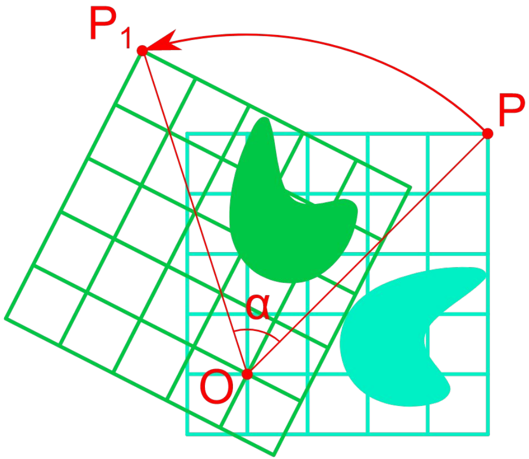
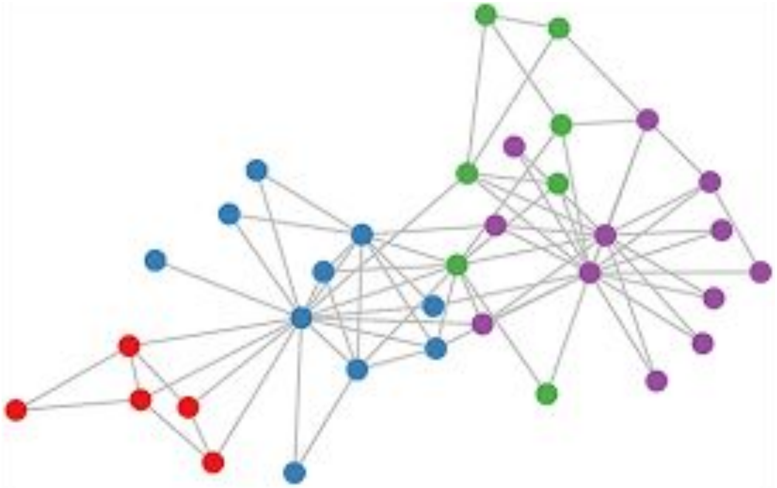




Proyecciones Multidimensionales

una mirada hacia el descubrimiento de patrones y análisis de datos en alta dimensión

SECTION

3

Prof. Dr. Diego Nascimento

UNIVERSIDAD DE ATACAMA
1857

DPTO. MATEMÁTICA | FACULTAD DE INGENIERÍA

OVERVIEW

MINI-COURSE SCHEDULE

SECTION #1

Exploratory Analysis

- Visual Data Mining

SECTION #2

Intro to Clustering in Data Modeling

- First steps to Overlook High-Dimensional Data

SECTION #3

Intro to Causal Discovery

- Multivariate Time Series Structure Estimation

“Correlation is not Causation”

What is Causation?

We may define a cause to be an object, followed by another, [...] where, if the first object had not been, the second never had existed // Podemos definir una causa como un objeto, seguido de otro, [...] donde, si el primer objeto no hubiera existido, el segundo nunca hubiera existido. (DAVID HUME).

COrrelation := is a linear measurement between TWO variables, then one can extend to

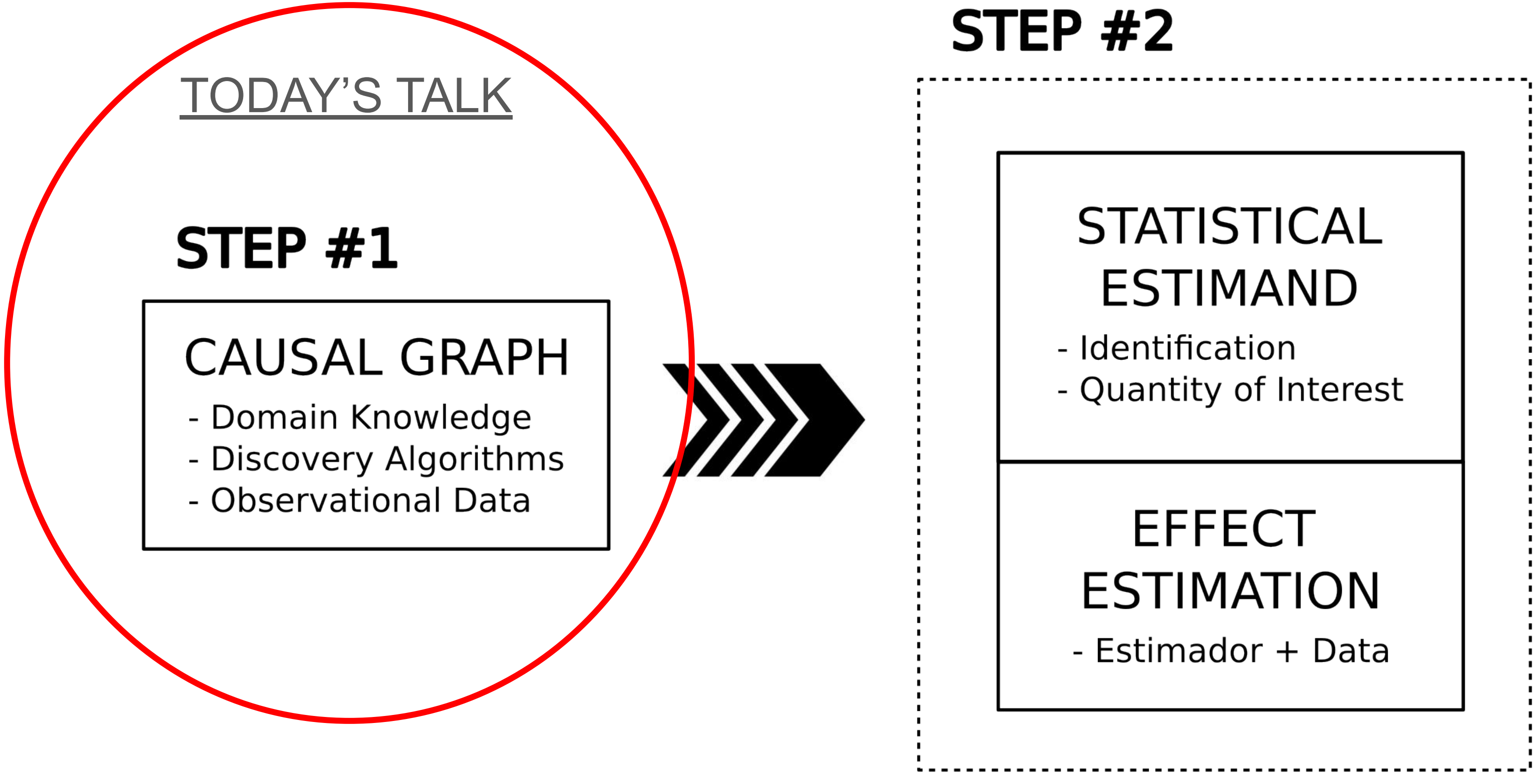
“Association is not Causation”

…another way to see it is that,

Association = Causation + Bias

(SPURIOUS)

CAUSAL INFERENCE –STEPS–



TODAY’S TALK

STEP #1

CAUSAL GRAPH

- Domain Knowledge
- Discovery Algorithms
- Observational Data

STEP #2

STATISTICAL ESTIMAND

- Identification
- Quantity of Interest

EFFECT ESTIMATION

- Estimador + Data

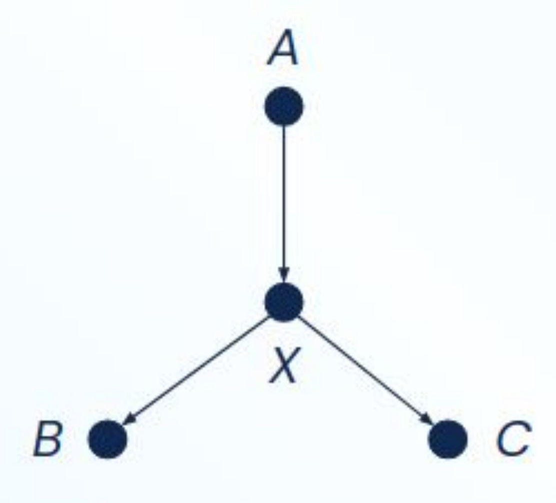
In Higher dimensions…

How can one decide if ANOTHER INFORMATION HELPs?

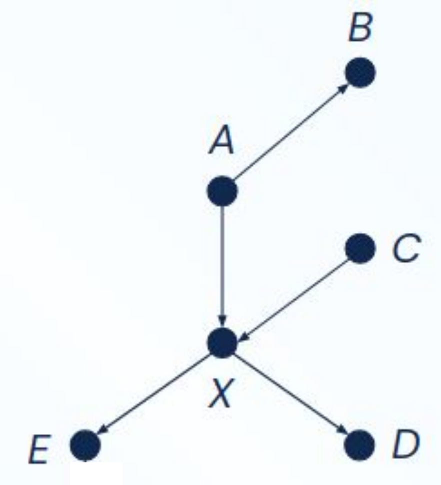
Causal Structure helps us solving this problem…

A graph G is composed by the combinations of a finite set Random Variables (which will be nodes) and a set of edges/links E (their relationship with directions).

A special class of graph, which holds only directed edges E ⊆ V × V, are called Directed Acyclic Graph (DAG). For instance, let's suppose G = (V, E) consists of



Graphical representation (DAGs) of a Causal relations



For instance, let's suppose G = (V, E) consists of

GRAPH TERMINOLOGY

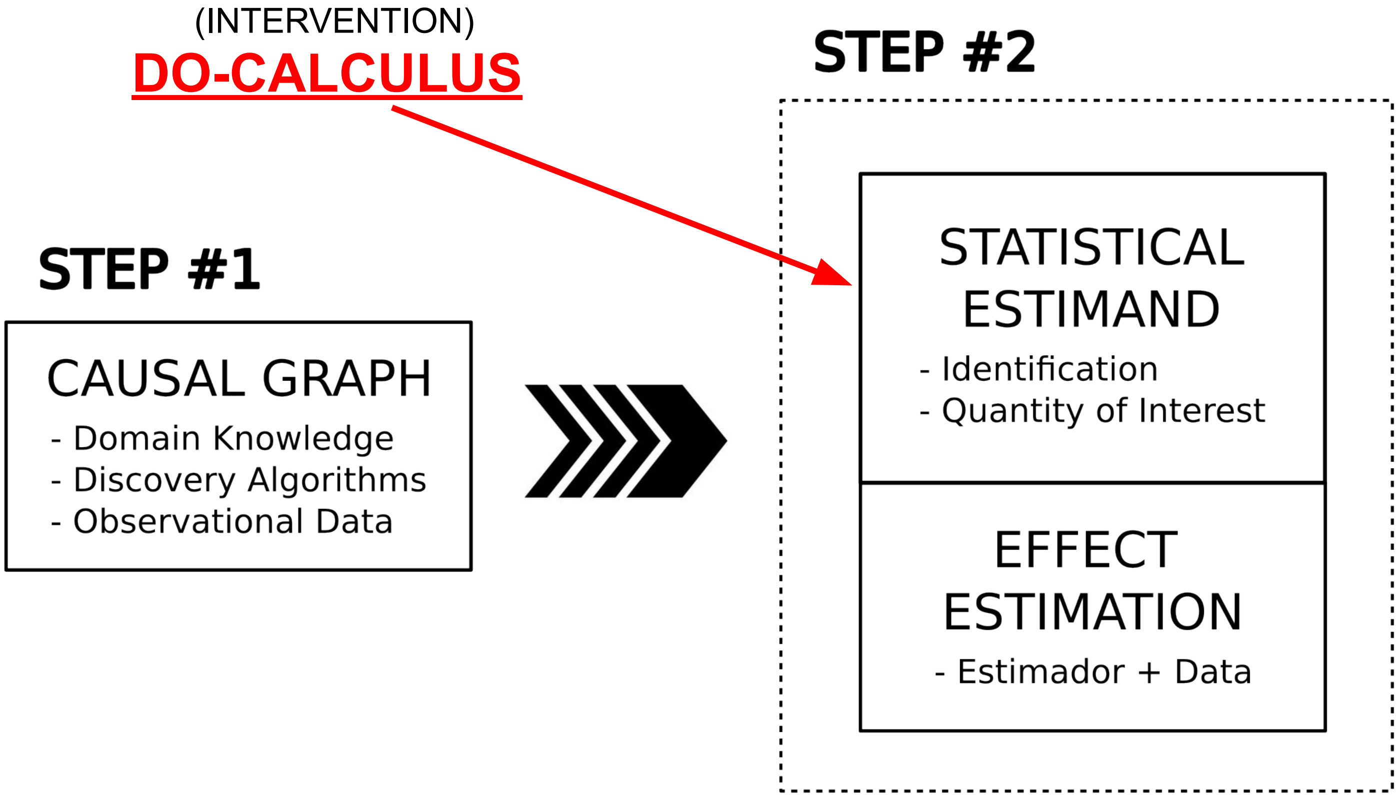
- The PARENT of a node
- The CHILD of a node
- DESCENDANTS of a node (nodes that can be reached by following directed edges starting from)
- NON-DESCENDANTS or Ancestors of a node (nodes that can not be reached from the selected node)
- ADJACENT (Nodes connected by the selected edge)

A Causal DAG is a DAG with causal meaning over the edges!

Assumptions:

1. CAUSAL MARKOV CONDITION
2. FAITHFULNESS
3. CAUSAL SUFFICIENCY
4. ACYCLICITY

So far, we only discuss about CAUSAL REPRESENTATION



(INTERVENTION)

DO-CALCULUS

STEP #1

CAUSAL GRAPH

- Domain Knowledge
- Discovery Algorithms
- Observational Data

STEP #2

STATISTICAL ESTIMAND

- Identification
- Quantity of Interest

EFFECT ESTIMATION

- Estimador + Data

RANDOMIZED CONTROLLED TRIAL (RCT)

GOLDEN STANDARD OF CAUSAL INFERENCE

- EXPERIMENTAL UNITS (PARTICIPANTS, MACHINES…) ARE RANDOMLY ASSIGNED TO DIFFERENT TREATMENT GROUPS.
- OUTCOMES OF THE GROUPS ARE COMPARED TO MEASURE AVERAGE CAUSAL EFFECTS.
- RANDOMIZATION ENSURES THAT THERE IS NO BIAS (COVARIATE BALANCE)

For instance, let’s suppose a feature is conditional on the treatment T,

$$\mathbb{P}(X|T=1) \overset{d}{=} \mathbb{P}(X|T=0) \to \mathbb{P}(X)$$

therefore, randomizing the treatment makes T independent of X. Then, discuss all experiment CAUSAL DISCOVERY!

Judea Pearl **DO-CALCULUS & its Assumptions**

(LINK BETWEEN STATISTICAL INFERENCE and CAUSAL INFERENCE)

Natural mechanisms that determine treatment are why biases can exist

- The Operator **do(X)** signifies that we are INTERVENING on a variable X

Distributions of the form P(Y|X) are called OBSERVATIONAL DISTRIBUTIONS.

Distributions of the form P(Y|do(X)) are called INTERVENTIONAL DISTRIBUTIONS.

Assumptions:

1. CONDITIONAL EXCHANGEABILITY (IGNORABILITY)
2. CONSISTENCY
3. FAITHFULNESS
4. POSITIVITY
5. CAUSAL STATIONARITY
6. CAUSAL SUFFICIENCY
7. CAUSAL MARKOV CONDITION

Representation of a probability as a CAUSAL PROCESS (DAG)

FACTORIZATION OF JOINT PROBABILITY DENSITY FUNCTION

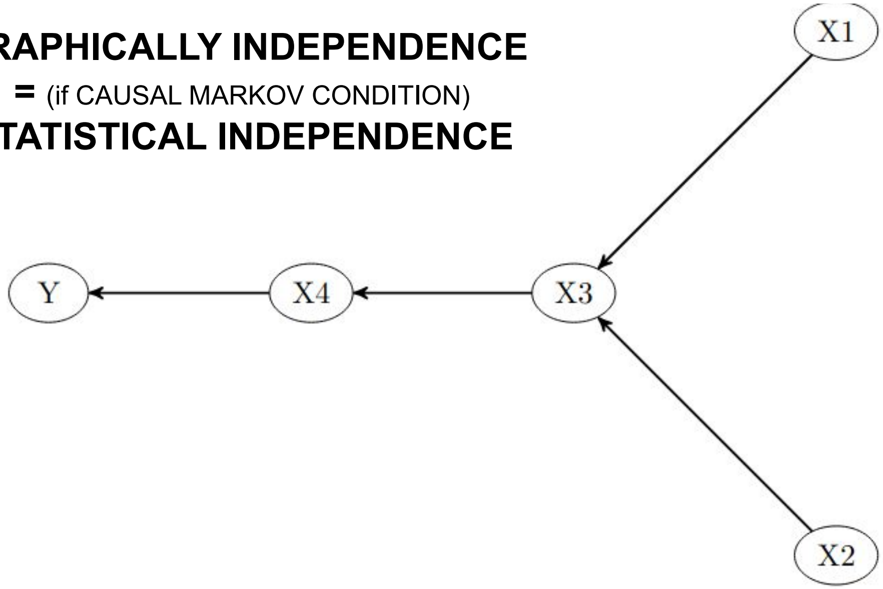
$$\begin{aligned}\mathbb{P}(X1, X2, X3, X4, Y) &= \mathbb{P}(X1)\mathbb{P}(X2|X1)\mathbb{P}(X3|X2, X1)\mathbb{P}(X4|X3, X2, X1)\mathbb{P}(Y|X4, X3, X2, X1)\\ &\overset{*}{=} \mathbb{P}(X1)\mathbb{P}(X2)\mathbb{P}(X3|X2, X1)\mathbb{P}(X4|X3)\mathbb{P}(Y|X4)\end{aligned}$$

* CONDITIONAL INDEPENDENCE

GRAPHICALLY INDEPENDENCE

= (if CAUSAL MARKOV CONDITION)

STATISTICAL INDEPENDENCE



RELATIONAL STRENGTH:

1. Association (1st level) - Conditional Probability

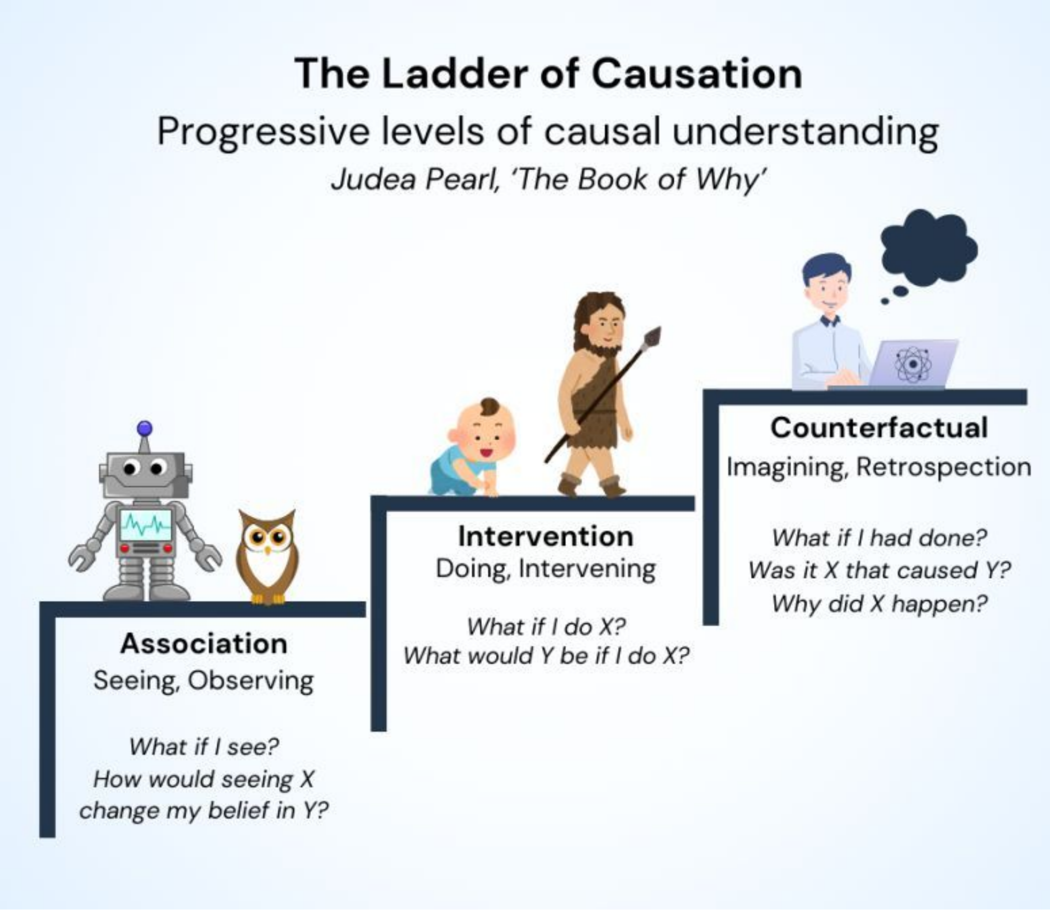
   Distribution form of Observational Distributions (Data Set only)

1. Intervention (2nd level) - Intervening (not only observing)

   Distribution form of Interventional Distribution (Do-Calculus)

1. Imagining Counterfactuals (3rd level) - Parallel World

The exact functional relationship X and Y



The Ladder of Causation

Progressive levels of causal understanding

Judea Pearl, 'The Book of Why'

Association
Seeing, Observing

What if I see?
How would seeing X change my belief in Y?

Intervention
Doing, Intervening

What if I do X?
What would Y be if I do X?

Counterfactual
Imagining, Retrospection

What if I had done?
Was it X that caused Y?
Why did X happen?

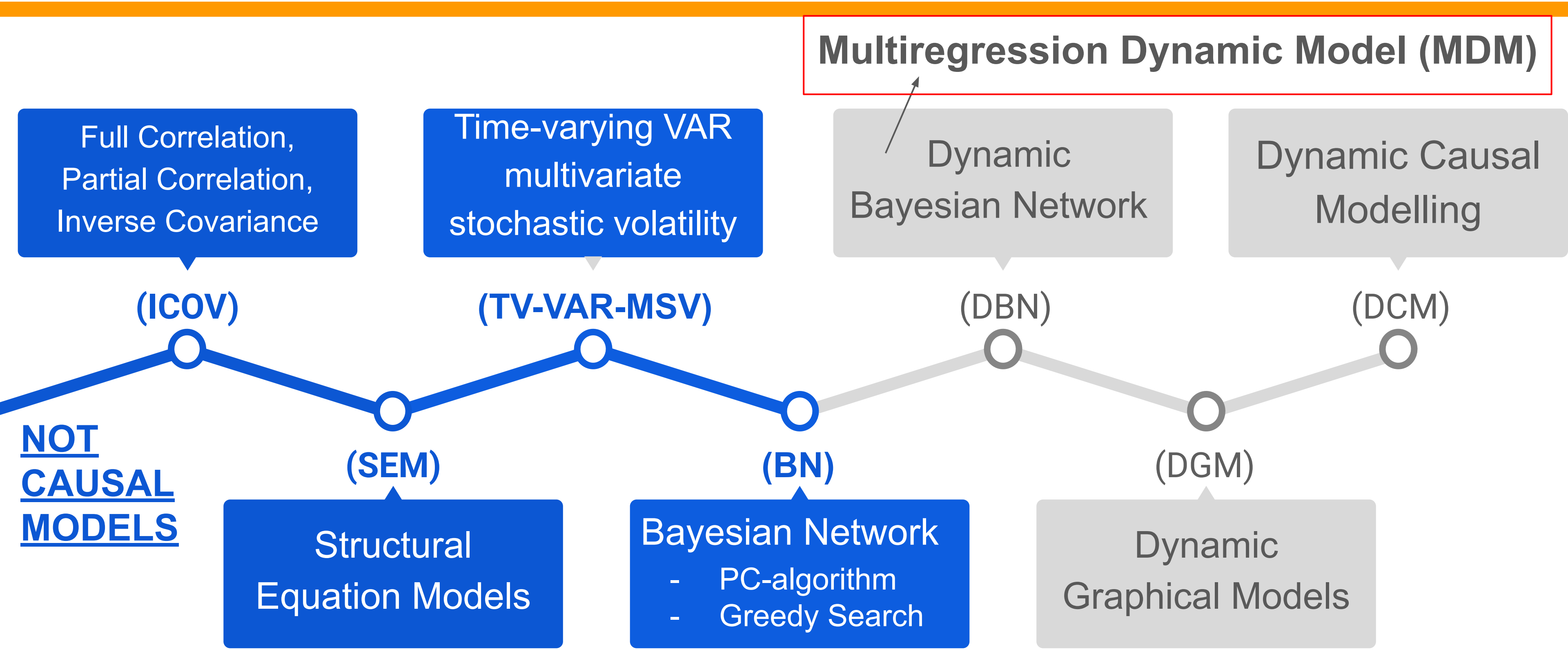
Causal Inference for TIME SERIES Analysis

Time series data is a collection of chronological observations which are generated by several domains such as medical and financial fields.

- Discovering the causal relations between the time series component
    1. Granger causality and conditional independence based
    2. Structural equation model based
    3. Dynamical based methods (Dynamic Bayesian Net, Gaussian Processes, ANN…)
- Estimate the impact of an intervention/treatment over time
    1. Time-invariant treatment effect
    2. Time-varying treatment
    3. Dynamic regimes

Moraffah, R., Sheth, P., Karami, M., Bhattacharya, A., Wang, Q., Tahir, A., ... & Liu, H. (2021).
Causal inference for time series analysis: Problems, methods and evaluation.
Knowledge and Information Systems, *63*, 3041-3085.

Time Series (TS) relational ESTIMATION methods



Multiregression Dynamic Model (MDM)

NOT CAUSAL MODELS

Full Correlation, Partial Correlation, Inverse Covariance

(ICOV)

(SEM)

Structural Equation Models

Time-varying VAR multivariate stochastic volatility

(TV-VAR-MSV)

(BN)

Bayesian Network
- PC-algorithm
- Greedy Search

Dynamic Bayesian Network

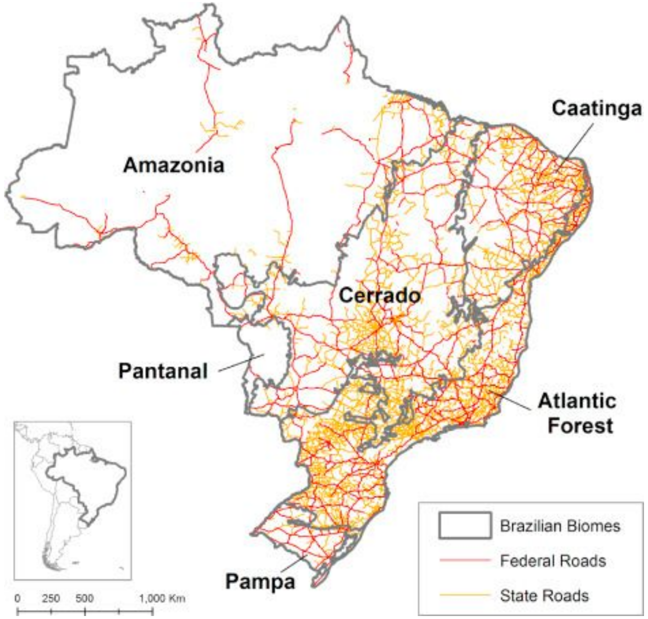
(DBN)

(DGM)

Dynamic Graphical Models

Dynamic Causal Modelling

(DCM)



Does geographic information enriches the Observed Outcome?

EXAMPLE #1 - CAUSAL DISCOVERY with TIME SERIES

Adjusted Gaussian-Autoregressive BN: a new spatiotemporal Bayesian network approach (SUBMITTED)

Federal University of Bahia (UFBA)

Federal University of Paraná (UFPR)

Federal University of Bahia (UFBA)

Federal University of Paraná (UFPR)

University of Atacama (CHILE)

BRAZILIAN COVID-19 case

Undeniable the COVID-19 epidemy was catastrophic worldwide. One big question related to the dynamic is how the virus interact across the areas?

Regardless the politics, the virus does not walk or flys, therefore is cared across the area. It can be useful to know how does people commute across the field.

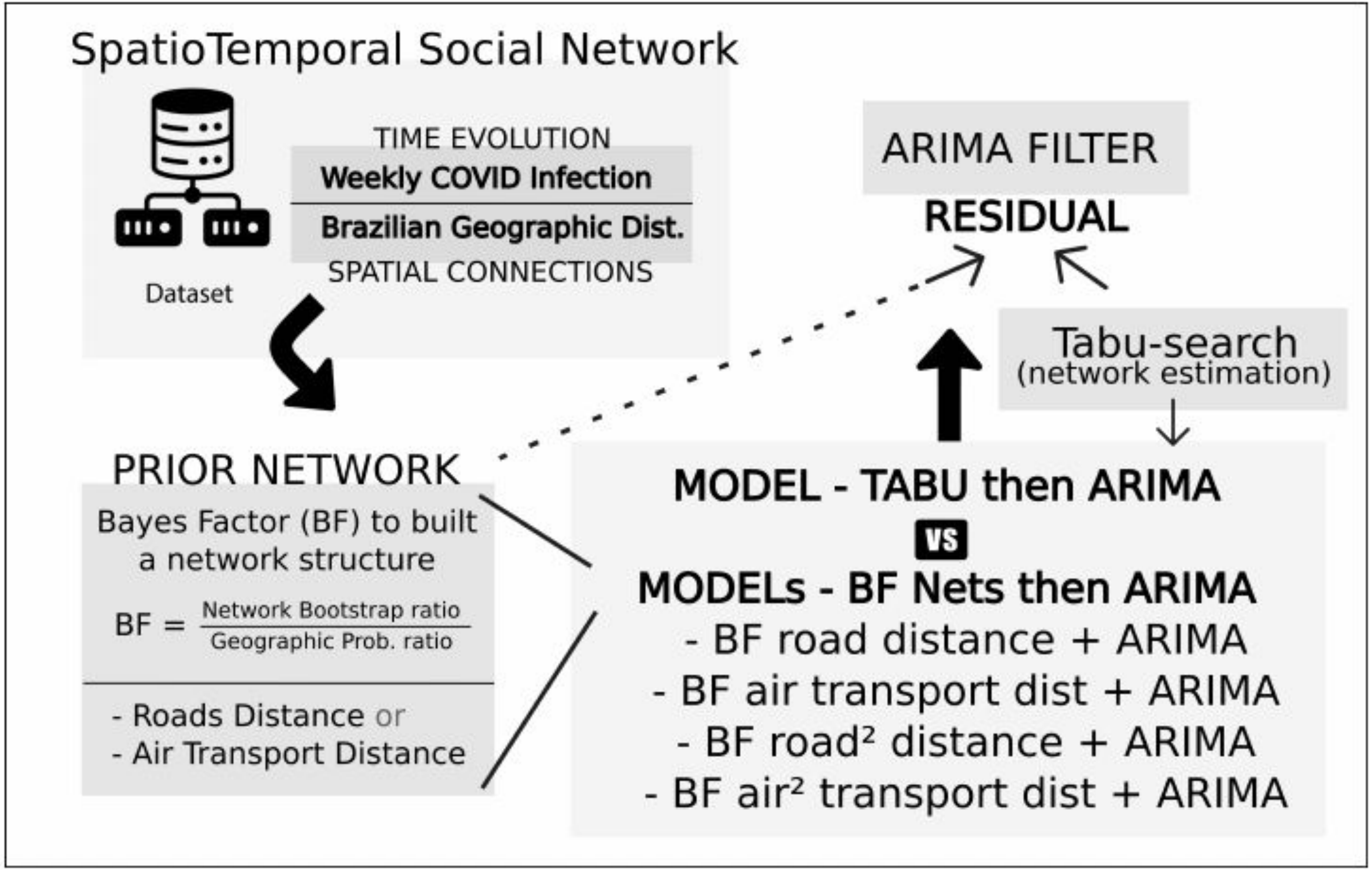
- This work was motivated by the Brazilian COVID-19 case, As a social network, it was considered 26 weekly Brazilian states and the Federal District, between 02/25/2020 and 02/15/2022.

Domain Knowledge was added as a geographical distance information using as prior causal structure.

Causal Discovery Study Case #1

Methodological Abstract

1. Calculate the distance across all capital cities (Road and Air distance)
2. Calculate the possible DAGs with Tabu Search algorithm (Score-based)
   a. The first way is the conventional (no information an the algorithm starts with no connections.
   b. The alternative way used the Bayes Factor to incorporate the Brazilian Logistic as an initial point.
3. Complement the static Bayesian Network, which the root node(s) are invariant, with a ARIMA modeling.



SpatioTemporal Social Network

Dataset

TIME EVOLUTION

Weekly COVID Infection

Brazilian Geographic Dist.

SPATIAL CONNECTIONS

PRIOR NETWORK

Bayes Factor (BF) to built a network structure

$$BF = \frac{\text{Network Bootstrap ratio}}{\text{Geographic Prob. ratio}}$$

- Roads Distance or
- Air Transport Distance

ARIMA FILTER

RESIDUAL

Tabu-search (network estimation)

MODEL - TABU then ARIMA

VS

MODELs - BF Nets then ARIMA

- BF road distance + ARIMA
- BF air transport dist + ARIMA
- BF road^2 distance + ARIMA
- BF air^2 transport dist + ARIMA

Review of – Bayesian Network

Let Y = {Y1 , Y2 , ..., Yk} be a set of k random variables in a BN, the Markov property is satisfied if the joint probability distribution of the variables can be written as the product of the INDEPENDENT conditional probabilities,

$$P(Y_1 = y_1, ..., Y_k = y_k) = \prod_{i=1}^{k} P(Y_i|par(Y_i)),$$

based on par(Yi) is the set of parent(s) of the variable Yi.

- Gaussian Bayesian Network is defined as a continuous and all conditional probability distributions are linear Gaussian (considering the model’s explanatory variables, X)

Review of – Bayes Factor

Bayes factor (BF), proposed by Jeffreys, is used to test the weight of evidence of one hypothesis over another of EACH PAIRWISE.

DATA

BOOTSTRAP STRENGTH SIMULATION

whereas,

$$\frac{P(H_0|D)/P(H_1|D)}{P(H_0)/P(H_1)},$$

$$P(H_0) = \frac{1/d_{ij}}{\sum_i 1/d_{ij}},$$

and

$P(H_1) = [1\text{-}P(H_0)]$.

DISTANCE BETWEEN REGIONS i and j

H. Jeffreys, Theory of Probability, 3rd ed. Oxford: Oxford University Press, 1961

Review of – ARIMA

Let W(t), t={1,…,n}, be random variable with conditional distribution following a Gaussian distribution

from the Bayesian Network model

$$\mathbb{E}[W(t)|\mathcal{F}(t-1)] = \mu + \Sigma_{i=1}^{p}\phi_i W(t-i) + \Sigma_{j=1}^{q}\theta_j \nu(t-j) + \nu(t)$$

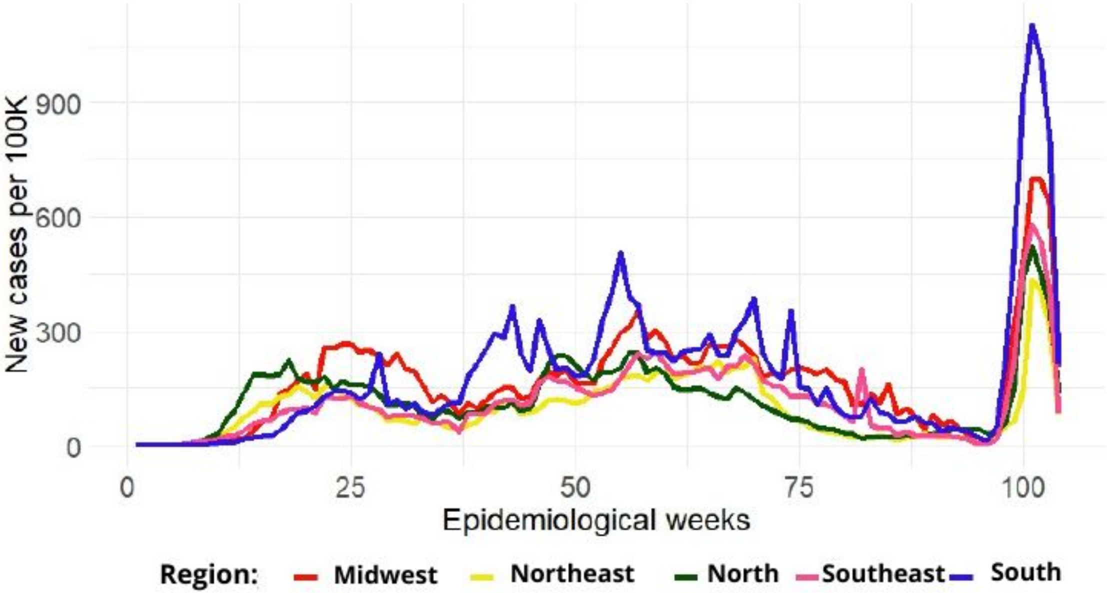
(smallest σ-algebra)

The autoregressive integrated moving average (ARIMA) model,

$$(1-\phi B)\,\nabla^d\, W(t) = (1-\theta B)\nu(t),$$

- Autoregressive (AR) p part indicates that the variables of interest are regressed by their past or lagged values
- Moving average (MA) q term indicates that the regression error is a combination of the errors of the autoregressive part
- Integrated (I) part is the d differentiation applied

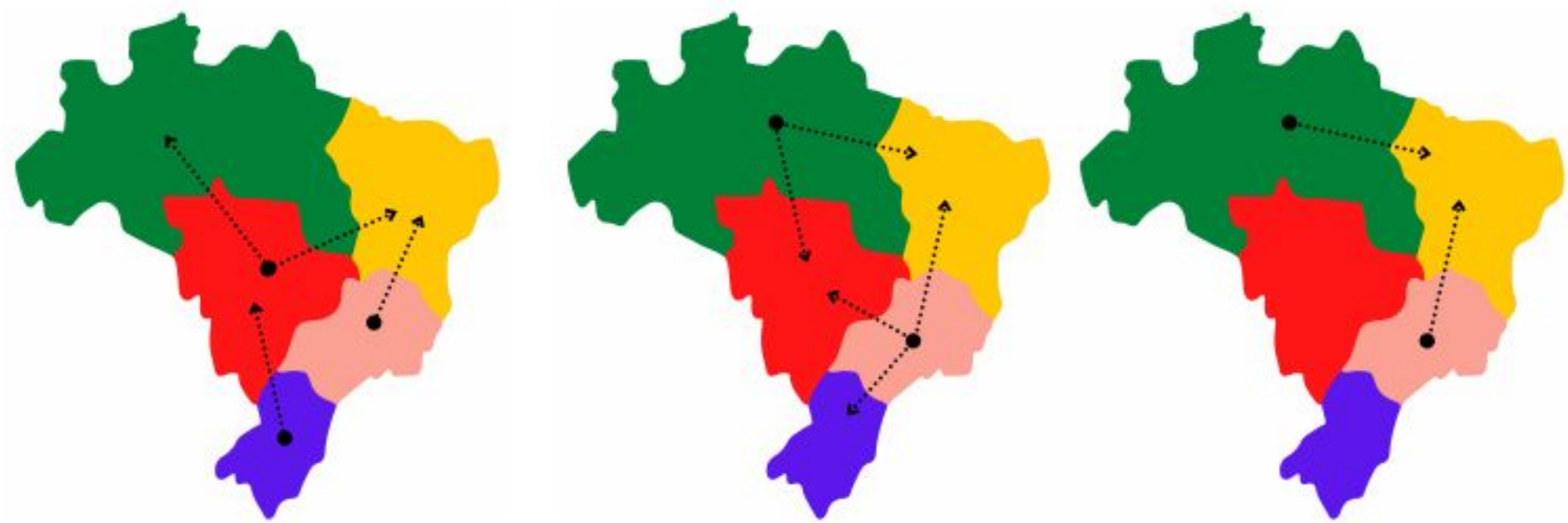
BRAZILIAN COVID-19 case



The period of analysis was from 02/25/2020 to 02/15/2022, and for each region, the number of new cases per 100,000 inhabitants in each of the 104 epidemiological weeks defined by the Ministry of Health was considered.

https://github.com/ProfNascimento/spatialBN

Bayesian Network Structure ONLY



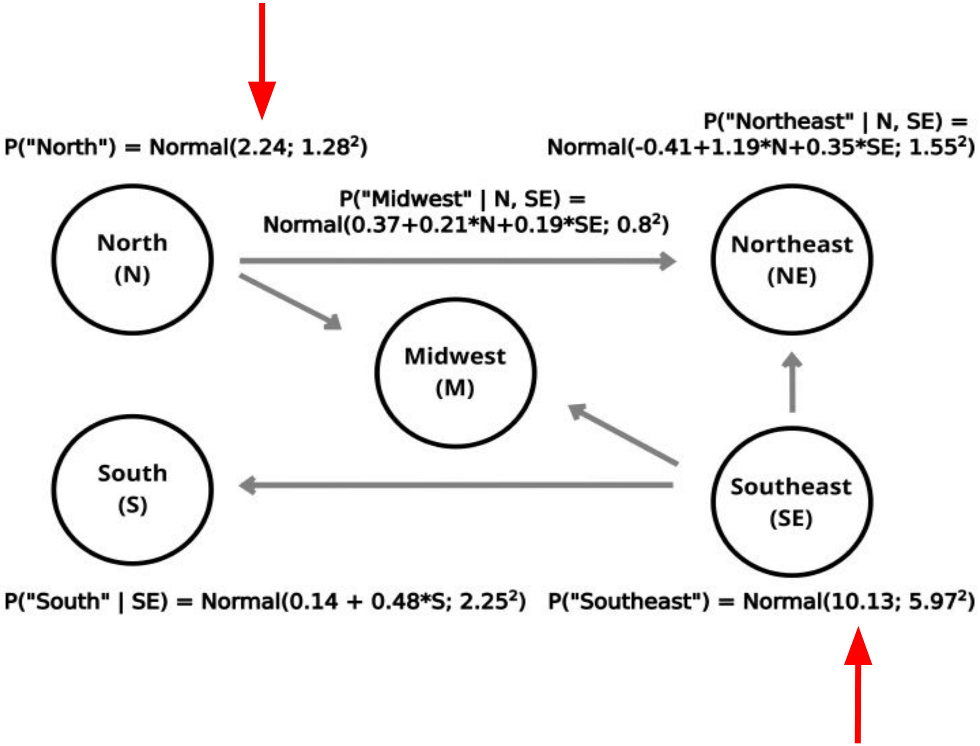
(a) Tabu

(b) Distances (BF=1)

(c) Distances (BF=4)

Bayesian Network Structure –ONLY–

$$(Y_N, Y_{NE}, Y_M, Y_{SE}, Y_S)^T \sim N_5(\underset{\sim}{\mu}, \Sigma)$$



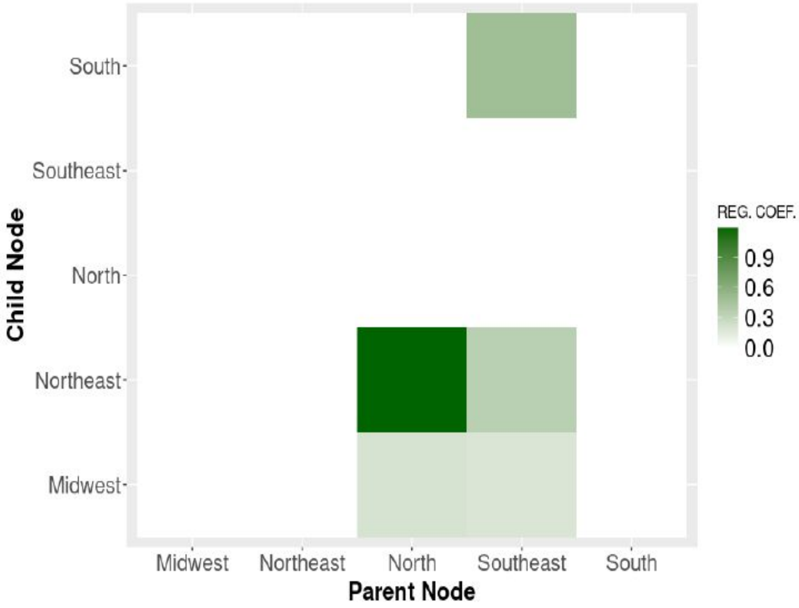
P("North") = Normal(2.24; 1.28^2)

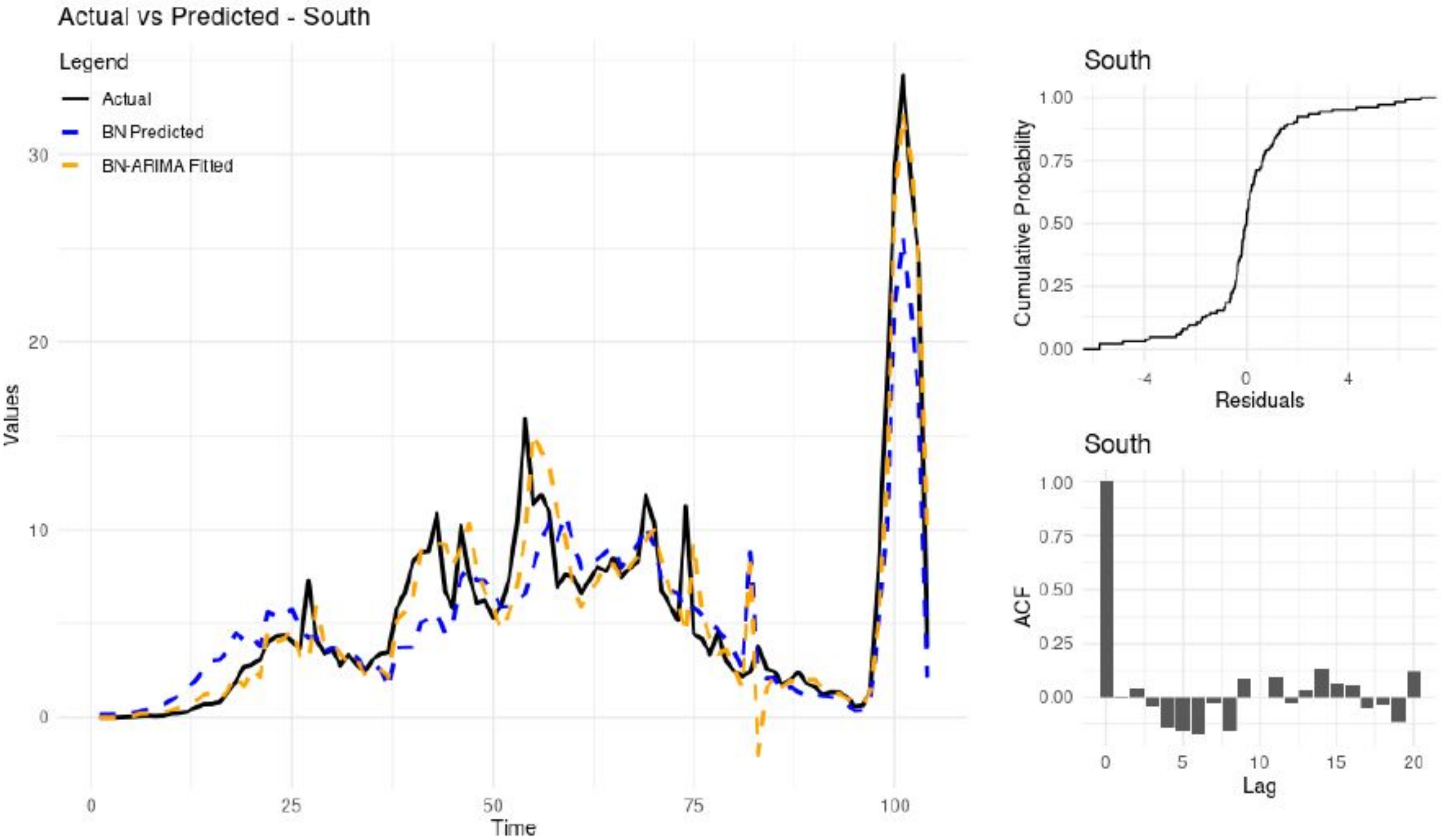
P("Northeast" | N, SE) = Normal(-0.41+1.19*N+0.35*SE; 1.55^2)

P("Midwest" | N, SE) = Normal(0.37+0.21*N+0.19*SE; 0.8^2)

P("South" | SE) = Normal(0.14 + 0.48*S; 2.25^2)

P("Southeast") = Normal(10.13; 5.97^2)




Actual vs Predicted - South
Legend
Actual
BN Predicted
BN-ARIMA Fitted
Values
0
10
20
30
0
25
50
75
100
Time
South
Cumulative Probability
0.00
0.25
0.50
0.75
1.00
-4
0
4
Residuals
South
ACF
0.00
0.25
0.50
0.75
1.00
0
5
10
15
20
Lag

Sort Discussion

Results from the proposed model unraveled the spatial dependence between regions, from the COVID-19 incidence spread contribution rate. From a statistical causal perspective via DAG, the BN roots are explainable:

- The North region, one presented higher mortality rates and was contagious.
- The Southeast region during the highest demographic concentration, therefore amount of infection.

The Brazilian **Northeast region** is the one that presents **greater weighted relations** with the other regions and is impacted by the spread of the disease due to connections.

Is grouping relevant information on a Brain Circuits Communication?

EXAMPLE #2 - CAUSAL DISCOVERY IN TIME SERIES

Evaluating brain group structure methods using hierarchical dynamic models (Pattern Recognition)

Federal University of Bahia (UFBA)

University of Sao Paulo (USP)

University of Atacama (CHILE)

Warwick University (UK)

Open University (UK)

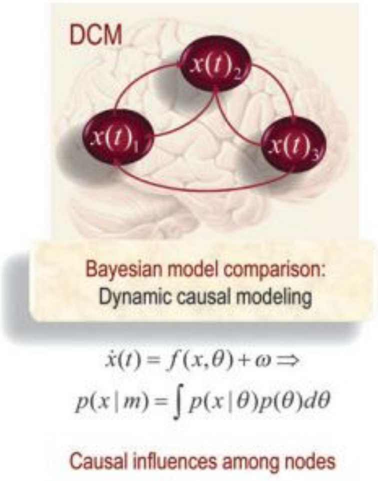
University of Sao Paulo (USP)

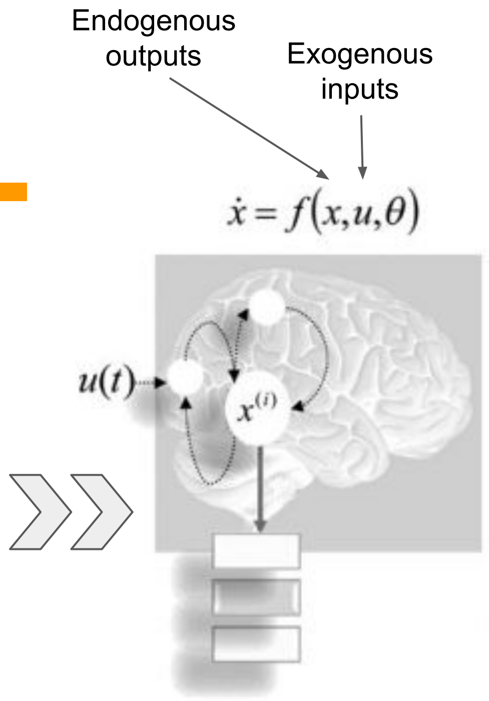
Oxford University (UK)

Based on Dynamic Causal Model (DCM)

The brain is a function ($\dot{\mathbf{x}}(\mathbf{t})$) formed by different regions ($\mathbf{x}$) that share information (Horwitz, 2003).

- **Functional Connectivity** represents the patterns of statistical dependence the activity of distinct brain regions
- **Effective Connectivity** represents the causal influences of the activity of one regions over another


DCM
$x(t)_2$
$x(t)_1$
$x(t)_3$
Bayesian model comparison:
Dynamic causal modeling
$\dot{x}(t) = f(x,\theta) + \omega \Rightarrow$
$p(x|m) = \int p(x|\theta)p(\theta)d\theta$
Causal influences among nodes



FONTE: Friston, 2011.

Taking the MULTIVARIATE NORMAL distribution as example, the functional connectivity is estimated by the variance-covariance matrix (uses **previous/full time information**). The effective connectivity would be the impact of the response variables $x(t)_i$, across each other ONLY in the **same time** (Causal Pattern), through Dynamic Factor Model (DFM) or Multiregression Dynamic Model (MDM).

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain connectivity*, *1*(1), 13-36.

Review of –Dynamic Linear Model (DLM)

$$Y_t(i) \text{ independent } Y_t^i | Parent(Y_t(i))$$

Estimation as a Kalman Filter

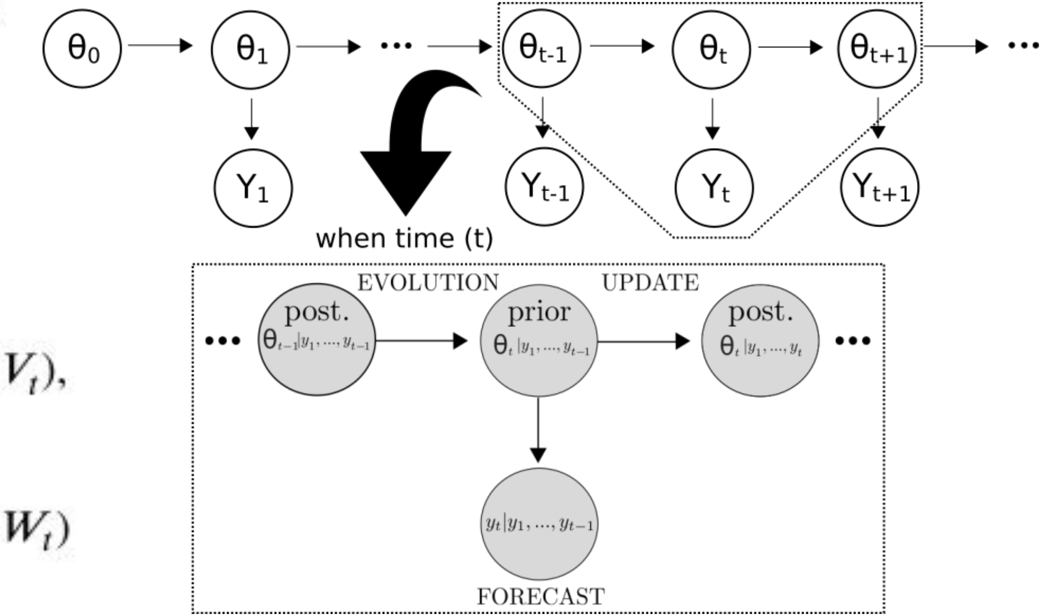
At time $t = 0$,

$$\theta_0 \sim N(\theta_0, \sigma_0^2)$$

after the time $t \geq 1$ then,

$$\underbrace{Y_t = F_t\theta_t + \upsilon_t}_{\text{observation equation}} , \quad \upsilon_t \sim N(0, V_t),$$

$$\underbrace{\theta_t = G_t\theta_{t-1} + \omega_t}_{\text{state equation}}, \quad \omega_t \sim N(0, W_t)$$



where Gt and Ft are known matrices (p × p and m × p) and the disturbance terms υt and ωt are two independent Gaussian random vectors, and θt is a Markov chain (West, M., & Harrison, J., 1989).

Causal Discovery Study Case #2

Methodological Abstract

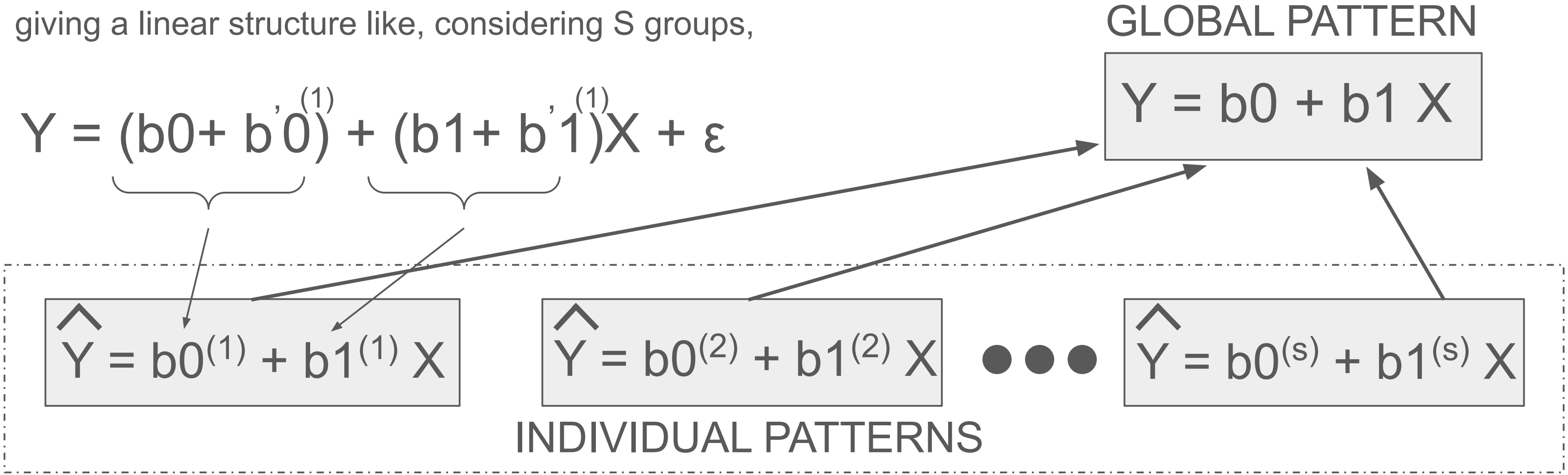
1. Calculate the possible DAG with MDM-IPA (PC + Score-based)
2. Calculate the Regression Dynamic of the estimated DAG via DLM
   a. Considering levels of hierarchy:
      i. COMMON BRAIN
      ii. SUBJECT
      iii. SESSION
3. Based on the joint Log Predictive Likelihood (LPL), difference graphical structures are calculated for each level.

THE STATISTICAL DISTRIBUTION THAT CONNECTS ALL ESTIMATED EFFECTS AND STRUCTURE WITH THE MULTIVARIATE NORMAL DIST.

Review of HIERARCHICAL MODELING (including random effects)

It means that the average process will be estimate (GLOBAL EFFECT) + each individual dynamic (LOCAL EFFECT), composing the total process mean as blocks.

For instance, the response variable (Y) is explained by (X)
giving a linear structure like, considering S groups,



GLOBAL PATTERN

$$Y = b0 + b1\,X$$

$$Y = \underbrace{(b0+ b'0^{(1)})} + \underbrace{(b1+ b'1^{(1)})}X + \varepsilon$$

$$\hat{Y} = b0^{(1)} + b1^{(1)}\,X \qquad \hat{Y} = b0^{(2)} + b1^{(2)}\,X \quad \bullet\bullet\bullet \quad \hat{Y} = b0^{(s)} + b1^{(s)}\,X$$

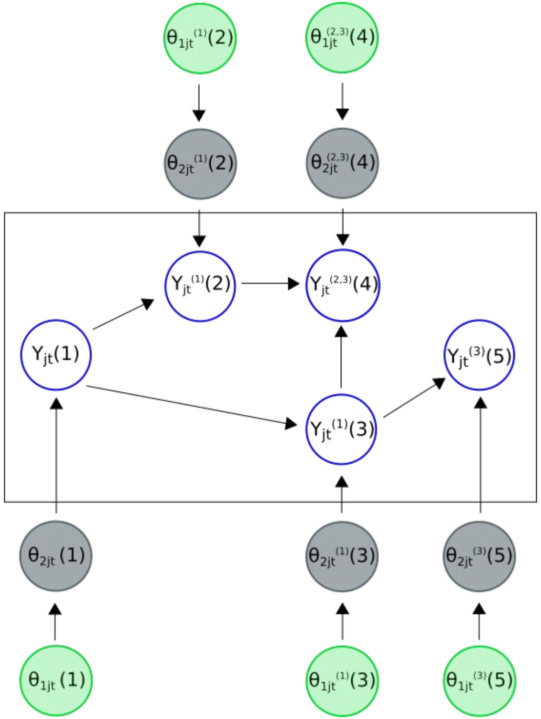
INDIVIDUAL PATTERNS

$\Sigma b0^{(i)}=0$ and $\Sigma b1^{(i)}=0$

http://mfviz.com/hierarchical-models/

The Brain Group Hierarchical Dynamic Model (BGHDM) for 3 levels [brain, subject and session], global dynamic (GS) with individual patterns (contributions - VTS + CS + IS), through a fix Network causal structure and a dynamic strength:

- Group-structure (GS) -> Treatment
- Virtual-typical-subject (VTS) -> Group
- Common-structure (CS) -> Subject
- Individual-structure (IS) -> Session

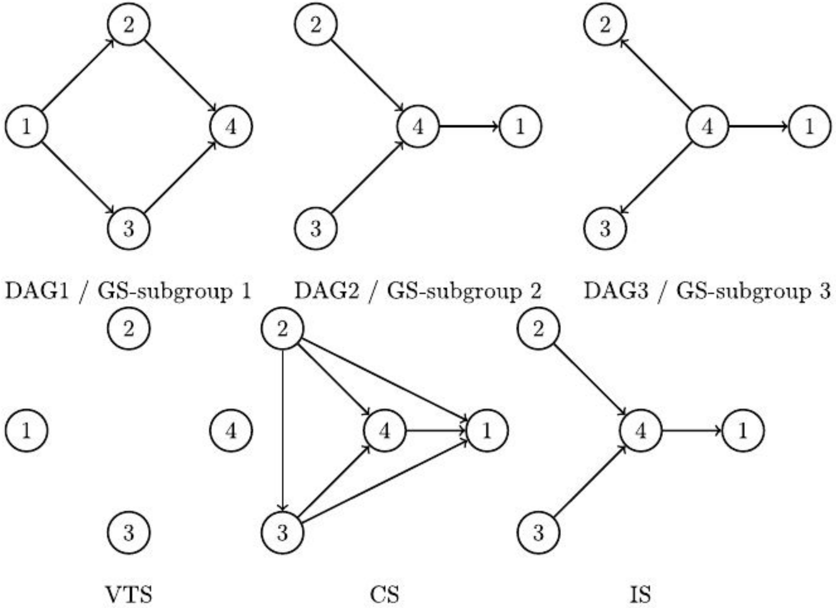


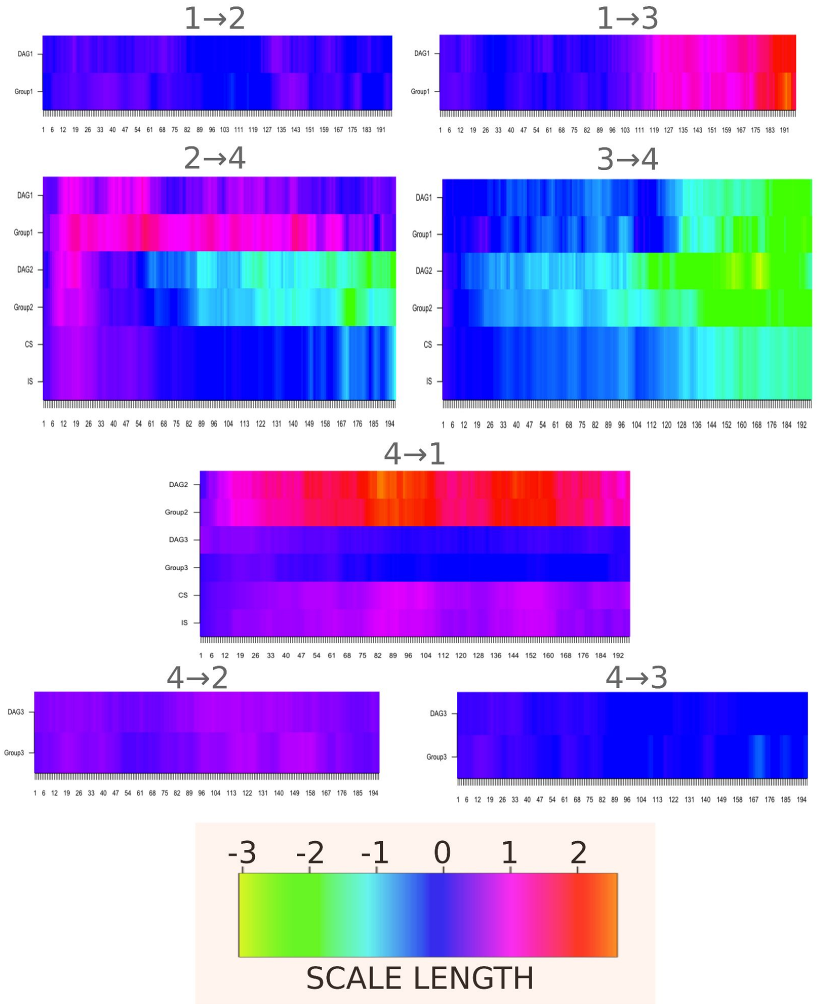
$$\mathrm{LPL}_i(m) = \sum_{r=1}^{n} \sum_{t=1}^{T} \log p_{tr}(y_{it}(r)|\mathbf{y}_i^{t-1}, Pa_i(r), m),$$

The Network is estimated as a the joint Log Predictive Likelihood (LPL)

SYNTHETIC DATA

It was simulated data from 3 different DAGs (DAG1, DAG2 and DAG3 in Fig. 2), using 10 subjects for each DAG, and considering 4 nodes and 197 time points





Then, NETWORK STRUCTURE + DYNAMIC STRENGTH, with a decomposition of **dissimilarity matrix**, d, to embed projection reduction space visualization, via MDM-Cluster analysis.

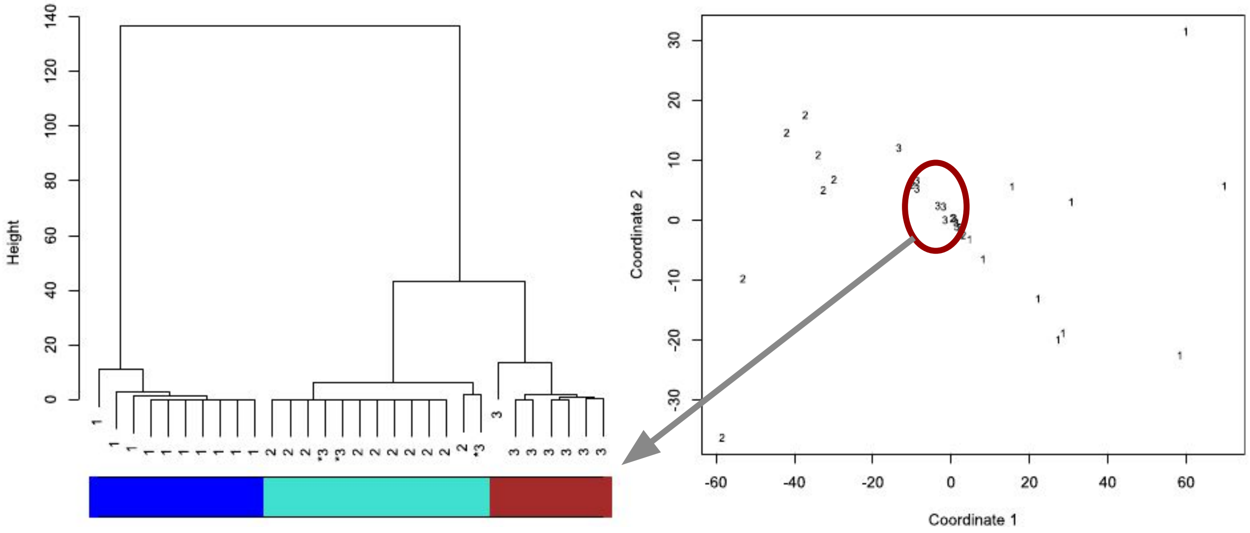
Using this method after estimating

- Individual networks (IS), Mi and Mj, and
- Pairwise group network (GS), mij, for every pair of subjects i and j,

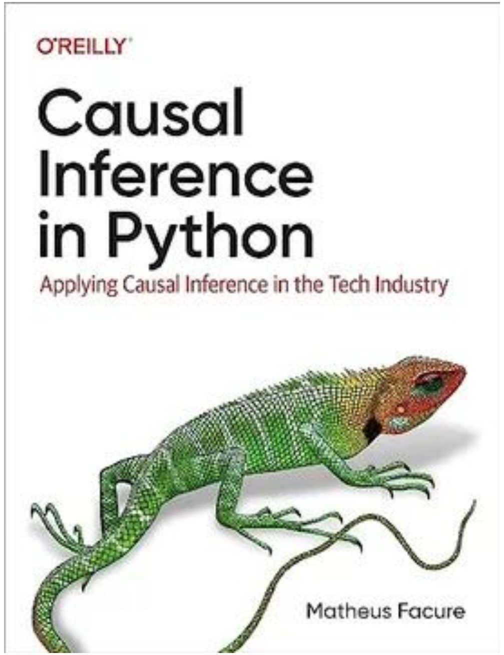
a separation measure, d(i, j).

$$d(i, j) = \mathrm{LPL}_i(M_i) + \mathrm{LPL}_j(M_j) - \mathrm{LPL}_i(m_{ij}) - \mathrm{LPL}_j(m_{ij}),$$

Synthetic data example,



FURTHER INVESTIGATION

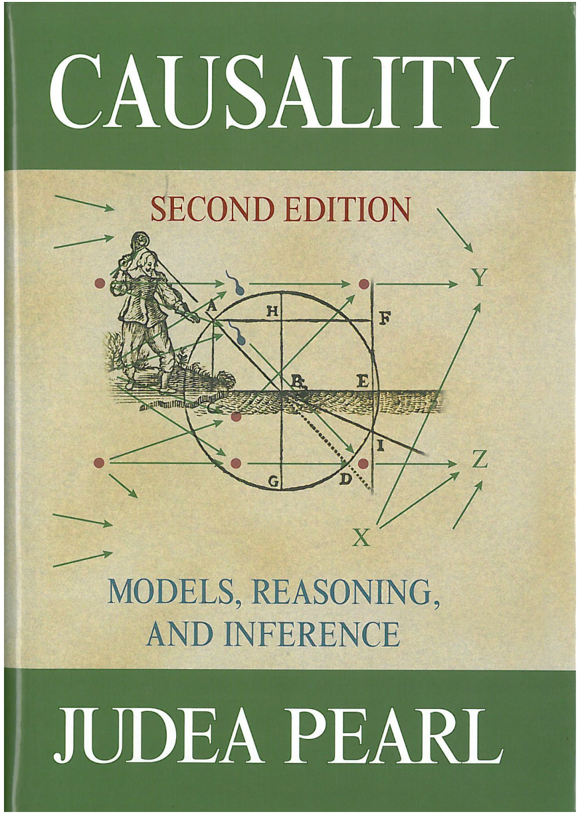


O'REILLY

Causal Inference in Python

Applying Causal Inference in the Tech Industry

Matheus Facure

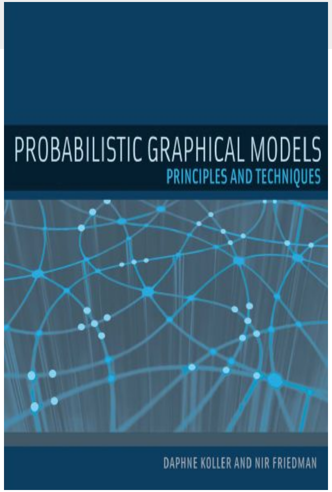


CAUSALITY

SECOND EDITION

MODELS, REASONING, AND INFERENCE

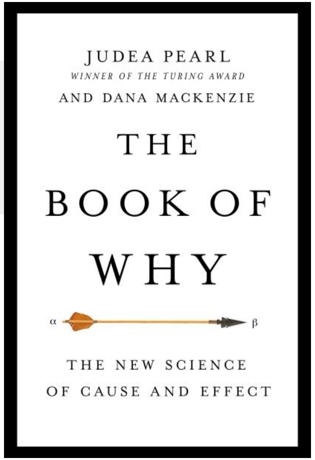
JUDEA PEARL



PROBABILISTIC GRAPHICAL MODELS

PRINCIPLES AND TECHNIQUES

DAPHNE KOLLER AND NIR FRIEDMAN



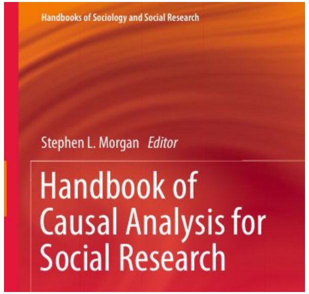
JUDEA PEARL

WINNER OF THE TURING AWARD

AND DANA MACKENZIE

THE BOOK OF WHY

THE NEW SCIENCE OF CAUSE AND EFFECT



Handbooks of Sociology and Social Research

Stephen L. Morgan *Editor*

Handbook of Causal Analysis for Social Research

Causal AI: An Introduction

udemy

UNIVERSIDAD
DE ATACAMA

Diego Nascimento
diego.nascimento@uda.cl
Ph.D. in Statistics (USP/UFSCar)