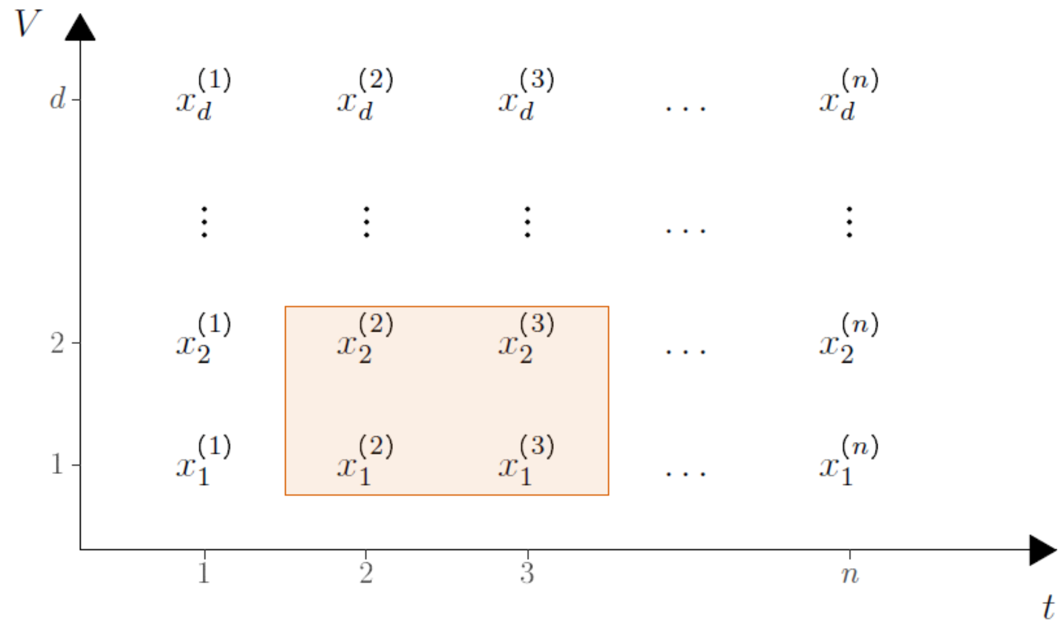
Estimation and model selection for mixing graphical models

Magno Severino & Florencia Leonardi

October 26, 2023

Agenda

- Introduction and motivation.
- Graph Estimator.
- Algorithms for estimation.
- Applications to real data.
- Directions for future work.



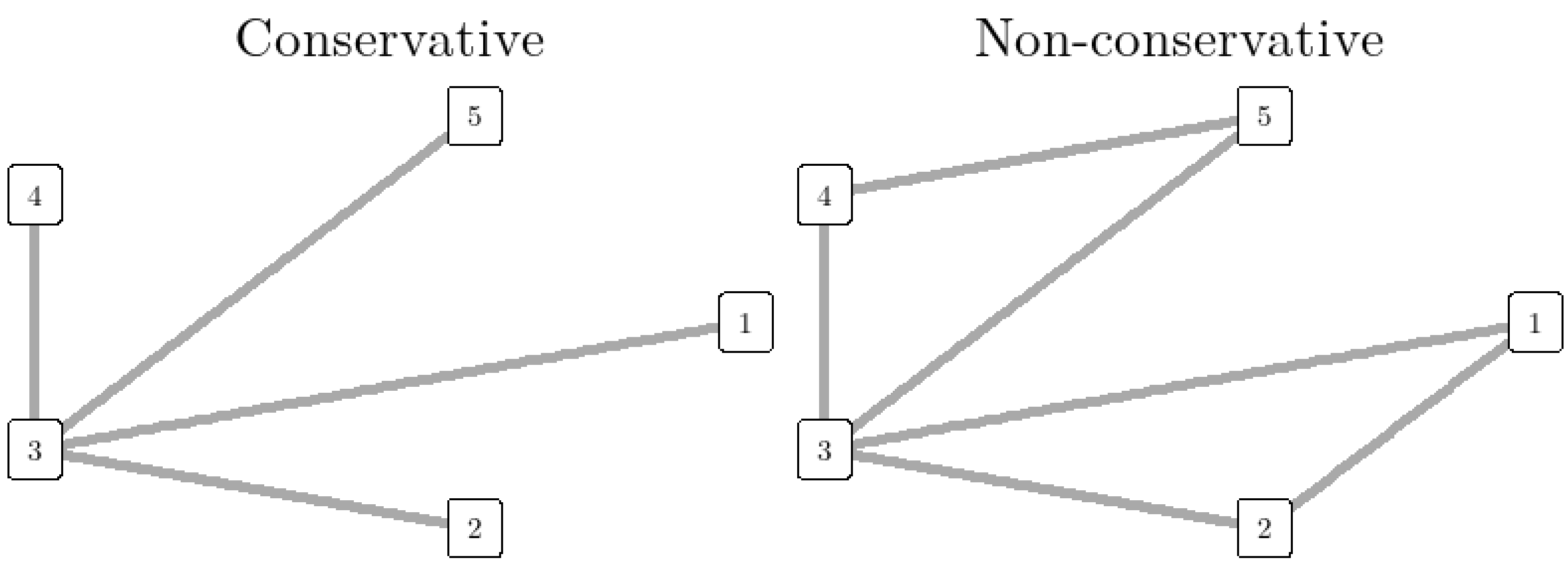
- Process $\mathbf{X} = \{X^{(i)} : -\infty < i < \infty\}$.
- $X^{(i)} = (X_1^{(i)}, \dots, X_d^{(i)})$, with set of vertices $V = \{1, \dots, d\}$.
- $X_v^{(i)} \in A$, A is a finite alphabet.
- Process observed at time $i \in \{1, \dots, n\}$.
- Subscript indicates the vertex and superscript the time the observation was taken.
- We assume that the process $\mathbf{X}$ has an underlying graph G^*.

Motivation 1: Leonardi et al. (2023)

Consider these estimated neighborhoods for a set of 5 discrete random variables:

$$\mathrm{ne}(\mathrm{X}_1) = \{2, 3\}, \qquad \mathrm{ne}(\mathrm{X}_2) = \{3\}, \qquad \mathrm{ne}(\mathrm{X}_3) = \{1, 2, 4, 5\},$$

$$\mathrm{ne}(\mathrm{X}_4) = \{3\}, \qquad \mathrm{ne}(\mathrm{X}_5) = \{3, 4\}.$$



Motivation 2: Leonardi et al. (2021)

- Leonardi et al. (2021) presents a model selection criterion for independence within random vectors.
- $\mathbf{Y} = \{Y^{(i)}\}$ is multivariate stochastic process.
- $Y^{(i)} = (Y_1^{(i)}, \ldots, Y_7^{(i)})$, with joint probability distribution $p(y_1, \ldots, y_7)$.
- The method decomposes the vector's distribution function into independent blocks.
- If the set of independence of $\mathbf{Y}$ is $U = \{3\}$, then

$$p(y_1, \ldots, y_7) = p(y_1, y_2, y_3)p(y_4, y_5, y_6, y_7).$$

- Therefore, $(Y_1, Y_2, Y_3) \perp (Y_4, Y_5, Y_6, Y_7)$.

Objectives of the research

- Proposal: Method to estimate the graph of conditional dependencies for multivariate stochastic processes with mixing conditions.
- Aims to overcome limitations of previous works:
  - Leonardi et al. (2023): estimator for iid data only
  - Leonardi et al. (2021): method assumes decomposition into subvectors with immediate neighbor dependencies.
- Proposed solution: penalized pseudo-likelihood criterion for entire graph estimation for multivariate processes with mixing conditions.
- Key advantages:
  - Handles non-iid data.
  - Provides a global estimation approach.

Mixing condition

Definition: For $i < j$, let $X^{(i:j)}$ denote the sequence of vectors $X^{(i)}, X^{(i+1)}, \dots, X^{(j)}$. We say the process $\mathbf{X} = \{X^{(i)}\colon -\infty < i < \infty\}$ satisfies a mixing condition with rate $\{\psi(\ell)\}_{\ell\in\mathbb{R}}$ if for each $k, m \in \mathbb{N}$ and each $x^{(1:k)} \in (A^d)^k, x^{(1:m)} \in (A^d)^m$ with $\mathbb{P}(X^{(1:m)} = x^{(1:m)}) > 0$, we have that

$$\Big|\mathbb{P}\big(X^{(n:(n+k-1))} = x^{(1:k)} \mid X^{(1:m)} = x^{(1:m)}\big) - \mathbb{P}\big(X^{(n:(n+k-1))} = x^{(1:k)}\big)\Big|$$

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

for $n \geq m + \ell$.

Regularized graph estimator

We take a **regularized pseudo maximum likelihood** approach to estimate the graph G^*, given a sample $x^{(1)}, \ldots, x^{(n)}$ of the stochastic process.

The pseudo log likelihood estimator is given by

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

The conditional probabilities are estimated from the observed data

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

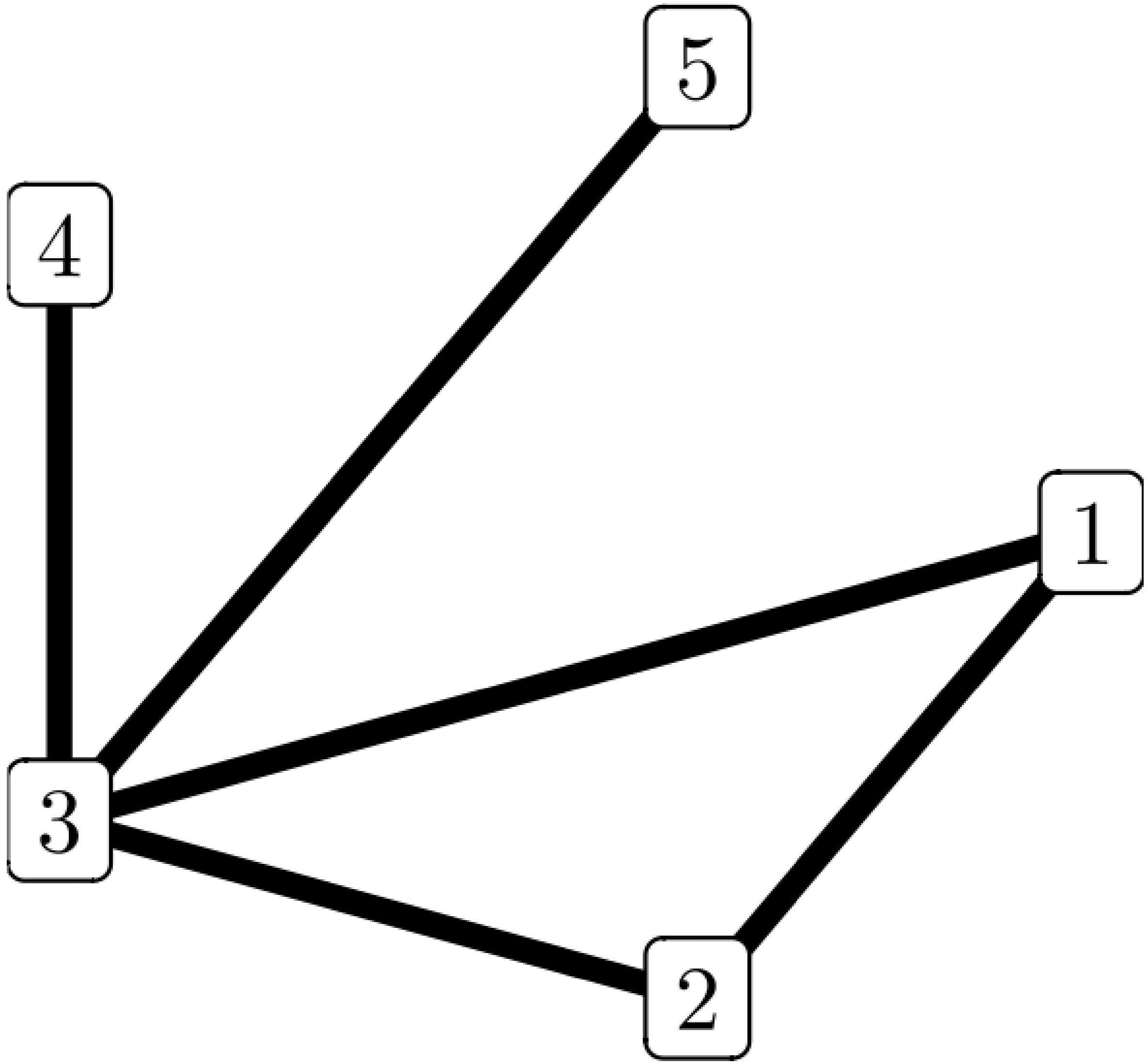
Therefore, the graph estimator is $\widehat{G} = \arg\max_G \Big\{ \log \widehat{L}(G) - \lambda_n \sum_{v \in V} |A|^{|G(v)|} \Big\}.$

We prove the consistency of this estimator.

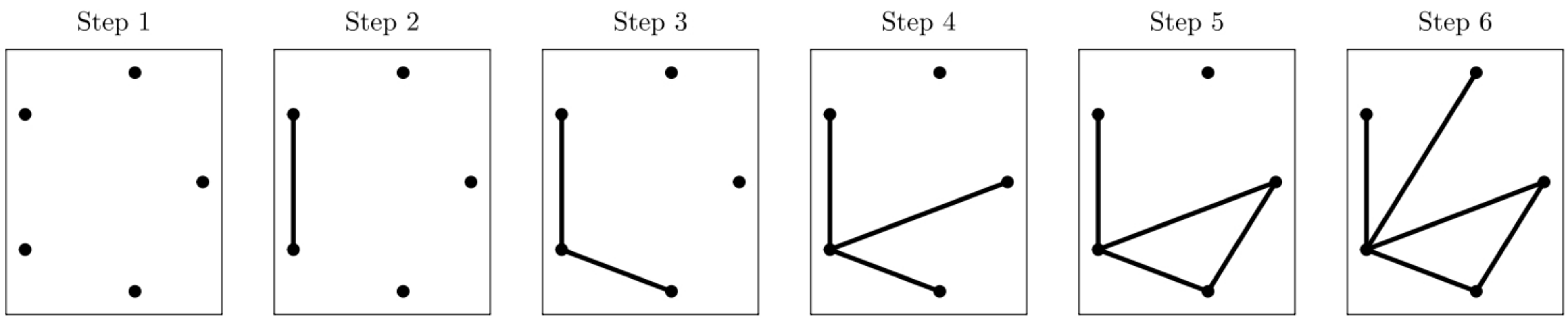
Algorithms for estimation

- Exact algorithm
- Simulated Annealing
- Stepwise selection algorithm

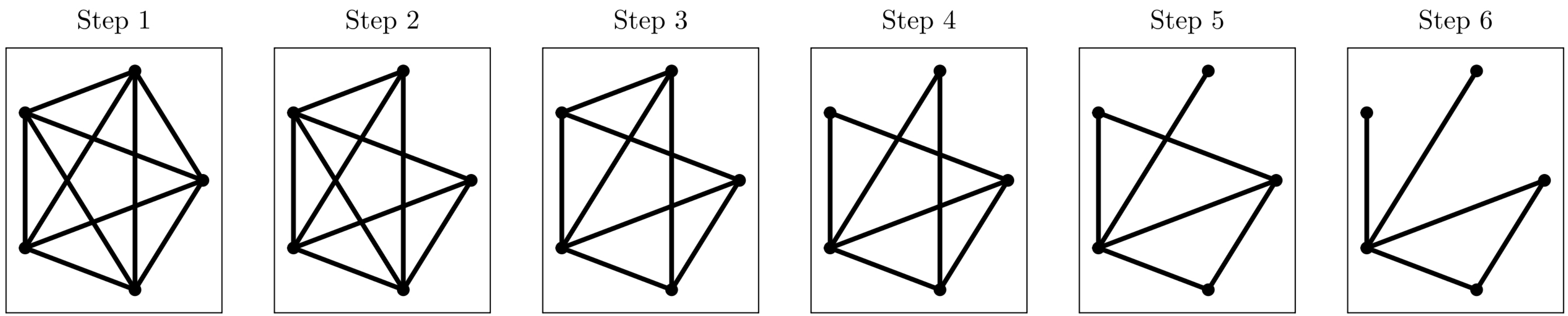
Illustrative example



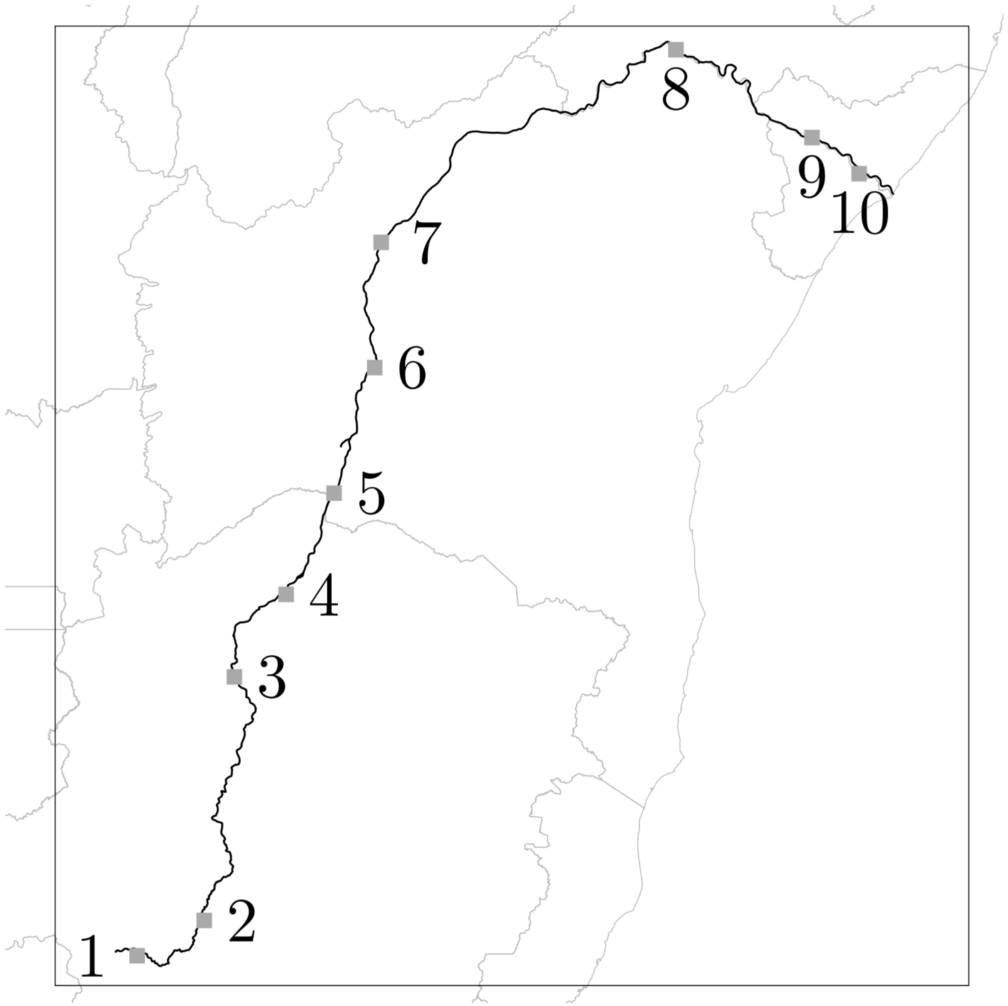
Forward stepwise algorithm

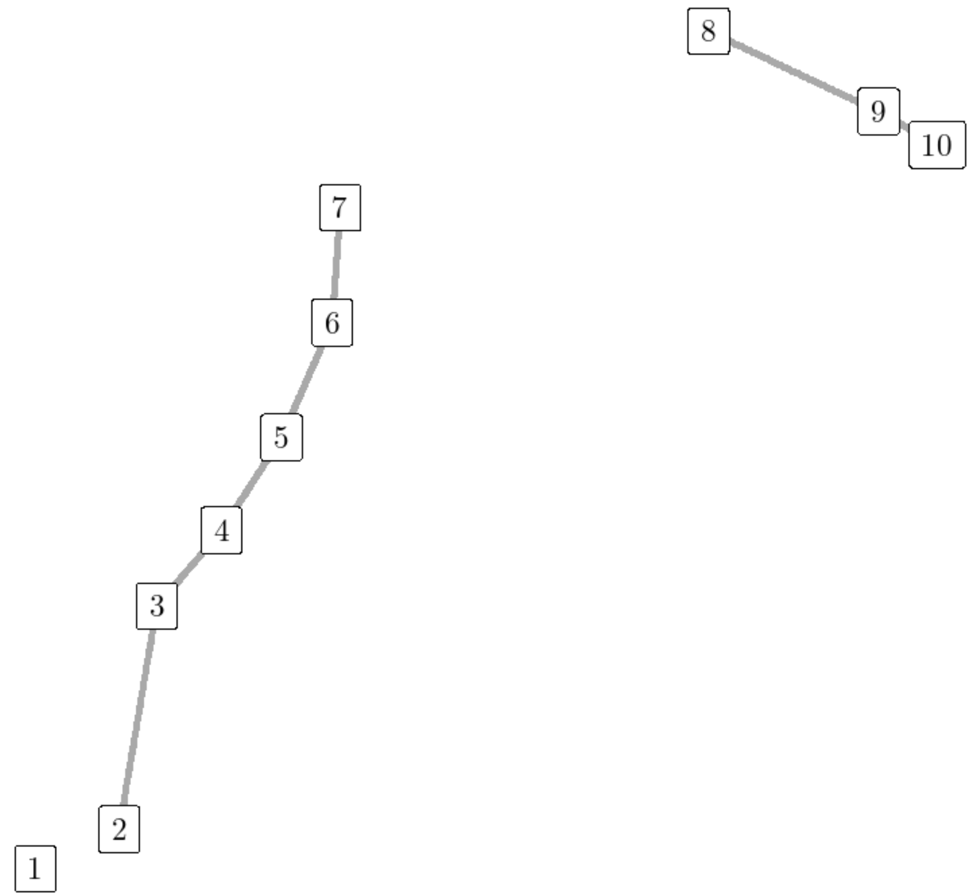


Backward stepwise algorithm



São Francisco river data





Directions for future work

- Generalize the approach to handle both discrete and continuous random variables.
  - It would broaden the applicability of the estimator.

- Adapt the model to estimate the direction of edges in graphs.
  - Address intricate dependencies among random variables.
  - Improve accuracy and insights in various applications.

Conclusion

- Study motivated by Leonardi et al. (2021) and Leonardi et al. (2023).
- Overcame limitations of previous works.
  - Handles non-iid data .
  - Provides a global estimation approach for the entire graph.
- Proved consistency and convergence rate of the proposed estimator.
- Proposed practical implementation methods: exact, simulated annealing, and stepwise algorithm.
- Factors affecting estimation process: number of nodes, alphabet size, and graph structure.

References

- Leonardi, F., Carvalho, R., and Frondana, I. (2023). Structure recovery for partially observed discrete markov random fields on graphs under not necessarily positive distributions. *Scandinavian Journal of Statistics*.
- Leonardi, F., Lopez-Rosenfeld, M., Rodriguez, D., Severino, M. T. F., and Sued, M. (2021). Independent block identification in multivariate time series. *Journal of Time Series Analysis*, 42(1):19–33.

Acknowledgements

This work was produced as part of the activities of the Research, Innovation and Dissemination Center for Neuromathematics (grant FAPESP 2013/07699-0).

The authors also received financial support from CAPES and CNPq.