CS839:
Probabilistic Graphical Models

Lecture 1: Introduction to Graphical Models

Theo Rekatsinas

Acknowledgement: adapted slides by Eric Xing

1. Introduction, admin & setup

Who am I…

Instructor (me) Theo Rekatsinas

- Faculty in the Computer Sciences and part of the UW-Database Group
- **Research**: data integration and cleaning, statistical analytics, and machine learning.
- thodrek@cs.wisc.edu
- Office hours: By appointment @CS 4361

Course Webpage:

https://thodrek.github.io/CS839_fall18/

Logistics

- Text books:
  - Probabilistic Graphical Models, by Daphne Koller and Nir Friedman
  - Introduction to Statistical Relational Learning, by Lise Getoor and Ben Taskar
- Office hours:
  - By appointment. Just send me an email
- Homework submission:
  - We will use Canvas

Assignments and Grading Logistics

- 3 homework assignments: 20% of grade
  - Theory exercises, implementation exercises
- Midterm: 30% of grade
  - In class exam
  - ~ weak #9
- Final project: 50% of grade
  - Project proposal: 10% of grade (~weak #9)
  - Proposal presentation: 10% of grade
  - Final report: 30% of grade (due on Dec 20th)
  - In groups of (up to) 3. Ideally it should be three. Groups should be formed in the first two weeks.

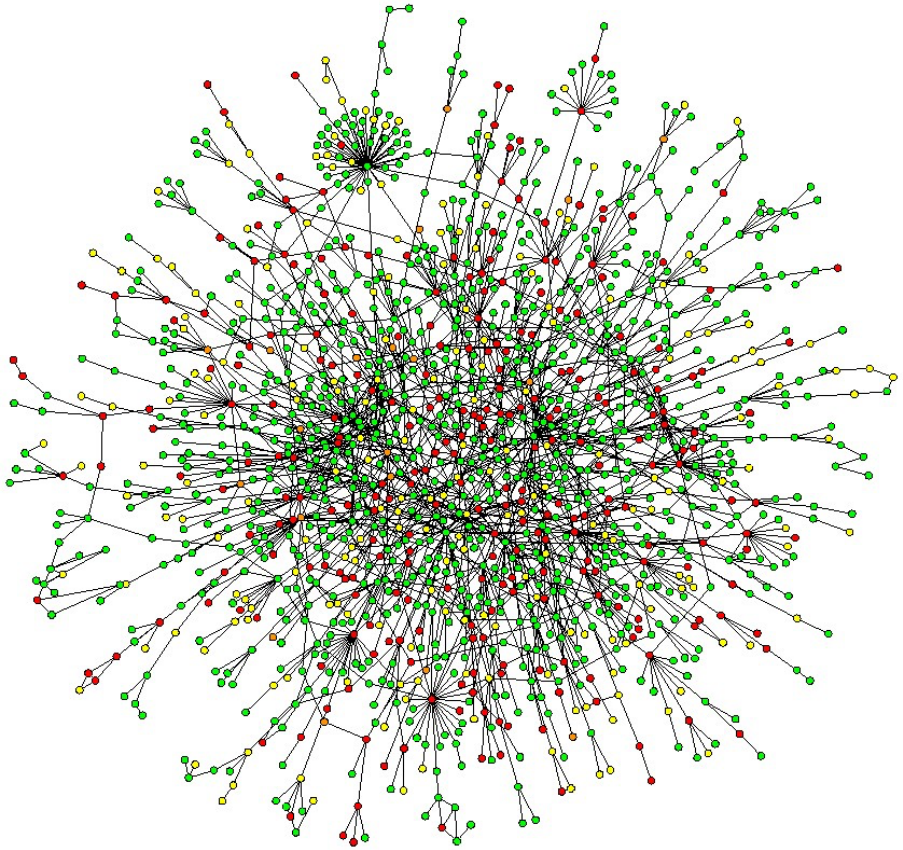
Project examples

- Applying PGM to the development of a real, substantial ML system
  - Build a web-scale fake news detector.
  - Build a story line tracking system for news media.
  - Design and implement the state-of-the-art knowledge base emgeddings
- Theory and/or algorithmic projects
  - A more efficient approximate inference algorithm.
  - When is inference in the presence of noisy observations hard?
  - When can we approximate PGMs with feed-forward networks?
- System's
  - Implement Markov logic on top of Pyro

2. Class overview

What are graphical models?

Graph



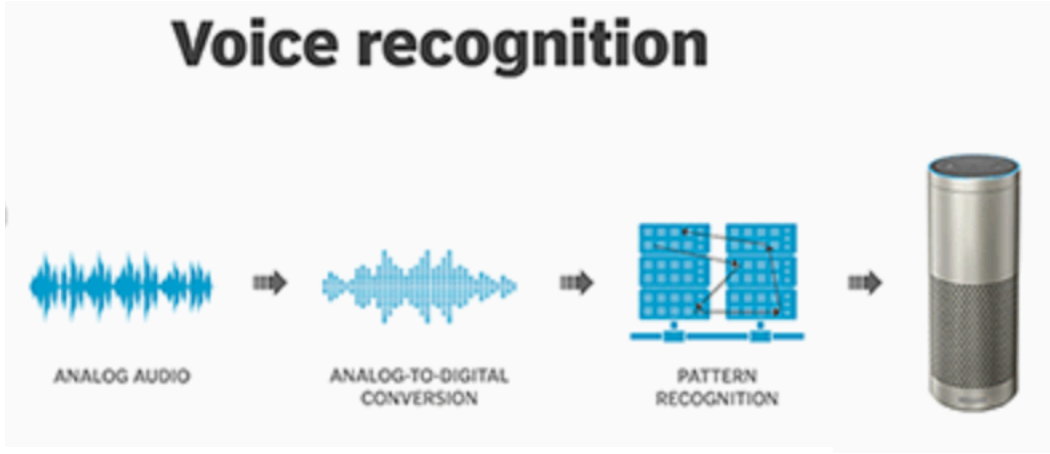
Model

$$\mathcal{M}$$

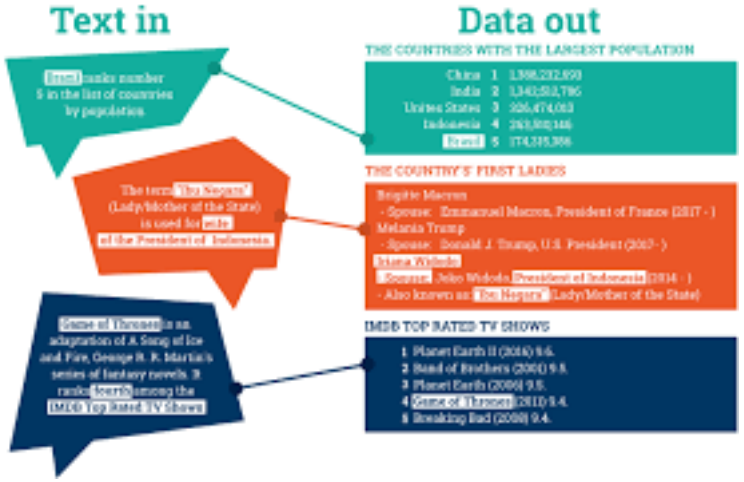
Data

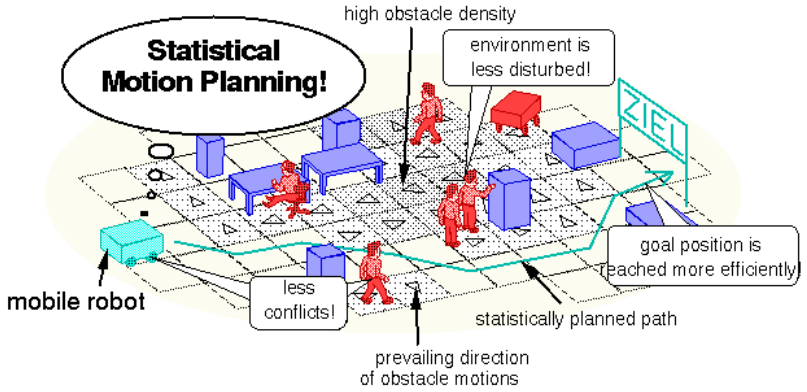
$$\mathcal{D} \equiv \{X_1^{(i)}, X_2^{(i)}, \ldots, X_m^{(i)}\}_{I=1}^{N}$$

PGMs allow us to reason about uncertainty

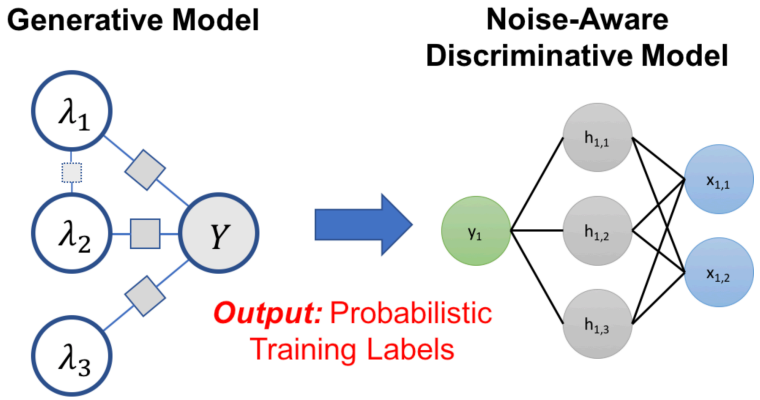


Information Extraction`

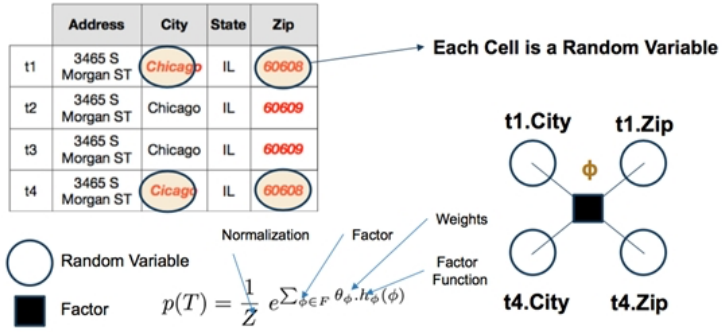




Weak Supervision



Data Cleaning



Fundamental Questions

- Representation
  - How to capture/model uncertainties in possible worlds?
  - How to encode our domain knowledge/assumptions/constraints?

- Example: Is your Grade independent of the Difficulty of the class?

Fundamental Questions

- Inference
  - How do we answer questions/queries according to the model in hand and the available data P(X|Data)

- Example: What will your Grade be if Difficulty is “high”?

Fundamental Questions

- Learning
  - What model is "right" for the data? $\mathcal{M} = \arg\max_{\mathcal{M} \in M} F(\mathcal{D}; \mathcal{M})$
- Example: What if we have (Difficulty = "Low", Intelligence = "High", Grade = "High") for person 1, (Difficulty = "High", Intelligence = "High", Grade = "Low") for person 2, etc?

Basic Probability Concepts

- Representation: What is the joint probability distribution on multiple variables?

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

  - How many state configurations we have in total?
  - Are they all needed to be represented?
  - What insights do we get from this model?

Basic Probability Concepts

- Learning: Where do we get all these probabilities?
  - Maximal-likelihood estimation? How many data do we need?
  - Are there other estimation principles?
  - Where do we put domain knowledge in terms of plausible relationships between variables, and plausible values of the probabilities?

Basic Probability Concepts

- Inference: If not all variables are observable, how to compute the conditional distribution of latent variables given evidence?

  - Assume X_1 is given. Computing $P(X_2|X_1)$ requires summing over all 2^6 configurations of the unobserved varibles.

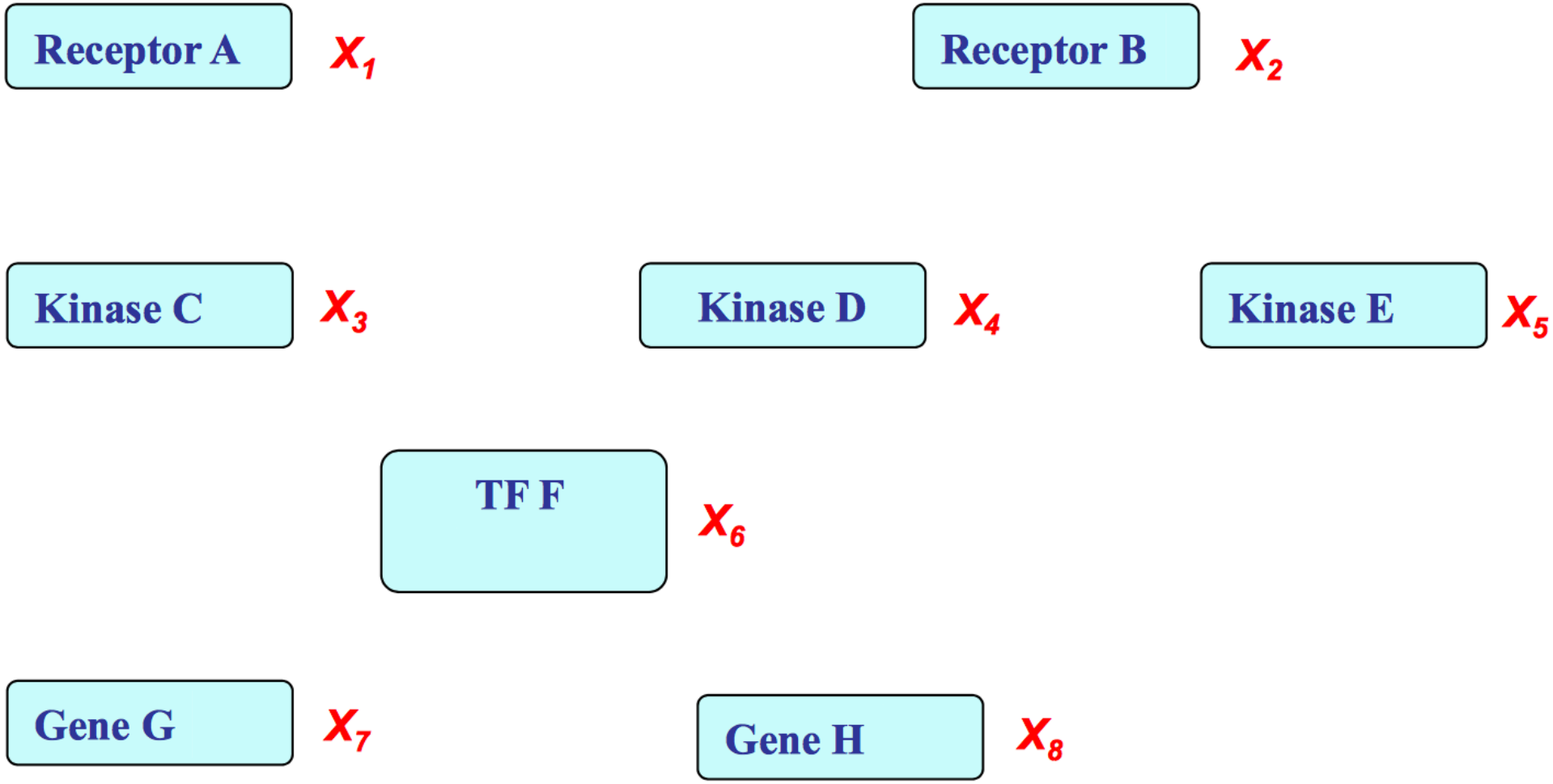
What is a graphical model?

A Multivariate Distribution in High-D Space

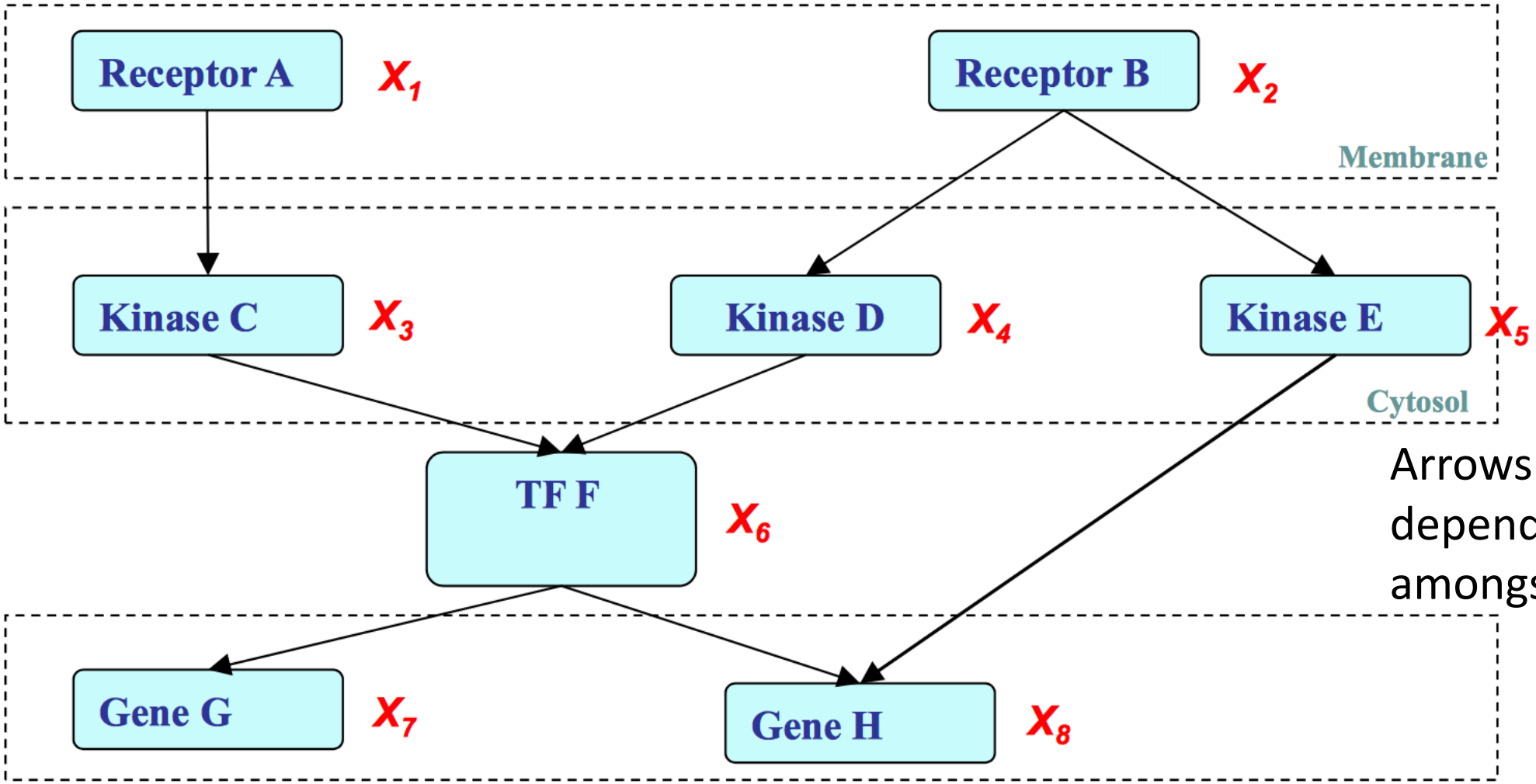
Example: A possible world for cellular signal transduction

Signal transduction is the process by which a chemical or physical signal is transmitted through a cell as a series of molecular events

Example: A possible world for cellular signal transduction



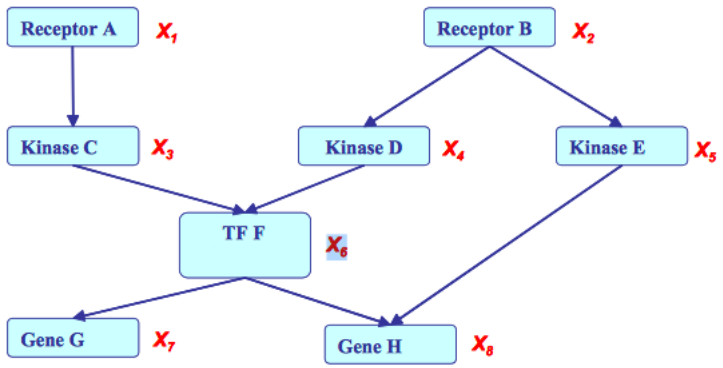
Structure Simplifies Representation



Arrows indicate dependencies amongst variables

Probabilistic Graphical Models

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

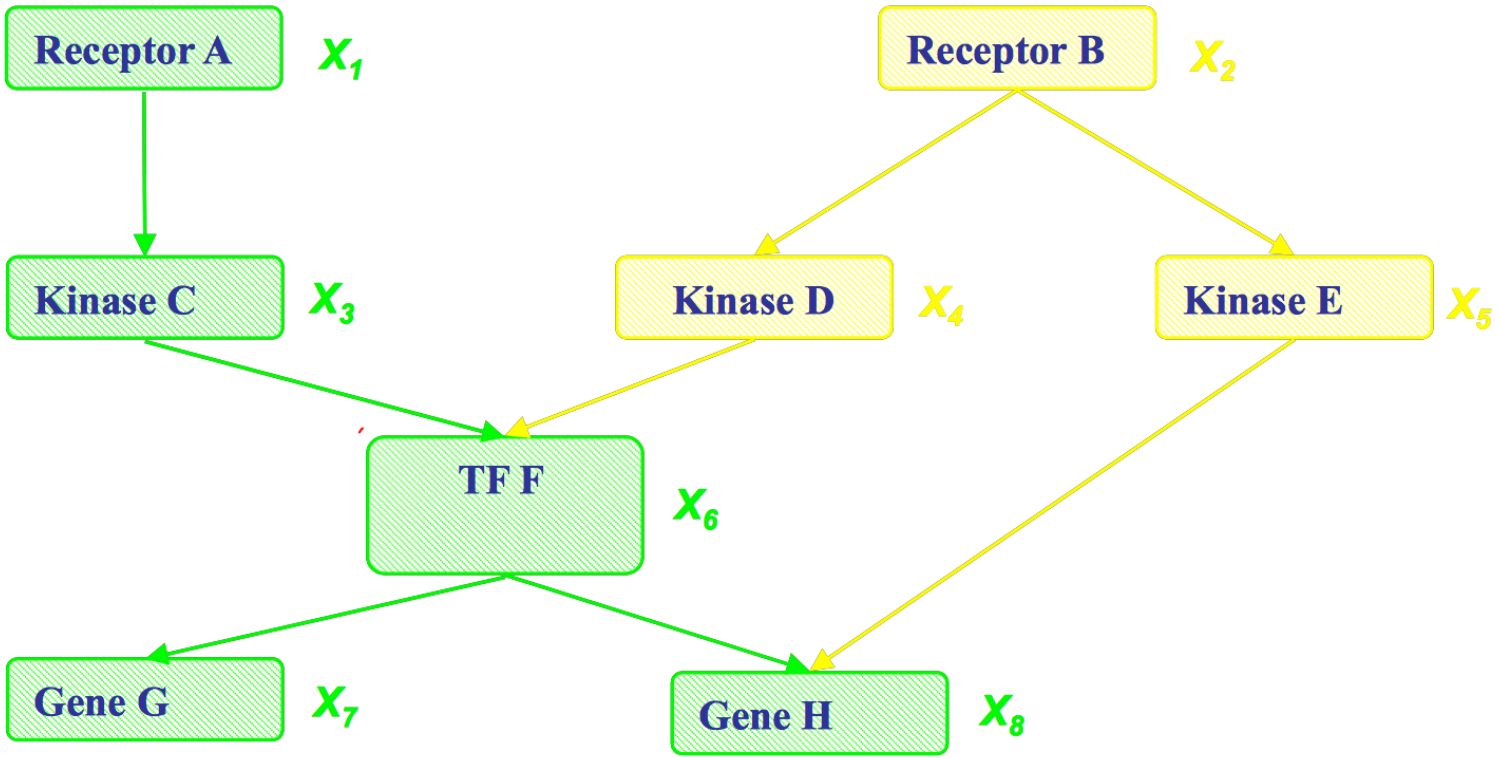


$$
\begin{aligned}
&P(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8) \\
=&P(X_1)P(X_2)P(X_3|X_1)P(X_4|X_2)P(X_5|X_2) \\
&P(X_6|X_3, X_4)P(X_7|X_6)P(X_8|X_5, X_6)
\end{aligned}
$$

- So, why a PGM? We can incorporate domain knowledge
  - 1+1+2+2+2+4+2+4=18, a 16-fold reduction from 28 in representation cost!

Other desired properties of PGMS

- Modularity – Allows us to integrate heterogeneous data



Other desired properties of PGMS

- Prior knowledge – Bayesian learning

Posterior probability

Likelihood

Prior probability

$$p(h \mid d) = \frac{p(d \mid h)\, p(h)}{\sum_{h' \in H} p(d \mid h')\, p(h')}$$

Sum over space of hypotheses

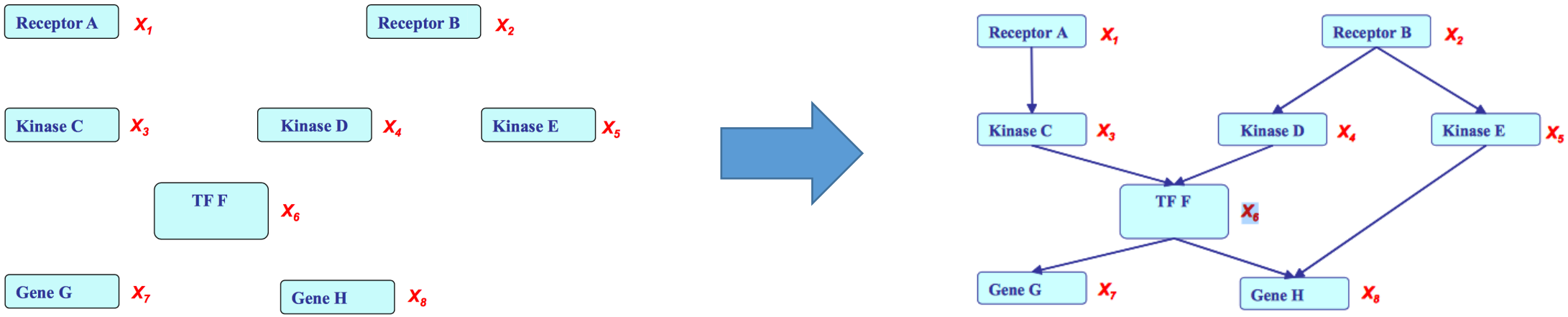
- Captures uncertainty in a more principled way – introduce priors

What is a graphical model?

Multivariate statistics + structure

What is a graphical model?

- **Informal:** It is a smart way to specify/compose/design exponentially-large probability distributions without paying an exponential cost, and at the same time endow the distributions with ***structured semantics***



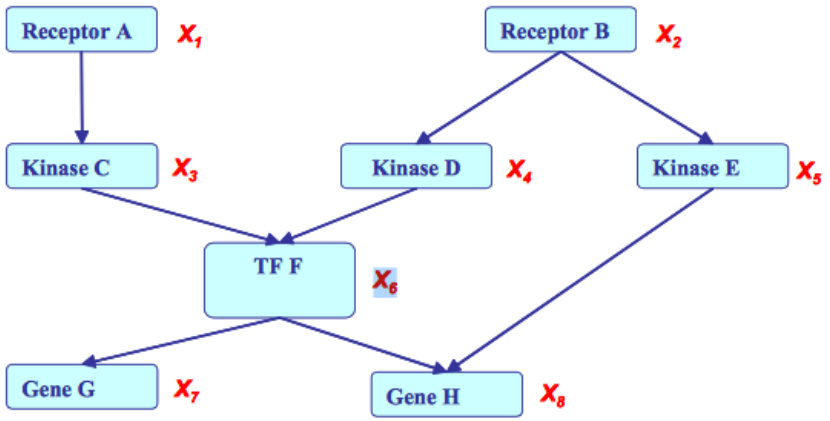
What is a graphical model?

- **More formal:** It refers to a family of distributions on a set of random variables that are compatible with all the probabilistic independence propositions encoded by a graph that connects these variables

Types of PGMs

- **Directed:** Bayesian networks
  - Directed edges give causality relationships

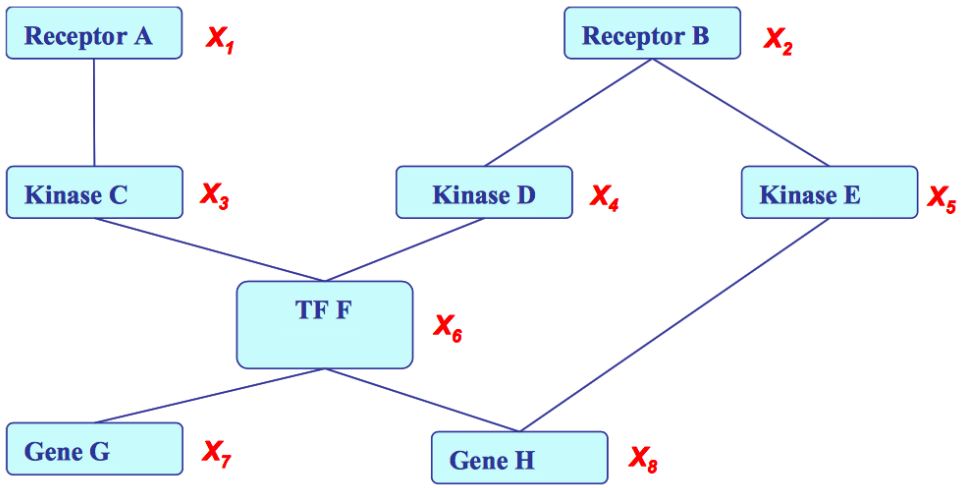
$$
\begin{aligned}
&P(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8) \\
=&P(X_1)P(X_2)P(X_3|X_1)P(X_4|X_2)P(X_5|X_2) \\
&P(X_6|X_3, X_4)P(X_7|X_6)P(X_8|X_5, X_6)
\end{aligned}
$$



Types of PGMs

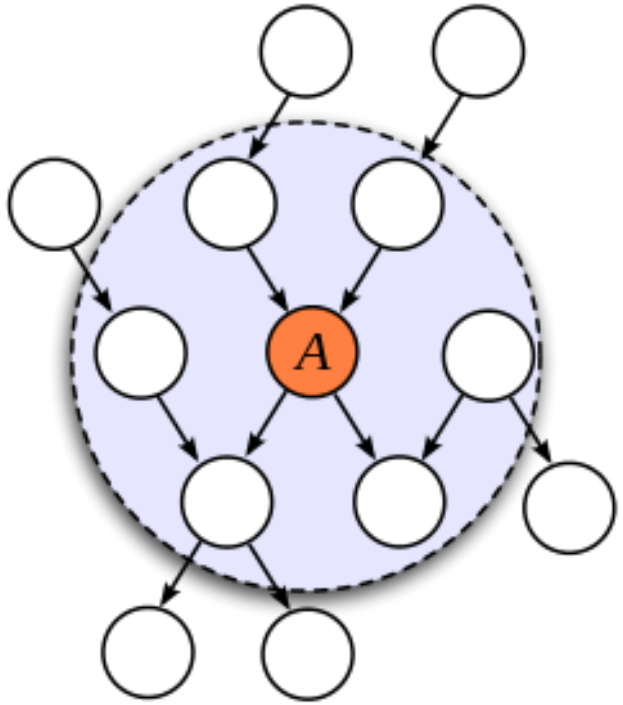
- **Undirected:** Markov random fields
  - Undirected edges simply give correlations between variables

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



Bayesian Networks

- Structure: DAGs
- Meaning: a node is **conditionally independent** of every other node in the network outside its **Markov blanket**



The Markov blanket of node includes its parents, children and the other parents of all of its children.

Bayesian Networks

- Structure: DAGs
- Meaning: a node is **conditionally independent** of every other node in the network outside its **Markov blanket**
- Local conditional distributions (**CPD**) and the **DAG** completely determine the **joint** distribution.
- Edges represent **causality** relationships, and facilitate a **generative** process

Markov Random Fields

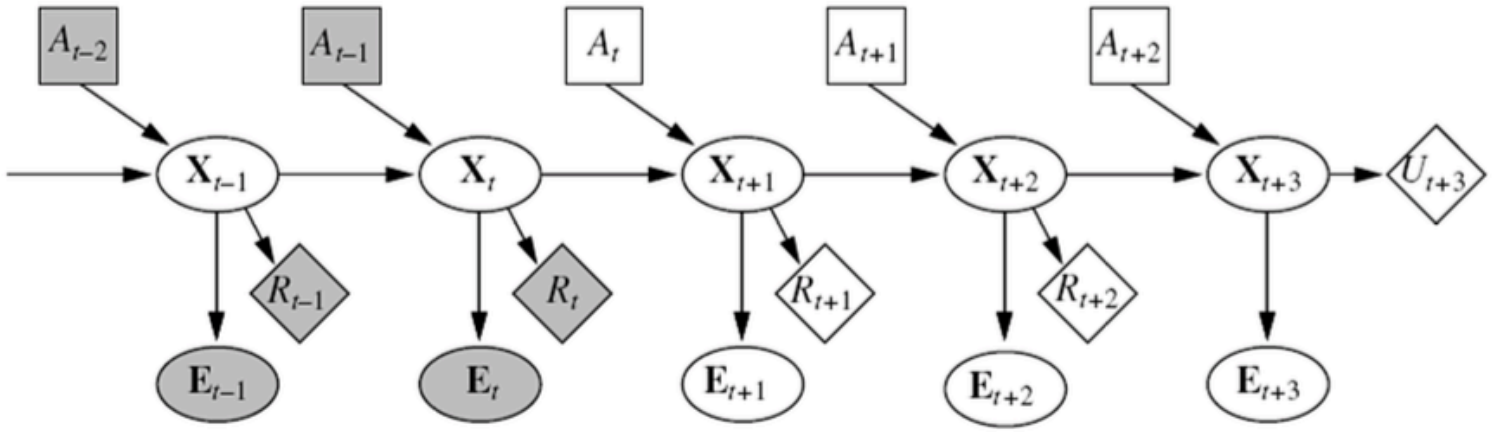
- Structure: undirected graph
- Meaning: a node is **conditionally independent** of every other node in the network given its **direct neighbors**
- Local contingency functions (**potentials**) and the **cliques** in the graph completely determine the **joint** dist.
- Edges represent **correlations** between variables, but no explicit way to generate samples

Well-known models as PGMs

- Density estimation
  - Parametric and non-parametric methods
- Regression
  - Linear, conditional mixture, non-parametric
- Classification
  - Generative and discriminative approaches
- Clustering

More complex models

- Partially observed Markov decision processes



More complex models

- Information Extraction

Variety Pack Filet Mignon and Porterhouse Steak Dog Food (1 Count) Price: $92.60 & **FREE Shipping**

Be the first to review this item

- 6 trays of Filet Mignon flavor in meaty juic
- Cesar pet food has an irresistible taste wit palatability to tempt even the fussiest dog
- Formulated to meet the nutritional levels by the AAFCO Dog Food Nutrient Profiles maintenance

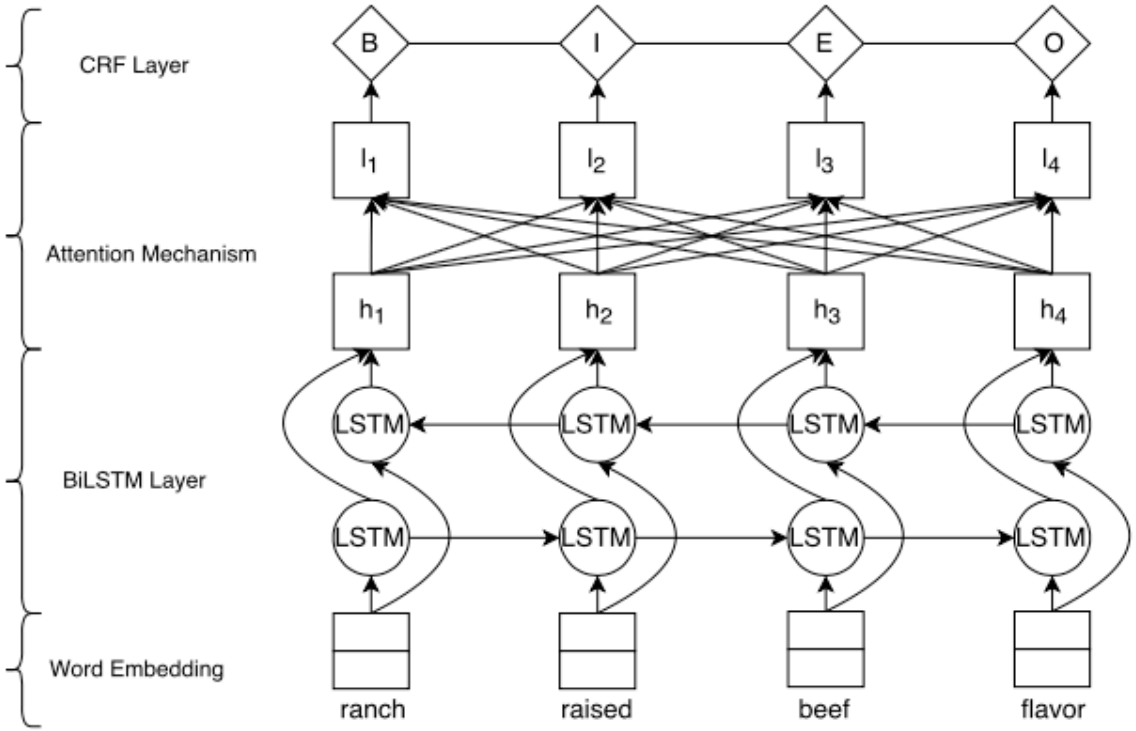
In stock.

Get it as soon as Wednesday, Feb. 14 when you choose **Two-Day Shipping** at checkout.

Ships from and sold by Cunningham Collective.

Product description

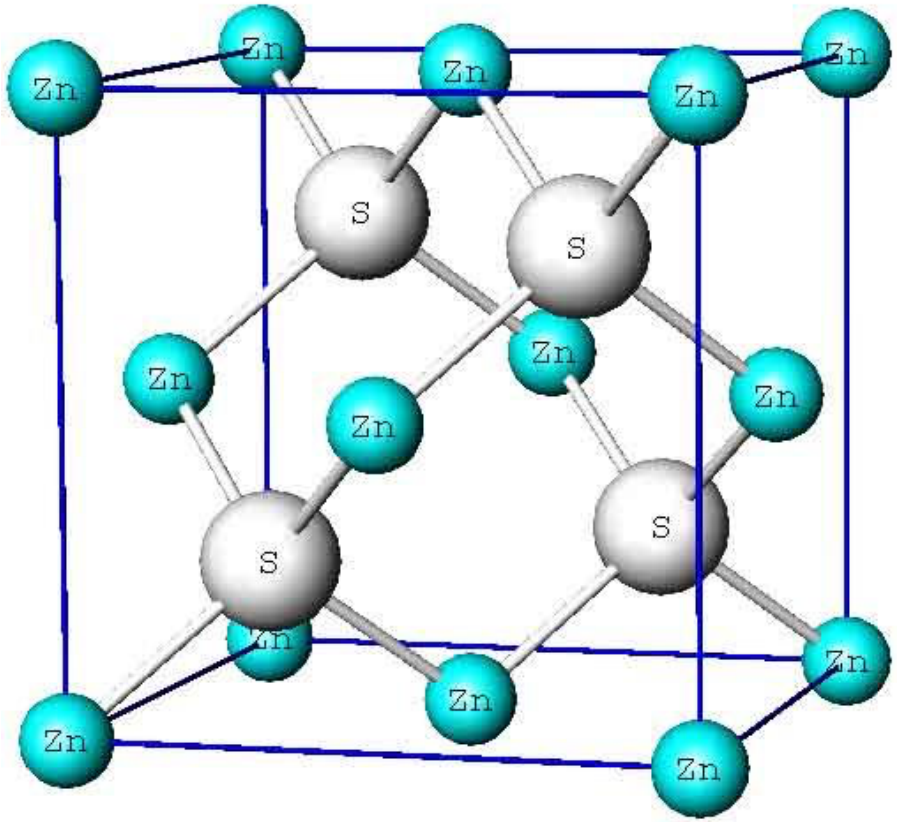
Variety pack includes: 6 trays of Filet Mignon flavor in meaty juices 6 trays of Porterhouse Steak flavor Cesar pet food has an irresistible taste with exceptional palatability to tempt even the fussiest dogs Fo the nutritional levels established by the AAFCO Dog Food Nutrient Profiles for maintenance Complete nutrition for small adult dogs Fortified with vitamins and minerals Packaged in convenient feeding tray peel-away freshness seals Includes 6 Each Chicken & Liver

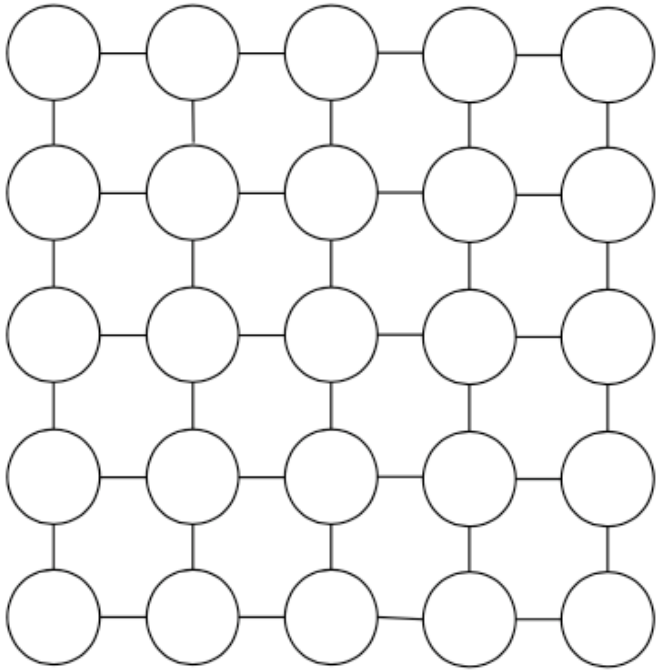


[OpenTag, Zheng et al., KDD 2018]

More complex models

- Solid state physics





Ising/Potts model

Applications of Graphical Models

- Machine learning
- Computational statistics
- Computer vision and graphics
- NLP
- Information extraction
- Robotic control
- Decision making under uncertainty
- Computational biology
- Medical diagnosis/prognosis
- Finance and economics
- Etc.

Why PGMs?

- Language for communication
- Language for computation
- Language for development

- Does it remind you of something?

Why PGMs?

- Probability theory: Formal framework to combine heterogeneous parts and ensure consistency.
- Graph structure: Appealing interface for modeling highly-interacting sets of variables. Interpretability and domain knowledge.
- Generalization: Many classical probabilistic systems are special cases of PGMs

PGMs in the Deep Learning era

- Probabilistic Models: Goal is to capture the joint distribution of input variables, output variables, latent variables, parameters and hyper-parameters. Everything is a random variables.
- Deep (Learning) Models: Hierarchical model structure where the output of one model becomes the input of the next higher level model. Targeted towards feature learning.

PGMs in the Deep Learning era

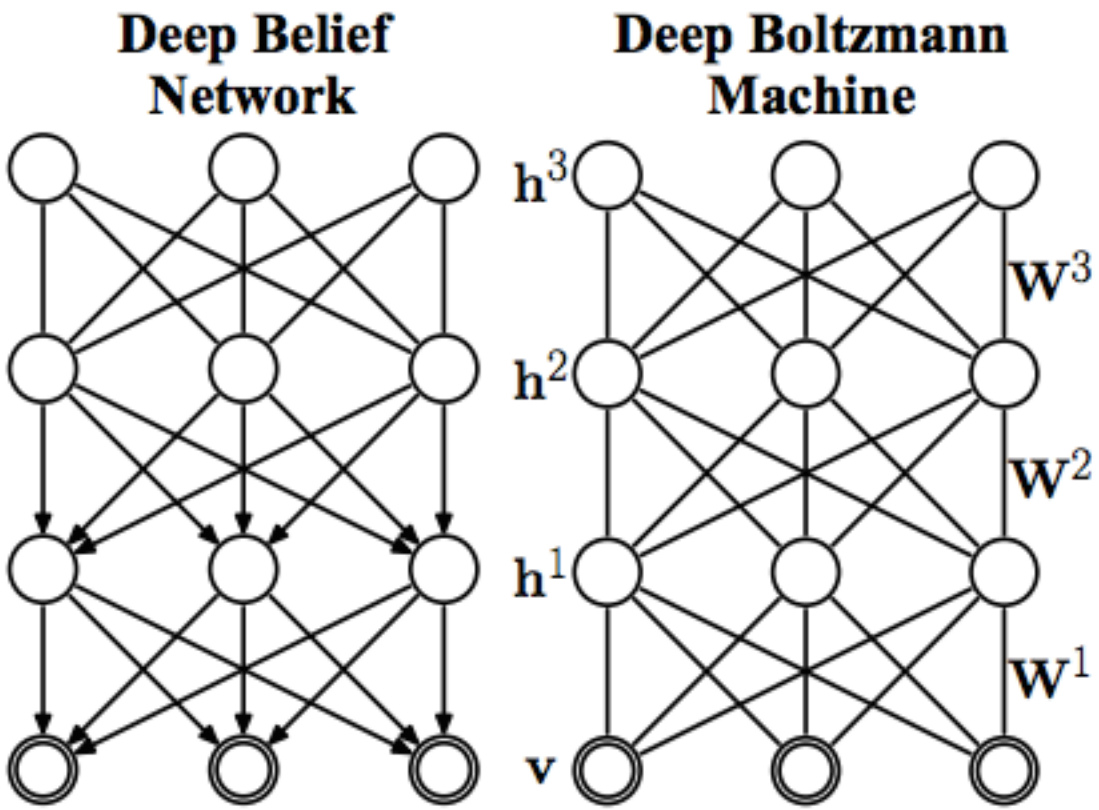
	Deep Learning	PGMs
Empirical goal:	e.g., Classification, feature learning	e.g., transfer learning, latent variable inference
Structure:	Graphical	Graphical
Objective:	Aggregated from local functions	Aggregated from local functions
Vocabulary:	Neuron, activation/gate function	Variables, potential functions
Algorithm:	Single inference algorithm, BP	Many algorithms, major focus of open research, approximate inference
Evaluation:	On end-performance	On almost every intermediate quantity (calibrated probabilities)
Implementation:	Many tricks ☺	Quite standardized

PGMs in the Deep Learning era

- Why Probabilistic Models?: Predictions from a probabilistic model that captures a principled notion of uncertainty. Decision making.
- Why Deep (Learning) Models: Feature learning. No assumptions for complex domains such as images and speech.

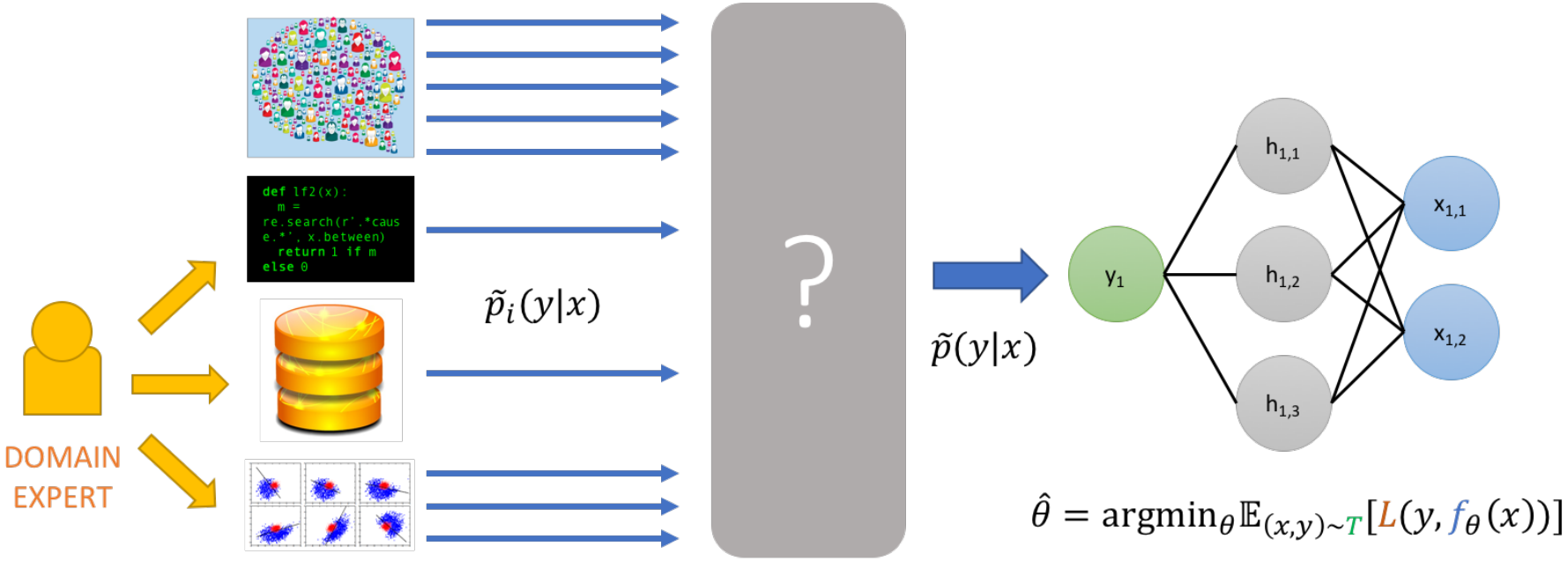
Combining PGMs and Deep Learning

- Deep Boltzmann Machines



Using PGMs to generate training data for DL

- Weak supervision/Data programming



Example Weak Supervision Sources

Technical Challenge: Integrating & Modeling Diverse Sources

Use Weak Supervision to Train End Model

Class Overview

- Fundamentals of PGMs:
  - Bayesian Networks and Markov Random Fields
  - Discrete, Continuous, and Hybrid models, exponential family
  - Basic representation, inference, and learning
  - Focus on specific networks: Multivariate Gaussian Models, Hidden Markov Models
- Advanced Topics:
  - Approximate inference
  - Bounded treewidth
  - Spectral methods for Graphical models
  - Structure learning
  - Relational representation learning and connections to deep learning
- Applications