icassp 2022

Singapore, China, Online

# Learning Sparse Graphs with a Core-periphery Structure

**Sravanthi Gurugubelli**
Indian Institute of Science

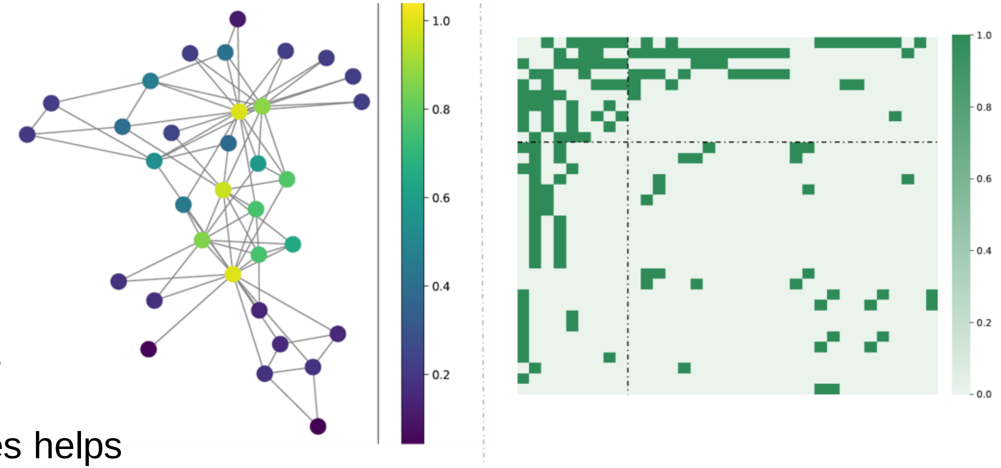
**Sundeep Prabhakar Chepuri**
Indian Institute of Science

# Core-Periphery Structure

- Core-periphery structure: Densely connected groups of core vertices and sparsely connected periphery vertices

- Core-periphery property is ubiquitous in
  - social networks
  - trade and transport networks
  - citation networks
  - communication networks
  - brain networks
  - genome-scale metabolic networks
  - protein–protein interaction networks

- Identifying the core and peripheral vertices helps in analyzing the central processes in networks

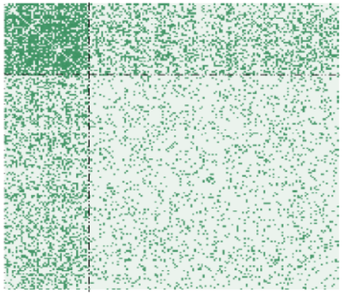


# Prior Art

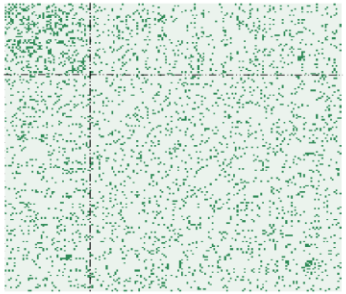
- Existing algorithms estimate core scores given the network topology
  - In many applications, we have access only to node attributes
  - The underlying graph structure may not always be available

- Conventional approaches to network topology inference do not readily incorporate a core-periphery structure

- We develop an approach that learns a core-periphery structured graph from node attributes so that the coreness of nodes are revealed implicitly



Adjacency matrix estimated from the proposed method



Adjacency matrix estimated from graphical lasso


. Rombach, M. Porter, J. Fowler, and P. Mucha, “Coreperiphery structure in networks.” SIAM J. Applied mathematics, vol. 74, no. 1, pp. 167–90, Oct. 2014.

. Della, F. Dercole, and P. C, “Profiling core-periphery network structure by random walkers.” Scientific reports, vol. 3, no. 1, pp. 1–8, Mar. 2013.

J. Jia and A. Benson, “Random spatial network models for core-periphery structure.” in In Proc. 12th ACM Int. Conf. on Web Search and Data Mining, New Orleans, USA, Jan. 2012.

D. Xiaowen, T. Dorina, R. Michael, and F. Pascal, “Learning graphs from data: A signal representation perspective,” IEEE Signal Process. Mag., vol. 36, no. 3, pp. 44–63, May 2019.

# Background: Gaussian Graphical Model

A weighted and an undirected graph: $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$

Feature matrix of $\mathcal{G}$: $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_d] \in \mathbb{R}^{N \times d}$

$$\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \ \forall i = 1, \ldots, d$$

Sparsity structure of the precision matrix $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$ encodes all the conditional dependencies between the $N$ variables associated with the vertices of $\mathcal{G}$

Graphical lasso learns the sparsity pattern in $\mathbf{\Theta}$ by solving

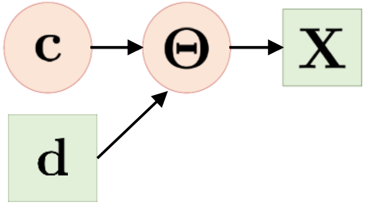
$$\underset{\mathbf{\Theta} \succeq 0}{\text{maximize}} \quad \log\det\mathbf{\Theta} - \mathrm{tr}(\mathbf{S}\mathbf{\Theta}) - \lambda\|\mathbf{\Theta}\|_1$$

$\mathbf{S}$: empirical covariance matrix

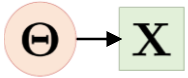
$\lambda > 0$: regularization parameter that controls the sparsity in $\mathbf{\Theta}$

Does not readily incorporate core-periphery structure!

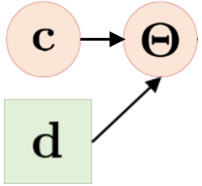
# Gaussian Graphical Model with a Core-periphery Structure




Modeling the dependence of the node attributes on the core scores through a latent graph structure




: $p(\mathbf{X}|\boldsymbol{\Theta}) \propto \det \boldsymbol{\Theta} \, \exp(-\operatorname{tr}(\mathbf{S}\boldsymbol{\Theta}))$

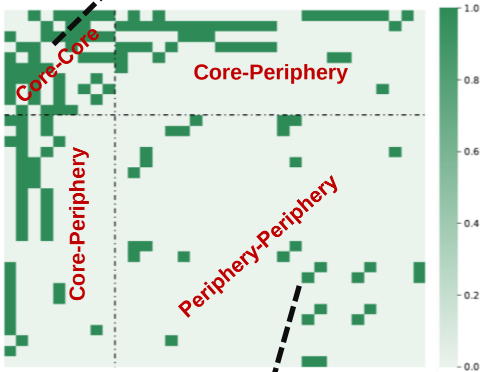



:

$$p(\boldsymbol{\Theta}; \mathbf{c}) = \prod_{i,j=1}^{N} p(\Theta_{ij}; c_i, c_j) = Z \prod_{i,j=1}^{N} \exp\left(-\lambda w_{ij} |\Theta_{ij}|\right)$$

$$w_{ij} = 1 - c_i - c_j + e \log(d_{ij})$$

$c_i + c_j - e \log(d_{ij})$ is large




$c_i + c_j - e \log(d_{ij})$ is small

# Proposed Learning Algorithm

We estimate the model parameters by maximizing the posterior distribution given data, i.e., by maximizing

$$l(\boldsymbol{\Theta}, \mathbf{c}) = \log p(\mathbf{X}|\boldsymbol{\Theta}) + \log p(\boldsymbol{\Theta}; \mathbf{c})$$

The proposed optimization problem:

$$\underset{\boldsymbol{\Theta} \succeq 0, \mathbf{c}}{\text{maximize}} \quad \boxed{\log \det \boldsymbol{\Theta} - \operatorname{tr}(\mathbf{S}\boldsymbol{\Theta})} \boxed{- \lambda \sum_{i,j=1}^{N} w_{ij} |\Theta_{ij}|}$$

$$\text{s. to} \quad w_{ij} = 1 - c_i - c_j + e \log(d_{ij})$$

$$w_{ij} > 0, \quad i, j = 1, 2, \ldots, N$$

$$\sum_{i=1}^{N} c_i = M, \quad c_i \in [0, 1], \quad i = 1, 2, \ldots, N$$

To prevent the case where all the weights tend to zero

To fix the scale of the core scores

# Proposed Learning Algorithm

known weights that
depend on $\mathbf{c}$

## Updating the graph

$$\underset{\boldsymbol{\Theta} \succeq 0}{\text{maximize}} \quad \log\det \boldsymbol{\Theta} - \mathrm{tr}(\mathbf{S}\boldsymbol{\Theta}) - \lambda \sum_{i,j=1}^{N} w_{ij} |\Theta_{ij}|$$

This is a convex program that can be solved using existing solvers, e.g., QUIC

## Updating the core scores

$$\underset{c_1, \cdots, c_N}{\text{maximize}} \quad \sum_{i,j=1}^{N} |\Theta_{ij}|(c_i + c_j)$$

$$\text{s. to} \quad \sum_{i=1}^{N} c_i = M, \;\; c_i \in [0,1],$$

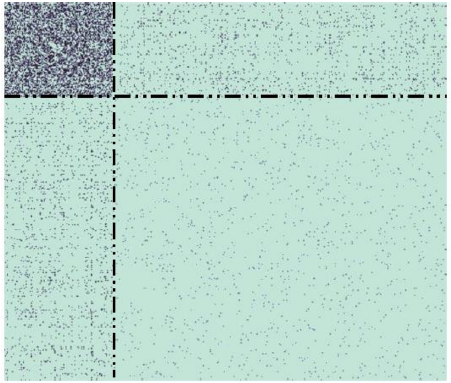
$$c_i + c_j < 1 + e\log(d_{ij}), \; i,j = 1, \cdots, N$$

This is a linear program that can be solved using standard off-the-shelf solvers

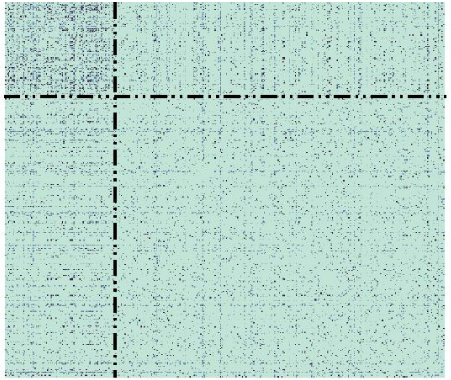
# Numerical Experiments: Model Evaluation

The core-periphery partitioning of the networks by the proposed method is similar to the others, in spite of not knowing the network directly!

$$\mathbf{\Theta}_{\text{ideal}} = \left[\begin{array}{c|c} \mathbf{1}_{tt} & \mathbf{0}_{t(N-t)} \\ \hline \mathbf{0}_{(N-t)t} & \mathbf{0}_{(N-t)(N-t)} \end{array}\right]$$



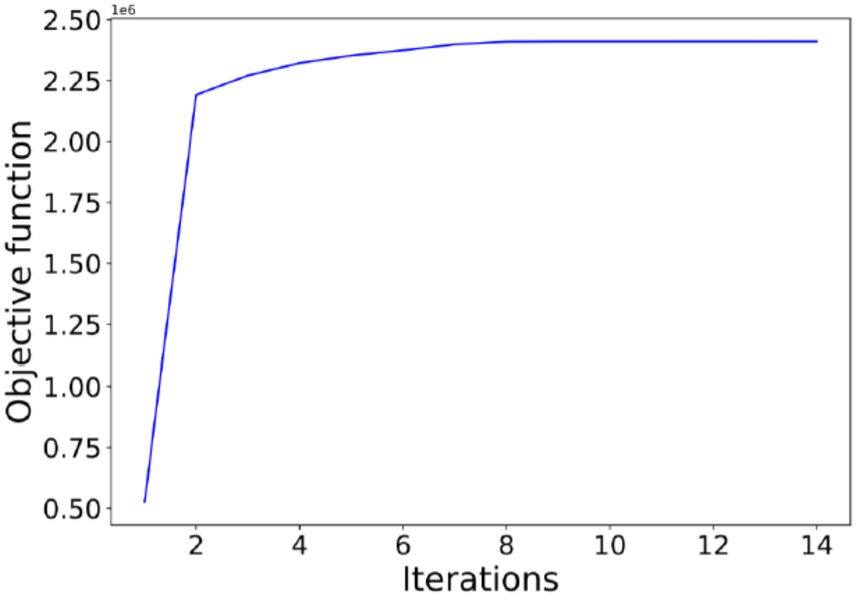
Adjacency matrix estimated from the proposed method



Adjacency matrix of the groundtruth network

| | | Proposed | MINRES | Rombach | RandomWalk | k-cores |
|---|---|---|---|---|---|---|
| Celegans | $\|\mathbf{\Theta}_0 - \mathbf{\Theta}_{\text{ideal}}\|_F^2$ | 41.940 | 41.821 | 39.076 | 40.877 | 39.051 |
| | $\||\mathbf{\Theta}| - \mathbf{\Theta}_{\text{ideal}}\|_F^2$ | 32.642 | 32.841 | 32.538 | 32.707 | 32.748 |
| Cora | $\|\mathbf{\Theta}_0 - \mathbf{\Theta}_{\text{ideal}}\|_F^2$ | 55.488 | 55.434 | 54.690 | 55.326 | 54.909 |
| | $\||\mathbf{\Theta}| - \mathbf{\Theta}_{\text{ideal}}\|_F^2$ | 47.626 | 55.722 | 47.884 | 47.983 | 47.625 |
| London underground | $\|\mathbf{\Theta}_0 - \mathbf{\Theta}_{\text{ideal}}\|_F^2$ | 79.216 | 79.249 | 78.7563 | 79.338 | 79.169 |
| | $\||\mathbf{\Theta}| - \mathbf{\Theta}_{\text{ideal}}\|_F^2$ | 78.818 | 78.905 | 78.811 | 78.858 | 78.856 |
| Twitter | $\|\mathbf{\Theta}_0 - \mathbf{\Theta}_{\text{ideal}}\|_F^2$ | 134.692 | 137.142 | 124.112 | 131.278 | 129.221 |
| | $\||\mathbf{\Theta}| - \mathbf{\Theta}_{\text{ideal}}\|_F^2$ | 110.526 | 111.837 | 111.427 | 111.429 | 110.526 |

# Numerical Experiments: Convergence



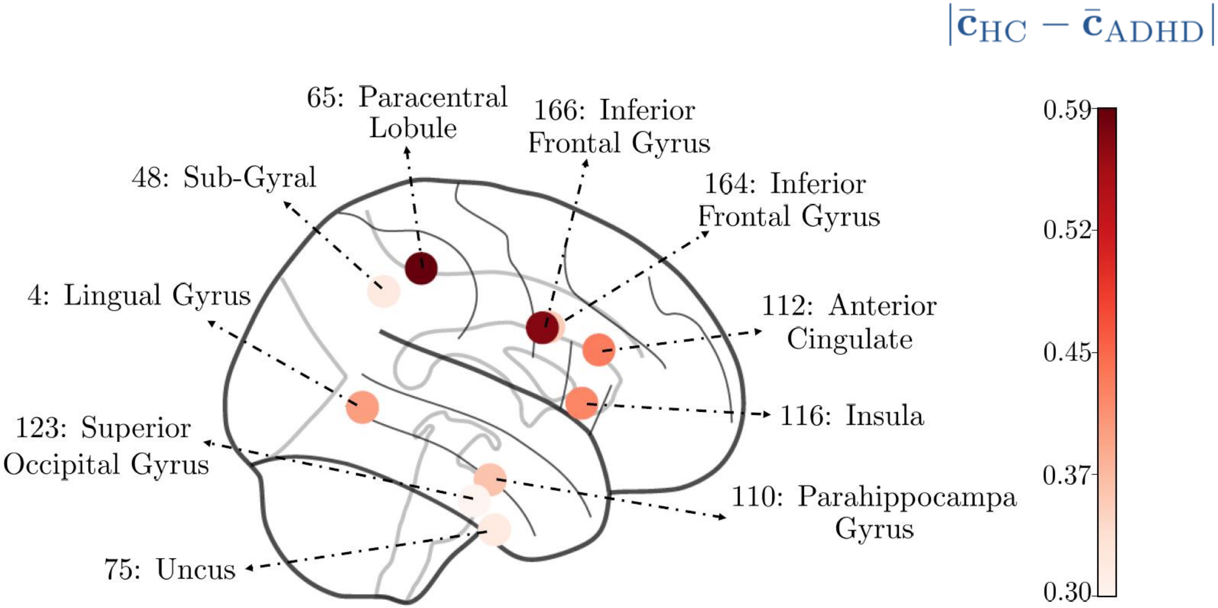

The proposed algorithm converges
in about 10 iterations

# Numerical Experiments: Brain Network Analysis

Data:
Functional MRI time series for the regions of interest in the cc200 parcellation for a total of 79 individuals
- 42 healthy subjects
- 37 subjects with ADHD

$\bar{\mathbf{c}}_{\mathrm{HC}}$ : average of the core score vectors of healthy subjects

$\bar{\mathbf{c}}_{\mathrm{ADHD}}$ : average of the core score vectors of subjects with ADHD




The regions with a large difference in the cores scores of the two groups coincide with the regions that have differences in activation for healthy individuals and patients with ADHD

S. Dickstein, K. Bannon, C. F. Xavier, and M. Milham, “The neural correlates of attention deficit hyperactivity disorder: An ale meta-analysis.” J. Child Psychology and Psychiatry, vol. 47, no. 10, pp. 1051–62, Nov. 2006.

# Conclusions

- We developed a generative model to relate node attributes to the core scores of vertices through a latent graph structure.
- We presented a joint estimator to simultaneously infer the vertex core scores and a sparse graph whose sparsity pattern is determined by the core scores.
- We presented a block coordinate ascent algorithm to solve the proposed estimation problem.
- We demonstrated via numerical experiments that the proposed method learns a core-periphery structured graph from only the node attributes while learning core scores on par with methods that use the ground truth network as input.
- We also applied our method to fMRI data to infer the regions that are the most affected in subjects with ADHD.

# Thank you!

This work is supported in part by the Pratiksha Trust Fellowship and SERB SRG/2019/000619.