TECHNISCHE UNIVERSITÄT DARMSTADT

Thesis (B.Sc. / M.Sc.)
Copula-Guided Causal Simulators

Department 18
Electrical Engineering and Information Technology
Self-Organizing Systems Lab

Prof. Dr. Heinz Koeppl
Head of lab

Philipp Froehlich
Project supervisor

S3|06 206
Merckstrasse 25
64283 Darmstadt

philipp.froehlich@tu-darmstadt.de
https://www.bcs.tu-darmstadt.de

January 15, 2026

Background: Supervised causal learning (SCL) trains neural networks to recover causal graphs from data by generating large numbers of synthetic pairs $(G, \mathcal{D}_G)$, where G is a graph (typically a DAG) and $\mathcal{D}_G$ is a dataset sampled from a structural causal model (SCM) Markov to G.

Most existing SCL work uses *hand-crafted* SCM generators: graphs are sampled from simple random graph models, and mechanisms are linear or drawn from small families of toy nonlinear functions. This often yields unrealistic dependencies and poor out-of-distribution performance when applied to real data.

A promising alternative is to *learn local dependence structure from real datasets*, in a graph-agnostic way, and then plug this realistic "coupling library" into arbitrary graphs. Copulas and vine copulas are a natural tool here: they separate marginals from dependence and allow us to model flexible bivariate and conditional bivariate relationships. If we restrict the maximum node degree (e.g. to 3–4 parents), we can hope to build fast, stable, and realistic generators based on pairwise and conditional pair copulas.

The core idea of this thesis is therefore: *Estimate a collection of pairwise and conditional pair copulas from real data, and use them as local building blocks to generate synthetic datasets for arbitrary graphs (directed or undirected) with bounded degree.*

Objective: Design and implement a *graph-conditioned data generator* that:

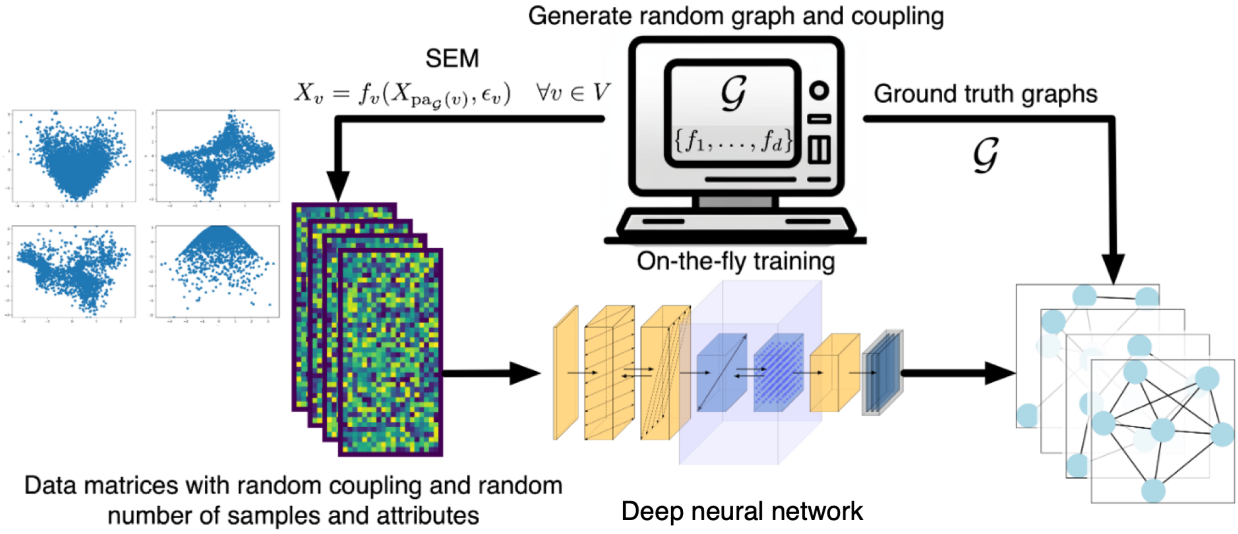
- learns realistic local dependence patterns from real-world datasets using pairwise and conditional pair copulas, and
- composes these into node-wise conditionals or clique potentials for user-specified graphs with bounded degree,

with the goal of producing stable, realistic synthetic data for training supervised causal learning methods.

Prerequisites:

- Background in electrical engineering, mathematics, physics, computer science, or a related field.
- Solid mathematical skills.
- Basic programming skills.

TECHNISCHE UNIVERSITÄT DARMSTADT



Supervised causal learning approach. We generate synthetic training data by first creating random ground-truth graphs. Then, observational data matrices are generated using SEMs given the ground-truth graphs. A mapping from observational data to the graph structure is learned via a deep neural network. Image from Froehlich and Koeppl [2024].

The Self-Organizing-Systems Lab combines practical applications of machine learning with strong theoretical foundations. The project is embedded in an interdisciplinary research team and offers the opportunity to contribute to publication-oriented work.

For further information, please contact Philipp Froehlich.

References

Philipp Froehlich and Heinz Koeppl. Graph structure inference with BAM: Neural dependency processing via bilinear attention. *Advances in Neural Information Processing Systems*, 37:128847–128885, 2024.