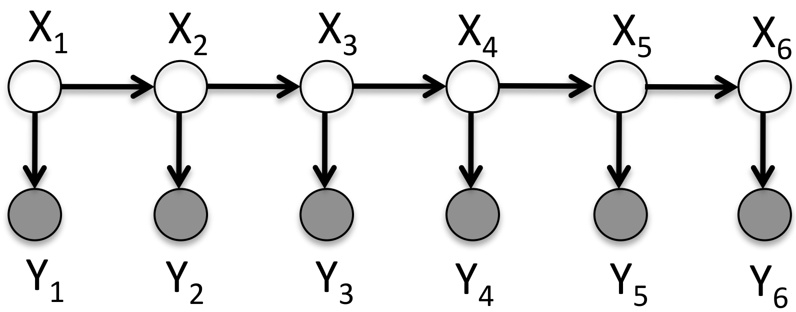
Bayesian networks
Lecture 24

David Sontag
New York University

Hidden Markov models

- We can represent a hidden Markov model with a graph:



Shading in denotes *observed* variables

$$\Pr(x_1, \ldots x_n, y_1, \ldots, y_n) = \Pr(x_1) \Pr(y_1 \mid x_1) \prod_{t=2}^{n} \Pr(x_t \mid x_{t-1}) \Pr(y_t \mid x_t)$$

- There is a 1-1 mapping between the graph structure and the factorization of the joint distribution

Bayesian networks

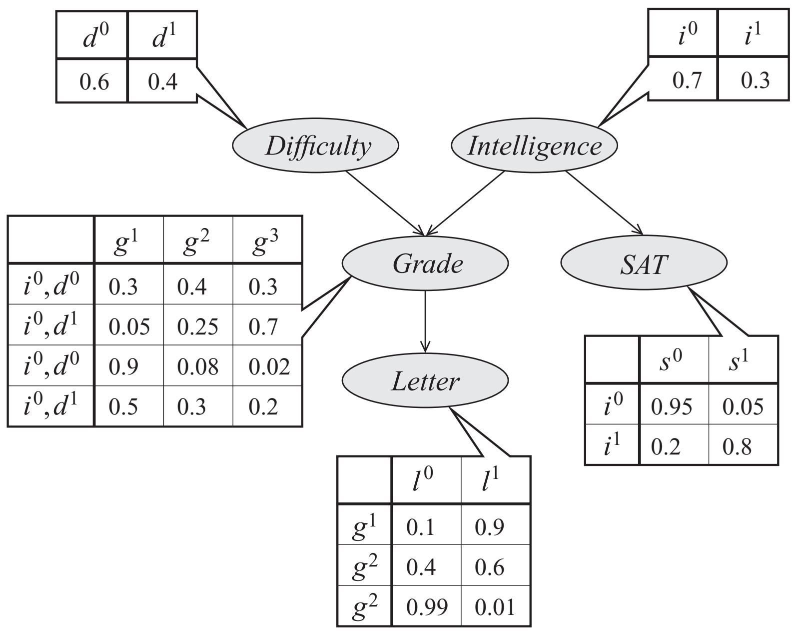
- A **Bayesian network** is specified by a directed *acyclic* graph G=(V,E) with:
  - One node *i* for each random variable X_i
  - One conditional probability distribution (CPD) per node, $p(x_i \mid \mathbf{x}_{Pa(i)})$, specifying the variable's probability conditioned on its parents' values

- Corresponds 1-1 with a particular factorization of the joint distribution:

$$p(x_1, \ldots x_n) = \prod_{i \in V} p(x_i \mid \mathbf{x}_{\mathrm{Pa}(i)})$$

Example

- Consider the following Bayesian network:



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^0, d^0	0.9	0.08	0.02
i^0, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^2	0.99	0.01

- What is its joint distribution?

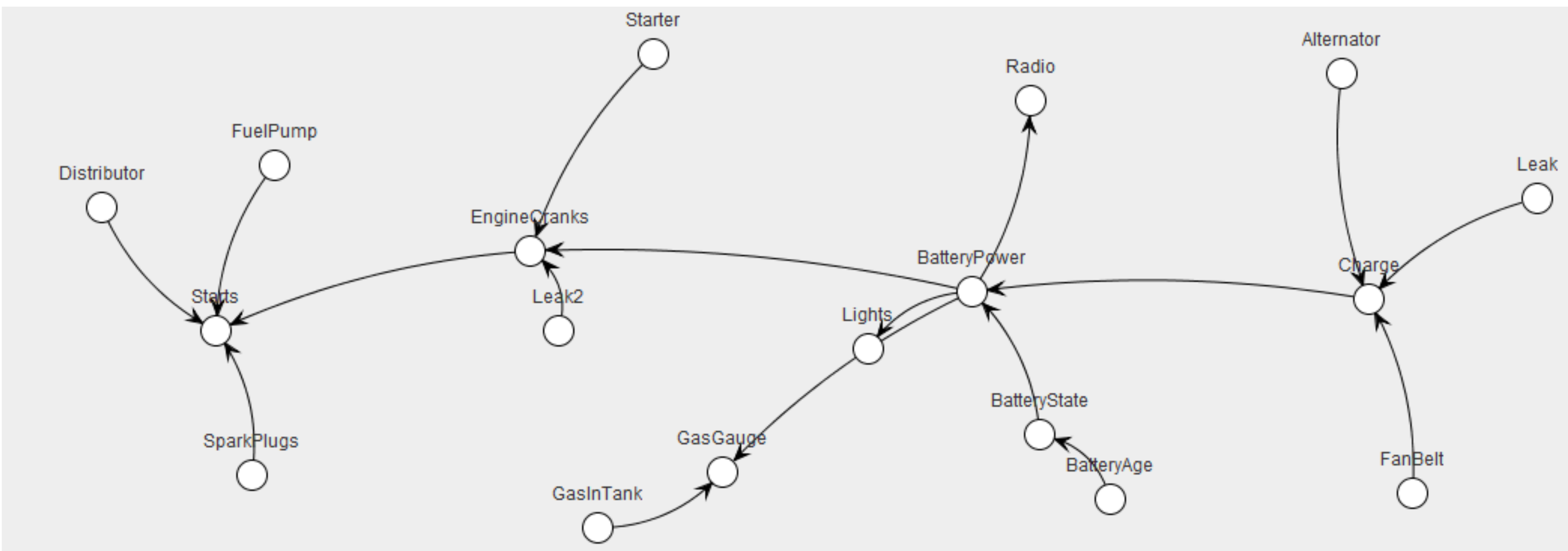
$$p(x_1, \ldots x_n) = \prod_{i \in V} p(x_i \mid \mathbf{x}_{\mathrm{Pa}(i)})$$

$$p(d, i, g, s, l) = p(d)p(i)p(g \mid i, d)p(s \mid i)p(l \mid g)$$

More examples

$$p(x_1, \dots x_n) = \prod_{i \in V} p(x_i \mid \mathbf{x}_{\mathrm{Pa}(i)})$$

Will my car start this morning?



Heckerman *et al.*, Decision-Theoretic Troubleshooting, 1995

More examples

$$p(x_1, \ldots x_n) = \prod_{i \in V} p(x_i \mid \mathbf{x}_{\mathrm{Pa}(i)})$$

What is the differential diagnosis?

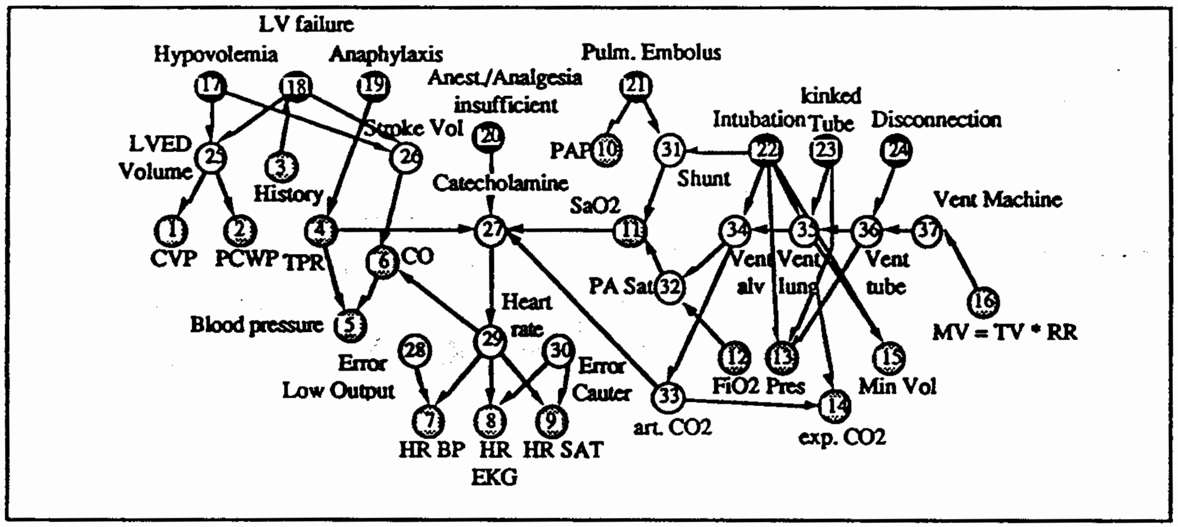
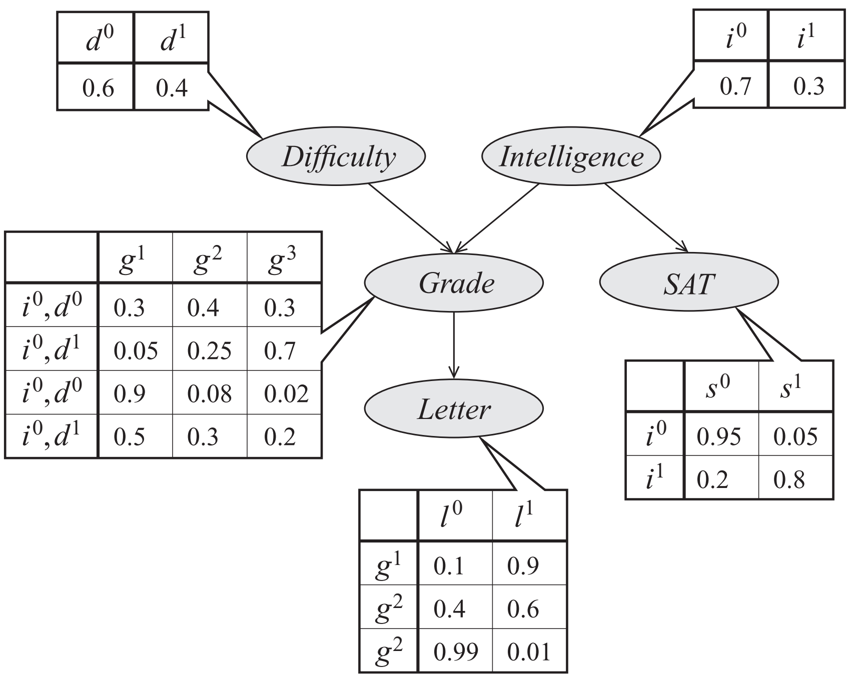


Fig. 1 The ALARM network representing causal relationships is shown with diagnostic (●), intermediate (○) and measurement (◎) nodes. CO: cardiac output, CVP: central venous pressure, LVED volume: left ventricular end-diastolic volume, LV failure: left ventricular failure, MV: minute ventilation, PA Sat: pulmonary artery oxygen saturation, PAP: pulmonary artery pressure, PCWP: pulmonary capillary wedge pressure, Pres: breathing pressure, RR: respiratory rate, TPR: total peripheral resistance, TV: tidal volume

Beinlich *et al.*, The ALARM Monitoring System, 1989

Conditional independencies



The network structure implies several conditional independence statements:

$$D \perp I$$

$$G \perp S \mid I$$

$$D \perp L \mid G$$

$$L \perp S \mid G$$

$$L \perp S \mid I$$

$$D \perp S$$

If two variables are (conditionally) independent, structure has no edge between them

Photo-Essay

BIRTH:

September 4, 1936, Tel Aviv.

EDUCATION:

B.S., Electrical Engineering (Technion, 1960); M.S., Electronics (Newark College of Engineering, 1961); M.S., Physics (Rutgers University, 1965); Ph.D., Electrical Engineering (Polytechnic Institute of Brooklyn, 1965).

EXPERIENCE:

Research Engineer, New York University Medical School (1960–1961); Instructor,

JUDEA PEARL

United States – 2011

CITATION

For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning.

SHORT ANNOTATED BIBLIOGRAPHY | ACM DL AUTHOR PROFILE | ACM TURING AWARD LECTURE VIDEO | RESEARCH SUBJECTS | ADDITIONAL MATERIALS

Judea Pearl created the representational and computational foundation for the processing of information under uncertainty.

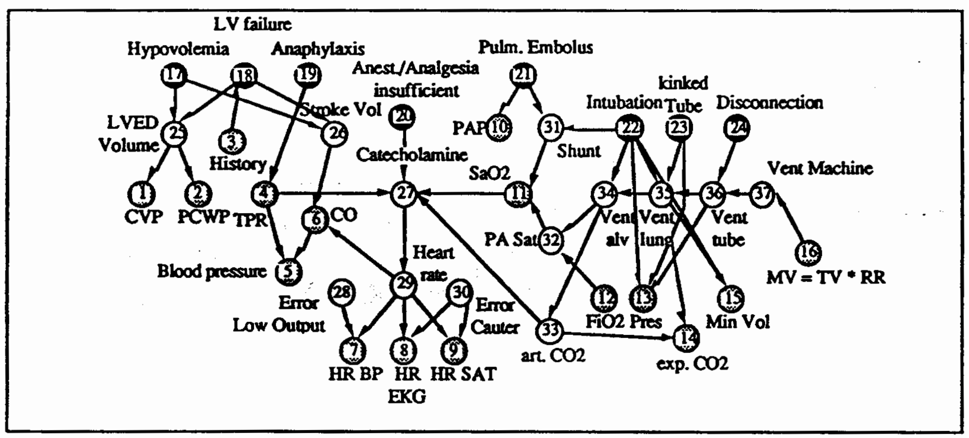
He is credited with the invention of *Bayesian networks*, a mathematical formalism for defining complex probability models, as well as the principal algorithms used for inference in these models. This work not only revolutionized the field of artificial intelligence but also became an important tool for many other branches of engineering and the natural sciences. He later created a mathematical framework for causal inference that has had significant impact in the social sciences.

Judea Pearl was born on September 4, 1936, in Tel Aviv, which was at that time administered under the British Mandate for Palestine. He grew up in Bnei Brak, a Biblical town his grandfather went to reestablish in 1924. In 1956, after serving in the Israeli army and joining a Kibbutz, Judea decided to study engineering. He attended the Technion, where he met his wife, Ruth, and received a B.S. degree in Electrical Engineering in 1960. Recalling the Technion faculty members in a 2012 interview in the *Technion Magazine*, he emphasized the thrill of discovery:

Inference in Bayesian networks

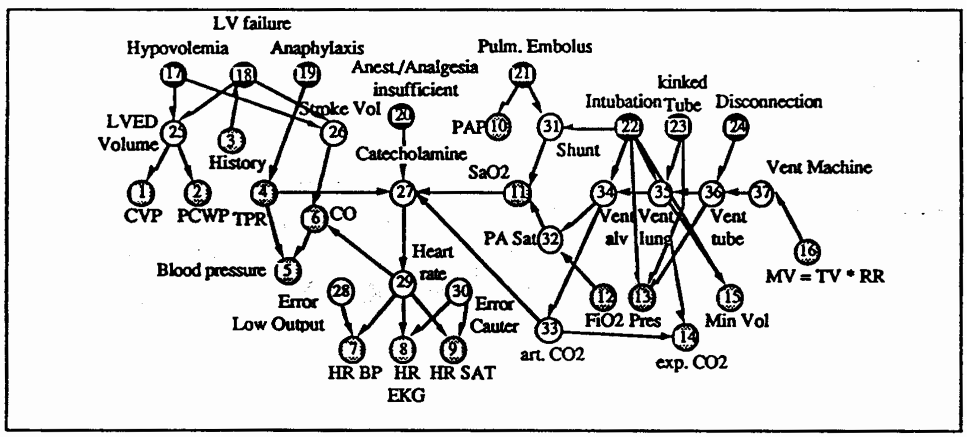
- Computing marginal probabilities in **tree** structured Bayesian networks is easy
  - The algorithm called "belief propagation" generalizes what we showed for hidden Markov models to arbitrary trees

- Wait… this isn't a tree! What can we do?



Inference in Bayesian networks

- In some cases (such as this) we can *transform* this into what is called a “junction tree”, and then run belief propagation



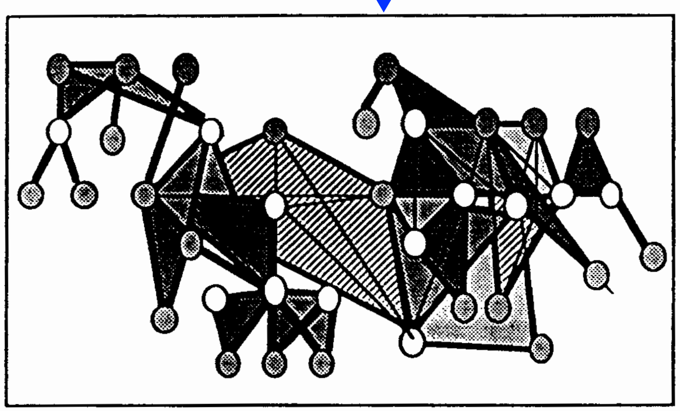
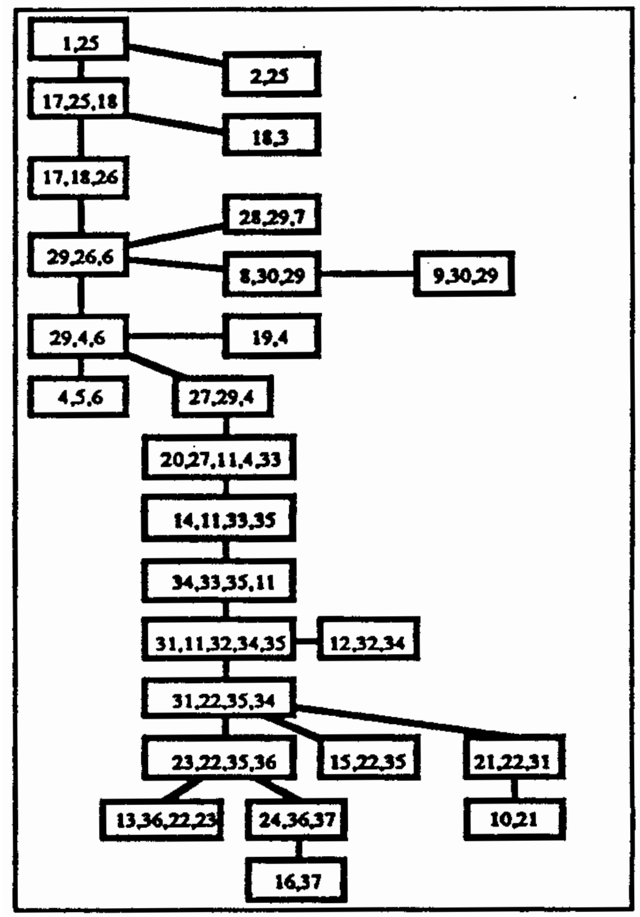


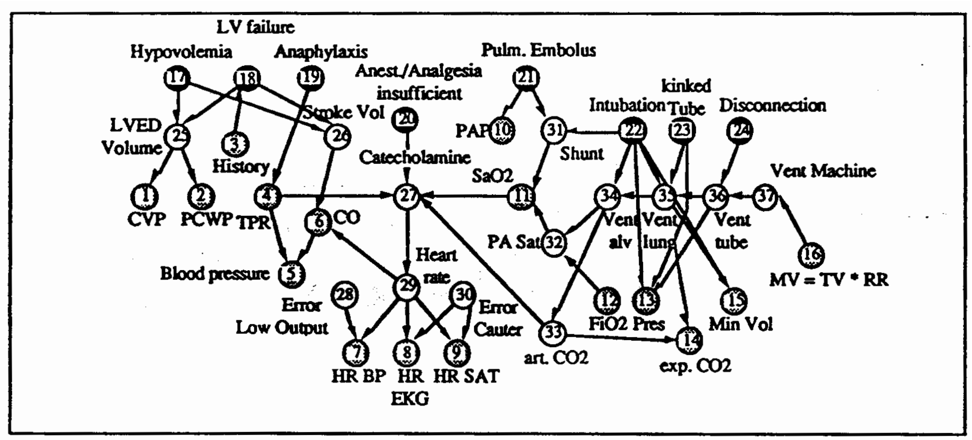
Fig. 7

Spiegelhalter's algorithm rearranges the ALARM network by triangulation and clique formation. The cliques are shaded differently to make them visible.



Approximate inference

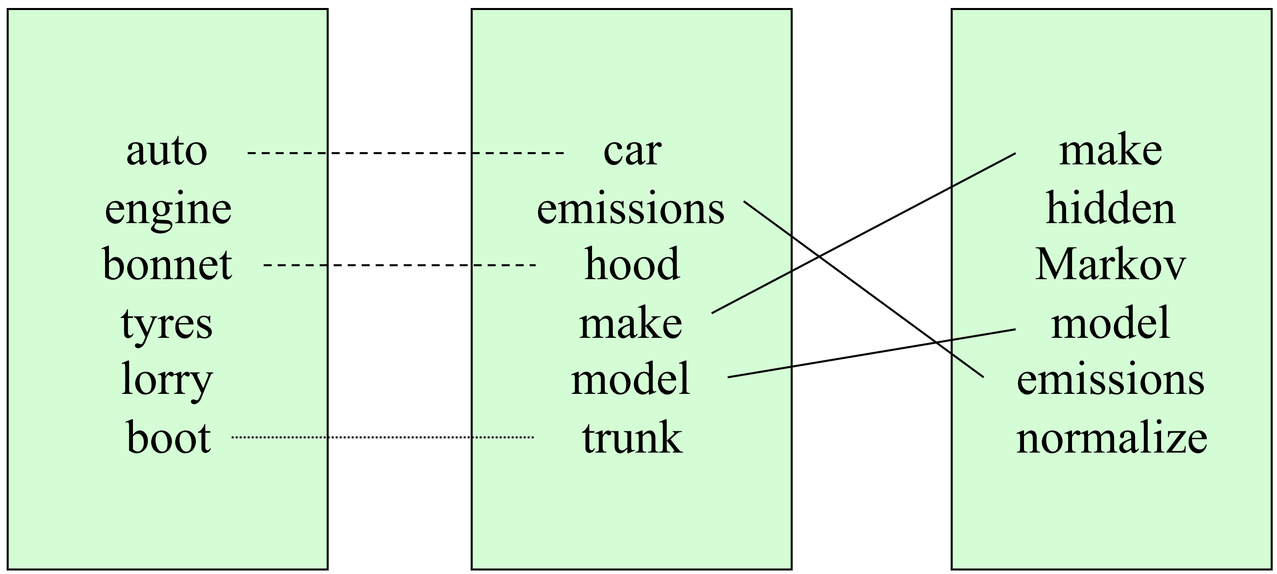
- There is also a wealth of **approximate** inference algorithms that can be applied to Bayesian networks such as these



- Markov chain Monte Carlo algorithms repeatedly sample assignments for estimating marginals
- Variational inference algorithms (which are deterministic) attempt to fit a simpler distribution to the complex distribution, and then computes marginals for the simpler distribution

Dimensionality reduction of text data

- The problem with using a bag of words representation:



Synonymy

Large distance, but related

Polysemy

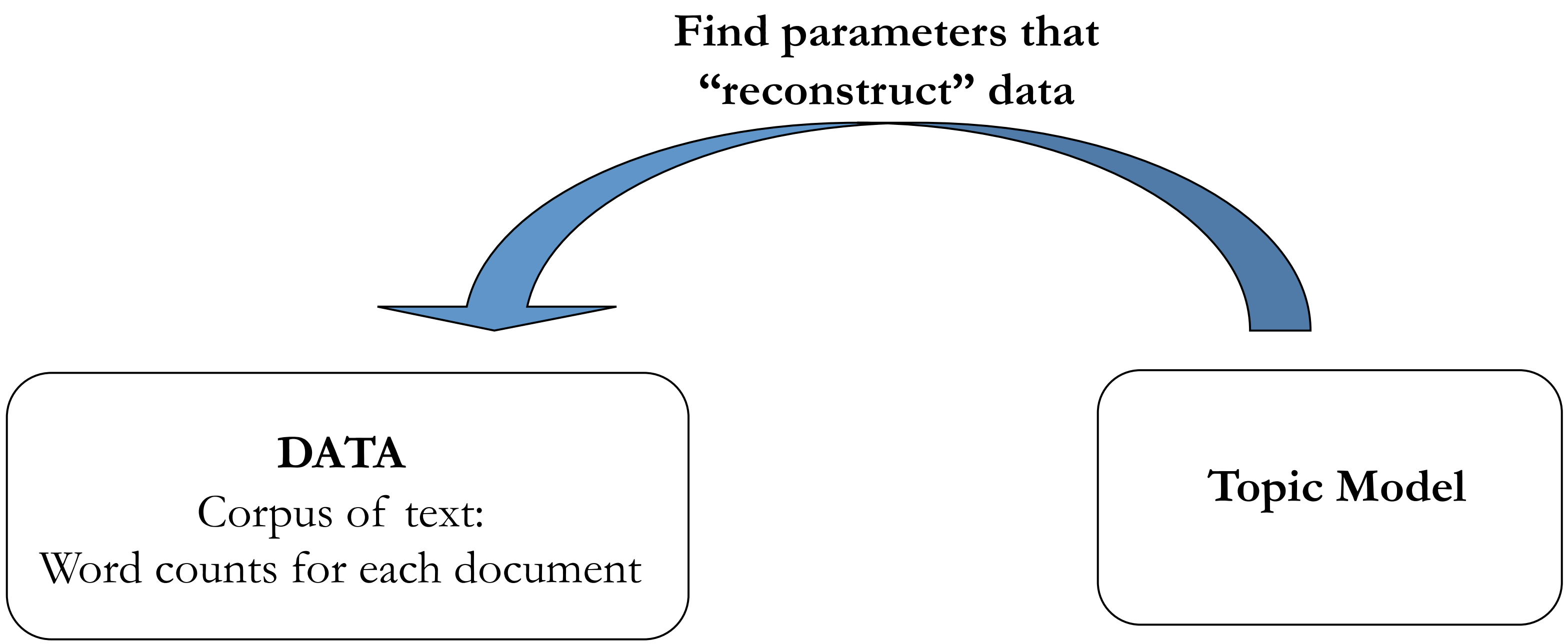
Small distance, but not related

[Example from Lillian Lee]

Probabilistic Topic Models

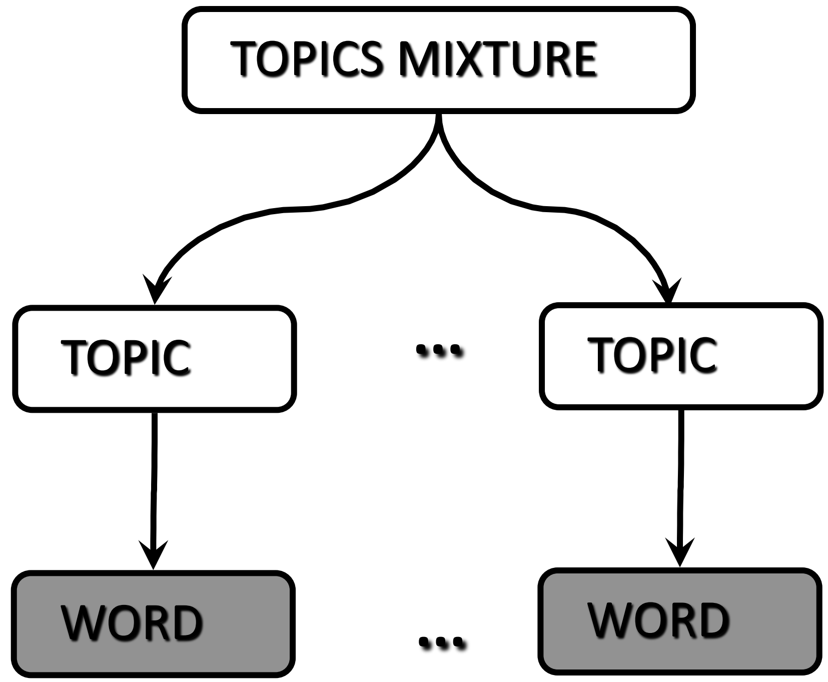
- A probabilistic version of SVD (called LSA when applied to text data)
- Originated in domain of statistics & machine learning
  - (e.g., Hoffman, 2001; Blei, Ng, Jordan, 2003)
- Extracts topics from large collections of text
- Topics are interpretable unlike the arbitrary dimensions of LSA

Model is Generative

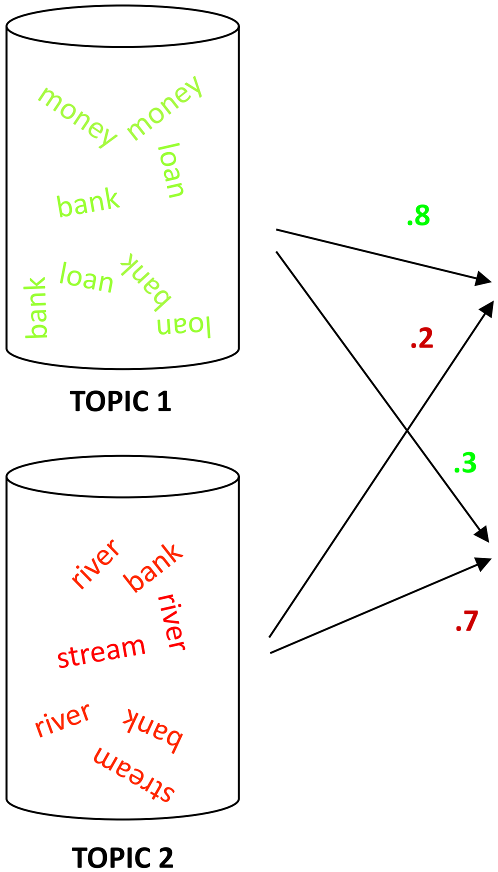


Document generation as a probabilistic process

1. for each document, choose a mixture of topics
2. For every word slot, sample a topic [1..T] from the mixture
3. sample a word from the topic



Example



DOCUMENT 1: money^1 bank^1 bank^1 loan^1 river^2 stream^2 bank^1 money^1 river^2 bank^1 money^1 bank^1 loan^1 money^1 stream^2 bank^1 money^1 bank^1 bank^1 loan^1 river^2 stream^2 bank^1 money^1 river^2 bank^1 money^1 bank^1 loan^1 bank^1 money^1 stream^2

DOCUMENT 2: river^2 stream^2 bank^2 stream^2 bank^2 money^1 loan^1 river^2 stream^2 loan^1 bank^2 river^2 bank^2 bank^1 stream^2 river^2 loan^1 bank^2 stream^2 bank^2 money^1 loan^1 river^2 stream^2 bank^2 stream^2 bank^2 money^1 river^2 stream^2 loan^1 bank^2 river^2 bank^2 money^1 bank^1 stream^2 river^2 bank^2 stream^2 bank^2 money^1

Mixture components

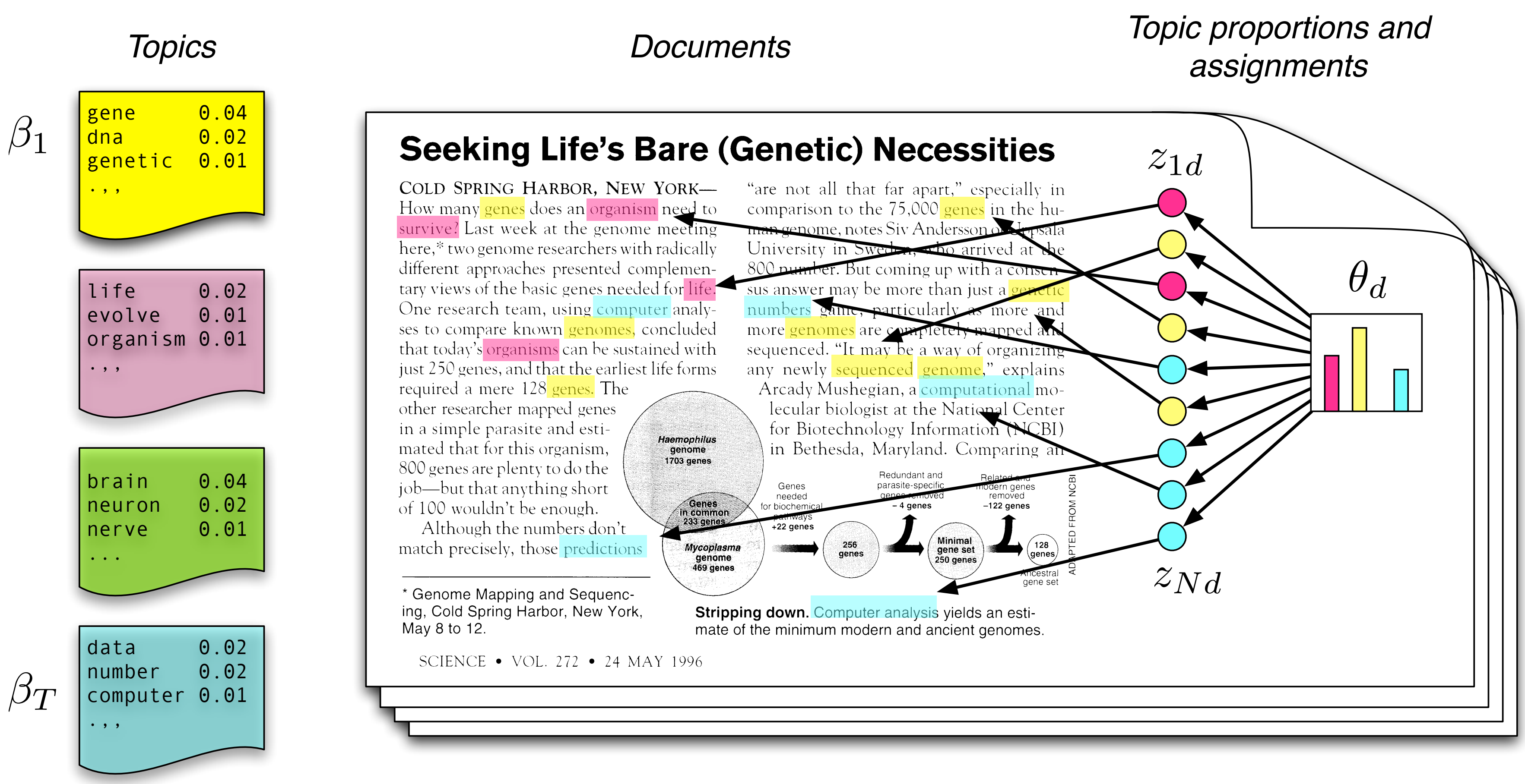
Mixture weights

Bayesian approach: use priors

Mixture weights $\sim$ Dirichlet(α)

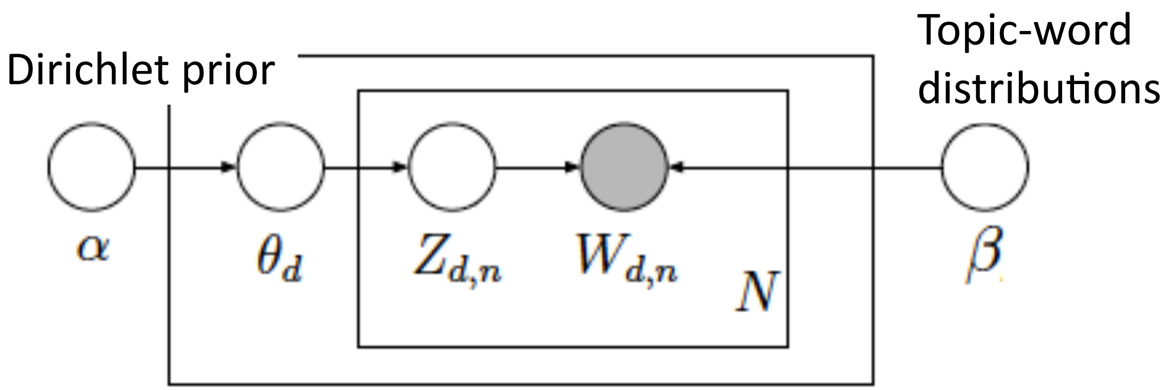
Mixture components $\sim$ Dirichlet(β)

Latent Dirichlet allocation



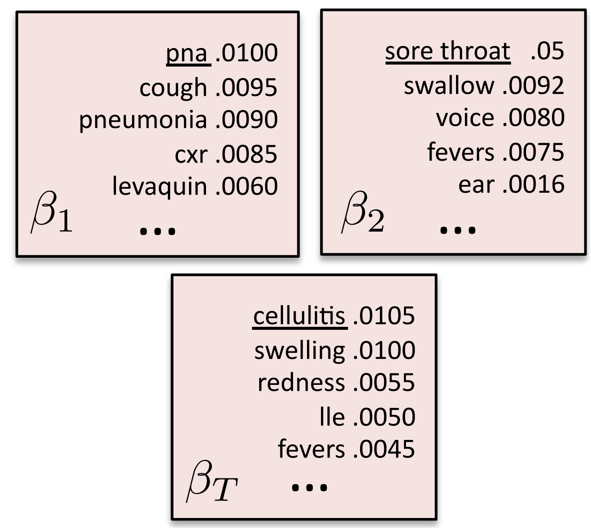
(Blei, Ng, Jordan JMLR '03)

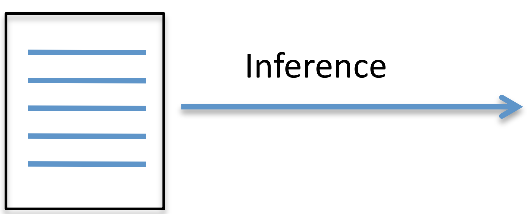
Latent Dirichlet allocation



Graphical model for Latent Dirichlet Allocation (LDA)

Topic word distributions





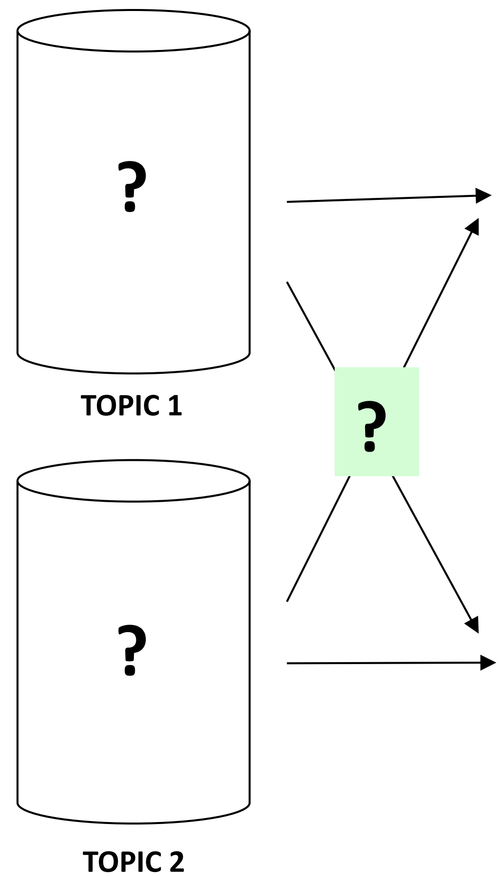
Triage note

	θ_d
Pneumonia	0.50
Common cold	0.49
Diabetes	0.01

Low Dimensional representation: distribution of topics for the note

(Blei, Ng, Jordan JMLR ‘03)

Inverting the model (learning)



DOCUMENT 1: money$^?$ bank$^?$ bank$^?$ loan$^?$ river$^?$ stream$^?$ bank$^?$ money$^?$ river$^?$ bank$^?$ money$^?$ bank$^?$ loan$^?$ money$^?$ stream$^?$ bank$^?$ money$^?$ bank$^?$ bank$^?$ loan$^?$ river$^?$ stream$^?$ bank$^?$ money$^?$ river$^?$ bank$^?$ money$^?$ bank$^?$ loan$^?$ bank$^?$ money$^?$ stream$^?$

DOCUMENT 2: river$^?$ stream$^?$ bank$^?$ stream$^?$ bank$^?$ money$^?$ loan$^?$ river$^?$ stream$^?$ loan$^?$ bank$^?$ river$^?$ bank$^?$ bank$^?$ stream$^?$ river$^?$ loan$^?$ bank$^?$ stream$^?$ bank$^?$ money$^?$ loan$^?$ river$^?$ stream$^?$ bank$^?$ stream$^?$ bank$^?$ money$^?$ river$^?$ stream$^?$ loan$^?$ bank$^?$ river$^?$ bank$^?$ money$^?$ bank$^?$ stream$^?$ river$^?$ bank$^?$ stream$^?$ bank$^?$ money$^?$

Mixture components

Mixture weights

Example of learned representation

Paraphrased note:

*"Patient has **URI [upper respiratory infection] symptoms** like **cough, runny nose, ear pain**. **Denies fevers**. **history of seasonal allergies**"*

Cold/URI

Allergy

Inferred Topic Distribution