

MaRDI's Zenodo Community for Graphical Modeling and Causal Inference

Leopold Mareis^{1*}, Stephan Haug¹, Mathias Drton¹

Motivation

Topical Focus: **Graphical modeling** is a statistical field of research where relationships in the data distribution are represented by an associated graph.

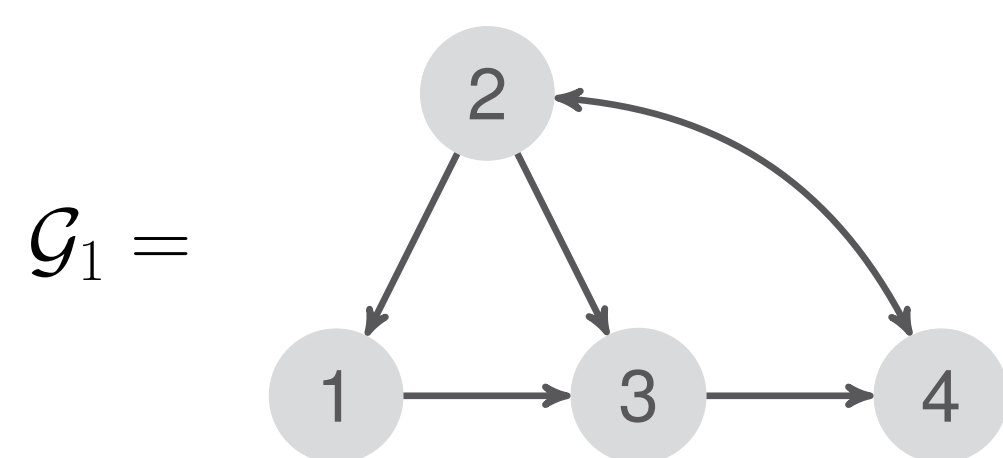


Figure 1 Acyclic directed mixed graph on X_1, X_2, X_3, X_4 implying $X_1 \perp\!\!\!\perp X_4 \mid X_{2,3}$. X_2 and X_4 are affected by an unobserved confounder.

💡 **Structure learning** aims to uncover underlying relationships by estimating a graph $\mathcal{G}$.

💡 **Causal inference** identifies and estimates related distributions that are derived by intervening on the given graph. Often the graph is interpreted as a causal mechanism generating the data distribution.

➔ **Useful for:** Network discovery, Efficient distribution estimation, Integration of domain knowledge, Robust reasoning, Explainable results, Confounder controlling, ...

Target Audience: Researchers and practitioners working on statistical methods for structure learning and causal inference in graphical models.

Challenge:

- ❓ Underlying ground truth graphs for benchmarking are often unknown.
- ❓ Interventional distributions can not be verified from an observational dataset.

Current dataset practice:

- ⚡ Simulation studies are non-standardized.
- ⚡ Rare ground truth datasets are not well documented and are held by independent sources.

Goal is to provide ...

- ➔ ... *long term storage* for well-curated topical datasets,
- ➔ ... *standardized* curation guidelines for harmonized documentation,
- ➔ ... *analysis notebooks* for benchmarking and reproducibility,

to enable **FAIR** [Findable, Accessible, Interoperable, and Reusable] aligned dataset availability to the community.

The GMCI Zenodo Community

The **Graphical Modeling and Causal Inference** (GMCI) community is a centralized hub for curated, topical **datasets** and accompanying analysis **notebooks**. Notebooks can present datasets or methodology.

Available metadata:

- Stable digital object identifier (DOI)
- Standardized description
- License information
- Literature references and referencing literature
- Notebooks rendered and presented on <https://st-mardi.quarto.pub/gmci/>

Application domains: Biology, Environmental Science, Economics, Medicine, Metrology, Psychology, Pharmacology, Political Science, ...

From contribution to discovery:

- Submit your contribution to the GMCI Community through the standardized submission procedure.
- Moderators assist with the procedure and create necessary embeddings.
- Researchers discover your contribution through related datasets, notebooks, and rich metadata.
- The GMCI community ensures long-term visibility and accessibility.



Zenodo GMCI Community



GMCI Website

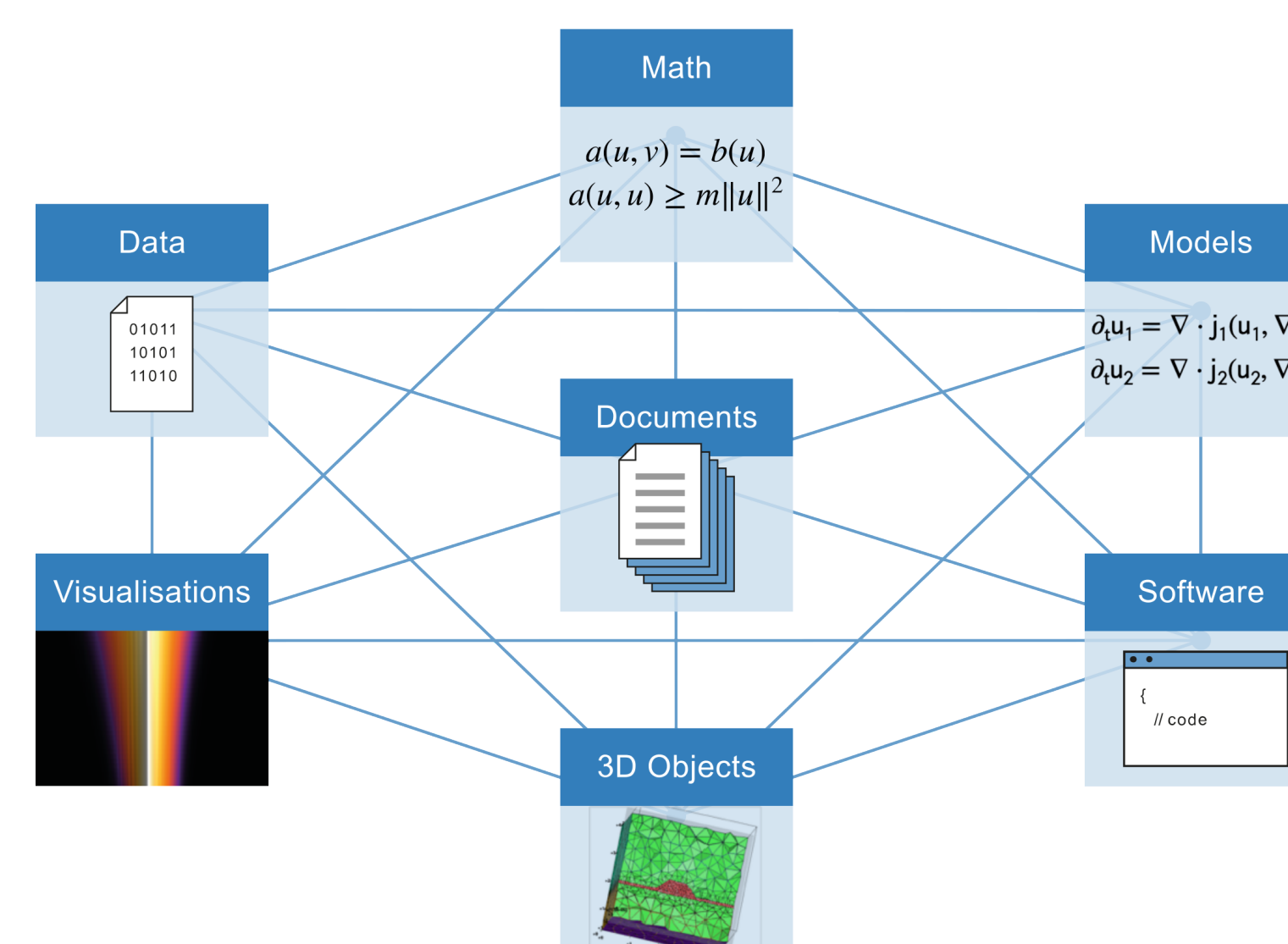


Submission Instructions

Contributions of insightful and high-quality datasets and notebooks from the broader academic community are encouraged.

Embedding in MaRDI and Conclusion

The **MaRDI** consortium within the NFDI aims to develop a robust Mathematical Research Data **Infrastructure** that would be useful within mathematics and other disciplines as well as non-scientific fields. MaRDI seeks to set **standards and confirmable workflows** for certified Mathematical Research Data and provide **services** to both the mathematical and wider scientific community.



The Mathematical Research Data Initiative (MaRDI) knowledge graph. Source: mardi4nfdi.de

➔ Through MaRDI's **knowledge graph** with currently over 5 million nodes, each GMCI submission becomes part of an evolving digital ecosystem of publications, software, and datasets.

💡 Why it matters. GMCI ...

- hosts open-access data with transparent metadata, supporting reproducible research.
- provides a framework for benchmarking in graphical modeling and causal inference.
- ensures long-term storage with DOIs to ensure citation and discoverability.
- builds a bridge between statistical methods and open science platforms.
- presents educational implementations as references for researchers.
- facilitates open collaboration opportunities within and beyond the statistical community.

! **Get Involved:** Explore and contribute to the community by scanning the QR code.



MaRDI Website



L. Mareis' LinkedIn

¹ Research Group Mathematical Statistics, Technical University of Munich, Germany.
* Contact: leopold.mareis@tum.de