Causal Modelling, DAGs, & SEMs for Marine Scientists

From Theory to Practice

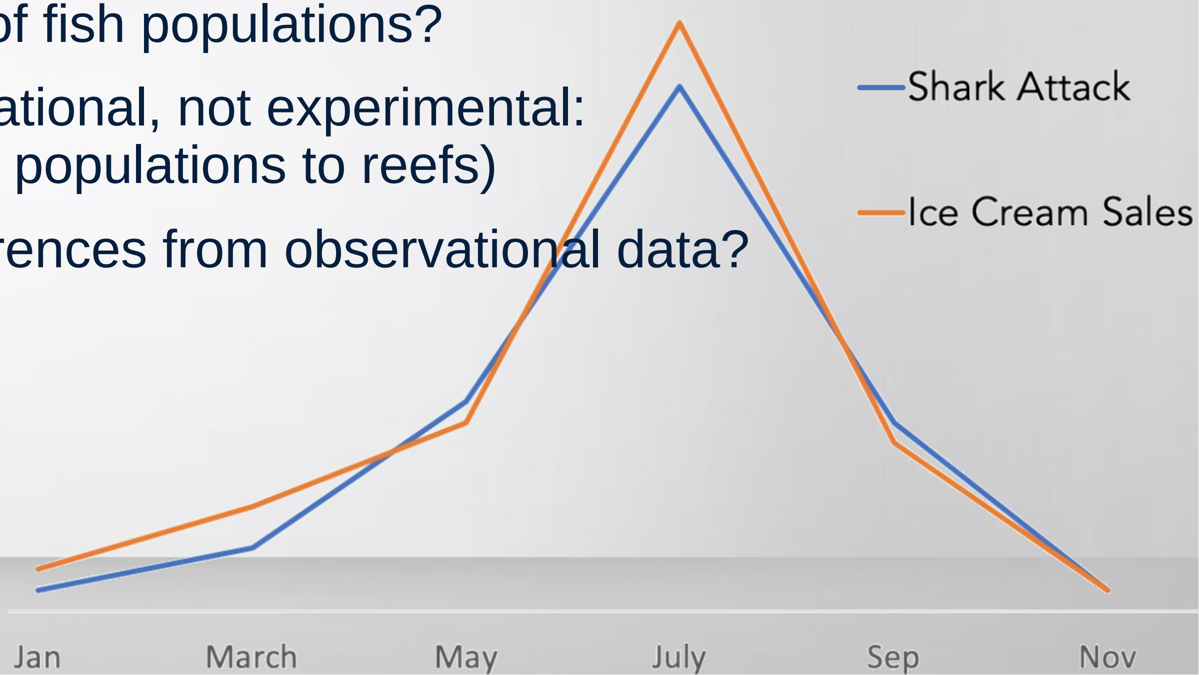
Simon Dedman | Florida International University

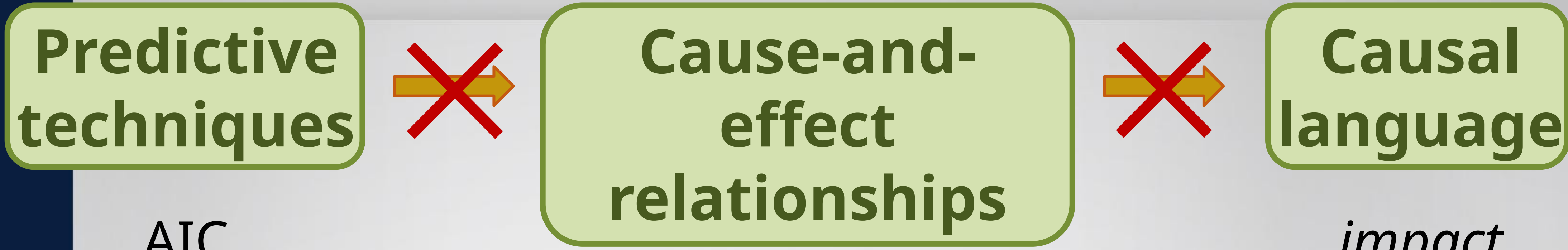
Today, now, 2026

Why Causal Inference?

The Fundamental Problem

- Correlation does not imply causation... but we need causal answers.
- Marine ecology is fundamentally about causal questions:
  - Do sharks control reef health from the top down?
  - Does fishing cause ecosystem degradation?
  - Do MPAs cause recovery of fish populations?
- Most ecological data are observational, not experimental:
  (we can’t randomly assign shark populations to reefs)
- So how do we make causal inferences from observational data?





Predictive techniques

Cause-and-effect relationships

Causal language

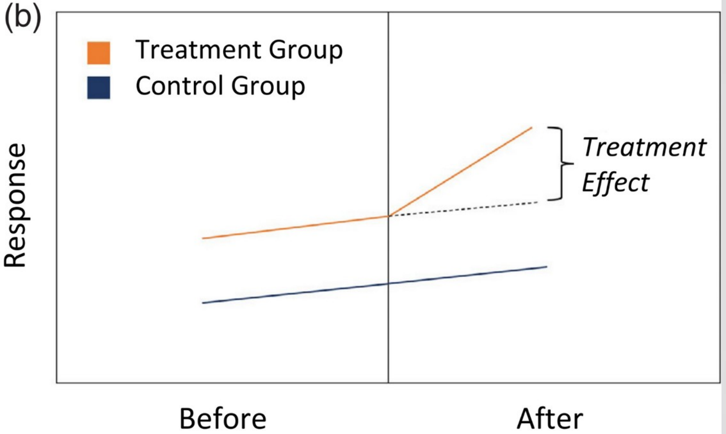
AIC
BIC
WAIC
LOO-CV

X Y

impact
effect
driver
influence

Predictive inference	Causal inference
“Which model best predicts Y?”	“What is the effect of X on Y?”
Associative/correlative	Causal

The Causal Revolution



- Judea Pearl's 'Ladder of Causation' (Pearl 2009, The Book of Why 2018):
- Rung 1: Association (seeing) -- What if I see…? Patterns in data; how variables are related. $P(B|A)$: probability of seeing B given we see A.
- Rung 2: Intervention (doing) -- What if I do? $P(B|do(A))$: probability of B if I do A? (e.g. BACI, before/after control/intervention)
- Rung 3: Counterfactual (imagining) -- What if I had done C? $P(B_x|A)$: probability of B being B_x given we observed A? “you never would have gotten lung cancer, if you had never smoked”.
- Most statistics operate on Rung 1
- Causal inference tools let us climb to Rung 2 with observational data
- Key frameworks: Pearl's do-calculus, Rubin's potential outcomes
- McElreath (2020): DAGs as core tool in Statistical Rethinking
- Ecology adoption: Schoolmaster et al. (2020), Arif & MacNeil (2022, 2023)

Ecology's Causal Challenge

- What We Want:
  - Mechanistic understanding
  - Causal effects of management
  - Predictions under intervention
  - Evidence-based conservation

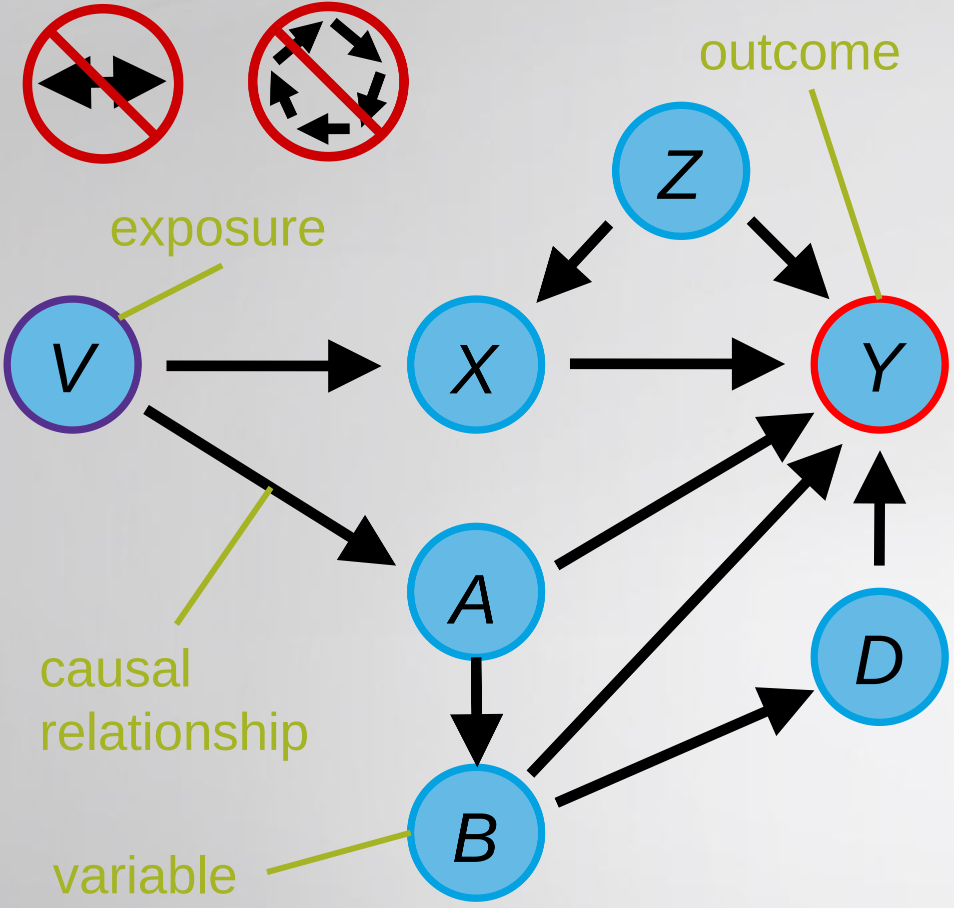
Directed Acyclic Graphs DAGs

What is a DAG?

- Directed: each arrow points from cause to effect
- Acyclic: no variable can indirectly cause itself (no feedback loops)
- Graph: nodes (variables) connected by edges (arrows)
- Encodes your causal assumptions about the system
- Not derived from data: comes from domain knowledge
- Qualitative causal model, not statistical
- Once you have a DAG, it tells you exactly what to control for (what to include in subsequent statistical models)
- Key references: Pearl (2009), Elwert (2013), Textor et al. (2016)

Create DAG

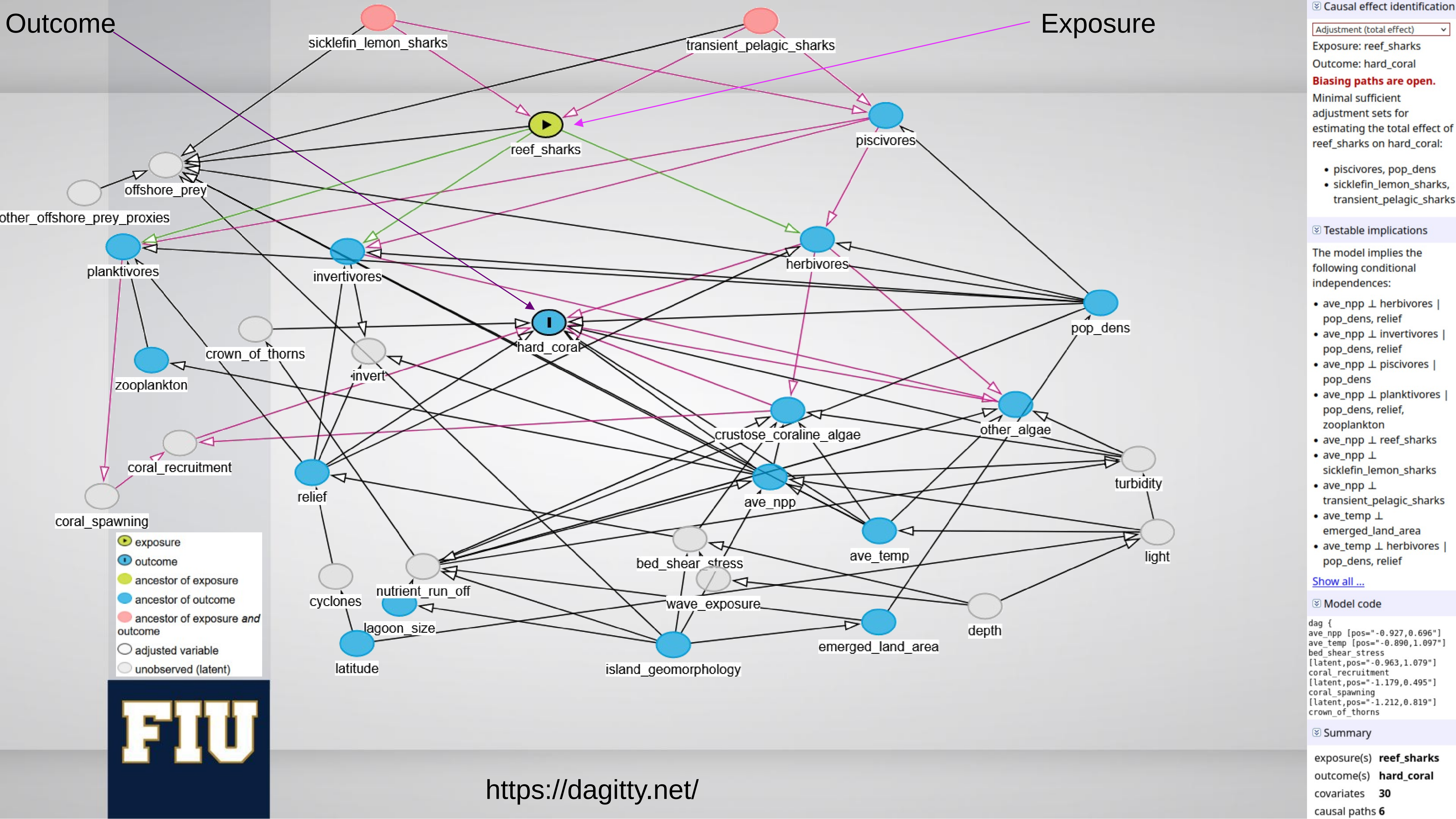
Directed Acyclic Graph



- Sources of knowledge for DAG construction:
  - Published literature on the system
  - Domain expertise and ecological theory
  - Prior experimental evidence
  - Known physical/chemical mechanisms
- Practical tips:
  - Start with the exposure-outcome pair of interest
  - Ask: 'What else affects both the exposure and outcome?'
  - Include temporal ordering (causes precede effects)
  - Don't include variables that are consequences of both X and Y
  - Build progressively: add one relationship at a time
  - Use dagitty.net for interactive DAG building
- When presenting:
  - Color-code arrows (green = positive, red = negative)
  - Keep the DAG as simple as possible

Outcome

Exposure



Causal effect identification

Adjustment (total effect)

Exposure: reef_sharks

Outcome: hard_coral

Biasing paths are open.

Minimal sufficient adjustment sets for estimating the total effect of reef_sharks on hard_coral:

- piscivores, pop_dens
- sicklefin_lemon_sharks, transient_pelagic_sharks

Testable implications

The model implies the following conditional independences:

- ave_npp ⊥ herbivores | pop_dens, relief
- ave_npp ⊥ invertivores | pop_dens, relief
- ave_npp ⊥ piscivores | pop_dens
- ave_npp ⊥ planktivores | pop_dens, relief, zooplankton
- ave_npp ⊥ reef_sharks
- ave_npp ⊥ sicklefin_lemon_sharks
- ave_npp ⊥ transient_pelagic_sharks
- ave_temp ⊥ emerged_land_area
- ave_temp ⊥ herbivores | pop_dens, relief

Show all ...

Model code

```
dag {
ave_npp [pos="-0.927,0.696"]
ave_temp [pos="-0.890,1.097"]
bed_shear_stress
[latent,pos="-0.963,1.079"]
coral_recruitment
[latent,pos="-1.179,0.495"]
coral_spawning
[latent,pos="-1.212,0.819"]
crown_of_thorns
```

Summary

exposure(s)	**reef_sharks**
outcome(s)	**hard_coral**
covariates	**30**
causal paths	**6**

https://dagitty.net/

Worked Example: Caudal Fin Morphology

- Aitchison et al.: What drives caudal fin aspect ratio (CFAR) in sharks?

Step 1: Build the DAG

- 8 variables: Avg Temperature, Temp Variability, Dissolved O2,
- Water Column Position, Max Body Size, Gill Surface Area,
- Gill Slit Height, Diet Level → CFAR
- Data: Sharks of the World, IUCN Red List, NOAA, VertLife phylogeny

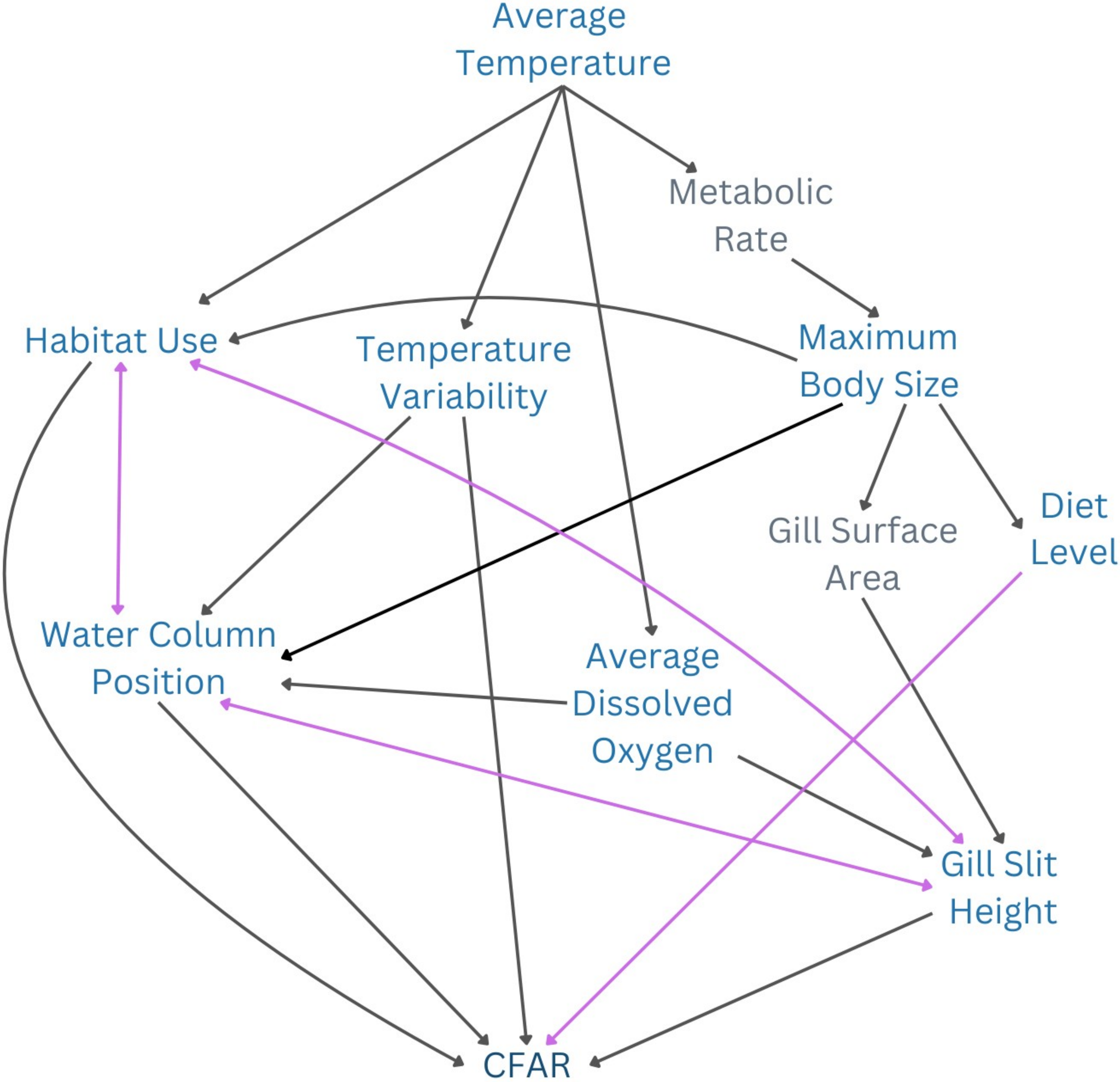
Step 2: Argue about the DAG

- Simplified to 5 key variables after expert discussion
- Used phylopath (van der Bijl 2018) to account for phylogeny

Step 3: Run statistical models (brms)

- Results: Max Body Size (+0.30) and Avg Temperature (-0.21)
- strongest predictors of CFAR
- Path diagram: arrows color-coded & width scaled by effect size

Rachel’s DAG



Average Temperature

Metabolic Rate

Habitat Use

Temperature Variability

Maximum Body Size

Gill Surface Area

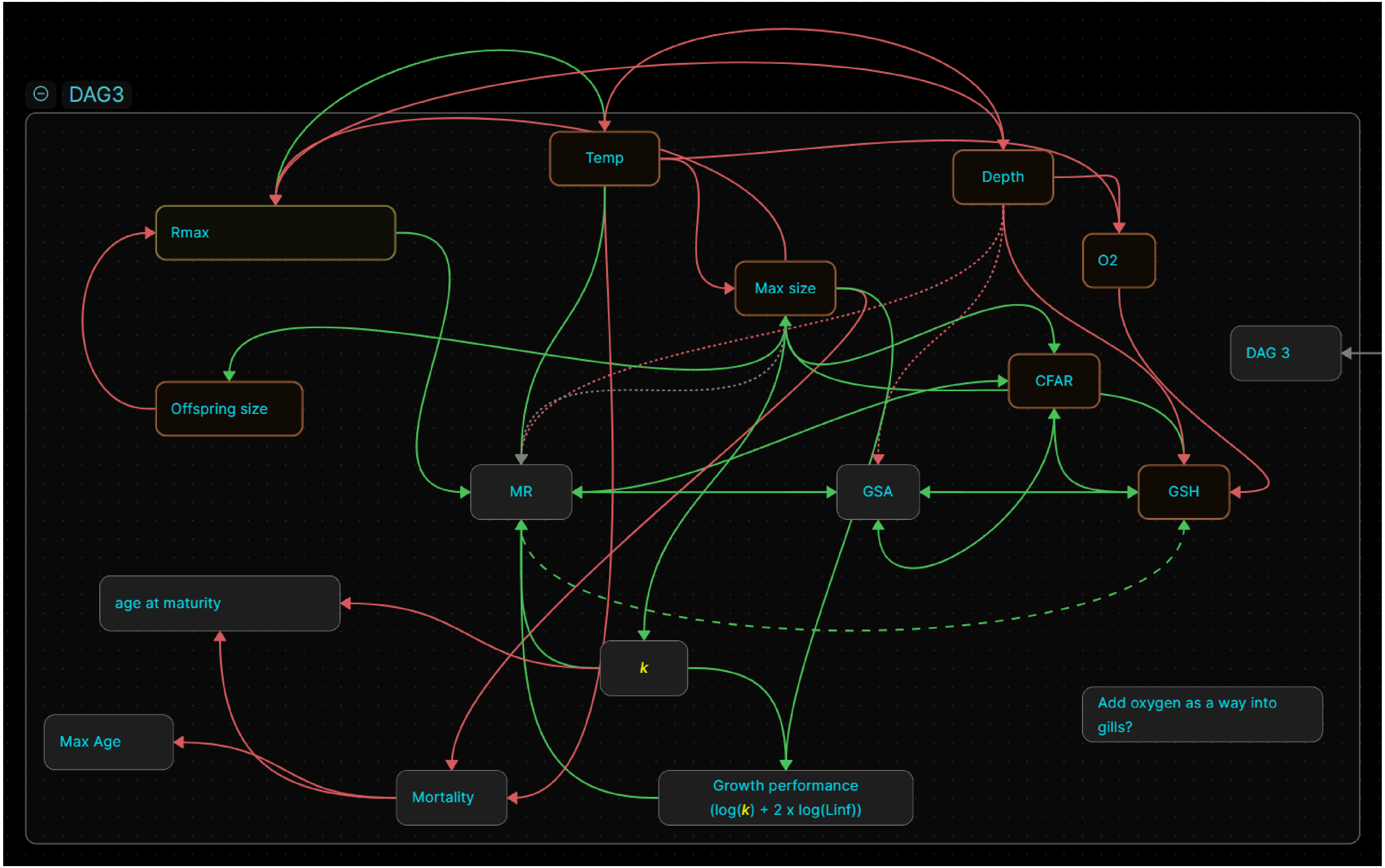
Diet Level

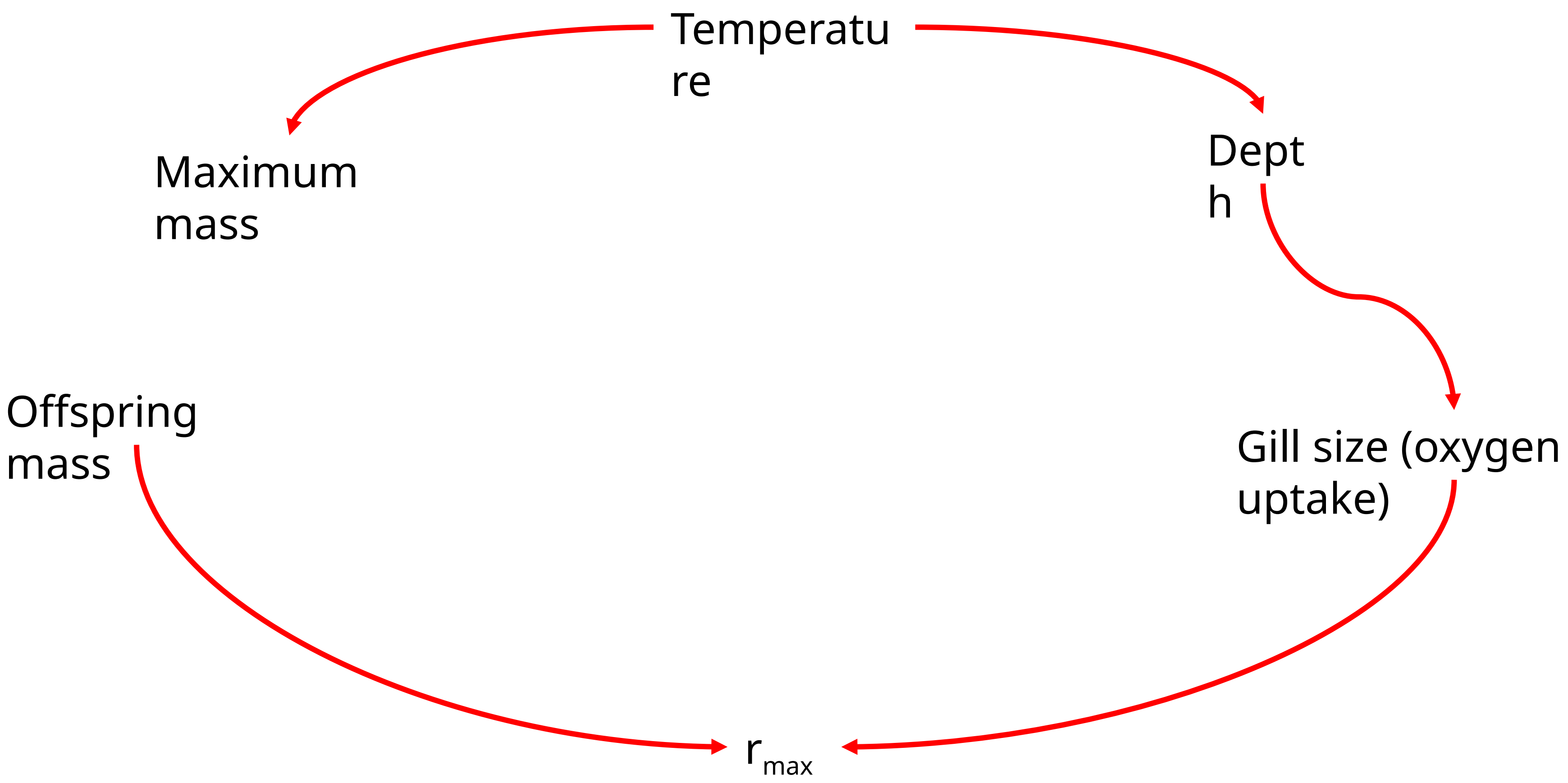
Water Column Position

Average Dissolved Oxygen

Gill Slit Height

CFAR


DAG3
Temp
Depth
Rmax
O2
Max size
DAG 3
CFAR
Offspring size
MR
GSA
GSH
age at maturity
k
Max Age
Mortality
Growth performance
(log(k) + 2 x log(Linf))
Add oxygen as a way into gills?



Temperature

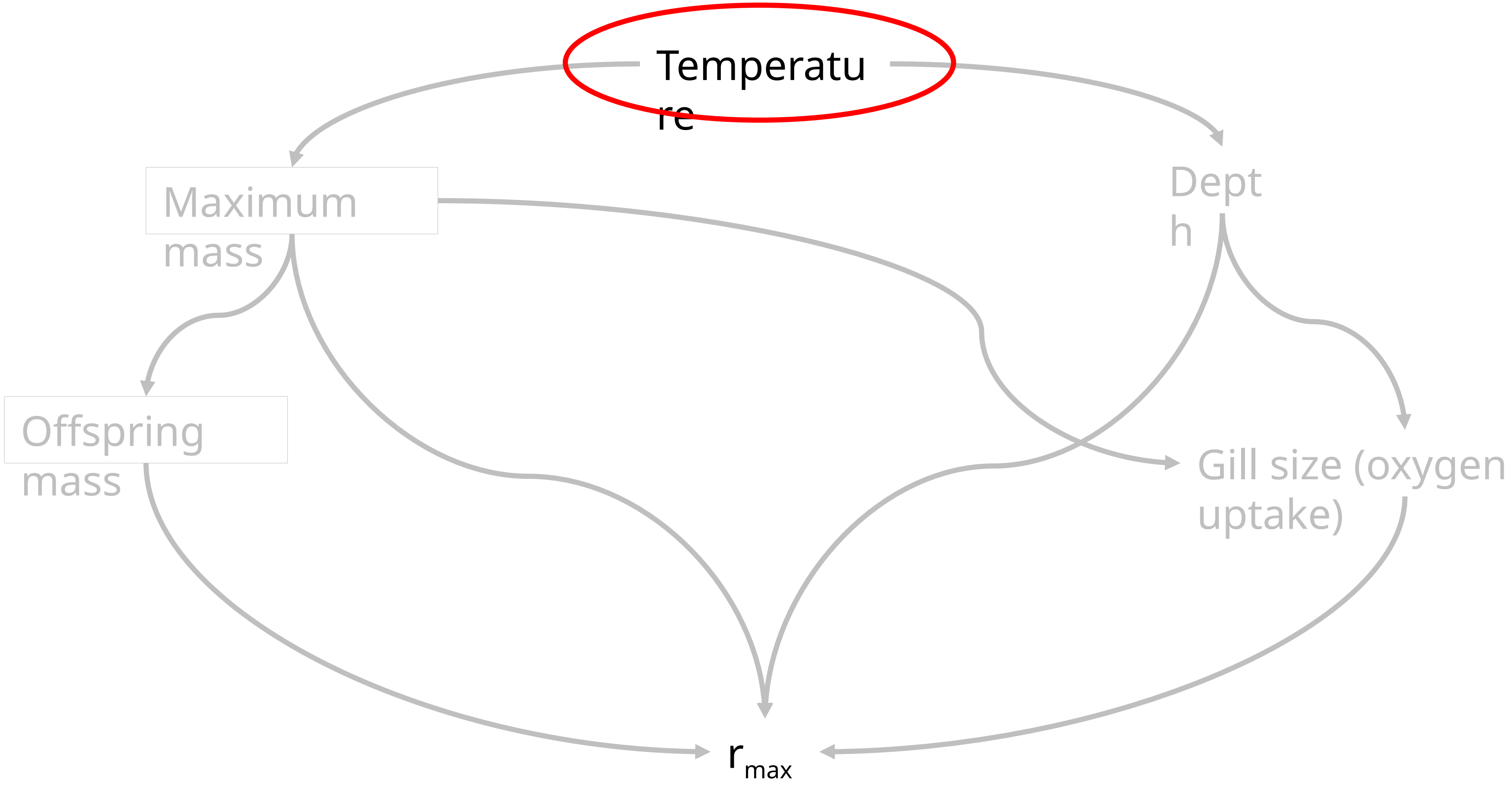
Maximum mass

Depth

Offspring mass

Gill size (oxygen uptake)

r_{max}



Temperature

Maximum mass

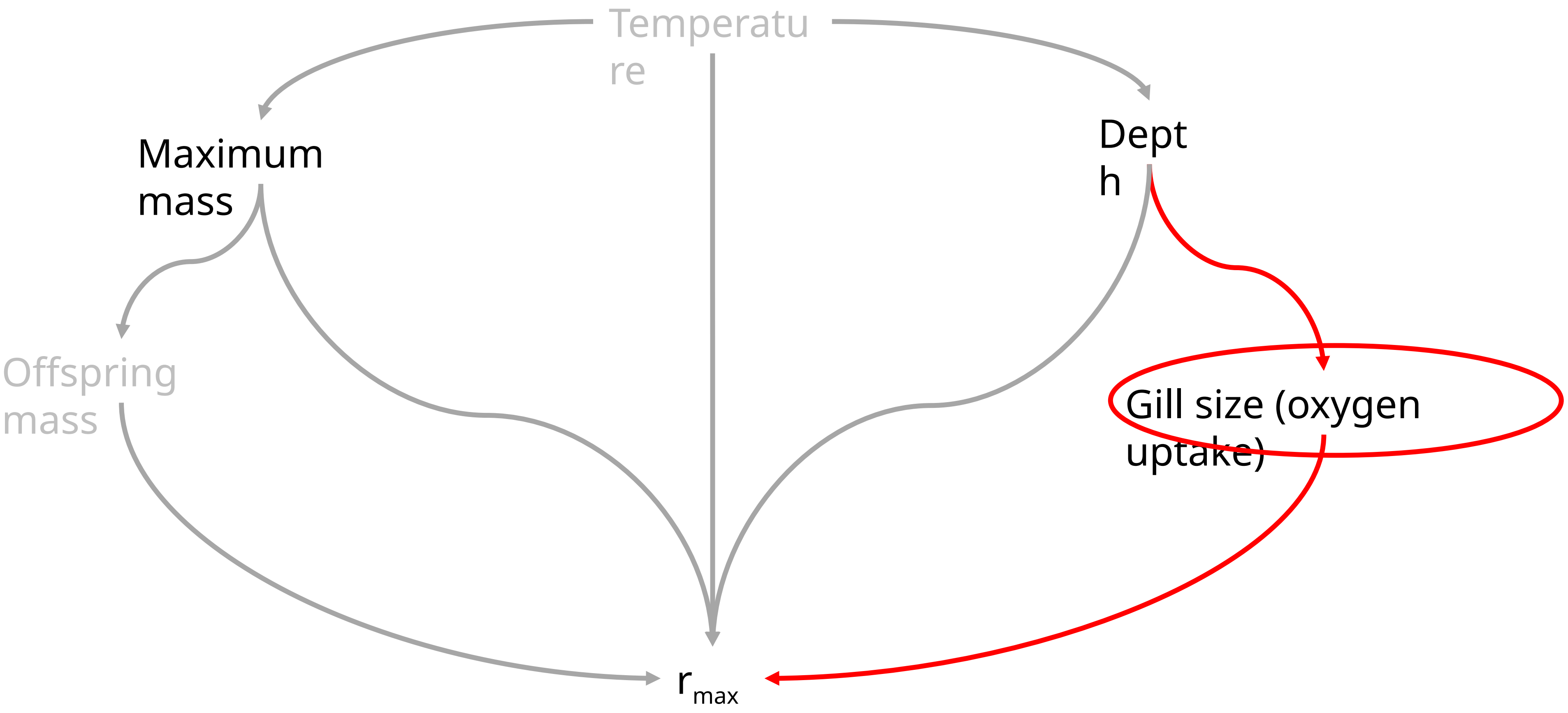
Depth

Offspring mass

Gill size (oxygen uptake)

r_{max}

$$ln(r_{max}) \sim \beta_0 + \beta_1 * ln(temperature)$$



Temperatu
re

Maximum
mass

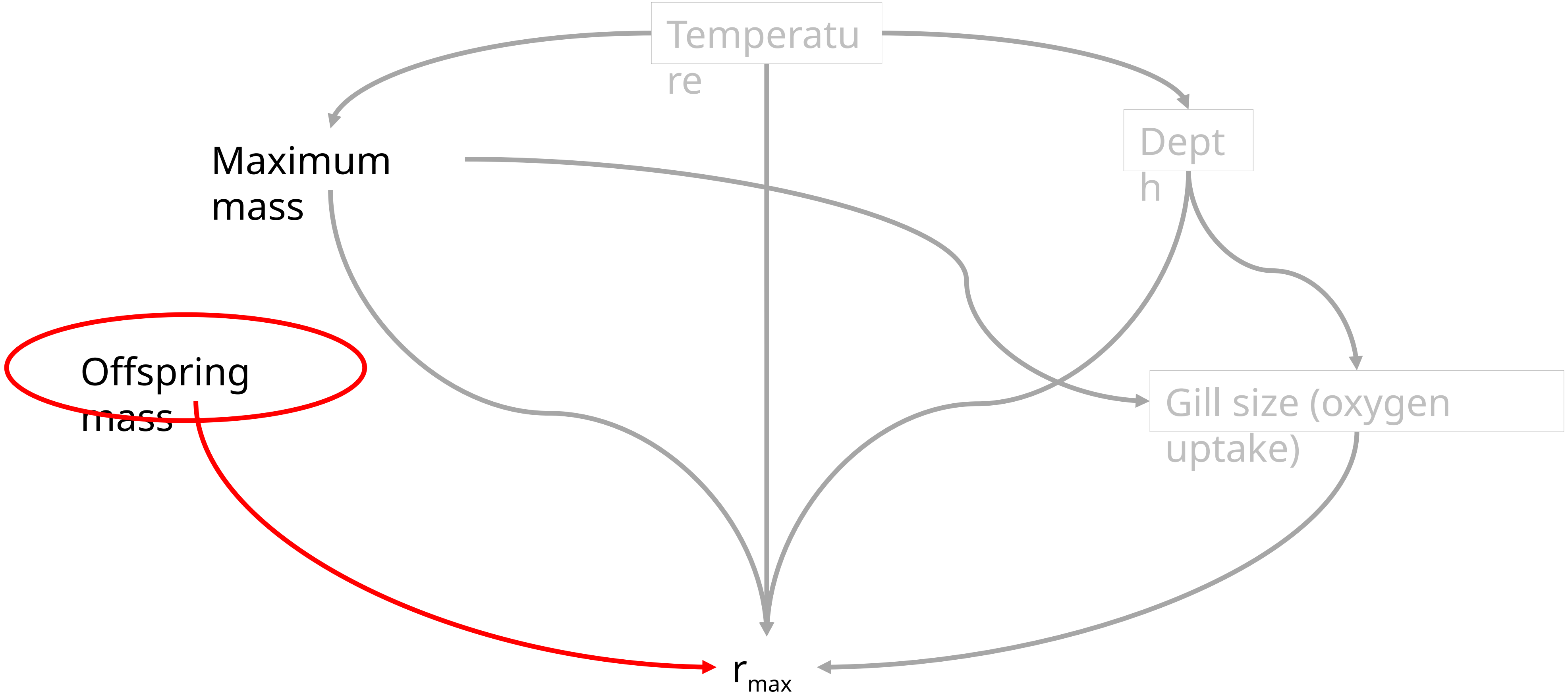
Dept
h

Offspring
mass

Gill size (oxygen
uptake)

r_{max}

$$ln(r_{max}) \sim \beta_0 + \beta_1 * ln(\max mass) + \beta_2 * ln(depth) + \beta_3 * ln(gill\ size)$$



Temperature

Maximum mass

Depth

Offspring mass

Gill size (oxygen uptake)

r_{max}

$$ln(r_{max}) \sim \beta_0 + \beta_1 * ln(\max mass) + \beta_2 * ln(offsp.mass)$$

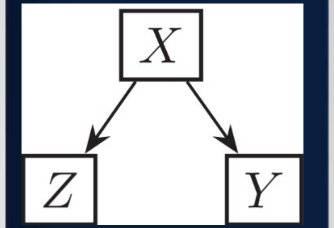
Building a DAG: Key Principles

- 1. Include ALL common causes of any two variables in the DAG
    - (Completeness criterion: Spirtes et al. 2000)
- 2. Variables can be measured OR unmeasured
    - Mark unmeasured variables clearly
- 3. Use domain knowledge, prior literature, and logic
- 4. Be explicit about what you assume does NOT have an effect
    - (Missing arrows are strong assumptions!)
- 5. Iterate: check assumptions, refine, repeat. 'Argue' about DAG with colleagues: leads to meaningful simplification
    - (cf. Aitchison et al. -- shark CFAR example: 8 variables reduced to 5)
- 6. DAGs are qualitative -- they show presence/direction, not strength
- 7. The DAG is your most important modelling decision

Workshop Exercise 1: Build a DAG: 10-15 mins

- Think about a system YOU study. Identify:
- 1. Your exposure (treatment/predictor of interest) right of glm
- 2. Your outcome (response variable) left of glm
- 3. All common causes of exposure and outcome
- 4. Any mediators between exposure and outcome
- 5. Any variables that are caused by BOTH exposure and outcome
- Draw the DAG on paper or at dagitty.net
- Check: Are there any backdoor paths? Dagitty.net
- What is the minimal sufficient adjustment set?
- Discuss with your neighbour:
- Have you included all common causes?
- Are there unmeasured confounders?
- Are there any colliders you might accidentally condition on?

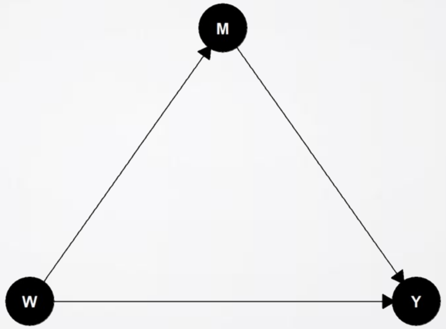
The Fork: Confounding, Backdoor Paths



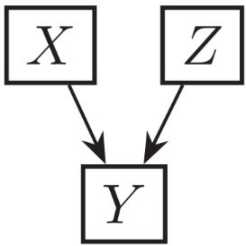
- Example: Ice cream sales and shark attacks are correlated
  - Ice Cream <- Temperature -> Shark attacks
  - Temperature is the confounder (common cause)
- Marine example: Shark abundance and coral cover are correlated
  - Sharks <- Productivity -> Coral
  - Could productivity be driving both?
- Rule: To estimate Z -> Y, you MUST block backdoor paths
- A backdoor path flows through a common cause (X)
- Block it by conditioning on the common cause (add X to the model)
  - To estimate the effect of shark abundance on coral cover, include productivity in your statistical model.

The Pipe/Chain/Mediator

- Lies between exposure and outcome: X -> Z -> Y
- Z mediates the effect of X on Y
- Pipe: DON’T condition on (include) Z if you want the total effect of X on Y
- Triangle: Total effect = Direct Effect (W → Y) + Indirect Effect (W → Z → Y)



The Collider: A Dangerous Trap



- The name: In DAGs, arrowheads ‘collide’ on the bottom node.
- McElreath's example: Trustworthiness and Newsworthiness
  - Grant panels select on BOTH Trustworthiness and Newsworthiness
  - If you only look at (filter for) funded studies:
  - Highly newsworthy studies can get funded regardless of trustworthiness. Highly trustworthy studies can get funded regardless of newsworthiness.
  - Can create spurious negative correlation. Berkson's paradox / selection-distortion effect.
- Marine example: Body size and swimming speed both affect survival
  - Body Size -> Survival <- Swimming Speed
  - If you *only* study survivors, size and speed can appear negatively correlated:
  - If you survived just because you’re fast, or just because you’re big, then fast+small and big+slow survivors can create a negative relationship.
- Rule: NEVER condition on a collider (or its descendants)
- Doing so OPENS a path that was previously blocked

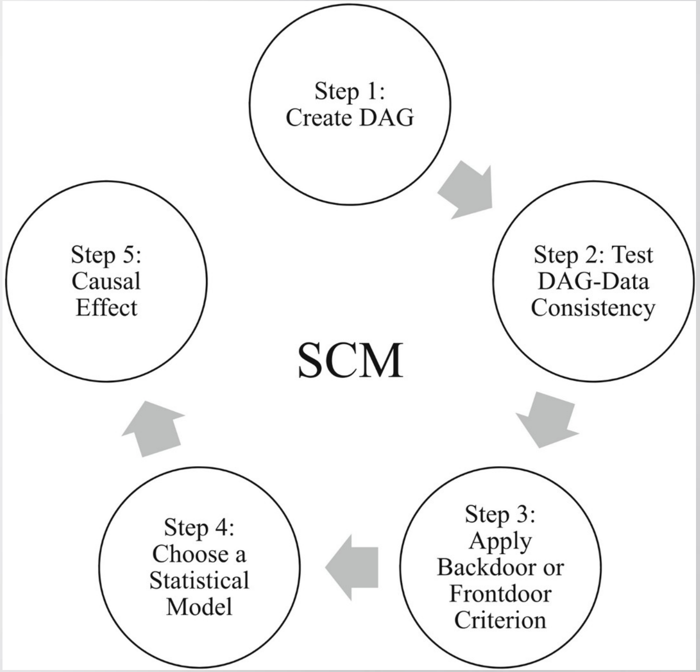
Structural Causal Models

The Schoolmaster et al. 2022 Framework

Three-step approach for observational causal inference in ecology:

- Step 1: BUILD YOUR DAG
    - Encode domain knowledge as a directed acyclic graph
    - Include all common causes, measured and unmeasured
- Step 2: CHECK DAG-DATA CONSISTENCY
    - Test implied conditional independences against data
    - Use dagitty R package for formal tests
    - If independencies fail, refine the DAG structure
- Step 3: IDENTIFY ADJUSTMENT SETS*
    - Apply the backdoor criterion*
    - Determine what to control for in statistical model*
    - Fit the model with the correct adjustment set*

*Dagitty does this for you



Step 2: DAG-Data Consistency

- A DAG implies many conditional independencies
  - e.g., 'X is independent of Y, given Z'
- These can be tested against your observational data
- Using dagitty in R:
  - library(dagitty)
  - dag <- dagitty('dag{ X -> Y; Z -> X; Z -> Y }')
  - impliedConditionalIndependencies(dag)
  - localTests(dag, data = your_data)
- If tests PASS: supports your DAG structure
- If tests FAIL: indicates problems with asserted structure
  - Missing arrow? Add a connection
  - Wrong direction? Reverse an arrow
  - Iterate until the DAG passes all tests

Workshop Exercise 2: dagitty in R

- Define a simple marine ecology DAG
- library(dagitty)
- dag <- dagitty('dag{
-   Fishing -> Sharks
-   Fishing -> Teleosts
-   Sharks -> Teleosts
-   Sharks -> CoralHealth
-   Teleosts -> CoralHealth
-   Temperature -> Sharks
-   Temperature -> CoralHealth
- }')
- What should we control for?
- adjustmentSets(dag, exposure='Sharks', outcome='CoralHealth')
- Test implied independencies
- impliedConditionalIndependencies(dag)

Step 3: The Backdoor Criterion

- To estimate the causal effect of X on Y:
    - Block all BACKDOOR PATHS from X to Y
    - A backdoor path has an arrow pointing INTO X
- The adjustment set = variables to include as covariates
- Using dagitty:
    - `adjustmentSets(dag, exposure = 'X', outcome = 'Y')`
- Rules:
    - Block all backdoor (confounding) paths
    - Do NOT condition on mediators (if you want total effect)
    - Do NOT condition on colliders
    - Do NOT condition on descendants of treatment
- Then fit your statistical model with only those covariates

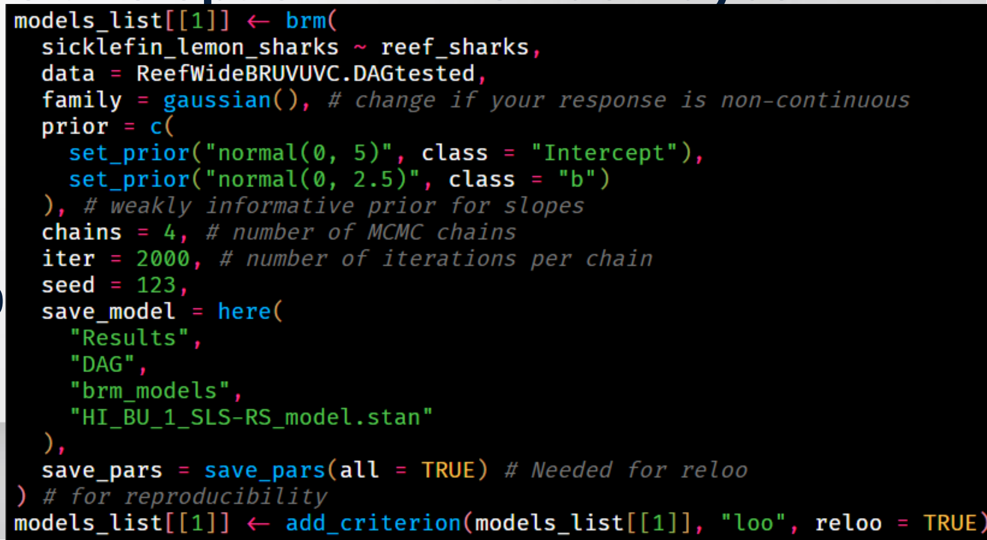
Structural Equation Modelling

What is an SEM?

- A set of simultaneous regression equations
- Each equation represents one arrow (or set of arrows) in a DAG
- SEMs quantify the STRENGTH of relationships encoded in a DAG
- Key distinction (Kunicki et al. 2023):
  - DAG = qualitative causal structure (what causes what)
  - SEM = quantitative estimation (how strong are the effects)
- SEM approaches in ecology:
  - Classical SEM (covariance-based): lavaan
  - Piecewise SEM: piecewiseSEM (Lefcheck 2016)
  - **Bayesian SEM: brms (Buerkner 2017)**
  - DAG-informed regression: individual models + adjustment sets

Bayesian SEMs with brms

- brms (Bayesian Regression Models using Stan; Buerkner 2017)
- For each relationship in your DAG:
- 1. Identify the adjustment set (from dagitty)
- 2. Fit: response ~ predictor + adjustment_variables (dagitty tells you)
- 3. Use appropriate family (Gaussian, Student-t, etc.)
- 4. Set informative priors where possible
- 5. Check convergence (Rhat, ESS, trace plots): Model summary
- 6. Evaluate with Bayesian LOO R-squared (Gelman et al. 2019). Compare different DAGs, compare different adjustment sets for the same DAG. **SEM packages have DAG comparison functions. Phylopath has this option. BRMS has Bayes R2. brms::loo_R2(model)**
- Advantages for ecology:
- Handles small sample sizes well
- Full uncertainty quantification via post.distrib
- Flexible: non-linear eff, splines, random eff



```r
models_list[[1]] ← brm(
  sicklefin_lemon_sharks ~ reef_sharks,
  data = ReefWideBRUVUVC.DAGtested,
  family = gaussian(), # change if your response is non-continuous
  prior = c(
    set_prior("normal(0, 5)", class = "Intercept"),
    set_prior("normal(0, 2.5)", class = "b")
  ), # weakly informative prior for slopes
  chains = 4, # number of MCMC chains
  iter = 2000, # number of iterations per chain
  seed = 123,
  save_model = here(
    "Results",
    "DAG",
    "brm_models",
    "HI_BU_1_SLS-RS_model.stan"
  ),
  save_pars = save_pars(all = TRUE) # Needed for reloo
) # for reproducibility
models_list[[1]] ← add_criterion(models_list[[1]], "loo", reloo = TRUE
```

Workshop Exercise 3: Fit a Bayesian SEM

- Using brms for one relationship from your DAG
- library(brms)
- Example: Effect of sharks on herbivores, controlling for temperature
- model <- brm(
-   Herbivores ~ s(Sharks, k=3) + Temperature,
-   data = reef_data,
-   family = student(),
-   prior = c(
-     prior(normal(0, 5), class = 'Intercept'),
-     prior(exponential(1), class = 'sds')
-   ), chains = 4, iter = 2000)
- Evaluate: Bayesian LOO R-squared
- loo_R2(model)

Top-Down vs Bottom-Up

The Problem: Directionality in Food Webs

- A fundamental question in ecology:
- Are ecosystems controlled from the top down (predators)?
- Or from the bottom up (primary productivity)?
- For coral reefs: Do sharks control reef health, or does reef health support sharks?
- Standard DAGs assume CONSISTENT directionality
- But food webs have arrows going BOTH ways
- Top-down: Sharks -> Prey teleosts -> Herbivores -> Coral
- Bottom-up: Coral -> Herbivores -> Prey teleosts -> Sharks
- This was previously not testable within the DAG/SEM framework
- Pearl's framework requires acyclicity -- no feedback loops

Opposing DAGs as Null Models

- Solution: Create TWO separate DAGs for the same system
- DAG 1 (Top-Down): All trophic arrows point downward
- Sharks -> Piscivores -> Prey teleosts -> Herbivores -> Coral
- DAG 2 (Bottom-Up): All trophic arrows point upward
- Coral -> Herbivores -> Prey teleosts -> Piscivores -> Sharks
- Both DAGs share the same non-trophic connections
- (environmental drivers, spatial factors, etc.)
- Distinct models from dagitty though: different dependencies.
- Neither DAG is expected to be 'correct' -- they are null models
- Compare RELATIVE SUPPORT for top-down vs bottom-up
- using Bayesian LOO R-squared for each relationship
- This reveals which direction better explains the data

Case Study: French Polynesian Coral Reefs

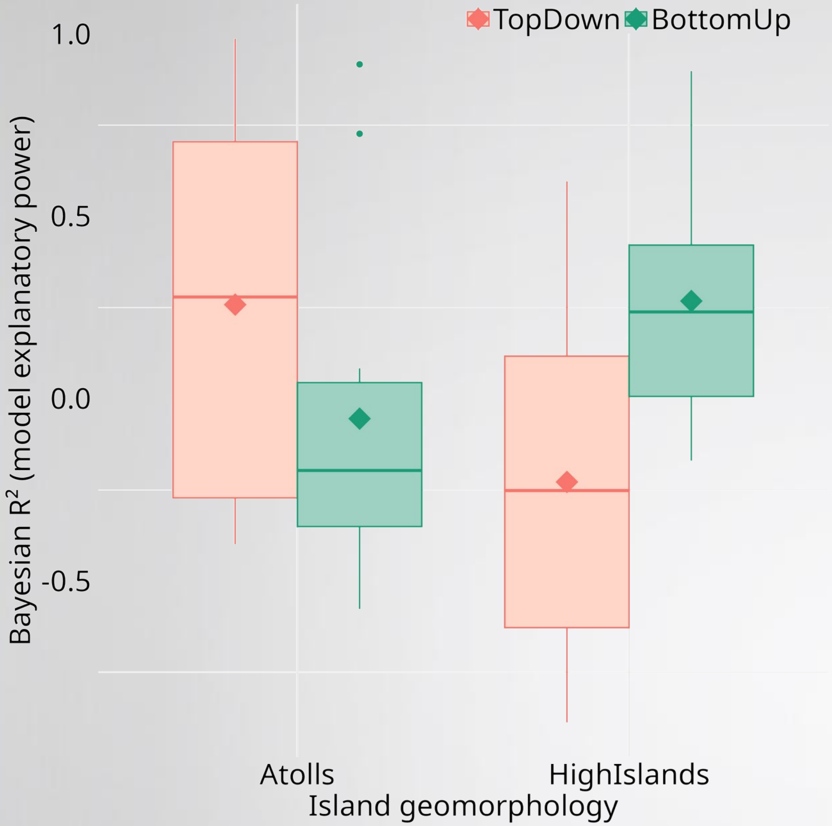
- 24 reefs across 12 islands/atolls in the world's largest shark sanctuary
- Shark sanctuary is a dubiously meaningful term but shark abundances here are some of the highest in the world.
- BRUV + UVC sampling of sharks, teleosts, and benthic cover
- Natural experiment: atolls vs high islands
- Atolls: higher chlorophyll-a, more sharks, more productive
- High islands: lower productivity, fewer mesopredators
- Exploitation Ecosystems Hypothesis (EEH; Oksanen et al. 1981):
- Top-down control should be STRONGER in more productive systems
- Prediction: Atolls = top-down dominant, High islands = bottom-up dominant
- 11 trophic relationships tested in each direction (44 models total)

Our DAG

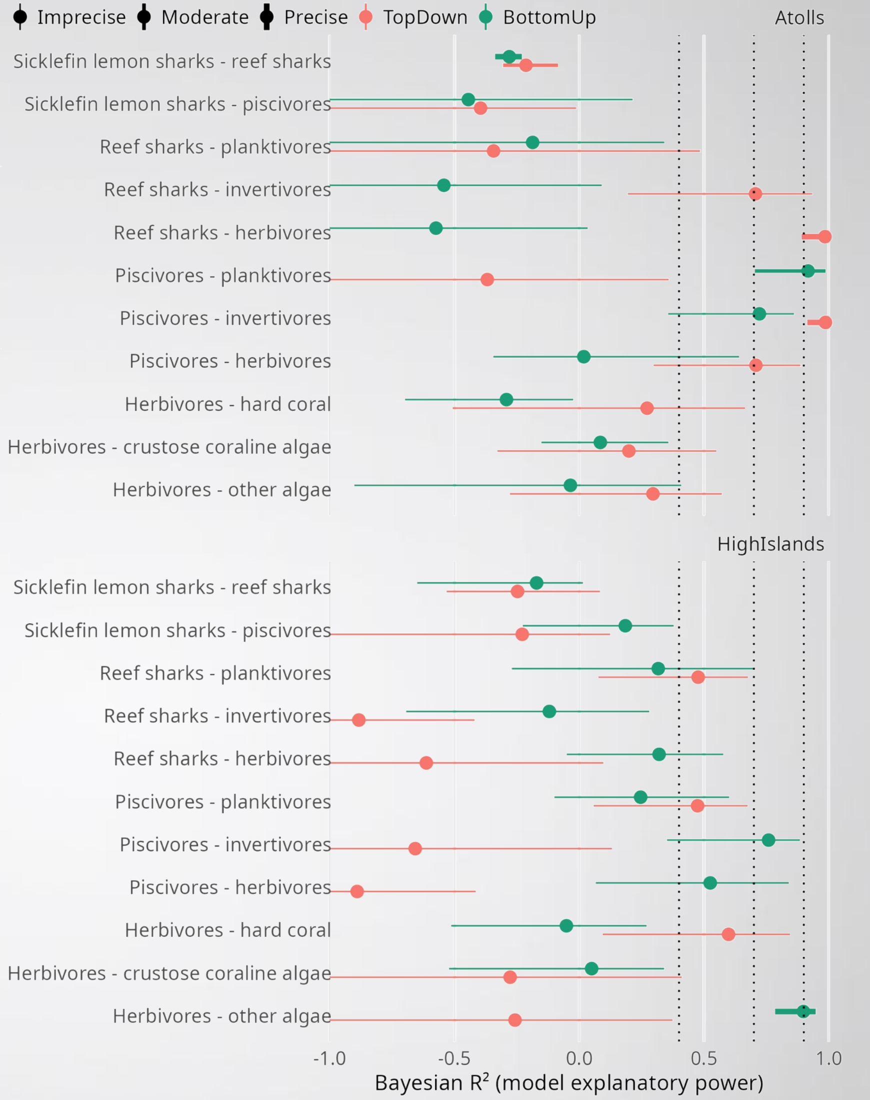
- Full DAG includes ~30 nodes:
- Measured (blue): shark species, teleost groups, benthic cover, environment
- Unmeasured (grey): wave exposure, recruitment, historical disturbance
- 11 key trophic relationships tested for top-down vs bottom-up:
- Large mesopredators <-> Reef sharks
- Large mesopredators <-> Piscivores
- Reef sharks <-> Planktivores, Invertivores, Herbivores
- Piscivores <-> Planktivores, Invertivores, Herbivores
- Herbivores <-> Hard coral, CCA, Other algae
- Each relationship fitted with brms using cubic splines
- Backdoor adjustment sets from dagitty

Results: Top-Down vs Bottom-Up

- ATOLLS:
- Mean TD R-sq: 0.26 (weak+)
- Mean BU R-sq: -0.05 (none)
- TD precision better
- Strong TD connections:
- - Piscivores on invertivores (0.99)
- - Sharks on herbivores (0.98)
- - Piscivores on herbivores (0.71)
- KEY: Mesopredators control prey teleosts top-down



TopDown
BottomUp
1.0
0.5
0.0
-0.5
Bayesian R^2 (model explanatory power)
Atolls
HighIslands
Island geomorphology



Imprecise
Moderate
Precise
TopDown
BottomUp
Atolls
Sicklefin lemon sharks - reef sharks
Sicklefin lemon sharks - piscivores
Reef sharks - planktivores
Reef sharks - invertivores
Reef sharks - herbivores
Piscivores - planktivores
Piscivores - invertivores
Piscivores - herbivores
Herbivores - hard coral
Herbivores - crustose coraline algae
Herbivores - other algae
HighIslands
Sicklefin lemon sharks - reef sharks
Sicklefin lemon sharks - piscivores
Reef sharks - planktivores
Reef sharks - invertivores
Reef sharks - herbivores
Piscivores - planktivores
Piscivores - invertivores
Piscivores - herbivores
Herbivores - hard coral
Herbivores - crustose coraline algae
Herbivores - other algae
-1.0
-0.5
0.0
0.5
1.0
Bayesian R^2 (model explanatory power)

Part 6: Tools, Tips & Resources

R Packages for DAGs & SEMs1

- DAG Construction & Testing:
- dagitty -- Draw, test, and analyze DAGs (Textor et al. 2016)
- ggdag -- ggplot2-based DAG visualisation
- daggity.net -- Interactive web tool for DAG building
- SEM Fitting:
- **brms -- Bayesian regression models via Stan (Buerkner 2017)**
- piecewiseSEM -- Piecewise structural equation modelling
- lavaan -- Classical covariance-based SEM
- blavaan -- Bayesian lavaan
- phylopath -- Phylogenetic path analysis (van der Bijl 2018)
- Accounts for shared evolutionary history across species
- Model Evaluation:
- loo -- Leave-one-out cross-validation (Vehtari et al. 2017)
- bayesplot -- Posterior diagnostic plots

dagitty: Your Most Important Tool

- dagitty.net -- Free web-based tool for DAG analysis
- What it does:
- Draw DAGs interactively
- Automatically derive adjustment sets (backdoor criterion)
- List all implied conditional independencies
- Test DAG-data consistency
- Identify instrumental variables
- Find minimal sufficient adjustment sets
- R interface: library(dagitty)
- dag <- dagitty('dag{ X -> Y <- Z; Z -> X }')
- adjustmentSets(dag, exposure='X', outcome='Y')
- impliedConditionalIndependencies(dag)
- localTests(dag, data=mydata)

Common Pitfalls & How to Avoid Them

- 1. Conditioning on a collider

  Tip: Draw the DAG BEFORE running any regressions
- 2. Omitting important confounders

  Tip: Include ALL common causes, even unmeasured ones
- 3. Controlling for mediators (blocking your own causal path)

  Tip: Be clear whether you want total or direct effects
- 4. Treating the DAG as derived from data

  Tip: DAGs encode prior knowledge; data tests consistency
- 5. Ignoring DAG-data consistency failures

  Tip: Failed independencies mean your DAG needs revision
- 6. Over-reliance on p-values in adjustment

  Tip: Control based on DAG structure, not statistical significance

Tips from Practitioners

- From Arif & MacNeil (2023):
- Start simple, add complexity only as needed
- Document your DAG assumptions explicitly
- Multiple adjustment sets may exist -- check robustness
- From McElreath (2020):
- The data alone can never tell you which DAG is right
- Scientific knowledge eliminates impossible DAGs
- Markov equivalence: multiple DAGs can imply the same data
- From Aitchison et al. (shark CFAR morphology):
- Argue about the DAG: present to colleagues, invite criticism
- Discussion reveals unnecessary complexity (8 vars reduced to 5)
- phylopath handles phylogenetic non-independence across species
- Show results as posterior density plots with credible intervals
- From our experience (Dedman et al., in prep):
- Complex ecological DAGs need iterative refinement
- Bayesian approaches handle small marine datasets well
- Splines can capture non-linear ecological relationships
- Additional tips from colleagues

Uncertain Underpinnings: Circular Causal Confusion

- Correlative papers say component relationships are causal,
  then we have dagitty run its do() calculus, eliminate backdoor paths & circularities to create a final parsimonious DAG
  with conditional independences to check. If they pass:
  you can run [brms] models to generate results that are causal.
  But:
- 1. Are they causal (partially) contingent on proven causality of component relationships?
- 2. If those proofs come from correlational studies are the final results still causal?
- 3. We say it's causal -> do() calculus due diligence + independences -> causal results?
- Also: Can we ascribe strength of confidence of individual relationship based on the quality of the original study? Weighting?
- Those relationship data are each normalised, scaled, & centred no? If all 0-1, what scope to weight?
  And what implication for downstream formulae?

Part 7: Workshop & Further Reading

Essential Reading

Introductory:

- Rohrer (2018) Thinking clearly about correlations & causation
- Arif & MacNeil (2022) Utilizing causal diagrams...
- Williams et al. (2018) DAGs: a tool for causal studies in paediatrics

Core methodology:

- Arif & MacNeil (2023) Applying the structural causal model framework...
- McElreath (2020) Statistical Rethinking, Chs 5-6
- Pearl (2009) Causality
- Barrett, McGowan & Gerke (2025) Causal Inference in R (r-causal.org)

Ecological applications:

- Schoolmaster et al. (2020) Graphical causal models in ecology
- Laubach et al. (2021) Biology guides model selection & causal inference
- Kimmel et al. (2021) Causal assumptions in ecological experiments
- van der Bijl (2018) phylopath: phylogenetic path analysis in R

Summary & Take-Home Messages

- 1. Draw a DAG before you run a regression

  Your causal assumptions should be explicit and testable
- 2. Use the backdoor criterion to determine what to control for

  Not everything should go in the model!
- 3. Test your DAG against data (dagitty conditional independencies)
- 4. Bayesian SEMs work well for small ecological datasets
- 5. Opposing DAGs can test directional hypotheses (e.g., TD vs BU)

  A new approach for food web ecology
- 6. DAGs are the most important statistical tool you're not using

  -- adapted from McElreath

References

- Arif, S., & MacNeil, M. A. (2022). Predictive models aren’t for causal inference. Ecology Letters, 25(8), 1741–1745.
- Barrett, M., et al. (2025). ggdag: Analyze and create elegant directed acyclic graphs. R package. CRAN.
- Grace, J. B., et al. (2012). Guidelines for a graph-theoretic implementation of structural equation modeling. Ecosphere, 3(8), art73.
- Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. Epidemiology, 10(1), 37–48.
- Laubach, Z. M., et al. (2021). A biologist’s guide to model selection and causal inference. Proceedings of the Royal Society B, 288(1943), 20202815.
- Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modelling in R. Methods in Ecology and Evolution, 7(5), 573–579.
- Pearl, J. (2009). Causality: Models, Reasoning, and Inference (2nd ed.). Cambridge University Press.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. Journal of Statistical Software, 48(2), 1–36.
- Schoolmaster, D. R., et al. (2020). A graphical causal model for resolving species identity effects and biodiversity–ecosystem function correlations. Ecology, 101(8), e03070.
- Shipley, B. (2016). Cause and Correlation in Biology (2nd ed.). Cambridge University Press.
- Textor, J., et al. (2016). Robust causal inference using directed acyclic graphs: the R package dagitty. International Journal of Epidemiology, 45(6), 1887–1894.
- van der Bijl, W. (2018). phylopath: Easy phylogenetic path analysis in R. PeerJ, 6, e4718.
- Wright, S. (1921). Correlation and causation. Journal of Agricultural Research, 20(7), 557–585.

https://www.github.com/SimonDedman/statscourse
Slides & code available on/before 1st April

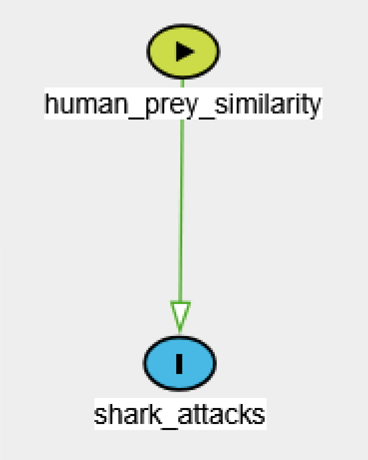
Thank You

Simon Dedman, simondedman@gmail.com, simondedman.com
Florida International University

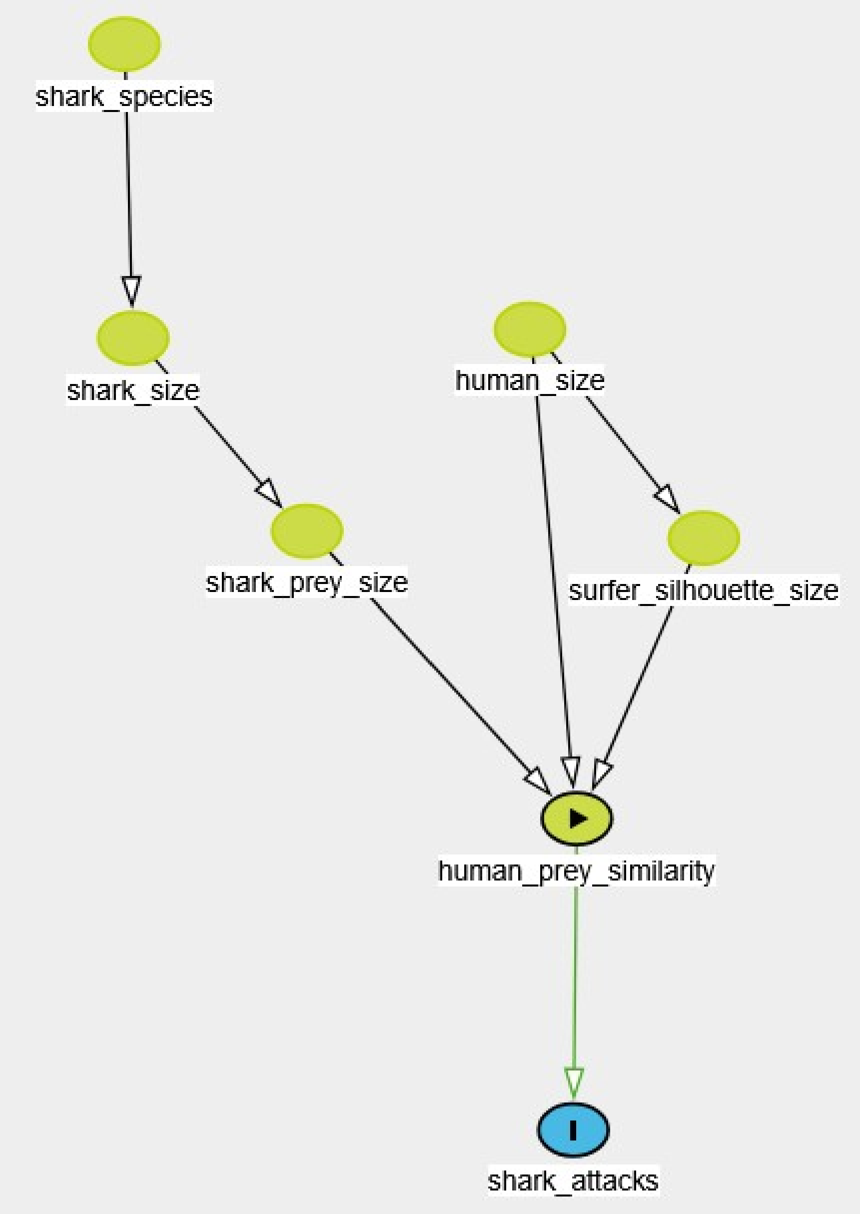
DAG Workshop/Project

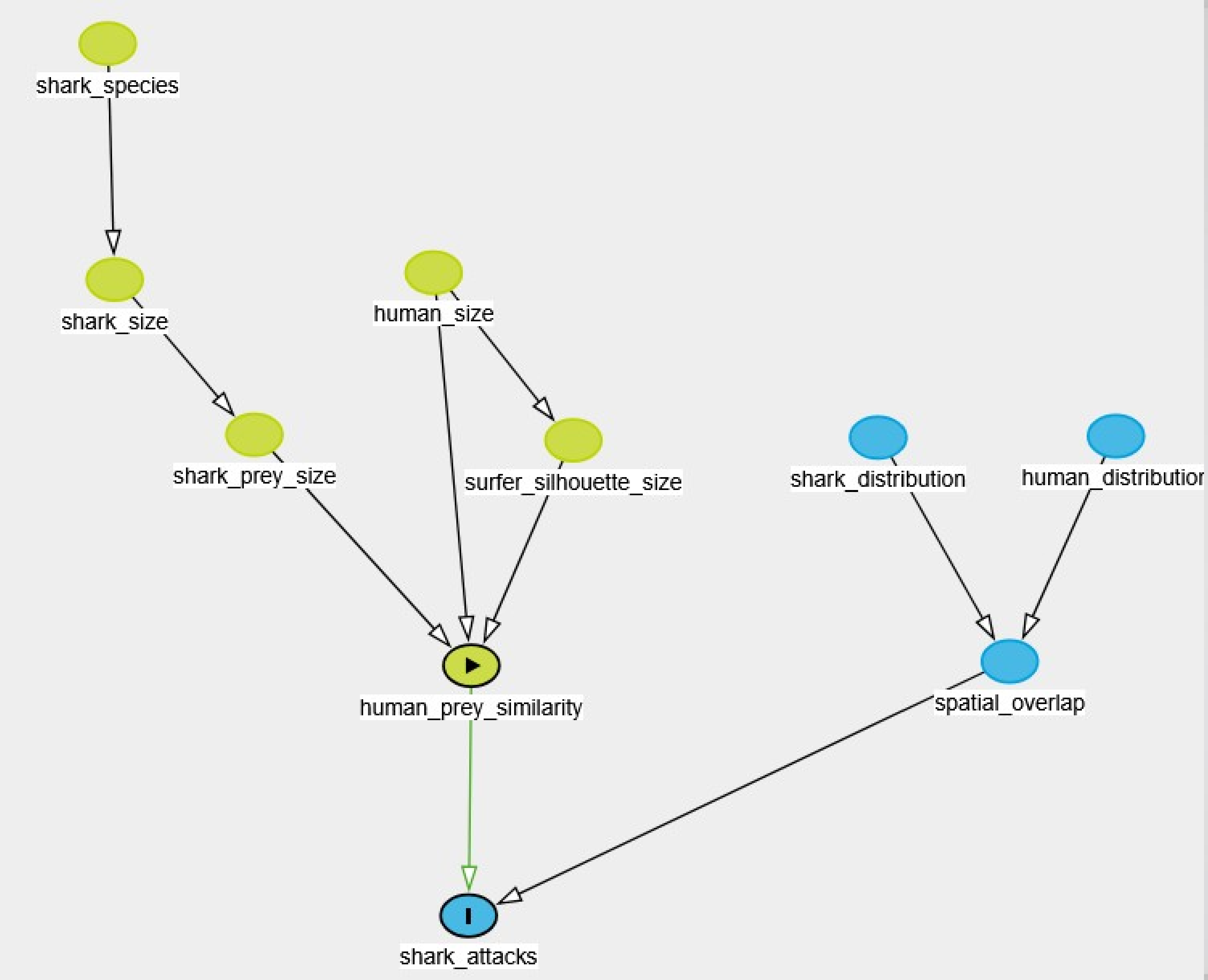
- Read Eric Clua’s humour/ opinion/ mistaken identity rebuttal paper.
- Make a DAG of ALL causes of shark bites.
  - Where does mistaken identity fit?
  - What does this tell us about shark bite patterns for shark species, location, time, etc?
- Team exercise to develop this, model it, write it up, publish.

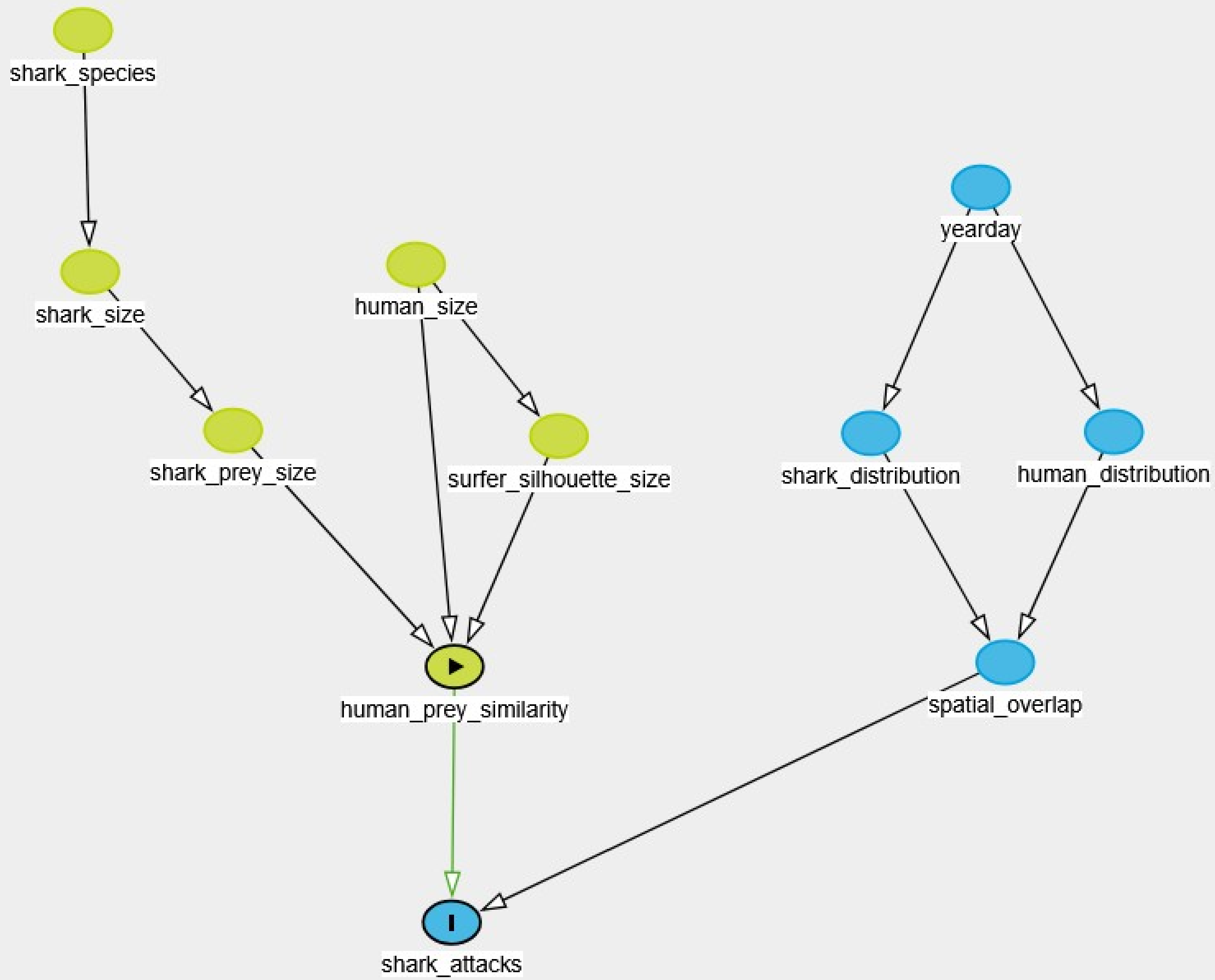
Mistaken Identity Hypothesis: human / shark-prey similarity (e.g. seal/surfer)

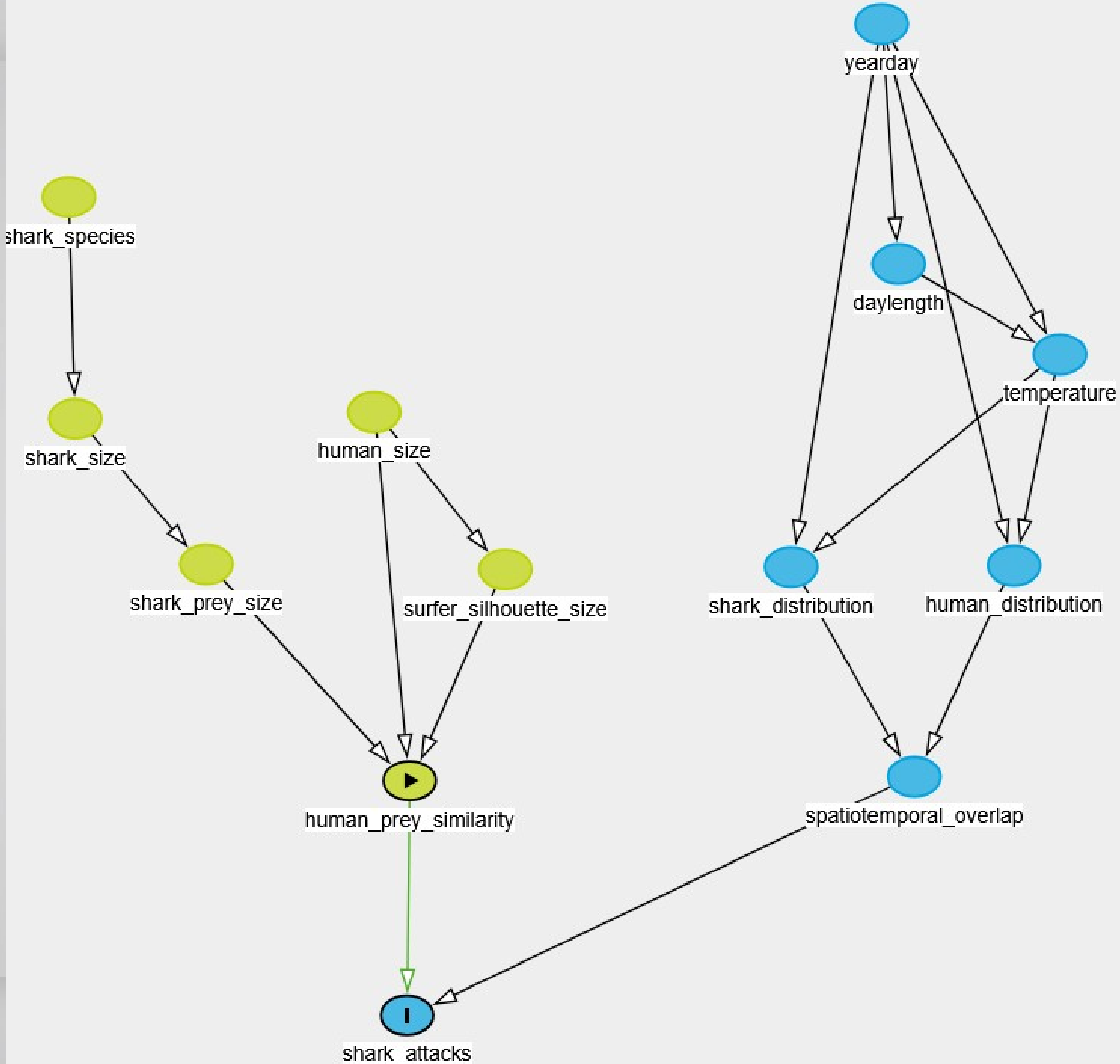


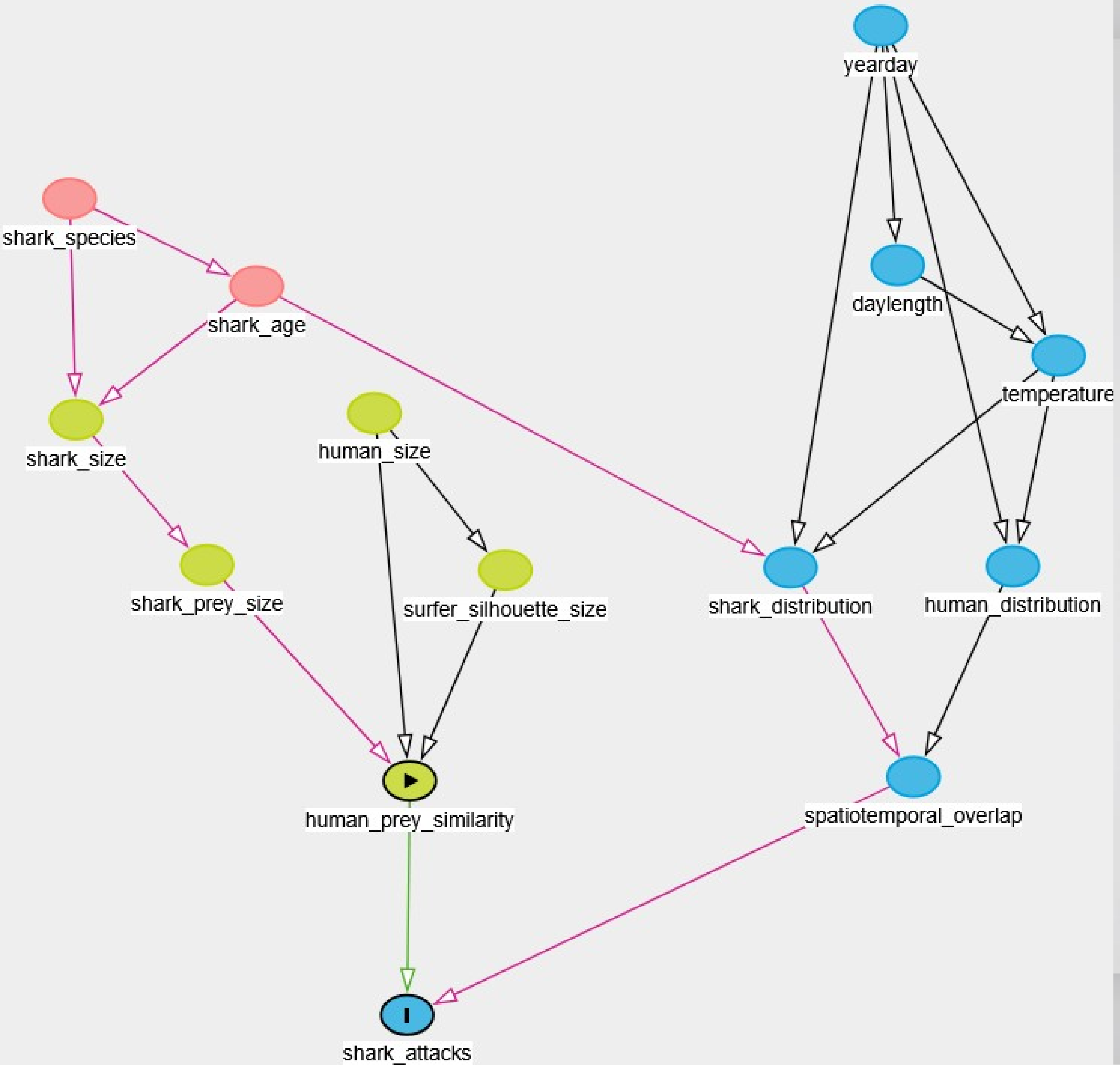
Mistaken Identity Hypothesis: human / shark-prey similarity (e.g. seal/surfer)

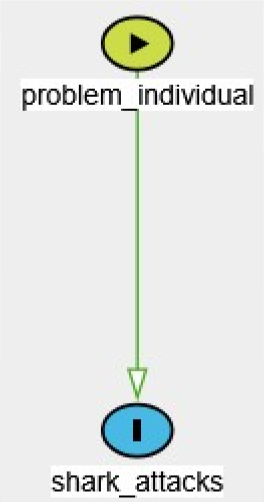


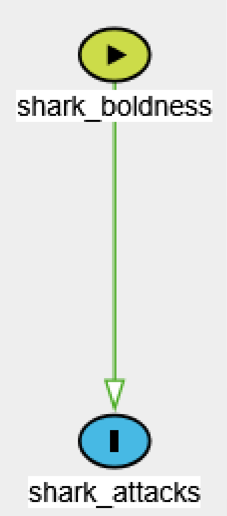

shark_species
shark_size
human_size
shark_prey_size
surfer_silhouette_size
shark_distribution
human_distributio
human_prey_similarity
spatial_overlap
shark_attacks

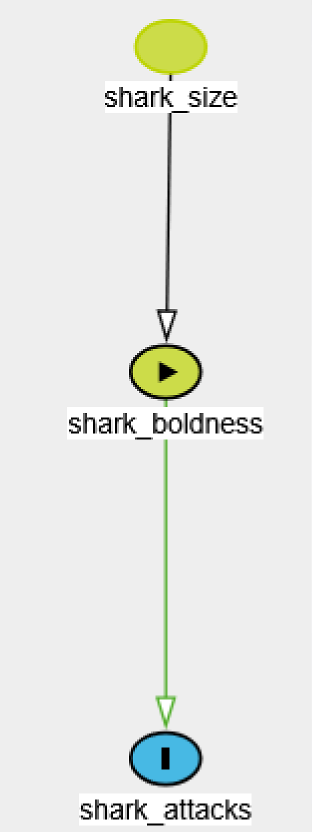

shark_species
shark_size
human_size
shark_prey_size
surfer_silhouette_size
human_prey_similarity
yearday
shark_distribution
human_distribution
spatial_overlap
shark_attacks

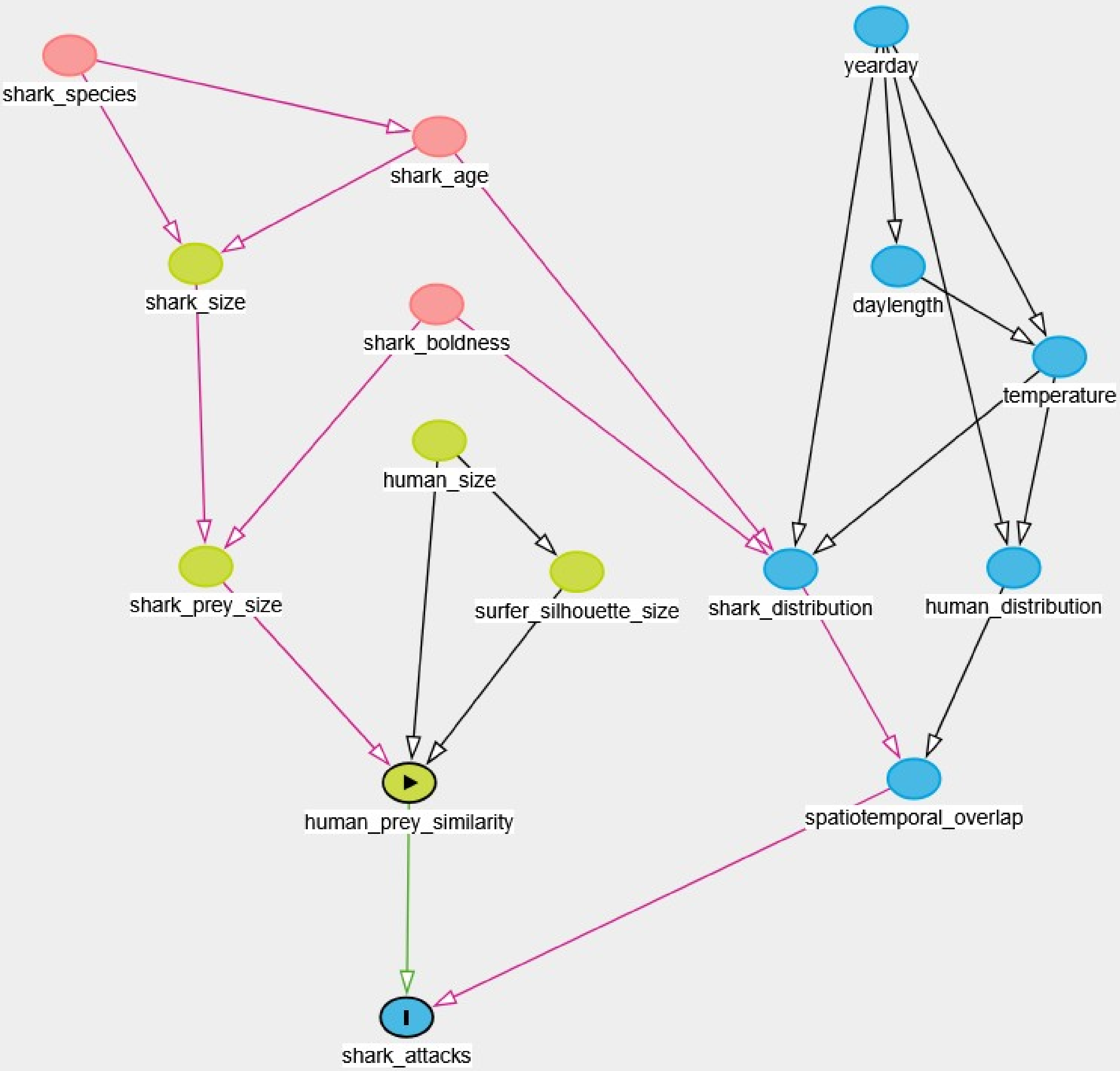

yearday
shark_species
daylength
temperature
shark_size
human_size
shark_prey_size
surfer_silhouette_size
shark_distribution
human_distribution
human_prey_similarity
spatiotemporal_overlap
shark_attacks


yearday
shark_species
daylength
shark_age
temperature
shark_size
human_size
shark_prey_size
surfer_silhouette_size
shark_distribution
human_distribution
human_prey_similarity
spatiotemporal_overlap
shark_attacks


problem_individual
shark_attacks


shark_boldness
shark_attacks


shark_size
shark_boldness
shark_attacks



shark_species

shark_age

shark_size

shark_boldness

human_size

shark_prey_size

surfer_silhouette_size

human_prey_similarity

shark_attacks

yearday

daylength

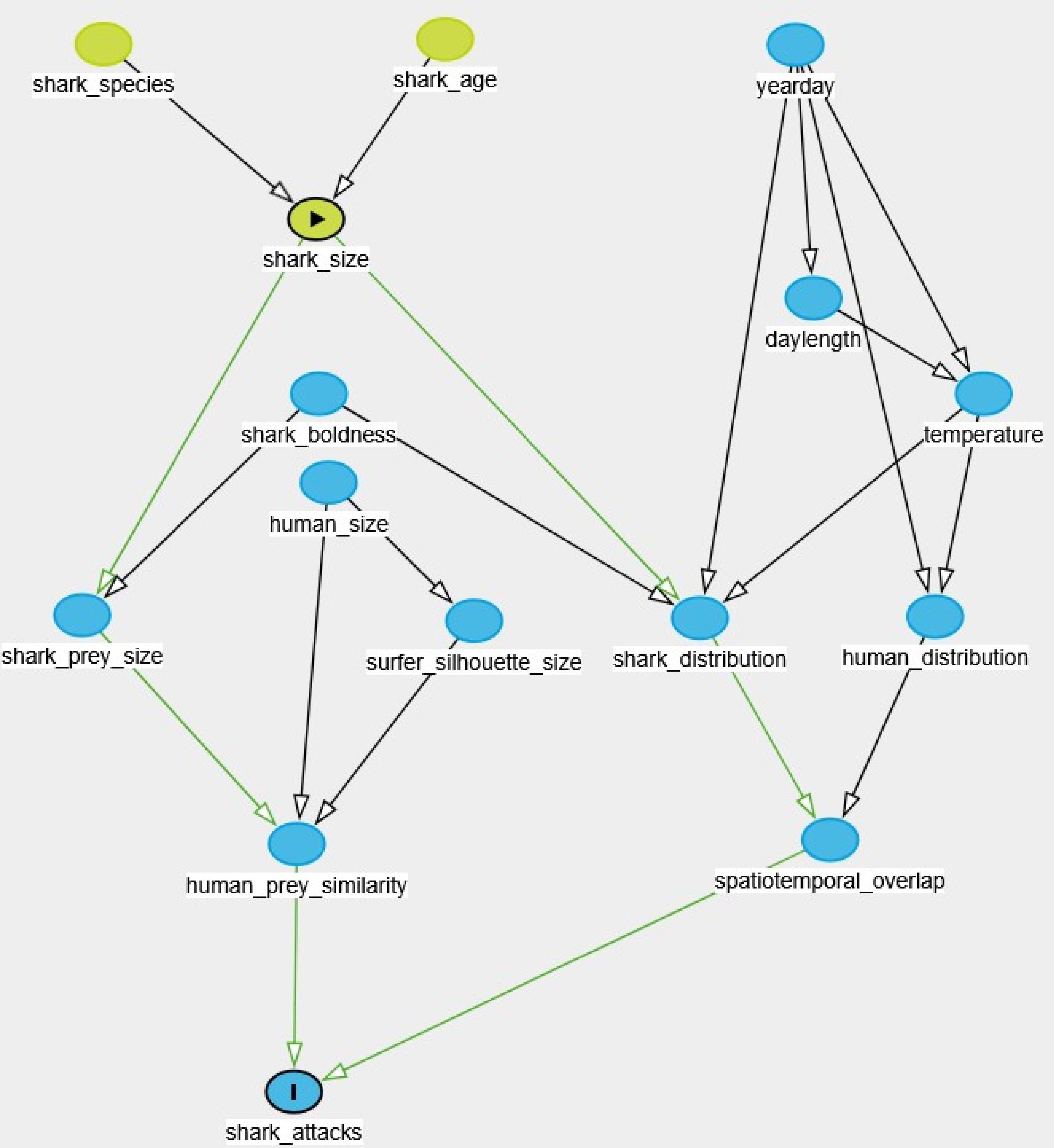
temperature

shark_distribution

human_distribution

spatiotemporal_overlap

FIU



Exposure: shark_size
Outcome: shark_attacks

No open biasing paths.

No adjustment is necessary to estimate the total effect of shark_size on shark_attacks.

Testable implications

The model implies the following conditional independences:

- shark_species ⊥ shark_prey_size | shark_size
- shark_species ⊥ human_size
- shark_species ⊥ surfer_silhouette_size
- shark_species ⊥ human_prey_similarity | shark_prey_size
- shark_species ⊥ human_prey_similarity | shark_size
- shark_species ⊥ shark_distribution | shark_size
- shark_species ⊥ human_distribution
- shark_species ⊥ shark_attacks | human_prey_similarity, spatiotemporal_overlap
- shark_species ⊥ shark_attacks | shark_prey_size, spatiotemporal_overlap
- shark_species ⊥ shark_attacks | human_distribution, human_prey_similarity, shark_distribution
- shark_species ⊥ shark_attacks | human_prey_similarity, shark_distribution, temperature, yearday
- shark_species ⊥ shark_attacks | human_distribution, shark_distribution, shark_prey_size