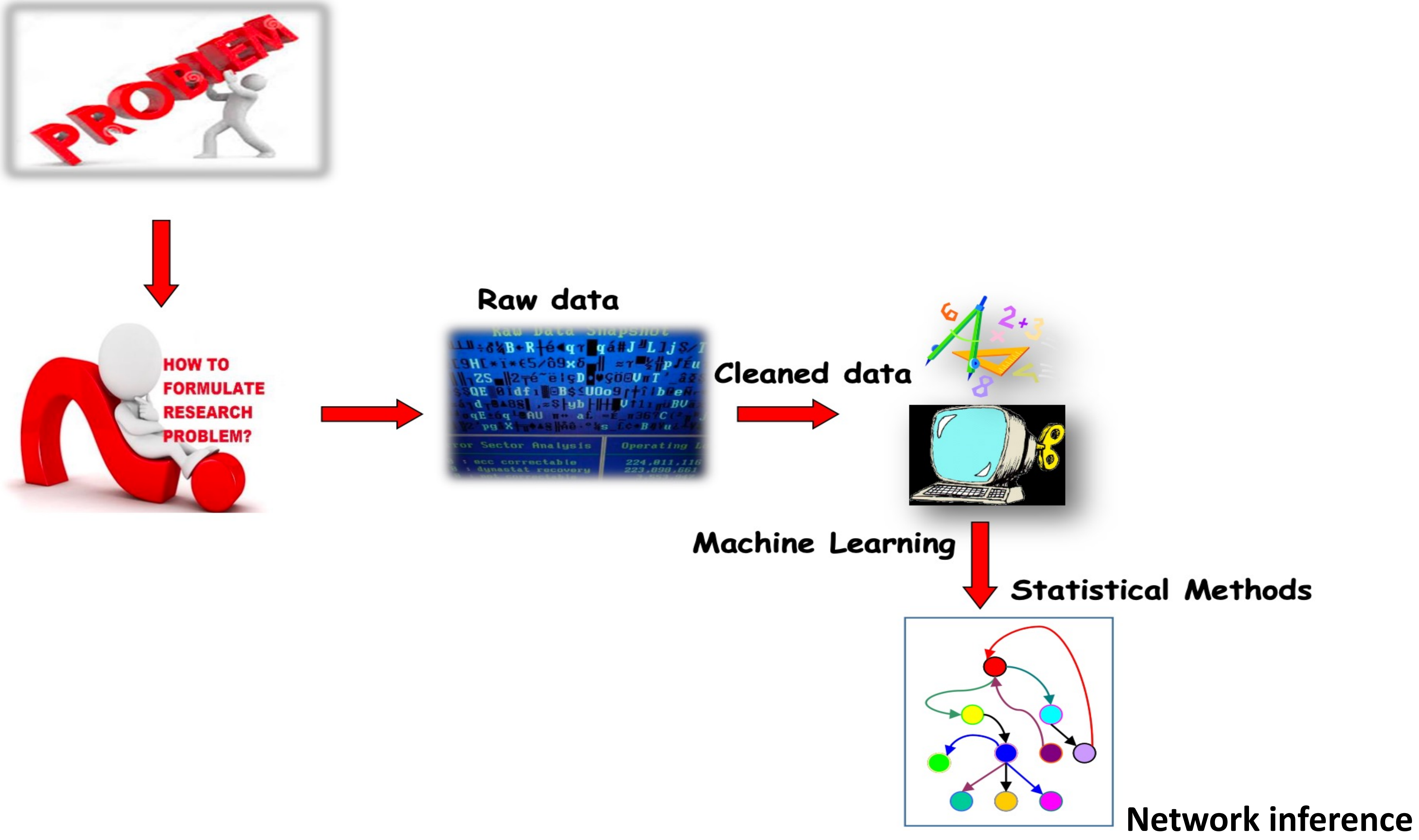
Bayesian networks

Mahdi Shafiee Kamalabad
Utrecht University

In a nutshell



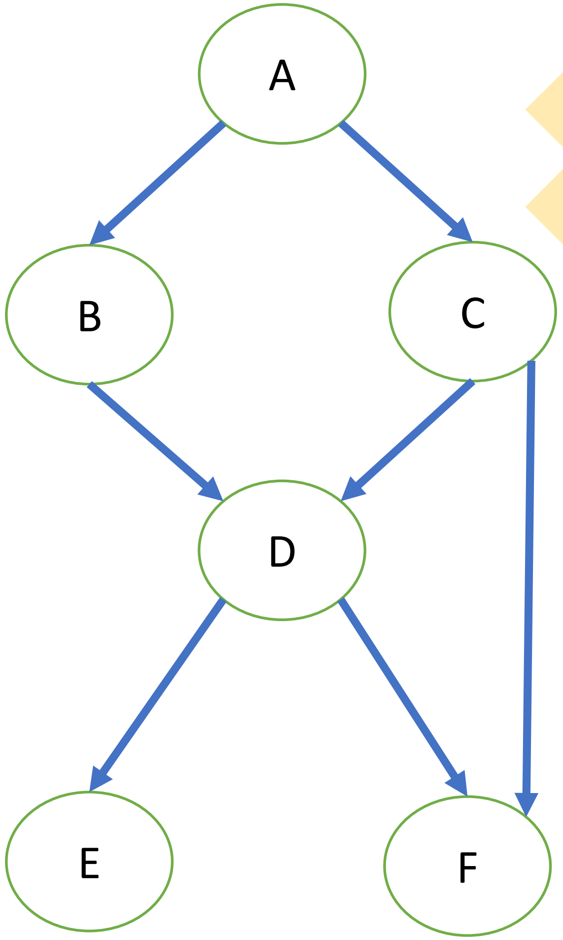
So, we aim to estimate the network structure, determining what depends on what and how, in the form of a network.

Discrete Data: Train Use Survey

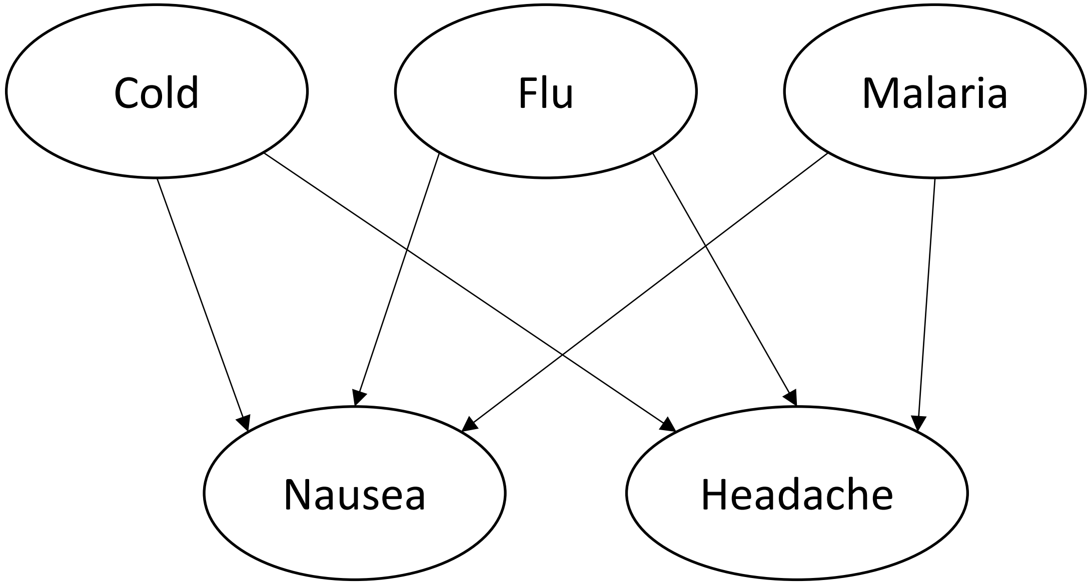
- "Age”   "Residence"   "Education"   "Occupation”   "Sex"   "Travel"
- "adult"   "big”   "high"   "emp”   "F”   "car"
- "adult"   "small"   "uni"   "emp"   "M"   "car"
- "adult"   "big"   "uni"   "emp"   "F"   "train"
- "adult"   "big"   "high"   "emp"   "M"   "car"
- "adult"   "big"   "high"   "emp"   "M"   "car"
- "adult"   "small"   "high"   "emp"   "F"   "train"
- "adult"   "big"   "high"   "emp"   "F"   "car"
- "young"   "big"   "uni"   "emp"   "F"   "train"

Bayesian networks (BNs)

- **Marriage between** graph theory and probability theory.
- **Nodes** represent variables and **edges** represent (conditional) dependence between variables.
- Bayesian Networks are **directed networks.**
- **If the directions of the dependencies are important in your project, go for a Bayesian network.**



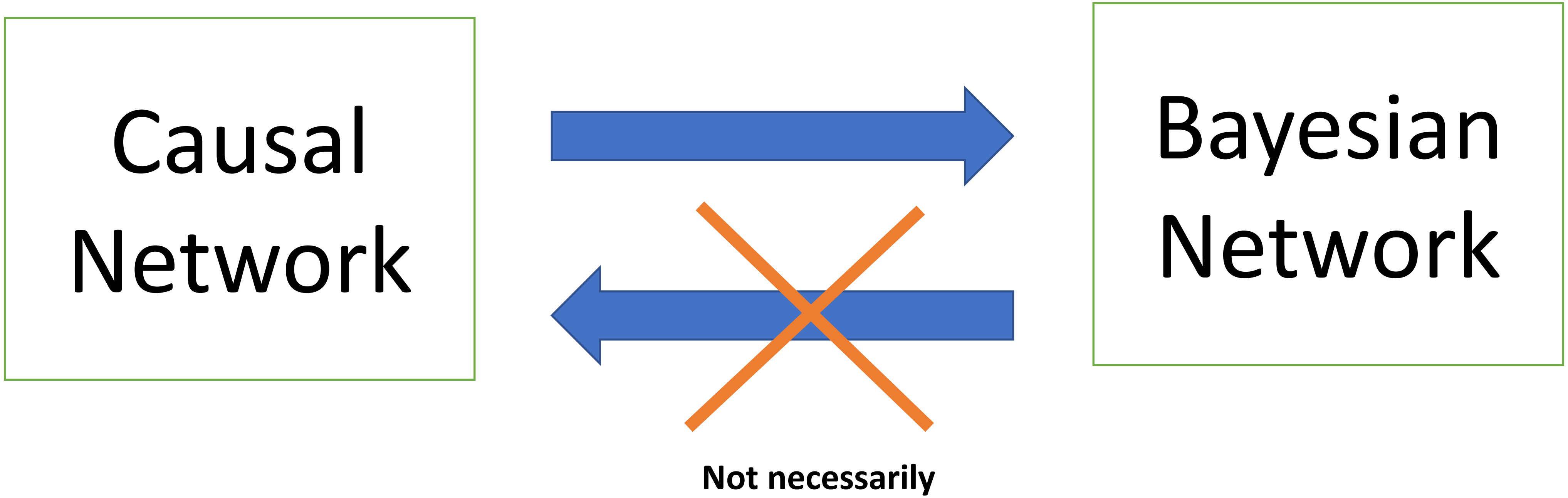
Example 1:



- A Bayesian network could represent the **probabilistic relationships** between **diseases** and **symptoms**.

- Bayesian network with causes (diseases) Cold, Flu, and Malaria and effects (symptoms) Nausea and Headache.

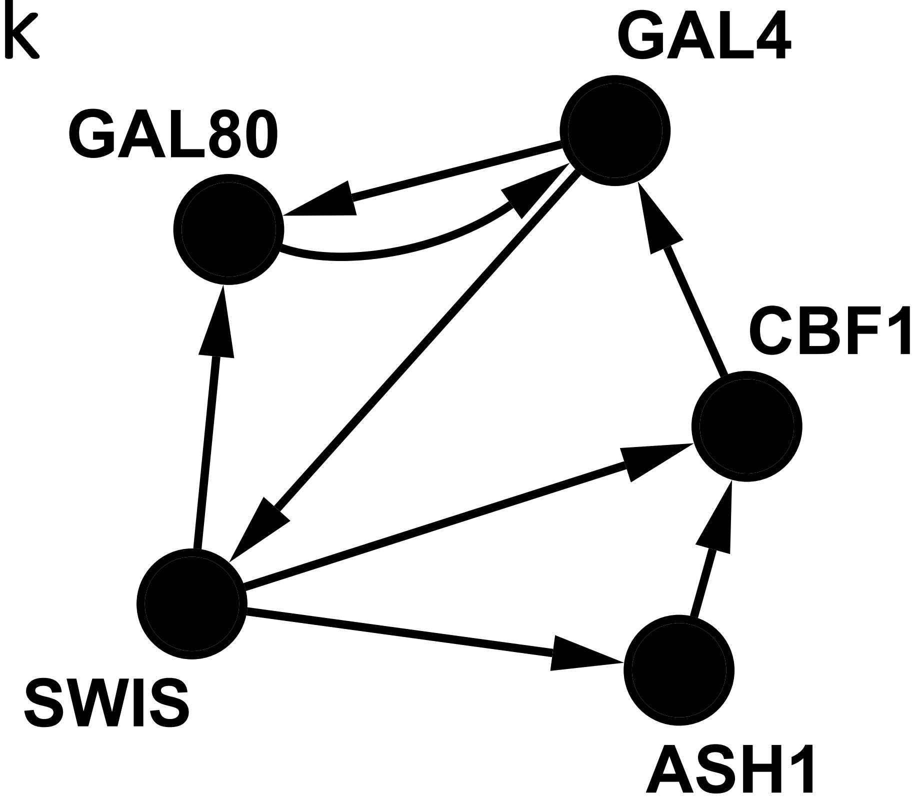
Note

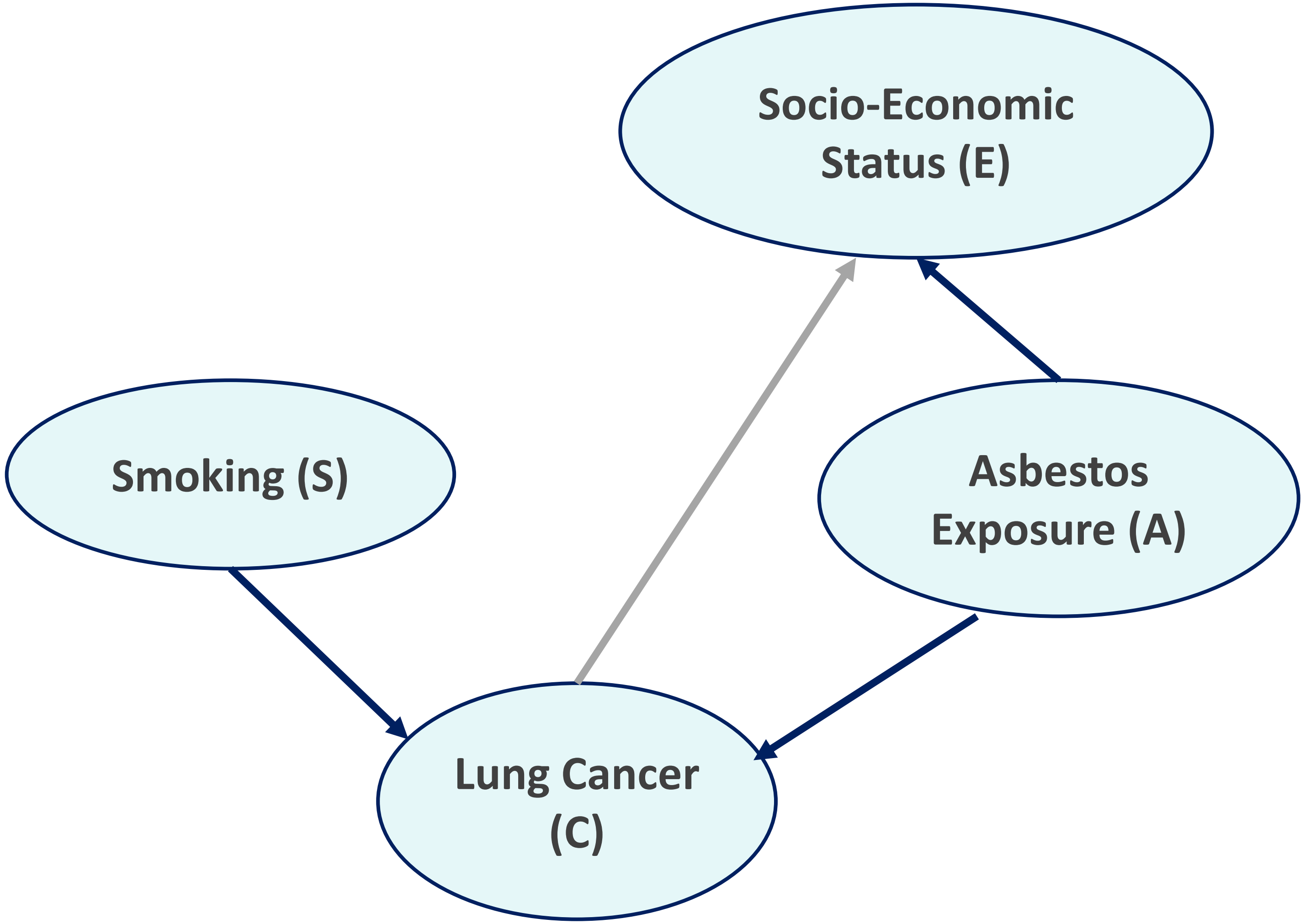


Example 2: Gene regularity network

Network of n = 5 genes in Saccharomyces cerevisiae (yeast). The data obtained from synthetically designed yeast cells grown with different carbon sources: galactose ("switch on") or glucose ("switch off "), Cantone et al. (2009).

- **<u>Dynamic</u> Bayesian Network Models**



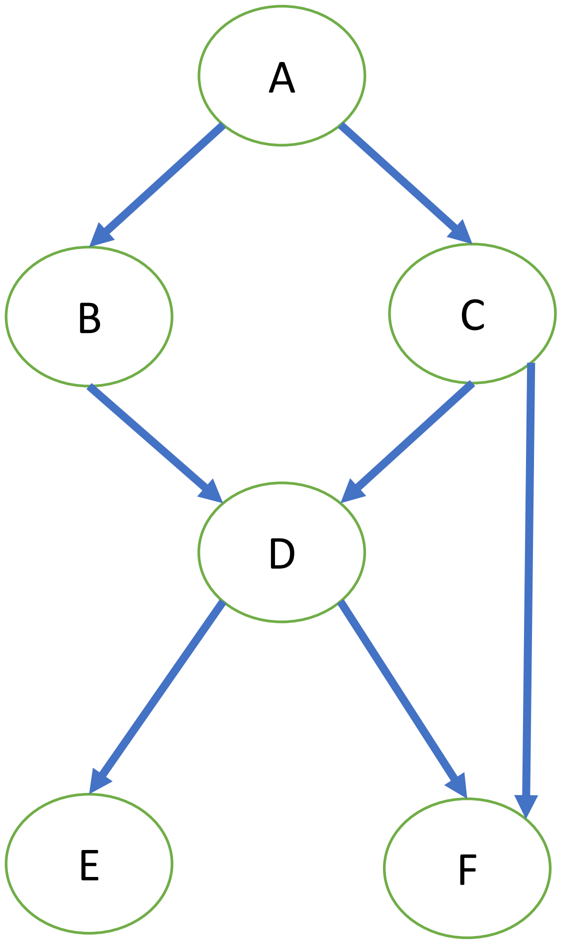

Socio-Economic Status (E)
Smoking (S)
Asbestos Exposure (A)
Lung Cancer (C)

Some applications of BNs

- Biology (Gene Regulatory Network, …)
- Medicine
- Document Classification. ...
- Image Processing. ...
- Spam Filter
- ….

See here:

https://data-flair.training/blogs/bayesian-network-applications/

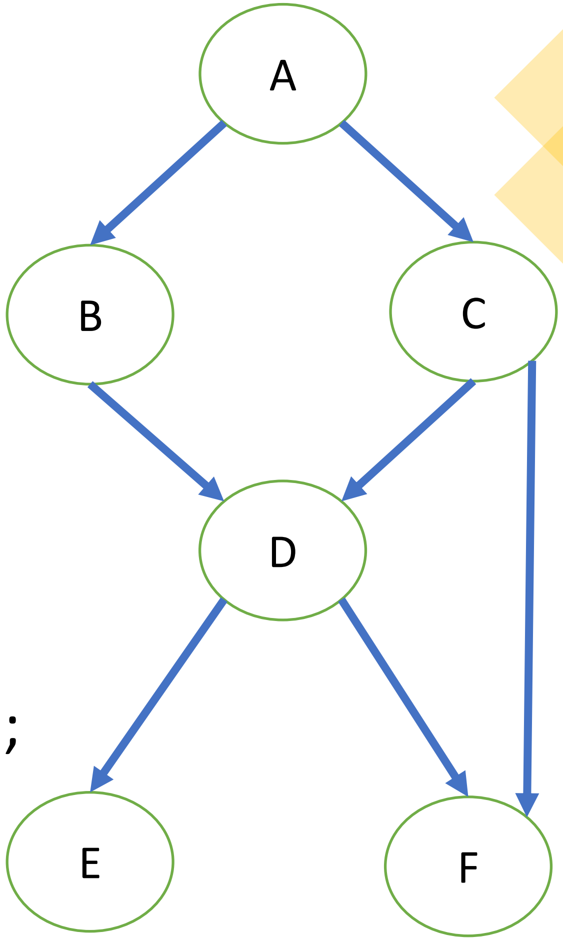


Bayesian networks

- **<u>The first component</u>** of a BN is a <u>graph</u>.

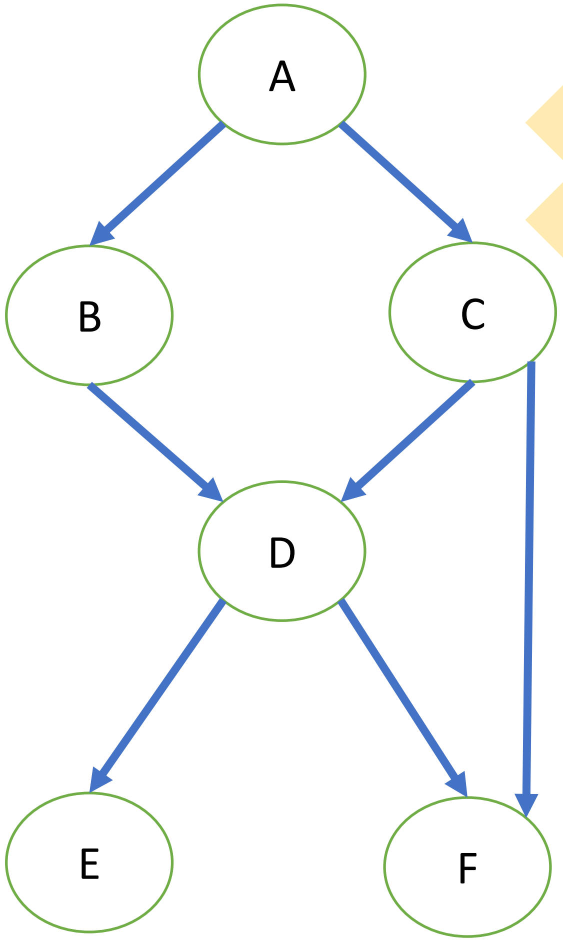
  A graph G is a mathematical object with:
- a set of nodes $V = \{v_1, \dots, v_N\}$;
- a set of **arcs** A which are identified by pairs of nodes in V, e.g. $a_{ij} = (v_i, v_j)$.

- **The second component** of a BN is the <u>probability</u> distribution $P(X)$, should be such that the BN:
- can be learned efficiently from data;
- is flexible (distributional assumptions should not be too strict);
- is easy to query to perform inference.

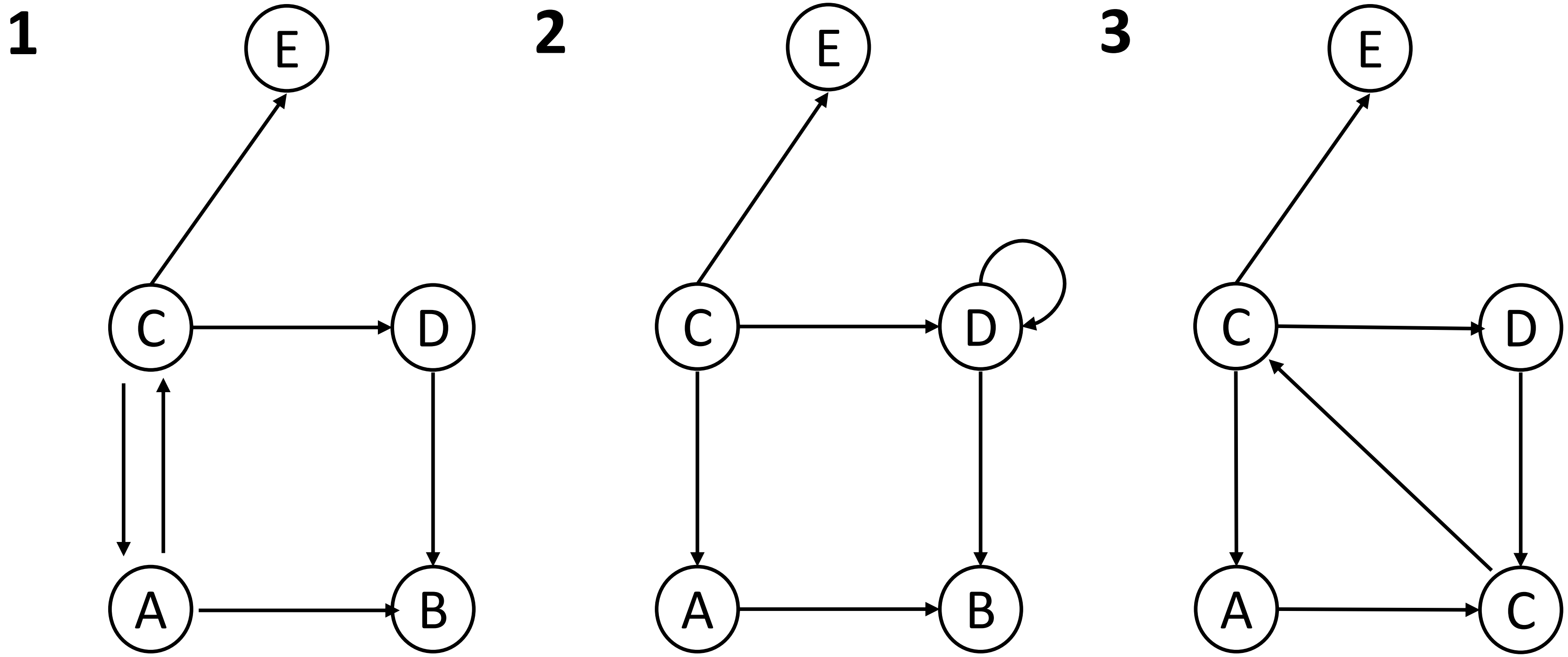


Static Bayesian network as Directed Acyclic Graph (DAG)

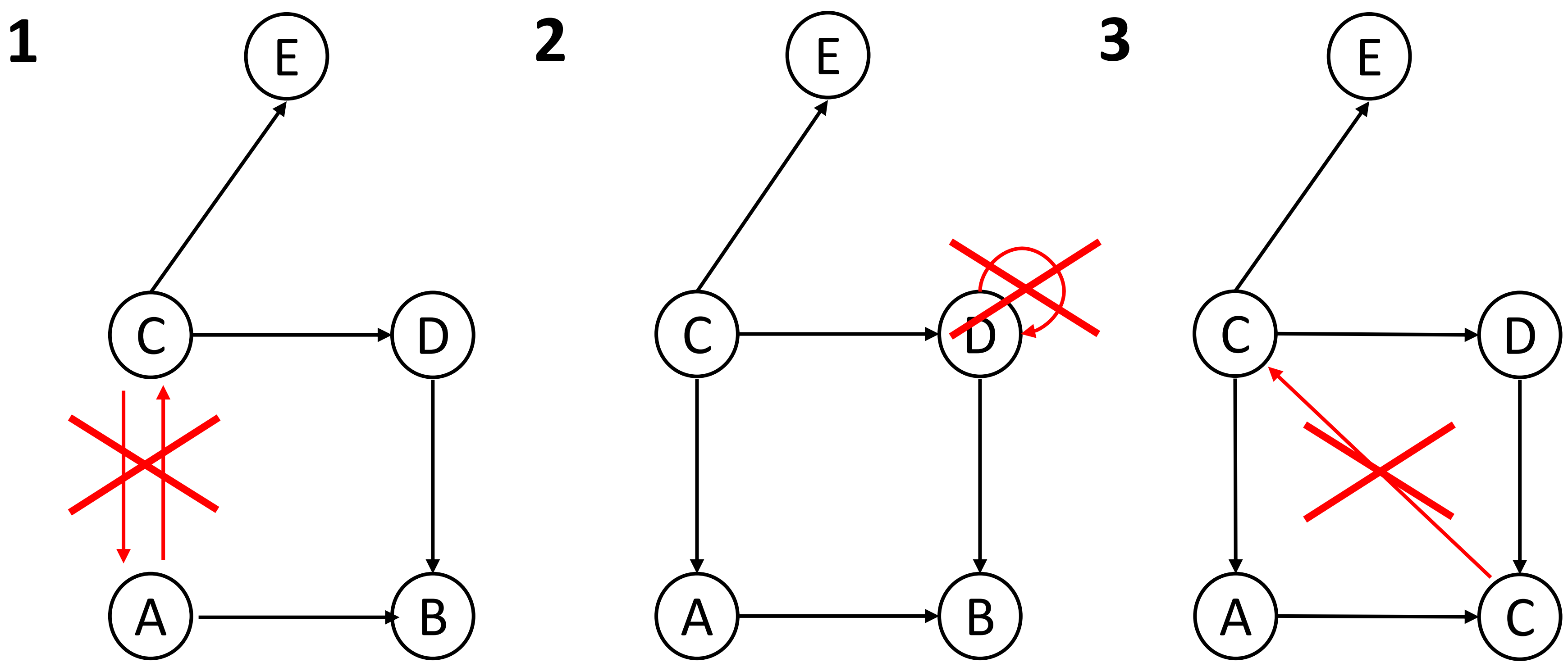
- contains only **directed** arcs and <u>does not contain any loops/cycle</u>.



DAG: Which one is a Directed acyclic graph(No loops/no cycle)?

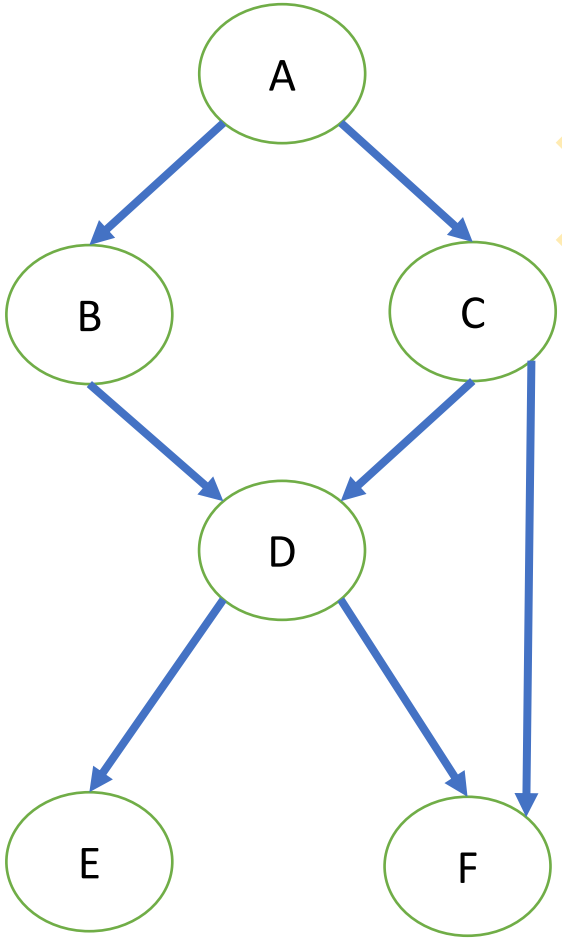


DAG: Directed acyclic graph: No loops/no cycle



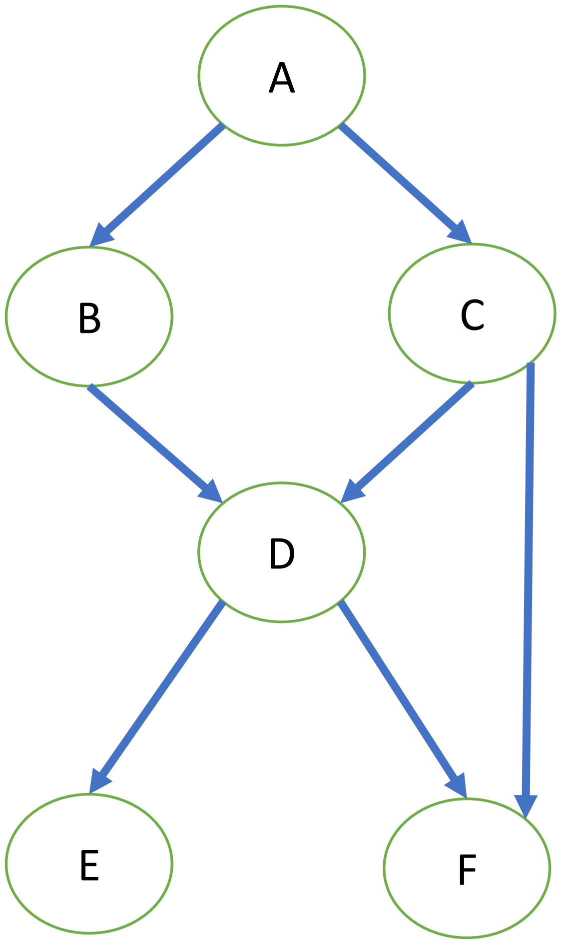
Interpretation of edges in BNs

- If there is an **edge** from one variable to another, **the later depends on the former**.
- Variables that **are not linked (lack of edges)** are **conditionally independent**.



Bayesian networks

- A is a parent node of B.
- B is a child node of A.
- The parent node set of D is the set $\{B,C\}$.
- D is a common child node of B and C.
- A has no parents. That is the parent set of A is an empty set.
- (B) C a is (co-)parent node of D (another parent).



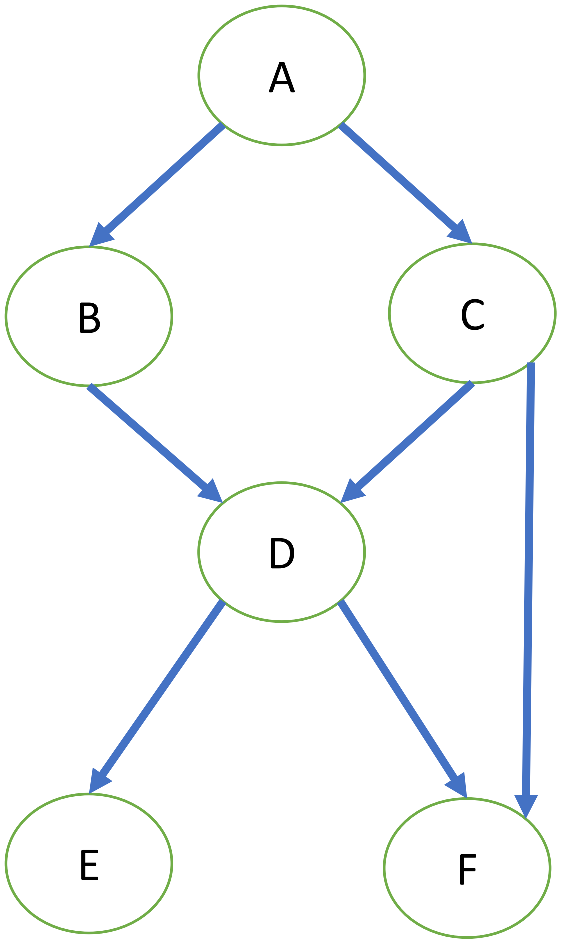
Bayesian networks

- Node X is an **ancestor** of node Y

if there is a **path** of directed edges leading from X to Y:

$$\boldsymbol{X} \rightarrow \cdots \rightarrow \boldsymbol{Y}$$

- Y is then called a **descendant** of X.



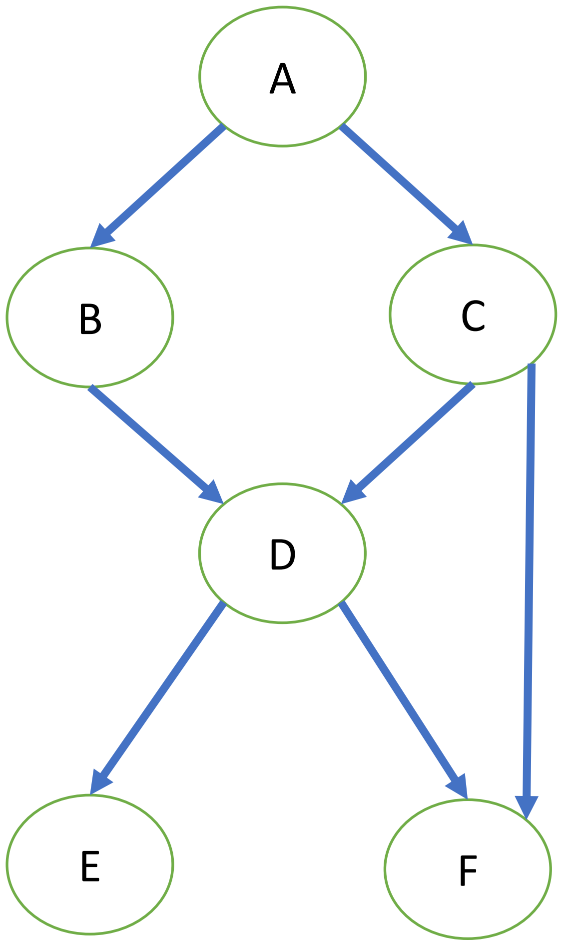
Bayesian networks and Markov assumption

- **Markov assumption** leads to a factorization of the joint probability distribution:

$$P(A, B, C, D, E, F) =$$

$$P(A)\, P(B|A)\, P(C|A)\, P(D|B,C)\, P(E|D)\, P(F|C,D)$$

P(child| parent(s))



- **Markov assumption:**

Every **node** in a Bayesian network is **conditionally independent** of its non-descendants, given its parents (only the parents).

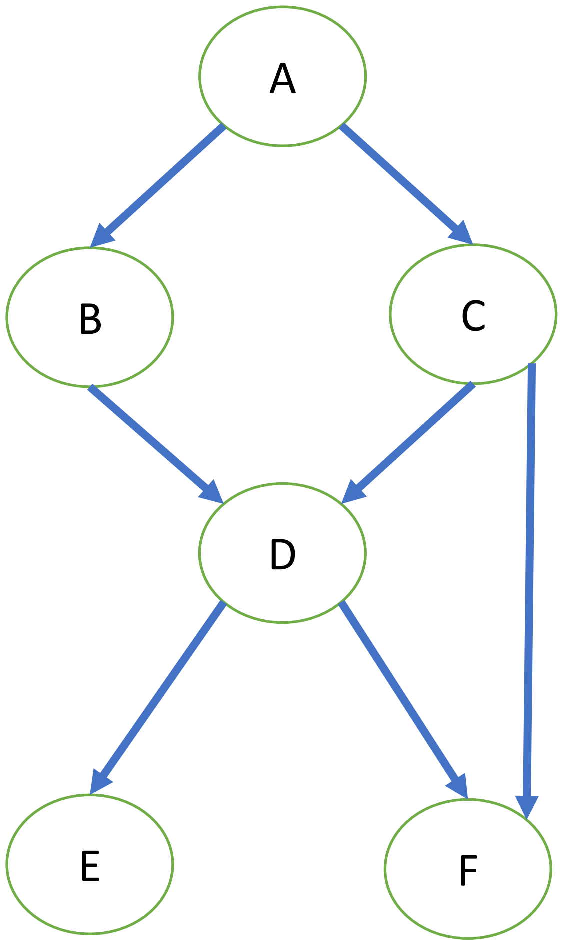
Bayesian networks and Markov assumption

Which node is conditionally independent of node D given D's parent nodes?

a. $A \perp\!\!\!\perp D | \{B, C\}$

b. $E \perp\!\!\!\perp D | \{B, C\}$

c. $F \perp\!\!\!\perp D | \{B, C\}$



Example: Train Use Survey

- Consider a survey whose aim is to **investigate the usage patterns** of different means of **transport**, with a **focus** on **cars** and **trains**.

- Such surveys are used to assess **customer satisfaction** across different social groups, to evaluate public policies or for urban planning.

- Some other real-world examples can also be found, in **Kenett et al. (2012).**

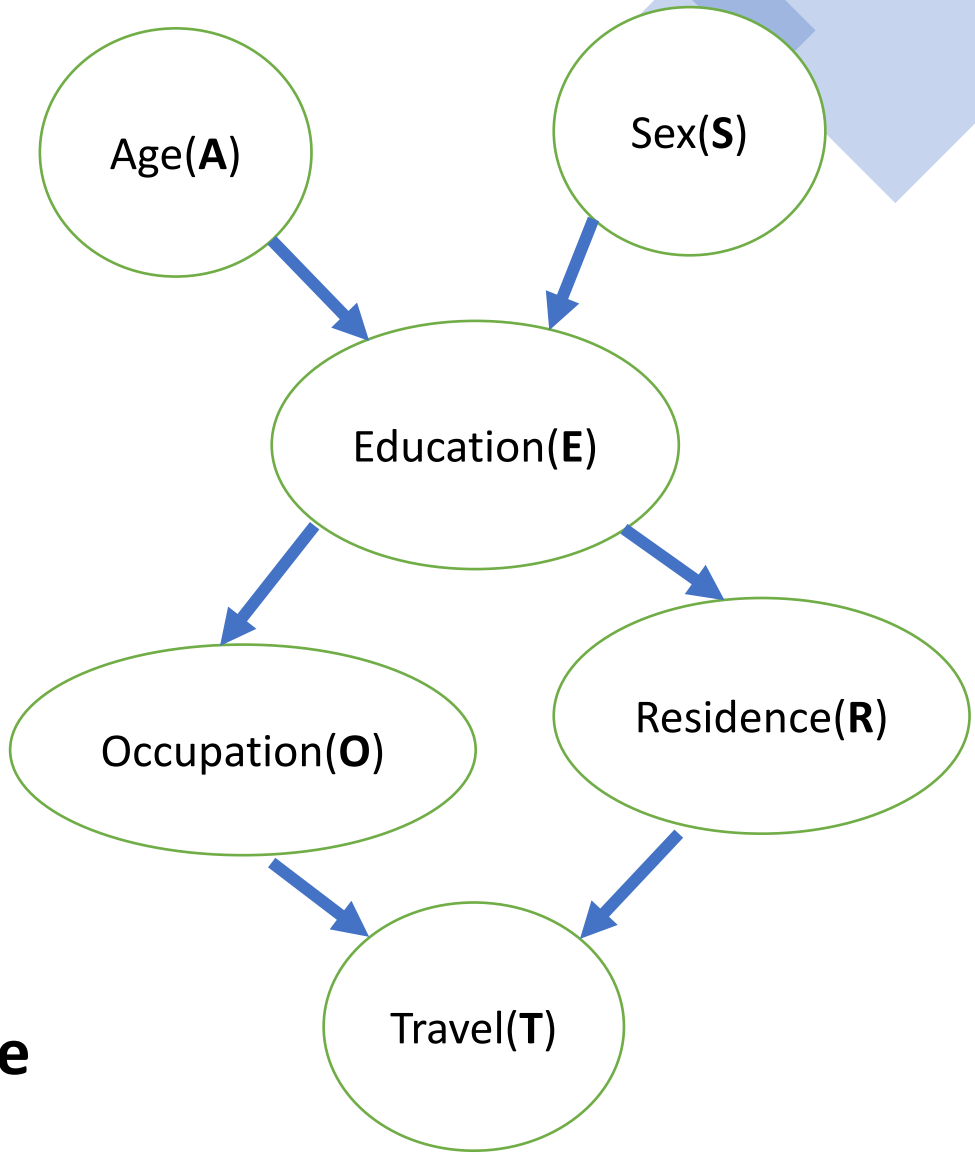
Data: Train Use Survey

Age	Residence	Education	Occupation	Sex	Travel
• "adult"	"big”	"high"	"emp”	"F”	"car"
• "adult"	"small"	"uni"	"emp"	"M"	"car"
• "adult"	"big"	"uni"	"emp"	"F"	"train"
• "adult"	"big"	"high"	"emp"	"M"	"car"
• "adult"	"big"	"high"	"emp"	"M"	"car"
• "adult"	"small"	"high"	"emp"	"F"	"train"
• "adult"	"big"	"high"	"emp"	"F"	"car"
• "young"	"big"	"uni"	"emp"	"F"	"train"

Example: Train Use Survey

- **Age** and **sex** are not influenced by any other variable.
- **Age** and **sex** have a direct influence on **Education.**
- **Education** strongly influences both **occupation** and residence
- **Transports** are directly influenced by both **occupation** and **residence**.

- This **DAG** represents the **dependence relationships** between: **Age** , **Sex**, **Education**, **Occupation**, **Residence** and **Travel**.



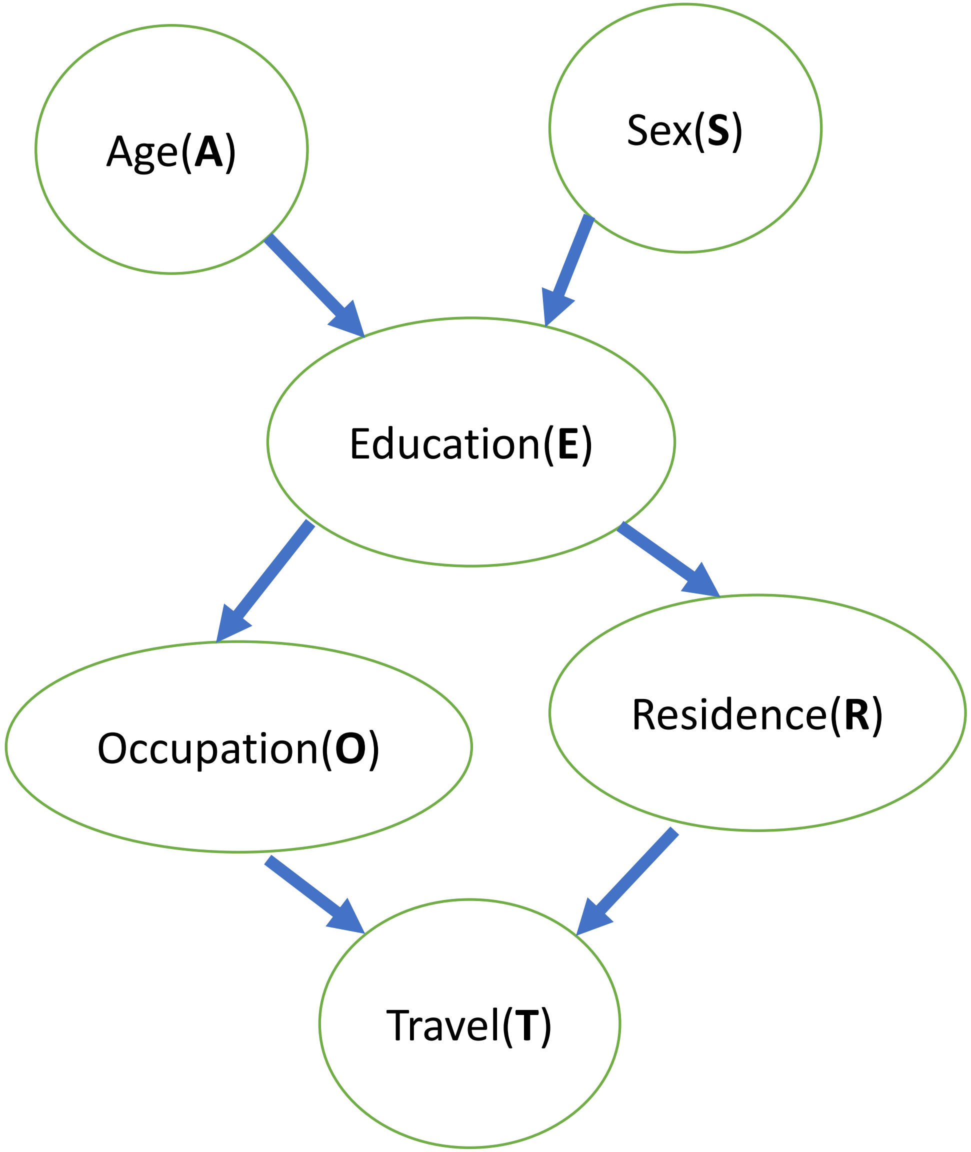
Example: Train Use Survey

The probabilistic relationship:

[A] [S] [E|A:S] [O|E] [R|E] [T|O:R]

[child|parents]

This type of representation
is what we will see when we use
bnlearn package in practical.



- Think about the application of this concept in your field for a few minutes and discuss that in pairs.

- Consider the following questions for reflection in your field:

✓ What variables are present in your potential project?

✓ Why are these variables important?

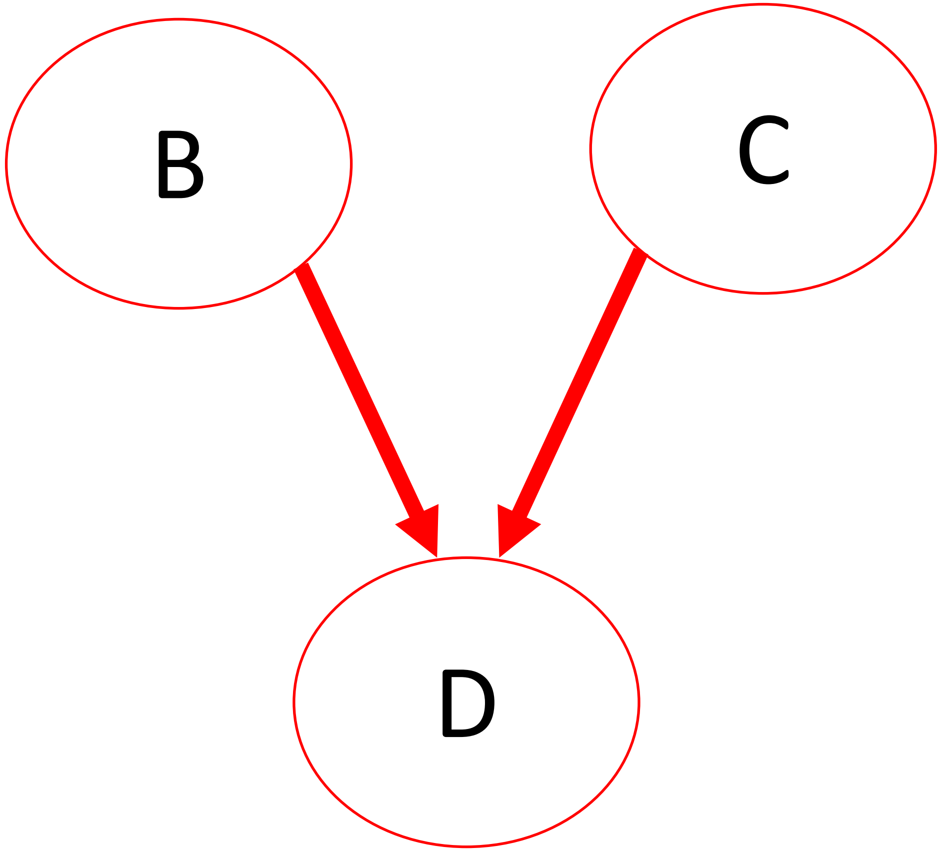
✓ What motivates your interest in understanding their interdependencies?

✓ How does this understanding contribute to your work or goals?

Some terminologies

v-structure

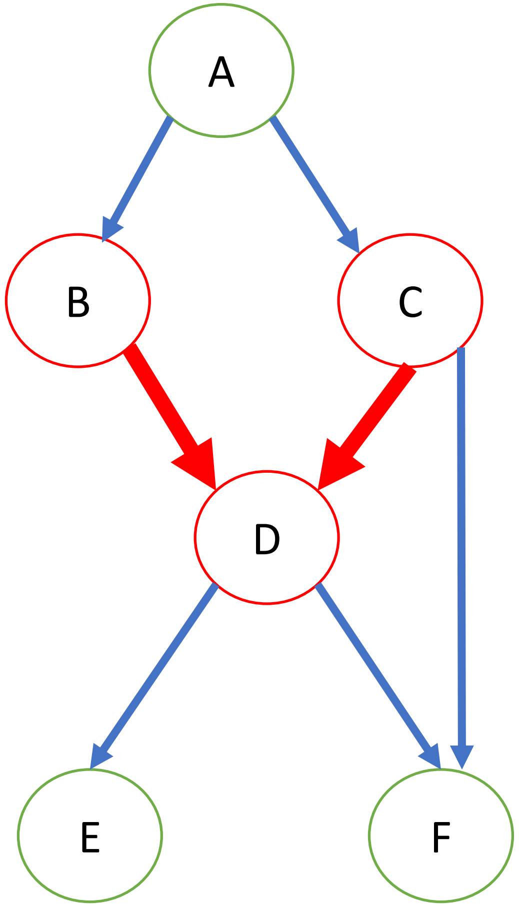
Important for interpretation



$B \longrightarrow D \longleftarrow C$

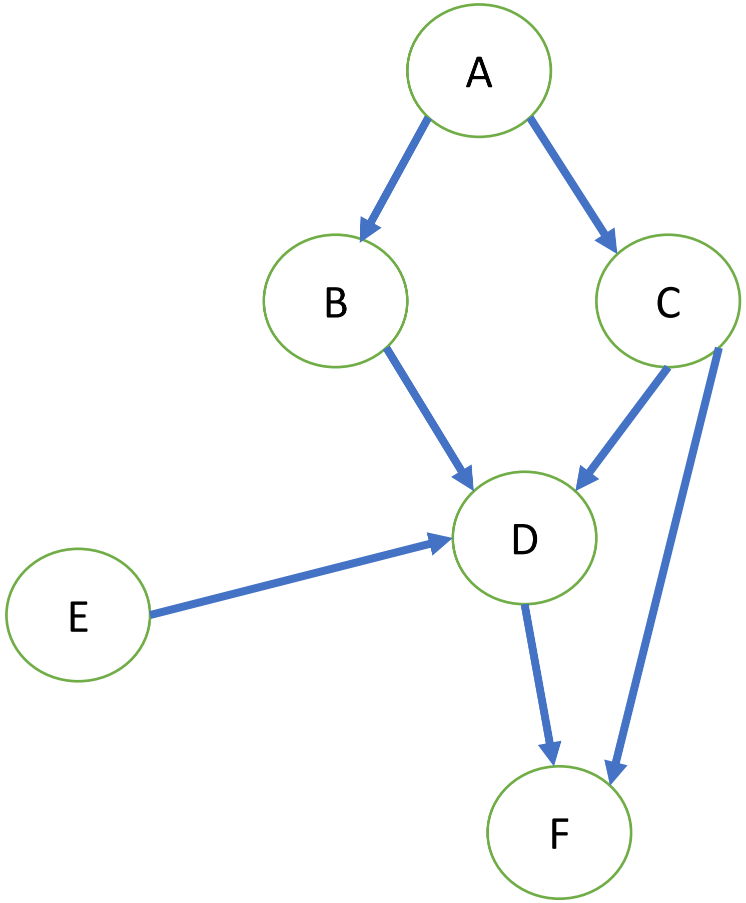
v-structure
Important for interpretation

- D has the parent nodes B and C, and there is **no connection** between the nodes **B** and **C**.



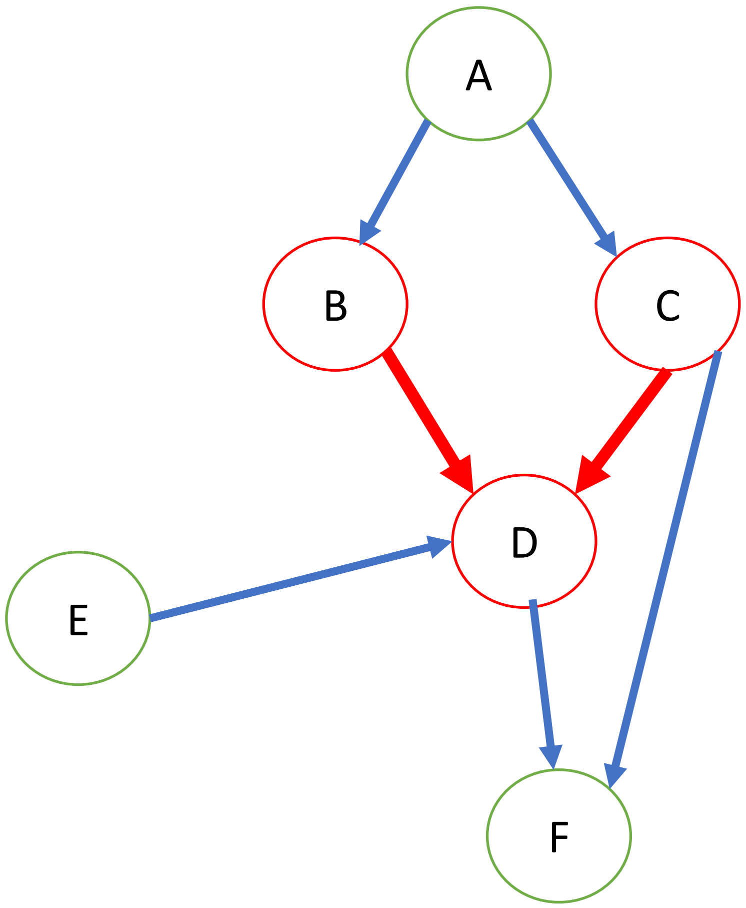
v-structure

How many v-sructures do you see?



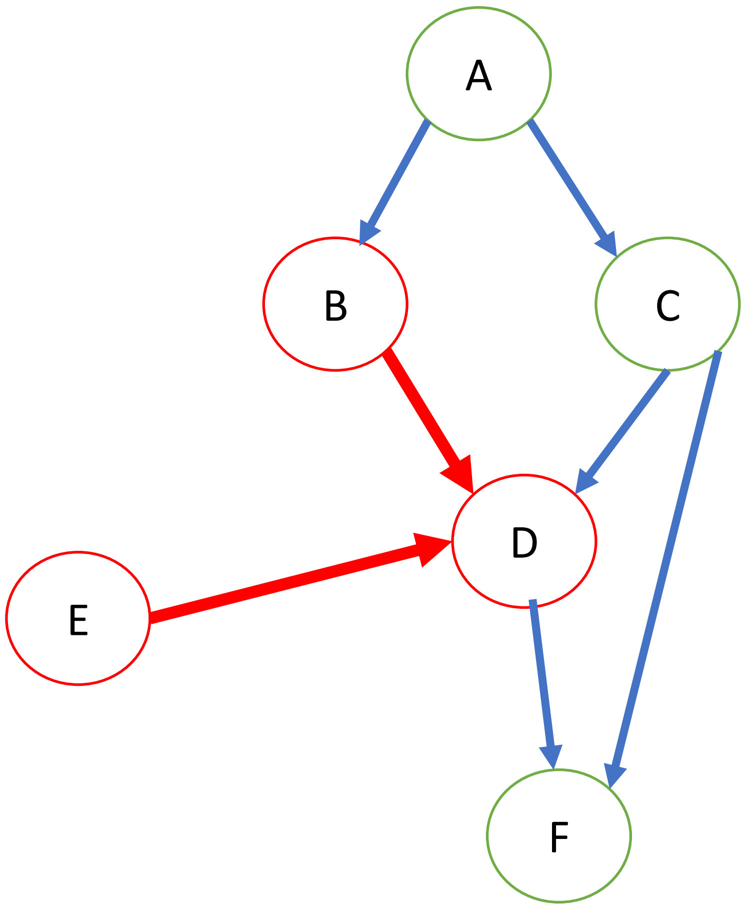
v-structure

1



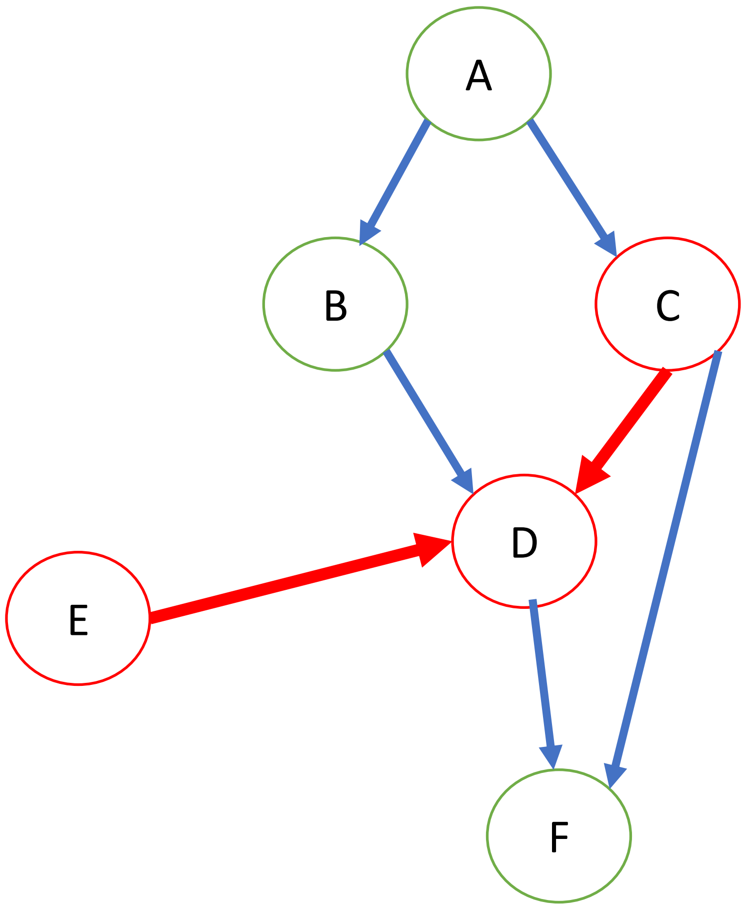
v-structure

2



v-structure

3



Markov Blanket

Graph **G** has **n** nodes, $\boldsymbol{X_1}, \ldots, \boldsymbol{X_n}$.

The Markov blanket of the node X_i $(i = 1, \ldots, n)$ includes:

- **all parent** nodes of X_i

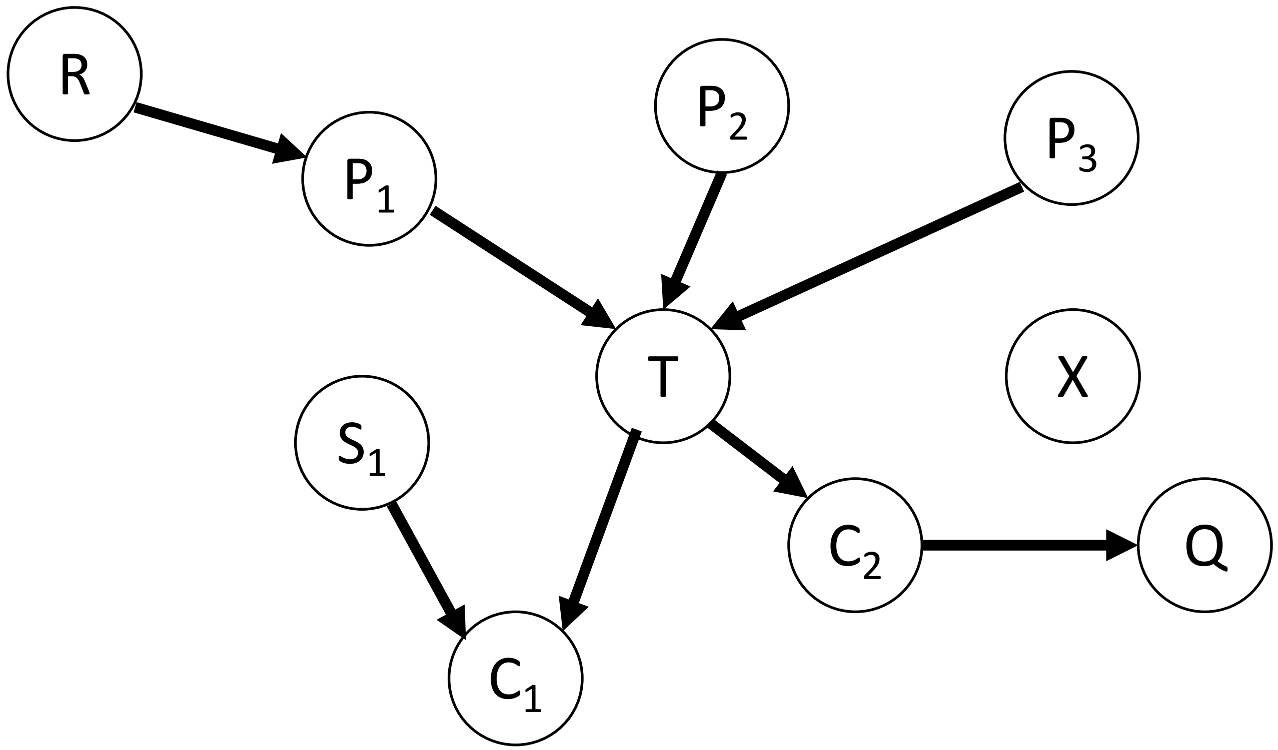
- **all child** nodes of X_i

- **all "co-parent"** nodes of X_i

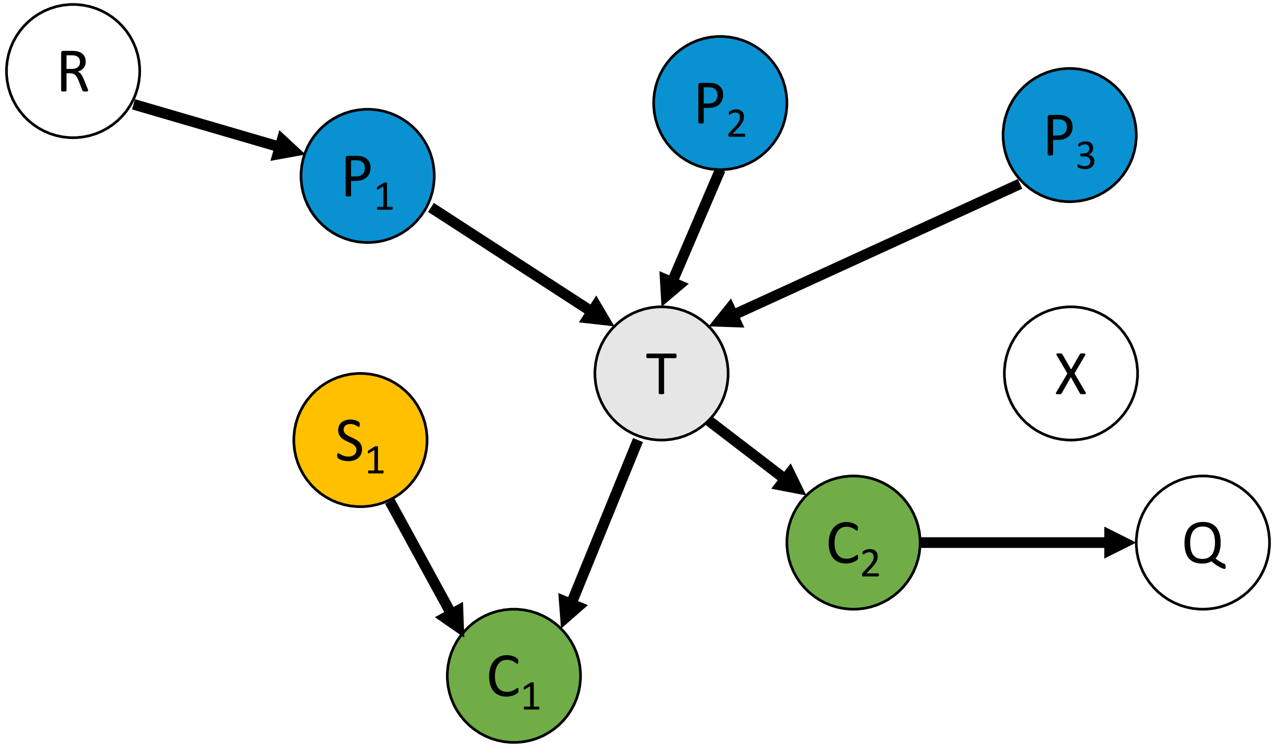
We denote the **Markov Blanket** of X_i symbolically as $\mathbf{MB}(\boldsymbol{X_i})$.

Using **bnlearn** package, **mb()** function can be used to show the Markov blankets.

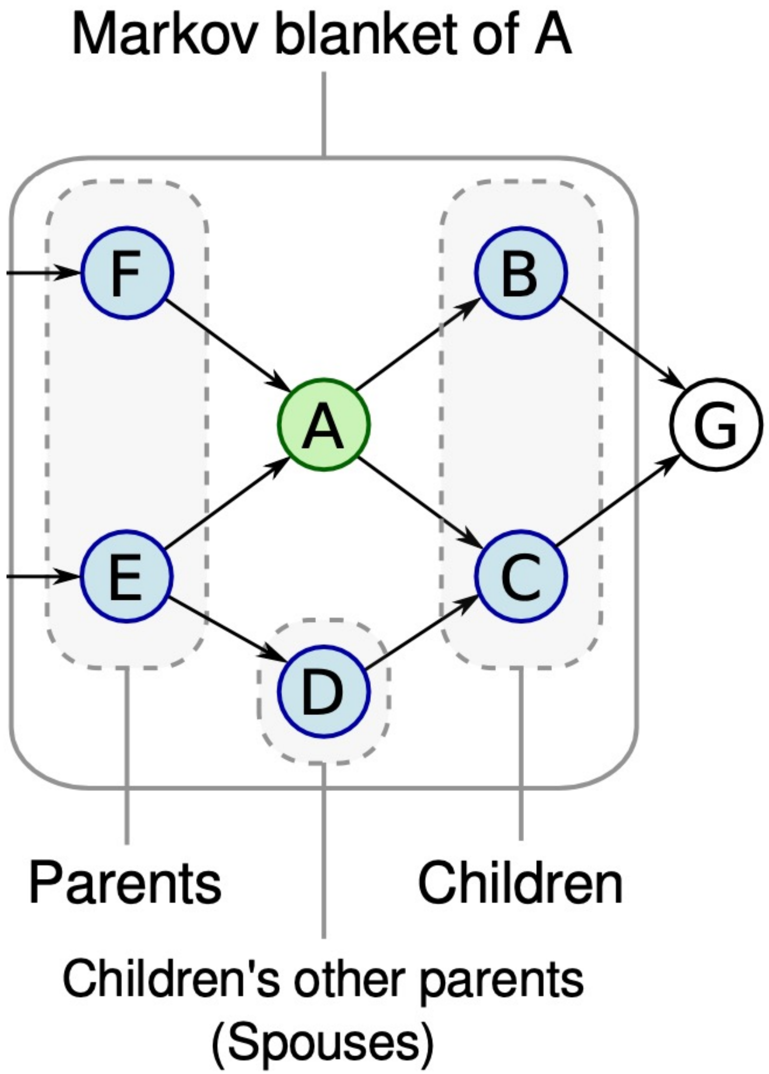
Markov Blanket of T?



Markov Blanket



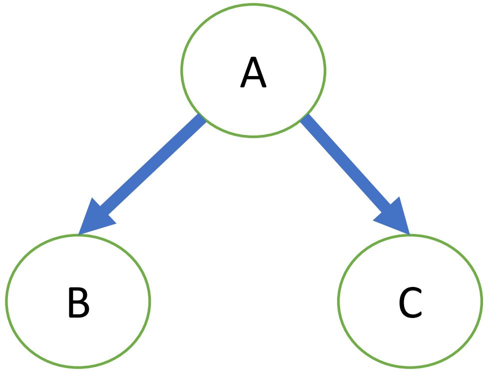
Example Markov blanket of a target T, consisting of **three parents (blue nodes), two children (green)** and **one spouse (orange)**. All other nodes are conditionally independent of T given MB(T) .



We can easily use the DAG to solve the **feature selection** problem.

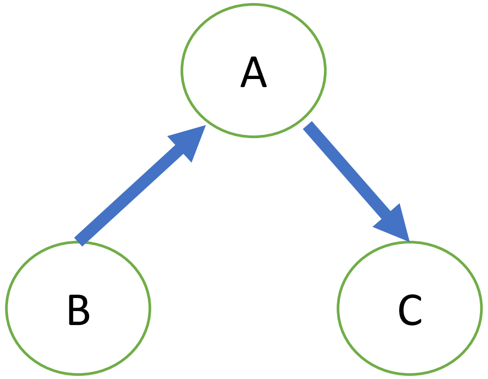
We can restrict ourselves to the Markov blanket to perform any kind of inference on the target node and disregard the rest.

Fundamental connections



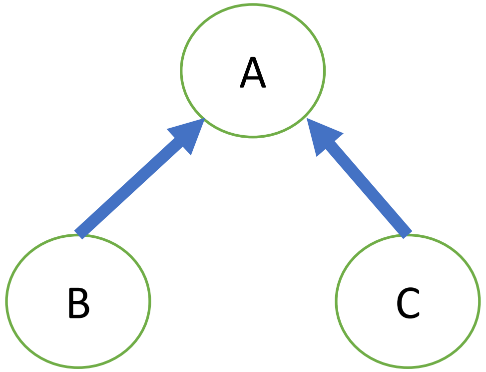
B ← A → C

Divergent connection



B → A → C

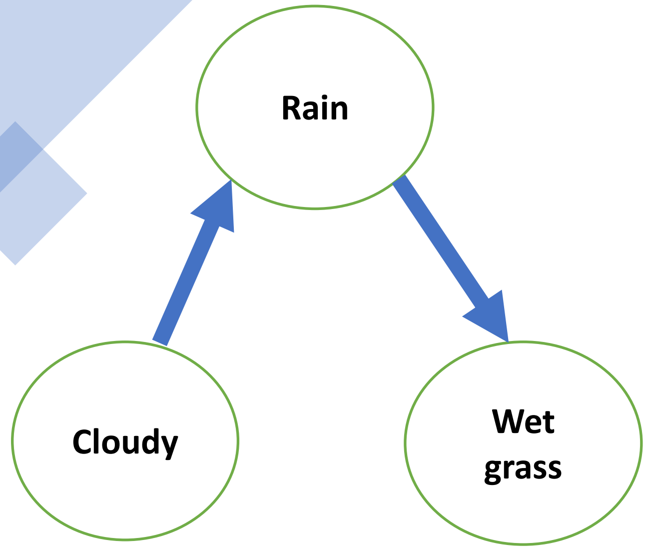
Serial connection



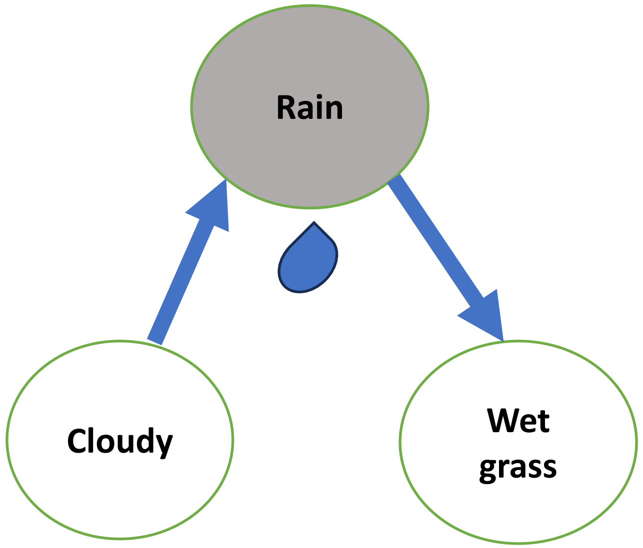
B → A ← C

Convergent connection

Some examples:



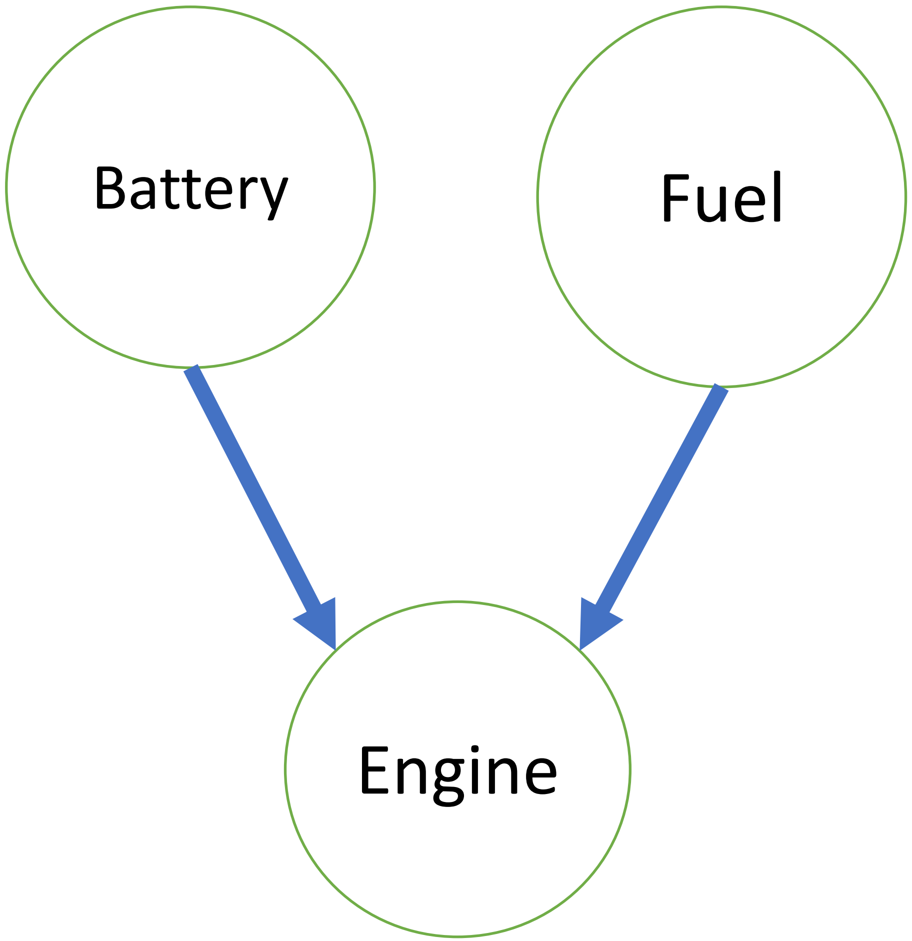
- Both variables **Cloudy** and **Wet grass** are **statistically dependent**.
- **Cloudiness** increases the probability of **rain** and thus indirectly the probability of a **wet grass**.
- **Conditional** on the variable **Rain**, the two variables **Cloudy and wet grass** are stochastically independent of each other.



1) If it is known whether it rains or not, the state of cloudiness has no influence on the probability that the grass is wet.

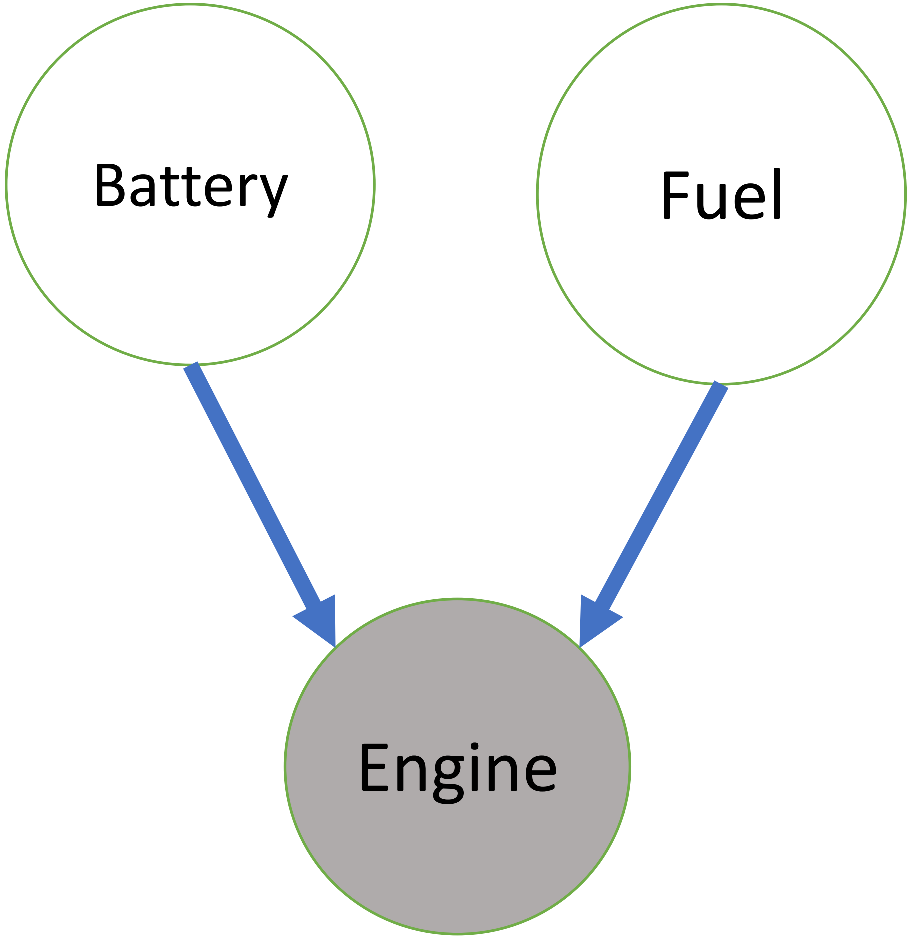
2) **If it is known whether it rains or not**, the condition of the grass has no influence on the probability of the state of the clouds.

Some examples:



- The binary variable **Battery** indicates if the car battery is working or not.
- The binary variable **Fuel** indicates whether the tank of the car is empty or not.
- The binary variable **Engine** indicates whether the car can be started or not.
- **Battery** and **Fuel** are **stochastically independent**.

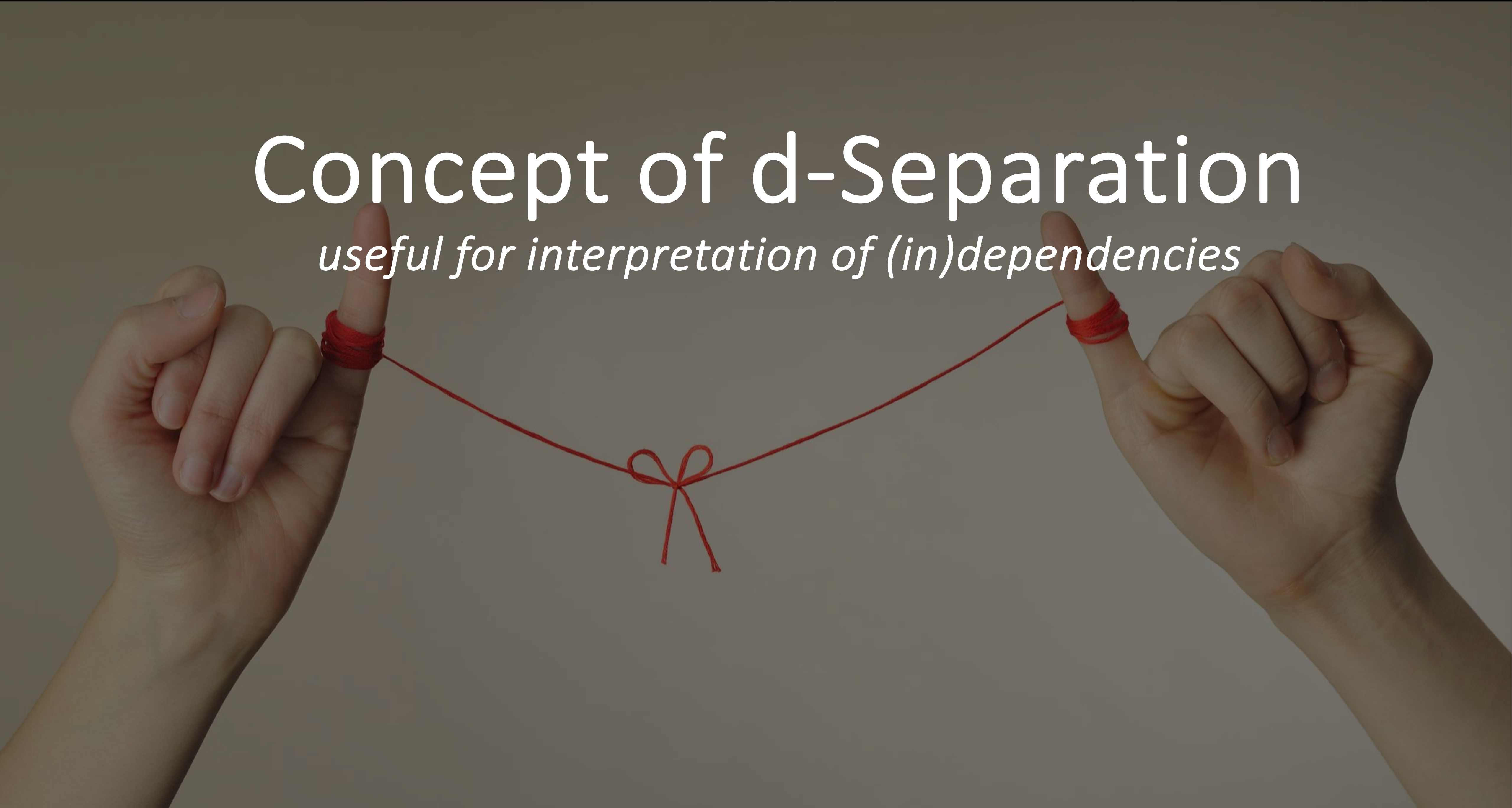
Some examples:



- However, **conditional** on the variable **Engine**, the two variables **Battery** and **Fuel** become **stochastically dependent**.
- When the car cannot be started, the probability that the fuel tank of the car is empty increases with the information that the battery is operating.

Concept of d-Separation

useful for interpretation of (in)dependencies



Concept of d-Separation

- In Bayesian networks, **the (in-)dependence relations** between the nodes (or variables) are easily obtained with the help of the **d-separation**.
- If two nodes/variables are d-separated, they are conditionally independent.
- It is very useful for interpretation.
- **dsep()** in **bnlearn** package.

Path

- **Definition of directed path**

There is a directed path from node $\mathbf{X_i}$ to node $\mathbf{X_j}$ in a graph G if one can move from $\mathbf{X_i}$ to $\mathbf{X_j}$ by **following directed edges** (according to their edge directions).

$$\mathbf{X_i} \rightarrow \ldots \rightarrow \mathbf{X_j}$$

- **Definition of (any) path (path, trail)**

There is a path (trail) between the nodes $\boldsymbol{X}_{\mathrm{i}}$ and $\mathbf{X_j}$, if the two nodes are connected to each other through **a sequence of edges** (does not matter in which direction).

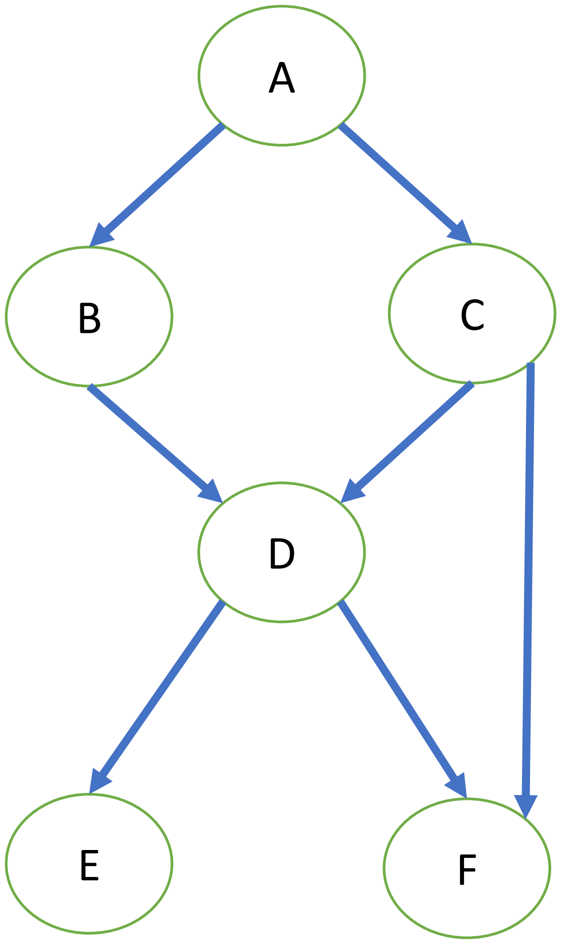
- In **a path or trail**, each node can appear **only once.**

$$\mathbf{X_i} \rightarrow \mathbf{X_{i+1}} \leftarrow \ldots \rightarrow \mathbf{X_j}$$

- Are these valid paths?

1. A→C→F←D←B←A←C
2. A→D←B←C
3. C→F←D←B←A

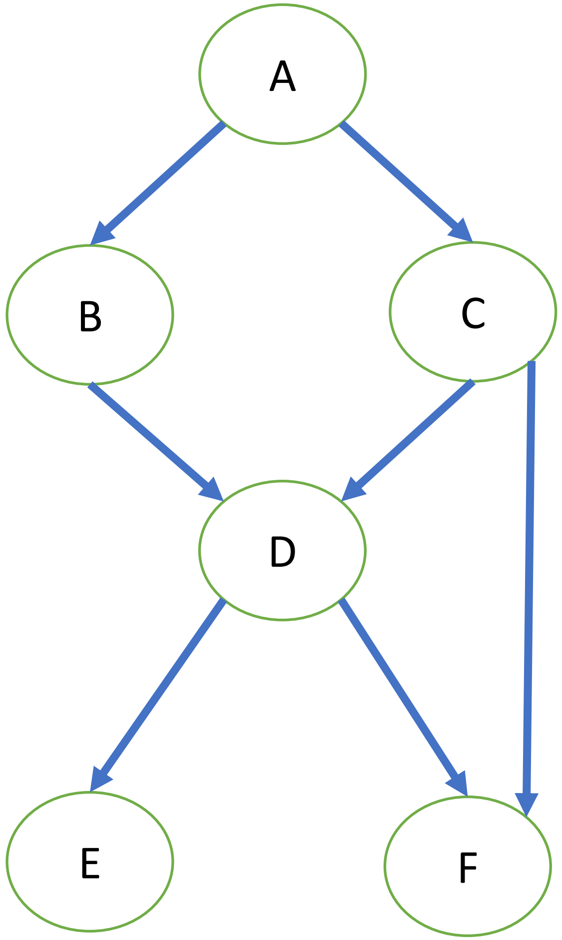
- One more paths?



Example of (directed) paths

- Examples of directed paths:
- A→B→D→F
- A→C→D

- We have the paths (trails):
- A→B→D←C
- B→D← C
- A→C→F←D←B



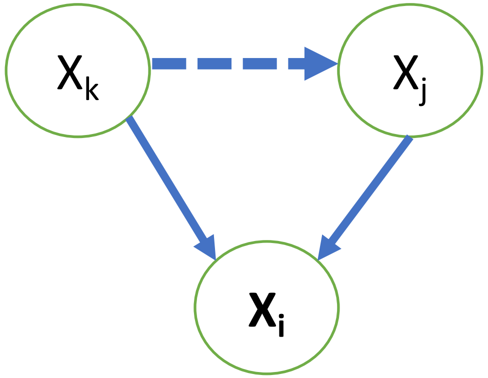
Collider

In a trail / path the node X_i $(i = 1, \ldots, n)$ is a collider if two edges converge on X_i.

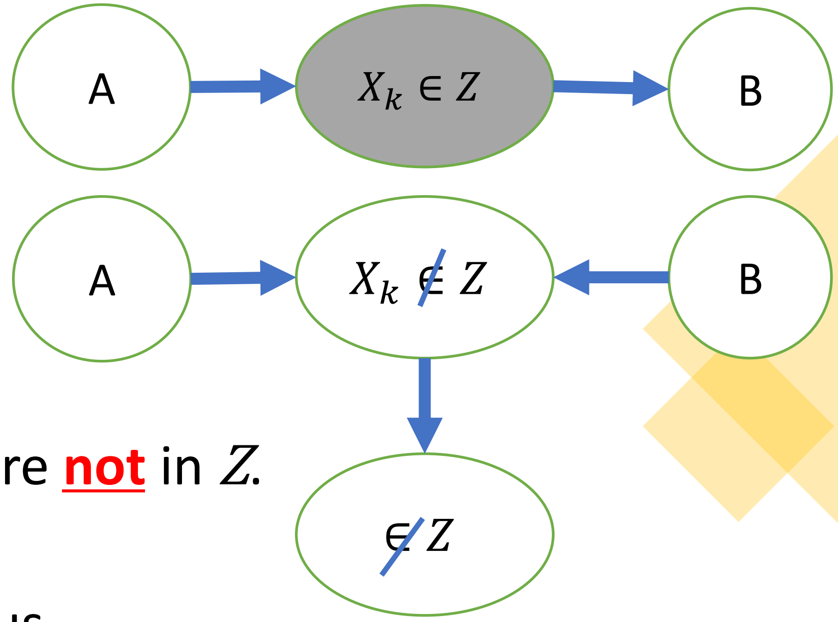
E.g.: $\mathbf{X_w} \rightarrow \mathbf{X_k} \rightarrow \mathbf{X_i} \leftarrow \mathbf{X_j} \rightarrow \mathbf{X_m}$

- **Note**: This definition does not require that $\mathbf{X_k} \rightarrow \mathbf{X_i} \leftarrow \mathbf{X_j}$ is a v-structure.

The nodes X_k and X_j can be in a parent-child relationship.



Blocked path



- We consider the nodes $\boldsymbol{A}$ and $\boldsymbol{B}$ and a subset Z , where $\boldsymbol{A}$ and $\boldsymbol{B}$ are **not** in Z.

- Being in Z means the node is observed. That is, we know its status.

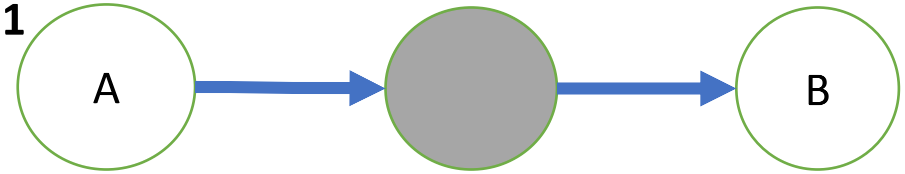
- A path (trail) between $\boldsymbol{A}$ and $\boldsymbol{B}$ is **blocked** when the trail leads through any node X_k and:

(1) X_k **is not** a collider and X_k **is an element** of Z (X_k is observed).

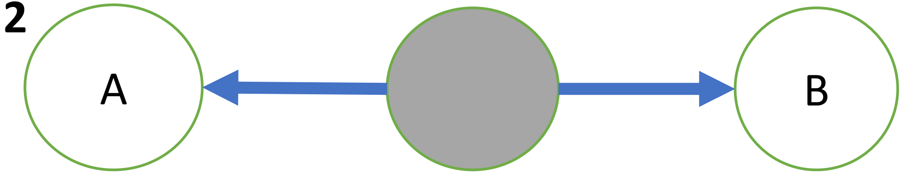
(2) X_k **is** a collider and neither X_k nor a descendant of X_k is an element of Z (**not** observed).

Which ones are blocked?

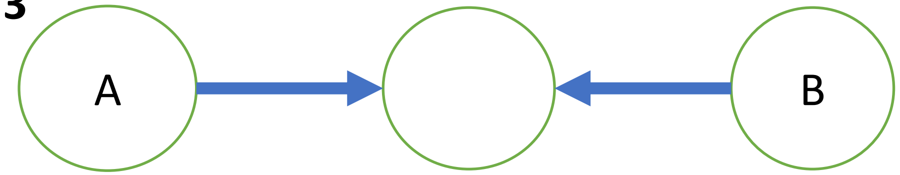
1



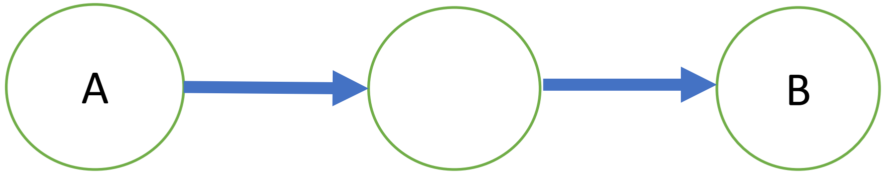
2



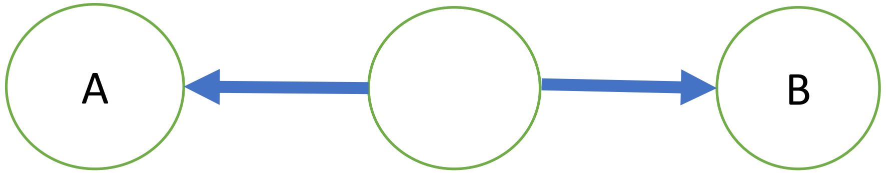
3



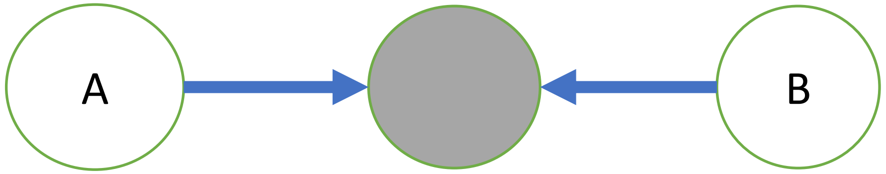
4

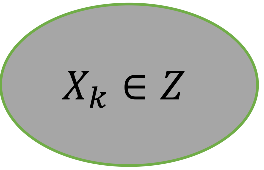


5

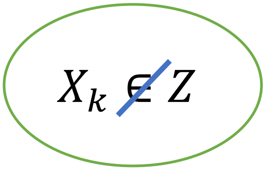


6



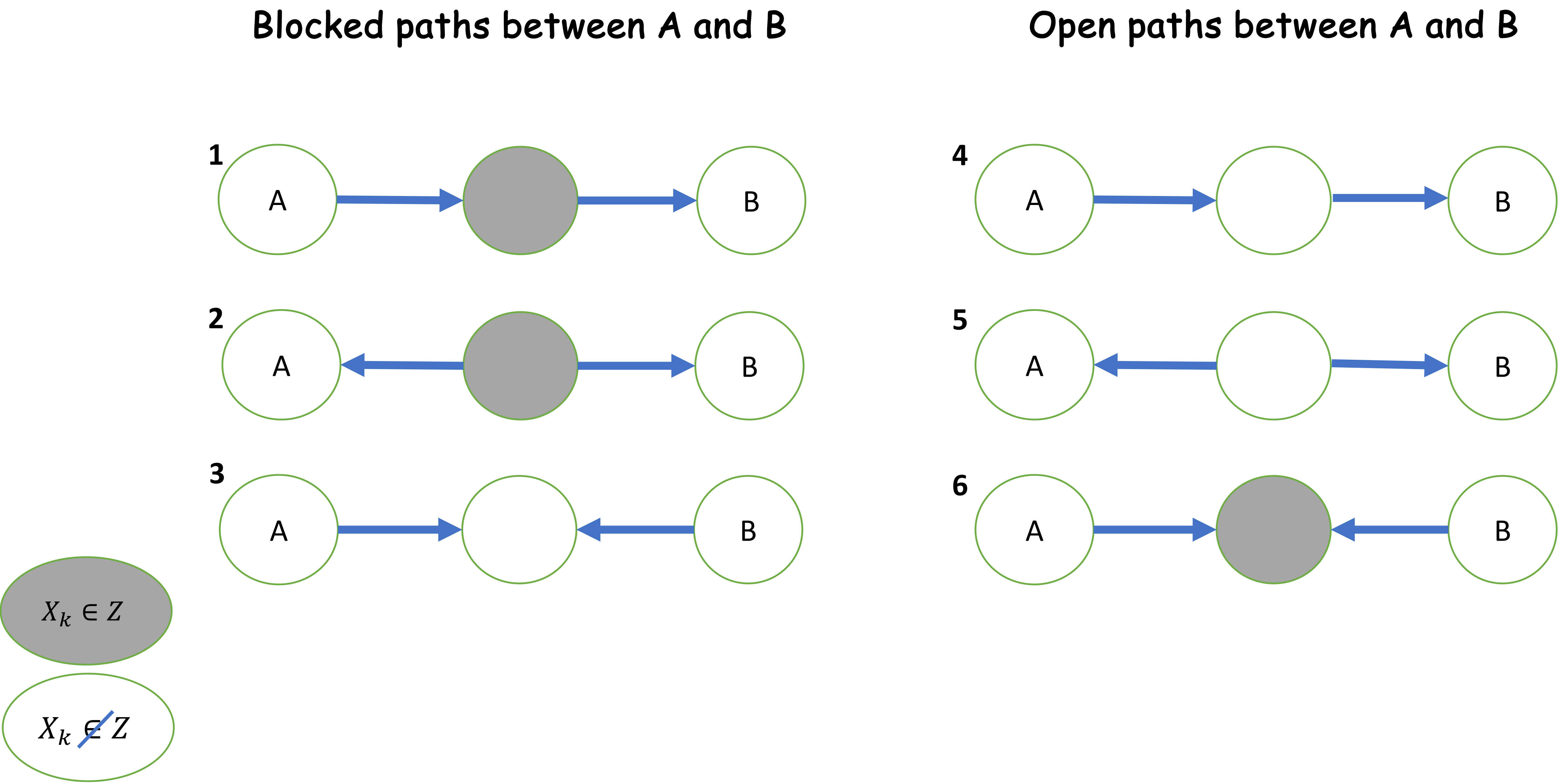


observed



unobserved

The filled (grey) nodes are elements of the set Z (known). The empty (white) nodes are not elements of Z.



The filled (grey) nodes are elements of the set Z (known). The empty (white) nodes are not elements of Z.

Definition: d-Separation

d-Separation

- **The nodes $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$ are d-separated with respect to $\mathcal{Z}$, when every path between $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$ is blocked (conditional on $\mathcal{Z}$).**

- When X_i and X_j are d-separated given $\mathcal{Z}$, then X_i and X_j are **stochastically independent (conditional on $\mathcal{Z}$).**

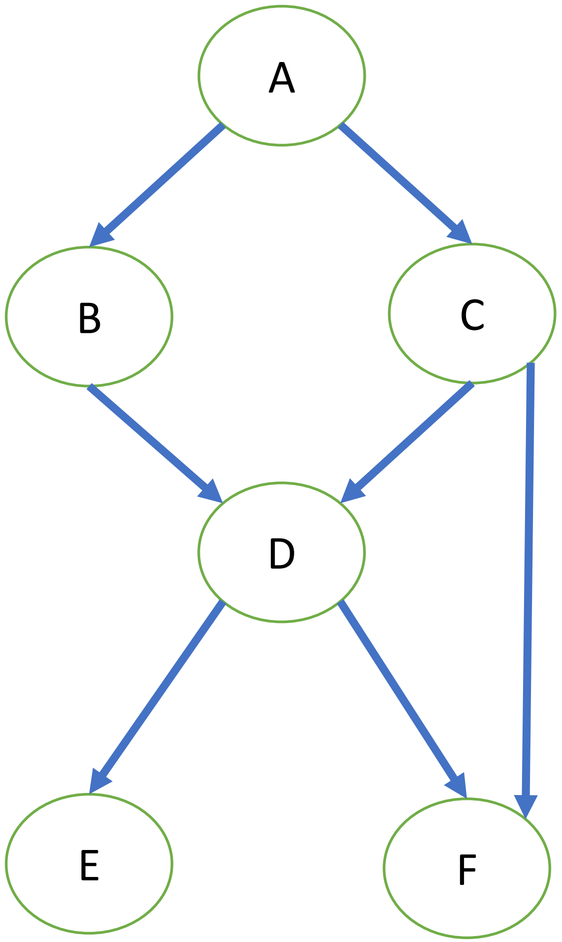
Example

A and D are d-separated conditional on Z={B,C}?

A and F are d-separated conditional on Z={D,C}?

B and C are d-separated conditional on Z={A}?

B and C are d-separated conditional on Z={A,D}?



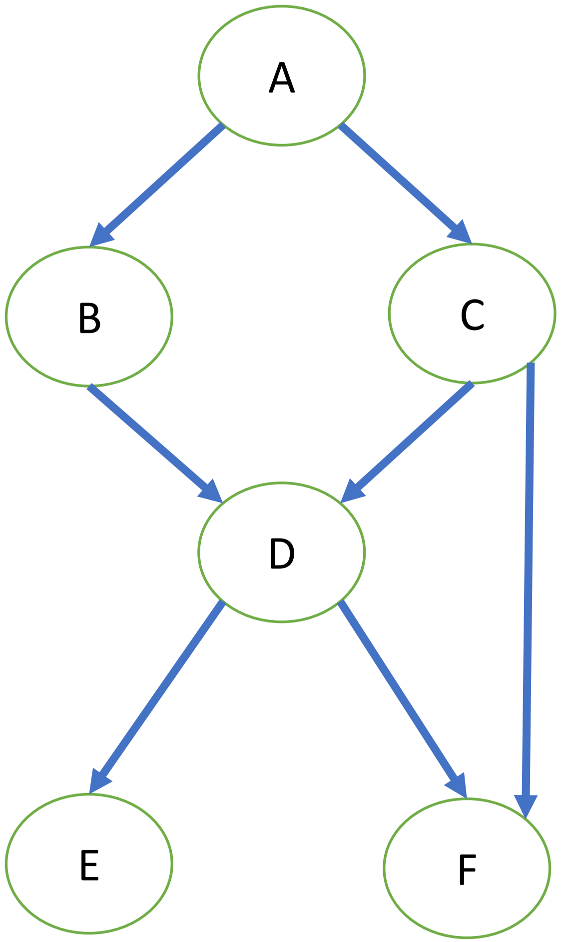
Example

A and D are d-separated conditional on Z={B,C}.

A and F are d-separated conditional on Z={D,C}.

B and C are d-separated conditional on Z={A}.

B and C are **not** d-separated conditional on Z={A,D}   **why?**

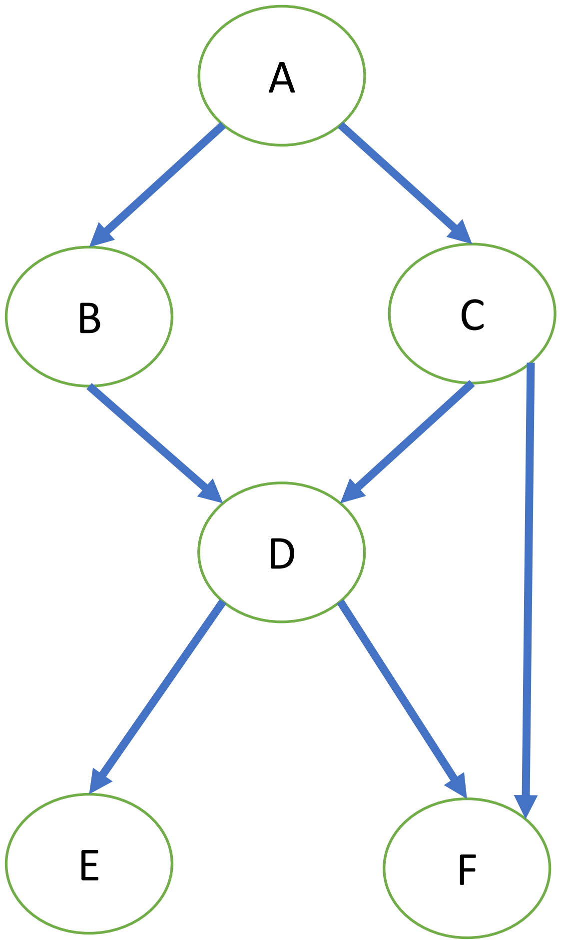


- **The software does everything for you.**

Static Bayesian networks

- **The first component** of a BN is a graph. A graph G is a mathematical object with:
- a set of nodes $V = \{v_1, \ldots, v_N\}$;
- a set of arcs A which are identified by pairs for nodes in V, e.g. $a_{ij} = (v_i, v_j)$.

- The second component of a BN is the probability distribution $P(X)$, should be such that the BN:
- can be learned efficiently from data;
- is flexible (distributional assumptions should not be too strict);
- is easy to query to perform inference.



- Second component:

The probability distribution $P(X)$

Types of BN:

- The **three most common choices** in the literature (by far), are:

- **Discrete BNs:** X and the $(X_i \mid X_i's\ parents)$ are **multinomial**;

- **Gaussian BNs (GBNs):** X is **multivariate normal** and the $(X_i \mid X_i's\ parents)$ are univariate normal;

- **Conditional linear Gaussian BNs (CLGBNs):** CLGBNs contain both discrete and continuous nodes and combine **discrete BNs** and **GBNs** to obtain a mixture-of-Gaussians network.

Discrete BNs

- The joint probability distribution is a **<u>multinomial distribution</u>,** assigning a probability to each **combination of states of variables.**

Multinomial distribution

- A multinomial distribution involves a process that has a set of k possible results (X_1, X_2, X_3, ..., X_k) with associated probabilities (p_1, p_2, p_3, ..., p_k) such that $\sum p_i = 1$.

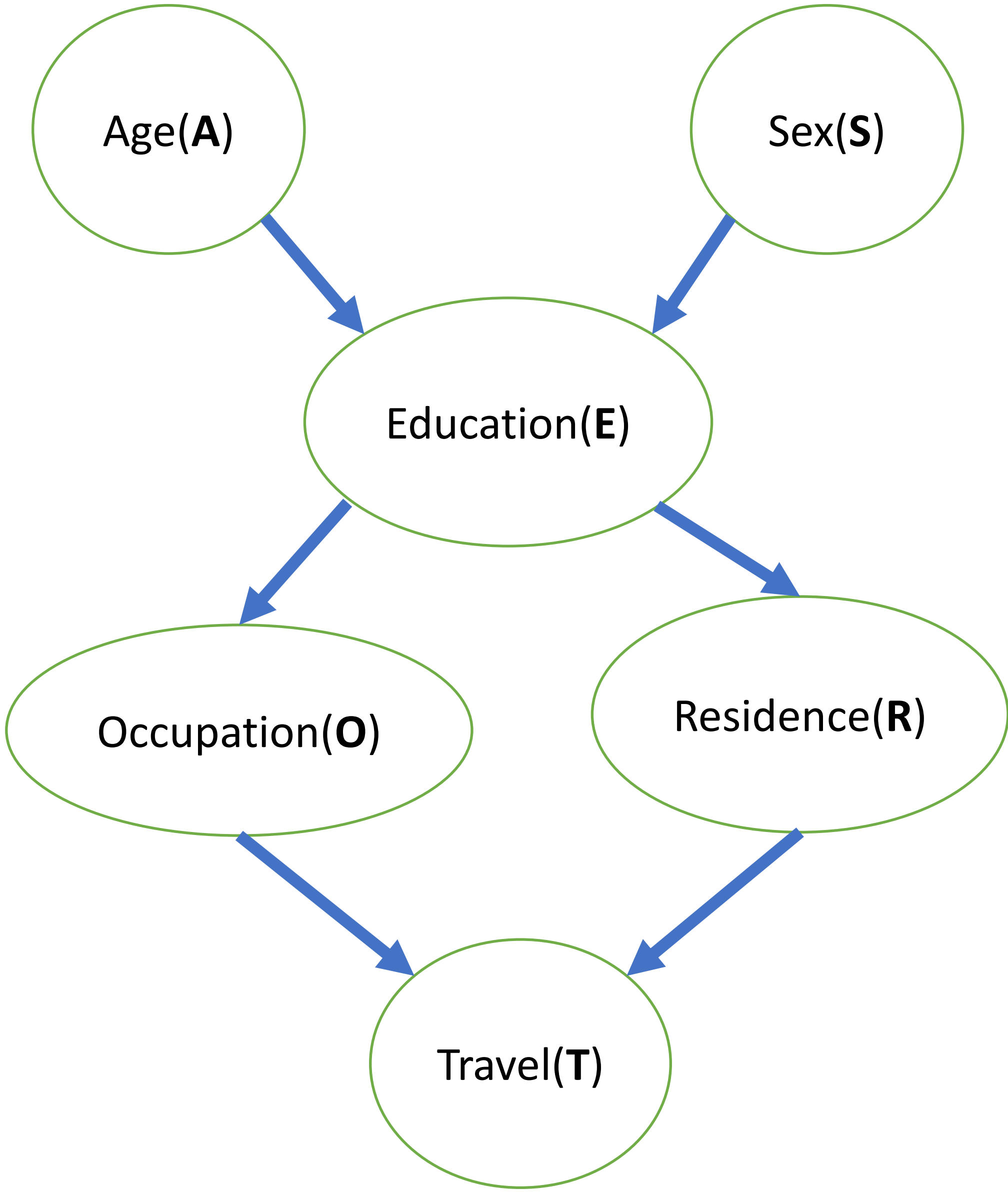
- Then for n repeated trials of the process, let x_i indicate the number of times that the result X_i occurs, subject to the restraints that $0 \leq x_i \leq n \; and \; \sum x_i = n$.

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1! \, x_2! \, \ldots \, x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

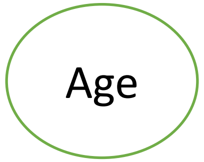
Using BNs it can be decomposed.

Example of Discrete BNs

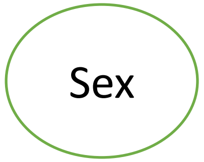
- **Age** and sex are not influenced by any other variable.
- **Age** and sex have a direct influence on **Education.**
- **Education** strongly influences both **occupation** and **residence.**
- **Transports** are directly influenced by both **occupation** and residence.
- $\mathbf{2^6 - 1 = 63}$



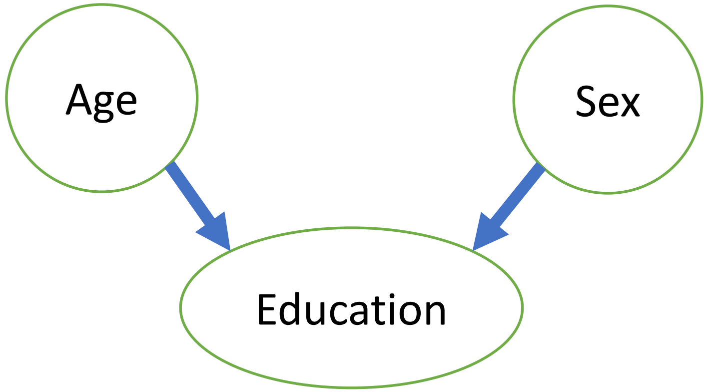
Young	Adult	Old
0.3	0.5	0.2



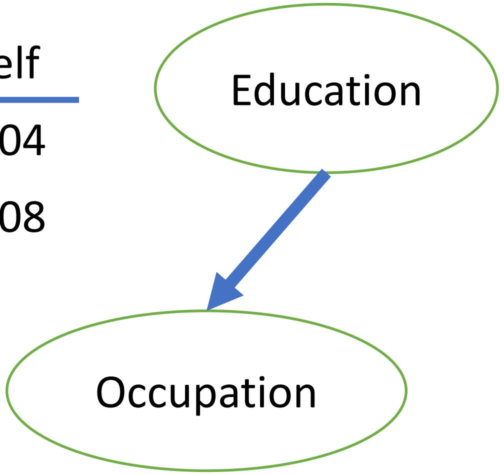
M	F
0.6	0.4



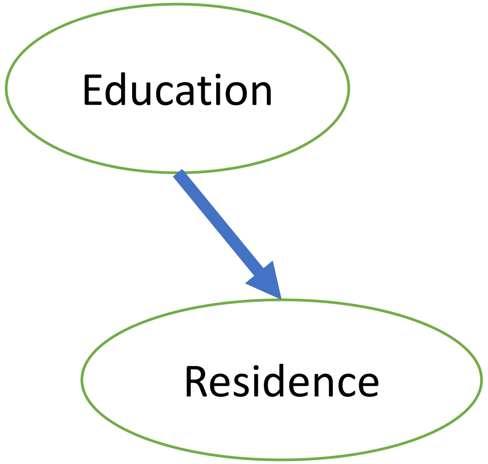
	High	Uni
Young& M	0.75	0.25
Adult& M	0.72	0.28
Old& M	0.88	0.12
Young& F	0.64	0.36
Adult& F	0.70	0.30
Old& F	0.90	0.10



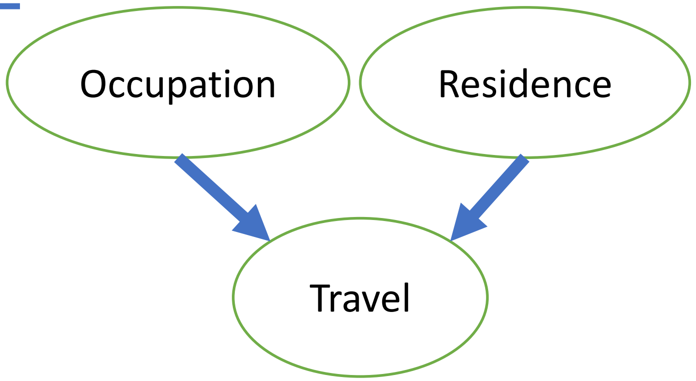
	Emp	Self
High	0.96	0.04
Uni	0.92	0.08



	Small	Big
High	0.25	0.75
Uni	0.20	0.80



	Car	Train	Other
Small& Emp	0.48	0.42	0.10
Small& Self	0.56	0.36	0.08
Big& Emp	0.58	0.24	0.18
Big & Self	0.70	0.21	0.09

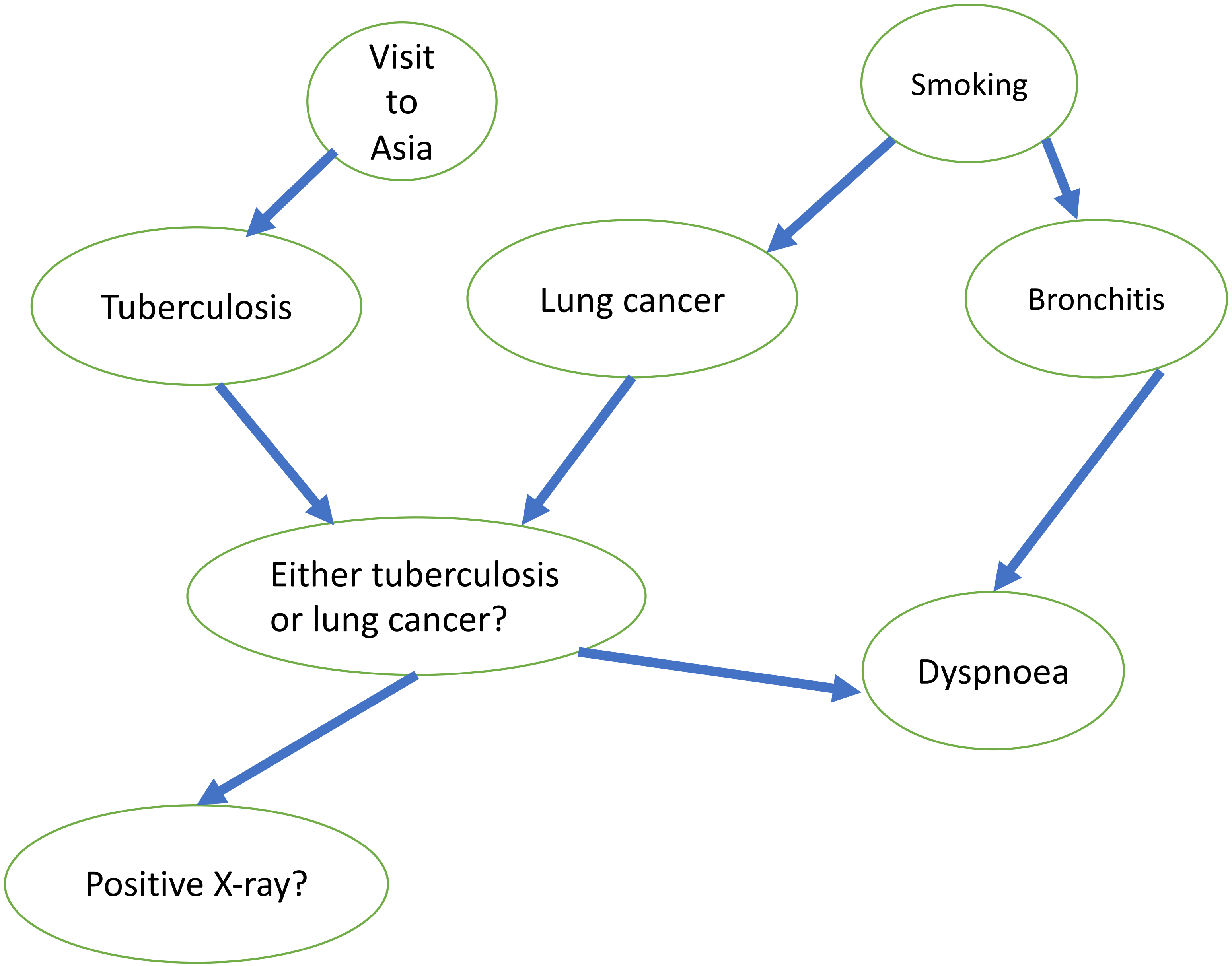


Overall, fewer parameters:

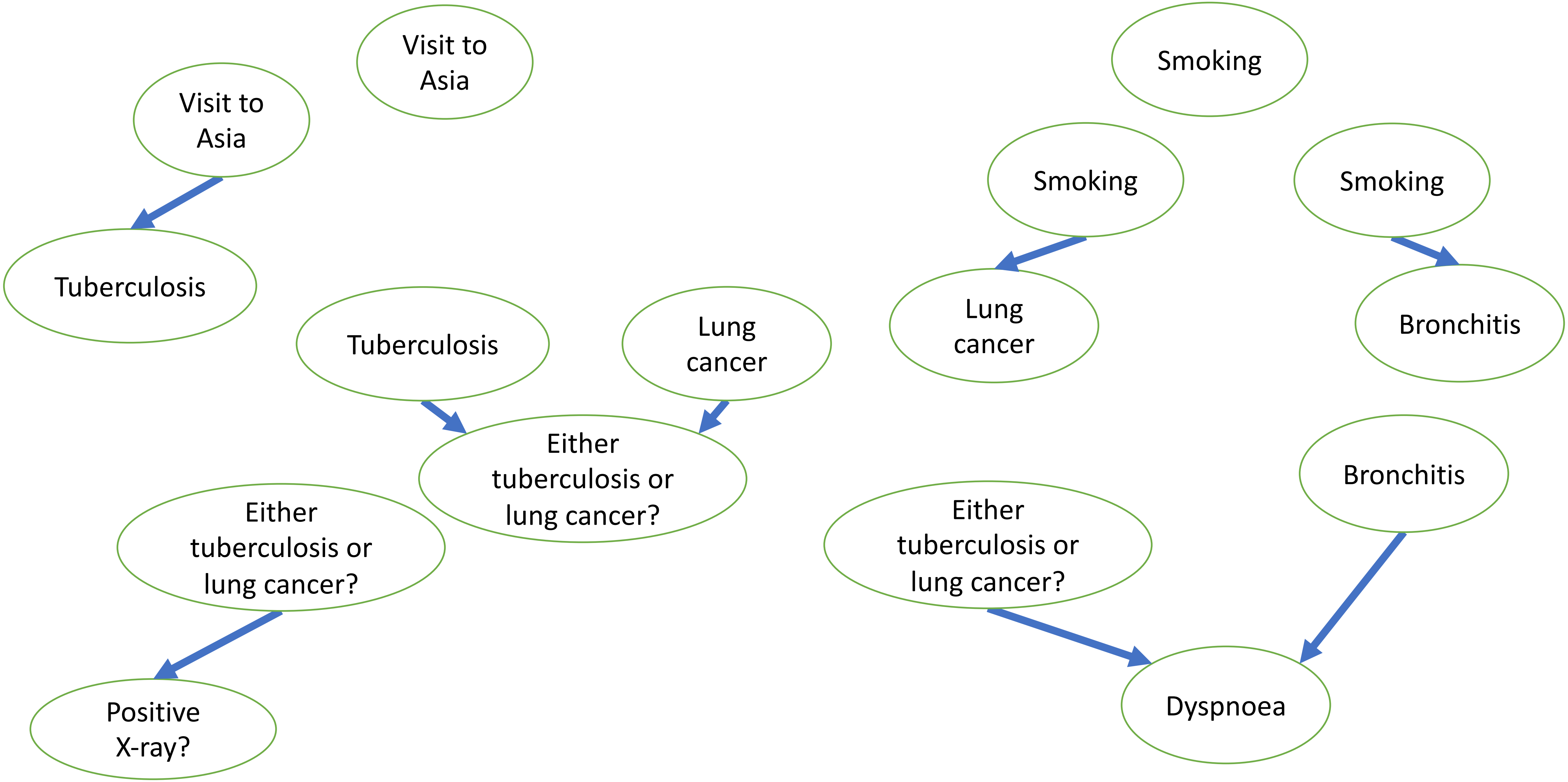
16 parameters only !

A **classic example of BN** is **the ASIA network** from Lauritzen & Spiegelhalter (1988), which includes a collection of binary variables. It describes a simple diagnostic problem for tuberculosis and lung cancer. Total parameters of X :

$\mathbf{2^8 - 1 = 255}$



- **Overall parameters of the $X_i \mid X_I$'s parents : 18**



Learning the dag structure:

- It is not always possible or desired to rely on prior knowledge of the phenomenon we are modeling to decide which arcs are present in the graph and which are not.

- **Therefore, the structure of the DAG itself maybe the object of our investigation.**

- E.g., in genetics and systems biology reconstructing the molecular pathways and networks underlying complex disease and metabolic processes (Sachs et al., (2005)).

Learning the dag structure:

- Several algorithms have been presented in literature for this problem,
- Despite the variety of theoretical backgrounds and terminology they can all be traced to only three approaches:

- *constraint-based,*
- *score-based* and
- *hybrid*.

Constraint-based :

Conditional independence test

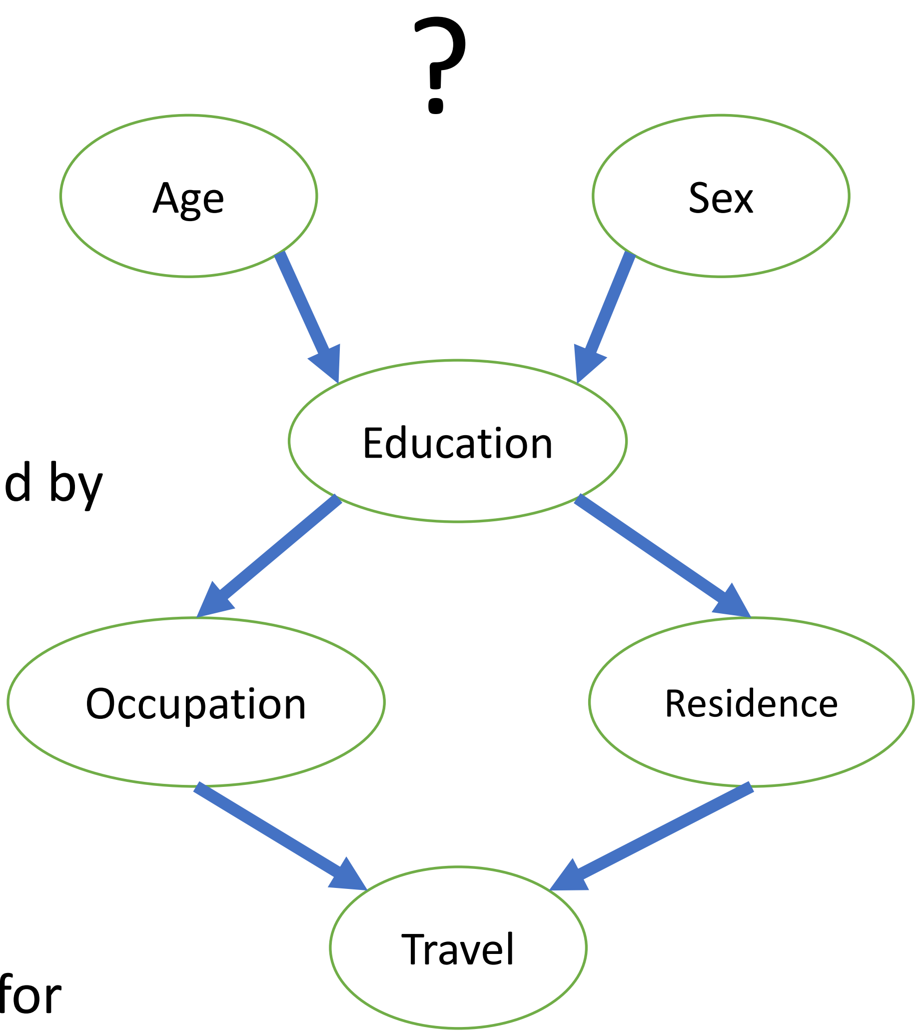
(Pearl 1990, Verma and Pearl, 1991)

- It focuses on presence of individual arcs.
- It can be used to assess whether the dependency is supported by the data.

- H_0: T $\perp\!\!\!\perp_{\boldsymbol{P}}$ E |{O,R}, they are independent, no edge
- H_1: T $\not\perp\!\!\!\perp_{\boldsymbol{P}}$ E |{O,R}, they are dependent, edge

- If the **null hypothesis** is rejected, the arcs can be considered for **inclusion** otherwise for **exclusion**.

- This approach operates edgewise.

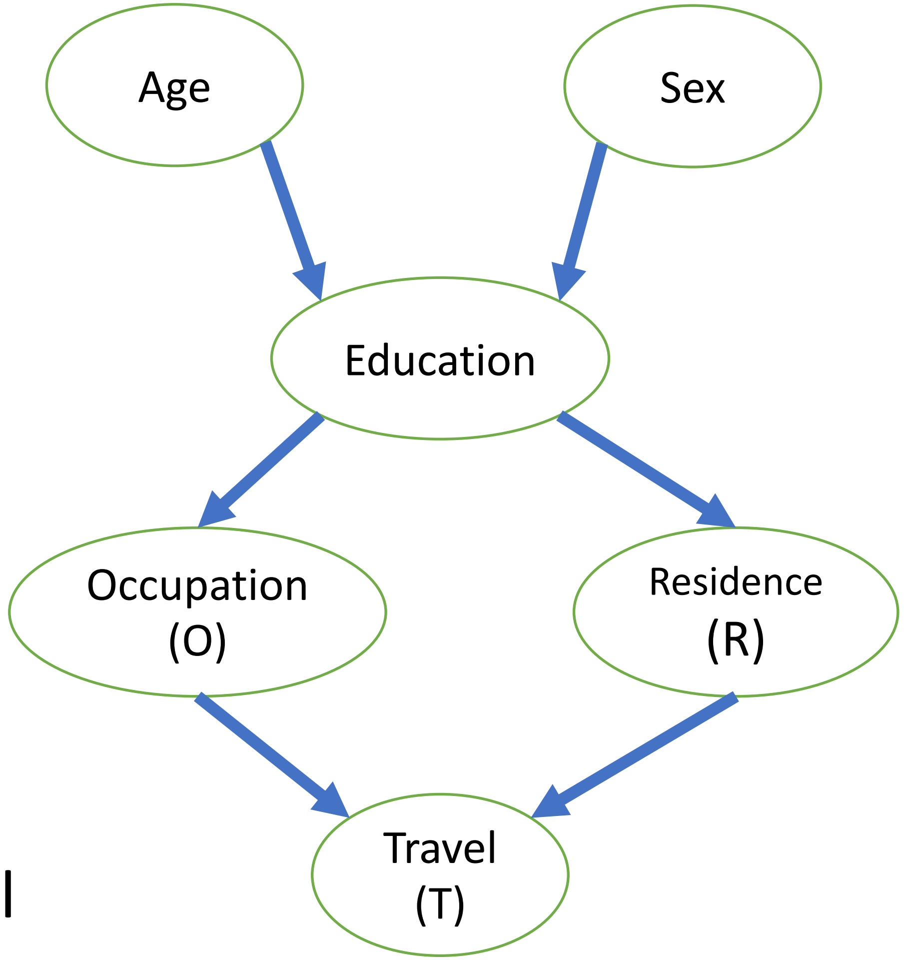


Conditional independence test

- H_0: T $\perp\!\!\!\perp_{\boldsymbol{P}}$ E |{O,R},
- H_1: T $\not\perp\!\!\!\perp_{\boldsymbol{P}}$ E |{O,R}

- Performing this test H_0 by adapting the **log-likelihood ratio**, G^2, or **Pearson's** X^2 test.

- If the **null hypothesis** is rejected, the arcs can be considered for **inclusion** otherwise for **exclusion** (small p-value).



Conditional independence test

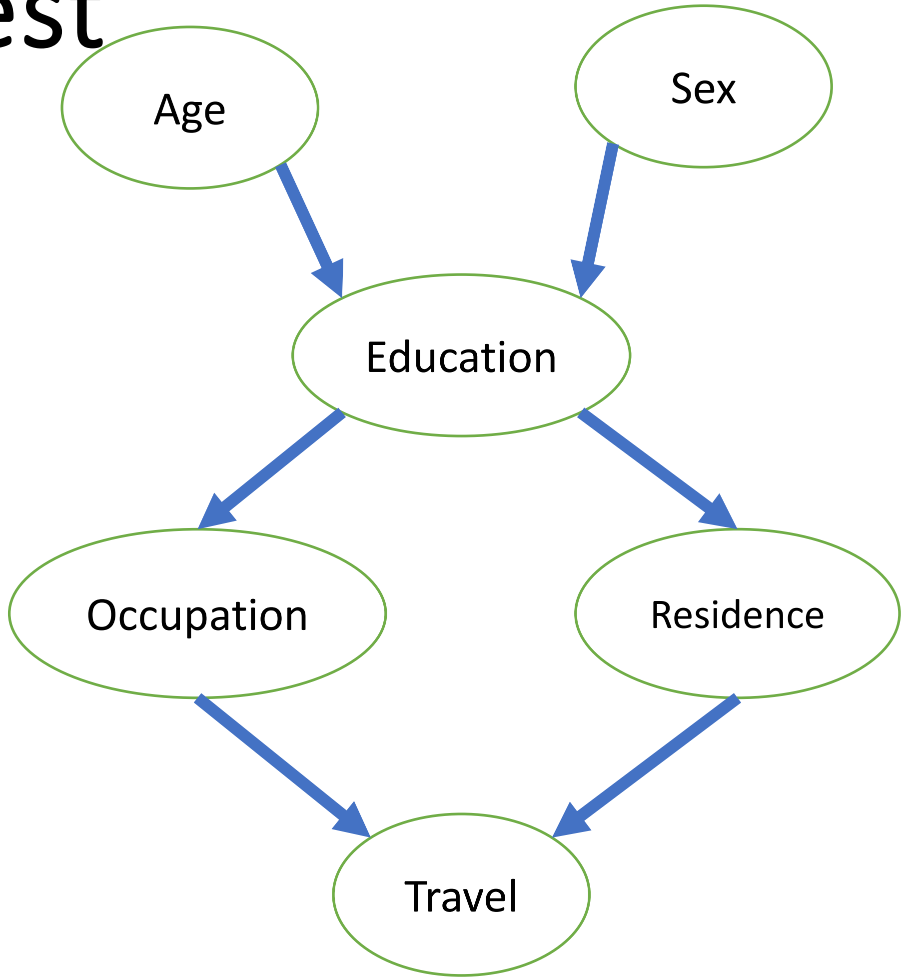
- **the null hypothesis** ***<u>is not</u>*** rejected,

  p-value > α ---- there is no edge.

- **The null hypothesis** is rejected

  p-value < α----- there is an edge.

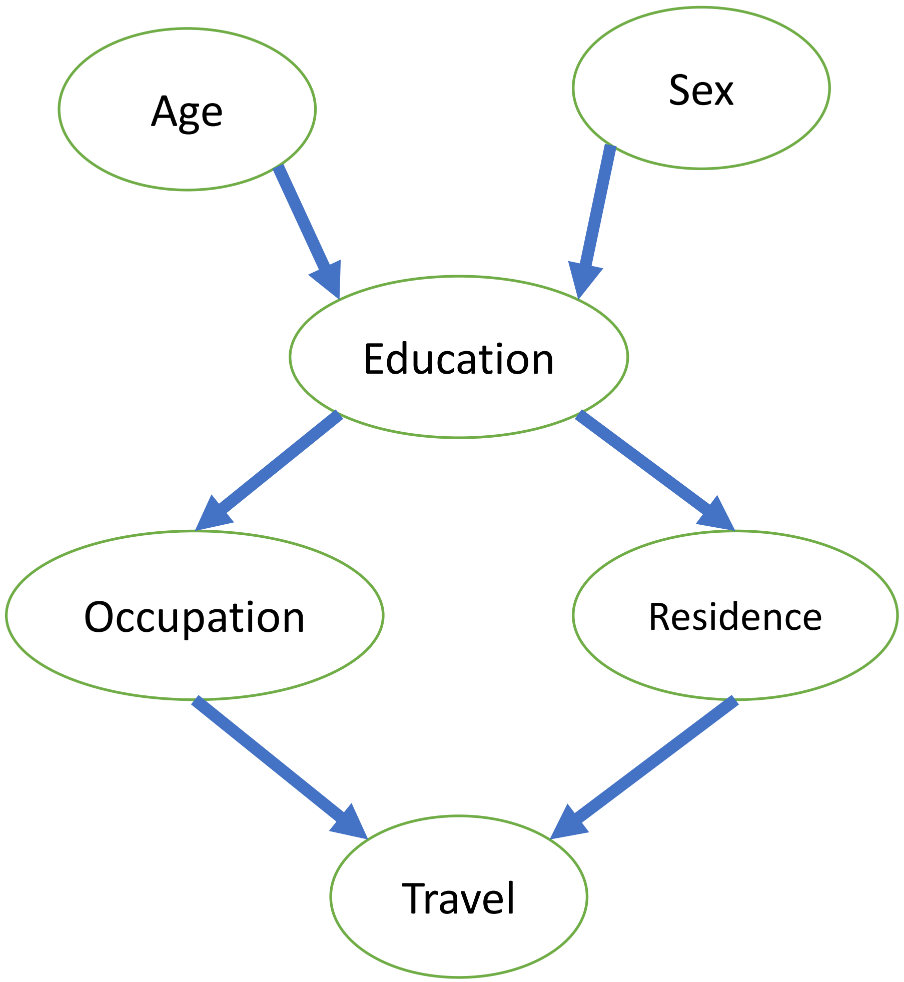
- α is usually 0.05 or 0.01.
- We will discuss this in practical.



Network score

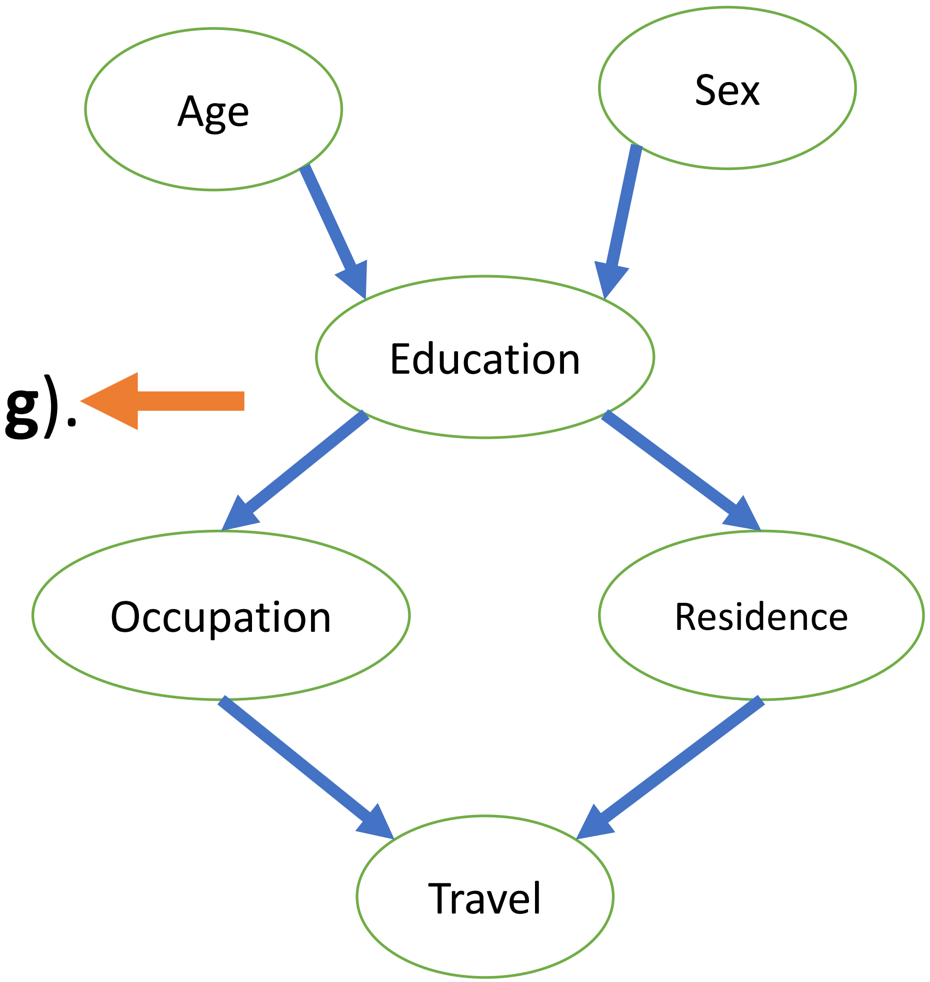
- Each candidate BN is assigned a network score reflecting its **goodness of fit.**

- Bayesian information criterion (**BIC**).
- Bayesian Dirichlet equivalent uniform (**BDeu**).



Algorithms that search for the DAG given the data (maximize a given network score)

- **Greedy search algorithm (such as Hill-climbing).**



- **Genetic algorithm.**
- **Simulated annealing** (Bouckaert, 1995).

Continuous (Gaussian) Bayesian Network

Continuous BNs

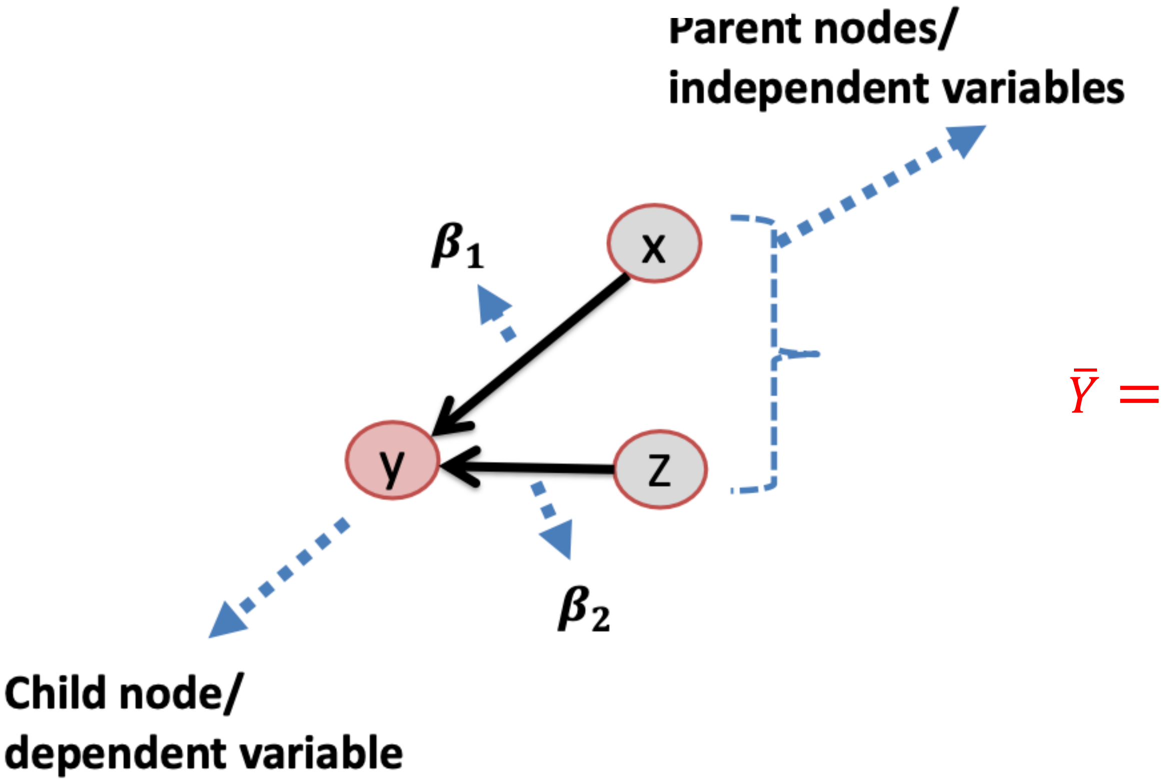
- Every node follows a **normal distribution**.

- Nodes without any parents (root nodes), are discribed by the univariate normal distribution.

- The local distribution of each node can be equivalently expressed as **a Guassian linear model** which includes an intercept and the node's parents as explanatory variables (predictors), **without any intraction term**.

- **Child~ $\boldsymbol{\beta_1}$*Parent1 + $\boldsymbol{\beta_2}$ *parent2 +…**

Edges: presence or absence of a coefficient



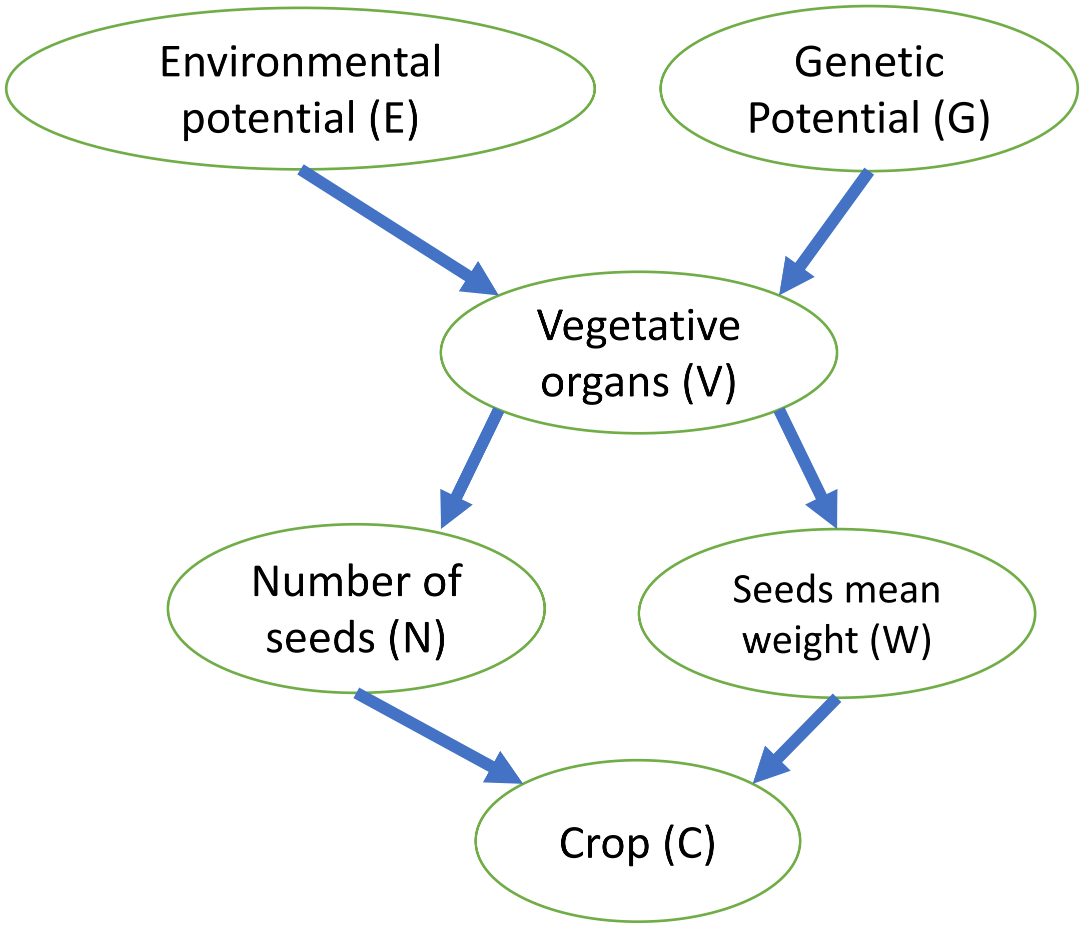
$$\bar{Y} = 0.1\boldsymbol{X} + 0.8\boldsymbol{Z}$$

$\overline{Child}$ ~ β_1 *Parent1 + β_2 *parent2 +…

Example of continuous BNs

For the analysis of a particular plant:

- **Genetic Potential(G):** Genotype effect (a single score).
- **Environmental potential (E)**: Environmental (location and season) effect (a single score).
- **Vegetative organs (V):** Roots, stems, etc., grow and accumulate reserves exploited for reproduction and summarises all the information available on constituted reserves.
- **Number of seeds (N)** is determined at the flowering time.
- **Seeds mean weight (W)** is assessed in the plant’s life.
- **Crop (C):** The harvasted grain mass.



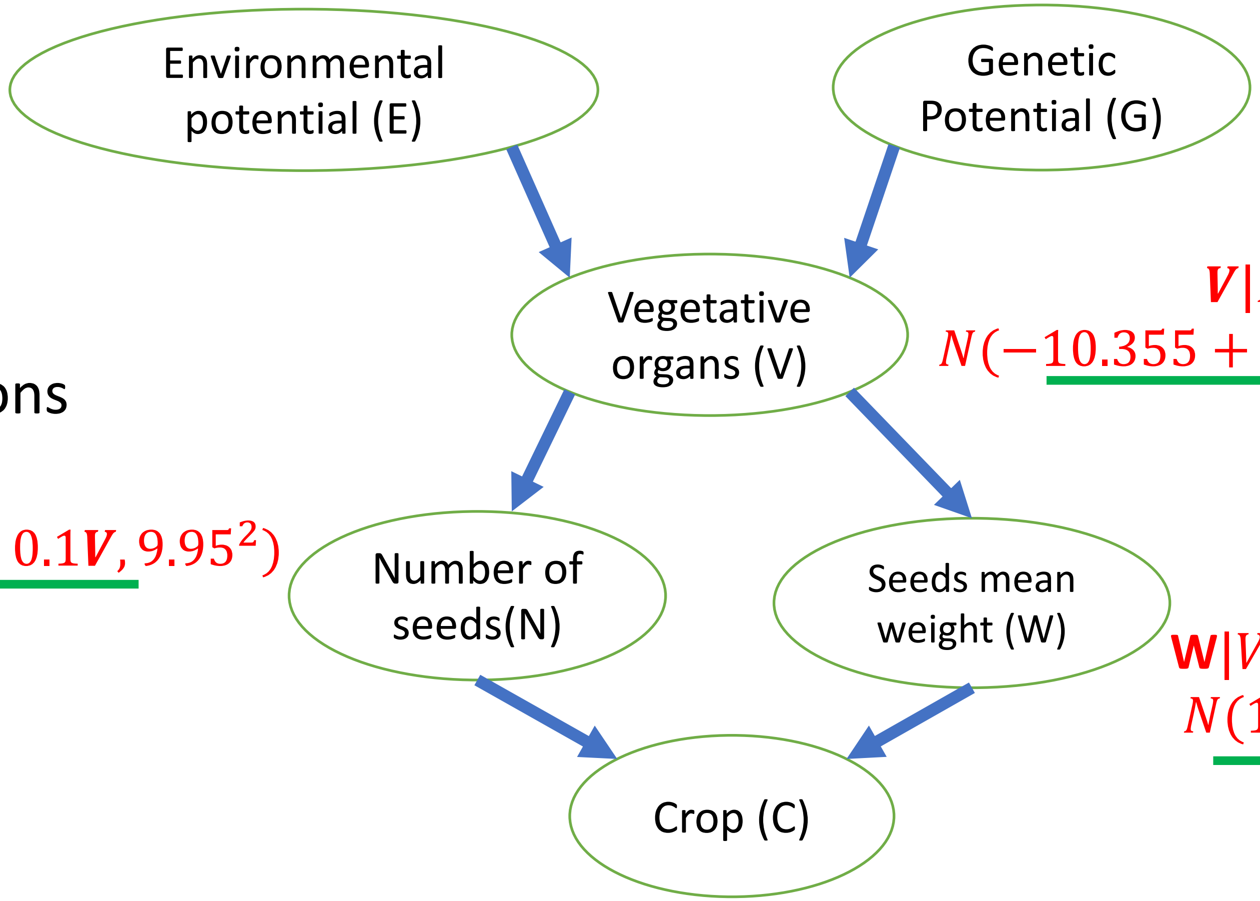
Example

- Six variables and six arcs corresponding to the direct dependencies linking them.

- The local probability distributions are shown for each node.

$E \sim N(50, 10^2)$

$\mathrm{G} \sim N(50, 10^2)$



$\boldsymbol{V}|E, G \sim N(\underline{-10.355 + 0.5\boldsymbol{G} + 0.77\boldsymbol{E}}, 5^2)$

$\mathbf{N}|V \sim N(\underline{45 + 0.1\boldsymbol{V}}, 9.95^2)$

$\mathbf{W}|V \sim N(\underline{15 + 0.7\boldsymbol{V}}, 7.171^2)$

$\mathbf{C}|N, W \sim N(\underline{0.3\boldsymbol{N} + 0.3\boldsymbol{W}}, 6.25^2)$

N means normal/ Gaussian distribution.

$\overline{Child}$ ~ β_1 *Parent1 + β_2 *parent2 +…

Some remarks

- WHY linear dependencies?
- Closed-form results for many inference procedures.

- Relatively **simple models** often perform better than very **sophisticated ones**.

- $\overline{Child}$ ~ β_1 *Parent1 + β_2 *parent2 +…

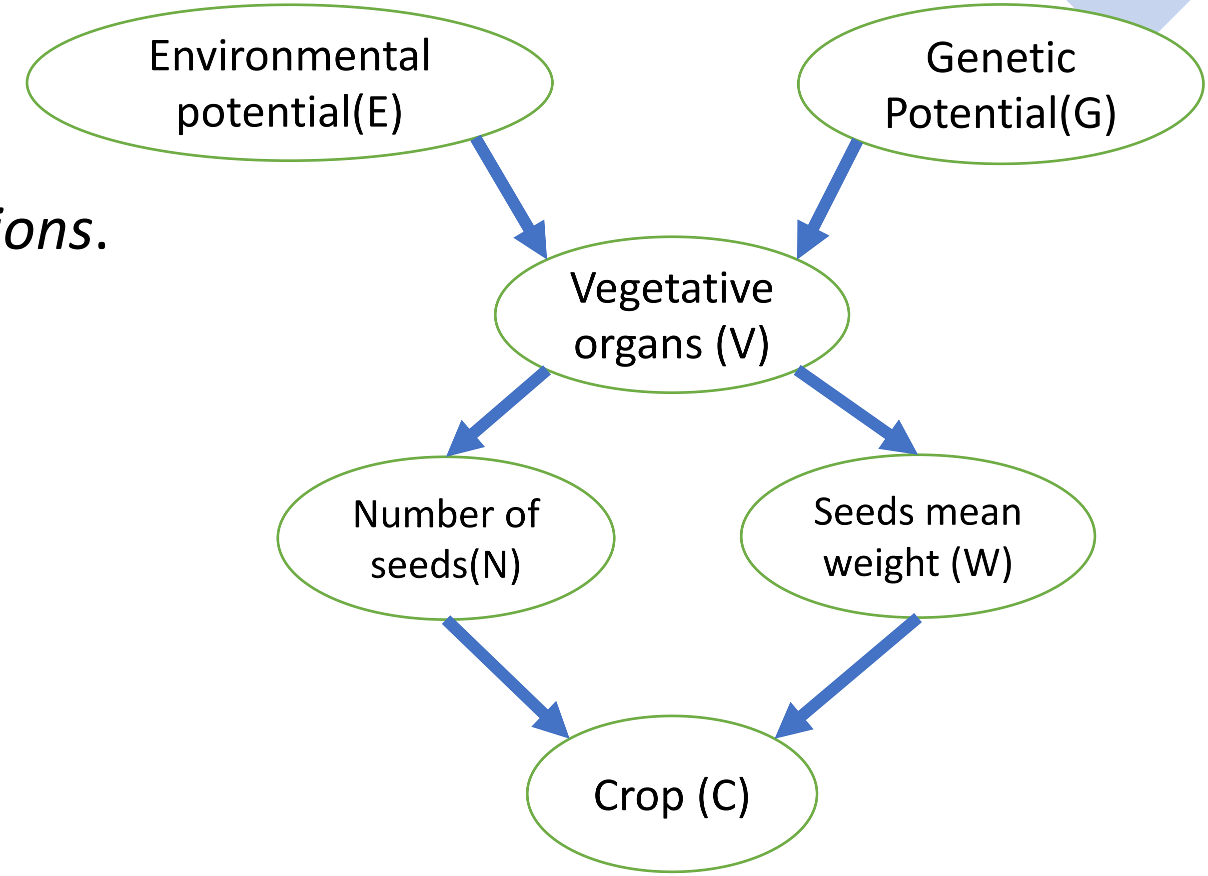
Learning the DAG Structure: Tests and Scores

- Often expert knowledge on the data is not detailed enough to completely specify the structure of the DAG. In such cases, if sufficient data are available, we can infer a sparse BN.

- The two classes of criteria used to learn the structure of the DAG are:
  - *constraint-based*
  - *score-based*

Conditional Independence Tests

- **Most common**: exact test for *partial correlations*.
- H_0: C $\perp\!\!\!\perp_{\boldsymbol{P}}$ W |N $\longrightarrow$ **no edge**
- H_1: C $\not\perp\!\!\!\perp_{\boldsymbol{P}}$ W |N $\longrightarrow$ edge between **C** and **W**



Conditional Independence Tests using <u>bnlearn</u> package

- ✓ Using "bnlearn" package
- ✓ Test for partial correlations
- ✓ Computing the corresponding statistics.
- ✓ P-value

- ✓ **The null hypothesis** ***<u>is not</u>*** rejected :

  p-value > α ---- there is no edge.

- ✓ **The null hypothesis** is rejected:

  p-value < α----- there is an edge.

- ✓ We will discuss this in practical.

Network Scores

Same as Discrete BN:

- **BIC**
- **BGe (**Bayesian Guassian equvalent score)

✓Search for the best network structure.

Summary

- Bayesian networks

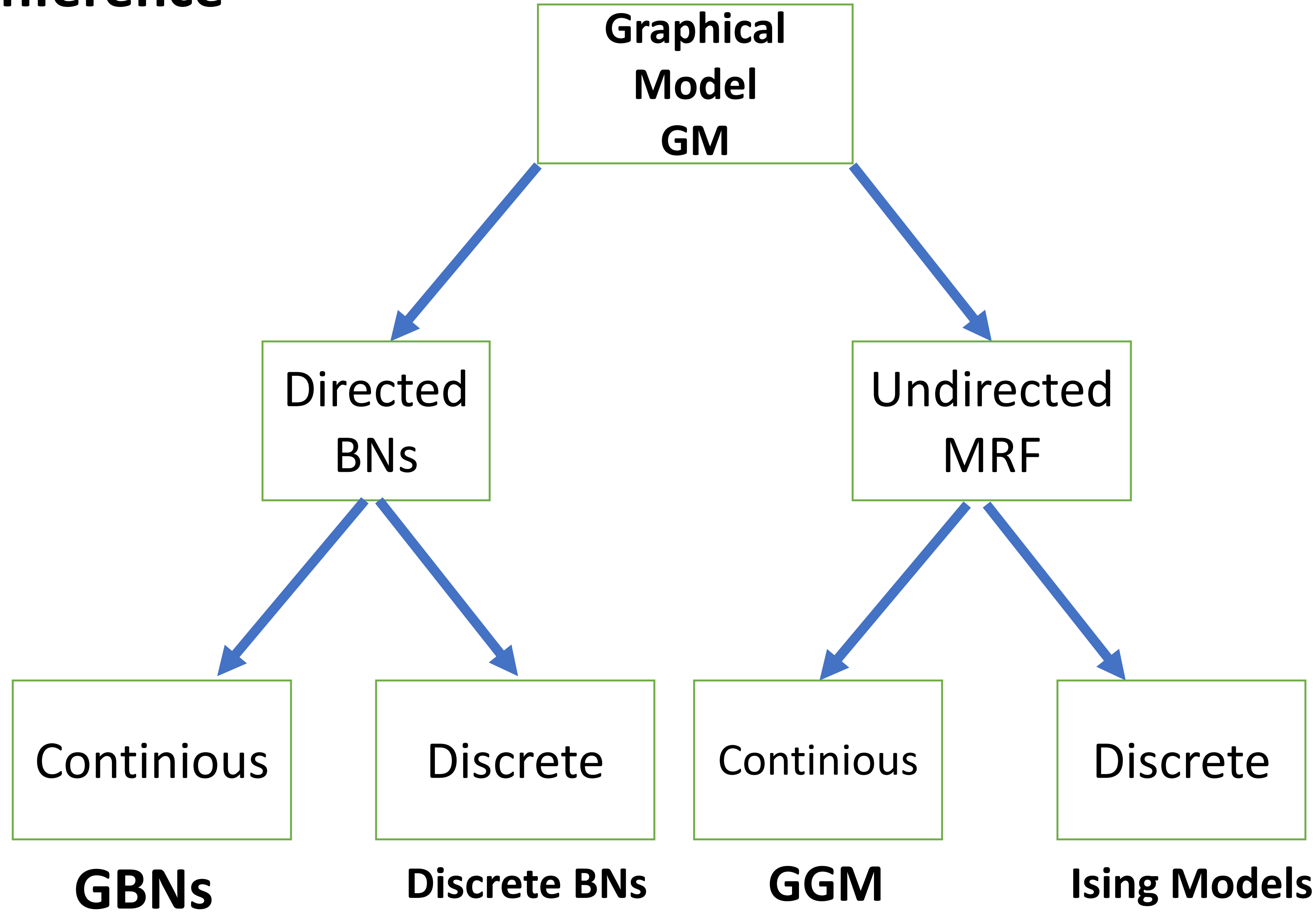
- are a combination of a DAG and a global distribution, both defined on the same variables.

- Provide a systematic decomposition of the global distribution into lower-dimensional local distributions, in a divide-and-conquer way.

- Provide a principled solution to the problem of feature selection using Markov blankets.

- ***Can be a very useful tool for Network reconstruction.***

Network Inference

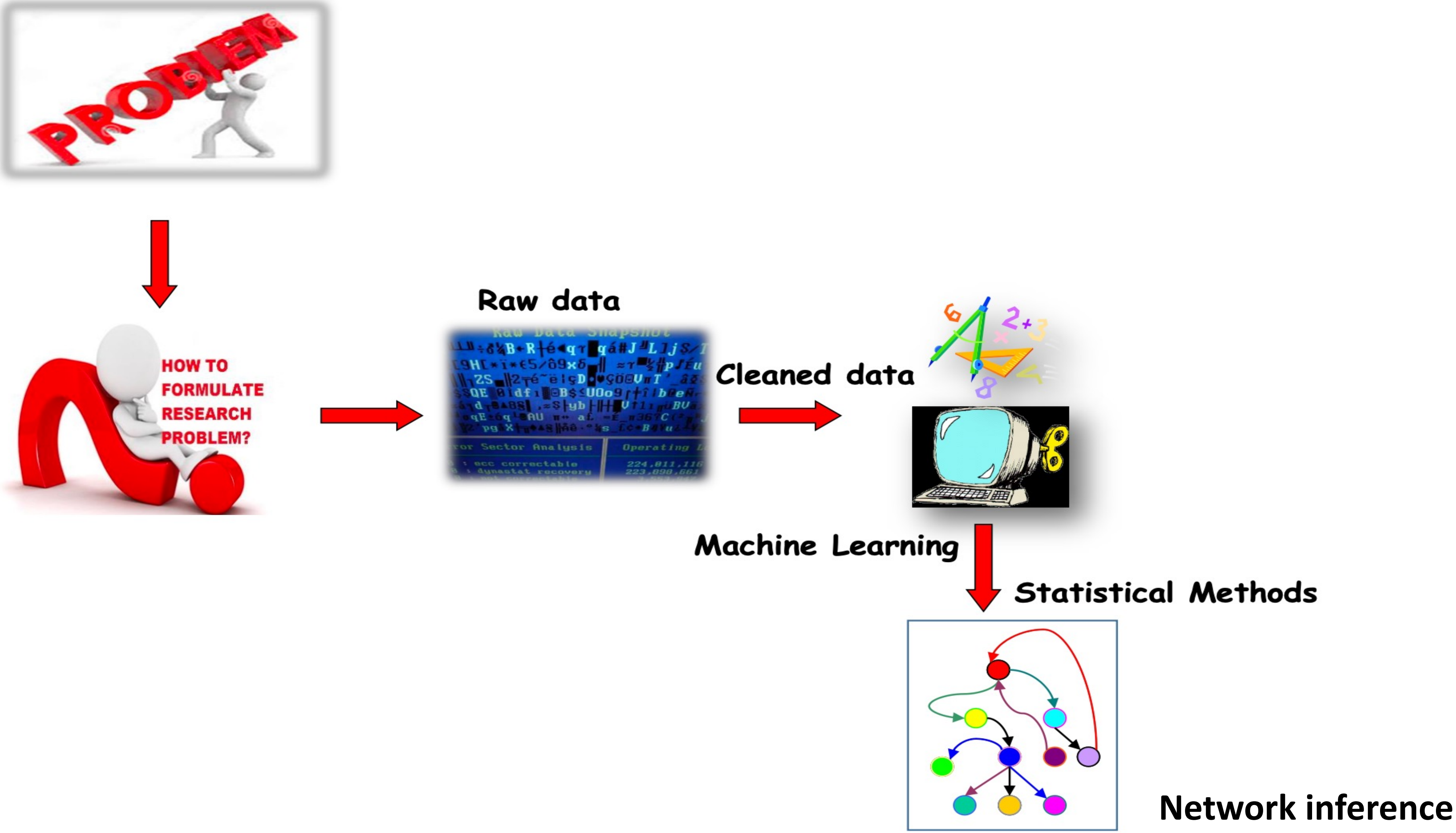


Finally, practically: *MRFs vs BNs*

- **MRFs** have more power than **BNs**, but are more difficult to interpret and deal with computationally.

- A general rule of thumb is to use **Bayesian networks** whenever possible, and only switch to MRFs if there is no natural way to model the problem with a directed graph.

In a nutshell



- Think about the application of this concept in your field for a few minutes and discuss that in pairs.

- Consider the following questions for reflection in your field:

✓What variables are present in your project?

✓Why are these variables important?

✓What motivates your interest in understanding their interdependencies?

✓How does this understanding contribute to your work or goals?

References

- Dechter, R. (2019). ***Reasoning with Probabilistic and Deterministic Graphical Models****: Exact Algorithms*. Morgan & Claypool publishers. https://doi.org/10.2200/S00893ED2V01Y201901AIM041
- Højsgaard, S., Edwards, D., & Lauritzen, S. (2012). ***Graphical Models with R***. Springer New York, NY. https://doi.org/10.1007/978-1-4614-2299-0
- Koller, D.& Friedman, N. (2010). ***Probabilistic Graphical Models: Principles and Techniques***. The MIT Press Cambridge, Massachusetts.
- Nagarajan, R., Scutari, M. & Lébre, S. (2013). ***Bayesian Networks in R****: with Applications in Systems Biology*. Springer New York, NY. https://doi.org/10.1007/978-1-4614-6446-4