



Scalable Machine Learning

9. Graphical Models

Alex Smola

Yahoo! Research and ANU

<http://alex.smola.org/teaching/berkeley2012>

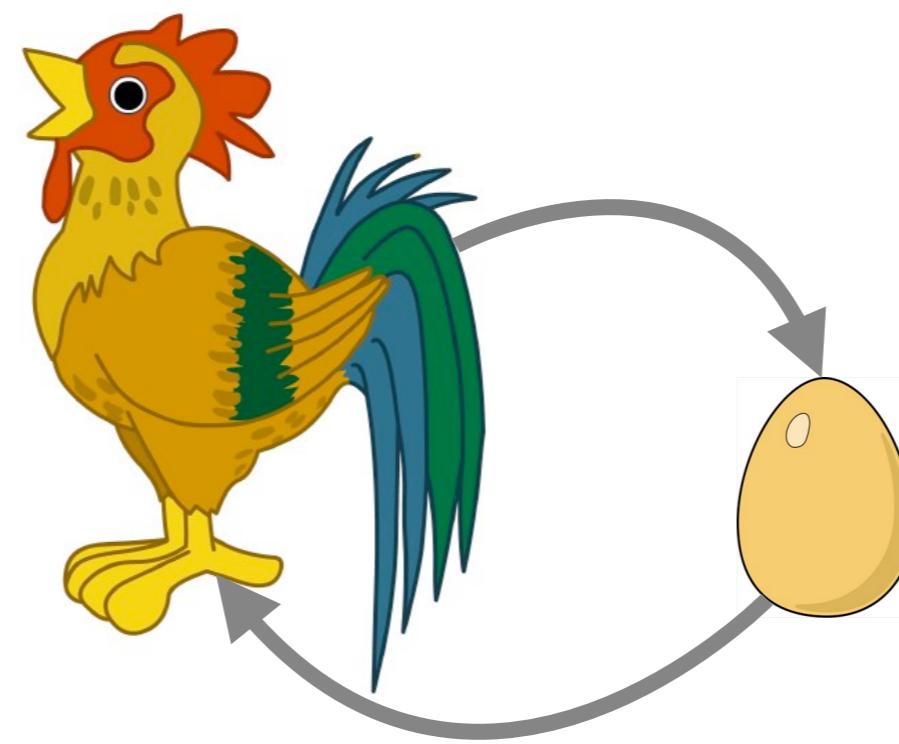
Stat 260 SP 12

Significant content courtesy of Yehuda Koren

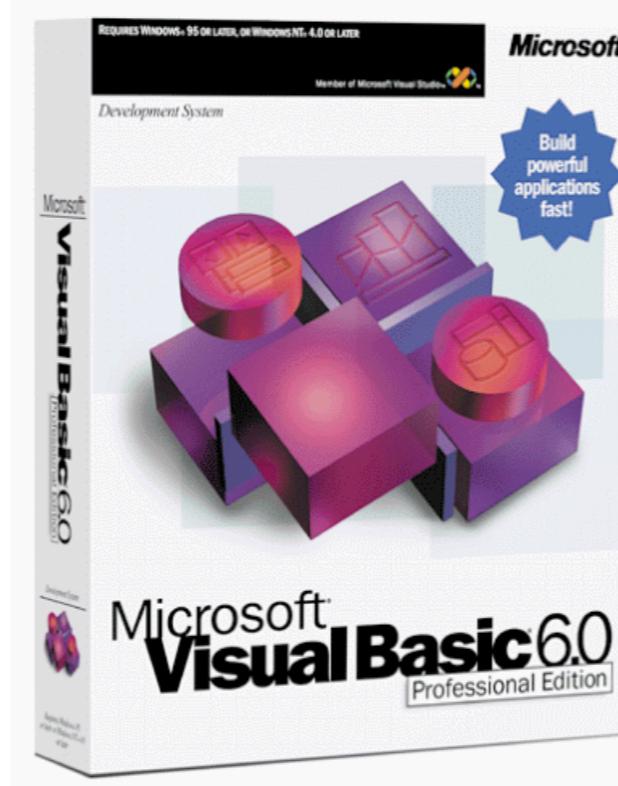
Outline

- Directed Graphical Models
 - Dependence
 - Inference for fully observed models
 - Incomplete information / variational and sampling inference
- Undirected Graphical Models
 - Hammersley Clifford decomposition
 - Conditional independence
 - Junction trees
- Dynamic Programming
 - Generalized Distributive Law
 - Naive Message Passing
- Inference techniques
 - Sampling (Gibbs and Monte Carlo)
 - Variational methods (EM, extensions, dual decomposition)

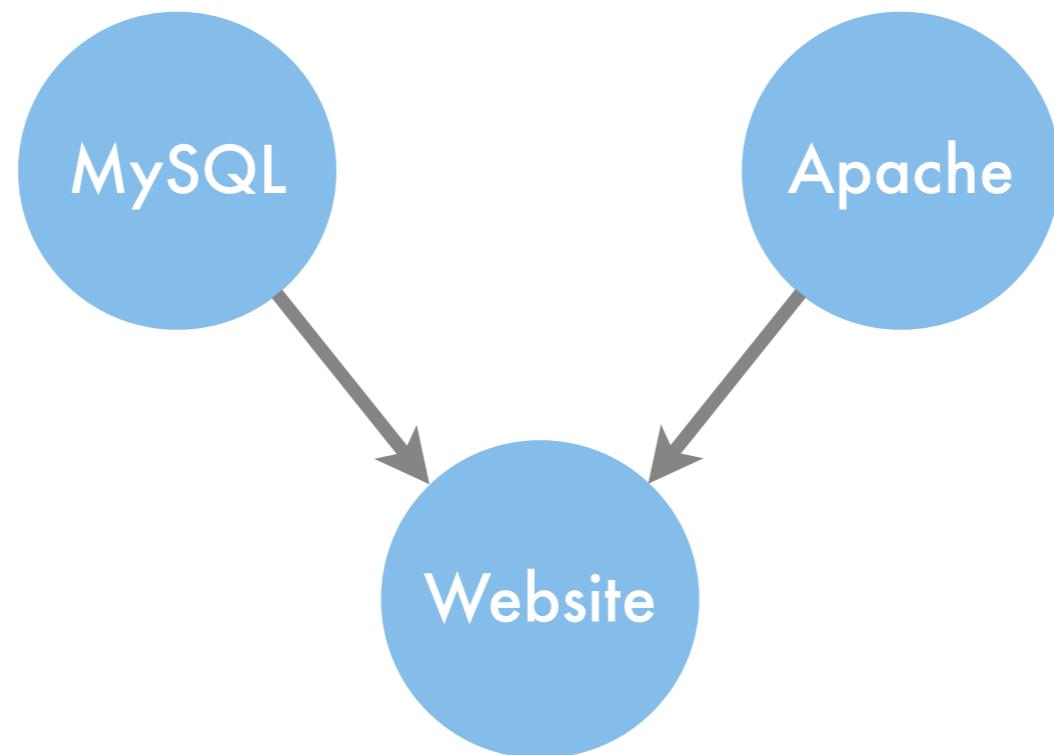
Directed Graphical Models



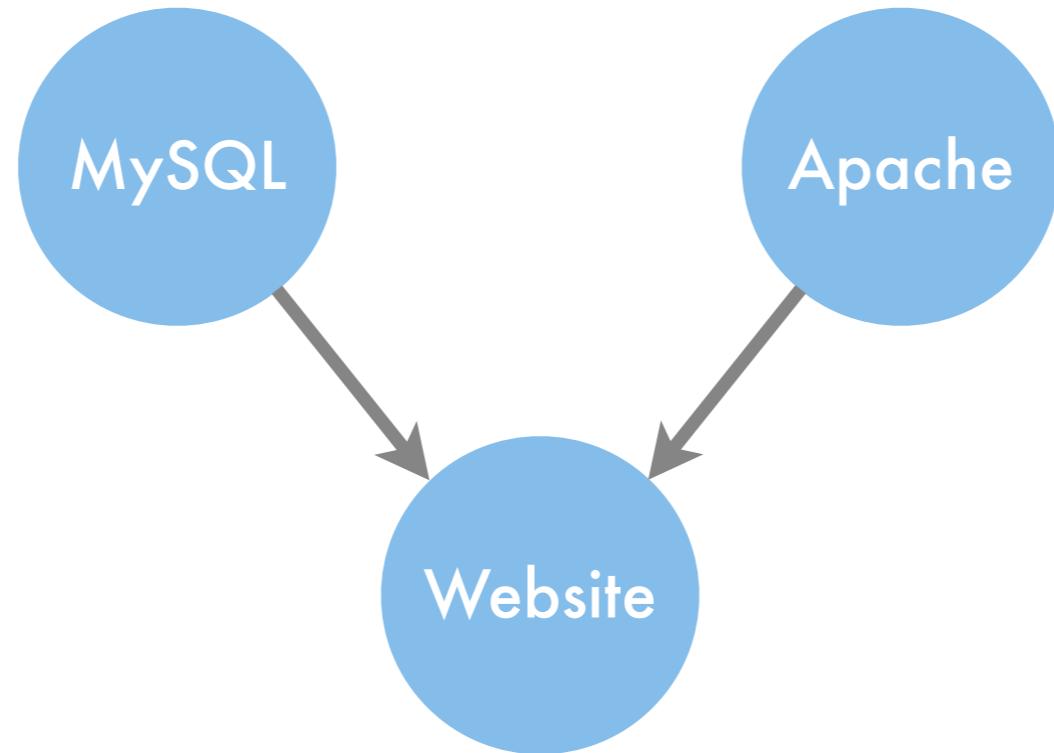
Basics



... some Web 2.0 service



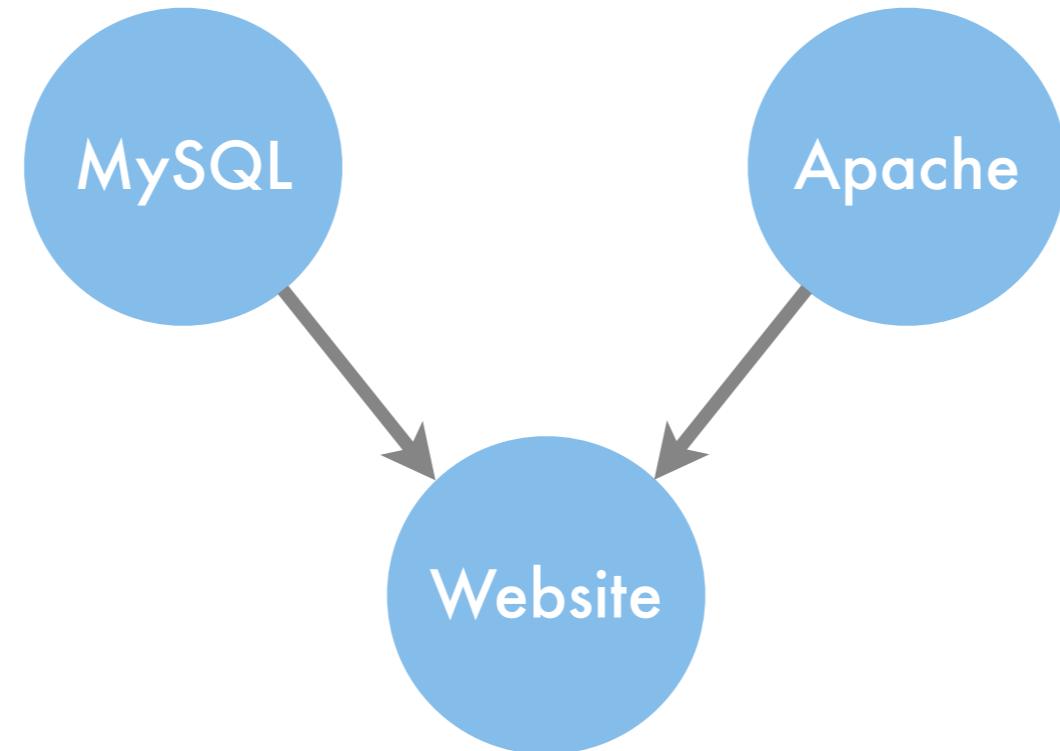
... some Web 2.0 service



- **Joint distribution (assume a and m are independent)**

$$p(m, a, w) = p(w|m, a)p(m)p(a)$$

... some Web 2.0 service



- Joint distribution (assume a and m are independent)

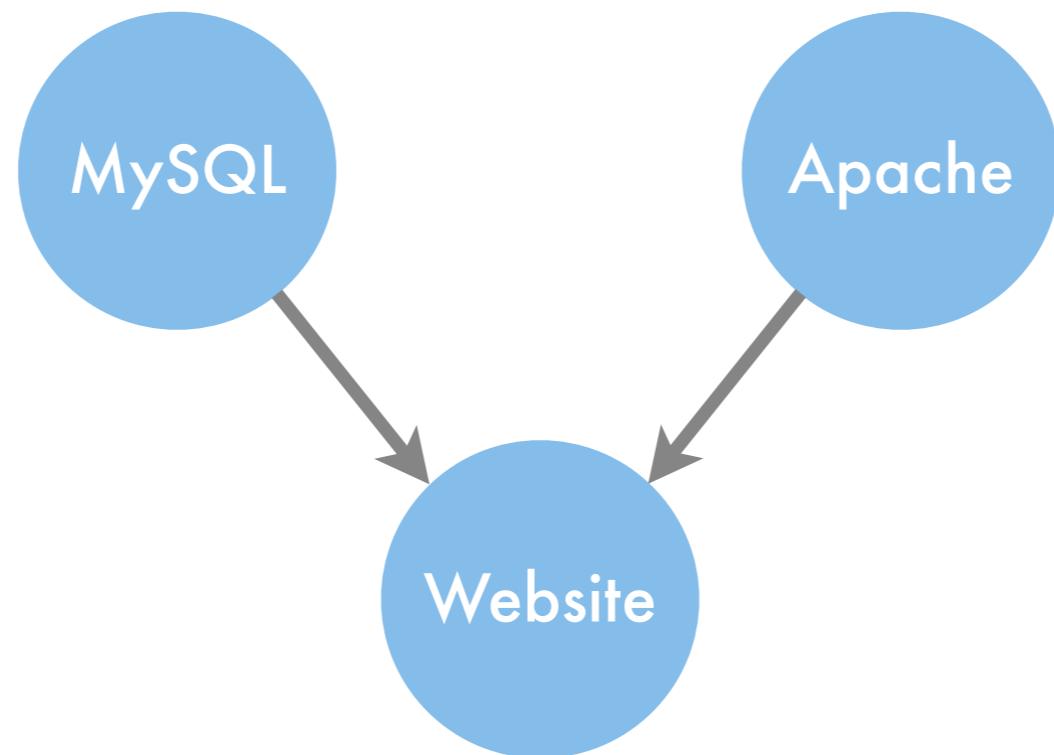
$$p(m, a, w) = p(w|m, a)p(m)p(a)$$

- Explaining away

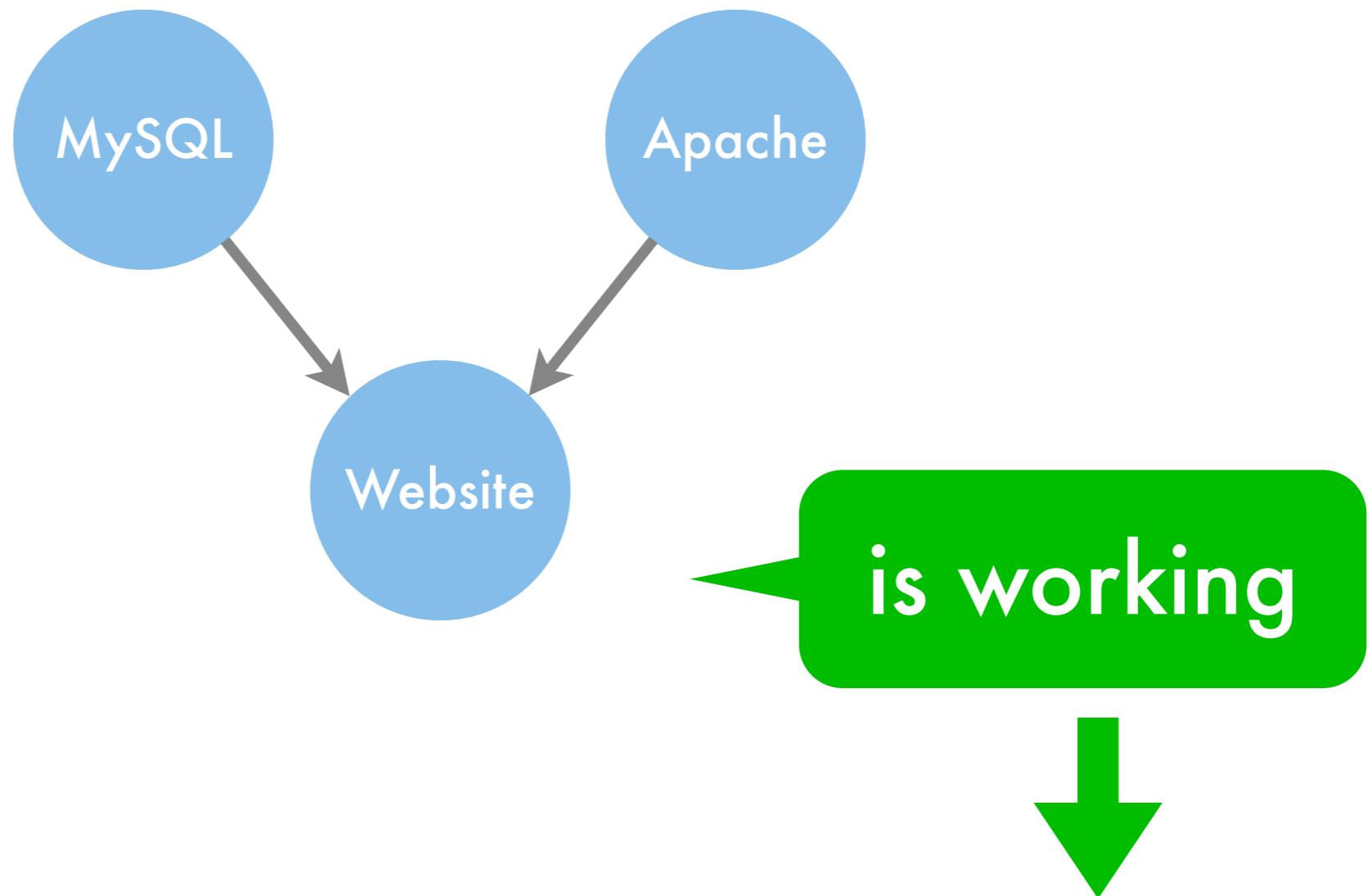
$$p(m, a|w) = \frac{p(w|m, a)p(m)p(a)}{\sum_{m', a'} p(w|m', a')p(m')p(a')}$$

a and m are dependent conditioned on w

... some Web 2.0 service

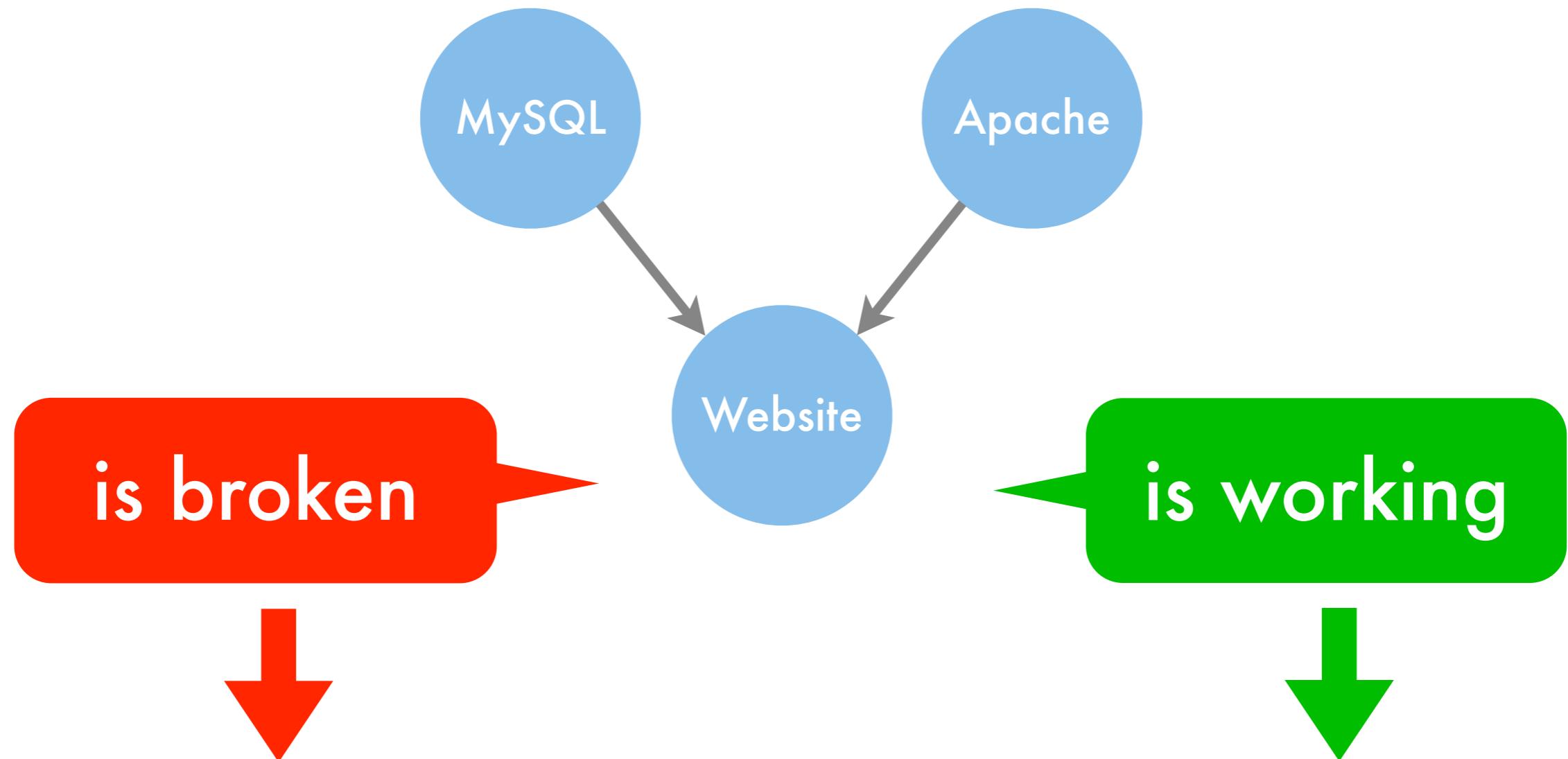


... some Web 2.0 service



MySQL is working
Apache is working

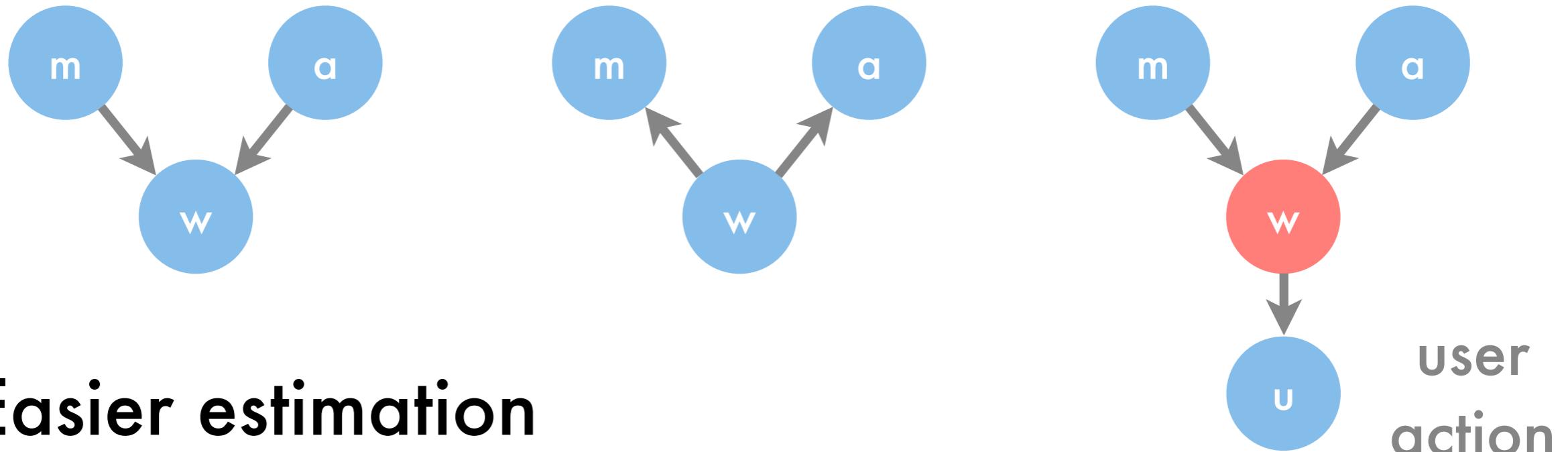
... some Web 2.0 service



At least one of the
two services is broken
(not independent)

MySQL is working
Apache is working

Directed graphical model

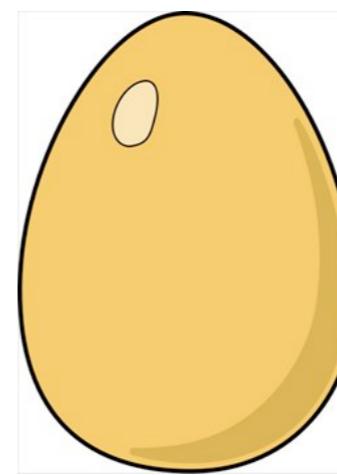


- Easier estimation
 - 15 parameters for full joint distribution
 - $1+1+4+1$ for factorizing distribution
- Causal relations
- Inference for unobserved variables

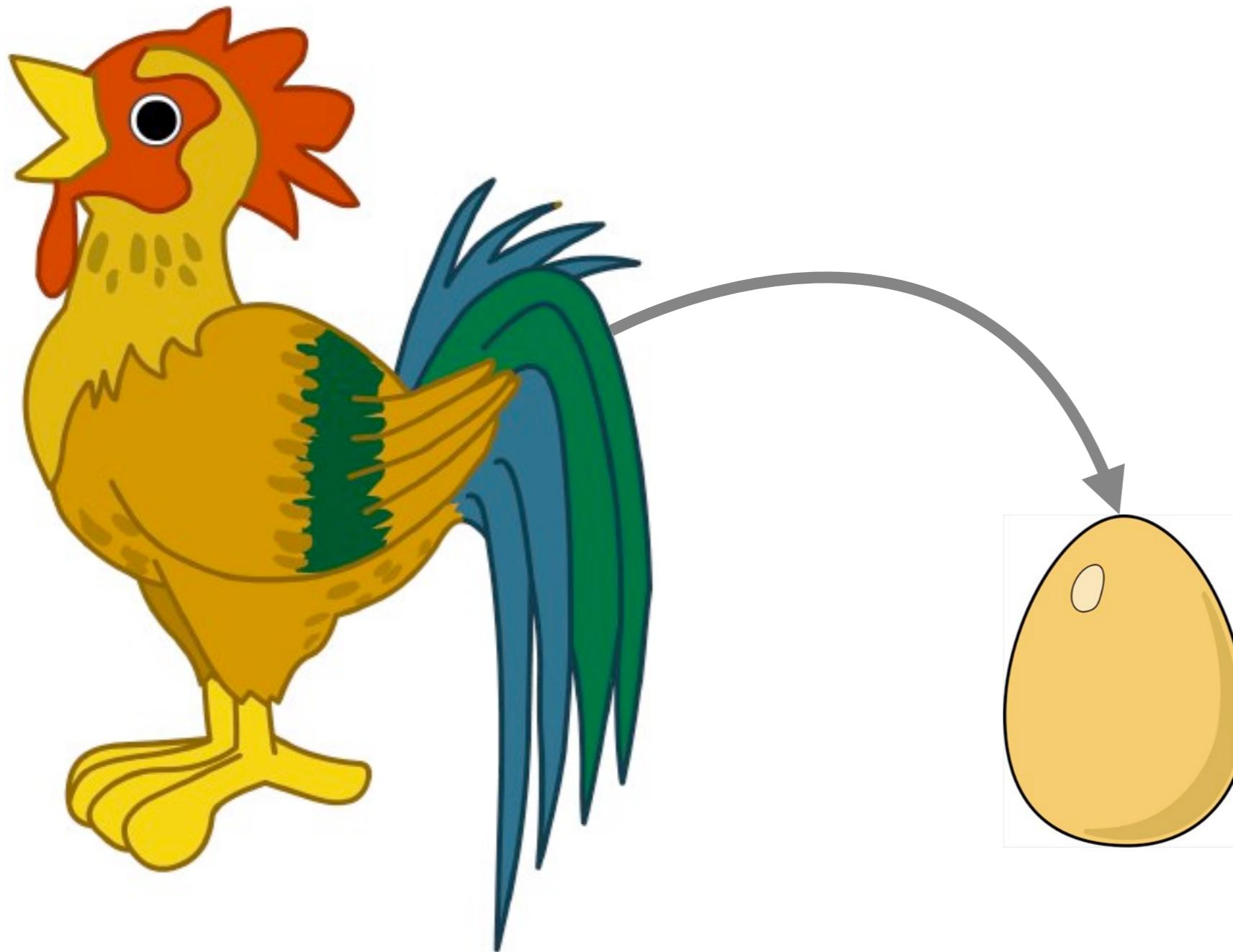
No loops allowed



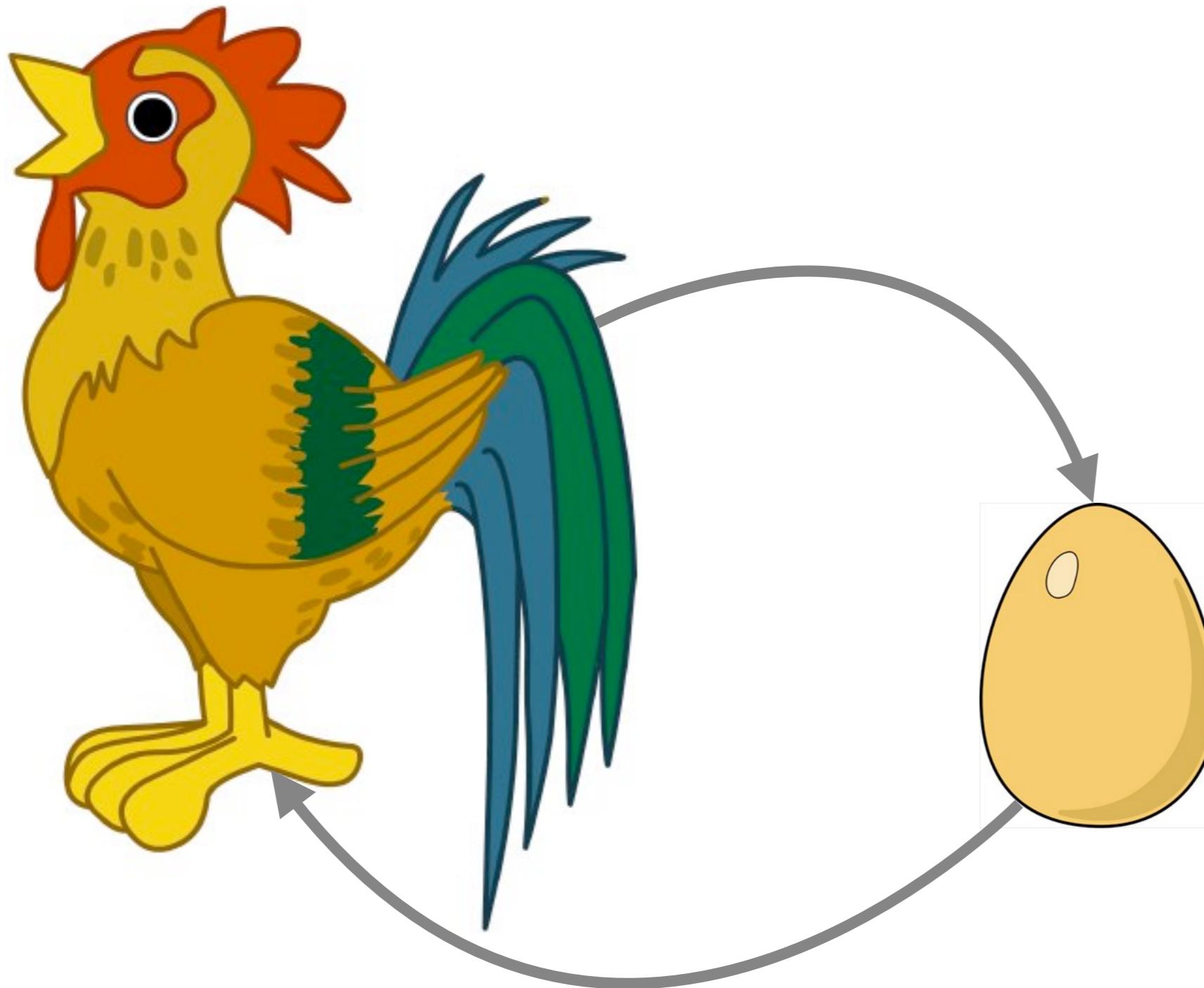
No loops allowed



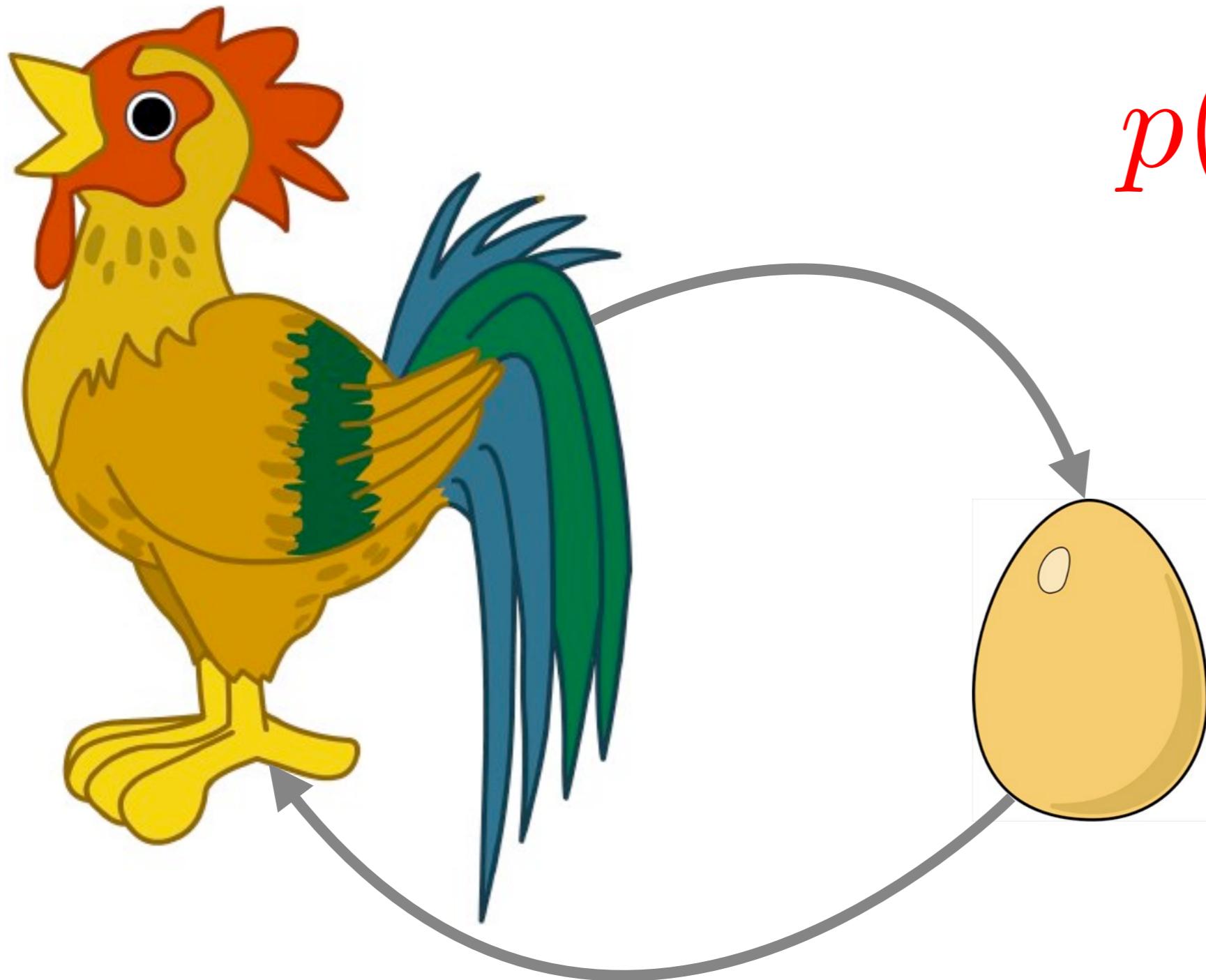
No loops allowed



No loops allowed

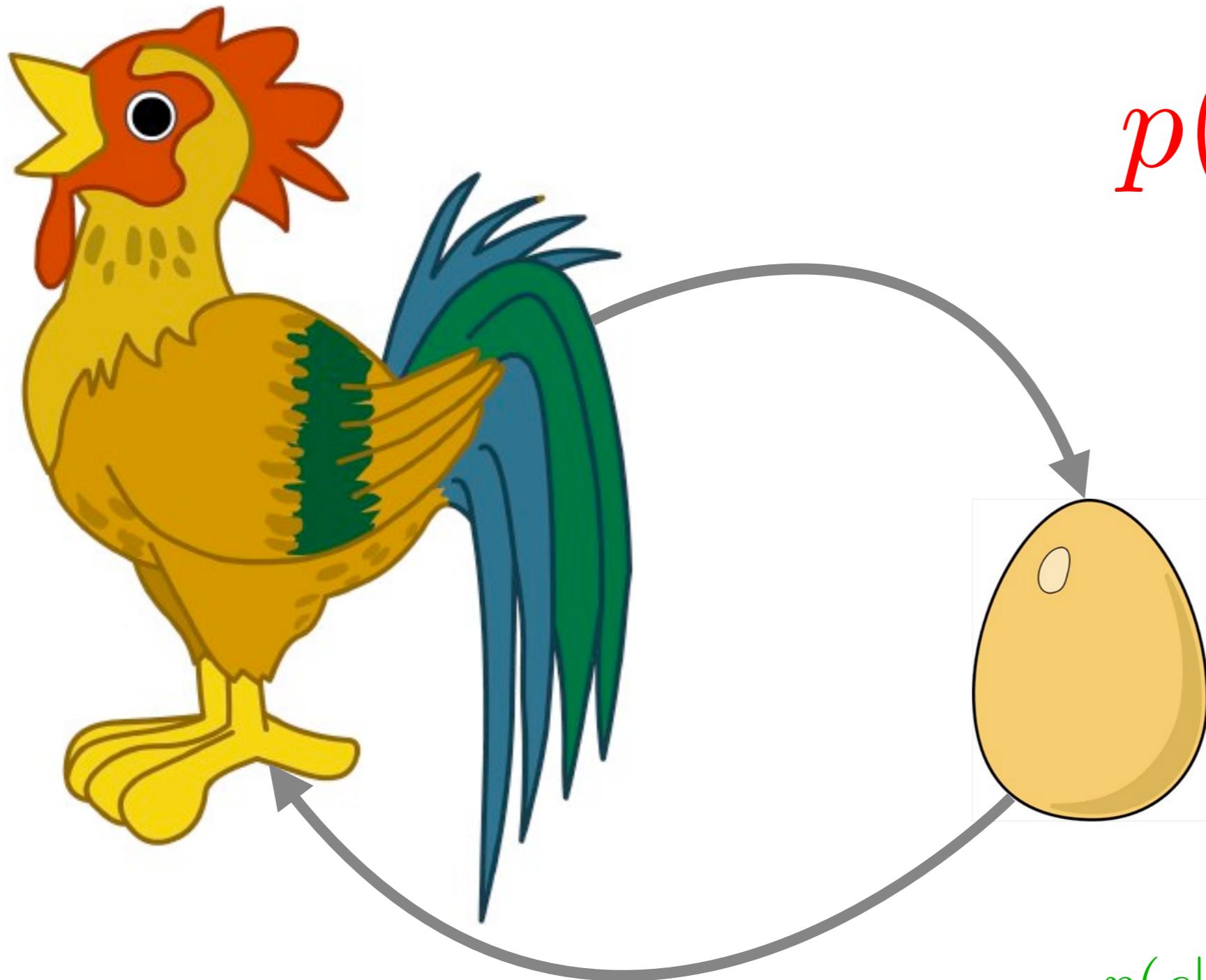


No loops allowed



$$p(c|e)p(e|c)$$

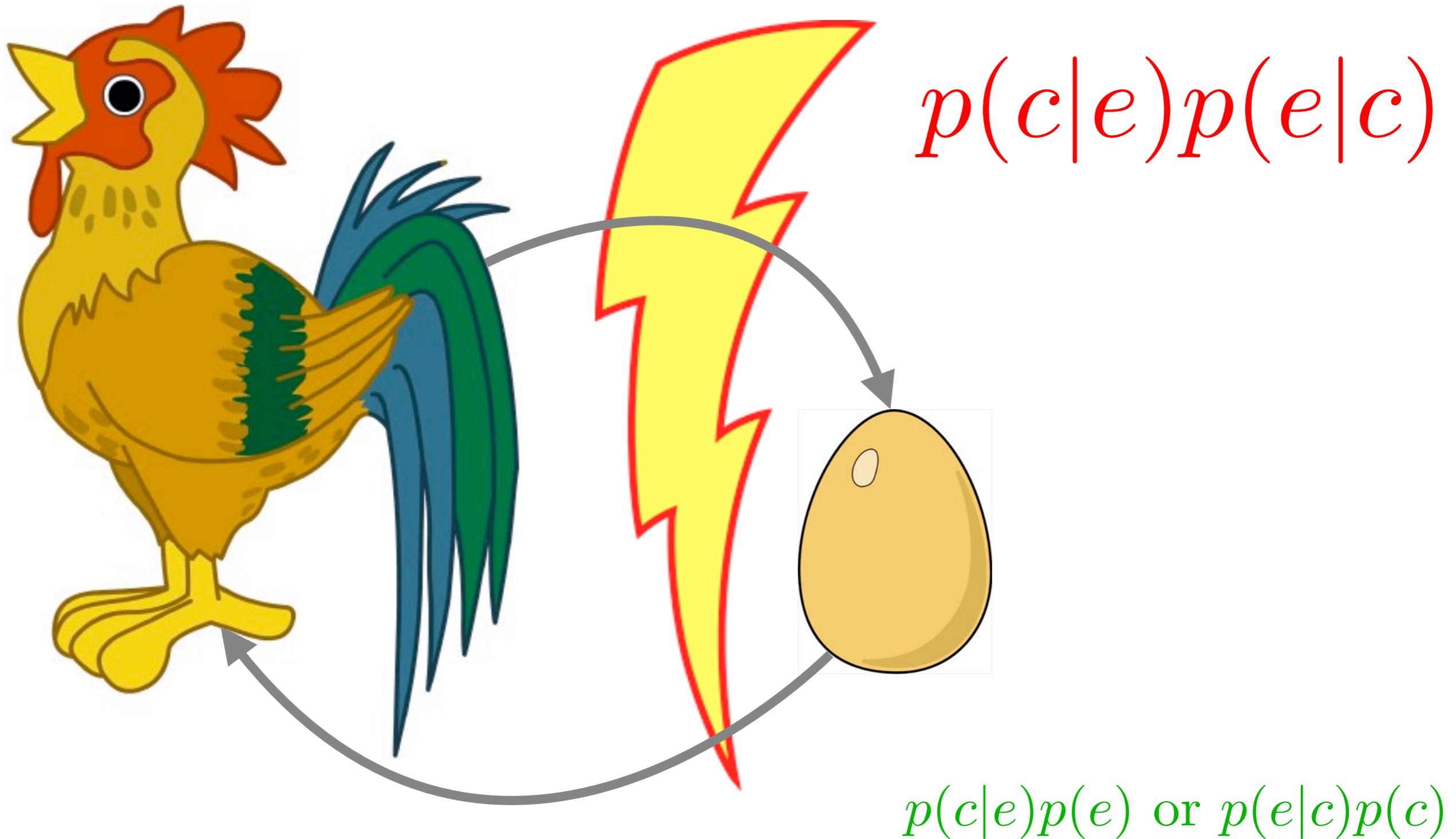
No loops allowed



$p(c|e)p(e|c)$

$p(c|e)p(e)$ or $p(e|c)p(c)$

No loops allowed



Directed Graphical Model

- Joint probability distribution

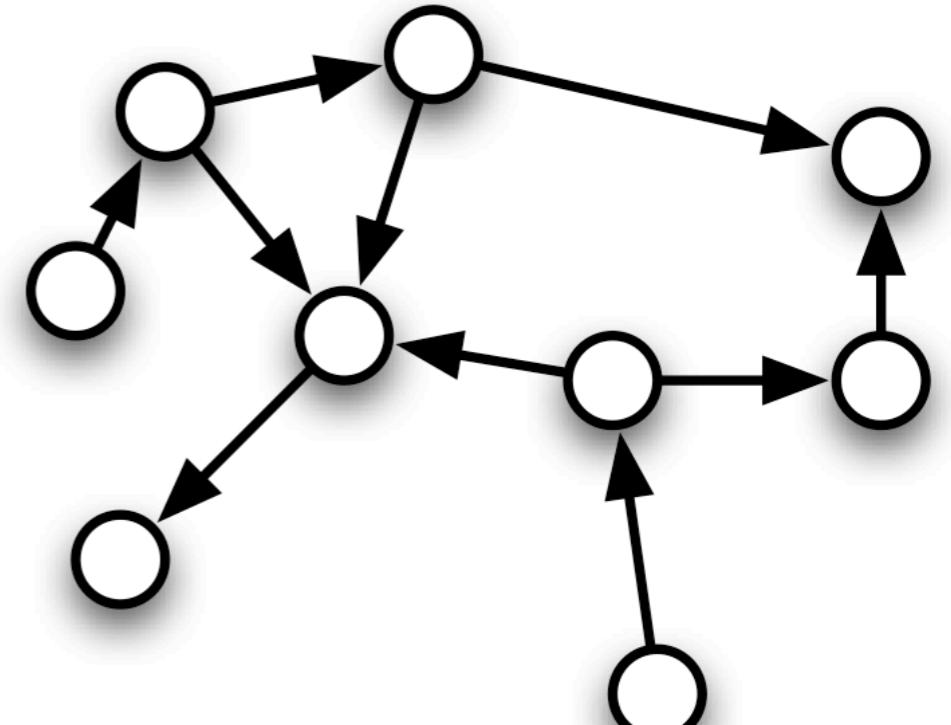
$$p(x) = \prod_i p(x_i | x_{\text{parents}(i)})$$

- Parameter estimation

- If x is fully observed the likelihood breaks up

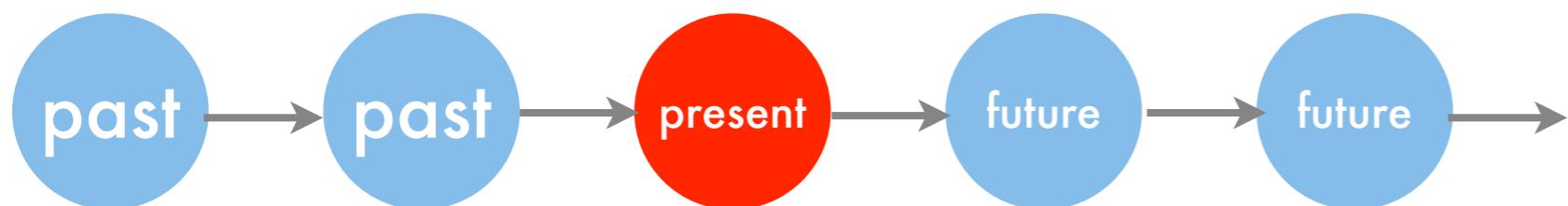
$$\log p(x|\theta) = \sum_i \log p(x_i | x_{\text{parents}(i)}, \theta)$$

- If x is partially observed things get interesting maximization, EM, variational, sampling ...



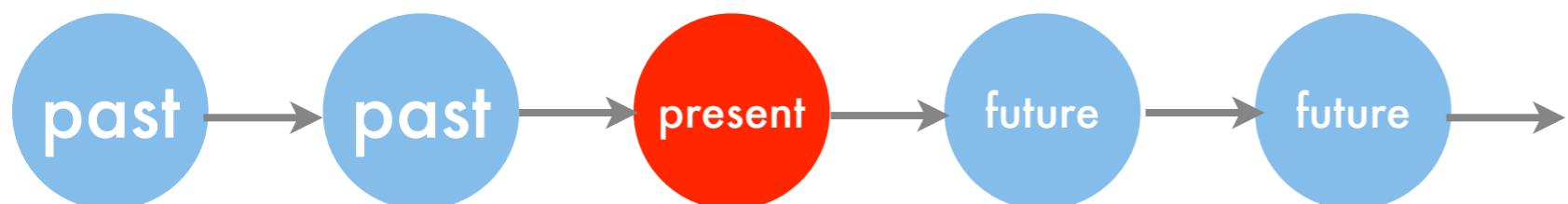
Chains

Markov Chain

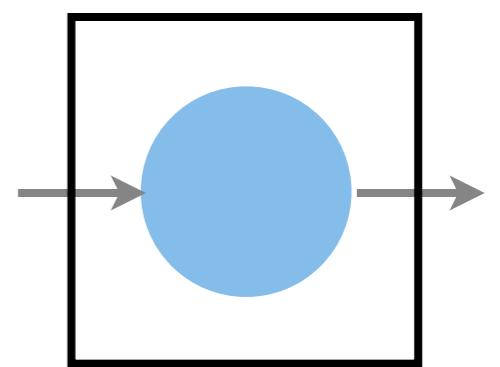


Chains

Markov Chain

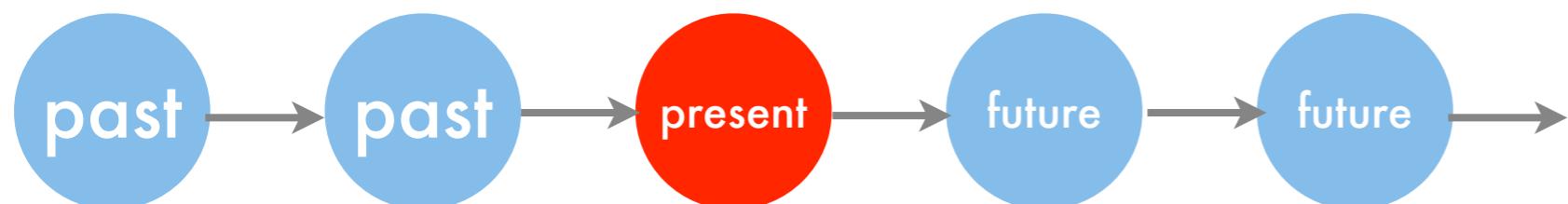


Plate

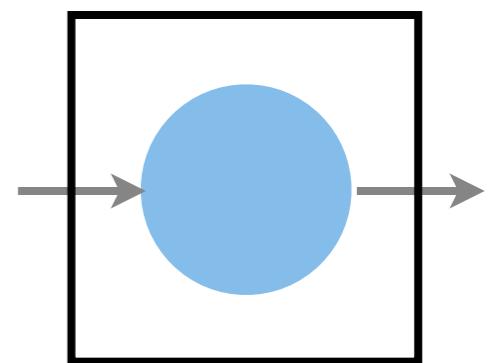


Chains

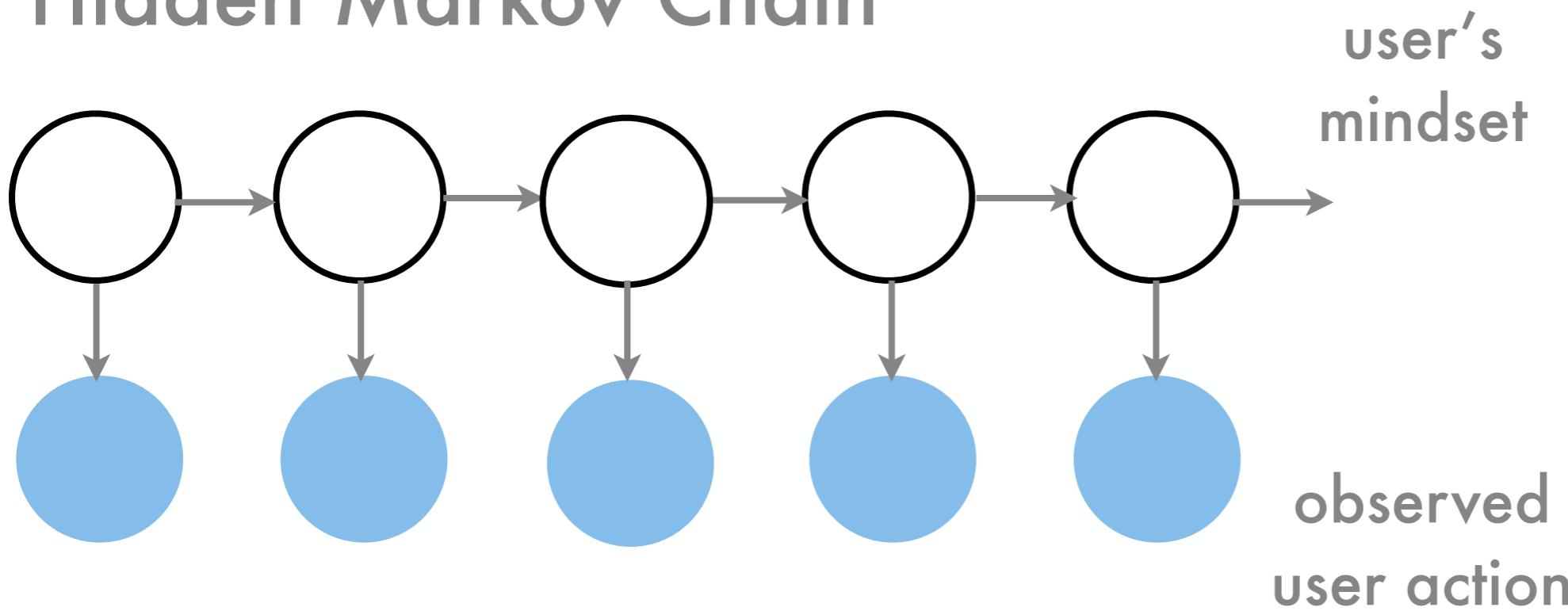
Markov Chain



Plate

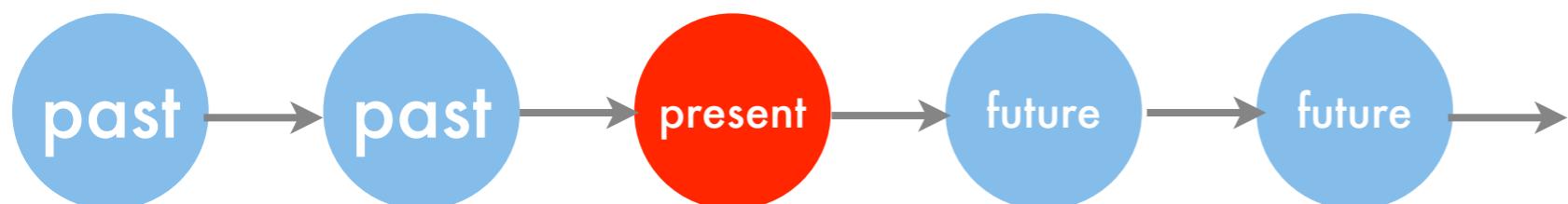


Hidden Markov Chain

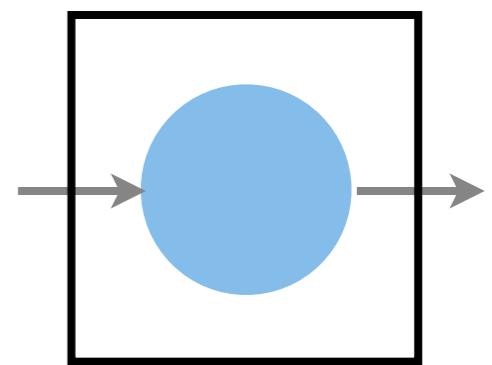


Chains

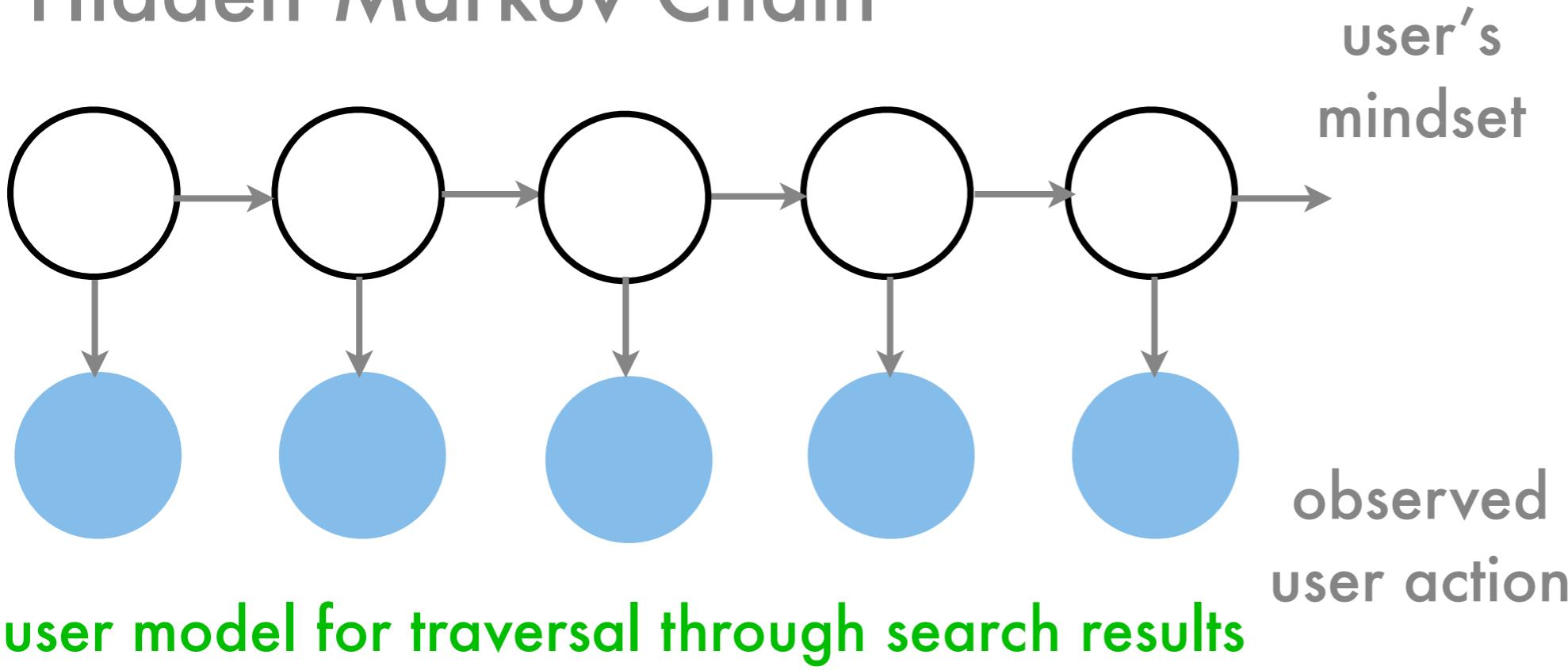
Markov Chain



Plate

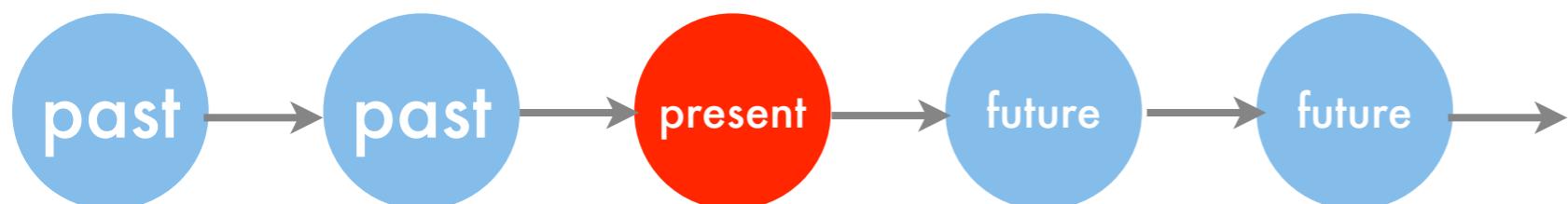


Hidden Markov Chain

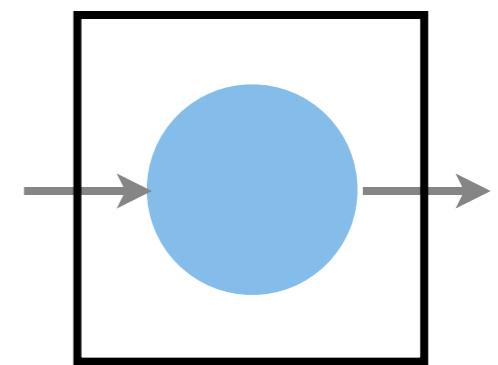


Chains

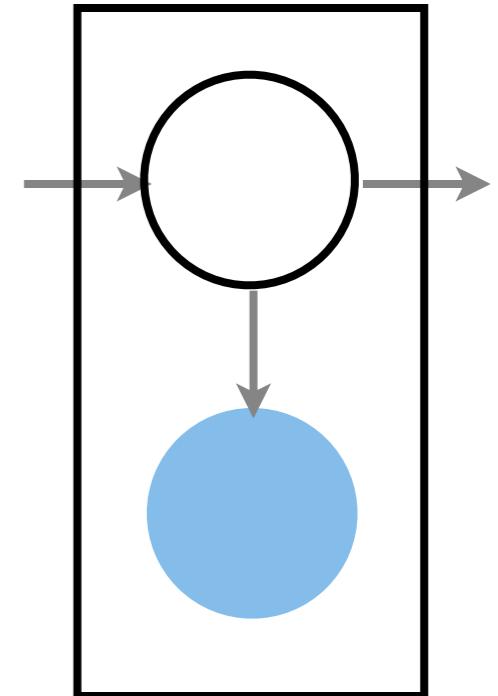
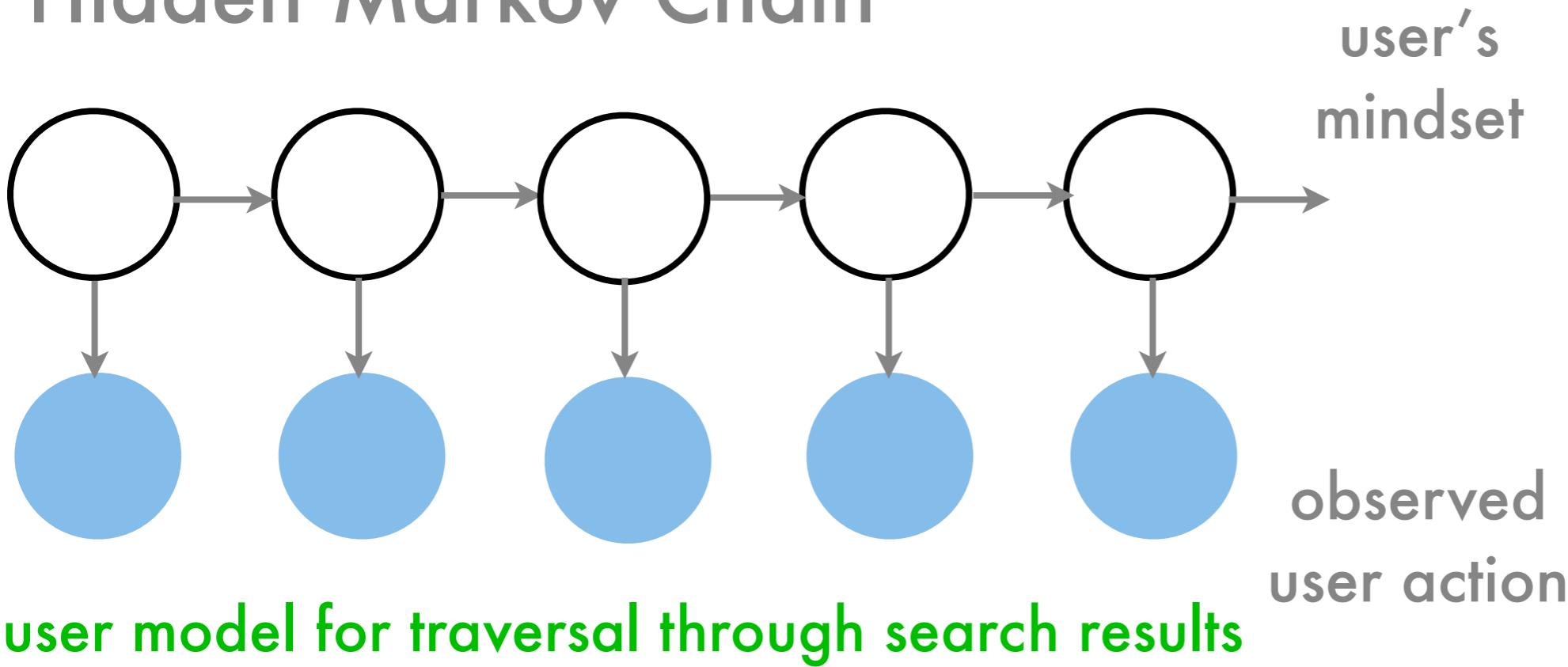
Markov Chain



Plate



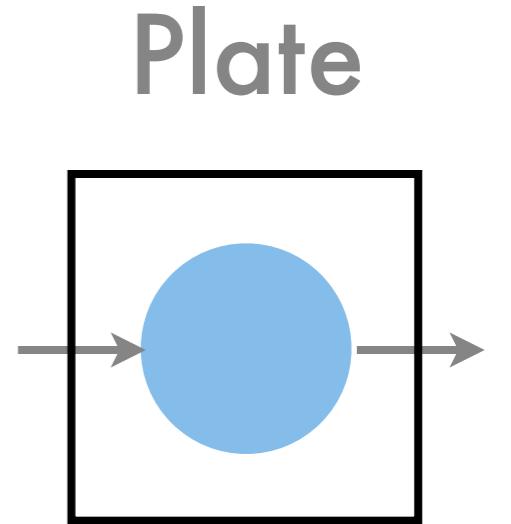
Hidden Markov Chain



Chains

Markov Chain

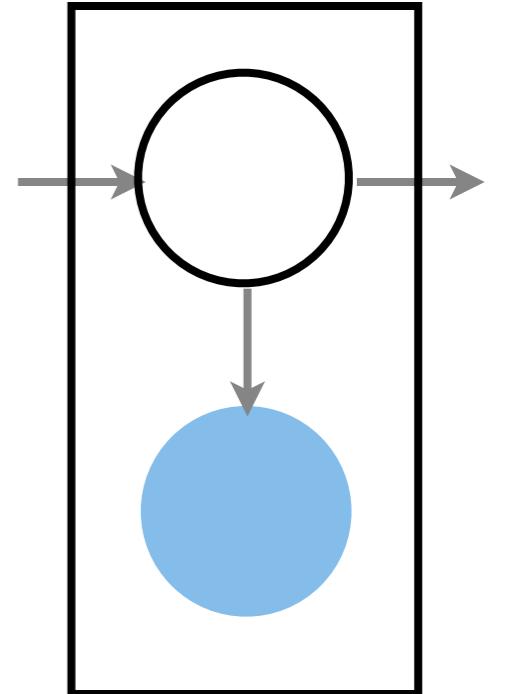
$$p(x; \theta) = p(x_0; \theta) \prod_{i=1}^{n-1} p(x_{i+1} | x_i; \theta)$$



Hidden Markov Chain

$$p(x, y; \theta) = p(x_0; \theta) \prod_{i=1}^{n-1} p(x_{i+1} | x_i; \theta) \prod_{i=1}^n p(y_i | x_i)$$

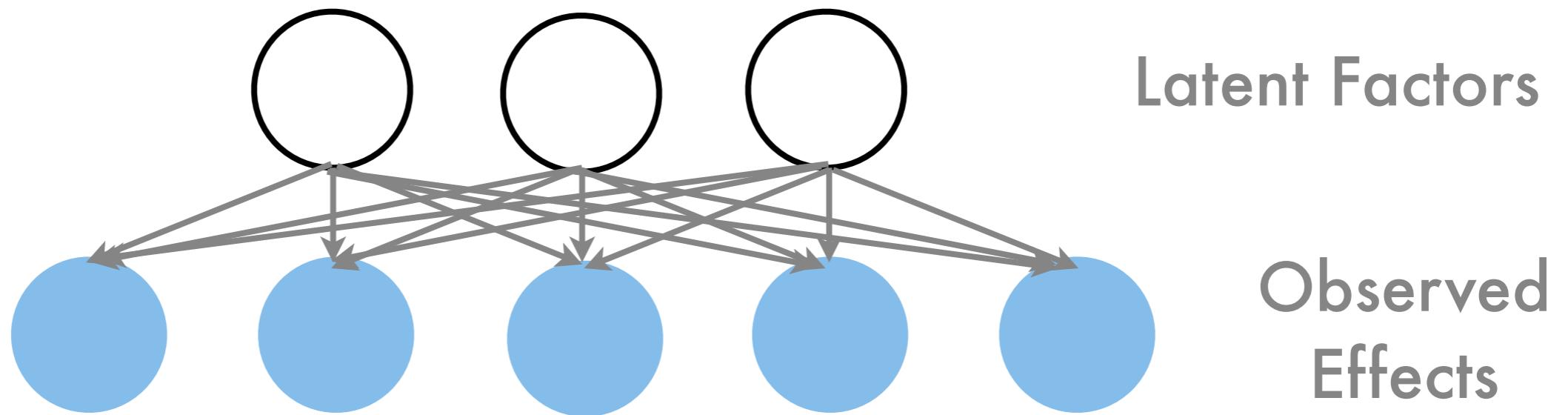
user's
mindset



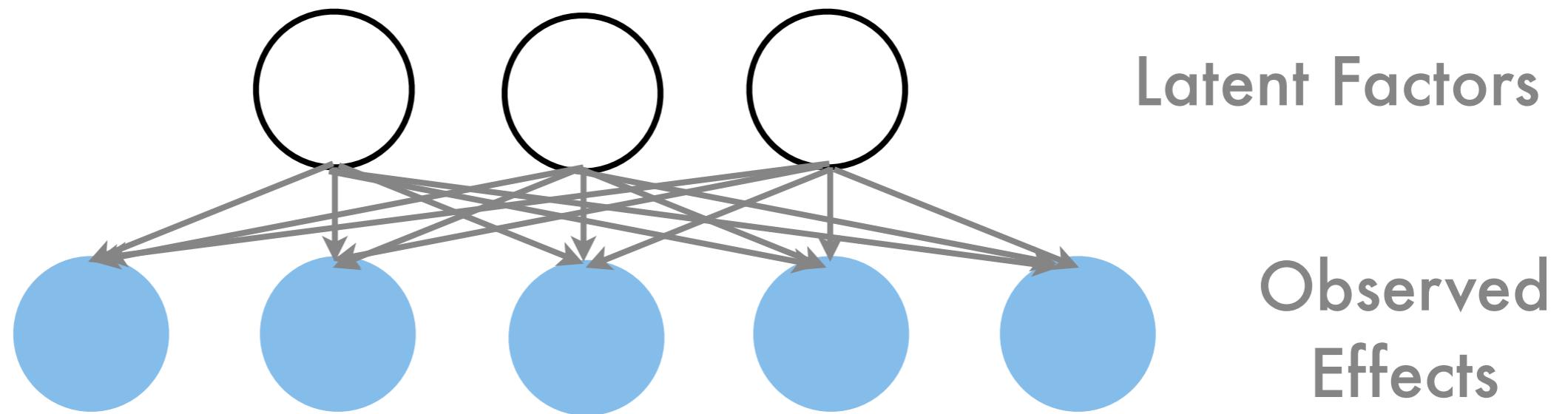
observed
user action

user model for traversal through search results

Factor Graphs

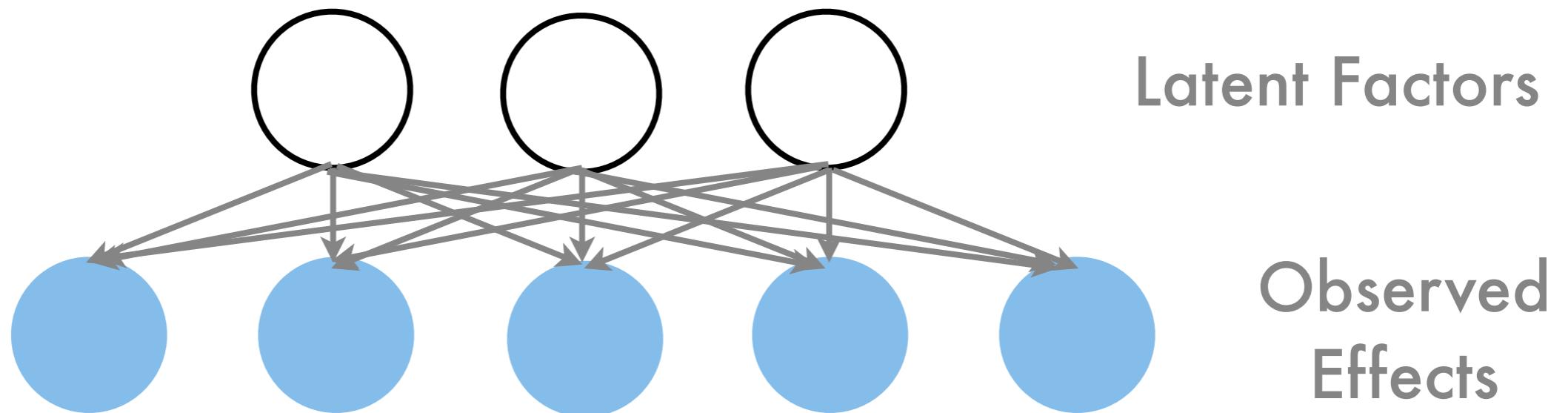


Factor Graphs



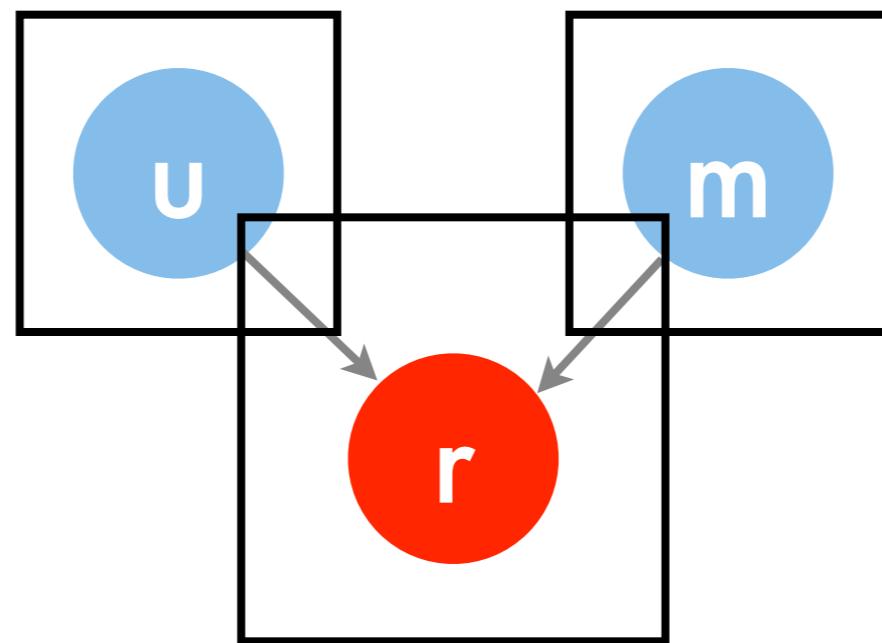
- **Observed effects**
Click behavior, queries, watched news, emails

Factor Graphs

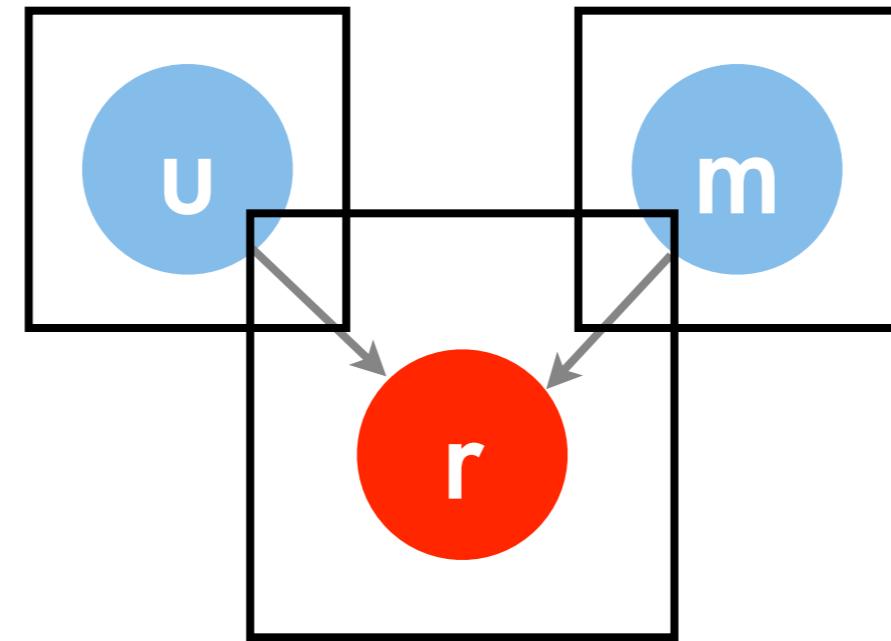


- Observed effects
Click behavior, queries, watched news, emails
- Latent factors
User profile, news content, hot keywords, social connectivity graph, events

Recommender Systems

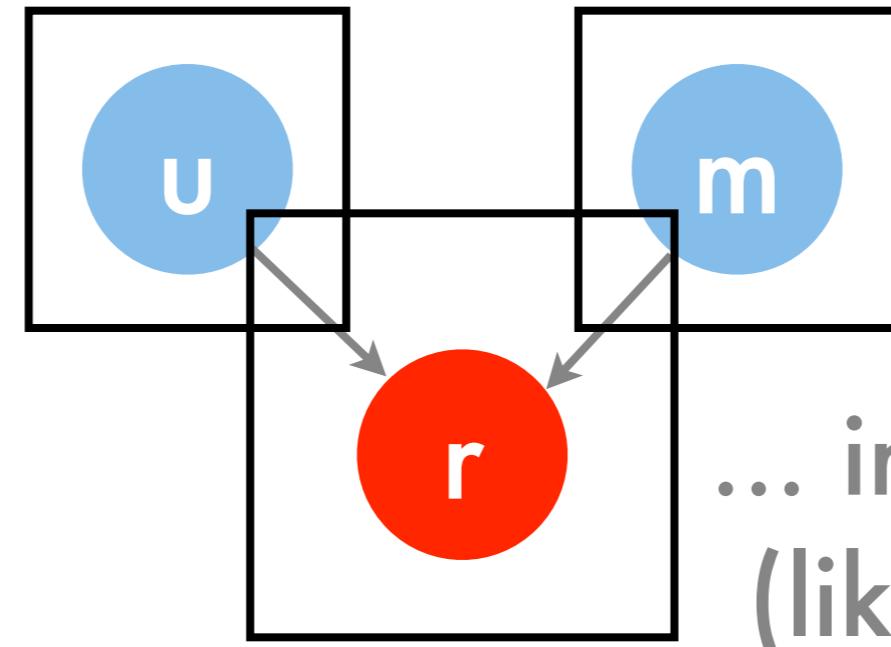


Recommender Systems



- Users u
- Movies m
- Ratings r (but only for a subset of users)

Recommender Systems

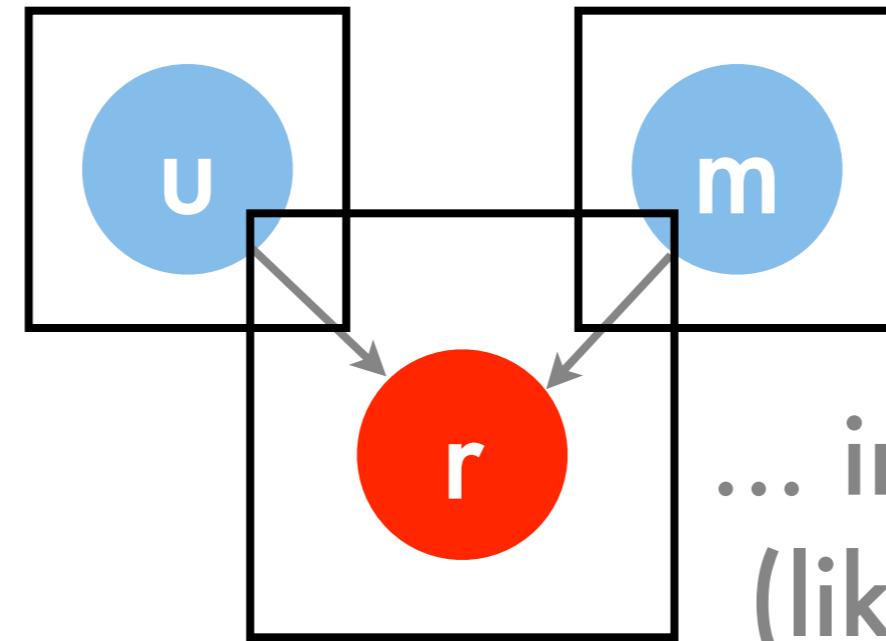


... intersecting plates ...
(like nested for loops)

- Users u
- Movies m
- Ratings r (but only for a subset of users)

Recommender Systems

news,
SearchMonkey
answers
social
ranking
OMG
personals



... intersecting plates ...
(like nested for loops)

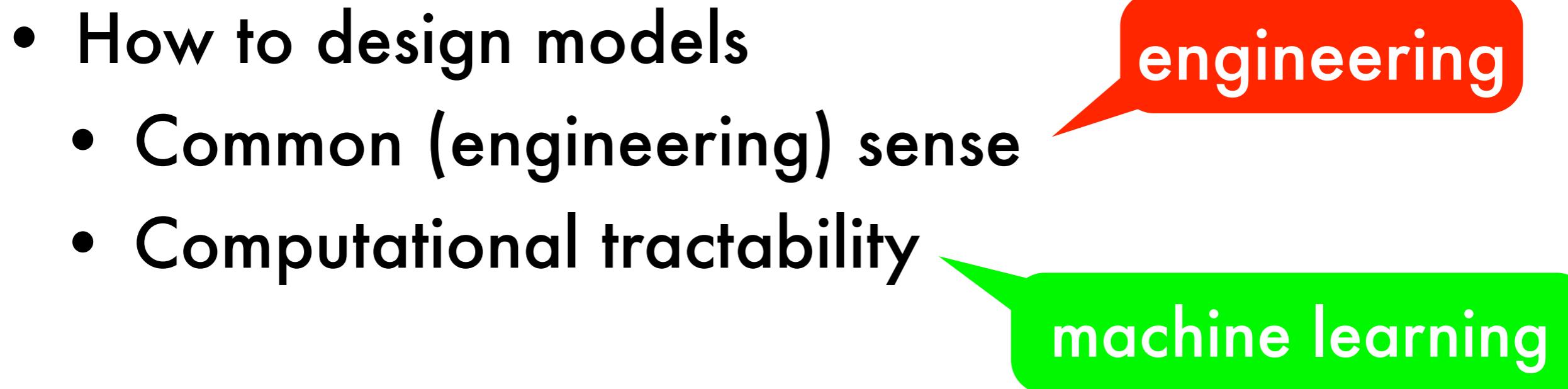
- Users u
- Movies m
- Ratings r (but only for a subset of users)

Challenges

engineering

machine learning

Challenges

- How to design models
 - Common (engineering) sense
 - Computational tractability
- 
- engineering
- machine learning

Challenges

- How to design models
 - Common (engineering) sense
 - Computational tractability
 - Dependency analysis
 - Bayes ball (not in this lecture)
- 
- 

Challenges

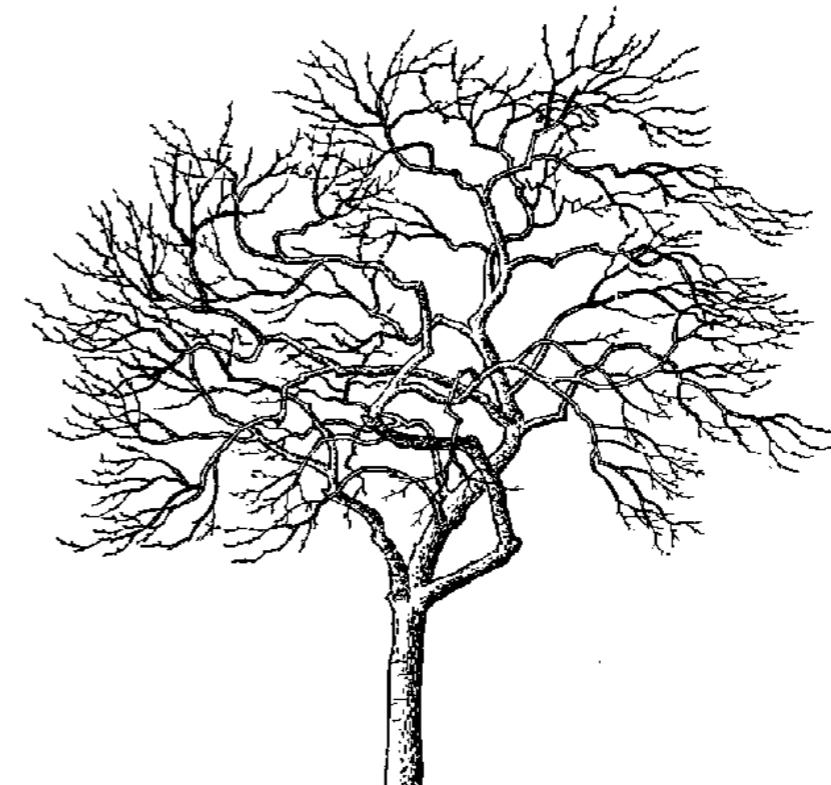
- How to design models
 - Common (engineering) sense
 - Computational tractability
 - Dependency analysis
 - Bayes ball (not in this lecture)
- Inference
 - Easy for fully observed situations
 - Many algorithms if not fully observed
 - Dynamic programming / message passing

engineering

machine learning

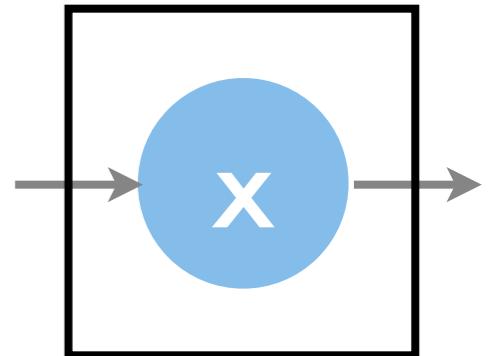
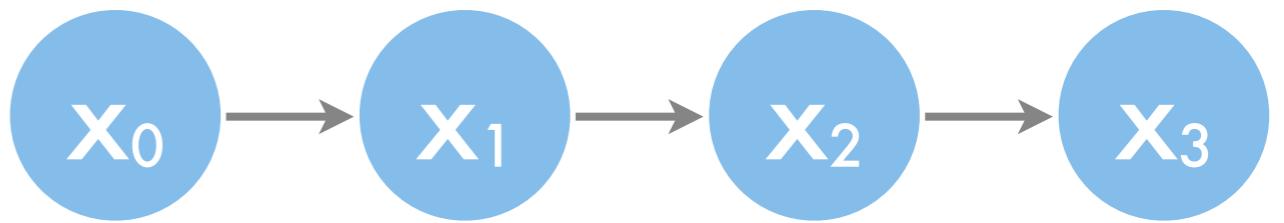
Dynamic Programming

Chains and Trees



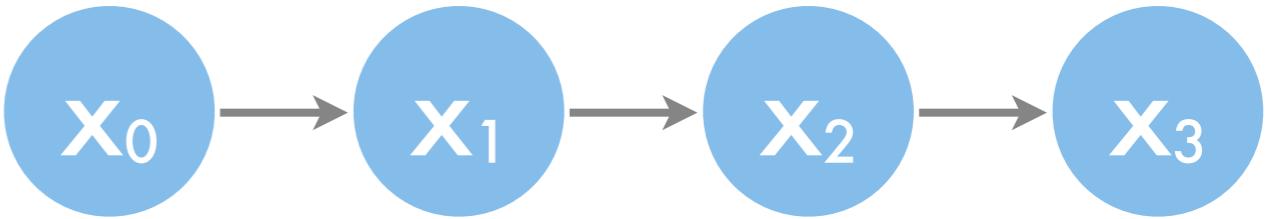
Chains

$$p(x; \theta) = p(x_0; \theta) \prod_{i=1}^{n-1} p(x_{i+1} | x_i; \theta)$$

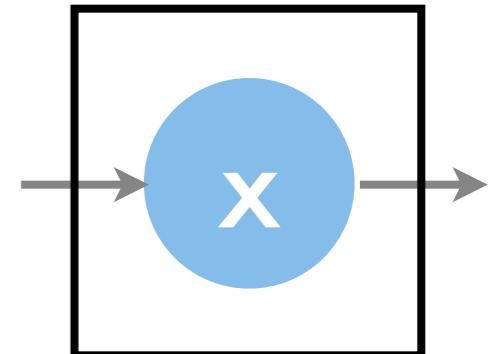


Chains

$$p(x; \theta) = p(x_0; \theta) \prod_{i=1}^{n-1} p(x_{i+1} | x_i; \theta)$$

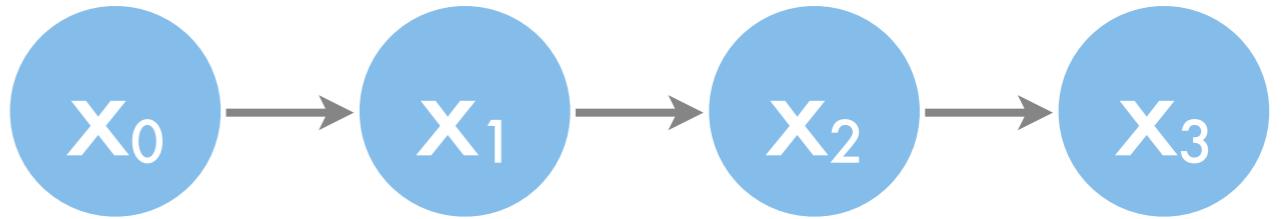


$$p(x_i) = \sum_{x_0, \dots x_{i-1}, x_{i+1} \dots x_n} \underbrace{p(x_0)}_{:= l_0(x_0)} \prod_{j=1}^n p(x_j | x_{j-1})$$



Chains

$$p(x; \theta) = p(x_0; \theta) \prod_{i=1}^{n-1} p(x_{i+1} | x_i; \theta)$$



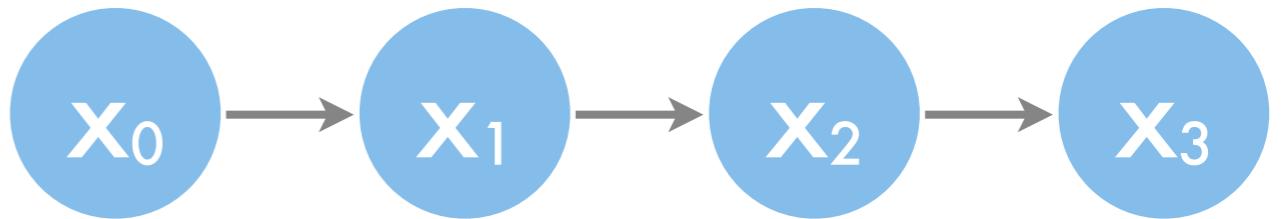
$$p(x_i) = \sum_{x_0, \dots, x_{i-1}, x_{i+1} \dots x_n} \underbrace{p(x_0)}_{:= l_0(x_0)} \prod_{j=1}^n p(x_j | x_{j-1})$$

$$= \sum_{x_1, \dots, x_{i-1}, x_{i+1} \dots x_n} \underbrace{\sum_{x_0} [l_0(x_0) p(x_1 | x_0)]}_{:= l_1(x_1)} \prod_{j=2}^n p(x_j | x_{j-1})$$

```
graph LR; subgraph Sum [ ]; direction TB; l1[l1(x1)] --- prod1[prod]; prod1 --- x((x)); end
```

Chains

$$p(x; \theta) = p(x_0; \theta) \prod_{i=1}^{n-1} p(x_{i+1} | x_i; \theta)$$



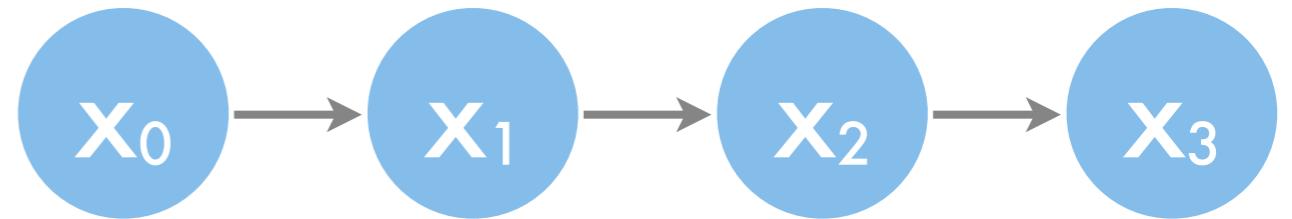
$$p(x_i) = \sum_{x_0, \dots, x_{i-1}, x_{i+1} \dots x_n} \underbrace{p(x_0)}_{:=l_0(x_0)} \prod_{j=1}^n p(x_j | x_{j-1})$$

$$= \sum_{x_1, \dots, x_{i-1}, x_{i+1} \dots x_n} \underbrace{\sum_{x_0} [l_0(x_0) p(x_1 | x_0)]}_{:=l_1(x_1)} \prod_{j=2}^n p(x_j | x_{j-1}) \rightarrow \boxed{x}$$

$$= \sum_{x_2, \dots, x_{i-1}, x_{i+1} \dots x_n} \underbrace{\sum_{x_1} [l_1(x_1) p(x_2 | x_1)]}_{:=l_2(x_2)} \prod_{j=3}^n p(x_j | x_{j-1})$$

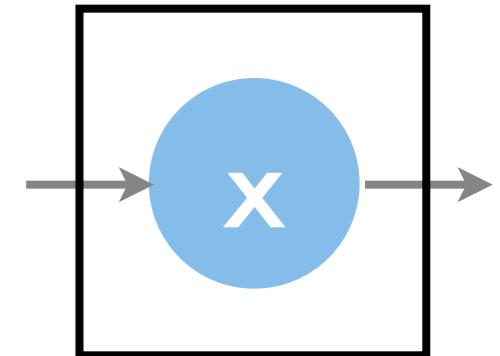
Chains

$$p(x; \theta) = p(x_0; \theta) \prod_{i=1}^{n-1} p(x_{i+1} | x_i; \theta)$$



$$p(x_i) = l_i(x_i) \sum_{x_{i+1} \dots x_n} \prod_{j=i}^{n-1} p(x_{j+1} | x_j)$$

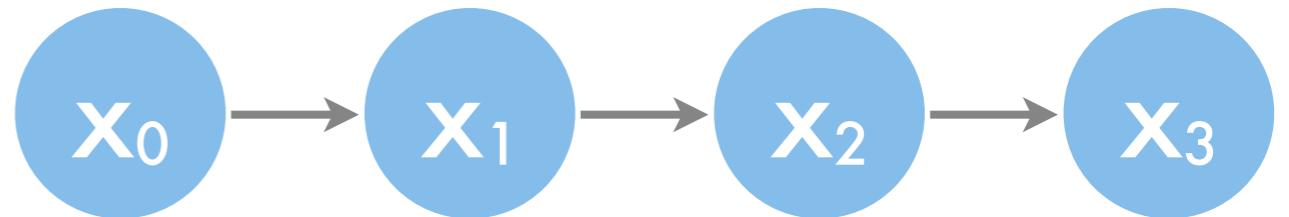
$$= l_i(x_i) \sum_{x_{i+1} \dots x_{n-1}} \prod_{j=i}^{n-2} p(x_{j+1} | x_j) \underbrace{\sum_{x_n} p(x_n | x_{n-1})}_{:= r_{n-1}(x_{n-1})}$$



$$= l_i(x_i) \sum_{x_{i+1} \dots x_{n-2}} \prod_{j=i}^{n-3} p(x_{j+1} | x_j) \underbrace{\sum_{x_{n-1}} p(x_{n-1} | x_{n-2}) r_{n-1}(x_{n-1})}_{:= r_{n-2}(x_{n-2})}$$

Chains

$$p(x; \theta) = p(x_0; \theta) \prod_{i=1}^{n-1} p(x_{i+1} | x_i; \theta)$$



- **Forward recursion**

$$l_0(x_0) := p(x_0) \text{ and } l_i(x_i) := \sum_{x_{i-1}} l_{i-1}(x_{i-1}) p(x_i | x_{i-1})$$

- **Backward recursion**

$$r_n(x_n) := 1 \text{ and } r_i(x_i) := \sum_{x_{i+1}} r_{i+1}(x_{i+1}) p(x_{i+1} | x_i)$$

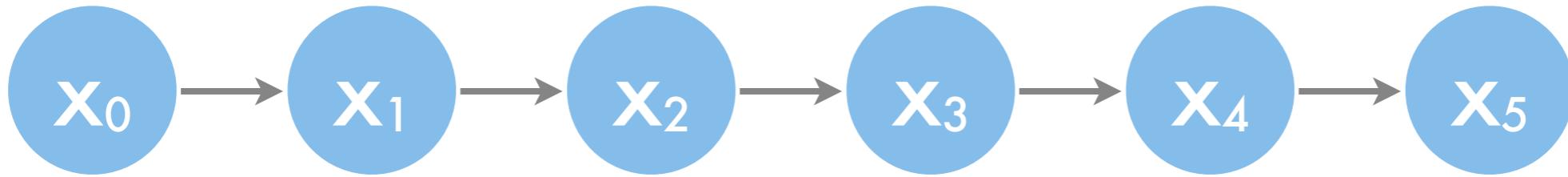
- **Marginalization & conditioning**

$$p(x_i) = l_i(x_i) r_i(x_i)$$

$$p(x_{-i} | x_i) = \frac{p(x)}{p(x_i)}$$

$$p(x_i, x_{i+1}) = l_i(x_i) p(x_{i+1} | x_i) r_i(x_{i+1})$$

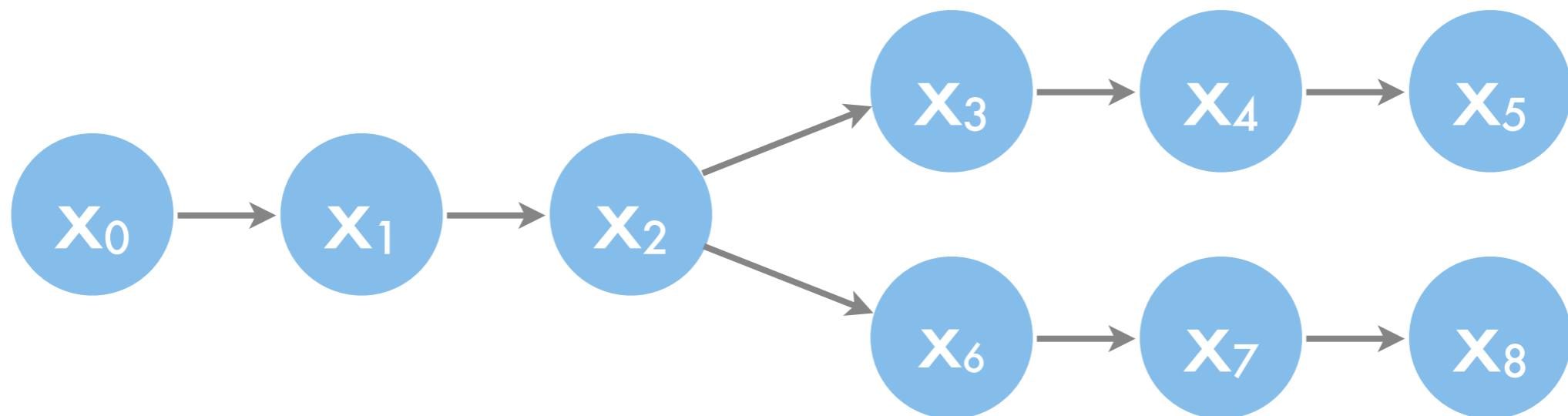
Chains



- Send forward messages starting from left node
→ $m_{i-1 \rightarrow i}(x_i) = \sum_{x_{i-1}} m_{i-2 \rightarrow i-1}(x_{i-1}) f(x_{i-1}, x_i)$
- Send backward messages starting from right node

$$m_{i+1 \rightarrow i}(x_i) = \sum_{x_{i+1}} m_{i+2 \rightarrow i+1}(x_{i+1}) f(x_i, x_{i+1})$$
 ←

Trees



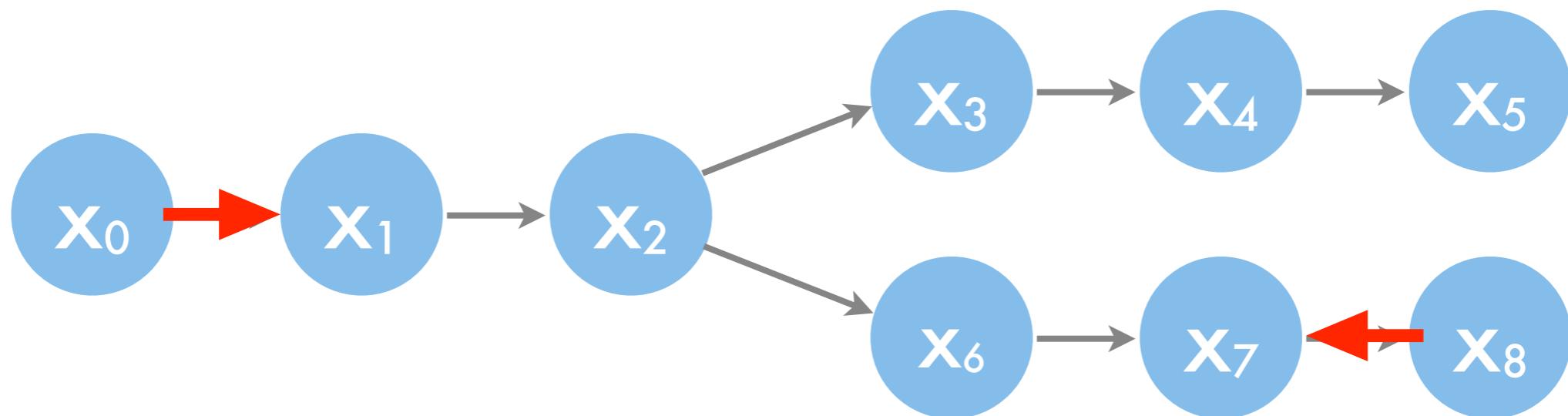
- Forward/Backward messages as normal for chain
- When we have more edges for a vertex use ...

$$m_{2 \rightarrow 3}(x_3) = \sum_{x_2} m_{1 \rightarrow 2}(x_2) m_{6 \rightarrow 2}(x_2) f(x_2, x_3)$$

$$m_{2 \rightarrow 6}(x_6) = \sum_{x_2} m_{1 \rightarrow 2}(x_2) m_{3 \rightarrow 2}(x_2) f(x_2, x_6)$$

$$m_{2 \rightarrow 1}(x_1) = \sum_{x_2} m_{3 \rightarrow 2}(x_2) m_{6 \rightarrow 2}(x_2) f(x_1, x_2)$$

Trees



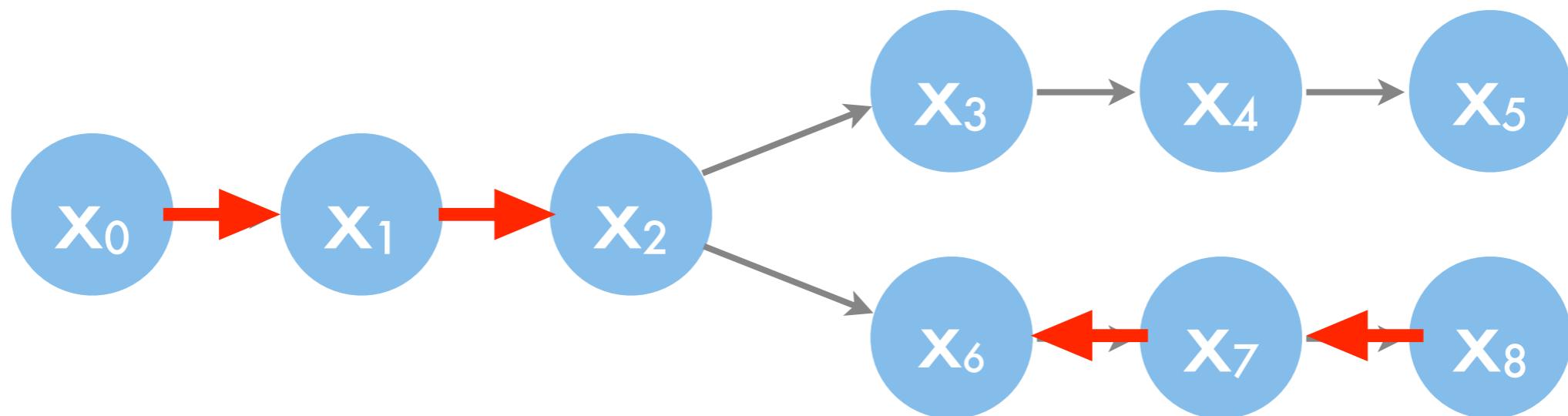
- Forward/Backward messages as normal for chain
- When we have more edges for a vertex use ...

$$m_{2 \rightarrow 3}(x_3) = \sum_{x_2} m_{1 \rightarrow 2}(x_2) m_{6 \rightarrow 2}(x_2) f(x_2, x_3)$$

$$m_{2 \rightarrow 6}(x_6) = \sum_{x_2} m_{1 \rightarrow 2}(x_2) m_{3 \rightarrow 2}(x_2) f(x_2, x_6)$$

$$m_{2 \rightarrow 1}(x_1) = \sum_{x_2} m_{3 \rightarrow 2}(x_2) m_{6 \rightarrow 2}(x_2) f(x_1, x_2)$$

Trees



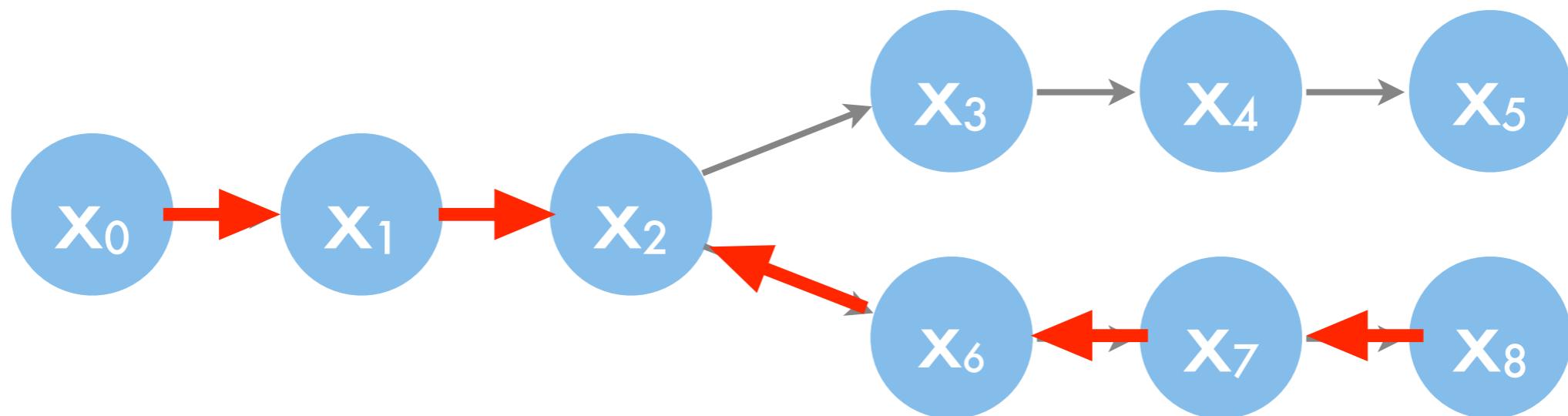
- Forward/Backward messages as normal for chain
- When we have more edges for a vertex use ...

$$m_{2 \rightarrow 3}(x_3) = \sum_{x_2} m_{1 \rightarrow 2}(x_2) m_{6 \rightarrow 2}(x_2) f(x_2, x_3)$$

$$m_{2 \rightarrow 6}(x_6) = \sum_{x_2} m_{1 \rightarrow 2}(x_2) m_{3 \rightarrow 2}(x_2) f(x_2, x_6)$$

$$m_{2 \rightarrow 1}(x_1) = \sum_{x_2} m_{3 \rightarrow 2}(x_2) m_{6 \rightarrow 2}(x_2) f(x_1, x_2)$$

Trees



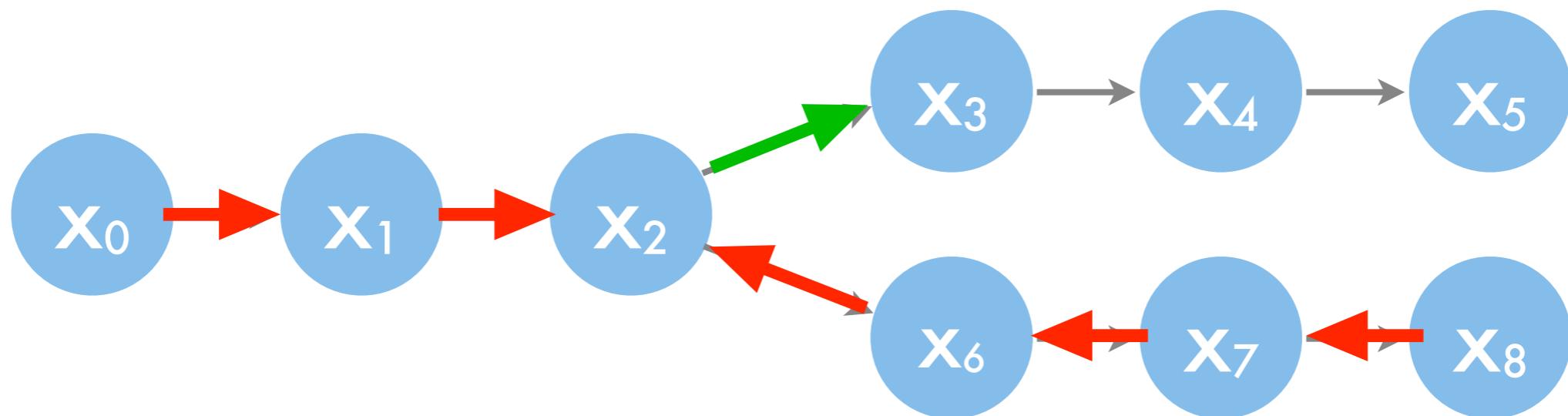
- Forward/Backward messages as normal for chain
- When we have more edges for a vertex use ...

$$m_{2 \rightarrow 3}(x_3) = \sum_{x_2} m_{1 \rightarrow 2}(x_2) m_{6 \rightarrow 2}(x_2) f(x_2, x_3)$$

$$m_{2 \rightarrow 6}(x_6) = \sum_{x_2} m_{1 \rightarrow 2}(x_2) m_{3 \rightarrow 2}(x_2) f(x_2, x_6)$$

$$m_{2 \rightarrow 1}(x_1) = \sum_{x_2} m_{3 \rightarrow 2}(x_2) m_{6 \rightarrow 2}(x_2) f(x_1, x_2)$$

Trees



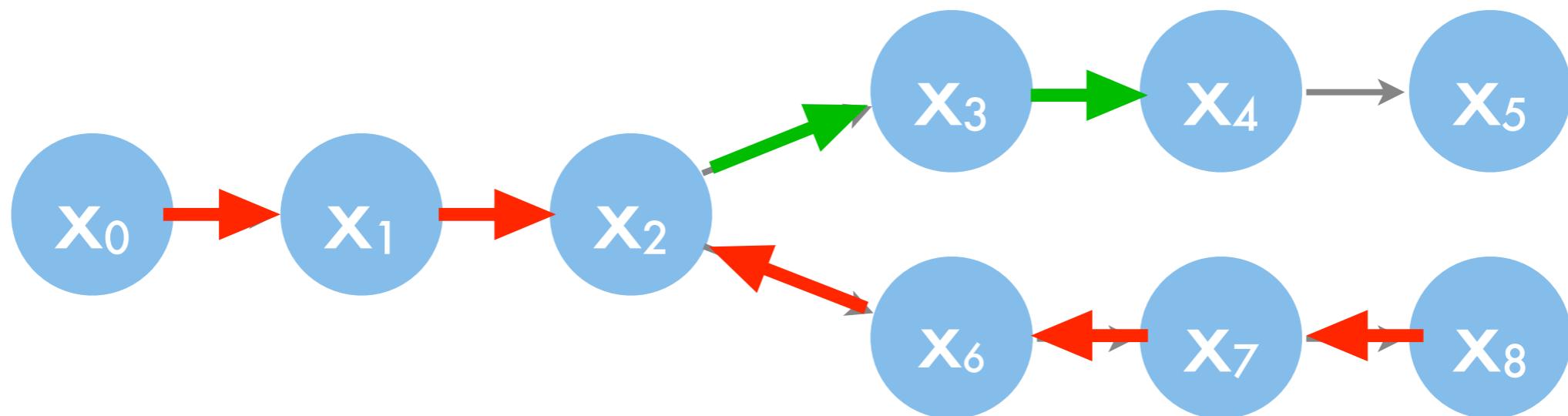
- Forward/Backward messages as normal for chain
- When we have more edges for a vertex use ...

$$m_{2 \rightarrow 3}(x_3) = \sum_{x_2} m_{1 \rightarrow 2}(x_2) m_{6 \rightarrow 2}(x_2) f(x_2, x_3)$$

$$m_{2 \rightarrow 6}(x_6) = \sum_{x_2} m_{1 \rightarrow 2}(x_2) m_{3 \rightarrow 2}(x_2) f(x_2, x_6)$$

$$m_{2 \rightarrow 1}(x_1) = \sum_{x_2} m_{3 \rightarrow 2}(x_2) m_{6 \rightarrow 2}(x_2) f(x_1, x_2)$$

Trees



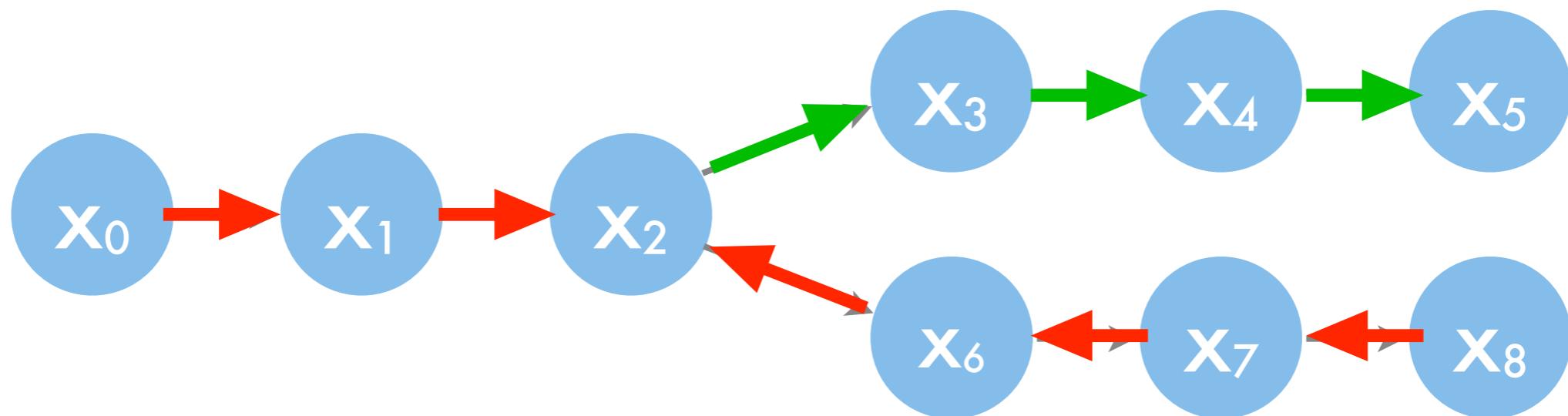
- Forward/Backward messages as normal for chain
- When we have more edges for a vertex use ...

$$m_{2 \rightarrow 3}(x_3) = \sum_{x_2} m_{1 \rightarrow 2}(x_2) m_{6 \rightarrow 2}(x_2) f(x_2, x_3)$$

$$m_{2 \rightarrow 6}(x_6) = \sum_{x_2} m_{1 \rightarrow 2}(x_2) m_{3 \rightarrow 2}(x_2) f(x_2, x_6)$$

$$m_{2 \rightarrow 1}(x_1) = \sum_{x_2} m_{3 \rightarrow 2}(x_2) m_{6 \rightarrow 2}(x_2) f(x_1, x_2)$$

Trees



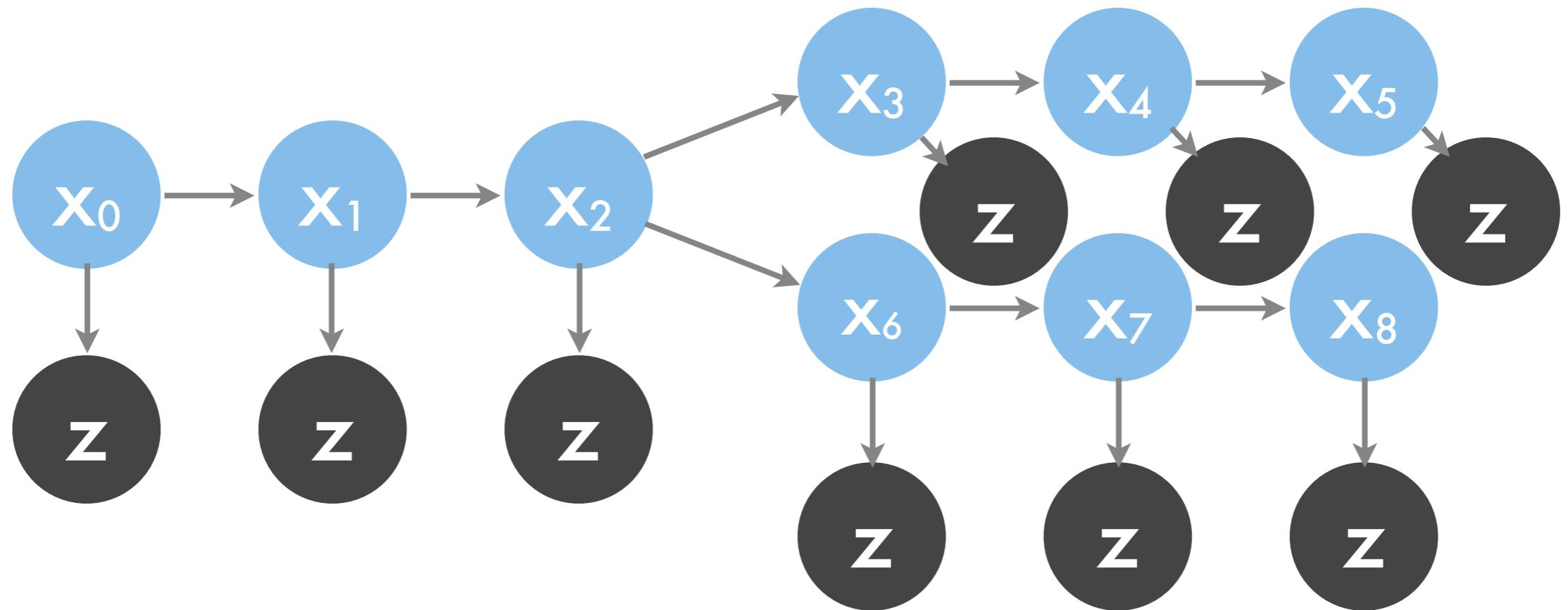
- Forward/Backward messages as normal for chain
- When we have more edges for a vertex use ...

$$m_{2 \rightarrow 3}(x_3) = \sum_{x_2} m_{1 \rightarrow 2}(x_2) m_{6 \rightarrow 2}(x_2) f(x_2, x_3)$$

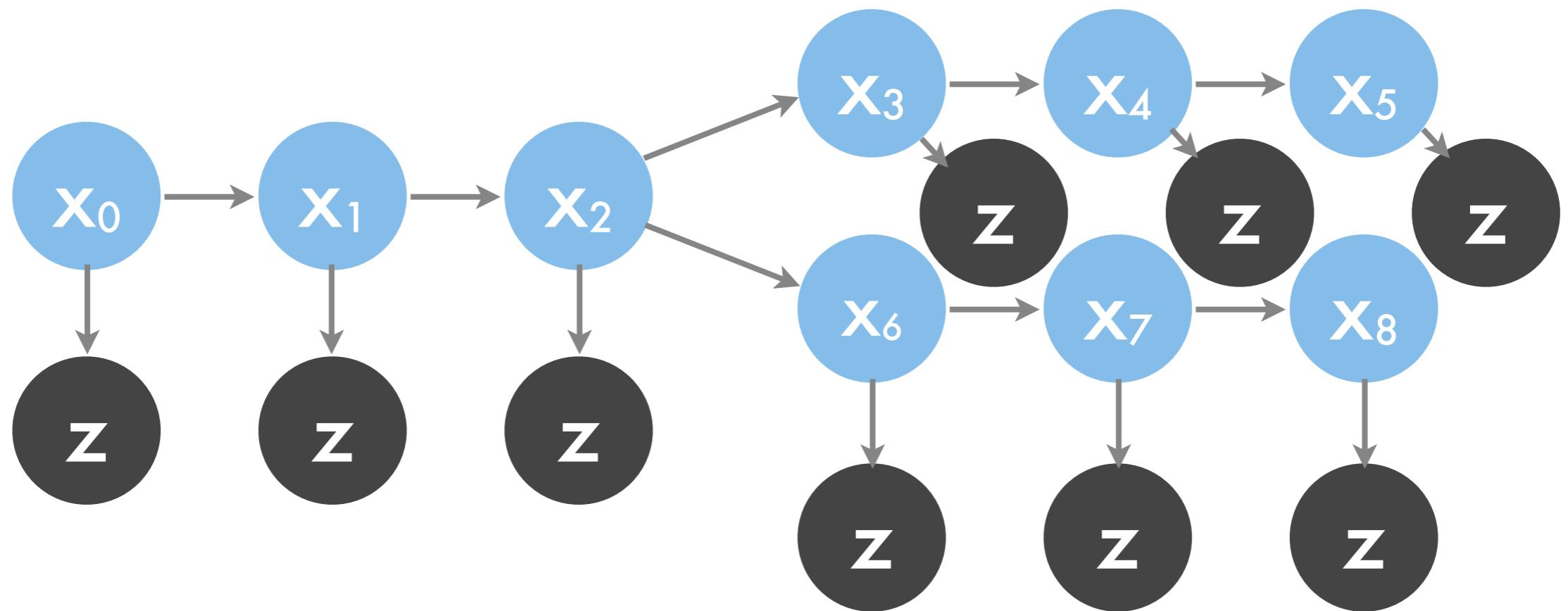
$$m_{2 \rightarrow 6}(x_6) = \sum_{x_2} m_{1 \rightarrow 2}(x_2) m_{3 \rightarrow 2}(x_2) f(x_2, x_6)$$

$$m_{2 \rightarrow 1}(x_1) = \sum_{x_2} m_{3 \rightarrow 2}(x_2) m_{6 \rightarrow 2}(x_2) f(x_1, x_2)$$

Trees



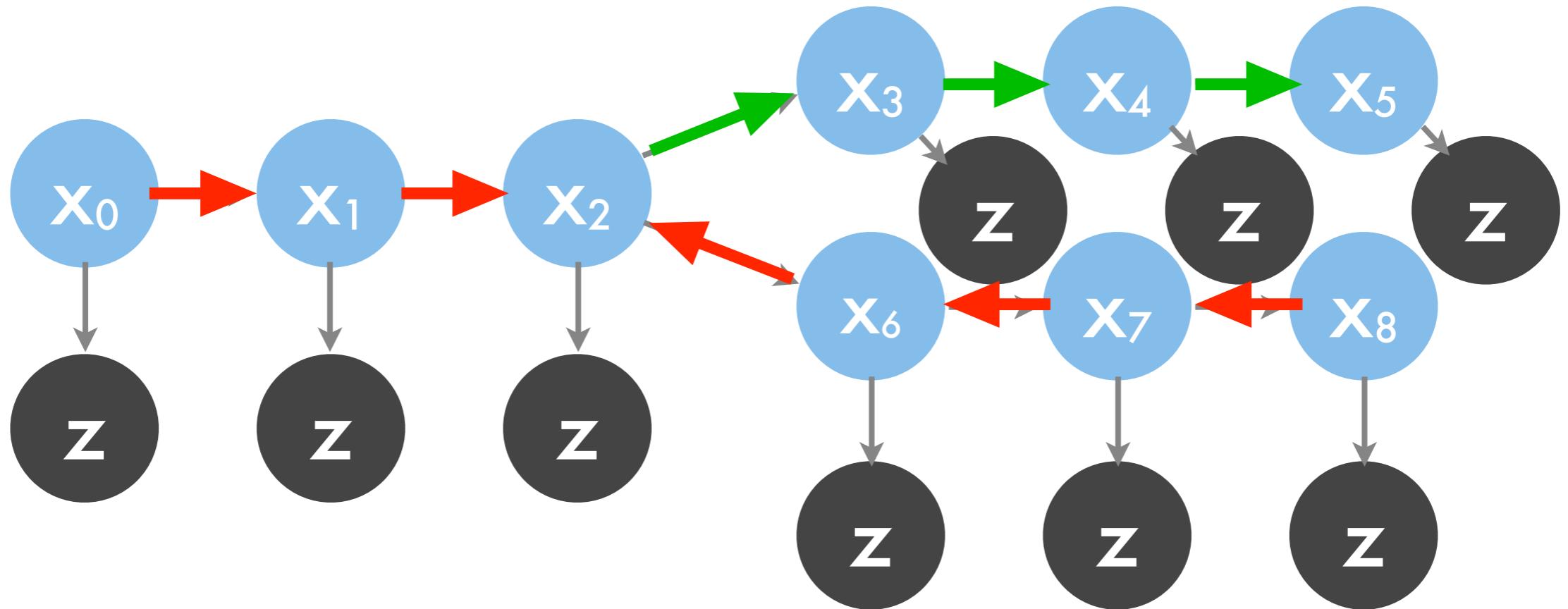
Trees



- Joint distribution over latent state and observations
- To compute conditional probability we need to normalize

$$p(x, z) = p(x) \prod_i p(z_i | x_i) = \prod_{i,j \in T} f(x_i, x_j) \prod_i g(x_i, z_i)$$

Trees



$$p(x_i | \text{rest}) \propto \sum_{x^{-i}} \left[\prod_{j, k \in T} f(x_j, x_k) \prod_j g(x_j) \right]$$

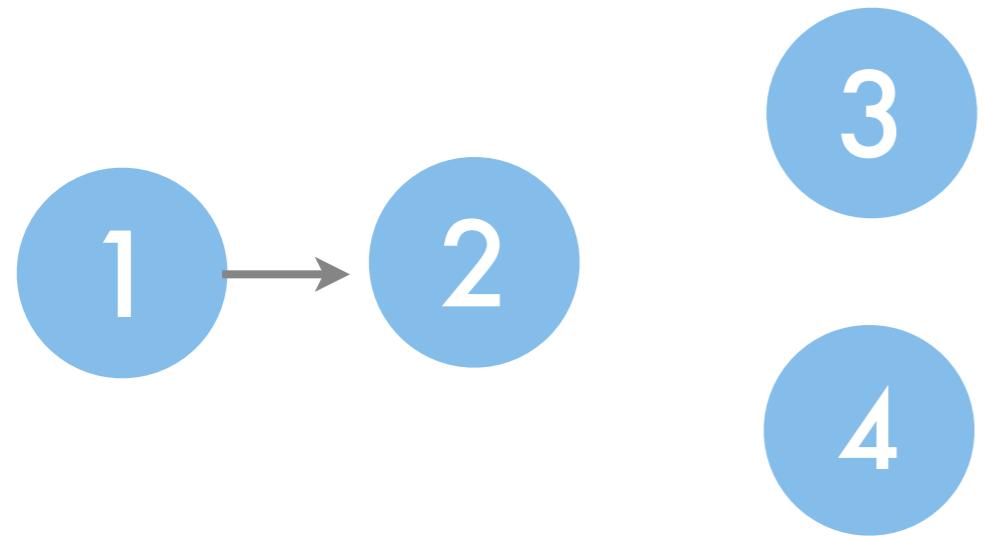
$$= g(x_i) \prod_{(j, i)} m_{j \rightarrow i}(x_i)$$

Junction Trees



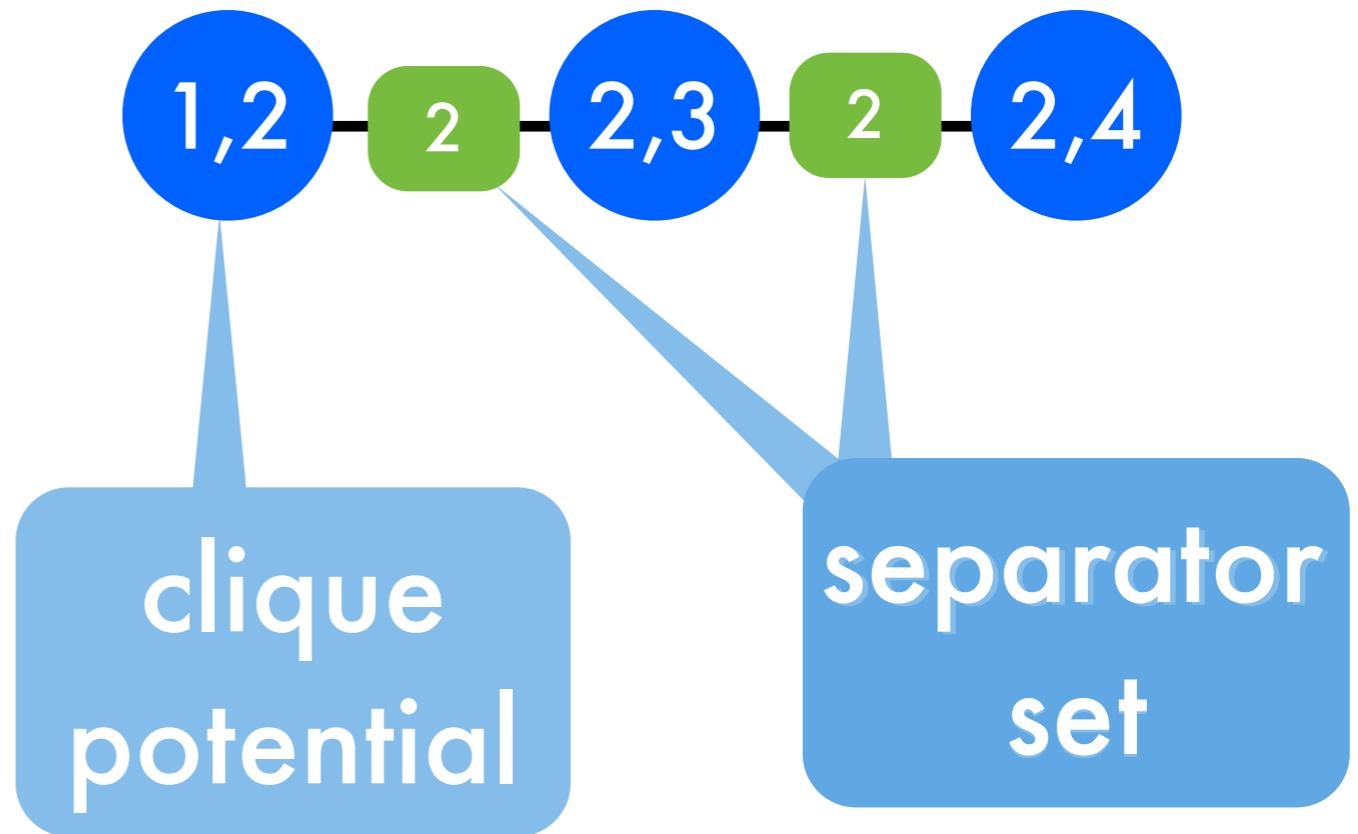
Junction Trees

$$f(x_1, x_2)f(x_2, x_3)f(x_2, x_4)$$



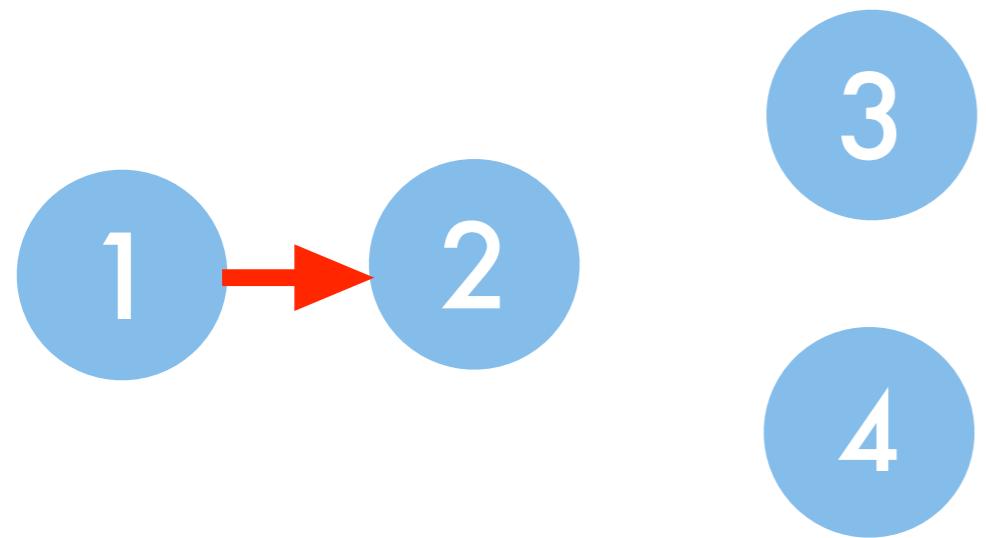
$$m_{i \rightarrow j}(x_j) = \sum_{x_i} f(x_i, x_j) \prod_{l \neq j} m_{l \rightarrow i}(x_j)$$

clique
potential



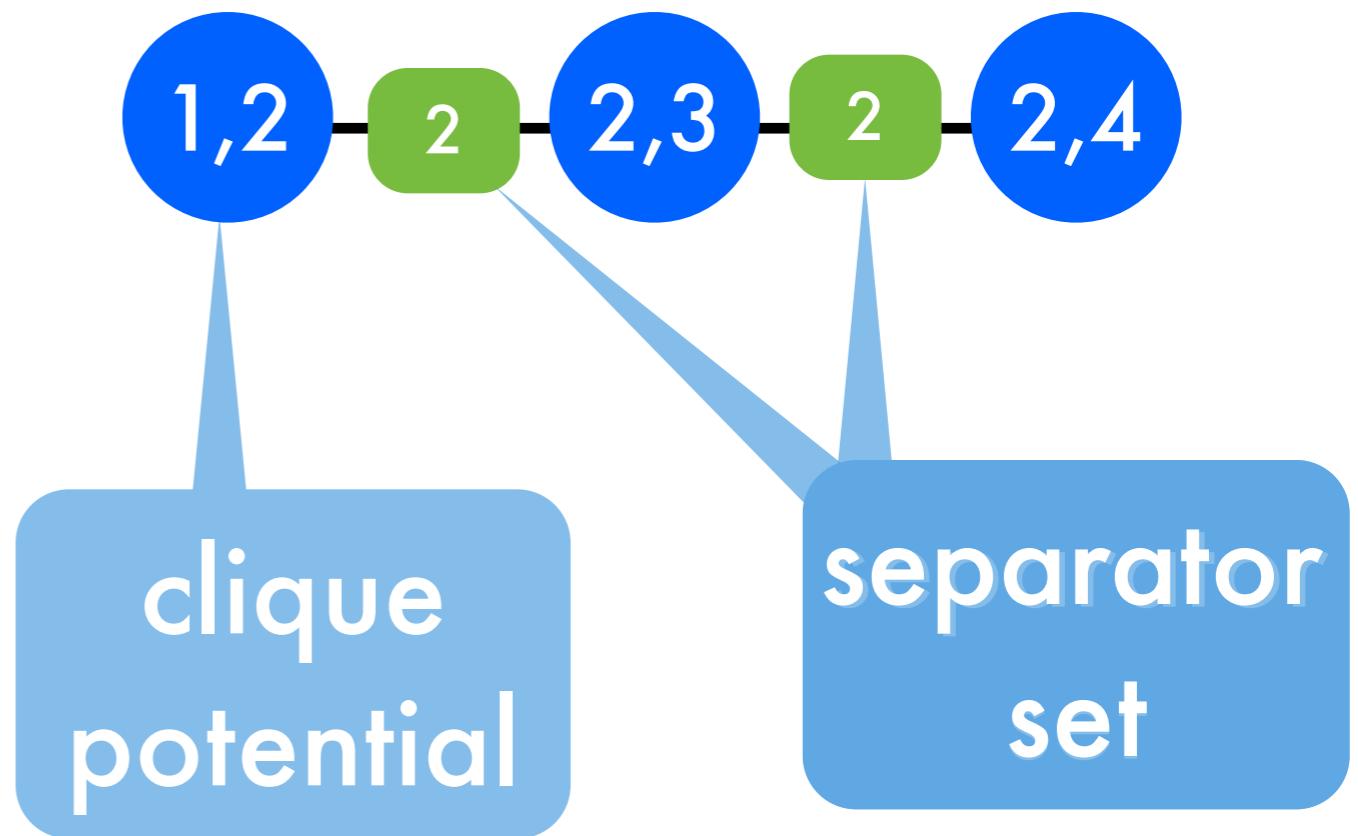
Junction Trees

$$f(x_1, x_2)f(x_2, x_3)f(x_2, x_4)$$



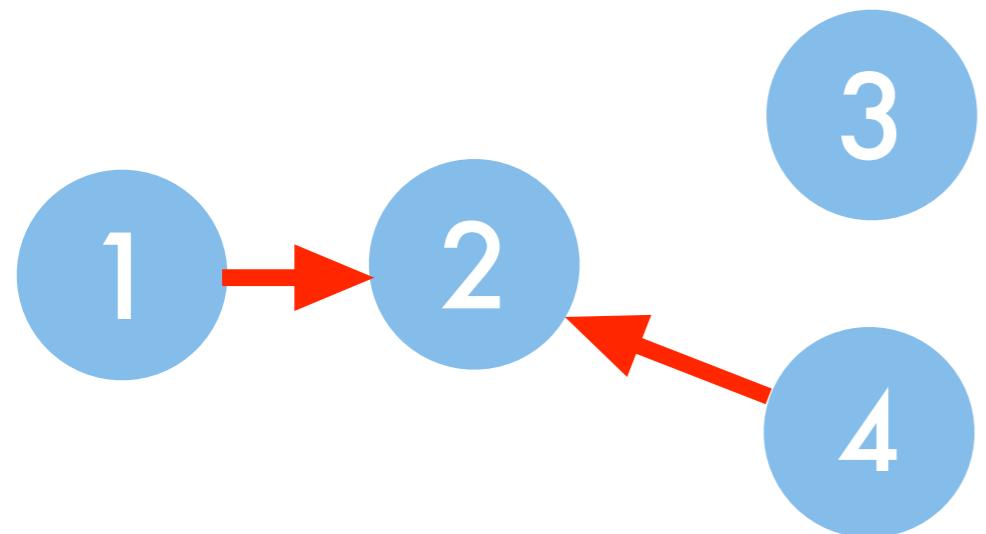
$$m_{i \rightarrow j}(x_j) = \sum_{x_i} f(x_i, x_j) \prod_{l \neq j} m_{l \rightarrow i}(x_j)$$

clique
potential



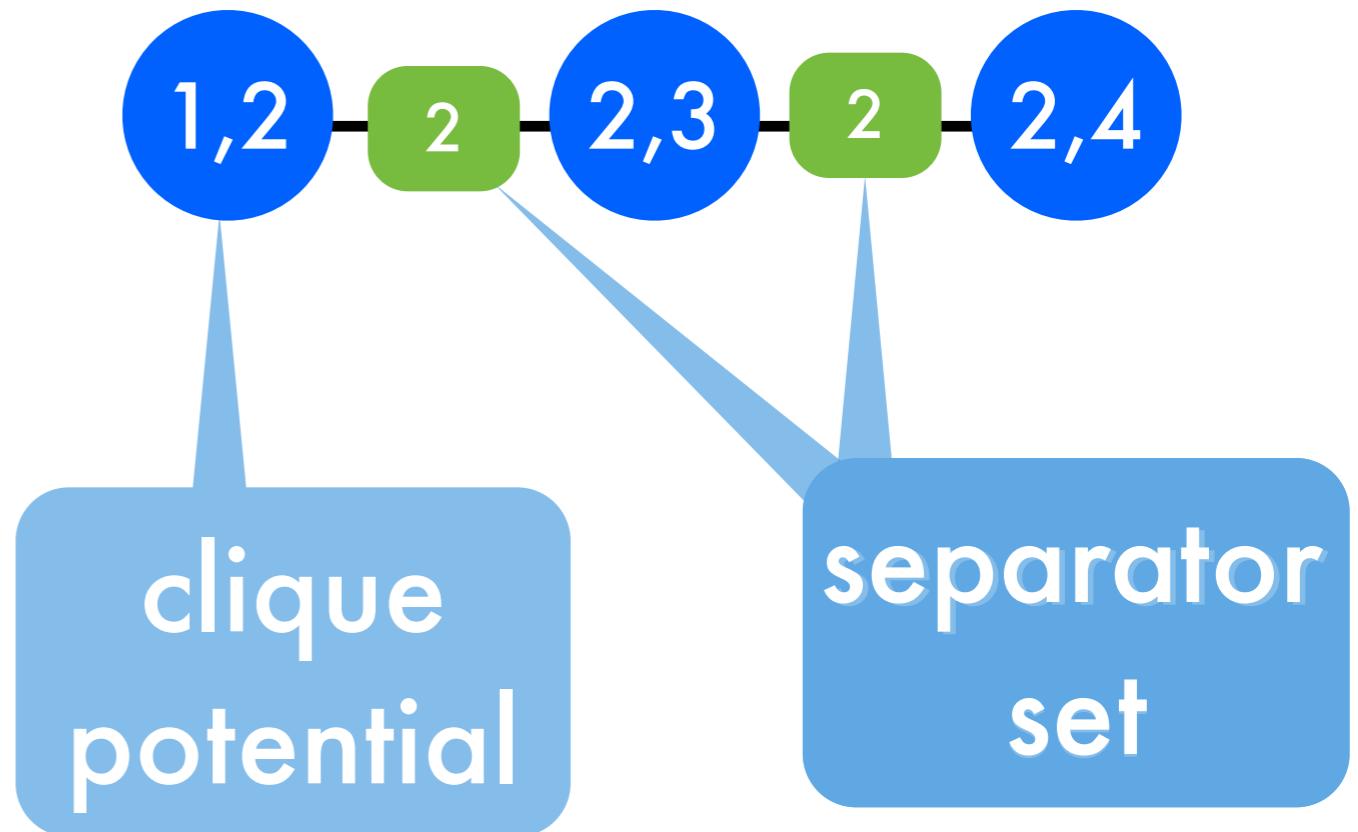
Junction Trees

$$f(x_1, x_2)f(x_2, x_3)f(x_2, x_4)$$



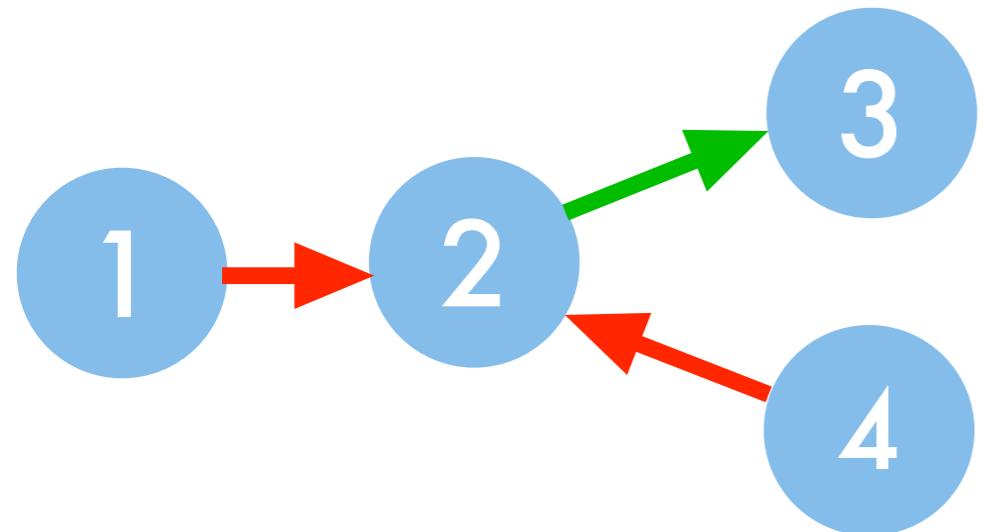
$$m_{i \rightarrow j}(x_j) = \sum_{x_i} f(x_i, x_j) \prod_{l \neq j} m_{l \rightarrow i}(x_j)$$

clique
potential



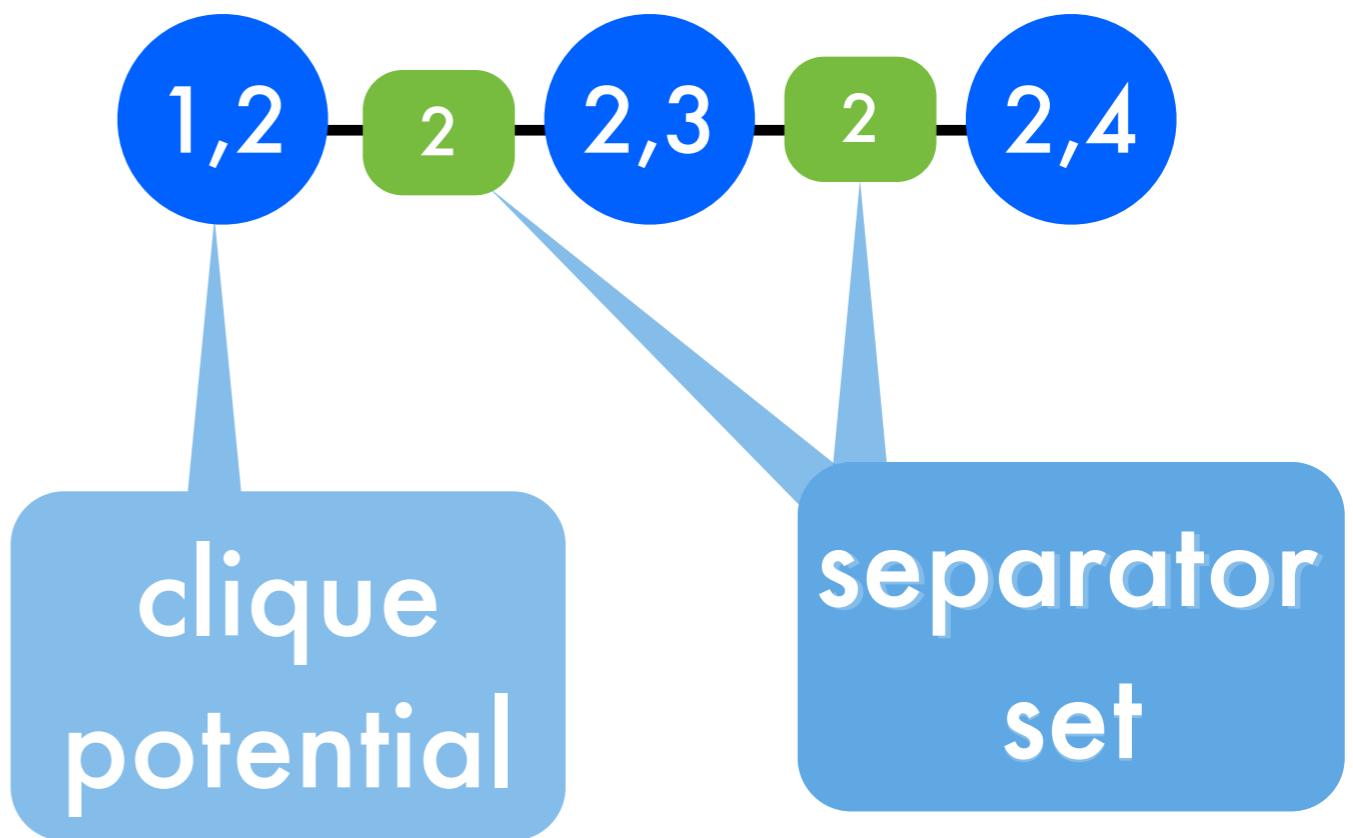
Junction Trees

$$f(x_1, x_2)f(x_2, x_3)f(x_2, x_4)$$

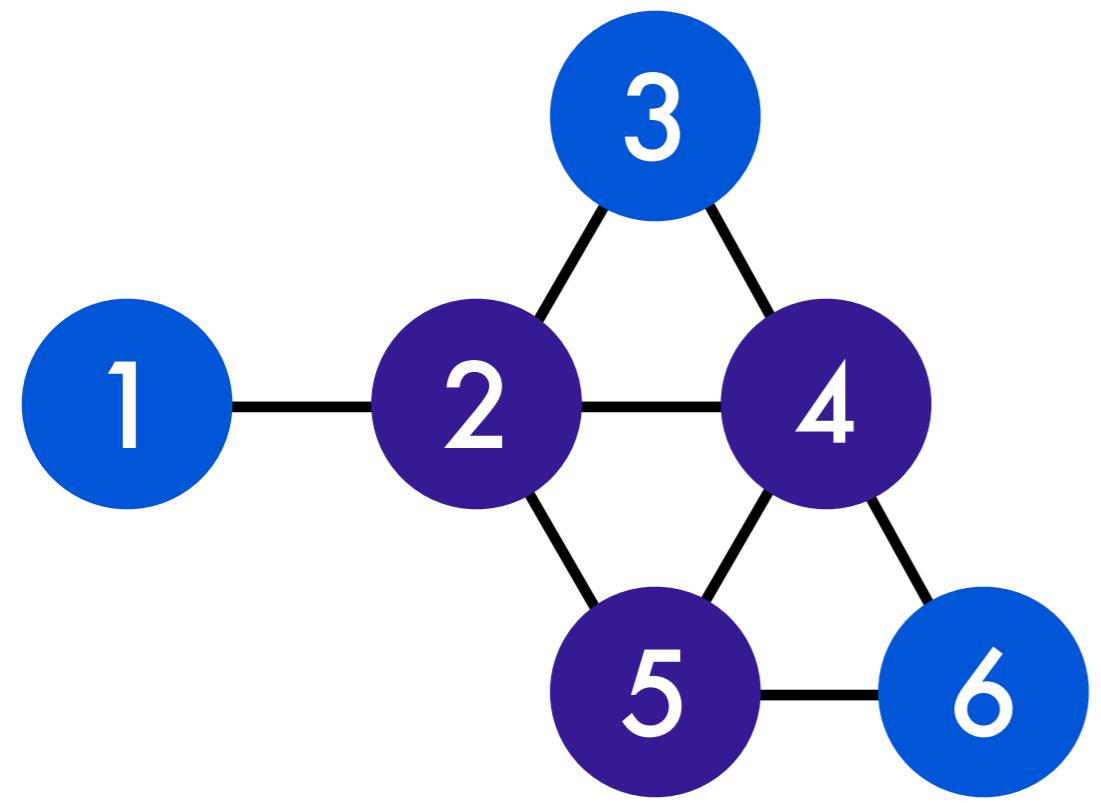


$$m_{i \rightarrow j}(x_j) = \sum_{x_i} f(x_i, x_j) \prod_{l \neq j} m_{l \rightarrow i}(x_j)$$

clique
potential

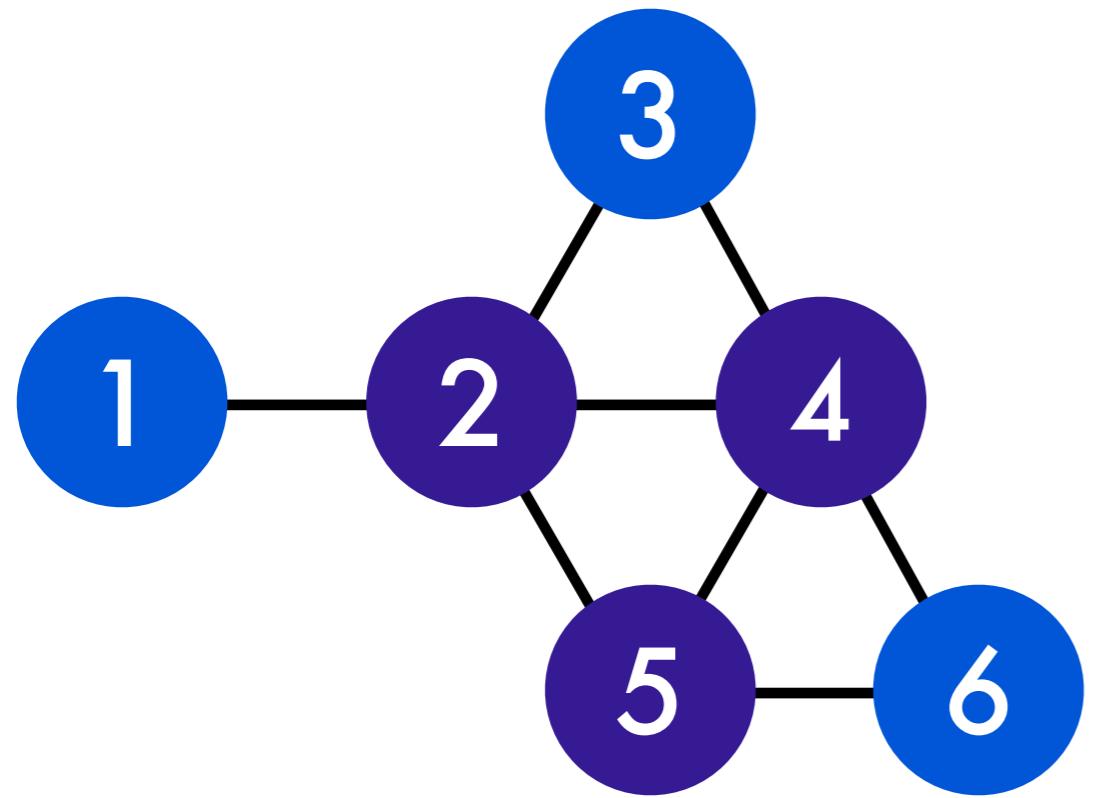


Junction Trees

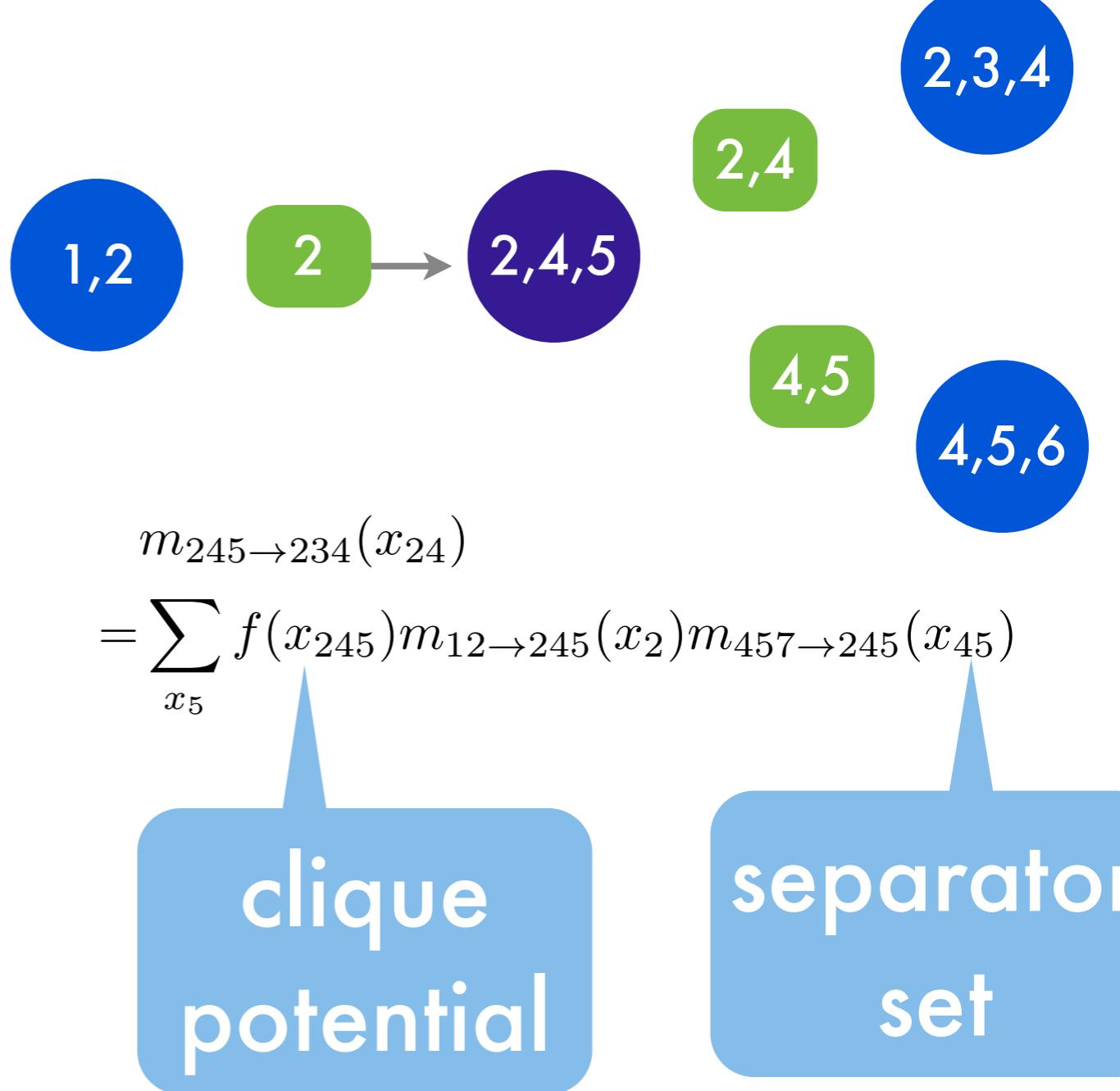


dependency
graph

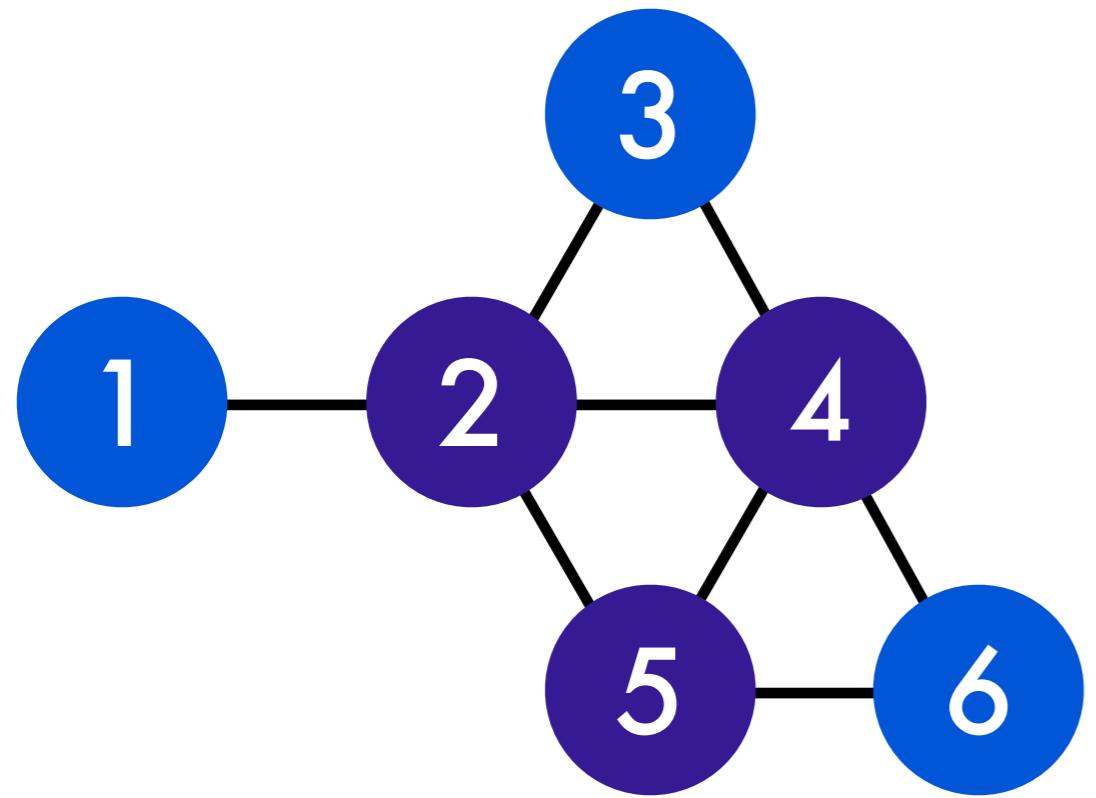
Junction Trees



dependency
graph



Junction Trees



dependency
graph



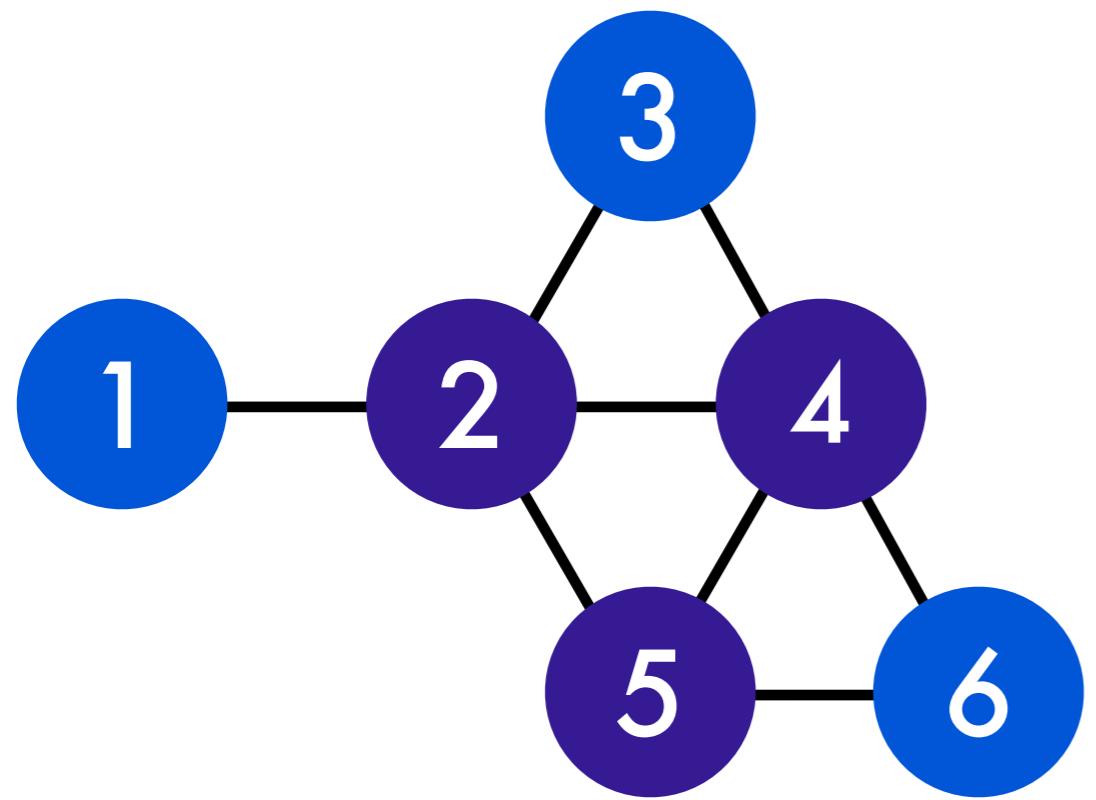
$$m_{245 \rightarrow 234}(x_{24})$$

$$= \sum_{x_5} f(x_{245}) m_{12 \rightarrow 245}(x_2) m_{457 \rightarrow 245}(x_{45})$$

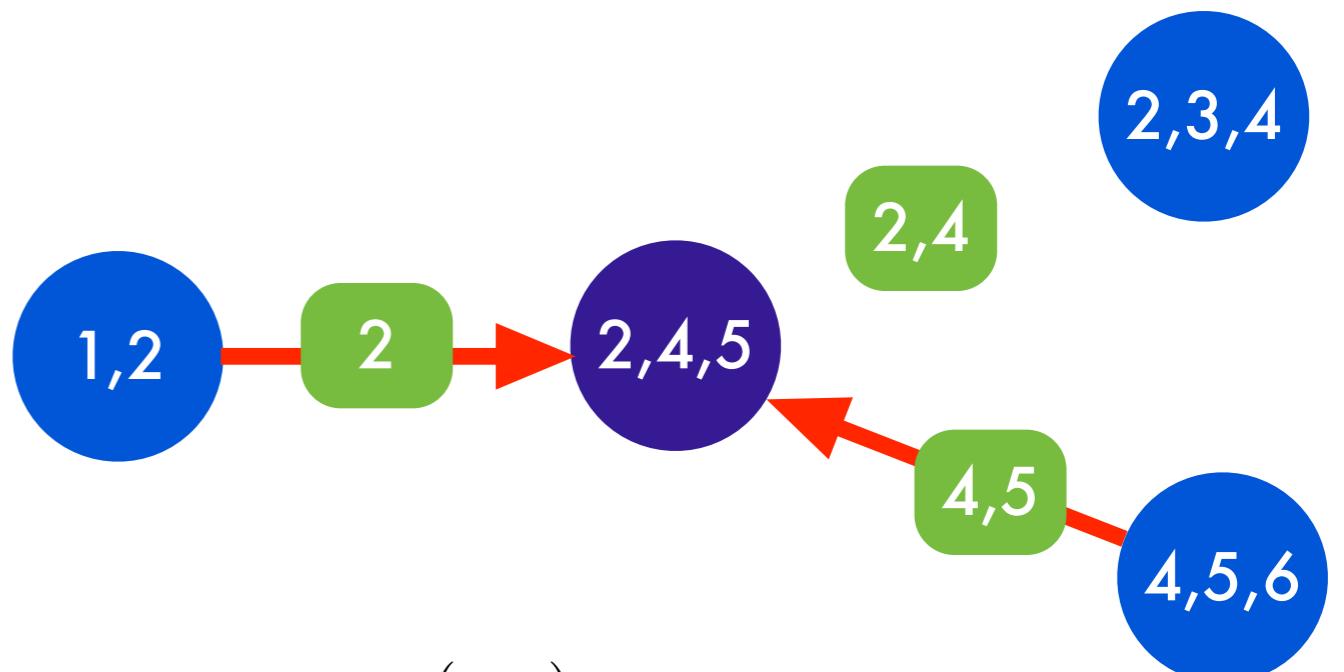
clique
potential

separator
set

Junction Trees



dependency
graph



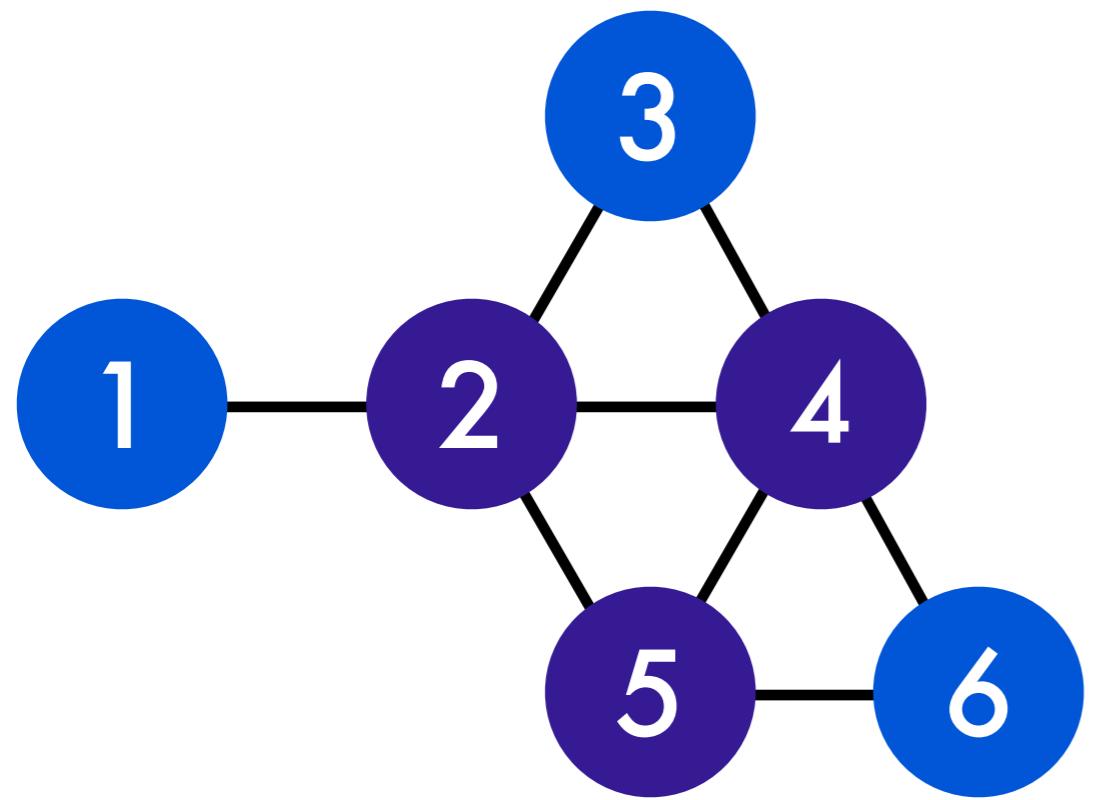
$$m_{245 \rightarrow 234}(x_{24})$$

$$= \sum_{x_5} f(x_{245}) m_{12 \rightarrow 245}(x_2) m_{457 \rightarrow 245}(x_{45})$$

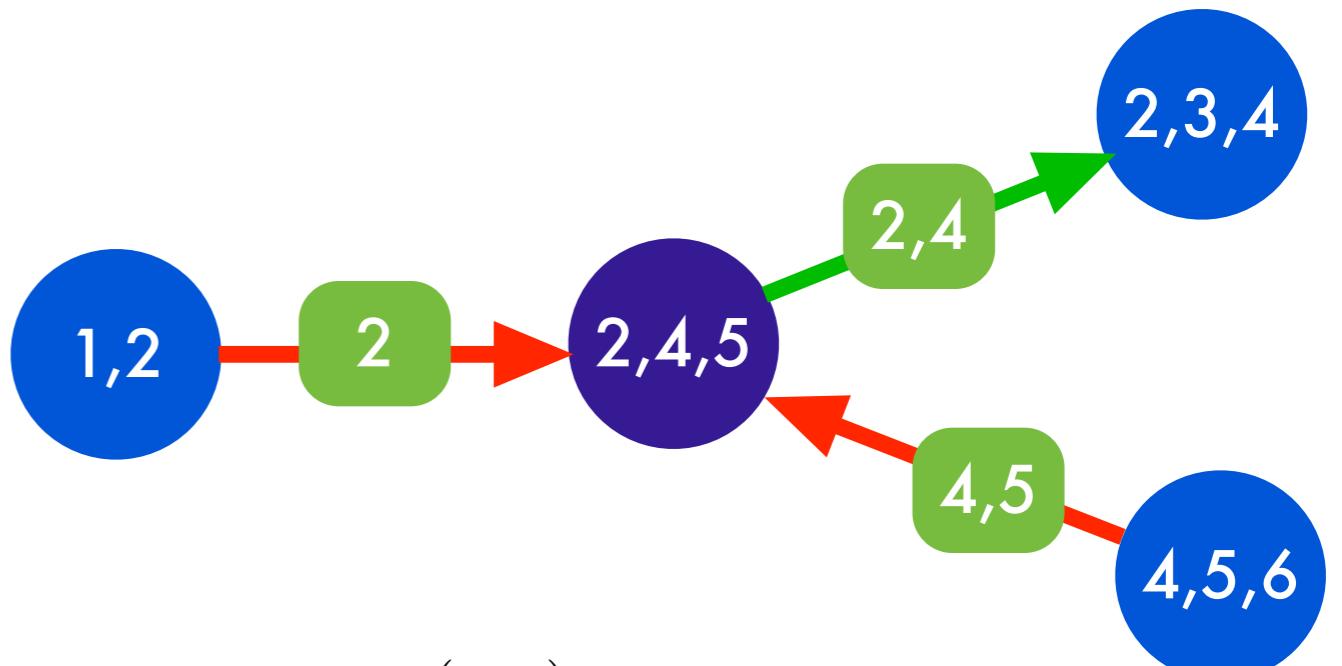
clique
potential

separator
set

Junction Trees



dependency
graph



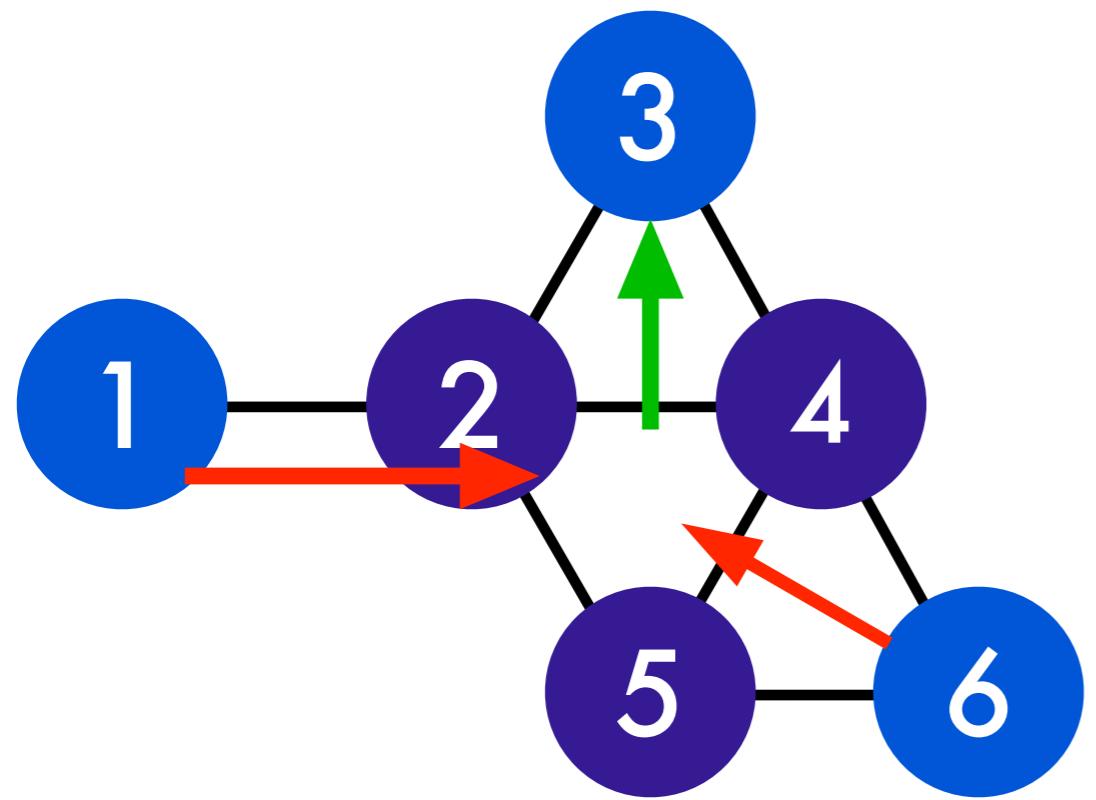
$$m_{245 \rightarrow 234}(x_{24})$$

$$= \sum_{x_5} f(x_{245}) m_{12 \rightarrow 245}(x_2) m_{457 \rightarrow 245}(x_{45})$$

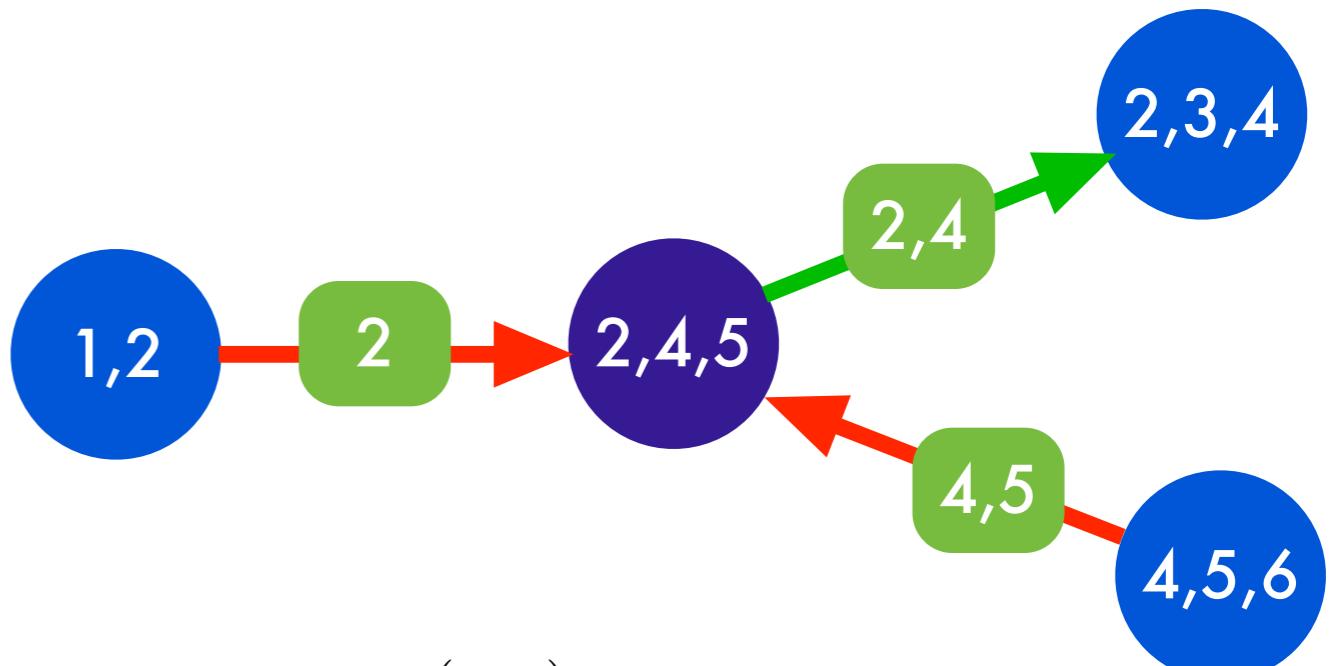
clique
potential

separator
set

Junction Trees



dependency
graph



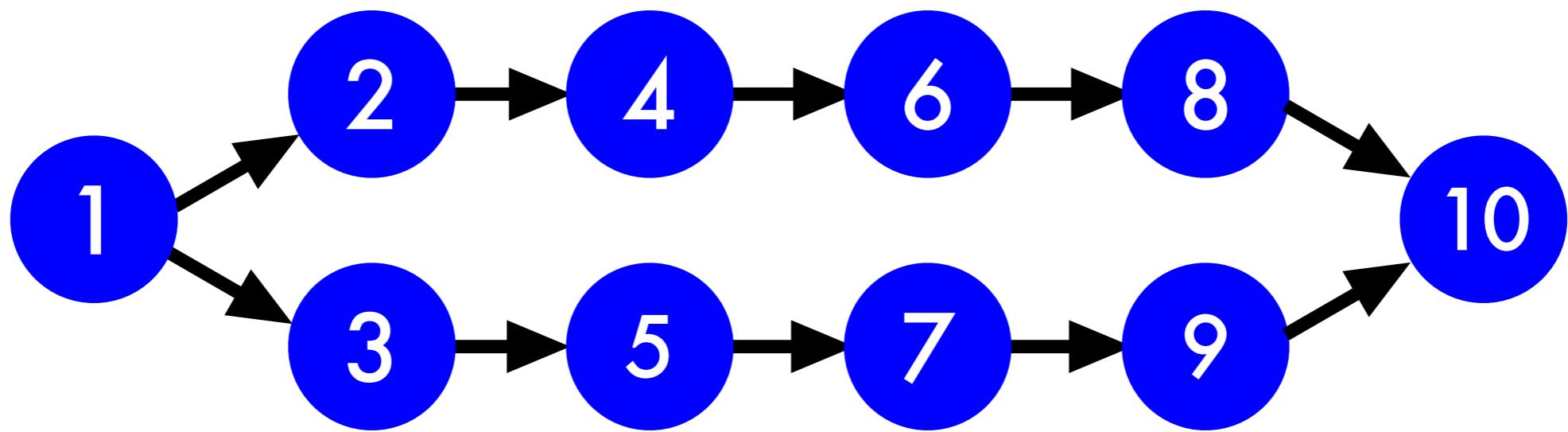
$$m_{245 \rightarrow 234}(x_{24})$$

$$= \sum_{x_5} f(x_{245}) m_{12 \rightarrow 245}(x_2) m_{457 \rightarrow 245}(x_{45})$$

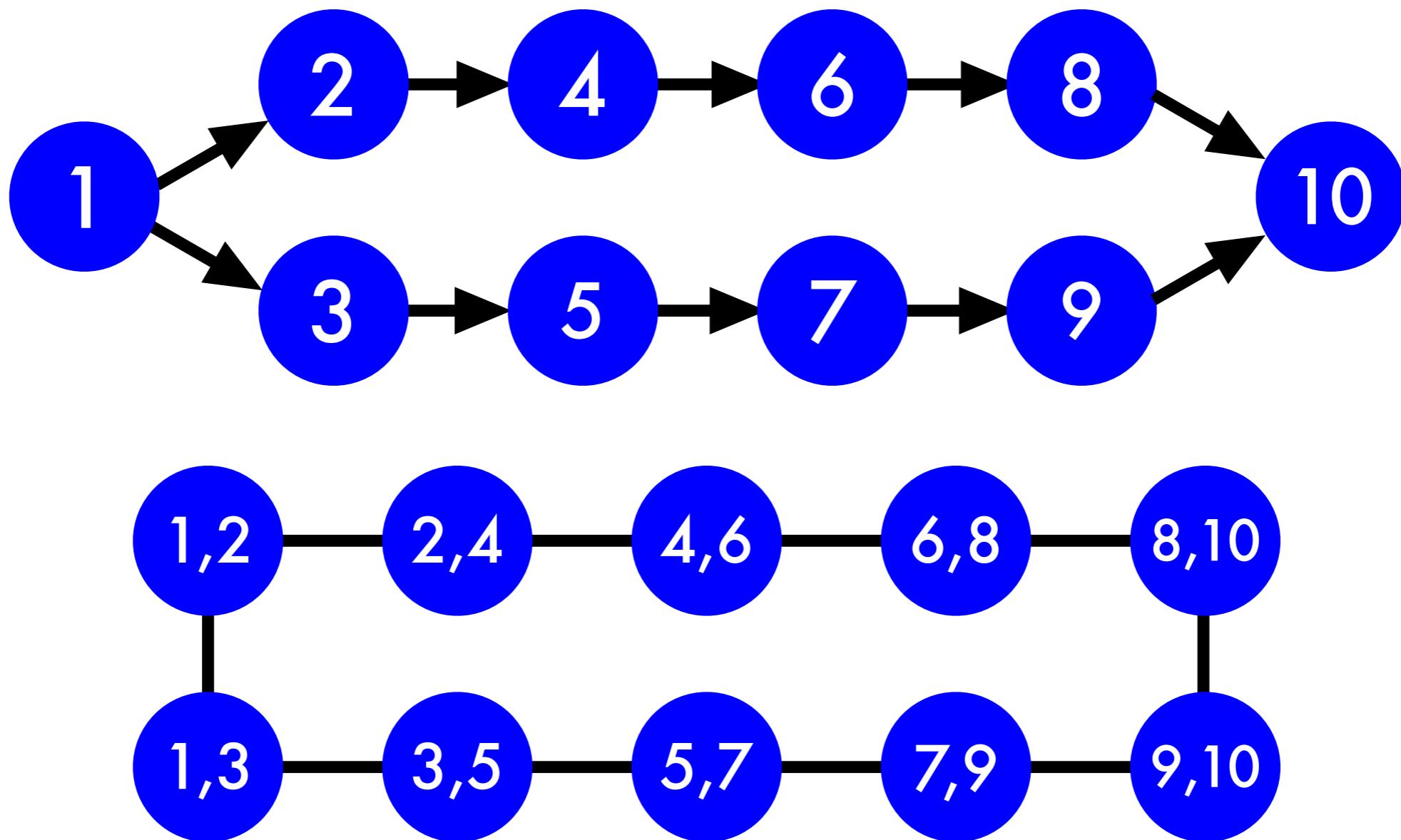
clique
potential

separator
set

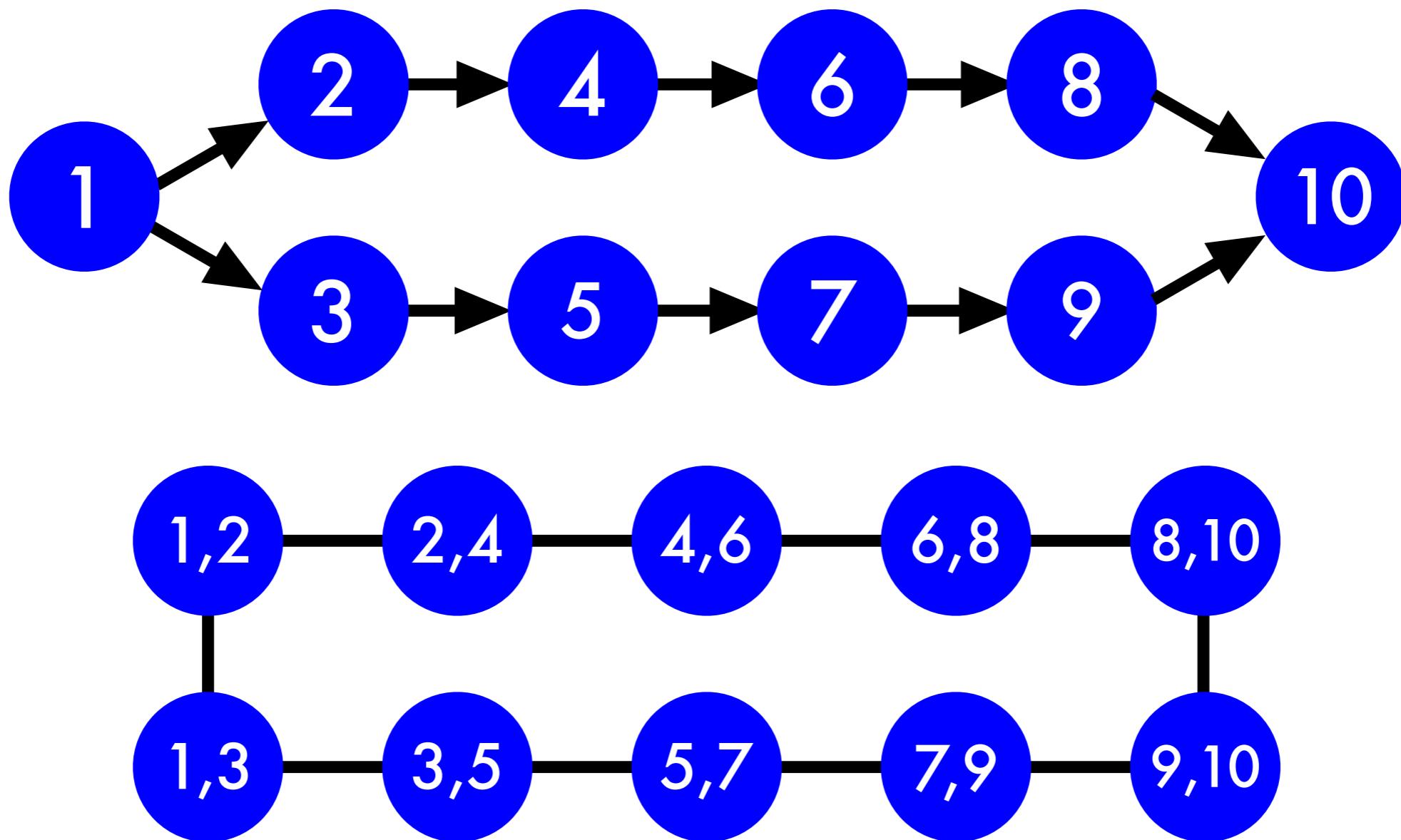
Caution



Caution

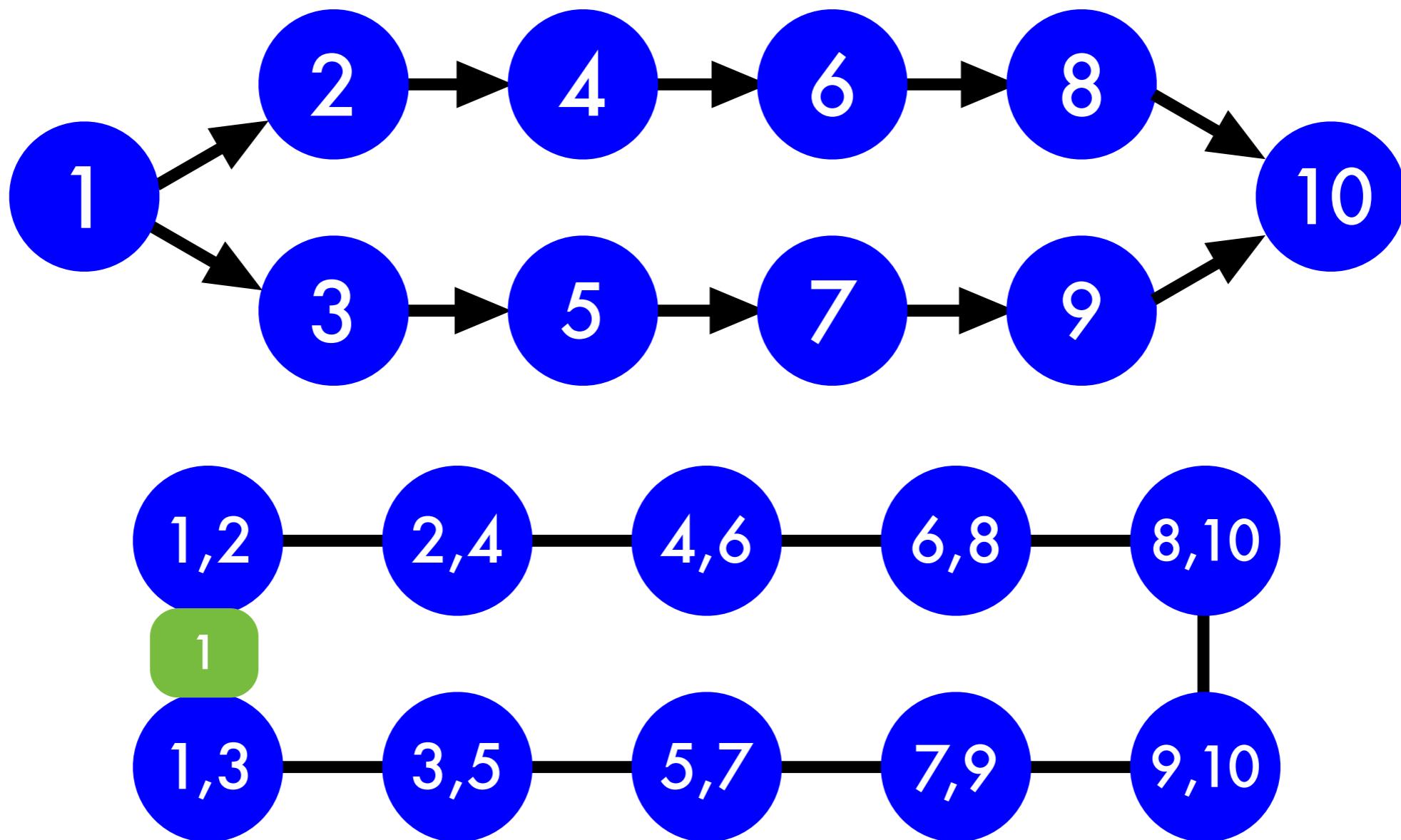


Caution



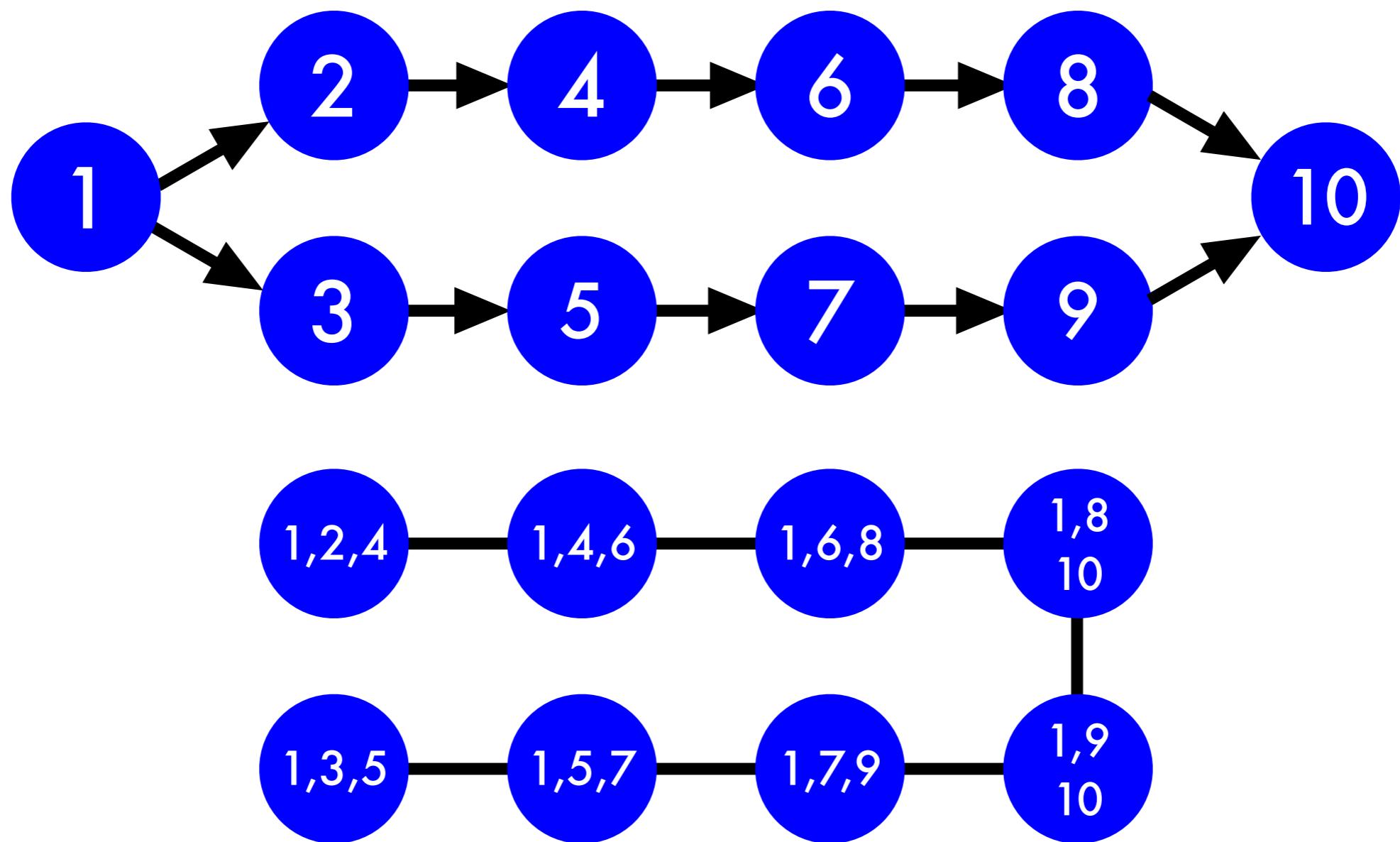
This is not a tree

Caution

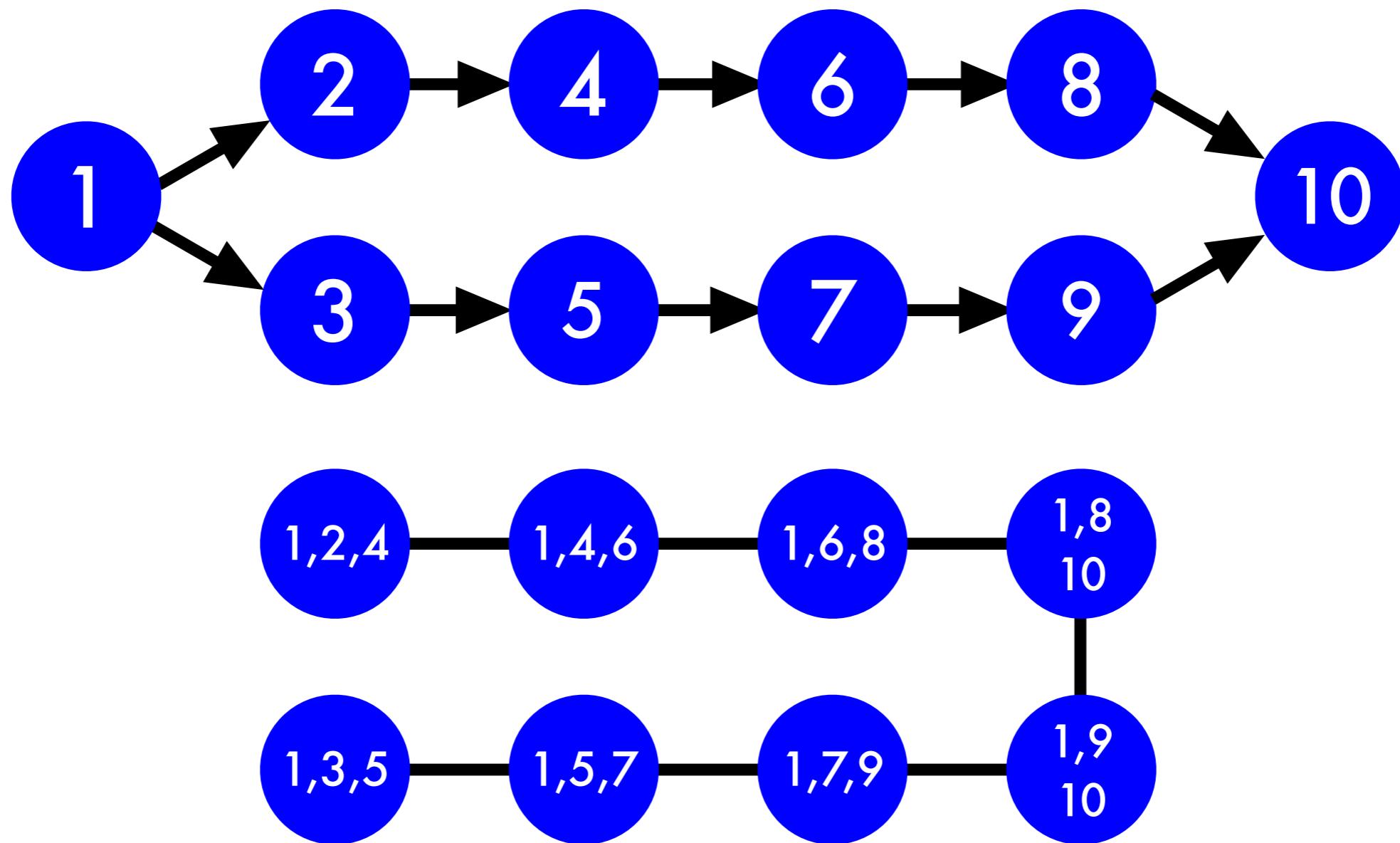


This is not a tree

Graph triangulation

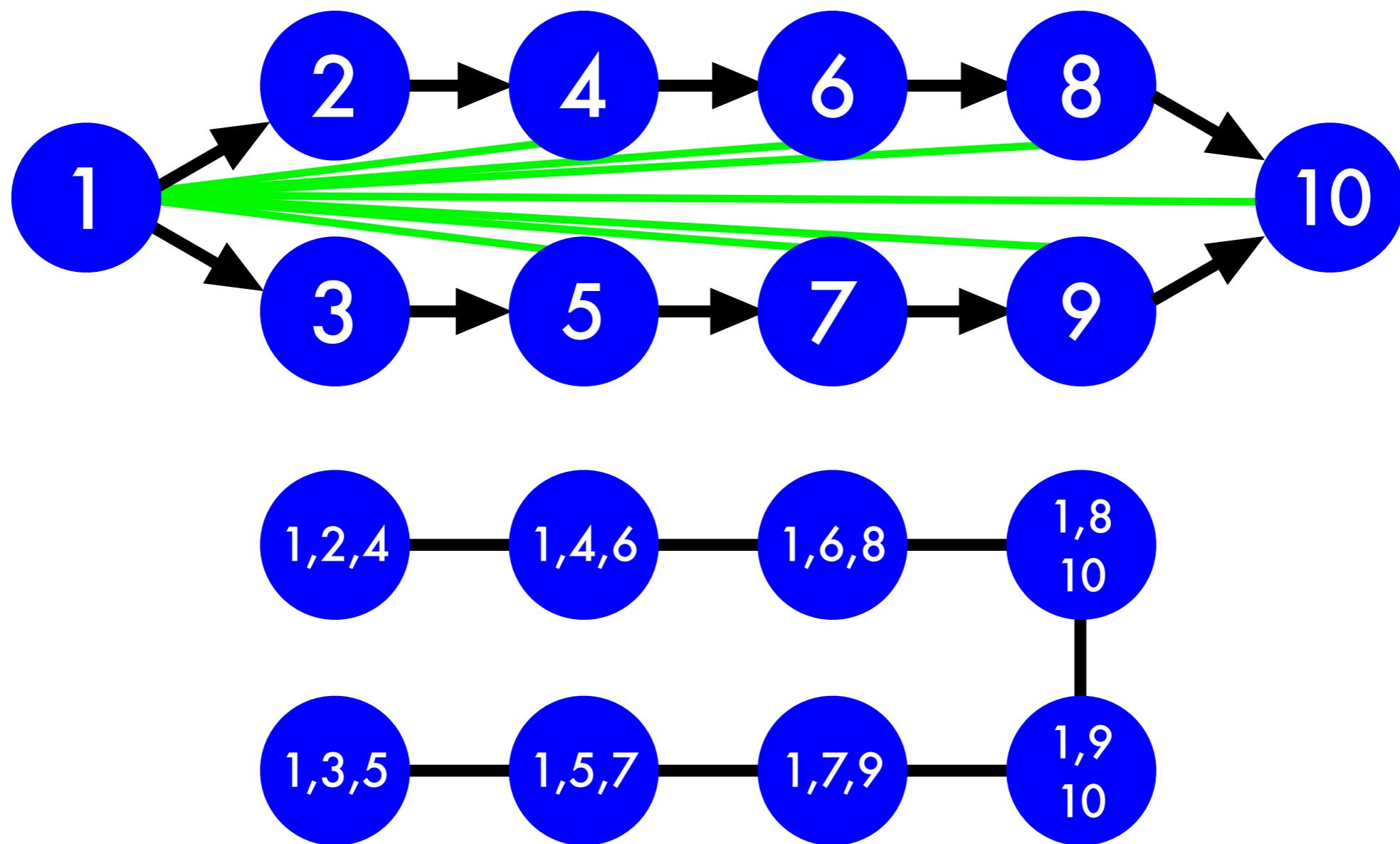


Graph triangulation



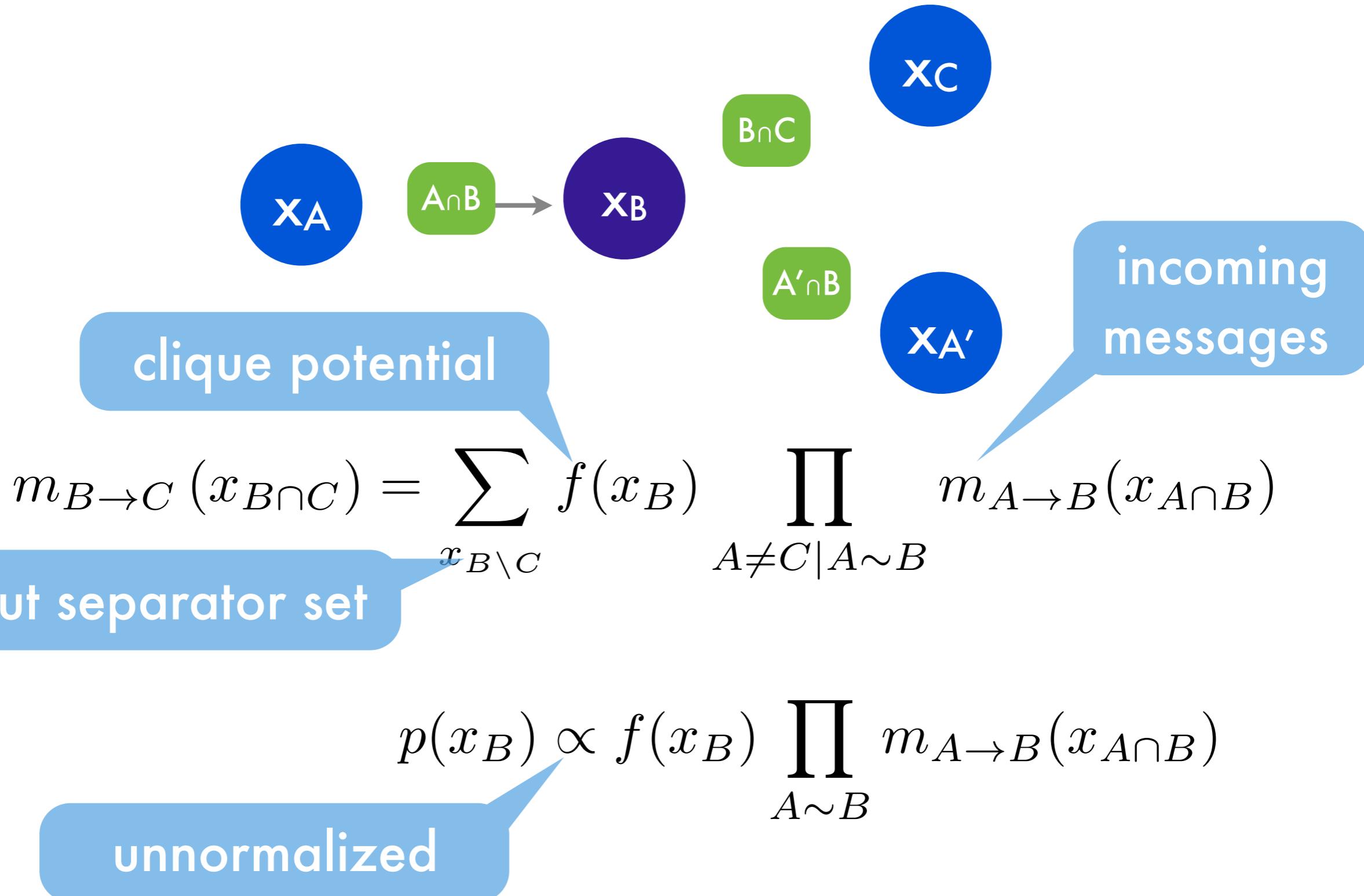
Separator set increases

Graph triangulation

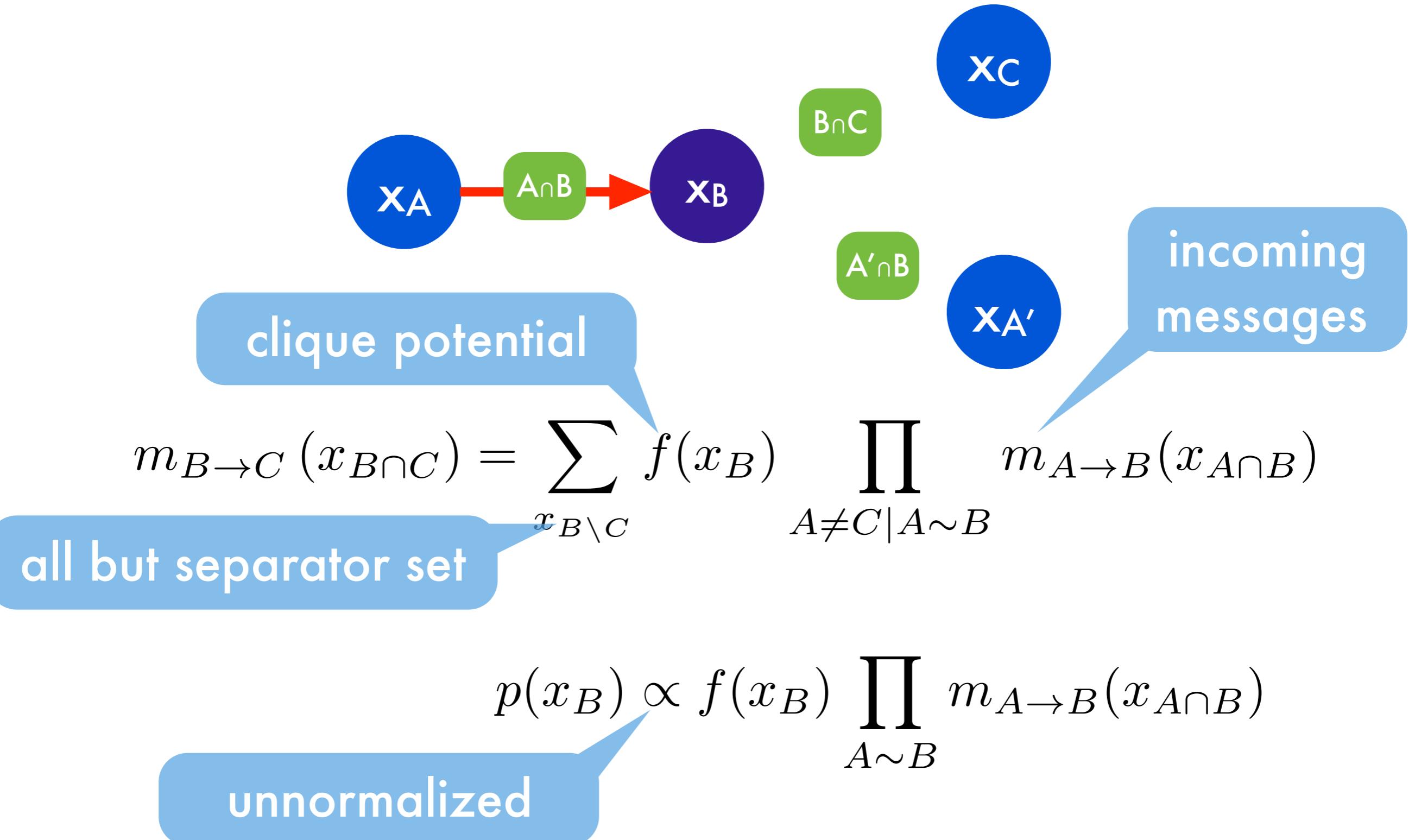


Separator set increases

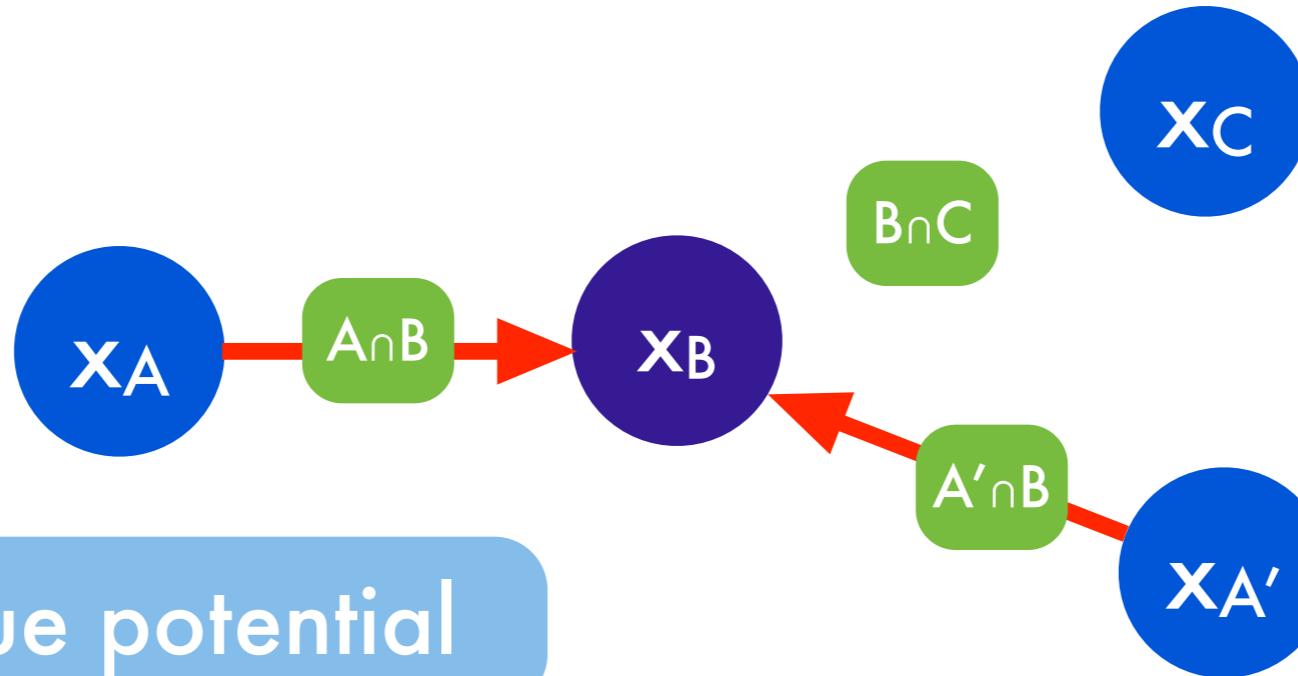
Update equations



Update equations



Update equations



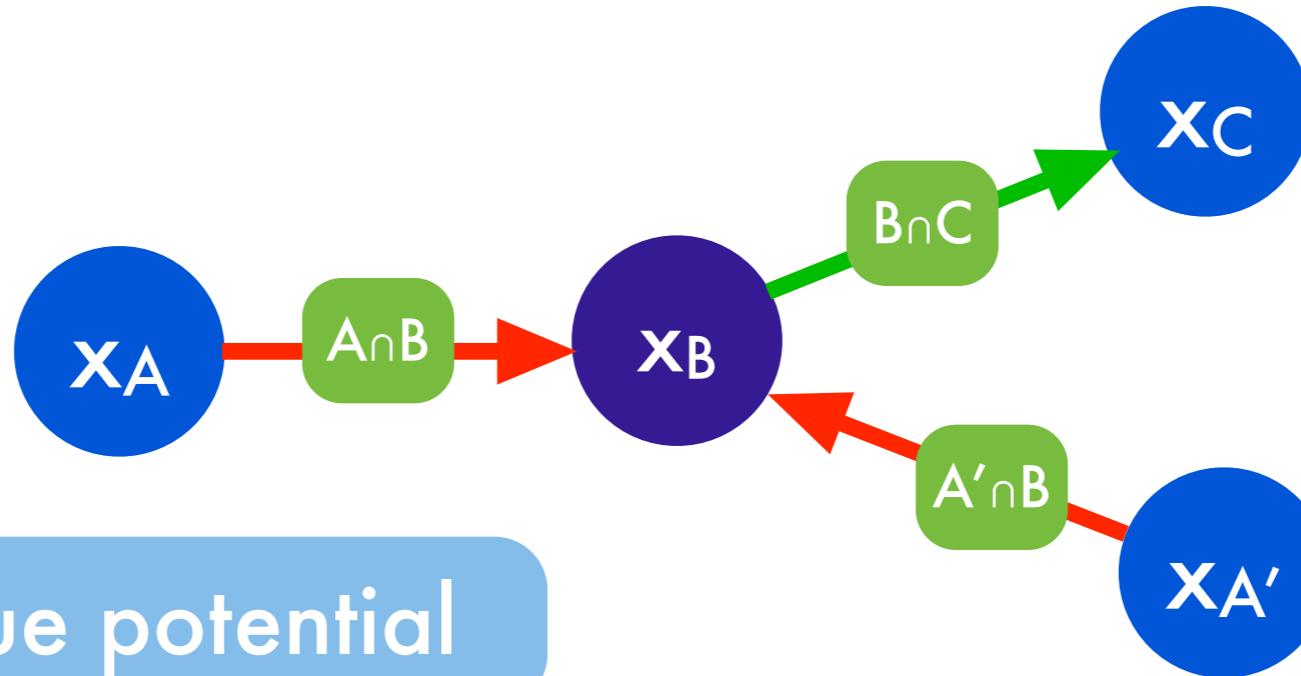
$$m_{B \rightarrow C}(x_{B \cap C}) = \sum_{x_{B \setminus C}} f(x_B) \prod_{A \neq C | A \sim B} m_{A \rightarrow B}(x_{A \cap B})$$

all but separator set

$$p(x_B) \propto f(x_B) \prod_{A \sim B} m_{A \rightarrow B}(x_{A \cap B})$$

unnormalized

Update equations



$$m_{B \rightarrow C}(x_{B \cap C}) = \sum_{x_{B \setminus C}} f(x_B) \prod_{A \neq C | A \sim B} m_{A \rightarrow B}(x_{A \cap B})$$

all but separator set

$$p(x_B) \propto f(x_B) \prod_{A \sim B} m_{A \rightarrow B}(x_{A \cap B})$$

unnormalized

Generalize Distributive Law



Generalized Distributive Law

- Key Idea

Dynamic programming uses only sums and multiplications, hence replace them with equivalent operations from other semirings

- Semiring

- ‘addition’ and ‘summation’ equivalent

- **Associative law** $(a + b) + c = a + (b + c)$

- **Distributive law** $a(b + c) = ab + ac$

Generalized Distributive Law

- Integrating out probabilities (sum, product)

$$a \cdot (b + c) = a \cdot b + a \cdot c$$

- Finding the maximum (max, +)

$$a + \max(b, c) = \max(a + b, a + c)$$

- Set algebra (union, intersection)

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

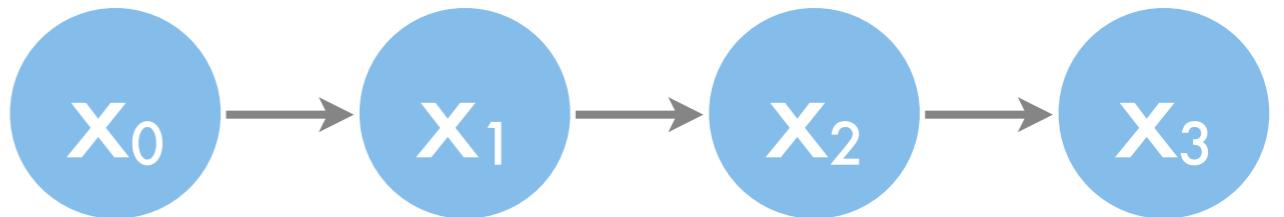
- Boolean semiring (AND, OR)

- Probability semiring ($\log +$, $+$)

- Tropical semiring (\min , $+$)

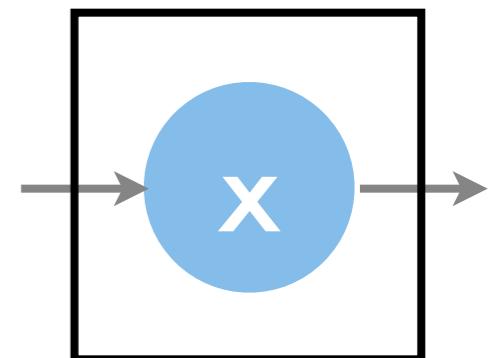
Chains ... again

$$\bar{s} = \max_x s(x_0) + \sum_{i=1}^{n-1} s(x_{i+1}|x_i)$$



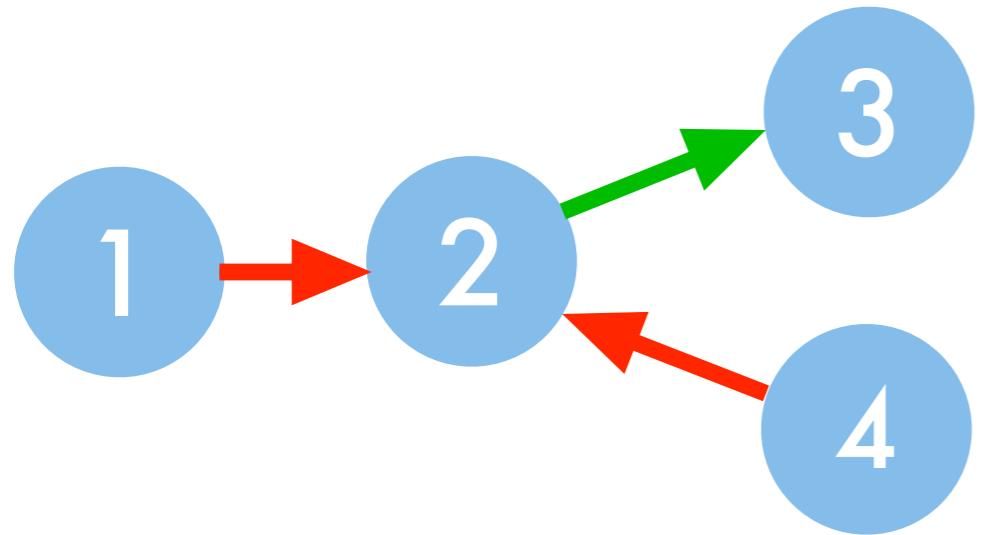
$$\bar{s} = \max_{x_0 \dots n} \underbrace{s(x_0)}_{:=l_0(x_0)} + \sum_{j=1}^n s(x_j|x_{j-1})$$

$$= \max_{x_1 \dots n} \underbrace{\max_{x_0} [l_0(x_0)s(x_1|x_0)]}_{:=l_1(x_1)} + \sum_{j=2}^n s(x_j|x_{j-1})$$



$$= \max_{x_2 \dots n} \underbrace{\max_{x_1} [l_1(x_1)s(x_2|x_1)]}_{:=l_2(x_2)} + \sum_{j=3}^n s(x_j|x_{j-1})$$

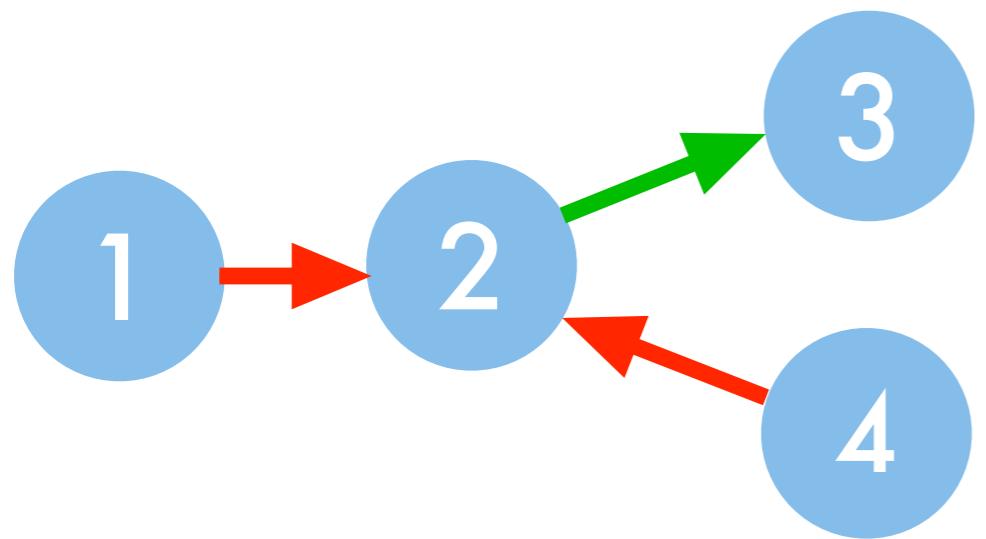
Junction Trees



$$m_{i \rightarrow j}(x_j) = \max_{x_i} f(x_i, x_j) + \sum_{l \neq j} m_{l \rightarrow i}(x_j)$$

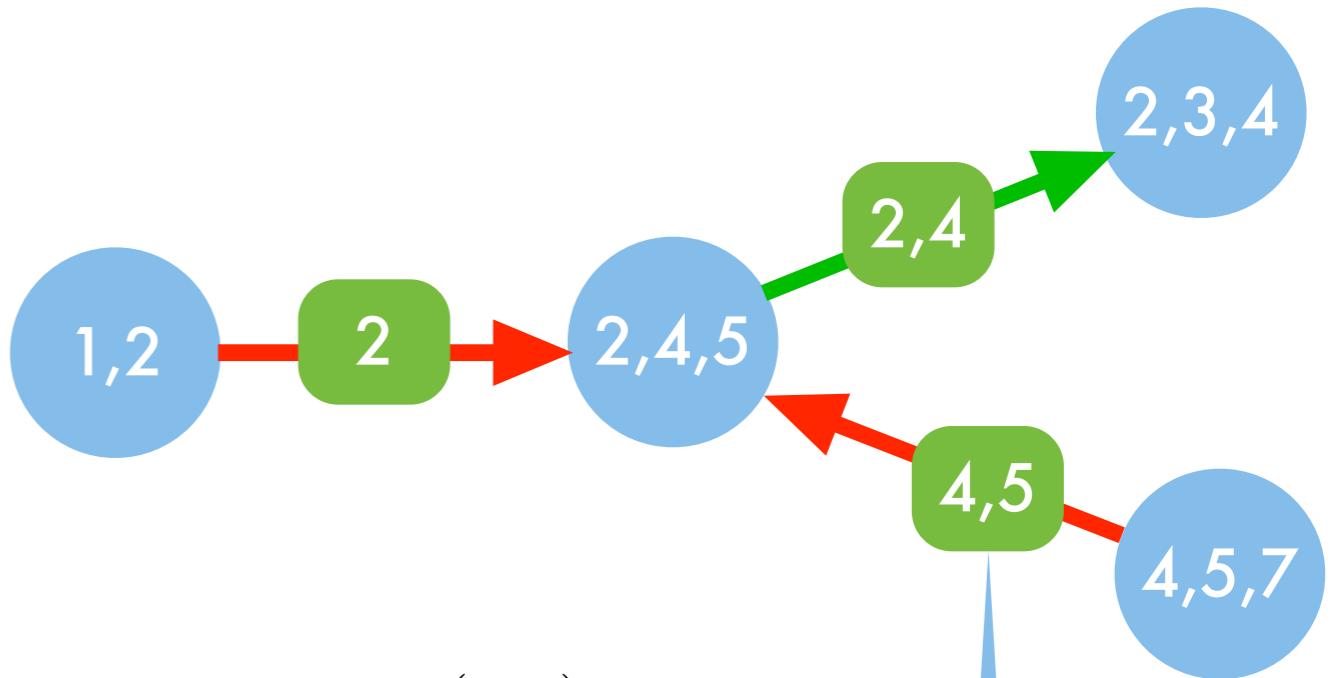
clique
potential

Junction Trees



$$m_{i \rightarrow j}(x_j) = \max_{x_i} f(x_i, x_j) + \sum_{l \neq j} m_{l \rightarrow i}(x_j)$$

clique
potential



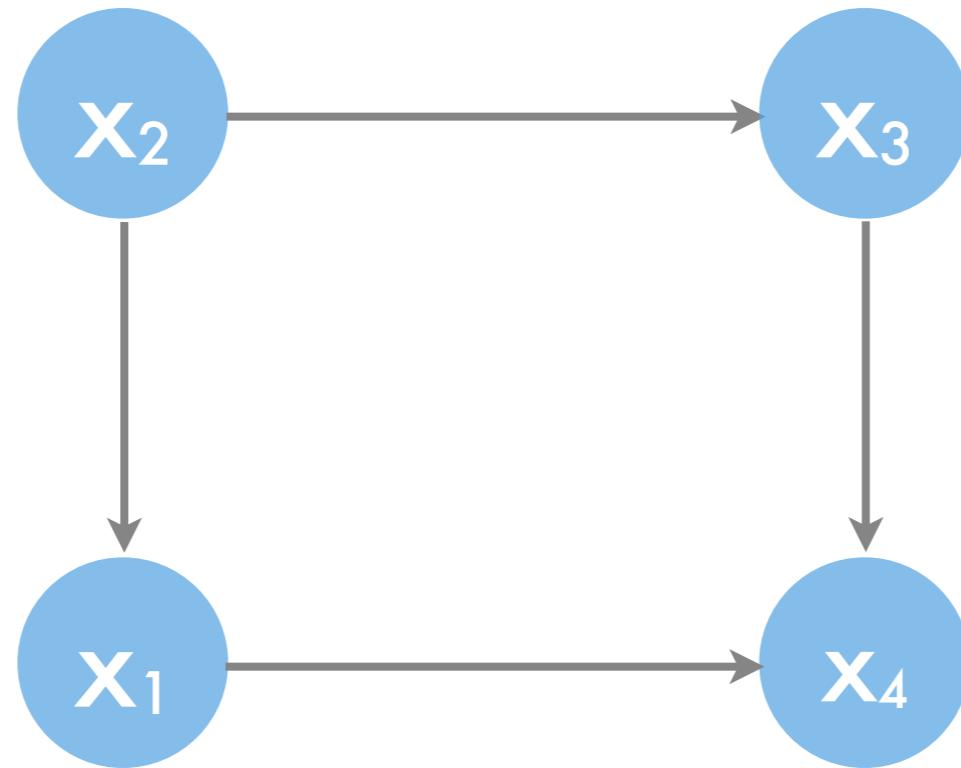
$$\begin{aligned} m_{245 \rightarrow 234}(x_{24}) \\ = \max_{x_5} f(x_{245}) + m_{12 \rightarrow 245}(x_2) + m_{457 \rightarrow 245}(x_{45}) \end{aligned}$$

clique
potential

separator
set

No loops allowed

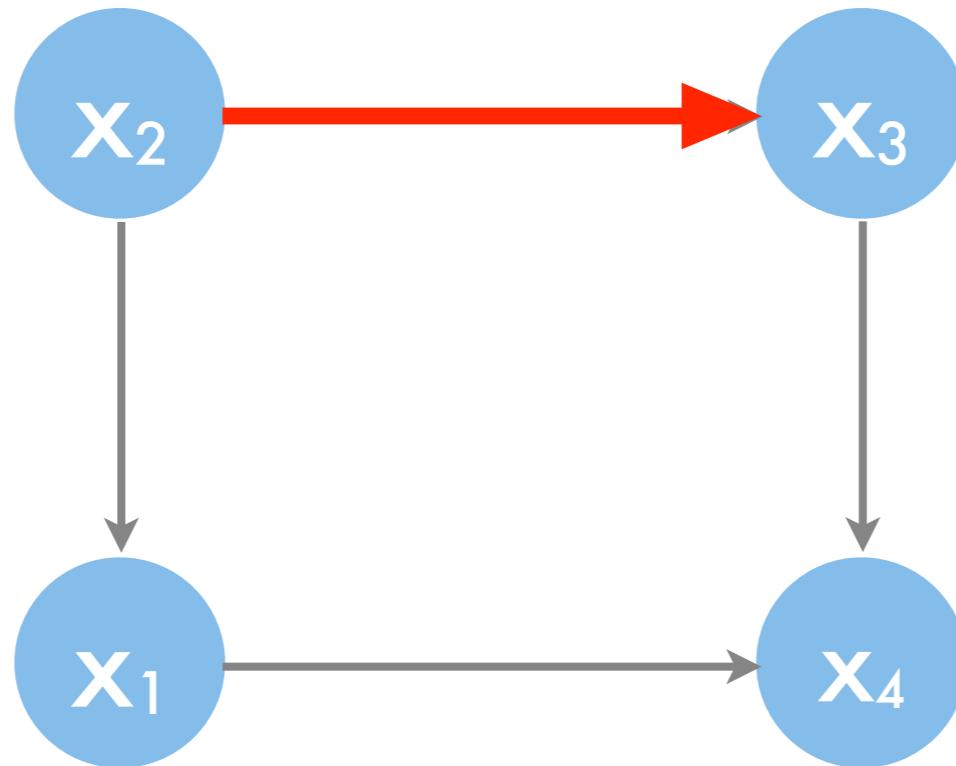
$$s(x_1, x_2) + s(x_2, x_3) + s(x_3, x_4) + s(x_4, x_1)$$



Often use it anyway – Loopy Belief Propagation
(Turbo Codes, Markov Random Fields, etc.)

No loops allowed

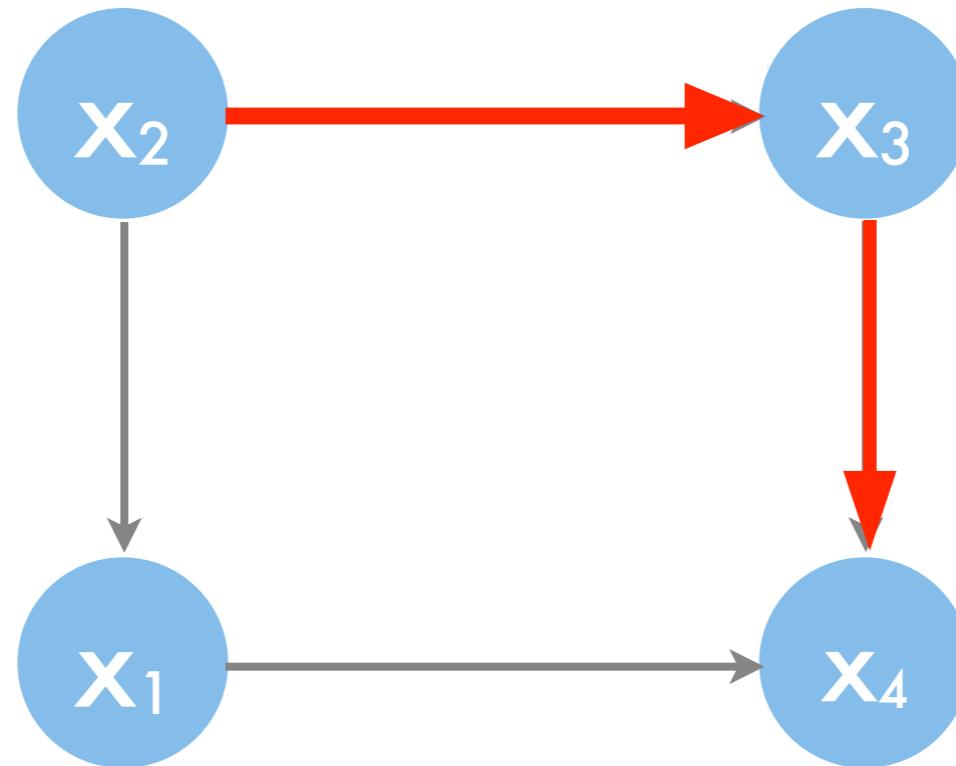
$$s(x_1, x_2) + s(x_2, x_3) + s(x_3, x_4) + s(x_4, x_1)$$



Often use it anyway – Loopy Belief Propagation
(Turbo Codes, Markov Random Fields, etc.)

No loops allowed

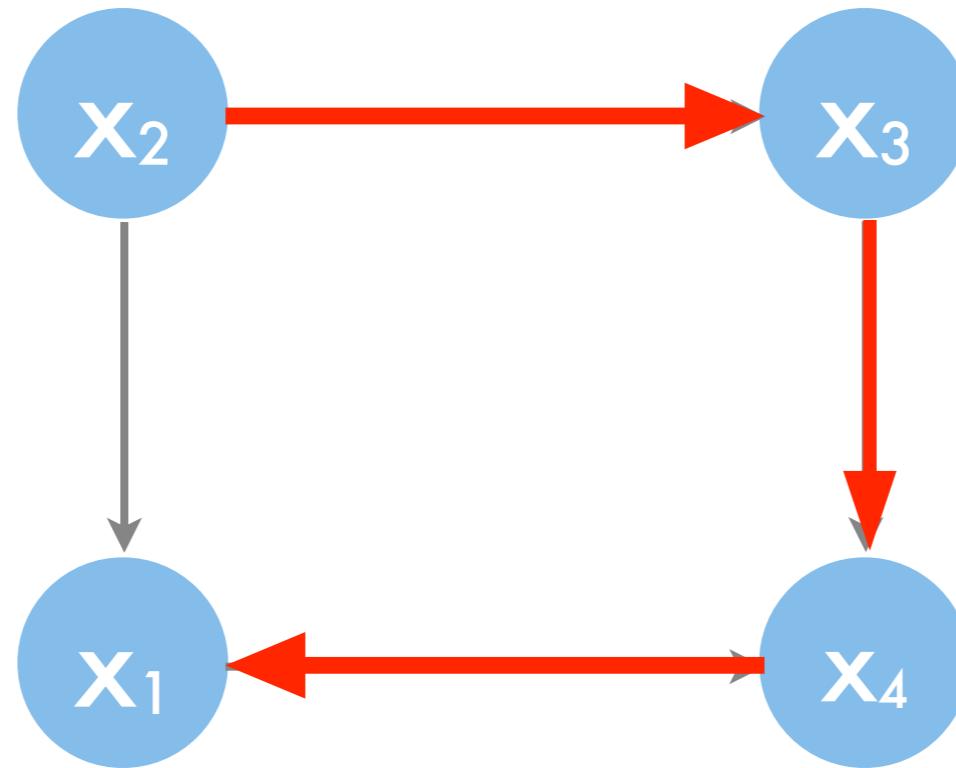
$$s(x_1, x_2) + s(x_2, x_3) + s(x_3, x_4) + s(x_4, x_1)$$



Often use it anyway – Loopy Belief Propagation
(Turbo Codes, Markov Random Fields, etc.)

No loops allowed

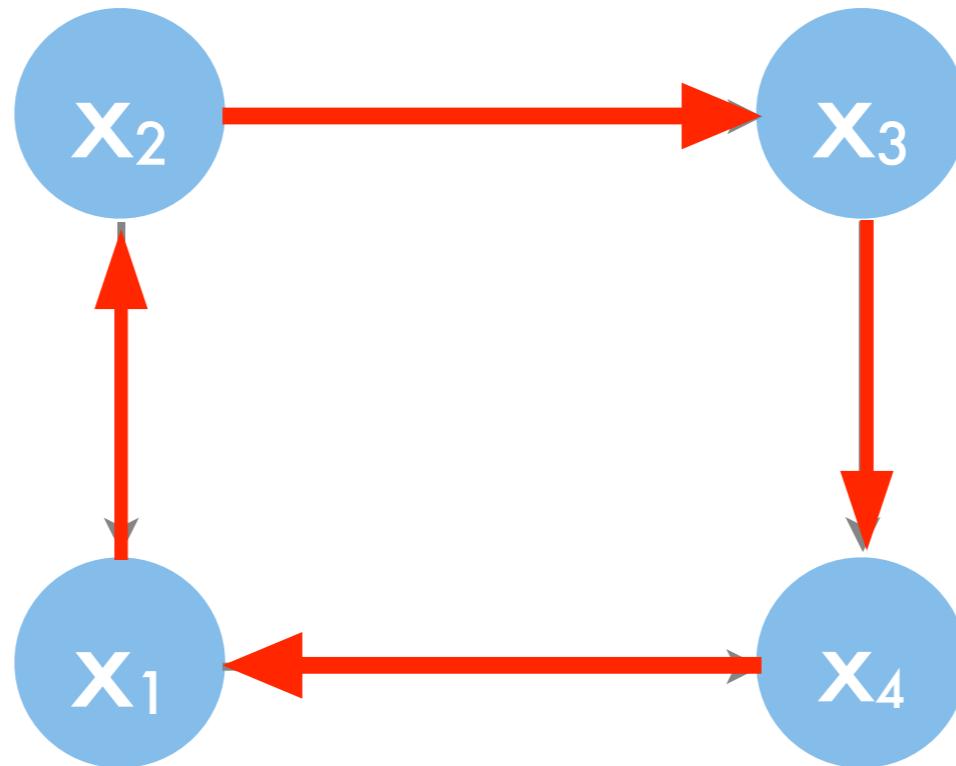
$$s(x_1, x_2) + s(x_2, x_3) + s(x_3, x_4) + s(x_4, x_1)$$



Often use it anyway – Loopy Belief Propagation
(Turbo Codes, Markov Random Fields, etc.)

No loops allowed

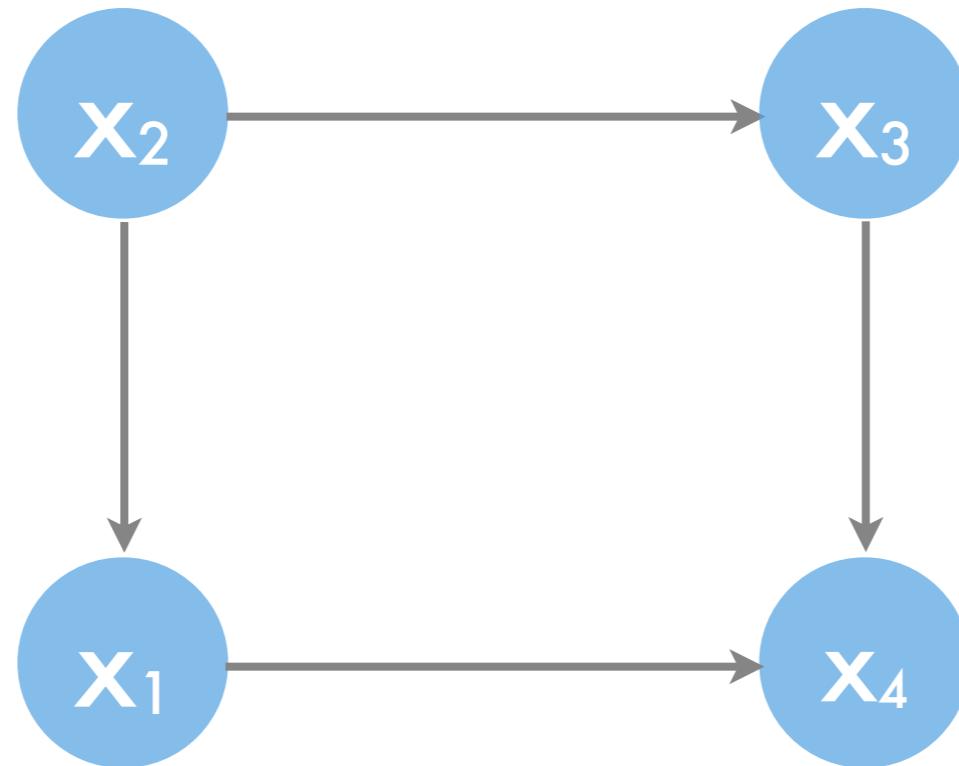
$$s(x_1, x_2) + s(x_2, x_3) + s(x_3, x_4) + s(x_4, x_1)$$



Often use it anyway – Loopy Belief Propagation
(Turbo Codes, Markov Random Fields, etc.)

No loops allowed

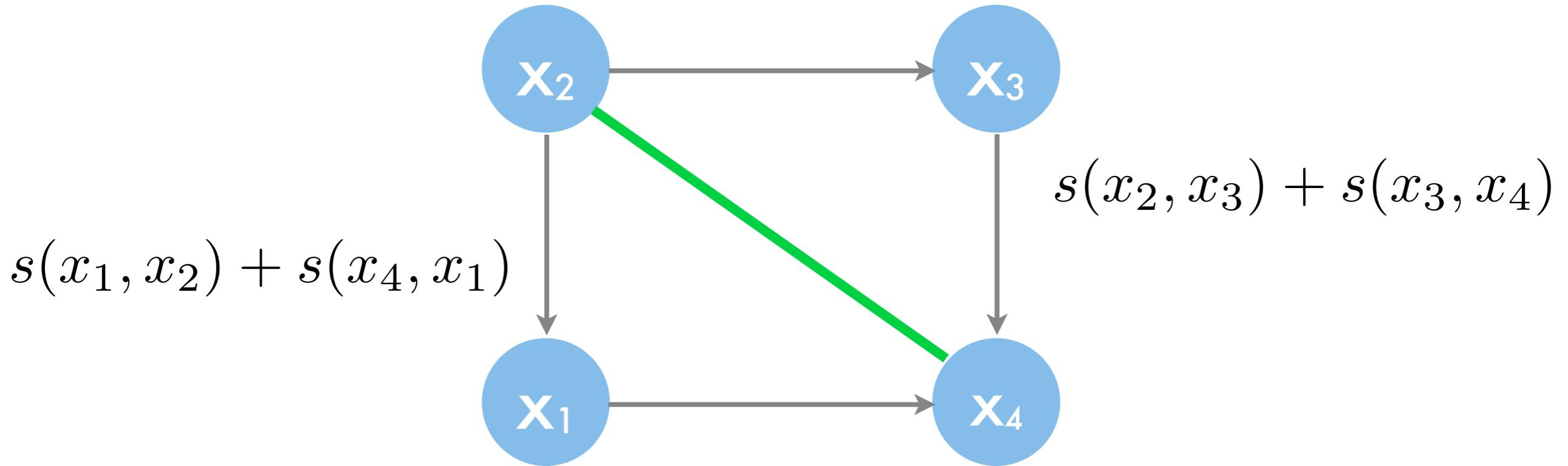
$$s(x_1, x_2) + s(x_2, x_3) + s(x_3, x_4) + s(x_4, x_1)$$



Often use it anyway – Loopy Belief Propagation
(Turbo Codes, Markov Random Fields, etc.)

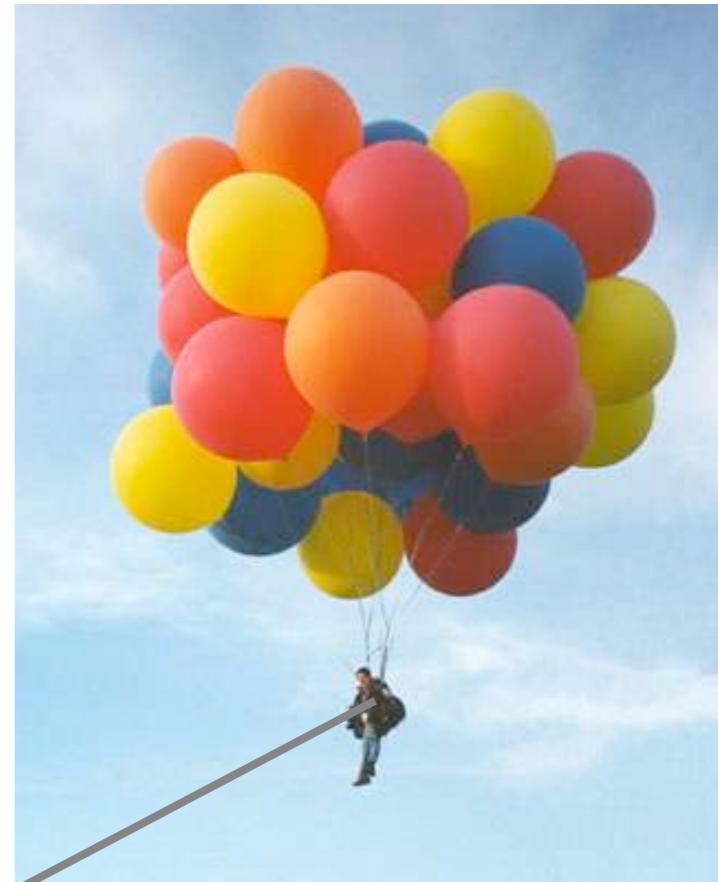
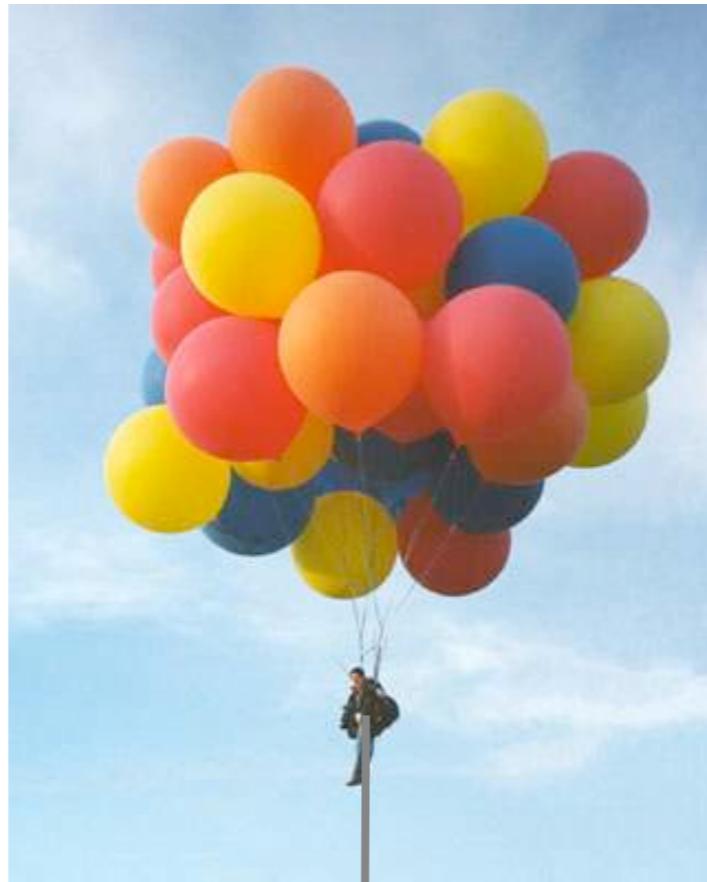
No loops allowed

$$s(x_1, x_2) + s(x_2, x_3) + s(x_3, x_4) + s(x_4, x_1)$$



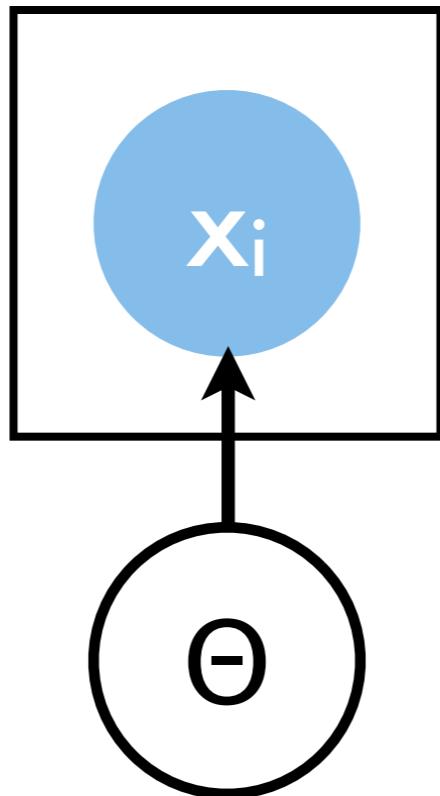
Often use it anyway – Loopy Belief Propagation
(Turbo Codes, Markov Random Fields, etc.)

Clustering



Basic Idea

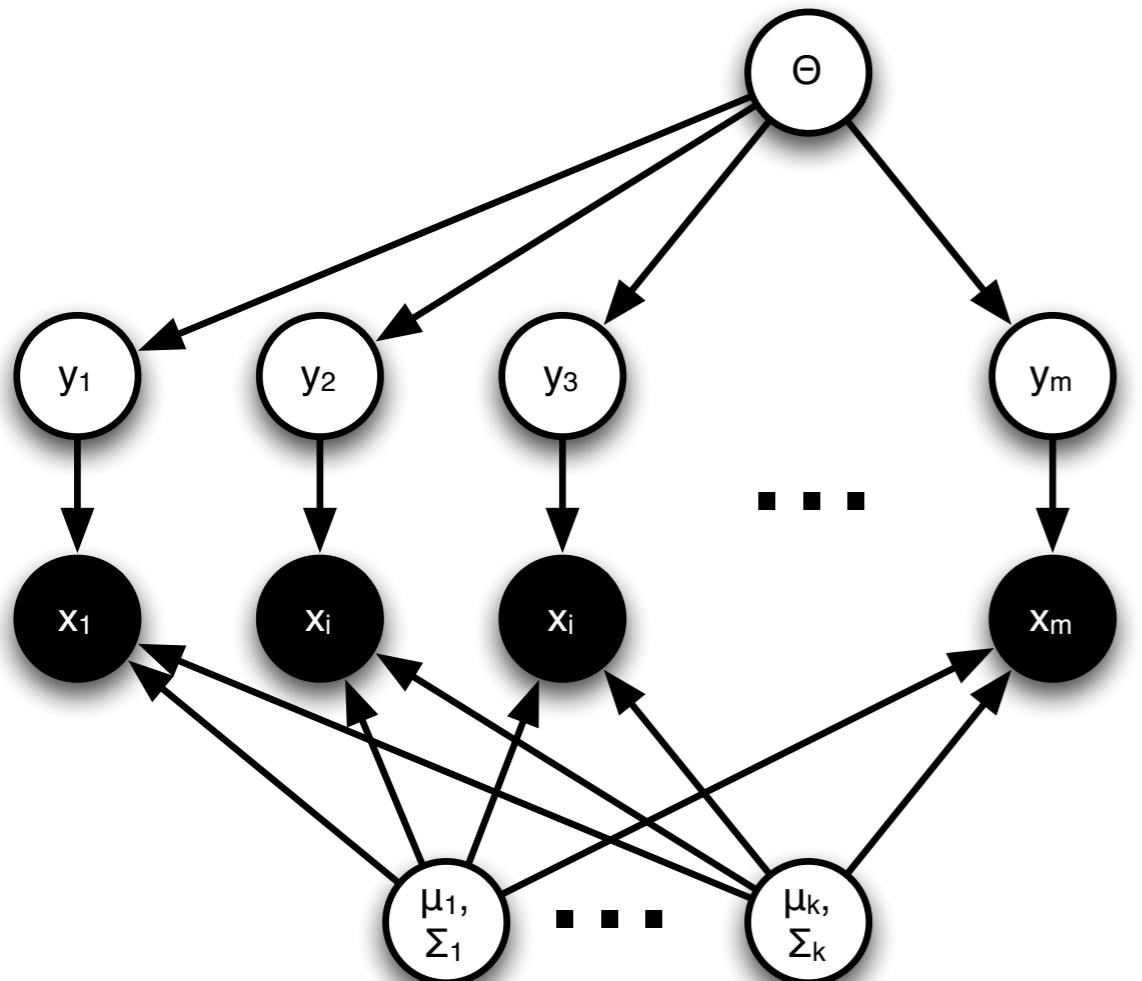
Density Estimation



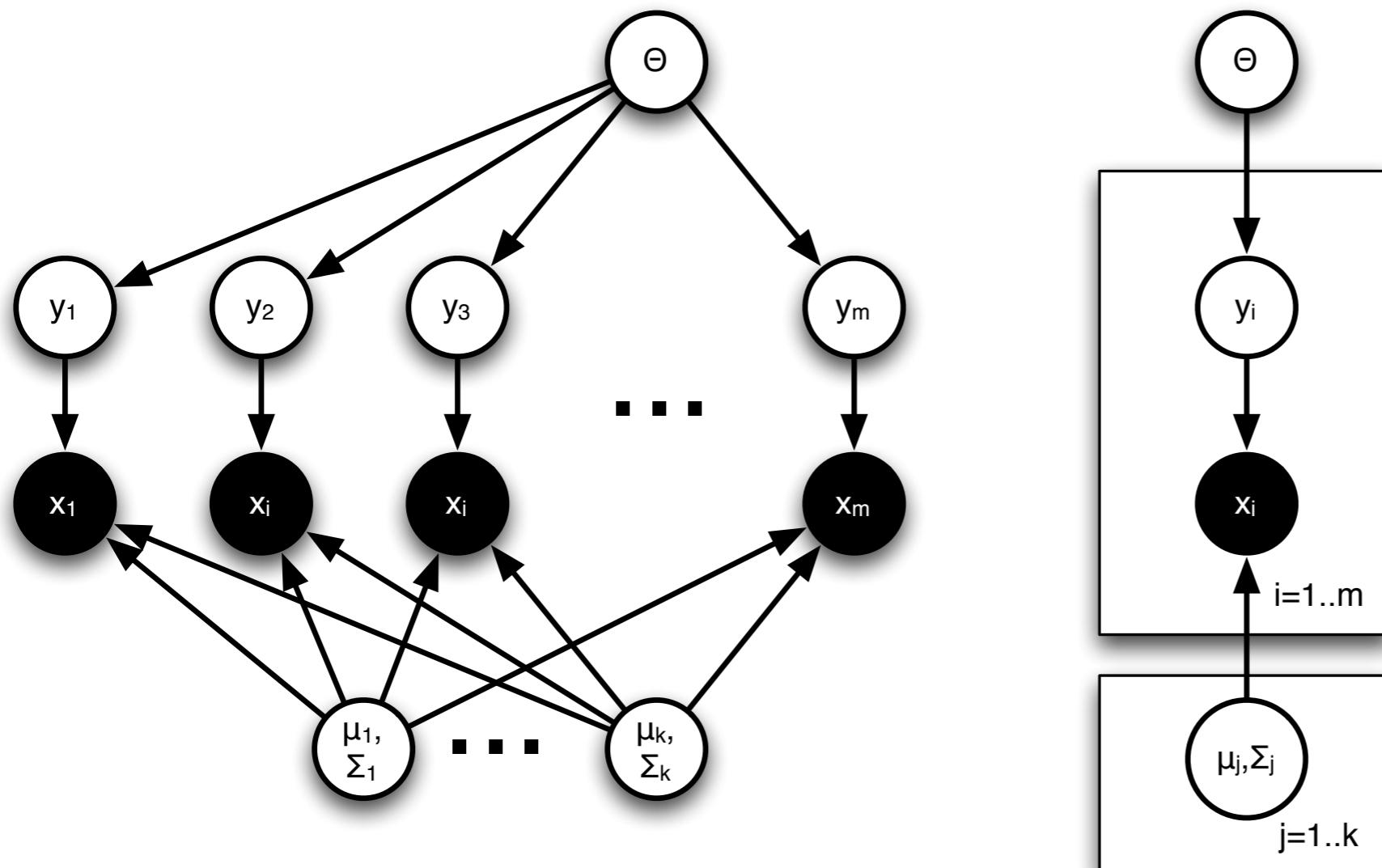
$$p(X|\theta) = \prod_{i=1}^m p(x_i|\theta)$$

- Draw latent parameter Θ
- For all i draw observed x_i given Θ
- What if the basic model doesn't fit all data?

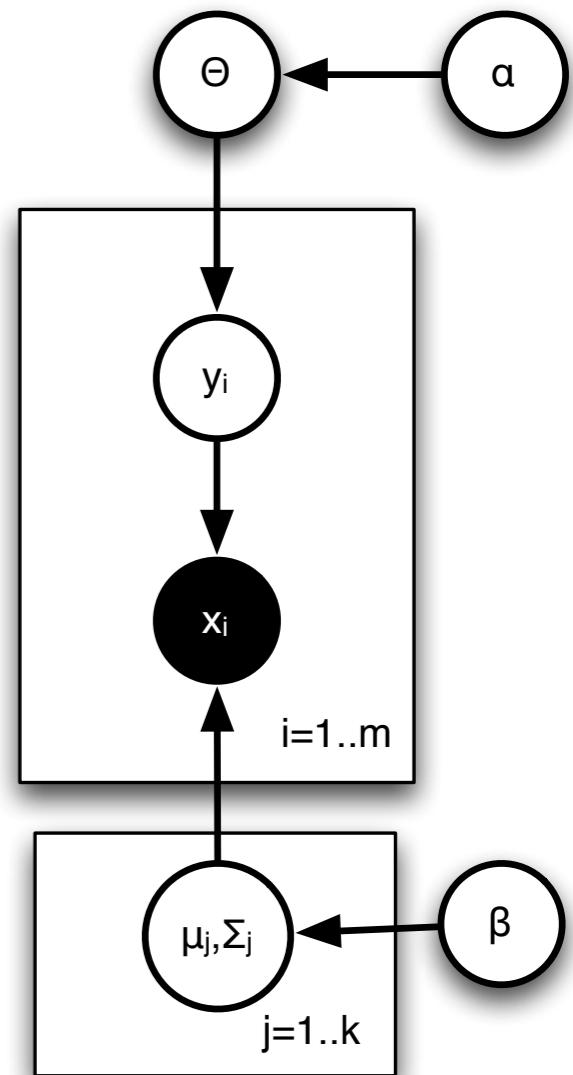
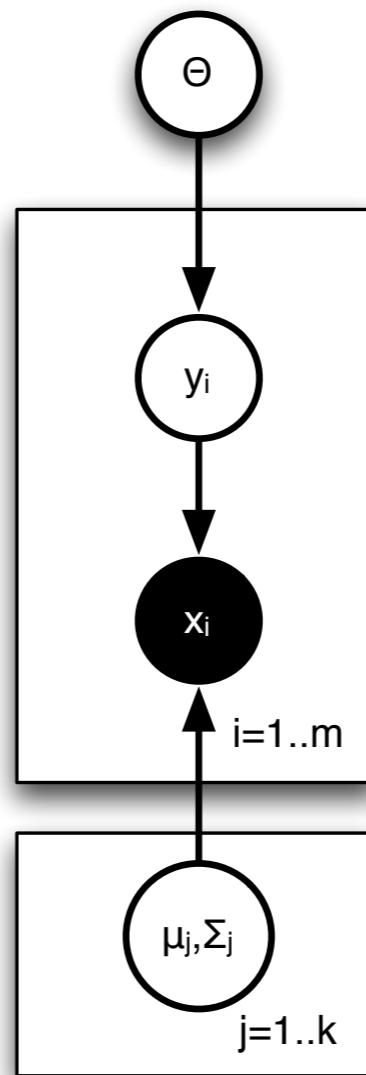
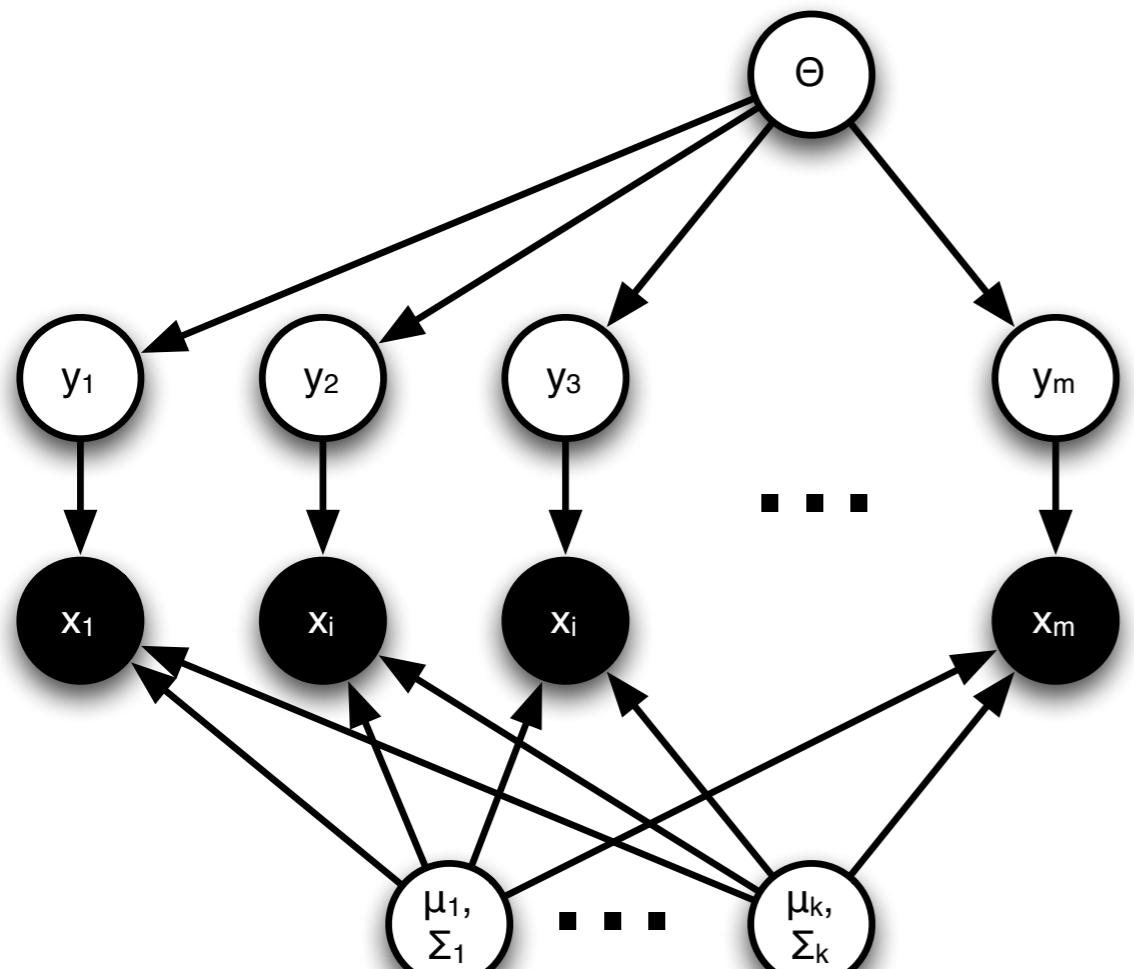
One size doesn't fit all



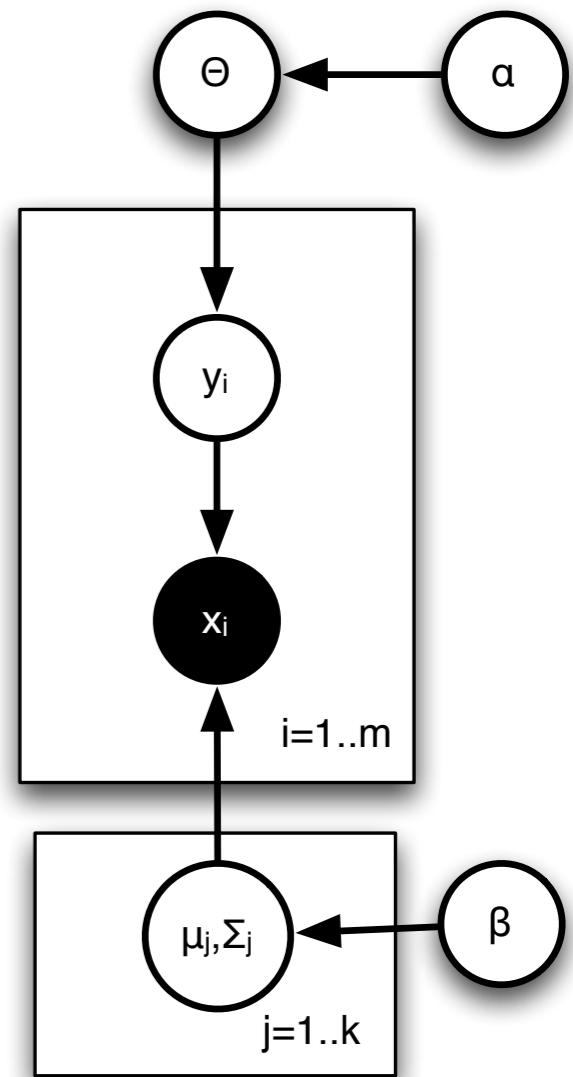
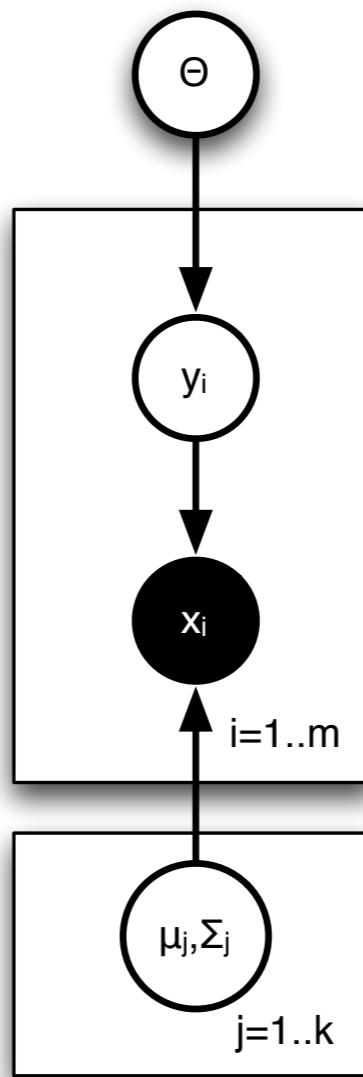
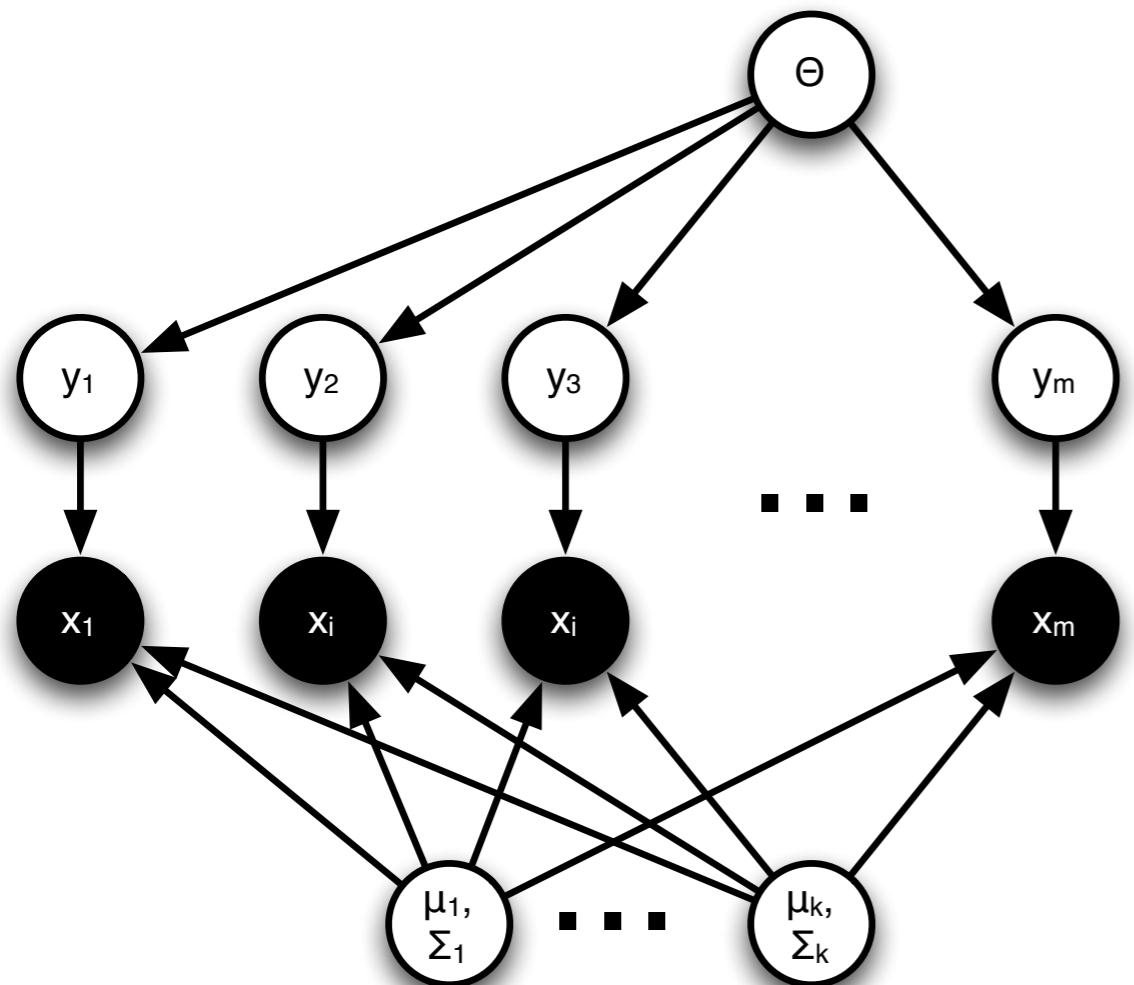
One size doesn't fit all



One size doesn't fit all



One size doesn't fit all



$$p(X, Y | \theta, \sigma, \mu) = \prod_{i=1}^n p(x_i | y_i, \sigma, \mu) p(y_i | \theta)$$

What can we cluster?

What can we cluster?

The diagram illustrates various entities that can be clustered, arranged in a grid-like structure:

- Row 1:** mails, text, urls, products
- Row 2:** news, users
- Row 3:** queries, locations
- Row 4:** spammers, ads, events
- Row 5:** abuse

Mixture of Gaussians

- Draw cluster label y from discrete distribution
- Draw data x from Gaussian for given cluster y
- Prior for discrete distribution - Dirichlet
- Prior for Gaussians - Gauss-Wishart distribution
- Problem: we don't know y
 - If we knew the parameters we could get y
 - If we knew y we could get the parameters

k-means

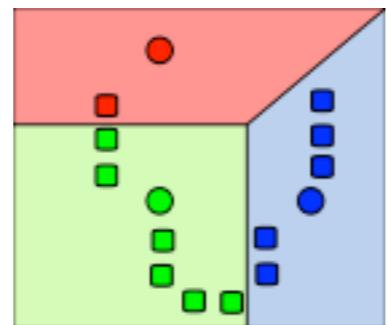
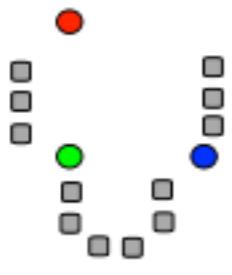
- Fixed uniform variance for all Gaussians
- Fixed uniform distribution over clusters
- Initialize centers with random subset of points
- Find most likely cluster y for x
$$y_i = \operatorname*{argmax}_y p(x_i|y, \sigma, \mu)$$
- Find most likely center for given cluster
$$\mu_y = \frac{1}{n_y} \sum_i \{y_i = y\} x_i$$
- Repeat until converged

k-means

- Pro
 - simple algorithm
 - can be implemented by MapReduce passes
- Con
 - no proper probabilistic representation
 - can get stuck easily in local minima

k-means

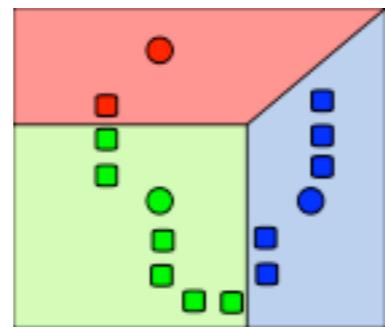
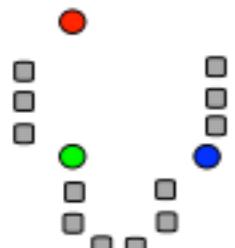
partitioning



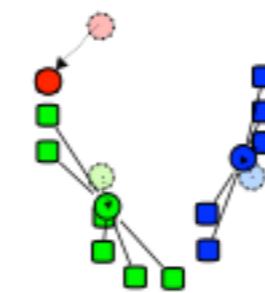
initialization

k-means

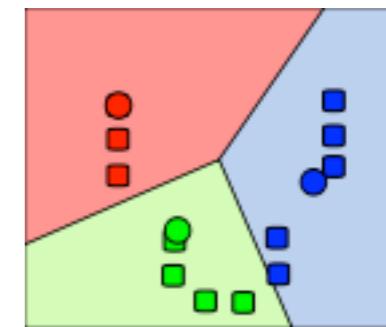
partitioning



initialization



partitioning



update

Variational Inference

Expectation Maximization

- Optimization Problem

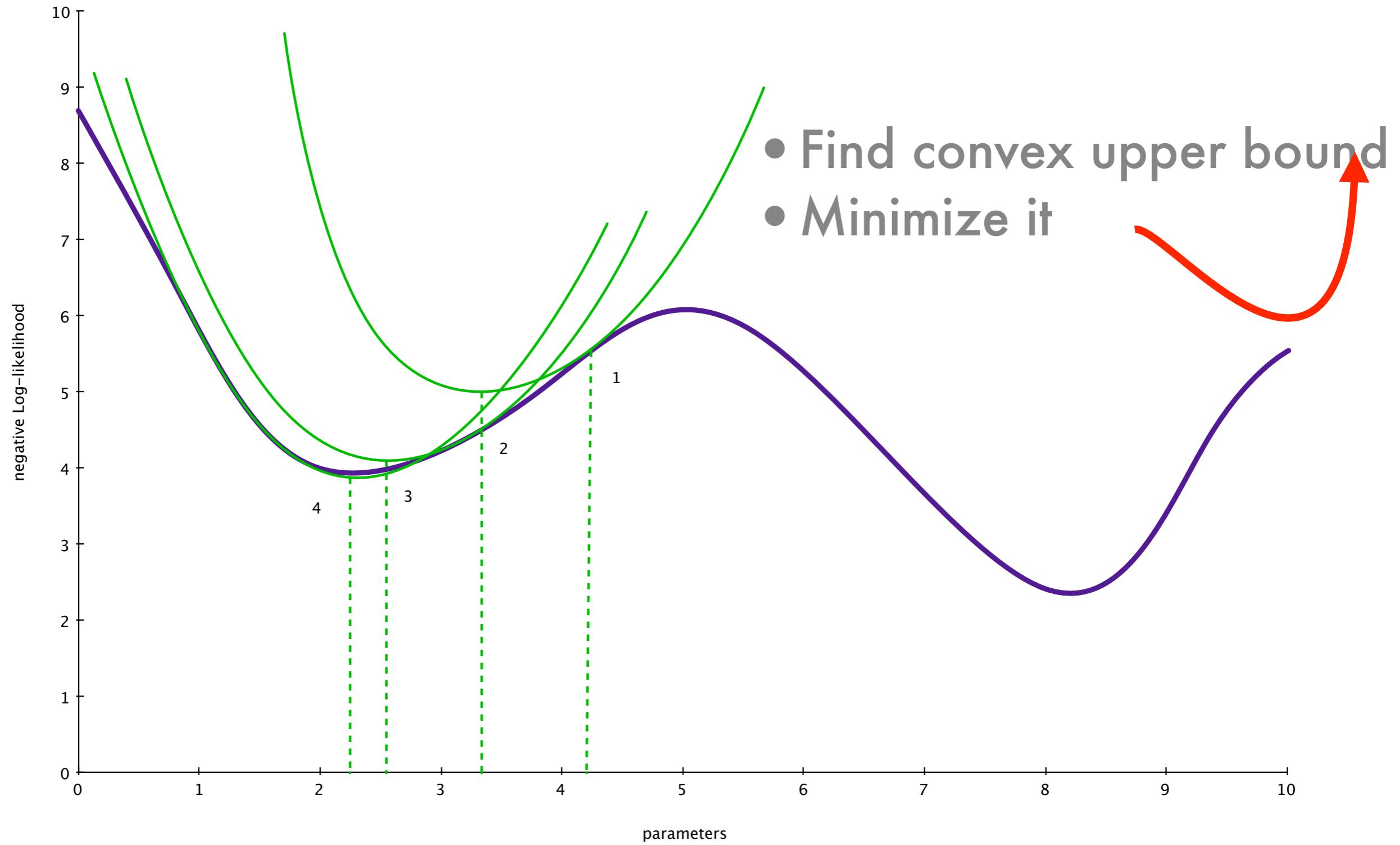
$$\underset{\theta, \mu, \sigma}{\text{maximize}} p(X|\theta, \sigma, \mu) = \underset{\theta, \mu, \sigma}{\text{maximize}} \sum_Y \prod_{i=1}^n p(x_i|y_i, \sigma, \mu)p(y_i|\theta)$$

This problem is nonconvex and difficult to solve

- Key idea

If we knew $p(y|x)$ we could estimate the remaining parameters easily and vice versa

Nonconvex Minimization



Expectation Maximization

Expectation Maximization

- **Variational Bound**

$$\begin{aligned}\log p(x; \theta) &\geq \log p(x; \theta) - D(q(y) \| p(y|x; \theta)) \\&= \int dq(y) [\log p(x; \theta) + \log p(y|x; \theta) - \log q(y)] \\&= \int dq(y) \log p(x, y; \theta) - \int dq(y) \log q(y)\end{aligned}$$

This inequality is tight for $p(y|x) = q(y)$

Expectation Maximization

- **Variational Bound**

$$\begin{aligned}\log p(x; \theta) &\geq \log p(x; \theta) - D(q(y) \| p(y|x; \theta)) \\&= \int dq(y) [\log p(x; \theta) + \log p(y|x; \theta) - \log q(y)] \\&= \int dq(y) \log p(x, y; \theta) - \int dq(y) \log q(y)\end{aligned}$$

This inequality is tight for $p(y|x) = q(y)$

- **Expectation step**

$$q(y) = p(y|x; \theta)$$

Expectation Maximization

- **Variational Bound**

$$\begin{aligned}\log p(x; \theta) &\geq \log p(x; \theta) - D(q(y) \| p(y|x; \theta)) \\&= \int dq(y) [\log p(x; \theta) + \log p(y|x; \theta) - \log q(y)] \\&= \int dq(y) \log p(x, y; \theta) - \int dq(y) \log q(y)\end{aligned}$$

This inequality is tight for $p(y|x) = q(y)$

- **Expectation step**

$$q(y) = p(y|x; \theta)$$

find bound

Expectation Maximization

- **Variational Bound**

$$\begin{aligned}\log p(x; \theta) &\geq \log p(x; \theta) - D(q(y) \| p(y|x; \theta)) \\ &= \int dq(y) [\log p(x; \theta) + \log p(y|x; \theta) - \log q(y)] \\ &= \int dq(y) \log p(x, y; \theta) - \int dq(y) \log q(y)\end{aligned}$$

This inequality is tight for $p(y|x) = q(y)$

- **Expectation step**

find bound

$$q(y) = p(y|x; \theta)$$

- **Maximization step**

$$\theta^* = \operatorname{argmax}_{\theta} \int dq(y) \log p(x, y; \theta)$$

Expectation Maximization

- **Variational Bound**

$$\begin{aligned}\log p(x; \theta) &\geq \log p(x; \theta) - D(q(y) \| p(y|x; \theta)) \\ &= \int dq(y) [\log p(x; \theta) + \log p(y|x; \theta) - \log q(y)] \\ &= \int dq(y) \log p(x, y; \theta) - \int dq(y) \log q(y)\end{aligned}$$

This inequality is tight for $p(y|x) = q(y)$

- **Expectation step**

find bound

$$q(y) = p(y|x; \theta)$$

- **Maximization step**

maximize it

$$\theta^* = \operatorname{argmax}_{\theta} \int dq(y) \log p(x, y; \theta)$$

Expectation Step

- Factorizing distribution

$$q(Y) = \prod_i q_i(y)$$

- E-Step

$q_i(y) \propto p(x_i|y_i, \mu, \sigma)p(y_i|\theta)$ hence

$$m_{iy} := \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_y|^{\frac{1}{2}}} \exp \left[-\frac{1}{2}(x_i - \mu_y)\Sigma_y^{-1}(x_i - \mu_y) \right] p(y)$$

$$q_i(y) = \frac{m_{iy}}{\sum_{y'} m_{iy'}}$$

Maximization Step

- Log-likelihood

$$\log p(X, Y | \theta, \mu, \sigma) = \sum_{i=1}^n \log p(x_i | y_i, \mu, \sigma) + \log p(y_i | \theta)$$

- Cluster distribution
(weighted Gaussian MLE)

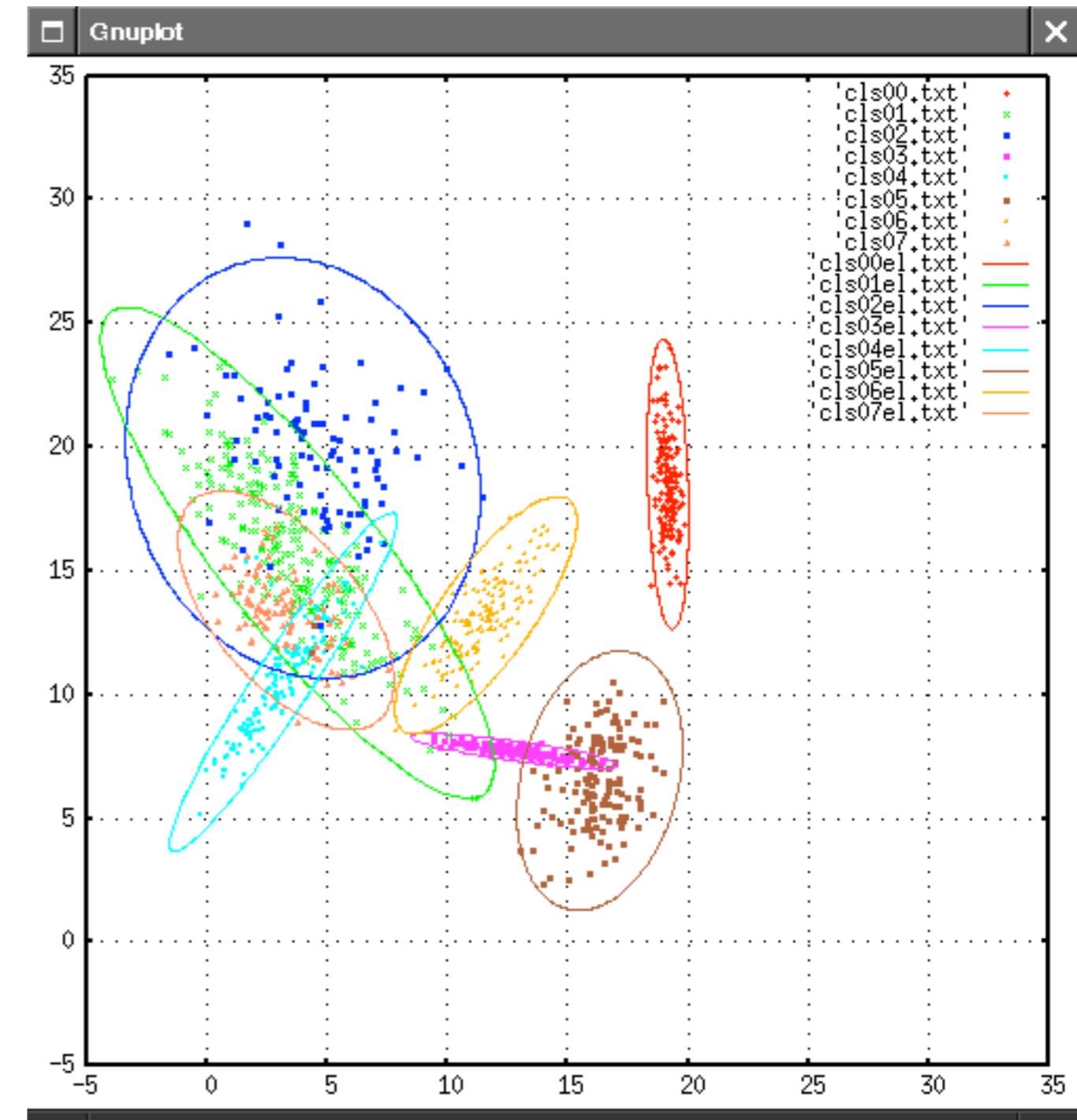
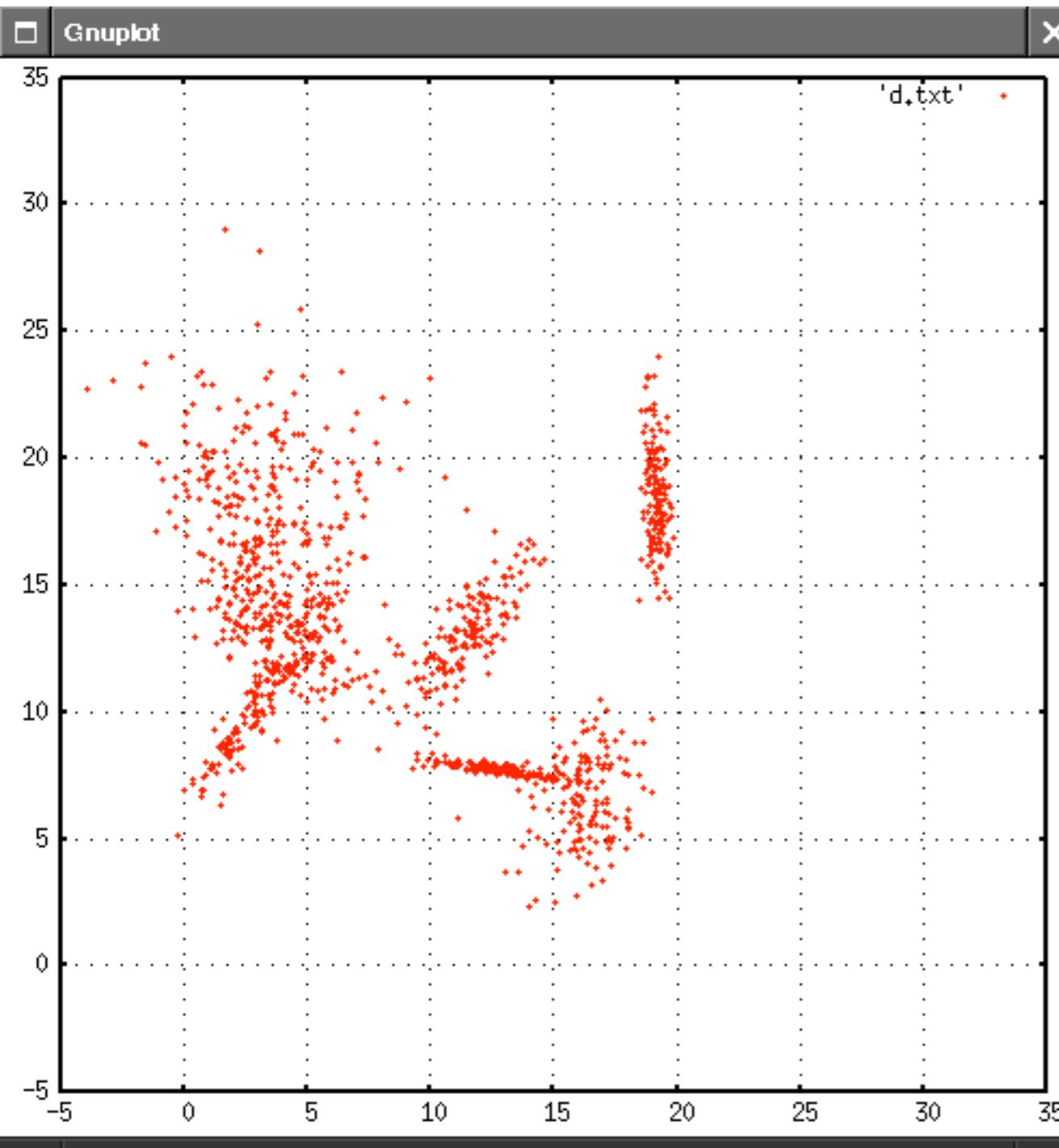
$$n_y = \sum_i q_i(y)$$

$$\mu_y = \frac{1}{n_y} \sum_{i=1}^n q_i(y) x_i$$
$$\Sigma_y = \frac{1}{n_y} \sum_{i=1}^n q_i(y) x_i x_i^\top - \mu_y \mu_y^\top$$

- Cluster probabilities

$$\theta^* = \operatorname{argmax}_{\theta} \sum_{i=1}^n \sum_y q_i(y) \log p(y_i | \theta) \text{ hence } p(y | \theta) = \frac{n_y}{n}$$

EM Clustering in action



Problem

**Estimates will diverge
(infinite variance, zero probability, tiny clusters)**

Solution

- Use priors for μ, σ, θ
 - Dirichlet distribution for cluster probabilities
 - Gauss-Wishart for Gaussian
- Cluster distribution

$$n_y = n_0 + \sum_i q_i(y)$$

$$\mu_y = \frac{1}{n_y} \sum_{i=1}^n q_i(y) x_i$$

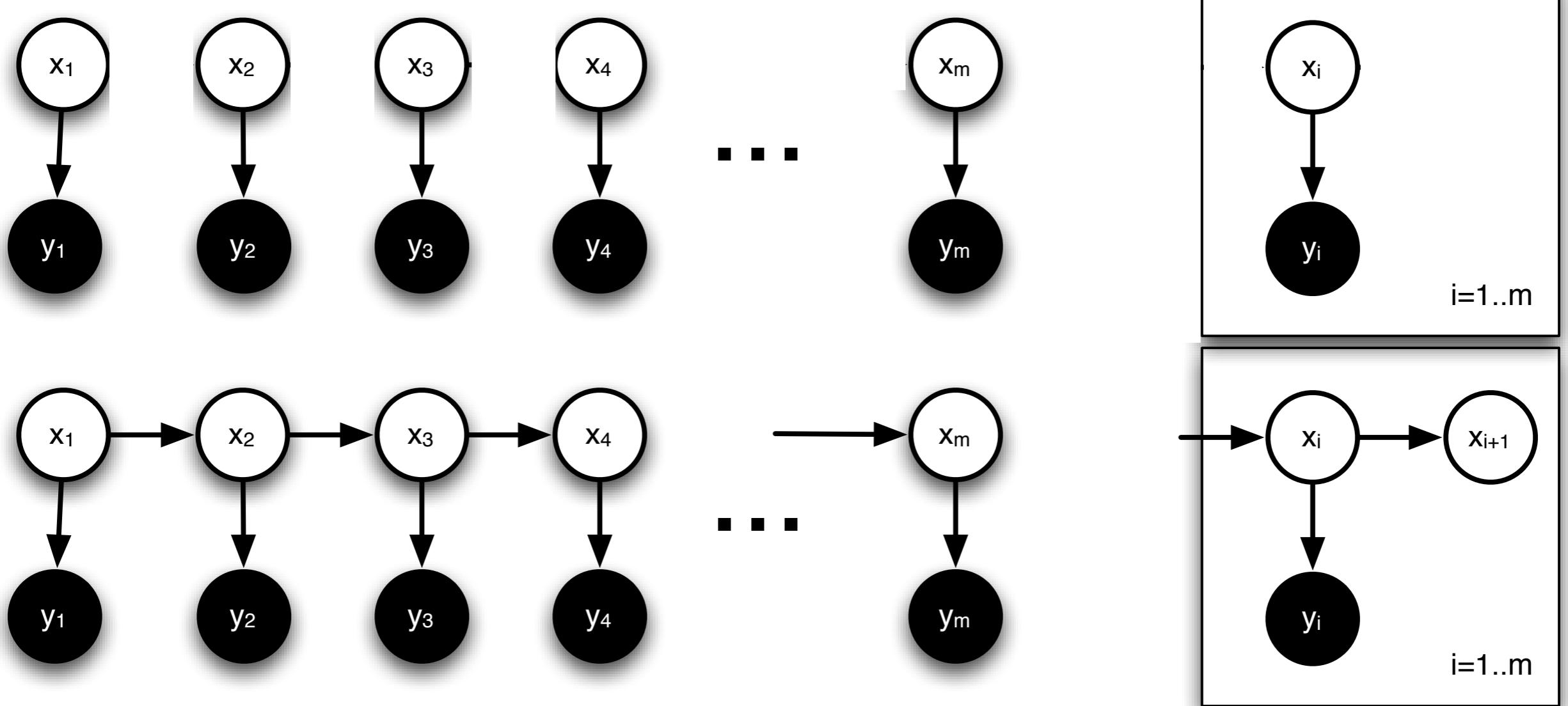
$$\Sigma_y = \frac{1}{n_y} \sum_{i=1}^n q_i(y) x_i x_i^\top + \frac{n_0}{n_y} \mathbf{1} - \mu_y \mu_y^\top$$

- Cluster probabilities

$$p(y|\theta) = \frac{n_y}{n + k \cdot n_0}$$

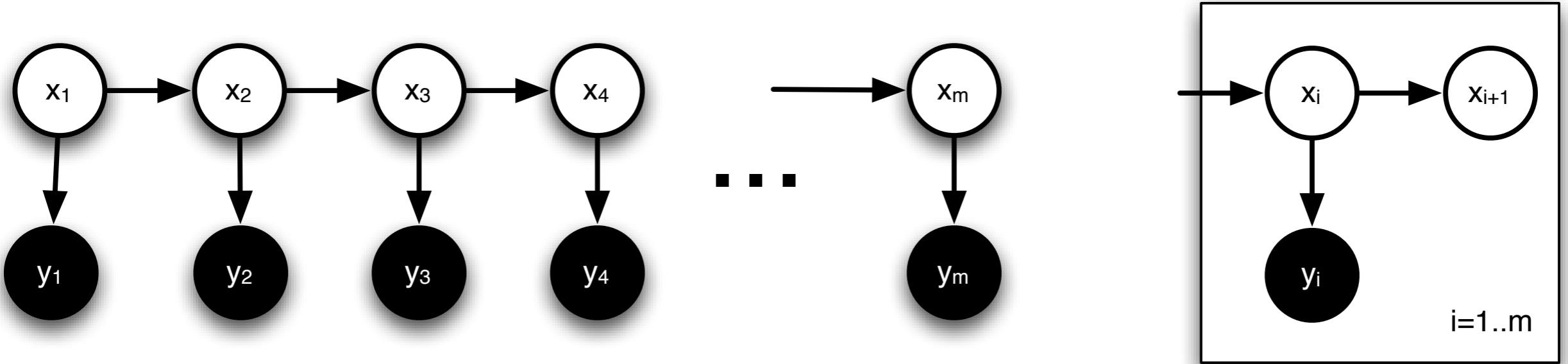
Hidden Markov Models

Clustering and Hidden Markov Models



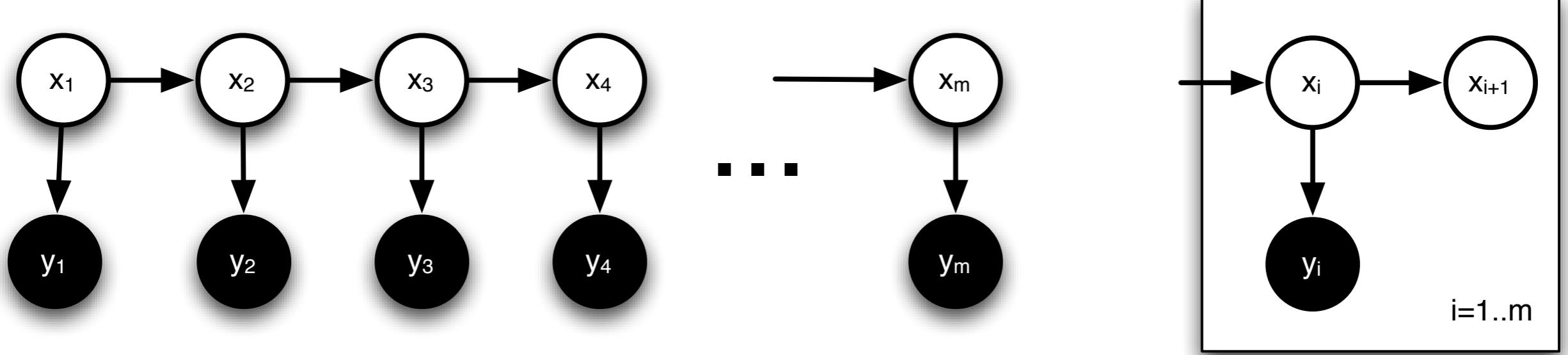
- **Clustering - no dependence between observations**
- **Hidden Markov Model - dependence between states**

Applications



- Speech recognition (sound | text)
- Optical character recognition (writing | text)
- Gene finding (DNA sequence | genes)
- Activity recognition (accelerometer | activity)

Inference



$$p(x, y) = p(y_1) \left[\prod_{i=1}^{m-1} p(y_{i+1}|y_i) p(x_i|y_i) \right] p(x_m|y_m)$$

- Summing over y possible via dynamic programming
- Log-likelihood is nonconvex

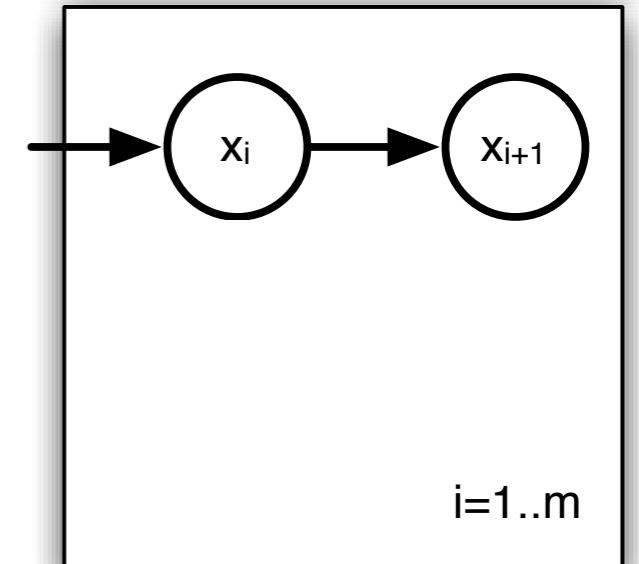
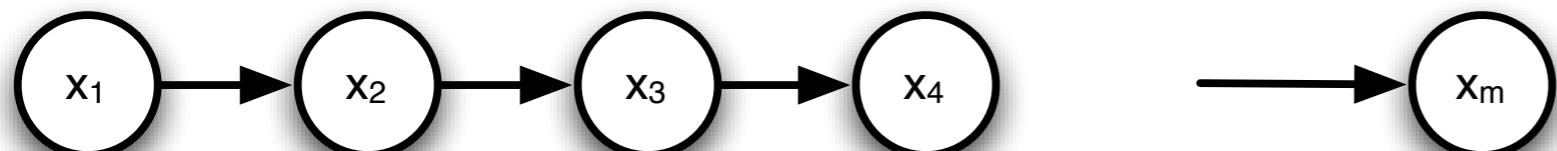
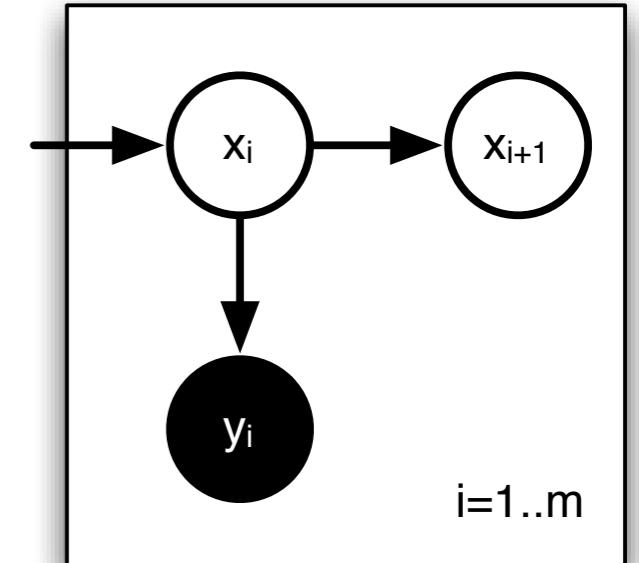
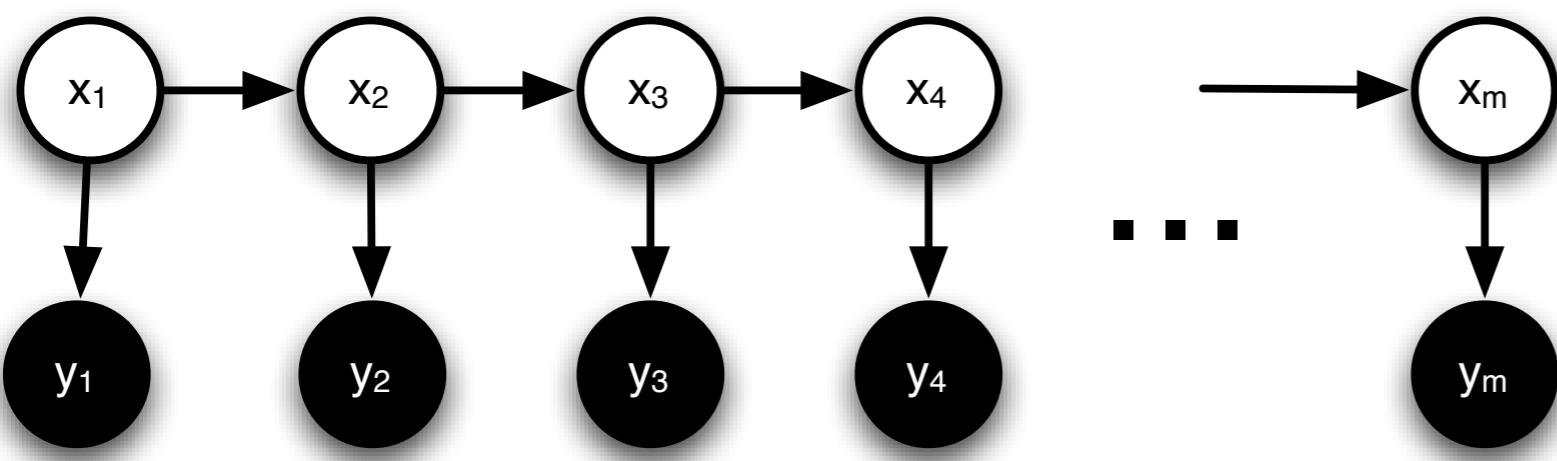
Variational Approximation

- Lower bound on log-likelihood

$$\log p(x; \theta) \geq \int dq(y) \log p(x, y; \theta) - \int dq(y) \log q(y)$$

- Key insight - inequality holds for any q
 - Find q within subset Q to tighten inequality
 - Find parameters to maximize for fixed q
- Inference for graphical models where joint probability computation is infeasible

Variational Approximation



- Variational approximation via

$$q(y) = q(y_1) \prod_{x=2}^m q(y_i | y_{i-1})$$

- Compute $p(x|y)$ via dynamic programming

Variational Method

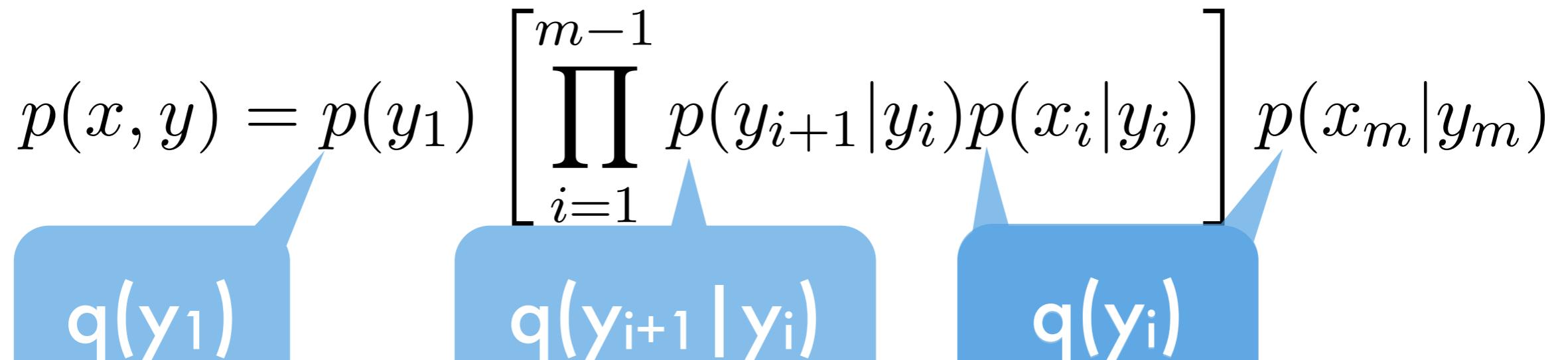
- Initialize parameters somehow

- Set $q(x) = p(x|y)$

Dynamic programming yields chain

- Maximizing the log-likelihod w.r.t. q

$$\log p(x; \theta) \geq \int dq(y) \log p(x, y; \theta) - \int dq(y) \log q(y)$$

$$p(x, y) = p(y_1) \left[\prod_{i=1}^{m-1} p(y_{i+1}|y_i) p(x_i|y_i) \right] p(x_m|y_m)$$


q(y₁) q(y_{i+1} | y_i) q(y_i)

Parameter Estimation

$$\mathbf{E}_{y \sim q} [\log p(x, y; \theta)] = \mathbf{E}_{y_1 \sim q} \log p(y_1; \theta) + \sum_{i=1}^m \mathbf{E}_{y_i \sim q} \log p(x_i | y_i; \theta)$$

$$+ \sum_{i=1}^{m-1} \mathbf{E}_{y_{i+1}, y_i \sim q} \log p(y_{i+1} | y_i; \theta)$$

- $p(y_1)$

Since we have $\mathbf{E}_{q(y_1)} [\log p(y_1)]$ set $p(y_1) = q(y_1)$

- $p(x_i | y_i)$

Same as clustering

e.g. for Gaussians

$$\mu_y = \frac{1}{n_y} \sum_{i=1}^n q_i(y) x_i$$

$$\Sigma_y = \frac{1}{n_y} \sum_{i=1}^n q_i(y) x_i x_i^\top - \mu_y \mu_y^\top$$

Parameter Estimation

$$\begin{aligned}\mathbf{E}_{y \sim q} [\log p(x, y; \theta)] &= \mathbf{E}_{y_1 \sim q} \log p(y_1; \theta) + \sum_{i=1}^{m-1} \mathbf{E}_{y_i \sim q} \log p(x_i | y_i; \theta) \\ &\quad + \sum_{i=1}^{m-1} \mathbf{E}_{y_{i+1}, y_i \sim q} \log p(y_{i+1} | y_i; \theta)\end{aligned}$$

- Maximum likelihood estimate for $p(y' | y)$

$$\sum_{i=1}^{m-1} q(y_{i+1} = a, y_i = b) \log p(a | b)$$

$$\text{hence } p(a | b) = \frac{\sum_{i=1}^{m-1} q(y_{i+1} = a, y_i = b)}{\sum_{i=1}^{m-1} q(y_i = b)}$$

effective sample

Smoothed Estimates

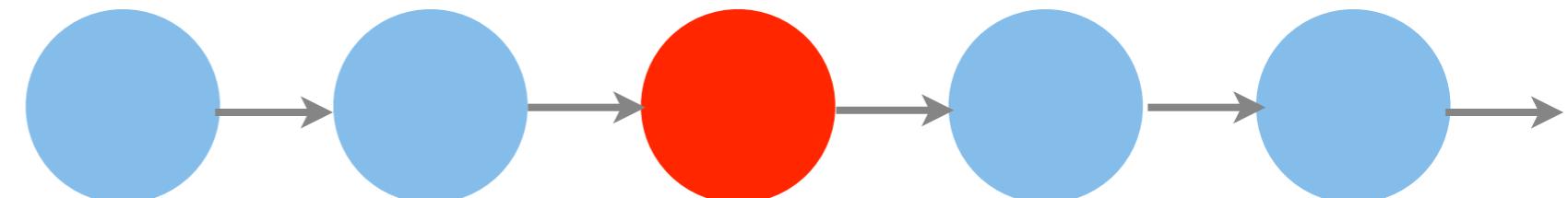
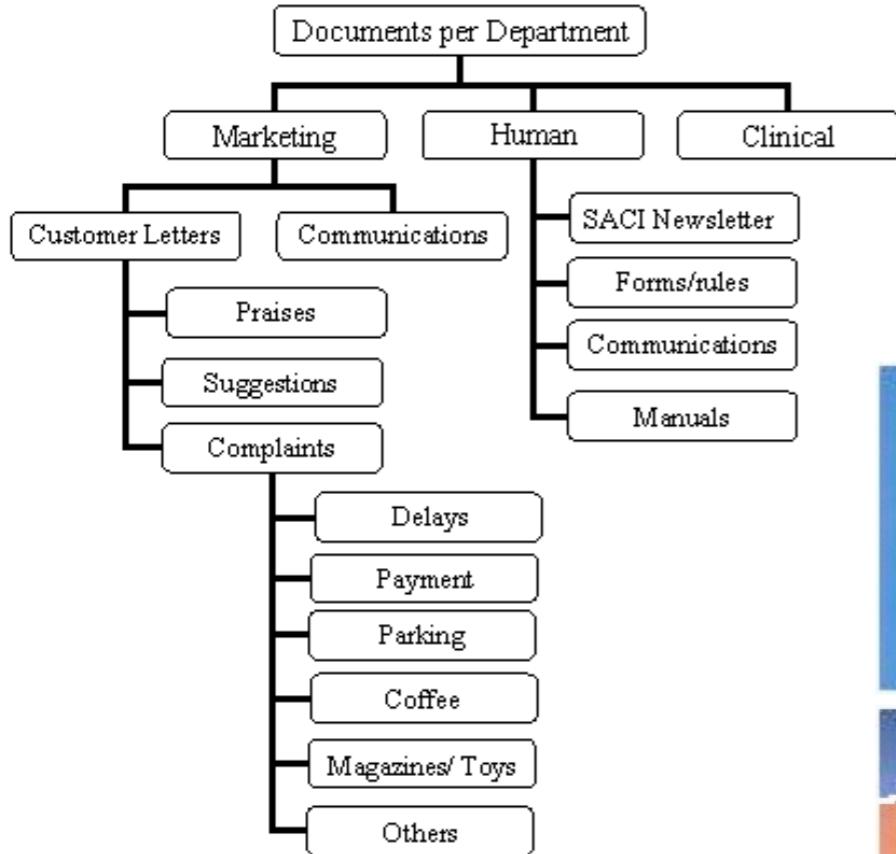
- Laplace prior on latent state distribution
 - Uniform distribution over states
 - Alternatively assume that state remains

$$p(a|b) = \frac{n_{a|b} + \sum_{i=1}^{m-1} q(y_{i+1} = a, y_i = b)}{n_b + \sum_{i=1}^{m-1} q(y_i = b)}$$

transition
smoother

aggregate
mass

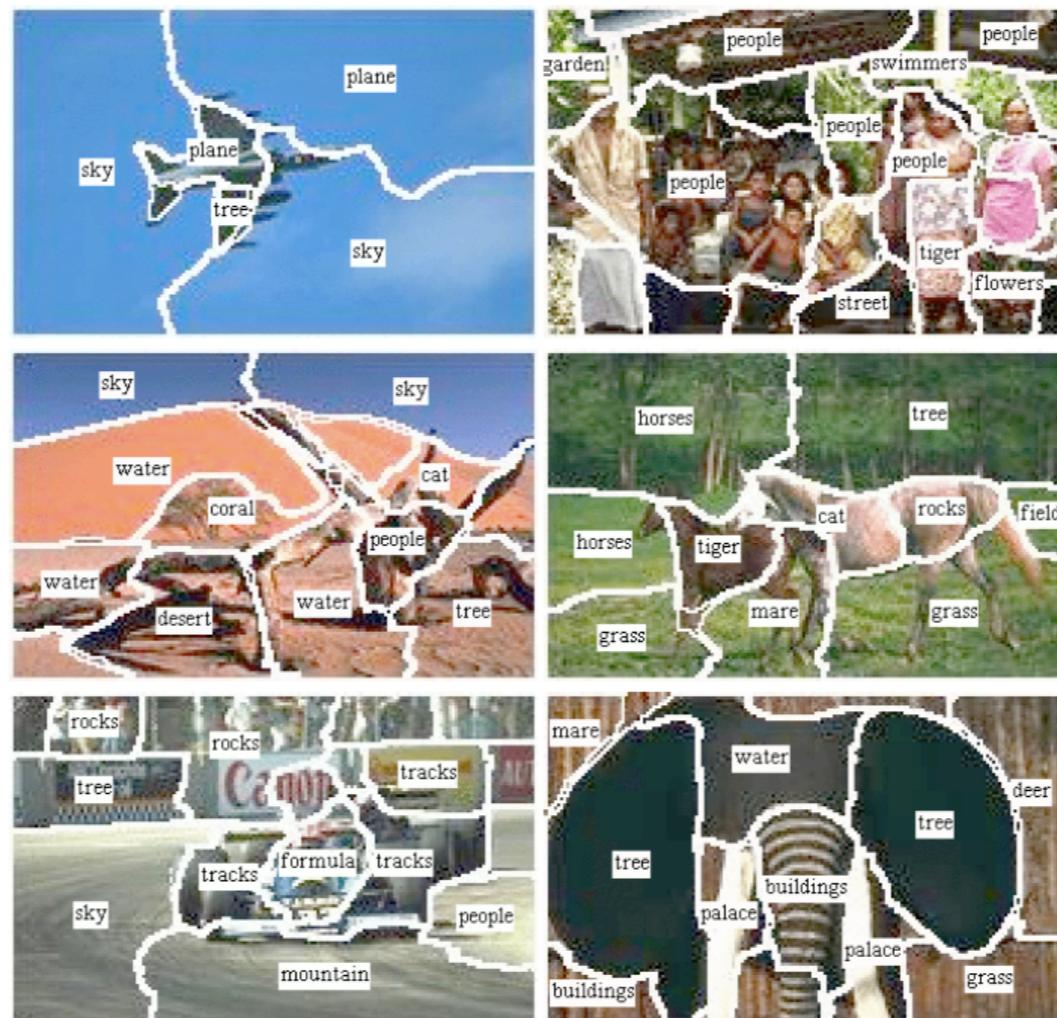
Beyond mixtures



chains

taxonomies

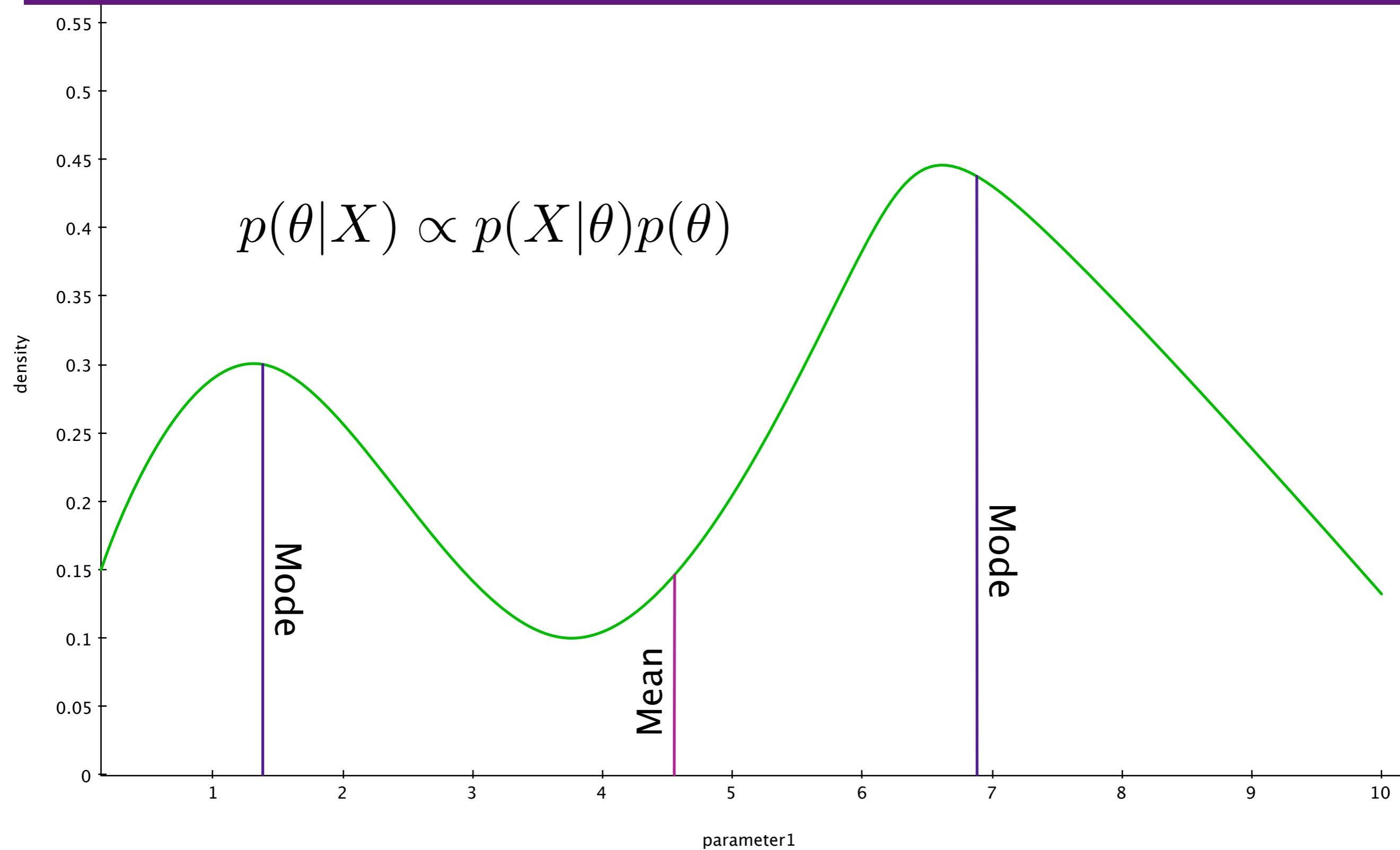
topics



Sampling

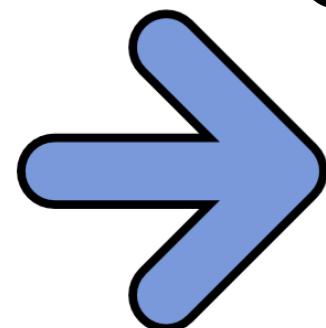


Is maximization (always) good?

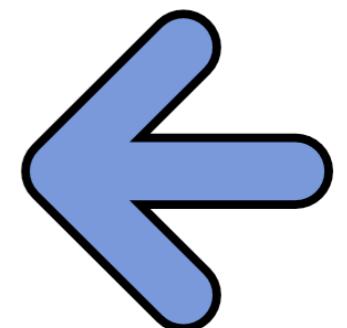


Sampling

- Key idea
 - Want **accurate** distribution of the posterior
 - **Sample** from posterior distribution rather than **maximizing** it
- Problem - direct sampling is usually intractable
- Solutions
 - Markov Chain Monte Carlo (complicated)
 - Gibbs Sampling (somewhat simpler)



$x \sim p(x|x')$ and then $x' \sim p(x'|x)$



Gibbs sampling

- Gibbs sampling:
 - In most cases direct sampling not possible
 - Draw one set of variables at a time

		
	0.45	0.05
	0.05	0.45

Gibbs sampling

- Gibbs sampling:
 - In most cases direct sampling not possible
 - Draw one set of variables at a time

		
	0.45	0.05
	0.05	0.45

(b,g) - draw $p(.,g)$

Gibbs sampling

- Gibbs sampling:
 - In most cases direct sampling not possible
 - Draw one set of variables at a time

		
	0.45	0.05
	0.05	0.45

(b,g) - draw $p(.,g)$
(g,g) - draw $p(g,.)$

Gibbs sampling

- Gibbs sampling:
 - In most cases direct sampling not possible
 - Draw one set of variables at a time

		
	0.45	0.05
	0.05	0.45

(b,g) - draw $p(.,g)$
(g,g) - draw $p(g,.)$
(g,b) - draw $p(.,g)$

Gibbs sampling

- Gibbs sampling:
 - In most cases direct sampling not possible
 - Draw one set of variables at a time

		
	0.45	0.05
	0.05	0.45

(b,g) - draw $p(.,g)$
(g,g) - draw $p(g,.)$
(g,g) - draw $p(.,g)$
(b,g) - draw $p(b,.)$

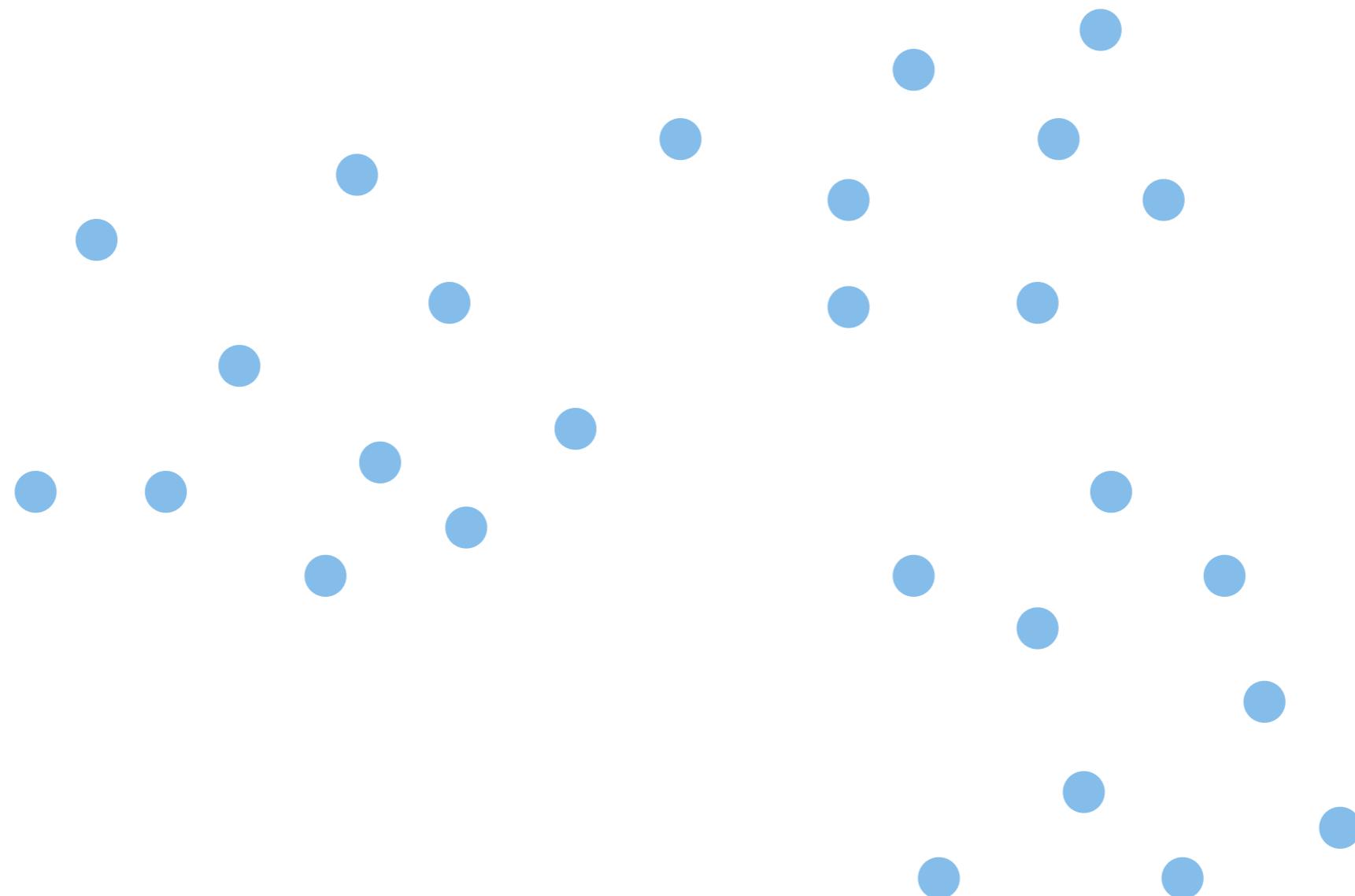
Gibbs sampling

- Gibbs sampling:
 - In most cases direct sampling not possible
 - Draw one set of variables at a time

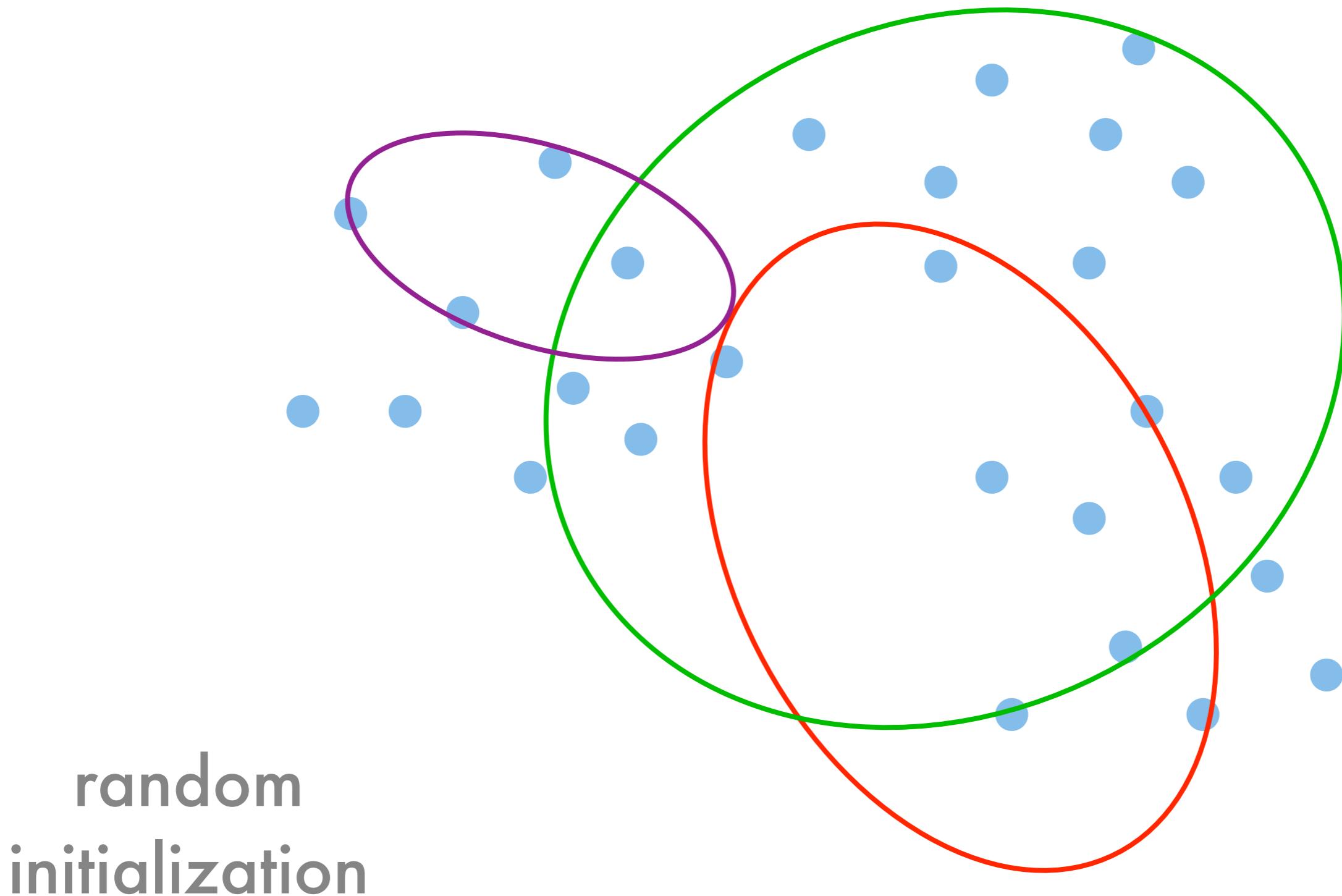
		
	0.45	0.05
	0.05	0.45

(b,g) - draw $p(.,g)$
(g,g) - draw $p(g,.)$
(g,g) - draw $p(.,g)$
(b,g) - draw $p(b,.)$
(b,b) ...

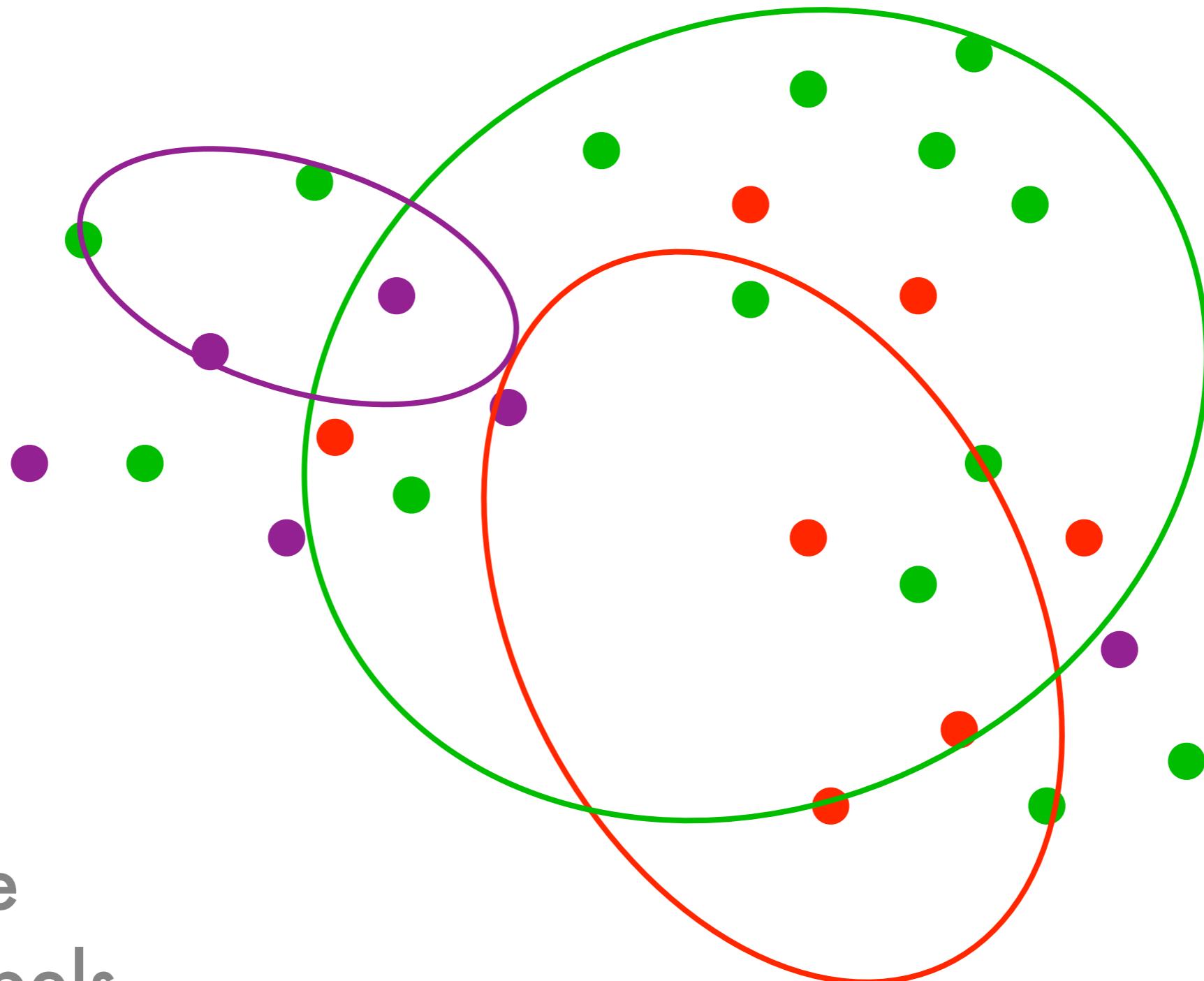
Gibbs sampling for clustering



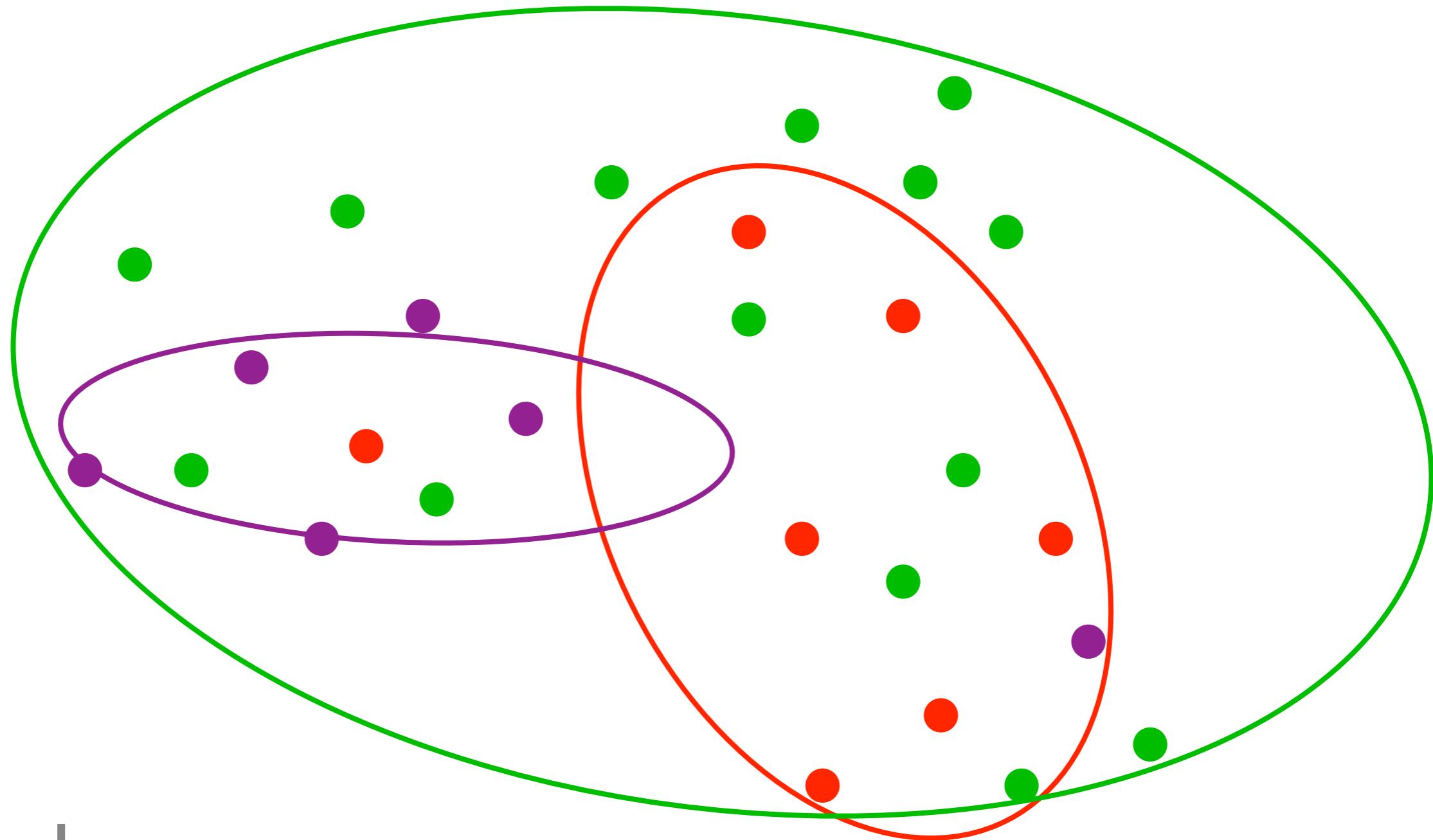
Gibbs sampling for clustering



Gibbs sampling for clustering

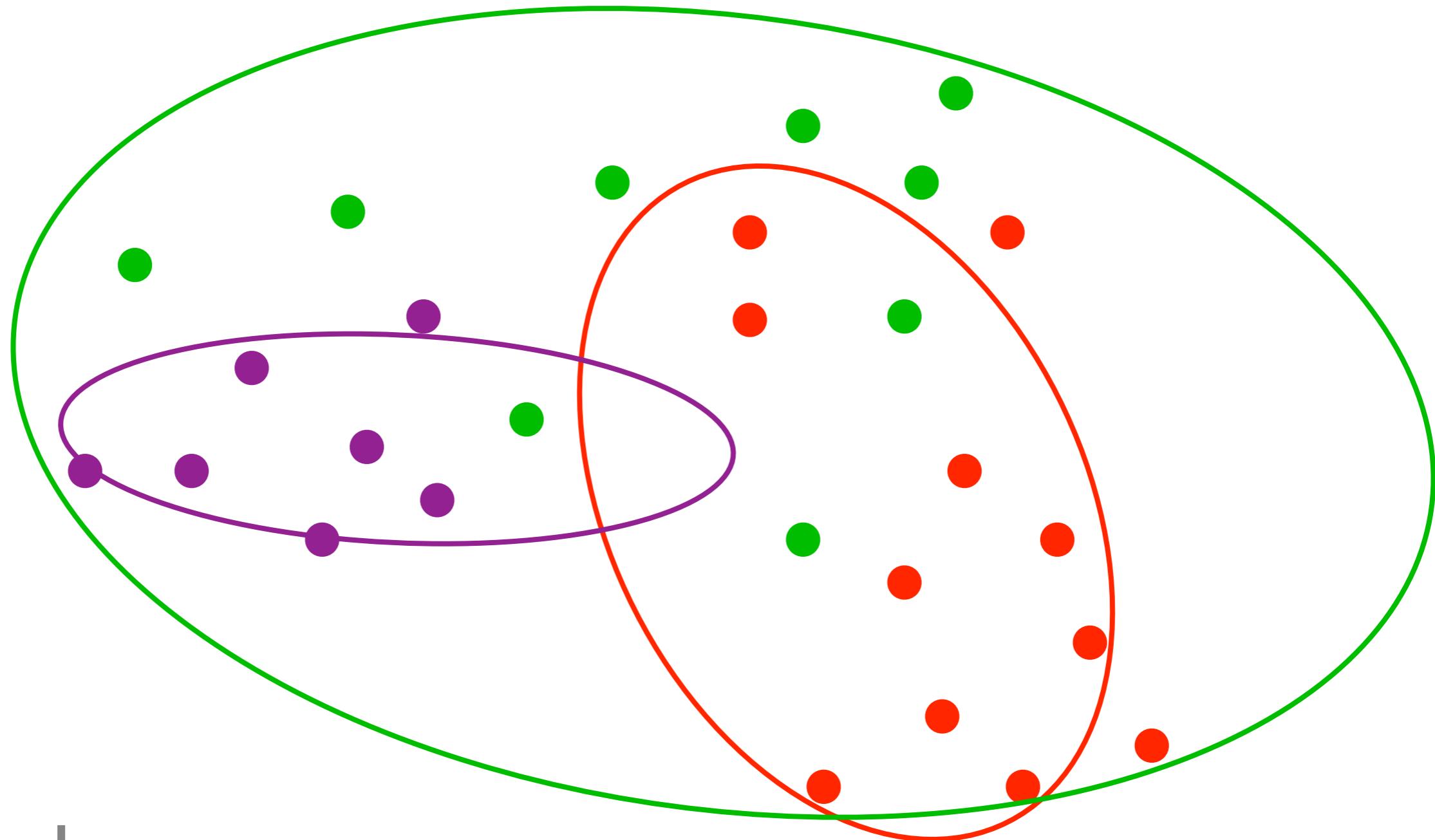


Gibbs sampling for clustering



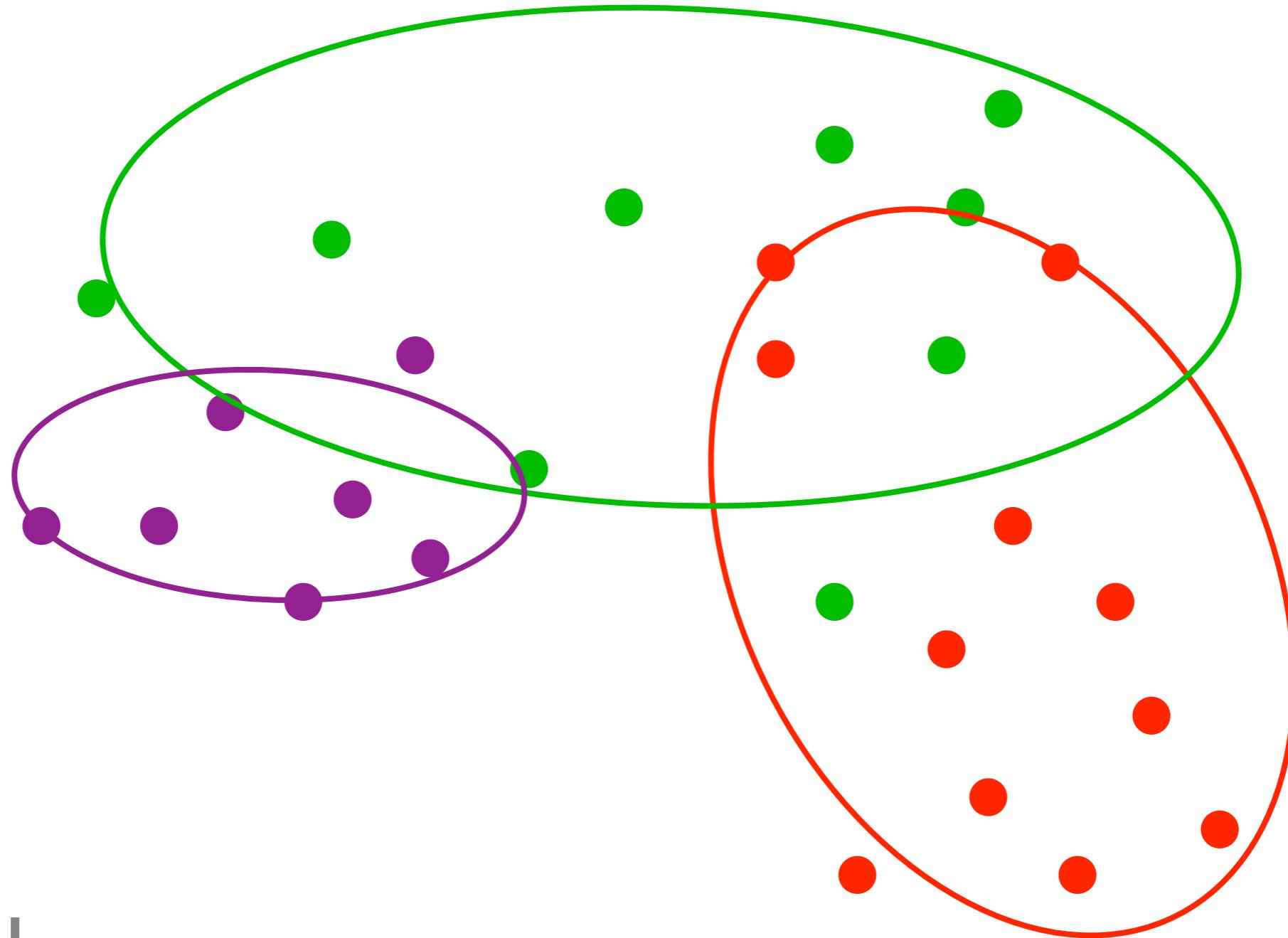
resample
cluster model

Gibbs sampling for clustering



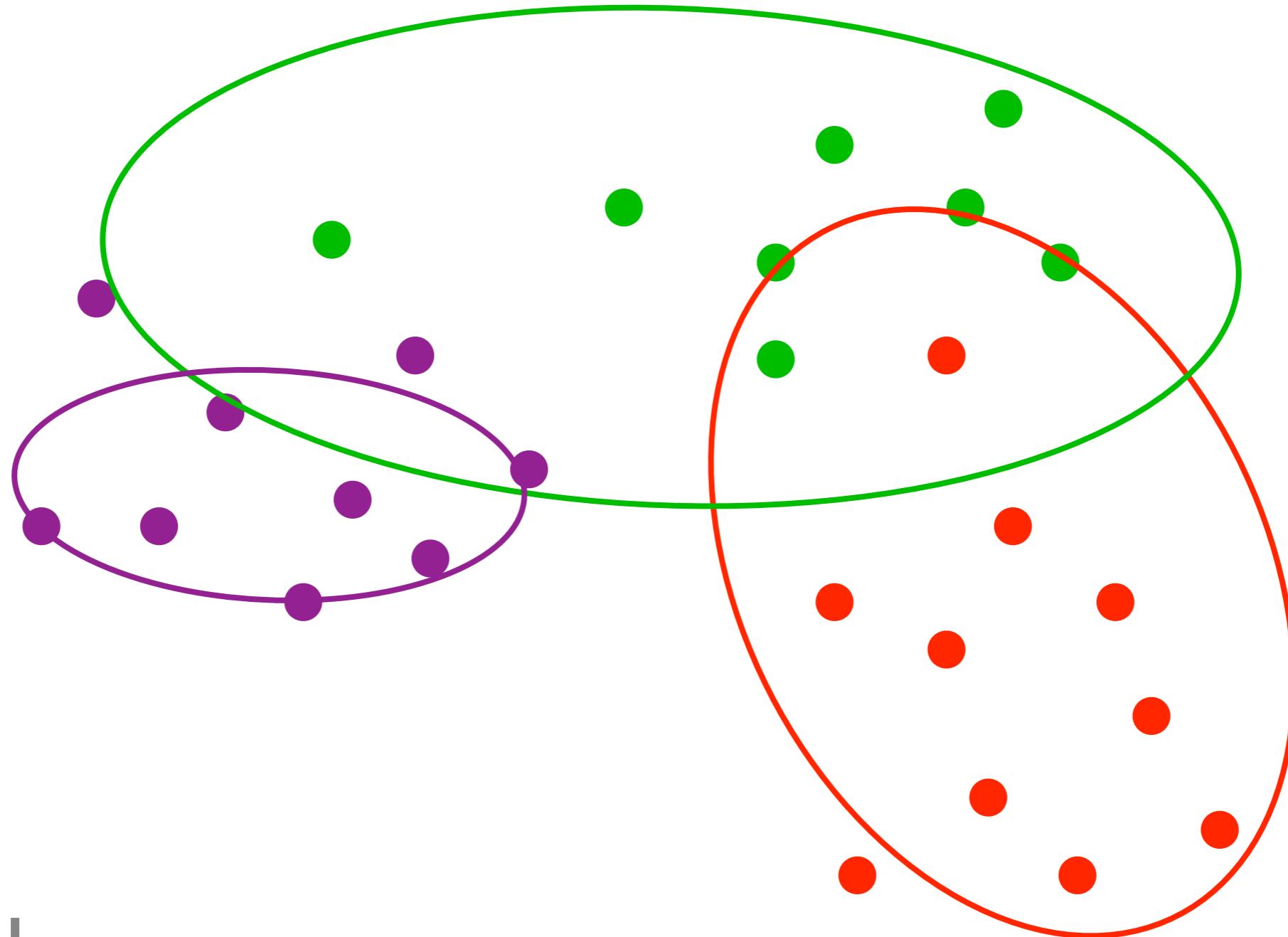
resample
cluster labels

Gibbs sampling for clustering



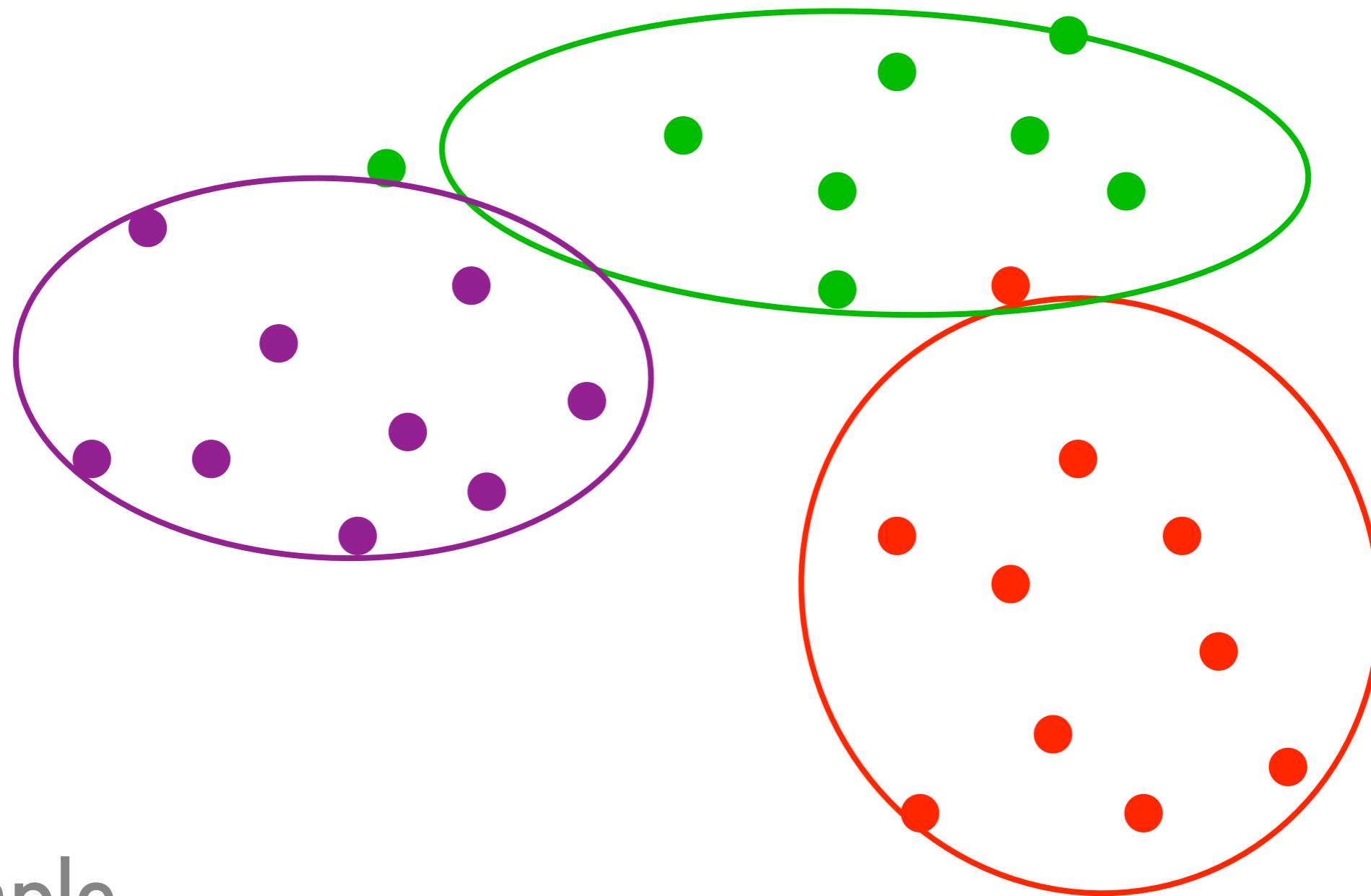
resample
cluster model

Gibbs sampling for clustering



resample
cluster labels

Gibbs sampling for clustering



resample

cluster model

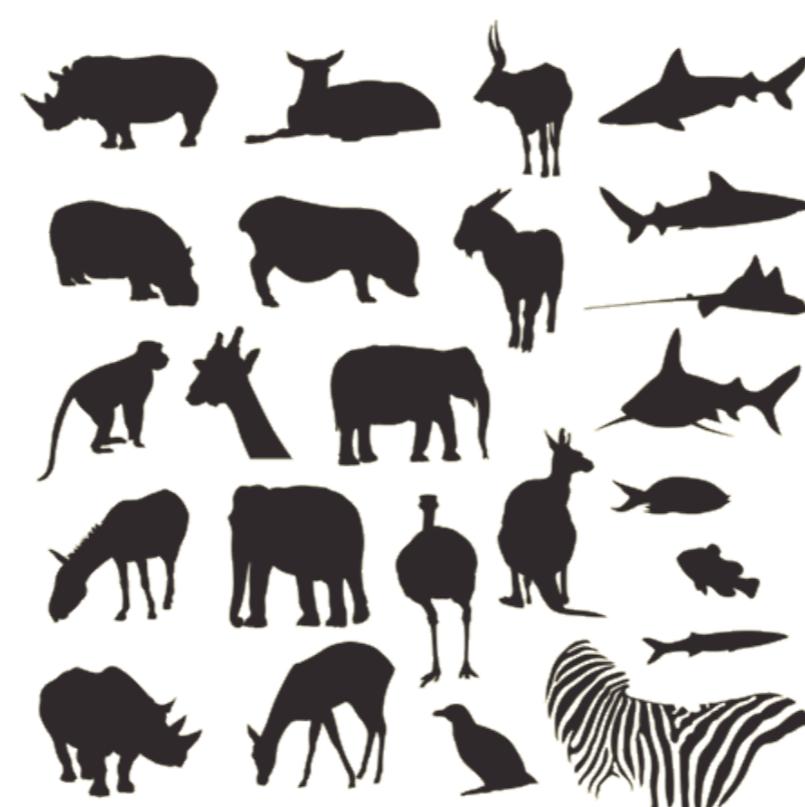
e.g. Mahout Dirichlet Process Clustering

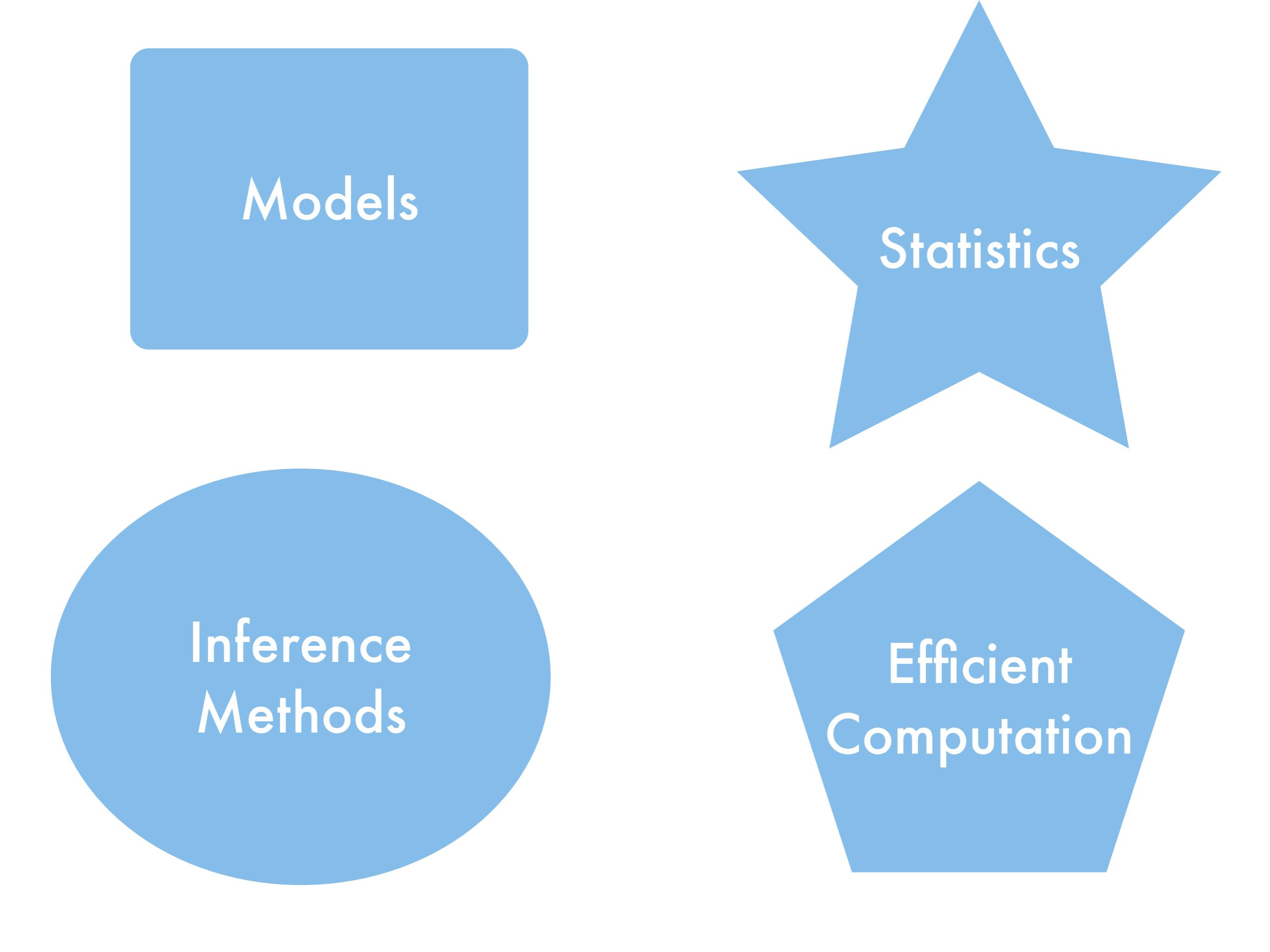
Inference Algorithm \neq Model

Inference Algorithm \neq Model

Corollary: EM \neq Clustering

Graphical Models Zoology





Models

Inference
Methods

Statistics

Efficient
Computation

Inference
Methods

Models

Efficient
Computation

Statistics

l_1, l_2 Priors

Conjugate
Prior

l_1, l_2 Priors

Exponential
Families

Mixtures
Clusters

Chains
HMM

Matrix
Factorization

Models

Factor
Models

directed
undirected

MRF
CRF

ℓ_1, ℓ_2 Priors

Conjugate
Prior

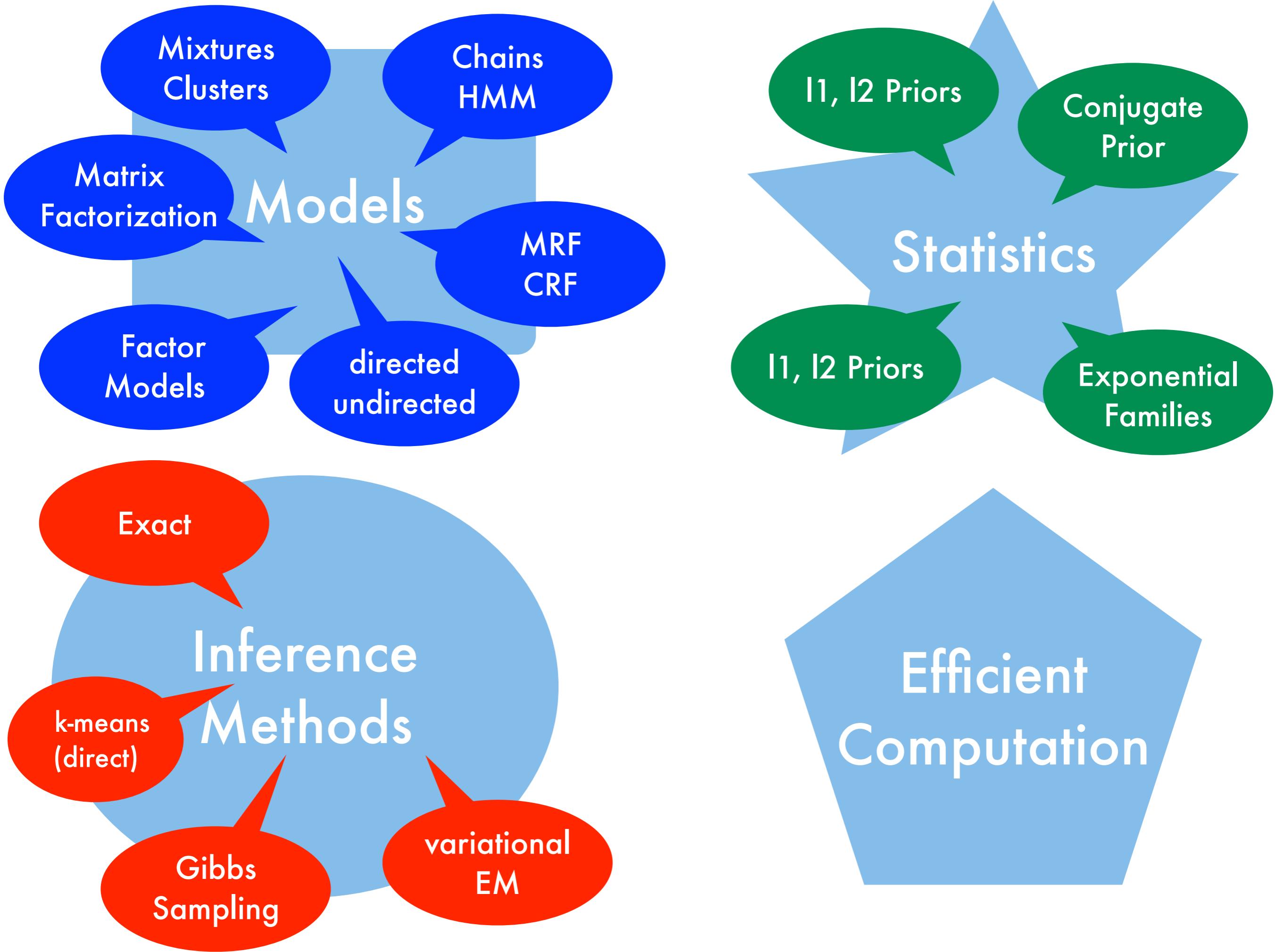
Statistics

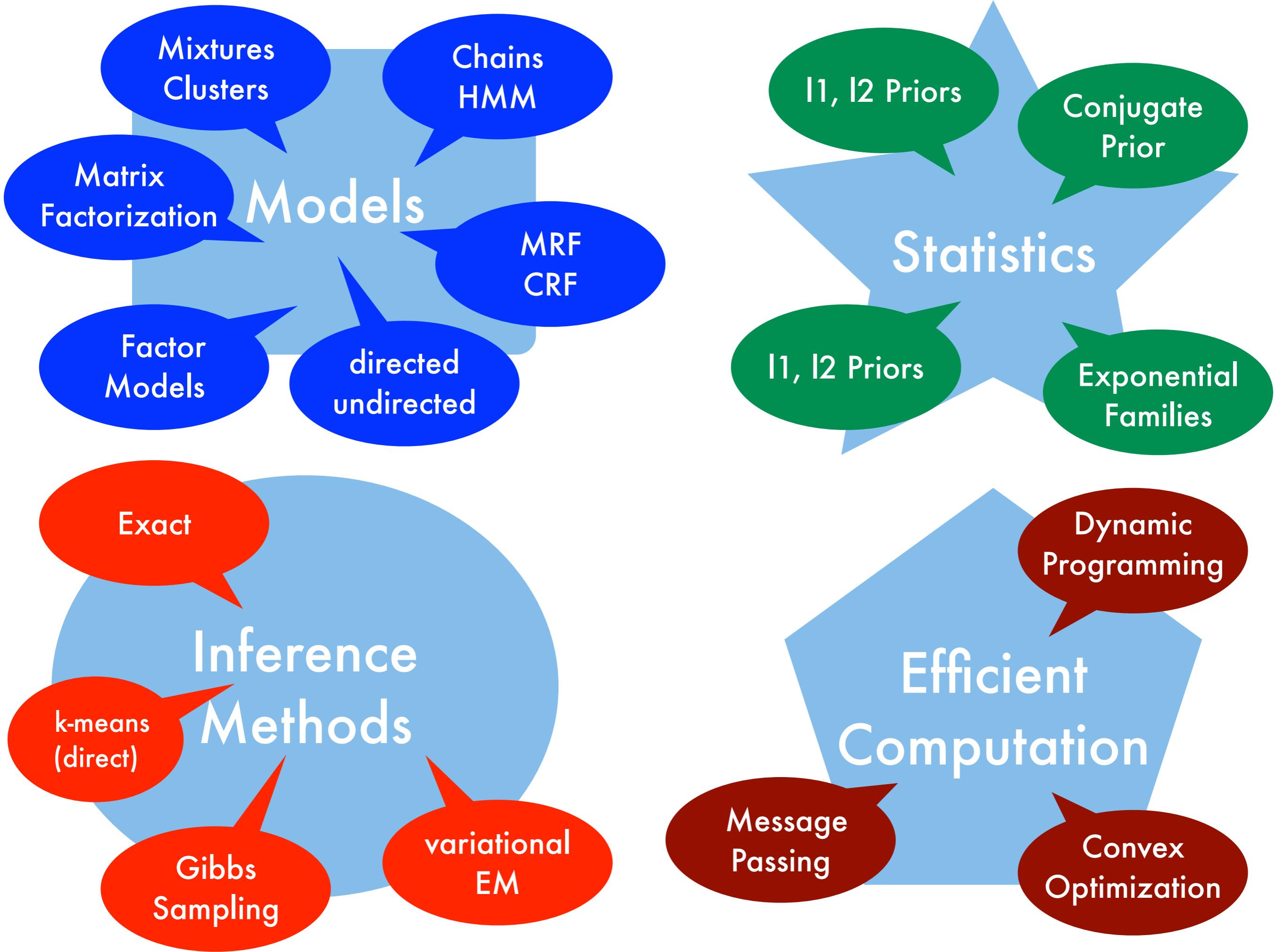
ℓ_1, ℓ_2 Priors

Exponential
Families

Inference
Methods

Efficient
Computation





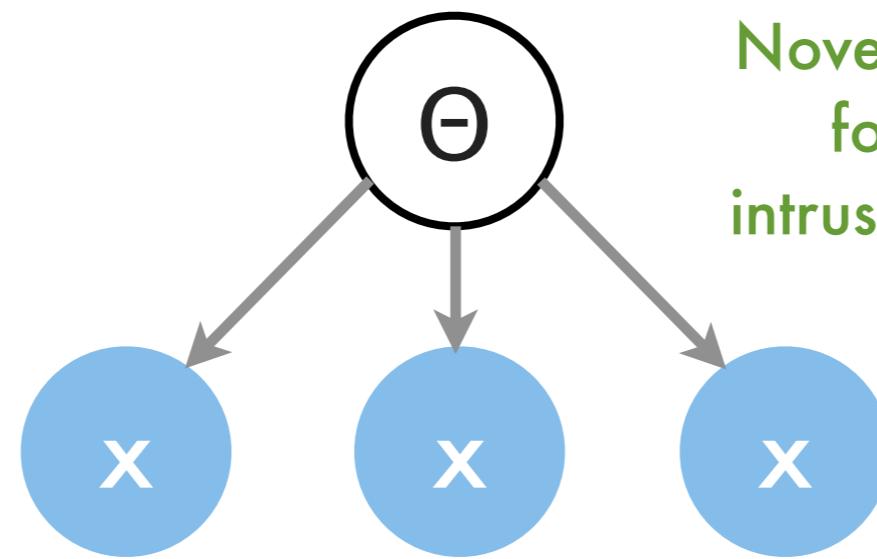
YAHOO![®]

YAHOO!

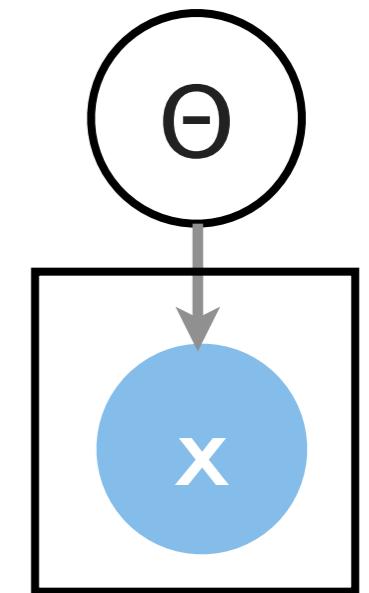


'Unsupervised' Models

Density
Estimation



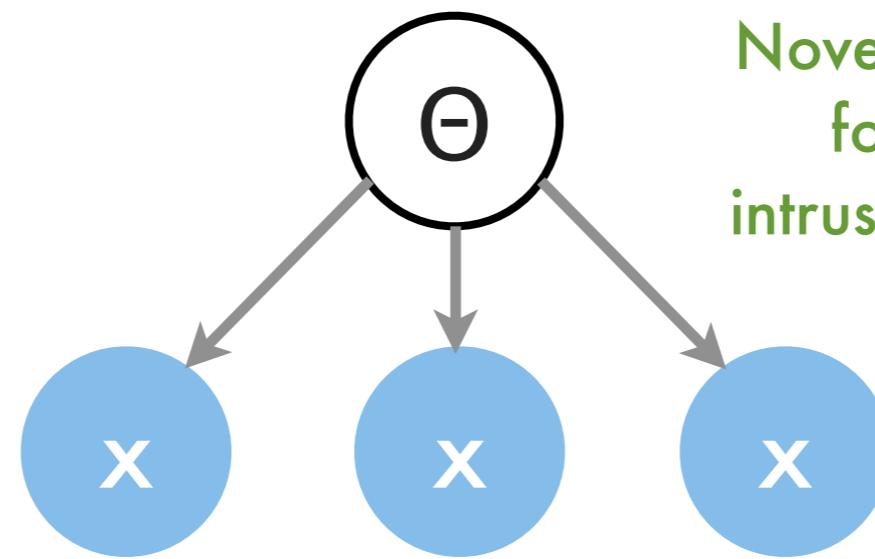
Novelty Detection
forecasting
intrusion detection



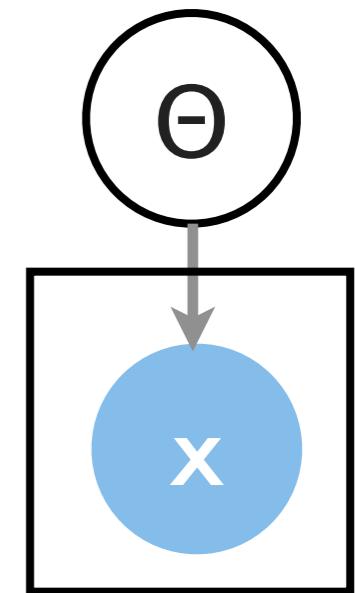
webpages
news
users
ads
queries
images

'Unsupervised' Models

Density
Estimation

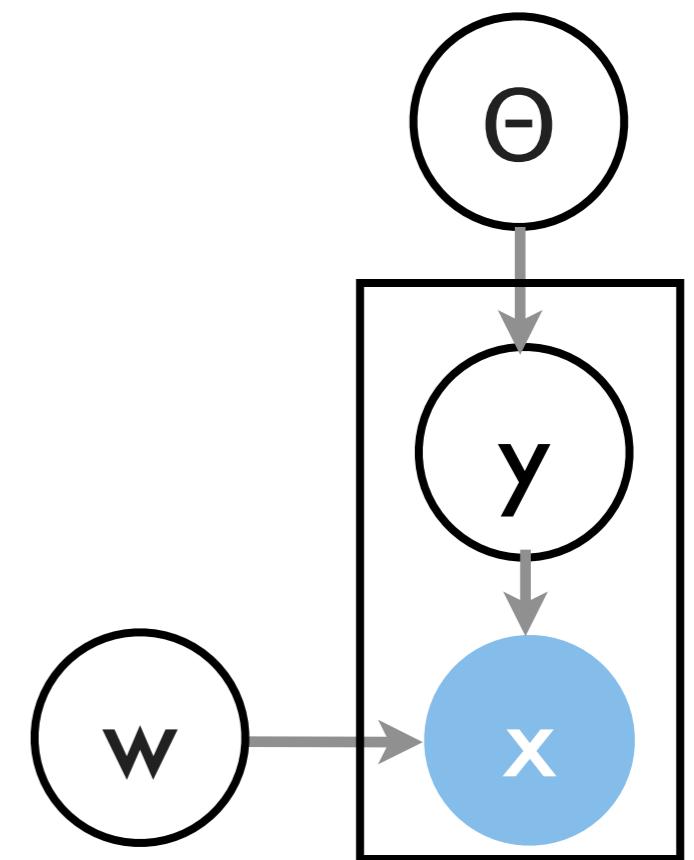
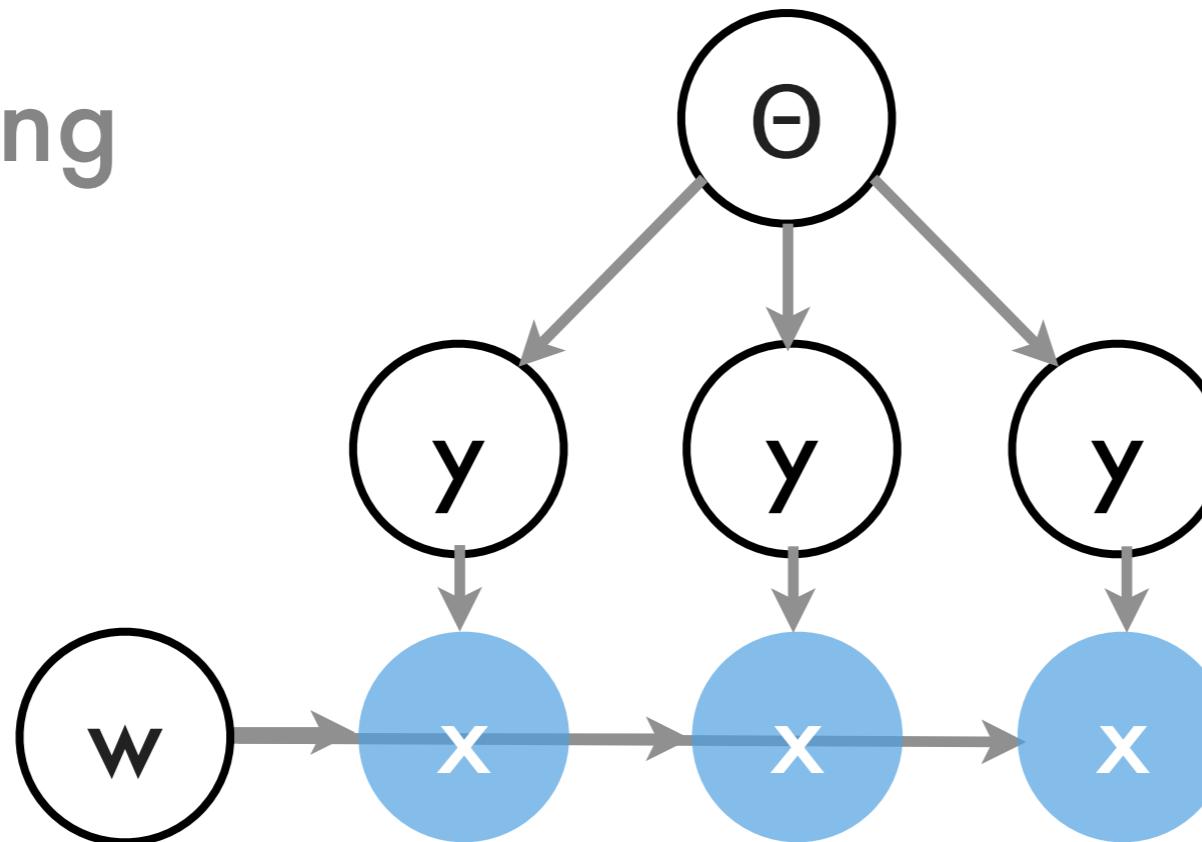


Novelty Detection
forecasting
intrusion detection



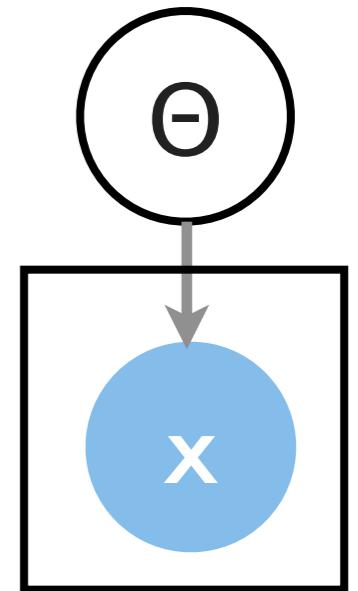
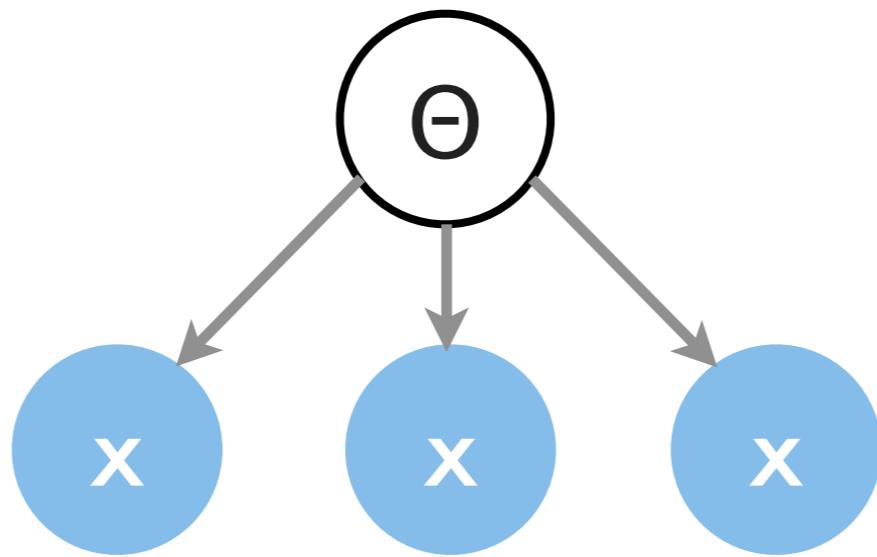
Clustering

webpages
news
users
ads
queries
images

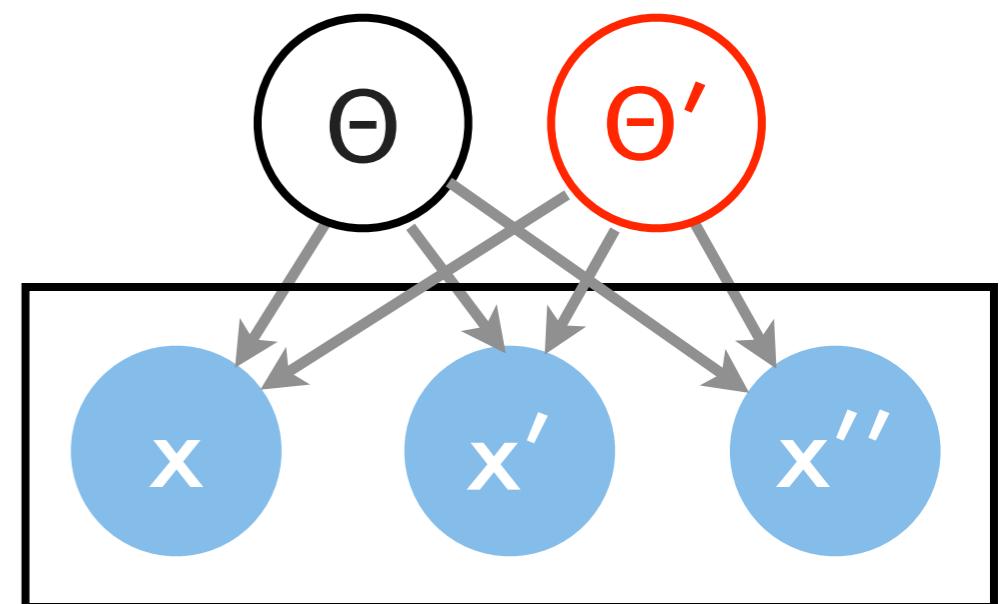


'Unsupervised' Models

Density
Estimation



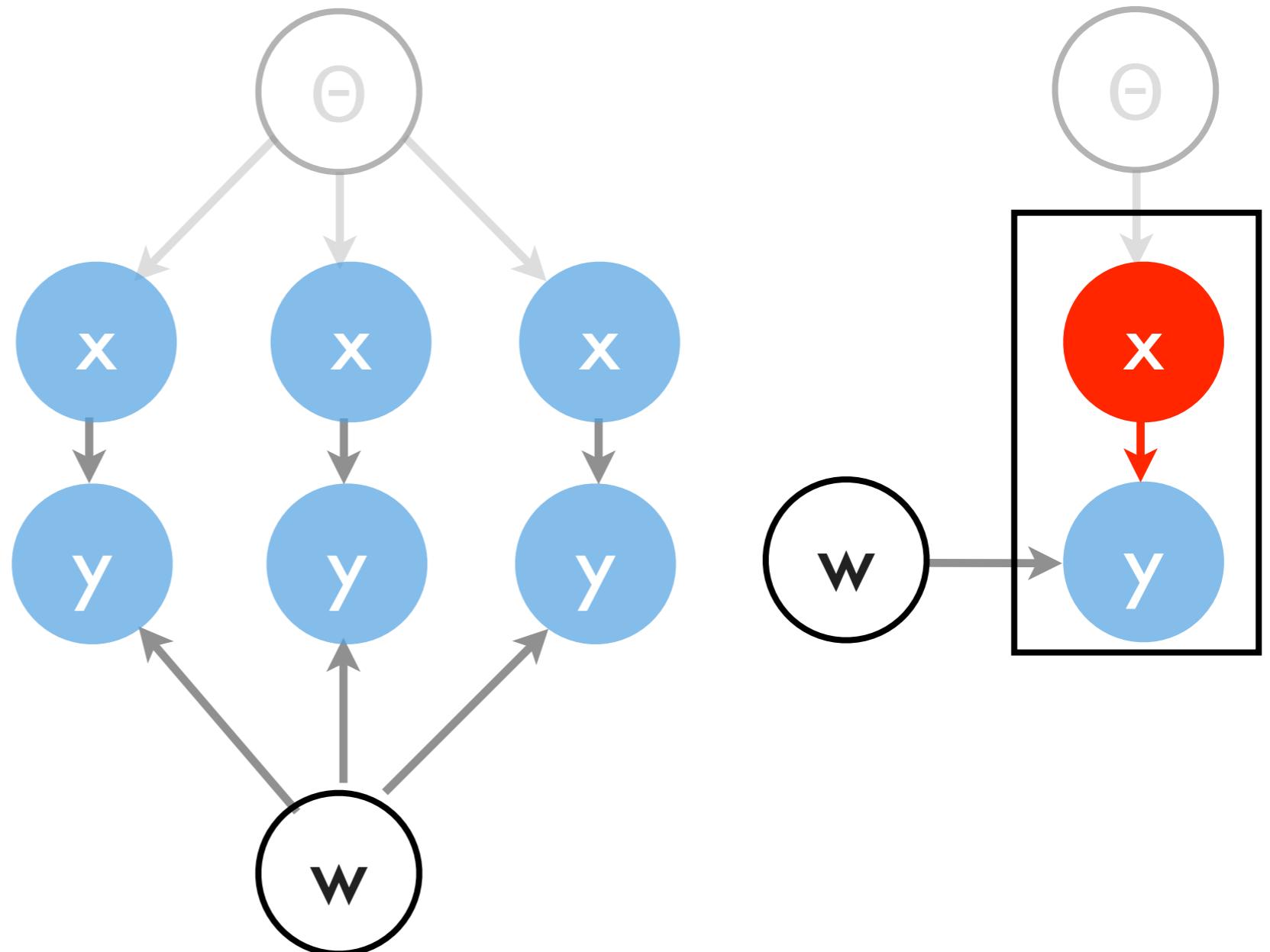
Factor
Analysis



'Supervised' Models

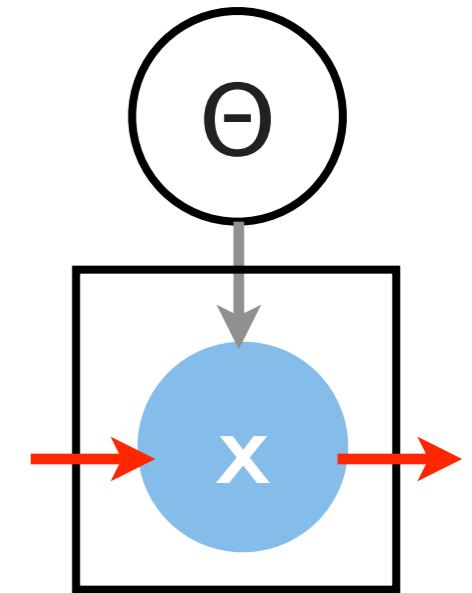
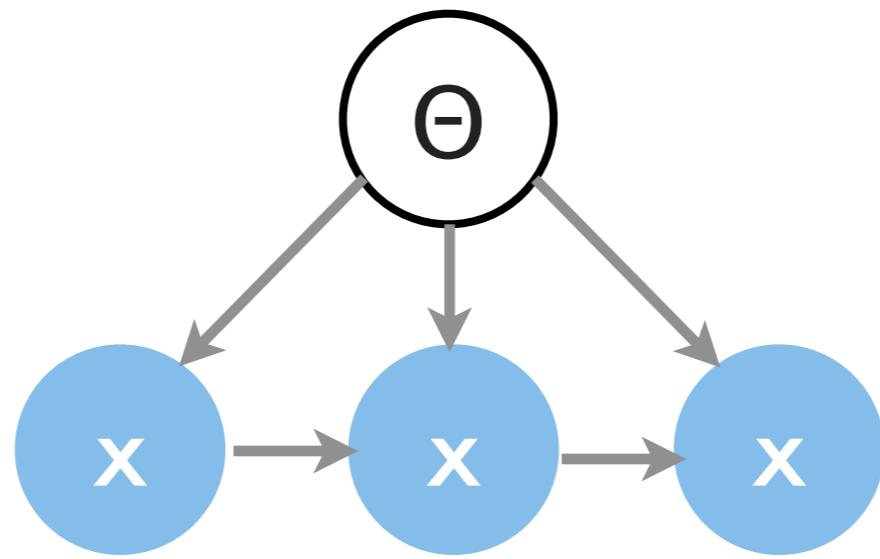
Classification
Regression

spam filtering
tiering
crawling
categorization
bid estimation
tagging



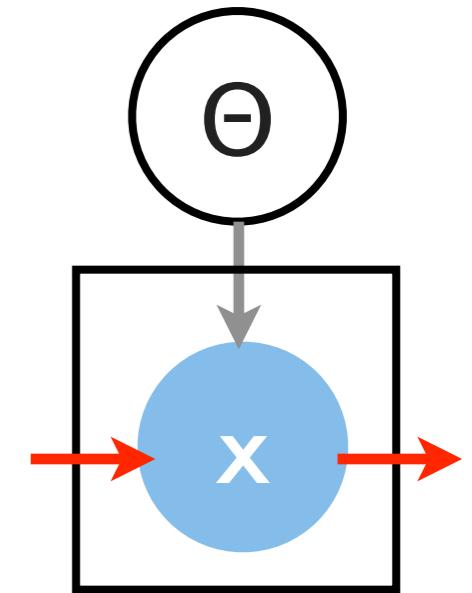
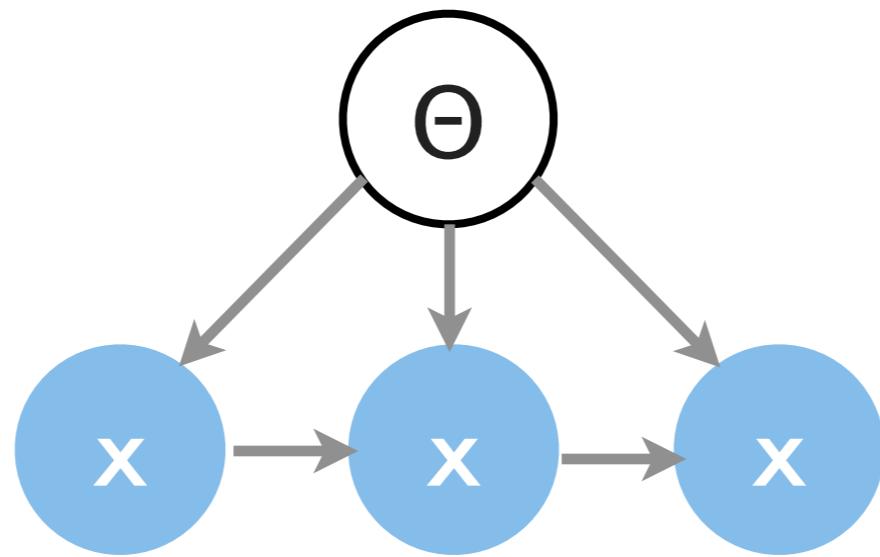
Chains

Markov
Chain

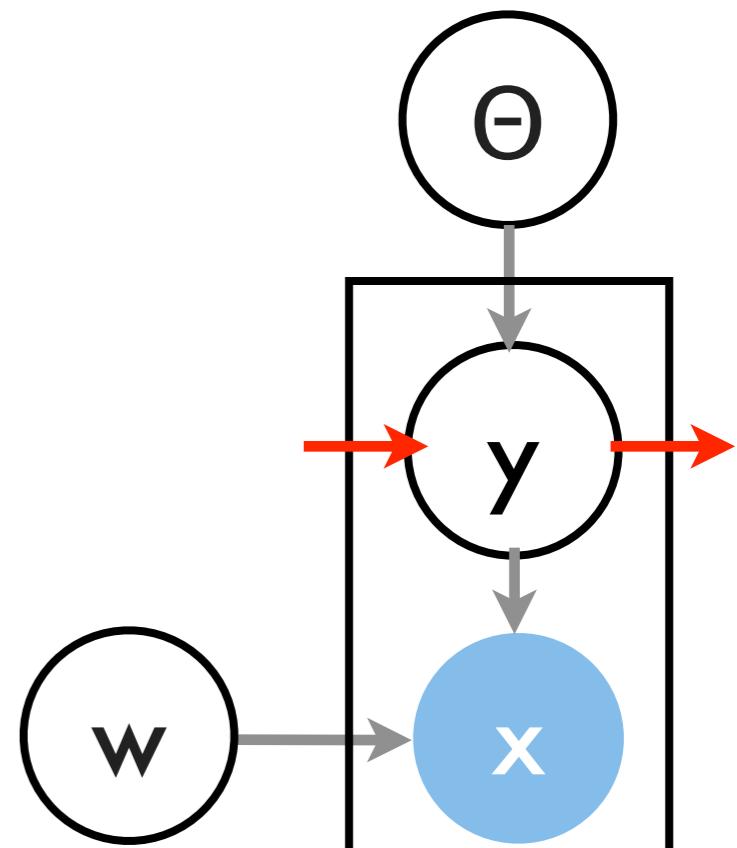
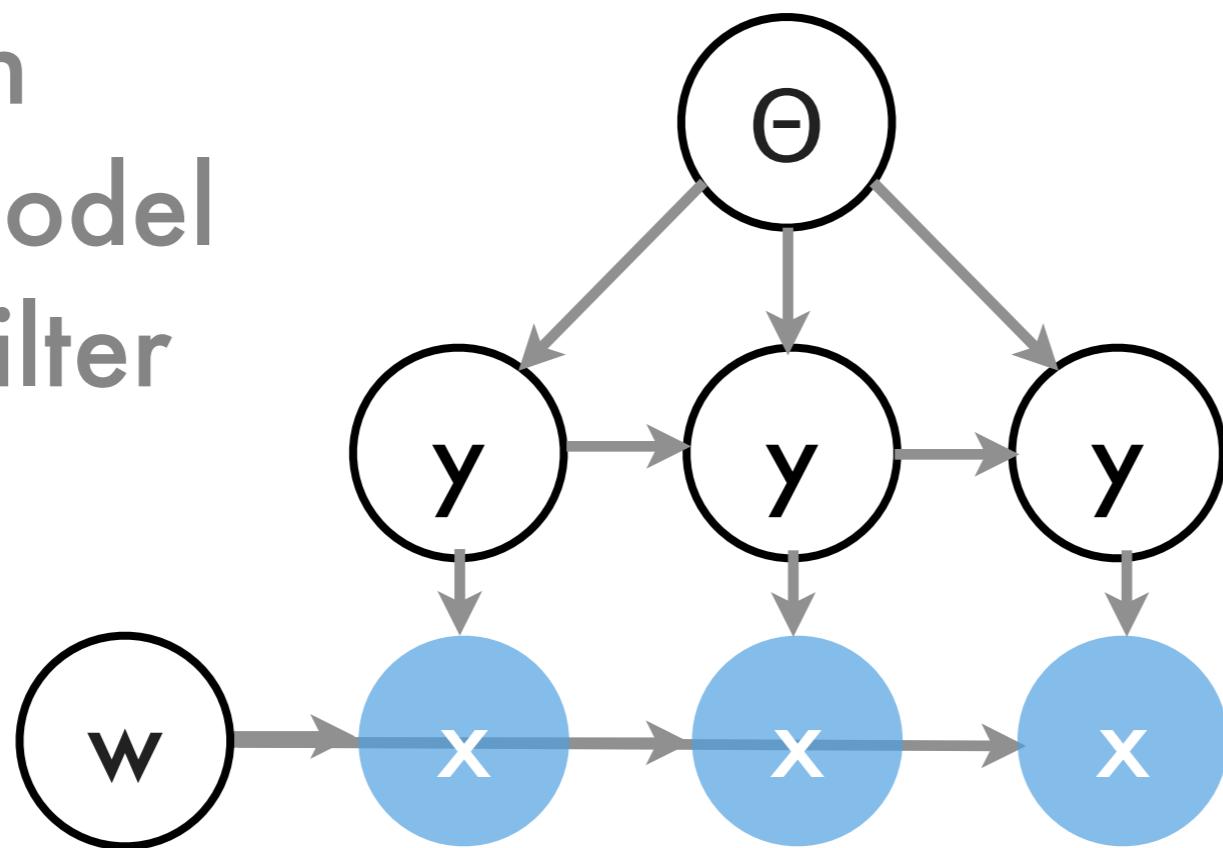


Chains

Markov
Chain



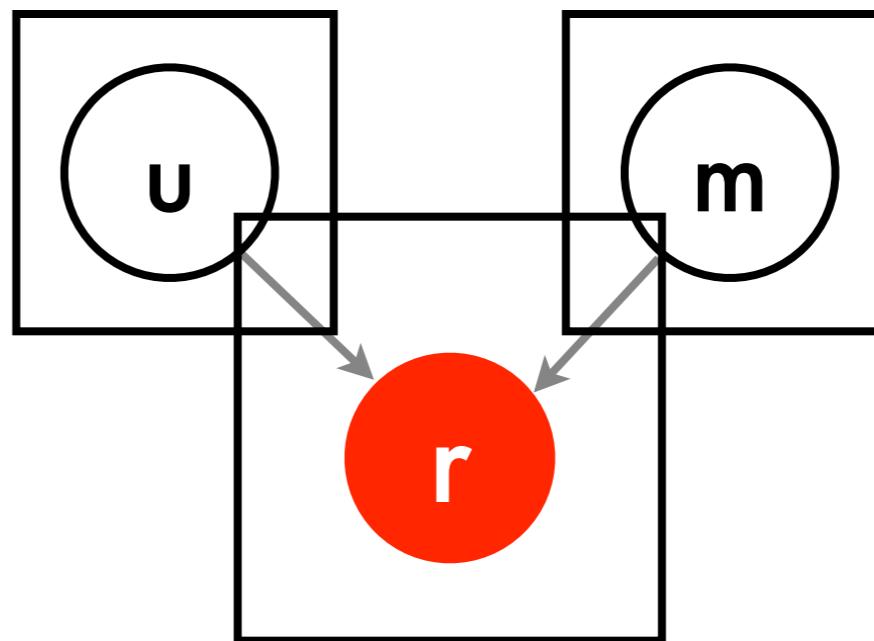
Hidden
Markov Model
Kalman Filter



Collaborative Models

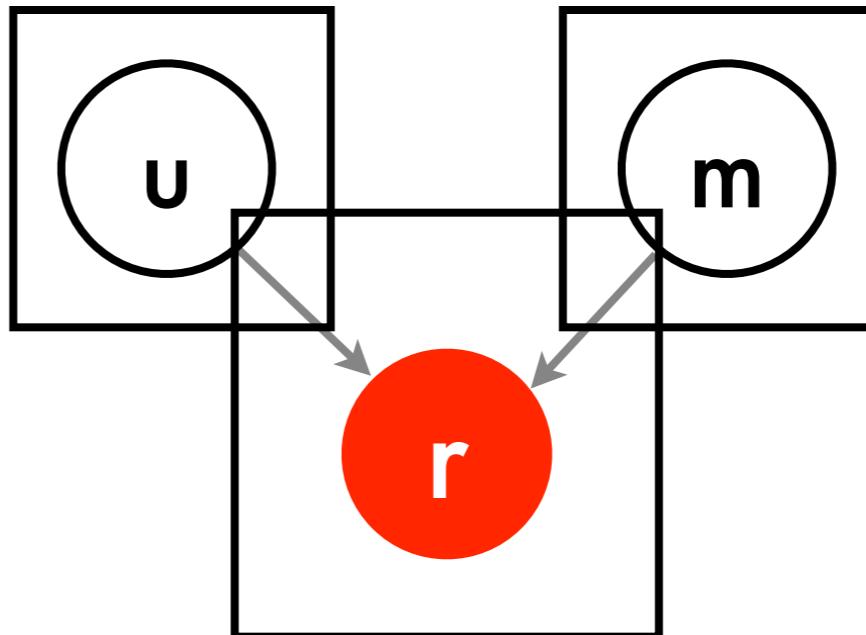
Collaborative Models

Collaborative
Filtering

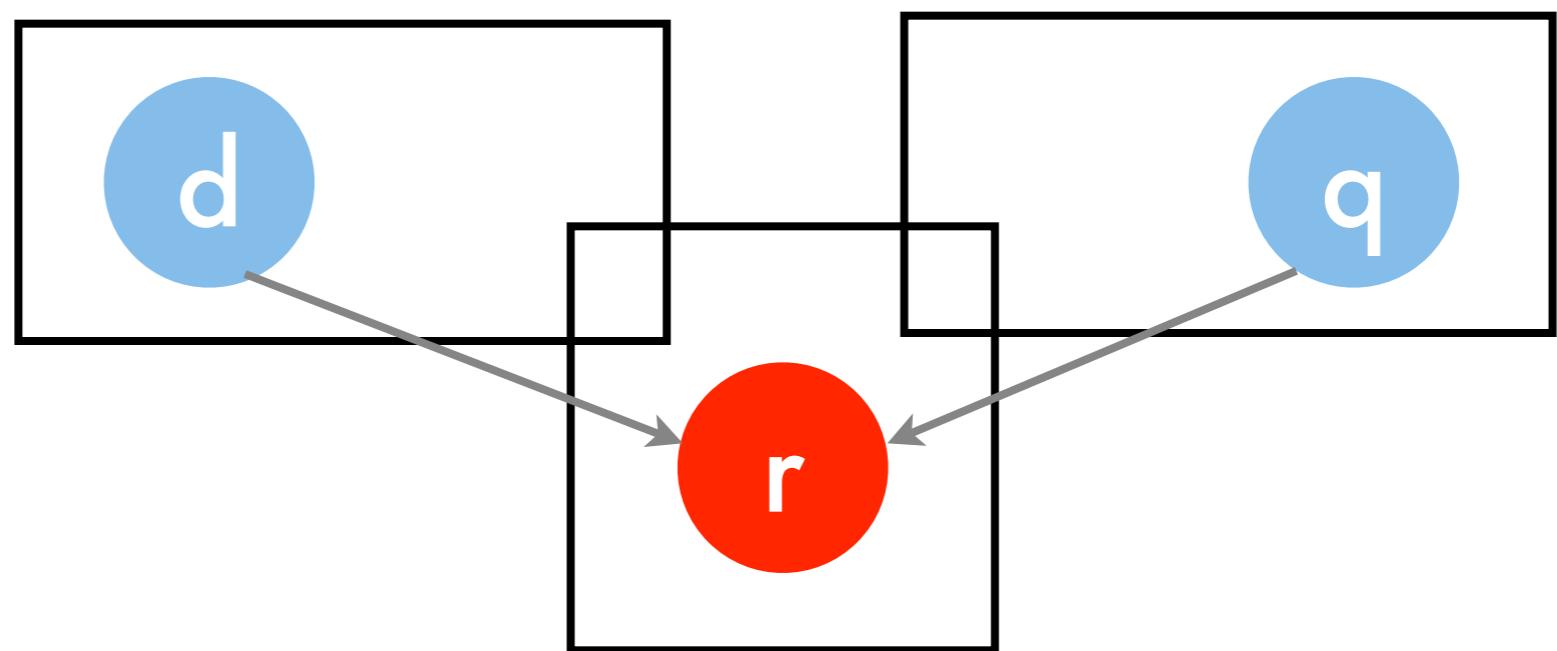


Collaborative Models

Collaborative
Filtering

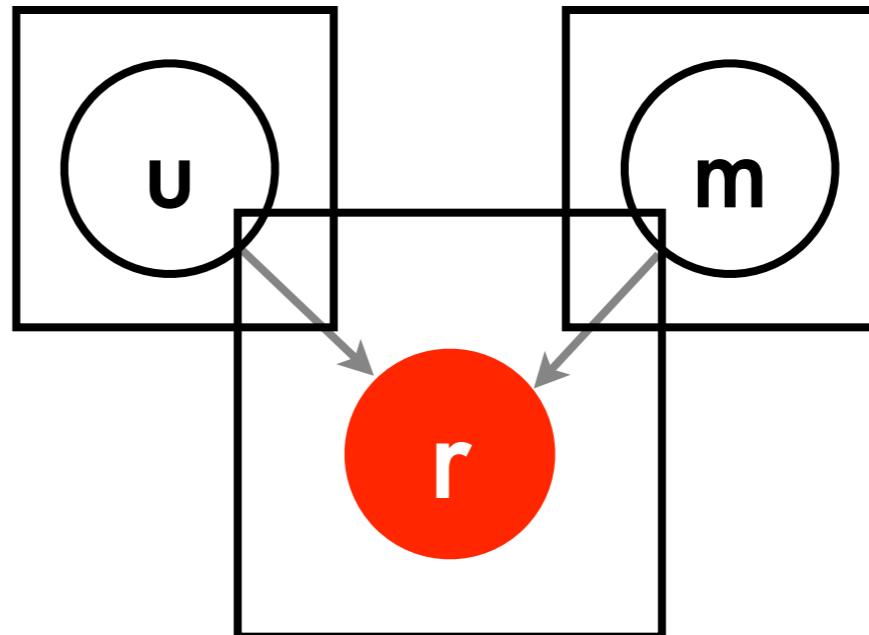


Current
Webpage
Ranking

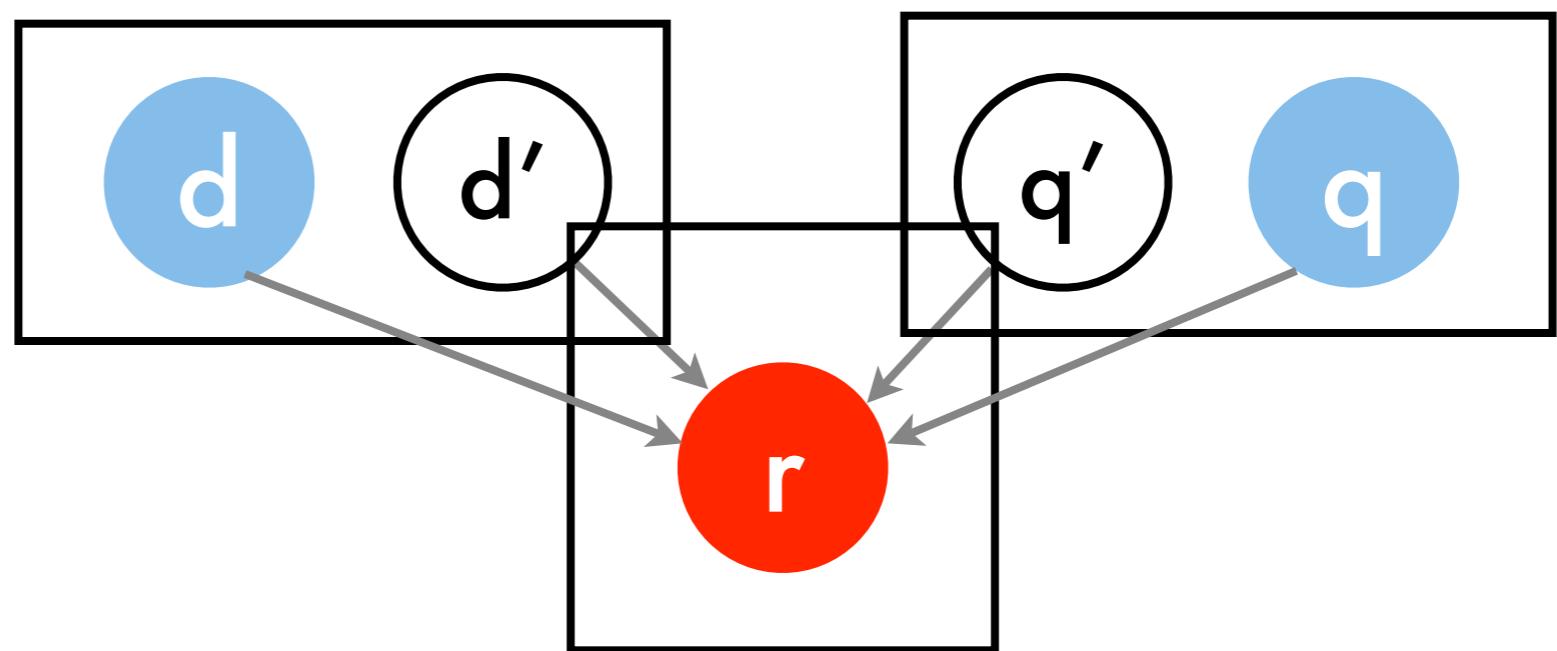


Collaborative Models

Collaborative
Filtering

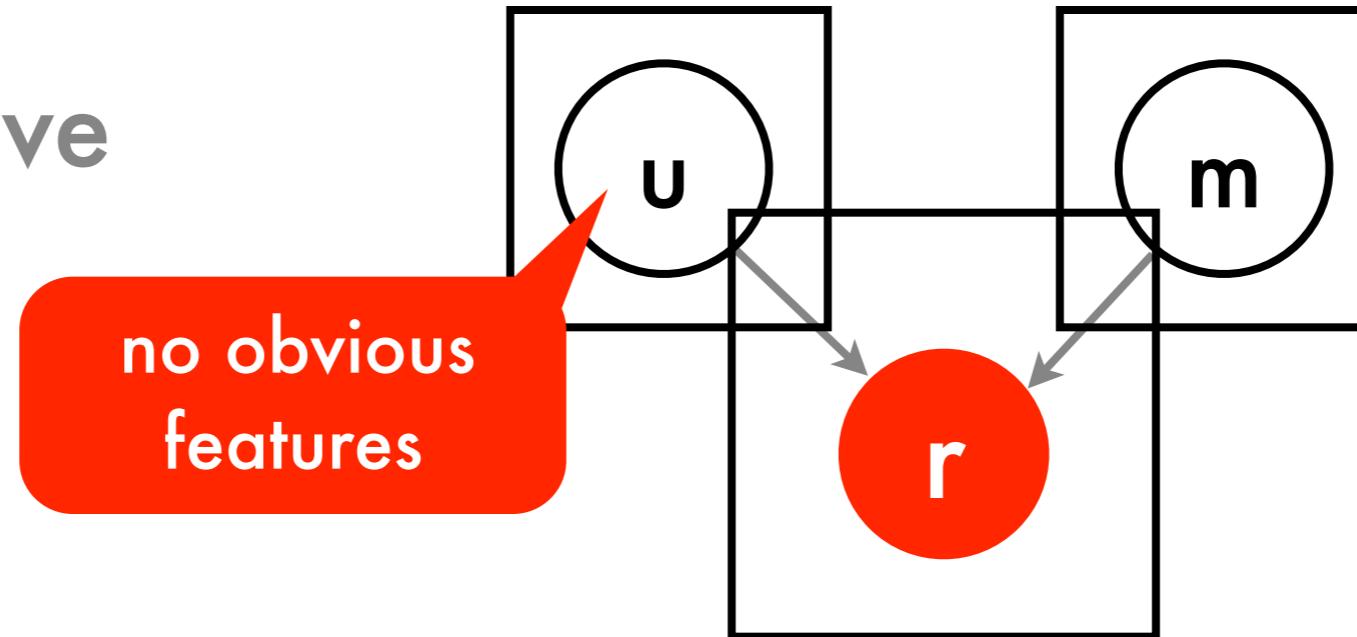


Webpage
Ranking

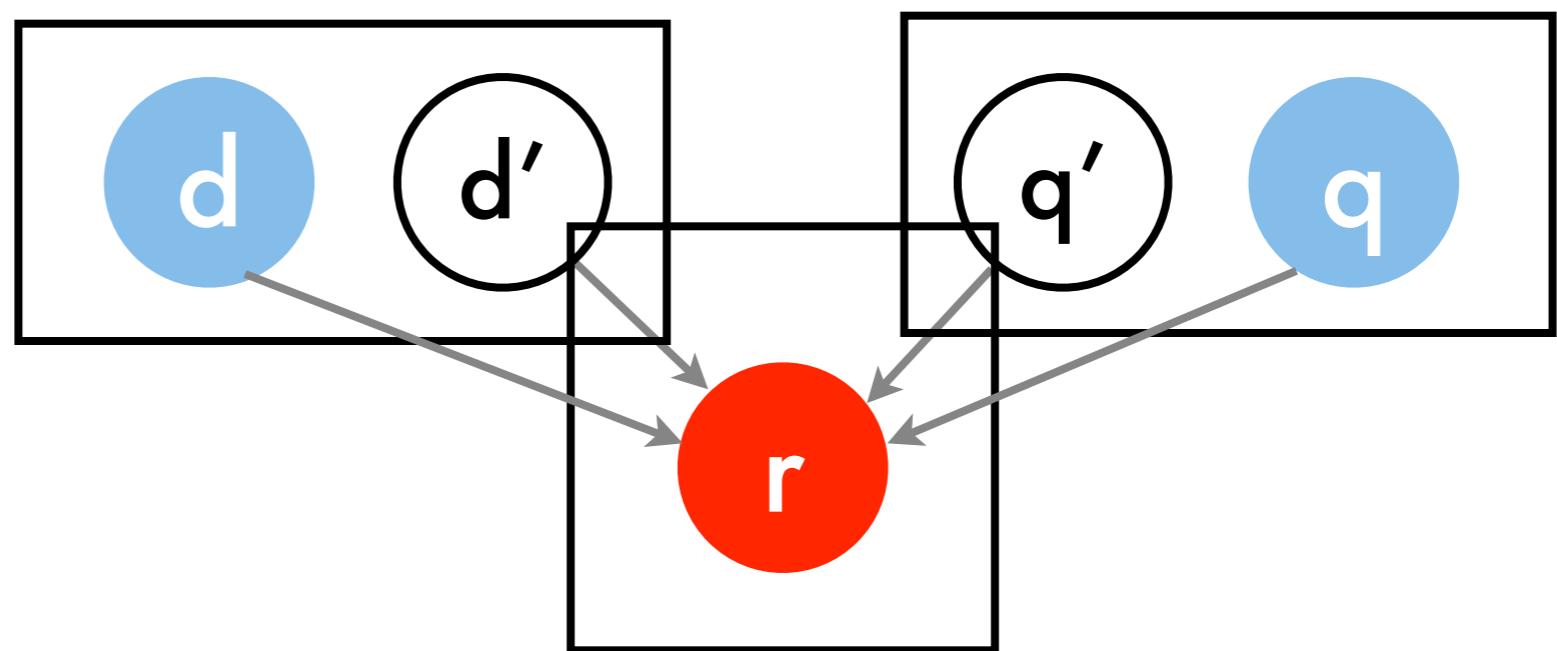


Collaborative Models

Collaborative
Filtering

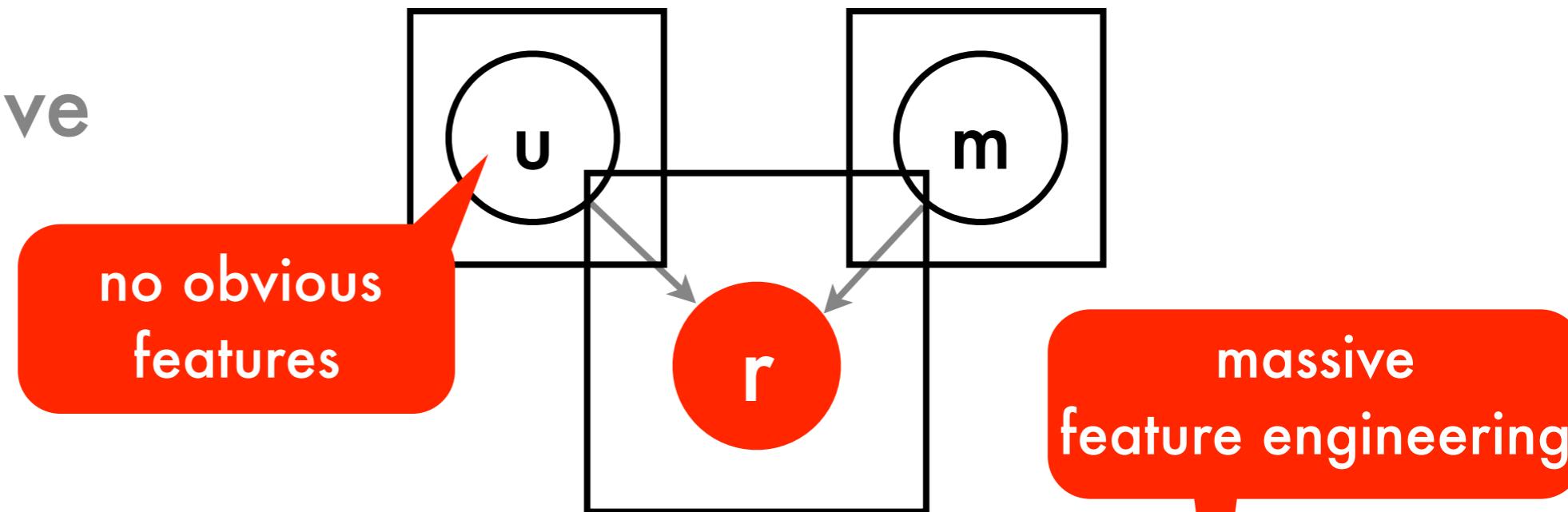


Webpage
Ranking

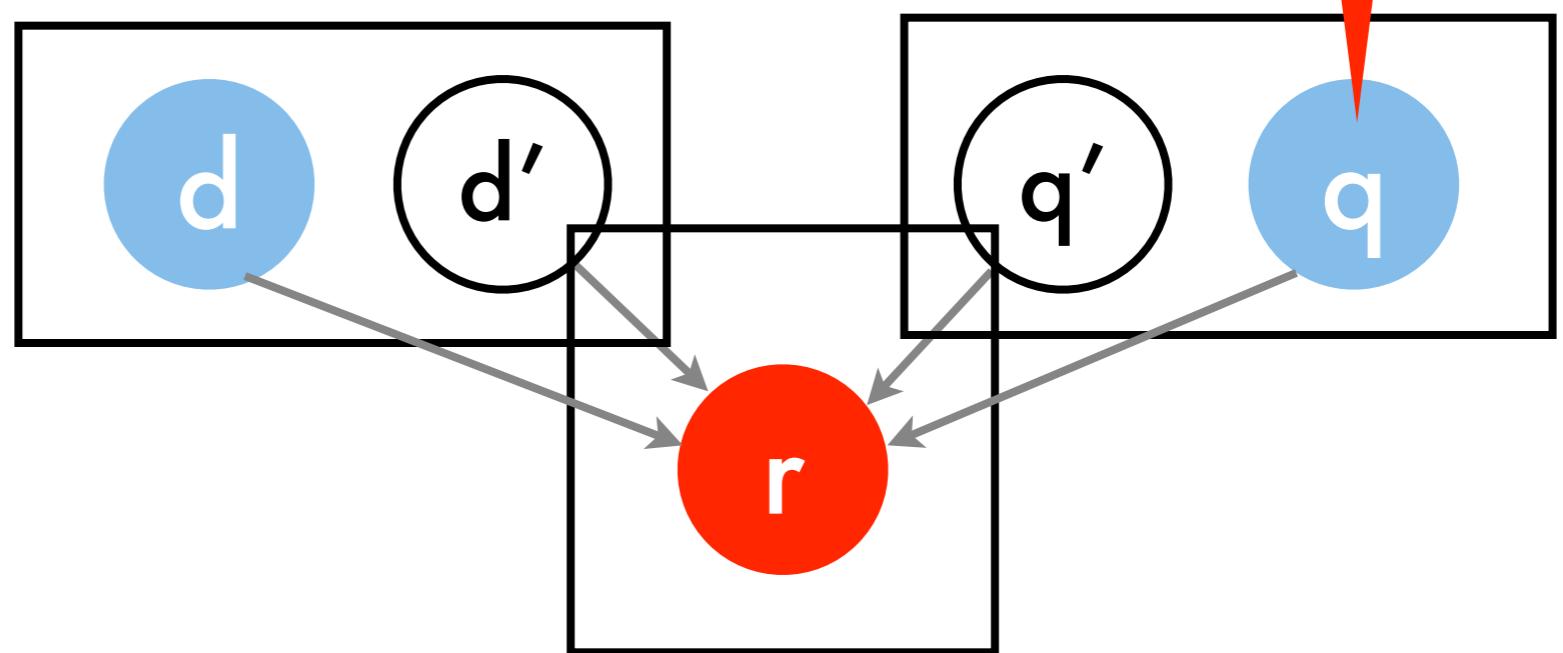


Collaborative Models

Collaborative
Filtering

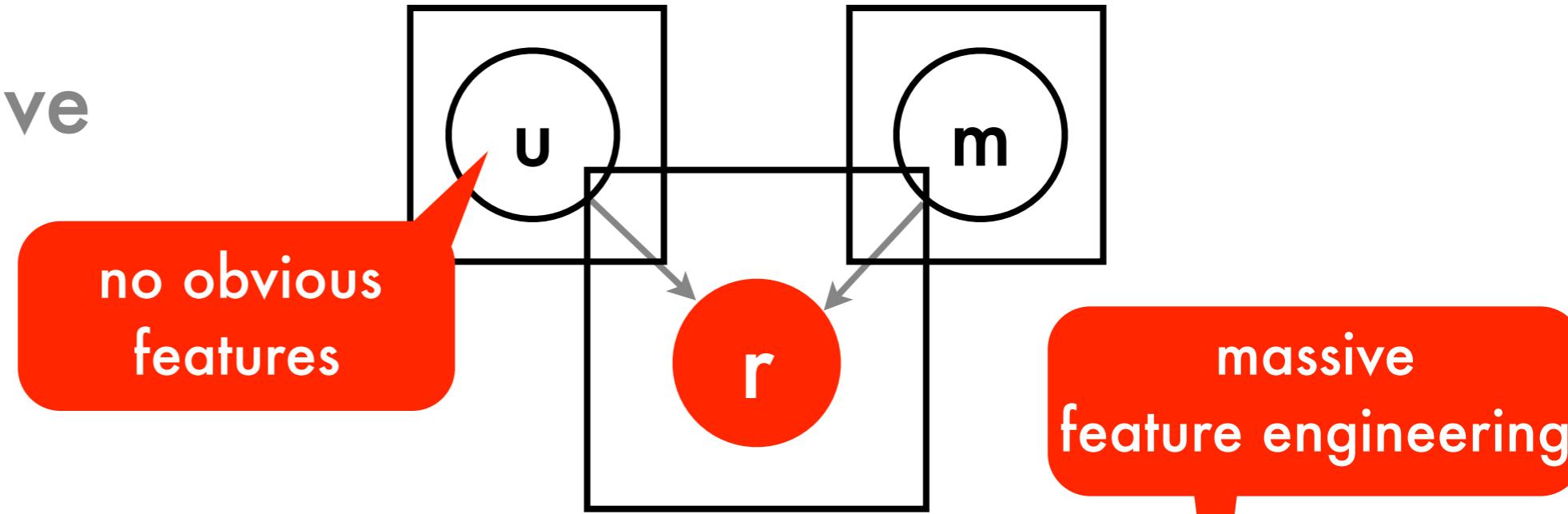


Webpage
Ranking

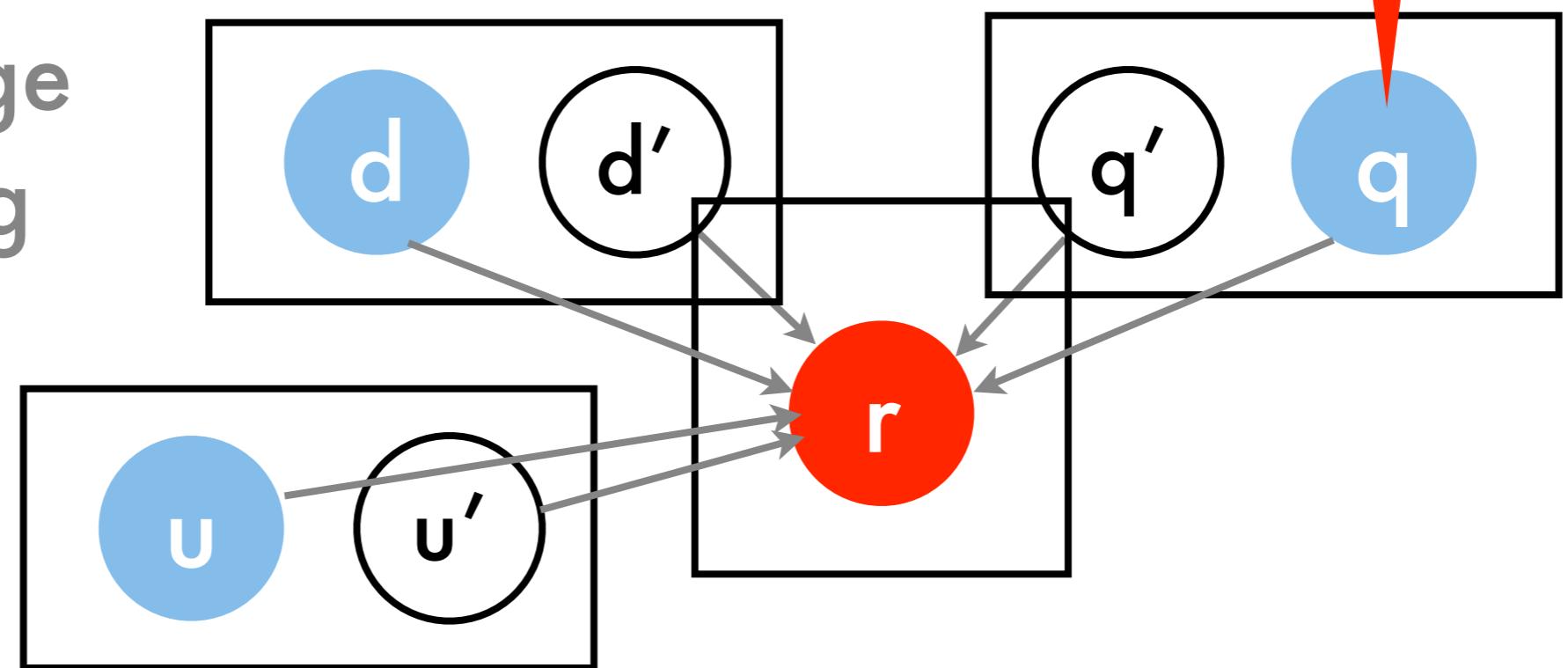


Collaborative Models

Collaborative
Filtering

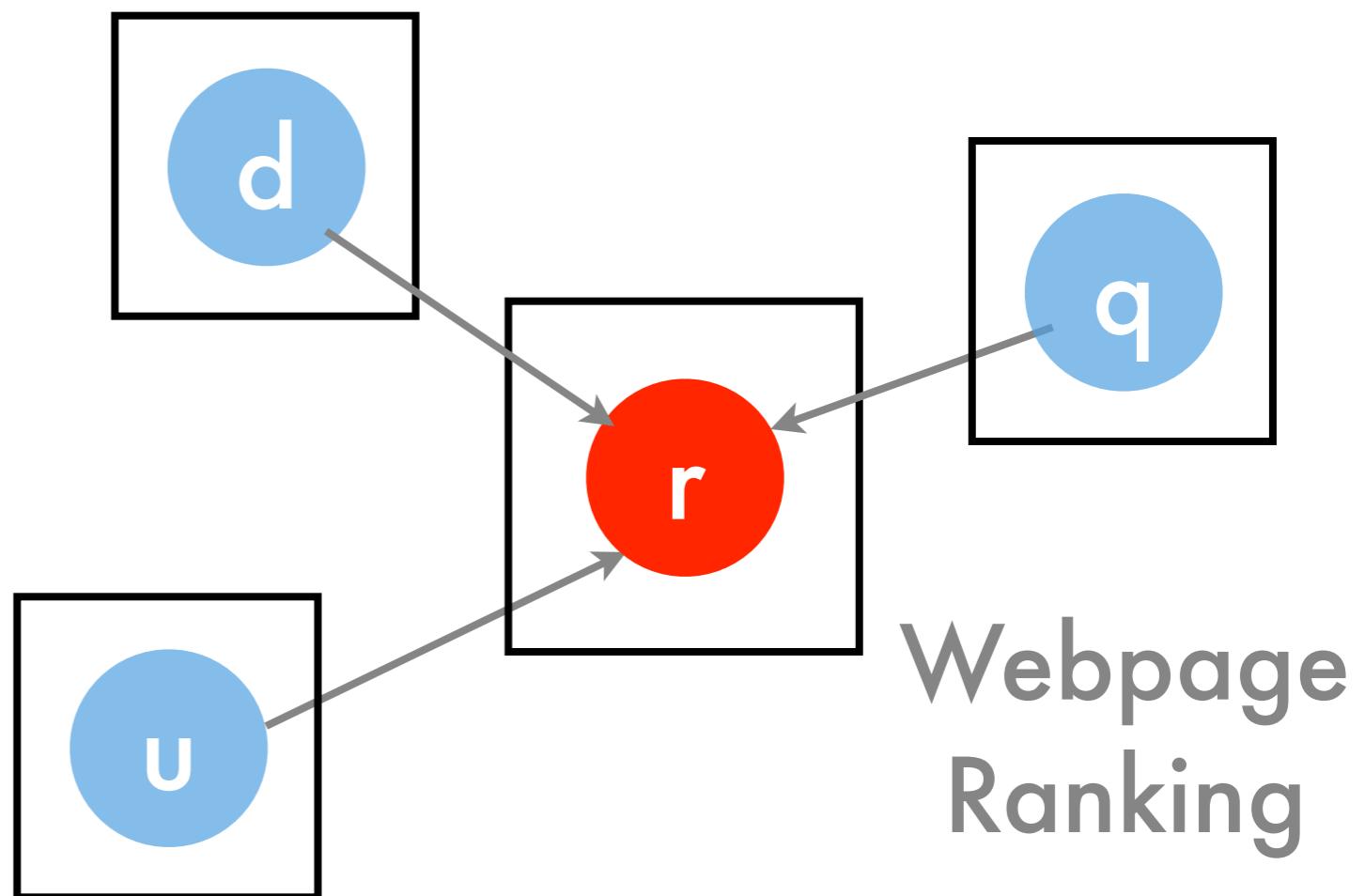


Webpage
Ranking

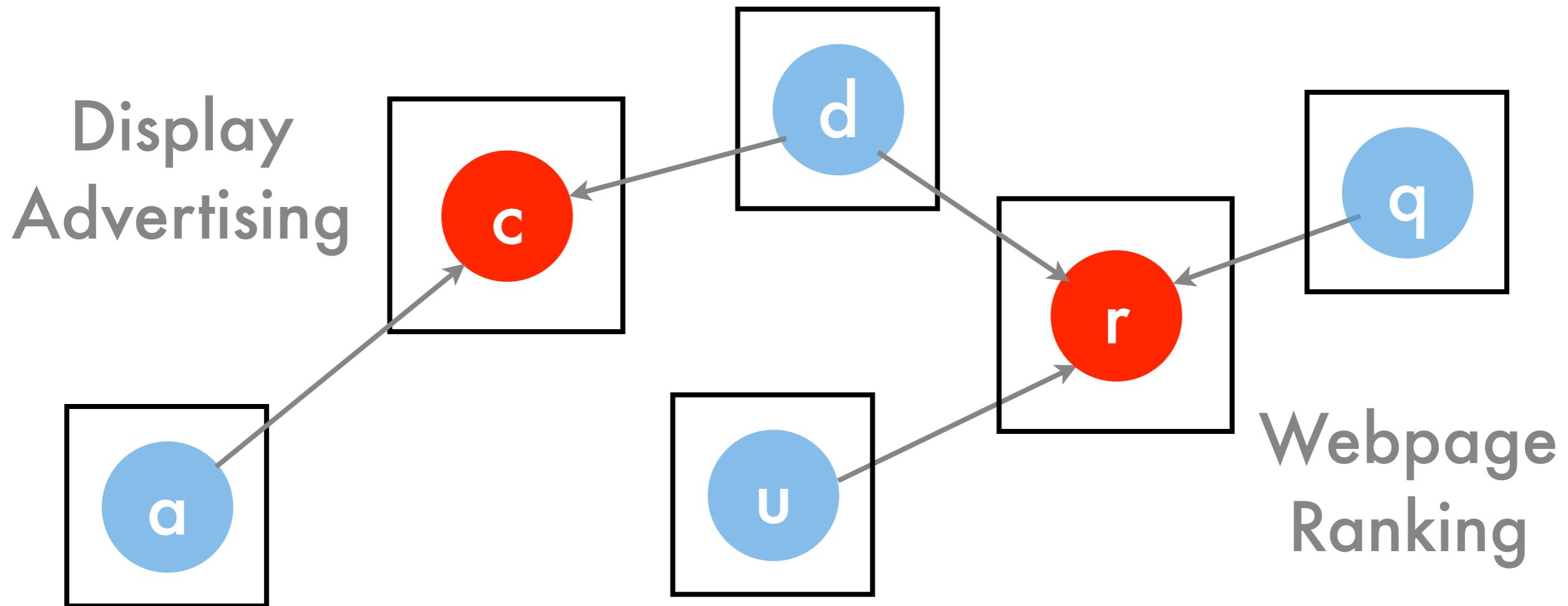


personalized

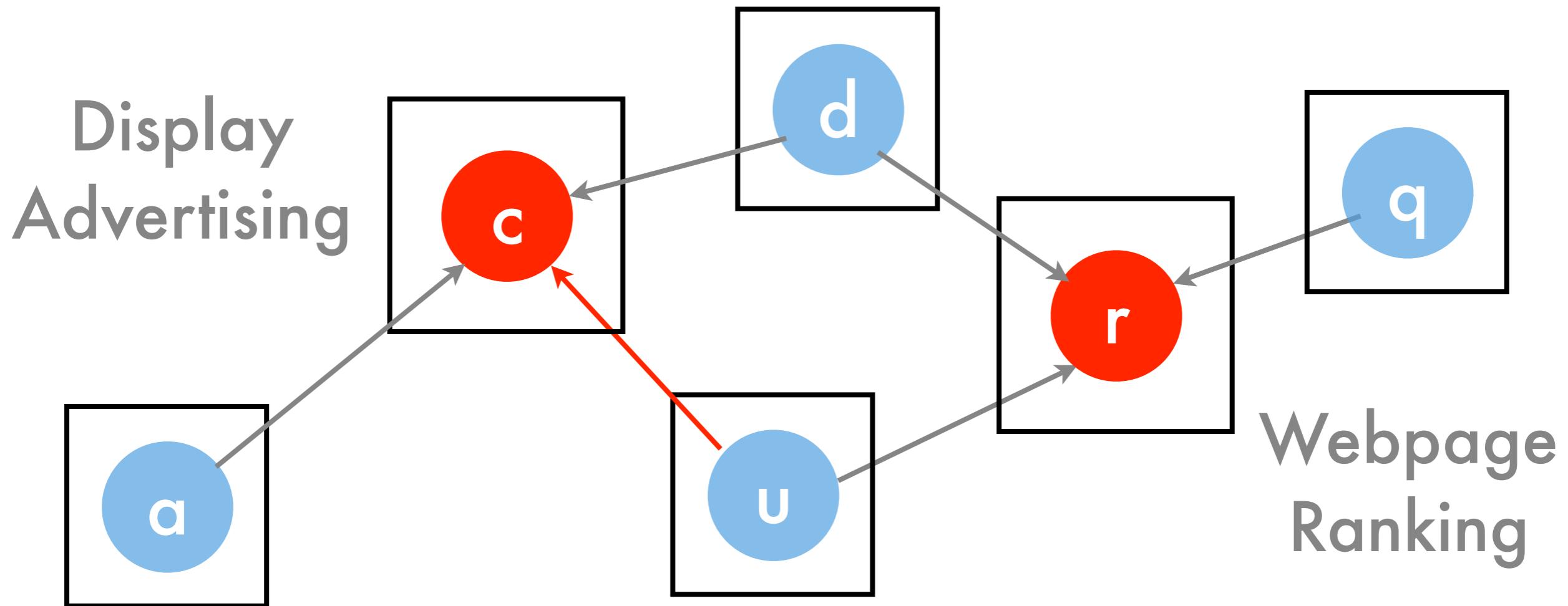
Data Integration



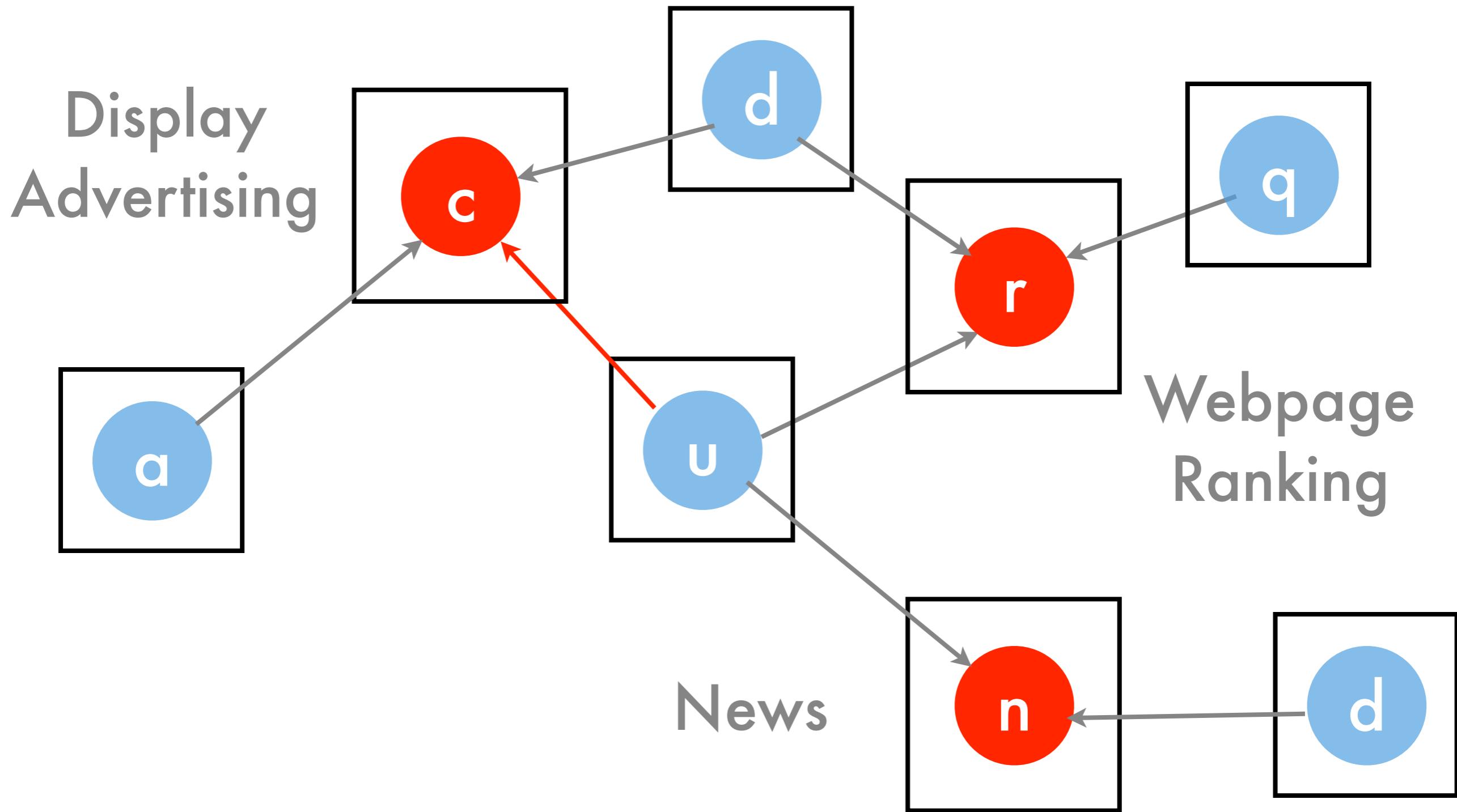
Data Integration



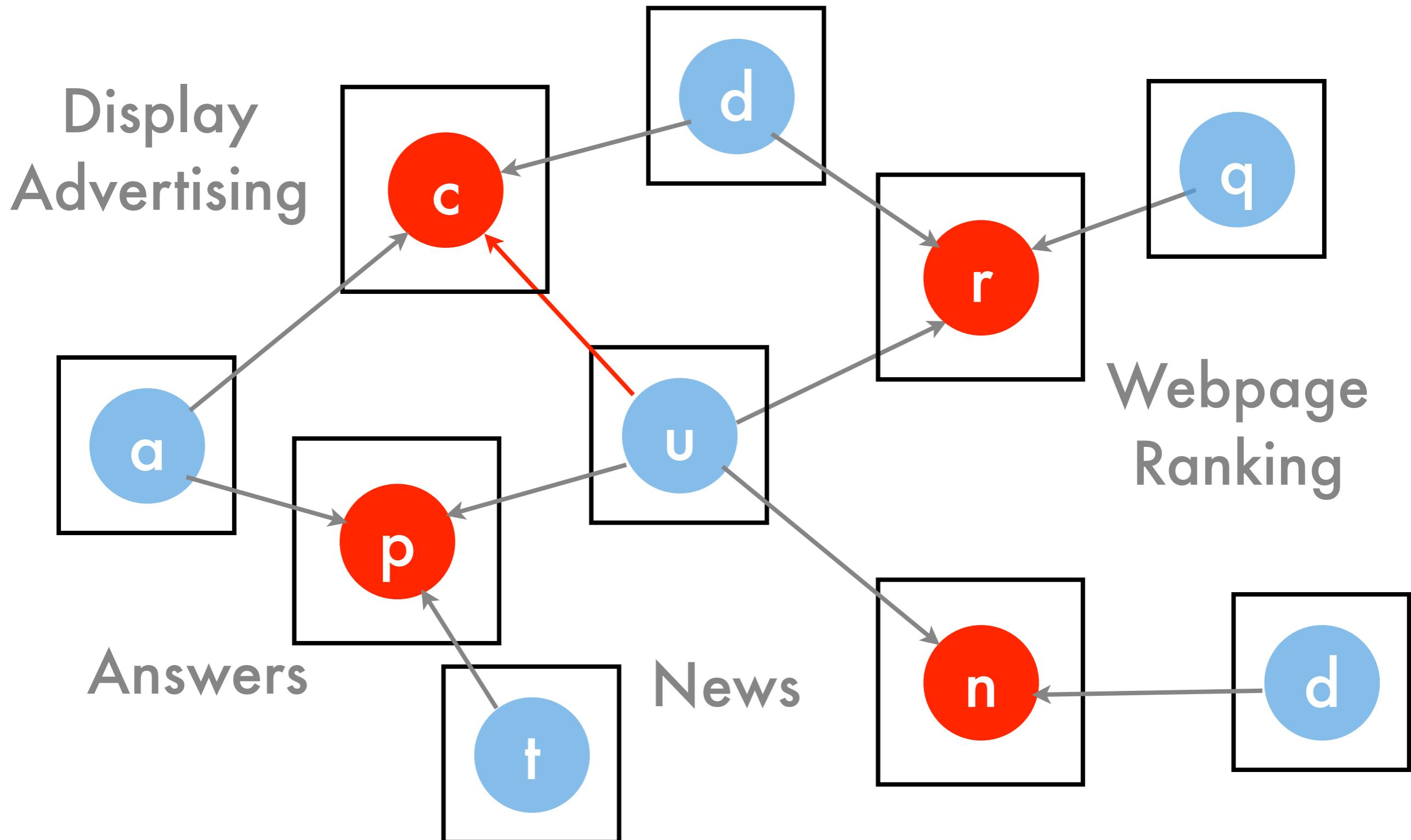
Data Integration



Data Integration



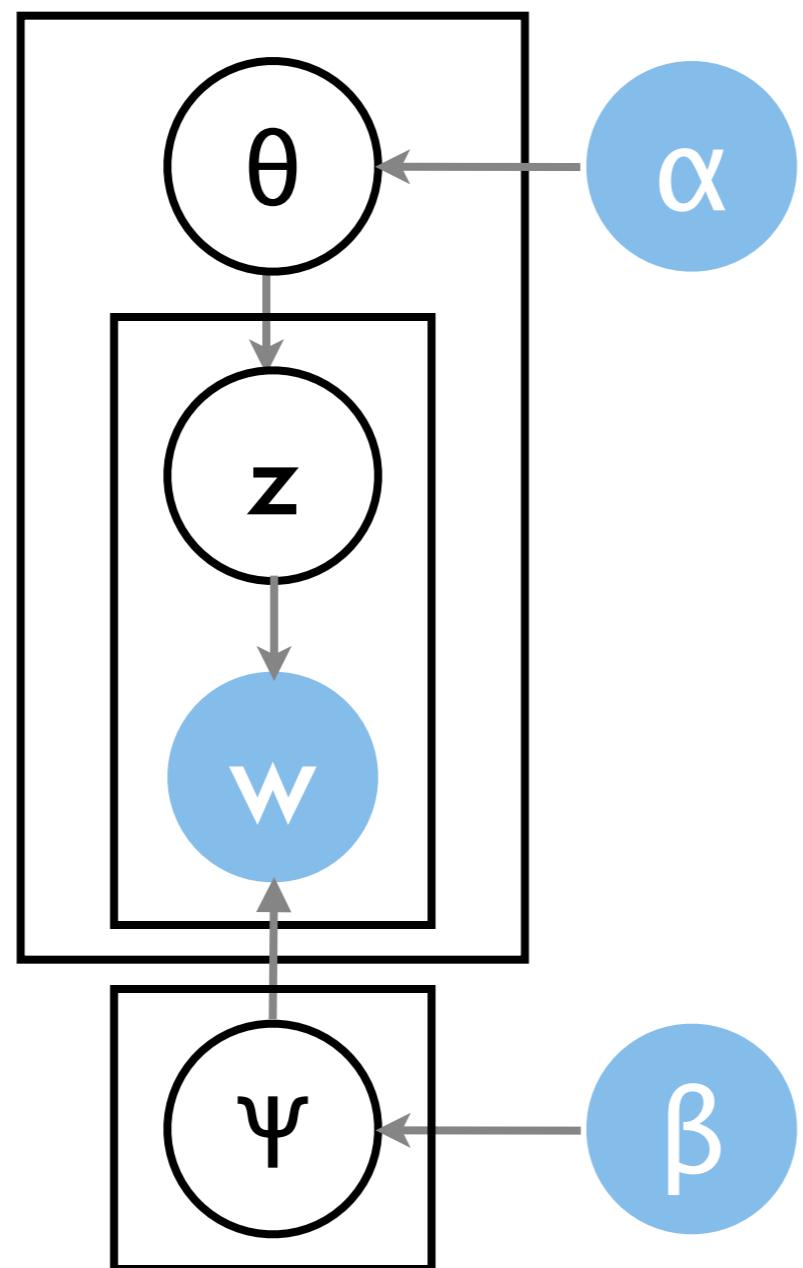
Data Integration



Topic Models

Topic Models

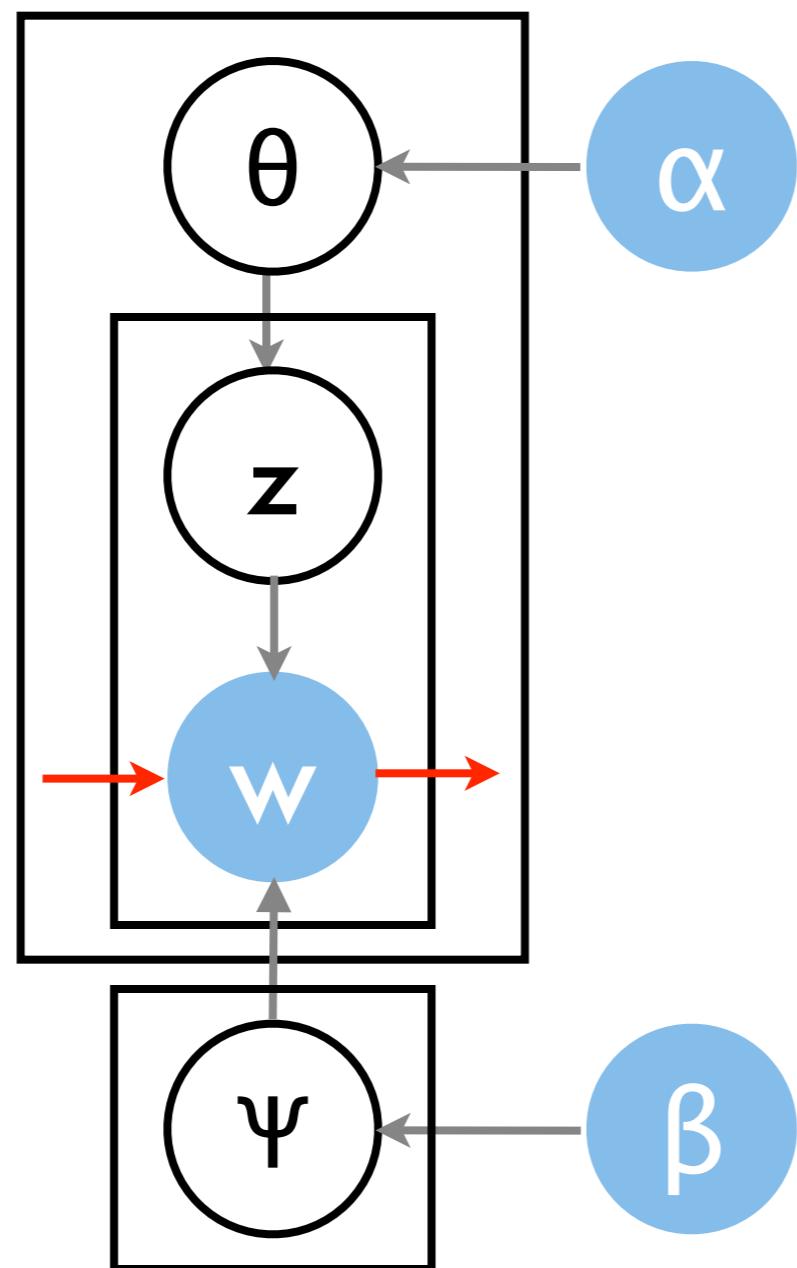
Topic
Models



Topic Models

Topic
Models

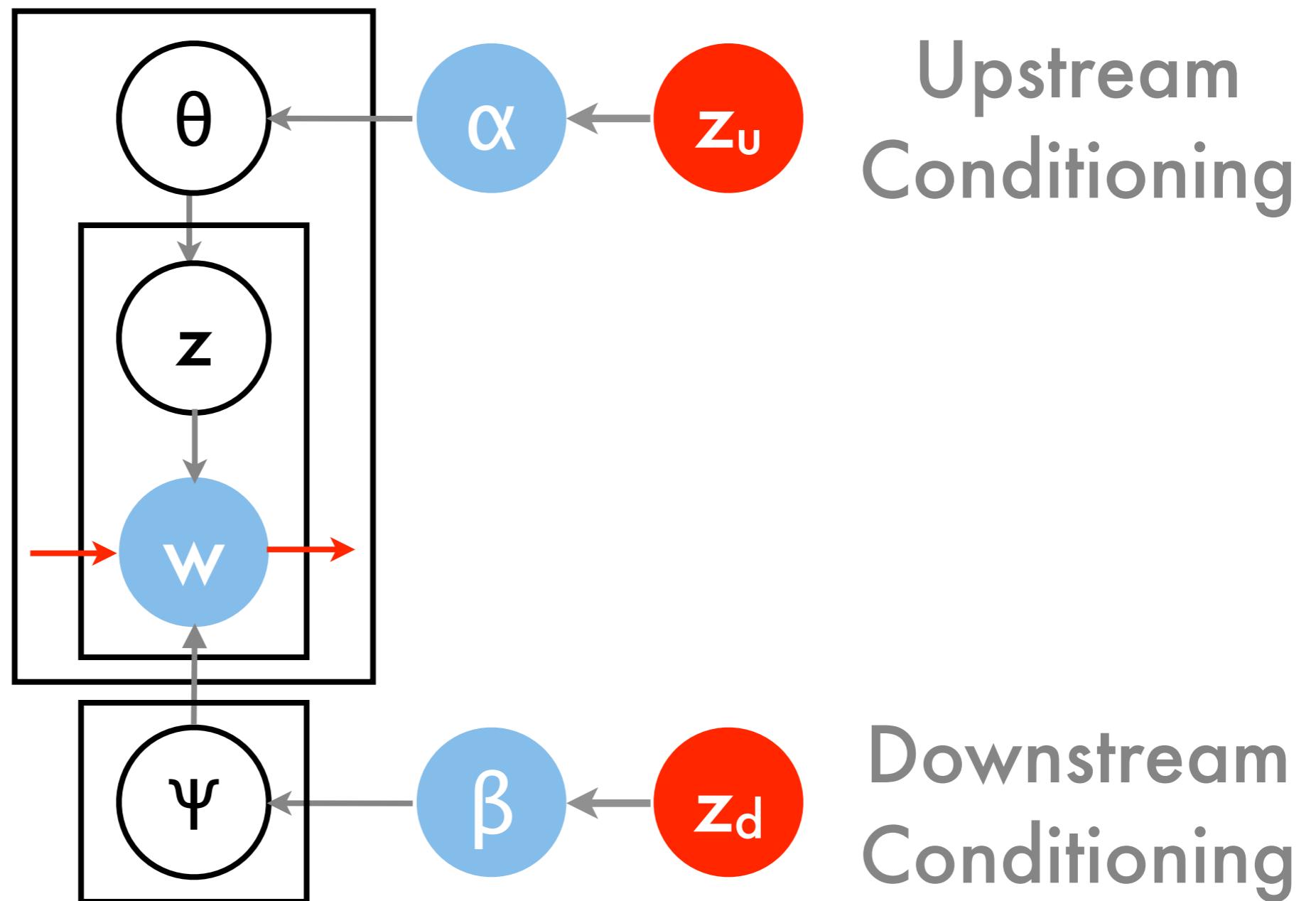
Simplical
Mixtures



Topic Models

Topic
Models

Simplical
Mixtures



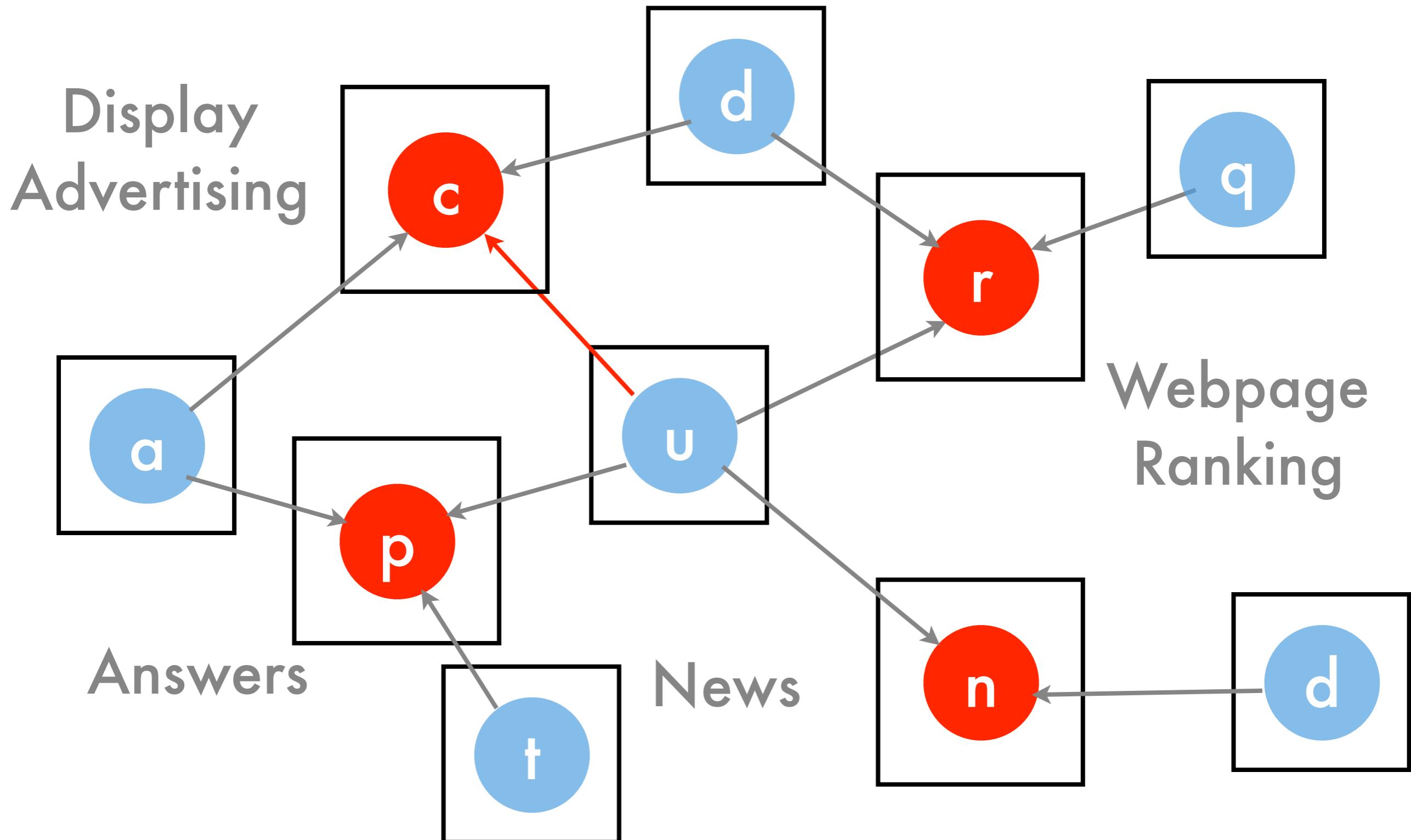
Undirected Graphical Models

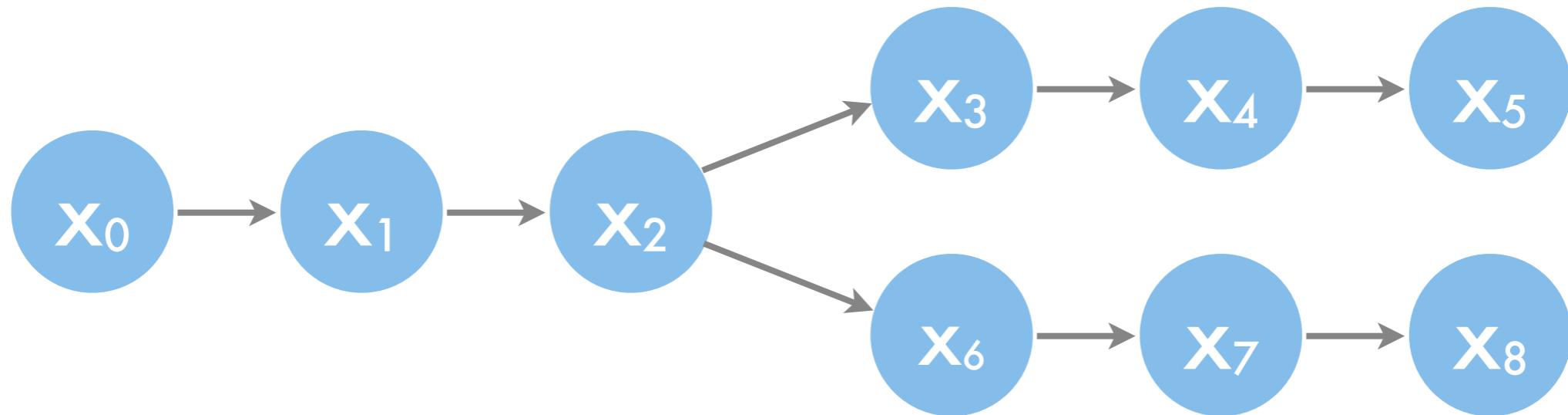
Review

YAHOO!

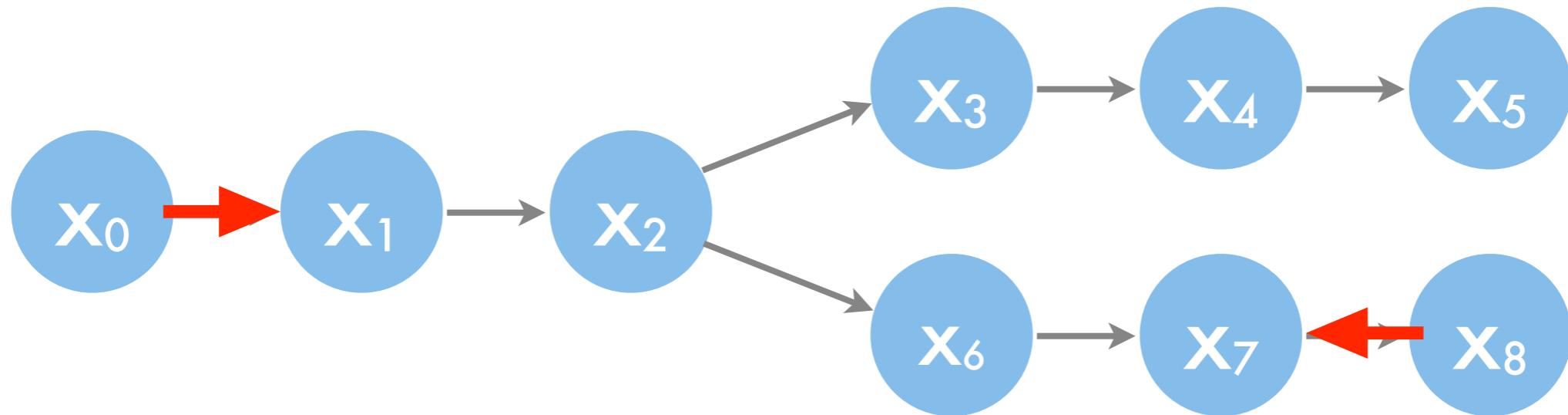


Data Integration

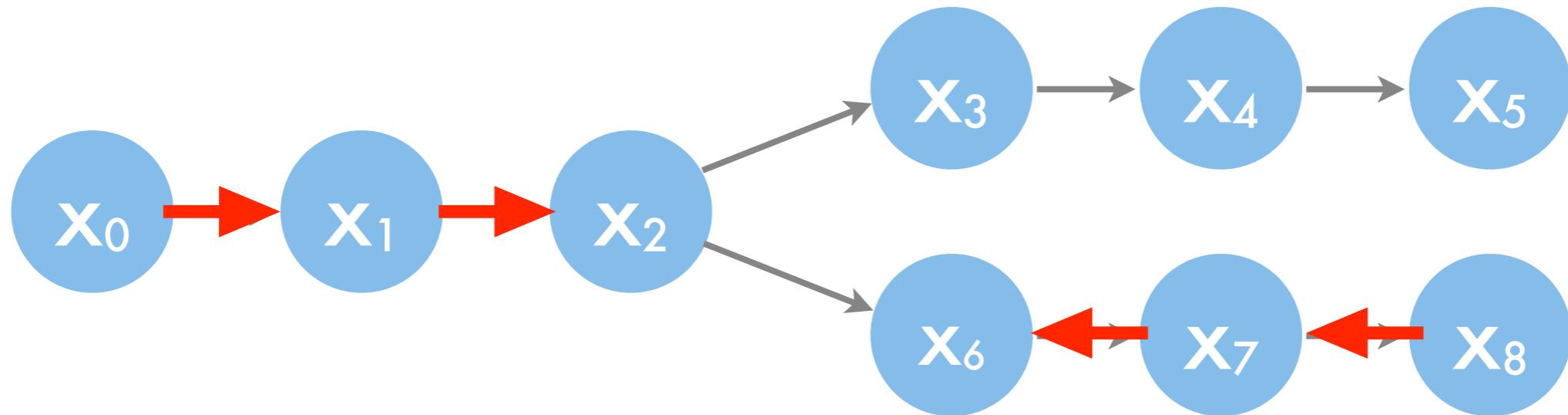




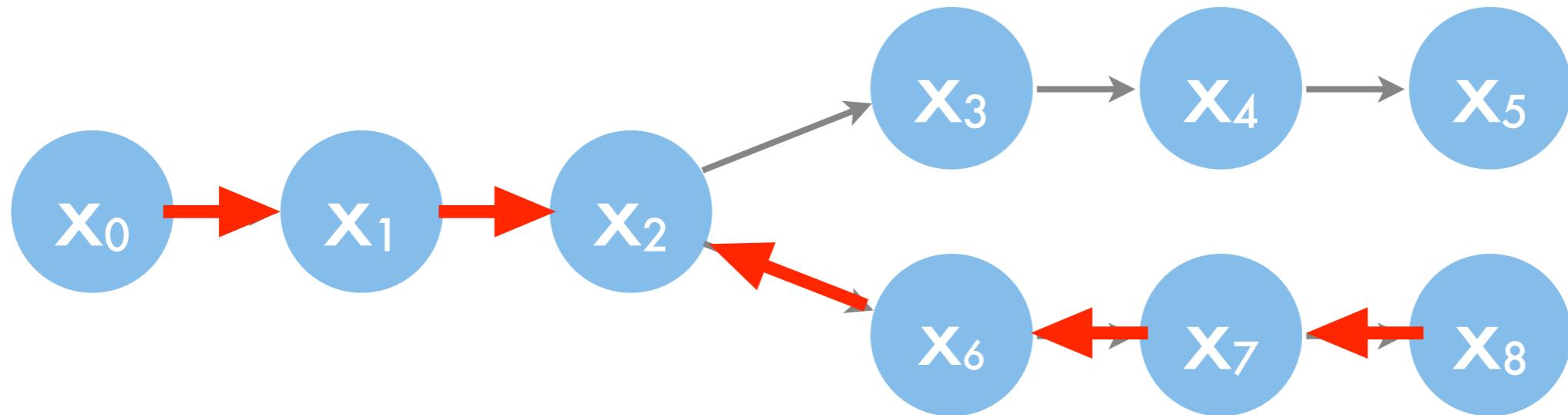
- Forward/Backward messages as normal for chain
- When we have more edges for a vertex use
 - For each outgoing message, send it once you have all other incoming messages
 - **PRINCIPLED HACK**
If no message received yet, set it to 1 altogether



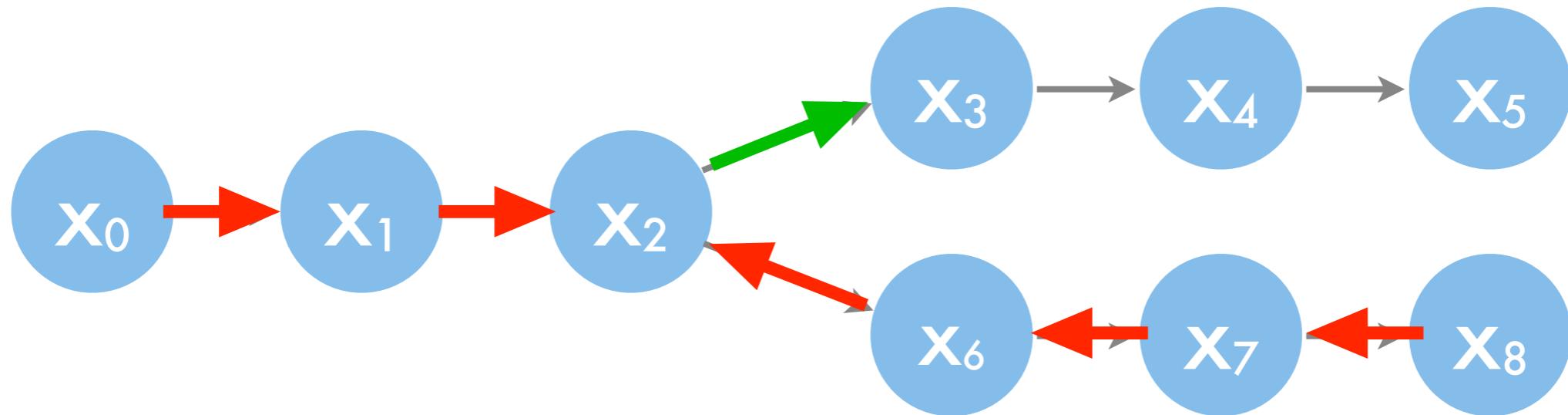
- Forward/Backward messages as normal for chain
- When we have more edges for a vertex use
 - For each outgoing message, send it once you have all other incoming messages
 - **PRINCIPLED HACK**
If no message received yet, set it to 1 altogether



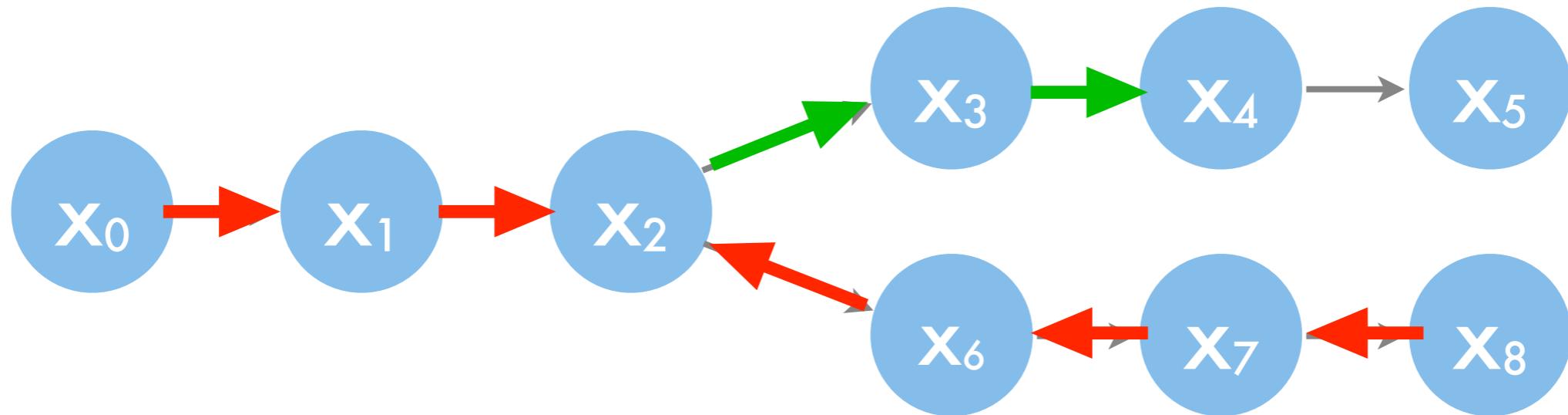
- Forward/Backward messages as normal for chain
- When we have more edges for a vertex use
 - For each outgoing message, send it once you have all other incoming messages
 - **PRINCIPLED HACK**
If no message received yet, set it to 1 altogether



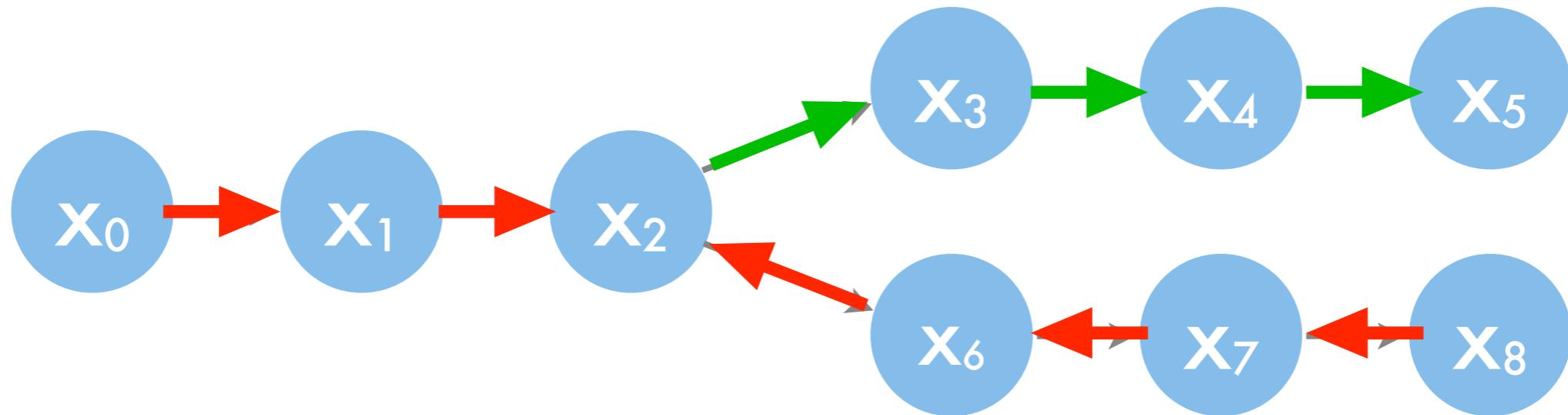
- Forward/Backward messages as normal for chain
- When we have more edges for a vertex use
 - For each outgoing message, send it once you have all other incoming messages
 - **PRINCIPLED HACK**
If no message received yet, set it to 1 altogether



- Forward/Backward messages as normal for chain
- When we have more edges for a vertex use
 - For each outgoing message, send it once you have all other incoming messages
 - **PRINCIPLED HACK**
If no message received yet, set it to 1 altogether



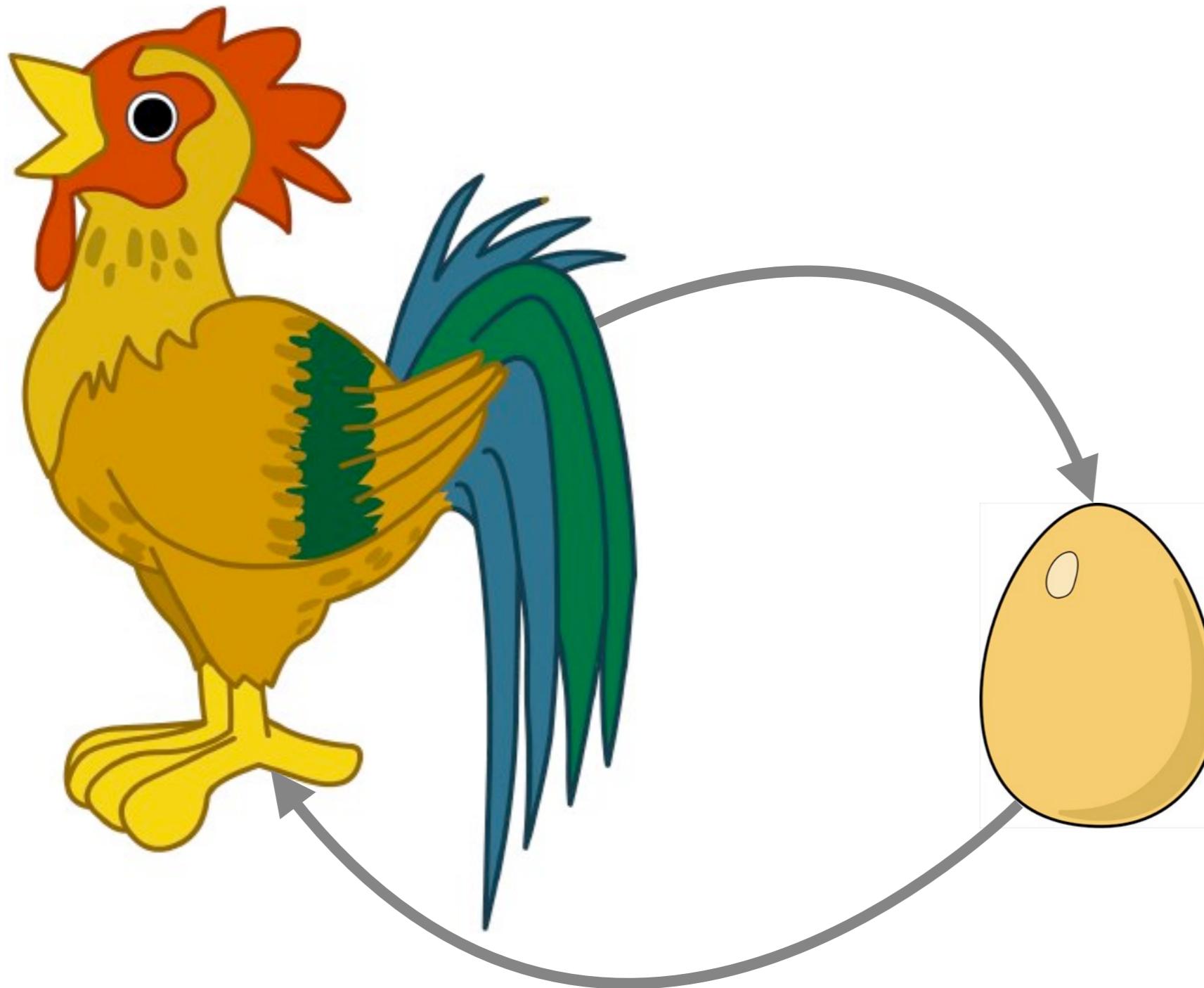
- Forward/Backward messages as normal for chain
- When we have more edges for a vertex use
 - For each outgoing message, send it once you have all other incoming messages
 - **PRINCIPLED HACK**
If no message received yet, set it to 1 altogether



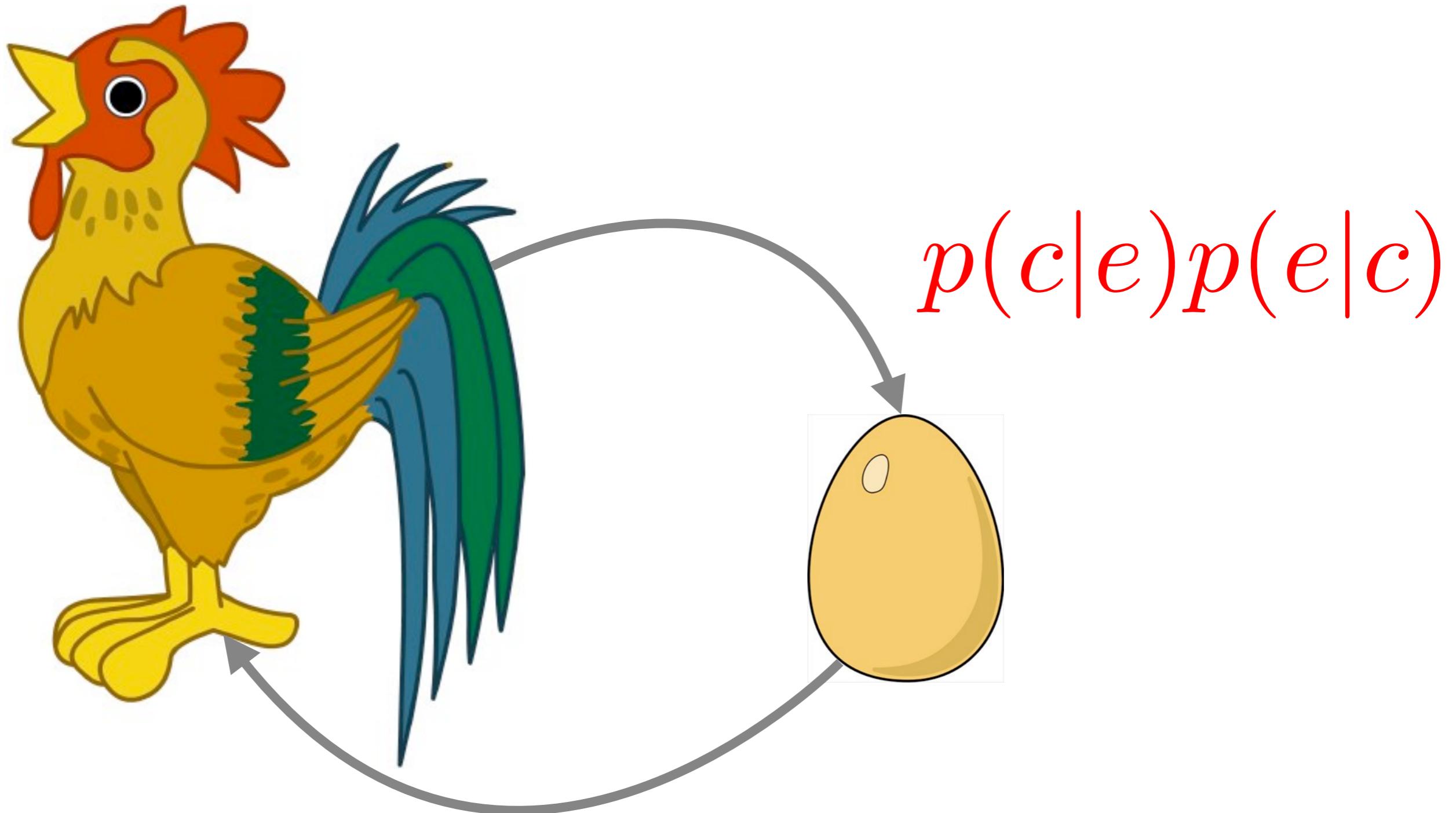
- Forward/Backward messages as normal for chain
- When we have more edges for a vertex use
 - For each outgoing message, send it once you have all other incoming messages
 - **PRINCIPLED HACK**
If no message received yet, set it to 1 altogether

Blunting the arrows ...

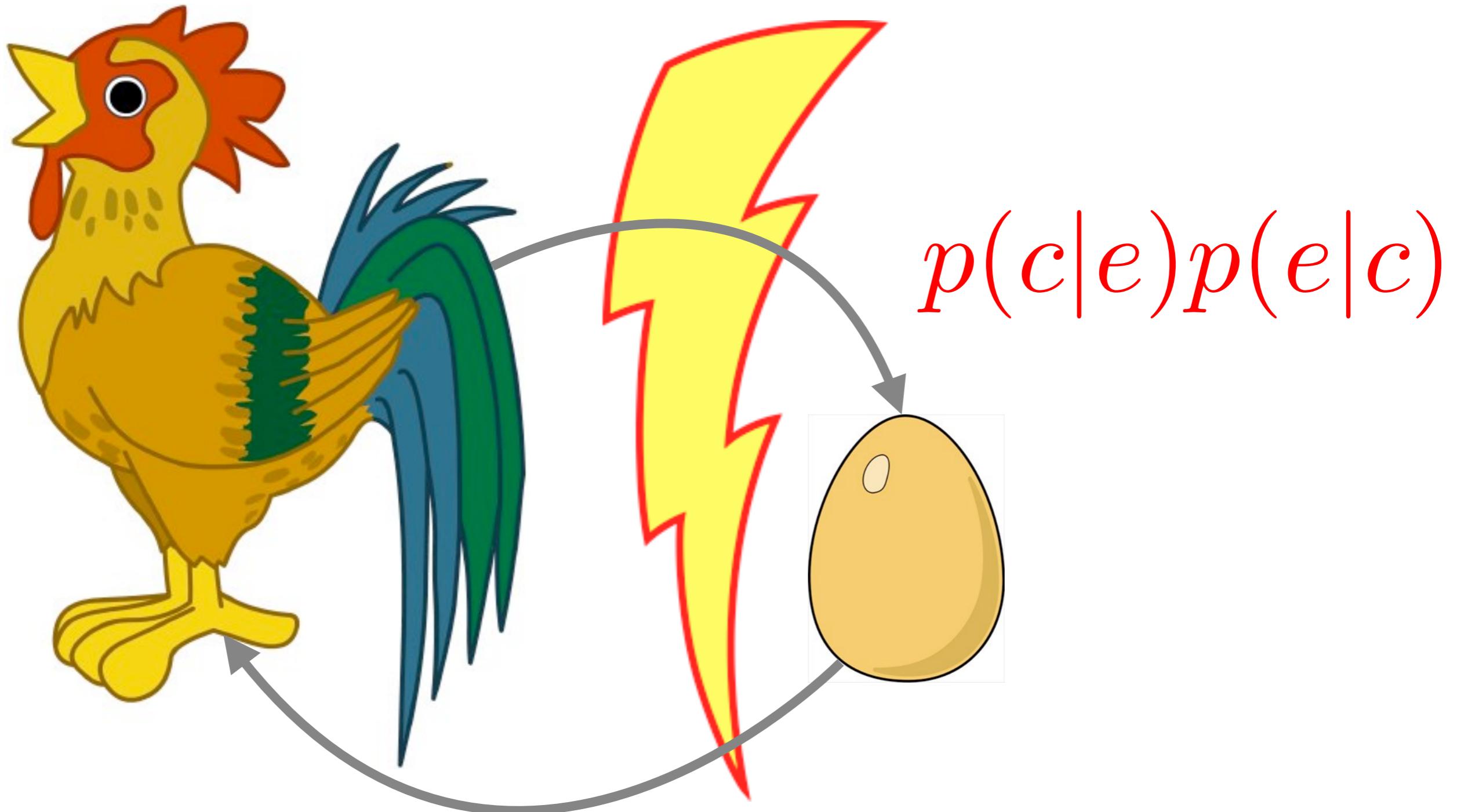
Chicken and Egg



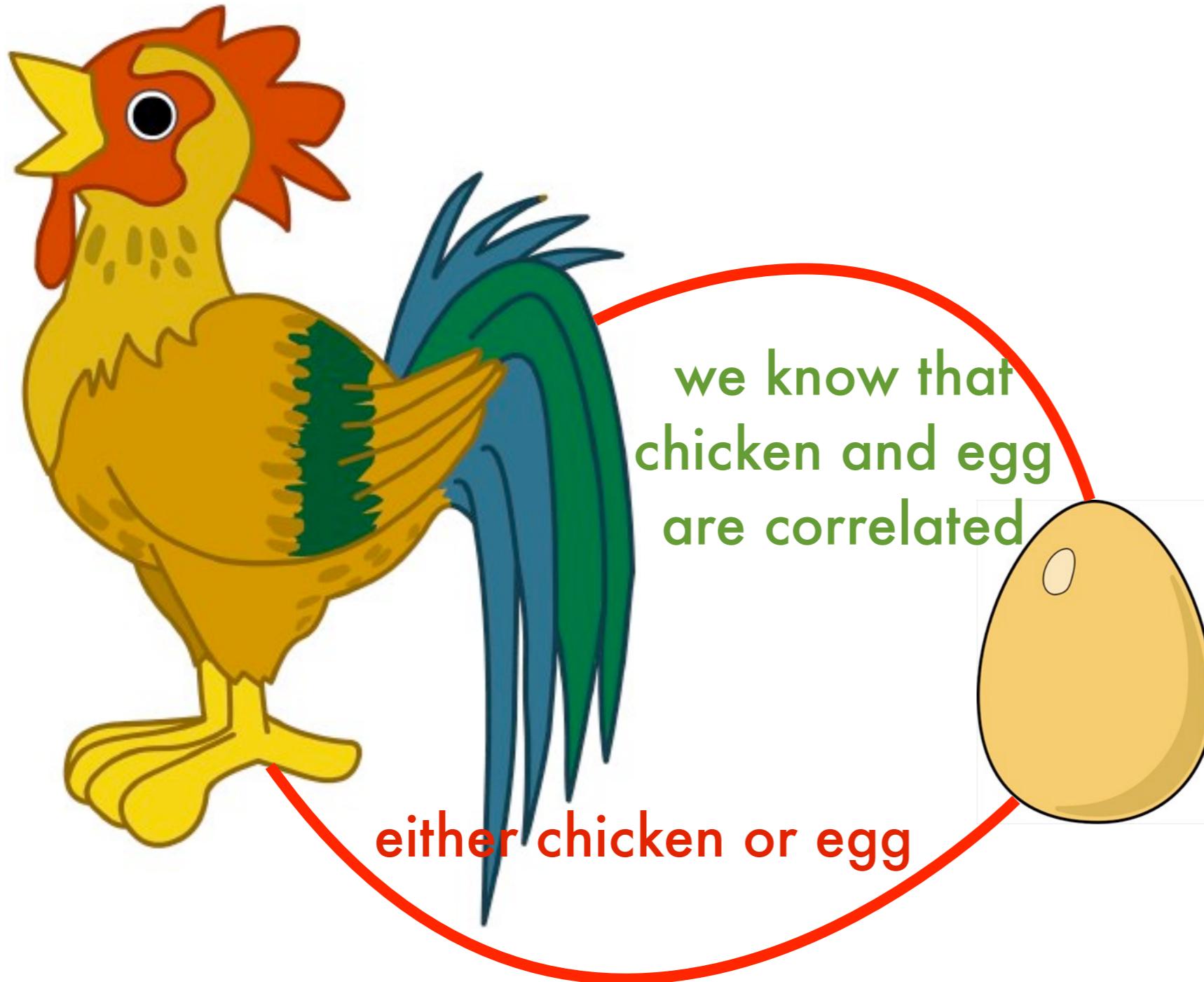
Chicken and Egg



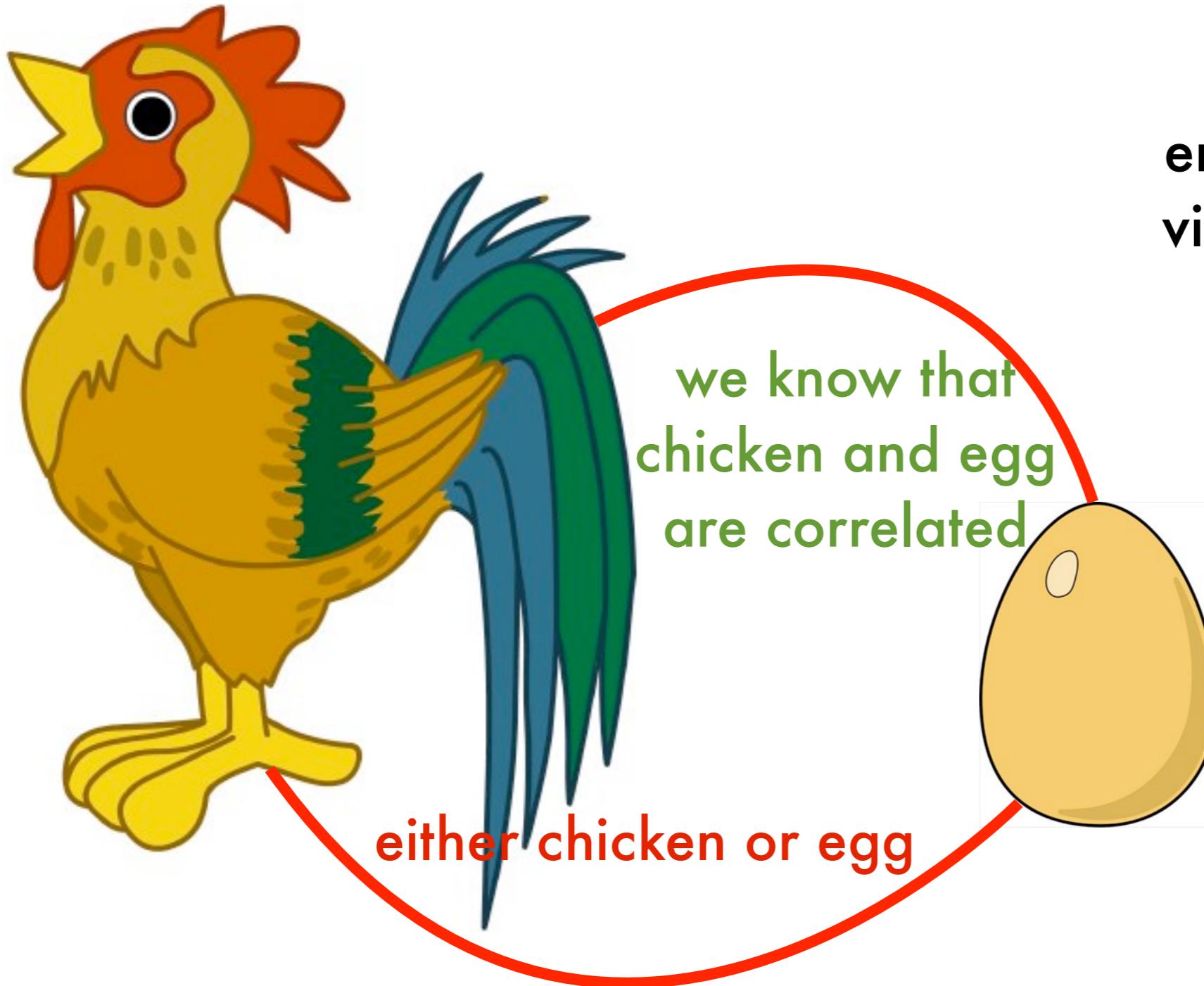
Chicken and Egg



Chicken and Egg



Chicken and Egg



encode the correlation
via the clique potential
between c and e

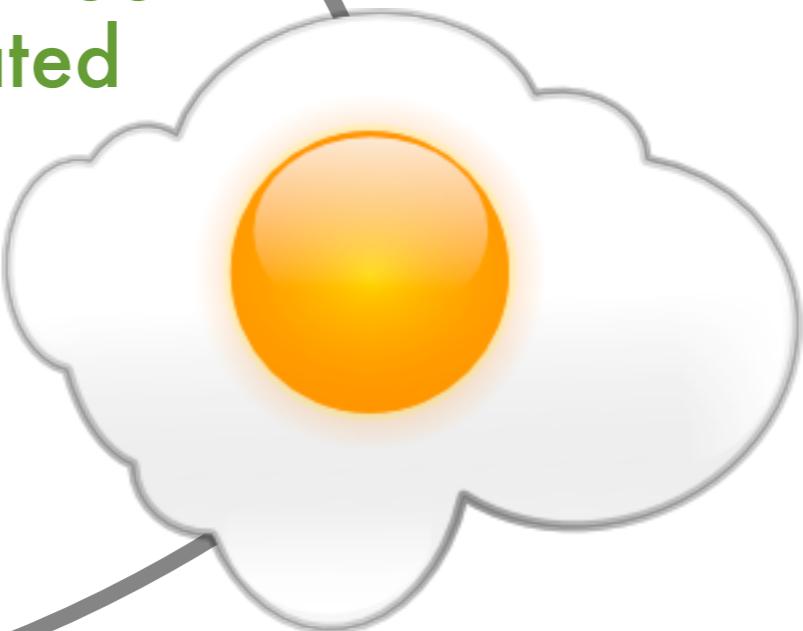
$$p(c, e) \propto \exp \psi(c, e)$$

Chicken and Egg



we know that
chicken and egg
are correlated

either chicken or egg



Chicken and Egg

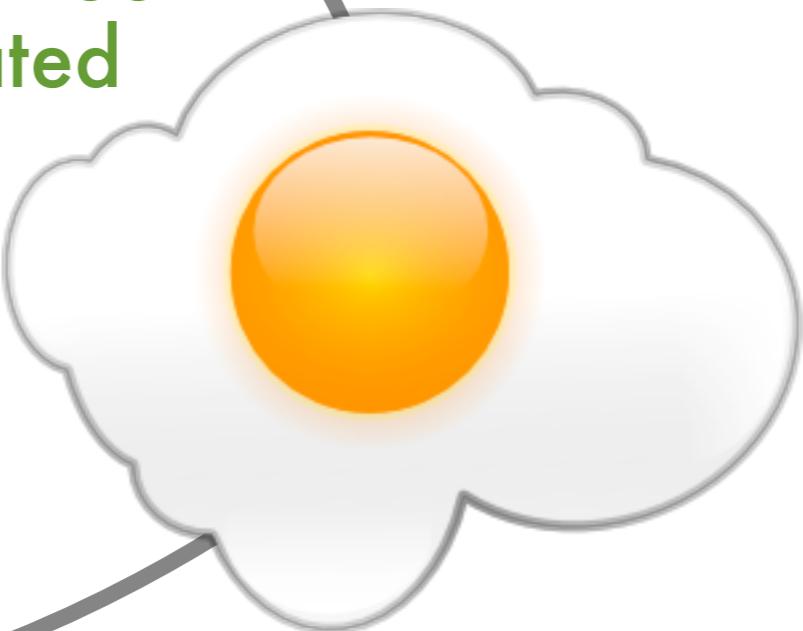
$$p(c, e) = \frac{\exp \psi(c, e)}{\sum_{c', e'} \exp \psi(c', e')}$$

$$= \exp [\psi(c, e) - g(\psi)] \text{ where } g(\psi) = \log \sum_{c, e} \exp \psi(c, e)$$

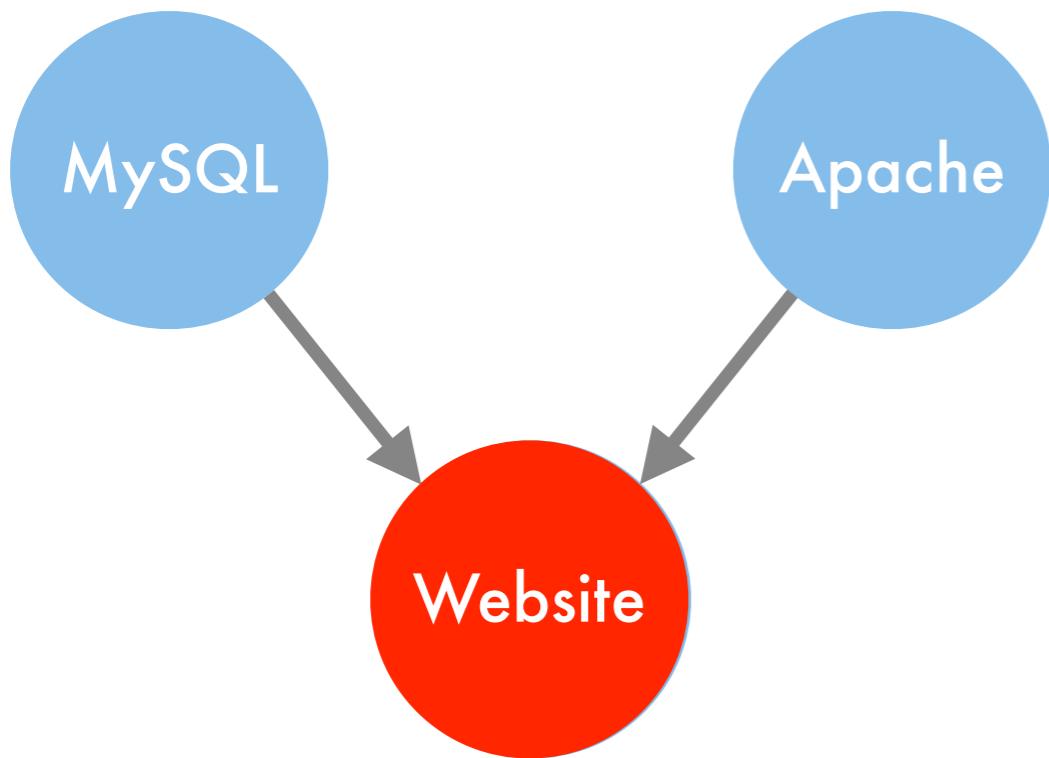


we know that
chicken and egg
are correlated

either chicken or egg



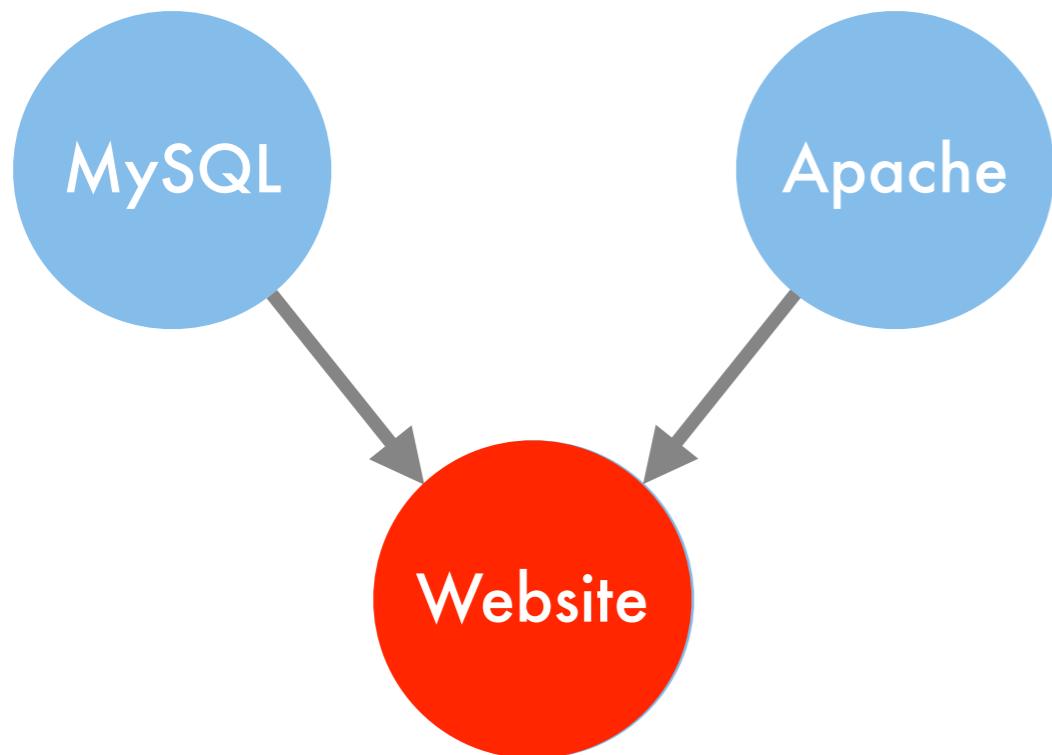
... some Yahoo service



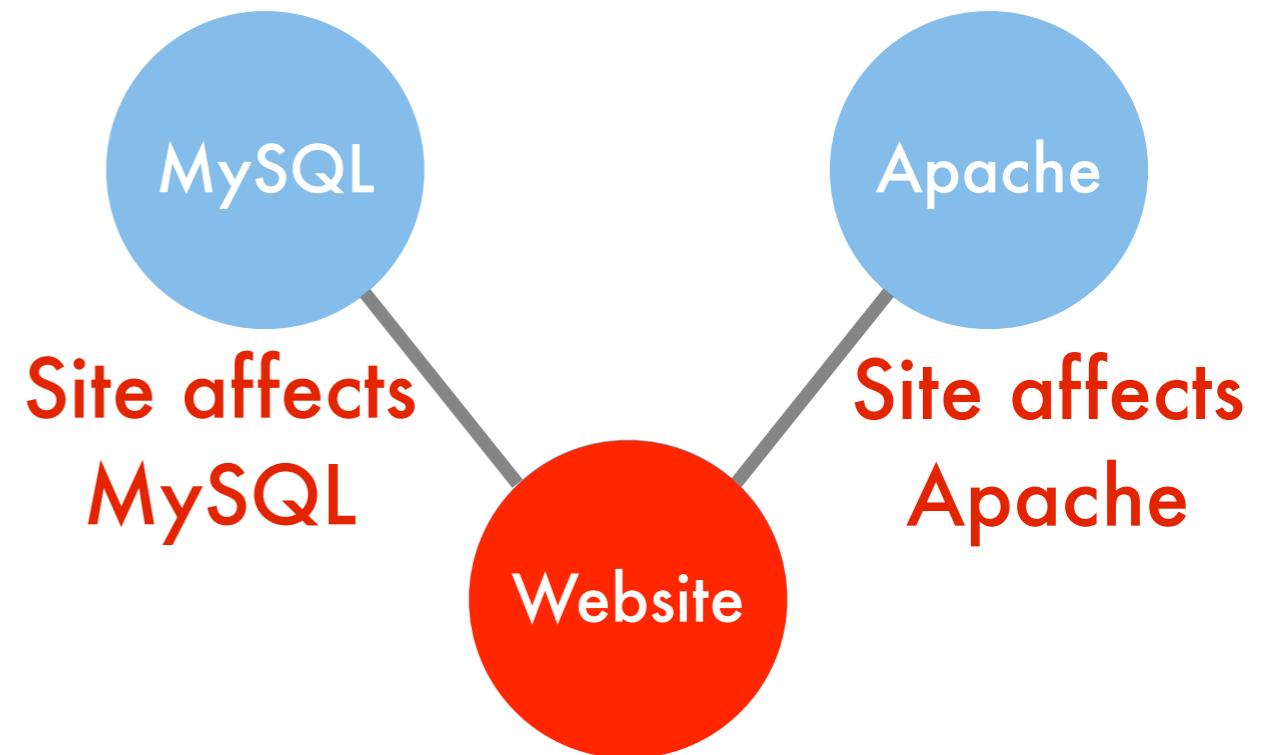
$$p(w|m, a)p(m)p(a)$$

$$m \not\perp a|w$$

... some Yahoo service

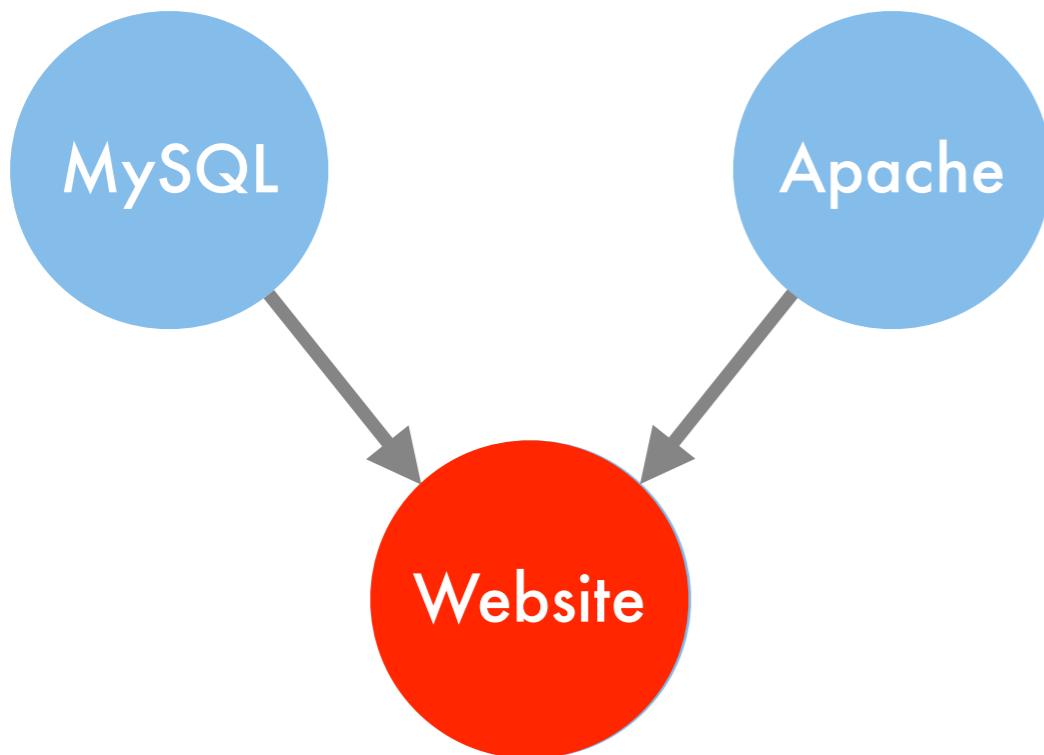


$$p(w|m, a)p(m)p(a)$$
$$m \not\perp\!\!\! \perp a|w$$



$$p(m, w, a) \propto \phi(m, w)\phi(w, a)$$
$$m \perp\!\!\! \perp a|w$$

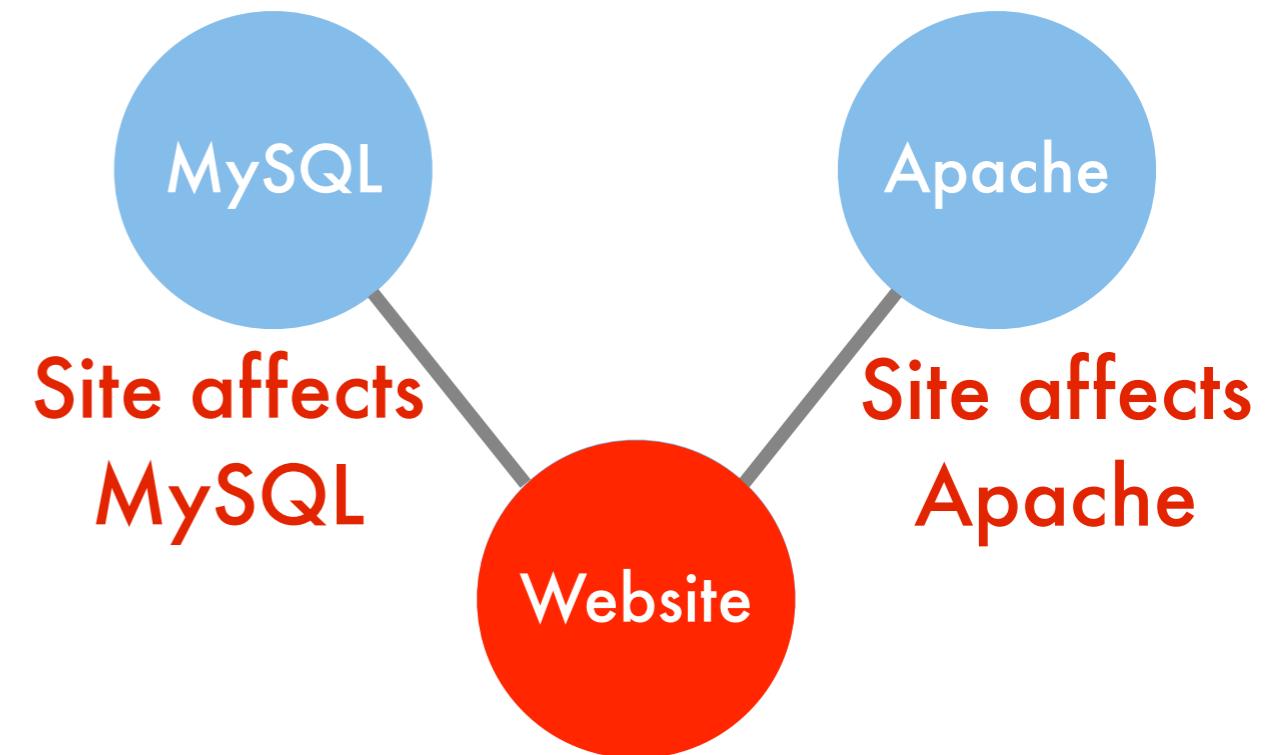
... some Yahoo service



$$p(w|m, a)p(m)p(a)$$

$$m \perp\!\!\! \perp a|w$$

easier
“debugging”

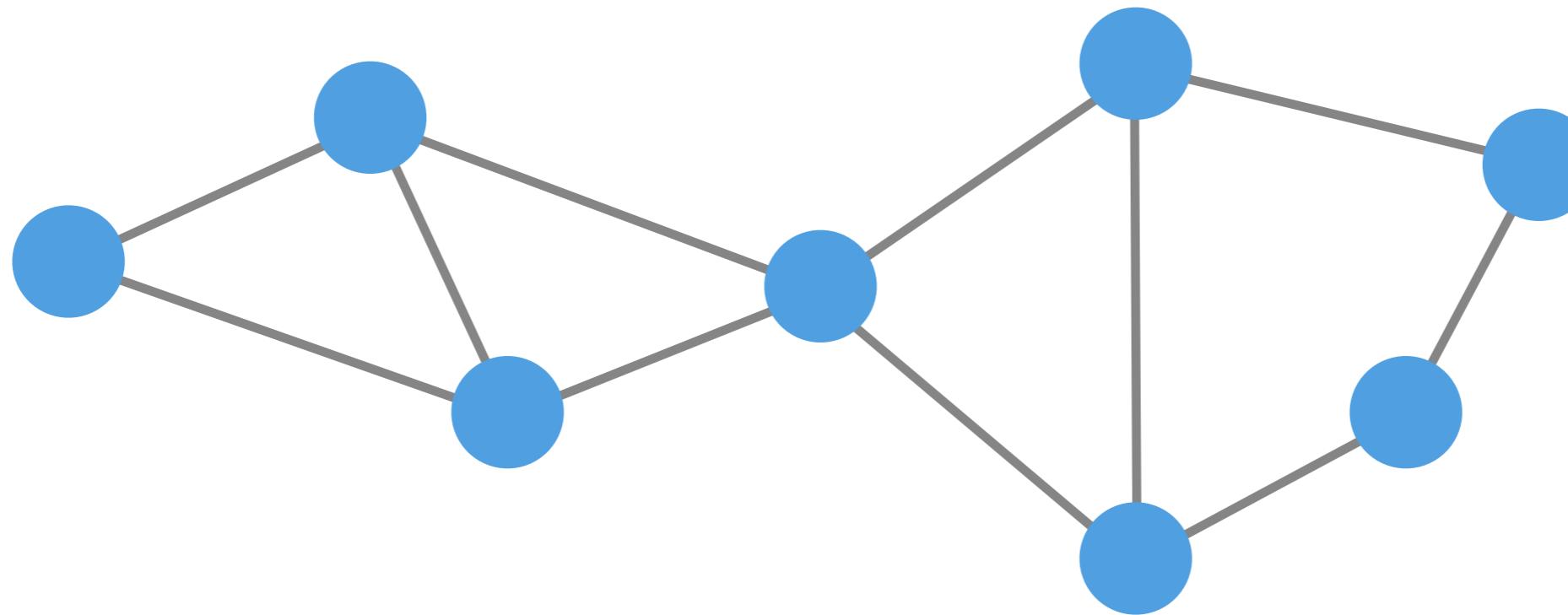


$$p(m, w, a) \propto \phi(m, w)\phi(w, a)$$

$$m \perp\!\!\! \perp a|w$$

easier
“modeling”

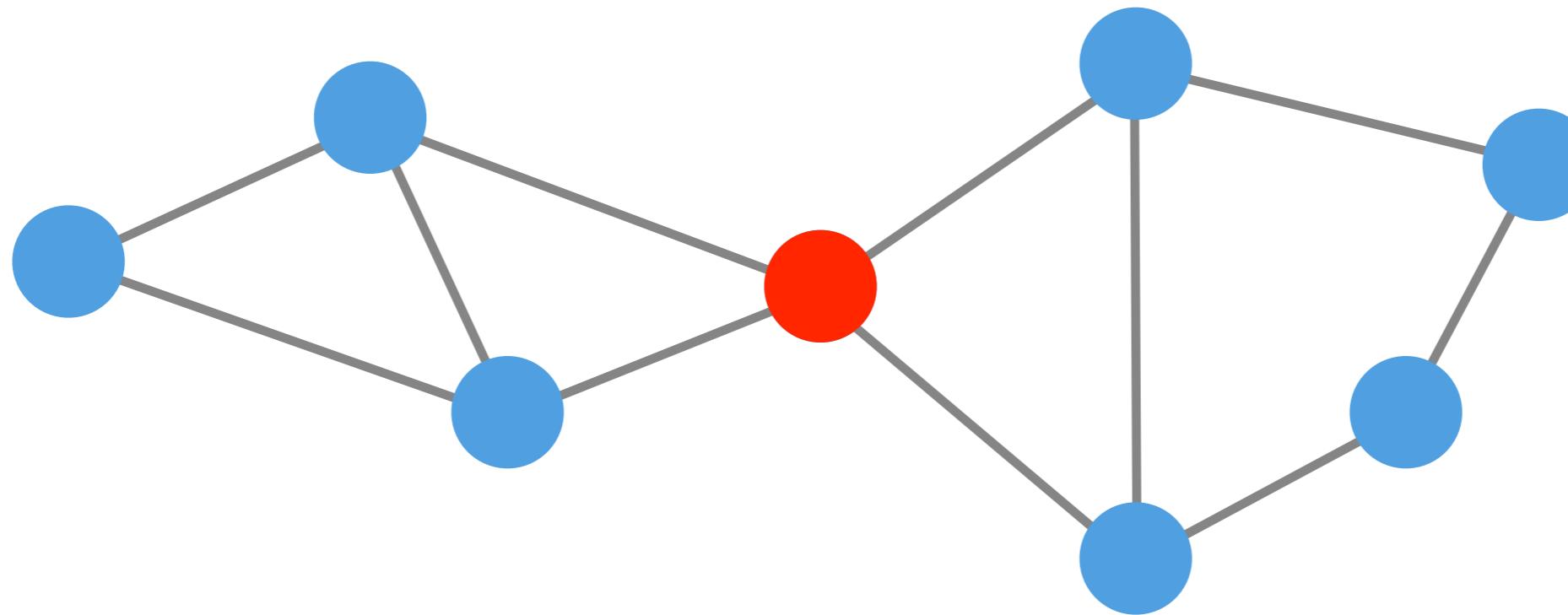
Undirected Graphical Models



Key Concept

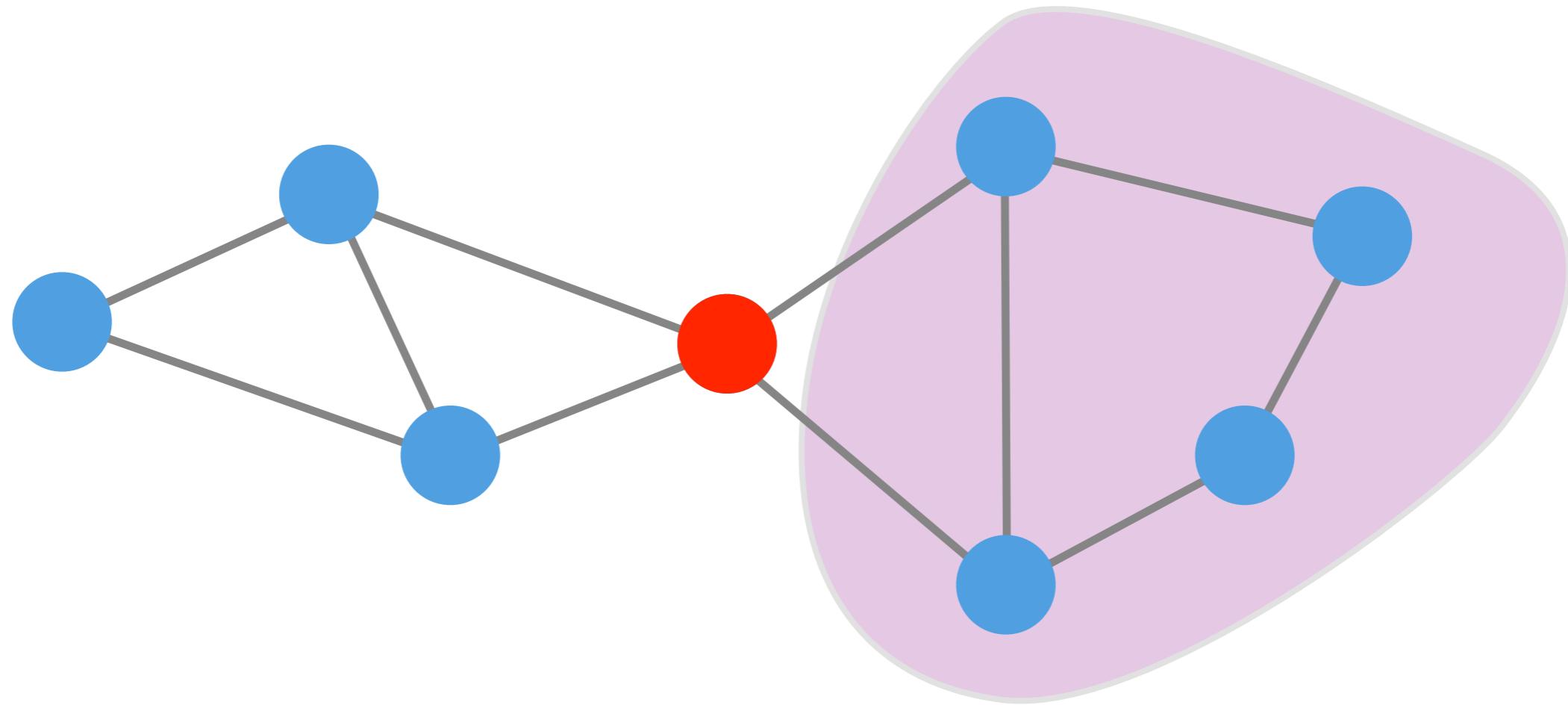
Observing nodes makes remainder
conditionally independent

Undirected Graphical Models



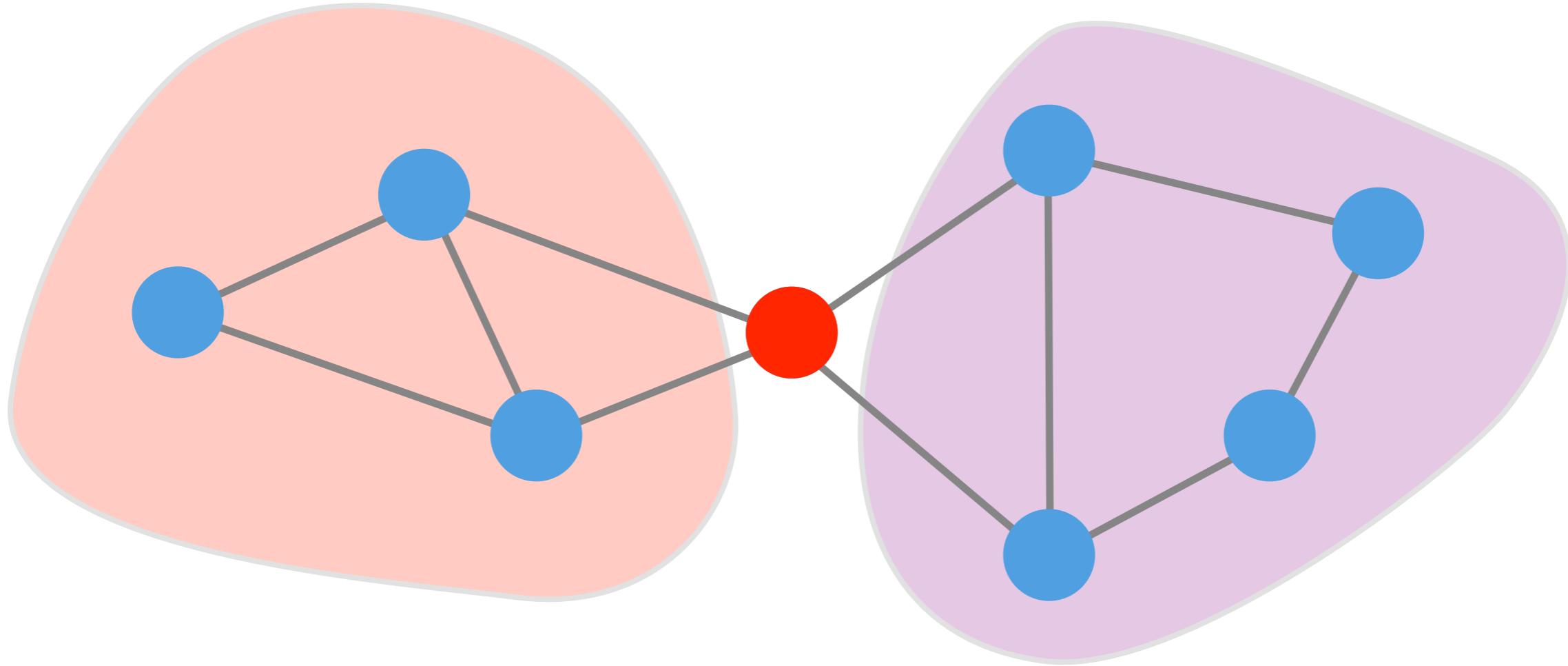
Key Concept
Observing nodes makes remainder
conditionally independent

Undirected Graphical Models



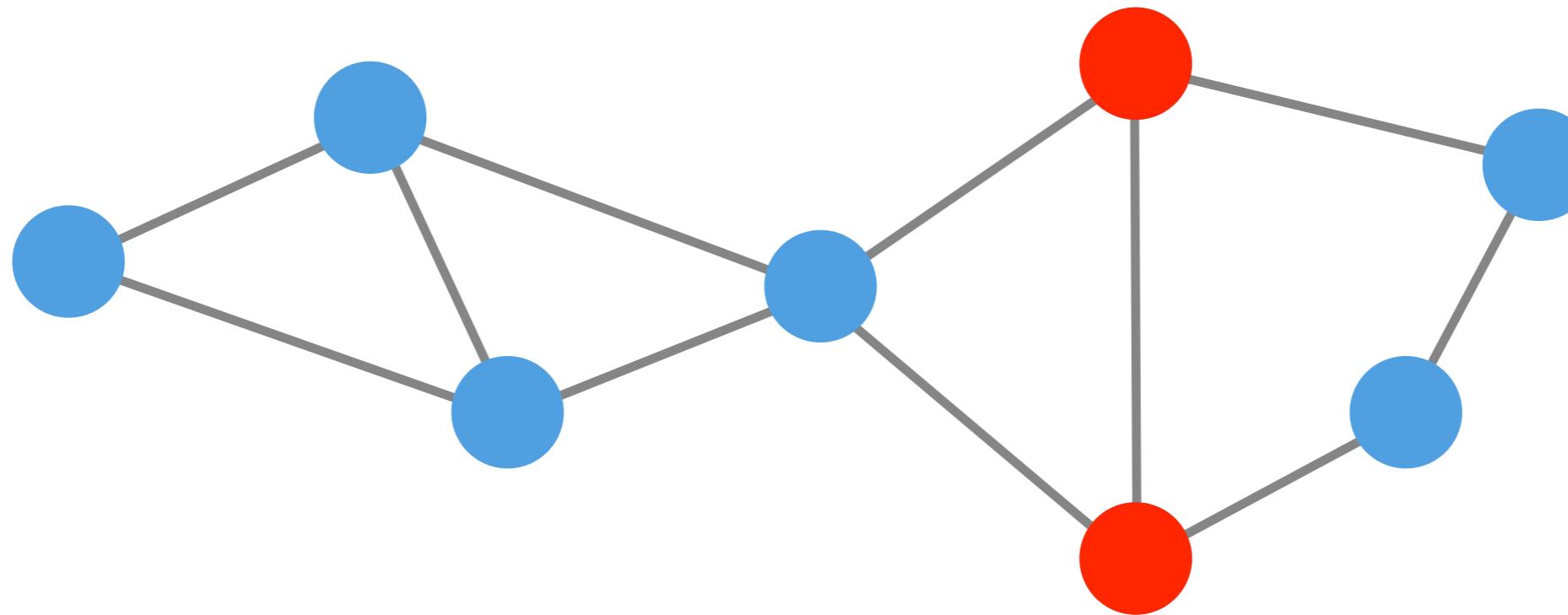
Key Concept
Observing nodes makes remainder
conditionally independent

Undirected Graphical Models



Key Concept
Observing nodes makes remainder
conditionally independent

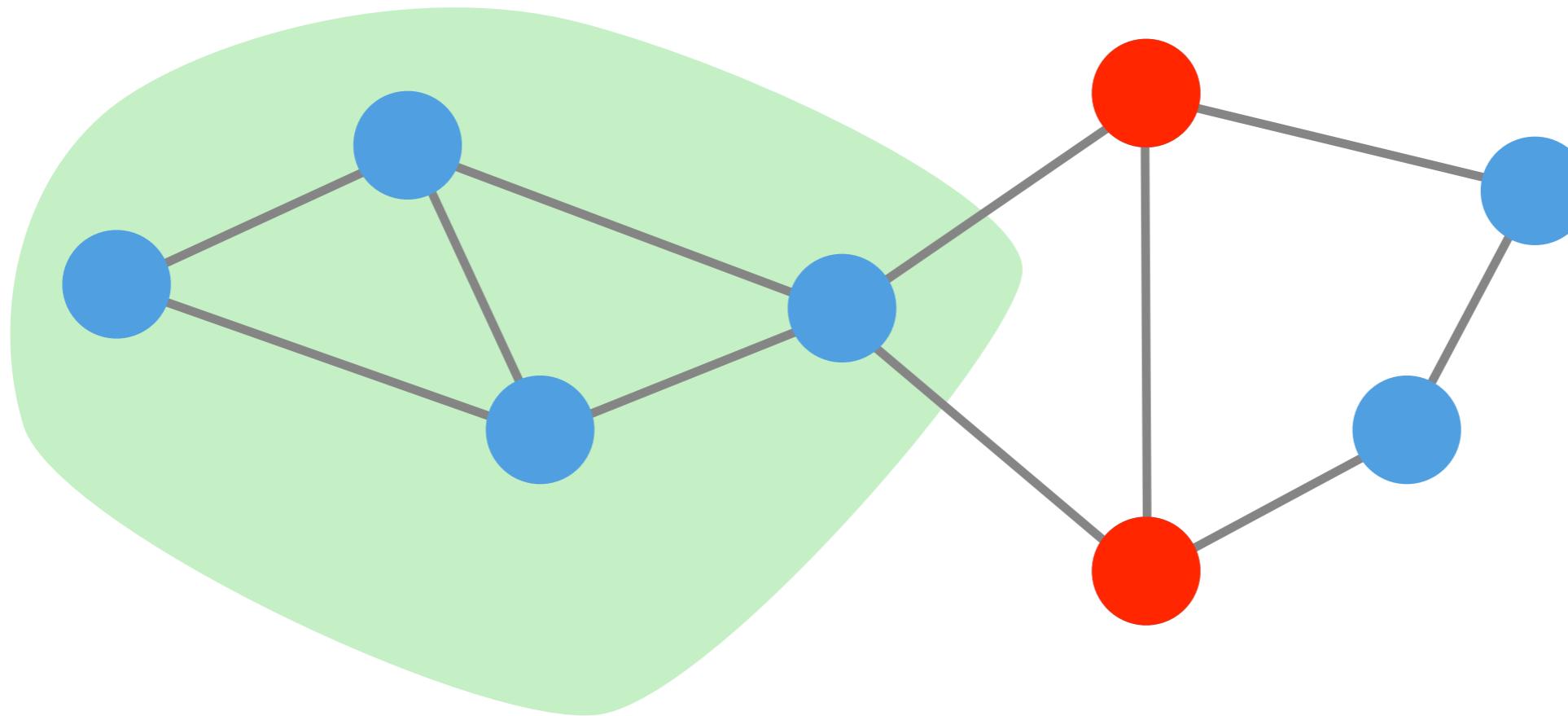
Undirected Graphical Models



Key Concept

Observing nodes makes remainder
conditionally independent

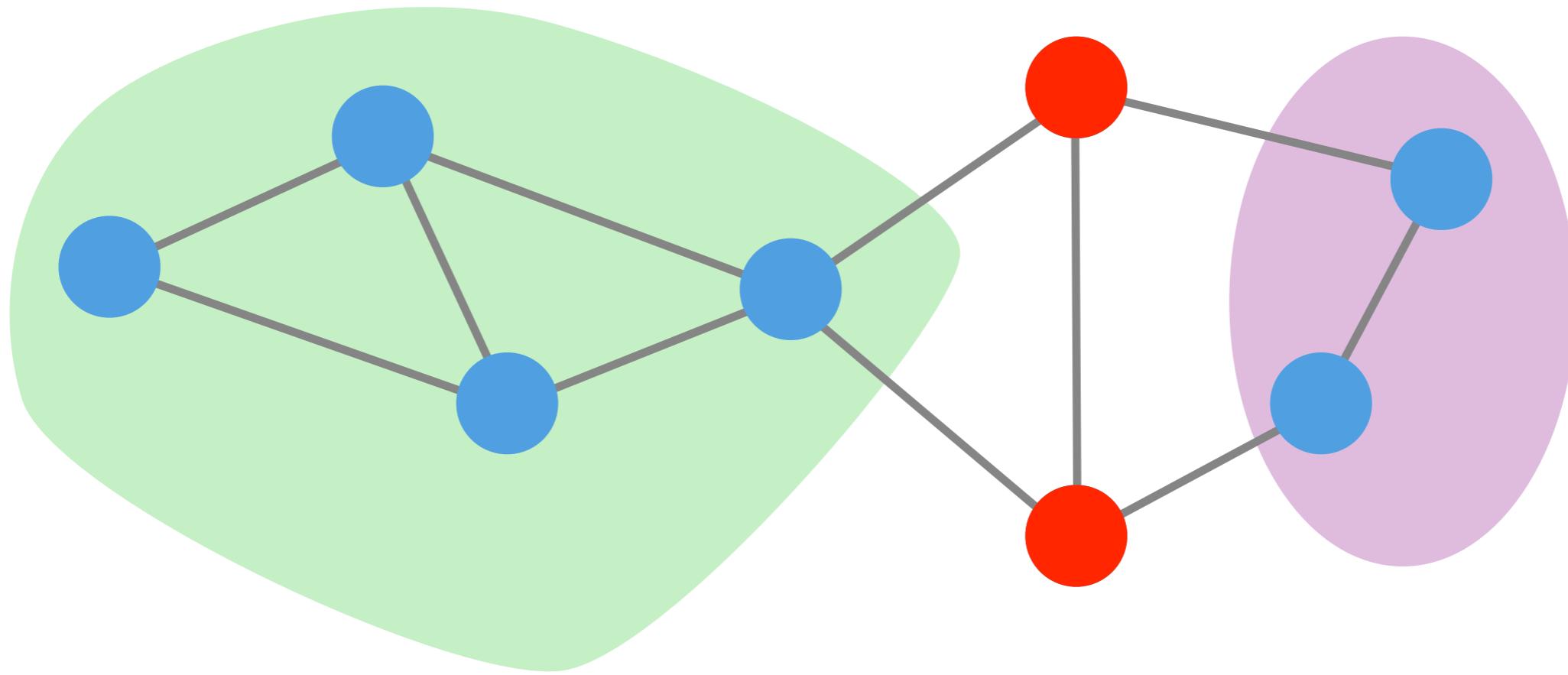
Undirected Graphical Models



Key Concept

Observing nodes makes remainder
conditionally independent

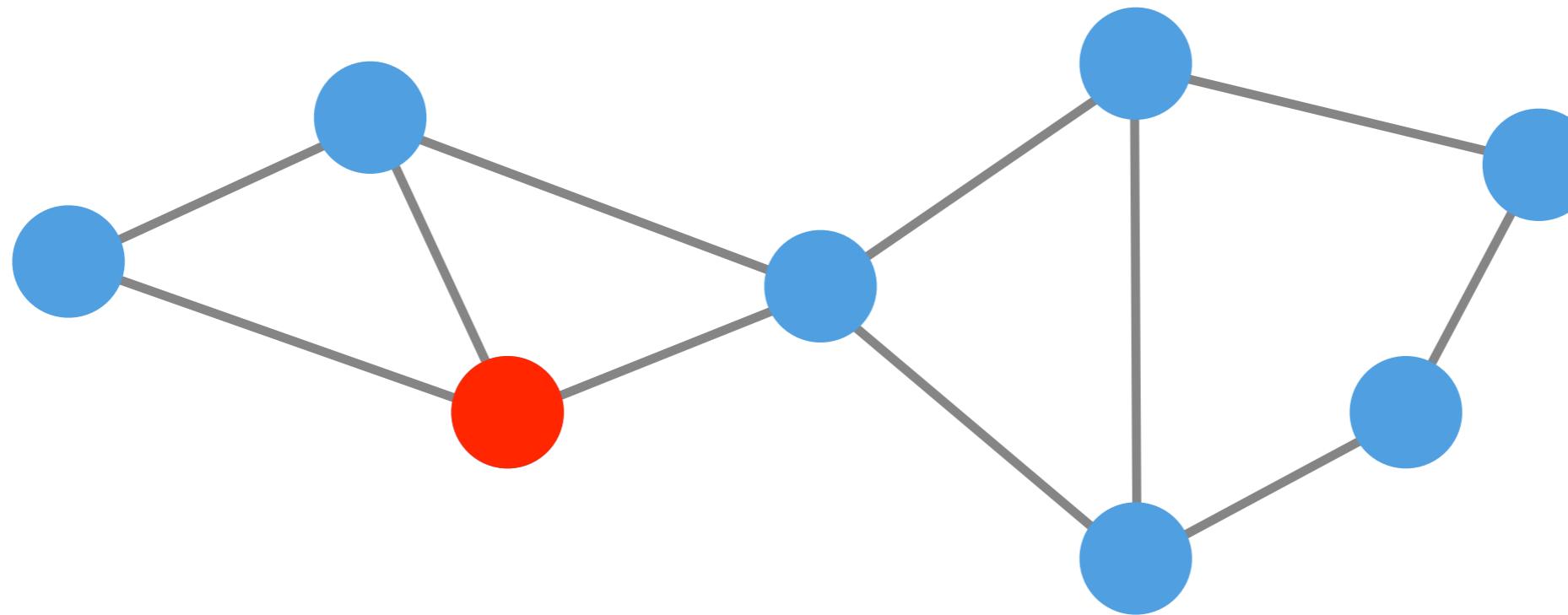
Undirected Graphical Models



Key Concept

Observing nodes makes remainder
conditionally independent

Undirected Graphical Models



Key Concept

Observing nodes makes remainder
conditionally independent

Cliques

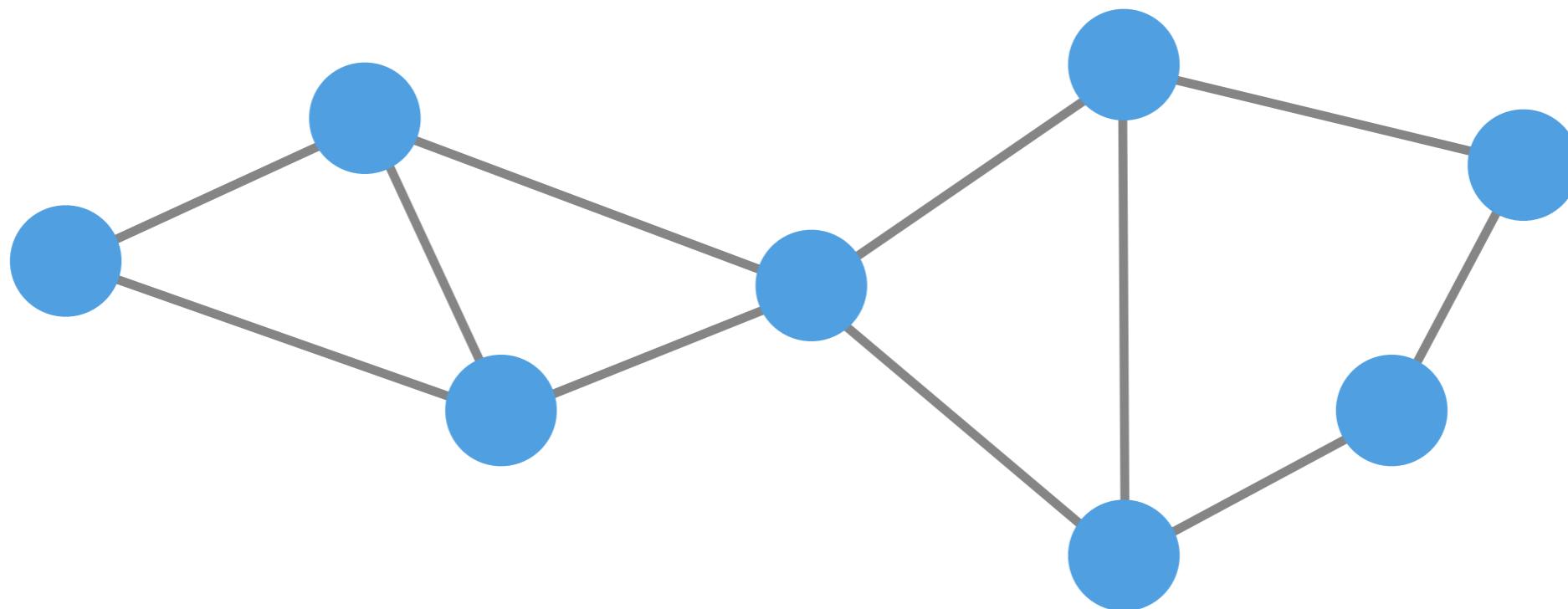


Cliques



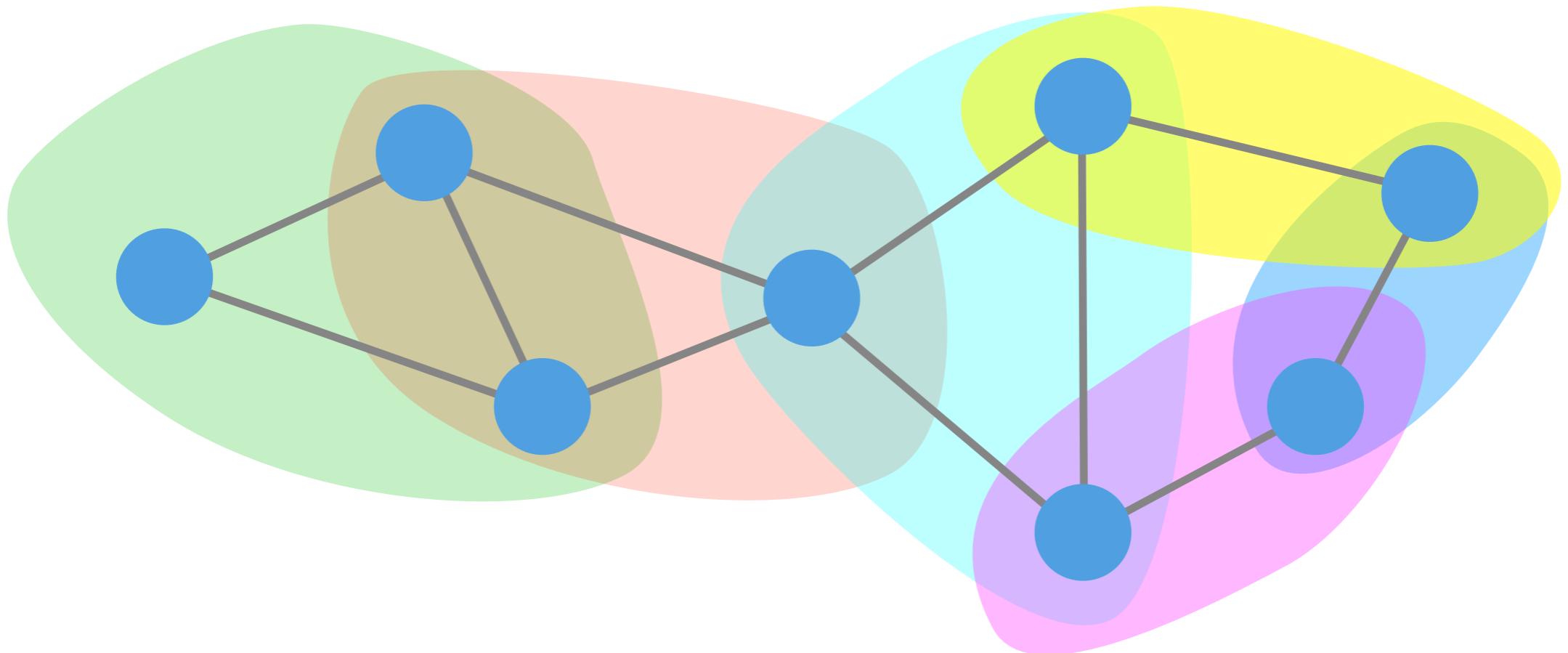
maximal fully connected subgraph

Cliques



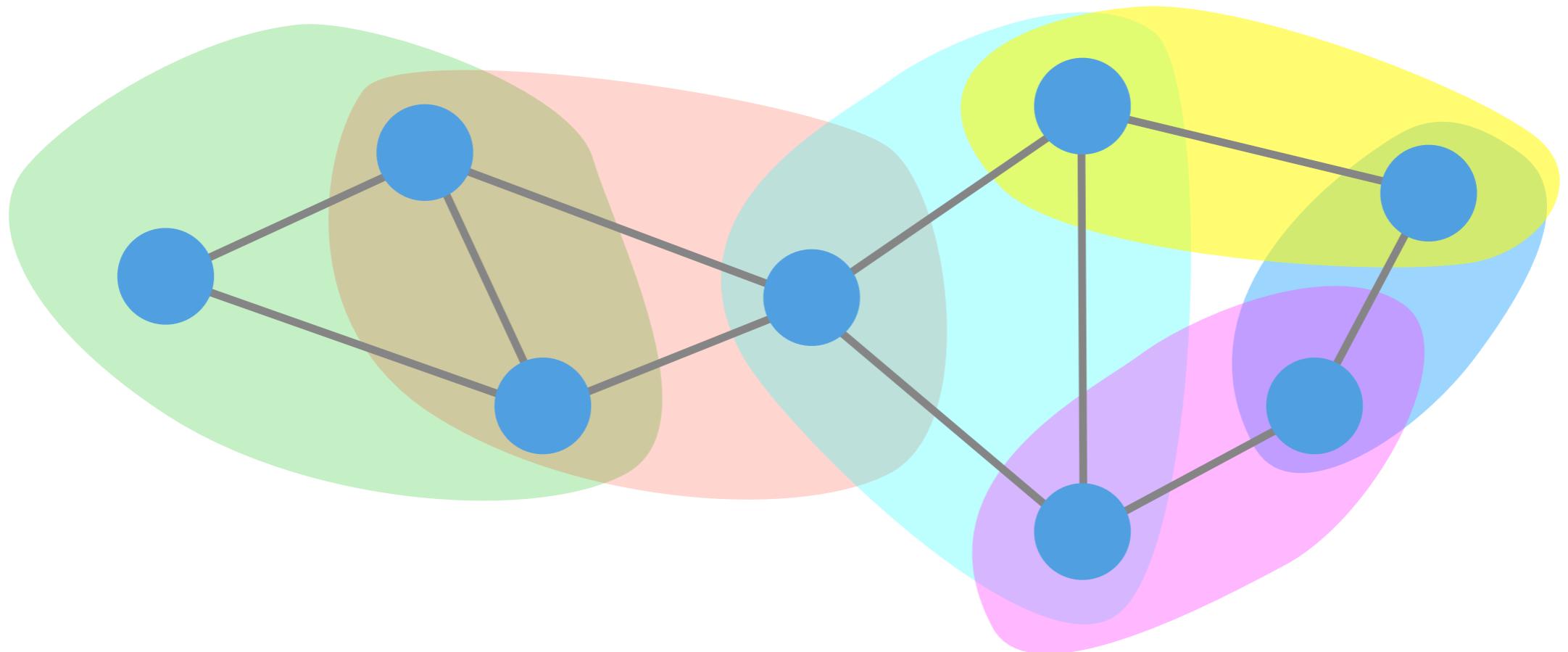
maximal fully connected subgraph

Cliques



maximal fully connected subgraph

Hammersley Clifford Theorem



If density has full support then it decomposes into products of clique potentials

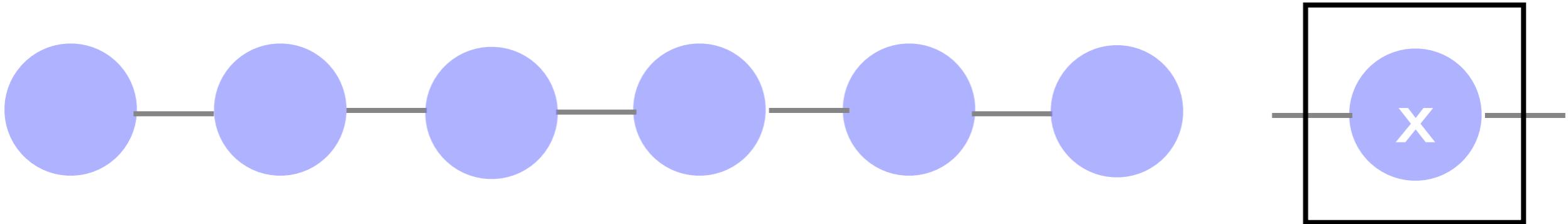
$$p(x) = \prod_c \psi_c(x_c)$$

Directed vs. Undirected

- Causal description
- Normalization automatic
- Intuitive
- Requires knowledge of dependencies
- Conditional independence tricky (Bayes Ball algorithm)
- Noncausal description (correlation only)
- Intuitive
- Easy modeling
- Normalization difficult
- Conditional independence easy to read off (graph connectivity)

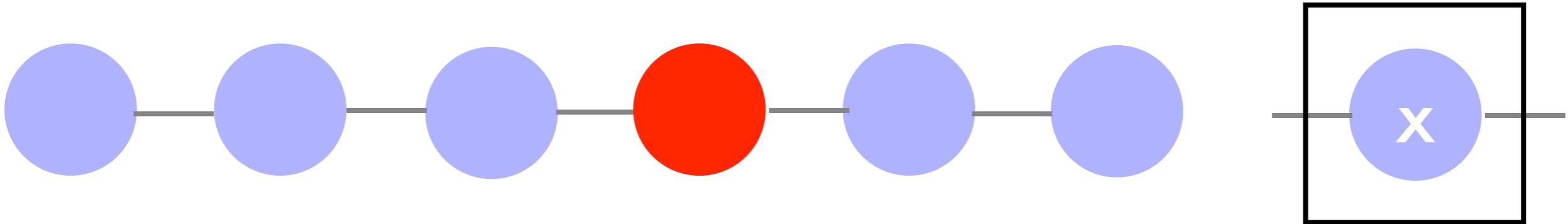
Examples

Chains



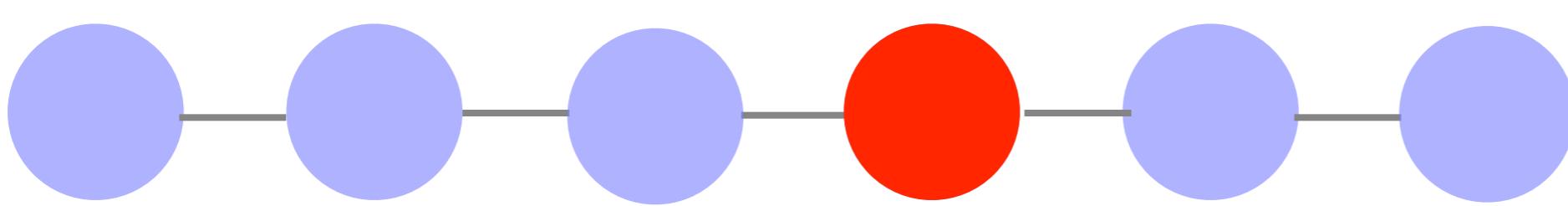
$$p(x) = \prod_i \psi_i(x_i, x_{i+1})$$

Chains

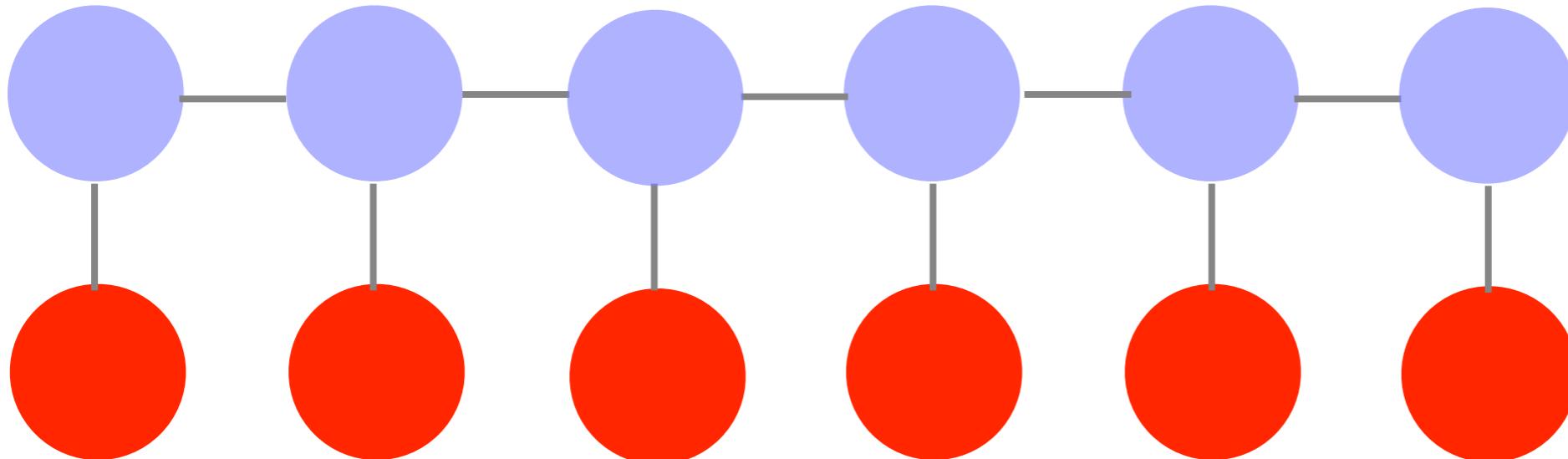


$$p(x) = \prod_i \psi_i(x_i, x_{i+1})$$

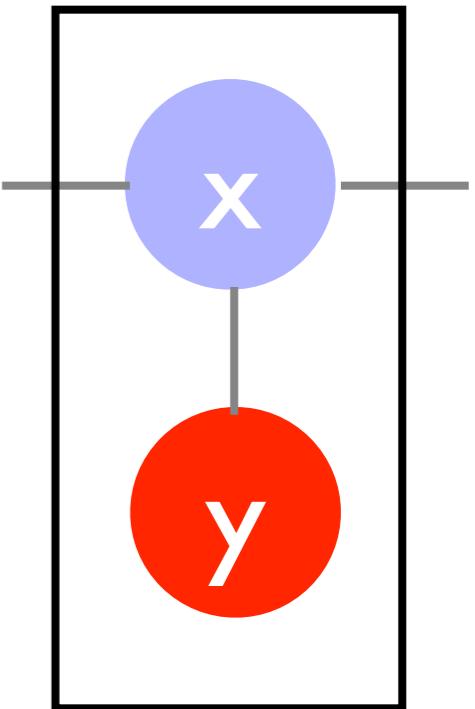
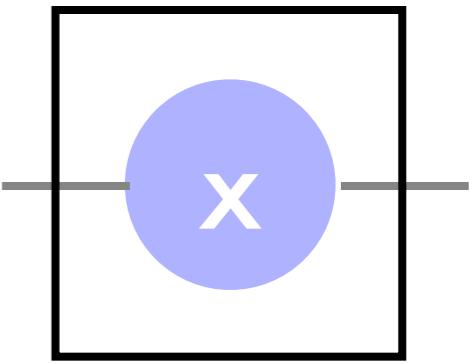
Chains



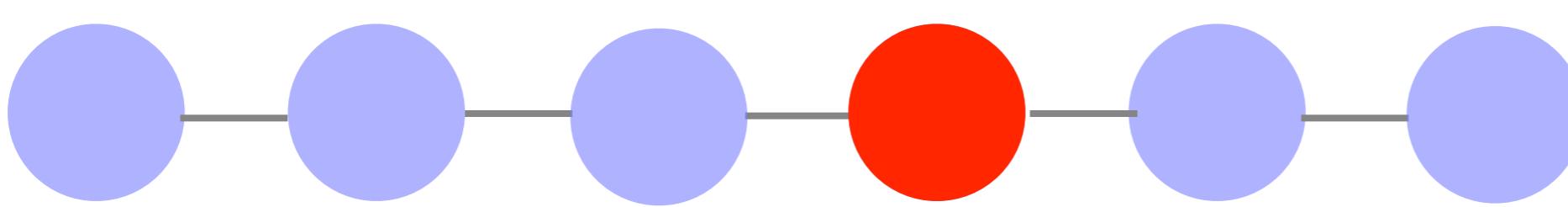
$$p(x) = \prod_i \psi_i(x_i, x_{i+1})$$



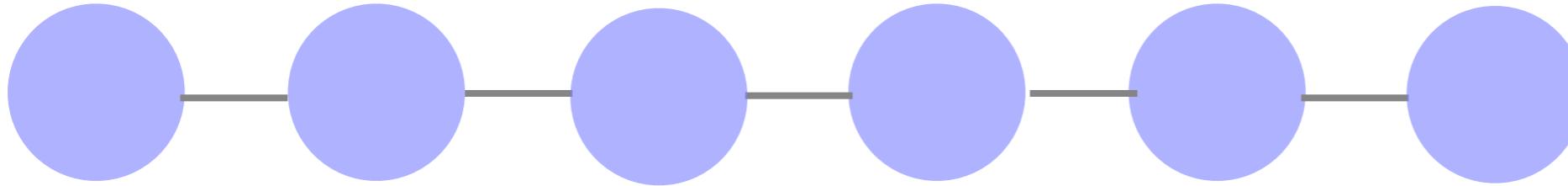
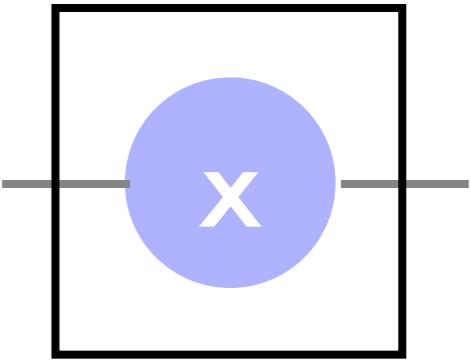
$$p(x, y) = \prod_i \psi_i^x(x_i, x_{i+1}) \psi_i^{xy}(x_i, y_i)$$



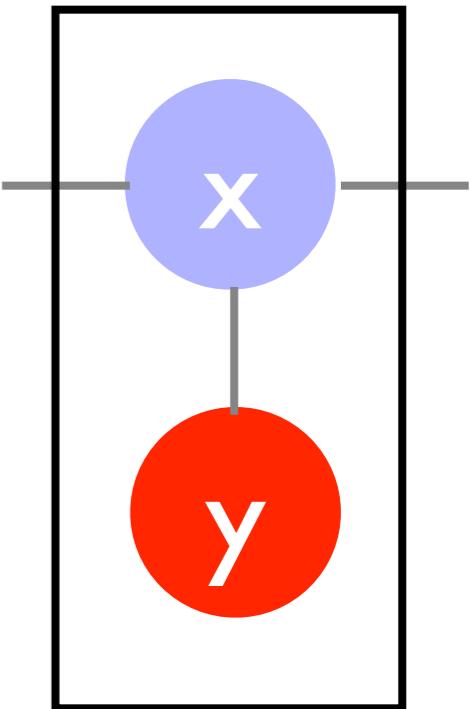
Chains



$$p(x) = \prod_i \psi_i(x_i, x_{i+1})$$

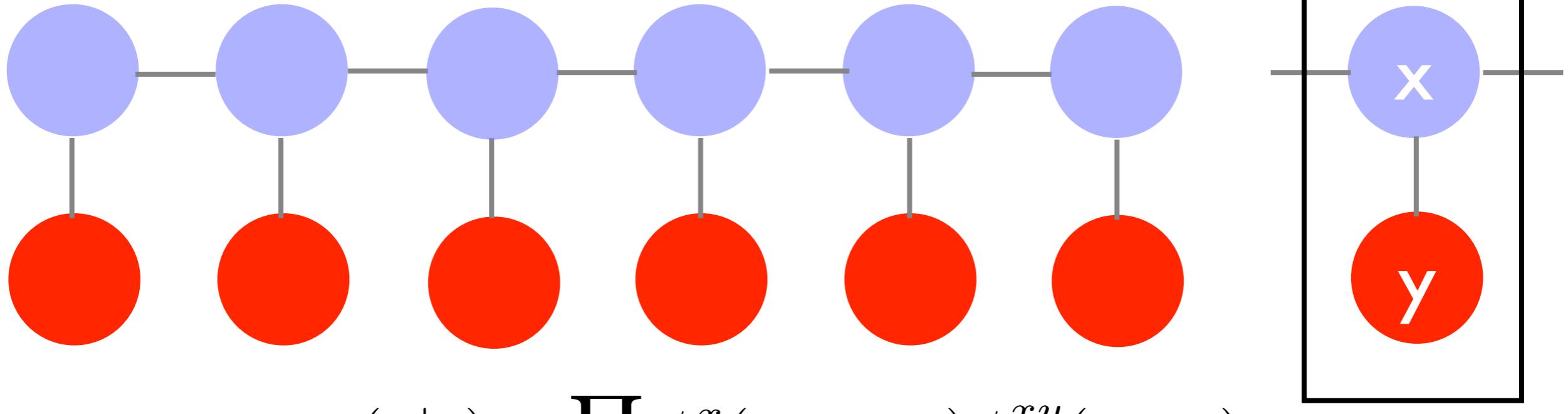


$$p(x|y) \propto \prod_i \underbrace{\psi_i^x(x_i, x_{i+1})}_{=:f_i(x_i, x_{i+1})} \underbrace{\psi_i^{xy}(x_i, y_i)}_{}$$



$$p(x, y) = \prod_i \psi_i^x(x_i, x_{i+1}) \psi_i^{xy}(x_i, y_i)$$

Chains



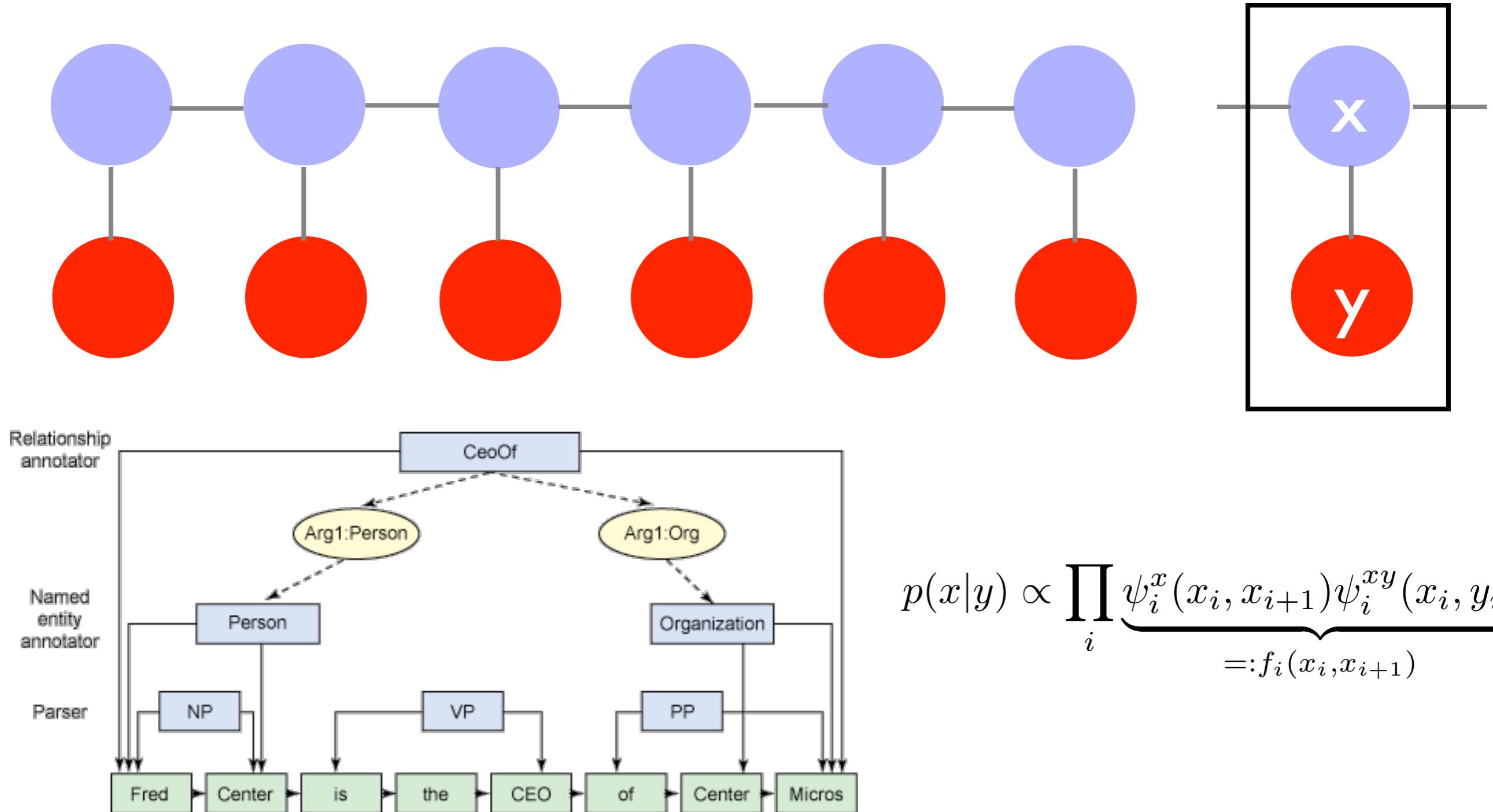
$$p(x|y) \propto \prod_i \underbrace{\psi_i^x(x_i, x_{i+1}) \psi_i^{xy}(x_i, y_i)}_{=: f_i(x_i, x_{i+1})}$$

Dynamic Programming

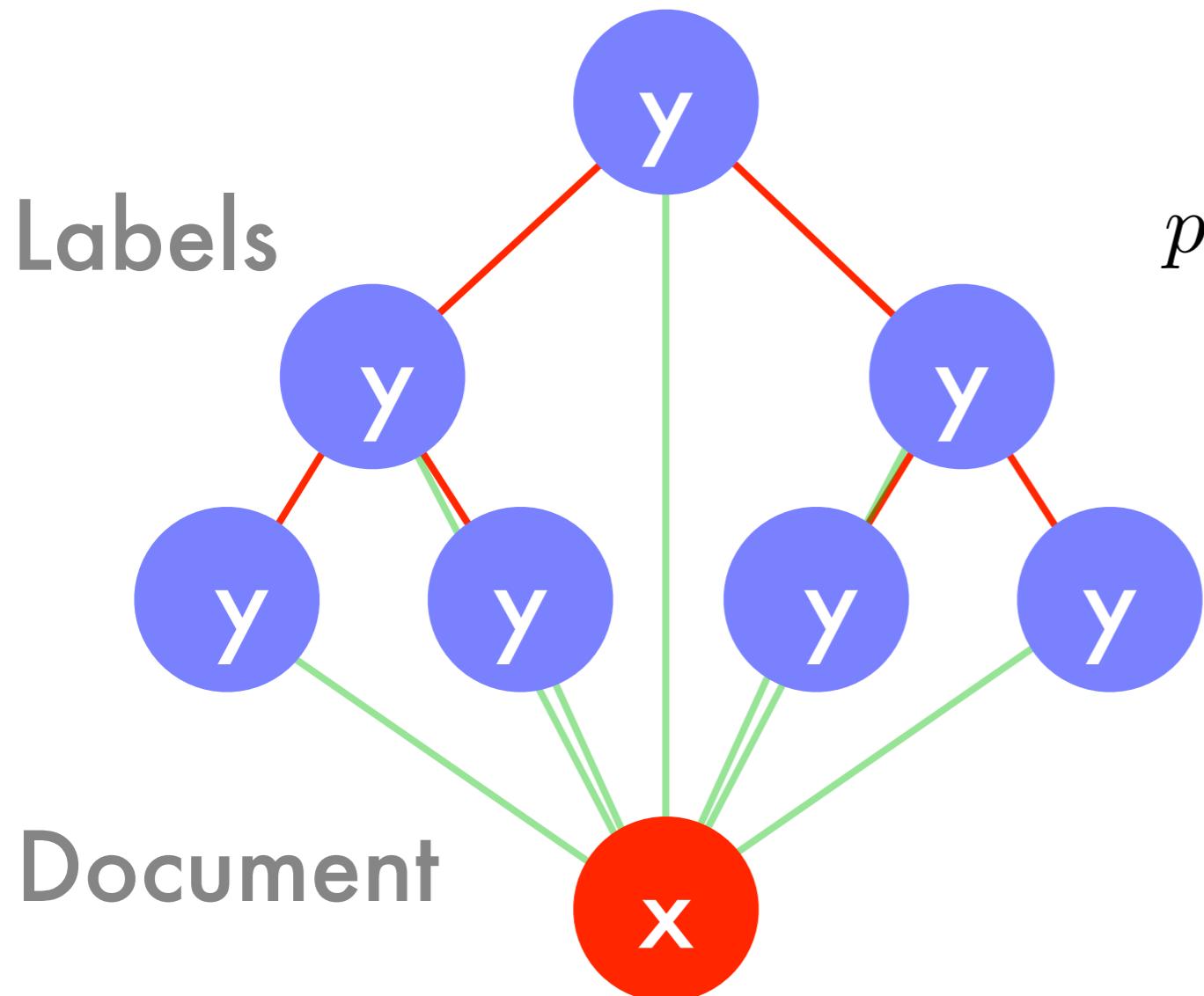
$$l_1(x_1) = 1 \text{ and } l_{i+1}(x_{i+1}) = \sum_{x_i} l_i(x_i) f_i(x_i, x_{i+1})$$

$$r_n(x_n) = 1 \text{ and } r_i(x_i) = \sum_{x_{i+1}} r_{i+1}(x_{i+1}) f_i(x_i, x_{i+1})$$

Named Entity Tagging



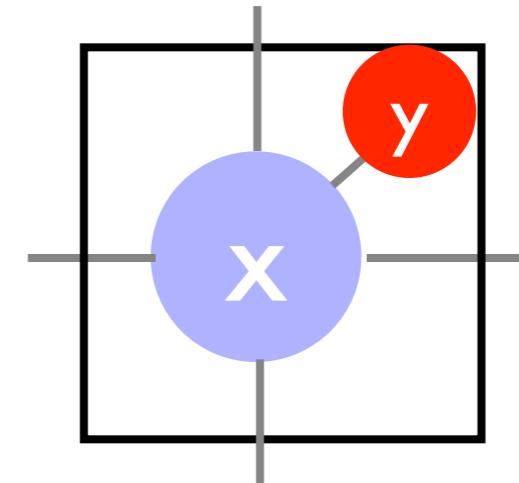
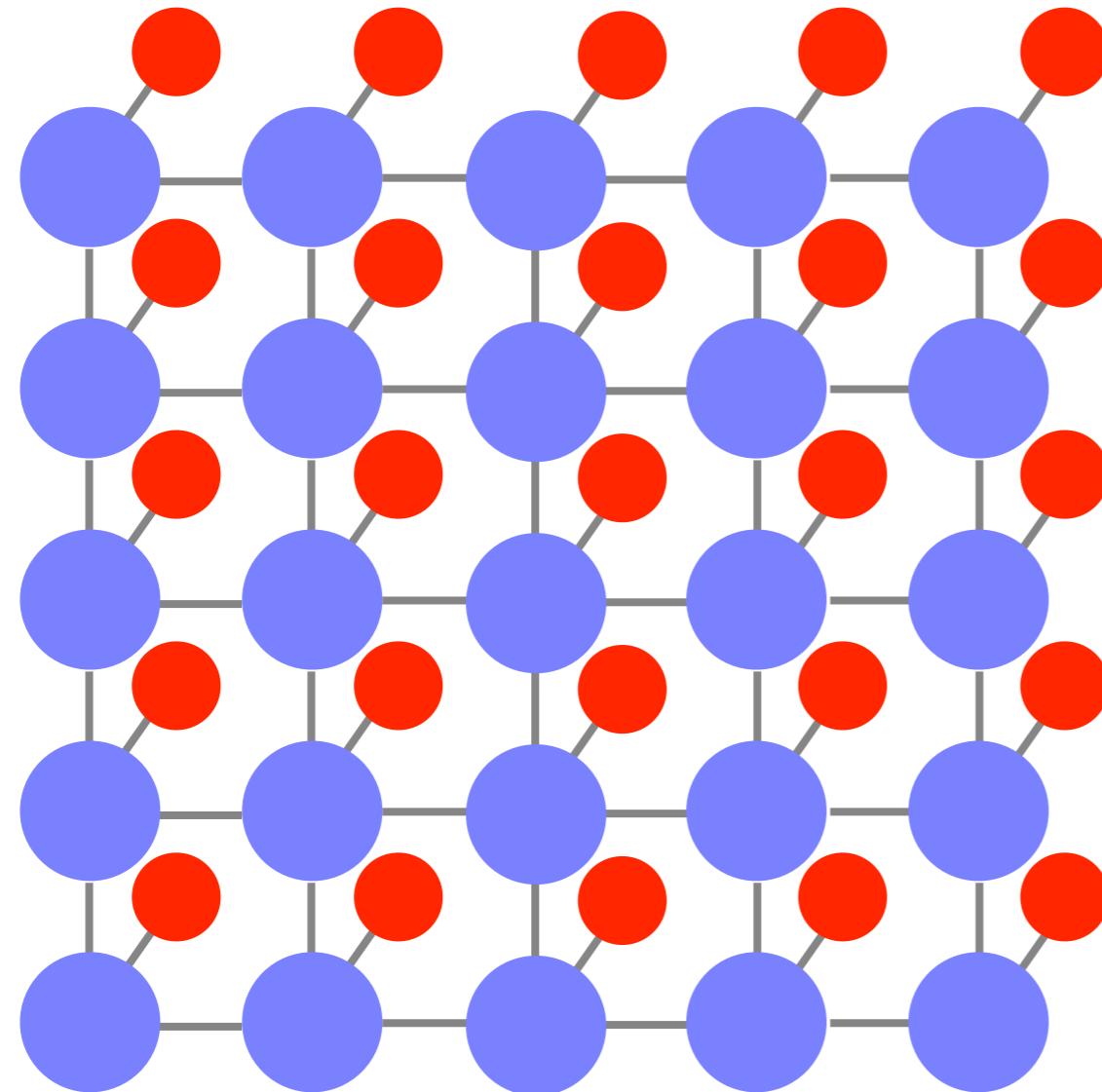
Trees + Ontologies



$$p(y|x) = \prod_i \psi(y_i, y_{\text{parent}(i)}, x)$$

- Ontology classification (e.g. YDir, DMOZ)

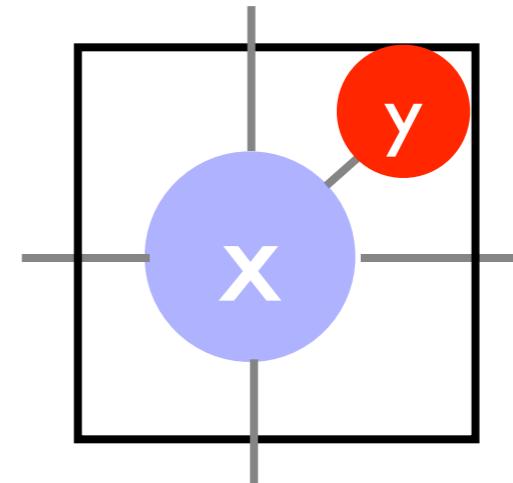
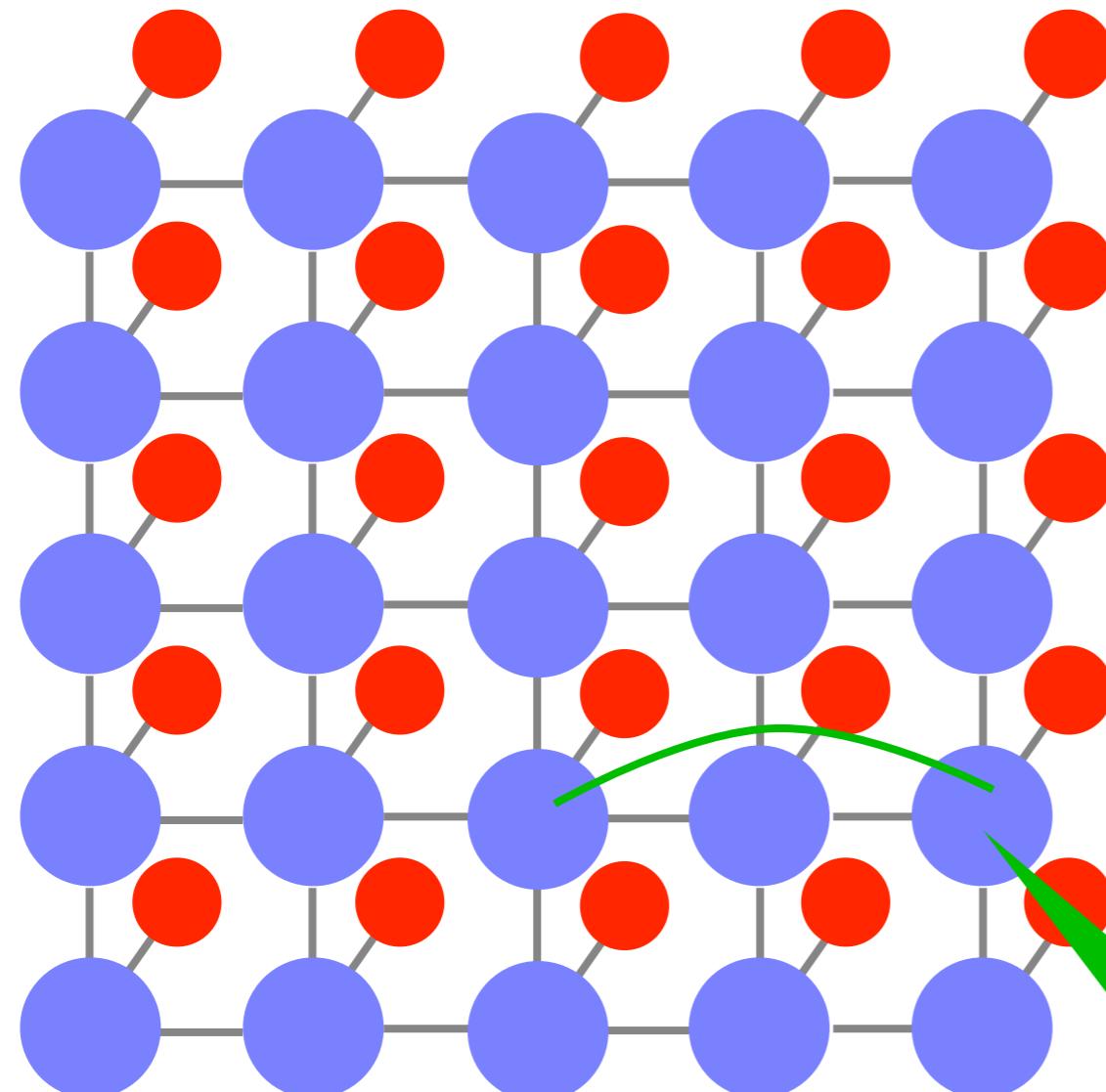
Spin Glasses + Images



observed pixels
real image

$$p(x|y) = \prod_{ij} \psi^{\text{right}}(x_{ij}, x_{i+1,j}) \psi^{\text{up}}(x_{ij}, x_{i,j+1}) \psi^{xy}(x_{ij}, y_{ij})$$

Spin Glasses + Images

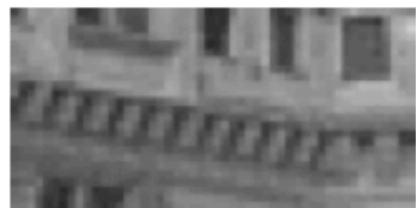
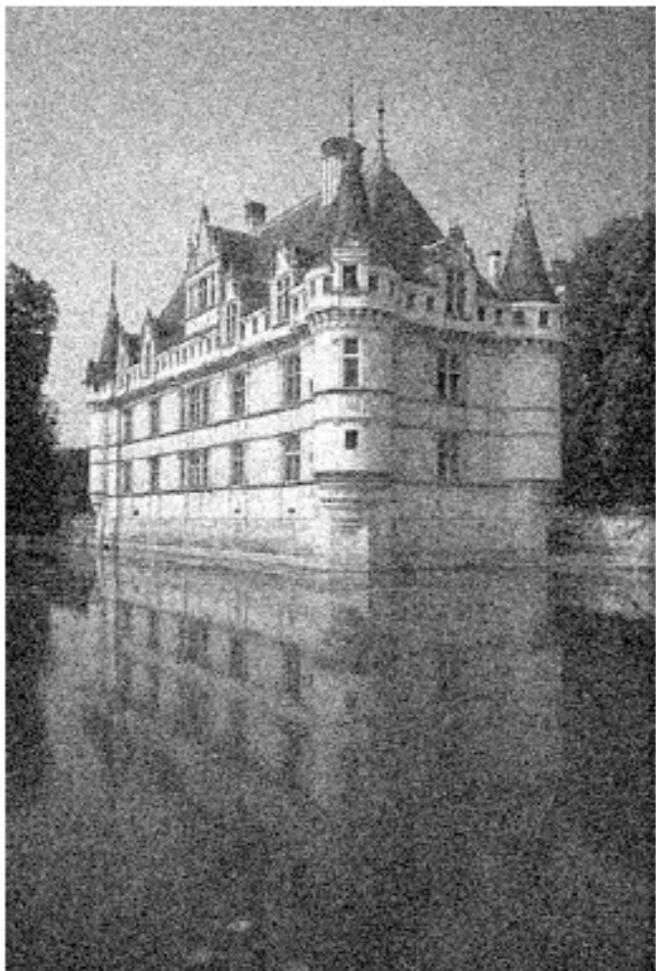
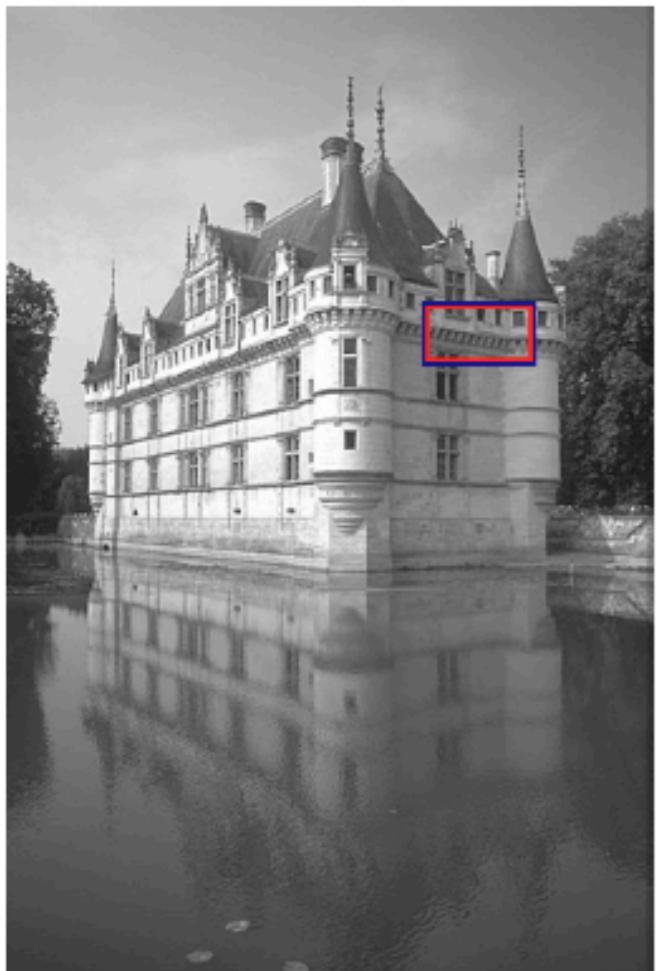


observed pixels
real image

long range interactions

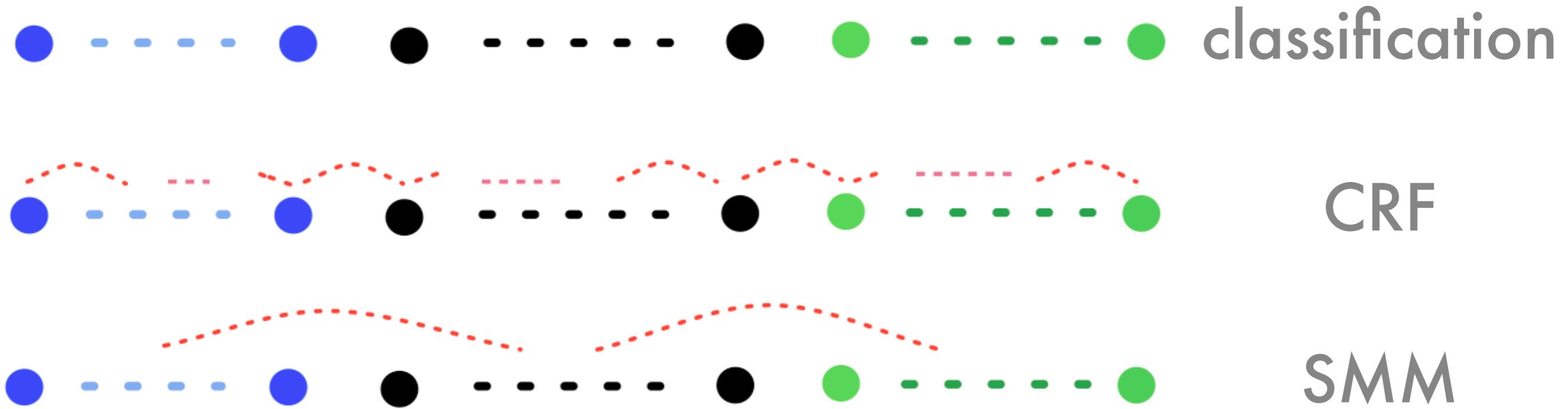
$$p(x|y) = \prod_{ij} \psi^{\text{right}}(x_{ij}, x_{i+1,j}) \psi^{\text{up}}(x_{ij}, x_{i,j+1}) \psi^{xy}(x_{ij}, y_{ij})$$

Image Denoising



Li&Huttenlocher, ECCV'08

Semi-Markov Models

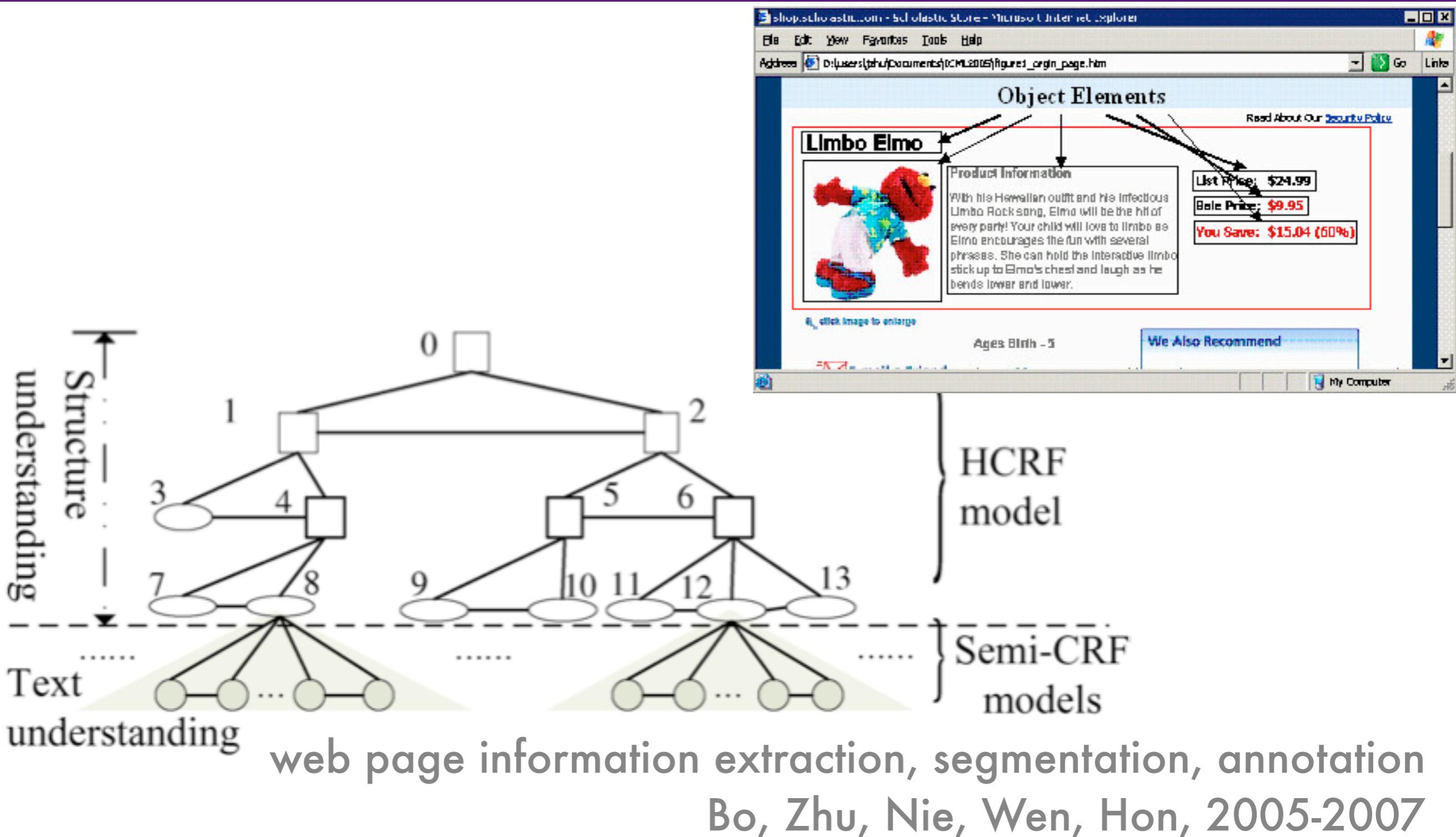


- **Flexible length of an episode**
- **Segmentation between episodes**

phrase segmentation, activity recognition, motion data analysis

Shi, Smola, Altun, Vishwanathan, Li, 2007-2009

2D CRF for Webpages



Exponential Families and Graphical Models

Exponential Family Reunion

- **Density function**

$$p(x; \theta) = \exp (\langle \phi(x), \theta \rangle - g(\theta))$$

$$\text{where } g(\theta) = \log \sum_{x'} \exp (\langle \phi(x'), \theta \rangle)$$

- **Log partition function generates cumulants**

$$\partial_\theta g(\theta) = \mathbf{E} [\phi(x)]$$

$$\partial_\theta^2 g(\theta) = \text{Var} [\phi(x)]$$

- **g is convex (second derivative is p.s.d.)**

Log Partition Function

$$p(x|\theta) = e^{\langle \phi(x), \theta \rangle - g(\theta)}$$

Unconditional model

$$g(\theta) = \log \sum_x e^{\langle \phi(x), \theta \rangle}$$

$$\partial_\theta g(\theta) = \frac{\sum_x \phi(x) e^{\langle \phi(x), \theta \rangle}}{\sum_x e^{\langle \phi(x), \theta \rangle}} = \sum_x \phi(x) e^{\langle \phi(x), \theta \rangle - g(\theta)}$$

$$p(y|\theta, x) = e^{\langle \phi(x,y), \theta \rangle - g(\theta|x)}$$

Conditional model

$$g(\theta|x) = \log \sum_y e^{\langle \phi(x,y), \theta \rangle}$$

$$\partial_\theta g(\theta|x) = \frac{\sum_y \phi(x, y) e^{\langle \phi(x,y), \theta \rangle}}{\sum_y e^{\langle \phi(x,y), \theta \rangle}} = \sum_y \phi(x, y) e^{\langle \phi(x,y), \theta \rangle - g(\theta|x)}$$

Estimation

- Conditional log-likelihood

$$\log p(y|x; \theta) = \langle \phi(x, y), \theta \rangle - g(\theta|x)$$

- Log-posterior (Gaussian Prior)

$$\log p(\theta|X, Y) = \sum_i \log(y_i|x_i; \theta) + \log p(\theta) + \text{const.}$$

$$= \left\langle \sum_i \phi(x_i, y_i), \theta \right\rangle - \sum_i g(\theta|x_i) - \frac{1}{2\sigma^2} \|\theta\|^2 + \text{const.}$$

- First order optimality conditions

maxent
model

$$\sum_i \phi(x_i, y_i) = \sum_i \mathbf{E}_{y|x_i} [\phi(x_i, y)] + \frac{1}{\sigma^2} \theta$$

expensive

prior

Logistic Regression

- **Label space**

$$\phi(x, y) = y\phi(x) \text{ where } y \in \{\pm 1\}$$

- **Log-partition function**

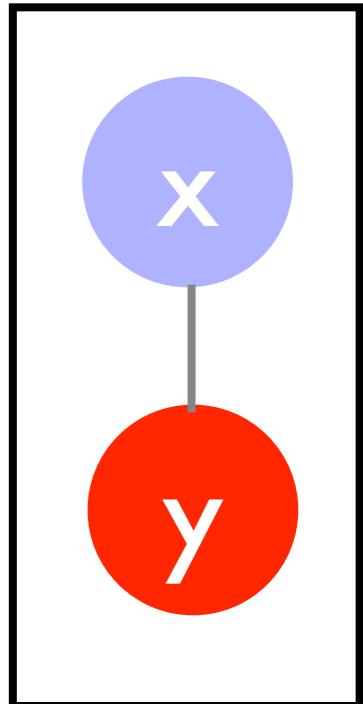
$$g(\theta|x) = \log \left[e^{1 \cdot \langle \phi(x), \theta \rangle} + e^{-1 \cdot \langle \phi(x), \theta \rangle} \right] = \log 2 \cosh \langle \phi(x), \theta \rangle$$

- **Convex minimization problem**

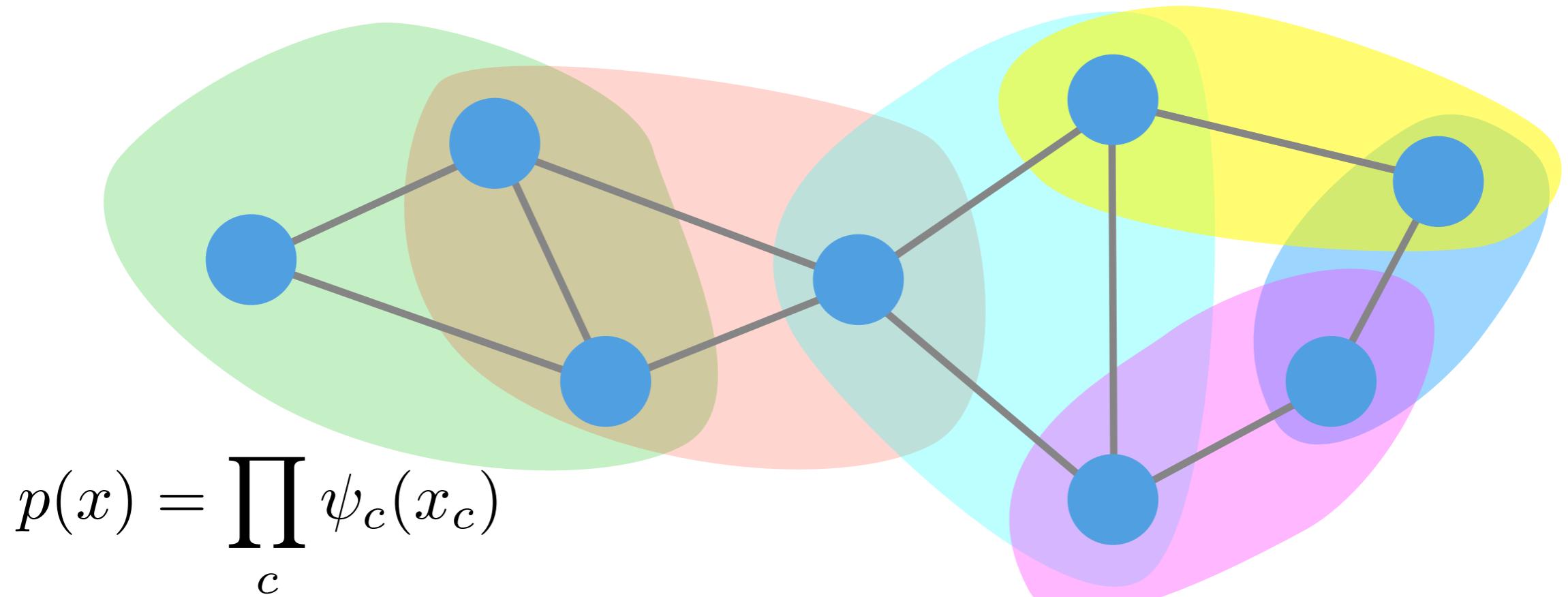
$$\underset{\theta}{\text{minimize}} \frac{1}{2\sigma^2} \|\theta\|^2 + \sum_i \log 2 \cosh \langle \phi(x_i), \theta \rangle - y_i \langle \phi(x_i), \theta \rangle$$

- **Prediction**

$$p(y|x, \theta) = \frac{e^{y\langle \phi(x), \theta \rangle}}{e^{\langle \phi(x), \theta \rangle} + e^{-\langle \phi(x), \theta \rangle}} = \frac{1}{1 + e^{-2y\langle \phi(x), \theta \rangle}}$$



Exponential Clique Decomposition


$$p(x) = \prod_c \psi_c(x_c)$$

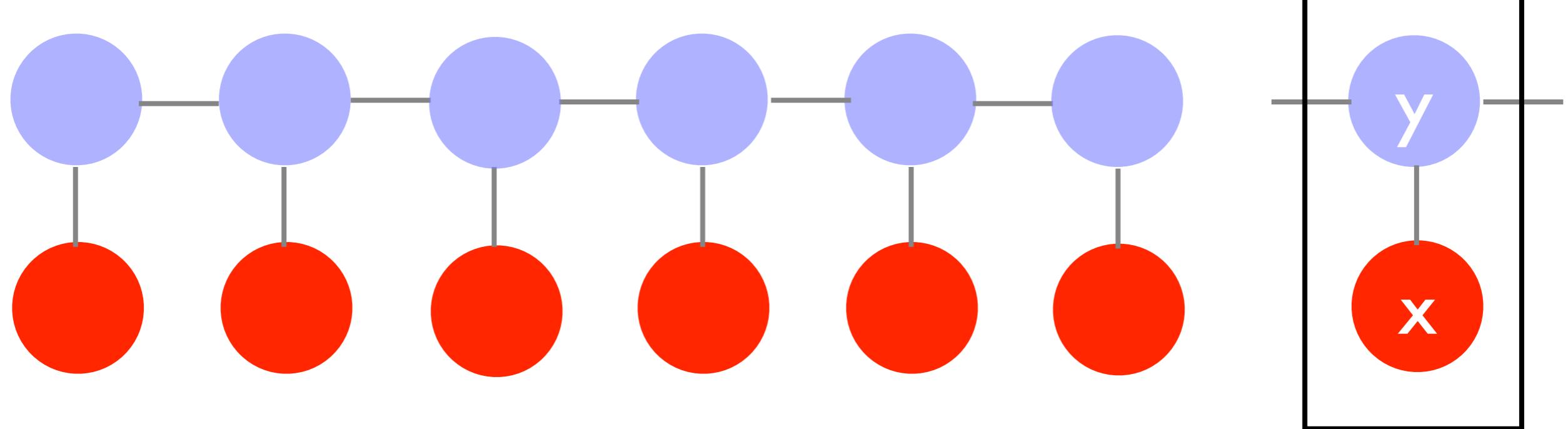
Theorem: Clique decomposition holds in sufficient statistics

$$\phi(x) = (\dots, \phi_c(x_c), \dots) \text{ and } \langle \phi(x), \theta \rangle = \sum_c \langle \phi_c(x_c), \theta_c \rangle$$

Corollary: we only need expectations on cliques

$$\mathbf{E}_x[\phi(x)] = (\dots, \mathbf{E}_{x_c} [\phi_c(x_c)], \dots)$$

Conditional Random Fields



$$\phi(x) = (y_1 \phi_x(x_1), \dots, y_n \phi_x(x_n), \phi_y(y_1, y_2), \dots, \phi_y(y_{n-1}, y_n))$$

$$\langle \phi(x), \theta \rangle = \sum_i \langle \phi_x(x_i, y_i), \theta_x \rangle + \sum_i \langle \phi_y(y_i, y_{i+1}), \theta_y \rangle$$

$$g(\theta|x) = \sum_y \prod_i f_i(y_i, y_{i+1}) \text{ where}$$

$$f_i(y_i, y_{i+1}) = e^{\langle \phi_x(x_i, y_i), \theta_x \rangle + \langle \phi_y(y_i, y_{i+1}), \theta_y \rangle}$$

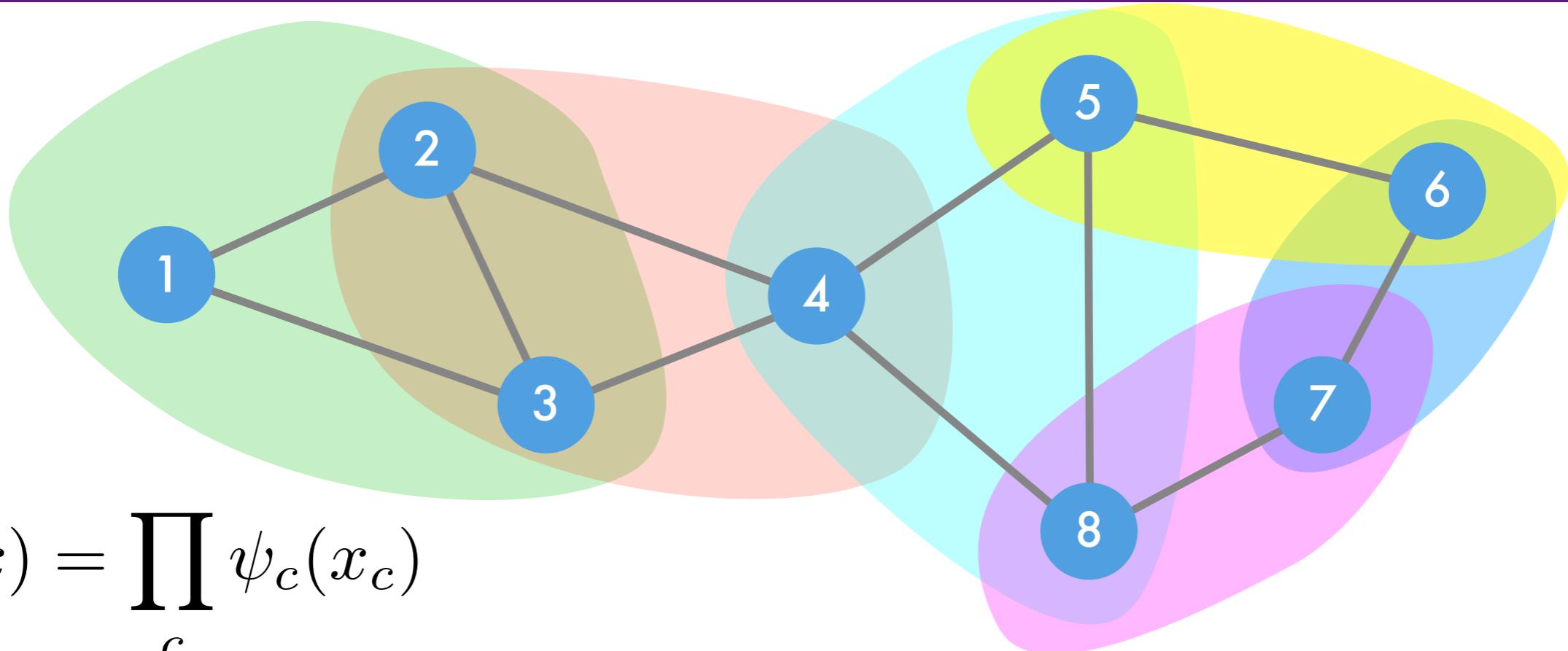
dynamic
programming

Conditional Random Fields

- Compute distribution over marginal and adjacent labels
- Take conditional expectations
- Take update step (batch or online)
- More general techniques for computing normalization via message passing ...

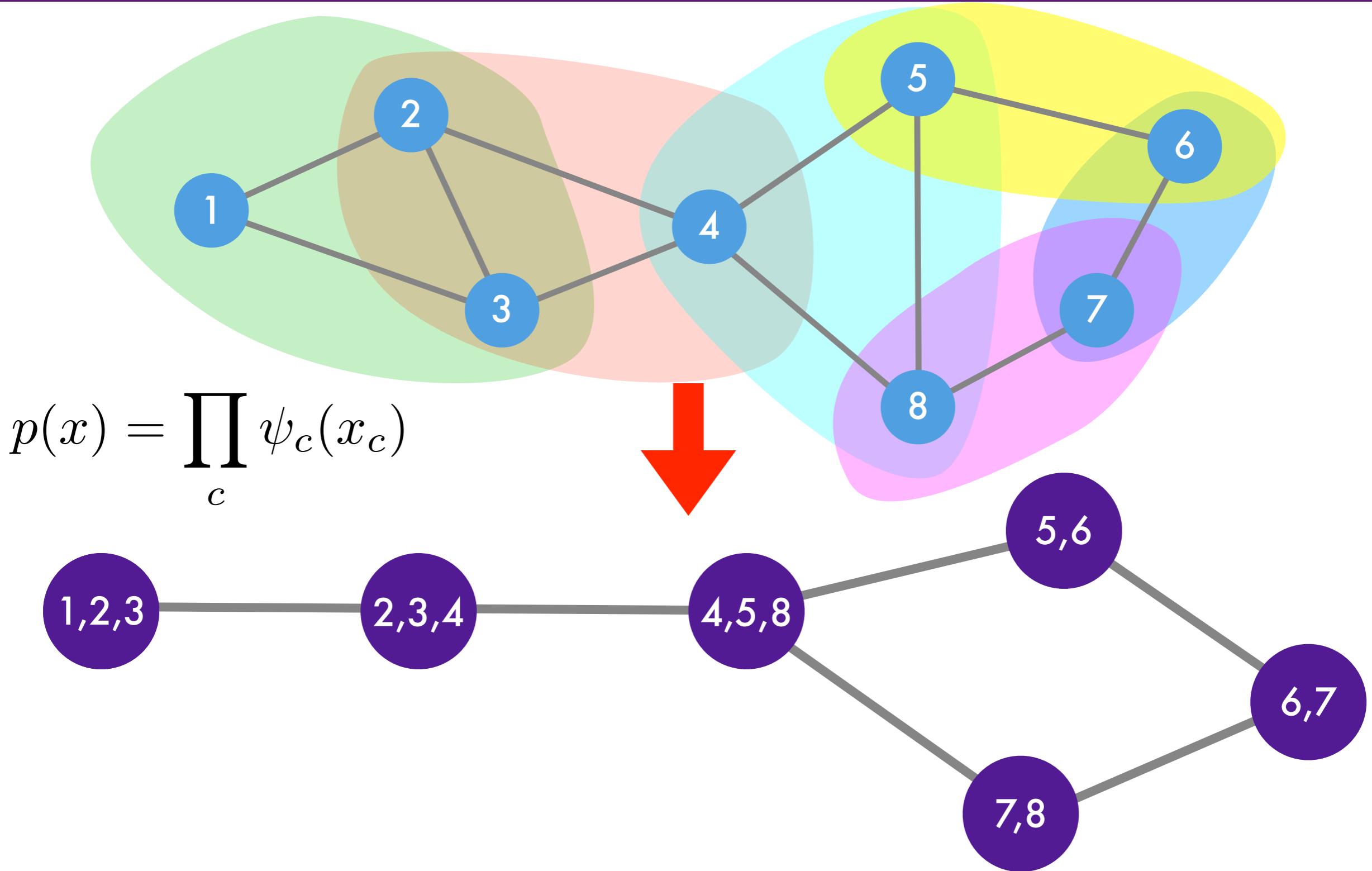
Dynamic Programming + Message Passing

Clique Graph

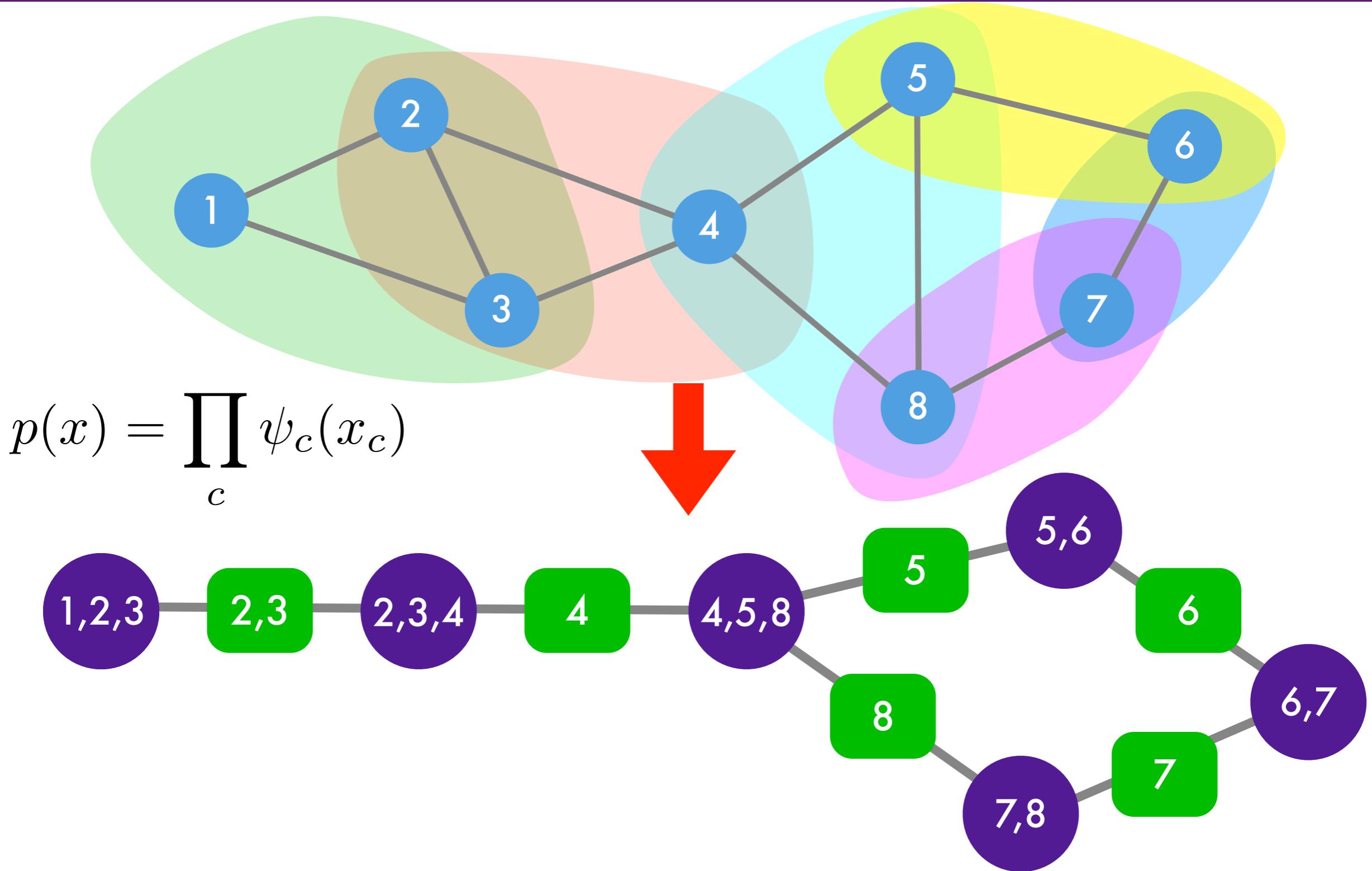


$$p(x) = \prod_c \psi_c(x_c)$$

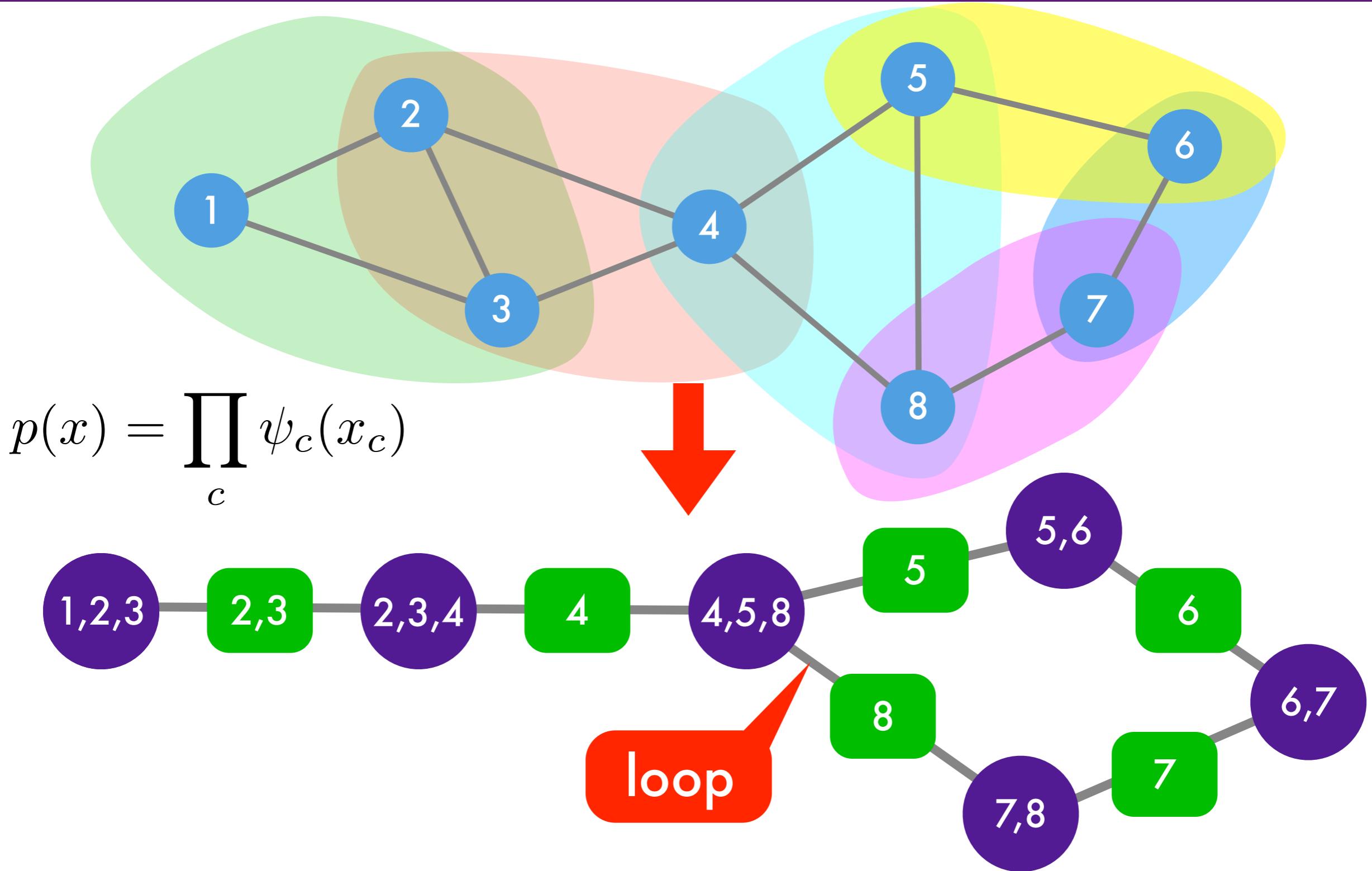
Clique Graph



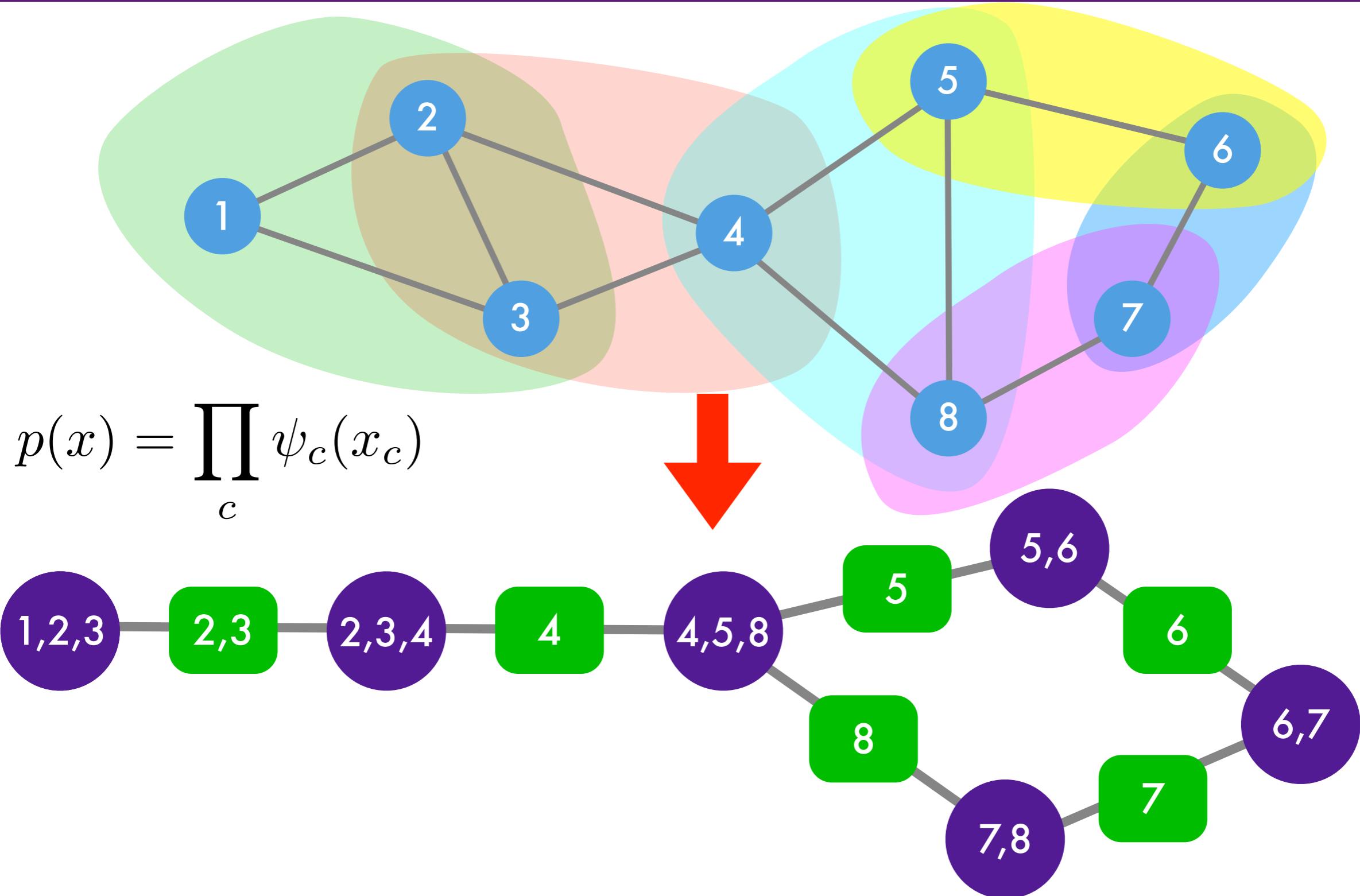
Clique Graph



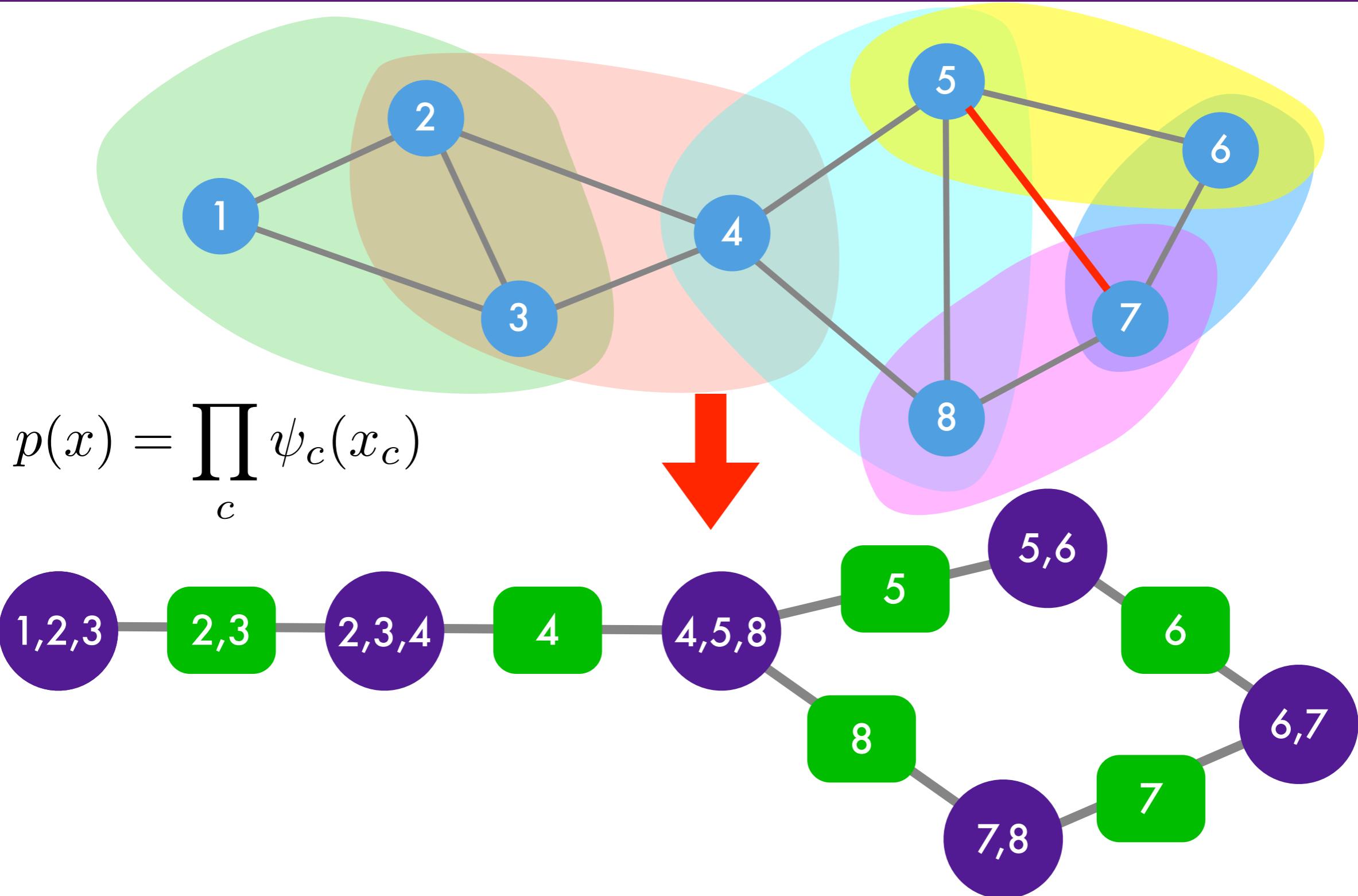
Clique Graph



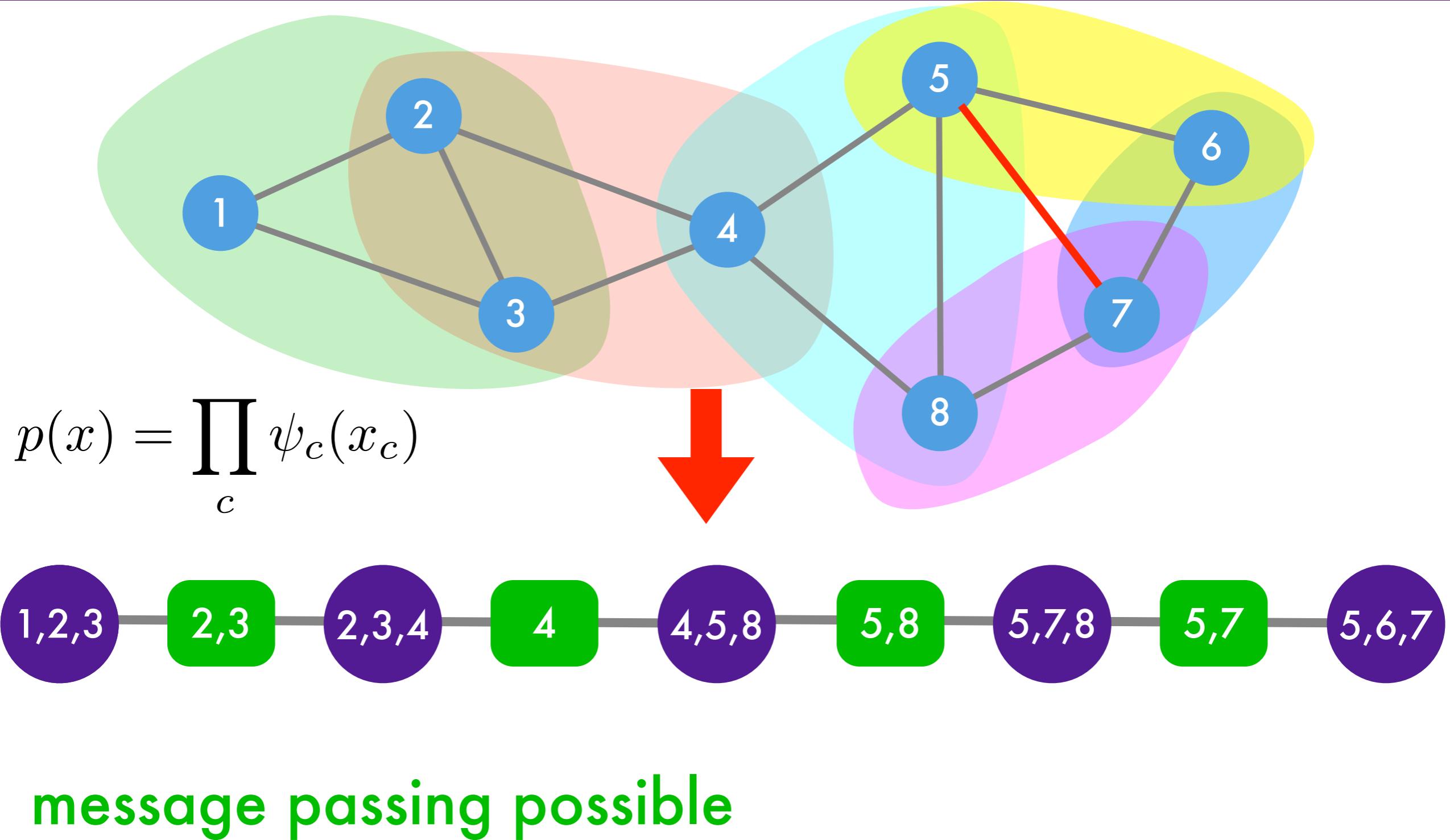
Junction Tree / Triangulation



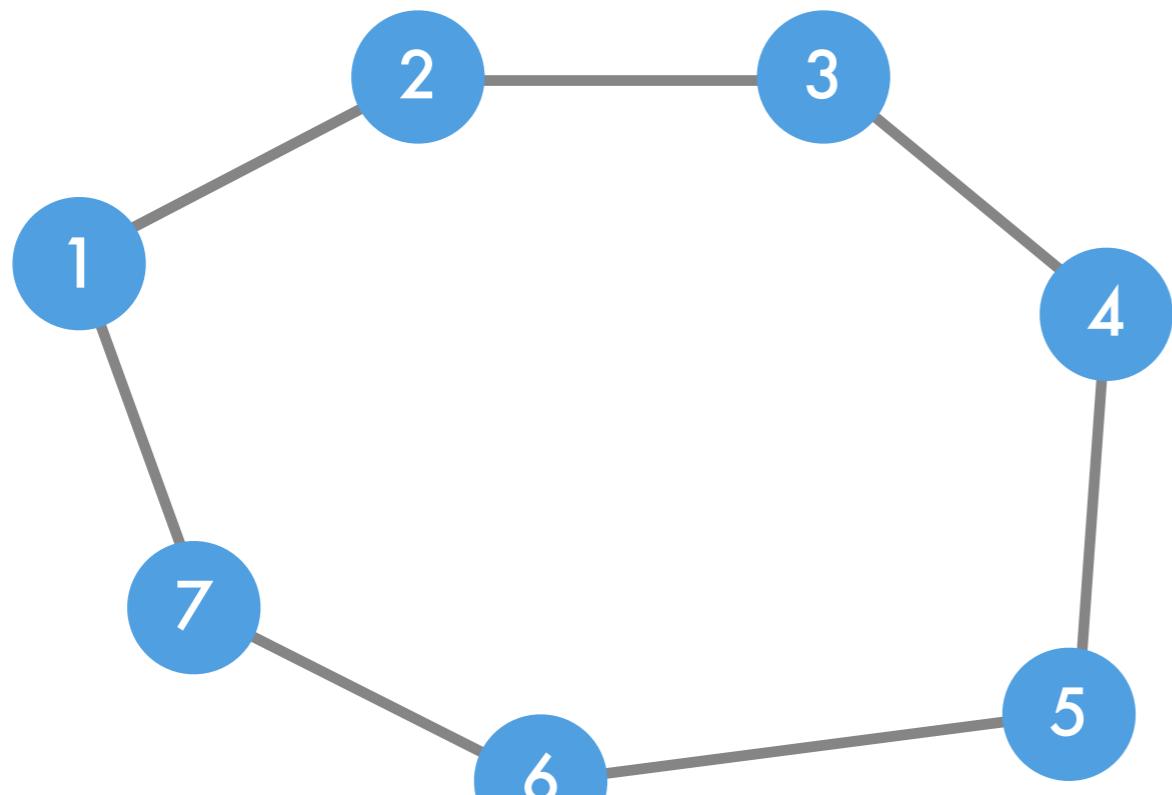
Junction Tree / Triangulation



Junction Tree / Triangulation

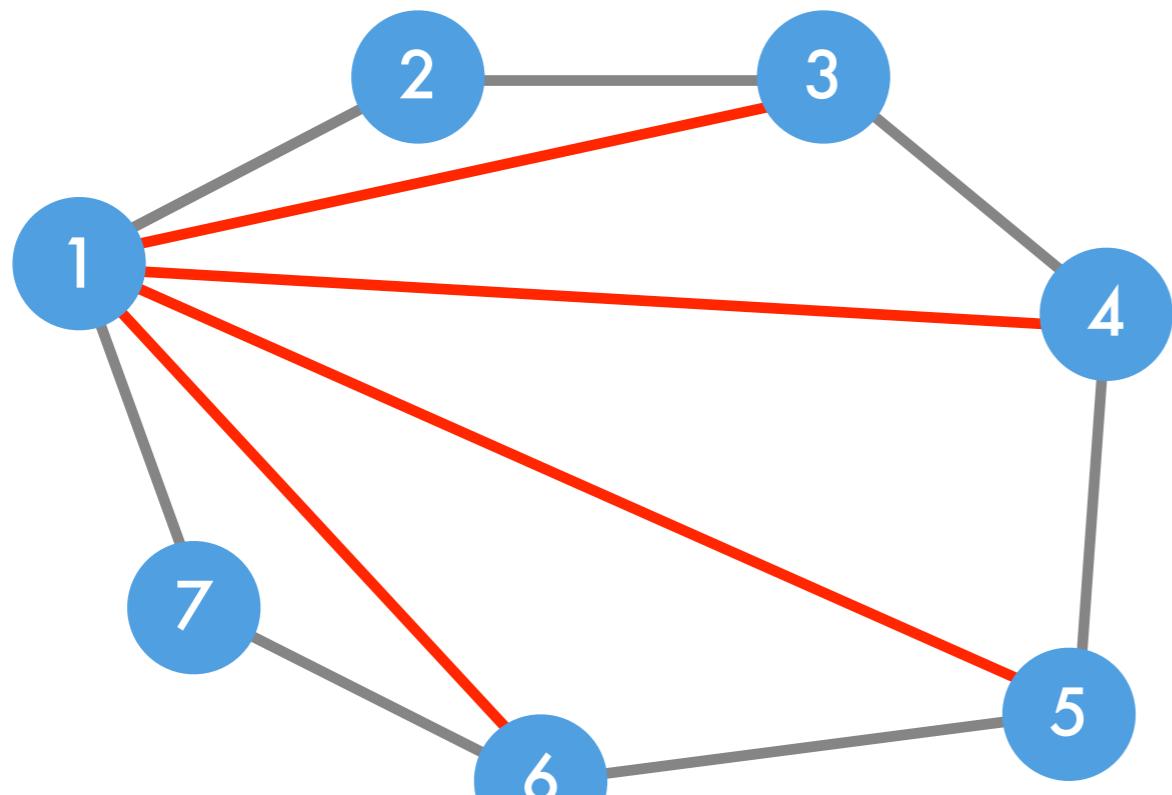


Triangulation Examples



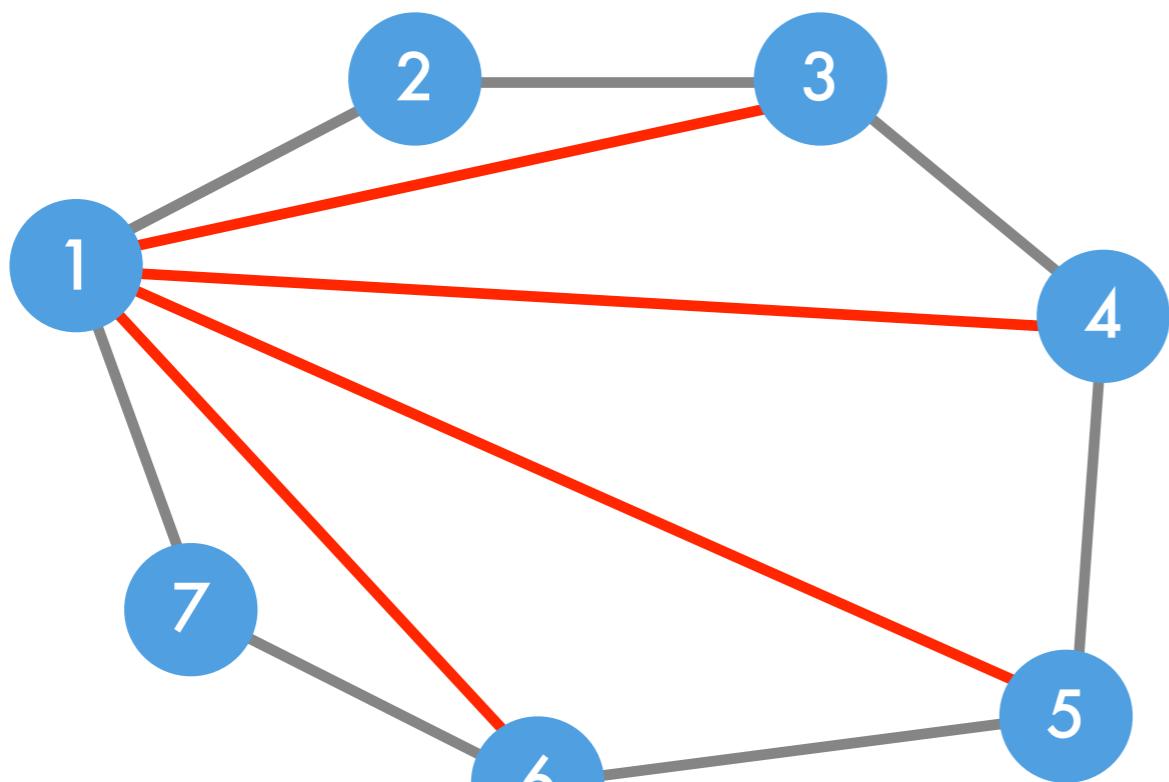
- **Clique size increases**
- **Separator set size increases**

Triangulation Examples



- **Clique size increases**
- **Separator set size increases**

Triangulation Examples



- Clique size increases
- Separator set size increases



Message Passing



- **Joint Probability**

$$p(x) \propto \psi(x_1, x_2, x_3) \psi(x_1, x_3, x_4) \psi(x_1, x_4, x_5) \psi(x_1, x_5, x_6) \psi(x_1, x_6, x_7)$$

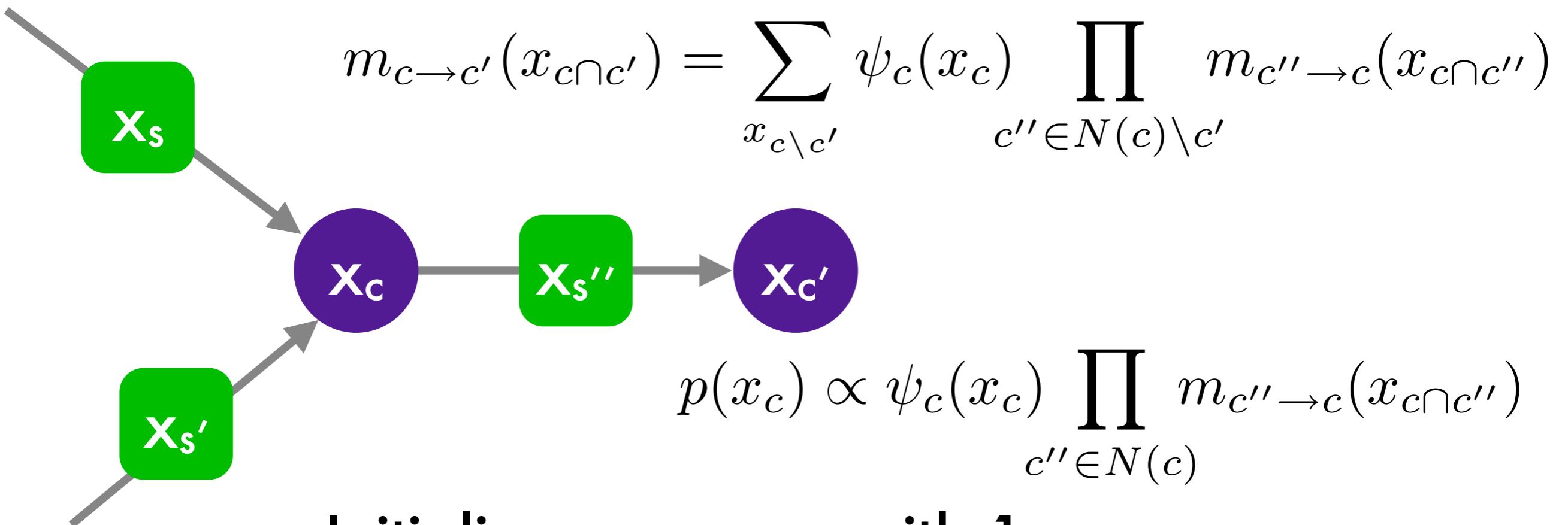
- **Computing the normalization**

$$m_{\rightarrow}(x_1, x_3) = \sum_{x_2} \psi(x_1, x_2, x_3)$$

$$m_{\rightarrow}(x_1, x_4) = \sum_{x_3} m_{\rightarrow}(x_1, x_3) \psi(x_1, x_3, x_4)$$

$$m_{\rightarrow}(x_1, x_5) = \sum_{x_4} m_{\rightarrow}(x_1, x_4) \psi(x_1, x_4, x_5)$$

Message Passing



- Initialize messages with 1
- Guaranteed to converge for (junction) trees
- Works well in practice even for loopy graphs
- Only local computations are required

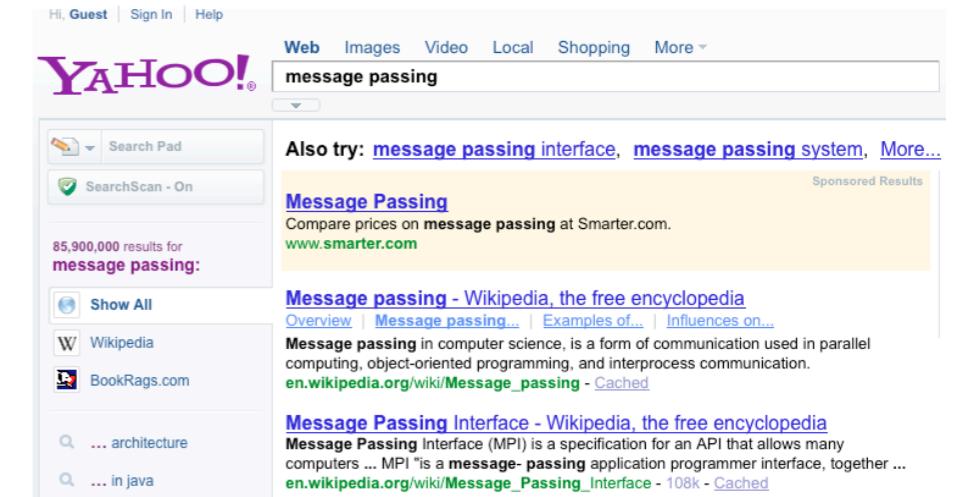
Message Passing in Practice

- Incoming messages contain aggregate uncertainty from neighboring random variables
- Message passing combines and transmits this information **in both directions**

crawler

tiering

phase 1
ranker



Message Passing in Practice

- Incoming messages contain aggregate uncertainty from neighboring random variables
- Message passing combines and transmits this information **in both directions**



Part 5 - Scalable Topic Models

Topic models

Grouping objects

Grouping objects

The image displays three distinct web pages, each featuring a prominent red speech bubble containing the word "Singapore".

- Singapore Airlines:** The top navigation bar includes links for Help, Site Map, Contact Us, Singapore (highlighted in yellow), Change Location, and a search bar. Below the navigation is a menu with tabs: The Experience, Flights & Fares, Before You Fly, Loyalty Programmes, and Promotions.
- National University of Singapore (NUS):** The header features the NUS logo and name. It includes links for myEMAIL, IVLE, LIBRARY, MAPS, CALENDAR, SITEMAP, CONTACT, and e-CARDS. A search bar is also present. The main menu includes tabs: ABOUT NUS, GLOBAL, ADMISSIONS, ENTERPRISE, CAMPUS LIFE, GIVING, and CAREERS@NUS.
- Chijmes:** The page features a large image of a historic building at night. Text on the left reads: "Discover a century of resplendent living history behind the cloistered walls." Below this is another text block: "Chijmes, a premier lifestyle destination in Singapore". At the bottom, it says "Owned by: SUNTEC Real Estate Investment Trust", "Managed by: ARA Asset Resource Asia", and "Property Manager: PAC Pacific Asset Management Pte Ltd".

A large red speech bubble is overlaid on the NUS page, containing the word "Singapore". Another red speech bubble is overlaid on the Chijmes page, also containing the word "Singapore".

Bottom Right Corner: The Yahoo! logo is visible in the bottom right corner.

Page Footer: The footer of the Chijmes page includes links for STAFF, ALUMNI, and VISITORS. The footer of the NUS page includes links for myEMAIL, IVLE, LIBRARY, MAPS, CALENDAR, SITEMAP, CONTACT, and e-CARDS.

Page Bottom: The footer of the NUS page includes links for myEMAIL, IVLE, LIBRARY, MAPS, CALENDAR, SITEMAP, CONTACT, and e-CARDS.

Grouping objects

UNITED

My profile | Worldwide sites | Customer service

Planning & booking | Reservations & check-in | Mileage Plus® | Services & information | Search site

#**ON TIME** United. #1 in on-time arrivals. [Details](#)

Flights | Check-in | Flight status

BOOK FLIGHT | REDEEM MILES

From (Find airport) To (Find airport)

Search nearby airports Search nearby airports

Roundtrip One-way > Multicity

Departing Anytime

Returning Anytime

Search by Schedule & price Price > Flexible

Adult (child or senior?)

Cabin Refundable

Promotion code or Electronic certificate More info

+ Log in to view all seating options

>> Advanced Search

Cars | Hotels | Vacations

Use 30% fewer miles on your next United flight.

Save now on Saver Awards for flights 700 miles or less. [Learn more](#)

United news and deals

- > Travel waiver issued due to Hurricane Earl
- > E-Fares: Save on weekend getaways
- > Opt to send your bags ahead
- > Wireless check-in, paperless boarding
- > Receive deal alerts: Follow us on Twitter
- > Take our survey & you could win miles

United-Continental merger [Learn more about the merger](#)

© 1998-2010 Chez Panisse Restaurant

A STAR ALLIANCE MEMBER

Singapore Change Location Search

Before You Fly Loyalty Programmes Promotions

CALENDAR SITEMAP CONTACT e-CARDS

GIVING CAREERS@NUS

Search ANU...

WEB CONTACTS MAP GO

The Australian National University

CURRENT STUDENTS RESEARCH & EDUCATION ABOUT ANU STAFF

Forests renew after Black Saturday fires

School of Music at Floriade

Undergraduate studies

Higher Degree Research

Grouping objects

UNITED

Planning & booking | Reservations & check-in | Mileage Plus® | Services & information | Search site

#ON TIME United, #1 in on-time arrivals. Details

Flights | Check-in | Flight status

BOOK FLIGHT | REDEEM MILES

From (Find airport) To (Find airport)

Search nearby airports | Roundtrip | One-way | Multicity

Departing Anytime

Returning Anytime

Search by Schedule & price | Price | Flexible

Adult (child or senior?)

Cabin Economy | Refundable

Promotion code or Electronic certificate More info

Log in to view all seating options

Advanced Search | Search

Cars | Hotels | Vacations

Use 30% fewer miles on your next United flight.

%

Save now on Saver Awards for flights 700 miles or less. Learn more

Start earning miles today. Join Mileage Plus united.com benefits and features

> Low Fare Guarantee | > Why united.com? Now!

Travel information

> Updates to baggage & standby policies | > View travel requirements and regulations

Play now. Win now. Opt-in mobile and online games. The gold medal is a million miles.

Earn up to 30,000 Bonus Miles

United news and deals

- > Travel waiver issued due to Hurricane Earl
- > E-Fares: Save on weekend getaways
- > Opt to send your bags ahead
- > Wireless check-in, paperless boarding
- > Receive deal alerts: Follow us on Twitter
- > Take our survey & you could win miles

United-Continental merger Learn more about the merger

About United | Investor relations | Business resources | Careers | Site map

A STAR ALLIANCE MEMBER

SGD 093

Book Now | Show Schedule

SIA Holidays

SGD 824*

Book Now

Singapore - Shanghai SGD 824*

Book Now

Singapore - Sydney SGD 983*

Book Now

Singapore - Bangkok SGD 395*

Book Now

(Ball)

Singapore - Hong Kong SGD 546*

Book Now

Singapore - Taipei SGD 768*

Book Now

Singapore - Tokyo (Haneda) SGD 983*

Book Now

Singapore - London

chez Panisse

RESERVATIONS
RESTAURANT & CAFÉ

MENUS
RESTAURANT • CAFÉ
MONDAY NIGHTS • WINE LIST

ABOUT
CHEZ PANISSE • ALICE WATERS
OUR CHEFS • FRIENDS • PRESS
FOUNDATION & MISSION

SPECIAL EVENTS
CALENDAR

STORE
BOOKS • POSTERS • GIFTS

CONTACT
INFORMATION
DIRECTIONS • MAILING LIST



EXPLORE ANU » A-Z INDEX »

Search ANU... WEB CONTACTS MAP GO

ANU
THE AUSTRALIAN NATIONAL UNIVERSITY

HOME FUTURE STUDENTS CURRENT STUDENTS RESEARCH & EDUCATION ABOUT ANU STAFF

Ash forests rise and rise again

A new book that graphically documents the spectacular natural recovery of Victoria's ash forests after the Black Saturday bushfires also argues that wildfires are typical natural disturbances in these environments.

» read more

Forests renew after Black Saturday fires School of Music at Floriade Undergraduate studies Higher Degree Research

KRISFLYER

Singapore - Bangkok SGD 395*	Book Now
Singapore - Hong Kong SGD 546*	Book Now
Singapore - Taipei SGD 768*	Book Now
Singapore - Tokyo (Haneda) SGD 983*	Book Now
Singapore - London	

PROSPECTIVE STUDENTS CURRENT STUDENTS STAFF ALUMNI VISITORS

at NUS WATCH THE VIDEO

Joint Evacuation Exercises

- 7 & 14 Sept 2010
- 10am - 12pm
- Heng Mui Keng Terrace & vicinity

MORE DETAILS

YAHOO!

Grouping objects

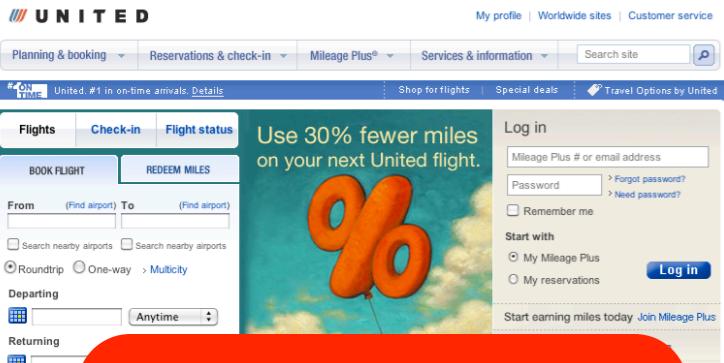
The screenshot shows the United Airlines website's homepage. It features a search bar at the top with fields for 'From' and 'To' airports, and dropdowns for 'Departing' and 'Returning' dates. Below this is a large orange speech bubble containing the word 'airline'. To the left of the speech bubble is a promotional banner for 'Saver Awards' flights 700 miles or less, featuring a large orange percentage sign. To the right is a section titled 'Travel information' with a photo of a smiling woman.

The screenshot shows the ANU website's homepage. At the top, there are links for 'EXPLORE ANU' and 'A-Z INDEX'. A search bar is located at the top right. Below the header is a large banner with a photo of a tree and the text 'Ash forests rise and rise again'. To the right of the banner is a section for 'Higher Degree Research'. A large red speech bubble containing the word 'university' is overlaid on the bottom right of the page.

The screenshot shows the Chez Panisse website. On the left is a sidebar with links for 'RESERVATIONS', 'MENUS', 'ABOUT', 'SPECIAL EVENTS', 'STORE', and 'CONTACT'. The main content area features a large photo of the restaurant's exterior and interior. A large red speech bubble containing the word 'restaurant' is overlaid on the bottom center of the page.

YAHOO!

Grouping objects



UNITED

Planning & booking | Reservations & check-in | Mileage Plus® | Services & information | Search site

#1 in on-time arrivals. Details

Flights | Check-in | Flight status

BOOK FLIGHT | REDEEM MILES

From (Find airport) To (Find airport)

Search nearby airports | Search nearby airports

Roundtrip | One-way | Multicity

Depart Anytime

Returning

Use 30% fewer miles on your next United flight.

%

Log in

Mileage Plus # or email address

Password

Forgot password? Need password?

Remember me

Start with

My Mileage Plus | My reservations

Log in

Start earning miles today. Join Mileage Plus

RESERVATIONS

RESTAURANT & CAFÉ

MENUS

RESTAURANT • CAFÉ

MONDAY NIGHTS • WINE LIST

ABOUT

CHEZ PANISSE • ALICE WATERS

OUR CHEFS • FRIENDS • PRESS

FOUNDATION & MISSION

SPECIAL EVENTS

CALENDAR

STORE

BOOKS • POSTERS • GIFTS

CONTACT

INFORMATION

DIRECTIONS • MAILING LIST

Directions Reservations Contact

© 1998-2010 Chez Panisse Restaurant & Café. All Rights Reserved.

USA



EXPLORE ANU » A-Z INDEX »

Search ANU... WEB CONTACTS MAP GO

The Australian National University

ANU THE AUSTRALIAN NATIONAL UNIVERSITY

HOME FUTURE STUDENTS

Ash forests rise and rise again

A new book that graphically documents the recovery of Victoria's ash forests after bushfires also argues that wildfires disturbances in these environments

» read more

Forests renew after Black Saturday fires School of Music at Floriade Undergraduate studies Higher Degree Research

STAFF

Australia



NUS National University of Singapore

myEMAIL IVLE LIBRARY MAPS CALENDAR SITEMAP CONTACT e-CARDS

Search search for... in NUS Websites GO

SINGAPORE AIRLINES

The Experience Flights

ABOUT NUS GLOBAL ADMISSIONS EDUCATION RESEARCH ENTERPRISE CAMPUS LIFE GIVING CAREERS@NUS

Book a Flight Check In Flight Status My Bookings

Help Site Map Cor

Round Trip One Way Stopover/Multi-city

From: Departure City Depart: KrisFly

To: Destination City Return: Sal

Must travel on these dates

Adults: Children (2-11): Infants:

Need Help? View Book A Flight

SIA Holidays Hotel Bookings

A Leading Global University Centred in Asia

CLICK HERE TO FIND OUT MORE

Flame Arrival Ceremony at NUS WATCH THE VIDEO

Joint Evacuation Exercises

7 & 14 Sept 2010 10am - 12pm

Heng Mui Keng Terrace & vicinity

MORE DETAILS

CHIJMES

restau

Discover living here

Chijmes, a premier lifestyle destination in Singapore

Owned by: SUNTEC Real Estate Investment Trust Managed by: ARA Property Manager: APC

Copyright © 2006 Chijmes. All rights reserved.

Feedback Terms & Conditions

ALUMNI VISITORS

YAHOO!

Singapore

Topic Models

UNITED

Planning & booking | Reservations & check-in | Mileage Plus® | Services & information | Search site

#1 United, #1 in on-time arrivals. Details

Flights | Check-in | Flight status

BOOK FLIGHT | REDEEM MILES

From (Find airport) To (Find airport)

Departing Anytime

Returning Anytime

Search by Schedule & price | Price | Flexi

Adult 1 (child or senior?)

Cabin Economy | Refundable

Promotion code or Electronic certificate

Log in to view all seating options

Advanced Search | Search

Cars | Hotels | Vacations

Use 30% fewer miles on your next United flight.

%

REDEEM MILES

From (Find airport) To (Find airport)

Search nearby airports | Search nearby airports

Roundtrip | One-way | Multicity

Departing Anytime

Returning Anytime

Search by Schedule & price | Price | Flexi

Adult 1 (child or senior?)

Cabin Economy | Refundable

Promotion code or Electronic certificate

Log in to view all seating options

Advanced Search | Search

Cars | Hotels | Vacations

About United | Investor relations | Business resources | Careers | Site map

A STAR ALLIANCE MEMBER

USA
airline

EXPLORE ANU » A-Z INDEX »

Search ANU... WEB CONTACTS

ANU THE AUSTRALIAN NATIONAL UNIVERSITY

HOME FUTURE STUDENTS CURRICULUM

Ash forests rise and rise again

A new book that graphically documents the recovery of Victoria's ash forests after the bushfires also argues that wildfires are typical disturbances in these environments.

» read more

Forests renew after Black Saturday fires School of Music at Floriade Undergraduate studies Higher Degree Research

Australia
university

Help | Site Map | Contact Us | Singapore | Change Location | Search

SINGAPORE AIRLINES

The Experience | Flights & Fares | Before You Fly | Loyalty Programmes | Promotions

Book a Flight | Check In | Flight Status | My Bookings

Member Log-In

Round Trip | One Way | Stopover/Multi-city

From: Departure City

To: Destination City

Must travel on these dates

Adults: Children (2-11): Infants:

Need Help? View Book A Flight Guide | SIA Holidays | Hotel Bookings

Singapore - Bangkok SGD 395* Book Now

Singapore - Hong Kong SGD 546* Book Now

Singapore - Taipei SGD 768* Book Now

Singapore - Tokyo (Haneda) SGD 983* Book Now

Singapore - Sydney Book Now

Singapore - London Book Now

Singapore
airline

NUS National University of Singapore

myEMAIL IVLE LIBRARY MAPS CALENDAR SITEMAP CONTACT e-CARDS

Search search for... in NUS Websites GO

ABOUT NUS GLOBAL ADMISSIONS EDUCATION RESEARCH ENTERPRISE CAMPUS LIFE GIVING CAREERS@NUS

A Leading Global University

Home Arrival Ceremony

Joint Evacuation Exercises

PROSPECTIVE STUDENTS CURRENT STUDENTS STAFF ALUMNI VISITORS

Singapore
university

chez Panisse

RESERVATIONS RESTAURANT & CAFÉ

MENUS RESTAURANT • CAFÉ MONDAY NIGHTS • WINE LIST

ABOUT CHEZ PANISSE • ALICE WATERS OUR CHEFS • FRIENDS • PRESS FOUNDATION & MISSION

SPECIAL EVENTS CALENDAR

STORE BOOKS • POSTERS • GIFTS

CONTACT INFORMATION DIRECTIONS • MAILING LIST

Directions Reservations Contact

USA
food

I.J.M.E.S. restaurants • bars • shops

Discover a century of resplendent living history behind the cloisters

Chijmes, a premier lifestyle destination in Singapore

Owned by: SUNTEC Real Estate Investment Trust Managed by: ARA APAC Investment Management Pte Ltd Property Manager: APAC

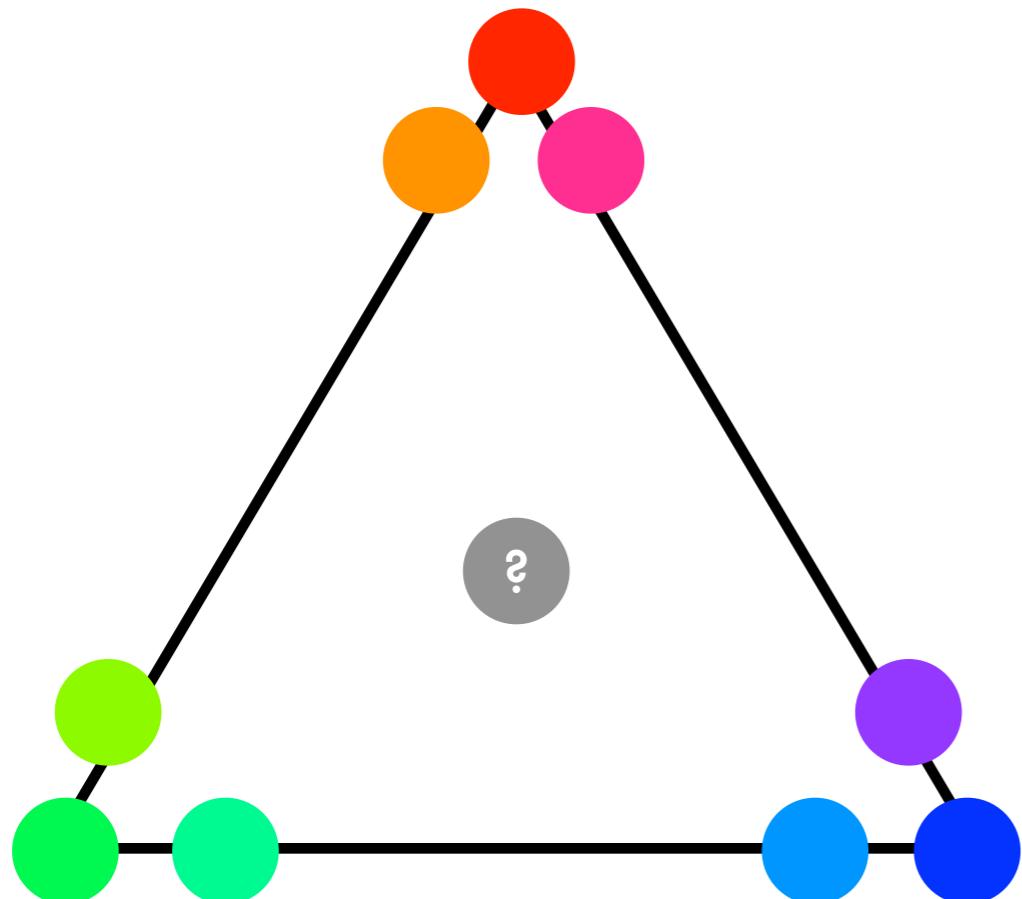
Copyright © 2006 Chijmes. All rights reserved.

Feedback | Terms & Conditions

Singapore
food

Clustering & Topic Models

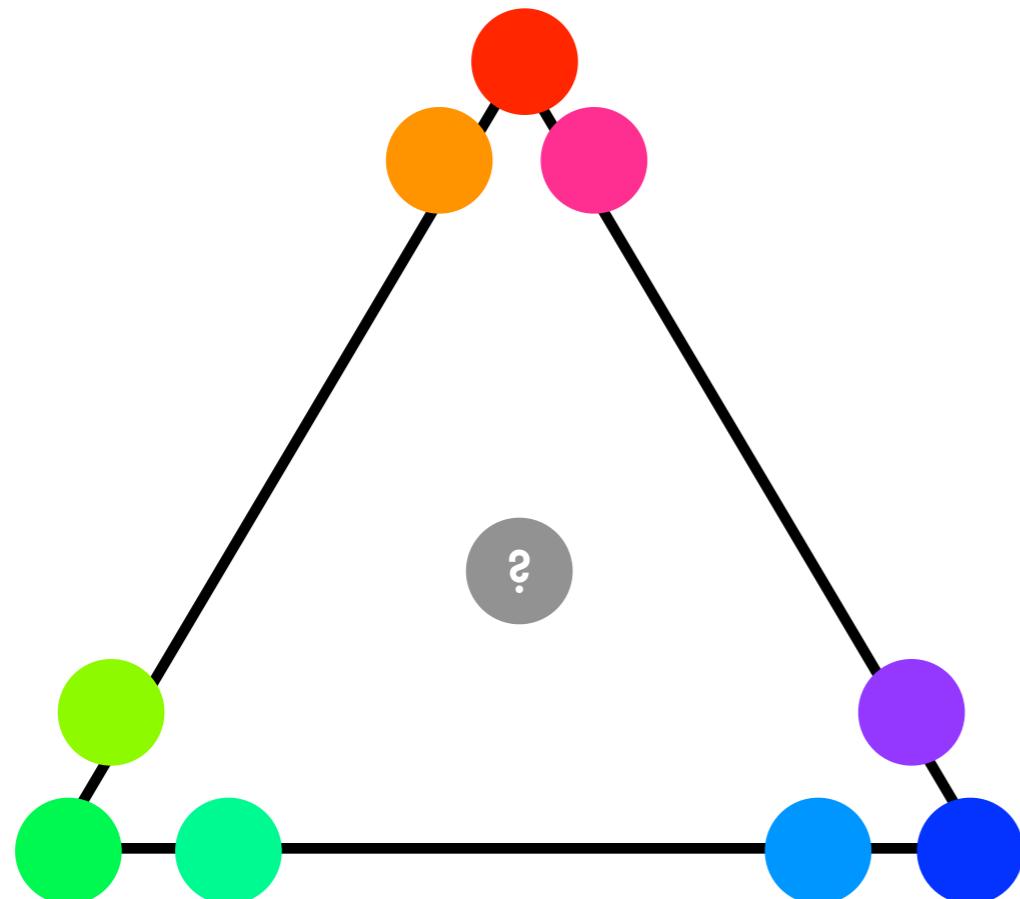
Clustering



group objects
by prototypes

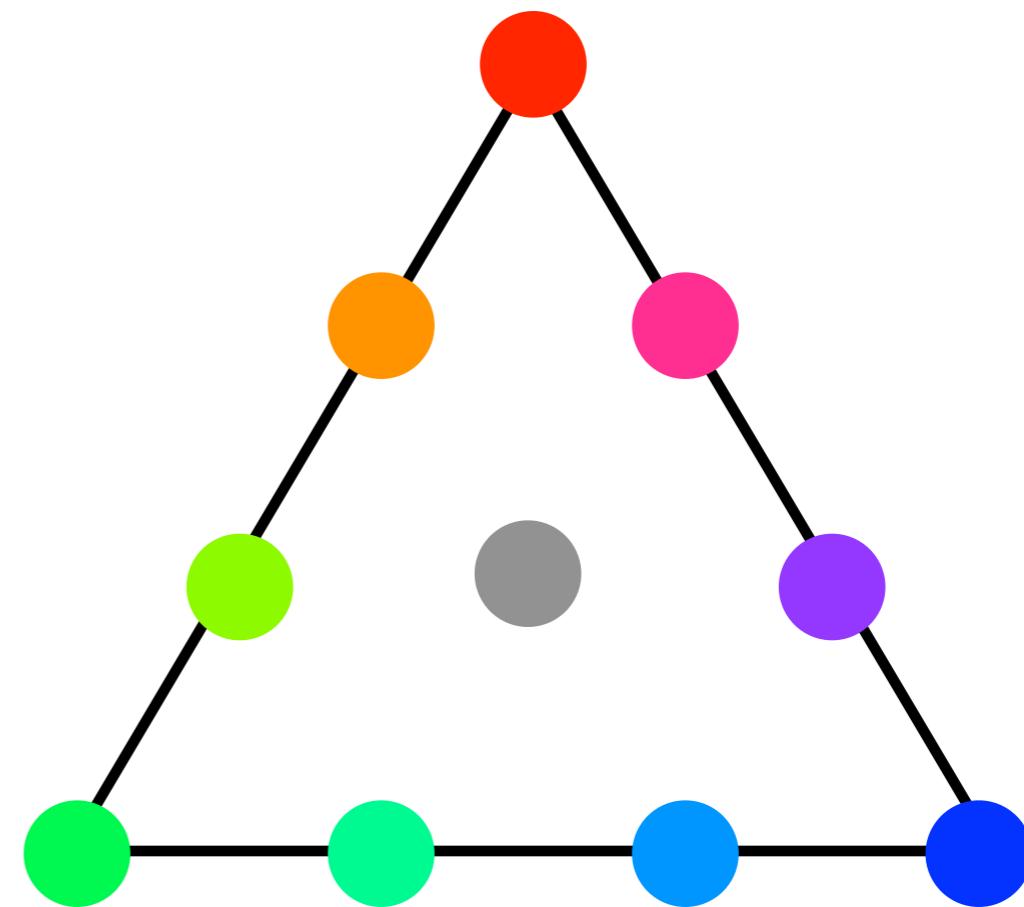
Clustering & Topic Models

Clustering



group objects
by prototypes

Topics

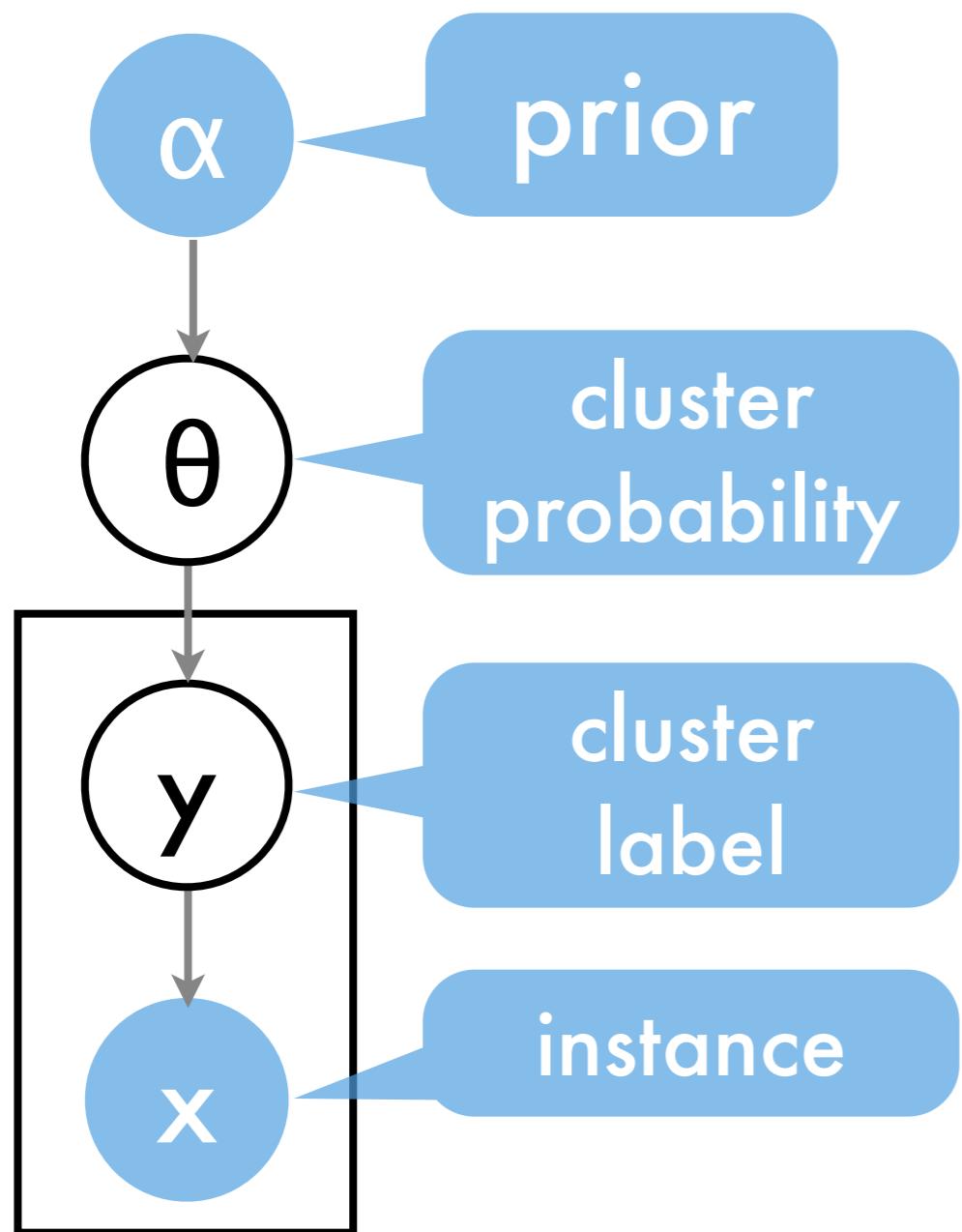


decompose objects
into prototypes

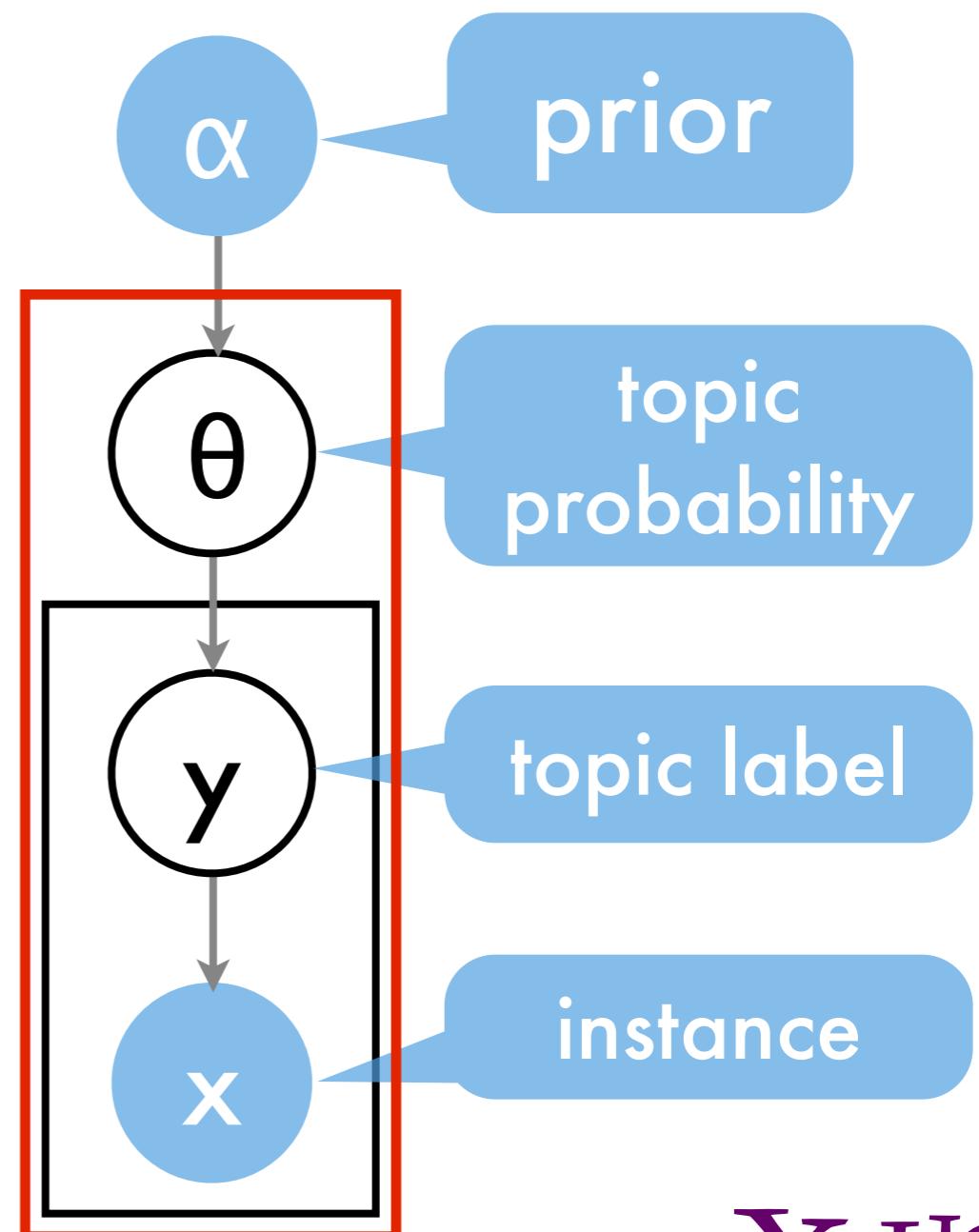
YAHOO!

Clustering & Topic Models

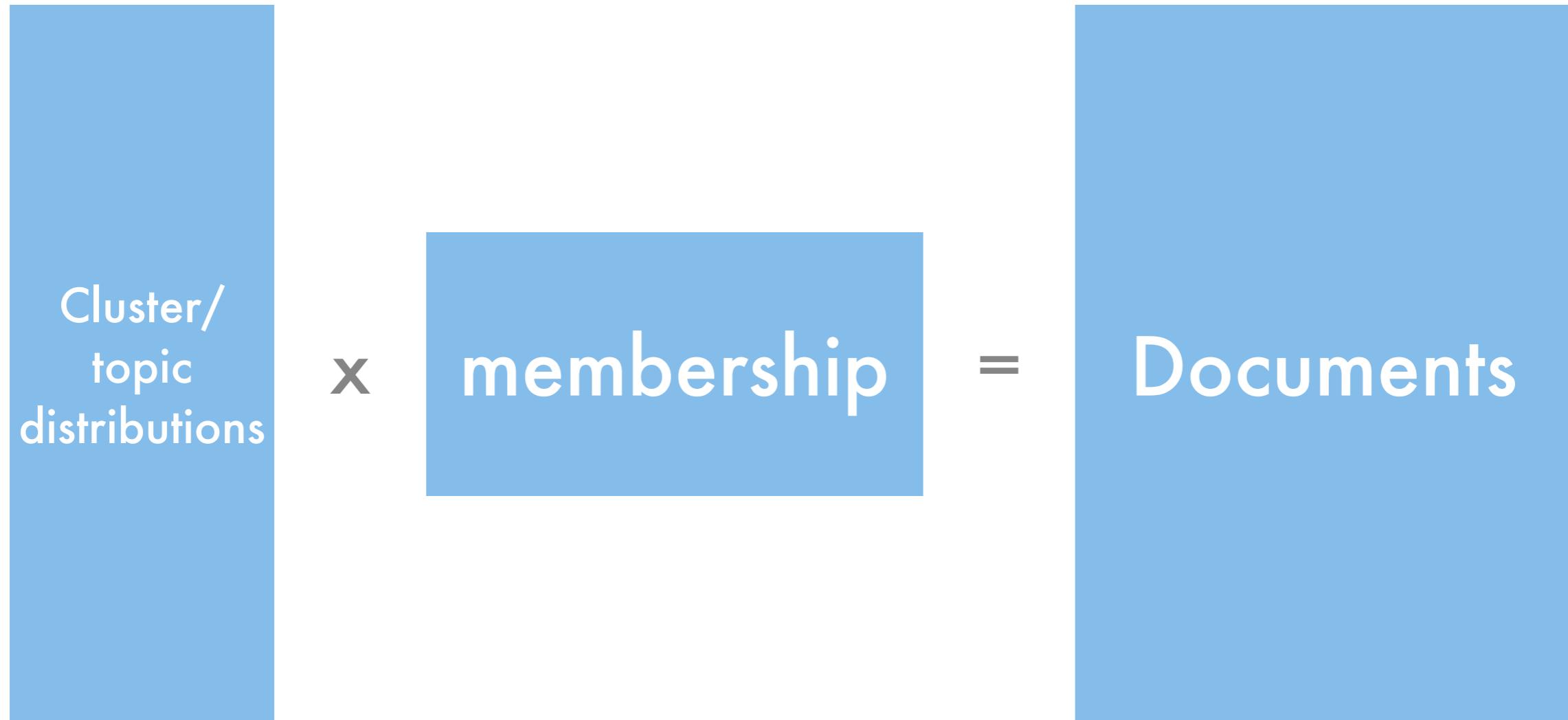
clustering



Latent Dirichlet Allocation



Clustering & Topic Models



clustering: (0, 1) matrix

topic model: stochastic matrix

LSI: arbitrary matrices

Topics in text

The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. “Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services,” Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center’s share will be \$200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive \$400,000 each. The Juilliard School, where music and the performing arts are taught, will get \$250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual \$100,000 donation, too.

Collapsed Gibbs Sampler

Joint Probability Distribution

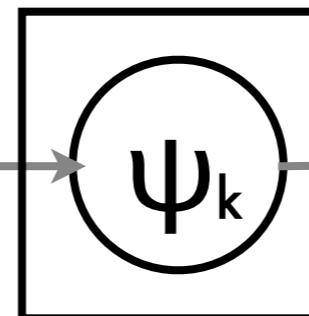
$$p(\theta, z, \psi, x | \alpha, \beta)$$

$$= \prod_{k=1}^K p(\psi_k | \beta) \prod_{i=1}^m p(\theta_i | \alpha)$$

$$\prod_{i,j}^{m, m_i} p(z_{ij} | \theta_i) p(x_{ij} | z_{ij}, \psi)$$

language prior

$$\beta \rightarrow \psi_k$$



α

θ_i

z_{ij}

x_{ij}

topic
probability

topic label

instance

Joint Probability Distribution

sample Ψ
independently

$$p(\theta, z, \psi, x | \alpha, \beta)$$

$$= \prod_{k=1}^K p(\psi_k | \beta) \prod_{i=1}^m p(\theta_i | \alpha)$$

$$\prod_{i,j}^{m, m_i} p(z_{ij} | \theta_i) p(x_{ij} | z_{ij}, \psi)$$

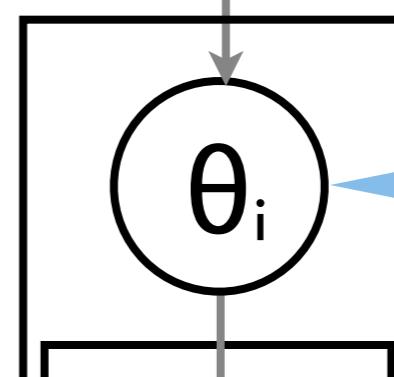
sample z
independently

language prior

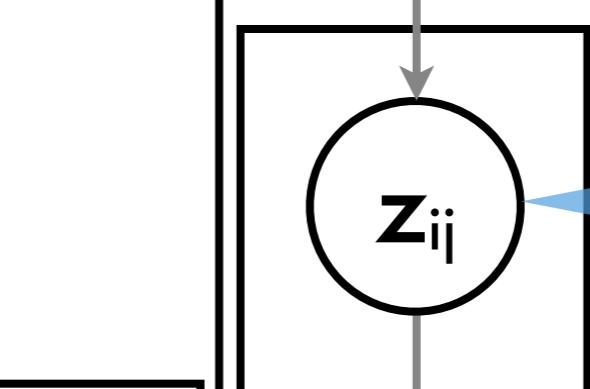
$$\beta$$

sample θ
independently

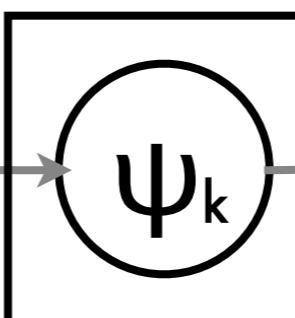
$$\alpha$$



topic
probability



topic label



instance

Joint Probability Distribution

sample Ψ
independently

$$p(\theta, z, \psi, x | \alpha, \beta)$$

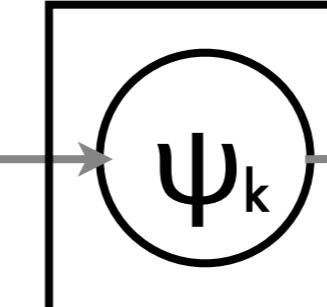
$$= \prod_{k=1}^K p(\psi_k | \beta) \prod_{i=1}^m p(\theta_i | \alpha)$$

$$\prod_{i,j}^{m, m_i} p(z_{ij} | \theta_i) p(x_{ij} | z_{ij}, \psi)$$

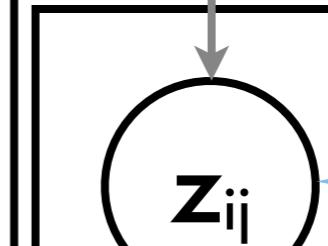
sample z
independently

language prior

$$\beta$$



$$x_{ij}$$



$$\alpha$$

$$\theta_i$$

sample θ
independently

slow

topic
probability

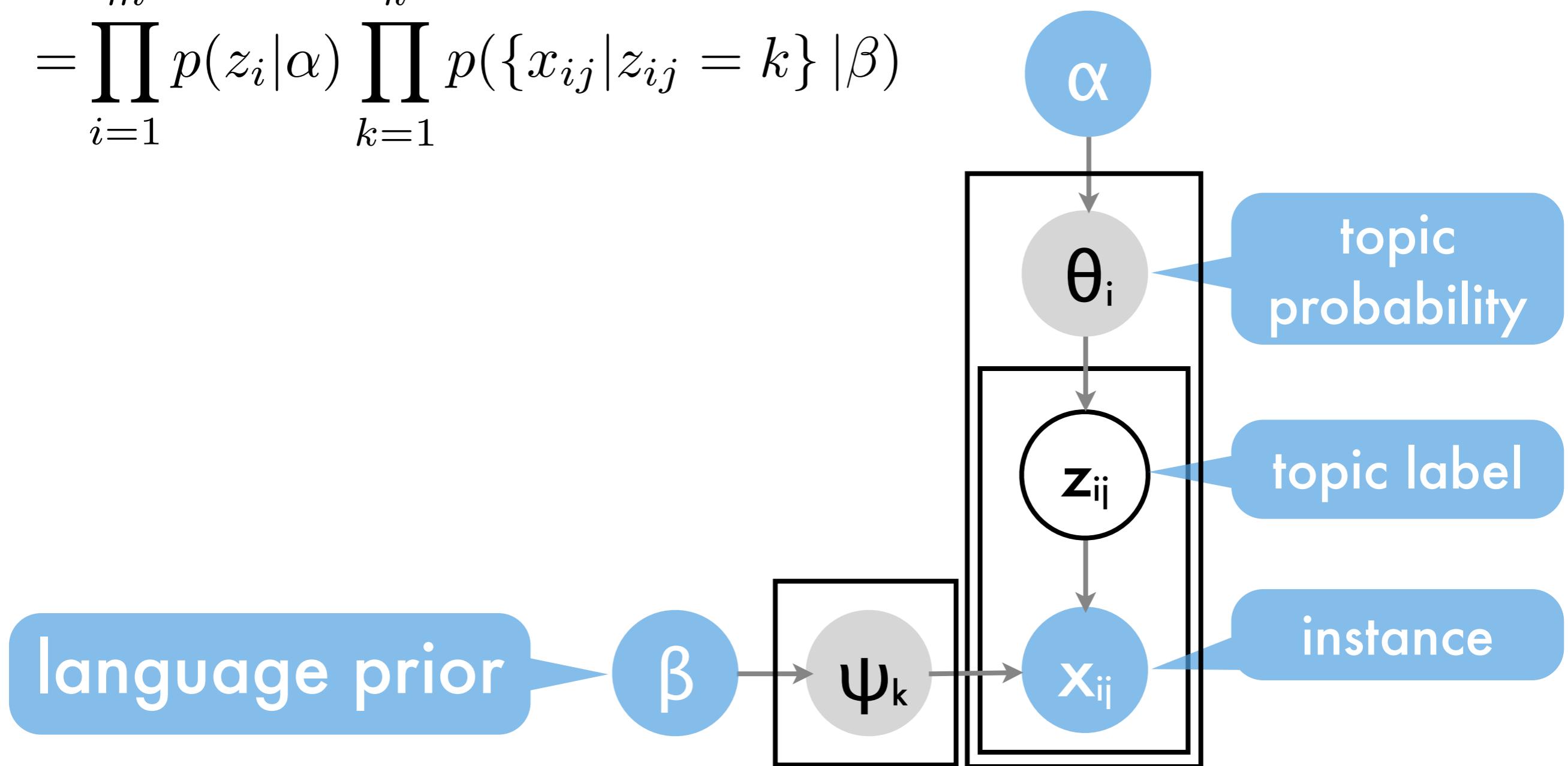
topic
label

instance

Collapsed Sampler

$$p(z, x | \alpha, \beta)$$

$$= \prod_{i=1}^m p(z_i | \alpha) \prod_{k=1}^K p(\{x_{ij} | z_{ij} = k\} | \beta)$$



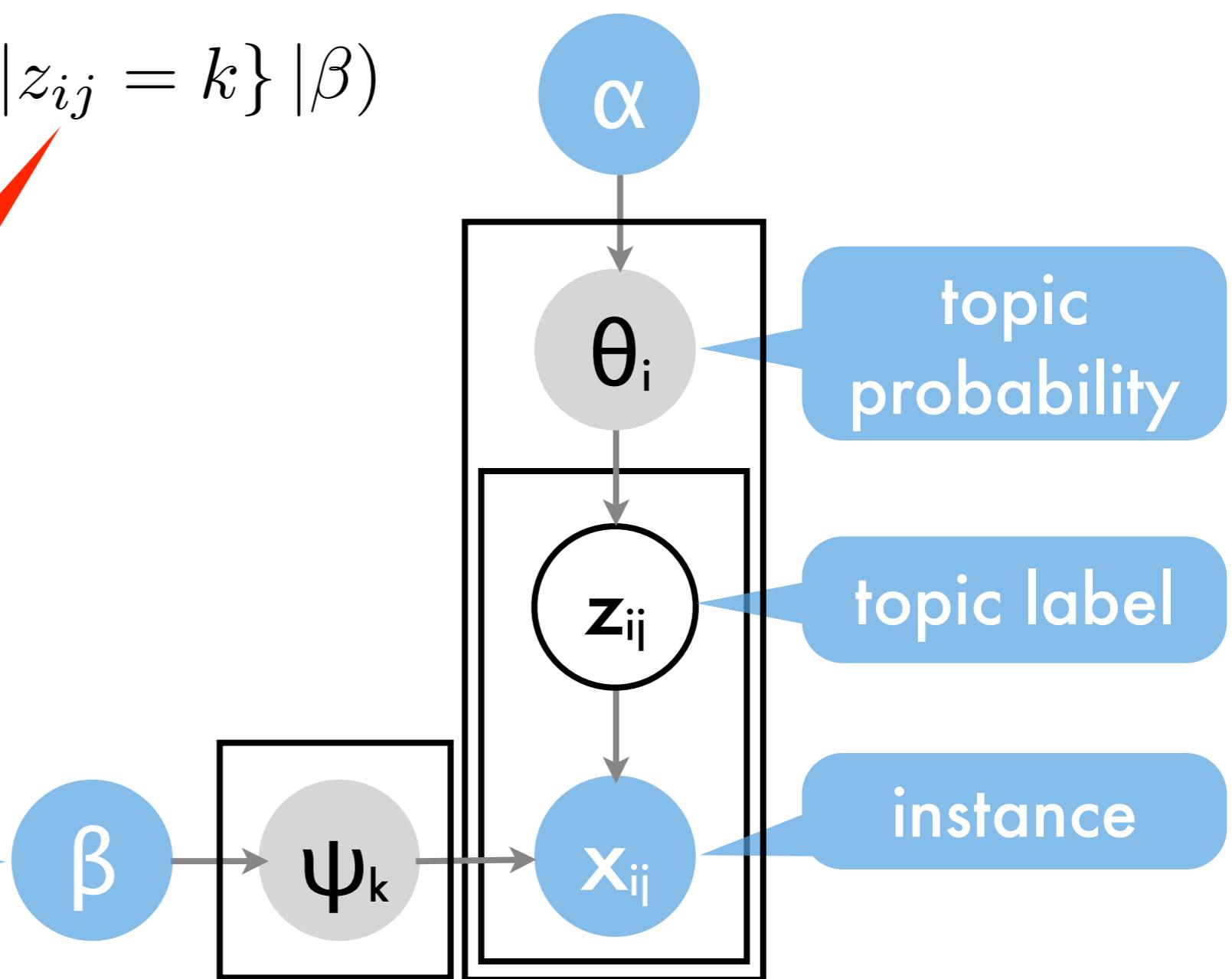
Collapsed Sampler

$$p(z, x | \alpha, \beta)$$

$$= \prod_{i=1}^m p(z_i | \alpha) \prod_{k=1}^k p(\{x_{ij} | z_{ij} = k\} | \beta)$$

sample z
sequentially

language prior

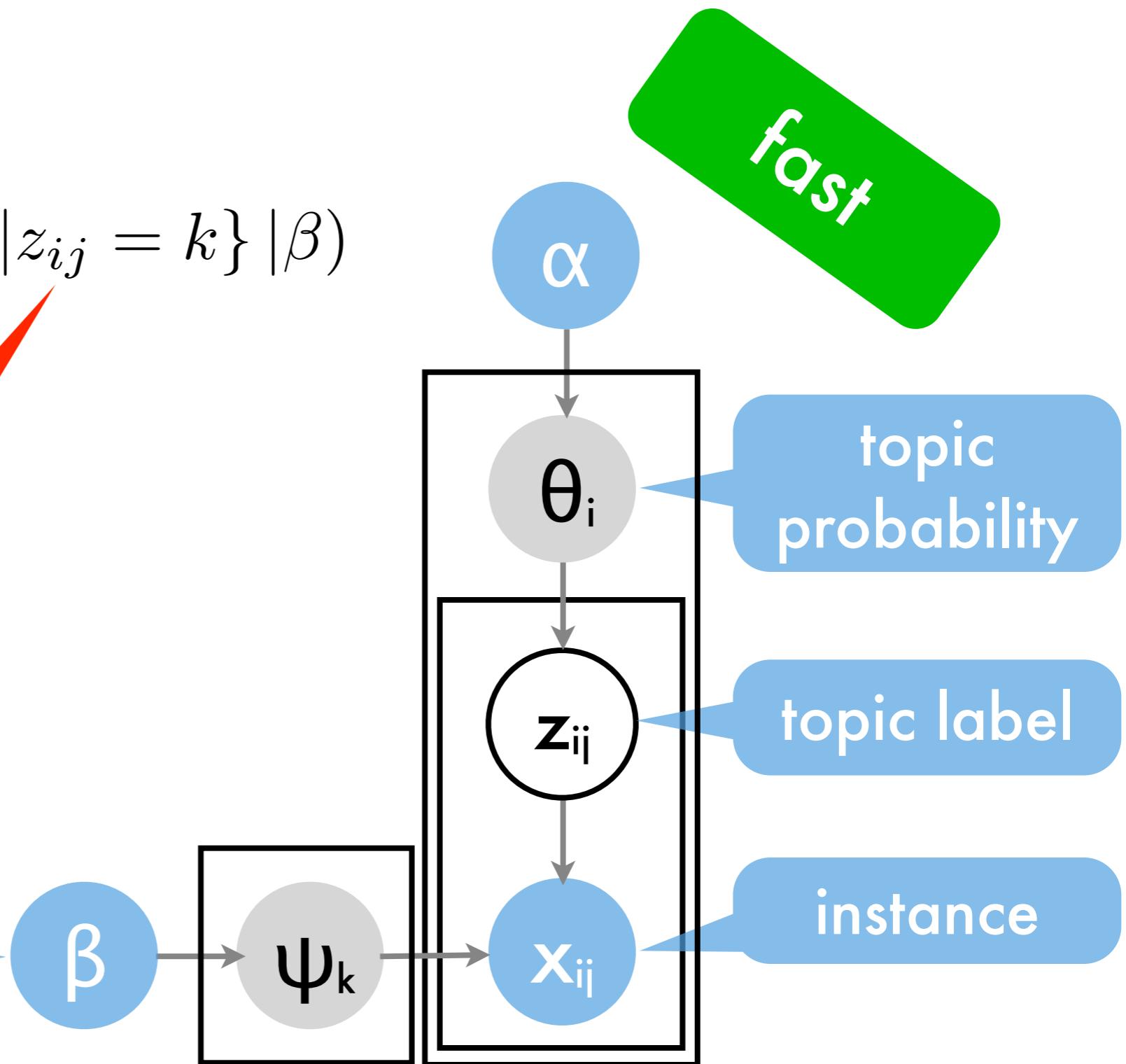


Collapsed Sampler

$$p(z, x | \alpha, \beta) = \prod_{i=1}^m p(z_i | \alpha) \prod_{k=1}^K p(\{x_{ij} | z_{ij} = k\} | \beta)$$

sample z
sequentially

language prior

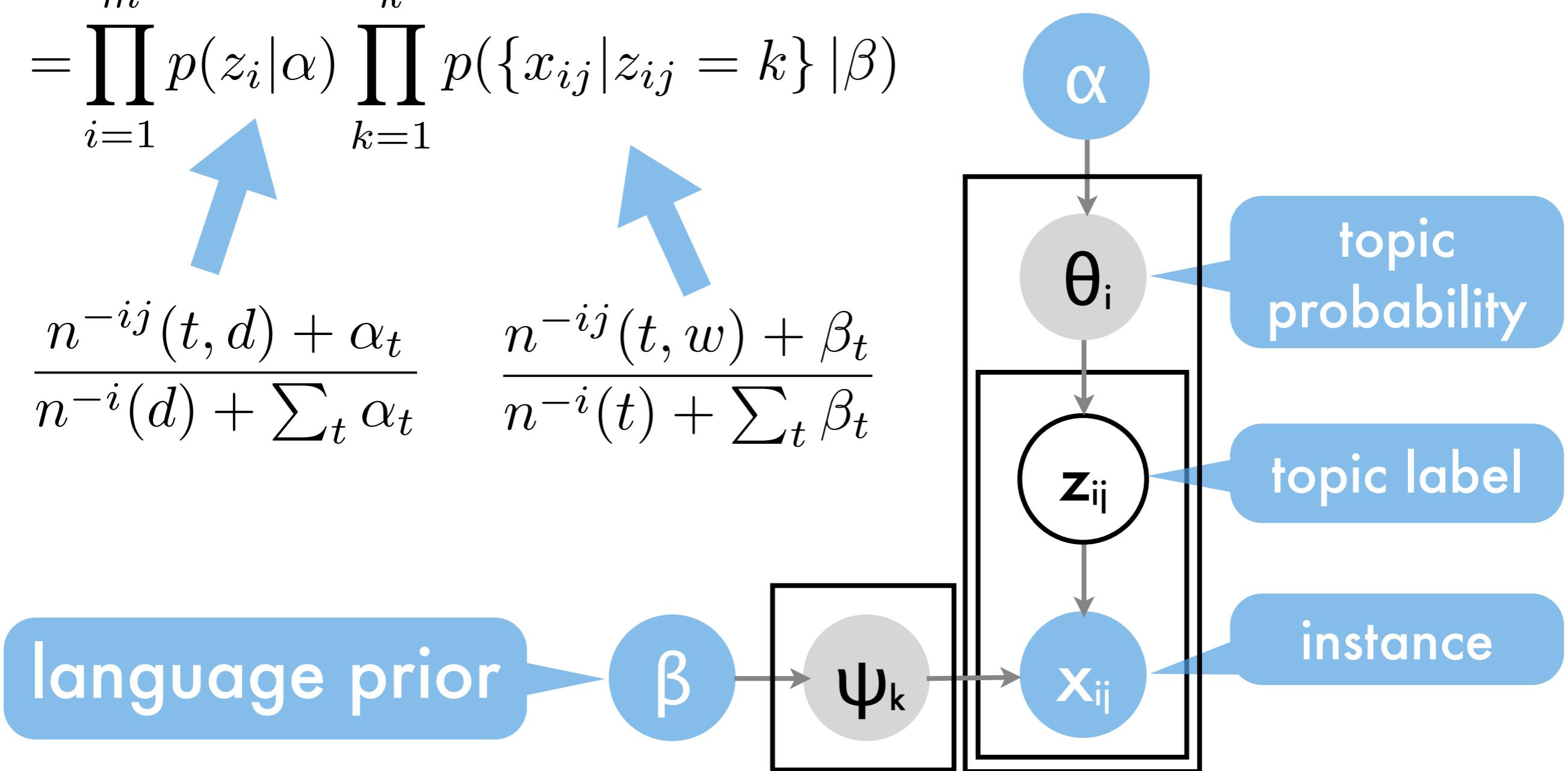


Collapsed Sampler

$$p(z, x | \alpha, \beta)$$

$$= \prod_{i=1}^m p(z_i | \alpha) \prod_{k=1}^k p(\{x_{ij} | z_{ij} = k\} | \beta)$$
$$\frac{n^{-ij}(t, d) + \alpha_t}{n^{-i}(d) + \sum_t \alpha_t}$$
$$\frac{n^{-ij}(t, w) + \beta_t}{n^{-i}(t) + \sum_t \beta_t}$$

Griffiths & Steyvers, 2005



Collapsed Sampler

$$p(z, x | \alpha, \beta)$$

$$= \prod_{i=1}^m p(z_i | \alpha) \prod_{k=1}^k p(\{x_{ij} | z_{ij} = k\} | \beta)$$

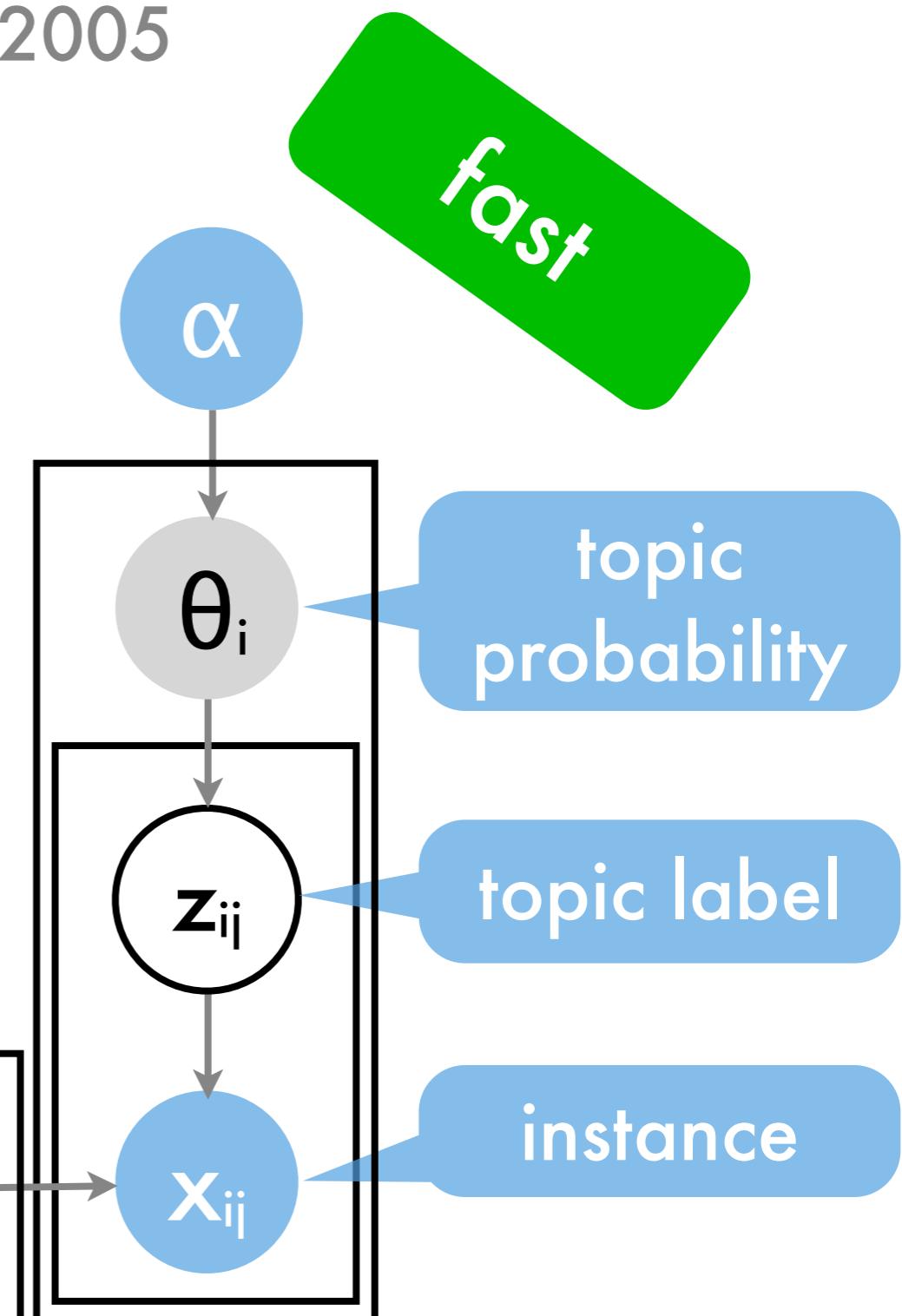
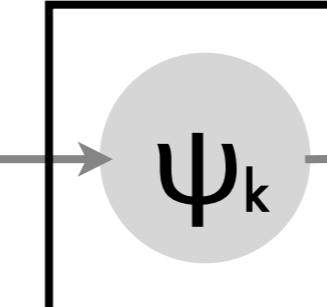
$$\frac{n^{-ij}(t, d) + \alpha_t}{n^{-i}(d) + \sum_t \alpha_t}$$

Griffiths & Steyvers, 2005

$$\frac{n^{-ij}(t, w) + \beta_t}{n^{-i}(t) + \sum_t \beta_t}$$

language prior

$$\beta$$



Sequential Algorithm (Gibbs sampler)

- For 1000 iterations do
 - For each document do
 - For each word in the document do
 - Resample topic for the word
 - Update local (document, topic) table
 - Update CPU local (word, topic) table
 - Update global (word, topic) table

Sequential Algorithm (Gibbs sampler)

- For 1000 iterations do
 - For each document do
 - For each word in the document do
 - Resample topic for the word
 - Update local (document, topic) table
 - Update CPU local (word, topic) table
 - Update global (word, topic) table



this kills parallelism

State of the art

UMass Mallet, UC Irvine, Google

- For 1000 iterations do
 - For each document do
 - For each word in the document do
 - Resample topic for the word
 - Update local (document, topic) table
 - Update CPU local (word, topic) table
 - Update global (word, topic) table

$$p(t|w_{ij}) \propto \beta_w \frac{\alpha_t}{n(t) + \bar{\beta}} + \beta_w \frac{n(t, d = i)}{n(t) + \bar{\beta}} + \frac{n(t, w = w_{ij}) [n(t, d = i) + \alpha_t]}{n(t) + \bar{\beta}}$$

State of the art

UMass Mallet, UC Irvine, Google

- For 1000 iterations do
 - For each document do
 - For each word in the document do
 - Resample topic for the word
 - Update local (document, topic) table
 - Update CPU local (word, topic) table
 - Update global (word, topic) table

$$p(t|w_{ij}) \propto \beta_w \frac{\alpha_t}{n(t) + \bar{\beta}} + \beta_w \frac{n(t, d = i)}{n(t) + \bar{\beta}} + \frac{n(t, w = w_{ij}) [n(t, d = i) + \alpha_t]}{n(t) + \bar{\beta}}$$

slow

YAHOO!

State of the art

UMass Mallet, UC Irvine, Google

- For 1000 iterations do
 - For each document do
 - For each word in the document do
 - Resample topic for the word
 - Update local (document, topic) table
 - Update CPU local (word, topic) table
 - Update global (word, topic) table

changes rapidly

$$p(t|w_{ij}) \propto \beta_w \frac{\alpha_t}{n(t) + \bar{\beta}} + \beta_w \frac{n(t, d = i)}{n(t) + \bar{\beta}} + \frac{n(t, w = w_{ij}) [n(t, d = i) + \alpha_t]}{n(t) + \bar{\beta}}$$

slow

State of the art

UMass Mallet, UC Irvine, Google

- For 1000 iterations do
 - For each document do
 - For each word in the document do
 - Resample topic for the word
 - Update local (document, topic) table
 - Update CPU local (word, topic) table
 - Update global (word, topic) table

changes rapidly

$$p(t|w_{ij}) \propto \beta_w \frac{\alpha_t}{n(t) + \bar{\beta}} + \beta_w \frac{n(t, d = i)}{n(t) + \bar{\beta}} + \frac{n(t, w = w_{ij}) [n(t, d = i) + \alpha_t]}{n(t) + \bar{\beta}}$$

slow

moderately fast

YAHOO!

State of the art

UMass Mallet, UC Irvine, Google

- For 1000 iterations do
 - For each document do
 - For each word in the document do
 - Resample topic for the word
 - Update local (document, topic) table
 - Update CPU local (word, topic) table
 - Update global (word, topic) table

table out
of sync

memory
inefficient

blocking

network
bound

changes rapidly

$$p(t|w_{ij}) \propto \beta_w \frac{\alpha_t}{n(t) + \bar{\beta}} + \beta_w \frac{n(t, d = i)}{n(t) + \bar{\beta}} + \frac{n(t, w = w_{ij}) [n(t, d = i) + \alpha_t]}{n(t) + \bar{\beta}}$$

slow

moderately fast

YAHOO!

Our Approach

- For 1000 iterations do (independently per computer)
 - For each thread/core do
 - For each document do
 - For each word in the document do
 - Resample topic for the word
 - Update local (document, topic) table
 - Generate computer local (word, topic) message
 - In parallel update local (word, topic) table
 - In parallel update global (word, topic) table

Our Approach

- For 1000 iterations do (independently per computer)
 - For each thread/core do
 - For each document do
 - For each word in the document do
 - Resample topic for the word
 - Update local (document, topic) table
 - Generate computer local (word, topic) message
 - In parallel update local (word, topic) table
 - In parallel update global (word, topic) table

network
bound

concurrent
cpu hdd net

YAHOO!

Our Approach

- For 1000 iterations do (independently per computer)
 - For each thread/core do
 - For each document do
 - For each word in the document do
 - Resample topic for the word
 - Update local (document, topic) table
 - Generate computer local (word, topic) message
 - In parallel update local (word, topic) table
 - In parallel update global (word, topic) table

network
bound

memory
inefficient

concurrent
cpu hdd net

minimal
view

Our Approach

- For 1000 iterations do (independently per computer)
 - For each thread/core do
 - For each document do
 - For each word in the document do
 - Resample topic for the word
 - Update local (document, topic) table
 - Generate computer local (word, topic) message
 - In parallel update local (word, topic) table
 - In parallel update global (word, topic) table

network
bound

memory
inefficient

table out
of sync

concurrent
cpu hdd net

minimal
view

continuous
sync

Our Approach

- For 1000 iterations do (independently per computer)
 - For each thread/core do
 - For each document do
 - For each word in the document do
 - Resample topic for the word
 - Update local (document, topic) table
 - Generate computer local (word, topic) message
 - In parallel update local (word, topic) table
 - In parallel update global (word, topic) table

network
bound

memory
inefficient

table out
of sync

blocking

concurrent
cpu hdd net

minimal
view

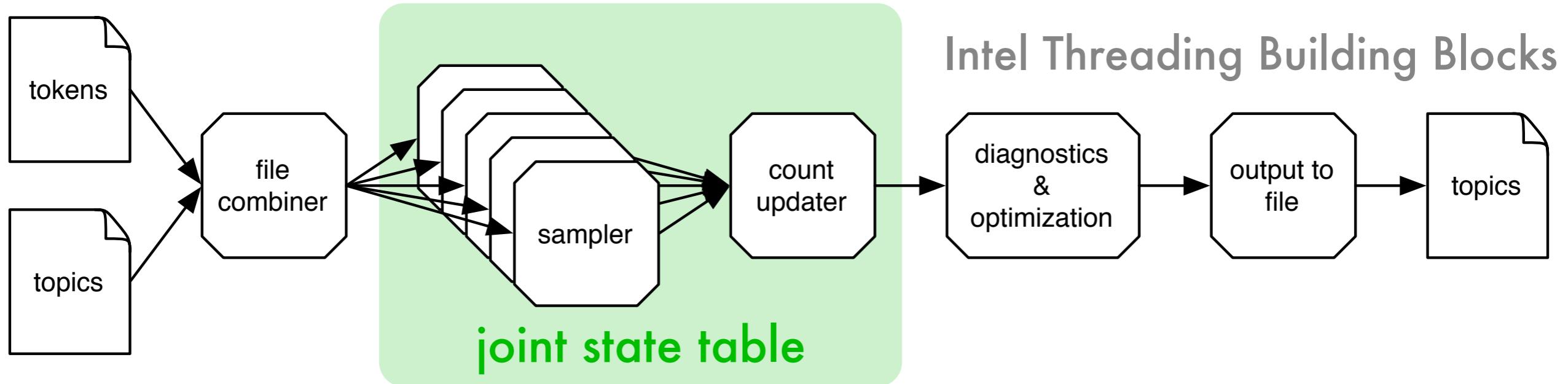
continuous
sync

barrier
free

FOO!

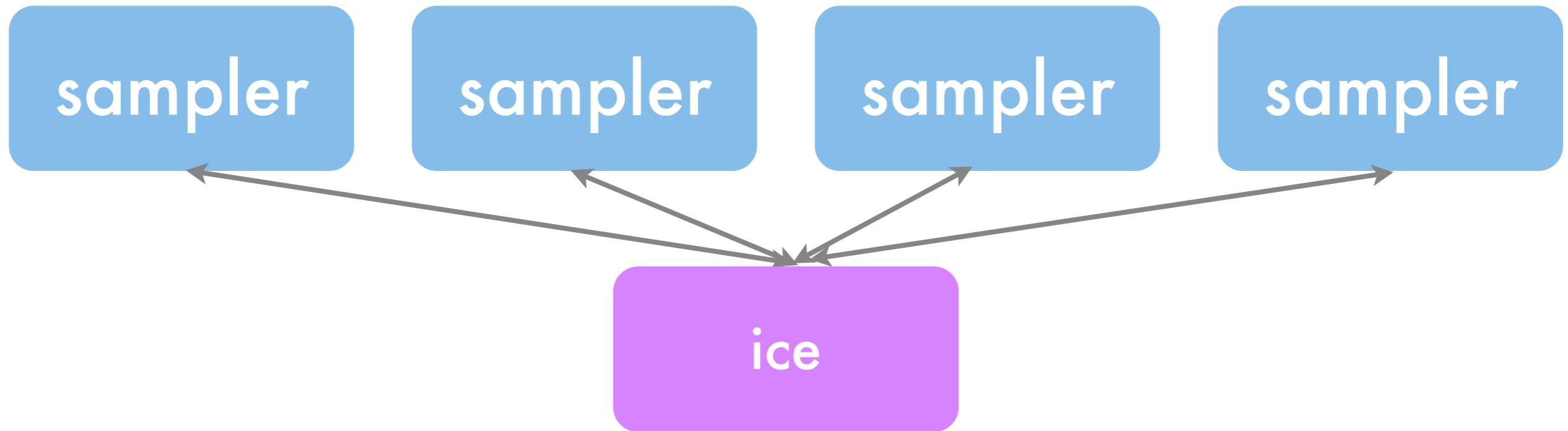
Architecture details

Multicore Architecture



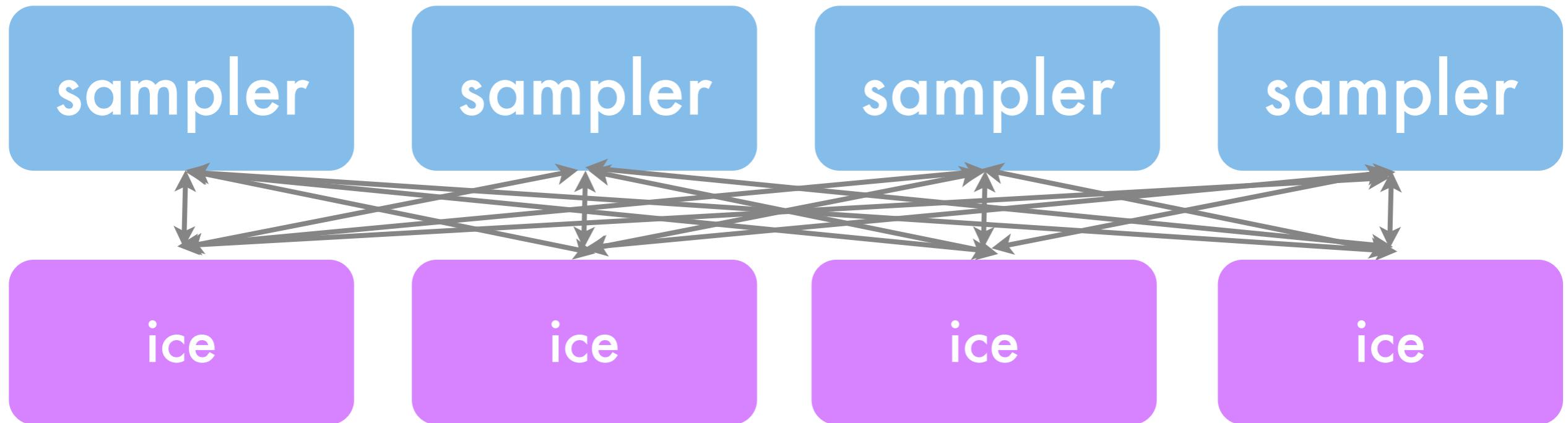
- Decouple multithreaded sampling and updating
(almost) avoids stalling for locks in the sampler
- Joint state table
 - much less memory required
 - samplers synchronized (10 docs vs. millions delay)
- Hyperparameter update via stochastic gradient descent
- No need to keep documents in memory (streaming)

Cluster Architecture



- Distributed (key,value) storage via memcached
- Background asynchronous synchronization
 - single word at a time to avoid deadlocks
 - no need to have joint dictionary
 - uses disk, network, cpu simultaneously

Cluster Architecture



- Distributed (key,value) storage via ICE
- Background asynchronous synchronization
 - single word at a time to avoid deadlocks
 - no need to have joint dictionary
 - uses disk, network, cpu simultaneously

Making it work

- **Startup**
 - Randomly initialize topics on each node
(read from disk if already assigned - hotstart)
 - Sequential Monte Carlo for startup **much faster**
 - Aggregate changes on the fly
- **Failover**
 - State constantly being written to disk
(worst case we lose 1 iteration out of 1000)
 - Restart via standard startup routine
 - Achilles heel: need to restart from checkpoint if even a single machine dies.

Easily extensible

- Better language model (topical n-grams)
can process millions of users (vs 1000s)
- Conditioning on side information (upstream)
estimate topic based on authorship, source,
joint user model ...
- Conditioning on dictionaries (downstream)
integrate topics between different languages
- Time dependent sampler for user model
approximate inference per episode

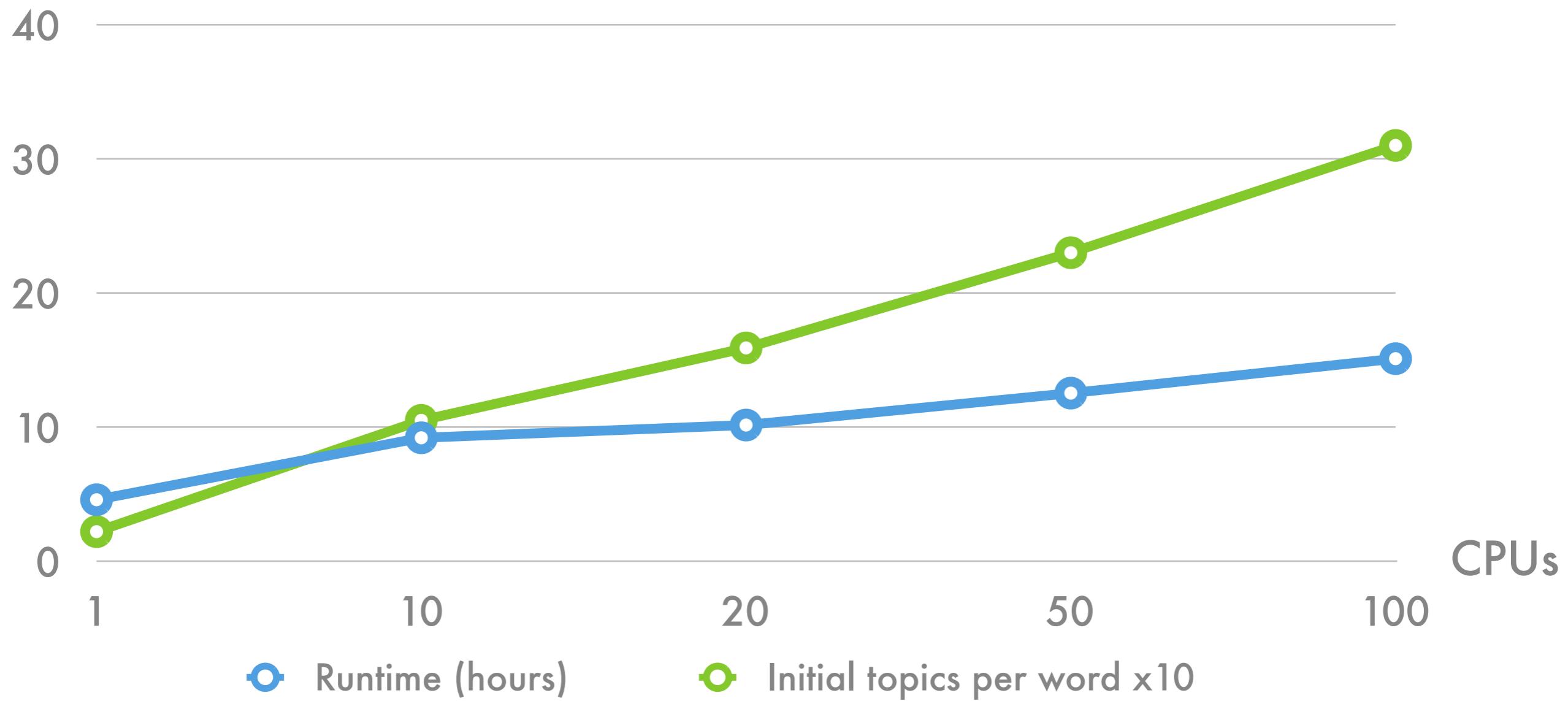
	Google LDA	Mallet	Irvine'08	Irvine'09	Yahoo LDA
Multicore	no	yes	yes	yes	yes
Cluster	MPI	no	MPI	point 2 point	memcached
State table	dictionary split	separate sparse	separate	separate	joint sparse
Schedule	synchronous exact	synchronous exact	synchronous exact	asynchronous approximate messages	asynchronous exact

Speed

- **1M documents per day** on 1 computer
(1000 topics per doc, 1000 words per doc)
- **350k documents per day** per node
(context switches & memcached & stray reducers)
- 8 Million docs (Pubmed)
(sampler does not burn in well - too short doc)
 - Irvine: **128 machines, 10 hours**
 - Yahoo: **1 machine, 11 days**
 - Yahoo: **20 machines, 9 hours**
- 20 Million docs (Yahoo! News Articles)
 - Yahoo: 100 machines, 12 hours

Scalability

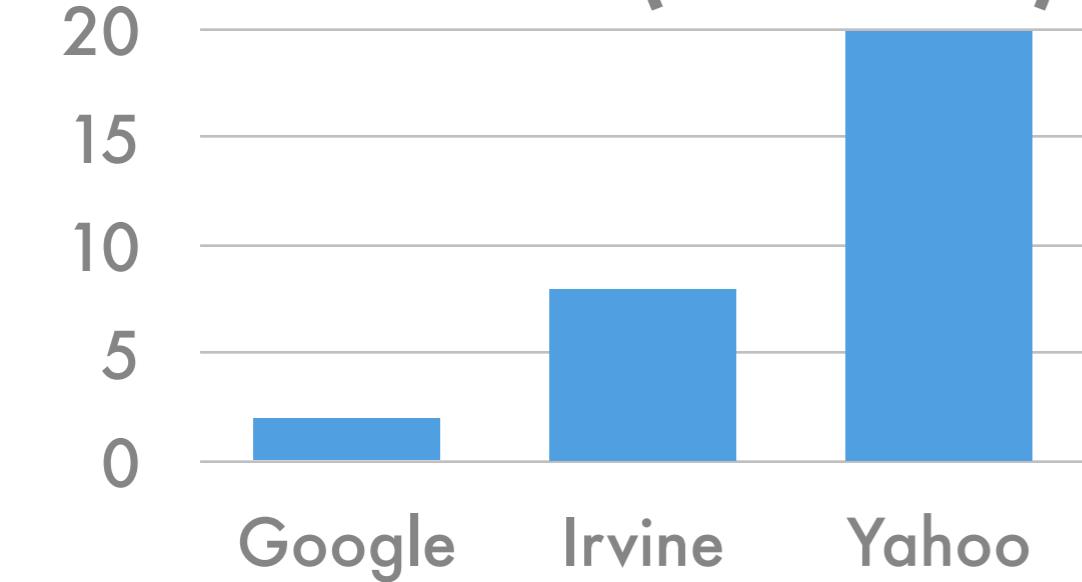
200k documents/computer



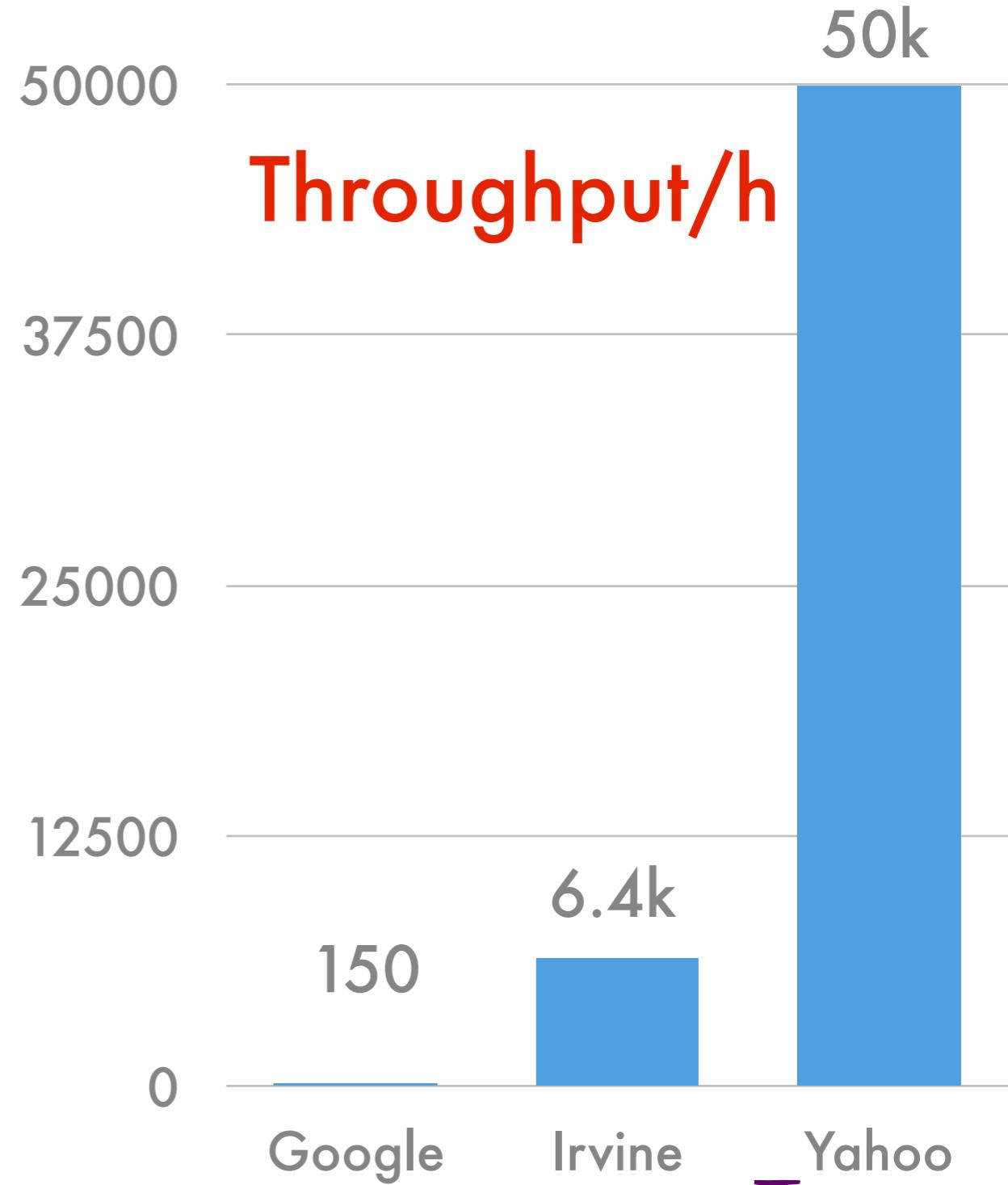
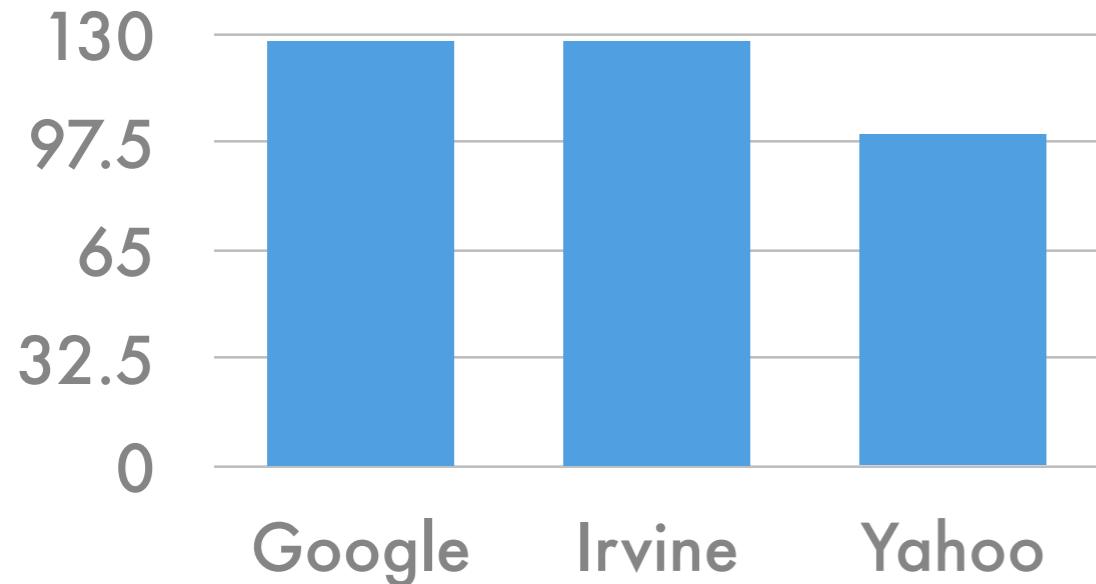
Likelihood even improves with parallelism!
-3.295 (1 node) -3.288 (10 nodes) -3.287 (20 nodes)

The Competition

Dataset size (millions)



Cluster size

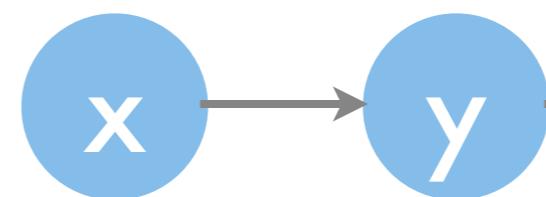
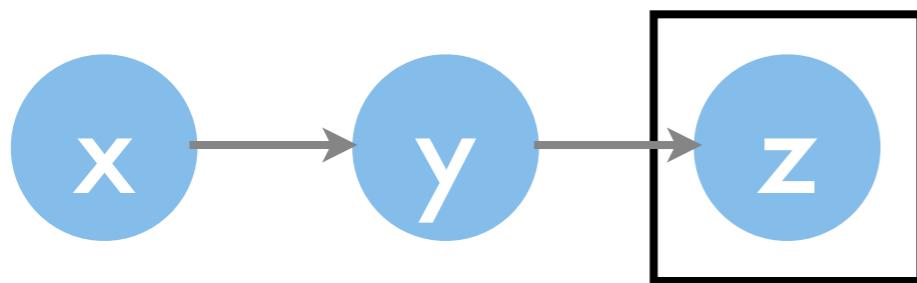


YAHOO!

Design Principles

Variable Replication

- Global shared variable



computer

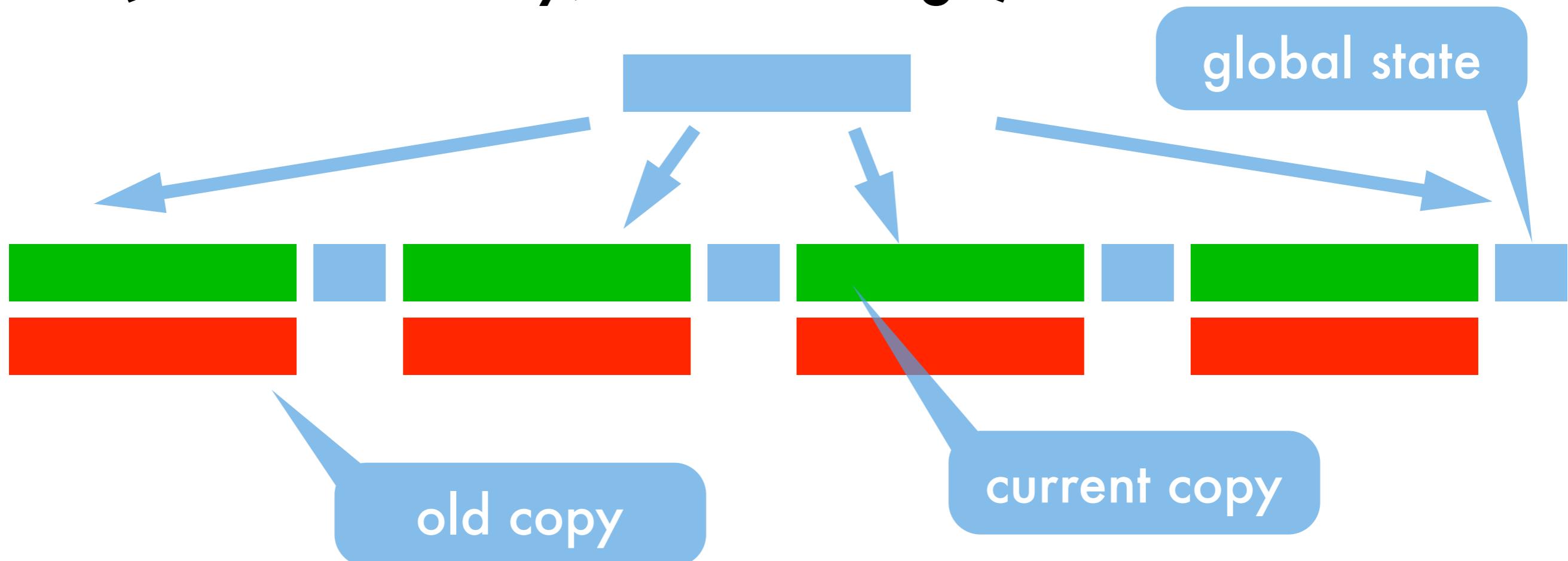
synchronize

local copy

- Make local copy
 - Distributed (key,value) storage table for global copy
 - Do all bookkeeping locally (store old versions)
 - Sync local copies asynchronously using message passing (no global locks are needed)
- **This is an approximation!**

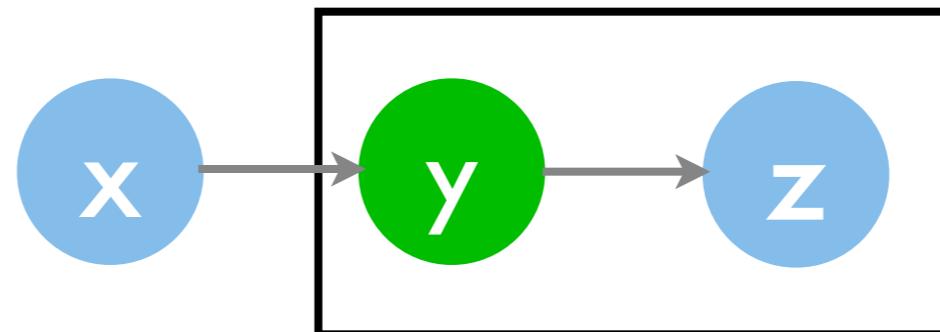
Asymmetric Message Passing

- Large global shared state space
(essentially as large as the memory in computer)
- Distribute global copy over several machines
(distributed key,value storage)

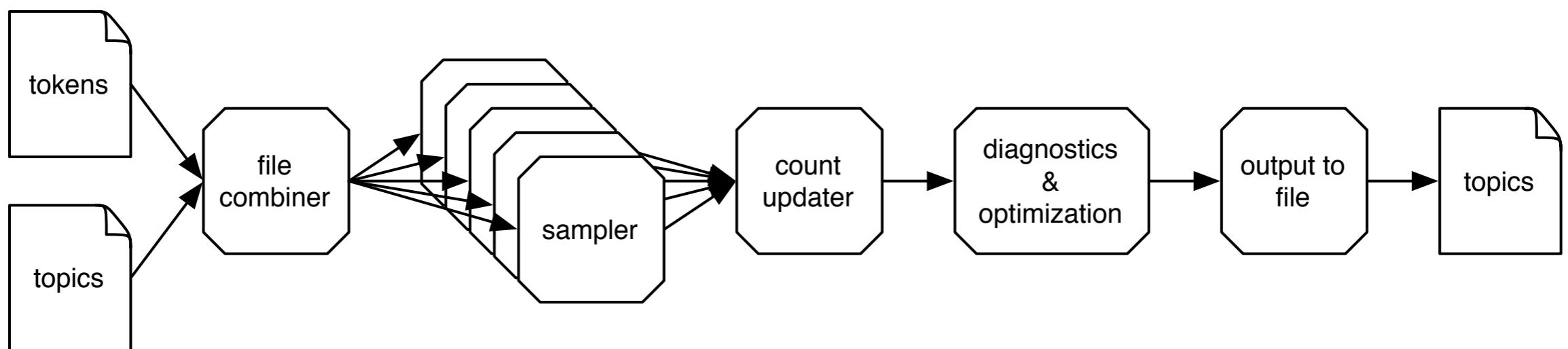


Out of core storage

- Very large state space



- Gibbs sampling requires us to traverse the data sequentially many times (think 1000x)
- Stream local data from disk and update coupling variable each time local data is accessed
- This is exact

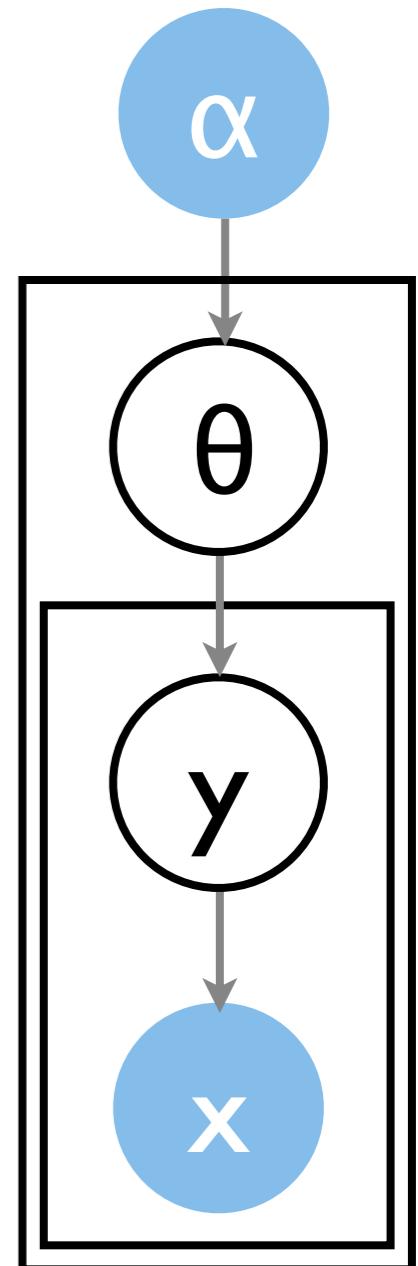


Part 6 - Advanced Modeling

Advances in Representation

Extensions to topic models

- Prior over document topic vector
 - Usually as Dirichlet distribution
 - Use correlation between topics (CTM)
 - Hierarchical structure over topics
- Document structure
 - Bag of words
 - n-grams (Li & McCallum)
 - Simplicial Mixture (Girolami & Kaban)
- Side information
 - Upstream conditioning (Mimno & McCallum)
 - Downstream conditioning (Petterson et al.)
 - Supervised LDA (Blei and McAulliffe 2007; Lacoste, Sha and Jordan 2008; Zhu, Ahmed and Xing 2009)



Correlated topic models

- Dirichlet distribution
 - Can only model which topics are hot
 - Does not model relationships between topics

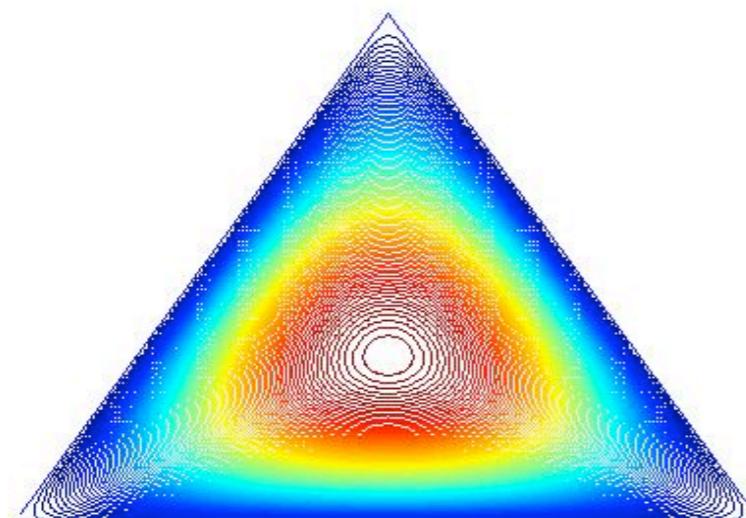
Correlated topic models

- Dirichlet distribution
 - Can only model which topics are hot
 - Does not model relationships between topics
- Key idea
 - We expect to see documents about sports and health but not about sports and politics
 - Uses a logistic normal distribution as a prior
- Conjugacy is no longer maintained
- Inference is harder than in LDA

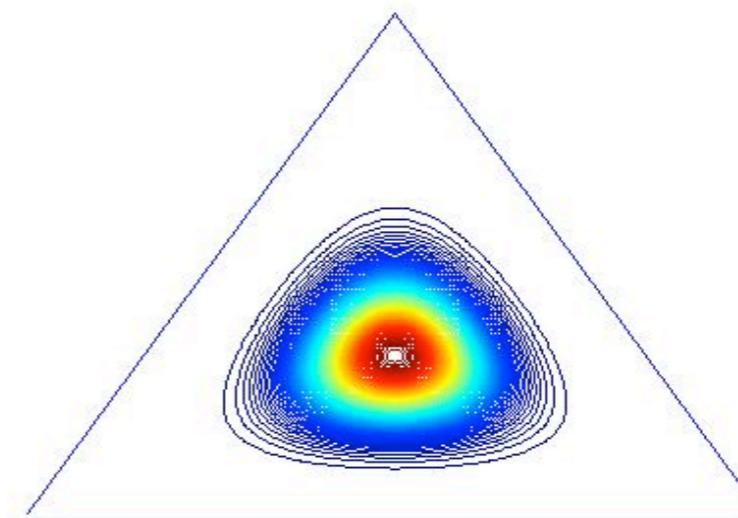
Blei & Lafferty 2005; Ahmed & Xing 2007

Dirichlet prior on topics

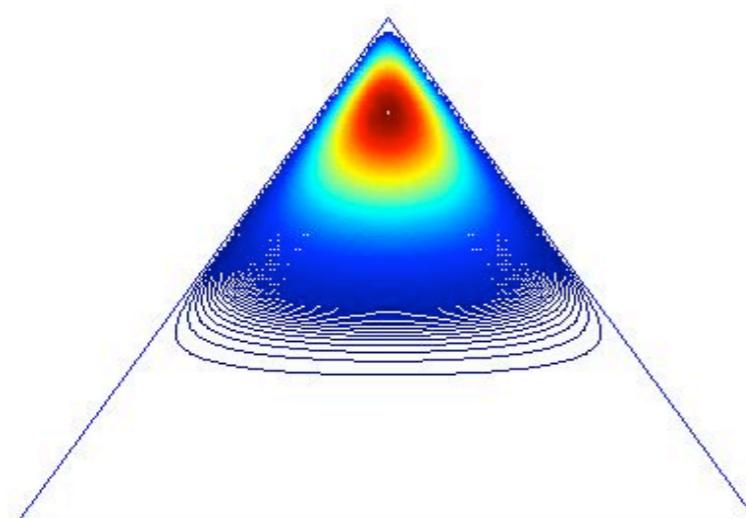
Alpha =[2.00 2.00 2.00]



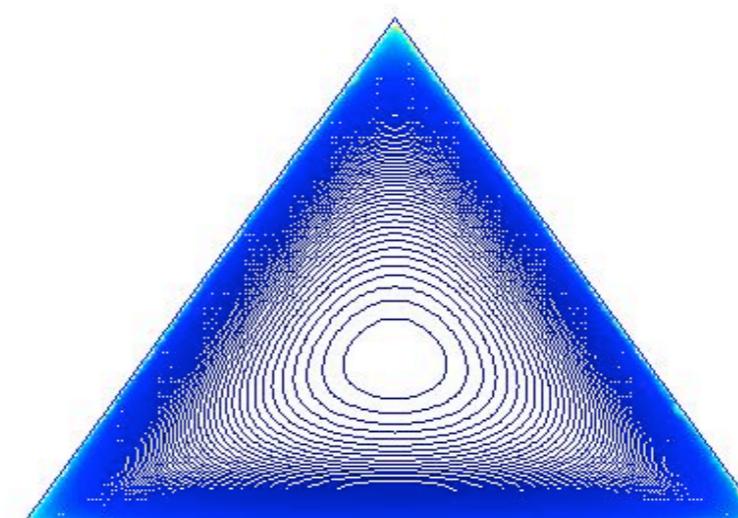
Alpha =[10.00 10.00 10.00]



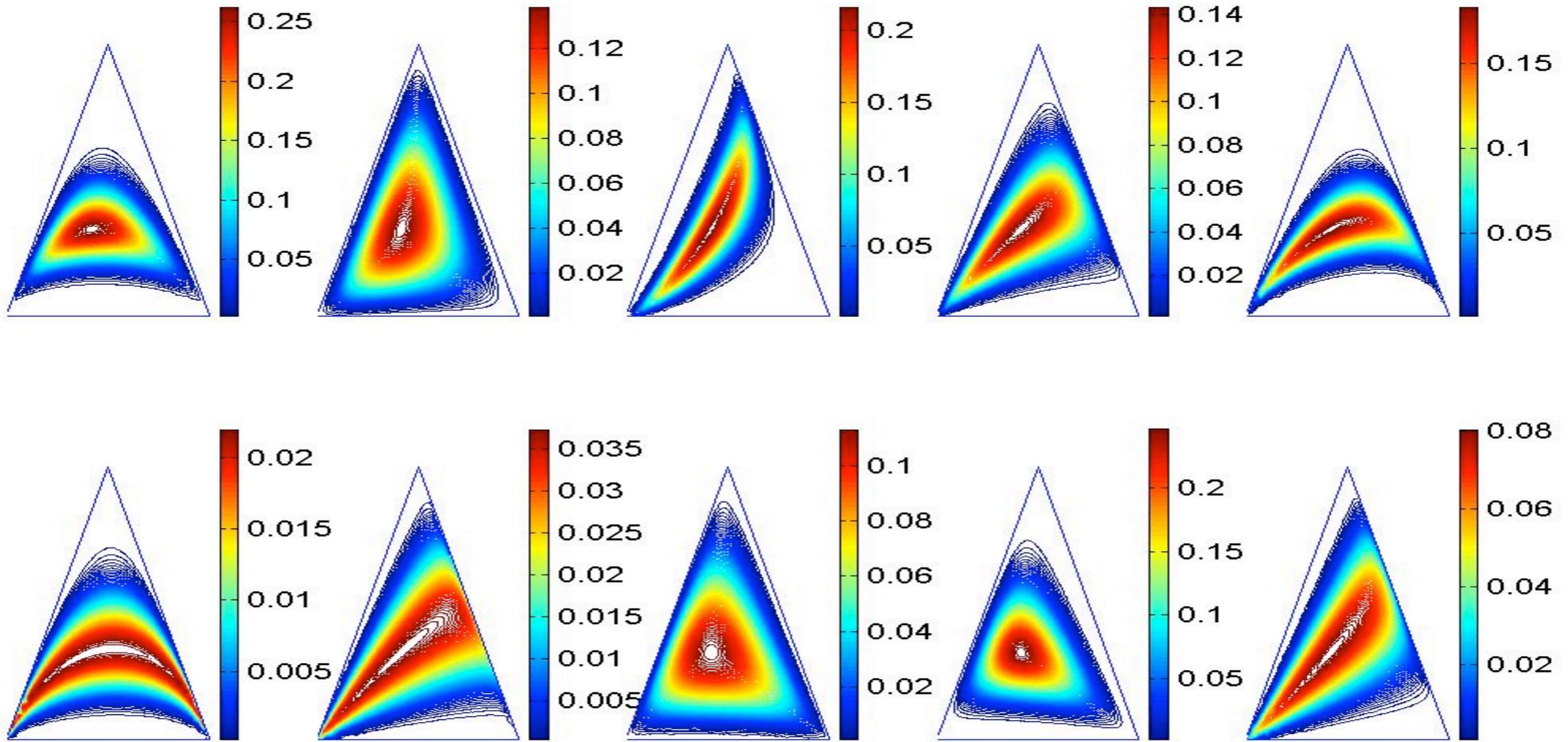
Alpha =[2.00 10.00 2.00]



Alpha =[0.90 0.90 0.90]

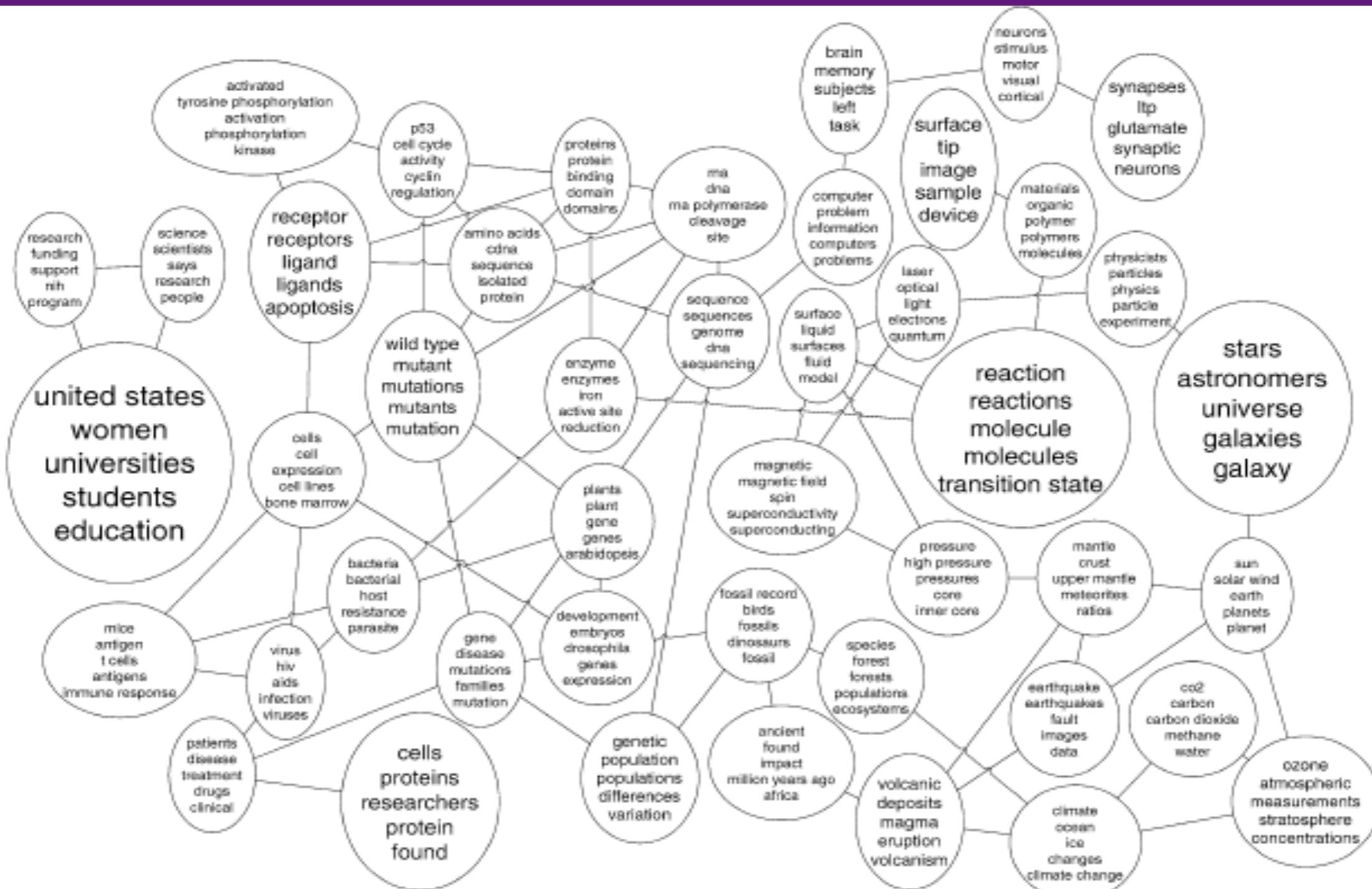


Log-normal prior on topics



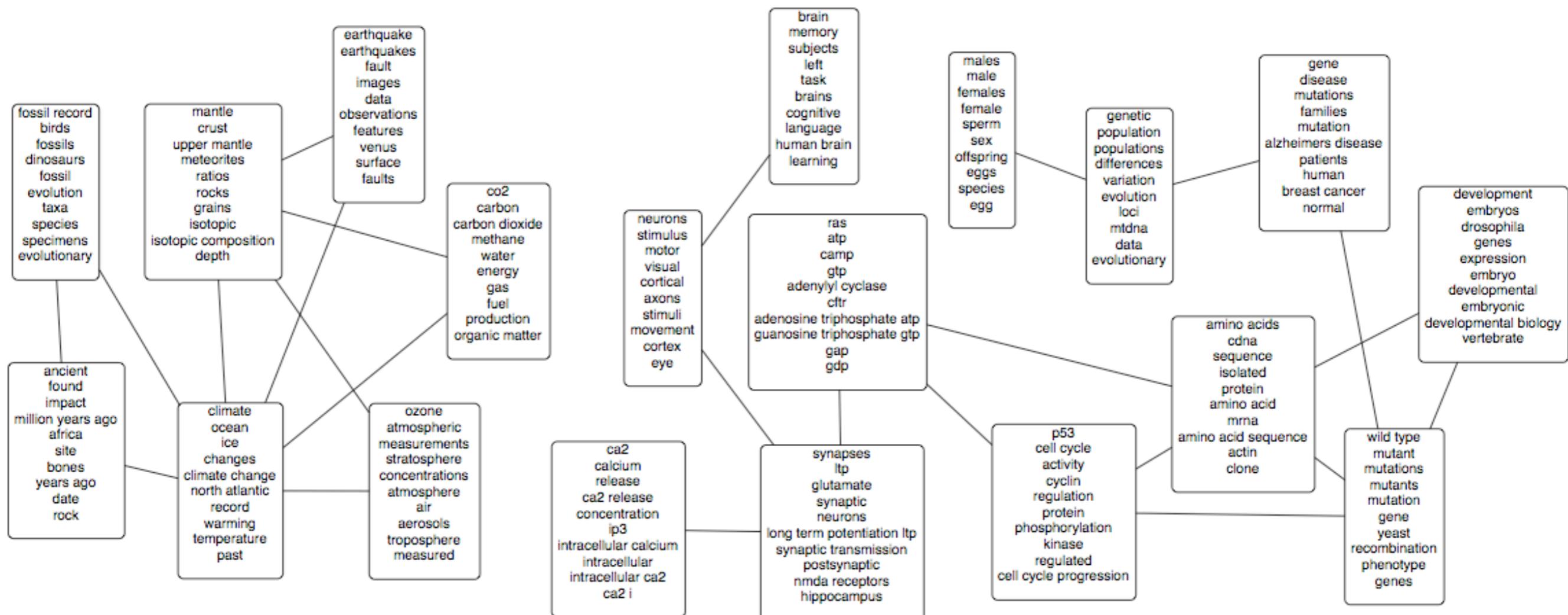
$$\theta = e^{\eta - g(\eta)} \text{ with } \eta \sim \mathcal{N}(\mu, \Sigma)$$

Correlated topics



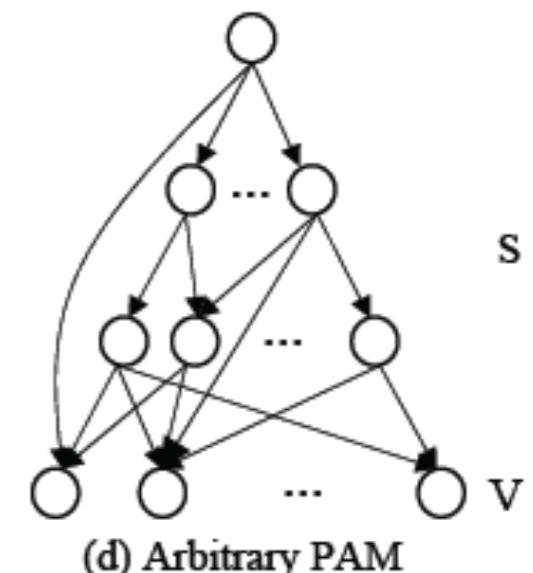
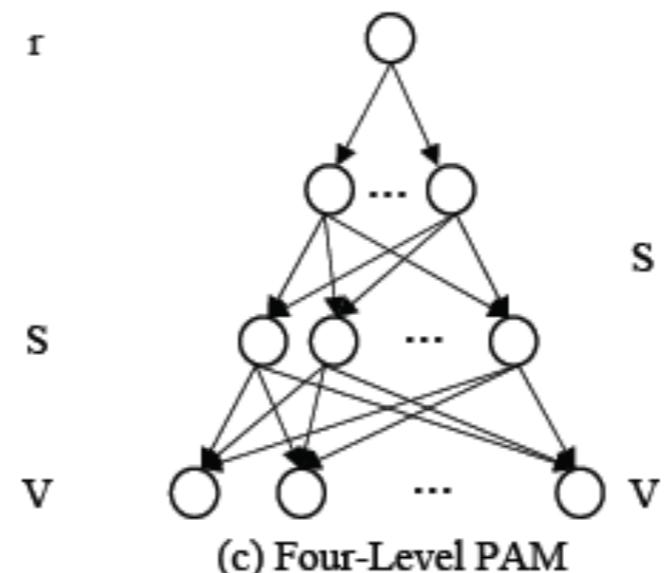
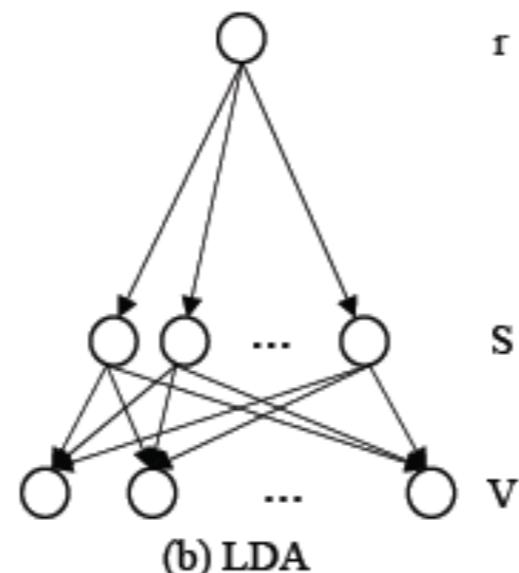
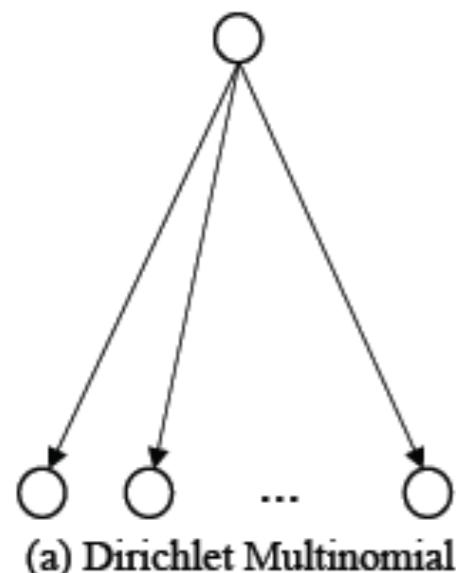
Blei and Lafferty 2005

Correlated topics

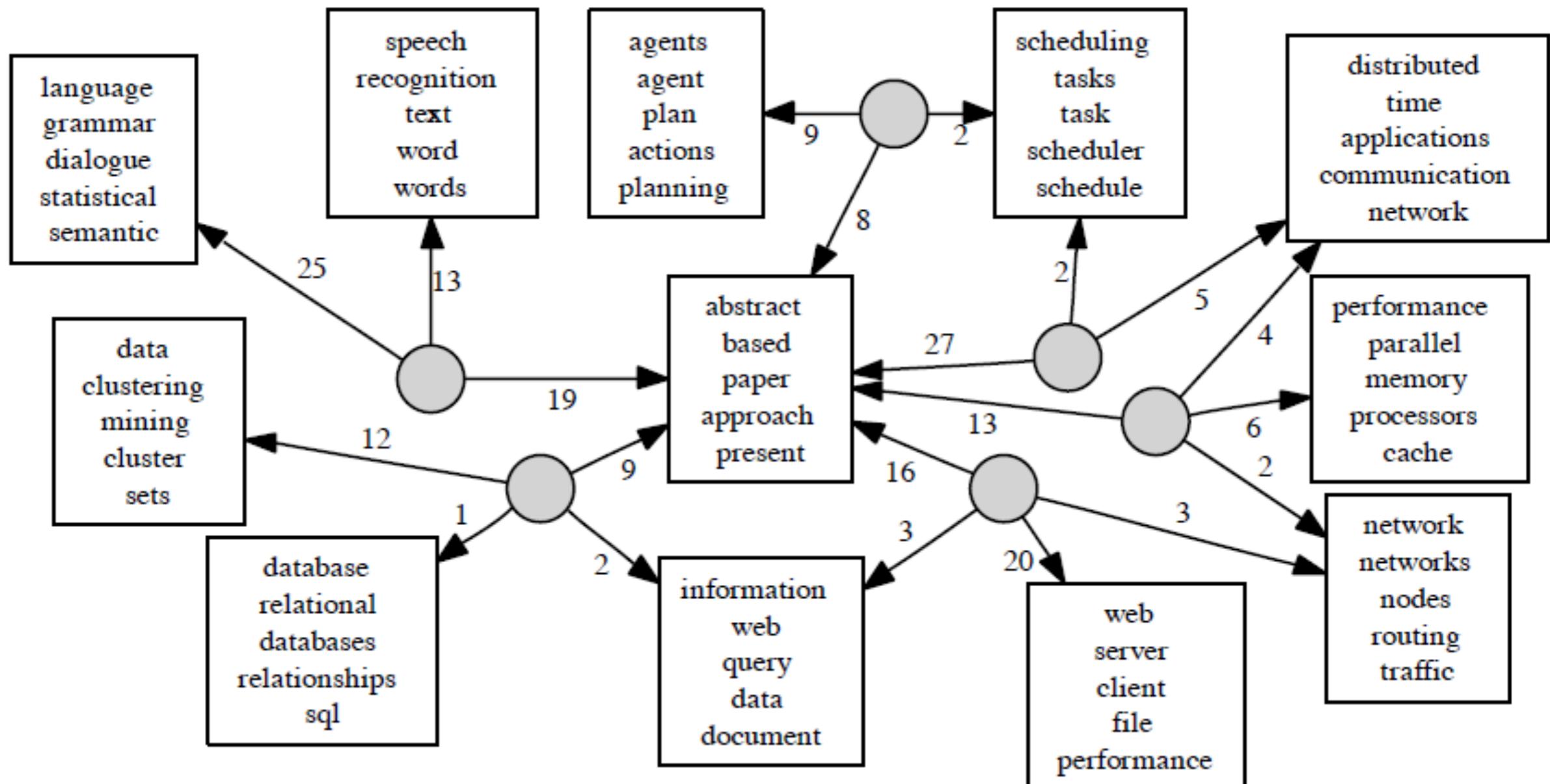


Pachinko Allocation

- Model the prior as a Directed Acyclic Graph
- Each document is modeled as multiple paths
- To sample a word, first select a path and then sample a word from the final topic
- The topics reside on the leaves of the tree



Pachinko Allocation

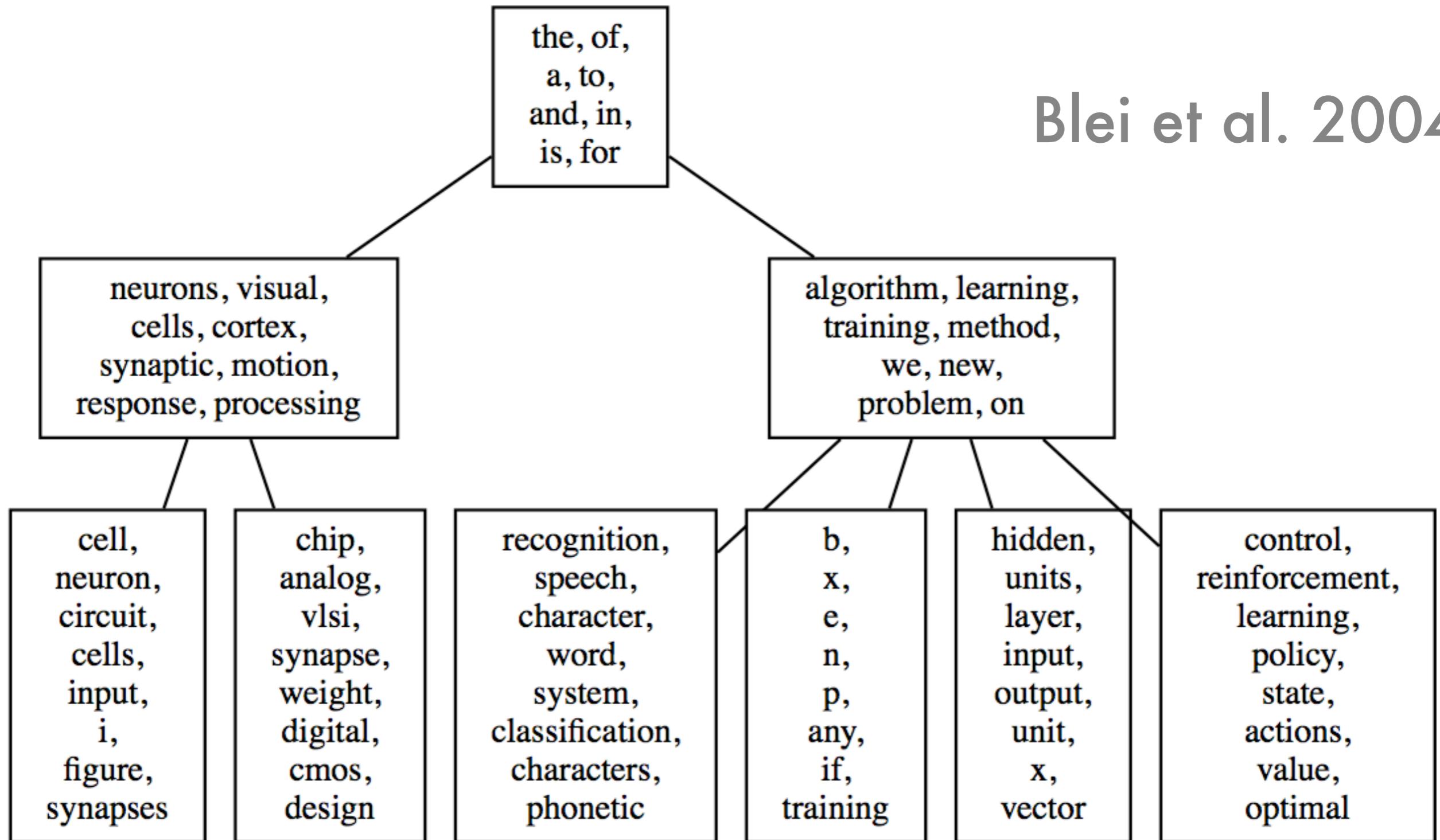


Topic Hierarchies

- Topics can appear **anywhere** in the tree
- Each document is modeled as
 - Single path over the tree (Blei et al., 2004)
 - Multiple paths over the tree (Mimno et al., 2007)

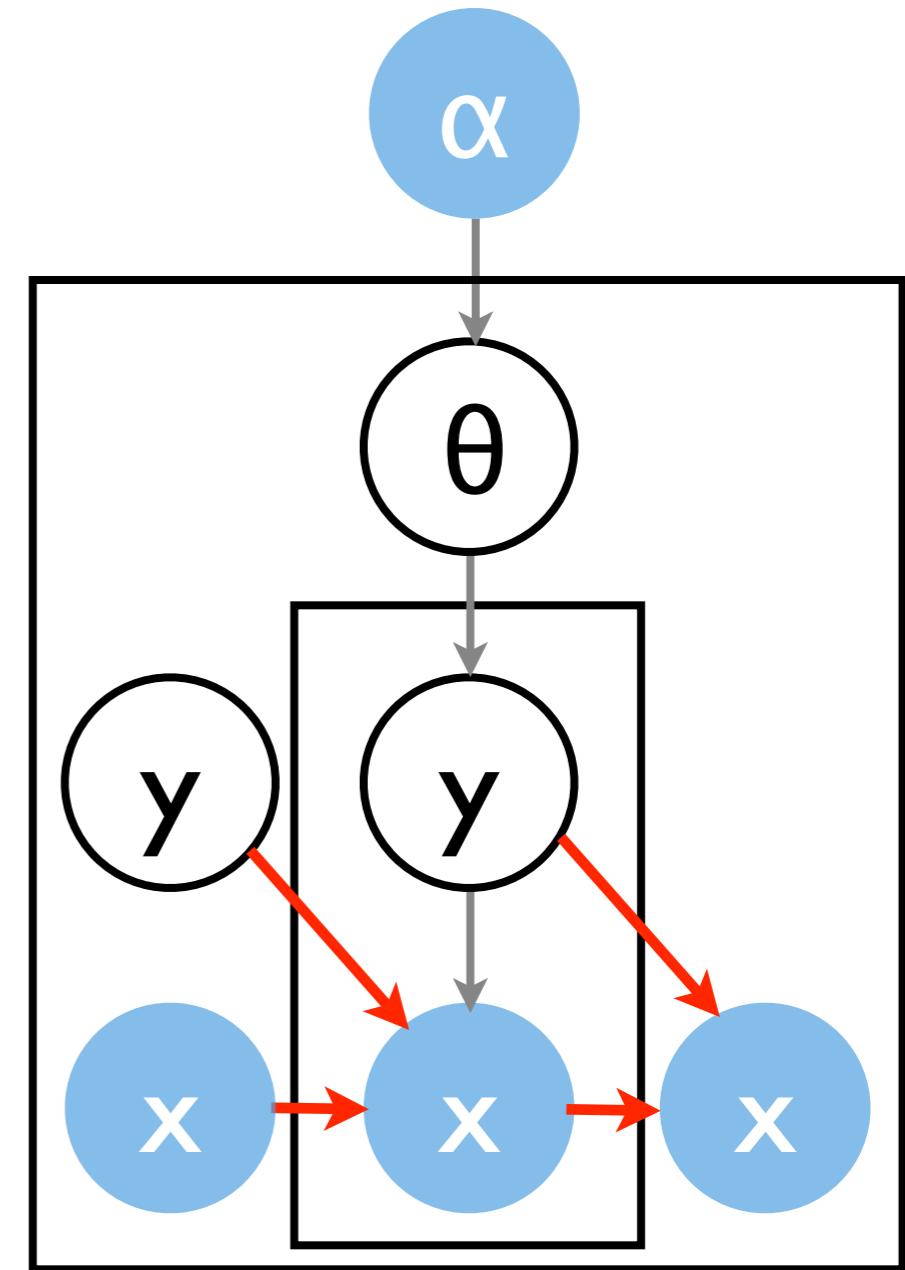
Topic Hierarchies

Blei et al. 2004



Topical n-grams

- Documents as bag of words
- Exploit sequential structure
- N-gram models
 - Capture longer phrases
 - Switch variables to determine segments
 - Dynamic programming needed



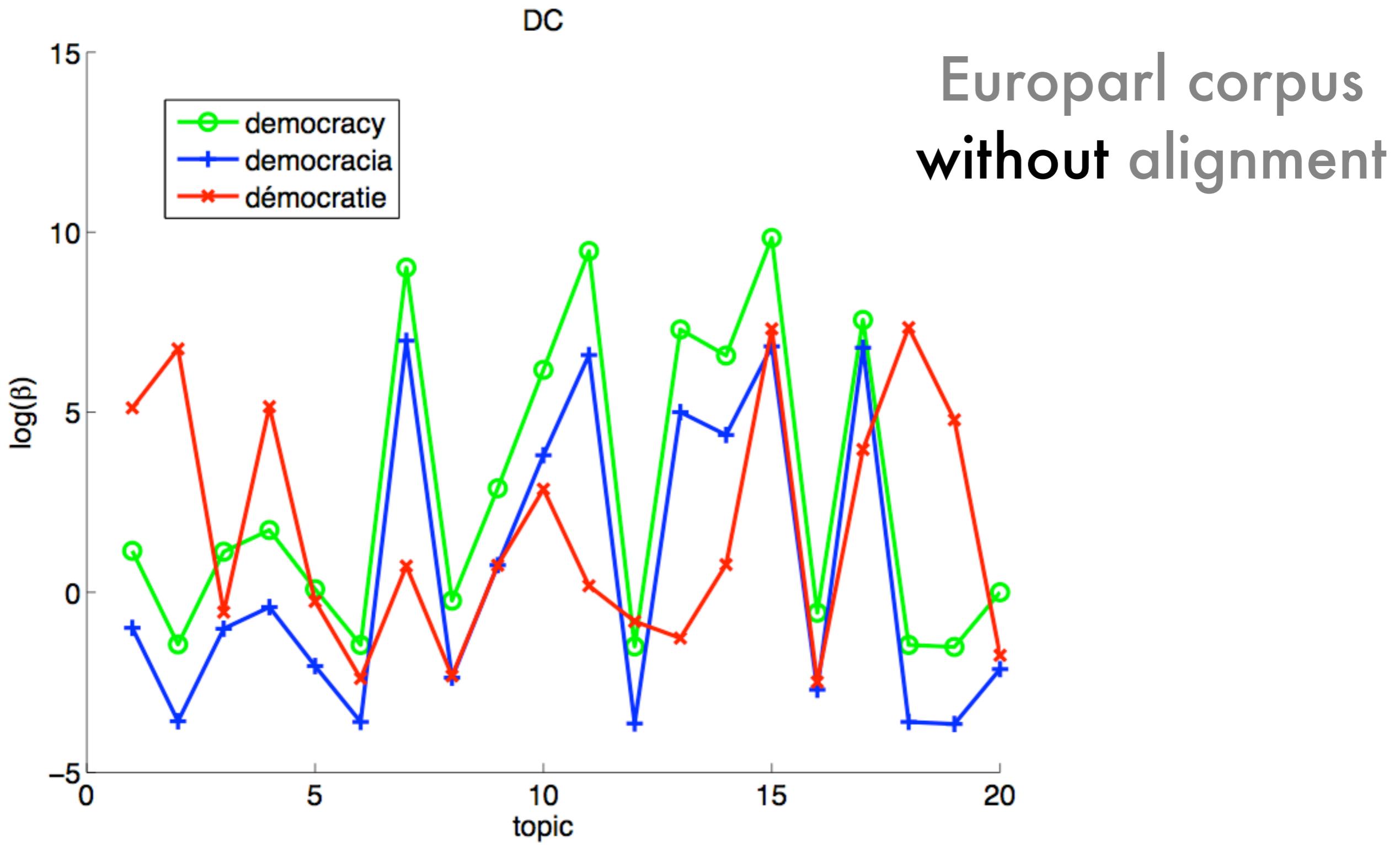
Topic n-grams

Speech Recognition			Support Vector Machines		
LDA	n-gram (2+)	n-gram (1)	LDA	n-gram (2+)	n-gram (1)
recognition	speech recognition	speech	kernel	support vectors	kernel
system	training data	word	linear	test error	training
word	neural network	training	vector	support vector machines	support
face	error rates	system	support	training error	margin
context	neural net	recognition	set	feature space	svm
character	hidden markov model	hmm	nonlinear	training examples	solution
hmm	feature vectors	speaker	data	decision function	kernels
based	continuous speech	performance	algorithm	cost functions	regularization
frame	training procedure	phoneme	space	test inputs	adaboost
segmentation	continuous speech recognition	acoustic	pca	kkt conditions	test
training	gamma filter	words	function	leave-one-out procedure	data
characters	hidden control	context	problem	soft margin	generalization
set	speech production	systems	margin	bayesian transduction	examples
probabilities	neural nets	frame	vectors	training patterns	cost
features	input representation	trained	solution	training points	convex
faces	output layers	sequence	training	maximum margin	algorithm
words	training algorithm	phonetic	svm	strictly convex	working
frames	test set	speakers	kernels	regularization operators	feature
database	speech frames	mlp	matrix	base classifiers	sv
mlp	speaker dependent	hybrid	machines	convex optimization	functions

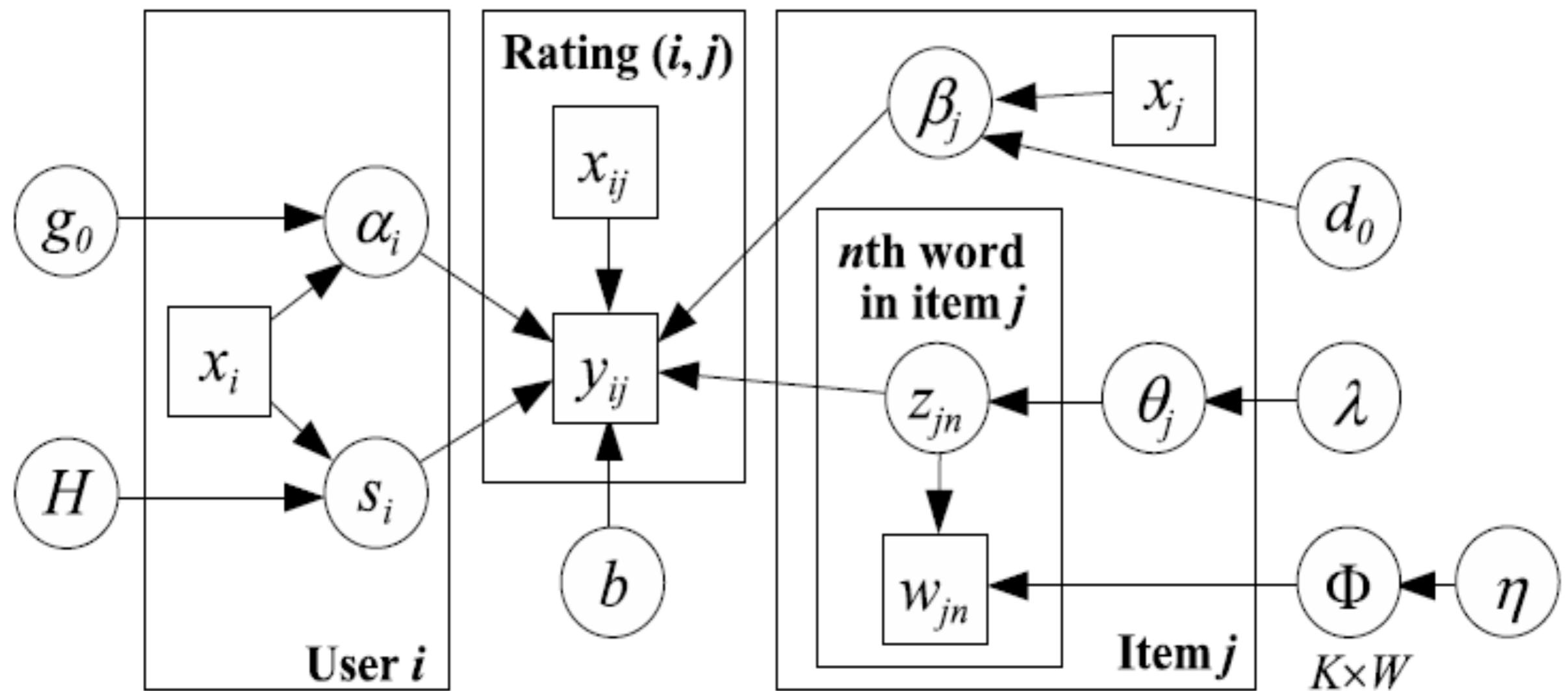
Side information

- Upstream conditioning (Mimno et al., 2008)
 - Document features are informative for topics
 - Estimate topic distribution e.g. based on authors, links, timestamp
- Downstream conditioning (Petterson et al., 2010)
 - Word features are informative on topics
 - Estimate topic distribution for words e.g. based on dictionary, lexical similarity, distributional similarity
- Class labels (Blei and McAulliffe 2007; Lacoste, Sha and Jordan 2008; Zhu, Ahmed and Xing 2009)
 - Joint model of unlabeled data and labels
 - Joint likelihood - **semisupervised learning done right!**

Downstream conditioning

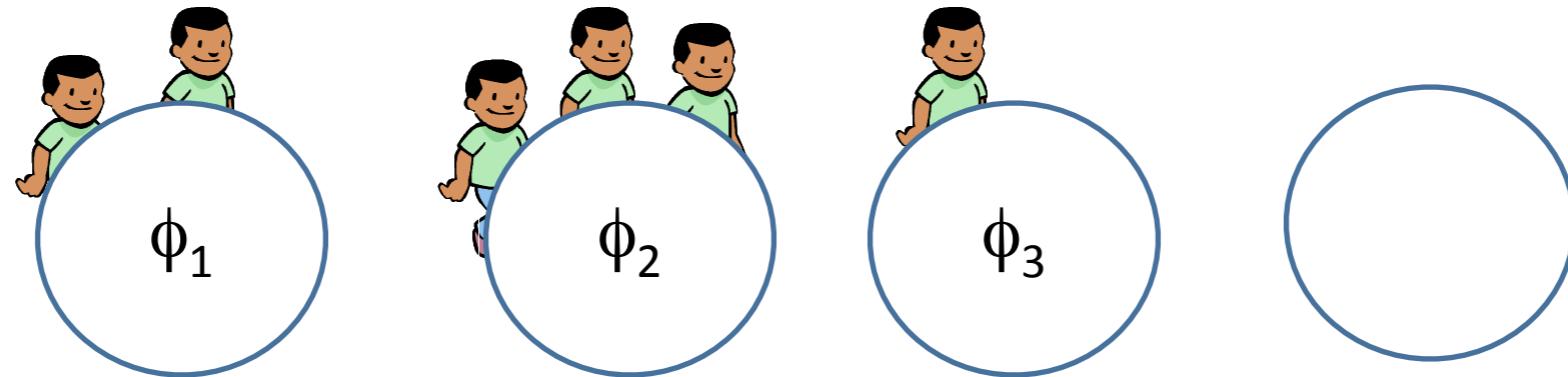


Recommender Systems



Agarwal & Chen, 2010

Chinese Restaurant Process



Problem

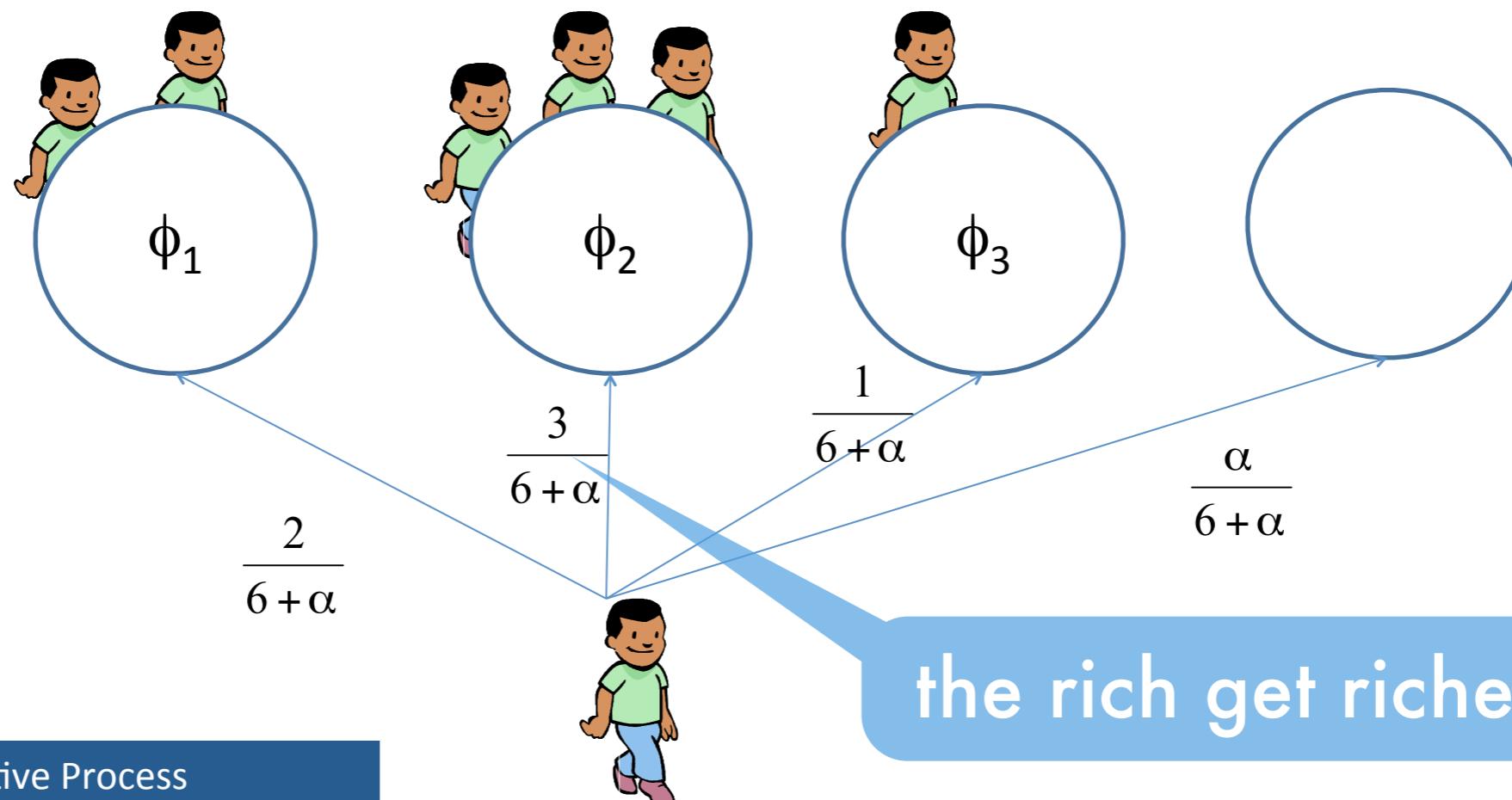
- How many clusters should we pick?
- How about a prior for infinitely many clusters?
- Finite model

$$p(y|Y, \alpha) = \frac{n(y) + \alpha_y}{n + \sum_{y'} \alpha_{y'}}$$

- Infinite model
Assume that the total smoother weight is constant

$$p(y|Y, \alpha) = \frac{n(y)}{n + \sum_{y'} \alpha_{y'}} \text{ and } p(\text{new}|Y, \alpha) = \frac{\alpha}{n + \alpha}$$

Chinese Restaurant Metaphor



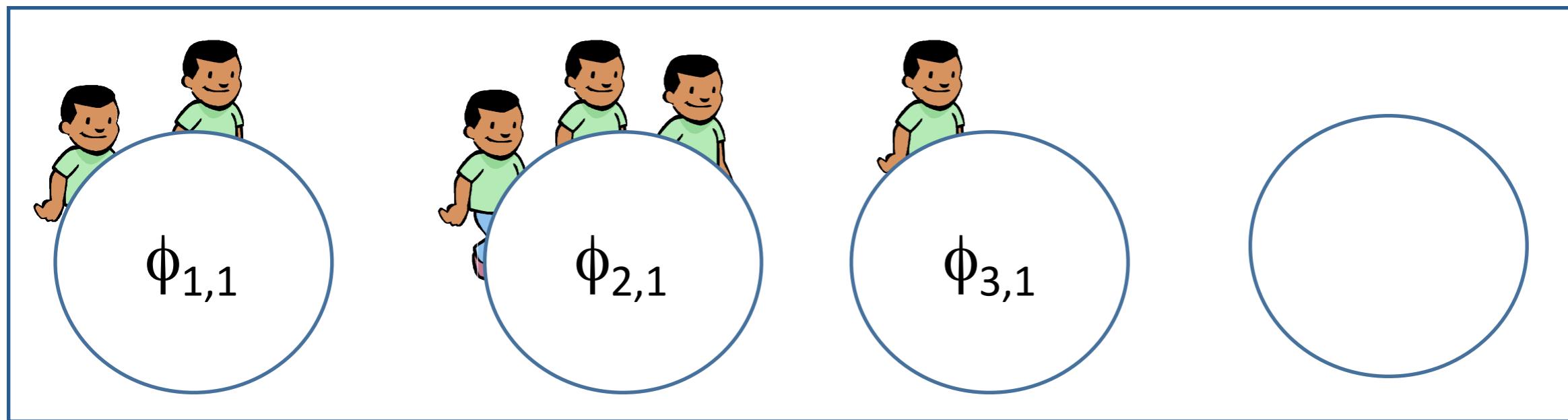
Generative Process

- For data point x_i
 - Choose table $j \propto m_j$ and Sample $x_i \sim f(\phi_j)$
 - Choose a new table $K+1 \propto \alpha$
 - Sample $\phi_{K+1} \sim G_0$ and Sample $x_i \sim f(\phi_{K+1})$

Evolutionary Clustering

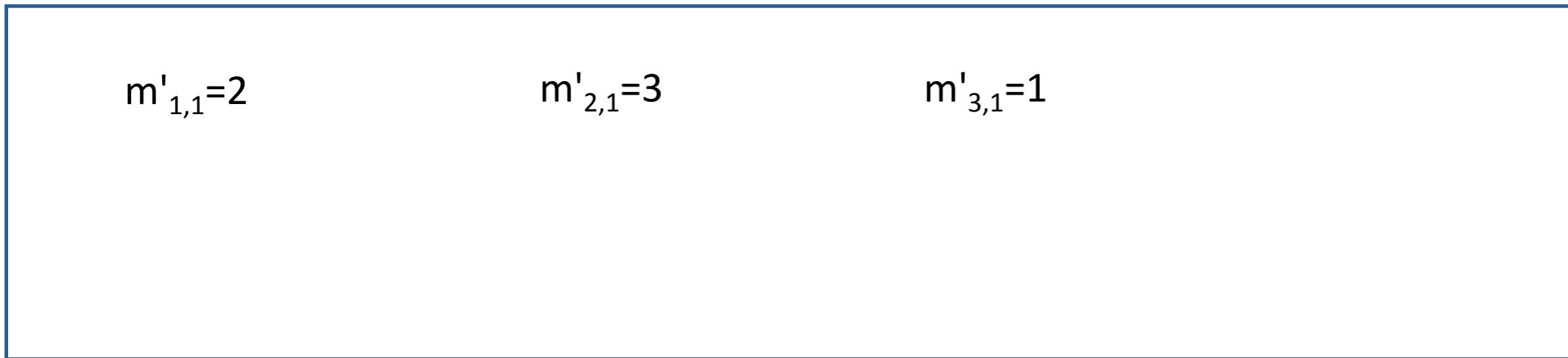
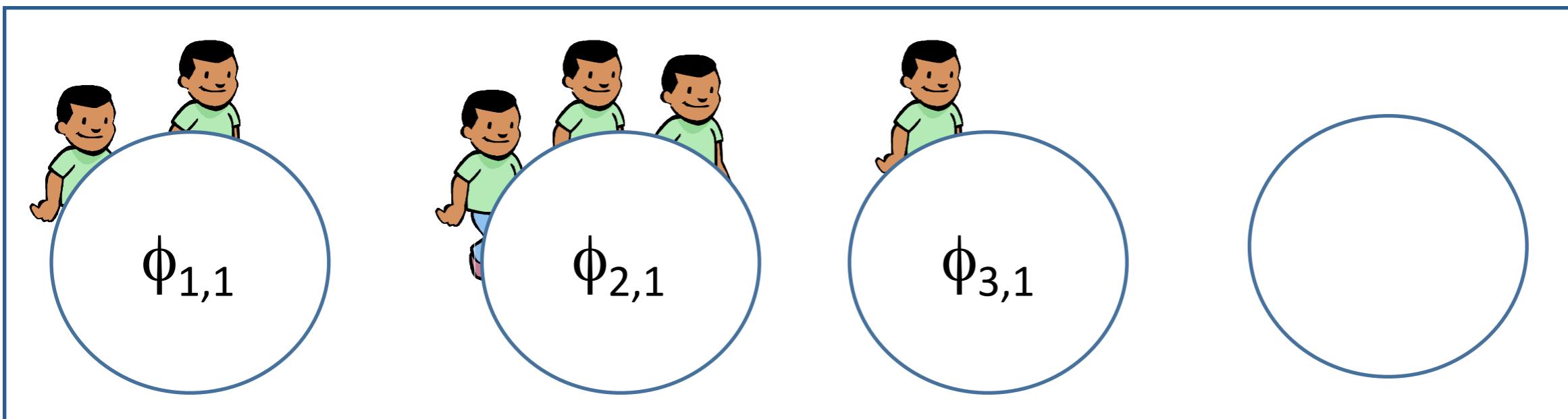
- Time series of objects, e.g. news stories
- Stories appear / disappear
- Want to keep track of clusters automatically

Recurrent Chinese Restaurant Process

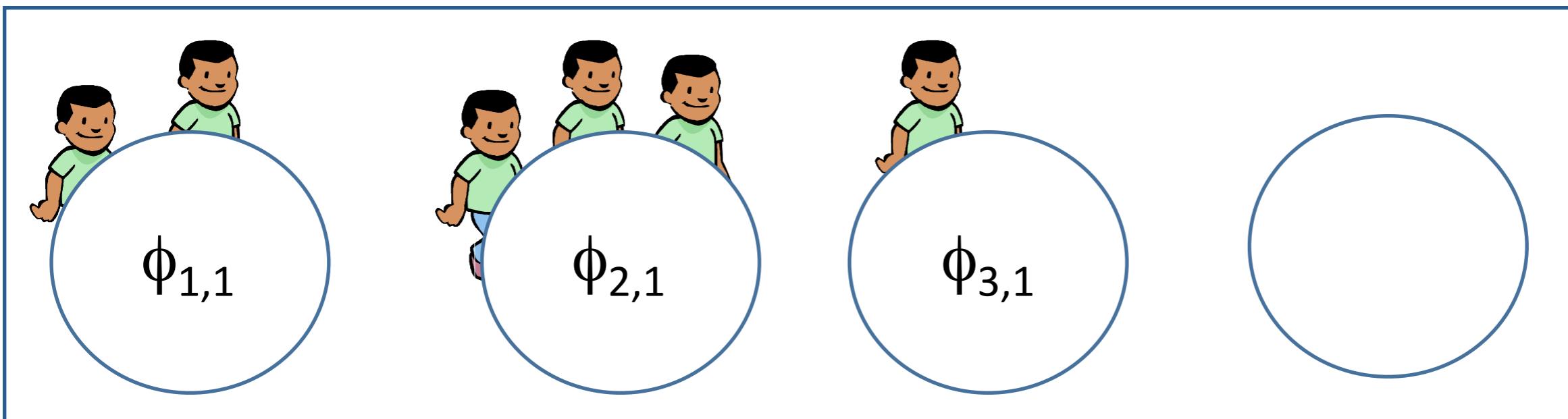


$T=1$

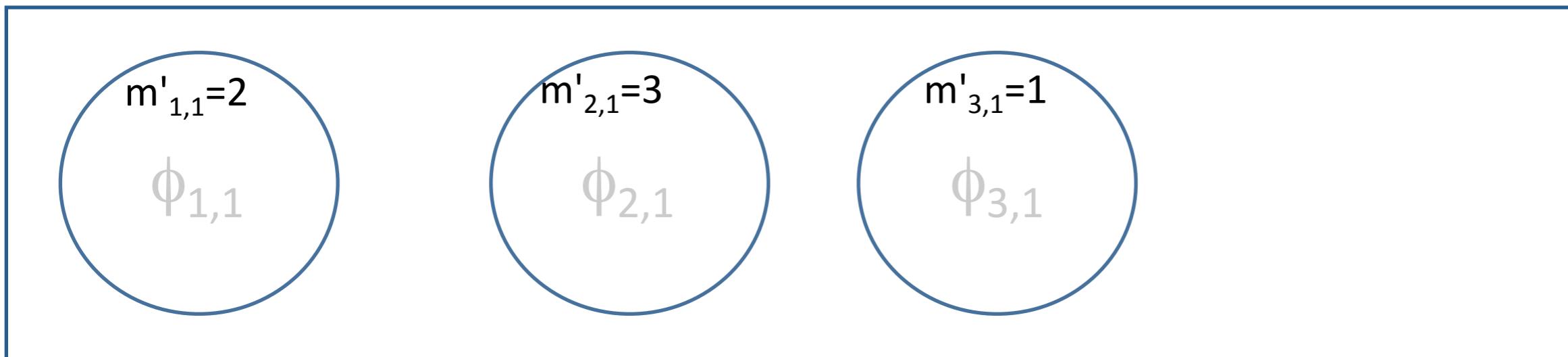
Recurrent Chinese Restaurant Process



Recurrent Chinese Restaurant Process



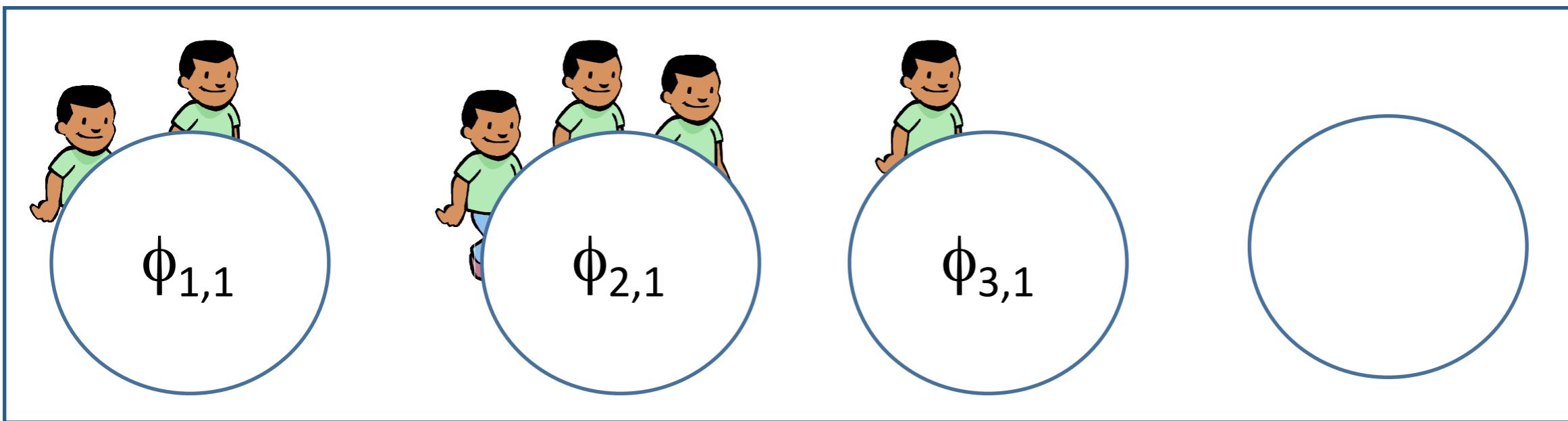
$T=1$



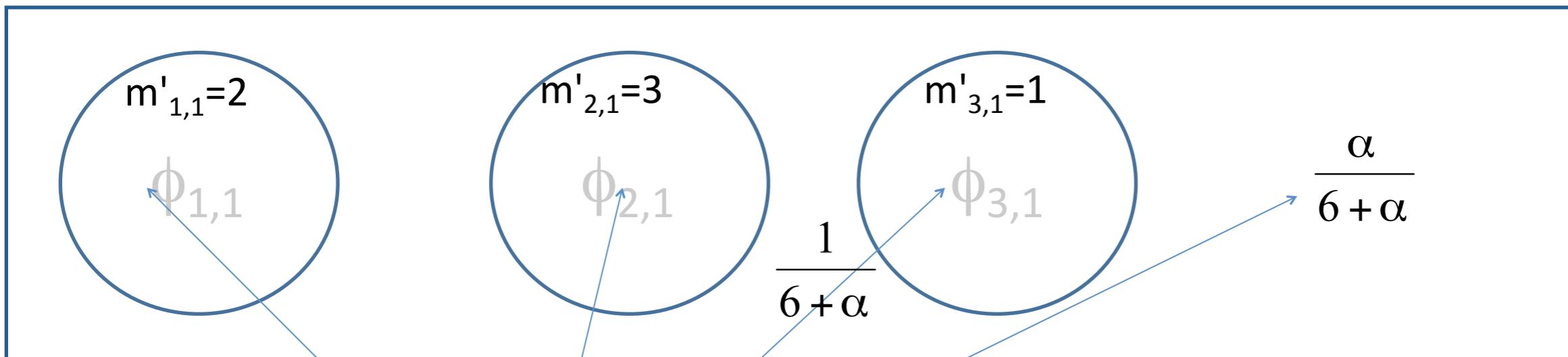
$T=2$



Recurrent Chinese Restaurant Process



$T=1$



$T=2$

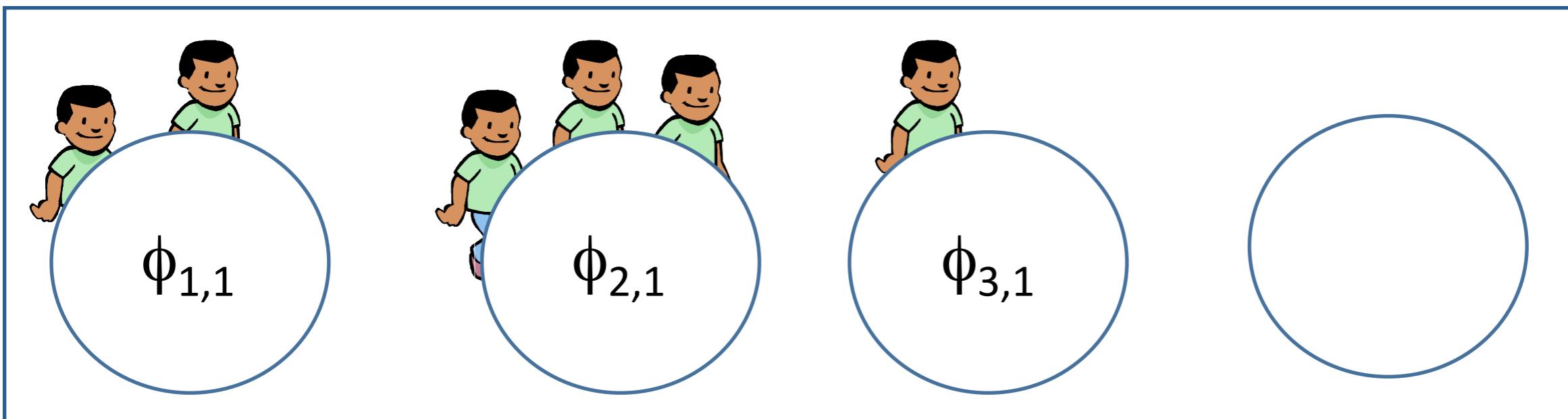
$$\frac{2}{6+\alpha}$$

$$\frac{3}{6+\alpha}$$

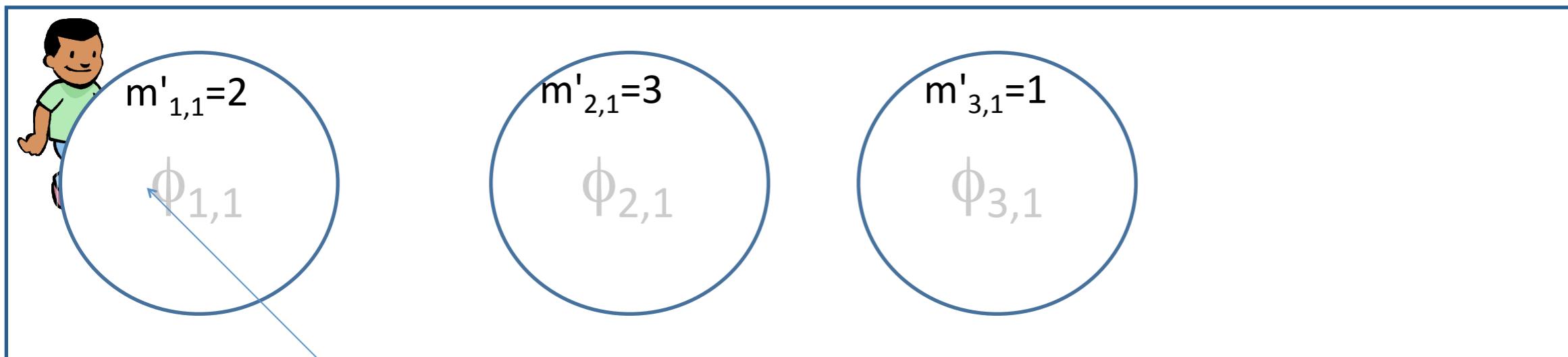
$$\frac{1}{6+\alpha}$$



Recurrent Chinese Restaurant Process



$T=1$

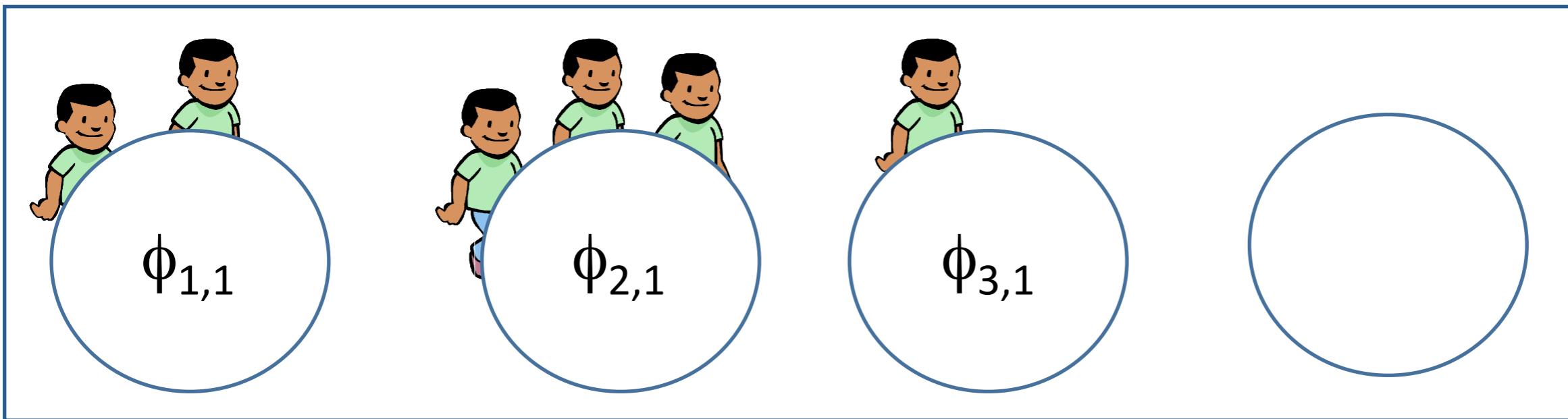


$T=2$

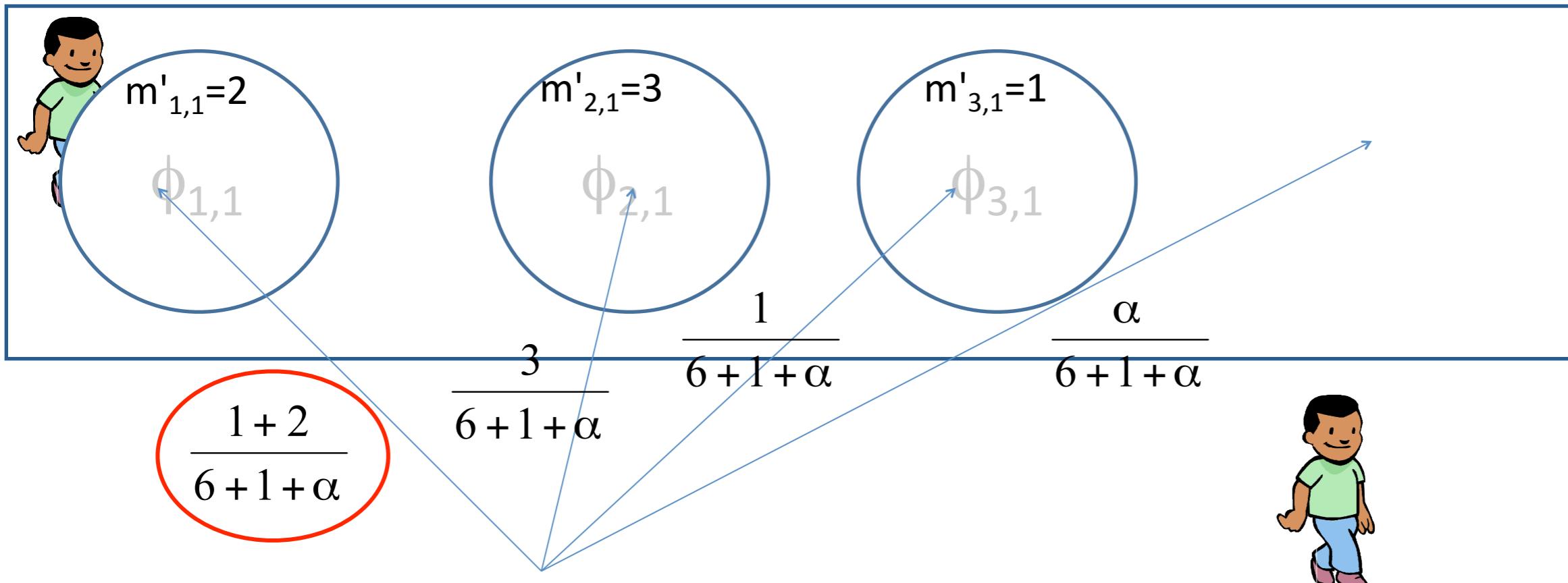
$$\frac{2}{6 + \alpha}$$

Sample $\phi_{1,2} \sim P(\cdot | \phi_{1,1})$

Recurrent Chinese Restaurant Process

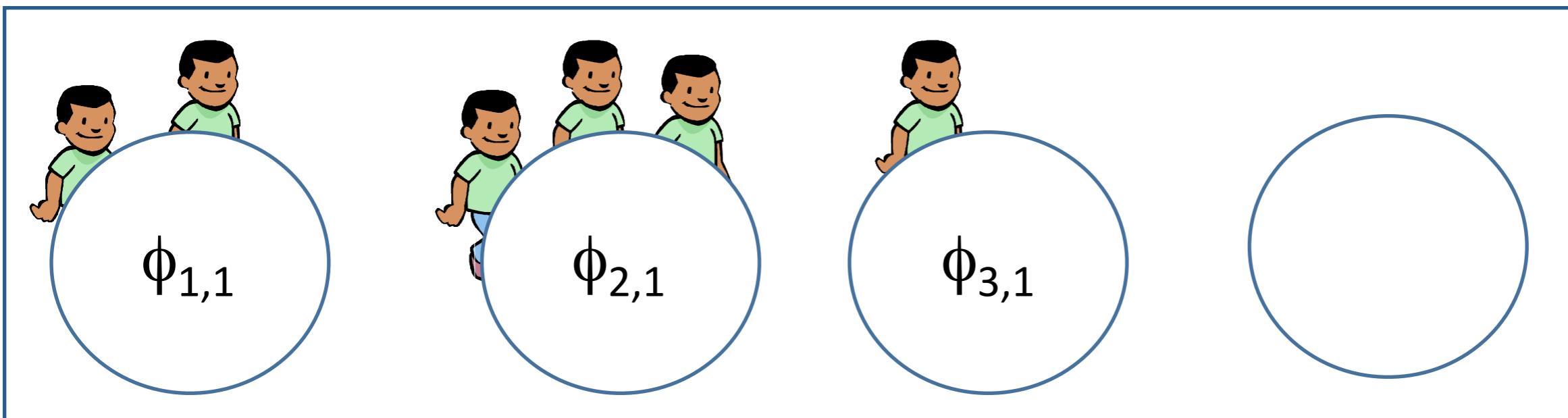


$T=1$

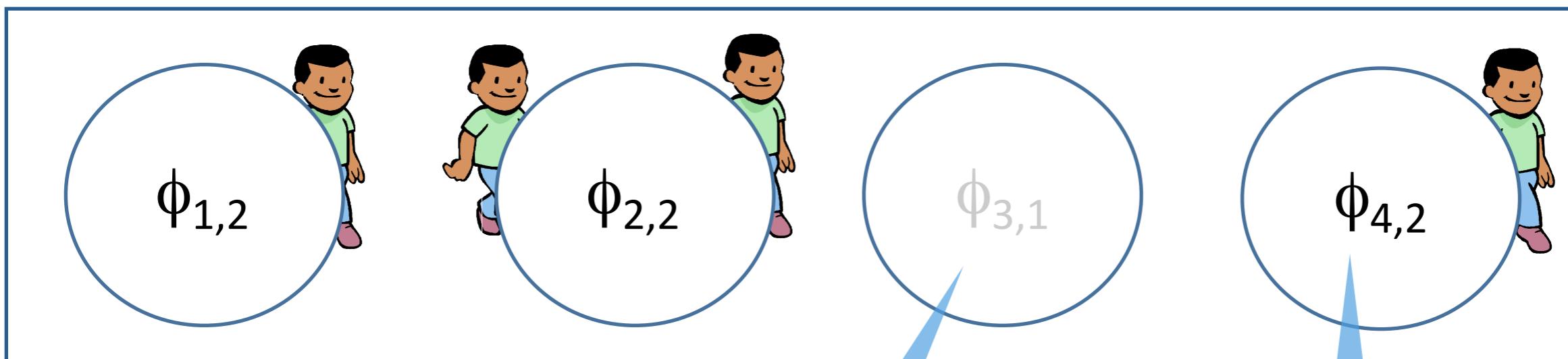


$T=2$

Recurrent Chinese Restaurant Process



$T=1$

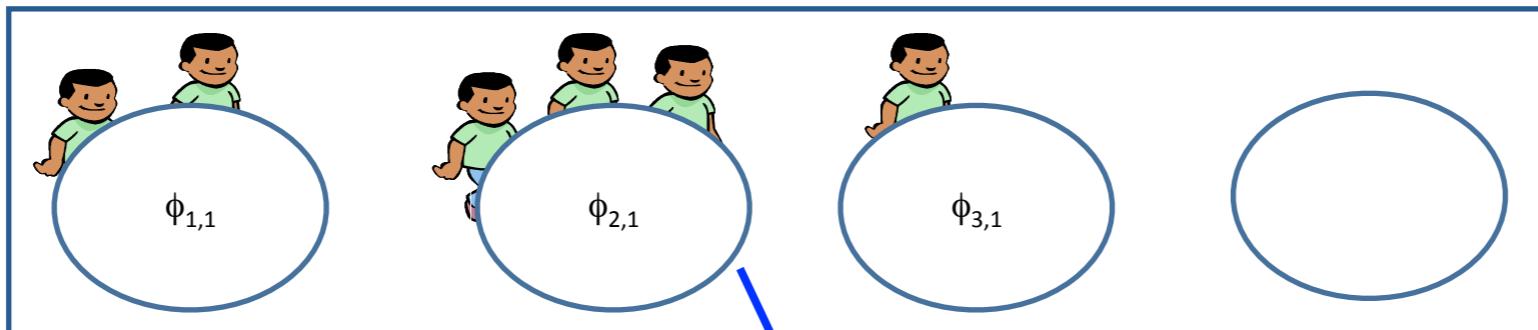


$T=2$

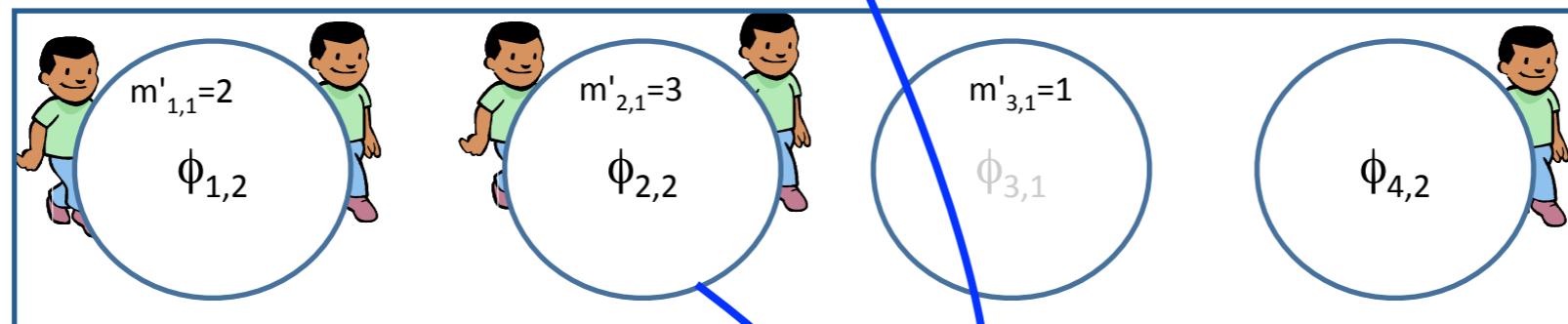
dead cluster

new cluster

Longer History



$T=1$



$T=2$



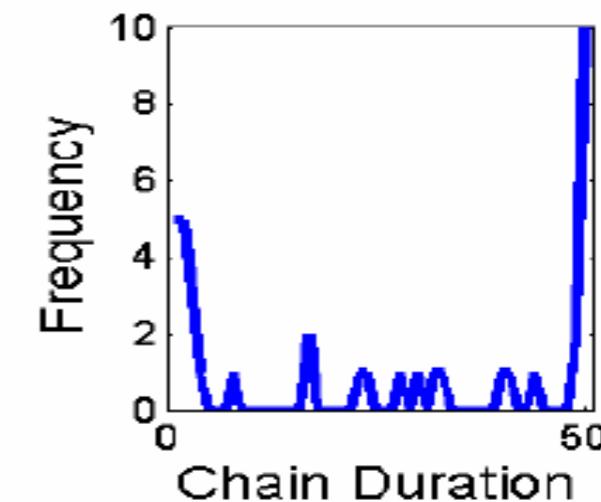
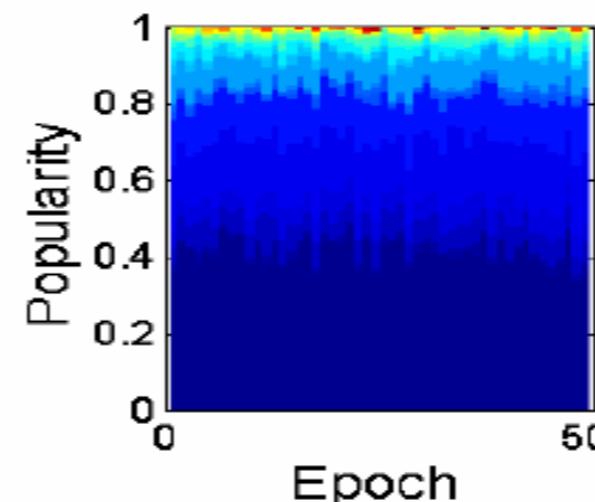
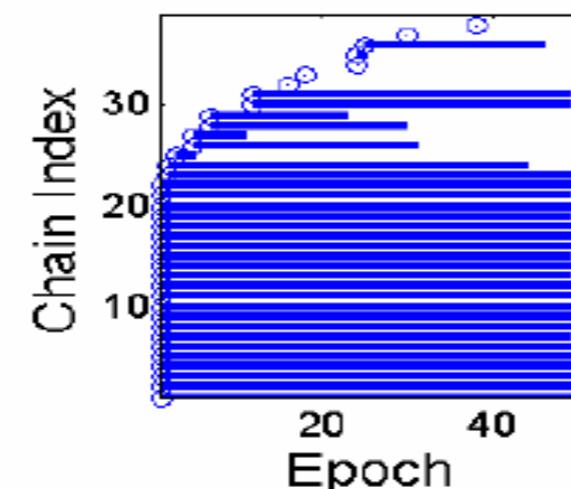
$T=3$

TDPM Generative Power

DPM

$W = \infty$

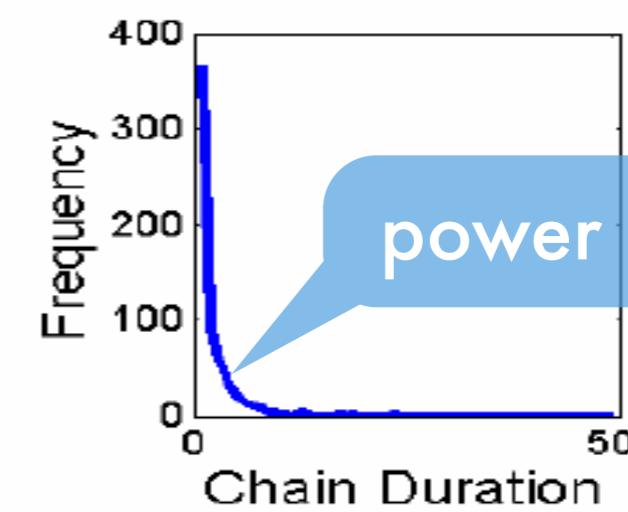
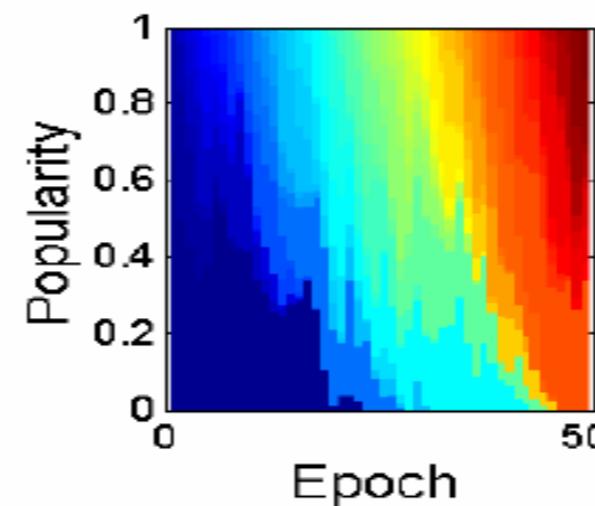
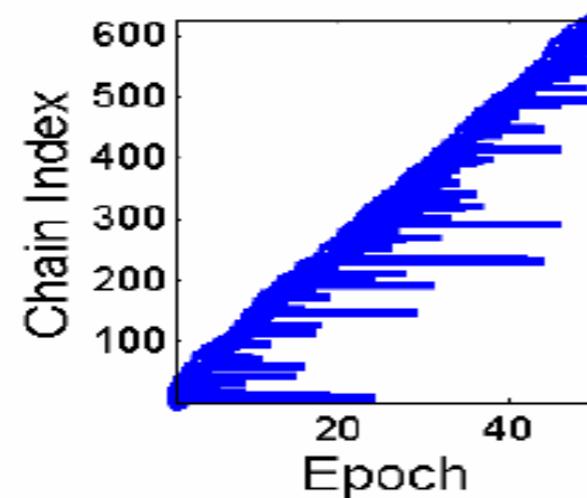
$\lambda = \infty$



TDPM

$W=4$

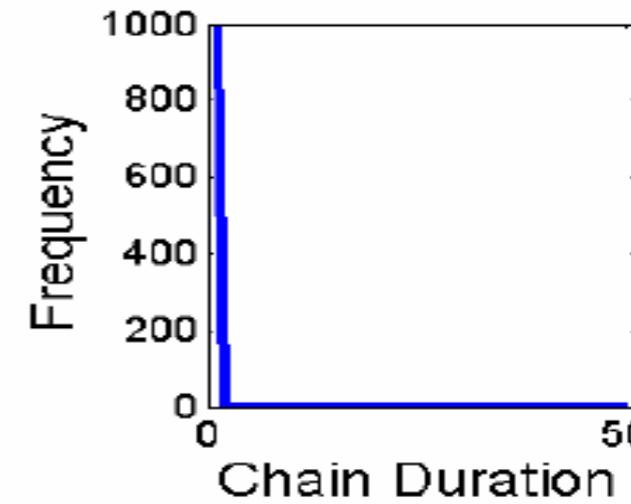
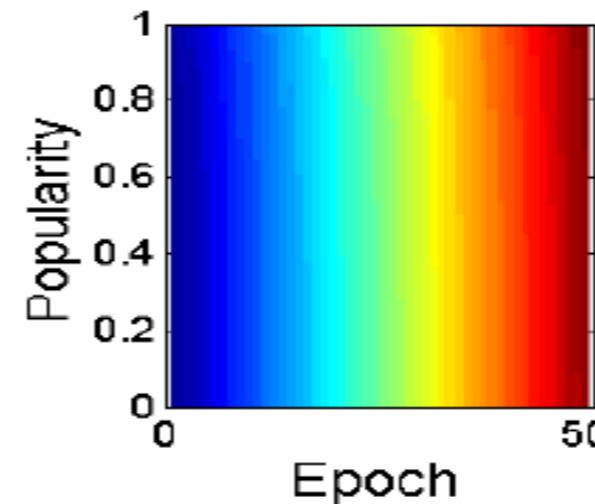
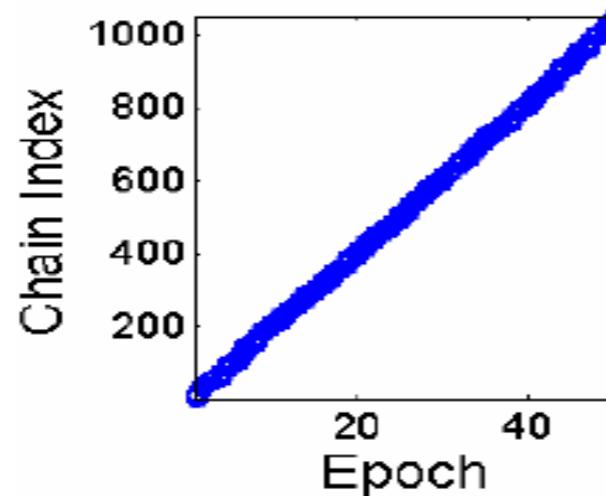
$\lambda = .4$



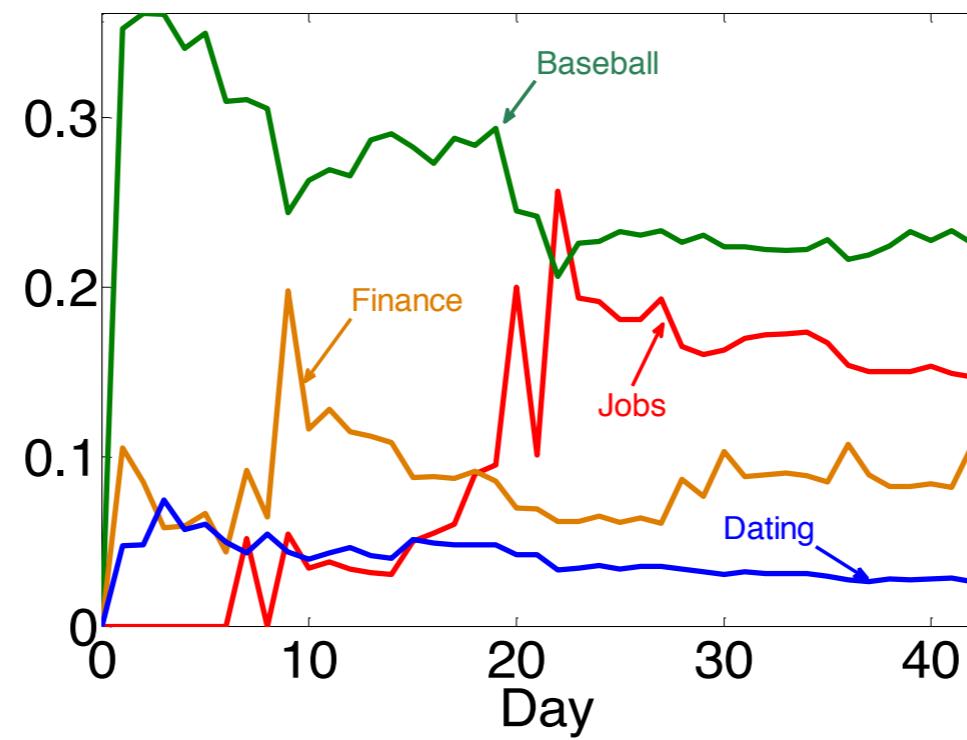
Independent
DPMs

$W= 0$

$\lambda = ?$ (any)



User modeling



Buying a camera

time

Buying a camera

YAHOO! Web Images Video Local Shopping News More ▾

panasonic lx5

Search In: the Web pages in English, French, German, Italian and Spanish

Also try: [panasonic lx5](#), [more...](#)

Panasonic LX5 Cheap
Best Value for Panasonic LX5. Find NexTag Sellers
www.NexTag.com

Panasonic Lumix DMC-LX5 Review (white)
\$434.00 as of Oct 17, 2010 Despite its shortcomings the Panasonic Lumix DMC- LX5 delivers an exceller fastest in its class ...
reviews.cnet.com/digital-cameras/panasonic-lumix-this-site

Panasonic LX5 | Get The Lowest Price On
Panasonic LX5 with 14.1MP captures enough detail.
Panasonic LX5 Camera
www.panasoniclx5.com - [Cached](#) - [More from this s](#)

Panasonic Lumix DMC-LX5 White Digital Camera
shopping.yahoo.com
The Panasonic Lumix DMC-LX5 is a compact digital photo enthusiasts the ideal way for capturing profess photos and High De...
Price: **\$434 to \$513.99**
[Reviews](#) | [Price & Details](#) | [Specs](#)

Sponsored Results

amazon.com Prime
Shop All Departments Search Electronics

Alexander Smola: Hello, Alexander Smola. We have [recommendations](#) for you. ([Not Alexander?](#))
Alexander's Amazon.com | Today's Deals | [Gifts & Wish Lists](#) | [Gift Cards](#)

Instant Order Update for Alexander Smola. You purchased this item on October 6, 2010. [View Details](#)
Color: Black

Member: Alexander Smola

Alexander Smola: This item is eligible for Amazon Prime. [Click here to turn on 1-Click](#) and make Prime even better for you. (With 1-Click enabled, you can always use the regular shopping cart as well.)

Panasonic Lumix DMC-LX5 10.1 MP Digital Camera with 3.8x Optical Image Stabilized Zoom and 3.0-Inch LCD (Black) by [Panasonic](#)
 (40 customer reviews)
List Price: **\$499.00**
Price: **\$444.95** & eligible for free shipping with **Amazon Prime**
You Save: **\$54.05 (11%)**

time

Buying a camera

YAHOO! Web Images Video Local Shopping News More ▾

panasonic lx5

Search In: the Web pages in English, French, German, Italian and Spanish

Also try: [panasonic lx5](#), [more...](#)

Panasonic LX5 Cheap
Best Value for Panasonic LX5. Find NexTag Sellers
www.NexTag.com

Panasonic Lumix DMC-LX5 Review (white)
\$434.00 as of Oct 17, 2010 Despite its shortcomings the Panasonic Lumix DMC- LX5 delivers an exceller fastest in its class ...
reviews.cnet.com/digital-cameras/panasonic-lumix-this-site

Panasonic LX5 | Get The Lowest Price On
Panasonic LX5 with 14.1MP captures enough detail.
Panasonic LX5 Camera
www.panasoniclx5.com - [Cached](#) - [More from this s](#)

Panasonic Lumix DMC-LX5 White Digital Camera
shopping.yahoo.com

Sponsor Results

show ads now

time

amazon.com Prime

Sponsored Results

Hello, Alexander Smola. We have [recommendations](#) for you. ([Not Alexander?](#))
Alexander's Amazon.com | Today's Deals | Gifts & Wish Lists | Gift Cards

Shop All Departments Search Electronics GO

Camera & Photo All Electronics Brands Bestsellers Digital SLRs & Lenses Point-And-Shoots Camcorders

Instant Order Update for Alexander Smola. You purchased this item on October 6, 2010. [View Details](#)

Color: Black

Prime Member: Alexander Smola

Alexander Smola: This item is eligible for Amazon Prime. [Click here to turn on 1-Click](#) and make Prime even better for you. (With 1-Click enabled, you can always use the regular shopping cart as well.)

Panasonic Lumix DMC-LX5 10.1 MP Digital Camera with 3.8x Optical Image Stabilized Zoom and 3.0-Inch LCD (Black) by [Panasonic](#)

★★★★★ (40 customer reviews)

List Price: \$499.00
Price: **\$444.95** & eligible for free shipping with **Amazon Prime**
You Save: \$54.05 (11%)

Buying a camera

YAHOO! Web Images Video Local Shopping News More ▾

panasonic lx5

Search In: the Web pages in English, French, German, Italian and Spanish

Also try: [panasonic lx5](#), [more...](#)

Panasonic LX5 Cheap
Best Value for Panasonic LX5. Find NexTag Sellers
www.NexTag.com

Panasonic Lumix DMC-LX5 Review (white)
\$434.00 as of Oct 17, 2010 Despite its shortcomings the Panasonic Lumix DMC- LX5 delivers an exceller fastest in its class ...
reviews.cnet.com/digital-cameras/panasonic-lumix-this-site

Panasonic LX5 | Get The Lowest Price On
Panasonic LX5 with 14.1MP captures enough detail.
Panasonic LX5 Camera
www.panasoniclx5.com - [Cached](#) - [More from this s](#)

Panasonic Lumix DMC-LX5 White Digital Camera
shopping.yahoo.com

Sponsored Results

amazon.com Prime
Shop All Departments Search Electronics

Member: Alexander Smola

Instant Order Update for Alexander Smola. You purchased this item on October 6, 2010. [View Details](#)

Alexander Smola: This item is eligible for Amazon Prime. [Click here to turn on 1-Click](#) and make Prime even better for you. (With 1-Click enabled, you can always use the regular shopping cart as well.)

Panasonic Lumix DMC-LX5 10.1 MP Digital Camera with 3.8x Optical Image Stabilized Zoom and 3.0-Inch LCD (Black) by [Panasonic](#)

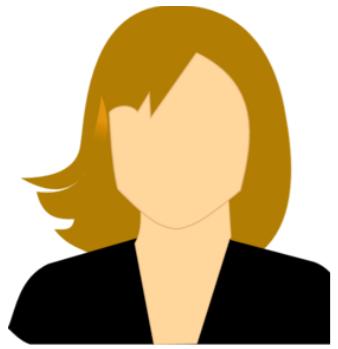
★★★★★ (40 customer reviews)

List Price: \$499.00
Price: \$444.95 & eligible for [Amazon Prime](#)
You Save: \$54.05 (11%)

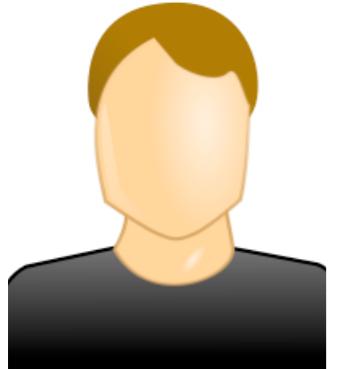
show ads now

time

too late

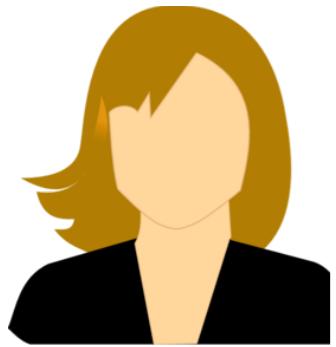


Car
Deals
van



job
Hiring
diet

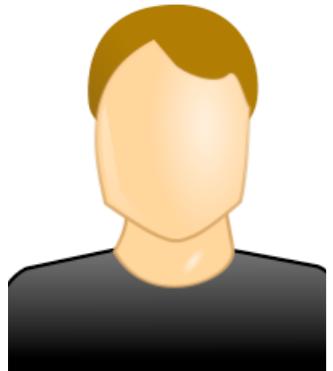




Car
Deals
van

Auto
Price
Used
inspection

Movies
Theatre
Art
gallery



job
Hiring
diet

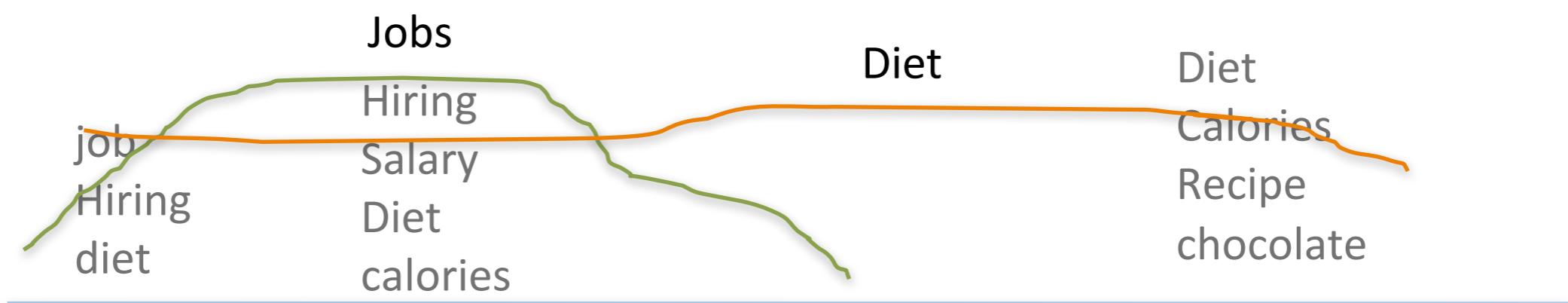
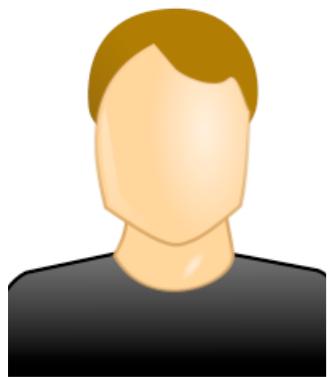
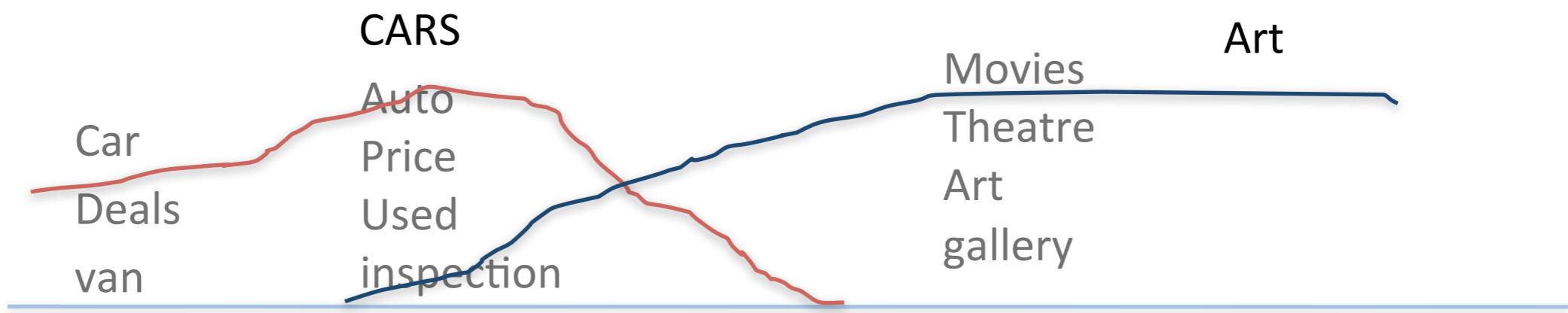
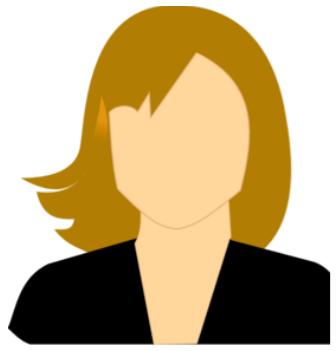
Hiring
Salary
Diet
calories

Diet
Calories
Recipe
chocolate



Flight
London
Hotel
weather

School
Supplies
Loan
college



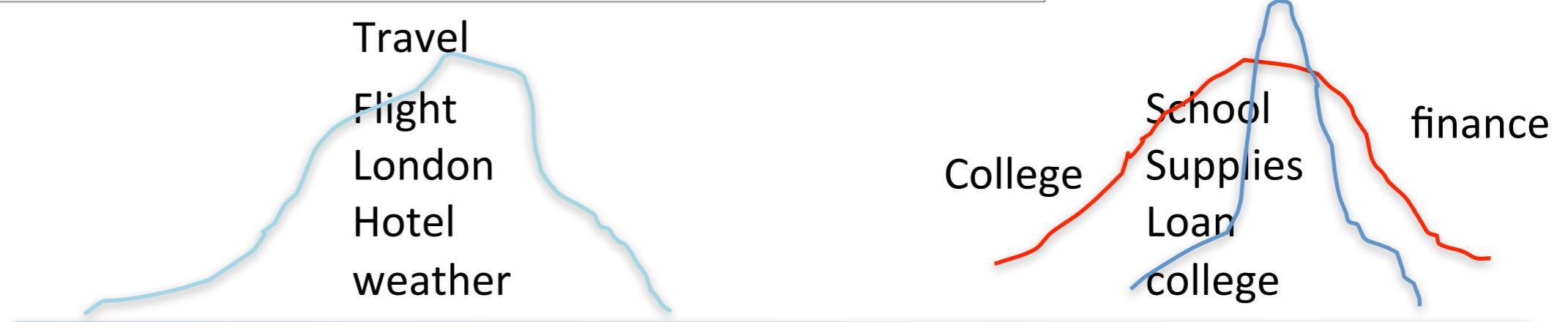
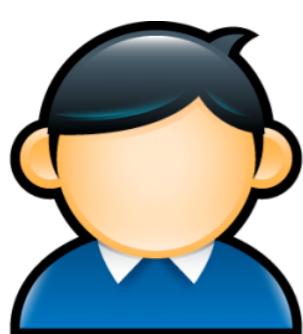
User modeling

Input

- Queries issued by the user or Tags of watched content
- Snippet of page examined by user
- Time stamp of each action (day resolution)

Output

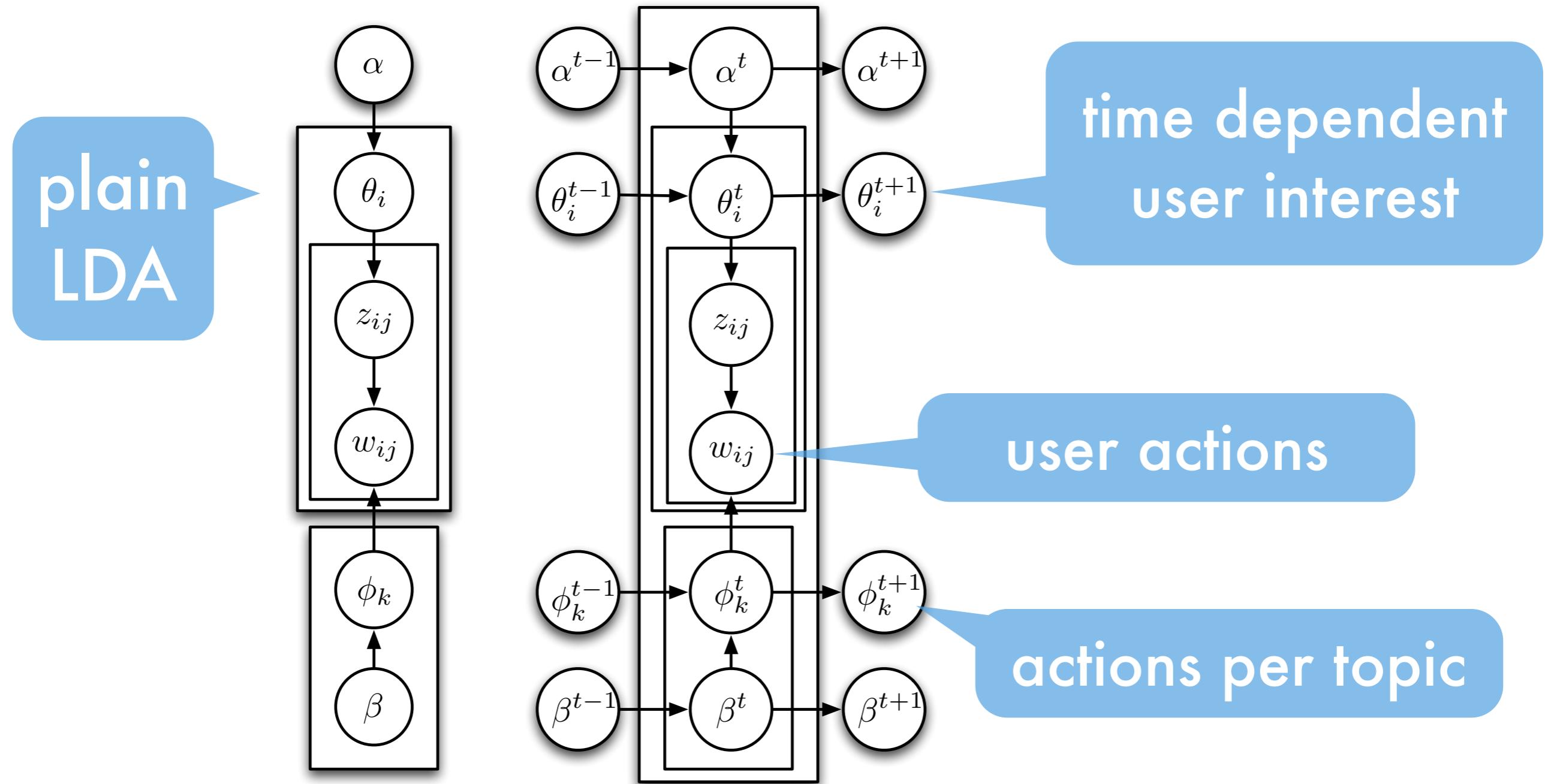
- Users' daily distribution over intents
- Dynamic intent representation

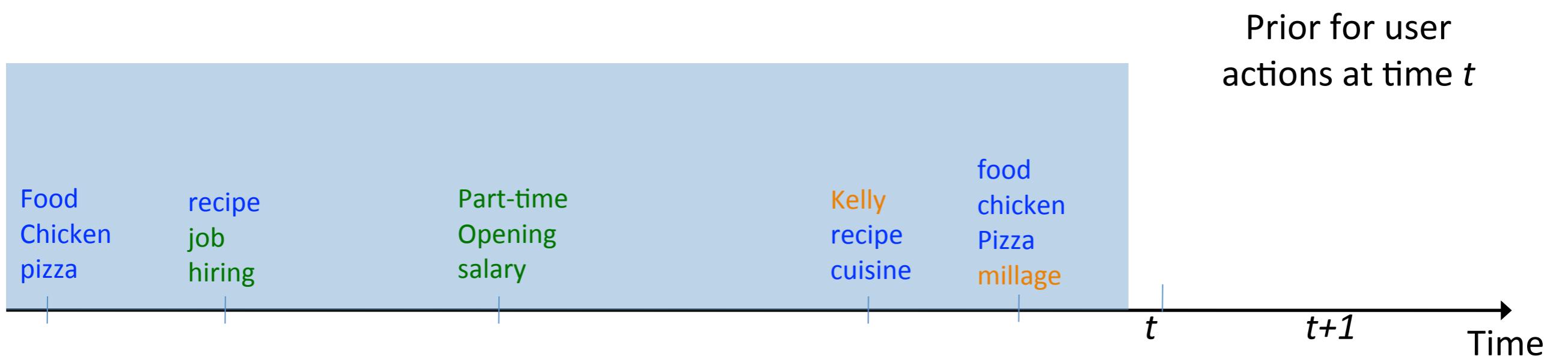


Time dependent models

- LDA for topical model of users where
 - User interest distribution changes over time
 - Topics change over time
- This is like a Kalman filter except that
 - Don't know what to track (a priori)
 - Can't afford a Rauch-Tung-Striebel smoother
 - Much more messy than plain LDA

Graphical Model





Diet

- Recipe
- Chocolate
- Pizza
- Food
- Chicken
- Milk
- Butter
- Powder

Cars

- Car
- Blue
- Book
- Kelley
- Prices
- Small
- Speed
- large

Job

- job
- Career
- Business
- Assistant
- Hiring
- Part-time
- Receptionist

Finance

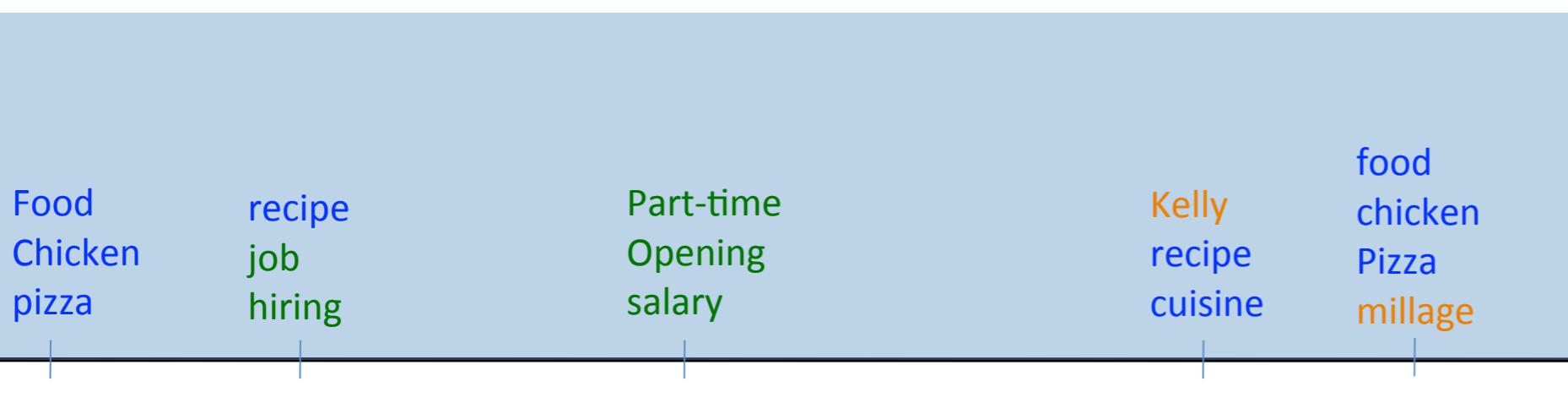
- Bank
- Online
- Credit
- Card
- debt
- portfolio
- Finance
- Chase

All



Long-term

Prior for user
actions at time t



Diet

Recipe
Chocolate
Pizza
Food
Chicken
Milk
Butter
Powder

Cars

Car
Blue
Book
Kelley
Prices
Small
Speed
large

Job

job
Career
Business
Assistant
Hiring
Part-time
Receptionist

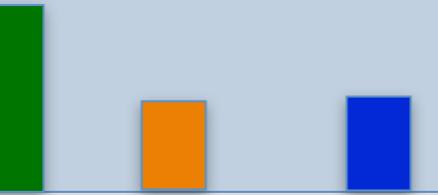
Finance

Bank
Online
Credit
Card
debt
portfolio
Finance
Chase

All



month



Long-term

Prior for user
actions at time t



Diet

Recipe
Chocolate
Pizza
Food
Chicken
Milk
Butter
Powder

Cars

Car
Blue
Book
Kelley
Prices
Small
Speed
large

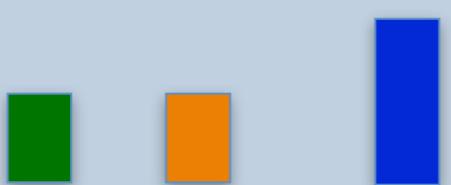
Job

job
Career
Business
Assistant
Hiring
Part-time
Receptionist

Finance

Bank
Online
Credit
Card
debt
portfolio
Finance
Chase

All



month



week



Long-term

short-term

Prior for user
actions at time t

Food
Chicken
pizza

recipe
job
hiring

Part-time
Opening
salary

Kelly
recipe
cuisine

food
chicken
Pizza
millage

t

$t+1$

Time

Diet

Recipe
Chocolate
Pizza
Food
Chicken
Milk
Butter
Powder

Cars

Car
Blue
Book
Kelley
Prices
Small
Speed
large

Job

job
Career
Business
Assistant
Hiring
Part-time
Receptionist

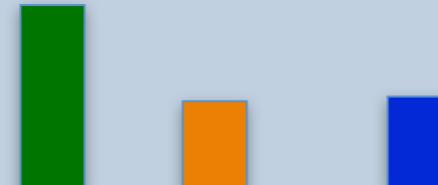
Finance

Bank
Online
Credit
Card
debt
portfolio
Finance
Chase

All



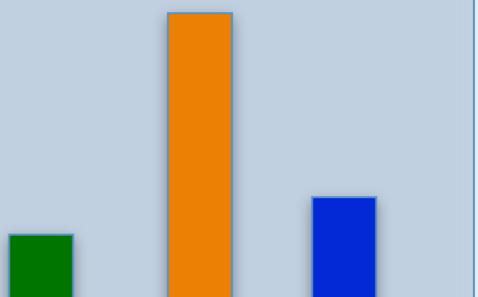
month



Long-term

short-term

week



Prior for user
actions at time t

Food
Chicken
pizza

recipe
job
hiring

Part-time
Opening
salary

Kelly
recipe
cuisine

food
chicken
Pizza
millage

t

$t+1$

Time

Diet

Recipe
Chocolate
Pizza
Food
Chicken
Milk
Butter
Powder

Cars

Car
Blue
Book
Kelley
Prices
Small
Speed
large

Job

job
Career
Business
Assistant
Hiring
Part-time
Receptionist

Finance

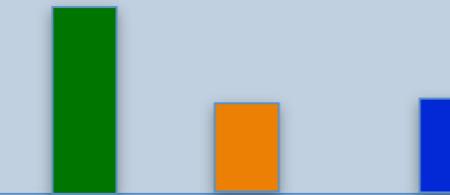
Bank
Online
Credit
Card
debt
portfolio
Finance
Chase

All



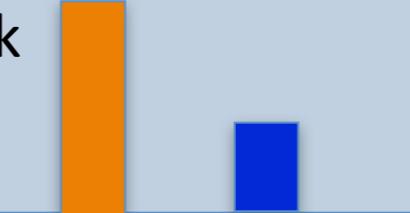
μ^3

month



μ^2

week



μ

Long-term

short-term

Prior for user
actions at time t



Diet

Recipe
Chocolate
Pizza
Food
Chicken
Milk
Butter
Powder

Cars

Car
Blue
Book
Kelley
Prices
Small
Speed
large

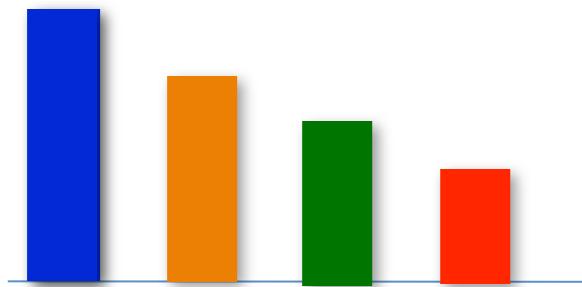
Job

job
Career
Business
Assistant
Hiring
Part-time
Receptionist

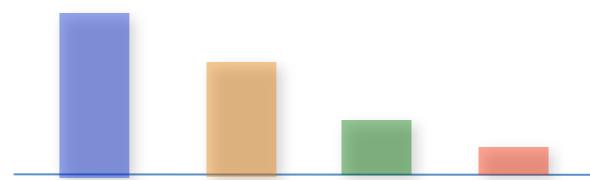
Finance

Bank
Online
Credit
Card
debt
portfolio
Finance
Chase

At time t



At time t+1

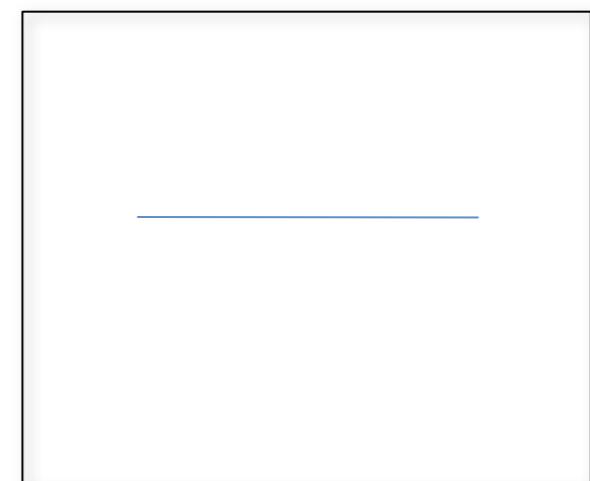


Recipe
Chocolate
Pizza
Food
Chicken
Milk
Butter
Powder

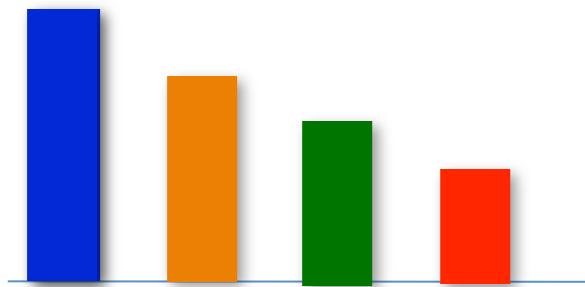
Car
Altima
Accord
Blue
Book
Kelley
Prices
Small
Speed

job
Career
Business
Assistant
Hiring
Part-time
Receptionist

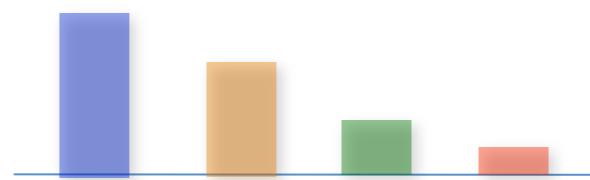
Bank
Online
Credit
Card
debt
portfolio
Finance
Chase



At time t



At time t+1

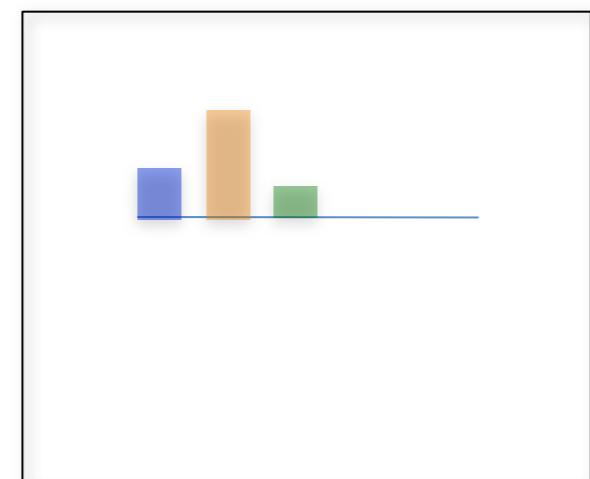


Recipe
Chocolate
Pizza
Food
Chicken
Milk
Butter
Powder

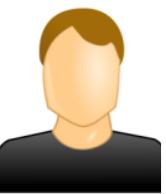
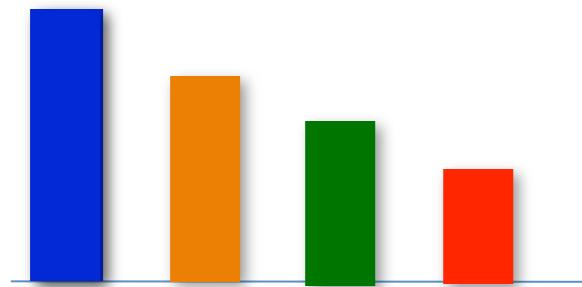
Car
Altima
Accord
Blue
Book
Kelley
Prices
Small
Speed

job
Career
Business
Assistant
Hiring
Part-time
Receptionist

Bank
Online
Credit
Card
debt
portfolio
Finance
Chase

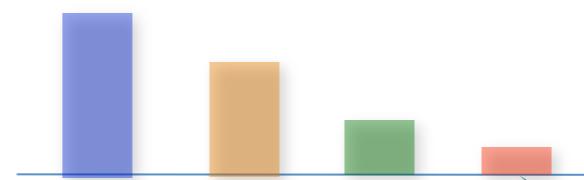


At time t



Food Chicken
Pizza mileage

At time t+1



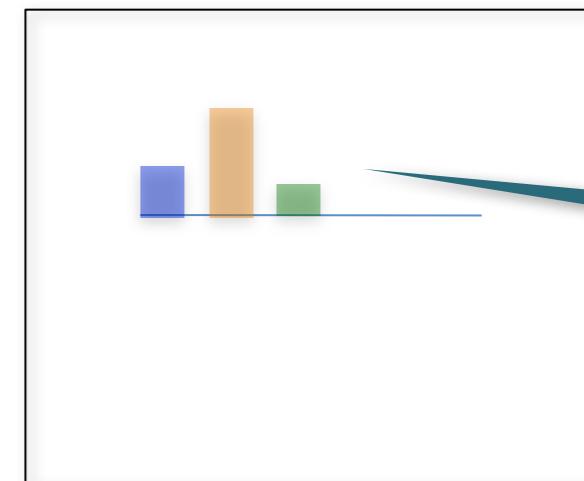
Recipe
Chocolate
Pizza
Food
Chicken
Milk
Butter
Powder

Car
Altima
Accord
Blue
Book
Kelley
Prices
Small
Speed

job
Career
Business
Assistant
Hiring
Part-time
Receptionist

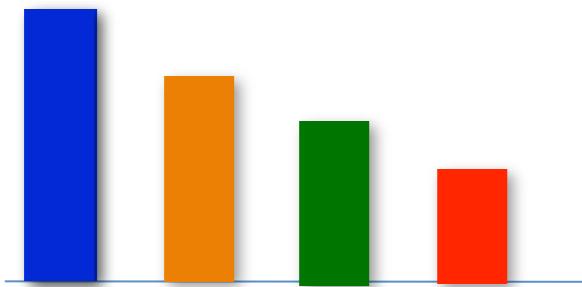
Bank
Online
Credit
Card
debt
portfolio
Finance
Chase

priors



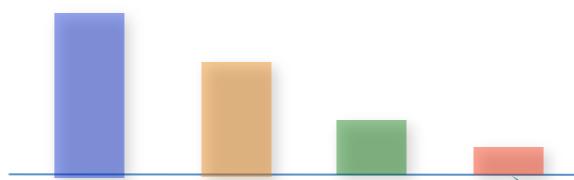
Car speed offer
Camry accord career

At time t



Food Chicken
Pizza mileage

At time t+1



Recipe
Chocolate
Pizza
Food
Chicken
Milk
Butter
Powder

Car
Altima
Accord
Blue
Book
Kelley
Prices
Small
Speed

job
Career
Business
Assistant
Hiring
Part-time
Receptionist

Bank
Online
Credit
Card
debt
portfolio
Finance
Chase

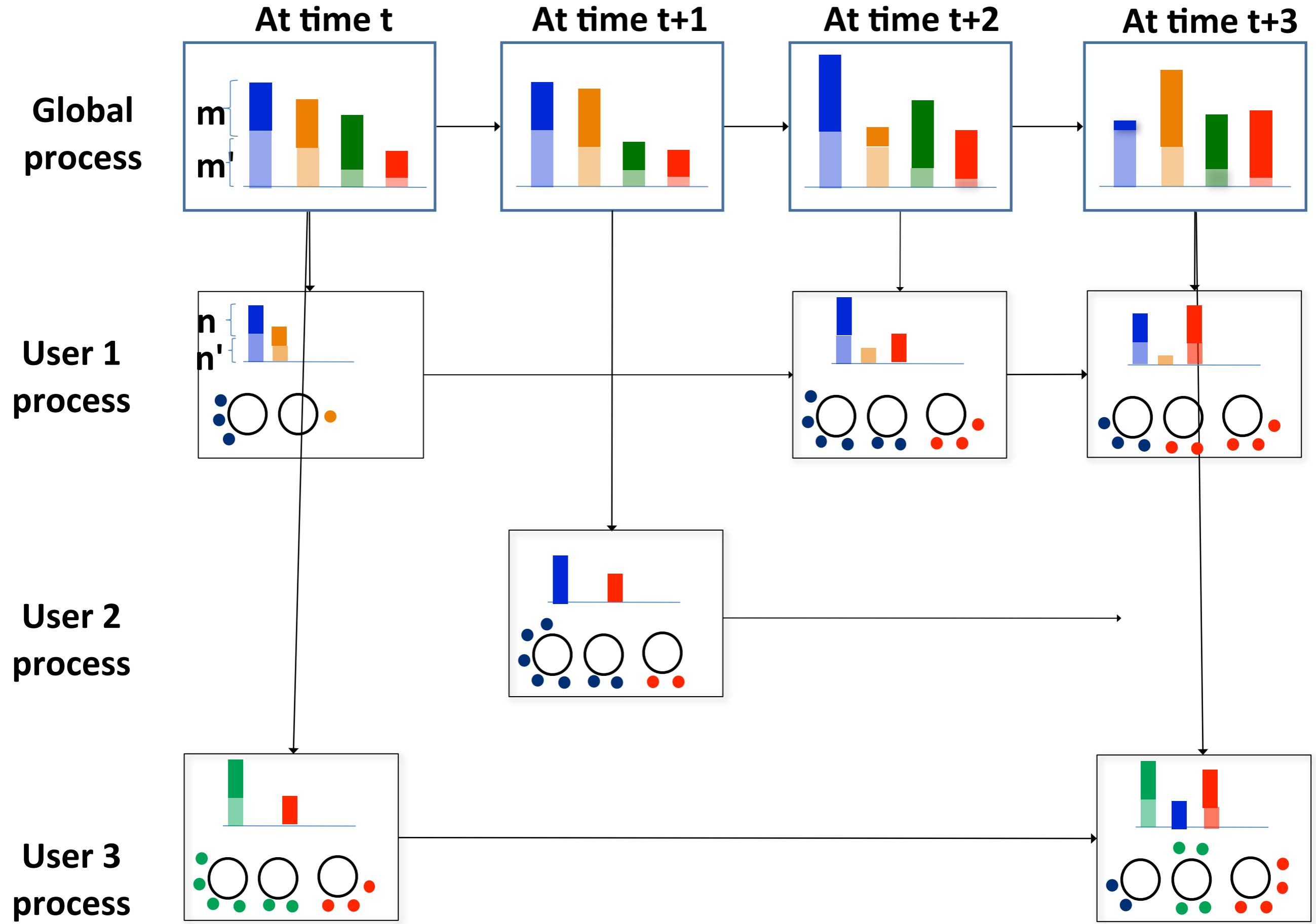
priors

Generative Process

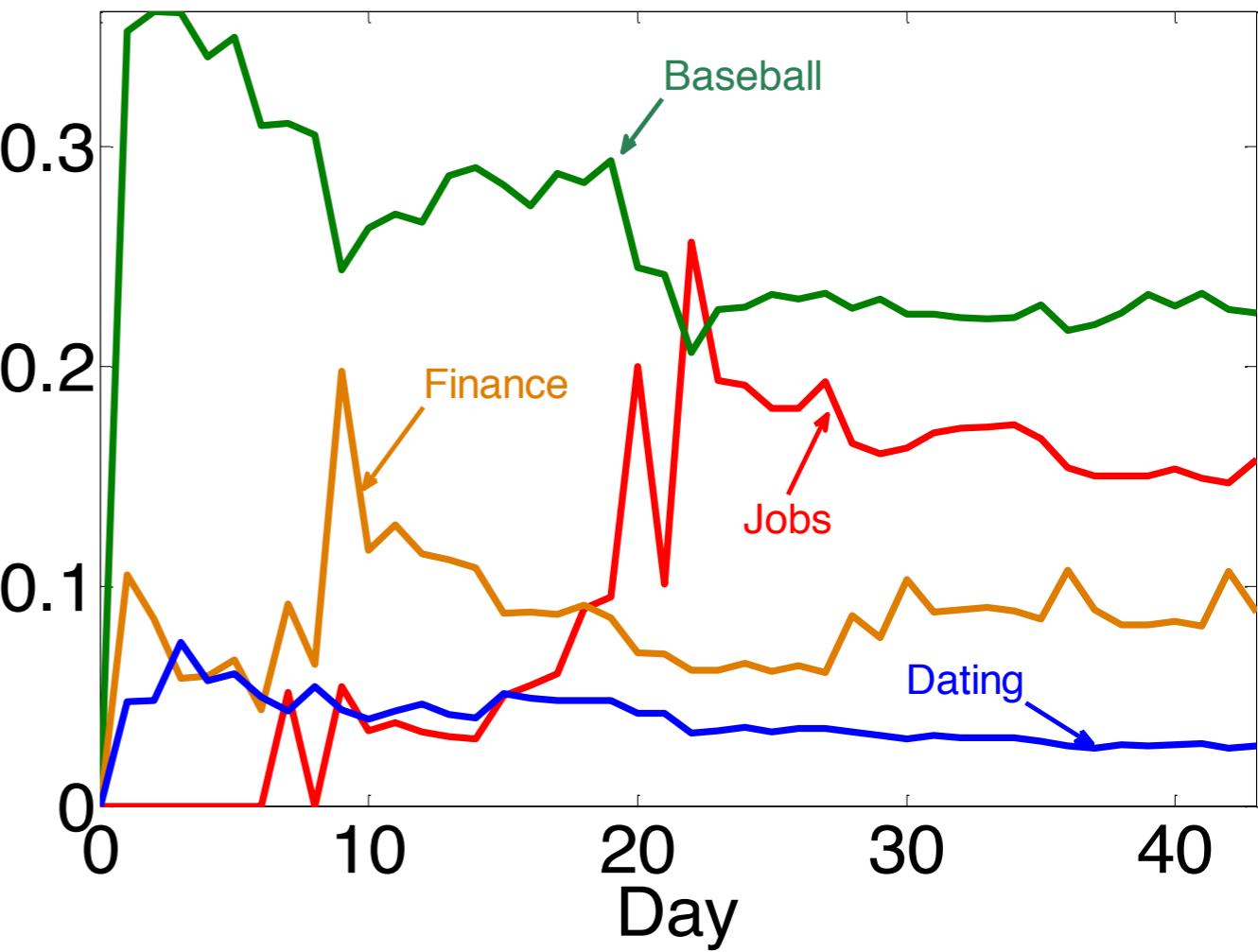
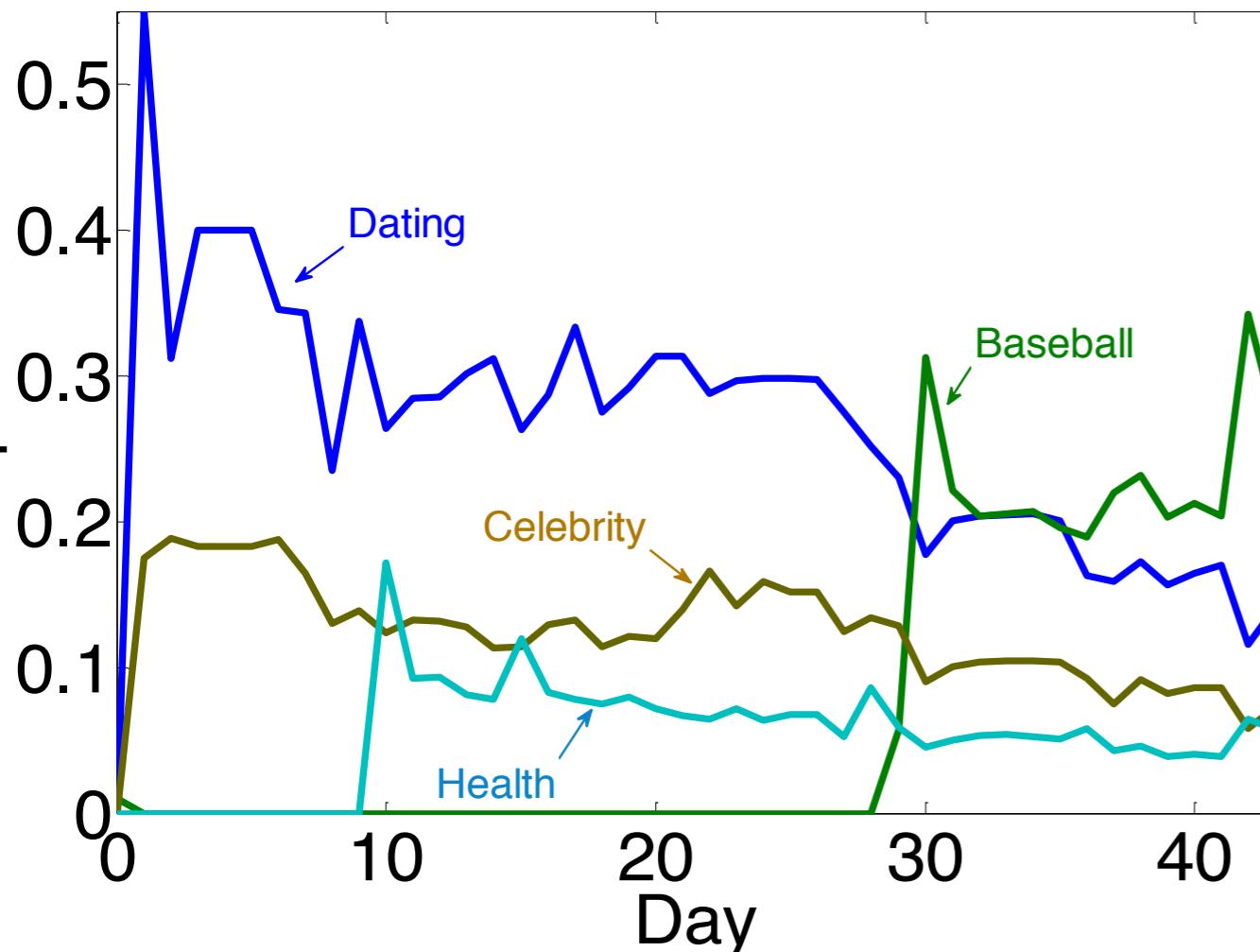
- For each user interaction
 - Choose an intent from local distribution
 - Sample word from the topic's word-distribution
 - Choose a new intent $\propto \alpha$
 - Sample a new intent from the global distribution
 - Sample word from the new topic word-distribution



Car speed offer
Camry accord career



Sample users



Dating

women
men
dating
singles
personals
seeking
match

Baseball

League
baseball
basketball,
doublehead
Bergesen
Griffey
bullpen
Greinke

Celebrity

Snooki
Tom
Cruise
Katie
Holmes
Pinkett
Kudrow
Hollywood

Health

skin
body
fingers
cells
toes
wrinkle
layers

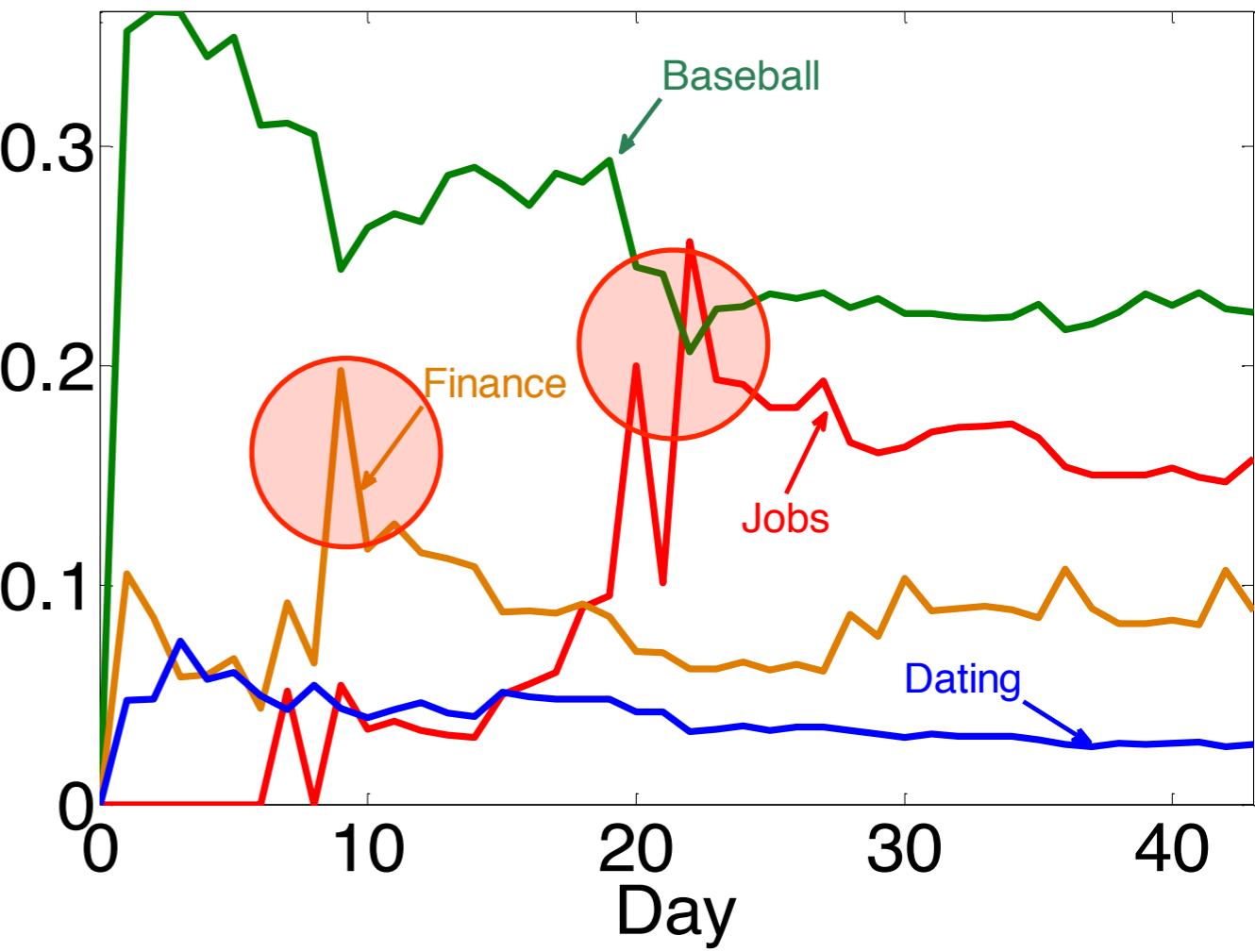
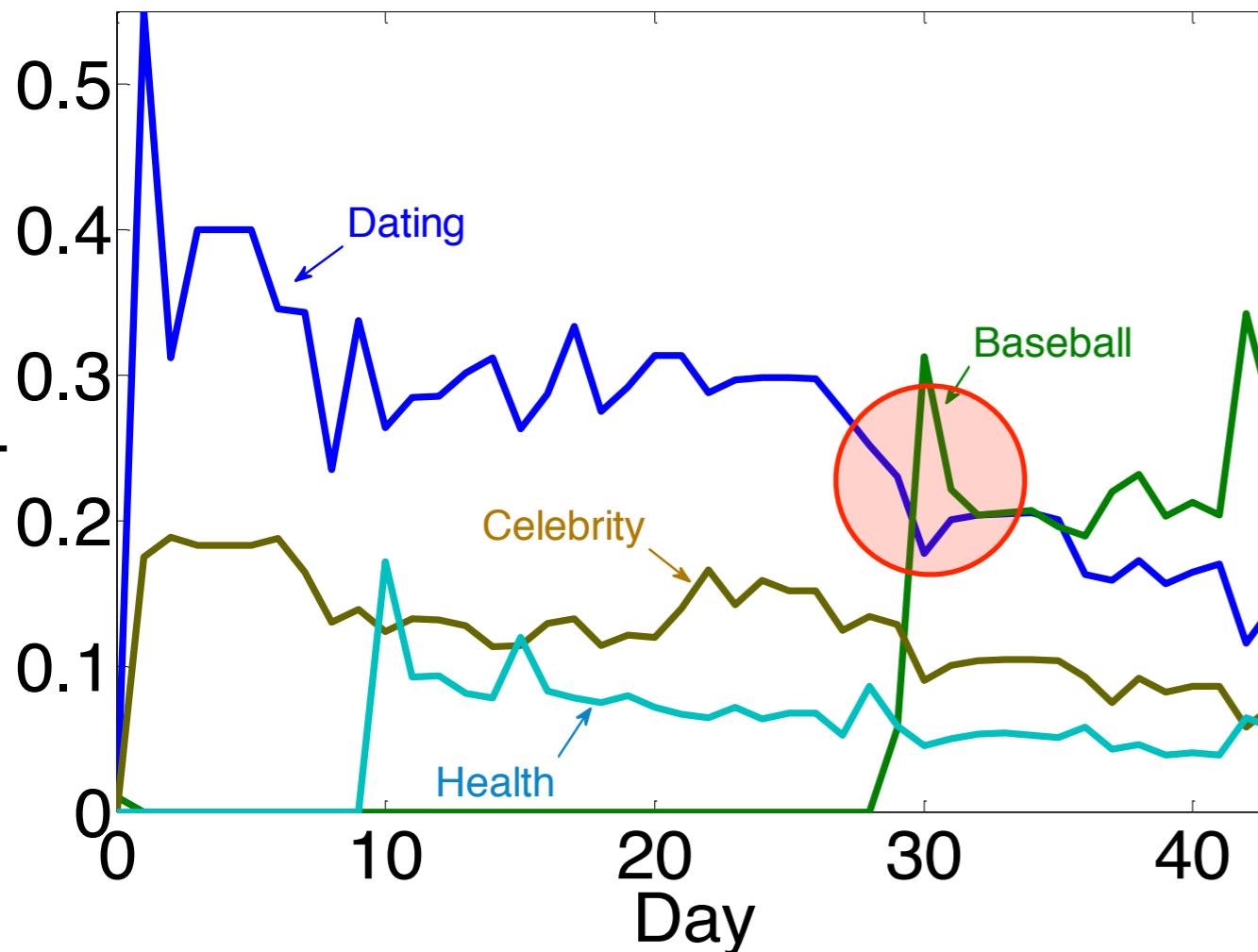
Jobs

job
career
business
assistant
hiring
part-time
receptionist

Finance

financial
Thomson
chart
real
Stock
Trading
currency

Sample users



Dating

women
men
dating
singles
personals
seeking
match

Baseball

League
baseball
basketball,
doublehead
Bergesen
Griffey
bullpen
Greinke

Celebrity

Snooki
Tom
Cruise
Katie
Holmes
Pinkett
Kudrow
Hollywood

Health

skin
body
fingers
cells
toes
wrinkle
layers

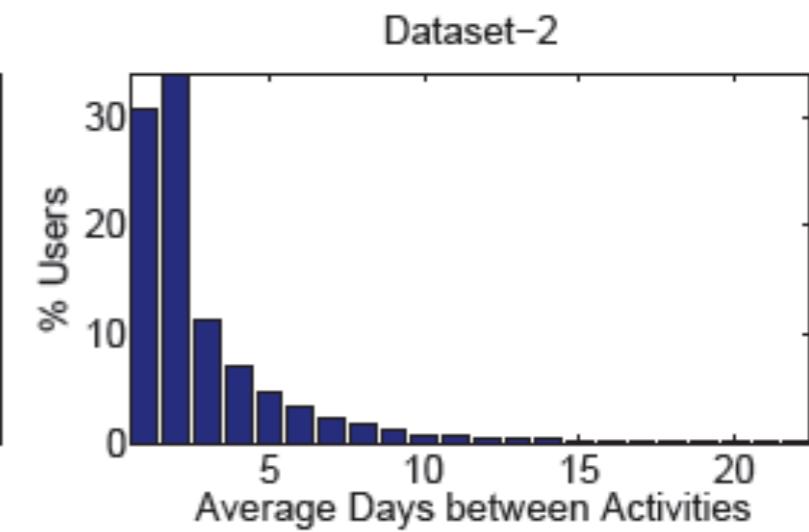
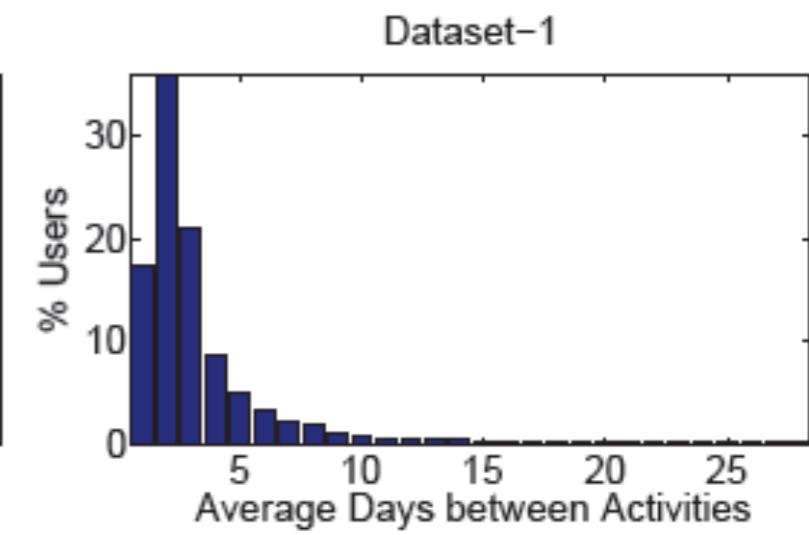
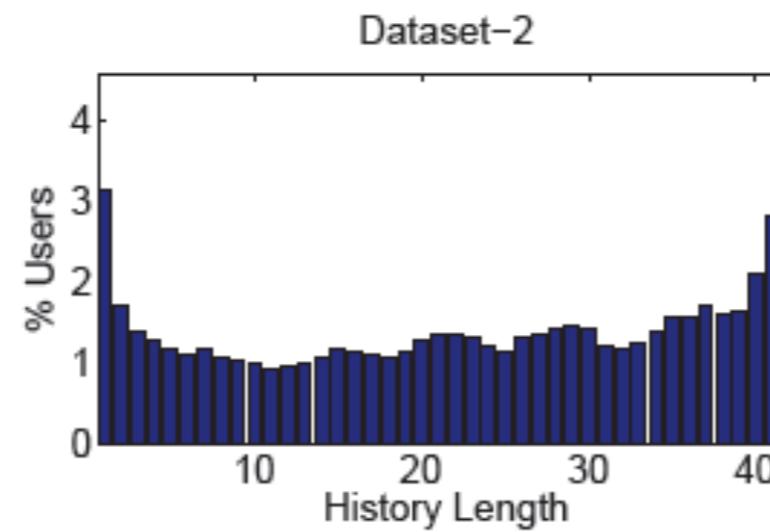
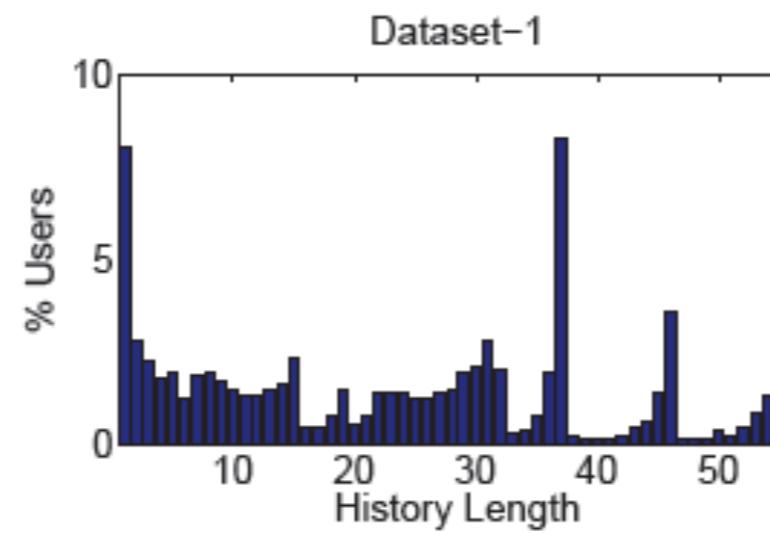
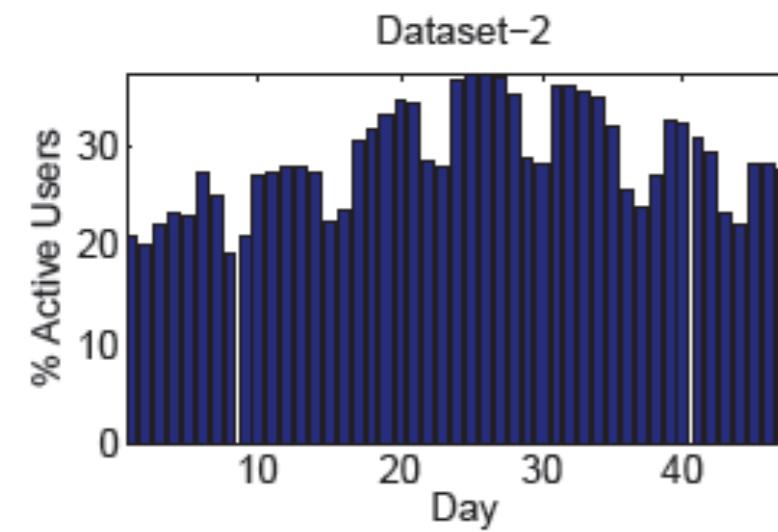
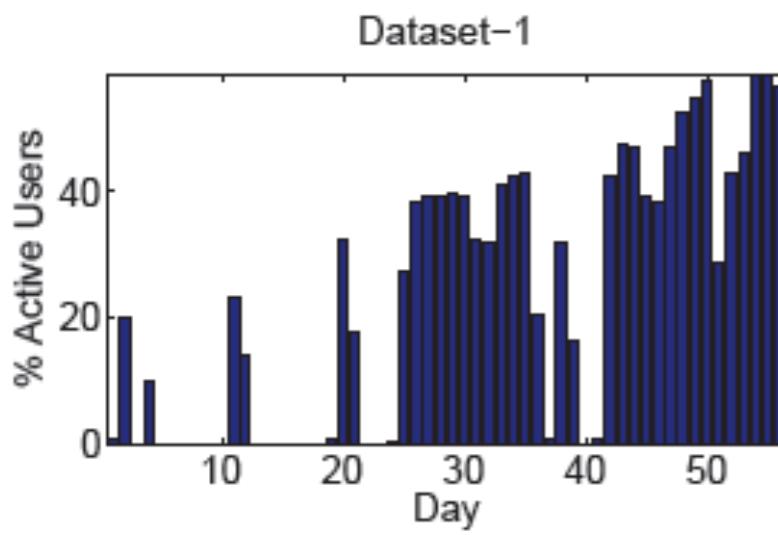
Jobs

job
career
business
assistant
hiring
part-time
receptionist

Finance

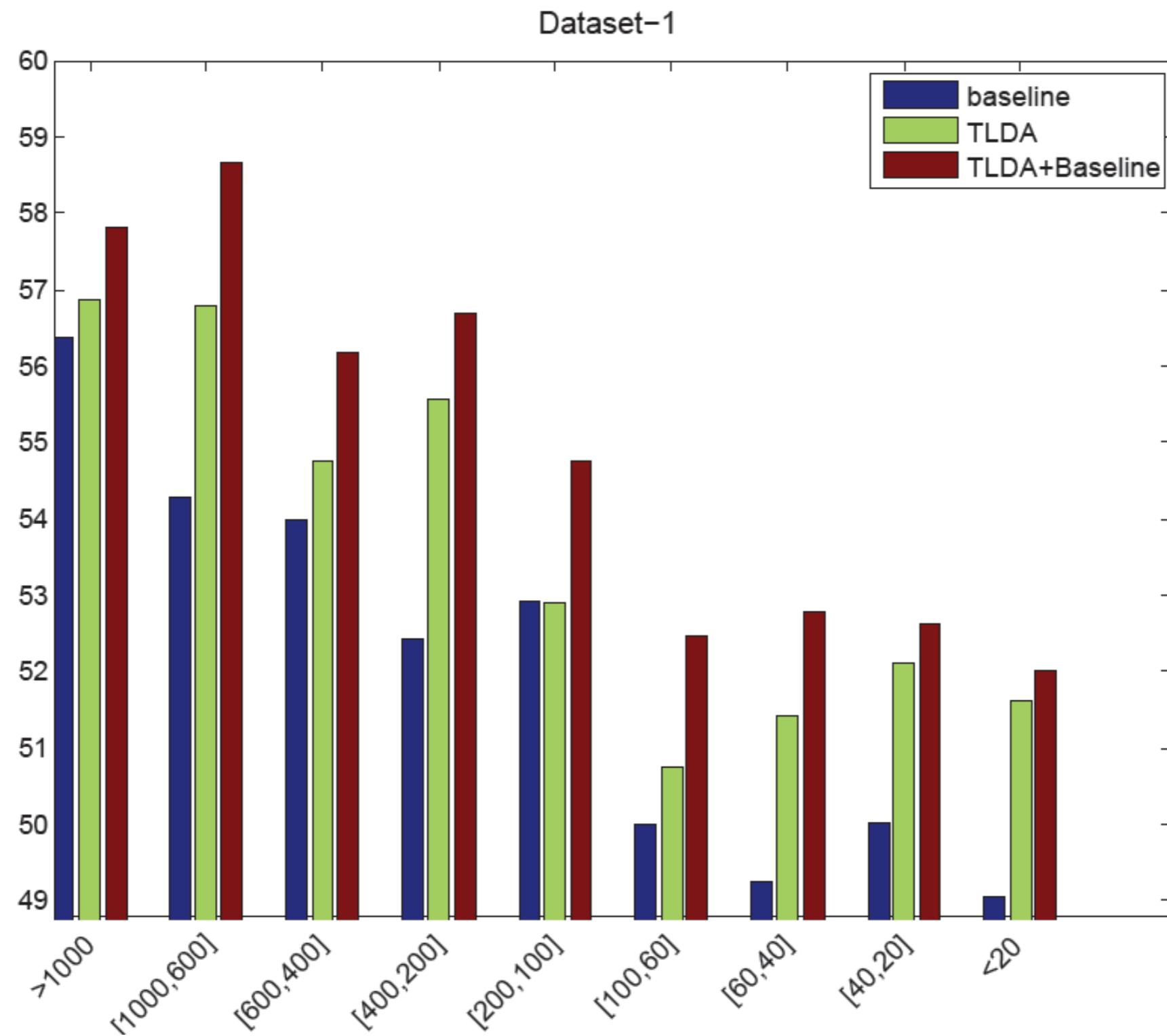
financial
Thomson
chart
real
Stock
Trading
currency

Data



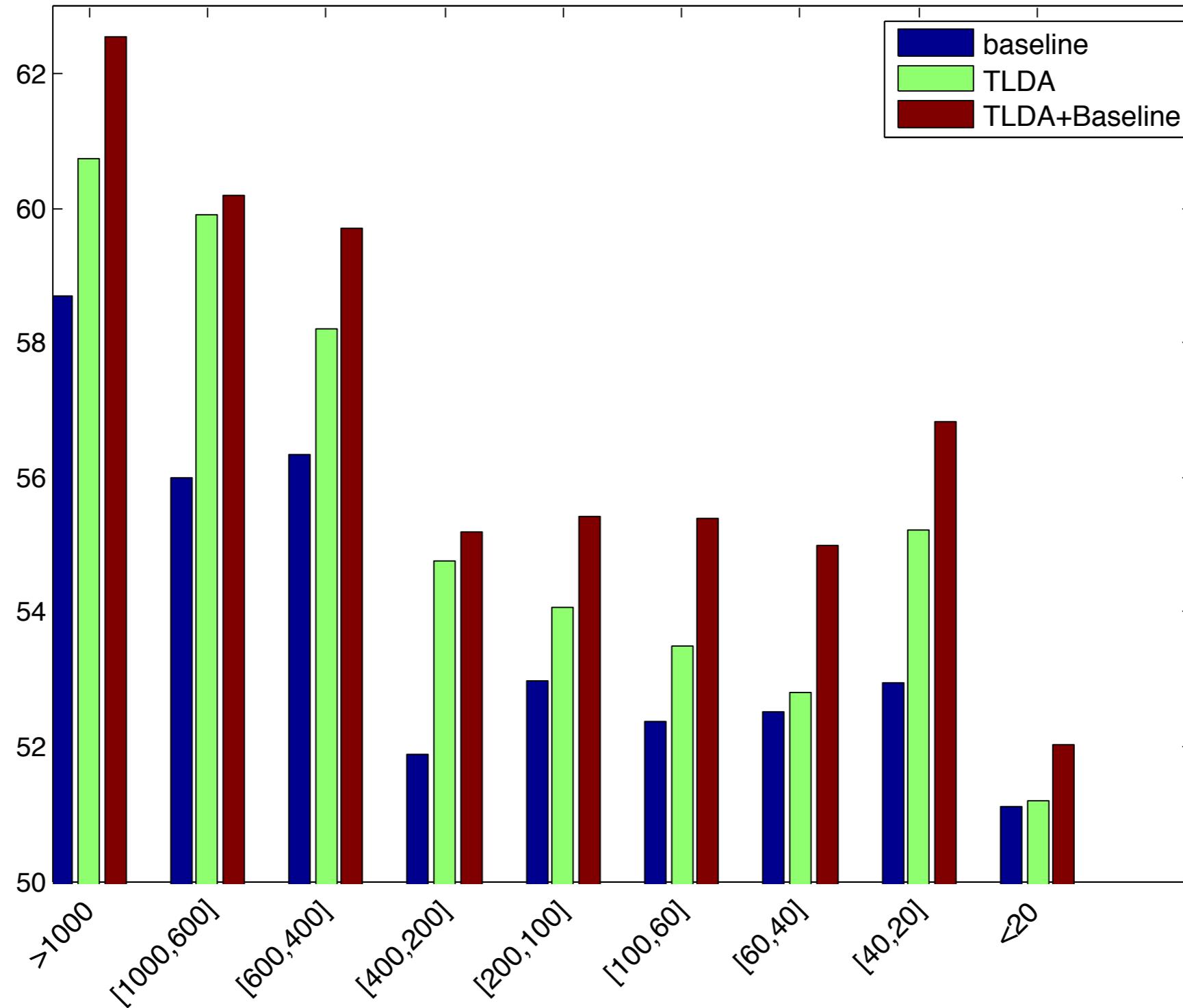
dataset	# days	# users	# campaigns	size
1	56	13.34M	241	242GB
2	44	33.5M	216	435GB

ROC score improvement

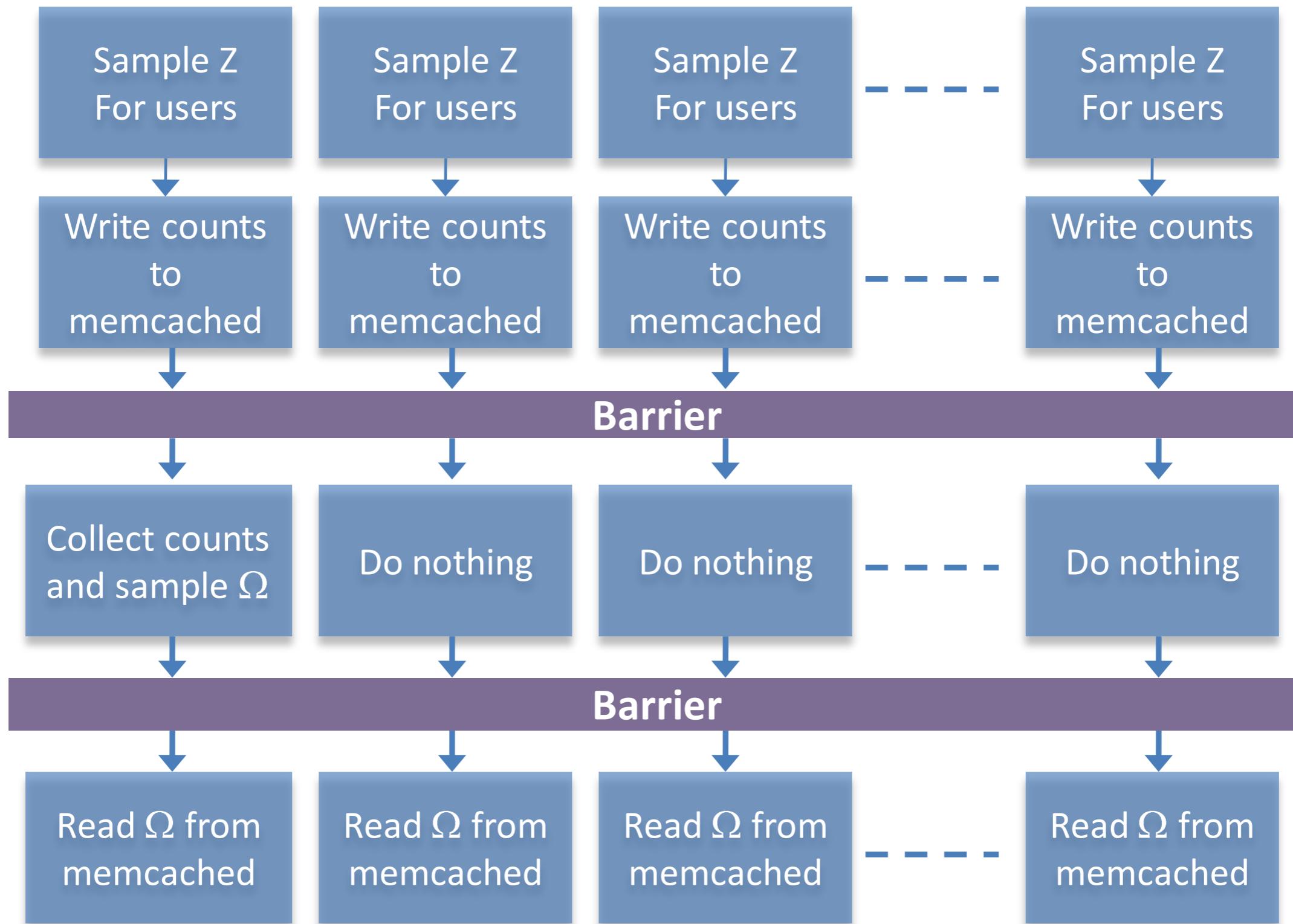


ROC score improvement

Dataset-2



LDA for user profiling



News

News Stream

News Stream



Add-ons turn tax cut bill into 'Christmas tree'

AP - 1 hr 32 mins ago

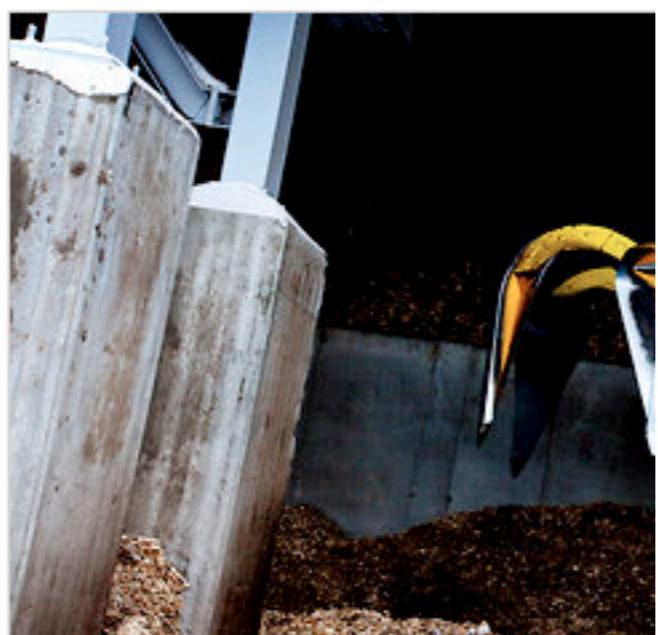
WASHINGTON - In the

... - - - - -

AP

BEYOND FOSSIL FUELS

Using Waste, Swedish



Republicans is becoming a political issue for both candidates and lawmakers. But it's not just Republicans who are worried. Bill Clinton even banned wood stoves in his home. Full Story »

- Video: Gibbs: I Have a Plan
- Slideshow: President Obama's Energy Plan
- Related: Tax fight

News Stream

Suit to Recover Madoff's Money Calls Austrian an Accomplice

By DIANA B. HENRIQUES and PETER LATTMAN

Sonja Kohn, an Austrian banker, is accused of masterminding a 23-year conspiracy that played a central role in financing the gigantic Ponzi scheme.

Post a Comment

er

Print

ovember,
ase food

China says inflation up 5.1 percent

AP Associated Press

b Buzz up! 19 votes | f Share

By CARA ANNA, Associated Press

BEIJING - China's inflation surprised officials said Saturday, despite supplies and end diesel shortages.

The 5.1 percent inflation rate was driven by a 11.7 percent jump in food prices year on year.

The news comes as China's leaders meet for the top economic planning conference of the year and as financial markets watch for a widely anticipated interest rate hike to help bring rapid economic growth to a more sustainable level.

"I think this means that an interest rate hike of 25 basis points is very likely by the end of the year," said CLSA analyst Andy Rothman.



Wall Street Video: Charting Consumer Sentiment CNBC



Wall Street Video: Bright Future TheStreet.com

RELATED QUOTES

^DJI	11,410.32	+40.26
^GSPC	1,240.40	+7.40
^IXIC	2,637.54	+20.87



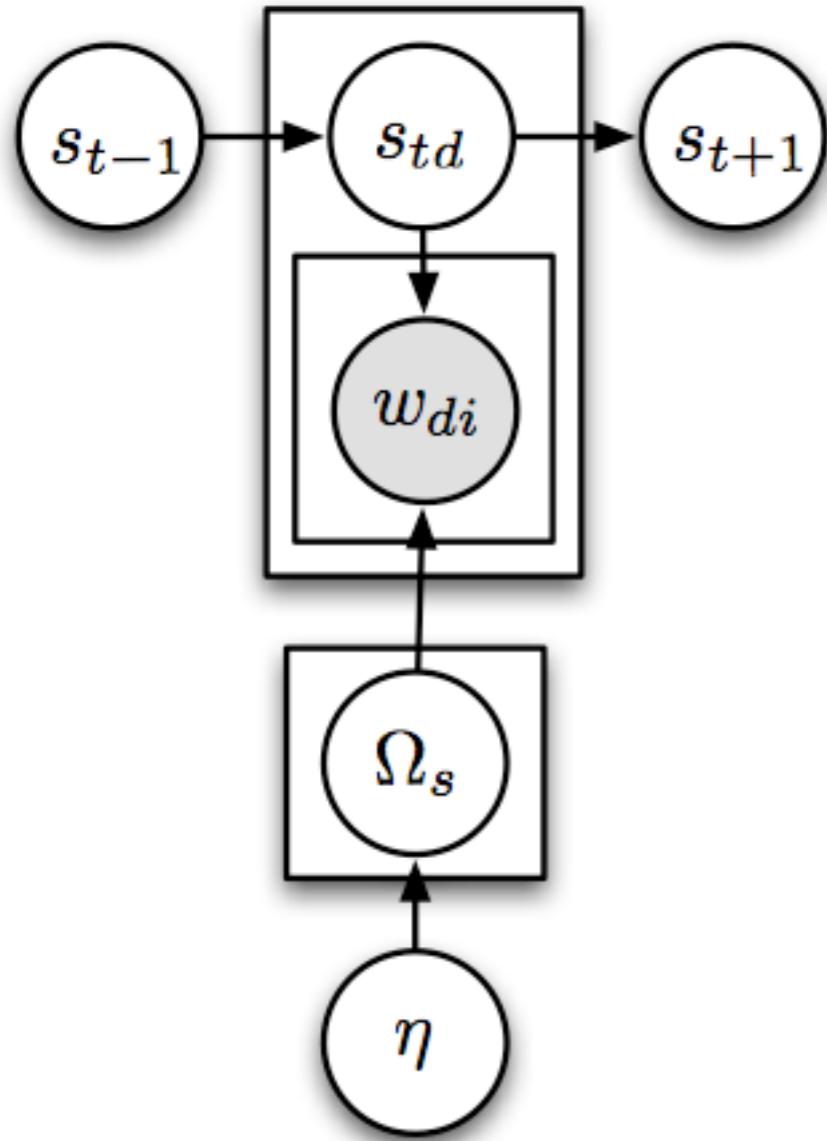
Johan Spanner for The New York Times

As part of its citywide system, Kristianstad burns wood waste like tree prunings and scraps from flooring factories to power an underground district heating grid.

News Stream

- Over 1 high quality news article per second
- Multiple sources (Reuters, AP, CNN, ...)
- Same story from multiple sources
- Stories are related
- Goals
 - Aggregate articles into a storyline
 - Analyze the storyline (topics, entities)

Clustering / RCRP



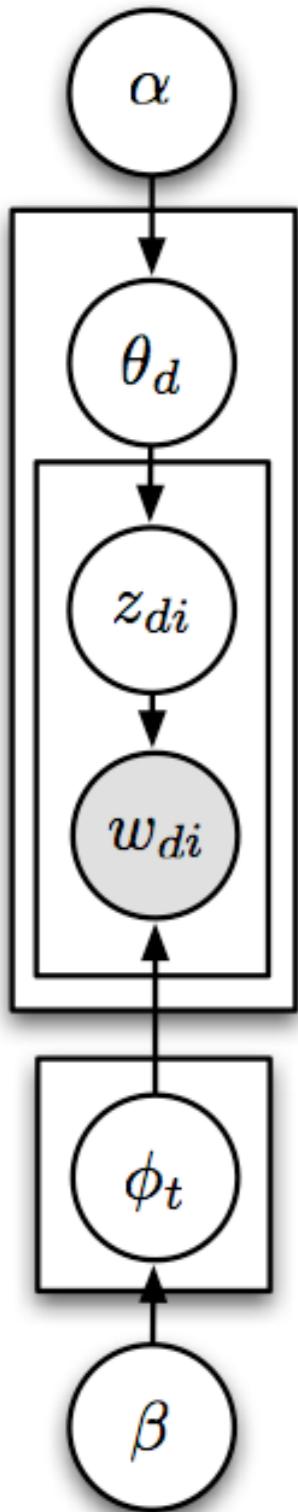
- Assume active story distribution at time t
- Draw story indicator
- Draw words from story distribution
- Down-weight story counts for next day

Ahmed & Xing, 2008

Clustering / RCRP

- Pro
 - Nonparametric model of story generation
(no need to model frequency of stories)
 - No fixed number of stories
 - Efficient inference via collapsed sampler
- Con
 - We learn nothing!
 - No content analysis

Latent Dirichlet Allocation



- Generate topic distribution per article
- Draw topics per word from topic distribution
- Draw words from topic specific word distribution

Blei, Ng, Jordan, 2003

Latent Dirichlet Allocation

- Pro
 - Topical analysis of stories
 - Topical analysis of words (meaning, saliency)
 - More documents improve estimates
- Con
 - No clustering

More Issues

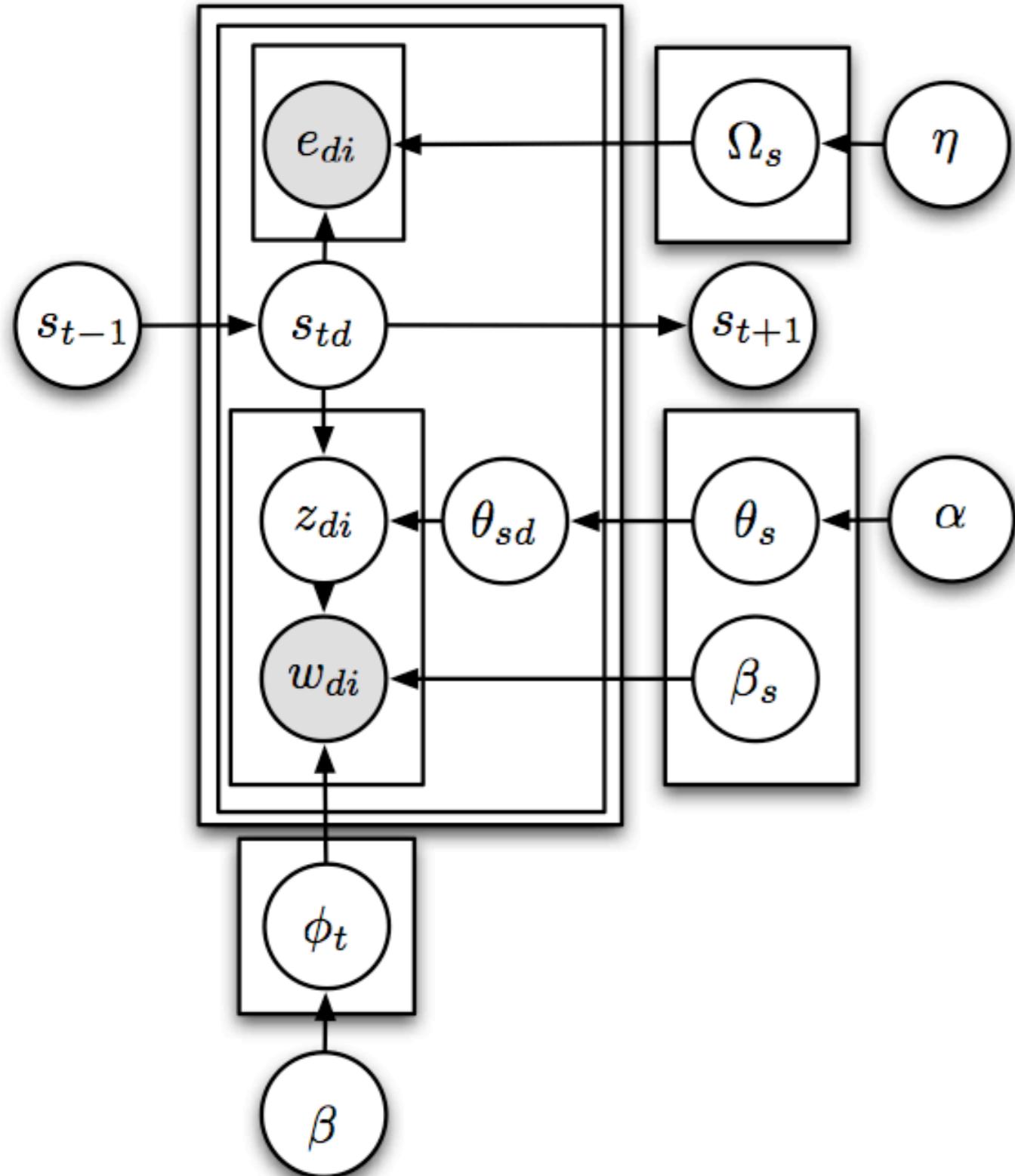


More Issues

- **Named entities are special, topics less**
(e.g. Tiger Woods and his mistresses)
- **Some stories are strange**
(topical mixture is not enough - dirty models)
- **Articles deviate from general story**
(Hierarchical DP)

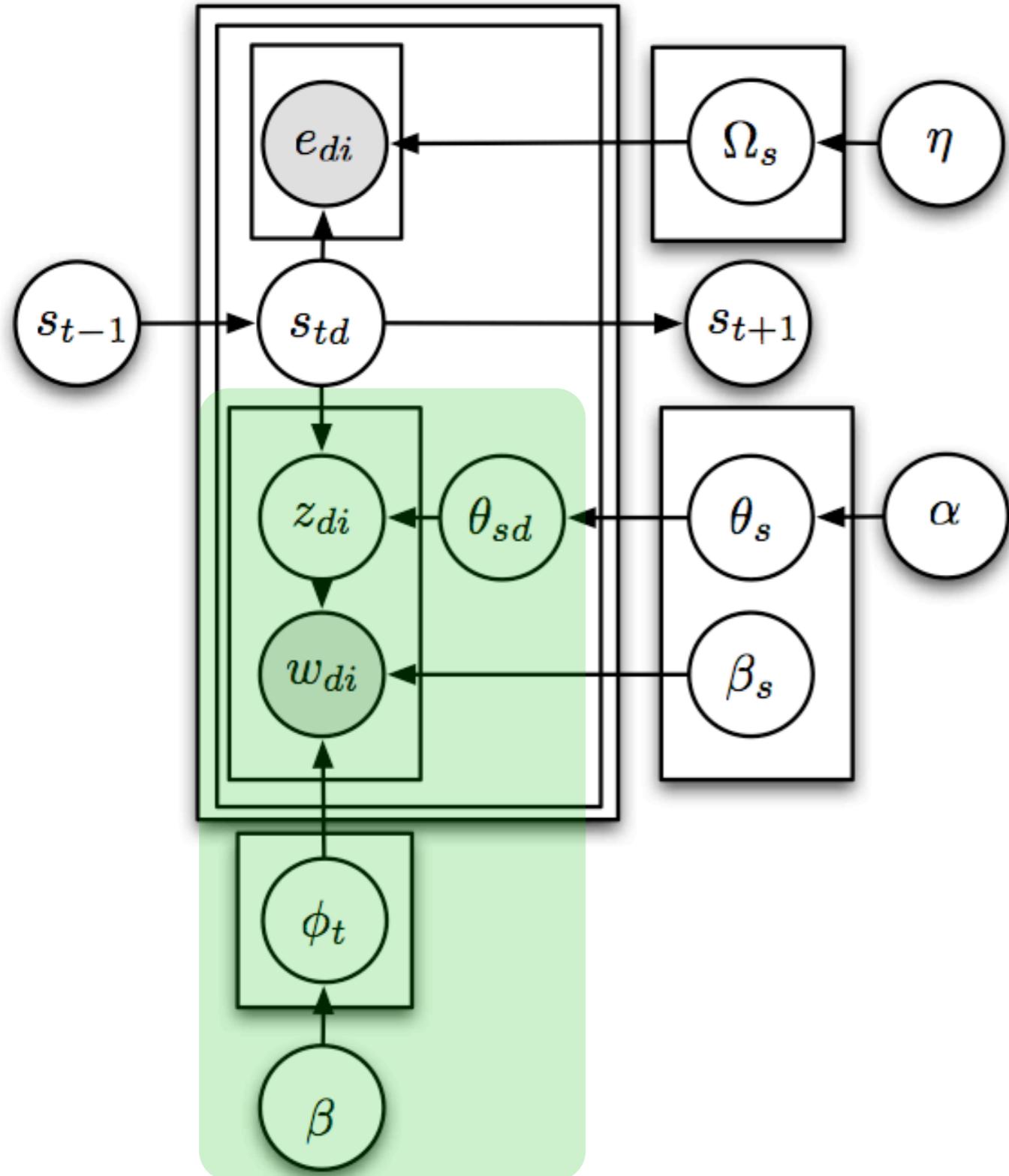
Storylines

Storylines Model



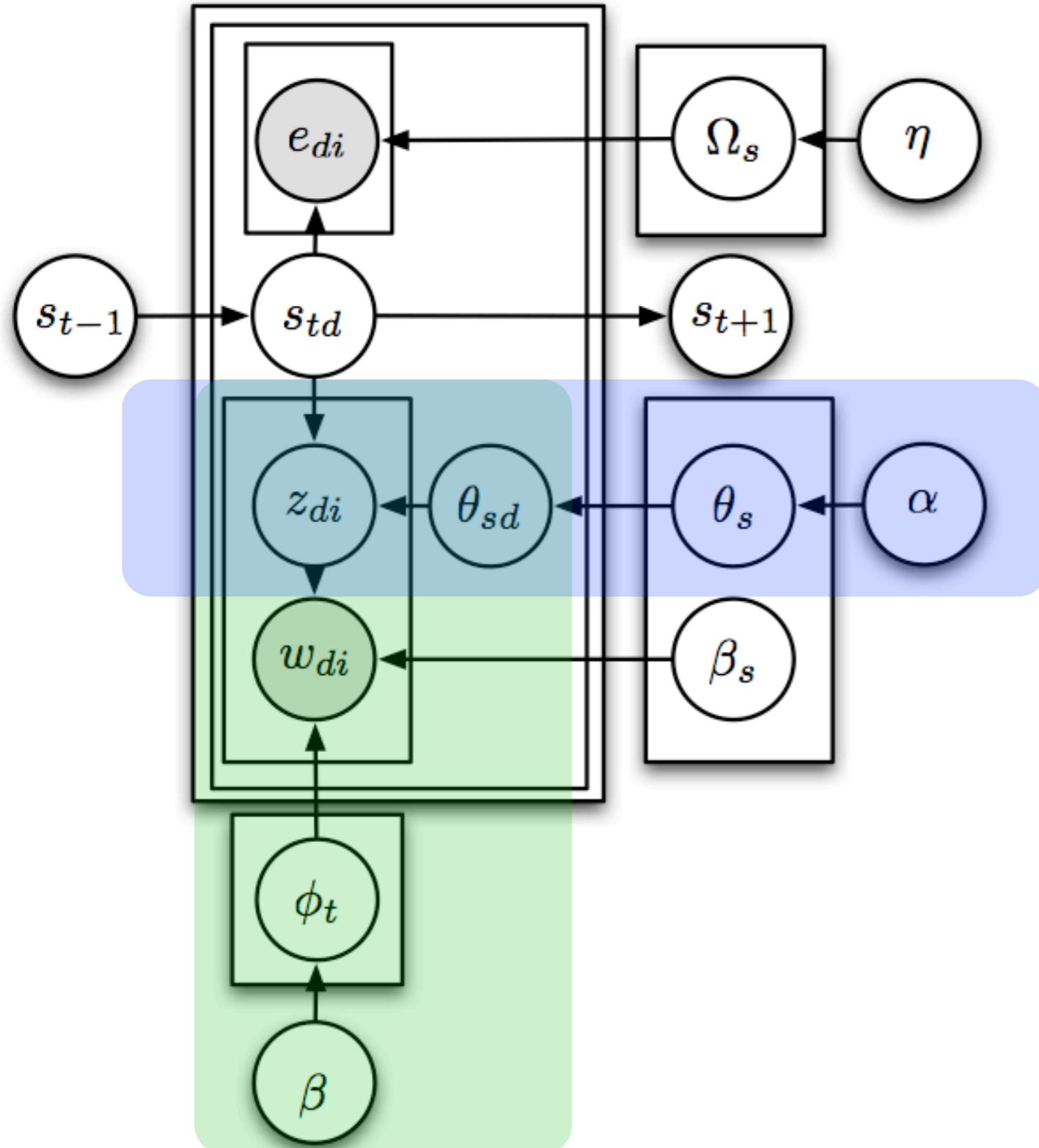
- Topic model
- Topics per cluster
- RCRP for cluster
- Hierarchical DP for article
- Separate model for named entities
- Story specific correction

Storylines Model



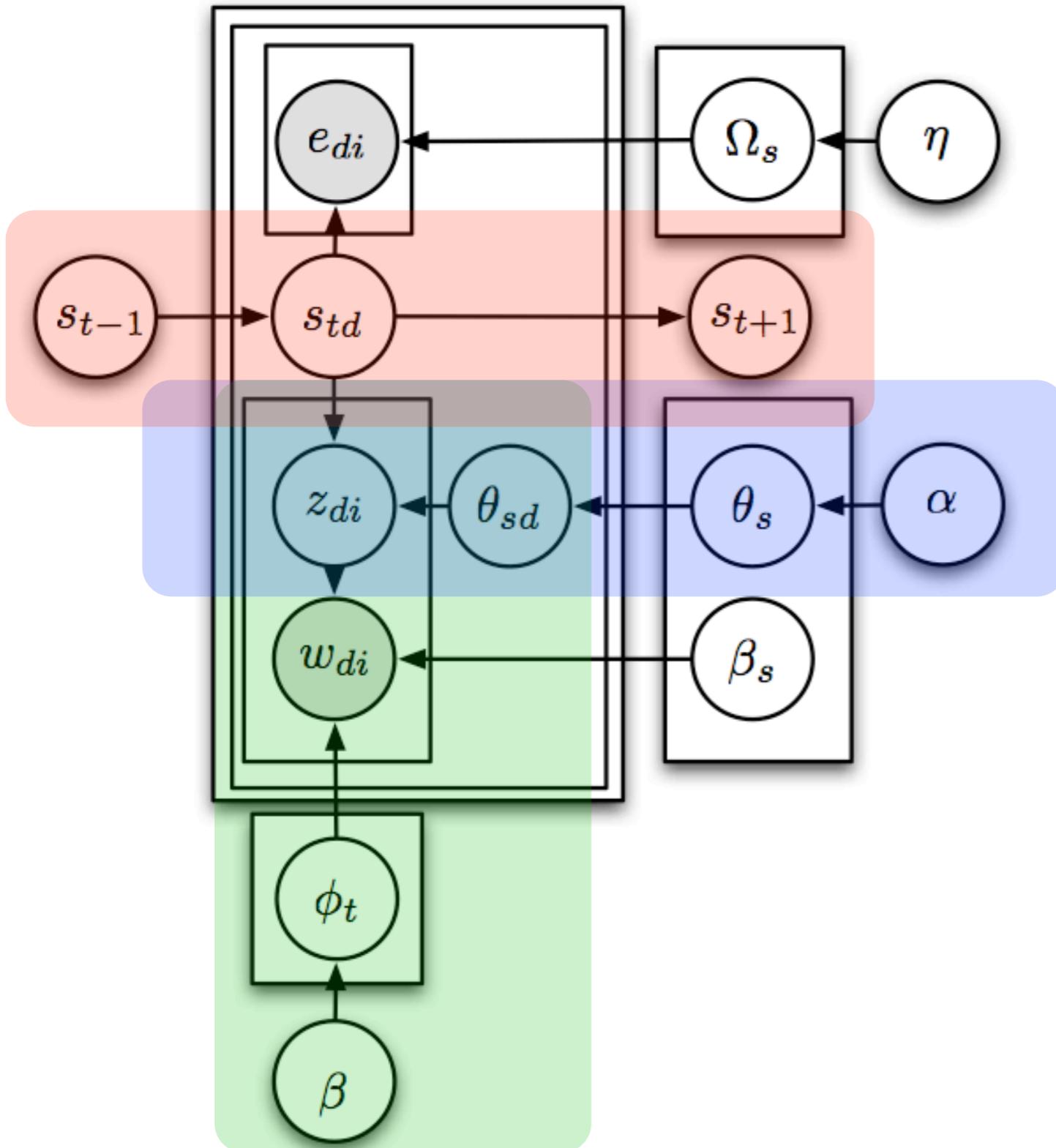
- Topic model
- Topics per cluster
- RCRP for cluster
- Hierarchical DP for article
- Separate model for named entities
- Story specific correction

Storylines Model



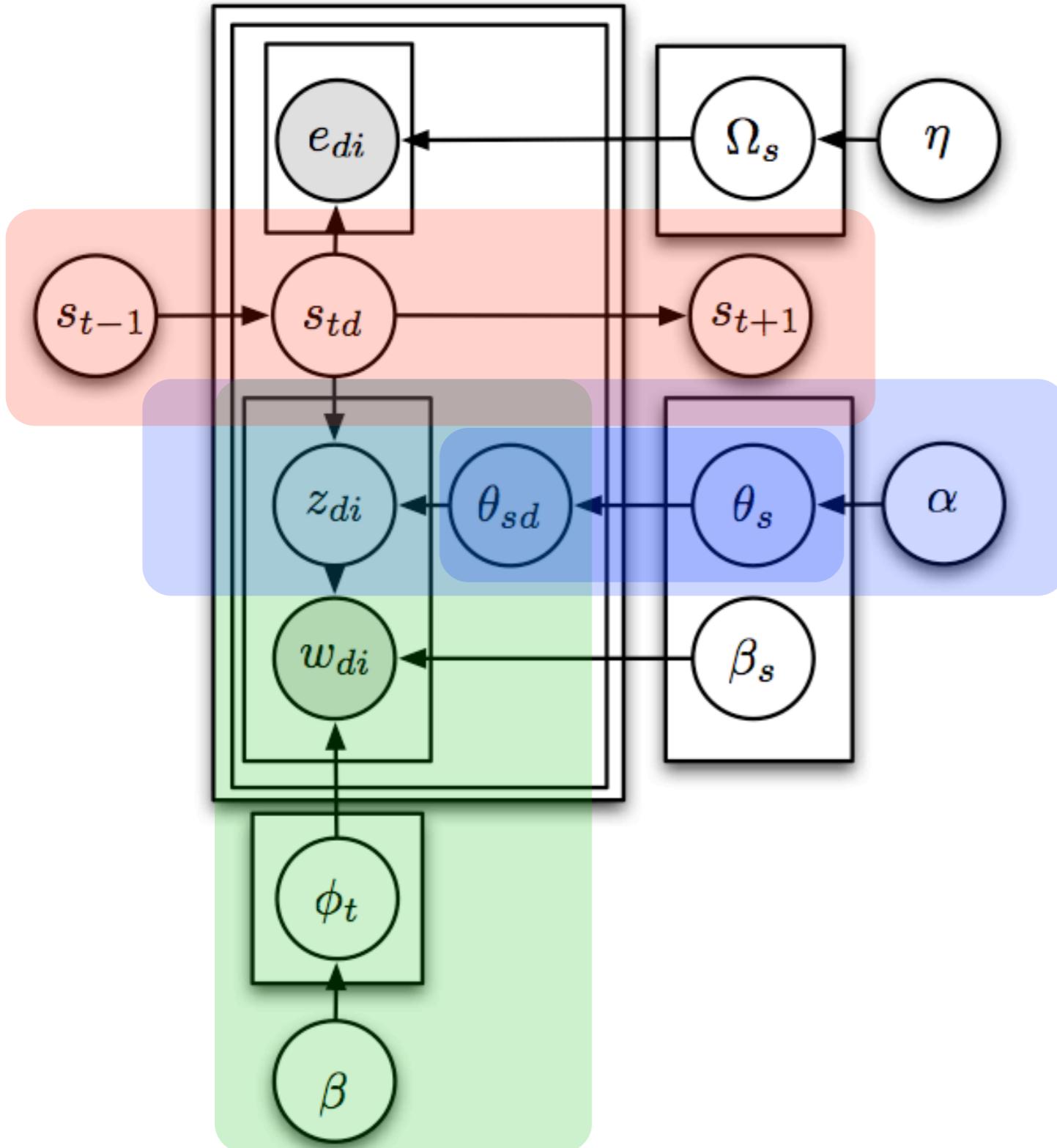
- Topic model
- Topics per cluster
- RCRP for cluster
- Hierarchical DP for article
- Separate model for named entities
- Story specific correction

Storylines Model



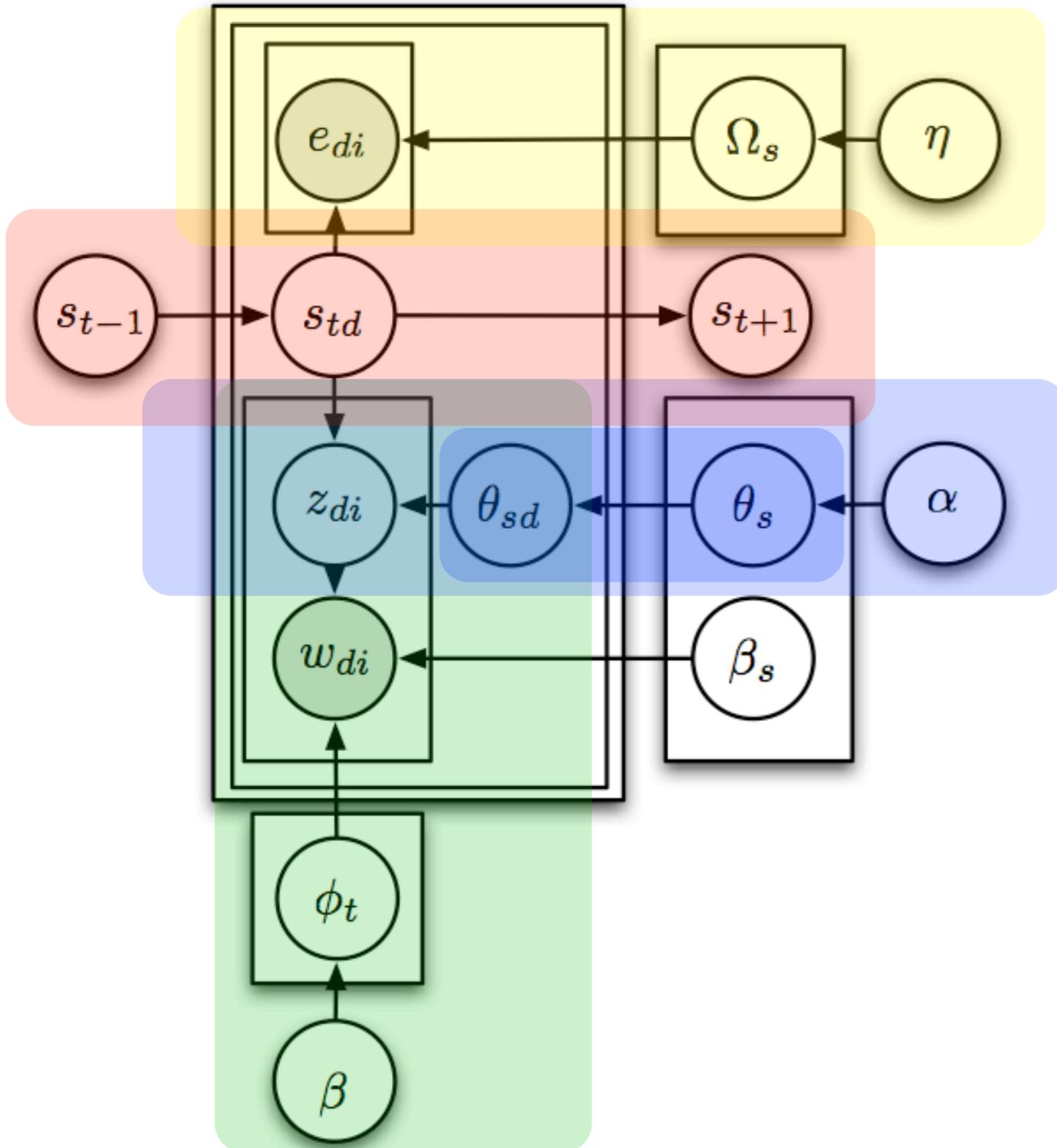
- Topic model
- Topics per cluster
- RCRP for cluster
- Hierarchical DP for article
- Separate model for named entities
- Story specific correction

Storylines Model



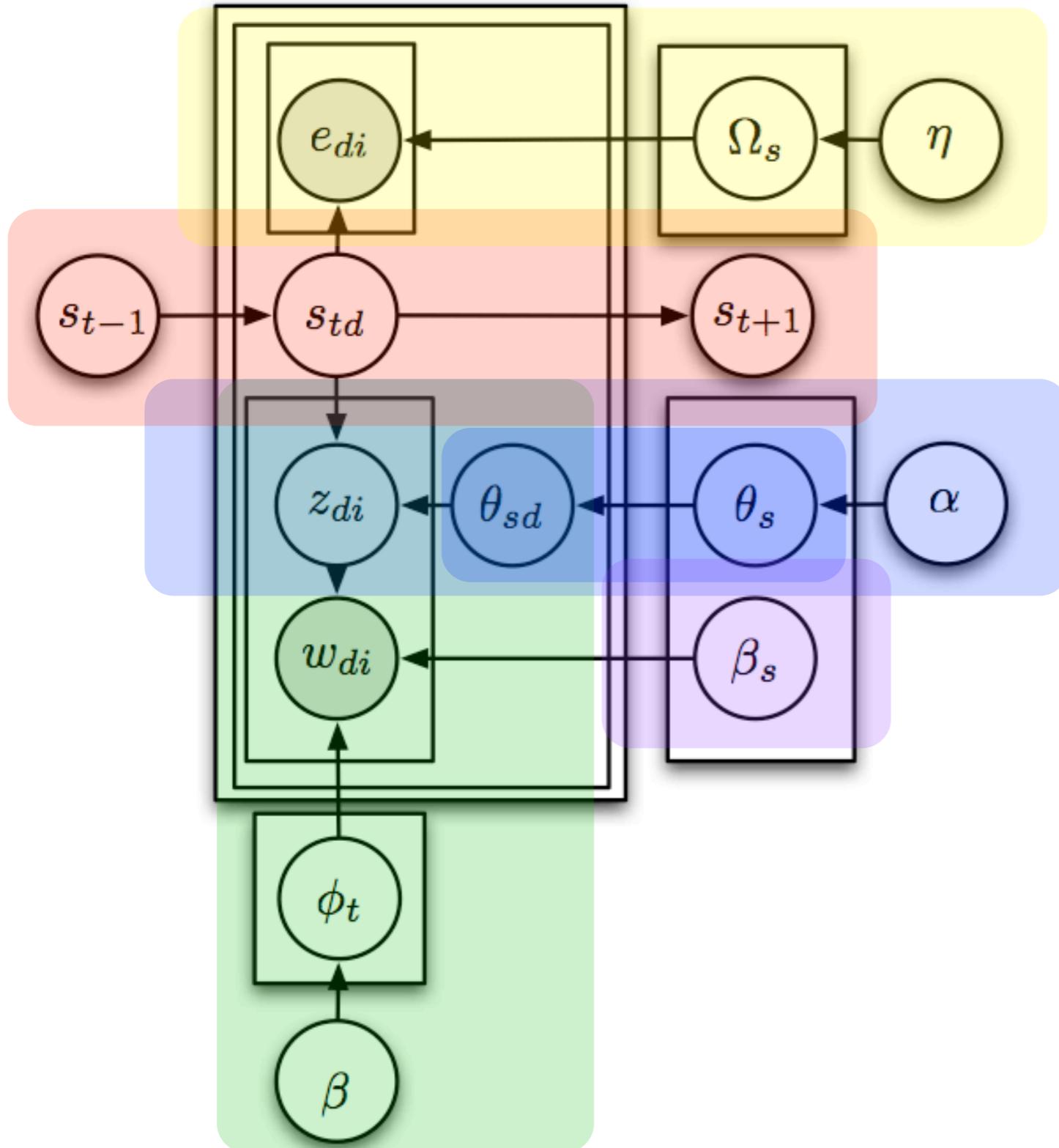
- Topic model
- Topics per cluster
- RCRP for cluster
- Hierarchical DP for article
- Separate model for named entities
- Story specific correction

Storylines Model



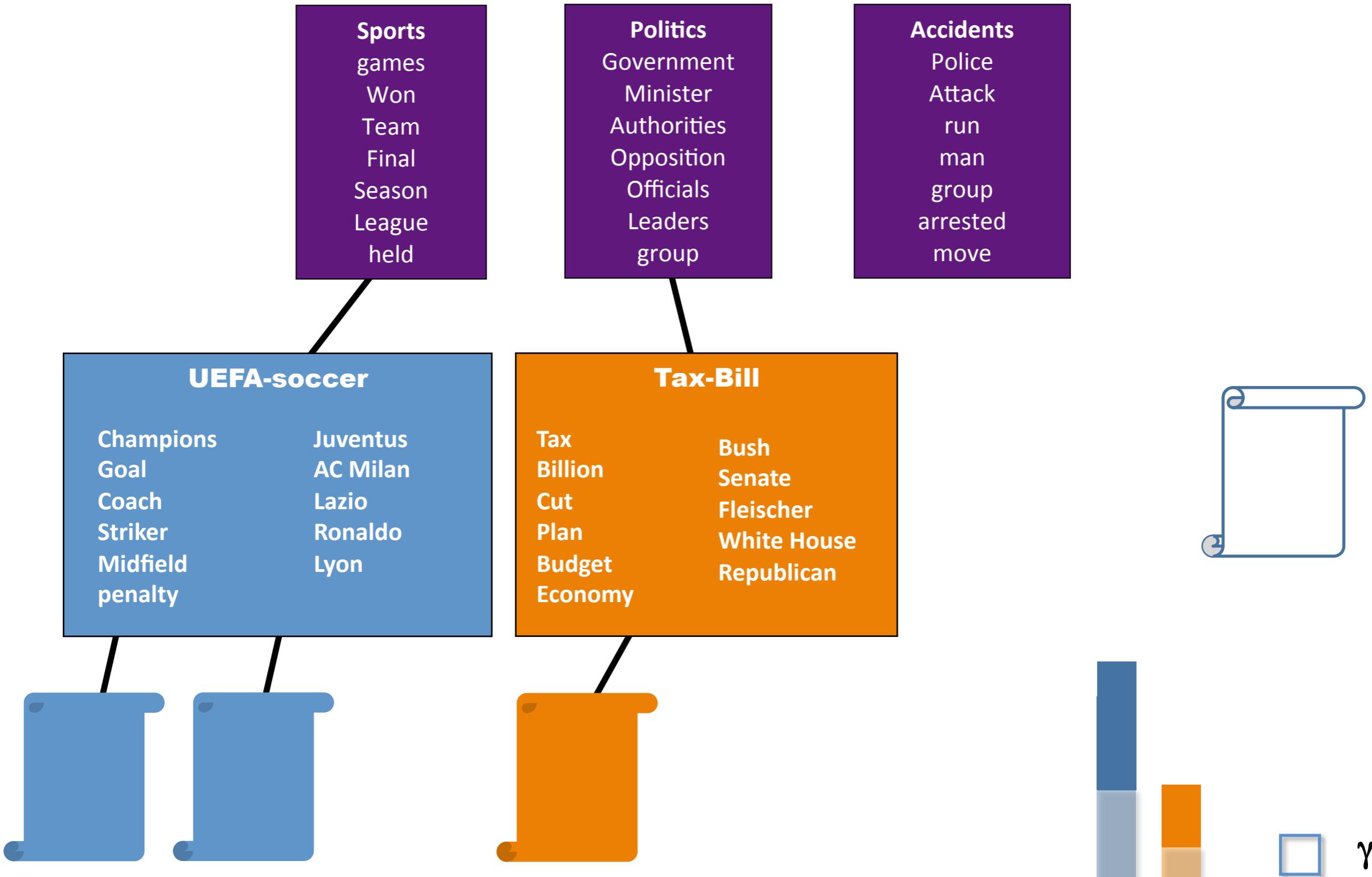
- Topic model
- Topics per cluster
- RCRP for cluster
- Hierarchical DP for article
- Separate model for named entities
- Story specific correction

Storylines Model

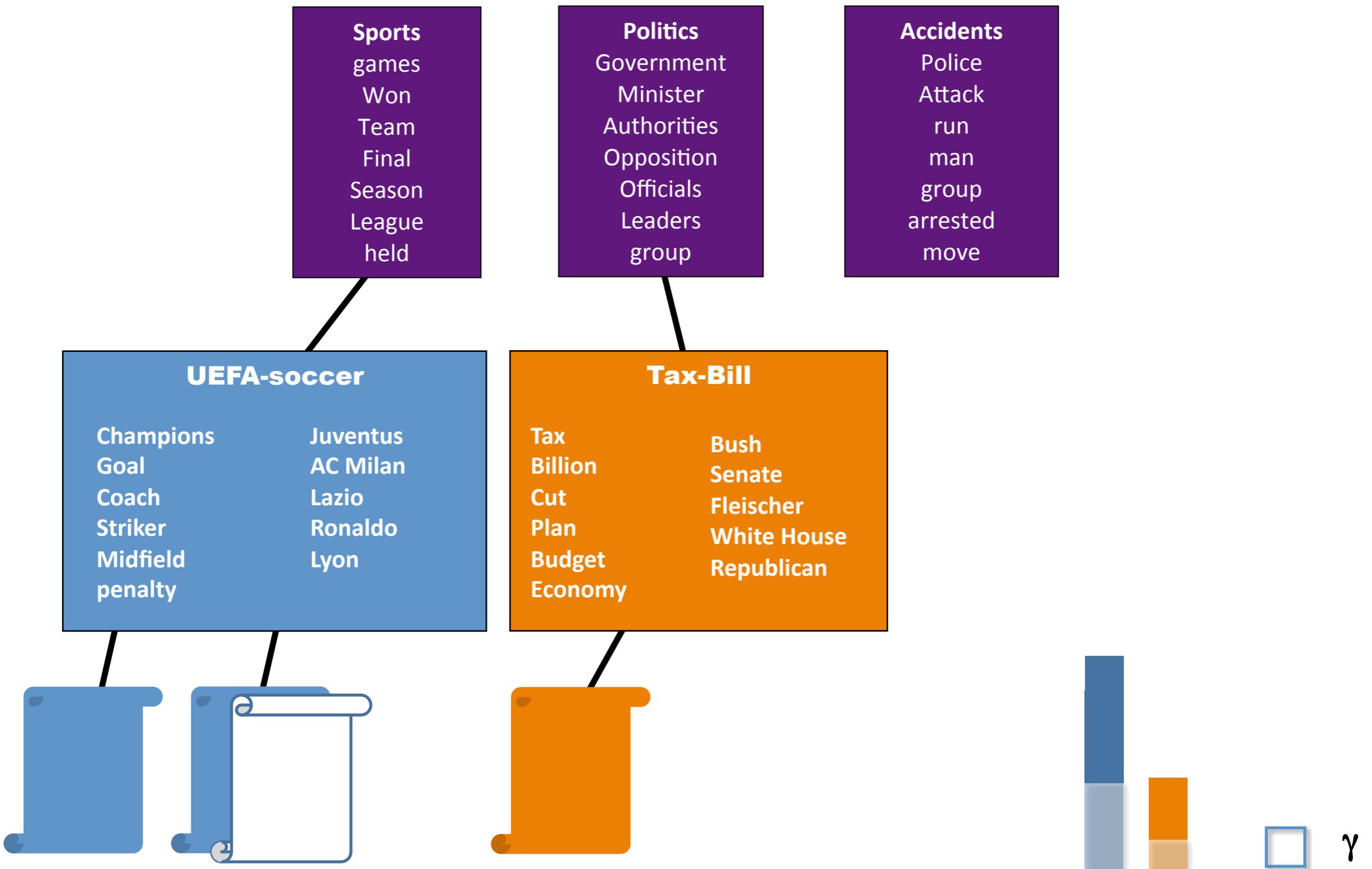


- Topic model
- Topics per cluster
- RCRP for cluster
- Hierarchical DP for article
- Separate model for named entities
- Story specific correction

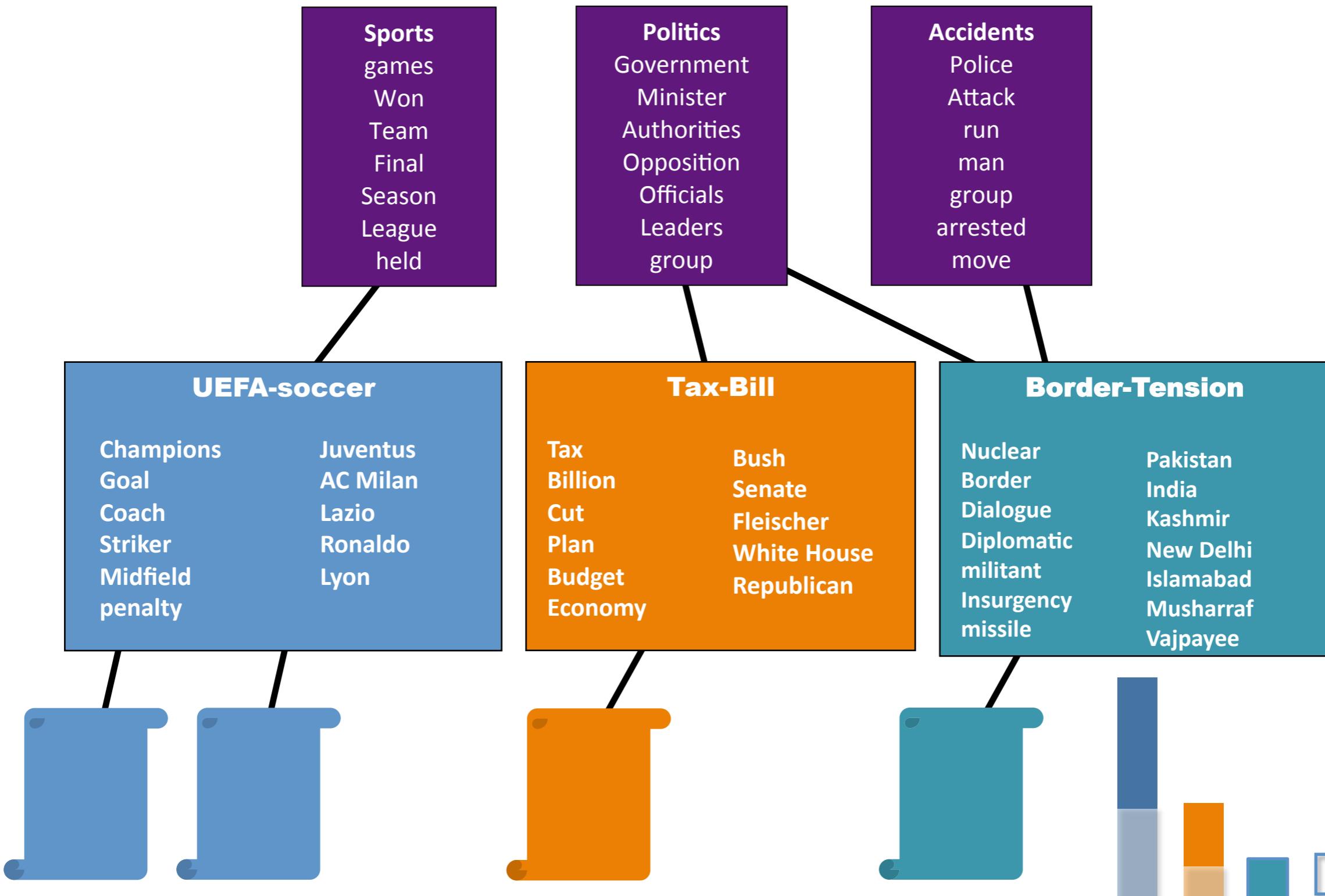
Dynamic Cluster-Topic Hybrid



Dynamic Cluster-Topic Hybrid



Dynamic Cluster-Topic Hybrid



Inference

- We receive articles as a stream
Want topics & stories now
- Variational inference infeasible
(RCRP, sparse to dense, vocabulary size)
- We have a ‘tracking problem’
 - Sequential Monte Carlo
 - Use sampled variables of surviving particle
 - Use ideas from Cannini et al. 2009

Particle Filter

- Proposal distribution - draw stories s , topics z

$$p(s_{t+1}, z_{t+1} | x_{1\dots t+1}, s_{1\dots t}, z_{1\dots t})$$

using Gibbs Sampling for each particle

- Reweight particle via

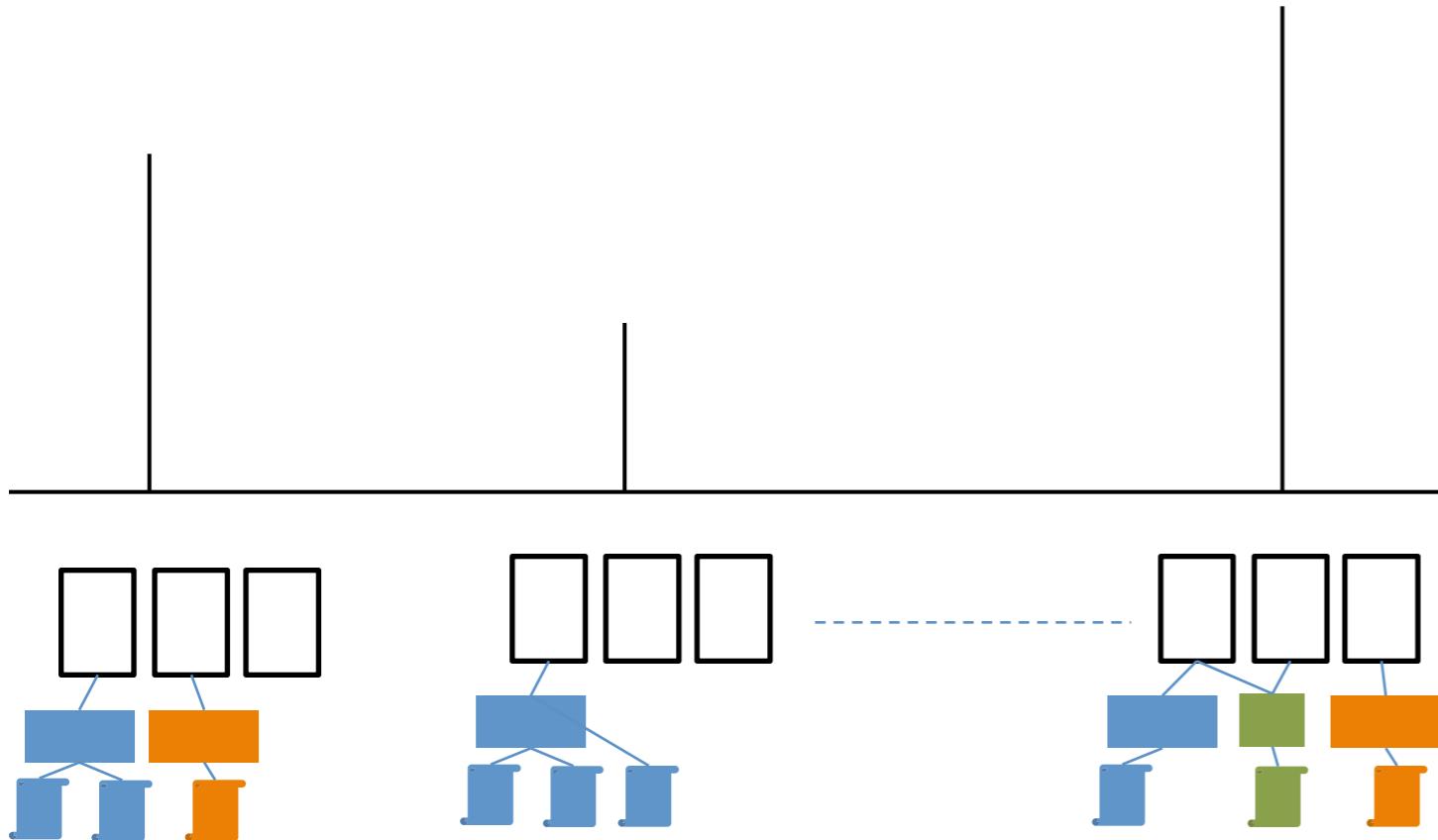
new data

past state

$$p(x_{t+1} | x_{1\dots t}, s_{1\dots t}, z_{1\dots t})$$

- Resample particles if ℓ_2 norm too large
(resample some assignments for diversity, too)
- Compare to multiplicative updates algorithm
In our case predictive likelihood yields weights

Particle Filter



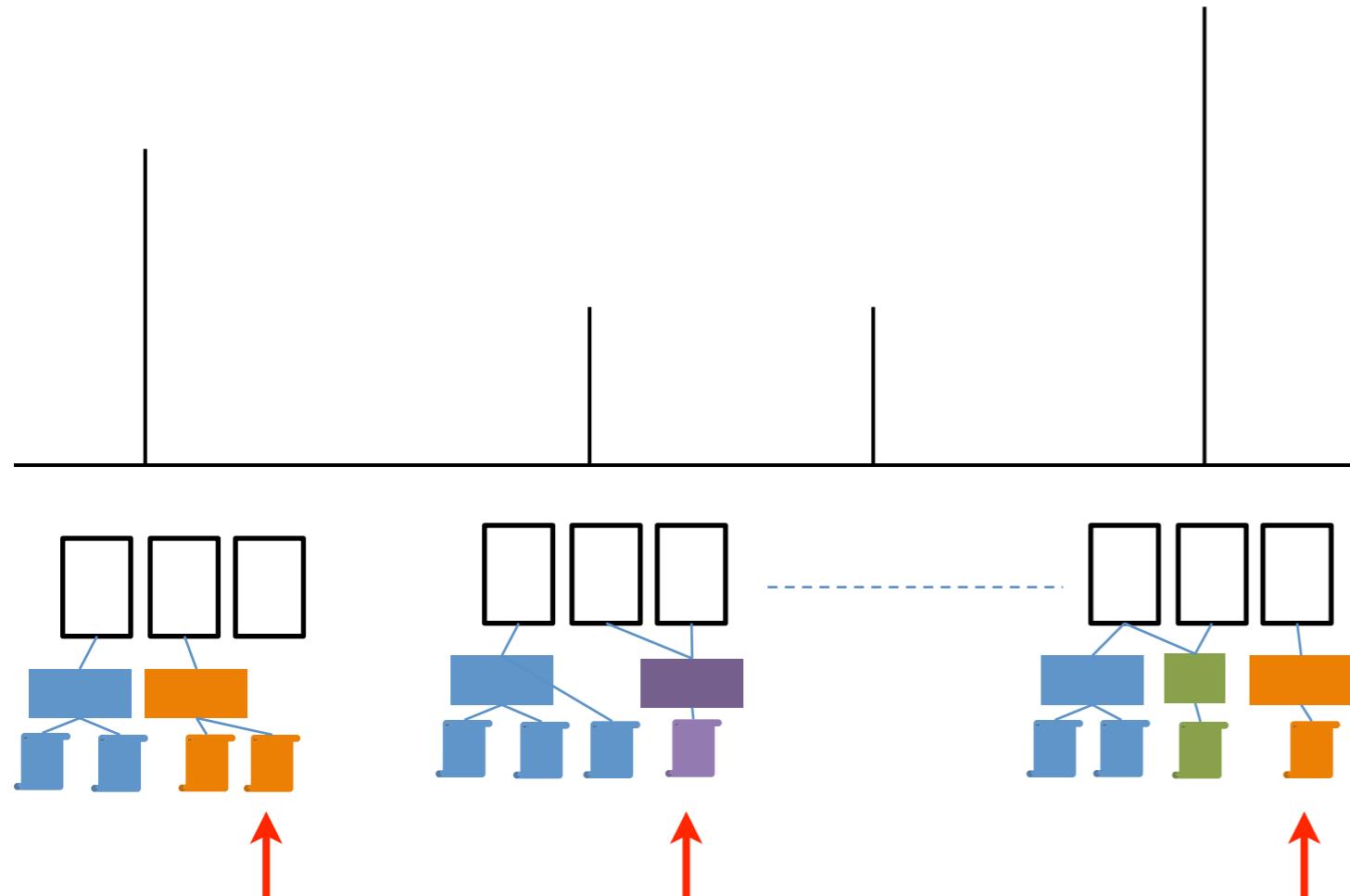
Algorithm 1 A Particle Filter Algorithm

```

Initialize  $\omega_1^f$  to  $\frac{1}{F}$  for all  $f \in \{1, \dots, F\}$ 
for each document  $d$  with time stamp  $t$  do
  for  $f \in \{1, \dots, F\}$  do
    Sample  $s_{td}^f, z_{td}^f$  using MCMC
     $\omega^f \leftarrow \omega^f P(x_{td}|z_{td}^f, s_{td}^f, x_{1:t,d-1})$ 
  end for
  Normalize particle weights
  if  $\|\omega_t\|_2^{-2} < \text{threshold}$  then
    resample particles
    for  $f \in \{1, \dots, F\}$  do
      MCMC pass over 10 random past documents
    end for
  end if
end for
  
```

- s and z are tightly coupled
- Alternative to MCMC
 - Sample s then sample z (high variance)
 - Sample z then sample s (doesn't make sense)
- Idea (following a similar trick by Jain and Neal)
 - Run a few iterations of MCMC over s and z
 - Take last sample as the proposed value

Particle Filter



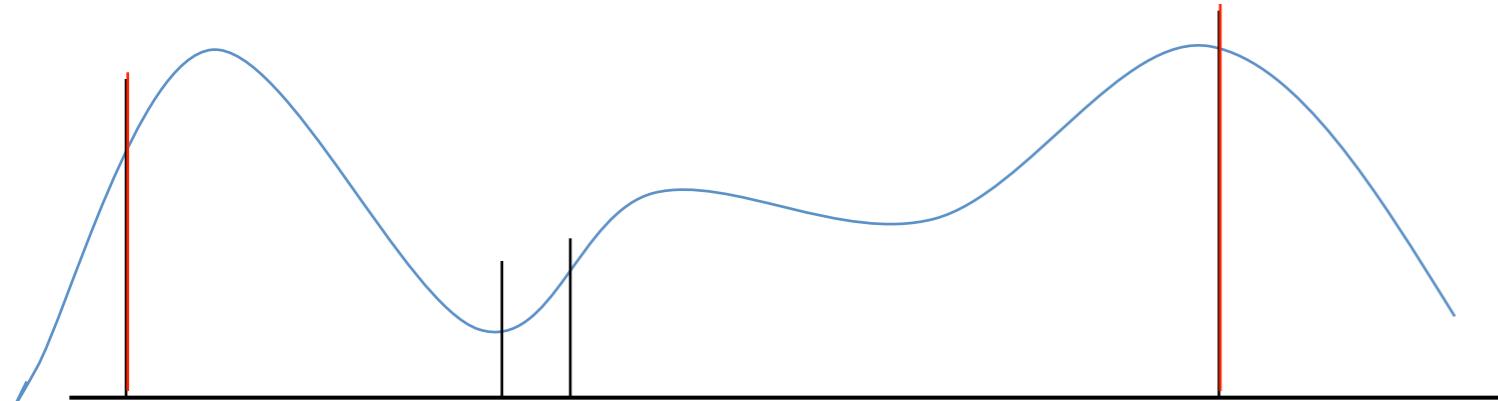
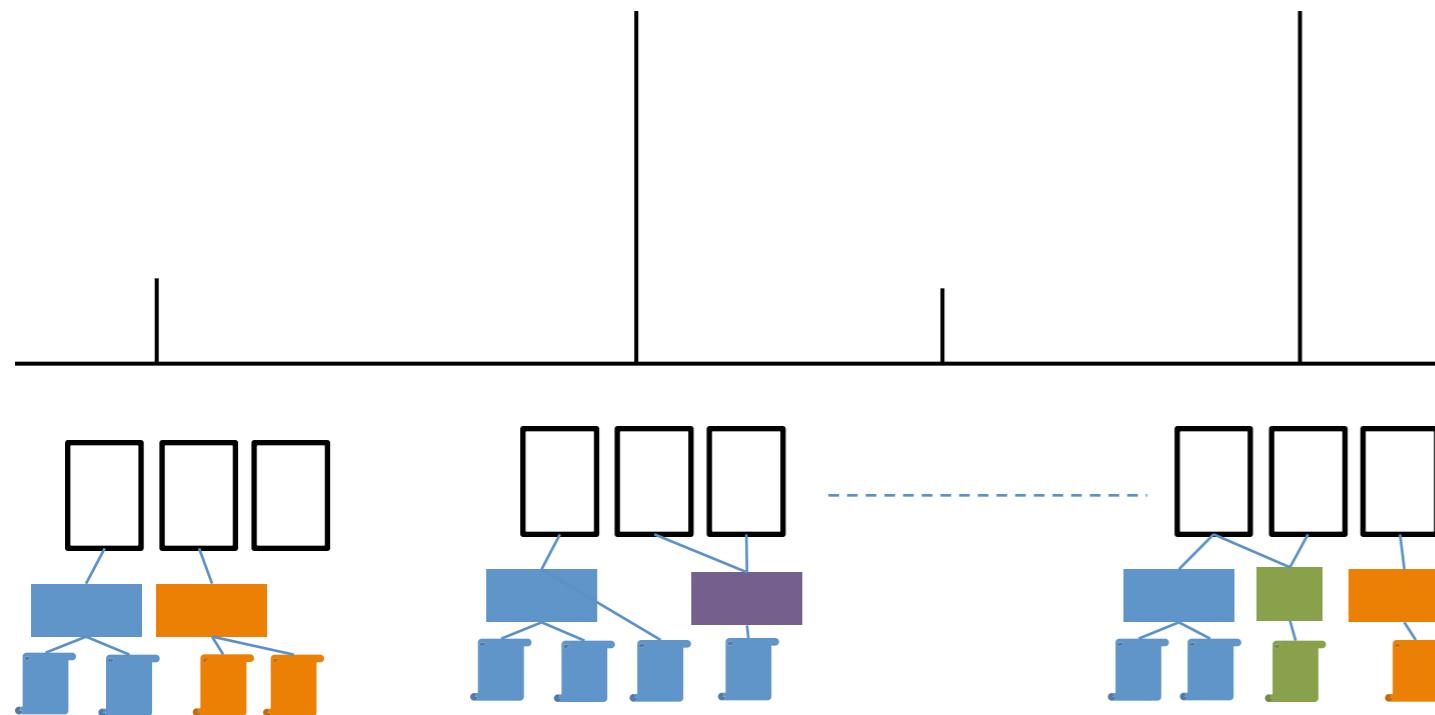
Algorithm 1 A Particle Filter Algorithm

```
Initialize  $\omega_1^f$  to  $\frac{1}{F}$  for all  $f \in \{1, \dots, F\}$ 
for each document  $d$  with time stamp  $t$  do
    for  $f \in \{1, \dots, F\}$  do
        Sample  $s_{td}^f, z_{td}^f$  using MCMC
         $\omega^f \leftarrow \omega^f P(x_{td}|z_{td}^f, s_{td}^f, x_{1:t,d-1})$ 
    end for
    Normalize particle weights
    if  $\|\omega_t\|_2^{-2} < \text{threshold}$  then
        resample particles
    for  $f \in \{1, \dots, F\}$  do
        MCMC pass over 10 random past documents
    end for
    end if
end for
```

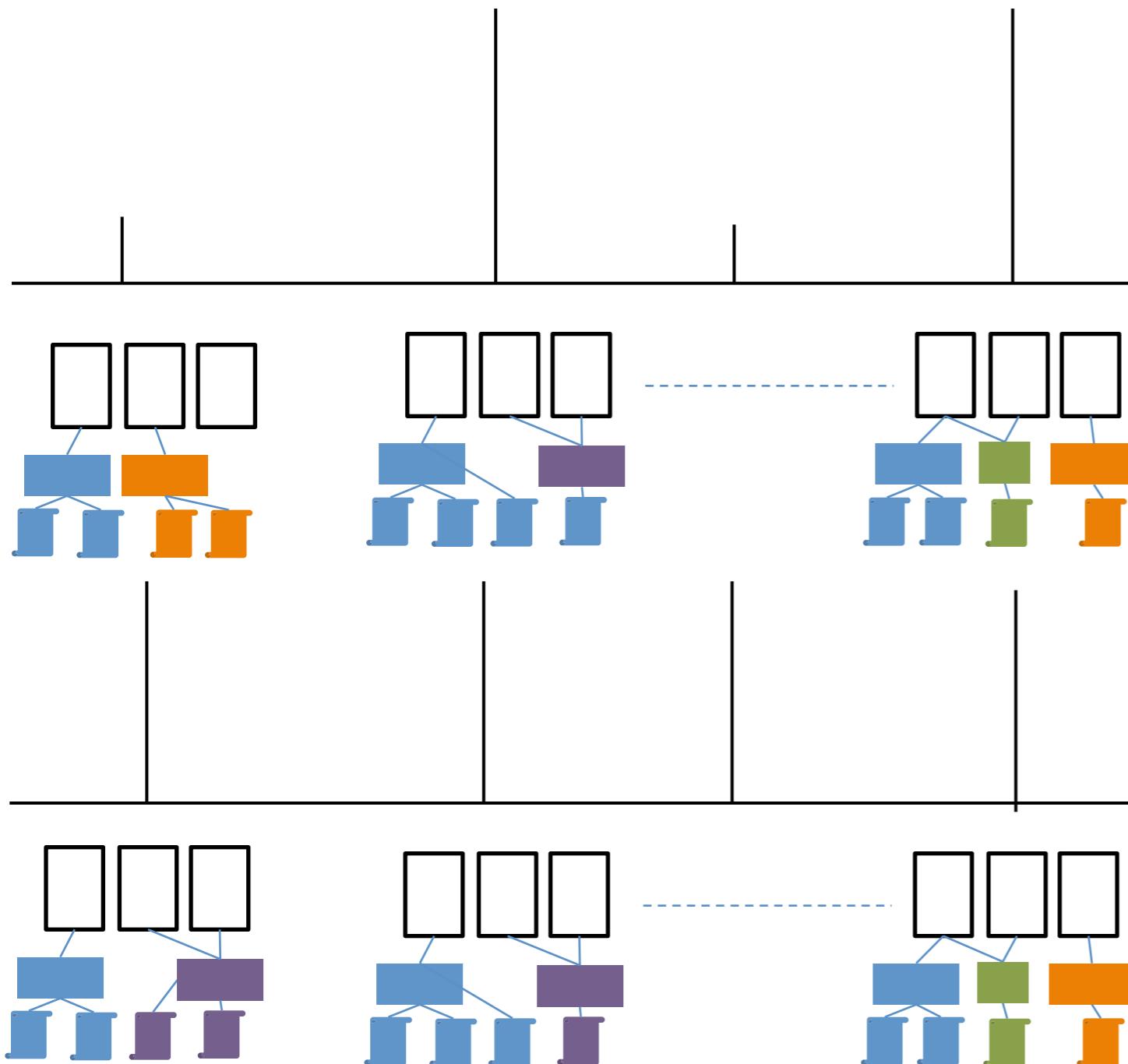
Particle Filter

Algorithm 1 A Particle Filter Algorithm

```
Initialize  $\omega_1^f$  to  $\frac{1}{F}$  for all  $f \in \{1, \dots, F\}$ 
for each document  $d$  with time stamp  $t$  do
    for  $f \in \{1, \dots, F\}$  do
        Sample  $s_{td}^f, z_{td}^f$  using MCMC
         $\omega^f \leftarrow \omega^f P(x_{td}|z_{td}^f, s_{td}^f, x_{1:t,d-1})$ 
    end for
    Normalize particle weights
    if  $\|\omega_t\|_2^{-2} < \text{threshold}$  then
        resample particles
        for  $f \in \{1, \dots, F\}$  do
            MCMC pass over 10 random past documents
        end for
    end if
end for
```



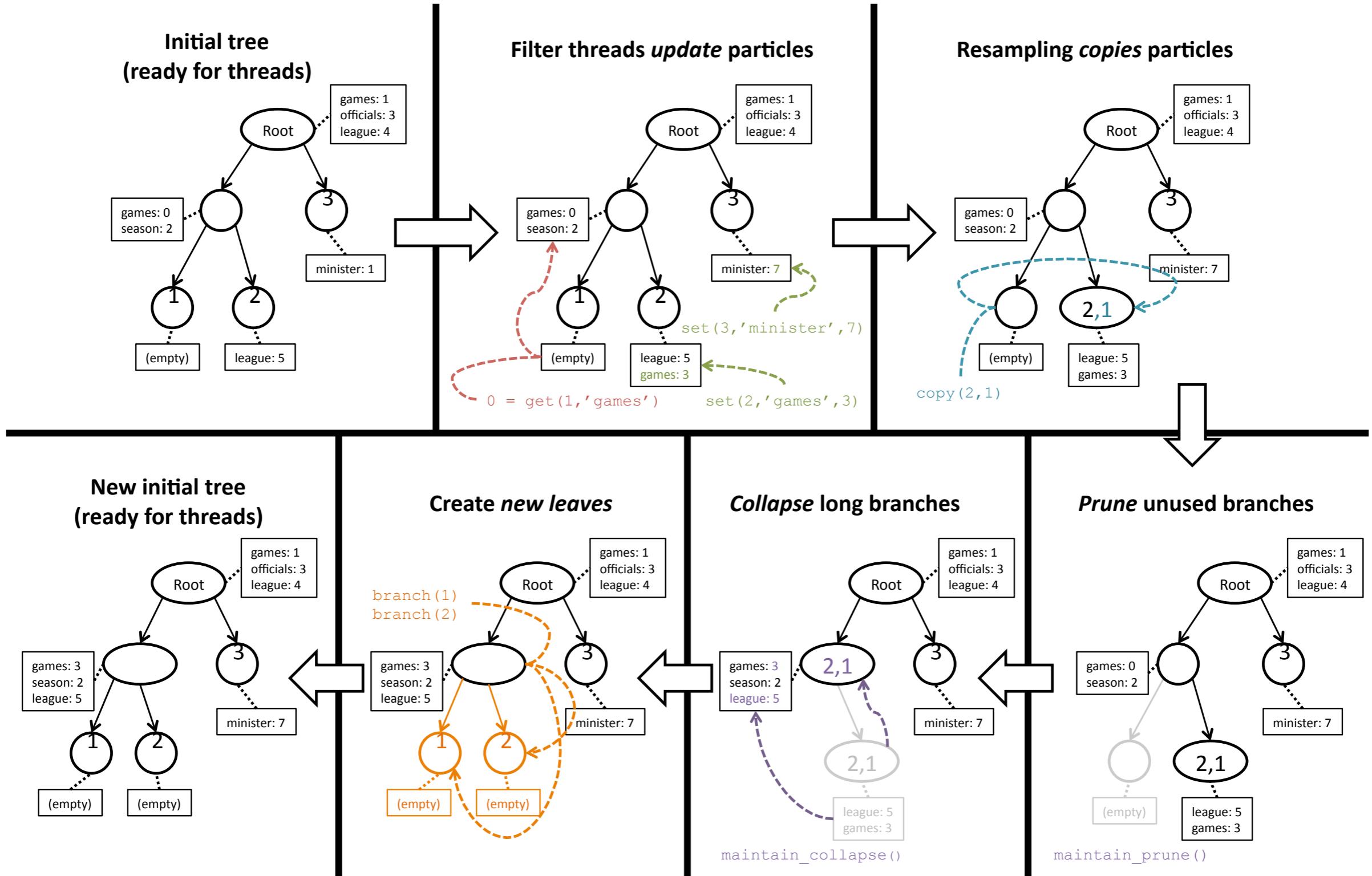
Particle Filter



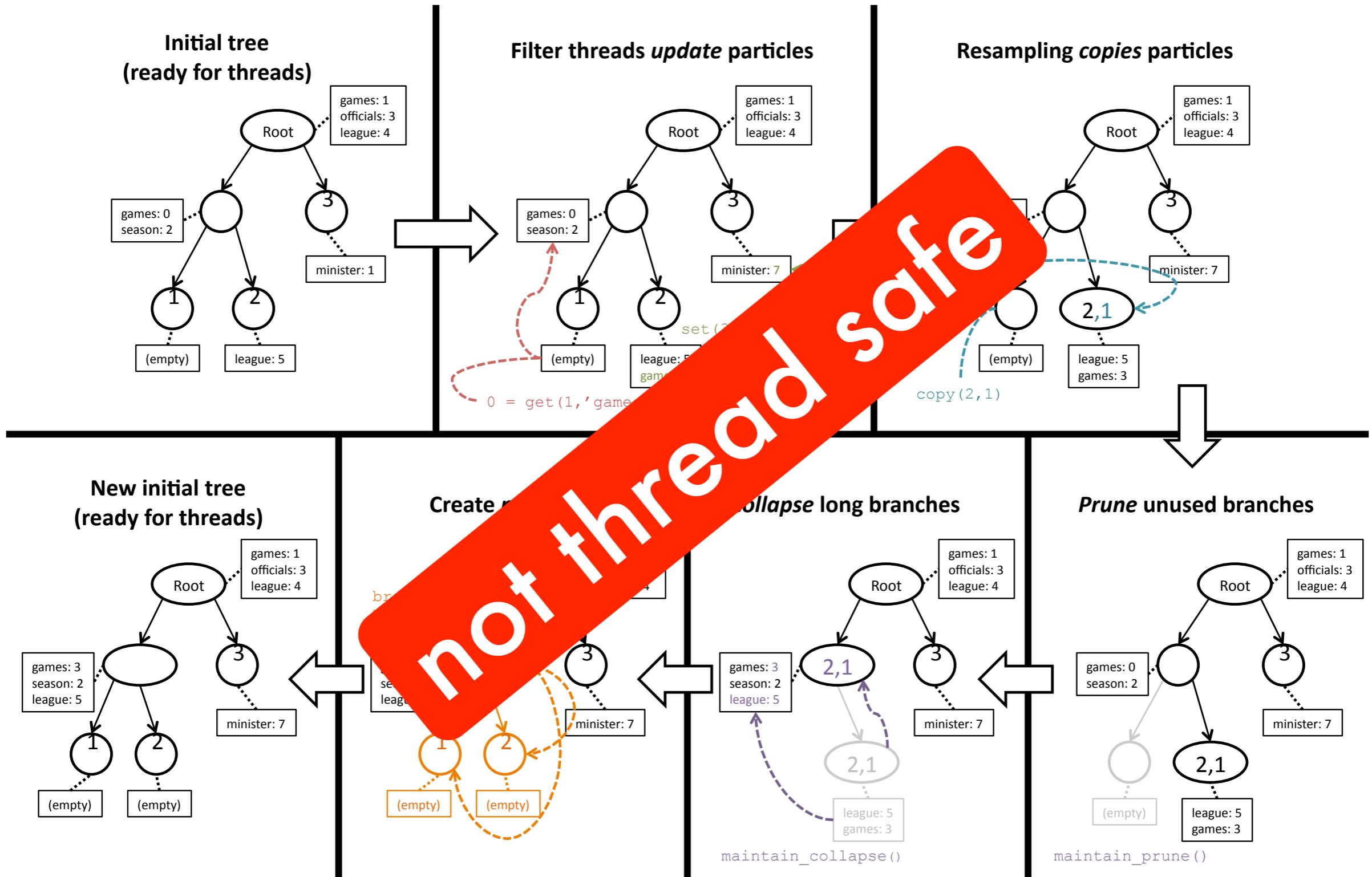
Algorithm 1 A Particle Filter Algorithm

```
Initialize  $\omega_1^f$  to  $\frac{1}{F}$  for all  $f \in \{1, \dots, F\}$ 
for each document  $d$  with time stamp  $t$  do
    for  $f \in \{1, \dots, F\}$  do
        Sample  $s_{td}^f, z_{td}^f$  using MCMC
         $\omega^f \leftarrow \omega^f P(x_{td}|z_{td}^f, s_{td}^f, x_{1:t,d-1})$ 
    end for
    Normalize particle weights
    if  $\|\omega_t\|_2^{-2} < \text{threshold}$  then
        resample particles
    for  $f \in \{1, \dots, F\}$  do
        MCMC pass over 10 random past documents
    end for
end if
end for
```

Inheritance Tree

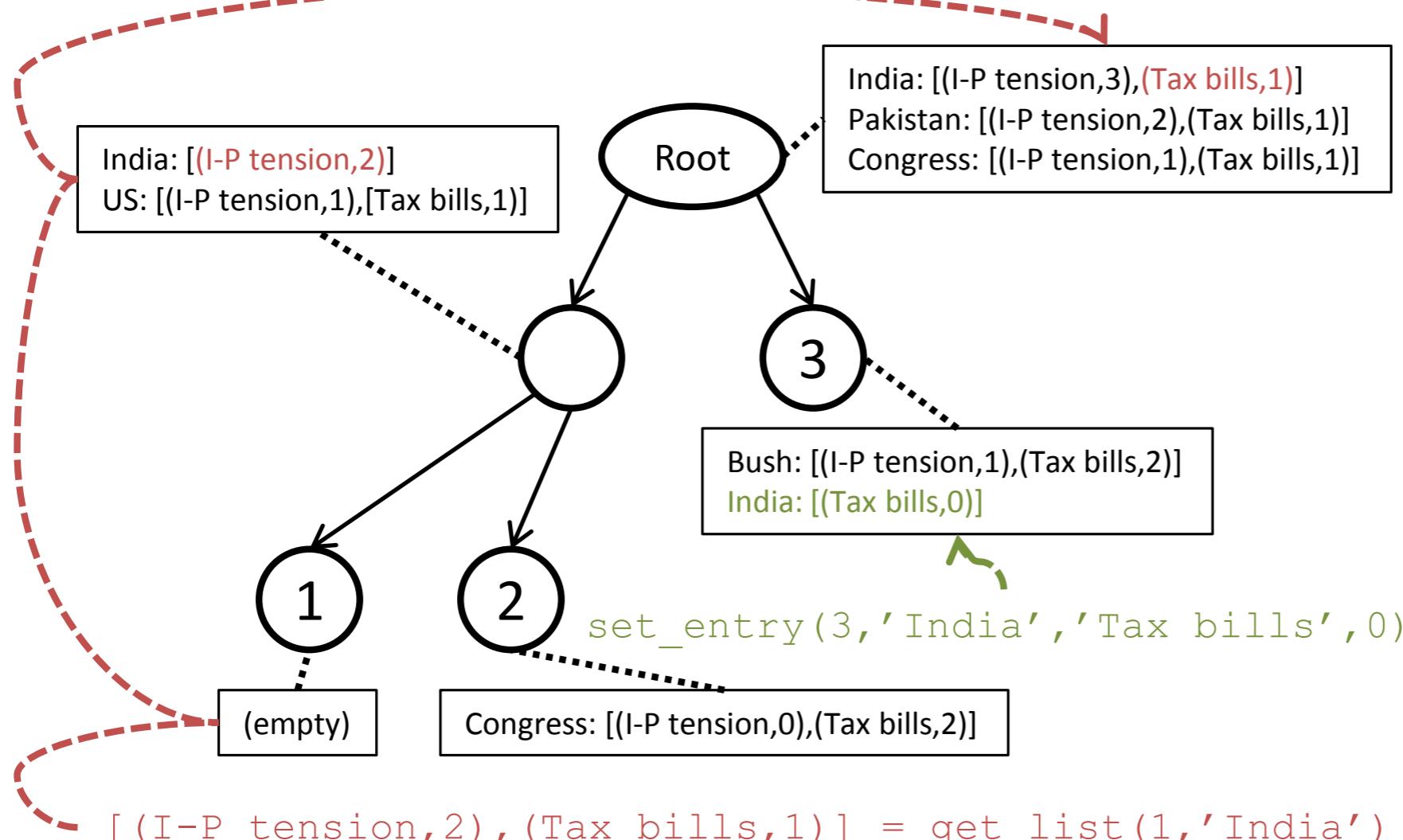


Inheritance Tree



Extended Inheritance Tree

Extended Inheritance Tree



write only in
the leaves
(per thread)

Note: “I-P tension” is short for “India-Pakistan tension”

Results

Ablation studies

- TDT5 (Topic Detection and Tracking)
macro-averaged minimum detection cost: **0.714**
- Removing features

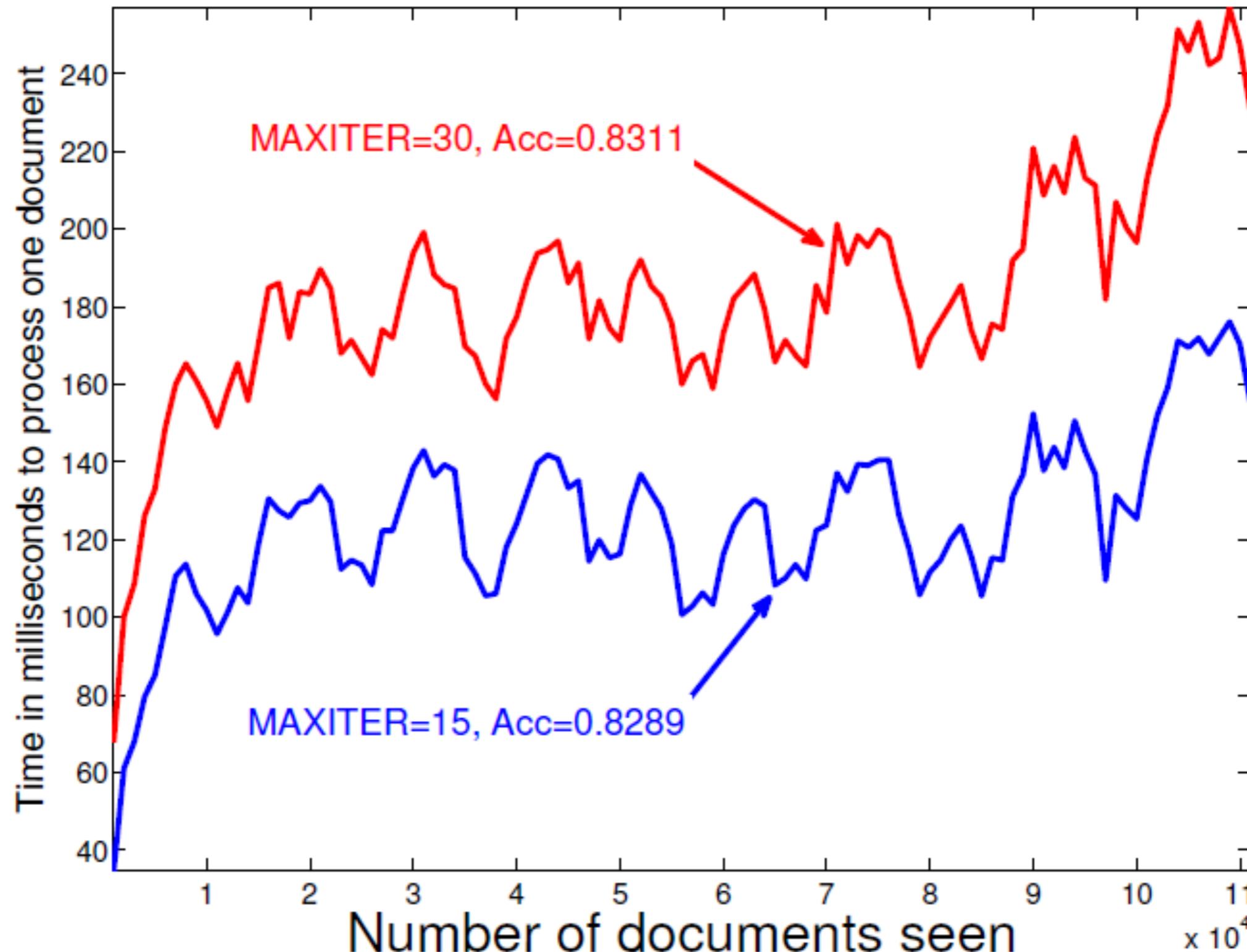
time	entities	topics	story words
0.84	0.90	0.86	0.75

Comparison

Sample No.	Sample size	Num Words	Num Entities	Story Acc.	LSHC Acc.
1	111,732	19,218	12,475	0.8289	0.738
2	274,969	29,604	21,797	0.8388	0.791
3	547,057	40,576	32,637	0.8395	0.800

Hashing &
correlation clustering

Time-Accuracy trade off



Stories

TOPICS

STORYLINES

Sports

games
won
team
final
season
league
held

Politics

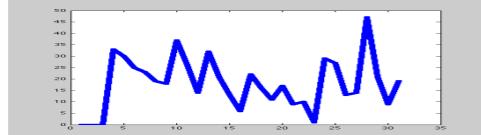
government
minister
authorities
opposition
officials
leaders
group

Unrest

police
attack
run
man
group
arrested
move

UEFA-soccer

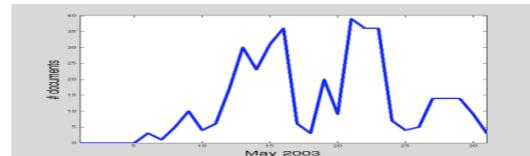
champions *Juventus*
goal *AC Milan*
leg *Real Madrid*
coach *Milan*
striker *Lazio*
midfield *Ronaldo*
penalty *Lyon*



Tax bills

tax
billion
cut
plan
budget
economy
lawmakers

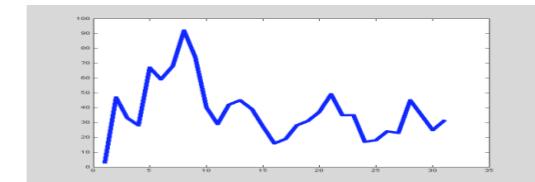
Bush
Senate
US
Congress
Fleischer
White House
Republican



India-Pakistan tension

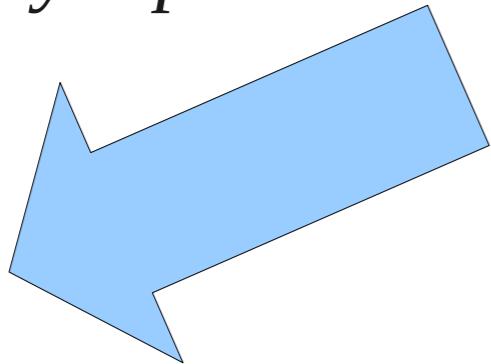
nuclear
border
dialogue
diplomatic
militant
insurgency
missile

Pakistan
India
Kashmir
New Delhi
Islamabad
Musharraf
Vajpayee



Related Stories

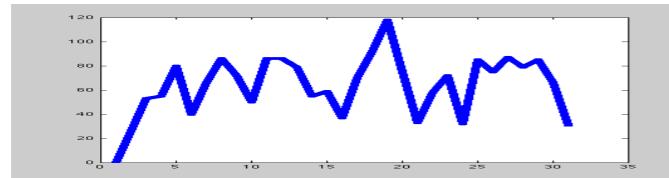
“Show similar stories by topic”



Middle-east conflict

Peace
Roadmap
Suicide
Violence
Settlements
bombing

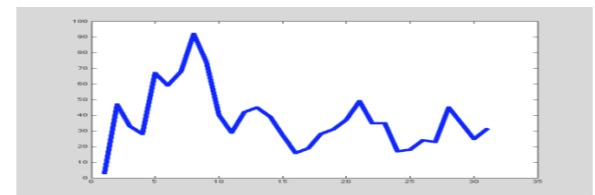
*Israel
Palestinian
West bank
Sharon
Hamas
Arafat*



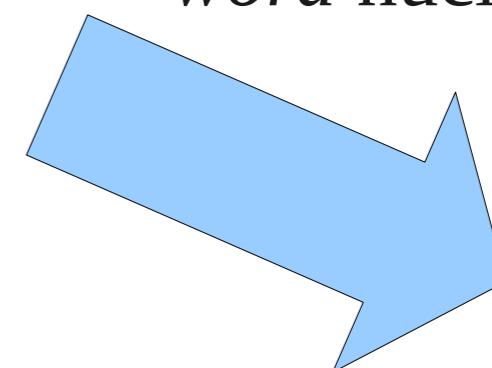
India-Pakistan tension

nuclear
border
dialogue
diplomatic
militant
insurgency
missile

*Pakistan
India
Kashmir
New Delhi
Islamabad
Musharraf
Vajpayee*



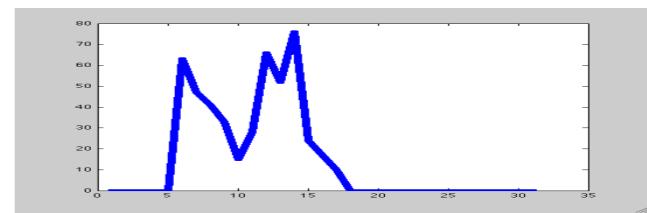
“Show similar stories, require the word nuclear”



North Korea nuclear

nuclear
summit
warning
policy
missile
program

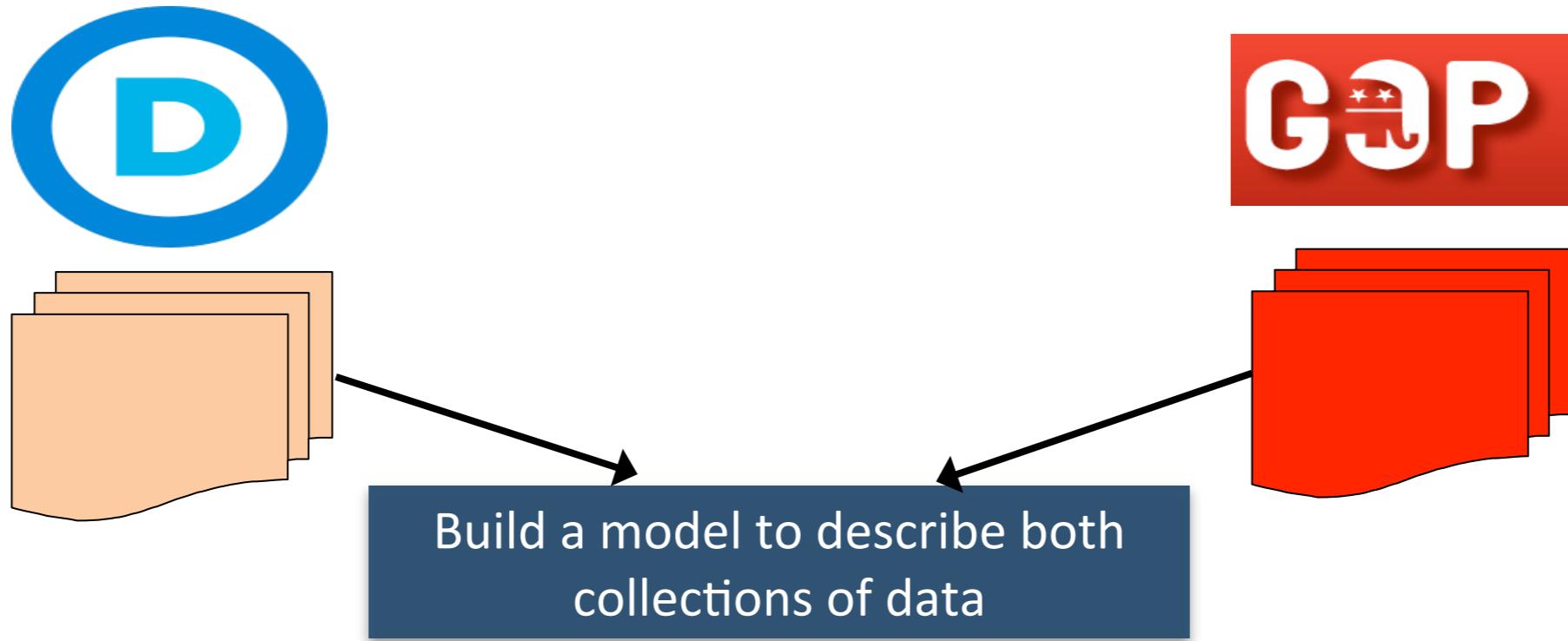
*North Korea
South Korea
U.S
Bush
Pyongyang*



Detecting Ideologies

Ahmed and Xing, 2010

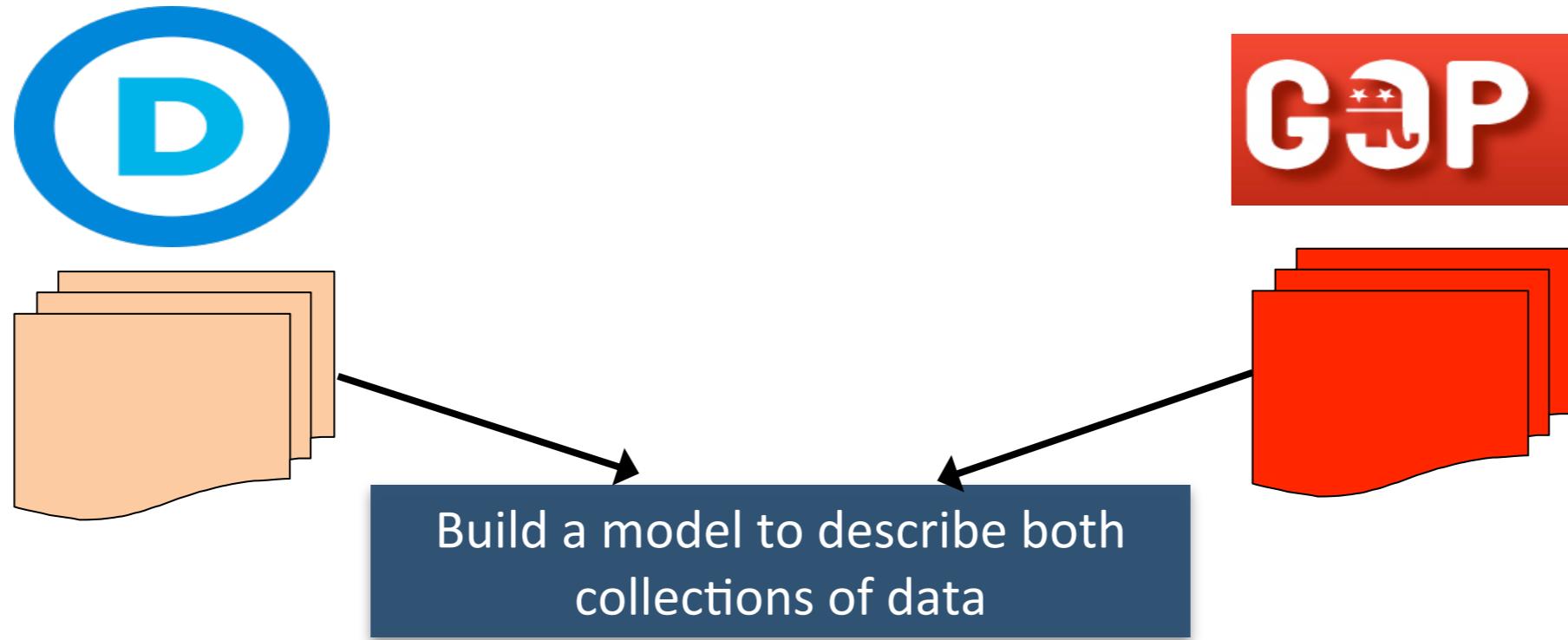
Ideologies



Visualization

- How does each ideology **view** mainstream events?
- On which topics do they **differ**?
- On which topics do they **agree**?

Ideologies

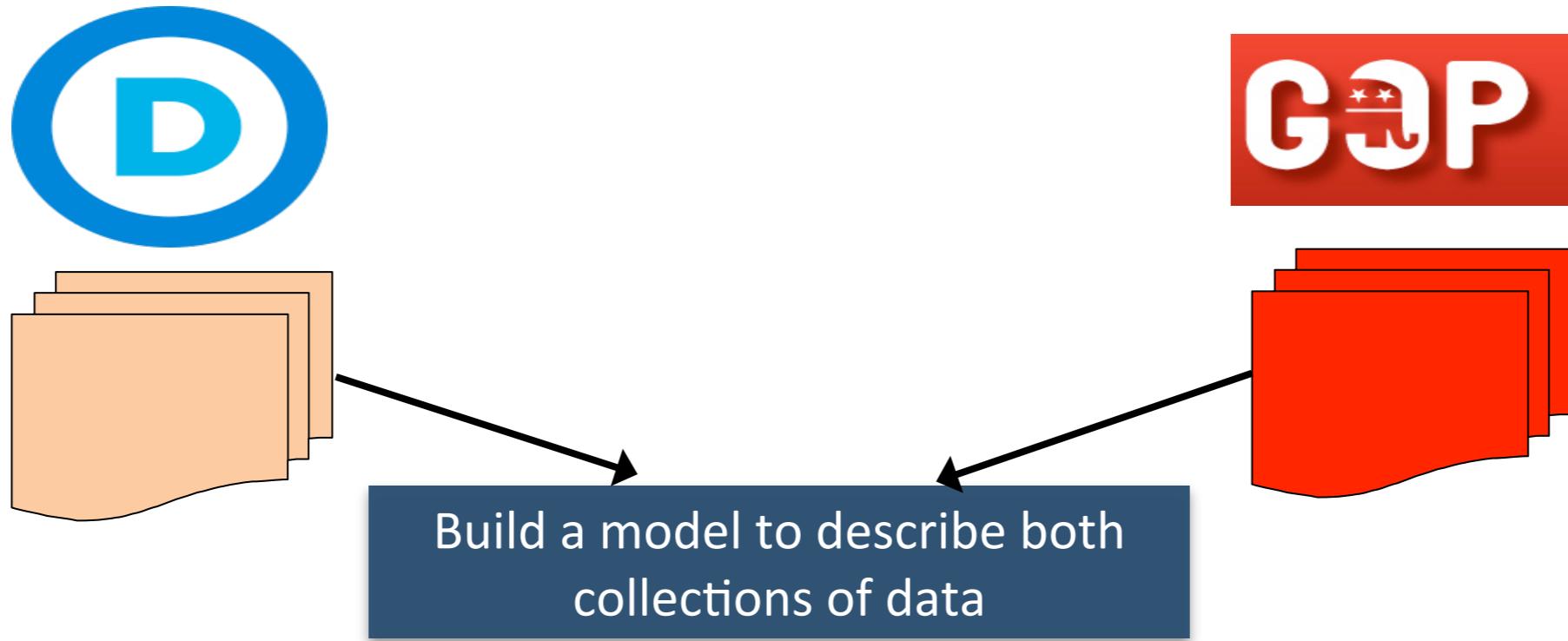


Visualization

Classification

- Given a **new** news article or a blog post, the system should infer
 - From which **side** it was written
 - **Justify** its answer on a topical level (view on abortion, taxes, health care)

Ideologies



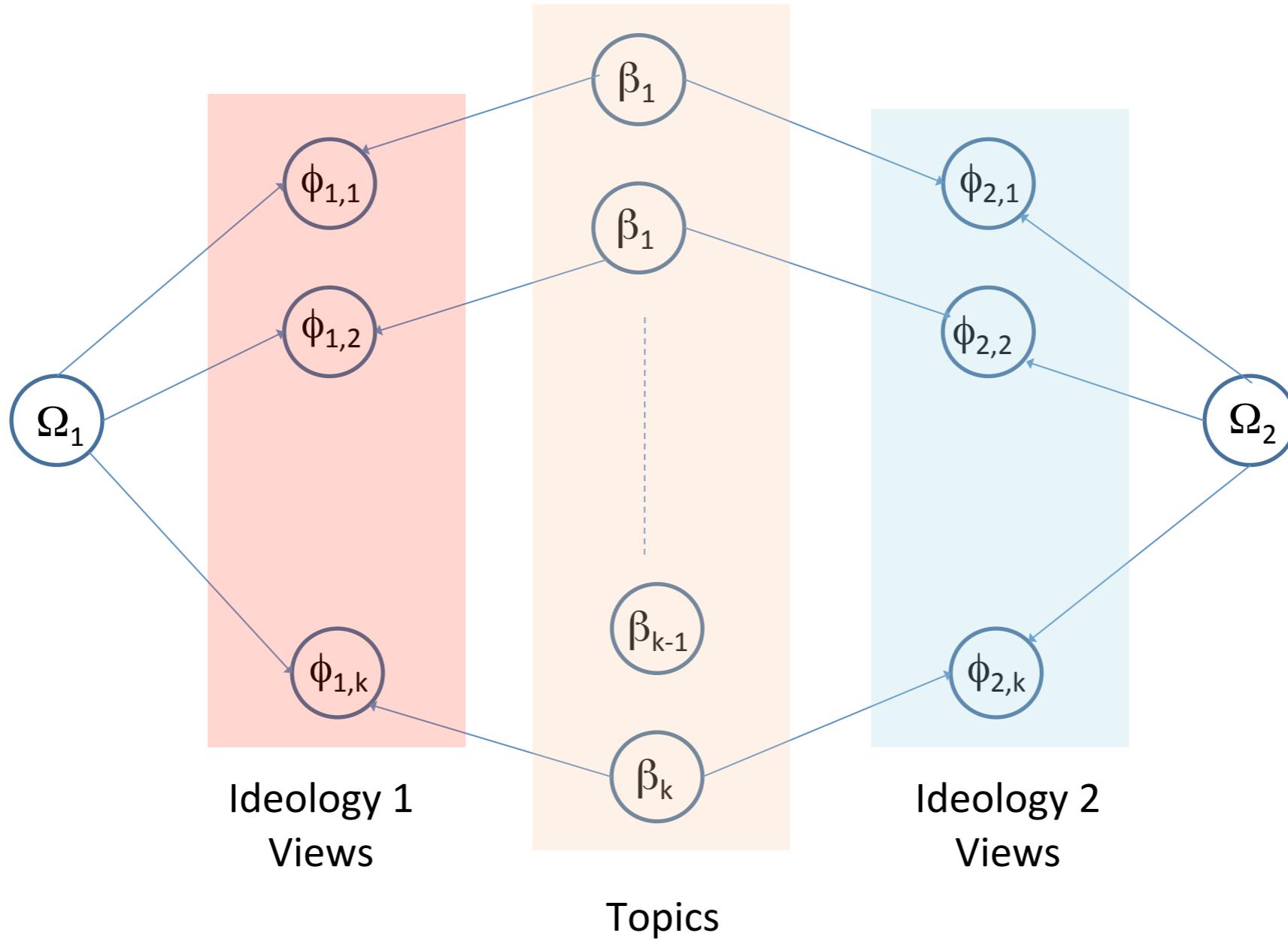
Visualization

Classification

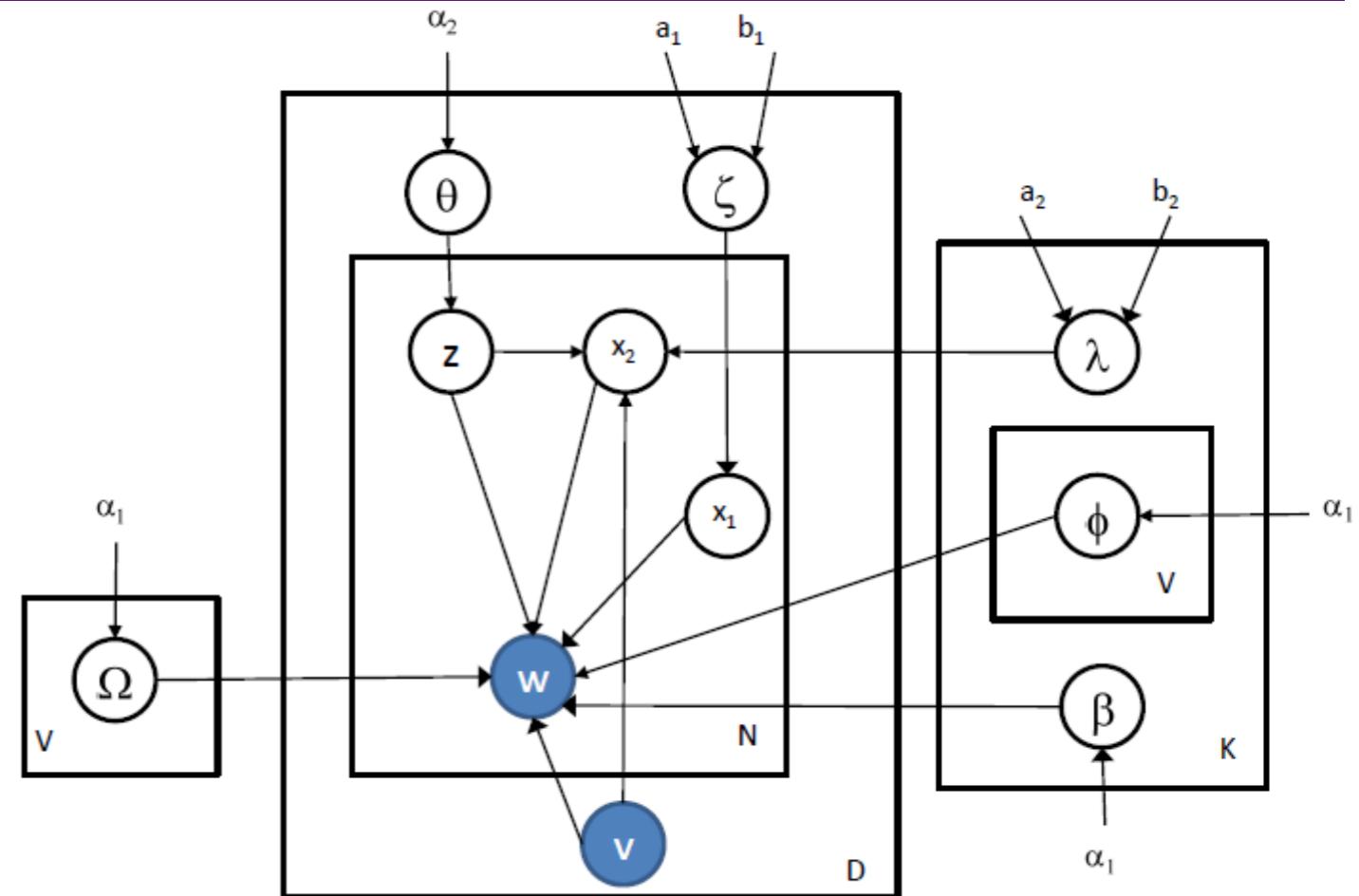
Structured browsing

- Given a **new** news article or a blog post, the user can ask for :
 - Examples of other articles from the same ideology about the same topic
 - Documents that could exemplify **alternative** views from **other ideologies**

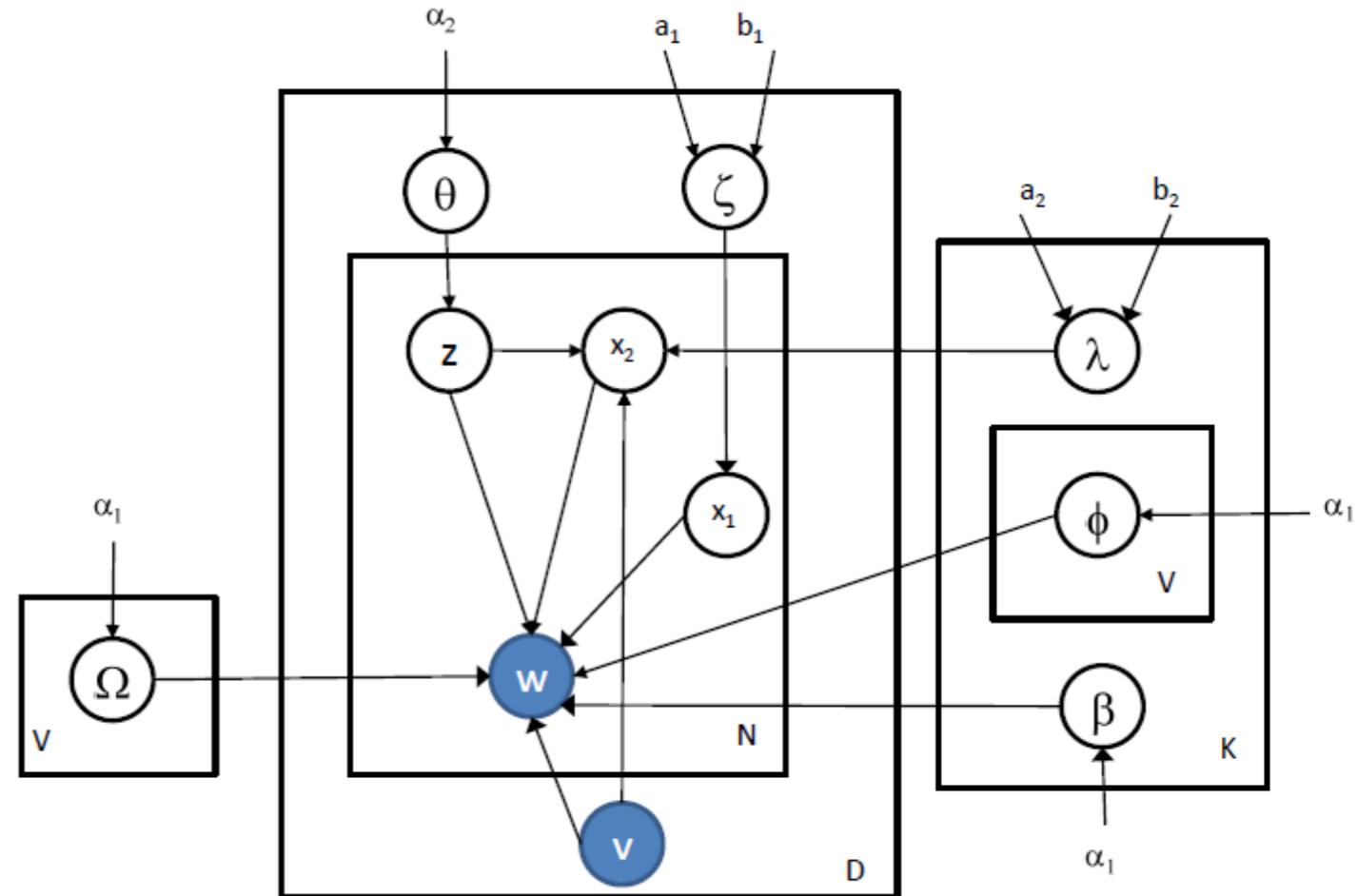
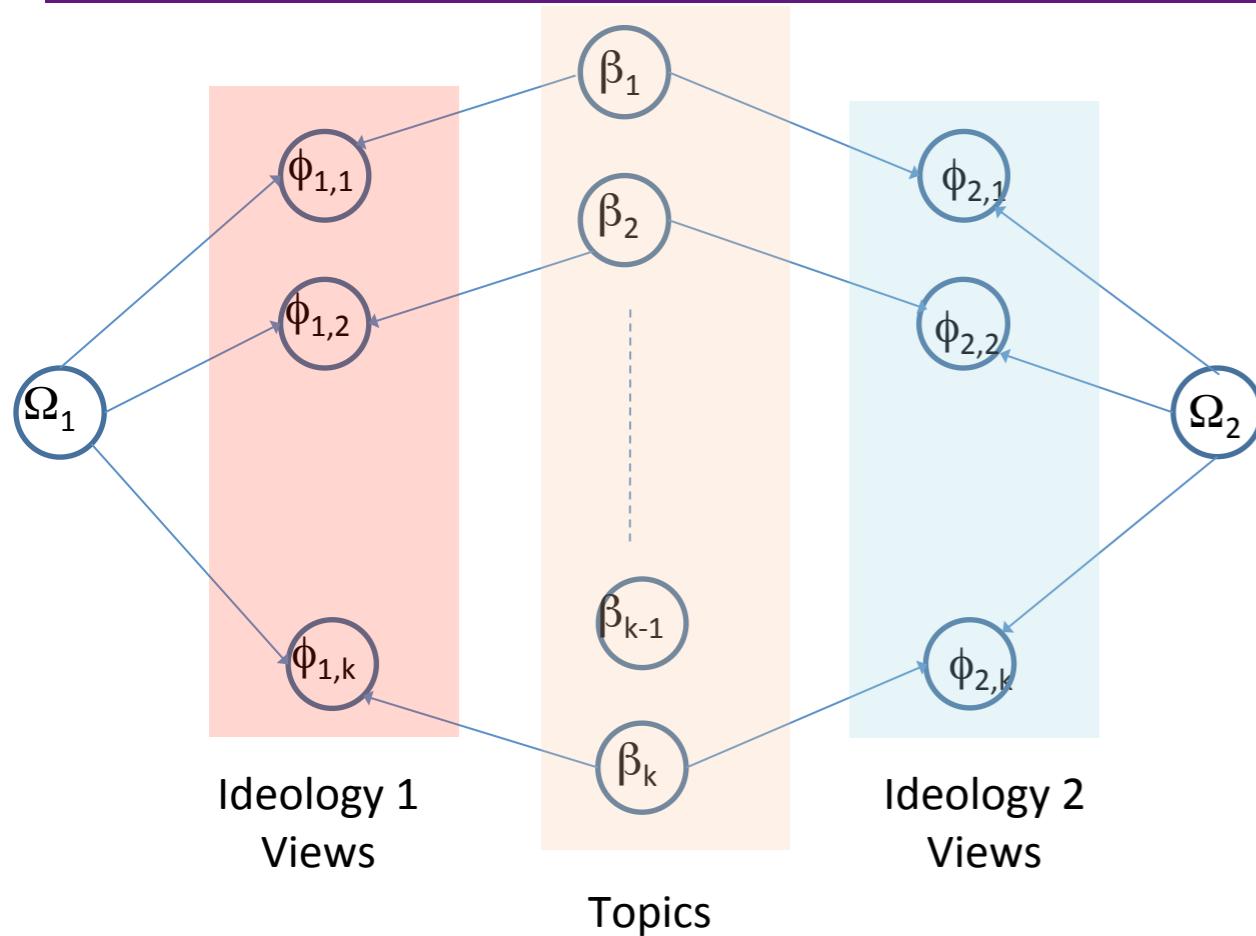
Building a factored model



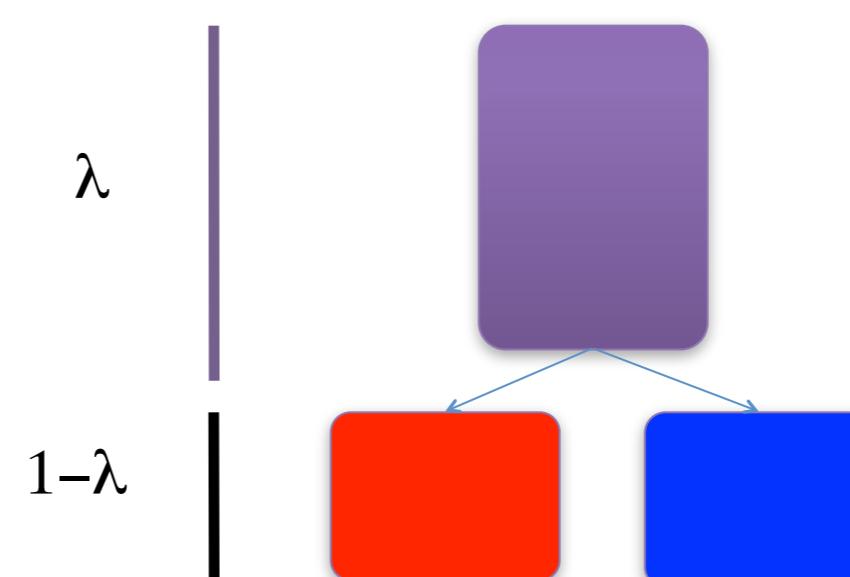
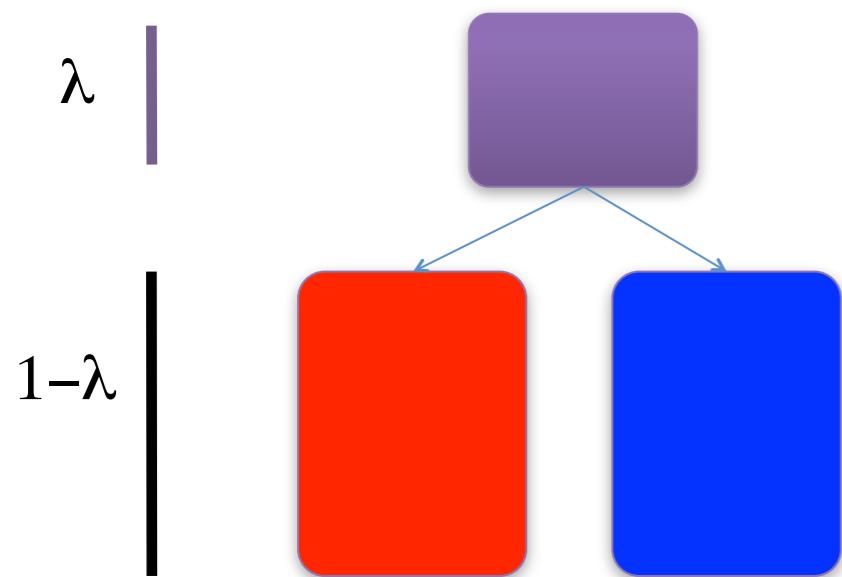
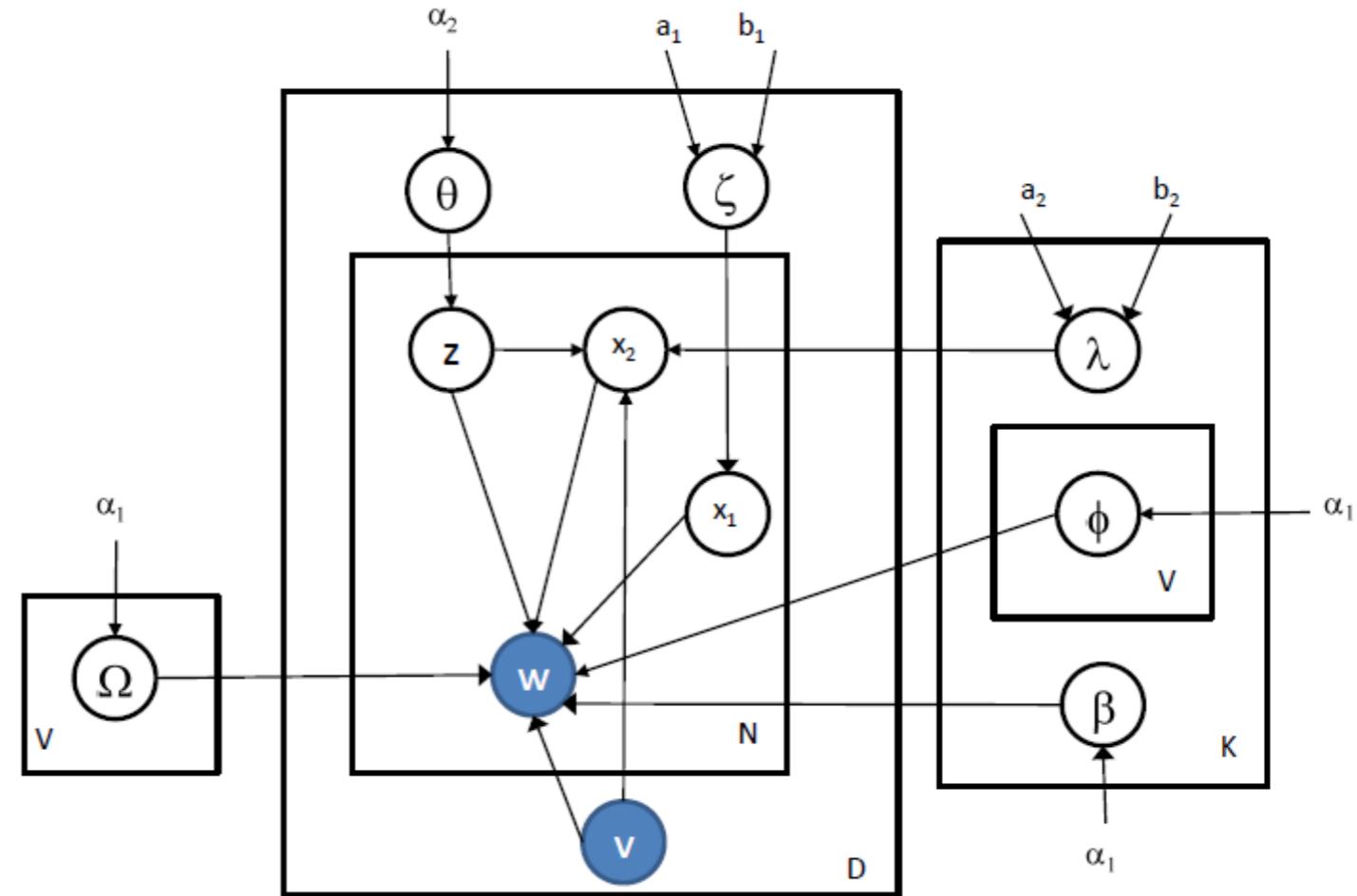
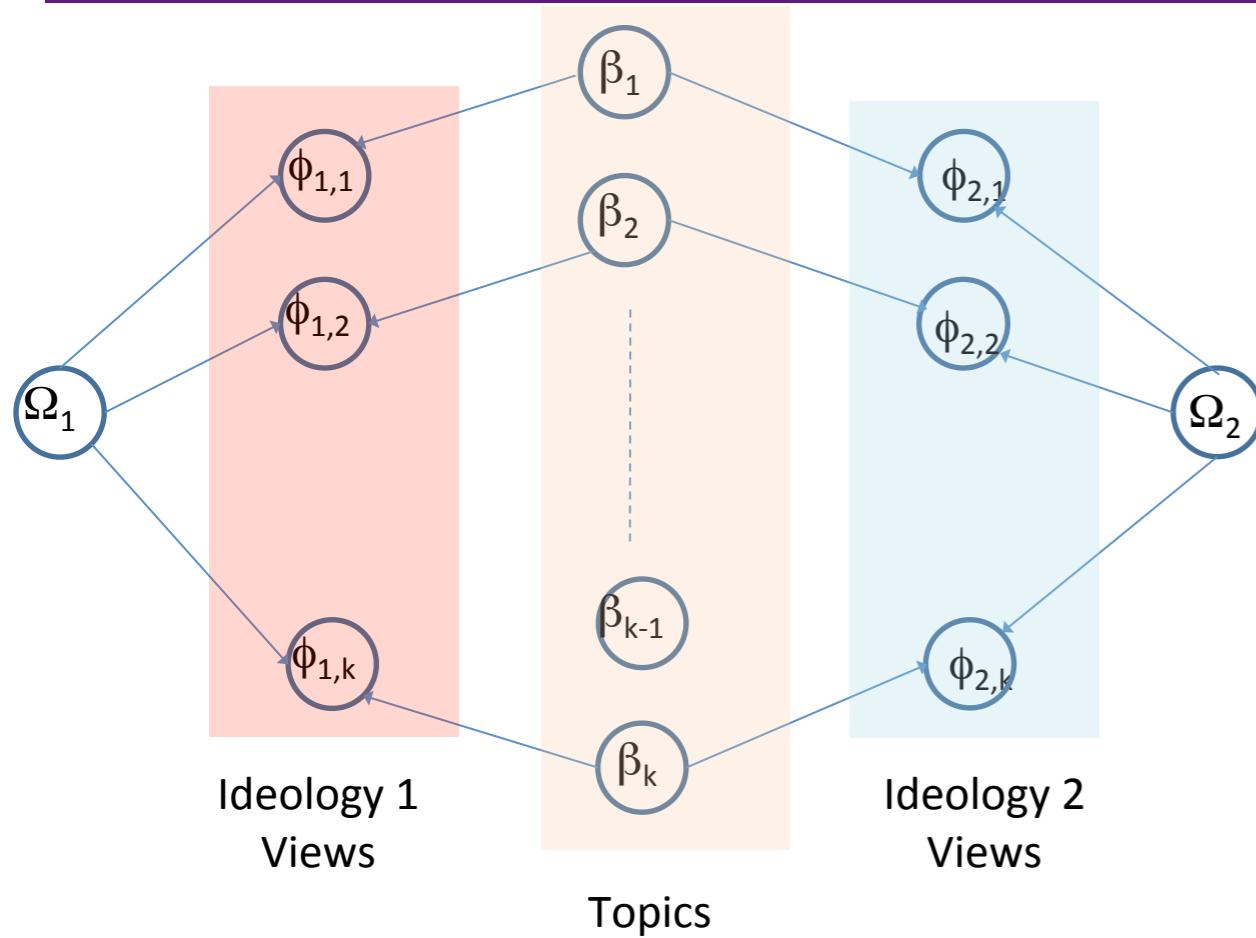
Building a factored model



Building a factored model



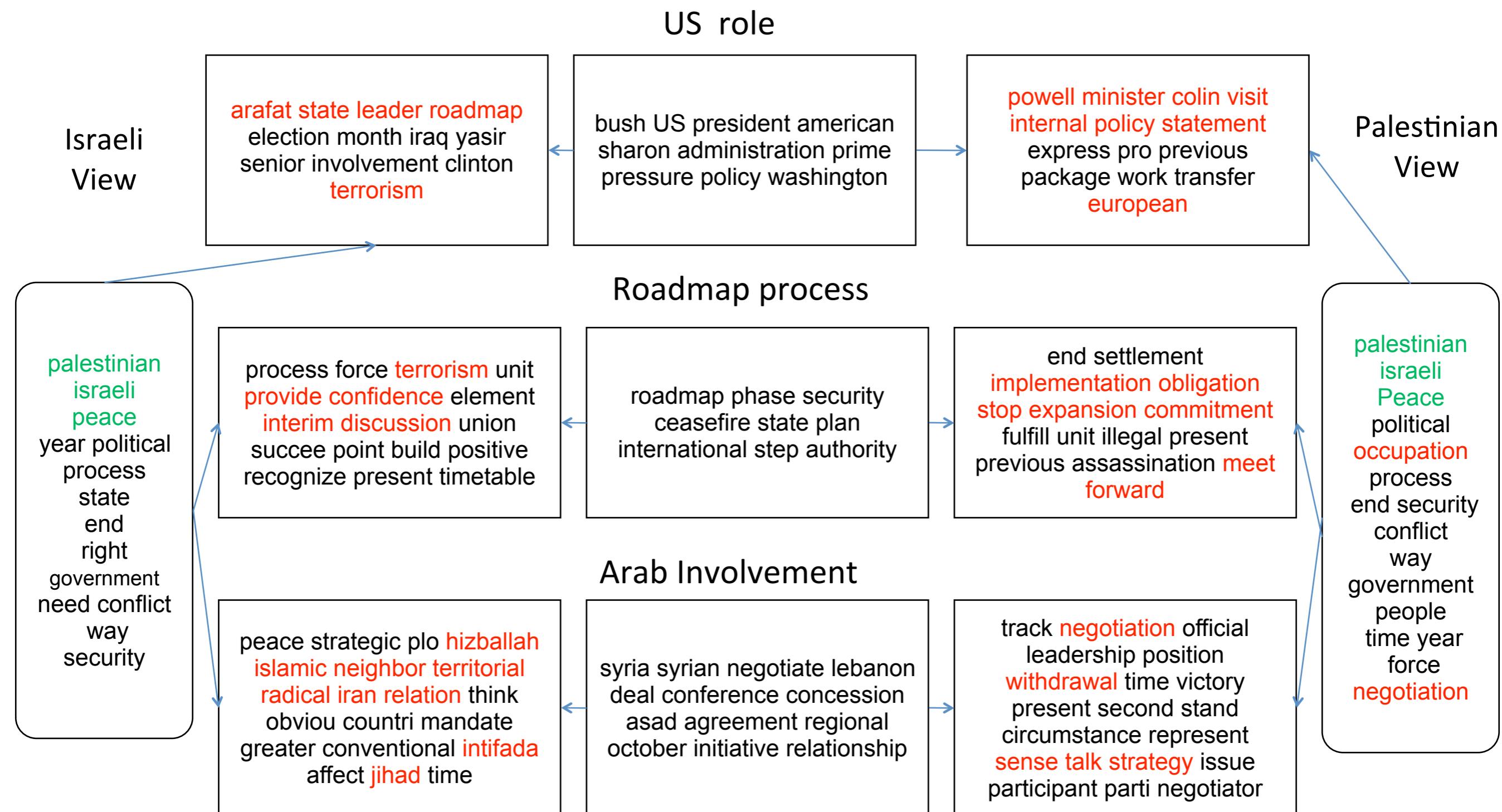
Building a factored model



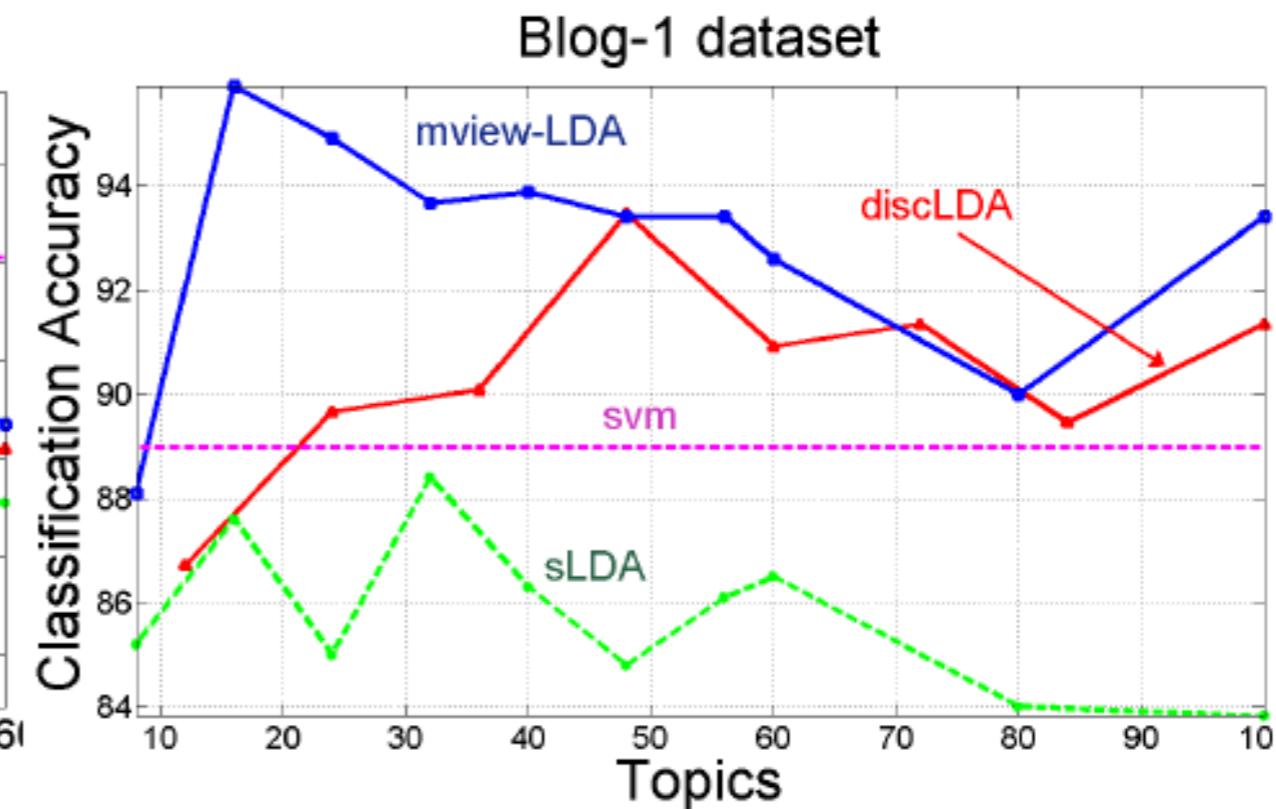
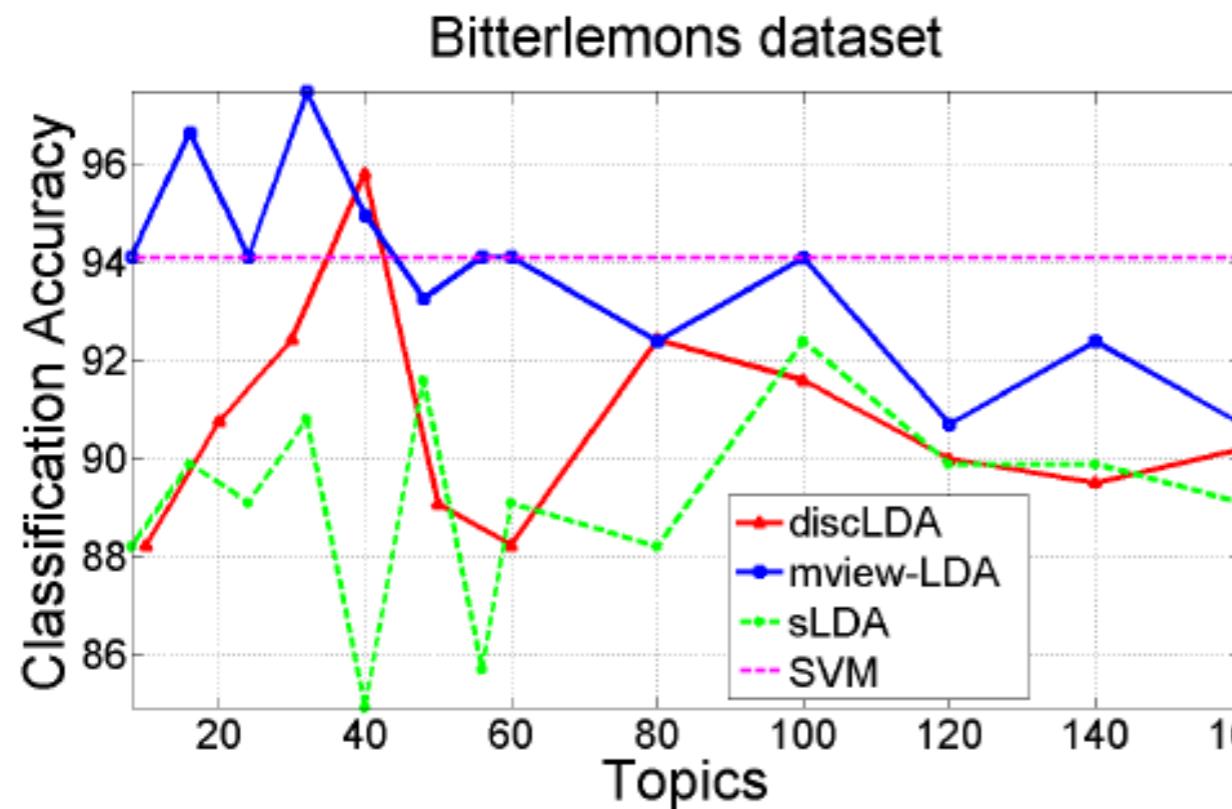
Data

- **Bitterlemons:**
 - Middle-east conflict, document written by Israeli and Palestinian authors.
 - ~300 documents form each view with average length 740
 - Multi author collection
 - 80-20 split for test and train
- **Political Blog-1:**
 - American political blogs (Democrat and Republican)
 - 2040 posts with average post length = 100 words
 - Follow test and train split as in (Yano et al., 2009)
- **Political Blog-2 (test generalization to a new writing style)**
 - Same as 1 but 6 blogs, 3 from each side
 - ~14k posts with ~200 words per post
 - 4 blogs for training and 2 blogs for test

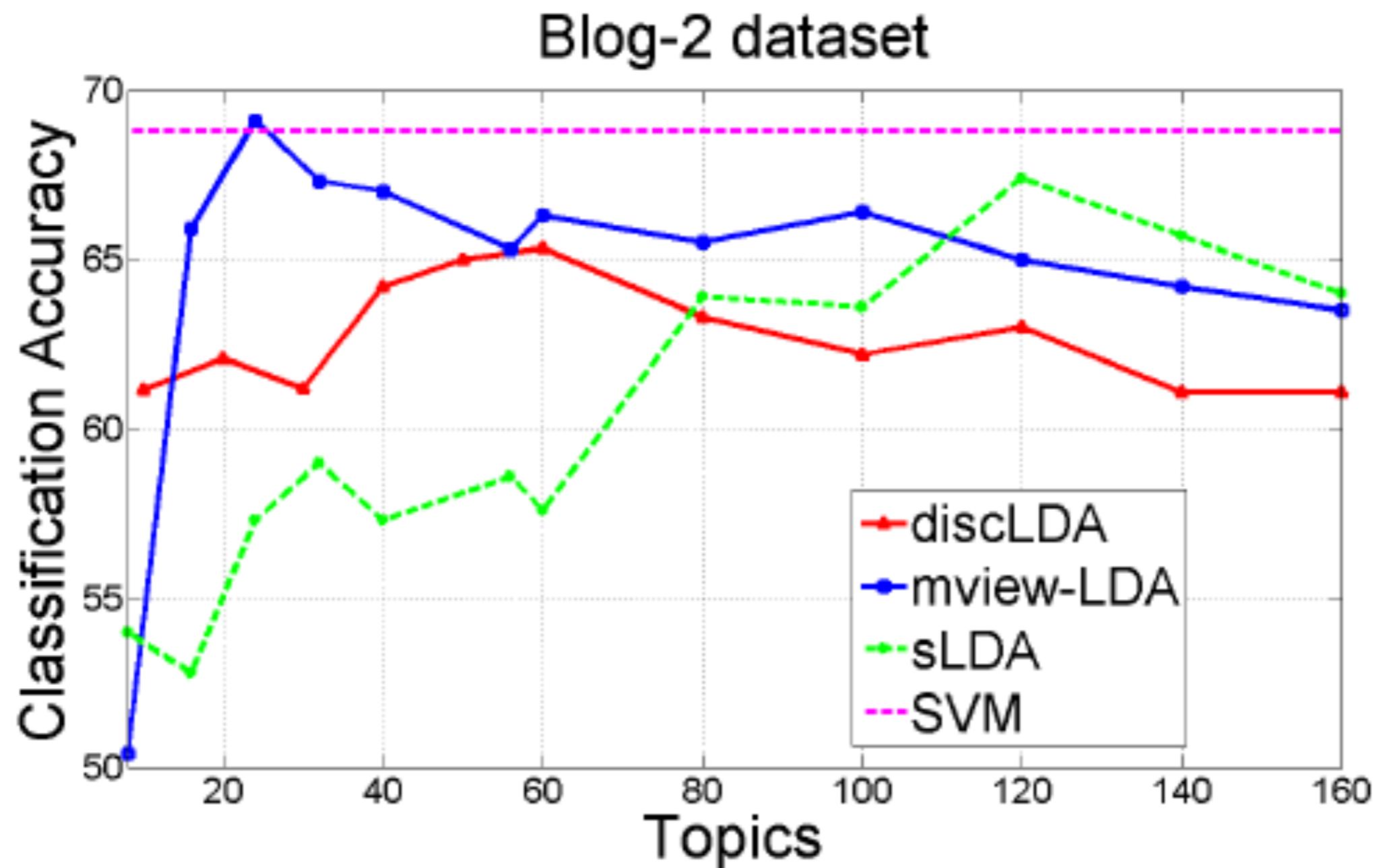
Bitterlemons dataset



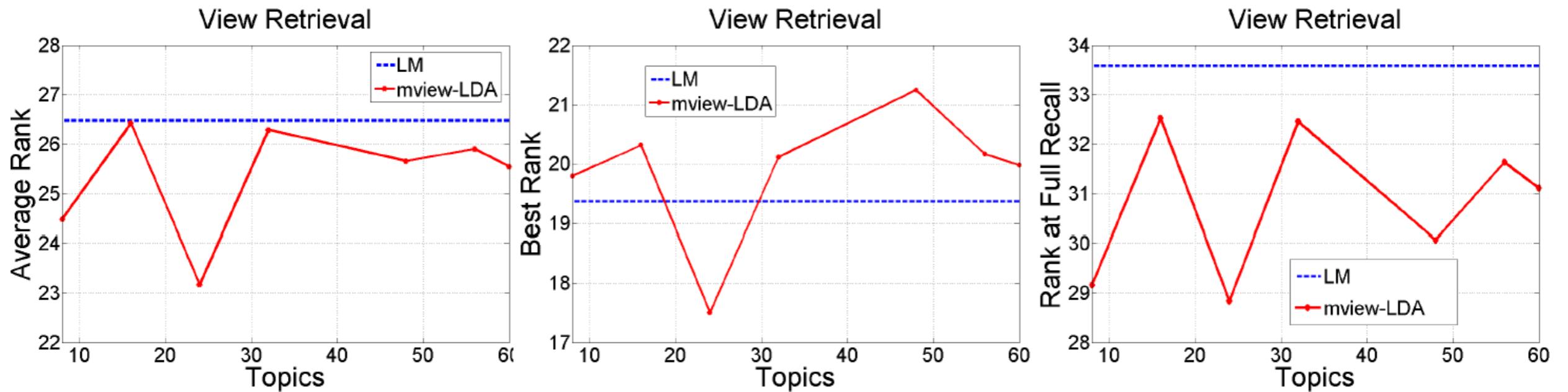
Classification accuracy



Generalization to new blogs

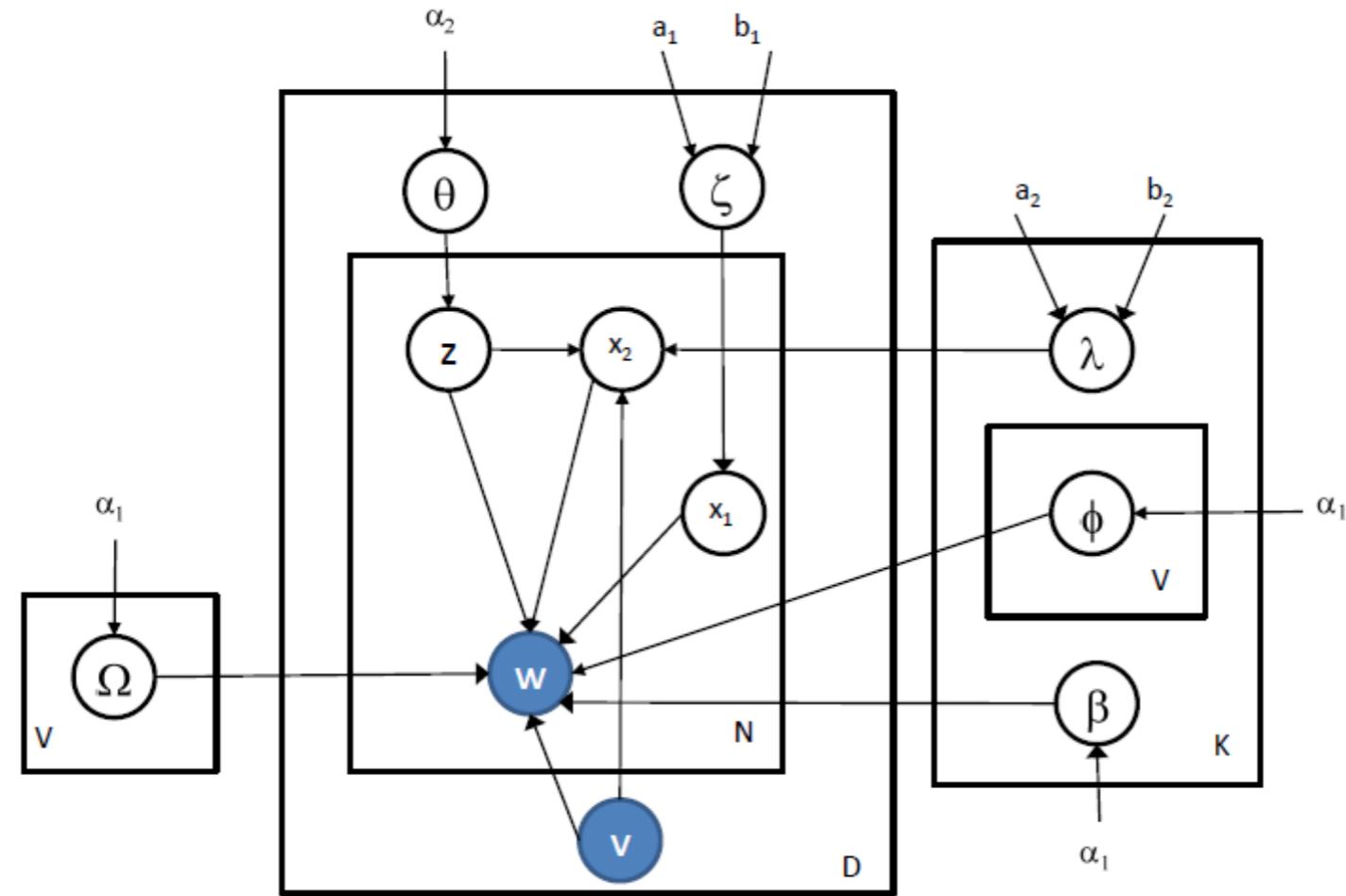


Finding alternate views

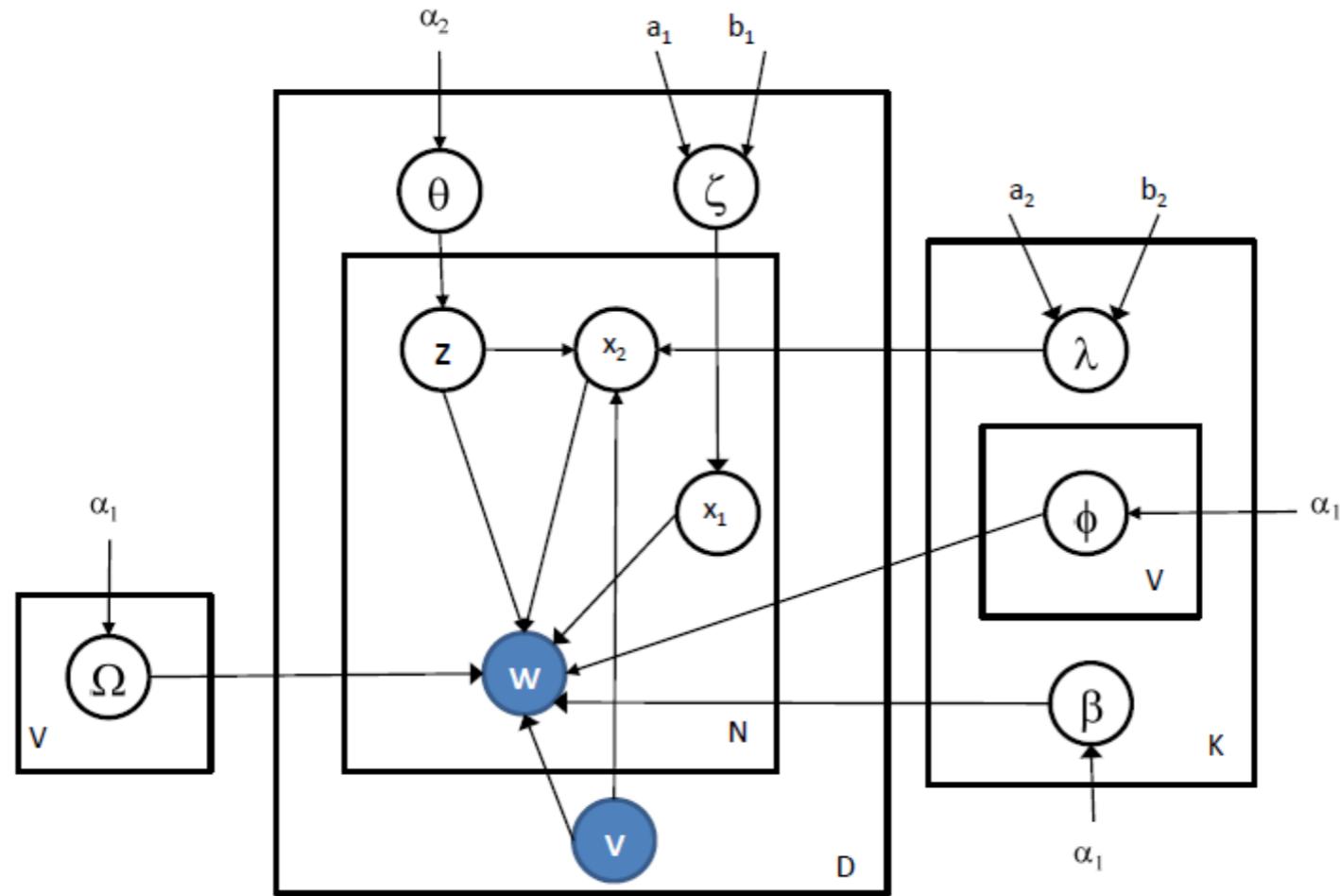


- Given a document written in one ideology, retrieve the equivalent
- Baseline: SVM + cosine similarity

Unlabeled data

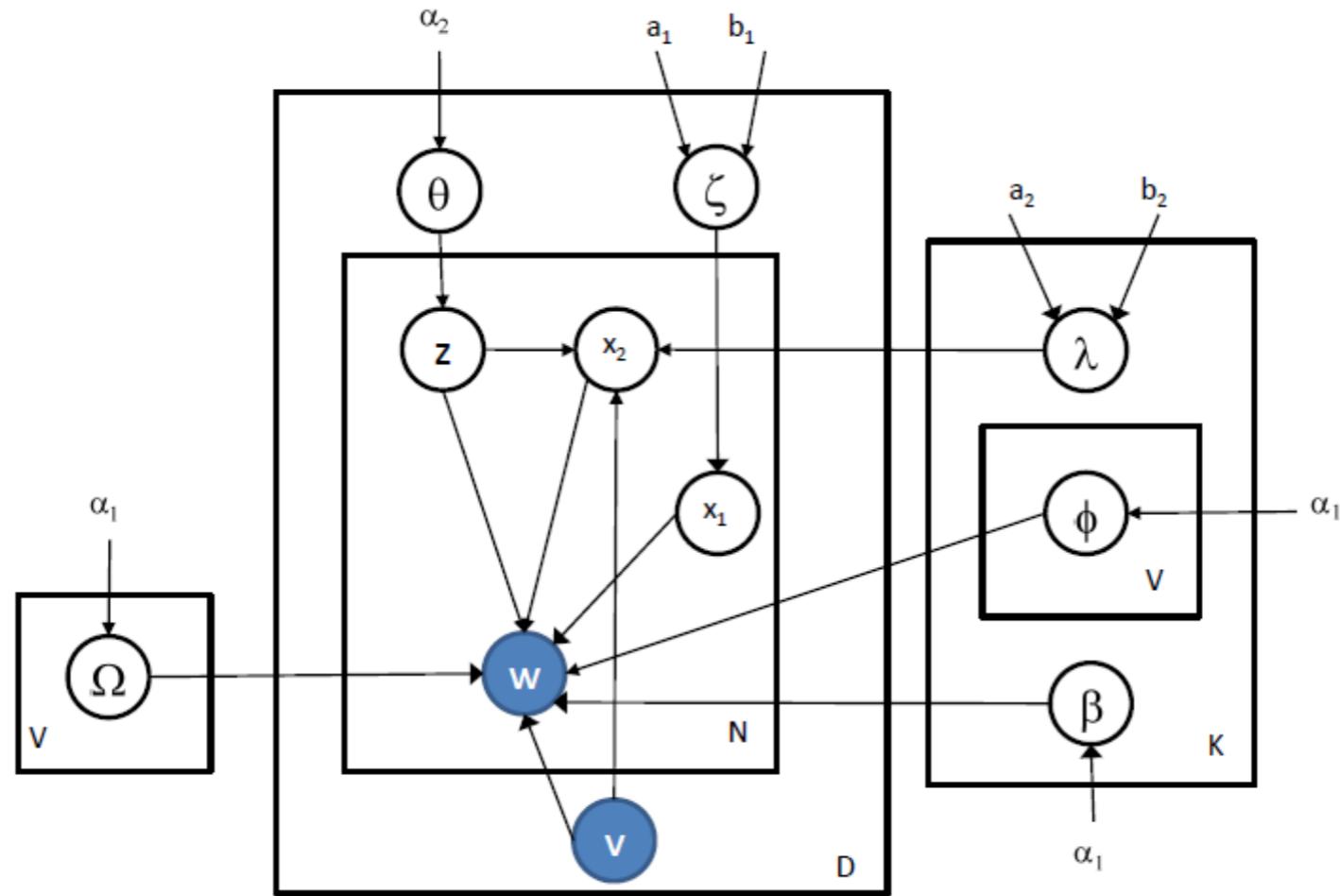


Unlabeled data



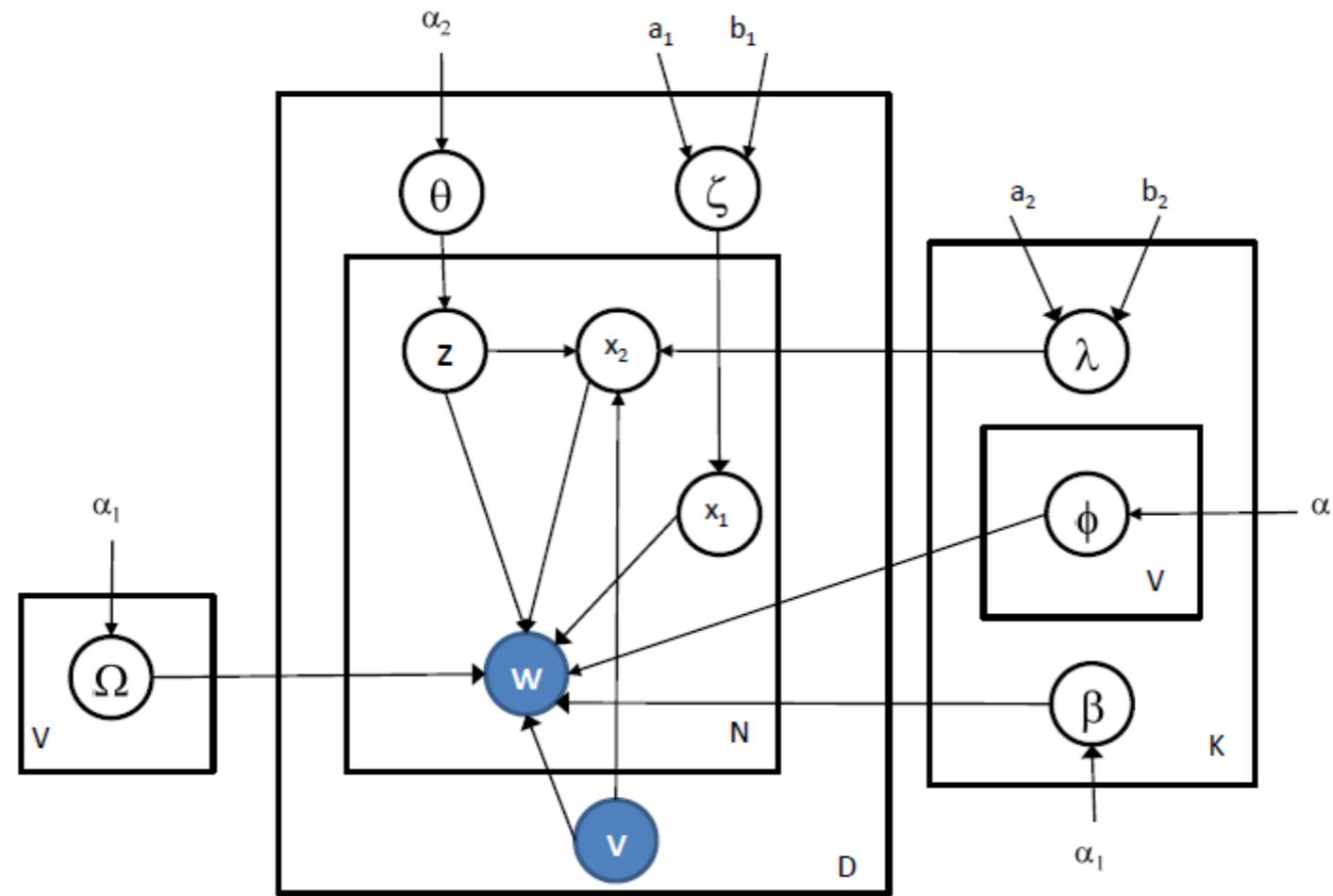
- In theory this is simple
 - Add a step that samples the document view (v)
 - **Doesn't mix** in practice because tight coupling between v and (x_1, x_2, z)

Unlabeled data



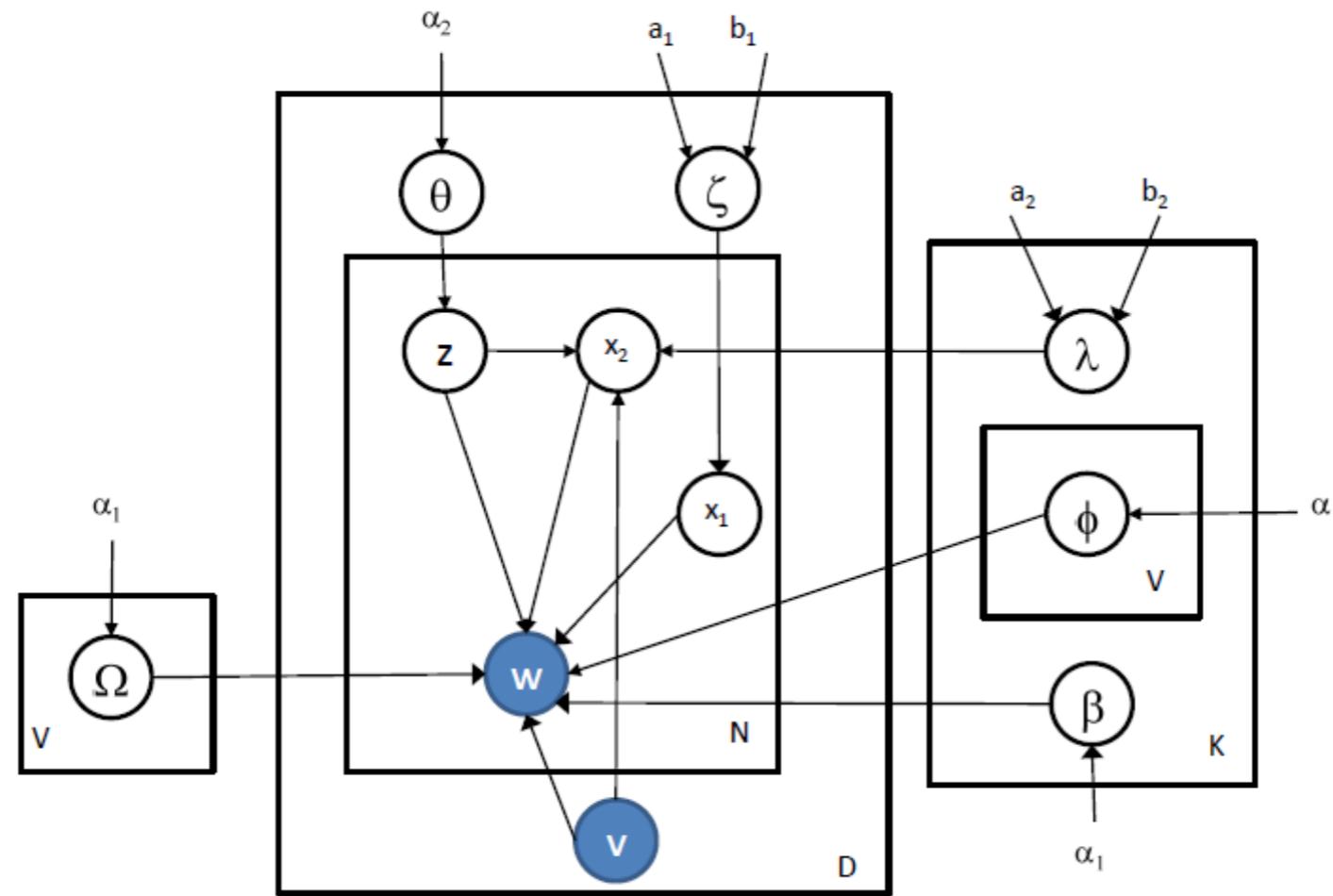
- In theory this is simple
 - Add a step that samples the document view (v)
 - **Doesn't mix** in practice because tight coupling between v and (x_1, x_2, z)
- Solution

Unlabeled data



- In theory this is **simple**
 - Add a step that samples the document view (v)
 - **Doesn't mix** in practice because tight coupling between v and (x_1, x_2, z)
- Solution
 - Sample v and (x_1, x_2, z) as a block using a Metropolis-Hastings step

Unlabeled data



- In theory this is **simple**
 - Add a step that samples the document view (v)
 - **Doesn't mix** in practice because tight coupling between v and (x_1, x_2, z)
- Solution
 - Sample v and (x_1, x_2, z) as a block using a Metropolis-Hastings step
 - This is a **huge proposal!**