The Geometry of Community Detection

Galen Reeves

Department of ECE and Department of Statistical Science
Duke University

September 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

Our Goals

- Broad class of network models (beyond symmetric models)
- Multivariate measures of performance
- Exact expressions for asymptotic performance
- Explore computational gaps

Paper on arxiv
`https://arxiv.org/abs/1907.02496`

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

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
- $\boldsymbol{X} = (X_1, \dots, X_n)$ contains vertex labels

How much do we learn?

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

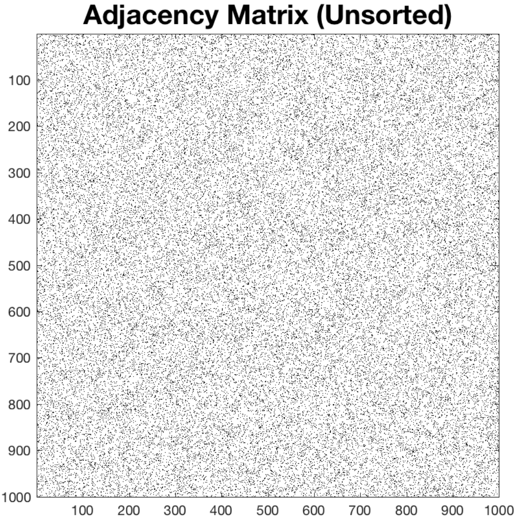
$$I(\boldsymbol{X}; \boldsymbol{G}) = \mathbb{E}\left[\log \frac{P(\boldsymbol{X}, \boldsymbol{G})}{P(\boldsymbol{X})P(\boldsymbol{G})}\right]$$

How well can we recover labels?

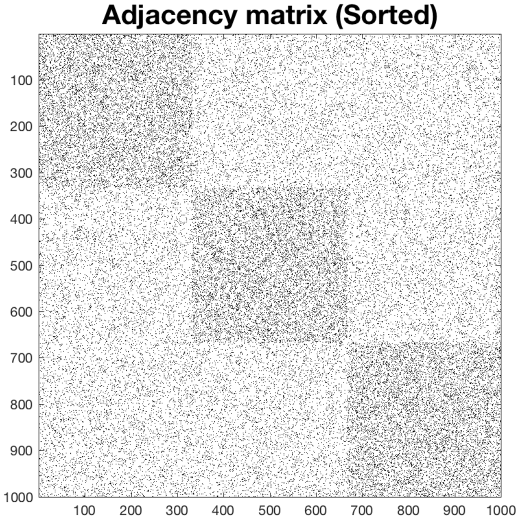
$$\hat{\boldsymbol{X}} \quad \text{verus} \quad \boldsymbol{X}$$

Table of Contents

A network with three similar communities



A network with three similar communities



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
- F. Krzakala, J. Xu, 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, 2017
- F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, 2018

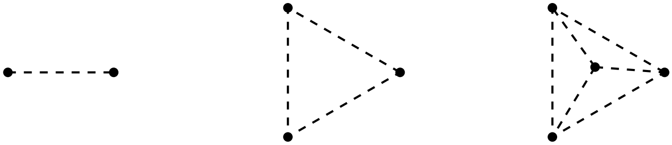
Parameterization of symmetric SBM

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

Parameterization of 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:

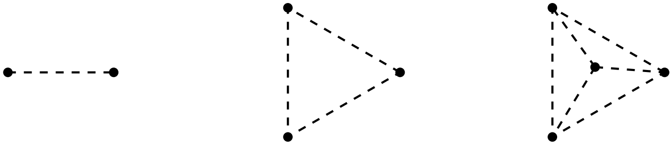


$k = 2$ $k = 3$ $k = 4$

Parameterization of 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 $\mathsf{Cov}(X_i) = I$, and hence

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

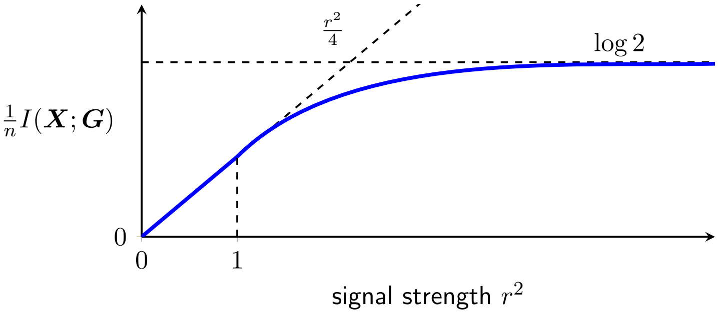
Parameterization of symmetric SBM

Entries of adjacency matrix are conditionally independent:

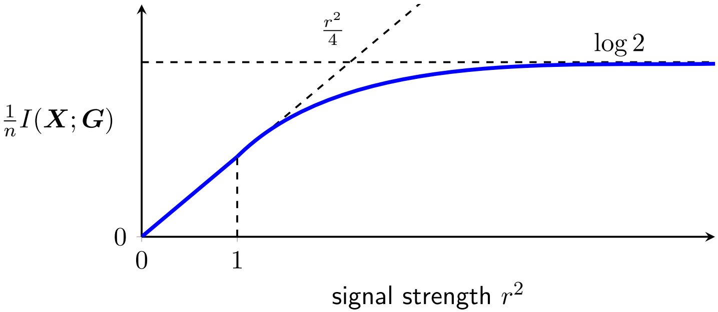
$$G_{ij} \sim \text{Bernoulli}\left(\frac{d}{n} + \frac{r\sqrt{d(1-d/n)}}{n} X_i^\top X_j\right), \quad i < j$$

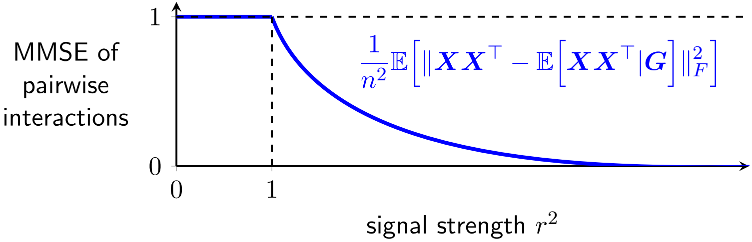
- d is the expected degree of each node
- r characterizes "community structure"
  - $r = 0 \implies$ no dependence
  - $r > 0 \implies$ assortative
  - $r < 0 \implies$ disassortative

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



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



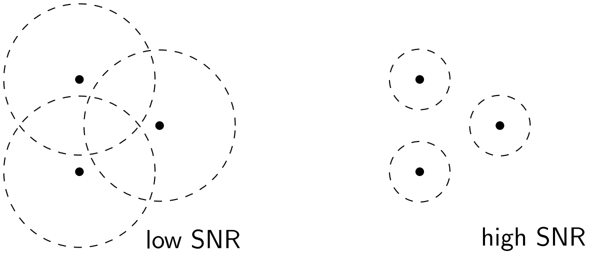


Where do these formulas come from?

The signal-plus-noise model

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

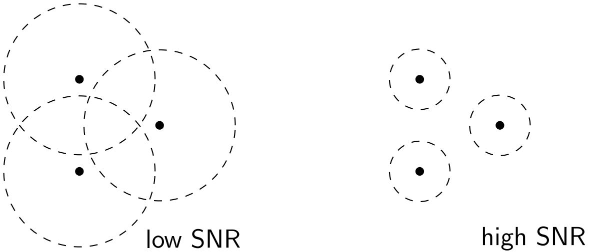
where $s > 0$ is the signal-to-noise ratio.



The signal-plus-noise model

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

where $s > 0$ is the signal-to-noise ratio.



Mutual information and MMSE functions (can approximate numerically)

$$I_X(s) = I(X_1; Y_1)$$
$$m_X(s) = \mathbb{E}\Big[\|X_1 - \mathbb{E}[X_1 \mid Y_1]\|^2\Big]$$

Theorem (Deshpande et al. 2015)

Assume $k = 2$ *communities. Let* $s^* \geq 0$ *be global minimizer of*

$$I_X(s) + \frac{1}{4}\Big(r - \frac{s}{r}\Big)^2$$

Then, MMSE of pairwise interactions satisfies

$$\frac{1}{n^2}\mathbb{E}\Big[\Big\|\boldsymbol{X}\boldsymbol{X}^\top - \mathbb{E}\big[\boldsymbol{X}\boldsymbol{X}^\top \mid \boldsymbol{G}\big]\Big\|_F^2\Big] = 1 - \frac{(s^*)^2}{r^4} + o_{n,d}(1),$$

Theorem (Deshpande et al. 2015)

Assume $k = 2$ communities. Let $s^* \geq 0$ be global minimizer of

$$I_X(s) + \frac{1}{4}\Big(r - \frac{s}{r}\Big)^2$$

Then, MMSE of pairwise interactions satisfies

$$\frac{1}{n^2}\mathbb{E}\Big[\Big\|\boldsymbol{X}\boldsymbol{X}^\top - \mathbb{E}\Big[\boldsymbol{X}\boldsymbol{X}^\top \mid \boldsymbol{G}\Big]\Big\|_F^2\Big] = 1 - \frac{(s^*)^2}{r^4} + o_{n,d}(1),$$

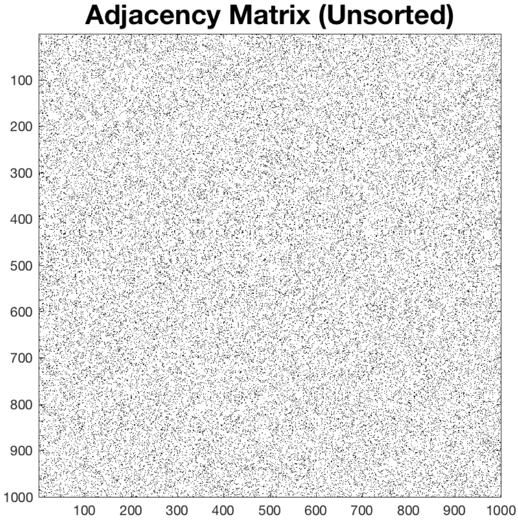
Furthermore

$$\inf_{\hat{\boldsymbol{X}}(\cdot)} \mathbb{E}\Big[\min_{\pi \in \{+1,-1\}} \frac{1}{n}\Big\|\boldsymbol{X} - \pi\hat{\boldsymbol{X}}(\boldsymbol{G})\Big\|^2\Big] \leq m(s^*) + o_{n,d}(1),$$

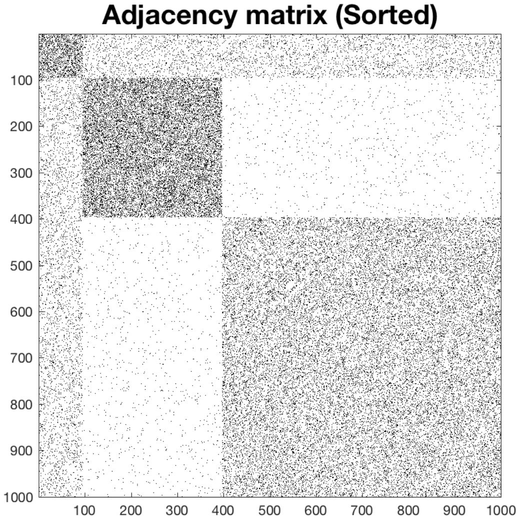
where the minimum over π resolves label invariance problem.

Table of Contents

A degree-balance network with three communities



A degree-balance network with three communities



Parameterization of degree-balanced SBM

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

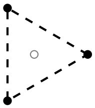
$$p = (p_1, \ldots, p_k)$$

Parameterization of degree-balanced SBM

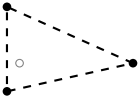
Vertex labels $X_1, \ldots, X_n$ are i.i.d. over k communities with

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

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



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



$p = (0.1, 0.3, 0.6)$

Parameterization of 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:

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

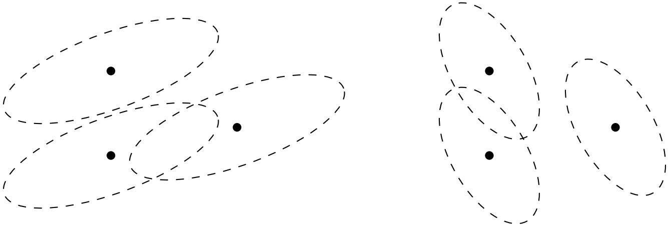
Trace corresponds to usual (scalar) MMSE

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

The signal-plus-noise model with matrix SNR

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

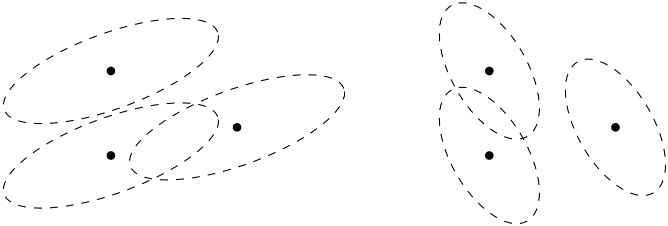
where S is positive semidefinite signal-to-noise ratio matrix.



The signal-plus-noise model with matrix SNR

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

where S is positive semidefinite signal-to-noise ratio matrix.



Mutual information and MMSE functions (can approximate numerically)

$$I_X(S) = I(X_1; Y_1)$$
$$M_X(S) = \mathbb{E}\Big[(X_1 - \mathbb{E}[X_1 \mid Y_1])(X_1 - \mathbb{E}[X_1 \mid Y_1])^\top\Big]$$

Theorem (R., Mayya, Volfovsky, 2019)

Assume that R is definite and let S^* be the global minimizer of

$$I_X(S) + \frac{1}{4}\,\mathrm{tr}\big((R - R^{-1}S)^2\big).$$

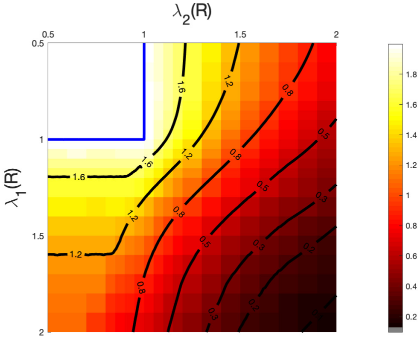
Then, the MMSE matrix satisfies[1]

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

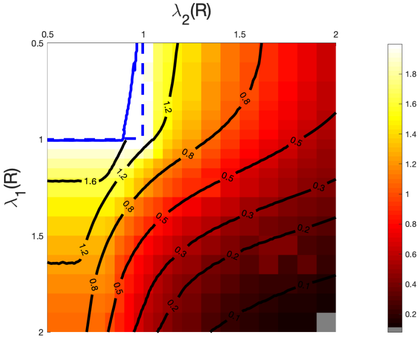
where $o_{n,d}(1)$ denotes a symmetric matrix that converges to zero as $n, d \to \infty$.

[1] after resolving label invariances

Comparison of theoretical and empirical results



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



$p = (0.6, 0.3, 0.1)$

Bound on tr(MMSE($\boldsymbol{X} \mid \boldsymbol{G}$)) (contour lines) and empirical MSE of belief propagation (heat map) with $n = 10^5$, average degree $d = 30$.

Table of Contents

Express MMSE as gradient of mutual information

Signal-plus-noise problem (in matrix notation)

$$\underbrace{\boldsymbol{Y}}_{n\times\ell} = \underbrace{\boldsymbol{X}}_{n\times\ell}\underbrace{S^{1/2}}_{\ell\times\ell} + \underbrace{\text{Gaussian noise}}_{n\times\ell}$$

Express MMSE as gradient of mutual information

Signal-plus-noise problem (in matrix notation)

$$\underbrace{\boldsymbol{Y}}_{n\times\ell} = \underbrace{\boldsymbol{X}}_{n\times\ell}\underbrace{S^{1/2}}_{\ell\times\ell} + \underbrace{\text{Gaussian noise}}_{n\times\ell}$$

By Matrix I-MMSE relation [R. Pfister, Dytso 2018],

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

Holds for any joint distribution on $(\boldsymbol{X}, \boldsymbol{G})$!

Express MMSE as gradient of mutual information

Signal-plus-noise problem (in matrix notation)

$$\underbrace{\boldsymbol{Y}}_{n\times\ell} = \underbrace{\boldsymbol{X}}_{n\times\ell}\underbrace{S^{1/2}}_{\ell\times\ell} + \underbrace{\text{Gaussian noise}}_{n\times\ell}$$

By Matrix I-MMSE relation [R. Pfister, Dytso 2018],

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

Holds for any joint distribution on $(\boldsymbol{X}, \boldsymbol{G})$!

$$\text{Goal:} \quad \lim_{n\to\infty}\frac{1}{n}I(\boldsymbol{X};\boldsymbol{G},\boldsymbol{Y}), \qquad S \succeq 0$$

Channel universality

Comparison between SBM and noisy matrix estimation

$$\boldsymbol{G} \sim \text{Bernoulli}\left(\frac{d}{n} + \frac{\sqrt{d(1-d/n)}}{n}\boldsymbol{X}R\boldsymbol{X}^\top\right)$$

$$\boldsymbol{Z} = \sqrt{\frac{t}{n}}\boldsymbol{X}R\boldsymbol{X}^T + \text{Gaussian noise}$$

Channel universality

Comparison between SBM and noisy matrix estimation

$$\boldsymbol{G} \sim \text{Bernoulli}\left(\frac{d}{n} + \frac{\sqrt{d(1-d/n)}}{n}\boldsymbol{X}R\boldsymbol{X}^\top\right)$$

$$\boldsymbol{Z} = \sqrt{\frac{t}{n}}\boldsymbol{X}R\boldsymbol{X}^T + \text{Gaussian noise}$$

Theorem (R., Mayya, Volfovsky, 2019)

If $t = 1$, *then*

$$\lim_{n,d\to\infty} \frac{1}{n}|I(\boldsymbol{X};\boldsymbol{G},\boldsymbol{Y}) - I(\boldsymbol{X};\boldsymbol{Y},\boldsymbol{Z})| = 0$$

Channel universality

Comparison between SBM and noisy matrix estimation

$$\boldsymbol{G} \sim \text{Bernoulli}\left(\frac{d}{n} + \frac{\sqrt{d(1-d/n)}}{n}\boldsymbol{X}R\boldsymbol{X}^\top\right)$$

$$\boldsymbol{Z} = \sqrt{\frac{t}{n}}\boldsymbol{X}R\boldsymbol{X}^T + \text{Gaussian noise}$$

Theorem (R., Mayya, Volfovsky, 2019)

If $t = 1$*, then*

$$\lim_{n,d\to\infty} \frac{1}{n}|I(\boldsymbol{X};\boldsymbol{G},\boldsymbol{Y}) - I(\boldsymbol{X};\boldsymbol{Y},\boldsymbol{Z})| = 0$$

Generalizes prior work [Deshpande et al. 2015], [Krzakala et al. 2016]

Interpolation via mutual information

$$\boldsymbol{Y} = \boldsymbol{X} S^{1/2} + \text{Gaussian noise}$$
$$\boldsymbol{Z} = \sqrt{\frac{t}{n}} \boldsymbol{X} R \boldsymbol{X}^T + \text{Gaussian noise}$$

Define interpolating function $\mathcal{I} : \mathbb{S}_+^d \times [0, \infty) \to [0, \infty)$

$$\mathcal{I}(S, t) = \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{Y}, \boldsymbol{Z})$$

Interpolation via mutual information

$$\boldsymbol{Y} = \boldsymbol{X} S^{1/2} + \text{Gaussian noise}$$
$$\boldsymbol{Z} = \sqrt{\frac{t}{n}} \boldsymbol{X} R \boldsymbol{X}^T + \text{Gaussian noise}$$

Define interpolating function $\mathcal{I} : \mathbb{S}_+^d \times [0,\infty) \to [0,\infty)$

$$\mathcal{I}(S,t) = \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{Y}, \boldsymbol{Z})$$

- $t = 0$ is signal-plus-noise model:

$$\mathcal{I}(S,0) = \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{Y})$$

Interpolation via mutual information

$$\boldsymbol{Y} = \boldsymbol{X} S^{1/2} + \text{Gaussian noise}$$

$$\boldsymbol{Z} = \sqrt{\frac{t}{n}} \boldsymbol{X} R \boldsymbol{X}^T + \text{Gaussian noise}$$

Define interpolating function $\mathcal{I} : \mathbb{S}_+^d \times [0, \infty) \to [0, \infty)$

$$\mathcal{I}(S, t) = \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{Y}, \boldsymbol{Z})$$

- $t = 0$ is signal-plus-noise model:

$$\mathcal{I}(S, 0) = \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{Y})$$

- $t = 1$ is desired goal (by channel universality):

$$\mathcal{I}(S, 1) = \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{G}, \boldsymbol{Y}) + o_{n,d}(1)$$

Estimation inequality for gradients

By I-MMSE relation,

$$\nabla_S \mathcal{I}(S,t) = \frac{1}{2}\,\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{Y}, \boldsymbol{Z})$$
$$\nabla_t \mathcal{I}(S,t) = \frac{1}{4}\frac{1}{n^2}\mathbb{E}\left[\left\|\boldsymbol{X} R \boldsymbol{X}^\top - \mathbb{E}\left[\boldsymbol{X} R \boldsymbol{X}^\top \mid \boldsymbol{G}\right]\right\|^2\right]$$

Estimation inequality for gradients

By I-MMSE relation,

$$\nabla_S \mathcal{I}(S,t) = \frac{1}{2}\,\mathsf{MMSE}(\boldsymbol{X} \mid \boldsymbol{Y}, \boldsymbol{Z})$$
$$\nabla_t \mathcal{I}(S,t) = \frac{1}{4}\frac{1}{n^2}\mathbb{E}\Big[\Big\|\boldsymbol{X}R\boldsymbol{X}^\top - \mathbb{E}\Big[\boldsymbol{X}R\boldsymbol{X}^\top \mid \boldsymbol{G}\Big]\Big\|^2\Big]$$

Rearranging terms + Jensen's inequality yields

$$\nabla_t \mathcal{I}(S,t) \le \frac{1}{4} g\big(2\nabla_S \mathcal{I}(S,t)\big),$$

where $g : \mathbb{S}_+^d \to [0,\infty)$ is given by

$$g(U) = \frac{1}{n^2}\,\mathrm{tr}\Big(\mathbb{E}\Big[\big(R\boldsymbol{X}^T\boldsymbol{X}\big)^2\Big]\Big) - \mathrm{tr}\left(\left(R\left(\frac{1}{n}\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - U\right)\right)^2\right).$$

Information inequality via duality

Conjugate function

$$\mathcal{J}(U,t) = \sup_{S \succeq 0} \left\{ \mathcal{I}(S,t) - \frac{1}{2}\langle S, U \rangle \right\}$$

Dual variable U corresponds to MMSE matrix

$$U = \nabla_S \mathcal{I}(S^*,t) \quad \iff \quad \mathcal{J}(U,t) = \mathcal{I}(S^*,t) - \frac{1}{2}\langle S^*, U \rangle$$

Information inequality via duality

Conjugate function

$$\mathcal{J}(U,t) = \sup_{S \succeq 0} \left\{ \mathcal{I}(S,t) - \frac{1}{2}\langle S, U \rangle \right\}$$

Dual variable U corresponds to MMSE matrix

$$U = \nabla_S \mathcal{I}(S^*,t) \quad \iff \quad \mathcal{J}(U,t) = \mathcal{I}(S^*,t) - \frac{1}{2}\langle S^*, U \rangle$$

By envelope theorem + inequality for gradients

$$\nabla_t \mathcal{J}(U,t) = \nabla_t \mathcal{I}(S^*,t) \le \frac{1}{4} g(\nabla_S \mathcal{I}(S^*,t)) = \frac{1}{4} g(U)$$

Information inequality via duality

Conjugate function

$$\mathcal{J}(U,t) = \sup_{S \succeq 0} \left\{ \mathcal{I}(S,t) - \frac{1}{2}\langle S, U \rangle \right\}$$

Dual variable U corresponds to MMSE matrix

$$U = \nabla_S \mathcal{I}(S^*, t) \quad \iff \quad \mathcal{J}(U,t) = \mathcal{I}(S^*, t) - \frac{1}{2}\langle S^*, U \rangle$$

By envelope theorem + inequality for gradients

$$\nabla_t \mathcal{J}(U,t) = \nabla_t \mathcal{I}(S^*, t) \le \frac{1}{4} g(\nabla_S \mathcal{I}(S^*, t)) = \frac{1}{4} g(U)$$

Integrating gives simple upper bound:

$$\mathcal{J}(U,1) \le \mathcal{J}(U,0) + \frac{1}{4} g(U)$$

Information inequality via duality (cont.)

For any joint distribution on $(\boldsymbol{X}, \boldsymbol{G})$,

$$\mathcal{I}(S, 1) = \inf_{U \succeq 0} \left\{ \mathcal{J}(U, 1) + \frac{1}{2} \langle S, U \rangle \right\}$$

$I(\cdot, 1)$ is concave

.

Information inequality via duality (cont.)

For any joint distribution on $(\boldsymbol{X}, \boldsymbol{G})$,

$$\mathcal{I}(S, 1) = \inf_{U \succeq 0} \left\{ \mathcal{J}(U, 1) + \frac{1}{2} \langle S, U \rangle \right\} \qquad I(\cdot, 1) \text{ is concave}$$

$$\leq \inf_{U \succeq 0} \left\{ \mathcal{J}(U, 0) + \frac{1}{4} g(U) + \frac{1}{2} \langle S, U \rangle \right\} \qquad \text{from gradient inq.}$$

.

Information inequality via duality (cont.)

For any joint distribution on $(\boldsymbol{X}, \boldsymbol{G})$,

$$
\begin{aligned}
\mathcal{I}(S,1) &= \inf_{U \succeq 0} \left\{ \mathcal{J}(U,1) + \frac{1}{2}\langle S, U \rangle \right\} && I(\cdot, 1) \text{ is concave} \\
&\le \inf_{U \succeq 0} \left\{ \mathcal{J}(U,0) + \frac{1}{4} g(U) + \frac{1}{2}\langle S, U \rangle \right\} && \text{from gradient inq.} \\
&= \inf_{U \succeq 0} \inf_{\Delta \succeq 0} \left\{ \mathcal{J}(U,0) + \frac{1}{4} h(\Delta) + \frac{1}{2}\langle S + \Delta, U \rangle \right\} && h \text{ conjugate to } g
\end{aligned}
$$

.

Information inequality via duality (cont.)

For any joint distribution on $(\boldsymbol{X}, \boldsymbol{G})$,

$$\begin{aligned}
\mathcal{I}(S,1) &= \inf_{U \succeq 0} \left\{ \mathcal{J}(U,1) + \frac{1}{2}\langle S, U \rangle \right\} && I(\cdot,1) \text{ is concave} \\
&\le \inf_{U \succeq 0} \left\{ \mathcal{J}(U,0) + \frac{1}{4} g(U) + \frac{1}{2}\langle S, U \rangle \right\} && \text{from gradient inq.} \\
&= \inf_{U \succeq 0} \inf_{\Delta \succeq 0} \left\{ \mathcal{J}(U,0) + \frac{1}{4} h(\Delta) + \frac{1}{2}\langle S + \Delta, U \rangle \right\} && h \text{ conjugate to } g \\
&= \inf_{\Delta \succeq 0} \left\{ \mathcal{I}(S+\Delta, 0) + \frac{1}{4} h(\Delta) \right\}.
\end{aligned}$$

Information inequality via duality (cont.)

For any joint distribution on $(\boldsymbol{X}, \boldsymbol{G})$,

$$\begin{aligned}
\mathcal{I}(S,1) &= \inf_{U\succeq 0}\left\{\mathcal{J}(U,1)+\frac{1}{2}\langle S,U\rangle\right\} && I(\cdot,1) \text{ is concave}\\
&\le \inf_{U\succeq 0}\left\{\mathcal{J}(U,0)+\frac{1}{4}g(U)+\frac{1}{2}\langle S,U\rangle\right\} && \text{from gradient inq.}\\
&= \inf_{U\succeq 0}\inf_{\Delta\succeq 0}\left\{\mathcal{J}(U,0)+\frac{1}{4}h(\Delta)+\frac{1}{2}\langle S+\Delta,U\rangle\right\} && h \text{ conjugate to } g\\
&= \inf_{\Delta\succeq 0}\left\{\mathcal{I}(S+\Delta,0)+\frac{1}{4}h(\Delta)\right\}.
\end{aligned}$$

If rows of $\boldsymbol{X}$ are IID with $\mathbb{E}\left[X_1 X_1^\top\right] = I$, then

$$\mathcal{I}(S,0) = I_X(S), \qquad h(\Delta) = \operatorname{tr}\left((R - R^{-1}\Delta)^2\right) + o_n(1)$$

Final steps in proof

We showed that for all $S \succeq 0$,

$$\limsup_{n,d\to\infty} \frac{1}{n} I(\boldsymbol{X};\boldsymbol{G},\boldsymbol{Y}) \leq \inf_{\Delta\succeq 0}\left\{ I_X(S+\Delta,0) + \frac{1}{4}\operatorname{tr}\big((R - R^{-1}\Delta)^2\big)\right\}$$

Final steps in proof

We showed that for all $S \succeq 0$,

$$\limsup_{n,d\to\infty} \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{G}, \boldsymbol{Y}) \le \inf_{\Delta \succeq 0} \left\{ I_X(S + \Delta, 0) + \frac{1}{4} \operatorname{tr}\big((R - R^{-1}\Delta)^2\big) \right\}$$

Matching lower bound is much more difficult! We use matrix estimation result of Lelarge, Miolane 2018 to show inequality is tight for $S = 0$.

$$\lim_{n,d\to\infty} \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{G}) = \inf_{\Delta \succeq 0} \left\{ I_X(\Delta, 0) + \frac{1}{4} \operatorname{tr}\big((R - R^{-1}\Delta)^2\big) \right\}$$

Final steps in proof

We showed that for all $S \succeq 0$,

$$\limsup_{n,d\to\infty} \frac{1}{n} I(\boldsymbol{X};\boldsymbol{G},\boldsymbol{Y}) \leq \inf_{\Delta\succeq 0}\left\{ I_X(S+\Delta,0) + \frac{1}{4}\operatorname{tr}\big((R - R^{-1}\Delta)^2\big)\right\}$$

Matching lower bound is much more difficult! We use matrix estimation result of Lelarge, Miolane 2018 to show inequality is tight for $S = 0$.

$$\lim_{n,d\to\infty} \frac{1}{n} I(\boldsymbol{X};\boldsymbol{G}) = \inf_{\Delta\succeq 0}\left\{ I_X(\Delta,0) + \frac{1}{4}\operatorname{tr}\big((R - R^{-1}\Delta)^2\big)\right\}$$

Comparing these expressions gives asymptotic upper bound on the gradient of the mutual information, which is the MMSE matrix.

Table of Contents

Conclusion

Contributions

- Study broad class of network models beyond the symmetric SBM.
- Characterize asymptotic performance via MMSE matrix.
- Novel interpolation method.
- Explore computation gap in asymmetric networks.

Conclusion

Contributions

- Study broad class of network models beyond the symmetric SBM.
- Characterize asymptotic performance via MMSE matrix.
- Novel interpolation method.
- Explore computation gap in asymmetric networks.

Future directions

- Bridge gaps in theory / statistics / applied network inference
- Use geometric insight to inform methodology
- Covariate information and other types of community structure.

Paper on arxiv

`https://arxiv.org/abs/1907.02496`