Learning Shared Subgraphs in Ising Model Pairs

Burak Varıcı Saurabh Sihag Ali Tajer

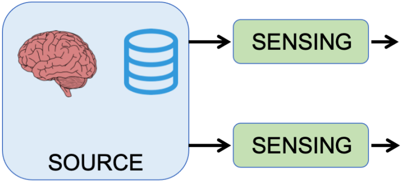
Rensselaer Polytechnic Institute

AISTATS 2021

Motivation

Multiple Information Networks

- Graphical models describe complex dependency structures.
- **Shared subgraphs** - joint information
  - Multiple brain imaging techniques.
  - Finding similar molecular structures for drug discovery.



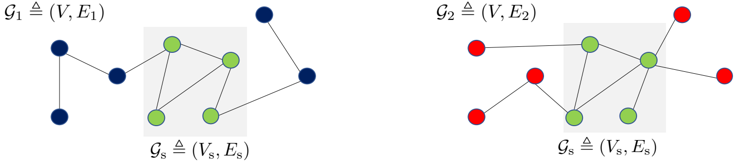
$$\mathbf{X}_1 \triangleq [X_1^1, \ldots, X_1^p], \quad \mathbf{X}_1 \sim f_1$$

$$\mathbf{X}_2 \triangleq [X_2^1, \ldots, X_2^p], \quad \mathbf{X}_2 \sim f_2$$

Each information layer $\rightarrow$ distinct representation

Shared Structure Learning

- Graph samples: $\mathbf{X}_i \triangleq [X_i^1, \ldots, X_i^p], \quad \mathbf{X}_i \sim f_i$ for $i \in \{1, 2\}$.
- Objective: Observe $\mathbf{X}_1, \mathbf{X}_2 \rightarrow$ estimate $\mathcal{G}_s \triangleq (V_s, E_s)$.
- Baseline: estimate $E_1, E_2 \rightarrow E_s = E_1 \cap E_2$: vastly inefficient
- Joint learning of $\mathcal{G}_s$ only.
- Ising Model: $f(\mathbf{X}) = \frac{1}{Z} \exp\left(\sum_{(u,v)\in E} \lambda X^u X^v\right)$



Problem Formulation

- **Goal.** Learning *only* the shared subgraph $\mathcal{G}_s$.
- Graph decoder $\psi_{\rm s}$: maps the data samples to class of subgraphs.

$$\psi_{\rm s} : \mathcal{X}^{n\times p} \times \mathcal{X}^{n\times p} \rightarrow \mathcal{I}_p^{\rm s}$$

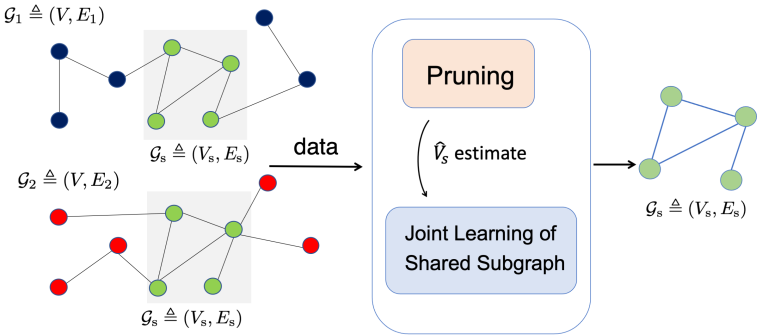
- Exact recovery: Perfectly learn $\mathcal{G}_s$

$$\mathsf{P}_{\rm L}(\mathcal{I}_p^{\rm s}) \triangleq \max_{\mathcal{G}_1,\mathcal{G}_2\in\mathcal{I}_p} \mathbb{P}(|E_s \Delta \hat{E}_s| \neq 0)$$

- Vertex sample complexity: $\mathsf{N}(n_{\rm T}) = \sum_{k=1}^{n_{\rm T}} |\hat{V}_s(k)|$

Algorithmic Framework

- An adaptive algorithm focused on learning only $\mathcal{G}_s$.
  - Pruning step and joint learning step.



Algorithmic Framework

Similarity-based Pruning

- Form coarse estimates $\hat{V}_s(k), \hat{E}_s(k)$ at each iteration k via:

$$\min_{i\in\{1,2\}} \bar{\mathbb{E}}_k[X_i^u X_i^v] > \tanh\lambda - \sqrt{\alpha\log p/2k}, \quad \forall k$$

- Importance: Narrow down sampling to **only** V_s adaptively.

$$\mathbb{P}(V_s \subseteq \hat{V}_s(k)) \geq 1 - 2p^{2-\alpha}, \quad \forall k$$

- Sample complexity: $k = O(\frac{\alpha\log p}{\lambda^2})$ in correlation decay regime ensures

$$\mathbb{P}(\hat{V}_s(k) = V_s) \geq 1 - 2p^{2-\alpha}.$$

- **Joint** multiplicative updates at every iteration

$$w_i^{uv}(k+1) = w_i^{uv}(k) \cdot \exp\left(\frac{\beta}{2}(\ell_1^{uv}(k) + \ell_2^{uv}(k))\right), \qquad u, v \in \hat{V}_s(k)$$

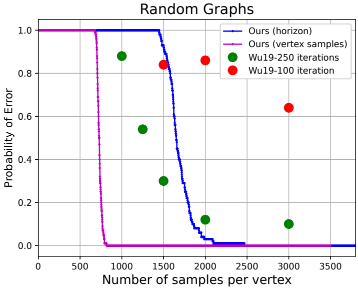
- Importance: Joint updates improve learning of E_s, and samples from only $\hat{V}_s(k)$ suffice.
- Sample complexity: When $\mathcal{G}_s$ is *isolated*, and pruning localizes V_s, for ensuring $\mathrm{P_L}(\mathcal{I}_p^{\mathrm{s}}) \leq (1 - \frac{2}{\rho})$,

Joint (ours): $O\left(\frac{1}{\lambda^2}\exp(\lambda d)\log\frac{\rho q}{\lambda}\right)$ where $q = |V_s| < p$

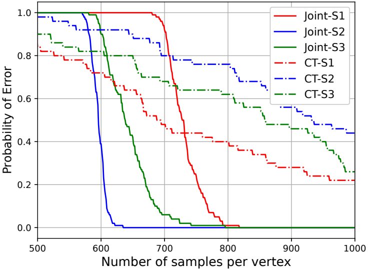
Independent: $O\left(\frac{1}{\lambda^2}\exp(\lambda d)\log\frac{\rho p}{\lambda}\right)$ (Klivans and Meka, 2017)

Simulations

- Erdős-Rényi random graphs with $p = 200$ vertices, $|V_s| = 20$.
- Baseline approach: learn E_1, E_2 separately and form $E_s = E_1 \cap E_2$.



Ours vs. Sparse logistic regression (Wu et al. NeurIPS'19)



Ours vs. Correlation Thresholding algorithm (Anandkumar et al. 2010)

Conclusions

- Novel problem of learning the shared structure of two graphs.
- An algorithmic framework and its evaluation in various regimes.
- Sample complexity analysis for specific settings.