

Magno  
Severino

---

**Projeto de Pesquisa**

---

# Inference and Model Selection for Continuous and Infinite Markov Random Fields on Graphs

Área: 1.a. Inferência para Processos Estocásticos.

Área correlata: 3.a. Aprendizagem Estatística e Ciência de Dados.

# Agenda

- Introduction
- Theoretical Background
- Research Proposal
- Viability

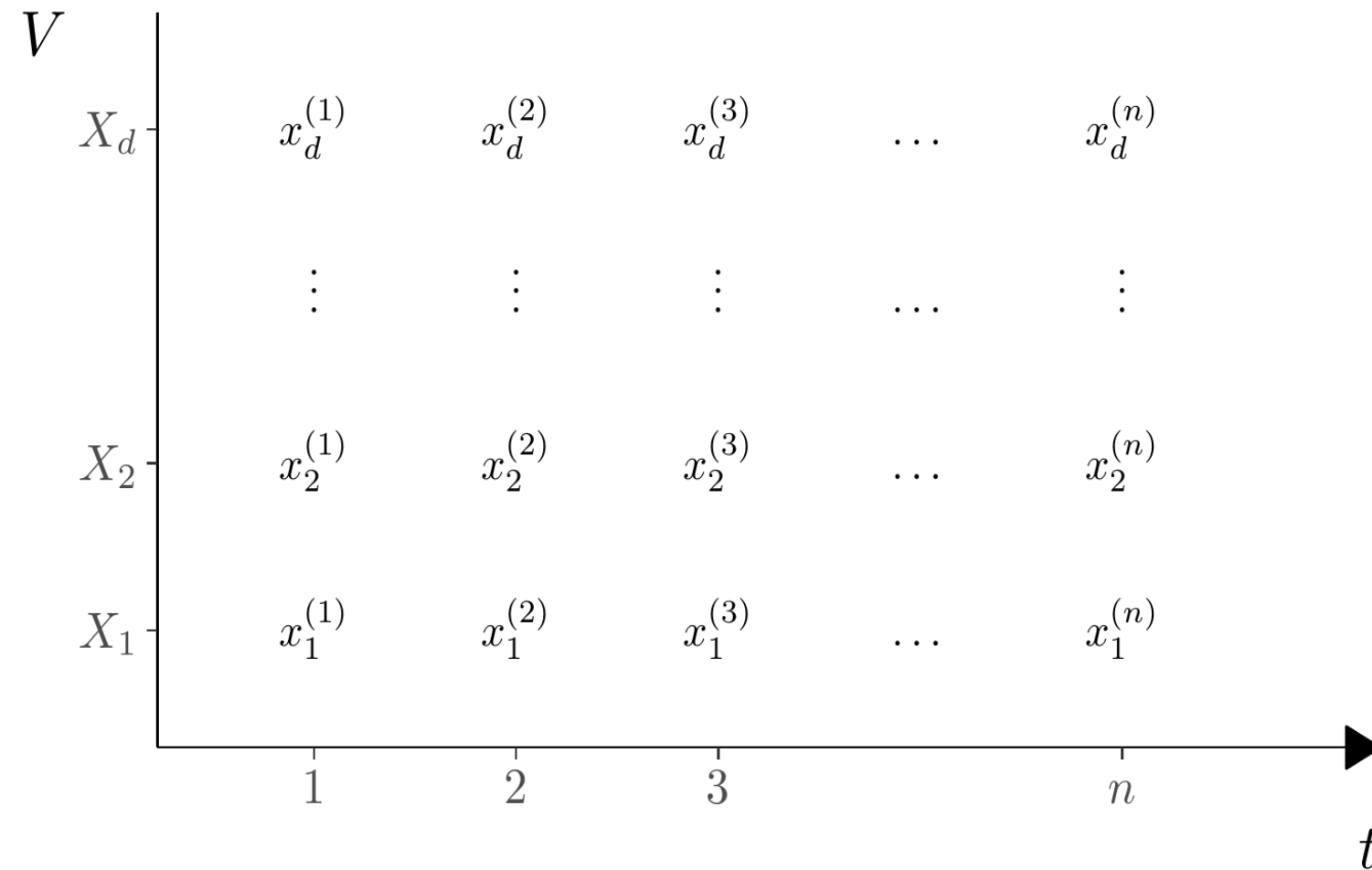
# Introduction

- Previous research (Severino, 2025):
  - Focus on **discrete** vector-valued stochastic processes,
  - Model selection criteria for **graphs** under **mixing conditions**,
  - Graphs with **fixed** number of nodes.
- Research proposals:
  - **Proposal 1:** Extending models to **countably infinite vertex sets**,
  - **Proposal 2:** Adapting the methodology for **continuous data**.

# Background and Definitions

# Vector-Valued Stochastic Processes

- $\mathbf{X}^{(\mathbf{i})} = (\mathbf{X}_1^{(\mathbf{i})}, \mathbf{X}_2^{(\mathbf{i})}, \dots, \mathbf{X}_d^{(\mathbf{i})}) \in \mathbf{A}^d$ , where  $A$  a finite alphabet.
- Invariant distribution  $\pi$ , probability space  $((A^d)^\mathbb{N}, \mathcal{F}, \mathbb{P})$
- Process  $\mathbf{X} = \{\mathbf{X}^{(\mathbf{i})} : \mathbf{i} \in \mathbb{N}\}$  is stationary and ergodic.
- We assume the process  $\mathbf{X}$  has an underlying graph  $G^* = (V, E^*)$  encoding the conditional dependencies in  $\pi$ .

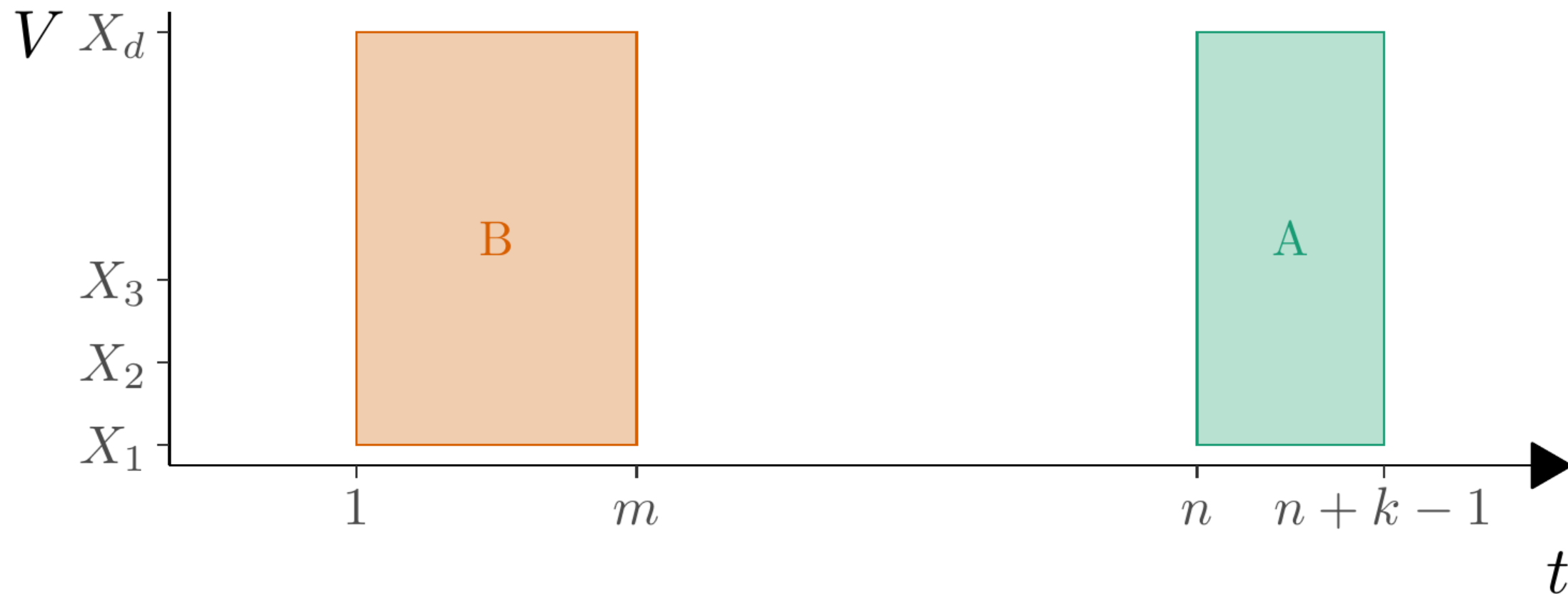


# Mixing Condition

- $X^{(i:j)}$  denote the sequence of vectors  $X^{(i)}, X^{(i+1)}, \dots, X^{(j)}$ .
- $\mathbf{X} = \{\mathbf{X}^{(\mathbf{i})} : -\infty < \mathbf{i} < \infty\}$  satisfies a *mixing condition* with rate  $\{\psi(\ell)\}_{\ell \in \mathbb{R}}$  if

$$\left| \mathbb{P}\left(X^{(n:(n+k-1))} = x^{(1:k)} \mid X^{(1:m)} = x^{(1:m)}\right) - \mathbb{P}\left(X^{(n:(n+k-1))} = x^{(1:k)}\right) \right| \leq \psi(n-m) \mathbb{P}\left(X^{(n:(n+k-1))} = x^{(1:k)}\right),$$

for  $n \geq m + \ell$  and for each  $k, m \in \mathbb{N}$  and each  $x^{(1:k)} \in (A^d)^k, x^{(1:m)} \in (A^d)^m$  with  $\mathbb{P}(X^{(1:m)} = x^{(1:m)}) > 0$ .



# Empirical Probabilities

Given a process  $\mathbf{X} \in \{a, b\}^3$  and the sample of size 5 below. Then

$X_1$	$X_2$	$X_3$
$b$	$a$	$a$
$a$	$b$	$b$
$b$	$b$	$a$
$a$	$a$	$a$
$b$	$a$	$a$

$$\hat{\pi}(\{X_1 = a\}) = \frac{2}{5}, \quad \hat{\pi}(\{X_1 = a, X_3 = a\}) = \frac{1}{5},$$

$$\hat{\pi}(\{X_1 = b\} | \{X_2 = a, X_3 = a\}) = \frac{2}{3}.$$

$$\text{as } \hat{\pi}(\{X_2 = a, X_3 = a\}) > 0.$$

Formally, assume we observe a sample of size  $n$  of the process, denoted by  $\{x^{(i)} : i = 1, \dots, n\}$ . Then, for any  $W \subset V$  and any  $a_W \in A^W$ ,

$$\hat{\pi}(a_W) = \frac{N(a_W)}{n}; \quad \hat{\pi}(a_{W'} | a_W) = \frac{\hat{\pi}(a_{W' \cup W})}{\hat{\pi}(a_W)},$$

for  $\hat{\pi}(a_W) > 0$ , two disjoint subsets  $W, W' \subset V$ , and configurations  $a_W \in A^W, a_{W'} \in A^{W'}$ .



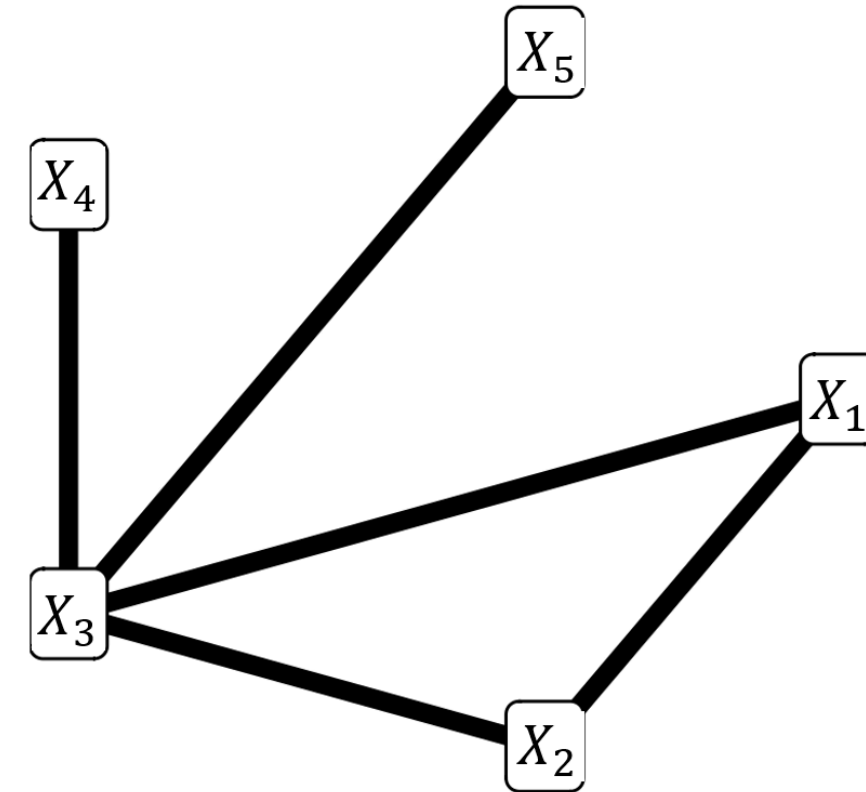
# Empirical Probabilities

Given a graph  $G = (V, E)$  and  $v \in V$ , define

$$G(v) = \{u \in V : (u, v) \in E\},$$

the set of neighbors of  $v$  in graph  $G$ .

For  $v = X_1$ , then  $G(v) = \{X_2, X_3\}$ .



Then

$$\hat{\pi}(a_v | a_{G(v)}) = \frac{\hat{\pi}(a_{\{v\} \cup G(v)})}{\hat{\pi}(a_{G(v)})}.$$

# Graph Estimator

Given any graph  $G = (V, E)$  and a sample of the process, we define the log-pseudo-likelihood function by

$$\log L(G) = \sum_{v \in V} \sum_{(a_v \in A)} \sum_{a_{G(v)} \in A^{|G(v)|}} N(a_v, a_{G(v)}) \log \pi(a_v | a_{G(v)}).$$

As the conditional probabilities  $\pi$  are not known, we can estimate them from the data

$$\log \hat{L}(G) = \sum_{v \in V} \sum_{(a_v \in A)} \sum_{a_{G(v)} \in A^{|G(v)|}} N(a_v, a_{G(v)}) \log \hat{\pi}(a_v | a_{G(v)}).$$

# Graph Estimator Consistency


**Theorem (Severino, 2025, Stochastic Processes and their Applications):**

Let  $\{X^{(i)} : i \in \mathbb{N}\}$  be a stationary process that satisfies the mixing condition presented before with rate  $\psi(\ell) = O(1/\ell^{1+\epsilon})$  for some  $\epsilon > 0$ .

Then, by taking  $\lambda_n = c \log n$ , for  $c > 0$ , we have that

$$\hat{G} = \arg \max_G \left\{ \log \hat{L}(G) - \lambda_n \sum_{v \in V} |A|^{G(v)} \right\}$$

satisfies  $\hat{G} = G^*$  eventually almost surely as  $n \rightarrow \infty$ .

 **Proposal 1: model selection  
for Markov random fields with  
countable infinite set of vertices  
on graphs under a mixing  
condition**

# Proposal 1

- **Extending model selection in Markov Random Fields (MRFs)**
  - Focus on graphs with **countably infinite** vertices.
  - Applications in **neural networks** and **social networks**.
- **Motivation and existing methods**
  - **Leonardi et al. (2023)**: Penalized pseudo-likelihood for discrete MRFs.
  - Graph estimated based on local neighborhood estimation.
  - **Severino (2025)**: Developed theoretical results for global estimation of discrete MRFs over finite graphs.
- **Research goal**
  - Generalize the results from **finite** to **countably infinite graphs**.
  - Improve global estimation, possibly reducing errors from local neighborhood estimation.

# Proposal 1: Estimation Approach and Applications

- **Proposed estimation framework**
  - Let  $V$  be infinite and  $V_n, n \in \mathbb{N}$  be a sequence of finite subsets of  $V$ .
  - Assume  $V_n \uparrow V$  as  $n \rightarrow \infty$ .
  - Sample:  $\{\mathbf{X} = \{X_v : v \in V_n\}\}$ , assuming that  $\deg(v)$  is finite.
  - Adaptation of key theorems to handle countably infinite vertex sets.
- **Algorithm development and evaluation**
  - Implement graph estimator in **R package**.
  - Performance assessed through **extensive simulation studies**.
- **Real-world applications**
  - **Social interaction networks**: Capturing dependencies in large-scale social systems.
  - **Online social networks**: Understanding **information diffusion & social influence**.

# Proposal 2: Model selection for continuous Markov random fields on graphs under a mixing condition

# Proposal 2

- **Extending previous work on MRFs**
  - **Severino (2025):** Applied MRF model to **São Francisco River** water flow data.
  - Discretization was required, introducing potential limitations.
  - **Goal: Generalize the approach to continuous stochastic processes.**
- **Theoretical adaptations needed**
  - Modify key results from discrete to **continuous random variables**.
  - Replace summations with integrals in **penalized pseudo-likelihood function**.
  - Adapt **consistency and convergence proofs** for continuous measurements.
- **Key benefits**
  - Expands applicability to **environmental monitoring, finance, and signal processing**.
  - Enables more accurate inference from **continuous data sources**.



# Proposal 2: Implementation & Applications

- **Algorithm development and valuation**
  - Implement adapted algorithms in **R package**.
  - Assess performance through **extensive simulation studies**.
  - Validate robustness and accuracy using real-world datasets.
- **Broader applications**
  - **Bioinformatics**: Model gene regulatory networks and protein interactions.
  - **Economics**: Capture dependencies in financial markets and economic indicators.
  - Expands to **various fields** beyond traditional MRF applications.

## **Future Research Direction**

- **Combine Proposal 1 (countable infinite graphs) and Proposal 2 (continuous data).**
- **Develop a unified method for estimating MRFs with continuous variables and countable infinite vertices.**
- **Provide a versatile tool for modeling complex stochastic processes.**

# Viability

# ☰ Relevance to Data Science and Statistical Learning

- **Growing Importance of Data Science**
  - Essential for innovation and decision-making across industries.
  - Need for advanced statistical methodologies to handle complex datasets.
- **Unsupervised Learning & Graphical Models**
  - Graphical model estimation enables the discovery of hidden structures in data.
  - Adaptations of methods for **continuous and dependent data processes** and **large-scale networks**.

# 🎯 Expected Outcomes

- **Theoretical Contributions**

- Extension of Markov Random Fields (MRFs) theory under mixing conditions.
- Application to graphs with countably infinite vertices and continuous data.
- Submission to leading international journals.

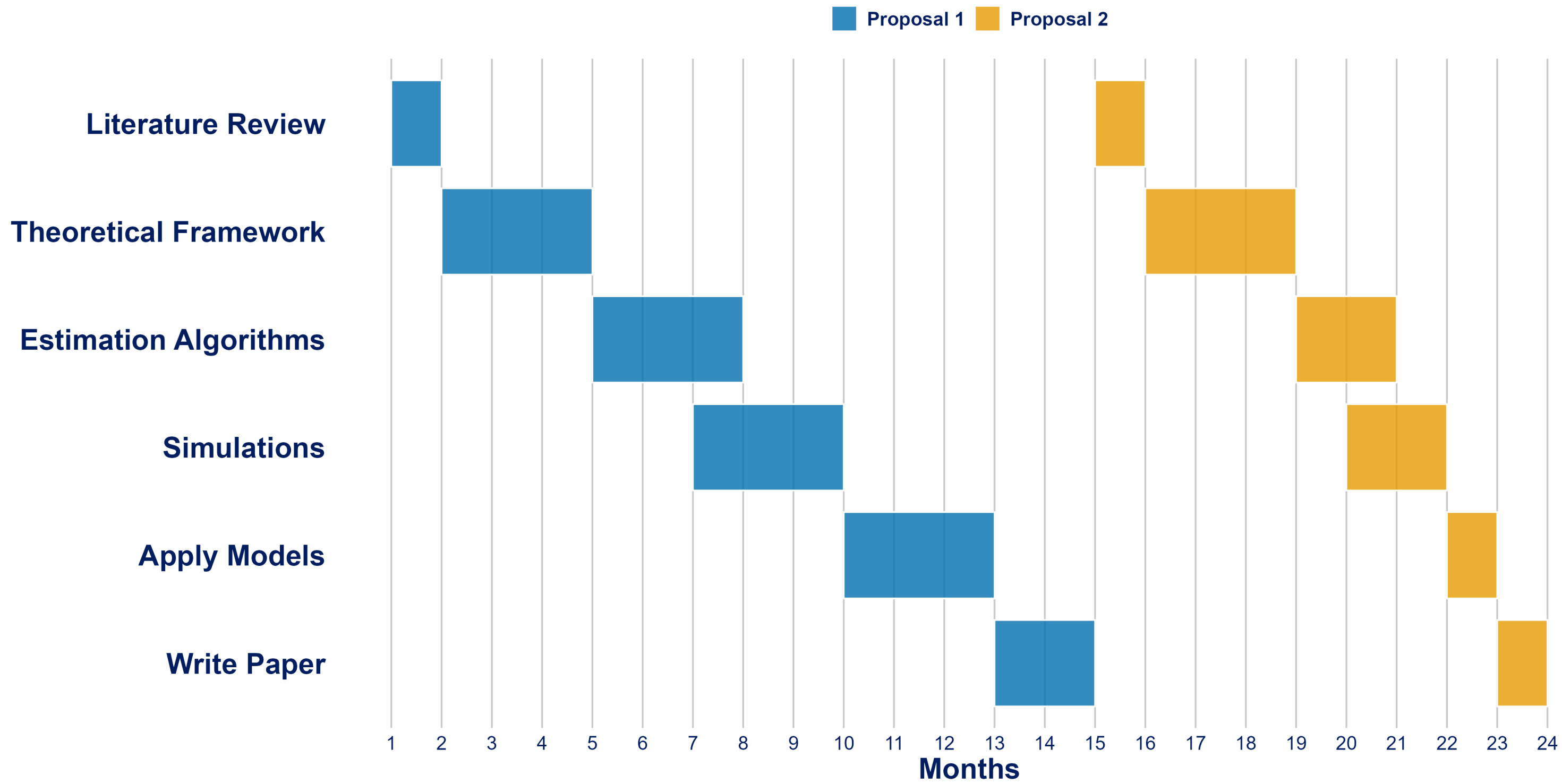
- **Algorithm Development**

- Robust, scalable algorithms for complex graphical models.
- R packages with open-source access.
- Comprehensive documentation with practical examples for ease of adoption.

- **Broader Scientific Impact**

- Increased applicability in domains like **neuromathematics, finance, and ecology**.
- Potential for interdisciplinary collaborations (UFRJ, UFRN, UBA, Neuromat researches).

# Timeline



# \$ Budget

- Minimal financial costs.
- Theoretical development **does not require** any additional resources.
- Simulation and real data application of the developed algorithms will be conducted using existing computer server **resources at the department.**
- Anticipated expenses: presenting the research at international scientific conferences.

# References

- **Severino, M. T. F., & Leonardi, F. (2025).** *Model selection for Markov random fields on graphs under a mixing condition.* Stochastic Processes and their Applications.
- Leonardi, F., Lopez-Rosenfeld, M., Rodriguez, D., **Severino, M. T. F., & Sued, M. (2021).** *Independent block identification in multivariate time series.* Journal of Time Series Analysis.
- Leonardi, F., Carvalho, R., & Frondana, I. (2023). *Structure recovery for partially observed discrete Markov random fields on graphs under not necessarily positive distributions.* Scandinavian Journal of Statistics.
- Lauritzen, S. L. (1996). *Graphical models (Vol. 17).* Clarendon Press.
- Oodaira, H., & Yoshihara, K. I. (1971). *The law of the iterated logarithm for stationary processes satisfying mixing conditions.* Kodai Mathematical Seminar Reports.





# Rate of convergence of the empirical probabilities

Based on the Law of the Iterated Logarithm for stationary polynomial mixing processes proved in Oodaira and Yoshihara (1971), we can derive the rate of convergence of the empirical probabilities to the true probabilities of the process.

**Proposition 1 (Typicality):** Assume the process  $\{X^{(i)}: i \in \mathbb{Z}\}$  satisfies the mixing condition with rate  $\psi(\ell) = O(1/\ell^{1+\epsilon})$ , for some  $\epsilon > 0$ . Then, for any  $W \subset V$  and  $\delta > 0$ ,

$$\left| \hat{\pi}(a_W) - \pi(a_W) \right| < \sqrt{\frac{\delta \log n}{n}},$$

eventually almost surely as  $n \rightarrow \infty$ .

**Proposition 2 (Conditional typicality):** Then for any  $\delta > 0$ , any disjoint sets  $W, W' \subset V$  and any  $a_W \in A^W$  and  $a_{W'} \in A^{W'}$  we have that

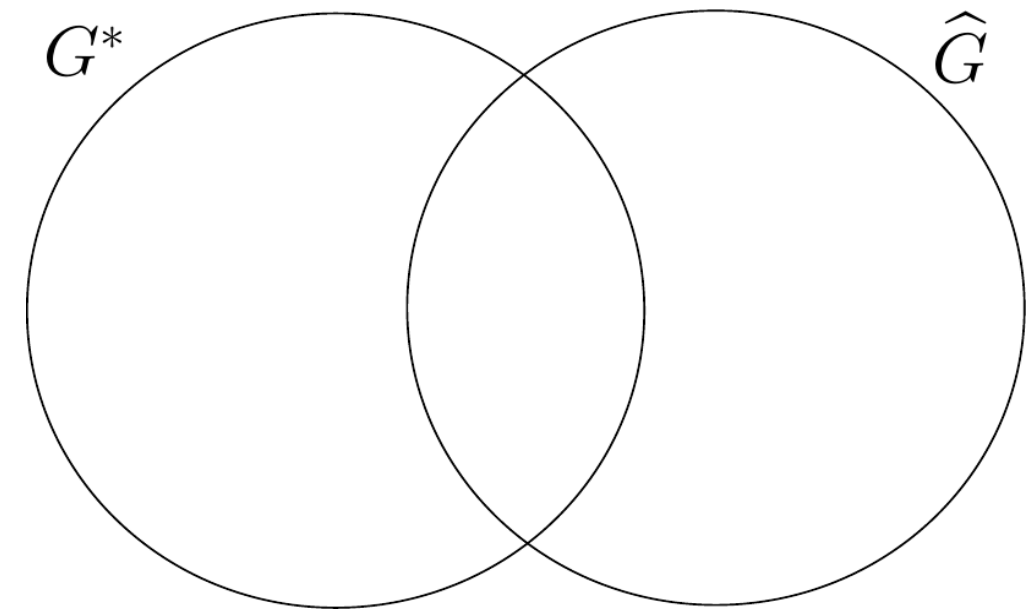
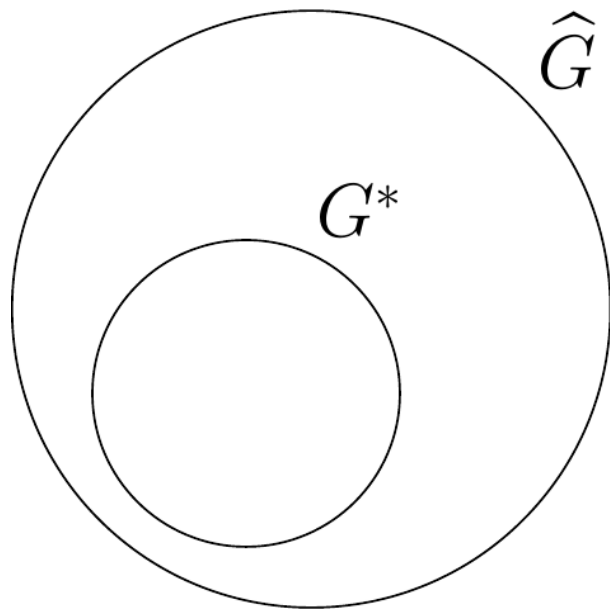
$$\left| \hat{\pi}(a_W | a_{W'}) - \pi(a_W | a_{W'}) \right| < \sqrt{\frac{\delta \log n}{N(a_W)}},$$

eventually almost surely as  $n \rightarrow \infty$ .

# Intuitive Overview

review of Theorem Proof {smaller visibility="uncounted"}

Consider  $\{\hat{G} \neq G^*\} = \{G^* \subsetneq \hat{G}\} \cup \{G^* \not\subset \hat{G}\}$ .



We prove that, eventually almost surely as  $n \rightarrow \infty$ , neither of the cases above can happen, which implies that  $\hat{G} = G^*$ .

# Algorithms for Estimation

Define

$$H(G) = \log \hat{L}(G) - \lambda_n \sum_{v \in V} |A|^{G(v)}.$$

**Focus:** determine the maximal value of  $H(\cdot)$  and identify the argument at which this maximum occurs.

# Exact Algorithm

- Let  $H(G) = \log \hat{L}(G) - \lambda_n \sum_{v \in V} |A|^{G(v)}$ .
- Define

$$\mathcal{G} = \left\{ G = (V, E) : E \subseteq \{V \times V\} \setminus \{(v, v) : v \in V\} \right\}.$$

- Set

$$\hat{G} = \arg \max_{G \in \mathcal{G}} H(G).$$

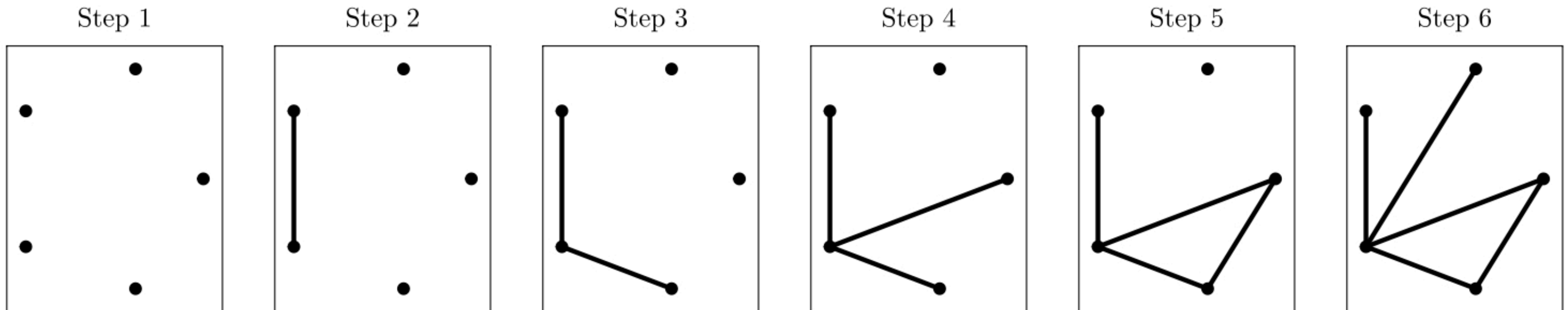
- Drawback: computational complexity

$$|\mathcal{G}| = 2^{d(d-1)/2},$$

$$|V| = d.$$

# Forward Stepwise Algorithm

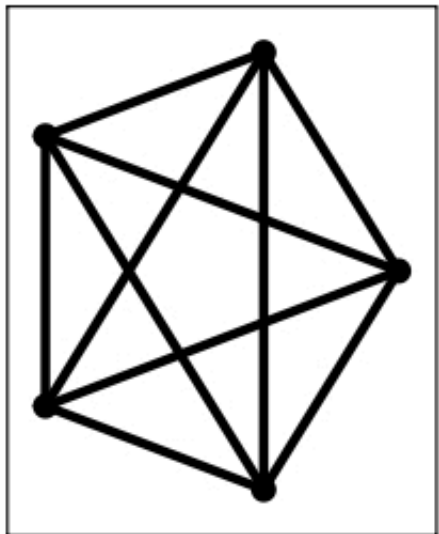
- Starts with an empty graph.
- Adds edges one at a time, selecting the edge that maximizes improvement in fit.
- Stops when no further enhancement in fit is achieved.



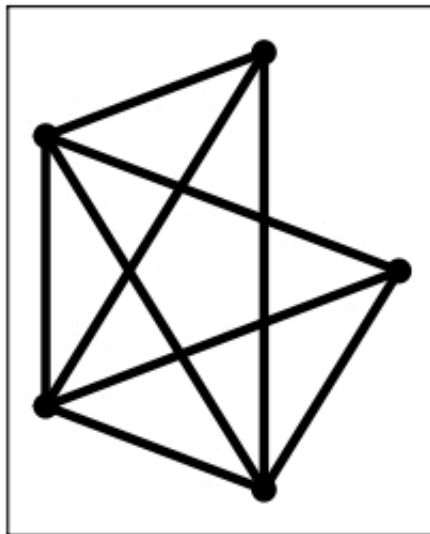
# Backward Stepwise Algorithm

- Begins with a complete graph.
- Removes edges one at a time, selecting the edge that maximizes improvement in fit.
- Stops when no further enhancement in fit is achieved.

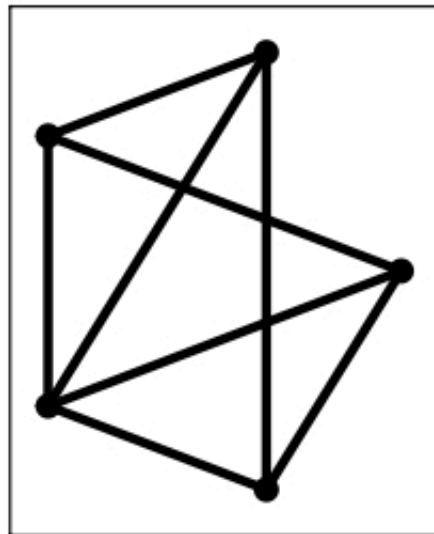
Step 1



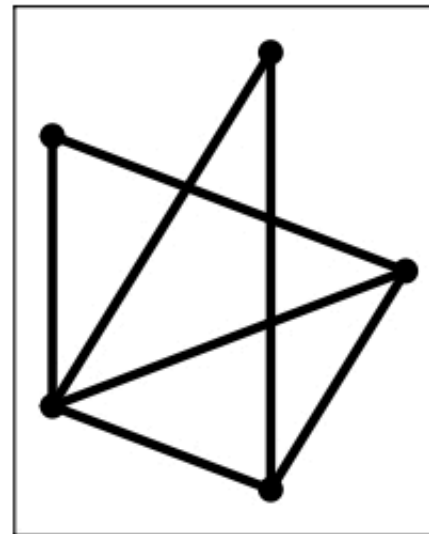
Step 2



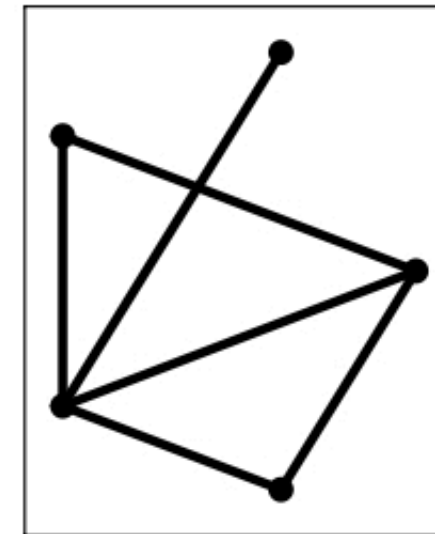
Step 3



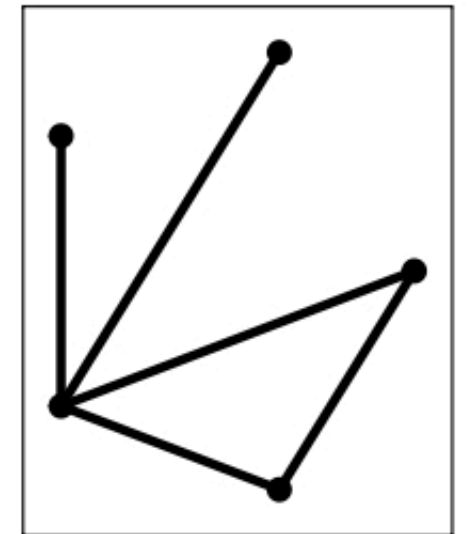
Step 4



Step 5

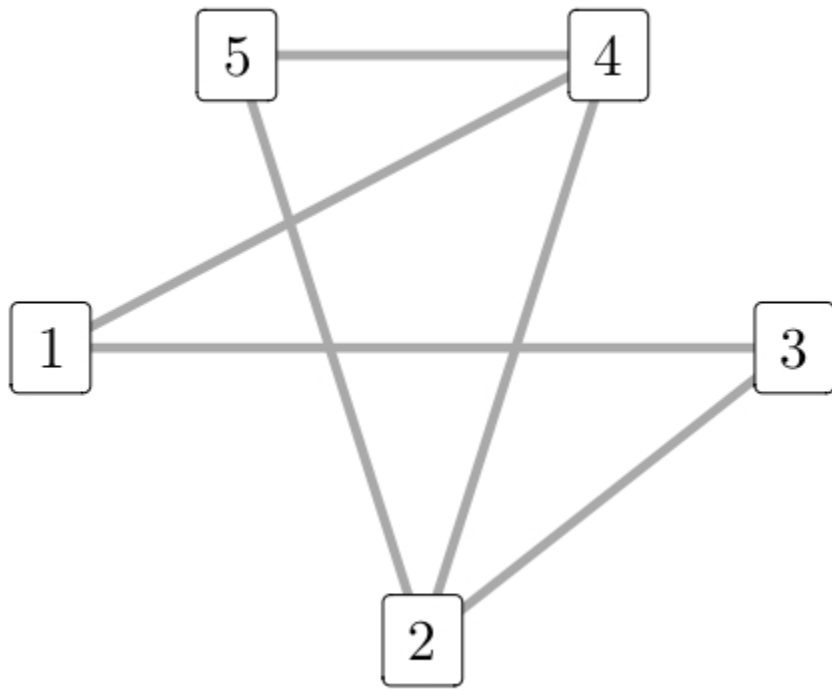


Step 6

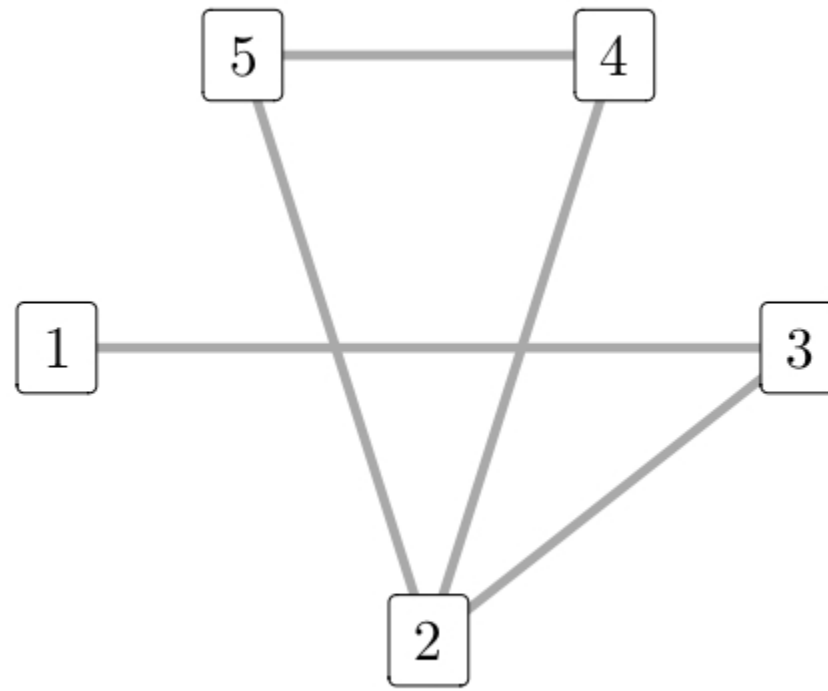


# Simulated Annealing Algorithm

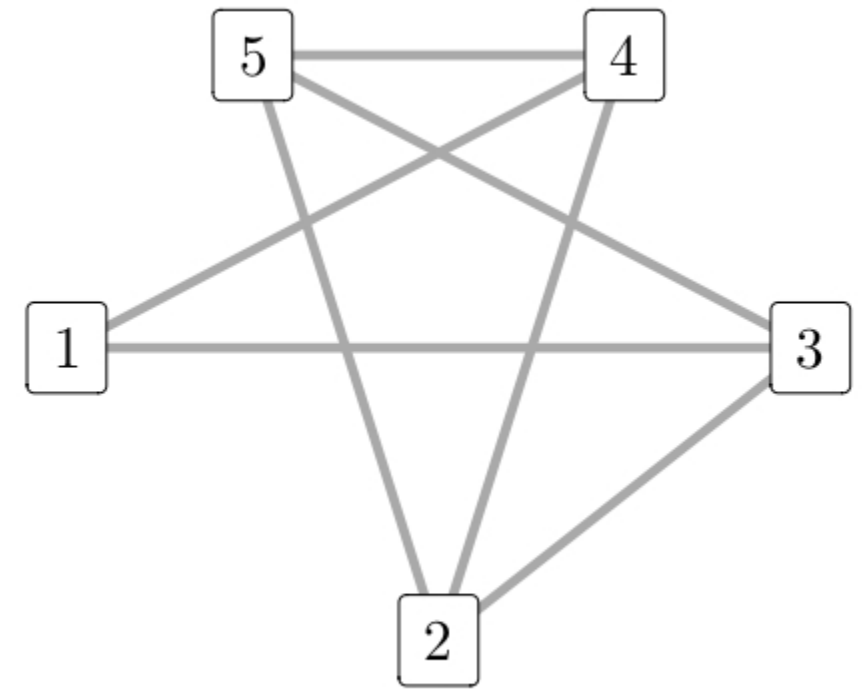
- $H(G) = \log \hat{L}(G) - \lambda_n \sum_{v \in V} |A|^{G(v)}.$
- Definition of a neighbor of a graph.



$G_1$



$G_2$



$G_3$



# Simulation Study

- **Objective:** Assess the performance of the proposed algorithms via simulation studies.
- **Scenario 1: Fixed True Graph**
  - Generate synthetic data based on fixed true graph.
  - Assess estimator's performance under stability conditions.
- **Scenario 2: Random Graphs with Varying Edge Number**
  - Generate random graphs with varying edge numbers.
  - Evaluate estimator's performance across graphs with different edge configuration.

# Example 11 - Scenario 1

- Consider  $\mathbf{X} = (X_1, \dots, X_5)$ .

- Joint probability function of these variables:

$$p(x_1, x_2, x_3, x_4, x_5) = p(x_3)p(x_1|x_3)p(x_2|x_1, x_3)p(x_4|x_3)p(x_5|x_3).$$

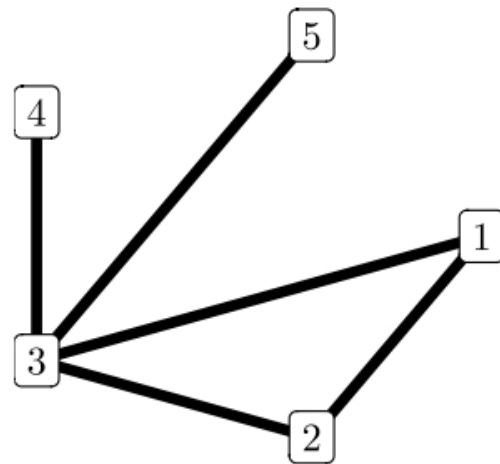
- Conditional probabilities:

$$p(x_1|x_2, x_3, x_4, x_5) = p(x_1|x_2, x_3),$$

$$p(x_2|x_1, x_3, x_4, x_5) = p(x_2|x_1, x_3),$$

$$p(x_4|x_1, x_2, x_3, x_5) = p(x_4|x_3),$$

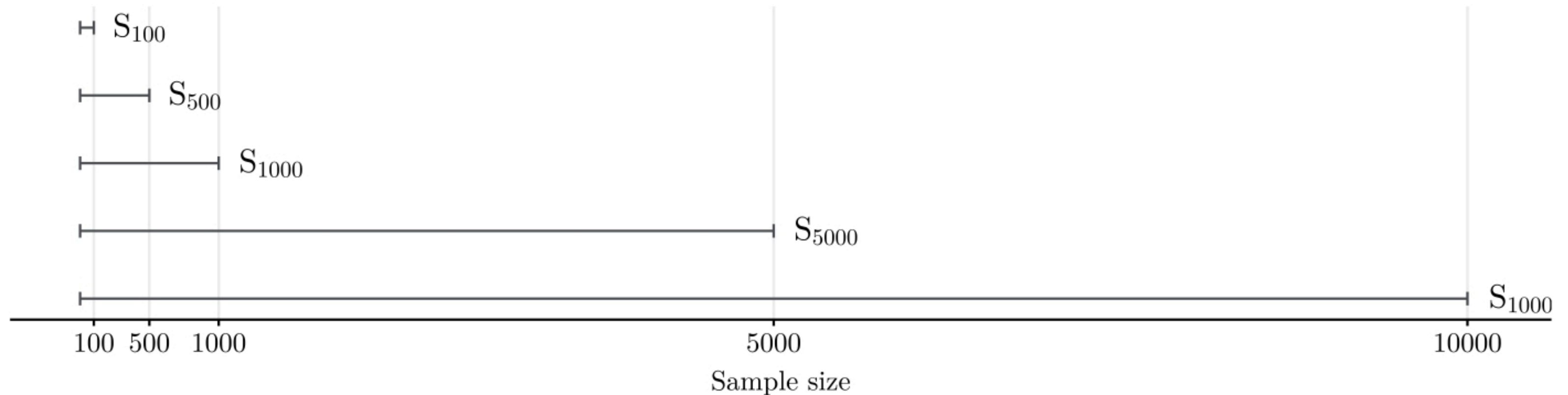
$$p(x_5|x_1, x_2, x_3, x_4) = p(x_5|x_3).$$



- Aim:** assess the performance of the algorithms for estimation.
- Data generation:** Gibbs Sampler algorithm.

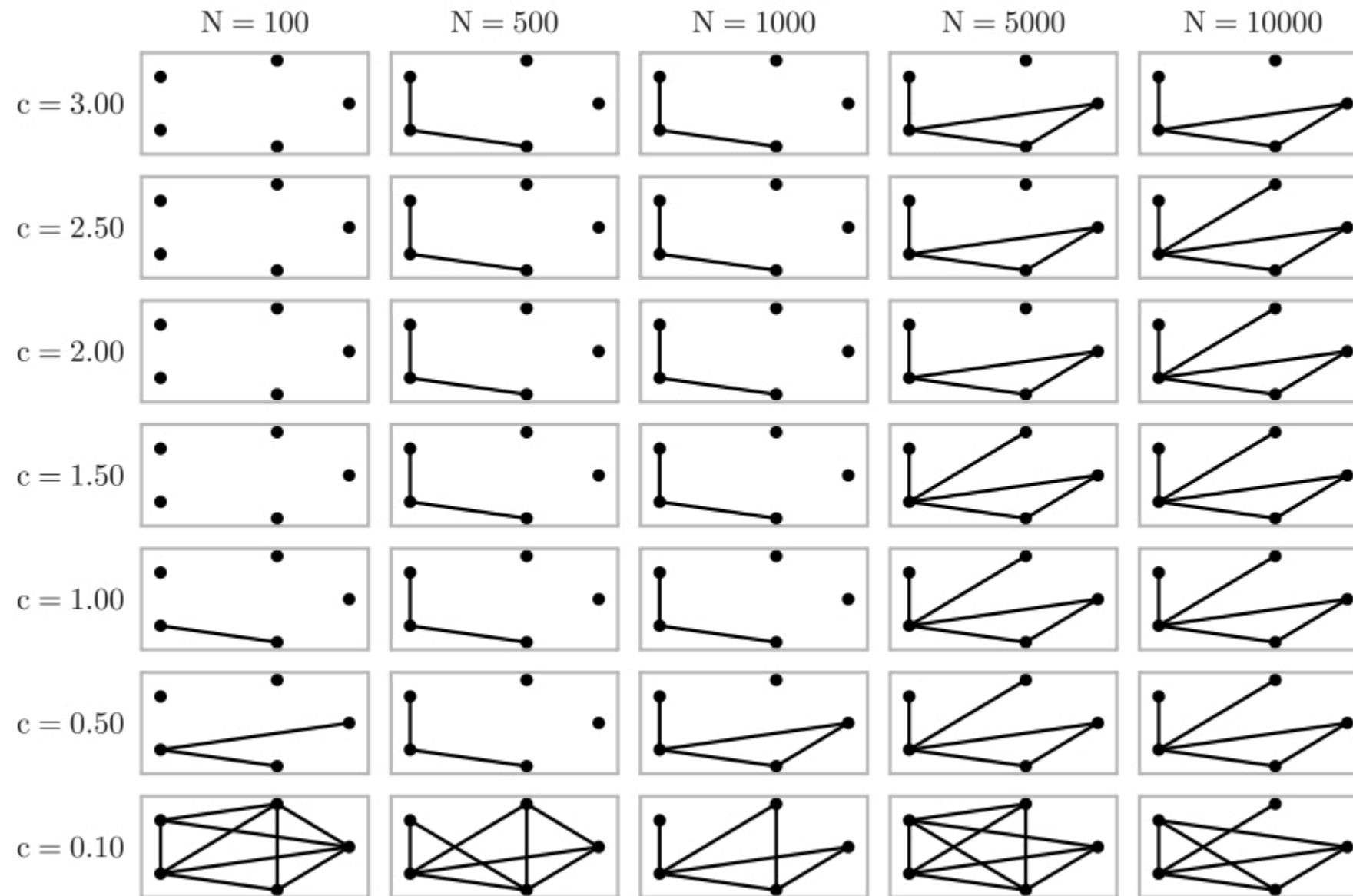
# Example 11 - Sampling Scheme

- The Gibbs sampler (Geman and Geman, 1984 and Gelfand and Smith, 1990):
  - iterations: 15,000,
  - burn-in period: 5,000,
  - final sample: 10,000.
- Smaller samples were extracted from the initial sample  $s_1, \dots, s_{10,000}$ , with sizes  $N \in \{100, 500, 1,000, 5,000, 10,000\}$ .



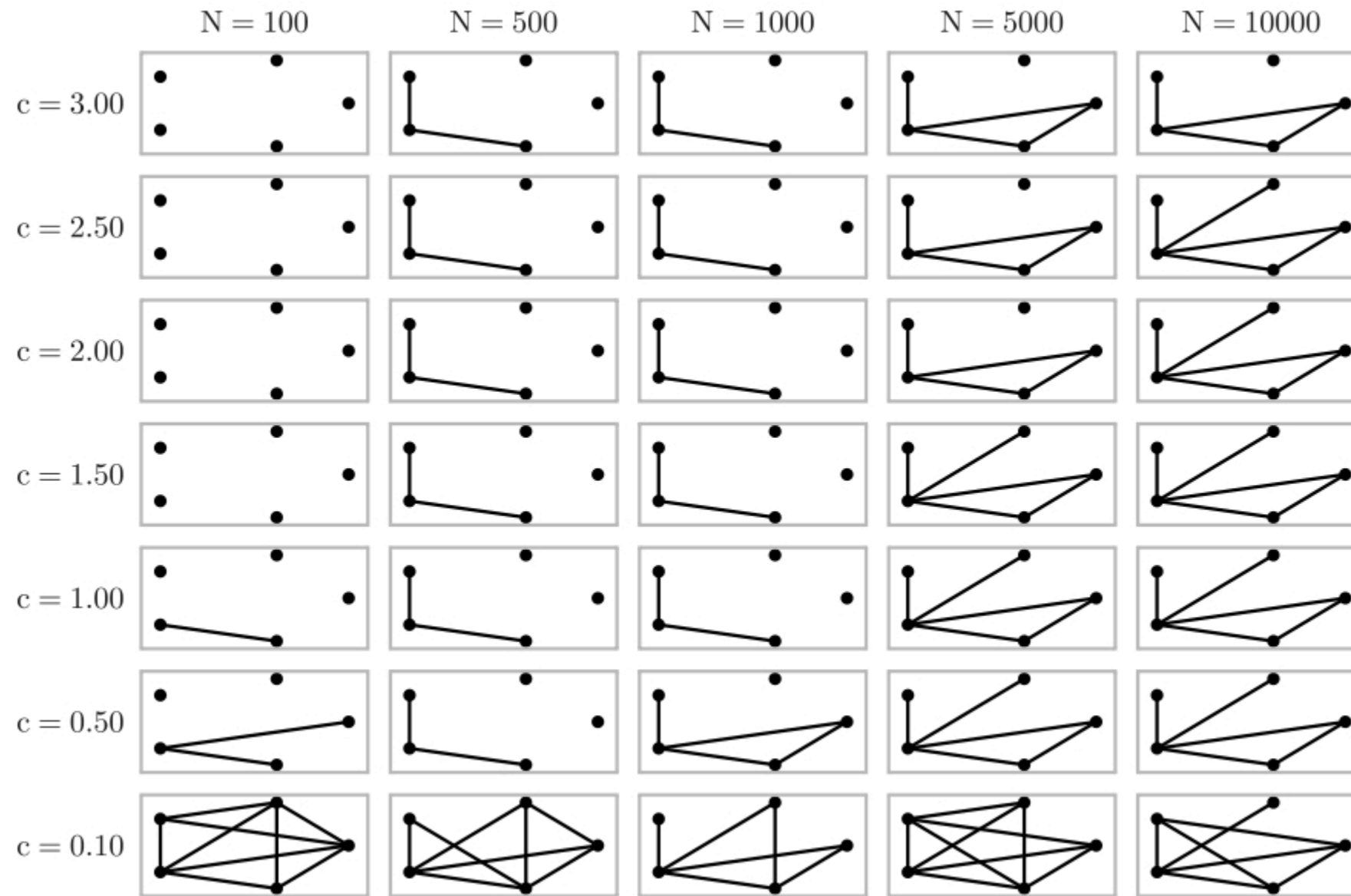
# Example 11 - Exact Algorithm

Effects of sample size and penalizing constant value. Here,  $\lambda_N = c \log N$ .



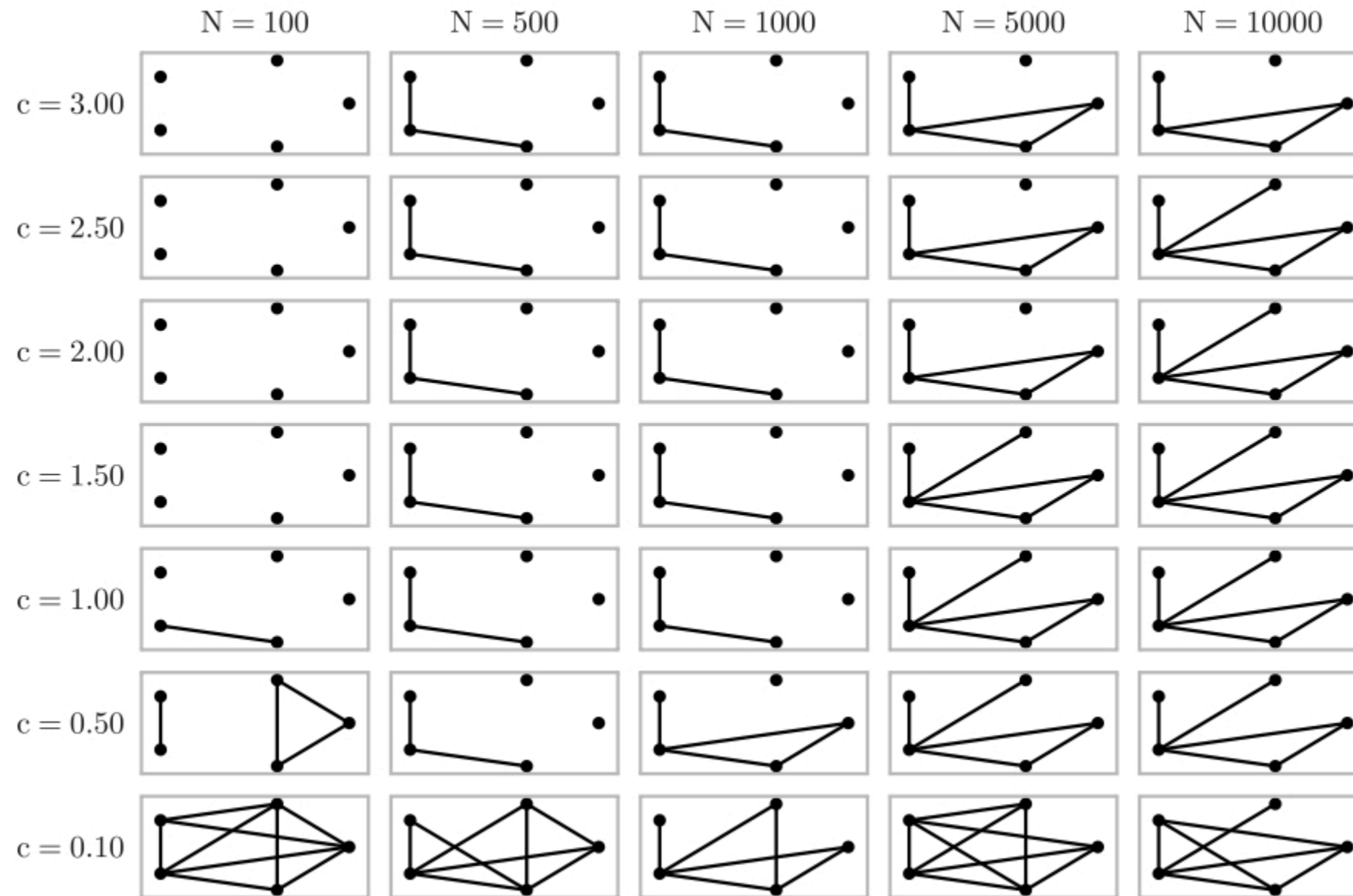
# Example 11 - Forward Stepwise

Effects of sample size and penalizing constant value. Here,  $\lambda_N = c \log N$ .



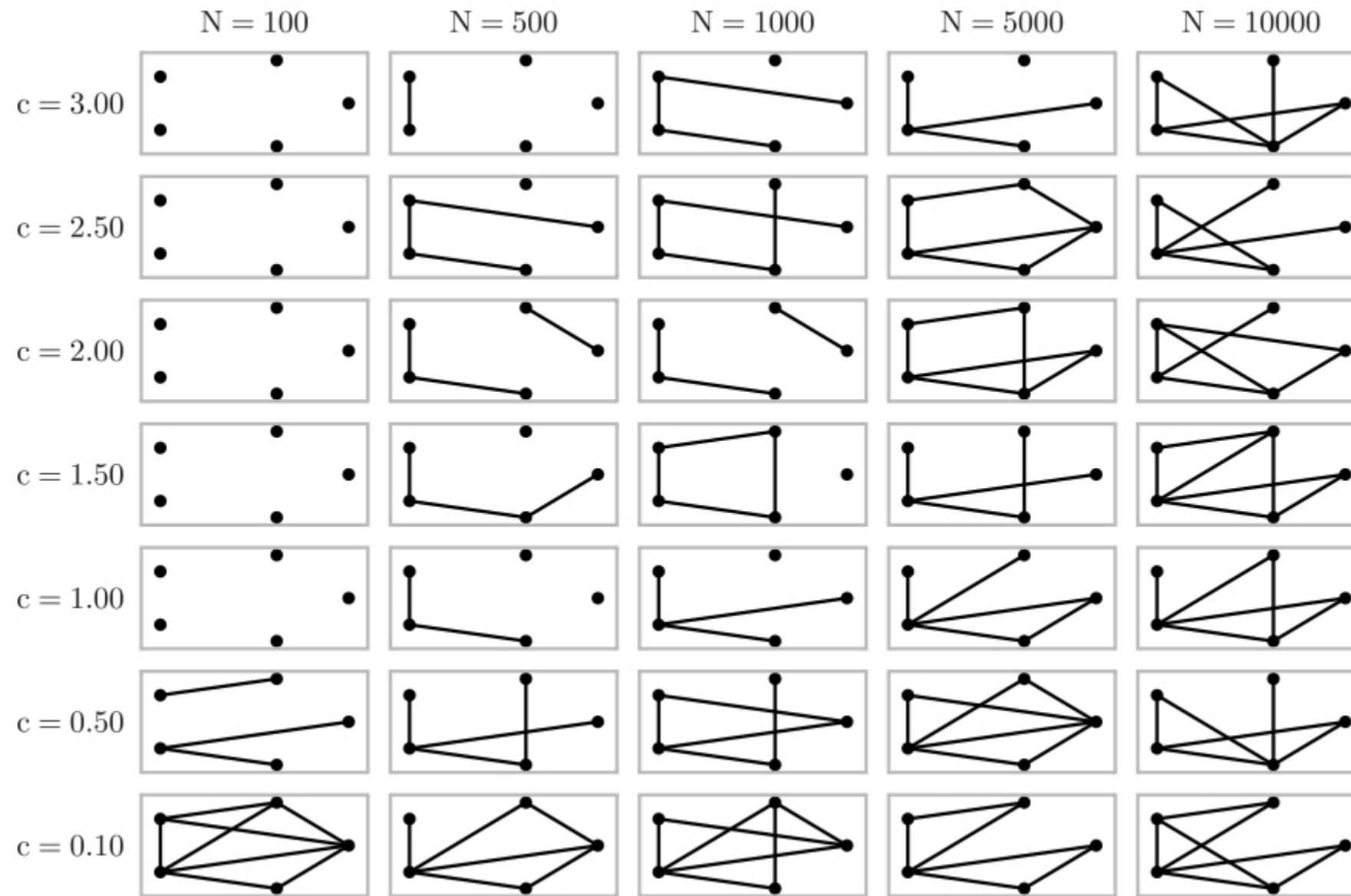
# Example 11 - Backward Stepwise

Effects of sample size and penalizing constant value. Here,  $\lambda_N = c \log N$ .



# Example 11 - Simulated Annealing

Effects of sample size and penalizing constant value. Here,  $\lambda_N = c \log N$ .



# Example 11 - Several Replications

- **Objective:** generate several samples and assess the performance of the algorithms.
- **Metrics:** underestimation error ( $ue$ ), overestimation error ( $oe$ ), and total error ( $te$ ).

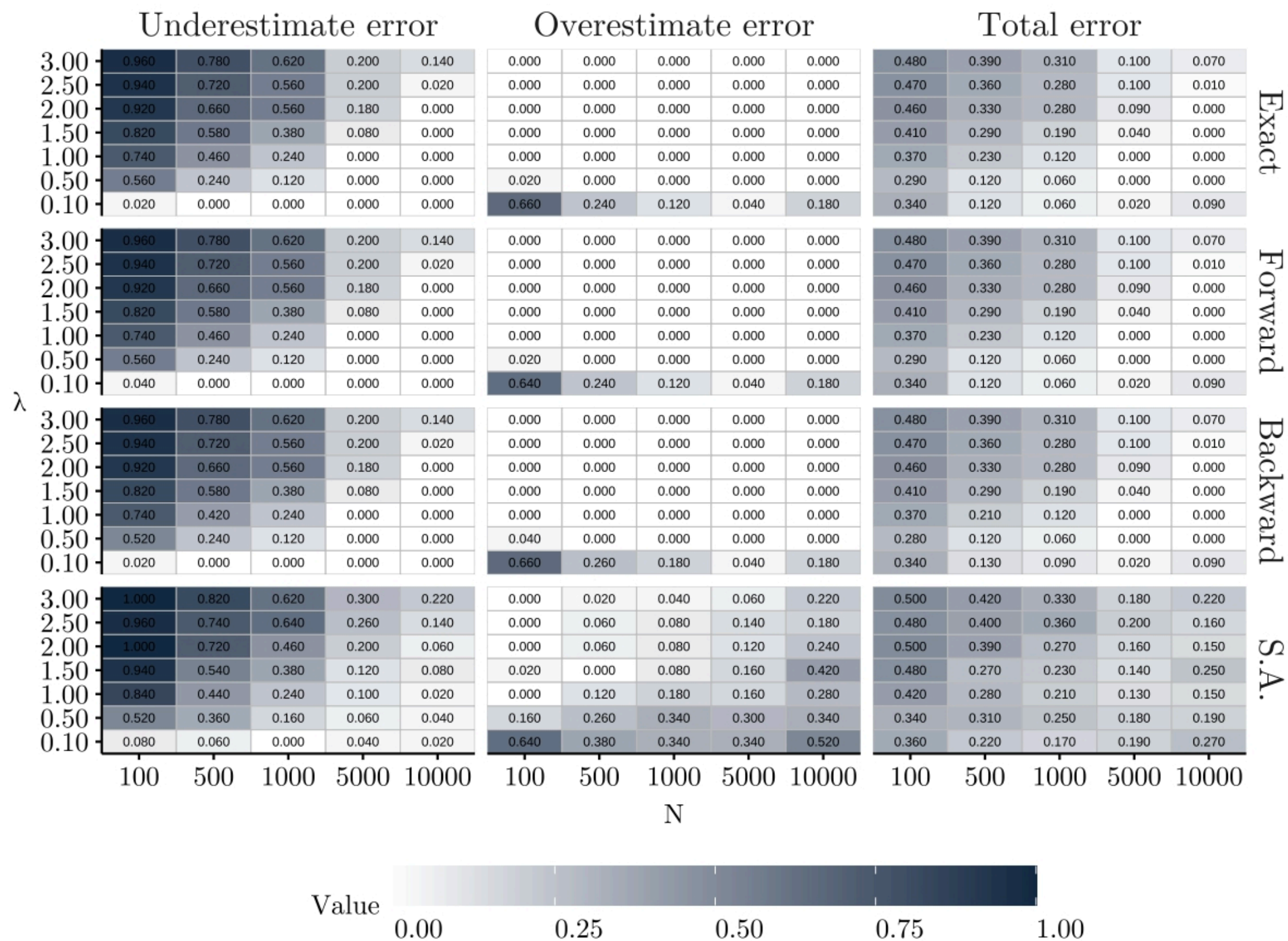
$$ue(G, \hat{G}) = \frac{\sum_{(v,w)} \mathbf{1}\{(v,w) \in E \text{ and } (v,w) \notin \hat{E}\}}{\sum_{(v,w)} \mathbf{1}\{(v,w) \in E\}},$$

$$oe(G, \hat{G}) = \frac{\sum_{(v,w)} \mathbf{1}\{(v,w) \notin E \text{ and } (v,w) \in \hat{E}\}}{\sum_{(v,w)} \mathbf{1}\{(v,w) \notin E\}},$$

$$te(G, \hat{G}) = \frac{ue \sum_{(v,w)} \mathbf{1}\{(v,w) \notin E\} + oe \sum_{(v,w)} \mathbf{1}\{(v,w) \in E\}}{|V|(|V|-1)/2}.$$



# Example 11 - Several Replications



# The Choice of Penalizing Constant $c$

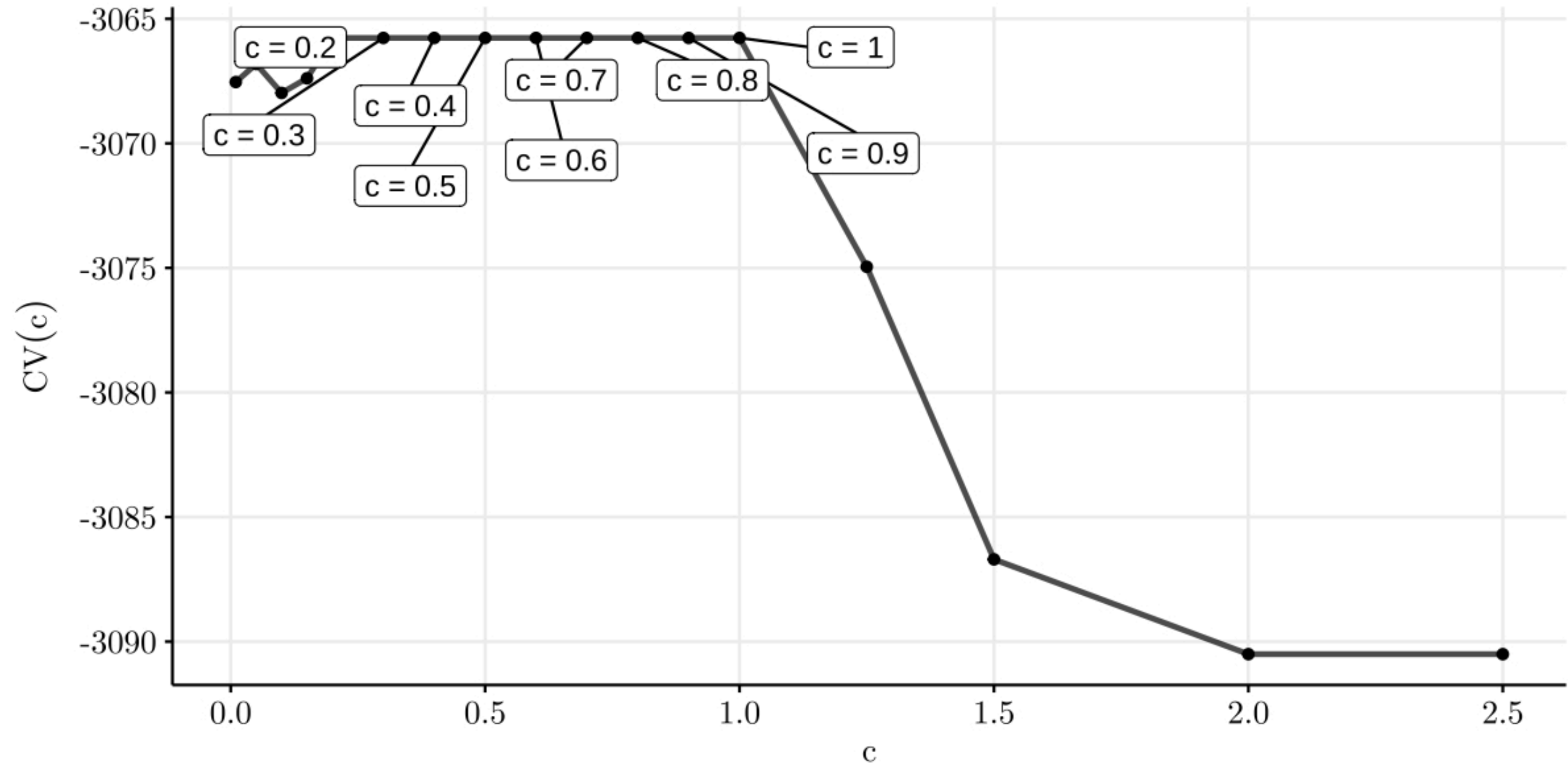
Utilize  $k$ -fold cross-validation to assess model performance and select optimal penalizing constant values.

Iteration 1	Val.	Train	Train	Train	Train
Iteration 2	Train	Val.	Train	Train	Train
Iteration 3	Train	Train	Val.	Train	Train
Iteration 4	Train	Train	Train	Val.	Train
Iteration 5	Train	Train	Train	Train	Val.

Compute pseudo-log-likelihood over validation sets for different constant values.

$$\text{CV}_k^{(i)}(c) = \sum_{v \in V} \sum_{(a_v \in A)} \sum_{a_{\hat{G}(v)} \in A^{|\hat{G}(v)|}} N_i(a_v, a_{\hat{G}(v)}) \log \hat{\pi}(a_v | a_{\hat{G}(v)}).$$

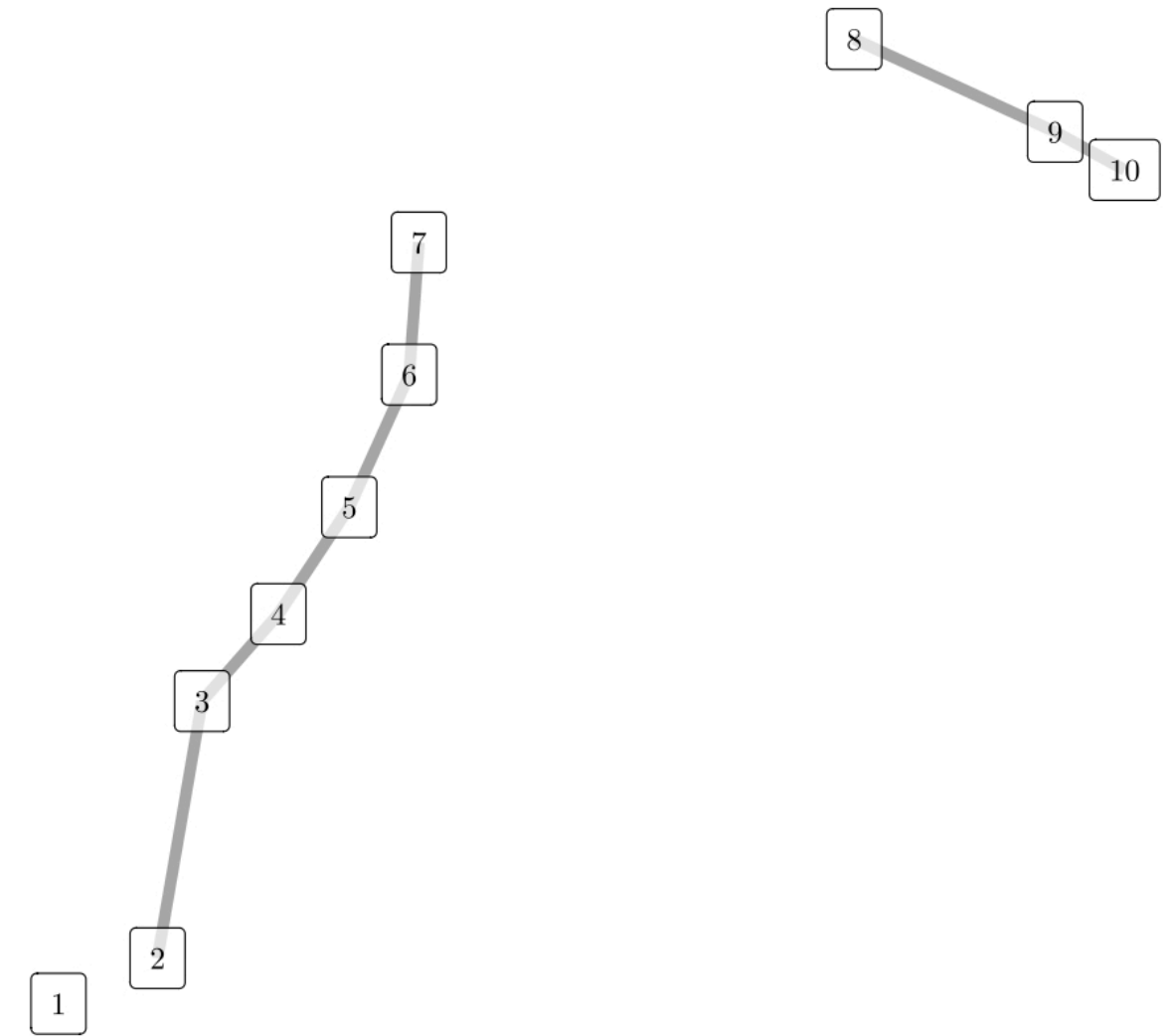
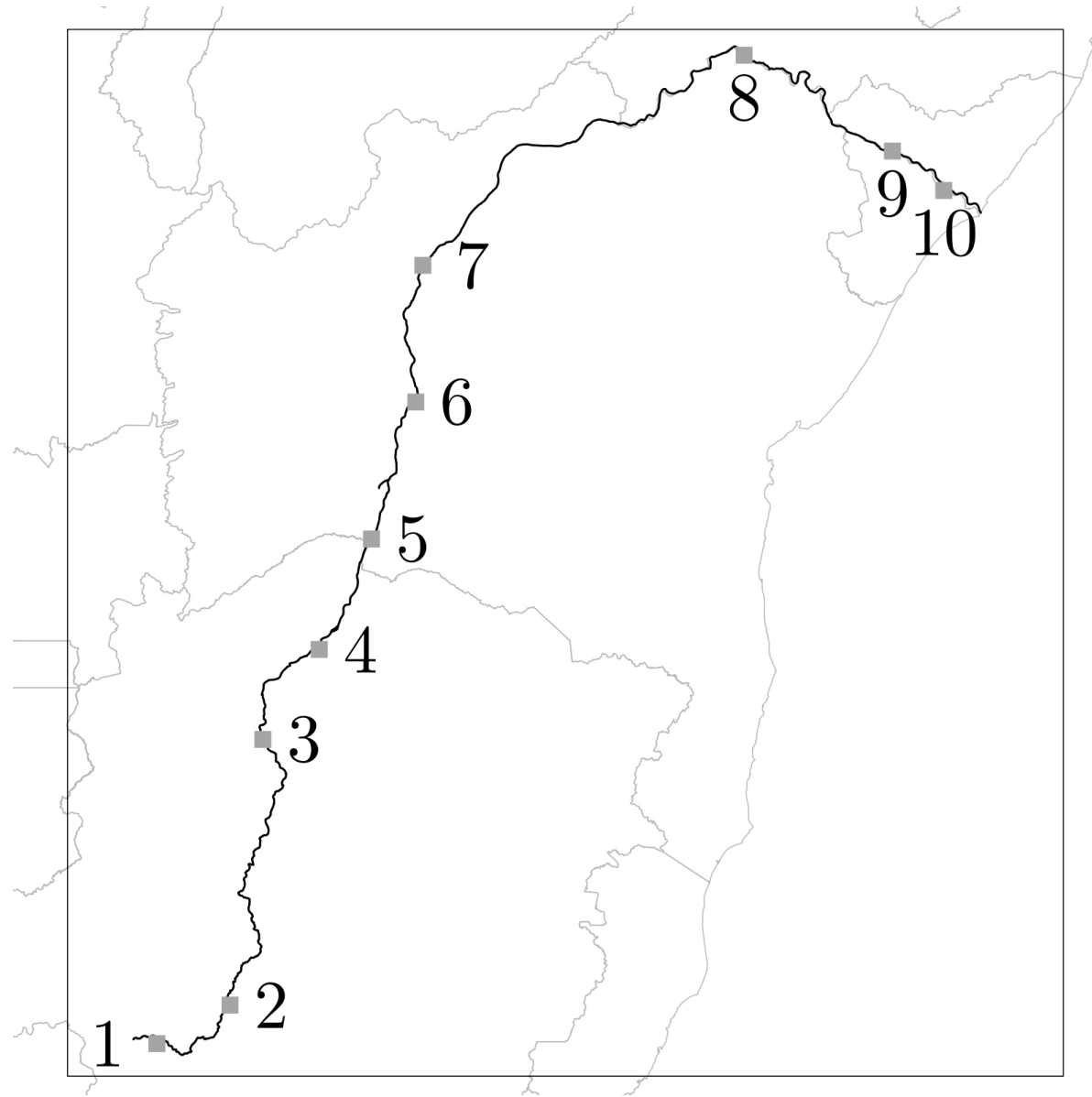
# The Choice of Penalizing Constant $c$



5-fold cross-validation error for Example 11, considering a sample of size 5,000 and set of values for penalizing constant.

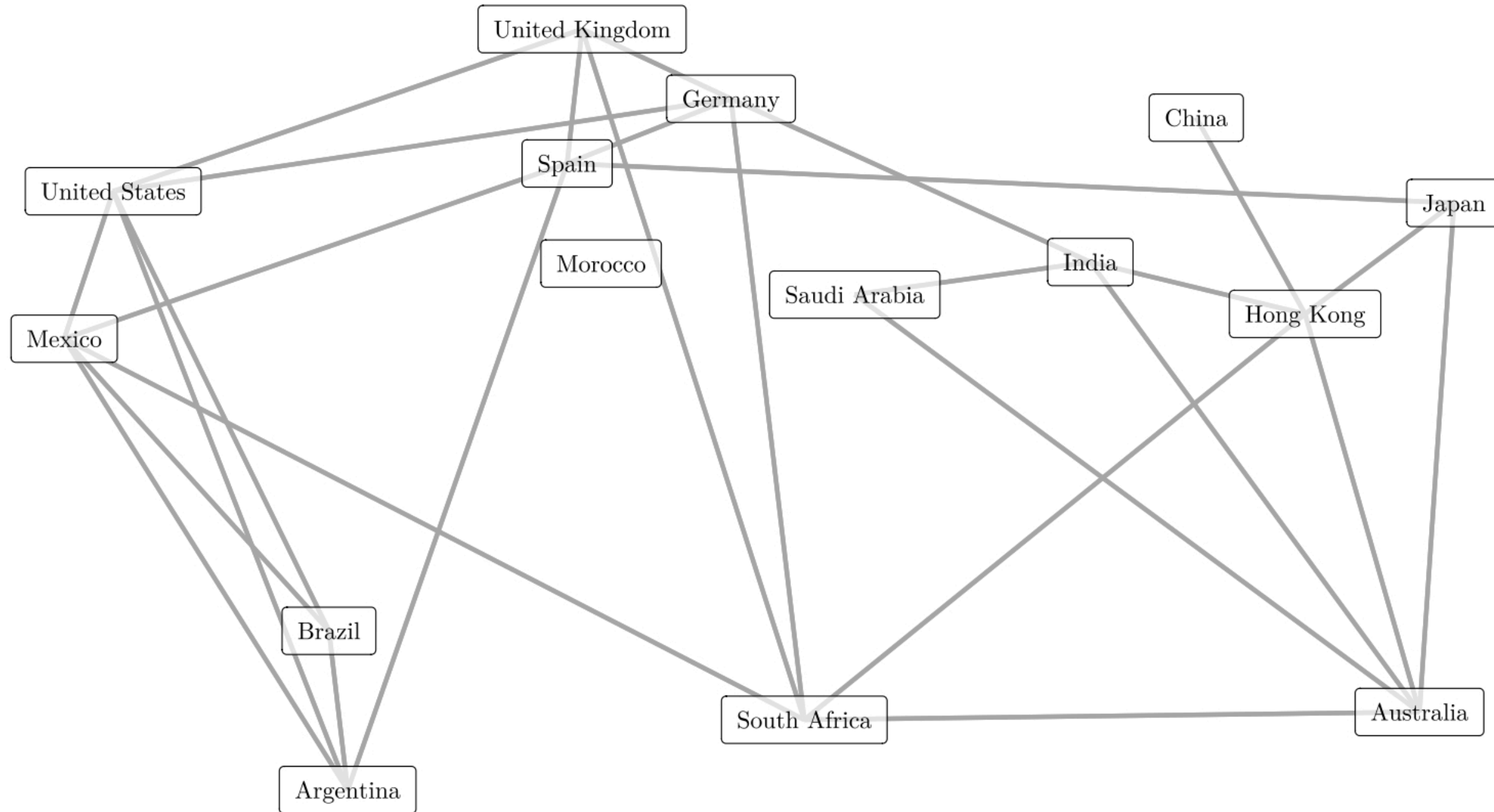
# São Francisco River Data

- Water volume measured at  $d$  stations along the river's course, denoted as  $X_u$ , where  $u = 1, \dots, d$ .
- Vector  $\mathbf{X}$  observed at discrete time intervals (10-day mean), from January 1977 to December 2016.
- Process  $\mathbf{X}^n = \{\mathbf{X}^{(i)} : 1 \leq i \leq n\}$ ,  $\mathbf{X}^{(i)} = (X_1^{(i)}, \dots, X_d^{(i)})$ .
- Forward stepwise algorithm and a 5-fold cross-validation approach below.



Results similar to Leonardi et al (2021).

# Stock Exchange Data



Estimated graph, considering the penalizing constant value chosen by cross-validation.

