Learning the Structure of Deep, Sparse Graphical Models

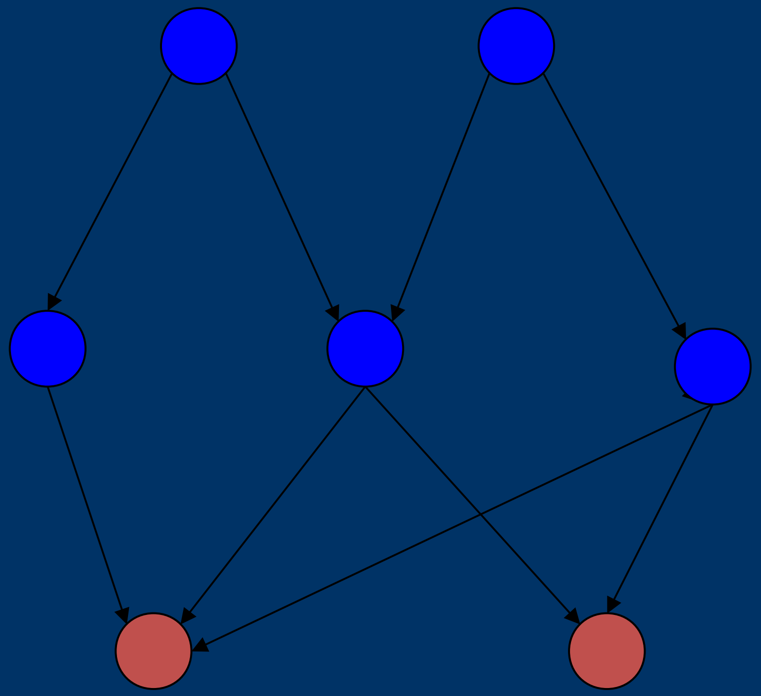
Ryan Prescott Adams, Hanna M. Wallach, Zoubin Ghahramani

Presented by: Soumya Ghosh

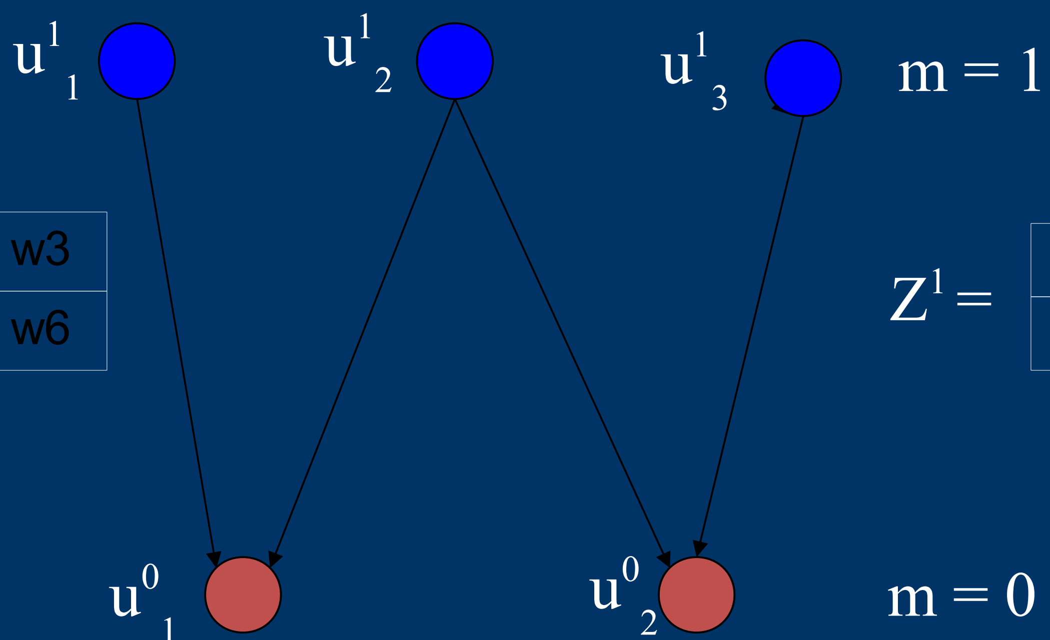
Slides courtesy: Hanna Wallach

Introduction - Belief networks

- Belief Networks = Directed Graphical models.
- Generative model of data.
- Various models covered in class have been belief nets, with fixed known structure.
- This paper aims to learn the structure* in addition to inferring latent variables.



Belief networks - Notation



$W^1 =$

w1	w2	w3
w4	w5	w6

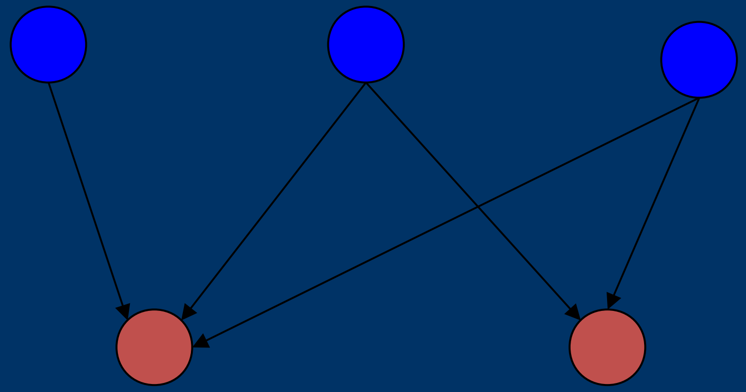
$Z^1 =$

1	1	0
0	1	1

$$u_2^0 = \sigma(w_5 * u_2^1 + w_6 * u_3^1 + \gamma^0 + \epsilon)$$

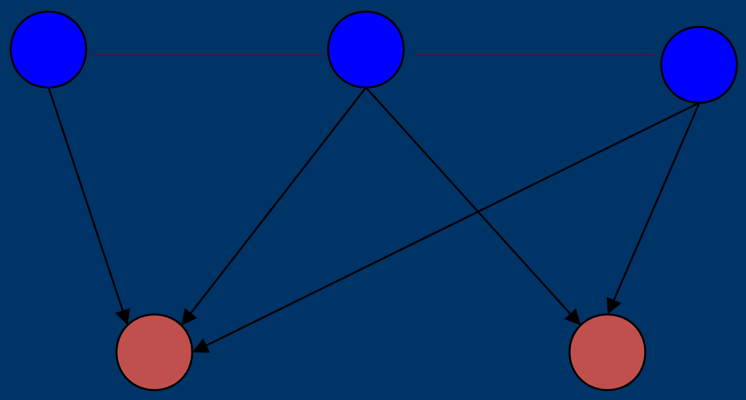
Single Layer networks

- A visible layer and just one hidden layer.
- No intra layer connections are allowed.



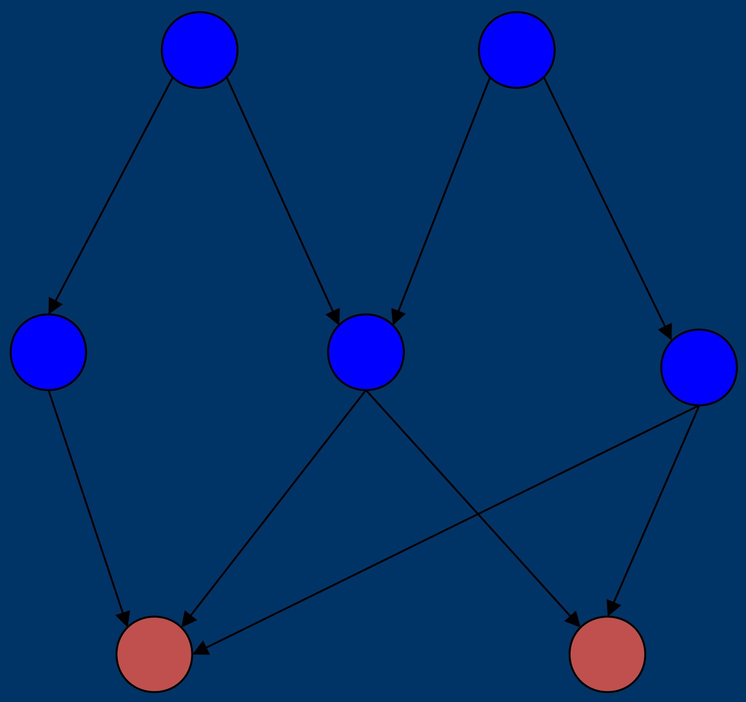
Single Layer networks

- A visible layer and just one hidden layer.
- No intra layer connections are allowed.
- Doesn't model relationships amongst latent variables



Single Layer networks

- A visible layer and just one hidden layer.
- No intra layer connections are allowed.
- Doesn't model relationships amongst latent variables
- Introduce additional layers – *Deep belief networks*



Natural Modeling Questions Arise!

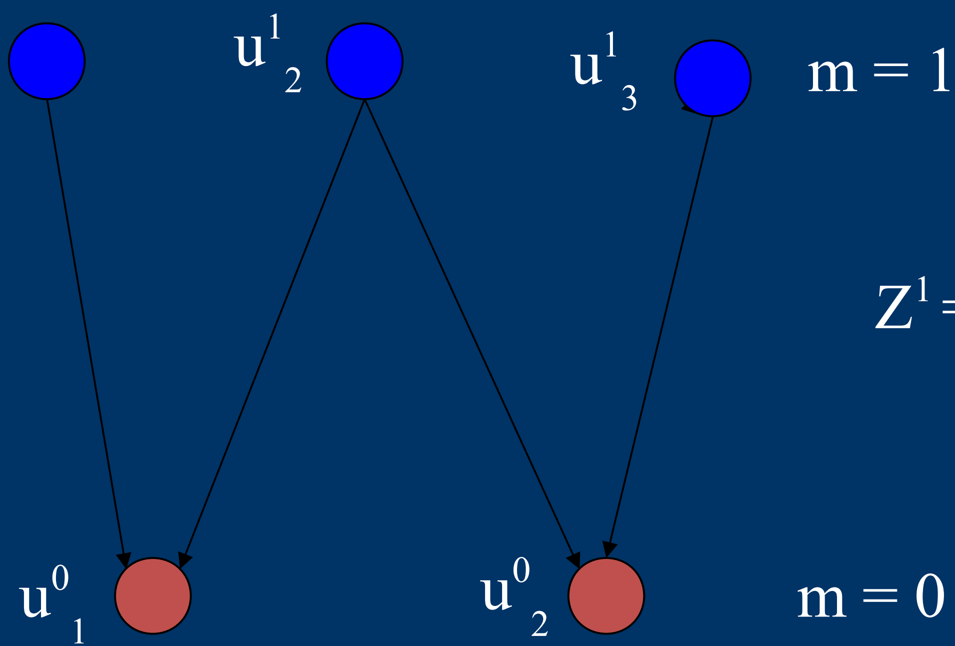
1) How many units per layer?

2) How many connections between layers?

3) How many layers?

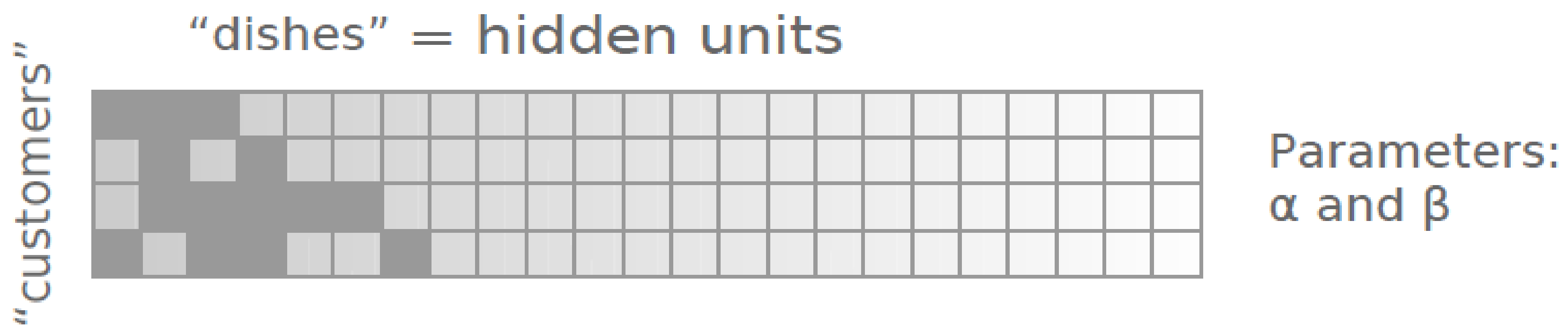
- Paper addresses these questions.

Single layer belief network



$$Z^1 = \begin{array}{|c|c|c|} \hline 1 & 1 & 0 \\ \hline 0 & 1 & 1 \\ \hline \end{array}$$

Infinite Hidden Units -Indian Buffet Process

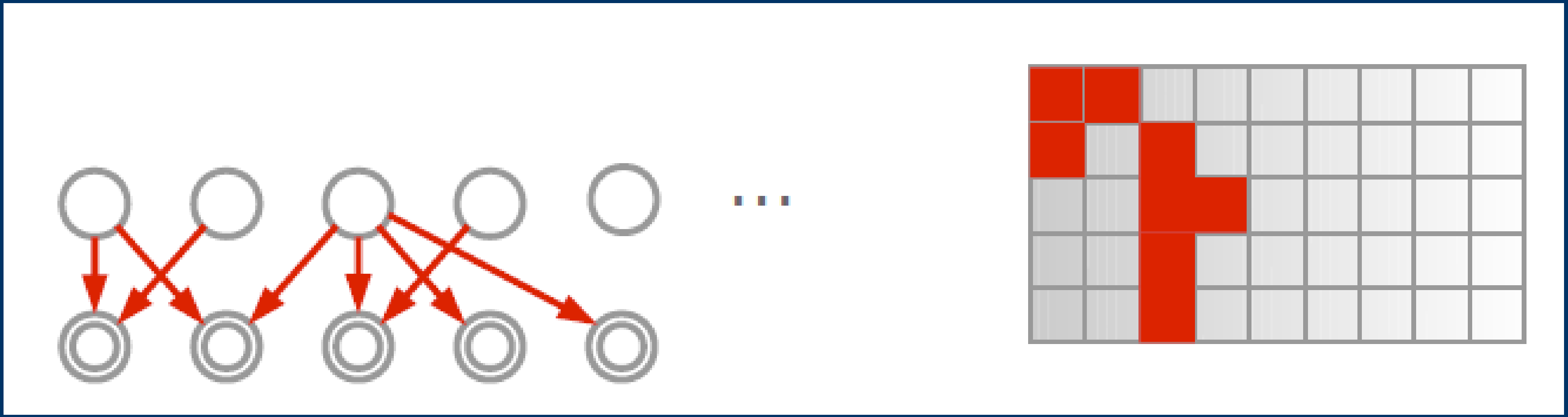


- First customer tries Poisson(α) dishes
- n^{th} customer tries:
  - Previously-tasted dish k with probability n_k / (β + n - 1)
  - Poisson(αβ / (β + n – 1)) completely new dishes

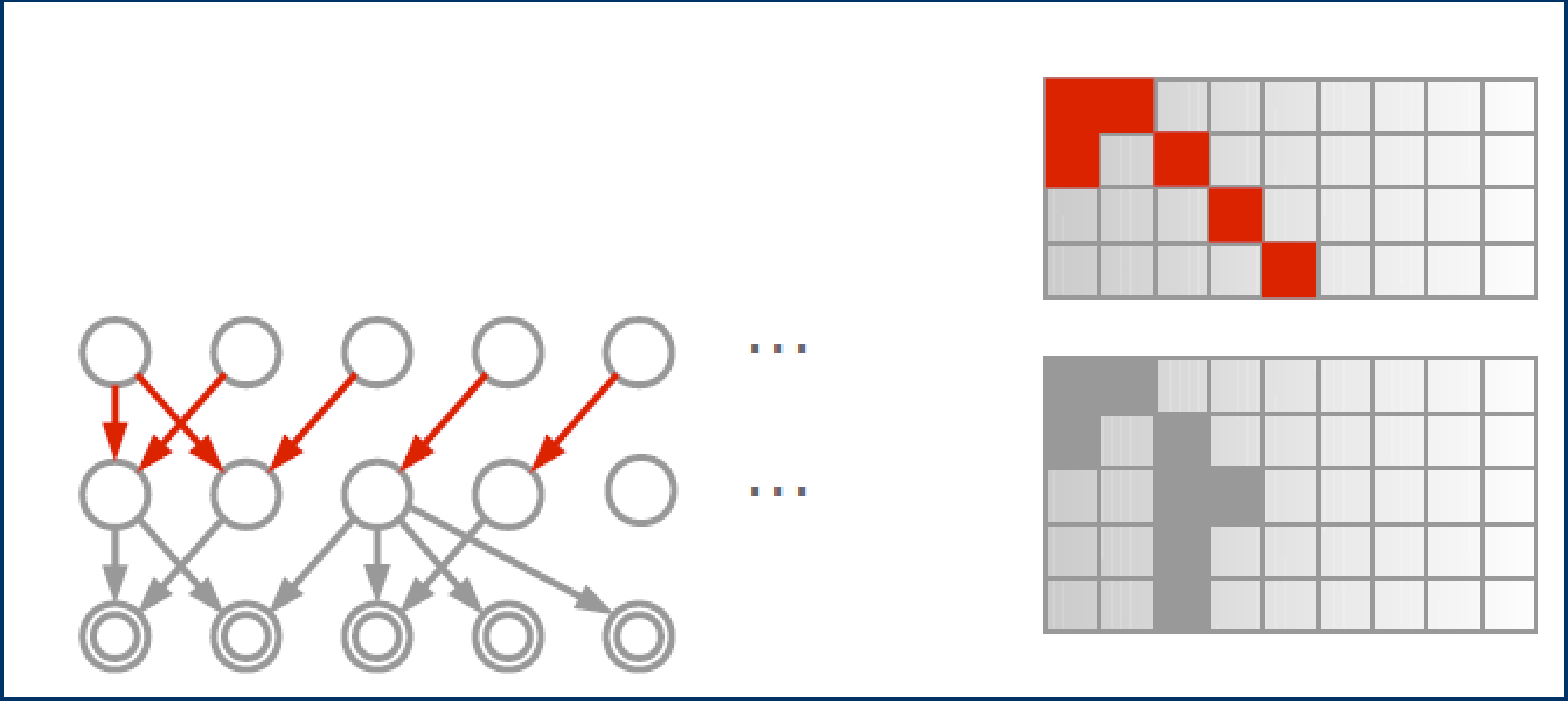
Multi Layered Belief Network

- Use one IBP for each layer.
- Could fix the number of layers, but how about a infinite number of layers?
- Cascading Indian Buffet Process – Infinite sequence of binary matrices.
- Unbounded number of layers each with unbounded number of hidden units.

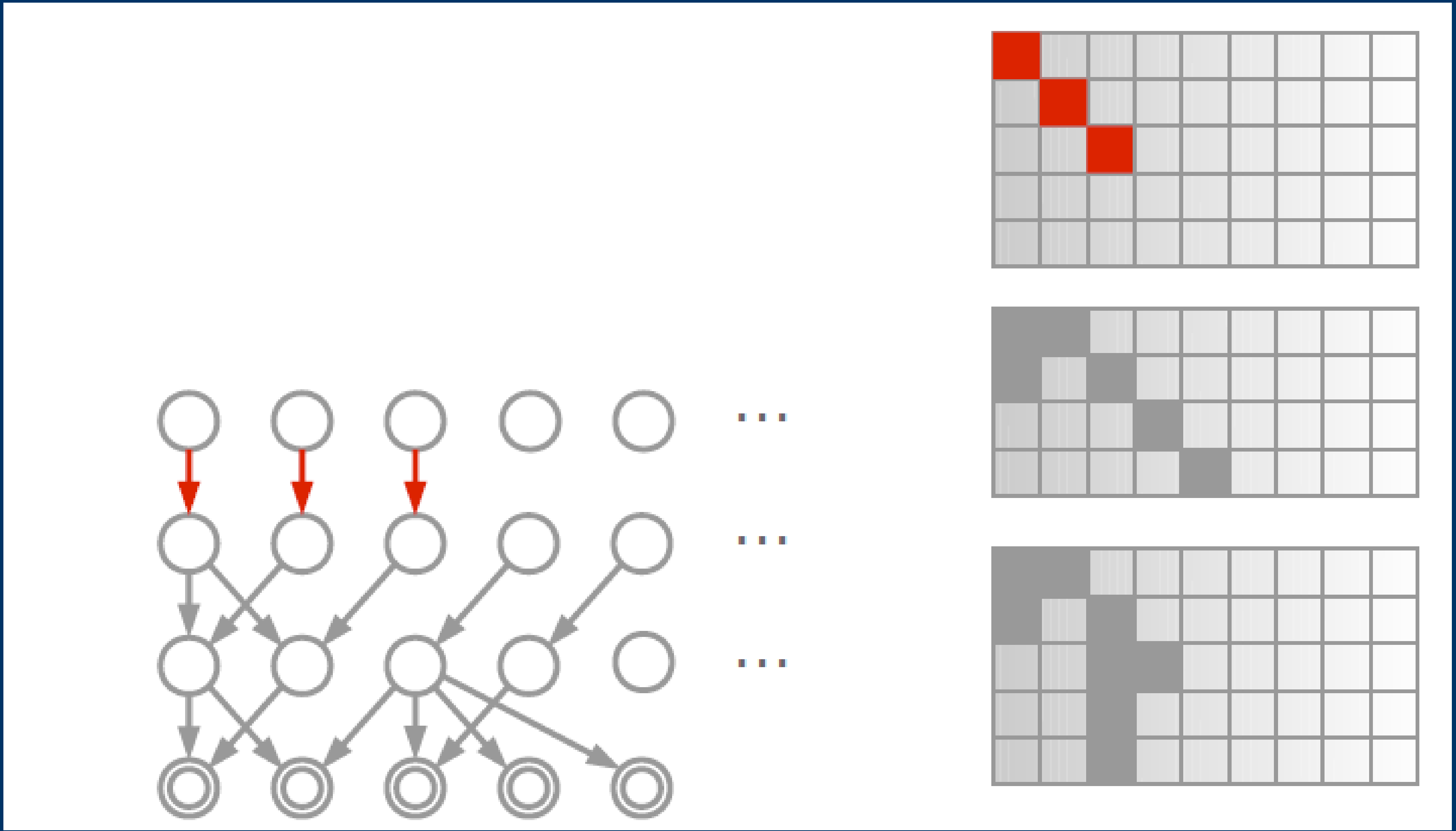
Infinite belief network



Infinite belief network



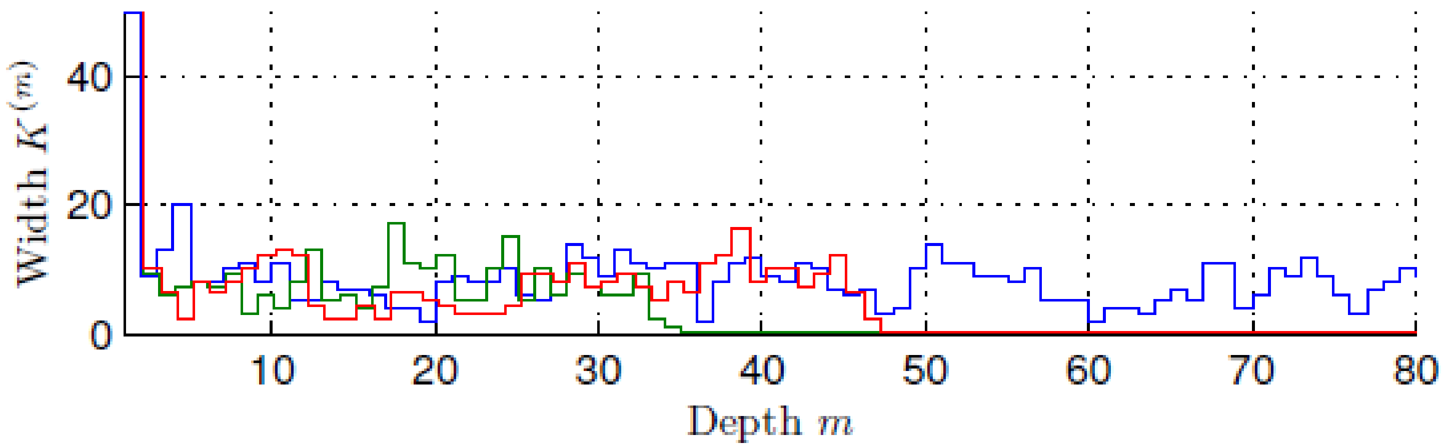
Infinite belief network



Cascading Indian Buffet Process

- Extends the IBP
  - Each dish in the restaurant of layer 'm' is a customer in the restaurant of layer 'm+1'.
- Interestingly, the authors prove that eventually the recursion terminates.
  - Eventually, there is a layer with no units.

Cascading Indian Buffet Process

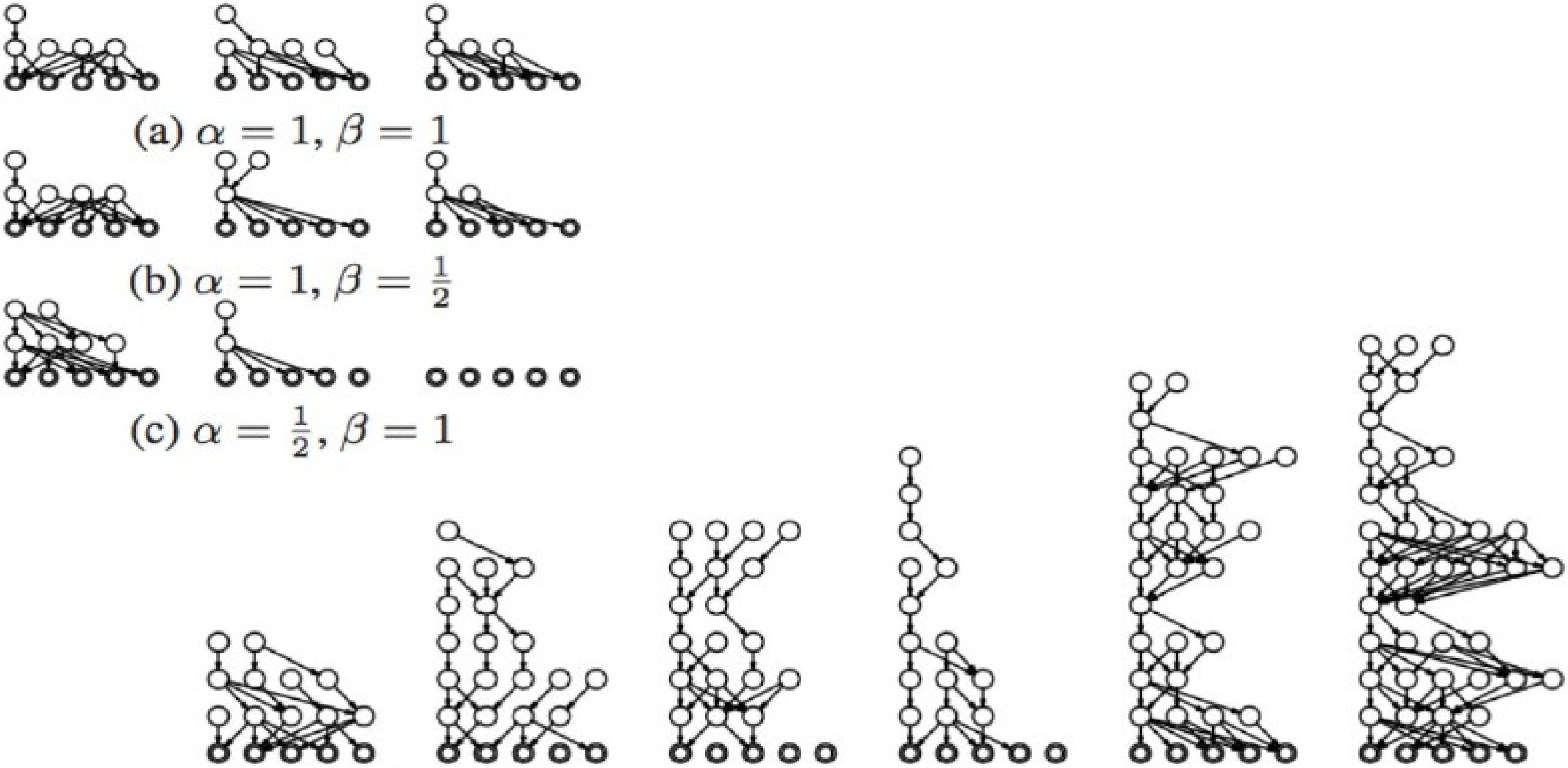


(a) Example traces with $K^{(0)} = 50$

CIBP - Properties

- For a unit in layer m+1
  - Expected # of parents = α
  - Expected # of children = $K / \sum_{k=1}^{K} \frac{\beta}{\beta+k-1}$
- \alpha controls the width of a layer and \beta the number of edges.
- Each layer has it's own \alpha and \beta parameters.

Samples from the CIBP prior



(a) $\alpha = 1, \beta = 1$

(b) $\alpha = 1, \beta = \frac{1}{2}$

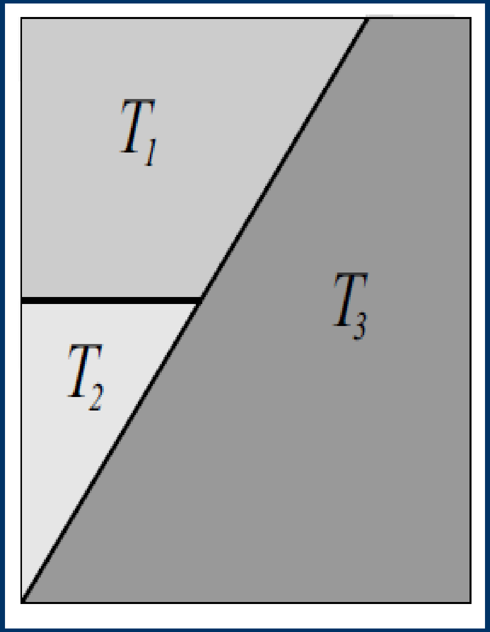
(c) $\alpha = \frac{1}{2}, \beta = 1$

(d) $\alpha = 1, \beta = 2$

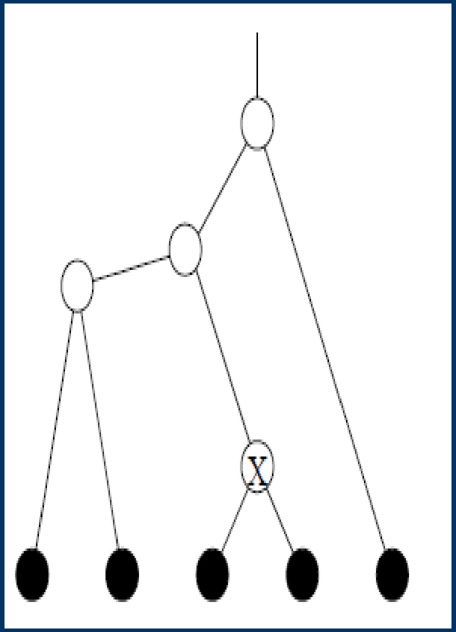
(e) $\alpha = \frac{3}{2}, \beta = 1$

Big Picture

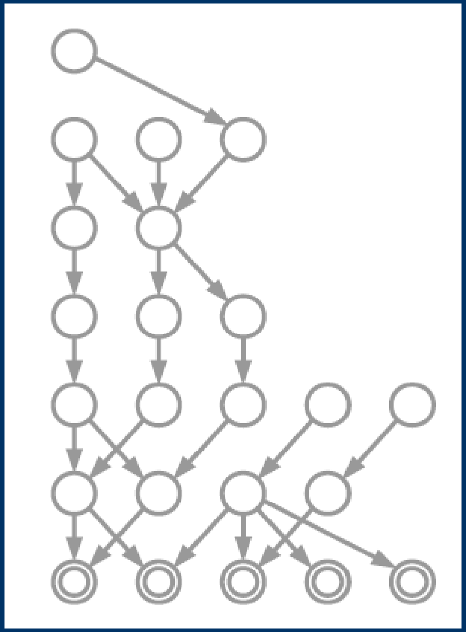
DP



DDT, Kingman's Coalescent



CIBP



Figures courtesy : Erik Sudderth, Radford Neal

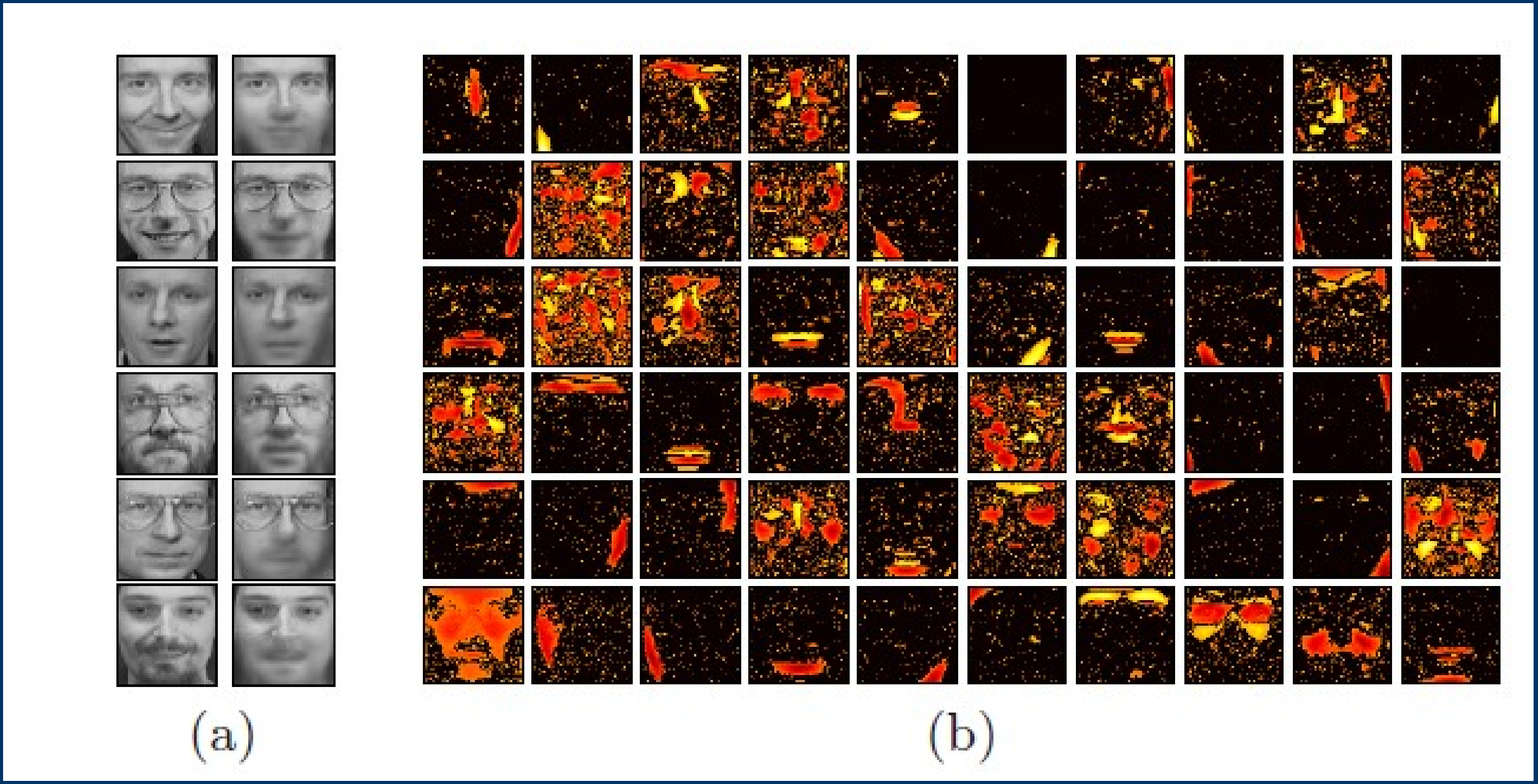
Inference

- MCMC

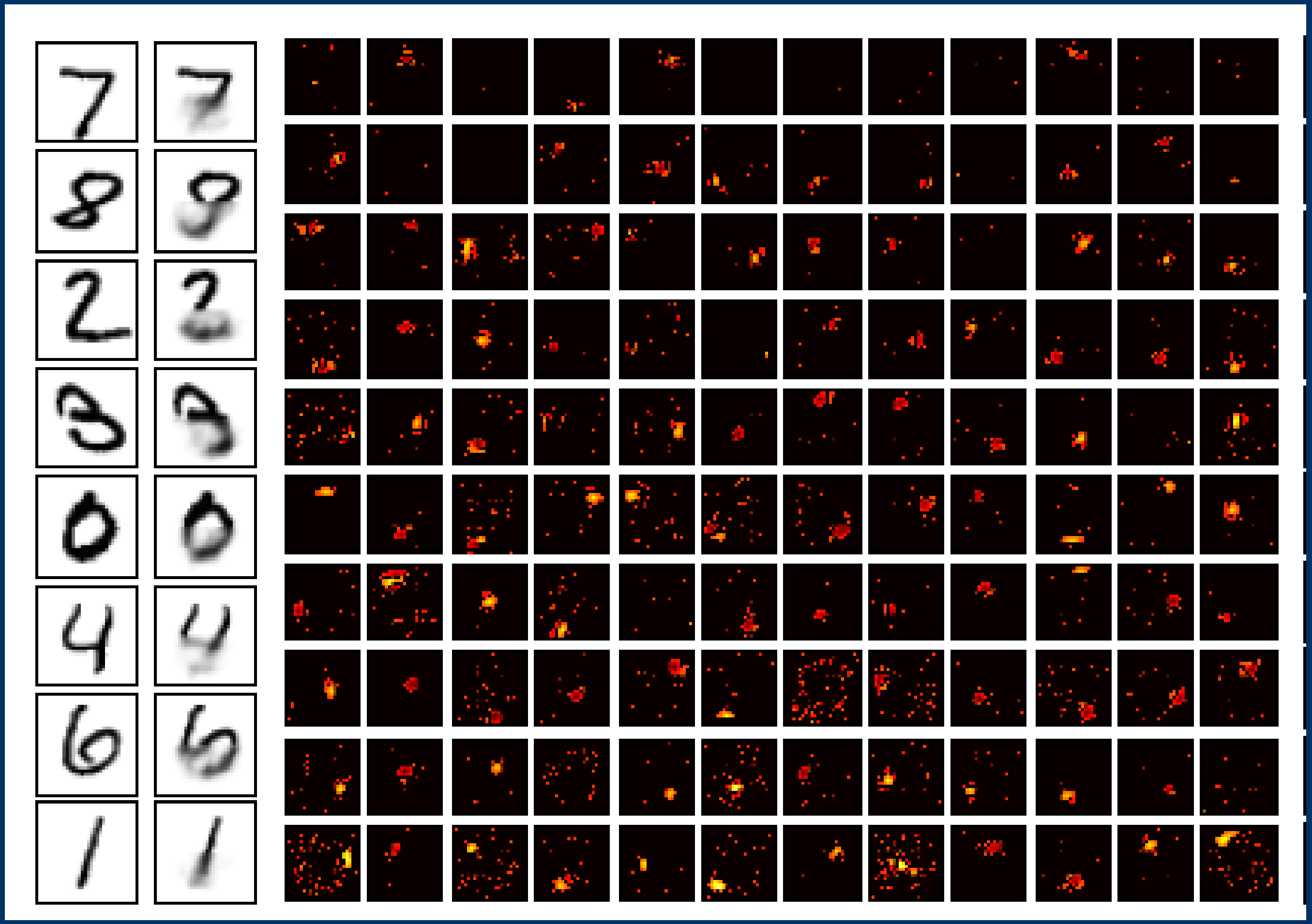
- Conditioned on the structure (Z0, Z1, …) inference is identical to finite belief networks.

- Updating structure.
    - Edges added/deleted using MH.

Results – Image Reconstruction



Results – Digit Reconstruction



Conclusion

- Deep belief networks + Bayesian nonparametrics.
- Introduces a prior over a recursive sequence of binary matrices.
- Allows for unbounded number of units and unbounded number of layers in belief networks.