2 Graphical Models and Their Applications

In **Part 1**, we introduced

probabilities + **graphs** + **structural equations** = **Causal Networks**

Independence

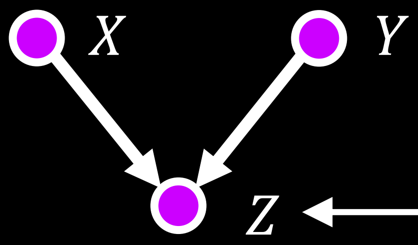
Algebraic Equalities

$P(x|y) = P(x) \quad \forall\, x, y$

$P(y|x) = P(y) \quad \forall\, x, y$

$P(x, y) = P(x)\, P(y) \quad \forall\, x, y$

Graphical Representation

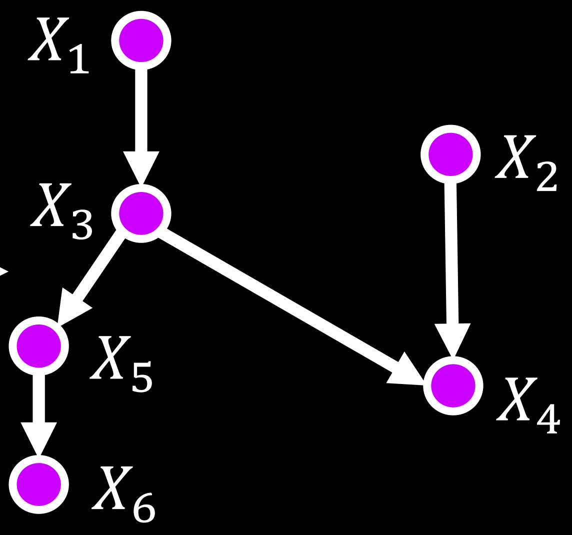


Embedded Probabilistic Information

$Z = 2X + Y$

The researcher who has scientific knowledge in the form of **structural equation model** is able to predict patterns of independencies in the data, based solely on the structure of the model's graph, without relying on any quantitative information carried by the equations or by the distributions of errors.

$$M = \langle U, V, F \rangle$$



X_1	X_2	X_3	X_4	X_5	X_6
0	1	1	1	0	1
0	1	0	1	1	1
1	0	1	1	0	0
0	0	1	0	1	1
0	1	0	0	0	0
1	1	0	1	0	1
1	0	1	1	1	1
0	1	1	1	0	1

Data Set

patterns of independencies in the data

$$X_6 \perp X_3 \mid X_5$$

$$X_6 \perp X_1 \mid X_5$$

$$X_6 \perp X_1 \mid X_3$$

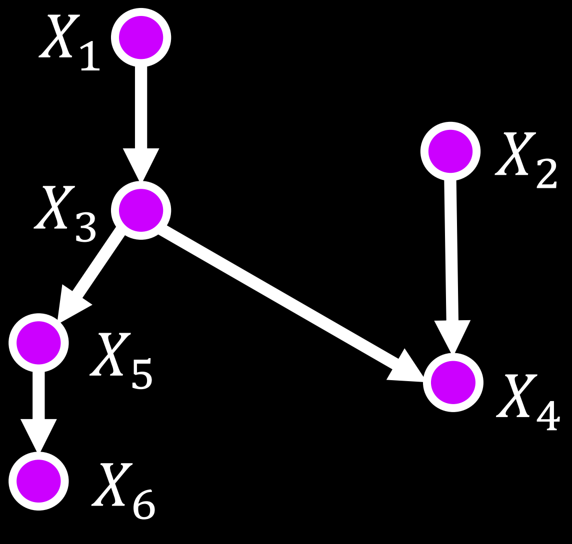
$$X_1 \perp X_2$$

2.1 Connecting Models to Data

Conversely, it means that observing **patterns of independencies** in the data enables us to say something about whether a hypothesized model is correct.

X_1	X_2	X_3	X_4	X_5	X_6
0	1	1	1	0	1
0	1	0	1	1	1
1	0	1	1	0	0
0	0	1	0	1	1
0	1	0	0	0	0
1	1	0	1	0	1
1	0	1	1	1	1
0	1	1	1	0	1

Data Set



Hypothesized Model

represent the causal

story behind the data

X_1	X_2	X_3	X_4	X_5	X_6
0	1	1	1	0	1
0	1	0	1	1	1
1	0	1	1	0	0
0	0	1	0	1	1
0	1	0	0	0	0
1	1	0	1	0	1
1	0	1	1	1	1
0	1	1	1	0	1

Data Set

$$M = \langle U, V, F \rangle$$

Causal Model

mechanism by which data

were generated

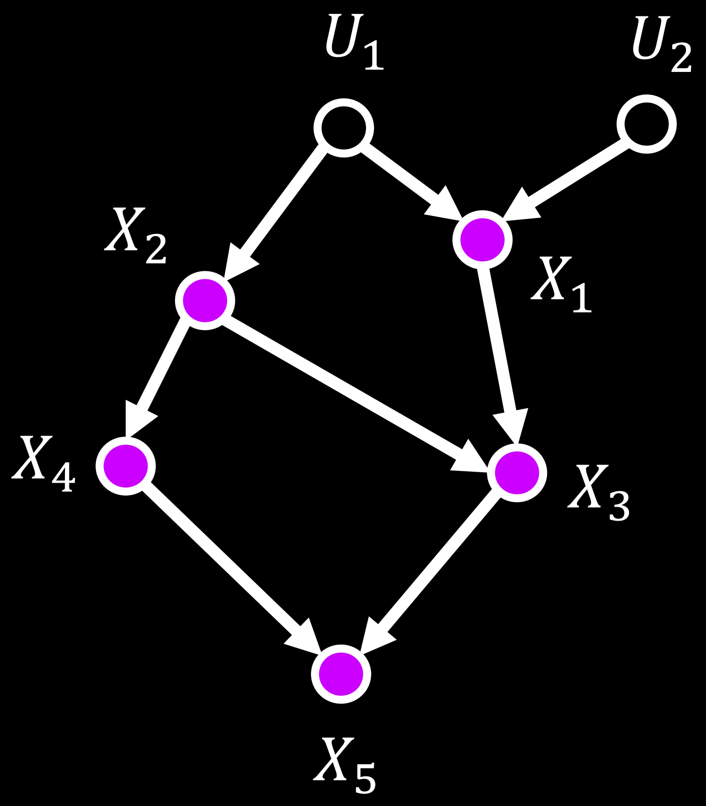
Given a truly complete causal model for, say, math test score in high school juniors, and given complete list of values for every exogenous variable in that model, we could theoretically generate a data point (i.e., a test score), for each individual (student).

$$M = \langle U, V, F \rangle$$

$$U = \{U_1, U_2\}$$

$$V = \{X_1, X_2, X_3, X_4, X_5\}$$

$$F = \{f_1, f_2, f_3, f_4, f_5\}$$



math test score

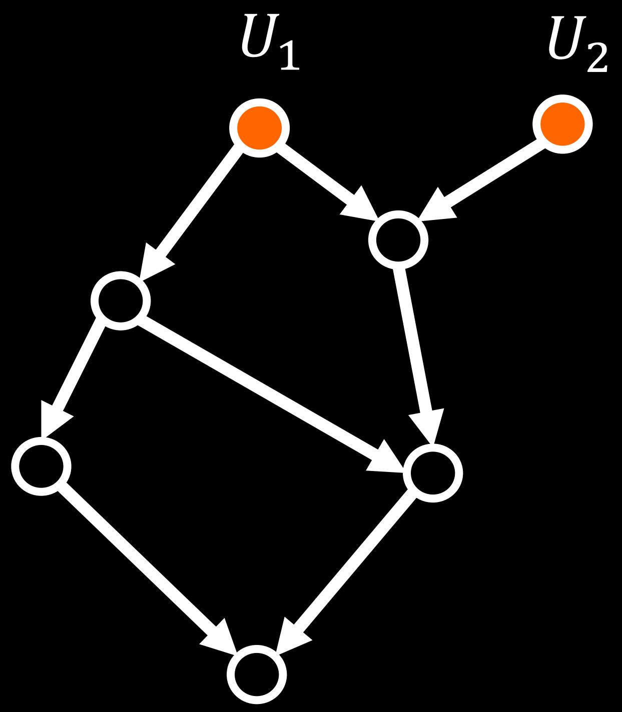
Given a truly complete causal model for, say, math test score in high school juniors, and given complete list of values for every exogenous variable in that model, we could theoretically generate a data point (i.e., a test score), for each individual (student).

$$M = \langle U, V, F \rangle$$

$$U = \{U_1, U_2\}$$

$$V = \{X_1, X_2, X_3, X_4, X_5\}$$

$$F = \{f_1, f_2, f_3, f_4, f_5\}$$



math test score

2.2 Chain and Forks

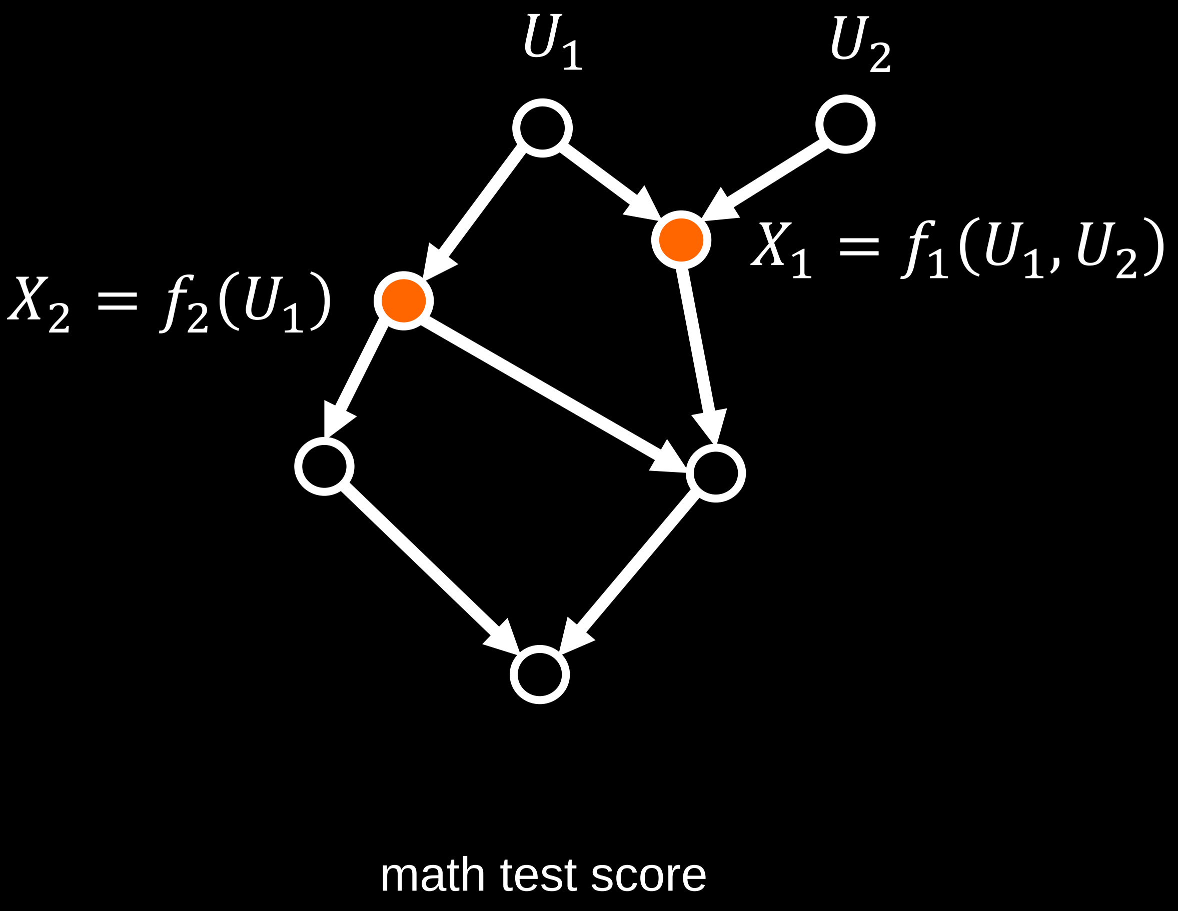
Given a truly complete causal model for, say, math test score in high school juniors, and given complete list of values for every exogenous variable in that model, we could theoretically generate a data point (i.e., a test score), for each individual (student).

$$M = \langle U, V, F \rangle$$

$$U = \{U_1, U_2\}$$

$$V = \{X_1, X_2, X_3, X_4, X_5\}$$

$$F = \{f_1, f_2, f_3, f_4, f_5\}$$



2.2 Chain and Forks

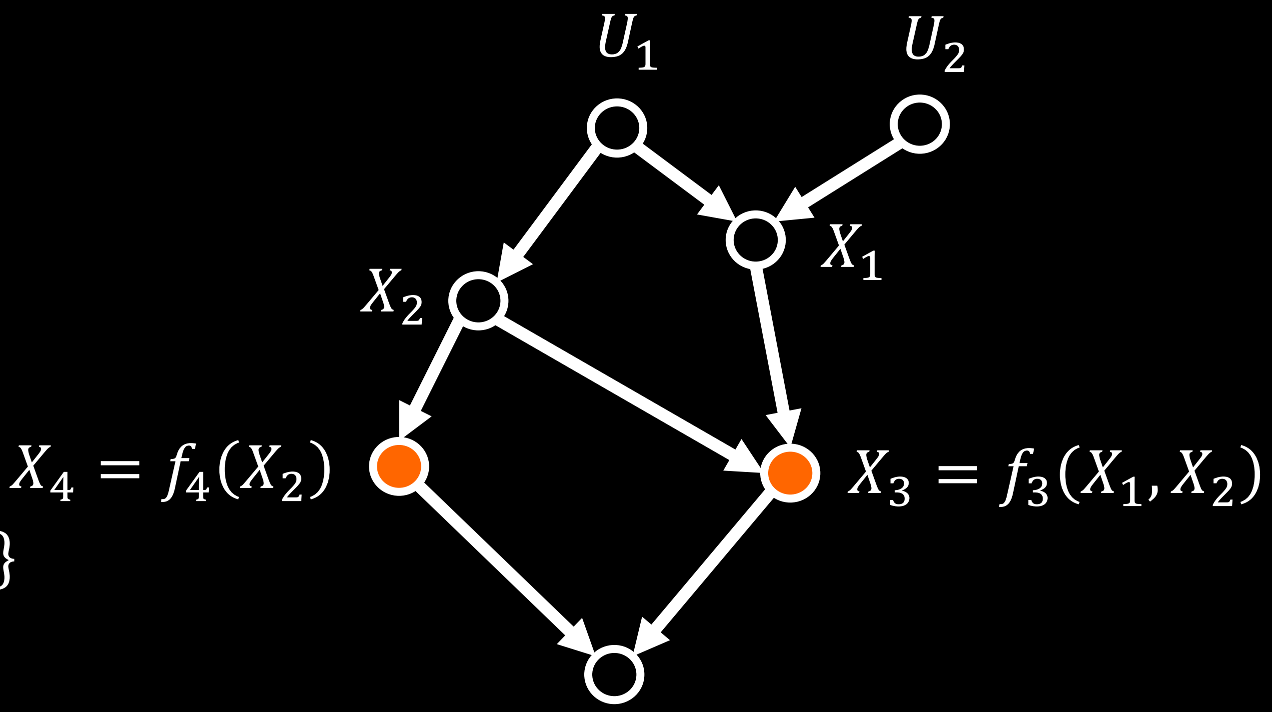
Given a truly complete causal model for, say, math test score in high school juniors, and given complete list of values for every exogenous variable in that model, we could theoretically generate a data point (i.e., a test score), for each individual (student).

$$M = \langle U, V, F \rangle$$

$$U = \{U_1, U_2\}$$

$$V = \{X_1, X_2, X_3, X_4, X_5\}$$

$$F = \{f_1, f_2, f_3, f_4, f_5\}$$



math test score

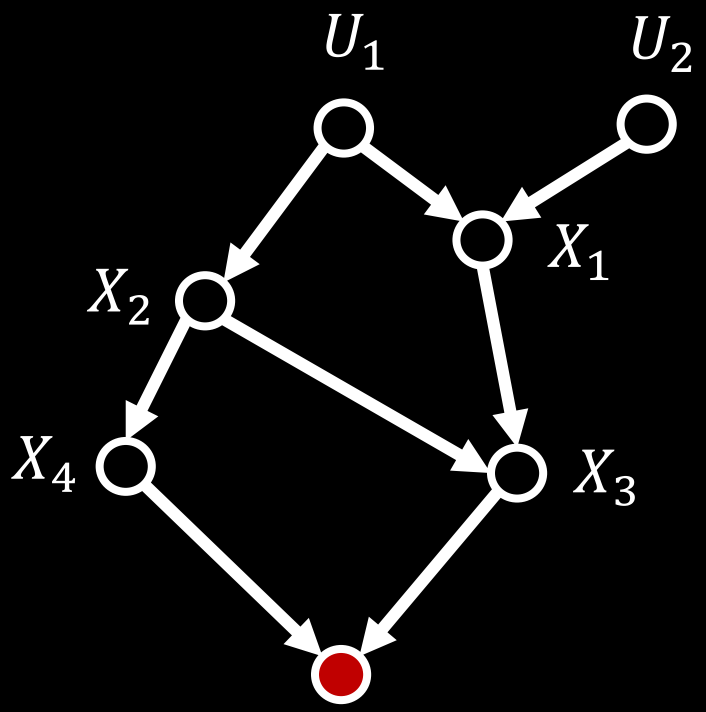
Given a truly complete causal model for, say, math test score in high school juniors, and given complete list of values for every exogenous variable in that model, we could theoretically generate a data point (i.e., a test score), for each individual (student).

$$M = \langle U, V, F \rangle$$

$$U = \{U_1, U_2\}$$

$$V = \{X_1, X_2, X_3, X_4, X_5\}$$

$$F = \{f_1, f_2, f_3, f_4, f_5\}$$



$$X_5 = f_5(X_3, X_4)$$

math test score

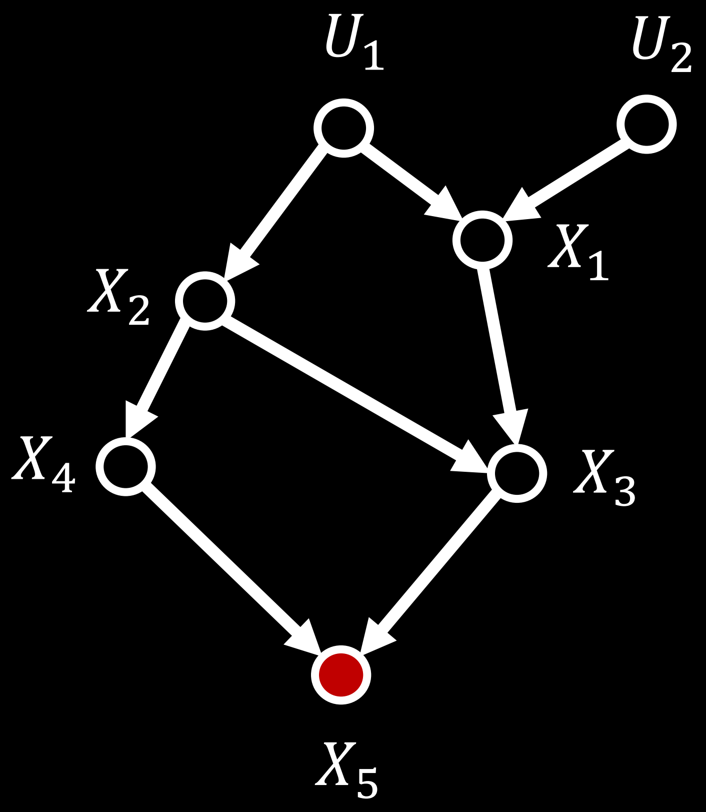
Given a truly complete causal model for, say, math test score in high school juniors, and given complete list of values for every exogenous variable in that model, we could theoretically generate a data point (i.e., a test score), for each individual (student).

$$M = \langle U, V, F \rangle$$

$$U = \{U_1, U_2\}$$

$$V = \{X_1, X_2, X_3, X_4, X_5\}$$

$$F = \{f_1, f_2, f_3, f_4, f_5\}$$



math test score

We can compute test score for each student.

This would necessitate specifying all factors that may have an effect on a student's test score, an unrealistic task.

2.2 CHAIN AND FORKS

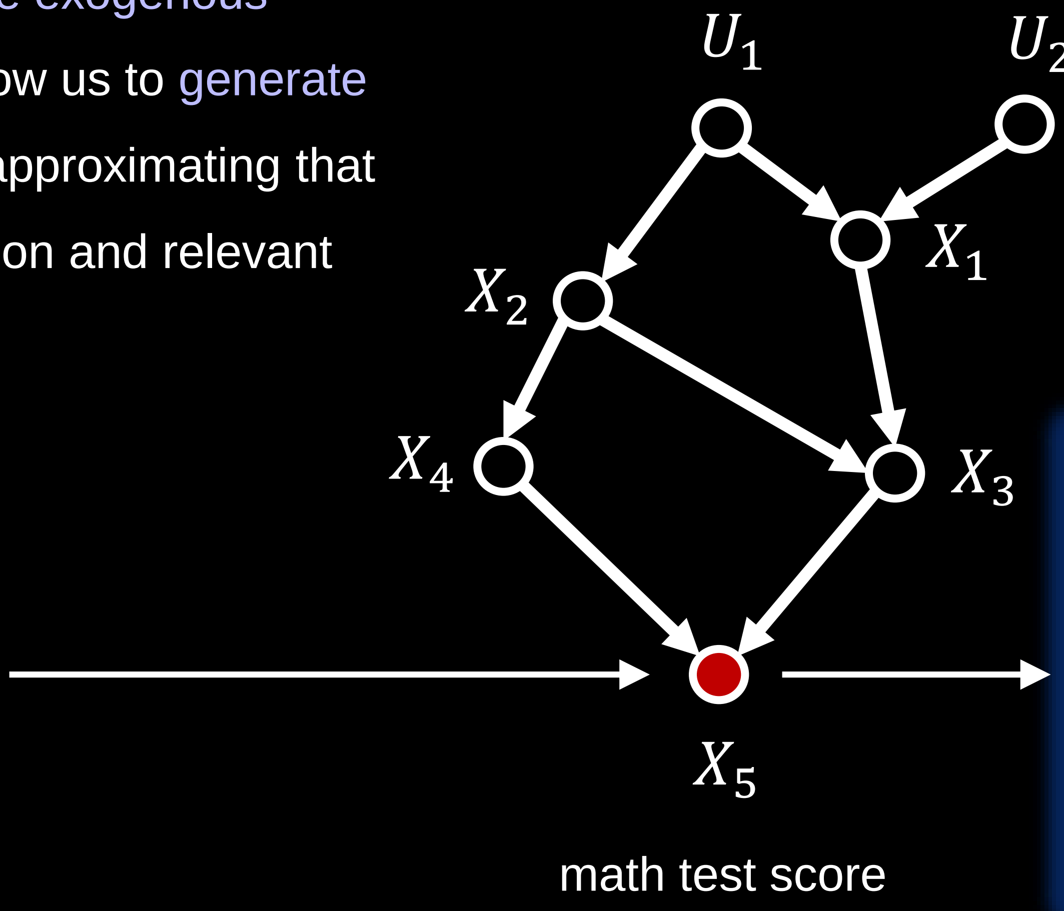
In many cases, we will not have such a precise knowledge about a model *M*.

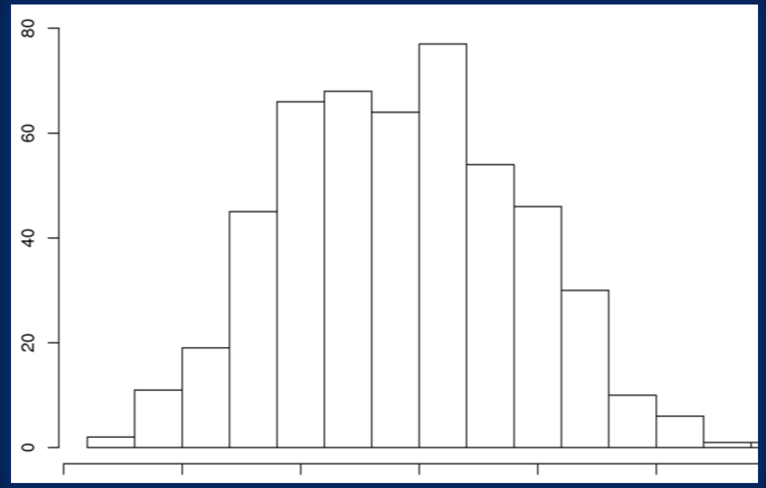
$$M = \langle U, V, F \rangle$$

probability distribution

We might instead have a probability distribution characterizing the exogenous variables *U*, which would allow us to generate a distribution of test scores approximating that of the entire student population and relevant subgroups of students.

For each student we do not get the corresponding math test score X_5 but we get a distribution for its value





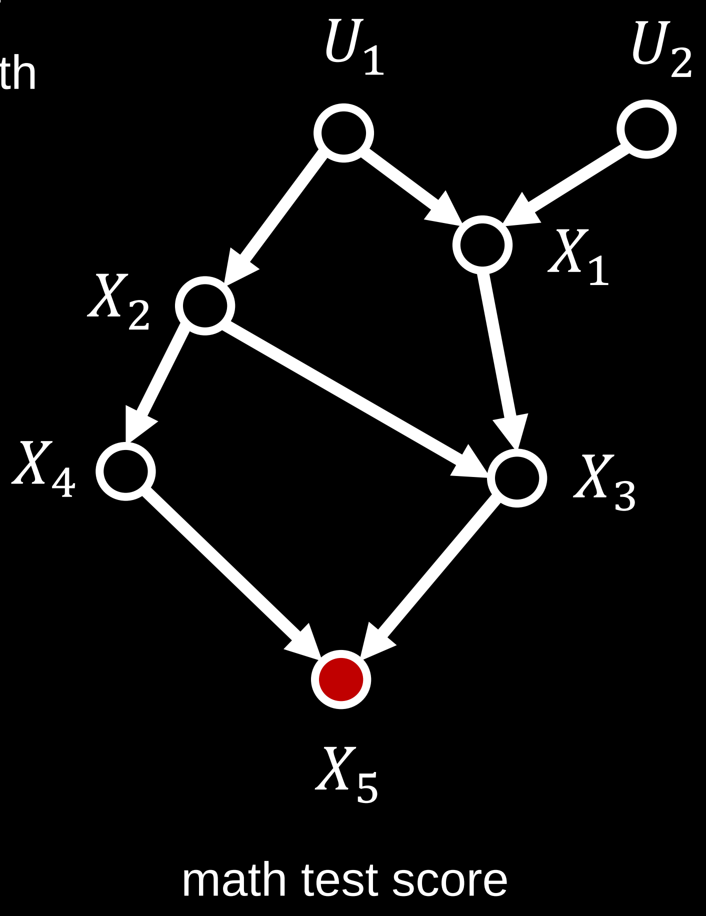
Suppose, however, that we do not have even a probabilistically specified causal model, but only a graphical structure of the model.

We know which variables are caused by which other variables, but we do not know the strength or nature of the relationships.

$$M = \langle U, V, F \rangle \qquad F = ???$$

Unspecified Causal Model

Even with such limited information, we can discern a great deal about the data set generated by the model.

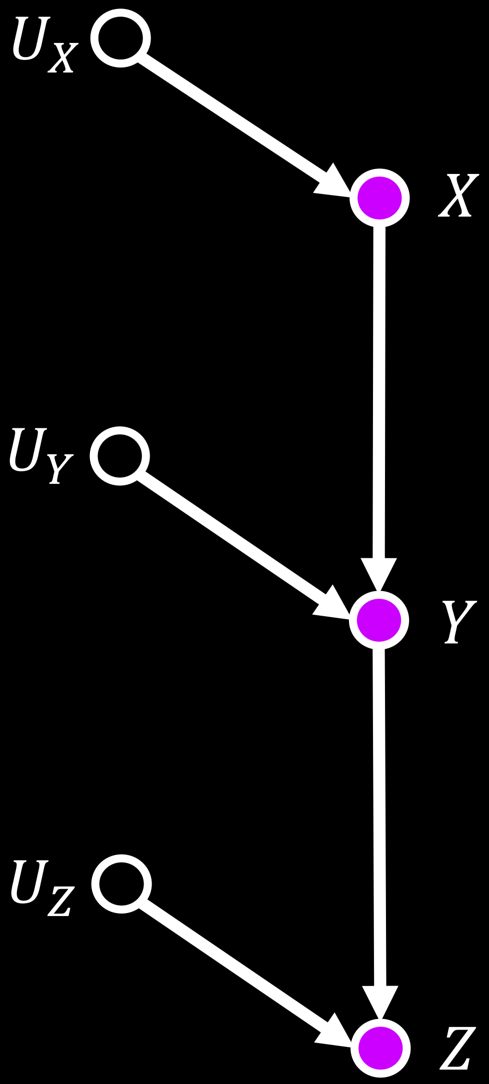


We can learn which variables in the data set are independent of each other and which are independent of each other conditional on other variables.

These independencies will be true of every data set generated by a causal model with that graphical structure, regardless of the specific functions attached to the SCM.

Consider the following three hypothetical SCMs, all share the same graphical model.

SCM 2.2.1 (School Funding, SAT Scores, and College Acceptance)



$f_X: X = U_X$

$f_Y: Y = \frac{x}{3} + U_Y$

$f_Z: Z = \frac{y}{16} + U_Z$

Figure 2.1

X: High School's funding in dollars

Y: Average SAT Score

Z: College acceptance rate

$$M_1 = \langle U, V, F_1 \rangle$$

$$U = \{U_X, U_Y, U_Z\}$$

$$V = \{X, Y, Z\}$$

$$F_1 = \{f_X, f_Y, f_Z\}$$

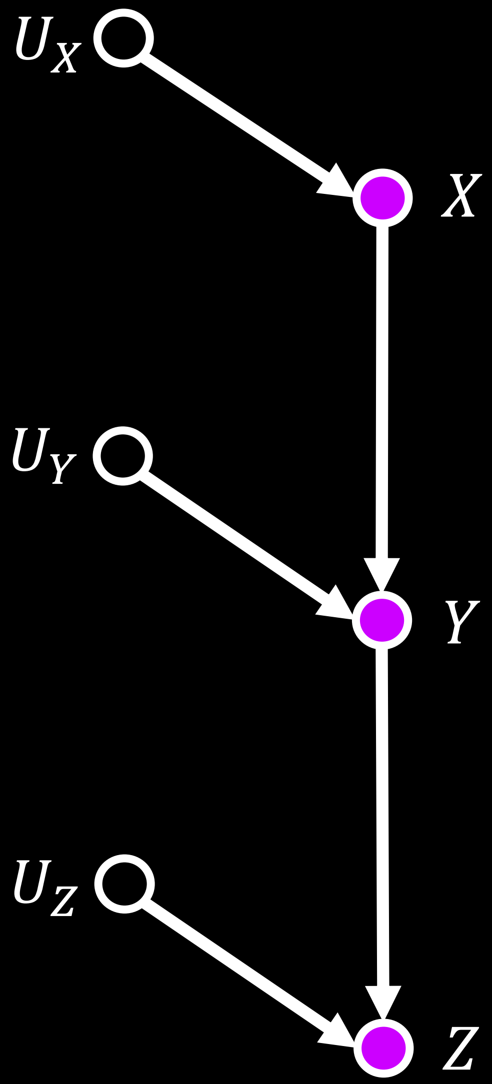


Figure 2.1

SCM 2.2.2 (Switch, Circuit, and Light Bulb)

$$M_2 = \langle U, V, F_2 \rangle$$

$$U = \{U_X, U_Y, U_Z\}$$

$$V = \{X, Y, Z\}$$

$$F_2 = \{f_X, f_Y, f_Z\}$$

X: State of a light switch

Y: State of the associated electrical circuit

Z: State of a light bulb

$$f_X: X = U_X$$

$$f_Y: Y = \begin{cases} closed & \text{IF } (X = up \text{ AND } U_Y = 0) \text{ OR } (X = down \text{ AND } U_Y = 1) \\ open & \text{otherwise} \end{cases}$$

$$U_X = \{up, down\} \qquad X = \{up, down\}$$

$$U_Y = \{0,1\} \qquad Y = \{closed, open\}$$

$$U_Z = \{0,1\} \qquad Z = \{on, off\}$$

$$f_Z: Z = \begin{cases} on & \text{IF } (Y = closed \text{ AND } U_Z = 0) \text{ OR } (Y = open \text{ AND } U_Z = 1) \\ off & \text{otherwise} \end{cases}$$

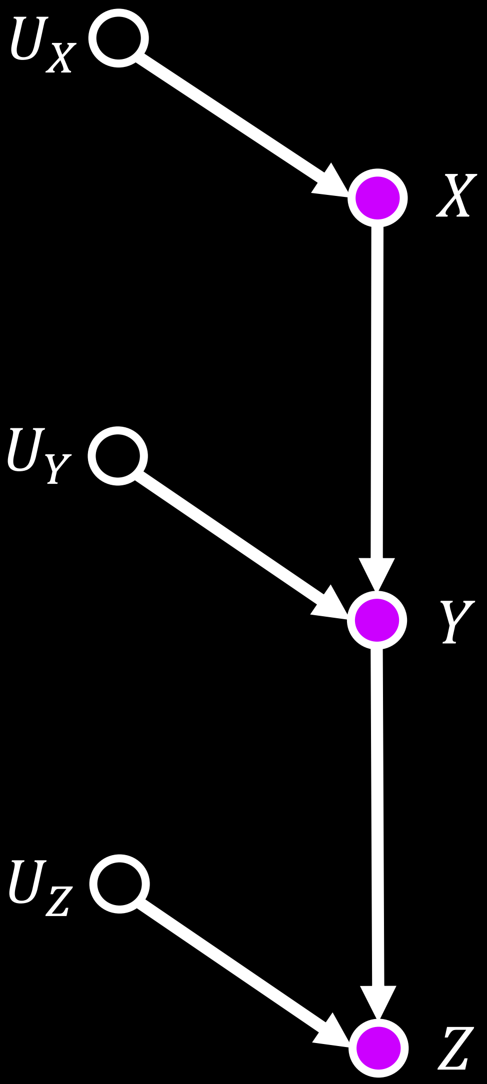


Figure 2.1

$f_X: X = U_X$

$f_Y: Y = 84 - x + U_Y$

$f_Z: Z = \frac{100}{y} + U_Z$

SCM 2.2.3 (Work Hours, Training, and Race Time)

X: Hours of work at jobs each week

Y: Hours of training each week

Z: Completion time of the race

$$M_3 = \langle U, V, F_3 \rangle$$

$$U = \{U_X, U_Y, U_Z\}$$

$$V = \{X, Y, Z\}$$

$$F_3 = \{f_X, f_Y, f_Z\}$$

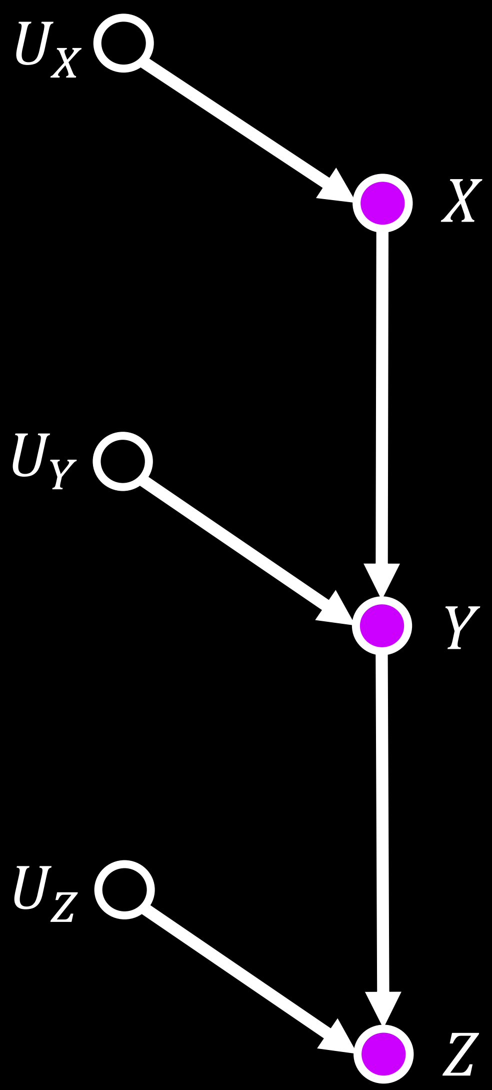


Figure 2.1

$f_X: X = U_X$

$f_Y: Y = \frac{x}{3} + U_Y$

$$f_Y: Y = \begin{cases} closed & \text{IF } (X = up \text{ AND } U_Y = 0) \text{ OR } (X = down \text{ AND } U_Y = 1) \\ open & \text{otherwise} \end{cases}$$

$f_Y: Y = 84 - x + U_Y$

$f_Z: Z = \frac{y}{16} + U_Z$

$$f_Z: Z = \begin{cases} on & \text{IF } (Y = closed \text{ AND } U_Z = 0) \text{ OR } (Y = open \text{ AND } U_Z = 1) \\ off & \text{otherwise} \end{cases}$$

$f_Z: Z = \frac{100}{y} + U_Z$

$M_1 = \langle U, V, F_1 \rangle$

$M_2 = \langle U, V, F_2 \rangle$

$M_3 = \langle U, V, F_3 \rangle$

$U = \{U_X, U_Y, U_Z\}$

$V = \{X, Y, Z\}$

2.2 Chain and Forks

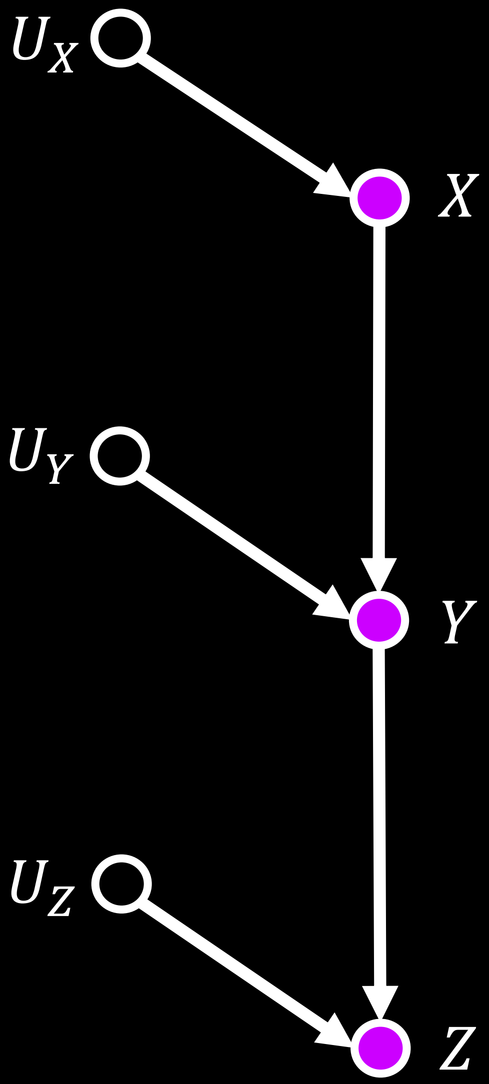


Figure 2.1

$f_X: X = U_X$

Exogenous variables stand for any **unknown** or **random effects** that may alter the relationship between endogenous variables.

$$f_Y: Y = \frac{x}{3} + U_Y$$

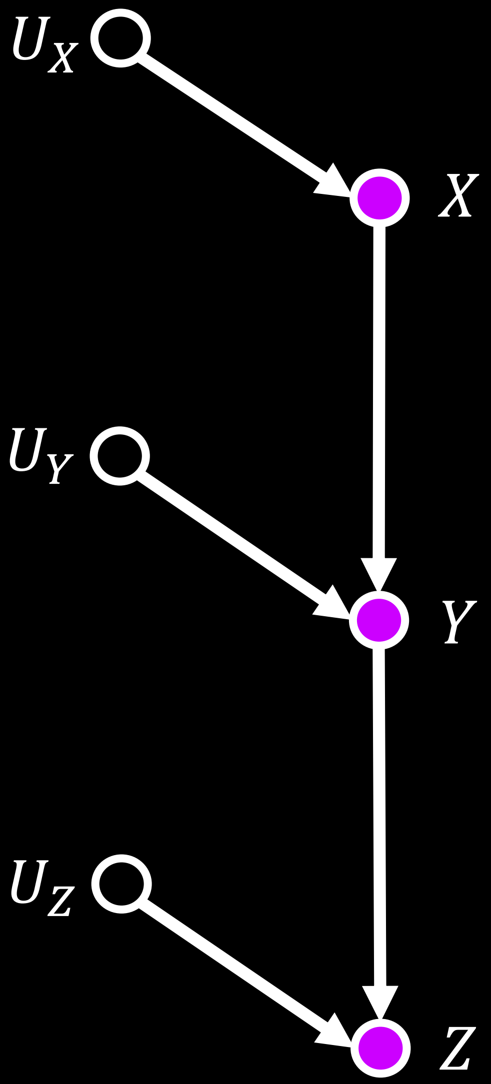
$$f_Y: Y = \begin{cases} closed & \text{IF } (X = up \text{ AND } U_Y = 0) \text{ OR } (X = down \text{ AND } U_Y = 1) \\ open & \text{otherwise} \end{cases}$$

$$f_Y: Y = 84 - x + U_Y$$

$$f_Z: Z = \frac{y}{16} + U_Z$$

$$f_Z: Z = \begin{cases} on & \text{IF } (Y = closed \text{ AND } U_Z = 0) \text{ OR } (Y = open \text{ AND } U_Z = 1) \\ off & \text{otherwise} \end{cases}$$

$$f_Z: Z = \frac{100}{y} + U_Z$$



$f_X: X = U_X$

$f_Y: Y = \frac{x}{3} + U_Y$

$f_Y: Y = 84 - x + U_Y$

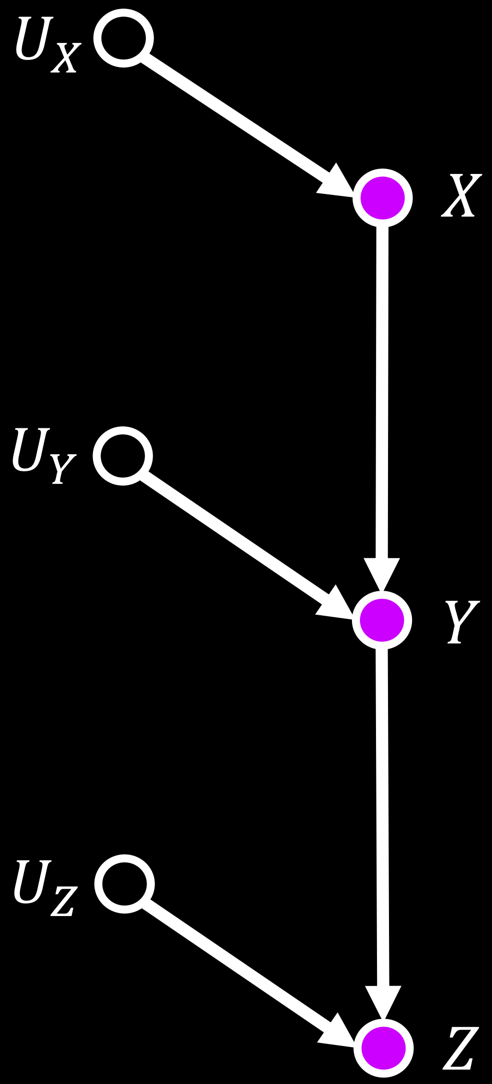
$f_Z: Z = \frac{y}{16} + U_Z$

$f_Z: Z = \frac{100}{y} + U_Z$

Figure 2.1

U_Y and U_Z are additive factors that account for variations among individuals.

2.2 CHAIN AND FORKS



$f_X: X = U_X$

U_Y and U_Z take value 1 if there is some unobserved anomaly, and 0 if there is none.

$$f_Y: Y = \begin{cases} closed & \text{IF } (X = up \text{ AND } U_Y = 0) \text{ OR } (X = down \text{ AND } U_Y = 1) \\ open & \text{otherwise} \end{cases}$$

$$f_Z: Z = \begin{cases} on & \text{IF } (Y = closed \text{ AND } U_Z = 0) \text{ OR } (Y = open \text{ AND } U_Z = 1) \\ off & \text{otherwise} \end{cases}$$

Figure 2.1

2.2 Chain and Forks

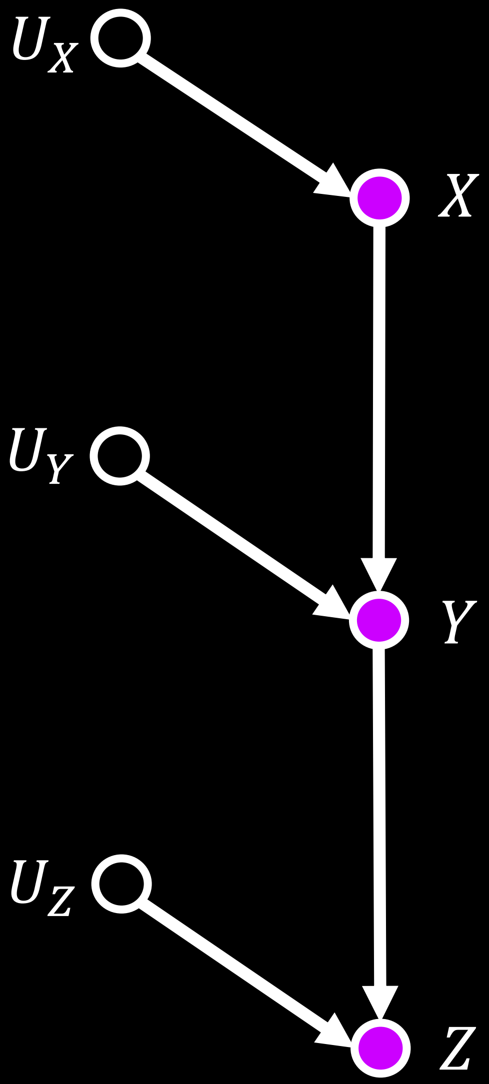


Figure 2.1

$$f_X: X = U_X$$

$$f_Y: Y = \frac{x}{3} + U_Y$$

$$f_Y: Y = 84 - x + U_Y$$

$$f_Z: Z = \frac{y}{16} + U_Z$$

$$f_Z: Z = \frac{100}{y} + U_Z$$

M_1 and M_3 deal with continuous variables.

M_2 deals with categorical variables.

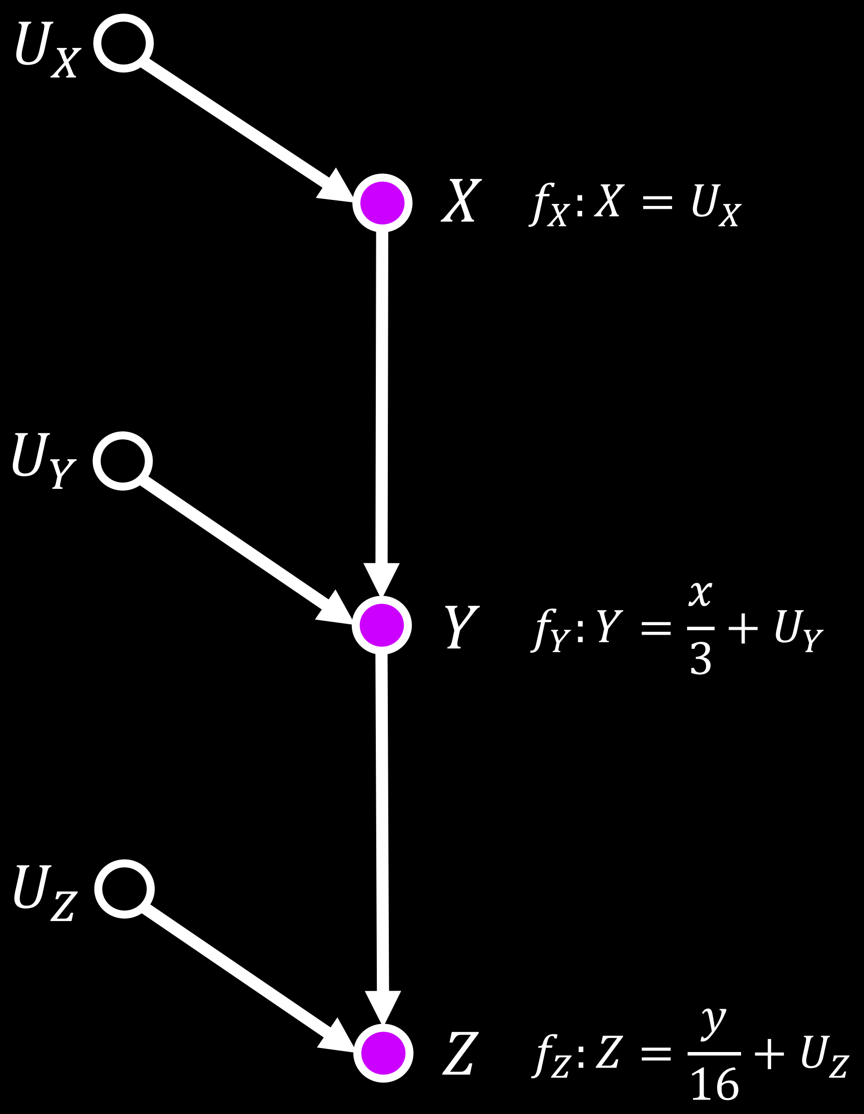


Figure 2.1

In 2.2.1, the relationships between variables are all positive, i.e.,

- the higher the value of the parent variable, the higher the values of the child variable.

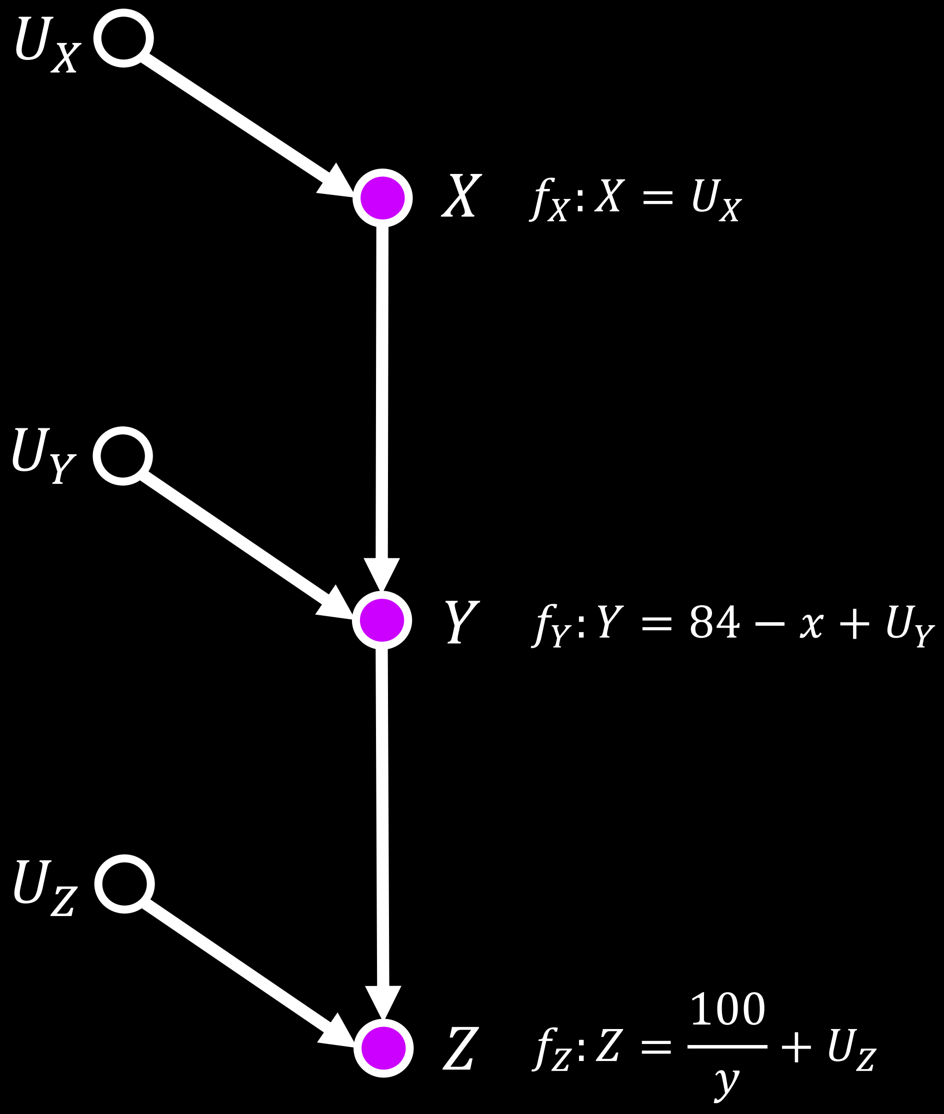
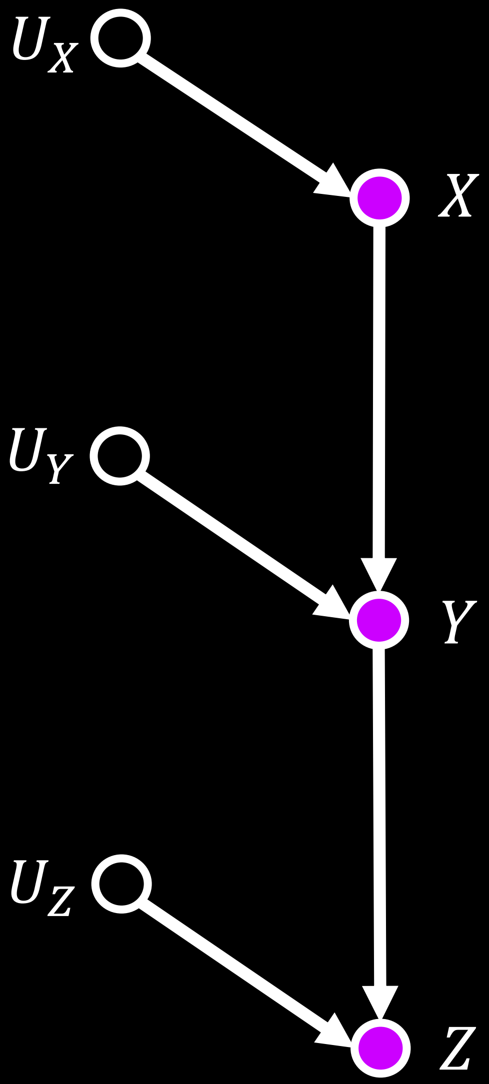


Figure 2.1

In 2.2.3, for variables *Y* and *Z*, the correlation between them and their parents are all negative, i.e.,

- the higher the value of the parent variable, the lower the value of the child variable.

2.2 Chain and Forks



$f_X: X = U_X$

In 2.2.2, the correlations between the variables are not linear at all, but logical.

$$f_Y: Y = \begin{cases} closed & \text{IF } (X = up \text{ AND } U_Y = 0) \text{ OR } (X = down \text{ AND } U_Y = 1) \\ open & \text{otherwise} \end{cases}$$

$$f_Z: Z = \begin{cases} on & \text{IF } (Y = closed \text{ AND } U_Z = 0) \text{ OR } (Y = open \text{ AND } U_Z = 1) \\ off & \text{otherwise} \end{cases}$$

Figure 2.1

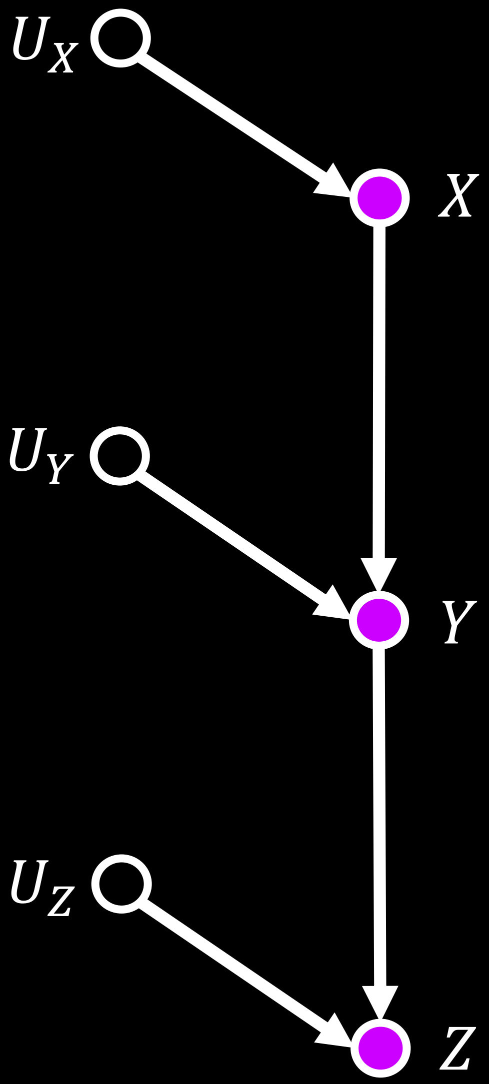


Figure 2.1

The three SCMs share no functions, except for f_X, but they share the same graphical structure.

The data sets generated by all three SCMs must share certain independencies, and we can predict those independencies simply by examining the graphical model to the left.

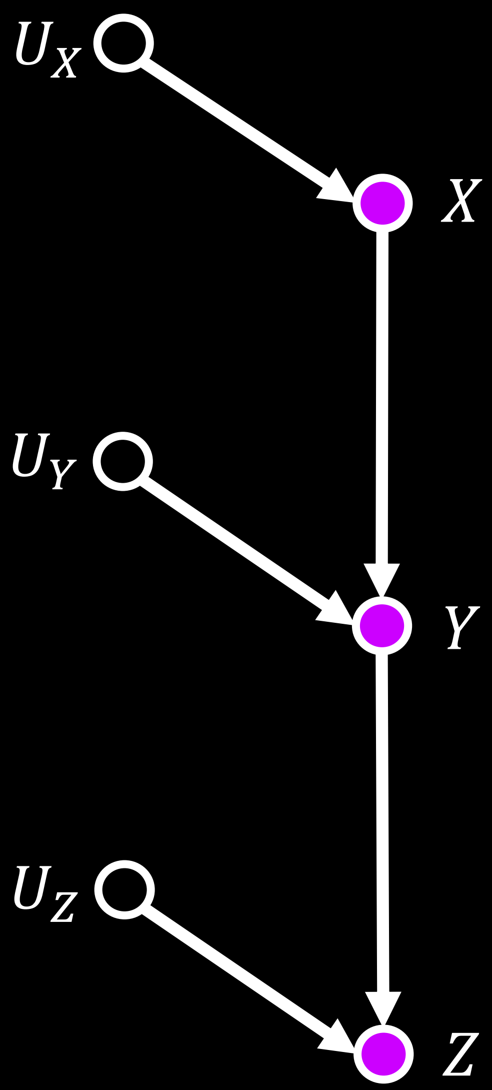


Figure 2.1

The independencies shared by data sets generated by these three SCMs, and the dependencies that are likely shared by all such SCMs are the following:

1. ***Z* and *Y* are likely dependent**, i.e., for some pair of values *z*, *y*

$$P(Z = z|Y = y) \neq P(Z = z)$$

2. ***Y* and *X* are likely dependent**, i.e., for some pair of values *y*, *x*

$$P(Y = y|X = x) \neq P(Y = y)$$

3. ***Z* and *X* are likely dependent**, i.e., for some pair of values *z*, *x*

$$P(Z = z|X = x) \neq P(Z = z)$$

4. ***Z* and *X* are independent, conditional on *Y***, i.e., for all values *x*, *y*, *z*

$$P(Z = z|X = x, Y = y) = P(Z = z|Y = y)$$

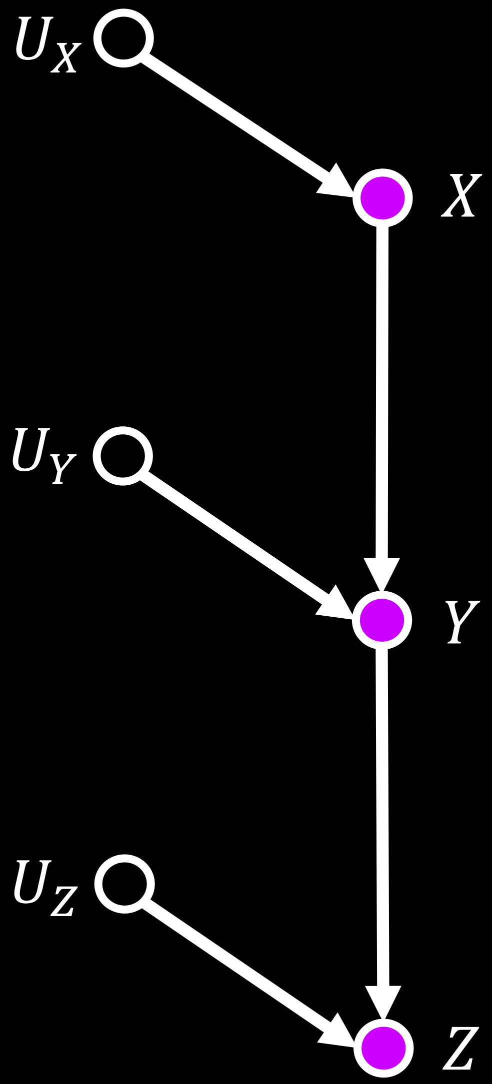


Figure 2.1

To understand why these independencies and dependencies hold, let's examine the graphical model.

- Any two variables with an edge between them are likely dependent.

  An arrow from one variable to another indicates that the first variable causes the second, that is, the value of the first variable is part of the function that determines the value of the second variable.

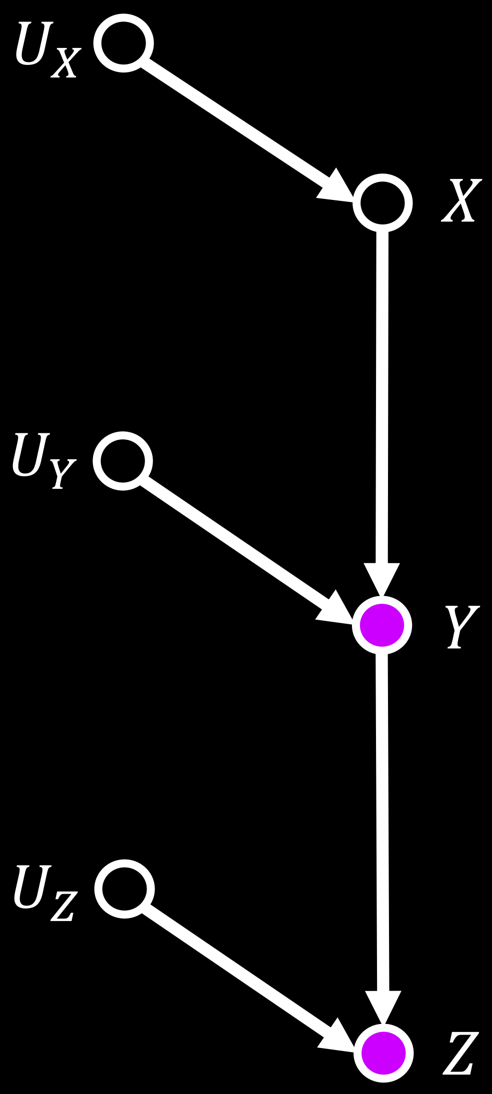


Figure 2.1

To understand why these independencies and dependencies hold, let's examine the graphical model.

- Any two variables with an edge between them are likely dependent.

An arrow from one variable to another indicates that the first variable causes the second, that is, the value of the first variable is part of the function that determines the value of the second variable.

Y causes *Z* → *Z* depends on *Y*, there is some case in which changing the value of *Y* changes the value of *Z*.

$$f_Z: Z = \frac{y}{16} + U_Z$$

$$f_Z: Z = \begin{cases} on & \text{IF } (Y = closed \text{ AND } U_Z = 0) \text{ OR } (Y = open \text{ AND } U_Z = 1) \\ off & \text{otherwise} \end{cases}$$

$$f_Z: Z = \frac{100}{y} + U_Z$$

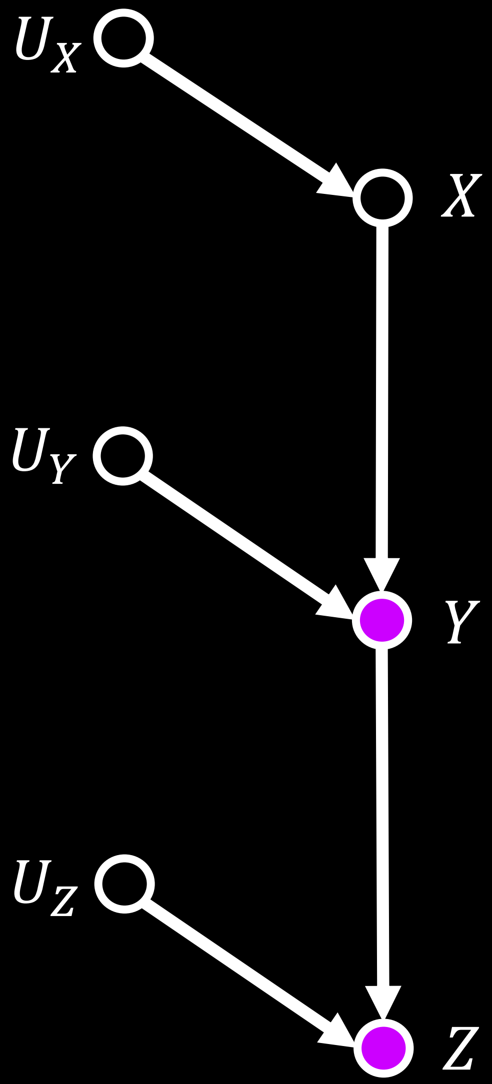


Figure 2.1

To understand why these independencies and dependencies hold, let's examine the graphical model.

- Any two variables with an edge between them are likely dependent.

When we examine those variables in the data set, the probability that one variable takes a given value will change, given that we know the value of the other variable.

The probability that *Z* takes value *z*, when we know that the value of the variable *Y* is equal to *y*, is different from the probability that *Z* takes the value *z*, when we do not know which value *Y* takes, i.e.,

$$P(Z = z|Y = y\,) \neq P(Z = z)$$

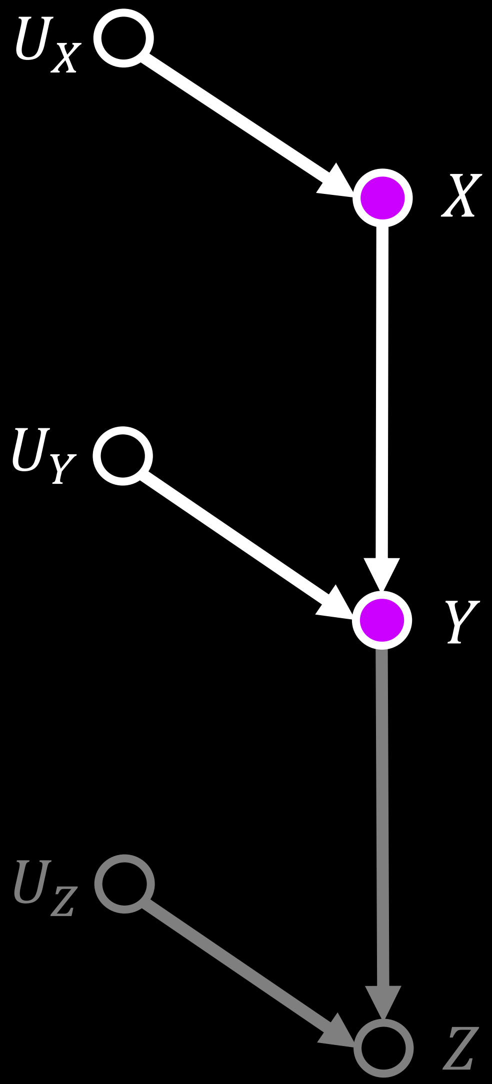


Figure 2.1

Therefore, we can conclude that in a typical causal model, regardless of the specific functions, two variables connected by an edge are likely dependent.

- X and Y are likely dependent

$$f_Y: Y = \frac{x}{3} + U_Y$$

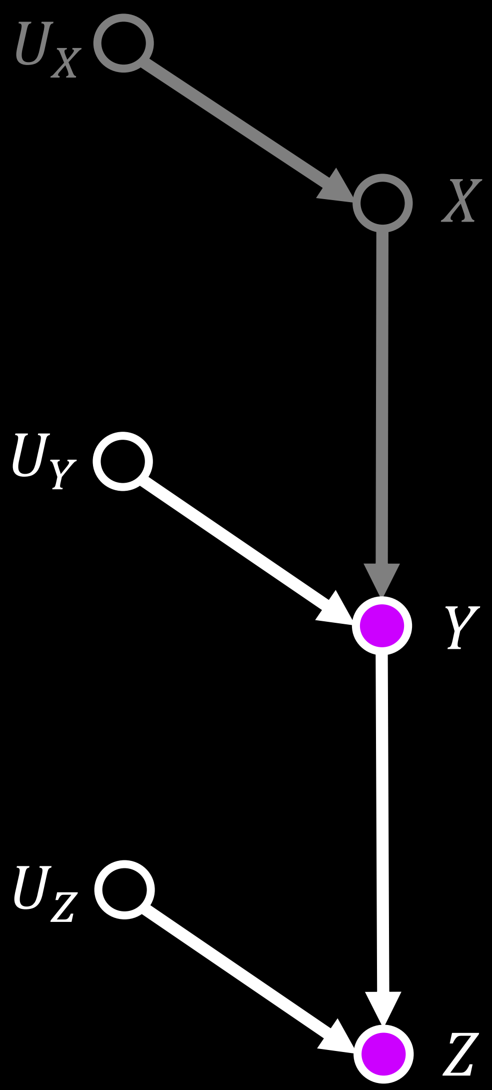


Figure 2.1

Therefore, we can conclude that in a typical causal model, regardless of the specific functions, two variables connected by an edge are likely dependent.

- X and Y are likely dependent $\quad f_Y: Y = \frac{x}{3} + U_Y$
- Y and Z are likely dependent $\quad f_Z: Z = \frac{y}{16} + U_Z$

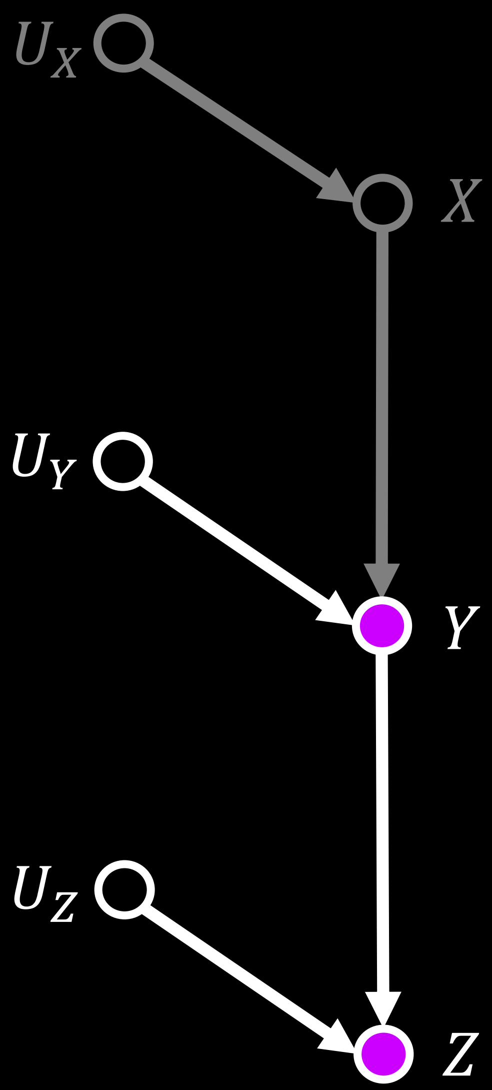


Figure 2.1

Therefore, we can conclude that in a typical causal model, regardless of the specific functions, two variables connected by an edge are likely dependent.

- X and Y are likely dependent $\quad f_Y: Y = \frac{x}{3} + U_Y$
- Y and Z are likely dependent $\quad f_Z: Z = \frac{y}{16} + U_Z$

Why in general we say **likely dependent** and not simply **dependent**?

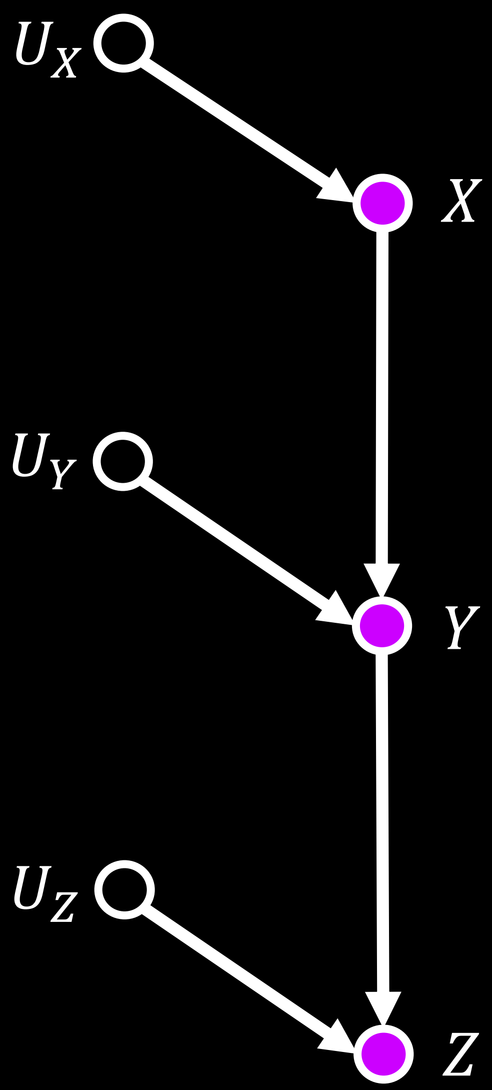


Figure 2.1

Therefore, we can conclude that in a typical causal model, regardless of the specific functions, two variables connected by an edge are likely dependent.

- X and Y are likely dependent $\qquad f_Y: Y = \frac{x}{3} + U_Y$
- Y and Z are likely dependent $\qquad f_Z: Z = \frac{y}{16} + U_Z$

Why in general we say **likely dependent** and not simply **dependent**?

Consider X and U_Y be fair coins, and let $Y = 1$ if and only if $X = U_Y$, then

$$P(Y = 1|X = 1) = P(Y = 1|X = 0) = P(Y = 1) = \frac{1}{2}.$$

Such pathological cases require precise numerical probabilities to achieve independence

$$P(X = 1) = P(U_X = 1) = \frac{1}{2}$$

they are rare, and can be ignored for all practical purposes.

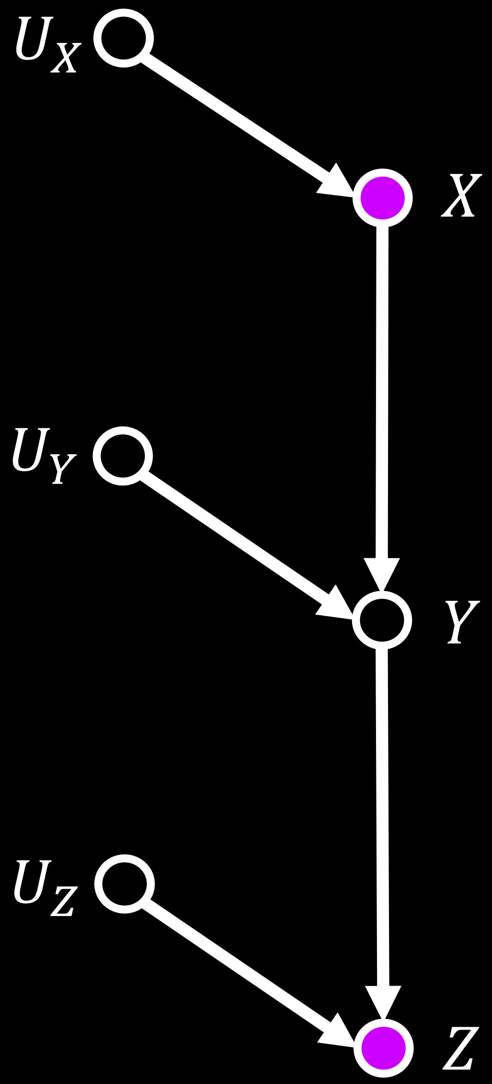


Figure 2.1

Furthermore, in a typical causal model, regardless of the specific functions, two variables connected by a directed path are likely dependent.

$$f_Y: Y = \frac{x}{3} + U_Y$$

$$f_Z: Z = \frac{y}{16} + U_Z = \frac{\frac{x}{3} + U_Y}{16} + U_Z$$

- X and Z are likely dependent

There are pathological cases in which the above is not true!!!

This the reason why we say likely dependent and not just dependent.

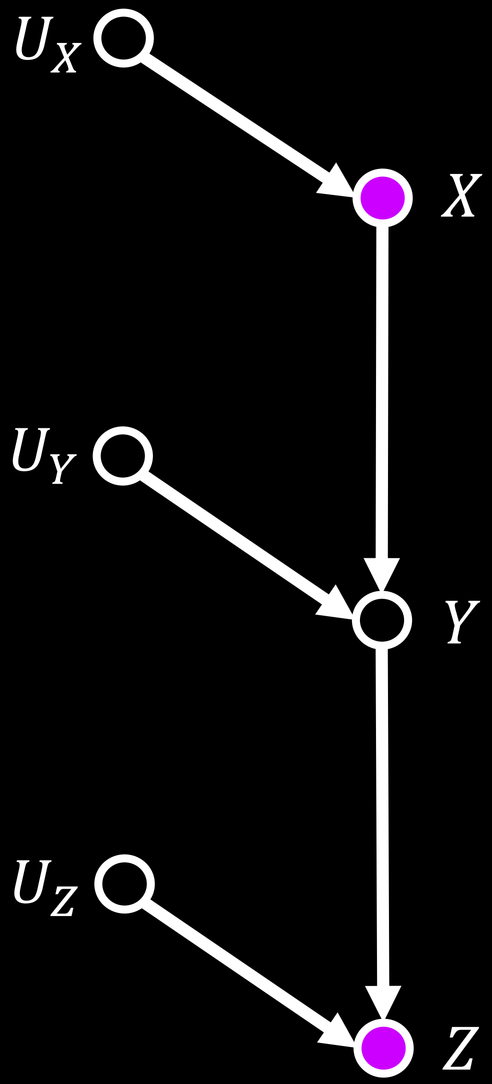


Figure 2.1

SCM 2.2.4 (Pathological Case of Intransitive Dependence)

$$M_4 = \langle U, V, F_4 \rangle$$

$$U = \{U_X, U_Y, U_Z\}$$

$$V = \{X, Y, Z\}$$

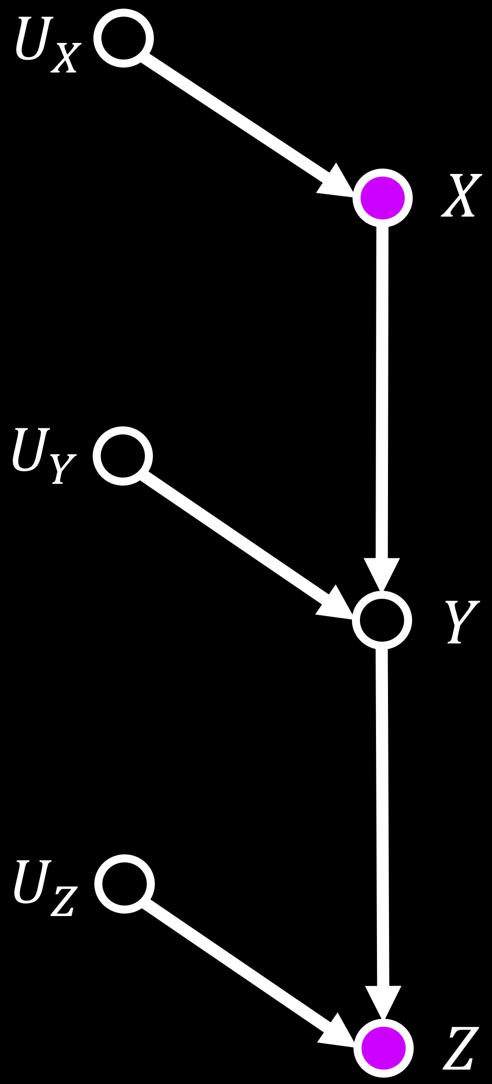
$$F_4 = \{f_X, f_Y, f_Z\}$$

$$U_X = \{1,2\} \qquad X = \{1,2\}$$

$$U_Y = \{1,2\} \qquad Y = \{a, b, c\}$$

$$U_Z = \{1,2\} \qquad Z = \{i, j\}$$

SCM 2.2.4 (Pathological Case of Intransitive Dependence)



$$f_X: X = U_X$$

$$f_Y: Y = \begin{cases} a & \text{IF } (X = 1 \text{ AND } U_Y = 1) \\ b & \text{IF } (X = 2 \text{ AND } U_Y = 1) \\ c & \text{IF } U_Y = 2 \end{cases}$$

$$f_Z: Z = \begin{cases} i & \text{IF } (Y = c \; OR \; U_Z = 1) \\ j & \text{IF } (Y \neq c \; AND \; U_Z = 2) \end{cases}$$

Figure 2.1

$$M_4 = \langle U, V, F_4 \rangle$$

$$U = \{U_X, U_Y, U_Z\}$$

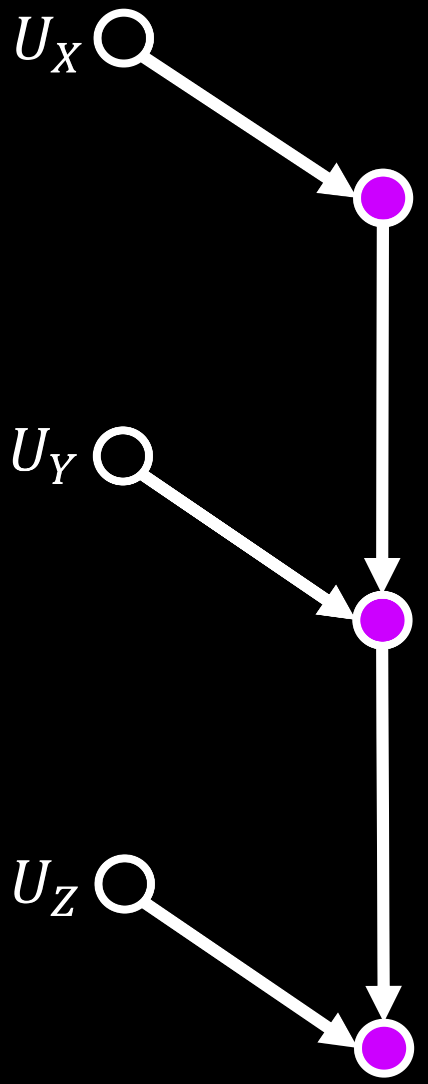
$$V = \{X, Y, Z\}$$

$$F_4 = \{f_X, f_Y, f_Z\}$$

$$U_X = \{1,2\} \qquad X = \{1,2\}$$

$$U_Y = \{1,2\} \qquad Y = \{a, b, c\}$$

$$U_Z = \{1,2\} \qquad Z = \{i, j\}$$



$$f_X: X = U_X$$

$$f_Y: Y = \begin{cases} a & \text{IF } (X = 1 \text{ AND } U_Y = 1) \\ b & \text{IF } (X = 2 \text{ AND } U_Y = 1) \\ c & \text{IF } U_Y = 2 \end{cases}$$

$$f_Z: Z = \begin{cases} i & \text{IF } (Y = c \ OR \ U_Z = 1) \\ j & \text{IF } (Y \neq c \ AND \ U_Z = 2) \end{cases}$$

Figure 2.1

SCM 2.2.4 (Pathological Case of Intransitive Dependence)

In this case, no matter what value U_Y and U_Z take, *X* will have no effect on the value that *Z* takes;

- changes in *X* account for variation in *Y* between *a* and *b*, but *Y* does not affect Z unless it takes value *c*.

Therefore, *X* and *Z* vary independently in this model.

Intransitive Case

By now it should be clear why previously we talked about **likely dependent** and not about dependent.

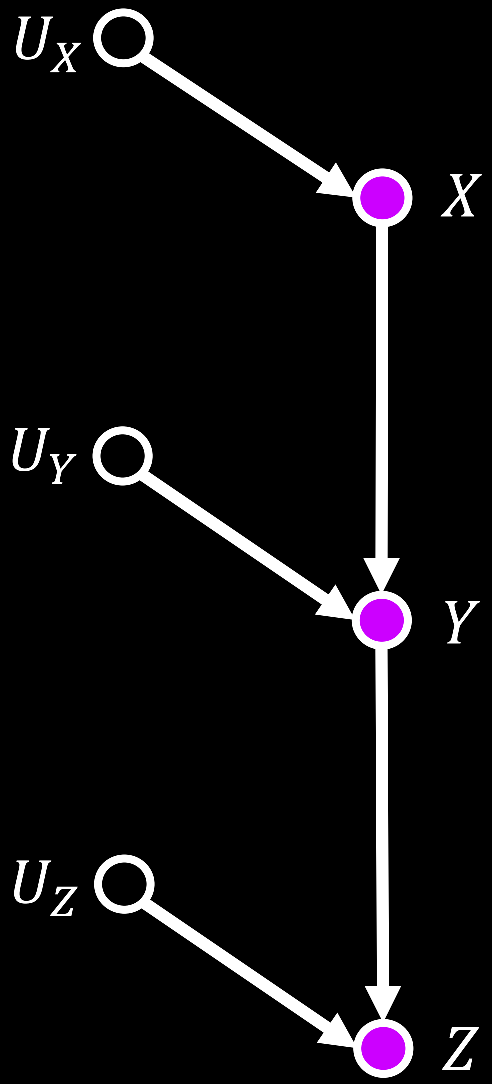


Figure 2.1

To summarize we have

- X and Y are likely dependent $\quad f_Y: Y = \frac{x}{3} + U_Y$
- Y and Z are likely dependent $\quad f_Z: Z = \frac{y}{16} + U_Z$
- X and Z are likely dependent $\quad f_Z: Z = \frac{y}{16} + U_Z = \frac{\frac{x}{3} + U_Y}{16} + U_Z$

2.2 Chain and Forks

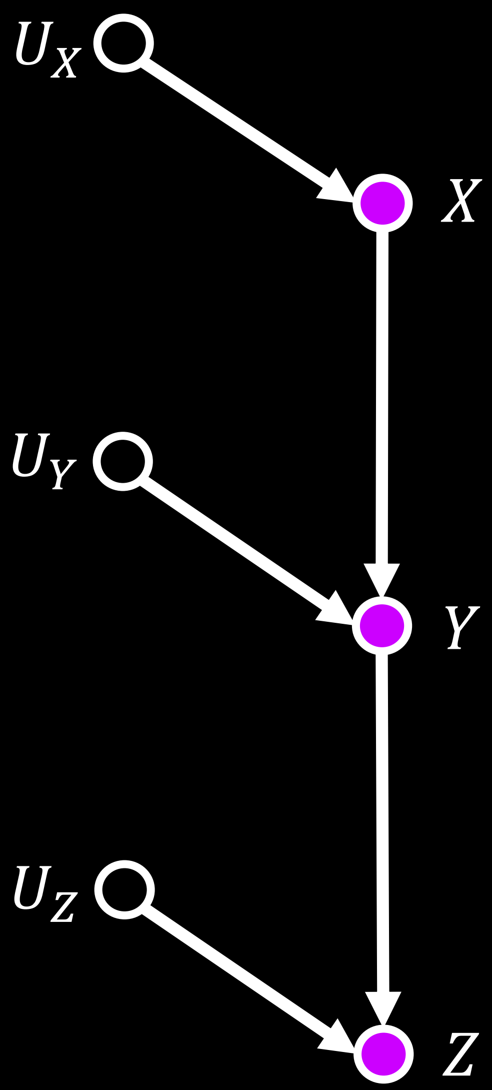


Figure 2.1

Now, let's consider the following point

- ***Z* and *X* are independent, conditionally on *Y***, i.e., for all values *x, y, z*

$$P(Z = z|X = x, Y = y) = P(Z = z|Y = y) \qquad Z \perp X \,|Y$$

Remember that when we condition on *Y*, we filter the data into groups based on the value of *Y*.

U_x	U_y	U_z	X	Y	Z
1	0	0	1	1	1
1	0	0	1	1	1
1	0	1	1	1	2
2	0	0	2	2	2
1	-1	0	1	2	2
1	-1	0	1	2	2
2	1	0	2	1	1
2	1	0	2	1	1
2	1	1	2	1	2
2	0	0	2	2	2

Table filtering by Y = 1

U_x	U_y	U_z	X	Y	Z
1	0	0	1	1	1
1	0	0	1	1	1
1	0	1	1	1	2
2	1	0	2	1	1
2	1	0	2	1	1
2	1	1	2	1	2

Table filtering by Y = 2

U_x	U_y	U_z	X	Y	Z
2	0	0	2	2	2
1	-1	0	1	2	2
1	-1	0	1	2	2
2	0	0	2	2	2

$$X = U_X; \qquad Y = X - U_Y; \qquad Z = Y + U_Z;$$

2.2 Chain and Forks

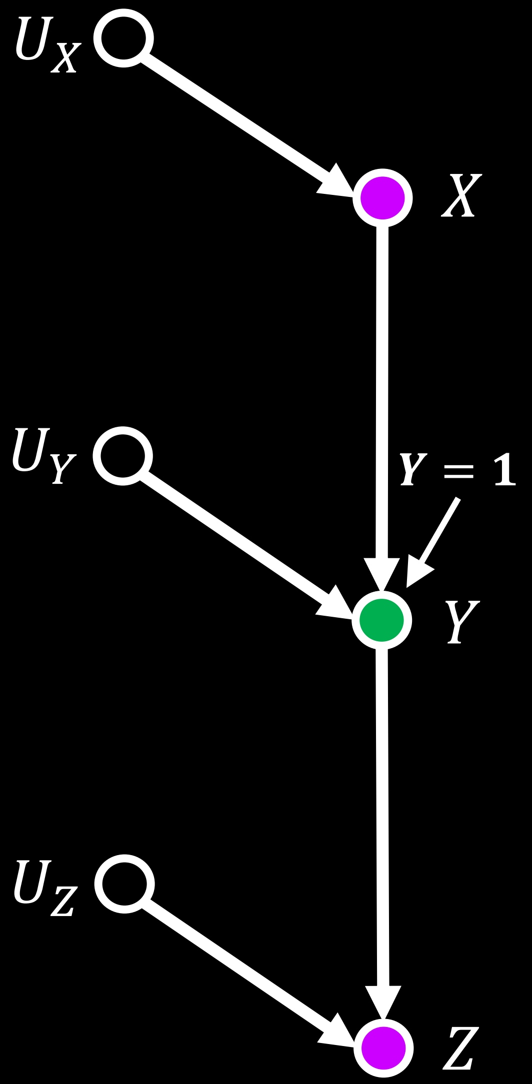


Figure 2.1

We compare *X* and *Z* on all the cases where *Y* = 1, and on all the cases where *Y* = 2.

Let's assume that we are looking at the cases where *Y* = 1.

We want to know whether, in these cases only, the value of *Z* is independent of the value of *X*.

Previously, we determined that *X* and *Z* are likely dependent, because when the value of *X* changes, the value of *Y* likely changes, and when the value of *Y* changes, the value of *Z* is likely to change.

U_x	U_y	U_z	X	Y	Z
1	0	0	1	1	1
1	0	0	1	1	1
1	0	1	1	1	2
2	1	0	2	1	1
2	1	0	2	1	1
2	1	1	2	1	2

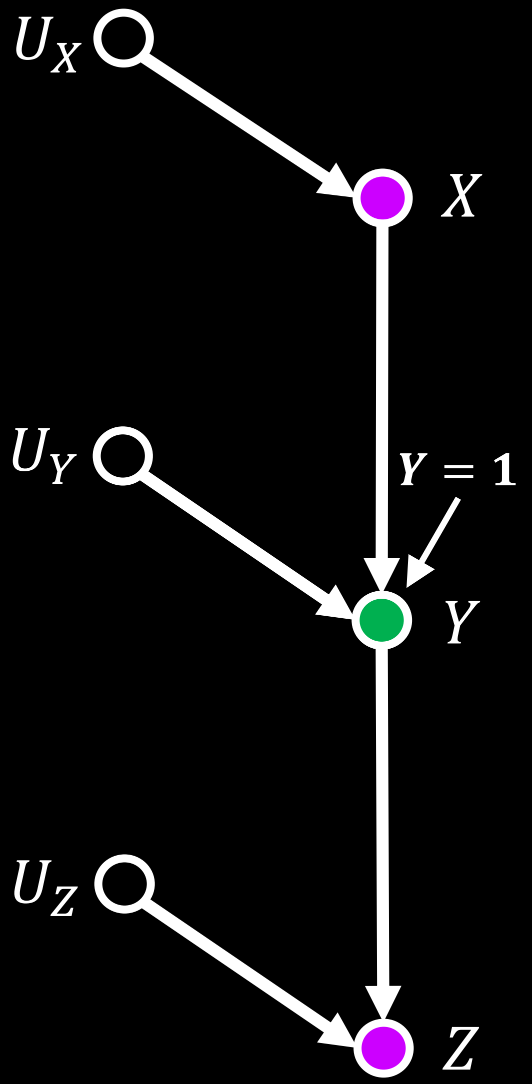


Figure 2.1

We compare *X* and *Z* on all the cases where *Y* = 1, and on all the cases where *Y* = 2.

Let's assume that we are looking at the cases where *Y* = 1.

We want to know whether, in these cases only, the value of *Z* is independent of the value of *X*.

$$P(Z = z|X = x, Y = y) = P(Z = z|Y = y), \qquad \forall\, x, z \quad Y = 1$$

U_x	U_y	U_z	*X*	*Y*	*Z*
1	0	0	**1**	**1**	**1**
1	0	0	**1**	**1**	**1**
1	0	1	**1**	**1**	**2**
2	1	0	**2**	**1**	**1**
2	1	0	**2**	**1**	**1**
2	1	1	**2**	**1**	**2**

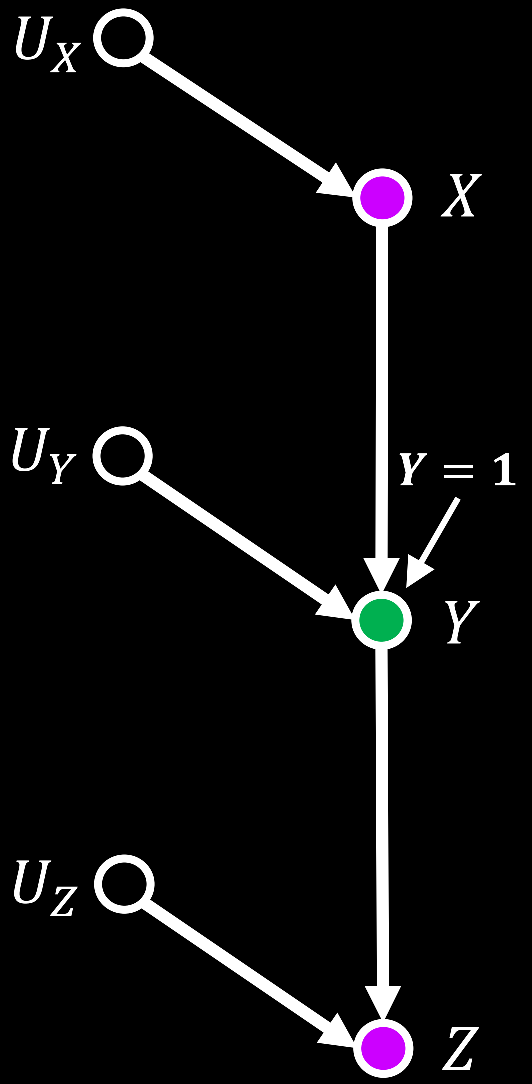


Figure 2.1

We compare *X* and *Z* on all the cases where *Y* = 1, and on all the cases where *Y* = 2.

Let's assume that we are looking at the cases where *Y* = 1.

We want to know whether, in these cases only, the value of *Z* is independent of the value of *X*.

$$P(Z = z|X = x, Y = y) = P(Z = z|Y = y), \qquad \forall\, x, z \quad Y = 1$$

$$P(Z = 1|X = 1, Y = 1) = \frac{2}{3}$$

$$P(Z = 2|X = 1, Y = 1) = \frac{1}{3}$$

$X = 1, Y = 1$

U_x	U_y	U_z	X	Y	Z
1	0	0	**1**	**1**	**1**
1	0	0	**1**	**1**	**1**
1	0	1	**1**	**1**	**2**
2	1	0	**2**	**1**	**1**
2	1	0	**2**	**1**	**1**
2	1	1	**2**	**1**	**2**

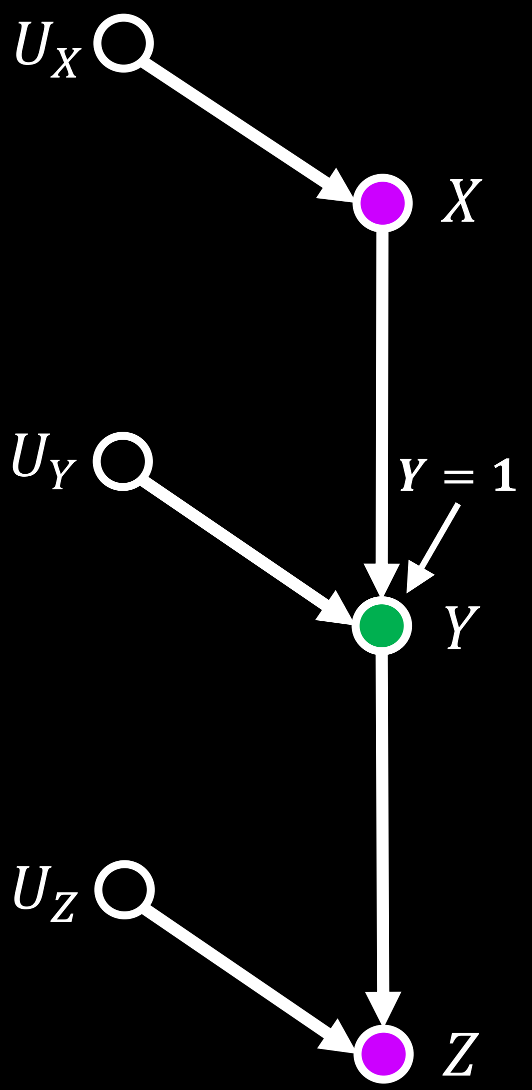


Figure 2.1

We compare *X* and *Z* on all the cases where *Y* = 1, and on all the cases where *Y* = 2.

Let's assume that we are looking at the cases where *Y* = 1.

We want to know whether, in these cases only, the value of *Z* is independent of the value of *X*.

$$P(Z = z|X = x, Y = y) = P(Z = z|Y = y), \qquad \forall\, x, z \quad Y = 1$$

$$P(Z = 1|X = 1, Y = 1) = \frac{2}{3} \qquad P(Z = 1|X = 2, Y = 1) = \frac{2}{3}$$

$$P(Z = 2|X = 1, Y = 1) = \frac{1}{3} \qquad P(Z = 2|X = 2, Y = 1) = \frac{1}{3}$$

$X = 2, Y = 1$

U_x	U_y	U_z	**X**	**Y**	**Z**
1	0	0	**1**	**1**	**1**
1	0	0	**1**	**1**	**1**
1	0	1	**1**	**1**	**2**
2	1	0	**2**	**1**	**1**
2	1	0	**2**	**1**	**1**
2	1	1	**2**	**1**	**2**

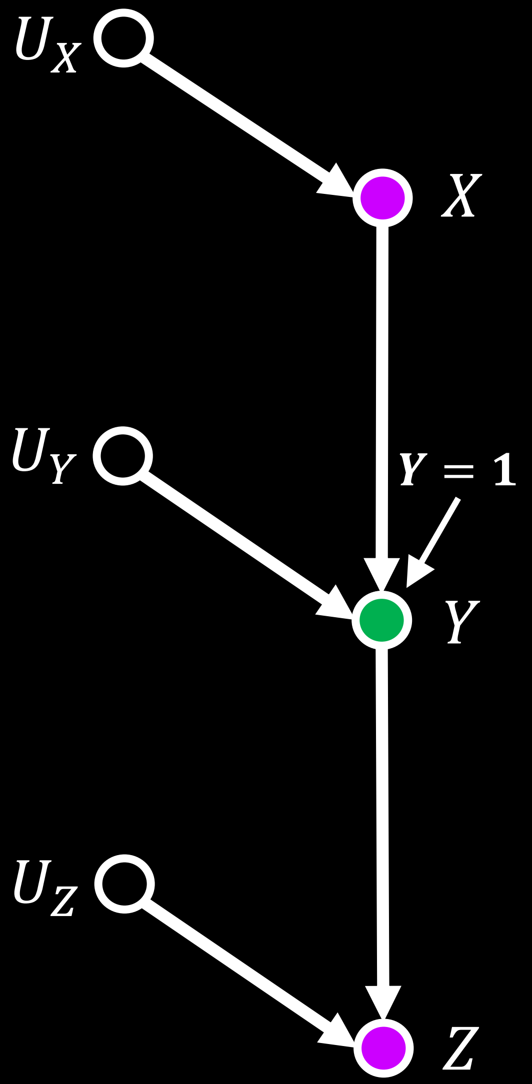


Figure 2.1

We compare X and Z on all the cases where $Y = 1$, and on all the cases where $Y = 2$.

Let's assume that we are looking at the cases where $Y = 1$.

We want to know whether, in these cases only, the value of Z is independent of the value of X.

$$P(Z = z|X = x, Y = y) = P(Z = z|Y = y), \qquad \forall\, x, z \quad Y = 1$$

$$P(Z = 1|X = 1, Y = 1) = \frac{2}{3} \qquad P(Z = 1|X = 2, Y = 1) = \frac{2}{3} \qquad P(Z = 1|Y = 1) = \frac{4}{6} = \frac{2}{3}$$

$$P(Z = 2|X = 1, Y = 1) = \frac{1}{3} \qquad P(Z = 2|X = 2, Y = 1) = \frac{1}{3} \qquad P(Z = 2|Y = 1) = \frac{2}{6} = \frac{1}{3}$$

$Y = 1$

U_x	U_y	U_z	X	Y	Z
1	0	0	1	1	1
1	0	0	1	1	1
1	0	1	1	1	2
2	1	0	2	1	1
2	1	0	2	1	1
2	1	1	2	1	2

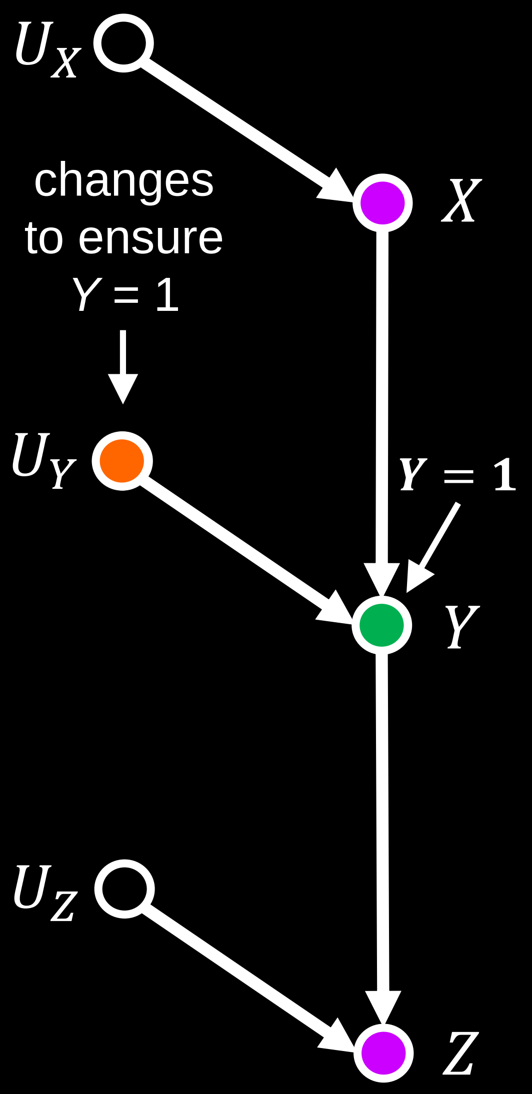


Figure 2.1

We compare X and Z on all the cases where $Y = 1$, and on all the cases where $Y = 2$.

Let's assume that we are looking at the cases where $Y = 1$.

We want to know whether, in these cases only, the value of Z is independent of the value of X.

Now, however, examining only the cases where $Y = 1$, when we select cases with different values of X, the value of U_Y changes so as to keep Y at $Y = 1$, but since Z depends only on Y and U_Z, not on U_Y, the value of Z remains unaltered.

$$Y = X - U_Y$$

U_x	U_y	U_z	X	Y	Z
1	0	0	1	1	1
1	0	0	1	1	1
1	0	1	1	1	2
2	1	0	2	1	1
2	1	0	2	1	1
2	1	1	2	1	2

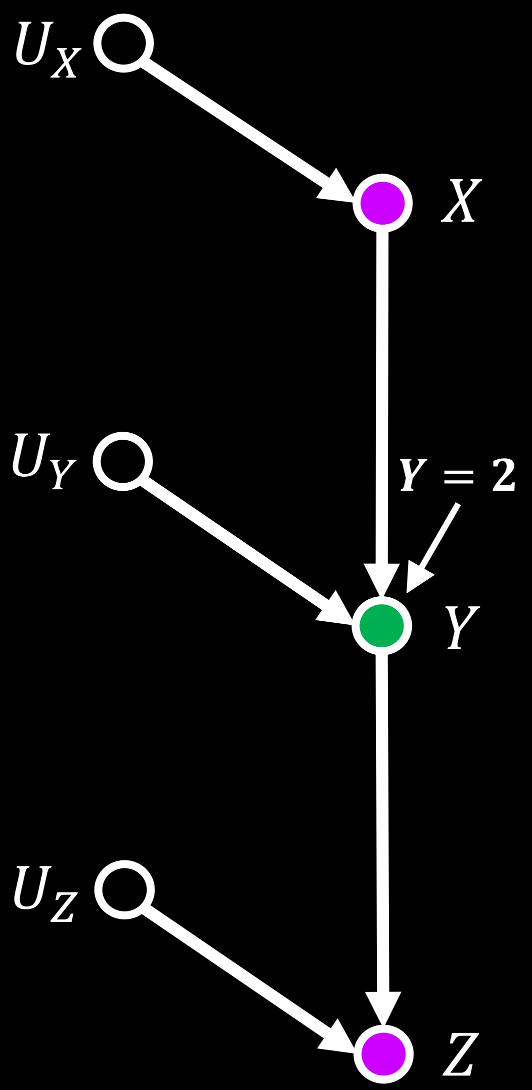


Figure 2.1

So, in the case where $Y = 1$, X is independent of Z.

This is of course true no matter which specific value of Y we condition on, i.e. $Y = 2$.

$$P(Z = z|X = x, Y = y) = P(Z = z|Y = y), \qquad \forall\, x, z \quad Y = 2$$

$$P(Z = 1|X = 1, Y = 2) = \frac{0}{2} = 0 \qquad P(Z = 1|X = 2, Y = 2) = \frac{0}{2} = 0 \qquad P(Z = 1|Y = 2) = \frac{0}{4} = 0$$

$$P(Z = 2|X = 1, Y = 2) = \frac{2}{2} = 1 \qquad P(Z = 2|X = 2, Y = 2) = \frac{2}{2} = 1 \qquad P(Z = 2|Y = 2) = \frac{4}{4} = 1$$

U_x	U_y	U_z	**X**	**Y**	**Z**
2	0	0	**2**	**2**	**2**
1	-1	0	**1**	**2**	**2**
1	-1	0	**1**	**2**	**2**
2	0	0	**2**	**2**	**2**

Therefore, X is independent of Z, conditionally on Y.

2.2 Chain and Forks

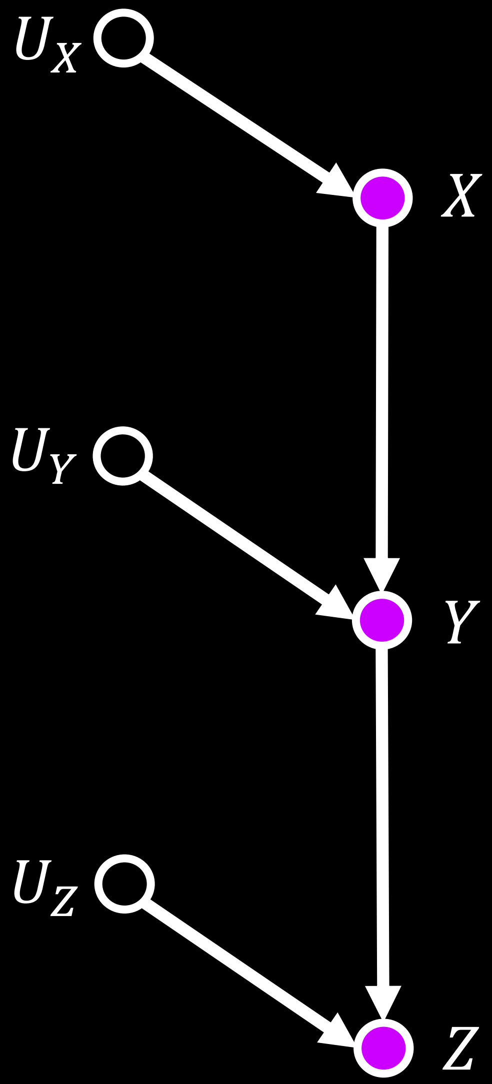


Figure 2.1

This configuration of variables, three nodes and two edges, with one edge directed into and one edge directed out of the middle variable, is called a **chain**.

In any graphical model, given any two variables *X* and *Y*, if the only path between *X* and *Y* is composed entirely of chains, then *X* and *Y* are Independent conditional on any intermediate variable on the path.

This independence relation holds regardless of the functions that connect the variables.

The previous, gives the following rule: **WARNING** → *X*, *Y* and *Z* in **Rule 1** do not refer to **Figure 2.1**.

Rule 1 (**Conditional Independence in Chains**)

Two variables, *X* and *Y*, are conditionally independent given *Z*, if there is only one unidirectional path between *X* and *Y*, and *Z* is any set of variables that intercepts that path.

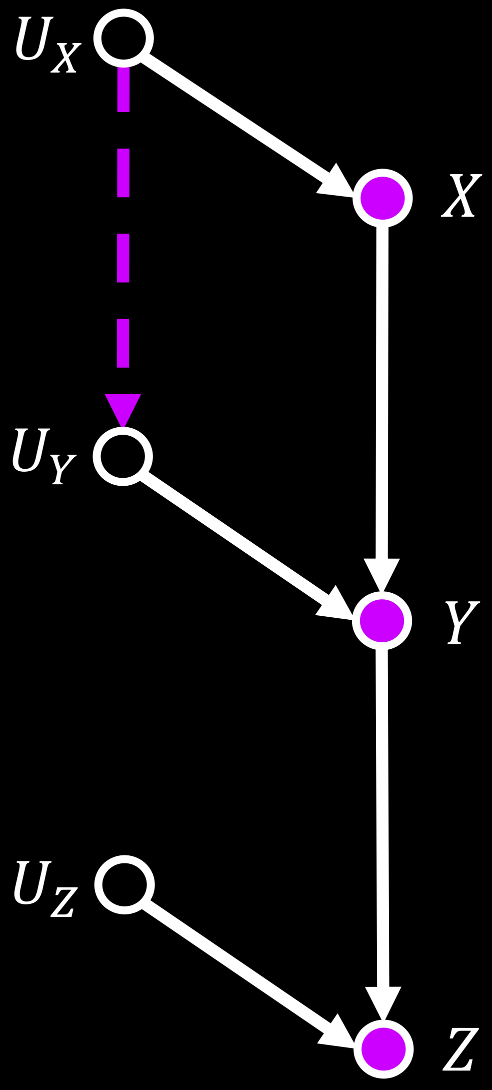


Figure 2.1

If, for instance, U_X were a cause of U_Y (dashed arrow in **Figure 2.1**), then conditioning on Y would not necessarily make X and Z independent, because variations in X could still be associated with variations in Y, through their error terms.

Rule 1 only holds when we assume that the error terms U_X, U_Y, and U_Z are independent of each other.

The previous, gives the following rule:

WARNING → X, Y and Z in **Rule 1** do not refer to **Figure 2.1**.

Rule 1 (**Conditional Independence in Chains**)

Two variables, X and Y, are conditionally independent given Z, if there is only one unidirectional path between X and Y, and Z is any set of variables that intercepts that path.

Now, consider the graphical model in **Figure 2.2.**

This structure might represent, for example, the causal mechanism that connects a day's temperature in a city in degrees Fahrenheit (*X*), the number of sales at a local ice cream shop on that day (*Y*), and the number of violent crimes in the city on that day (*Z*).

SCM 2.2.5 (Temperature, Ice Cream Sales, and Crime)

$$M = \langle U, V, F \rangle$$

$$U = \{U_X, U_Y, U_Z\}$$

$$V = \{X, Y, Z\}$$

$$F_3 = \{f_X, f_Y, f_Z\}$$

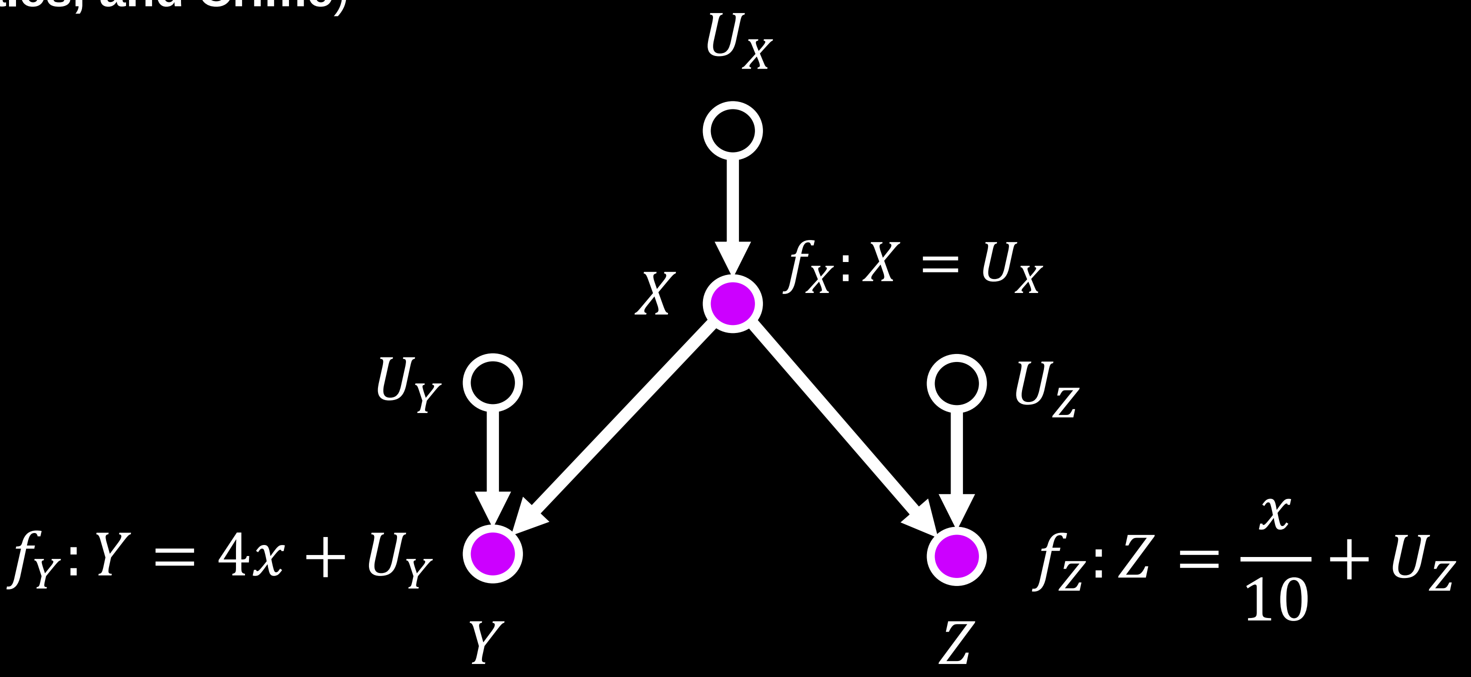


Figure 2.2

Now, consider the graphical model in **Figure 2.2.**

Causal mechanism that connects the state (*up* or *down*) of a switch (*X*), the state (*on* or *off*) of one light bulb (*Y*), and the state (*on* or *off*) of a second light bulb (*Z*).

SCM 2.2.6 (**Switch and Two Light Bulbs**)

$$M = \langle U, V, F \rangle$$

$$f_Y : Y = \begin{cases} on & \text{IF } (X = up \text{ AND } U_Y = 0) \text{ OR } (X = down \text{ AND } U_Y = 1) \\ off & \text{otherwise} \end{cases}$$

$$f_Z : Z = \begin{cases} on & \text{IF } (X = up \text{ AND } U_Z = 0) \text{ OR } (X = down \text{ AND } U_Z = 1) \\ off & \text{otherwise} \end{cases}$$

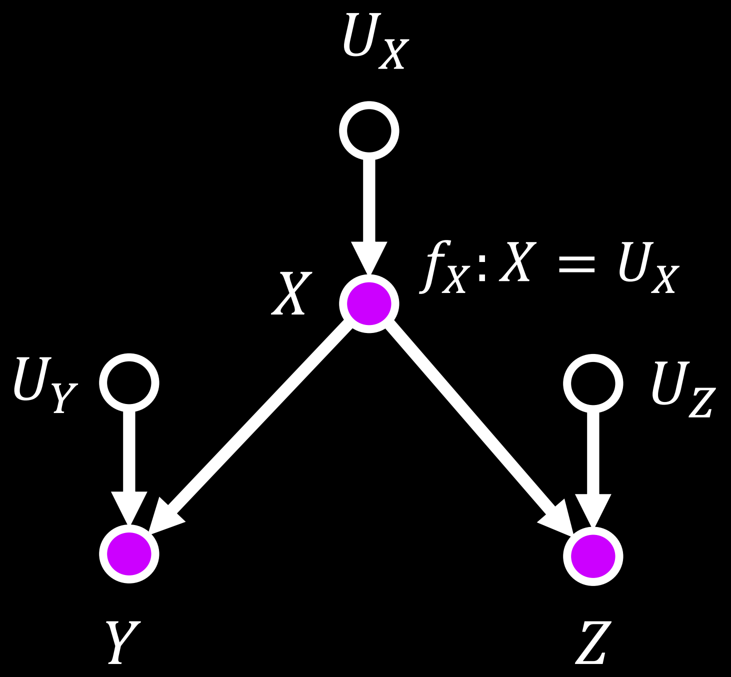


Figure 2.2

If we assume that the error terms U_X, U_Y, and U_Z are independent, then by examining the graphical model in **Figure 2.2**, we can determine that the **SCMs 2.2.5/2.2.6** share the following dependencies and independencies:

1. ***X* and *Y* are likely dependent**, i.e., for some pair of values *x*, *y*

$$P(X = x|Y = y) \neq P(X = x)$$

2. ***X* and *Z* are likely dependent**, i.e., for some pair of values *x*, *z*

$$P(X = x|Z = z) \neq P(X = x)$$

3. ***Z* and *Y* are likely dependent**, i.e., for some pair of values *z*, *y*

$$P(Z = z|Y = y) \neq P(Z = z)$$

4. ***Y* and *Z* are independent, conditional on *X***, i.e., for all values *x*, *y*, *z*

$$P(Y = y|X = x, Z = z) = P(Y = y|X = x)$$

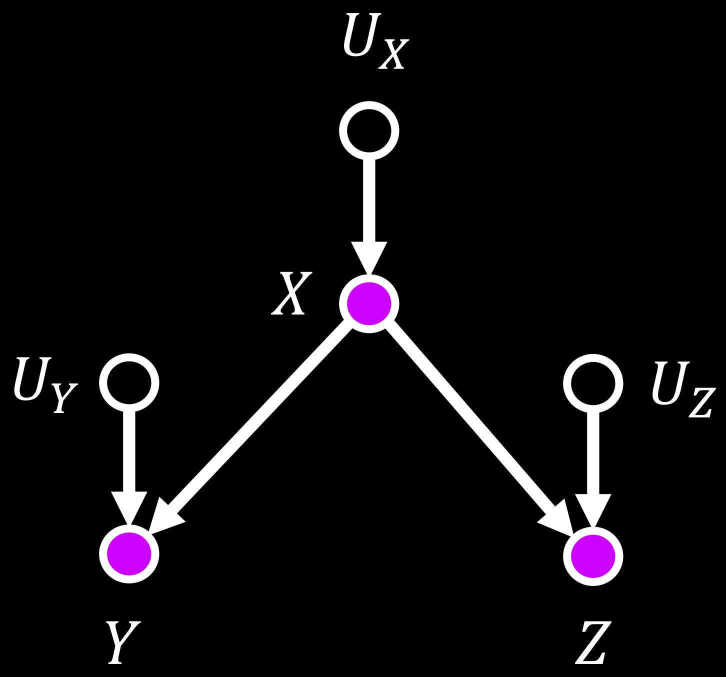


Figure 2.2

2.2 Chain and Forks

If we assume that the error terms U_X, U_Y, and U_Z are independent, then by examining the graphical model in **Figure 2.2**, we can determine that the **SCMs 2.2.5/2.2.6** share the following dependencies and independencies:

1. ***X* and *Y* are likely dependent**, i.e., for some pair of values *x*, *y*

$$P(X = x|Y = y) \neq P(X = x)$$

2. ***X* and *Z* are likely dependent**, i.e., for some pair of values *x, z*

$$P(X = x|Z = z) \neq P(X = x)$$

Follow, once again, by the fact that *Y* and *Z* are both directly connected to *X* by an arrow.

When the value of *X* changes, the values of both *Y* and *Z* **likely change**.

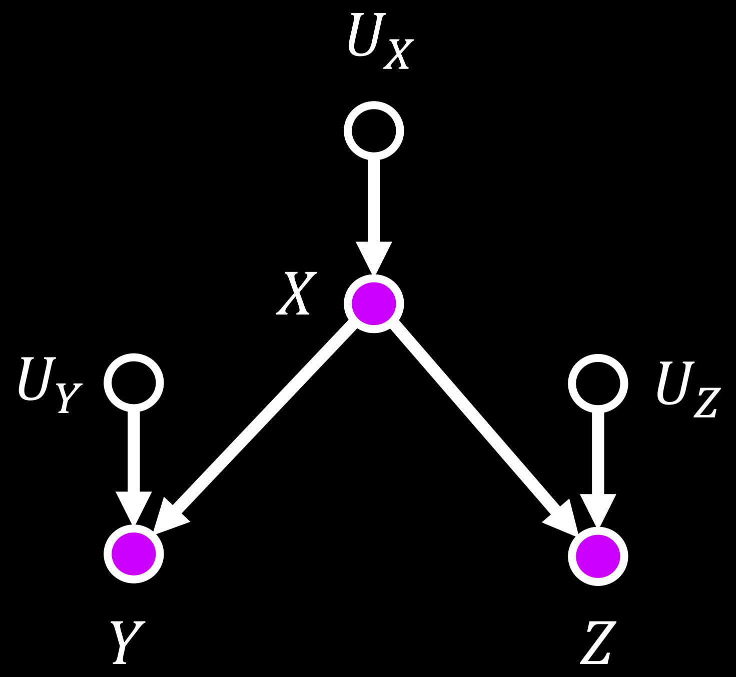


Figure 2.2

If we assume that the error terms U_X, U_Y, and U_Z are independent, then by examining the graphical model in **Figure 2.2**, we can determine that the **SCMs 2.2.5/2.2.6** share the following dependencies and independencies:

This tells us something further, however: If *Y* changes when *X* changes, and *Z* changes when *X* changes, then it is likely (though not certain) that *Y* changes together with *Z*, and vice versa.

Therefore, since a change in the value of *Y* gives us information about an associated change in the value of *Z*, *Y* and *Z* are likely dependent variables.

3. ***Z* and *Y* are likely dependent**, i.e., for some pair of values *z*, *y*

$$P(Z = z|Y = y) \neq P(Z = z)$$

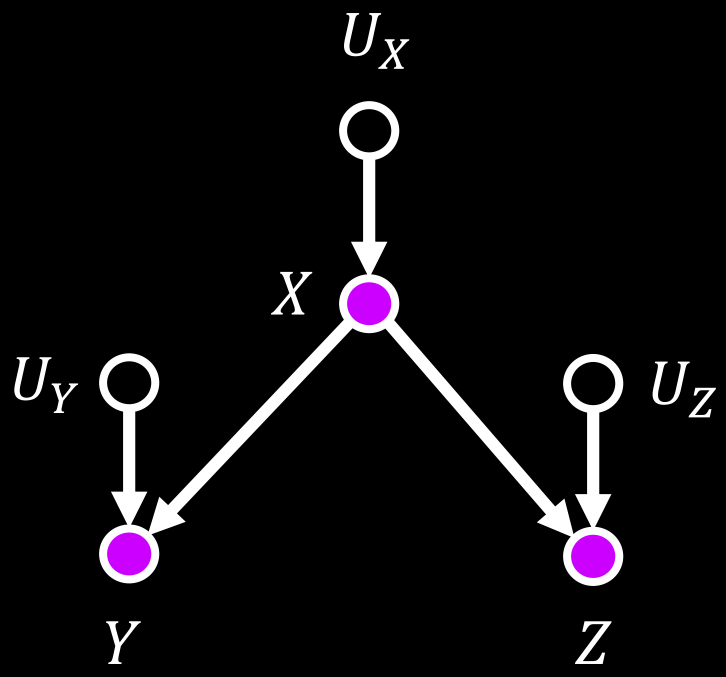


Figure 2.2

If we assume that the error terms U_X, U_Y, and U_Z are independent, then by examining the graphical model in **Figure 2.2**, we can determine that the **SCMs 2.2.5/2.2.6** share the following dependencies and independencies:

Why, then, are Y and Z independent conditional on X?

What happens when we condition on X?

We filter the data based on the value of X. So now, we are only comparing cases where the value of X is constant.

Since X does not change, the values of Y and Z do not change in accordance with it, they change only in response to U_Y and U_Z, which we have assumed independent.

Therefore, any additional changes in the values of Y and Z must be independent of each other.

4. **Y and Z are independent, conditional on X,** i.e., for all values *x, y, z*

$$P(Y = y|X = x, Z = z) = P(Y = y|X = x)$$

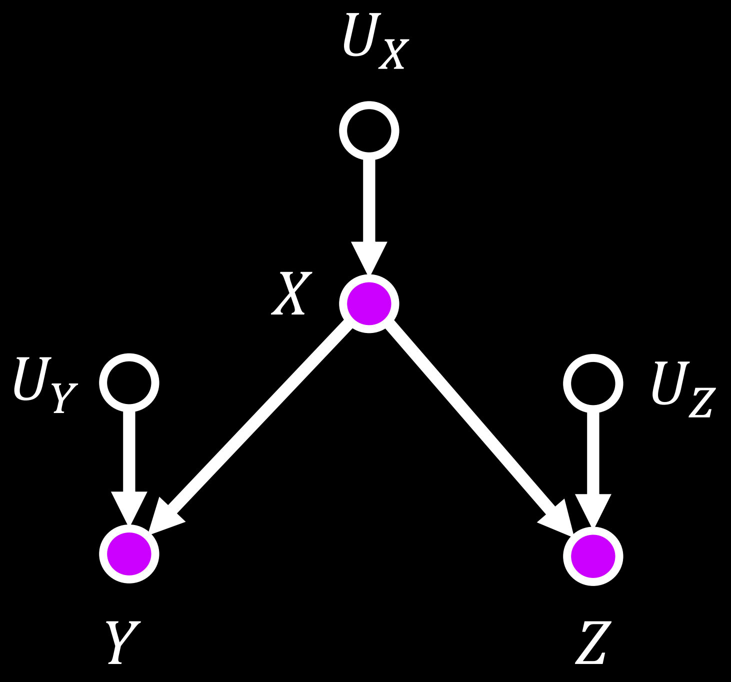


Figure 2.2

This configuration of variables, three nodes, with two arrows emanating from the middle variable, is called a **fork**.

The middle variable in a fork (X) is a **common cause** of the other two variables (Y and Z), and of any of their descendants.

If two variables share a common cause, and if that common cause is part of the only path between them, then analogous reasoning to the above tells us that these dependencies and conditional independencies are true of those variables.

Therefore, we come by another rule:

Rule 2 (Conditional Independence in Forks)

If a variable X is a common cause of variables Y and Z, and there is only one path between Y and Z, then Y and Z are independent conditional on X.

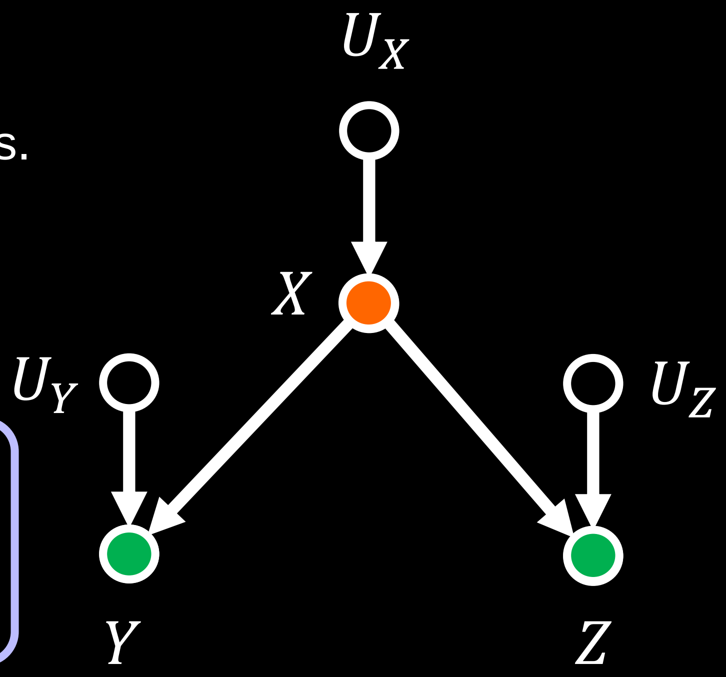


Figure 2.2

So far we have looked at two simple configurations of edges and nodes that can occur on a path between two variables: chains and forks.

There is a third such configuration that we speak of separately, because it carries with it unique considerations and challenges.

The third configuration contains a **collider** node, and it occurs when one node receives edges from two other nodes.

The simplest graphical causal model containing a **collider** is illustrated in **Figure 2.3**, representing a common effect *Z*, of two causes *X* and *Y*.

As is the case with every graphical causal model, all **SCMs** that have **Figure 2.3** as their graph share a set of dependencies and independencies that we can determine from the graphical model alone.

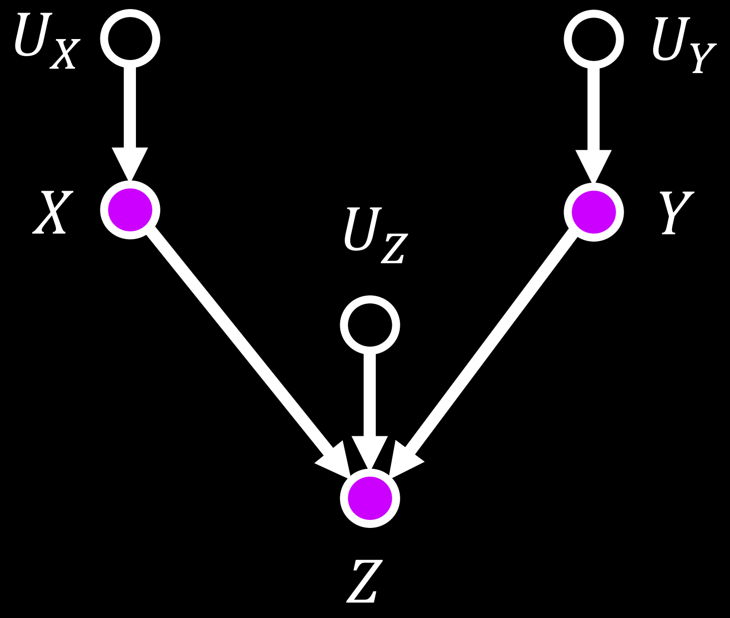


Figure 2.3

Assume that the error terms U_X, U_Y, and U_Z are independent, then by examining the graphical model in **Figure 2.3**, we can determine that any corresponding **SCMs** share the following dependencies and independencies:

1. ***X* and *Z* are likely dependent**, i.e., for some pair of values *x*, *z*

$$P(X = x|Z = z) \neq P(X = x)$$

2. ***Y* and *Z* are likely dependent**, i.e., for some pair of values *y*, *z*

$$P(Y = y|Z = z) \neq P(Y = y)$$

3. ***X* and *Y* are independent**, i.e., for all pairs of values *x*, *y*

$$P(X = x|Y = y) = P(X = x)$$

4. ***X* and *Y* are likely dependent, conditional on *Z*,** i.e., for some values *x*, *y*, *z*

$$P(X = x|Y = y, Z = z) \neq P(X = x|Z = z)$$

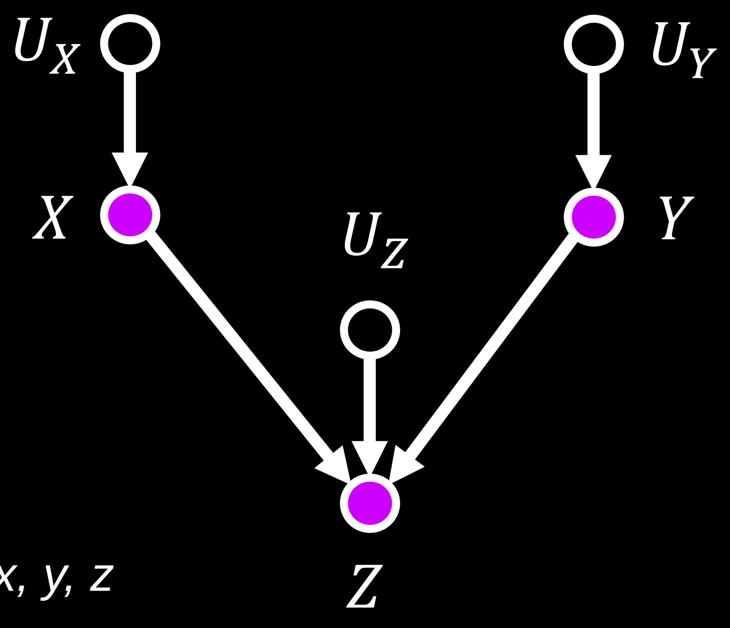


Figure 2.3

The truth of the first two points was established in **Section 2.2 Chain and Forks**.

1. ***X* and *Z* are likely dependent**, i.e., for some pair of values *x*, *z*

$$P(X = x|Z = z\,) \neq P(X = x)$$

2. ***Y* and *Z* are likely dependent**, i.e., for some pair of values *y*, *z*

$$P(Y = y|Z = z\,) \neq P(Y = y)$$

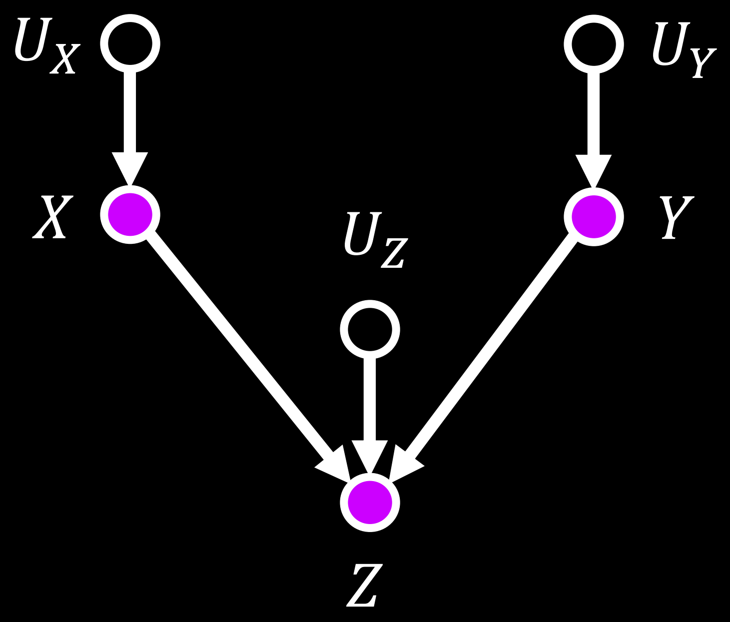


Figure 2.3

The third point is self-evident,

- neither X nor Y is a descendant or an ancestor of the other,
- nor do they depend for their value on the same variable,
- X and Y respond only to U_X and U_Y, which we assumed to be independent,

so there is no causal mechanism by which variations in the value of X should be associated with variations in the value of Y.

3. **X and Y are independent**, i.e., for all pairs of values x, y

$$P(X = x|Y = y) = P(X = x)$$

This independence also reflects our understanding of how causation operates in time; events that are independent in the present do not become dependent merely because they may have common effects in the future.

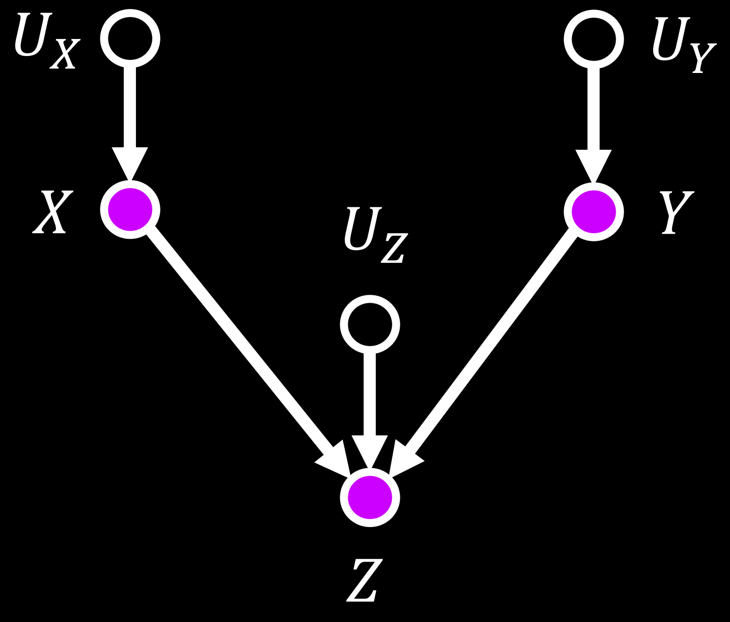


Figure 2.3

Why, then, does Point 4 hold?

Why would two independent variables suddenly become dependent when we condition on their common effect?

To answer this question, we return again to the definition of conditioning as filtering by the value of the conditioning variable.

When we condition on *Z*, we limit our comparison to cases in which *Z* takes the same value.

But remember that *Z* depends, for its value, on *X* and *Y*. So, when comparing cases where *Z* takes some value, any change in value of *X* must be compensated for by a change in the value of *Y*, otherwise, the value of *Z* would change as well.

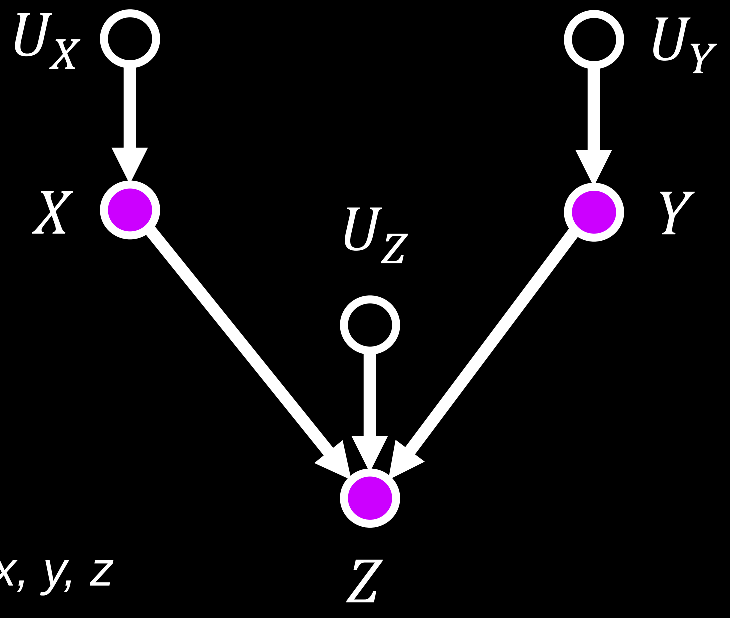


Figure 2.3

4. ***X* and *Y* are likely dependent, conditional on *Z*,** i.e., for some values *x, y, z*

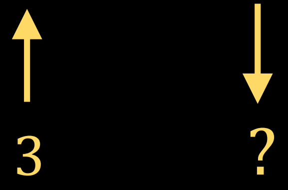
$$P(X = x|Y = y, Z = z) \neq P(X = x|Z = z)$$

Why, then, does Point 4 hold?

Why would two independent variables suddenly become dependent when we condition on their common effect?

The reasoning behind this attribute of colliders, that conditioning on a collider node produces a dependence between the node's parents, can be difficult to grasp at first.

$Z = X + Y$ $\quad$ $Y = Z - 3$



From the *X* value we learn nothing about the *Y* value, because the two numbers are independent

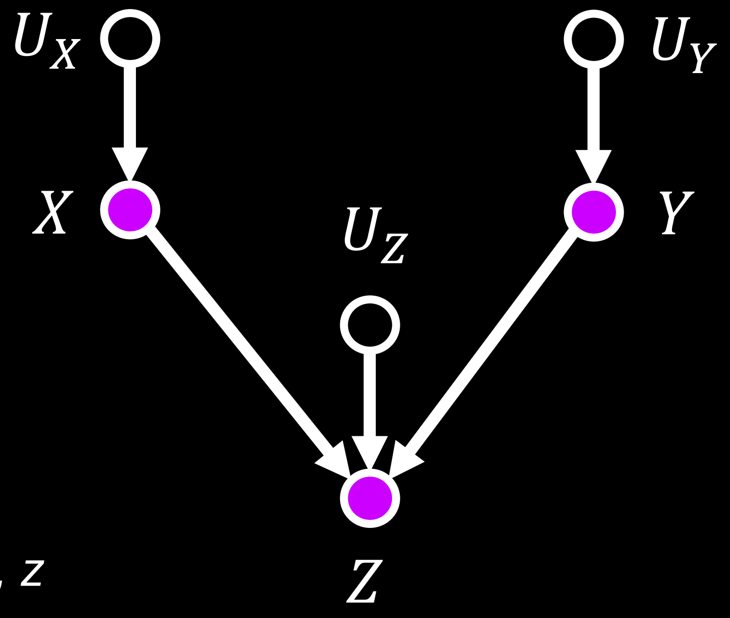


Figure 2.3

4. *X* and *Y* are likely dependent, conditional on *Z*, i.e., for all values *x, y, z*

$$P(X = x|Y = y, Z = z) \neq P(X = x|Z = z)$$

Why, then, does Point 4 hold?

Why would two independent variables suddenly become dependent when we condition on their common effect?

The reasoning behind this attribute of colliders, that conditioning on a collider node produces a dependence between the node's parents, can be difficult to grasp at first.

$Z = X + Y$ $\quad$ $Y = 10 - 3$

↑ ↑ ↓

10 $\quad$ 3 $\quad$ 7

Thus X and Y are dependent, given that Z = 10.

We are implicitly assuming U_X, U_Y, and U_Z to be zero.

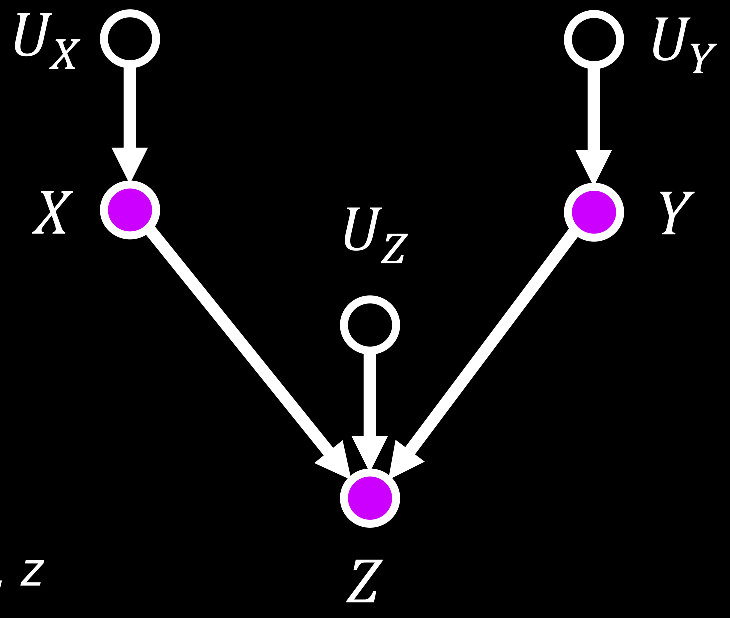


Figure 2.3

4. ***X* and *Y* are likely dependent, conditional on *Z*,** i.e., for all values *x*, *y*, *z*

$$P(X = x|Y = y, Z = z) \neq P(X = x|Z = z)$$

Consider a simultaneous (independent) toss of **two fair coins** and a **bell** that rings whenever at least one of the coin lands on heads.

X Y Z

$X, Y \in \{heads, tails\}$ $Z \in \{silence, rings\}$

We know the following:

- **coin 1** landed on **heads** (*X* = *heads*)

Tells us nothing about the outcome of the toss of coin 2 (*Y*)

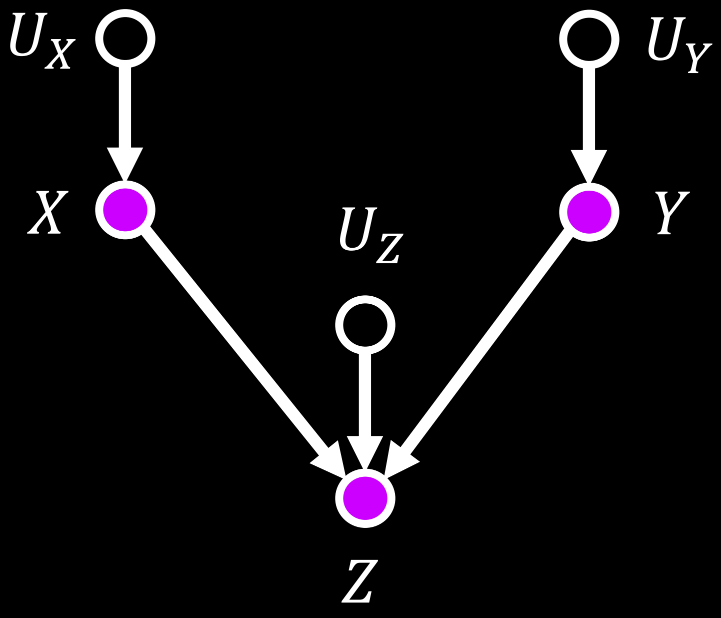


Figure 2.3

Consider a simultaneous (independent) toss of **two fair coins** and a **bell** that rings whenever at least one of the coin lands on heads.

X $\quad$ Y $\quad$ Z

$X, Y \in \{heads, tails\}$ $\quad$ $Z \in \{silence, rings\}$

We know the following:

- we hear the **bell ringing** *(Z = rings)*
- **coin 1** landed on **tails** *(X = tails)*

Tells us that coin 2 must have landed on heads, i.e., *Y = heads*.

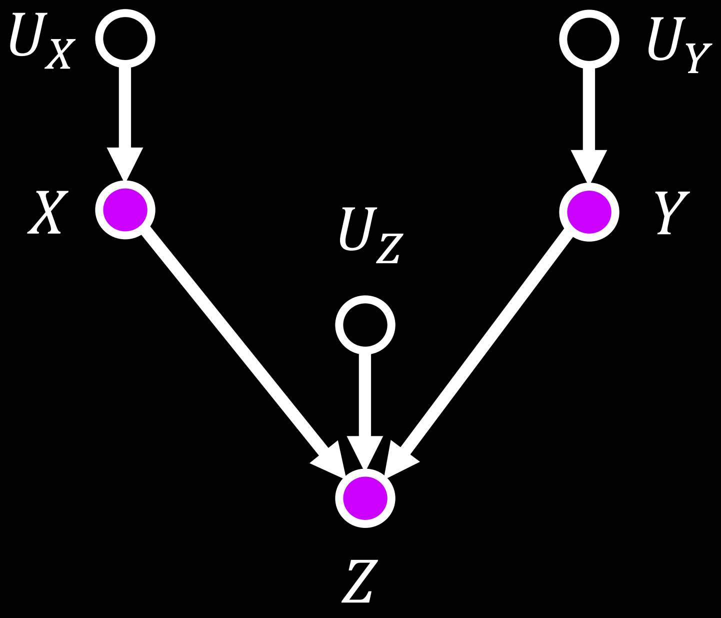


Figure 2.3

$X \qquad Y \qquad Z$

$X, Y \in \{heads, tails\} \qquad Z \in \{silence, rings\}$

Consider a simultaneous (independent) toss of **two fair coins** and a **bell** that rings whenever at least one of the coin lands on heads.

We know the following:

- we hear the **bell ringing** (*Z = rings*)
- we know that **coin 2** landed on **heads** (*Y = heads*)

$$P(X = heads|Z = rings\ ) \neq P(X = heads|Y = heads, Z = rings\ )$$

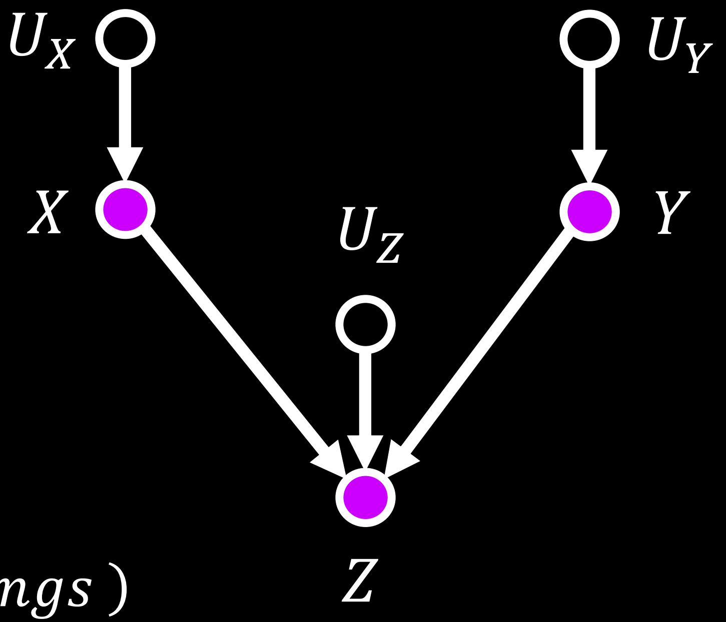


Figure 2.3

To see the latter calculation, consider the initial probabilities in **Table 2.1**. We see that

$$P(X = heads) = 0.5$$

$$P(X = tails) = 0.5$$

Table 2.1 Probability distribution for two flips of a fair coin, with X representing flip one, Y representing flip two, and Z representing a bell that rings if either flip results in heads.

X Coin 1	*Y* Coin 2	*Z* Bell	*P(X, Y, Z)*
heads	heads	rings	0.25
heads	tails	rings	0.25
tails	heads	rings	0.25
tails	tails	silence	0.25

To see the latter calculation, consider the initial probabilities in **Table 2.1**. We see that

$$P(X = heads) = 0.5 = P(X = heads|Y = tails)$$

$$P(X = tails) = 0.5$$

Table 2.1 Probability distribution for two flips of a fair coin, with X representing flip one, Y representing flip two, and Z representing a bell that rings if either flip results in heads.

X Coin 1	*Y* Coin 2	*Z* Bell	*P(X, Y, Z)*
heads	tails	rings	0.25
tails	tails	silence	0.25

To see the latter calculation, consider the initial probabilities in **Table 2.1**. We see that

$$P(X = heads) = 0.5 = P(X = heads|Y = tails) = P(X = heads|Y = heads)$$

$$P(X = tails) = 0.5$$

Table 2.1 Probability distribution for two flips of a fair coin, with X representing flip one, Y representing flip two, and Z representing a bell that rings if either flip results in heads.

X Coin 1	*Y* Coin 2	*Z* Bell	*P(X, Y, Z)*
heads	heads	rings	0.25
tails	heads	rings	0.25

To see the latter calculation, consider the initial probabilities in **Table 2.1**. We see that

$$P(X = heads) = 0.5 = P(X = heads|Y = tails) = P(X = heads|Y = heads)$$

$$P(X = tails) = 0.5 = P(X = tails|Y = tails)$$

Table 2.1 Probability distribution for two flips of a fair coin, with X representing flip one, Y representing flip two, and Z representing a bell that rings if either flip results in heads.

X Coin 1	*Y* Coin 2	*Z* Bell	*P(X, Y, Z)*
heads	tails	rings	0.25
tails	tails	silence	0.25

To see the latter calculation, consider the initial probabilities in **Table 2.1**. We see that

$$P(X = heads) = 0.5 = P(X = heads|Y = tails) = P(X = heads|Y = heads)$$

$$P(X = tails) = 0.5 = P(X = tails|Y = tails) = P(X = tails|Y = heads)$$

Table 2.1 Probability distribution for two flips of a fair coin, with X representing flip one, Y representing flip two, and Z representing a bell that rings if either flip results in heads.

X Coin 1	*Y* Coin 2	*Z* Bell	*P(X, Y, Z)*
heads	heads	rings	0.25
tails	heads	rings	0.25

To see the latter calculation, consider the initial probabilities in **Table 2.1**. We see that

$$P(X = heads) = 0.5 = P(X = heads|Y = tails) = P(X = heads|Y = heads)$$

$$P(X = tails) = 0.5 = P(X = tails|Y = tails) = P(X = tails|Y = heads)$$

***X* and *Y* are independent**

Table 2.1 Probability distribution for two flips of a fair coin, with X representing flip one, Y representing flip two, and Z representing a bell that rings if either flip results in heads.

X Coin 1	*Y* Coin 2	*Z* Bell	*P(X, Y, Z)*
heads	heads	rings	0.25
heads	tails	rings	0.25
tails	heads	rings	0.25
tails	tails	silence	0.25

Now let's condition on Z = *rings* and Z = *silence*, the resulting data subsets are shown in **Table 2.2**.

Table 2.2 Conditional probability distribution for the distribution in **Table 2.1.**

X Coin 1	*Y* Coin 2	*P(X, Y|Z=silence)*
heads	heads	0
heads	tails	0
tails	heads	0
tails	tails	1

X Coin 1	*Y* Coin 2	*P(X, Y|Z=rings)*
heads	heads	0.333
heads	tails	0.333
tails	heads	0.333
tails	tails	0

$$P(X = heads|Z = rings\ ) =$$

Now let's condition on *Z* = *rings* and *Z* = *silence*, the resulting data subsets are shown in **Table 2.2**.

$$P(X = heads|Z = rings\ ) =$$

X Coin 1	*Y* Coin 2	*P(X, Y|Z=rings)*
heads	heads	0.333
heads	tails	0.333
tails	heads	0.333
tails	tails	0

2.3 Colliders

Now let’s condition on *Z* = *rings* and *Z* = *silence*, the resulting data subsets are shown in **Table 2.2**.

$$P(X = heads|Z = rings\ ) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

$$P(X = heads|Y = heads, Z = rings\ ) =$$

X Coin 1	*Y* Coin 2	*P(X, Y|Z=rings)*
heads	heads	0.333
heads	tails	0.333
tails	heads	0.333
tails	tails	0

Now let's condition on *Z* = *rings* and *Z* = *silence*, the resulting data subsets are shown in **Table 2.2**.

$$P(X = heads|Z = rings) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

$$P(X = heads|Y = heads, Z = rings) = \frac{1}{2}$$

X Coin 1	*Y* Coin 2	*P(X, Y|Z=rings)*
heads	heads	0.333
heads	tails	0.333
tails	heads	0.333
tails	tails	0

Now let's condition on *Z* = *rings* and *Z* = *silence*, the resulting data subsets are shown in **Table 2.2**.

Given *Z* = *rings*, the probability of *X* = *heads is*

$$P(X = heads|Z = rings) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

However, when we learn that *Y* = *heads*, the probability of *X* = *heads* changes as follows

$$P(X = heads|Y = heads, Z = rings) = \frac{1}{2}$$

$$P(X = heads|Y = heads, Z = rings) \neq P(X = heads|Z = rings)$$

X Coin 1	*Y* Coin 2	*P(X, Y|Z=rings)*
heads	heads	0.333
heads	tails	0.333
tails	heads	0.333
tails	tails	0

Therefore, we conclude that *X* and *Y* are dependent given *Z* = *rings*.

Table 2.2 Conditional probability distribution for the distribution in **Table 2.1.**

X Coin 1	*Y* Coin 2	*P(X, Y|Z=silence)*
heads	heads	0
heads	tails	0
tails	heads	0
tails	tails	1

A more pronounced dependence occurs, of course, when the bell does not ring (Z = *silence*), because then we know that both coins must have landed on tails.

Just as **conditioning on a collider** makes previously independent variables dependent, so too does **conditioning on any descendant of a collider**.

X Y Z

To see why this is true, let's return to our example of two independent coins and a bell.

Suppose we do not hear the bell directly, but instead rely on a witness (*W*) who is somewhat unreliable;

- whenever the bell does not ring (*Z* = *silence*), 50% chance that the witness will falsely report that it did (*W* = 1), and 50% chance that will correctly report that it did not (*W* = 0).
- whenever the bell rings (*Z* = *rings*), 100% chance that the witness will report that it did (*W* = 1).

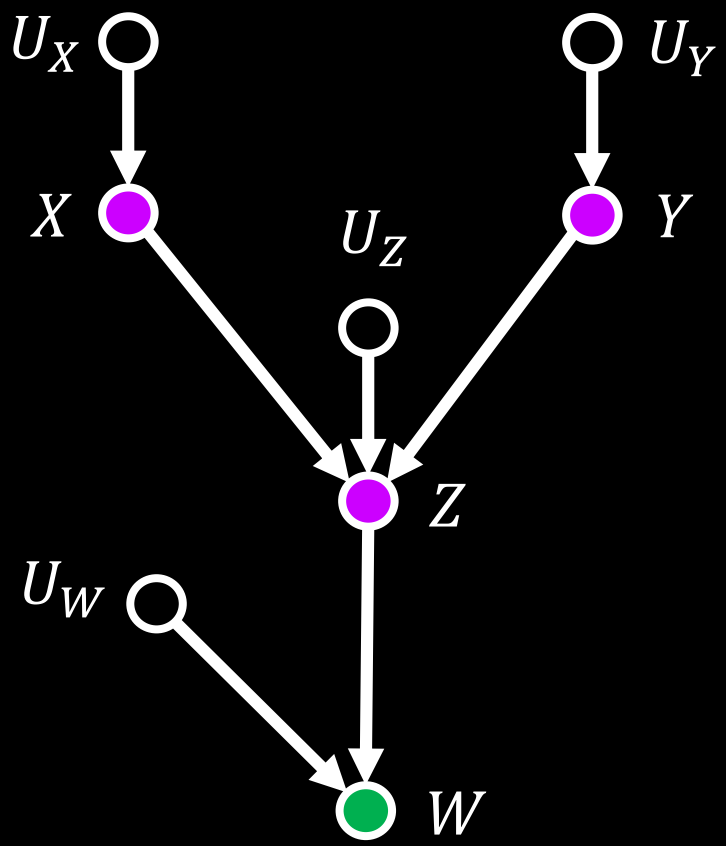


Figure 2.4

Probabilities for all combinations of *X*, *Y* and *W* are shown in **Table 2.3**.

Table 2.3 Probability distribution for two flips of a fair coin and a bell that rings if either flip results in heads, with *X* representing flip one, *Y* representing flip two, and *W* representing a witness who, with variable reliability, reports whether or not the bell has rung.

X Coin 1	*Y* Coin 2	*W* witness	*P(X, Y, W)*
heads	heads	1	0.250
heads	tails	1	0.250
tails	heads	1	0.250
tails	tails	1	0.125
tails	tails	0	0.125

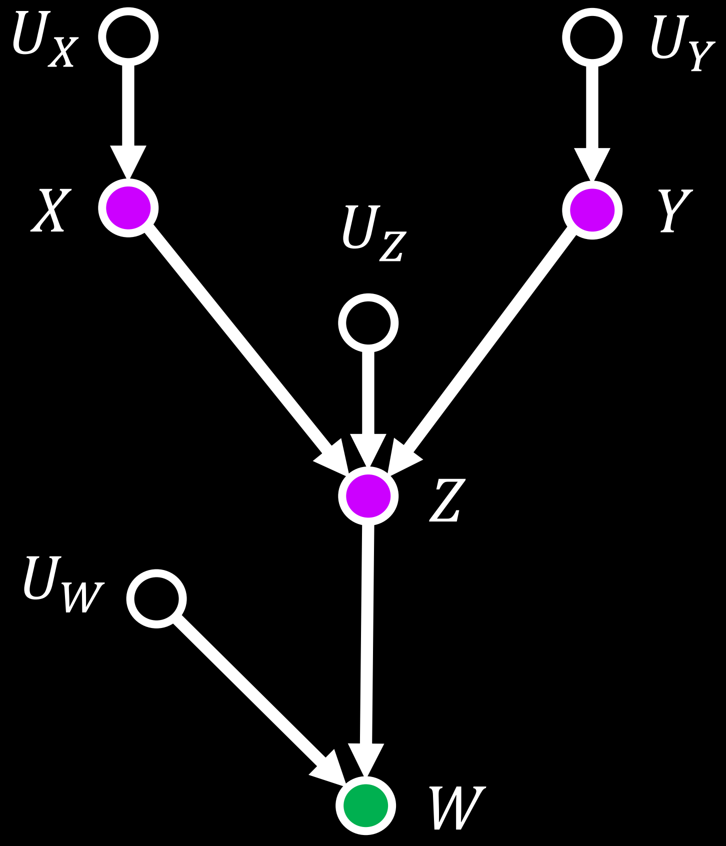


Figure 2.4

How do we get **Table 2.3 ?**

Table 2.1 Probability distribution for two flips of a fair coin, with X representing flip one, Y representing flip two, and Z representing a bell that rings if either flip results in heads.

X Coin 1	*Y* Coin 2	*Z* Bell	*P(X, Y, Z)*
heads	heads	rings	0.25
heads	tails	rings	0.25
tails	heads	rings	0.25
tails	tails	silence	0.25

Table 2.3 Probability distribution for two flips of a fair coin and a bell that rings if either flip results in heads, with *X* representing flip one, *Y* representing flip two, and *W* representing a witness who, with variable reliability, reports whether or not the bell has rung.

X Coin 1	*Y* Coin 2	*W* witness	*P(X, Y, W)*
heads	heads	1	0.250
heads	tails	1	0.250
tails	heads	1	0.250
tails	tails	1	0.125
tails	tails	0	0.125

2.3 Colliders

How do we get **Table 2.3 ?**

Table 2.1 Probability distribution for two flips of a fair coin, with X representing flip one, Y representing flip two, and Z representing a bell that rings if either flip results in heads.

X Coin 1	*Y* Coin 2	*Z* Bell	*P(X, Y, Z)*
heads	heads	rings	0.25
heads	tails	rings	0.25
tails	heads	rings	0.25

Table 2.3 Probability distribution for two flips of a fair coin and a bell that rings if either flip results in heads, with *X* representing flip one, *Y* representing flip two, and *W* representing a witness who, with variable reliability, reports whether or not the bell has rung.

X Coin 1	*Y* Coin 2	*W* witness	*P(X, Y, W)*
heads	heads	1	0.250
heads	tails	1	0.250
tails	heads	1	0.250

100% chance that

W = 1

2.3 COLLIDERS

How do we get **Table 2.3 ?**

Table 2.1 Probability distribution for two flips of a fair coin, with X representing flip one, Y representing flip two, and Z representing a bell that rings if either flip results in heads.

X Coin 1	***Y*** Coin 2	***Z*** Bell	***P(X, Y, Z)***
tails	tails	silence	0.25

Table 2.3 Probability distribution for two flips of a fair coin and a bell that rings if either flip results in heads, with *X* representing flip one, *Y* representing flip two, and *W* representing a witness who, with variable reliability, reports whether or not the bell has rung.

X Coin 1	***Y*** Coin 2	***W*** witness	***P(X, Y, W)***
tails	tails	1	0.125
tails	tails	0	0.125

50% chance that *W* = 1

50% chance that *W* = 0

Based on **Table 2.3** we can easily check that

$$P(X = heads|Y = heads) = \frac{0.250}{0.250 + 0.250} = 0.5$$

Table 2.3 Probability distribution for two flips of a fair coin and a bell that rings if either flip results in heads, with *X* representing flip one, *Y* representing flip two, and *W* representing a witness who, with variable reliability, reports whether or not the bell has rung.

X Coin 1	*Y* Coin 2	*W* witness	*P(X, Y, W)*
heads	heads	1	0.250
tails	heads	1	0.250

Based on **Table 2.3** we can easily check that

Table 2.3 Probability distribution for two flips of a fair coin and a bell that rings if either flip results in heads, with *X* representing flip one, *Y* representing flip two, and *W* representing a witness who, with variable reliability, reports whether or not the bell has rung.

X Coin 1	*Y* Coin 2	*W* witness	*P(X, Y, W)*
heads	heads	1	0.250
heads	tails	1	0.250

$$P(X = heads|Y = heads) = \frac{0.250}{0.250 + 0.250} = 0.5$$

$$P(X = heads) = 0.250 + 0.250 = 0.5$$

$$P(X = heads|Y = heads) = P(X = heads) = 0.5$$

Based on **Table 2.3** we can easily check that

Table 2.3 Probability distribution for two flips of a fair coin and a bell that rings if either flip results in heads, with *X* representing flip one, *Y* representing flip two, and *W* representing a witness who, with variable reliability, reports whether or not the bell has rung.

X Coin 1	*Y* Coin 2	*W* witness	*P(X, Y, W)*
heads	heads	1	0.250
heads	tails	1	0.250
tails	heads	1	0.250
tails	tails	1	0.125
tails	tails	0	0.125

In the same way we can show the following

$$P(X = heads|Y = heads\ ) = P(X = heads\ ) = 0.5$$

$$P(X = heads|Y = tails\ ) = P(X = heads\ ) = 0.5$$

$$P(X = tails|Y = heads\ ) = P(X = tails\ ) = 0.5$$

$$P(X = tails|Y = tails\ ) = P(X = tails\ ) = 0.5$$

Therefore, we conclude that *X* and *Y* are independent.

Based on **Table 2.3** we can easily check that

Table 2.3 Probability distribution for two flips of a fair coin and a bell that rings if either flip results in heads, with *X* representing flip one, *Y* representing flip two, and *W* representing a witness who, with variable reliability, reports whether or not the bell has rung.

X Coin 1	*Y* Coin 2	*W* witness	*P(X, Y, W)*
heads	heads	1	0.250
heads	tails	1	0.250

$$P(X = heads|W = 1\,) = \frac{0.25 + 0.25}{0.25 + 0.25 + 0.25 + 0.125} = 0.571$$

Based on **Table 2.3** we can easily check that

Table 2.3 Probability distribution for two flips of a fair coin and a bell that rings if either flip results in heads, with *X* representing flip one, *Y* representing flip two, and *W* representing a witness who, with variable reliability, reports whether or not the bell has rung.

X Coin 1	*Y* Coin 2	*W* witness	*P(X, Y, W)*
heads	heads	1	0.250
tails	heads	1	0.250

$$P(X = heads|W = 1) = \frac{0.25 + 0.25}{0.25 + 0.25 + 0.25 + 0.125} = 0.571$$

$$P(X = heads|Y = heads, W = 1) = \frac{0.25}{0.25 + 0.25} = 0.5$$

$$P(X = heads|Y = heads, W = 1) \neq P(X = heads|W = 1)$$

Therefore, we conclude that *X* and *Y* are dependent when conditioning on *W* = 1.

To summarize

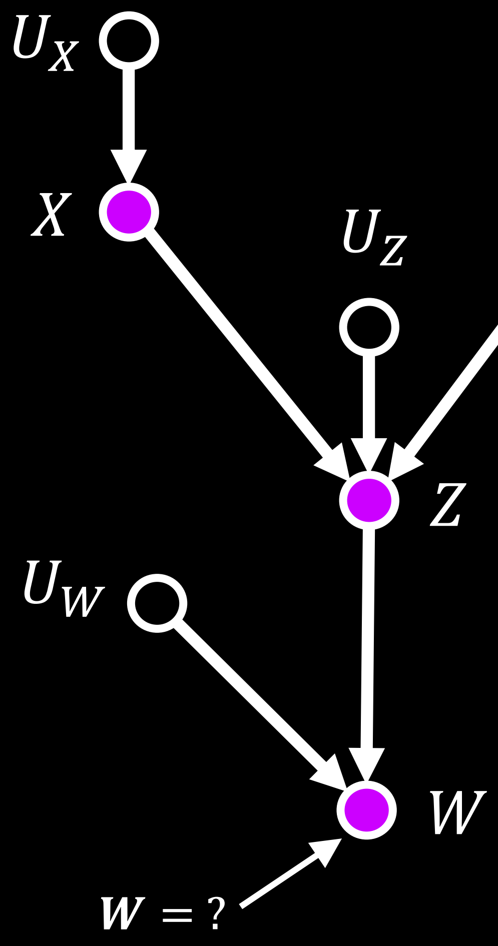


Figure 2.4

X and *Y* are independent when we do not know the witness report *W*, i.e., when the state of the variable *W* is unknown (not given) (*W* = ?)

X and *Y* are dependent when we know that the witness reports the bell rings (*W* = 1), i.e., *X* and *Y* become dependent after we know that the witness reports the bell is ringing (*W* = 1).

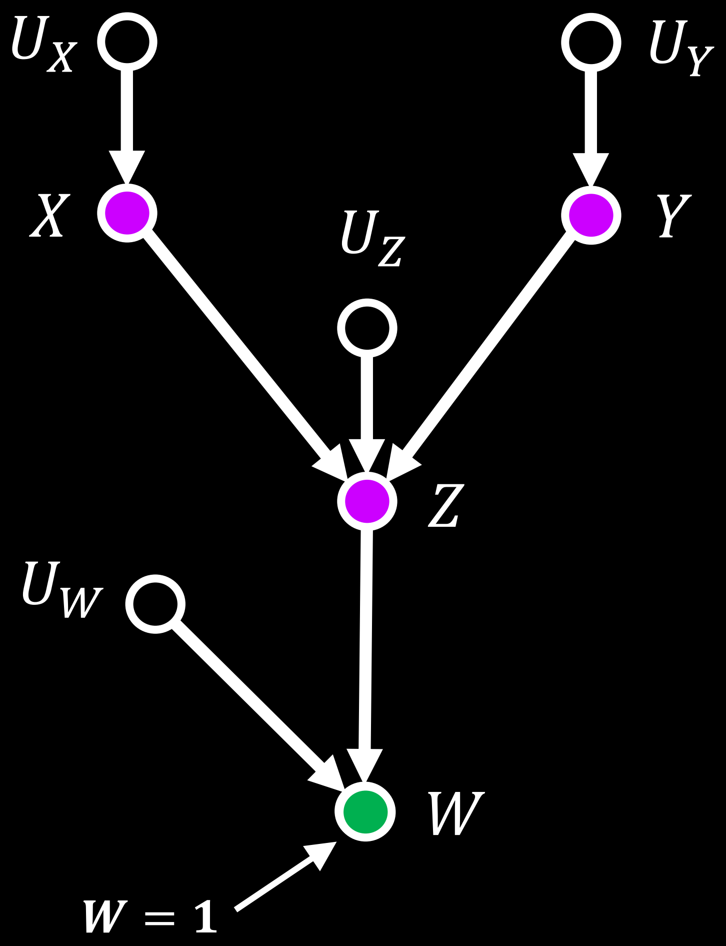


Figure 2.4

All these considerations lead us to the third rule:

Rule 3 (Conditional Independence in Colliders)

If a variable Z is the collision node between two variables X and Y, and there is only one path between Y and X, then X and Y are unconditionally independent but are dependent conditional on Z and any descendants of Z.

Extremely important to the study of causality, i.e., it allows to:

- test whether a causal model could have generated a data set
- discover models from data
- fully resolve the Simpson's paradox by determining which variables to measure
- estimate causal effect under **confounding**

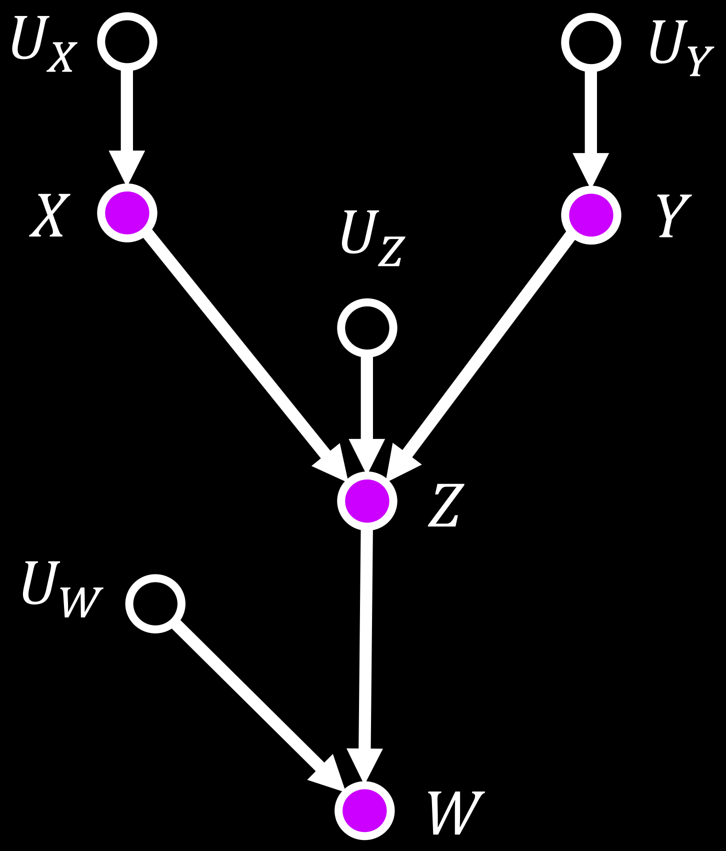


Figure 2.4

Inquisitive students may wonder why it is that dependencies associated with conditioning on a collider are so surprising to most people — as in, for example, the Monty Hall example.

The reason is that humans tend to associate dependence with causation.

Accordingly, they assume (wrongly) that statistical dependence between two variables can only exist if there is a causal mechanism that generates such dependence; that is, either one of the variables causes the other or a third variable causes both.

In the case of a collider, they are surprised to find a dependence that is created in a third way, thus **violating the assumption** of "**no correlation without causation.**"

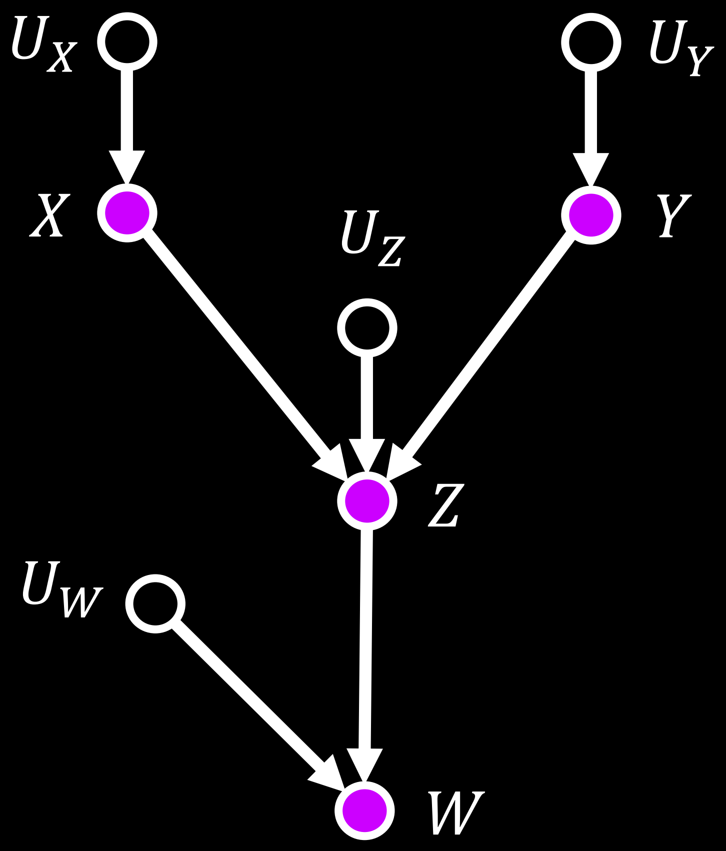


Figure 2.4

Causal models are generally not as simple as the cases we have examined so far. Specifically, it is rare for a graphical model to consist of a single path between variables.

In most graphical models, pairs of variables will have multiple possible paths connecting them, and each path will traverse a variety of chains, forks, and colliders.

The questions remains whether there is a criterion or process that can be applied to a graphical causal model of any complexity in order to predict dependencies that are shared by all data sets generated by that graph.

Rule 1 + Rule 2 + Rule 3 = d-separation

Allows, to determine, for **any pair of nodes**, whether the nodes are **d-connected**, meaning there exists a connecting path between them, or **d-separated**, meaning there exists no such path.

nodes *X* and *Y* are **d-separated**

↓

variables *X* and *Y* are independent

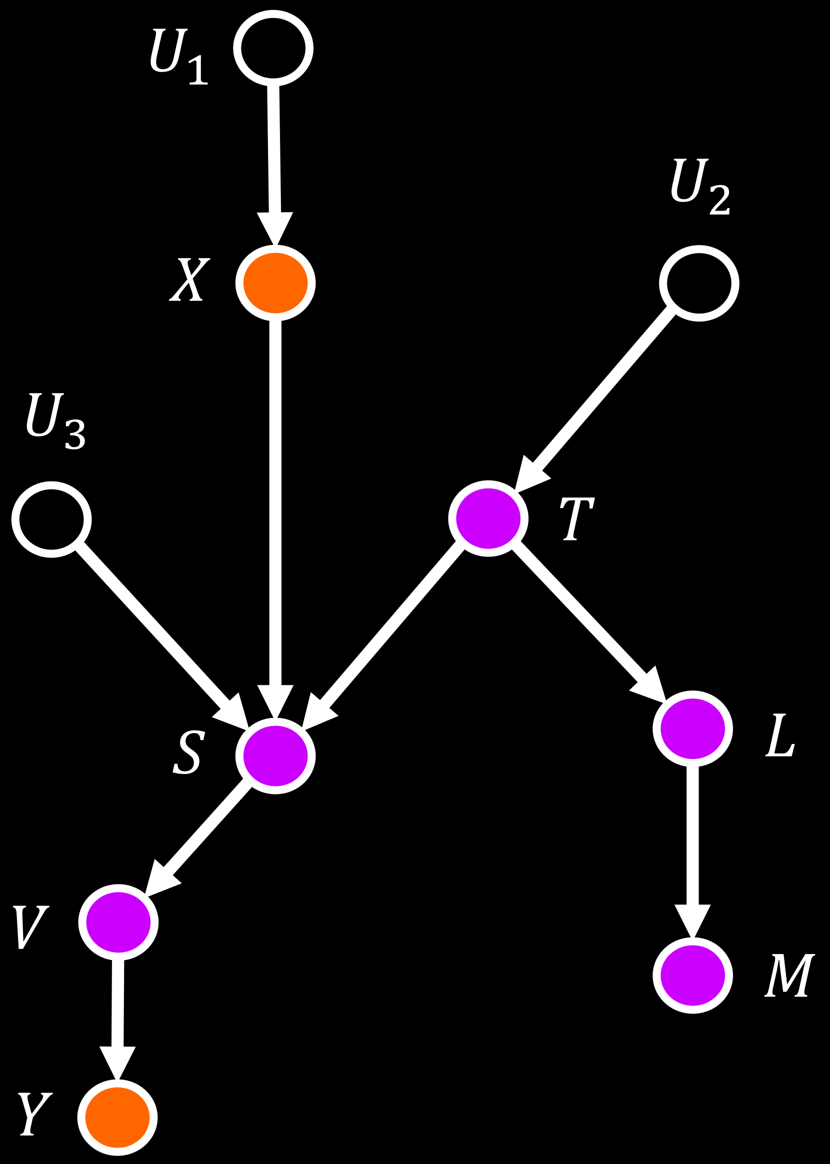
nodes *X* and *Y* are **d-connected**

↓

variables *X* and *Y* are possibly, or most likely dependent

Two nodes ***X*** and ***Y*** are **d-separated** if every path between them (should any exist) is **blocked**.

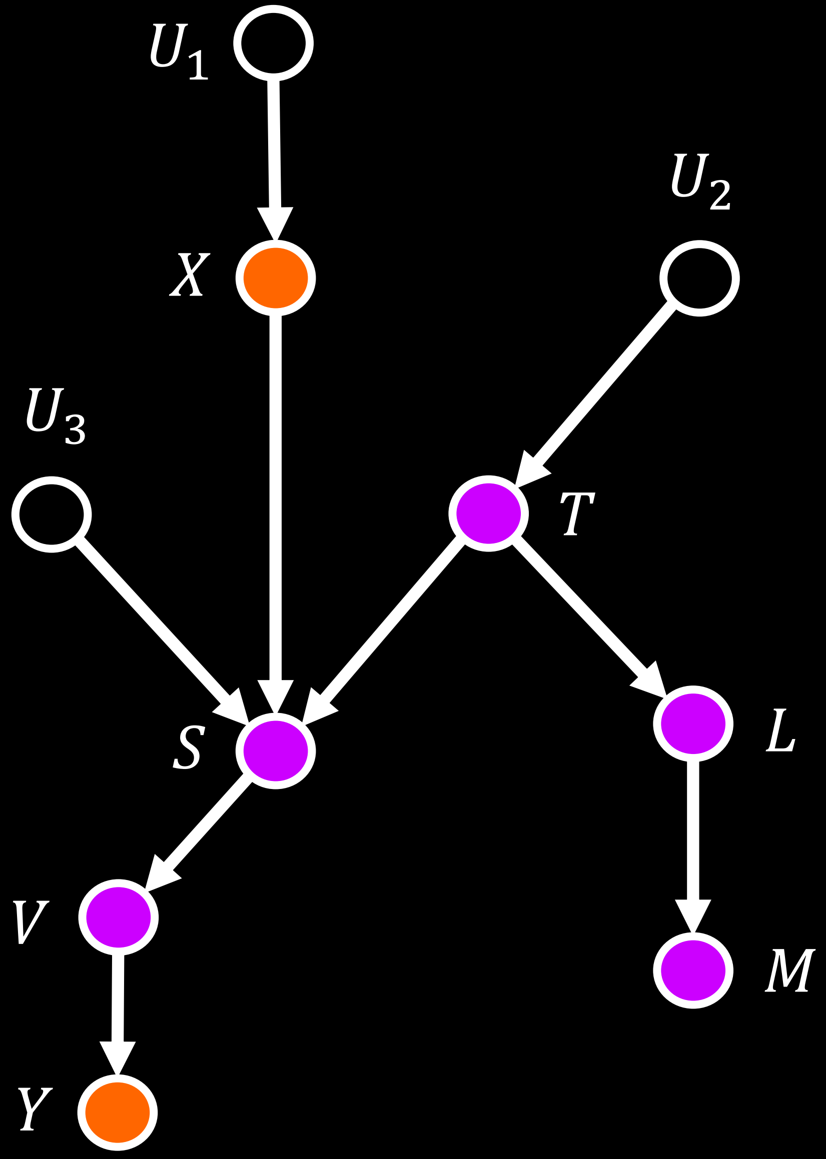
If even one path between *X* and *Y* is **unblocked**, ***X*** and ***Y*** are **d-connected**.



Two nodes ***X*** and ***Y*** are **d-separated** if every path between them (should any exist) is **blocked**.

If even one path between *X* and *Y* is **unblocked**, ***X*** and ***Y*** are **d-connected**.

- ***X*** and ***Y*** are **d-separated** if **the following path**



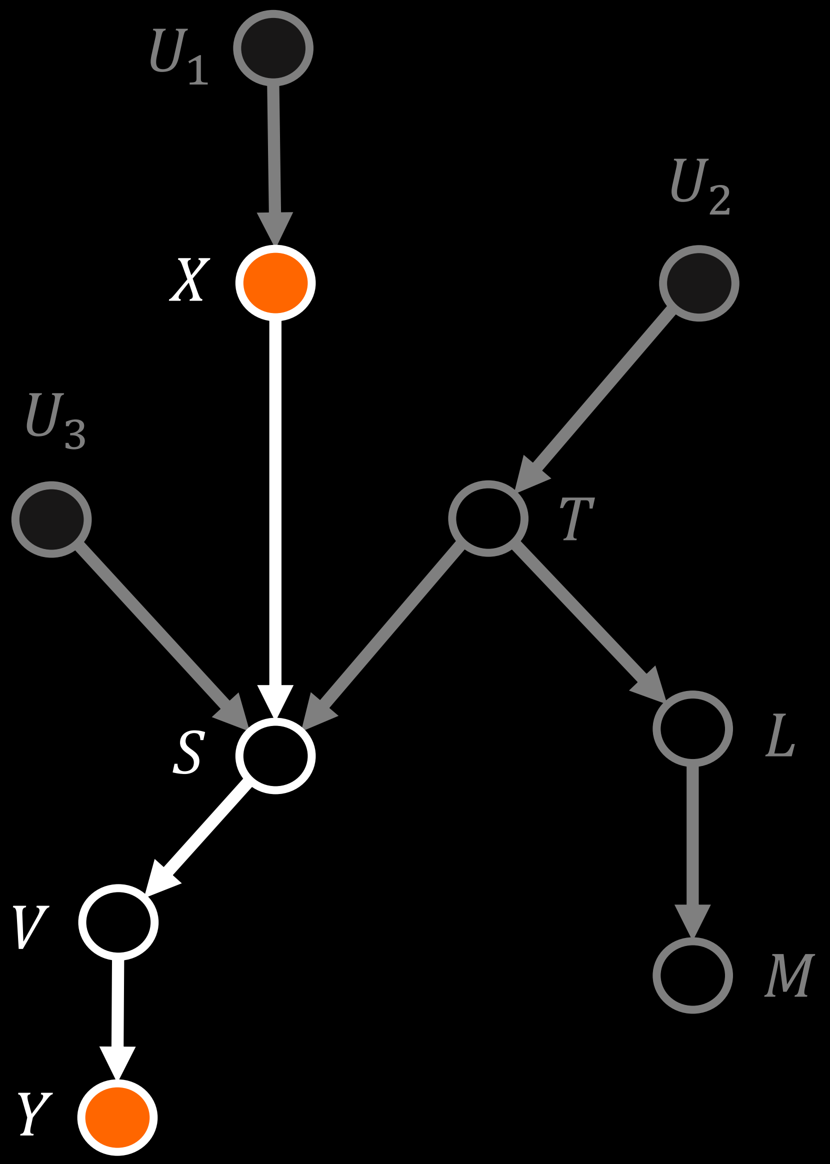
Two nodes ***X*** and ***Y*** are **d-separated** if every path between them (should any exist) is **blocked**.

If even one path between *X* and *Y* is **unblocked**, ***X*** and ***Y*** are **d-connected**.

- ***X*** and ***Y*** are **d-separated** if **the following path**

$$X \rightarrow S \rightarrow V \rightarrow Y$$

  is **blocked**.



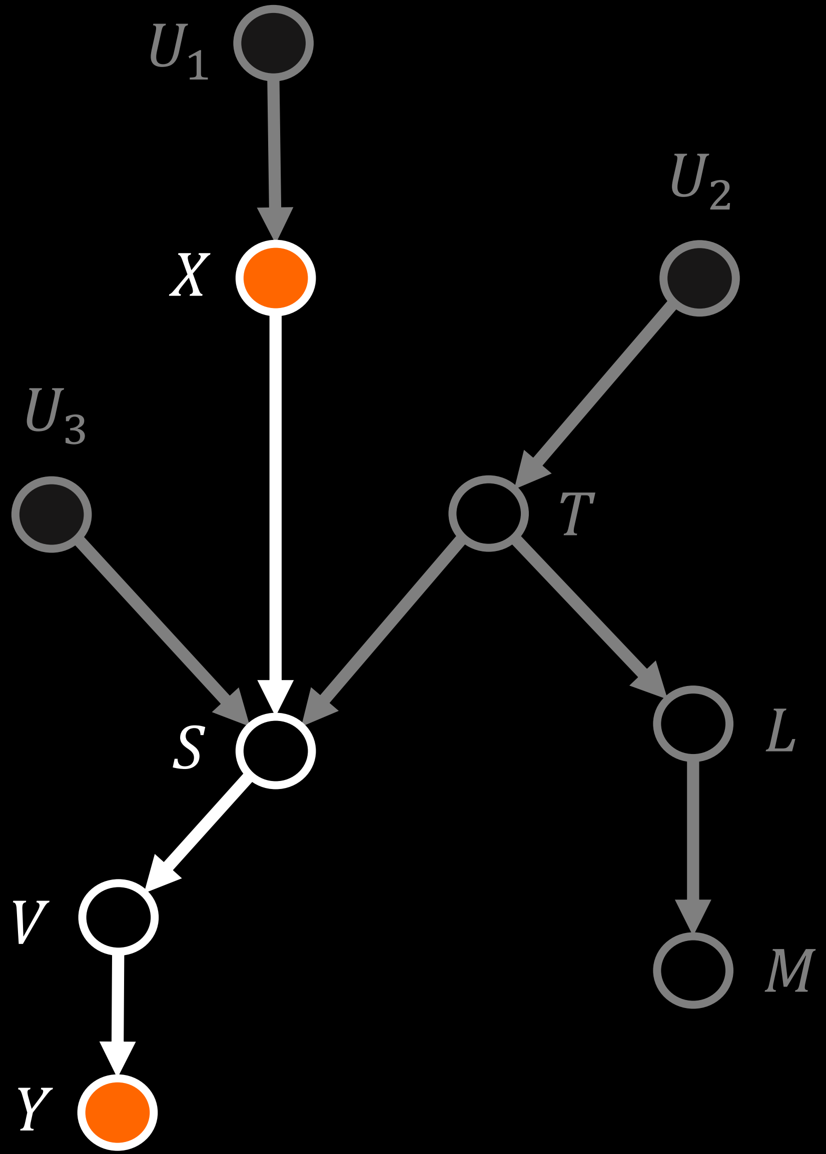
Two nodes ***X*** and ***Y*** are **d-separated** if every path between them (should any exist) is **blocked**.

If even one path between *X* and *Y* is **unblocked**, ***X*** and ***Y*** are **d-connected**.

Water flows through Pipes

Path = Pipe

Dependence = Water

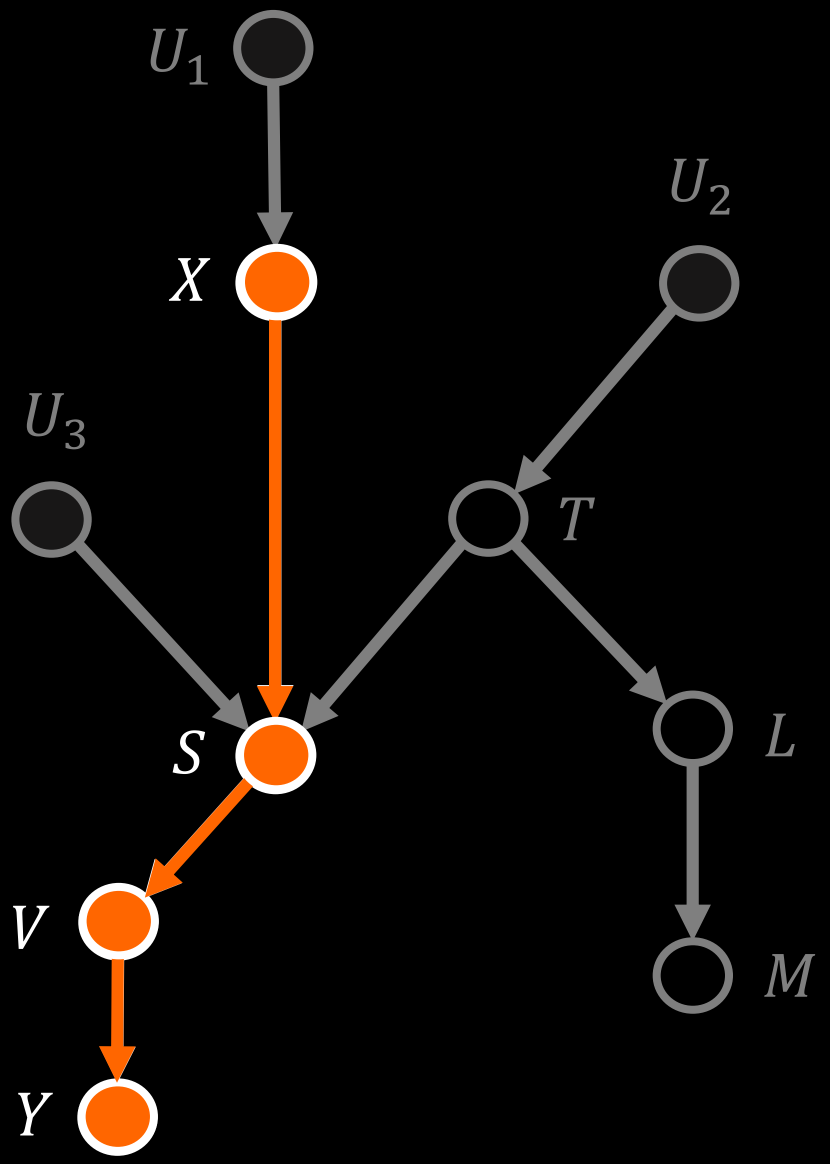


If even one pipe in unblocked, some water can pass from one place to another, and if a single path is clear, the variables at either end will be dependent.

However, a pipe need only be blocked in one place to stop the flow of water through it, it takes only one node to block the passage of dependence in an entire path.

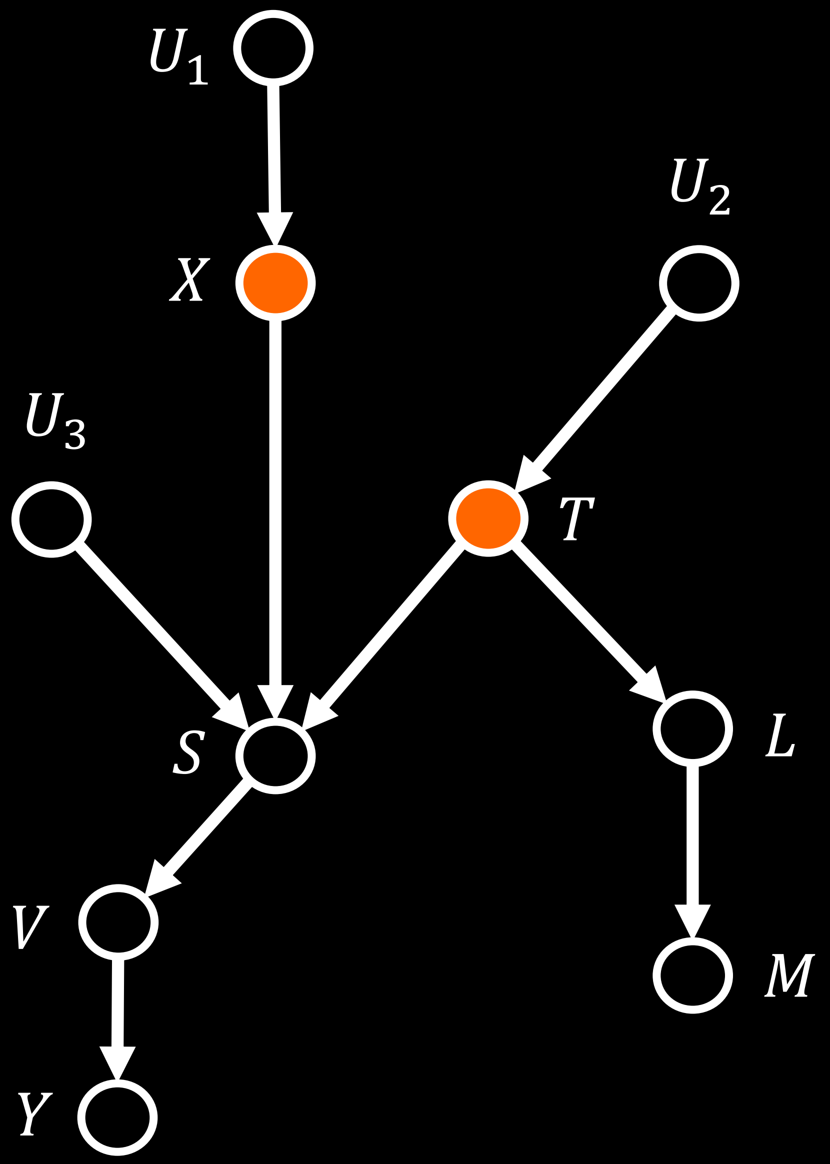
There are certain **kinds of nodes that can block a path**, depending on whether we are performing **unconditional or conditional d-separation**.

If we are not conditioning on any variable, then only colliders can block a path.



Unconditional d-separation

Consider nodes *X* and *T*



There are certain **kinds of nodes that can block a path**, depending on whether we are performing **unconditional or conditional d-separation**.

If we are not conditioning on any variable, then only colliders can block a path.

Unconditional d-separation

Consider nodes *X* and *T*

The path

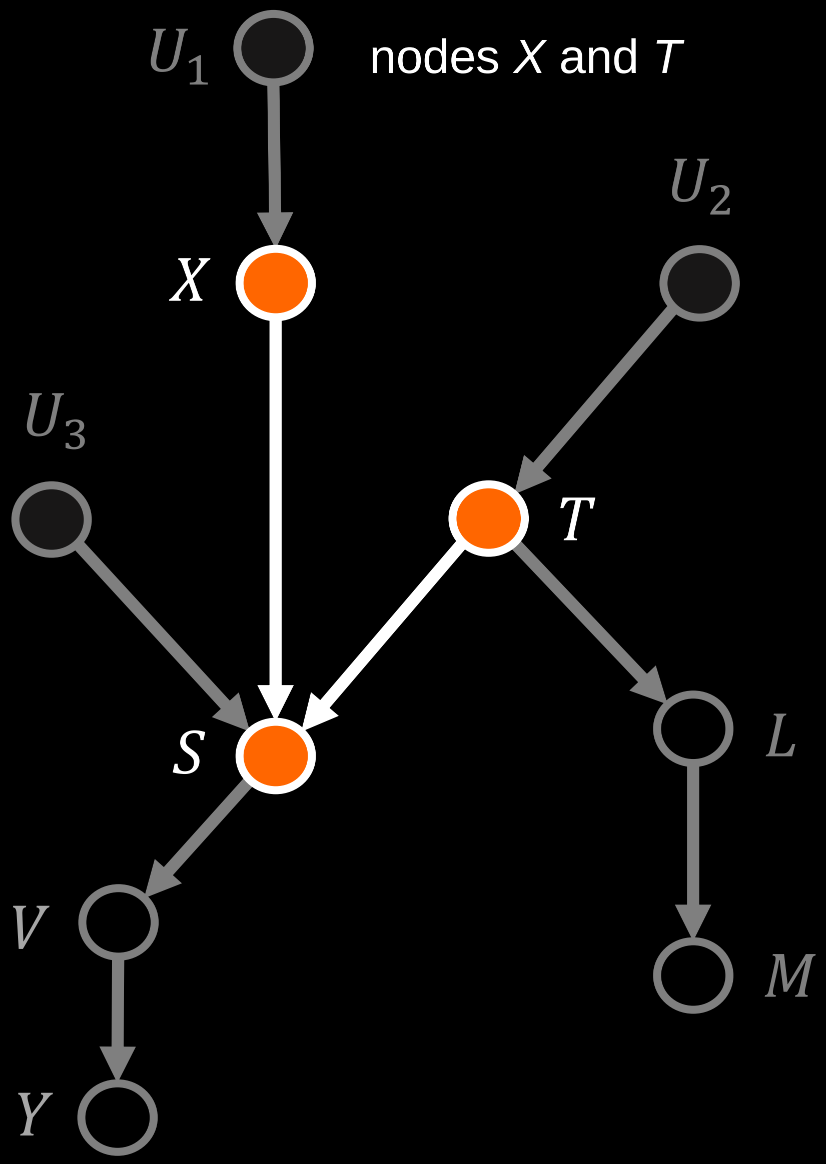
$$X \rightarrow S \leftarrow T$$

is blocked by collider *S*.

There are certain **kinds of nodes that can block a path**, depending on whether we are performing **unconditional or conditional d-separation**.

If we are not conditioning on any variable, then only colliders can block a path.

we consider nodes *X* and *T*

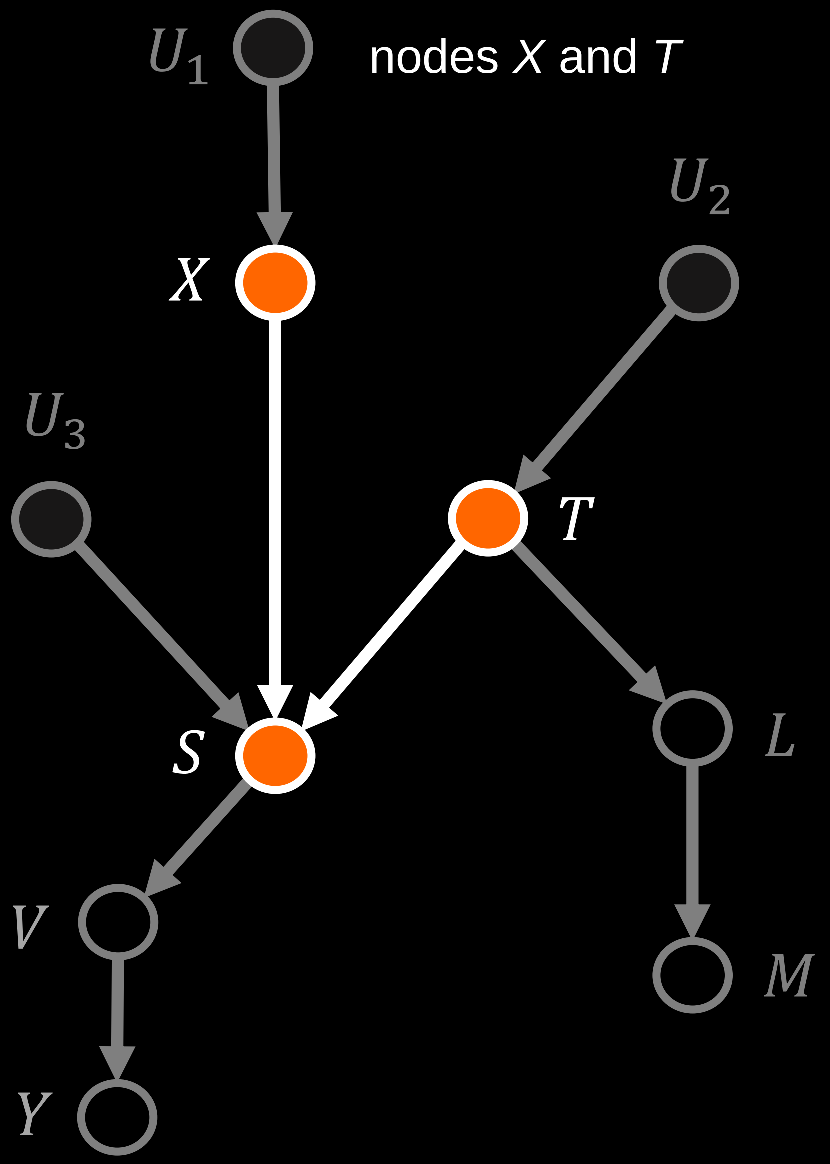


Unconditional d-separation

The reasoning for this is fairly straightforward as we saw in Section 2.3, unconditional dependence can't pass through a collider, i.e. the collider blocks the path.

So if every path between two nodes X and Y has a collider in it, then X and Y cannot be unconditionally dependent; they must be marginally independent.

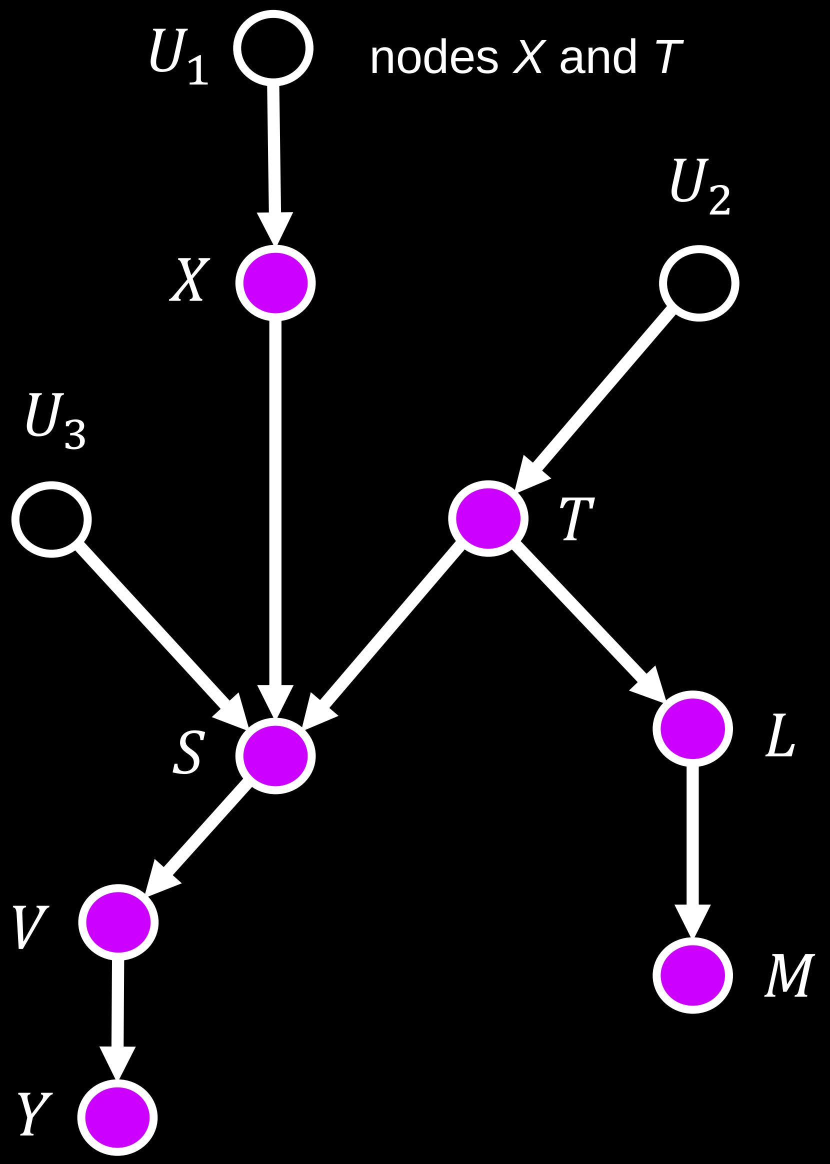
we consider nodes X and T



Conditional d-separation

If, however, we are conditioning on a set of nodes *Z*, then the following kinds of nodes can block a path:

- a collider that is not conditioned on (I.e., not in *Z*), and that has no descendants in *Z*.

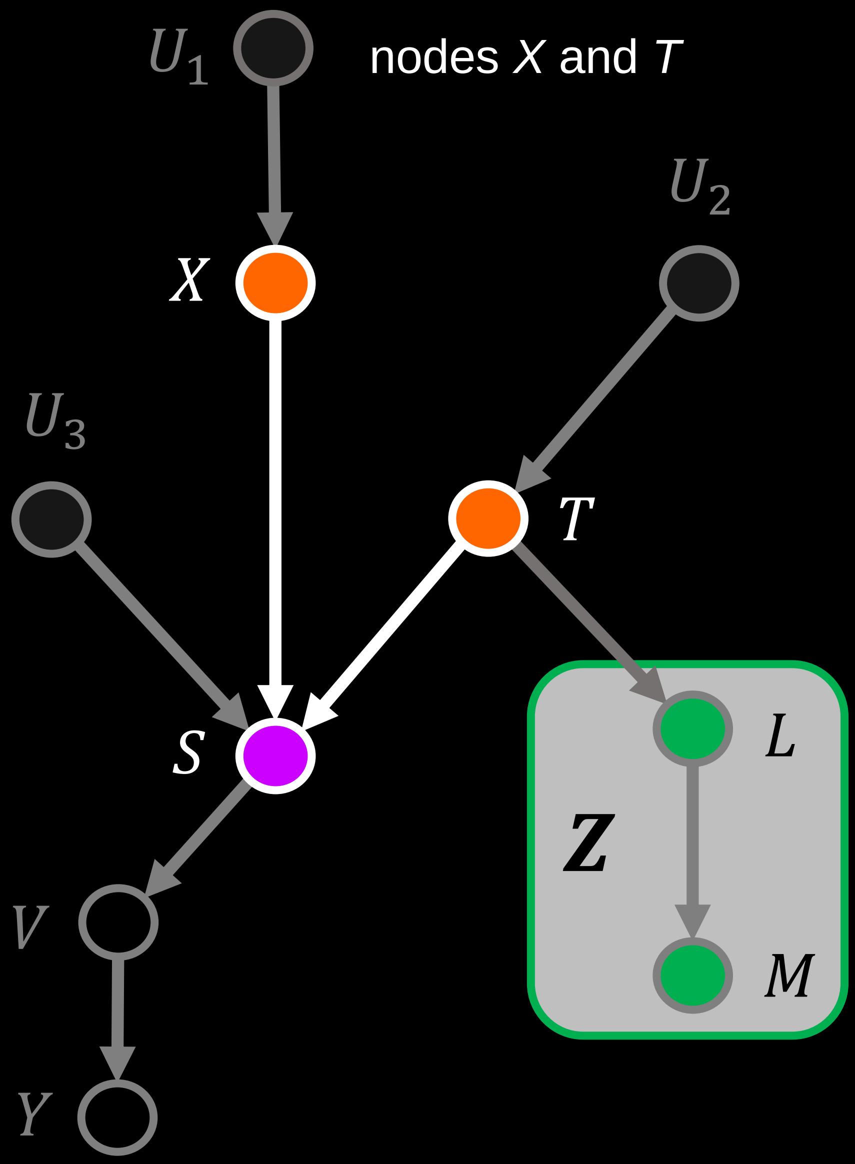


Conditional d-separation

If, however, we are conditioning on a set of nodes *Z*, then the following kinds of nodes can block a path:

- a collider that is not conditioned on (I.e., not in *Z*), and that has no descendants in *Z*.

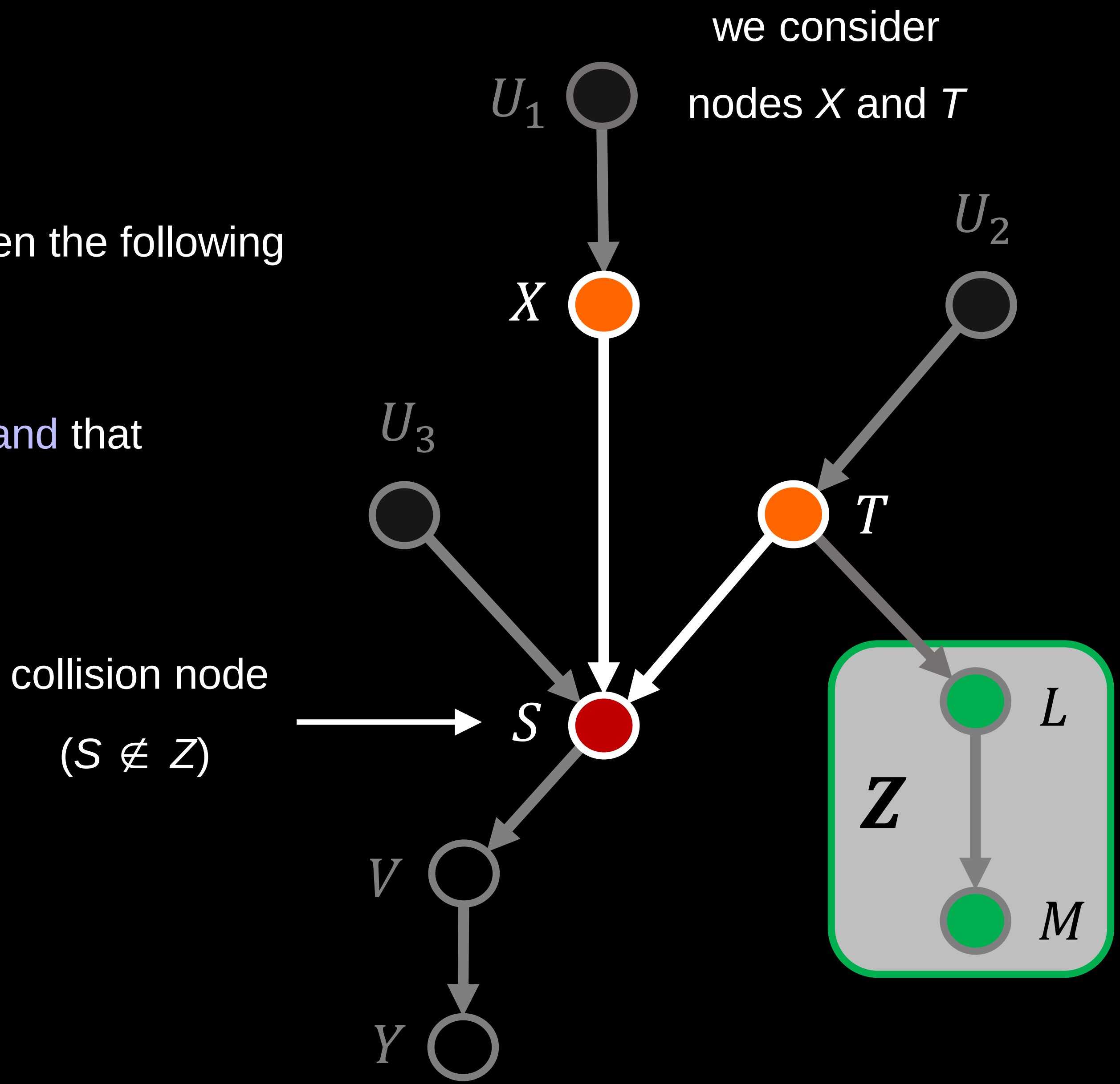
we consider nodes *X* and *T*



Conditional d-separation

If, however, we are conditioning on a set of nodes Z, then the following kinds of nodes can block a path:

- a collider that is not conditioned on (I.e., not in Z), and that has no descendants in Z.



Conditional d-separation

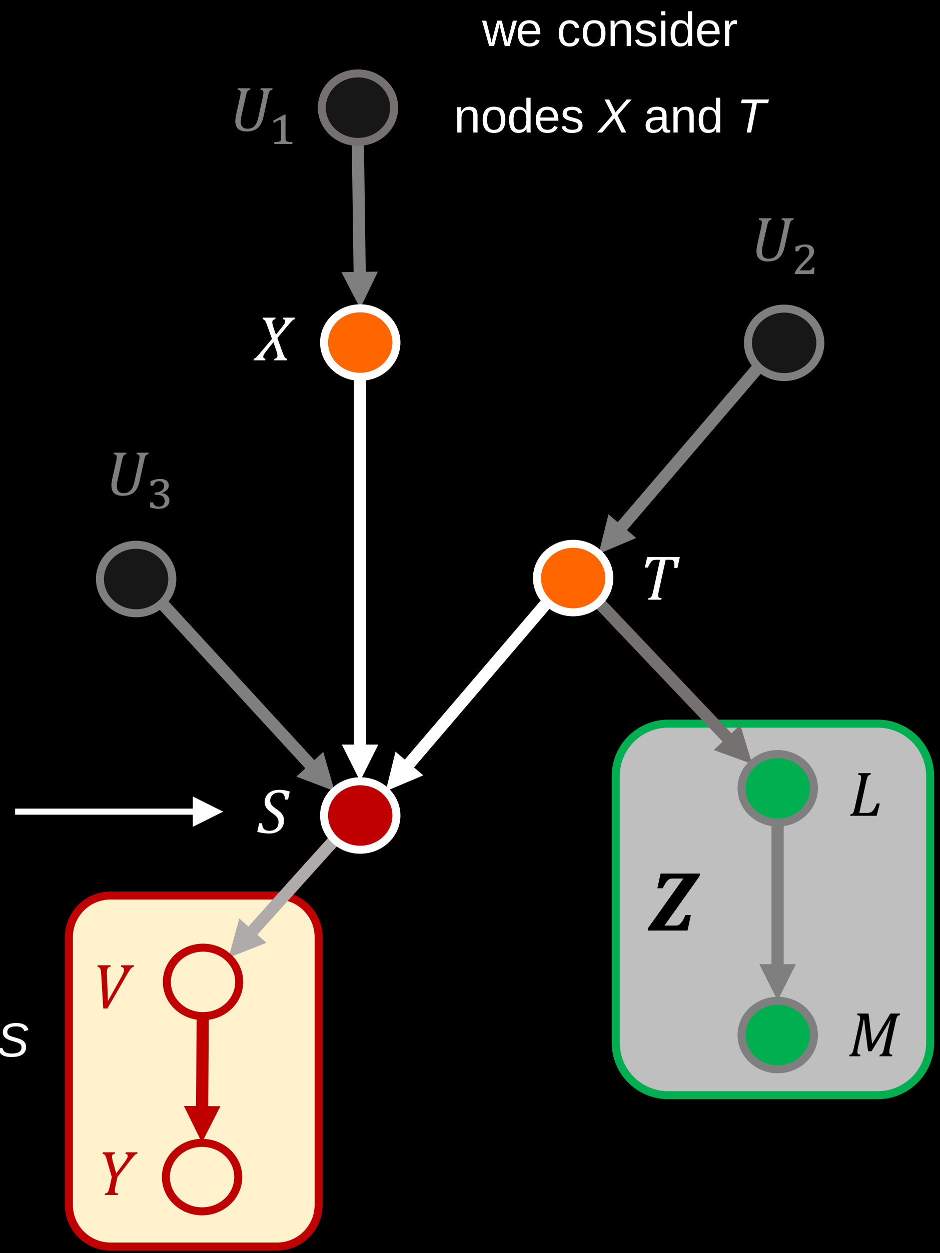
If, however, we are conditioning on a set of nodes Z, then the following kinds of nodes can block a path:

- a collider that is not conditioned on (I.e., not in Z), and that has no descendants in Z.

S blocks the path $X \rightarrow S \leftarrow T$

collision node ($S \notin Z$)

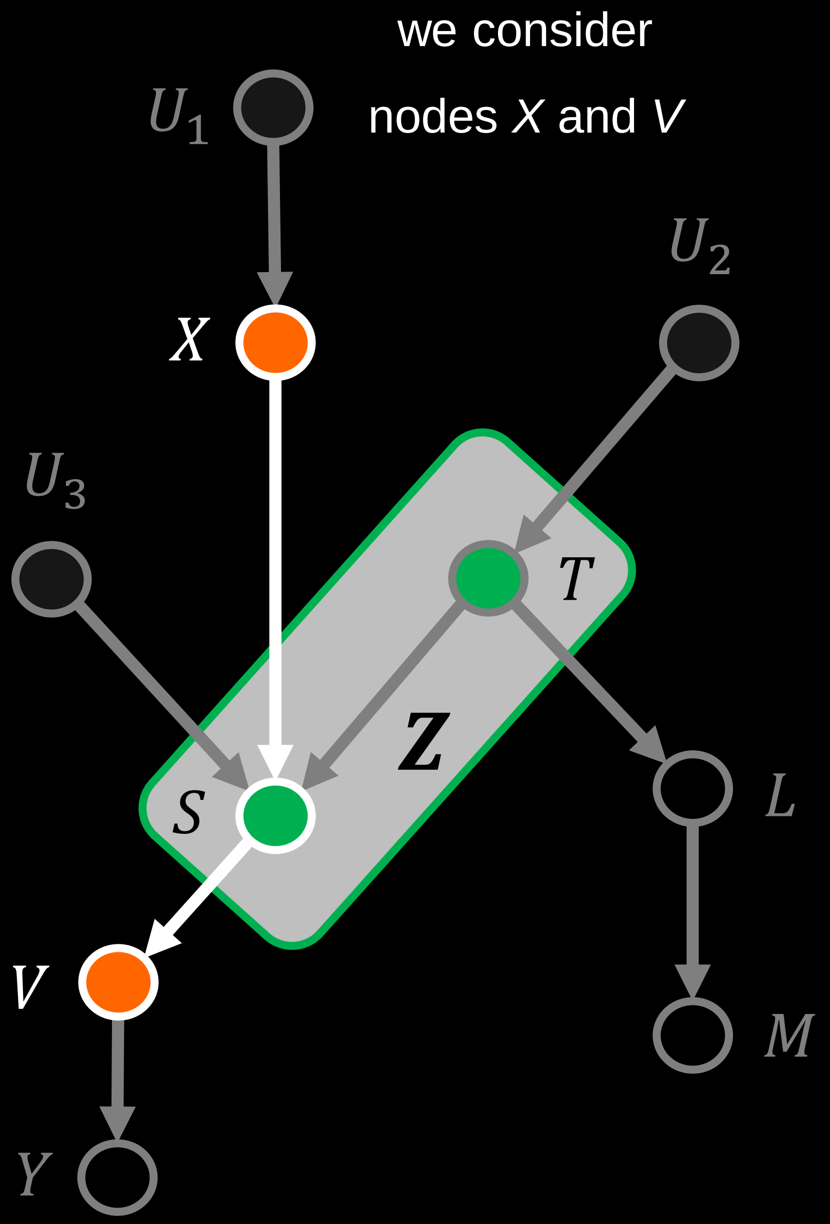
descendants of S ($V, Y \notin Z$)



Conditional d-separation

If, however, we are conditioning on a set of nodes *Z*, then the following kinds of nodes can block a path:

- a collider that is not conditioned on (I.e., not in *Z*), and that has no descendants in *Z*.
- a chain or fork whose middle node is in *Z*

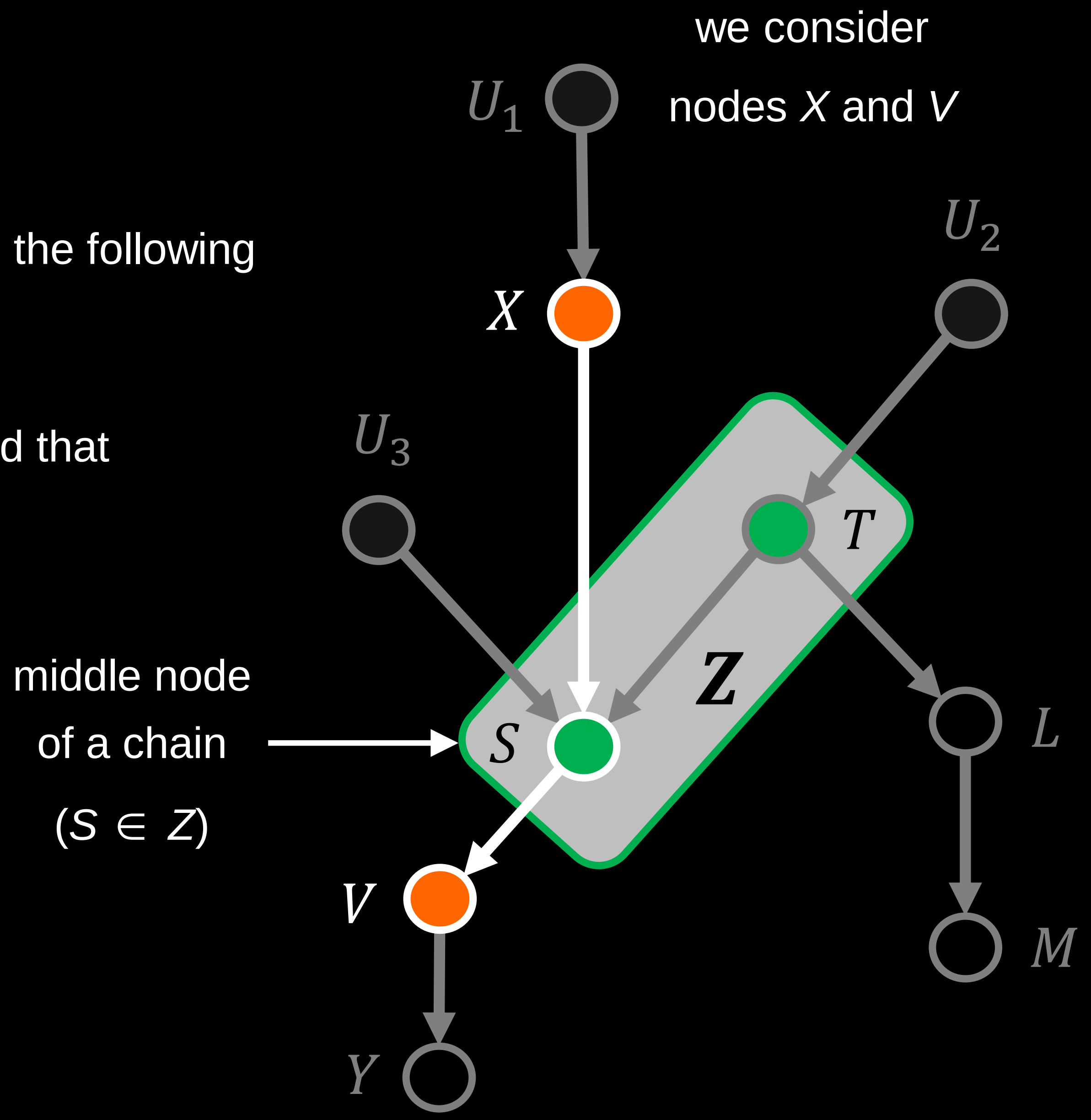


Conditional d-separation

If, however, we are conditioning on a set of nodes *Z*, then the following kinds of nodes can block a path:

- a collider that is not conditioned on (I.e., not in *Z*), and that has no descendants in *Z*.
- a chain or fork whose middle node is in *Z*

S blocks the path $X \to S \to V$

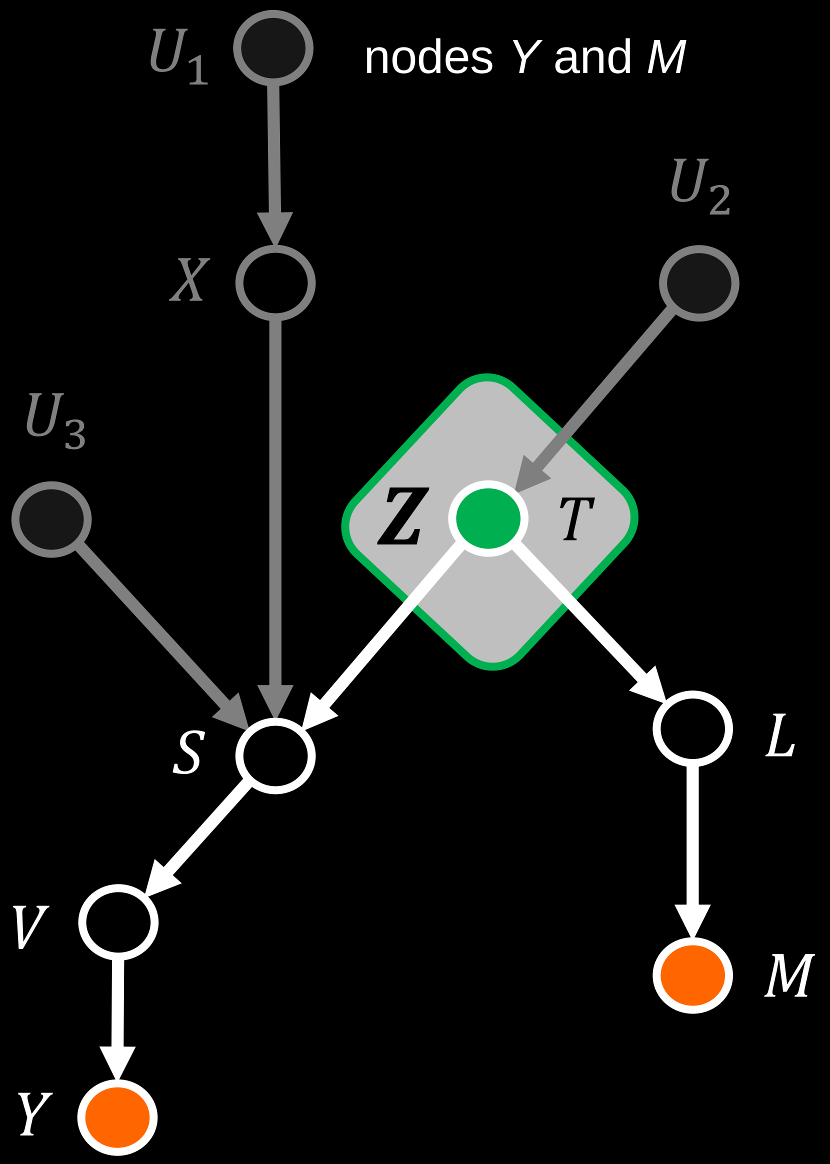


Conditional d-separation

If, however, we are conditioning on a set of nodes *Z*, then the following kinds of nodes can block a path:

- a collider that is not conditioned on (I.e., not in *Z*), and that has no descendants in *Z*.
- a chain or fork whose middle node is in *Z*

we consider nodes *Y* and *M*

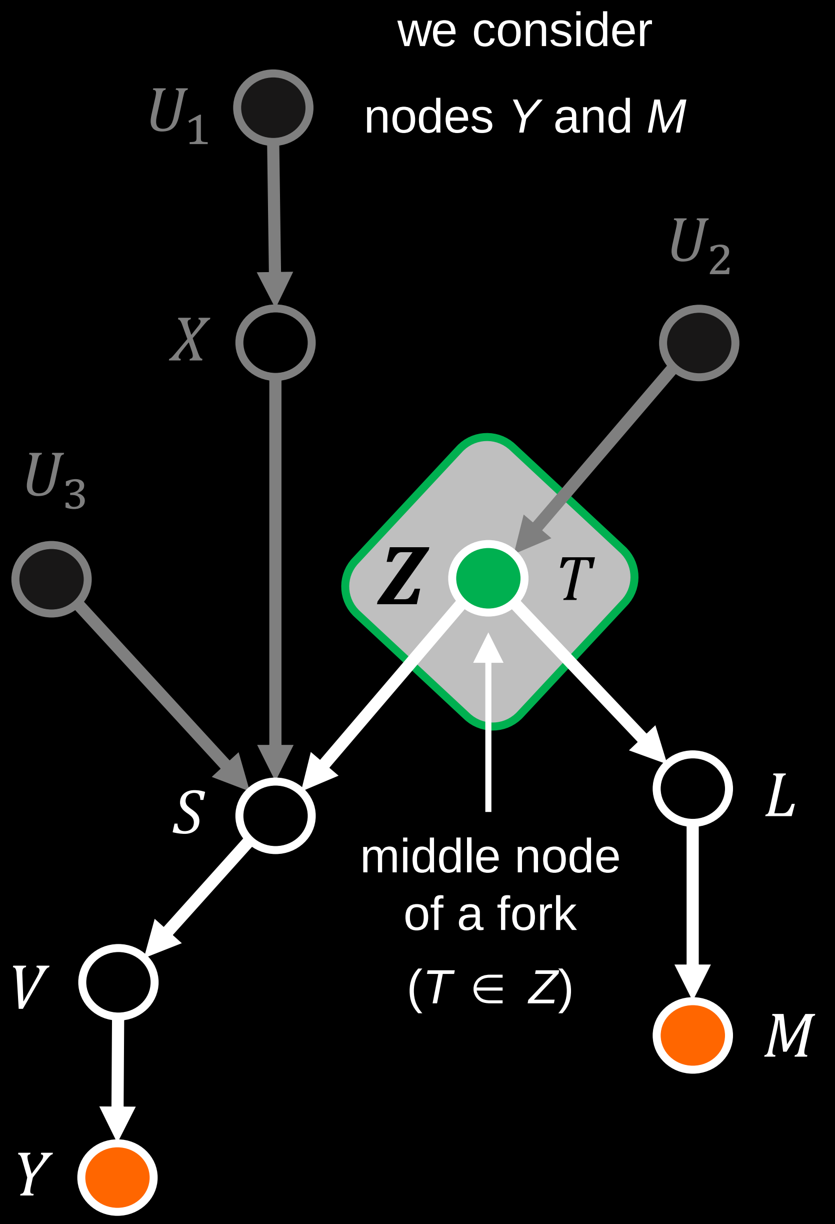


Conditional d-separation

If, however, we are conditioning on a set of nodes *Z*, then the following kinds of nodes can block a path:

- a collider that is not conditioned on (I.e., not in *Z*), and that has no descendants in *Z*.
- a chain or fork whose middle node is in *Z*

T blocks the path $Y \leftarrow V \leftarrow S \leftarrow T \rightarrow L \rightarrow M$

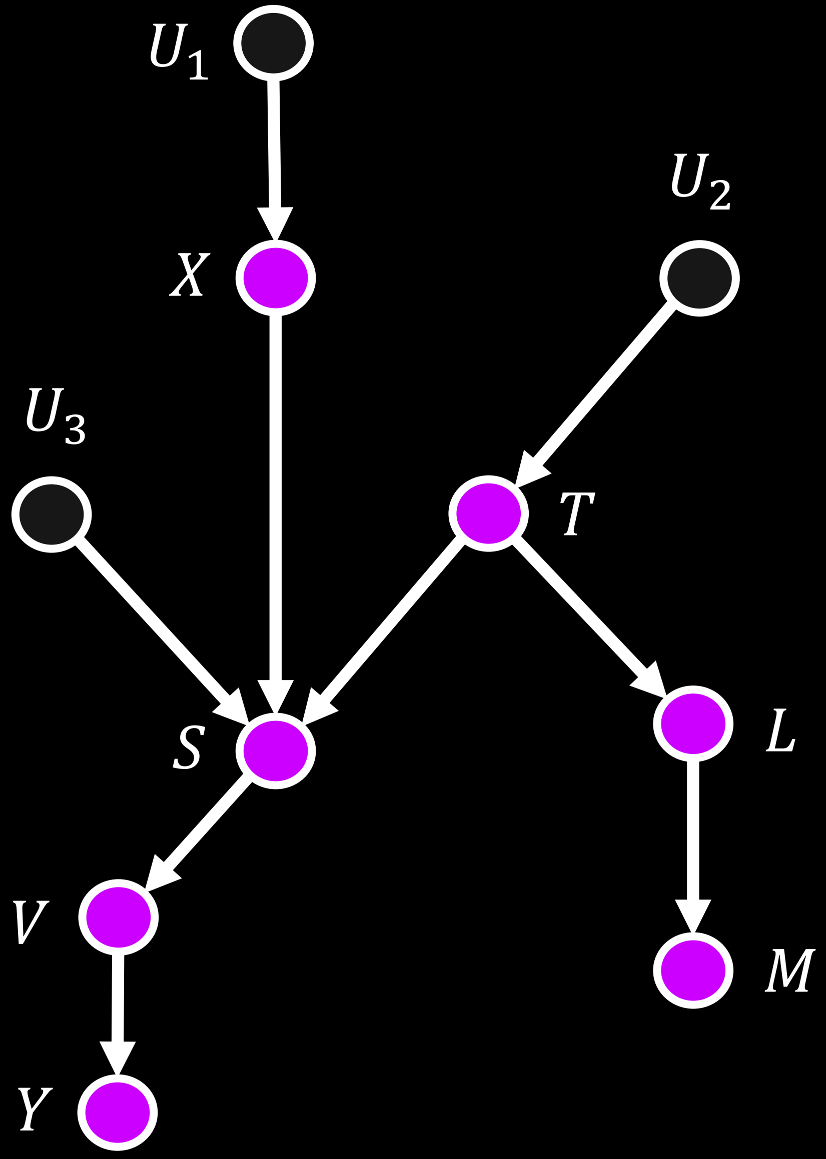


The reasoning behind these points goes back to what we learned in Sections 2.2 and 2.3.

- A collider does not allow dependence to flow between its parents, thus blocking the path,
- but Rule 3 tells us that when we condition on a collider or its descendants, the parent nodes may become dependent.

So

- a collider whose collision node is not in the conditioning set **Z** would block dependence from passing through a path,



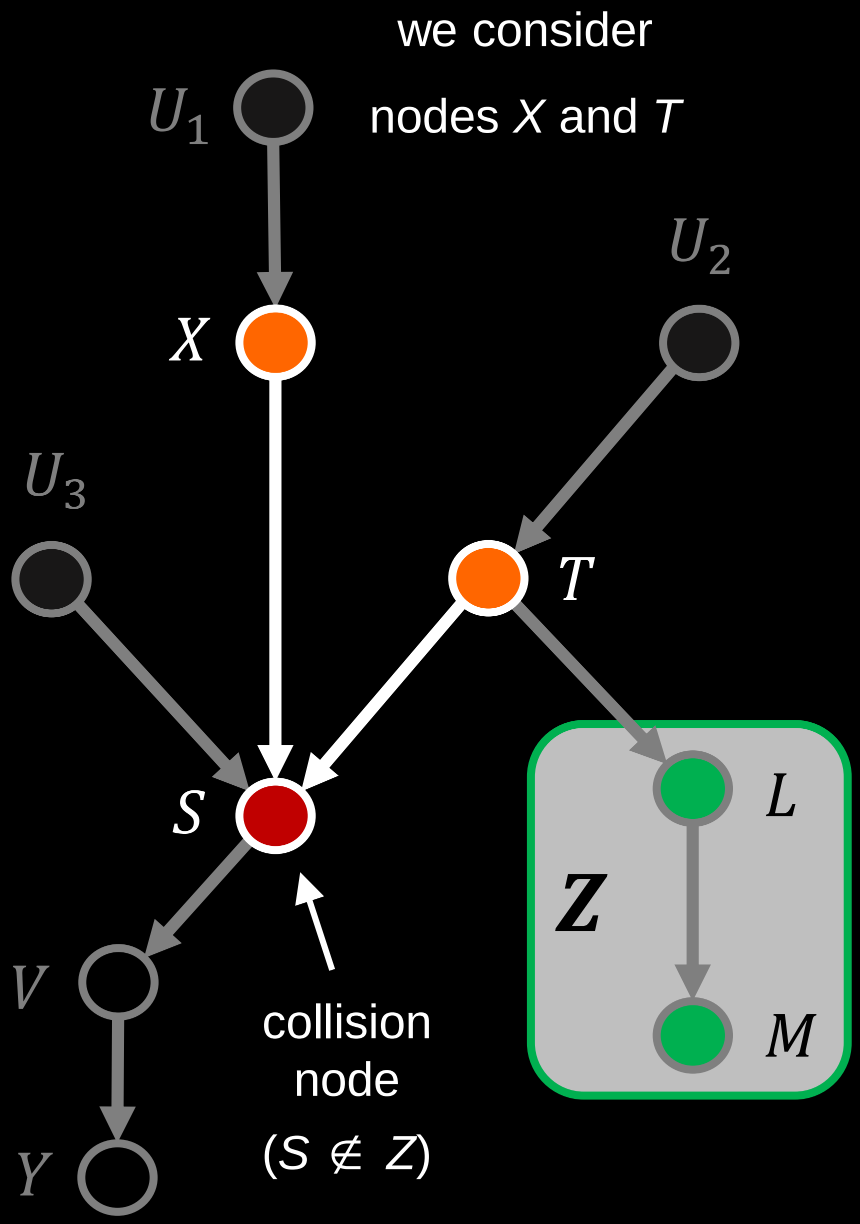
The reasoning behind these points goes back to what we learned in Sections 2.2 and 2.3.

- A collider does not allow dependence to flow between its parents, thus blocking the path,
- but Rule 3 tells us that when we condition on a collider or its descendants, the parent nodes may become dependent.

So

- a collider whose collision node is not in the conditioning set **Z** would block dependence from passing through a path,

S blocks the path $X \rightarrow S \leftarrow T$



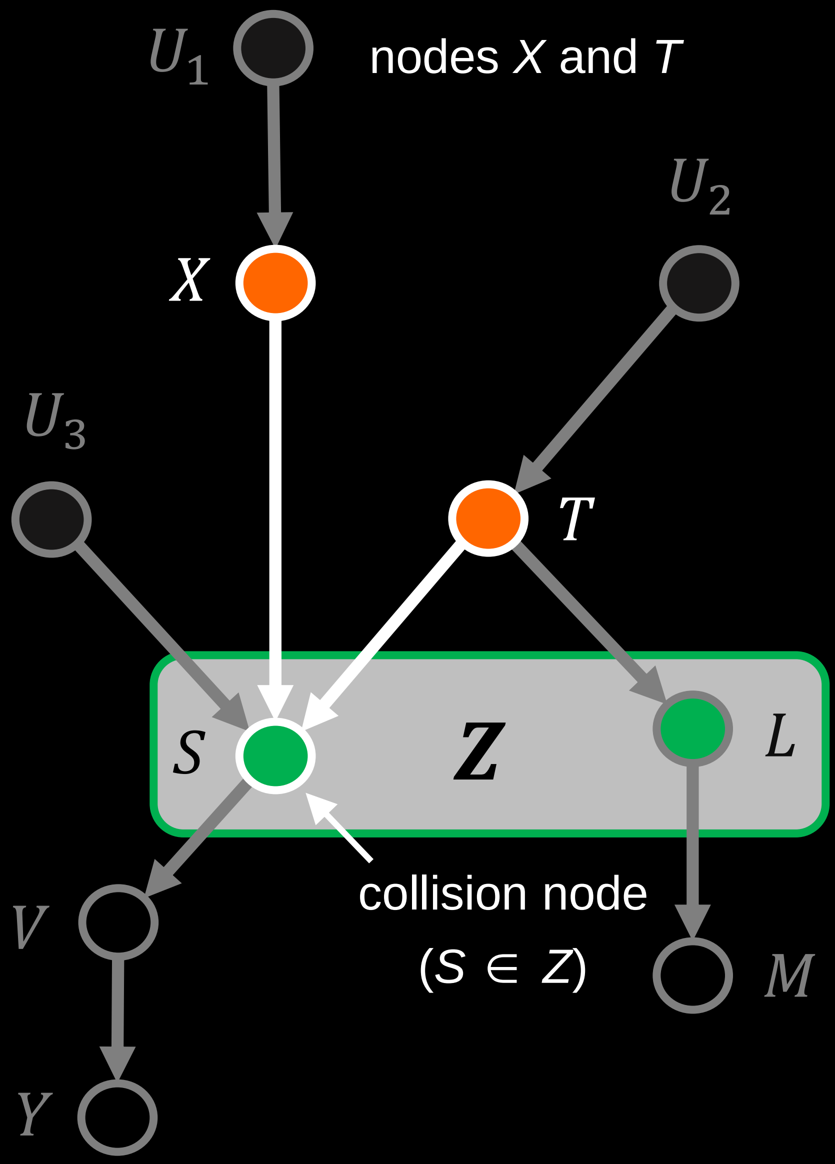
The reasoning behind these points goes back to what we learned in Sections 2.2 and 2.3.

- A collider does not allow dependence to flow between its parents, thus blocking the path,
- but Rule 3 tells us that when we condition on a collider or its descendants, the parent nodes may become dependent.

So

S does not block (opens) the path $X \rightarrow S \leftarrow T$

- a collider whose collision node or its descendants, is in the **conditioning set *Z*** would not block dependence passing through a path.



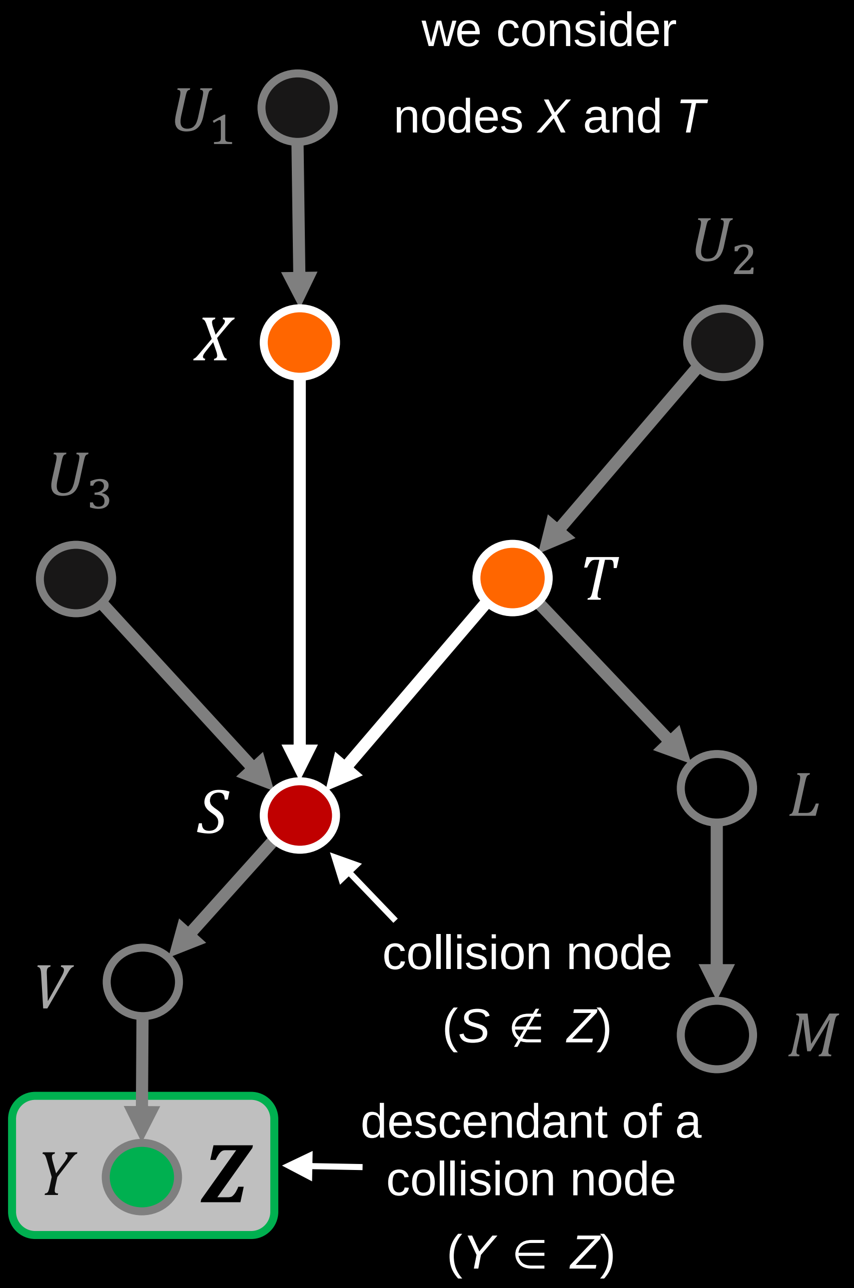
The reasoning behind these points goes back to what we learned in Sections 2.2 and 2.3.

- A collider does not allow dependence to flow between its parents, thus blocking the path,
- but Rule 3 tells us that when we condition on a collider or its descendants, the parent nodes may become dependent.

So

S does not block (opens) the path $X \rightarrow S \leftarrow T$

- a collider whose collision node or its descendants, is in the **conditioning set Z** would not block dependence passing through a path.

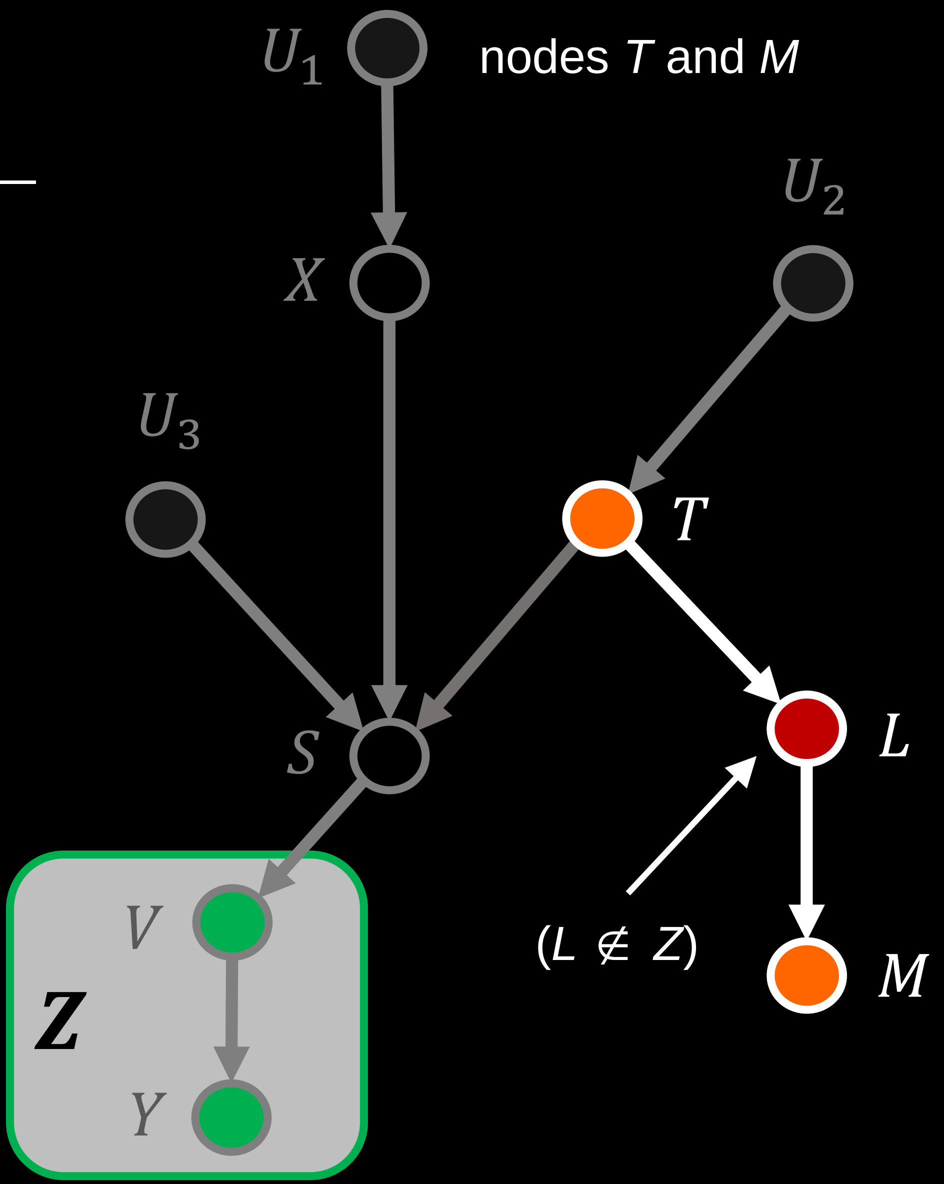


Conversely,

- dependence can pass through **noncolliders** — chains and forks —

L does not block the path $T \rightarrow L \leftarrow M$

dependence can pass from T to M and vice versa ← not conditioning on L ($L \notin Z$)

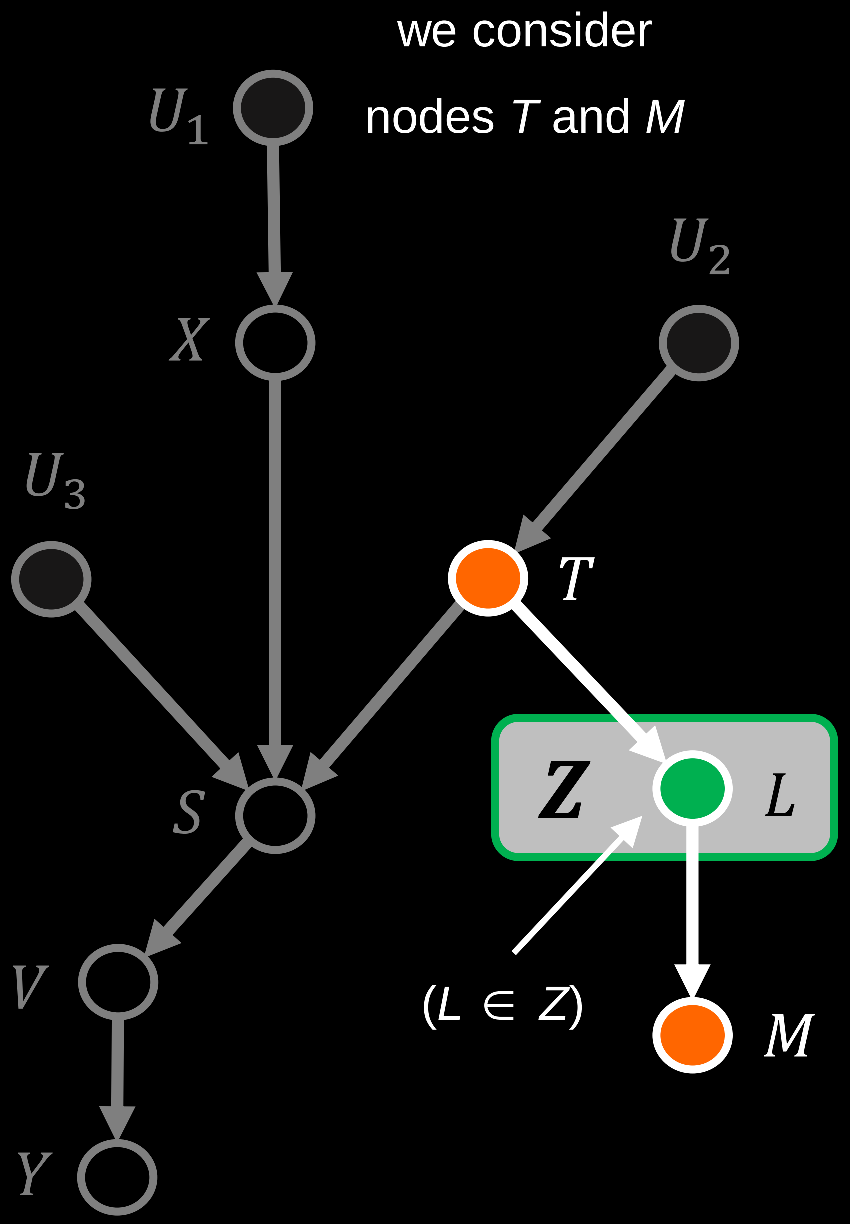


Conversely,

- dependence can pass through **noncolliders** — chains and forks —
- but **Rules 1 and 2** tell us that when we condition on them, the variables on either end of those paths become independent (when we consider one path at a time), and thus dependence can not pass through the path.

L blocks the path $T \rightarrow L \leftarrow M$

dependence can not pass from T to M and vice versa ← conditioning on L ($L \in Z$)



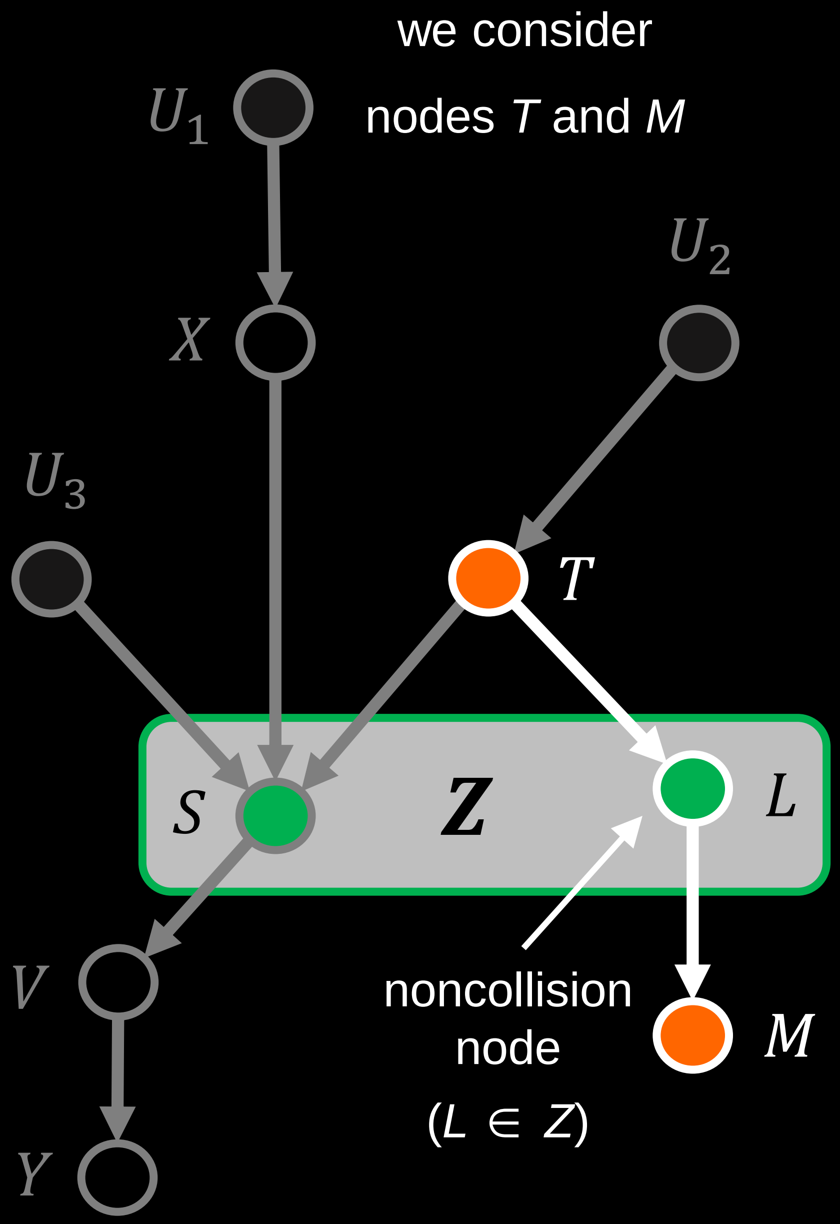
Conversely,

- dependence can pass through **noncolliders** — chains and forks —
- but **Rules 1 and 2** tell us that when we condition on them, the variables on either end of those paths become independent (when we consider one path at a time), and thus dependence can not pass through the path.

So

- any noncollision node in the conditioning set would block dependence,

L blocks the path $T \rightarrow L \leftarrow M$



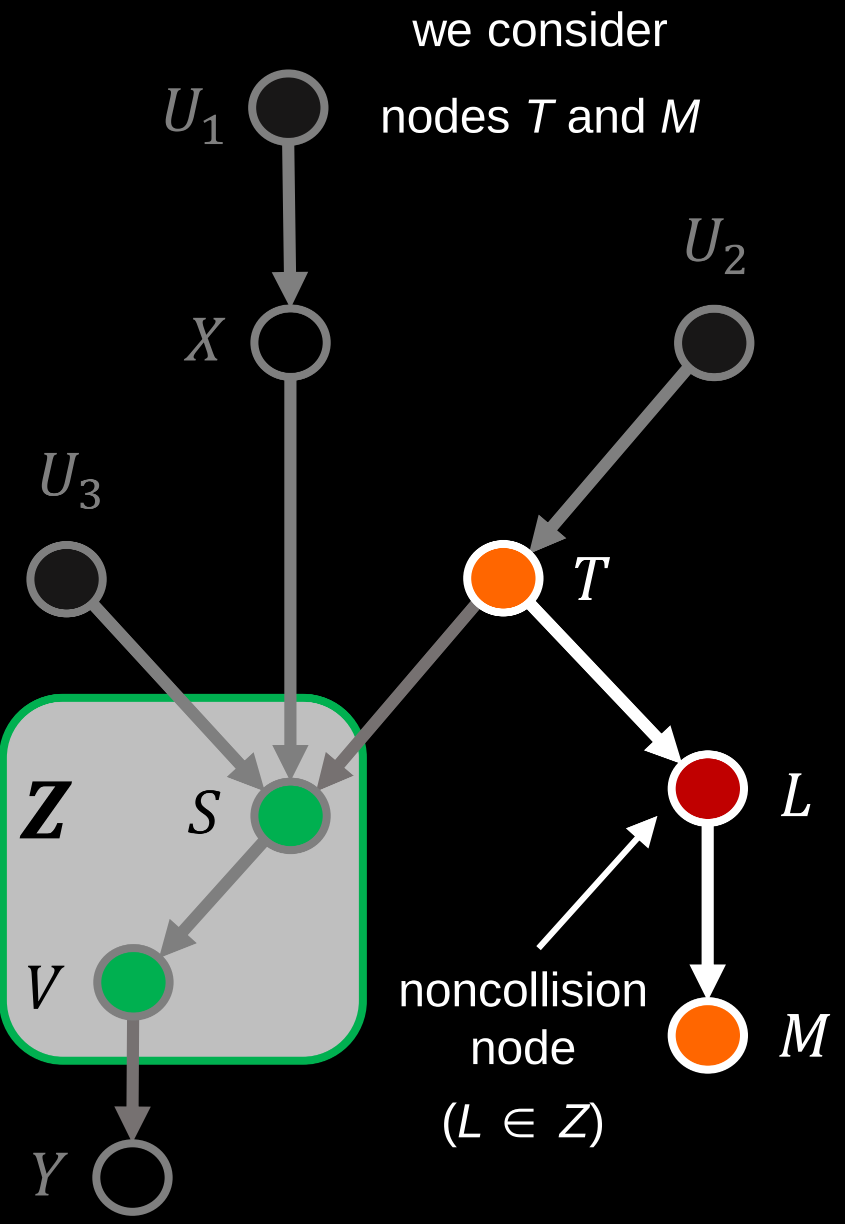
Conversely,

- dependence can pass through **noncolliders** — chains and forks —
- but **Rules 1 and 2** tell us that when we condition on them, the variables on either end of those paths become independent (when we consider one path at a time), and thus dependence can not pass through the path.

So

L does not block the path $T \rightarrow L \leftarrow M$

- any noncollision node that is not in the conditioning set would allow dependence through.



Definition 2.4.1 (d-separation)

A path *p* is blocked by a set of nodes *Z* if and only if

1. *p* contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node *B* is in *Z* (i.e., is conditioned on), or
2. *p* contains a collider $A \rightarrow B \leftarrow C$ such that the collision node *B* is not in *Z*, and no descendant of *B* is in *Z*.

If *Z* blocks every path between two nodes *X* and *Y*, then *X* and *Y* are d-separated, conditional on *Z*, and thus are independent conditional on *Z*.

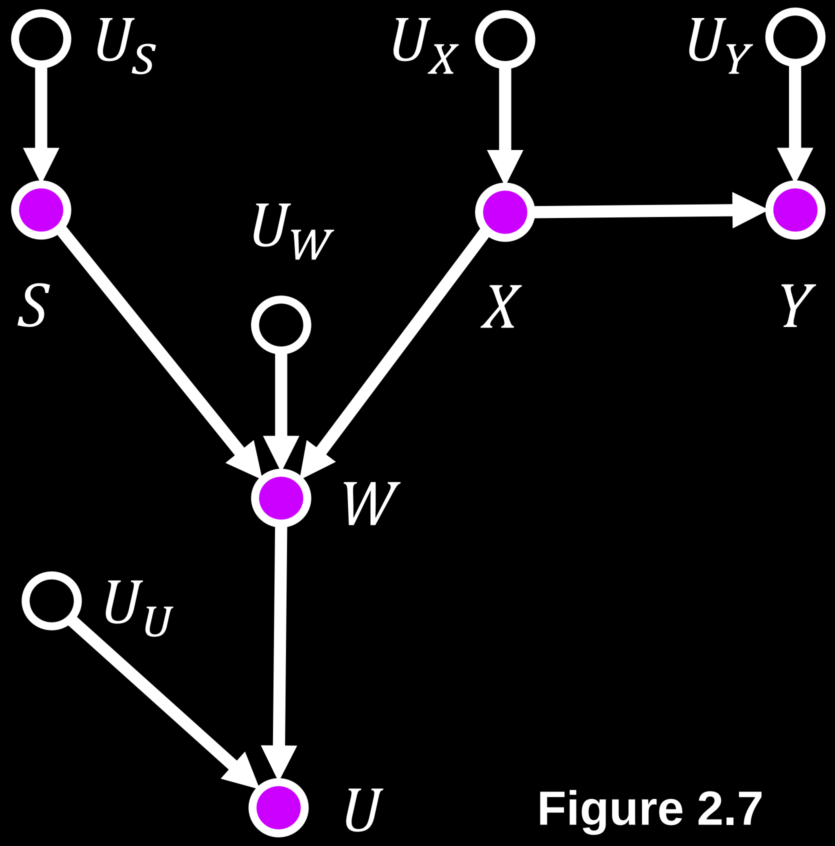


Figure 2.7

The variables might be discrete, continuous, or a mixture of the two; the relationships between them might be linear, exponential, or any of an infinite number of other relations. No matter the model, however, d-separation will always provide the same set of independencies in the data the model generates.

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

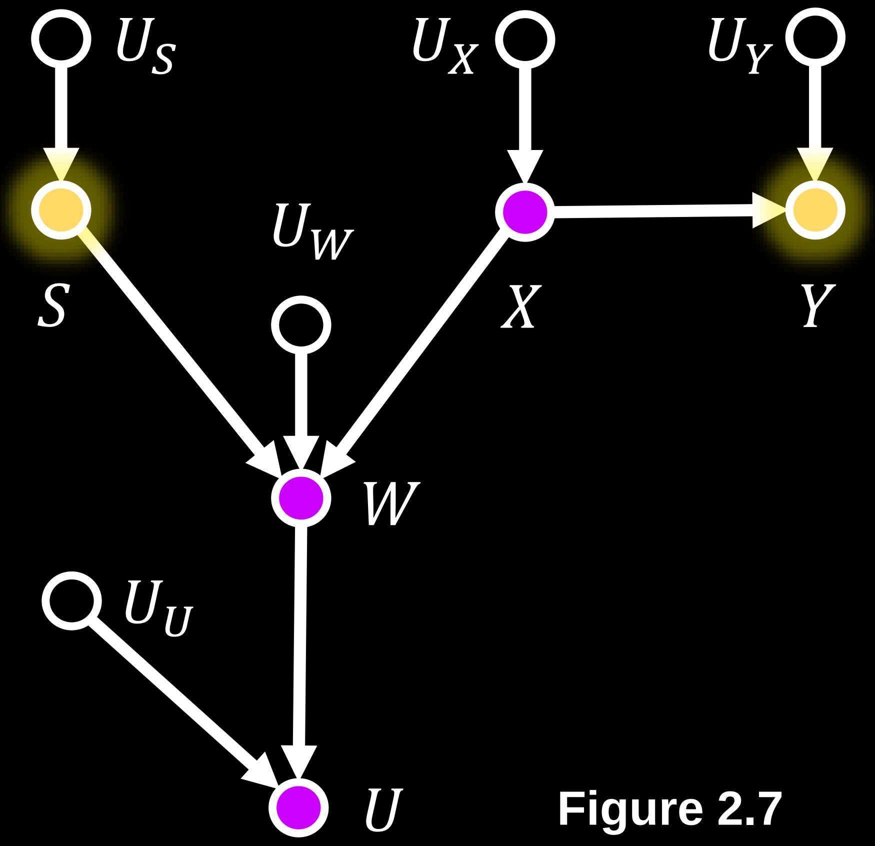


Figure 2.7

In particular, let's look at the relationship between S and Y.

empty conditioning set $Z = \{\varnothing\}$ → S and Y are d-separated → S and Y are unconditionally independent

Why?

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

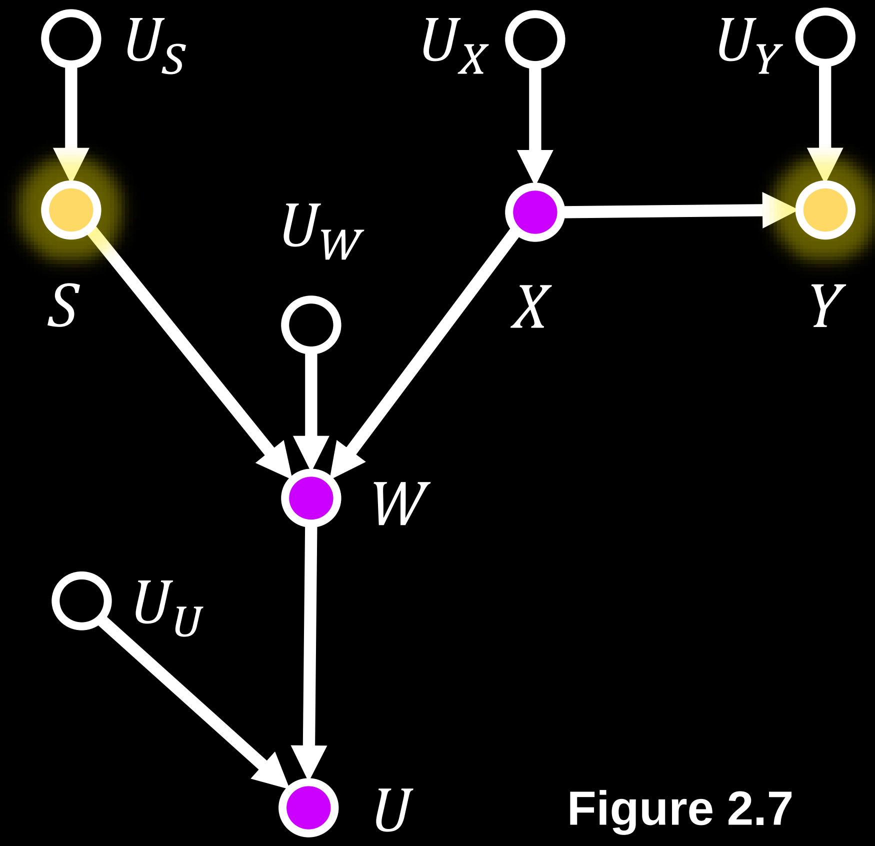


Figure 2.7

Because there is no unblocked path between S and Y.

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

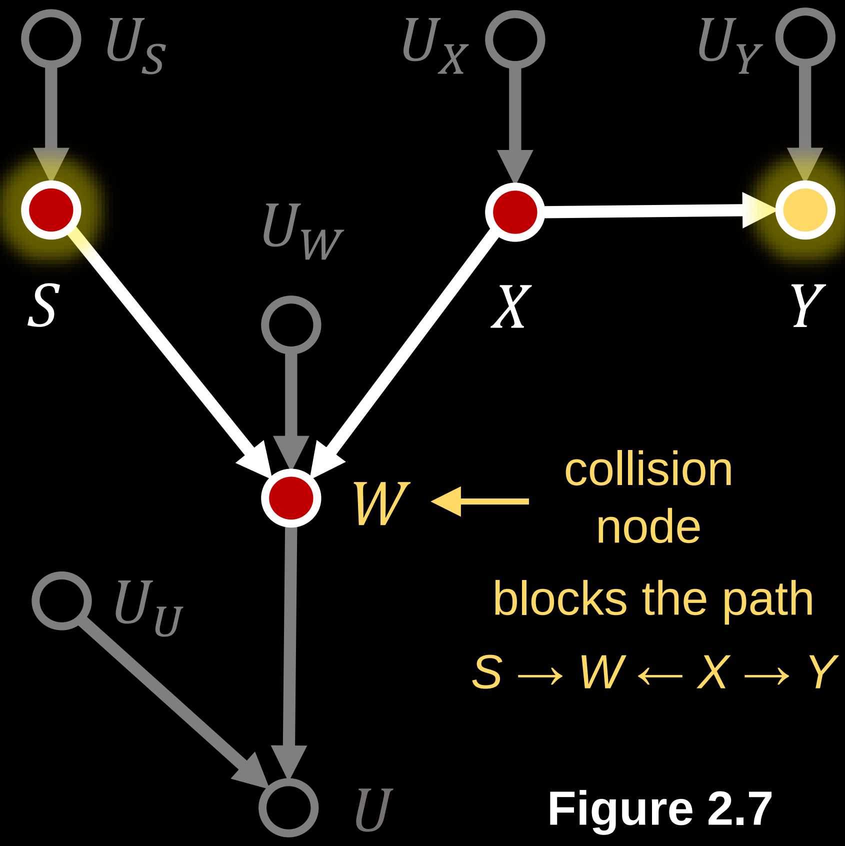


Figure 2.7

Because there is no unblocked path between S and Y.

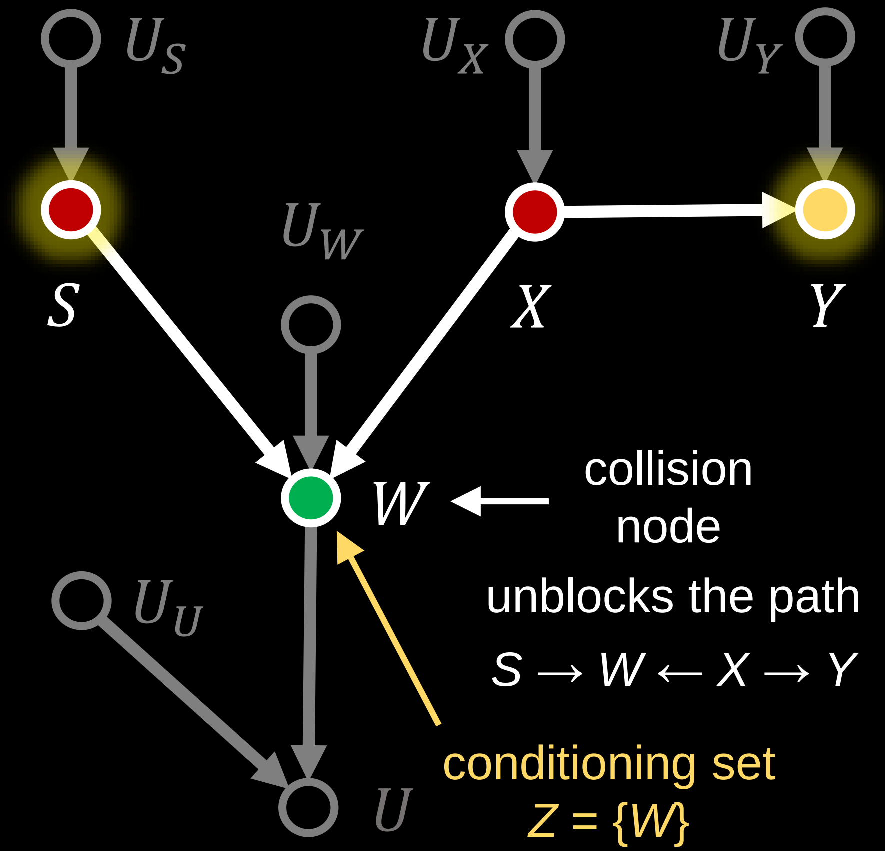
There is only one path between S and Y, and that path is blocked by a collider ($S \rightarrow W \leftarrow X$).

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.



But suppose we condition on W.

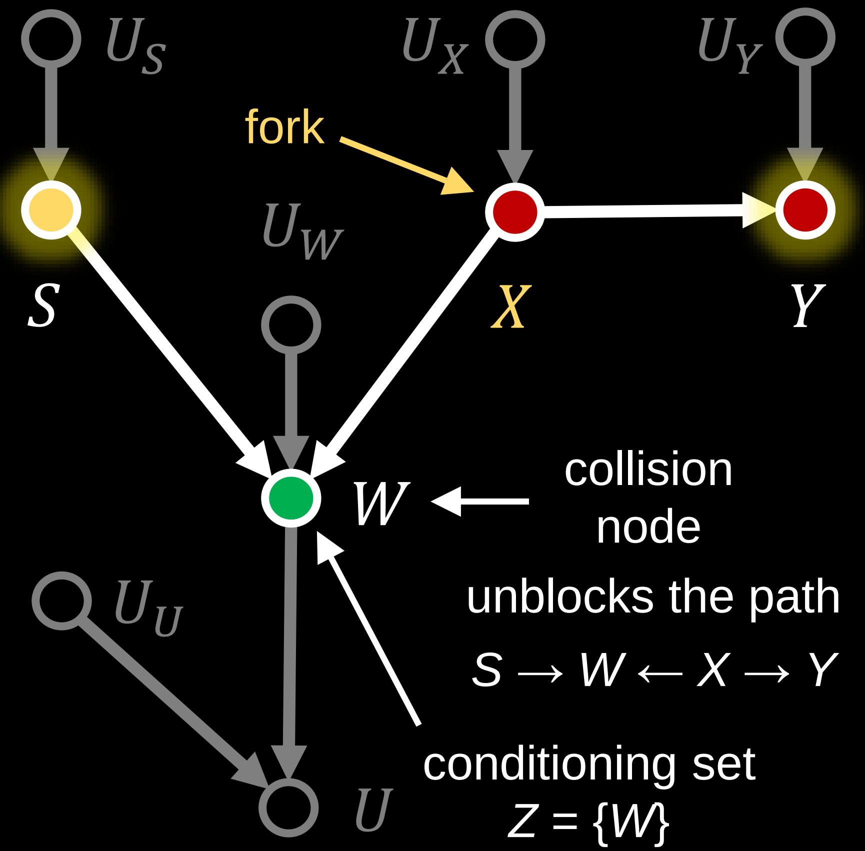
d-separation tells us that S and Y are d-connected, conditional on W.

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \to B \to C$ or a fork $A \leftarrow B \to C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \to B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.



The reason is that our conditioning set is now $Z = \{W\}$, and since the only path between S and Y contains a fork (X) that is not in that set, and the only collider (W) on the path is in that set, that path is not blocked.

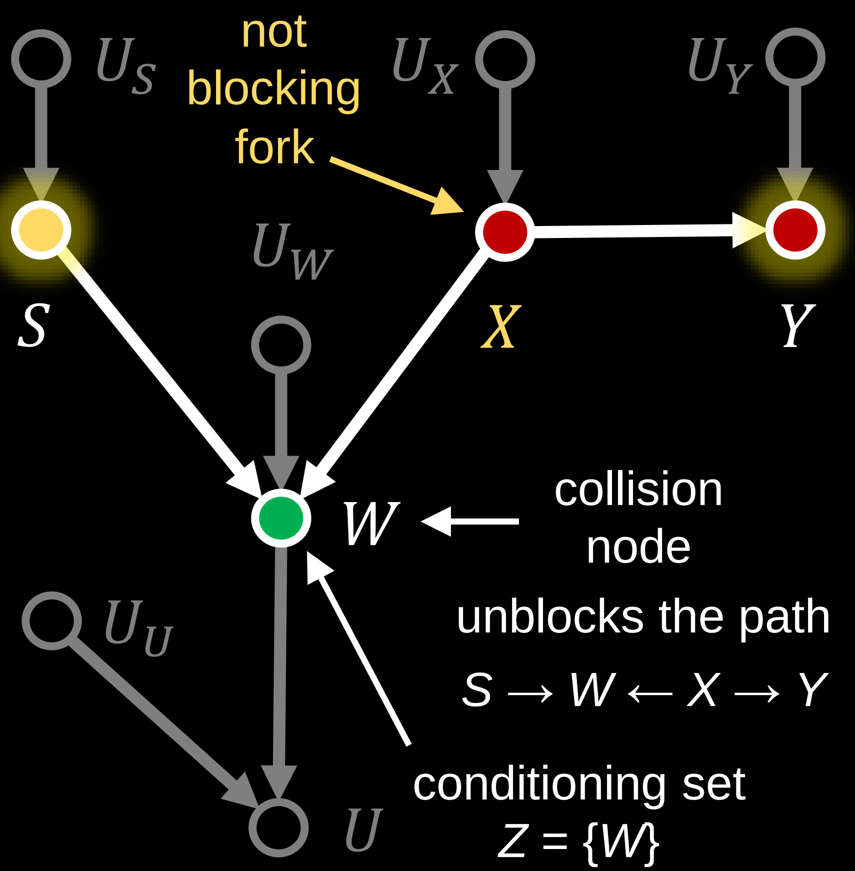
(Remember that conditioning on colliders "unblocks" them.)

Definition 2.4.1 (d-separation)

A path *p* is blocked by a set of nodes *Z* if and only if

1. *p* contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node *B* is in *Z* (i.e., is conditioned on), or
2. *p* contains a collider $A \rightarrow B \leftarrow C$ such that the collision node *B* is not in *Z*, and no descendant of *B* is in *Z*.

If *Z* blocks every path between two nodes *X* and *Y*, then *X* and *Y* are d-separated, conditional on *Z*, and thus are independent conditional on *Z*.



The reason is that our conditioning set is now $Z = \{W\}$, and since the only path between *S* and *Y* contains a fork (*X*) that is not in that set, and the only collider (*W*) on the path is in that set, that path is not blocked.

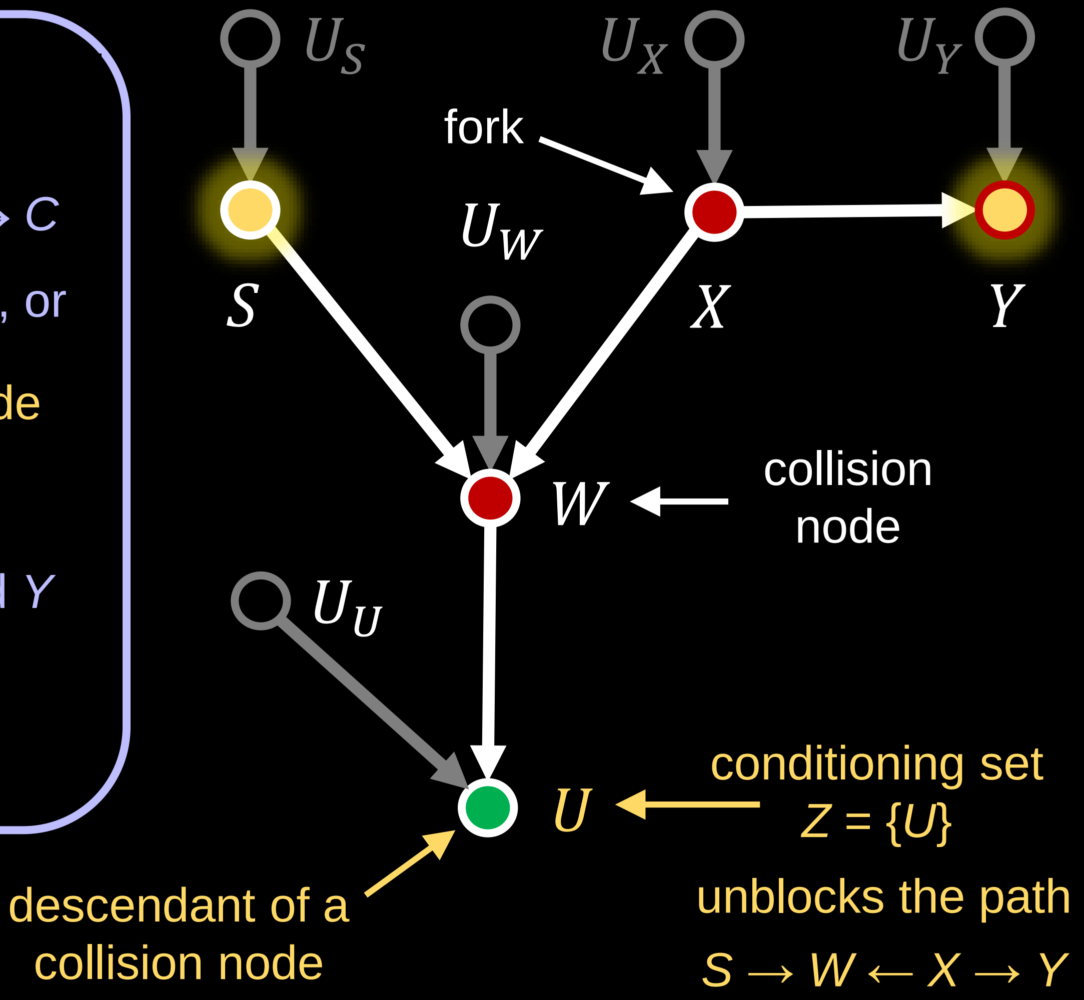
(Remember that conditioning on colliders "unblocks" them.)

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.



The same is true if we condition on U, because U is a descendant of a collider along the path between S and Y.

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

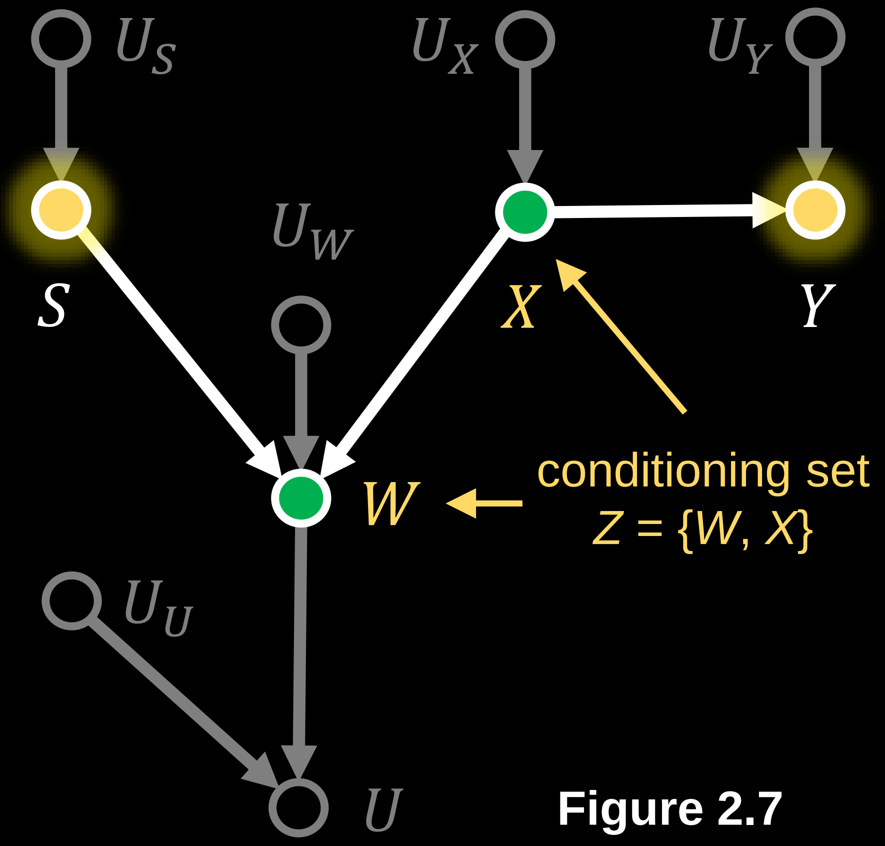


Figure 2.7

On the other hand, if we condition on the set $Z = \{W, X\}$, S and Y remain independent.

This time, the path between S and Y is blocked by the first criterion, rather than the second.

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

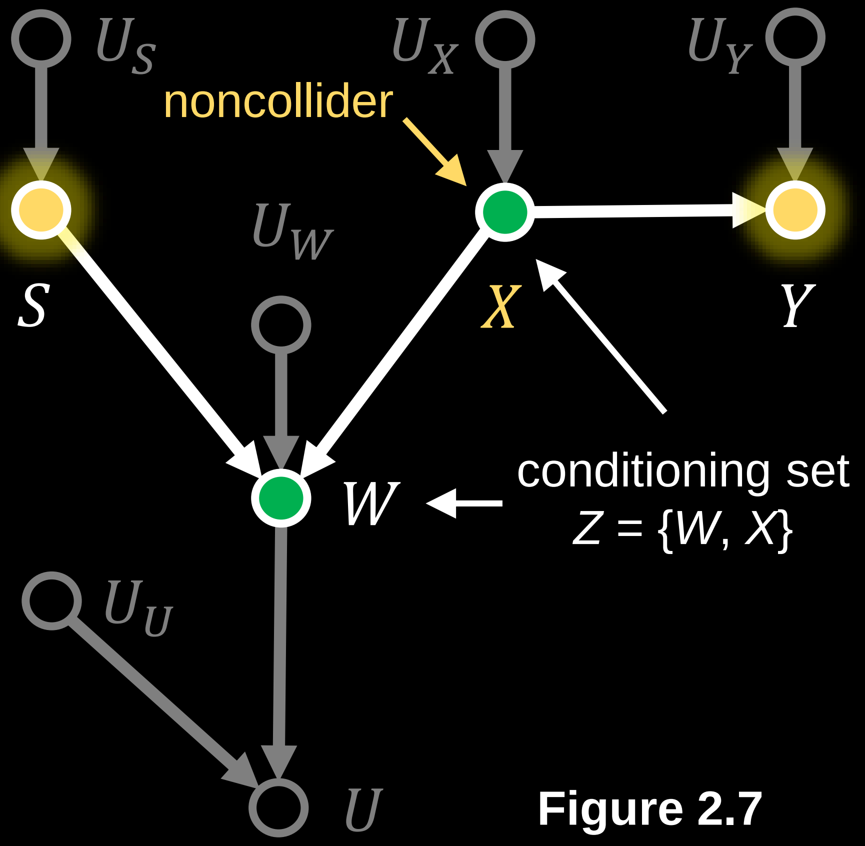


Figure 2.7

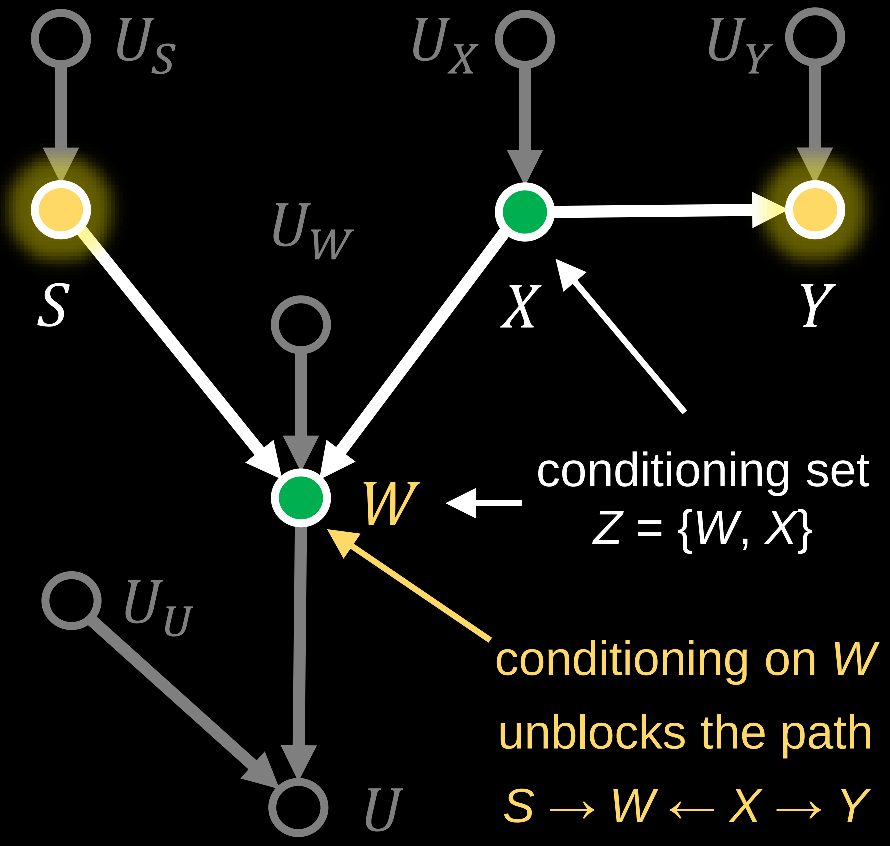
There is now a noncollider node (X) on the path that is in the conditioning set.

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.



Though

$$S \rightarrow W \leftarrow X \rightarrow Y$$

has been unblocked by conditioning on W, one blocked node is sufficient to block the entire path.

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \to B \to C$ or a fork $A \leftarrow B \to C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \to B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

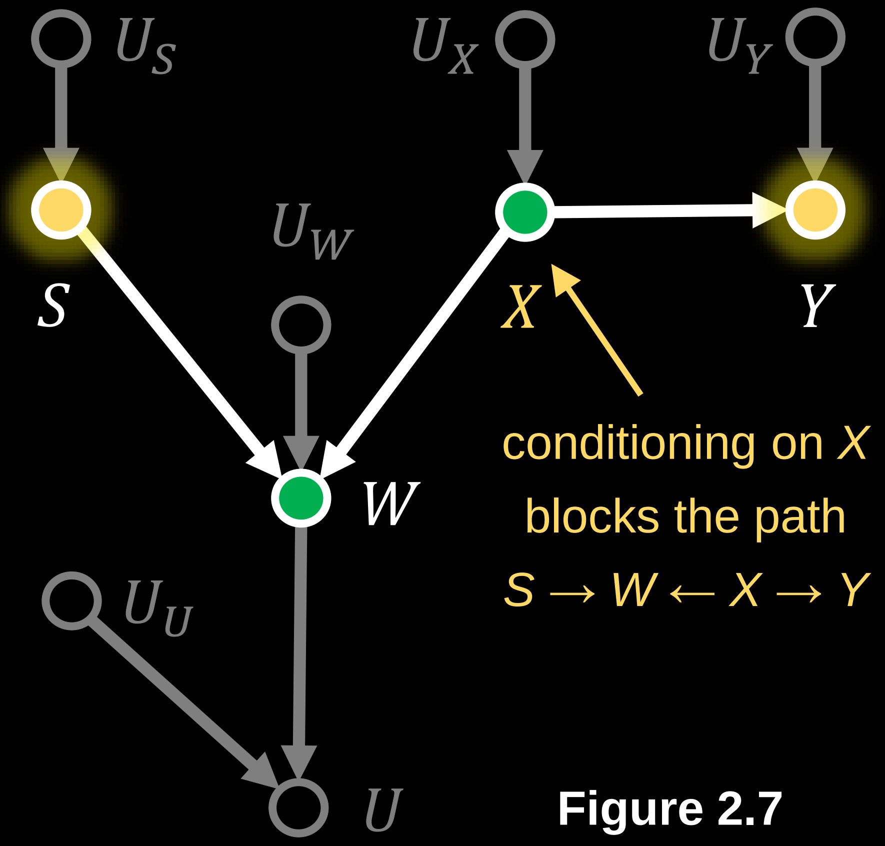


Figure 2.7

Since the only path between S and Y is blocked by this conditioning set Z, S and Y are d-separated conditional on $Z = \{W, X\}$.

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

Now, consider what happens when we add another path between S and Y, as in **Figure 2.8**.

S and Y are now unconditionally dependent.

Why?

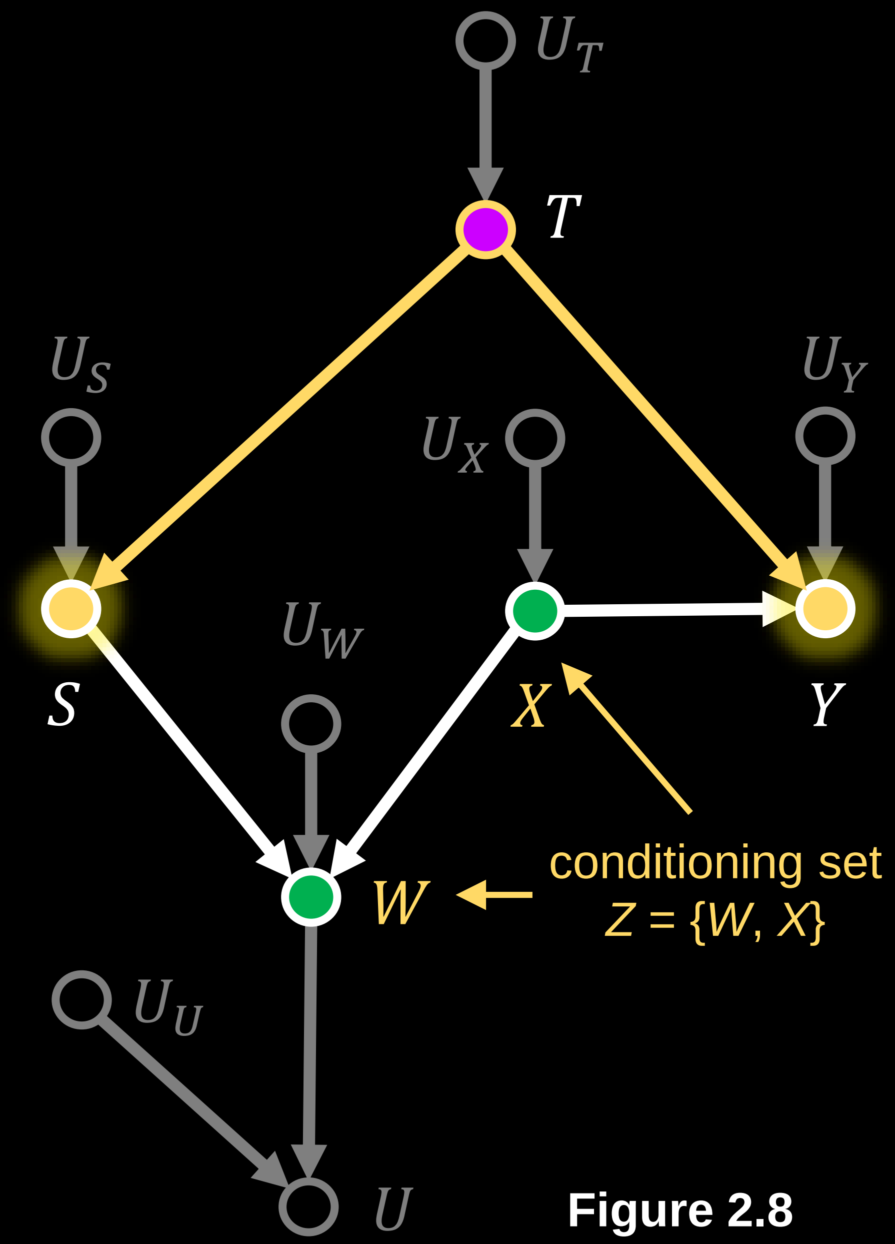


Figure 2.8

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

Because there is a path between them ($S \leftarrow T \rightarrow Y$) that contains no colliders, and the middle node T does not belong to the conditioning set $Z = \{W, X\}$.

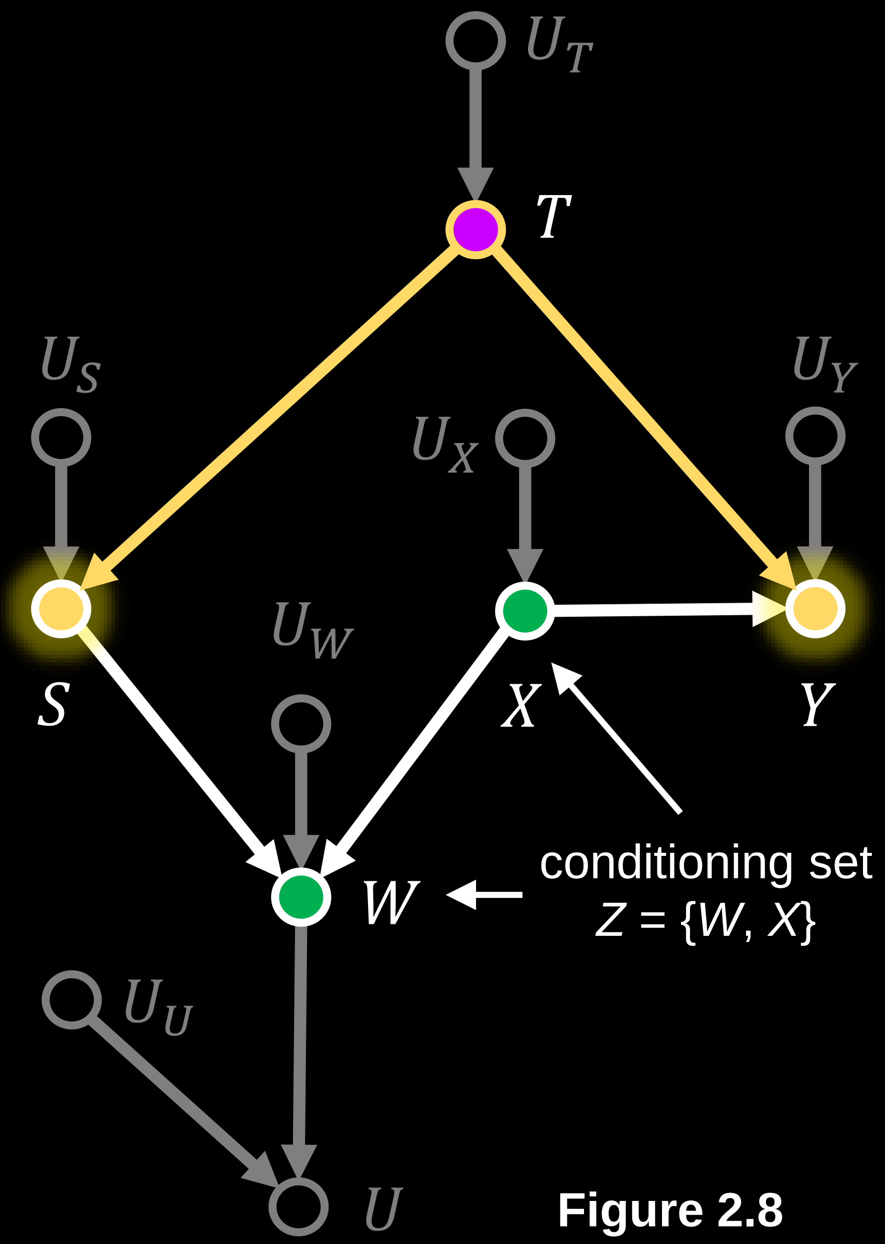


Figure 2.8

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

If we also condition on T, *i.e., if we set* $Z = \{W, X, T\}$, however, the path ($S \leftarrow T \rightarrow Y$) is blocked, and S and Y become independent again.

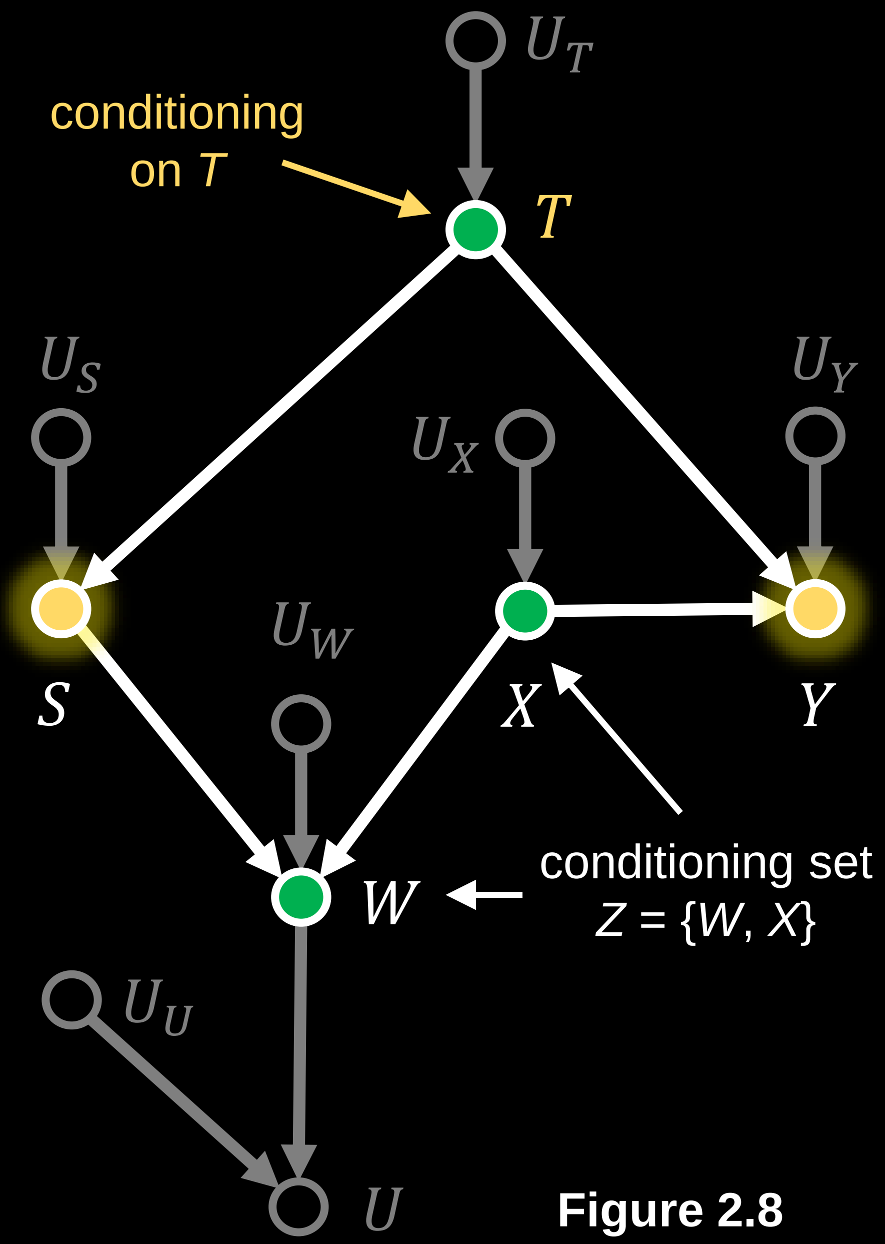


Figure 2.8

Definition 2.4.1 (d-separation)

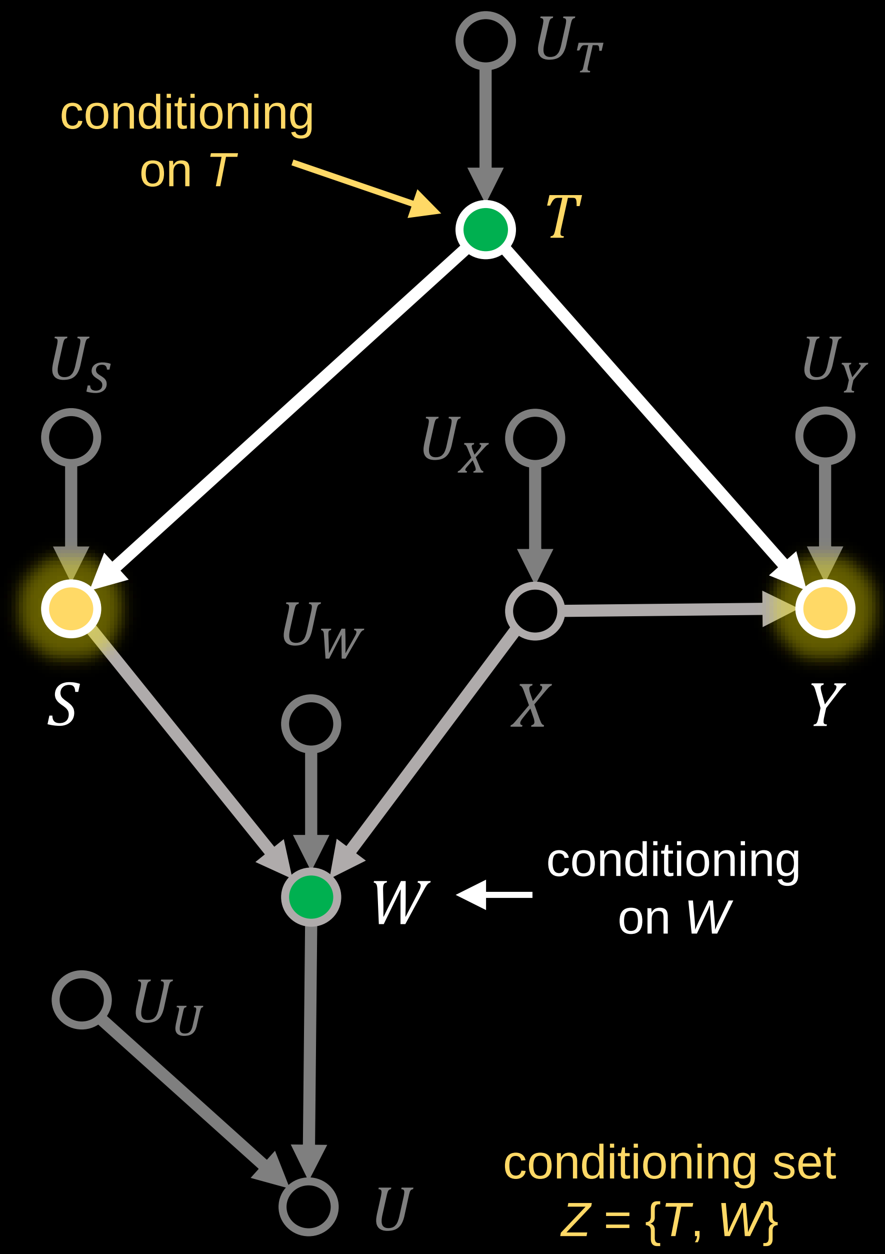
A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

Conditioning on $Z = \{T, W\}$, on the other hand, makes them d-connected again:

- conditioning on T blocks the path $S \leftarrow T \rightarrow Y$,



Definition 2.4.1 (d-separation)

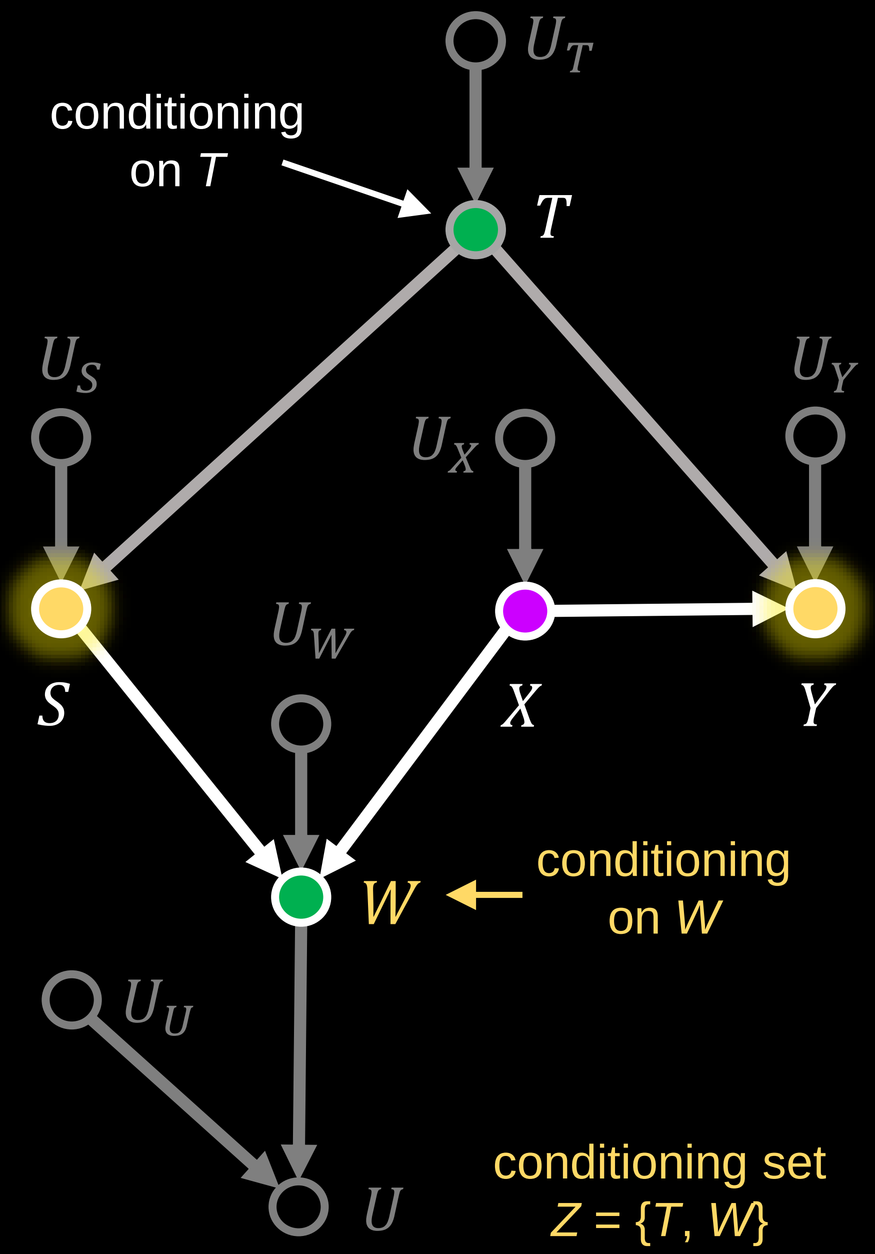
A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

Conditioning on $Z = \{T, W\}$, on the other hand, makes them d-connected again:

- conditioning on T blocks the path $S \leftarrow T \rightarrow Y$,
- but conditioning on W unblocks the path $S \rightarrow W \leftarrow X \rightarrow Y$.



Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

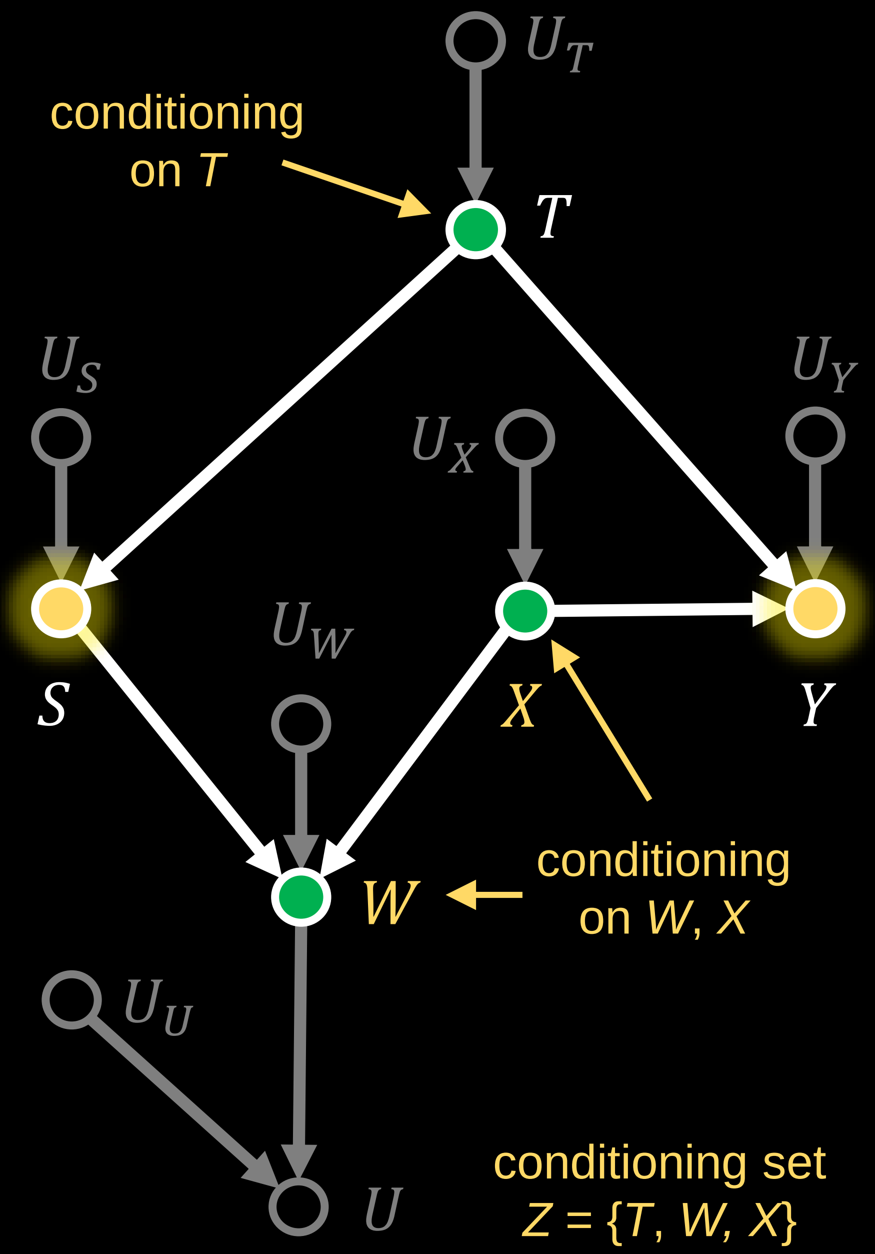
1. p contains a chain of nodes $A \to B \to C$ or a fork $A \leftarrow B \to C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \to B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

And if we add X to the conditioning set, making it

$$Z = \{T, W, X\},$$

S, and Y become independent yet again!



Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

In this graph, S and Y are d-connected (and therefore likely dependent) conditional on

W, U, $\{W, U\}$, $\{W, T\}$, $\{U, T\}$, $\{W, U, T\}$, $\{W, X\}$, $\{U, X\}$, and $\{W, U, X\}$.

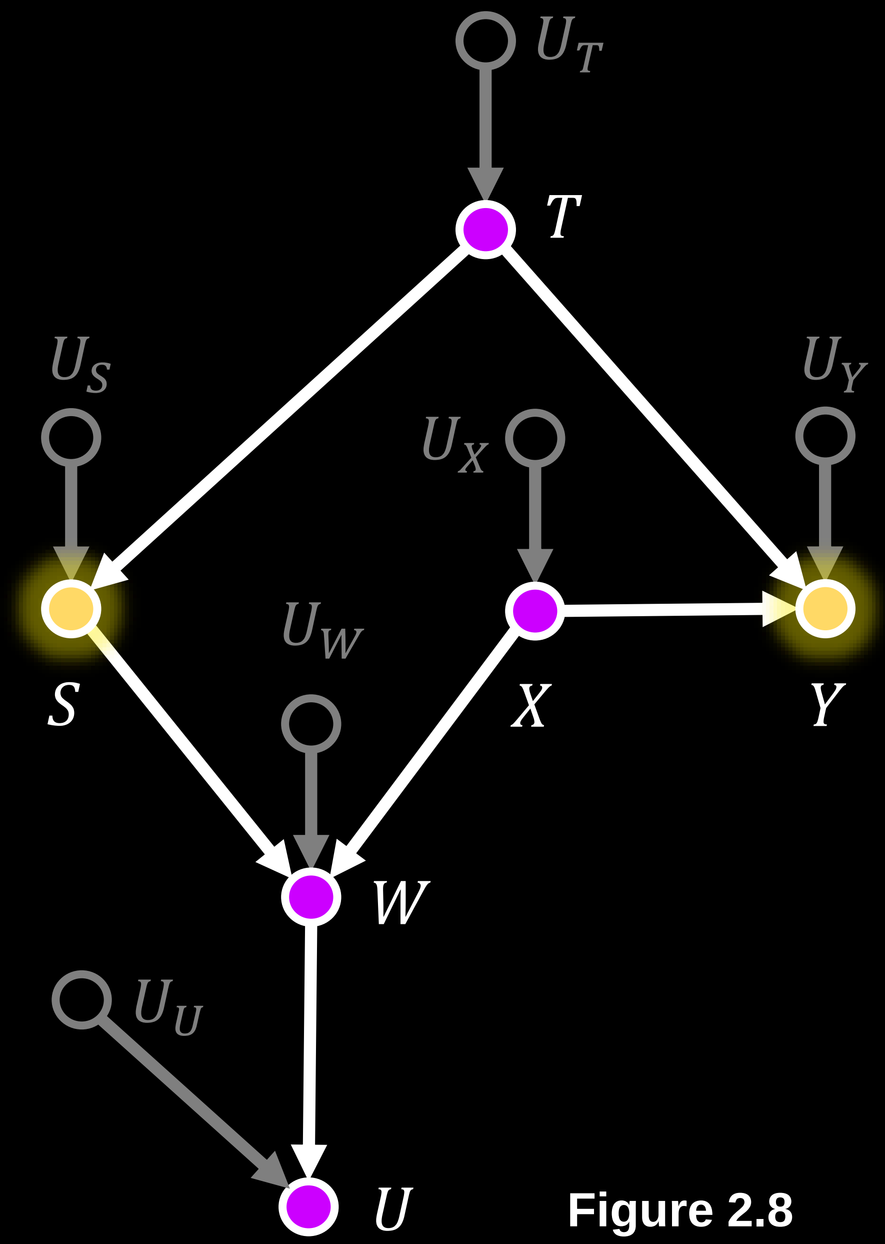


Figure 2.8

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

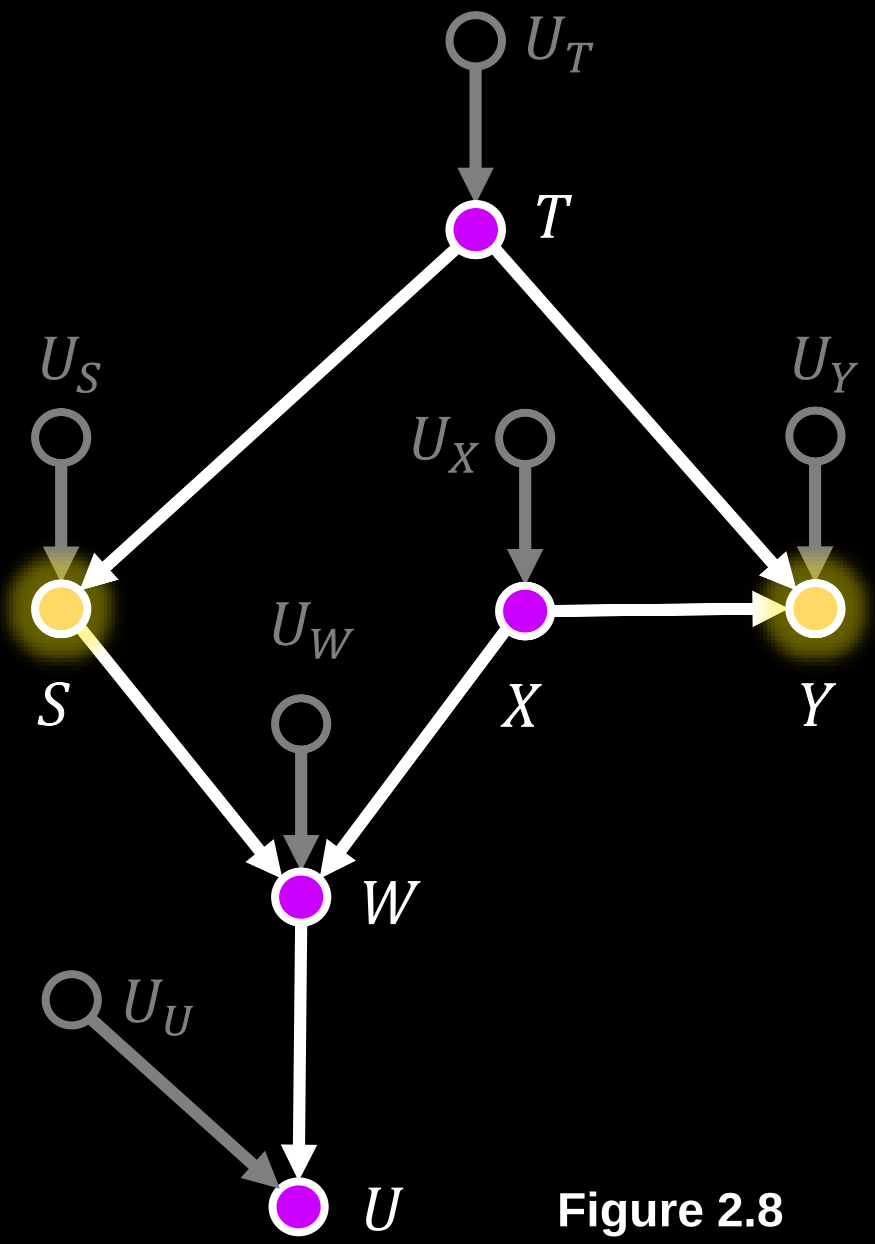


Figure 2.8

S and Y are d-separated (and therefore independent) conditional on:

T, $\{X, T\}$, $\{W, X, T\}$, $\{U, X, T\}$, and $\{W, U, X, T\}$.

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

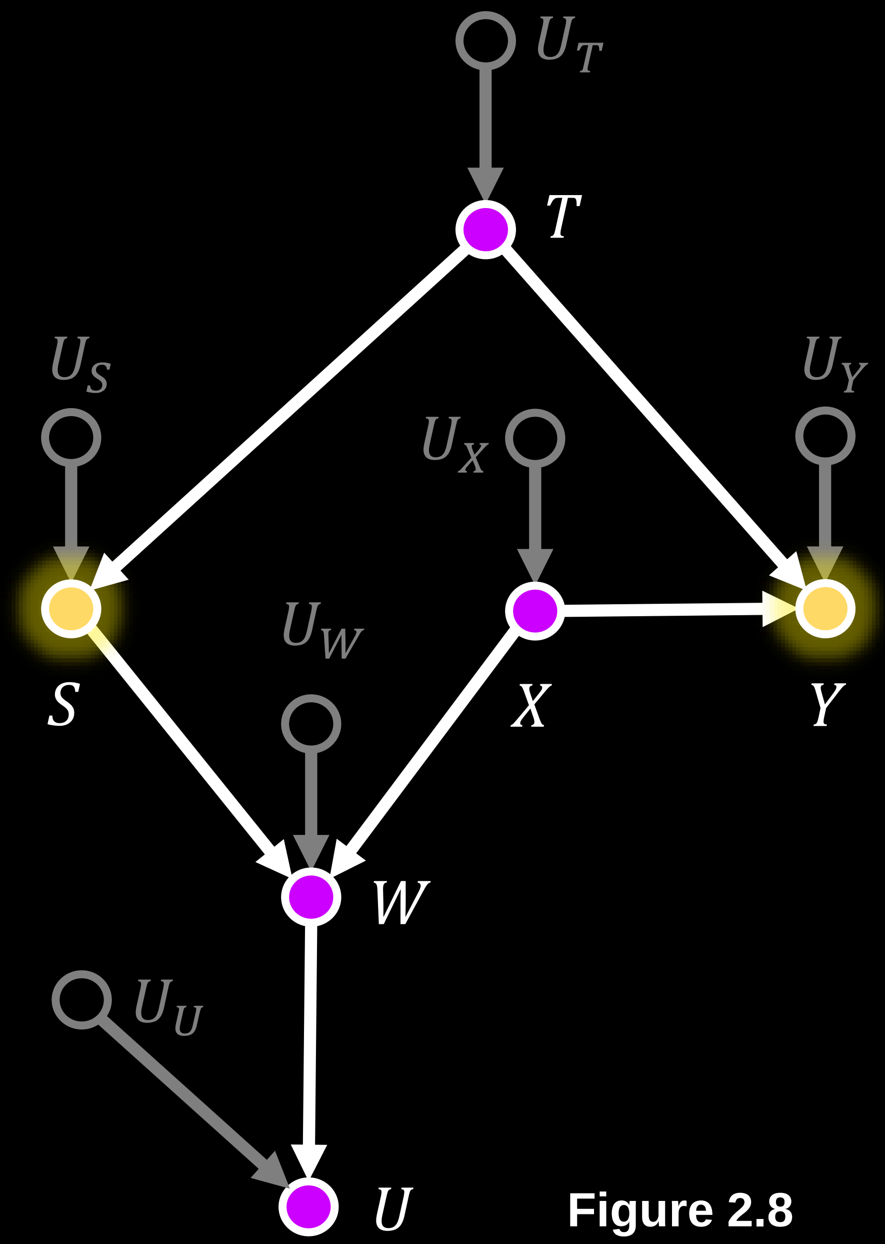


Figure 2.8

S and Y are d-separated (and therefore independent) conditional on:

T, $\{X, T\}$, $\{W, X, T\}$, $\{U, X, T\}$, and $\{W, U, X, T\}$.

Note that T is in every conditioning set that d-separates S and Y.

Definition 2.4.1 (d-separation)

A path p is blocked by a set of nodes Z if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., is conditioned on), or
2. p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

T is in every conditioning set that d-separates S and Y because T is the only node in a path that unconditionally d-connects S and Y, so unless it is conditioned on, S and Y will always be d-connected.

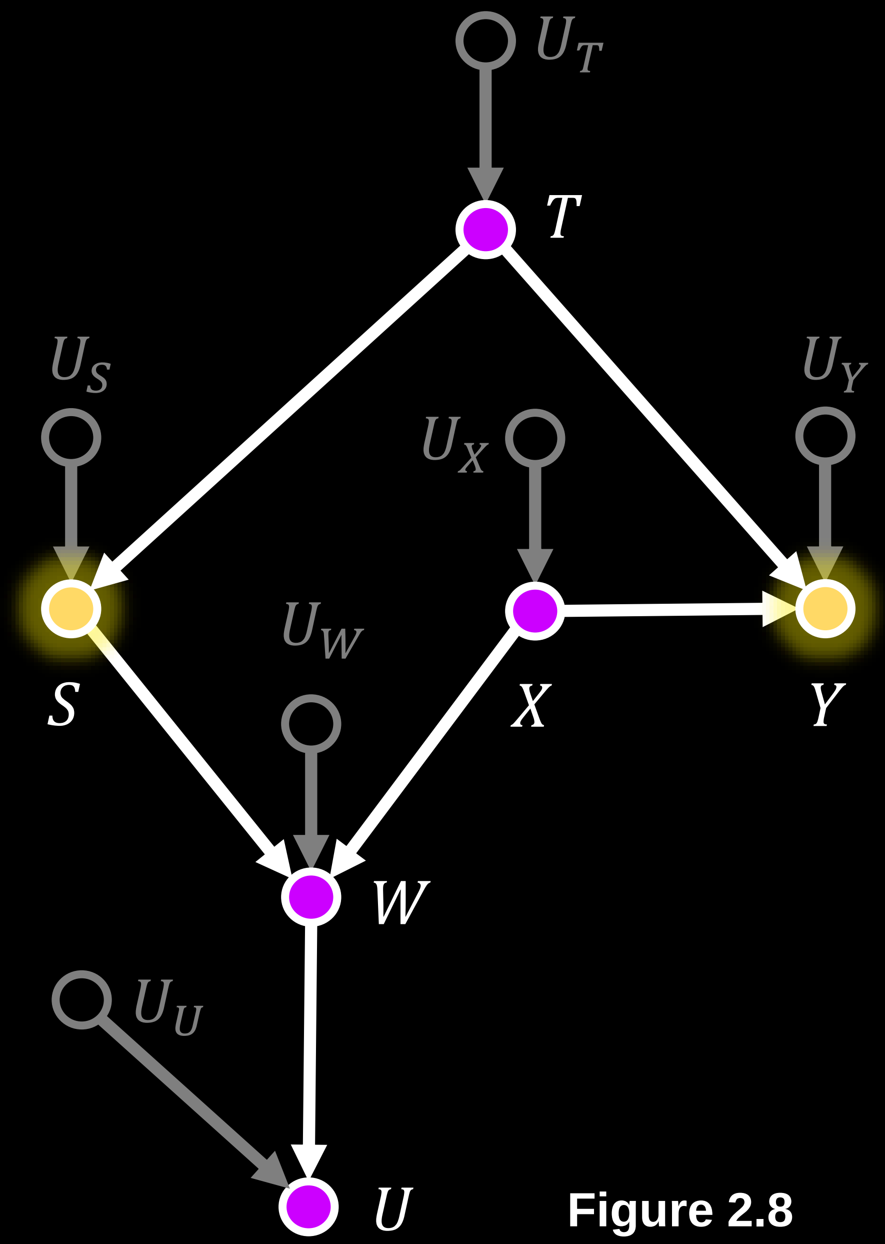
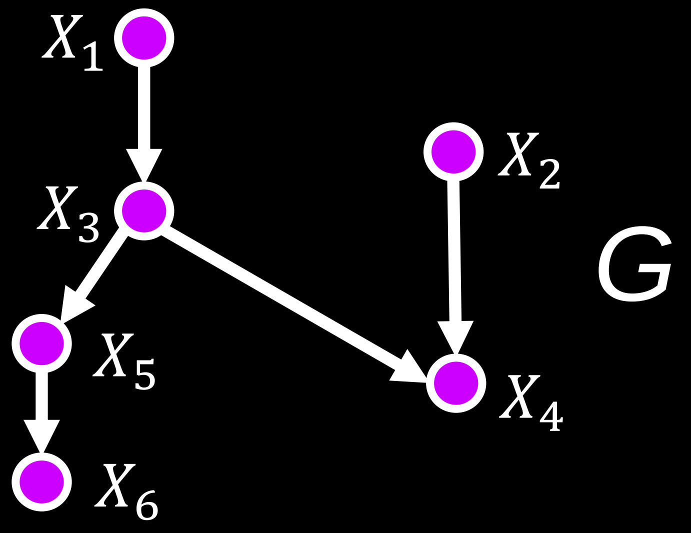


Figure 2.8

The preceding sections demonstrate that causal models have testable implications in the data sets they generate.

For instance, if we have a graph G that we believe might have generated a data set $\mathbb{S}$, d-separation will tell us which variables in G must be independent conditional on which other variables.

Conditional independence is something we can test for using a data set $\mathbb{S}$.



X_1	X_2	X_3	X_4	X_5	X_6
0	1	1	1	0	1
0	1	0	1	1	1
1	0	1	1	0	0
0	0	1	0	1	1
0	1	0	0	0	0
1	1	0	1	0	1
1	0	1	1	1	1
0	1	1	1	0	1

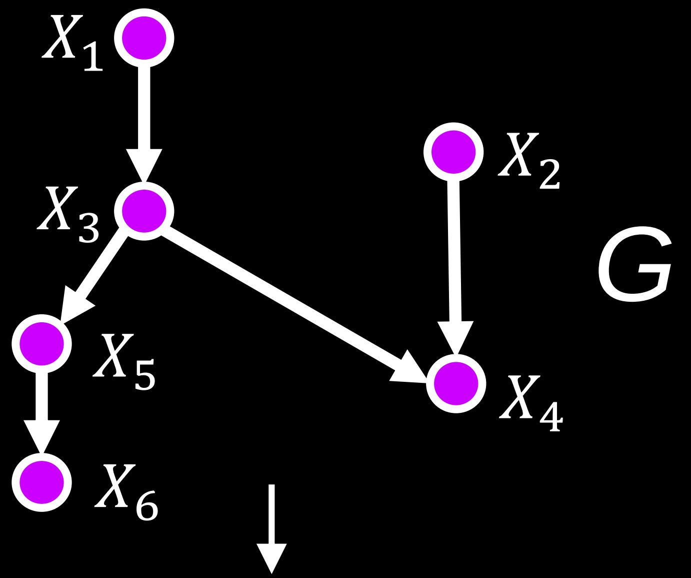
$\mathbb{S}$

The preceding sections demonstrate that causal models have testable implications in the data sets they generate.

For instance, if we have a graph *G* that we believe might have generated a data set $\mathbb{S}$, d-separation will tell us which variables in *G* must be independent conditional on which other variables.

Conditional independence is something we can test for using a data set $\mathbb{S}$.

- Suppose we list the d-separation conditions in *G*, and note that variables X_6 and X_3 must be independent conditional on X_5.

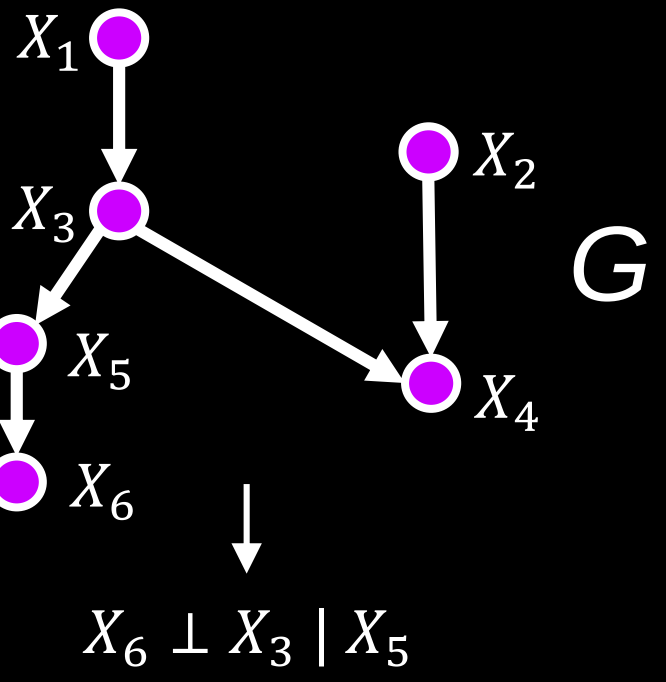


$X_6 \perp X_3 \mid X_5$

$X_6 \perp X_1 \mid X_3$

$X_1 \perp X_2$

$X_6 \perp X_1 \mid X_5$



The preceding sections demonstrate that causal models have testable implications in the data sets they generate.

For instance, if we have a graph *G* that we believe might have generated a data set $\mathbb{S}$, d-separation will tell us which variables in *G* must be independent conditional on which other variables.

Conditional independence is something we can test for using a data set $\mathbb{S}$.

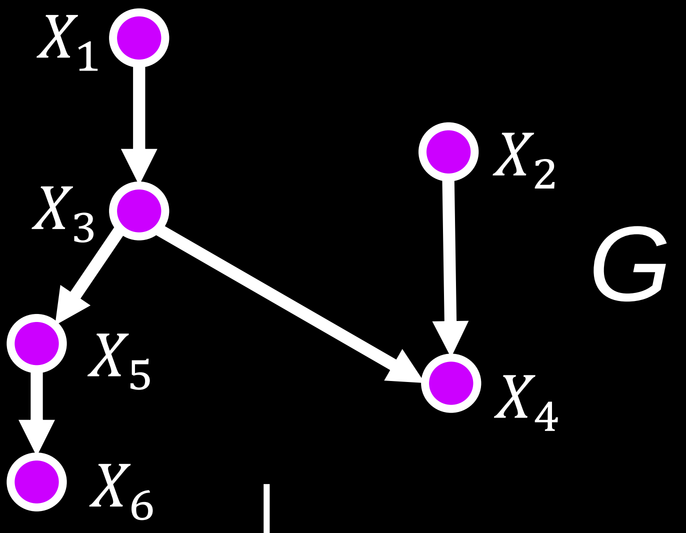
- Suppose we list the d-separation conditions in *G*, and note that variables X_6 and X_3 must be independent conditional on X_5.

The preceding sections demonstrate that causal models have testable implications in the data sets they generate.

For instance, if we have a graph *G* that we believe might have generated a data set $\mathbb{S}$, d-separation will tell us which variables in *G* must be independent conditional on which other variables.

Conditional independence is something we can test for using a data set $\mathbb{S}$.

- Suppose we list the d-separation conditions in *G*, and note that variables X_6 and X_3 must be independent conditional on X_5.
- Then, suppose we estimate the probabilities based on $\mathbb{S}$, and discover that the data suggests that X_6 and X_3 are not independent conditional on X_5

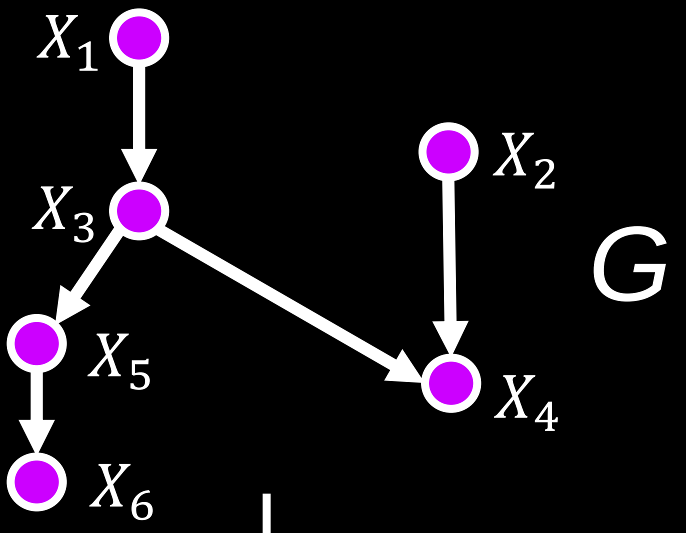


$$X_6 \perp X_3 \mid X_5$$

$$X_6 \not\perp X_3 \mid X_5$$

X_1	X_2	X_3	X_4	X_5	X_6
0	1	1	1	0	1
0	1	0	1	1	1
1	0	1	1	0	0
0	0	1	0	1	1
0	1	0	0	0	0
1	1	0	1	0	1
1	0	1	1	1	1
0	1	1	1	0	1

$\mathbb{S}$

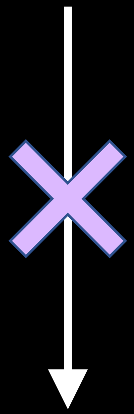


The preceding sections demonstrate that causal models have testable implications in the data sets they generate.

For instance, if we have a graph *G* that we believe might have generated a data set $\mathbb{S}$, d-separation will tell us which variables in *G* must be independent conditional on which other variables.

Conditional independence is something we can test for using a data set $\mathbb{S}$.

- Suppose we list the d-separation conditions in *G*, and note that variables X_6 and X_3 must be independent conditional on X_5.
- Then, suppose we estimate the probabilities based on $\mathbb{S}$, and discover that the data suggests that X_6 and X_3 are not independent conditional on X_5
- We can then reject *G* as a possible causal model for $\mathbb{S}$.



X_1	X_2	X_3	X_4	X_5	X_6
0	1	1	1	0	1
0	1	0	1	1	1
1	0	1	1	0	0
0	0	1	0	1	1
0	1	0	0	0	0
1	1	0	1	0	1
1	0	1	1	1	1
0	1	1	1	0	1

$\mathbb{S}$

We can demonstrate it on the causal model of **Figure 2.9**.

Among the many conditional independencies advertised by the model, we find that *W* and Z_1 are independent given *X*, because *X* d-separates *W* from Z_1.

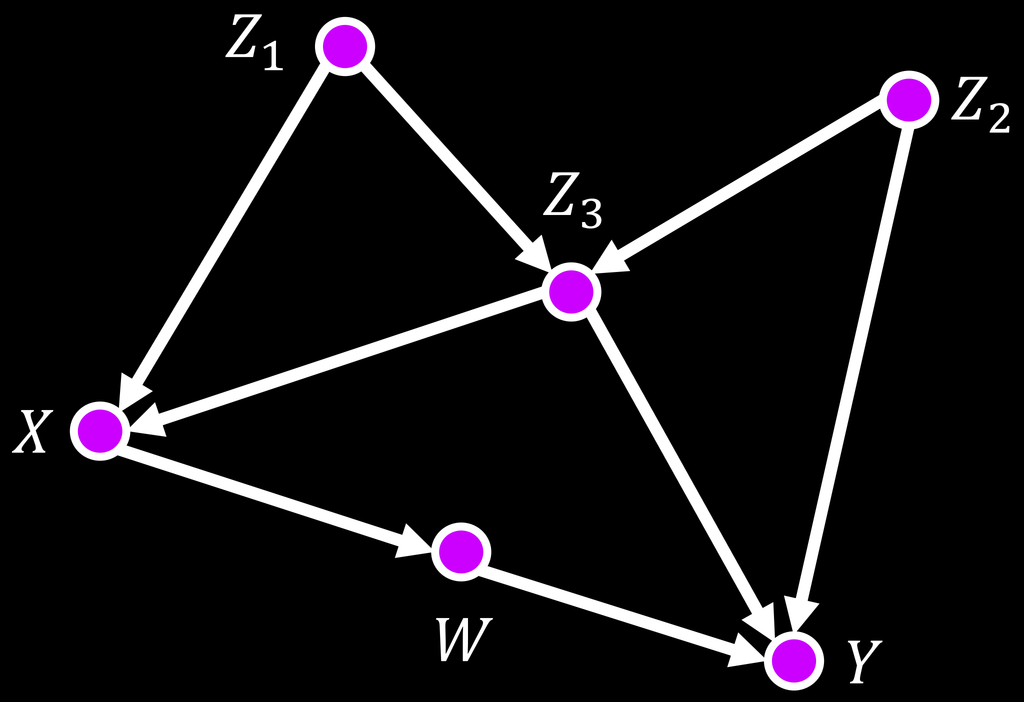


Figure 2.9

We can demonstrate it on the causal model of **Figure 2.9**.

Among the many conditional independencies advertised by the model, we find that *W* and Z_1 are independent given *X*, because *X* d-separates *W* from Z_1.

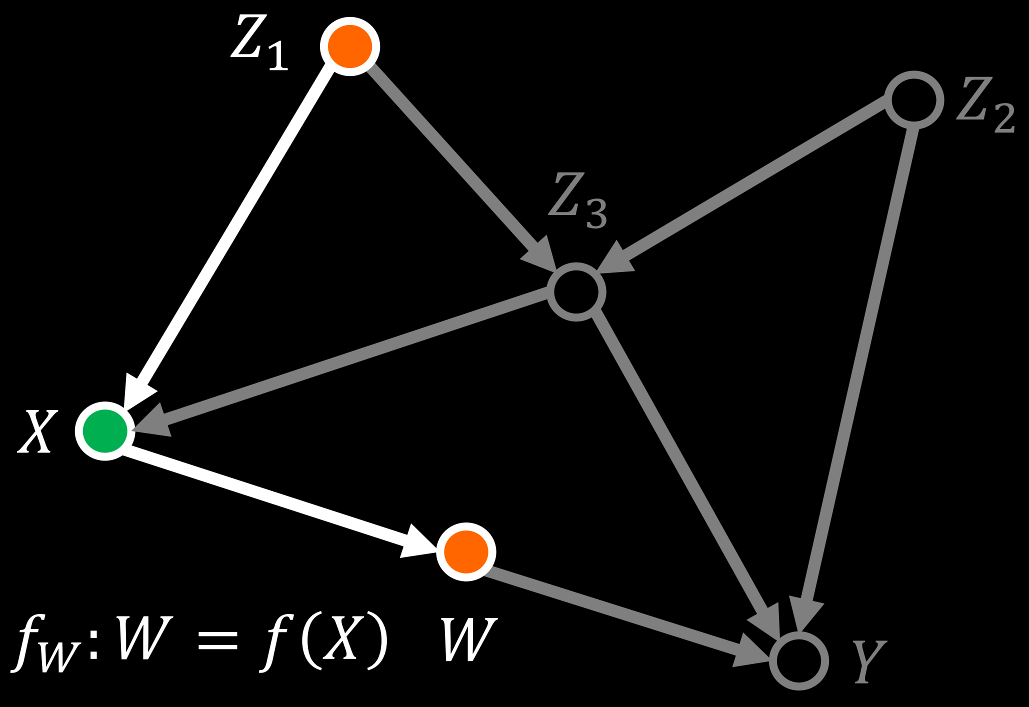


Figure 2.9

We can demonstrate it on the causal model of **Figure 2.9**.

Among the many conditional independencies advertised by the model, we find that *W* and Z_1 are independent given *X*, because *X* d-separates *W* from Z_1.

Now suppose we regress *W* on *X* and Z_1. Namely, we find the line

$$W = r_X X + r_1 Z_1$$

that best fits our data.

IF $r_1 \neq 0 \Rightarrow W \ depends \ on \ Z_1 \ given \ X$

and consequently, that the model in **Figure 2.9** is wrong.

[Conditional correlation implies conditional dependence.]

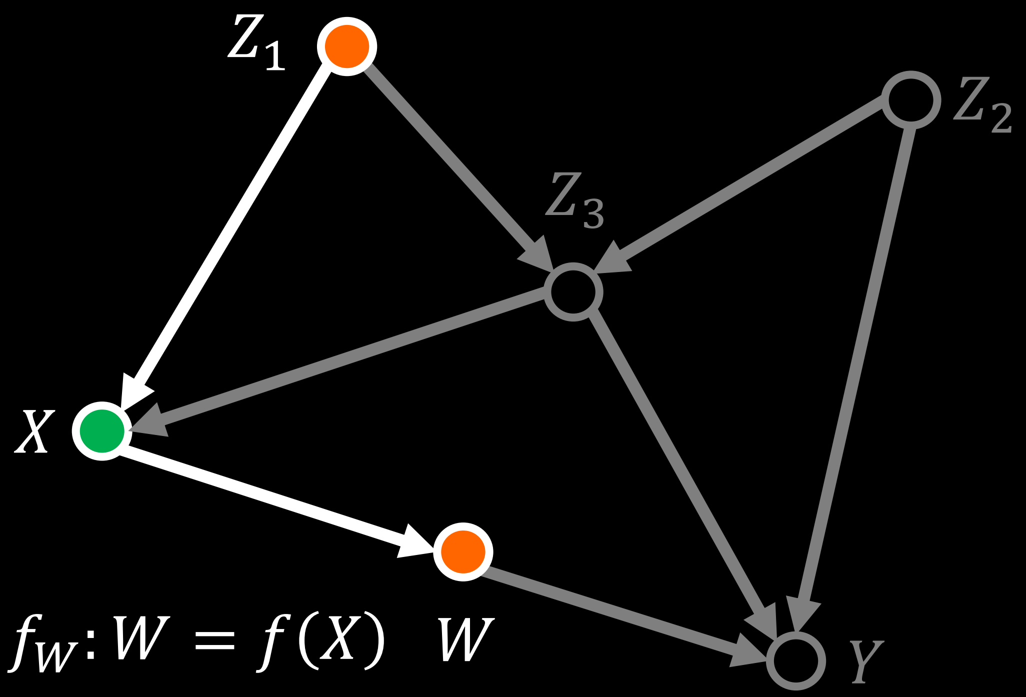


Figure 2.9

2.5 Model Testing and Causal Search

We can demonstrate it on the causal model of **Figure 2.9**.

Among the many conditional independencies advertised by the model, we find that *W* and Z_1 are independent given *X*, because *X* d-separates *W* from Z_1.

Now suppose we regress *W* on *X* and Z_1. Namely, we find the line

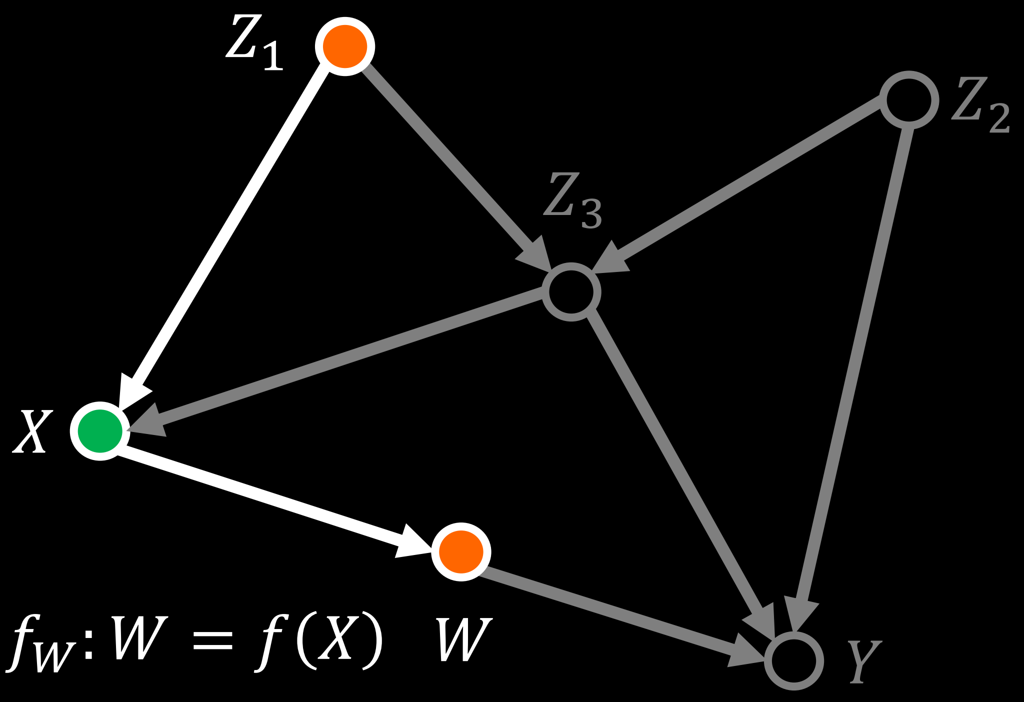
$$W = r_X X + r_1 Z_1 \qquad W \not\perp Z_1 \mid X$$

that best fits our data.

$$\text{IF } r_1 \neq 0 \Rightarrow W\ depends\ on\ Z_1\ given\ X$$

and consequently, that the model in **Figure 2.9** is wrong.

[Conditional correlation implies conditional dependence.]



$$f_W : W = f(X)$$

Figure 2.9

Not only do we know that the model in **Figure 2.9** is wrong, but we also know where it is wrong;

- the **true model** must have a path between *W* and Z_1 that is not d-separated by *X*.

Finally, this is a theoretical result that holds for all acyclic models with independent errors (*Verma and Pearl 1990*), and we also know that if every d-separation condition in the model matches a conditional independence in the data, then no further test can refute the model.

This means that, for any data set whatsoever, one can always find a set of functions *F* for the model and an assignment of probabilities to the *U* terms, so as to generate the data precisely.

There are other methods for testing the fitness of a model.

The standard way of evaluating fitness involves a **statistical hypothesis test** over the entire model, that is, we evaluate how likely it is for the observed samples to have been generated by the hypothesized model, as opposed to sheer chance.

However, since the model is not fully specified, we need to first estimate its parameters before evaluating that likelihood. This can be done (approximately) when we assume a linear and Gaussian model (i.e., all functions in the model are linear and all error terms are normally distributed), because, under such assumptions, the joint distribution (also Gaussian) can be expressed succinctly in terms of the model's parameters, and we can then evaluate the likelihood that the observed samples to have been generated by the fully parameterized model (Bollen 1989).

There are, however, a number of issues with this procedure:

- if any parameter cannot be estimated, then the joint distribution cannot be estimated, and the model cannot be tested. (this can occur when some of the error terms are correlated or, equivalently, when some of the variables are unobserved)
- this procedure tests models globally. If we discover that the model is not a good fit to the data, there is no way for us to determine why that is—which edges should be removed or added to improve the fit.
- when we test a model globally, the number of variables involved may be large, and if there is measurement noise and/or sampling variation associated with each variable, the test will not be reliable.

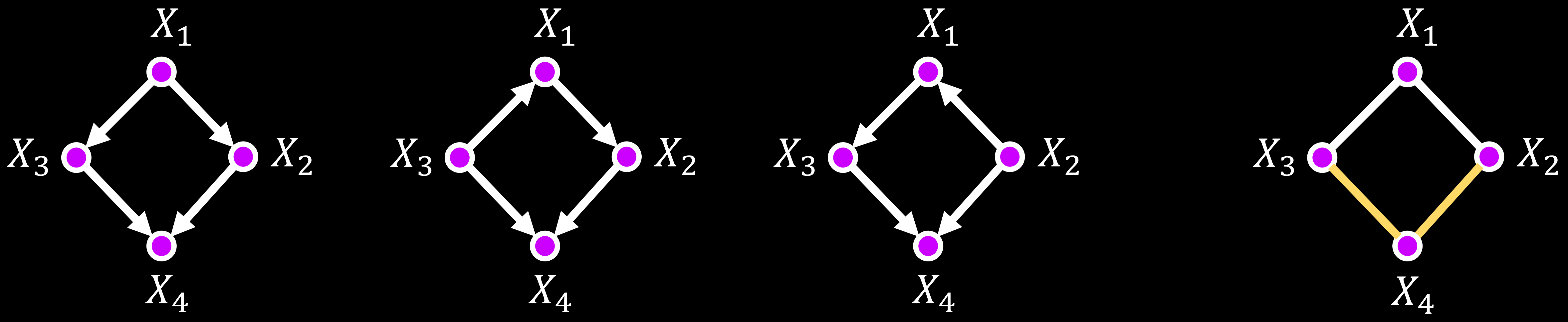
d-separation presents several advantages over this global testing method.

- **it is nonparametric**, meaning that it doesn't rely on the specific functions that connect variables; instead, it uses only the graph of the model in question,
- **it tests models locally, rather than globally**. This allows us to identify specific areas, where our hypothesized model is flawed, and to repair them, rather than starting from scratch on a whole new model. It also means that if, for whatever reason, we can't identify the coefficient in one area of the model, we can still get some incomplete information about the rest of the model. (As opposed to the first method, in which if we could not estimate one coefficient, we could not test any part of the model.)

If we had a computer, we could test and reject many possible models in this way, eventually whittling down the set of possible models to only a few whose testable implications do not contradict the dependencies present in the data set. It is a set of models, rather than a single model, because some graphs have indistinguishable implications. A set of graphs with indistinguishable implications is called an **equivalence class**.

Two graphs G_1 and G_2 are in the same **equivalence class** if they share a **common skeleton**—that is,

- the same edges, regardless of the direction of those edges—and
- if they share common v-structures, that is, colliders whose parents are not adjacent.

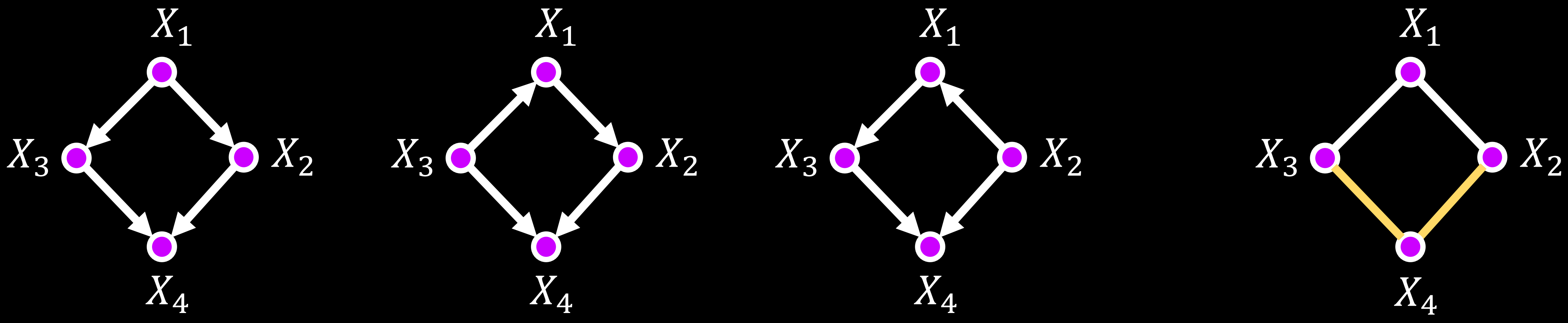


Three **equivalent graphs** and their **skeleton** with the common **v-structure** highlighted.

Any two graphs that satisfy this criterion have identical sets of d-separation conditions and, therefore, identical sets of testable implications (Verma and Pearl 1990).

Two graphs G_1 and G_2 are in the same **equivalence class** if they share a **common skeleton**—that is,

- the same edges, regardless of the direction of those edges—and
- if they share common v-structures, that is, colliders whose parents are not adjacent.

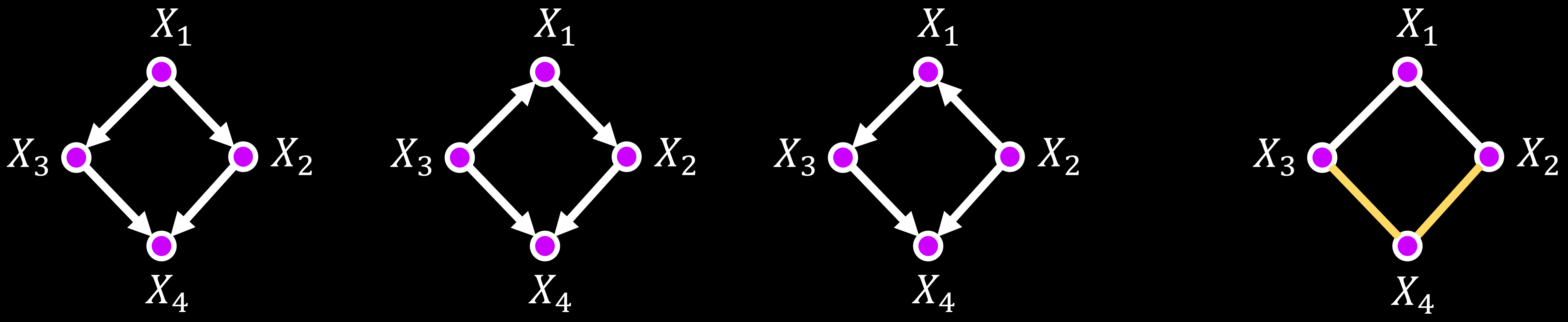


Three **equivalent graphs** and their **skeleton** with the common **v-structure** highlighted.

Any two graphs that satisfy this criterion have identical sets of d-separation conditions and, therefore, identical sets of testable implications (Verma and Pearl 1990).

The importance of this result is that it allows us to search a data set for the causal models that could have generated it. Thus, not only can we start with a causal model and generate a data set—but we can also start with a data set, and reason back to a causal model.

This is enormously useful, since the object of most data-driven research is exactly to find a model that explains the data.



Three **equivalent graphs** and their **skeleton** with the common **v-structure** highlighted.