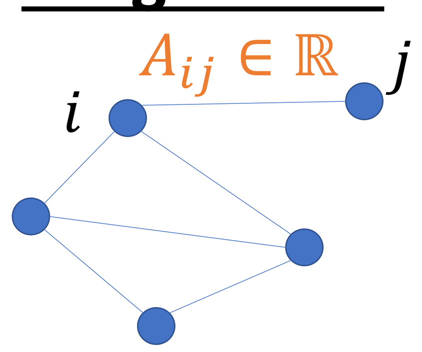
# Sparse Logistic Regression Learns All Discrete Pairwise Graphical Models

Shanshan Wu, Sujay Sanghavi, Alex Dimakis

*University of Texas at Austin*

## [ Problem Formulation ]

- **Ising model**:




Edge weight b/t $i$ & $j$ External field

$$\mathbb{P}[Z=z] \propto \exp\Big(\sum_{1\le i<j\le n} A_{ij} z_i z_j + \sum_{i\in[n]} \theta_i z_i\Big)$$

An undirected graph on $n$ nodes

A distribution over $Z \in \{-1,1\}^n$

- Ising model is a special case of discrete pairwise graphical models: the alphabet size $k=2$.

  **Note:** For simplicity, this poster focuses on Ising models. See our paper for results on general graphical models.

- The standard **structural learning** problem:

  Given $N$ i.i.d. samples from a graphical model

  🤔

  Can we recover the graph structure?

## [ Our Contributions ]

- **Our main results**: We prove that a classical algorithm $\ell_{2,1}$-constrained logistic regression can recover the edge weights of any discrete pairwise graphical model from i.i.d. samples.

| Algorithm | Sample complexity |
|---|---|
| Greedy algorithm [Hamilton et al.'17] | $O(\exp(\frac{k^{O(d)}\exp(d^2\lambda)}{\eta^{O(1)}})\ln(nk))$ |
| Sparsitron [Klivans and Meka'17] | $O(\frac{\lambda^2 k^5 \exp(14\lambda)}{\eta^4}\ln\left(\frac{nk}{\eta}\right))$ |
| $\ell_{2,1}$-constrained logistic regression [This work] | $O(\frac{\lambda^2 k^4 \exp(14\lambda)}{\eta^4}\ln(nk))$ |

**Table 1**. Sample complexity in terms of alphabet size $k$, model width $\lambda$, degree $d$, minimum edge weight $\eta$, and number of variables $n$. For Ising models ($k=2$), the algorithm reduces to $\ell_1$-constrained logistic regression.

## [ Algorithm ]

logistic: $\sigma(z) = \frac{1}{1+e^{-z}}$ $x' = [x,1] \in \{-1,1\}^n$

- **Fact:** $\mathbb{P}[Z_i = 1 | Z_{-i} = x] = \sigma(w^* \cdot x')$

  $w^* = 2[A_{i1}, \dots, A_{i(i-1)}, A_{i(i+1)}, \dots, A_{in}, \theta_i] \in \mathbb{R}^n$

- **Model width:** $\lambda = \max_{i\in[n]}(\sum_j |A_{ij}| + |\theta_i|)$ $\|w^*\|_1 \le 2\lambda$

- **Algorithm:**

  Estimate $A_{ij}$ to maximize the conditional likelihood ➡ $\ell_1$-constrained logistic regression

## [ Proof Sketch ]

- **Def**: $L$ and $\hat{L}$ are the expected & empirical logistic loss

  $$\hat{w} \in \operatorname{argmin} \hat{L}(w) \quad \text{s.t. } \|w\|_1 \le 2\lambda$$

- **Step 1:** Given $O(\lambda^2 \ln(n/\rho)/\gamma^2)$ samples, w.p. $\ge 1-\rho$

  $$L(\hat{w}) - L(w^*) \le \gamma.$$

  **Note:** This uses a generalization bound for logistic regression with an $\ell_1$ constraint on $w$ and an $\ell_\infty$ constraint on $x$ [Kakade et al.'09].

- **Step 2:** For any $w$,

  $$L(w) - L(w^*) \ge \mathrm{E}_X\big(\sigma(w\cdot X) - \sigma(w^*\cdot X)\big)^2.$$

  **Note:** This uses the property of logistic function and Pinsker's inequality.

- After **Step 1** and **Step 2**, we have

  $$\mathrm{E}_X\big(\sigma(\hat{w}\cdot X) - \sigma(w^*\cdot X)\big)^2 \le \gamma.$$

- **Step 3:** we use a result from [Klivans and Meka'17]:

  If samples ~ Ising model, and $\gamma = O(\epsilon^2/\exp(6\lambda))$,

  $$E_X\big(\sigma(w\cdot X) - \sigma(w^*\cdot X)\big)^2 \le \gamma$$

  $$\big\|w_{1:(n-1)} - w^*_{1:(n-1)}\big\|_\infty \le \epsilon$$

**Note:** The constrained logistic reg. can be approximately solved in $\tilde{O}(n^2)$ time with the same statistical guarantee. See our papers for details.

## [ General Alphabet Size ]

- Each edge now has a group of parameters
- Enforce sparsity in the group level

Alphabet size $k$ ➡

$$\hat{w} \in \operatorname*{argmin}_{w\in\mathbb{R}^{n\times k}} \hat{L}(w) \quad \text{s.t. } \|w\|_{2,1} \le 2\lambda\sqrt{k}$$

Group sparsity

## [ Experiments ]

- **Code:** https://github.com/wushanshan/GraphLearn
- **Learning Ising Models:**

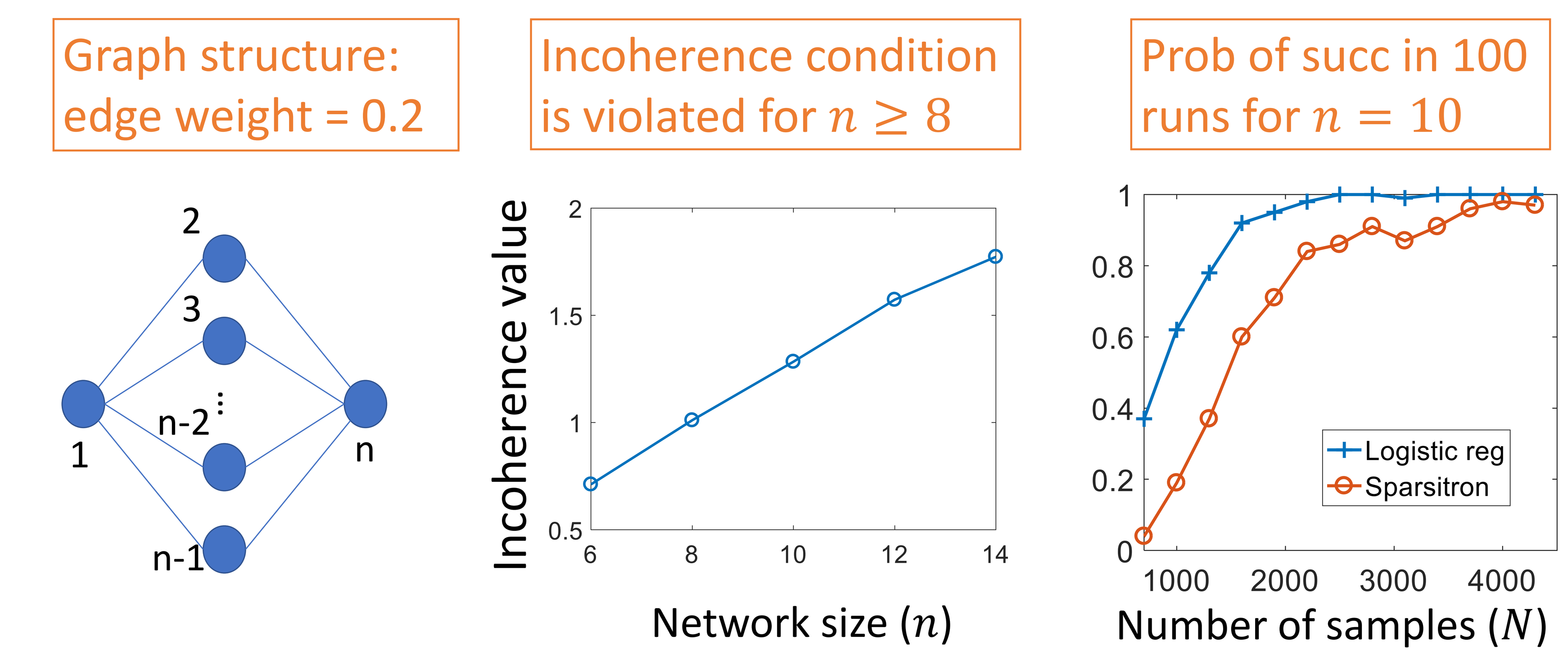



**Figure 1**. Sparse logistic regression can recover the graph even when the incoherence condition [Ravikumar et al.'10] is violated.

- **Learning general pairwise graphical models:**

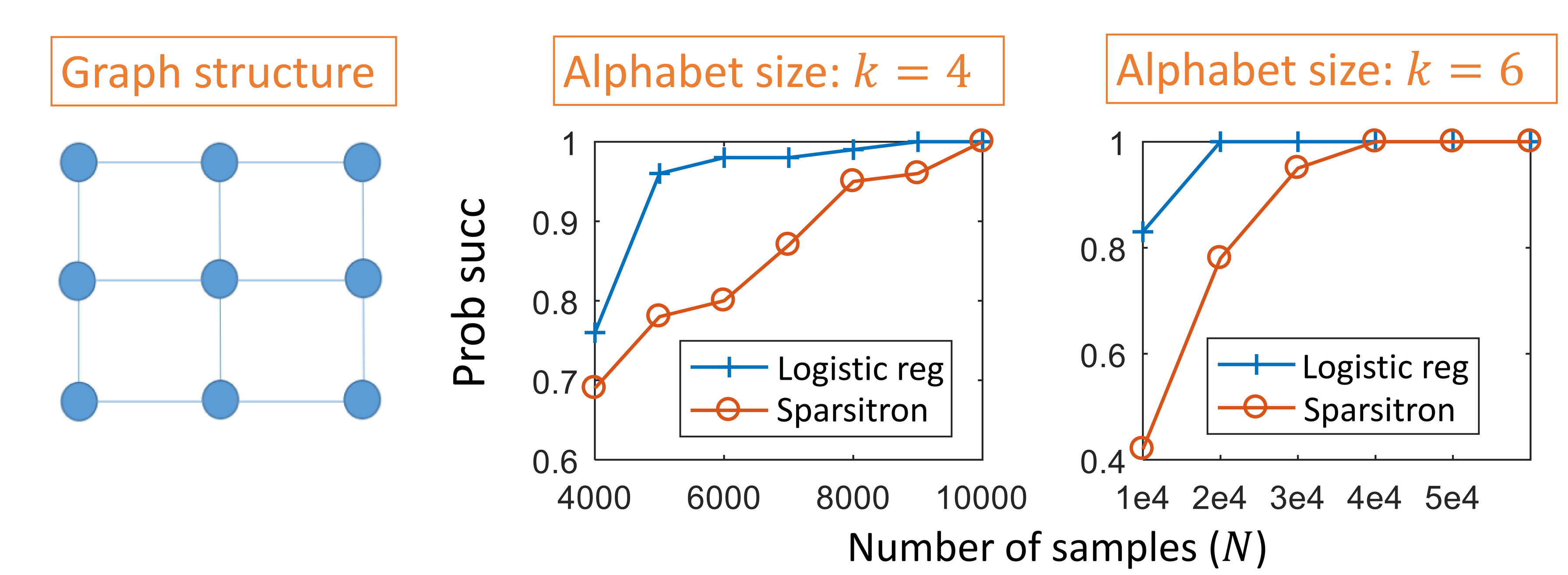



**Figure 2**. Sparse logistic regression requires fewer samples for graph recovery than the Sparsitron algorithm [Klivans and Meka'17] .