BAYESIAN NETWORKS MEET OBSERVATIONAL DATA

GILLES KRATZER, PHD

PROF. REINHARD FURRER

University of Zurich UZH

http://r-bayesian-networks.org/

gilles.kratzer@gmail.com

reinhard.furrer@math.uzh.ch

USER! CONFERENCE, ONLINE 07.07.2021

SCHEDULE



11:15

| Brief introduction on Additive Bayesian modelling

12:00

| Hands-on exercise: first analysis

13:00

| More advanced features of Additive Bayesian modelling

13:20

| Hands-on exercice: advanced features

13:40

| Wrap-up and discussion

13:45

MATERIAL

Material for the workshop

https://gilleskratzer.github.io/ABN-UseR-2021/

More ressources about ABN

http://r-bayesian-networks.org/

MOTIVATIONAL EXAMPLE: CREDIT CARD FRAUD DETECTION PREDICTION

Credit Card Fraud Detection
Using Bayesian and Neural Networks

Sam Maes Karl Tuyls Bram Vanschoenwinkel
Bernard Manderick
Vrije Universiteit Brussel - Department of Computer Science
Computational Modeling Lab (COMO)
Pleinlaan 2
B-1050 Brussel, Belgium
{sammaes@,ktuyls@,bvschoen@,bernard@arti.}vub.ac.be

Abstract

This paper discusses automated credit card fraud detection by means of machine learning. In an era of digitalization, credit card fraud detection is of great importance to financial institutions. We apply two machine learning techniques suited for reasoning under uncertainty: artificial neural networks and

do the fraud detection. After a process of learning, the program is supposed to be able to correctly classify a transaction it has never seen before as fraudulent or not fraudulent, given some features of that transaction.

The structure of this paper is as follows: first we introduce the reader to the domain of credit card fraud detection. In Sections 3 and 4 we briefly ex-

MOTIVATIONAL EXAMPLE: CREDIT CARD FRAUD DETECTION PREDICTION

Credit Card Fraud Detection Using Bayesian and Neural Networks

Sam Maes Karl Tuyls Bram Vanschoenwinkel

experiment	±10% false pos	±15% false pos
ANN-fig 2(a)	60% true pos	70% true pos
ANN-fig 2(a)	47% true pos	58% true pos
ANN-fig 2(c)	60% true pos	70% true pos
BBN-fig 2(e)	68% true pos	74% true pos
BBN-fig 2(g)	68% true pos	74% true pos

Table 1: This table compares the results achieved with ANN and BBN, for a false positive rate of respectively 10% and 15%.

Abstract

This paper discusses ... tection by means of ... of digitalization, cre... great importance to ... two machine learning techniques suited for reasoning under uncertainty: artificial neural networks and

...rocess of learning, ...e to correctly clas-... ...n before as fraud-... ...he features of that ... ...s follows: first we introduce the reader to the domain of credit card fraud detection. In Sections 3 and 4 we briefly ex-

MOTIVATIONAL EXAMPLE: VETERINARY EPIDEMIOLOGY DATA VISUALISATION

Contents lists available at SciVerse ScienceDirect

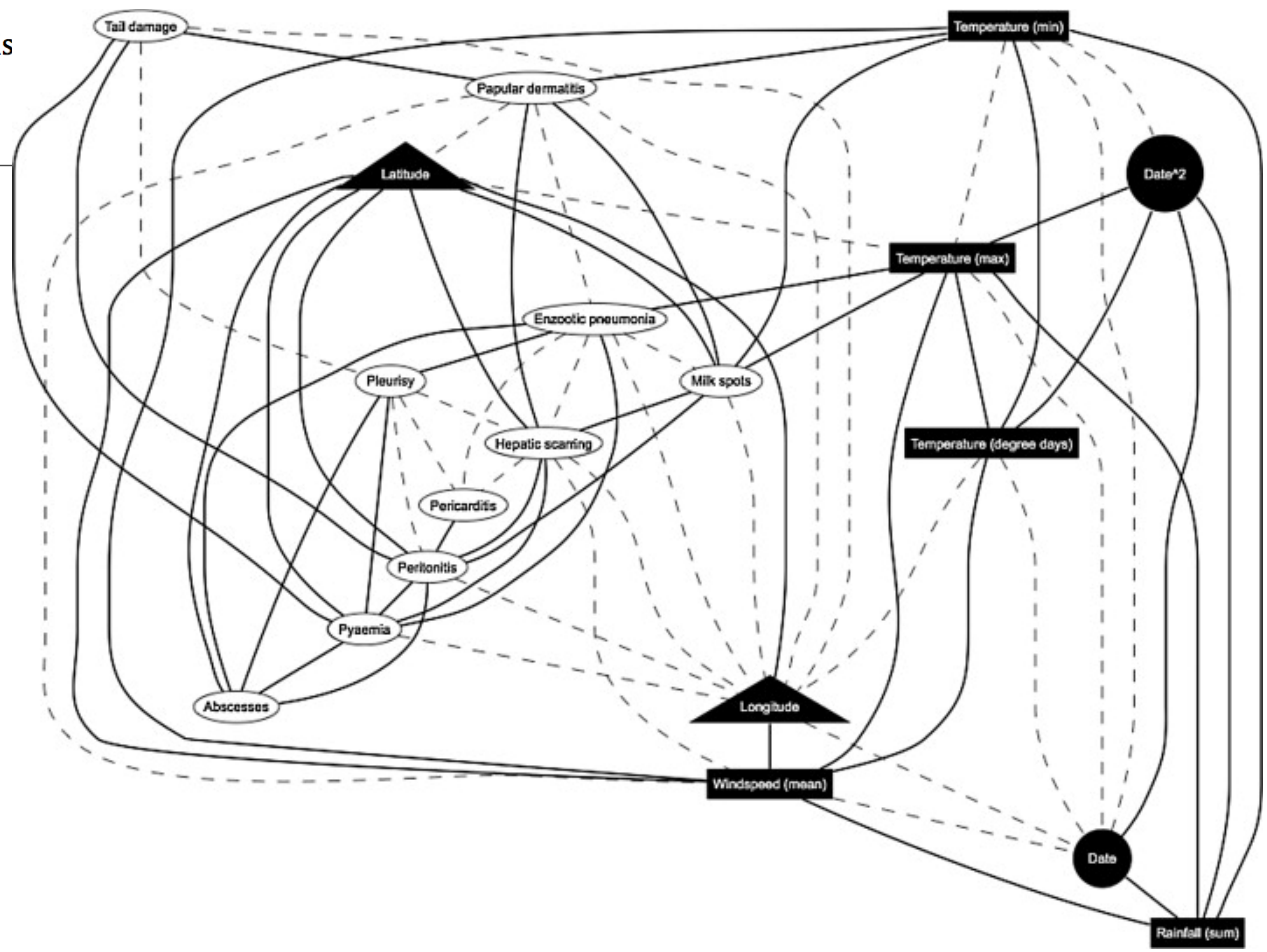
Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Using Bayesian networks to explore the role of weather as a potential determinant of disease in pigs

B.J.J. McCormick[a], M.J. Sanchez-Vazquez[b], F.I. Lewis

[a] Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA
[b] OIE Organisation Mondiale de la Santé Animale, 12, rue de Prony, 75017 Paris, France
[c] Section of Epidemiology, University of Zurich, Zurich, Switzerland



MOTIVATIONAL EXAMPLE: SOCIAL SCIENCES DATA INTERPRETATION

Discovering complex interrelationships between socioeconomic status and health in Europe: A case study applying Bayesian Networks

Javier Alvarez-Galvez [a, b, *]

[a] Loyola University Andalusia, Department of International Studies, Campus de Palmas Altas, Faculty of Political Sciences and Law, Seville 41014, Spain
[b] Complutense University of Madrid, Department of Sociology IV (Research Methodology and Communication Theory), Campus de Somosaguas, Faculty of Political

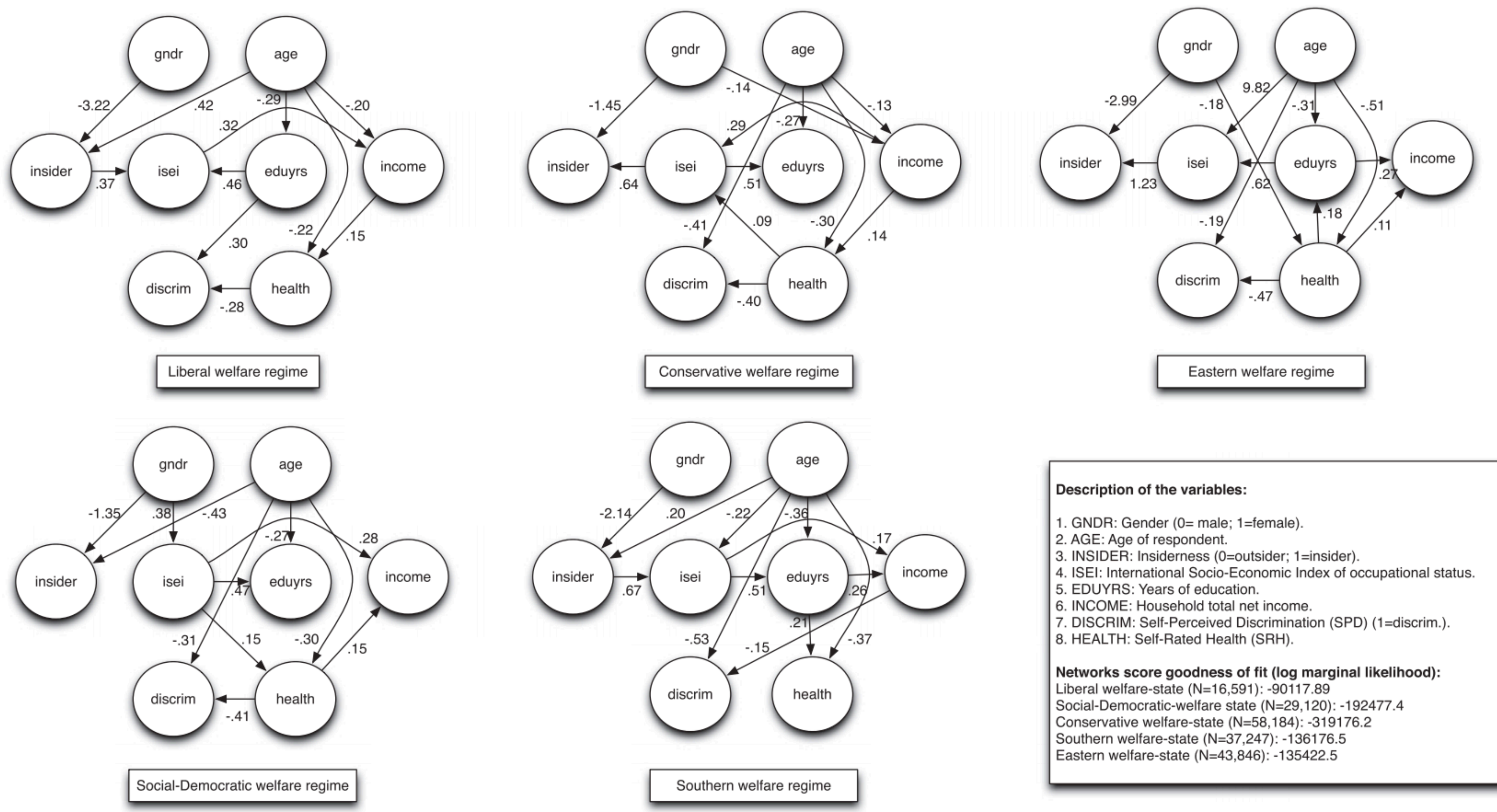
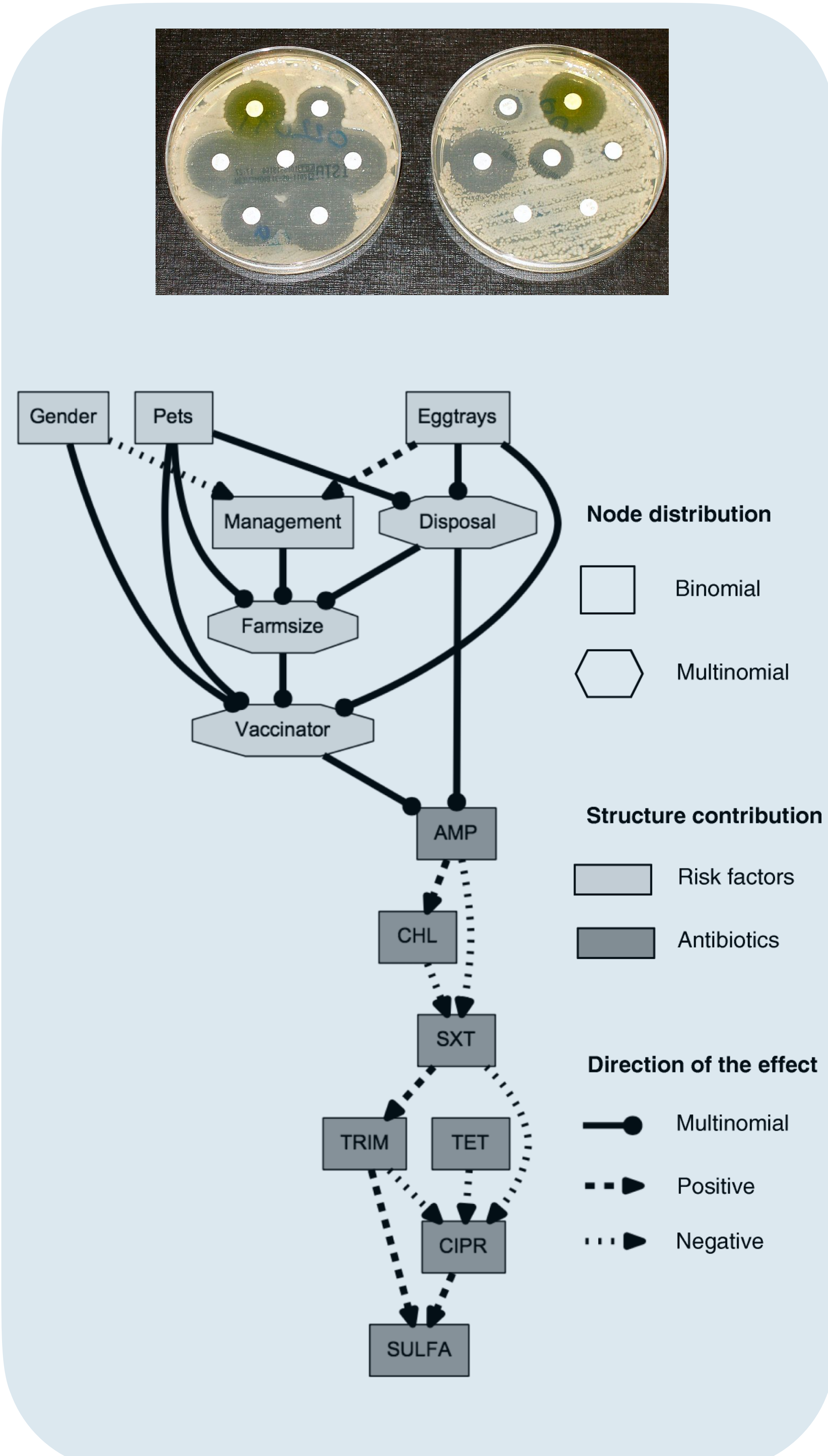
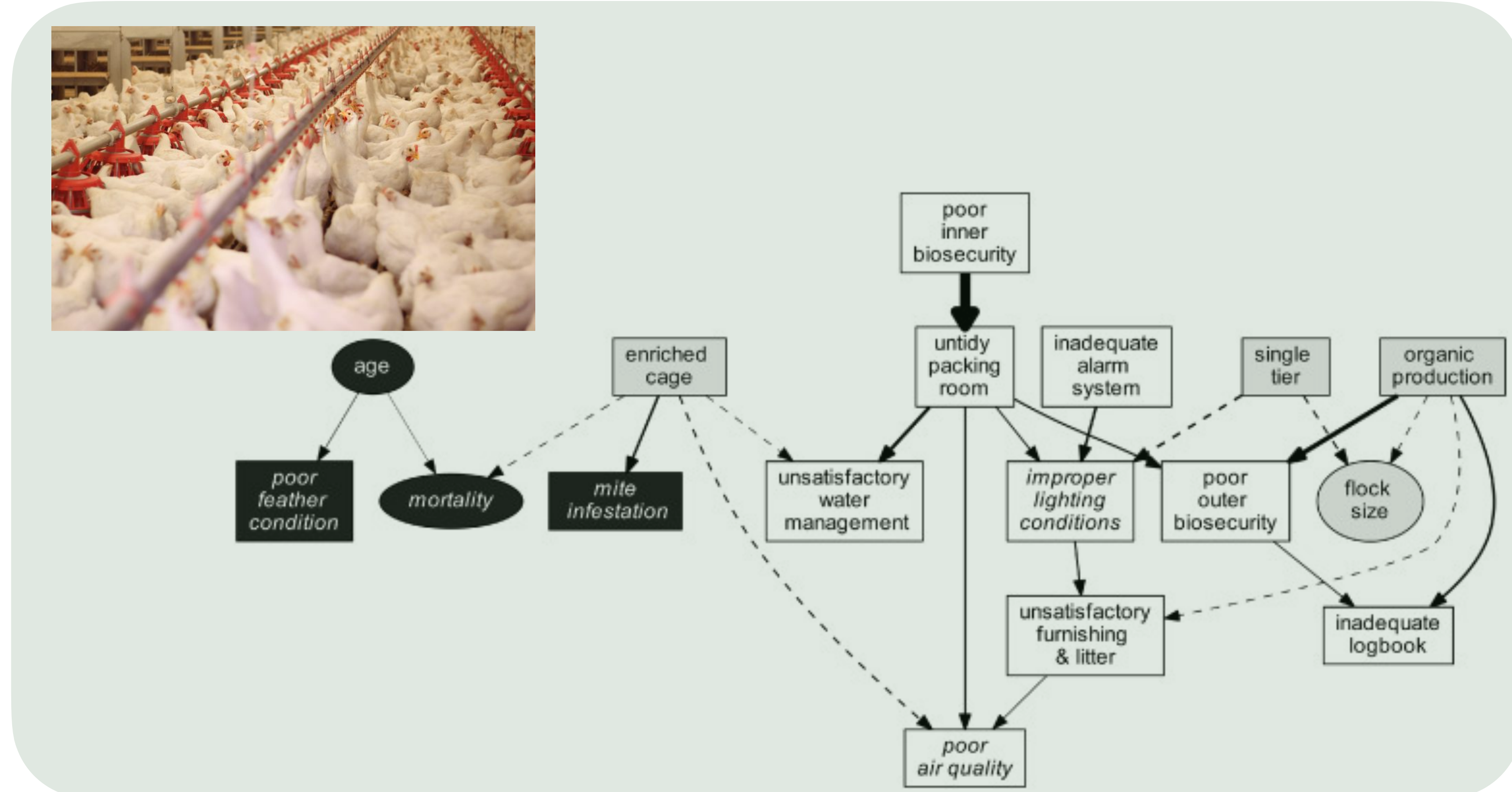
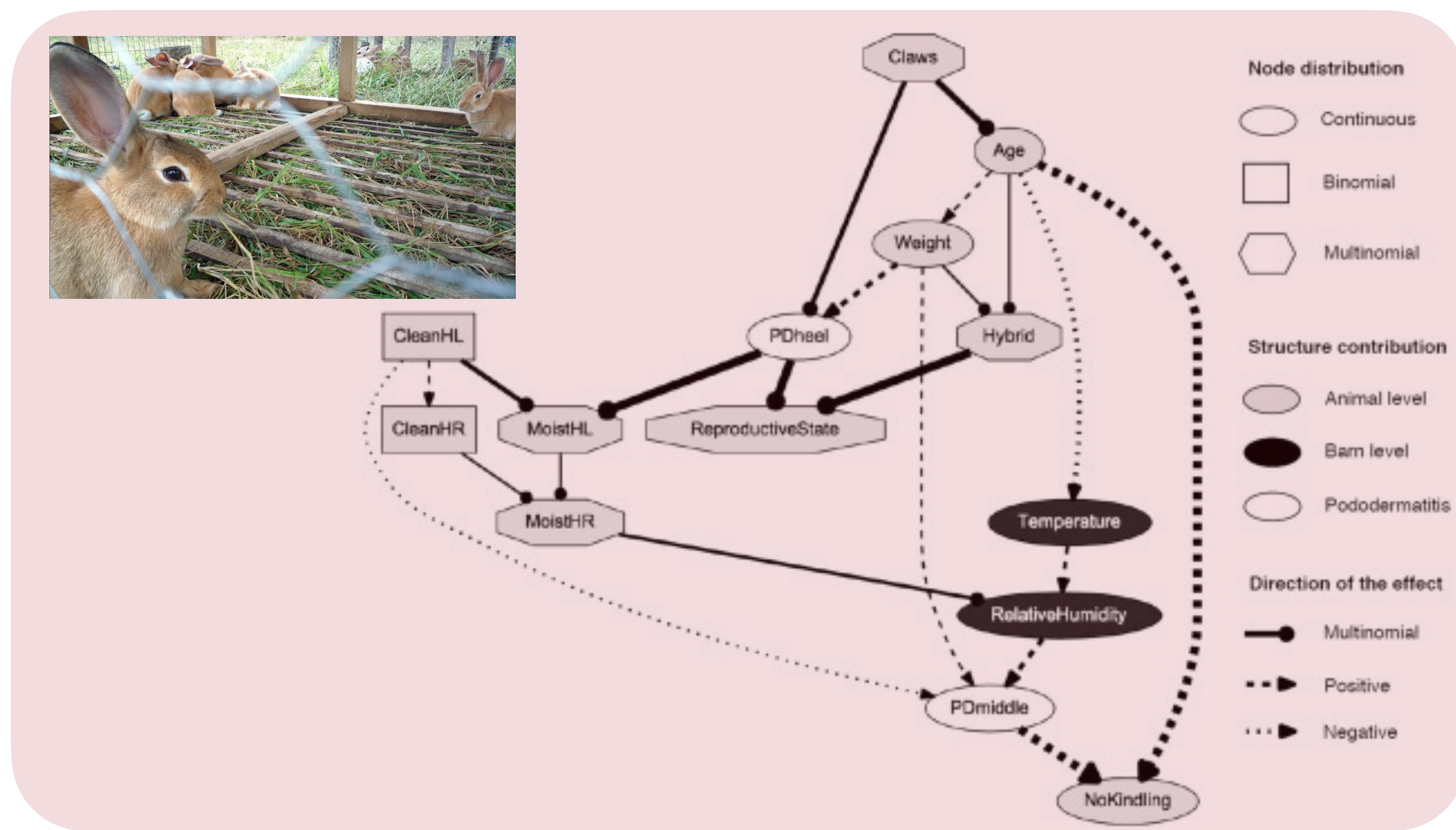


Fig. 1. Bayesian networks describing interrelationships between SES and health in five European welfare states.

EXAMPLE OF SYSTEMS EPIDEMIOLOGY DATA ANALYSED WITH ABN







University of Zurich[UZH]

EXAMPLE OF SYSTEMS EPIDEMIOLOGY DATA ANALYSED WITH ABN

Anti-microbial resistance

- Multi-drug resistant Salmonella isolates (7 a tics)
- 43 poultry farms in Uganda
- Risk factors: Management practice, farm size, etc …

MULTIPLE OUTCOMES

Hartnack and al. (2019) in BMC

Animal welfare

- Welfare control programme after ban of batt cage
- 193 different poultry farms in Sweden
- Welfare status depends on many inter-related varia
- Risk factors: Management practice, weather, etc …

MULTIDIMENSIONAL

Comin and al. (2019) in PVM

Technopathy in rabbit

- Longitudinal study on Pododermatitis in rabb
- 3 commercial farms in Switzerland
- Group housing on litter and plastic slats
- Main interest: Healing process

HYPOTHESIS GENERATION

Ruchti and al. (2019) in PVM

BAYESIAN NETWORKS IN THE MACHINE LEARNING WORLD

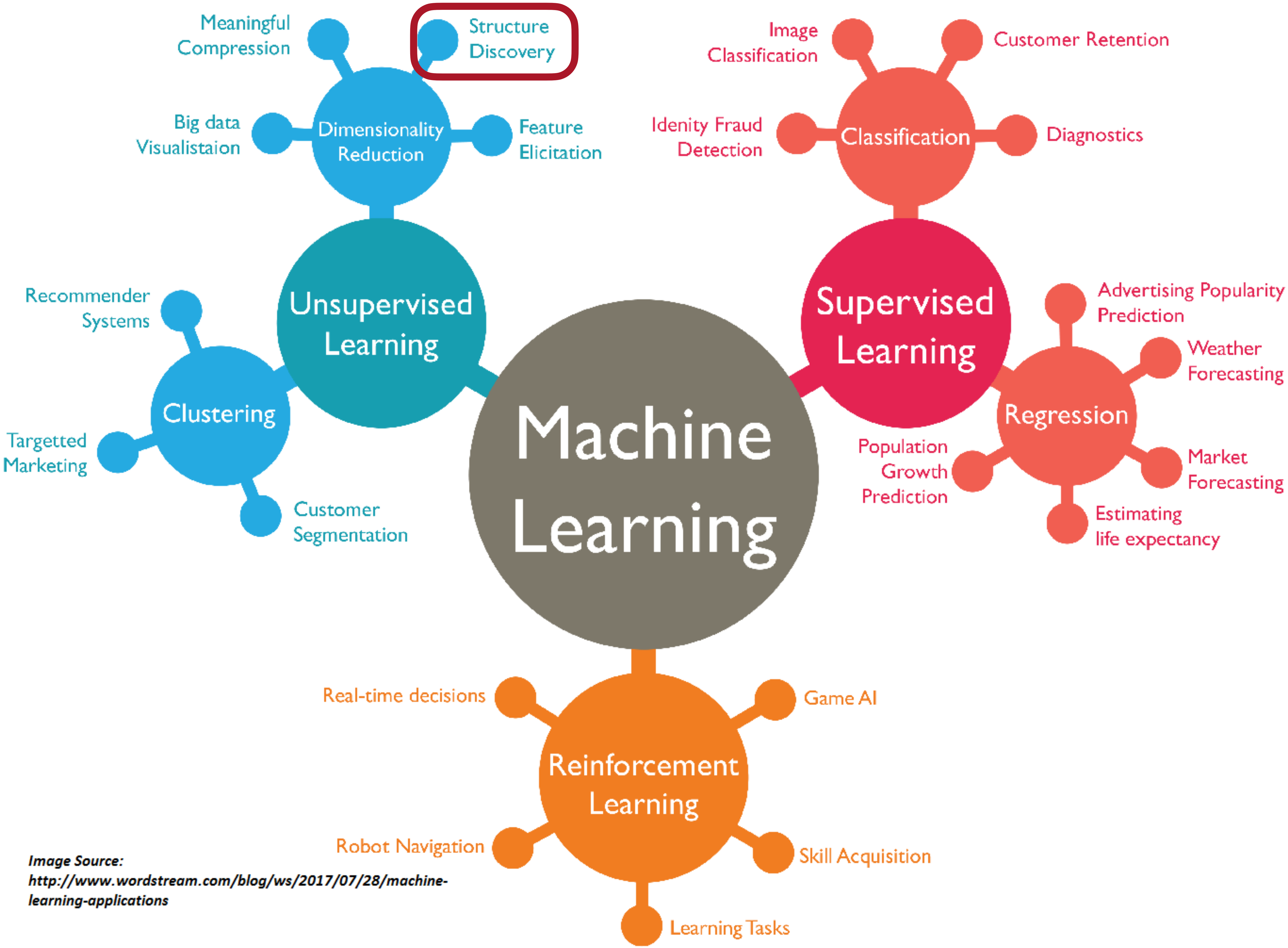


Image Source: http://www.wordstream.com/blog/ws/2017/07/28/machine-learning-applications

Objective of the workshop:

How to learn Bayesian networks from observational data?

Objectif of the workshop:

select

How to ~~learn~~ Bayesian networks from observational data?

Bayesian Networks are defined by two elements:

Network structure:

Directed Acyclic Graph (DAG): G = (V, A)

in which each node $v_i \in V$ corresponds to a random variable X_i

Probability distribution:

Probability distribution X with parameters Θ, which can be factorised into smaller local probability distributions according to the arcs $a_{ij} \in A$ present in the graph.

A BN encodes the factorisation of the joint distribution

$$P(\mathbf{X}) = \prod_{j=1}^{n} P(X_j \mid \mathbf{Pa}_j, \Theta_j), \text{ where } \mathbf{Pa}_j \text{ is the set of parents of } X_j$$

ABN WORKFLOW

1. From observational dataset deduce probabilistic model

   Epidemiological constrain: mixture of distributions

2. From probabilistic model deduce structure

EXPONENTIAL FAMILY

Observational dataset

1

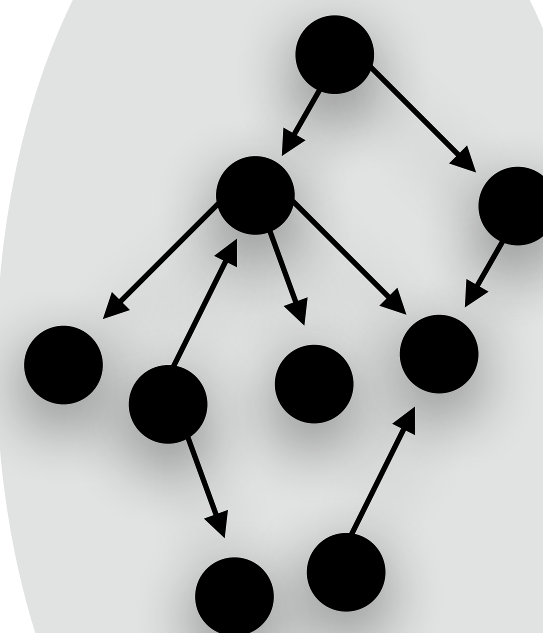
Probabilistic model

$$P(\mathbf{X}) = \prod_{j=1}^{n} P(X_j \mid \mathbf{Pa}_j, \Theta_j)$$

Computing directly

2

Network structure



COMBINATORIAL WALL

# Nodes	# DAGs	Inference
1 - 15 Nodes	$< 10^{41}$ DAGs	Exact inference
16 - 25 Nodes	$< 10^{100}$ DAGs	Exact inference possible
26 - 50 Nodes	$< 10^{400}$ DAGs	Approximate inference
51 - 100 Nodes	$< 10^{1700}$ DAGs	Approximate inference
101 - 1000 Nodes	$< 10^{100000}$ DAGs	(very) approximative inference

Approximations:

- limiting number of parents per node
- Decomposable scores/efficient algorithm
- Score equivalence

SOME ELEMENTS OF PROBABILITY THEORY

1 The conditional probability of A given B is: $P(A \mid B) = \dfrac{P(A, B)}{P(B)}$

2 Bayes theorem: $P(A \mid B) = \dfrac{P(B \mid A)P(A)}{P(B)}$

SOME ELEMENTS OF PROBABILITY THEORY

1 The conditional probability of A given B is: $P(A \mid B) = \dfrac{P(A, B)}{P(B)}$

2 Bayes theorem: $P(A \mid B) = \dfrac{P(B \mid A)P(A)}{P(B)}$

1 + 2

Let A, B and C non intersecting subsets of nodes in a DAG G

A is conditionally independent of B given C if: $P(A, B \mid C) = P(A \mid C)P(B \mid C)$

1 The conditional probability of A given B is: $P(A \mid B) = \dfrac{P(A, B)}{P(B)}$

2 Bayes theorem: $P(A \mid B) = \dfrac{P(B \mid A)P(A)}{P(B)}$

Let A, B and C non intersecting subsets of nodes in a DAG G

1 + 2

A is conditionally independent of B given C if: $P(A, B \mid C) = P(A \mid C)P(B \mid C)$

Theorem (Verma & Pearl, 1988): A is d-separated from B by C if, and only if, the joint distribution over all variables in the graph satisfies:

$$A \perp\!\!\!\perp_G B|C$$

Link between statistical statement (**conditionally independence**) and a graph propriety (**d-separation**)

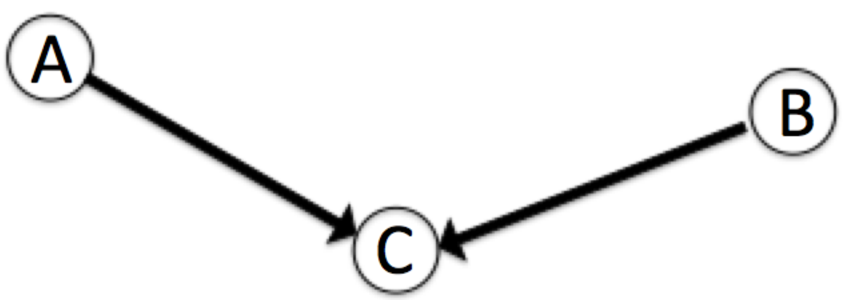
ELEMENT OF GRAPH THEORY

Let A, B and C non intersecting subsets of nodes in a DAG G

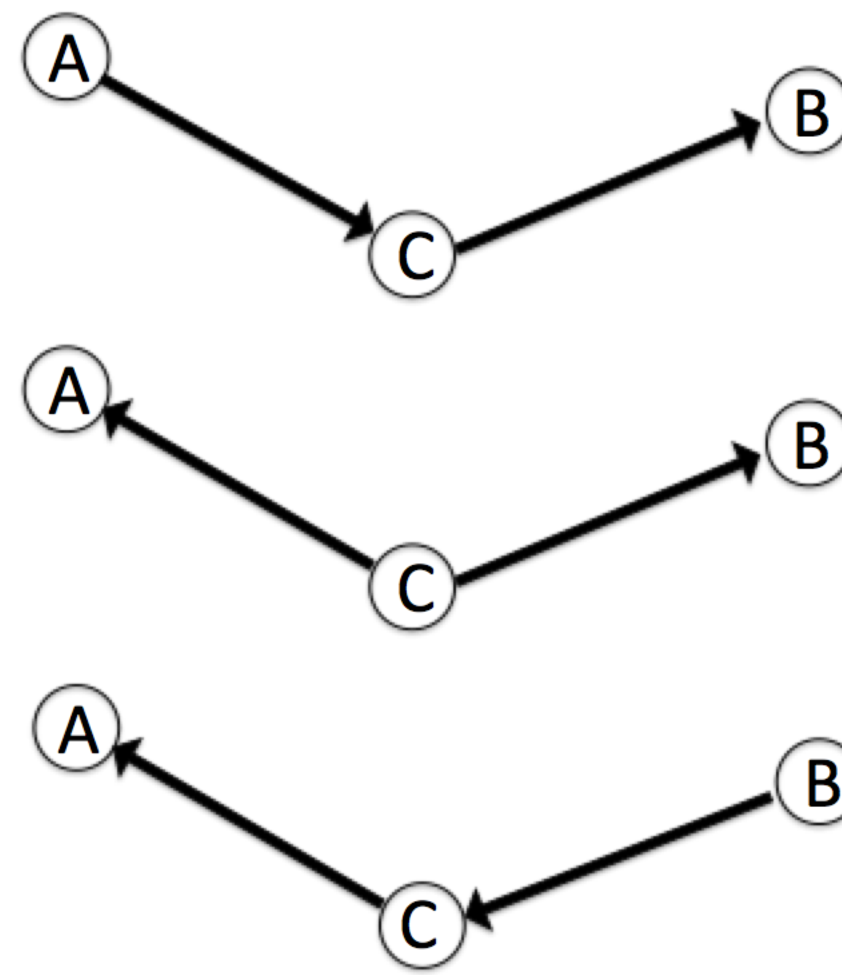
A is conditionally independent of B given C if: $A \perp\!\!\!\perp_P B|C$

$$P(A, B \mid C) = P(A \mid C)P(B \mid C)$$

$A \not\perp\!\!\!\perp_P B|C$



$A \perp\!\!\!\perp_P B|C$

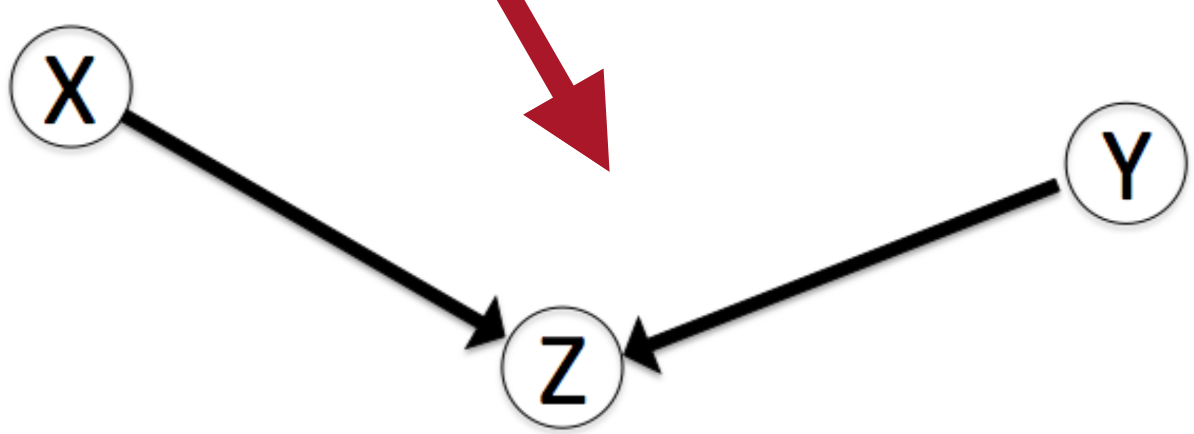


Constraint based algorithms

Learning independence relationships

$$P_{X \perp\!\!\!\perp Y | Z} < \alpha$$

$$X \perp\!\!\!\perp_{\mathcal{S}} Y | Z = X \perp Y | Z$$



Search-and-score algorithms

Maximum a posteriori score

Example of scoring functions:

- Bayesian versus ML scores
  - log marginal likelihood
  - Bayesian-Dirichlet (BDeu,BDs,BDe)
  - Bayesian Information Criterion (BIC)

LEARNING BAYESIAN NETWORKS VIA SEARCH-AND-SCORE ALGORITHMS

$$\mathcal{M} = (\mathcal{S}, \Theta_{\mathcal{M}})$$

Model selection

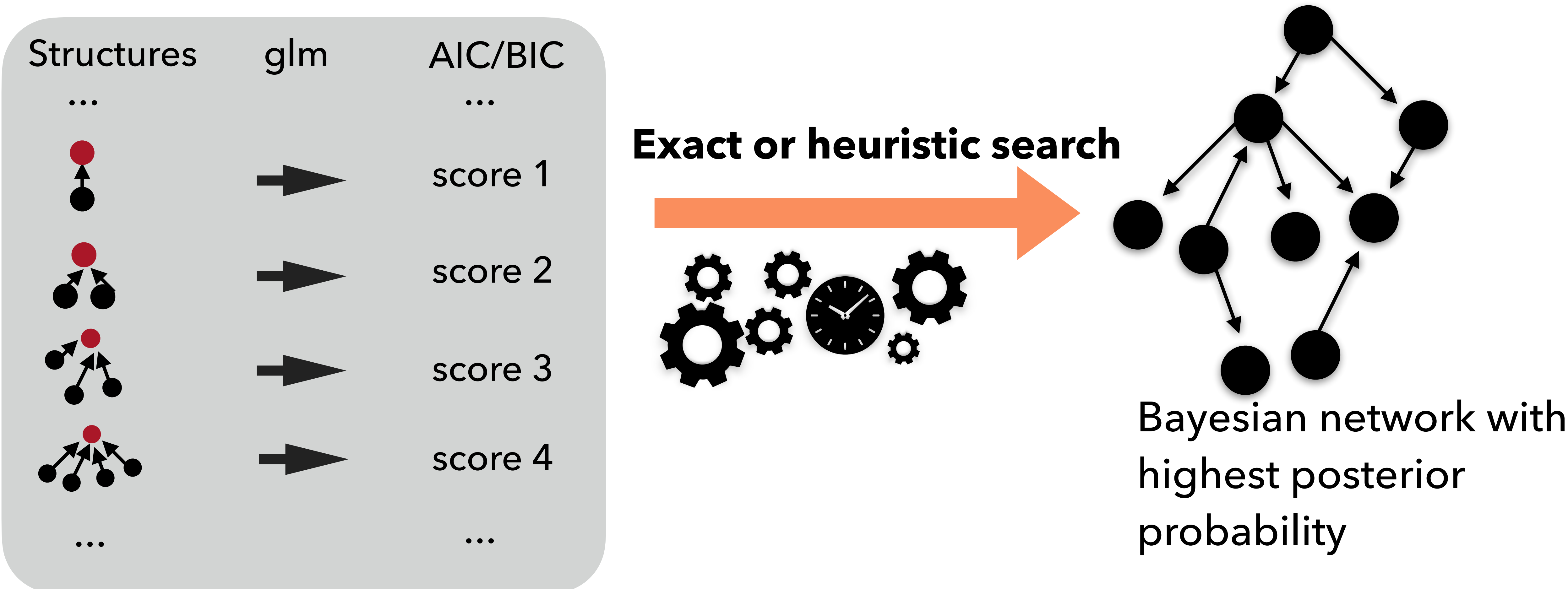
Structure learning

Parameter estimation

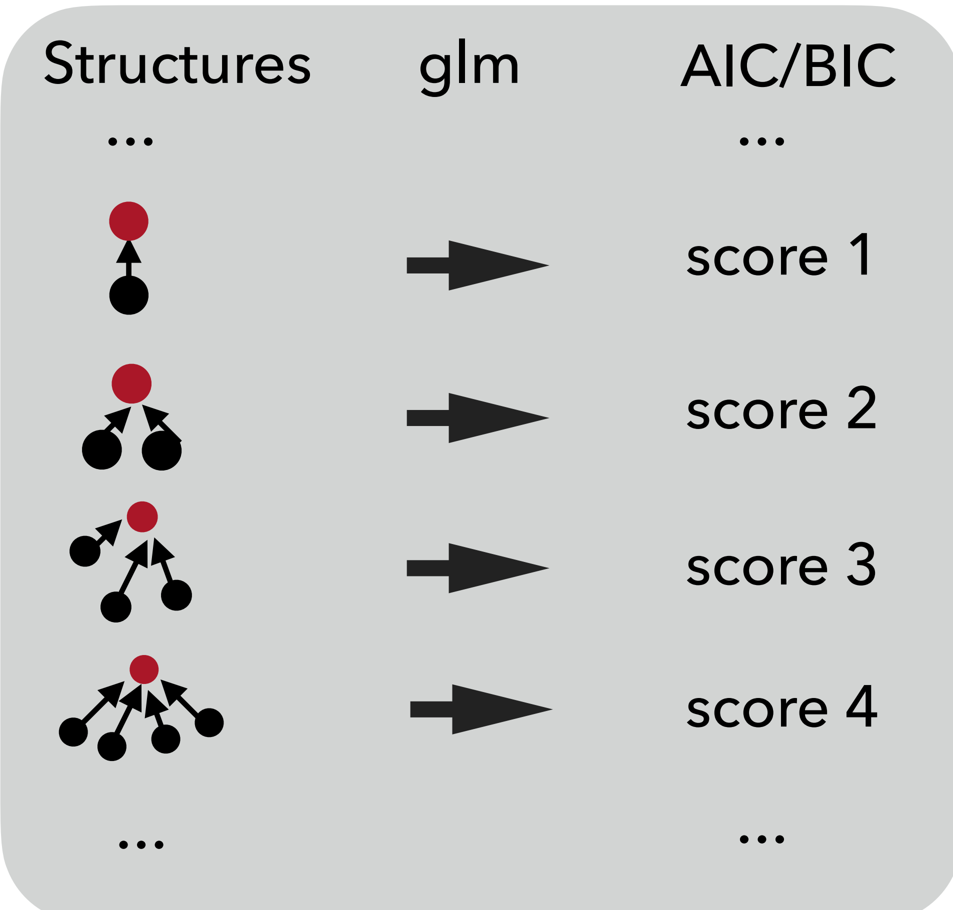
Parameter learning

$$P(\mathcal{M}|\mathcal{D}) = \underbrace{P(\Theta_{\mathcal{M}}, \mathcal{S}|\mathcal{D})}_{\text{model learning}} = \underbrace{P(\Theta_{\mathcal{M}}|\mathcal{S}, \mathcal{D})}_{\text{parameter learning}} \cdot \underbrace{P(\mathcal{S}|\mathcal{D})}_{\text{structure learning}}$$

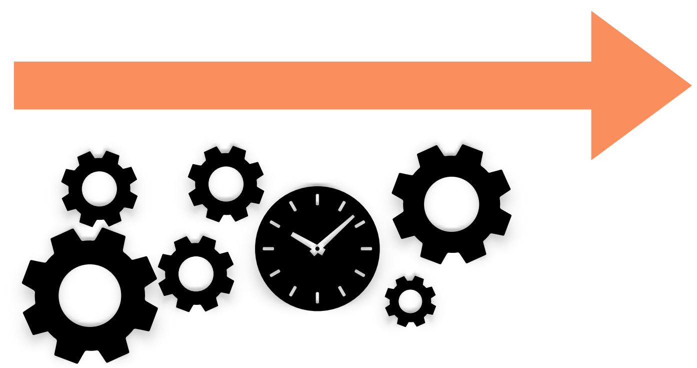
Search and score algorithm

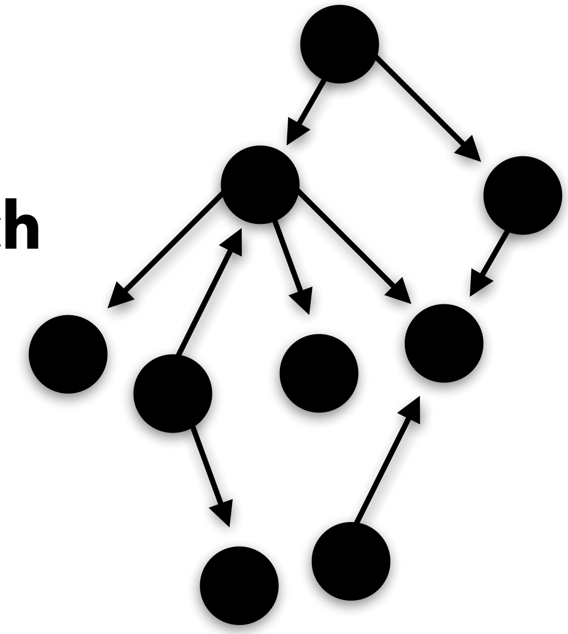


Search and score algorithm



Exact or heuristic search

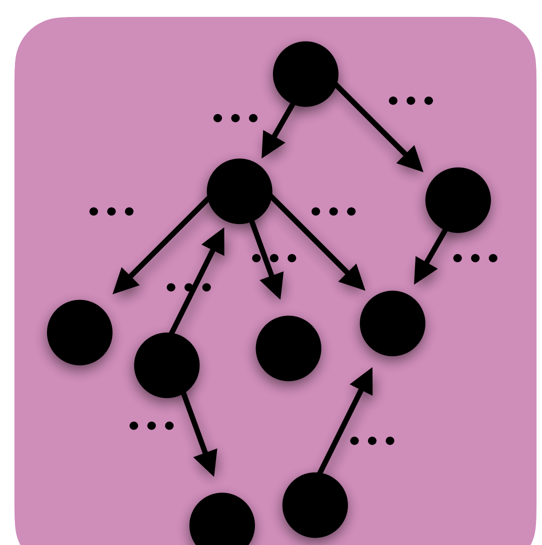




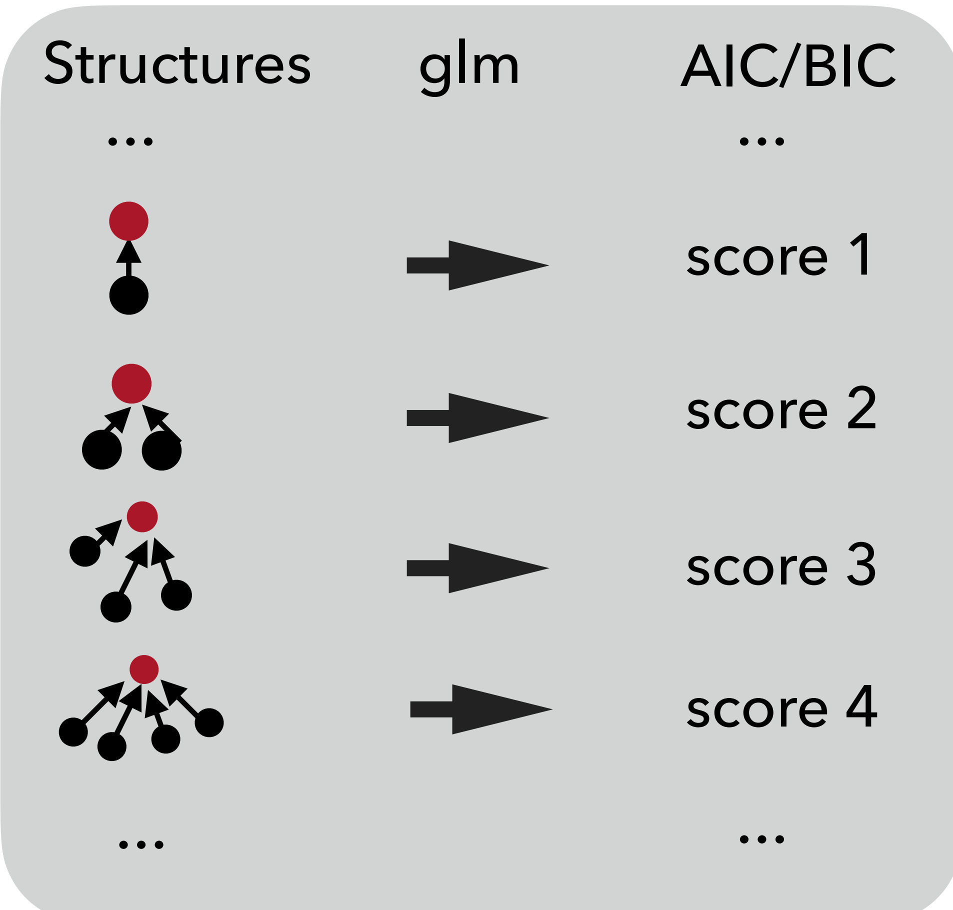
Bayesian network with highest posterior probability

Parameter estimation

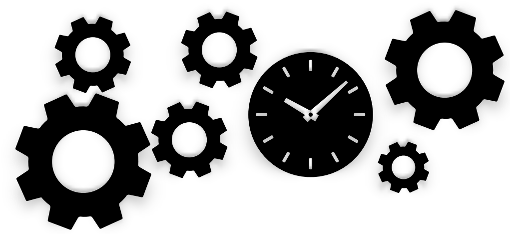
- compute marginal posterior density
- regression estimate



Search and score algorithm

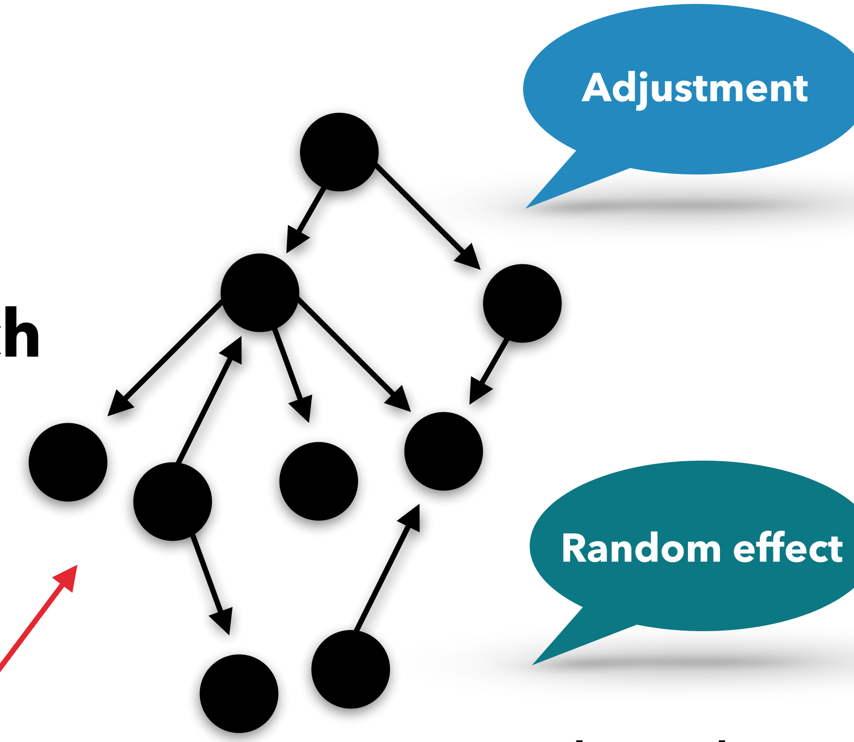


Exact or heuristic search



Causality!

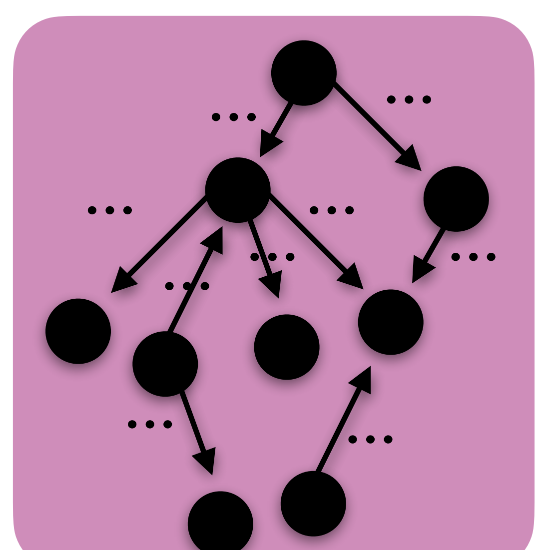
Ban/Retain structures



Bayesian network with highest posterior probability

Parameter estimation

- compute marginal posterior density
- regression estimate

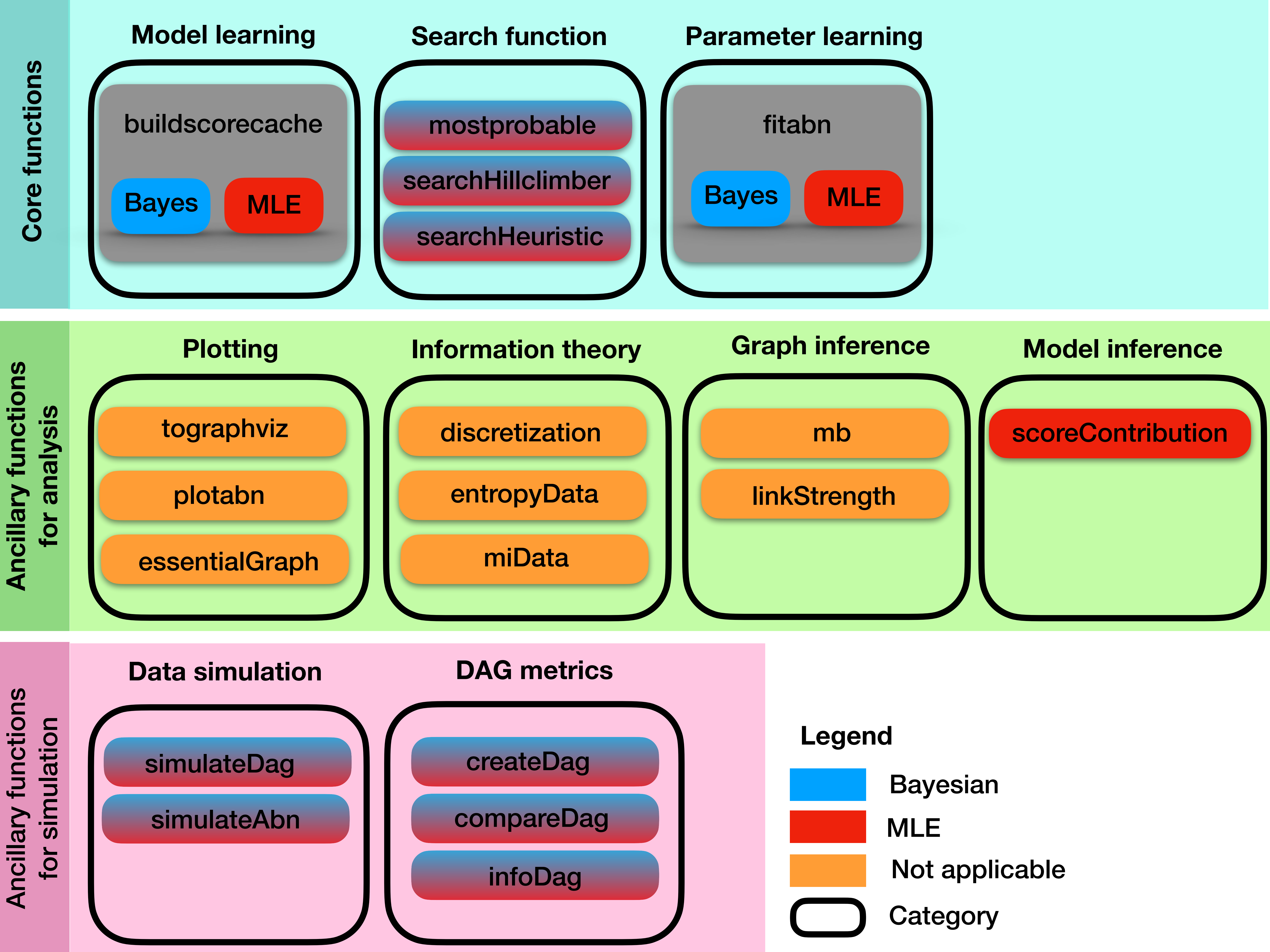


Using R

buildscorecache()

mostprobable()

fitabn()


Core functions
Model learning
buildscorecache
Bayes
MLE
Search function
mostprobable
searchHillclimber
searchHeuristic
Parameter learning
fitabn
Bayes
MLE
Ancillary functions for analysis
Plotting
tographviz
plotabn
essentialGraph
Information theory
discretization
entropyData
miData
Graph inference
mb
linkStrength
Model inference
scoreContribution
Ancillary functions for simulation
Data simulation
simulateDag
simulateAbn
DAG metrics
createDag
compareDag
infoDag
Legend
Bayesian
MLE
Not applicable
Category

SELECTED BIBLIOGRAPHY