# Structured Probabilistic Models for Deep Learning

Lecture slides for Chapter 16 of *Deep Learning*

www.deeplearningbook.org

Ian Goodfellow

2016-10-04

# Roadmap

- Challenges of Unstructured Modeling
- Using Graphs to Describe Model Structure
- Sampling from Graphical Models
- Advantages of Structured Modeling
- Structure Learning and Latent Variables
- Inference and Approximate Inference
- The Deep Learning Approach to Structured Probabilistic Modeling

# Tasks for Generative Models

- Density estimation
- Denoising
- Sample generation
- Missing value imputation
  - Conditional sample generation
  - Conditional density estimation

(Goodfellow 2017)

# Samples from a BEGAN

(Berthelot et al, 2017)

Images are 128 pixels wide, 128 pixels tall

R, G, and B pixel at each location.

# Cost of Tabular Approach

$$k^n$$

Number of variables
For BEGAN faces:
$128 \times 128 = 16384$

Number of values per variable
For BEGAN faces: 256

There are roughly *ten to the power of forty thousand* times more points in the discretized domain of the BEGAN face model than there are atoms in the universe.

(Goodfellow 2017)

# Tabular Approach is Infeasible

- Memory: cannot store that many parameters
- Runtime: inference and sampling are both slow
- Statistical efficiency: extremely high number of parameters requires extremely high number of training examples

# Roadmap

- Challenges of Unstructured Modeling
- Using Graphs to Describe Model Structure
- Sampling from Graphical Models
- Advantages of Structured Modeling
- Structure Learning and Latent Variables
- Inference and Approximate Inference
- The Deep Learning Approach to Structured Probabilistic Modeling

# Insight of Model Structure

- Most variables influence each other
- Most variables do not influence each other *directly*
- Describe influence with a graph
    - Edges represent *direct influence*
    - Paths represent *indirect influence*
- Computational and statistical savings come from *omissions of edges*

# Directed Models

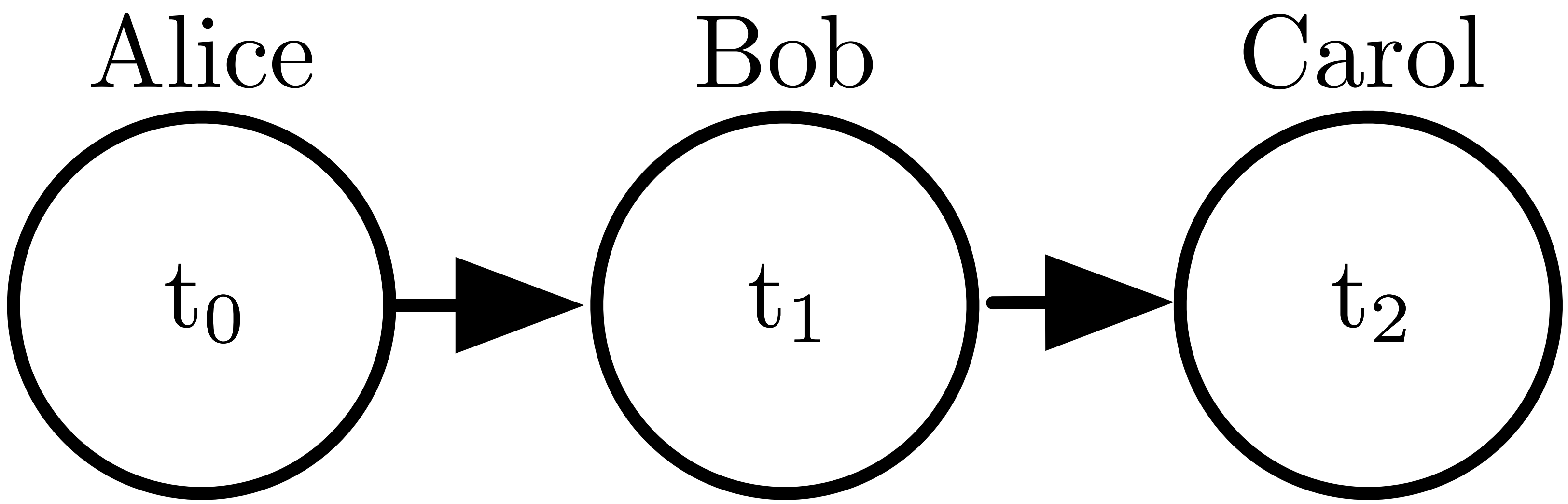



Figure 16.2

$$p(\mathbf{x}) = \Pi_i p(\mathrm{x}_i \mid Pa_{\mathcal{G}}(\mathrm{x}_i)). \tag{16.1}$$

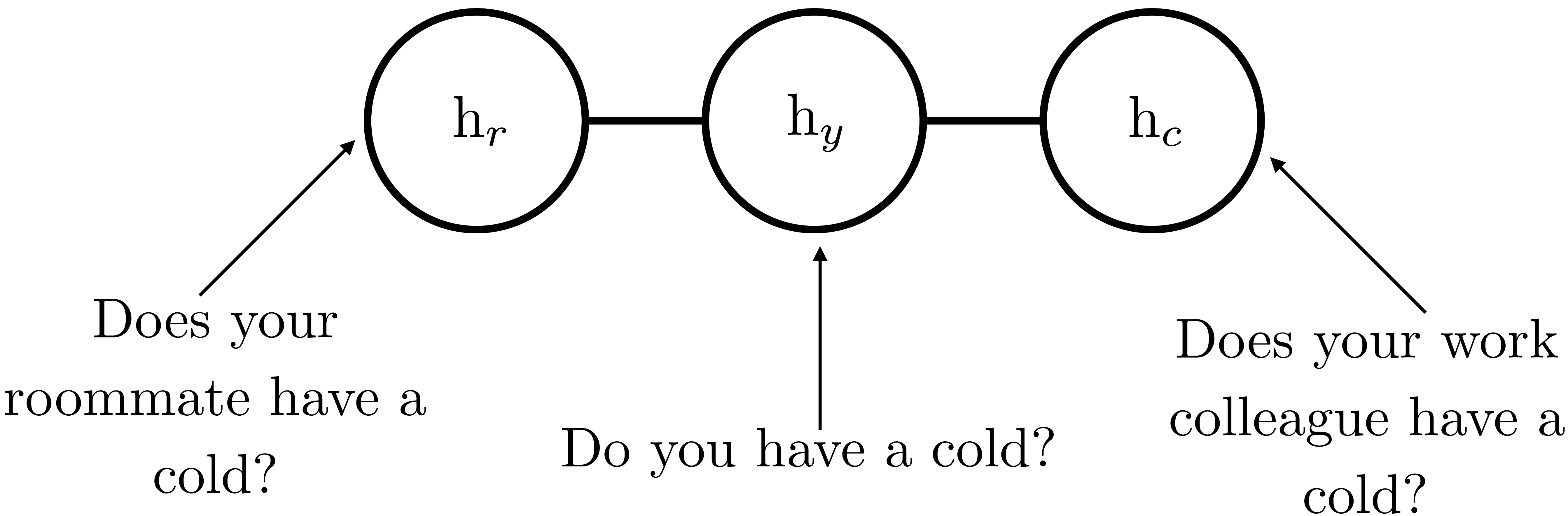
$$p(\mathrm{t}_0, \mathrm{t}_1, \mathrm{t}_2) = p(\mathrm{t}_0)p(\mathrm{t}_1 \mid \mathrm{t}_0)p(\mathrm{t}_2 \mid \mathrm{t}_1). \tag{16.2}$$

Directed models work best when influence clearly flows in one direction

(Goodfellow 2017)

# Undirected Models

Undirected models work best when influence has no clear direction or is best modeled as flowing in both directions




(Goodfellow 2017)

# Undirected Models

Unnormalized probability

$$\tilde{p}(\mathbf{x}) = \Pi_{\mathcal{C}\in\mathcal{G}}\phi(\mathcal{C}). \tag{16.3}$$

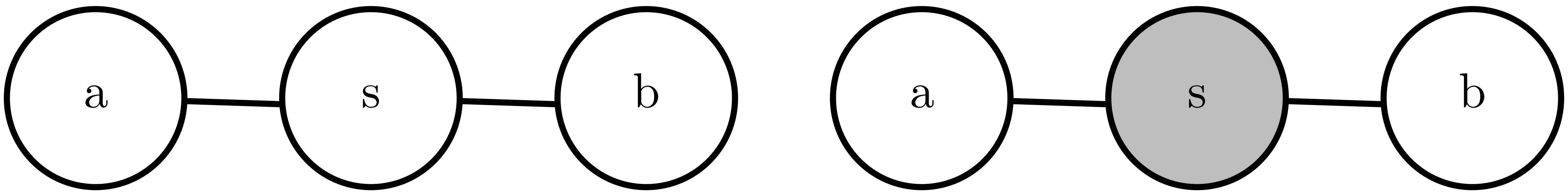
$$p(\mathbf{x}) = \frac{1}{Z}\tilde{p}(\mathbf{x}), \tag{16.4}$$

Partition function

$$Z = \int \tilde{p}(\mathbf{x})d\mathbf{x}. \tag{16.5}$$

(Goodfellow 2017)

# Separation

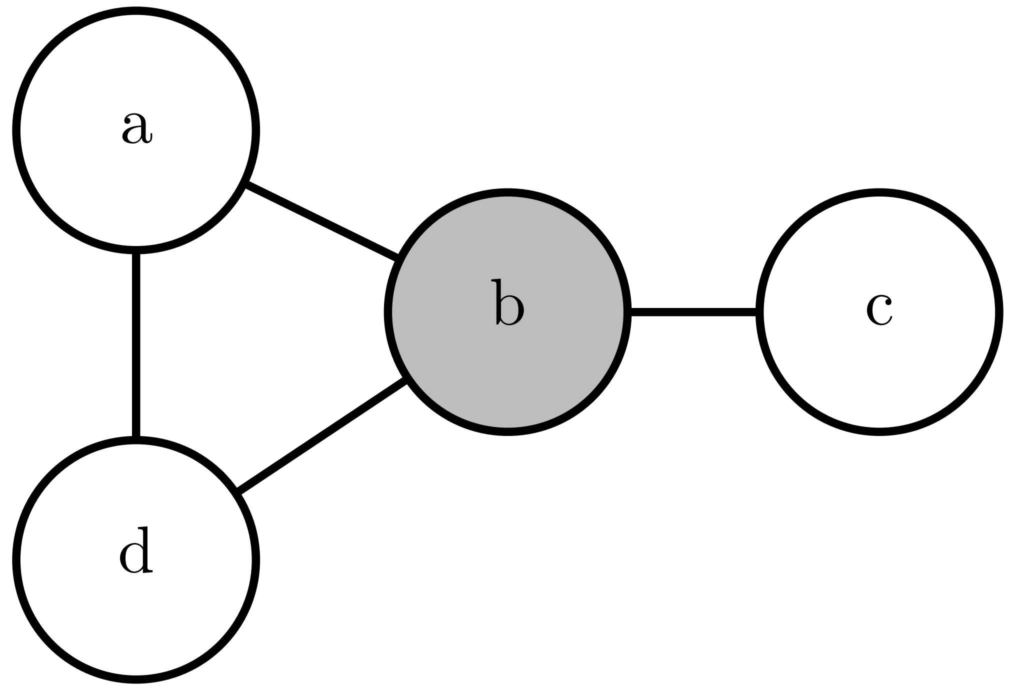



When s is not observed, influence can flow from a to b and vice versa through s.

When s is observed, it blocks the flow of influence between a and b: they are *separated*

(Goodfellow 2017)

# Separation example



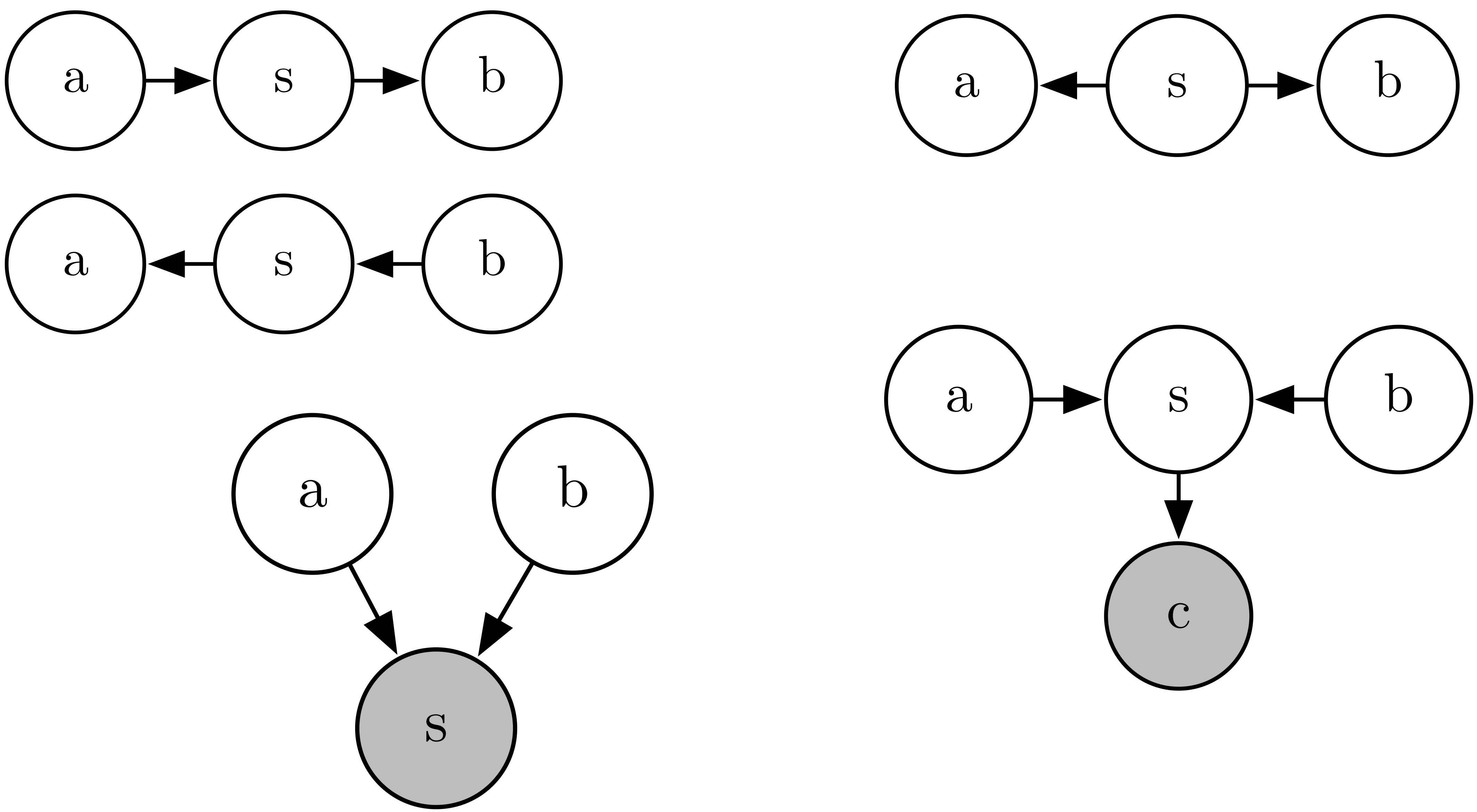

The nodes a and c are separated

One path between a and d is still active, though the other path is blocked, so these two nodes are not separated.

# d-separation

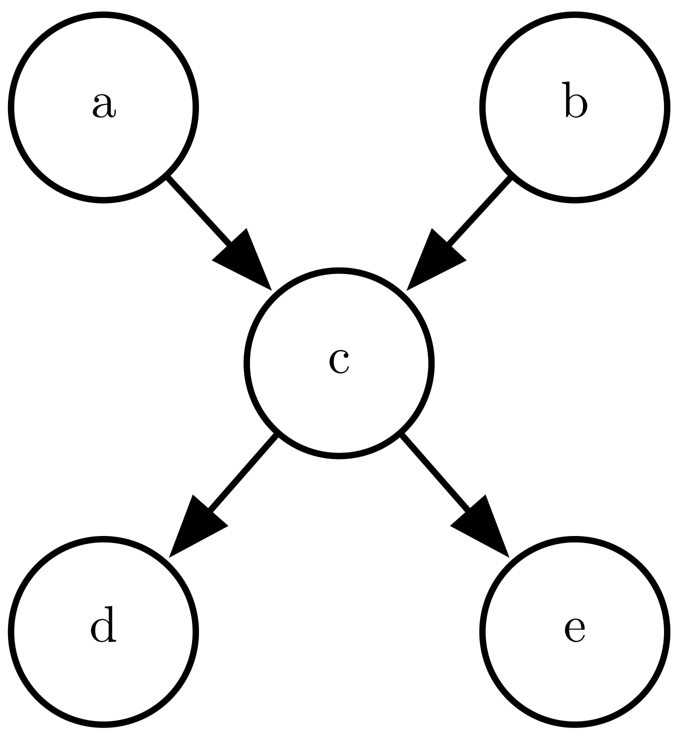
The flow of influence is more complicated for directed models

The path between a and b is active for all of these graphs:




(Goodfellow 2017)

# d-separation example



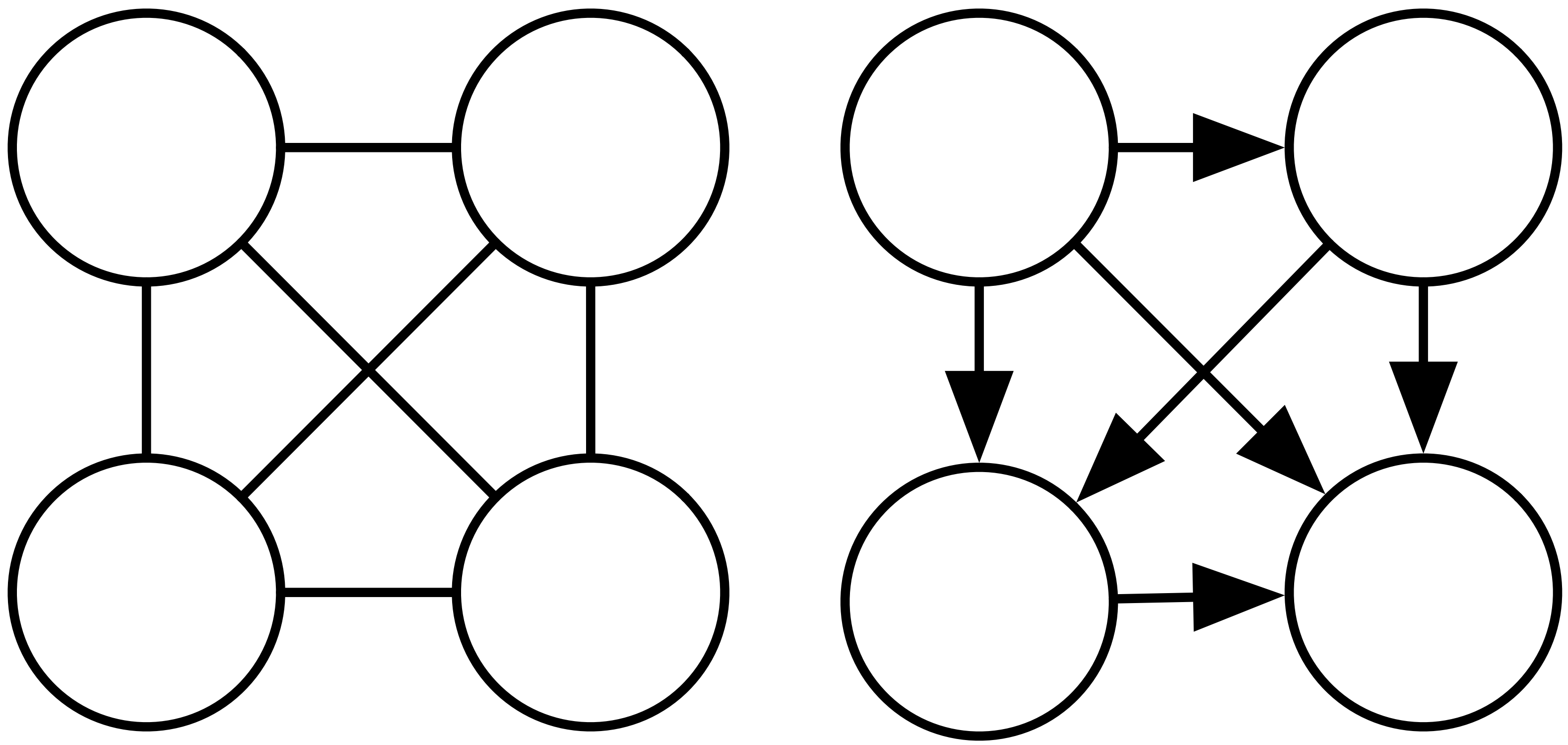

- a and b are d-separated given the empty set
- a and e are d-separated given c
- d and e are d-separated given c

Observing variables can activate paths!

- a and b are not d-separated given c
- a and b are not d-separated given d

(Goodfellow 2017)

# A complete graph can represent any probability distribution



The benefits of graphical models come from *omitting* edges

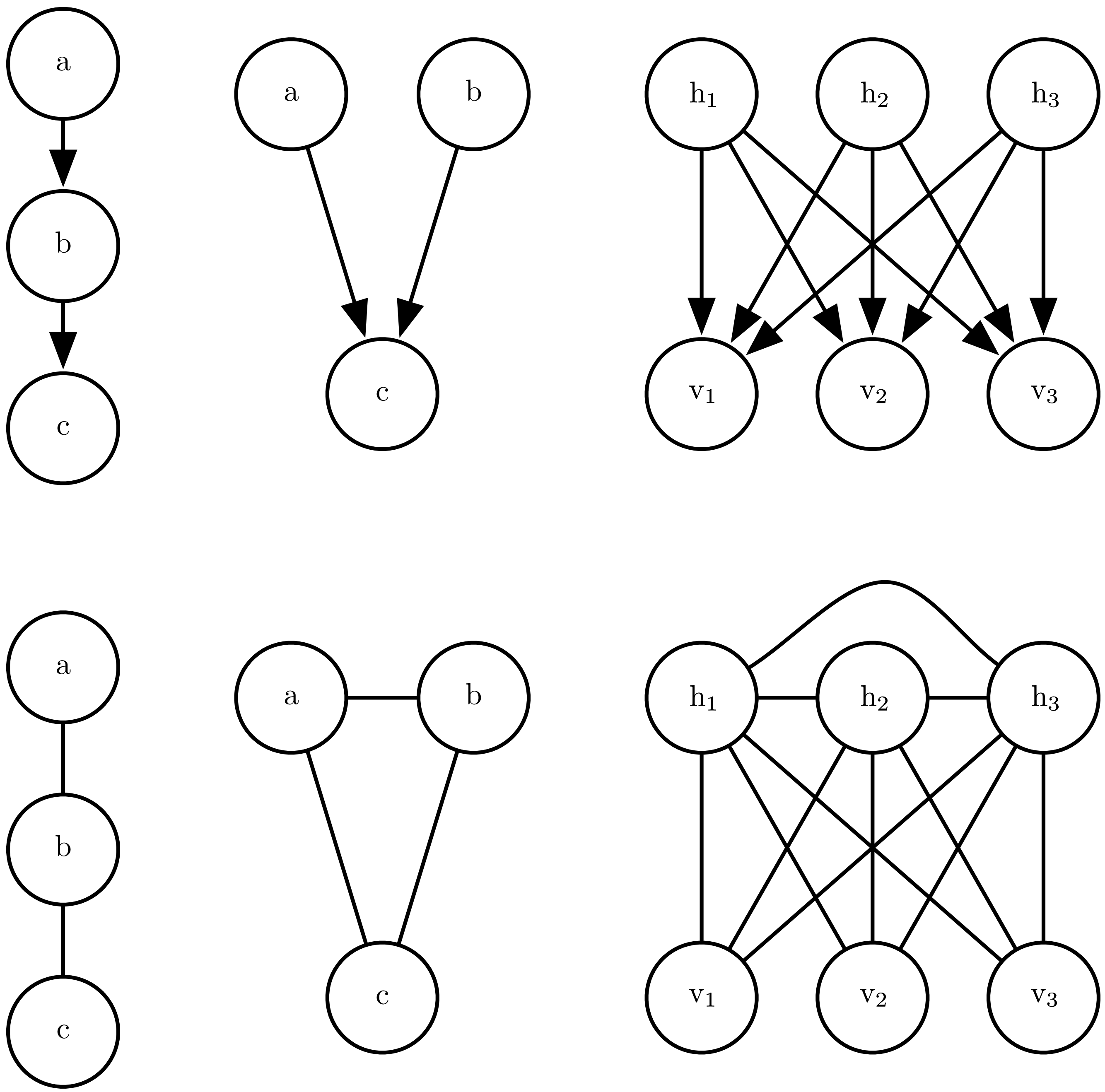
(Goodfellow 2017)

# Converting between graphs

- Any specific probability distribution can be represented by either an undirected or a directed graph

- Some probability distributions have conditional independences that one kind of graph fails to imply (the distribution is simpler than the graph describes; need to know the conditional probability distributions to see the independences)

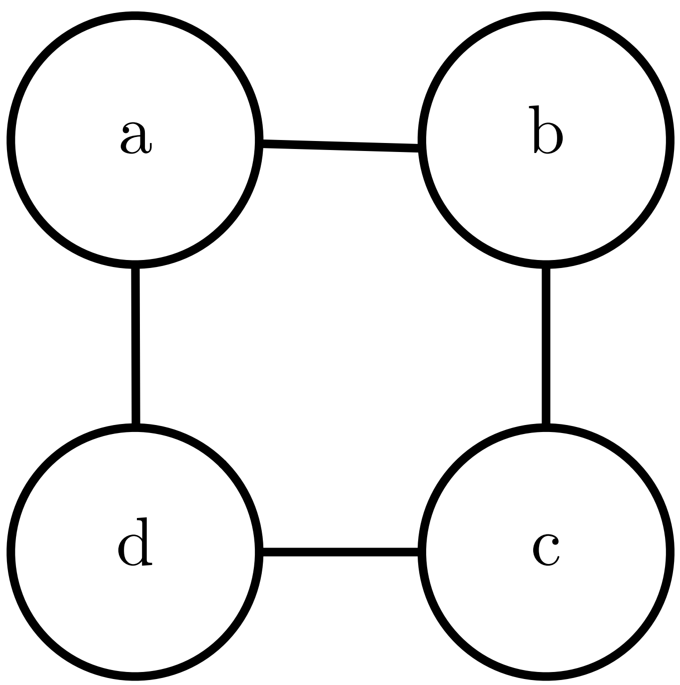
# Converting directed to undirected



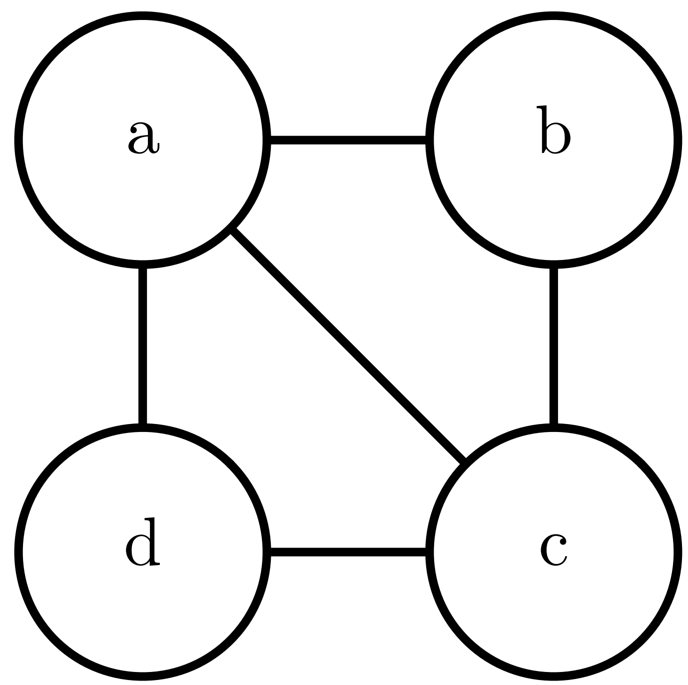

Must add an edge between unconnected coparents

(Goodfellow 2017)

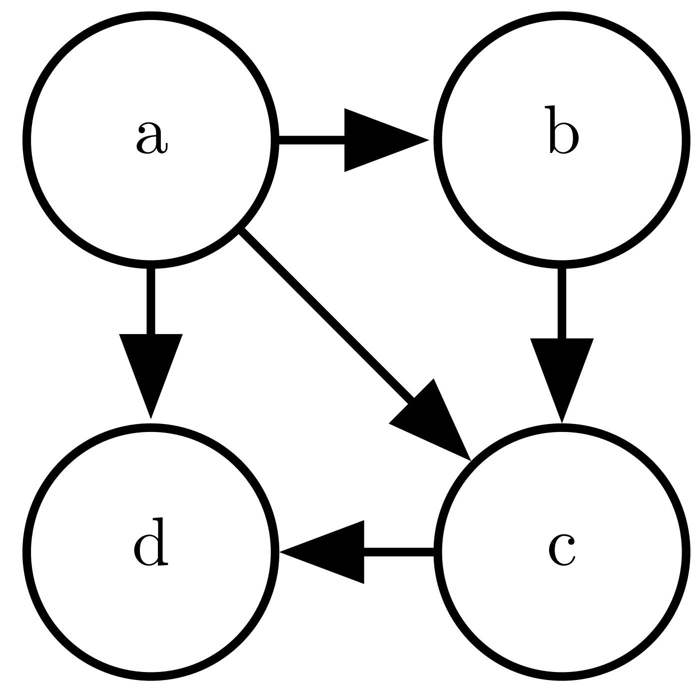
# Converting undirected to directed




No loops of length greater than three allowed!



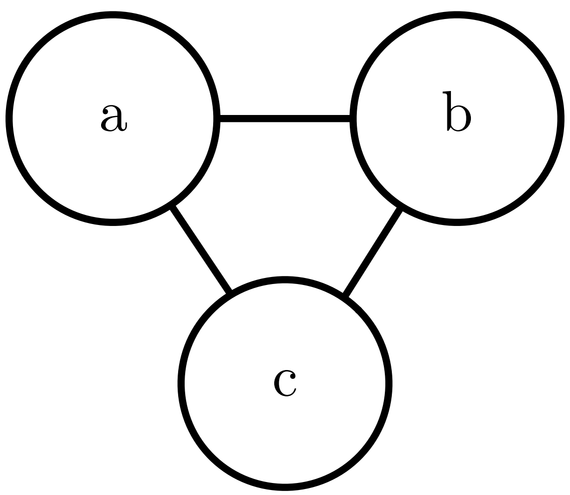

Add edges to triangulate long loops



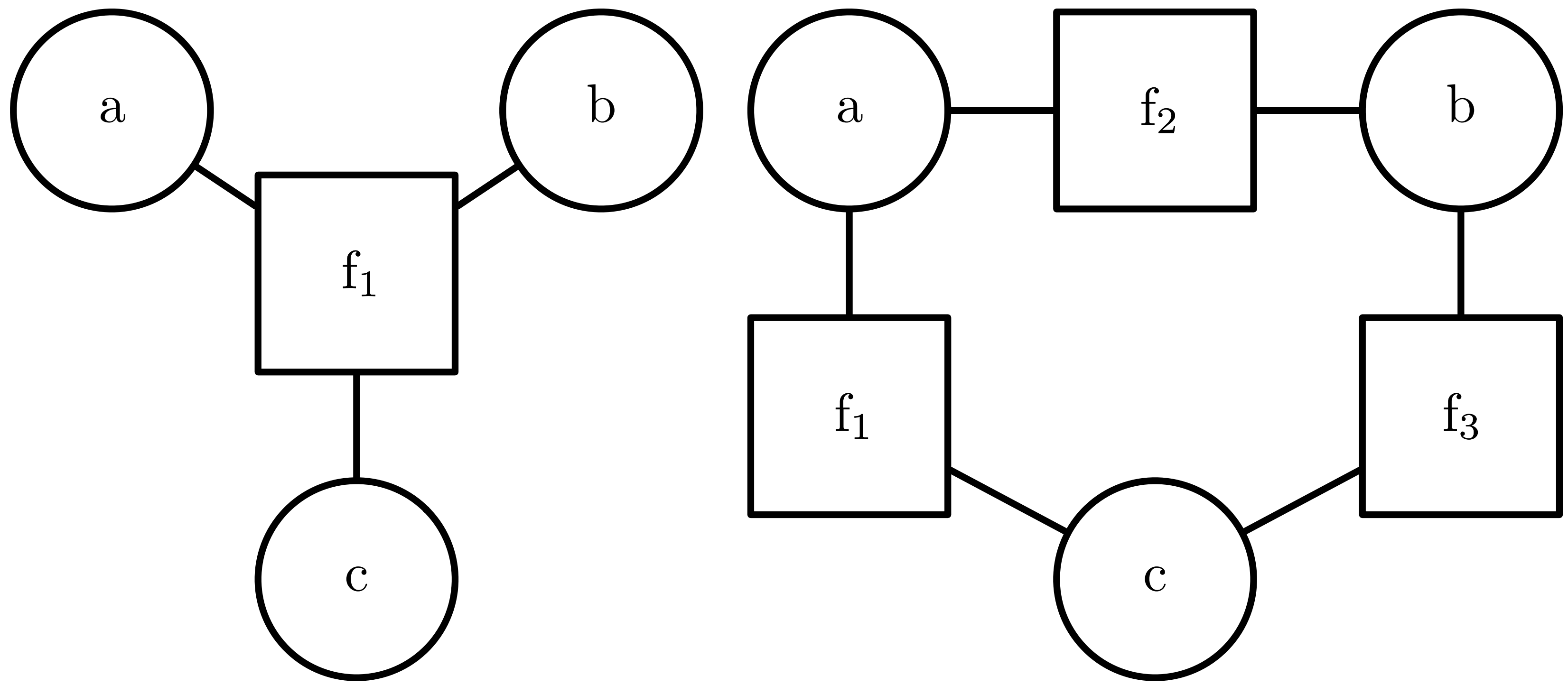

Assign directions to edges. No directed cycles allowed.

(Goodfellow 2017)

# Factor graphs are less ambiguous




Undirected graph: is this three pairwise potentials or one potential over three variables?




Factor graphs disambiguate by placing each potential in the graph

(Goodfellow 2017)

# Roadmap

- Challenges of Unstructured Modeling
- Using Graphs to Describe Model Structure
- Sampling from Graphical Models
- Advantages of Structured Modeling
- Structure Learning and Latent Variables
- Inference and Approximate Inference
- The Deep Learning Approach to Structured Probabilistic Modeling

# Sampling from directed models

- Easy and fast to draw fair samples from the whole model

- *Ancestral sampling*: pass through the graph in topological order. Sample each node given its parents.

- Harder to sample some nodes given other nodes, unless the observed nodes are at the start of the topology

# Sampling from undirected models

- Usually requires Markov chains
- Usually cannot be done exactly
- Usually requires multiple iterations even to approximate
- Described in Chapter 17

# Roadmap

- Challenges of Unstructured Modeling
- Using Graphs to Describe Model Structure
- Sampling from Graphical Models
- Advantages of Structured Modeling
- Structure Learning and Latent Variables
- Inference and Approximate Inference
- The Deep Learning Approach to Structured Probabilistic Modeling

# Tabular Case

- Assume each node has a tabular distribution given its parents
- Memory, sampling, inference are now exponential in *number of variables in factor with largest scope*
  - For many interesting models, this is very small
  - e.g., RBMs: all factor scopes are size 2 or 1
- Previously, these costs were exponential in *total number of nodes*
- Statistically, much easier to estimate this manageable number of parameters

# Roadmap

- Challenges of Unstructured Modeling
- Using Graphs to Describe Model Structure
- Sampling from Graphical Models
- Advantages of Structured Modeling
- Structure Learning and Latent Variables
- Inference and Approximate Inference
- The Deep Learning Approach to Structured Probabilistic Modeling

# Learning about dependencies

- Suppose we have thousands of variables
  - Maybe gene expression data
- Some interact
- Some do not
- We do not know which ahead of time

# Structure learning strategy

- Try out several graphs
- See which graph does best job of some criterion
  - Fitting training set with small model complexity
  - Fitting validation set
- Iterative search, propose new graphs similar to best graph so far (remove edge / add edge / flip edge)

# Latent variable strategy

- Use one graph structure
- Many latent variables
- Dense connections of latent variables to observed variables
- Parameters learn that each latent variable interacts strongly with only a small subset of observed variables
- Trainable just with gradient descent; no discrete search over graphs

# Roadmap

- Challenges of Unstructured Modeling
- Using Graphs to Describe Model Structure
- Sampling from Graphical Models
- Advantages of Structured Modeling
- Structure Learning and Latent Variables
- Inference and Approximate Inference
- The Deep Learning Approach to Structured Probabilistic Modeling

# Inference and Approximate Inference

- Inferring marginal distribution over some nodes or conditional distribution of some nodes given other nodes is #P hard

  - NP-hardness describes *decision* problems. #P-hardness describes *counting* problems, e.g., how many solutions are there to a problem where finding one solution is NP-hard

- We usually rely on *approximate* inference, described in chapter 19

# Roadmap

- Challenges of Unstructured Modeling
- Using Graphs to Describe Model Structure
- Sampling from Graphical Models
- Advantages of Structured Modeling
- Structure Learning and Latent Variables
- Inference and Approximate Inference
- The Deep Learning Approach to Structured Probabilistic Modeling

# Deep Learning Stylistic Tendencies

- Nodes organized into layers
- High amount of connectivity between layers
- Examples: RBMs, DBMs, GANs, VAEs

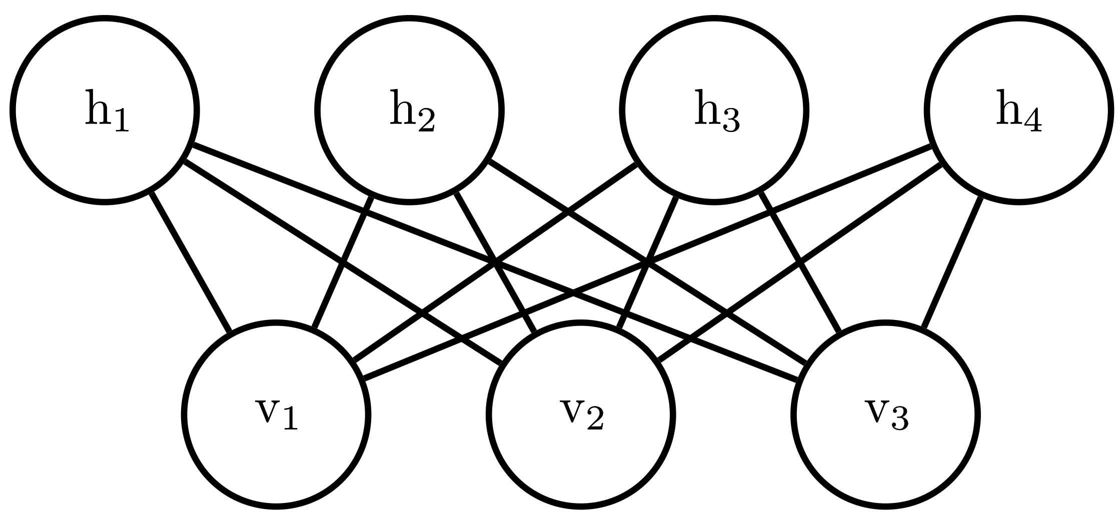
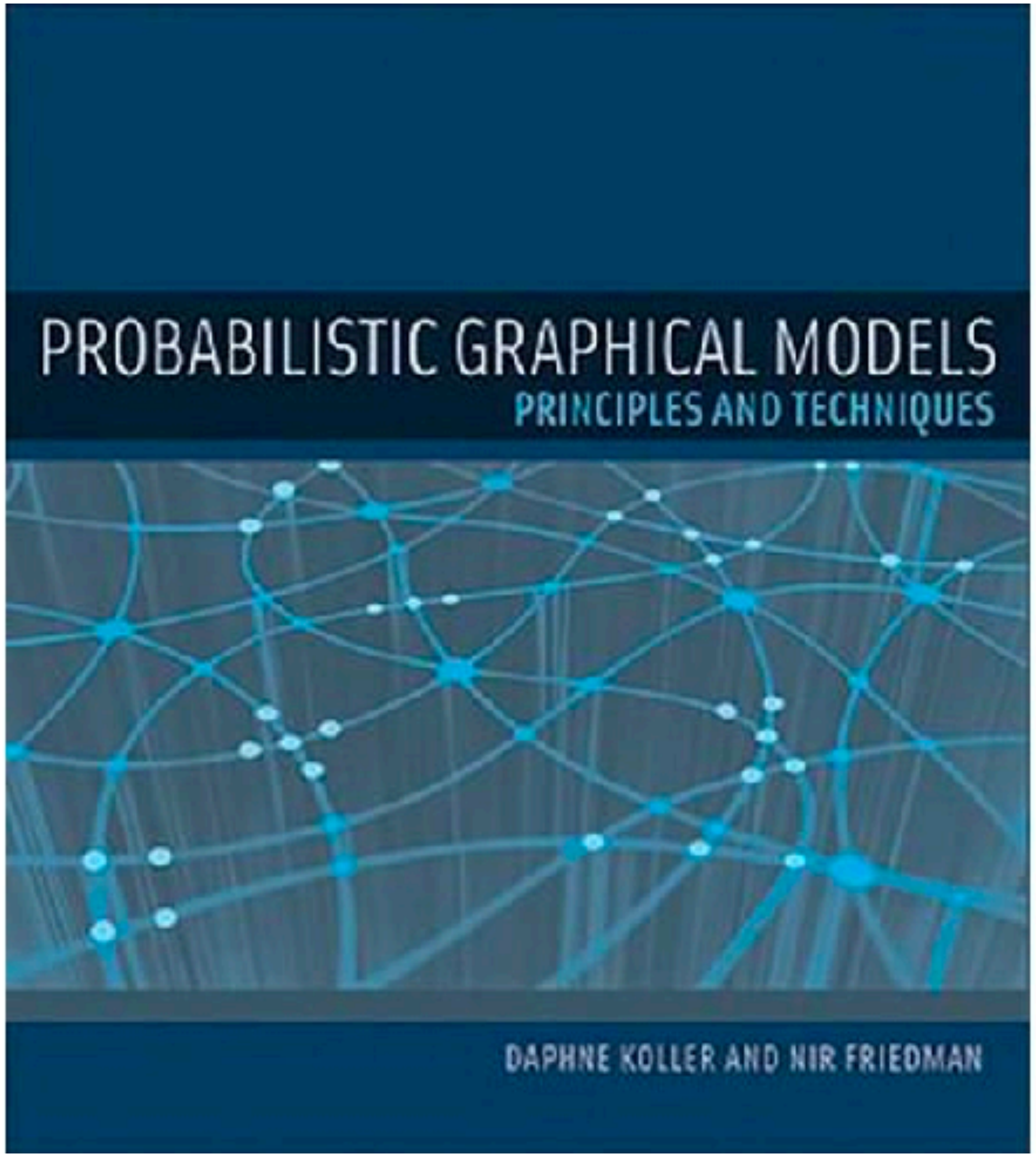



Figure 16.14: An RBM drawn as a Markov network.

(Goodfellow 2017)

# For more information...




(Goodfellow 2017)