

Graphical Models for Network Inference

Mahdi Shafiee Kamalabad
Utrecht University

General Learning Goals

1. Define probabilistic graphical models (PGMs).
2. Understand the definitions and roles of nodes and edges in PGMs.
3. Define Markov Random Fields (MRFs) and understand their role as undirected graphical models.
4. Explain the importance of estimating the network structure in real-world scenarios using MRFs.
5. Discuss applications of MRFs in various domains.
6. Define and differentiate between various types of dependencies between nodes.
7. Develop skills to interpret the network structure of PGMs.
8. Discuss methods to estimate network structures.

Graphical models for network Inference

- Concept of graphical models
- Types of Graphical models
 - Undirected Graphical Models or Markov random fields (MRFs)
 - Directed Graphical Models or Bayesian Networks (BNs)
- Types of Undirected Graphical models for network inference (will be discussed in the morning)
 - Continuous
 - Discrete
- Types of directed Graphical models for network inference (will be discussed in the afternoon)
 - Continuous
 - Discrete
 - Mixed

Let's first, 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?

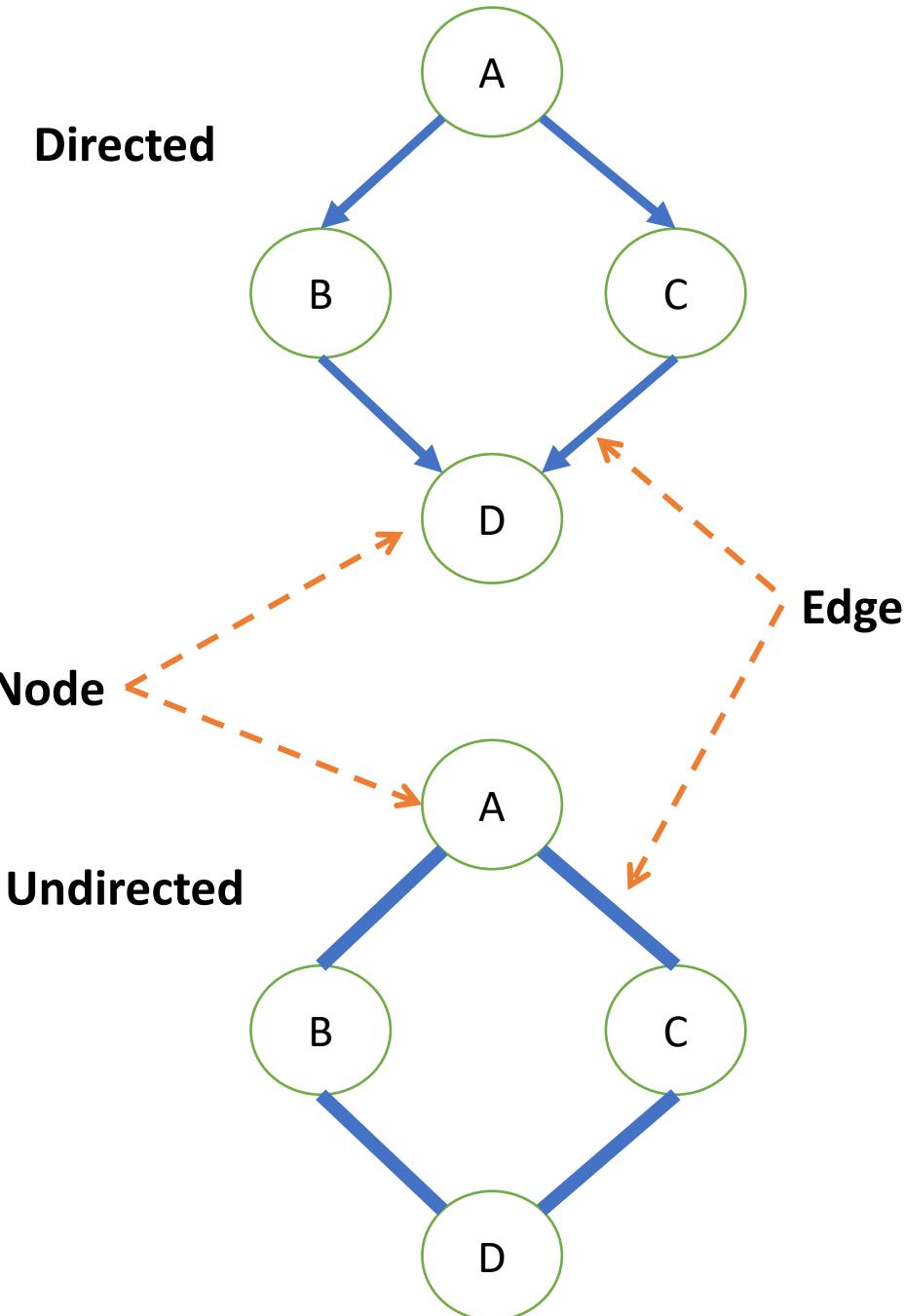
Graph

- We define a graph G by the following equation:

$$G = (V, E)$$

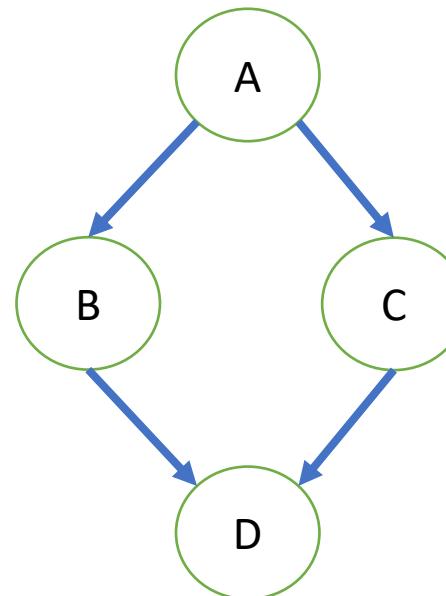
Here:

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



Edges and nodes in a graphical model

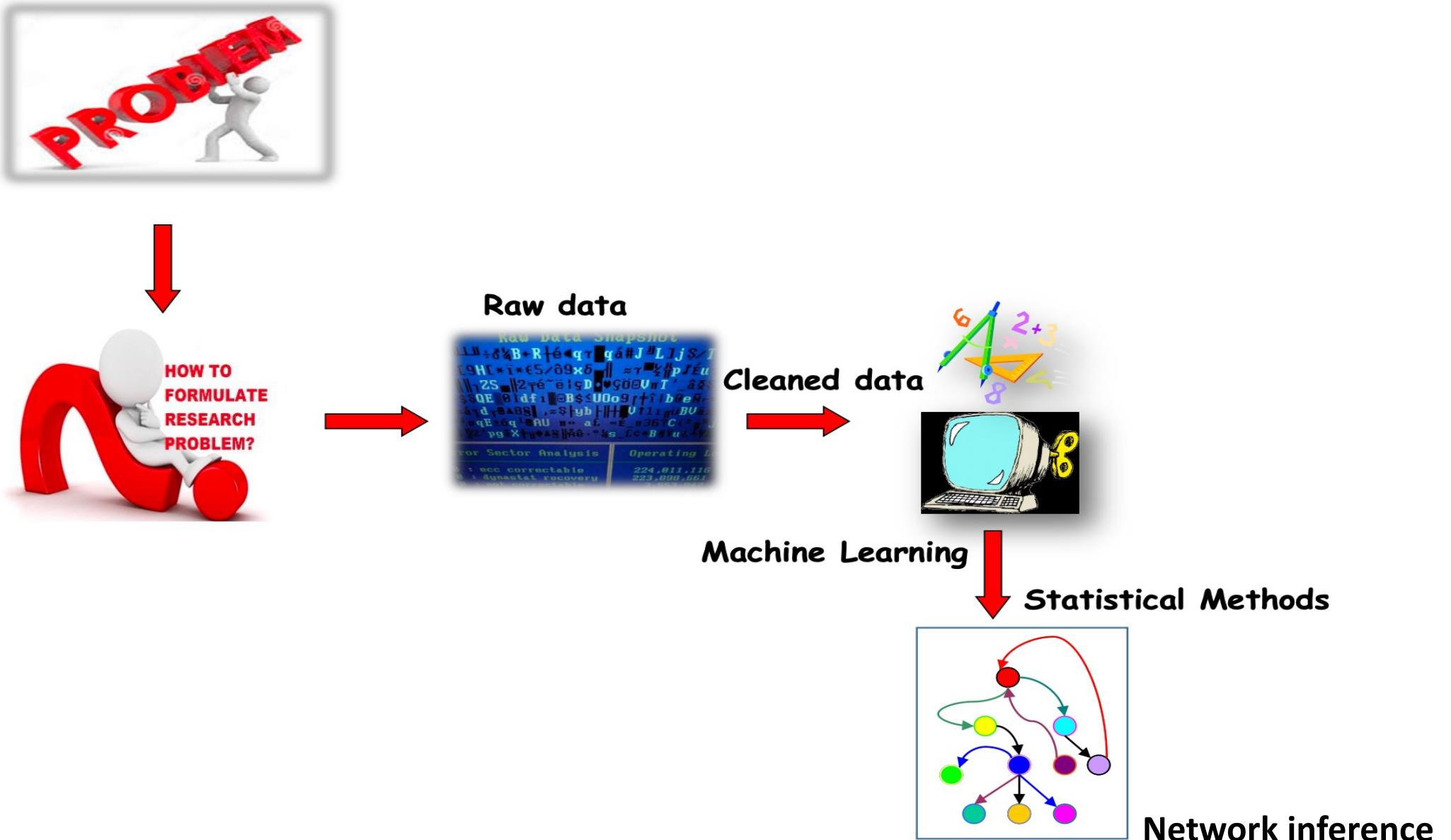
- Random variables as nodes.
- (Statistical) dependencies as edges.



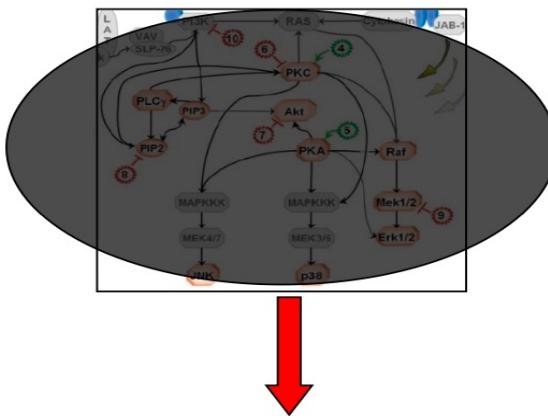
Here

Network inference = network structure estimation

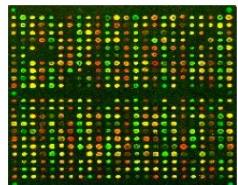
In a nutshell



possibly
completely
unknown



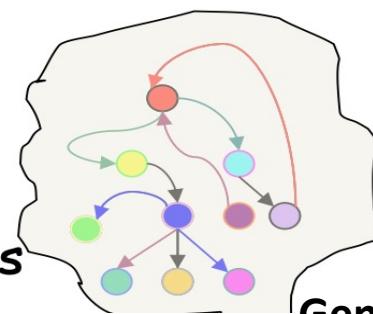
E.g.: Gene-
Microarray
experiments



data data



Machine Learning
statistical methods



Gene regularity Network

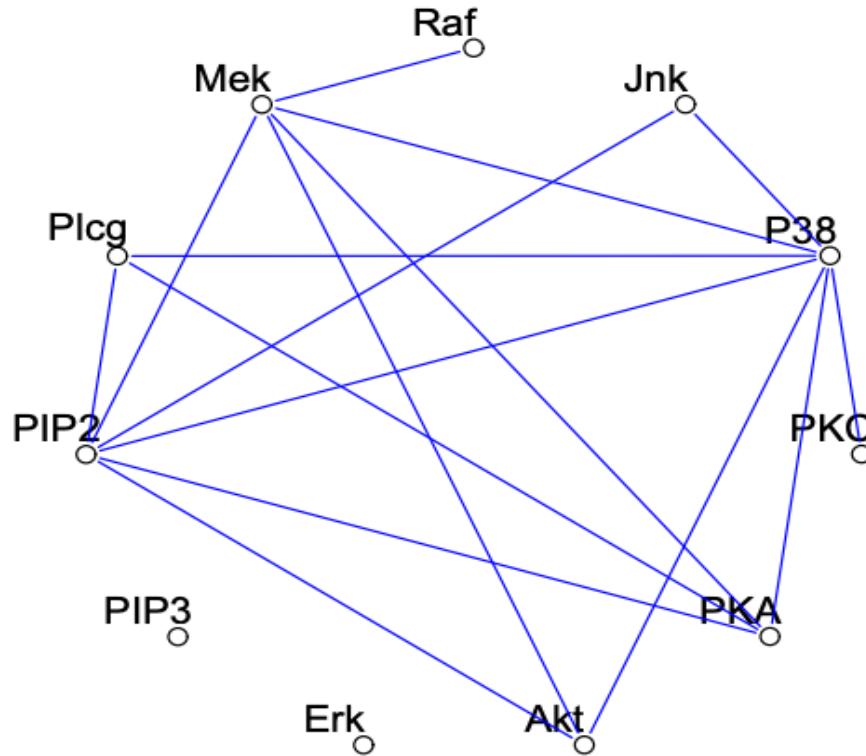
Motivation

- Discover the dependencies between components/nodes/variables/ unites/phenomena from data and represent them in the form of a network.
- Identifying the hub (active) nodes/variables in the network.
- Identifying the regulator nodes in the network.
- In some situations, it is helpful for casual interpretation.
- ...

Data structure: 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

Example 1: A flow cytometry dataset, Undirected network



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)**.

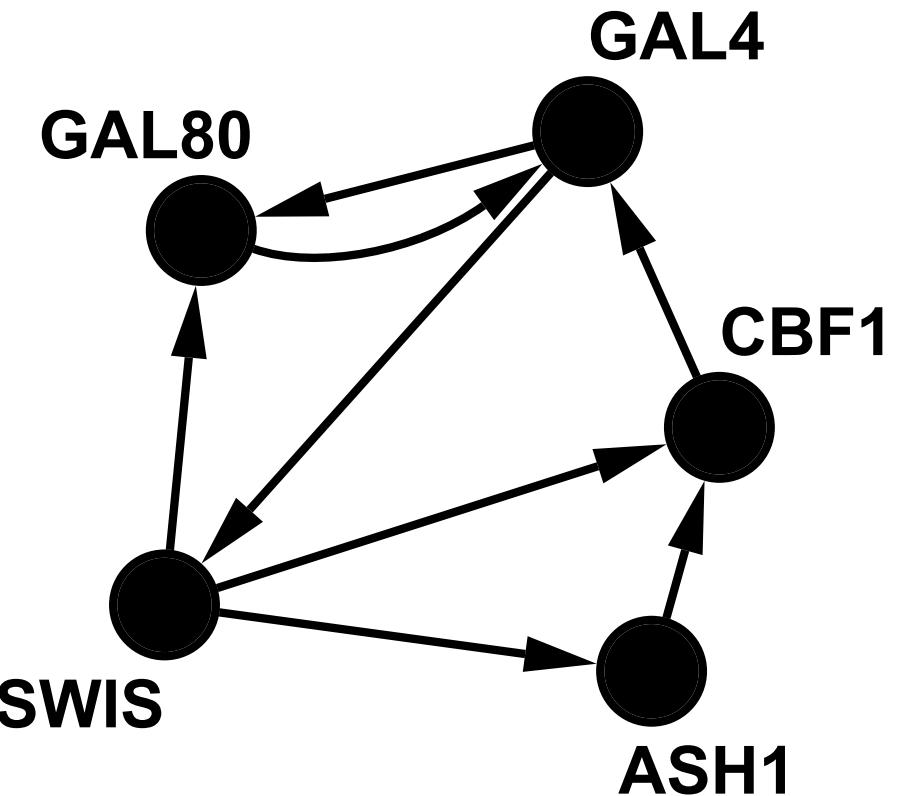
Gene expression data

• GAL80	GAL4	CBF1	ASH1	SWIS
• 3.4912	6.5525	2.6925	6.2262	3.9452
• 5.1125	5.9450	4.588	5.5015	3.3802
• 5.98875	5.9406	5.6675	5.9625	5.3575
• 4.33609	7.04375	6.88360	5.95651	4.7701
• 4.50636	7.07253	6.29784	4.99502	4.78430
• 4.41940	7.04159	6.84621	6.22766	4.78430
• 4.227071	6.88555	6.78362	6.21065	4.76709
• 4.001944	6.77326	7.14769	6.68846	5.52362
• 4.314800	6.61452	7.17859	6.52343	5.69920
• 4.931042	6.38479	6.44533	6.54546	5.17640
• 4.936013	6.51599	7.46192	6.223381	5.80765

Example 2: Yeast, directed network

Cantone et al. (2009) synthetically generated a small network of $n = 5$ genes in *Saccharomyces cerevisiae* (yeast). The network among those genes was obtained from synthetically designed yeast cells grown with different carbon sources: galactose ("switch on") or glucose ("switch off").

- Inferred from Dynamic Bayesian network models.



Data: 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	71.10	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 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.

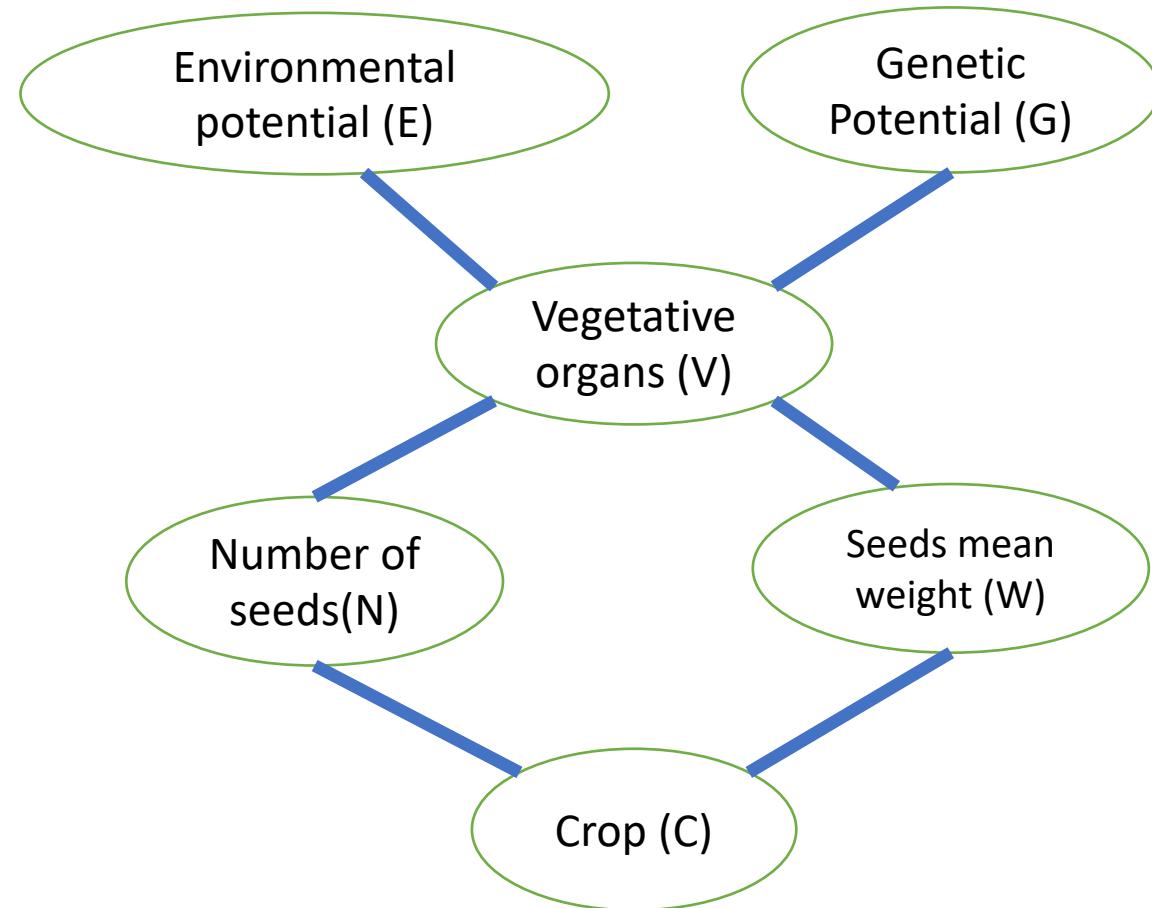
Number of seeds (N) is determined at the flowering time.

Seeds mean weight (W) is assessed in the plant's life.

Crop (C): The harvested grain mass.

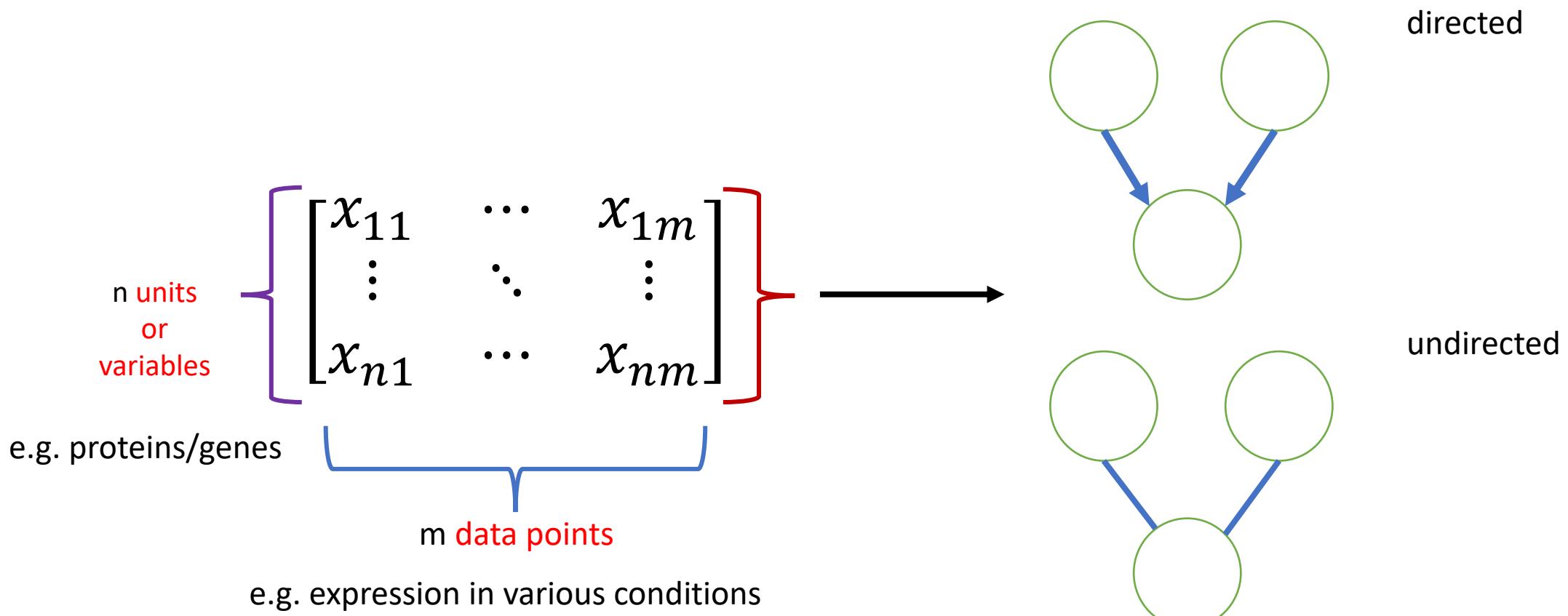
Analysis of a particular plant

- **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 harvested grain mass.

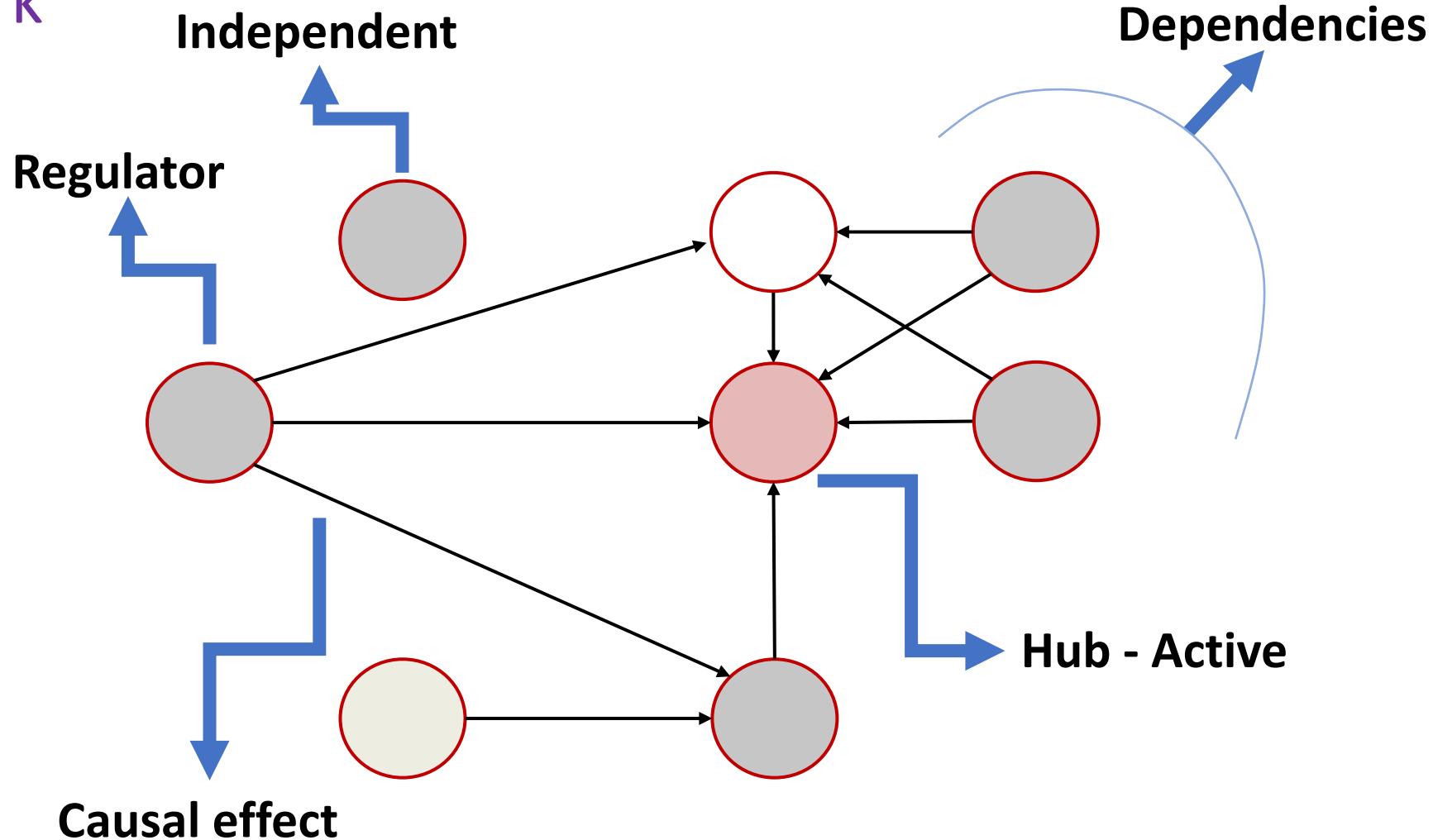


Aim

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



Network



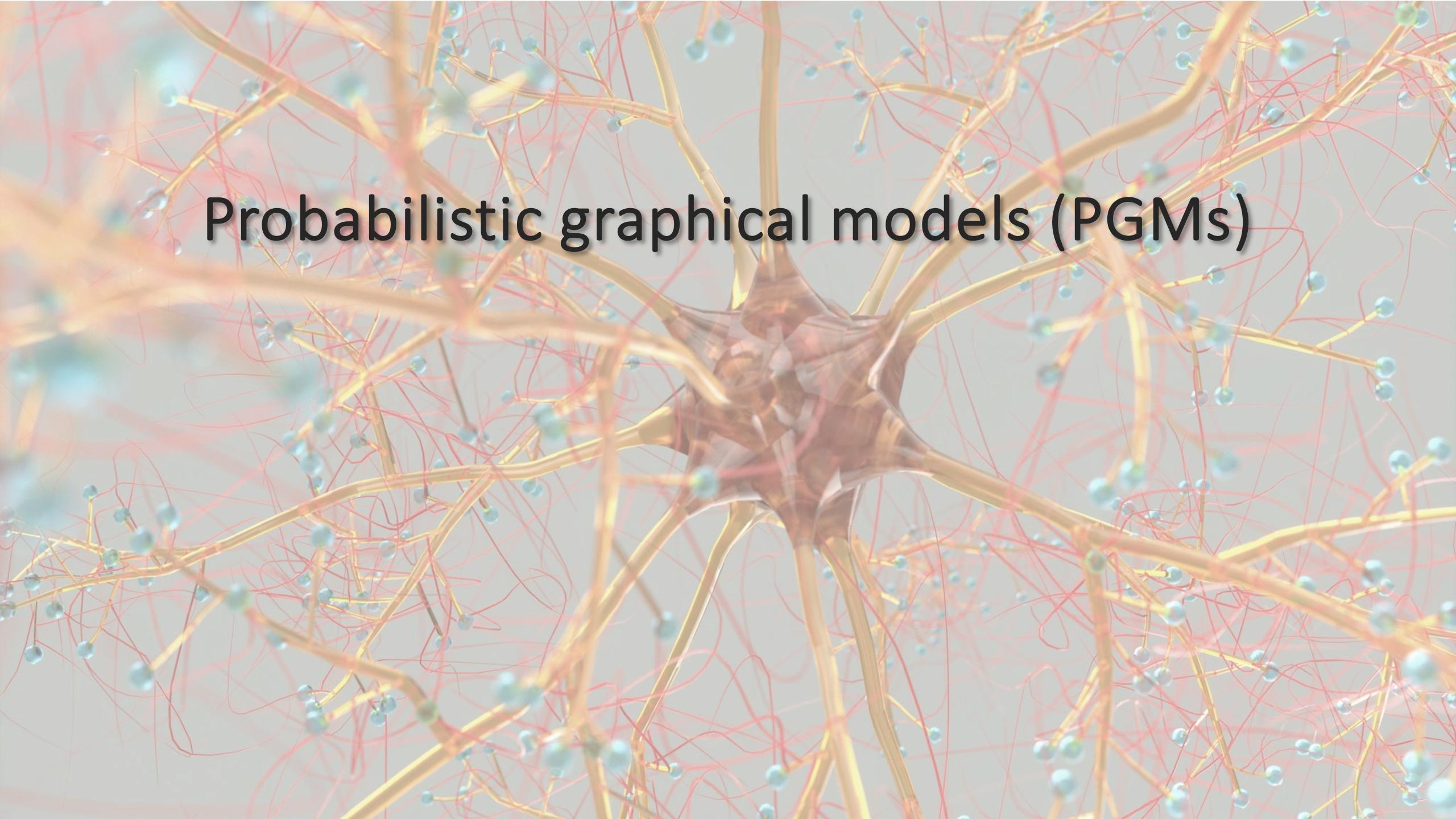
Example

- **A major goal in systems biology:**
Learning cellular network, such as gene regulatory transcription network.
- **Importance:**
E.g. identifying promising new drug targets as well as for developing personalized medical treatments.

Think about the application of this concept (e.g. in your field) for just a few minutes and discuss that in pairs.

- Consider the following questions for reflection in your field:
 - ✓ 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?

Probabilistic graphical models (PGMs)



Definition of Graphical Models

Michael Jordan, 1998

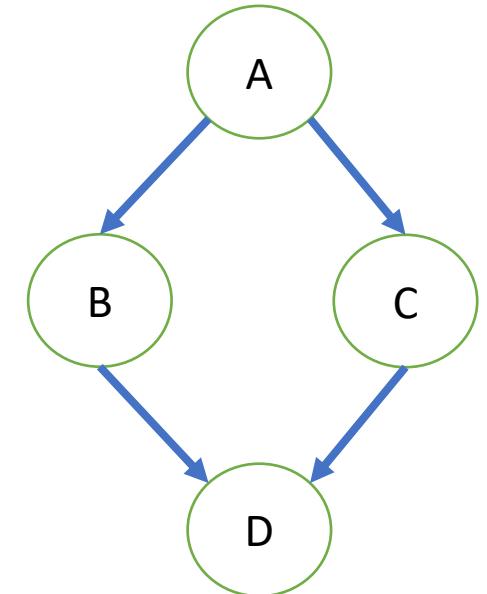
- A **Graphical model** is a **marriage** between **graph** theory and **probability** theory.
- A **graph** represents **relationships** among a set of **variables**.
- Probability helps estimate the graph (network) structure and answers queries.

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**

Edges and nodes in a graphical model

- **Random variables** as nodes.
- **Dependencies** as edges.
- **Graphical models** encode conditional (in)dependence assumptions between variables, indicating where edges should be placed.
- **What is conditional independence assumption?**



Independency

- The variables A and B are independent if

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

or

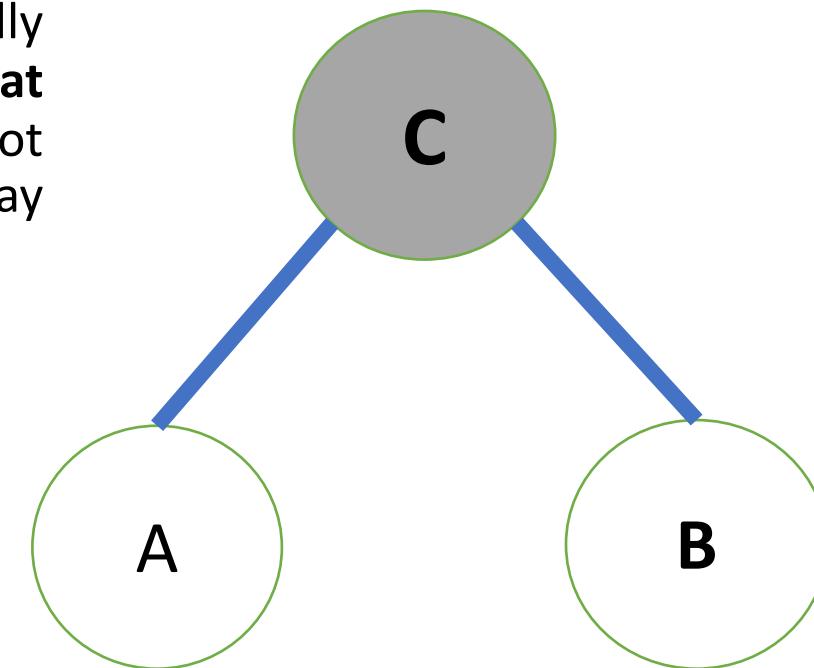
$$P(A, B) = P(A) P(B)$$

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

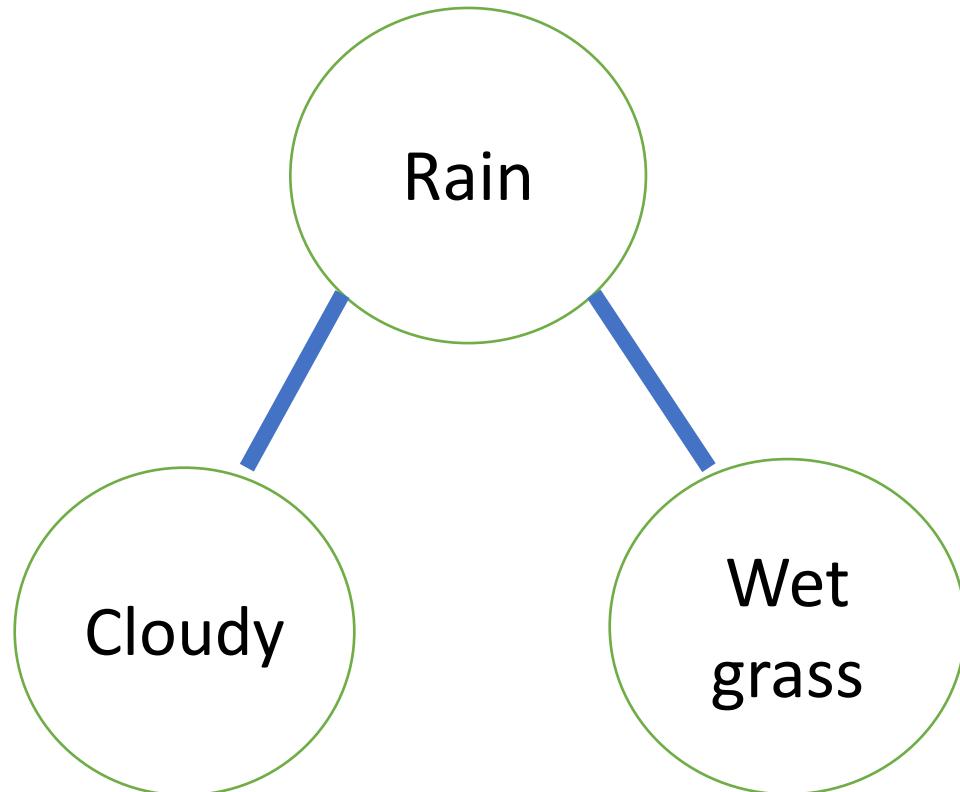
- Two nodes A and B given C are conditionally independent (no edge between them) if, **given that event C occurs**, the occurrence of event A does not give any information about B and the other way around.

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

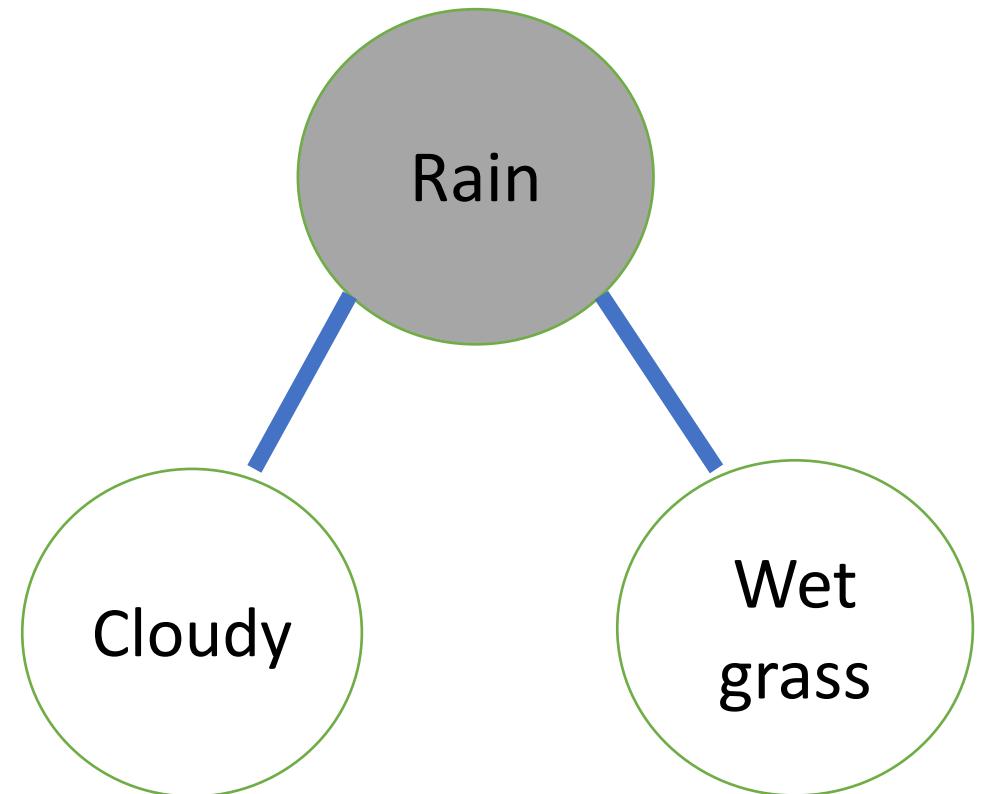


Conditional (in)dependency

we do not know whether it is rainy or not

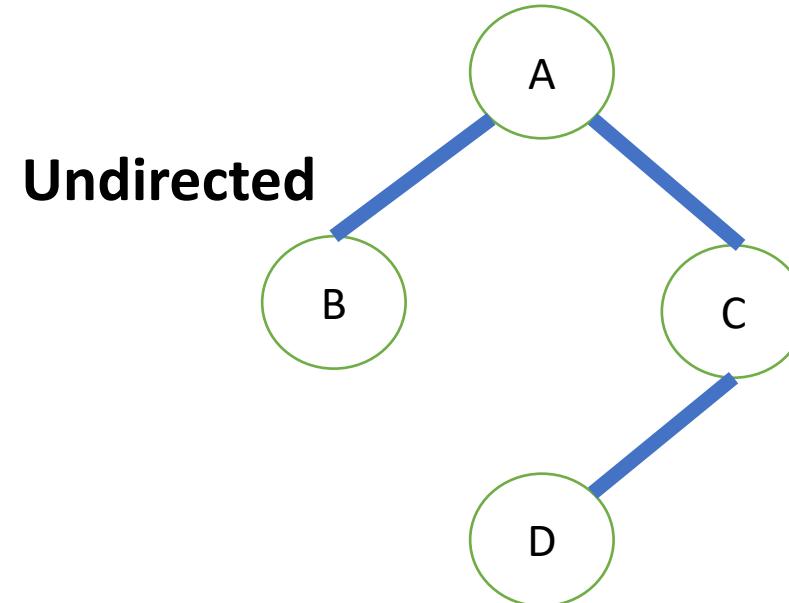
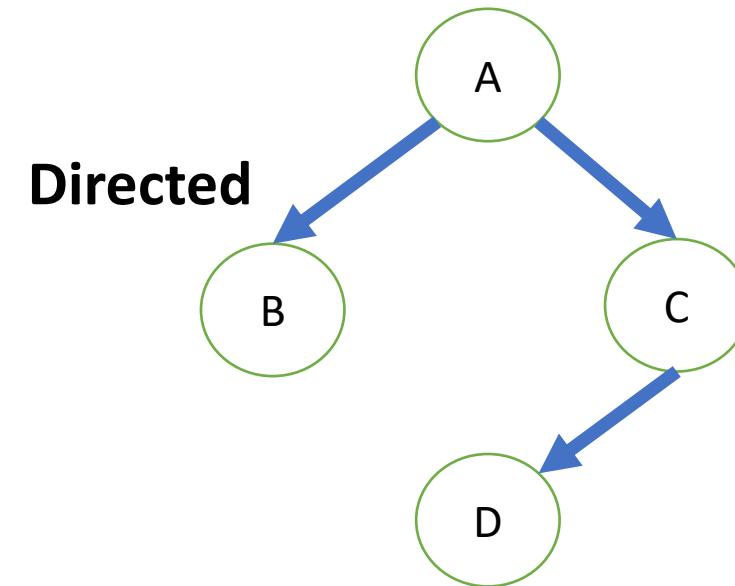


Conditional on rain: we know it is rainy or not



Edge and lack of edge

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



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 a directed graph).

They differ in the set of independences.

- **Combined**: chain graph

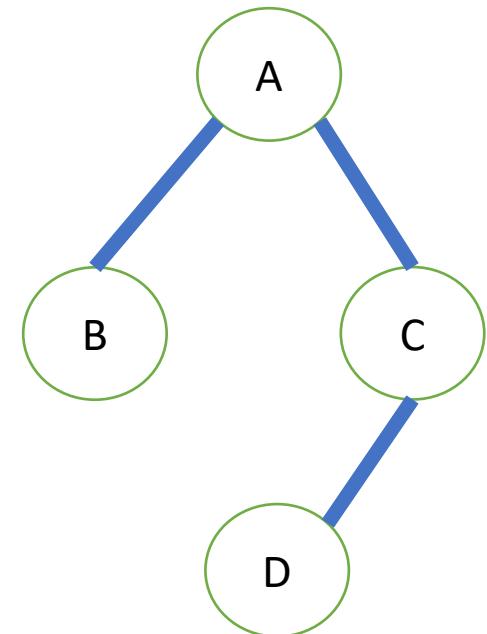
What are Undirected Graphical Models or **Markov Random Fields**?

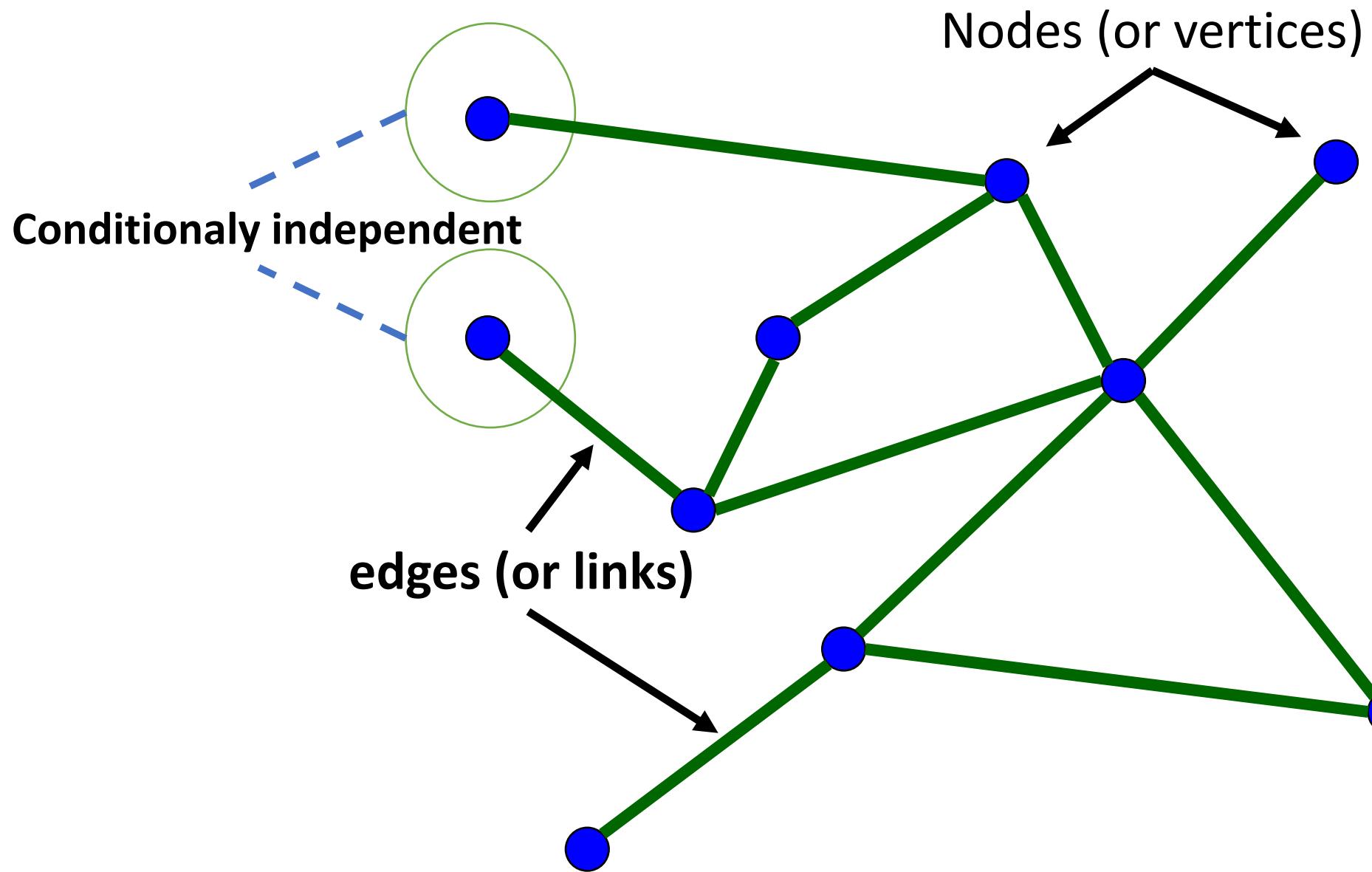
- Edges do not possess any form of orientation. They only show dependencies between nodes.
- Use undirected graphs where **relationships between two variables, do not have a clear directionality.**
- The edge (a,b) is identical to edge (b,a).

a ————— b

Conditional independency in a Markov Graph G

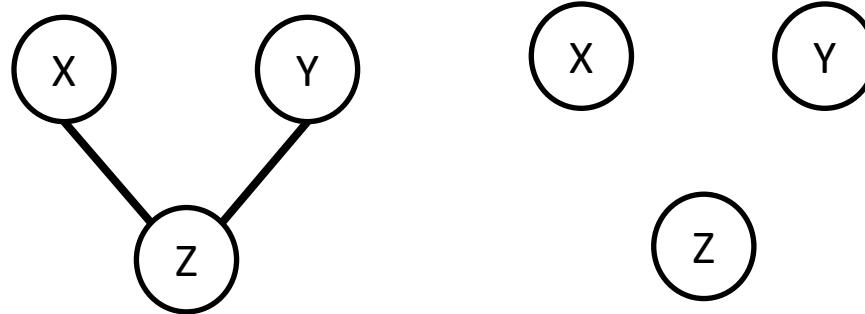
- The **absence of an edge** implies that the corresponding random variables are **conditionally independent** given the variables at the other vertices:
- **No edge joining B and D $\iff B \perp D \mid \text{rest}$** (pairwise Markov independencies of G).
- “Rest” refers to all other vertices in the graph.



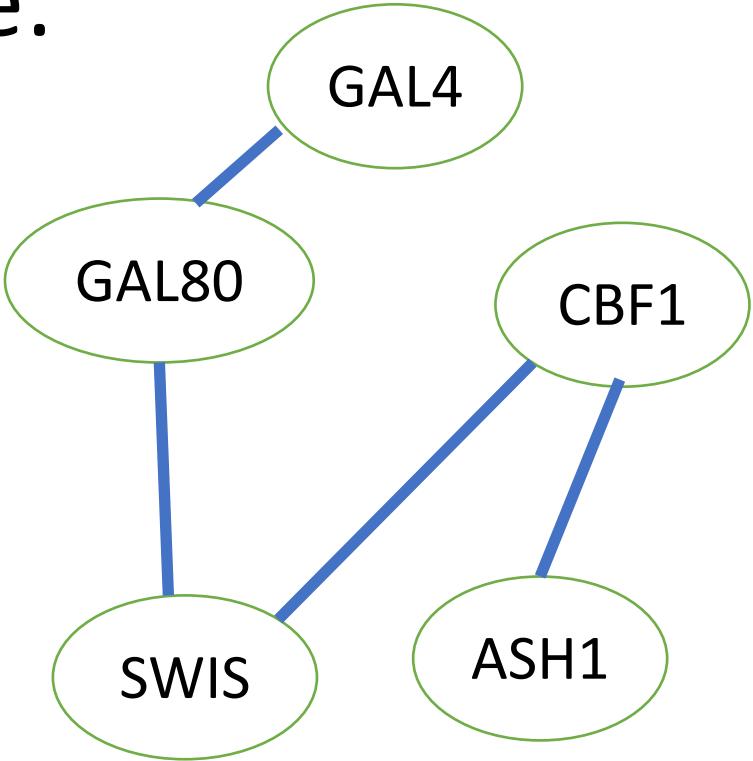


Undirected Graphical Models

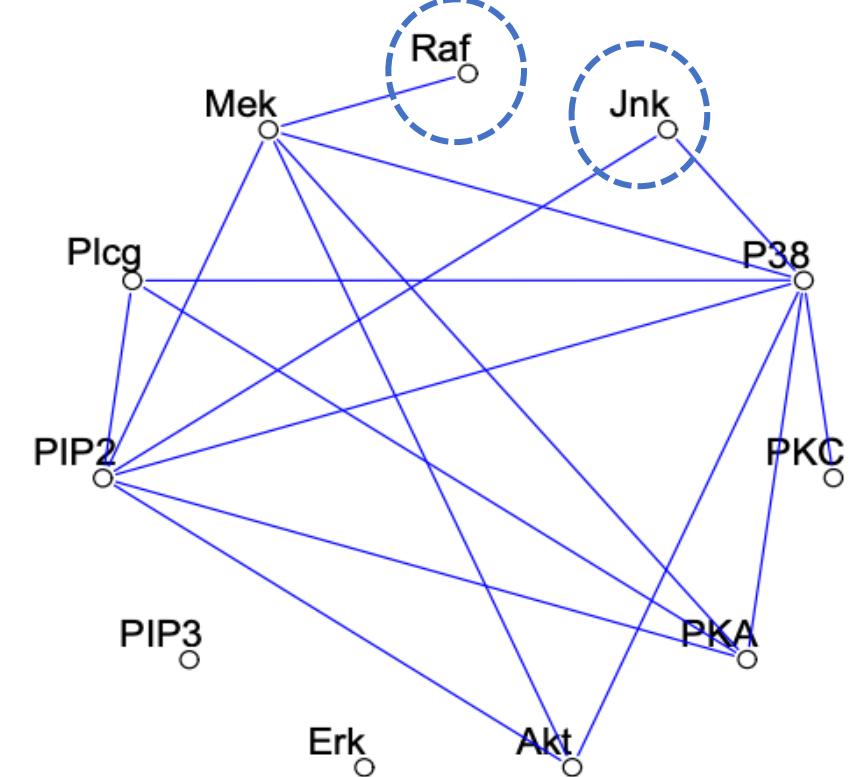
- Relationships that can be described with undirected graphs.


$$x \perp\!\!\!\perp y \mid z$$
$$\begin{matrix} ? \\ x \perp\!\!\!\perp y \end{matrix}$$

Example:



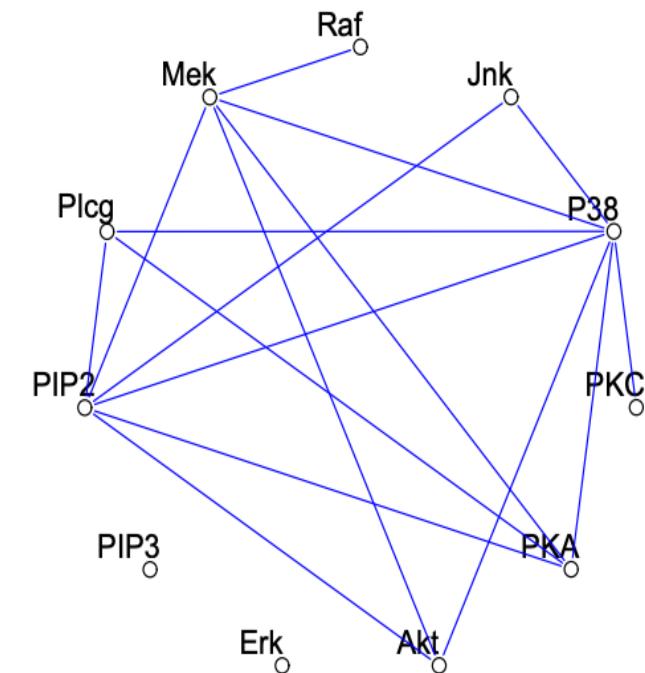
Estimated from yeast gene expression data.



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

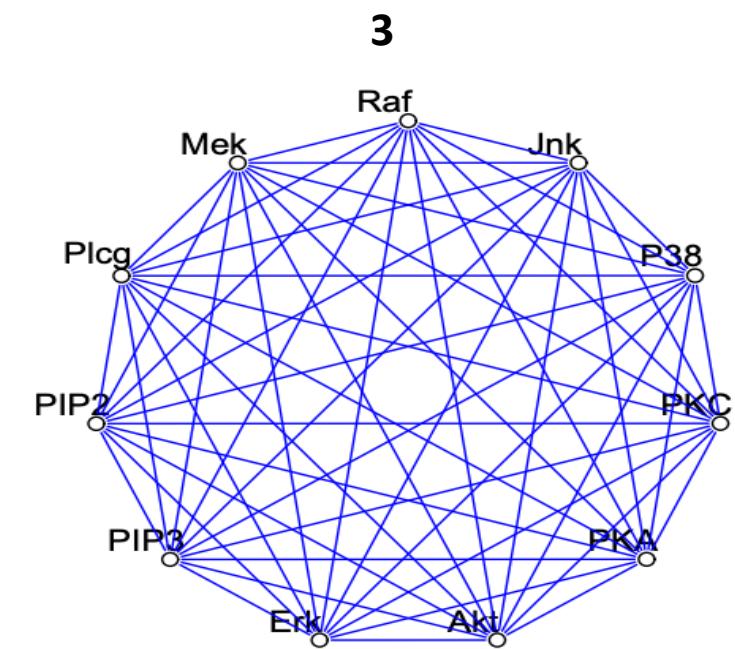
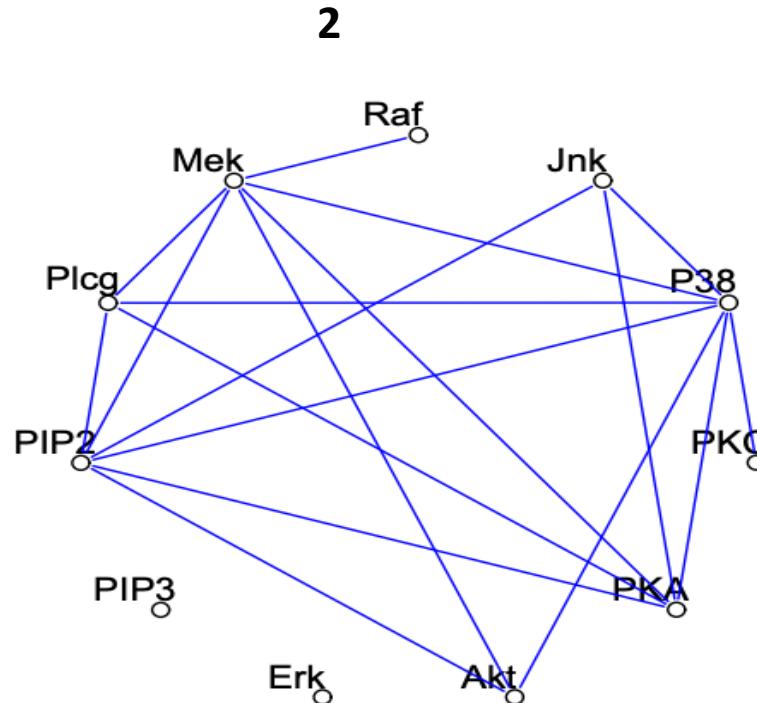
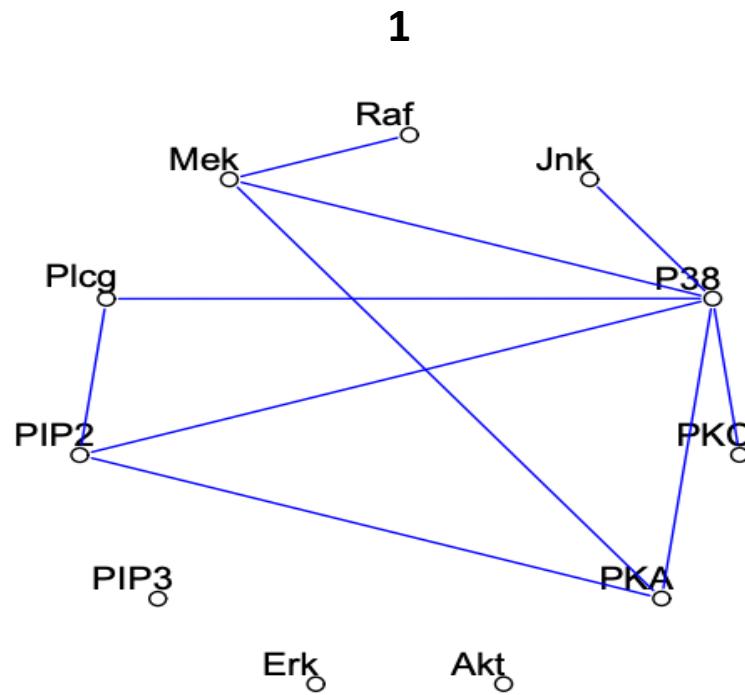
What can we do with graphical models

- Estimation of the structure of a graph. ←
- Estimation of the edge parameters from data.
- Computation of marginal vertex probabilities and expectations.



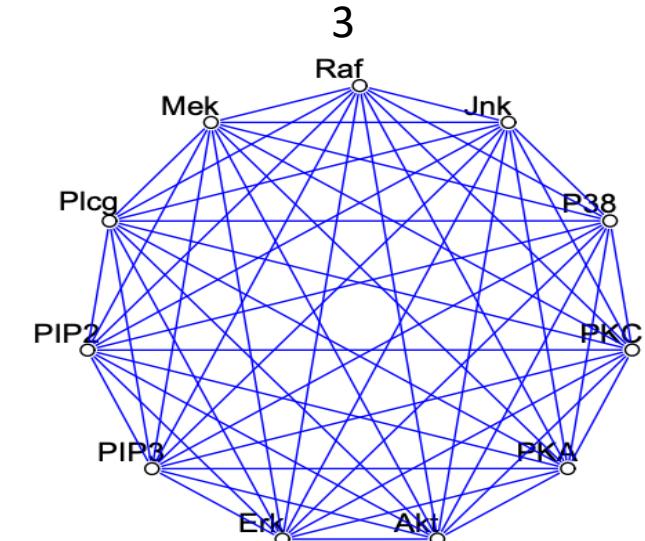
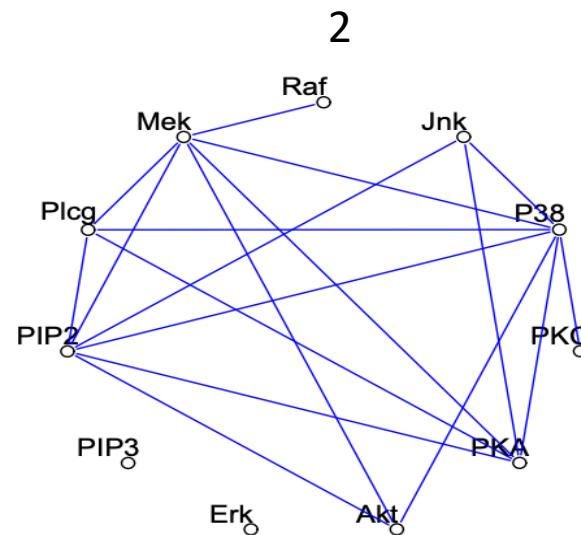
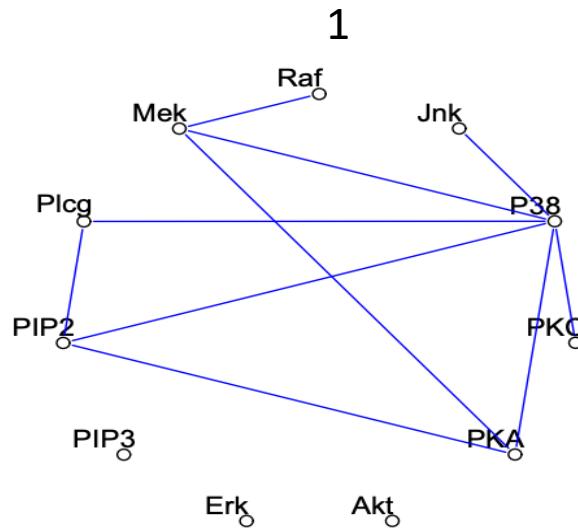
Hastie, T., Tibshirani, R. & Friedman, J. H.
(2009)

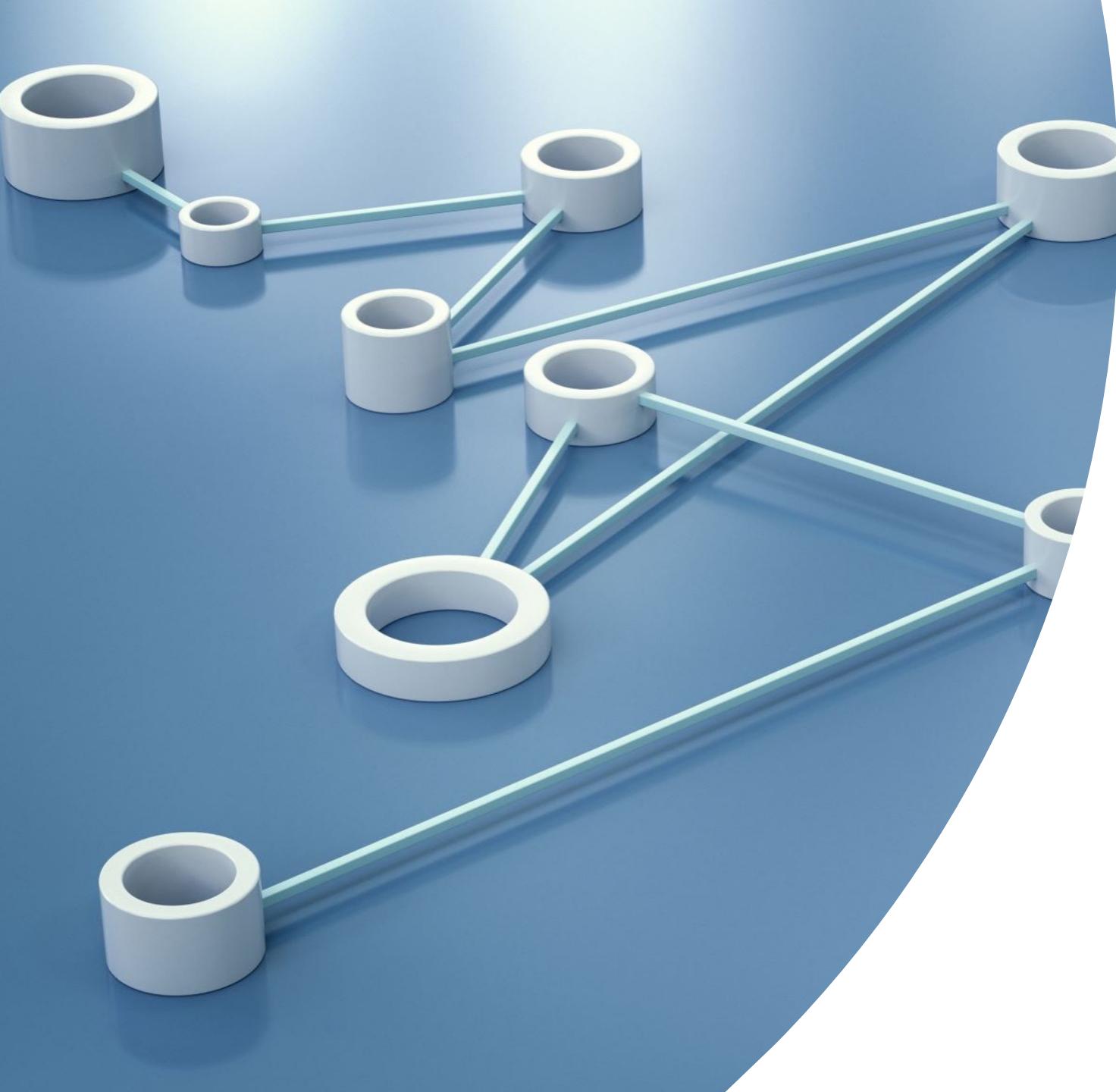
Imagine we estimated the network structure as below.
Which one is better?



Remark: Choosing the structure of a 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.

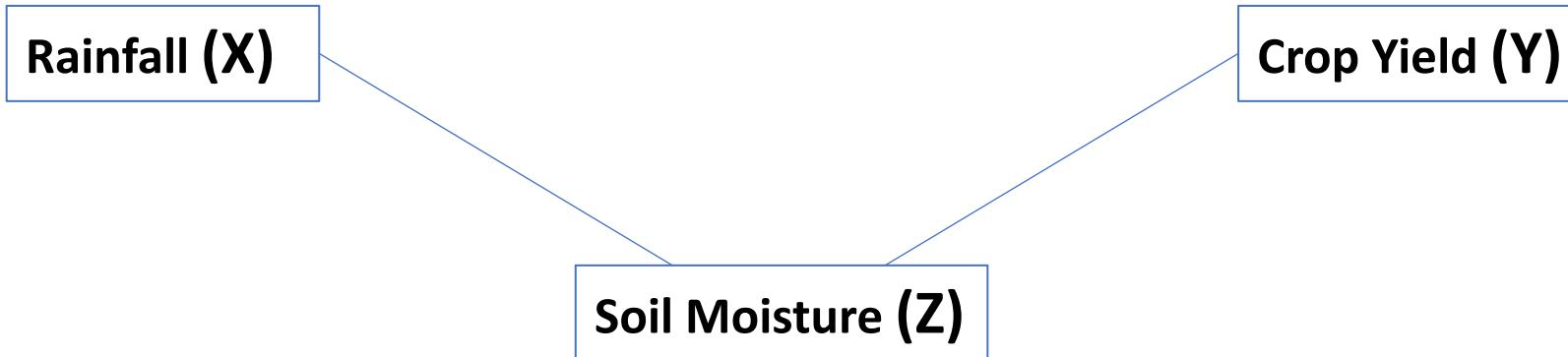




Some terminologies
useful for Network
interpretation

(for after network estimation)

Remark



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

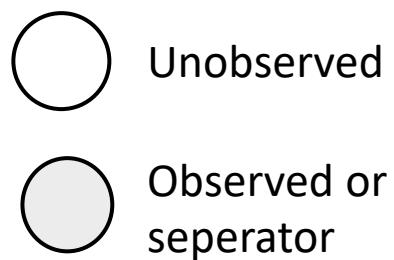
Unobserved ~= Unknown ~= Non-gray nodes

Soil Moisture (Z)

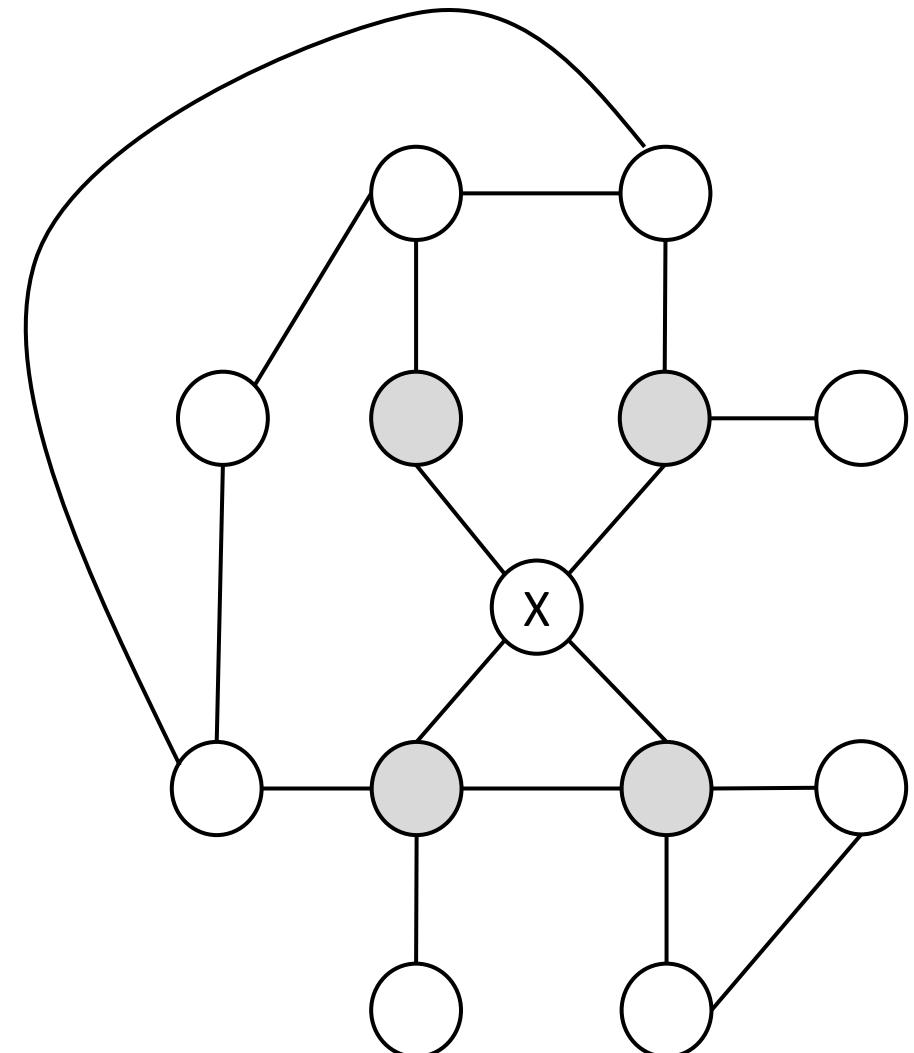
Observed ~= Known ~= Given ~= Gray nodes

Soil Moisture (Z)

Markov blanket/ Separators

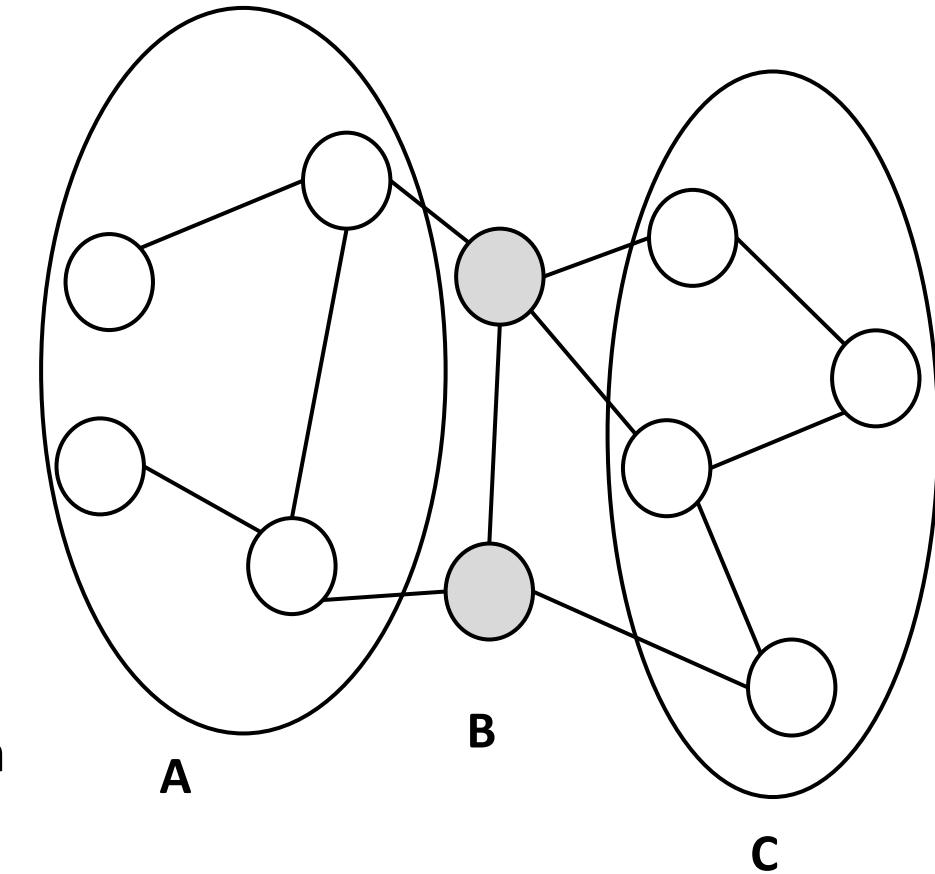


- If x 's **neighbours** are all observed, x is independent of all the others (they influence x only via its neighbours).
- In the undirected case the **Markov blanket** equals a node's neighbourhood.
- Useful for variables selection.



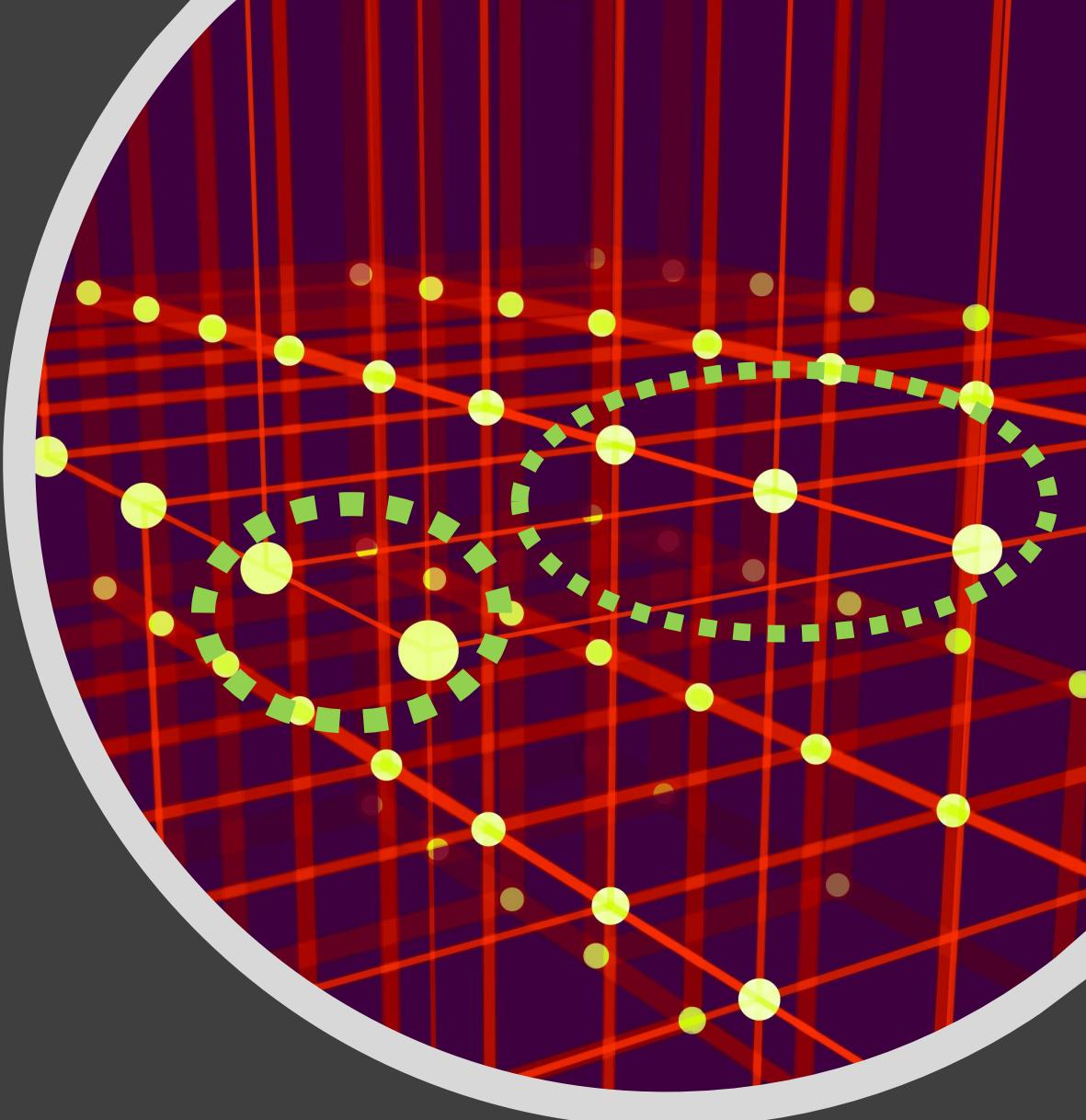
Separators and global Markov properties of G

- **B** is said to **separate** A and C if every path between A and C intersects a node in B, hence $A \perp\!\!\!\perp C | B$.
- Separators break the graph into conditionally independent pieces.
- These are known as the **global Markov properties of G**.
- This allows us to **decompose** graphs into **smaller more manageable pieces** and thus leads to essential simplifications in computation and interpretation.

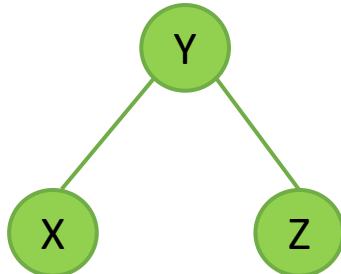


Clique

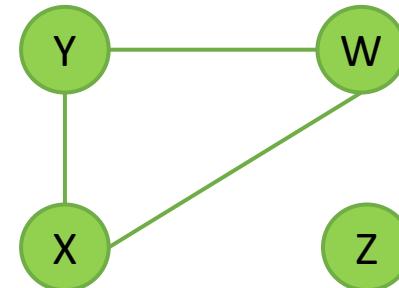
- We separate the graph into cliques (small pieces, including sufficient information).
- A clique is a **complete** subgraph— a set of vertices that are all adjacent to one another.



Examples



(a)



(b)

Maximal cliques:

(a) {X,Y}, {Y,Z} (b) {X,Y,W}, {Z}

Remark:

- If we can factorize an MRF according to the graph's cliques, we can define the probability of the entire graph, this helps in estimating the graph structure, and interpretation of the network.

Clique potential

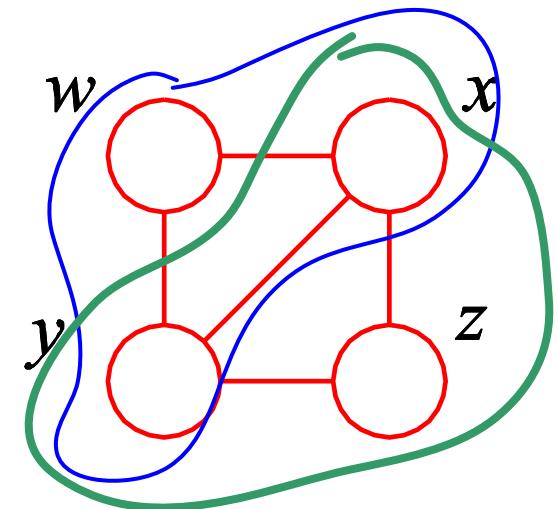
- Provided $p(x) > 0$, the joint distribution is a product of **non-negative functions** over the *cliques* of the graph

$$p(x) = \frac{1}{Z} \prod_C \psi_C(x_C)$$

where $\psi_C(x_C)$ is the *clique potential*, a function of only the values of the clique members in C

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

Z is a normalization constant.



Remark :Why are cliques important?

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

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

- Called **Ising models** or **Boltzmann machines**.
- The “nodes” are binary-valued.
- The Ising model is used to model the joint effects of pairwise interactions.

Undirected Markov networks with all discrete variables

- To fit an **L1-penalized logistic regression** model to each node as a function of the other nodes.
- And then optimization.
- Network structure.

Undirected Graphical Models for Continuous Variables

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

$$X \sim MN(\mu, \Sigma)$$

- Gaussian distribution represents second-order relationships (pairwise relationships between variables) and automatically encodes a pairwise Markov graph.
- All conditional distributions are also Gaussian.

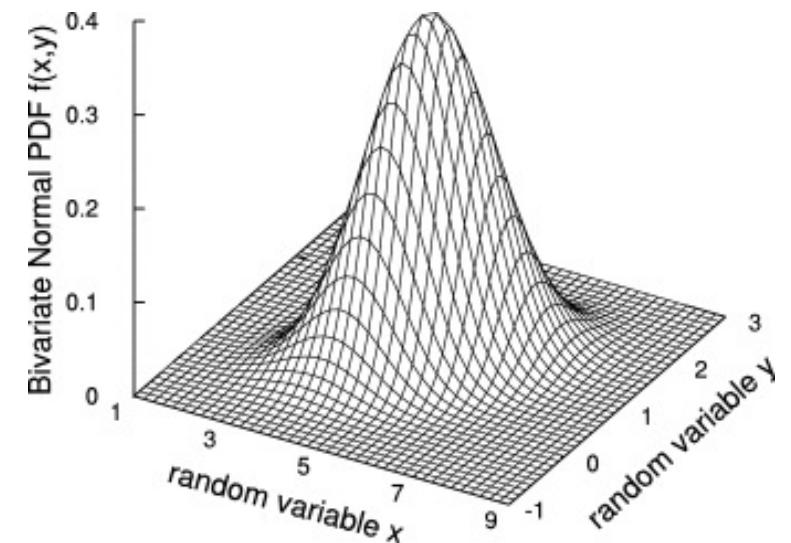
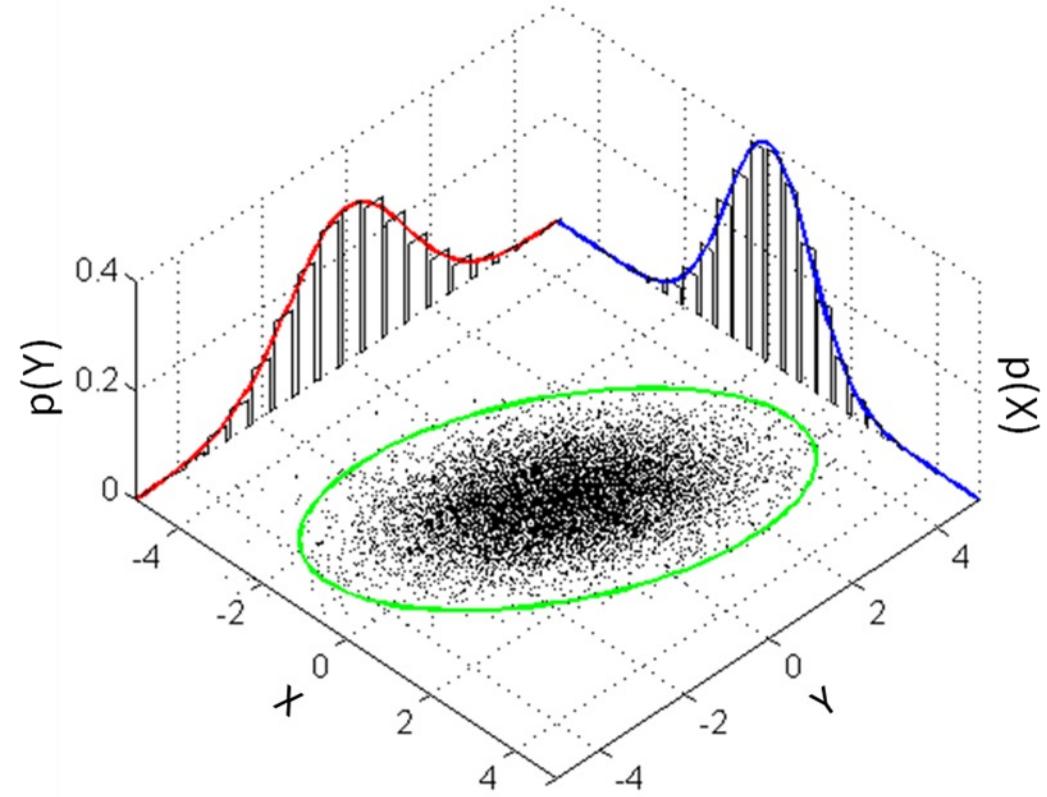
Undirected Graphical Models for Continuous Variables

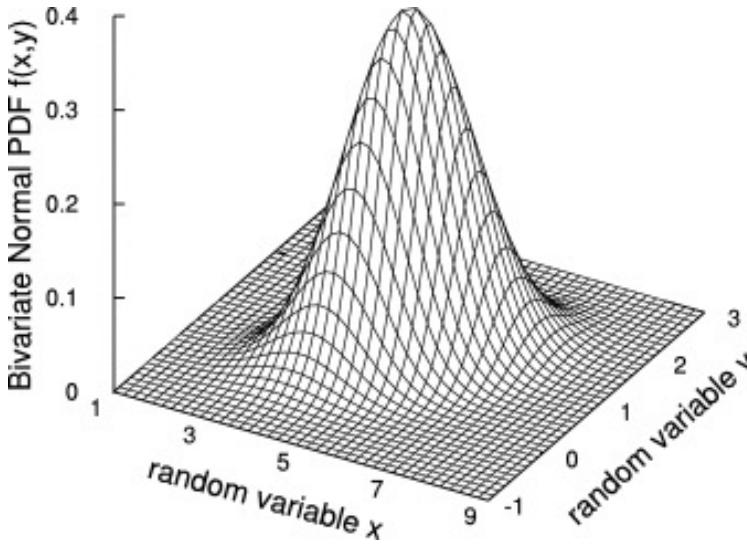
- The **Gaussian distribution** is almost always used for such graphical models, because of its convenient analytical properties.

$$f_X(x_1, \dots, x_k) = \frac{\exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)}{\sqrt{(2\pi)^k |\Sigma|}}$$

μ : mean parameter

Σ : covariance matrix





$$f_X(x_1, \dots, x_k) = \frac{\exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right)}{\sqrt{(2\pi)^k |\Sigma|}}$$

All pairwise relationships between variables can be extracted from the covariance matrix: Σ .

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 on how variables relate to one another.

- $cov_{x,y} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{N-1}$

$$\begin{matrix} & \begin{matrix} x & y \end{matrix} \\ \begin{matrix} x \\ y \end{matrix} & \begin{bmatrix} var(x) & cov(x, y) \\ cov(x, y) & var(y) \end{bmatrix} \end{matrix}$$

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

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

$$\begin{matrix} & & x & y & z \\ & x & \left[\begin{matrix} var(x) & cov(x,y) & cov(x,z) \\ cov(x,y) & var(y) & cov(y,z) \\ cov(x,z) & cov(y,z) & var(z) \end{matrix} \right] \\ y & & & & \\ z & & & & \end{matrix}$$

```
> 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

```
>
```

Covariance matrix Σ

Protein Data

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

$$COV_{x,y} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

```
> 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 **relationships** between variables (conditional dependencies).

$$\begin{matrix} & \begin{matrix} x & & y & & z \end{matrix} \\ \begin{matrix} x \\ y \end{matrix} & \begin{bmatrix} var(x) & cov(x, y) \\ cov(x, y) & var(y) \end{bmatrix}^{-1} & \begin{matrix} x \\ y \\ z \end{matrix} & \begin{bmatrix} var(x) & cov(x, y) & cov(x, z) \\ cov(x, y) & var(y) & cov(y, z) \\ cov(x, z) & cov(y, z) & var(z) \end{bmatrix}^{-1} \end{matrix}$$

Inverse variance-covariance matrix (precision matrix) Σ^{-1}

- If the ij^{th} component of Σ^{-1} is zero, then variables i and j are conditionally independent given the other variables (**there is no edge between them**).
- Thus, Σ^{-1} captures all the second-order (pairwise) information.

Why does inversion of a covariance matrix yield partial correlations between random variables?

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

(precision matrix) Σ^{-1}

solve(A)

	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

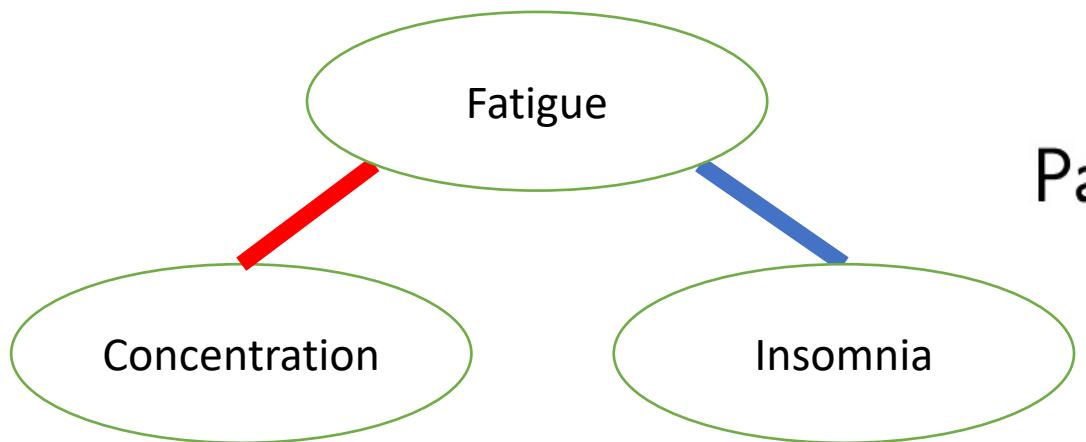
Partial correlation coefficients

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

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

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

Partial correlation (standardization)



Partial correlations:

$$\begin{matrix} & \text{C} & \text{F} & \text{I} \\ \text{C} & 1 & -0.25 & 0 \\ \text{F} & -0.25 & 1 & 0.30 \\ \text{I} & 0 & 0.30 & 1 \end{matrix}$$

Covariance Matrix Σ

```
> 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 covariance matrix (aka precision matrix) Σ^{-1}

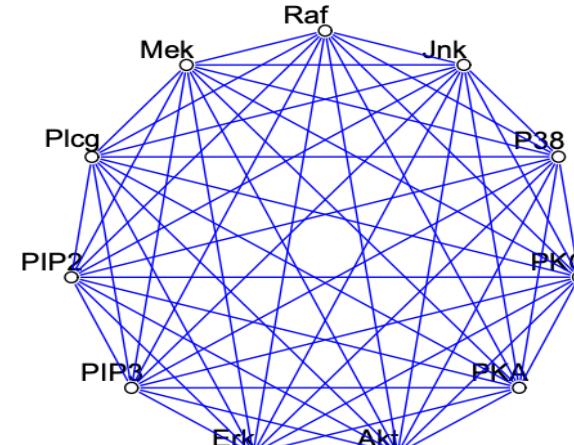
	Raf	Mek	Plcg	PIP2	PIP3	Erk	Akt	PKA	PKC	P38	Jnk
Raf	0.001560	-0.001886	-0.000081	0.000003	0.000014	-0.000142	0.000111	-4.0e-06	-0.000103	0.000184	-0.000065
Mek	-0.001886	0.003629	0.000102	-0.000033	-0.000059	-0.000016	0.000007	3.0e-06	0.000208	-0.000206	0.000124
Plcg	-0.000081	0.000102	0.004772	-0.000046	-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 matrix, standardized

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

Standardize

$$\text{Cor} (Y_i, Y_j | \mathbf{Y}^{-(i,j)}) = -\frac{\kappa_{ij}}{\sqrt{\kappa_{ii}} \sqrt{\kappa_{jj}}}$$



Partial correlation network

Covariance Matrix

$$\begin{matrix} & \begin{matrix} x & y & z \end{matrix} \\ \begin{matrix} x \\ y \\ z \end{matrix} & \left[\begin{matrix} var(x) & cov(x, y) & cov(x, z) \\ cov(x, y) & var(y) & cov(y, z) \\ cov(x, z) & cov(y, z) & var(z) \end{matrix} \right] \end{matrix}$$

Inverse covariance matrix (aka precision matrix) Σ^{-1}

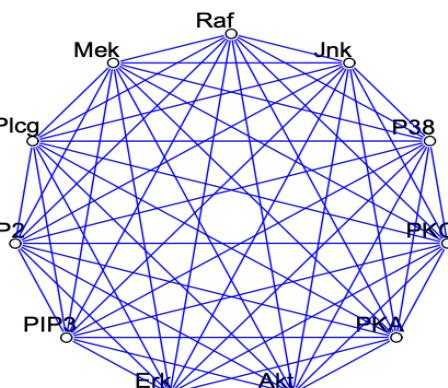
$$\begin{matrix} & \begin{matrix} x & y & z \end{matrix} \\ \begin{matrix} x \\ y \\ z \end{matrix} & \left[\begin{matrix} var(x) & cov(x, y) & cov(x, z) \\ cov(x, y) & var(y) & cov(y, z) \\ cov(x, z) & cov(y, z) & var(z) \end{matrix} \right] \end{matrix}^{-1}$$

Partial correlation matrix, standardized

$$\begin{matrix} & \begin{matrix} x & y & z \end{matrix} \\ \begin{matrix} x \\ y \\ z \end{matrix} & \left[\begin{matrix} P_{xx} & P_{xy,z} & P_{xz,y} \\ P_{xy,z} & P_{yy} & P_{yz,x} \\ P_{xz,y} & P_{yz,x} & P_{zz} \end{matrix} \right] \end{matrix}$$

Standardize

$$\text{Cor} (Y_i, Y_j | Y^{-(i,j)}) = -\frac{\kappa_{ij}}{\sqrt{\kappa_{ii}}\sqrt{\kappa_{jj}}}$$

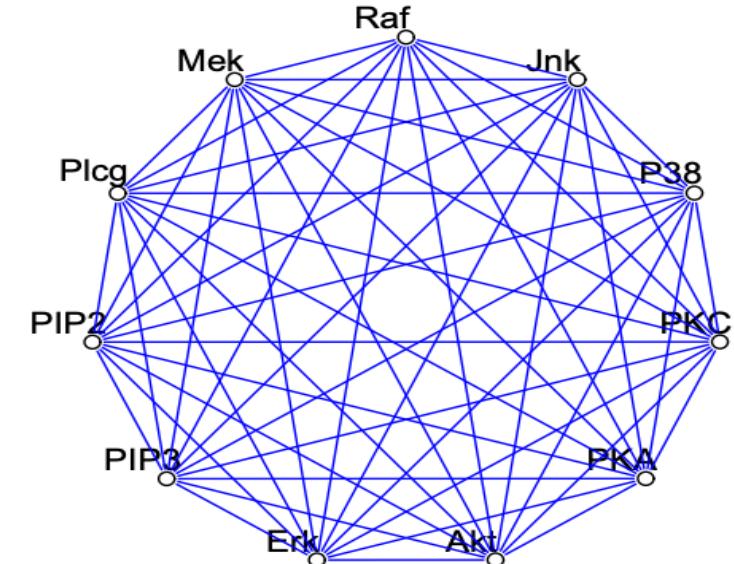


Partial correlation network

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	1.000000000	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	1.000000000	-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	1.0000000



What do you think?

Is that really helpful?

Really difficult for interpretation.

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

Remark

- In real-life datasets, especially those with many variables, it's common for the precision matrix to have very few exact zeros due to noise and other factors.

Network structure:

- **Partial correlation coefficients**



- **Network structure**

The partial correlation is a measure of their relationship given all the other variables in the data set.

Can we infer a network directly from **partial correlation coefficient?** Almost **NO**

Zero means there is no edge.

Non-zero means there is an edge.

How to estimate the Network structure?

- **Problem:** In most cases we do not know which edges to omit from our graph based on *partial correlation coefficients* and would like to discover this from the data.
- **Solution:** Shrink the small components partial correlation matrix to zero using **L1 (lasso) regularization**, as in **Graphical Lasso**.
- This gives sparse network. This is easy for interpretation, So a big advantage. 

Graphical Lasso, a tool for estimating the network structure

R package is glasso

Friedman et al. (2008)

- We controls the number of zeros in the precision matrix $\Theta = \Sigma^{-1}$.
- The idea is **to shrink low correlations to 0** such that they disappear from the graph.
- Fit a lasso regression between variables.

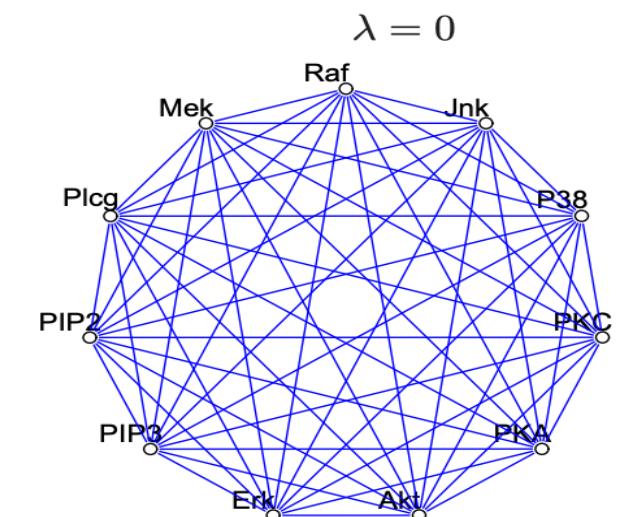
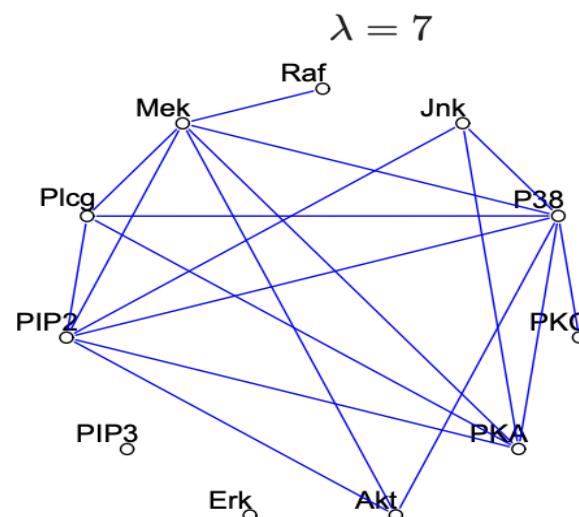
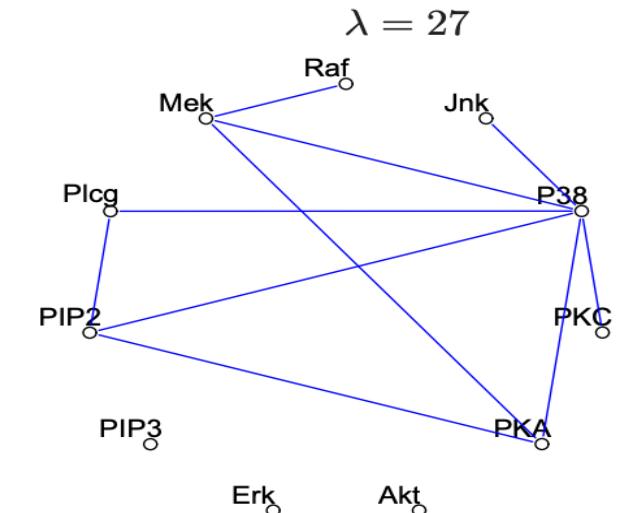
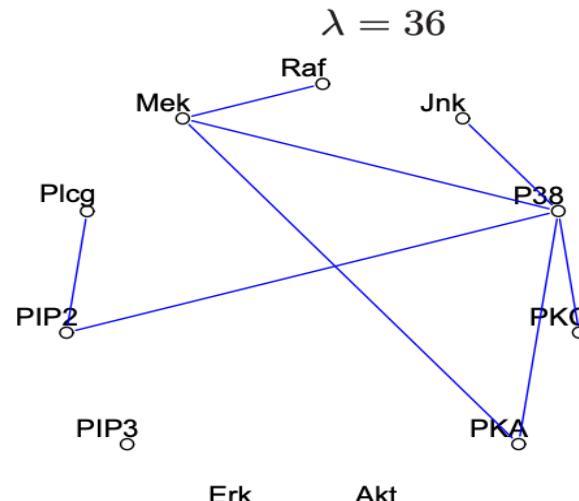
The graphical lasso

- 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.

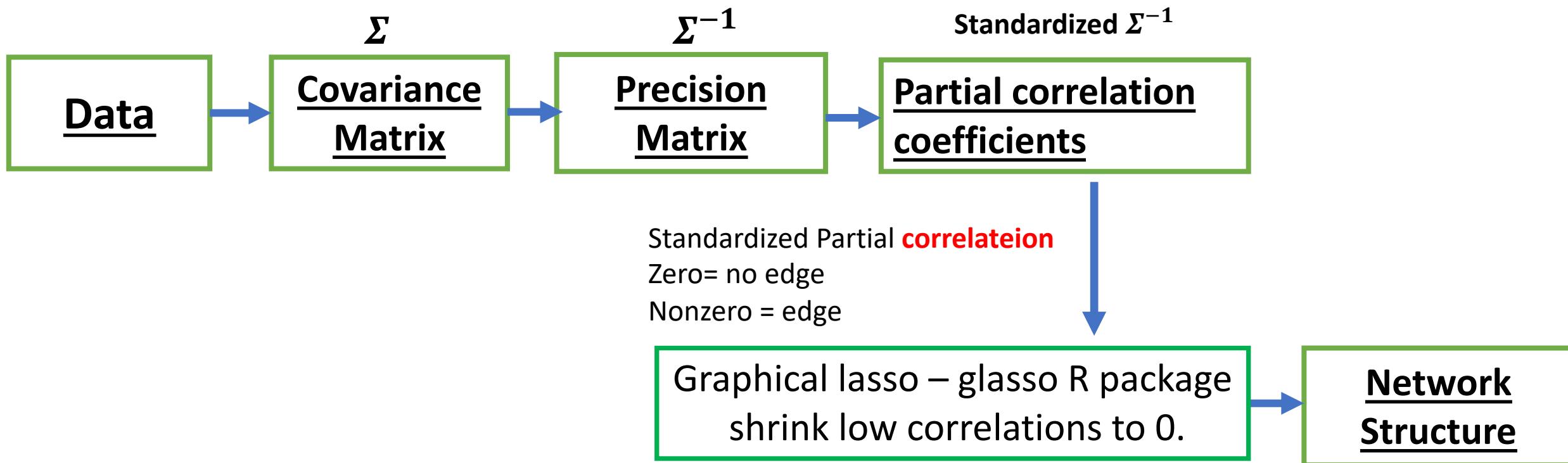
Graphical lasso - Penalty parameter

Friedman et al. (2008)

- Four different graphical lasso solutions for 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:

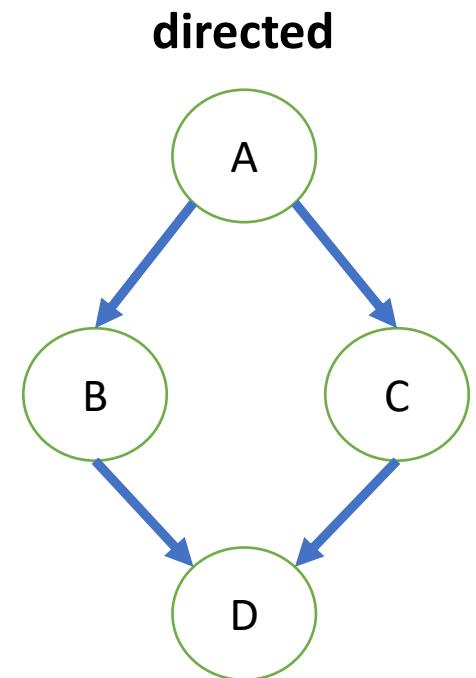
- MRFs can be applied to a wider range of problems in which there is no natural directionality associated with variable dependencies.
 - MRFs succinctly express certain dependencies (cycle) that BNs cannot easily describe (the converse is also true).
 - Conditional independence is easier to read off for MRFs.
 - In MRFs, there is only one type of neighbors (simpler Markov blanket).
-

Disadvantages of MRFs:

- Undirected models may be difficult to interpret.

Directed graphs

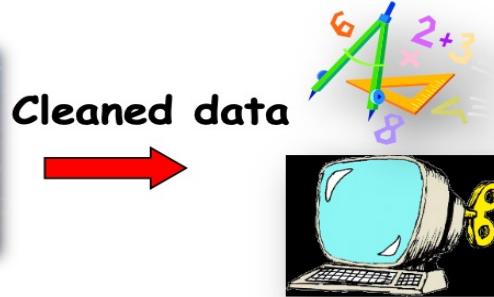
- All edges have a specific direction that depicts the nature and dependence of the relationship between the two vertices at the end of an edge.
- A graph that **does not** form any cycles is called a Directed Acyclic Graph(DAG).



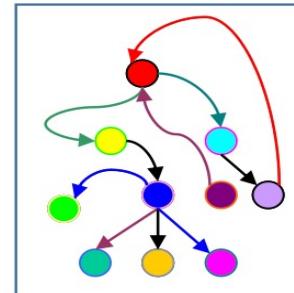
Directed graphs

- **Bayesian networks** are a type of Directed Acyclic Graphs that will be discussed in the afternoon.
- We will talk about this in the afternoon 😊.

In a nutshell



Machine Learning



Statistical Methods

Network inference

- 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>