

# Probabilistic Graphical Models

## Introduction to GM

Eric Xing

Lecture 1, January 14, 2019

Reading: see class homepage





# Logistics

- Class webpage: <https://sailinglab.github.io/pgm-spring-2019/>

10-708 PGM      [description](#) [lectures](#) [calendar](#) [notes](#) [homework](#) [project](#)

 **Probabilistic Graphical Models**  
10-708 • Spring 2019 • Carnegie Mellon University

Many of the problems in artificial intelligence, statistics, computer systems, computer vision, natural language processing, and computational biology, among many other fields, can be viewed as the search for a coherent global conclusion from local information. The probabilistic graphical models framework provides an unified view for this wide range of problems, enabling efficient inference, decision-making and learning in problems with a very large number of attributes and huge datasets. This graduate-level course will provide you with a strong foundation for both applying graphical models to complex problems and for addressing core research topics in graphical models.

---

- Time: Monday, Wednesday 12:00-1:20 pm
- Location: [Posner Hall 152](#)
- Recitations: TBA

---



Instructor [Eric P. Xing](#)  
Email: [epxing@cs.cmu.edu](mailto:epxing@cs.cmu.edu)  
Office hours: TBA



Head GSI [Maruan Al-Shedivat](#)  
Email: [alshedivat@cs.cmu.edu](mailto:alshedivat@cs.cmu.edu)  
Office hours: TBA

© Eric Xing @ CMU, 2005-2019

2



# Logistics

- ❑ Text books:
  - ❑ Daphne Koller and Nir Friedman, Probabilistic Graphical Models
  - ❑ M. I. Jordan, An Introduction to Probabilistic Graphical Models
- ❑ Mailing Lists:
  - ❑ To contact the instructors: 10708-instructor@cs.cmu.edu
  - ❑ Class announcements: Piazza
- ❑ TA:
  - ❑ Maruan Al-Shedivat, GHC 8229, Office hours: Wednesday 1:30-2:30 pm
  - ❑ Lisa Lee, GHC 8011, Office hours: Monday 9-10 am
  - ❑ Xun Zheng, GHC 8013, Office hours: Friday 4-5 pm
  - ❑ Hao Zhang, NSH 4225, Office hours: Friday 12-1 pm
  - ❑ Paul Liang, GHC 8011, Office hours: Thursday 5-6 pm
- ❑ Lecturer: Eric Xing, Office hours: Monday 1:30-2:30 pm
- ❑ Class Assistant:
  - ❑ Amy Protos, GHC 8001
- ❑ Instruction aids: Piazza





# Logistics

- ❑ 4 homework assignments: 40% of grade
  - ❑ Theory exercises, Implementation exercises
- ❑ Scribe duties: 10% (~once to twice for the whole semester)
- ❑ Class participation: 4% (more details on the course webpage)
  - ❑ Contribution to discussion on Piazza (up to 3.5%)
  - ❑ Complete mid-semester evaluation (0.5%)
- ❑ Final project: 46% of grade
  - ❑ Applying PGM to the development of a real, substantial ML system
    - ❑ Design and Implement a (record-breaking) distributed Logistic Regression, Gradient Boosted Tree, Deep Network, or Topic model on Petuum and apply to ImageNet, Wikipedia, and/or other data
    - ❑ Build a web-scale topic or story line tracking system for news media, or a paper recommendation system for conference review matching
    - ❑ An online car or people or event detector for web-images and webcam
    - ❑ An automatic “what’s up here?” or “photo album” service on iPhone
  - ❑ Theoretical and/or algorithmic work
    - ❑ a more efficient approximate inference or optimization algorithm, e.g., based on stochastic approximation, proximal average, or other new techniques
    - ❑ a distributed sampling scheme with convergence guarantee
  - ❑ 3-member team to be formed in the first three weeks, proposal, mid-way report, oral presentation & demo, final report, peer review → possibly conference submission !





# Past projects:

The screenshot shows a Mozilla Firefox browser window with the title "Grading - Mozilla". The address bar displays the URL <http://www.cs.cmu.edu/~epxing/Class/10708-07/project.html#list>. The main content area shows a course project page for "Probabilistic Graphical Models" (10-708, Fall 2007) at the School of Computer Science, Carnegie-Mellon University. The page includes the university's logo, a "Course Project" section describing the assignment requirements, and a "Project Proposal" section. The browser interface includes standard menu bars (File, Edit, View, Go, Bookmarks, Tools, Window, Help), a toolbar with icons for Back, Forward, Reload, Stop, Search, Print, and a status bar showing the URL.

## ❑ Award Winning Projects:

J. Yang, Y. Liu, E. P. Xing and A. Hauptmann, [Harmonium-Based Models for Semantic Video Representation and Classification](#), *Proceedings of The Seventh SIAM International Conference on Data Mining (SDM 2007 best paper)*

Manaal Faruqui, Jesse Dodge, Sujay Kumar Jauhar, Chris Dyer, Eduard Hovy, Noah A. Smith, [Retrofitting Word Vectors to Semantic Lexicons](#), NAACL 2015 best paper

Others ... such as KDD 2014 best paper

## ❑ Other projects:

Andreas Krause, Jure Leskovec and Carlos Guestrin, [Data Association for Topic Intensity Tracking](#), *23rd International Conference on Machine Learning (ICML 2006)*.

M. Sachan, A. Dubey, S. Srivastava, E. P. Xing and Eduard Hovy, [Spatial Compactness meets Topical Consistency: Jointly modeling Links and Content for Community Detection](#), *Proceedings of The 7th ACM International Conference on Web Search and Data Mining (WSDM 2014)*.

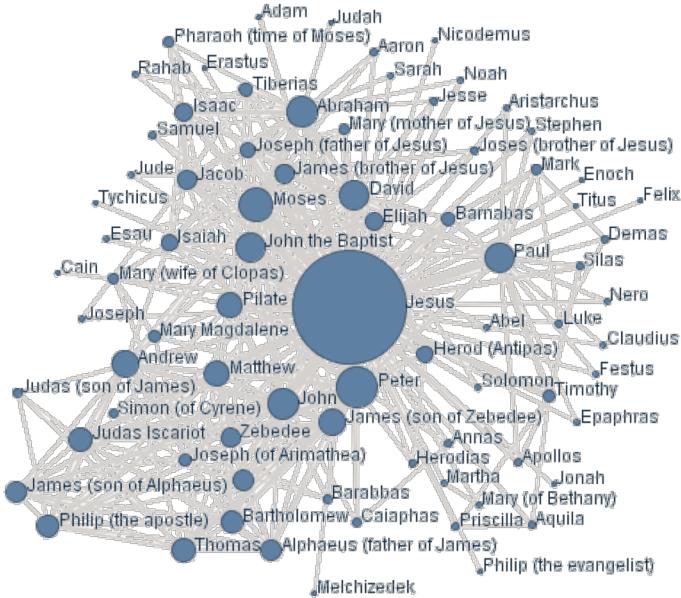
- ❑ We will have a prize for the best project(s) ...





# What Are Graphical Models?

Graph



Model

$$M_G$$

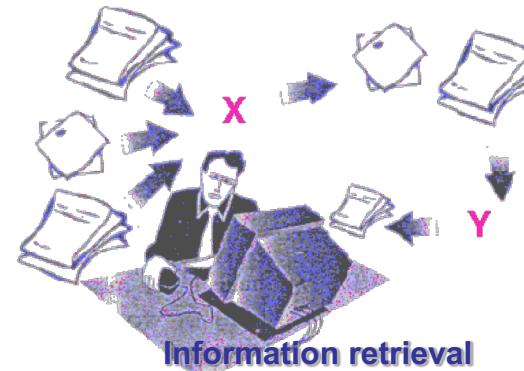
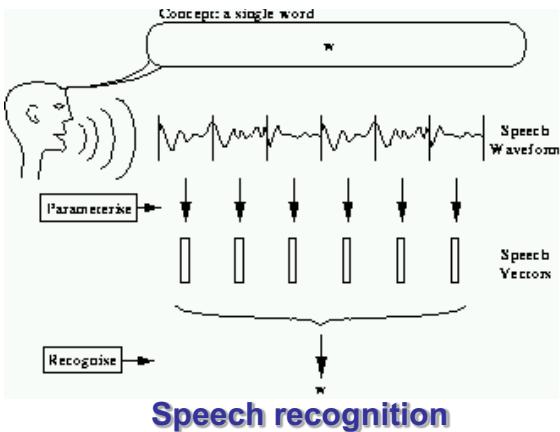
Data

$$D = \{X_1^{(i)}, X_2^{(i)}, \dots, X_m^{(i)}\}_{i=1}^N$$

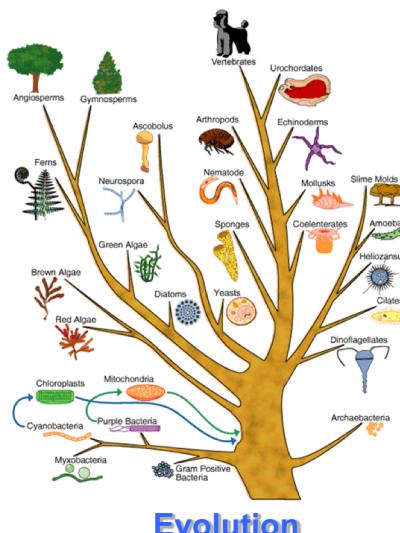
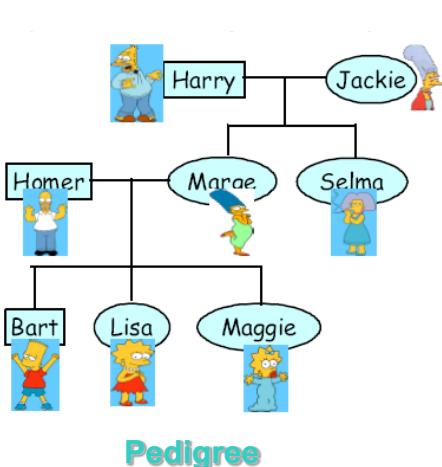




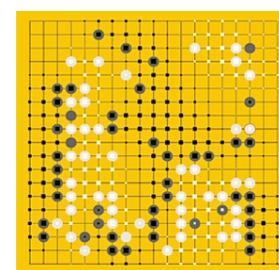
# Reasoning under uncertainty!



Computer vision



Evolution



Games



Planning



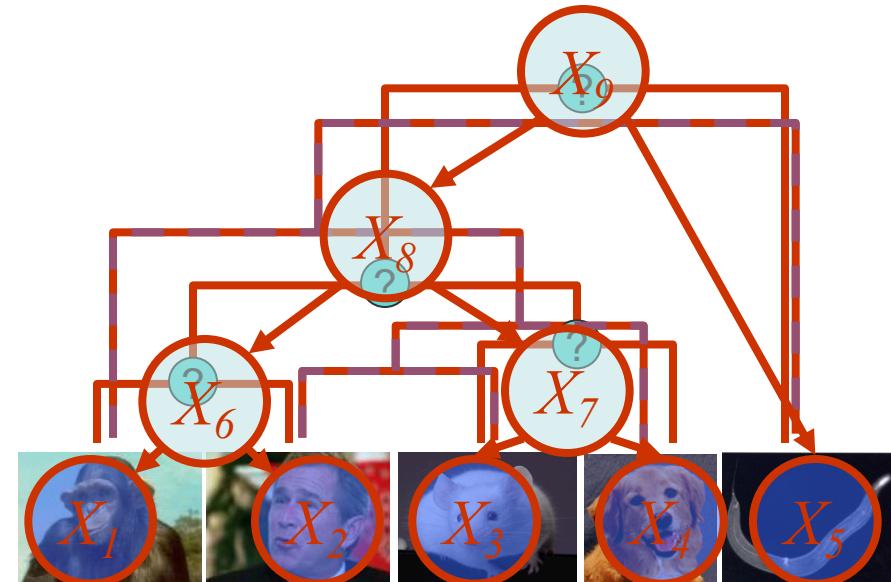
Robotic control





# The Fundamental Questions

- ❑ Representation
  - ❑ How to capture/model uncertainties in possible worlds?
  - ❑ How to encode our domain knowledge/assumptions/constraints?
- ❑ Inference
  - ❑ How do I answers questions/queries according to my model and/or based given data?  
e.g.:  $P(X_i | \mathcal{D})$
- ❑ Learning
  - ❑ What model is "right" for my data?  
e.g.:  $\mathcal{M} = \arg \max_{\mathcal{M} \in \mathcal{M}} F(\mathcal{D}; \mathcal{M})$



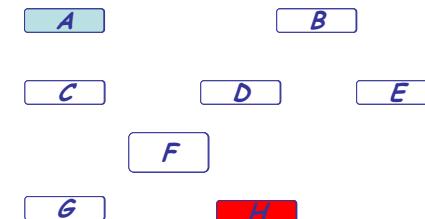


# Recap of Basic Prob. Concepts

- Representation: what is the joint probability dist. on multiple variables?

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

- How many state configurations in total? ---  $2^8$
  - Are they all needed to be represented?
  - Do we get any scientific/medical insight?
- Learning: where do we get all this probabilities?
- Maximal-likelihood estimation? but how many data do we need?
  - Are there other est. principles?
  - Where do we put domain knowledge in terms of plausible relationships between variables, and plausible values of the probabilities?
- Inference: If not all variables are observable, how to compute the conditional distribution of latent variables given evidence?
- Computing  $p(H|A)$  would require summing over all  $2^6$  configurations of the unobserved variables





# What is a Graphical Model?

--- Multivariate Distribution in High-D Space

- A possible world for cellular signal transduction:

Receptor A  $x_1$

Receptor B  $x_2$

Kinase C  $x_3$

Kinase D  $x_4$

Kinase E  $x_5$

TF F  $x_6$

Gene G  $x_7$

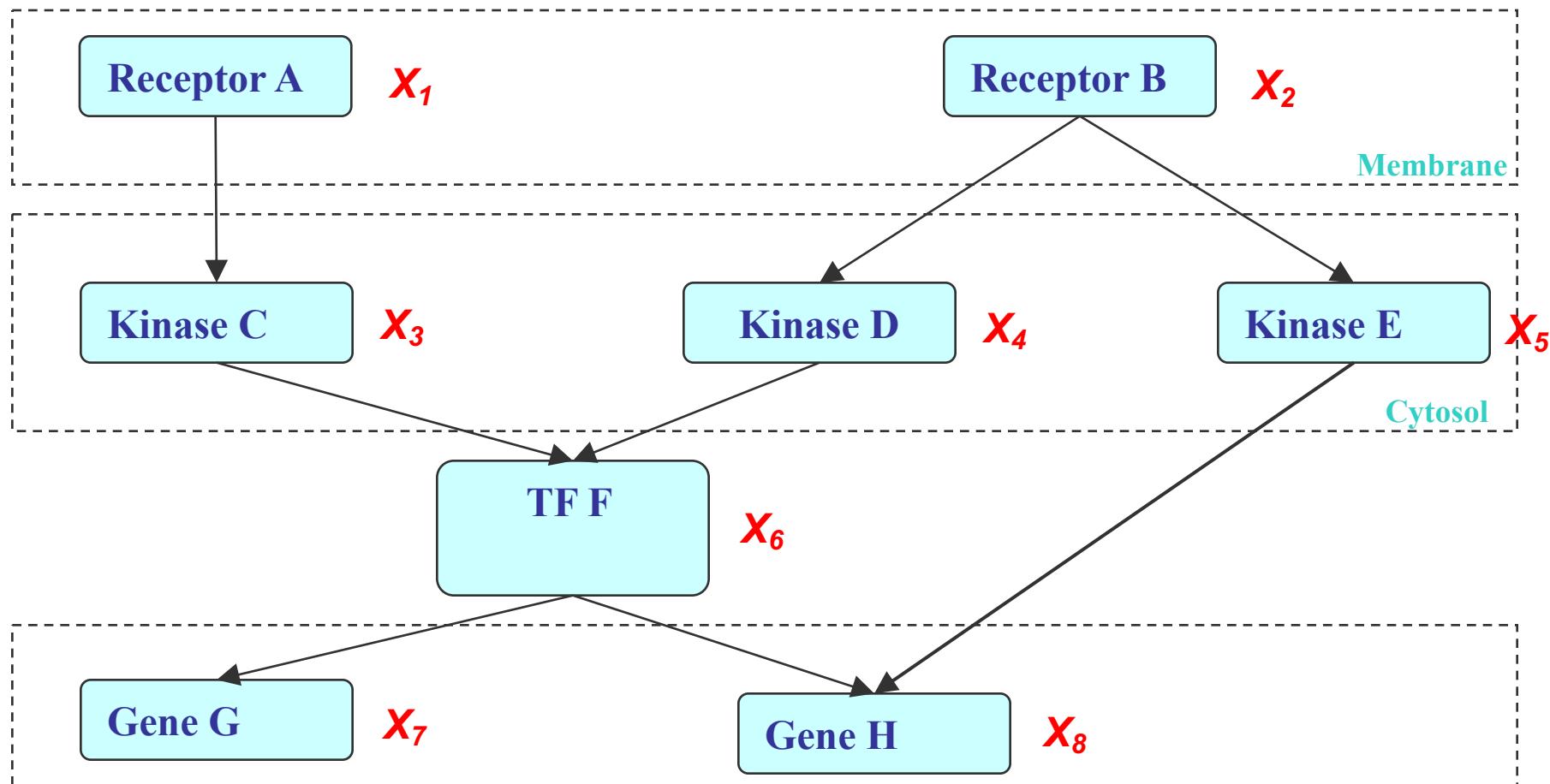
Gene H  $x_8$





# GM: Structure Simplifies Representation

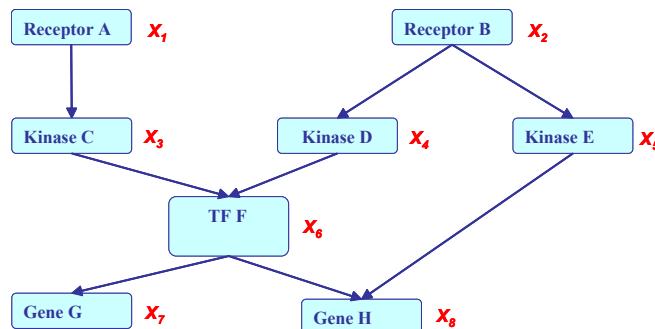
- Dependencies among variables





# 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.,



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

Stay tune for what are these independencies!

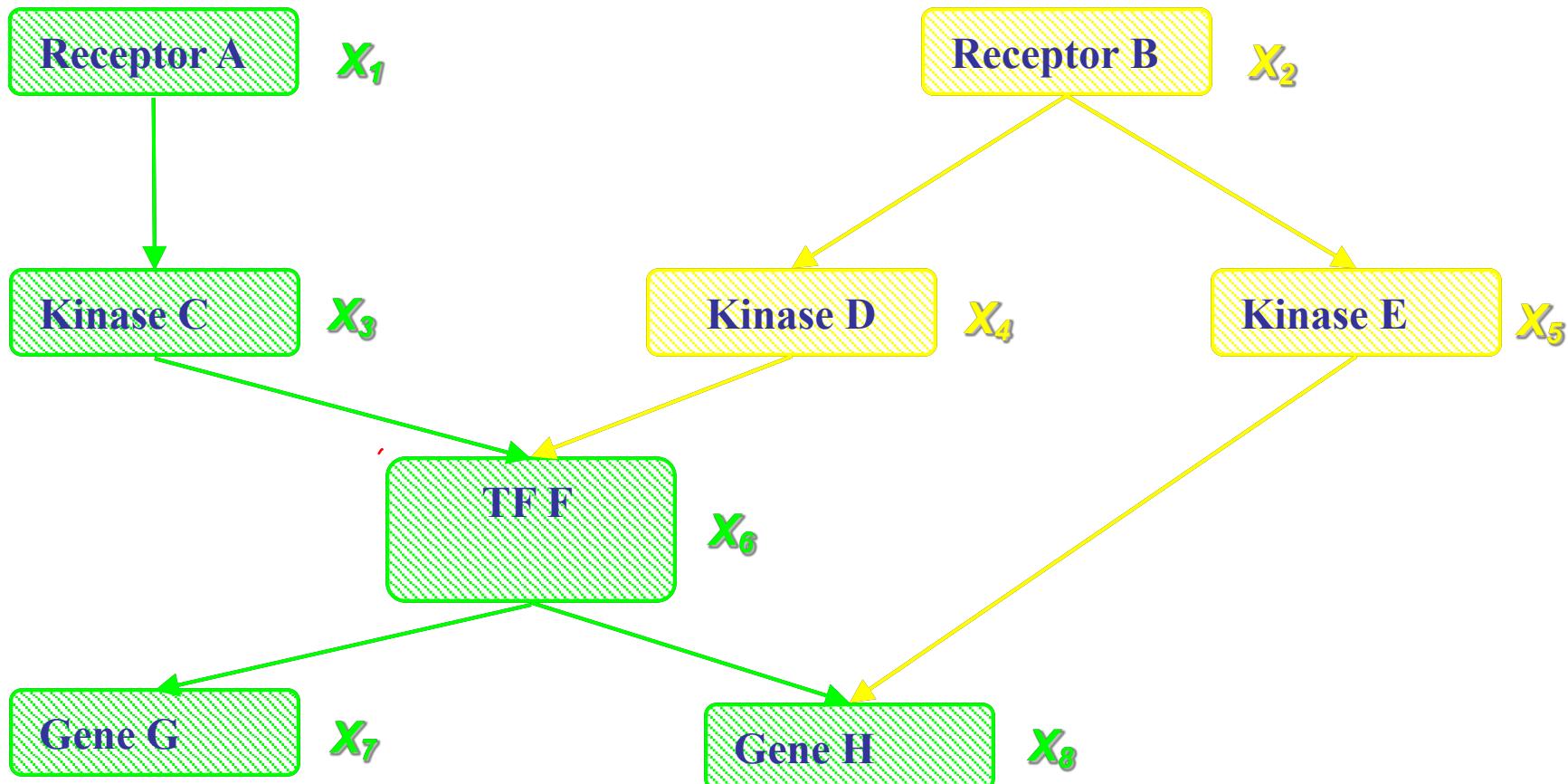
- Why we may favor a PGM?
  - Incorporation of domain knowledge and causal (logical) structures

1+1+2+2+2+4+2+4=18, a 16-fold reduction from  $2^8$  in representation cost !





# GM: Data Integration





# More Data Integration

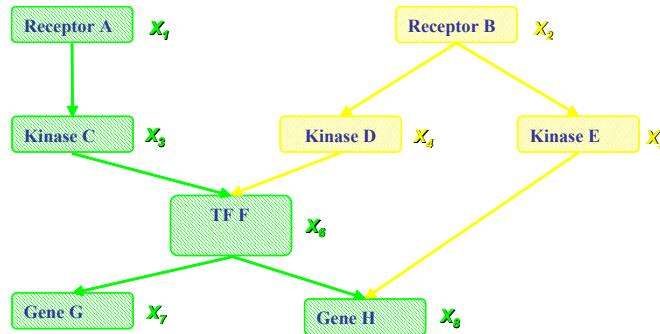
- ❑ Text + Image + Network → Holistic Social Media
  
- ❑ Genome + Proteome + Transcriptome + Phenome + ... → PanOmic Biology





# 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.,



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

- Why we may favor a PGM?
  - Incorporation of domain knowledge and causal (logical) structures  
 $2+2+4+4+4+8+4+8=36$ , an 8-fold reduction from  $2^8$  in representation cost !
  - Modular combination of heterogeneous parts – data fusion





# Rational Statistical Inference

## The Bayes Theorem:

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

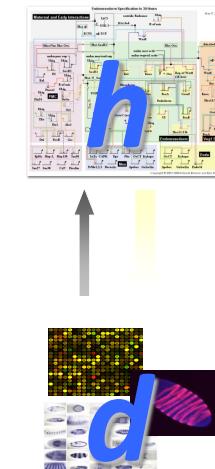
Sum over space of hypotheses

Posterior probability

Likelihood

Prior probability

$p(d)$



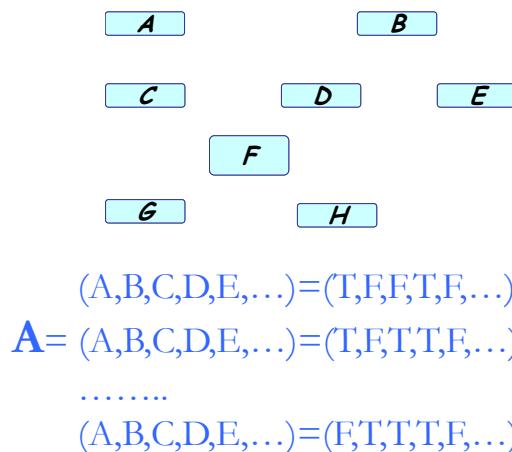
- ❑ This allows us to capture uncertainty about the model in a principled way
- ❑ But how can we specify and represent a complicated model?
  - ❑ Typically the number of genes need to be modeled are in the order of thousands!



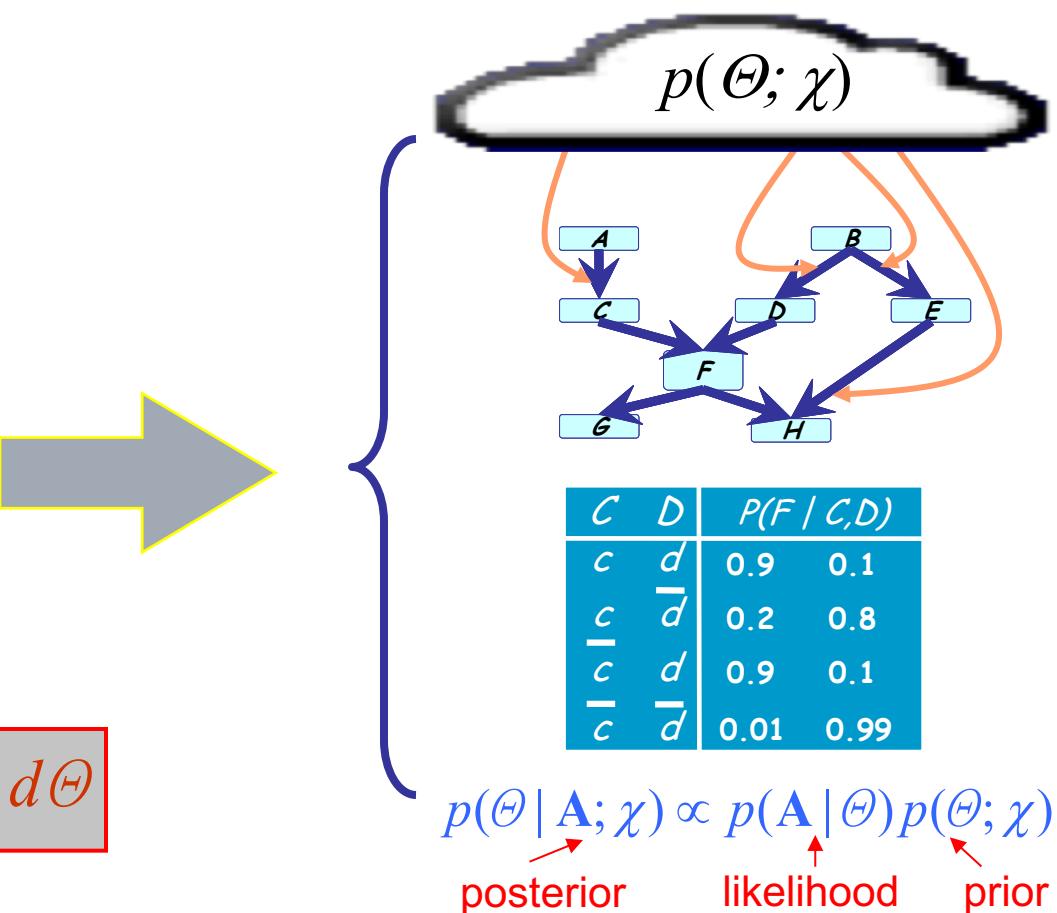


# GM: MLE and Bayesian Learning

- Probabilistic statements of  $\Theta$  is conditioned on the values of the observed variables  $A_{\text{obs}}$  and prior  $p(\Theta | \chi)$



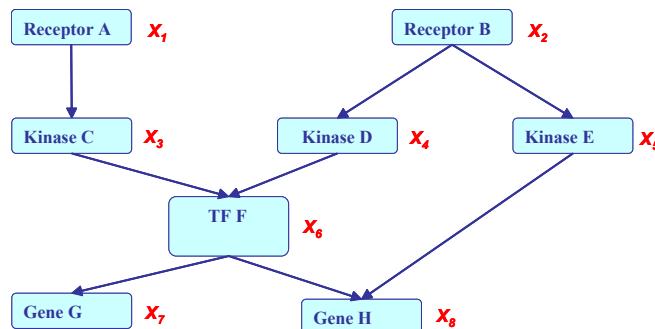
$$\Theta_{\text{Bayes}} = \int \Theta p(\Theta | \mathbf{A}, \chi) d\Theta$$





# 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.,



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

- Why we may favor a PGM?
  - Incorporation of domain knowledge and causal (logical) structures
  - Modular combination of heterogeneous parts – data fusion
  - Bayesian Philosophy  $2+2+4+4+4+8+4+8=36$ , an 8-fold reduction from  $2^8$  in representation cost !
  - Knowledge meets data





# So What Is a PGM After All?

In a nutshell:

**PGM = Multivariate Statistics + Structure**

**GM = Multivariate Obj. Func. + Structure**

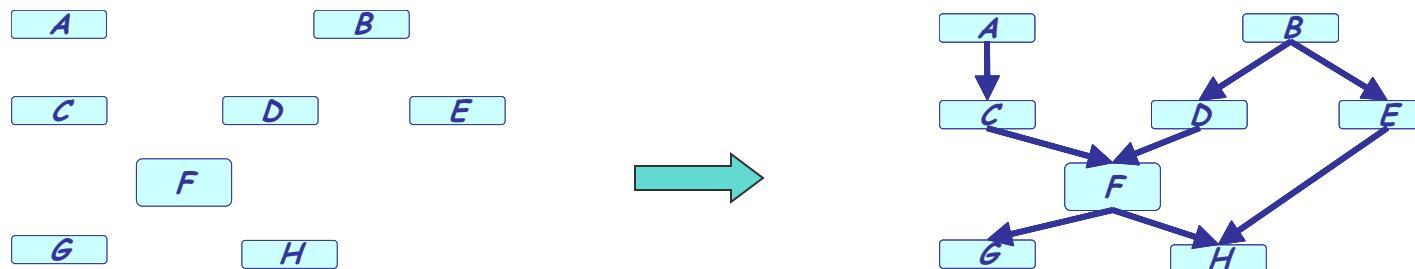
.





# So What Is a PGM After All?

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



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

$$\begin{aligned}P(X_{1:8}) &= P(X_1)P(X_2)P(X_3 | X_1X_2)P(X_4 | X_2)P(X_5 | X_2) \\&\quad P(X_6 | X_3, X_4)P(X_7 | X_6)P(X_8 | X_5, X_6)\end{aligned}$$

- ❑ A more formal description:
  - ❑ 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

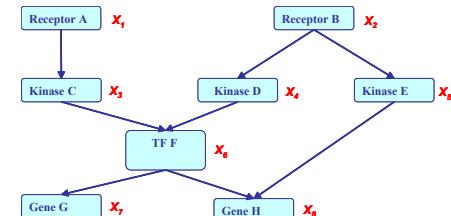




# Two types of GMs

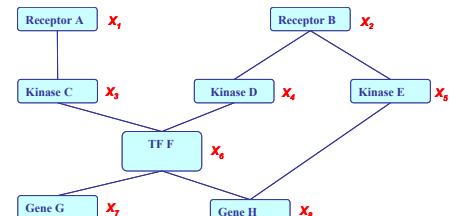
- Directed edges give causality relationships (Bayesian Network or Directed Graphical Model):

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



- Undirected edges simply give correlations between variables (Markov Random Field or Undirected Graphical model):

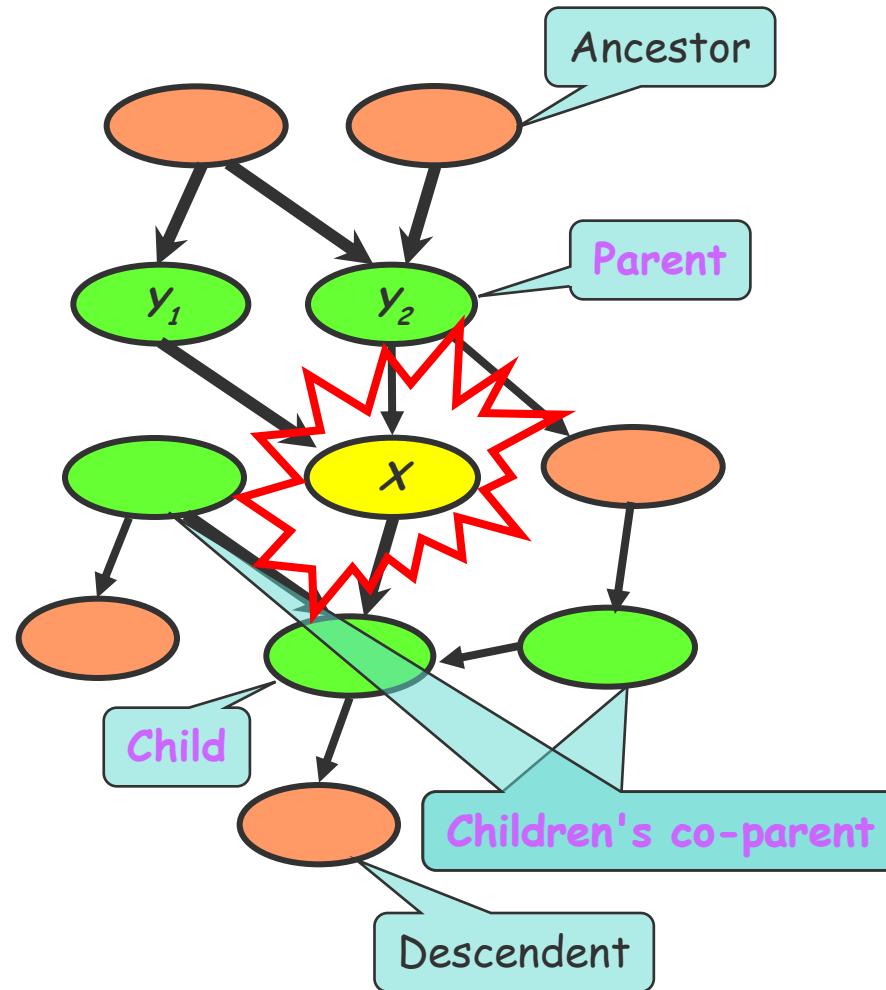
$$\begin{aligned} & P(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8) \\ &= \frac{1}{Z} \exp\{E(X_1)+E(X_2)+E(X_3, X_1)+E(X_4, X_2)+E(X_5, X_2) \\ &\quad + E(X_6, X_3, X_4)+E(X_7, X_6)+E(X_8, X_5, X_6)\} \end{aligned}$$





# Bayesian Networks

- Structure: *DAG*
- 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 dist.
- Give causality relationships, and facilitate a generative process



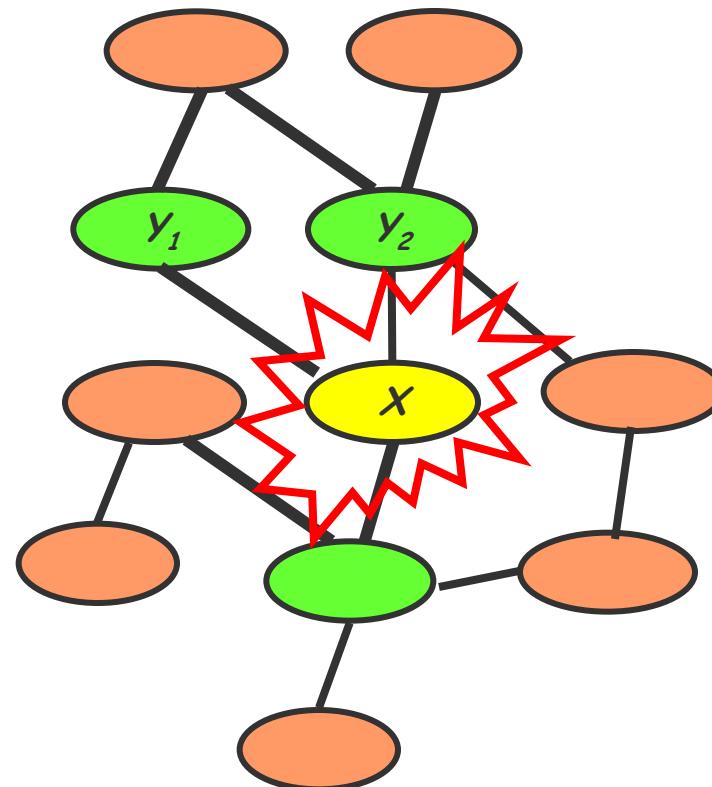
$$\overbrace{P(x|y), \neg} \equiv P(x|y)$$





# Markov Random Fields

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





# Towards structural specification of probability distribution

- Separation properties in the graph imply independence properties about the associated variables
- For the graph to be useful, any conditional independence properties we can derive from the graph should hold for the probability distribution that the graph represents
- The Equivalence Theorem

For a graph  $G$ ,

Let  $\mathcal{D}_1$  denote the family of all distributions that satisfy  $I(G)$ ,

Let  $\mathcal{D}_2$  denote the family of all distributions that factor according to  $G$ ,

Then  $\mathcal{D}_1 \equiv \mathcal{D}_2$ .





# GMs are your old friends

Density estimation

Parametric and nonparametric methods

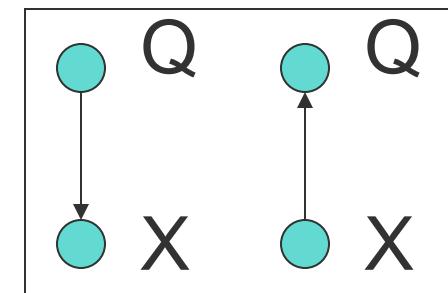
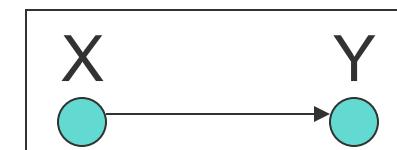
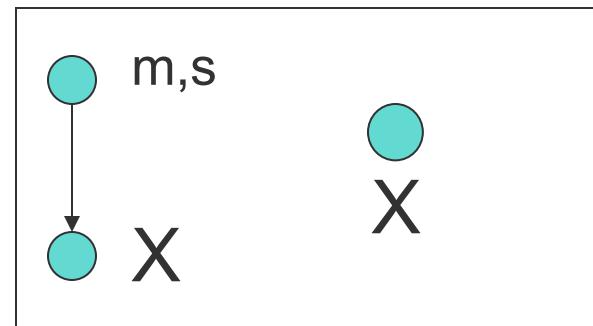
Regression

Linear, conditional mixture, nonparametric

Classification

Generative and discriminative approach

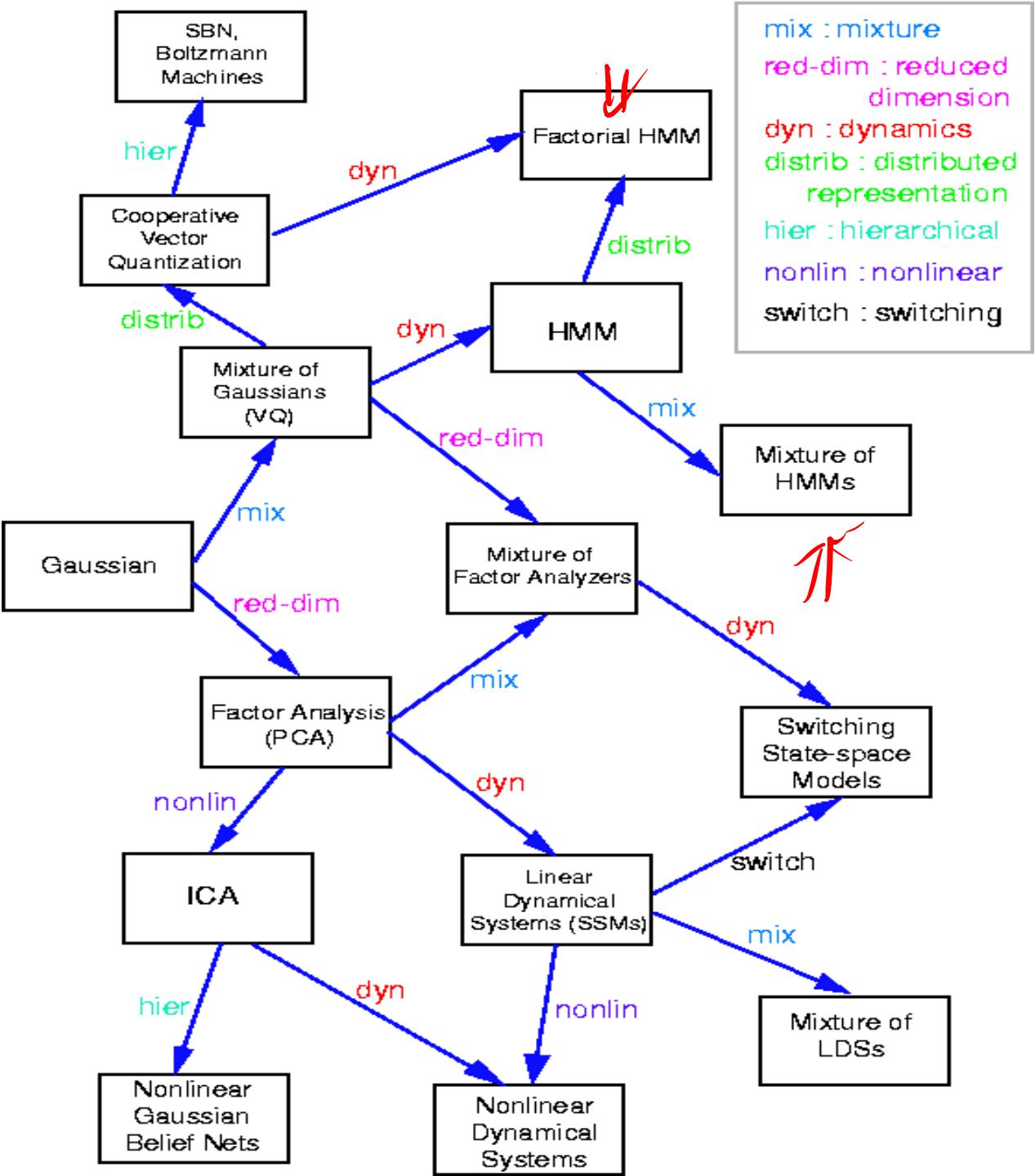
Clustering





# An (incomplete) genealogy of graphical models

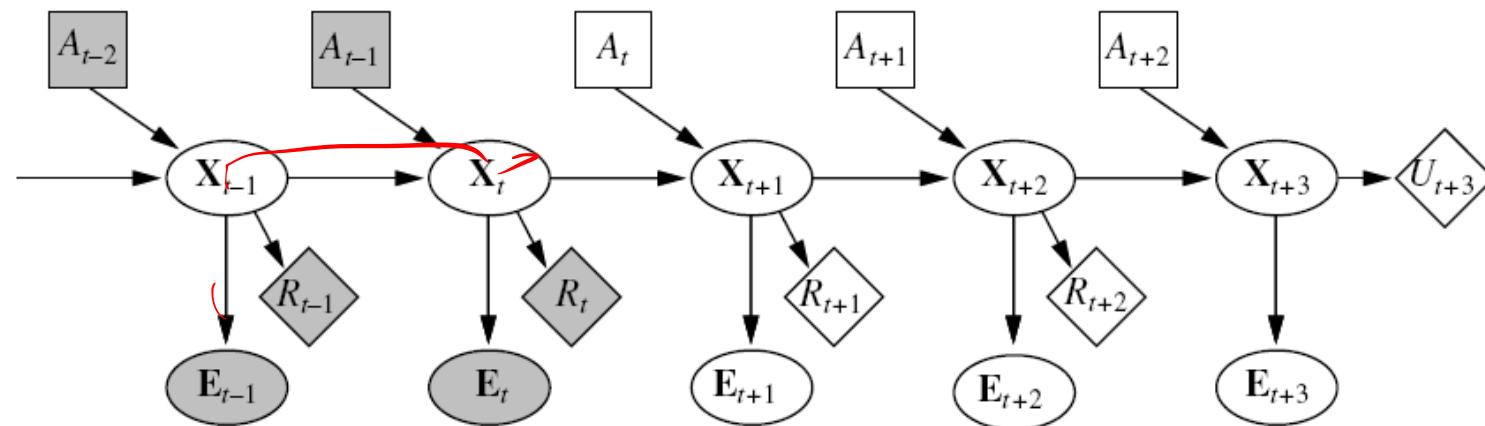
(Picture by Zoubin  
Ghahramani and  
Sam Roweis)





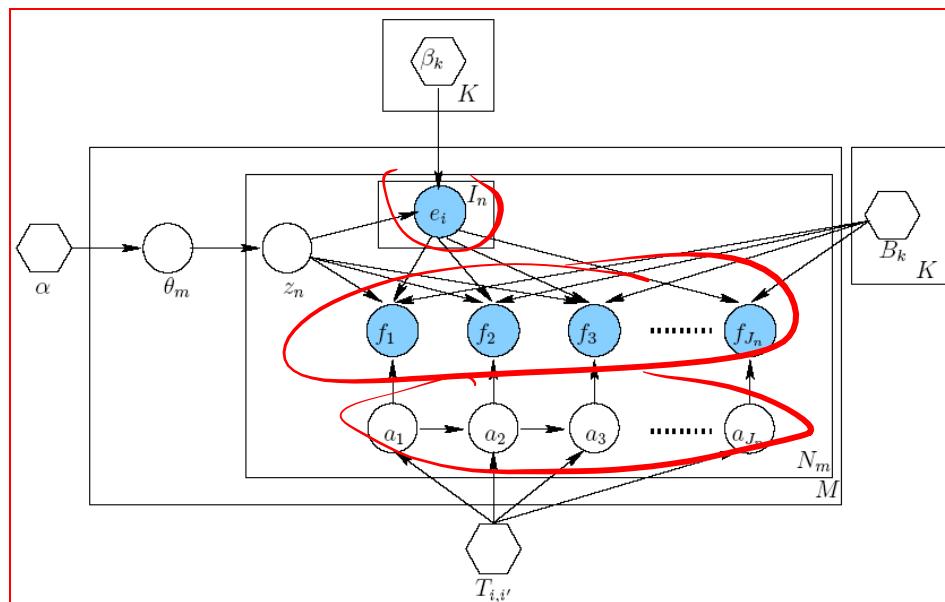
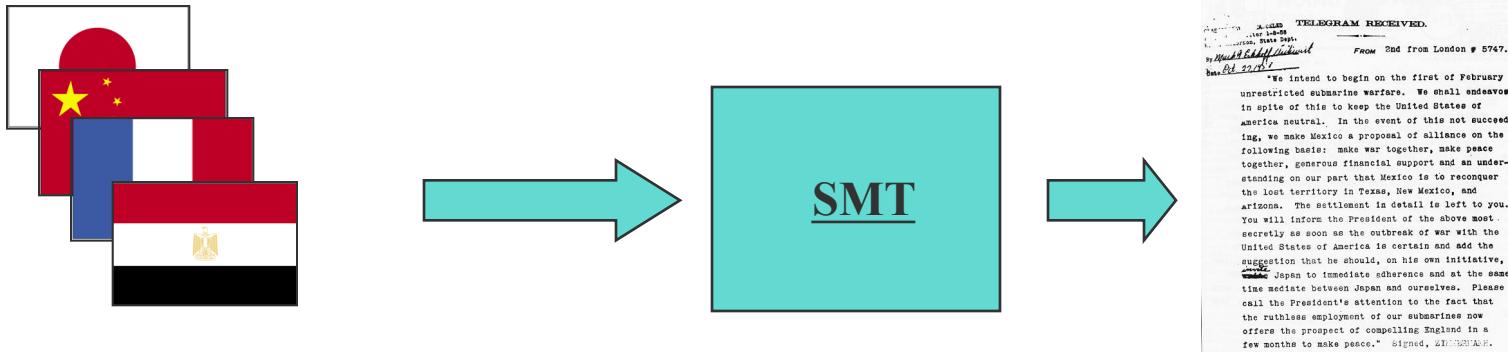
# Fancier GMs: reinforcement learning

- Partially observed Markov decision processes (POMDP)





# Fancier GMs: machine translation

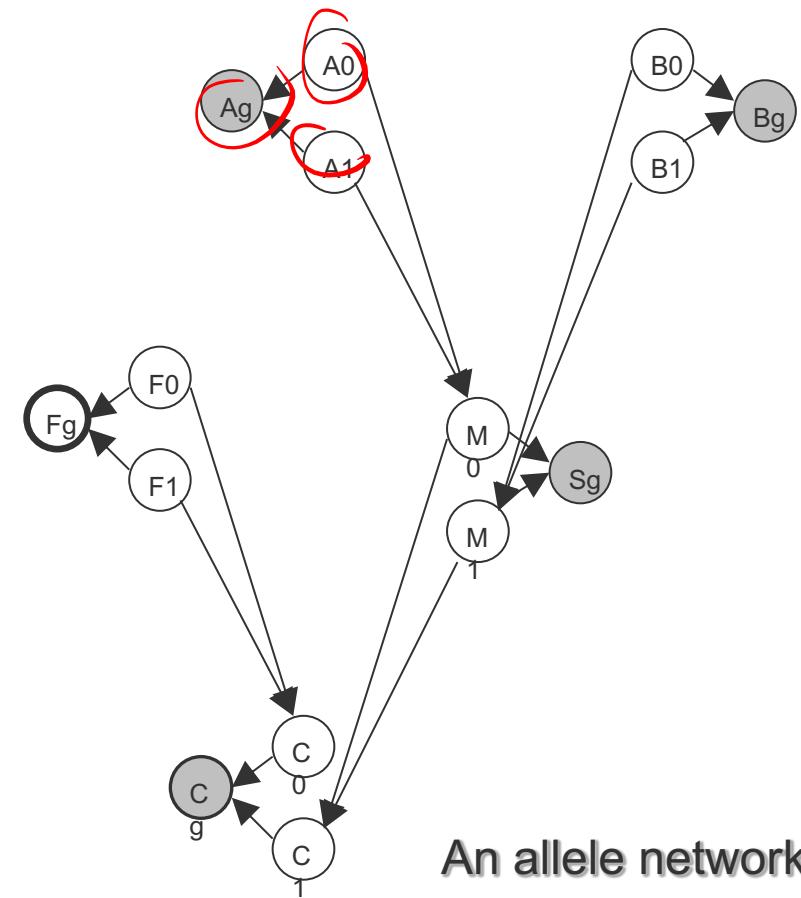
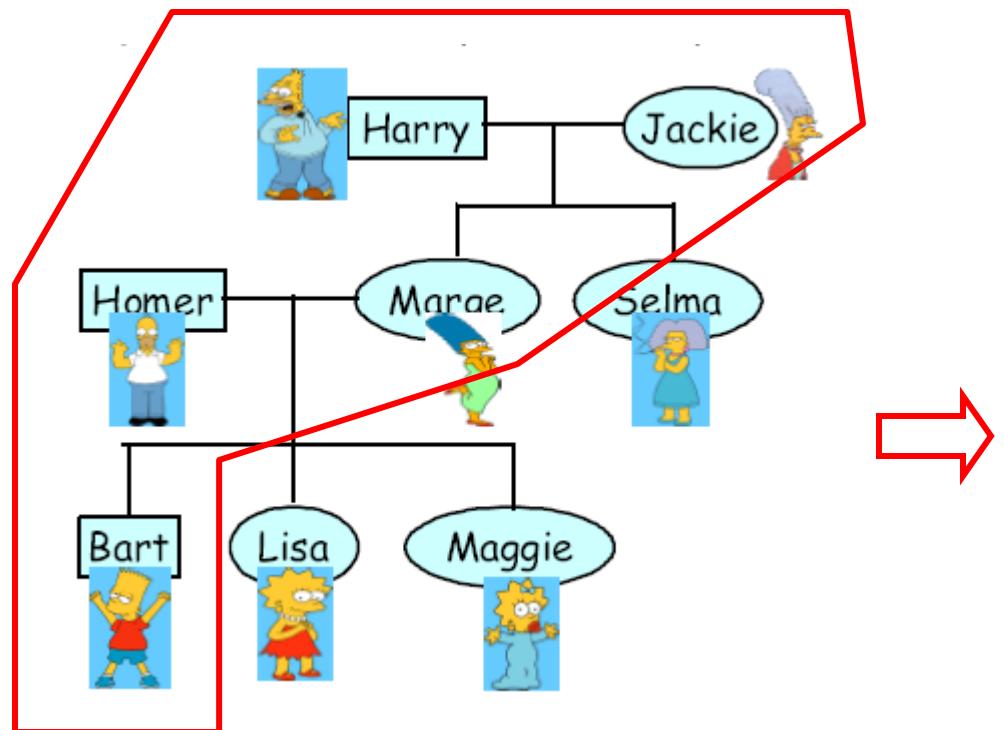


The HM-BiTAM model  
(B. Zhao and E.P Xing,  
ACL 2006)



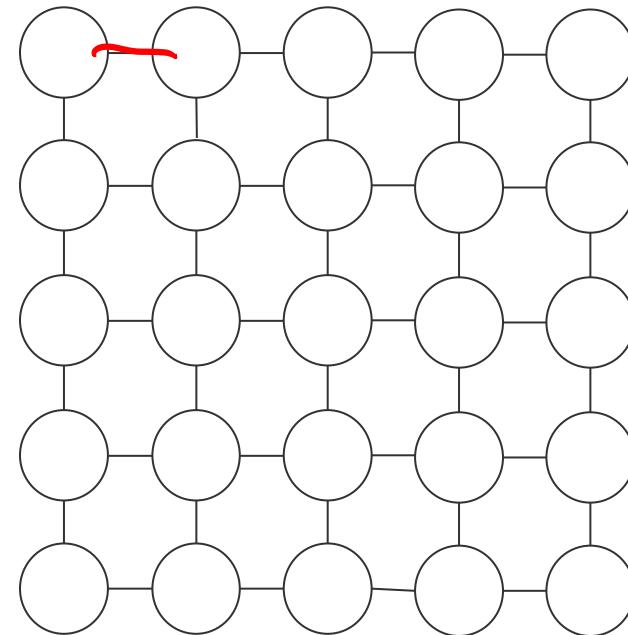
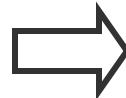
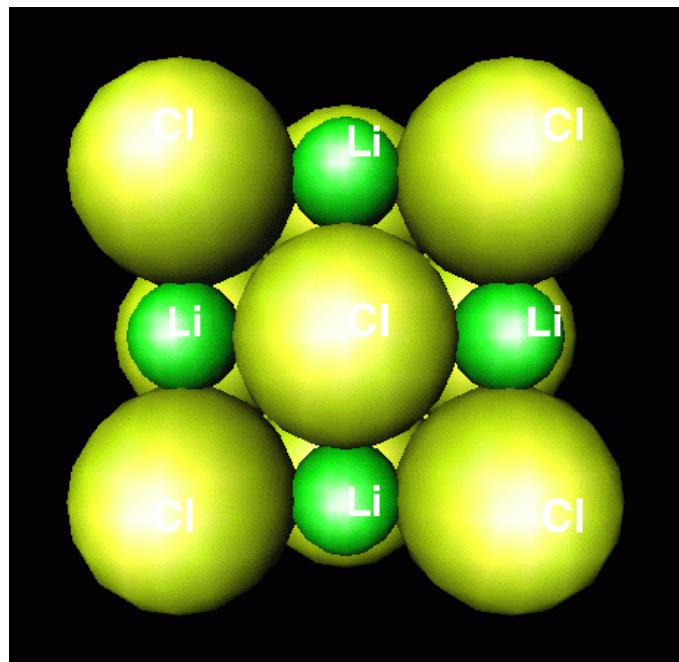


# Fancier GMs: genetic pedigree





# Fancier GMs: solid state physics



Ising/Potts model





# Application of GMs

- Machine Learning
- Computational statistics
  
- Computer vision and graphics
- Natural language processing
- Informational retrieval
- Robotic control
- Decision making under uncertainty
- Error-control codes
- Computational biology
- Genetics and medical diagnosis/prognosis
- Finance and economics
- Etc.





# Why graphical models

- A language for communication
- A language for computation
- A language for development
  
- Origins:
  - Wright 1920's
  - Independently developed by Spiegelhalter and Lauritzen in statistics and Pearl in computer science in the late 1980's





# Why graphical models

- **Probability theory** provides the **glue** whereby the parts are combined, ensuring that the system as a whole is consistent, and providing ways to interface models to data.
- The **graph theoretic** side of graphical models provides both an intuitively appealing interface by which humans can model highly-interacting sets of variables as well as a data structure that lends itself naturally to the design of efficient general-purpose algorithms.
- **Many of the classical multivariate probabilistic systems** studied in fields such as statistics, systems engineering, information theory, pattern recognition and statistical mechanics **are special cases of the general graphical model formalism**
- The graphical model framework provides a way to view all of these systems as instances of a **common underlying formalism**.

--- M. Jordan





# Plan for the Class

- ❑ Fundamentals of Graphical Models:
  - ❑ Bayesian Network and Markov Random Fields
  - ❑ Discrete, Continuous and Hybrid models, exponential family, GLIM
  - ❑ Basic representation, inference, and learning
- ❑ Advanced topics and latest developments
  - ❑ Approximate inference
    - ❑ Monte Carlo algorithms
    - ❑ Variational methods and theories
  - ❑ “Infinite” GMs: nonparametric Bayesian models
  - ❑ Optimization-theoretic formulations for GMs, e.g., Structured sparsity
  - ❑ **Deep Nets vs. GMs**
  - ❑ Nonparametric and spectral graphical models, where GM meets kernels and matrix algebra
  - ❑ Alternative GM learning paradigms,
    - ❑ e.g., Margin-based learning of GMs (where GM meets SVM)
    - ❑ e.g., Regularized Bayes: where GM meets SVM, and meets Bayesian, and meets NB ...
- ❑ Case studies: popular GMs and applications
  - ❑ Multivariate Gaussian Models
  - ❑ Conditional random fields
  - ❑ Mixed-membership, aka, Topic models





# Questions ?

