Roots Screening: One Shot Identification of Root Nodes in LiNGAM Models

Causal Networks Seminar

Francesco Canonaco[1, 2, *]

December 10, 2025

[1]Models and Algorithms for Data and Text Mining (MADLab), University of Milano-Bicocca, Milan, Italy

[2]Minutia.AI, Singapore

[*]Corresponding author: francesco.canonaco@minutia.ai

Table of Contents

A Walk Through the LiNGAM Landscape

LiNGAM Models

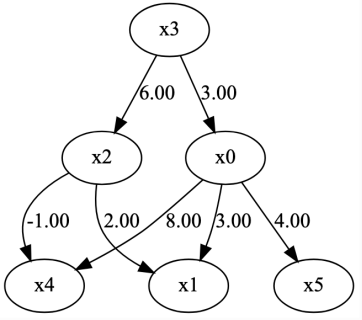
LiNGAM stands for **Li**near **N**on-**G**aussian **A**cyclic **M**odels. The basic LiNGAM model makes the following assumptions:

- Linearity
- Non-Gaussian continuous error variables (except at most one)
- Acyclicity
- No hidden common cause

LiNGAM Models

Python Code

```
import numpy as np
x3 = np.random.uniform(size=1000)
x0 = 3.0*x3 + np.random.uniform(size=1000)
x2 = 6.0*x3 + np.random.uniform(size=1000)
x1 = 3.0*x0 + 2.0*x2 + np.random.uniform(size=1000)
x5 = 4.0*x0 + np.random.uniform(size=1000)
x4 = 8.0*x0 - 1.0*x2 + np.random.uniform(size=1000)
```



LiNGAM Models

Observed variables are denoted by x_i and error variables by e_i and coefficients or connection strengths b_{ij}.

The error variables e_i are independent due to the assumption of no hidden common causes.

Given a causal order $k(i)(i = 1, \ldots, p)$. Then, a structural equation defining observed variables $x_i(i = 1, \ldots, p)$ can be obtained as follows:

$$x_i = \sum_{k(j)<k(i)} b_{ij} x_j + e_i.$$

LiNGAM Models: A Big Family

- ICALiNGAM
- DirectLiNGAM
- Multigroup DirectLiNGAM
- Nonlinear LiNGAM
- Longitudinal LiNGAM

And many more... For a complete list, visit https://lingam.readthedocs.io/en/latest/

DirectLiNGAM

DirectLiNGAM in a Nutshell

DirectLiNGAM algorithm can be divided in two parts:

- **Causal Order Estimation**:Estimation of the causal ordering, $k(i)(i = 1, \ldots, p)$.
- **Coefficients Estimation**: Once a causal ordering has been estimated, the coefficients estimation takes each observed variable as the objective variable and its *candidate parent variables* as the explanatory variables and perform a sparse regression analysis. [1]

DirectLiNGAM algorithm

DirectLiNGAM algorithm can be intuitively illustrade with a two-variable case.

$$x_1 = e_1$$
$$x_2 = b_{21}x_1 + e_2$$

where e_1 and e_2 are independent and following a non-Gaussian continuous distribution. The coefficient b_{21} is non-zero. In the example above x_1 is the cause and x_2 is the effect.

$$
\begin{aligned}
\mathrm{cov}(x_2, x_1) &= \mathbb{E}(x_2 x_1) - \mathbb{E}(x_2)\,\mathbb{E}(x_1) \\
&= \mathbb{E}\big[(b_{21}e_1 + e_2)e_1\big] - \mathbb{E}(b_{21}e_1 + e_2)\,\mathbb{E}(e_1) \\
&= b_{21}\mathbb{E}(e_1^2) + \mathbb{E}(e_1 e_2) - b_{21}\mathbb{E}(e_1)\mathbb{E}(e_1) - \mathbb{E}(e_1)\mathbb{E}(e_2) \\
&= b_{21}\mathbb{E}(e_1^2) + \mathbb{E}(e_1)\mathbb{E}(e_2) - b_{21}\mathbb{E}(e_1)^2 - \mathbb{E}(e_1)\mathbb{E}(e_2) \\
&= b_{21}\big\{\mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2\big\} \qquad (3.13)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}(x_1) &= \mathbb{E}(x_1^2) - \mathbb{E}(x_1)^2 \\
&= \mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2 \qquad (3.14)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{cov}(x_2, x_1) &= \mathbb{E}(x_2 x_1) - \mathbb{E}(x_2)\,\mathbb{E}(x_1) \\
&= \mathbb{E}\big[(b_{21} e_1 + e_2) e_1\big] - \mathbb{E}(b_{21} e_1 + e_2)\,\mathbb{E}(e_1) \\
&= b_{21}\mathbb{E}(e_1^2) + \mathbb{E}(e_1 e_2) - b_{21}\mathbb{E}(e_1)\mathbb{E}(e_1) - \mathbb{E}(e_1)\mathbb{E}(e_2) \\
&= b_{21}\mathbb{E}(e_1^2) + \mathbb{E}(e_1)\mathbb{E}(e_2) - b_{21}\mathbb{E}(e_1)^2 - \mathbb{E}(e_1)\mathbb{E}(e_2) \\
&= b_{21}\big\{\mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2\big\} \qquad (3.13)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}(x_1) &= \mathbb{E}(x_1^2) - \mathbb{E}(x_1)^2 \\
&= \mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2 \qquad (3.14)
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{cov}(x_2, x_1) &= \mathbb{E}(x_2 x_1) - \mathbb{E}(x_2)\,\mathbb{E}(x_1) \\
&= \mathbb{E}\big[(b_{21} e_1 + e_2) e_1\big] - \mathbb{E}(b_{21} e_1 + e_2)\,\mathbb{E}(e_1) \\
&= b_{21}\mathbb{E}(e_1^2) + \mathbb{E}(e_1 e_2) - b_{21}\mathbb{E}(e_1)\mathbb{E}(e_1) - \mathbb{E}(e_1)\mathbb{E}(e_2) \\
&= b_{21}\mathbb{E}(e_1^2) + \mathbb{E}(e_1)\mathbb{E}(e_2) - b_{21}\mathbb{E}(e_1)^2 - \mathbb{E}(e_1)\mathbb{E}(e_2) \\
&= b_{21}\big\{\mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2\big\} \qquad (3.13)
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{var}(x_1) &= \mathbb{E}(x_1^2) - \mathbb{E}(x_1)^2 \\
&= \mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2 \qquad (3.14)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{cov}(x_2, x_1) &= \mathbb{E}(x_2 x_1) - \mathbb{E}(x_2)\,\mathbb{E}(x_1) \\
&= \mathbb{E}\big[(b_{21} e_1 + e_2) e_1\big] - \mathbb{E}(b_{21} e_1 + e_2)\,\mathbb{E}(e_1) \\
&= b_{21}\mathbb{E}(e_1^2) + \mathbb{E}(e_1 e_2) - b_{21}\mathbb{E}(e_1)\mathbb{E}(e_1) - \mathbb{E}(e_1)\mathbb{E}(e_2) \\
&= b_{21}\mathbb{E}(e_1^2) + \mathbb{E}(e_1)\mathbb{E}(e_2) - b_{21}\mathbb{E}(e_1)^2 - \mathbb{E}(e_1)\mathbb{E}(e_2) \\
&= b_{21}\big\{\mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2\big\} \qquad (3.13)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}(x_1) &= \mathbb{E}(x_1^2) - \mathbb{E}(x_1)^2 \\
&= \mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2 \qquad (3.14)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{cov}(x_2, x_1) &= \mathbb{E}(x_2 x_1) - \mathbb{E}(x_2)\,\mathbb{E}(x_1) \\
&= \mathbb{E}\big[(b_{21}e_1 + e_2)e_1\big] - \mathbb{E}(b_{21}e_1 + e_2)\,\mathbb{E}(e_1) \\
&= b_{21}\mathbb{E}(e_1^2) + \mathbb{E}(e_1 e_2) - b_{21}\mathbb{E}(e_1)\mathbb{E}(e_1) - \mathbb{E}(e_1)\mathbb{E}(e_2) \\
&= b_{21}\mathbb{E}(e_1^2) + \mathbb{E}(e_1)\mathbb{E}(e_2) - b_{21}\mathbb{E}(e_1)^2 - \mathbb{E}(e_1)\mathbb{E}(e_2) \\
&= b_{21}\big\{\mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2\big\} \qquad (3.13)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}(x_1) &= \mathbb{E}(x_1^2) - \mathbb{E}(x_1)^2 \\
&= \mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2 \qquad (3.14)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{cov}(x_2, x_1) &= \mathbb{E}(x_2 x_1) - \mathbb{E}(x_2)\,\mathbb{E}(x_1) \\
&= \mathbb{E}\big[(b_{21} e_1 + e_2) e_1\big] - \mathbb{E}(b_{21} e_1 + e_2)\,\mathbb{E}(e_1) \\
&= b_{21}\mathbb{E}(e_1^2) + \mathbb{E}(e_1 e_2) - b_{21}\mathbb{E}(e_1)\mathbb{E}(e_1) - \mathbb{E}(e_1)\mathbb{E}(e_2) \\
&= b_{21}\mathbb{E}(e_1^2) + \mathbb{E}(e_1)\mathbb{E}(e_2) - b_{21}\mathbb{E}(e_1)^2 - \mathbb{E}(e_1)\mathbb{E}(e_2) \\
&= b_{21}\big\{\mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2\big\} \qquad (3.13)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}(x_1) &= \mathbb{E}(x_1^2) - \mathbb{E}(x_1)^2 \\
&= \mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2 \qquad (3.14)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{cov}(x_2, x_1) &= \mathbb{E}(x_2 x_1) - \mathbb{E}(x_2)\,\mathbb{E}(x_1) \\
&= \mathbb{E}\big[(b_{21} e_1 + e_2) e_1\big] - \mathbb{E}(b_{21} e_1 + e_2)\,\mathbb{E}(e_1) \\
&= b_{21}\mathbb{E}(e_1^2) + \mathbb{E}(e_1 e_2) - b_{21}\mathbb{E}(e_1)\mathbb{E}(e_1) - \mathbb{E}(e_1)\mathbb{E}(e_2) \\
&= b_{21}\mathbb{E}(e_1^2) + \mathbb{E}(e_1)\mathbb{E}(e_2) - b_{21}\mathbb{E}(e_1)^2 - \mathbb{E}(e_1)\mathbb{E}(e_2) \\
&= b_{21}\big\{\mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2\big\} \qquad (3.13)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}(x_1) &= \mathbb{E}(x_1^2) - \mathbb{E}(x_1)^2 \\
&= \mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2 \qquad (3.14)
\end{aligned}
$$

The formula 3.13 employs the property that the expectation of the product of two random variables is equal to the product of their expecations iff they are independent as in $\mathbb{E}(e_1 e_2) = \mathbb{E}(e_1)\mathbb{E}(e_2)$. Using equation 3.13 and 3.14 we can estimate b_{21}:

$$\frac{\mathrm{cov}(x_2, x_1)}{\mathrm{var}(x_1)} = \frac{b_{21}\{\mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2\}}{\mathbb{E}(e_1^2) - \mathbb{E}(e_1)^2} \tag{3.15}$$

$$= b_{21}. \tag{3.16}$$

Let's go back to our two-variable case:

$$
\begin{aligned}
x_1 &= e_1 \\
x_2 &= b_{21} x_1 + e_2
\end{aligned}
$$

The residual $r_2^{(1)}$ becomes:

$$
\begin{aligned}
r_2^{(1)} &= x_2 - \frac{\mathrm{cov}(x_2, x_1)}{\mathrm{var}(x_1)} x_1 \\
&= x_2 - b_{21} x_1 \\
&= e_2.
\end{aligned}
\tag{3.17}
$$

The lower (upper right) index of the residual represents the objective (explanatory) variable number.

Let's now do a linear regression analysis with x_1 using x_2 as the objective (explanatory) variable. Unlike the previous example we are inverting the order, the residual can be expressed as:

$$
\begin{aligned}
r_1^{(2)} &= x_1 - \frac{\mathrm{cov}(x_1, x_2)}{\mathrm{var}(x_2)} x_2 \\
&= x_1 - \frac{\mathrm{cov}(x_1, x_2)}{\mathrm{var}(x_2)} (b_{21} x_1 + e_2) \\
&= \left\{ 1 - \frac{b_{21}\,\mathrm{cov}(x_1, x_2)}{\mathrm{var}(x_2)} \right\} x_1 - \frac{\mathrm{cov}(x_1, x_2)}{\mathrm{var}(x_2)} e_2 \\
&= \left\{ 1 - \frac{b_{21}\,\mathrm{cov}(x_1, x_2)}{\mathrm{var}(x_2)} \right\} e_1 - \frac{b_{21}\,\mathrm{var}(x_1)}{\mathrm{var}(x_2)} e_2
\end{aligned}
\tag{3.18}
$$

From the model assumption b_{21} is non-zero, therefore the residual $r_1^{(2)}$ contains the error variable e_2. The variable x_2 also contains the error variable e_2. Given that they both contain the error variable e_2 the residual $r_1^{(2)}$ and the variable x_2 are dependent.

Conclusion: independece tells us something about the causal ordering.

Theorem 2.1

Assume a LiNGAM model. Assume the number of observations is sufficiently large that the estimation error can be ignored. Denote the residuals from the linear regression of x_i on x_j by $r_i^{(j)}$, i.e.,

$$r_i^{(j)} = x_i - \frac{\mathrm{cov}(x_i, x_j)}{\mathrm{var}(x_j)} x_j, \qquad (i, j = 1, \ldots, p;\ i \neq j). \tag{1}$$

The observed variable x_j may then be put in the first place of a causal ordering if and only if it is independent of all its residuals $r_i^{(j)}$ $(i = 1, \ldots, p;\ i \neq j)$.

DirectLiNGAM algorithm

To assess independence between x_j and its residuals $r_i^{(j)}$, DirectLiNGAM uses the mutual information:

$$I(x_j, r_i^{(j)}) = H(x_j) + H(r_i^{(j)}) - H([x_j, r_i^{(j)}]^T)$$
$$I(x_i, r_j^{(i)}) = H(x_i) + H(r_j^{(i)}) - H([x_i, r_j^{(i)}]^T)$$

The estimated value will never be exaclty zero even when the explanatory variable is independent of the residuals. Examining which variables (x_i,x_j) give a smaller value of the mutual information with the residual, when used as the explanatory variable, is common.

DirectLiNGAM algorithm

Practical issue: estimating the joint entropy $H([x_j, r_i^{(j)}]^T)$ is difficult and error-prone.

Solution (Hyvärinen & Smith, 2013):

- Standardize variables x_i, x_j.
- Compute only 1-dimensional entropies using the difference of the mutual informations.

This leads to an efficient difference measure:

$$m(x_i, x_j) = I(\tilde{x}_j, \tilde{r}_i^{(j)}) - I(\tilde{x}_i, \tilde{r}_j^{(i)})$$

Ordering rule:

$$m(x_i, x_j) < 0 \;\Rightarrow\; x_j \text{ earlier than } x_i.$$
$$m(x_i, x_j) > 0 \;\Rightarrow\; x_i \text{ earlier than } x_j.$$

Since now we have the difference of the mutual information we can use:

Ordering rule:

$$m(x_i, x_j) < 0 \Rightarrow x_j \text{ earlier than } x_i.$$

$$m(x_i, x_j) > 0 \Rightarrow x_i \text{ earlier than } x_j.$$

The measure $m(x_i, x_j)$ compares the mutual information in the two regression directions. For an explanatory variable x_i, summarize these values over all objective variables x_j $(j \neq i)$ as:

$$M(x_i; U) = -\sum_{j \in U} \min\big(0,\, m(x_i, x_j)\big)^2,$$

where U is the set of variable indices to be analyzed and $\min(0, m(x_i, x_j))$ denotes the smaller of 0 and $m(x_i, x_j)$.

A variable x_i that maximizes $M(x_i; U)$ is estimated to be a candidate for the first position in a causal ordering:

- If $m(x_i, x_j) > 0$, then x_i is earlier in the causal order than x_j, and $\min(0, m(x_i, x_j))^2 = 0$.
- If $m(x_i, x_j) = 0$, x_i and x_j are not causally ordered.
- If $m(x_i, x_j) < 0$, then x_j is earlier than x_i, giving a positive contribution $\min(0, m(x_i, x_j))^2$.

DirectLiNGAM algorithm

The causal ordering is built sequentially using $M(x_i; U)$. At each step, the variable with the largest score is selected as the next in the ordering. Its influence is removed by replacing all remaining variables with their residuals, and the selection process continues on the reduced dataset.

Improving DirectLiNGAM

Improving DirectLiNGAM via Roots Screening

Root variables can be placed first in the causal ordering because they are exogenous. When evaluating $M(x_i; U)$, only one root will be selected as the first candidate, but in principle any root could occupy that position.

Hypothesis: Root variables yield $M(x_i; U)$ values that cluster near zero, in contrast with non-root nodes, whose scores deviate more substantially.

Python Code Example

```
x1 = np.random.uniform(size = 10_000)
x2 = np.random.uniform(size = 10_000)
x4 = np.random.uniform(size = 10_000)
x0 = 3.0 * x1 + 4.0 * x2 + np.random.uniform(size = 10_000)
x3 = 6.0 * x0 + 4.0 * x4 + np.random.uniform(size = 10_000)
```

Improving DirectLiNGAM via Roots Screening

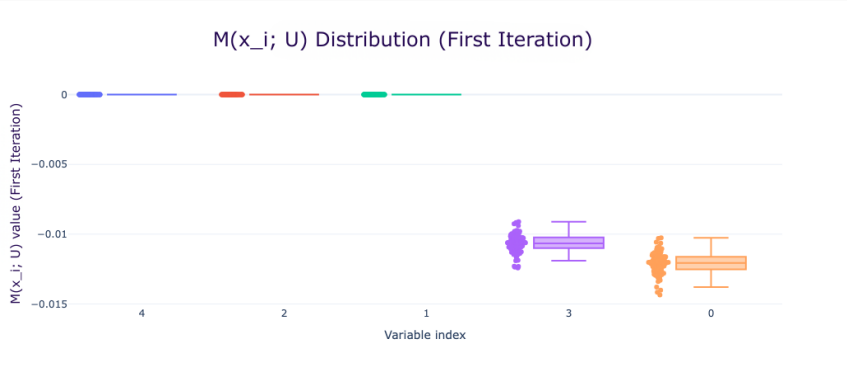


Figure 1: $M(x_i;U)$ values for each variable over 100 runs.

Improving DirectLiNGAM via Roots Screening

Goal: Given a sequence of positive values where each element is represented by $|M(x_i; U)|$, identify the position of the *largest multiplicative jump*.

Algorithm Overview

- Ensure the vector containing $|M(x_i; U)|$ has at least two elements.
- Add a tiny constant ε to avoid numerical issues when values are zero.
- Compute consecutive differences in $\log_{10}$ scale:

$$\Delta_i = \log_{10}(x_{i+1} + \varepsilon) - \log_{10}(x_i + \varepsilon).$$

- Ignore jumps starting from values too close to zero.
- Among the valid jumps, select the index with the largest positive log jump.

Output: the position of the most significant jump, i.e. the index where the ratio $\frac{x_{i+1}}{x_i}$ is maximal.

Parameter Configuration for Experiments

Experimental Grid

- **P_LIST**: $\{100, 300, 500\}$ – Number of variables in the network.
- **S_LIST**: $\{0.15\}$ – Sparsity level (probability of an edge).
- **N_LIST**: $\{10\,000\}$ – Number of samples.
- **ROOT_FRAC_LIST**: $\{0.1,\ 0.2,\ 0.3\}$ – Fraction of root nodes in the simulated DAG.
- **RUNS**: 20 – Number of experiments for each configuration.

Improving DirectLiNGAM via Roots Screening

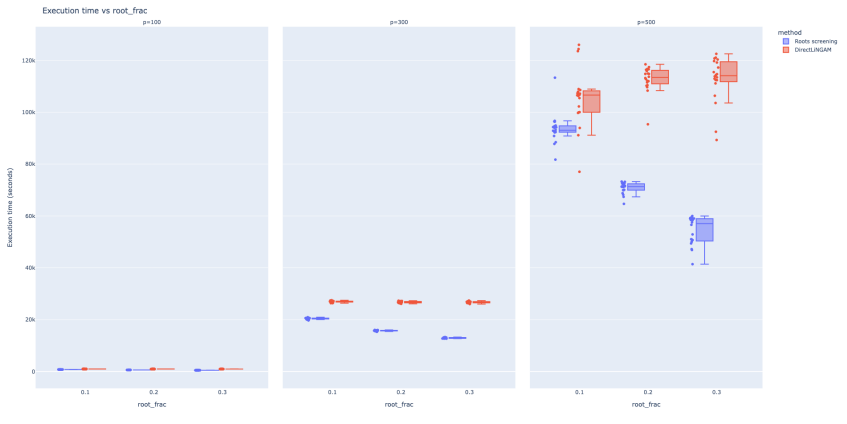


Figure 2: Execution time for each configuration over 20 runs.

Conclusion & Next Steps

In conclusion the root screening can be used to reduce the execution time of DirectLiNGAM in a high-variable number setting typical in microbiome data. Extend experiments to scale-free Networks such as Barabasi-Albert using LiNGAM as a data generation process. Assess structural hamming distance, mean squared error, precision and recall.

Thank You for your Attention!

MADLab

The Models and Algorithms for Data & Text Mining Laboratory is a research lab at University of Milano-Bicocca focused on Causal Networks, Bayesian Networks and Continuous-Time Bayesian Networks applied to Healthcare and Medicine.

More at: https://mad.disco.unimib.it/

References i

[1] S. Shimizu, *Statistical causal discovery: LiNGAM approach*. Springer, 2022.