# GMCI_MaRDI

MaRDI@TUM

2025-02-04

# Table of contents

# Graphical Modelling and Causal Inference

Within MaRDI, we aim to create a platform and public service to support and foster mathematical research. This comprises the development of good research practices, providing tools for establishing interconnections of research via unique digital tokens, and creating databases for immediate access to key resources of the mathematical research community.

We host the **Zenodo** community Graphical Modelling and Causal Inference. On this platform, we curate and present topical datasets and metadata. Exemplary statistical notebooks showcase advances in methodology and present new applications. The community supports content-moderation by our TA, and we will encourage and solicit submissions of datasets and notebooks by researchers from the broader academic community.

# 1 Zenodo Search

Searching in Zenodo is a logical 'OR' and not a logical 'AND'. We have thus implemented a search option generating URL for an appropriate Zenodo search with **javascript**. Select the desired attributes and click the button to open an external Zenodo page.

The search function is currently not working as planned. E.g. "Missing = Yes" does not return the correct

# 2 Custom Online Data Extraction

There are a lot of important areas for **Graphical Modelling and Causal Inference** on tabular data. Below is a list of **categories** and useful websites for custom data extraction:

---

## 2.1 Economic & Financial Data

Economic and financial data often involve causal relationships, such as how inflation affects employment or how policies impact market behavior.

**Possible datasets:**

- GDP, inflation, and trade data
- Stock market and financial transactions
- Household income and economic indicators

**Example sources:**

- World Bank Open Data
- OECD Data
- Federal Reserve Economic Data (FRED)

---

## 2.2 Social & Behavioral Data

Understanding human behavior, social interactions, and cultural influences can be crucial in causal inference.

**Possible datasets:**

- Social network analysis (Facebook, Twitter datasets)
- Behavioral economics studies
- Social mobility and inequality data

**Example sources:**

- General Social Survey (GSS)
- Pew Research Data
- European Social Survey (ESS)

---

## 2.3 Environmental & Ecological Data

Climate change, pollution, and biodiversity studies require causal modeling for impact assessment.

**Possible datasets:**

- Climate change and greenhouse gas emissions
- Biodiversity and species distribution
- Air and water quality

**Example sources:**

- NASA Earth Data
- Copernicus Climate Data
- Global Biodiversity Information Facility (GBIF)

---

## 2.4 Engineering & Sensor Data

Sensor-based datasets (IoT, industrial processes, robotics) can provide time-series data for causal discovery.

**Possible datasets:**

- Smart cities and traffic monitoring
- Industrial machine failure predictions
- Energy consumption and grid stability

**Example sources:**

- OpenML
- UCI Machine Learning Repository

---

## 2.5 Political & Legal Data

Causal models can analyze policy effects, election predictions, and legal case outcomes.

**Possible datasets:**

- Election results and voter behavior
- Government spending and policy impact
- Legal case rulings and sentencing patterns

**Example sources:**

- Congressional Voting Records
- Comparative Political Data Set (CPDS)
- World Justice Project Rule of Law Index

---

## 2.6 Transportation & Mobility Data

Causal analysis in urban planning, traffic patterns, and logistics is crucial for optimizing transport systems.

**Possible datasets:**

- Public transportation ridership data
- Road traffic and accident reports
- Flight delays and logistics performance

**Example sources:**

- U.S. Department of Transportation Open Data
- NYC Open Data (Taxi & Traffic Data)
- OpenStreetMap Traffic Data

---

## 2.7 Education Data

Educational outcomes are often influenced by multiple factors (socioeconomic status, school policies, etc.), making them ideal for causal modeling.

**Possible datasets:**

- Student performance and test scores
- School funding and academic achievement
- Higher education statistics

**Example sources:**

- National Center for Education Statistics (NCES)
- OECD Education at a Glance

---

## 2.8 Crime & Justice Data

Understanding the causal effects of policies, policing, and social factors on crime rates is crucial.

**Possible datasets:**

- Crime rates and law enforcement reports
- Prison population statistics
- Impact of social programs on crime reduction

**Example sources:**

- FBI Crime Data Explorer
- Bureau of Justice Statistics (BJS)
- UK Crime Data

---

## 2.9 Energy & Sustainability Data

Causal relationships between energy consumption, economic growth, and environmental policies are important.

**Possible datasets:**

- Renewable energy production
- Electricity grid performance
- Carbon footprint analysis

**Example sources:**

- International Energy Agency (IEA)
- Our World in Data – Energy
- U.S. Energy Information Administration (EIA)

## 2.10 Further Categories

We will add sources for the categories

*Gene expression data * health data * meterological data * psychological data * population data * biological data * pharmaceutical data

# 3 Gallery