PETUUM

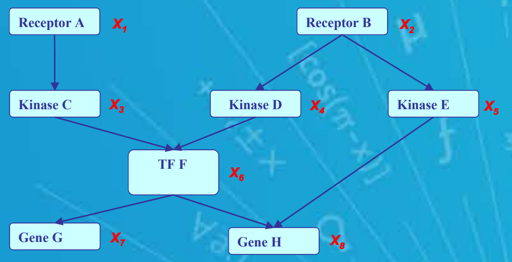
Carnegie Mellon University
School of Computer Science

Probabilistic Graphical Models

Introduction to GM

Eric Xing

Lecture 1, January 13, 2020



Reading: see class homepage

Logistics

- Class webpage: http://www.cs.cmu.edu/~epxing/Class/10708-20/

10-708

- Home
- Lectures
- Homework
- Project
- Textbook
- Policies

10-708 – Probabilistic Graphical Models

2020 Spring

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, enables 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.

- Instructor: Eric P. Xing (epxing@cs)
- Time: MW 12:00-1:20pm
- Location: Wean 7500
- Office Hours: TBA
- Piazza: https://www.piazza.com/cmu/spring2020/10708
- Gradescope: https://www.gradescope.com/courses/80181
- TAs (email, office hours):
  - Xun Zheng (xzheng1@andrew, TBA)
  - Ben Lengerich (blengeri@andrew, TBA)
  - Haohan Wang (haohanw@andrew, TBA)
  - Yiwen Yuan (yiweny@andrew, TBA)
  - Xiang Si (xsi@andrew, TBA)
  - Junxian He (junxian1@andrew, TBA)

Page generated 2020-01-08 21:27:37 EST, by jemdoc.

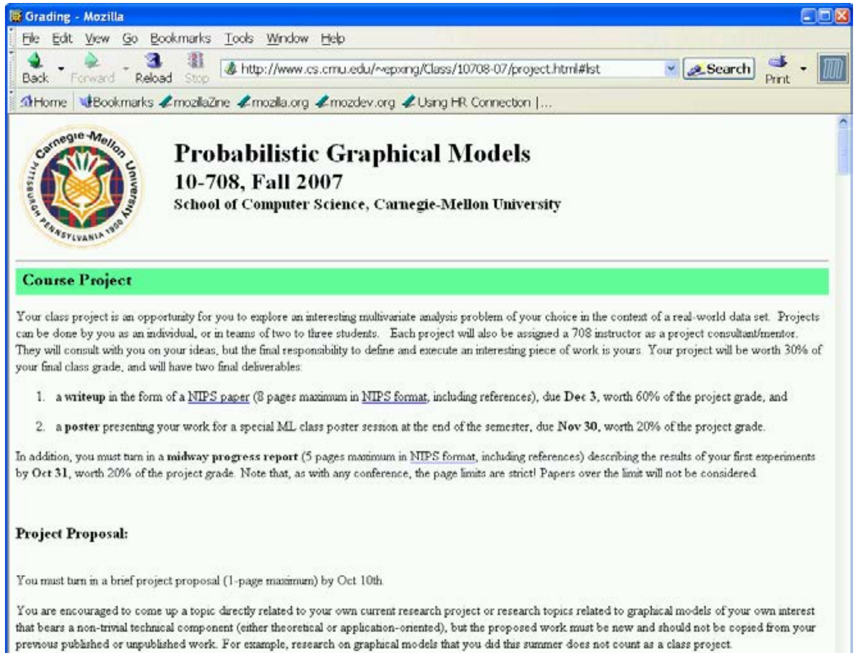
Logistics

- Textbooks:
  - Daphne Koller and Nir Friedman, Probabilistic Graphical Models
  - M. I. Jordan, An Introduction to Probabilistic Graphical Models (chapters will be made available)
- Class announcements and discussion: Piazza
- Homework submission: Gradescope
- TAs:
  - Xun Zheng
  - Ben Lengerich
  - Haohan Wang
  - Yiwen Yuan
  - Xiang Si
  - Junxian He
- Lecturer: Eric Xing
- Class Assistant: Amy Protos

Logistics

- 4 homework assignments: 50% of grade
  - Theory exercises, Implementation exercises
- Scribe duties: 10% (~once to twice for the whole semester)
- Final project: 40% 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 or 4-member team to be formed in the first three weeks, proposal, mid-way report, presentation & demo, final report → possibly conference submission !
- Bonus:
  - Contribution to discussion on Piazza
  - Complete mid-semester evaluation

Past projects:


Grading - Mozilla

File Edit View Go Bookmarks Tools Window Help

http://www.cs.cmu.edu/~epxing/Class/10708-07/project.html#lst

Probabilistic Graphical Models

10-708, Fall 2007

School of Computer Science, Carnegie-Mellon University

Course Project

Your class project is an opportunity for you to explore an interesting multivariate analysis problem of your choice in the context of a real-world data set. Projects can be done by you as an individual, or in teams of two to three students. Each project will also be assigned a 708 instructor as a project consultant/mentor. They will consult with you on your ideas, but the final responsibility to define and execute an interesting piece of work is yours. Your project will be worth 30% of your final class grade, and will have two final deliverables:

1. a **writeup** in the form of a NIPS paper (8 pages maximum in NIPS format, including references), due **Dec 3**, worth 60% of the project grade, and
2. a **poster** presenting your work for a special ML class poster session at the end of the semester, due **Nov 30**, worth 20% of the project grade.

In addition, you must turn in a **midway progress report** (5 pages maximum in NIPS format, including references) describing the results of your first experiments by **Oct 31**, worth 20% of the project grade. Note that, as with any conference, the page limits are strict! Papers over the limit will not be considered.

Project Proposal:

You must turn in a brief project proposal (1-page maximum) by Oct 10th.

You are encouraged to come up a topic directly related to your own current research project or research topics related to graphical models of your own interest that bears a non-trivial technical component (either theoretical or application-oriented), but the proposed work must be new and should not be copied from your previous published or unpublished work. For example, research on graphical models that you did this summer does not count as a class project.

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

- 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).

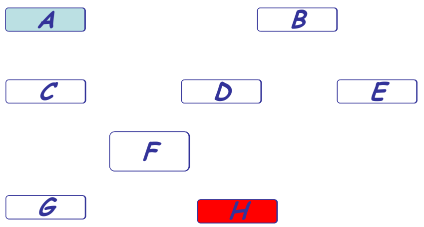
  M. Sachan, A. Dubey, S. Srivastava, E. P. Xing and Eduard Hovy, Spatial Compactness meets Topical Consistency: Jointly modeling Links and Content for Community Detection , *Proceedings of The 7th ACM International Conference on Web Search and Data Mining* (WSDM 2014).

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?

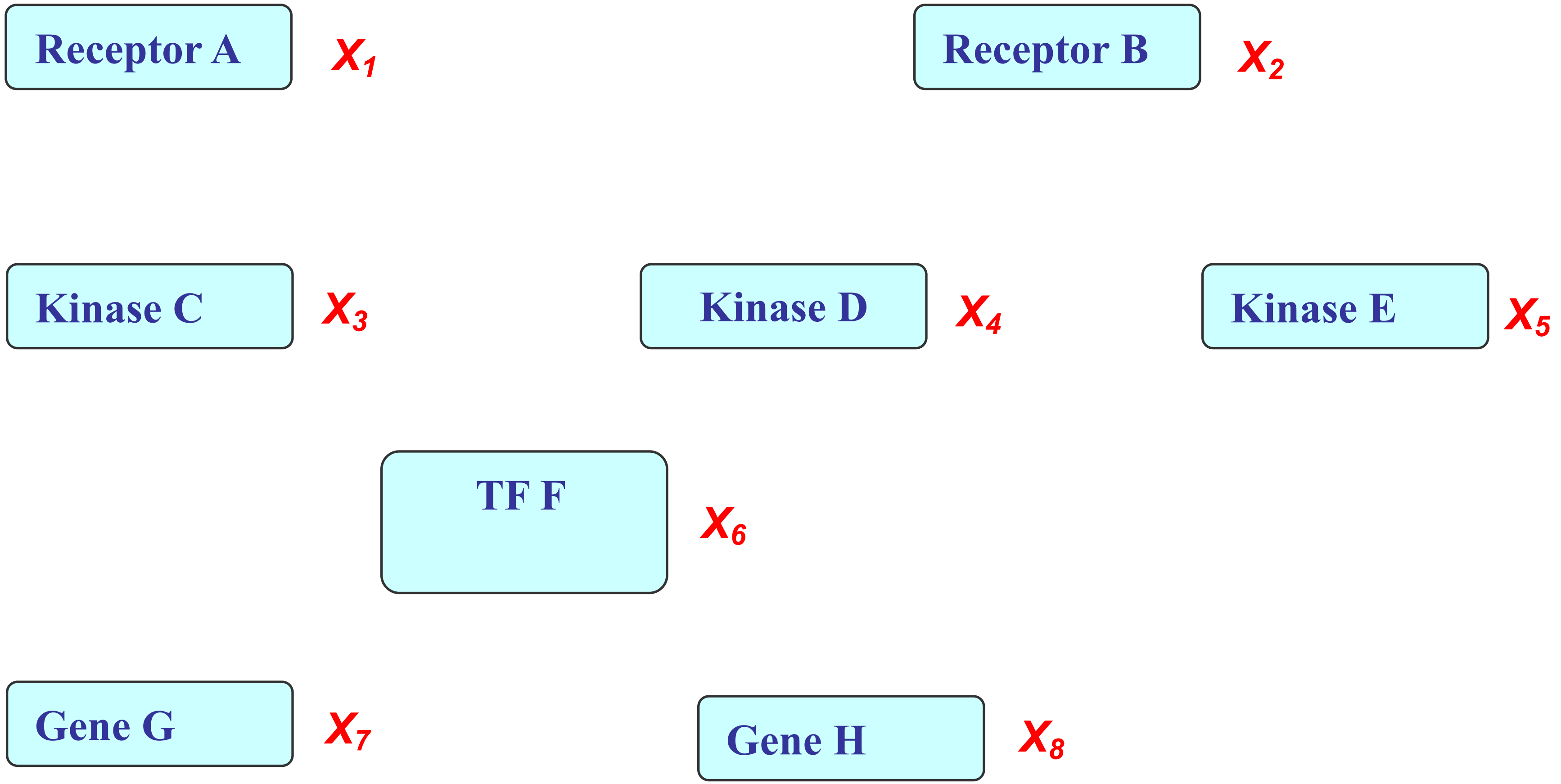


- Learning: where do we get all this probabilities?
  - Maximal-likelihood estimation? but how many data do we need?
  - Are there other est. principles?
  - Where do we put domain knowledge in terms of plausible relationships between variables, and plausible values of the probabilities?

- 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

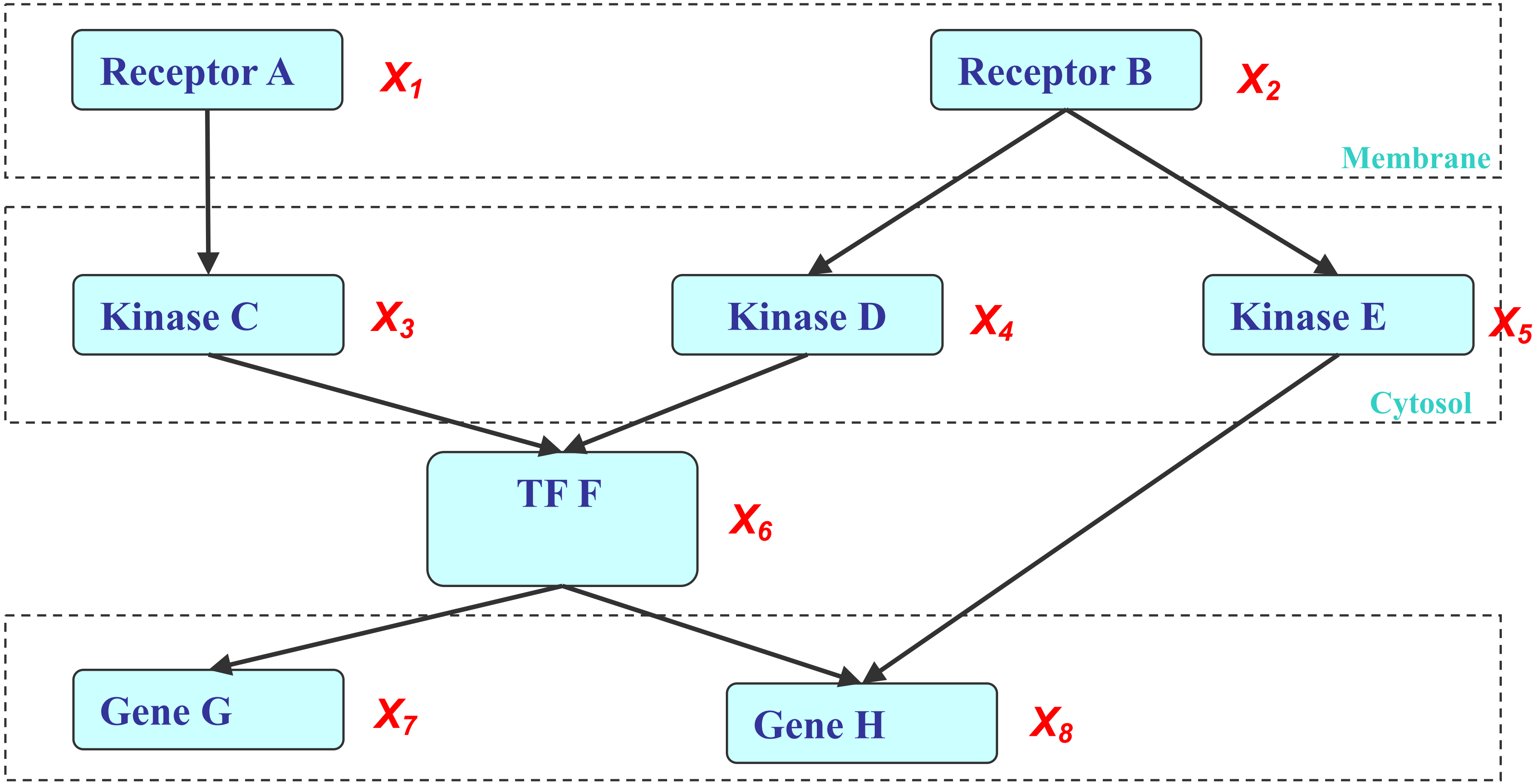
Multivariate Distribution in High-D Space

- A possible world for cellular signal transduction:



A Structured View From Domain Experts

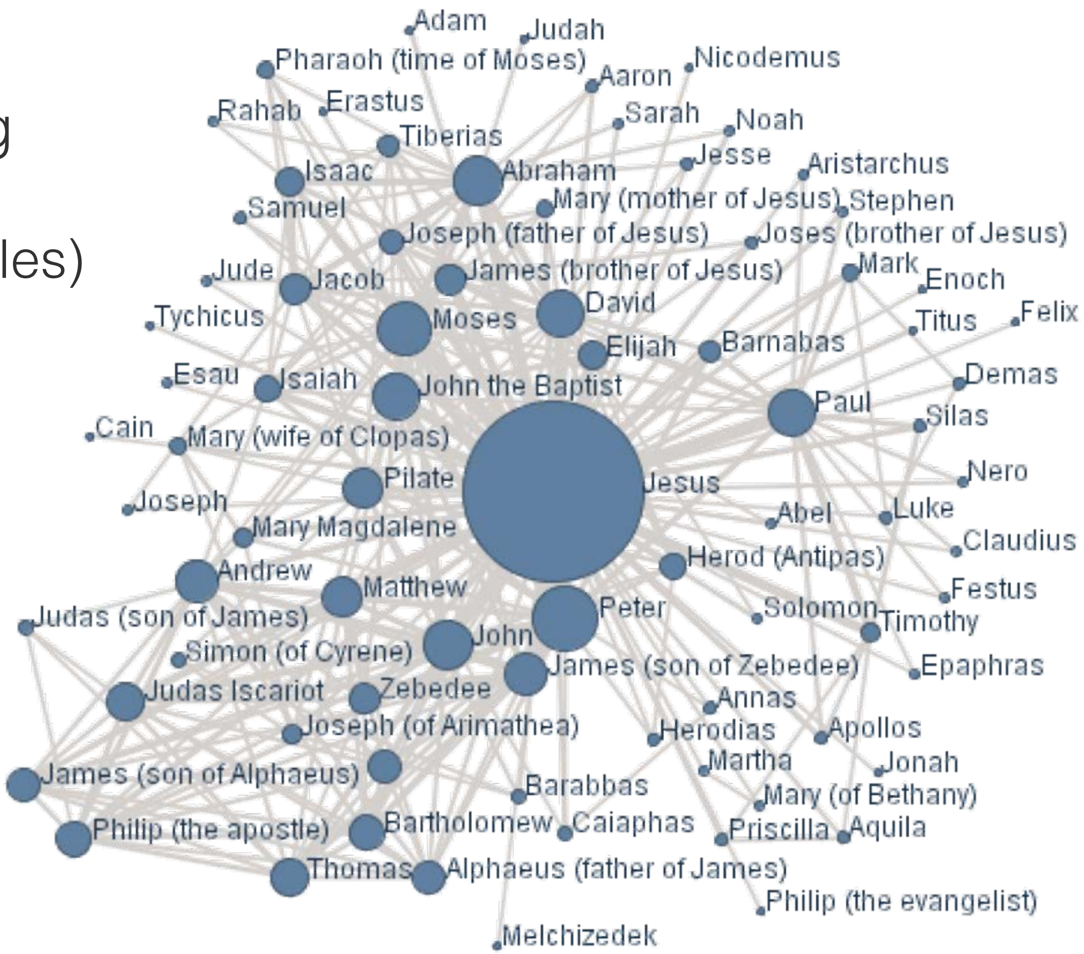
- Dependencies among variables



What are graphical models?

- Informally, a GM is just a graph representing relationship among random variables
  - Nodes: random variables (features, not examples)
  - Edges (or absence of edges): relationship

- Looks simple!
  - But detail matters, as always.
  - What exactly do we mean by relationship?



Relationship between two random variables

- Many types of relationships exist:
  - X and Y are correlated
  - X and Y are dependent
  - X and Y are independent
  - X and Y are partially correlated given Z
  - X and Y are conditionally dependent given Z
  - X and Y are conditionally independent given Z
  - X causes Y
  - Y causes X
  - ...
- Many of them can be measured by an “one number summary”

Measure of association between two random variables

- Measures of association:
  - Pearson's correlation
  - Mutual information
  - Hilbert-Schmidt Independence Criterion (HSIC)
  - Partial correlation
  - …

- Why study them? (rather than directly diving into graphical models?)
  - Gives better understanding of what graphical models really mean
  - Useful when estimating graph from data (later in the course)

Probability 101: Pearson's correlation

- Normalized covariance

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

- Captures linear dependency
  - Linear regression from X to Y gives $\beta = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$

- Important properties:
  - $X \perp\!\!\!\perp Y$ implies $\rho(X,Y) = 0$ (Why?)
  - $\rho(X,Y) = 0$ does **not** imply $X \perp\!\!\!\perp Y$ (Counterexamples?)

- Q1: Is there any measure that implies independence?
- Q2: What kind of dependency should they consider?

Strong measure of association

- Q1: Is there any measure that implies independence?
  - A1: Yes, many! We will mention two of them today.

- Q2: What kind of dependency should they consider?
  - A2: Nonlinear dependency.

- One way to construct such a measure of dependence:
  - If X ⫫ Y then joint pdf **factorizes** $P_{XY} = P_X P_Y$
  - Measure "**distance**" between P_{XY} and $P_X P_Y$
  - distance == 0 if and only if X ⫫ Y

Mutual information

- Distance between two distributions?
- Recall our old friend – the Kullback–Leibler divergence

$$\mathrm{KL}(P, Q) = \int_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)} dx$$

- Apply then we get another old friend – mutual information

$$I(X, Y) = \mathrm{KL}(P_{XY}, P_X P_Y)$$

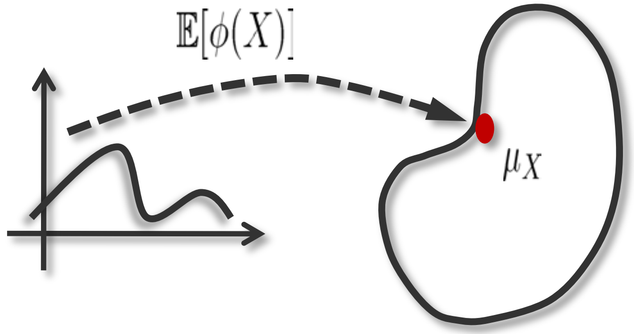
- Foundation of many topics later in the course
- I(X, Y) = 0 if and only if X ⫫ Y

Hilbert-Schmidt Independence Criterion (HSIC)

- A relatively recent(?) finding by Gretton et al. 2005
- Use maximum mean discrepancy (MMD) as the distance metric

$$\mathrm{HSIC}(X, Y) = \mathrm{MMD}(P_{XY}, P_X P_Y)$$
$$\mathrm{MMD}(P, Q) = \|\mu_k(P) - \mu_k(Q)\|_{\mathcal{H}_k}$$
$$\mu_k(P) = \mathbb{E}_{Z \sim P}[\phi(Z)] \quad \text{(kernel embedding of } P\text{)}$$
$$\phi(Z) = \text{feature map of kernel } k$$

- Looks scary! No need to know what it means for now.
  - Will cover later if anyone's interested ☺
- HSIC(X, Y) = 0 if and only if X ⫫ Y



But what do they have to do with graphical models?

- Marginal correlation/dependency graph for $\mathbf{X} = \{X_1, \ldots, X_d\}$
  - Most primitive form of graphical models one can think of
  - Connect variables that have nontrivial pairwise correlation/mutual information/HSIC/etc.

- Not very informative. Why?
  - X = height of a kid
  - Y = vocabulary of a kid
  - Z = age of a kid
  - Q1: What is the marginal dependency graph?
  - Q2: What is the graph that you think will make more sense?

Partial correlation: accounting for other variables

- Partial correlation between X and Y given a random vector Z
  - Correlation measured after eliminating linear effect of Z
  - i.e. correlation between residuals from regressing Z to X and Z to Y

$$\rho(X, Y|\mathbf{Z}) = \rho(e_X, e_Y) = \frac{\text{Cov}(e_X, e_Y)}{\sqrt{\text{Var}(e_X)]}\sqrt{\text{Var}(e_Y)}}$$

$$e_X = X - (\boldsymbol{\beta}_X^T \mathbf{Z} + \text{intercept}_X)$$

$$e_Y = Y - (\boldsymbol{\beta}_Y^T \mathbf{Z} + \text{intercept}_Y)$$

- Similar to Pearson's correlation:
  - X ⫫ Y | Z implies ρ(X,Y | Z) = 0
  - ρ(X,Y | Z) = 0 does not imply X ⫫ Y | Z

Partial correlation graphs

- Partial correlation graph for $\mathbf{X} = \{X_1, \ldots, X_d\}$
  - A more informative graphical model than marginal dependency graph
  - Connect variables with nontrivial partial correlation given the rest
  - Recall the height-vocab-age example (assuming everything is linear)

- A deeper look at the d x d partial correlation matrix R with

$$R_{ij} = \rho(X_i, X_j | \mathbf{X}_{-ij})$$

  - Looks scary! (So many regressions to run?!)
  - But turns out R is just some version of inverse covariance matrix Θ
  - Homework ☺

$$R_{ij} = -\frac{\Theta_{ij}}{\sqrt{\Theta_{ii}}\sqrt{\Theta_{jj}}}$$

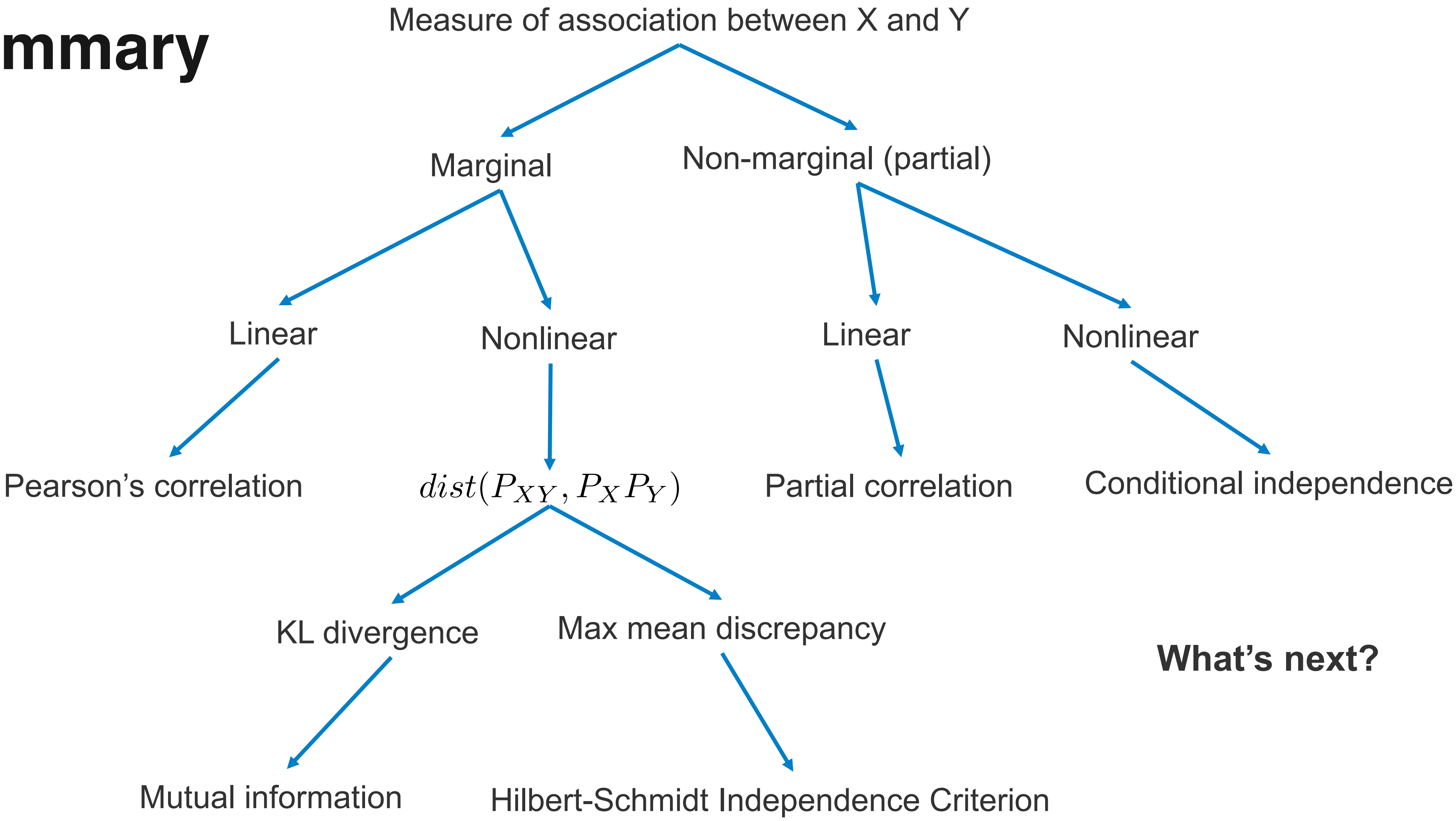
Conditional independence

- How do we measure conditional (in)dependence?
  - After seeing strong dependency measures and partial correlation, conditional independency appears to be harder than we thought…

- Ancient wisdom: if something is hard, assume Gaussian.

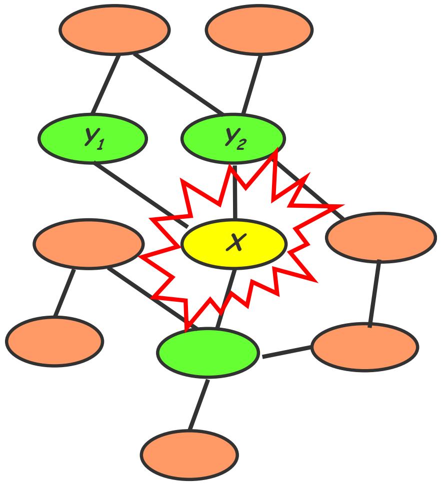
- If (X, Y, Z) are jointly Gaussian, $\rho(X,Y \mid Z) = 0$ if and only if $X \perp\!\!\!\perp Y \mid Z$
  - We will see later that many papers with Gaussian assumption rely on this fact, even though it is rarely explicitly stated

Short Summary



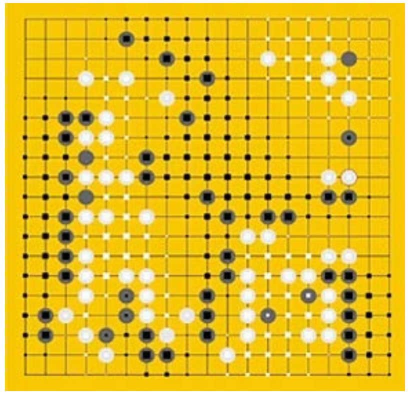
What's next?

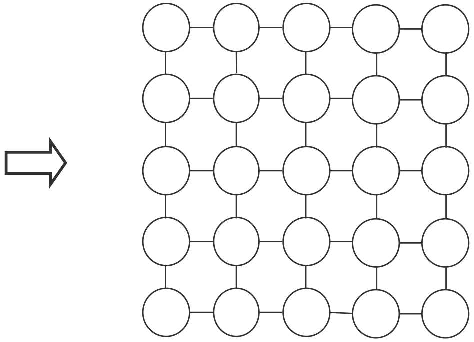
Lecture 2: Conditional independence graph



- Go by many different names
  - Conditional independence graphs (CIG)
  - Markov networks (MN)
  - Markov random fields (MRF)
  - Undirected graphical models (UG)

- Many interesting properties, widely used in physics, statistics, computer vision, NLP, deep learning, bioinformatics, coding theory, finance, …

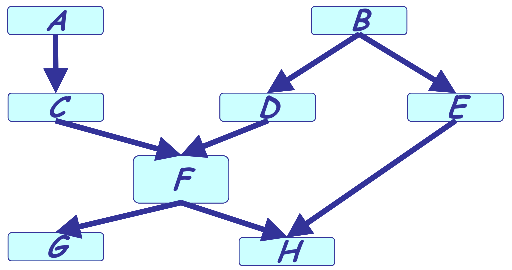




Ising/Potts model

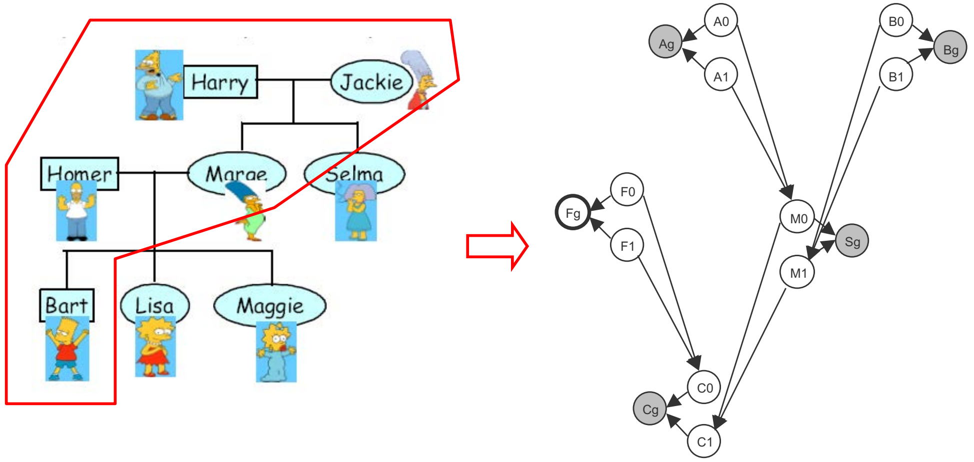
Lecture 3: Directed graphical models

- Another major class of models, also has many names:
  - Directed graphical models
  - Directed acyclic graphs (DAG) (cyclic model exists but hard to work with)
  - Bayesian networks (BN)
  - Structural equation models (SEM)
  - Structural causal models (SCM)
  - …



$$P(X_{1:8}) = P(X_1)P(X_2)P(X_3 \mid X_1 X_2)P(X_4 \mid X_2)P(X_5 \mid X_2)$$
$$P(X_6 \mid X_3, X_4)P(X_7 \mid X_6)P(X_8 \mid X_5, X_6)$$

- Powerful language to express structured knowledge



Lecture 4-13 (tentative): Inference and Learning

- Given a graphical model representing our knowledge

- Inference:
  - What is the marginal/conditional density?
  - What is the mean of the marginal/conditional?
  - What is the mode of the marginal/conditional?
  - Can we draw samples from the marginal/conditional?
  - …

- Learning: Statistical parameter estimation and model selection

Lecture 15-end (tentative): Modern GMs

- Relationship between deep learning and graphical models
- Deep generative models and their unified view
- Reinforcement learning as probabilistic inference
- GMs on functions and sets
- Bayesian nonparametrics
- Large-scale algorithms and systems

- 2-3 open slots:
  - We will list several candidate topics
  - Your voice matters!

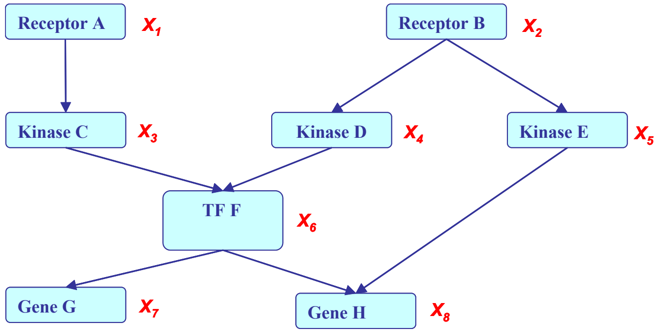
Why graphical models

- 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

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.,

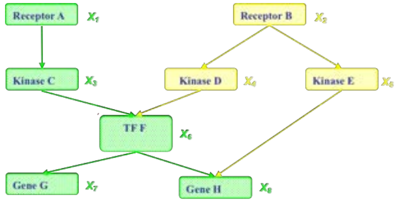


$$P(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8)$$
$$= P(X_1)\, P(X_2)\, P(X_3 | X_1)\, P(X_4 | X_2)\, P(X_5 | X_2)\, P(X_6 | X_3, X_4)\, P(X_7 | X_6)\, P(X_8 | X_5, X_6)$$

- 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



  - Bayesian Philosophy
    - Knowledge meets data

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.

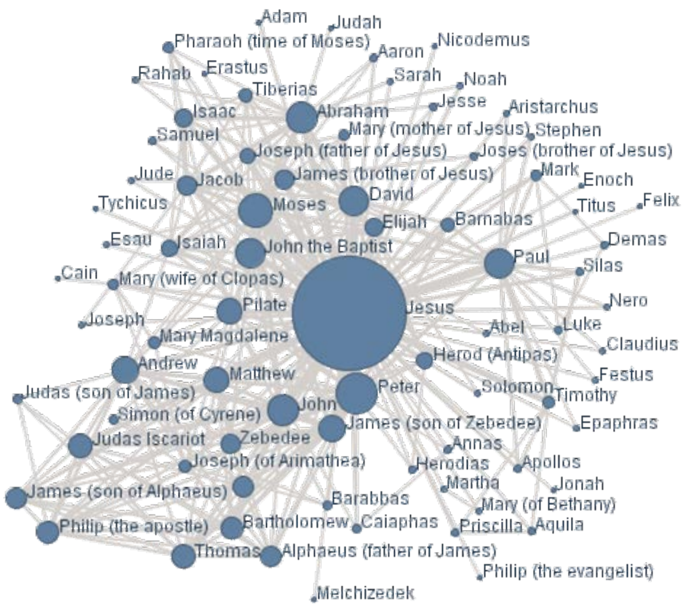
--- M. Jordan

Questions?

Appendix

What Are Graphical Models?

Graph



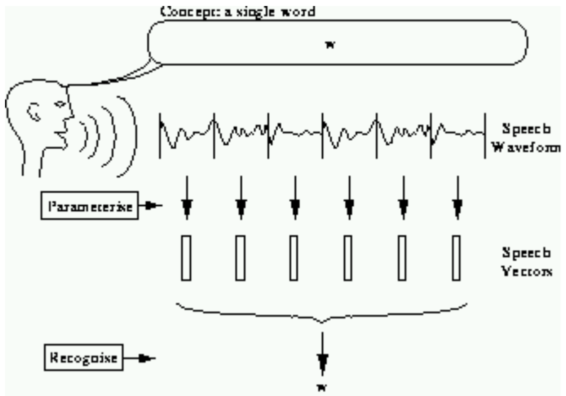
Model

$$M_G$$

Data

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

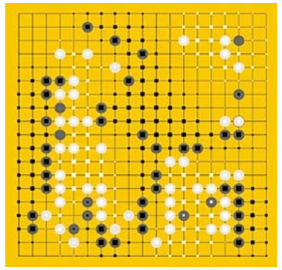
Reasoning under uncertainty!



Speech recognition

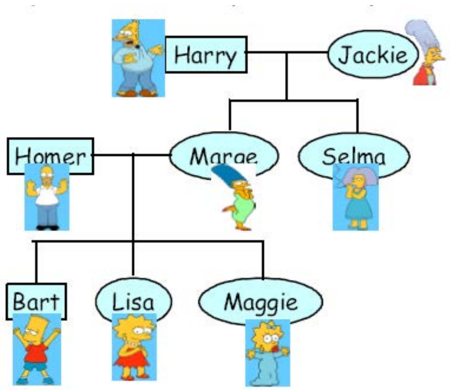
Information retrieval

Computer vision

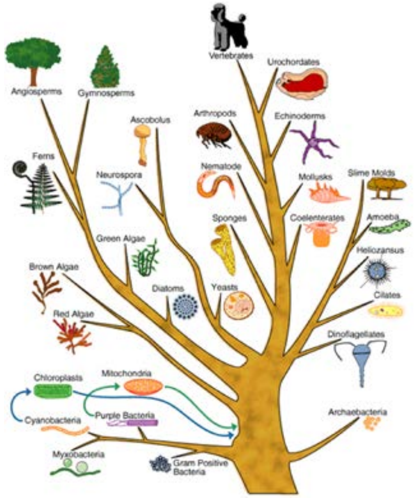


Games

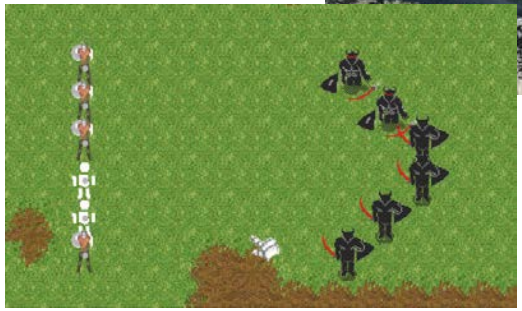
Robotic control



Pedigree



Evolution



Planning

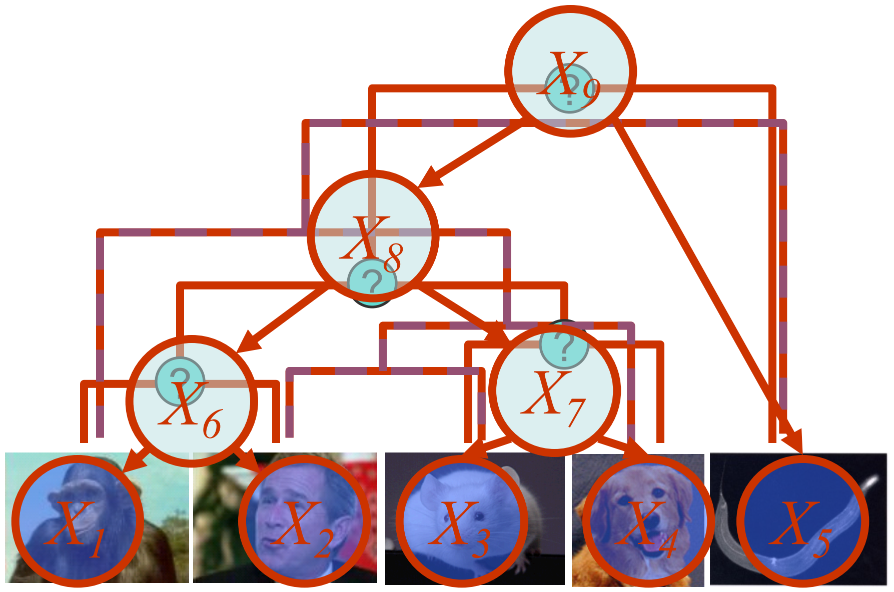
The Fundamental Questions

- Representation
  - How to capture/model uncertainties in possible worlds?
  - How to encode our domain knowledge/assumptions/constraints?
- Inference
  - How do I answer questions/queries according to my model and/or based given data?

$$\text{e.g.:}\ \ P(X_i \mid \mathcal{D})$$

- Learning
  - What model is "right" for my data?

$$\text{e.g.:}\ \ \mathcal{M} = \arg\max_{\mathcal{M} \in \mathbb{M}} F(\mathcal{D}; \mathcal{M})$$

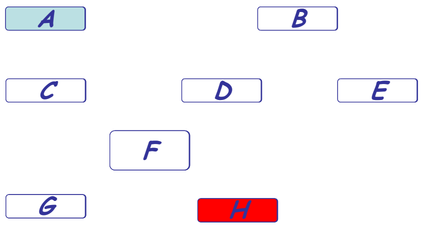


Recap of Basic Prob. Concepts

- 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?



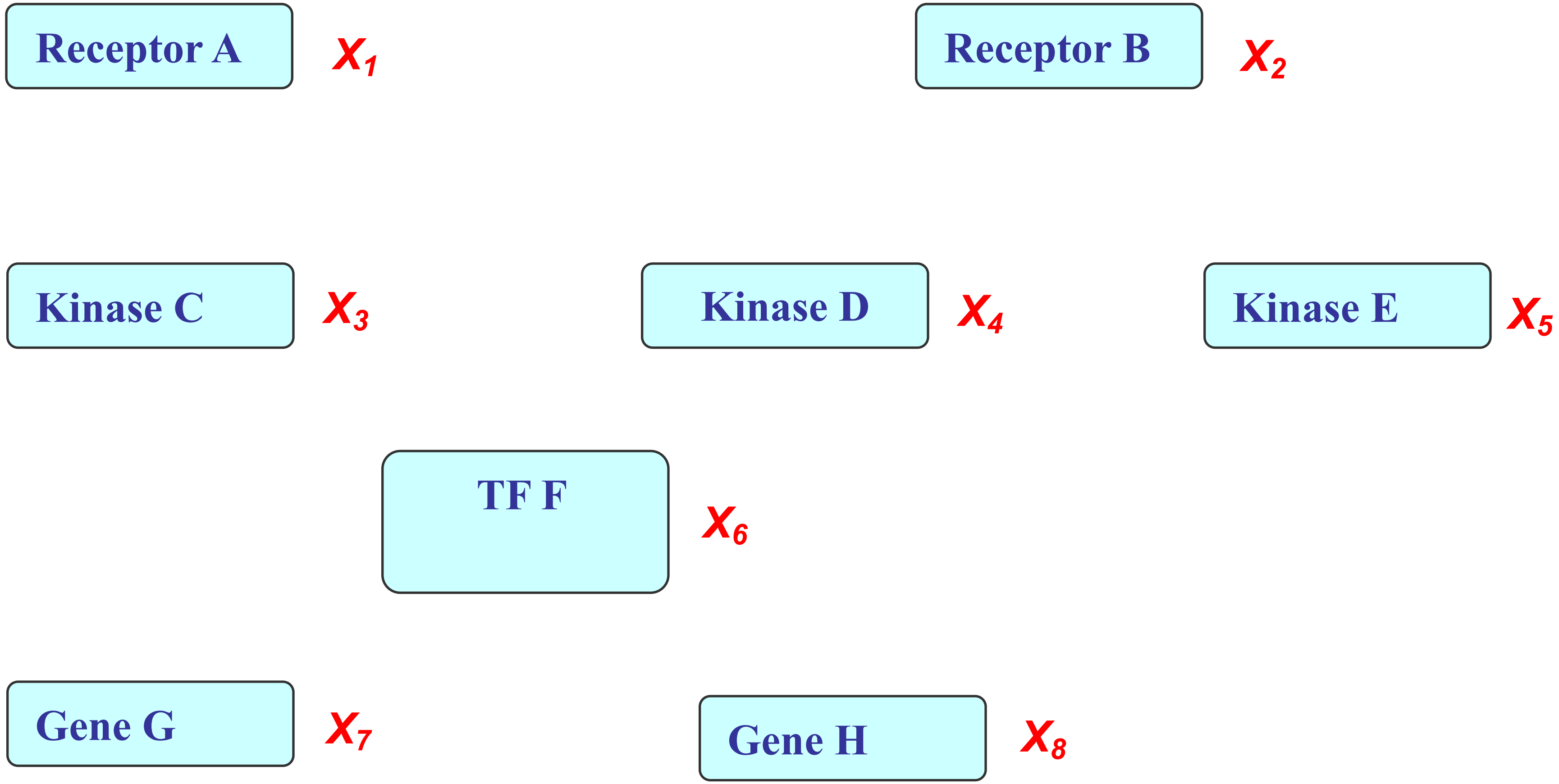
- Learning: where do we get all this probabilities?
  - Maximal-likelihood estimation? but how many data do we need?
  - Are there other est. principles?
  - Where do we put domain knowledge in terms of plausible relationships between variables, and plausible values of the probabilities?

- 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?

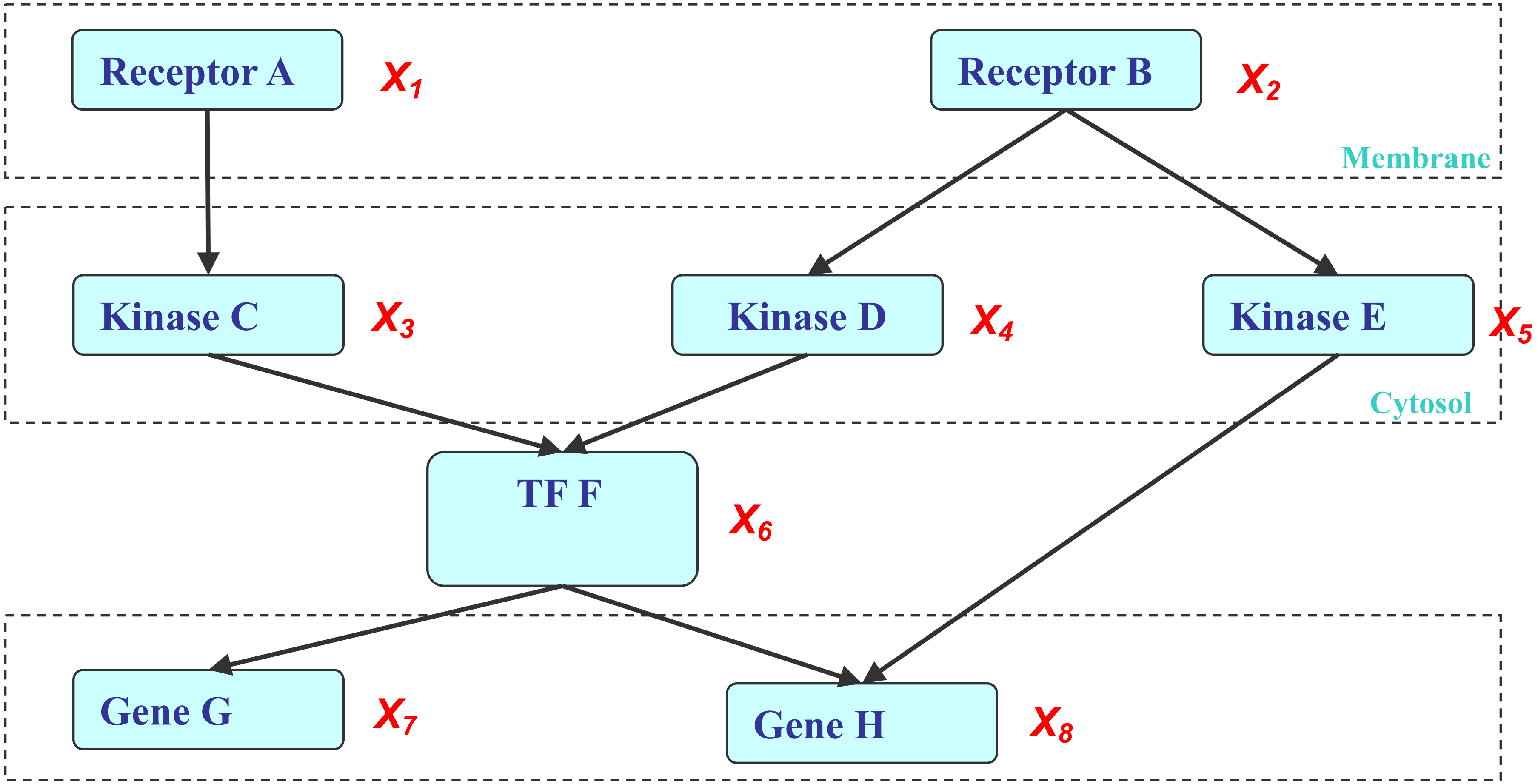
--- Multivariate Distribution in High-D Space

- A possible world for cellular signal transduction:



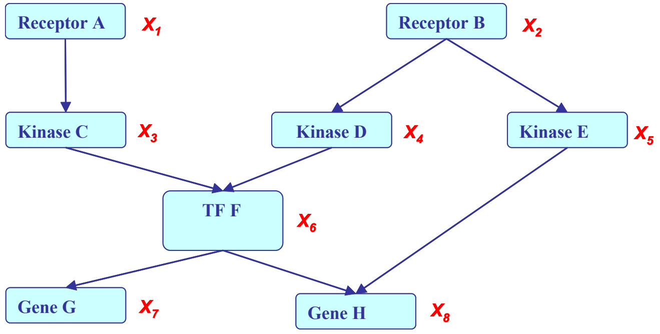
GM: Structure Simplifies Representation

- 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.,



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

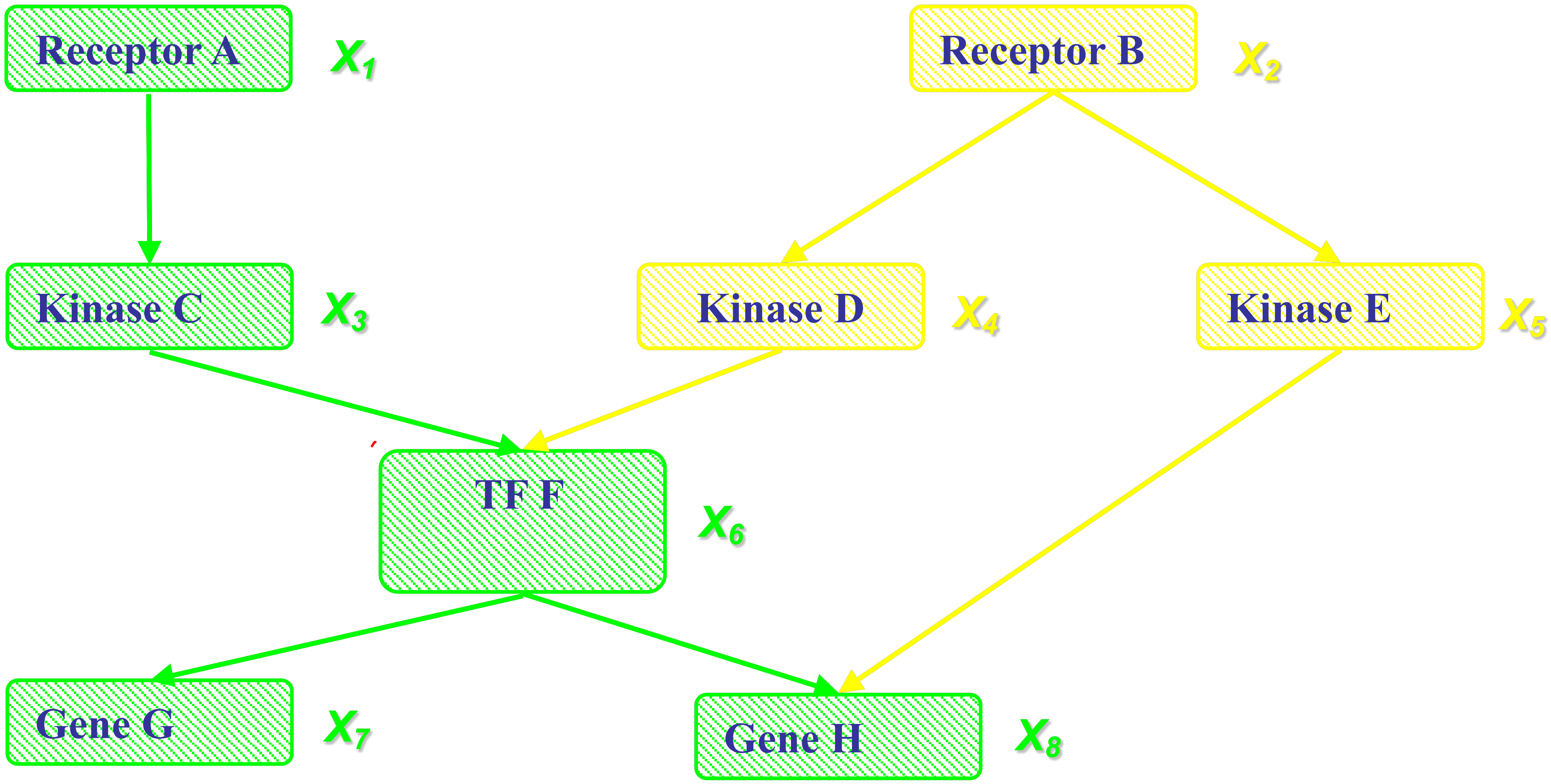
$$= P(X_1)\, P(X_2)\, P(X_3 | X_1)\, P(X_4 | X_2)\, P(X_5 | X_2)\, P(X_6 | X_3, X_4)\, P(X_7 | X_6)\, P(X_8 | X_5, X_6)$$

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

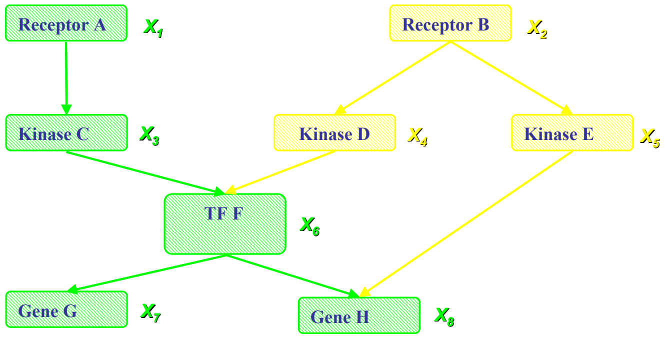


More Data Integration

- 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.,



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

- 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:

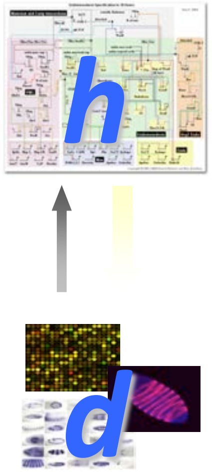
Posterior probability

Likelihood

Prior probability

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

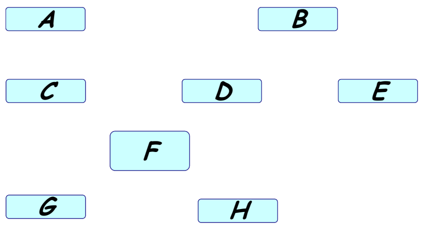
Sum over space of hypotheses



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

GM: MLE and Bayesian Learning

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



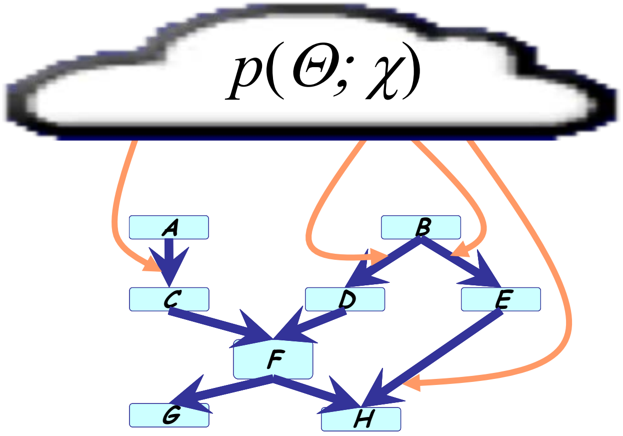
(A,B,C,D,E,…)=(T,F,F,T,F,…)

A= (A,B,C,D,E,…)=(T,F,T,T,F,…)

……..

(A,B,C,D,E,…)=(F,T,T,T,F,…)

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



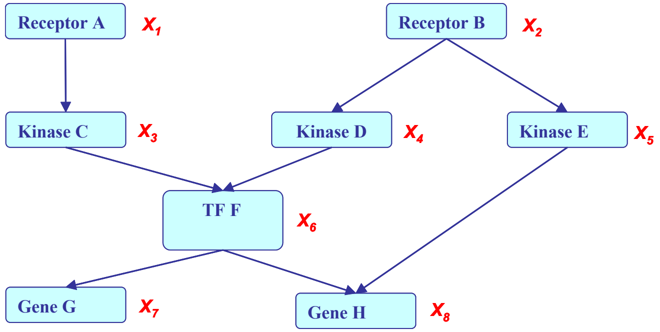
C	D	P(F | C,D)	
c	d	0.9	0.1
c	$\bar{d}$	0.2	0.8
$\bar{c}$	d	0.9	0.1
$\bar{c}$	$\bar{d}$	0.01	0.99

$$p(\Theta \mid \mathbf{A}; \chi) \propto p(\mathbf{A} \mid \Theta) \, p(\Theta; \chi)$$

posterior likelihood prior

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.,



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

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

- Why we may favor a PGM?
  - Incorporation of domain knowledge and causal (logical) structures
  - Modular combination of heterogeneous parts – data fusion
  - Bayesian Philosophy
    - Knowledge meets data

2+2+4+4+4+8+4+8=36, an 8-fold reduction from 2^8 in representation cost !

So What Is a PGM After All?

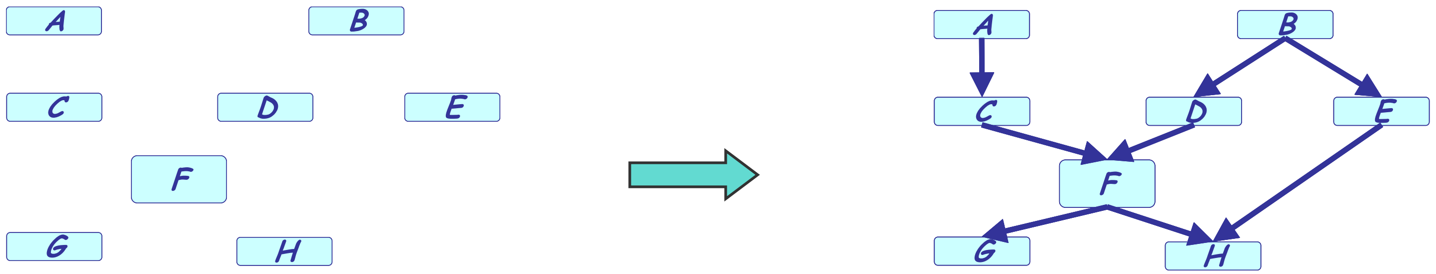
In a nutshell:

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*



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

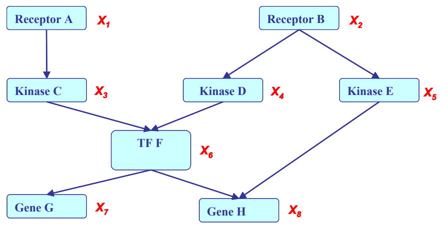
$$P(X_{1:8}) = P(X_1)P(X_2)P(X_3 \mid X_1X_2)P(X_4 \mid X_2)P(X_5 \mid X_2)$$
$$P(X_6 | X_3, X_4)P(X_7 | X_6)P(X_8 | X_5, X_6)$$

- 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

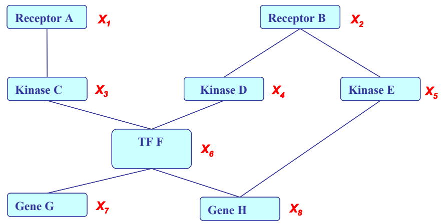
- Directed edges give causality relationships (Bayesian Network or Directed Graphical Model):

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



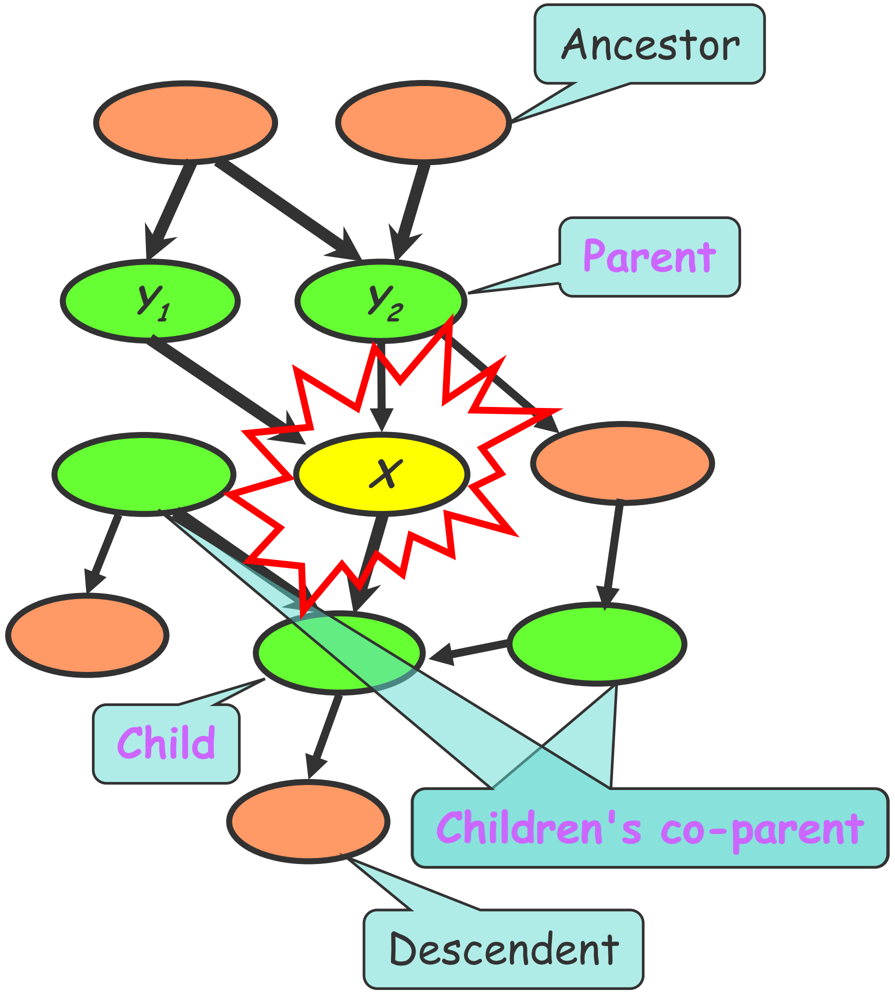
- Undirected edges simply give correlations between variables (Markov Random Field or Undirected Graphical model):

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



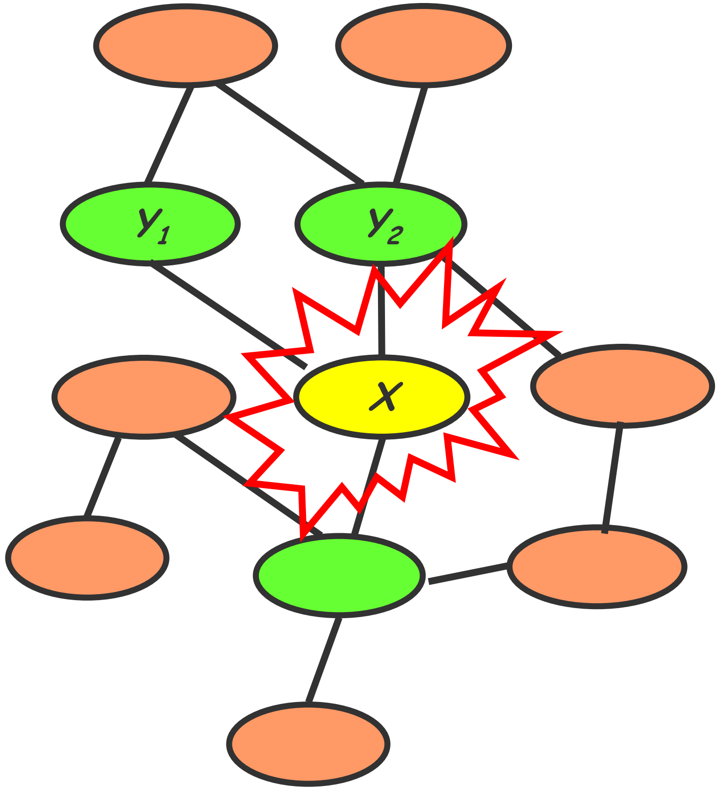
Bayesian Networks

- 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



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 D_1 denote the family of all distributions that satisfy I(G),

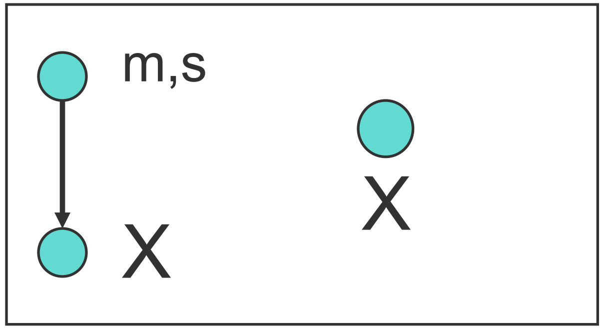
  Let D_2 denote the family of all distributions that factor according to G,

  Then $D_1 \equiv D_2$.

GMs are your old friends

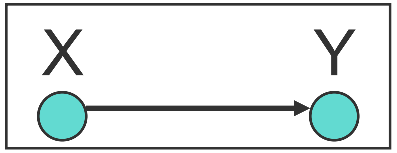
Density estimation

Parametric and nonparametric methods



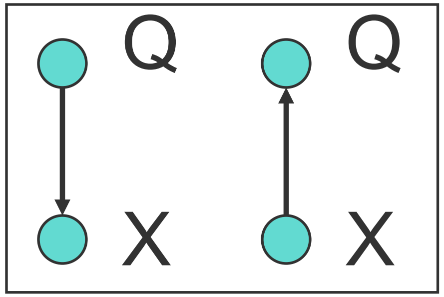
Regression

Linear, conditional mixture, nonparametric



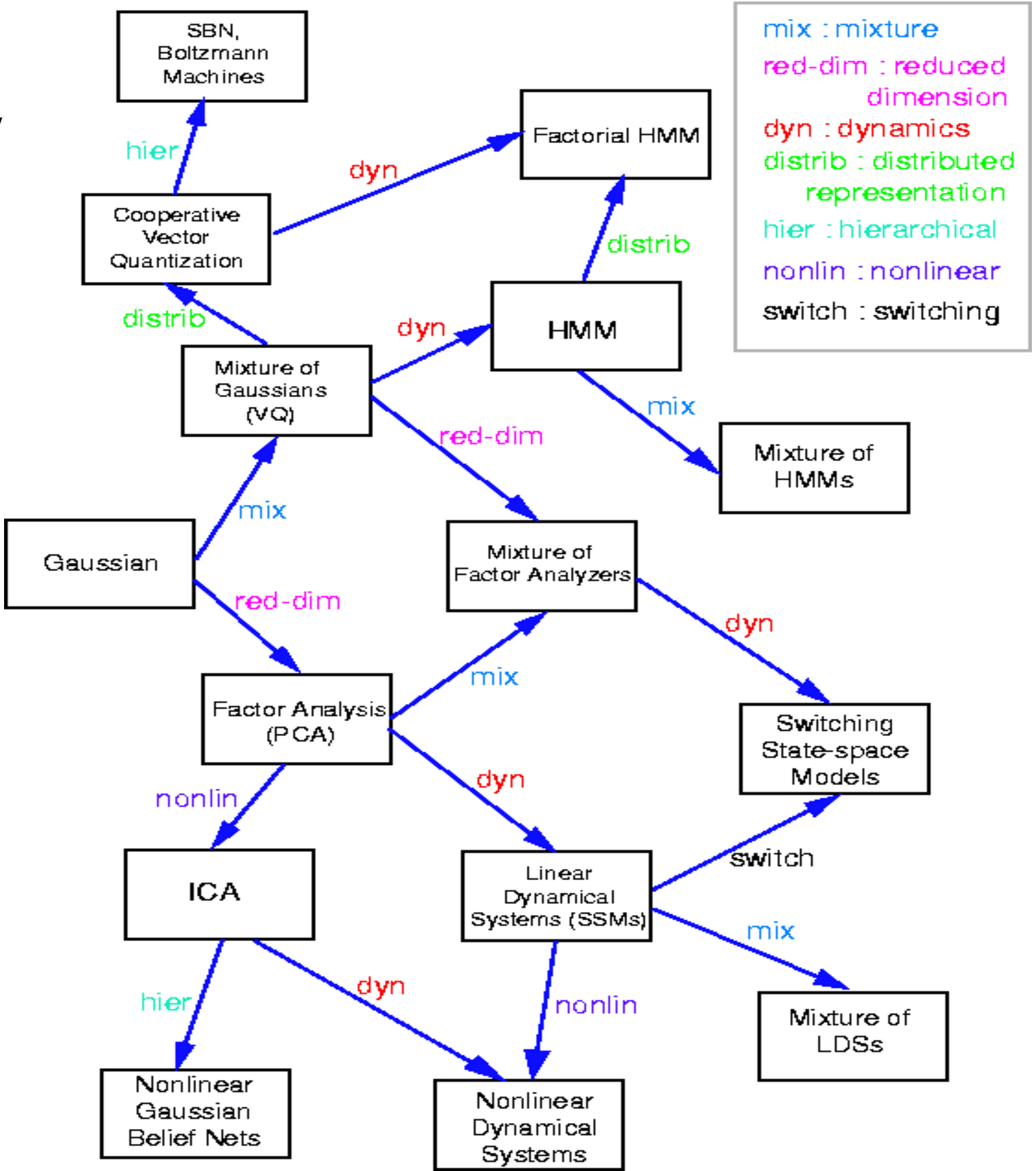
Classification

Generative and discriminative approach



Clustering

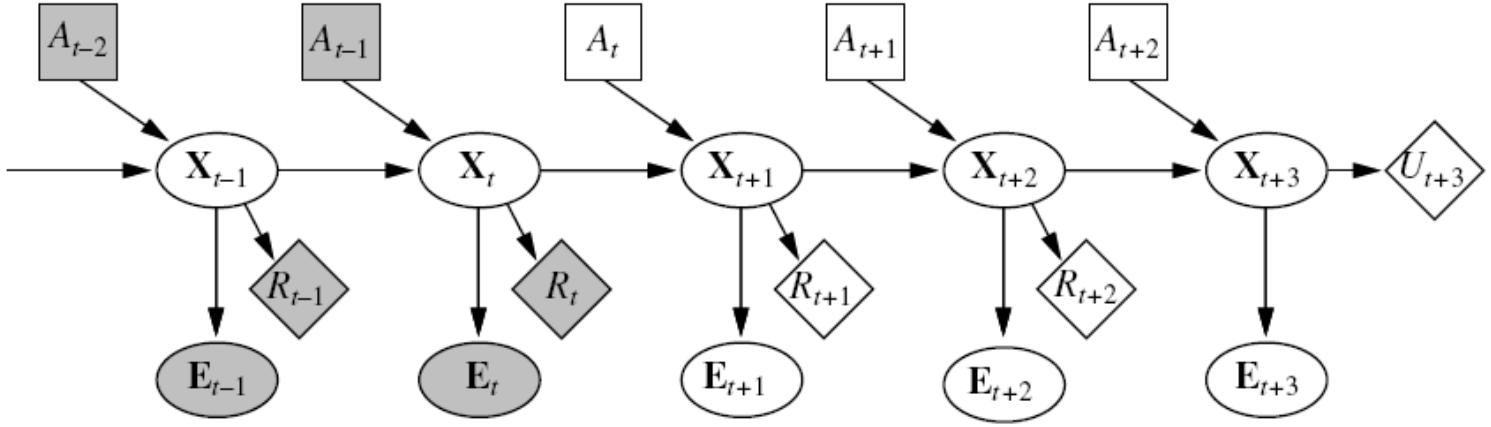
An (incomplete) genealogy of graphical models



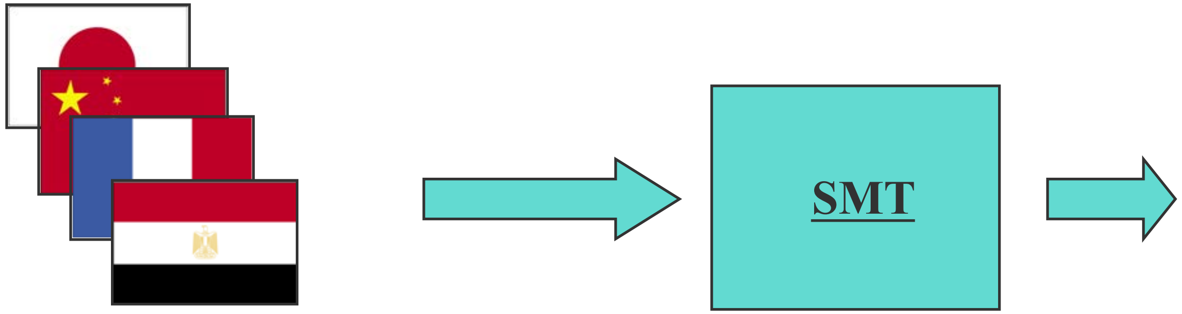
(Picture by Zoubin Ghahramani and Sam Roweis)

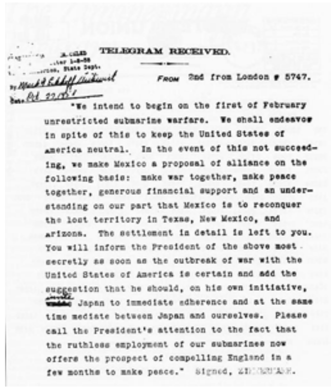
Fancier GMs: reinforcement learning

- Partially observed Markov decision processes (POMDP)



Fancier GMs: machine translation

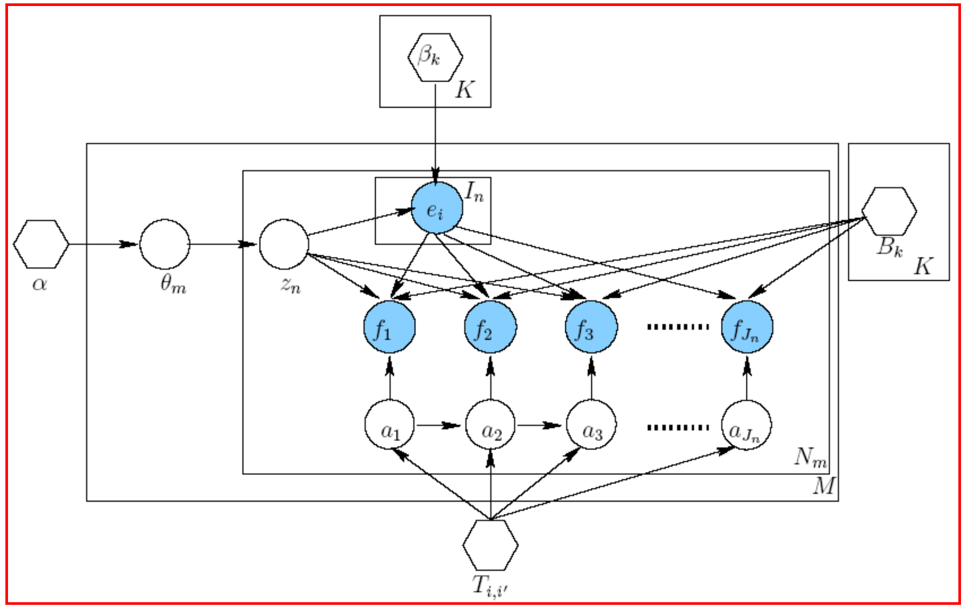




TELEGRAM RECEIVED.

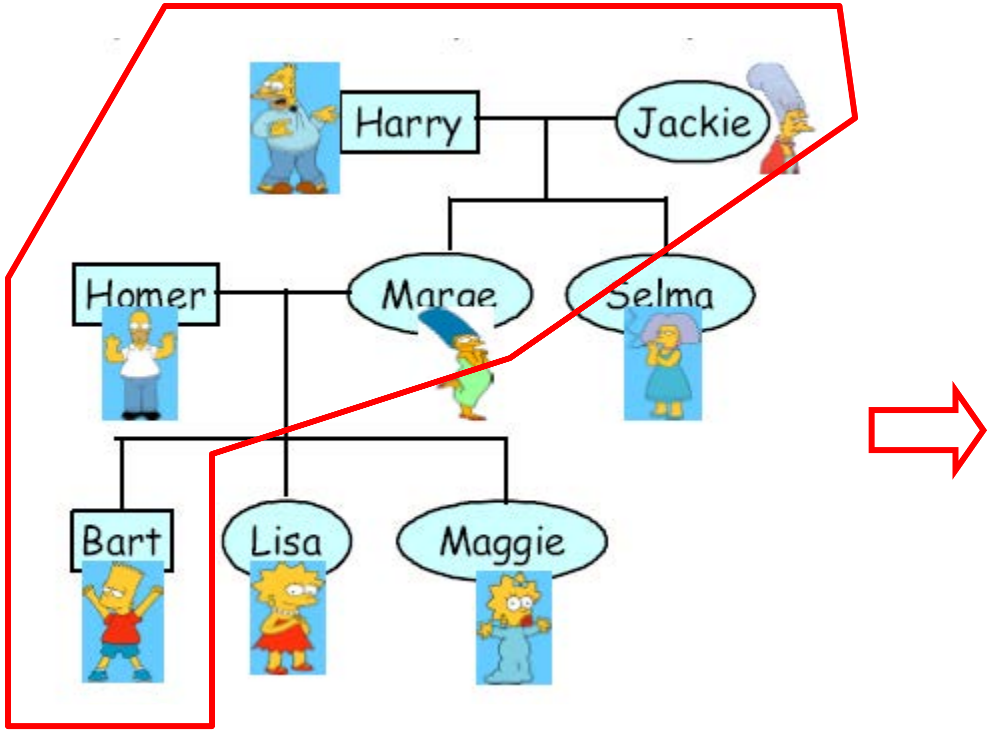
FROM 2nd from London # 5747.

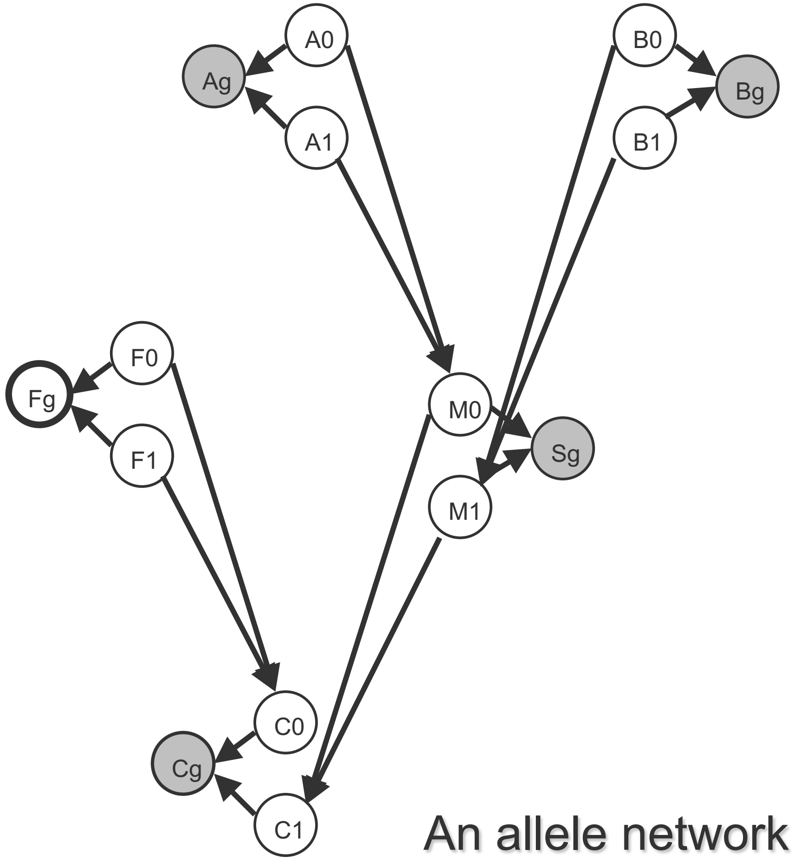
"We intend to begin on the first of February unrestricted submarine warfare. We shall endeavor in spite of this to keep the United States of America neutral. In the event of this not succeeding, we make Mexico a proposal of alliance on the following basis: make war together, make peace together, generous financial support and an understanding on our part that Mexico is to reconquer the lost territory in Texas, New Mexico, and Arizona. The settlement in detail is left to you. You will inform the President of the above most secretly as soon as the outbreak of war with the United States of America is certain and add the suggestion that he should, on his own initiative, invite Japan to immediate adherence and at the same time mediate between Japan and ourselves. Please call the President's attention to the fact that the ruthless employment of our submarines now offers the prospect of compelling England in a few months to make peace." Signed, [illegible].



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

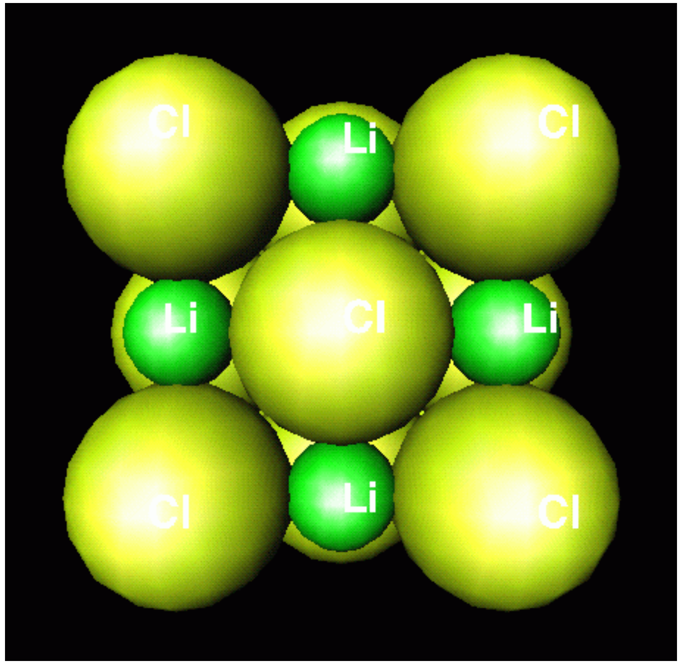
Fancier GMs: genetic pedigree

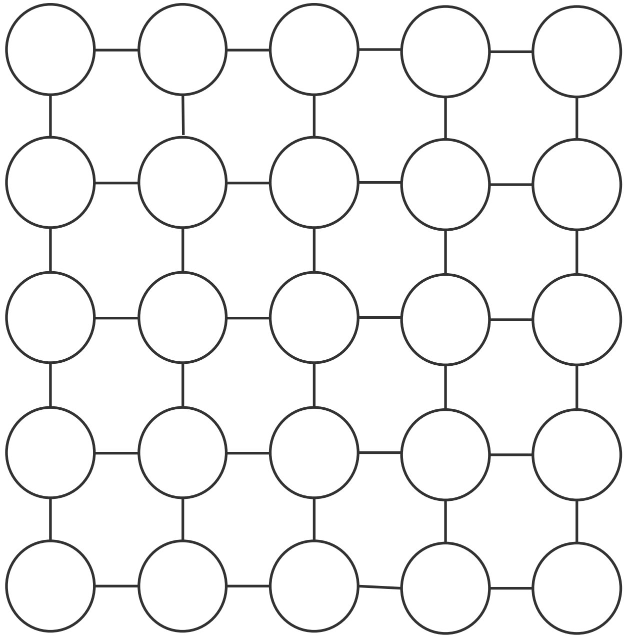




An allele network

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.