CS 276: Causal and Probabilistic Reasoning

Rina Dechter

Algorithmic Approach for Identification

Based on Elias Bareinboim slides

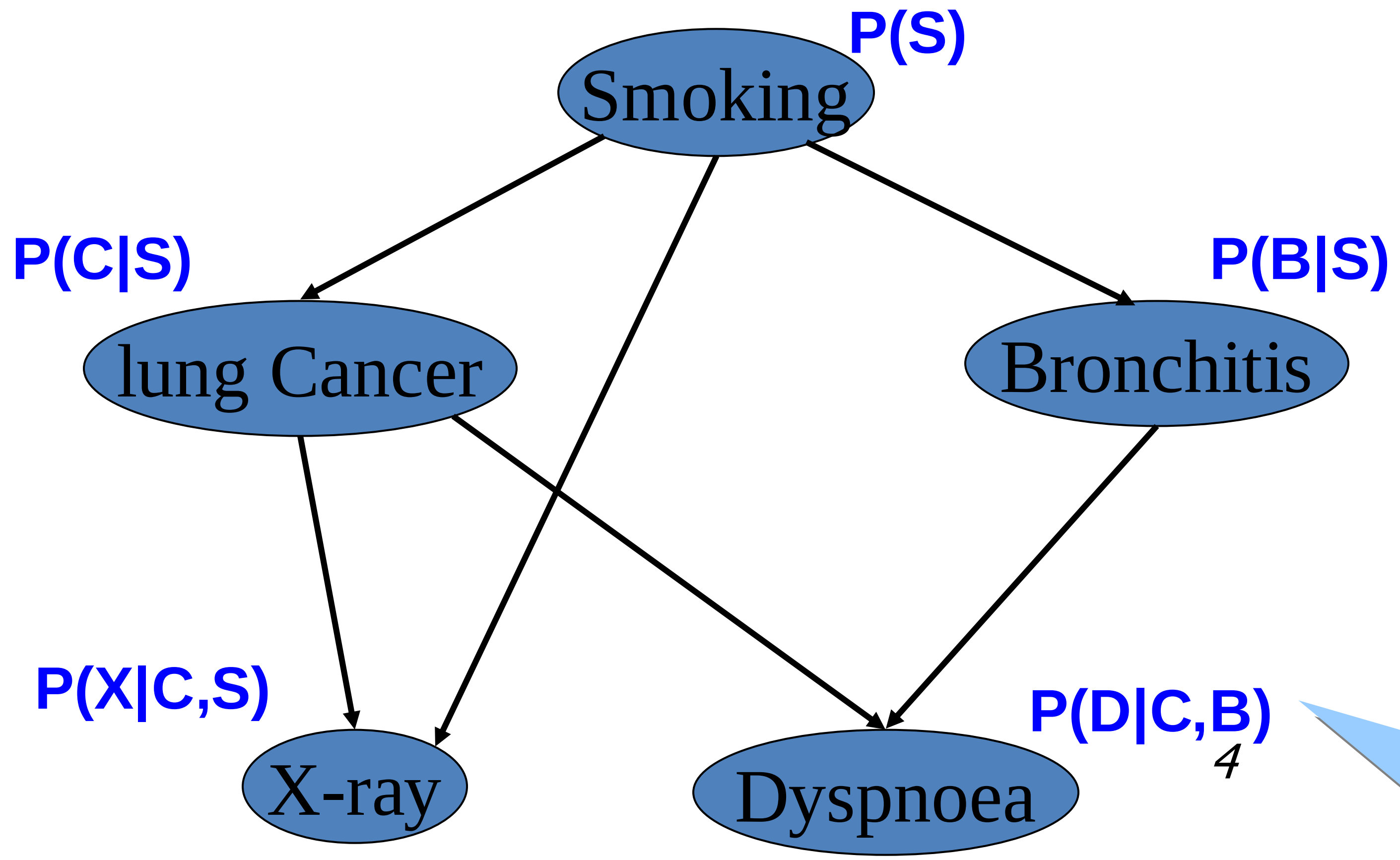
Roadmap

- Define a decomposition (factorization) of the probability distributions generated by a SCM, based on the corresponding causal diagram.
- Establish operations that allows us to identify particular components (factors) from a distribution.
- Express the target causal effect into factors and develop a systematic procedure to identify each one of them independently.

Factorizing Observational Distributions

Bayesian Networks: Example

(Pearl, 1988)



BN = (G, Θ)

CPD:

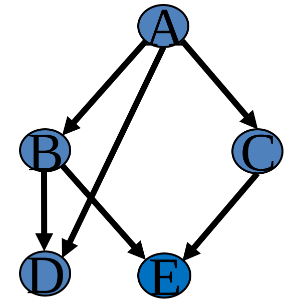
C	B	P(D|C,B)	
0	0	0.1	0.9
0	1	0.7	0.3
1	0	0.8	0.2
1	1	0.9	0.1

$P(S, C, B, X, D) = P(S)\, P(C|S)\, P(B|S)\, P(X|C,S)\, P(D|C,B)$

Belief Updating:

P (lung cancer=yes | smoking=no, dyspnoea=yes) = ?

Belief Updating



Algorithm *BE-bel* [Dechter 1996]

$$p(A|E=0) = \alpha \sum_{e,d,c,b} p(A)\, p(b|A)\, p(c|A)\, p(d|A,b)\, p(e|b,c)\, \mathbb{1}[e=0]$$

$\sum_b \prod$ ← Elimination & combination operators

bucket B: $p(b|A)\ p(d|b,A)\ p(e|b,c)$

bucket C: $p(c|A)\quad \lambda_{B\to C}(A,d,c,e)$

bucket D: $\lambda_{C\to D}(A,d,e)$

bucket E: $\mathbb{1}[E=0]\ \lambda_{D\to E}(A,e)$

bucket A: $p(A)\quad \lambda_{E\to A}(A)$

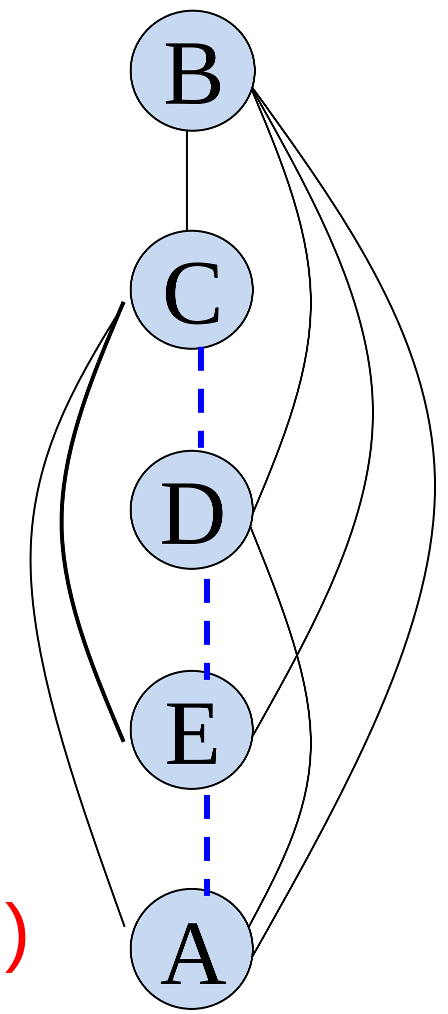
$p(E=0)$

$p(A|E=0) = p(A,E=0)\,/\,p(E=0)$

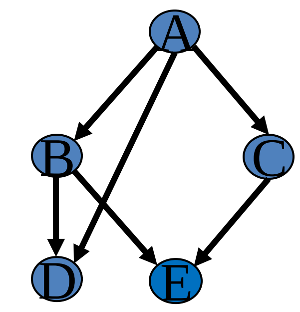
W*=4

"induced width"

(max clique size)



Bucket Elimination



Algorithm *BE-bel* [Dechter 1996]

$$p(A|E=0) = \alpha \sum_{e,d,c,b} p(A)\, p(b|A)\, p(c|A)\, p(d|A,b)\, p(e|b,c)\, \mathbb{1}[e=0]$$

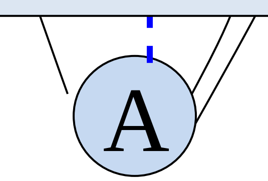
$\sum_b \prod$ ← Elimination & combination operators

Time and space exponential in the induced-width / treewidth

8

bucket A: $p(A)$ $\lambda_{E \to A}(A)$

induced width (max clique size)



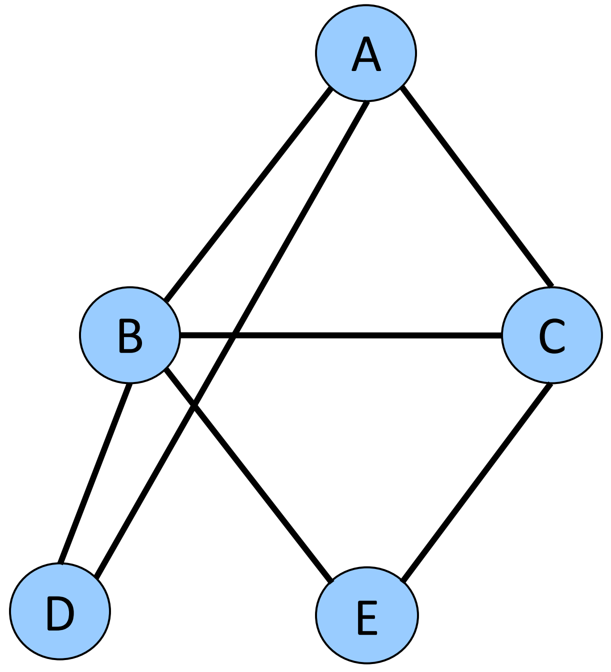
$p(E=0)$

$p(A|E=0) = p(A, E=0) \,/\, p(E=0)$

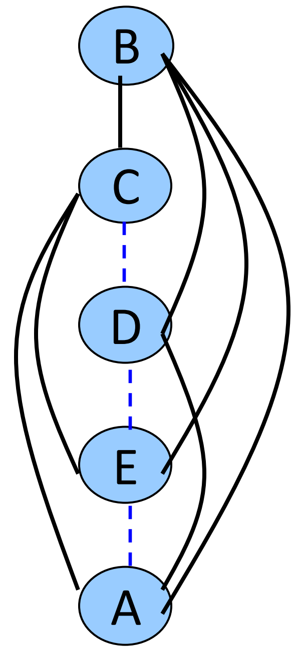
Induced Width (continued)

$w^*(d)$ – the induced width of the primal graph along ordering d

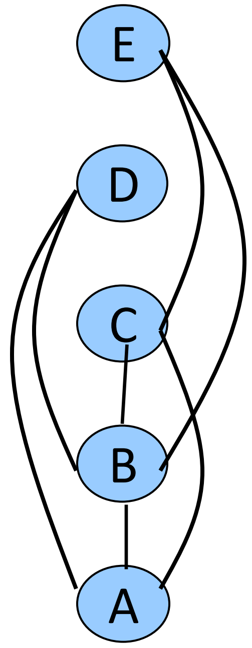
The effect of the ordering:



Primal (moraal) graph



$$w^*(d_1) = 4$$



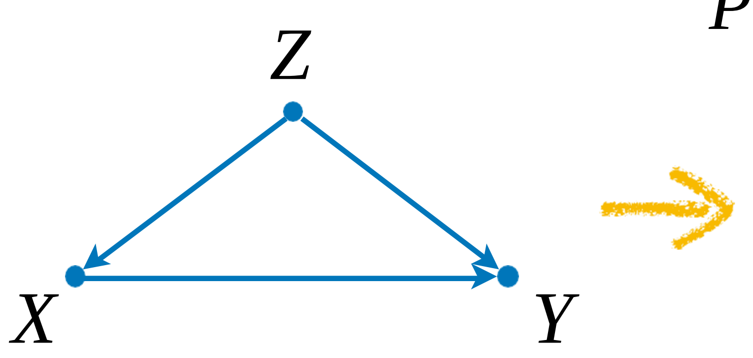
$$w^*(d_2) = 2$$

Back to SCM

Markovian Case

- The distribution $P(\mathbf{v})$ decomposes as:

$$P(\mathbf{v}) = \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{V_i \in \mathbf{V}} P(v_i \mid v_1, \ldots, v_{i-1}, \mathbf{u}) = \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{V_i \in \mathbf{V}} P(v_i \mid pa_i, u_i)$$



$$P(z, x, y) = \sum_{\mathbf{u}} P(\mathbf{u}) P(z \mid u_z) P(x \mid z, u_x) P(y \mid x, z, u_y)$$

$$= \left(\sum_{u_z} P(z \mid u_z) P(u_z) \right) \left(\sum_{u_x} P(x \mid z, u_x) P(u_x) \right) \left(\sum_{u_y} P(y \mid x, z, u_y) P(u_y) \right)$$

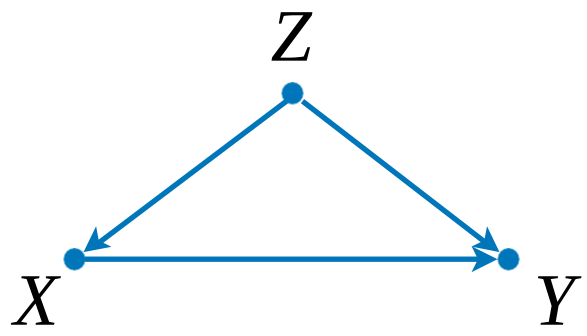
$$= P(z) P(x \mid z) P(y \mid x, z)$$

- In Markovian models, $P(v_i \mid pa_i)$ can be seen as "canonical factors".

Markovian Case

- Every $P(v_i|pa_i)$ is computable from $P(\mathbf{v})$, *i.e.*,

$$P(v_i \mid pa_i) = \frac{\sum_{v \setminus v_i, pa_i} P(v)}{\sum_{v \setminus pa_i} P(v)}$$



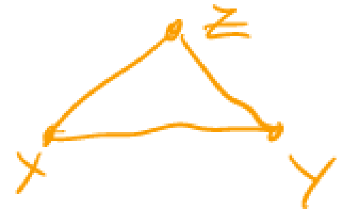
$$P(z, x, y) = P(z)P(x \mid z)P(y \mid x, z)$$

$$P(z) = \sum_{x,y} P(v)$$

$$P(y \mid x, z) = \frac{P(v)}{\sum_y P(v)}$$

$$P(x \mid z) = \frac{\sum_y P(v)}{\sum_{x,y} P(v)}$$

Markovian Case



U_Z: $P(U_Z)$, $P(Z|U_Z)$ →

U_X: $P(U_X)$, $P(X|U_X, Z)$

U_Y: $P(U_Y)$, $P(Y|Z, X, U_Y)$

X, Y, Z:

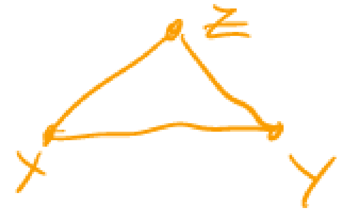
$$\lambda_{U_Z}(Z) = \sum_{U_Z} P(Z|U_Z) \cdot P(U_Z) = ?$$

$$\lambda_{U_X}(X, Z) = \sum_{U_X} P(X|U_X, Z) \cdot P(U_X) = ?$$

$$\lambda_{U_Y}(Y, X, Z) = \sum_{U_Y} P(Y|Z, X, U_Y) \cdot P(U_Y) = ?$$

$$P(X, Y, Z) = \lambda_{U_Z}(Z) \cdot \lambda_{U_X}(X, Z) \cdot \lambda_{U_Y}(Y, X, Z)$$

Markovian Case



U_Z: $P(U_Z)$, $P(Z|U_Z)$ →

U_X: $P(U_X)$, $P(X|U_X,Z)$

U_Y: $P(U_Y)$, $P(Y|Z,X,U_Y)$

X, Y, Z:

$$\lambda_{U_Z}(Z) = \sum_{U_Z} P(Z|U_Z)\cdot P(U_Z) = ?\ P(Z)$$

$$\lambda_{U_X}(X,Z) = \sum_{U_X} P(X|U_X,Z)\cdot P(U_X) = ?\ P(X|Z)$$

$$\lambda_{U_Y}(Y,X,Z) = \sum_{U_Y} P(Y|Z,X,U_Y)\cdot P(U_Y) = ?\ P(Y|Z,X)$$

$$P(X,Y,Z) = \lambda_{U_Z}(Z)\cdot\lambda_{U_X}(X,Z)\cdot\lambda_{U_Y}(Y,X,Z)$$

P(X,Y,Z) = P(Z)P(X|Z)P(Y|X,Y,Z)

Truncated Product in Semi-Markovian Models

The distribution generated by an intervention *do(X=x)* in a Semi-Markovian model *M* is given by the (generalized) truncated factorization product, namely,

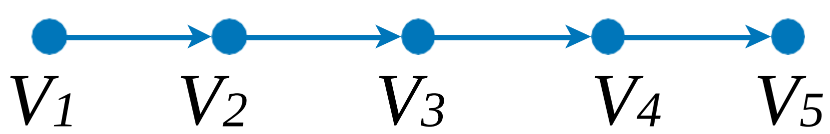
$$P(\mathbf{v} \mid do(\mathbf{x})) = \sum_{\mathbf{u}} \prod_{\{V_i \in \mathbf{V} \setminus \mathbf{X}\}} P\left(v_i \mid pa_i, u_i\right) P(\mathbf{u})$$

And the effect of such intervention on a set *Y* is

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{v} \setminus (\mathbf{y} \cup \mathbf{x})} \sum_{\mathbf{u}} \prod_{\{V_i \in \mathbf{V} \setminus \mathbf{X}\}} P\left(v_i \mid pa_i, u_i\right) P(\mathbf{u})$$

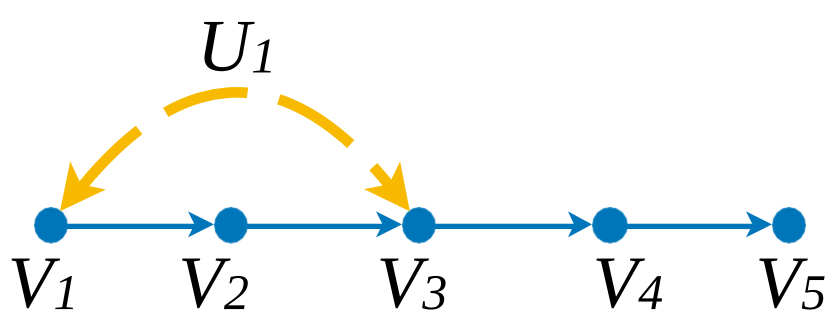
Semi-Markovian Case

- Start from a simple Markovian model:



$$P(\mathbf{v}) = P(v_1)P(v_2 \mid v_1)P(v_3 \mid v_2)P(v_4 \mid v_3)P(v_5 \mid v_4)$$

- Let's add an unobservable U_1, that affects two observables, and breaking Markovianity:



$$P(\mathbf{v}) = \sum_{u_1} P(u_1)P(v_1 \mid u_1)P(v_2 \mid v_1)P(v_3 \mid v_2, u_1)P(v_4 \mid v_3)P(v_5 \mid v_4)$$

$$= P(v_2 \mid v_1)P(v_4 \mid v_3)P(v_5 \mid v_4)\left(\sum_{u_1} P(u_1)P(v_1 \mid u_1)P(v_3 \mid v_2, u_1)\right)$$

Using Bucket Elimination

U_1: $P(U_1),\ P(V_1|U_1),\ P(V_3|V_2,U_1)$

V_3: $P(V_4|V_3)\ \lambda(V_1,V_3,V_2) =$

V_2: $P(V_2|V_1)\ \sum_{U_1} P(V_3|V_2,U_1)\cdot P(V_1|U_1)\cdot P(U_1)$

V_1:

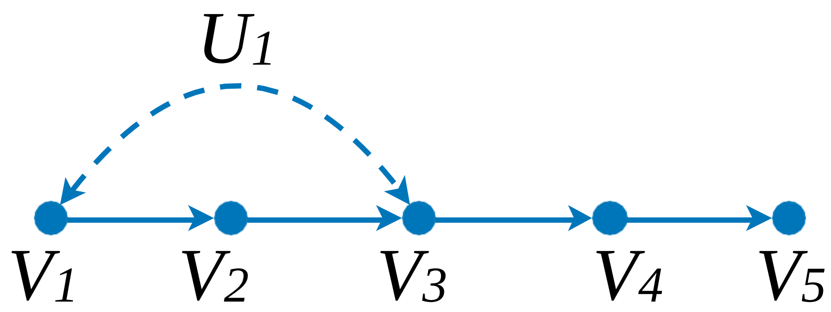
V_4: $P(V_5|V_4)$

V_5:

Can this be expressed using P(V) only?

Semi-Markovian Case

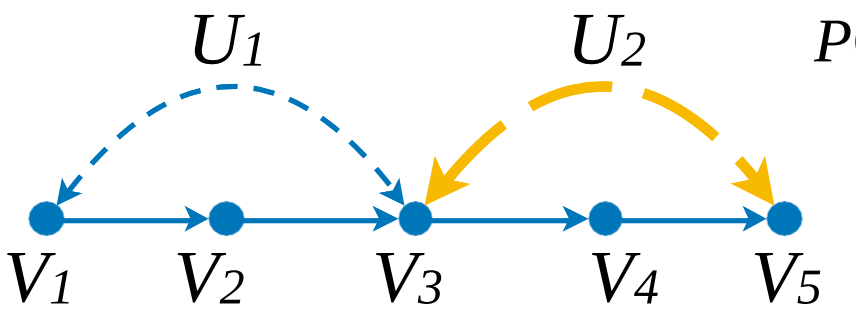
- From the previous model …



$$P(\mathbf{v}) = \sum_{u_1} P(u_1)P(v_1 | u_1)P(v_2 | v_1)P(v_3 | v_2, u_1)P(v_4 | v_3)P(v_5 | v_4)$$

$$= P(v_2 | v_1)P(v_4 | v_3)P(v_5 | v_4)\left(\sum_{u_1} P(u_1)P(v_1 | u_1)P(v_3 | v_2, u_1) \right)$$

- Add another unobservable U_2,

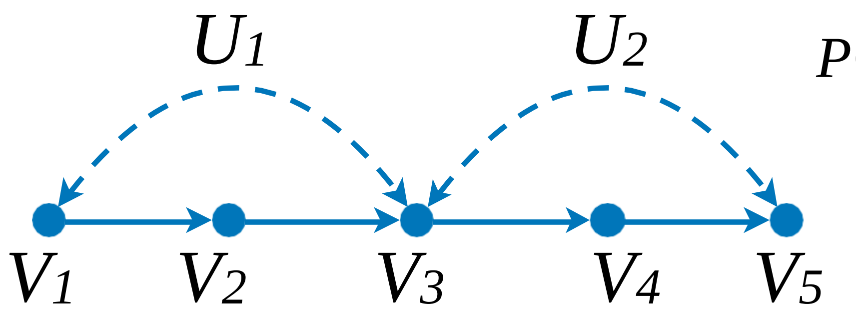


$$P(\mathbf{v}) = \sum_{u_1, u_2} P(u_1, u_2)P(v_1 | u_1)P(v_2 | v_1)P(v_3 | v_2, u_1, u_2)P(v_4 | v_3)P(v_5 | v_4, u_2)$$

$$= P(v_2 | v_1)P(v_4 | v_3)\left(\sum_{u_1, u_2} P(u_1, u_2)P(v_1 | u_1)P(v_3 | v_2, u_1, u_2)P(v_5 | v_4, u_2) \right)$$

Semi-Markovian Case

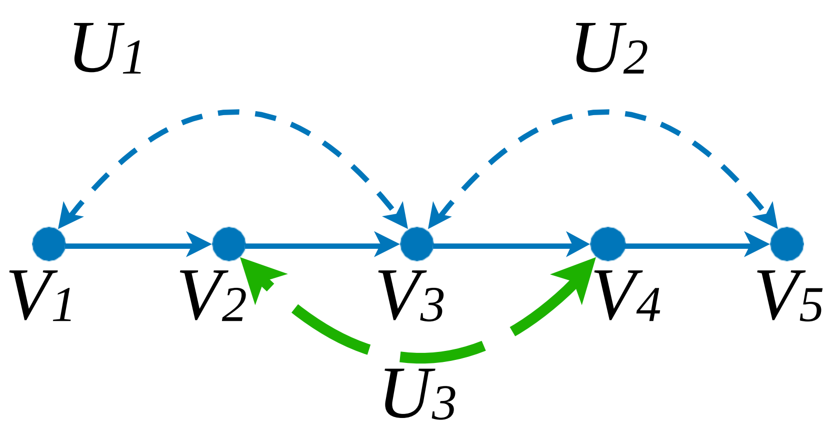
- From the previous model…



$$P(\mathbf{v}) = \sum_{u_1,u_2} P(u_1,u_2)P(v_1 \mid u_1)P(v_2 \mid v_1)P(v_3 \mid v_2,u_1,u_2)P(v_4 \mid v_3)P(v_5 \mid v_4,u_2)$$

$$= P(v_2 \mid v_1)P(v_4 \mid v_3)\left(\sum_{u_1,u_2} P(u_1,u_2)P(v_1 \mid u_1)P(v_3 \mid v_2,u_1,u_2)P(v_5 \mid v_4,u_2)\right)$$

- Let's add one more, U_3,

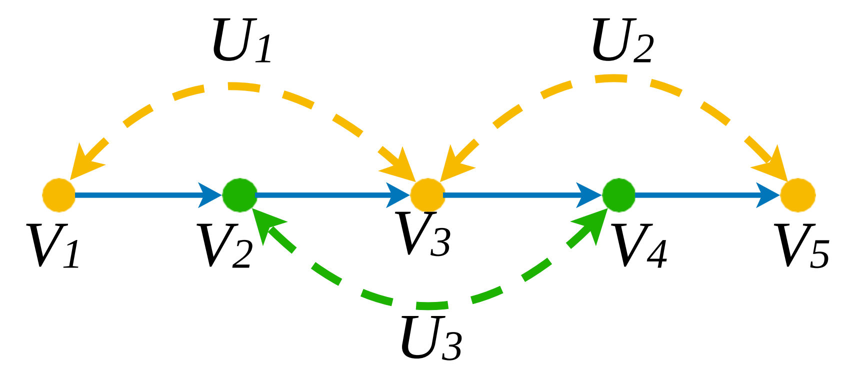


$$P(\mathbf{v}) = \sum_{u_1,u_2,u_3} P(u_1,u_2,u_3)P(v_1 \mid u_1)P(v_2 \mid v_1,u_3)P(v_3 \mid v_2,u_1,u_2) P(v_4 \mid v_3,u_3)P(v_5 \mid v_4,u_2)$$

$$= \left(\sum_{u_3} P(u_3)P(v_2 \mid v_1,u_3)P(v_4 \mid v_3,u_3)\right)\left(\sum_{u_1,u_2} P(u_1,u_2)P(v_1 \mid u_1)P(v_3 \mid v_2,u_1,u_2)P(v_5 \mid v_4,u_2)\right)$$

C-Factors

- Recall our example



$$P(\mathbf{v}) = \left(\sum_{u_3} P(u_3)P(v_2 \mid v_1, u_3)P(v_4 \mid v_3, u_3) \right) \left(\sum_{u_1,u_2} P(u_1, u_2)P(v_1 \mid u_1)P(v_3 \mid v_2, u_1, u_2)P(v_5 \mid v_4, u_2) \right)$$

- These factors made of sums may be long to write in terms of $P(\mathbf{v},\mathbf{u})$. However, their structure follows from the topology of the diagram, then we can abstract this concept out by defining a new function Q:

$$Q[\mathbf{C}](\mathbf{c}, pa_{\mathbf{c}}) = \sum_{u(\mathbf{C})} P(u(\mathbf{C})) \prod_{V_i \in \mathbf{C}} P(v_i \mid pa_i, u_i) \quad \text{where} \quad U(\mathbf{C}) = \bigcup_{V_i \in \mathbf{C}} U_i$$

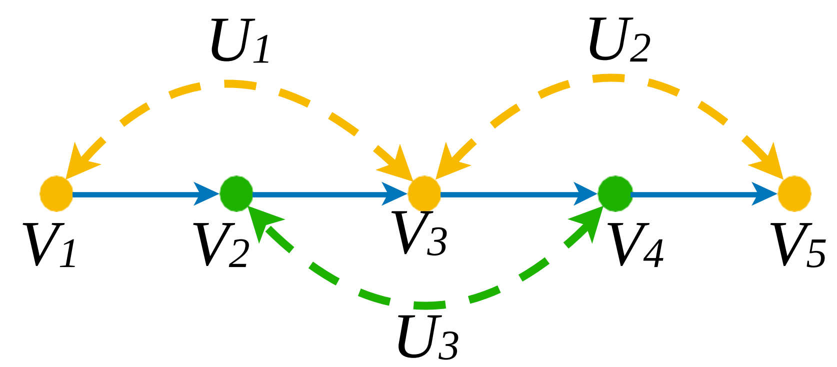
Then $P(\mathbf{v})$ can be re-written as

$$P(\mathbf{v}) = Q[V_2, V_4](v_2, v_4, v_1, v_3)Q[V_1, V_3, V_5](v_1, v_3, v_5, v_2, v_4)$$

C-factors

- For convenience $Q[\mathbf{C}](\mathbf{c}, pa_c)$ can be written just as $Q[\mathbf{C}]$
- Then, for our example, we can just write

$$P(\mathbf{v}) = Q[V_2, V_4]Q[V_1, V_3, V_5]$$



- No need to name the variables in $\boldsymbol{U}$ explicitly!
- Note that for the whole set of variables V

$$Q[\mathbf{V}] = \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{V_i \in \mathbf{V}} P(v_i \mid pa_i, u_i) = P(\mathbf{v})$$)

- For consistency define $Q[\varnothing]=1$

C-factors are Causal Effects

- Let $\boldsymbol{C} \subseteq \boldsymbol{V}$. Consider the causal effect of all other variables on $\boldsymbol{C}$, that is $P(\mathbf{c}|do(\mathbf{v}\backslash\mathbf{c}))$.

- By the truncated product we have

$$P(\mathbf{c} \,|\, do(\mathbf{v}\backslash\mathbf{c})) = \sum_{\mathbf{u}} \prod_{V_i \in \mathbf{C}} P(v_i \,|\, pa_i, u_i) P(\mathbf{u})$$

- All $\boldsymbol{U}$s that are not parents of any element in $\boldsymbol{C}$ can be summed out, hence

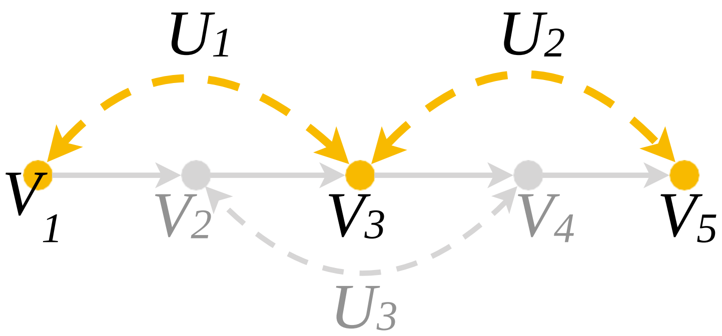
$$\boxed{P(\mathbf{c} \,|\, do(\mathbf{v}\backslash\mathbf{c}))} = \sum_{u(\mathbf{C})} P(u(\mathbf{C})) \prod_{V_i \in \mathbf{C}} P(v_i \,|\, pa_i, u_i) \boxed{= Q[C]}$$

This is a key connection between C-factors and causal effects.

Marginalizing Variables in C-factors

- To a certain extent, c-factors behave as its probabilistic counterparts.
- Consider the c-factor $Q[V_1,V_3,V_5]$ in our example

$$Q[V_1,V_3,V_5] = \sum_{u_1,u_2} P(u_1,u_2)P(v_1 \mid u_1)P(v_3 \mid v_2,u_1,u_2)P(v_5 \mid v_4,u_2)$$

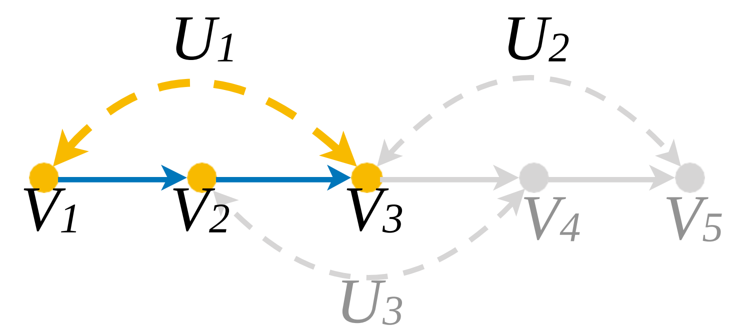


- Variables V_1,V_3,V_5 only appear in one term, because they are not the parent of any other variable in the factor. So, if we sum $Q[V_1,V_3,V_5]$ over any of the vraiables, for instance V_3, we have

$$\sum_{v_3} Q[V_1,V_3,V_5] = \sum_{v_3}\sum_{u_1,u_2} P(u_1,u_2)P(v_1,u_1)P(v_3 \mid v_2,u_1,u_2)P(v_5 \mid v_4,u_2)$$

$$= \sum_{u_1,u_2} P(u_1,u_2)P(v_1 \mid u_1)P(v_5 \mid v_4,u_2) = Q[V_1,V_5]$$

Marginalizing Variables in C-factors



- Consider now a different c-factor $Q[V_1,V_2,V_3]$,

  By Definition of Q:

$$Q[V_1, V_2, V_3] = \sum_{u_1,u_2,u_3} P(u_1,u_2,u_3)P(v_1 \mid u_1)P(v_2 \mid v_1,u_3)P(v_3 \mid v_2,u_1,u_2)$$

- In contrast to the previous case, here V_1 appears in two terms since it's a parent of another variable in the factor. So, if we sum $Q[V_1,V_2,V_3]$ over V_1, we have

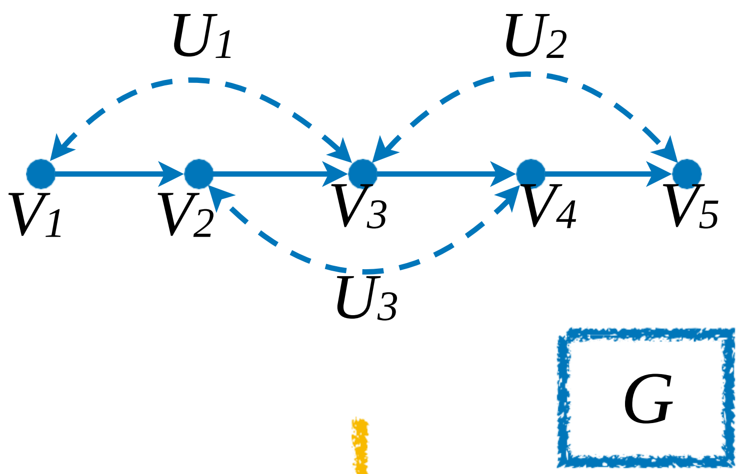
$$\sum_{v_1} Q[V_1, V_2, V_3] = \sum_{v_1} \sum_{u_1,u_2,u_3} P(u_1,u_2,u_3)P(v_1 \mid u_1)P(v_2 \mid v_1,u_3)P(v_3 \mid v_2,u_1,u_2)$$

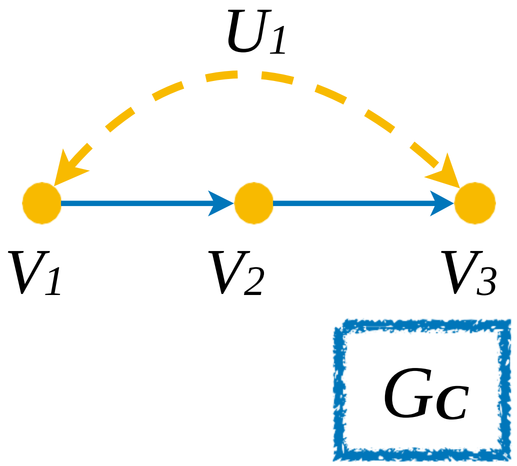
$$= \sum_{u_1,u_2} P(u_1,u_2)P(v_3 \mid v_2,u_1,u_2) \sum_{v_1,u_3} P(u_3)P(v_1 \mid u_1)P(v_2 \mid v_1,u_3) \neq Q[V_2, V_3]$$

Can we remove V1 here? Symbolically?

Marginalizing Variables in C-factors

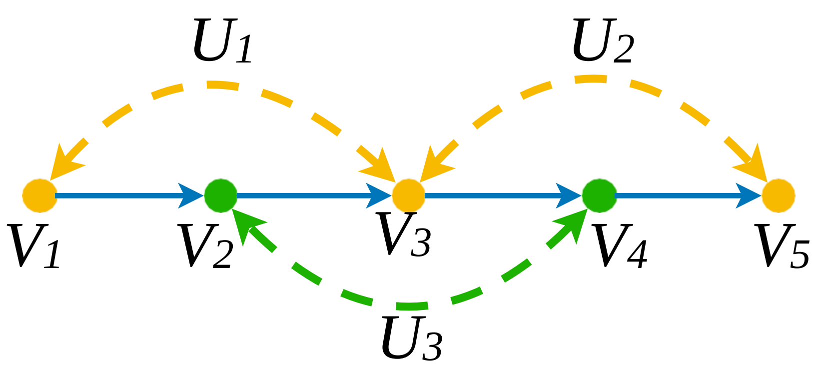
- Let $\boldsymbol{W} \subset \boldsymbol{C} \subseteq \boldsymbol{V}$, be two sets of variables.
- Lemma (ancestral-reduction). If $\boldsymbol{W}$ is ancestral, that is, it contains all $An(\boldsymbol{W})$ present in the subgraph made of the variables in $\boldsymbol{C}$, i.e., G_C.
- Then, $$Q[\mathbf{W}] = \sum_{\mathbf{c} \setminus \mathbf{w}} Q[\mathbf{C}]$$
- For example, for $\boldsymbol{C}=\{V_1,V_2,V_3\}$, in G_C
  - $\boldsymbol{W}=\{V_1,V_2\}$ is ancestral $$Q[V_1, V_2] = \sum_{v_3} Q[V_1, V_2, V_3]$$
  - $\boldsymbol{W}=\{V_1\}$ is ancestral $$Q[V_1] = \sum_{v_2, v_3} Q[V_1, V_2, V_3]$$
  - $\boldsymbol{W}=\{V_2,V_3\}$ is not ancestral





Confounded Components (C-Components)

- Recall our example



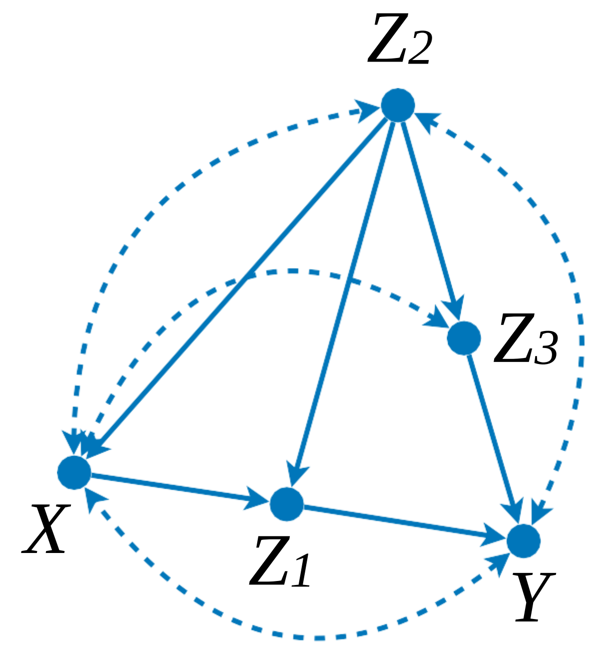
$$P(\mathbf{v}) = \left(\sum_{u_3} P(u_3) P(v_2 \mid v_1, u_3) P(v_4 \mid v_3, u_3) \right) \left(\sum_{u_1, u_2} P(u_1, u_2) P(v_1 \mid u_1) P(v_3 \mid v_2, u_1, u_2) P(v_5 \mid v_4, u_2) \right)$$

- Definition (C-component). When V_i and V_j share a common unobservable parent U, $P(v_i \mid pa_i, u_i)$ and $P(v_j \mid pa_j, u_j)$ are tied together by the sum over $U \in U_i \cap U_j$. Then, we say that V_i and V_j are in the same confounded component (C-Component, for short).

We are interested in C-Factors over C-Components

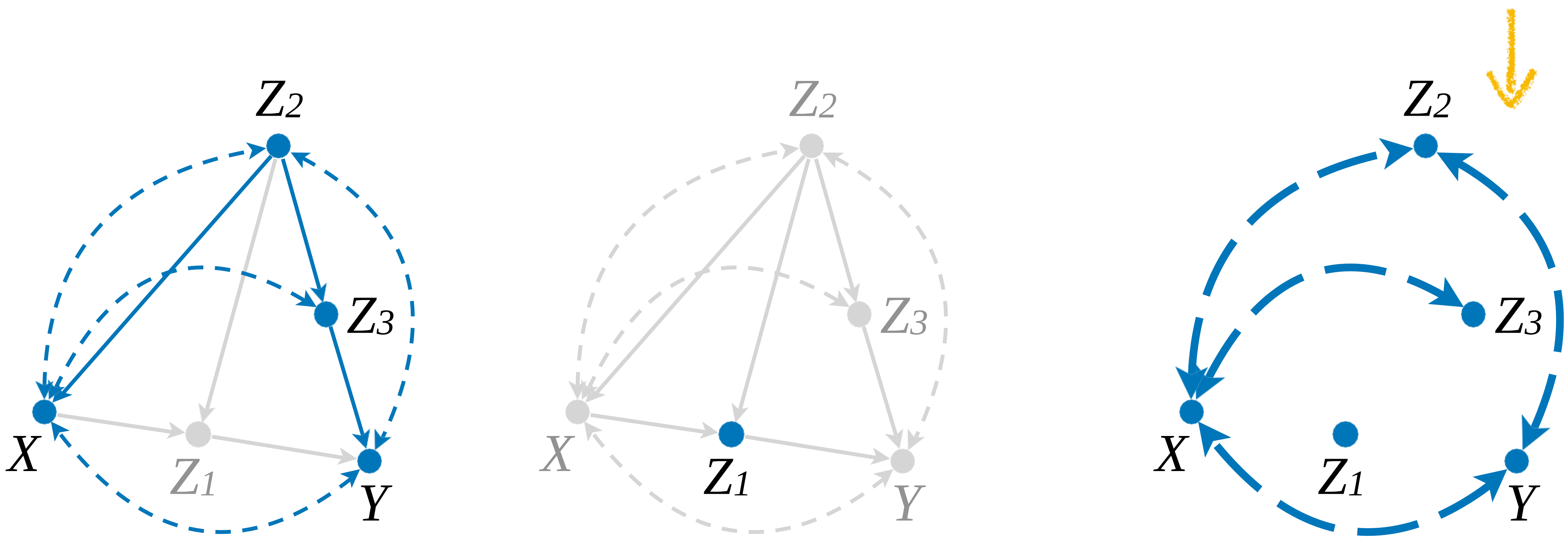
C-Component Factorization

- Consider another example

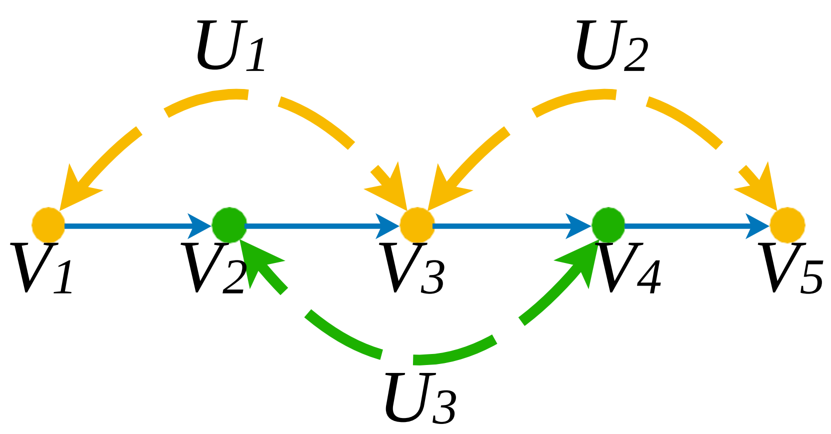


$$Q[\mathbf{V}] = Q[Z_2, Z_3, X, Y]Q[Z_1]$$

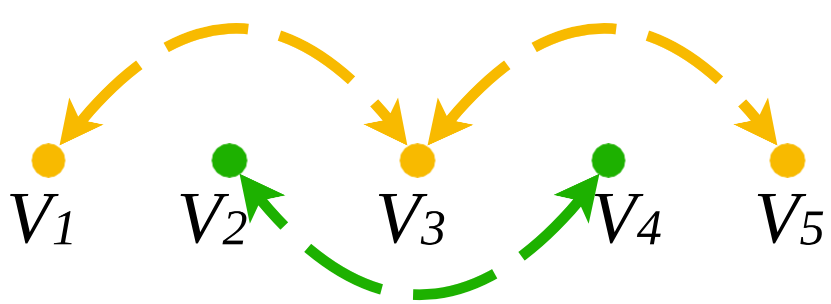
$$P(\mathbf{v}) = P(z_2, z_3, x, y \mid do(z_1))P(z_1 \mid do(z_2, z_3, x, y))$$



C-Component Relationship



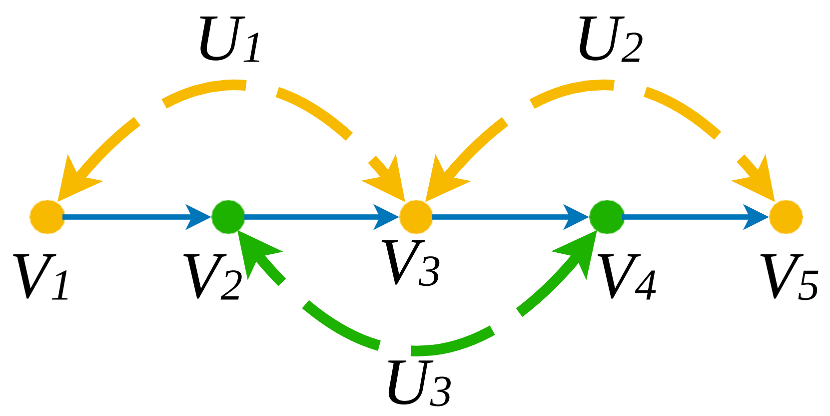
- V_1 is in the same c-component as V_3,
- V_3 is in the same c-component as V_5,
- By extension, V_1 is in the same c-component as V_5 too.
- V_2 is in the same c-component as V_4.



- To see it easily, consider the graph induced over the bidirected edges!
- Obs. The C-Component relation defines a partition over the observable variables, hence it is *Reflexive*, *Symmetric* and *Transitive*.

C-Component Factorization

- The distribution $P(\mathbf{v})$ factorizes into c-factors associated with the c-components of the graph.



$$Q_1 = \{V_2, V_4\} \quad Q_2 = \{V_1, V_3, V_5\}$$

$$P(\mathbf{v}) = \left(\sum_{u_3} P(u_3)P(v_2 \mid v_1, u_3)P(v_4 \mid v_3, u_3) \right) \left(\sum_{u_1,u_2} P(u_1, u_2)P(v_1 \mid u_1)P(v_3 \mid v_2, u_1, u_2)P(v_5 \mid v_4, u_2) \right)$$

$$P(\mathbf{v}) = Q[V_2, V_4]\, Q[V_1, V_3, V_5]$$

C-Component Factorization

- For any $\boldsymbol{H} \subseteq \boldsymbol{V}$, consider a graph $G_{\boldsymbol{H}}$.
- Let $H_1, H_2, \ldots, H_k$ be the c-components of $G_{\boldsymbol{H}}$.
- Then

$$Q[\mathbf{H}] = \prod_j Q[H_j]$$

And, the C- factor of any C component can be computed from Q(H). We will next see how.

C-Component Factorization

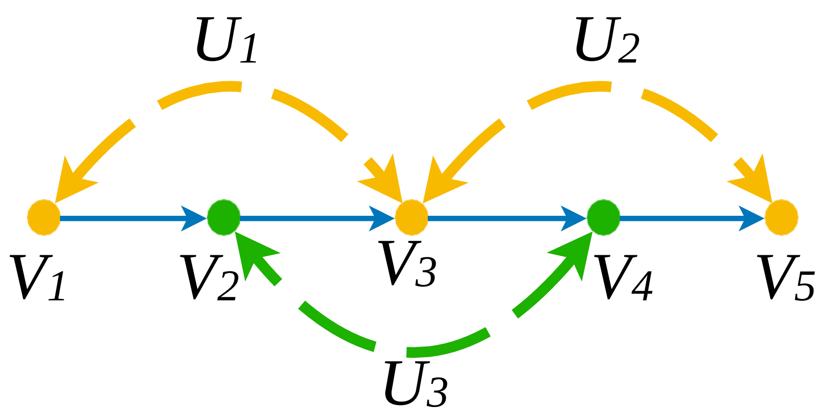
(Continued)

- Let $V_{h1} < V_{h2} < \ldots < V_{hn}$ be a topological order over the variables in $\boldsymbol{H}$ according to G.
- Let $H^{\leq i}$ be the variables in $\boldsymbol{H}$ that come before V_{hi}, including V_{hi}.
- Let $H^{>i}$ be the variables in $\boldsymbol{H}$ that come after V_{hi}.
- Then

$$Q[H_j] = \prod_{V_i \in H_j} \frac{Q[\mathbf{H}^{\leq i}]}{Q[\mathbf{H}^{\leq i-1}]} \qquad Q[\mathbf{H}^{\leq i}] = \sum_{h^{>i}} Q[\mathbf{H}]$$

C-Component Factorization

- Suppose $\boldsymbol{H}=\mathbf{V}=\{V_1,V_2,V_3,V_4,V_5\}$ is ancestral in G_C.



$$Q[\mathbf{V}] = Q[V_1, V_3, V_5]Q[V_2, V_4]$$

$$Q[V_1, V_3, V_5] = \frac{Q[V_1]}{Q[\varnothing]} \frac{Q[V_1, V_2, V_3]}{Q[V_1, V_2]} \frac{Q[V_1, V_2, V_3, V_4, V_5]}{Q[V_1, V_2, V_3, V_4]}$$

$$Q[V_2, V_4] = \frac{Q[V_1, V_2]}{Q[V_1]} \frac{Q[V_1, V_2, V_3, V_4]}{Q[V_1, V_2, V_3]}$$

$$Q[H_j] = \prod_{V_i \in H_j} \frac{Q[\mathbf{H}^{\leq i}]}{Q[\mathbf{H}^{\leq i-1}]}$$

C-Component Factorization

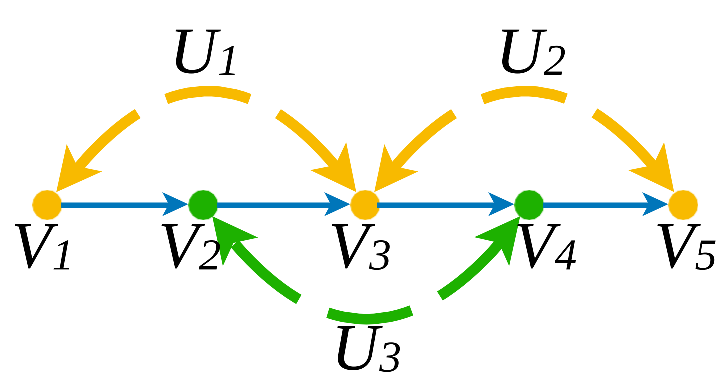
$$Q[V_1, V_3, V_5] = \frac{Q[V_1]}{Q[\emptyset]} \frac{Q[V_1, V_2, V_3]}{Q[V_1, V_2]} \frac{Q[V_1, V_2, V_3, V_4, V_5]}{Q[V_1, V_2, V_3, V_4]}$$

$$= \frac{\sum_{v_2,v_3,v_4,v_5} Q[\mathbf{V}]}{\sum_{v_1,v_2,v_3,v_4,v_5} Q[\mathbf{V}]} \frac{\sum_{v_4,v_5} Q[\mathbf{V}]}{\sum_{v_3,v_4,v_5} Q[\mathbf{V}]} \frac{\sum_{\emptyset} Q[\mathbf{V}]}{\sum_{v_5} Q[\mathbf{V}]}$$

$$= \frac{\sum_{v_2,v_3,v_4,v_5} P(\mathbf{v})}{\sum_{v_1,v_2,v_3,v_4,v_5} P(\mathbf{v})} \frac{\sum_{v_4,v_5} P(\mathbf{v})}{\sum_{v_3,v_4,v_5} P(\mathbf{v})} \frac{\sum_{\emptyset} P(\mathbf{v})}{\sum_{v_5} P(\mathbf{v})}$$

$$Q[\mathbf{H}^{\leq i}] = \sum_{h^{>i}} Q[\mathbf{H}]$$

$$= \frac{P(v_1)}{1} \frac{P(v_1, v_2, v_3)}{P(v_1, v_2)} \frac{P(v_1, v_2, v_3, v_4, v_5)}{P(v_1, v_2, v_3, v_4)}$$



$$= P(v_1)P(v_3 \mid v_1, v_2)P(v_5 \mid v_1, v_2, v_3, v_4)$$

C-Component Factorization

$$Q[V_2, V_4] = \frac{Q[V_1, V_2]}{Q[V_1]} \frac{Q[V_1, V_2, V_3, V_4]}{Q[V_1, V_2, V_3]}$$

$$= \frac{\sum_{v_3,v_4,v_5} Q[\mathbf{V}]}{\sum_{v_2,v_3,v_4,v_5} Q[\mathbf{V}]} \frac{\sum_{v_5} Q[\mathbf{V}]}{\sum_{v_4,v_5} Q[\mathbf{V}]}$$

$$= \frac{\sum_{v_3,v_4,v_5} P(\mathbf{v})}{\sum_{v_2,v_3,v_4,v_5} P(\mathbf{v})} \frac{\sum_{v_5} P(\mathbf{v})}{\sum_{v_4,v_5} P(\mathbf{v})}$$

$$= \frac{P(v_1, v_2)}{P(v_1)} \frac{P(v_1, v_2, v_3, v_4)}{P(v_1, v_2, v_3)}$$

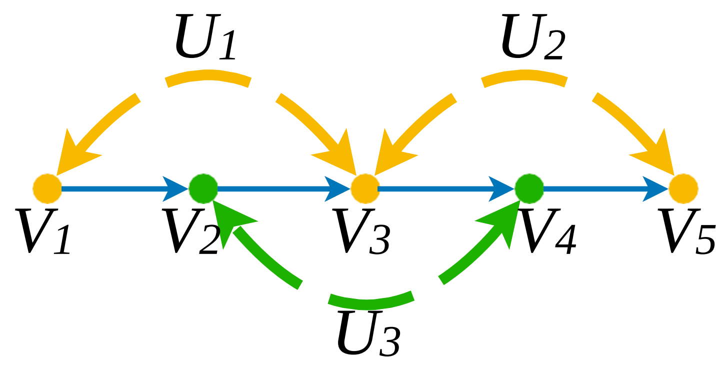
$$= P(v_2 \mid v_1) P(v_4 \mid v_1, v_2, v_3)$$

How to get just $Q[V_2]$ or $Q[V_4]$?

Both are ancestral in $G_{\{V_2,V_4\}}$!

$$Q[V_2] = \sum_{v_4} Q[V_2, V_4] = P(v_2 \mid v_1)$$

$$Q[V_4] = \sum_{v_2} Q[V_2, V_4] = \sum_{v_2} P(v_2 \mid v_1) P(v_4 \mid v_1, v_2, v_3)$$



C-Component Factorization

$$Q[V_2, V_4] = \frac{Q[V_1, V_2]}{Q[V_1]} \frac{Q[V_1, V_2, V_3, V_4]}{Q[V_1, V_2, V_3]}$$

$$= \frac{\sum_{v_3,v_4,v_5} Q[\mathbf{V}]}{\sum_{v_2,v_3,v_4,v_5} Q[\mathbf{V}]} \frac{\sum_{v_5} Q[\mathbf{V}]}{\sum_{v_4,v_5} Q[\mathbf{V}]}$$

$$= \frac{\sum_{v_3,v_4,v_5} P(\ldots}{\sum_{v_2,v_3,v_4,v_5} P\ldots}$$

$$= \frac{P(v_1, v_2)}{P(v_1)} \frac{P(v_1, v_2, v_3, v_4)}{P(v_1, v_2, v_3)}$$

$$= P(v_2 \mid v_1) P(v_4 \mid v_1, v_2, v_3)$$

How to get just $Q[V_2]$ or $Q[V_4]$?

Both are ancestral in $G_{\{V_2,V_4\}}$!

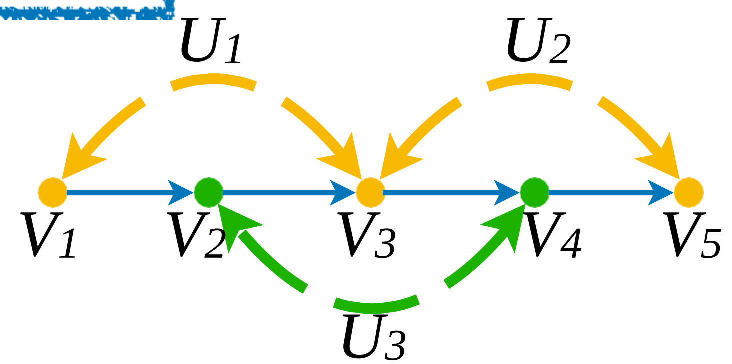
$$Q[V_2] = \sum_{v_4} Q[V_2, V_4]$$

$$= P(v_2 \mid v_1)$$

$$Q[V_4] = \ldots Q[V_2, V_4]$$

$$\ldots \mid v_1) P(v_4 \mid v_1, v_2, v_3)$$

Notice that these c-factors are expressible in terms of the obs. distribution (no ***U***-terms)!



C-factor Algebra - Summary

We have two basic operations over c-factors

1. Reduce to an ancestral set

$$Q[\mathbf{W}] = \sum_{\mathbf{c} \setminus \mathbf{w}} Q[\mathbf{C}]$$

If $\boldsymbol{W}$ is ancestral in G_C

2. Factorize into c-components

$$Q[\mathbf{H}] = \prod_j Q[H_j]$$

Where $H_1, \ldots H_k$, are the c-components in $G_{\boldsymbol{H}}$

Each Q[Hj] is identifiable from Q[H]

C-Factor (component)

Definition C-factor or C-component

A *c-component* (short for "confounded component," [3]) of variable set V on graph G consists of all the unobservable variables belonging to the same c-component related part of U and all observable variables that have an unobservable parent which is a member of that c-component.

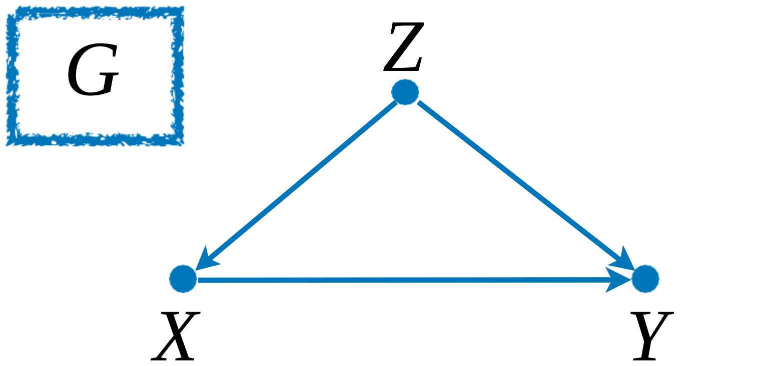
Definition of Ancestral set

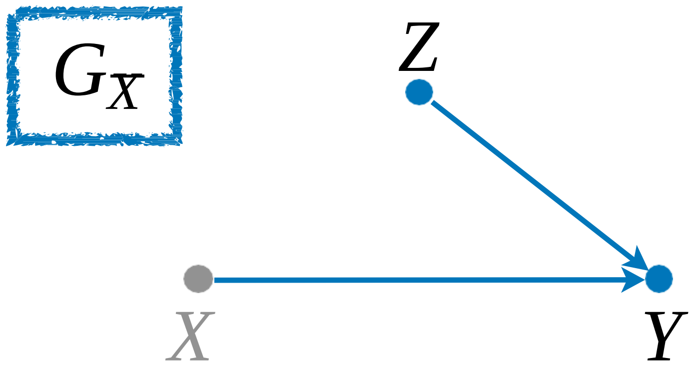
We conclude this section by giving several simple graphical definitions that will be needed later. For a given variable set $C \subseteq N$, let G_C denote the subgraph of G composed only of variables in C and all the bidirected links between variable pairs in C. We define $An(C)$ be the union of C and the set of observable ancestors of the variables in C in graph G and $De(C)$ be the union of C and the set of observable descendents of the variables in C in graph G.

An observable variable set $S \subseteq N$ in graph G is called an *ancestral set* if it contains all its own observed ancestors (i.e., $S = An(S)$).

Causal Effect in terms of C-Factors

- Consider an intervention $do(x)$





$$P(\mathbf{v}) = Q[Z]Q[X]Q[Y]$$

- We can get both *Q[Z]* and *Q[Y]* from *Q[**V**]* using c-component decomposition with *G* and *P(**v**)*.

$$Q[Z] = \frac{Q[Z]}{Q[\varnothing]} = \frac{\sum_{y,x} Q[Z, X, Y]}{\sum_{z,y,x} Q[Z, X, Y]} = P(z)$$

$$Q[Y] = \frac{Q[Z, X, Y]}{Q[Z, X]} = P(y \mid z, x)$$

$$P(y \mid do(x)) = \sum_{z} P(y, z \mid do(x))$$

$$= \sum_{z} Q[Y, Z]$$

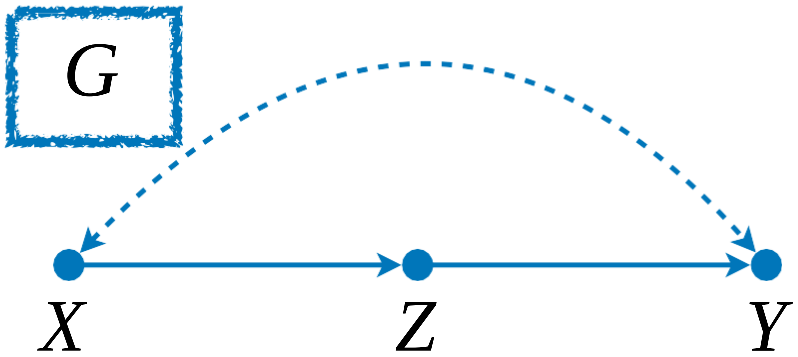
$$= \sum_{z} Q[Y]Q[Z]$$

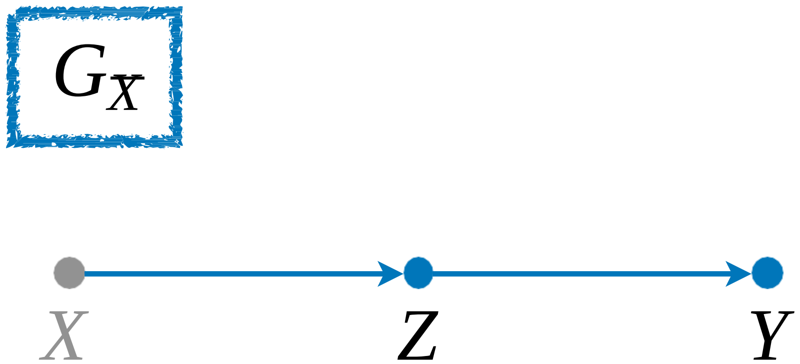
Back-door!

$$P(y \mid do(x)) = \sum_{z} P(y \mid z, x)P(z)$$

Causal Effect in terms of C-Factors

- Consider an intervention





$$P(\mathbf{v}) = Q[X,Y]\,Q[Z]$$

$$Q[Z] = \frac{Q[X, Z]}{Q[X]} = P(z \mid x)$$

$$P(y \mid do(x)) = \sum_z P(y, z \mid do(x))$$

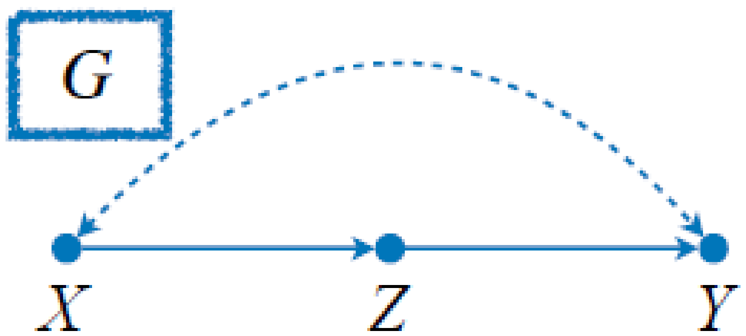
$$= \sum_z Q[Y, Z]$$

$$= \sum_z Q[Y]\,Q[Z]$$

- *Q[Z]* is the same in both
- Can we get *Q[Y]* from *Q[X,Y]* ?

Causal Effect in terms of C-Factors

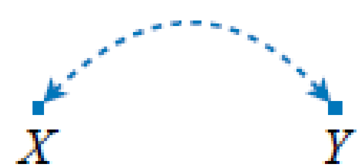
- Consider an intervention



$$Q[X,Y] = \frac{Q[X]}{Q[\varnothing]}\,\frac{Q[X,Z,Y]}{Q[X,Z]}$$

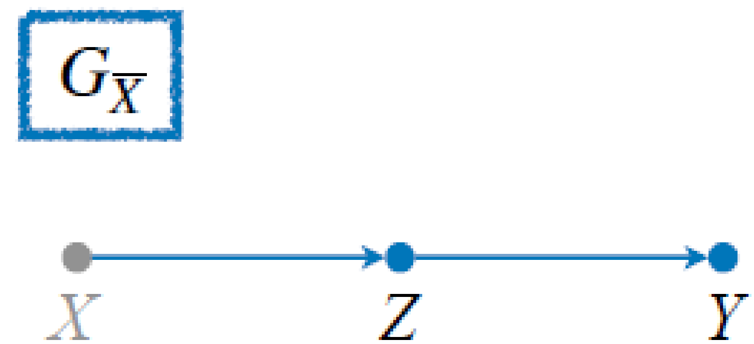
$$= P(x)P(y \mid x, z)$$

- $\{Y\}$ is ancestral in $G_{\{X,Y\}}$



$$Q[Y] = \sum_x Q[X,Y]$$

$$= \sum_x P(x)P(y \mid x, z)$$



$$P(y \mid do(x)) = \sum_z Q[Y]Q[Z]$$

$$= \sum_z \left(\sum_{x'} P(x')P(y \mid x', z) \right) P(z \mid x)$$

$$= \sum_z P(z \mid x) \sum_{x'} P(x')P(y \mid x', z)$$

Front-door!

A General Identification Algorithm

- Given G and the query variables $\boldsymbol{X}, \boldsymbol{Y}$

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{v} \setminus (\mathbf{x} \cup \mathbf{y})} Q[\mathbf{V} \setminus \mathbf{X}]$$

$$= \sum_{\mathbf{d} \setminus \mathbf{y}} Q[\mathbf{D}] \quad \text{where } \boldsymbol{D} = An(\boldsymbol{Y}) \text{ in } G_{\overline{\mathbf{X}}}$$

- Suppose the graph $G_{\boldsymbol{D}}$ has C-components $\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_k$, then

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{d} \setminus \mathbf{y}} \prod_i Q[D_i]$$

ID Algorithm

To compute $P_x(S) = P(S|do(X))$

INPUT: a set $S \subset V$.
OUTPUT: the expression for $P_x(s)$ or fail to determine.
Phase-1:

1. Find the c-components of G: $S^X, S_1, \ldots, S_k$, where $X \in S^X$.
2. Compute the c-factors $Q[S^X], Q[S_1], \ldots, Q[S_k]$
3. Let $D = An(S)_{G_{V\setminus\{X\}}}$, $D^X = D \cap S^X$.
4. Let the c-components of G_{D^X} be $D_j^X, j = 1, \ldots, l$.

Phase-2:
For each set D_j^X:
Compute $Q[D_j^X]$ *from* $Q[S^X]$ *by calling the function* $Identify(D_j^X, S^X, Q[S^X])$
If the function returns FAIL, then stop and output FAIL.

Phase-3:
Output $P_x(s) = \sum_{D\setminus S} \prod_j Q[D_j^X] \prod_i \sum_{S_i \setminus D} Q[S_i]$.

Completeness

Theorem [Huang and Valtorta, 2008]

The causal effect $P(\mathbf{y}|do(\mathbf{x}))$ is identifiable from causal diagram G and $P(\mathbf{v})$ *if and only if* each of the C-factors $Q[\mathbf{D_i}$ is identifiable by `Identify(Di,Ci,Q[Ci],G)`.

Where $\mathbf{C}_i$ is the C-component of G containing $\mathbf{D}_i$.

Examples of Estimand Expressions

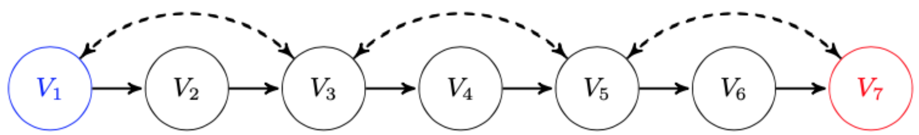
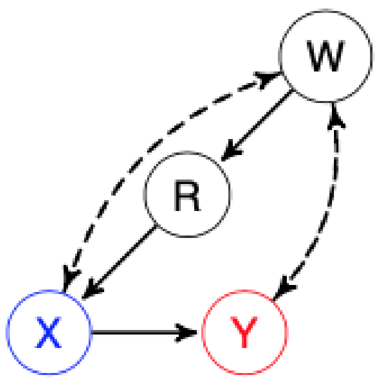
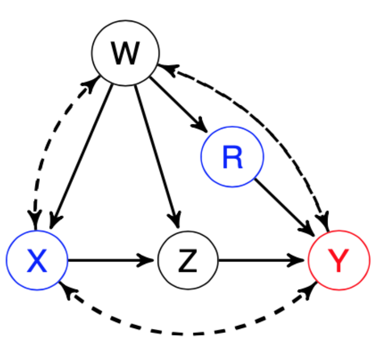


Figure 1: Chain Model with 7 observable variables and 3 latent variables

$$P(V_7 \mid do(V_1)) = \sum_{V_2,V_3,V_4,V_5,V_6} P(V_6 \mid V_1, V_2, V_3, V_4, V_5) P(V_4 \mid V_1, V_2, V_3) P(V_2 \mid V_1) \times \sum_{V_1'} P(V_7 \mid V_1', V_2, V_3, V_4, V_5, V_6) P(V_5 \mid V_1', V_2, V_3, V_4) P(V_3 \mid V_1', V_2) P(V_1')$$



(a) Model 1

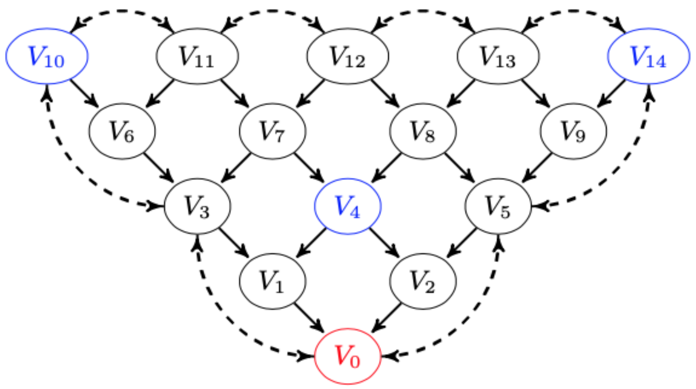


(c) Model 8

Estimand Expressions for Models 1 & 8 .

Model	Estimate of $P(Y \mid do(X))$
1	$\frac{\sum_W P(X,Y \mid R,W) P(W)}{\sum_W P(X \mid R,W) P(W)}$
8	$\sum_{R,W,Z} P(Z \mid R,W,X) P(R \mid W) \sum_x P(Y \mid R,W,x,Z) P(x \mid R,W) P(W)$

Examples of Estimand Expressions



(b) Cone Cloud, $n = 15$ (15-CC)

$$
\begin{aligned}
P(V_0|V_{14}, V_{10}, V_4) = \\
\sum_{V_1,V_2,V_3,V_5,V_6,V_7,V_8,V_9,V_{11},V_{12},V_{13},v'_{10},V'_{11},V'_{12},V'_{13},V'_{14}} & P(V_2|V_4, V_5, V_7, V_8, V_9, V_{11}, V_{12}, V_{13}, V_{14})\times \\
& P(V_9|V_{13}, V_{14})P(V_8|V_{12}, V_{13})P(V_1|V_3, V_4, V_6, V_7, V_8, V_{10}, V_{11}, V_{12}, V_{13})\times \\
& P(V_7|V_{11}, V_{12})P(V_6|V_{10}, V_{11})P(V_{11}, V_{12}, V_{13})\times \\
& P(V_0|V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8, V_9, V'_{10}, V_{11}, V_{12}, V_{13}, V'_{14})\times \\
& P(V_5|V_1, V_3, V_4, V_6, V_7, V_8, V_9, V'_{10}, V_{11}, V_{12}, V_{13}, V'_{14})\times \\
& P(V'_{14}|V_1, V_3, V_4, V_6, V_7, V_8, V'_{10}, V_{11}, V_{12}, V_{13})\times \\
& P(V_3, V_{13}|V_6, V_7, V'_{10}, V_{12}, V_{13})P(V'_{10}|V_7, V_{11}, V_{12})P(V_{11}, V_{12}) \qquad (7)
\end{aligned}
$$

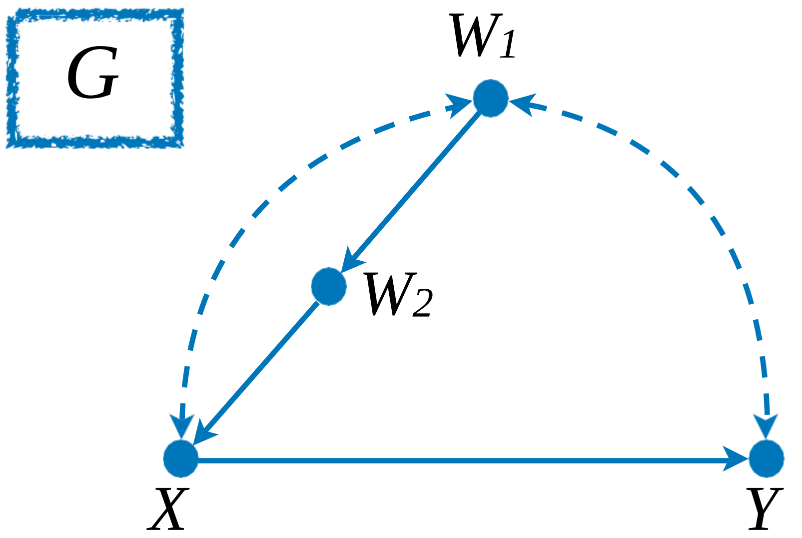
An estimand often corresponds to inference over a Bayesian network
Which is sometime very dense.

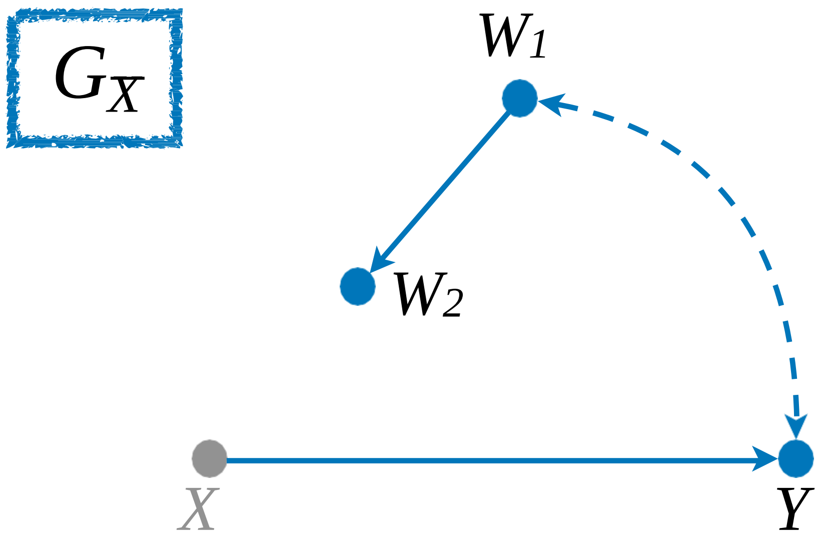
The treewidth of the above example is sqrt of n, when n is the number of variables

So, is evalution Exp(w)?.

Solving the Napkin

- Recall the Napkin graph from last time





$$P(\mathbf{v}) = Q[W_1, X, Y]Q[W_2]$$

$$P(y \mid do(x)) = \sum_{w_1, w_2} Q[W_1, W_2, Y]$$

- $Q[W_1,X,Y]$ is computable from $Q[\mathbf{V}]$
- Can we get $Q[Y]$ from $Q[W_1,X,Y]$?

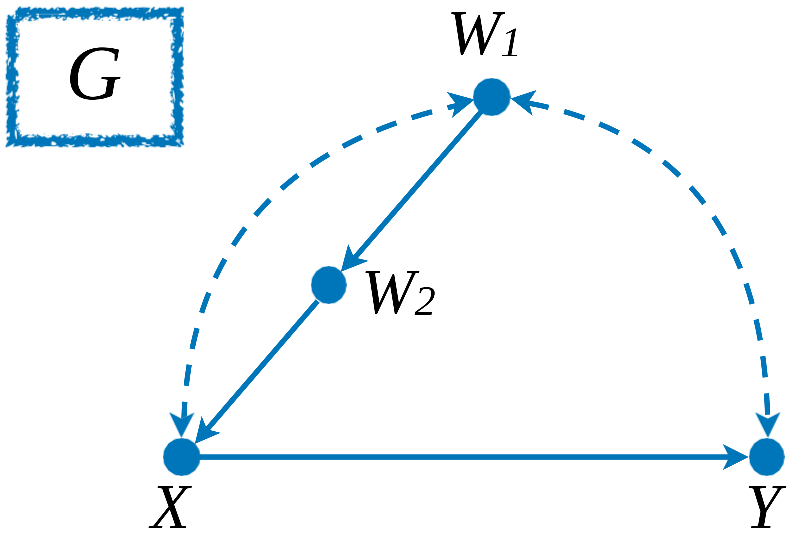
- $\{Y\}$ is ancestral in $G_{\{W_1,W_2,Y\}}$, so

$$P(y \mid do(x)) = Q[Y]$$

Solving the Napkin

- We can compute $Q[W_1,X,Y]$ from $Q[W_1,W_2,X,Y]$

$$P(y \mid do(x)) = Q[Y]$$



$$Q[W_1,X,Y] = \frac{Q[W_1]}{Q[\varnothing]} \frac{Q[W_1,W_2,X]}{Q[W_1,W_2]} \frac{Q[W_1,W_2,X,Y]}{Q[W_1,W_2,X]}$$

$$= \frac{P(w_1)}{1} \frac{P(w_1,w_2,x)}{P(w_1,w_2)} \frac{P(w_1,w_2,x,y)}{P(w_1,w_2,x)}$$

$$= P(w_1)P(x \mid w_1,w_2)P(y \mid w_1,w_2,x)$$

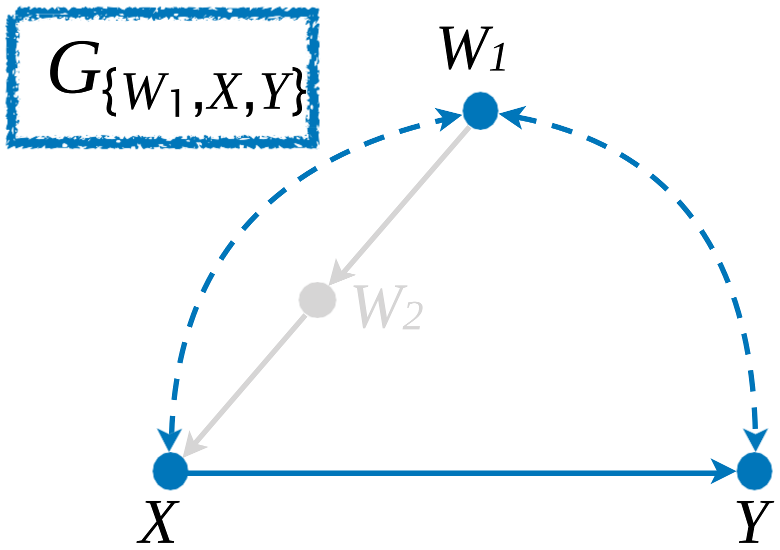
Solving the Napkin

$$P(y \mid do(x)) = Q[Y]$$

- $\{X,Y\}$ is ancestral in $G_{\{W_1,X,Y\}}$

$$Q[X,Y] = \sum_{w_1} Q[W_1,X,Y]$$

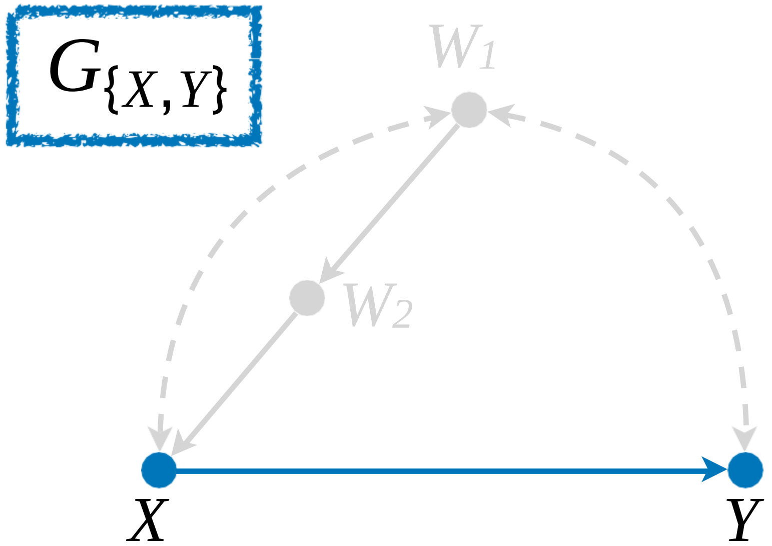
$$= \sum_{w_1} P(w_1)P(x \mid w_1,w_2)P(y \mid w_1,w_2,x)$$



Solving the Napkin

$$P(y \mid do(x)) = Q[Y]$$

- $G_{\{X,Y\}}$ has two c-components, hence



$$Q[X,Y] = Q[X]Q[Y]$$

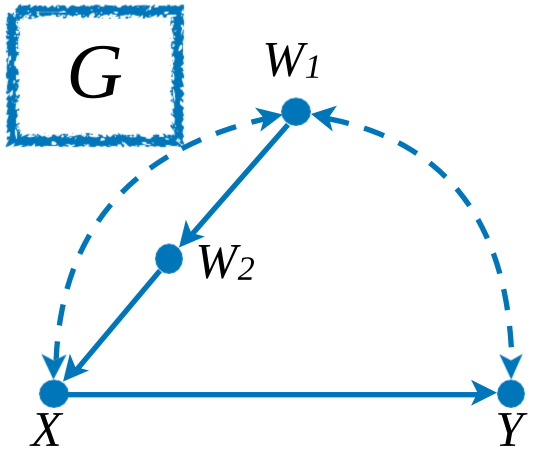
$$Q[Y] = \frac{Q[X,Y]}{Q[X]} = \frac{Q[X,Y]}{\sum_y Q[X,Y]}$$

$$= \frac{\sum_{w_1} P(w_1)P(x \mid w_1, w_2)P(y \mid w_1, w_2, x)}{\sum_{y,w_1} P(w_1)P(x \mid w_1, w_2)P(y \mid w_1, w_2, x)}$$

$$= \frac{\sum_{w_1} P(w_1)P(x \mid w_1, w_2)P(y \mid w_1, w_2, x)}{\sum_{w_1} P(w_1)P(x \mid w_1, w_2)} = P(y \mid do(x))$$

Solve the Napkin using Do-Calculus

- Let's see an equivalent do-calculus derivation



$$P(y \mid do(x)) = P(y \mid do(x, w_2, w_1))$$ Rule 3: $(Y \perp\!\!\!\perp W_2, W_1 \mid X)_{G_{X\overline{W_1, W_2}}}$

$$= P(y \mid x, do(w_2, w_1))$$ Rule 2: $(Y \perp\!\!\!\perp X \mid W_1, W_2)_{G_{\underline{X}\overline{W_1, W_2}}}$

$$= \frac{P(y, x \mid do(w_2, w_1))}{P(x \mid do(w_2, w_1))}$$ Conditional probability

$$= \frac{P(y, x \mid do(w_2))}{P(x \mid do(w_2))}$$ Rule 3: $(Y, X \perp\!\!\!\perp W_1 \mid W_2)_{G_{\overline{W_1}\overline{W_2}}}$

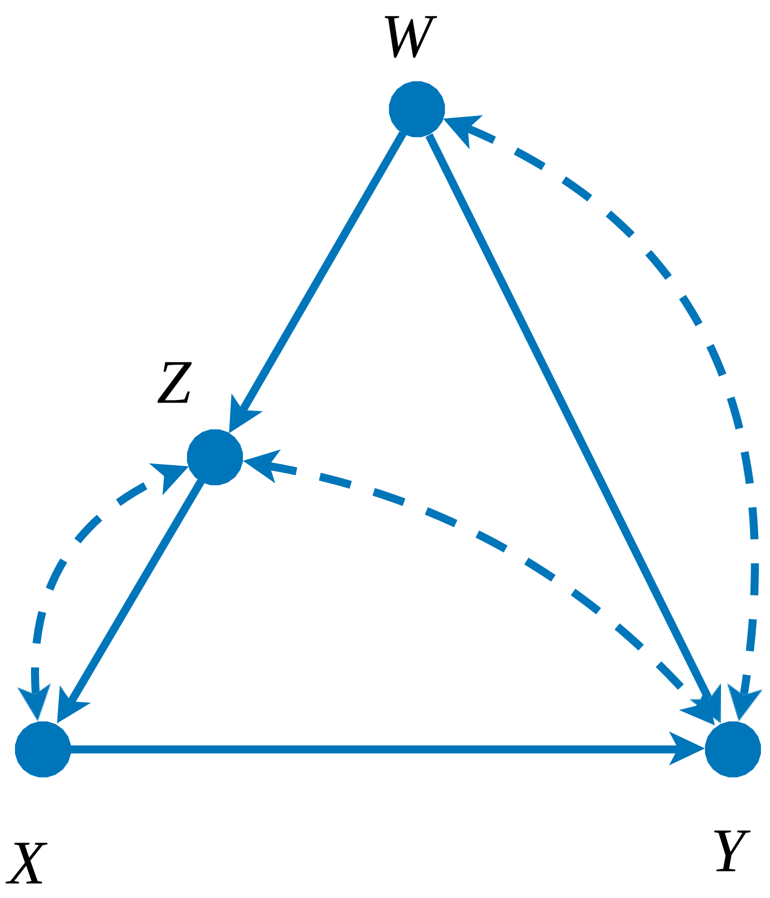
$$= \frac{\sum_{w_1} P(y, x \mid do(w_2), w_1) P(w_1 \mid do(w_2))}{\sum_{w_1} P(x \mid do(w_2), w_1) P(w_1 \mid do(w_2))}$$ Condition on W_1

$$= \frac{\sum_{w_1} P(y, x \mid w_2, w_1) P(w_1)}{\sum_{w_1} P(x \mid w_2, w_1) P(w_1)}$$ Rule 2: $(Y, X \perp\!\!\!\perp W_2 \mid W_1)_{G_{\underline{W_2}}}$

Rule 3: $(W_1 \perp\!\!\!\perp W_2)_{G_{\overline{W_2}}}$

Food for Thought

- Use the strategy discussed in this lecture to identify the effect $P(y|do(x))$ in the following causal diagram



References

- Tian, J. (2002). Studies in Causal Reasoning and Learning. Computer Science Department, University of California, Los Angeles, CA.
- Tian, J., & Pearl, J. (2002). A General Identification Condition for Causal Effects. In Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002) (pp. 567–573). Menlo Park, CA: AAAI Press/The MIT Press
- Huang, Y., & Valtorta, M. (2008). On the completeness of an identifiability algorithm for semi-Markovian models. Annals of Mathematics and Artificial Intelligence, 54(4), 363–408.