Information-Theoretic Limits of Network Inference Problems

Galen Reeves

Department of ECE and Department of Statistical Science
Duke University

GraphEx, April 2019

Table of Contents

Spectrum of research on network inference

flexible models
limited theory

toy models
extensive theory

corrected SBM
mixed-membership

SBM

k-community
symm. SBM

2-community
symm. SBM

Spectrum of research on network inference

flexible models
limited theory

toy models
extensive theory

corrected SBM
mixed-membership

SBM

k-community
symm. SBM

2-community
symm. SBM

applied network analysis
scientific modeling

statistical
modeling

computer science
information theory
statistical physics

The 'Bayes optimal' setting

Assume joint distribution on $(\boldsymbol{X}, \boldsymbol{G})$ where

- $\boldsymbol{G}$ is adjacency matrix of simple graph with n vertices
- $\boldsymbol{X} = (X_1, \dots, X_n)$ contains vertex labels

The 'Bayes optimal' setting

Assume joint distribution on $(\boldsymbol{X}, \boldsymbol{G})$ where

- $\boldsymbol{G}$ is adjacency matrix of simple graph with n vertices
- $\boldsymbol{X} = (X_1, \ldots, X_n)$ contains vertex labels

How much do we learn?

$$P(\boldsymbol{X} \mid \boldsymbol{G}) \quad \text{versus} \quad P(\boldsymbol{X})$$

$$I(\boldsymbol{X}; \boldsymbol{G}) = \mathbb{E}\left[\log \frac{P(\boldsymbol{X}, \boldsymbol{G})}{P(\boldsymbol{X})P(\boldsymbol{G})}\right]$$

The 'Bayes optimal' setting

Assume joint distribution on $(\boldsymbol{X}, \boldsymbol{G})$ where

- $\boldsymbol{G}$ is adjacency matrix of simple graph with n vertices
- $\boldsymbol{X} = (X_1, \ldots, X_n)$ contains vertex labels

How much do we learn?

$$P(\boldsymbol{X} \mid \boldsymbol{G}) \quad \text{versus} \quad P(\boldsymbol{X})$$

$$I(\boldsymbol{X}; \boldsymbol{G}) = \mathbb{E}\left[\log \frac{P(\boldsymbol{X}, \boldsymbol{G})}{P(\boldsymbol{X})P(\boldsymbol{G})}\right]$$

How well can we recover labels?

$$\hat{\boldsymbol{X}} \quad \text{verus} \quad \boldsymbol{X}$$

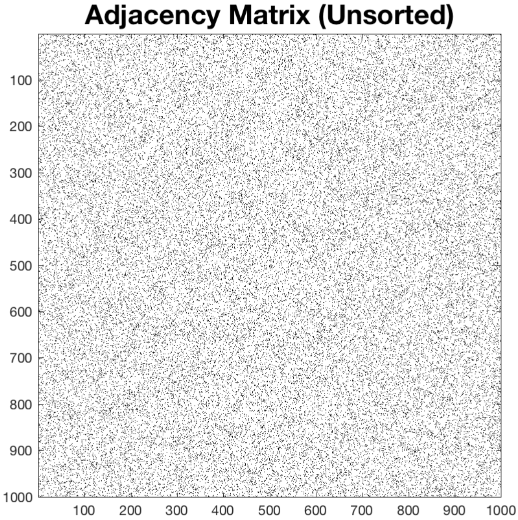
Table of Contents

Recent literature on detection / recovery thresholds

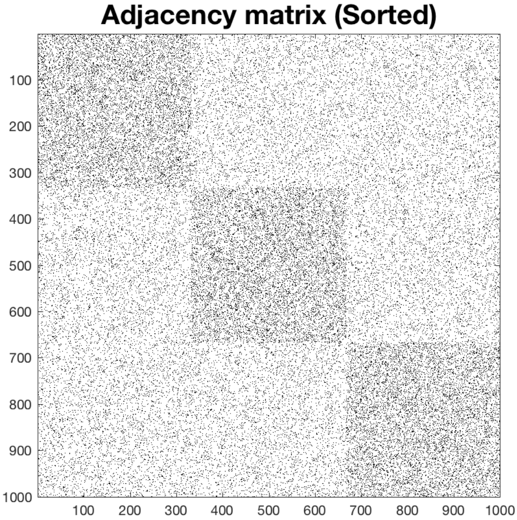
Incomplete list:

- A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, 2011
- E. Mossel, J. Neeman, and A. Sly, 2014
- Deshpande, E. Abbe, and A. Montanari, 2015
- J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová, 2016
- J. Banks, C. Moore, J. Neeman, and P. Netrapalli, 2017
- E. Abbe and C. Sandon, 2017
- F. Caltagirone, M. Lelarge and L. Miolane, 20'7
- F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, 2018

Symmetric SBM



Symmetric SBM



Symmetric SBM

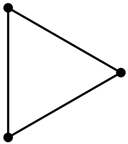
Vertex labels $X_1, \ldots, X_n$ are i.i.d. uniform over k communities.

Symmetric SBM

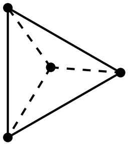
Vertex labels $X_1, \ldots, X_n$ are i.i.d. uniform over k communities.

Without loss of generality, labels can be represented as using k equidistant points in $(k-1)$-dimensional Euclidean space:

$k = 2$



$k = 3$

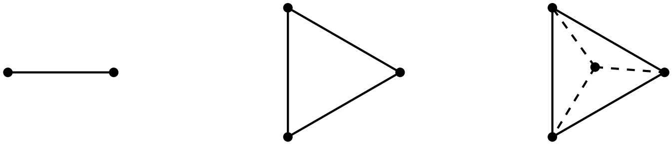


$k = 4$

Symmetric SBM

Vertex labels $X_1, \ldots, X_n$ are i.i.d. uniform over k communities.

Without loss of generality, labels can be represented as using k equidistant points in $(k-1)$-dimensional Euclidean space:



$k = 2$ $\quad$ $k = 3$ $\quad$ $k = 4$

Normalize such that $\mathbb{E}[X_i] = 0$ and $\mathrm{Cov}(X_i) = I$, and hence

$$X_i^\top X_j = \begin{cases} k-1, & X_i = X_j \\ -1, & X_i \neq X_j \end{cases}$$

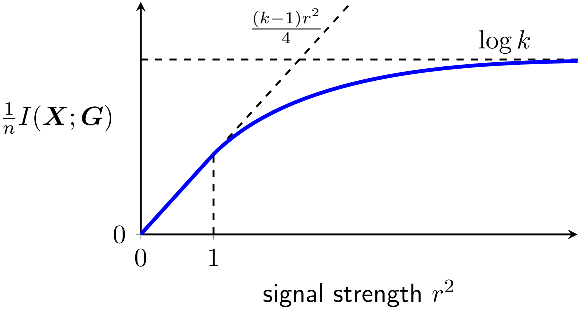
Symmetric SBM

Entries of adjacency matrix are conditionally independent:

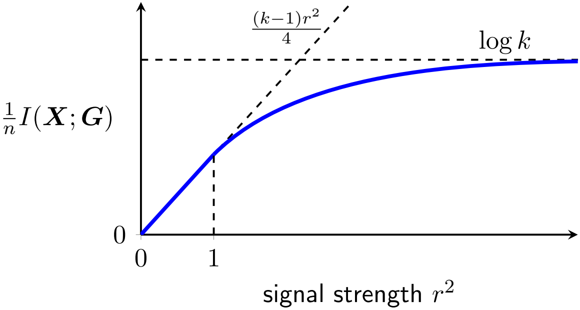
$$G_{ij} \sim \text{Bernoulli}\left(\frac{d}{n} + \frac{r\sqrt{d(1-d/n)}}{n} X_i^\top X_j\right), \quad i < j$$

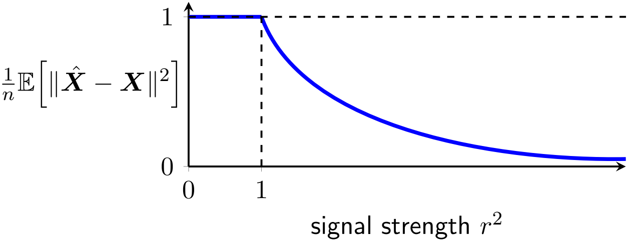
- d is the expected degree of each node
- r characterizes "community structure"
  - $r = 0 \implies$ no dependence
  - $r > 0 \implies$ assortative
  - $r < 0 \implies$ disassortative

Two-community SBM as $n, d \to \infty$ [Deshpande et al. 2015]



Two-community SBM as $n, d \to \infty$ [Deshpande et al. 2015]

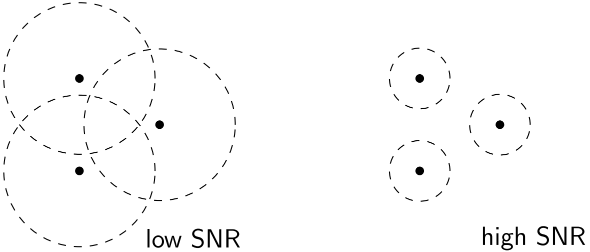




Signal-plus-noise problem

Gaussian noise model with signal-to-noise ratio $s \geq 0$:

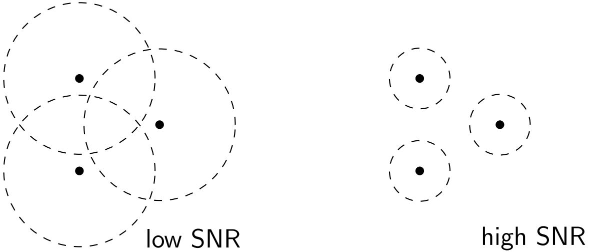
$$Y_i \sim \mathcal{N}(X_i, s^{-1}I), \quad i = 1, \ldots, n$$



Signal-plus-noise problem

Gaussian noise model with signal-to-noise ratio $s \geq 0$:

$$Y_i \sim \mathcal{N}(X_i, s^{-1} I), \quad i = 1, \ldots, n$$



MMSE function (can approximate via Monte Carlo integration)

$$M_{\boldsymbol{X}}(s) = \mathbb{E}\left[\| X_1 - \mathbb{E}[X_1 \mid Y_1] \|^2 \right]$$

Theorem

Let $s^* \geq 0$ be global minimizer of

$$I(X_1; Y_1) + \frac{k-1}{4}\left(r - \frac{s}{r}\right)^2$$

where Y_1 comes from signal-plus-noise problem. Then, the minimum MSE satisfies[1]

$$\frac{1}{n}\mathbb{E}\left[\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|^2\right] = M_{\boldsymbol{X}}(s^*) + o_{n,d}(1)$$

where $o_{n,d}(1) \to 0$ as $n, d \to \infty$.

[1] after resolving label invariances

Table of Contents

Spectrum of research on network inference

flexible models
limited theory

toy models
extensive theory

corrected SBM
mixed-membership

SBM

k-community
symm. SBM

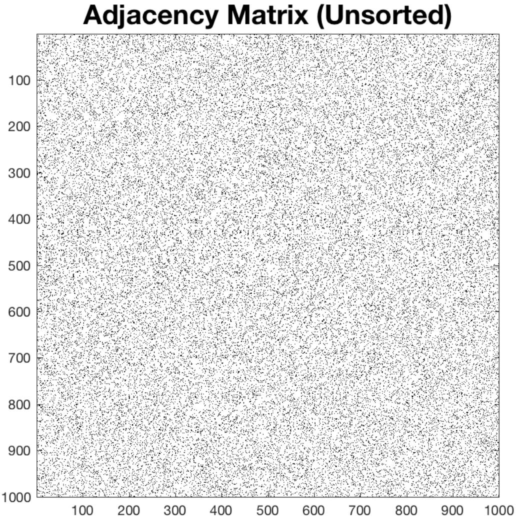
2-community
symm. SBM

applied network analysis
scientific modeling

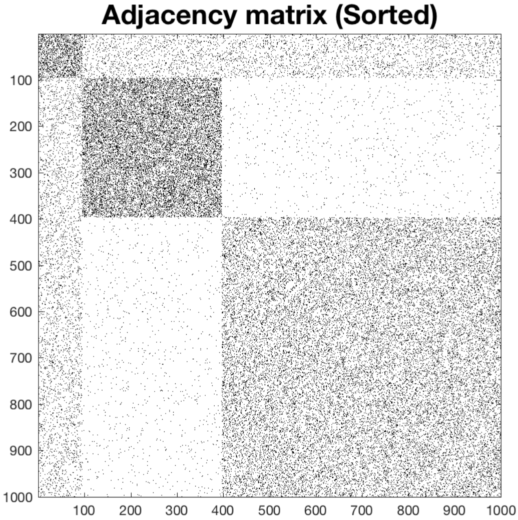
statistical
modeling

computer science
information theory
statistical physics

Degree-balanced SBM



Degree-balanced SBM



Degree-balanced SBM

Vertex labels $X_1, \ldots, X_n$ are i.i.d. over k communities with

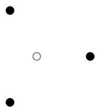
$$p = (p_1, \ldots, p_k)$$

Degree-balanced SBM

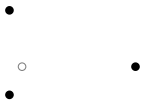
Vertex labels $X_1, \dots, X_n$ are i.i.d. over k communities with

$$p = (p_1, \dots, p_k)$$

Represent labels using k points in $(k-1)$ dimensions normalized to zero mean and identity covariance:



$p = (1/3, 1/3, 1/3)$



$p = (.1, .3, .6)$

Degree-balanced SBM

Entries of adjacency matrix are conditionally independent:

$$G_{ij} \sim \text{Bernoulli}\left(\frac{d}{n} + \frac{\sqrt{d(1-d/n)}}{n} X_i^\top R X_j\right), \quad i < j$$

- d is the expected degree of each vertex
- R is a symmetric matrix that characterizes "community structure"
  - $R = 0 \implies$ no dependence
  - $R \succ 0 \implies$ assortative
  - $R \prec 0 \implies$ disassortative

Multivariate performance metric

$$\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big[(X_i - \mathbb{E}[X_i \mid \boldsymbol{G}])(X_i - \mathbb{E}[X_i \mid \boldsymbol{G}])^\top\Big]$$

Data processing inequality + normalization of vertex labels:

$$0 \preceq \mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}) \preceq I_{k-1}$$

Multivariate performance metric

$$\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\Big[(X_i - \mathbb{E}[X_i \mid \boldsymbol{G}])(X_i - \mathbb{E}[X_i \mid \boldsymbol{G}])^\top\Big]$$

Data processing inequality + normalization of vertex labels:

$$0 \preceq \mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}) \preceq I_{k-1}$$

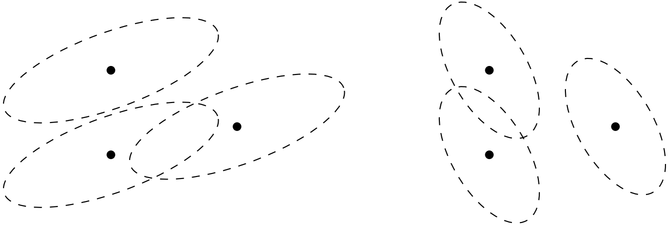
Trace corresponds to usual (scalar) MMSE

$$\mathrm{tr}(\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G})) = \frac{1}{n}\mathbb{E}\big[\|\boldsymbol{X} - \mathbb{E}[\boldsymbol{X} \mid \boldsymbol{G}]\|^2\big]$$

Signal-plus-noise problem

Gaussian noise model with positive semidefinite signal-to-noise ratio matrix $S \succeq 0$ given by:

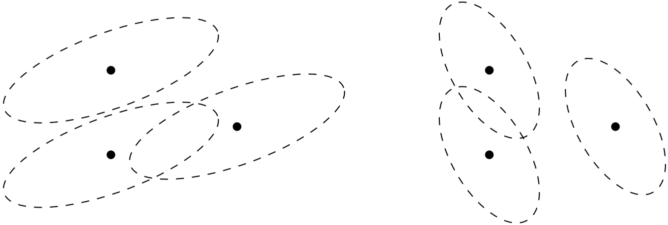
$$Y_i \sim \mathcal{N}(X_i, S^{-1}I), \quad i = 1, \ldots, n$$



Signal-plus-noise problem

Gaussian noise model with positive semidefinite signal-to-noise ratio matrix $S \succeq 0$ given by:

$$Y_i \sim \mathcal{N}(X_i, S^{-1} I), \quad i = 1, \ldots, n$$



MMSE function (can approximate via Monte Carlo integration)

$$M_{\boldsymbol{X}}(S) = \mathbb{E}\Big[(X_1 - \mathbb{E}[X_1 \mid Y_1])(X_1 - \mathbb{E}[X_1 \mid Y_1])^\top\Big]$$

Key identity

The matrix I-MMSE relation [Reeves et. al. 2018] states that

$$\nabla_S I(\boldsymbol{X}; \boldsymbol{G}, \boldsymbol{Y})\Big|_{S=0} = \frac{n}{2} \,\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G})$$

Theorem (R. , Mayya, Volfovsky, 2019)

Assume that R is definite and let $S^* \succeq 0$ be the global minimizer of

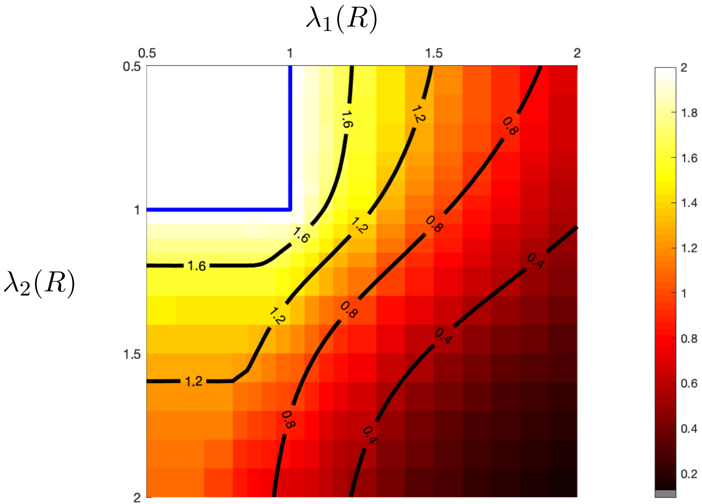
$$I(X_1; Y_1) + \frac{1}{4} \operatorname{tr}\big((R - R^{-1} S)^2\big).$$

Then, the MMSE matrix satisfies[2]

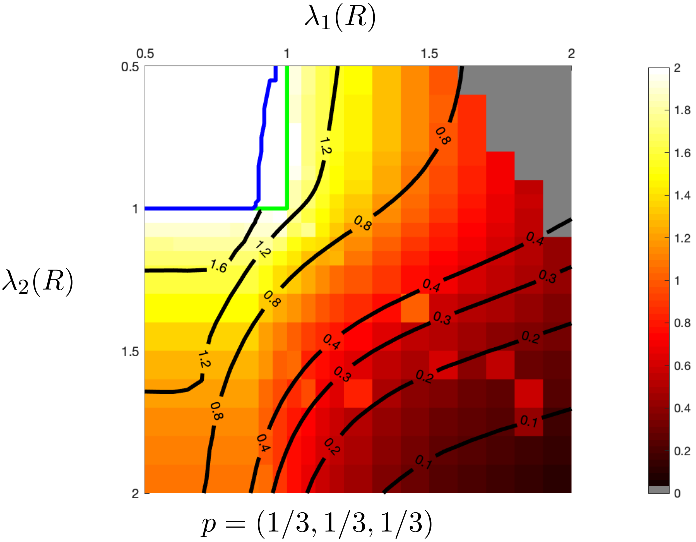
$$\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}) \preceq M_{\boldsymbol{X}}(S^*) + o_{n,d}(1),$$

where $o_{n,d}(1)$ denotes a symmetric matrix that converges to zero as $n, d \to \infty$.

[2] after resolving label invariances



Bound on $\mathrm{tr}(\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}))$ (contour lines) and empirical MSE of belief propagation (heat map) with $n = 100,000$, average degree $d = 10$, and $p = (1/3, 1/3, 1/3)$.



Bound on $\mathrm{tr}(\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{G}))$ (contour lines) and empirical MSE of belief propagation (heat map) with $n = 100,000$, average degree $d = 10$, and $p = (.1, .3, .6)$.

Table of Contents

Summary of ideas:

- Theoretical limits via mutual information and MMSE
- Symmetric SBM $\iff$ signal+noise model, scalar SNR
- General SBM $\iff$ signal + noise model, matrix SNR

Summary of ideas:

- Theoretical limits via mutual information and MMSE
- Symmetric SBM $\iff$ signal+noise model, scalar SNR
- General SBM $\iff$ signal + noise model, matrix SNR

Future directions

- Bridge gaps in theory / statistics / applied network inference
- Use geometric insight to inform methodology
- Impact of covariate information

[See posters by V. Mayya and H. Mathews]