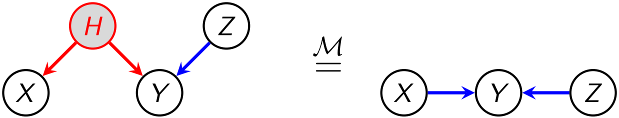
Marginal DAG models equivalent to another DAG

Robin J. Evans, University of Oxford

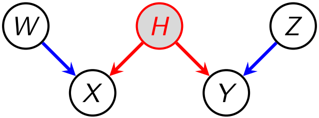
Quantum Physics and Statistical Causal Models
Simons Institute, April 2022

The Problem

When is the marginal model of a DAG equivalent to the Markov model of another DAG without any hidden variables?



These models are easily seen to be equivalent.



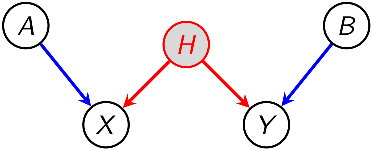
This model is known not to be equivalent to any DAG.

Literature

DAG models are often used as causal models, and we may not wish to assume **causal sufficiency** (i.e. all important variables are measured).

The problem of finding constraints in **marginalized DAG** models has a rich literature. In addition to the work of Robins, Pearl, Verma, Geiger, Richardson, Spirtes, Tian and others on finding **equality** contraints:

Bell (1964) was the first to propose inequalities on a DAG model, which he showed could be violated by quantum models.

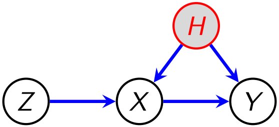


This was followed by Clauser et al. (1969) who developed the **CHSH inequality**.

Literature II

Pearl (1995) introduced the **instrumental inequality**, which gave a constraint on binary instrumental variable models.

Bonet (2001) expanded Pearl's work using computational algebra.



Later, Kang and Tian (2006), Evans (2012), Chaves et al. (2014), Fraser (2019), Kédagni and Mourifié (2020) and many others proposed **graphical approaches** to deriving inequalities.

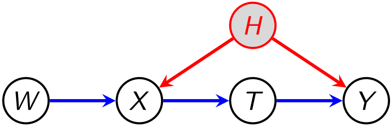
Most recently Wolfe et al. (2019) give the **inflation technique**, a complete method to obtain inequalities (Navascués and Wolfe, 2020).

Constraints

The marginal DAG model induces three kinds of constraints on P:

- **conditional independences**;
- **nested conditional independences**;
- **inequalities**.

$\mathcal{G}$



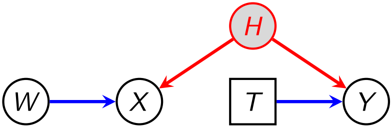
We notice that this graph has d-separation $W \perp_d T \mid X$.

Constraints

The marginal DAG model induces three kinds of constraints on P:

- **conditional independences**;
- **nested conditional independences**;
- **inequalities**.

$\mathcal{G}$



We notice that this graph has d-separation $W \perp_d T \mid X$.

If we consider

$$p^*(w, x, y \mid t) = \frac{p(w, x, t, y)}{p(t \mid x)}$$

then we find $Y \perp_d W \mid T$.
This is a **nested constraint** (or dormant independence).

This gives the 'Bell graph', so has a (classical and quantum) **inequality**.

Claim

We will show the following result:

Theorem (Main Result)

Given an mDAG $\mathcal{G}$, its marginal model $\mathcal{M}(\mathcal{G})$ is the same as $\mathcal{M}(\mathcal{D})$ for an ordinary DAG $\mathcal{D}$ **if and only if** there are no non-trivial inequality constraints implied by $\mathcal{G}$.

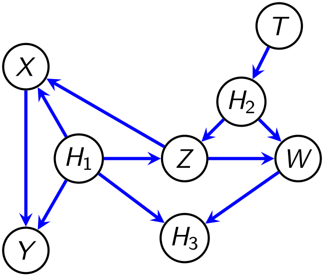
What is a **non-trivial** inequality constraint?
(Clearly probabilities are be non-negative.)

By this I mean an inequality constraint that is **not directly implied** by any of the equality constraints, nor the fact that this is a probability distribution.

The question of when $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{D})$ is of interest in semi-parametric efficiency theory, since the tangent space of DAGs are well understood.

DAG Models

A **directed acyclic graph** (DAG) is a graph with only directed edges that contains no directed cycle.



A distribution is in the model for a DAG $\mathcal{G}$ if each variable is a measurable function of its parents and some independent noise.

Equivalently, if for each vertex $Y \in \mathbf{V}$, we have

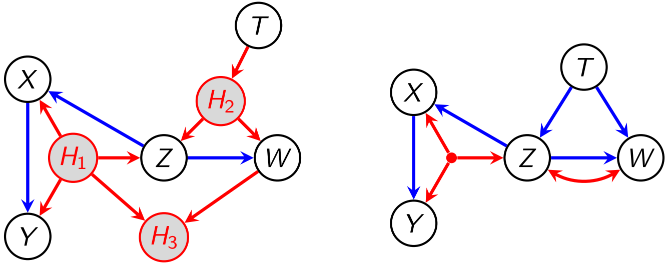
$$Y \perp\!\!\!\perp X_{\mathrm{pre}(Y;<)\setminus \mathrm{pa}(Y)} \mid X_{\mathrm{pa}(Y)}.$$

mDAGs

An mDAG is obtained from a DAG on $\mathbf{V} \dot{\cup} \mathbf{L}$ by marginalizing (E., 2016).

It consists of a DAG on $\mathbf{V}$ and a **simplicial complex** $\mathcal{B}$ (also over $\mathbf{V}$).

Example.

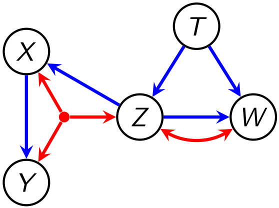


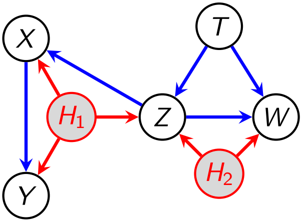
Here $\mathcal{B} = \{A : A \subseteq \{X, Y, Z\} \text{ or } A \subseteq \{W, Z\}\}$.

Canonical DAGs

Given an mDAG $\mathcal{G}$, its **canonical DAG**, $\bar{\mathcal{G}}$, is the one that takes the directed edges from $\mathcal{G}$ and adds in a latent for each facet of $\mathcal{B}$.

Example.





Marginal Model

Consider an mDAG $\mathcal{G}$ with vertices $\mathbf{V}$ and its canonical DAG $\bar{\mathcal{G}}$.

Then we say that a distribution P over $\mathbf{V}$ is in the **marginal model** for $\mathcal{G}$ if there exists a distribution Q such that:

- Q is Markov with respect to $\bar{\mathcal{G}}$;
- the margin of Q over $\mathbf{V}$ is P.

We write this collection of distributions (the **model**) as $\mathcal{M}(\mathcal{G})$.

Example. Consider the mDAG $X \leftrightarrow Y \rightarrow Z$. Then we have that

$$\mathcal{M}(\mathcal{G}) := \{P : X \perp\!\!\!\perp Z \mid Y\}.$$

Proof Sketch

Here is an outline of our proof strategy.

It is clear that there are no non-trivial inequalities in any DAG model.

For the other direction:

- first, we show that taking the **maximal arid projection** of the graph will only increase the size of the model;
- then we show that any model with a non-trivial nested constraint will imply a non-trivial inequality;

- then we show that taking the **maximal ancestral projection** will only increase the size of the model;
- then finally we show that any model with conditional independences not equivalent to a DAG will also induce a non-trivial inequality.

A Useful Result

Proposition

Let $\mathcal{G}$ be an mDAG and suppose that it contains a bidirected facet $B = C \dot{\cup} D$ such that:

- every bidirected face involving $c \in C$ is either a subset of B or all its vertices are parents of **every** $d \in D$;
- every element of $C \cup \text{pa}_{\mathcal{G}}(C)$ is also a parent of **every** $d \in D$.

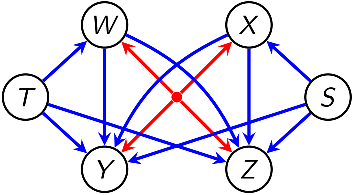
Then we can replace B with the separate facets C and D and the marginal model is preserved.

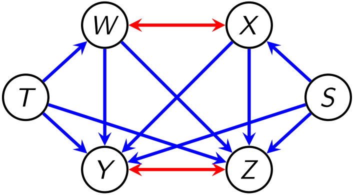
This a slight extension of Proposition 6.1* of Evans (2016).

(Wolfe and Ansanelli, personal communication.)

* Proposition 5 of the journal version.

Example

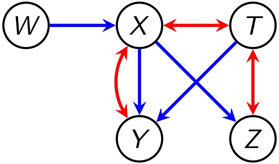




Intrinsic Sets

The **intrinsic closure** of a set is obtained by alternating between taking ancestors and the district (connected by bidirected paths) until we reach a fixed point.

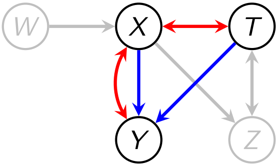
Example. Consider $\langle\{Y\}\rangle_{\mathcal{G}}$



Intrinsic Sets

The **intrinsic closure** of a set is obtained by alternating between taking ancestors and the district (connected by bidirected paths) until we reach a fixed point.

Example. Consider $\langle\{Y\}\rangle_{\mathcal{G}}$

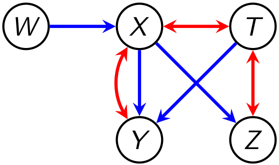


Hence $\langle\{Y\}\rangle_{\mathcal{G}} = \{X, T, Y\}$.

Intrinsic Sets

The **intrinsic closure** of a set is obtained by alternating between taking ancestors and the district (connected by bidirected paths) until we reach a fixed point.

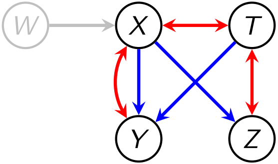
Example. Consider $\langle \{Y, Z\} \rangle_{\mathcal{G}}$



Intrinsic Sets

The **intrinsic closure** of a set is obtained by alternating between taking ancestors and the district (connected by bidirected paths) until we reach a fixed point.

Example. Consider $\langle \{Y, Z\} \rangle_{\mathcal{G}}$



Hence $\langle \{Y, Z\} \rangle_{\mathcal{G}} = \{X, T, Y, Z\}$.

Maximal Arid Graphs

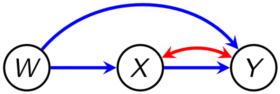
We say that an mDAG is **arid** if, for every vertex $v \in \mathbf{V}$, we have

$$\langle v \rangle_{\mathcal{G}} = \{v\}.$$

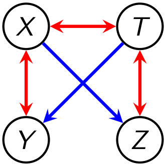
That is, every singleton vertex is **intrinsic**.

An arid graph is **maximal** if every pair of edges with no nested constraint is adjacent.

Examples.



A maximal graph that is not arid.



An arid graph that is not maximal.

Maximal Arid Projection

We can obtain an arid graph that is **nested Markov equivalent** by applying the **maximal arid projection**.

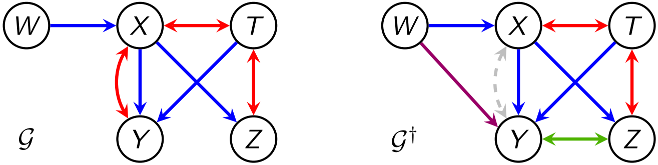
Steps:

- first, **add** directed edges between every $w \in \mathrm{pa}_{\mathcal{G}}(\langle v \rangle)$ and v;
- then for any non-adjacent v, w with bidirected-connected $\langle \{v, w\} \rangle_{\mathcal{G}}$ **add** a bidirected edge;
- then **remove** any bidirected edges where there is a directed edge.

Example. Consider the model below.

We have $\langle \{Y\} \rangle_{\mathcal{G}} = \{X, T, Y\}$, so add $W \to Y$.

Also $\langle \{Y, Z\} \rangle_{\mathcal{G}} = \{X, T, Y, Z\}$ so add $Y \leftrightarrow Z$.



Why are MArGs Sufficient?

First, taking the maximal arid projection is known not to change the nested constraints in the model (Shpitser et al., 2018).

Additionally:

Theorem

For any mDAG $\mathcal{G}$ we have $\mathcal{M}(\mathcal{G}) \subseteq \mathcal{M}(\mathcal{G}^\dagger)$.

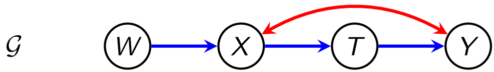
Hence, if $\mathcal{G}$ is a counterexample to our result, then so is $\mathcal{G}^\dagger$.

So we can consider only maximal arid graphs.

Nested Constraints

Suppose that there is a non-trivial nested constraint. This means that we 'create' an m-separation by **fixing** a vertex in the graph.

This requires that the graph is **not** ancestral, i.e. there is a vertex joined to one of its own ancestors by a latent variable.



After a fixing, we end up with a conditional model, and in this case it is the Bell graph.

Similar arguments show...

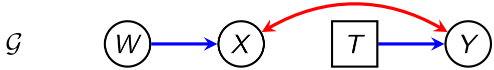
Proposition

For any MArG with non-trivial nested constraint, there is also a non-trivial inequality.

Nested Constraints

Suppose that there is a non-trivial nested constraint. This means that we 'create' an m-separation by **fixing** a vertex in the graph.

This requires that the graph is **not** ancestral, i.e. there is a vertex joined to one of its own ancestors by a latent variable.



After a fixing, we end up with a conditional model, and in this case it is the Bell graph.

Similar arguments show...

Proposition

For any MArG with non-trivial nested constraint, there is also a non-trivial inequality.

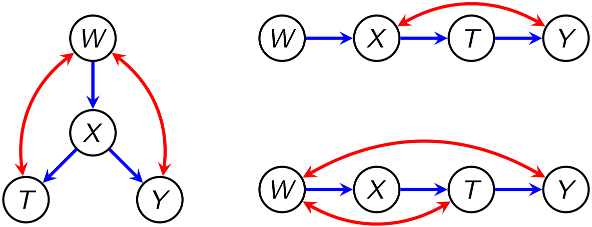
Proof Sketch

In order for there to be a (non-trivial) nested constraint, must have:

- two vertices (say, c, d) in the same district, c is a collider;
- a vertex v in $\mathrm{de}_{\mathcal{G}}(c) \cap \mathrm{an}_{\mathcal{G}}(d)$ **not** in the same district.

In addition, the new constraint involves a descendant of the fixed node.

After fixing there are generically three cases with a completely new independence:



Proof Sketch

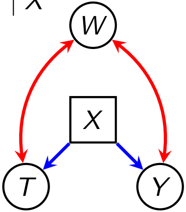
In order for there to be a (non-trivial) nested constraint, must have:

- two vertices (say, c, d) in the same district, c is a collider;
- a vertex v in $\mathrm{de}_{\mathcal{G}}(c) \cap \mathrm{an}_{\mathcal{G}}(d)$ **not** in the same district.

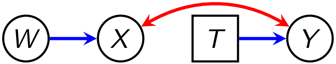
In addition, the new constraint involves a descendant of the fixed node.

After fixing there are generically three cases with a completely new independence:

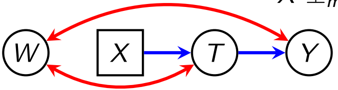
$T \perp_m Y \mid X$



$W \perp_m Y \mid T$



$X \perp_m Y \mid T$



Conditional Independence and MAGs

This means that any model that could be a counterexample **cannot** contain a nested constraint.

Therefore $\mathcal{G}$ is nested Markov equivalent to a **maximal ancestral graph** or MAG (Richardson and Spirtes, 2002).

This is a graph which is

- **maximal**: every pair of edges that cannot be m-separated is joined by an edge;
- **ancestral**: no vertex shares a latent parent with any of its ancestors.

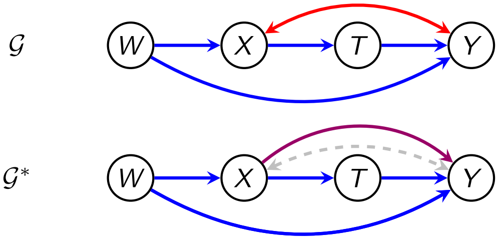
Maximal Ancestral Projection

We can obtain an ancestral graph that is **ordinary Markov equivalent** by applying the **maximal ancestral projection**.

Steps: for any pair v, w that cannot be m-separated:

- if $v \in \text{an}_{\mathcal{G}}(w)$ then **add** $v \to w$ (and vice versa);
- otherwise, **add** $v \leftrightarrow w$;
- then **remove** any bidirected edges between already adjacent nodes.

Example. Consider the model below. Note that everything is ordered, so we only add directed edges.



Why are MAGs Sufficient?

We know: non-trivial nested constraint implies a non-trivial inequality.

Thus any counterexample will only contain **conditional independence** equality constraints.

Taking the maximal ancestral projection is known to preserve these.

Additionally:

Theorem

For any mDAG with $\mathcal{G}$ we have $\mathcal{M}(\mathcal{G}) \subseteq \mathcal{M}(\mathcal{G}^*)$.

Hence, if $\mathcal{G}$ counterexample to our result, then so is $\mathcal{G}^*$.

So we can consider only maximal ancestral graphs.

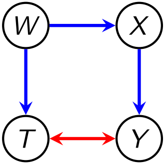
Graphs Inducing Inequalities

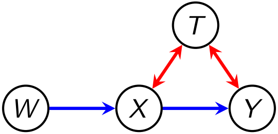
Consider the class of MAGs 'ordinary' Markov equivalent to $\mathcal{G}$.

This is equivalent to a DAG if and only if there is no edge that is **always** bidirected.

This can happen for one of two reasons:

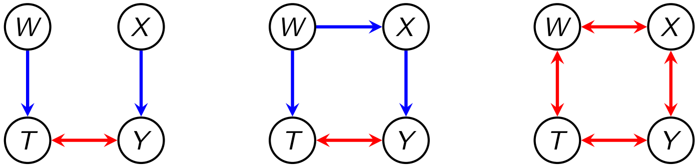
- both the end points are unshielded colliders;
- there is a discriminating path.



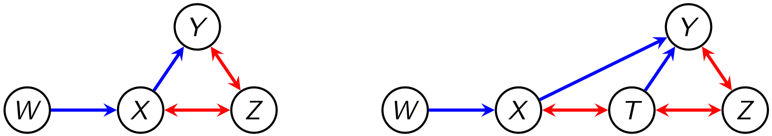


Inequalities

We know that a 4-chain of vertices joined by two unshielded colliders induces inequalities:



The same is true for discriminating path structures because there is an 'e-separation' between the W and Y, Z (E., 2012).



We see that $W \perp_e Y, Z$ in both cases.

Done!

We have shown that:

- if a graph is a counterexample to the claim, then so is its maximal arid closure;
- if a graph has a non-trivial nested constraint, then it also has an inequality;

- if a graph without nested constraints is a counterexample to the claim, then so is its maximal ancestral closure;
- if a nested-constraint free graph has a **definite** bidirected edge then there is non-trivial nested constraint.

Hence we have proven the main result!

Summary

We have given a result that shows an mDAG is distributionally equivalent to a DAG if and only if it implies no non-trivial inequalities.

As part of the proof, we have introduced:

- maximal arid projection;
- nested constraints;
- maximal ancestral projection;
- discriminating paths;
- e-separation.

We could *probably* simplify the proof if we understood the 'nested PAG' (we're working on this!)

Thank you!

References

Bell. On the Einstein-Podolsky-Rosen paradox, *Physics*, 1964.

Bonet. Instrumentality tests revisited, *UAI*, 2001.

Chaves et al. Inferring latent structures via information inequalities, *UAI*, 2014.

Clauser et al. Proposed experiment to test local hidden-variable theories, *Physical Review Letters*, 1969

Evans. Graphical methods for inequality constraints in marginalized DAGs, *MLSP*, 2012.

Evans. Graphs for margins of Bayesian Networks, *Scand. J. Statist.*, 2012.

Kang, Tian. Inequality constraints in causal models with hidden variables, *UAI*, 2006.

Kédagni and Mourifié. Generalized instrumental inequalities: testing the instrumental variable independence assumption. *Biometrika*, 2020.

Navascués and Wolfe. The inflation technique completely solves the causal compatibility problem, *J. Causal Inf.*, 2020.

Pearl. On the testability of causal models with latent and instrumental variables, *UAI*, 1995.

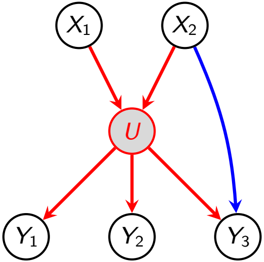
Richardson and Spirtes. Ancestral graph Markov properties, *Ann. Stat.*, 2002.

Shpitser et al. Acyclic linear SEMs obey the nested Markov property. *UAI*, 2018.

Wolfe et al. The inflation technique for causal inference, *J. Causal Inf.*, 2019.

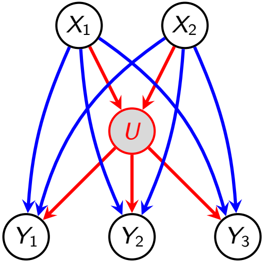
Sufficiency of Latent Projection

Suppose we wish to project out the vertex U.



Sufficiency of Latent Projection

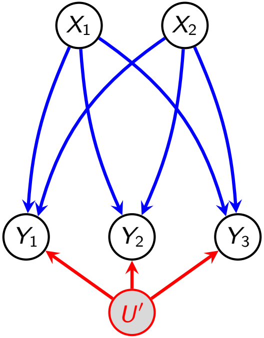
Suppose we wish to project out the vertex U.



Note that U can pass all information about its parents to its children, so we may just as well draw directed edges.

Sufficiency of Latent Projection

Suppose we wish to project out the vertex U.

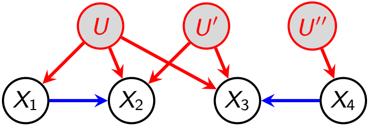


Note that U can pass all information about its parents to its children, so we may just as well draw directed edges.

Then, by a result of Chentsov (1982), it is equivalent to have U be independent of its parents over the observed margin.

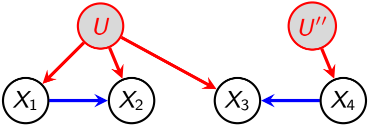
Sufficiency of Latent Projection

Then we need only consider exogenous latents with maximal child sets.



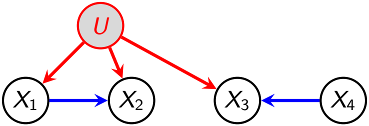
Sufficiency of Latent Projection

Then we need only consider exogenous latents with maximal child sets.



Sufficiency of Latent Projection

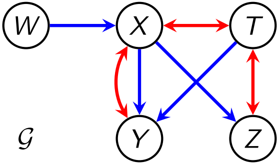
Then we need only consider exogenous latents with maximal child sets.



Singleton latents can be ignored.

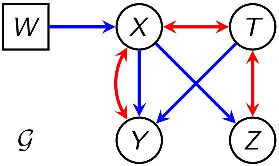
Intrinsic Closure

Suppose we want the intrinsic closure of $\{Y\}$.



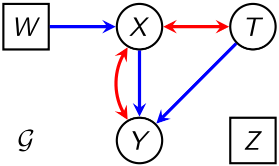
Intrinsic Closure

Suppose we want the intrinsic closure of $\{Y\}$.



Intrinsic Closure

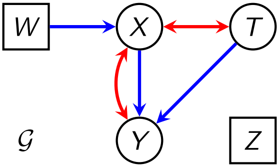
Suppose we want the intrinsic closure of $\{Y\}$.



Hence the intrinsic closure of Y is $\{X, T, Y\}$.

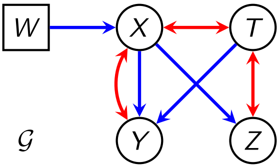
Intrinsic Closure

Suppose we want the intrinsic closure of $\{Y\}$.



Hence the intrinsic closure of Y is $\{X, T, Y\}$.

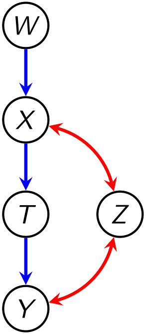
Now consider the intrinsic closure of $\{Y, Z\}$.



Clearly this is $\{X, T, Y, Z\}$, which is bidirected-connected.

The 16-not-18 Graph

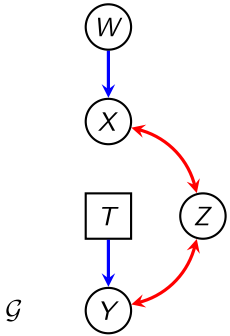
This graph has the m-separations $W \perp_m Y \mid T$ and $W \perp_m Z$, but no **joint** independence between W and Y, Z.



$\mathcal{G}$

The 16-not-18 Graph

This graph has the m-separations $W \perp_m Y \mid T$ and $W \perp_m Z$, but no **joint** independence between W and Y, Z.



Hence we can **uncover** the joint independence by fixing T, because now:

$$W \perp_m Y, Z \mid T.$$