Everything is about Graphs

(Except graphs··· graphs are supposed to be about causality!)

Eddie Xiao

PaL Lab Meeting Presentation, 13th April

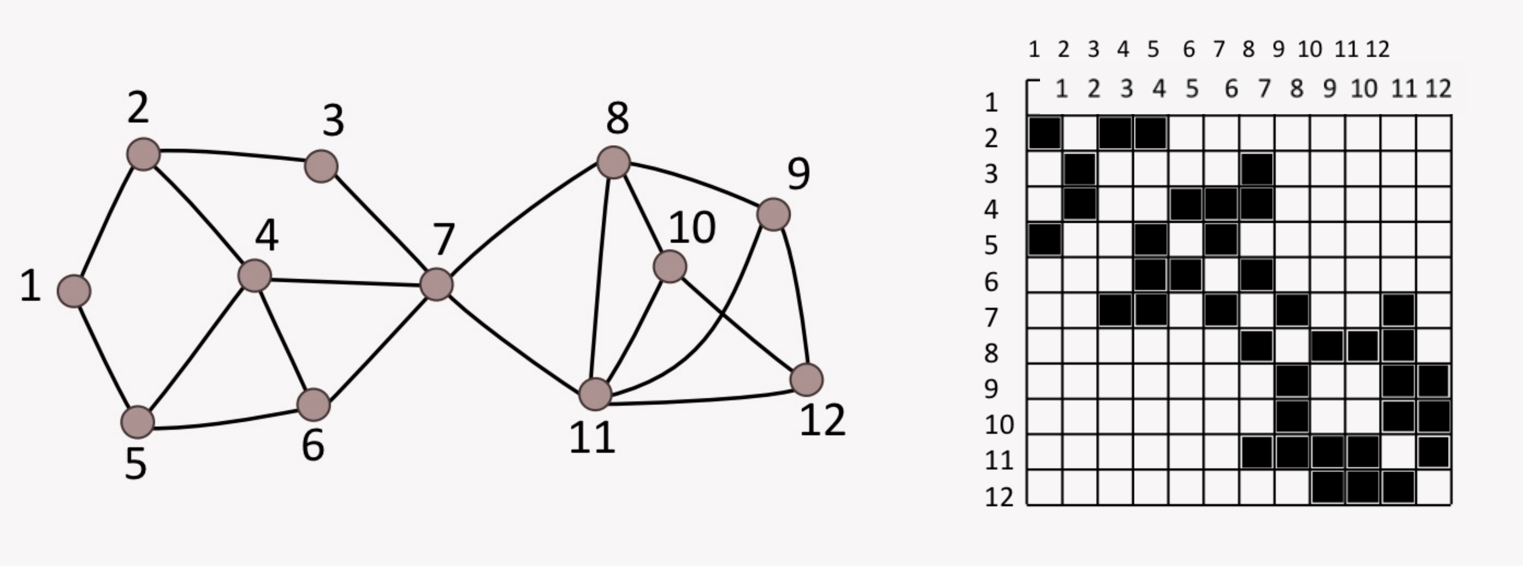
Disclaimer

(Acknowledgements)
-Statistical Rethinking from Richard McElreath
-Amy Orben, lecturer on Robust Behavioural Science

- Hidden theme: struggle with doing good scientific inferences

What is a Graph?

- Nodes & edges (links / paths / connections → relationships)



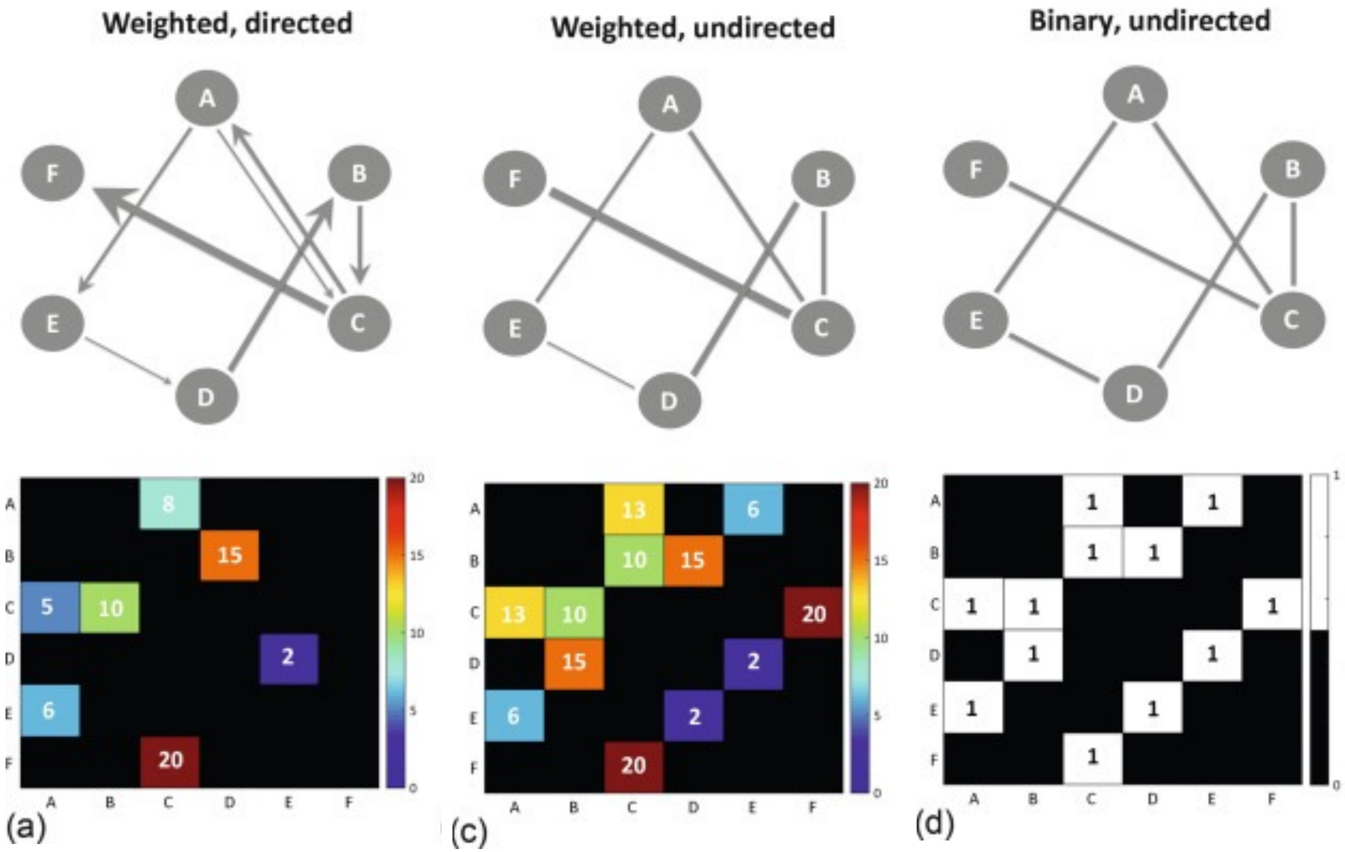
<=Adjacency matrix

(Figure credit: Petra Vertes)

What is a Graph?

- Nodes & edges (links / paths / connections → relationships)

Making stronger claims about the world →



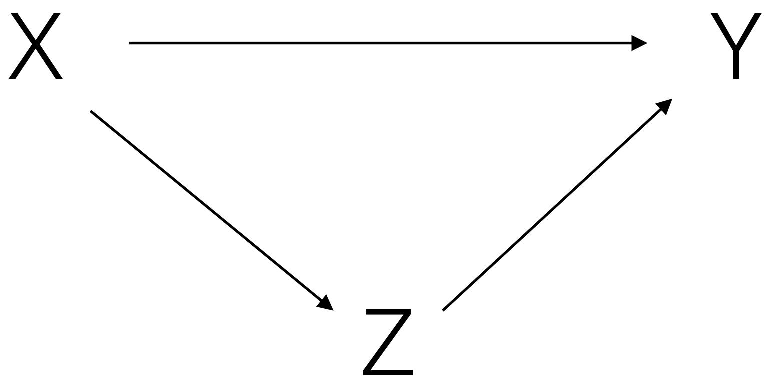
Less information, "easier" to study ←

(Fornito et al., 2016 *Fundamentals of Brain Network Analysis*)

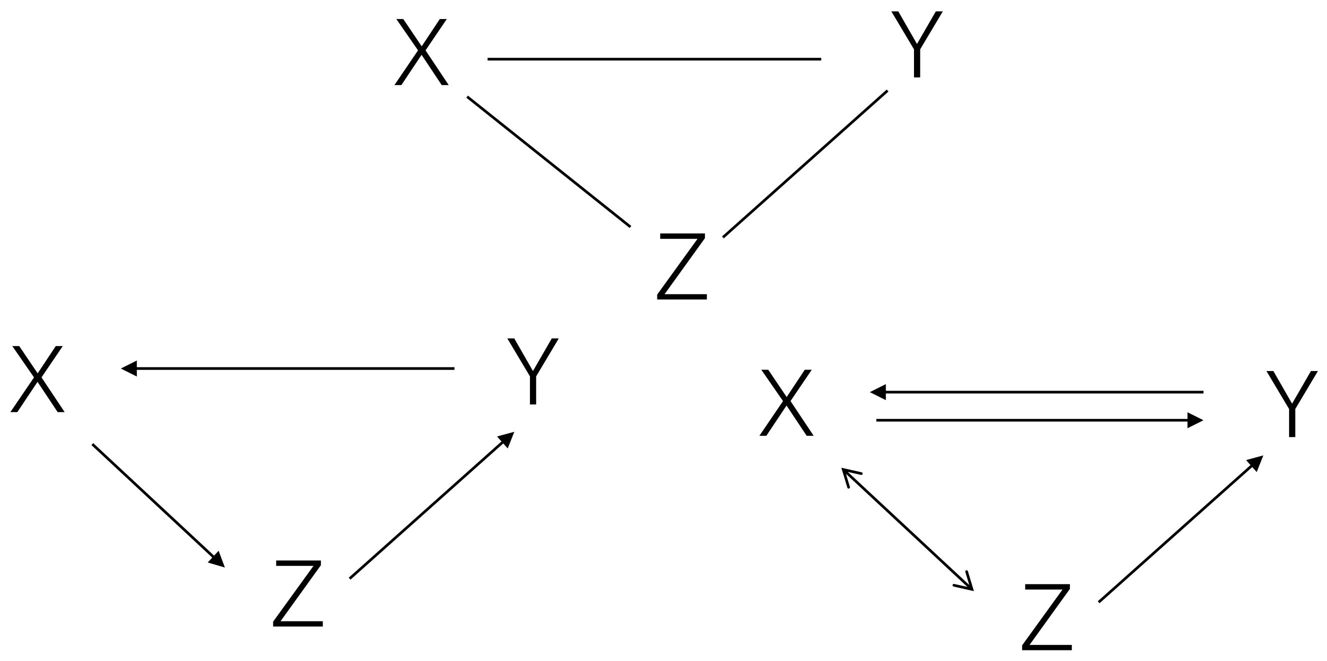
Cool phrase: Directed Acyclic Graph (DAG)

- Feature: cannot have loops / going in the reverse direction

(Causal implications: if we change Z, we influence Y, not X)



✅ DAG



❌ Not DAG

Multiple Regression & Semi-partial or Partial Correlation Coefficients

$$r_{xy.z} = \frac{r_{xy} - r_{xz} r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}$$ =>square it, get $\frac{A}{A+D}$

Partial

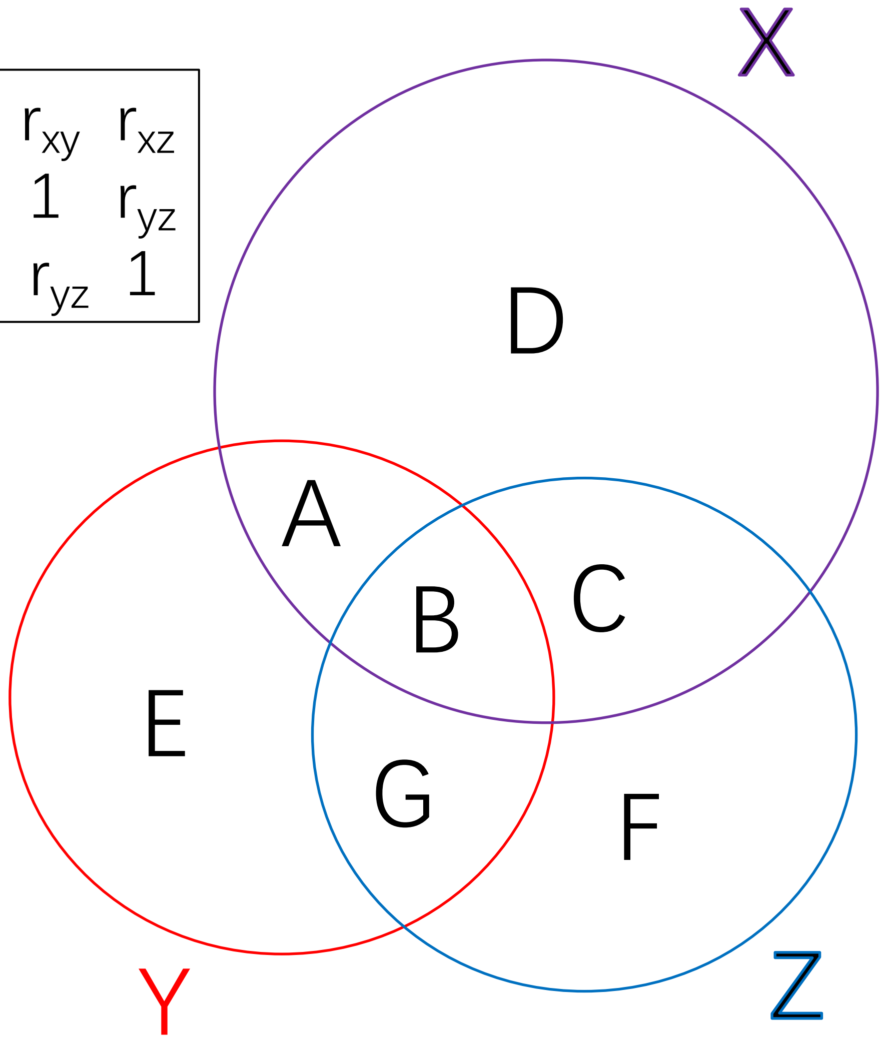
$$\begin{matrix} 1 & r_{xy} & r_{xz} \\ r_{xy} & 1 & r_{yz} \\ r_{xz} & r_{yz} & 1 \end{matrix}$$

$$r_{x(y.z)} = \frac{r_{xy} - r_{xz} r_{yz}}{\sqrt{1-r_{yz}^2}}$$ =>square it, get A

Semi-Partial

$$\beta_y = \frac{r_{xy} - r_{xz} r_{yz}}{1-r_{yz}^2}$$

For X ~ Y+Z, X being the DV
(X, Y & Z all standardised)



Examples on reading the Venn diagram:
-A + B: amount of variance in X explained by Y
-D: amount of variance in X not explained by Y or Z
-A + B + C: the amount of variance in X explained by Y and Z, i.e., the R^2 of X~Y+Z

(Kline SEM)

Application: STAI & STICSA (trait not state)

- Distinction between cognitive and somatic anxiety

Measure	1	2	3	4	5	6	7	8	9	10
1. STICSA State	(.92)									
2. STICSA State, Cognitive	.90	(.88)								
3. STICSA State, Somatic	.88	.59	(.88)							
4. STICSA Trait	.79	.74	.67	(.91)						
5. STICSA Trait, Cognitive	.71	.82	.43	.88	(.87)					
6. STICSA Trait, Somatic	.67	.47	.74	.87	.53	(.87)				
7. STAI State	.65	.63	.53	.58	.54	.47	(.95)			
8. STAI Trait	.60	.65	.42	.66	.70	.49	.71	(.93)		
9. DASS Anxiety	.67	.56	.65	.68	.51	.68	.52	.48	(.83)	
10. DASS Depression	.61	.64	.45	.58	.59	.42	.64	.68	.55	(.92)

Note. Cronbach's alphas are presented in parentheses on the diagonal. All correlations were significant at the .01 level. STICSA = State–Trait Inventory for Cognitive and Somatic Anxiety. STAI = State–Trait Anxiety Inventory. DASS = Depression Anxiety Stress Scales.

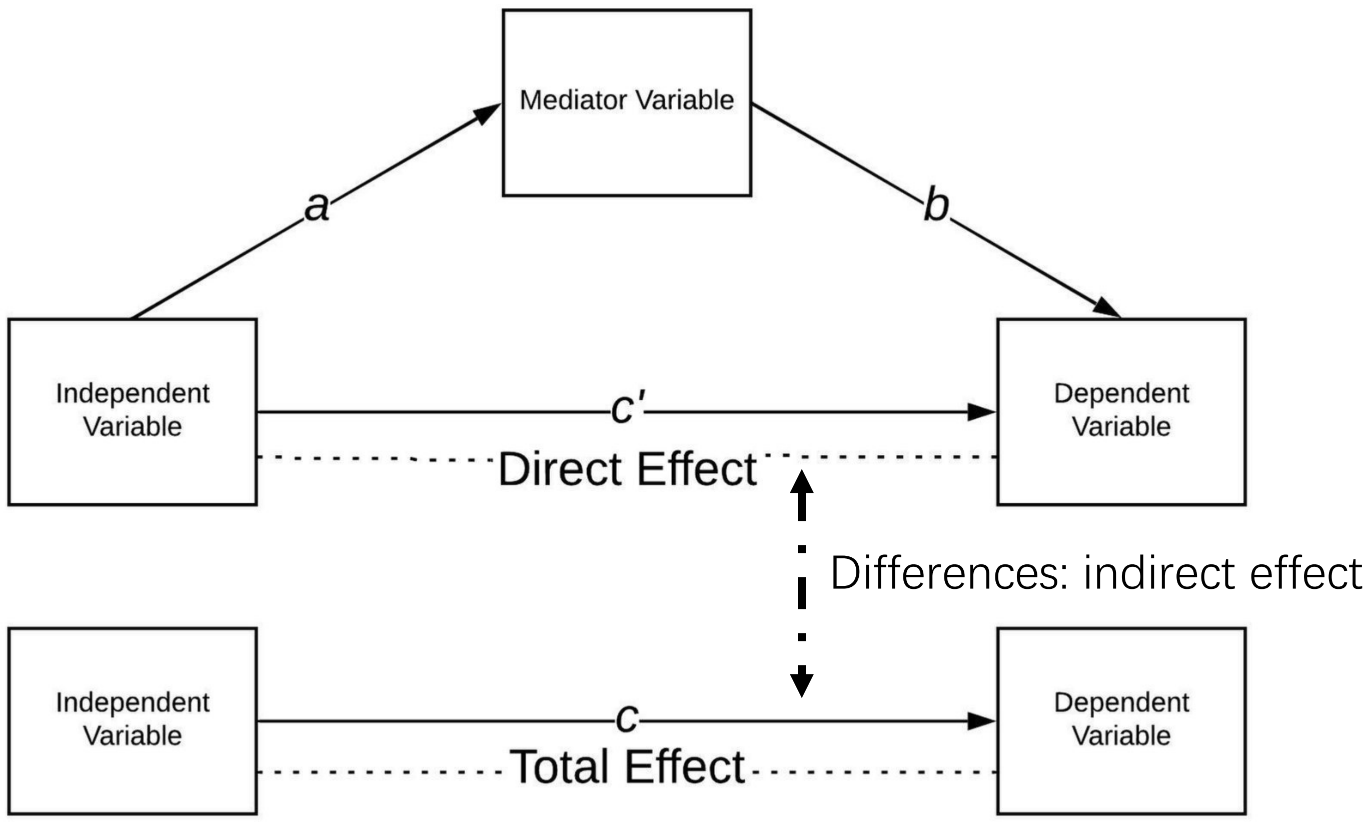
Grös et al., 2007
-Patient cohort
-N=567

STAI: how much variance explained by STICSA_Cognitive versus by STICSA_Somatic)

$r_{\text{STAI(Cognitive. Somatic)}} = 0.52$ (27%) $\quad r_{\text{STAI Cognitive. Somatic}} = 0.60$ (36%)

$r_{\text{STAI(Somatic.Cognitive)}} = 0.14$ (2%) $\quad r_{\text{STAI Somatic.Cognitive}} = 0.20$ (4%)

Path analysis from correlation matrices



(Wikipedia: Mediation Analysis)

$DV = c'*IV + b*MV$

$DV = c*IV$

$MV = a*IV$

(omitted intercept & error term)

Compute correlation coefficients:

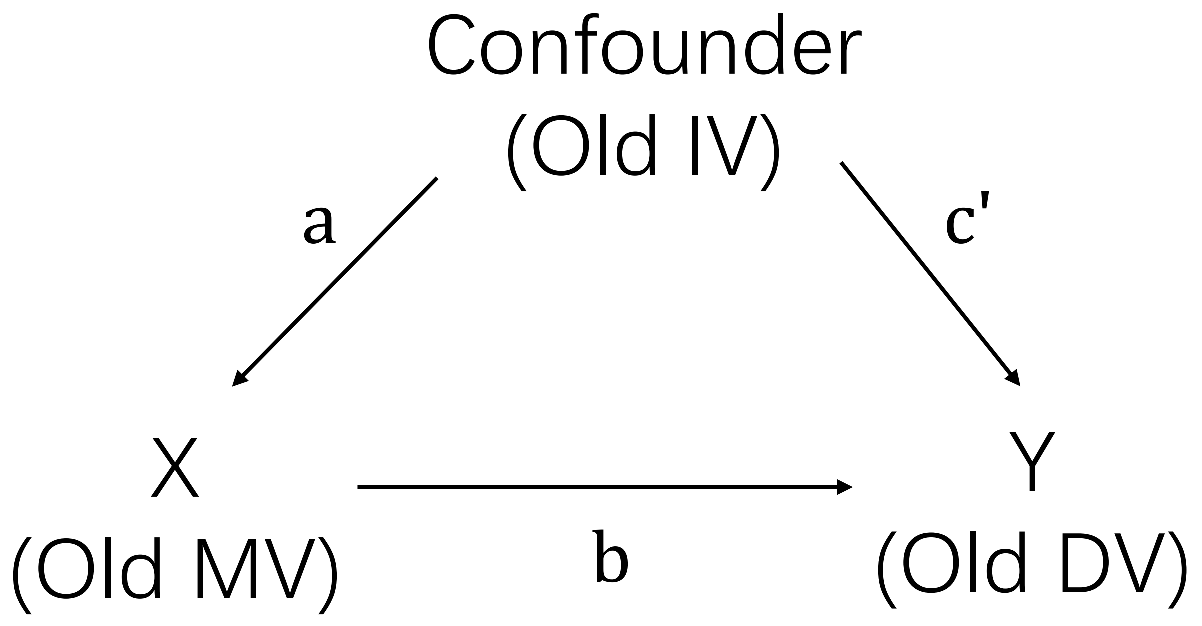
$r_{MV\text{-}IV} = a$ (if standardised)

$r_{DV\text{-}IV} = c = c' + a*b$

$r_{MV\text{-}DV} = b + a*c'$

Indirect effect

Path analysis: do some mental rotations



Path coefficients: the true relationships

$DV = c'^{*}IV + b^{*}MV$

$DV = c^{*}IV$

$MV = a^{*}IV$

(omitted intercept & error term)

Compute correlation coefficients:

$r_{MV\text{-}IV} = a$ (if standardised)

$r_{DV\text{-}IV} = c = c' + a^{*}b$

$r_{MV\text{-}DV} = b + a^{*}c'$

Spurious correlations!
(Extreme case: b =0)

Suppressors

- The **mathematical phenomenon** where the part(ial) correlation or beta regression coefficients change sign or increase from the absolute value of the original bivariate correlation

Kline (2016)

$r_{YX} = 0.0$ $r_{YZ} = 0.2$ $r_{XZ} = 0.1$

$r_{Y(X.Z)}$ = (0-0.1*0.2)/sqrt(1-0.1^2)= -0.2

Kline (2016) example for a good suppressor

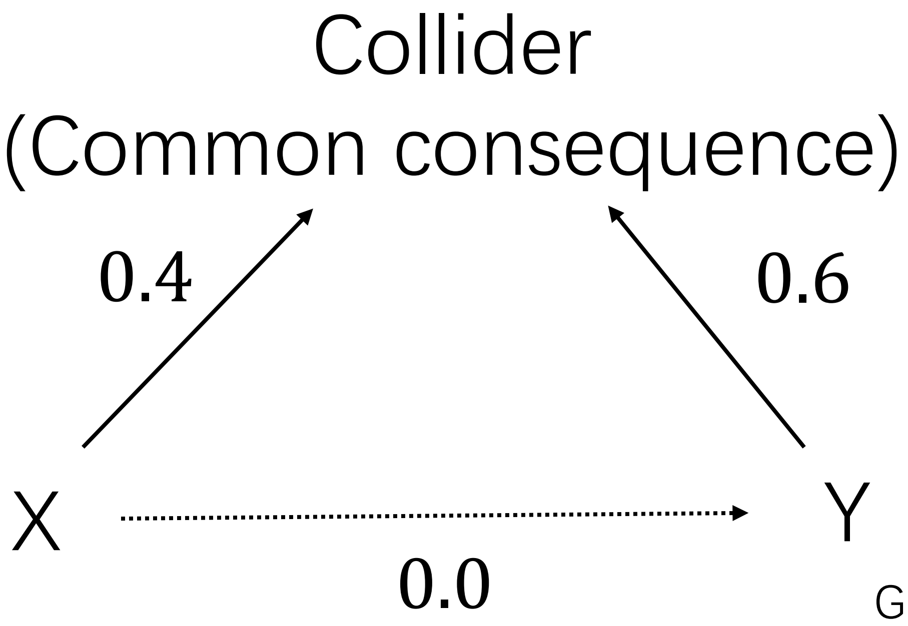
- Y: number of previous suicide attempts
- X1: amount of therapy
- X2: degree of depression
- Bivariate: $r_{Y\text{-}X1}$=0.19, $r_{Y\text{-}X2}$=0.49, $r_{X1\text{-}X2}$ = 0.70

- But if we do the regression…

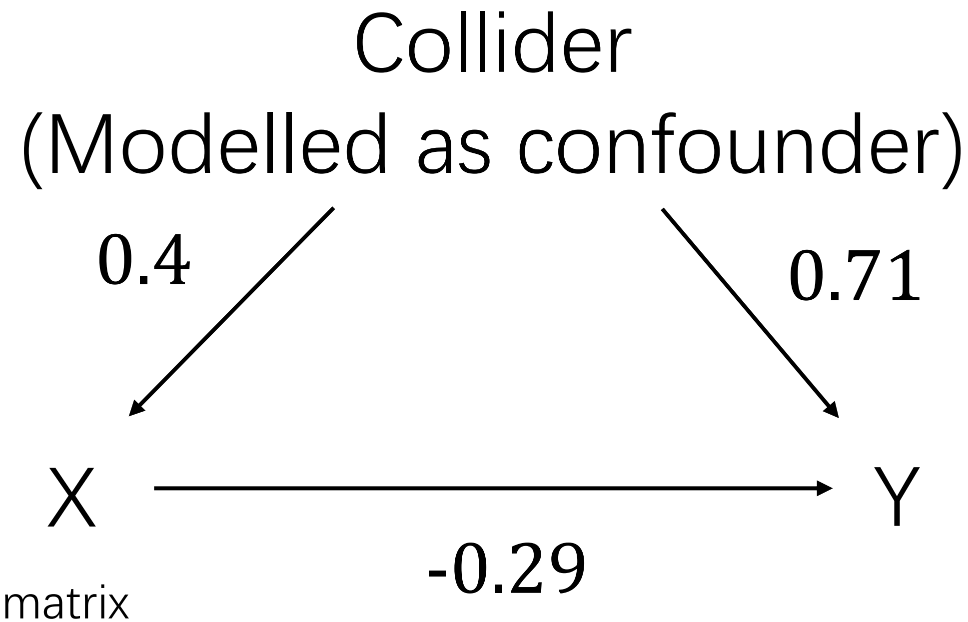
Suicide attempt ~ Depression + Therapy
Standardised regression coefficient of therapy: -0.3 <= negative effect!
Standardised regression coefficient of depression: 0.70

Colliders as hindrance to inference

- Colliders can appear as suppressors → spurious associations!
- Using the equations in slide 8 and 9, all variables standardised:



(Ground truth)



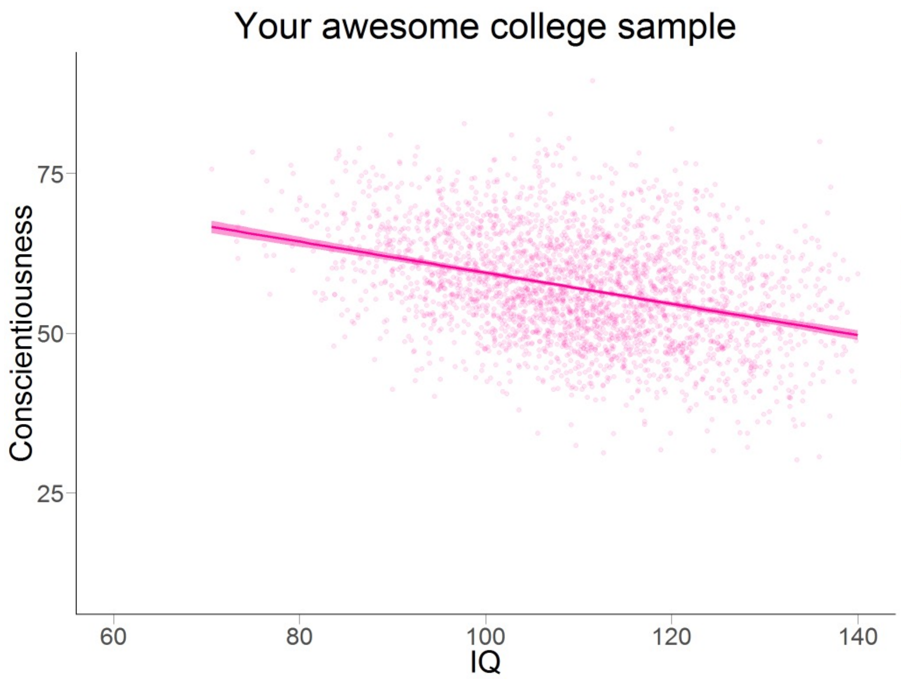
Generate correlation matrix

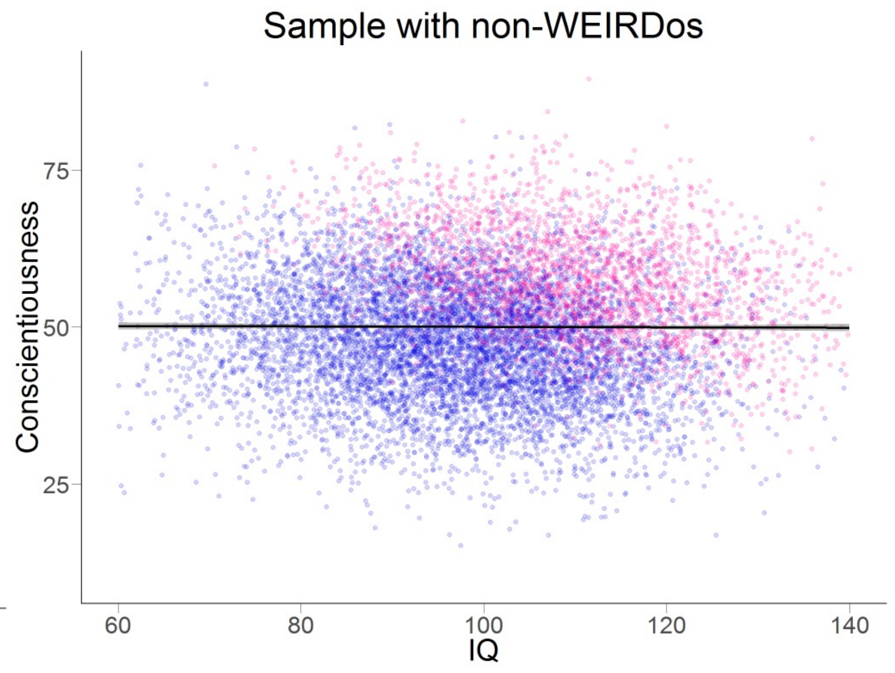
X	1.0000	0	0.4000
Y	0	1.0000	0.6000
C	0.4000	0.6000	1.0000

Colliders without statistical control

(Another thought experiment)

- Conscientiousness and IQ: -ve correlation?
- Conditioned on University Membership (typical pysch experiment)
  - Not problematic? Know X & Z, know Y → predictive value





<= Figures from Rohrer http://www.the100.ci/2017/03/14/that-one-weird-third-variable-problem-nobody-ever-mentions-conditioning-on-a-collider/

Summary

- Graphs can model the relationships amongst many variables
- Even a DAG with three nodes can get complicated…
- The possibility of colliders: do not throw everything in GLM blindly
  - (No panacea for spurious associations)

- The question remains: how might conditioning on the wrong variables distort *our* results? ("Who is safe?")

Thinking about what to think…
Should I really care about causal inferences?

- No causal claims made – free from colliders?
- The paradoxical curse & blessing of small effect sizes?
- (The actual hypotheses) we examine complicated interactions…

Thank you for listening!

Extra slides

The chiasma of causality & effect size?

	Large effect size	Small effect size
Casual inference	Easy science??	Ontology / Epistemic value (the TRUTH!)
Make predictions based on mere statistical associations	Geocentric / Real-world applications	Bad science??

Note:
-Instances where small effect sizes can have huge real-world impact?

Important Distinction: Statistical association vs Generative Process (Causal)

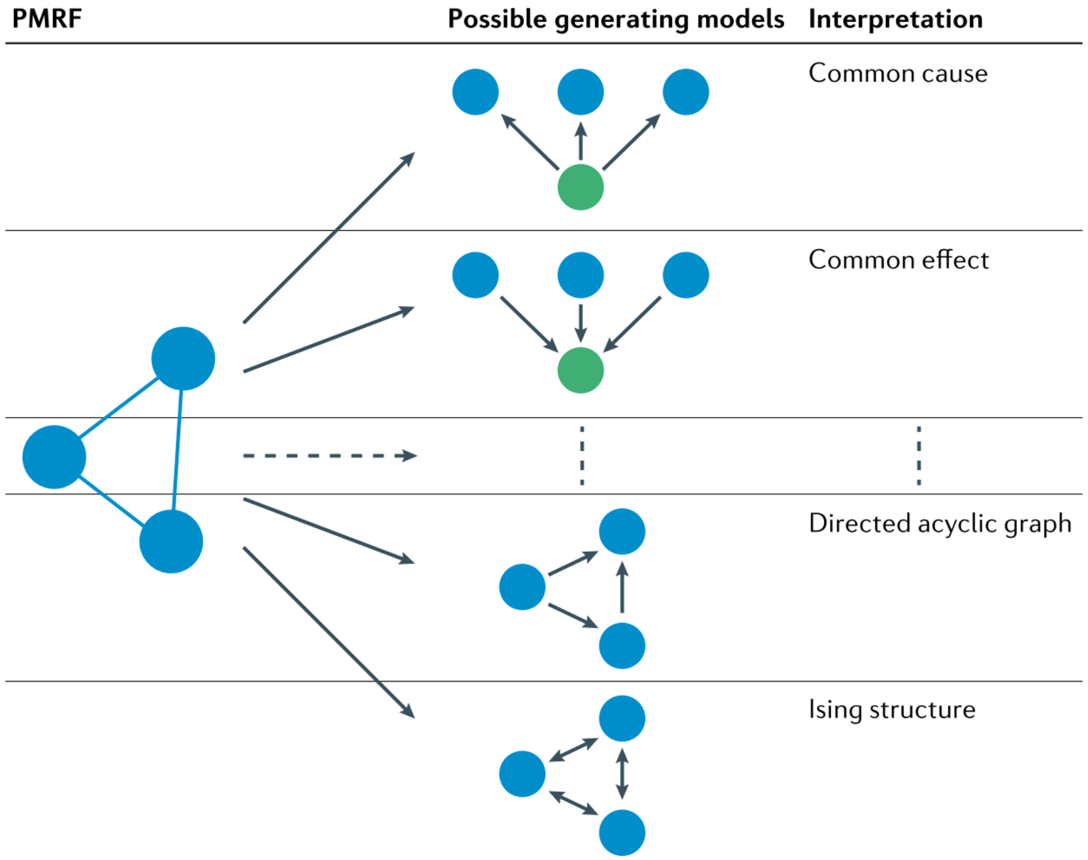


Fig. 9 | **From statistical model to causal inference.** Pairwise Markov random field (PMRF) (left) can be generated by alternative models (middle) that have different interpretations (right). Dashed lines represent range of models and interpretations not captured here.

(Borsboom et al., 2021)

Kline (2016)

$$r^2_{Y(X \cdot W)} = a = R^2_{Y \cdot X,W} - r^2_{WY} \tag{2.17}$$

$$r^2_{Y(W \cdot X)} = b = R^2_{Y \cdot X,W} - r^2_{XY}$$

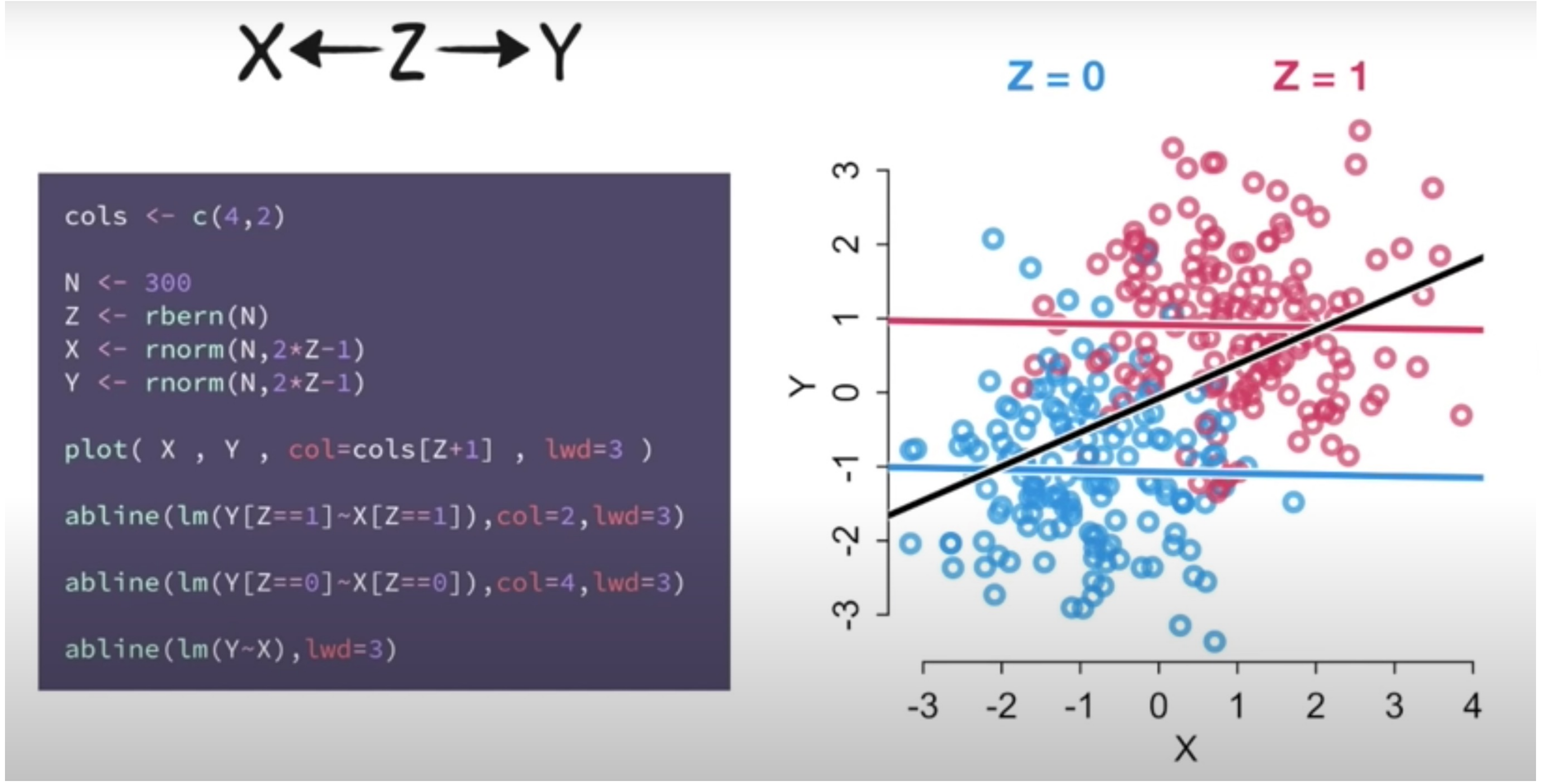
The squared partial correlations correspond to areas *a*, *b*, and *d* in Figure 2.3, and each estimates the proportion of variance in the criterion explained by one predictor but not the other. The formulas are

$$r^2_{XY \cdot W} = \frac{a}{a+d} = \frac{R^2_{Y \cdot X,W} - r^2_{WY}}{1 - r^2_{WY}} \tag{2.18}$$

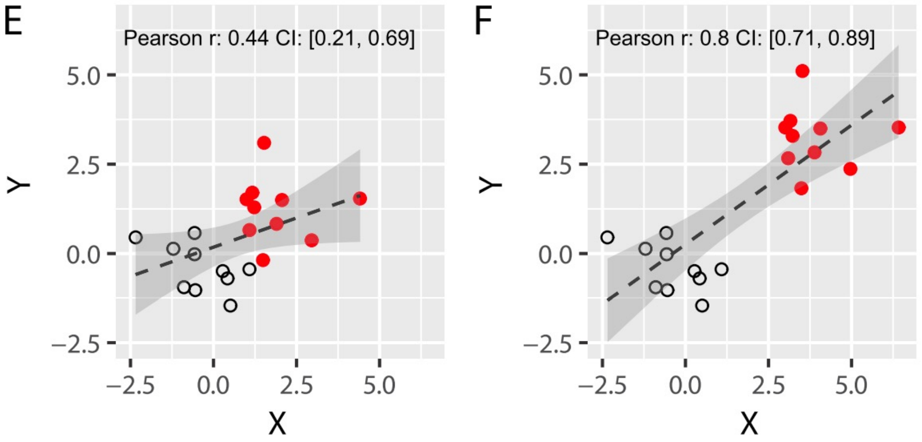
$$r^2_{WY \cdot X} = \frac{b}{b+d} = \frac{R^2_{Y \cdot X,W} - r^2_{XY}}{1 - r^2_{XY}}$$

For the data in Table 2.1, $r^2_{Y(X \cdot W)} = .327$ and $r^2_{XY \cdot W} = .435$. In words, predictor *X* uniquely explains .327, or 32.7% of the total variance of *Y* (squared part correlation). Of the variance in *Y* not already explained by *W*, predictor *X* accounts for .435, or 43.5% of the remaining variance (squared partial correlation). Exercise 7 asks you to calculate and interpret the corresponding results for the other predictor, *W*, and the same data.

Regression as stratification



(Figure from Statistical Rethinking)



Makin and Orban de Xivry (2019)