Separation in Graphs

Simple Graph

A simple graph (or just: graph) is a tuple $G = (V, E)$ where

$$V = \{A_1, \ldots, A_n\}$$

represents a finite set of vertices (or nodes) and

$$E \subseteq (V \times V) \setminus \{(A, A) \mid A \in V\}$$

denotes the set of edges.

It is called simple since there are no self-loops and no multiple edges.

Edge Types

Let **G** = (V, E) be a graph. An edge $e = (A, B)$ is called

directed if $(A, B) \in E \Rightarrow (B, A) \notin E$
Notation: $A \rightarrow B$

undirected if $(A, B) \in E \Rightarrow (B, A) \in E$
Notation: $A - B$ or $B - A$

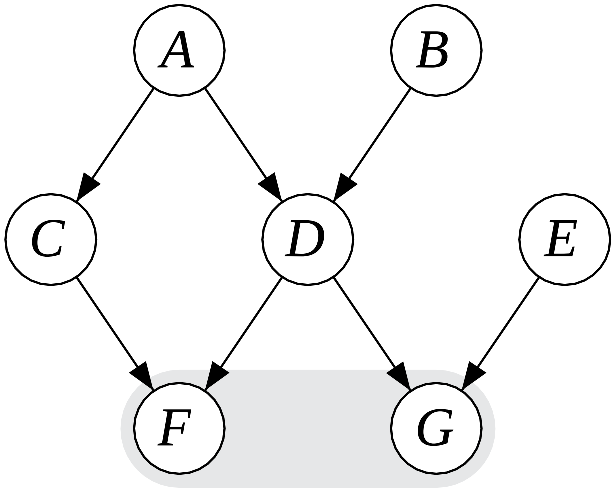
(Un)directed Graph

A graph with only (un)directed edges is called an (un)directed graph.

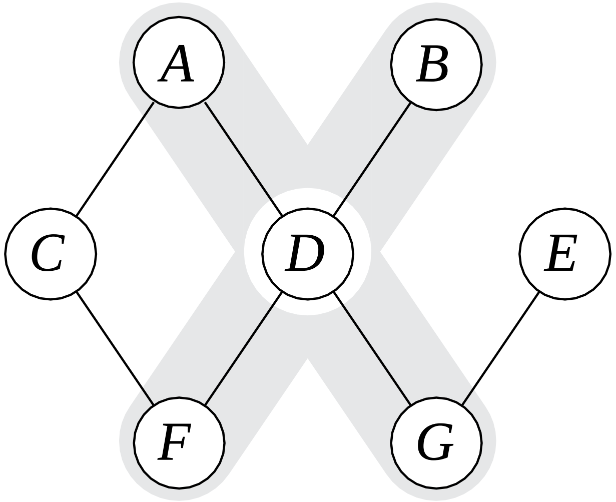
Adjacency Set

Let **G** = (V, E) be a graph. The set of nodes that is accessible via a given node $A \in V$ is called the **adjacency** set of A:

$$\text{adj}(A) = \{B \in V \mid (A, B) \in E\}$$



adj(D)



Paths

Let **G** = (V, E) be a graph. A series ρ of r pairwise different nodes

$$\rho = (A_{i_1}, \ldots, A_{i_r})$$

is called a **path** from A_i to A_j if

$$A_{i_1} = A_i, \quad A_{i_r} = A_j$$

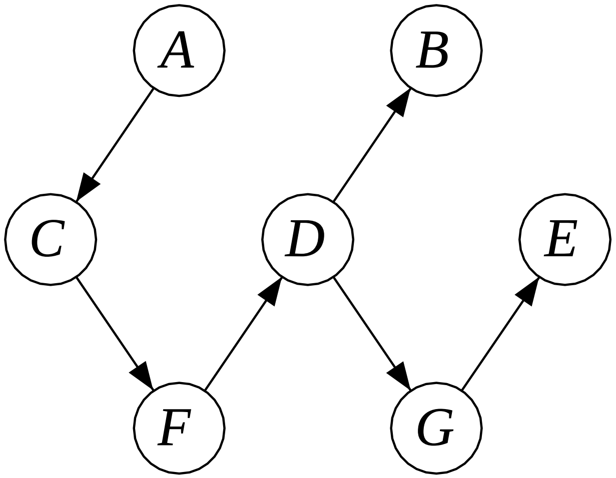
$$A_{i_{k+1}} \in \text{adj}(A_{i_k}), \quad 1 \leq k < r$$

A path with only undirected edges is called an **undirected path**

$$\rho = A_{i_1} - \cdots - A_{i_r}$$

whereas a path with only directed edges is referred to as a **directed path**

$$\rho = A_{i_1} \to \cdots \to A_{i_r}$$



If there is a directed path ρ from node A to node B in a directed graph **G** we write

$$A \overset{\rho}{\underset{\mathcal{G}}{\rightsquigarrow}} B$$

If the path ρ is undirected we denote this with

$$A \overset{\rho}{\underset{\mathcal{G}}{\leftrightsquigarrow}} B$$

Loop

Let **G** $= (V, E)$ be an undirected graph. A path

$$\rho = X_1 - \cdots - X_k$$

with $(X_k - X_1) \in E$ is called a loop.

Cycle

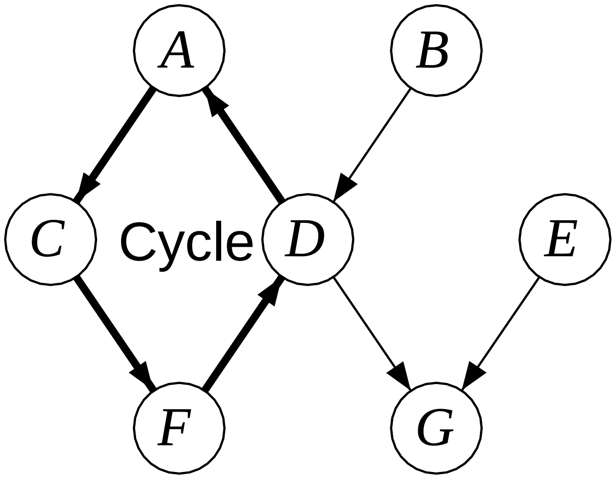
Let **G** $= (V, E)$ be a directed graph. A path

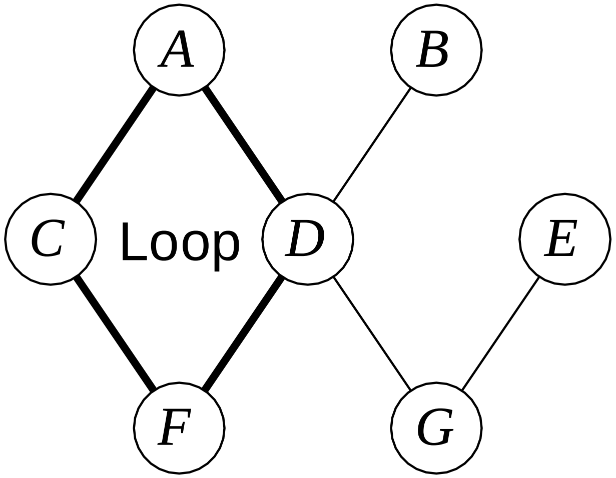
$$\rho = X_1 \rightarrow \cdots \rightarrow X_k$$

with $(X_k \rightarrow X_1) \in E$ is called a cycle.

Directed Acyclic Graph (DAG)

A directed graph **G** $= (V, E)$ is called **acyclic** if for every path $X_1 \rightarrow \cdots \rightarrow X_k$ in **G** the condition $(X_k \rightarrow X_1) \notin E$ is satisfied, i. e. it contains no cycle.





Let **G** $= (V, E)$ be a directed graph. For every node $A \in V$ we define the following sets:

Parents:

$$\text{parents}_{\mathbf{G}}(A) = \{B \in V \mid B \rightarrow A \in E\}$$

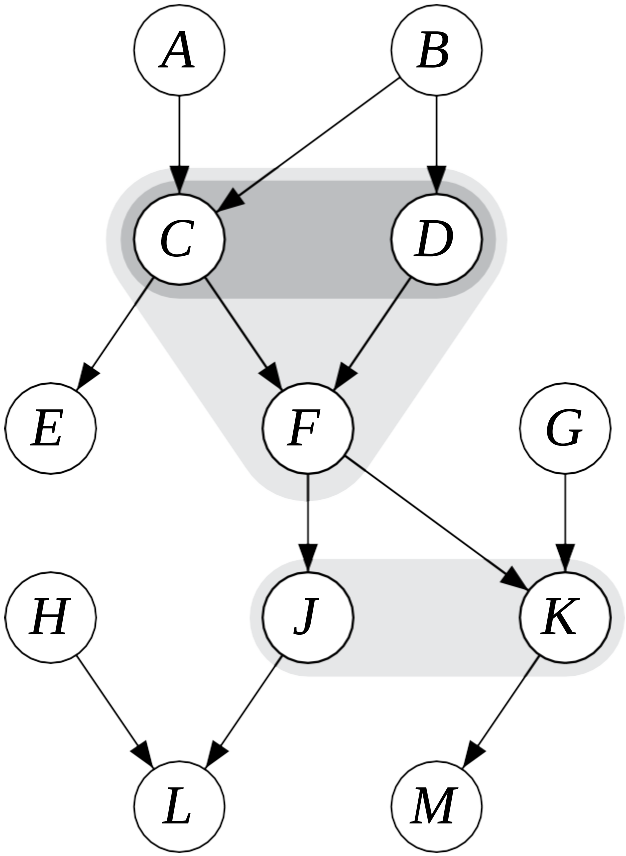
Children:

$$\text{children}_{\mathbf{G}}(A) = \{B \in V \mid A \rightarrow B \in E\}$$

Family:

$$\text{family}_{\mathbf{G}}(A) = \{A\} \cup \text{parents}_{\mathbf{G}}(A)$$

If the respective graph is clear from the context, the index **G** is omitted.



$$\text{parents}(F) = \{C, D\}$$
$$\text{children}(F) = \{J, K\}$$
$$\text{family}(F) = \{C, D, F\}$$

Let **G** = (V, E) be a DAG. For every node $A \in V$ we define the following sets:

Ancestors:

$$\text{ancs}_G(A) = \{B \in V \mid \exists \rho : B \overset{\rho}{\underset{\mathcal{G}}{\rightsquigarrow}} A\}$$

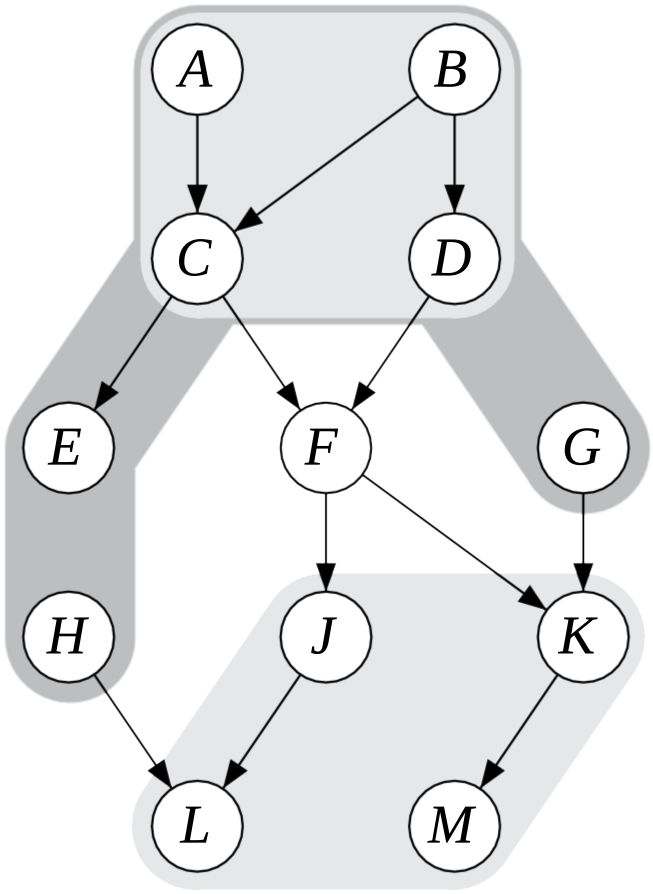
Descendants:

$$\text{descs}_G(A) = \{B \in V \mid \exists \rho : A \overset{\rho}{\underset{\mathcal{G}}{\rightsquigarrow}} B\}$$

Non-Descendants:

$$\text{non-descs}_G(A) = V \setminus \{A\} \setminus \text{descs}_{\mathbf{G}}(A)$$

If the respective graph is clear from the context, the index **G** is omitted.



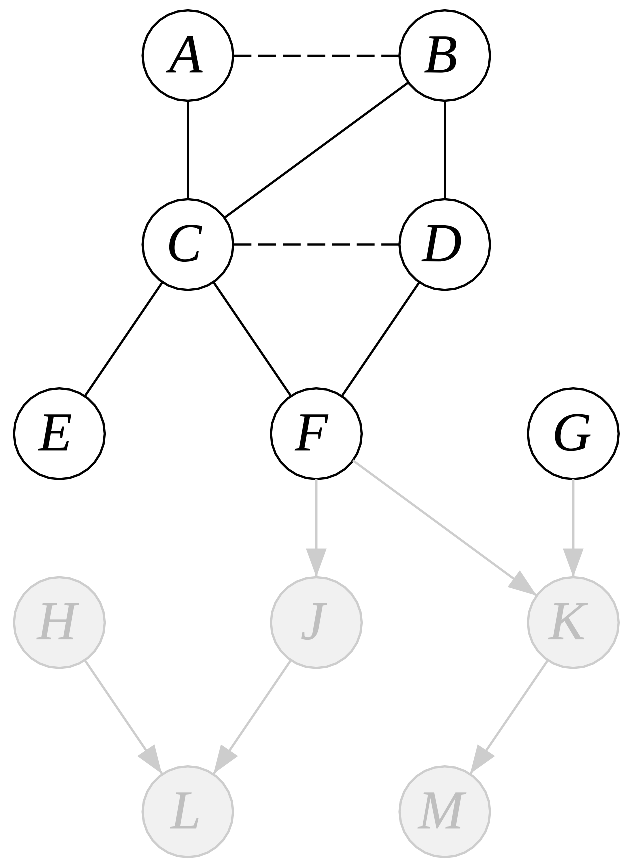
$$\text{ancs}(F) = \{A, B, C, D\}$$
$$\text{descs}(F) = \{J, K, L, M\}$$
$$\text{non-descs}(F) = \{A, B, C, D, E, G, H\}$$

Let G= (V, E) be a DAG.

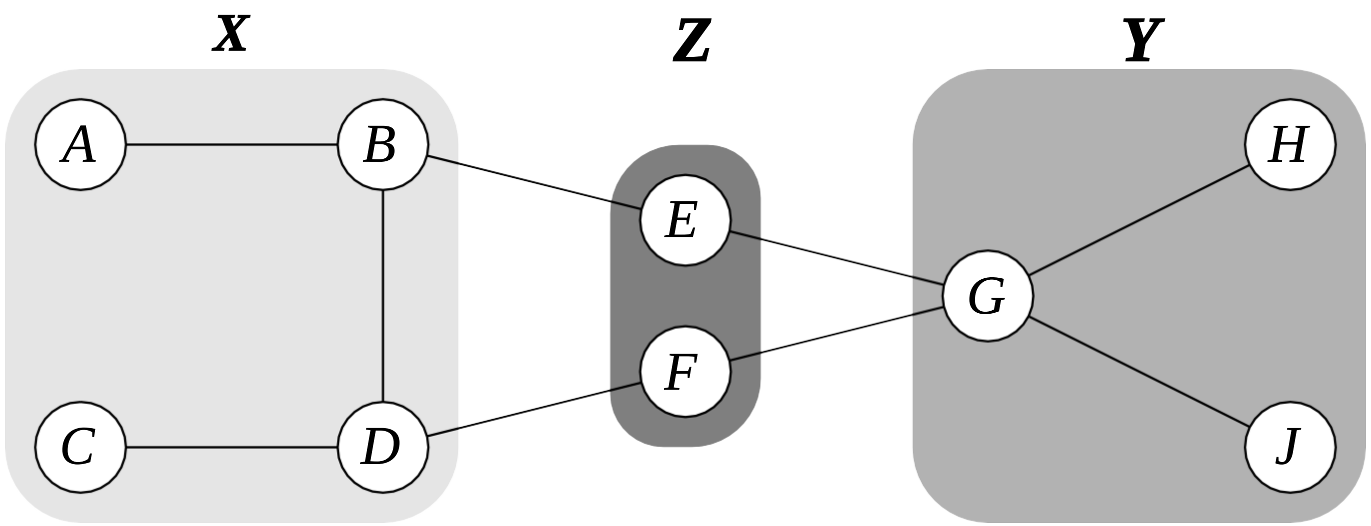
The **Minimal Ancestral Subgraph** of **G** given a set $M \subseteq V$ of nodes is the smallest subgraph that contains M and all ancestors of all nodes in M.

The **Moral Graph** of **G** is the undirected graph that is obtained by

1. connecting nodes that share a common child with an arbitrarily directed edge and,
2. converting all directed edges into undirected ones by dropping the arrow heads.



Moral graph of ancestral graph induced by the set $\{E, F, G\}$.



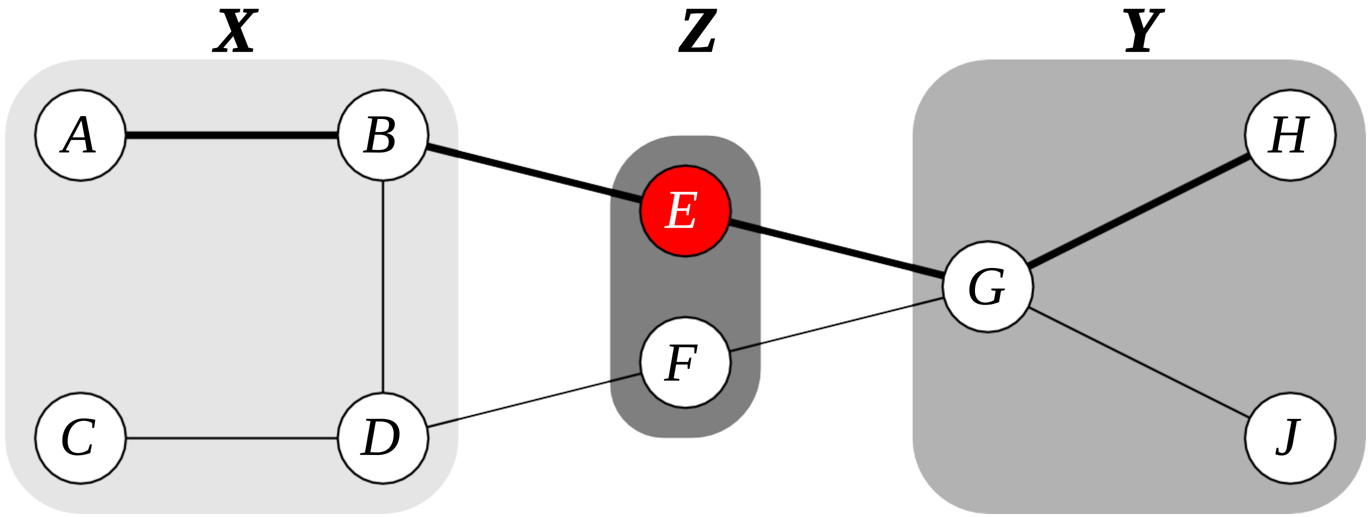
Let **G** = (V, E) be an undirected graph and $X, Y, Z \subseteq V$ three disjoint subsets of nodes. We agree on the following separation criteria:

1. Z u-separates X from Y — written as

$$X \perp\!\!\!\perp_{G} Y \mid Z,$$

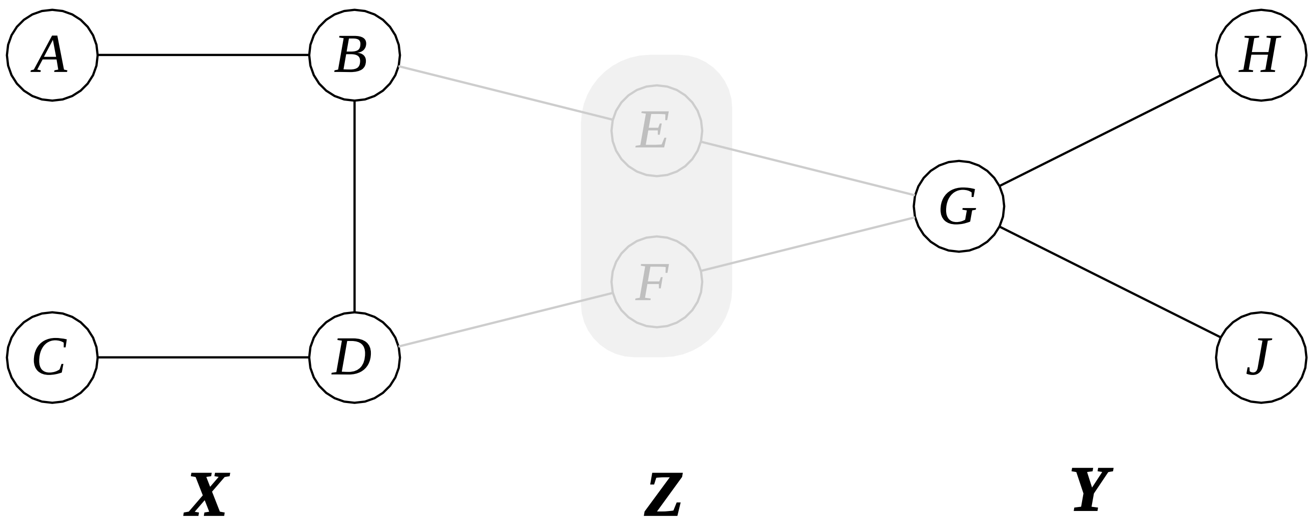
   if every possible path from a node in X to a node in Y is blocked.

2. A path is blocked if it contains one (or more) **blocking nodes**.
3. A node is a blocking node if it lies in Z.



e.g. path $A - B - E - G - H$ is blocked by $E \in Z$. It can be easily verified, that every path from X to Y is blocked by Z. Hence we have:

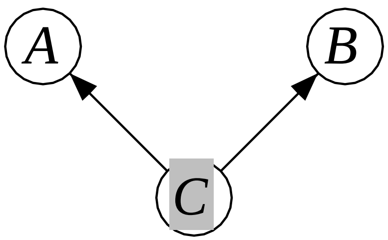
$$\{A, B, C, D\} \perp\!\!\!\perp \{G, H, J\} \mid \{E, F\}$$

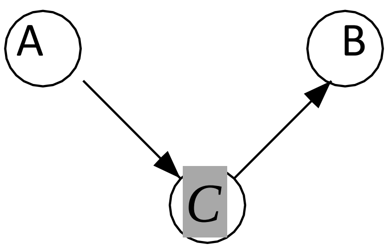


Another way to check for u-separation: Remove the nodes in Z from the graph (and all the edges adjacent to these nodes). X and Y are u-separated by Z if the remaining graph is disconnected with X and Y in two separate subgraphs.

Motivation: Separation in Directed Graphs

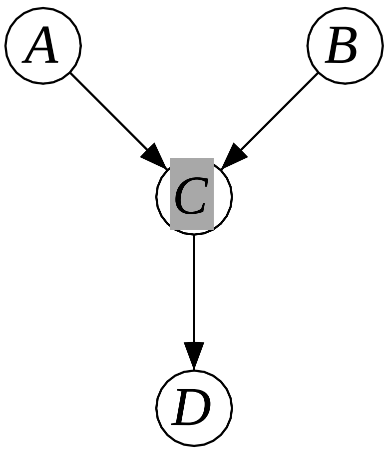
Idea: Separation in Directed Graphs should fit to the concept of Conditional independence





If C is instantiated with c
then A and B are conditional independent,
i.e. P(A,BIC=c)=P(AIC=c)P(BIC=c)

C separates A and B
C is a blocking node of the path A-B-C
(walking against the direction of the arrows is allowed)



A quality of ingredients
B cook's skill
C meal quality
D restaurant success

If C is instantiated with c
then A and B are conditional dependent

If D is instantiated with d
Then A and B are conditional dependent

C is no separator of A and B,
C is no blocking node

D is no separator of A and B,
C is no blocking node

Separation criterion for directed graphs.

We use the same principles as for u-separation. Two modifications are necessary: Directed paths may lead also in reverse to the arrows.

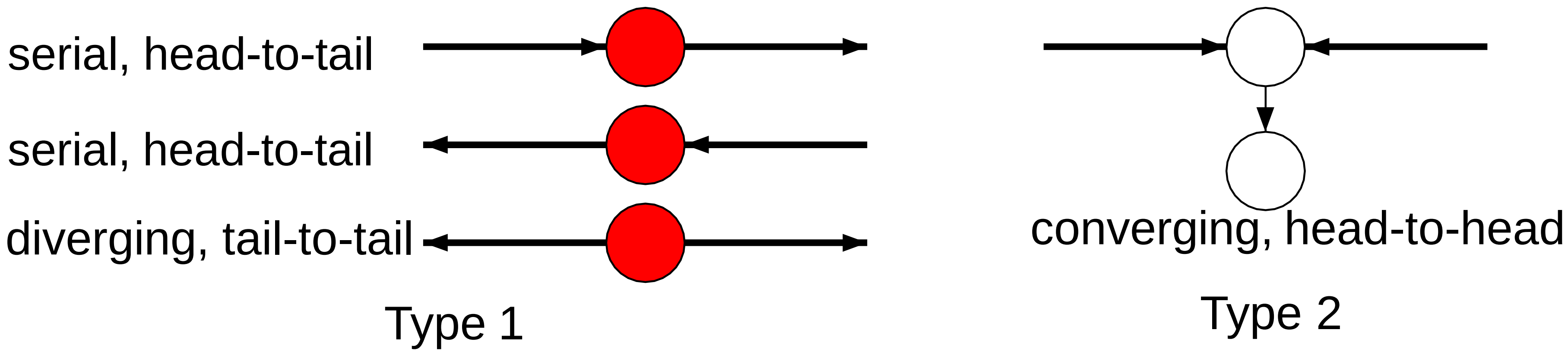
The blocking node condition is more sophisticated.

Blocking Node (in a directed path)

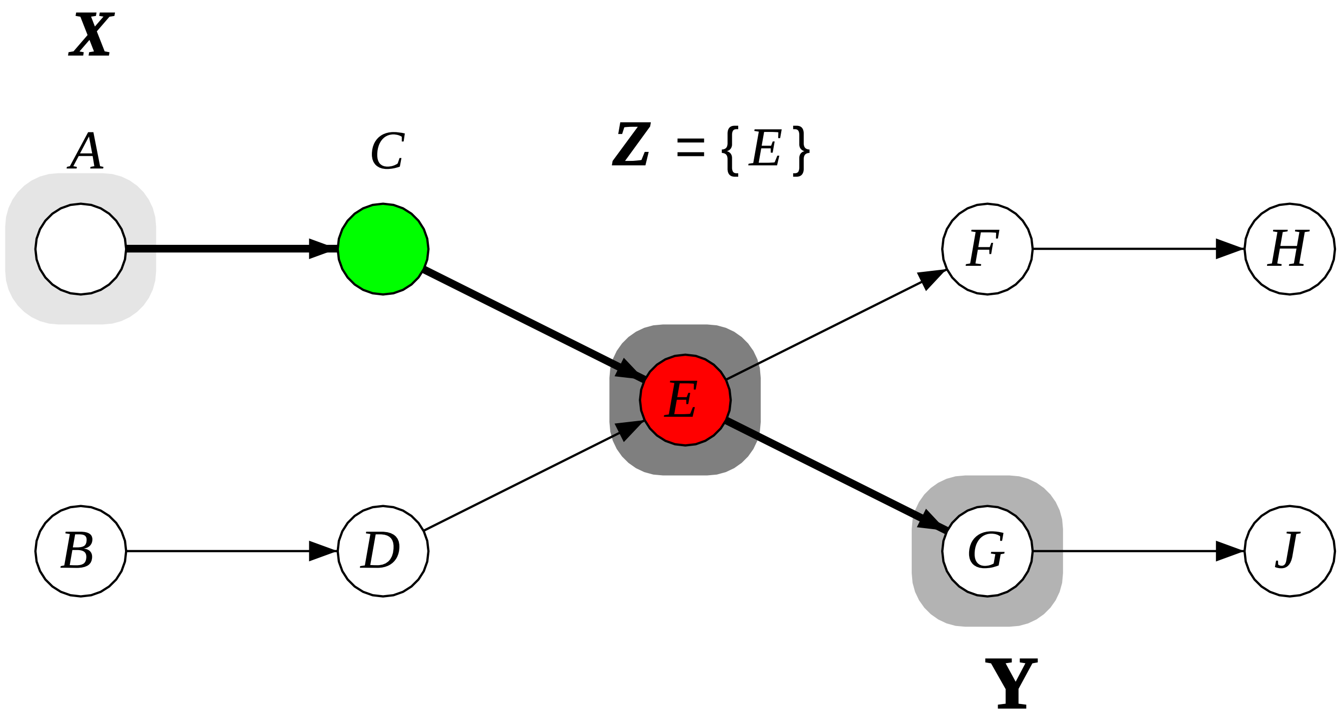
A node A is blocking if its edge directions **along the path**

are of type 1 and $A \in \boldsymbol{Z}$, or

are of type 2 and neither A nor one of its descendants is in $\boldsymbol{Z}$.



Examples for d-Separation $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$



Checking path $A \rightarrow C \rightarrow E \rightarrow G$

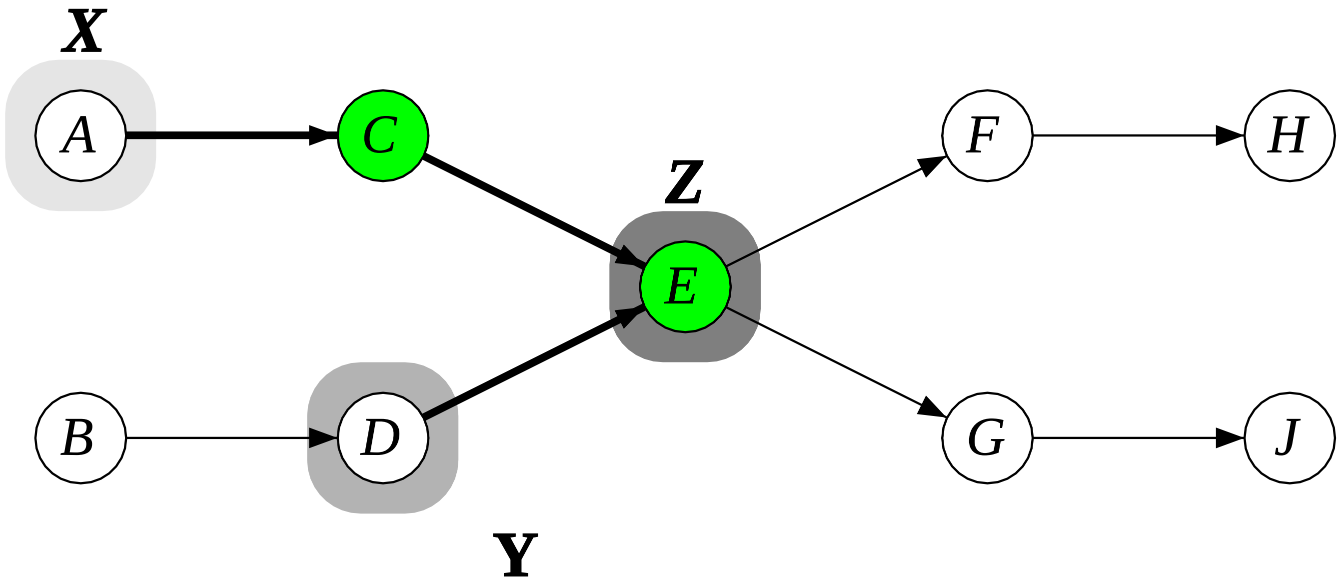
C is **serial** and not in Z: non-blocking

E is also **serial** but in Z: **blocking**

Path is blocked, no other paths between A and G are available

$$\Rightarrow \quad A \perp\!\!\!\perp G \mid E$$

Examples for d-Separation $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$



Checking path $A \rightarrow C \rightarrow E \leftarrow D$:

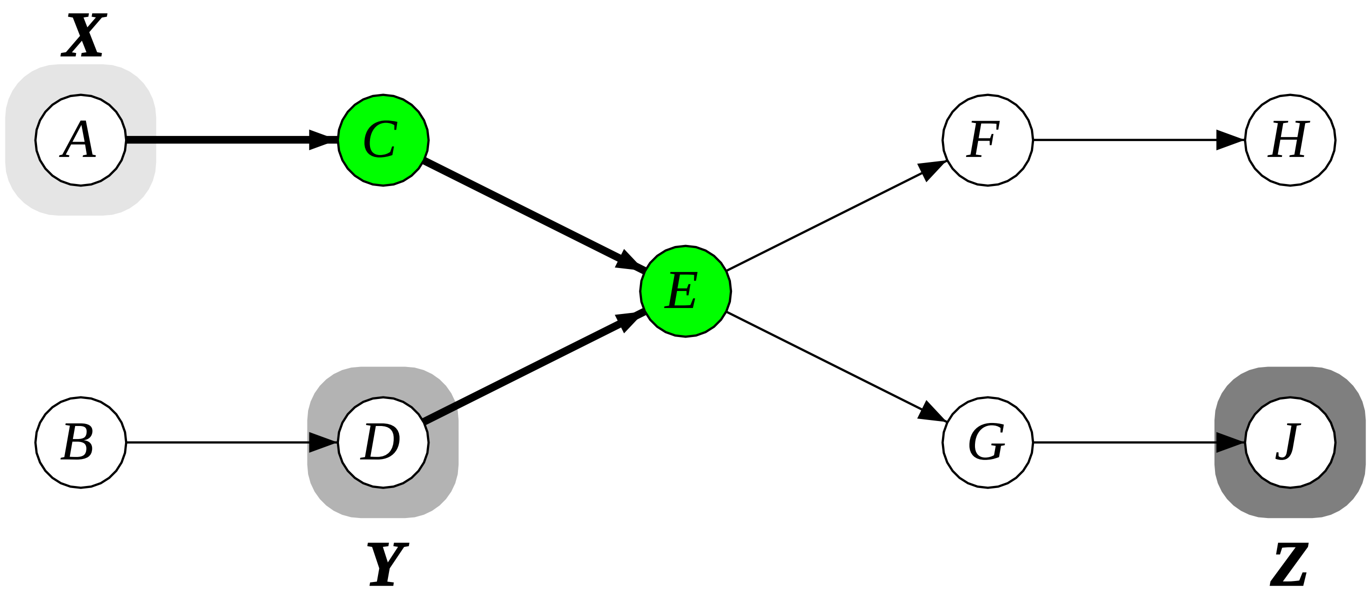
C is serial and not in Z: non-blocking

E is $\mathrm{converging}$ and in Z: non-blocking

$\Rightarrow$ Path is not blocked

$$A \not\perp\!\!\!\perp D \mid E$$

Examples for d-Separation $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$



Checking path $A \rightarrow C \rightarrow E \leftarrow D$:

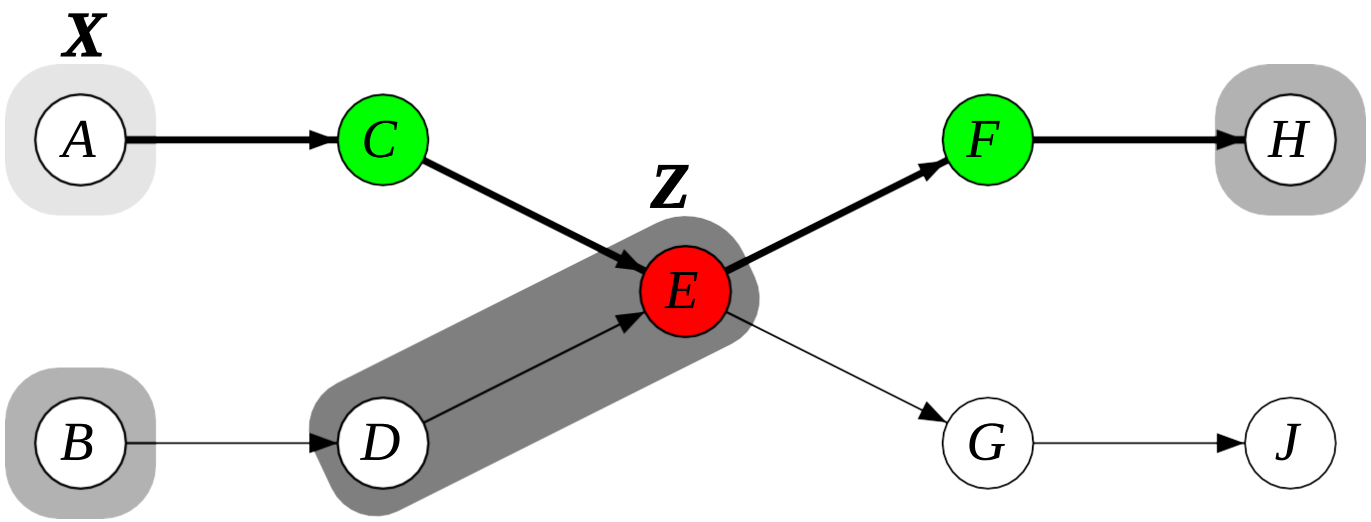
C is **serial** and not in Z: non-blocking

E is **converging** and not in Z but one of its descendants (J) is in Z: non-blocking

$\Rightarrow$ Path is not blocked

$$A \not\perp\!\!\!\perp D \mid J$$

Examples for d-Separation $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$



$\boldsymbol{Y} = \{B, H\}$

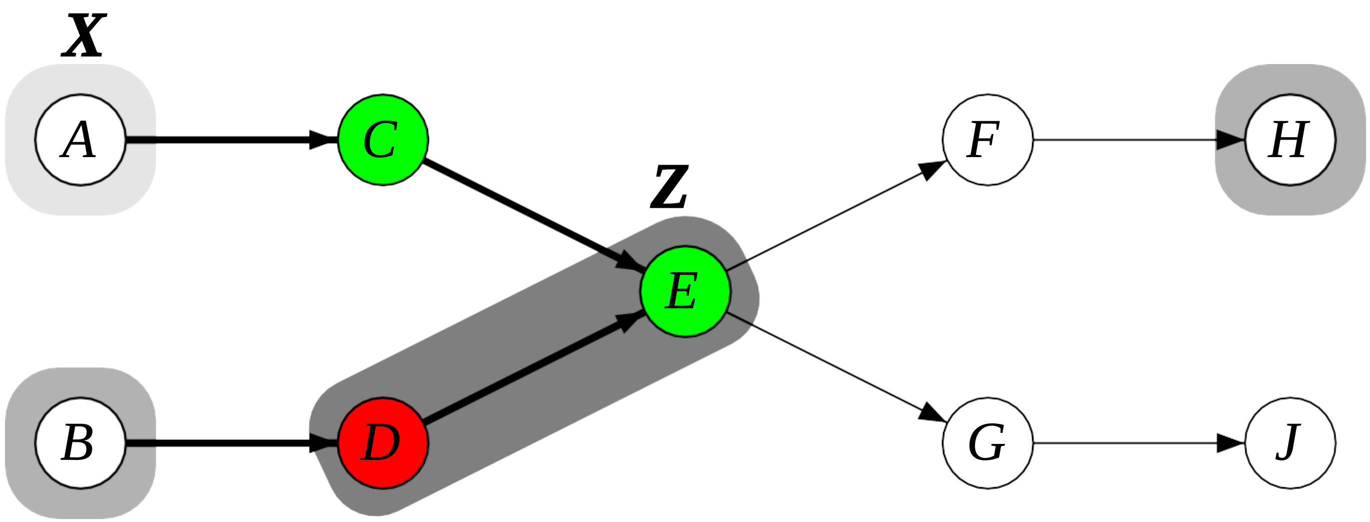
Checking path $A \rightarrow C \rightarrow E \rightarrow F \rightarrow H$:

C is **serial** and not in Z: non-blocking

E is **serial** and in Z: **blocking**

F is serial and not in Z: non-blocking

$\Rightarrow$ Path is blocked



$\boldsymbol{Y} = \{B, H\}$

Checking path $A \to C \to E \leftarrow D \to B$:

C is **serial** and not in Z: non-blocking

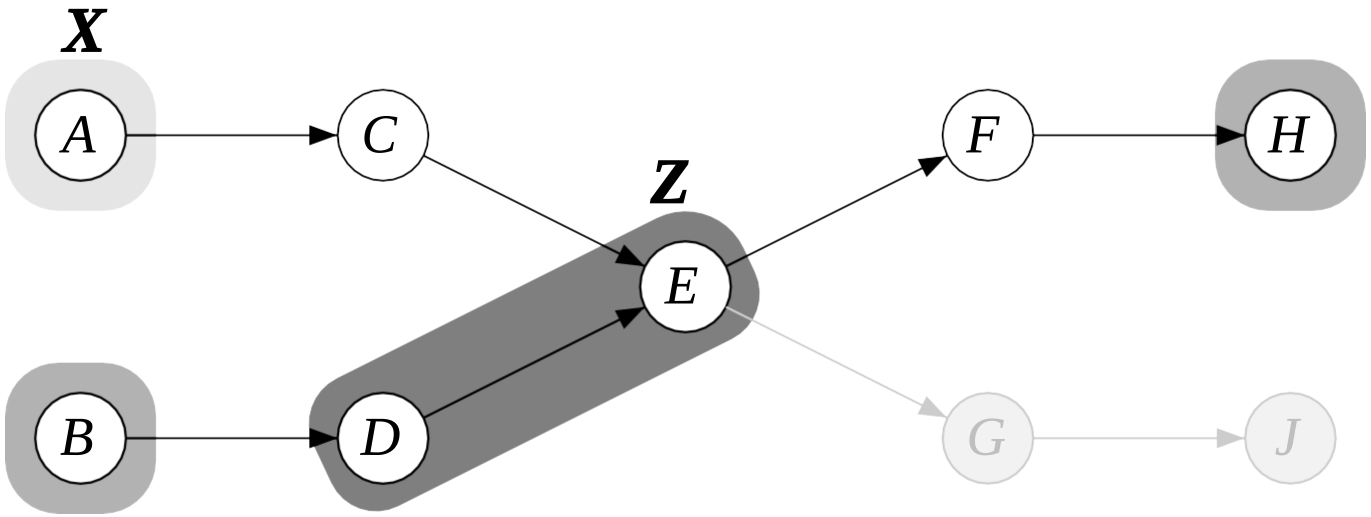
E is **converging** and in Z: non-blocking

D is **serial** and in Z: **blocking**

⇒ Path is blocked

$$A \perp\!\!\!\perp B, H \mid D, E$$

d-Separation: Alternative Way for Checking

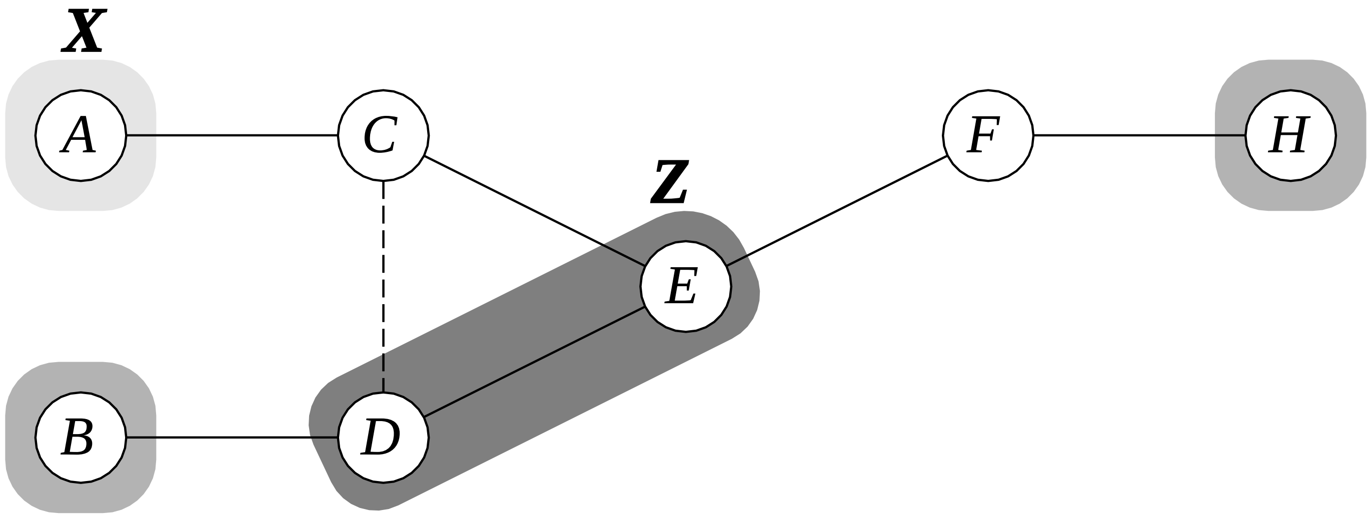


$\boldsymbol{Y} = \{B, H\}$

Steps

- Create the minimal ancestral subgraph induced by $X \cup Y \cup Z$

d-Separation: Alternative Way for Checking

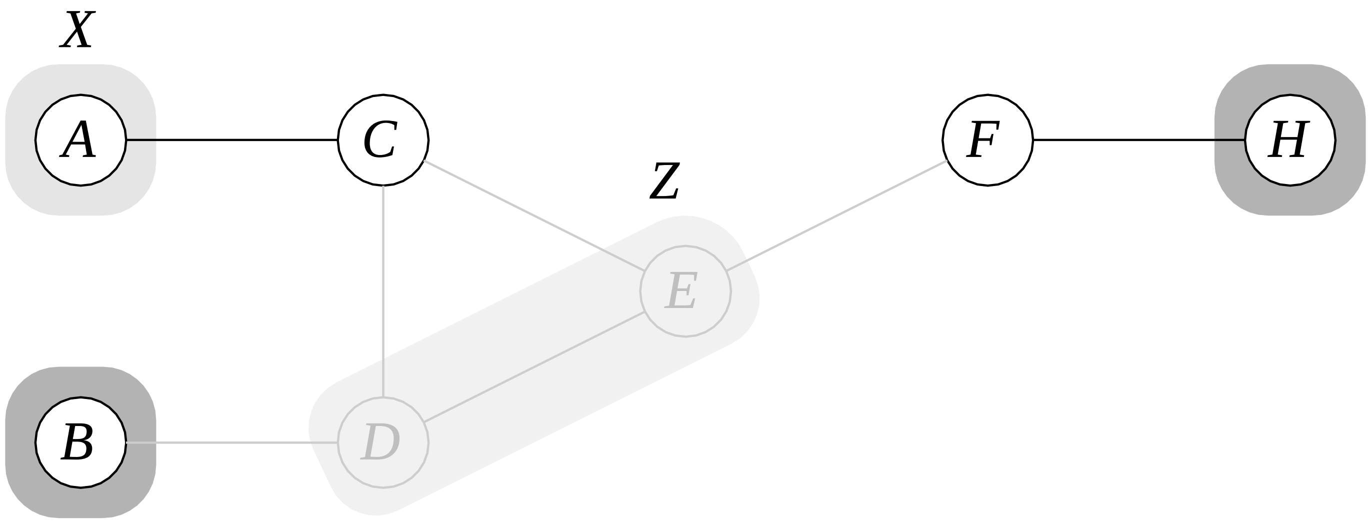


$\boldsymbol{Y} = \{B, H\}$

Steps

- Create the minimal ancestral subgraph induced by $X \cup Y \cup Z$
- Moralize that subgraph

d-Separation: Alternative Way for Checking



$Y = \{B, H\}$

Steps:

- Create the minimal ancestral subgraph induced by $X \cup Y \cup Z$
- Moralize that subgraph
- Check for u-Separation in that undirected graph

$$A \perp\!\!\!\perp H, B \mid D, E$$

Summary: d-Separation

Let $G = (V, E)$ a DAG and $X, Y, Z \in V$ three nodes.

a) A set $S \subseteq V \setminus \{X, Y\}$ *d-separates* X and Y, if S blocks all paths between X and Y. A path may also route in opposite edge direction.

b) A path π is d-separated by S if at least one pair of consecutive edges along π is blocked. There are the following blocking conditions:

1. $X \leftarrow Y \rightarrow Z$ tail-to-tail
2. $X \leftarrow Y \leftarrow Z$, $X \rightarrow Y \rightarrow Z$ head-to-tail
3. $X \rightarrow Y \leftarrow Z$ head-to-head

c) Two edges that meet tail-to-tail or head-to-tail in node Y are blocked if $Y \in S$.

d) Two edges meeting head-to-head in Y are blocked if neither Y nor its successors are in S.

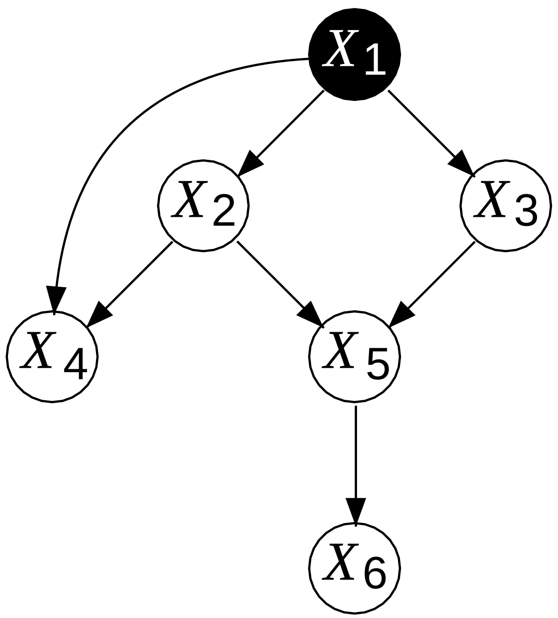
Theorem

If $S \subseteq V \backslash \{X, Y\}$ d-separates X and Y in a Bayesian network (V, E, P), then X and Y are conditionally independent given S:

$$P(X, Y \mid S) = P(X \mid S) \cdot P(Y \mid S)$$

Example

Paths: $\pi_1 = (X_2 - X_1 - X_3)$, $\pi_2 = (X_2 - X_5 - X_3)$
$\pi_3 = (X_2 - X_4 - X_1 - X_3)$, $S = \{X_1\}$



π_1 $X_2 \leftarrow X_1 \rightarrow X_3$ tail-to-tail
$X_1 \in S \Rightarrow \pi_1$ is blocked by S

π_2 $X_2 \rightarrow X_5 \leftarrow X_3$ head-to-head
$X_5, X_6 \notin S \Rightarrow \pi_2$ is blocked by S

π_3 $X_4 \leftarrow X_1 \rightarrow X_3$ tail-to-tail $X_2 \rightarrow X_4 \leftarrow X_1$ head-to-head both connections are blocked $\Rightarrow \pi_3$ is blocked

X_2 and X_3 are d-separated via $\{X_1\}$.

X_2 and X_3 are therefore conditionally independent given X_1

Algebraic structure of CI statements

Conditional independence statements can be characterised qualitatively,
e.g. without specifying the numerical values of probabilities.

Let $(\Omega, \mathcal{E}, P)$ be a probability space and W, X, Y, Z disjoint subsets of variables. If X and Y are conditionally independent given Z we write:

$$X \perp\!\!\!\perp_P Y \mid Z$$

Often, the following (equivalent) notation is used:

$$I_P(X \mid Z \mid Y) \quad \text{or} \quad I_P(X, Y \mid Z)$$

If the underlying space is known the index P is omitted.

(Semi-)Graphoid Axioms

Definition: Let V be a set of (mathematical) objects and $(\cdot \perp\!\!\!\perp \cdot \mid \cdot)$ a three-place relation of subsets of V. Furthermore, let W, X, Y, and Z be four disjoint subsets of V. The four statements

symmetry: $(X \perp\!\!\!\perp Y \mid Z) \Rightarrow (Y \perp\!\!\!\perp X \mid Z)$

decomposition: $(W \cup X \perp\!\!\!\perp Y \mid Z) \Rightarrow (W \perp\!\!\!\perp Y \mid Z) \wedge (X \perp\!\!\!\perp Y \mid Z)$

weak union: $(W \cup X \perp\!\!\!\perp Y \mid Z) \Rightarrow (X \perp\!\!\!\perp Y \mid Z \cup W)$

contraction: $(X \perp\!\!\!\perp Y \mid Z \cup W) \wedge (W \perp\!\!\!\perp Y \mid Z) \Rightarrow (W \cup X \perp\!\!\!\perp Y \mid Z)$

are called the **semi-graphoid axioms**. A three-place relation $(\cdot \perp\!\!\!\perp \cdot \mid \cdot)$ that satisfies the semi-graphoid axioms for all W, X, Y, and Z is called a **semi-graphoid**.

Note: The probability calculus satisfies the four **semi-graphoid** axioms, but not the additional fifth intersection axiom of a **graphoid**.

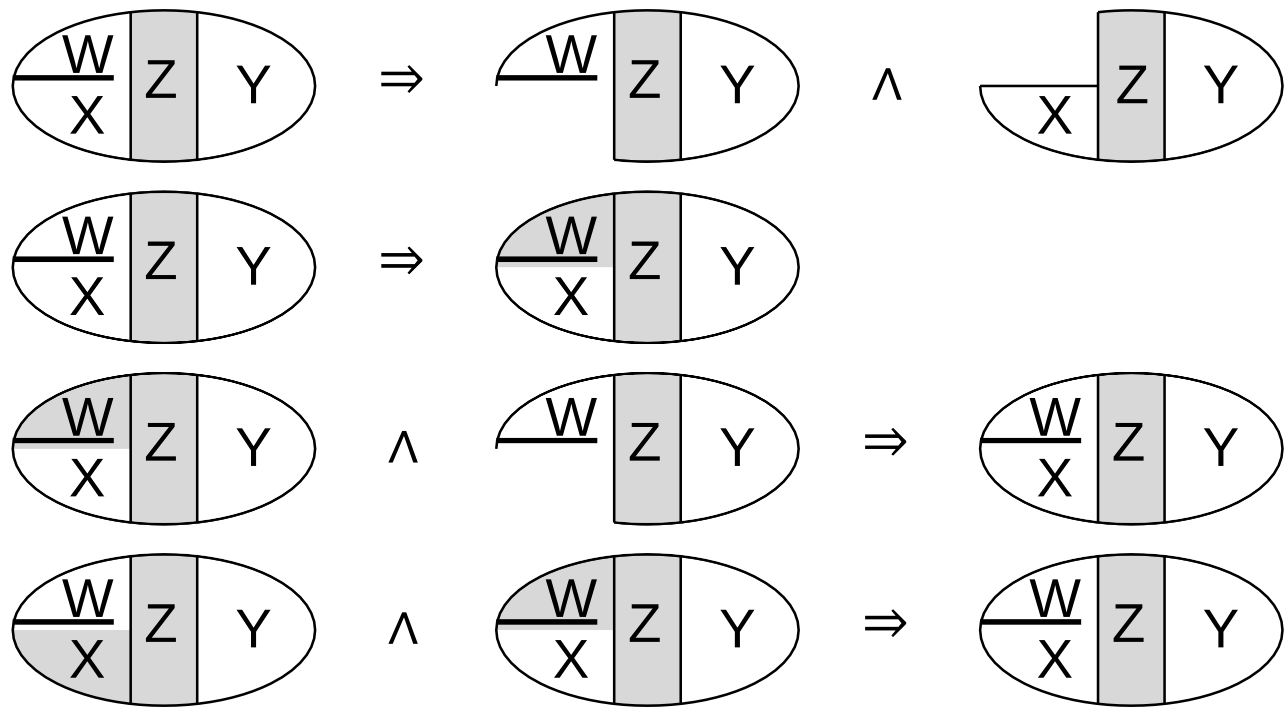
Illustration of the (Semi-)Graphoid Axioms

decomposition

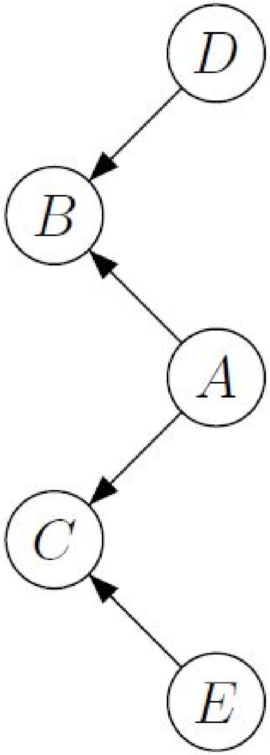
weak union

contraction

intersection



Example



$$
\begin{aligned}
& D \perp\!\!\!\perp A, C \mid \emptyset \quad \wedge \quad B \perp\!\!\!\perp C \mid A, D \\
\overset{\text{w. union}}{\Longrightarrow} \quad & D \perp\!\!\!\perp C \mid A \quad \wedge \quad B \perp\!\!\!\perp C \mid A, D \\
\overset{\text{symm.}}{\Longleftrightarrow} \quad & C \perp\!\!\!\perp D \mid A \quad \wedge \quad C \perp\!\!\!\perp B \mid A, D \\
\overset{\text{contr.}}{\Longrightarrow} \quad & C \perp\!\!\!\perp B, D \mid A \\
\overset{\text{decomp.}}{\Longrightarrow} \quad & C \perp\!\!\!\perp B \mid A \\
\overset{\text{symm.}}{\Longleftrightarrow} \quad & B \perp\!\!\!\perp C \mid A
\end{aligned}
$$

Independence Maps

Let $(\cdot \perp\!\!\!\perp_\delta \cdot \mid \cdot)$ be a three-place relation representing the set of **conditional independence statements** that hold in a given distribution δ over U.

A graph G=(U,E) over random variables U is called an **independence map (I-map)** for the joint probability space δ , if for all disjoint subsets X,Y,Z of U the property

$$\langle X \mid Z \mid Y \rangle_G \Rightarrow X \perp\!\!\!\perp_\delta Y \mid Z$$

holds.

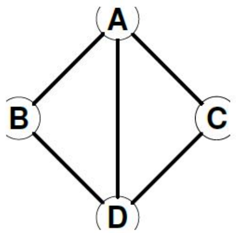
An I-map G for δ captures **only** conditional independences that are valid in δ.

An I-map G for δ is called a **perfect map**, if G captures **all** valid conditional independences in δ.

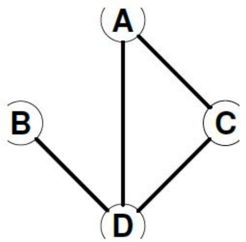
An I-map G for is called **minimal** iff no edge can be removed from G so that the resulting graph is still an I-map for δ.

These definitions hold for directed as well as undirected graphs.

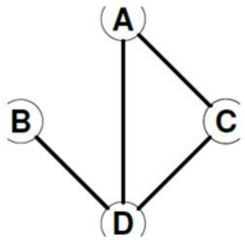
Independence Maps: Examples for undirected graphs



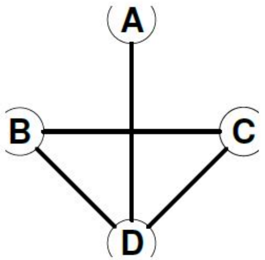
is not an I-map for $\{(A, B|\{C, D\}), (B, C|\{A, D\})$
$(B, A|\{C, D\}), (C, B|\{A, D\}\}$



is I-map for $\{(A, B|\{C, D\}), (B, C|\{A, D\})$
$(B, A|\{C, D\}), (C, B|\{A, D\}\}$



is a perfect I-map for $\{(A, B|\{C, D\}), (B, C|\{A, D\}), (B, \{A, C\}|D), (B, A|D), (B, C|D),$
$(B, A|\{C, D\}), (C, B|\{A, D\}), (\{A, C\}, B|D), (A, B|D), (C, B|D)\}$



is a minimal I-map for $\{(A, B|\{C, D\}), (A, C|\{B, D\}), (A, \{B, C\}|D), (A, B|D), (A, C|D),$
$(B, A|\{C, D\}), (C, A|\{B, D\}), (\{B, C\}, A|D), (B, A|D), (C, A|D)\}$

Independence Maps for Probability Spaces

If a probability P is given, then we can check for subsets X,Y,Z of random variables on P whether X and Y are conditionally independent with respect to Z. As the result we obtain a three-place relation representing a set of **conditional independence statements**

$$X \perp\!\!\!\perp_P Y \mid Z$$

A directed graph G=(U,E) over U is called an **independence map (I-map)** for P, if for all disjoint subsets X,Y,Z of U the property

$$\langle X \mid Z \mid Y \rangle_G \Rightarrow X \perp\!\!\!\perp_P Y \mid Z$$

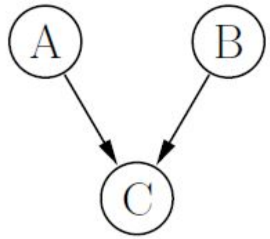
holds.

In an I-map every independence we can observe from G is encoded in P. In most cases the set of independencies we can see from the connectivity in the graph (via d-separation or u-separation) is only a part of the independencies the joint distribution P has. The “ultimate” connection between probability distributions and graphs requires the other implication direction to hold, namely for every conditional independence in the probability distribution to correspond to a separation in the graph. This connection has been called **faithfulness** of the probability distribution and the graph.

An I-map G for P is called a **perfect map**, if G captures **exactly the (**conditional) independences in P.

Limitations of Graph Representations

Perfect directed map, no perfect undirected map:

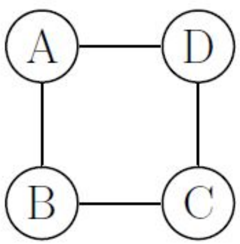


$A \perp\!\!\!\perp_p B \mid \emptyset$

$A \not\perp\!\!\!\perp_p B \mid C$

p_{ABC}	$A = a_1$		$A = a_2$	
	$B = b_1$	$B = b_2$	$B = b_1$	$B = b_2$
$C = c_1$	$^4/_{24}$	$^3/_{24}$	$^3/_{24}$	$^2/_{24}$
$C = c_2$	$^2/_{24}$	$^3/_{24}$	$^3/_{24}$	$^4/_{24}$

Perfect undirected map, no perfect directed map:



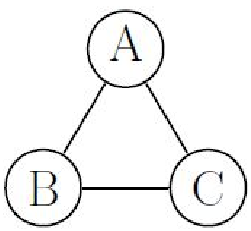
$B \perp\!\!\!\perp_p D \mid \{A,\ C\}$

$A \perp\!\!\!\perp_p C \mid \{B,\ D\}$

p_{ABCD}		$A = a_1$		$A = a_2$	
		$B = b_1$	$B = b_2$	$B = b_1$	$B = b_2$
$C = c_1$	$D = d_1$	$^1/_{47}$	$^1/_{47}$	$^1/_{47}$	$^2/_{47}$
	$D = d_2$	$^1/_{47}$	$^1/_{47}$	$^2/_{47}$	$^4/_{47}$
$C = c_2$	$D = d_1$	$^1/_{47}$	$^2/_{47}$	$^1/_{47}$	$^4/_{47}$
	$D = d_2$	$^2/_{47}$	$^4/_{47}$	$^4/_{47}$	$^{16}/_{47}$

Limitations of Graph Representations

There are also probability distributions for which
there exists neither a directed nor an undirected perfect map:



$$A \perp\!\!\!\perp_p B \mid \emptyset$$
$$A \perp\!\!\!\perp_p C \mid \emptyset$$
$$B \perp\!\!\!\perp_p C \mid \emptyset$$

p_{ABC}	$A = a_1$		$A = a_2$	
	$B = b_1$	$B = b_2$	$B = b_1$	$B = b_2$
$C = c_1$	$^2/_{12}$	$^1/_{12}$	$^1/_{12}$	$^2/_{12}$
$C = c_2$	$^1/_{12}$	$^2/_{12}$	$^2/_{12}$	$^1/_{12}$

In such cases either not all dependences or not all independences

can be captured by a graph representation.
In such a situation one usually decides to neglect some of the independence information, that is, to use only a (minimal) conditional independence graph.

This is sufficient for correct evidence propagation,
the existence of a perfect map is not required.