SGA: A Robust Algorithm for Partial Recovery of Tree-Structured Graphical Models with Noisy Samples

Anshoo Tandon,[1] Aldric J. Y. Han,[2] and Vincent Y. F. Tan[1,2]

Department of {[1]Electrical and Computer Engineering, [2]Mathematics}, National University of Singapore, Singapore

Emails: anshoo.tandon@gmail.com, e0175459@u.nus.edu, vtan@nus.edu.sg

Tree Ising Model

- Ising model for a d node tree with random variables $X_1, \ldots, X_d$:

$$P(\mathbf{x}) = \frac{1}{Z} \exp\Bigg(\sum_{\{i,j\} \in \mathcal{E}} \theta_{i,j}\, x_i x_j \Bigg),$$

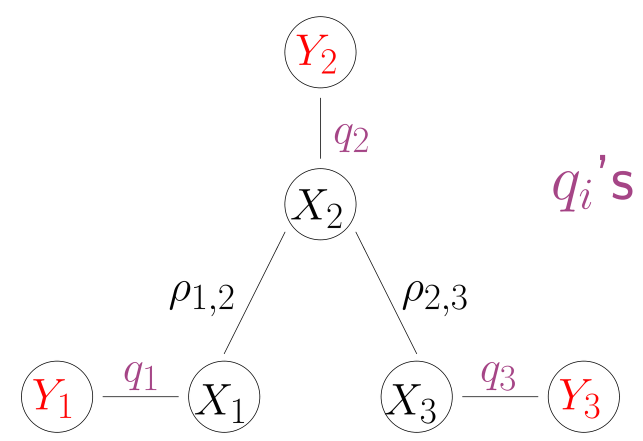
where $x_i \in \mathcal{X} = \{+1, -1\}$ with $\mathbb{E}[X_i] = 0$, $\mathcal{E}$ is the set of edges, $\theta_{i,j}$ are edge interaction parameters, and Z is the normalization constant

- For $\{i,j\} \in \mathcal{E}$, the correlation $\rho_{i,j} \triangleq \mathbb{E}[X_i X_j] = \tanh(\theta_{i,j})$

Non-Identically Distributed Noise

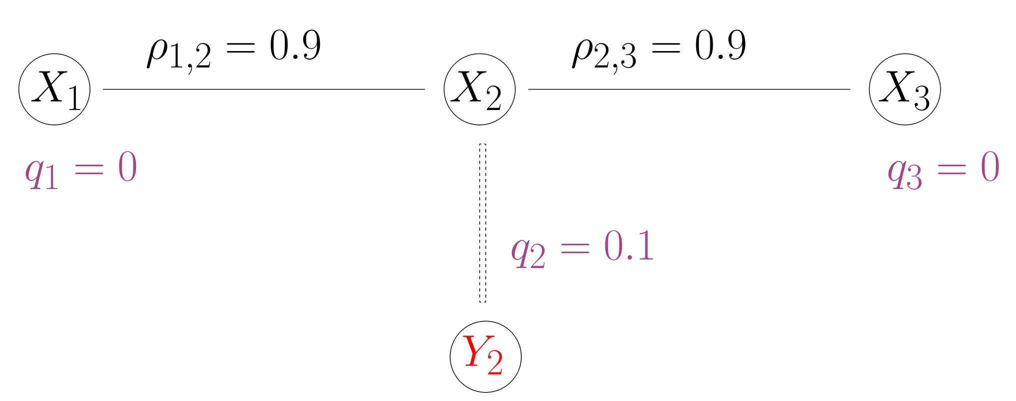
- Noise Model: For $i \in \{1, \ldots, d\}$, we observe $Y_i = X_i N_i$ where

$$\Pr(N_i = -1) = q_i, \quad \Pr(N_i = +1) = 1 - q_i$$



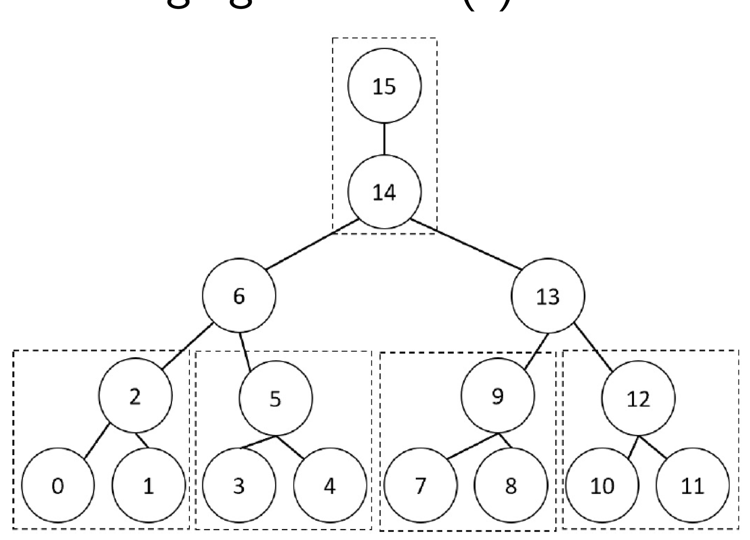
q_i's need not be the same!

- With non-identically distributed noise, the classical Chow-Liu algorithm may not be able to recover the tree-structure. For example:



Partial Tree Recovery

- Katiyar-Shah-Caramanis [arXiv, Jun. 2020] gave an algorithm for partial tree-structure recovery (up to an equivalence class)
- For a given tree T, the elements in the equivalence class $[\mathrm{T}]$ are obtained by interchanging leaf node(s) with their respective parent node



Katiyar, A., Shah, V., and Caramanis, C. *Robust estimation of tree structured Ising models*. arXiv:2006.05601 [stat.ML], Jun. 2020

SGA Algorithm

- SGA is adapted from the procedure by Katiyar-Shah-Caramanis [arXiv, Jun. 2020] for declaring any 4 nodes as star or non-star
- Overview of SGA algorithm via an example:
  - Let $\{X_1, X_2, X_3, X_4\}$ form a non-star with pair $\{X_1, X_2\}$
  - Let $\widehat{\rho}_{i,j}$ denote the empirical correlation between nodes i and j
  - Then, we would expect the following two equations to hold

$$\text{(i)} \;\; \frac{\widehat{\rho}_{1,3}\,\widehat{\rho}_{2,4}}{\widehat{\rho}_{1,2}\,\widehat{\rho}_{3,4}} < \frac{1+\rho_{\max}^2}{2}, \quad \text{and} \quad \text{(ii)} \;\; \frac{\widehat{\rho}_{1,4}\,\widehat{\rho}_{2,3}}{\widehat{\rho}_{1,2}\,\widehat{\rho}_{3,4}} < \frac{1+\rho_{\max}^2}{2}$$

  - The procedure by Katiyar et al. checks eq. (i) but ignores eq. (ii)
  - SGA computes the Geometric Average of (i) and (ii) to check if

$$\sqrt{\left|\frac{\widehat{\rho}_{1,3}\,\widehat{\rho}_{2,4}}{\widehat{\rho}_{1,2}\,\widehat{\rho}_{3,4}}\right| \cdot \left|\frac{\widehat{\rho}_{1,4}\,\widehat{\rho}_{2,3}}{\widehat{\rho}_{1,2}\,\widehat{\rho}_{3,4}}\right|} = \frac{\sqrt{|\widehat{\rho}_{1,3}\,\widehat{\rho}_{2,4}\,\widehat{\rho}_{1,4}\,\widehat{\rho}_{2,3}|}}{|\widehat{\rho}_{1,2}\,\widehat{\rho}_{3,4}|} \overset{?}{<} \frac{1+\rho_{\max}^2}{2}$$

- SGA has two useful properties:
  - Symmetry: Invariant to permutation of node indices
  - Robustness: Geometric Averaging of metrics makes it robust to noise

Partial Tree Recovery: Novel Converse Result

- Let $\mathcal{M}_n(q_{\max}, \rho_{\min}, \rho_{\max})$ denote the minimax error probability when $0 \le q_i \le q_{\max} < 0.5$, and $0 < \rho_{\min} \le |\rho_{i,j}| \le \rho_{\max} < 1$
- Converse Result: Let $\rho_q \triangleq (1 - 2q_{\max})\rho_{\min}$. If $d > 32$, and the number of samples n satisfy

$$n < \frac{\log(d)}{4\,(1-\rho_{\max})\,\rho_q\,\mathrm{atanh}(\rho_q)}$$

then we have $\mathcal{M}_n(q_{\max}, \rho_{\min}, \rho_{\max}) \ge 1/2$.

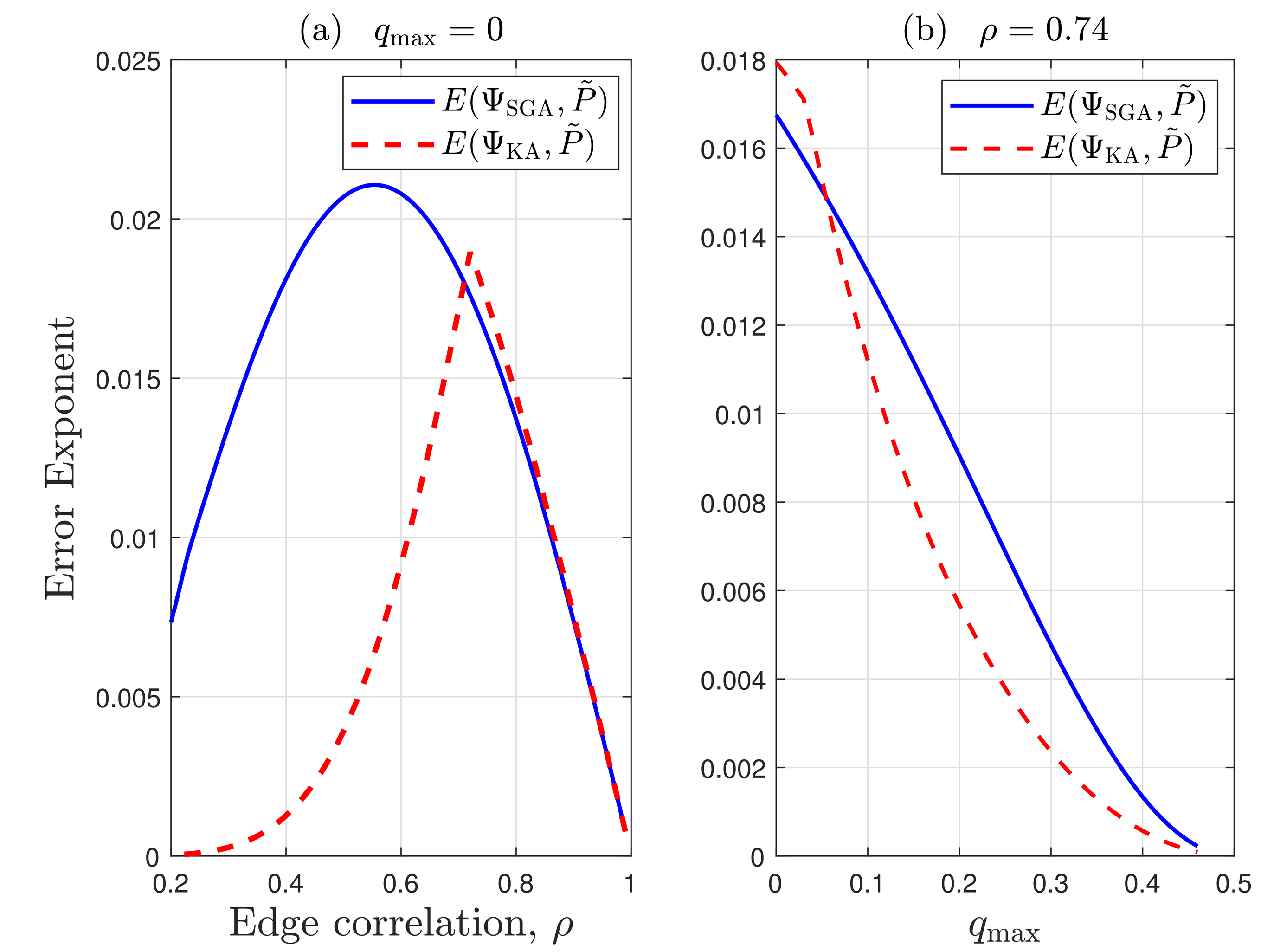
- Our proof has two key ingredients: (i) Choice of a sufficiently large number of 'close' tree structures whose equivalence classes are disjoint, (ii) Choice of noise parameters for different nodes that have a high impact on the error probability

Summary of Contributions

- We improve the sufficient sample complexity result of Katiyar et al. by reducing the dependence on minimum correlation from $\rho_{\min}^{-24}$ to $\rho_{\min}^{-8}$
- We present a modified procedure, SGA, for declaring a set of 4 nodes as star/non-star, that outperforms the algorithm by Katiyar et al.
- We provide an error exponent analysis that provides the intuition why SGA outperforms the algorithm by Katiyar et al.
- We present a novel converse result, quantifying necessary number of samples, for partial tree structure recovery under non-identical noise

Error Exponents for a 4-node Markov chain

Homogeneous Markov chain with edge correlation ρ

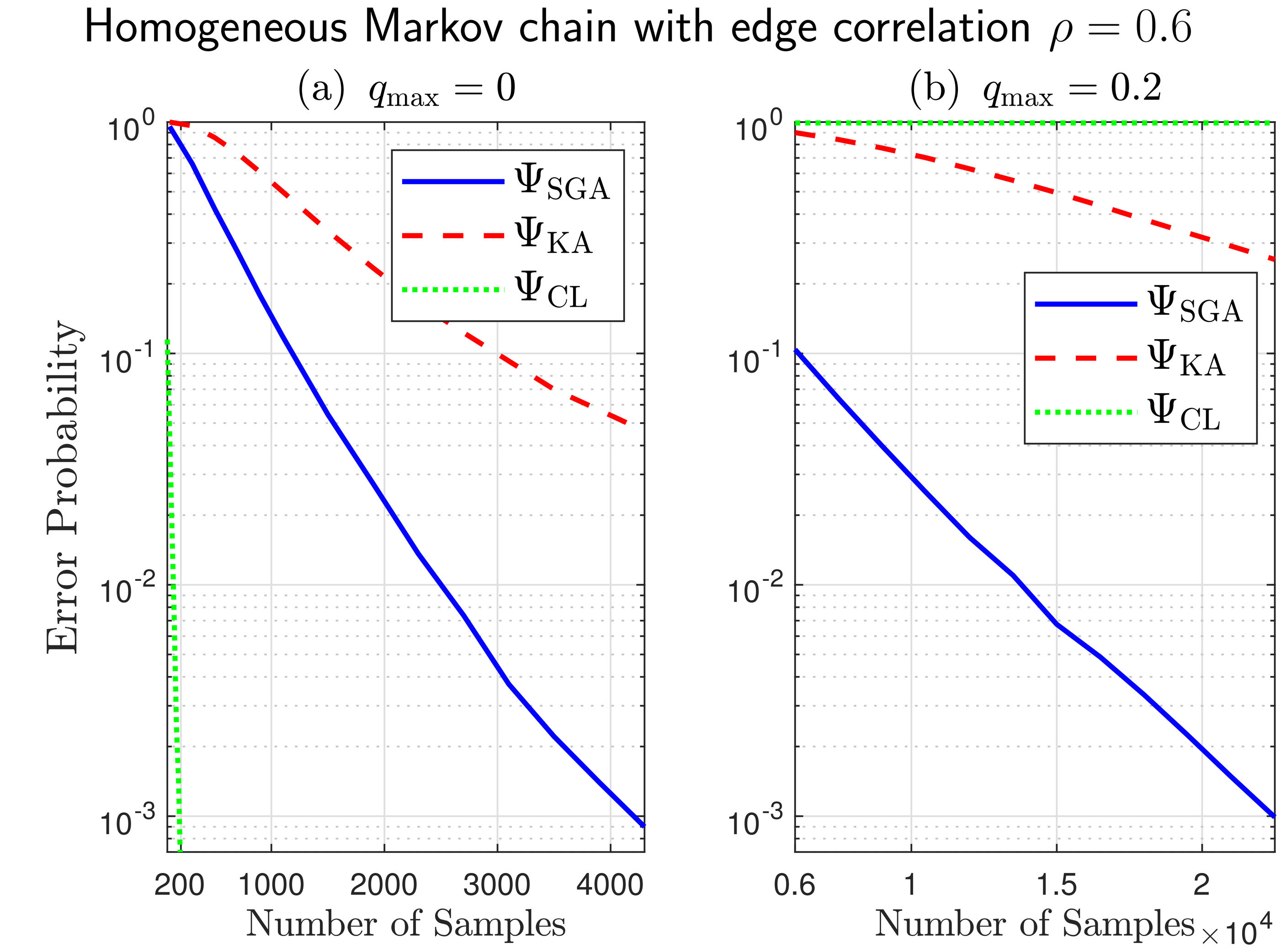


Ψ_{SGA}: SGA Algorithm, Ψ_{KA}: Algorithm by Katiyar et al.

- Error exponents quantify the exponential decay of error probability
- When ρ is relatively small, Ψ_{SGA} has a much larger error exponent (and hence better) compared to Ψ_{KA}

Simulation Results for a 12-node Markov chain

Homogeneous Markov chain with edge correlation $\rho = 0.6$



- For Fig. (b), $q_i = 0$ for odd indices and $q_i = 0.2$ for even indices
- The Chow-Liu algorithm, Ψ_{CL}, performs very well for the noiseless setting (a), but fails miserably in the noisy setting (b)
- Ψ_{SGA} performs robustly, and outperforms Ψ_{KA} both in (a) and (b)