Probabilistic Graphical Models

Dae-Ki Kang

Probabilistic Graphical Models

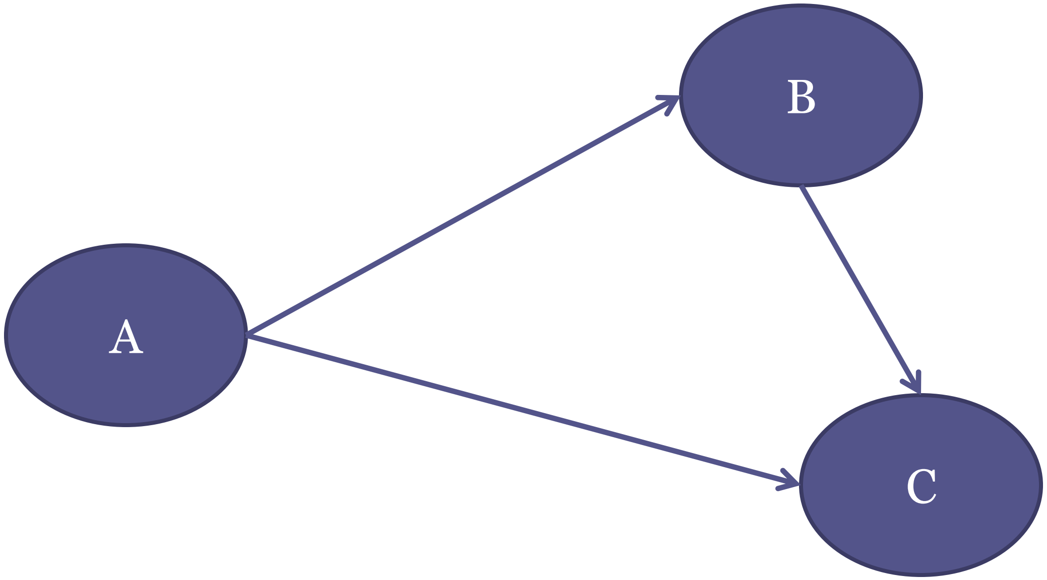
- = Probability Theory + Graph Theory

Probability

- Sum rule (or marginalization)
  - $P(A) = \sum_B P(A,B)$
- Product rule
  - $P(A,B) = P(B|A)P(A)$
- $P(A|B) = P(B|A)P(A)/P(B)$
- $P(A)= \sum_B P(A|B)P(B)$

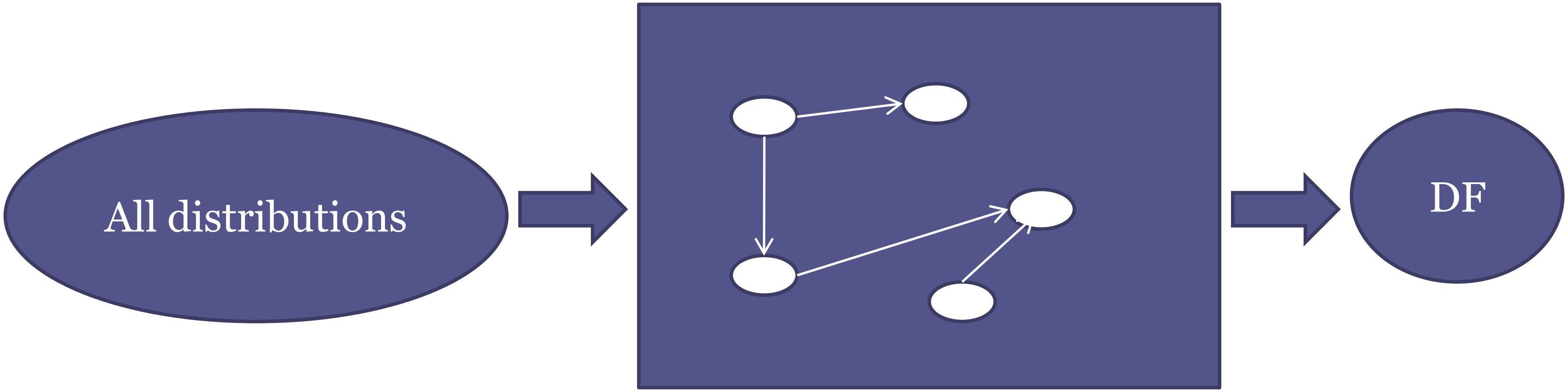
Decomposition

- $P(A,B,C) = P(A)P(B|A)P(C|A,B)$
- Chaining the product rule
  - $P(X1,X2,X3,..,Xn)=P(X2,X3,..,Xn|X1)P(X1)=P(X3,..,Xn|X2,X1)P(X2|X1)P(X1)=...=P(Xn|Xn-1...X2,X1)...P(X1)$



Directed Acyclic Graph

- General Factorization
  - $P(X1,...Xn)=\prod P(Xi|Parent\_i)$
- Directed Factorization (DF)



Example of Directed Graph

- Hidden Markov models
- Kalman filters
- Factor analysis
- Probabilistic principal component analysis
- Independent component analysis
- Mixtures of Gaussians
- Probabilistic expert systems
- Sigmoid belief networks
- Hierarchical mixtures of experts
- Etc.

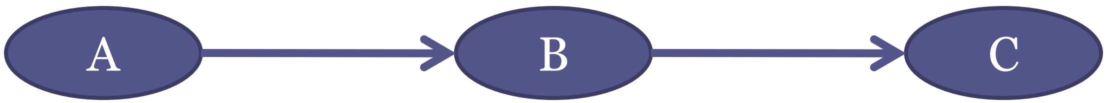
Conditional Intependence

- If A is independent of B given C
  - $A \perp B | C$
  - $P(A|B,C)=P(A|C)$
  - $P(A,B|C)=P(A|B,C)P(B|C)=P(A|C)P(B|C)$

Markov Properties

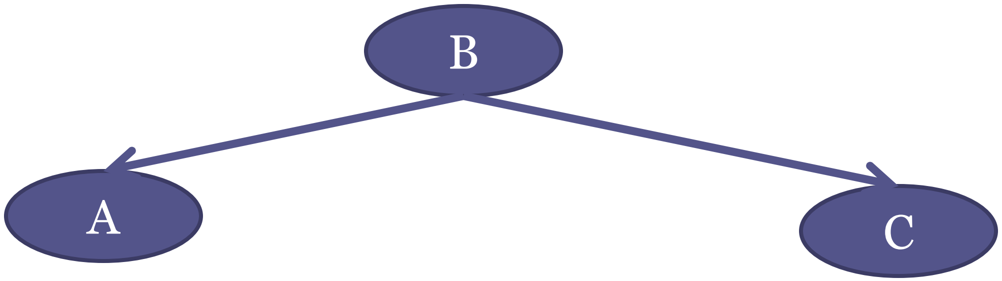
- We can determine the conditional independence properties of a distribution from its graph
  - d-separation

Markov Properties: Example 1



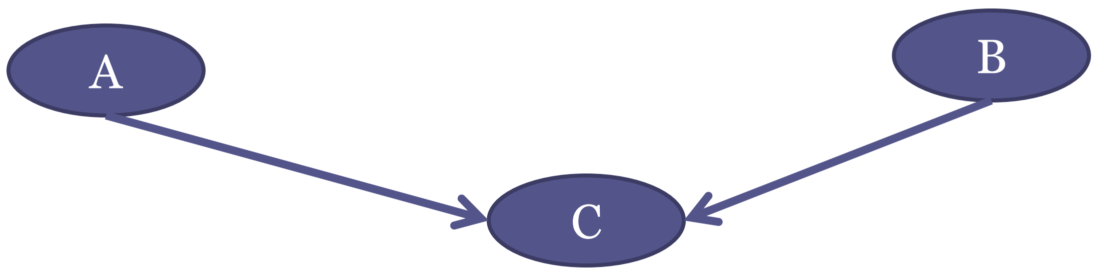
- head-to-tail
- P(A,B,C)=P(A)P(B|A)P(C|B)
- If we condition on node B
  - P(A,C|B)=P(A|B)P(C|B)
  - Therefore A⊥C|B
- If we haven't observed B
  - A¬⊥C|ϕ
- Observation of B "blocks the path" from A to C

Markov Properties: Example 2



- tail-to-tail
- P(A,B,C)=P(B)P(A|B)P(C|B)
- If we condition on node B
  - P(A,C|B)=P(A|B)P(C|B)
  - Therefore $A \perp C | B$
- If we haven't observed B
  - $A \neg\perp C | \phi$
- Observation of B "blocks the path" from A to C

Markov Properties: Example 3



- head-to-head
- $P(A,B,C)=P(A)P(B)P(C|A,B)$
- If we don't observe C
  - $P(A,B)=P(A)P(B)$
  - $A \perp B | \phi$
- If we condition on node C
  - $P(A,B|C) \neq P(A|C)P(B|C)$
  - Therefore $A \neg\perp B | C$
- Unobserved node C "blocks the path" from A to B
- Observation of C "unblocks the path" from A to B

Implication of d-separation

- A graphical property of Bayesian networks
- Implication: If two sets of nodes X and Y are d-separated in Bayesian networks by a third set Z (excluding X and Y), the corresponding variable sets X and Y are independent given the variables in Z

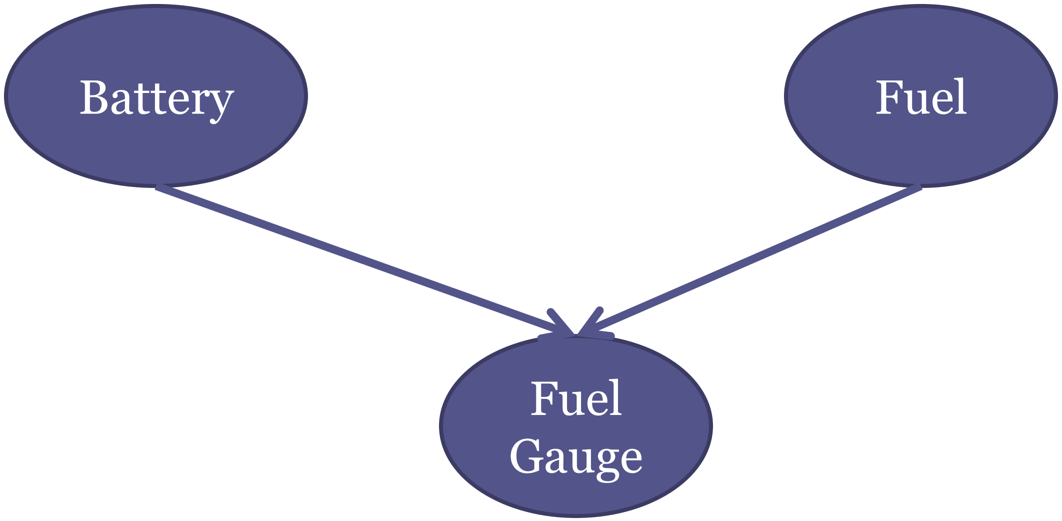
Definition of d-separation

- Two sets of nodes X and Y are d-separated in Bayesian networks by a third set Z (excluding X and Y) if and only if every path between X and Y is “blocked”, where the term “blocked” means that there is an intermediate variable V (distinct from X and Y) in Z such that:
  - The connection through V is “tail-to-tail” or “tail-to-head” and V is instantiated
  - Or, the connection through V is “head-to-head” and neither V nor any of V’s descendants have received evidence.

Markov blanket (MB)

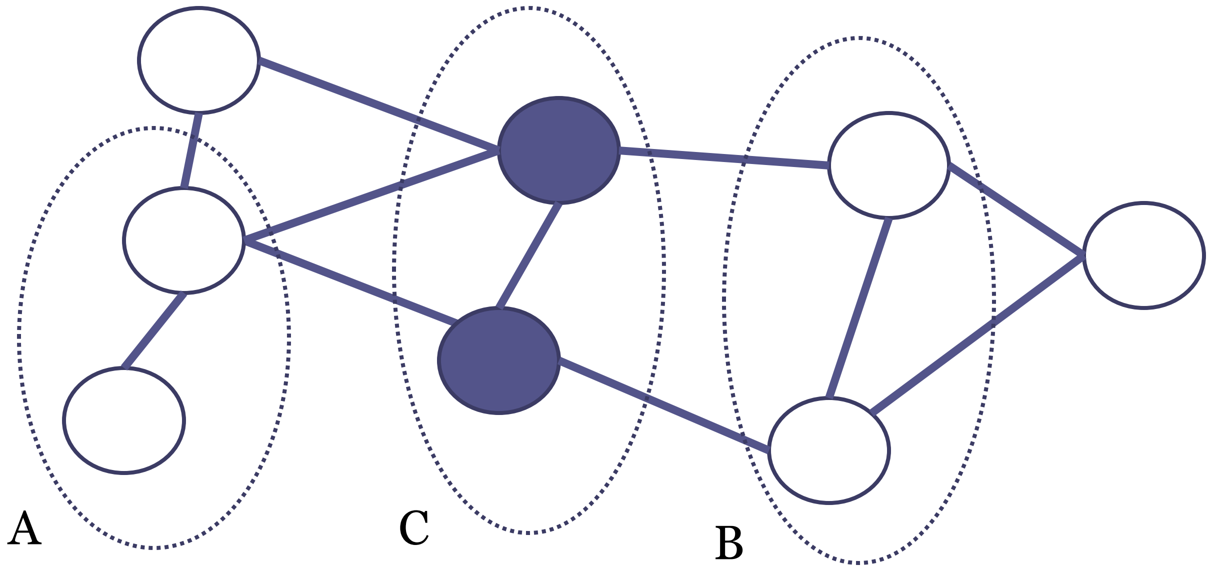
- The minimal set of nodes which d-separates node A from all other nodes is A's Markov blanket (MB)
- The Markov blanket MB(A) of node A in a Bayesian network is the set of nodes composed of A's parents, its children, and its children's parents

Explaining Away



- Explaining away is a common pattern of reasoning in which the confirmation of one cause of an observed or believed event reduces the need to invoke alternative causes.

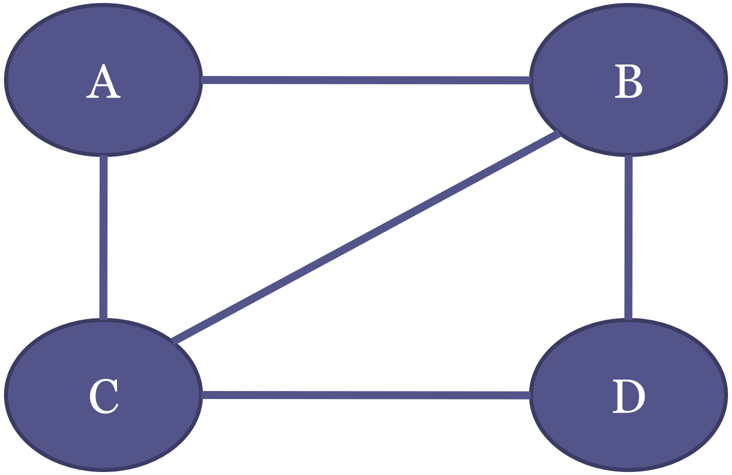
Separation for undirected graph



- Sets A and B of nodes are separated by a third set C if every path from any node in A to any node in B passes through a node in C

Complete and Clique

- A set of nodes is complete if there is a link from each node to every other node in the set;
- A clique is a maximal complete set of nodes
- Example: the following graph has cliques {A,B,C} and {B,C,D}

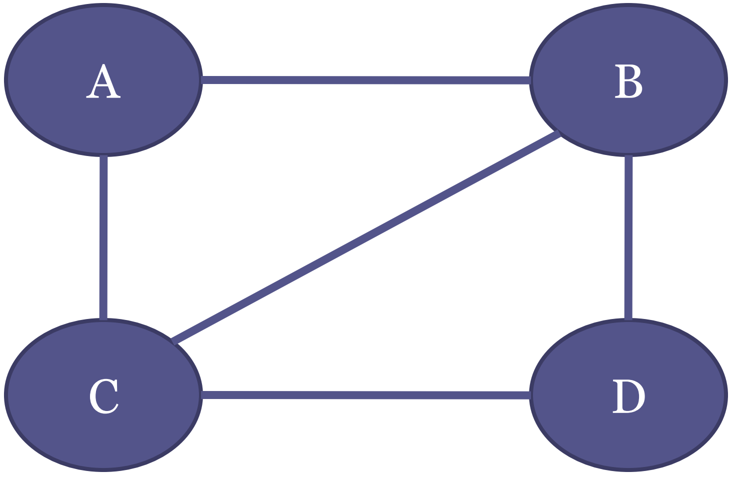


Clique potentials

- A probability distribution is said to factorize with respect to a given undirected graph if it can be expressed as the product of positive functions over the cliques of the graph
  - $P(X)=(1/Z) \prod_c \psi_c(X_c)$ where $\psi_c(X_c)$ are the clique potentials, and Z is a normalization constant

Separator

- A more general representation is the product of clique potentials divided by the separator potentials (a separator between two cliques is the set of nodes they have in common)
  - $P(X) = (\prod_C \psi_C(X_C)) / (Z \prod_S \psi_S(X_S))$
- The cliques are {A,B,C} and {B,C,D}, and the separator set is {B,C}



Undirected factorization

- A distribution which factorizes according to a particular graph is said to respect the undirected factorization property $\boldsymbol{F}$
- Theorem: for any graph and any distribution
  $\boldsymbol{F} \Rightarrow \boldsymbol{G}$
- Theorem (Hammersley-Clifford): for strictly positive distributions and arbitrary graphs $\boldsymbol{G} \Leftrightarrow \boldsymbol{F}$
- Also $\boldsymbol{G} \Leftrightarrow \boldsymbol{F}$ for any distribution if, and only if, the graph is triangulated