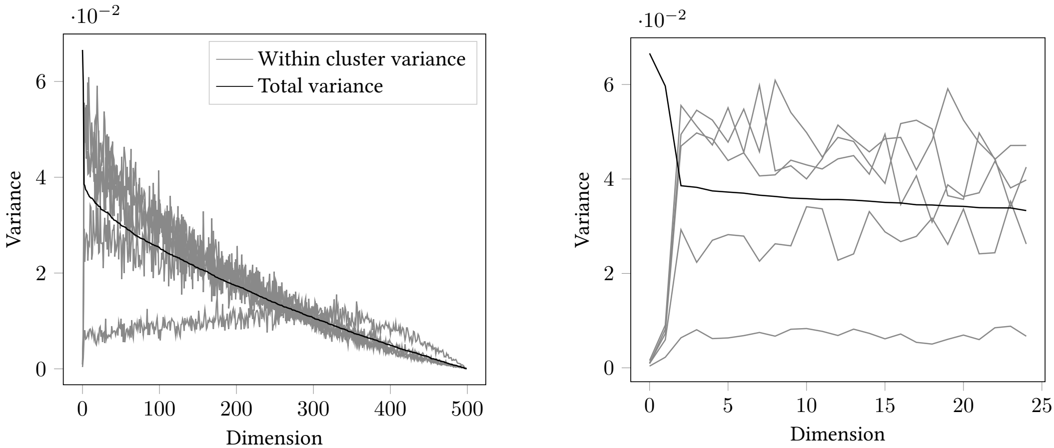


Figure 5. Within-block variance and total variance for the adjacency embedding obtained from a simulated SBM with $d = 2$, $K = 5$, $n = 500$, and well separated means $\boldsymbol{\mu}_1 = [0.7, 0.4]$, $\boldsymbol{\mu}_2 = [0.1, 0.1]$, $\boldsymbol{\mu}_3 = [0.4, 0.8]$, $\boldsymbol{\mu}_4 = [-0.1,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

DENVER, COLORADO JSM2019 JULY 27–AUGUST 1

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 $j > d$:

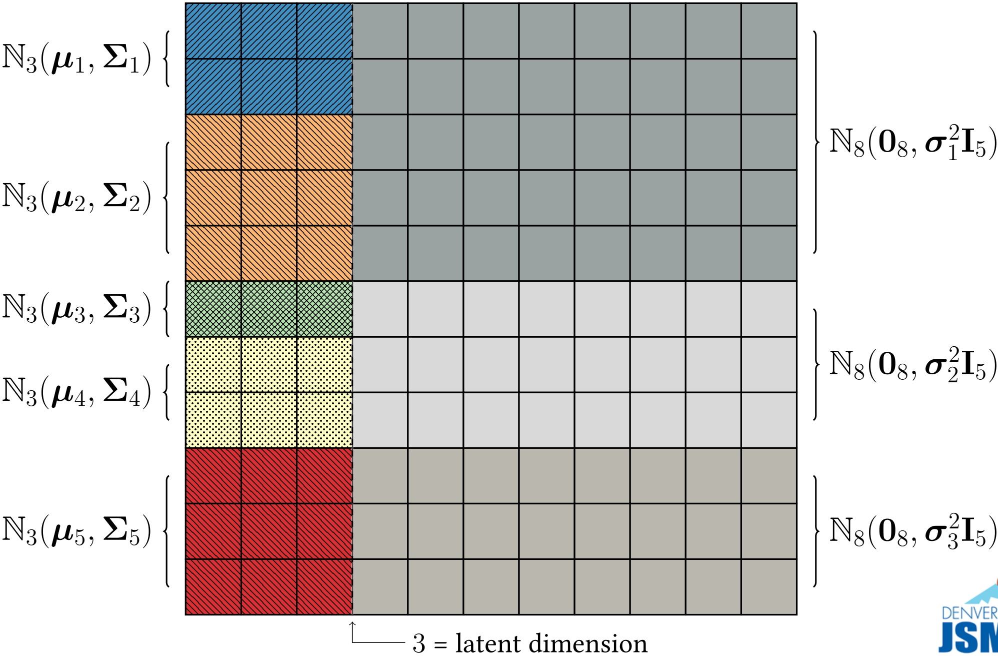
$$\boldsymbol{x}_i | d, z_i, v_{z_i}, \boldsymbol{\mu}_{z_i}, \boldsymbol{\Sigma}_{z_i}, \boldsymbol{\sigma}^2_{v_{z_i}} \stackrel{d}{\sim} \mathbb{N}_m \left(\begin{bmatrix} \boldsymbol{\mu}_{z_i} \\ \mathbf{0}_{m-d} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{z_i} & \mathbf{0} \\ \mathbf{0}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

$$v_k | K, H \stackrel{d}{\sim} \mathrm{Multinoulli}(\boldsymbol{\phi}), \ k = 1, \dots, K,$$

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

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

- The parameter v_k defines **clusters of clusters**.
- Empirical results show that the model is able to handle $d \ll m$.


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

Extension to directed and bipartite graphs

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

$$\hat{\mathbf{X}} = \hat{\mathbf{U}}\hat{\mathbf{D}}^{1/2} \oplus \hat{\mathbf{V}}\hat{\mathbf{D}}^{1/2} = \left[\hat{\mathbf{U}}\hat{\mathbf{D}}^{1/2} \quad \hat{\mathbf{V}}\hat{\mathbf{D}}^{1/2}\right] = \left[\hat{\mathbf{X}}_s \quad \hat{\mathbf{X}}_r\right].$$

  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 $\hat{\mathbf{U}} \in \mathbb{R}^{n\times d}$, $\hat{\mathbf{D}} \in \mathbb{R}_+^{d\times d}$ diagonal, and $\hat{\mathbf{V}} \in \mathbb{R}^{n\times d}$.
- Essentially, only three distributions change:

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

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

$$\sigma^2_{kj}|d,K \stackrel{iid}{\sim} \mathrm{Inv}\text{-}\chi^2(\lambda_0,\sigma_0^2),\ j=d+1,\dots,m.$$

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

Enron e-mail network: number of clusters

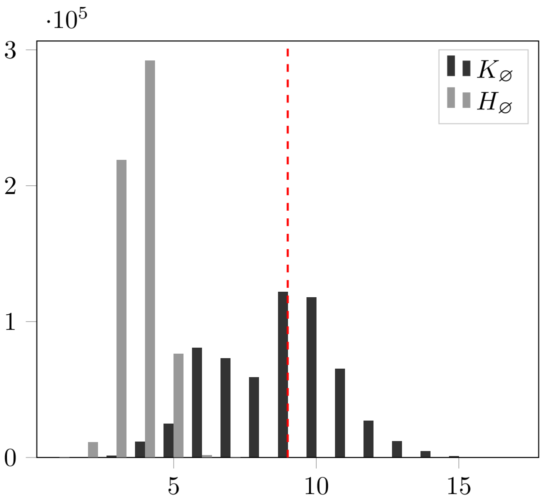


Figure 6. Posterior histogram of $K_\varnothing$ and $H_\varnothing$, unconstrained model, MAP for d in **red**.

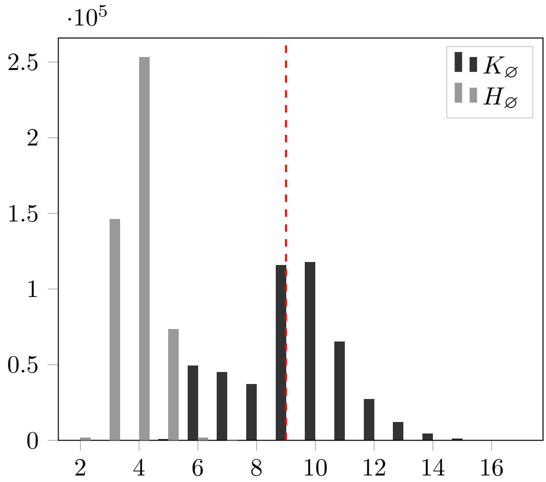


Figure 7. Posterior histogram of $K_\varnothing$ and $H_\varnothing$, constrained model, MAP for d in **red**.

Enron e-mail network: number of clusters

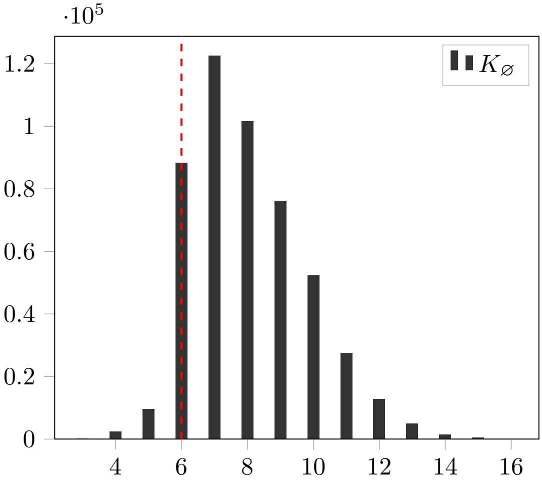


Figure 8. Posterior histogram of $K_\varnothing$, unconstrained model without second order clustering, MAP for d in **red**.

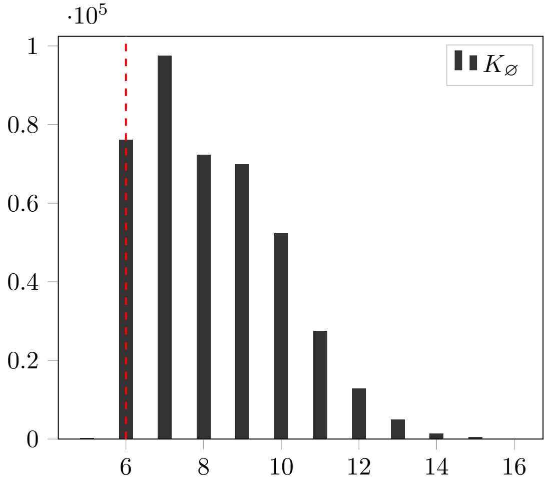


Figure 9. Posterior histogram of $K_\varnothing$, constrained model without second order clustering, MAP for d in **red**.

Enron e-mail network: scree-plot

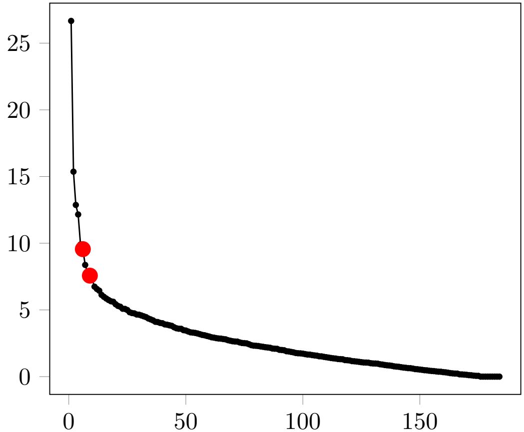


Figure 10. Singular values of the **adjacency** matrix.

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

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.

Scan me

DENVER, COLORADO
JSM2019
JULY 27–AUGUST 1

References

- 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.
- 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.