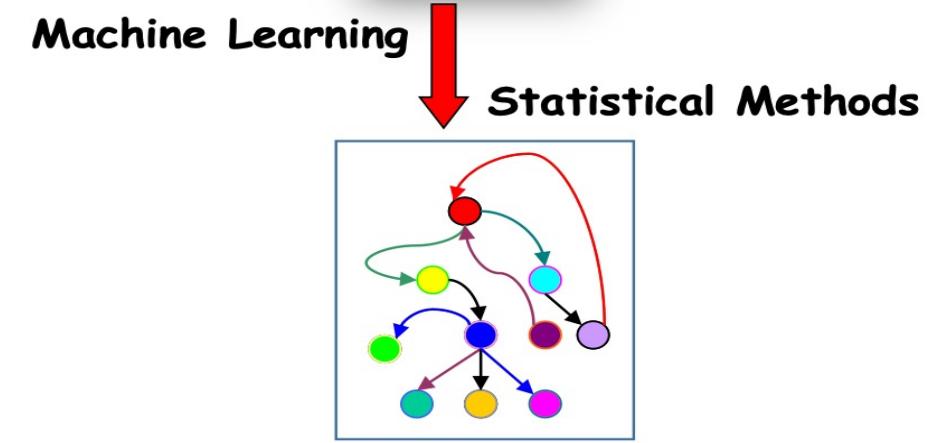
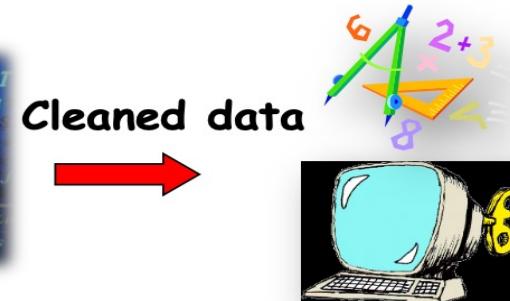


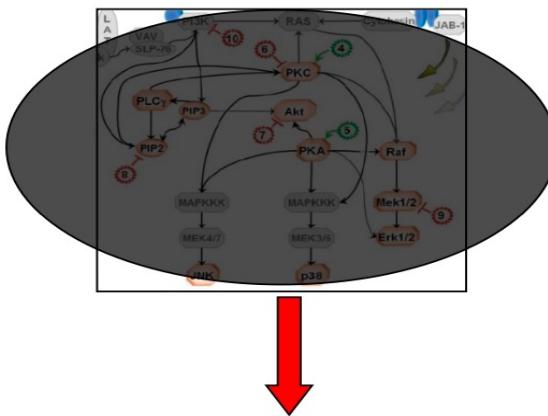
Graphical Models for Network Inference

Mahdi Shafiee Kamalabad

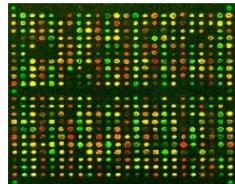
In a nutshell



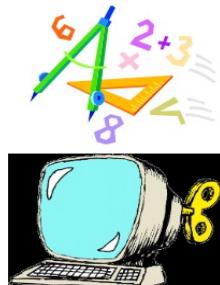
possibly
completely
unknown



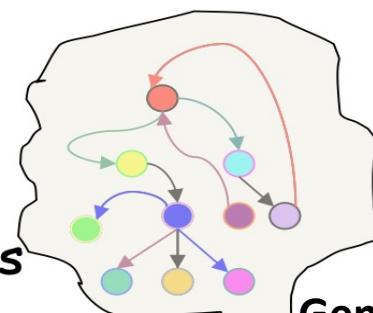
E.g.: Gene-
Microarray
experiments



data data



Machine Learning
statistical methods

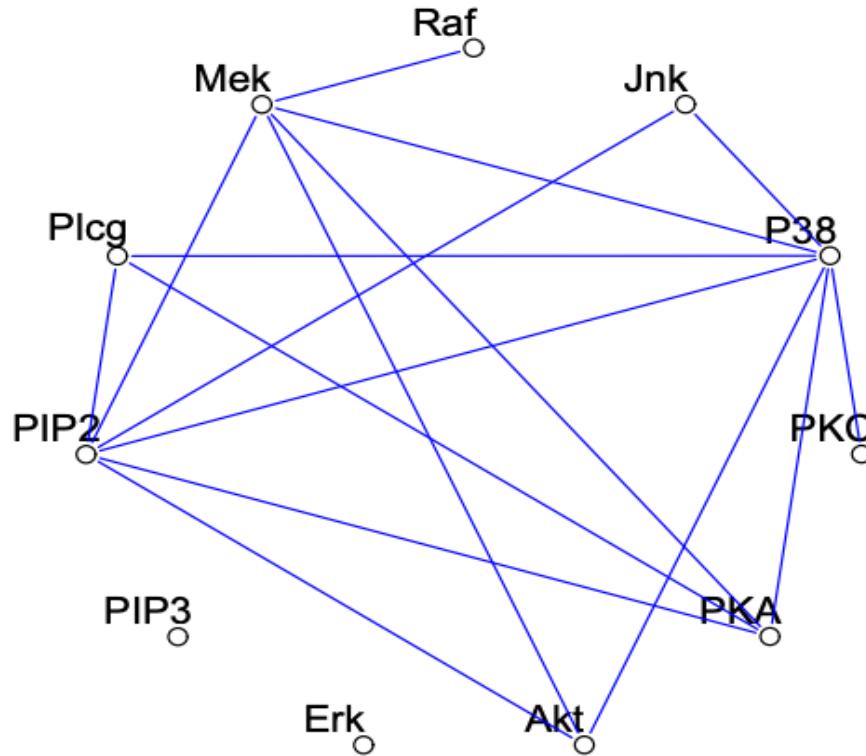


Gene regularity Network

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 we will discuss it later, **Sachs et al. (2003)**.

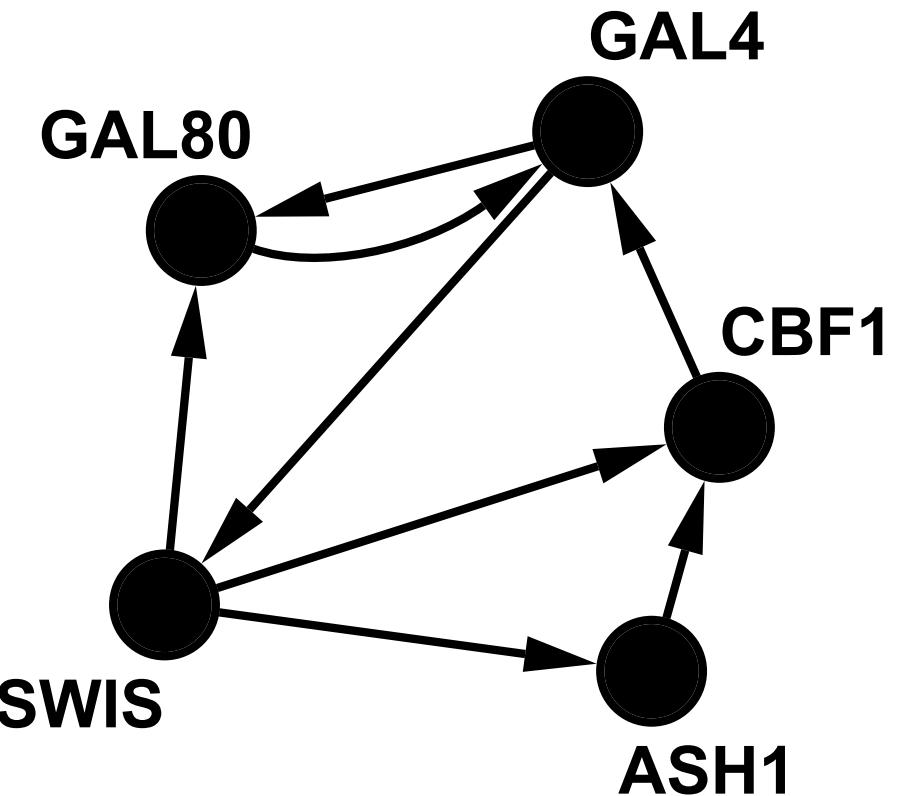
Gene expression data

| "GAL80" | "GAL4" | "CBF1" | "ASH1" | "SWIS" |
|---------------|-------------|-------------|------------|-------------|
| • 3.49 | 6.5525. | 2.6925 | 6.2262 | 3.945 |
| • 5.1125 | 5.9450 | 4.588 | 5.5015 | 3.3802808 |
| • 5.98875 | 5.940625 | 5.6675 | 5.9625 | 5.3575 |
| • 4.33609 | 7.04375 | 6.88360 | 5.95651 | 4.7701 |
| • 4.50636775 | 7.0725 6.49 | 6.29784 | 4.995024. | 4.78430925 |
| • 4.41940525 | 7.041591875 | 6.846219375 | 6.22766125 | 4.78430925 |
| • 4.227071 | 6.885554125 | 6.7836285 | 6.21065312 | 4.767093625 |
| • 4.001944625 | 6.77326375 | 7.147697875 | 6.68846375 | 5.523626375 |
| • 4.314800625 | 6.614521625 | 7.178597375 | 6.52343 | 5.69920225 |
| • 4.93104225 | 6.384797125 | 6.445337625 | 6.54546437 | 5.17640975 |
| • 4.936013625 | 6.51599725 | 7.46192 | 6.223381 | 5.807650375 |

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.

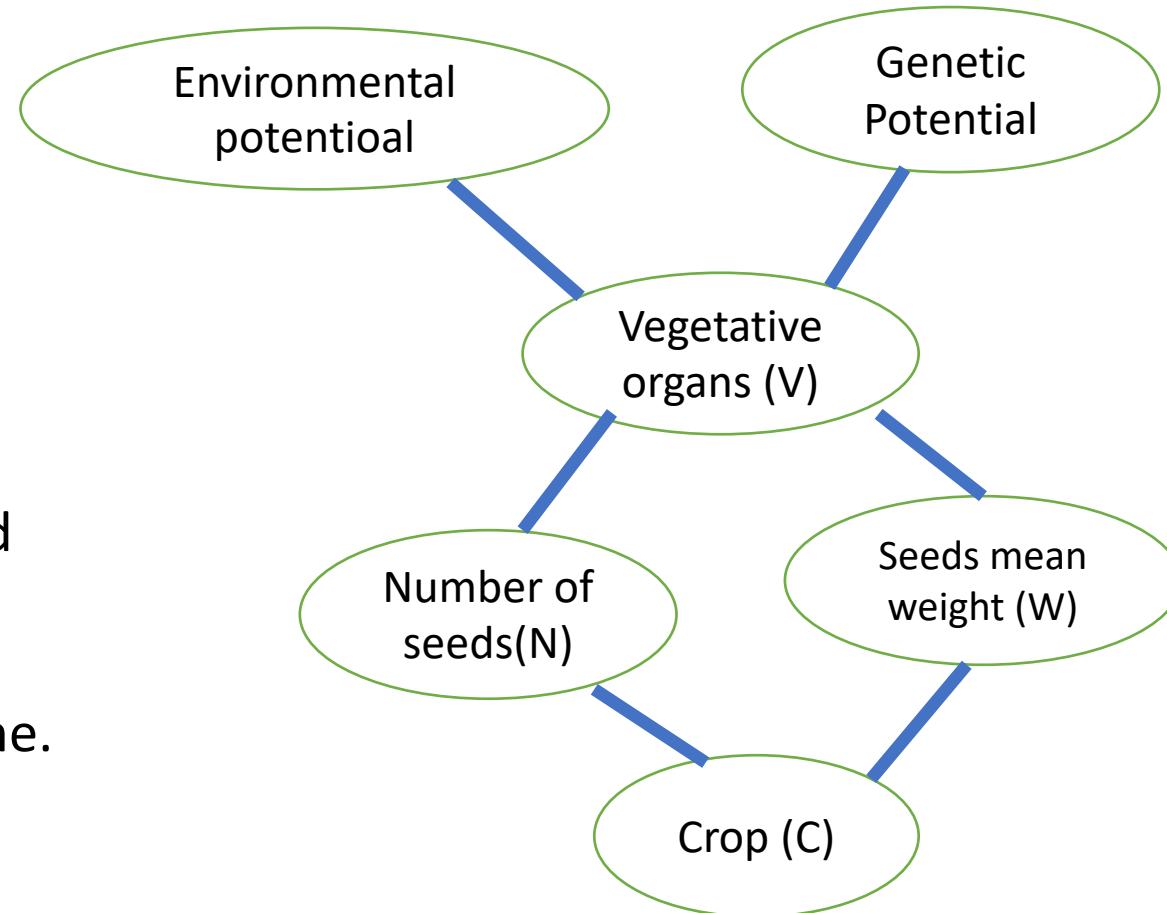


Data: Particular plant

- "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 71.098 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

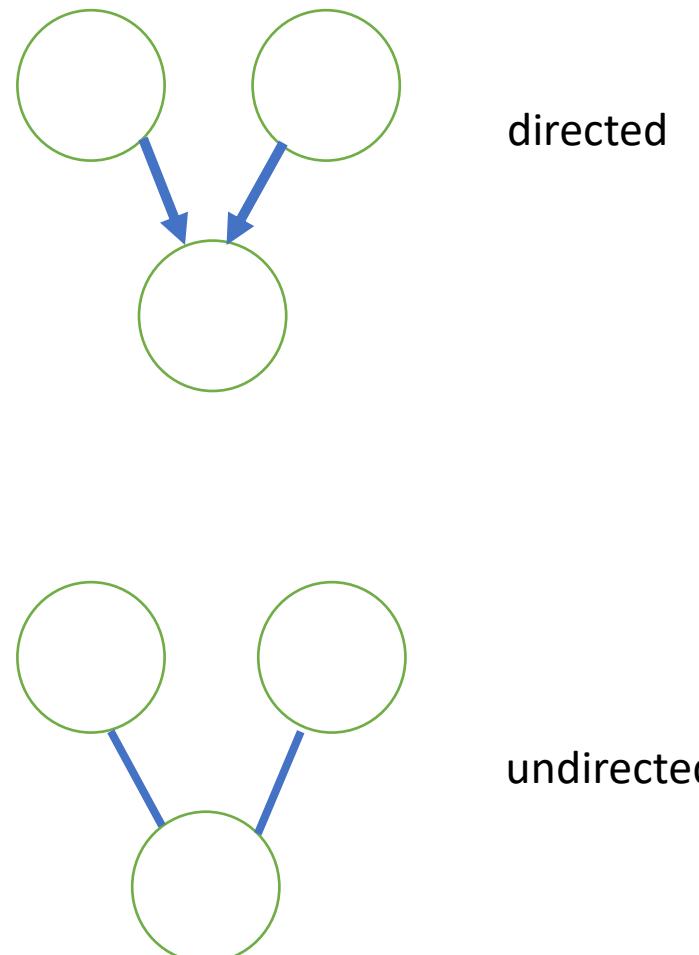
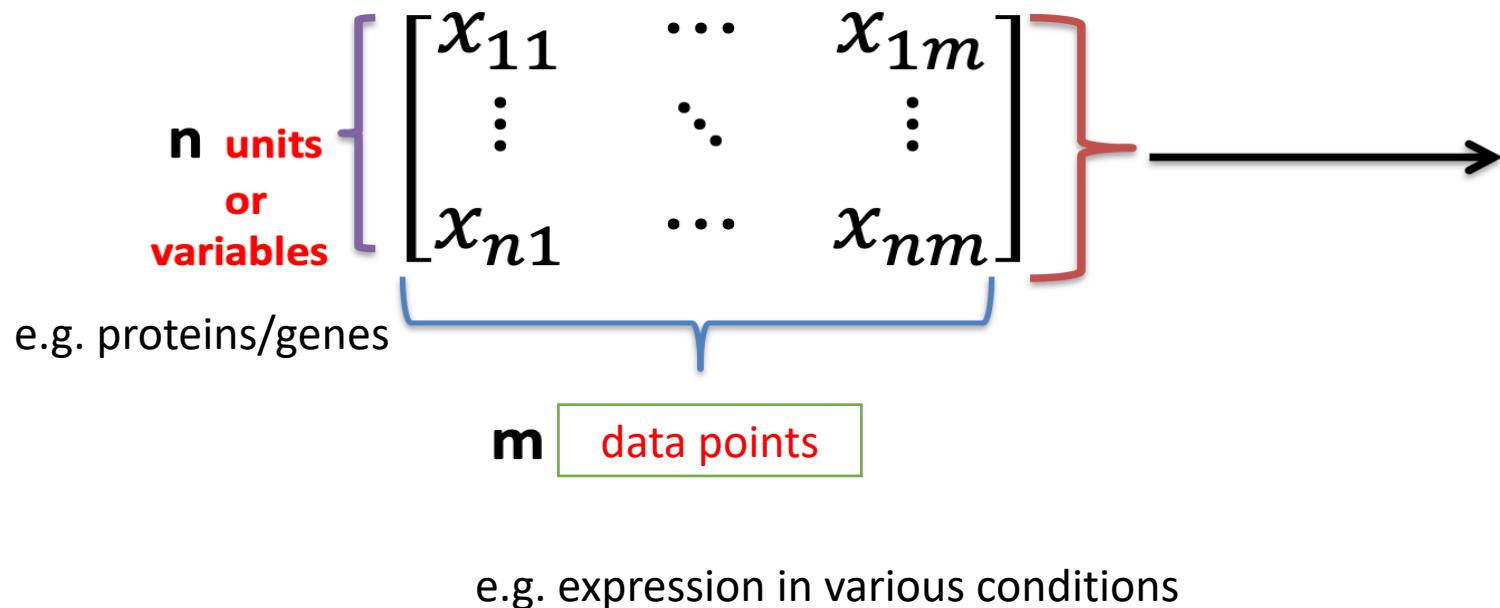
Analisis of a particular plant

- **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

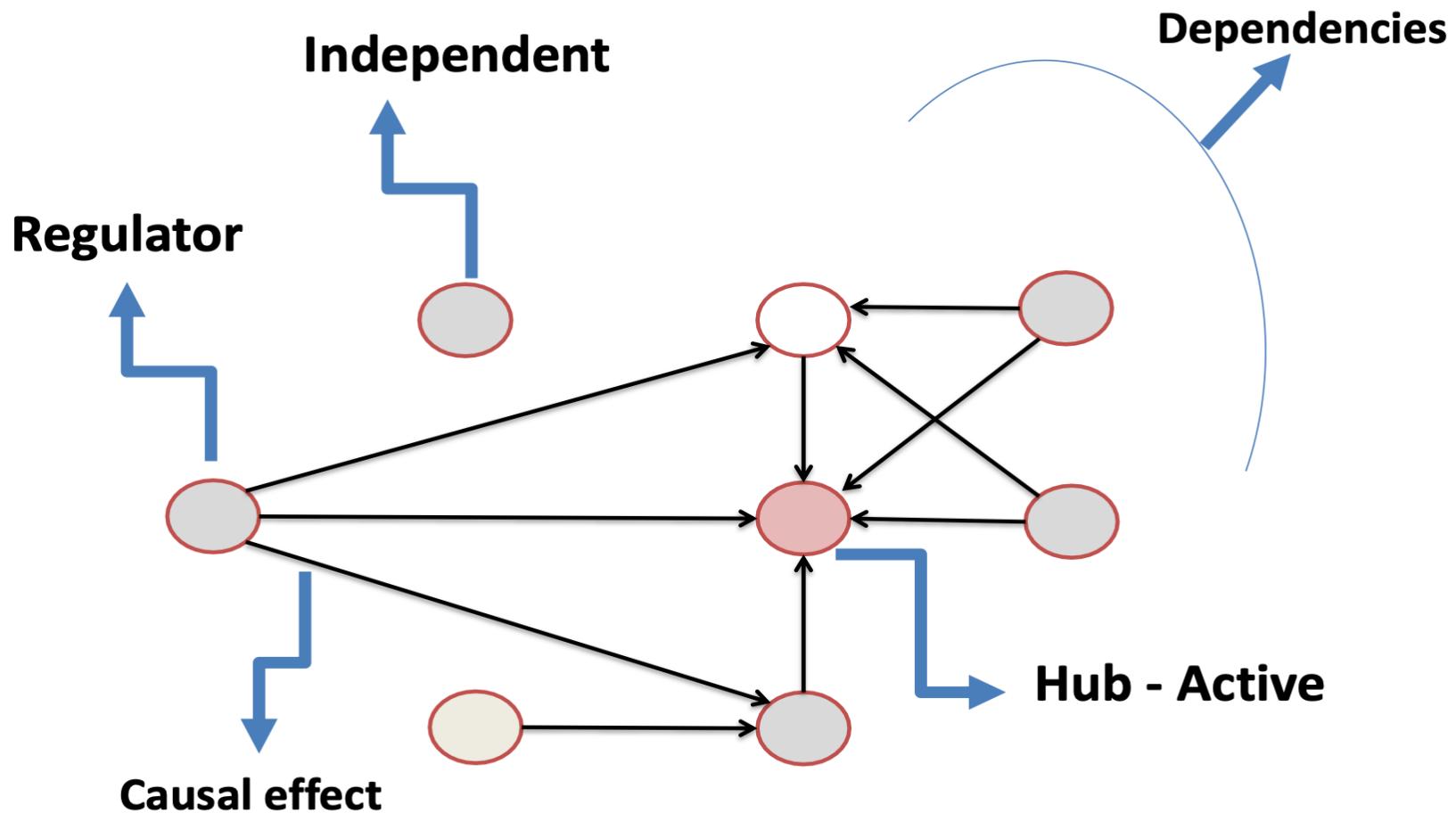
Learning the network structure from an n-by-m data matrix



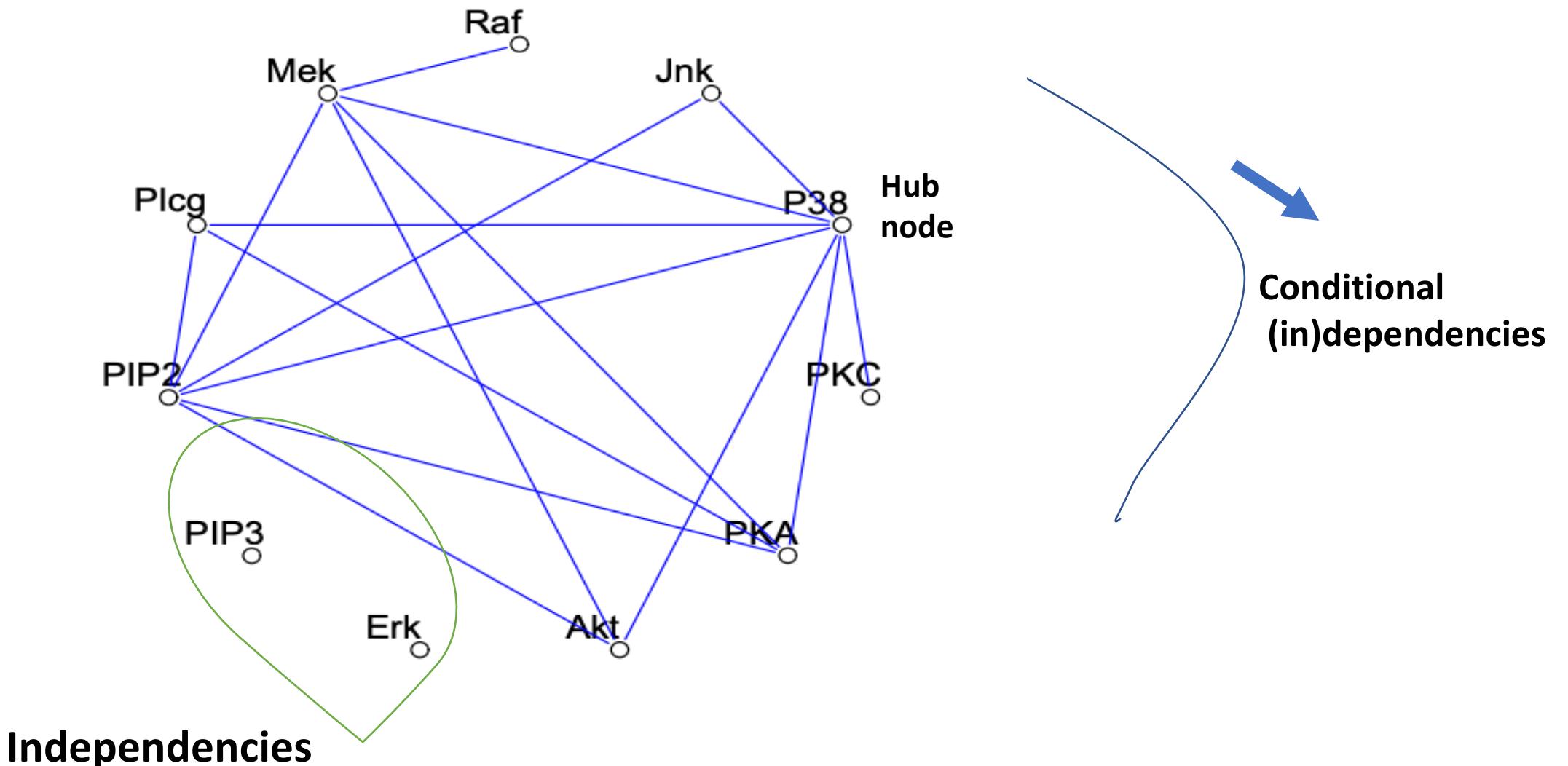
Motivation

- Discover the dependencies between components/nodes/variables/ phenomena from data and to 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.
- ...

Network



A flow cytometry dataset, **Protein-protein**



Note that this network can be sparser, we will see it later.

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.

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 (discussed morning)**
 - Continuous
 - Discrete
- **Types of directed Graphical models (discussed afternoon)**
 - Continuous
 - Discrete

Graphical models or probabilistic graphical models
(PGMs) or structured probabilistic models

Graphical Models

Michael Jordan, 1998

- A marriage between **probability** theory and **graph** theory.
- A probabilistic model for which a **graph** expresses the **conditional (in)dependence structure** between random variables.
- They are commonly used to specify the dependencies in a domain under study.
- A graph that represents relationships among a set of variables.

Graphical model

Michael Jordan, 1998

- Fundamental to the idea of the graphical model: The notion of **modularity**-- a complex system is built by combining simpler parts
- Probability theory **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.
- The graph theory provides
 - an intuitively appealing interface by which humans can **model highly-interacting sets of variables**
 - a **data structure** that lends itself naturally to the design of efficient general-purpose algorithm

Graphical Models Applications in Real Life

- Manufacturing
- Biology and system biology
- Medical Science
- Finance
- Economics
- Handwriting Recognition
- Telecommunication Network Diagnosis
- Object Recognition in Images
- ...

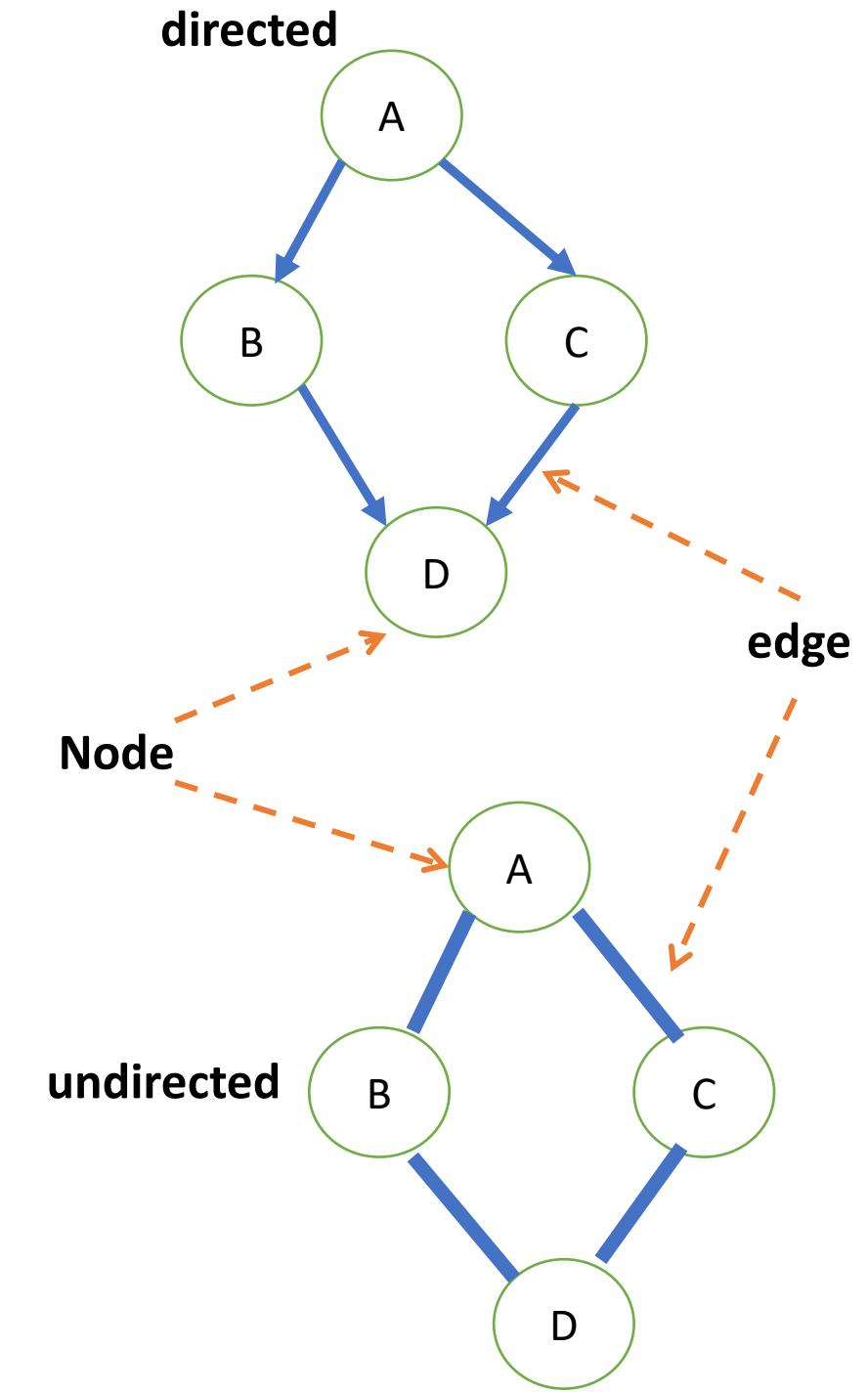
Graph

- We define a graph G by the following equation:

$$G = (V, E)$$

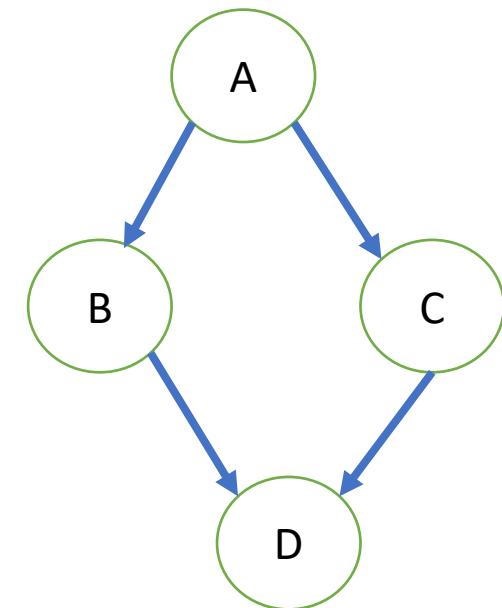
Here:

- 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.
- Conditional (in)dependence as (missing) edges.
- Graphical models : Encode conditional (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

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

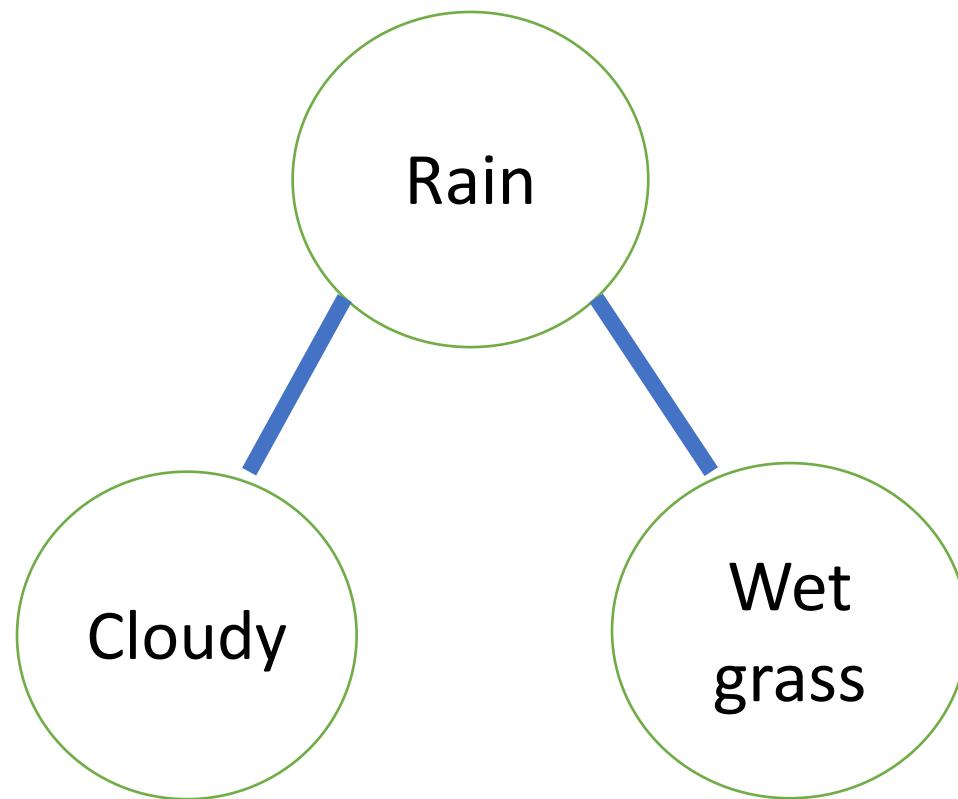
The occurrence of event A does not give any information about B and the occurrence of event B does not give any information about A.

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.

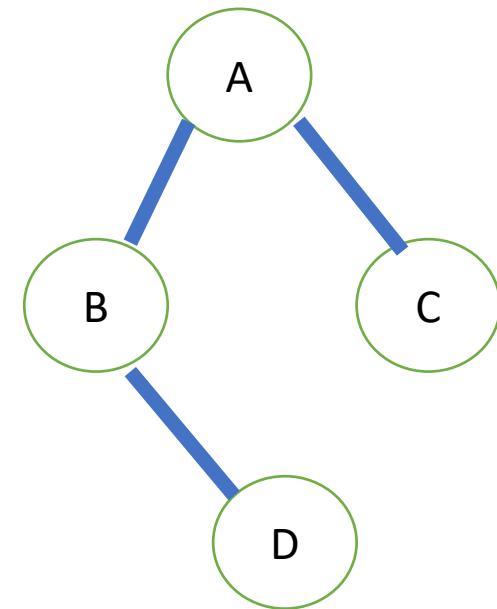
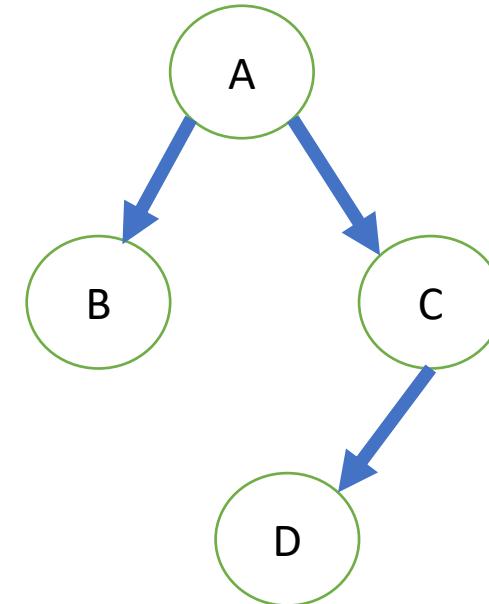
$$P(A,B|C)=P(A|C) P(B|C).$$

Conditional (in)dependency



Edge and lack of edge

- Edges show **conditional dependence.**
- Missing edges show **conditional independence.**



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).

They **differ** in the **set of independences** they can encode and the **factorization** of the distribution that they induce.

- **Combined**: chain graph

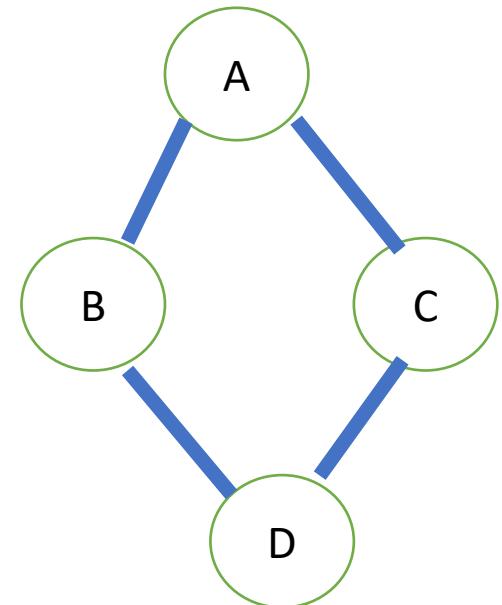
What are Undirected Graphical Models or Markov Random Fields?

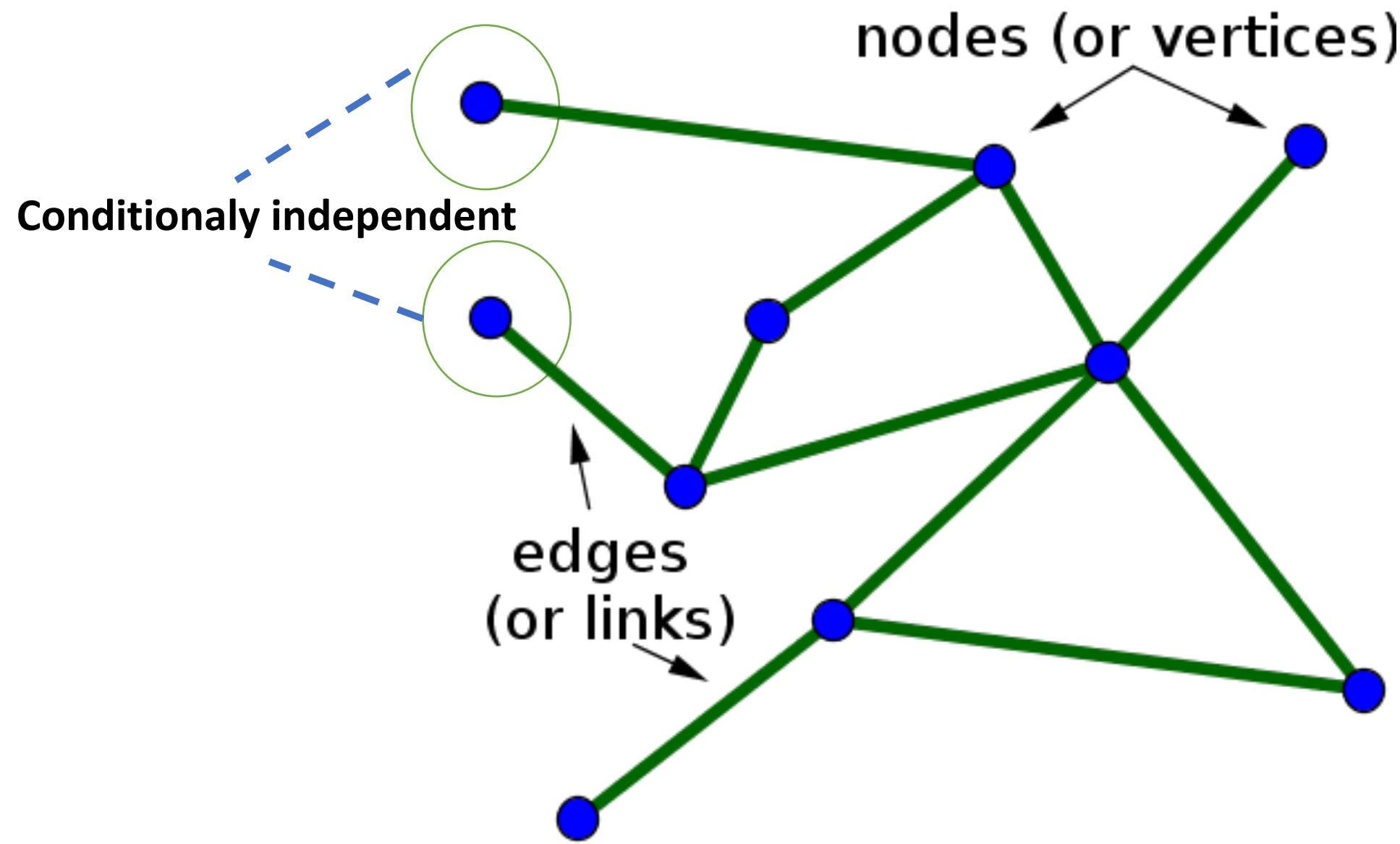
- Edges do not possess any form of orientation. They only show dependencies between nodes.
- Use undirected graphs where the **relationships between two variables, do not have a clear directionality.**
- The edge of (a,b) is identical to edge (b,a).

a ————— b

Conditional independency in a Markov Graph G

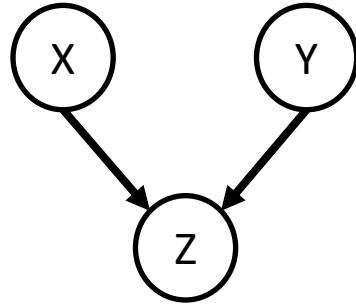
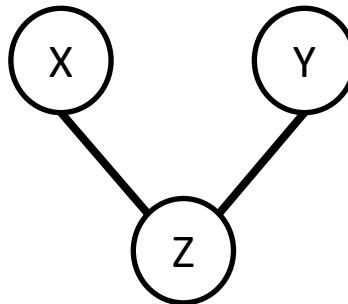
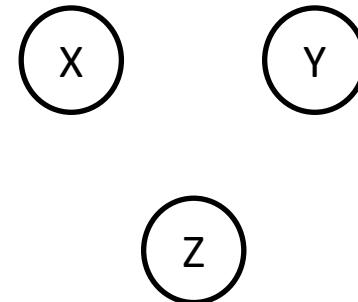
- The **absence of an edge** implies that the corresponding random variables are **conditionally independent** given the variables at the other vertices:
- **No edge joining X and Y $\iff X \perp Y \mid \text{rest}$**
(pairwise Markov independencies of G)
- “rest” refers to all of the other vertices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



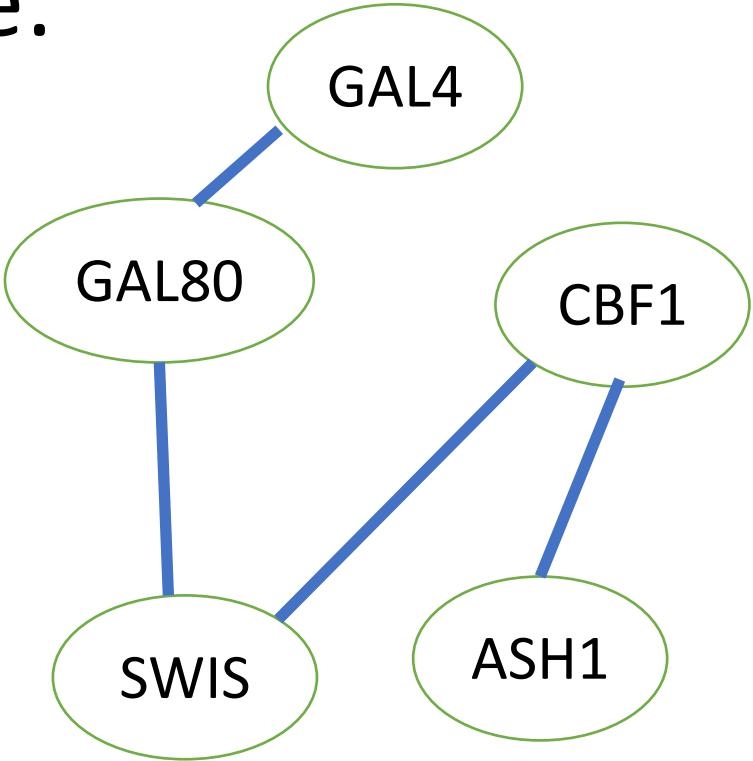


Undirected Graphical Models

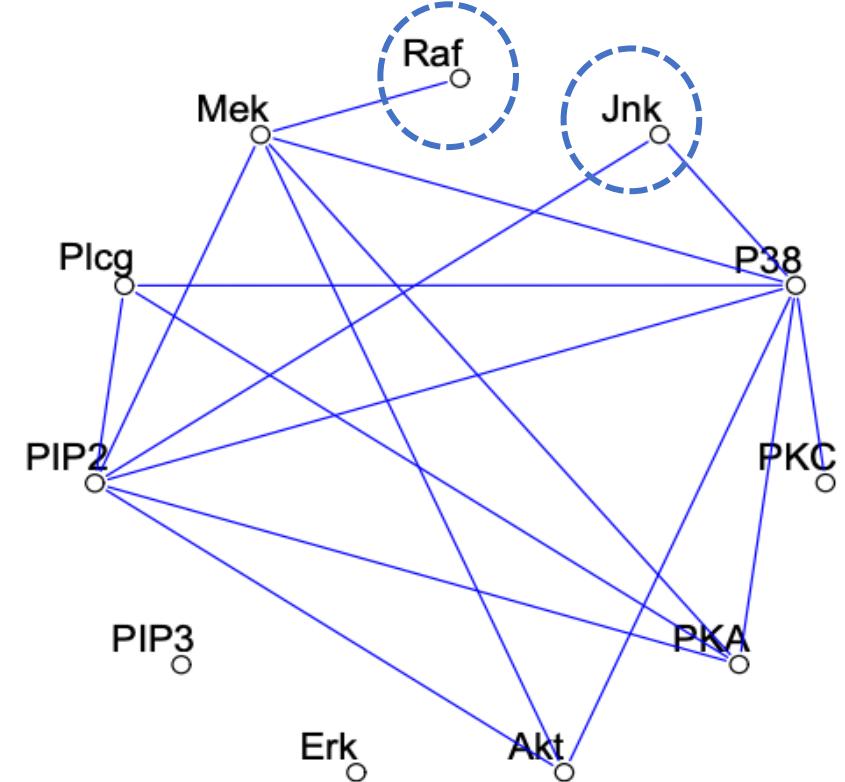
- Different relationships can be described with directed and undirected graphs.


$$\begin{aligned}x \perp\!\!\!\perp y \\ x \not\perp\!\!\!\perp y \mid z\end{aligned}$$

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

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

Example:



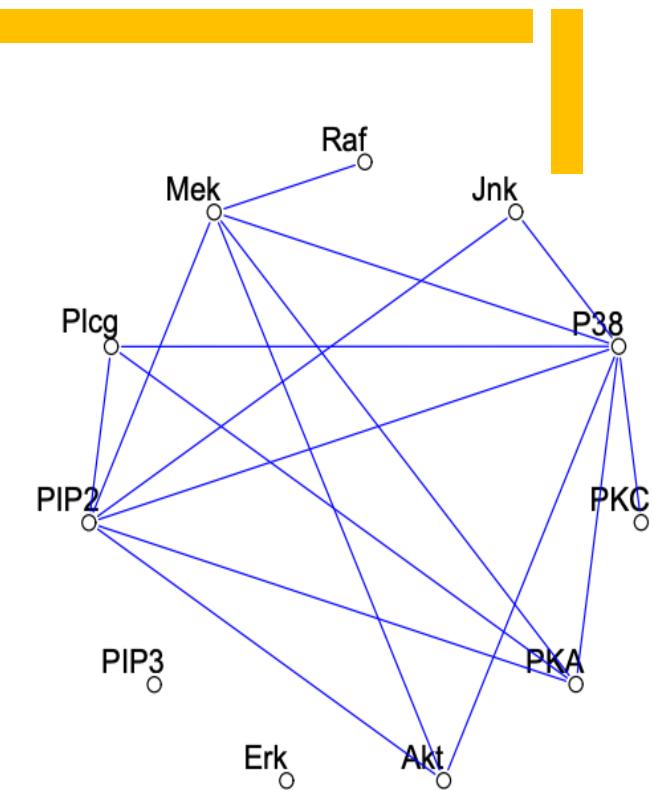
Estimated from yeast gene expression data



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

What can we do with graphical models

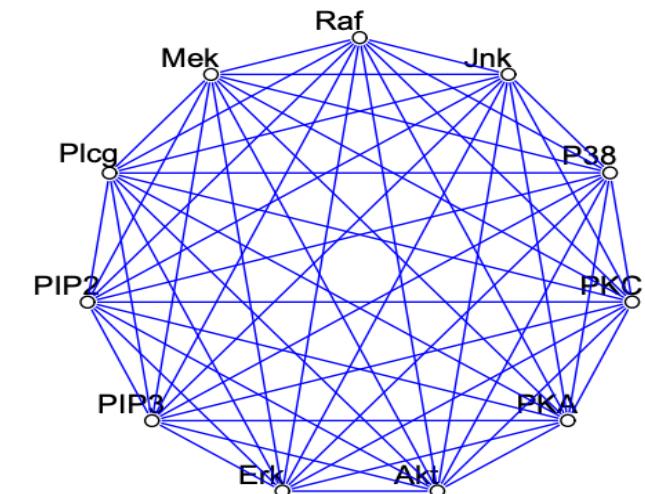
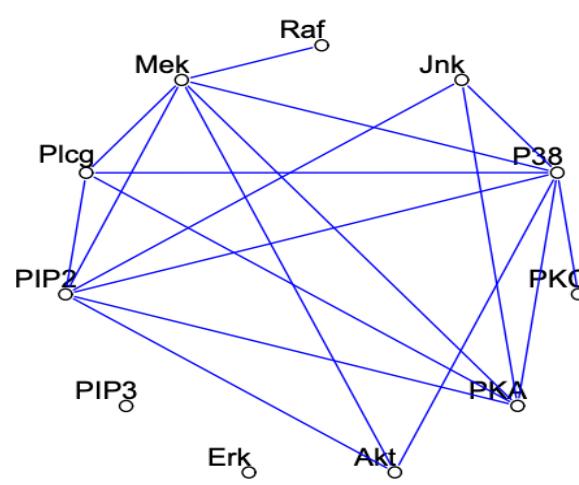
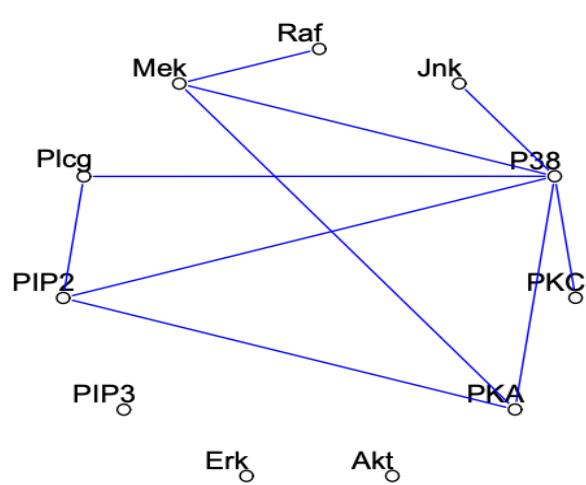
- Choosing the structure of the graph ←
- Estimation of the edge parameters from data
- Computation of marginal vertex probabilities and expectation

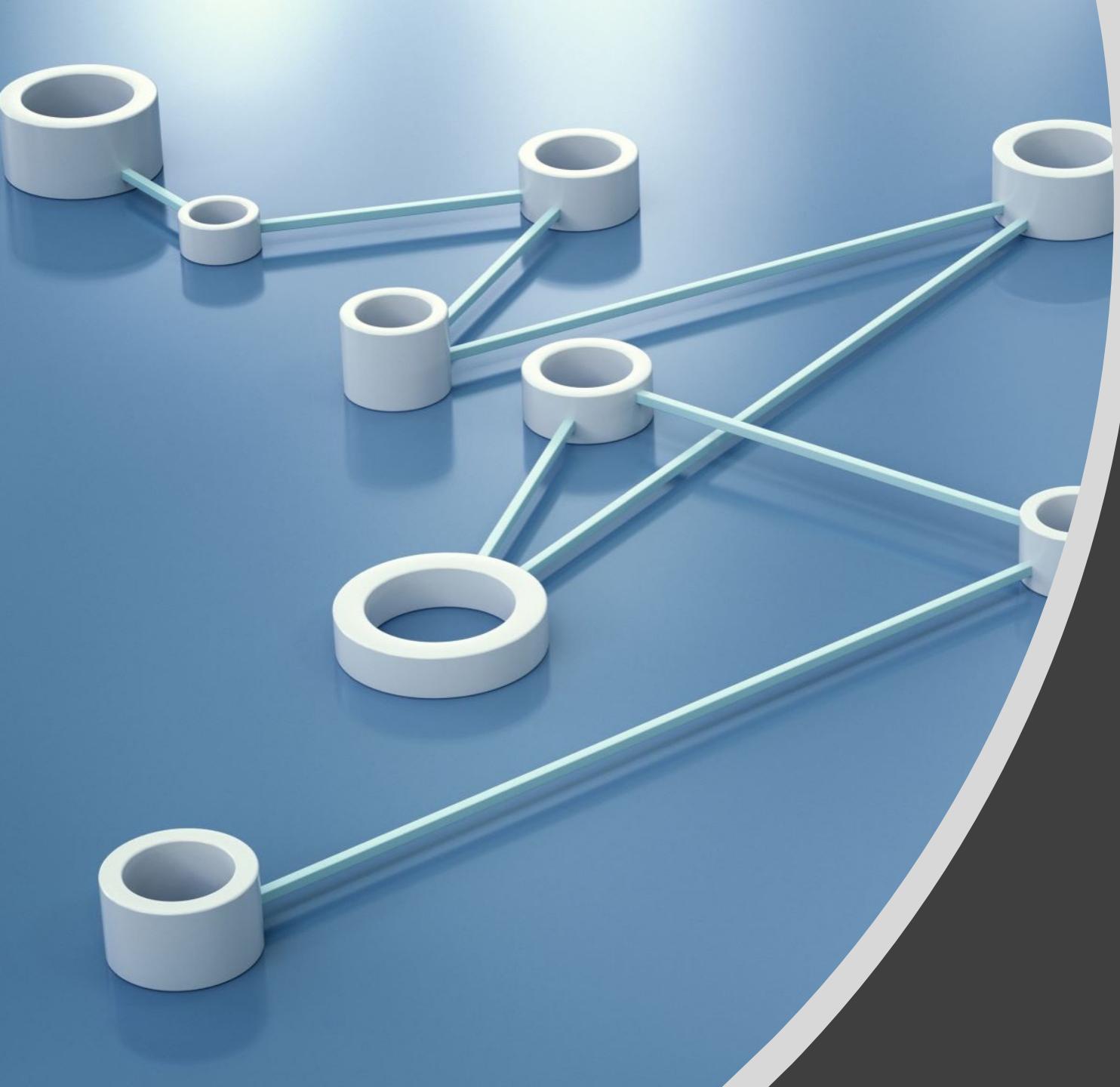


Hastie, T., Tibshirani, R., & Friedman, J. H. (2009)

Remark: Choosing the structure of the graph, Sparse graph

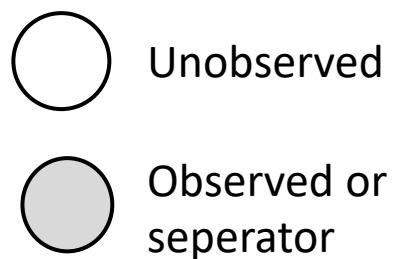
- Sparse graphs have a relatively small number of edges, and are convenient for interpretation.



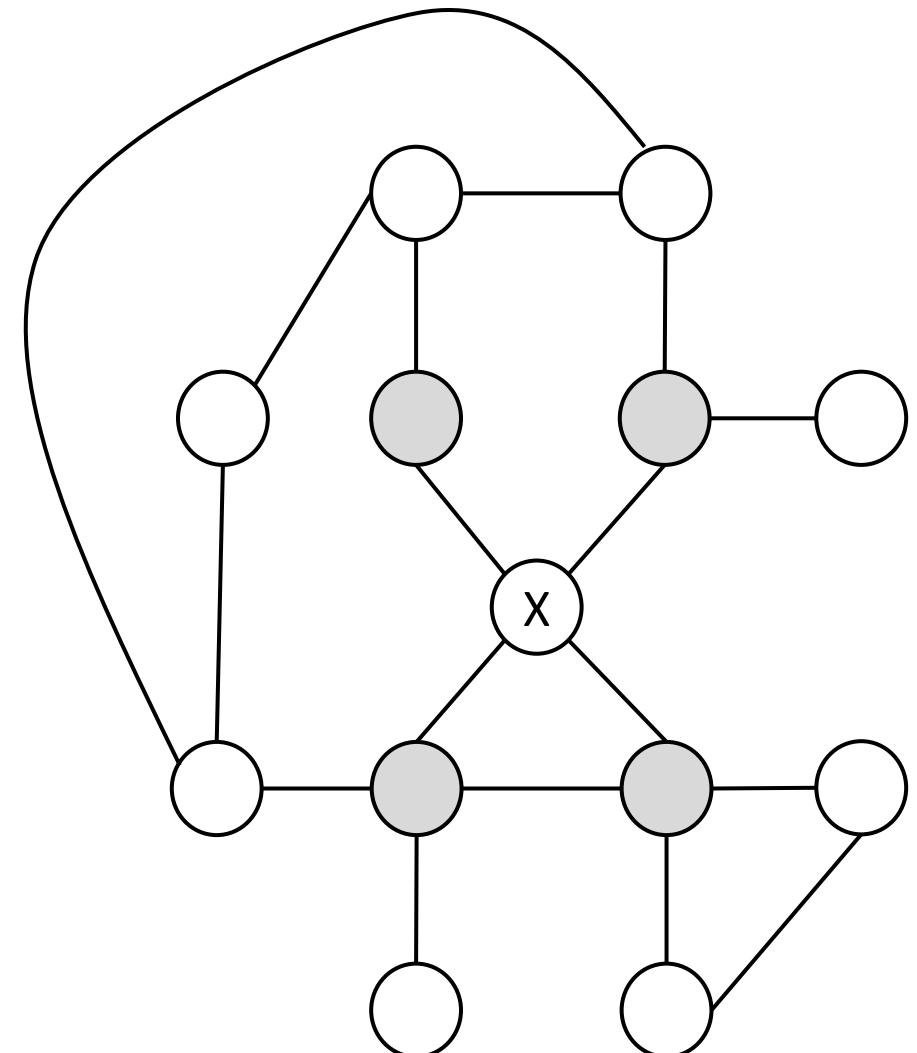


Some terminologies
useful for Network
interpretation

Marckov blanket/ Separators

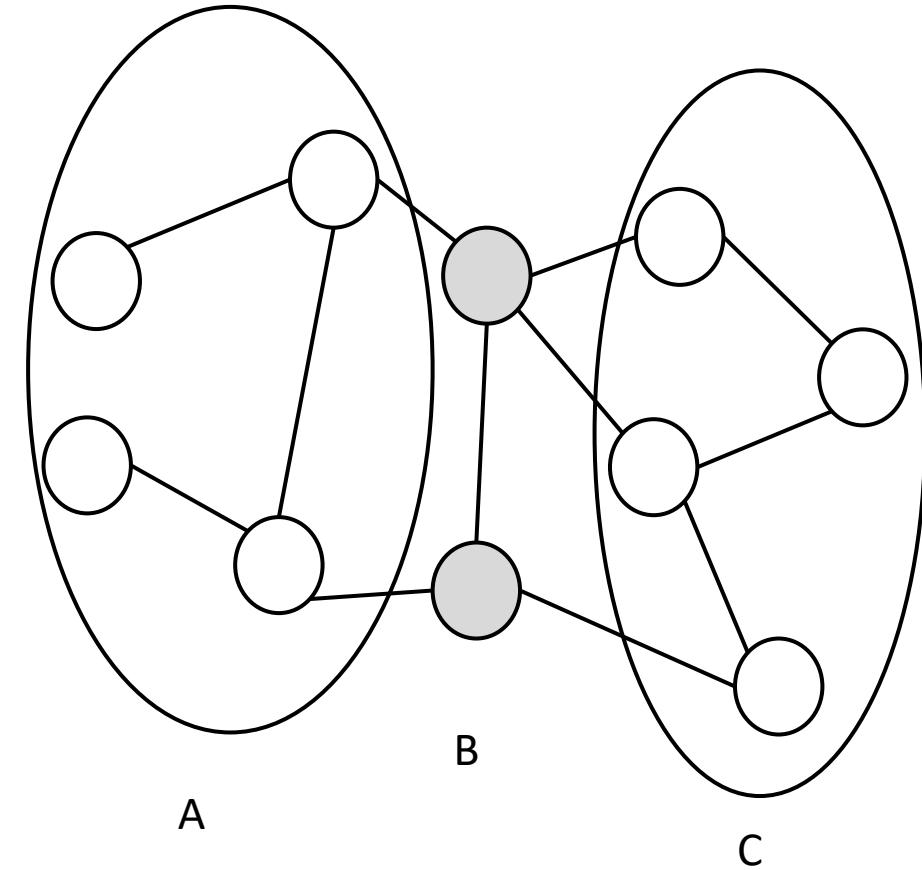


- Variables x, y are **dependent** if they are connected by a path of unobserved variables.
- If x 's **neighbours** are all observed, x is independent of all the others (they influence x only via its neighbours).
- In the undirected case the **Markov blanket** equals a node's neighbourhood.
- Useful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



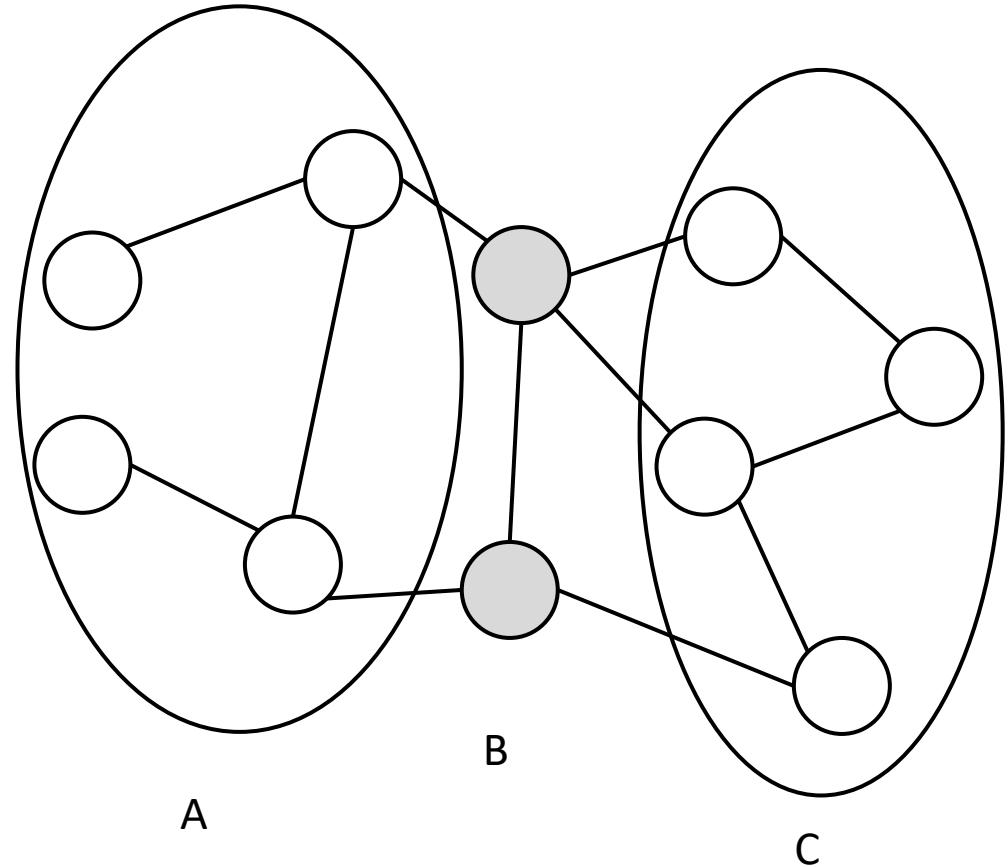
Separators and global Markov properties of G

- If a set of observed variables forms a cut-set between two halves of the graph, then variables in one half are independent from ones in the other.
- **B** is said to **separate** A and C if every path between A and C **intersects** a node in B, $A \perp C | B$.
- Separators break the graph into conditionally independent pieces.
- These are known as the **global Markov properties of G**.



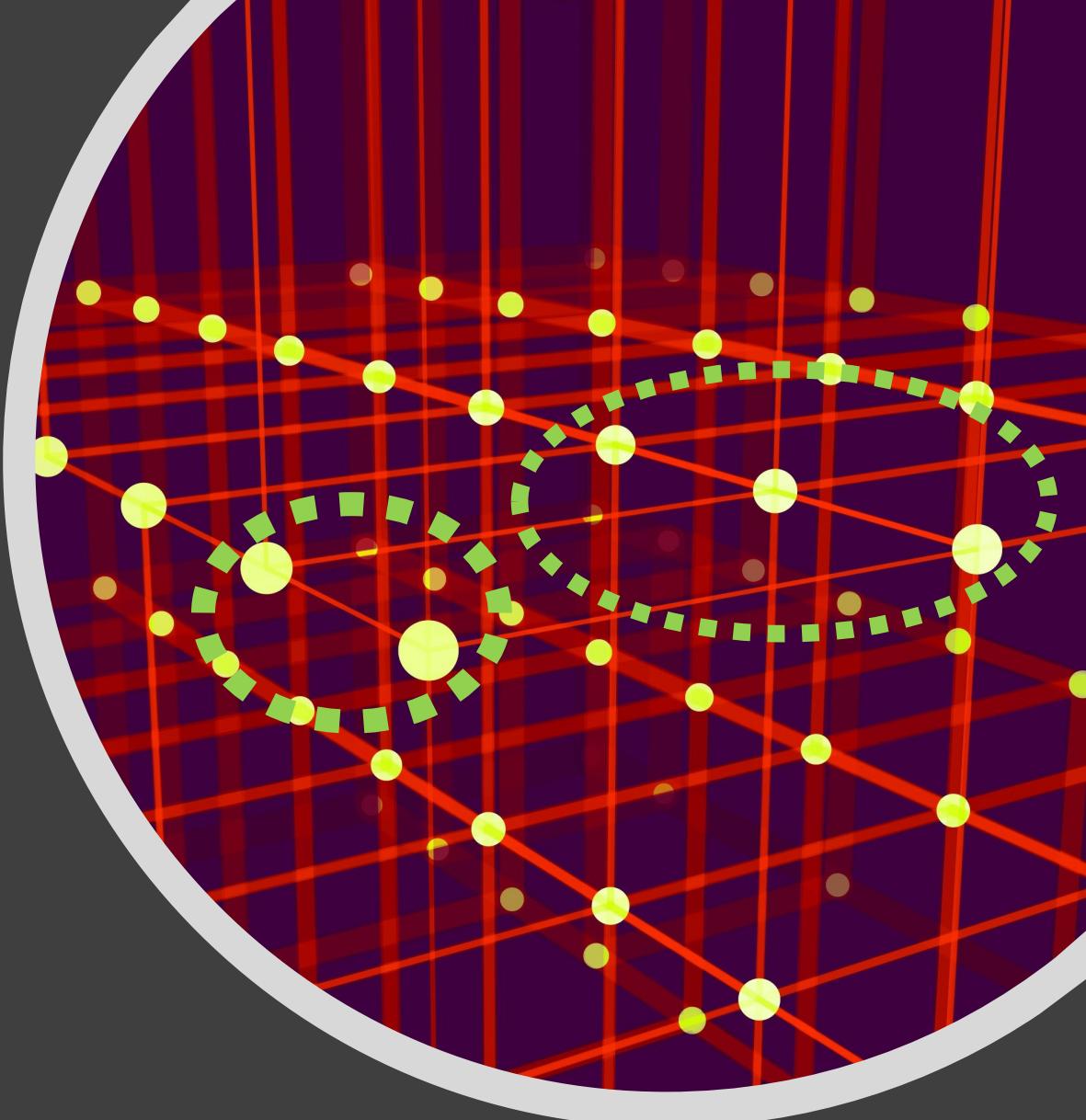
Remark:

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

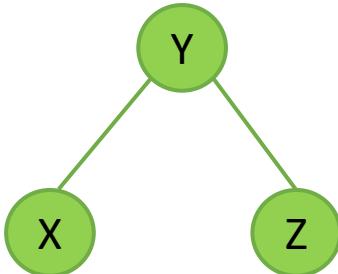


Clique

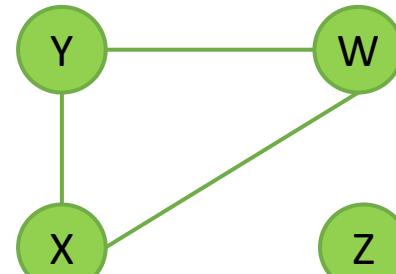
- Thus we separate the graph into cliques.
- A clique is a **complete** subgraph— a set of vertices that are all adjacent to one another.
- A subgraph is called **maximal** if it is **not a subgraph of a larger clique**, that is, a clique having more nodes.



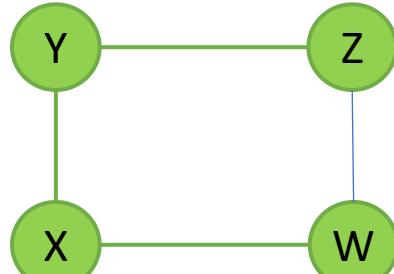
Examples



(a)



(b)



(c)



(d)

The maximal cliques:

(a) {X,Y}, {Y,Z} **(b)** {X,Y,W}, {Z} **(c)** {X,Y}, {Y,Z}, {Z,W}, {X,W}, and **(d)** {X,Y}, {Y,Z}, {Z,W}.

Remark:

- A commonly used class of Markov random fields are those that can be factorized according to the cliques of the graph.

Clique potential

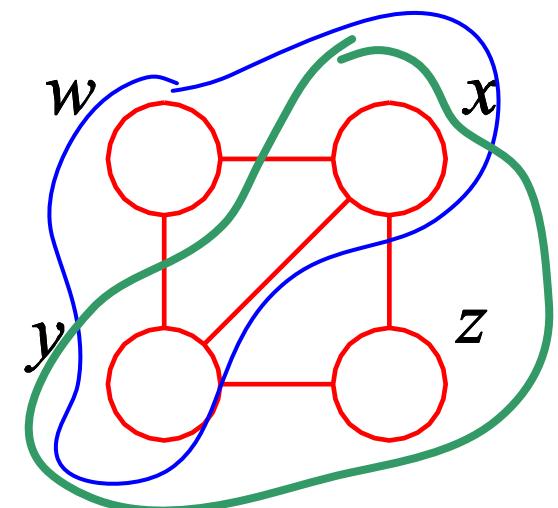
- Provided $p(x) > 0$ then joint distribution is product of **non-negative functions** over the *cliques* of the graph

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

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



Every MRF can be specified via clique potentials.

Remark:

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



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

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

- Pairwise Markov networks with binary 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

- The lasso penalty with binary pairwise Markov networks.
- 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 .

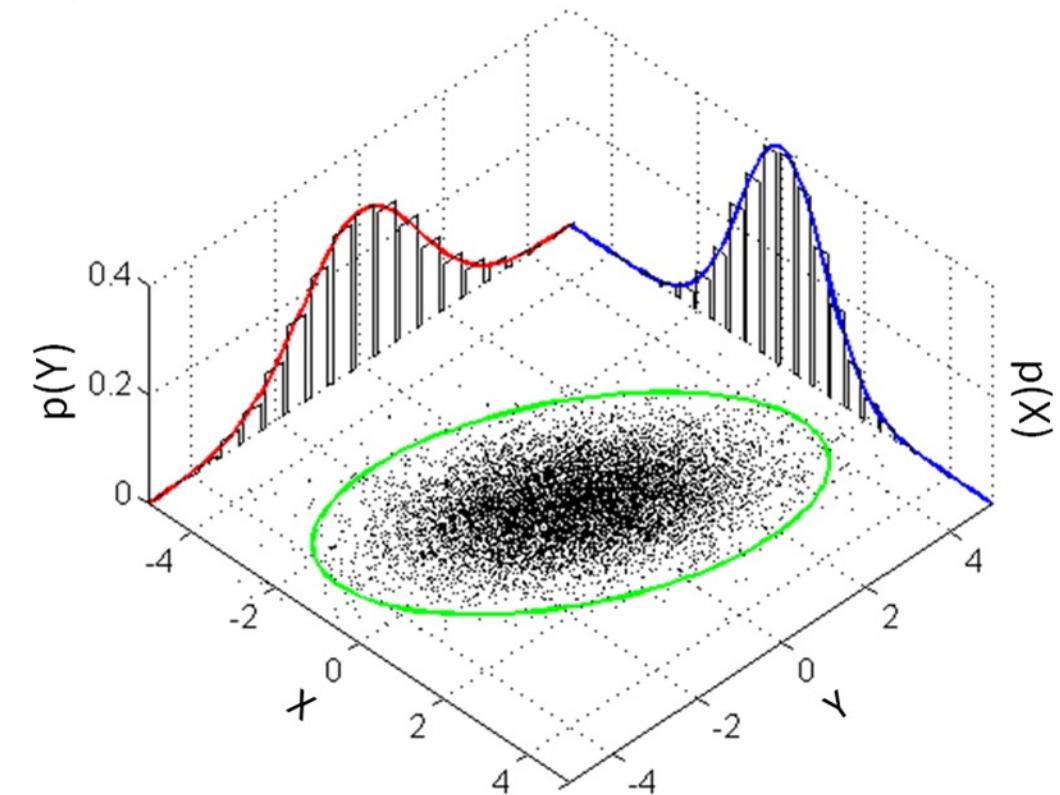
$$X \sim MN(\mu, \Sigma)$$

- Gaussian distribution represents second-order relationships 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.

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



$\boldsymbol{\mu}$: mean parameter

$\boldsymbol{\Sigma}$: covariance matrix



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

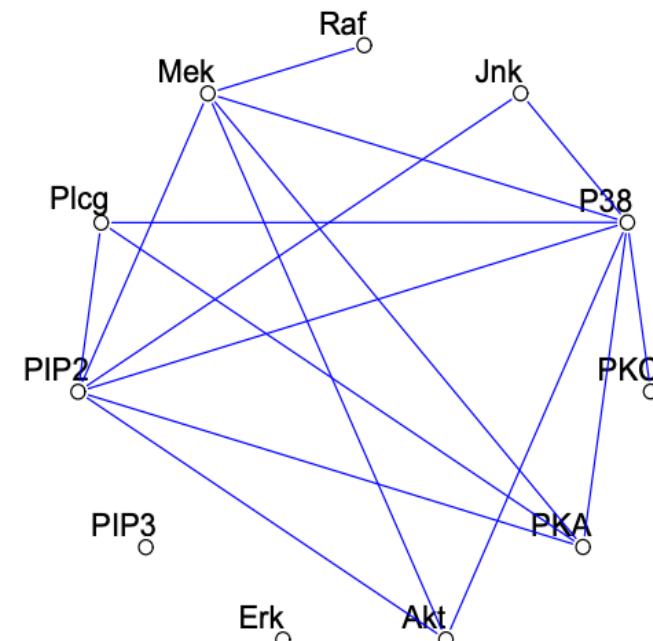
- Encodes all information how variables relate to one another.
- Encodes all conditional relationships

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

A flow cytometry dataset, Protein-protein

| Raf | Mek | Plcg | PIP2 | PIP3 | Erk | Akt | PKA |
|------|------|------|------|------|------|------|-----|
| 26.4 | 13.2 | 8.82 | 18.3 | 58.8 | 6.61 | 17 | 414 |
| 35.9 | 16.5 | 12.3 | 16.8 | 8.13 | 18.6 | 32.5 | 352 |
| 59.4 | 44.1 | 14.6 | 10.2 | 13 | 14.9 | 32.5 | 403 |
| 73 | 82.8 | 23.1 | 13.5 | 1.29 | 5.83 | 11.8 | 528 |
| 33.7 | 19.8 | 5.19 | 9.73 | 24.8 | 21.1 | 46.1 | 305 |
| 18.8 | 3.75 | 17.6 | 22.1 | 10.9 | 11.9 | 25.7 | 610 |
| 44.9 | 36.5 | 10.4 | 132 | 16.3 | 8.66 | 17.9 | 835 |
| 47.4 | 15 | 14.6 | 30.5 | 17.5 | 20.2 | 45.3 | 466 |
| 104 | 61.5 | 10.6 | 21.1 | 41.8 | 11.5 | 23.5 | 445 |

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



```
> 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 (*precision matrix*) Σ^{-1}

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

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

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

- contains information about the **partial covariances** between the variables (the covariances between pairs i and j, **conditioned** on all other variables).
- If the ij^{th} component of $\Theta = \Sigma^{-1}$ is zero, then variables i and j are conditionally independent, given the other variables.
- Thus Θ captures all the second-order information.

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.

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

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

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 | 0.198584117 |
| Jnk | 0.066953405 | -0.083976539 | 0.0621401887 | -0.011266330 | 0.064667726 | -0.041046519 | 0.041640264 | -0.0659514543 | -0.284095198 | 0.198584117 | 1.000000000 |

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.

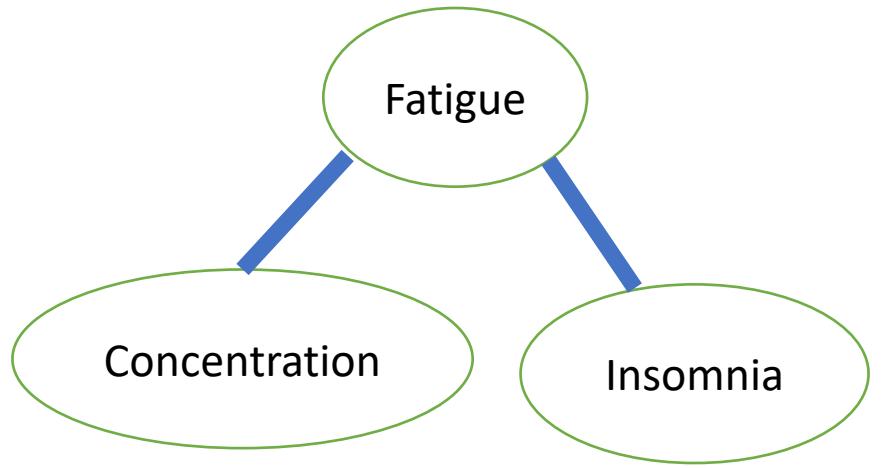
Zero mean there is not an edge.

Non-zero means there is an edge

Undirected Graphical Models for Continuous Variables

- The maximum likelihood estimate of Σ is simply S .
- S is the empirical covariance matrix.
- If some of the edges are missing; for the Gaussian distribution this implies that the corresponding entries of $\Theta = \Sigma^{-1} = S^{-1}$ are zero.
- Standardize Θ

Partial correlation (standardization)



Partial correlations:

$$\begin{bmatrix} 1 & -0.25 & 0 \\ -0.25 & 1 & 0.30 \\ 0 & 0.30 & 1 \end{bmatrix}$$

How to estimate the Network structure?

- In most cases we do not know which edges to omit from our graph, and would like to discover this from the data. It can be done by either

Trying to fully estimate Σ or $\Theta = \Sigma^{-1}$: a network with a lot of edges, **difficult for interpretation.**

Or

Estimating which components of θ_{ij} are nonzero, using of e.g. **L1 (lasso)** regularization (**Graphical lasso**), **sparse network**, **easy interpretation**, **A big advantage.**



Graphical Lasso, a tool for estimating the network structure

R package is glasso

Friedman et al. (2008)

- To estimate **an inverse covariance** matrix by 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 using each variable as the response and the others as predictors.
-
- The component θ_{ij} is estimated to be **nonzero** if the estimated coefficient of variable i on j is nonzero, (or the estimated coefficient of variable j on i is nonzero).
- This procedure consistently estimates the set of **nonzero elements of $\Theta = \Sigma^{-1}$** and **shrink low correlations to 0**.
- Maximizing the penalized log-likelihood

The graphical lasso

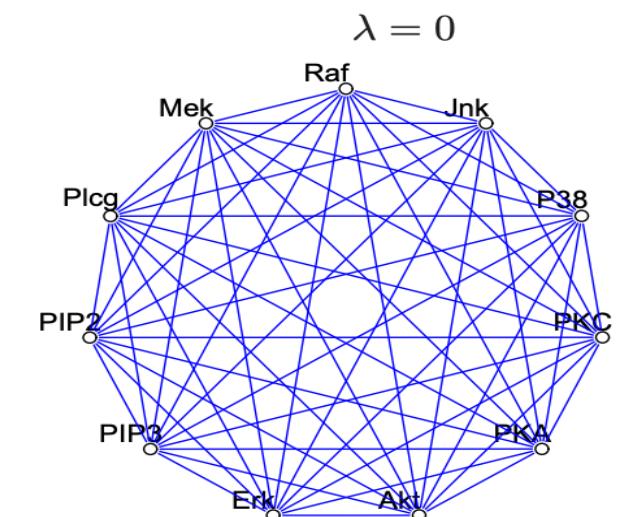
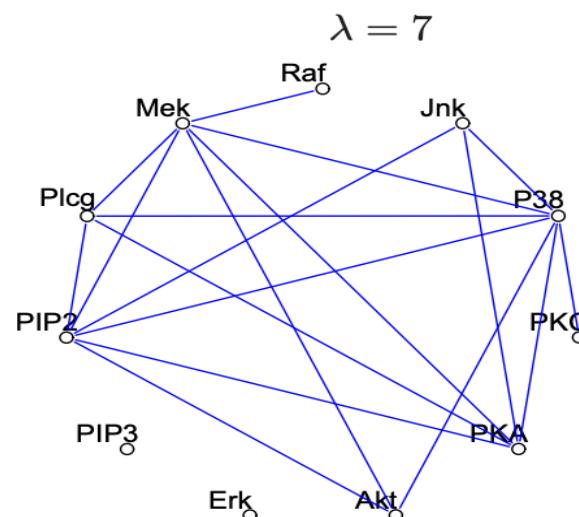
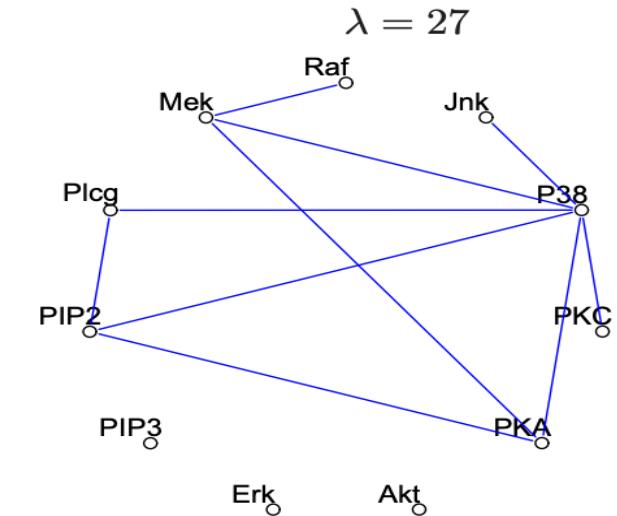
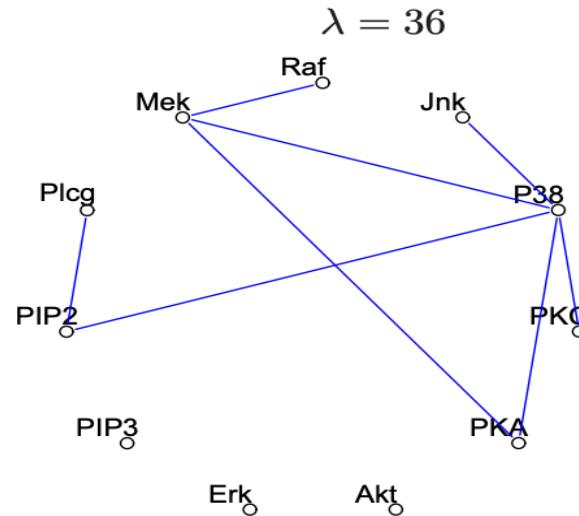
- The R package “**glasso**” is popular, fast, and allows one to efficiently build a path of models for different values of the tuning parameter.
- There are many other extensions that use **glasso** packages.
- In the practical we focus on **glasso** and **EBICglasso**.
- **Note:** graphical lasso can be also done in **Python** in scikit-learn
- <https://scikit-learn.org/stable/modules/generated/sklearn.covariance.GraphicalLasso.html#sklearn.covariance.GraphicalLasso>.

Graphical lasso - Penalty parameter

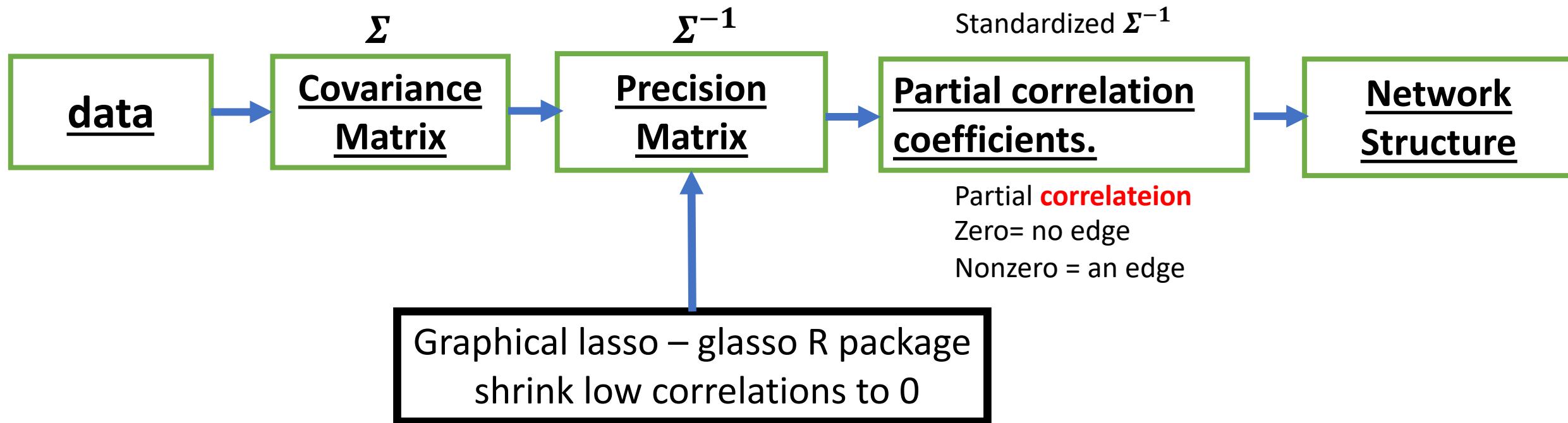
Friedman et al. (2008)

Four different graphical lasso solutions
for the flow-cytometry data.

The graph becomes more sparse as the
penalty parameter is increased.



In a nutshell: Network structure



Advantages of MRFs:

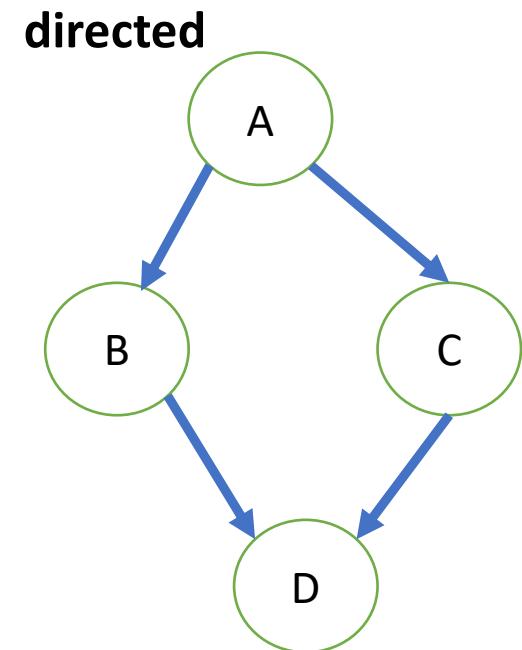
- **MRF** can be applied to a **wider** range of problems in which there is no natural directionality associated with variable dependencies.
 - **MRFs** succinctly express **certain dependencies** (cycle) that BN **cannot** easily describe (the converse is also true).
 - Conditional independence is easier to read off for MRFs.
 - **Without arrows, there is only one type of neighbors (simpler Markov blanket).**
-

Disadvantages of MRFs:

- Undirected models may be difficult to interpret.
- It is much easier to generate data from a **Bayesian network**, which is important in some real world applications.

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 Directed Acyclic Graph(DAG).



Directed graphs

- **Directed graphical models** represent probability distributions that can be **factored into products of conditional distributions**, and have the potential for causal interpretations.
- **Bayesian networks** are a type of Directed Acyclic Graphs.

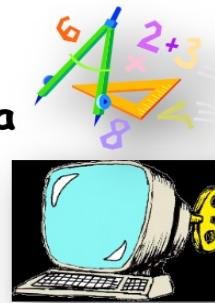
In a nutshell



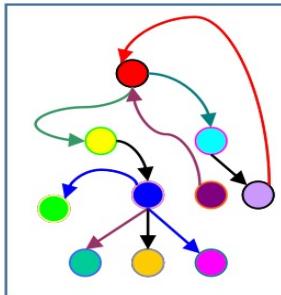
Raw data



Cleaned data



Machine Learning



Statistical Methods

Network inference

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>