Causal Discovery and Causal-driven Modeling in Astrophysics

2nd Annual Cosmic Horizons Conference, July 16th 2026

Zehao Jin (金泽灏)
Center for Astronomy and Astrophysics
Fudan University, Shanghai
zehaojin@fudan.edu.cn / zj448@nyu.edu
https://zehaojin.github.io/

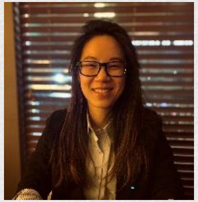
Main Collaborators

Astro side

Benjamin Davis
NYUAD →MSU

Tobias Buck
Heidelberg

Andrea Maccio
NYUAD



Yuxi(Lucy) Lu
OSU

ML4Astro

Yuan-sen Ting
OSU

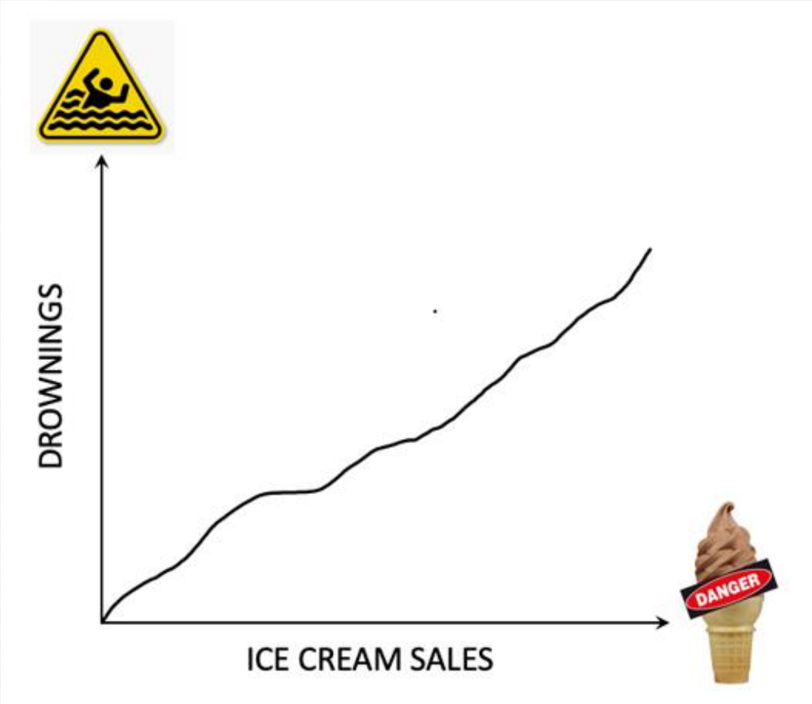
Mohamad Ali-Dib
NYUAD

ML side

Yujia Zheng
CMU→UIUC

Kun Zhang
CMU&MBZUAI

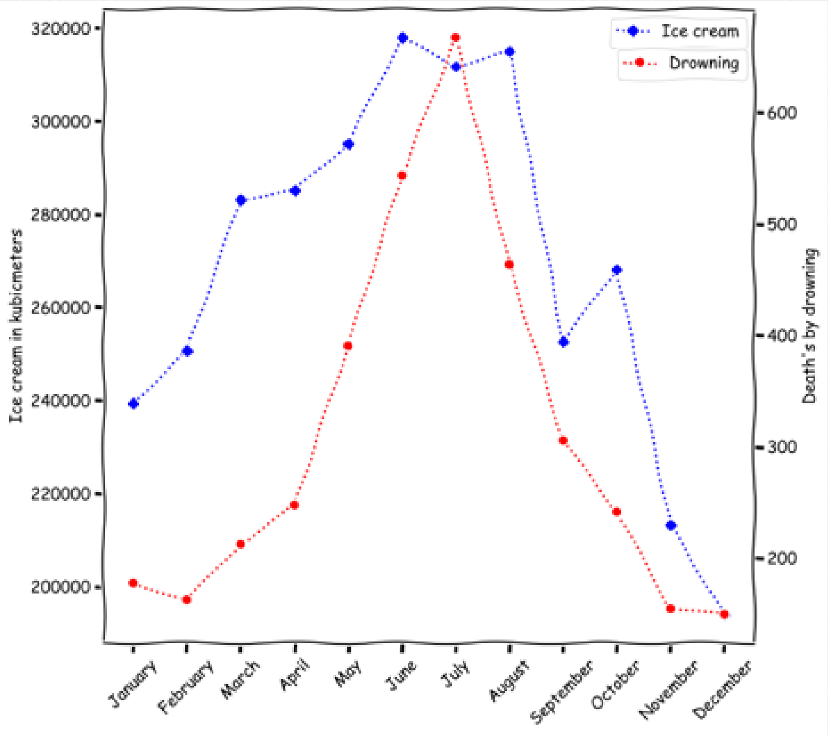
1 Correlation & Causation

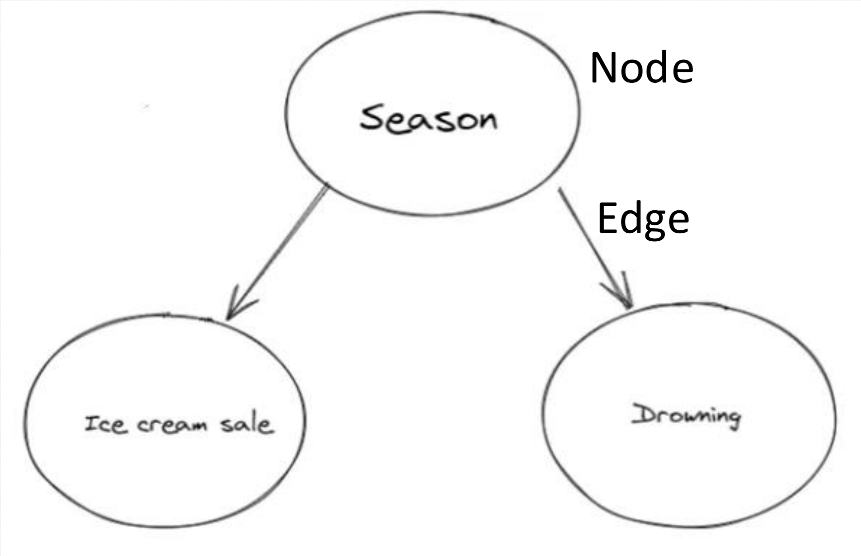


Does the correlation mean that ice cream can cause drowning?

Correlation≠Causation, but why is there such a correlation?

1 Correlation & Causation

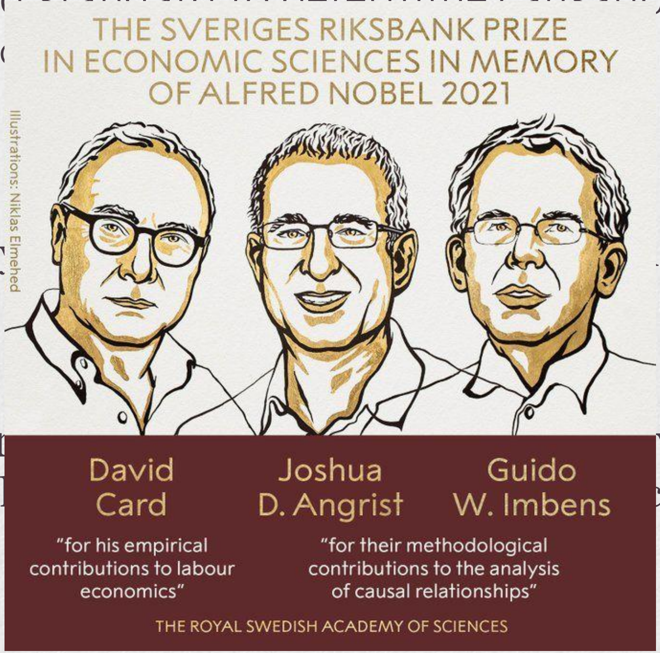




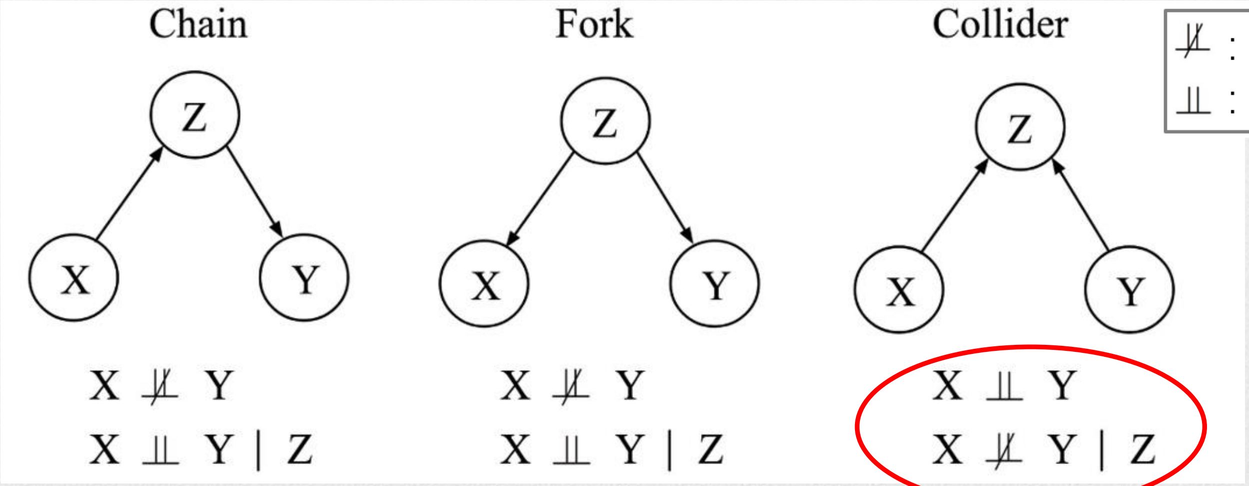
Directed Acyclic Graph (DAG)

1 Correlation & Causation

- Intervention is the gold standard to determine causation
  - randomized controlle
- However intervention
  - Especially in Astrophy tervene the universe
- Causal discovery: infer vational data
  - Already used in Geop e and Economics

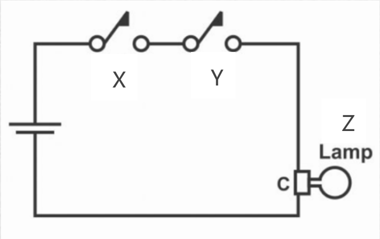


1 Correlation & Causation



$\not\perp\!\!\!\perp$: NOT independent, i.e. dependent
$\perp\!\!\!\perp$: independent

- **Correlation≠Causation**
- **Causation =(conditional) independence**
- **Conditional independence → Causation** (Causal discovery)



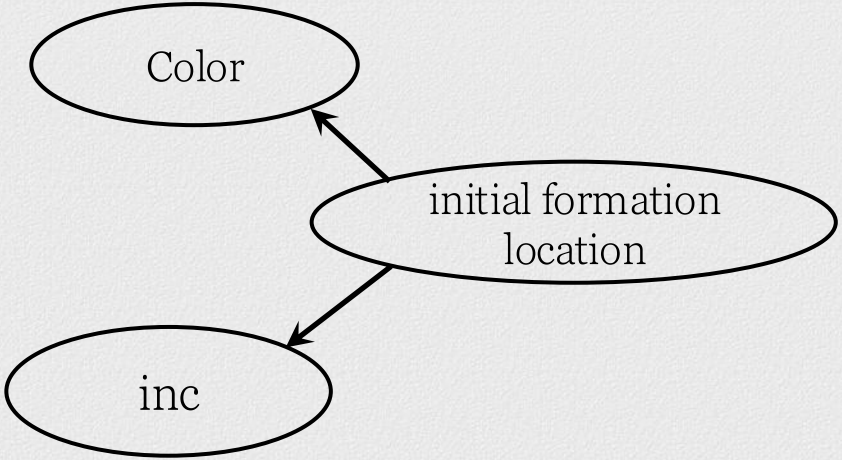
2 Planets: Origin of TNO colors

Causal evidence for the primordiality of colours in trans-Neptunian objects,

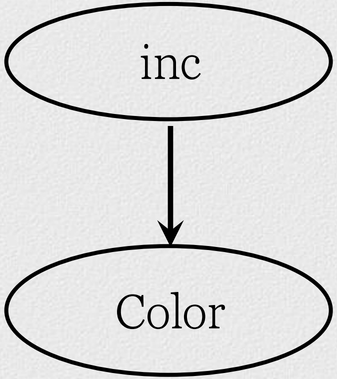
BL Davis* †, M Ali-Dib †, Y Zheng †, **Z Jin †**, K Zhang, AV Macciò, 2025, *MNRAS Letters*

Why Trans-Neptunian objects (TNOs)'s color and inclination are correlated?

The Primordial origin hypothesis



The collisional evolution hypothesis



The Fast Causal Inference (FCI) algorithm

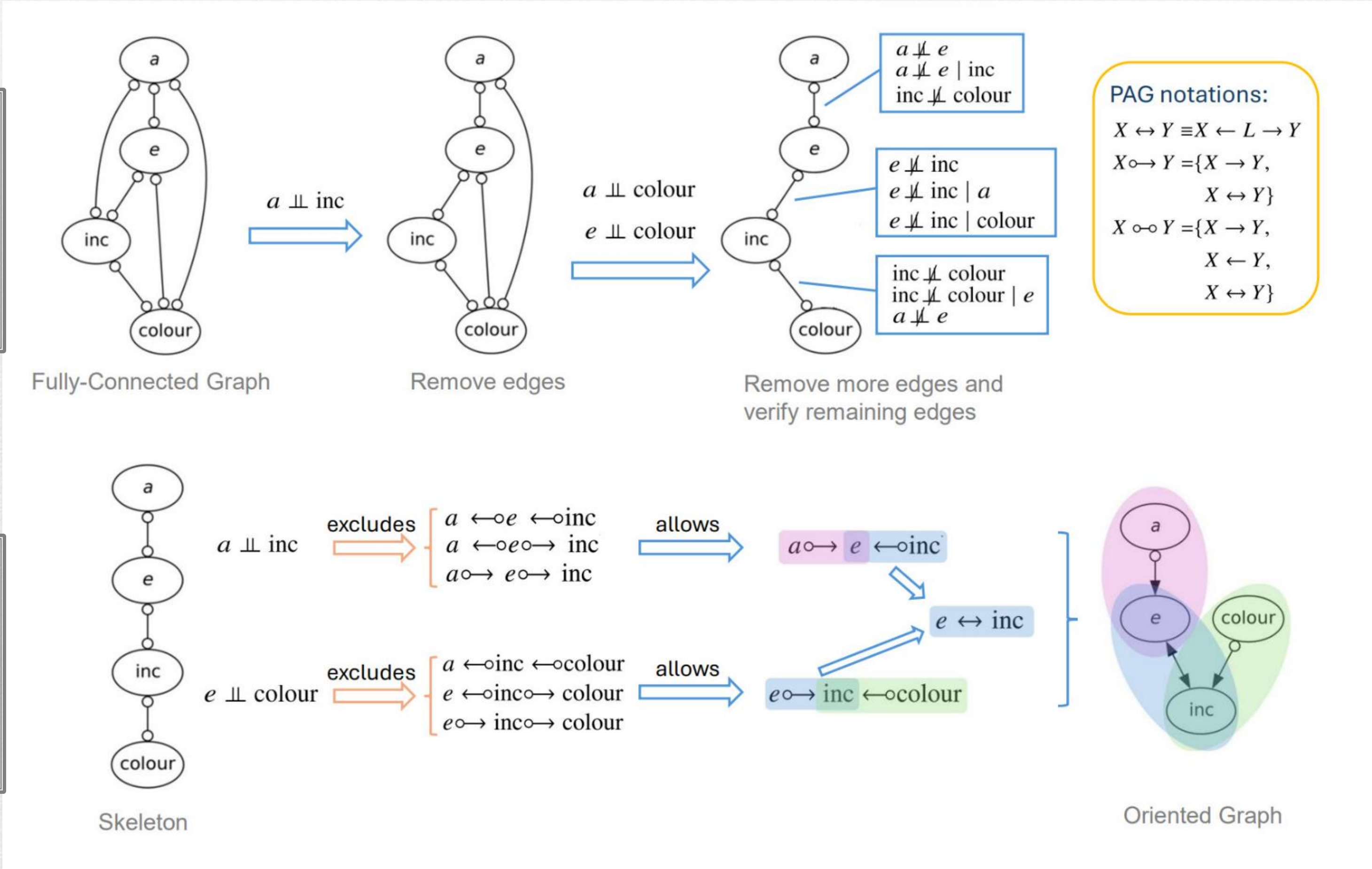
TNO data

Col-OSSOS & DES:

- a (semi-major axis)
- e (eccentricity)
- inc (inclination)
- color

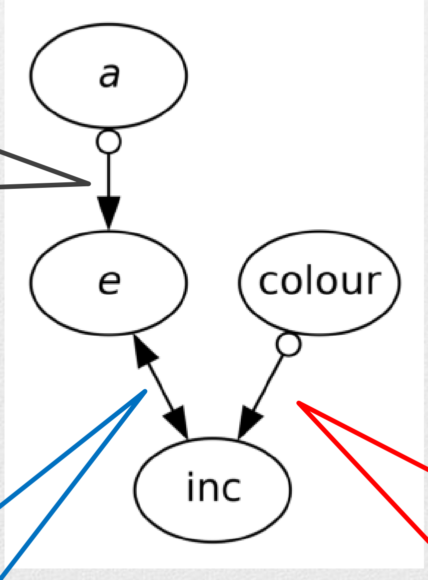
Considers possible unseen (latent) variables

through
Partial Ancestral Graph (PAG)



2 Planets: Origin of TNO colors

- a determines whether gets scattered by Neptune, and this scattering alters e
- a determines whether TNOs are inside an eccentricity-raising Mean motion resonances (MMRs)



PAG notations:

$X \leftrightarrow Y \equiv X \leftarrow L \rightarrow Y$

$X \circ\!\!\rightarrow Y = \{X \rightarrow Y,\ X \leftrightarrow Y\}$

$X \circ\!\!-\!\!\circ Y = \{X \rightarrow Y,\ X \leftarrow Y,\ X \leftrightarrow Y\}$

von Zeipel-Lidov-Kozai (vZLK) mechanism, perturbations from a third body – Neptune

If Neptune had not already been discovered in 1846, our result here would strongly suggest the presence of an unknown perturbing body. -- "Rediscovered" Neptune!

Consistent with <u>primordial origin hypothesis</u>
Rules out <u>collisional evolution hypothesis</u>

3 Stars: Galactic Archaeology

Causal Discovery of Latent Variables in Galactic Archaeology

Z Jin *†, Y Lu †, YS Ting, Y Zheng, T Buck, *ICML 2025 ML4Astro,* Journal version in prep.

- Galactic Archaeology:
  - We only observe the snapshot of the current Milky Way.
  - How to reconstruct the stellar migration histories in Milky Way?

- Observable Clues:
  - Chemical abundances
  - Kinodynamics

- Key Unobservable: Birth radius – the initial condition

3 Stars: Galactic Archaeology

Data

- Three "Milky Way" galaxies
  - 2 × NIHAO-UHD simulation of MW-mass galaxies
  - 1 × Real MW observation from APOKASC-3 sample

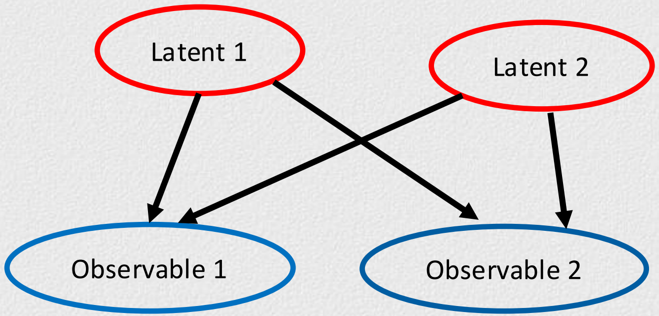
- Six observable properties of stars
  - [Fe/H] : metalicity
  - [Mg/H] : to trace evolution of alpha elements
  - Age
  - J_z : vertical action
  - J_ϕ : equivalent to z-component angular momentum L_z
  - e : eccentricity

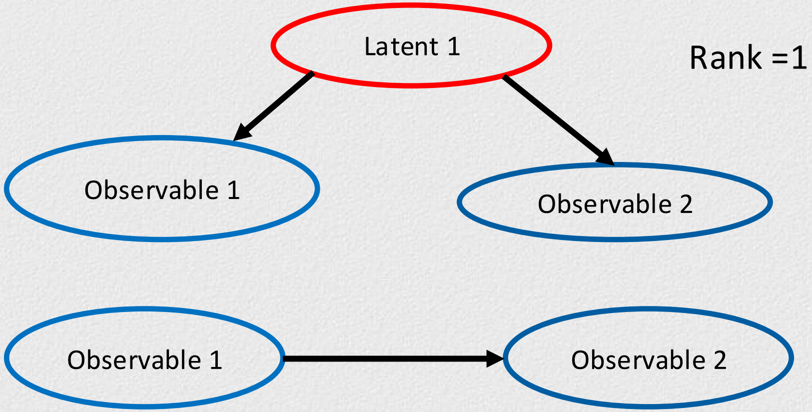
3 Stars：Galactic Archaeology

Method

- Rank-based Latent Causal Discovery (RLCD)
  - Assumes linear relationship between variables.
  - Rank tests can serve as conditional independence tests (~ test of DOF)
  - latent variables leave statistical footprints in rank deficiencies in the covariance matrix of observables

Rank =2

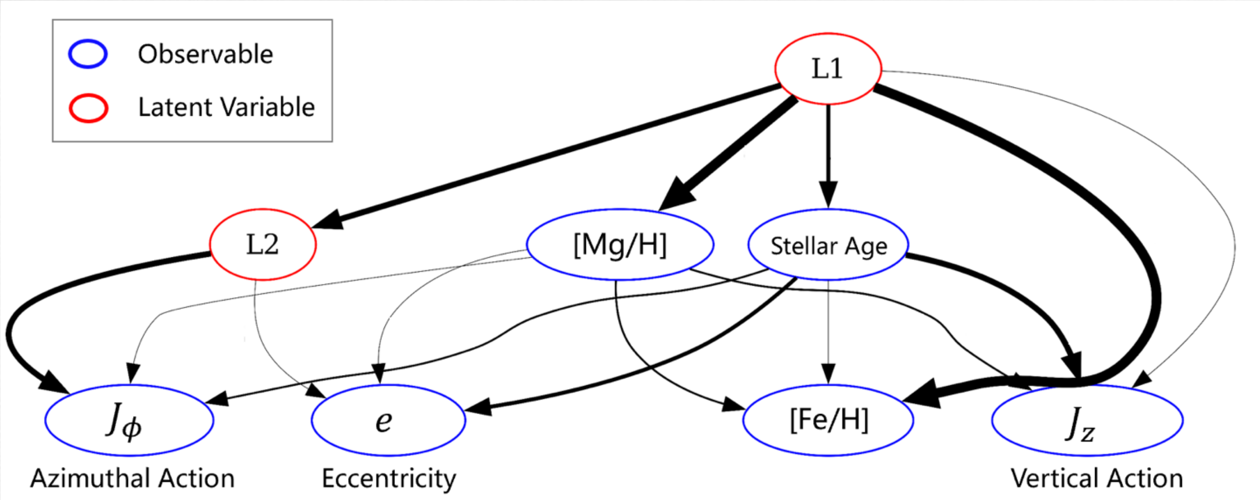




3 Stars: Galactic Archaeology

复旦大學

Result on first simulated galaxy

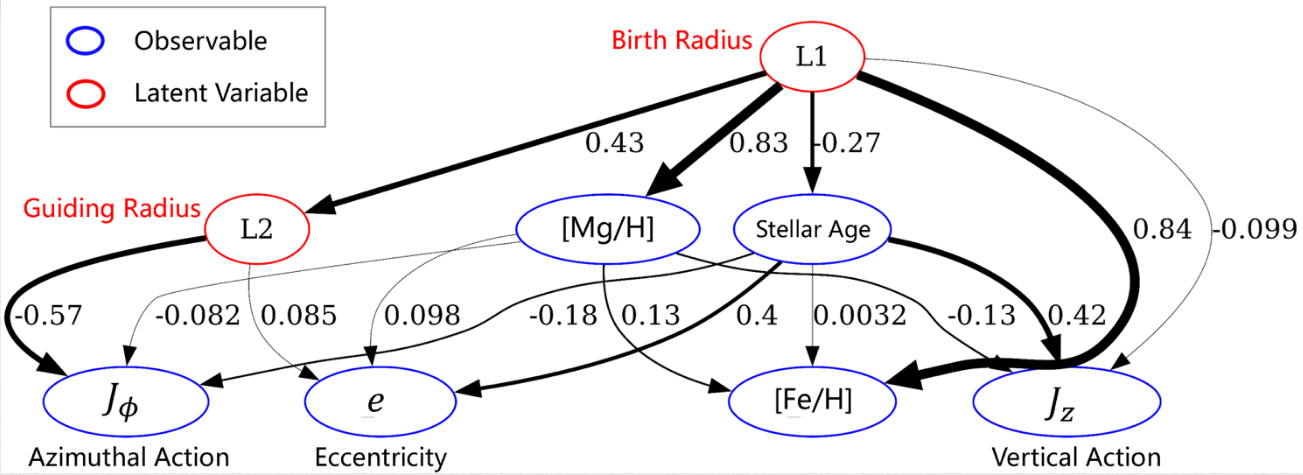


2 latent variables are found: L1 & L2

- L1 governs all other variables -- we hypothesize that L1 encodes info about stellar birth conditions (radial abundance, inside-out formation, vertical motions, gravitational potential). -- L1 = birth radius?
- L2 influences J_ϕ and e -- Stars orbits can heat up radially (increase in e) through gravitational scattering (“blurring”) -- L2 = guiding radius?

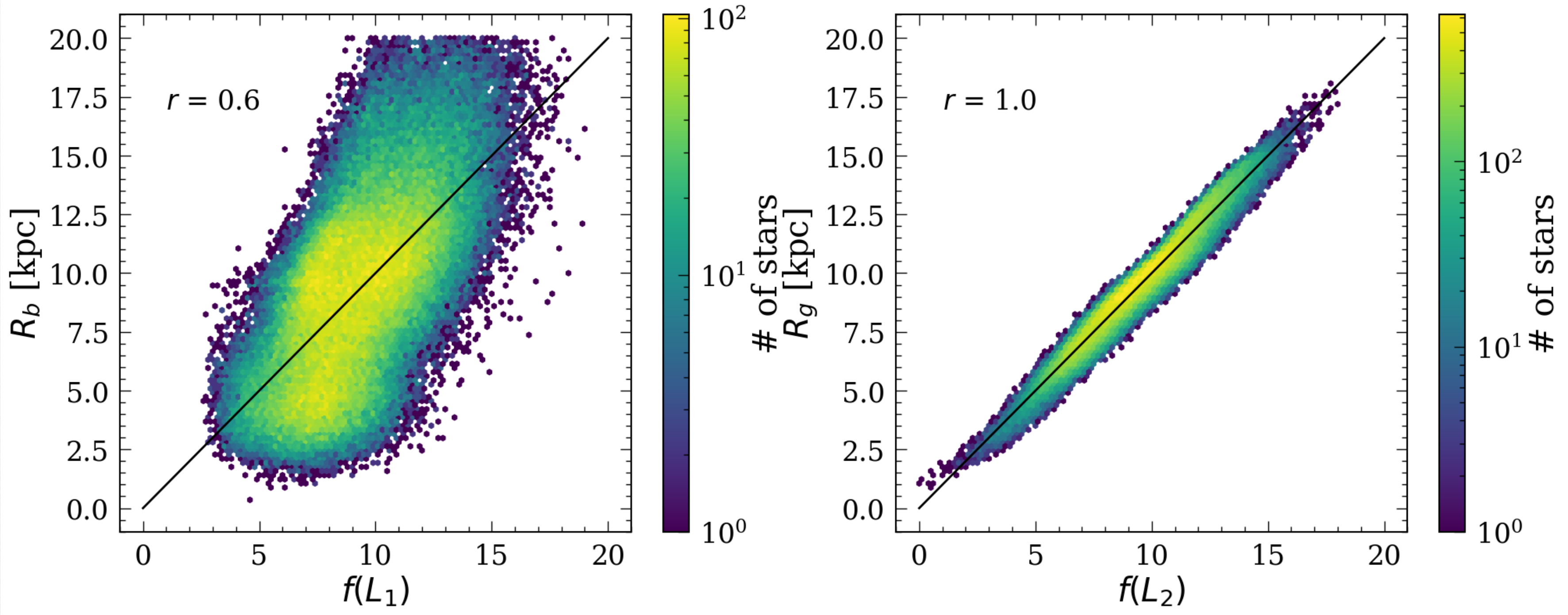
3 Stars: Galactic Archaeology

Result on first simulated galaxy



- To check our hypothesis, we calculate the coefficients on edges, and the value of L1 and L2 for each star
- Since in a simulated galaxy, we can compare L1 and L2 to the ground truth birth radius (R_b) and guiding radius (R_g)

3 Stars: Galactic Archaeology

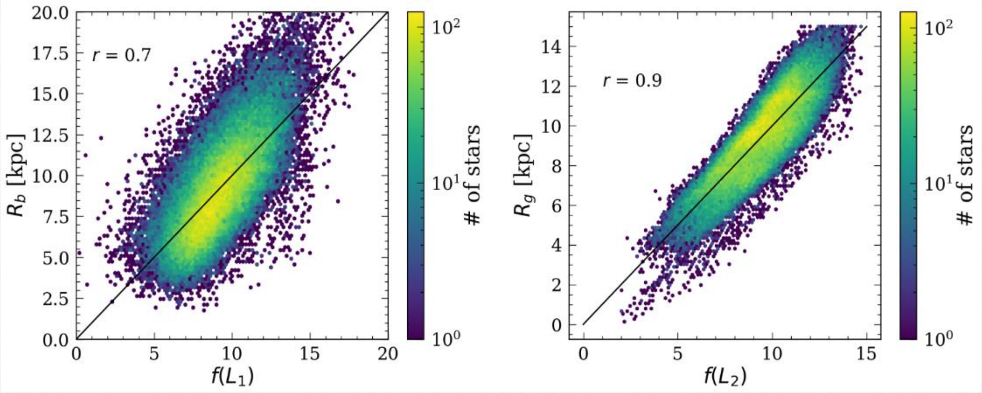


Extrapolation and Generalization

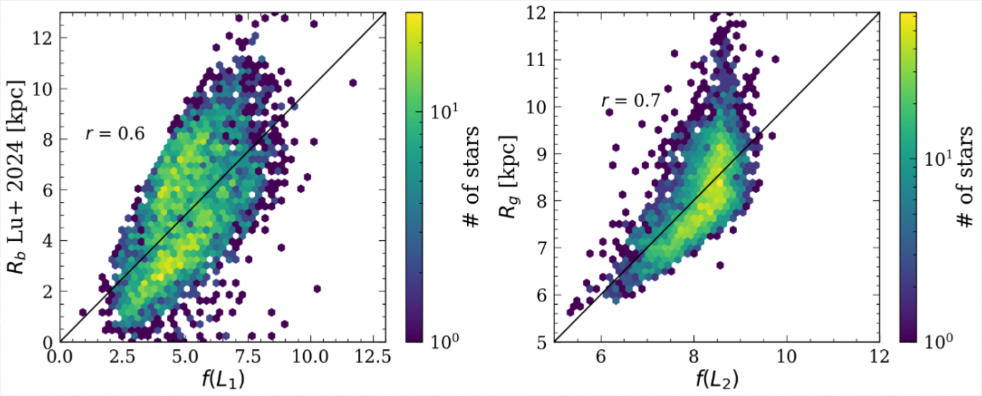
- If the model really learns the underlying physics, the same causal graph should generalize to the stellar dynamics of other galaxies, whether simulated or observationally inferred.
- The numerical values of the causal coefficients, may differ from galaxy to galaxy, as each has distinct global properties (total mass, size, etc)
- Extrapolate to another simulated galaxy and real MW observation
    - Keep the same causal model: the nodes, the number of edges, direction of edges are kept the same
    - The coefficients are re-estimated respectively

3 Stars: Galactic Archaeology

Causal model extrapolated to a new simulated galaxy



Causal model extrapolated to the real Milky Way



3 Stars: Galactic Archaeology

复旦大學

- Without any prior astro knowledge, the “AI scientist” draws a causal graph that consistent to our standing of stellar chemodynamics, with two physically meaningful latent nodes.
- Completely ignorant of birth radius, predicts the birth radius with a causal-driven model.
- The causal model demonstrates excellent generalizability across different simulations as well as real observation.

4 Conclusion

- Causal discovery in:
  - Planets: original of TNO colors
    -- *Distinguish different causation that shares similar correlation*
  - Stars: Galactic Archaeology
    -- *Causal discovery in the latent world, and highly generalizable causal model*

- Causal ML in Astronomy
  - Complementary to traditional approach
  - Causal ML x Deep learning
    - Inductive bias to NN architecture
    - Causal discovery to higher dimensional data:
      Causal representation learning

  - **Still a blue sea in Astro**

Thank you!

List of papers

Open Source code for every project

Intro to Causality hosted on my website

Zehao Jin
zehaojin@fudan.edu.cn
zj448@nyu.edu
July 16th 2026

1 Correlation & Causation

A side note: What exactly is causality?/How to understand X->Y ?

-- <u>Intervention</u> is the key

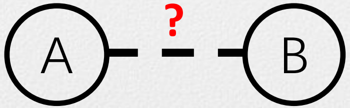
If one gets sick, his face will turn yellow

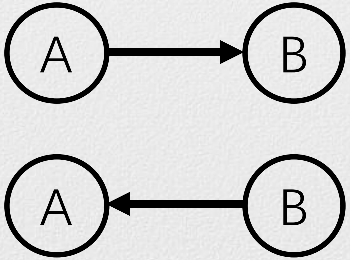
If we make sure his face is yellow, ...

Sick → Face yellow

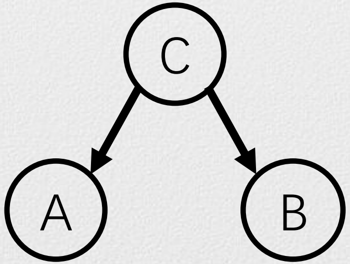
1 Correlation & Causation

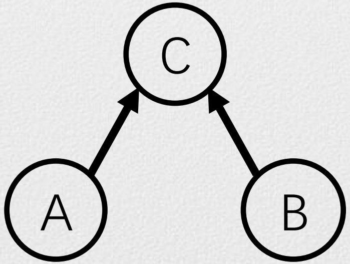
What is the causation behind a correlation?





"confounder"





How we know it is one causes another without having time-series data?

Age of the Universe: 1.38×10^{10} years
First human in Africa: 3×10^{5} years
-- Can't observe a galaxy to evolve during a grad student's lifetime

- Does A still cause B although there is a confounder?
- What is the confounder?
- What if we can't observe the confounder?

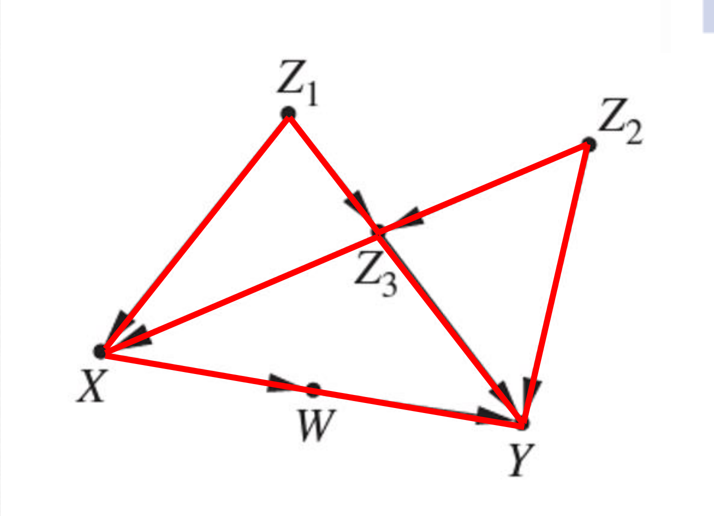
Biased dataset

Our dataset is always biased towards luminous and nearby objects.
i.e. Luminosity→visibility←distance

More complicated DAGs

Two variables are *independent* (*d-separated*) if every path between them is blocked.

	uncondition	condition
Chain	unblock	block
Fork	unblock	block
Collider	block	unblock



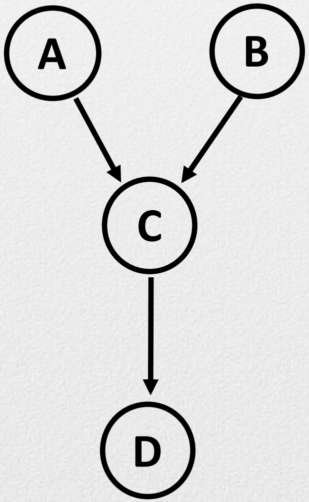
$$Z_2 \not\perp\!\!\!\perp X$$

$$Z_2 \perp\!\!\!\perp X \mid (Z_3, Z_1)$$

$$Z_2 \not\perp\!\!\!\perp X \mid (Z_3, Z_1, Y)$$

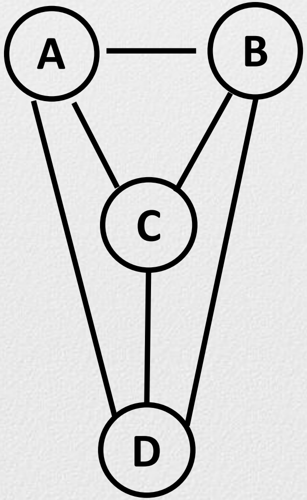
⋮

Peter-Clark (PC) Algorithm

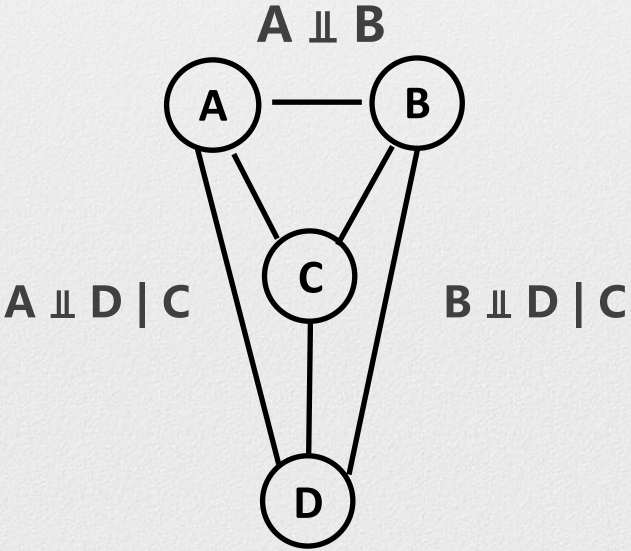


Ground Truth

Stage 1: Skeleton discovery

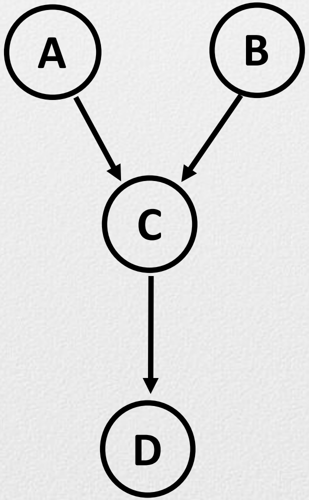


Fully connected graph



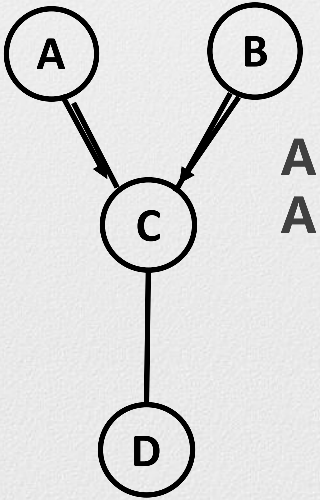
Skeleton

Peter-Clark (PC) Algorithm



Ground Truth

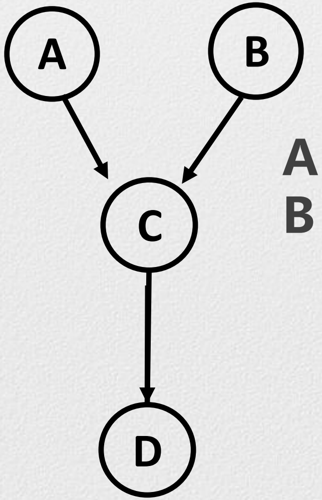
Stage 2: Orientation



$A \perp\!\!\!\perp B$

$A \not\perp\!\!\!\perp B \mid C$

Orientation



$A \not\perp\!\!\!\perp D$

$B \not\perp\!\!\!\perp D$

Orientation

Conditional independence tests

Conditional independencies	Conditional independence test p-value			
	Col-OSSOS		DES	
	Fisher-Z	KCI	Fisher-Z	KCI
$a \not\perp\!\!\!\perp e$	0.000	0.000	0.000	0.000
$a \perp\!\!\!\perp$ inc	0.144	0.421	0.006	0.568
$a \perp\!\!\!\perp$ colour	0.918	0.488	0.677	0.250
$e \not\perp\!\!\!\perp$ inc	0.005	0.085	0.000	0.001
$e \perp\!\!\!\perp$ colour	0.211	0.135	0.181	0.010
inc $\not\perp\!\!\!\perp$ colour	0.001	0.009	0.000	0.000
$a \not\perp\!\!\!\perp e \mid$ inc	0.000	0.068	0.000	0.001
$e \not\perp\!\!\!\perp$ inc $\mid a$	0.018	0.070	0.000	0.002
$e \not\perp\!\!\!\perp$ inc $\mid$ colour	0.010	0.041	0.000	0.003
inc $\not\perp\!\!\!\perp$ colour $\mid e$	0.003	0.004	0.000	0.000

2 Planets II: Cause of Hot-Jupiter Inflation

复旦大學

- Hot Jupiters often have radii larger than predicted by standard cooling–contraction models, but it remains unclear which process supplies or preserves the extra internal heat.

- Some source of external heat is needed
  - radiation from the star: thermal tides, kinetic/mechanical heating, ohmic dissipation
    - Planet radius (R_p), or fraction radius excess $\Delta_i \equiv \frac{R_p - R_0}{R_0}$ is likely linked to F $\propto T_{\rm eff}^4 P_{\rm orb}^{-4/3}$
  - Gravity from the star: gravitational tides
    - R_p or Δ_i is likely linked to $P_{\rm orb}$
  - Heat is locked inside the planet: delayed cooling, Layered convection, Potential-temperature advection
    - Weaker link to $T_{\rm eff}$ and $P_{\rm orb}$

Causal discovery to infer causal structure among R_p, M_p, $T_{\rm eff}$, $P_{\rm orb}$ in observation data

2 Planets II: Cause of Hot-Jupiter Inflation

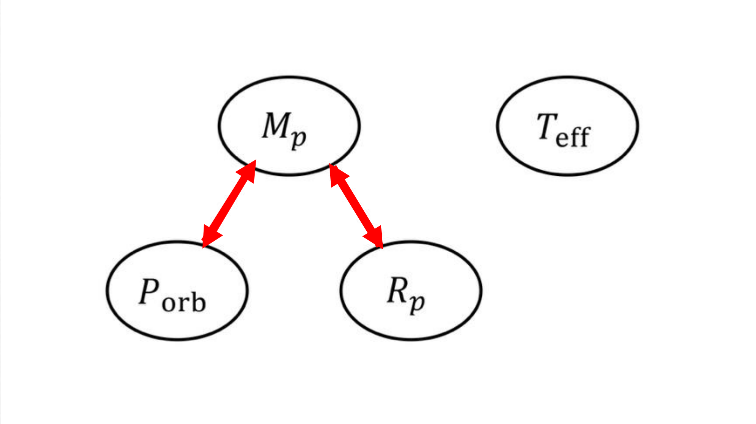
復旦大學

- Data: Encyclopedia of exoplanetary systems (J. Schneider et al. 2011)
    - 328 hot Jupiters
    - 179 super-Earths for sanity check
    - 4 nodes: R_p, M_p, T_{eff}, P_{orb}

- Causal discovery Method:
    - Go over possible causal graphs, checking the likelihood that the observed data is compatible with the conditional independence relationships implied by the causal graph, until arriving at a causal graph with the highest likelihood.
        - Specifically, a search strategy called Best Order Score Search (BOSS; B. Andrews et al. 2023)
        - A likelihood defined by the Fourier Feature Marginal Likelihood (FFML) score (J. D. Ramsey 2026)

2 Planets II: Cause of Hot-Jupiter Inflation

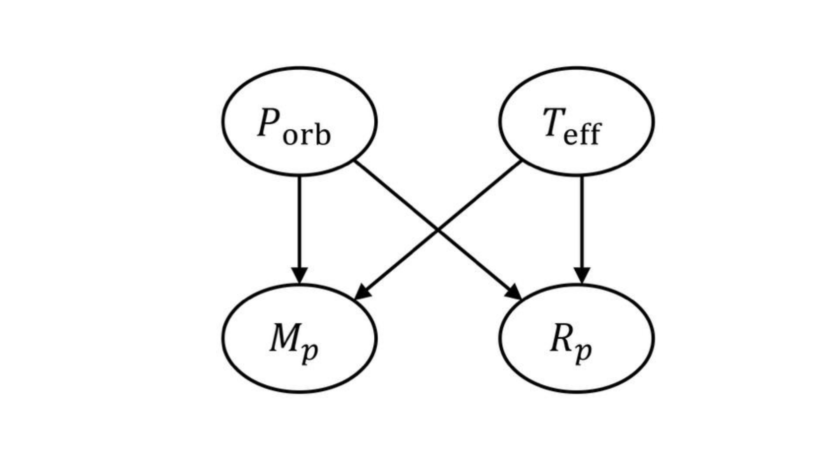
Sanity Check on Super Earth



- R_p governed by M_p
- $M_p - P_{\mathrm{orb}}$ reflects formation and migration in setting which planet masses are delivered to close-in orbits and survive there
- Isolated T_{eff} node since super-Earths have already lost their primordial envelope

Planets II: Cause of Hot-Jupiter Inflation

Hot Jupiters



- R_p driven by $T_{\rm eff}$ and $P_{\rm orb}$ ➔ irradiation-regulated inflation.
- The arrows into M_p: interpreted as population-level formation, migration, tidal-survival, photoevaporation/Roche-limit, and detection/selection effect.
- Purely period-controlled gravitational tides are disfavored.

2 Planets II: Cause of Hot-Jupiter Inflation

復旦大學

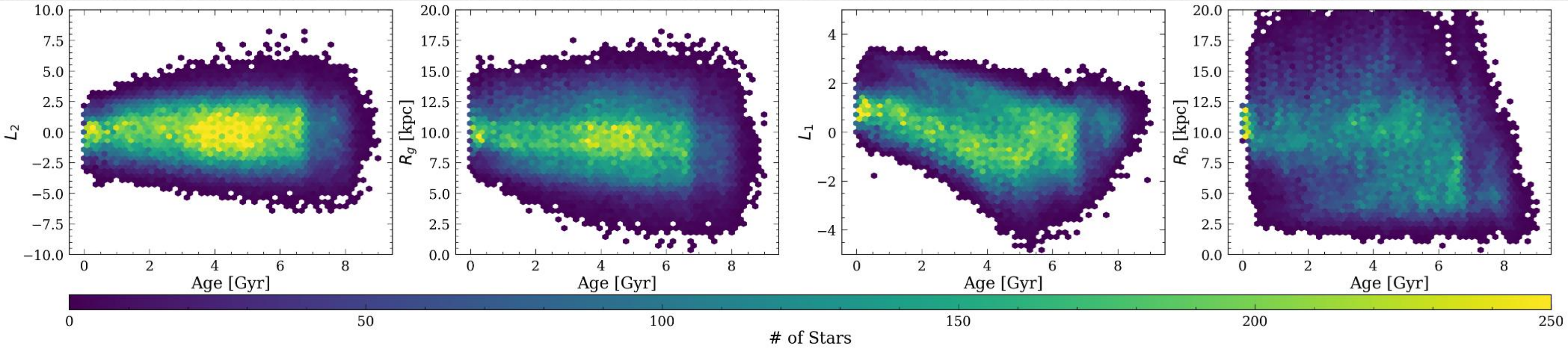
Table 1. Numerical Local Theoretical Scalings for the Radius-excess Law $\Delta_i \propto M_p^{\alpha_M} P_{\rm orb}^{\alpha_P} T_{\rm eff}^{\alpha_T}$.

Scenario	α_M	α_P	α_T	Dominant Causal Implication	Comparison to Data-driven Causal Graph
Ohmic dissipation, pre-turnover	−0.62	−0.33	+1.00	$M_p \to R_p$, $T_{\rm eff} \to R_p$	Dominant $M_p \to R_p$ not identified in observation
Ohmic dissipation, near efficiency peak	−0.62	−0.27	+0.80	$M_p \to R_p$, $T_{\rm eff} \to R_p$	Dominant $M_p \to R_p$ not identified in observation
Thermal tides, Gold-Soter	−0.42	−0.60	+0.60	$P_{\rm orb} \to R_p$, $T_{\rm eff} \to R_p$	Cleanest match
Thermal tides, dynamical	−0.44	−0.80	0.00	$M_p \to R_p$, $P_{\rm orb} \to R_p$	Missing $T_{\rm eff} \to R_p$
Kinetic / mechanical heating	−0.40	−0.27	+0.80	$M_p \to R_p$, $T_{\rm eff} \to R_p$	Dominant $M_p \to R_p$ not identified in observation
Eccentricity / obliquity tides	−0.44	−1.00	0.00	$M_p \to R_p$, $P_{\rm orb} \to R_p$	Missing $T_{\rm eff} \to R_p$
Radiative opacity / delayed cooling	−0.78	−0.08	+0.25	$M_p \to R_p$, $T_{\rm eff} \to R_p$	Missing $P_{\rm orb} \to R_p$
Layered convection, fixed layer strength	−1.06	0.00	0.00	$M_p \to R_p$	Missing $P_{\rm orb} \to R_p$, $T_{\rm eff} \to R_p$
Potential-temperature advection	−1.06	−0.33	+1.00	$M_p \to R_p$, $T_{\rm eff} \to R_p$	Dominant $M_p \to R_p$ not identified in observation

NOTE— The entries are the exponents in Eq. (1): α_M exponentiates M_p, α_P exponentiates $P_{\rm orb}$, and α_T exponentiates stellar $T_{\rm eff}$. The values assume $R_0 \propto M_p^{-0.06}$, $T_{\rm int} \propto M_p^{0.50}$ at fixed age for active-heating channels, $\chi = 0.20$ for active deep heating, fixed stellar mass and radius, fixed age and composition, fixed B, fixed e or obliquity, and fixed tidal Q'_p. The first ohmic row applies on the rising, pre-turnover branch; the second applies near the empirical heating-efficiency maximum. The dominant causal implication is the qualitative expected parent set; the mapping from exponent magnitude to graph edge is heuristic because causal discovery explores conditional dependence, not the local power-law slope.

- Gold–Soter thermal tides provide the closest single-mechanism match
- kinetic/mechanical heating and ohmic dissipation remain viable contributors
- Purely period-controlled gravitational tides and purely delayed cooling/layer convention are disfavored.

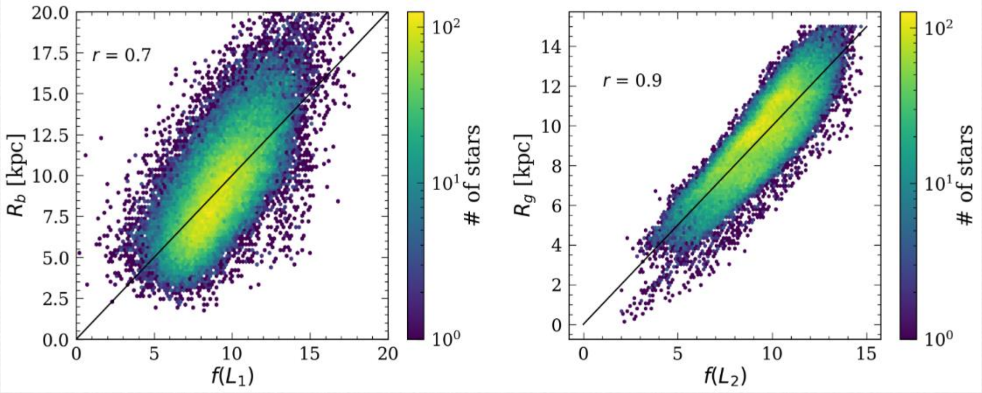
3 Stars: Galactic Archaeology

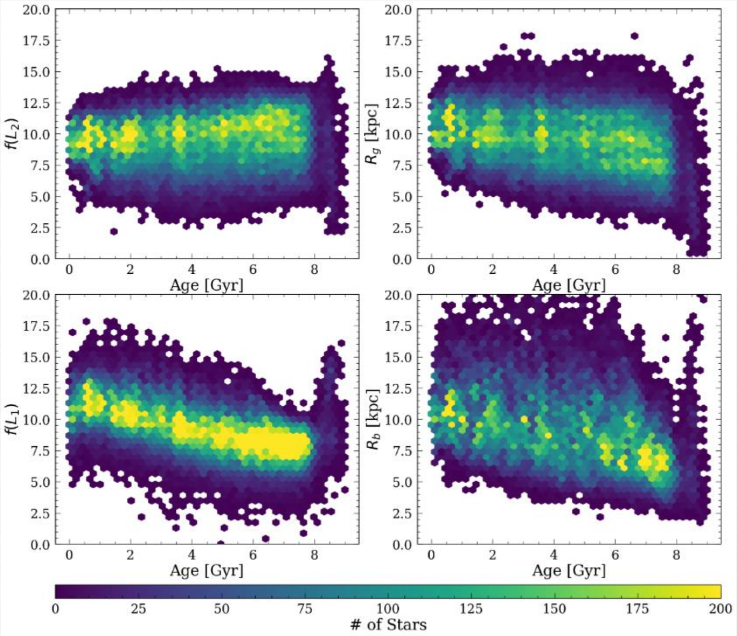


Notably, L1 (third plot) clearly captures the two star-forming sequences seen in R_b(fourth plot), which are produced by infalling gas from the merging satellite.

3 Stars: Galactic Archaeology

Causal model extrapolated to a new simulated galaxy

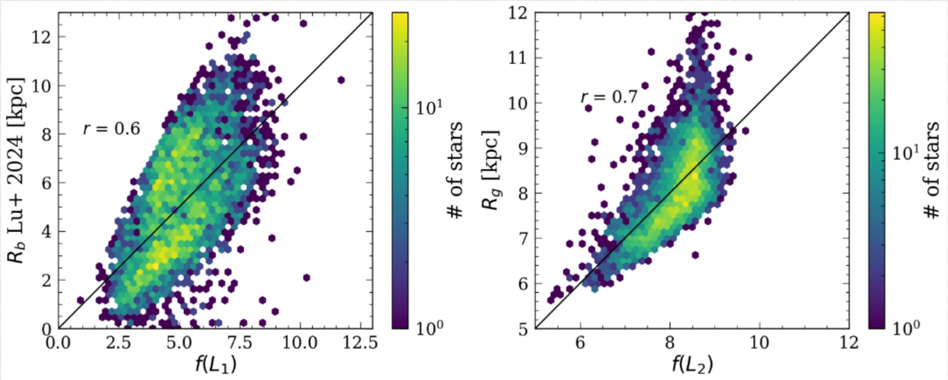


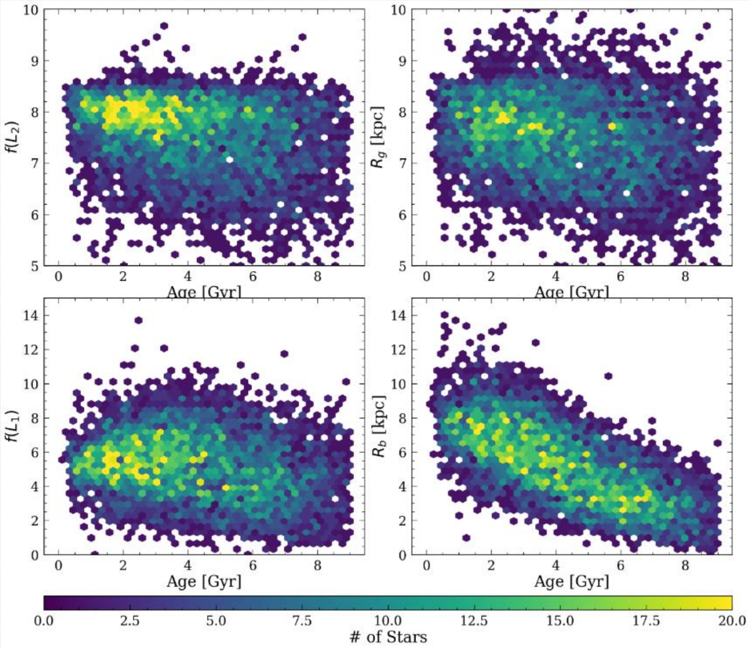


3 Stars: Galactic Archaeology

復旦大學

Causal model extrapolated to the real Milky Way



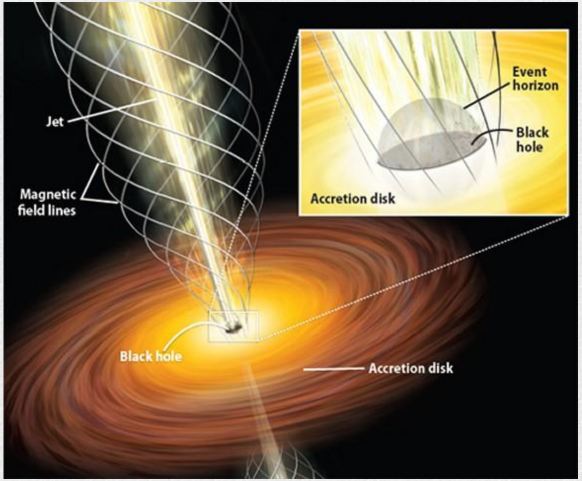


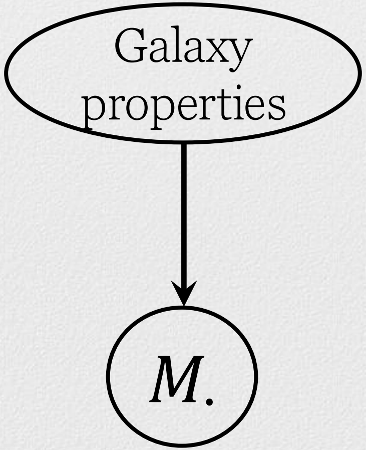
4 Galaxies：Galaxy – SMBH coevolution

Causal Discovery in Astrophysics: Unraveling Supermassive Black Hole and Galaxy Coevolution

Z Jin*, M Pasquato †, BL Davis †, T Deleu,..., Y Bengio, X Kang, AV Maccio, Y Hezaveh, 2025, *ApJ*

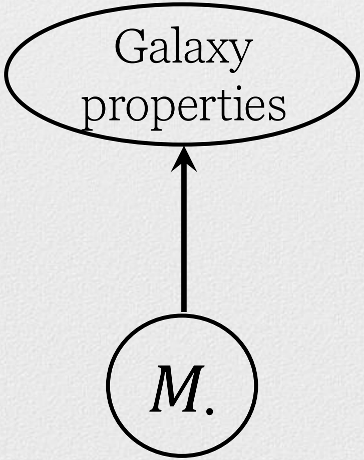
Causal Reversal in the $M_{\bullet}$–σ_0 Relation: Implications for High-redshift Supermassive Black Hole Mass Estimates, BL Davis, S More, **Z Jin***, M Pasquato, AV Macciò, F Yuan, 2026, ApJ





星系性质

黑洞质量



black hole accretion, Mergers
黑洞吸积，(星系/黑洞)并和

black hole feedback
黑洞反馈

4 Galaxies：Galaxy – SMBH coevolution

- Calculate P(Graph|Data) for all possible DAGs (1,138,779,265 in total)

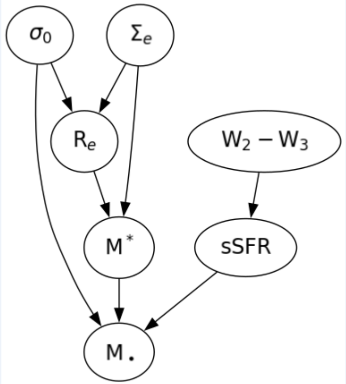
- P(Data|Graph) = Bayesian Gaussian Equivalent (BGe) score, by comparing the conditional independence found in the data vs. in the hypothetical graph

Real Data of 7 nodes:

$M_\bullet$ = 101 local, dynamical black hole mass

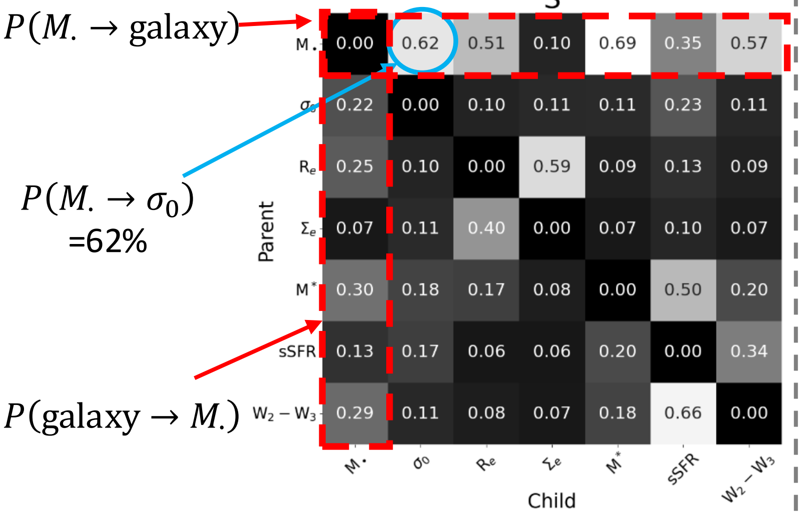
σ_0 = velocity dispersion
R_e = half-light radius
$\langle\Sigma_e\rangle$ = average surface density within R_e
M^* = stellar mass
$W_2 - W_3$ = WISE color
$sSFR$ = specific star-formation rate

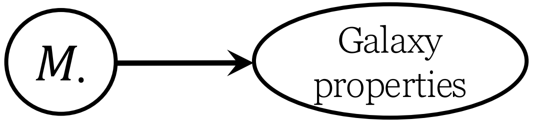
One example out of 1,138,779,265 DAGs



- GPU parallelization with JAX，
- GFlowNet estimation version also available for more nodes

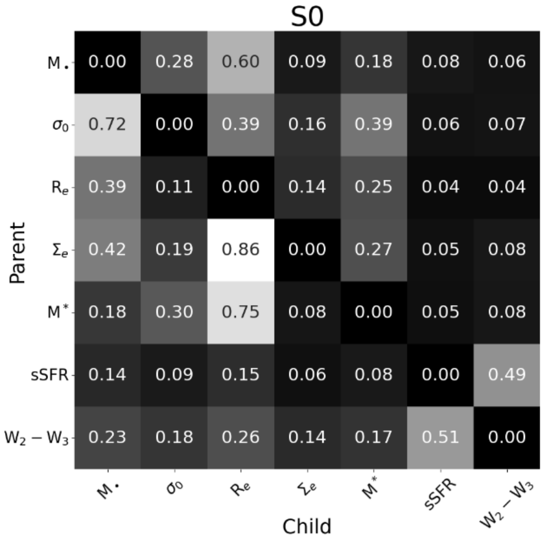
Spirals / Star forming galaxies





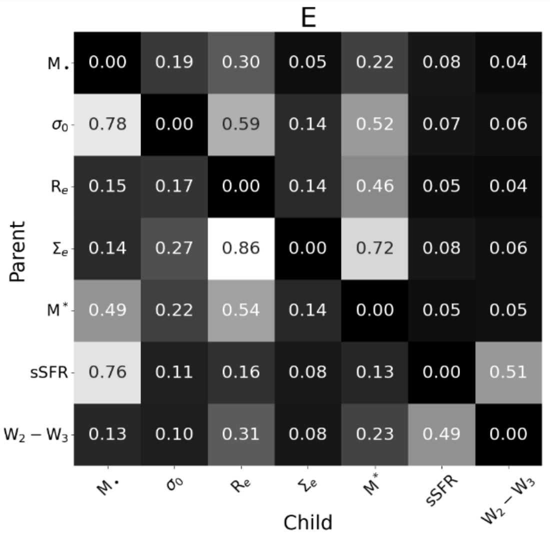
- Gas-rich
- BH feedback pushes gas out and progressively quench star formation.

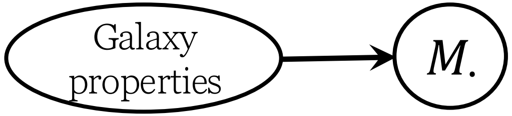
Lenticulars



Transition phase

Ellipticals / Quiescent galaxies



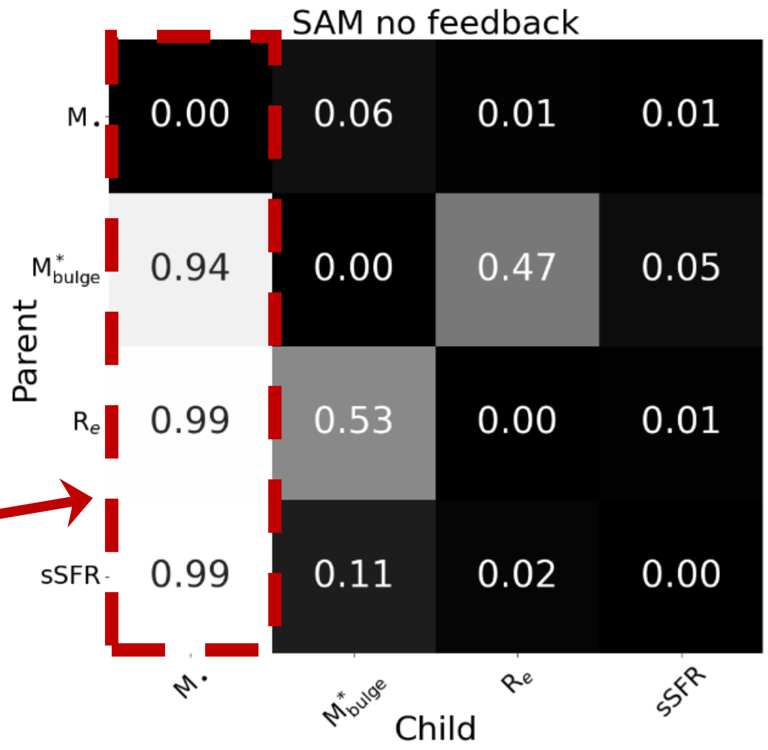


- Quenching is complete
- AGN feedback is still there but gas is depleted
- Galaxy influence SMBH by mergers

Sanity Check with Semi-Analytical Models (SAMs)

Manually shut down BH feedback in SAM, such that SMBH → Galaxy is not possible

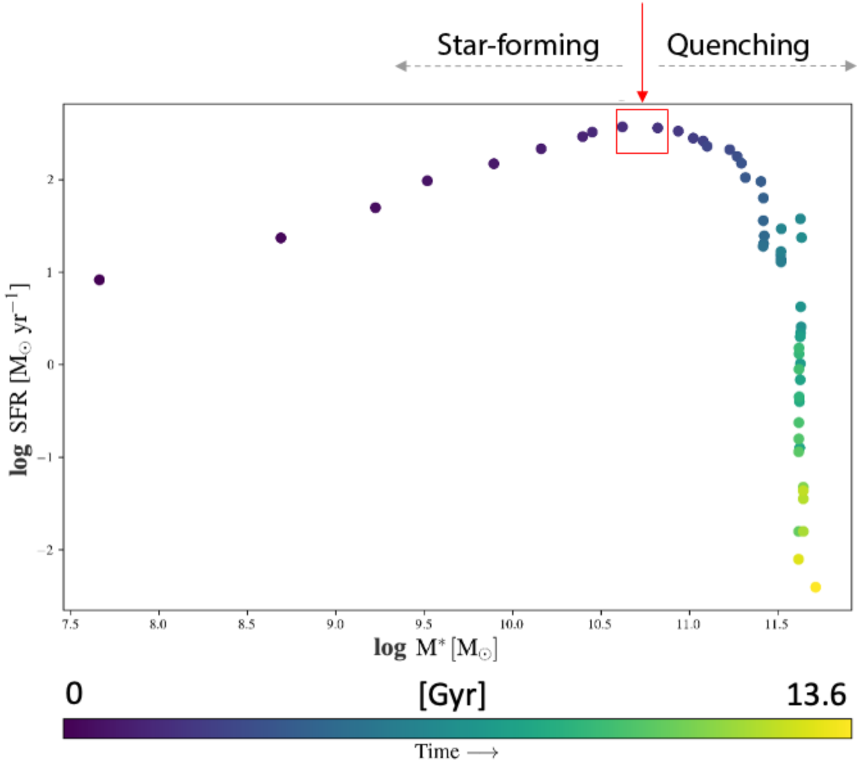
Should Only find Galaxy → SMBH



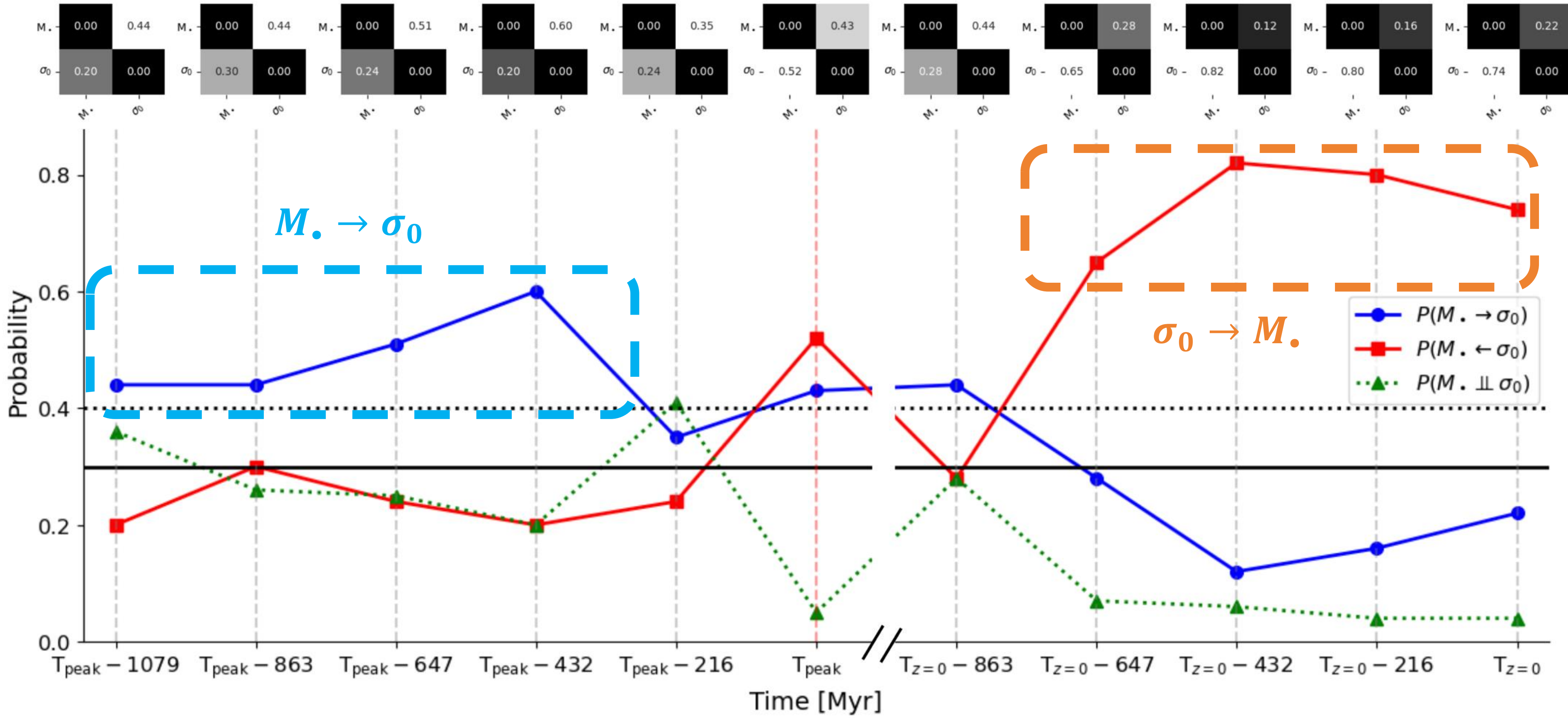
Causality as a function of time

- Data: NIHAO cosmological zoom-in hydrodynamical simulations
- 27 galaxies that quench by z = 0
- Each 64 snapshots
  - 216 Myr between each

- Align snapshots at $\mathrm{SFR}_{\mathrm{peak}}$
- Causal discovery on five snapshots before $\mathrm{SFR}_{\mathrm{peak}}$, and five last snapshots



The $M_\bullet$–σ_0 Causation as a Function of Time



Causally-informed *M*-σ Relations

Different Causal direction

⬇

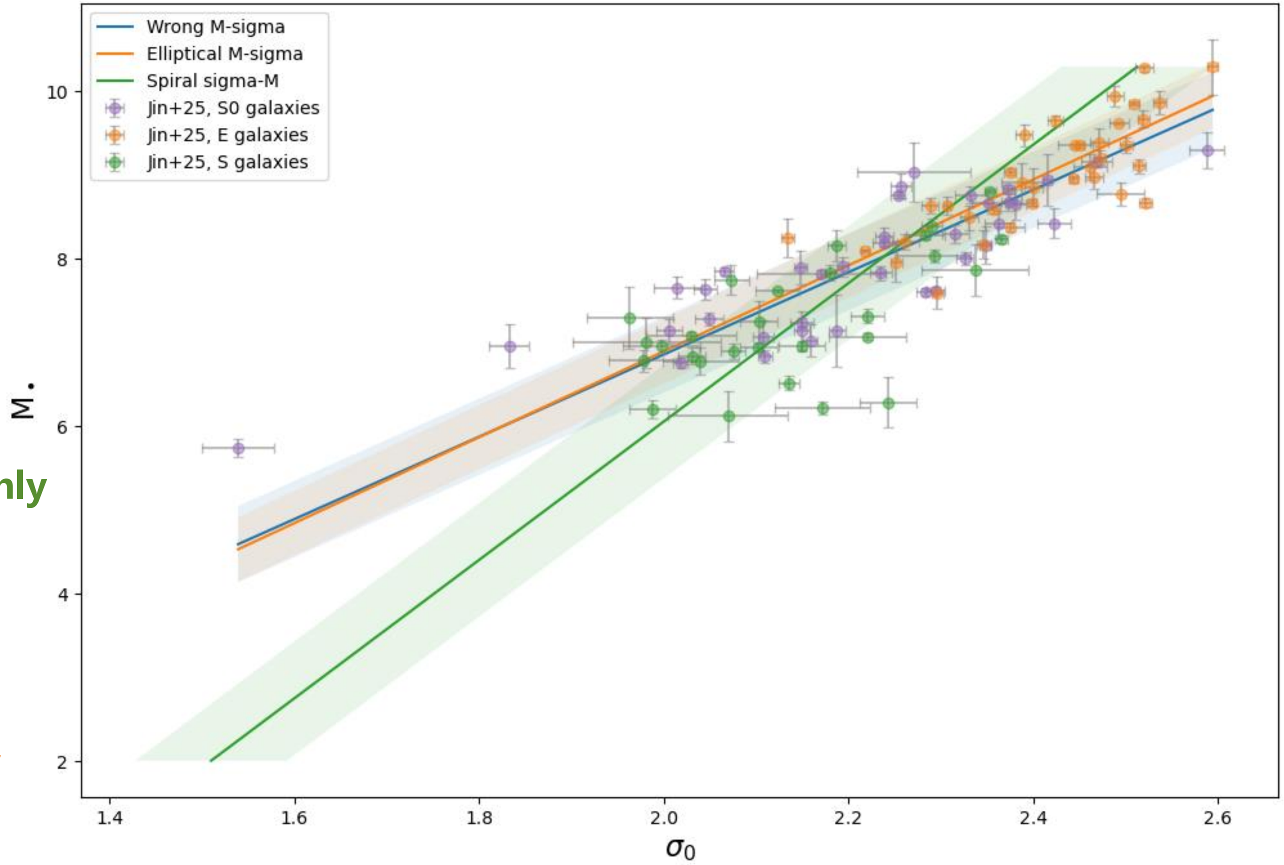
Different Physical mechanism

⬇

Different $M_\bullet - \sigma_0$ relation

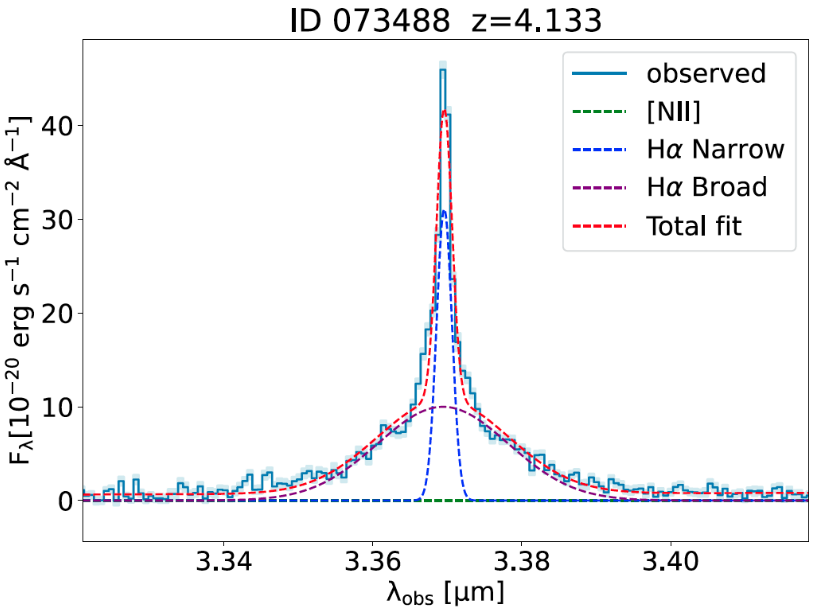
Spirals/star-forming:
SMBH → Galaxy
"$\sigma_0 - M_\bullet$ relation"
i.e. $\sigma_0 = \alpha\, M_\bullet + \beta$,
Fit with star forming galaxies only

Ellipticals/Quiescent:
Galaxy → SMBH
"$M_\bullet - \sigma_0$ relation"
i.e. $M_\bullet = \alpha\, \sigma_0 + \beta$,
Fit with Quiescent galaxies only



Estimation of High- z BH Masses

- Fit $M_\bullet - \sigma_0$ relation with dynamical BH masses at z = 0
- Reverberation Mapping (RM):
  $$GM_\bullet = f\,R_{BLR}(\Delta V)^2$$
  - "virial factor" f is calibrated from local $M_\bullet - \sigma_0$
- Single-epoch spectroscopic SMBH masses are correlated with RM
  - via the broad-line region (BLR) radius-luminosity relation.

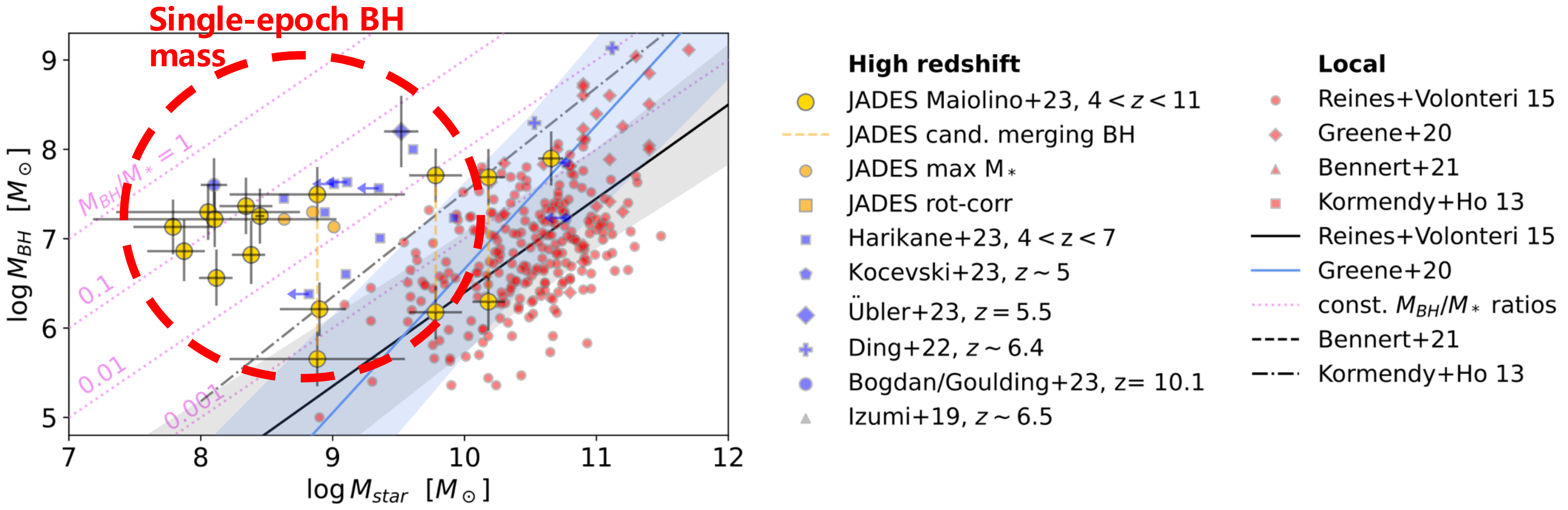


Maiolino et al. (2024)

$$\log\left(\frac{M_{\rm BH}}{M_\odot}\right) = 6.60 + 0.47\log\left(\frac{L_{\rm H\alpha}}{10^{42}\ {\rm erg/s}}\right) + 2.06\log\left(\frac{\rm FWHM_{H\alpha}}{10^3\ {\rm km/s}}\right)$$

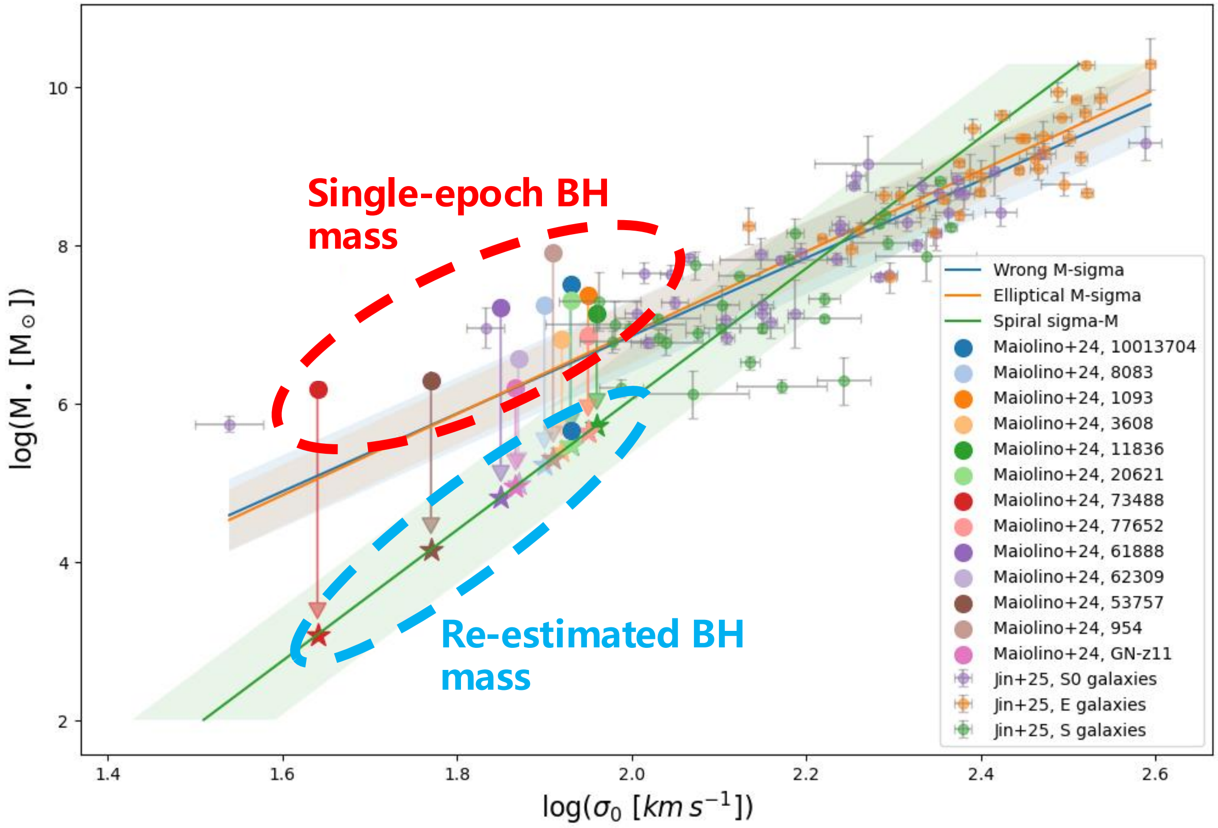
Reines & Volonteri (2015)

Revising black holes masses in 13 star-forming, late-type systems at 4.1 < z < 10.6 from Maiolino et al. (2024)



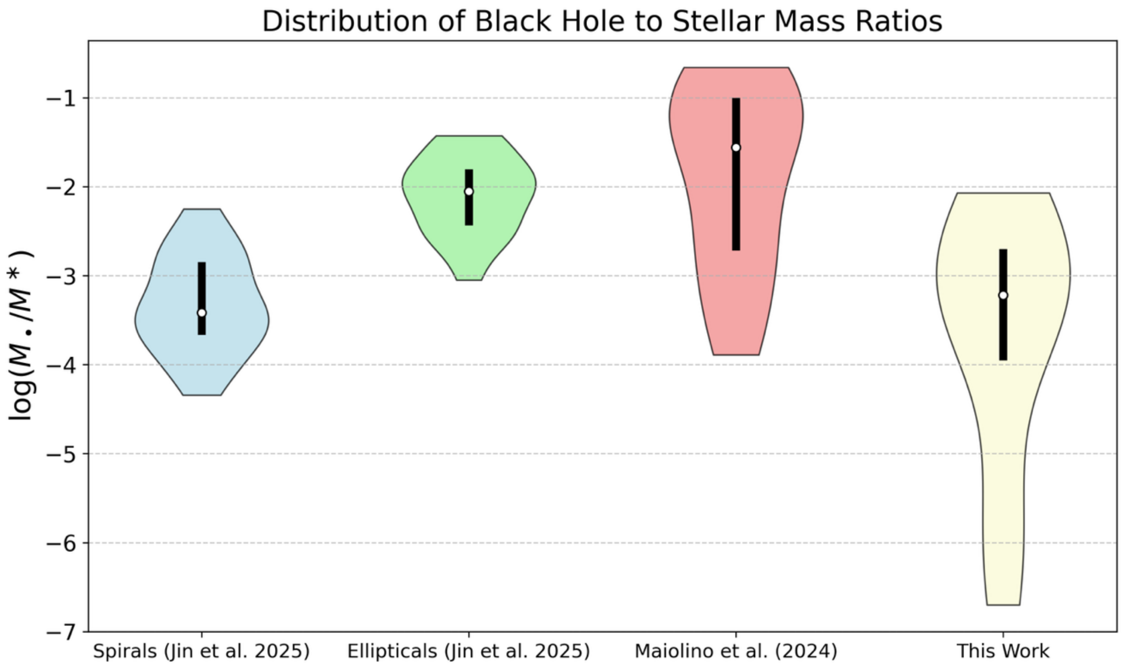
Maiolino et al. (2024)

Revising black holes masses in 13 star-forming, late-type systems at 4.1 < z < 10.6 from Maiolino et al. (2024)



$M_\bullet$ might be overestimated, especially on the lower end

Revising black holes masses in 13 star-forming, late-type systems at $4.1 < z < 10.6$ from Maiolino et al. (2024)



median reduction of $\log(M_{\bullet}/M^{*})$ = 1.93±0.51 dex

4 Galaxies: Galaxy – SMBH coevolution

复旦大学

- Spirals/star-forming: SMBH → Galaxy,
  - AGN feedback regulates
  - Fit $\sigma_0 - M_\bullet$ relation with spirals
- Ellipticals/Quiescent: Galaxy → SMBH
  - Gas poor, hierarchical assembly governs
  - Fit $M_\bullet - \sigma_0$ relation with ellipticals

- High- z BH Masses might be overestimated