Probabilistic Reasoning

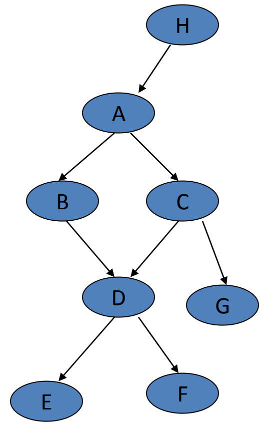
Unit # 3

d-Separation

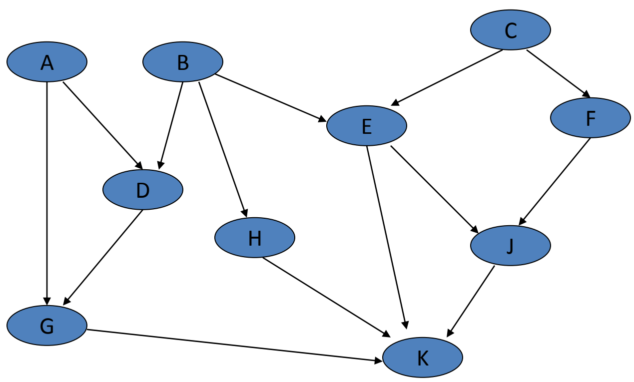
- d-separation is short for direction-dependent separation
- d-separation of vertices in a graph corresponds to conditional independence of the associated random variables
- d-separation is the mathematical basis for efficient inference algorithms in Bayesian networks
- Two nodes X and Y are d-separated by a set Z if all paths between X and Y are blocked by Z.
- When X and Y are d-separated by Z, no information can be transmitted between X and Y given Z. Hence X and Y are conditional independent given Z.

Markov Blanket

- **A node's Markov blanket consists of its parents, children, and other parents of its children (co-parents)**
  - **The Markov blanket of node B consists of all nodes whose local probability table mentions B and all nodes their local probability tables mention**
- **A node's Markov blanket d-separates it from all other nodes in the graph**
  - **A node is conditionally independent of all other nodes given its Markov blanket**

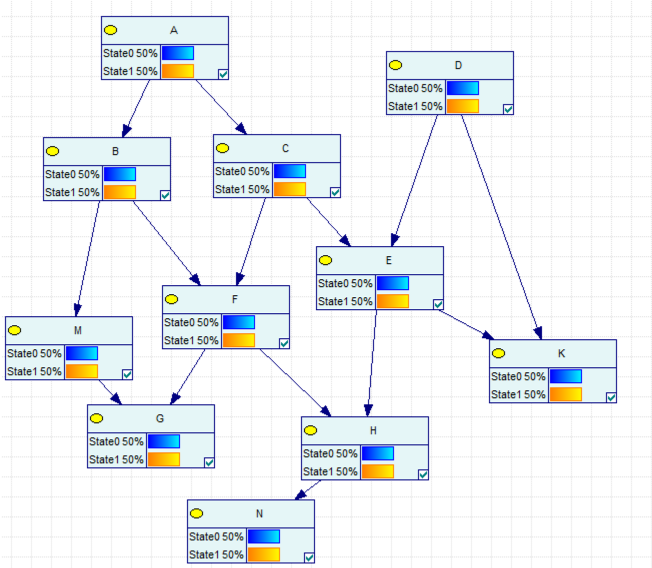


Inference Mechanism



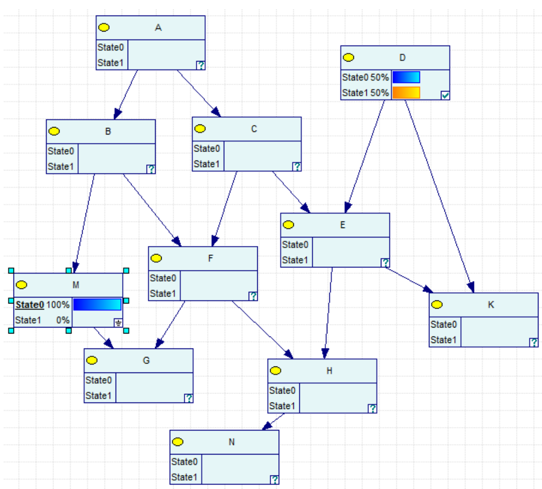
- A and K are independent?
- A and K are independent given G?
- What's the Markov blanket of H?
- A and B are independent given D?
- E and J are independent given K?
- G and B are independent given D?
- E and H are independent given B?
- E and H are independent given B and K?

Inference Mechanism Using GeNIe



Inference Mechanism Using GeNIe (Cont'd)

- When you enter evidence on a node (say E), GeNIe doesn't hide/change the probability of nodes which are independent of the node E.
- For instance, in this example, M and D are independent. Hence entering evidence on M doesn't change the probability of D.

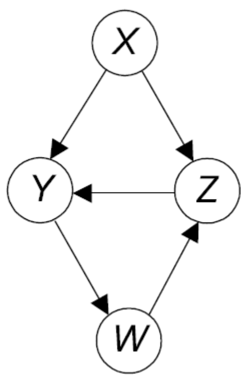


Directed Graph

- A **directed graph** is a pair (V, E), where V is a finite, nonempty set whose elements are called **nodes** (or vertices), and E is a set of ordered pairs of distinct elements of V.
- Elements of E are called **directed edges**, and if (X, Y) $\in$ E, we say there is an edge from X to Y.

Directed Graph (Cont'd)

- V = {X, Y, Z, W}
- E = {(X, Y), (X, Z), (Z, Y), (Y, W), (W, Z)}



Path and Cycle

- A **path** in a directed graph is a sequence of nodes $[X_1, X_2, ..., X_k]$ such that $(X_{i-1}, X_i) \in E$ for $2 \leq i \leq k$.
- A **cycle** in a directed graph is a path from a node to itself.

Directed Acyclic Graph

- A directed graph G is called a **directed acyclic graph (DAG)** if it contains no cycle.

More Definitions

- Given a DAG G = (V, E) and nodes X and Y in V,
  - Y is called a **parent** of X if there is an edge from Y to X,
  - Y is called a **descendent** of X and X is called an **ancestor** of Y if there is a path from X to Y, and
  - Y is called a **non-descendent** of X if Y is not a descendent of X and Y is not a parent of X.

Markov Condition and Bayesian Network

- Suppose we have a joint probability distribution P of the random variables in some set V and a DAG G = (V, E).
- We say that (G, P) satisfies the **Markov condition** if for each variable X $\in$ V, X is conditionally independent of the set of all of its non-descendents (ND) given the set of all its parents (PA).
- If (G, P) satisfies the Markov condition, we call (G, P) a **Bayesian network**.

Theorem

- (G, P) satisfies the Markov condition (and thus is a Bayesian network) if and only if P is equal to the product of its conditional distributions of all nodes given their parents in G, whenever these conditional distributions exist.

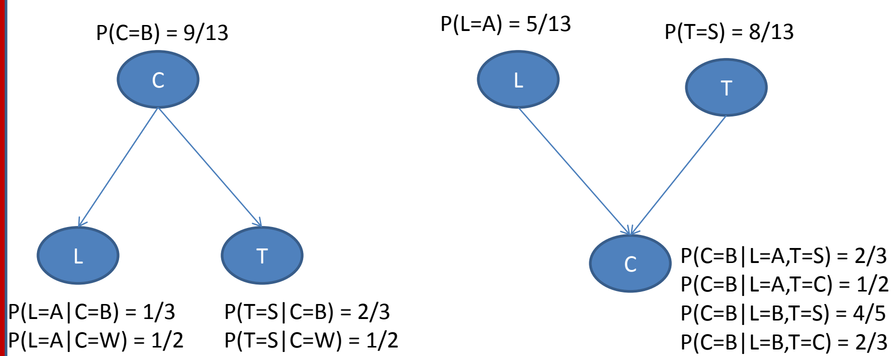
Caveat

- It is important to realize that we can't take just any DAG and expect a joint distribution to equal the product of its conditional distributions in the DAG.

Example Data

L	T	C
A	S	B
A	S	B
B	S	B
B	S	B
B	S	B
B	S	B
A	C	B
B	C	B
B	C	B
A	S	W
B	S	W
A	C	W
B	C	W

Example



- Which one of the above DAG satisfies the Markov condition?

Paradox

- Our goal is to represent a joint probability distribution succinctly using a DAG and conditional distributions for the DAG (a Bayesian network) rather than enumerating every value in the joint distribution.
- However, we don't know which DAG to use until we check whether the Markov condition is satisfied, and in general, we would need to have the joint distribution to check this.

Paradox (Cont'd)

- A common way out of this predicament is to construct a causal DAG, which is a DAG in which there is an edge from X to Y if X causes Y.
- A second way of obtaining the DAG is to learn it from data.

Causal DAGs

- A **causal graph** is a directed graph containing a set of causally related random variables V such that for every X, Y $\in$ V there is an edge from X to Y if and only if X is a direct cause of Y.
- By a **direct cause** we mean a manipulation of X results in a change in the probability distribution of Y.

Causal Markov Assumption

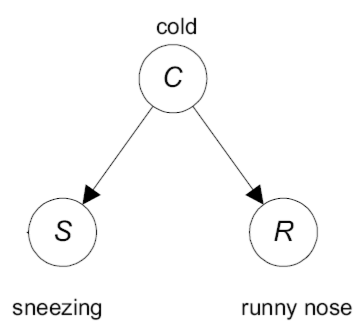
- If we assume the observed probability distribution P of a set of random variables V satisfies the Markov conditions with the causal DAG $\mathcal{G}$ containing the variables, we say we are making the **causal Markov assumption**, and we call ($\mathcal{G}$, P) a **causal network**.

Causal Markov Assumption (Cont'd)

- The causal Markov assumption is justified for a causal graph if the following conditions are satisfied:
  - There are no hidden common causes. That is, all common causes are represented in the graph.
  - There are no causal feedback loop. That is, our graph is a DAG.

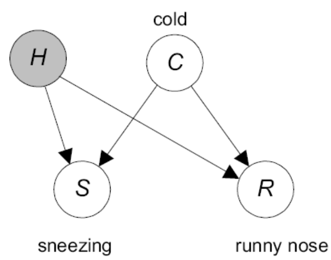
Example: Hidden Cause

- Suppose we wanted to create a causal DAG containing the variables cold (C), sneezing (S), and runny nose (R).
- Since a cold can cause both sneezing and a runny nose and neither of these conditions can cause each other, we would create this DAG.



Example: Hidden Cause (Cont'd)

- The causal Markov condition for that DAG would entail I_P (S, R|C).
- However, if there were a hidden common cause of S and R as depicted in this DAG, this conditional independency would not hold because even if the value of C were known, S would change the probability of H, which in turn would change the probability of R.
- Indeed, there is at least one other cause of sneezing and runny nose, namely hay fever.



Lesson Learned

- When making the causal Markov assumption, we must be ascertain that we have identified all common causes.

Markov Condition without Causality

- It has been argued that a causal DAG often satisfies the Markov conditions with the joint probability distribution of the random variables in the DAG.
- This does not mean that the edges in a DAG in a Bayesian network must be causal.
- A DAG can satisfy the Markov condition with the probability distribution of the variables in the DAG without the edges being causal.

Markov Condition without Causality (Cont'd)

- For instance, if we have the following causal DAG

$$A \rightarrow B \rightarrow C$$

- If we reverse the edges to obtain the DAG

$$C \rightarrow B \rightarrow A$$

- The new DAG would also satisfy the Markov condition with the probability distribution of the variables, yet the edges would not be causal.

Entailed Conditional Independences

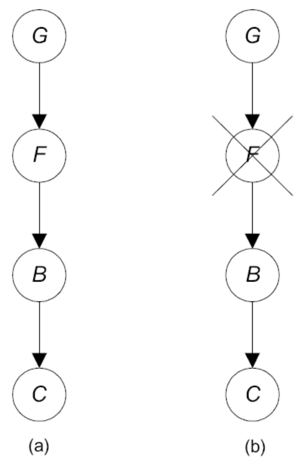
- If (G, P) satisfies the Markov conditions, are there any other conditional independencies which P must satisfy other than the one based on a node's parents?
- If such conditional independencies are present, they are called entailed conditional independencies.

Entailed Conditional Independences (Cont'd)

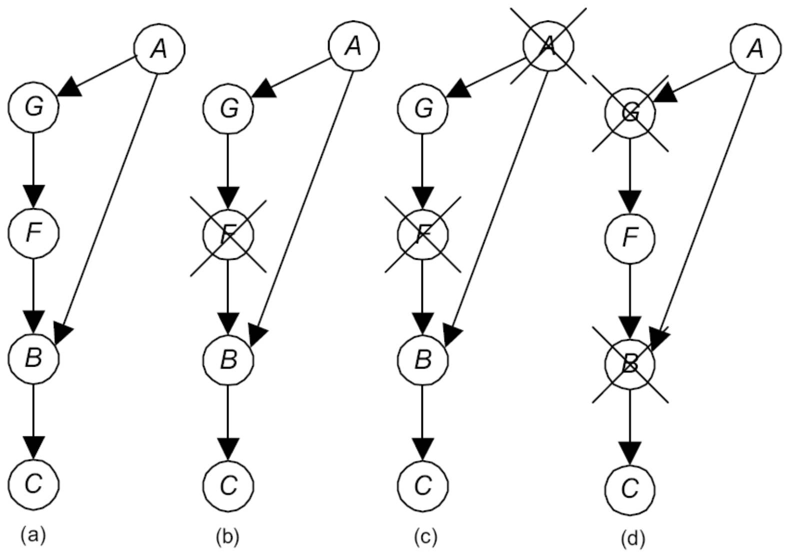
- We say a DAG entails a conditional independency if every probability distribution, which satisfies the Markov condition with the DAG, must have the conditional independency.

Example

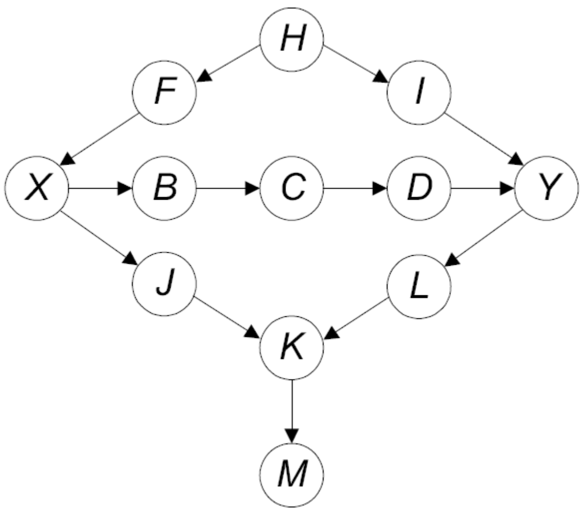
- Suppose some distribution P satisfies the Markov condition with the DAG in (a).
- We know
  - I_P (C, {F, G}|B)
  - I_P (B, G|F)
- Can we conclude?
  - I_P (C, G|F)



Example II



Example III



d-Separation

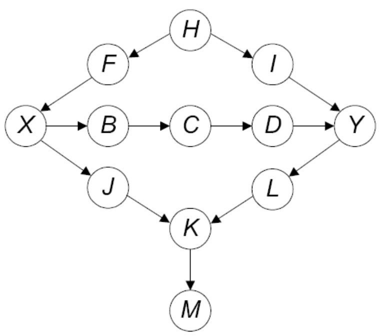
- Three types of chains
  - Causal path (A → B → C)
  - Common cause (A ← B → C)
  - Common effect (A → B ← C)
- Suppose we have a DAG G = (V, E) and a subset of nodes W ⊆ V. Then X and Y are d-separated by W if every chain between X and Y is blocked by W.

d-Separation (Theorem)

- Suppose we have a DAG G = (V, E) and three subsets of nodes X, Y, and W $\subseteq$ V. Then G entails the conditional independency I_P (X, Y|W) if and only if X and Y are d-separated by W.

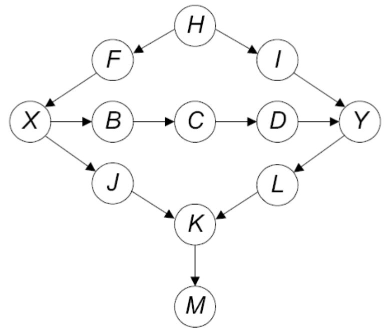
d-Separation Example

- The chain [X, B, C, D, Y] is a causal path from X to Y.
- In general, there is a dependency between X and Y on this chain, and the instantiation of any intermediate cause on the chain blocks the dependency.



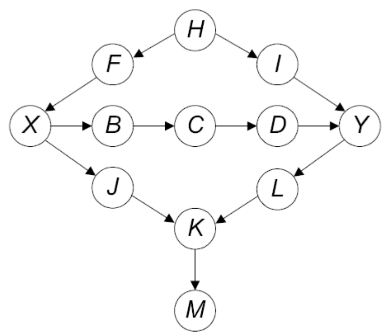
d-Separation Example (Cont’d)

- The chain [X, F, H, I, Y] is a chain in which H is a common cause of X and Y .
- In general, there is a dependency between any X and Y on this chain, and the instantiation of the common cause H or either of the intermediate causes F and I blocks the dependency.



d-Separation Example (Cont’d)

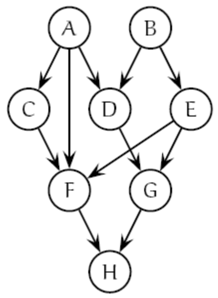
- The chain [X, J, K, L, Y] is a chain in which X and Y both cause K.
- There is no dependency between X and Y on this chain.
- However, if we instantiate K or M, in general a dependency would be created.
- We would then need to also instantiate J or L.



d-Separation Example - Conclusion

- To render X and Y conditionally independent, we need to instantiate at least one variable on all the chains that transmit a dependency between X and Y.
- So, we would need to instantiate at least one variable on the chain [X, B, C, D, Y], at least one variable on the chain [X, F, H, I, Y], and, if K or M are instantiated, at least one other variable on the chain [X, J, K, L, Y].

d-Separation – Example II



(1) C and G are d-connected

(2) C and E are d-separated

(3) C and E are d-connected given evidence on G

(4) A and G are d-separated given evidence on D and E

(5) A and G are d-connected given evidence on D

d-Separation – Example II (Cont'd)

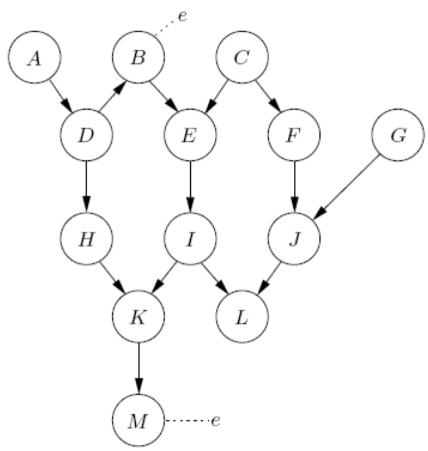
- Are C and G d-separated?
  - we find that there is a diverging connection C ← A → D allowing transmission of information from C to D via A. Second, there is a serial connection A → D → G allowing transmission of information from A to G via D. So, information can thus be transmitted from C to G via A and D, meaning that C and G are not d-separated (i.e., they are d-connected).

d-Separation – Example II (Cont'd)

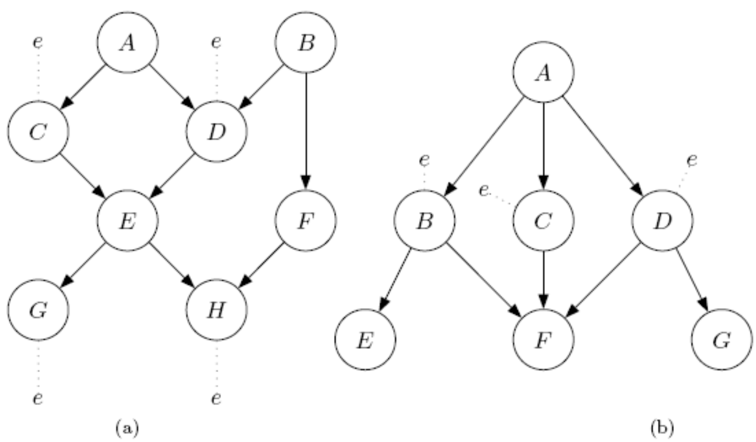
- Are C and E d-separated?
  - C and E are d-separated, since each path from C to E contains a converging connection, and since no evidence is available, each such connection will not allow transmission of information.
  - Given evidence on one or more of the variables in the set {D, F, G,H}, C and E will, however, become d-connected. For example, evidence on H will allow the converging connection D → G ← E to transmit information from D to E via G, as H is a child of G. Then information may be transmitted from C to E via the diverging connection C ← A → D and the converging connection D → G ← E.

d-Separation – Example III

- Are A and G independent?
- Are A and J independent given L?
- Are E and G independent given J?



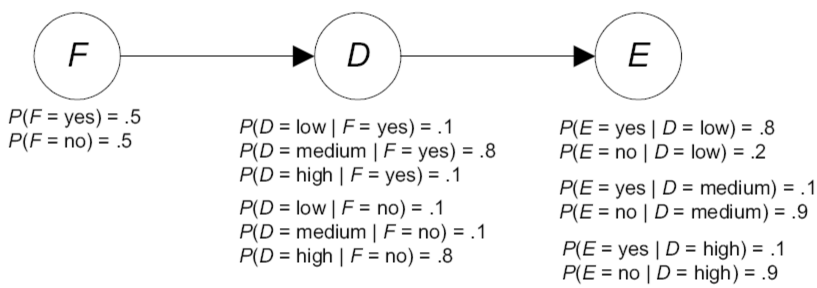
d-Separation – Example IV



Faithful and Unfaithful Probability Distribution

- A DAG entails a conditional independency if every probability distribution, which satisfies the Markov conditions with the DAG, must have the conditional independence.
- This definition DOES NOT say that if some particular probability distribution P satisfies the Markov condition with a DAG G, then P cannot have conditional independencies that G does not entail.
- It only says that P must have all the conditional independence that are entailed.

Example



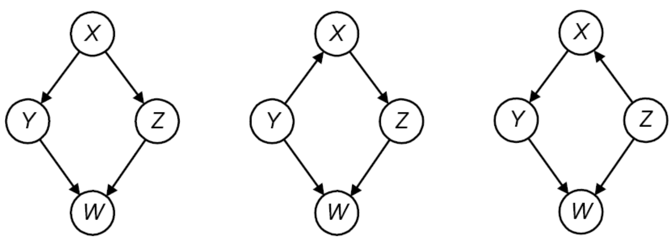
- The conditional probability distributions result in I_P (E, F) although the structure doesn't reveal this independence.

Faithfulness

- Suppose we have a joint probability distribution P of the random variables in some set V and a DAG G = (V, E). We say that (G, P) satisfies the faithfulness condition if **all and only** the conditional independencies in P are entailed by G. Furthermore, we say that P and G are faithful to each other.

Markov Equivalence

- Many DAGs are equivalent in the sense that they have the same d-separations, which means they entail the same conditional independencies.
- For example, each of the DAGs below contains the d-separations $I_G(Y, Z|X)$ and $I_G(X,W|\{Y,Z\})$ and these are the only d-separations each has.



Markov Equivalence (Cont'd)

- Let G_1 = (V, E1) and G_2 = (V, E2) be two DAGs containing the same set of variables V.
- Then G_1 and G_2 are called Markov equivalent if for every three mutually disjoint subsets A, B, C $\subseteq$ V, A and B are d-separated by C in G_1 if and only if A and B are d-separated by C in G_2. That is

$$I_{G_1}(A, B|C) \Longleftrightarrow I_{G_2}(A, B|C)$$

- **Theorem:** Two DAGs are Markov equivalent if and only if they entail the same conditional independencies

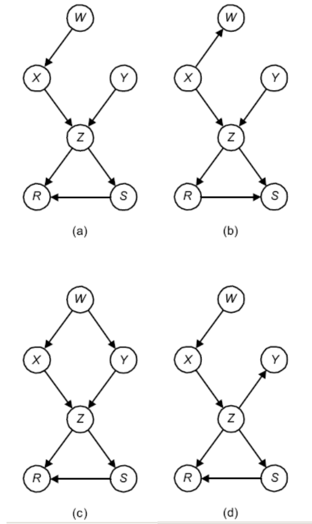
Uncouple head-to-head Meeting

- We say there is an uncoupled head-to-head meeting at X on a chain in a DAG G if there is a head-to-head meeting Y → X ← Z at X and there is no edge between Y and Z.

Theorem and Example

- Two DAGs G_1 and G_2 are Markov equivalent if and only if they have the same links (edges without regard for direction) and the same set of **uncoupled head-to-head meetings**.

- (a) and (b) are Markov equivalent.
- (c) and (d) are not – neither to (a) or (b) nor to each other.



Markov Equivalence (Cont'd)

- The previous Theorem easily enables us to develop an algorithm for determining whether two DAGs are Markov equivalent.
- The algorithm would simply check whether the two DAGs have the same links and the same set of uncoupled head-to-head meetings.
- The theorem also gives us a simple way to represent a Markov equivalence class with a single graph.
- We can represent a Markov equivalence class with a graph that has the same links and the same uncoupled head-to-head meeting as the DAGs in the class. Any assignment of directions to the undirected edges in this graph that does not create a new uncoupled head-to-head meeting or a directed cycle yields a member of the equivalence class.