COS513: FOUNDATIONS OF PROBABILISTIC MODELS
LECTURE 7

YUHUI LUO AND MASON SIMON

1. Local and Global Consistency

Local consistency in a graphical model means that two neighboring nodes will have the same marginals. **Global consistency** means that two nodes anywhere in the graph have the same marginals.

For a junction tree, local consistency implies global consistency.

2. The Junction Tree Algorithm

Theorem: Define

$$p(x) = \frac{\prod_{c \in \mathcal{C}} \Psi_c(x_c)}{\prod_{s \in \mathcal{C}} \Phi_s(x_s)},$$

where $\Psi(x_c)$ are the clique potentials and $\Phi(x_s)$ are the separator potentials. When the junction tree algorithm terminates, the potentials are equal to marginal probabilities. That is:

$$\Psi_c(x_c) = p(x_c)$$

and

$$\Phi_s(x_s) = p(x_s).$$

Proof:
First, if $\Psi_c(x_c) = p(x_c)$, then $\Phi_s(x_s) = p(x_s)$ by local consistency.

The remaining proof is by induction: We first assume that this holds for a tree of size N, where N represents the number of cliques. Now, consider a tree of size $N+1$ as depicted in Figure 1.

The additional clique C^* is represented by the node on the left, which is split into two subsets: S^*, which are the elements shared with its neighbor, and R, which are the remaining elements. We further define T to be the elements included in the dashed box, which is basically everything except R and S^*.

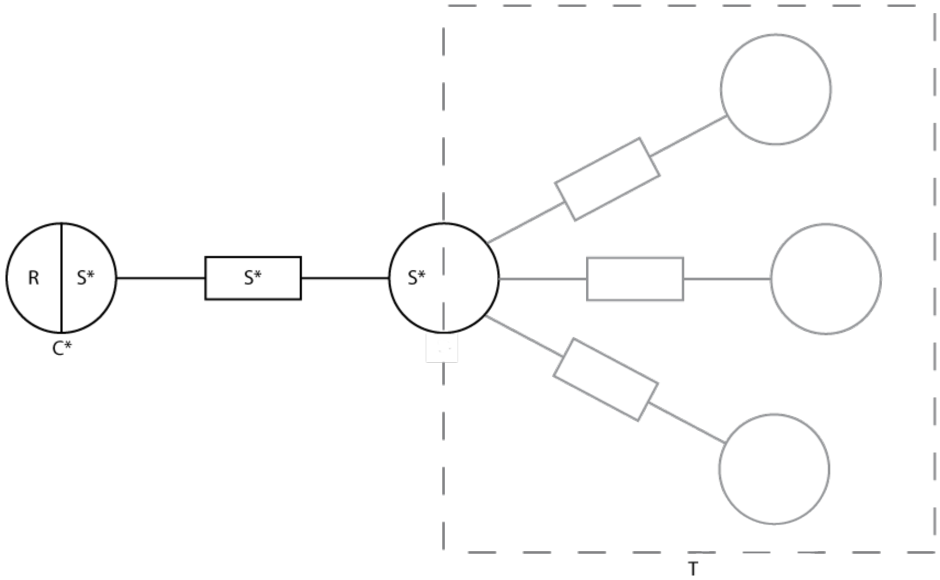

FIGURE 1. A tree of size $N+1$, formed by adding the node C^* to the tree T of size N

By the chain rule,

$$\begin{aligned} p(x) = p(x_R, x_{S^*}, x_T) &= p(x_{S^*}, x_T)p(x_R|x_{S^*}, x_T) \\ &= p(x_{S^*}, x_T)p(x_R|x_{S^*}). \end{aligned}$$

The last line is obtained from the fact R and T are conditionally independent given S^* since they are separated by S^*.

Now we look at:

$$\begin{aligned} p(x_T, x_{S^*}) &= \sum_{x_R} p(x) \\ &= \sum_{x_R} \frac{\prod_c \Psi_c(x_c)}{\prod_s \Phi_s(x_s)} \\ &= \frac{\sum_{x_R} \Psi_{C^*}(x_{C^*})}{\Phi_{S^*}(x_{S^*})} \frac{\prod_{c \notin C^*} \Psi_c(x_c)}{\prod_{s \notin C^*} \Phi_s(x_s)}. \end{aligned}$$

That final step followed because the only clique potential that depends on x_R is Ψ_{C^*}.

By local consistency:

$$\sum_{x_R} \Phi_{C^*}(x_{C^*}) = \Phi_{S^*}(x_S^*),$$

since everything that is in C^* except R is S^*. Because of that equality, the first fraction in $p(x_T, x_{S^*})$ equals 1, so we're left with:

$$p(x_T, x_{S^*}) = \frac{\prod_{c \notin C^*} \Psi_c(x_c)}{\prod_{s \notin C^*} \Phi_s(x_s)}.$$

Note that x_R does not appear in that equation, so $p(x_T, x_{S^*})$ does not depend on x_R. This implies that x_R will not send any messages into T. Because T is a graph of size N, the inductive hypothesis tells us that all of the clique potentials in T are marginals, i.e. $\Psi_c(x_c) = p(x_c)$ for $c \notin C^*$.

Now we can compute $\Phi_{S^*}(x_S^*)$ by marginalizing the other variables out of x_D:

$$\begin{aligned}
\Phi_{S^*}(x_S^*) &= \sum_{x_D \setminus x_{S^*}} \Psi_D(x_D) \\
&= \sum_{x_D \setminus x_{S^*}} p(x_D) \\
&= p(x_{S^*})
\end{aligned}$$

By the definition of the joint distribution,

$$\begin{aligned}
p(x_R | x_{S^*}) &= \frac{\Psi_{C^*}(x_{C^*})}{\Phi_{S^*}(x_{S^*})} \\
&= \frac{\Psi_{C^*}(x_{C^*})}{p(x_{S^*})} \\
p(x_R | x_{S^*}) p(x_{S^*}) &= \Psi_{C^*}(x_{C^*}) \\
p(x_R, x_{S^*}) &= \Psi_{C^*}(x_{C^*}) \\
p(x_{C^*}) &= \Psi_{C^*}(x_{C^*}).
\end{aligned}$$

And that shows that the $N + 1$th clique (C^*), and the separator between that clique and the tree of size N (S^*), both have potentials which are marginals. By the inductive hypothesis, the tree T of size N already had clique and separator potentials which were marginals, so now every clique and separator in the new tree of size $N + 1$ has a potential which is a marginal. QED.

3. How to build a Junction Tree

Now that you know the neat properties of the Junction Tree algorithm, you may wonder how to actually build a Junction Tree when given a graphical model. It's not hard.

(1) Moralize your graph (if undirected).
(2) Introduce evidence.
(3) Triangulate (the Graph Eliminate algorithm from chapter 3 in ITGM will do this).
(4) Construct the junction tree by choosing the spanning tree of this triangulated graph that maximizes the sum of sizes of the separator sets.

4. Recommended Reading Materials and Tools

Here are a list of recommended reading materials and tools for the midterm report and final project.

- **Machine Learning** JMLR, NIPS, ICML, UAI, CVPR, EMNLP
- **Statistics** JASA, BA (Bayesian Analysis), AAS (Annals of Applied Statistics), AS (Annals of Statistics - Theoretical)
- **Books**
  - Gelman: Bayesian Data Analysis, Hierarchical and Multi-Level Modeling
  - Tibshirani, Friedman, Hastie: Elements of Statistical Learning + online errata
  - Bishop: Pattern Recognition and Machine Learning + online errata
  - Manning and Schutze
  - Robert and Casella: MCMC
  - Mackay: Information Theory (free and online)
- **Languages**: C, Python (Cython), R (RCommander, RJava).

5. Frequently Asked Questions about R

Q: Can you make wonderful plots with R?
A: Yes, check out the R graph gallery.
Q: Is R's built-in text editor inferior to Notepad?
A: Yes, use a real text editor instead of that thing!
Q: "On the topic of R, is it easy to inject C code into the nuts...?" - Francisco Pereira
A: Yes, it is easy to call C code from R.