

Probabilistic Graphical Models

Aayush Bajaj

August 31, 2026

Contents

1 the pages	1
2 reading order	2

a probabilistic graphical model is a marriage of graph theory and probability: the graph records which variables talk to each other directly, and that structure alone dictates how compactly a joint distribution can be stored, which independencies it must satisfy, and how expensive it is to reason with. one set of ideas — factors, elimination, message passing, sampling — then serves diagnosis networks, spam filters, speech recognisers, image denoisers and gene-regulation maps alike.

this section grew out of unsw's comp9418 (advanced topics in statistical machine learning), which is taught from darwiche's modeling and reasoning with bayesian networks.

the course also covers graph neural networks — those live with the deep learning notes, not here

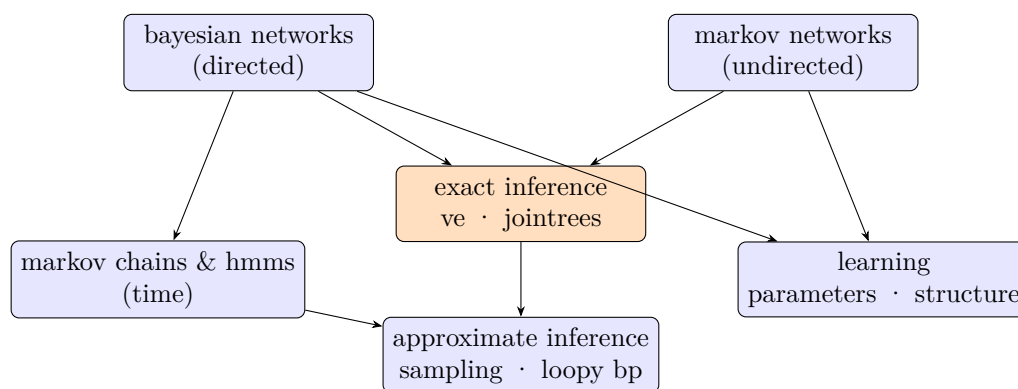


Figure 1: the map of this section: two representation languages, the inference algorithms that run on both, the temporal specialisations, and learning closing the loop from data back to models.

1 the pages

- bayesian networks — the directed language: cpts, the chain rule, d-separation and its valves, markov blankets, naive bayes and friends, gaussian networks, a word on causality.
- markov networks — the undirected language: gibbs distributions and the partition function, hammersley-clifford, factor graphs, energy minimisation, crfs.
- exact inference — factors and variable elimination, elimination orders and treewidth,

the jointree algorithm, mpe and map queries.

- approximate inference — the sampling ladder from forward sampling to gibbs, and loopy belief propagation.
- markov chains and hidden markov models — sequence models: stationary distributions, forward and viterbi, particle filters, dbns and kalman.
- learning graphical models — mle and smoothing, em for missing data, chow-liu trees and score-based structure search.

2 reading order

the pages are written to be read in the order listed: representation first (directed, then undirected), then the inference stack (exact, then approximate), then time, then learning. each page closes with a curated *results* list — the statements worth being able to reproduce cold — and a typeset pdf for printing.