INFO 410/510 Bayesian Modelling and Inference

Lecture 6 – Graphical Models and Bayesian inference

Dr. Kunal Arekar
College of Information
University of Arizona, Tucson

September 15, 2025

Recap:
– Reviewed basic probability (and a bit of calculus)
– Discussed some of the general structure of distributions (marginal and conditional independence!)
– Representations of some common discrete and continuous
distributions
– Conjugacy and some popular conjugate pairs

• **We will now introduce a general framework for describing (potentially large) joint probability distributions among random variables with component structure.**

Random Variables and Their Parameterization

$$Bern(x|\mu) = \mu^x(1-\mu)^{(1-x)}$$

$$\mathcal{N}(\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$Bin(m|N,\mu) = \binom{N}{m} \cdot \mu^m(1-\mu)^{(N-m)}$$

$$x \sim \mathcal{N}(\mu, \sigma)$$

$$Beta(\mu|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\mu^{\alpha-1}(1-\mu)^{\beta-1}$$

$$x|\mu, \sigma$$

In Bayesian updating, a random variable's parameters are themselves random variables

$$Cat(\mathbf{x}|\boldsymbol{\mu}) = \prod_{k=1}^{K} \boldsymbol{\mu}_k^{x_k}$$

$$Mult(\underbrace{m_1, m_2, ..., m_K}_{\mathbf{m}}|\boldsymbol{\mu}, N) = \binom{N}{m_1, m_2, ..., m_K} \prod_{k=1}^{K} \boldsymbol{\mu}_k^{m}$$

$$Dir(\boldsymbol{\mu}|\boldsymbol{\alpha}) = \frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\Gamma(\alpha_1)\Gamma(\alpha_2)...\Gamma(\alpha_K)} \prod_{k=1}^{K} \boldsymbol{\mu}_k^{\alpha_k-1}$$

Graphical Models

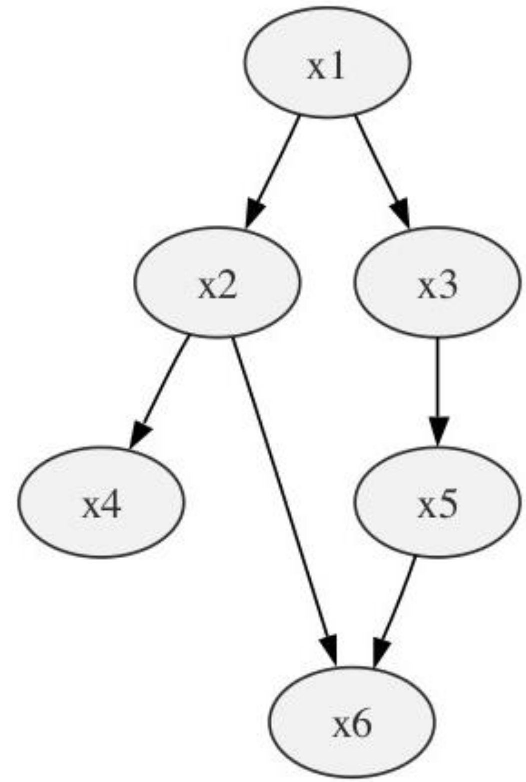
• Graphical representation of statistical models

• Nodes
– Random variables (or groups of them)

• Edges
– Probabilistic relationships between nodes

Introduction to graphical models

- Marriage between probability theory and graph theory
- Provides a tool for dealing with uncertainty and complexity
  dealing with uncertainty using probability theory
  coping with complexity using graph theory
- A graphical model formalizes the structure of the dependencies between random variables
- By encapsulating dependencies and conditional independencies, graphical models simplify the specification of joint probability distributions, making it easier to build, interpret, and compute Bayesian models
- Allows for practical reasoning about extensive network of random variables



Node – random variables
Edges – prob. relationship between nodes

$$p(x1, x2, x3, x4, x5, x6) = p(x1)p(x2|x1)p(x3|x1)p(x4|x2)p(x5|x3)p(x6|x2, x5)$$

Types of graphical models

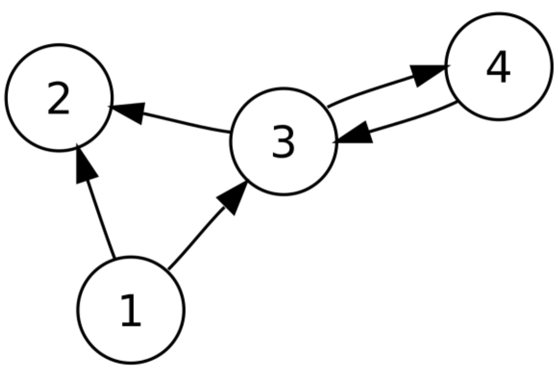
A graph is a collection of nodes and edges

Nodes are vertices that correspond to objects.

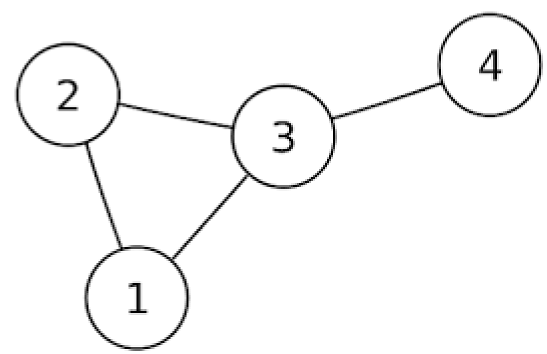
Edges are the connections between objects.

The graph edges sometimes have Weights, which indicate the strength (or some other attribute) of each connection between the nodes.

Directed graphical models



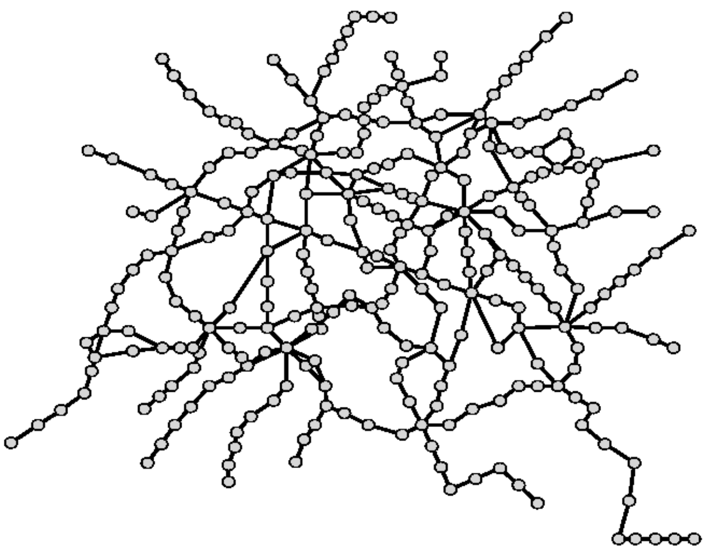
Undirected graphical models

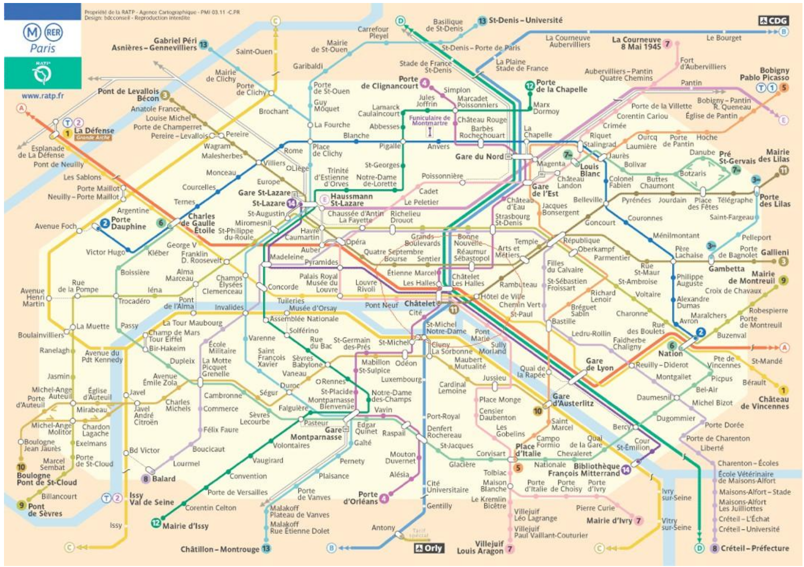


Set of vertices connected pairwise by edges.

Why study graph algorithms?

- Thousands of practical applications.
- Hundreds of graph algorithms known.
- Interesting and broadly useful abstraction. Models a wide range of systems
- Challenging branch of computer science and discrete math.

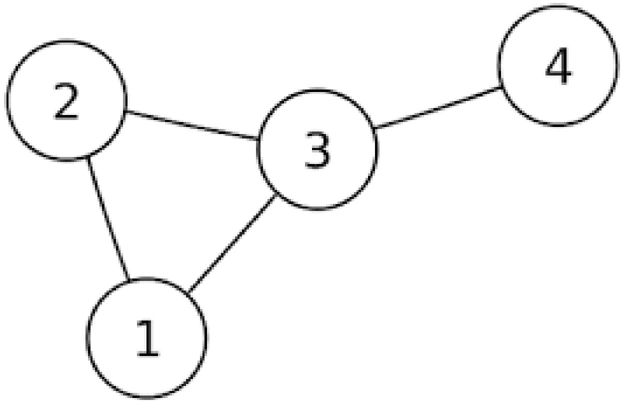




Undirected graphs

An undirected graph is a type of graph where the edges have no specified direction assigned to the them

- Edges in an undirected graph are bidirectional in nature
- no concept of a "parent" or "child" vertex
- may contain loops, which are edges that connect a vertex to itself
- Degree of each vertex is the same as the total no of edges connected to it



Pros:

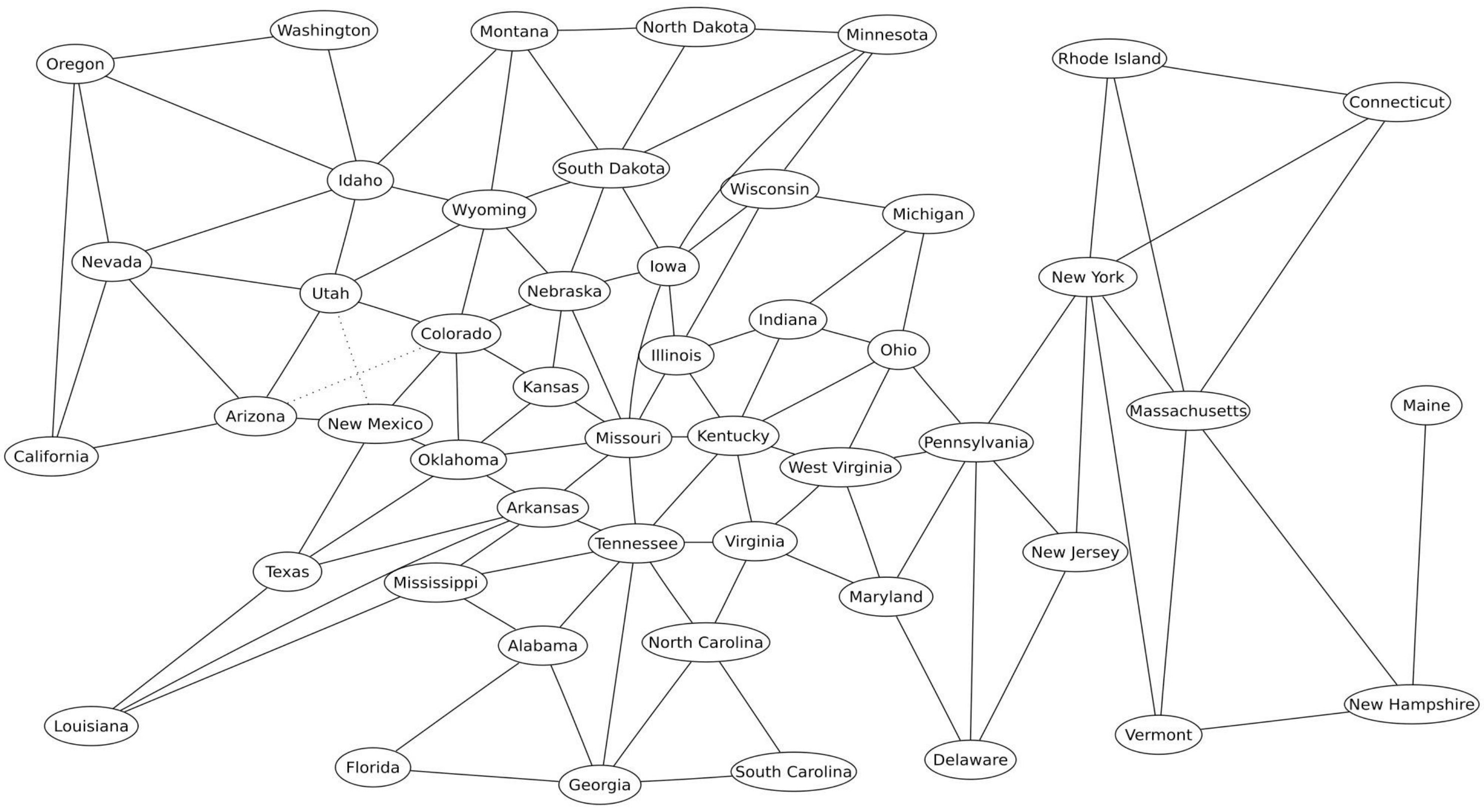
- provides simplistic structures
- provides flexibility

Cons:

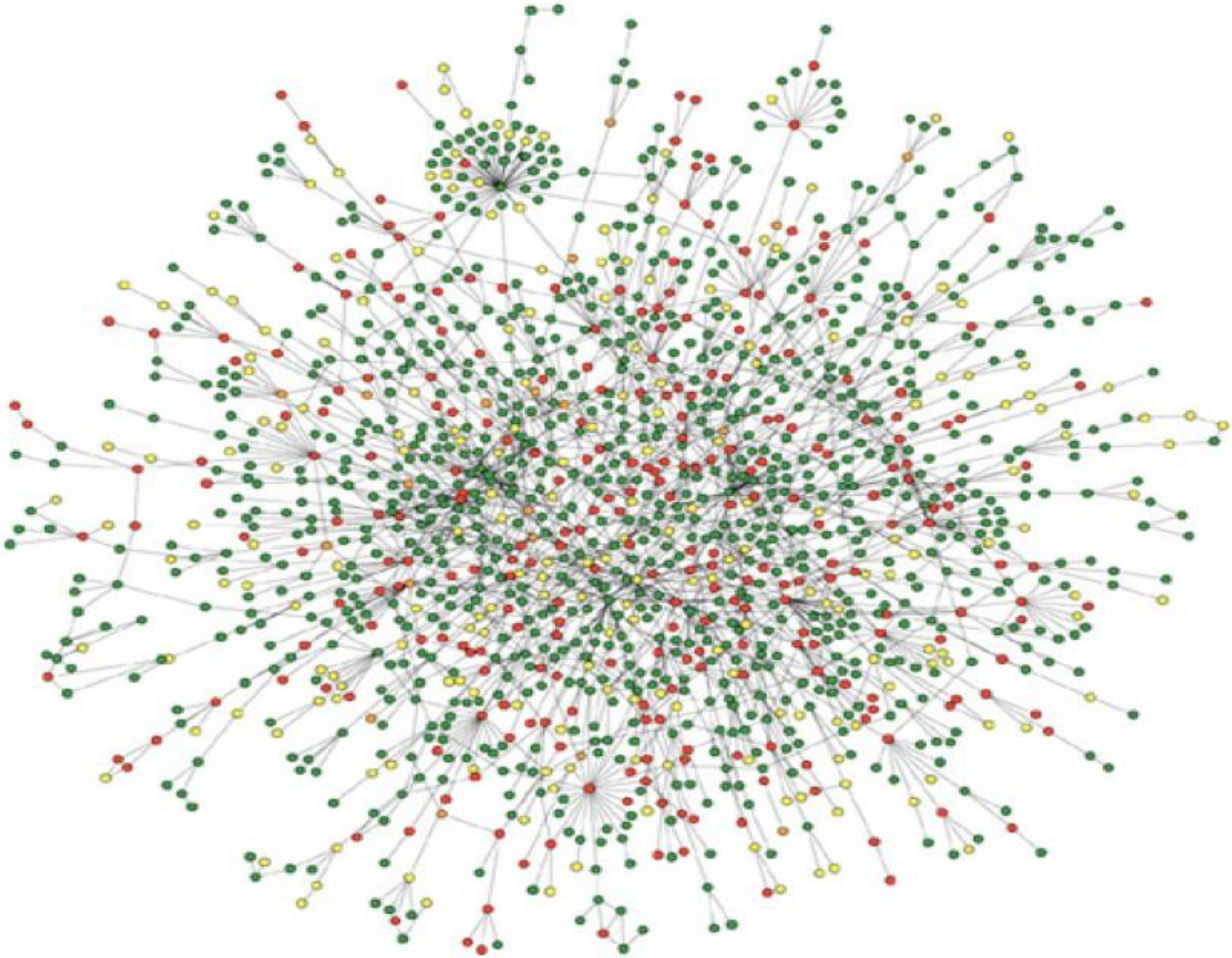
- undesirable for some applications; lack of directionality
- provides limited information

Undirected graphs

Border graph of 48 contiguous United States

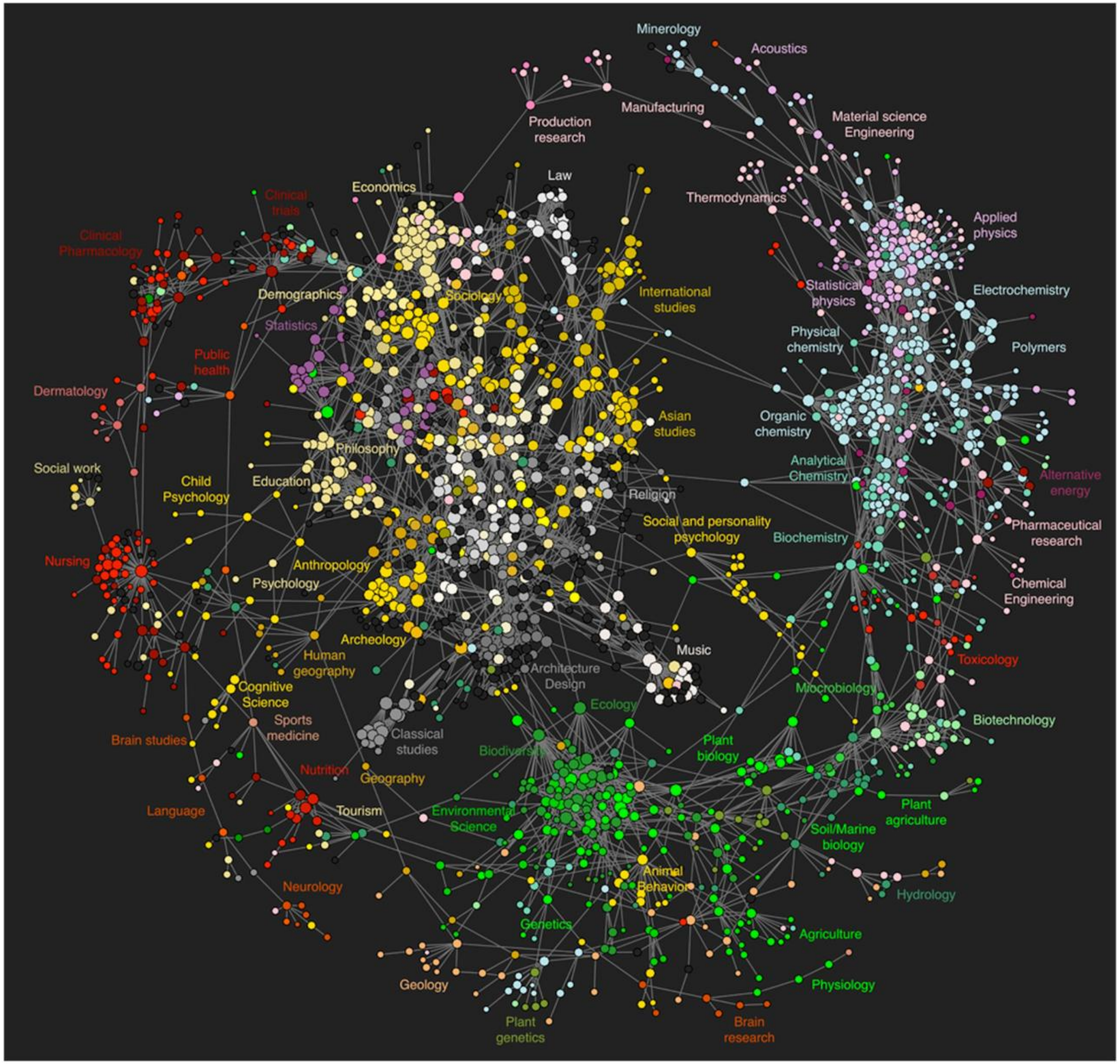


Protein-protein interaction network



Reference: Jeong et al, Nature Review | Genetics

Map of science clickstreams

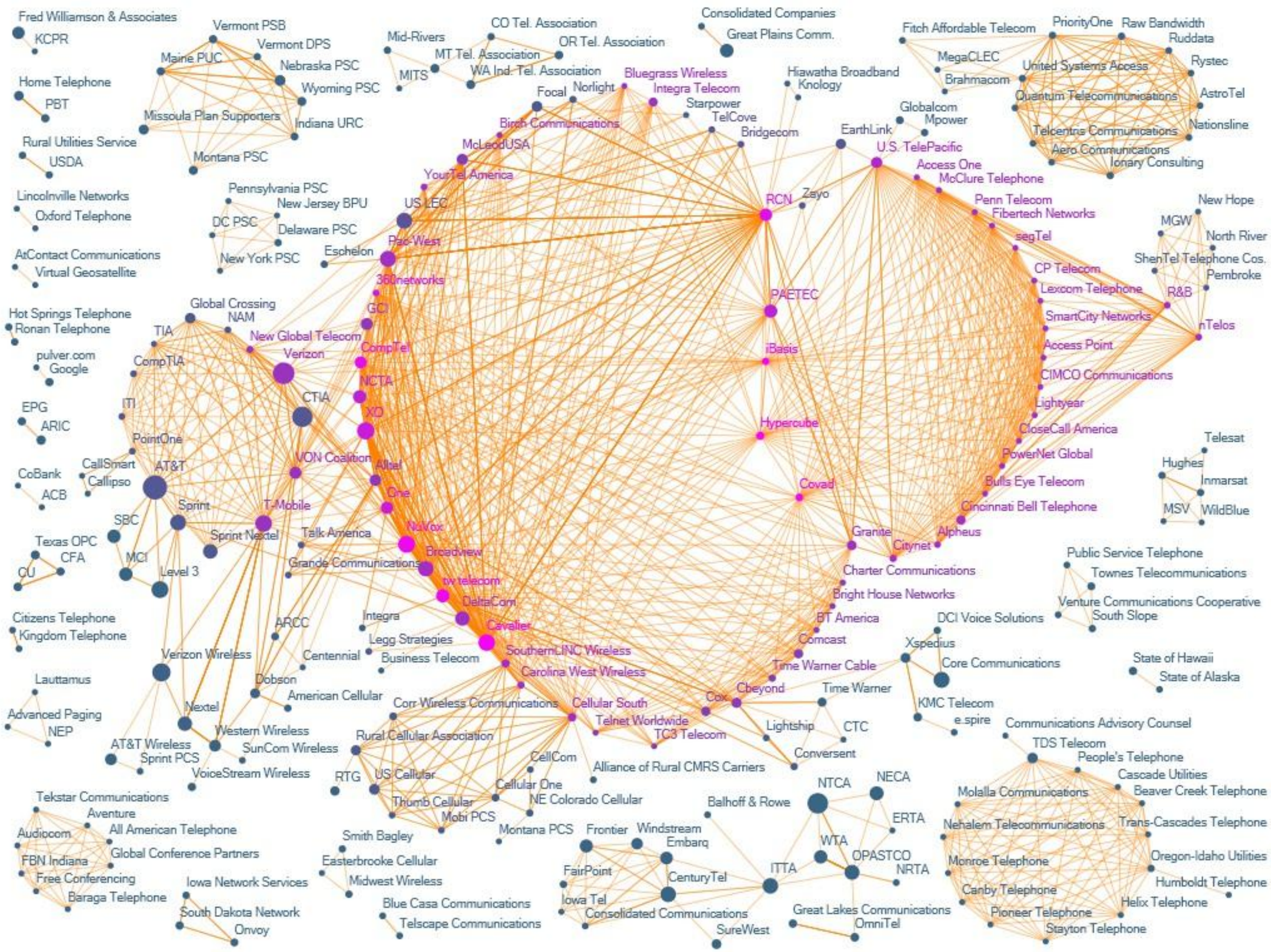


http://www.plosone.org/article/info:doi/10.1371/journal.pone.0004803

10 million Facebook friends

"Visualizing Friendships" by Paul Butler

The evolution of FCC lobbying coalitions



"The Evolution of FCC Lobbying Coalitions" by Pierre de Vries in JoSS Visualization Symposium 2010

Framingham heart study

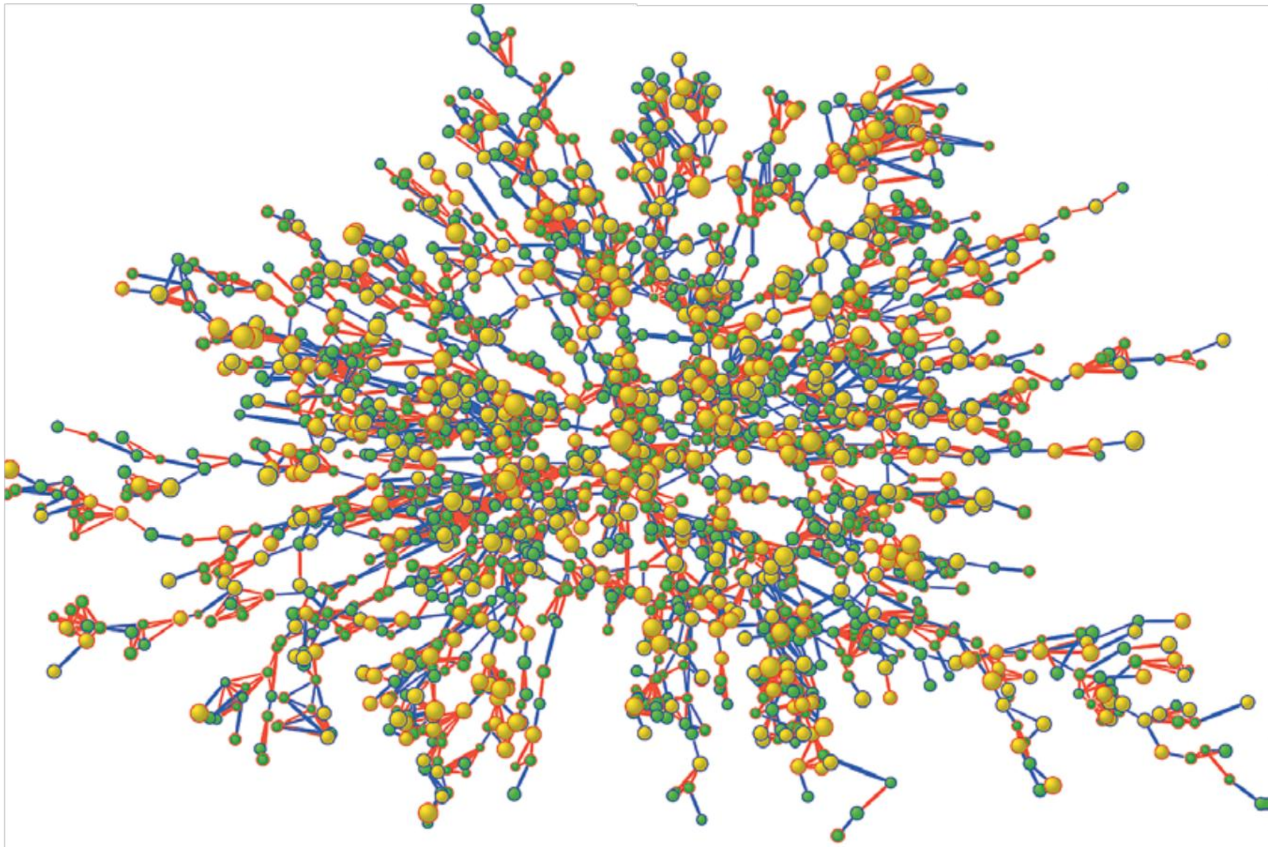
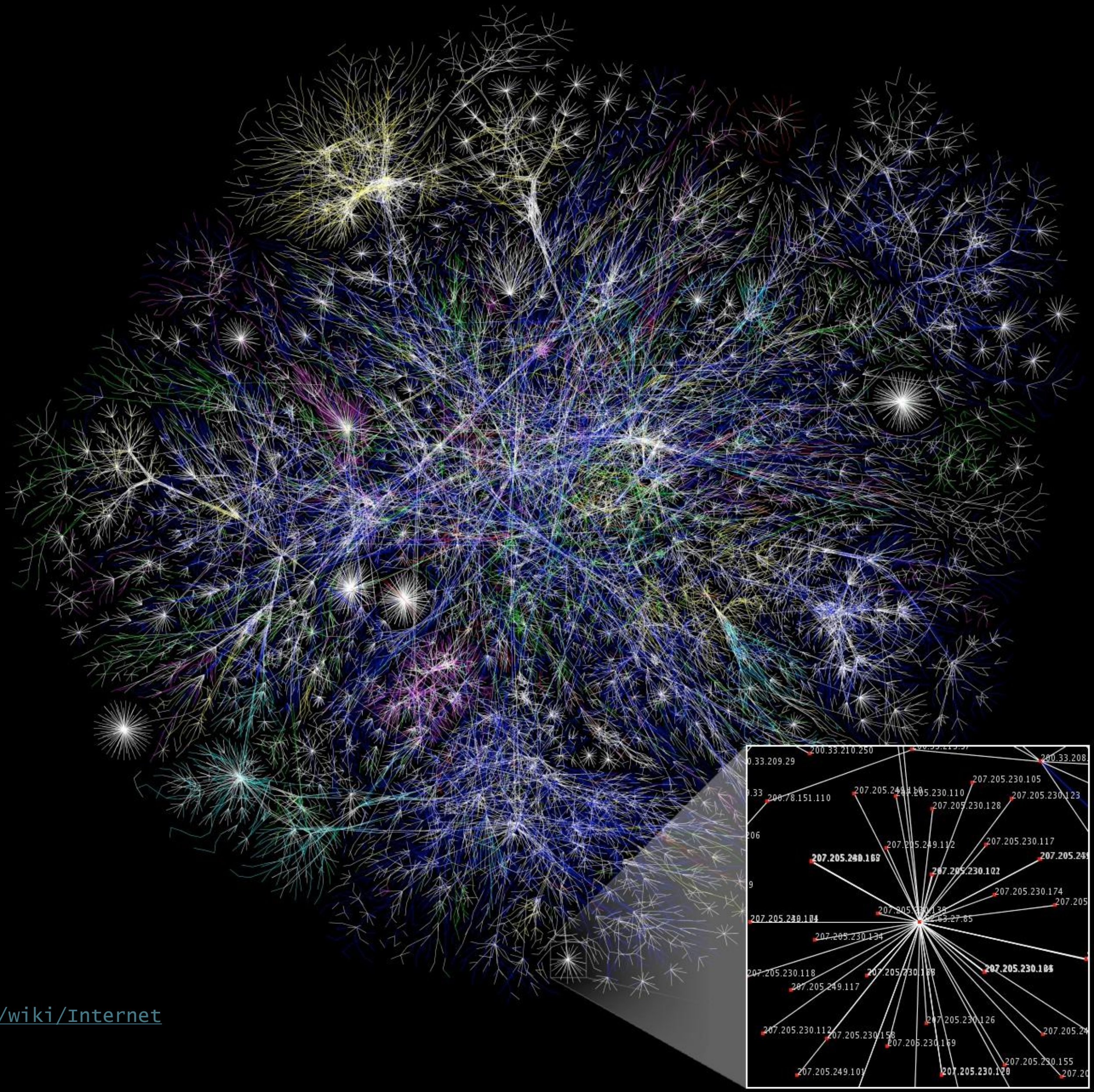


Figure 1. Largest Connected Subcomponent of the Social Network in the Framingham Heart Study in the Year 2000.
Each circle (node) represents one person in the data set. There are 2200 persons in this subcomponent of the social network. Circles with red borders denote women, and circles with blue borders denote men. The size of each circle is proportional to the person's body-mass index. The interior color of the circles indicates the person's obesity status: yellow denotes an obese person (body-mass index, ≥30) and green denotes a nonobese person. The colors of the ties between the nodes indicate the relationship between them: purple denotes a friendship or marital tie and orange denotes a familial tie.

"[The Spread of Obesity in a Large Social Network over 32 Years]" **by Christakis and Fowler in New England Journal of Medicine, 2007**

The Internet as mapped by the Opte Project



http://en.wikipedia.org/wiki/Internet

Graph applications

graph	vertex	edge
communication	telephone, computer	fiber optic cable
circuit	gate, register, processor	wire
mechanical	joint	rod, beam, spring
financial	stock, currency	transactions
transportation	intersection	street
internet	class C network	connection
game	board position	legal move
social relationship	person	friendship
neural network	neuron	synapse
protein network	protein	protein-protein interaction
molecule	atom	bond

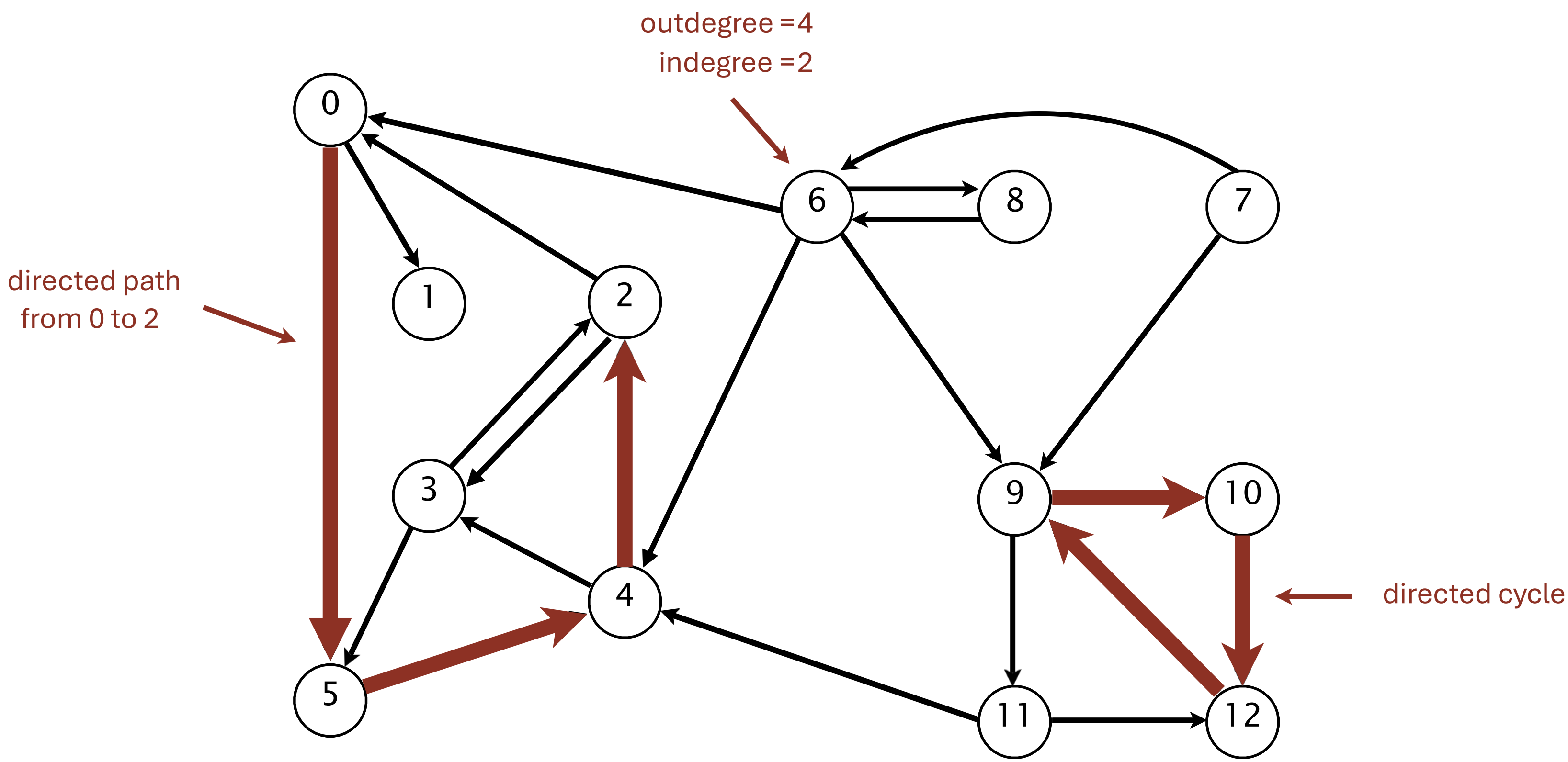
Directed graphs

A directed graph (or digraph) is a set of vertices and a collection of directed edges that each connects an ordered pair of vertices.

- Directed graph
  - Directed Acyclic Graphs (DAGs) — Directed edges with no cycles, eg. Bayesian network
  - Cyclic Directed Graphs — Graphical model with a cycle e.g. feedback loop
  - Multigraphs — Multiple edges between nodes e.g. transportation networks
  - Weighted Directed Graphs — E.g., Network routing problems, where weights represent distances, costs, or time
  - Trees (Rooted Directed Trees) — E.g., Organizational charts, evolutionary trees
  - Flow Networks — Directed graphs with a source and a sink e.g. Water or traffic flow networks
  - Polytrees

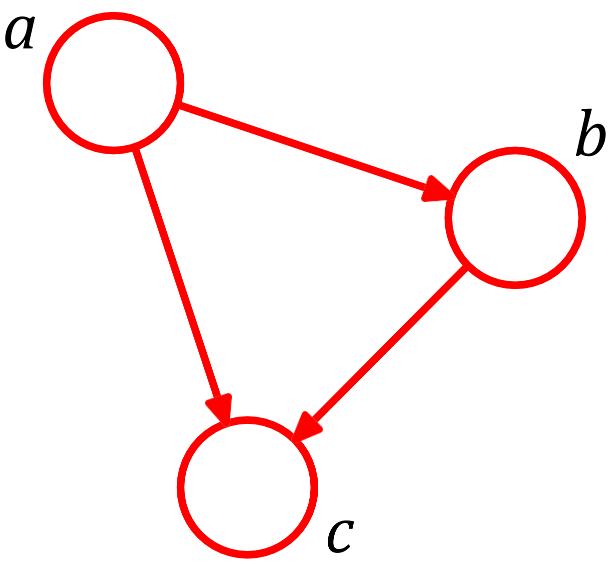
Directed graphs

Digraph - Set of vertices connected pairwise by directed edges.



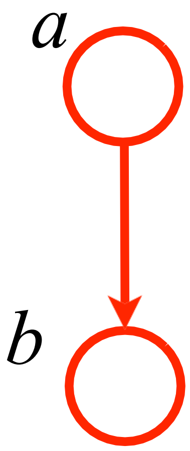
Directed Graphical Models

- Nodes represent random variables
- Edges between nodes have directed links
- No cycles!



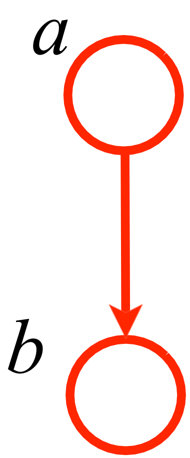
Directed Graphical Models

- Nodes represent random variables
- Edges between nodes have directed links
- No cycles!
- The graph represents a **factorization** of the joint probability of all of the random variables represented by the nodes.



  - An arrow from one node (a) to another another (b) indicates the second node (b) is conditioned on the first (a).
  - In other words, if you have information about (a), then you have information about (b).
  - The arrows tell you about information flow.

Directed Graphical Models



- Here we have two nodes, *a* and *b*.
- This is a representation of the joint distribution *p*(*a*,*b*).
- In particular, it is equivalent to writing:

$$p(a, b) = p(b|a)p(a)$$

- **Ancestral sampling** version of the story:

To sample from $p(a, b)$

First sample $\tilde{a}$ from $p(a)$

Then sample $\tilde{b}$ from $p(b|\tilde{a})$

Directed Graphical Models

- Now let's consider three random variables: a, b, and c
- General model is $p(a,b,c)$
- What are the possible relationships of a, b, and c?

Directed Graphical Models

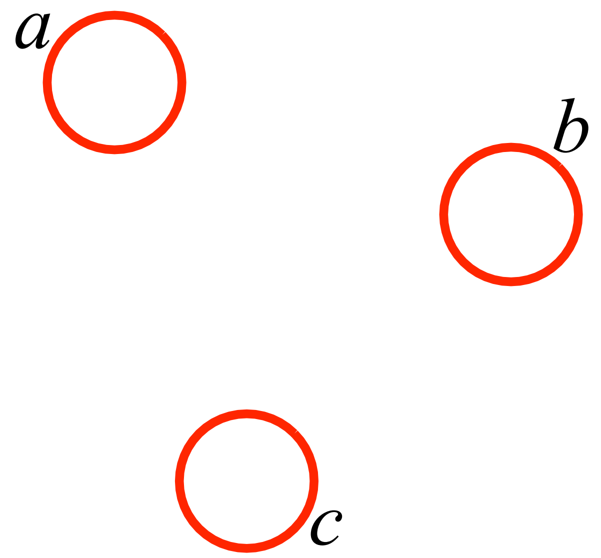
- Now let's consider three random variables: *a*, *b*, and *c*

- General model is $p(a,b,c)$

- What are the possible relationships of *a*, *b*, and *c*?
  - Independence: $p(a,b,c) = p(a)p(b)p(c)$
  - Some structure: $p(a,b,c) = p(a)p(b|a)p(c|a)$
  - Arbitrary relationship

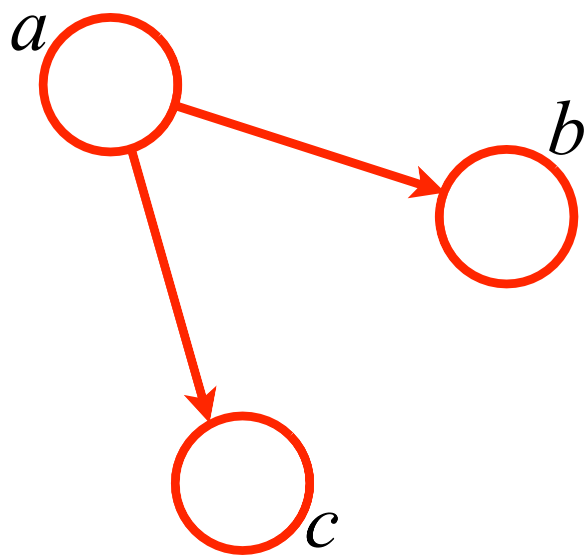
$$p(a,b,c) = p(a)p(b)p(c)$$

$$p(a,b,c) = p(a)p(b)p(c)$$

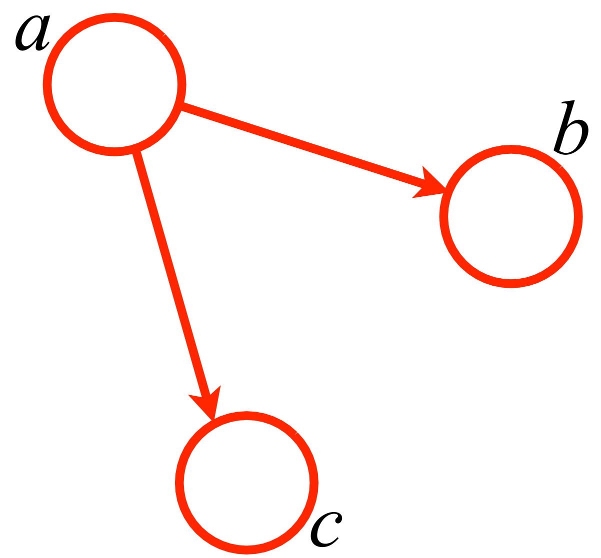


$$p(a,b,c) = p(a)p(b|a)p(c|a)$$

$$p(a,b,c) = p(a)p(b|a)p(c|a)$$

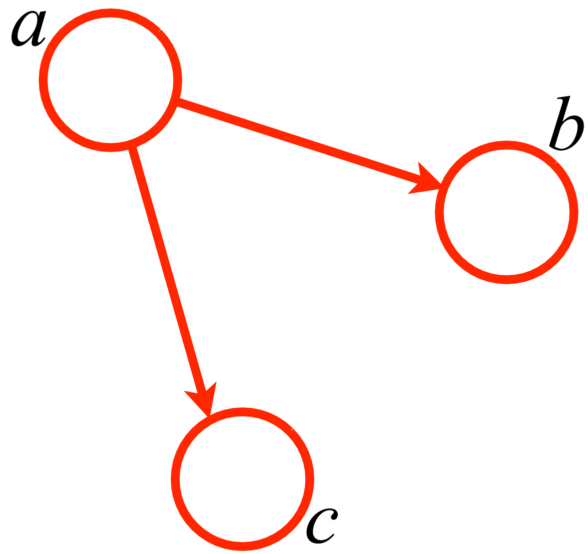


$$p(a,b,c) = p(a)p(b|a)p(c|a)$$



What is the independence relationship here?

$$p(a,b,c) = p(a)p(b|a)p(c|a)$$



What is the independence relationship here?

$$b \perp\!\!\!\perp c \quad a$$

$p(a,b,c)$ with no structure

p(*a*,*b*,*c*) with no structure

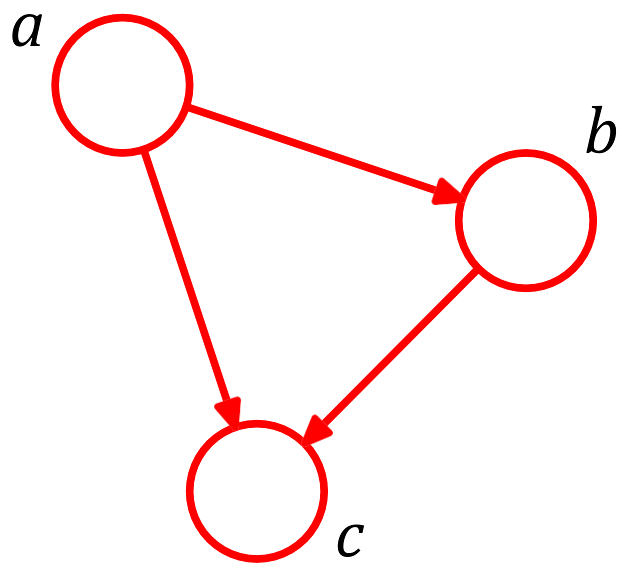
$$p(a,b,c) = p(a)p(a|b)p(c|a,b)$$

$$p(a,b,c) = p(a)p(c|a)p(b|a,c)$$

$$p(a,b,c) = p(b)p(c|b)p(a|c,b)$$

$$\cdots$$

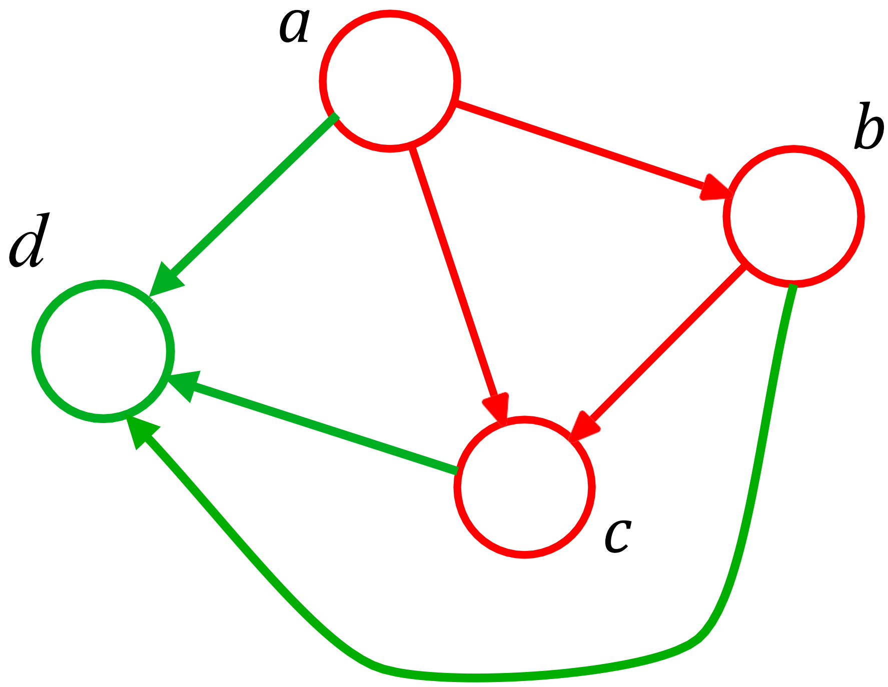
$$p(a, b, c) = p(a)p(b \mid a)p(c \mid a, b)$$



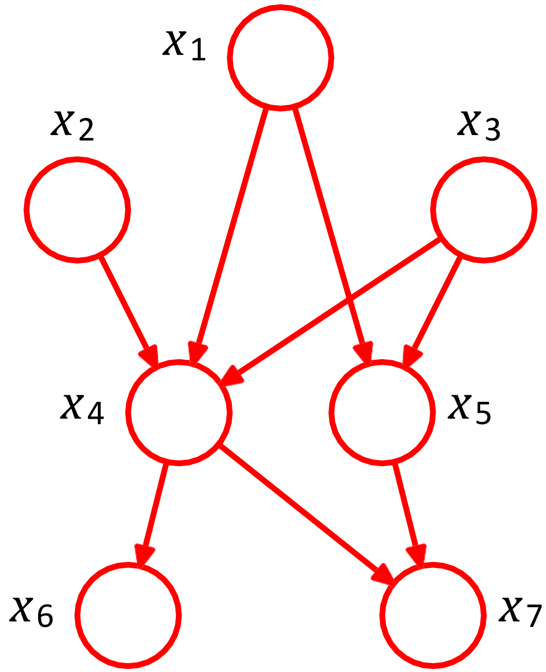
Note that the graph is fully connected

$$p(a,b,c,d) = p(d \mid a,b,c)\,p(a,b,c)$$

$$p(a,b,c,d) = p(d \mid a,b,c)\, p(a,b,c)$$

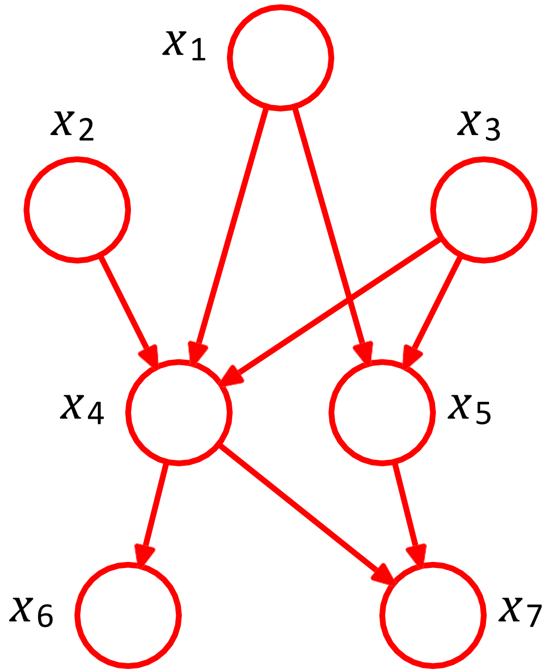


Another Example



What is the algebraic form?

Another Example



What is the algebraic form?

$$p(x_1)p(x_2)p(x_3)p(x_4|x_1,x_2,x_3)p(x_5|x_1,x_3)p(x_6|x_4)p(x_7|x_4,x_5)$$

Univariate Gaussian with Known Variance

$$D = \{x_1, x_2, x_3, ..., x_N\}$$

$$p(D, \mu) = p(\mu) \prod_{n-1}^{N} p(x_n|\mu)$$

where

$$p(x_n|\mu) = \mathcal{N}(x_n|\mu; \sigma^2)$$

Univariate Gaussian with Known Variance

$$D = \{x_1, x_2, x_3, ..., x_N\}$$

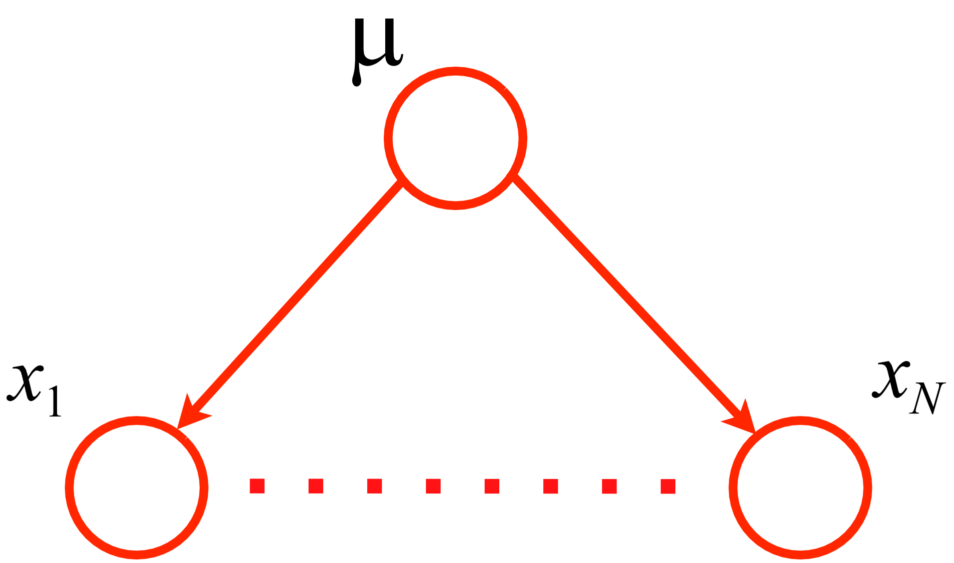
$$p(D, \mu) = p(\mu) \prod_{n-1}^{N} p(x_n|\mu)$$

where

$$p(x_n|\mu) = \mathcal{N}(x_n|\mu; \sigma^2)$$

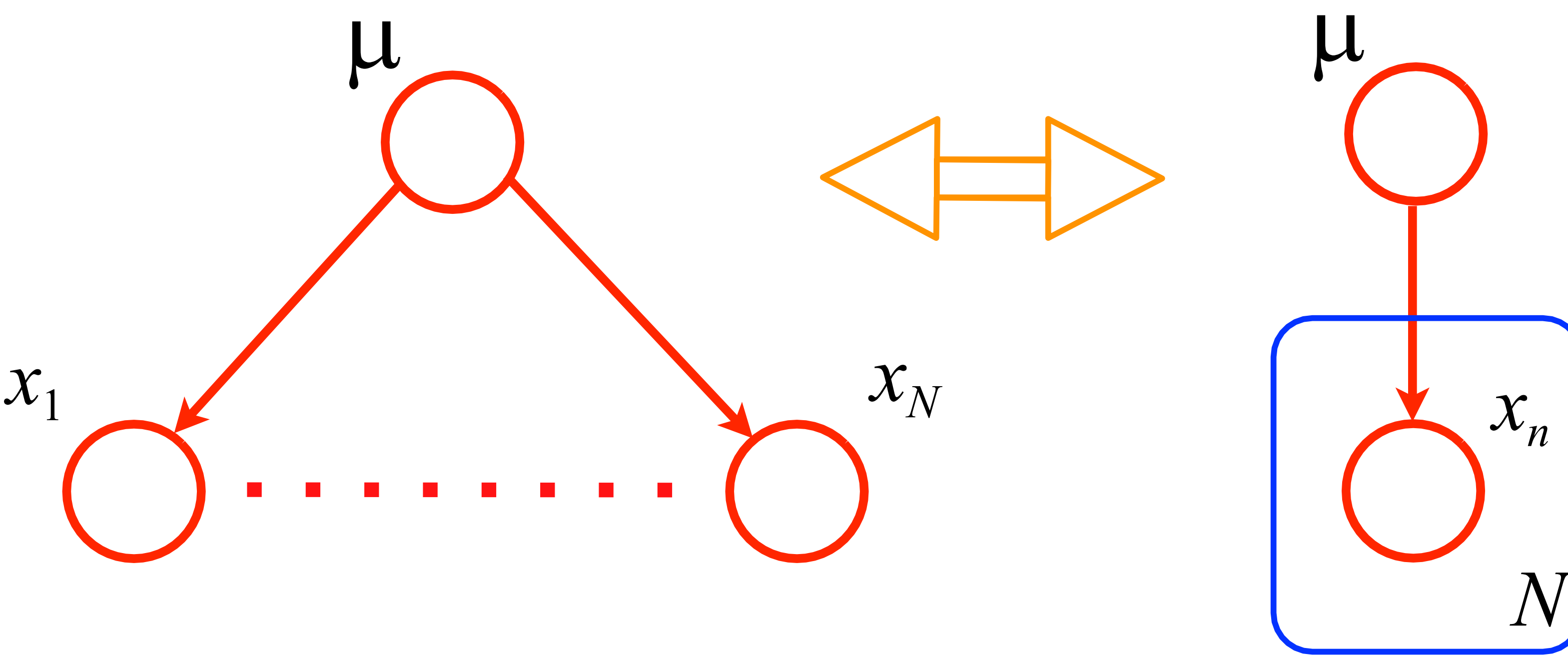
$$p(D, \mu) = p(\mu|\mu_0, \sigma_0^2) \prod_{n=1}^{N} p(x_n|\mu; \sigma^2)$$

Univariate Gaussian with Known Variance



$$p(D, \mu) = p(\mu | \mu_0, \sigma_0^2) \prod_{n=1}^{N} p(x_n | \mu; \sigma^2)$$

Univariate Gaussian with Known Variance



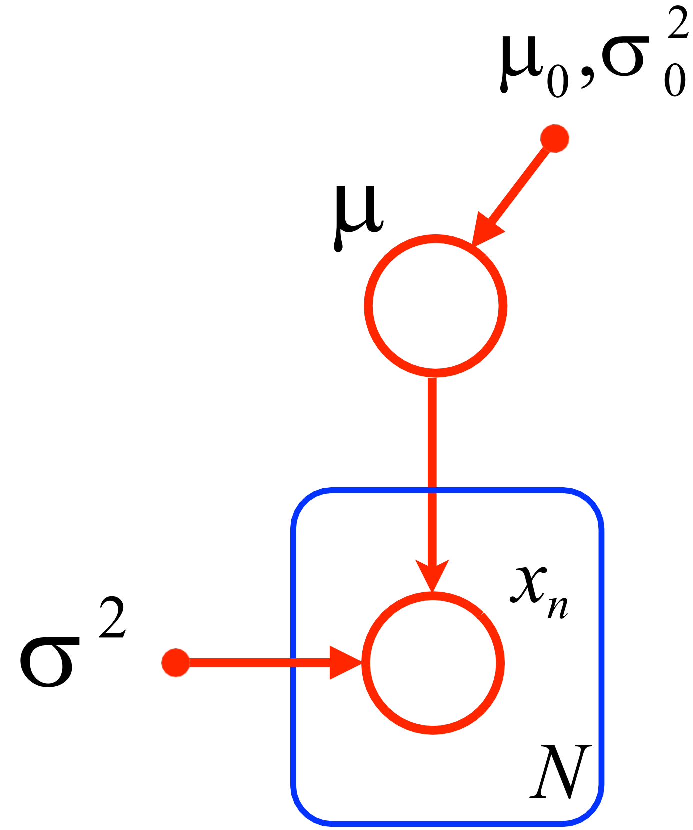
More compact notation: “plate” representation

$$p(D, \mu) = p(\mu | \mu_0, \sigma_0^2) \prod_{n=1}^{N} p(x_n | \mu; \sigma^2)$$

Deterministic Parameters

Our univariate Gaussian has some known parameters: the variance and the (ideally) conjugate prior on the mean.

If we wish to illustrate them, we use a small filled in circle.

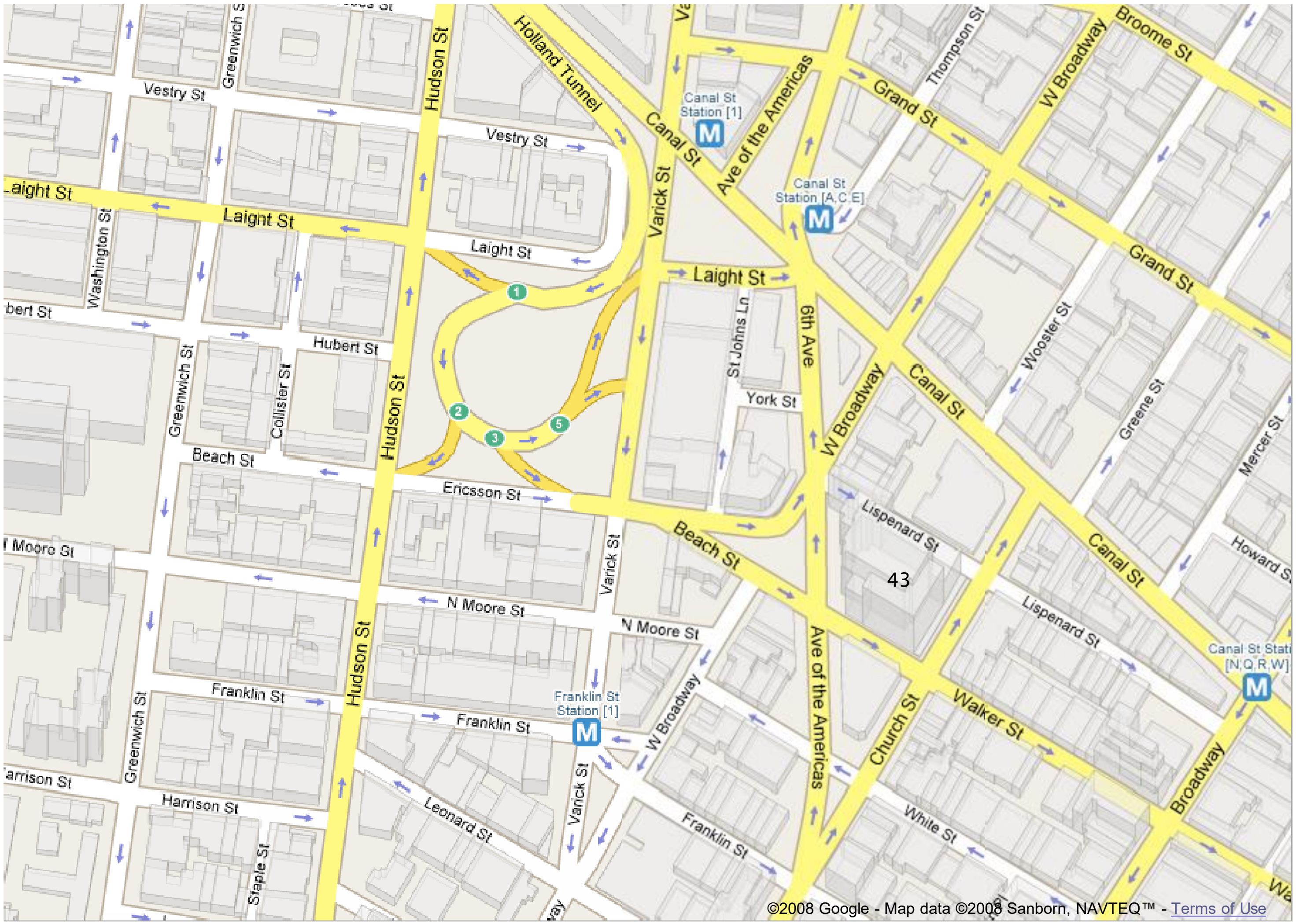


More compact notation: "plate" representation

$$p(D, \mu) = p(\mu | \mu_0, \sigma_0^2) \prod_{n=1}^{N} p(x_n | \mu; \sigma^2)$$

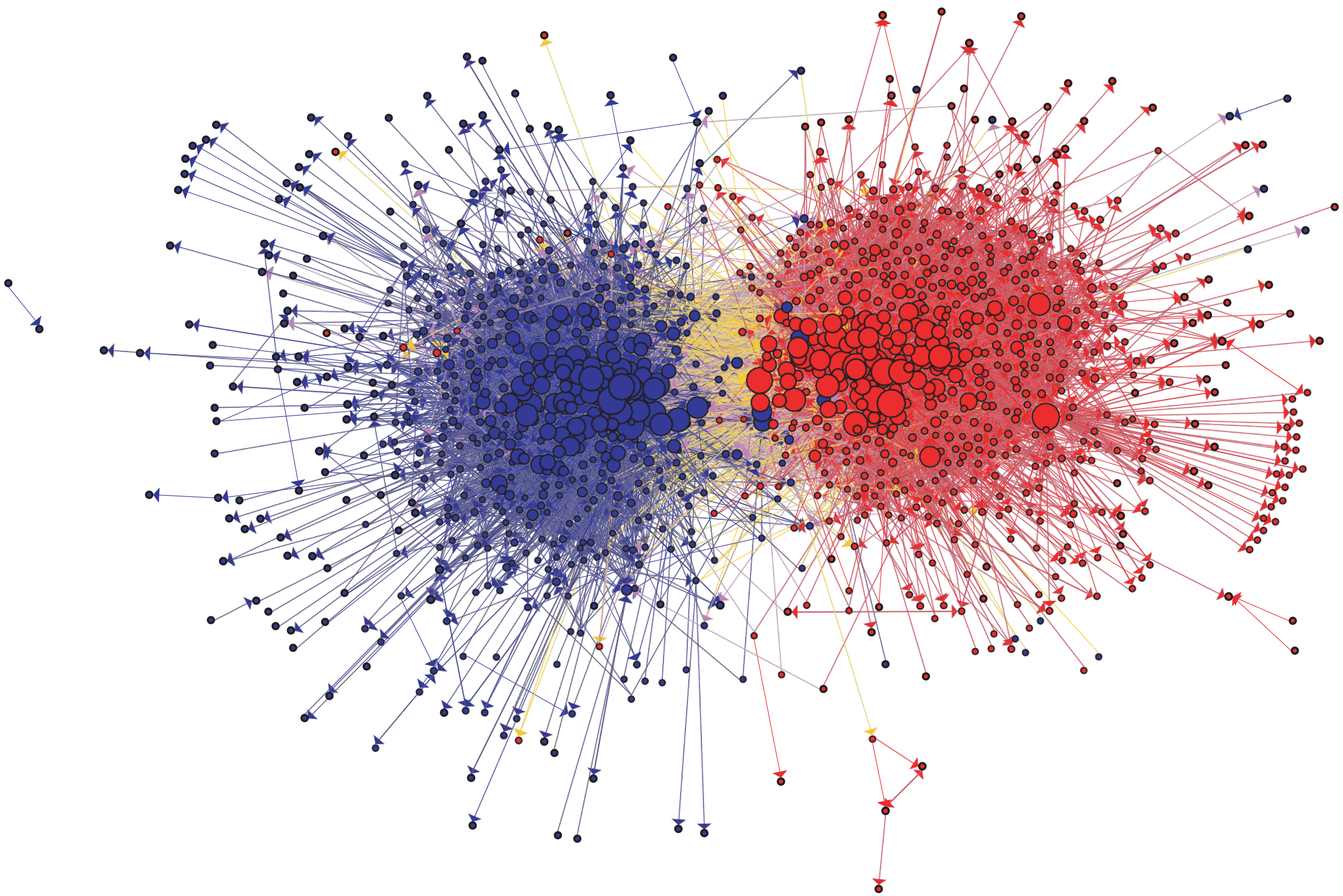
Road network

Vertex = intersection; edge = one-way street.



Political blogosphere graph

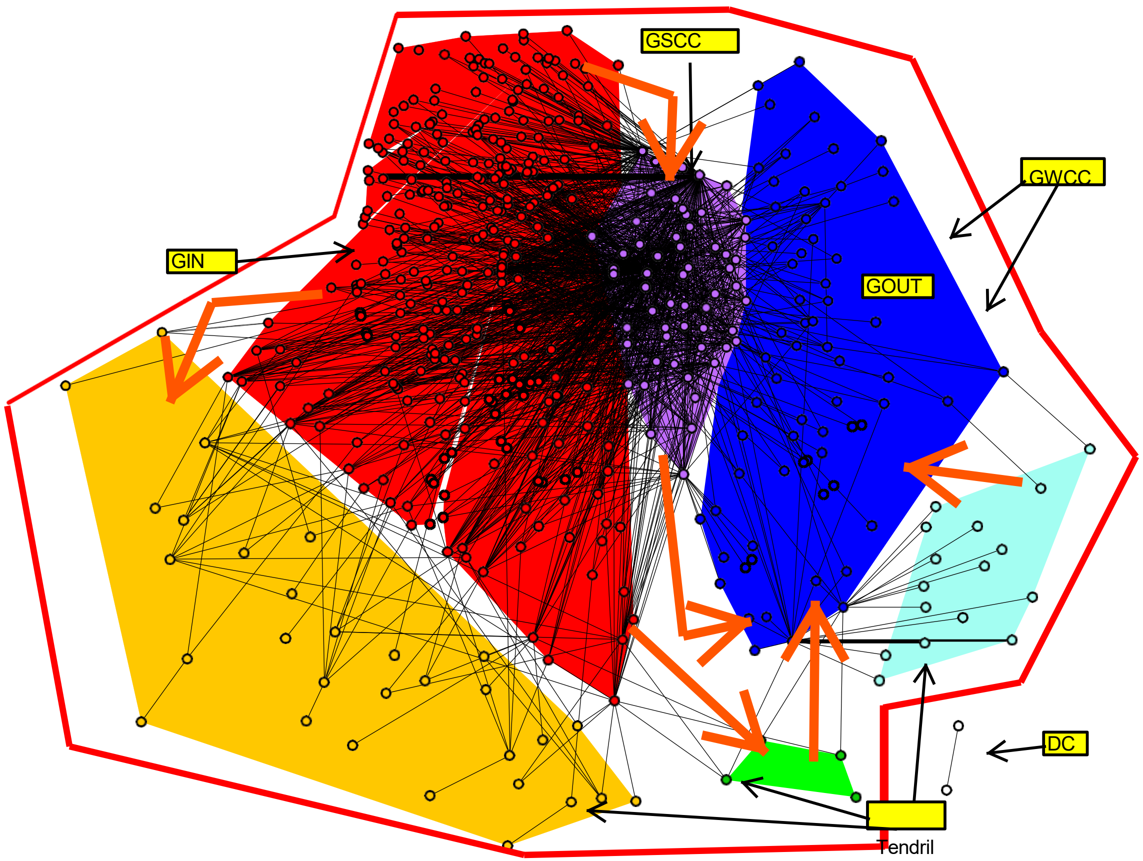
Vertex = political blog; edge = link.



The Political Blogosphere and the 2004 U.S. Election: Divided They Blog, Adamic and Glance, 2005

Overnight interbank loan graph

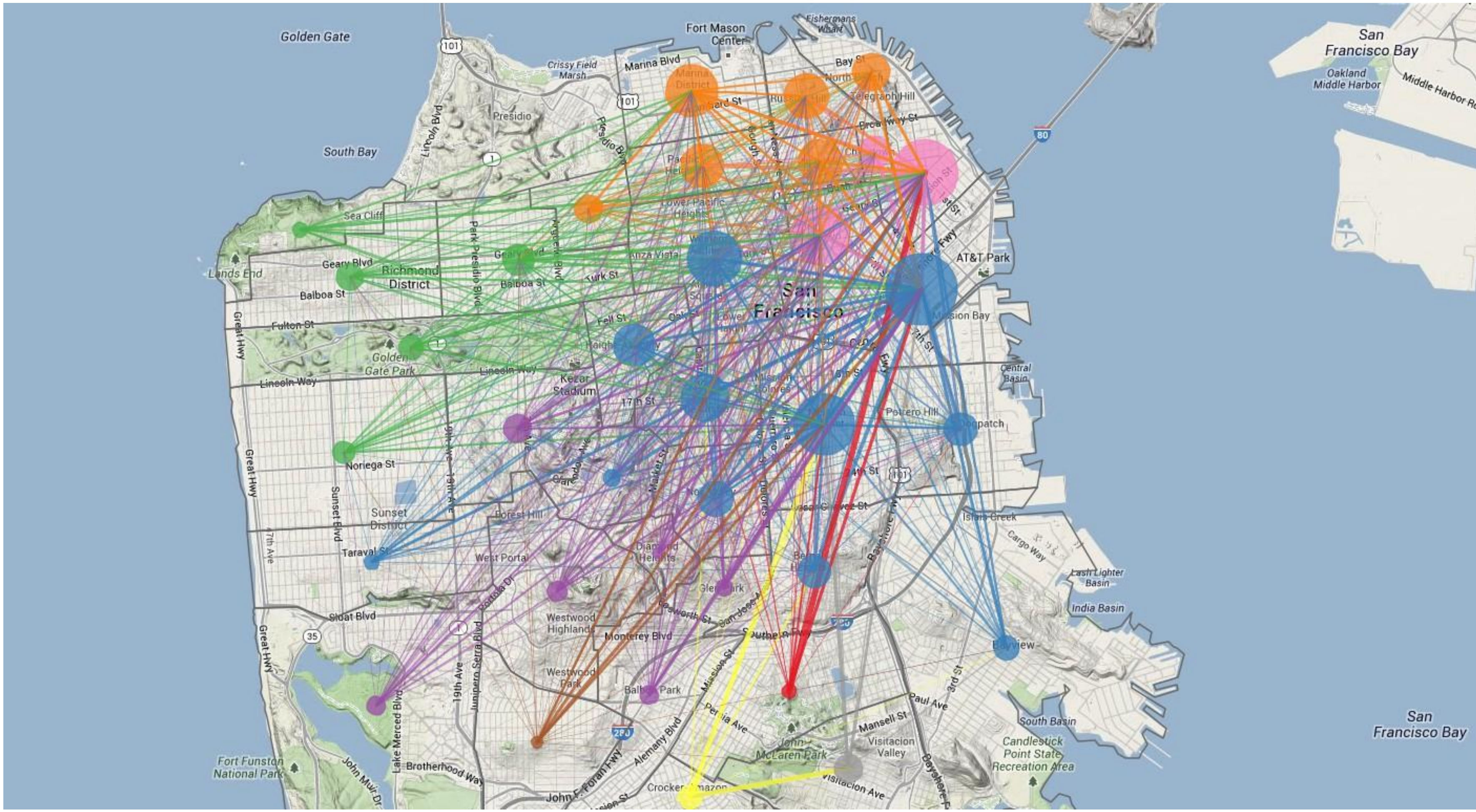
Vertex = bank; edge = overnight loan.



The Topology of the Federal Funds Market, Bech and Atalay, 2008

Uber taxi graph

Vertex = taxi pickup; edge = taxi ride.

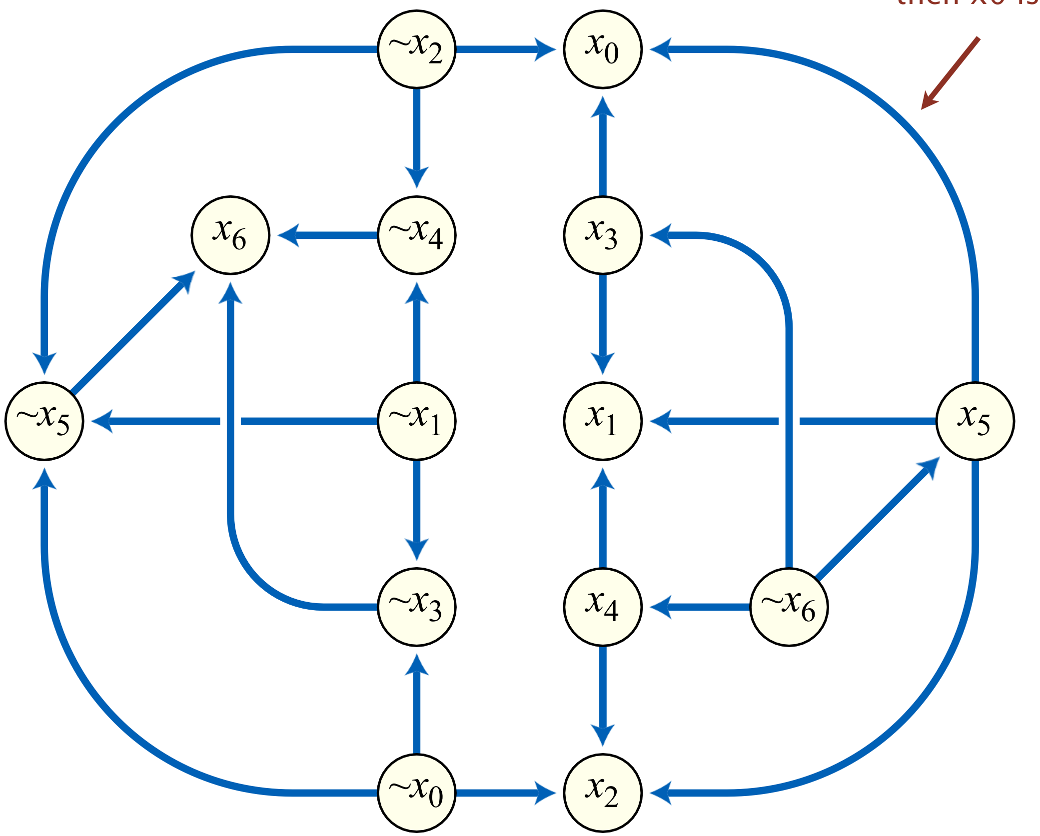


http://blog.uber.com/2012/01/09/uberdata-san-franciscomics/

Implication graph

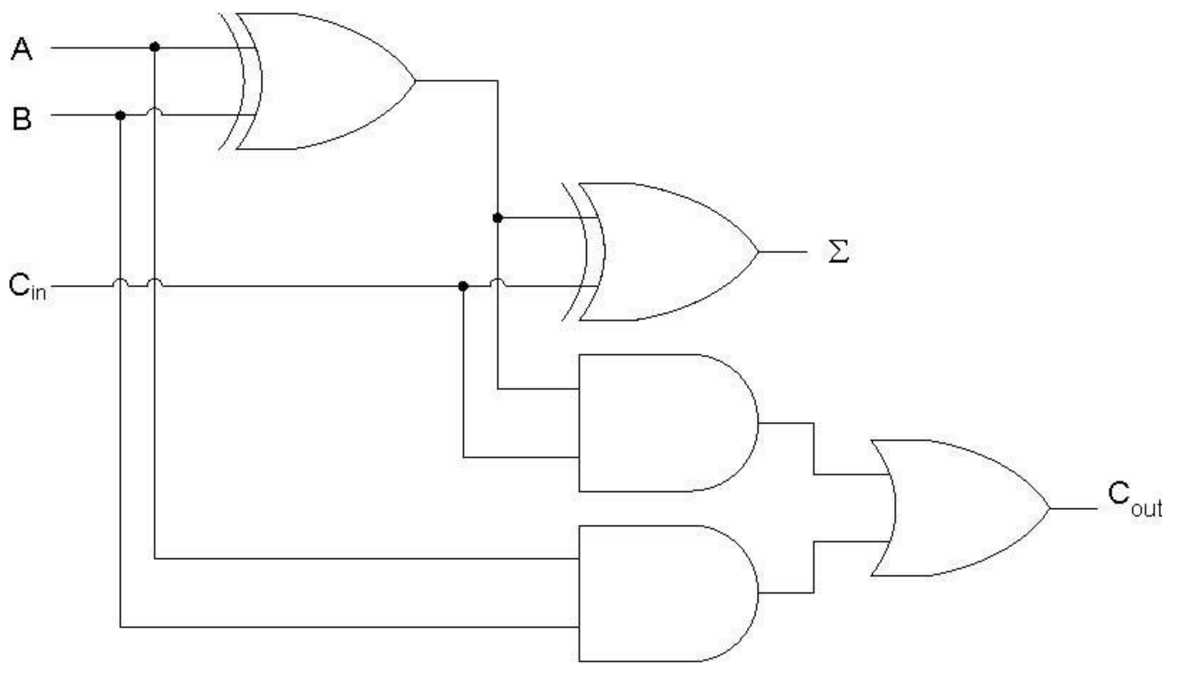
Vertex = variable; edge = logical implication.

if x5 is true,
then x0 is true



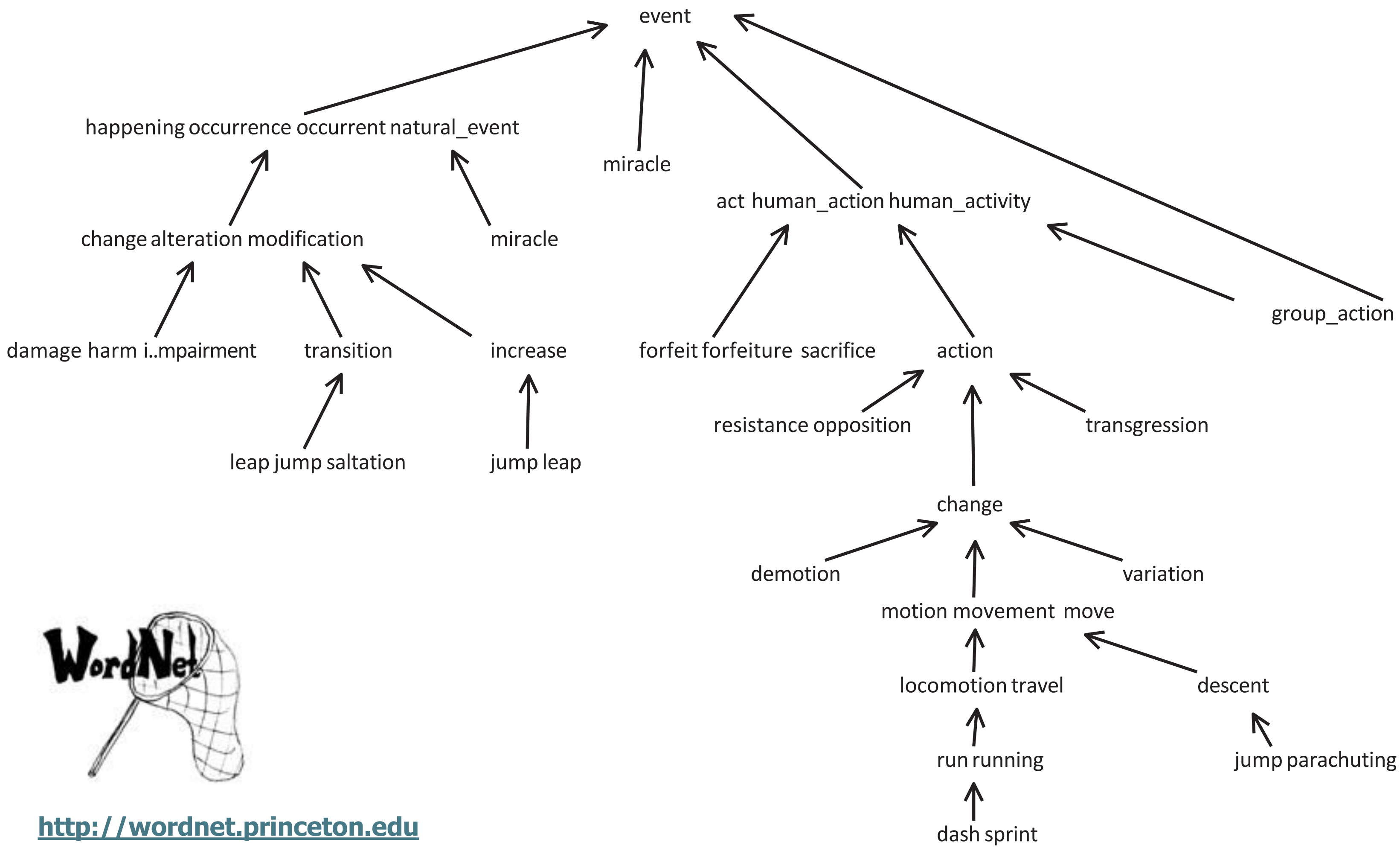
Combinational circuit

Vertex = logical gate; edge = wire.



WordNet graph

Vertex = synset; edge = hypernym relationship.



http://wordnet.princeton.edu

Digraph applications

digraph	vertex	directed edge
transportation	street intersection	one-way street
web	web page	hyperlink
food web	species	predator-prey relationship
WordNet	synset	hypernym
scheduling	task	precedence constraint
financial	bank	transaction
cell phone	person	placed call
infectious disease	person	infection
game	board position	legal move
citation	journal article	citation
object graph	object	pointer
inheritance hierarchy	class	inherits from
control flow	code block	jump

Directed graphs

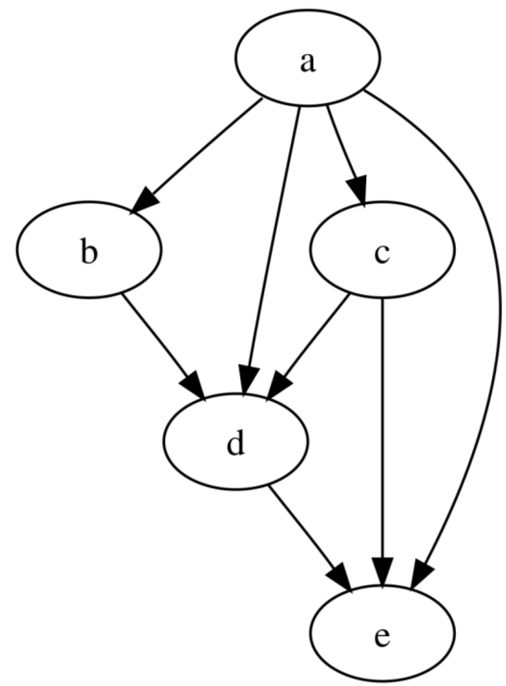
Directed acyclic graphs –

A Directed Acyclic Graph (DAG) is a directed graph that does not contain any cycles.

Two important features:

Directed Edges: each edge has a direction, from one vertex (node) to another.

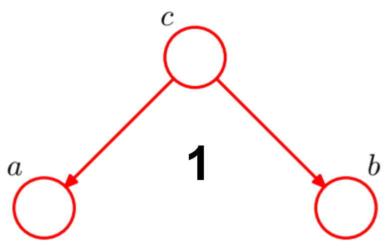
Acyclic: no cycles or closed loops within the graph



E.g., Bayesian networks

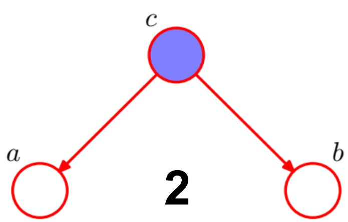
Conditional independence

the probability of the hypothesis given the uninformative observation is equal to the probability without that observation



$$p(a,b,c) = p(c)p(a|c)p(b|c)$$

c is not observed



$$p(a,b|c) = p(a|c)p(b|c)$$

c is observed

1 – With no conditioning, no independence $a \not\perp\!\!\!\perp b$ a is not independent of b

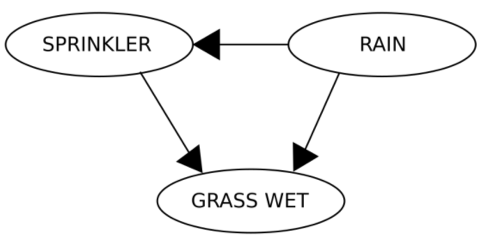
2 – With conditioning, we have independence $a \perp\!\!\!\perp b \mid c$ a is independent of b given c

Directed graphs

Directed acyclic graphs –

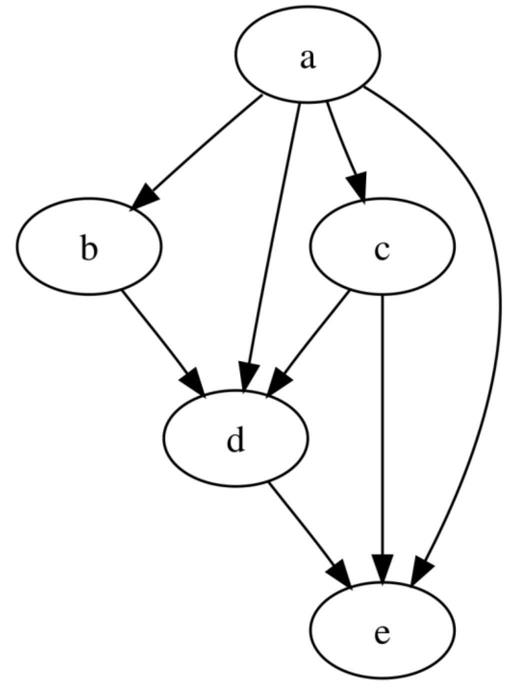
A Directed Acyclic Graph (DAG) is a directed graph that does not contain any cycles.

	SPRINKLER	
RAIN	T	F
F	0.4	0.6
T	0.01	0.99



RAIN	
T	F
0.2	0.8

		GRASS WET	
SPRINKLER	RAIN	T	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

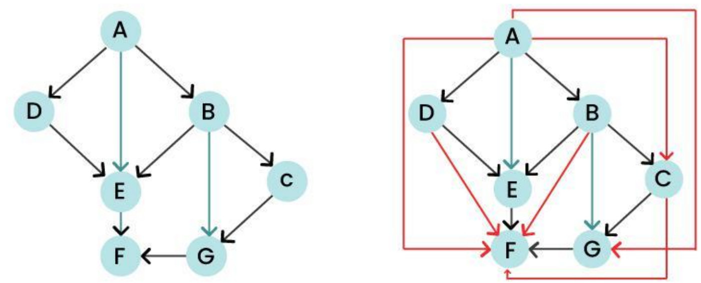


Properties of Directed Acyclic Graph DAG:

Reachability Relation

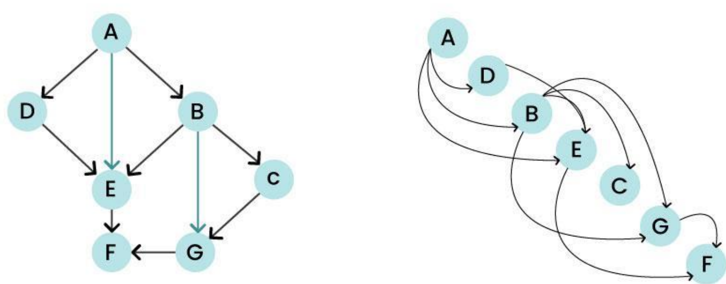
you can follow the direction of edges in the graph to get from B to A

Transitive Closure



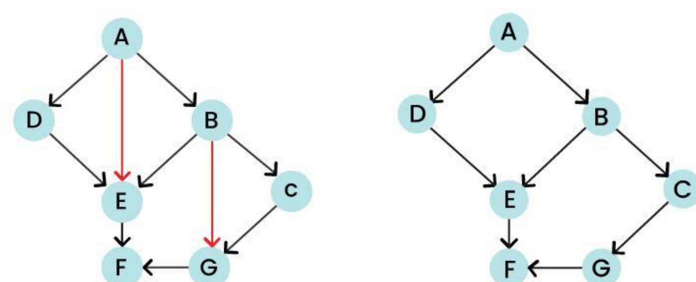
A Directed Acyclic Graph → Its Transitive Closure

Topological Ordering



A Directed Acyclic Graph → Its Topological Sort/Order

Transitive Reduction

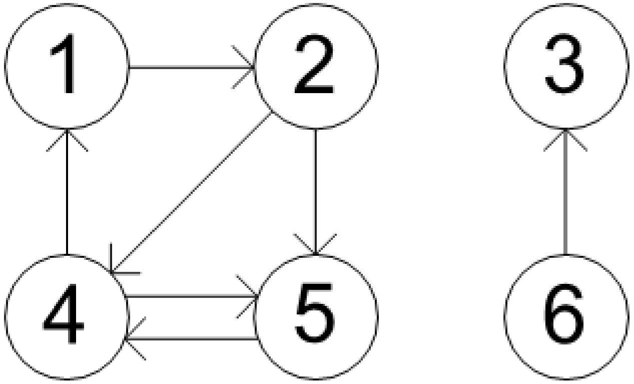


A Directed Acyclic Graph → Its Transitive Reduction

Incidence

Incident from & to: In a directed graph G = (V, E), an edge (u, v) is incident from or leaves u, and is incident to or enters v.

Example: In the digraph

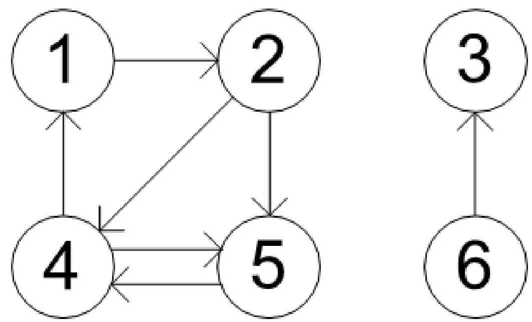


the edge (1, 2) is incident from 1 and is incident to 2

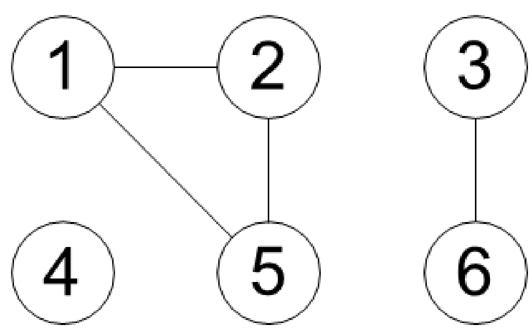
Adjacent

A vertex v is adjacent to u if there is an edge from u to v.

For example, in the digraph below, 2 is adjacent to 1.



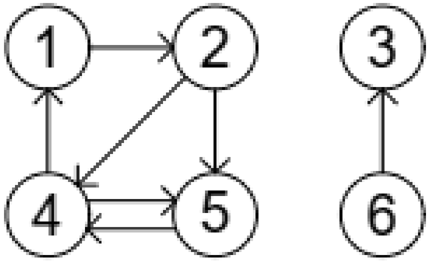
In the undirected graph below, 1 and 2 are adjacent vertices.



Degree of Vertices and Graphs – Continued

In a digraph, the outdegree of a vertex is the number of edges leaving it, and the indegree of a vertex is the number of edges entering it. The degree of a vertex in a directed graph is its indegree plus outdegree.

Example: For the directed graph



we have

indegree(1) = outdegree(1) = 1,

indegree(2) = 1, outdegree(2) = 2,

indegree(5) = 2, outdegree(5) = 1

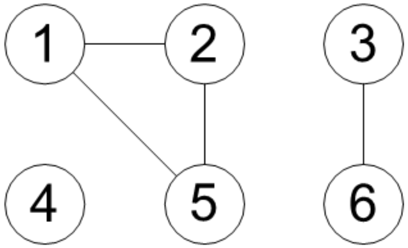
Degree Formula: For a digraph $G = (V, E)$,

$$\sum_{v \in V} \text{indegree}(v) = \sum_{v \in V} \text{outdegree}(v) = |E|.$$

Degree of Vertices and Graphs

Degrees: The degree of a vertex in an undirected graph is the number of edges incident on it, with a loop being counted twice. The (total) degree is the sum of the degrees of all vertices.

Example: For the undirected graph



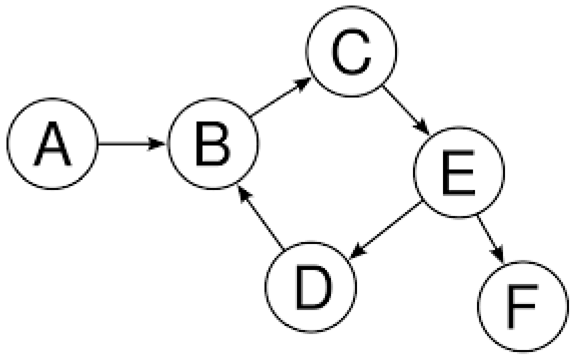
we have

$$\deg(1) = \deg(2) = \deg(5) = 2,$$
$$\deg(3) = \deg(6) = 1, \ \deg(4) = 0.$$

So $\deg(G) = \sum_{i=1}^{6} \deg(i) = 8.$

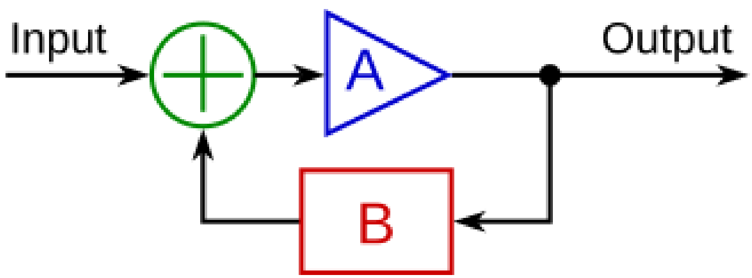
Degree Formula: For a graph $G = (V, E)$, $\deg(G) = \sum_{v \in V} \deg(v) = 2|E|.$

Cyclic Directed Graphs –



Edges have direction with one or more cycles/loops in the graph

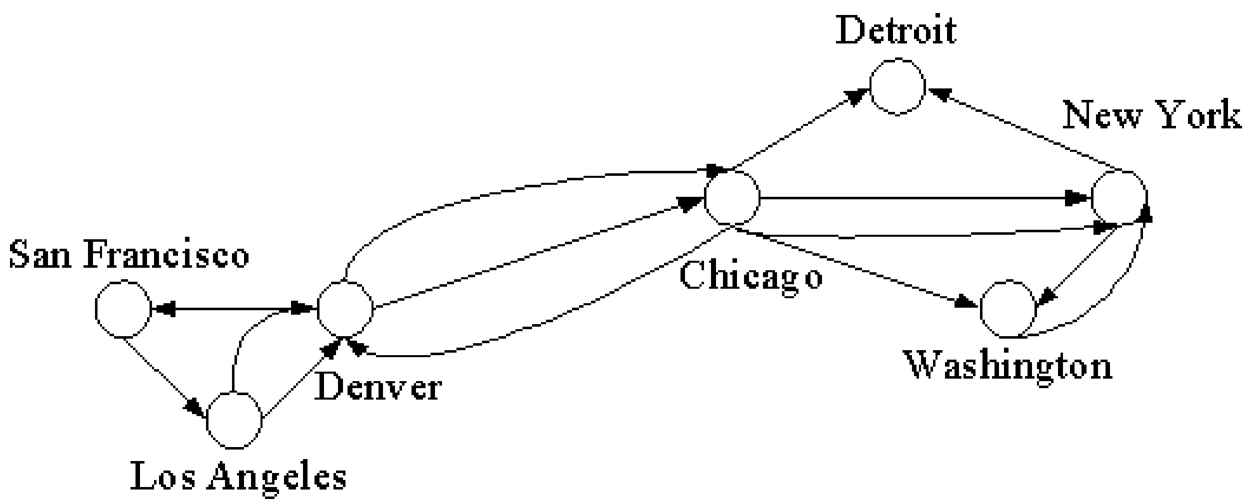
may have multiple cycles of different lengths and shapes. Some cycles may be contained within other cycles



E.g. Feedback loop

Multigraphs –

Directed graphs that allow multiple edges (parallel edges) between the same pair of nodes

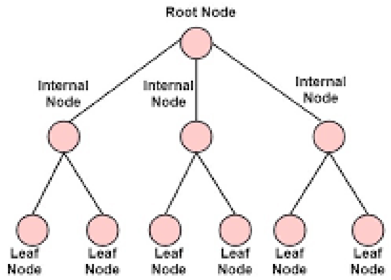


E.g. Transportation system between cities

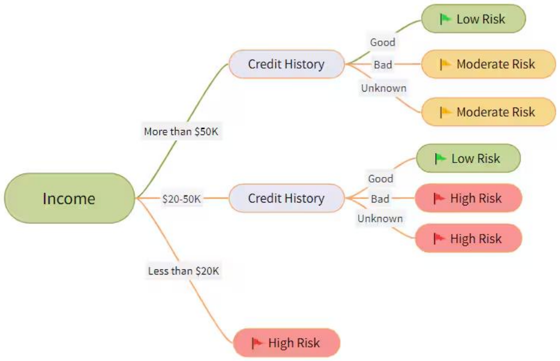
Trees (Rooted Directed Trees) –

A tree is a special type of graph that is connected and contains no cycles.

A rooted tree is a tree graph where one node is explicitly designated as the root, and all edges are directed (or implicitly understood) away from the root



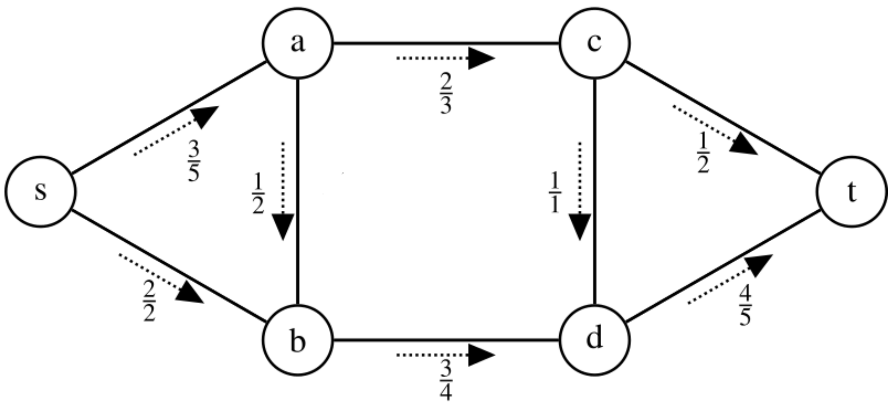
Single root: There exists a unique node designated as the root
Hierarchy: Every node (except the root) has a unique parent
Parent-child relationship: Each edge defines a relationship between parent and child nodes
Depth of a node = number of edges from the root to that node
Height of a tree = maximum depth of any node



E.g. Decision trees

Flow Networks/transportation networks –

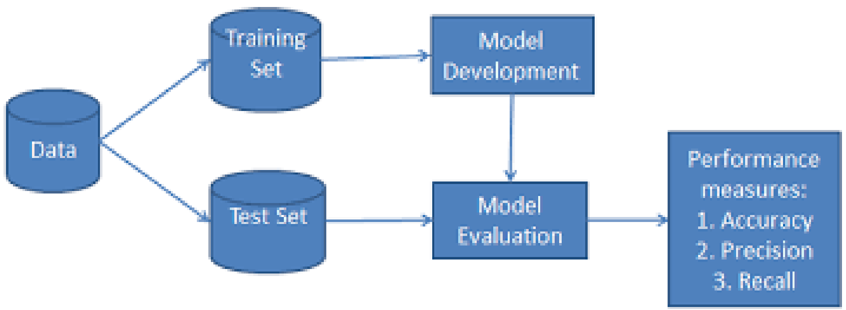
Directed graphs with a source and a sink, used to model the flow of resources through a network.

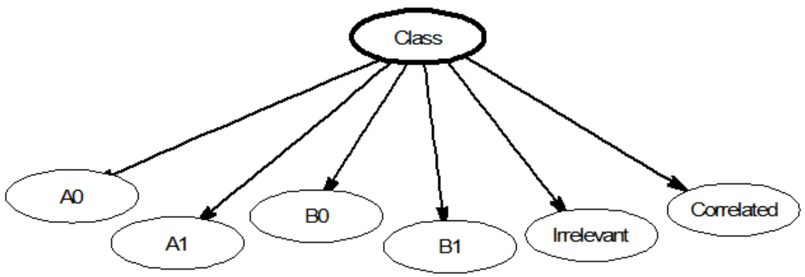


E.g., Water flow networks, flow of electricity in electrical distribution system, where the flow capacity along each edge is considered, Packet transmission in networks.

Graphical models in Bayesian modeling

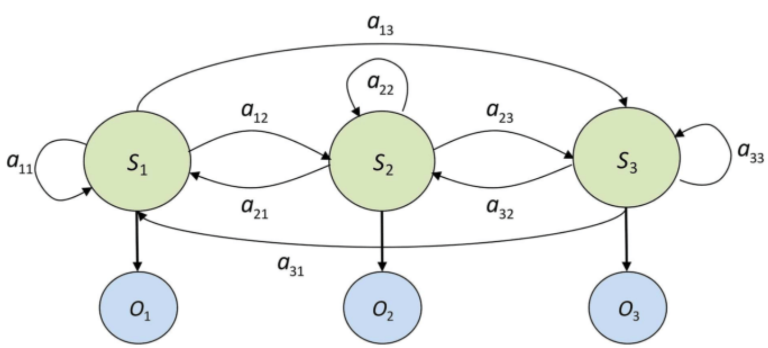
Naive Bayes Classifier - This is a simple graphical model where each feature is conditionally independent of the others given the class label. This structure simplifies the calculation of the posterior probabilities in classification tasks.

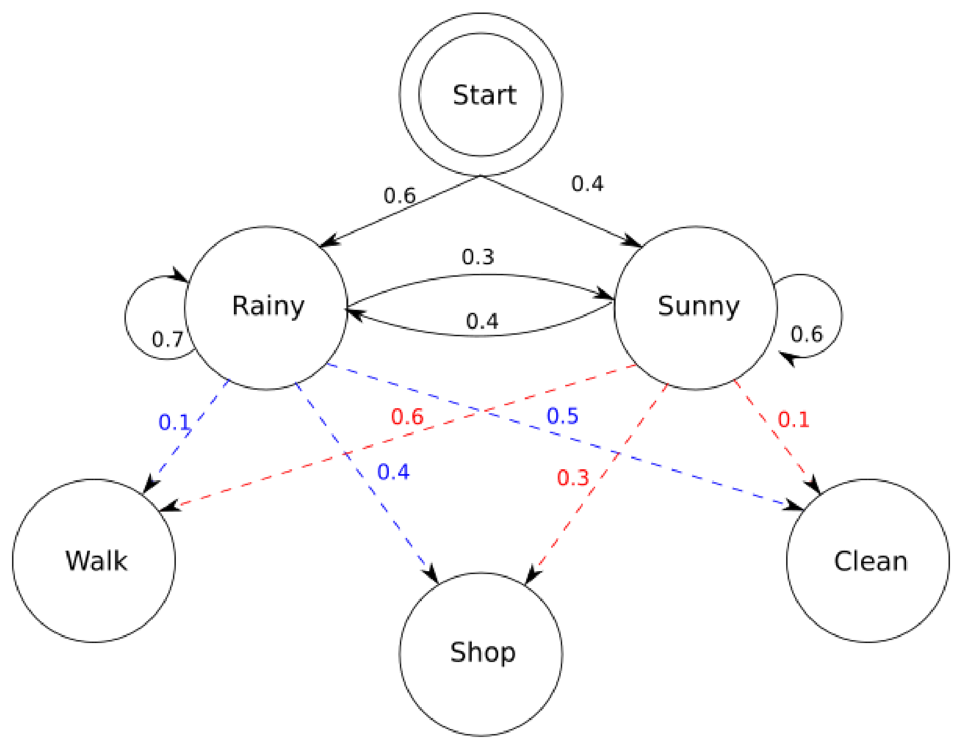




E.g. Spam Filtering: Email systems use Naive Bayes classifiers to filter out spam by analyzing the probability of certain words occurring in spam versus non-spam emails.

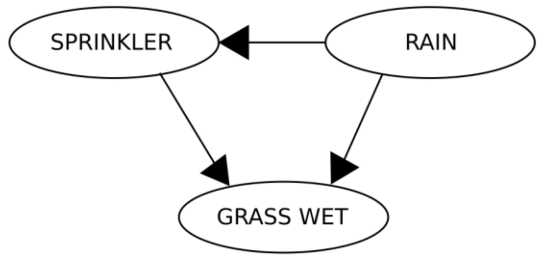
Hidden Markov Models (HMMs) - graphical model to represent sequences of observed data where the current observation depends on a hidden state, and the hidden states form a Markov chain.





Bayesian Networks – These are directed acyclic graphs where nodes represent variables and edges represent probabilistic dependencies.

	SPRINKLER	
RAIN	T	F
F	0.4	0.6
T	0.01	0.99



RAIN	
T	F
0.2	0.8

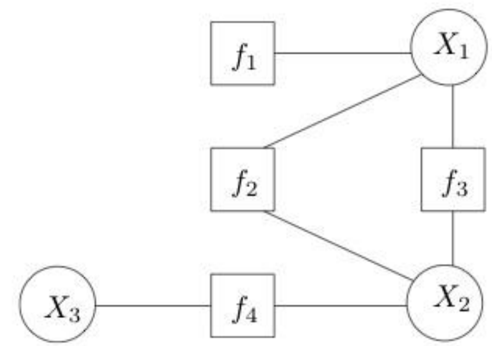
		GRASS WET	
SPRINKLER	RAIN	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

Applications -
In medical diagnostics, Bayesian networks help in assessing the likelihood of diseases based on symptoms and test results.
Bayesian networks are used in finance for credit scoring and risk assessment by modeling the dependencies among financial variables.

Factor Graphs –

Used in more complex models like those involving error-correcting codes or complex inference problems. They decompose a joint probability distribution into factors, which can simplify computations, especially in large networks.

- variables and factors
- variables - unknown quantities
- factors - functions on subsets of the variables
- Edges are always between factors and variables, indicate that a particular factor depends on a particular variable.



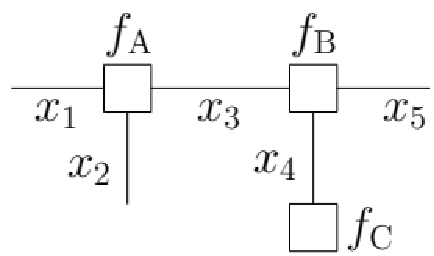
Application –
decoding of error-correcting codes in digital communications.

Factor Graphs –

A factor graph represents the factorization of a function of several variables.

Example:

$$f(x_1, x_2, x_3, x_4, x_5) = f_A(x_1, x_2, x_3) \cdot f_B(x_3, x_4, x_5) \cdot f_C(x_4).$$



Rules:

- A node for every factor.
- An edge or half-edge for every variable.
- Node g is connected to edge x if variable x appears in factor g.