Fujitsu Causal Discovery

A novel interactive platform for conditional causal discovery

J. Forde G. Mendez A. Okubo D. Quigley R. Sakamoto

Fujitsu Team, G-RIPS 2024

AM: Fabiana Ferracina
IMs: Hiroyuki Higuchi, Jorge Gutierrez

August 8, 2024

Overview

1. Project statement
2. Problem 1: graphs are inherently complex
   - Data visualization
   - Interactivity
   - Model builder
3. Problem 2: graphs are difficult to compare
   - Rate
   - Graph hierarchies
   - Statistical validity
4. Question: how to make decisions
5. Future directions
6. Appendices

Project statement

Causality: the statement such as 'A causes B'; we denote it as a relation $A \to B$. Hence, causal relations generate a **causal graph**

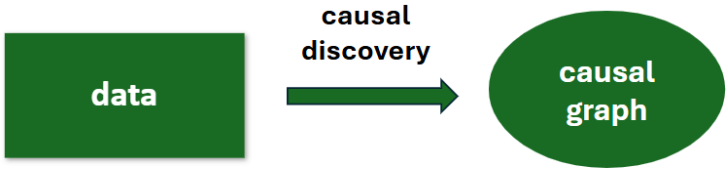
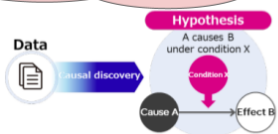


Figure: Causal discovery

Problem statement and proposed solution

I want to make high quality wine for Sendai during Tanabata, but I only trust advice from G-RIPS.

How can we best present a discoverable, convincing set of causal models?



Problem statement and proposed solution

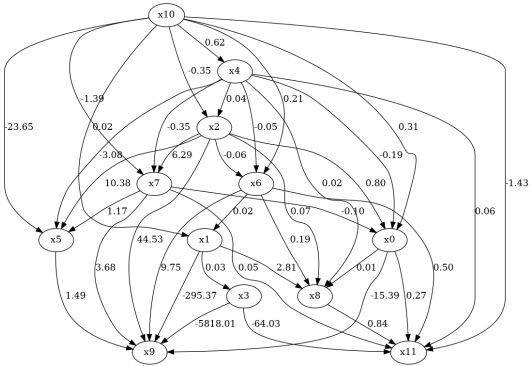


Figure: Problem 1: graphs are inherently complicated

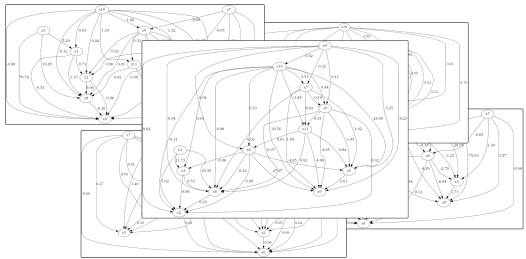


Figure: Problem 2: graphs are difficult to compare

CVD

- **C**onvincingness: Extent to which a suggested model/explanation matches (or exceeds) a user's expectation.
- **V**ariety: A set of unique 'equally good' explanations/models.
- **D**iscoverability: Extent to which there exist unexpected causal relations between features and outcomes.

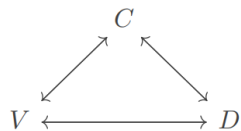
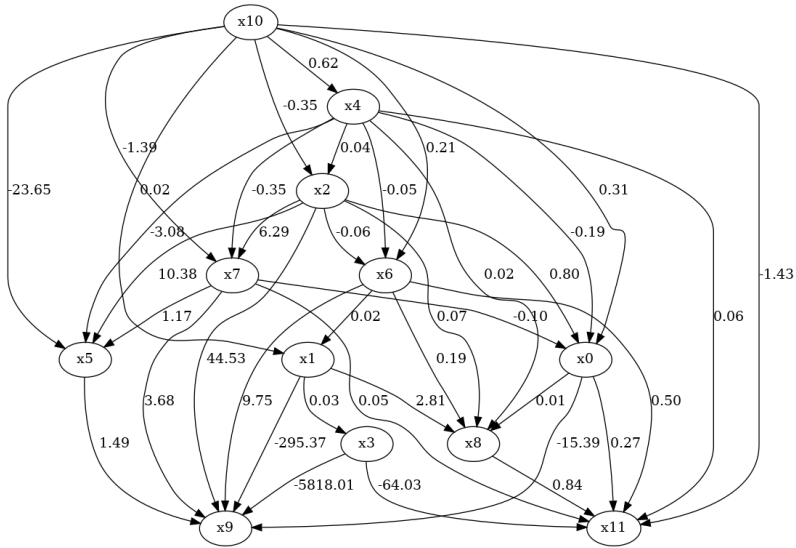


Figure: CVD triangle

Problem 1: graphs are inherently complex

Intra-graph analysis

1. Visualization
2. Interactivity
3. Model builder



Data visualization

Data visualization

My experience suggests a strong positive link between **free sulfur dioxide (FSD) concentration** and **quality.**

Does my data support this?

Hypothesis drives causal discovery journey.

Intra-Analysis	Observations:	Actions:
➔*Data Visualization*	- STD of density is extremely low - FSD is highly correlated to ~5 other features - FSD, others have skewed distributions	- Consider removing some features - **Investigate data imbalances** (red v. white?) and evaluate modifications

**STD=standard deviation*

Data visualization

View Data Description

	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality_bin
	6497	6497	6497	6497	6497	6497	6497
36109	30.525319378174544	115.7445744189626	0.9946966338309989	3.2185008465445586	0.5312682776666153	10.491800831149455	0.6330614129598277
45907	17.7493997720025	56.52185452263028	0.0029986730037190339	0.1607872021039876 4	0.14880587361449027	1.1927117488868981	0.4820066597331709
	1	6	0.98711	2.72	0.22	8	0
	17	77	0.99234	3.11	0.43	9.5	0
	29	118	0.99489	3.21	0.51	10.3	1
	41	156	0.99699	3.32	0.6	11.3	1
	289	440	1.03898	4.01	2	14.9	1

Pairwise Scatter Plots

Select variables to plot

Pairwise Correlation Heatmap

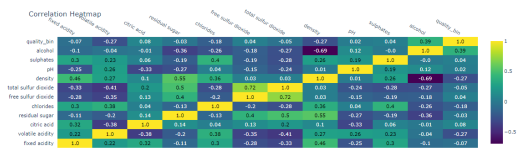


Figure: Density has a low STD and FSD shows significant collinearity.

Data Distribution

free sulfur dioxide

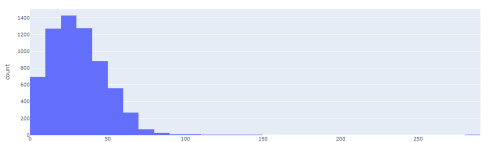
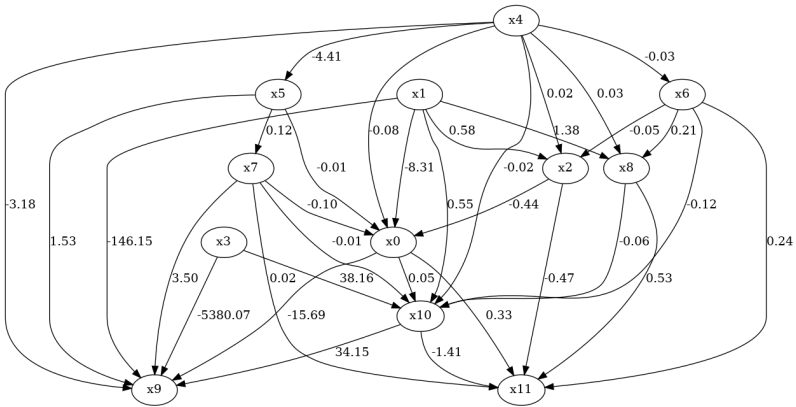


Figure: FSD distribution is slightly skewed.

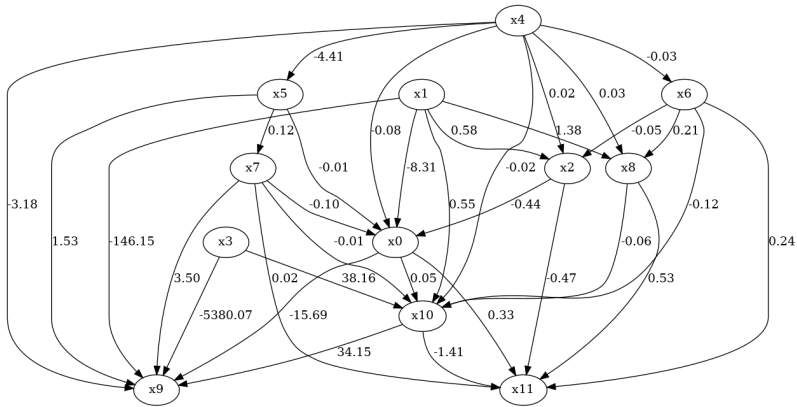
Interactivity

Interactivity



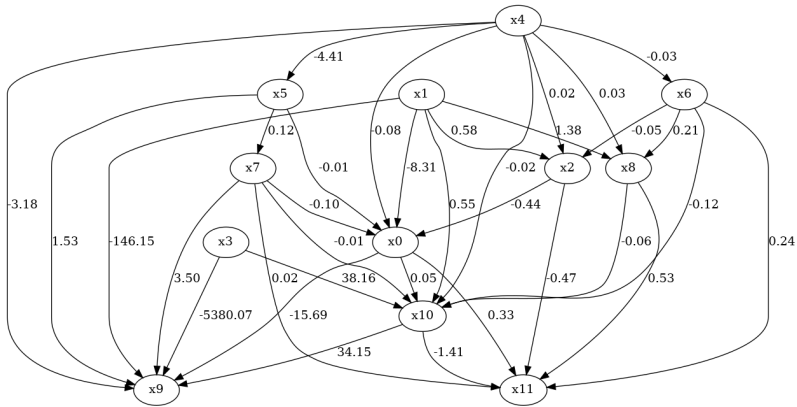
Interactivity

- Color gradient



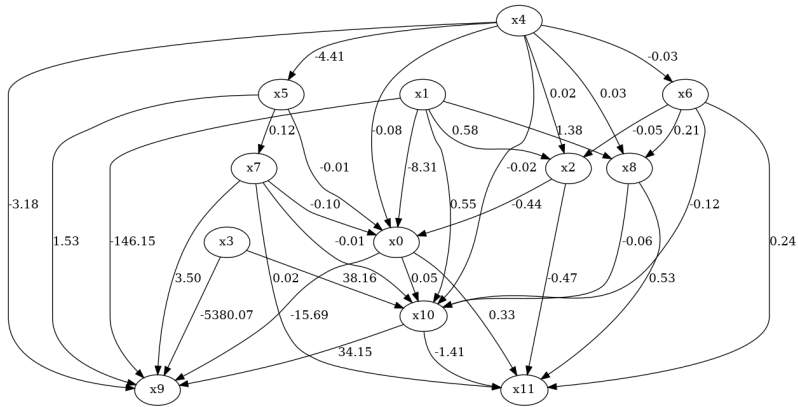
Interactivity

- Color gradient
- Edge styles

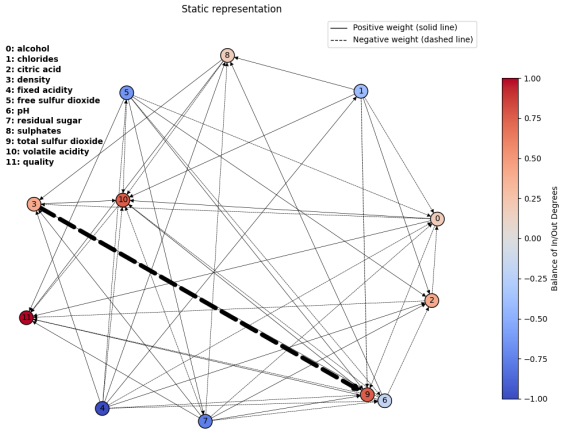


Interactivity

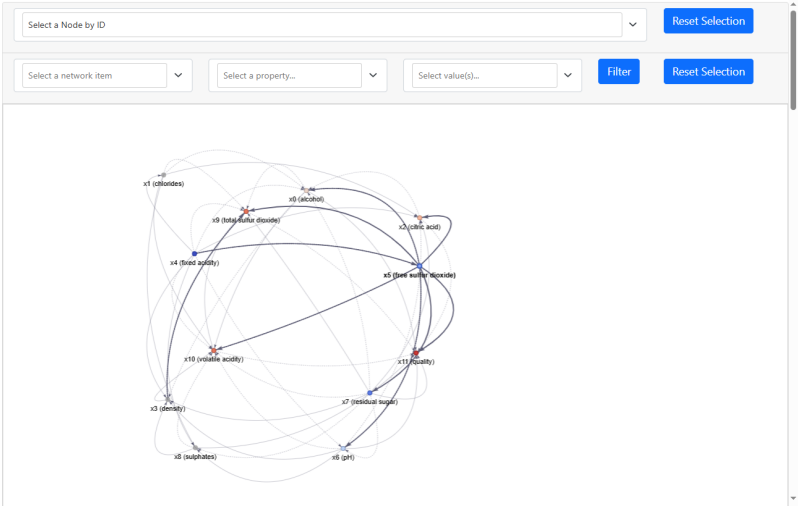
- Color gradient
- Edge styles
- Filter weights



Interactivity



Interactivity



Model builder

Intra-graph analysis

It appears my data has some limitations.
Can I see the effect on the causal model to
determine how significant any confounders are?

Explore causal models to find new, convincing trends.

Intra-Analysis
➔*Interactivity & Model Builder*

Observations:
- FSD is in an important condition
- FSD has weak negative effect on quality

Actions:
- **Consider other conditions** that lead to high quality wine to supplement hypotheses

Intra-graph analysis

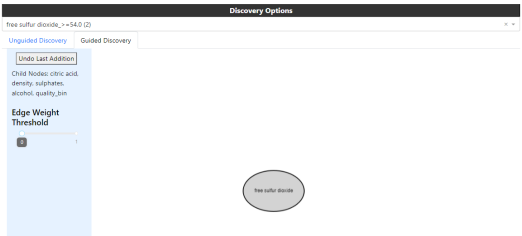


Figure: Selection of a model populates the parent node(s).

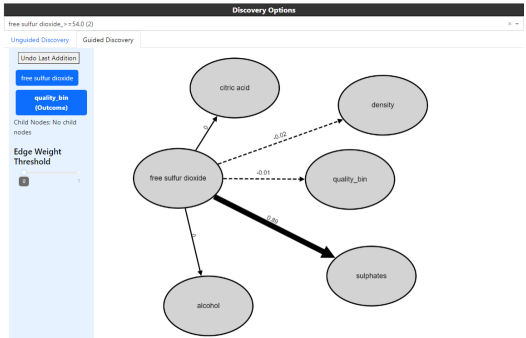
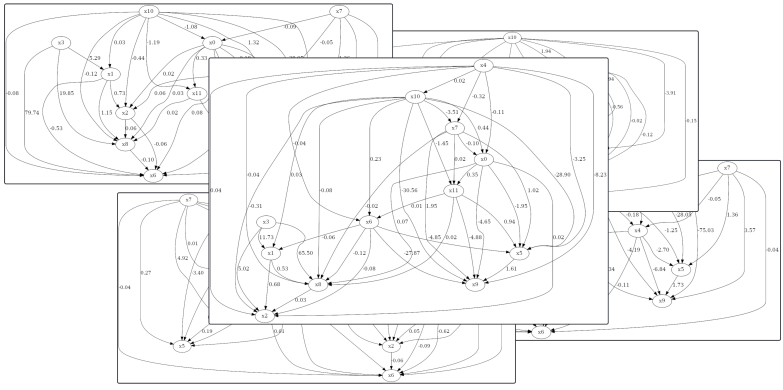


Figure: Selection of a node reveals next generation and edge information.

Problem 2: graphs are difficult to compare

Inter-graph analysis

1. Rate
2. Hierarchies
3. Statistical validity



Rate

Rate

Consider: *graph ordering by some indices* may reduce feeling overwhelmed by amount of information and various conclusions from causal graphs.

X: condition
X comes along with **rate** $R(X)$.

Roughly speaking, $R(X)$: the proportion of wines with high quality.

Rate

Summary

Selected Condition: free sulfur dioxide_>=54.0 | Instances meeting condition: 675 / 6497 (Ratio: 0.10)

Feature	Range	Average	Mode
fixed acidity	4.2 - 11.8	6.78	6.8
volatile acidity	0.105 - 0.735	0.28	0.26
citric acid	0.0 - 1.0	0.36	0.28
residual sugar	0.9 - 23.5	9.08	12.9
chlorides	0.009 - 0.346	0.05	0.045
free sulfur dioxide	54.0 - 289.0	64.32	54
total sulfur dioxide	80.0 - 440.0	180.57	178
density	0.9874 - 1.00369	1.00	0.9976
pH	2.86 - 3.76	3.17	3.16
sulphates	0.25 - 0.99	0.50	0.5
alcohol	8.0 - 13.5	9.85	9.4
quality_bin	0 - 1	0.55	1

Figure: FSD>=54.0 has a 10% level of support and 0.55 average quality.

Summary

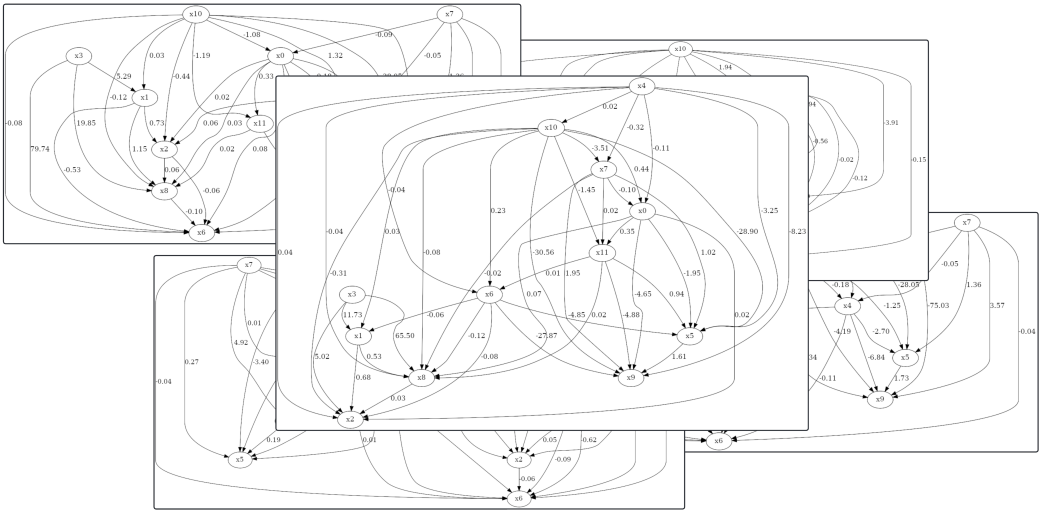
Selected Condition: residual sugar_<17.8 | Instances meeting condition: 6389 / 6497 (Ratio: 0.98)

Feature	Range	Average	Mode
fixed acidity	3.8 - 15.9	7.22	6.8
volatile acidity	0.08 - 1.58	0.34	0.28
citric acid	0.0 - 1.66	0.32	0.3
residual sugar	0.6 - 17.75	5.20	2
chlorides	0.009 - 0.611	0.06	0.044
free sulfur dioxide	1.0 - 289.0	30.36	29
total sulfur dioxide	6.0 - 440.0	114.93	111
density	0.98711 - 1.00369	0.99	0.9972
pH	2.72 - 4.01	3.22	3.16
sulphates	0.22 - 2.0	0.53	0.5
alcohol	8.0 - 14.9	10.51	9.5
quality_bin	0 - 1	0.64	1

Figure: RS<17.8 has a 98% level of support and 0.64 average quality. It may be worth investigating this model further.

Graph hierarchies

Graph hierarchies



Graph hierarchies

1. Microscopic - local structure (individual vertices)
   - hierarchical levels (HL)
   - influence centrality (IC)
2. Mesoscopic - groups or communities
   - hierarchical difference (HD)
3. Macroscopic - global structure
   - hierarchical incoherence (HI)
   - democracy coefficient (DC)

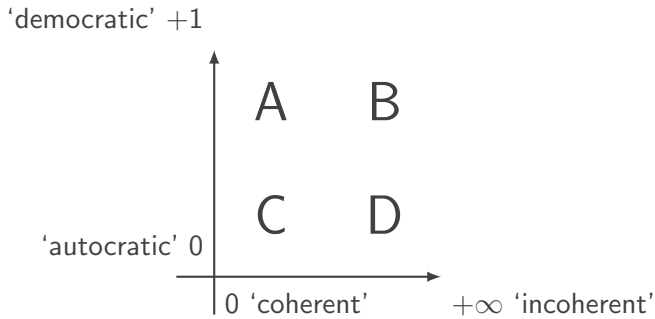
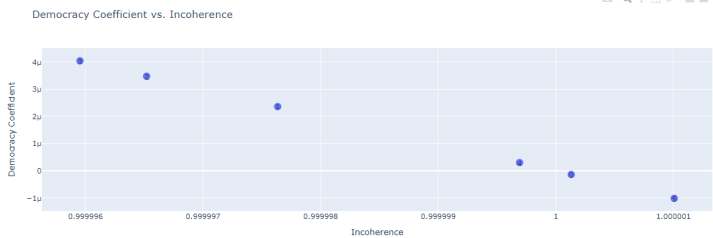


Figure: Democracy coefficient and coherence metrics

Graph hierarchies

View Metrics and Scatter Plot



Adjacency Matrix	Incoherence Score	Democracy Coefficient
adjacencyMatrix_0	1.000000731545235	-1.3154523492708847e-7
adjacencyMatrix_1	0.9999979349802251	0.000002365[illegible]1074908106
adjacencyMatrix_2	0.9999995200527225	0.000003479[illegible]2774516005
adjacencyMatrix_3	0.9999997538454203	0.000004046[illegible]4570717687
adjacencyMatrix_4	0.9999996026531833	3.0734681[illegible]6366725e-7
adjacencyMatrix_5	1.000001076017423	-0.000001007[illegible]7422644412

Graph hierarchies

If I should reconsider FSD as the focus of my study, **what should I look for next?**

New inquiries drive discovery from a variety of models.

Inter-Analysis

➔*Hierarchical and statistical metrics*

Observations:
- Highly incoherent, highly democratic model
- RMSEA and AIC are highly correlated
- FSD condition has a low level of support

Actions:
- FSD condition uses 10% of data vs. residual sugar_<17.8 has 98%; **consider higher rates for model selection**

**RMSEA=Root Mean Square Error; AIC=Akaike Information Criterion;*

Statistical validity

Statistical validity

Although there are many model fit evaluations, we mainly use

- comparative fit index (**CFI**)
- root mean square error of approximation (**RMSEA**)
- Akaike information criterion (**AIC**)

for simplicity.

Table: Conditions with statistical evaluations and graph hierarchical values

cond	CFI	RMSEA	AIC	HI (back)	DC (back)	HI (fwrd)	DC (fwrd)
1	0.929	0.162	16.4	2.21	0.000121	0.263	1.16e-05
2	0.0475	0.338	2.12	1.12	0.000323	0.264	0.000152
3	0.440	0.275	11.8	1.24	0.000123	0.649	0.000753

Is there any correlation between statistical and graph hierarchical values?

If there is some correlation, it could provide some support for the indicator.

Table: Correlation coefficients between model fit and graph hierarchical values.

	CFI	RMSEA	AIC	HI (back)	DC (back)	HI (fwrd)	DC (fwrd)
CFI	1.00						
RMSEA	-0.66	1.00					
AIC	0.64	-0.96	1.00				
HI (back)	-0.01	0.01	-0.02	1.00			
DC (back)	0.01	0.01	0.00	0.02	1.00		
HI (fwrd)	0.01	0.01	-0.01	0.03	0.02	1.00	
DC (fwrd)	-0.02	0.03	-0.04	0.06	0.18	0.09	1.00

$\rightarrow$ Provide model fit indicators after choosing a model according to graph hierarchical values.

- **Bootstrapping** is a resampling method, and calculates an estimator for each resample to obtain variances, confidence intervals, etc.

Here we calculate the probability of occurrence of edges and evaluate the confidence of edges. (0:alcohol, 3:density, 5:free sulfur dioxide, 11:quality)

Table: Example of paths from 0 to 11 and their probabilities.

path	effect[1]	probability
[0, 11]	0.344	1
[0, 3, 11]	-0.05	0.03
[0, 5, 11]	-0.004	0.02

[1] effect: median indirect effects occurring in the path

Statistical validity

For summary

- Statistical evaluations can help some who know statistics, but otherwise overwhelm.
- Providing them improves convincingness.
- Bootstrapping for each model would be beneficial, but not practical as it would take time to do and it is difficult to compare with other models.

Question: how to make decisions

Causality exploration

Now that I know which models I trust and why, **how can I begin developing the best wine in Tohoku?**

Strategy development follows causal discovery.

Decision-Making

➔*Actionability and Further Exploration*

Observations:
- FSD and CA are most actionable; I can find similar models with higher rates of support
- Consider restarting and controlling for skewed variables and collinearity

Actions:
- Reconsider the quality of the dataset and prior knowledge assumptions
- **Use the "recipe" provided by the model to predict wine quality** from other features

**CA=Citric Acid*

Video demo

Click for video demo

Future directions

Future directions

For prediction, adding, removing, or conditioning some variables to a model can sometimes produce discrepancies between the coefficients of the independent variables and the outcome, these discrepancies are known as biases. Identification is being able to express the correct causal path.

1. Conditioning can induce or remove biases from the causal graphs. $\Rightarrow$ V and D could be hindered by conditioning
2. Lack of standard methodology to tackle identification or bias issues on our graphs. $\Rightarrow$ C is greatly hindered if results are biased.

DAG identification basics

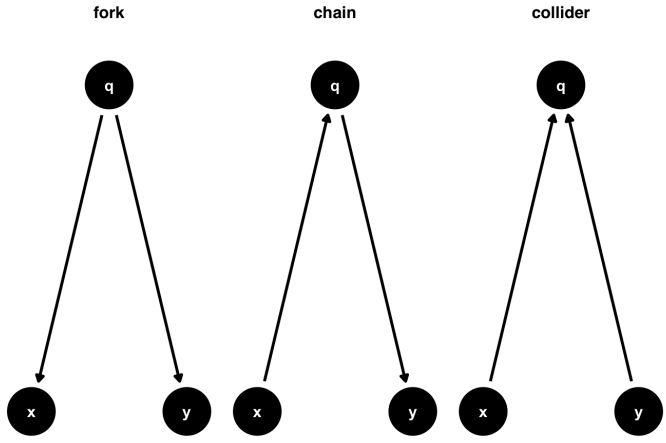


Figure: Three basic relationships in DAG

- **Selection Bias** classic example (Van der Weele, 2014): if this path is negative, the association between S and Y given L = 1 could become negative even if the direct causal effect is positive

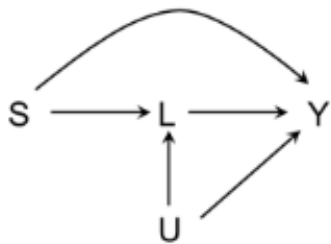


Figure: S is maternal smoking, L is low birth-weight, U is malnutrition, and Y is neonatal mortality.

Good control

Total Sulfur Dioxide $\geq 130 \wedge \leq 172$ and Sulphates $\geq 0.5 \wedge \leq 1.1$

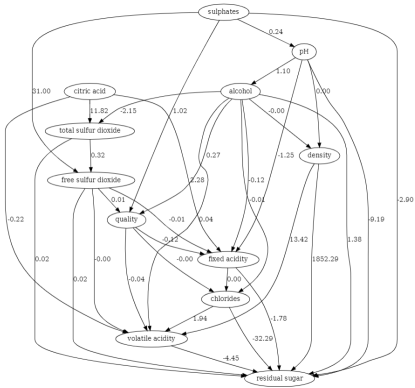


Figure: Four Condition Causal Graphs

Recommendations

- Implementation of identification
- Model supplementation
- Programming package implementation
- Bootstrapping

Thank you!

Thank you for your time and we look forward to answering your questions.

Appendices

Hierarchical metrics

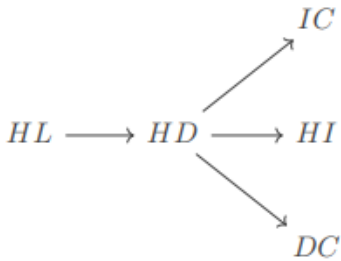


Figure: Relationships between hierarchical metrics

- A is the adjacency matrix
- $G = (V, E)$ is a graph with vertices $v \in V$ and edges $(i, j) \in E$ with associated weights w_{ij}
- $\mathrm{So}(G)$ is the set of all source vertices; $\mathrm{Sk}(G)$ is the set of all sink vertices

Table: Relevant definitions for hierarchical structures

weighted in-degree		weighted out-degree	
$d_i = \sum_j w_{ij}$	for vertex i	$\delta_i = \sum_j w_{ij}$	for vertex i
$d = (d_1, d_2, \ldots, d_n)$	is vector	$\delta = (\delta_1, \delta_2, \ldots, \delta_n)$	is vector
$L = \mathrm{diag}(d) - A$	is Laplacian	$\Lambda = \mathrm{diag}(\delta) - A$	is Laplacian

Hierarchical levels

Hierarchical levels (HL) grades vertices based on how far they are from sources $V \in \mathrm{So}(G)$ or sinks $V \in \mathrm{Sk}(G)$

Forward: $g := \mathrm{argmin}_{x \in T} \|x\|_2$, where $T = \mathrm{argmin}_{x \in \mathbb{R}^n} \left\|L^T x - d\right\|_2$

Backward: $\gamma := \mathrm{argmin}_{x \in S} \|x\|_2$, where $S = \mathrm{argmin}_{x \in \mathbb{R}^n} \left\|\Lambda^T x - \delta\right\|_2$

Difference: $h = \frac{1}{2}(g - \gamma)$

HL vector follows from the minimum Euclidean norm $\|x\|$ under the constraint that x minimizes $\left\|L^T x - D\right\|_2$ or $\left\|\Lambda^T x - \delta\right\|_2$.

Hierarchical differences

Hierarchical differences (HD) assign grades to edges via differences in HL.

Forward: $\mathrm{FHD}_{ij}(G) = \{g_j - g_i\}$

Backward: $\mathrm{BHD}_{ij}(G) = \{\gamma_i - \gamma_j\}$

HD evaluates the difference in HL between connected vertices, indicating directionality and magnitude of influence by directly comparing the HL of two connected vertices.

Influence centrality

Influence centrality (IC) measures the extent to which a vertex is an influencer of the graph by characterizing how significant the vertex is.

$$\text{Forward: } \eta_f(G,j) = 1 - \text{mean}\left(\text{FHD}\left(G,j\right)\right) = 1 - \frac{\sum_{i:(i,j)\in E} w_{ij}\cdot \text{FHD}_{ij}\left(G,j\right)}{\sum_{i:(i,j)\in E} w_{ij}}$$

$$\text{Backward: } \eta_b(G,j) = 1 - \text{mean}\left(\text{BHD}\left(G,j\right)\right) = 1 - \frac{\sum_{i:(i,j)\in E} w_{ij}\cdot \text{BHD}_{ij}\left(G,j\right)}{\sum_{i:(i,j)\in E} w_{ij}}$$

IC determines the degree to which a vertex acts as a source of influence (forward centrality) or resists influence (backward centrality) by checking each vertex's HL relative to those of its neighbors by measuring the weighted average of HD for a given vertex.

Hierarchical metrics - microscopic

A positive $\eta_f(G, j)$ indicates that j is an influencer, with HL higher than those from which it receives influence; a positive $\eta_b(G, j)$ indicates that j is resistant to influence, with HL lower than those it influences.

Democracy coefficient

Democracy coefficient (DC) measures the extent to which influencers are being influenced, and checks for relative "democractic" and "autocratic" behavior.

$$\text{Forward: } \eta_f(G) = 1 - \text{mean}\left(\text{FHD}\left(G\right)\right) = 1 - \frac{\sum_{(i,j)\in E} w_{ij} \cdot \text{FHD}_{ij}\left(G\right)}{\sum_{(i,j)\in E} w_{ij}}$$

$$\text{Backward: } \eta_b(G) = 1 - \text{mean}\left(\text{BHD}\left(G\right)\right) = 1 - \frac{\sum_{(i,j)\in E} w_{ij} \cdot \text{BHD}_{ij}\left(G\right)}{\sum_{(i,j)\in E} w_{ij}}$$

DC checks for equitability and uniformity of influence and control distribution among vertices by comparing their average HD across all edges to a baseline of zero.

Hierarchical metrics - macroscopic

If $\eta(G) \to +1$, then G is more "democractic", in that there is a more equitable distribution of influence (more variables have a say); if $\eta(G) \to 0$, then G is more "autocratic", in that there is a less equitable distribution of influence (fewer variables have a say).

Hierarchical incoherence

Hierarchical incoherence (HI) measures how neatly the graph structure is partitioned into levels.

$$\text{Forward: } \rho_f(G) = [\text{var}(\text{FHD}(G))]^{\frac{1}{2}} = \left[\frac{\sum_{(i,j)\in E} w_{ij}\cdot\left(\text{FHD}_{ij}(G) - \text{mean}(\text{FHD}(G))^2\right)}{\sum_{(i,j)\in E} w_{ij}}\right]^{\frac{1}{2}}$$

$$\text{Backward: } \rho_f(G) = [\text{var}(\text{FHD}(G))]^{\frac{1}{2}} = \left[\frac{\sum_{(i,j)\in E} w_{ij}\cdot\left(\text{BHD}_{ij}(G) - \text{mean}(\text{BHD}(G))^2\right)}{\sum_{(i,j)\in E} w_{ij}}\right]^{\frac{1}{2}}$$

HI checks for variability or inconsistency in HD across G by evaluating the spread or dispersion of HD from their mean value, indicating the extent of consistency of influence or control among nodes.

Hierarchical metrics - macroscopic

If $\rho(G) \to +\infty$, then G is more "incoherent", in that there are more disparities in hierarchical levels, and less uniform and equitable distribution of influence; if $\rho(G) \to 0$, then G is more "coherent", in that there are more uniform structure and minimal differences in hierarchical levels.

Table: Correlation coefficients with model fit evaluations

	CFI	GFI	AGFI	NFI	TLI	RMSEA	AIC	BIC	LogLik
CFI	1.00	1.00	1.00	1.00	1.00	-0.66	0.64	0.47	-0.64
GFI	1.00	1.00	1.00	1.00	1.00	-0.66	0.64	0.47	-0.64
AGFI	1.00	1.00	1.00	1.00	1.00	-0.66	0.64	0.47	-0.64
NFI	1.00	1.00	1.00	1.00	1.00	-0.66	0.64	0.47	-0.64
TLI	1.00	1.00	1.00	1.00	1.00	-0.66	0.64	0.47	-0.64
RMSEA	-0.66	-0.66	-0.66	-0.66	-0.66	1.00	-0.96	-0.71	0.97
AIC	0.64	0.64	0.64	0.64	0.64	-0.96	1.00	0.81	-0.99
BIC	0.47	0.47	0.47	0.47	0.47	-0.71	0.81	1.00	-0.76
LogLik	-0.64	-0.64	-0.64	-0.64	-0.64	0.97	-0.99	-0.76	1.00

The user

- Domain experts
- Users affected by model decisions
- Scientists/developers
- Managers/executive board
- Regulatory entities/agencies

Definition (**Agency model**)

An **agency model** is a quadruple of the form:

$$M_{agency} = \langle M, a, p, i \rangle$$

where M is the model being implemented or designed, a is an agent using or implementing a model M, p is the patient or audience with or for whom the agent is working and/or presenting, and i is the instrument to use or access the underlying model M.

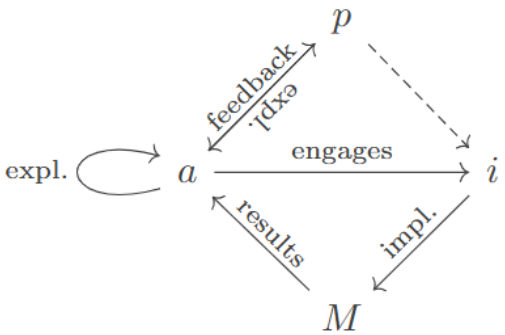


Figure: Agency model representation

Accessibility

1. **Perceivable** - Information must be available to users in ways they can perceive with their senses, using assistive technologies as necessary
2. **Operable** - Components must work with both keyboards and assistive devices
3. **Understandable** - Content needs to be clear and limit ambiguity
4. **Robust** - Documents must maximize compatibility with both current and future technologies like screen readers

Accessibility

Definition (Inclusive design)

Inclusive design is a design methodology that enables and draws on the full range of human diversity.

Definition (Accessibility)

Accessibility refers to the qualities that make an experience open to all; it is a professional discipline aimed at achieving an experience open to all.

Definition (Permanent disability)

Permanent disabilities are conditions that persist over time and significantly impact how individuals interact with (digital) content.

Definition (Temporary disability)

Temporary disabilities are impairments that arise from injuries and illnesses that affect users' otherwise abled abilities for a certain period of time.

Definition (Situational disability)

Situational disabilities are barriers or impedances that arise due to environmental or situational factors that affect users' otherwise abled abilities for a certain period of time.

Accessibility

- Contrast ratios for foreground and background elements
- Full navigation by a keyboard alone, integrated with assistive technologies (AT)
- Captions and tagged elements for multimedia and machine readability
- Functionality that uses multipoint or path-based gestures can be operated with a single pointer

- **Bias amplification**

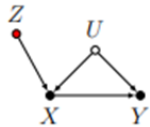


Figure: Controlling for Z will fail to deconfound the effect of X on Y

- **Over-control bias**

Figure: Controlling for Z will block the effect we want to estimate.

- **Blocking Backdoor path**

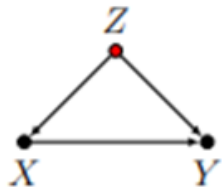


Figure: Z is a common cause of X and Y, blocking the backdoor path gives an unbiased estimate

- **Blocking Backdoor path of a mediator**

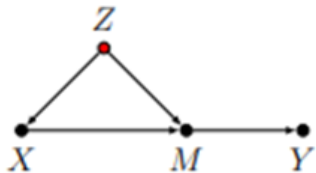


Figure: Common causes of X and a mediator also confound the effect of X on Y.

Prior knowledge

We create prior knowledge so that x0, x1 and x4 are sink variables.

The elements of prior knowledge matrix are defined as follows:

- `0` : x_i does not have a directed path to x_j
- `1` : x_i has a directed path to x_j
- `-1` : No prior knowledge is available to know if either of the two cases above (0 or 1) is true.

```
prior_knowledge = make_prior_knowledge(
    n_variables=6,
    sink_variables=[0, 1, 4],
)
print(prior_knowledge)
```

```
[[-1  0 -1 -1  0 -1]
 [ 0 -1 -1 -1  0 -1]
 [ 0  0 -1 -1  0 -1]
 [ 0  0 -1 -1  0 -1]
 [ 0  0 -1 -1 -1 -1]
 [ 0  0 -1 -1  0 -1]]
```

Figure: LiNGAM allows the user to set known relations prior to causal discovery (LiNGAM Documentation 1.9.0., July 2024).

```
# Draw a graph of prior knowledge
make_prior_knowledge_graph(prior_knowledge)
```

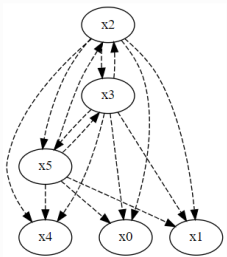


Figure: x0, x1, and x4 have been set as sink variables.

Prior knowledge

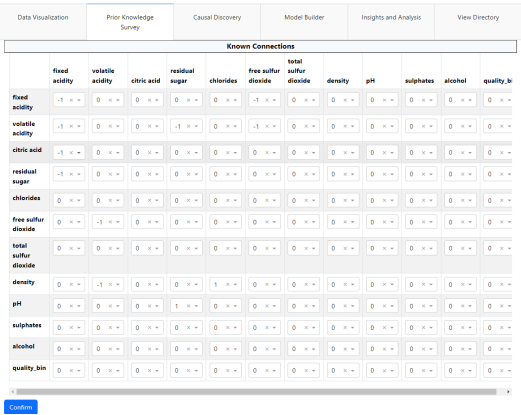


Figure: Prior knowledge is set in the interface via dropdown selections, based on the user's primary or secondary knowledge.

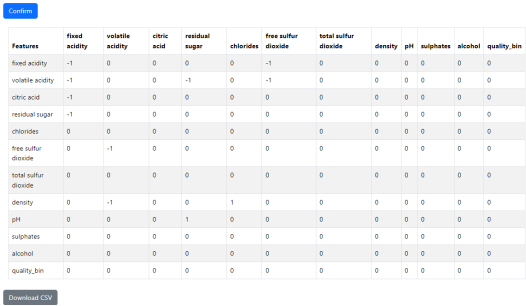


Figure: Relations can be exported to a .csv file for input into the CCD platform.

Future directions (interface)

- Complete What-If Analysis to test model sensitivity
- Enable full functionality for all condition combinations up to k=4
- Allow for analysis of bootstrapping for each model
- Integrate WL engine
- Investigate correlation between score and average rank for each feature
- Add menus (zoom, pan, etc.) and keyboard controls to increase accessibility in model explorer and model builder sections
- Add documentation, README, and tooltips throughout
- Integrate a Q & A bot or discovery assistant to guide the analysis
- Automatically apply suggestion to data and allow user to save results for comparison (e.g., effects if highly correlated variables are removed?)
- Incorporate new ordering or recommendation algorithms