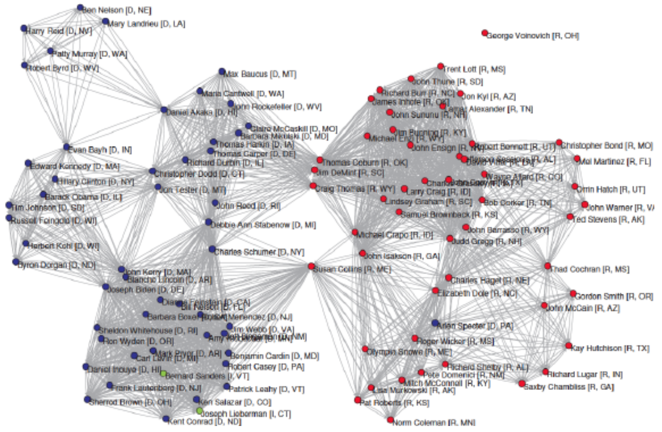
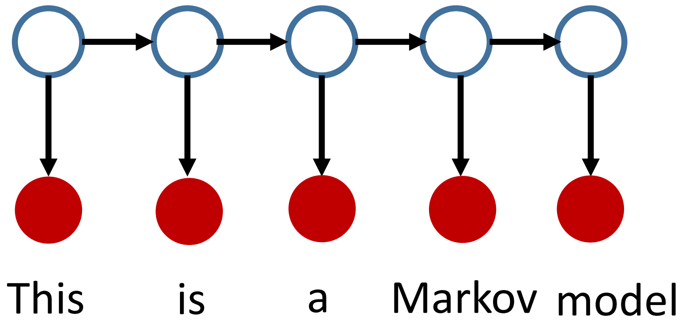
Sparse Logistic Regression Learns All Discrete Pairwise Graphical Models

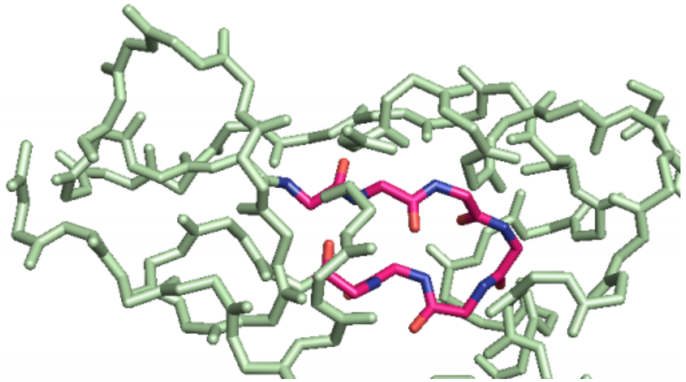
Shanshan Wu, Sujay Sanghavi, Alex Dimakis

University of Texas at Austin

Graphical models are used to describe complex dependency structures







Social network analysis

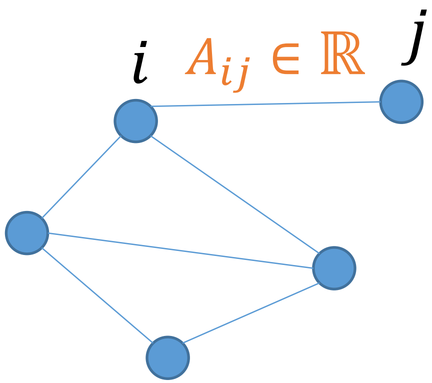
Natural language processing

Biology

J. Guo et al., “Estimating heterogeneous graphical models for discrete data with an application to roll call voting”. Annuals of Statistics, 2015.
H. Kamisetty et al., “Free Energy Estimates of All-atom Protein Structures Using Generalized Belief Propagation”, RECOMB 2007

Discrete pairwise graphical model

- **Binary case** (aka 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 $Z \in \{-1,1\}^n$

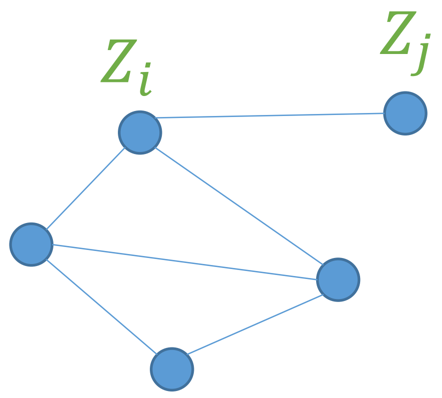
- **Non-binary case** (alphabet size k): $Z \in \{1,2,\ldots,k\}^n$

The structure learning problem

Given: i.i.d. samples from an unknown graphical model

Goal: Recover the graph, i.e., identify the edges

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	……	Z_n
Sample 1	[-1	1	-1	-1	-1	1	……	1]
Sample 2	[1	-1	-1	1	-1	-1	……	1]
…				…				



- Algorithms: Ravikumar et al.'2010, Jalali et al.'2011, Bresler'2015, Vuffray et al.'2016, Lokhov et al.'2018, Hamilton et al.'2017, Klivans and Meka'2017, Rigollet and Hütter'2019, Vuffray et al.'2019 …

A simple approach...

Maximize the conditional log-likelihood

Binary case

ℓ_1-regularized logistic regression
[Ravikumar et al.'10]

Non-binary case

$\ell_{2,1}$-regularized logistic regression
[Jalali et al.'11]

Limitation of [Ravikumar et al.'10, Jalali et al.'11]

Assuming that the graphical models satisfy
an incoherence condition,

sparse logistic regression provably recover
the graph structure.

Our contribution

~~Assuming that the graphical models satisfy an incoherence condition,~~

For all graphical models,

sparse logistic regression provably recover the graph structure.

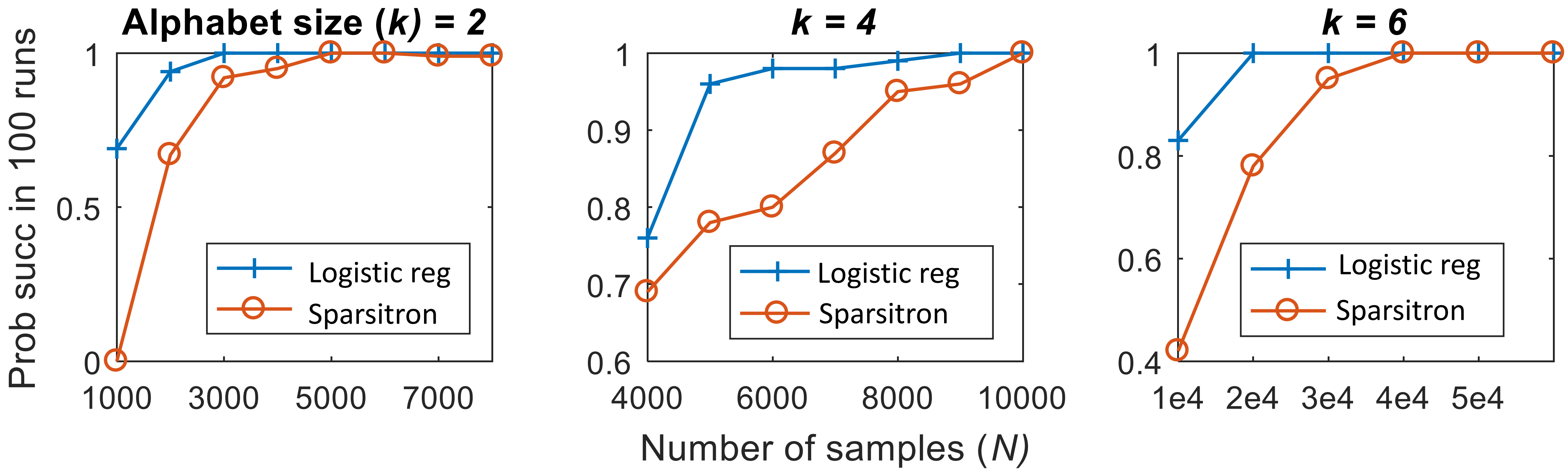
Our contribution

- Let n = # variables, alphabet size k, width λ, minimum edge weight η

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 [Our work]	$O(\frac{\lambda^2 k^4 \exp(14\lambda)}{\eta^4}\ln(nk))$

Improves from k^5 to k^4!

Experiments (grid graph)



Sparse logistic regression requires fewer samples for graph recovery.

Poster #183

Today 10:45 AM -- 12:45 PM

@East Exhibition Hall B + C