

## Lecture 11: Exact Inference: Belief Propagation

*Lecturer: Sasha Rush Scribes: Ismail Ben Attallah, Hao Wu, Raphael Rovinov, Ziliang Che, Jiaoyang Huang*

### 11.1 Simple Markov Chain

For a simple Markov Chain with time limit  $T$  and  $V$  classes per node:

$$\begin{aligned} p(y, t) &= \exp\left(\sum_t \theta_t^T(y_{t-1}, y_t) + \theta_t^o(y_t) - A(\theta)\right) \\ &\propto \prod_t \psi_t(y_{t-1}, y_t) \psi_t(y_t) \end{aligned}$$

$$y_1 \longrightarrow y_2 \longrightarrow y_3 \longrightarrow \dots \longrightarrow y_s \longrightarrow \dots \longrightarrow y_T$$

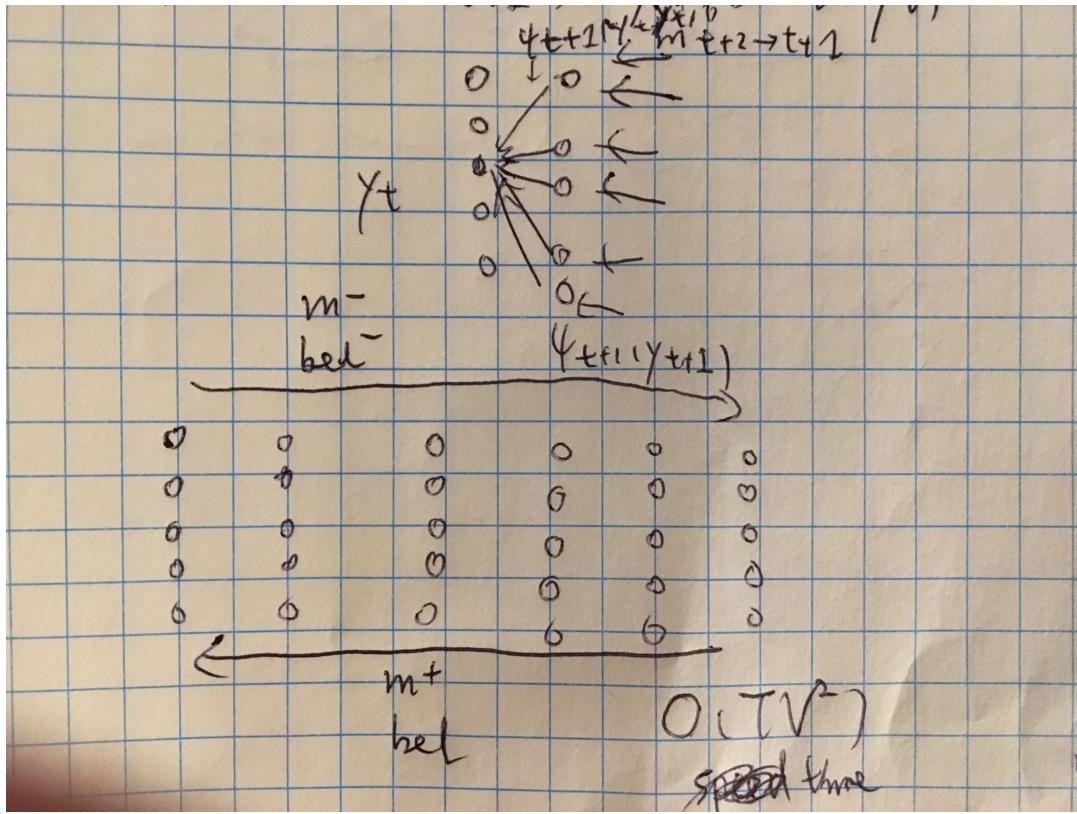
#### 11.1.1 Distributive Property

Marginal:

$$\begin{aligned} p(y_s = v) &= \sum_{y_{1:T}, y'_s=v} \prod_t \psi_t(y'_{t-1}, y'_t) \psi_t(y'_t) / Z(\theta) \\ &= \sum_{y'_T} \psi_T(y'_T) \sum_{y_{T-1}} \psi_{T-1}(y'_{T-1}) \psi_T(y_{T-1}, y'_T) \sum_{y'_2} \dots \sum_{y'_1} \psi_2(y'_2) \sum_{y'_1} \psi_1(y'_1, y'_2) \psi_1(y'_1) \end{aligned}$$

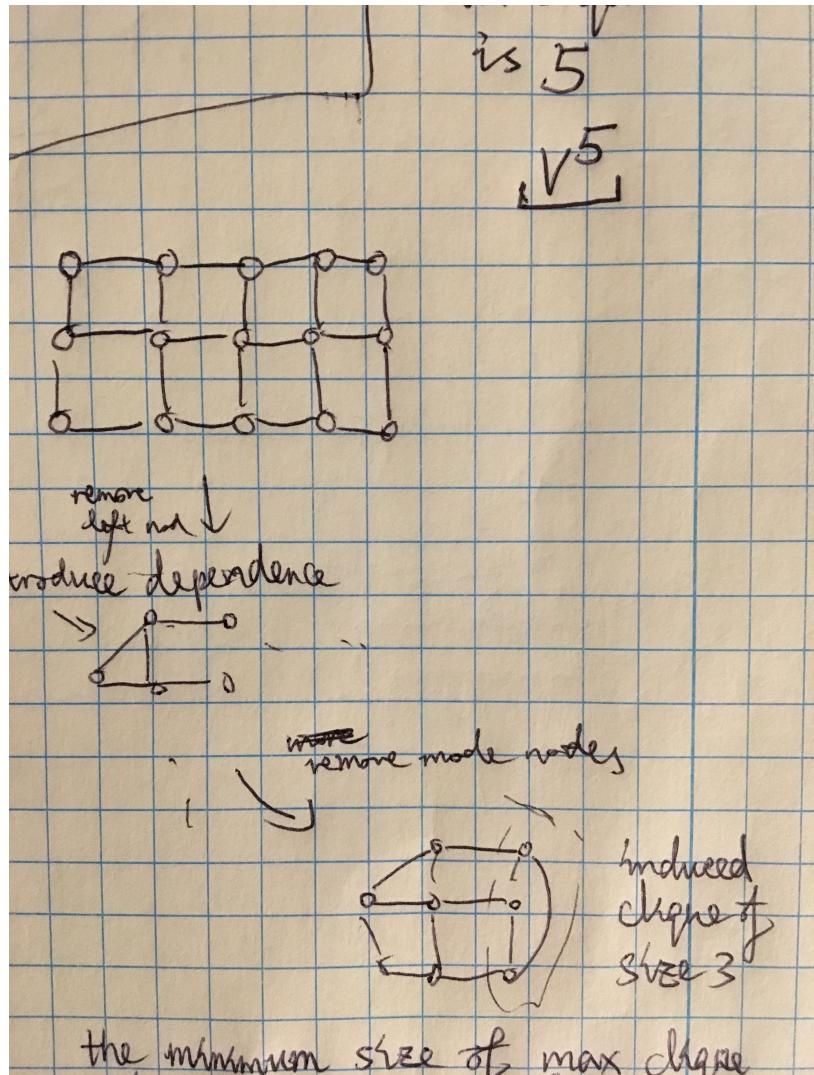
#### 11.1.2 Compute All Marginals with Dynamic Programming

$$\begin{aligned} \text{bel}_t^-(y_t) &\propto \psi_t(y_t) \mathbf{m}_{t-1 \rightarrow t}^-(y_t) \\ \mathbf{m}_{t-1 \rightarrow t}^- &= \sum_{y_{t-1}} \psi_{t-1}(y_{t-1}, y_t) \text{bel}_{t-1}^-(y_{t-1}) \\ \mathbf{m}_{t+1 \rightarrow t}^+ &= \sum_{y_{t+1}} \psi_{t+1}(y_{t+1}, y_t) \psi_{t+1}(y_{t+1}) \mathbf{m}_{t+2 \rightarrow t+1}^+(y_{t+1}) \\ p(y_t) &= \text{bel}_t(y_t) \propto \mathbf{m}_{t+1 \rightarrow t}(y_t) \text{bel}_t^-(y_t) \end{aligned}$$



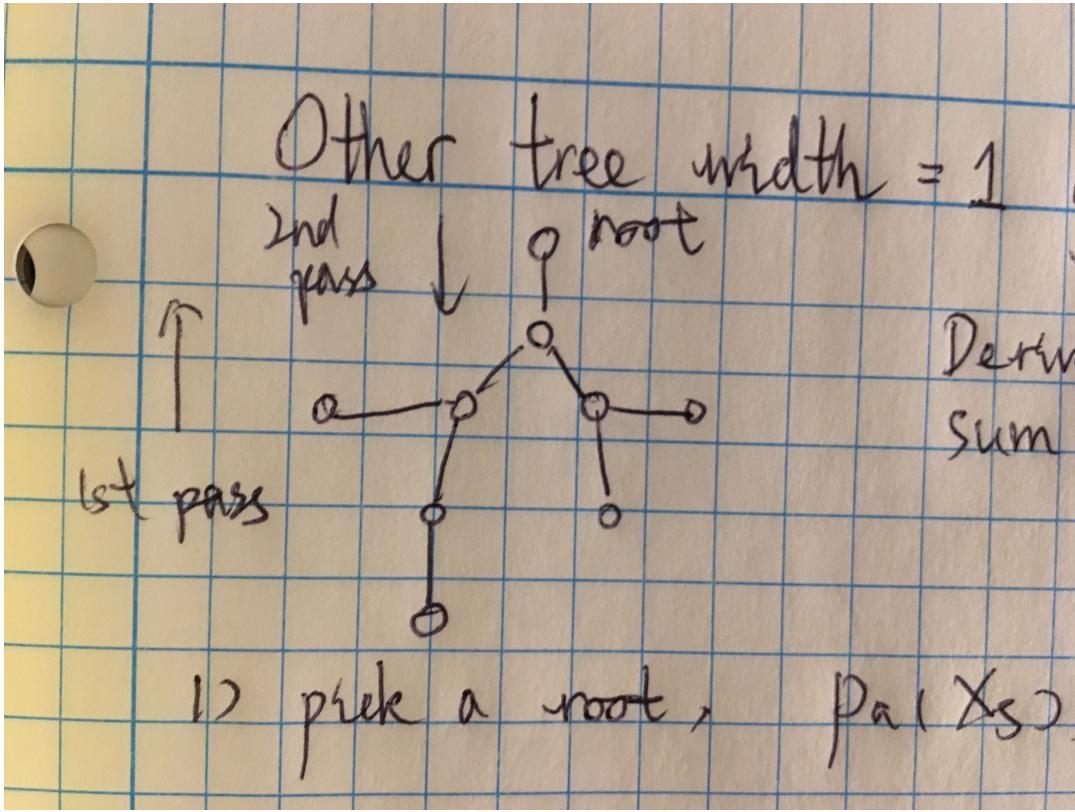
## 11.2 Other Graphs

$$p(y_s = v) = \sum_{y', y'_s = v} \prod_c \psi_c(y'_c)$$



Computing this sum product might be hard. Here we define the minimum size of maximum clique induced -1 to be the *treewidth* of the graph.

### 11.2.1 Compute Marginals for treewidth = 1 Graph



Here we derive the generalization of the forward-backward sum-product algorithm

1. Pick a root  $s$ ,  $\text{pa}(x_s), \text{ch}(x_s)$
2. Upward pass:

$$\begin{aligned} m_{s \rightarrow t}(x_t) &= \sum_{x_s} \psi_{s-t}(x_s, x_t) \text{bel}_s^-(x_s) \\ \text{bel}_t^-(x_t) &\propto \psi_t(x_t) \prod_{s \in \text{ch}(t)} m_{s-t}^-(x_t) \end{aligned}$$

3. Downward pass:

$$\begin{aligned} \text{bel}_s(x_s) &\propto \text{bel}_s^-(x_s) \prod_{t \in \text{pa}(s)} m_{t \rightarrow s}^+(x_s) \\ m_{t \rightarrow s}^+(x_s) &= \sum_{x_t} \psi_{s-t}(x_s, x_t) \psi_t(x_t) \prod_{c \in \text{ch}(t), c \neq s} m_c^-(x_c) = m_t^-(x_t) \end{aligned}$$

### 11.3 Parallel Protocol for Sum Product

$$\begin{aligned} \text{bel}_s(x_s) &\propto \psi_s(x_s) \prod_{t \in \text{nbr}(s)} m_{t \rightarrow s}(x_s) \\ m(s \rightarrow t) &= \sum_{x_s} \psi_s(x_s) \psi_{s-t} \prod_{u \in \text{nbr}(s), u \neq t} m_u^-(x_s) \end{aligned}$$

## 11.4 Final Notes

For this lecture we have been utilizing  $+$  and  $\times$  and distributive property

### 11.4.1 Commutative Semi-ring

$$\begin{array}{lll} + & \max & \cap \\ \times & \times & \cup \\ \vee & \vee & \vee \\ \text{marginal} & \text{argmax} & \text{satisfying assignment} \end{array}$$