...itional Inference
...ference in loopy models
...roximate inference

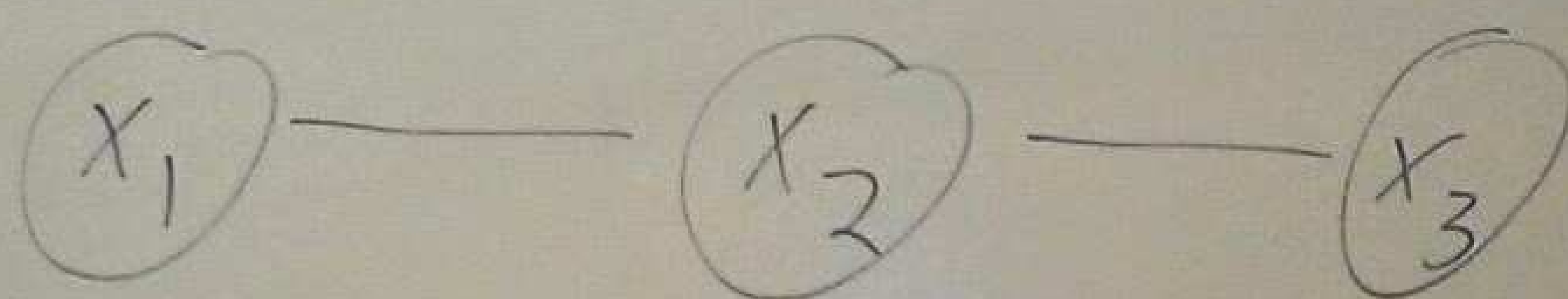
- Midterm comments
- Marked A6 due now
- A7 due next Wed.
- Please pick coding project (matLearn.zip up tonight)
- Optional "bonus" lecture Dec. 1 (location/time TBA)
  "Monte Carlo" methods.
- Course evaluations online.

Common mid-term errs:

- "cross"-validation
- $\|x\|^2$ is not a norm,

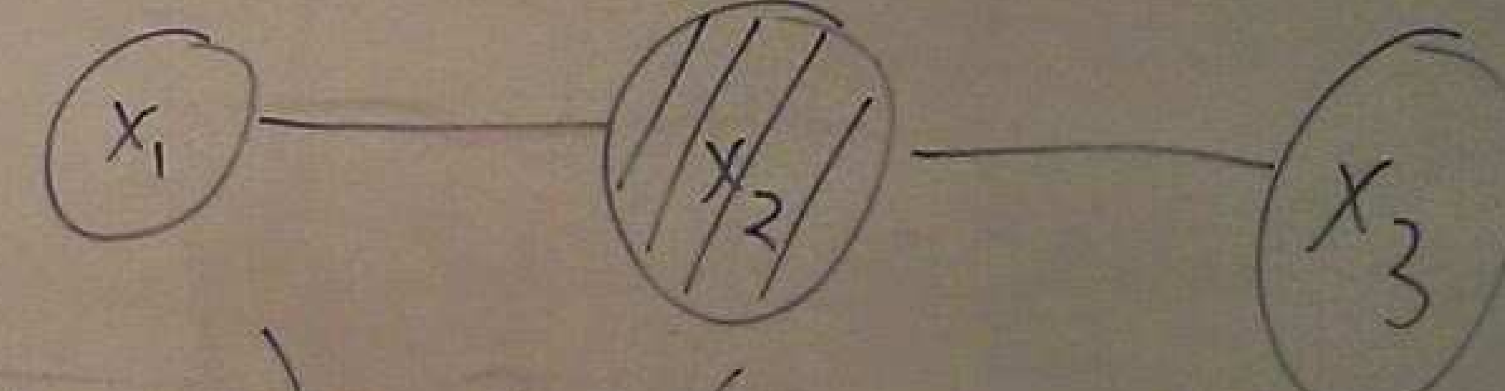
  composition of convex w/ convex
  is not generally convex

Conditioning in UGMs



$$p(x) = \frac{\phi_{12}(x_1, x_2)\,\phi_{23}(x_2, x_3)}{\underbrace{\sum_{x_1', x_2', x_3'} \phi_{12}(x_1', x_2')\,\phi_{23}(x_2', x_3')}_{Z}}$$

Condition on x_2:

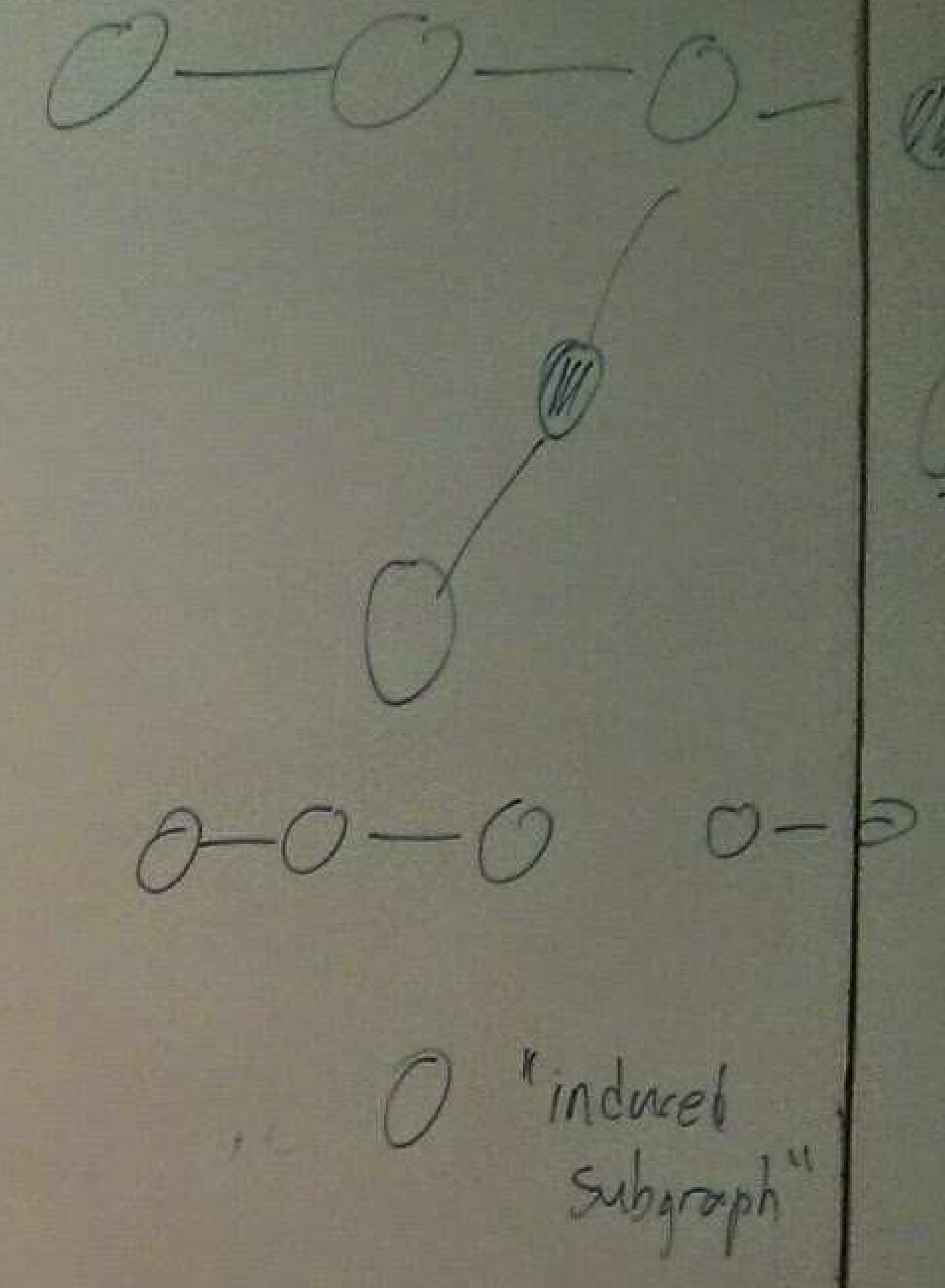


$$p(x_1, x_3 \mid x_2) = \frac{p(x_1, x_2, x_3)}{p(x_2)} = \frac{p(x_1, x_2, x_3)}{\sum_{x_1', x_3'} p(x_1', x_2, x_3')}$$

$$= \frac{\left(\dfrac{\phi_{12}(x_1, x_2)\,\phi_{23}(x_2, x_3)}{\underbrace{\sum_{x'} \phi_{12}(x_1', x_2)\,\phi_{23}(x_2, x_3')}_{Z}}\right)}{\left(\dfrac{\sum_{x'} \phi_{12}(x_1', x_2')\,\phi_{23}(x_2', x_3')}{\sum_{x_1', x_3'} \phi_{12}(x_1', x_2)\,\phi_{23}(x_2, x_3')}\right)}$$

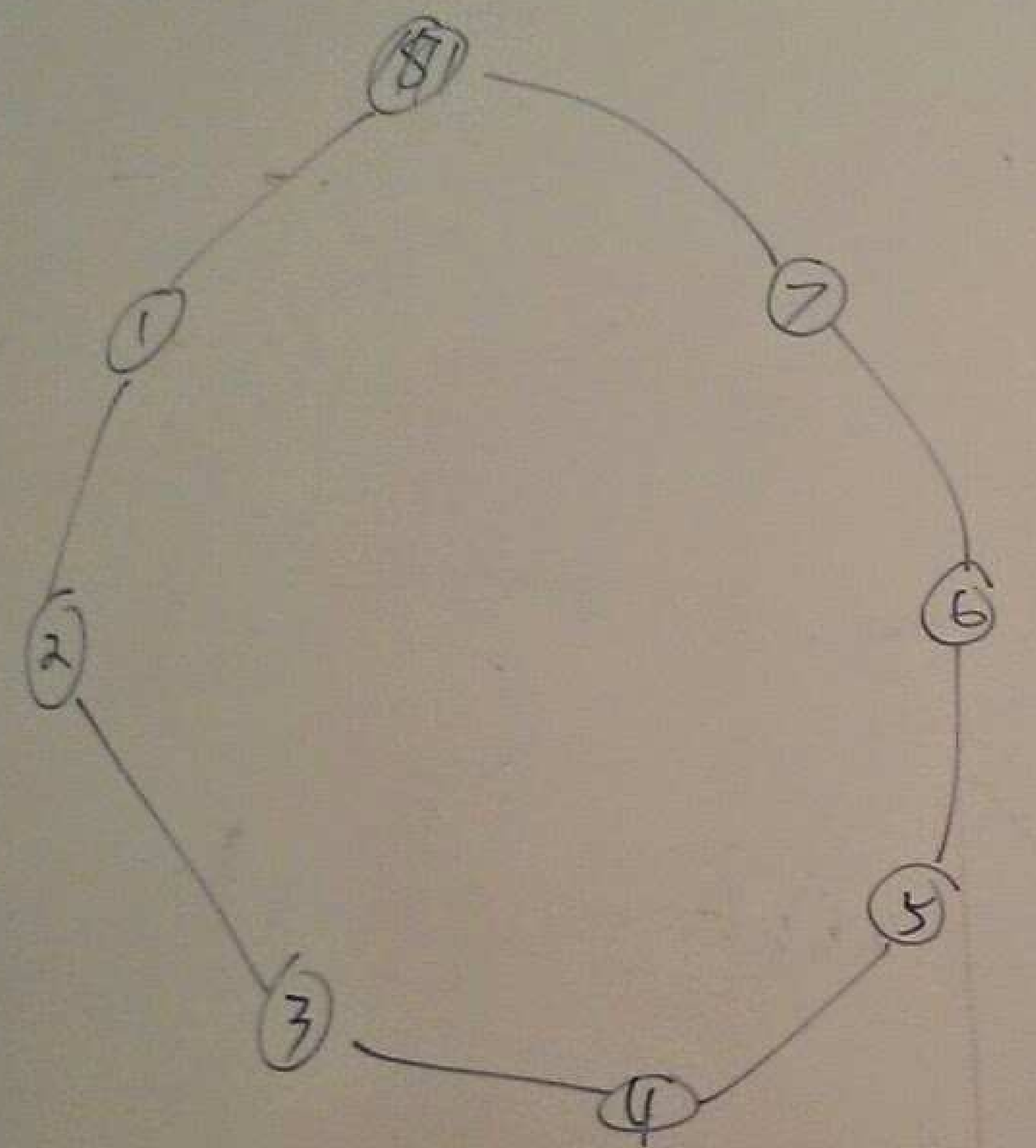
$$= \frac{\phi_{12}(x_1, x_2)\,\phi_{23}(x_2, x_3)}{\sum_{x_1', x_3'} \phi_{12}(x_1', x_2)\,\phi_{23}(x_2, x_3')} = \frac{\tilde{\phi}_1(x_1)\,\tilde{\phi}_3(x_3)}{\sum_{x_1', x_3'} \tilde{\phi}_1(x_1')\,\tilde{\phi}_3(x_3')} \Rightarrow \text{UG}$$

$\tilde{\phi}_1(x_1) = \phi_{12}(x_1, x_2) \quad \tilde{\phi}_3(x_3) = \phi_{23}(x_2, x_3)$

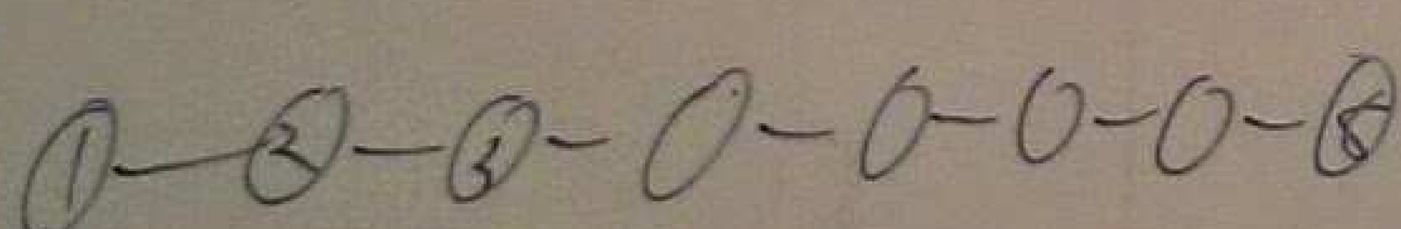


Cutset Conditioning

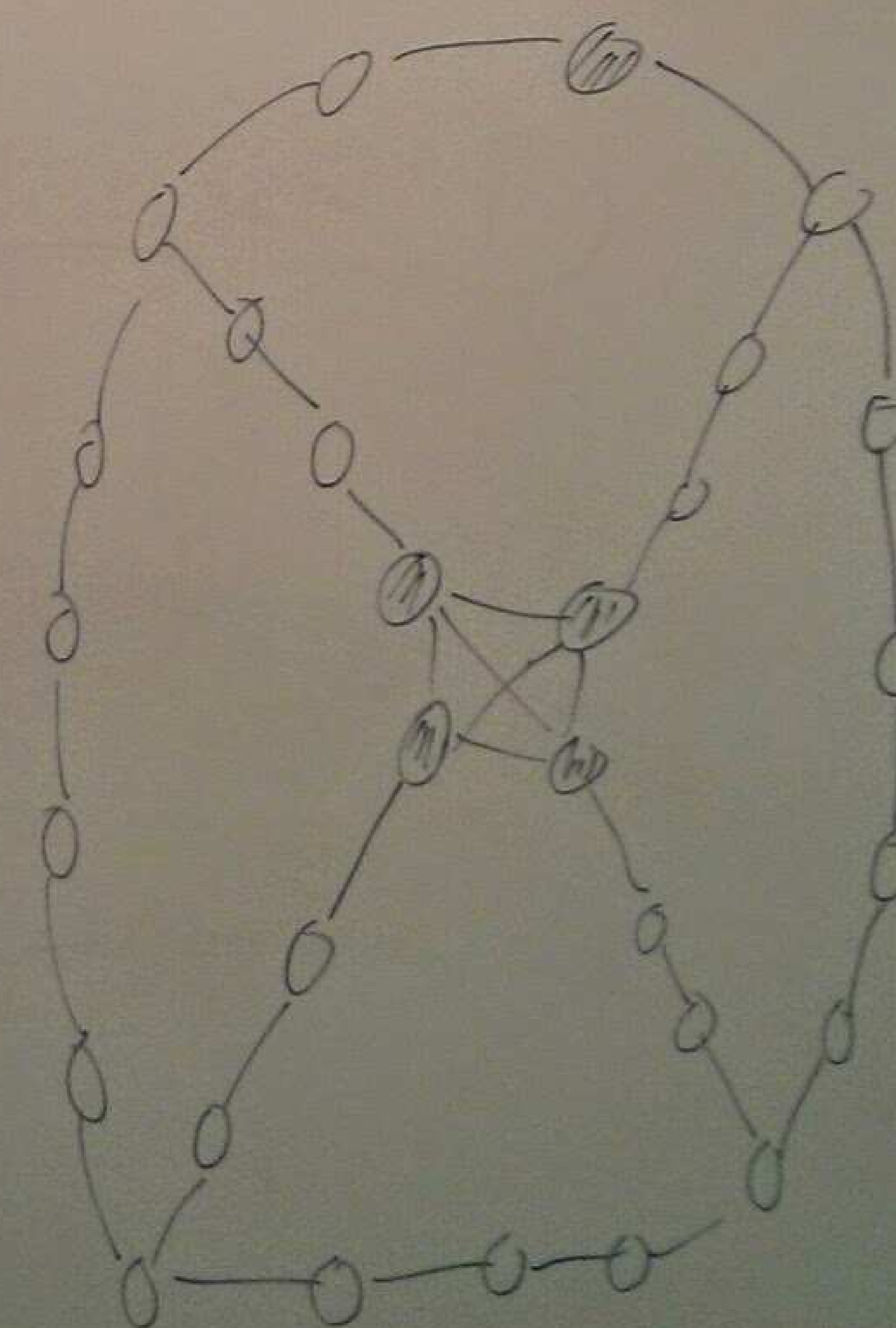
"UBC CZD Loop"



}2



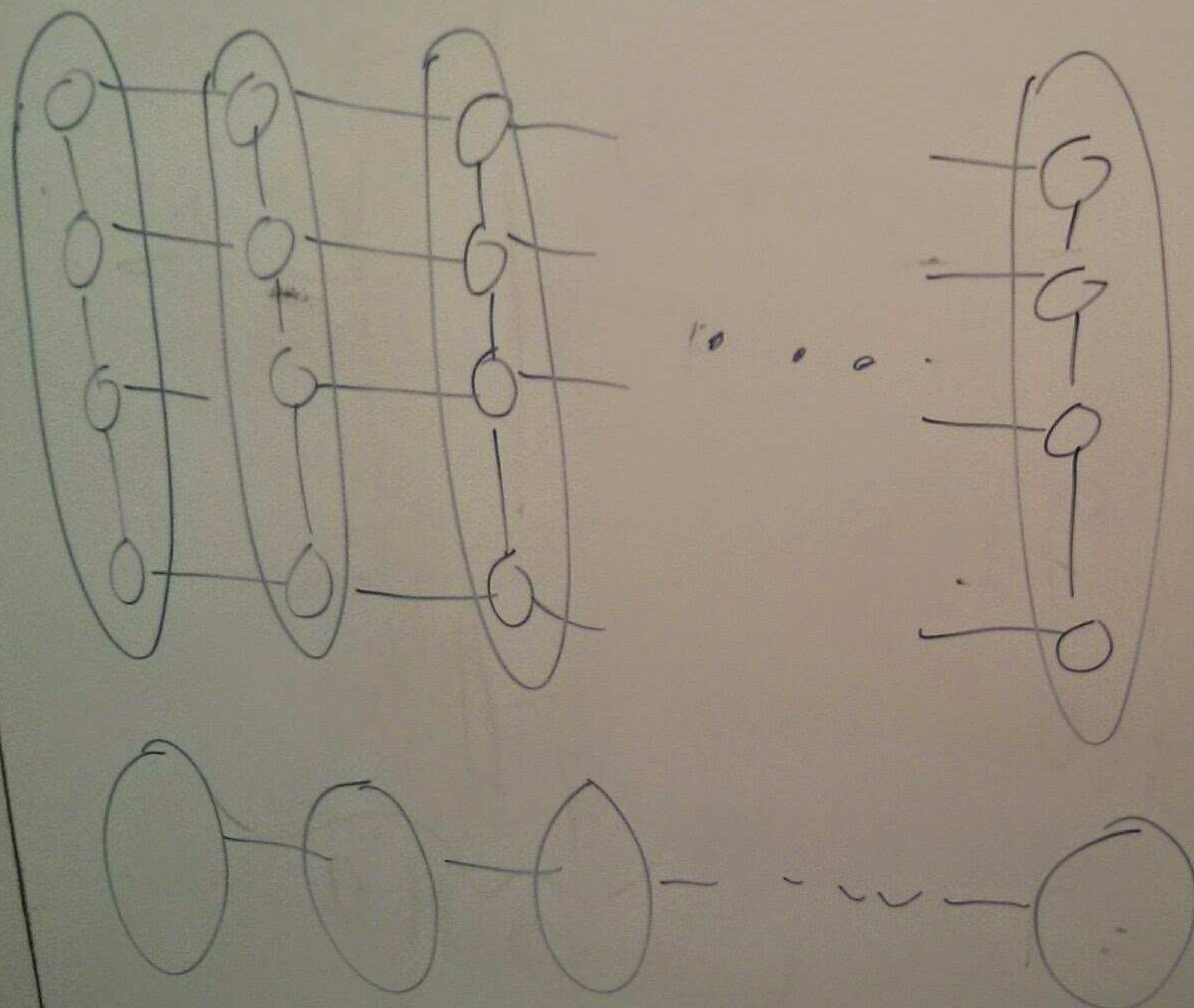
"Paris metro"



2^5

UGM: x_1 x_2

Super Nodes



"time-series with multiple variables"

"plane infection model"

Conditional Inference
Inference in loopy models
Approximate inference

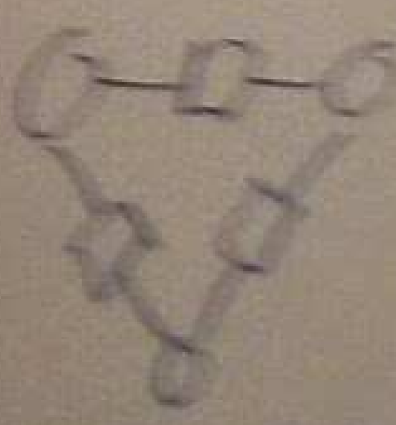
⊛ General graphs: "junction tree algorithm"

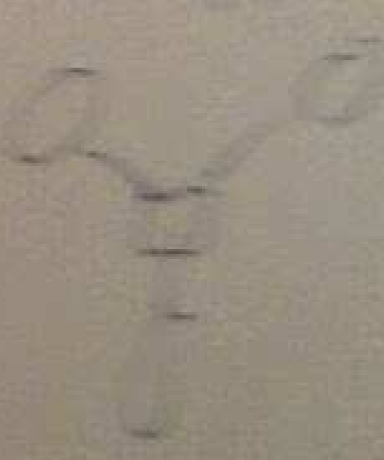
- runtime is exponential in "treewidth" of the graph

⊛ Other graphical models

- decomposable models
- factor graphs

$\psi_{12}\,\psi_{13}\,\psi_{23}$ $\qquad$ ψ_{123}





- hierarchical log-linear models: $\psi_{123} \Rightarrow$ also have $\psi_{12}\,\psi_{13}\,\psi_{23}$ $\Rightarrow$ also have $\psi_1\,\psi_2\,\psi_3$
- chain graphs

- ancestral graphs

Variational Inference

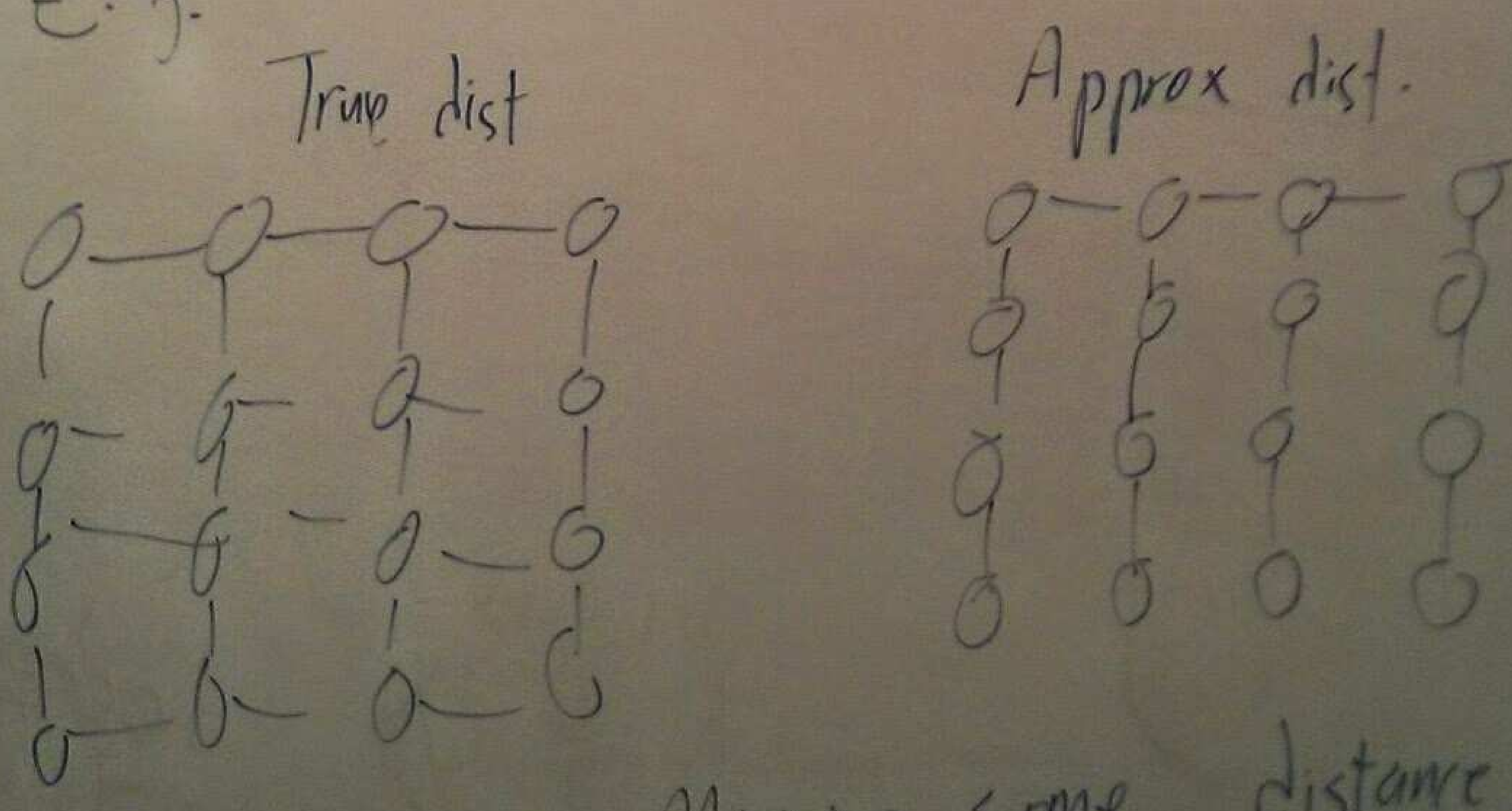
Why? - graphical models w/ high-treewidth

- Bayesian models w/ non-conjugate priors

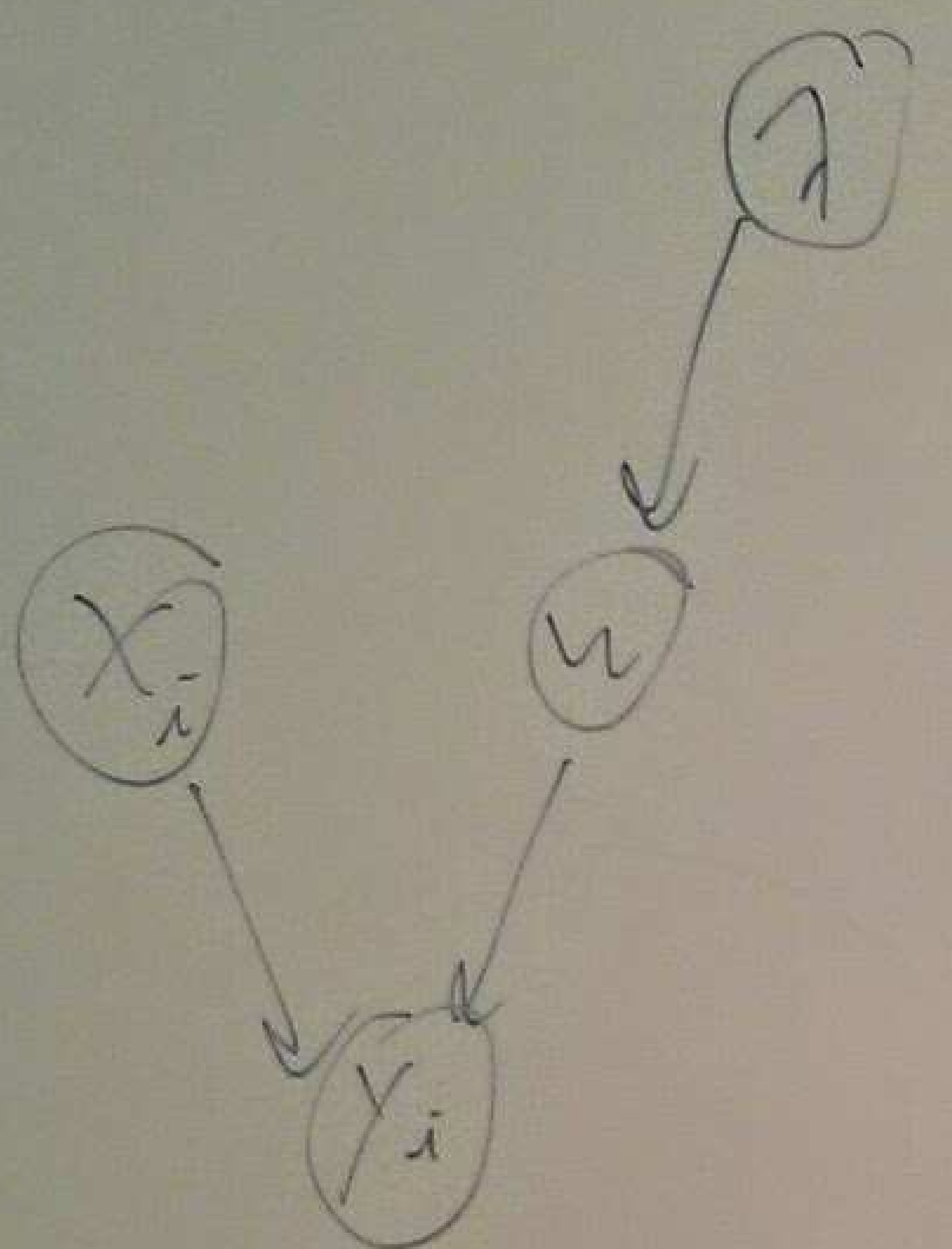
What? - pick approximation $q(x)$ from tractable

(Gaussian, independent Bernoullis,

E.g.

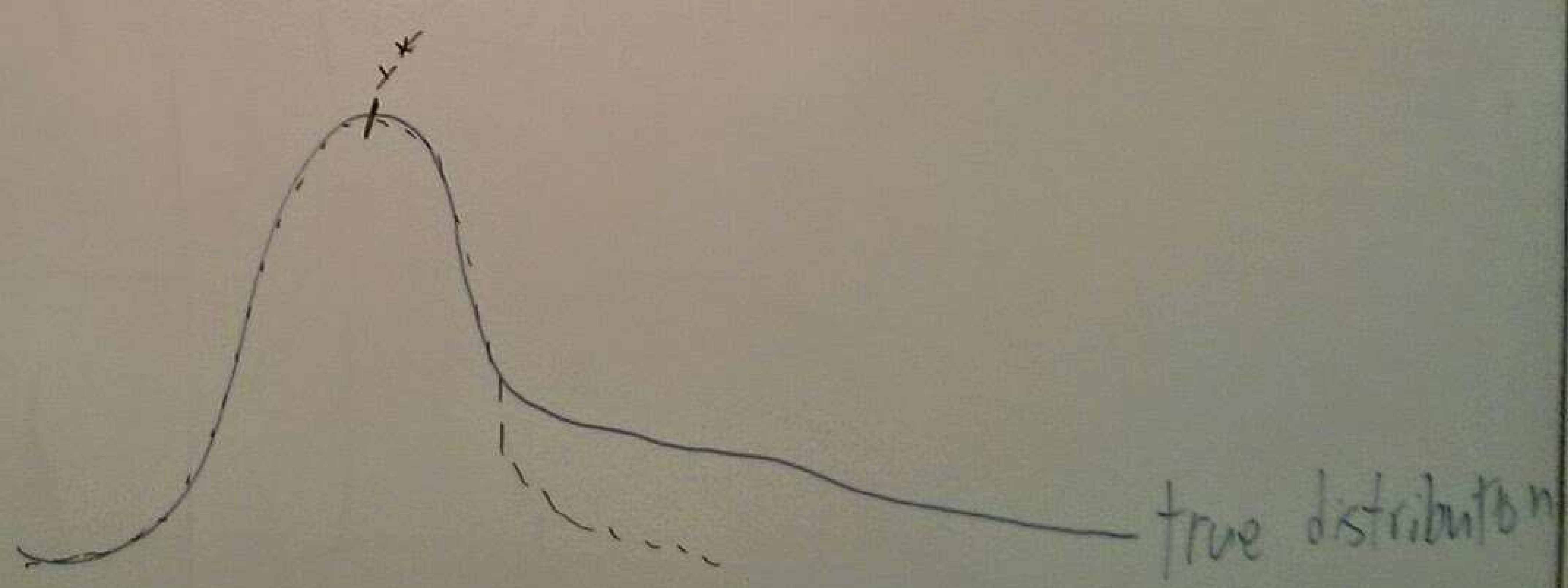


Minimize some distance between $q(x)$ and $p(x)$



e family.
(tree-structural
graphical model)

<u>Laplace approximation</u>: use x^* as mean, and



q: Gaussian

tional Inference
rence in loopy models
oximate inference

Common strategy:

- minimize KL-divergence

$$KL(q \,\|\, p) = \sum_x q(x) \log \frac{q(x)}{p(x)}$$

usually use $p(x) = \frac{\tilde{p}(x)}{Z}$

gives upper-bound on $-\log Z$

"Mean-field" methods ("variational Bayes")

x_1 x_2 x_3 x_4

$$q(x) = \prod_{j=1}^{d} q_j(x_j)$$

Usually, update each q_j sequentially.
(c.f., coordinate descent)

$$q_j(x_j) \leftarrow \frac{1}{Z_j} \exp\Big(\underbrace{E_{-q_j}[\log \tilde{p}(x)]}_{\text{generally easy}} \Big)$$

Ising model: $\psi_{jk} = W_{jk} x_j x_k$

MF

$$q_j(x_j) \propto \exp\Big(x_j \sum_{(j,k)\in E} W_{jk} \mu_k\Big)$$