CSE8803/CX4803

Machine Learning in Computational Biology

Lecture 16: Graphical Models

Yunan Luo

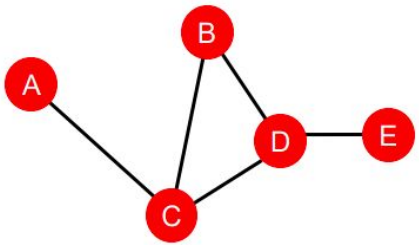
Slides credit: Xiuwei Zhang, Yunan Luo

Node centrality: betweenness centrality

- A node is important if it lies on many shortest paths between other nodes

$$c_v = \sum_{s \neq v \neq t} \frac{\#(\text{shortest paths betwen } s \text{ and } t \text{ that contain } v)}{\#(\text{shortest paths between } s \text{ and } t)}$$

- **Unnormalized** version:



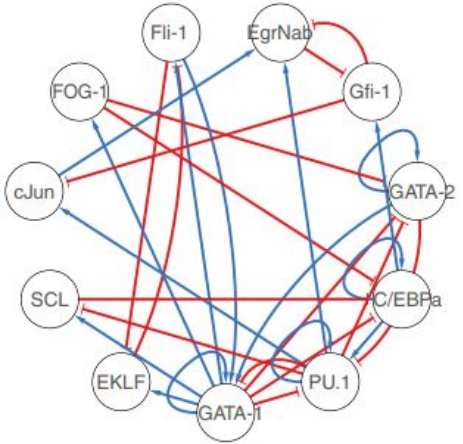
$$c_A = c_B = c_E = 0$$
$$c_C = 3$$
(A-C-B, A-C-D, A-C-D-E)

$$c_D = 3$$
(A-C-D-E, B-D-E, C-D-E)

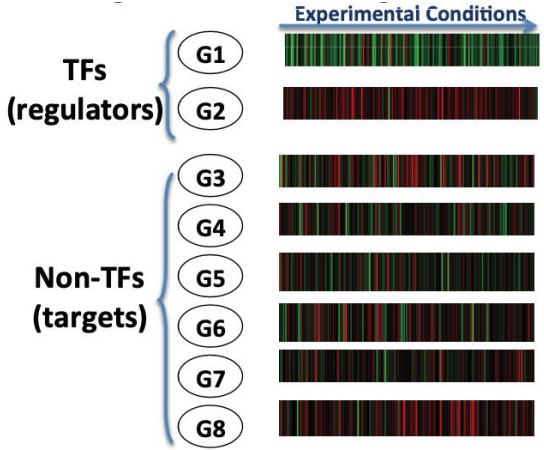
- **Normalized** version:
  - Undirected graph
    - Normalized by (N-1)(N-2)/2
  - Directed graph
    - Normalized by (N-1)(N-2)
  - E.g.:

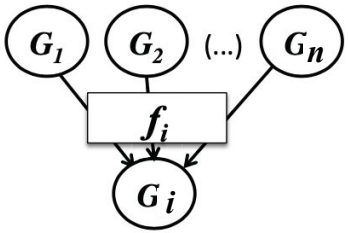
$$c_D = \frac{3}{\binom{N-1}{2}} = 0.5$$

Computationally infer regulatory networks from gene-expression data



Pratapa et al, *Nature Methods*, 2020





Gene expression prediction:

$$G_i = f_i(G_j)$$

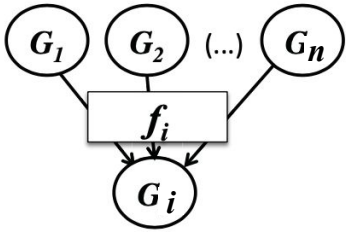
$$j \in \{1, \ldots, n\} - i$$

Slide credit: Soheil Feizi

Modeling Gene Regulatory Networks

- Regression Model;
- Bayesian Networks;
- Boolean Networks;
- Differential Equations;

Regression models for GRN inference



Gene expression prediction:

$$G_i = f_i(G_j)$$
$$j \in \{1, \ldots, n\} - i$$

$f_i()$ is often assumed to be linear.

$G_i = \alpha_{1i} G_1 + \alpha_{2i} G_2 + \alpha_{3i} G_3 + \ldots + \alpha_{ni} G_n + \varepsilon$

We need to learn the α parameters.

Objective:

Find parameters that minimize "residual sum of squares" between observed (G_i) and predicted expression levels G_i'.

$$RSS = \sum_{t=1}^{m} \sum_{i=1}^{n} (G_i - G_i')^2$$

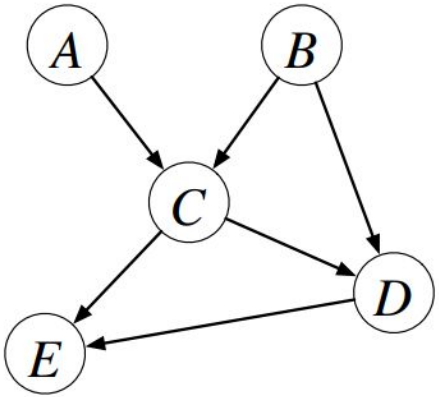
- n: number of genes; m: number of conditions/cells
- Can add regularization term to force small weights to be 0

Haury et al. TIGRESS: Trustful Inference of Gene REgulation using Stability Selection. *BMC Syst. Biol.* **6**, 145 (2012)

Modeling Gene Regulatory Networks

- Regression Model;
- Bayesian Networks;
- Boolean Networks;
- Differential Equations;

Representing knowledge through probabilistic graphical models (PGM)



Nodes: random variable

Edges: statistical dependencies between the variables

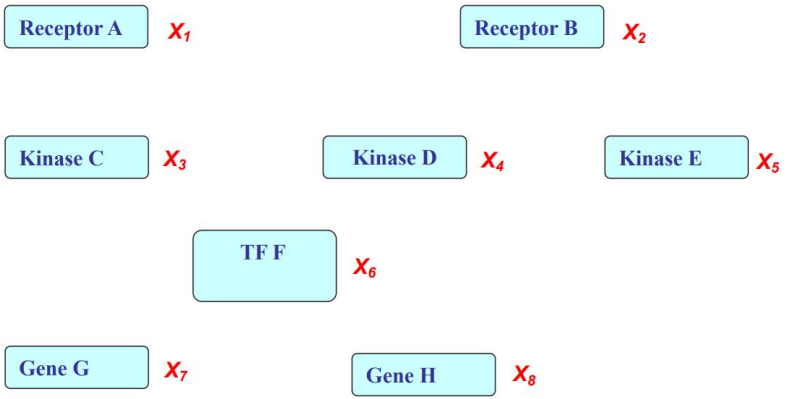
Slide credit: Zoubin Ghahramani

Representation of joint probability distribution

- Representation: the joint probability distribution on multiple **binary** variables?

$$P(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8)$$

- State configurations in total
  - 2^8



cellular signal transduction

- Are they all needed to be represented?
- Do we get any scientific/medical insight?

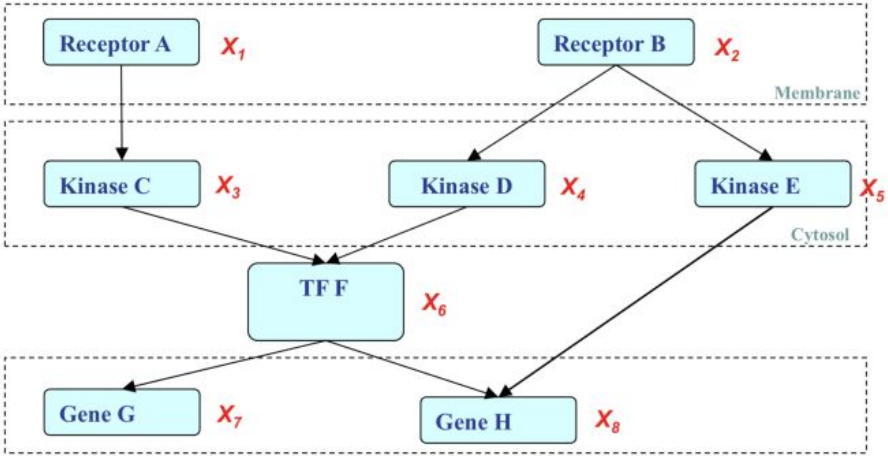
Slide credit: Eric Xing

Graphical Model: Structure Simplifies Representation

- If Xi's are conditionally independent (as described by a GM), the joint can be factored to a product of simpler terms, e.g.,

$$P(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8)$$

$$= P(X_1)\, P(X_2)\, P(X_3|\, X_1)\, P(X_4|\, X_2)\, P(X_5|\, X_2)$$
$$P(X_6|\, X_3, X_4)\, P(X_7|\, X_6)\, P(X_8|\, X_5, X_6)$$



- State configurations in total
  - 2+2+4+4+4+8+4+8=36
  - 8x reduction from 2^8 in representation cost!

Slide credit: Eric Xing

Why do we need GMs

- Graphs are an **intuitive** way of representing and visualising the relationships between many variables. (Examples: family trees, electric circuit diagrams, neural networks)
- A graph allows us to abstract out the **conditional independence** relationships between the variables from the details of their parametric forms. Thus we can answer questions like: “Is A dependent on B given that we know the value of C?” just by looking at the graph
- Graphical models allow us to define general **message-passing algorithms** that implement probabilistic inference efficiently. Thus we can answer queries like “What is p(A|C = c)?” without enumerating all settings of all variables in the model.

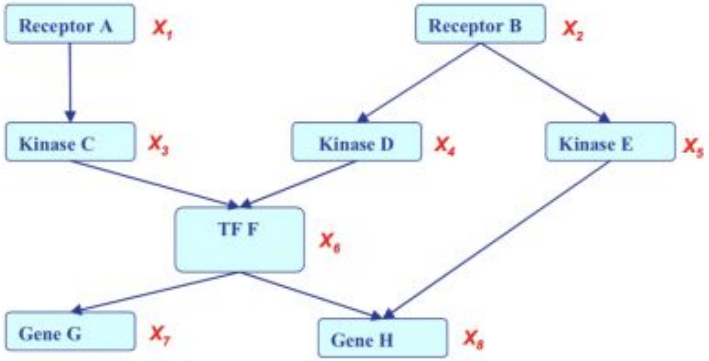
Graphical models = statistics × graph theory × computer science

Slide credit: Zoubin Ghahramani

Two types of GMs

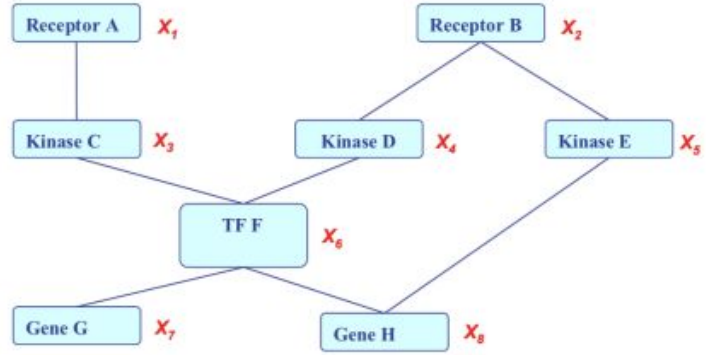
- Directed edges give causality relationships
  - Also called **Bayesian Network** or **Directed Graphical Model**

$$P(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8)$$
$$= P(X_1)\,P(X_2)\,P(X_3|X_1)\,P(X_4|X_2)\,P(X_5|X_2)\,P(X_6|X_3, X_4)\,P(X_7|X_6)\,P(X_8|X_5, X_6)$$



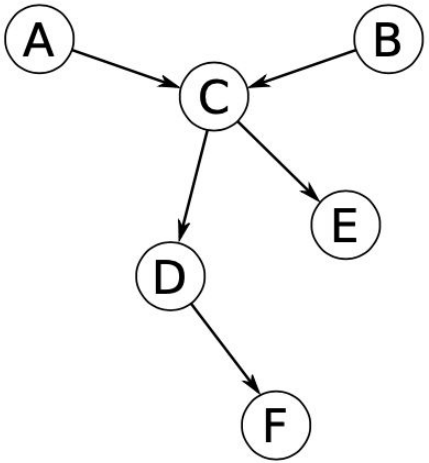
- Undirected edges simply give correlations between variables
  - Also called **Markov Random Field** or **Undirected Graphical Model**

$$P(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8)$$
$$= 1/Z \exp\{E(X_1)+E(X_2)+E(X_3, X_1)+E(X_4, X_2)+E(X_5, X_2) + E(X_6, X_3, X_4)+E(X_7, X_6)+E(X_8, X_5, X_6)\}$$



Bayesian networks vs GRNs

Directed acyclic graph (DAG) that represents dependency between variables.

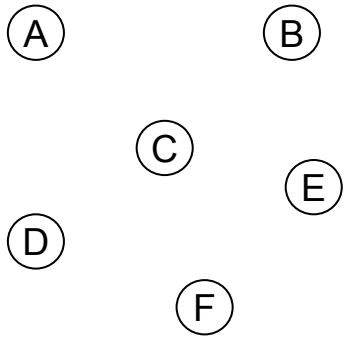


Bayesian net vs. GRN:

node → gene
edge → regulatory interaction

GRN inference:

Structure learning of Bayesian nets.



Bayesian networks

Structure learning vs Parameter learning

What are the parameters when the structure is given?

First, we focus on parameter learning

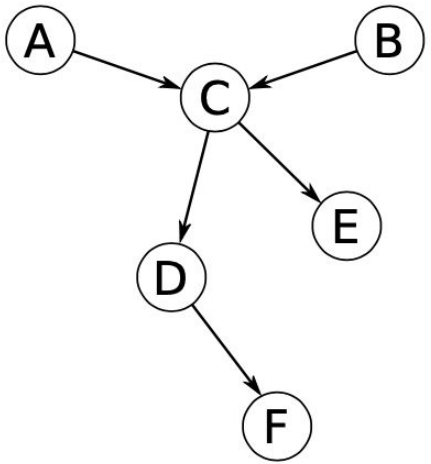
Bayesian networks

Structure learning vs Parameter learning

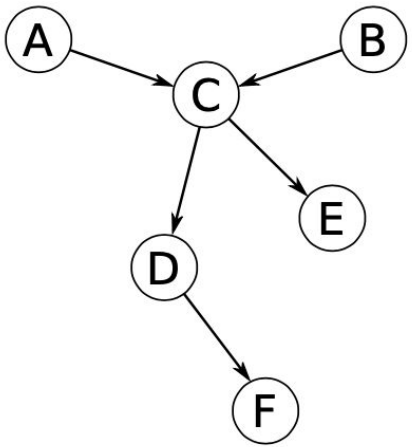
What are the parameters when the structure is given?

First, we focus on parameter learning

The probability function of each node, depending on other nodes in the graph



Conditional independence in BN



Graphical separation

Probabilistic independence

$$
\begin{aligned}
&A \perp\!\!\!\perp_G B \\
&A \perp\!\!\!\perp_G D \mid C \\
&B \perp\!\!\!\perp_G D \mid C \\
&A \perp\!\!\!\perp_G E \mid C \\
&B \perp\!\!\!\perp_G E \mid C \\
&D \perp\!\!\!\perp_G E \mid C \\
&C \perp\!\!\!\perp_G F \mid D \\
&\dots
\end{aligned}
\qquad \Longrightarrow \qquad
\begin{aligned}
&A \perp\!\!\!\perp_P B \\
&A \perp\!\!\!\perp_P D \mid C \\
&B \perp\!\!\!\perp_P D \mid C \\
&A \perp\!\!\!\perp_P E \mid C \\
&B \perp\!\!\!\perp_P E \mid C \\
&D \perp\!\!\!\perp_P E \mid C \\
&C \perp\!\!\!\perp_P F \mid D \\
&\dots
\end{aligned}
$$

Slide credit: Marco Scutari

Probability Toolbox Review

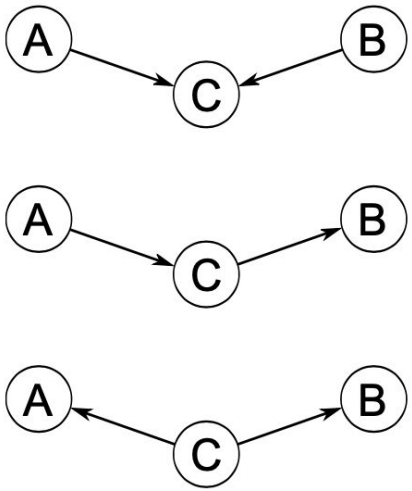
- Definition of conditional probability

$$P(A \mid B) = \frac{P(A, B)}{P(B)}$$

- Product rule

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

Conditional independence in BN



Case 1:
P(A,B,C) = P(C|A,B)P(A,B)
= P(C|A,B)P(A)P(B)

Case 2:
P(A,B,C) = P(B,C|A)P(A)
= P(B|A,C)P(C|A)P(A)
= P(B|C)P(C|A)P(A)

Case 3:
P(A,B,C) = P(A,B|C)P(C)
= P(A|C)P(B|C)P(C)

Example: proof of conditional independence of A, B in case 3

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

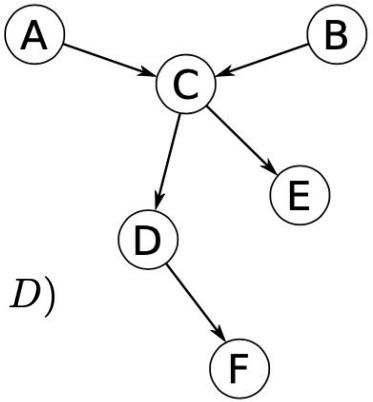
Conditional independence in BN

Each node is dependent only on its parents

Define:

$\mathrm{P}(\mathbf{X})$ - the global probability of all variables

$\mathrm{P}(\mathbf{X}) = \mathrm{P}(\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{E}, \mathrm{F})$



$$\mathrm{P}(\mathbf{X}) = \mathrm{P}(A)\,\mathrm{P}(B)\,\mathrm{P}(C \mid A, B)\,\mathrm{P}(D \mid C)\,\mathrm{P}(E \mid C)\,\mathrm{P}(F \mid D)$$

In general:

$$p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i | X_{\mathsf{pa}(i)})$$

where pa(i) is the parents of node i

Each variable is conditionally independent of its non-descendants, given its parents.

Slide credit: Marco Scutari

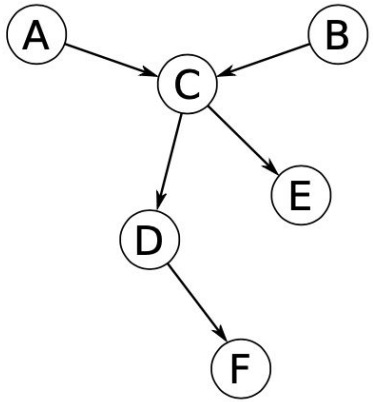
Conditional independence in BN

$$\mathrm{P}(\mathbf{X}) = \mathrm{P}(A)\,\mathrm{P}(B)\,\mathrm{P}(C \mid A, B)\,\mathrm{P}(D \mid C)\,\mathrm{P}(E \mid C)\,\mathrm{P}(F \mid D)$$

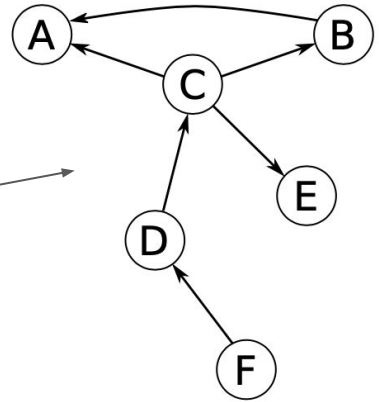
Does a factorization of $\mathrm{P}(\mathbf{X})$ uniquely define a DAG?

No.

Bayes Theorem: $P(A \mid B) = \dfrac{P(B \mid A) \cdot P(A)}{P(B)}$



$$\begin{aligned}\mathrm{P}(\mathbf{X}) &= \mathrm{P}(A)\,\mathrm{P}(B)\,\mathrm{P}(C \mid A, B)\,\mathrm{P}(D \mid C)\,\mathrm{P}(E \mid C)\,\mathrm{P}(F \mid D) \\ &= \mathrm{P}(A \mid B, C)\,\mathrm{P}(B \mid C)\,\mathrm{P}(C \mid D)\,\mathrm{P}(D \mid F)\,\mathrm{P}(E \mid C)\,\mathrm{P}(F)\end{aligned}$$



Slide credit: Marco Scutari

Bayesian networks

According to what type of values are used for the variables:

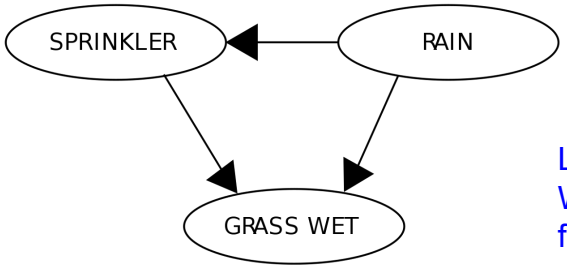
- Discrete BN
- Continuous BN (Gaussian BN)

Discrete BN: conditional probability tables (CPT)

Joint probability function: $\Pr(G, S, R) = \Pr(G \mid S, R)\Pr(S \mid R)\Pr(R)$

$\Pr(S \mid R)$

RAIN	SPRINKLER T	SPRINKLER F
F	0.4	0.6
T	0.01	0.99



RAIN T	RAIN F
0.2	0.8

$\Pr(R)$

SPRINKLER	RAIN	GRASS WET T	GRASS WET F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

Learn parameters:
When all variables are observed, simply count frequencies

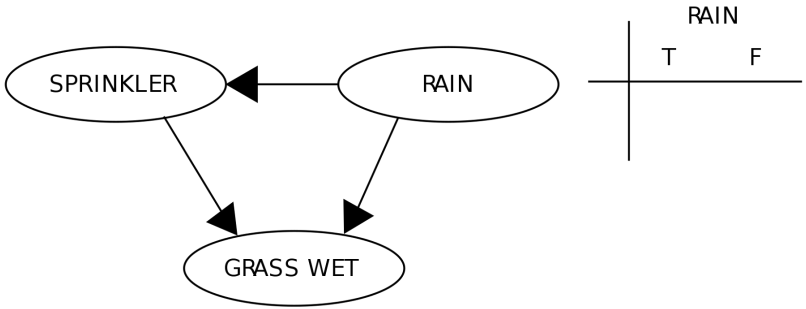
Sprinkler	Rain	GrassW
F	T	T
F	T	T
F	T	F
F	T	T

Discrete BN: conditional probability tables (CPT)

Joint probability function: $\Pr(G, S, R) = \Pr(G \mid S, R)\Pr(S \mid R)\Pr(R)$

$\Pr(S \mid R)$

RAIN	SPRINKLER T	SPRINKLER F
F		
T		



$\Pr(R)$

RAIN T	RAIN F

Sprinkler	Rain	GrassW
F	T	T
F	T	T
F	T	F
F	F	T
T	F	T
T	T	F
T	T	T

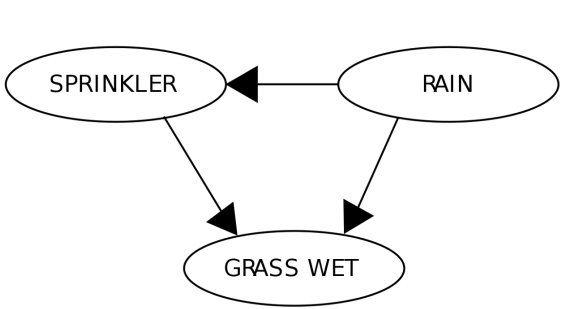
Pr (Rain=T) = 5/7
Pr (Rain=F) = 2/7

Discrete BN: conditional probability tables (CPT)

Joint probability function: $\Pr(G, S, R) = \Pr(G \mid S, R)\Pr(S \mid R)\Pr(R)$

$\Pr(S \mid R)$

RAIN	SPRINKLER T	SPRINKLER F
F		
T		



RAIN T	RAIN F

$\Pr(R)$

Sprinkler	Rain	GrassW
F	T	T
F	T	T
F	T	F
F	F	T
T	F	T
T	T	F
T	T	T

Joint probability:
Pr (S=F, R=T) = 3/7
Pr (S=F, R=F) = 1/7
Pr (S=T, R=T) = 2/7
Pr (S=T, R=F) = 1/7

conditional probability:
Pr (S=F | R=T) = 3/5
Pr (S=T | R=T) = 2/5
...

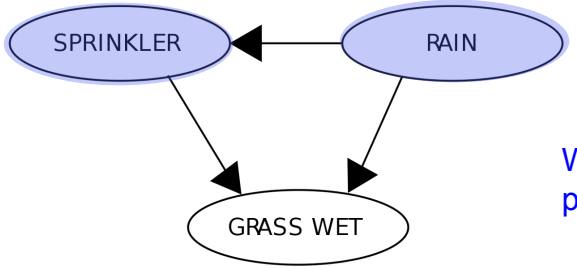
Pr(S=F | R=T) = Pr(S=F, R=T)/Pr(R=T)
=3/7 / (5/7) = 3/5

Discrete BN: conditional probability tables (CPT)

Joint probability function: $\Pr(G, S, R) = \Pr(G \mid S, R)\Pr(S \mid R)\Pr(R)$

$\Pr(S \mid R)$

RAIN	SPRINKLER T	SPRINKLER F
F	0.4	0.6
T	0.01	0.99



RAIN T	RAIN F
0.2	0.8

$\Pr(R)$

SPRINKLER	RAIN	GRASS WET T	GRASS WET F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

When there are hidden variables: inference problem

$$\Pr(R = T \mid G = T) = \frac{\Pr(G = T, R = T)}{\Pr(G = T)}$$

$$= \frac{\sum_{x \in \{T,F\}} \Pr(G = T, S = x, R = T)}{\sum_{x,y \in \{T,F\}} \Pr(G = T, S = x, R = y)}$$

Bayesian networks vs HMM

Probabilistic graphical models

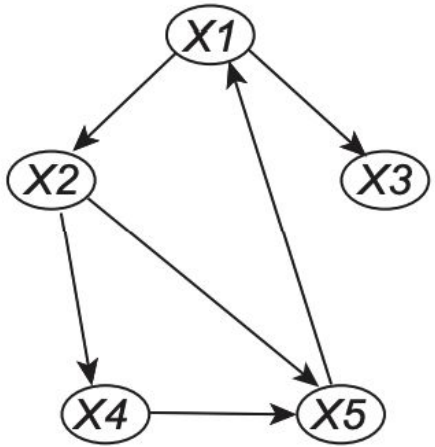
HMM: linear structure

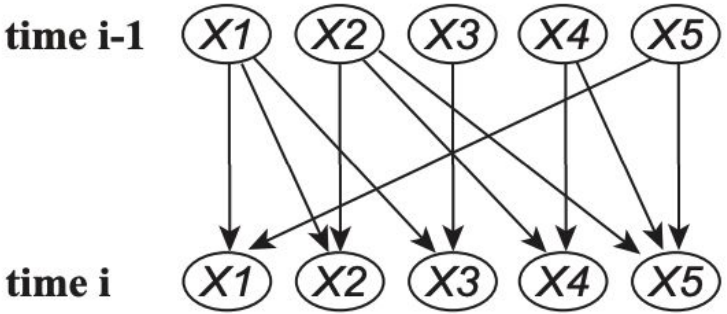
Bayesian network: graph structure

“Unroll” BN along time: dynamic Bayesian network

Models time series data

Can take into account cycles



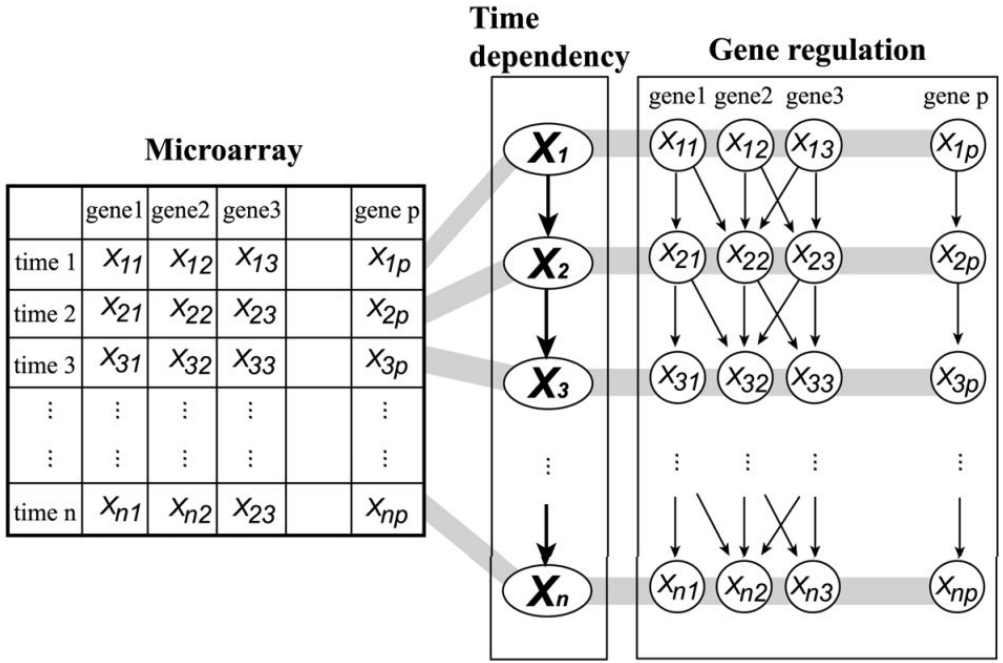


Kim, S. Y., Imoto, S. & Miyano, S. Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Brief. Bioinform.* **4**, 228–235 (2003)

“Unroll” BN along time: dynamic Bayesian network

Models time series data

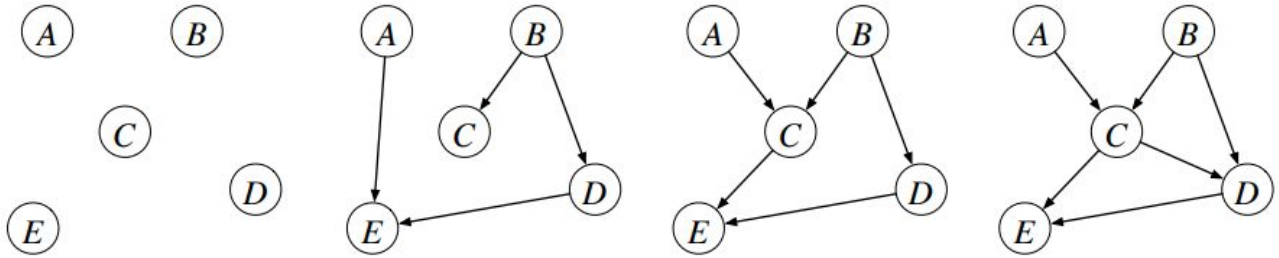
Can take into account cycles



Kim, S. Y., Imoto, S. & Miyano, S. Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Brief. Bioinform.* **4**, 228–235 (2003)

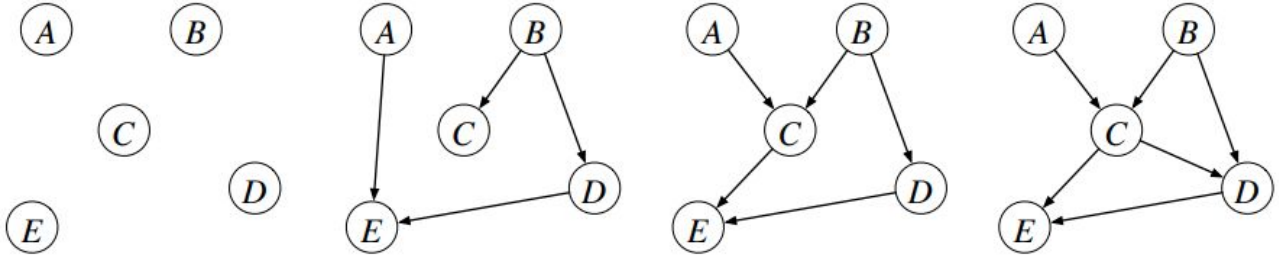
Structure learning in Bayesian Networks

Given a data set of observations of (A, B, C, D, E) can we learn the structure of the graphical mode?



Let G denote the graph structure = the set of edges.

Structure learning in Bayesian Networks



- **Constraint-Based Learning**: Use statistical tests of marginal and conditional independence. Find the set of DAGs whose d-separation relations match the results of conditional independence tests.
- **Score-Based Learning**: Use a global score such as the BIC score or Bayesian marginal likelihood. Find the structures that maximize this score.

Modeling GRN with differential equations

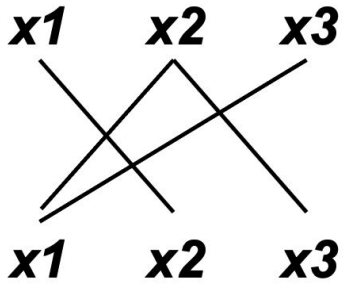
$d\mathbf{x}/dt = \mathbf{A}\mathbf{x} + \mathbf{u}$

Three genes: x_1, x_2, x_3

x1 activates *x2*

x2 activates *x1* and *x3*

x3 inhibits *x1*



$$dx_1/dt = a_{12}x_2 - a_{13}x_3$$

$$dx_2/dt = a_{21}x_1$$

$$dx_3/dt = a_{32}x_2$$

$d\mathbf{x}/dt = \mathbf{A}\mathbf{x} + \mathbf{u}$

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & a_{12} & -a_{13} \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Slide credit: Trey Ideker

Modeling GRN with differential equations

$$\frac{d\mathbf{r}}{dt} = f(\mathbf{p}) - V\mathbf{r} \qquad \frac{d\mathbf{p}}{dt} = L\mathbf{r} - U\mathbf{p}$$

n	The number of genes in the genome;
$\mathbf{r}$	mRNA concentrations, n-dimensional vector-valued functions of t;
$\mathbf{p}$	Protein concentrations, n-dimensional vector-valued functions of t;
$f(\mathbf{p})$	Transcription functions, n-dimensional vector polynomials on $\mathbf{p}$;
L	Translational constants, $n \times n$ non-degenerate diagonal matrix;
V	Degradation rates of mRNAs; $n \times n$ non-degenerate diagonal matrix;
U	Degradation rates of Proteins, $n \times n$ non-degenerate diagonal matrix;

Chen, T., He, H. L. & Church, G. M. MODELING GENE EXPRESSION WITH DIFFERENTIAL EQUATIONS. in *Biocomputing '99* 29–40 (WORLD SCIENTIFIC, 1998).

Summary of today

- Graphical models portrays the sparse dependencies of variables
- Bayesian networks
- Applications in gene regulatory network