# LEARNING BRAIN NETWORK DIFFERENCES USING STATISTICAL METHODS

Parinthorn Manomaisaowapak

Advisor: Assist. Prof. Jitkomut Songsiri

Department of Electrical Engineering, Faculty of Engineering

Chulalongkorn University

# PRESENTATION OUTLINE

- Objectives
- Introduction
- Background
- Methodology
- Results
- Conclusion

# OBJECTIVES

There are two objectives of this project,

- To estimate brain network using Granger causality concept from EEG or fMRI data.
- To compare brain network difference between control group and patient group.

**Already completed** in semester 1 (statistical framework).

We further explore formulations of sparse estimation on **simulated data.**

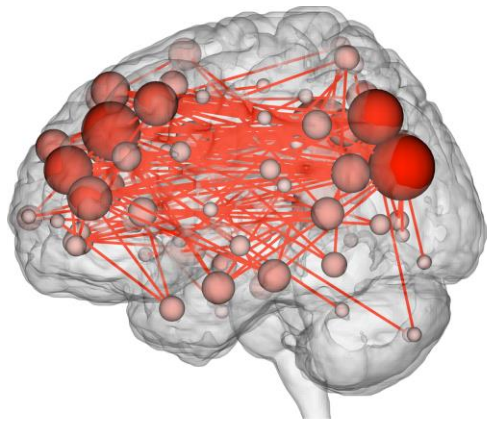
**This presentation** covered only our **sparse estimation framework.**

# INTRODUCTION

**What is sparse estimation ?**

- Human brains have **large amount of regions**.

- We aim to estimate simpler model to have only **important connection** between regions.



**The sparse estimation of brain connectivity.**



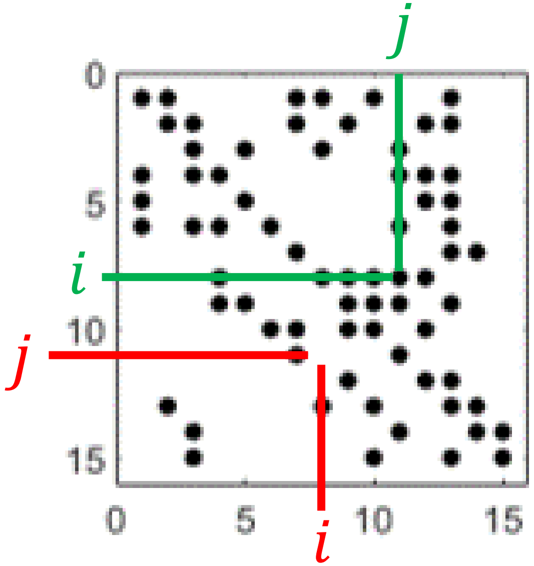
http://theconversation.com/brain-activity-is-as-unique-and-identifying-as-a-fingerprint-48723

# INTRODUCTION

## Effective brain connectivity : Measure of causality

Vector autoregressive (VAR) model

- Dynamic model → Mimicking brain signals.
- Granger causality → Represented as Granger Causality (GC) matrix.




GC matrix

Causal interaction has **direction**

Brain region $j$th $\underset{?}{\rightleftarrows}$ Brain region $i$th

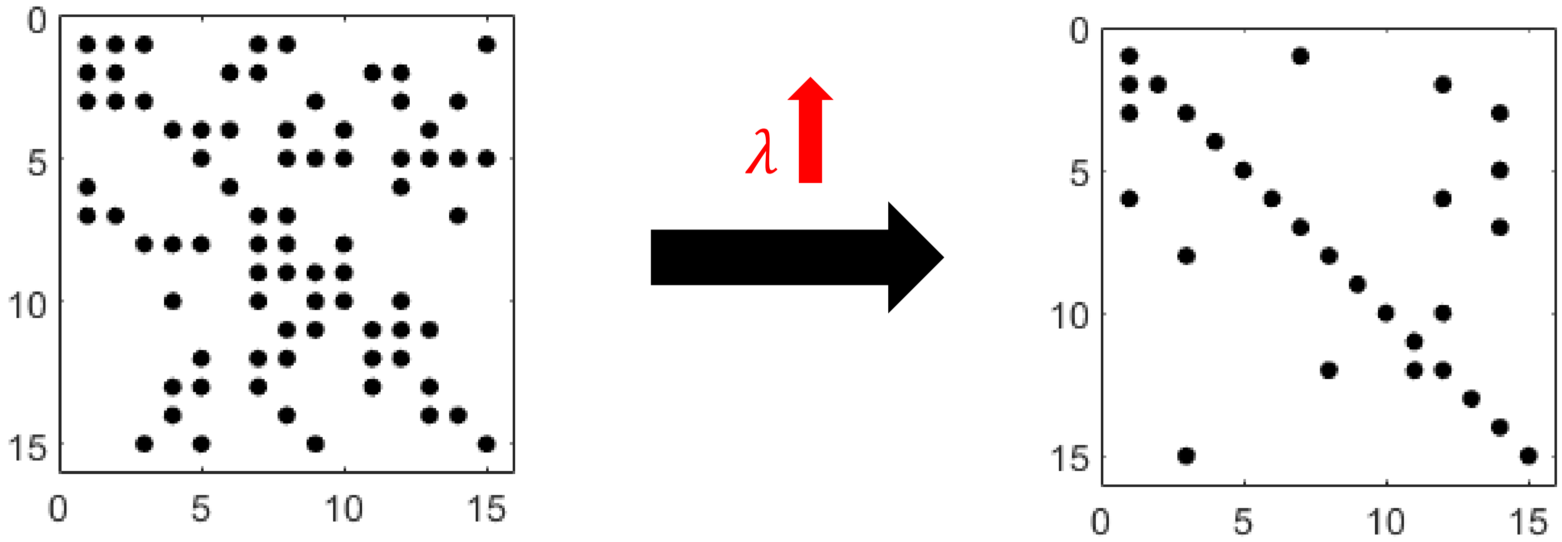
# BACKGROUND

**Sparse estimation formulation in general form.**

Model parameter

$$\min_{\theta} \underbrace{f(\theta)}_{\text{Fitting term}} + \lambda \underbrace{g(\theta)}_{\text{Regularization term}}$$

Regularization term → **Connection** between region **penalization**

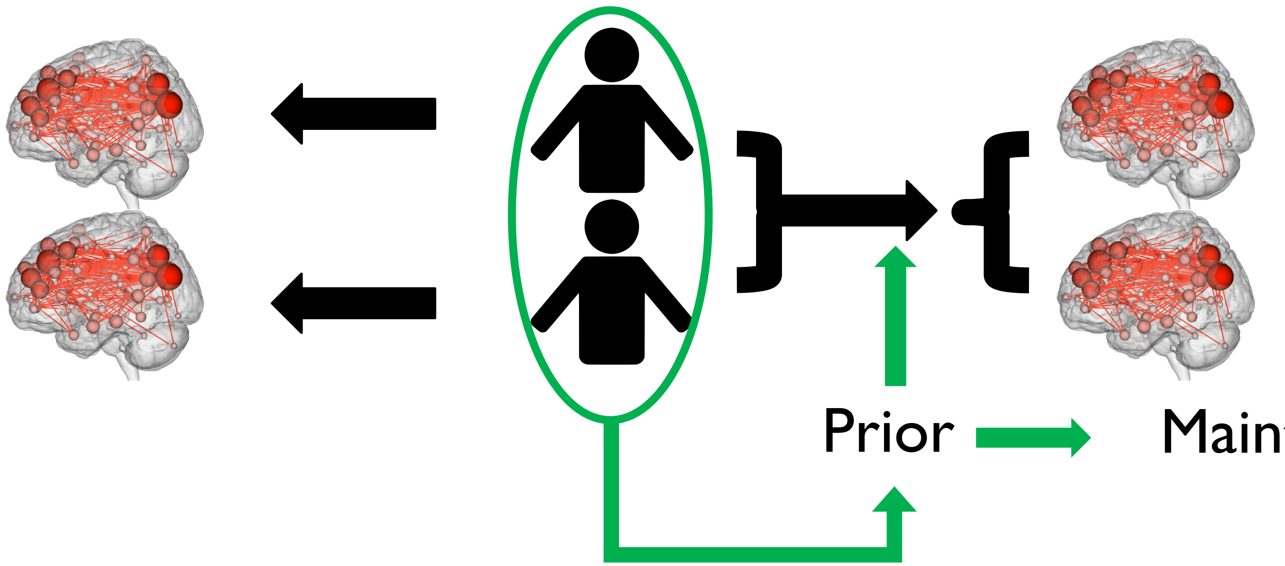



Granger causality (GC) matrix

# BACKGROUND

What if we have **multiple brain connectivity** to determine ?

Estimate **separately** vs Estimate **jointly**




Maintain accuracy with lower sample

→We can add prior knowledge on relation between those multiple models in joint estimation.

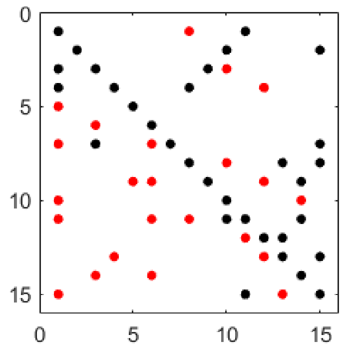
# BACKGROUND

For example, two models can be estimated from

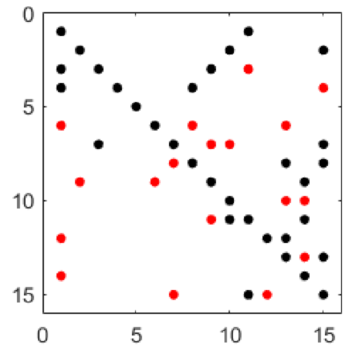
$$\min_{\theta_1,\theta_2} \overbrace{f_1(\theta_1) + f_2(\theta_2) + \lambda_1(g_1(\theta_1) + g_1(\theta_2))}^{\text{A. Individual estimation}} + \overbrace{\lambda_2 g_2(\theta_1, \theta_2)}^{\text{B. Joint estimation}}$$

where $g_1$ aims to promote different sparsity in each model.

$g_2$ aims to promote common sparsity across all models.




Model 1




Model 2

**But how to interpret these idea by VAR model ?**

Zero in VAR coefficients $\not\Rightarrow$ Zero in GC matrix

Not a necessary condition

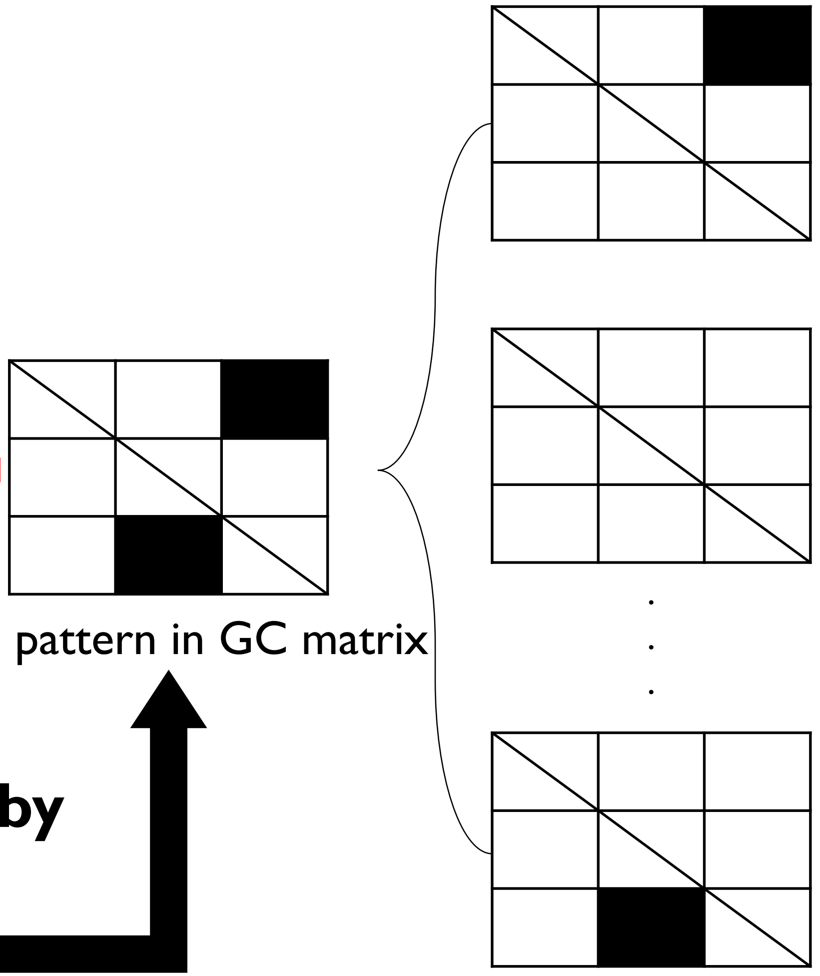
# BACKGROUND

$k$th VAR models can be described by

$$y^{(k)}(t) = \sum_{q=1}^{p} (A_q)^{(k)} y^{(k)}(t-q) + e(t)$$

which can be efficiently estimated by ordinary least square.

$$\left[\hat{A}_1^{(k)} \dots, \hat{A}_p^{(k)}\right] = A^{(k)} = \underset{A^{(k)}}{\operatorname{argmin}} \ (1/2) \left\| Y^{(k)} - A^{(k)} H^{(k)} \right\|_2^2$$




Zero pattern in GC matrix

Zero pattern in VAR model

**We can determine zero pattern in GC matrix by**

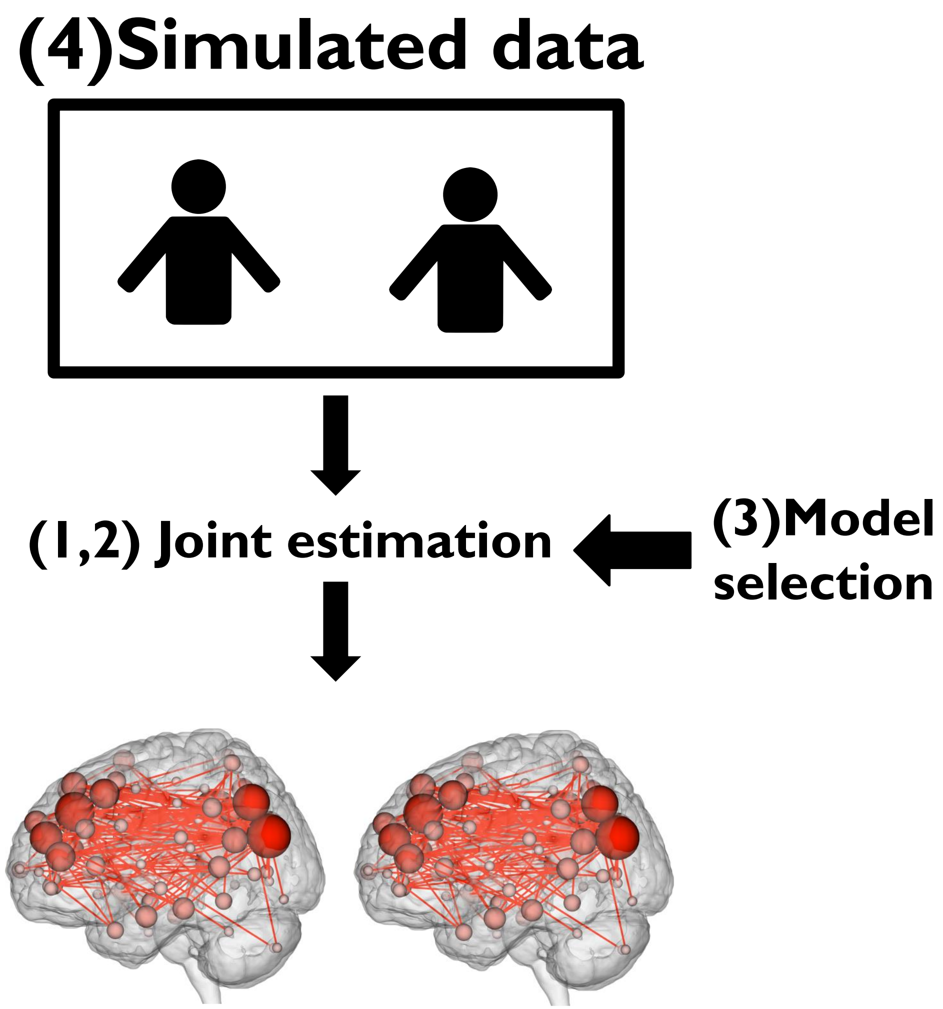
$$\boxed{F_{ij}^{(k)} \Leftrightarrow \left(A_q^{(k)}\right)_{ij} = 0; q = 1,2, \dots, p}$$

Where $F^{(k)}$ represents GC matrix of $k$th model.

# METHODOLOGY

Our methodology split into 4 parts,

1. Jointly sparse VAR estimation of brain networks.
2. Algorithm.
3. Model selection for learning brain networks.
4. Simulated data generation.



# METHODOLOGY | 1.JOINT ESTIMATION

We proposed three formulations,

Formulation **C** **Common**

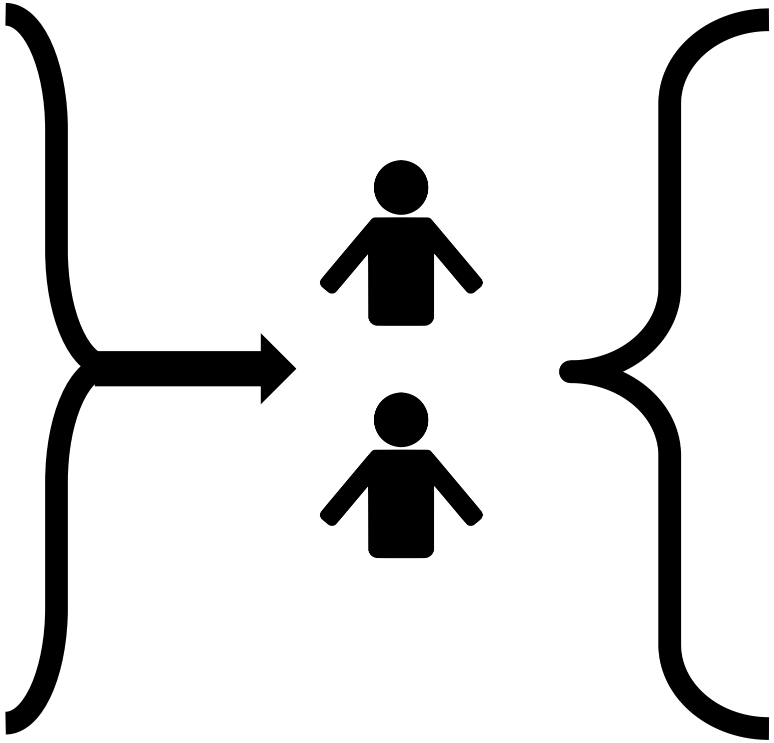
- **<u>Common</u>** pattern.

Formulation **D** **Differential**

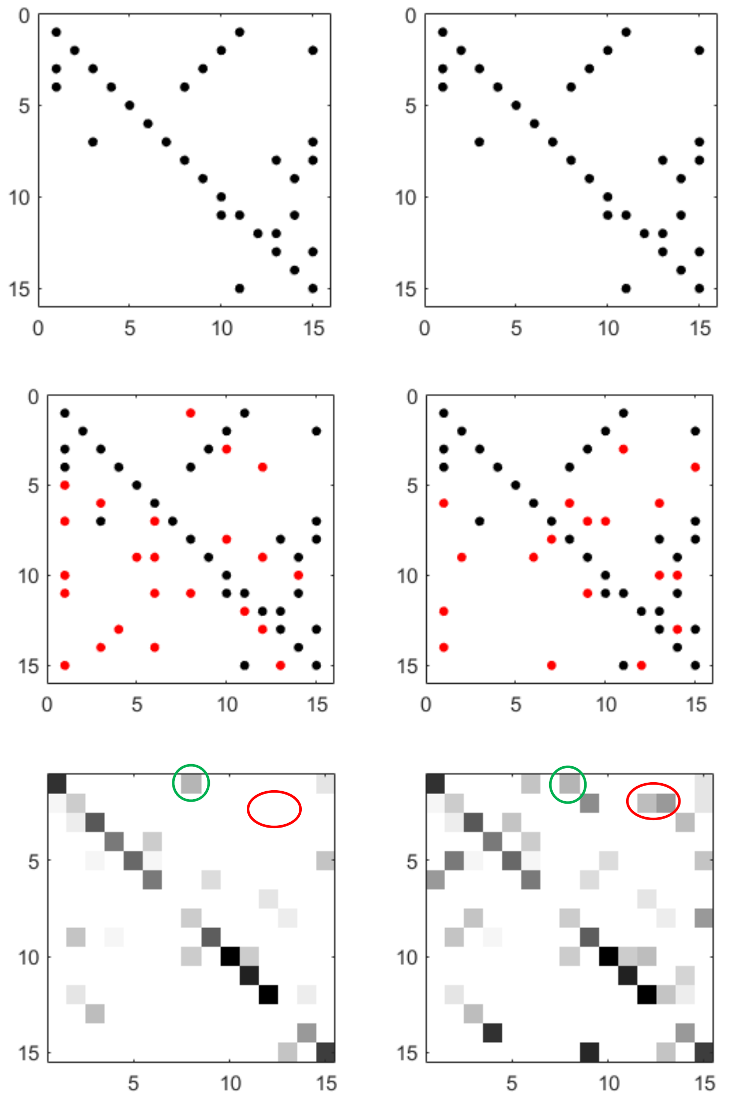
- **<u>Common</u>** pattern
- **<u>Different</u>** pattern

Formulation **S** **Similar**

- **<u>Shared VAR coefficients</u>** value
- Different pattern



Model 1 Model 2



Our formulation properties.

- Convex problem
- Have smooth fitting term
- But non-smooth in regularization term at zero

Gradient method does not work

Convex Programming in **CVX toolbox** have memory limitations.

We use ADMM (Alternating direction method of multipliers) solver

Require two predetermined tuning-parameters

# METHODOLOGY | 3.MODEL SELECTION

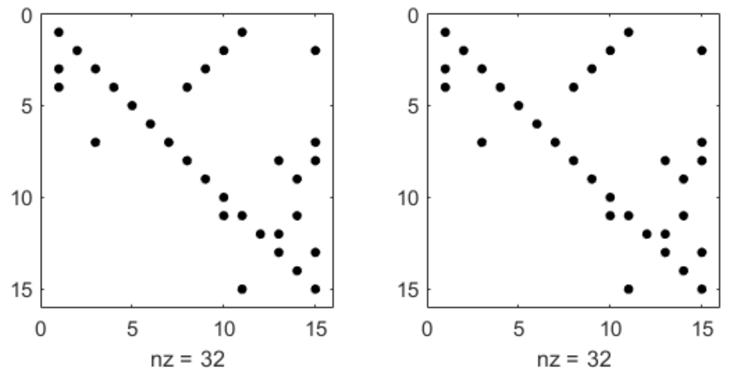
We used BIC criteria to find optimal tuning-parameters

$$BIC(\lambda_1, \lambda_2) = -2\,\mathcal{L} + \log(\mathrm{N}) \cdot \mathrm{df}$$

Log-likelihood for VAR model.
(Fitness of models)

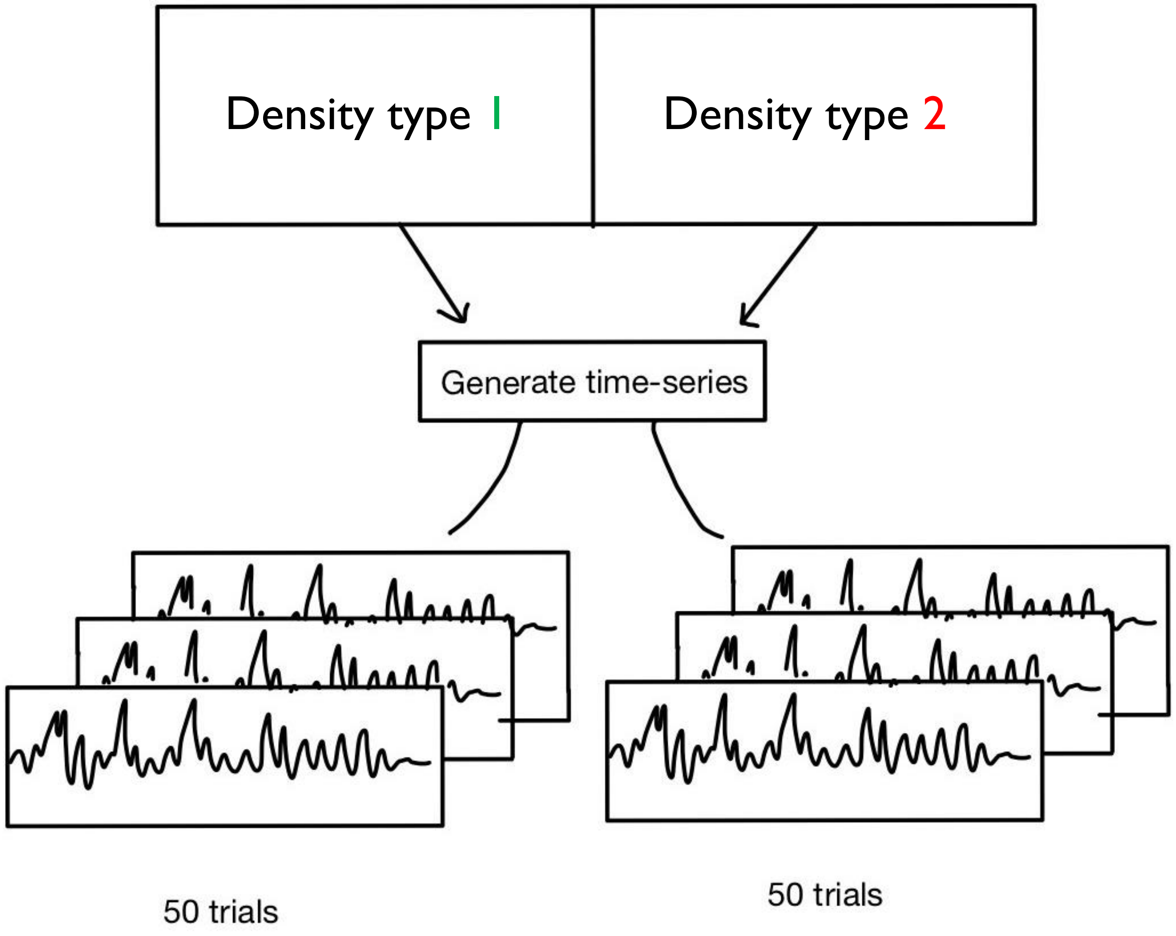
Effective degree of freedom
(Complexity of models).

# off-diagonal nonzero estimated parameters

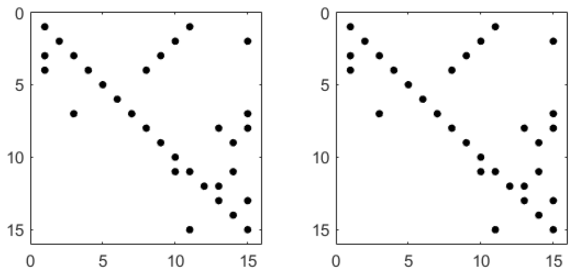


$n = 15, p = 3, K = 4$

We randomized stable VAR coefficients.

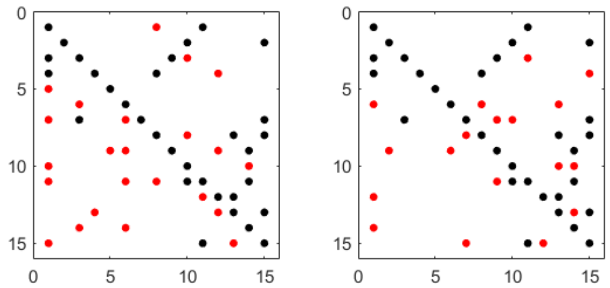



1. Common type ground truth



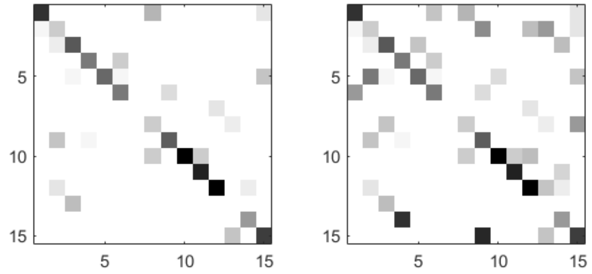
- Graph density : **0.1**, **0.3**
- **No** difference connection

2. Differential type ground truth



- Graph density : **0.1**
- Difference connection density : **0.1**, **0.3**

3. Similar type ground truth



- Graph density : **0.1**
- Difference connection density : **0.1**, **0.3**

We setup **three** experiments, each with different data assumptions.

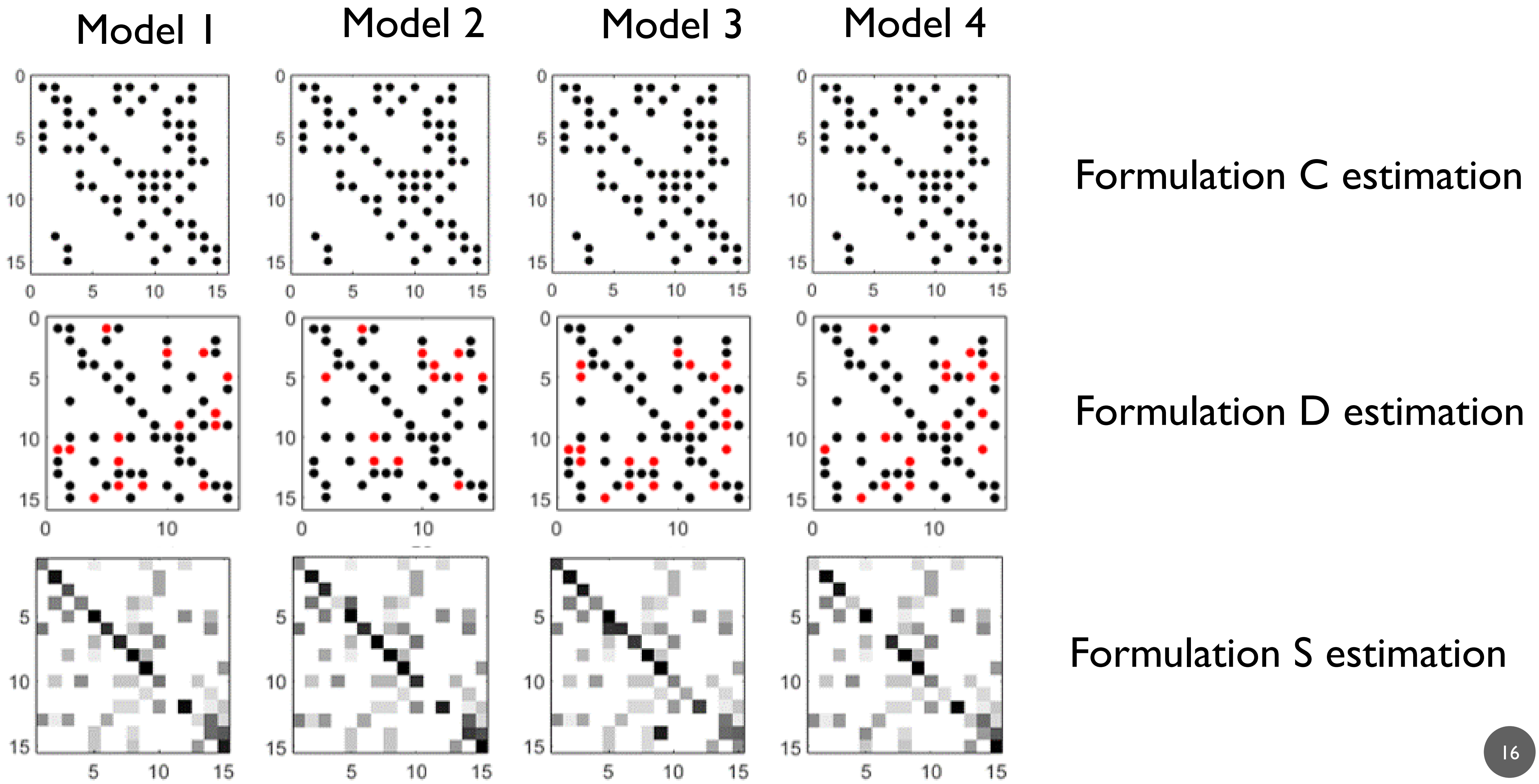
| Experiment 1 | Experiment 2 | Experiment 3 |
| --- | --- | --- |
| • Common type | • Differential type | • Similar type |

Estimate with formulation C, D, S

**TPR**, **FPR**, **ACC**, Area under ROC **(AUC)**, **parameter bias**

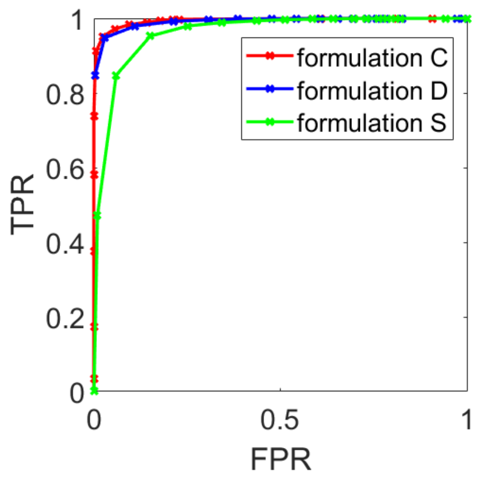
Calculated from VAR coefficients

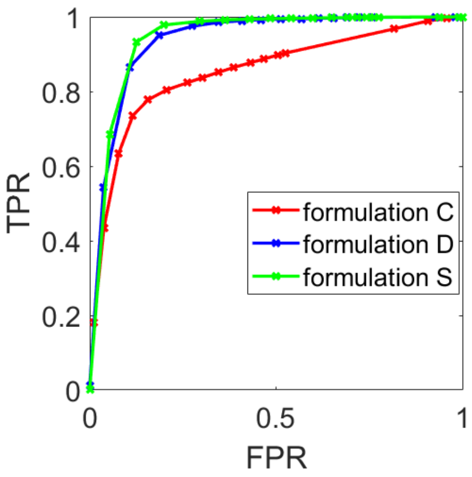
# RESULTS | GC MATRIX EXAMPLES

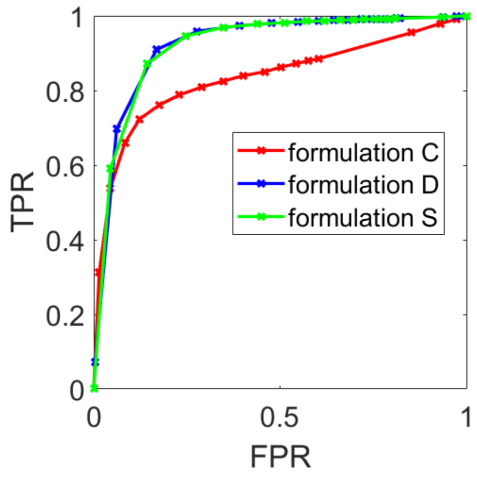


# RESULTS | ROC COMPARISON

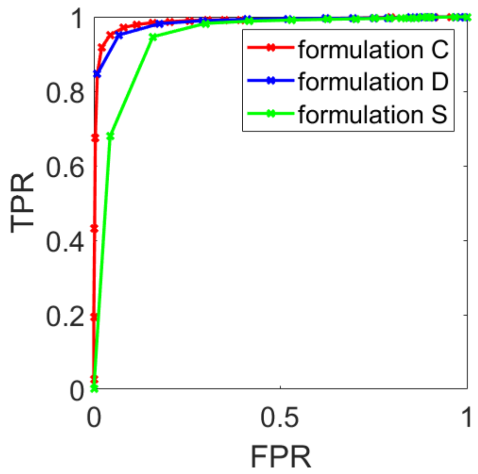
Density=0.1

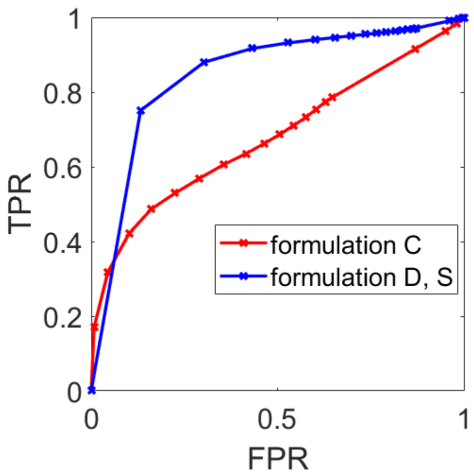


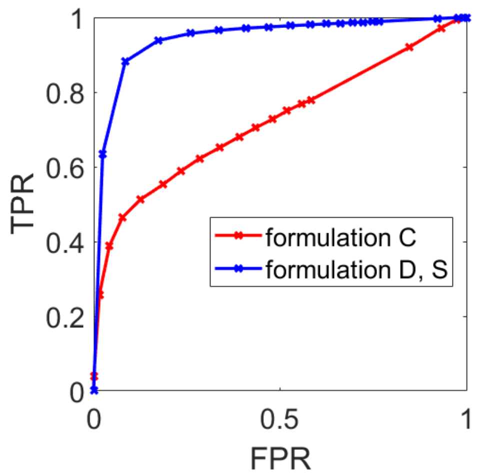




Density=0.3




Experiment 1 | Experiment 2 | Experiment 3

Common type | Differential type | Similar type

# CONCLUSION

- We developed three sparse estimation formulations depended on prior knowledge.

- Each assumption can be interpreted as group-level brain connectivity
- and individual-level connections.

- Each formulation performed best **if the assumptions on data are true**.

# Q&A

# SUPPLEMENTARY: FORMULATION COST FUNCTION

$$(1/2)\ \left\|Y^{(k)} - A^{(k)}H^{(k)}\right\|_2^2$$

Least square (individual)

$$\sum_{k=1}^{K}(1/2)\ \left\|Y^{(k)} - A^{(k)}H^{(k)}\right\|_2^2$$

Least square (joint)

$$A^{(k)} = \left[\hat{A}_1^{(k)} \dots, \hat{A}_p^{(k)}\right]$$

$$B_{ij}^{(k)} = [(A_1^{(k)})_{ij} \dots (A_p^{(k)})_{ij}]$$

$$C_{ij} = [B_{ij}^{(1)} \dots\ B_{ij}^{(K)}]$$

Regularization

$$\sum_{k=1}^{K}\sum_{i\neq j}\left\|C_{ij}\right\|_2$$

Formulation C

$$\sum_{k=1}^{K}\sum_{i\neq j}\left\|B_{ij}^{(k)}\right\|_2$$

$$\sum_{k=1}^{K}\sum_{i\neq j}\left\|C_{ij}\right\|_2$$

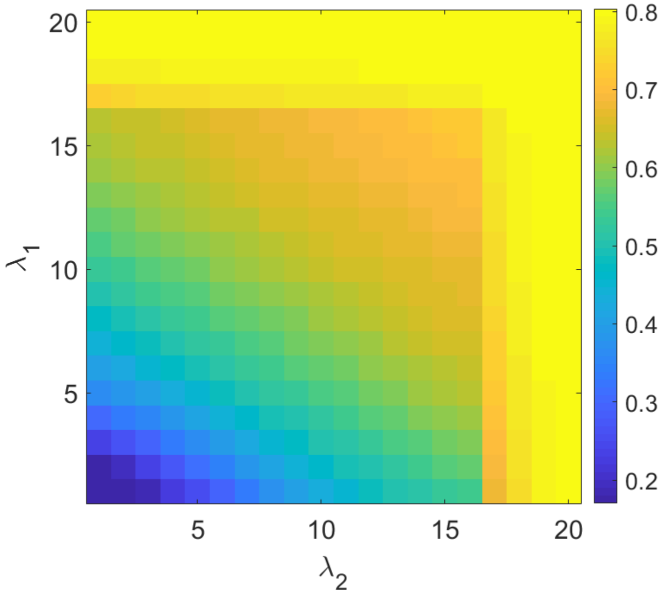
Formulation D

$$\sum_{k=1}^{K}\sum_{i\neq j}\left\|B_{ij}^{(k)}\right\|_2$$

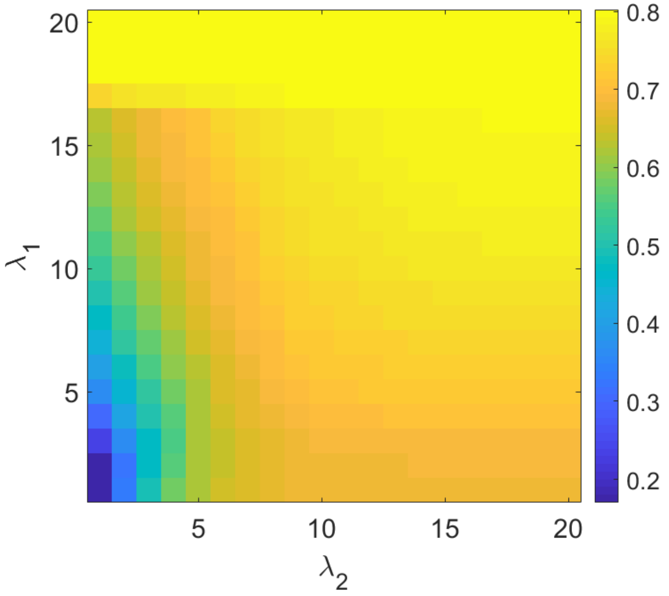
$$\sum_{k<k'}\sum_{i\neq j}\left\|B_{ij}^{k} - B_{ij}^{k'}\right\|_2$$

Formulation S

# SUPPLEMENTARY: BIAS HEAT MAP




Parameter bias of formulation D




Parameter bias of formulation S

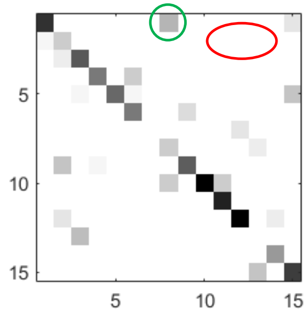
# RESULTS

## COMPARISON BETWEEN $df$.

| # Nonzero | | Formulation S | | | |
|---|---|---|---|---|---|
| | | TPR | FPR | ACC | bias |
| Experiment 1 | density = 0.1 | 0.987 | 0.206 | 0.969 | 0.112 |
| | density = 0.3 | 0.966 | 0.193 | 0.925 | 0.156 |
| Experiment 2 | density = 0.1 | 0.945 | 0.178 | 0.922 | 0.128 |
| | density = 0.3 | 0.858 | 0.169 | 0.846 | 0.165 |
| Experiment 3 | density = 0.1 | 0.973 | 0.195 | 0.945 | 0.121 |
| | density = 0.3 | 0.926 | 0.107 | 0.917 | 0.130 |

| # Nonzero and similar | | Formulation S | | | |
|---|---|---|---|---|---|
| | | TPR | FPR | ACC | bias |
| Experiment 1 | density = 0.1 | 0.995 | 0.303 | 0.967 | 0.129 |
| | density = 0.3 | 0.974 | 0.225 | 0.923 | 0.176 |
| Experiment 2 | density = 0.1 | 0.968 | 0.268 | 0.924 | 0.150 |
| | density = 0.3 | 0.889 | 0.180 | 0.858 | 0.165 |
| Experiment 3 | density = 0.1 | 0.972 | 0.197 | 0.944 | 0.125 |
| | density = 0.3 | 0.919 | 0.135 | 0.904 | 0.145 |

$k = 1$




$k = 2$