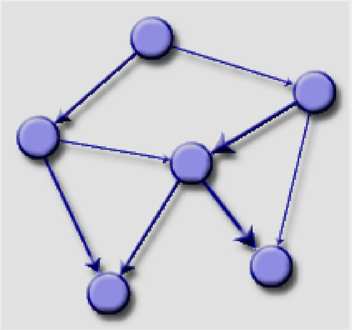


Created by S.S. in Jan 2008

# Discovery of Linear Acyclic Models Using Independent Component Analysis

Shohei Shimizu, Patrik Hoyer,
Aapo Hyvarinen and Antti Kerminen

LiNGAM homepage: http://www.cs.helsinki.fi/group/neuroinf/lingam/

# Independent Component Analysis

(Hyvarinen et al., 2001)

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

$$m \times 1 \qquad m \times n \quad n \times 1$$

- A is unknown, cov(s)=I
  - Typically, A is square (m=n)
- $s_i$ are assumed to be non-Gaussian and mutually independent
- Non-Gaussian and independence are the key assumptions in ICA
- Estimable including the rotation (Comon, 1994)

# Linear acyclic models

(Bollen, 1989; Pearl, 2000; Spirtes et al., 2000)

- Continuous variables
- Directed acyclic graph (DAG)
- The value assigned to each variable is a linear combination of those previously assigned, plus a disturbance term ei
- Disturbances (errors) are independent and have non-zero variances

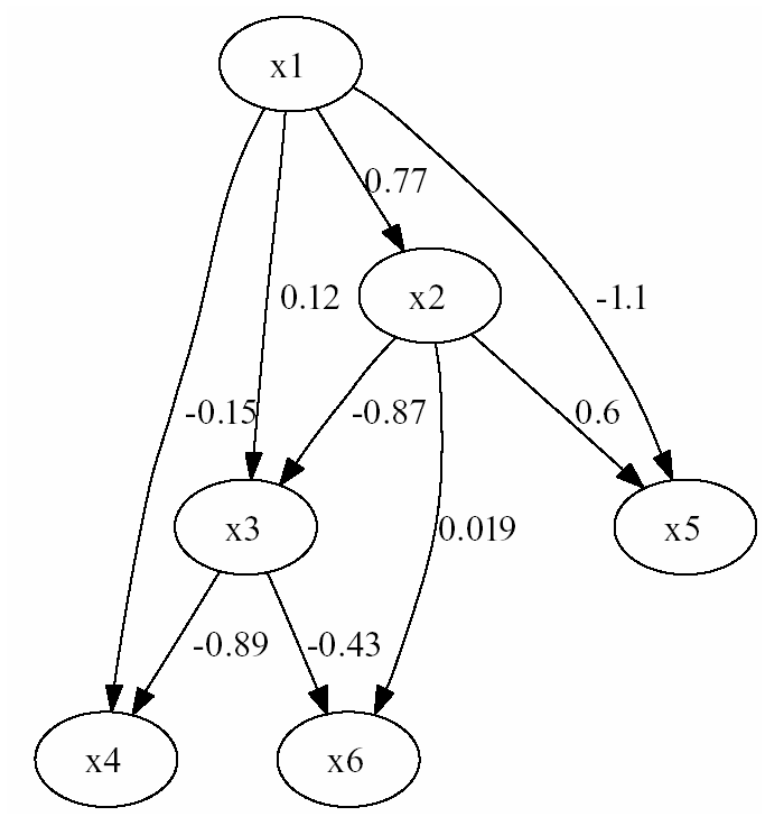
$$x_i = \sum_{k(j)<k(i)} b_{ij} x_j + e_i \quad \text{or} \quad \mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e}$$

# Our goal

- We know
  - Data X is generated by $\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e}$
- We do NOT know
  - Connection strengths: B
  - Order: k(i)
  - Disturbances: ei
- What we observe is data X only

- **Goal**
  - Estimate B and k using data X only!

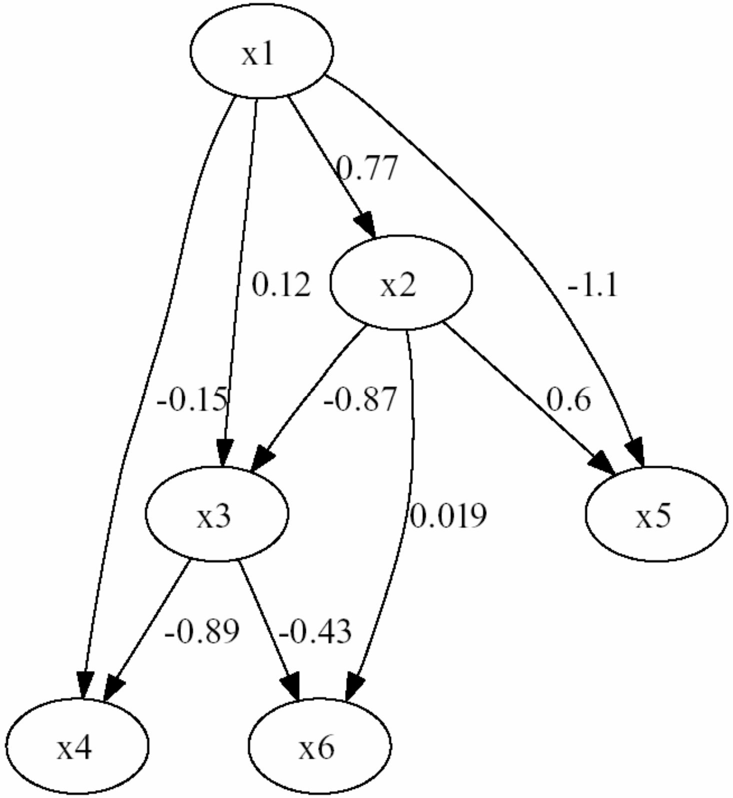
# Can we recover the original network?



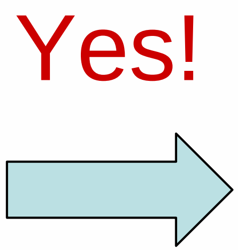


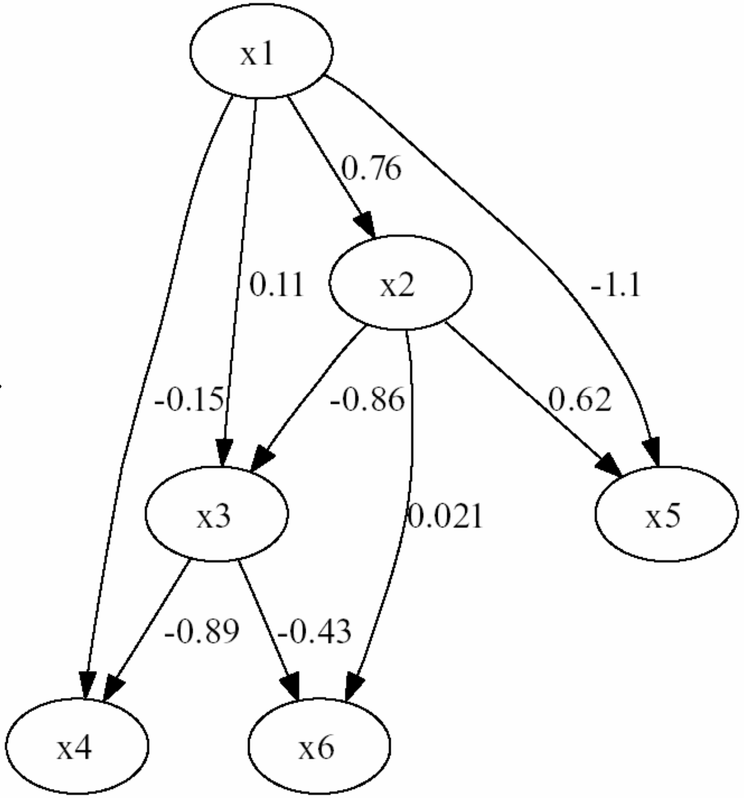
**Original network**

# Can we recover the original network using ICA?




**Original network**

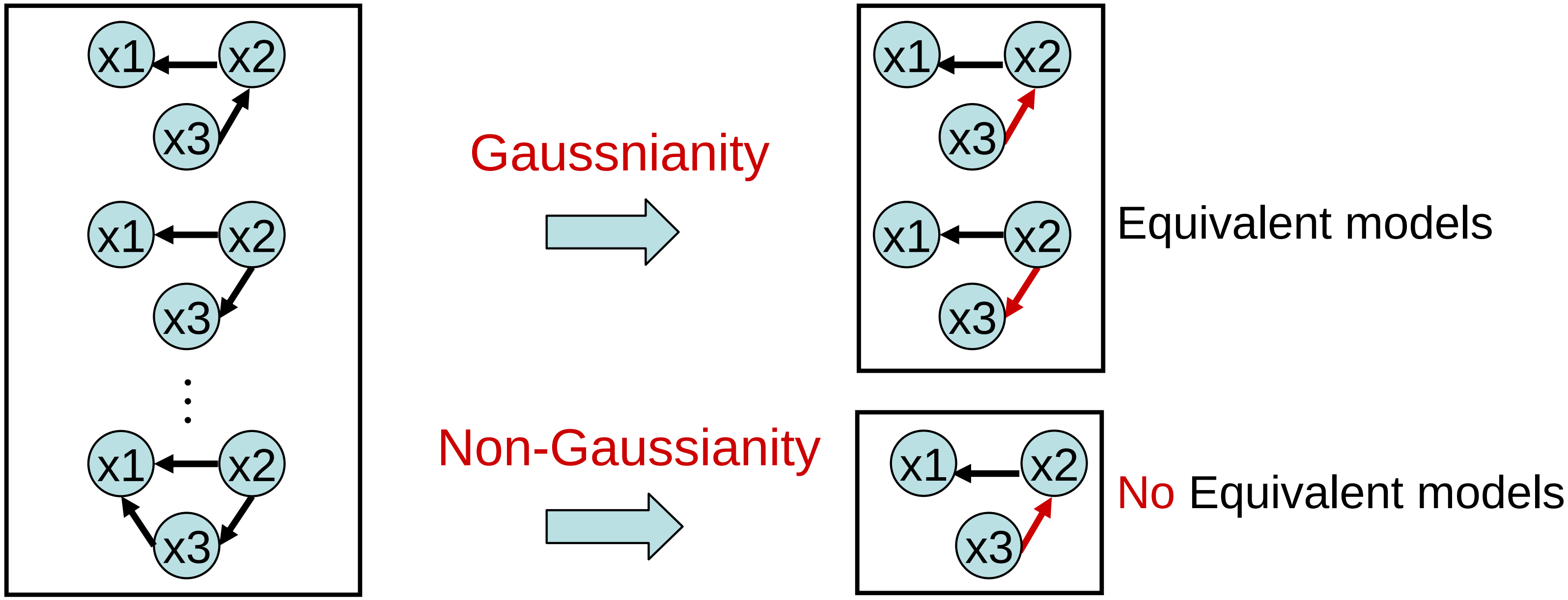



**Estimated network**

# Discovery of linear acyclic models from non-experimental data

- Existing methods (Bollen,1989; Pearl, 2000; Spirtes et al., 2000)
  - Gaussian assumption on disturbances ei
  - Produce many equivalent models
- Our LiNGAM approach (Shimizu et al, UAI2005, 2006 JMLR)
  - Replace Gaussian assumption by non-Gaussian assumption
  - Can identify the connection strengths and structure

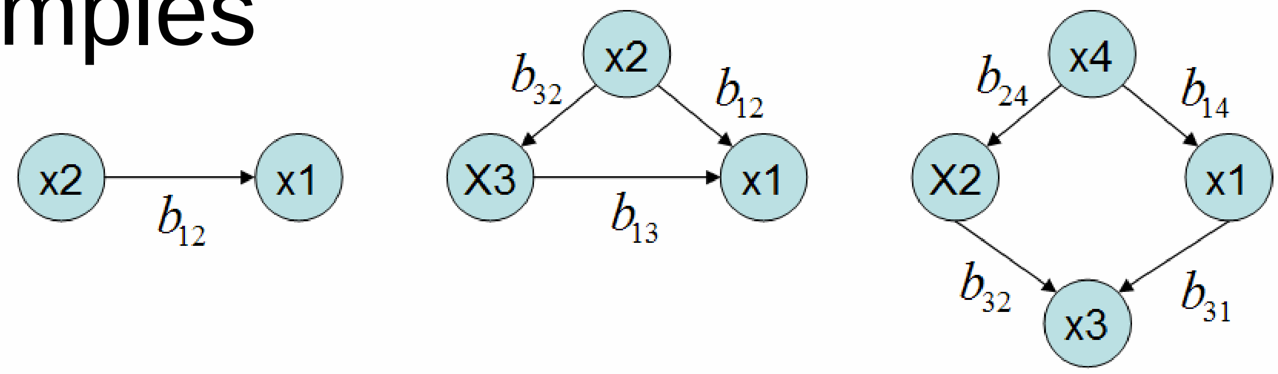


# Linear Non-Gaussian Acyclic Models (LiNGAM)

- As usual, linear acyclic models, but disturbances $e_i$ are assumed to be non-Gaussian:

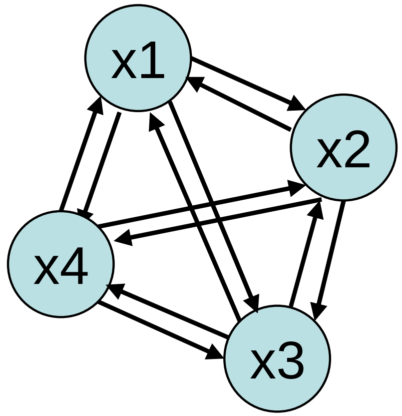
$$x_i = \sum_{k(j)<k(i)} b_{ij} x_j + e_i \quad \text{or} \quad \mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e}$$

- Examples



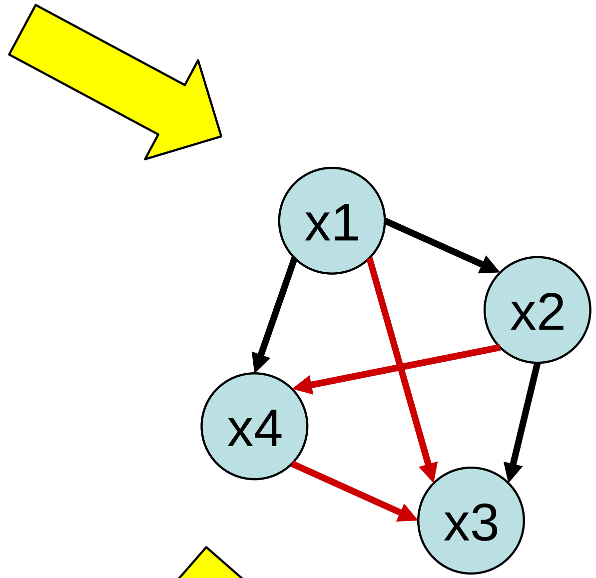
# Outline of LiNGAM algorithm

**1. Estimate B by ICA**
**+ post-processing**

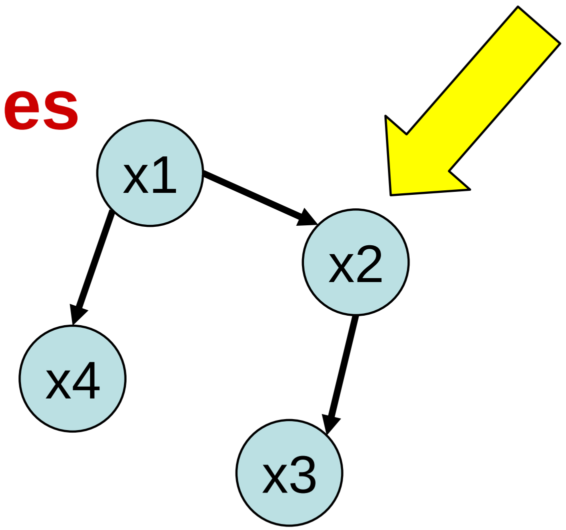



**2. Find an order k(i) (DAG)**

Non-significant edges




**3. Prune non-significant edges**



# Key ideas

- Observed variables $x_i$ are linear combinations of non-Gaussian independent disturbances $e_i$

$$\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e}$$

$$\Rightarrow \ \mathbf{x} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{e}$$

$$= \mathbf{A}\mathbf{e}$$

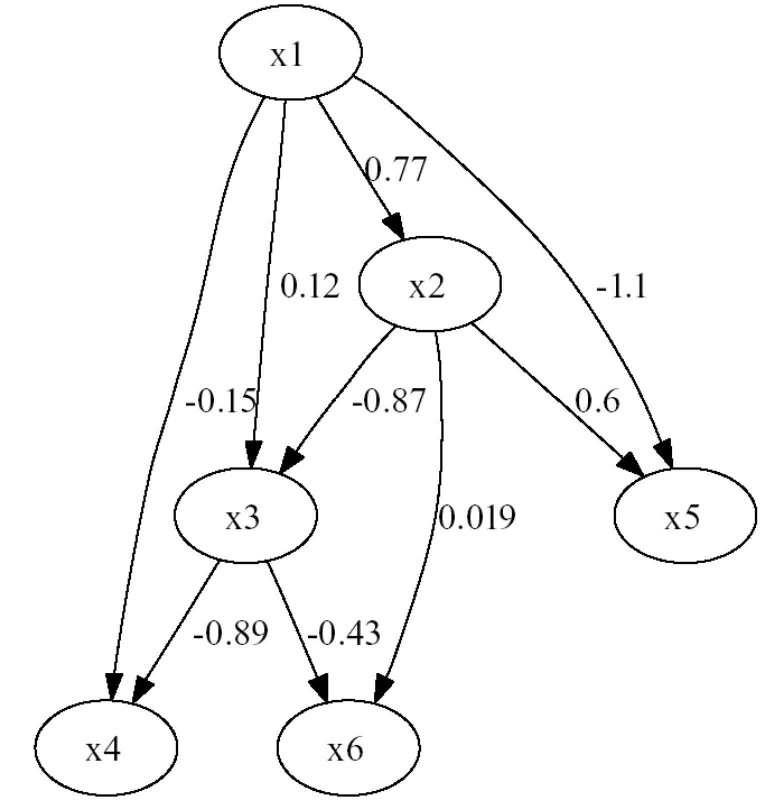
  – The classic case of ICA (Independent Component Analysis)

- Permutation indeterminacy in ICA can be solved
  – Can be shown that the correct permutation is the only one which has no zeros in the diagonal (Shimizu et al., 2005; 2006)

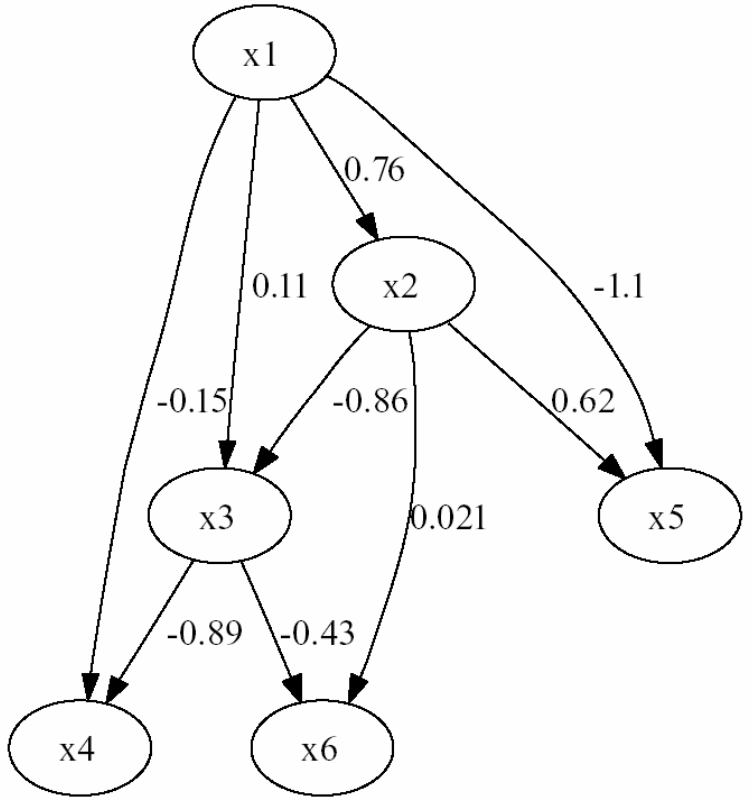
- Pruning edges can be done by many existing methods
  – Imposing sparseness, testing, model fit etc.

# Examples of estimated networks

- All the edges correctly identified
- All the connection strengths approximately correct



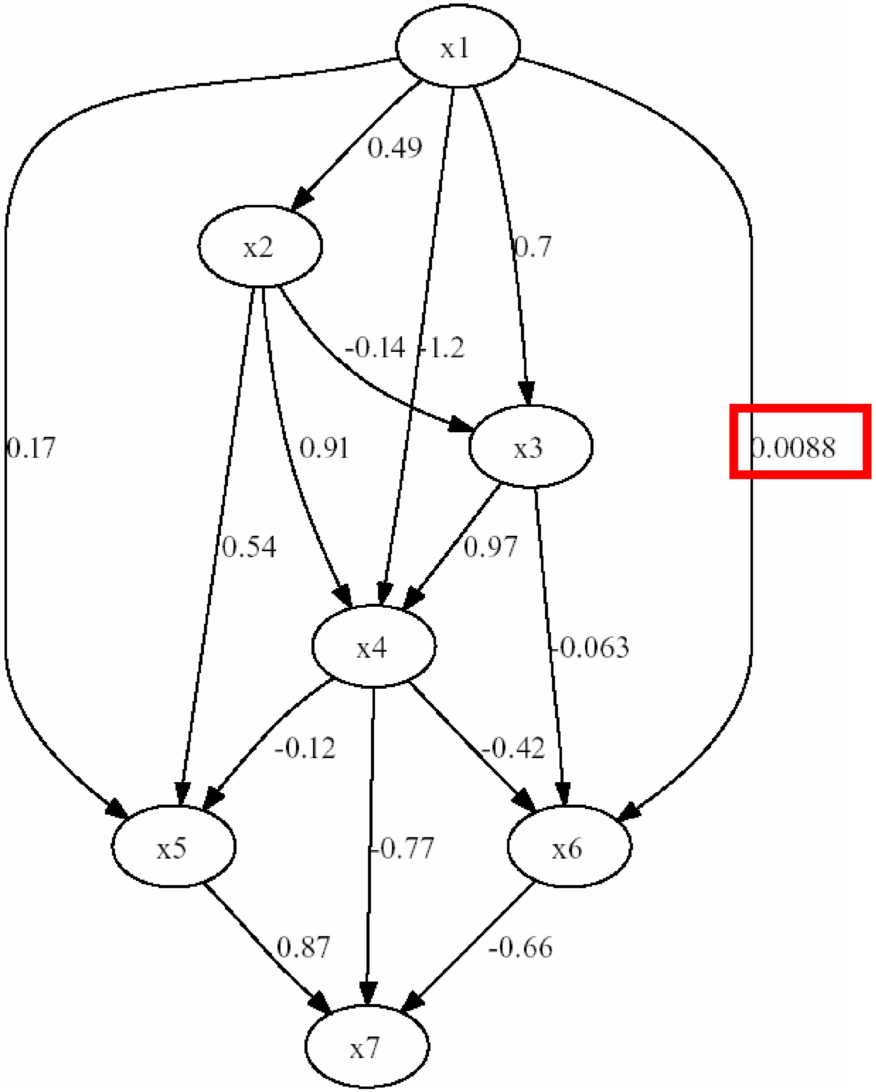

**Original network**



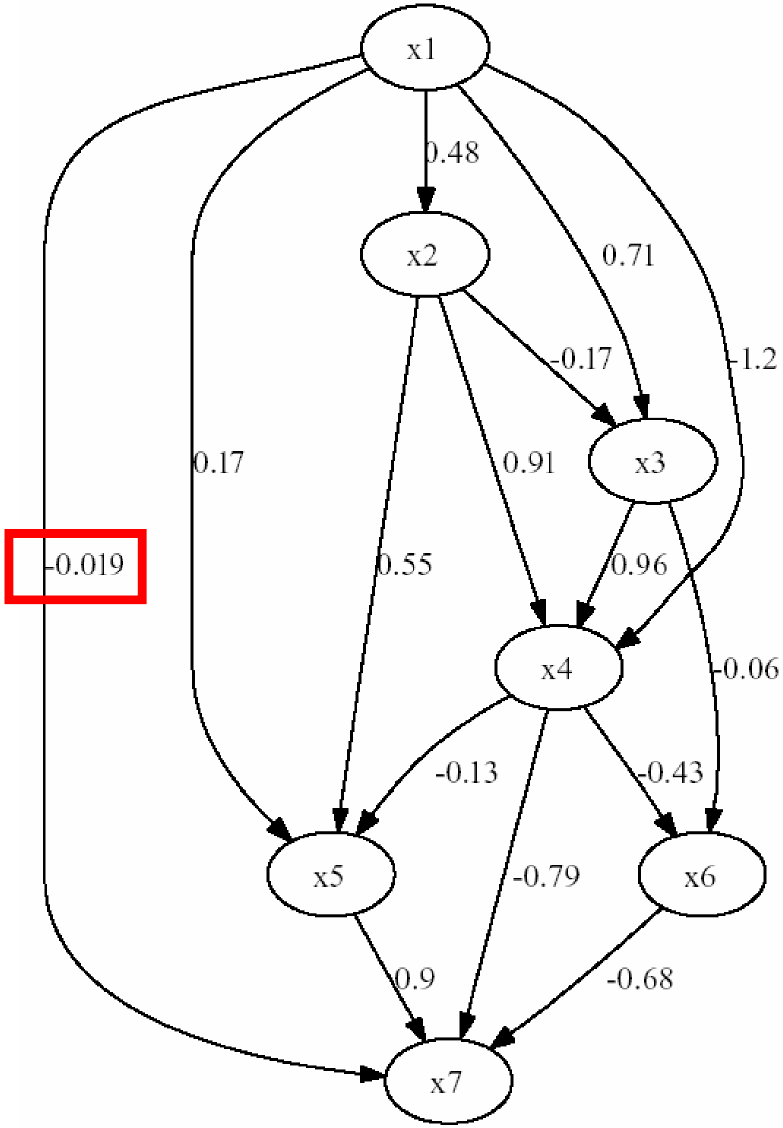

**Estimated network**

# What kind of mistakes LiNGAM might make?

- One falsely added edge (x1→x7, -0.019)
- One missing edge (x1→x6, 0.0088)




**Original network**




**Estimated network**

# Real data example 1

- Galton's height data (Galton, 1886)
  - x1: child height
  - x2: `midparent' height
    - (father's height + 1.08 mother's height)/2
  - 928 observations

- Estimates:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 0.67 \\ -0.012 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

- Estimated direction:

$$x_1 \leftarrow x_2 \qquad (\text{child} \leftarrow \text{midparent})$$

# Real data example 2

- Fuller's corn data (Fuller, 1987)
  - x1: Yield of corn
  - x2: Soil Nitrogen
  - 11 observations

- Estimates:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 0.34 \\ 0.014 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

- Estimated direction:

$$x_1 \leftarrow x_2 \qquad (\text{Yield} \leftarrow \text{Nitrogen})$$

# Real data example 3

- Cars data (Ezekiel, 1930)
  - x1: Speed
  - x2: Distance taken to stop
  - 50 observations

- Estimates:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 0.072 \\ 3.10 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

- Estimated direction:

$$x_1 \rightarrow x_2 \qquad (\text{Speed} \rightarrow \text{Distance})$$

# Conclusions

- Discovery of linear acyclic models from non-experimental data is an important topic of current research
- A common assumption is linear-Gaussianity, but this leads to a number of indistinguishable models
- A non-Gaussian assumption allows all the connection strengths and structure of linear acyclic models to be identified
- Basic method: ICA + permutation + pruning
- Matlab/Octave code: http://www.cs.helsinki.fi/group/neuroinf/lingam/

# References


- S. Shimizu, A. Hyvärinen, Y. Kano, and P. O. Hoyer **Discovery of non-gaussian linear causal models using ICA**. In Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI2005), pp. 526-533, 2005.

- S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen. **A linear non-gaussian acyclic model for causal discovery**. *Journal of Machine Learning Research*, 7: 2003--2030, 2006.