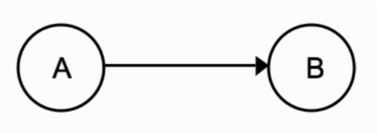
graphical models

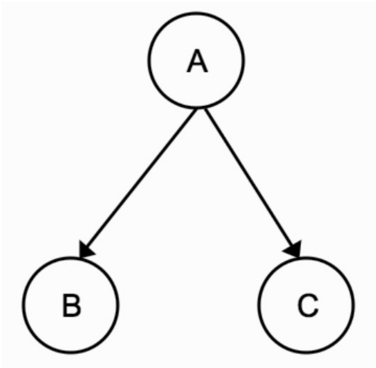
Graphical models use graph theory with all its underlying computational and mathematical machinery in order to explain probabilistic models.

‘A’ causes ‘B’

graph tells us that the probability of A and B is the same
as the probability of A times the probability of B conditioned on
A: $P(a, b) = P(b|a)P(a)$.

Two types: Directed graph, Adirected graph



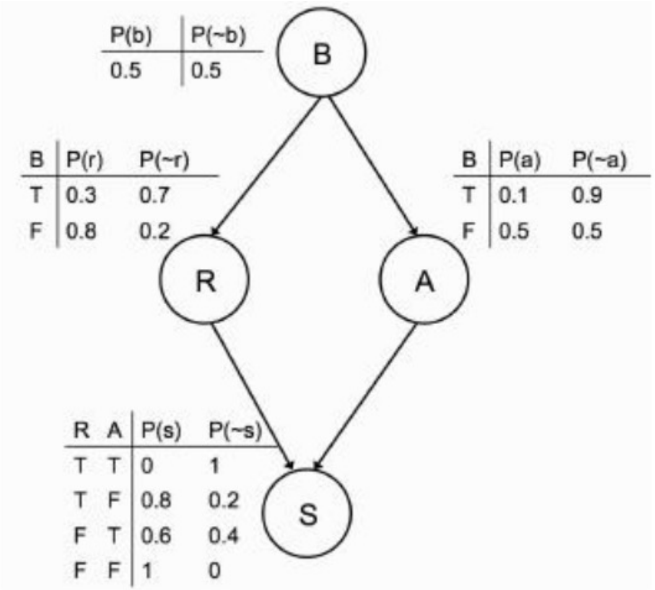


16.1 BAYESIAN NETWORKS

Directed graphs must not contain cycles, that is, there cannot be any loops in the graphs. These graphs go by the rather unlovely name of DAGs: directed, acyclic graphs, but for graphical models, when they are paired with the conditional probability tables, they are called Bayesian networks.

16.1.1 Example: Exam Fear
It is a handy guide to whether or not you will be scared (S) before an exam based on whether or not the course was boring (‘B’), which was the key factor you used to decide whether or not to attend lectures (‘A’) and revise (‘R’)

Perform inference to decide the likelihood of you being scared before the exam (‘S’)



Two kinds of inference: depending on whether the observations that are made come from the top of the graph or the bottom.

Conditional probability:

Probability of being scared

$$
\begin{aligned}
P(s) &= \sum_{b,r,a} P(b,r,a,s) \\
&= \sum_{b,r,a} P(b) \times P(r|b) \times P(a|b) \times P(s|r,a) \\
&= \sum_{b} P(b) \times \sum_{r,a} P(r|b) \times P(a|b) \times P(s|r,a).
\end{aligned}
$$

$$
\begin{aligned}
P(s) &= 0.3 \times 0.1 \times 0 + 0.3 \times 0.9 \times 0.8 + 0.7 \times 0.1 \times 0.6 + 0.7 \times 0.9 \times 1 \\
&= 0.328.
\end{aligned}
$$

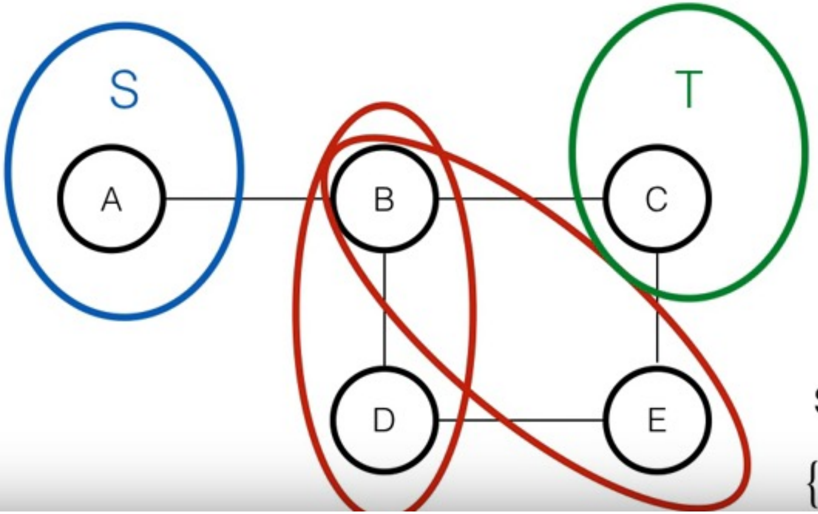
Markov Random Fields

- Any two subsets S and T of variables are conditionally independent given a **separating subset**
  - All paths between S and T must travel through the separating subset

paths:

A-B-C

A-B-D-E-C



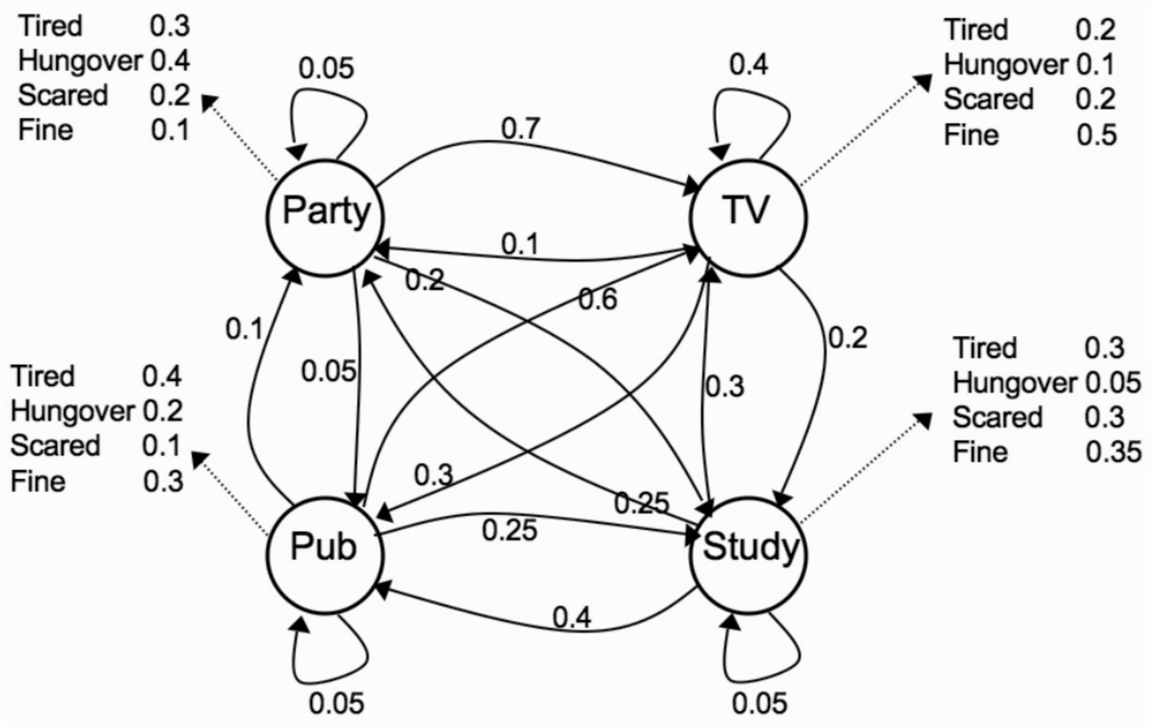
separating subsets

{B,D}, {B,E}, {B,D,E}

HIDDEN MARKOV MODELS (HMMS)

-It is used in speech processing and in a lot of statistical work.

- Its power comes from the fact that it deals with situations where you have a Markov model, but you do not know exactly which state of the Markov model you are in. Instead, you see observations that do not uniquely identify the state.



Observation guess probability

$o(t)$	$\omega(t)$	$P(o_k(t)\|\omega_j(t))$ $b_j(o_k)$
ω_j	ω_i	$P(\omega_j(t+1)\|\omega_i(t))$ $a_{i,j}$

I know that you did something last night, so $\sum_j a_{i,j} = 1$

and I know that I will make some observation (since if you aren't in the lecture I'll assume you were too tired), so $\sum_k b_j(o_k) = 1$