
CS 295: Causal Reasoning

Rina Dechter

Algorithmic Approach for Identification¹

Based on Elias Bareinboim slides

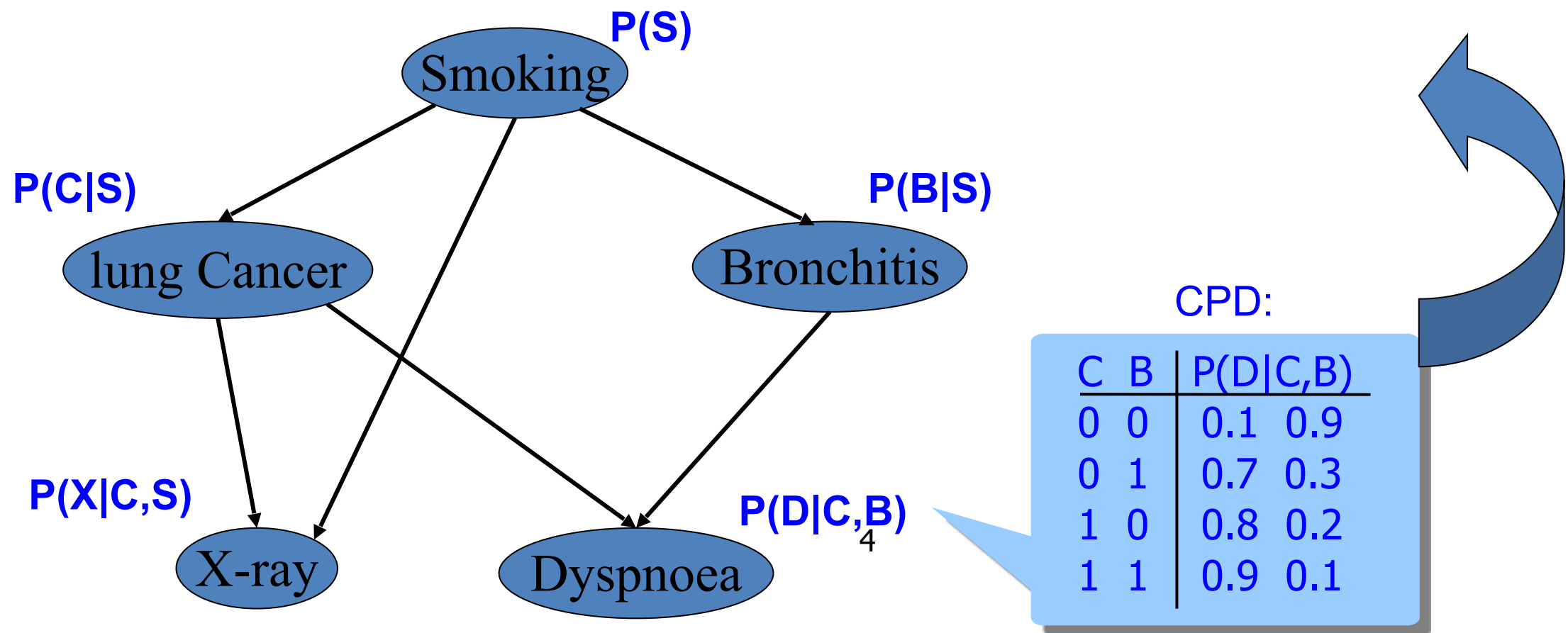
Roadmap

- 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)



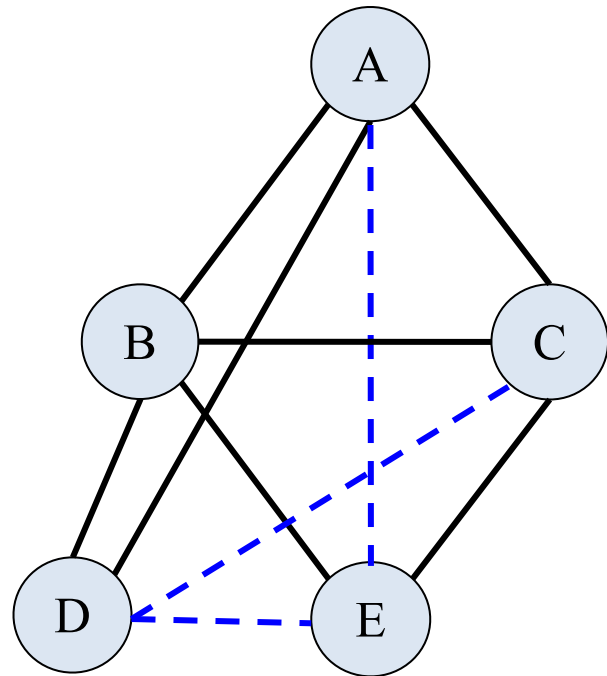
$$P(S, C, B, X, D) = P(S) P(C|S) P(B|S) P(X|C,S) P(D|C,B)$$

Belief Updating:

$P(\text{lung cancer=yes} \mid \text{smoking=no, dyspnoea=yes}) = ?$

Belief Updating

$$p(X \mid \text{Evidence}) = ?$$



“primal” graph

$$p(A \mid E = 0)$$

$$\propto p(A, E = 0)$$

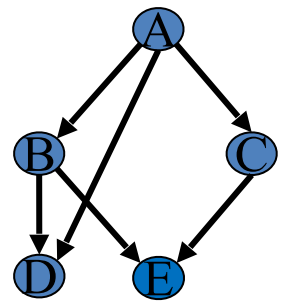
$$= \sum_{e,d,c,b} p(A) \underbrace{p(b|A)} \underbrace{p(c|A) p(d|b, A) p(e|b, c)} \mathbb{1}[e = 0]$$

$$p(A) \sum_e \sum_d \sum_c p(c|A) \mathbb{1}[e = 0] \sum_b \underbrace{p(b|A) p(d|b, A) p(e|b, c)}_{\lambda_{B \rightarrow C}(a, d, c, e)}$$

Variable Elimination

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:

$$p(c|A) \lambda_{B \rightarrow C}(A, d, c, e)$$

bucket D:

$$\lambda_{C \rightarrow D}(A, d, e)$$

bucket E:

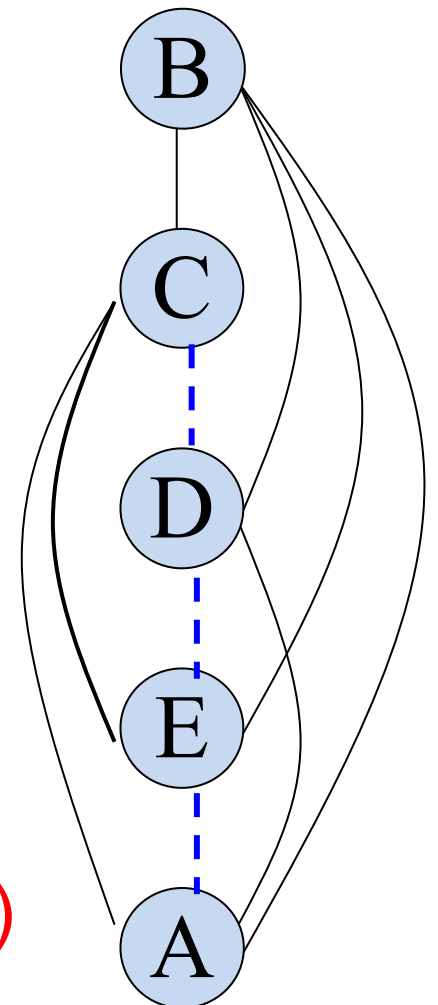
$$\mathbb{1}[E=0] \lambda_{D \rightarrow E}(A, e)$$

bucket A:

$$p(A) \lambda_{E \rightarrow A}(A)$$

$$p(E=0)$$

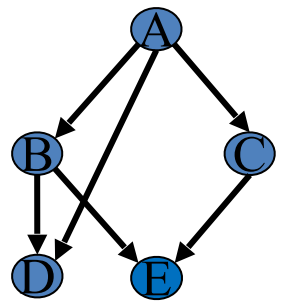
$W^*=4$
“induced width”
(max clique size)



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

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

7

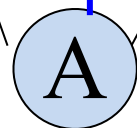
bucket A:

$p(A)$

$\lambda_{E \rightarrow A}(A)$

(max clique size)

$p(E=0)$

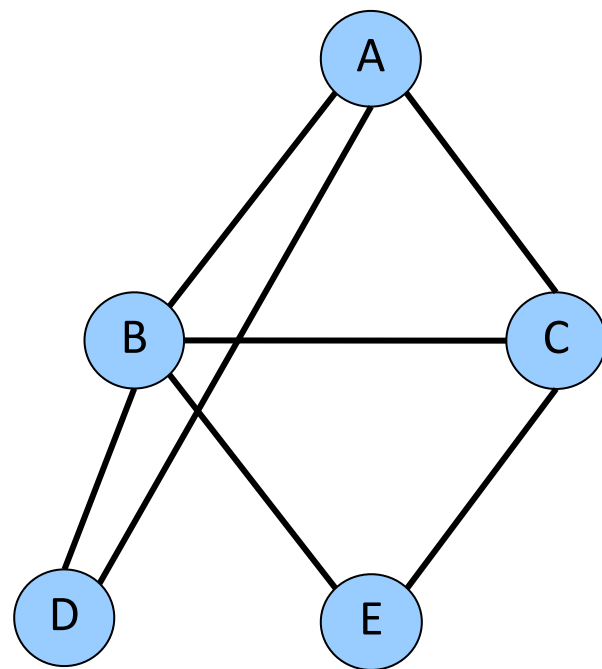


$$p(A|E=0) = p(A, E=0) / p(E=0)$$

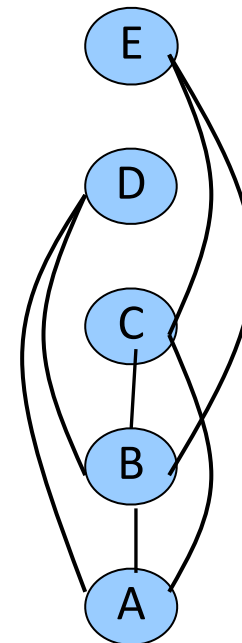
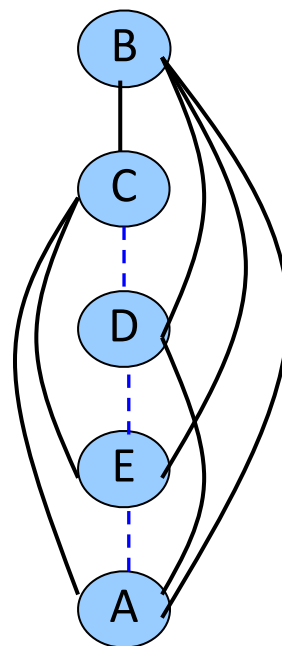
295, Winter 2023

Induced Width (continued)

The effect of the ordering:

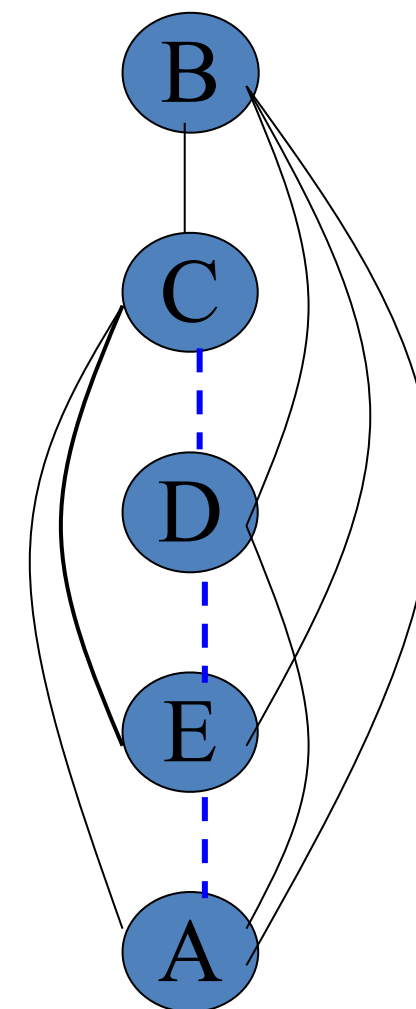
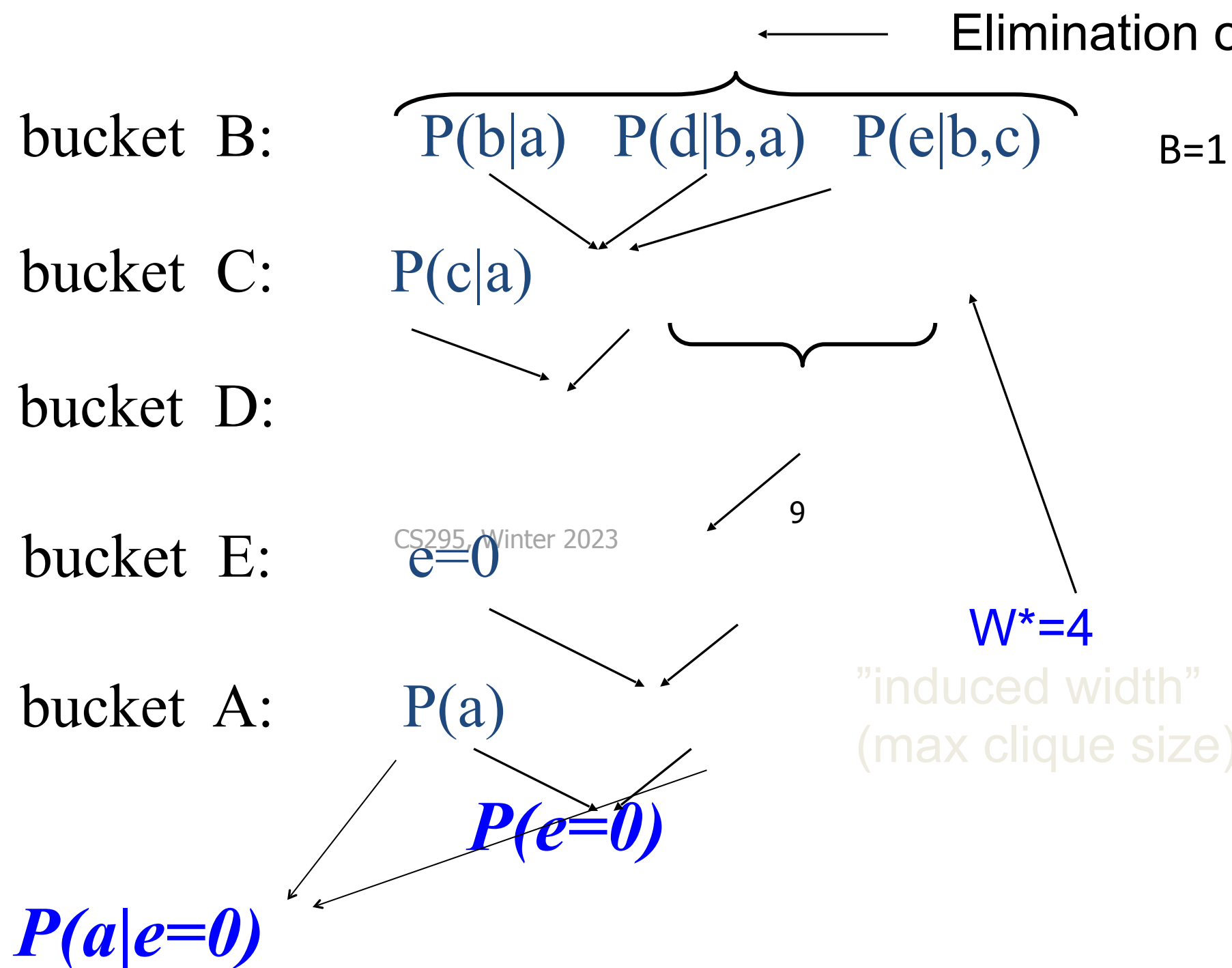
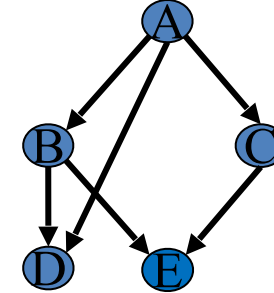


Primal (moraal) graph



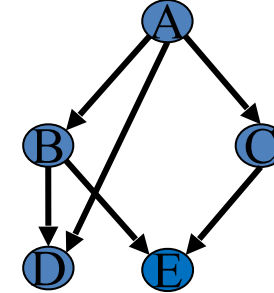
The impact of evidence?

Algorithm *BE-bel*

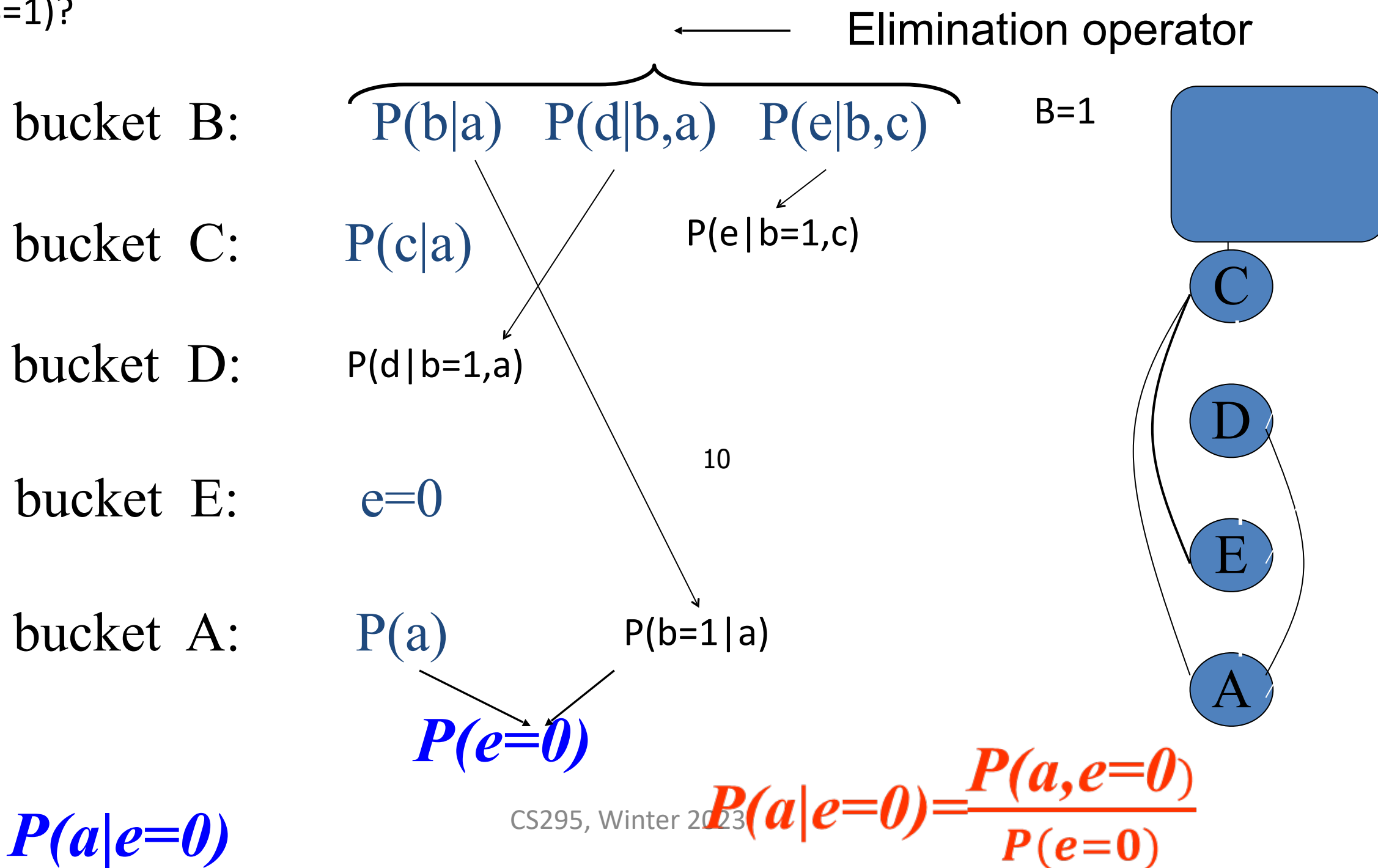


The impact of evidence?

Algorithm *BE-bel*



$P(A|E=0, B=1)$?



Back to SCM

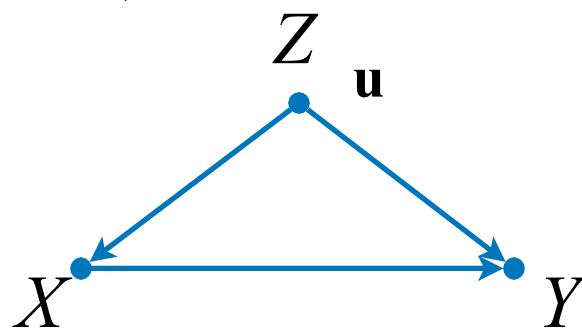
11

Markovian Case

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

$$P(\mathbf{v}) = \sum P(\mathbf{u}) \prod P(v_i \mid v_1, \dots, v_{i-1}, \mathbf{u}) = \sum P(\mathbf{u}) \prod P(v_i \mid pa_i,$$

$u_i)$



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

$u_y)$

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

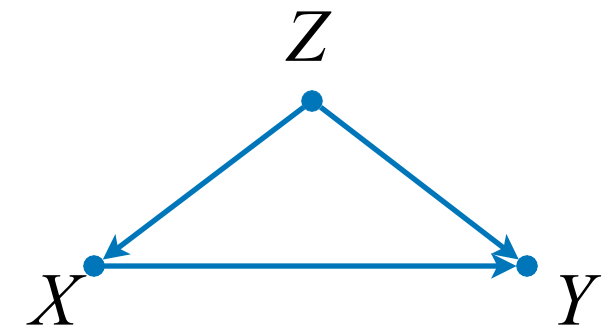
$$= P(z) P(x \mid z) P(y \mid x, z)$$

- 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.,

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



$$P(z, x, y) = P(z)P(x | z)P(y | x, z)$$

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

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

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

Markovian Case



$$U_Z: P(U_Z), P(Z|U_Z) \rightarrow$$

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

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

$$X, Y, Z: \begin{cases} \lambda_{U_Z}(Z) = \sum_{U_Z} P(Z|U_Z) \cdot P(U_Z) = ? \\ \lambda_{U_X}(X, Z) = \sum_{U_X} P(X|U_X, Z) \cdot P(U_X) = ? \\ \lambda_{U_Y}(Y, X, Z) = \sum_{U_Y} P(Y|Z, X, U_Y) \cdot P(U_Y) = ? \end{cases}$$

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

Markovian Case



$$U_Z: P(U_Z), P(Z|U_Z) \rightarrow$$

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

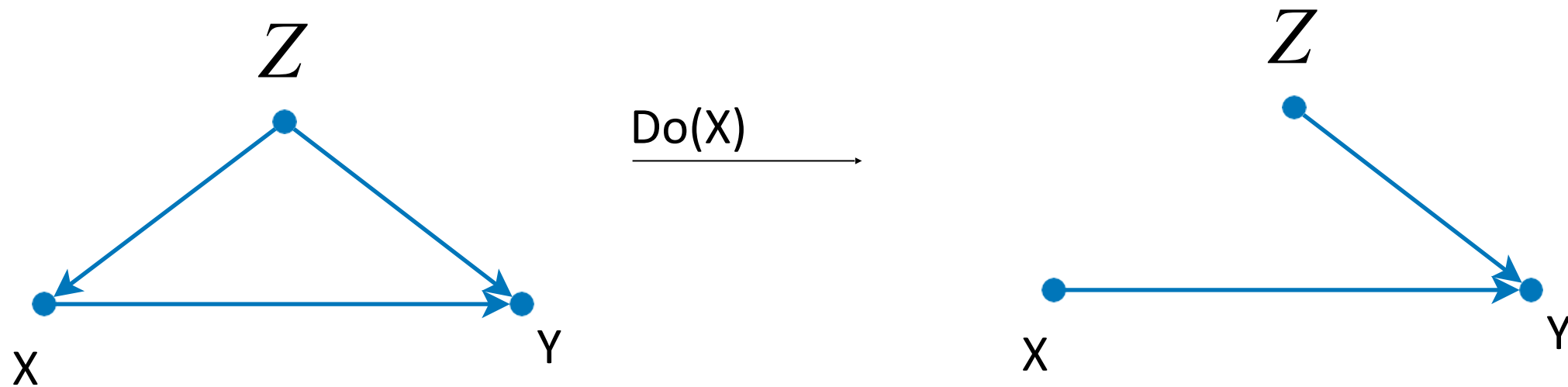
$$U_Y: P(U_Y), P(Y|Z, X, U_Y)$$

$$X, Y, Z: \begin{cases} \lambda_{U_Z}(Z) = \sum_{U_Z} P(Z|U_Z) \cdot P(U_Z) = ? \\ \lambda_{U_X}(X, Z) = \sum_{U_X} P(X|U_X, Z) \cdot P(U_X) = ? \\ \lambda_{U_Y}(Y, X, Z) = \sum_{U_Y} P(Y|Z, X, U_Y) \cdot P(U_Y) = ? \end{cases}$$

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

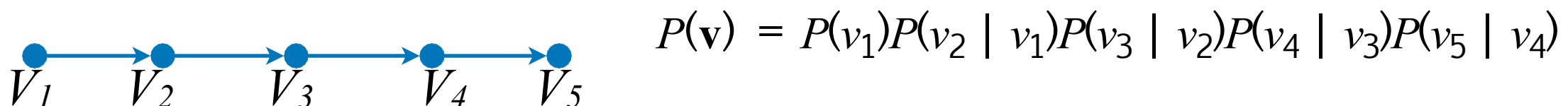
$$P(X, Y, Z) = P(Z)P(X|Z)P(Y|X, Z)$$

Markovian Case

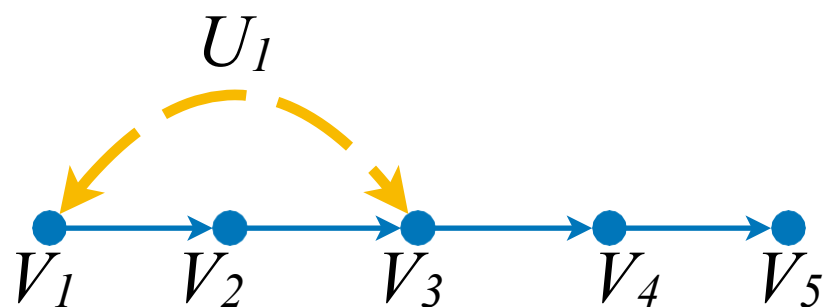


Semi-Markovian Case

- Start from a simple Markovian model:



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



$$\begin{aligned}
 P(\mathbf{v}) &= \sum_{u_1} P(u_1)P(v_1 | u_1)P(v_2 | v_1)P(v_3 | v_2, u_1)P(v_4 | v_3)P(v_5 | v_4) \\
 &= P(v_2 | v_1)P(v_4 | v_3)P(v_5 | v_4) \left(\sum_{u_1} P(u_1)P(v_1 | u_1)P(v_3 | v_2, u_1) \right)
 \end{aligned}$$

Using Bucket Elimination



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

$$V_3: \underbrace{P(V_4|V_3)}_{\lambda(V_1, V_3, V_2)} =$$

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

$$V_1:$$

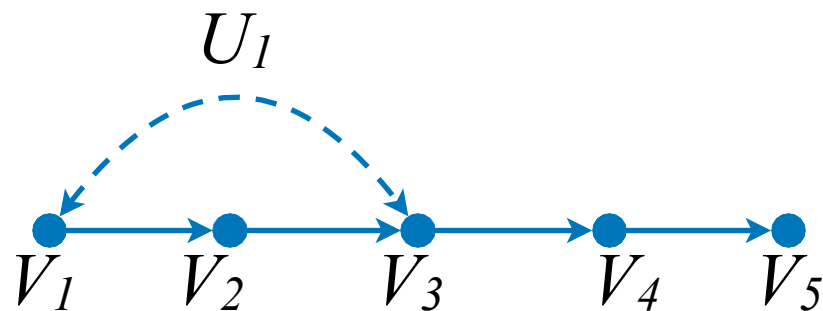
$$V_4: P(V_5|V_4)$$

$$V_5:$$

Can this be
expressed using
 $P(V)$ only?

Semi-Markovian Case

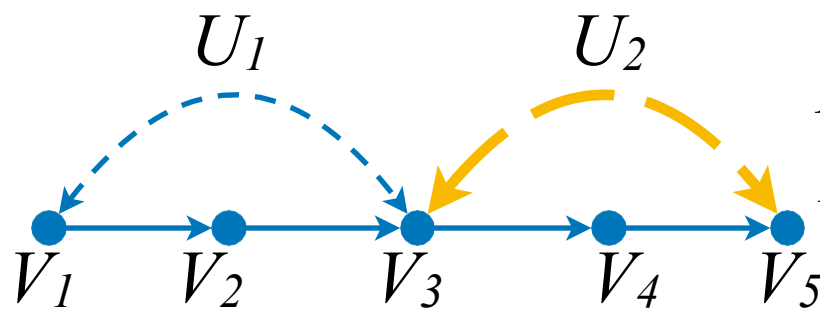
- From the previous model ...



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

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

- Add another unobservable U_2 ,

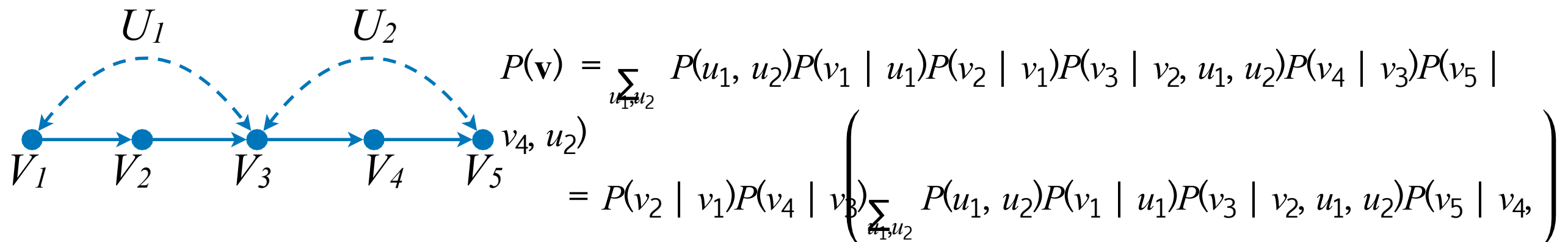


$$P(\mathbf{v}) = \sum_{u_1, u_2} P(u_1, u_2) P(v_1 | u_1) P(v_2 | v_1) P(v_3 | v_2, u_1, u_2) P(v_4 | v_3) P(v_5 | v_4, u_2)$$

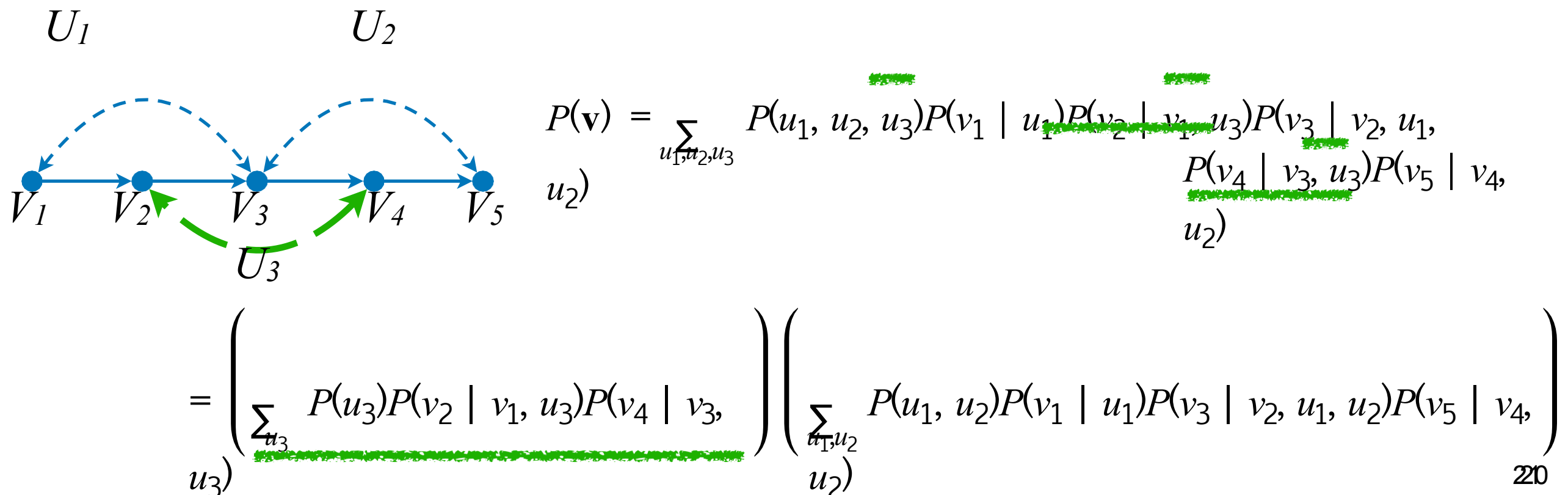
$$= P(v_2 | v_1) P(v_4 | v_3) \sum_{u_1, u_2} P(u_1, u_2) P(v_1 | u_1) P(v_3 | v_2, u_1, u_2) P(v_5 | v_4, u_2)$$

Semi-Markovian Case

- From the previous model...

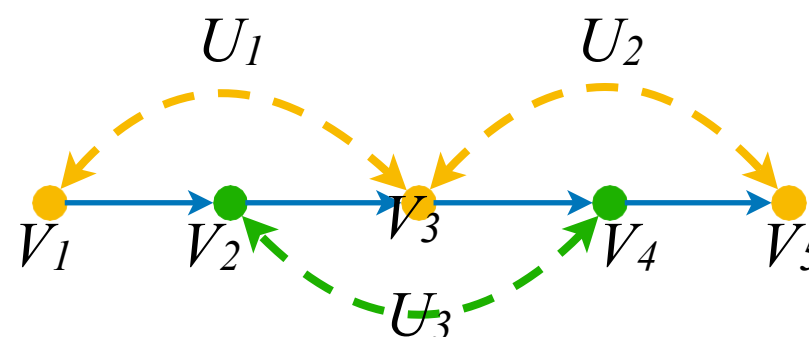


- Let's add one more, U_3 ,



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 :

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

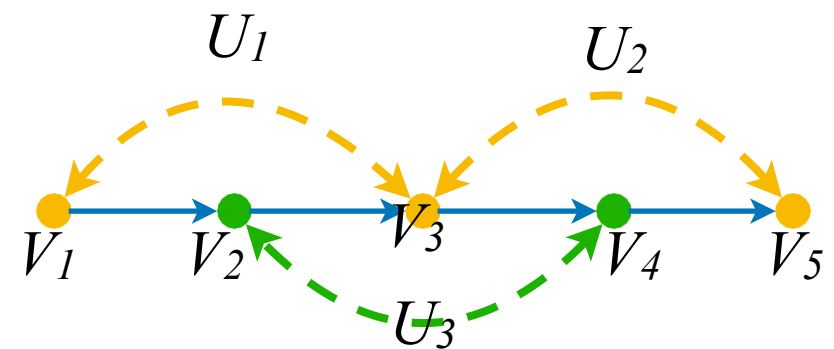
- Then $P(\mathbf{v})$ can be re-written as

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

C-factors

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

$$P(\mathbf{v}) = \underbrace{Q[V_2, V_4]}_{\text{green}} \underbrace{Q[V_1, V_3, V_5]}_{\text{yellow}}$$



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

$$\frac{Q[\mathbf{V}]}{P(\mathbf{v})} = \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{V_i \in \mathbf{V}} P(v_i \mid pa_i, u_i) =$$
- For consistency define $Q[\emptyset]=1$

C-factors are Causal Effects

- Let $C \subseteq V$. Consider the causal effect of all other variables on C , that is $P(\mathbf{c} | do(\mathbf{v} \setminus \mathbf{c}))$.
- By the truncated product we have

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

- All U s that are not parents of any element in C can be summed out, hence

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



On the completeness of an identifiability algorithm for semi-Ma models

Article *in* Annals of Mathematics and Artificial Intelligence · December 2008

DOI: 10.1007/s10472-008-9101-x · Source: DBLP



The Q function (from Valtorta)

$$P_{n \setminus s}(s) = \sum_{U_k \in U} \prod_{V_i \in S} P(v_i | pa(V_i)) \prod_{V_j \in U} P(v_j | pa(V_j)) \quad (8)$$

In formula 8, the subscript $n \setminus s$ indicates a configuration of the variable or variables in the set $N \setminus S$. For convenience and for uniformity with [2], we define

$$Q[S] = P_{N \setminus S}(S) \quad (9)$$

In a semi-markovian model:-

$$P_t(s) = Q[s] = P_{n \setminus s}(s) = \sum_{U_k \in U} \prod_{V_i \in S} P(v_i | pa(V_i)) \prod_{V_j \in U} P(v_j) \quad (11)$$

We define the *c-component relation* on the unobserved variable set U of graph G as follows. For any $U_1 \in U$ and $U_2 \in U$, they are related under the c-component relation if and only if one of conditions below is satisfied:

- (i) there is an edge between U_1 and U_2 ,
- (ii) U_1 and U_2 are both parents of the same observable node,
- (iii) both U_1 and U_2 are in the c-component relation with respect to another node $U_3 \in U$.

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 descendants of the variables in C in graph G .

An observable variable set S is called an ancestral set if it contains all its own

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

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

- 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 them, for instance V_3 , we have

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

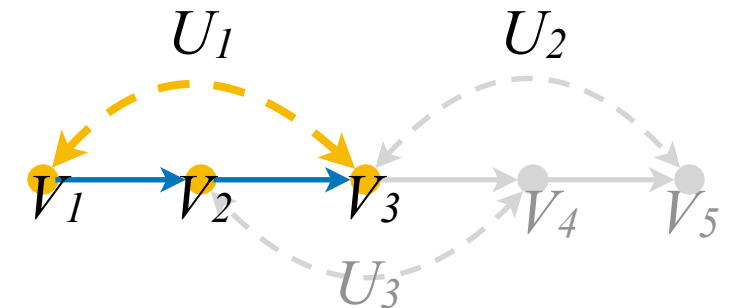
$$= \sum_{u_1, u_2} P(u_1, u_2) P(v_1 | u_1) P(v_5 | v_4, u_2) = Q[V_1, V_5]$$

Marginalizing Variables in C-factors

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

By Definition of Q:

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



- 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

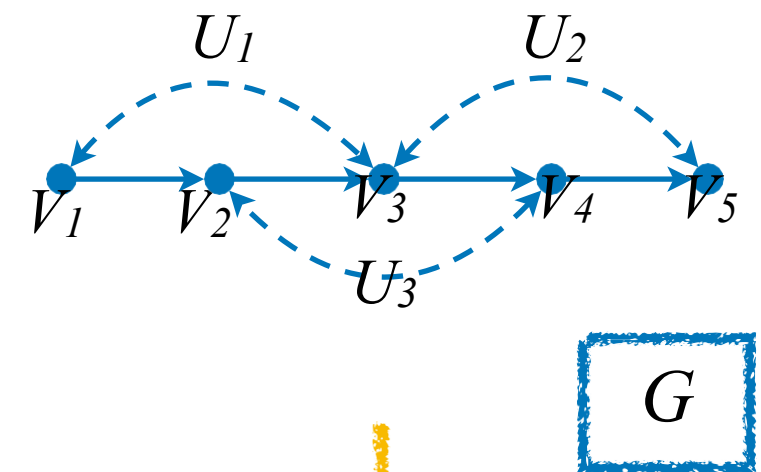
$$\sum_{v_1} Q[V_1, V_2, V_3] = \sum_{v_1} \sum_{u_1 u_2 u_3} P(u_1, u_2, u_3) P(v_1 | u_1) P(v_2 | v_1, u_3) P(v_3 | v_2, u_1, u_2)$$

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

Can we remove V1 here? Symbolically?

What variables can be marginalized in the logic of C-factors?

- Let $W \subset C \subseteq V$, be two sets of variables.
- Lemma (ancestral-reduction). If W is **ancestral**, that is, it contains all $An(W)$ present in the subgraph made of the variables in C , i.e., G_C .

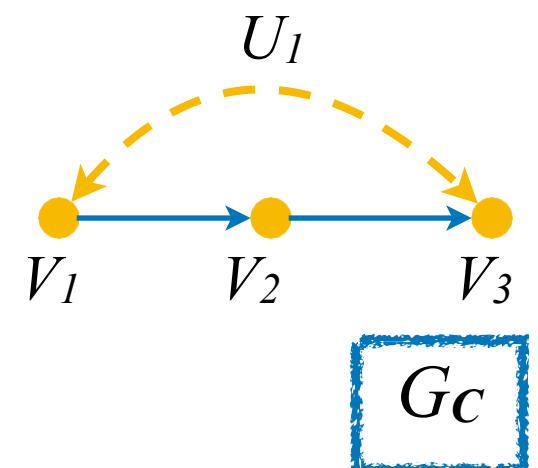


- Then, $Q[W] = \sum_C Q[C]$
- For example, for $C = \{V_1, V_2, V_3\}$, in G_C

- $W = \{V_1, V_2\}$ is ancestral $Q[V_1, V_2] = \sum_{v_3} Q[V_1, V_2, V_3]$

- $W = \{V_1\}$ is ancestral $Q[V_1] = \sum_{v_2, v_3} Q[V_1, V_2, V_3]$

- $W = \{V_2, V_3\}$ is not ancestral



Theorem 1 *Let $W \subseteq C \subseteq N$. If W is an ancestral set in G_C , then*

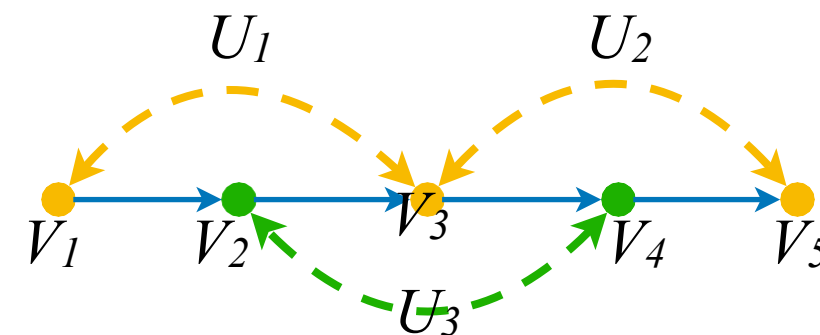
$$\sum_{V_i \in C \setminus W} Q[C] = Q[W] \quad (12)$$

We recall that subgraph G_C includes all variables in C and the subset of the unobservable variables in G for which their children are all in C . The lemma says that in such a subgraph, if W is a set of observable variables whose ancestor set includes no other observable variables in the subgraph, then $Q[W]$ can be calculated directly from $Q[C]$ by marginalizing variables in $C \setminus W$. In

So if $Q(C)$ is identifiable then $Q(W)$ is identifiable.

Confounded Components (C-Components)

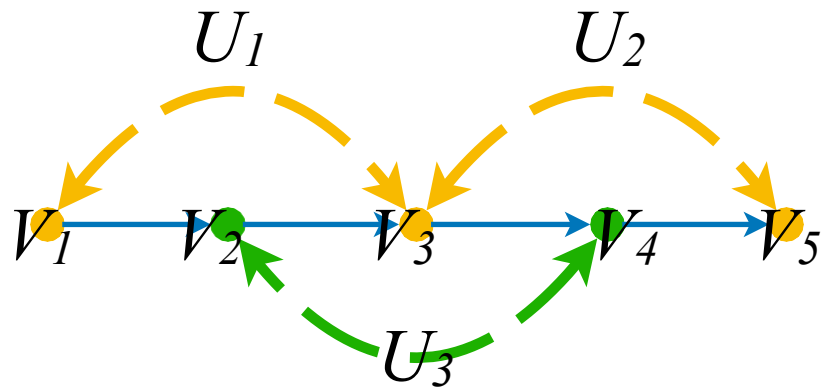
- Recall our example



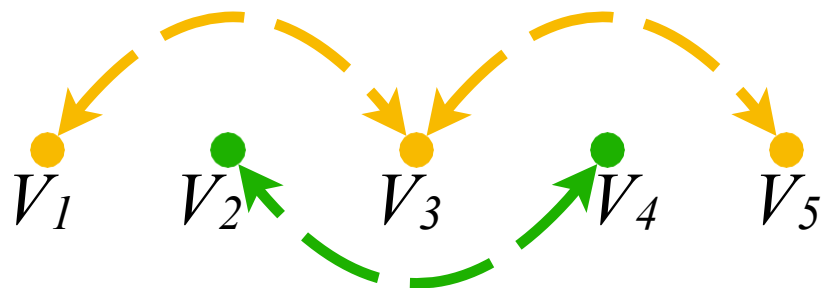
$$P(\mathbf{v}) = \left(\sum_{\substack{u_3 \\ \text{green}}} P(u_3) P(v_2 \mid v_1, u_3) P(v_4 \mid v_3, u_3) \right) \left(\sum_{\substack{u_1, u_2 \\ \text{yellow}}} P(u_1, u_2) P(v_1 \mid u_1) P(v_3 \mid v_2, u_1, u_2) P(v_5 \mid v_4, u_2) \right)$$

- 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).

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 .

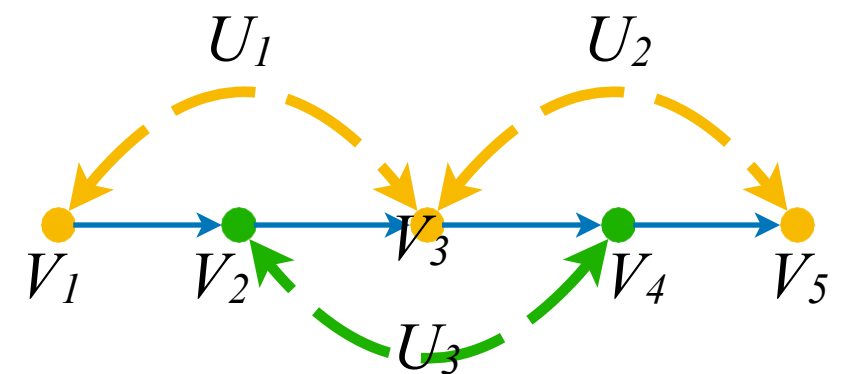


- To see it easily, consider the graph induced over the **bidirected edges**!
- Obs. The C-Component relation defines a partition over the observable variables, hence it is *Reflexive*, *Symmetric* and *Transitive*.

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 $H \subseteq V$, consider a graph G_H .
- Let $H_1, H_2, \dots, H_k$ be the c-components of G_H .
- Then

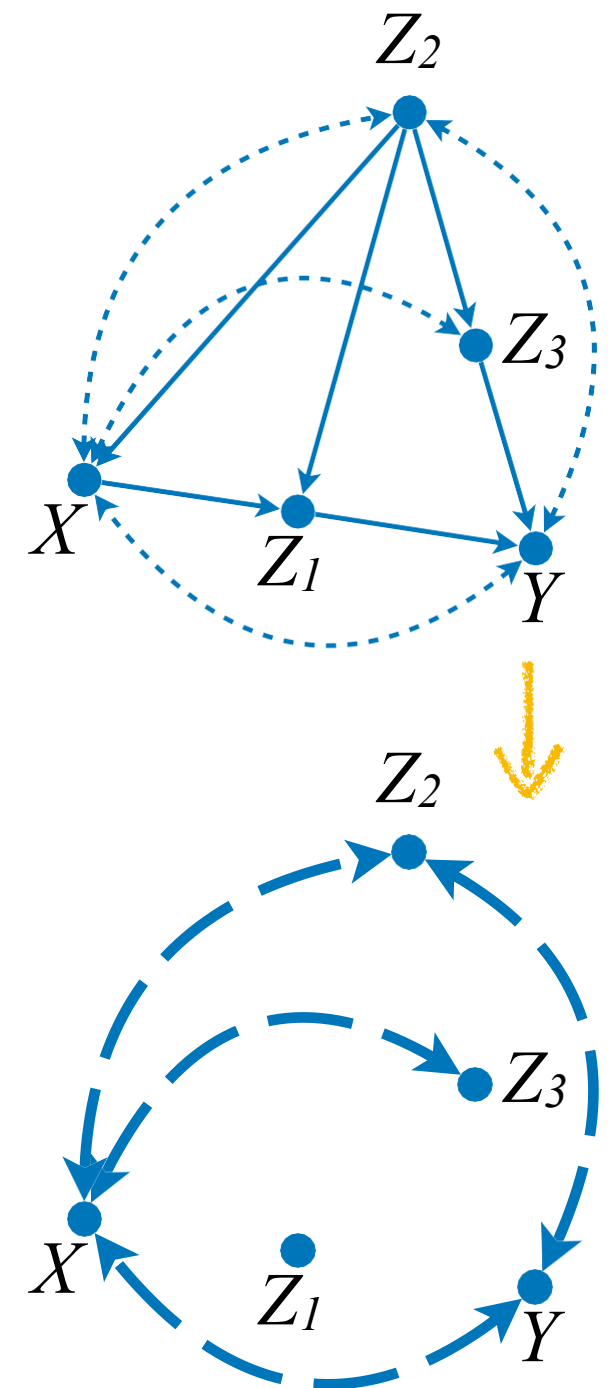
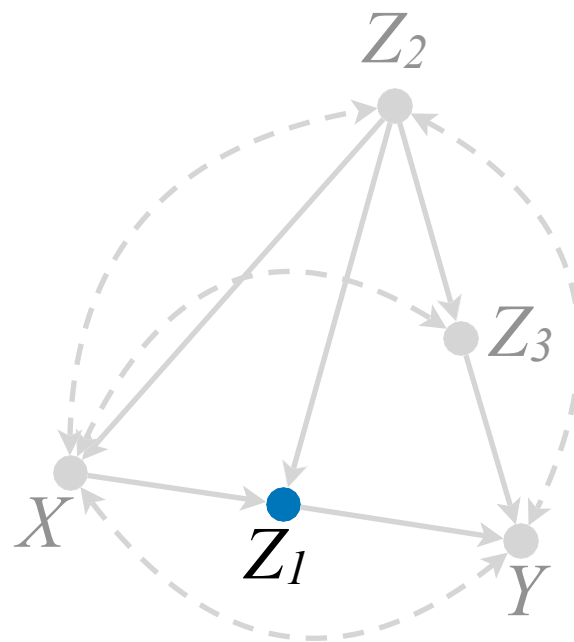
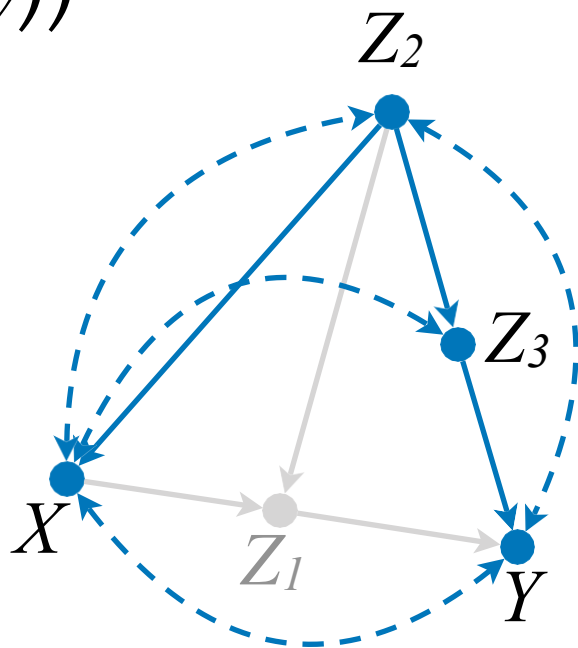
$$Q[\mathbf{H}] = \prod_j Q[H_j]$$

C-Component Factorization

- Consider another example

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

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



C-Component Factorization

- For any $H \subseteq V$, consider a graph G_H .
- Let $H_1, H_2, \dots, H_k$ be the c-components of G_H .
- Then

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

C-Component Factorization

(Continued)

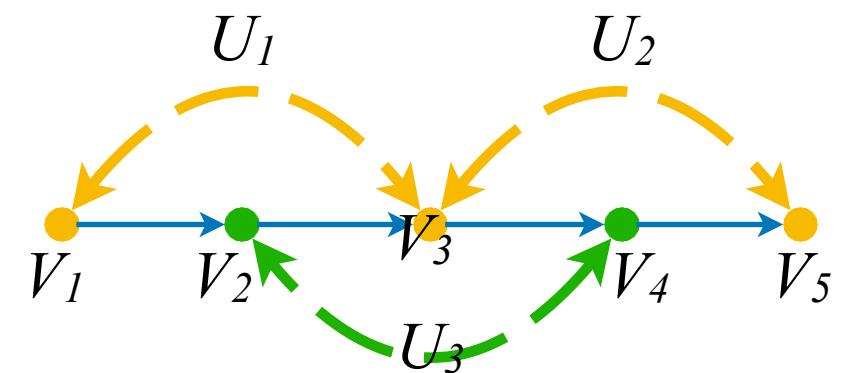
- Let $V_{h1} < V_{h2} < \dots < V_{hn}$ be a topological order over the variables in $\mathbf{H}$ according to G .
- Let $H^{\leq i}$ be the variables in $\mathbf{H}$ that come before V_{hi} , including V_{hi} .
- Let $H^{> i}$ be the variables in $\mathbf{H}$ that come after V_{hi} .
- Then

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

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

C-Component Factorization

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



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

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

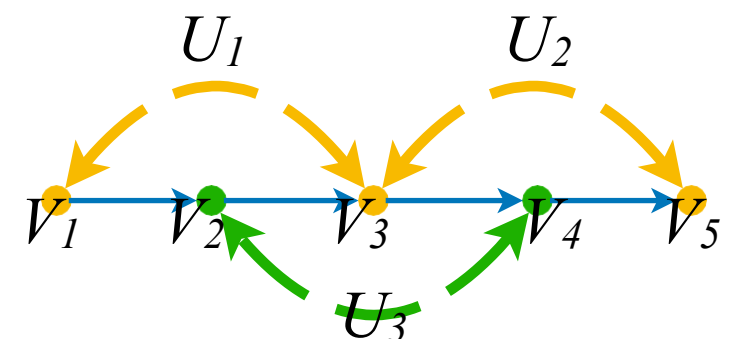
$$Q[V_2, V_4] = \frac{Q[V_5]}{Q[V_1, V_2]} \frac{Q[\emptyset]}{Q[V_1, V_2, V_3, V_4]} \frac{Q[V_1, V_2, V_3, V_4]}{Q[V_1, V_2, V_3]} \frac{Q[V_1, V_2]}{Q[V_1]}$$

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

C-Component Factorization

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

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

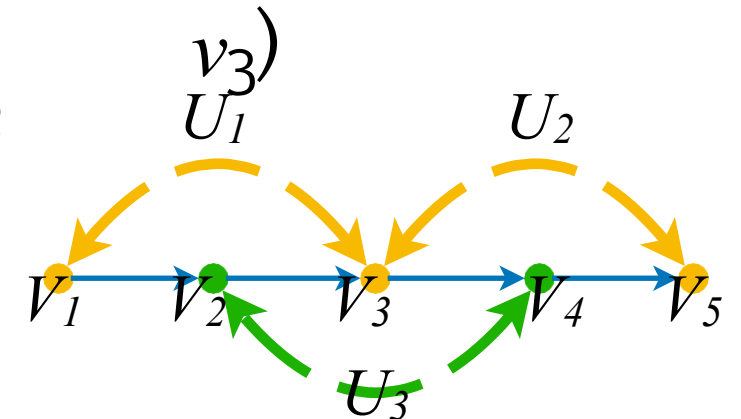


C-Component Factorization

$$\begin{aligned}
 Q[V_2, V_4] &= \frac{Q[V_1, V_2]}{Q[V_1]} \frac{Q[V_1, V_2, V_3, V_4]}{Q[V_1, V_2, V_3]} \\
 &= \frac{\sum_{v_3, v_4, v_5} Q[\mathbf{v}]}{\sum_{v_2, v_3, v_4, v_5} Q[\mathbf{v}]} \frac{\sum_{v_5} Q[\mathbf{v}]}{\sum_{v_4, v_5} Q[\mathbf{v}]} \\
 &= \frac{\sum_{v_3, v_4, v_5} P(\mathbf{v})}{\sum_{v_2, v_3, v_4, v_5} P(\mathbf{v})} \frac{\sum_{v_5} P(\mathbf{v})}{\sum_{v_4, v_5} P(\mathbf{v})} \\
 &= \frac{P(v_1, v_2)}{P(v_1)} \frac{P(v_1, v_2, v_3, v_4)}{P(v_1, v_2, v_3)} \\
 &= P(v_2 \mid v_1) P(v_4 \mid v_1, v_2, v_3)
 \end{aligned}$$

How to get just $Q[V_2]$ or $Q[V_4]$?
Both are ancestral in $G_{\{V_2, V_4\}}$!

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



C-Component Factorization

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

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

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

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

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

$$= P(v_2 | v_1) P(v_4 | v_1, v_2, v_3)$$

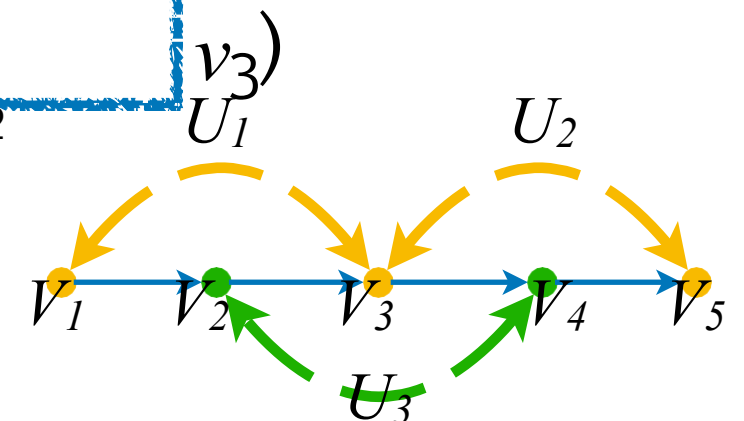
$$= P(v_2 | v_1) P(v_4 | v_1, v_2, v_3)$$

How to get just $Q[V_2]$ or $Q[V_4]$?
Both are ancestral in $G_{\{V_2, V_4\}}$!

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

$$= P(v_2 | v_1)$$

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

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

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

2. Factorize into c-components

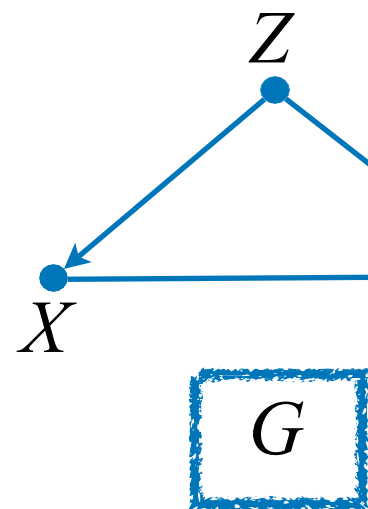
$$Q[\mathbf{H}] = \prod Q[H_j]$$

Where $H_1, \dots, H_k$, are the c-components in G_H

Expressing Causal Queries in terms of C-factors

Causal Effect in terms of C-Factors

- Consider a model with the following causal diagram

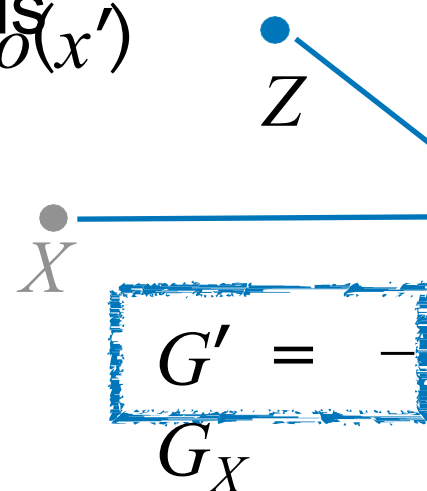


$$P(z, x, y) = \sum_{\mathbf{u}} P(z, x, y, \mathbf{u})$$

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

$$P(z, x, y) = Q[Z] Q[X | Z] Q[Y | Z, X] \mathbb{1}[f_z(u_z) = z] \mathbb{1}[f_x(z, u_x) = x] \mathbb{1}[f_y(x, z, u_y) = y]$$

- Let $M' = M_{x'}$, then the corresponding causal diagram $G' = G_X$ is



$$P'(z, x, y) = \sum_{\mathbf{u}} P'(z, x, y, \mathbf{u}) \mathbb{1}[x = x']$$

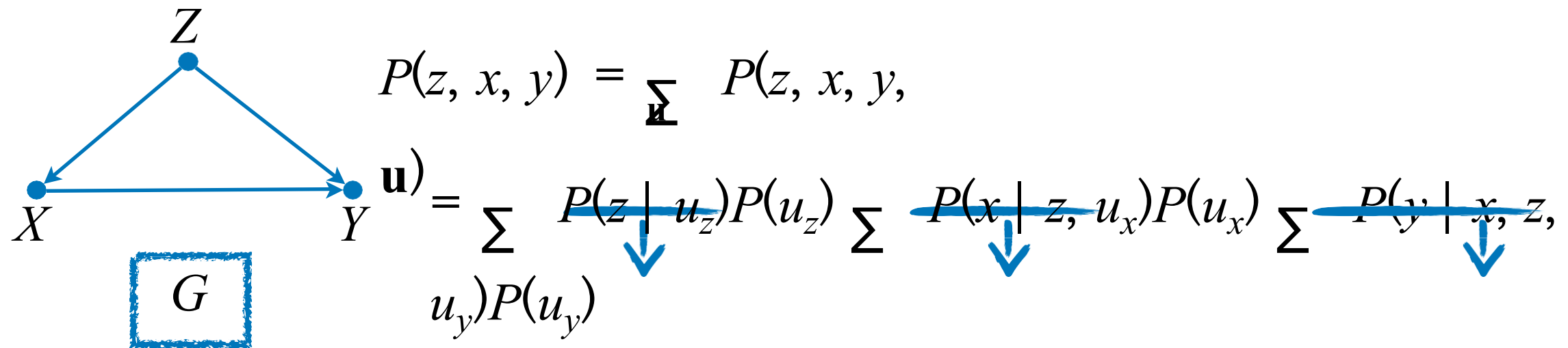
$$= \sum_{u_z} P'(z | u_z) P(u_z) P'(x) \sum_{u_y} P'(y | x, z, u_y) P(u_y)$$

$$= Q[Z] \mathbb{1}[x = x'] P(z, y | do(x')) = P'(z, x', y) =$$

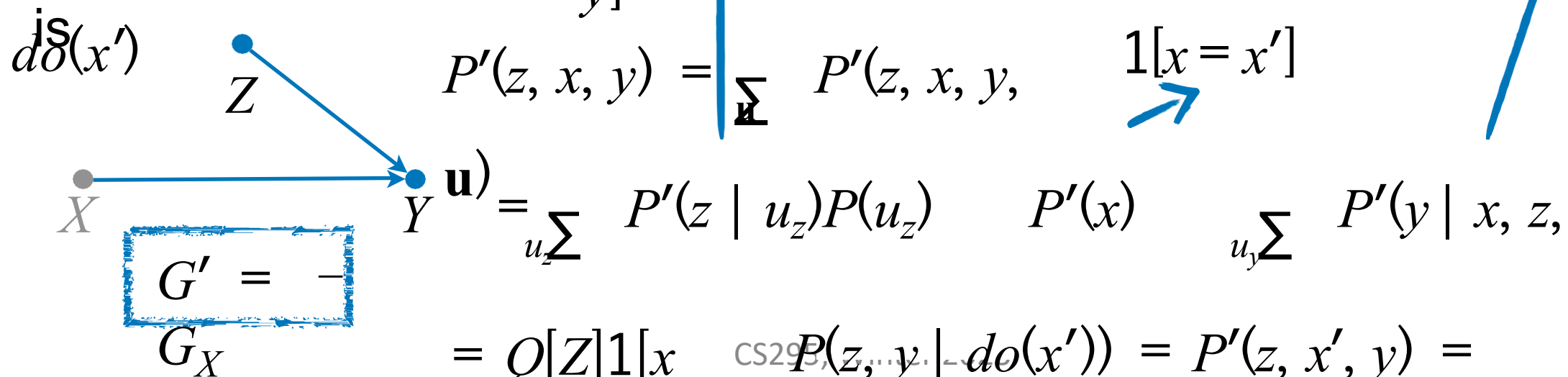
$$Q[Z] Q[Y | Z]$$

Causal Effect in terms of C-Factors

- Consider a model with the following causal diagram

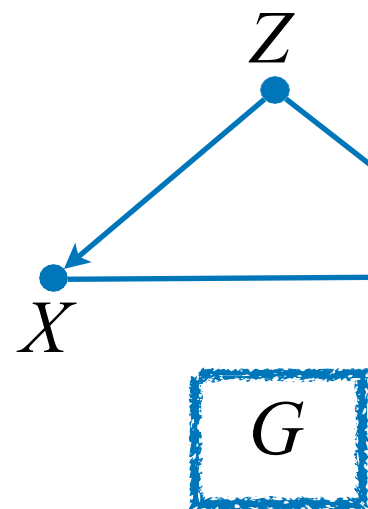


- Let $M' = M_{x'}$, then the corresponding causal diagram $G' = G_X$



Causal Effect in terms of C-Factors

- Consider a model with the following causal diagram



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

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

$$P(z, x, u_y) = Q[Z] Q[X | Z] Q[Y | Z, X]$$

$$1[f_z(u_z) = z] 1[f_x(z, u_x) = x] 1[f_y(x, z, u_y) = y]$$

- Let $M' = M_{x'}$, then the corresponding causal diagram $G' = G_X$

$do(x')$



$$P'(z, x, y) = \sum_{\mathbf{u}} P'(z, x, y, \mathbf{u}) 1[x = x']$$

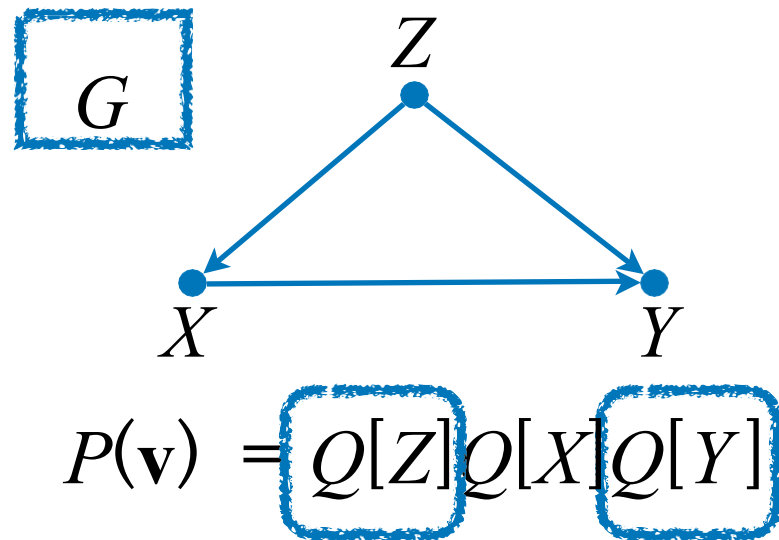
$$= \sum_{u_z} P'(z | u_z) P(u_z) P'(x) \sum_{u_y} P'(y | x, z, u_y) P(u_y)$$

$$G' = G_X$$

$$= Q[Z] 1[x = x'] P(z, y | do(x')) = P'(z, x', y) =$$

Causal Effect in terms of C-Factors

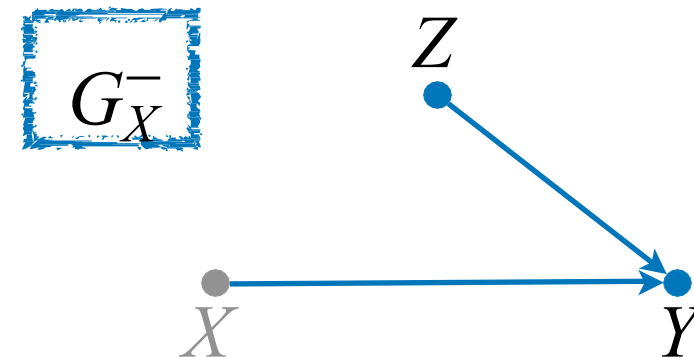
- Consider an intervention $do(x)$



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

$$Q[Z] = \frac{Q[Z]}{Q[\emptyset]} = \frac{\sum_{y,x} Q[Z, X, Y]}{\sum_{z,y,x} Q[Z, X, Y]} =$$

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



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

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

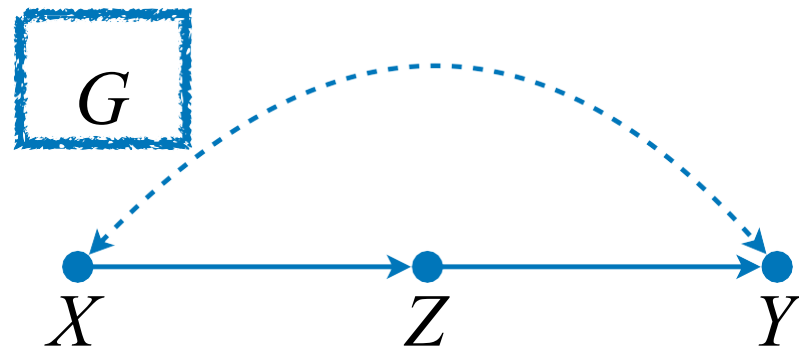
Back-door!

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

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

Causal Effect in terms of C-Factors

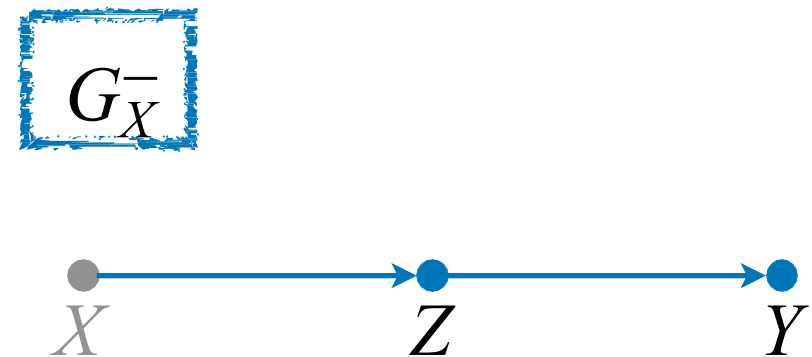
- Consider an intervention



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

$$Q[Z] = \frac{Q[X, Z]}{Q[X]} = P(Z |$$

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



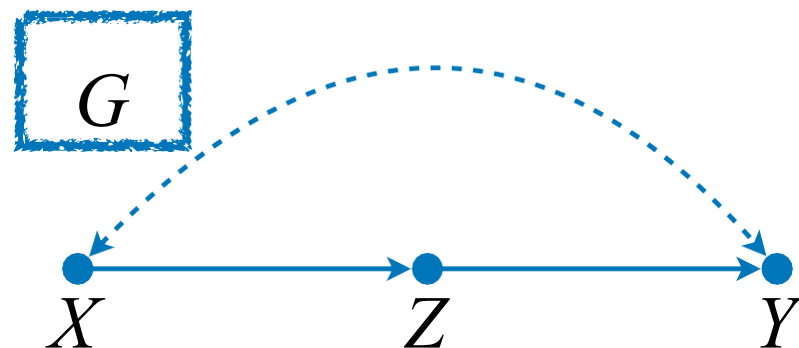
$$P(y | do(x)) = \sum_z P(y, z | do(x))$$

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

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

Causal Effect in terms of C-Factors

- Consider an intervention

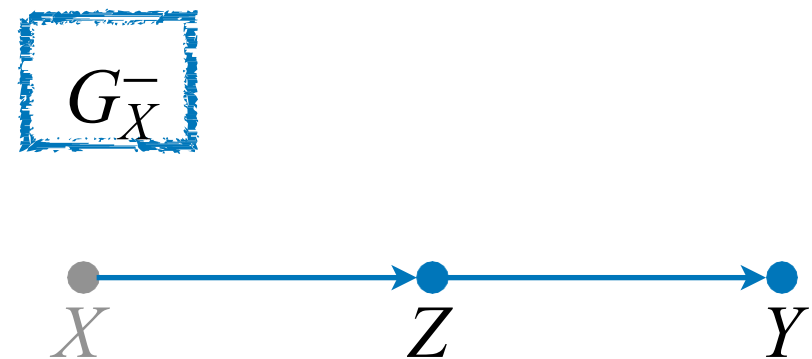


$$Q[X, Y] = \frac{Q[X]}{Q[\emptyset]} \frac{Q[X, Z, Y]}{Q[X, Z]}$$

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



$$\begin{aligned} Q[Y] &= \sum_x Q[X, Y] \\ &= \sum_x P(x) P(y | x, \end{aligned}$$



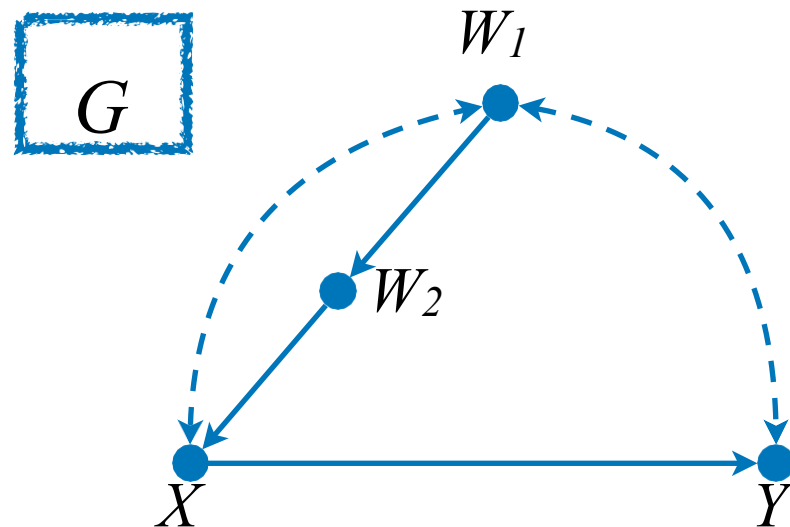
$$\begin{aligned} P(y | do(x)) &= \sum_z Q[Y] Q[Z] \\ &= \sum_z \left(\sum_{x'} P(x') P(y | x', z) \right) P(z | x) \end{aligned}$$

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

Front-door!

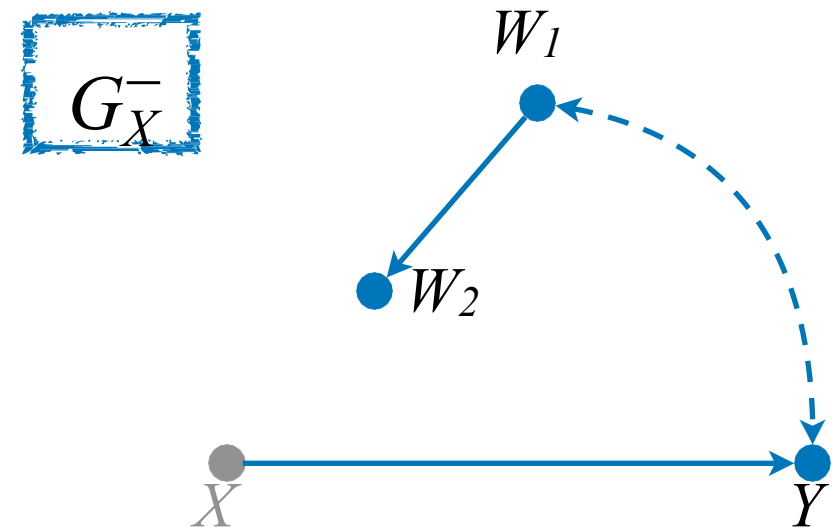
Solving the Napkin

- Recall the Napkin graph from last time



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

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



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

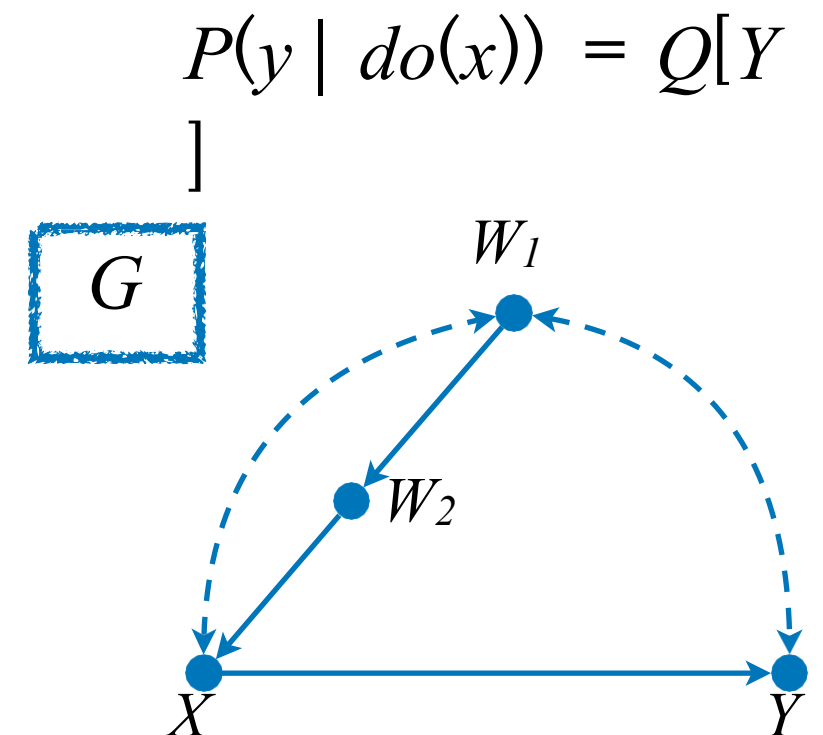
- $\{Y\}$ is ancestral in $G_{\{W_1, W_2, Y\}}$, so

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

Solving the Napkin

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

$$\begin{aligned}
 Q[W_1, X, Y] &= \frac{Q[W_1]}{Q[\emptyset]} \frac{Q[W_1, W_2, X]}{Q[W_1, W_2]} \frac{Q[W_1, W_2, X, Y]}{Q[W_1, W_2, X]} \\
 &= \frac{P(w_1)}{1} \frac{P(w_1, w_2, x)}{P(w_1, w_2)} \frac{P(w_1, w_2, x, y)}{P(w_1, w_2, x)} \\
 &= P(w_1)P(x \mid w_1, w_2)P(y \mid w_1, w_2, x)
 \end{aligned}$$

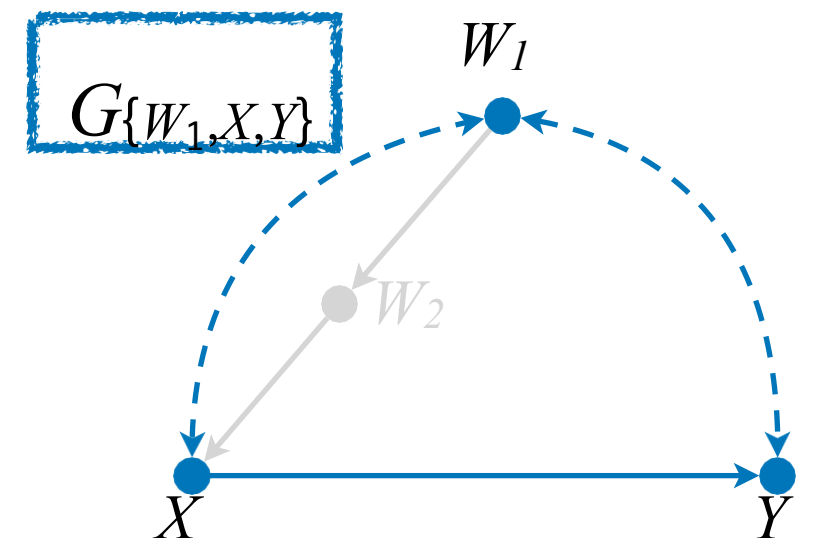


Solving the Napkin

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

$$\begin{aligned} Q[X, Y] &= \sum_{w_1} Q[w_1, X, Y] \\ &= \sum_{w_1} P(w_1) P(x \mid w_1, w_2) P(y \mid w_1, w_2, x) \end{aligned}$$

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



Solving the Napkin

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

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

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

$$= \frac{\sum_{w_1} P(w_1) \sum_x Q[X, Y] P(x | w_1, w_2) P(y | w_1, w_2, x)}{\sum_{w_1} P(w_1) \sum_x Q[X, Y] P(x | w_1, w_2) P(y | w_1, w_2, x)}$$

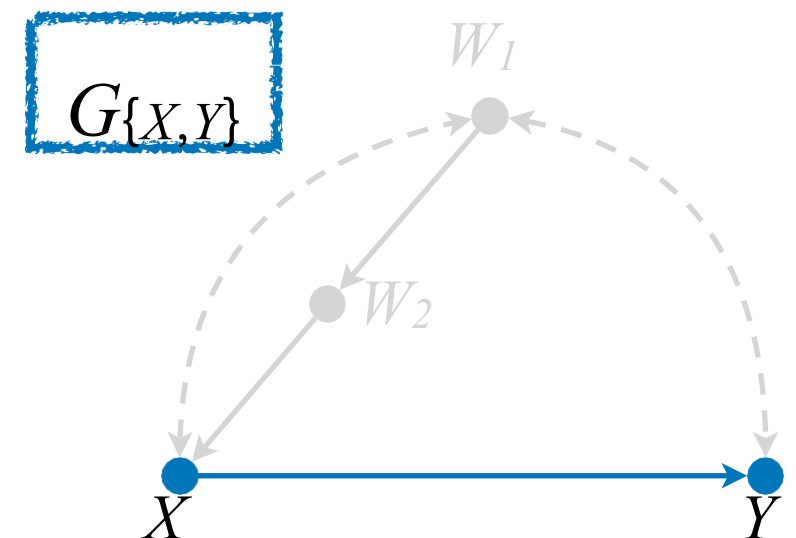
$$= \frac{\sum_{w_1} P(w_1) \sum_x P(x | w_1, w_2) P(y | w_1, w_2, x)}{\sum_{w_1} P(w_1) \sum_x P(x | w_1, w_2) P(y | w_1, w_2, x)}$$

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

$$do(x)) \quad \sum_{w_1} P(w_1) P(x | w_1, w_2)$$

$$= P(y |$$

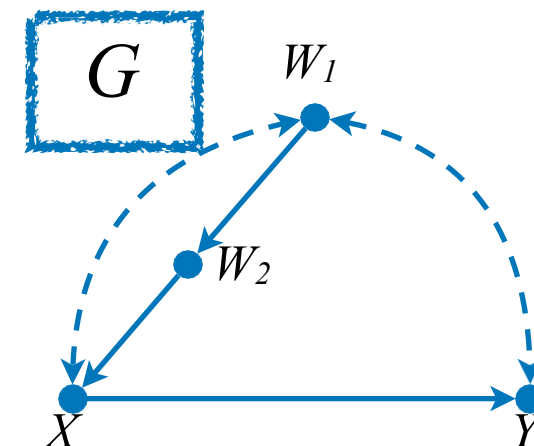
$$P(y | do(x)) = Q[Y]$$



Solve the Napkin using Do-Calculus

- Let's see an equivalent do-calculus derivation

$$P(y \mid do(x)) = P(y \mid do(x, w_2, w_1)) \quad \text{Rule 3: } (Y \perp\!\!\!\perp W_2, W_1 \mid X)_{G_{\overline{XW_1, W_2}}}$$



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

$$= \frac{P(y, x \mid do(w_2, w_1))}{P(x \mid do(w_2, w_1))} \quad \text{Rule 3: } (Y, X \perp\!\!\!\perp W_1 \mid W_2)_{G_{\overline{W_1W_2}}}$$

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

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

$$= \sum_{w_1} P(x \mid w_2, w_1) P(w_1) \quad \text{Rule 3: } (W_1 \perp\!\!\!\perp W_2)_{G_{\overline{W_2}}}$$

A General Approach

A General Identification Algorithm

- Given G and the query variables X, Y

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

$$= \sum_{\mathbf{D} \perp\!\!\!\perp \mathbf{y}} Q[\mathbf{D}] \quad \text{where } \mathbf{D} = An(\mathbf{Y}) \text{ in } G_{\mathbf{X}}$$

- Suppose the graph $G_{\mathbf{D}}^{\mathbf{y}}$ has C-components $\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_k$, then

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{d} \perp\!\!\!\perp \mathbf{y}} \prod_i Q[\mathbf{D}_i]$$

A General Identification Algorithm

Function Identify(C, T, Q)

INPUT: $C \subseteq T \subseteq V$, $Q = Q[T]$. Assuming G_T is composed of one single c-component.

OUTPUT: Expression for $Q[C]$ in terms of Q or fail to determine.

Let $A = An(C)_{G_T}$.

- IF $A = C$, output $Q[C] = \sum_{T \setminus C} Q$.
- IF $A = T$, output FAIL.
- IF $C \subset A \subset T$
 1. Assume that in G_A , C is contained in a c-component T' .
 2. Compute $Q[T']$ from $Q[A] = \sum_{T \setminus A} Q$ by Lemma 11.
 3. Output Identify($C, T', Q[T']$).

Figure 5.9: A function determining if $Q[C]$ is computable from $Q[T]$.

Completeness

Theorem [Huang and Valtorta, 2008]

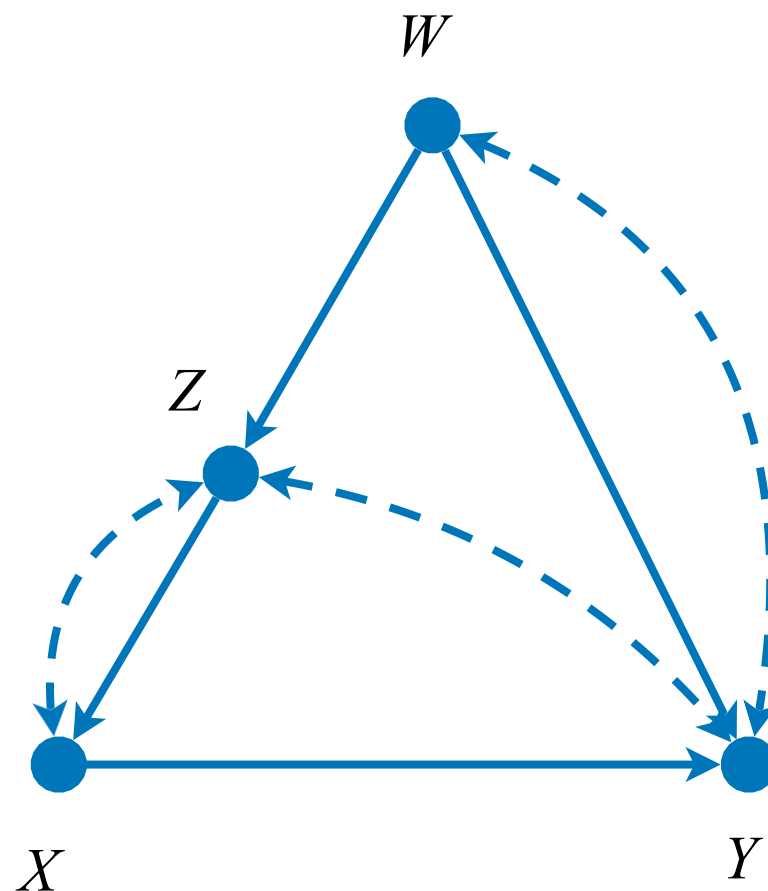
The causal effect $P(\mathbf{y}|\mathit{do}(\mathbf{x}))$ is identifiable from causal diagram G and $P(\mathbf{v})$ *if and only if* each of the C-factors $Q[\mathbf{D}_i]$ is identifiable by

`Identify( $\mathbf{D}_i, \mathbf{C}_i, Q[\mathbf{C}_i], G$ )`.

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

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.