Graphical Models for Network Inference

Mahdi Shafiee Kamalabad

Utrecht University

Graphical Models for Network Inference

Today's roadmap

Time	Focus	Topics
Intro	What is a graphical model?	Core idea & notation
Morning	**Undirected graphical models** (Markov Random Fields)	• Continuous variables • Discrete variables
Afternoon	**Directed graphical models** (Bayesian Networks)	• Continuous variables • Discrete variables • Mixed / hybrid models

First, let’s not forget our goal.

We would like to infer a network from data that shows the dependencies between all variables.

So, what is a network?

Network, graph, which one?

Graph

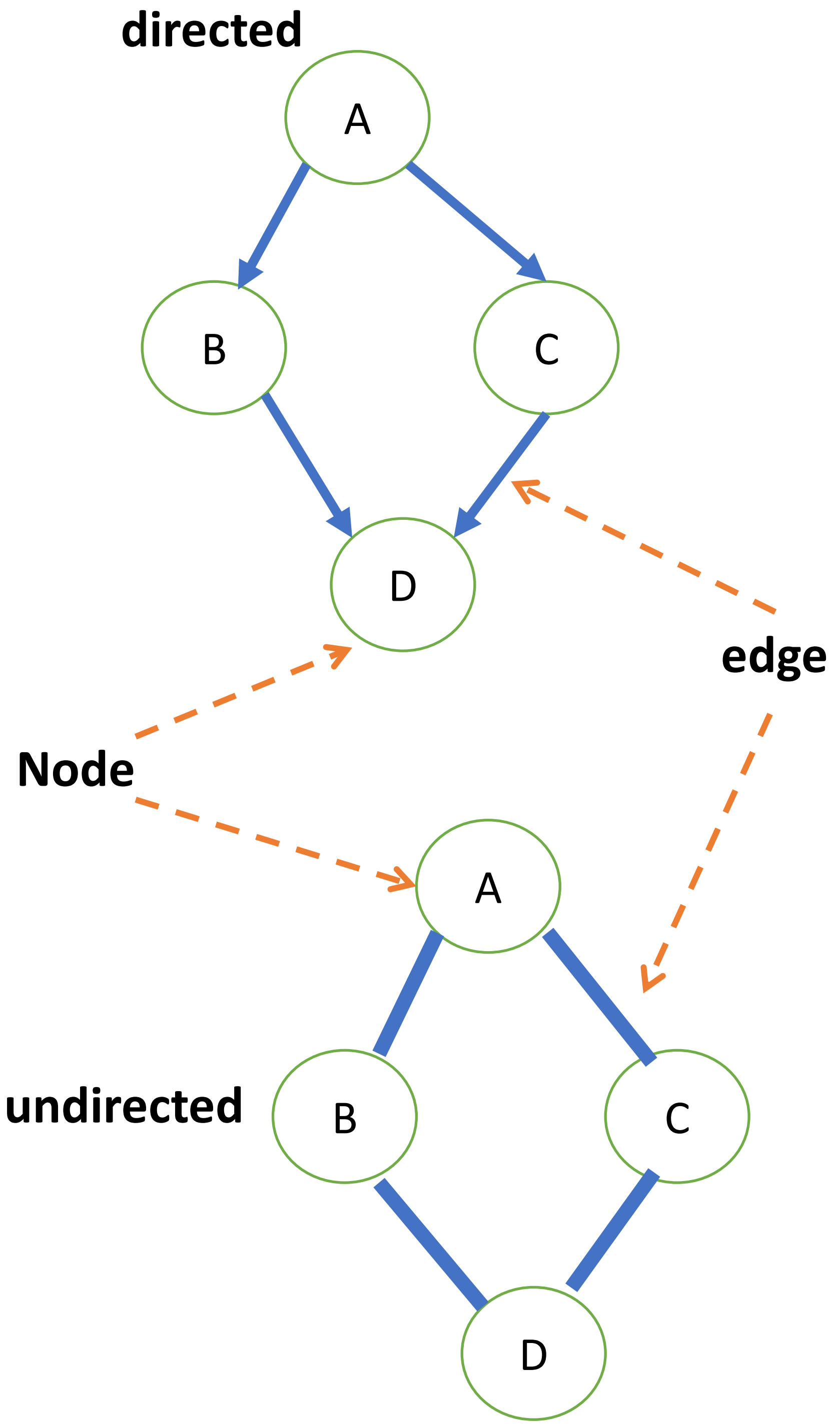
- We define a graph G by the following equation:

  $G = (V, E)$

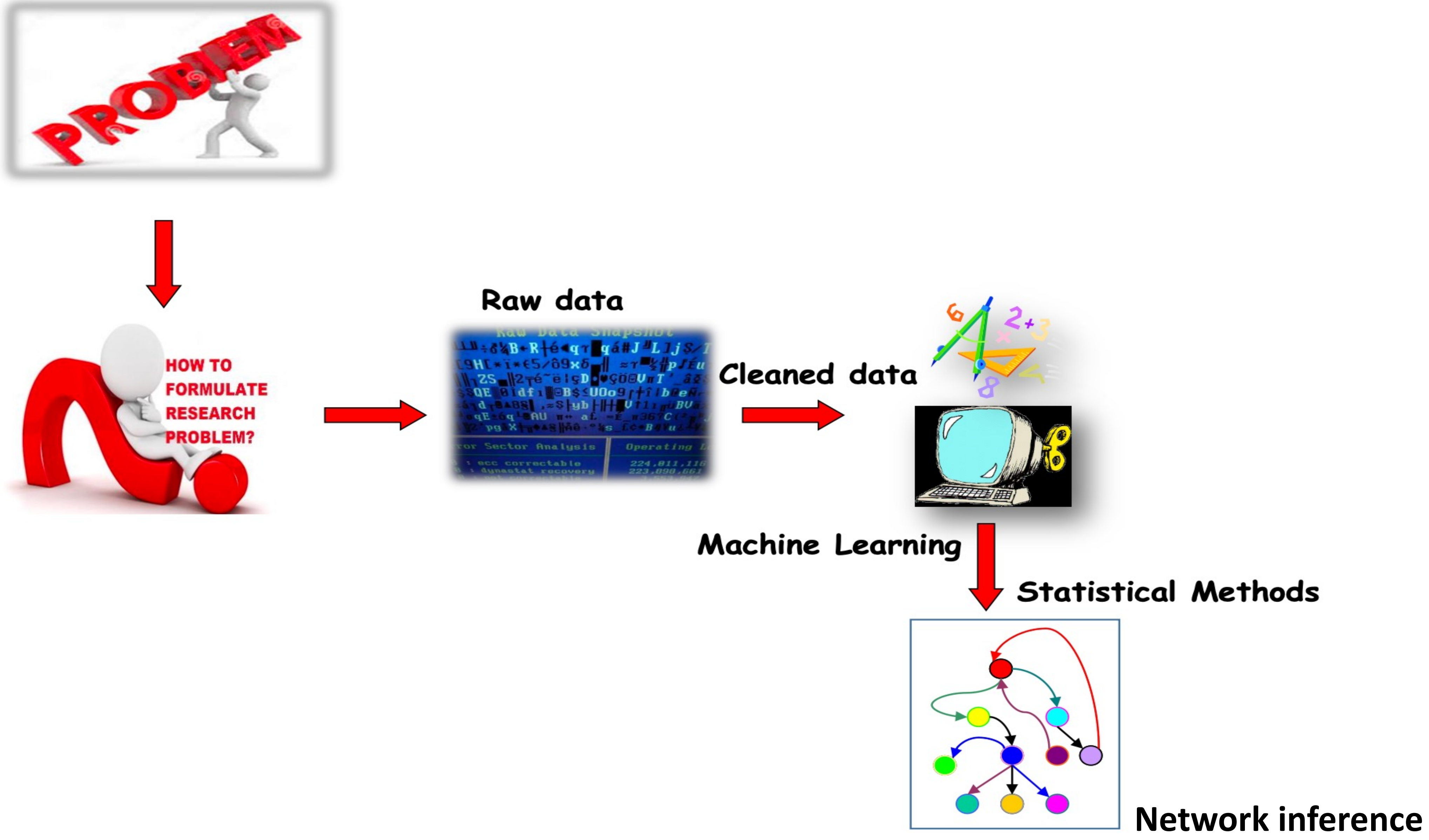
Here:

- V is a finite **set of vertices** or nodes
- $E \subseteq V \times V$ is a finite **set of edges**, links, or arcs.

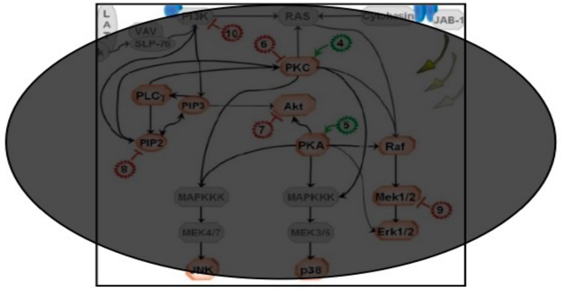
- Today we assume **nodes are variables** and **edges represent statistical relationships**.
- We use **graphical models** to **learn** the underlying network structure.



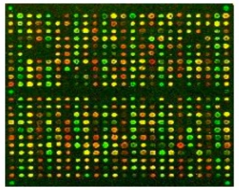
In a nutshell



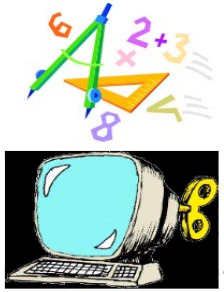
possibly completely unknown



E.g.: Gene-Microarry experiments

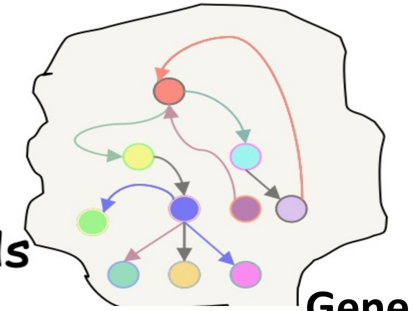


data data



Machine Learning

statistical methods



Gene regularity Network

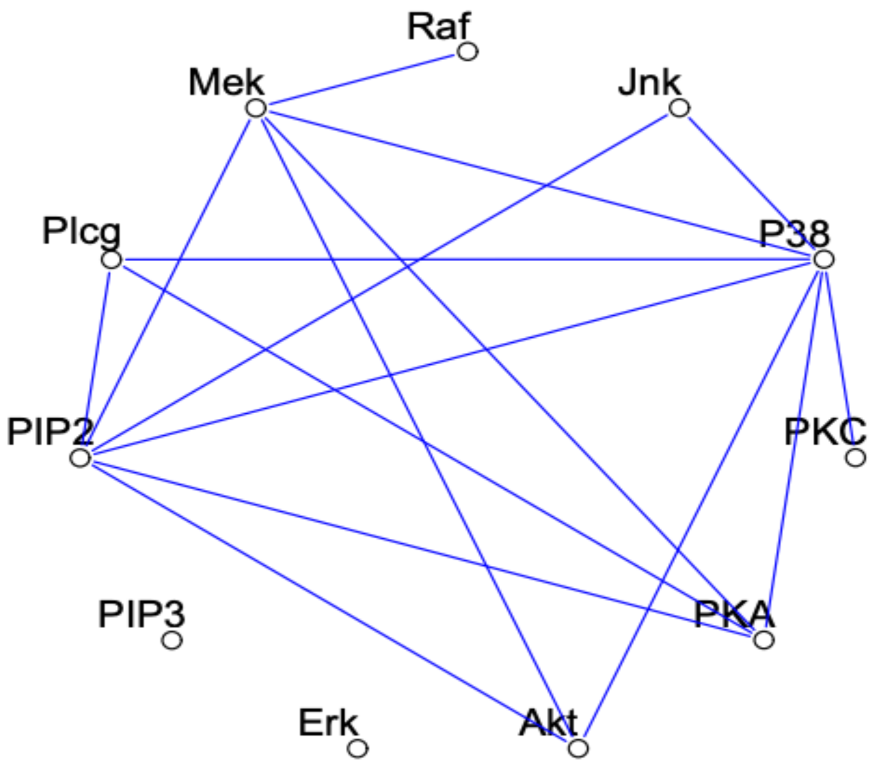
How should the data structure be designed?

A flow cytometry dataset, Protein-protein

RQ: How do proteins interact and regulate each other in single cells based on flow cytometry measurements?

Raf	Mek	Plcg	PIP2	PIP3	Erk	Akt	PKA	PKC	P38	Jnk
26.4	13.2	8.82	18.3	58.8	6.61	17	414	17	44.9	40
35.9	16.5	12.3	16.8	8.13	18.6	32.5	352	3.37	16.5	61.5
59.4	44.1	14.6	10.2	13	14.9	32.5	403	11.4	31.9	19.5
73	82.8	23.1	13.5	1.29	5.83	11.8	528	13.7	28.6	23.1
33.7	19.8	5.19	9.73	24.8	21.1	46.1	305	4.66	25.7	81.3
18.8	3.75	17.6	22.1	10.9	11.9	25.7	610	13.7	49.1	57.8
44.9	36.5	10.4	132	16.3	8.66	17.9	835	15	35.9	18.1
47.4	15	14.6	30.5	17.5	20.2	45.3	466	6.44	24.4	20
104	61.5	10.6	21.1	41.8	11.5	23.5	445	29.2	61	25.3

Example 1: A flow cytometry dataset, **Undirected** netwok



RQ: How do proteins interact and regulate each other in single cells based on flow cytometry measurements?

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009)

A sparse undirected graph, estimated from a flow cytometry dataset, with p = 11 proteins measured on N = 7466 cells. The network structure was estimated using the graphical lasso which will be discussed later, **Sachs et al. (2003)**.

Example3 : Analysis of a particular plant

"C"	"E"	"G"	"N"	"V"	"W"
48.82.	51.48.	42.64.	54.09	42.96	41.95
48.85	73.42	40.97	60.06	65.28	48.96
67.02	71.098	52.52	51.64	63.22	62.03
37.83	49.33.	56.15.	49.00.	47.75	38.77
55.30	49.27	63.54	54.62	60.56	56.66
56.12	48.71	66.02	43.95	55.53	52.38

Genetic Potential(G): Genotype effect (a single score)

Environmental potentioal(E): Environmental (location and season) effect (a single score).

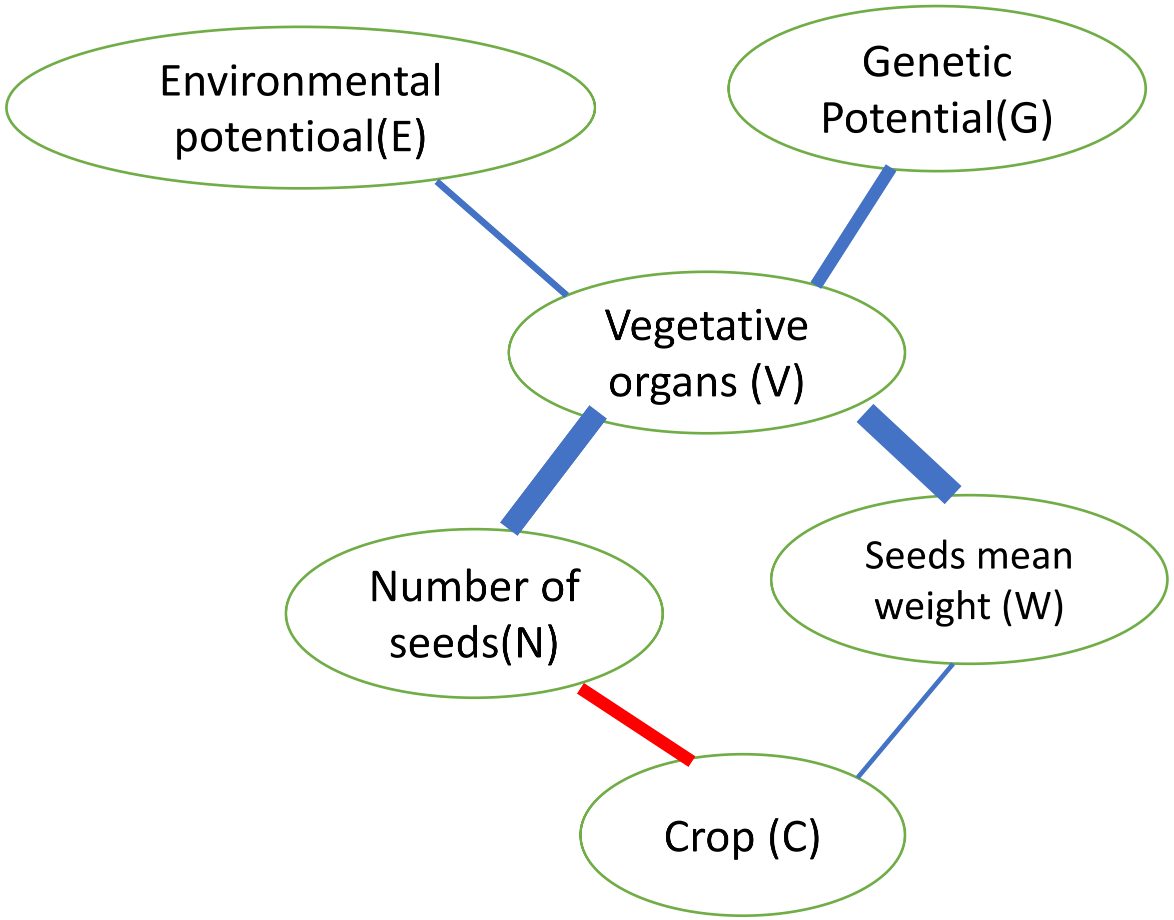
Vegetative organs (V): Roots, stems, etc., grow and accumulate reserves exploited for reproduction and summarises all the information available on constituted reserves.

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.

Example3 : Analysis of a particular plant



Aim

Learning the network structure from an n-by-m data matrix

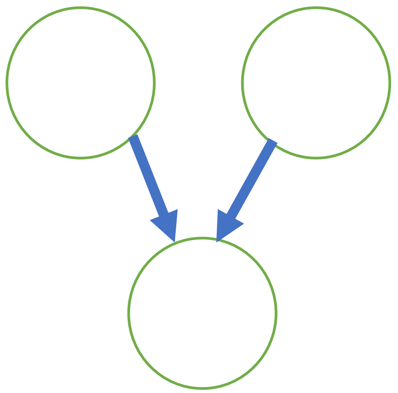
n **units or variables**

e.g. proteins/genes

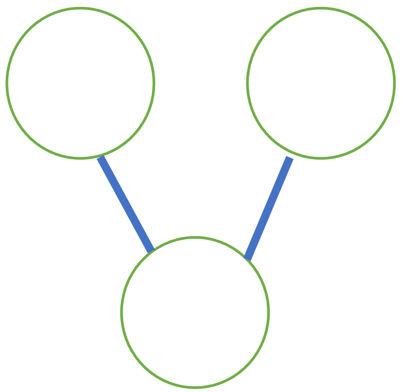
$$\begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}^t \longrightarrow$$

m data points

e.g. expression in various conditions



directed



undirected

Motivation

- **Discover statistical dependencies**
  Identify how variables relate to each other based on data and represent these relationships as a network.

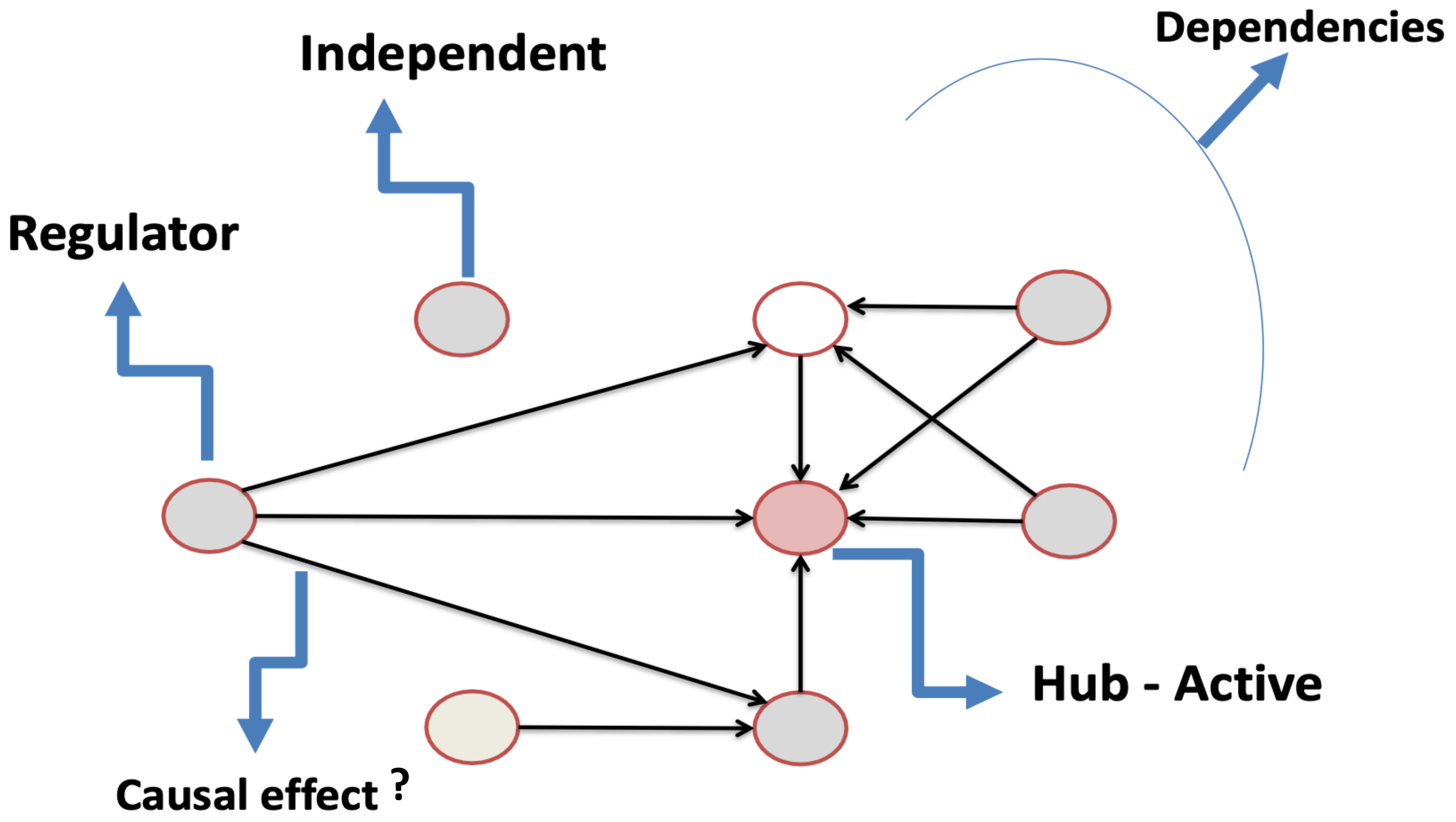
- **Quantify interaction strength**
  Capture not just connections, but the **magnitude of dependencies** between nodes.

- **Identify key players**
  Detect **hub nodes,** i.e., highly connected or active variables that may play central roles.

- **Reveal regulatory structure**
  Pinpoint **regulator nodes** that influence many others in the network.

- **Enable causal insights** *(in some cases)*
  Certain network models allow for **interpreting causal relationships** under specific assumptions.
- **Prediction**
- **…**

Network


Independent
Dependencies
Regulator
Hub - Active
Causal effect ?

Example: Why Learn Cellular Networks?

- A key goal in systems biology:

Understanding cellular networks, such as gene regulatory or signalling pathways.

- Why it matters:

For example,

- To identify **promising drug targets**
- To support **personalized medicine**
- To better understand **complex biological processes**

Example: How key macroeconomic indicators influence one another over time.

Node (random variable)	What it measures
GDP_growth	Quarterly real GDP growth rate
Inflation	CPI inflation rate
Unemployment	Jobless rate
InterestRate	Central bank policy rate
ExchangeRate	Domestic currency per USD
StockReturn	Broad market index return

- Why it matters:

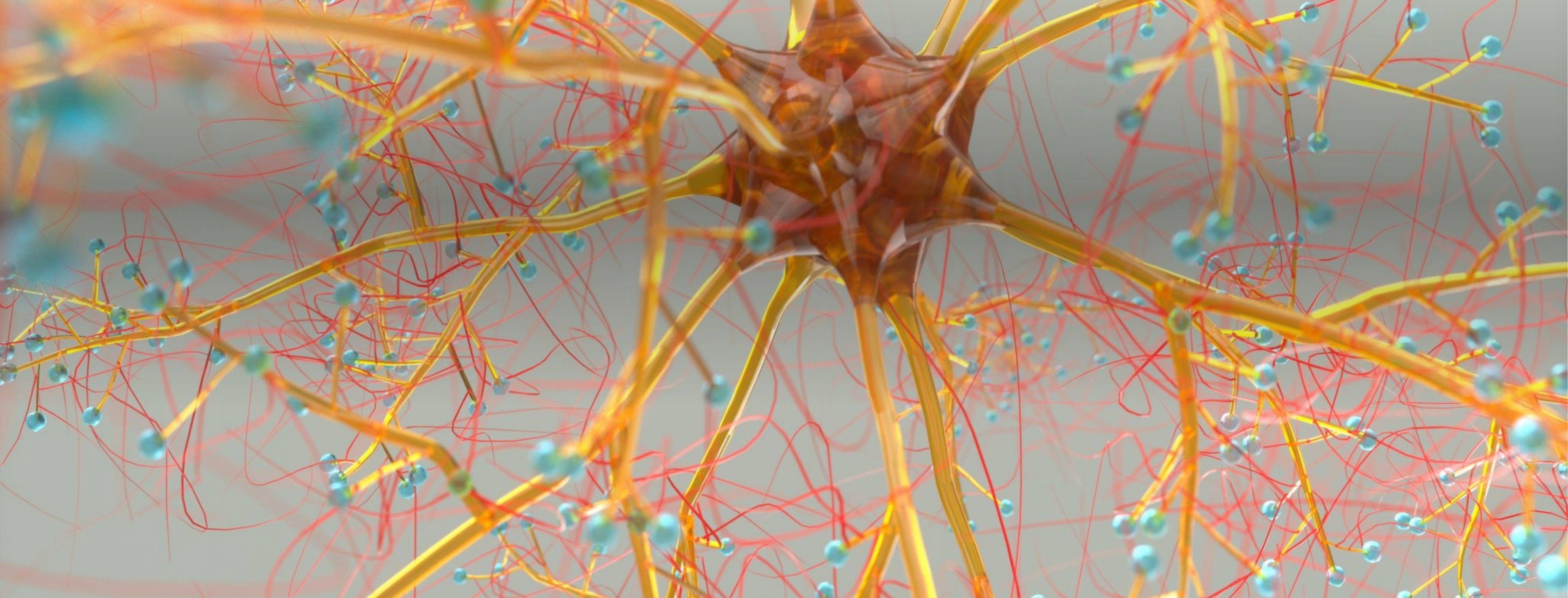
- **Explicit relationships:** Encode which variables directly influence one another through graph structure
- **Causal inference (with assumptions):** Estimate the effect of interventions (e.g., “What happens to inflation if we cut rates?”)
- **Structure learning:** Discover from data which indicators are conditionally independent vs. directly linked
- **Joint uncertainty propagation:** Track how uncertainty in one node ripples through to all others in the network
- **…**

Take a few minutes to consider how this concept might apply to your own field or to any imagined scenario and then share your thoughts with your group.

- Consider the following questions:

  ✓Which variables are present in your field?
  ✓Why are these variables important?
  ✓What motivates your interest in understanding their interdependencies?
  ✓How does this understanding contribute to your work or goals?

Graphical models **or**
Probabilistic graphical models (PGMs) **or**
Structured probabilistic models



Definition of Graphical Models
Michael Jordan, 1998

- A **Graphical model** is **a marriage** between **graph** theory and **probability** theory.
- A **graph** represents **relationships** among a set **of variables**.
- **Probability** helps to estimate the graph (network) structure.

Graphical model

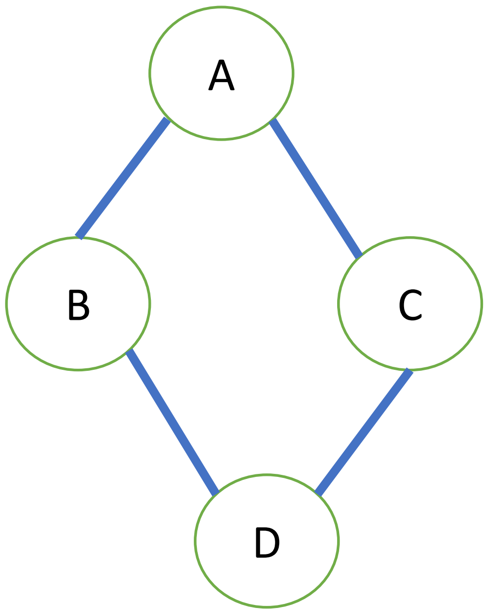
Michael Jordan, 1998

- **Probability:** provides the glue whereby the parts are combined, ensuring that the system as a whole is consistent, and providing ways to **interface models** to data.
- **Modularity** : a complex system is built by combining simpler parts.
- **Graph theory:** an intuitively appealing interface for modelling highly-interacting sets of variables

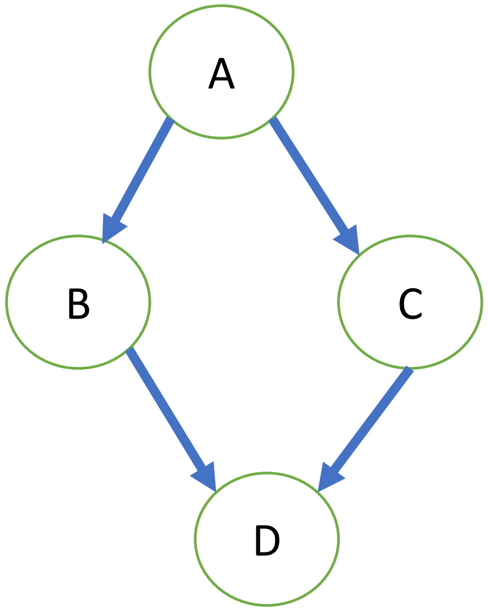
Graphical Models Applications in Real Life

- Biology and system biology
- Medical Science
- Finance
- Economics
- Handwriting Recognition
- Telecommunication Network Diagnosis
- Object Recognition in Images
- Relationship between symptoms and diseases
- **And many more**

<u>Edges</u> and <u>nodes</u> in a graphical model



- **Random variables** as **<u>nodes</u>**.
- **Statistical relationships** as **<u>edges.</u>**
- **Graphical models** : Encode <u>conditional</u> (in)dependence assumptions between variables.
- **What is conditional independence assumption?**



Independency

- The variables A and B are independent if

$$P(A)= P(A|B)$$

or

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

- The occurrence of event A does not give any information about B and the occurrence of event B does not give any information about A.

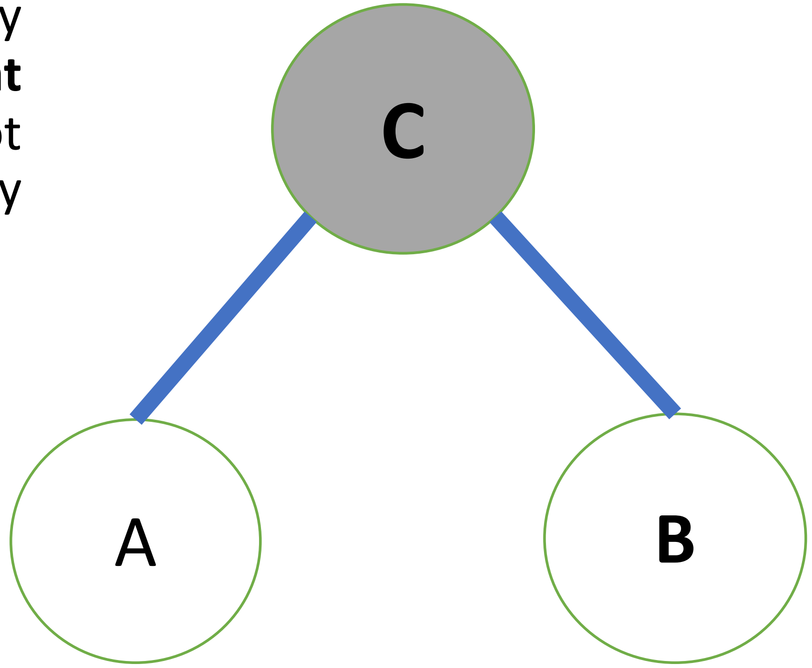
Conditional (in)dependency: (missing) edges

- Two nodes A and B ***given C*** are conditionally independent (no ***edge*** between them) if, **given that event C occurs**, the occurrence of event A does not give any information about B and the other way around.

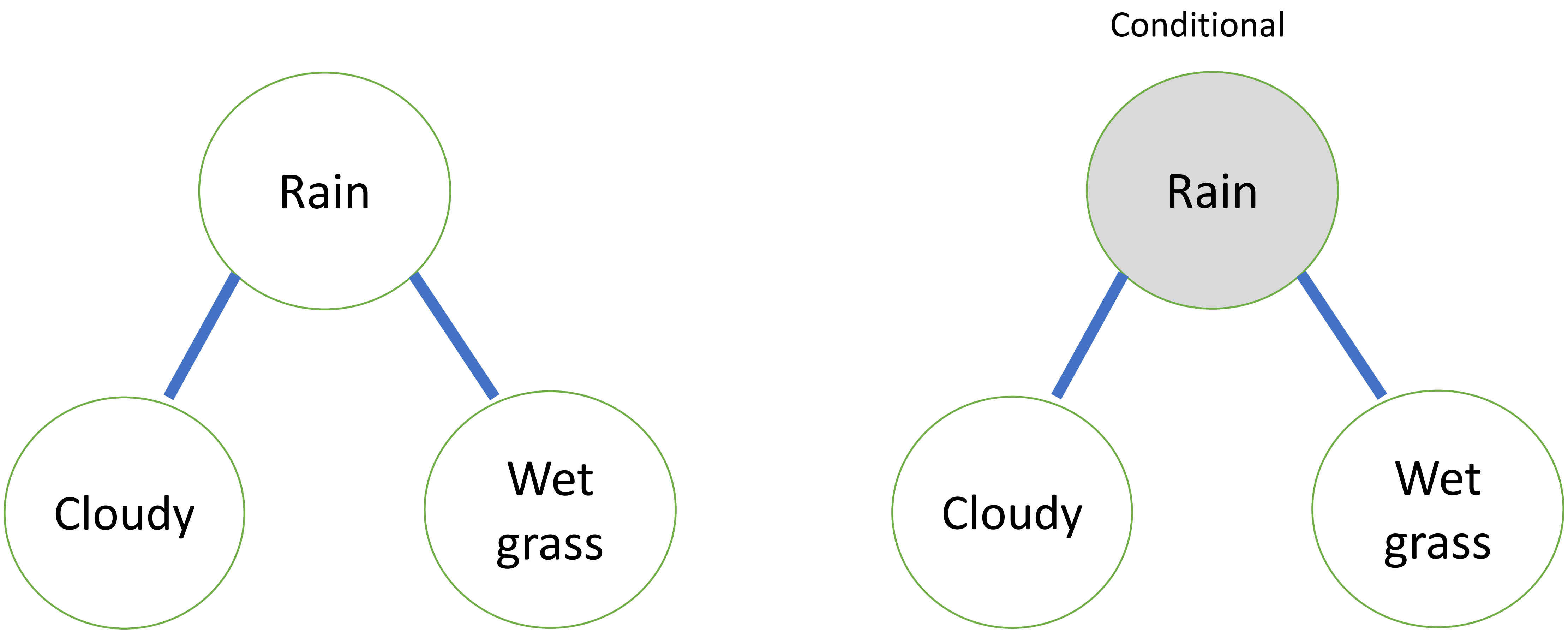
$$P(A,B|\mathbf{C})=P(A|\mathbf{C})\ P(B|\mathbf{C})$$

Knowing C makes A and B **irrelevant to each other**.

No edge between A and B in the graph.



Conditional (in)dependency

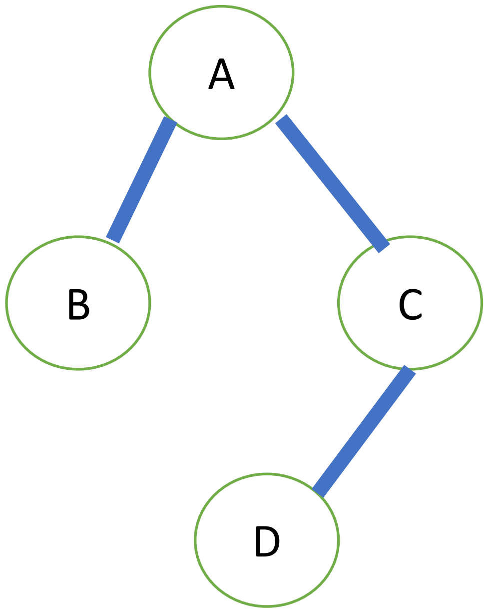


Once you know the “explainer” (rain), the other variables (cloudy & wet grass) don’t inform each other anymore.

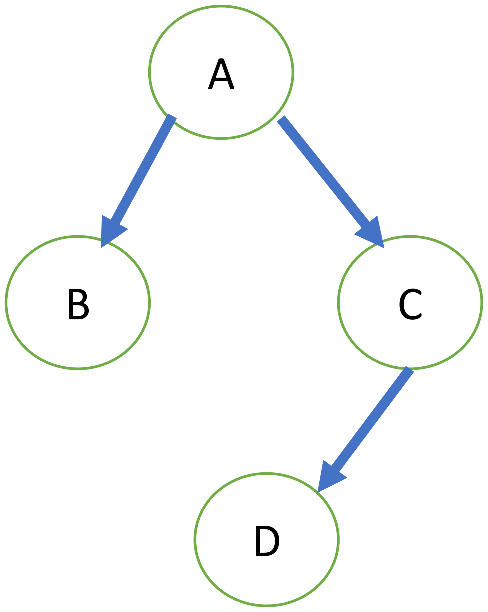
Edge and lack of edge

- Edges show **dependency.**
- **Missing edges** show **independency.**
- **But conditional.**

Undirected



Directed



Types of Graphical models:

- **Undirected Graphical Models or Markov Random Fields (MRFs) or Markov graph** (based on an undirected graph).

- **Directed Graphical Models or Bayesian Networks** (based on the directed graphs).

- **Combined**: chain graph

They differ in the set of independences.

What are <u>Undirected Graphical</u> Models or Markov Random Fields?

- Edges do not possess any form of orientation. They only show dependencies between nodes.
- Use undirected graphs where the **relationships between two variables, <u>do not</u> have a clear directionality**.
- The edge of (a,b) is identical to edge (b,a).

a —— **b**

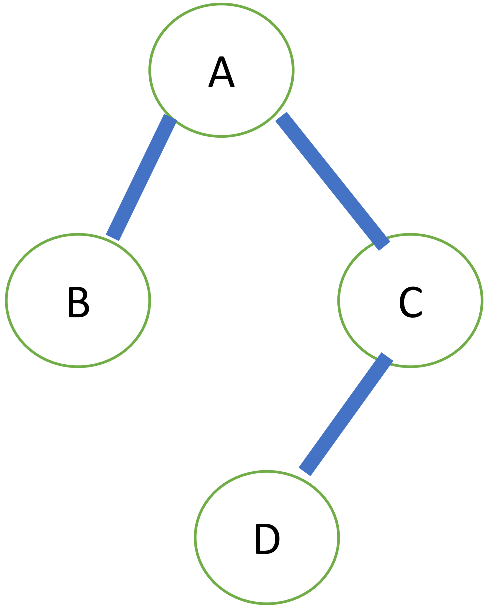
Conditional independency in a Markov Graph G

- The ***absence of an edge*** implies that the corresponding random variables are conditionally independent given the variables at the other vertices:

- **No edge joining X and Y** $\Longleftrightarrow X \perp Y \mid rest$

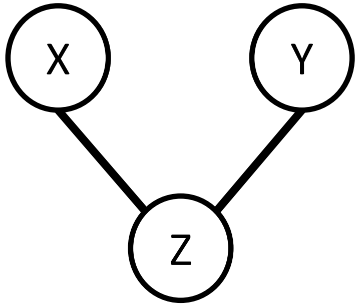
  (**pairwise Markov independencies of G**)

- "rest" refers to all the other vertices in the graph.

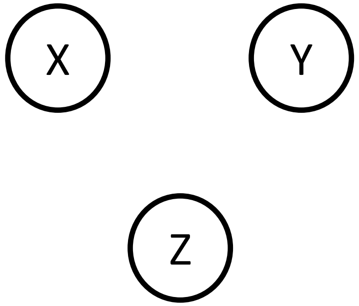


Undirected Graphical Models

- The relationships that can be described with undirected graphs.

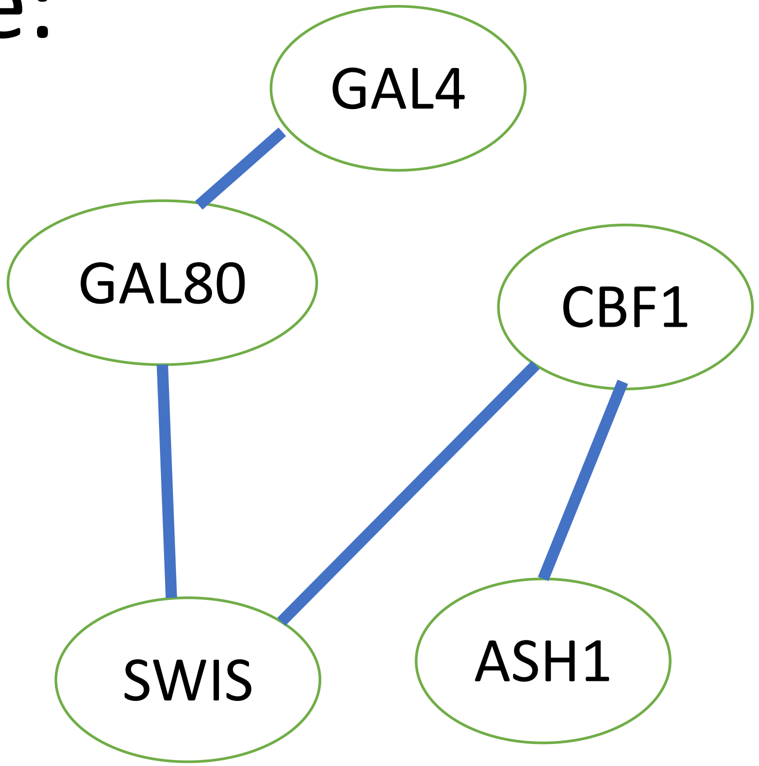


$x \perp\!\!\!\perp y \mid z$

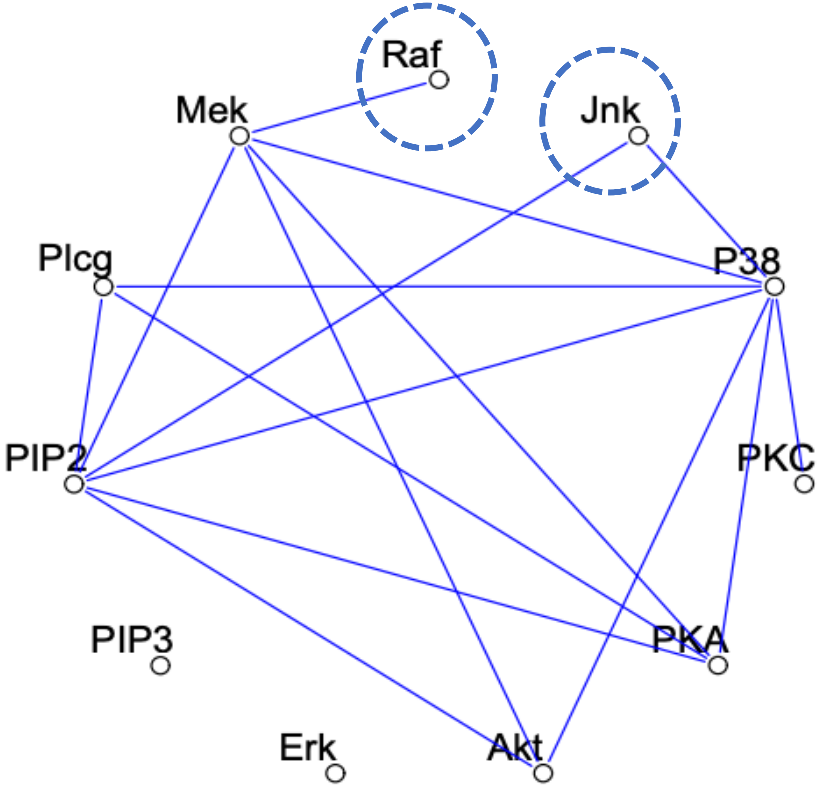


$x \perp\!\!\!\perp y$

Example:



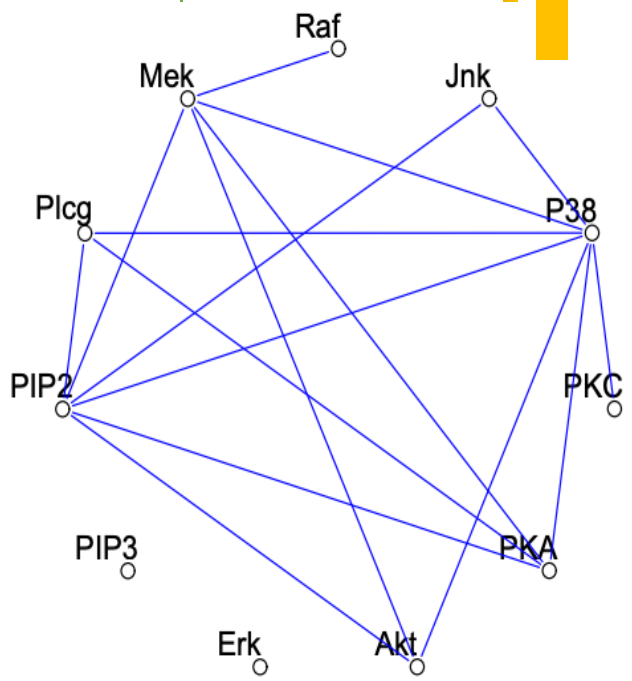
Estimated from yeast gene expression data



Estimated from a flow cytometry dataset. The network structure was estimated using the graphical lasso which will be discussed later, **Sachs et al. (2003)**.

What can we do with graphical models

- **<u>Estimation of the structure of the graph</u>**
- Estimation of the edge parameters from data
- Computation of marginal vertex probabilities and expectations.
- **Ask "what-if" or causal questions** (e.g., effect of interventions in a DAG)
- …



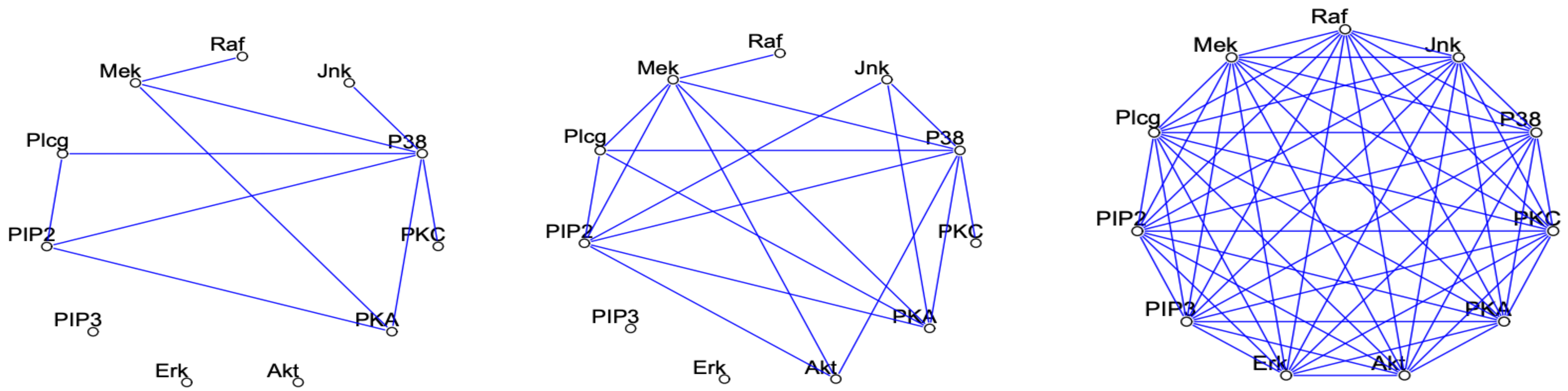
Hastie, T., Tibshirani, R., & Friedman, J. H. (2009)

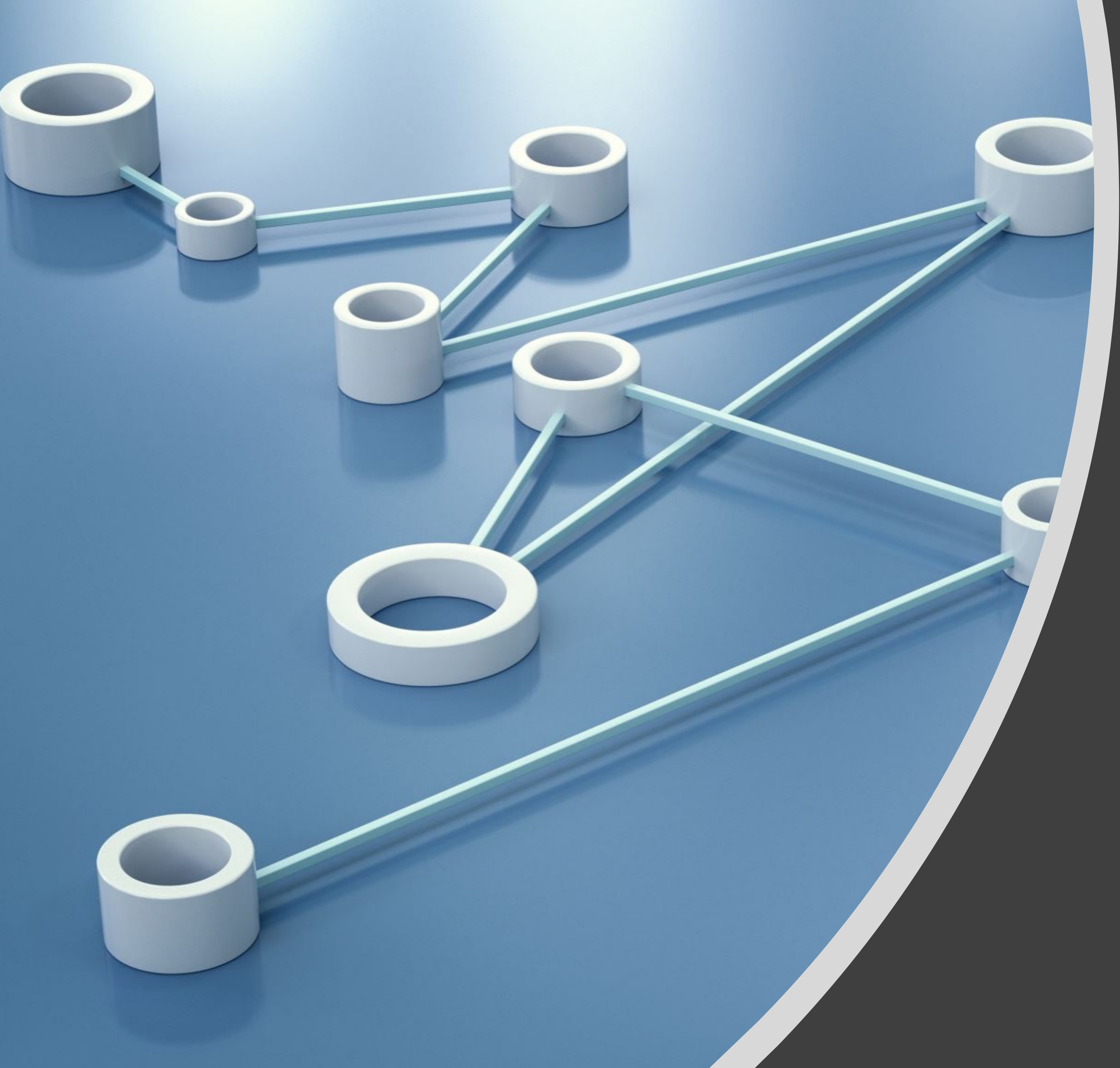
- **Choosing the structure of the graph: What kind of graph should we use?**

- Should the graph be **sparse** **or** **densely connected**?

Remark: Choosing the structure of the graph: Sparse graph

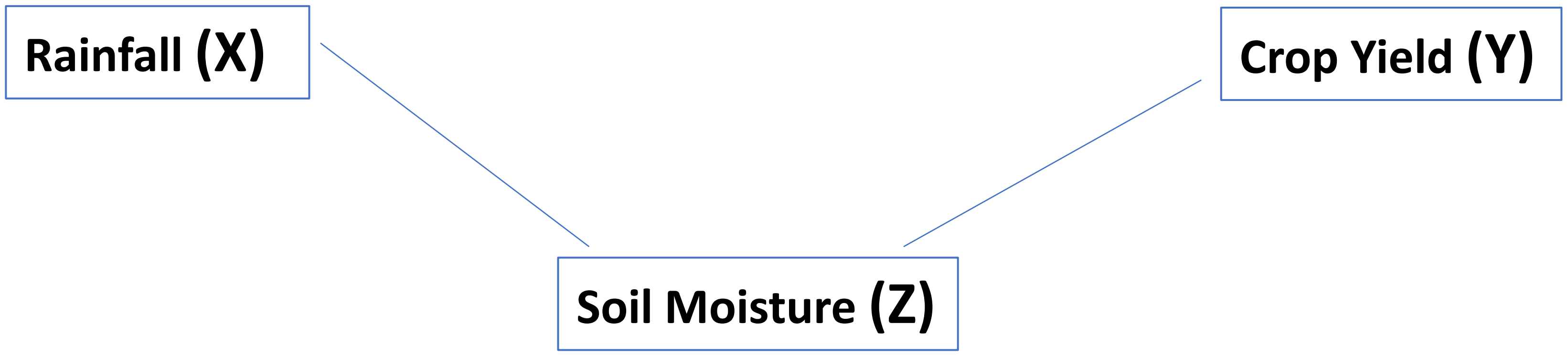
- Sparse graphs have relatively few edges.
- They are convenient for interpretation.
- We need to determine the graph’s sparsity or let the data guide us.





Some terminologies useful for Network interpretation

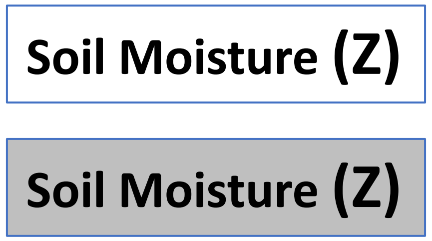
Remark



Variables **X,Y** are **dependent** if they are connected by a path of **<u>unobserved</u>** variables.

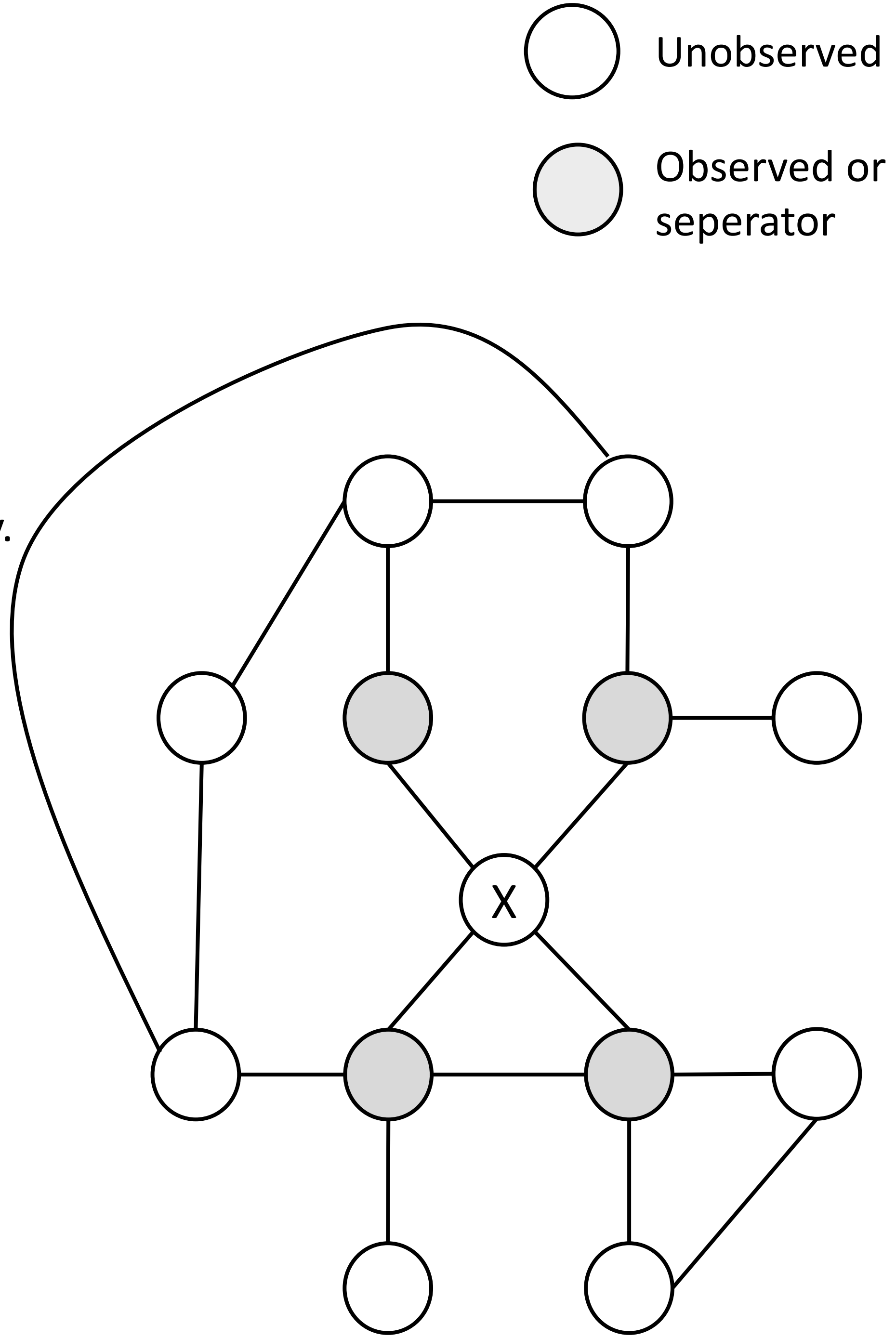
<u>Unobserved ~= Unknown ~= Non-gray nodes</u>

<u>Observed ~= Known ~= Given ~= Gray nodes</u>



Marckov blanket/ Seperators

- If x's **neighbours** are all observed, x is conditionally independent of all the others (they influence x only via its neighbours). This is local Markov property.
- In the undirected case the **Markov blanket** equals a node's neighbourhood.
- Useful for variables selection (impose sparsity).
- This is **crucial** because it explains **how graphical models reduce complexity** and enable efficient **inference and learning**.
- Instead of using all the variables in the graph to compute something about *x*, **you may only need its neighbours**. This **dramatically reduces computation** in large graphs.
- The neighbours of *x* directly **explain or influence** *x*.
  You know where the influence comes from, it's **localized**, not scattered.

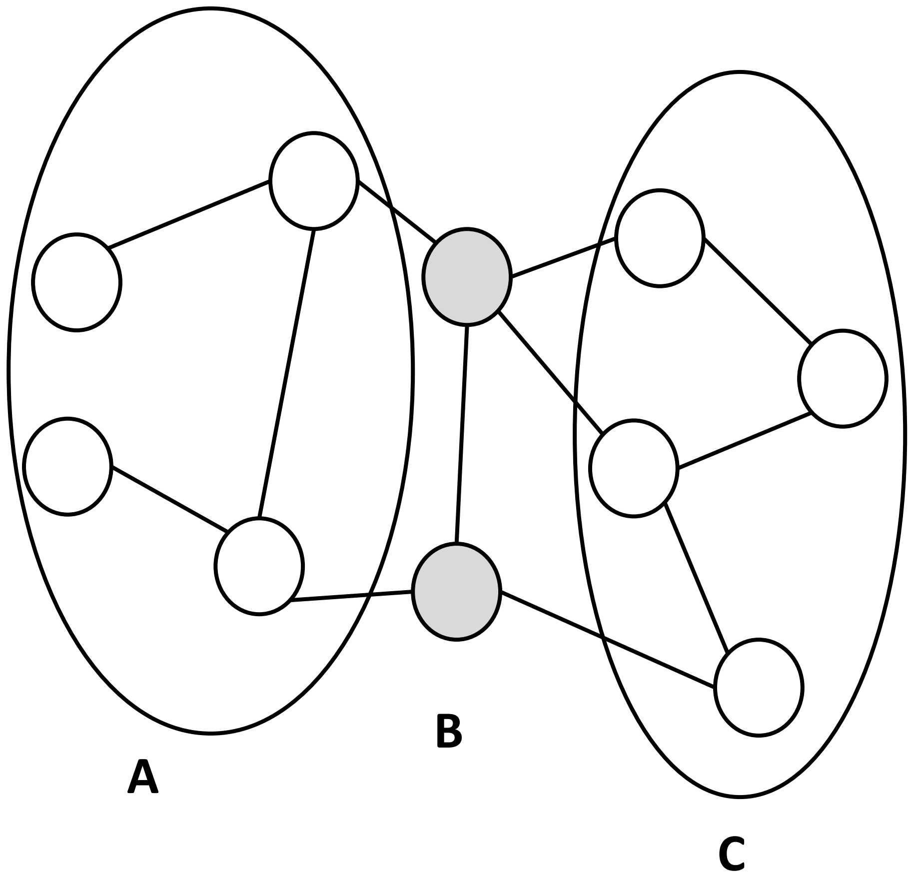


Separators and global Markov properties of G

- **B** is said to **separate A** and **C** if every path between A and C **intersects** a node in B, $A \perp C|\mathbf{B}$.

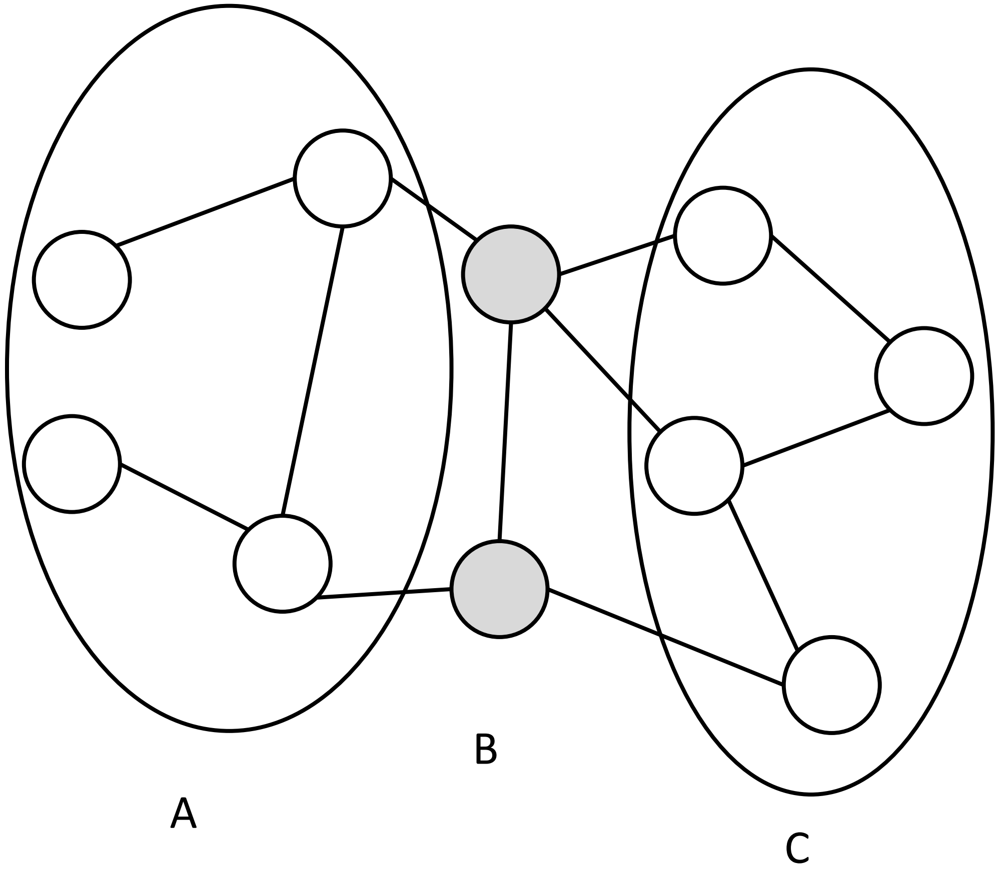
- Separators break the graph into conditionally independent pieces.

- This is known as the **global Markov properties of G.**
- Think of B as a **"firewall"**: if you observe **B**, then **A** and **C** can't "talk to each other" anymore; they're independent.



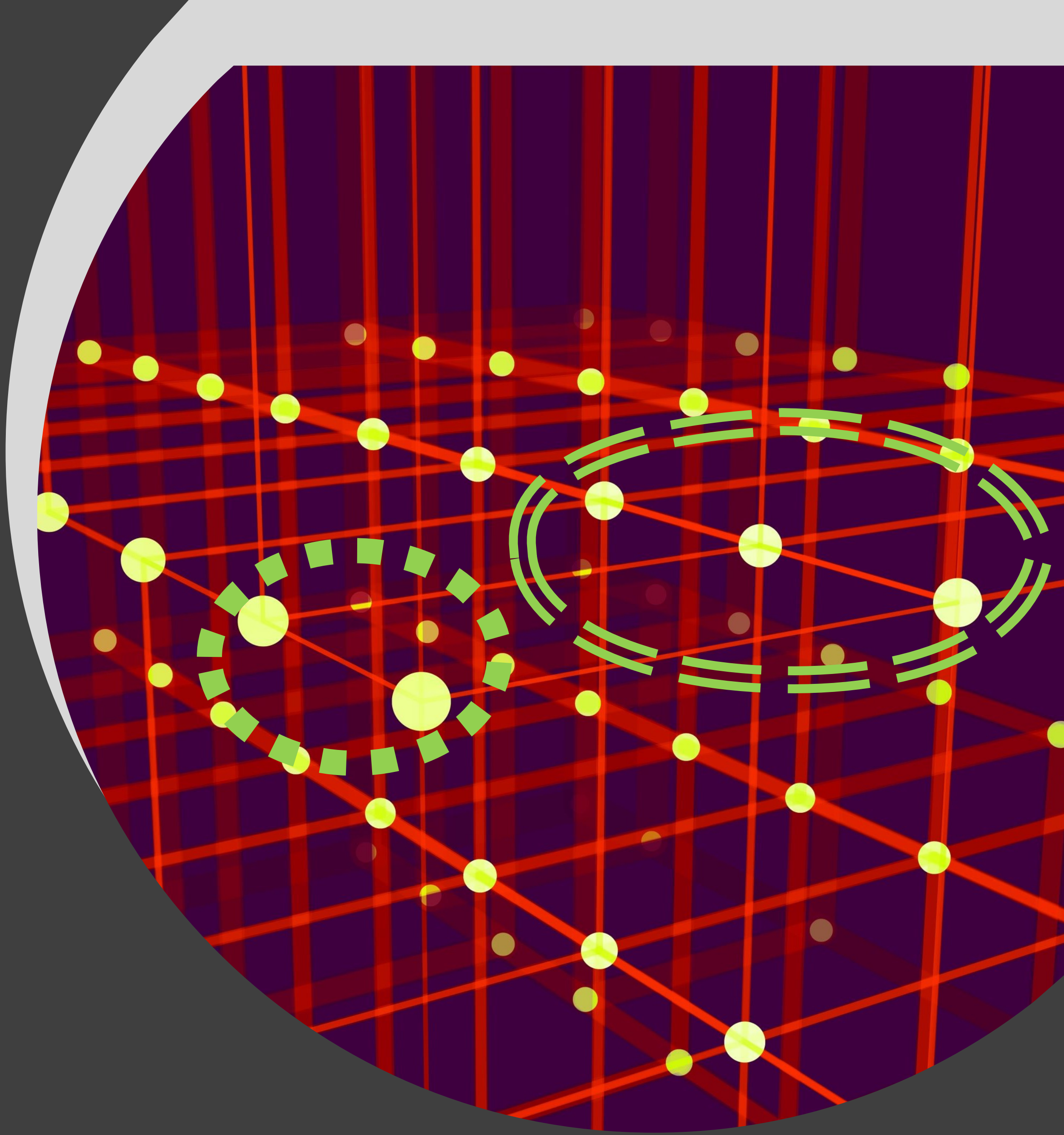
Remark:

- The global Markov property allows us to **decompose** graphs into **smaller more manageable pieces** and thus leads to essential simplifications in computation and interpretation.

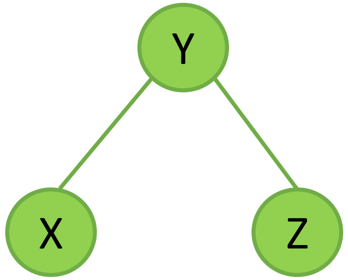


Clique

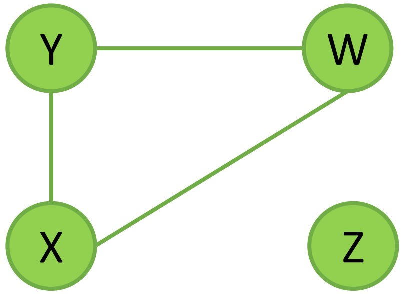
- We separate the graph into cliques (small pieces, including sufficient information).
- A clique is a **complete** subgraph- a set of vertices that are all adjacent to one another.



Examples: How many cliques can you identify in each graph?



(a)



(b)

Why Clique?

- **If we can factorize an MRF according to the cliques**, (next slide):

- This makes it **much easier to compute**, **learn** from data, and **estimate the structure** of the whole graph.

- We don’t have to model everything at once, just model each **clique**, then combine them.

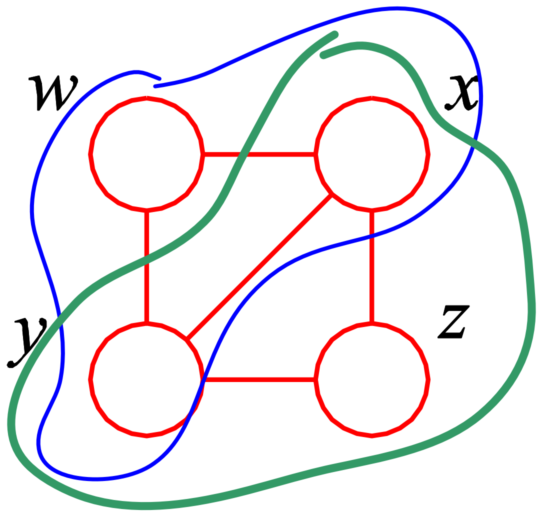
- Many of the **methods for estimation and computation on graphs** first decompose the graph into its **cliques**.

Clique potential

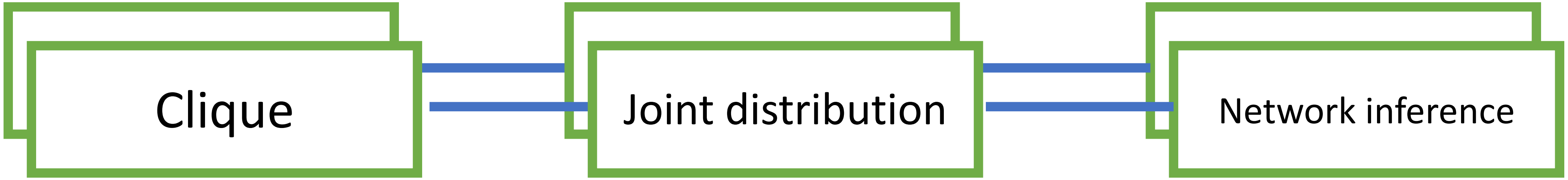
- Provided $\boldsymbol{p}(.)$ **> 0** then joint distribution is product of **non-negative functions** over the *cliques* of the graph and $\boldsymbol{\psi_C}(.)$ is the *clique potential*, a function of only the values of the clique members in C

$$p(w,x,y,z) = \frac{1}{Z}\psi_A(w,x,y)\psi_B(x,y,z)$$

Z is a normalization constant.



Remark :Why are cliques important?



Cliques are often used to analyse the general structure of a network

How to infer a Network Structure?

Types of Undirected Graphical models

- **Undirected Graphical Models for Discrete Variables**

  A special case of loglinear models for multiway contingency tables (Bishop et al., 1975, e.g.), Ising Model.

- **Undirected Graphical Models for Continuous Variables**

  - Markov networks where all the variables are continuous.
  - The Gaussian distribution is almost always used for such graphical models, because of its convenient analytical properties.

Undirected Markov networks with all discrete variables

1. Called Ising models or Boltzmann machines

2. The “nodes” are binary-valued.

3. The Ising model is used to model the joint effects of pairwise interactions.
4. Fit an L1-penalized logistic regression model to each node as a function of the other nodes.

5. Then optimization.

6. And finally, network structure estimation.

https://jwmi.github.io/ASM/Murphy%20chapter%2019.pdf

Undirected Graphical Models for Continuous Variables

- The observations have **a multivariate Gaussian** distribution with mean **μ** and covariance matrix **Σ**.

$$X \sim MN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- Gaussian distribution represents second-order relationships (pairwise relationships between variables) and automatically encodes a pairwise Markov graph.

- All conditional distributions are also Gaussian.

Undirected Graphical Models for Continuous Variables

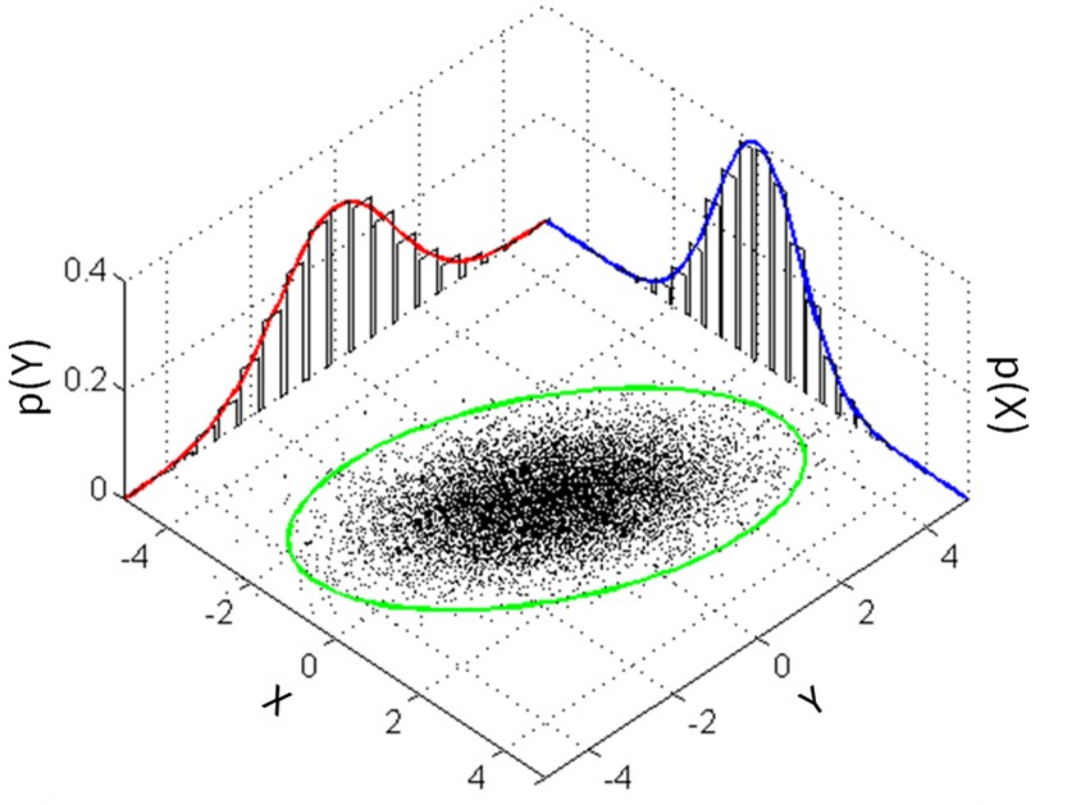
- The **Gaussian distribution** is almost always used for such graphical models, because of its convenient analytical properties.

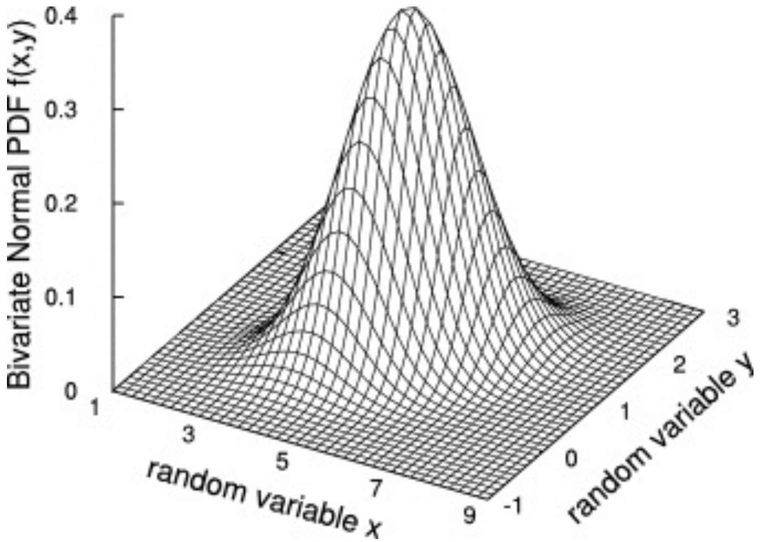
$$f_{\mathbf{X}}(x_1, \ldots, x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}$$

$\boldsymbol{\mu}$: **mean parameter**

$\boldsymbol{\Sigma}$: **covariance matrix**

All the pairwise relationships between variables can be extracted from covariance matrix: $\boldsymbol{\Sigma}$.





Remark: Covariance

- Covariance indicates **the relationship of two variables whenever one variable changes**.

- If an increase in one variable results in an increase in the other variable, both variables are said to have a positive covariance.

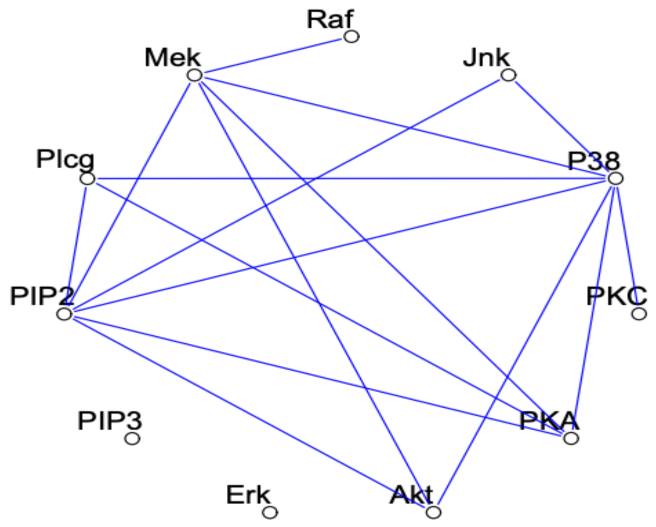
Covariance matrix Σ

- Encodes all information how variables relate to one another.

$$\begin{array}{c c} & \begin{array}{cc} x & y \end{array} \\ \begin{array}{c} x \\ y \end{array} & \begin{bmatrix} var(x) & cov(x,y) \\ cov(x,y) & var(y) \end{bmatrix} \end{array}$$

$$\begin{array}{c c} & \begin{array}{ccc} x & y & z \end{array} \\ \begin{array}{c} x \\ y \\ z \end{array} & \begin{bmatrix} var(x) & cov(x,y) & cov(x,z) \\ cov(x,y) & var(y) & cov(y,z) \\ cov(x,z) & cov(y,z) & var(z) \end{bmatrix} \end{array}$$

A flow cytometry dataset, Protein-protein



Raf	Mek	Plcg	PIP2	PIP3	Erk	Akt	PKA	PKC	P38	Jnk
26.4	13.2	8.82	18.3	58.8	6.61	17	414	17	44.9	40
35.9	16.5	12.3	16.8	8.13	18.6	32.5	352	3.37	16.5	61.5
59.4	44.1	14.6	10.2	13	14.9	32.5	403	11.4	31.9	19.5
73	82.8	23.1	13.5	1.29	5.83	11.8	528	13.7	28.6	23.1
33.7	19.8	5.19	9.73	24.8	21.1	46.1	305	4.66	25.7	81.3
18.8	3.75	17.6	22.1	10.9	11.9	25.7	610	13.7	49.1	57.8
44.9	36.5	10.4	132	16.3	8.66	17.9	835	15	35.9	18.1
47.4	15	14.6	30.5	17.5	20.2	45.3	466	6.44	24.4	20
104	61.5	10.6	21.1	41.8	11.5	23.5	445	29.2	61	25.3

Covariance matrix Σ

$$
\begin{array}{c|ccc}
 & x & y & z \\
\hline
x & & & \\
y & & & \\
z & & &
\end{array}
\quad
\begin{array}{c}
\\
\end{array}
$$

$$
\begin{array}{cc}
 & \begin{array}{ccc} \quad x \quad & \quad y \quad & \quad z \quad \end{array} \\
\begin{array}{c} x \\ y \\ z \end{array} & \begin{bmatrix} var(x) & cov(x,y) & cov(x,z) \\ cov(x,y) & var(y) & cov(y,z) \\ cov(x,z) & cov(y,z) & var(z) \end{bmatrix}
\end{array}
$$

```
> round(cov(Protein_data))
      Raf  Mek Plcg  PIP2 PIP3   Erk   Akt    PKA  PKC P38   Jnk
Raf  1751  911   15   248   55   -72  -144    -89  -24 -51     7
Mek   911  753    3   218   52   -36   -76   -161   -7 -14   -53
Plcg   15    3  215   127   68     3     4    -41   -1  -1    45
PIP2  248  218  127  8745  875   -74  -147  -1631  -16 -19    49
PIP3   55   52   68   875 1170   -95  -148   -726   12  33   103
Erk   -72  -36    3   -74  -95  8081 11291  14958    2  -7   -89
Akt  -144  -76    4  -147 -148 11291 16043  23591    8  -5  -121
PKA   -89 -161  -41 -1631 -726 14958 23591 183013  113 248 -1058
PKC   -24   -7   -1   -16   12     2     8    113  134 166  -103
P38   -51  -14   -1   -19   33    -7    -5    248  166 377   -20
Jnk     7  -53   45    49  103   -89  -121  -1058 -103 -20  1857
>
```

Inverse variance-covariance matrix, Σ^{-1} : ***precision matrix***

- This is a treasure.
- Called a **Gaussian graphical model (GGM).**
- Encodes **an undirected network.**
- Shows the **relationships** between variables (conditional dependencies).

$$\begin{array}{cc} & \begin{array}{cc} \mathrm{x} & \mathrm{y} \end{array} \\ \begin{array}{c} \mathrm{x} \\ \mathrm{y} \end{array} & \begin{bmatrix} var(x) & cov(x,y) \\ cov(x,y) & var(y) \end{bmatrix}^{\mathbf{-1}} \end{array} \qquad \begin{array}{cc} & \begin{array}{ccc} \mathrm{x} & \mathrm{y} & \mathrm{z} \end{array} \\ \begin{array}{c} \mathrm{x} \\ \mathrm{y} \\ \mathrm{z} \end{array} & \begin{bmatrix} var(x) & cov(x,y) & cov(x,z) \\ cov(x,y) & var(y) & cov(y,z) \\ cov(x,z) & cov(y,z) & var(z) \end{bmatrix}^{\mathbf{-1}} \end{array}$$

Inverse variance-covariance matrix (precision matrix): $\boldsymbol{\Sigma}^{-1}$

- Contains information about the **partial covariances** between the variables.
- If the ij^{th} component of $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$ is zero, then variables *i* and *j* are conditionally independent, given the other variables (no edge).
- Thus, Θ captures all the second-order (pairwise) information.

Precision matrix: Σ^{-1}

solve(…)

	Raf	Mek	Plcg	PIP2	PIP3	Erk	Akt	PKA	PKC	P38	Jnk
Raf	0.001560	-0.001886	-0.000081	0.000003	0.000014	-0.000142	0.000111	-4.0e-06	-0.000103	0.000184	-0.000065
Mek	-0.001886	0.003629	0.000102	-0.000033	-0.000059	-0.000016	0.000007	3.0e-06	0.000208	-0.000206	0.000124
Plcg	-0.000081	0.000102	0.004772	-0.000046	-0.000234	0.000015	-0.000015	0.0e+00	-0.000122	0.000070	-0.000105
PIP2	0.000003	-0.000033	-0.000046	0.000125	-0.000090	-0.000008	0.000005	1.0e-06	0.000015	0.000006	0.000003
PIP3	0.000014	-0.000059	-0.000234	-0.000090	0.000946	-0.000017	0.000016	2.0e-06	-0.000059	-0.000065	-0.000049
Erk	-0.000142	-0.000016	0.000015	-0.000008	-0.000017	0.008694	-0.006263	9.7e-05	0.000516	-0.000237	0.000094
Akt	0.000111	0.000007	-0.000015	0.000005	0.000016	-0.006263	0.004588	-8.0e-05	-0.000386	0.000185	-0.000069
PKA	-0.000004	0.000003	0.000000	0.000001	0.000002	0.000097	-0.000080	8.0e-06	0.000012	-0.000010	0.000005
PKC	-0.000103	0.000208	-0.000122	0.000015	-0.000059	0.000516	-0.000386	1.2e-05	0.017758	-0.007771	0.000925
P38	0.000184	-0.000206	0.000070	0.000006	-0.000065	-0.000237	0.000185	-1.0e-05	-0.007771	0.006080	-0.000379
Jnk	-0.000065	0.000124	-0.000105	0.000003	-0.000049	0.000094	-0.000069	5.0e-06	0.000925	-0.000379	0.000598

Partial correlation coefficients

- The **precision matrix** can be **standardized** to ***partial correlation coefficients***. **WHY?**

$$\mathrm{Cor}\left(Y_i, Y_j \mid \boldsymbol{Y}^{-(i,j)}\right) = -\frac{\kappa_{ij}}{\sqrt{\kappa_{ii}}\sqrt{\kappa_{jj}}}$$

- These are used to draw a network.
- GGM is also often called a ***partial correlation network***

https://stats.stackexchange.com/questions/140080/why-does-inversion-of-a-covariance-matrix-yield-partial-correlations-between-ran

partial correlation network, standardized

Pcor(data)

	Raf	Mek	Plcg	PIP2	PIP3	Erk	Akt	PKA	PKC	P38	Jnk
Raf	1.000000000	0.792857559	0.0296930157	-0.007029587	-0.011502333	0.038440491	-0.041308209	0.0367288040	0.019575831	-0.059685607	0.06695341
Mek	0.792857559	1.000000000	-0.0245441925	0.048772186	0.031814663	0.002778468	-0.001611057	-0.0181521456	-0.025877776	0.043816669	-0.08397654
Plcg	0.029693016	-0.024544192	1.0000000000	0.059422515	0.110235462	-0.002271585	0.003188654	0.0006681285	0.013305696	-0.013006298	0.06214019
PIP2	-0.007029587	0.048772186	0.0594225153	1.000000000	0.260917332	0.007382425	-0.005952502	-0.0244760835	-0.009921673	-0.007377868	-0.01126633
PIP3	-0.011502333	0.031814663	0.1102354623	0.260917332	1.000000000	0.005780849	-0.007779322	-0.0227838867	0.014416045	0.027224087	0.06466773
Erk	0.038440491	0.002778468	-0.0022715847	0.007382425	0.005780849	1.000000000	0.991569139	-0.3708765874	-0.041549776	0.032569018	-0.04104652
Akt	-0.041308209	-0.001611057	0.0031886538	-0.005952502	-0.007779322	0.991569139	1.000000000	0.4200682087	0.042762841	-0.034993207	0.04164026
PKA	0.036728804	-0.018152146	0.0006681285	-0.024476083	-0.022783887	-0.370876587	0.420068209	1.0000000000	-0.033324708	0.047368197	-0.06595145
PKC	0.019575831	-0.025877776	0.0133056963	-0.009921673	0.014416045	-0.041549776	0.042762841	-0.0333247077	1.000000000	0.747924935	-0.28409520
P38	-0.059685607	0.043816669	-0.0130062982	-0.007377868	0.027224087	0.032569018	-0.034993207	0.0473681967	0.747924935	1.000000000	0.19858412
Jnk	0.066953405	-0.083976539	0.0621401887	-0.011266330	0.064667726	-0.041046519	0.041640264	-0.0659514543	-0.284095198	0.198584117	1.00000000

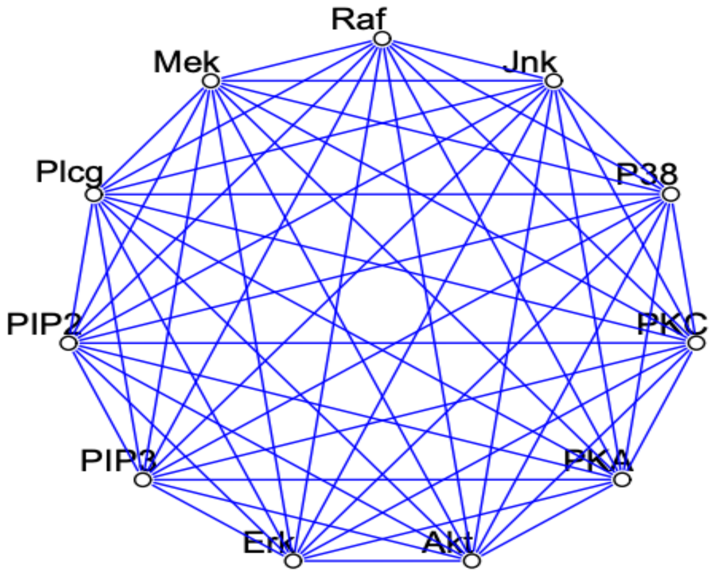
What do you think? Can you interpret that from a network perspective?

- **Zeros** show the **absence** of edges.
- **Non-zeros** show the **presence** of edges.

partial correlation network, standardized

Pcor(data)

	Raf	Mek	Plcg	PIP2	PIP3	Erk	Akt	PKA	PKC	P38	Jnk
Raf	1.000000000	0.792857559	0.0296930157	-0.007029587	-0.011502333	0.038440491	-0.041308209	0.0367288040	0.019575831	-0.059685607	0.06695341
Mek	0.792857559	1.000000000	-0.0245441925	0.048772186	0.031814663	0.002778468	-0.001611057	-0.0181521456	-0.025877776	0.043816669	-0.08397654
Plcg	0.029693016	-0.024544192	1.0000000000	0.059422515	0.110235462	-0.002271585	0.003188654	0.0006681285	0.013305696	-0.013006298	0.06214019
PIP2	-0.007029587	0.048772186	0.0594225153	1.000000000	0.260917332	0.007382425	-0.005952502	-0.0244760835	-0.009921673	-0.007377868	-0.01126633
PIP3	-0.011502333	0.031814663	0.1102354623	0.260917332	1.000000000	0.005780849	-0.007779322	-0.0227838867	0.014416045	0.027224087	0.06466773
Erk	0.038440491	0.002778468	-0.0022715847	0.007382425	0.005780849	1.000000000	0.991569139	-0.3708765874	-0.041549776	0.032569018	-0.04104652
Akt	-0.041308209	-0.001611057	0.0031886538	-0.005952502	-0.007779322	0.991569139	1.000000000	0.4200682087	0.042762841	-0.034993207	0.04164026
PKA	0.036728804	-0.018152146	0.0006681285	-0.024476083	-0.022783887	-0.370876587	0.420068209	1.0000000000	-0.033324708	0.047368197	-0.06595145
PKC	0.019575831	-0.025877776	0.0133056963	-0.009921673	0.014416045	-0.041549776	0.042762841	-0.0333247077	1.000000000	0.747924935	-0.28409520
P38	-0.059685607	0.043816669	-0.0130062982	-0.007377868	0.027224087	0.032569018	-0.034993207	0.0473681967	0.747924935	1.000000000	0.19858412
Jnk	0.066953405	-0.083976539	0.0621401887	-0.011266330	0.064667726	-0.041046519	0.041640264	-0.0659514543	-0.284095198	0.198584117	1.00000000

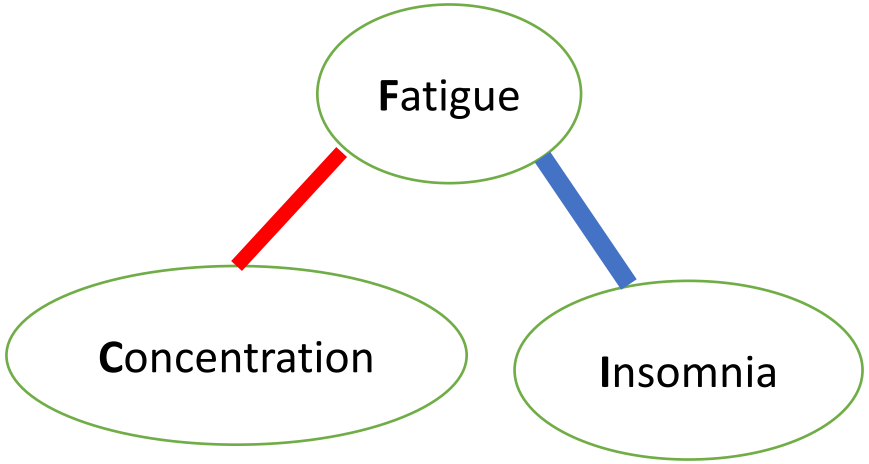


Zero mean there is not an edge.
Non-zero means there is an edge

Is that really helpful? Can we infer a network directly from **partial correlation coefficient**? **No**

Why? **Really difficult for interpretation**

Partial correlation
(standardization)



$$\text{Partial correlations:}\ \begin{array}{c c} & \begin{array}{ccc} \textbf{C} & \textbf{F} & \textbf{I} \end{array} \\ \begin{array}{c} \textbf{C} \\ \textbf{F} \\ \textbf{I} \end{array} & \begin{bmatrix} 1 & -0.25 & 0 \\ -0.25 & 1 & 0.30 \\ 0 & 0.30 & 1 \end{bmatrix} \end{array}$$

So, what should we do?

Let's review what we have done so far…

- We estimated **Σ** **(S: the** sample covariance matrix**)**.
- We approximated the precision matrix Σ^{-1} (by $\boldsymbol{S}^{-1}$).
- Standardize $\boldsymbol{S}^{-1}$

In real-life datasets, especially those with many variables:

> ➢**Standardized** $\boldsymbol{S}^{-1}$ contains few (if any) exact zeros due to noise.
> ➢Small values in **Standardized** $\boldsymbol{S}^{-1}$ might not mean *real connections.*

What should we do?

Shrink or remove small coefficients to recover a clearer graph
→ This leads to **sparser, easier-to-interpret models**

How can we shrink small values?

- We want the model to **learn the important ones from the data.**

- It can be done by **L1 (lasso)** regularization (**Graphical lasso**),

- The final advantage: We get a **sparse graph** that is **easier to interpret.**

Graphical Lasso, a <u>tool</u> for estimating the network structure

R package is glasso

Friedman et al. (2008)

- The idea is to shrink **low** correlations to **0** such that they disappear from the graph.

- Fit <u>**a lasso regression**</u> between variables.

- This procedure consistently estimates the set of **nonzero elements of** estimated $\boldsymbol{\Sigma}^{-1}$ and <u>**shrink low correlations to 0**</u>.

The graphical lasso

- The R package "**glasso**" is popular, fast, and allows one to efficiently build a path of models for different values of the tuning parameter.

- There are many other extensions that use **glasso** packages.

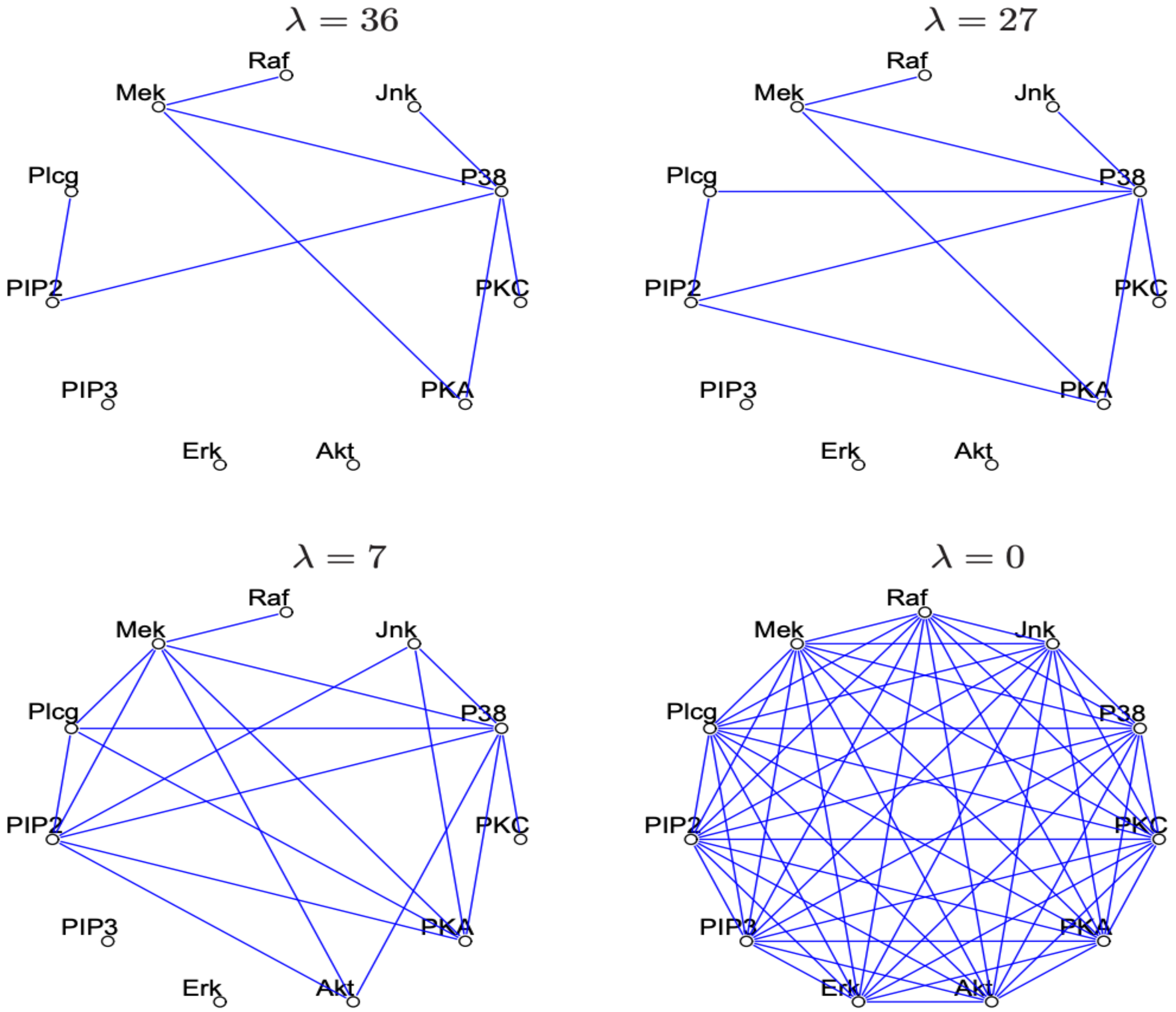
- In the practical we focus on **glasso** and **EBICglasso**.

- **Note:** graphical lasso can be also done in **Python** in scikit-learn
- https://scikit-learn.org/stable/modules/generated/sklearn.covariance.GraphicalLasso.html#sklearn.covariance.GraphicalLasso.
- You can download the .ipynb file of this from the website of the course.

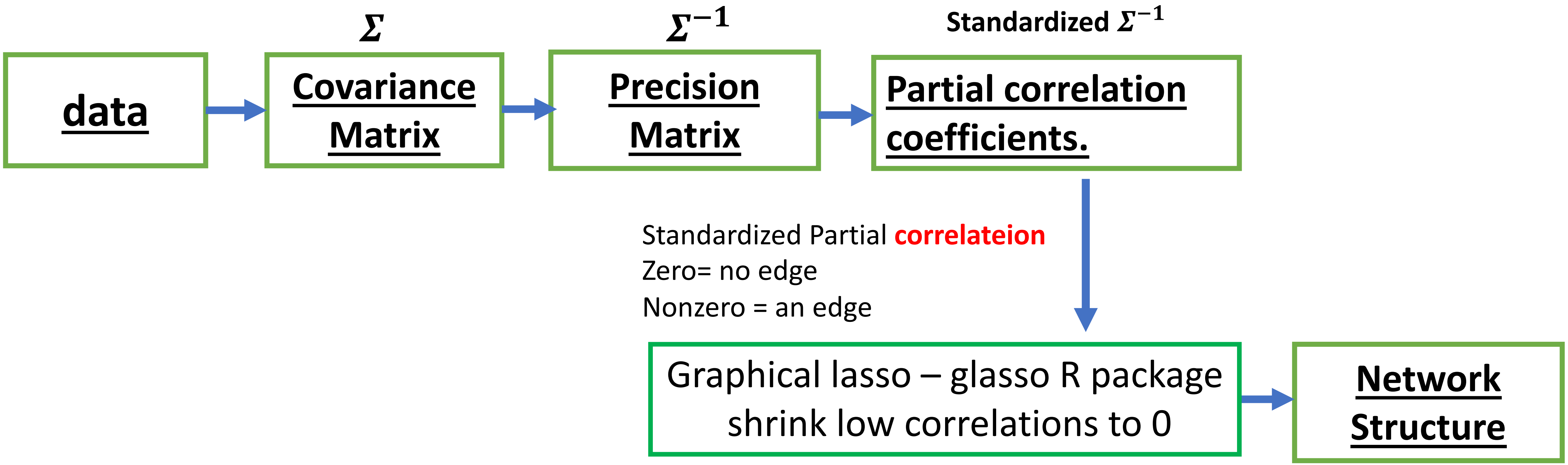
Graphical lasso - Penalty parameter λ

Friedman et al. (2008)

- Remember again the flow-cytometry data.
- The graph becomes sparser as the **penalty parameter,** λ, is increased.
- We can infer λ from the data using cross-validation.



In a nutshell: Network structure



Main advantages of MRFs:

- **MRFs works without direction** – perfect when there’s no obvious direction.
- **MRFs** succinctly express **certain dependencies** (cycle) that BNs (Bayesian Networks) **cannot** describe.
- Conditional independence is easier to read off for MRFs.
- **There is only one type of neighbors (simpler Markov blanket).**

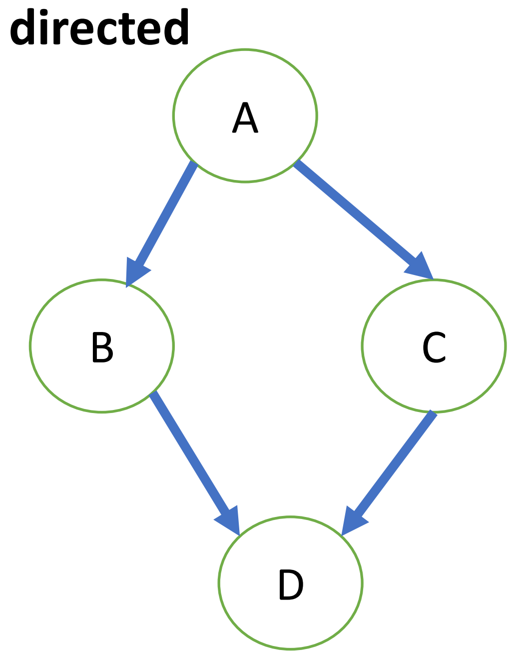
Disadvantages of MRFs:

- Difficult to interpret.
- **No causal story** at all – edges are symmetric.

In the afternoon: Directed graphs

- All the edges have a specific direction that depicts the nature and dependence of the relationship between the two vertices at the end of the edge.

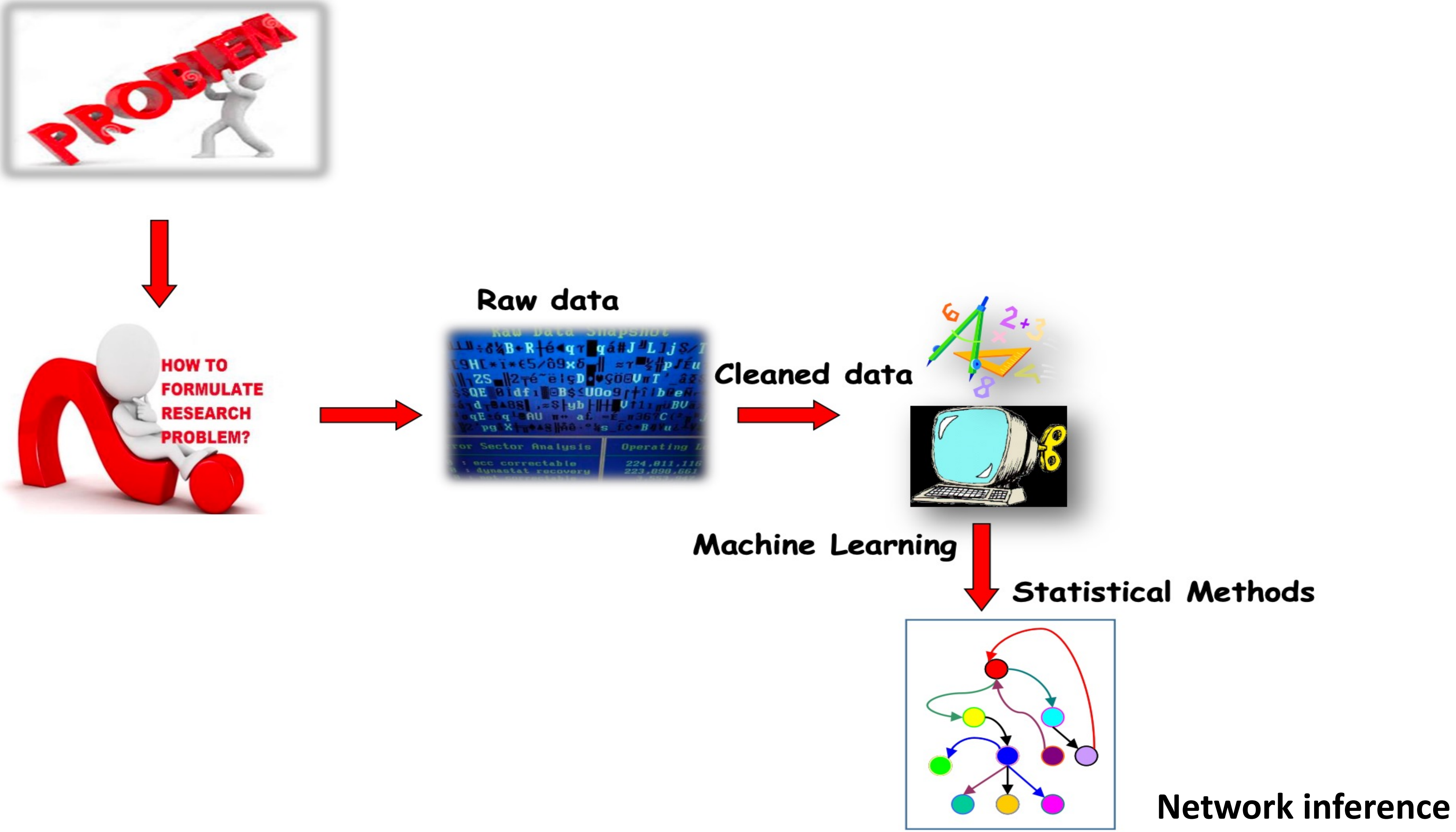
- A graph that **does not form any cycle** is called a **<u>Directed Acyclic Graph (DAG)</u>.**



Directed graphs

- Bayesian networks are a type of **Directed Acyclic Graphs (DAG)** that will be discussed in the afternoon.

In a nutshell



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

- Consider the following questions again:

✓What variables are present in your field?

✓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