10-701 Recitation 10: Graphical Models

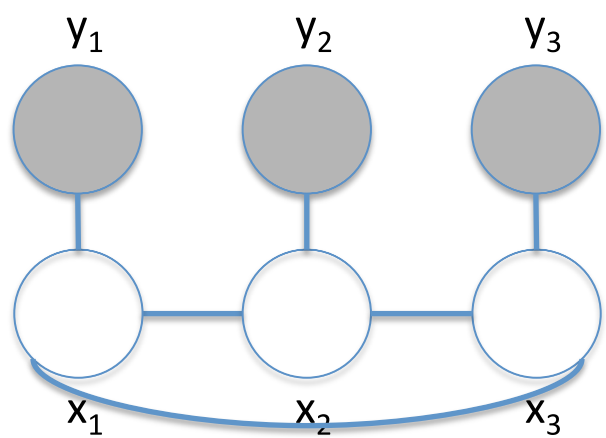
Dougal Sutherland

Graphical models

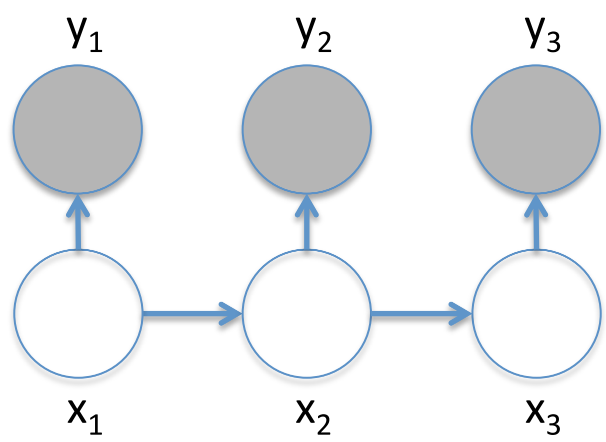
- Way to write a probability distribution
  - Exploit conditional independences to avoid writing out exponentially big tables

- Advantages over traditional ML methods:
  - Reason about the interactions of lots of complex things with one declarative model
  - Modular
  - Incorporate prior information really easily

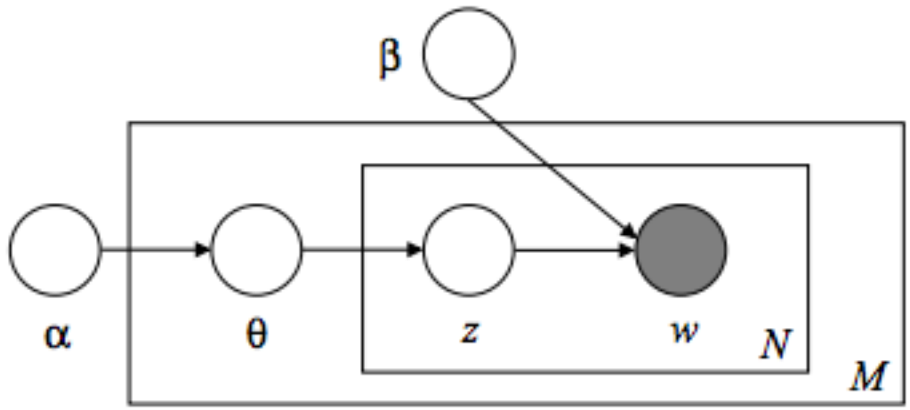
Undirected graphical models



Directed graphical models

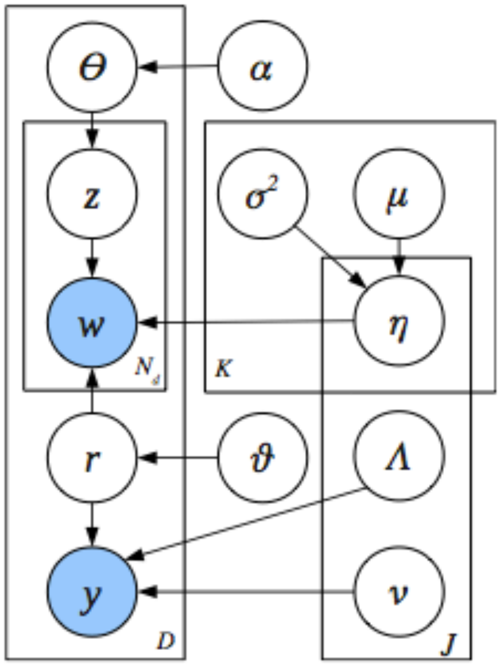


Latent Dirichlet Allocation



Blei, Ng, and Jordan, JMLR 2003

More complicated topic models



$\boldsymbol{\mu}_k$	log of base topic k's distribution over word types
σ_k^2	variance parameter for regional variants of topic k
$\boldsymbol{\eta}_{jk}$	region j's variant of base topic $\boldsymbol{\mu}_k$
$\boldsymbol{\theta}_d$	author d's topic proportions
r_d	author d's latent region
$\boldsymbol{y}_d$	author d's observed GPS location
$\boldsymbol{\nu}_j$	region j's spatial center
Λ_j	region j's spatial precision
z_n	token n's topic assignment
w_n	token n's observed word type
$\boldsymbol{\alpha}$	global prior over author-topic proportions
$\boldsymbol{\vartheta}$	global prior over region classes

Figure 1: Plate diagram for the geographic topic model, with a table of all random variables. Priors (besides α) are omitted for clarity, and the document indices on z and w are implicit.

Eisenstein, O'Connor, Smith, and Xing, EMNLP 2010

More complicated topic models

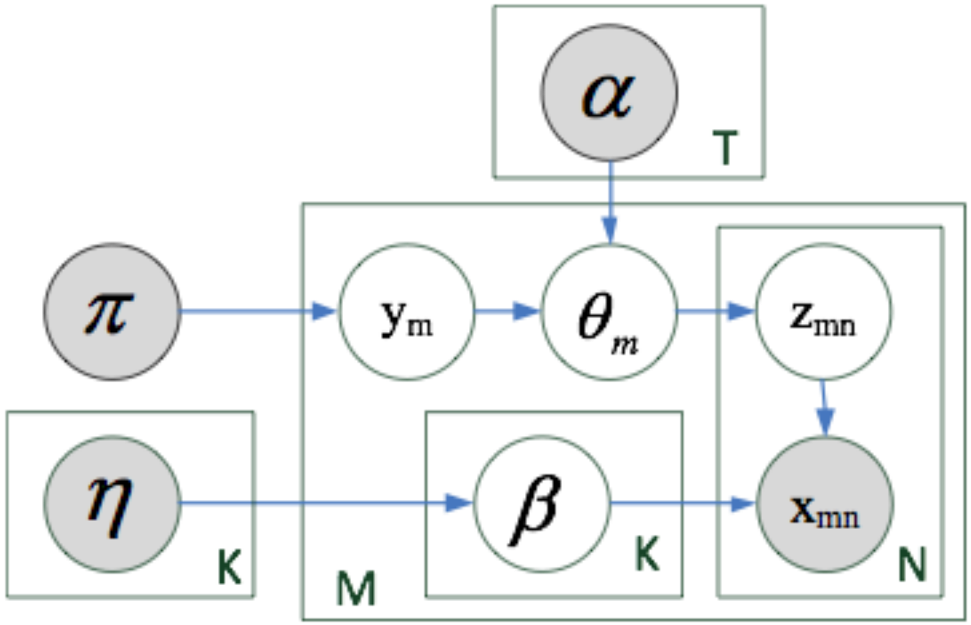
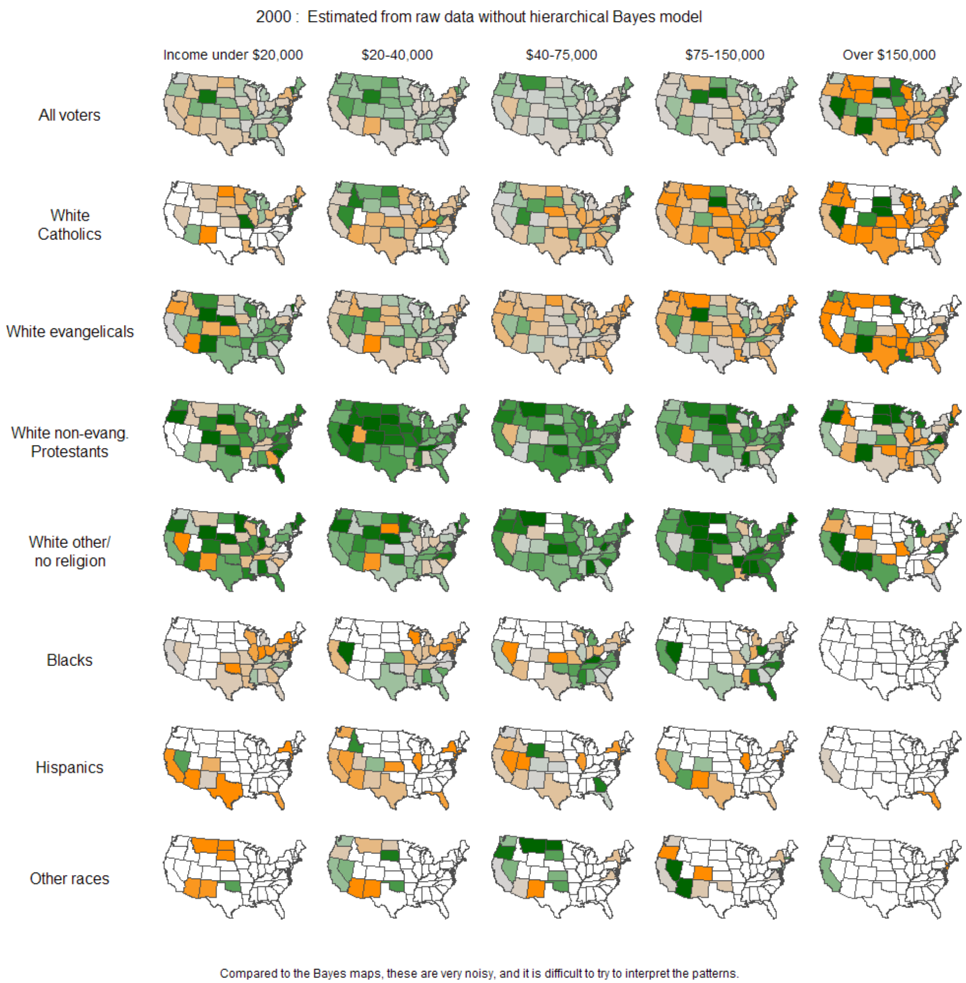
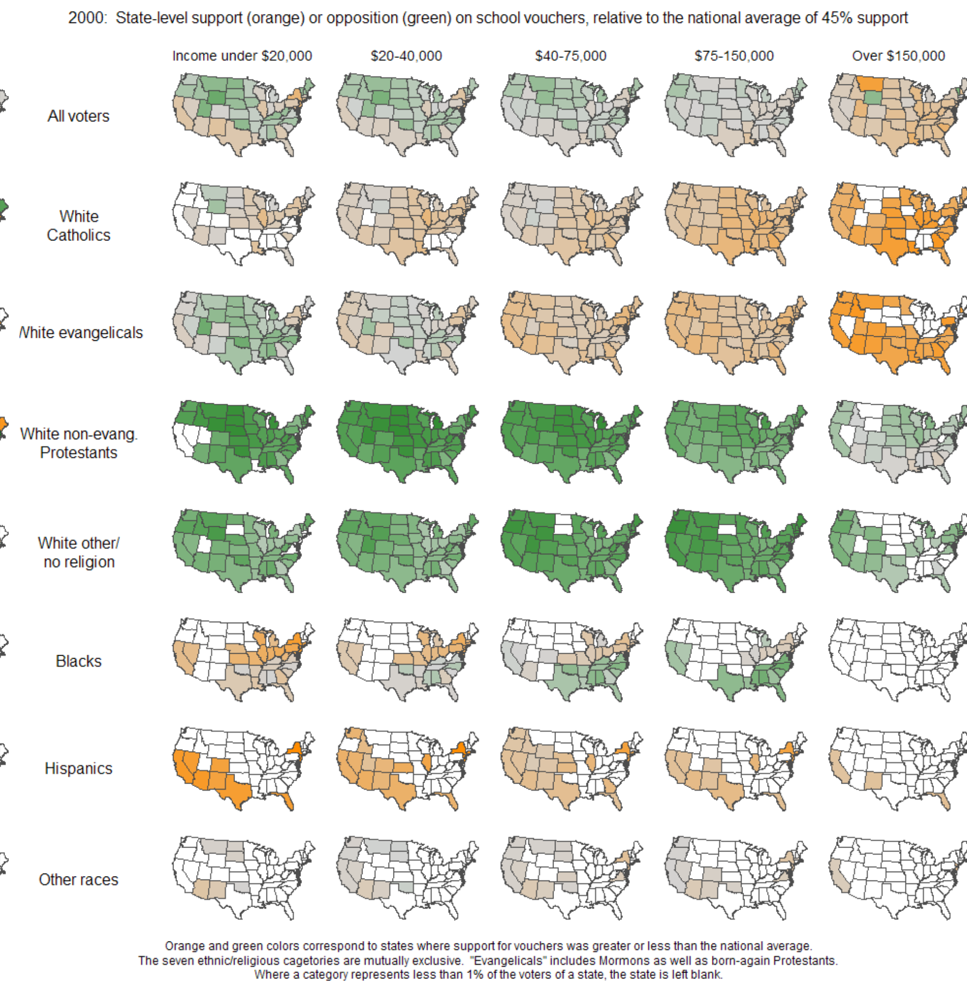


Figure 1: The Flexible Genre Model (FGM).

Xiong, Póczos, and Schneider, NIPS 2011

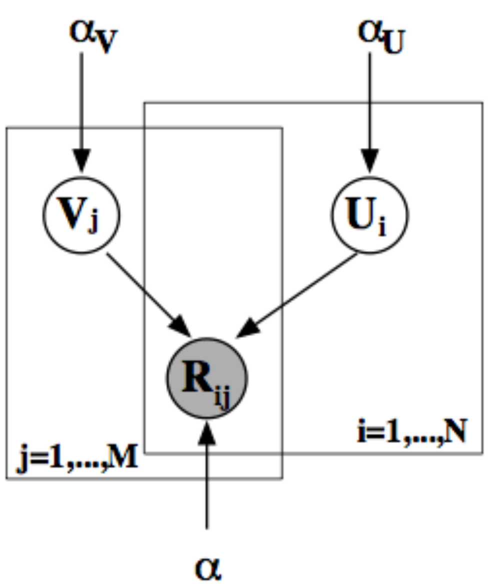
Hierarchical statistical models

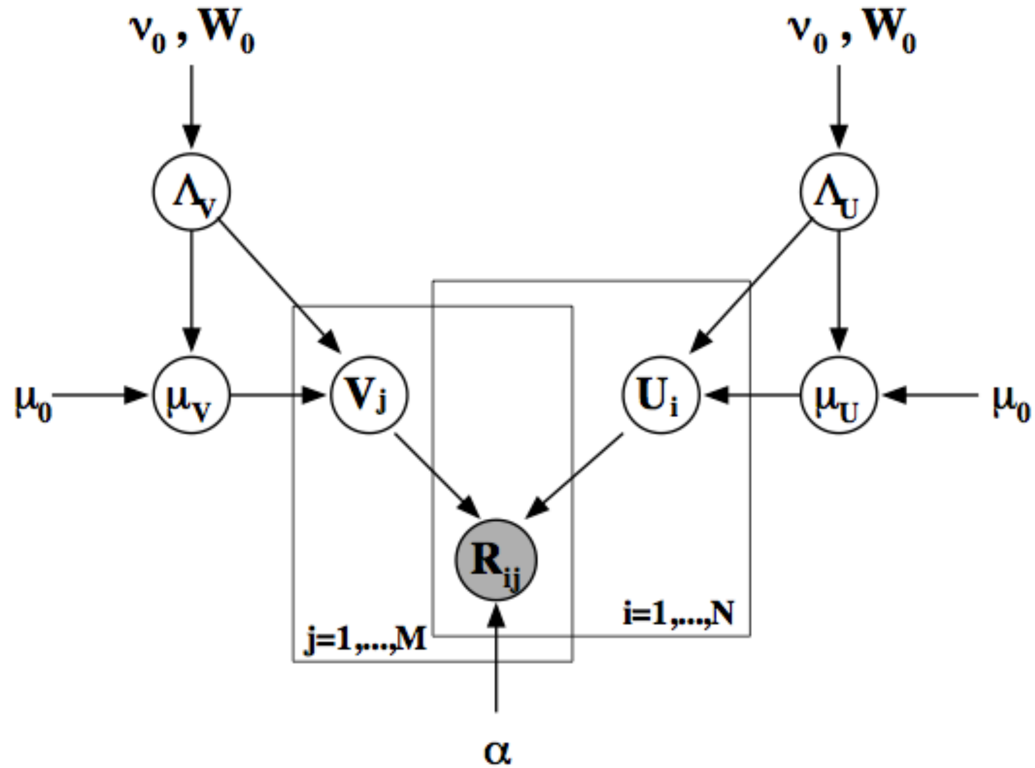




Gelman, http://andrewgelman.com/2009/07/15/hard_sell_for_b/

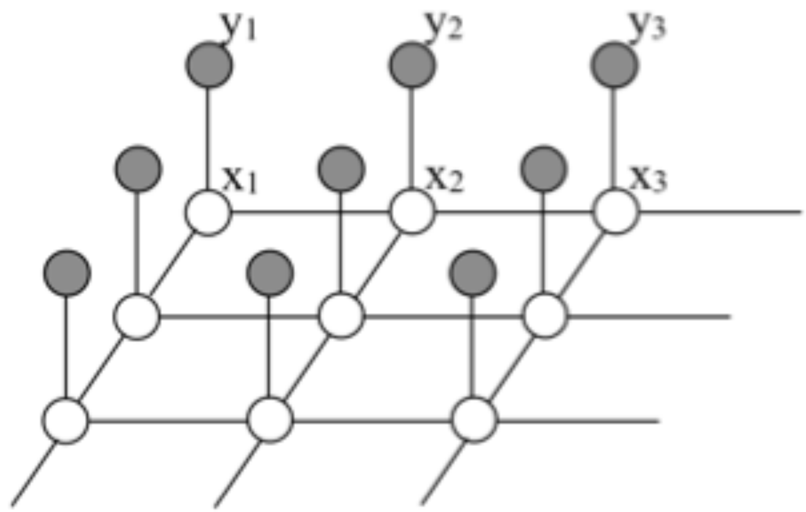
Probabilistic Matrix Factorization

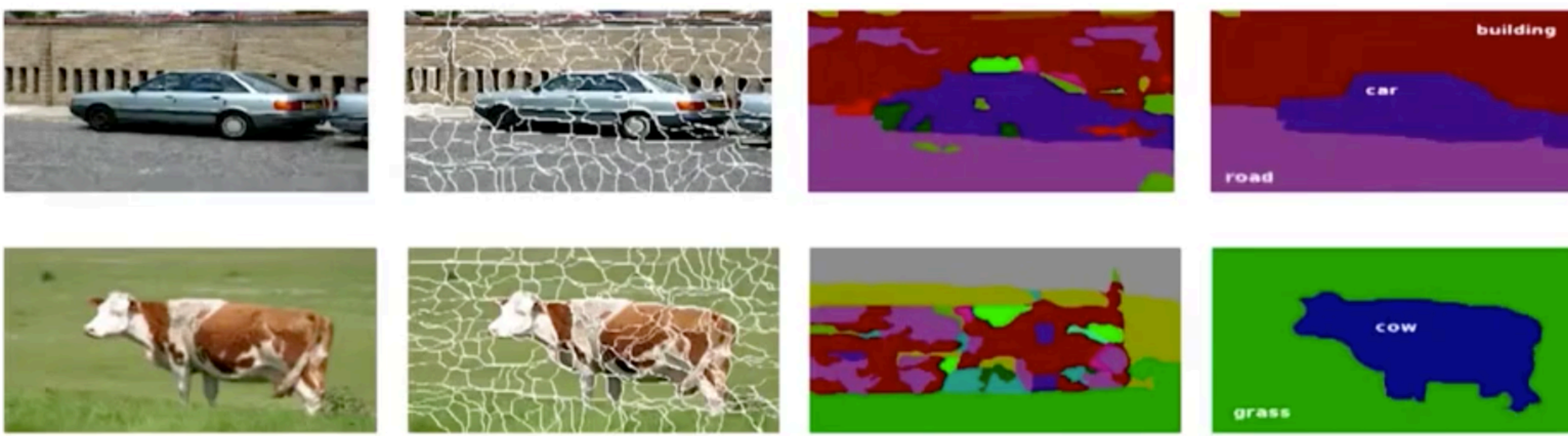




Salakhutdinov and Mnih, NIPS 2007 / ICML 2008

Image segmentation





Inference

- Asking probability questions about the model

- Exact methods:
  - Message passing (with junction tree)
  - Combinatorial graph-based methods

- Approximate methods:
  - Loopy belief propagation
  - Variational approximations
  - Sampling (via MCMC)

Inference: loopy belief propagation

- Just apply belief propagation to general graph
  - Without doing junction tree
- Approximate
  - Doesn’t always converge
  - Sometimes converges to wrong answers
  - Often works pretty well and fast

Inference: variational approximations

- Choose a family where inference is easy
  - Parameters are jointly normal
  - Fully-factorized: all params are independent
- Find the closest member of that family:
  - Call real model *p*; we're finding an approximation *q*
  - Minimize KL($q \| p$)

Inference: sampling via MCMC

- Get approximate samples from the model
  - Probabilistic questions => frequency questions
- Importance sampling:
  - Sample from marginals on hidden variables
  - Weight samples by their likelihood
- Metropolis-Hastings:
  - From the current point, propose a new point
  - Accept with probability based on relative likelihoods
- Gibbs:
  - Change one variable based on its conditional given all the others

Learning

- How do we choose the model parameters?
- Fully observed:
  - Directed models:
    - For each node *X*, fit the distribution conditional on its parents from the given samples
    - MAP is max $P(X \mid par(X))\, P(X)$
  - Undirected models:
    - Maximum likelihood over full model is convex
    - Requires inference step at each iteration

Learning with hidden variables

- Expectation maximization:
  - E step: impute hidden variables with current model
    - Inference problem
  - M step: update the fully-observed model
    - Learning problem with techniques we just talked about

- Converges to a local optimum

Structure learning

- What if we don't want to specify the relationships ourselves?
- Could maximize likelihood over model structure
  - Problem: there are a lot of models
    - With 10 variables, ~ a quintillion possible DAGs
    - Need to do local search over structures
  - Problem: guaranteed to overfit
    - Making model more complex can only increase likelihood
    - Need to penalize model complexity