

## Joint AMS/BMI/CS/ECE Seminar

### Continuous Optimization for Learning Bayesian Networks

**Speaker:** Dr. Yue Yu, Associate Professor, Department of Mathematics, Lehigh University

<https://www.lehigh.edu/~yuy214/>

**Time:** Wednesday, October 12, 10-11am

**Location (in-person seminar):** NCS 120, New Computer Science Building

**Zoom link:**

<https://stonybrook.zoom.us/j/9121212121?pwd=RjVlVXg3YUhudzZZQ3pheHUydTJBUT09>

Meeting ID: 251 330 7585; Passcode: 437562

**Abstract:** Bayesian networks are directed probabilistic graphical models used to compactly model joint probability distributions of data. Automatic discovery of their directed acyclic graph (DAG) structure is important to causal inference tasks. However, learning a DAG from observed samples of an unknown joint distribution is generally a challenging combinatorial problem, owing to the intractable search space superexponential in the number of graph nodes. A recent breakthrough formulates the problem as a continuous optimization with a structural constraint that ensures acyclicity (NOTEARS, Zheng et al., 2018), which enables a suite of continuous optimization techniques to be used and employs an augmented Lagrangian method to apply the constraint.

In this talk, we take a step further to propose new continuous optimization algorithms and models aiming to improve NOTEARS on both efficiency and accuracy. We first show that the Karush-Kuhn-Tucker (KKT) optimality conditions for the NOTEARS formulation cannot be satisfied except in a trivial case, which explains a behavior of the slow convergence. We then derive the KKT conditions for an equivalent reformulation, show that they are indeed necessary, and relate them to explicit constraints that certain edges are absent from the graph. Informed by the KKT conditions, a local search post-processing algorithm is proposed and shown to substantially and universally improve the learning accuracy, typically by a factor of 2 or more. Second, we consider a reformulation of the DAG space, and propose a new framework for DAG structure learning by searching in this equivalent set of DAGs. A fast projection method is developed based on this continuous optimization approach without constraint. Experimental studies on benchmark datasets demonstrate that our method provides comparable accuracy but better efficiency, often by more than one order of magnitude. Last, we develop a variational autoencoder structure parameterized by a graph neural network architecture, which we coin DAG-GNN, to capture complex nonlinear mappings and data types. We demonstrate that the proposed method is capable of handling datasets with either continuous or discrete variables, and it learns more accurate graphs for nonlinearly generated samples.

References:

- [1] Zheng, X., Aragam, B., Ravikumar, P., Xing, E. P. DAGs with NO TEARS: Continuous Optimization for Structure Learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.
- [2] Yu, Y., Chen, J., Gao, T., Yu, M. DAG-GNN: DAG structure learning with graph neural networks. In *International Conference on Machine Learning* (pp. 7154-7163). PMLR. 2019
- [3] Wei, D., Gao, T., Yu, Y. DAGs with No Fears: A Closer Look at Continuous Optimization for Learning Bayesian Networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020
- [4] Yu, Y., Gao, T., Yin, N., Ji, Q. DAGs with No Curl: An Efficient DAG Structure Learning Approach. In *International Conference on Machine Learning* (pp. 12156-12166). PMLR., 2021