

Causal Evidence for the Primordality of Colors in Trans-Neptunian Objects

Benjamin L. Davis*¹ Mohamad Ali-Dib*¹ Yujia Zheng*² Zehao Jin*^{1,3} Kun Zhang^{2,4}
Andrea Valerio Macciò¹

Abstract

The origins of the colors of Trans-Neptunian Objects (TNOs) represent a crucial unresolved question, central to understanding the history of our Solar System. Recent observational surveys have revealed correlations between the eccentricity and inclination of TNOs and their colors. This has rekindled the long-standing debate on whether these colors reflect the conditions of TNO formation or their subsequent collisional evolution. In this study, we address this question with 98.7% certainty, using a model-agnostic, data-driven approach based on causal graphs. First, as a sanity check, we demonstrate how our model can replicate the currently accepted paradigms of TNOs' dynamical history, blindly and without any orbital modeling or physics-based assumptions. In fact, our causal model (with no knowledge of the existence of Neptune) predicts the existence of an unknown perturbing body, i.e., Neptune. We then show how this model predicts, with high certainty, that the color of TNOs is the root cause of their inclination distribution, rather than the other way around. This strongly suggests that the colors of TNOs reflect an underlying dynamical property, most likely their formation location. Moreover, our causal model excludes formation scenarios that invoke substantial color modification by subsequent irradiation. We therefore conclude that the colors of TNOs are predominantly primordial.

1. Introduction

Trans-Neptunian Objects (TNOs) are invaluable probes into the history and evolution of our Solar System (Morbidei & Nesvorný, 2020). However, the wealth of information they encode is often difficult to decipher. This includes intrinsic characteristics such as their sizes and correlated properties such as their orbits and surface photometric colors. The last two have long been closely examined in an effort to unravel the relation between them (Jewitt & Luu, 2001).

Although the history of these studies is long, here we focus on Marsset et al. (2019) who found a strong correlation between the inclination and colors of TNOs. These results were expanded by Ali-Dib et al. (2021), who found an analogous trend where the eccentricity of VROs is cutoff at 0.42. They concluded that, in causality theory jargon, eccentricity (e) and inclination (i) are caused by the colors, which are indicative of the formation location. The primordial origin hypothesis of the TNO color diversity argues that TNO colors reflect compositional gradients in the protoplanetary disk, preserved since formation (Nesvorný et al., 2020; Ali-Dib et al., 2021). However, in alternatively, many works (Luu & Jewitt, 1996; Stern, 2002) argued that collisional evolution is the origin of TNO colors, where collisions expose fresh subsurface ices or organic materials, altering albedo and spectral slopes. A third possibility proposed that initially diverse bulk compositions undergo selective volatile evaporation post-formation, establishing steep compositional gradients across the primordial disk that, coupled with subsequent UV photolysis and particle irradiation, yield distinct surface chemistries (Brown et al., 2011; Wong & Brown, 2017).

In this paper, we use a purely data-driven, model-agnostic, statistical causal discovery method to study the relationships between the dynamical parameters of TNOs, and between those and the TNO colors. We show that not only this technique allows us to derive some of the main lines of the current consensus on the origins of TNOs, but also that it elucidates the direction of causality between the dynamical parameters and colors of TNOs, and predicts the existence of an unknown perturbing body, i.e., Neptune.

*Equal contribution ¹Center for Astrophysics and Space Science (CASS), New York University Abu Dhabi, PO Box 129188, Abu Dhabi, UAE ²Carnegie Mellon University, Pittsburgh, PA, USA ³Center for Astronomy and Astrophysics and Department of Physics, Fudan University, Shanghai 200438, People's Republic of China ⁴Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE. Correspondence to: Benjamin L. Davis <ben.davis@nyu.edu>.

2. Methodology

2.1. Data

Our dataset is based on (but not exclusively) the Col-OSSOS survey (Schwamb et al., 2019). It was taken from Marsset et al. (2019) and Ali-Dib et al. (2021). It consists of a total of 229 TNOs including hot classicals, centaurs, and resonant/scattered objects, in a dataset for which discovery biases were modeled. For each TNO, we have three orbital elements: semimajor axis (a), eccentricity (e), and inclination (i); and we have spectral slope (i.e., color).

A fundamental assumption of this work is that colors are primordial, and thus strongly correlated to the initial location of a TNO. Hereafter, we treat colors as a proxy for the initial semimajor axis of the objects. See Fig. 1 for a pairplot showing all the pairwise relations between our data. Additionally, Fig. 1 shows the subpopulation in our data by separating each TNO by its classifications as either a Classical (48), Resonant (102), Centaur (36), Scattered (28), or Detached (15) object.

2.2. Causal Discovery: Identifying Cause-effect Relationships

Identifying cause-effect relationships is crucial for moving beyond mere correlation to uncover the underlying causal mechanisms governing a system. Traditionally, causal relationships are established through interventions or randomized experiments, where one variable is explicitly manipulated while all others are held constant, and the resulting effects are observed. However, such interventions are infeasible in fields like astronomy, where the “test subjects” exist at unreachable astronomical distances. Consequently, advanced methods are required to infer causal relationships from purely observational data—an endeavor that lies at the core of causal discovery (Spirtes et al., 2001).

The foundation of causal discovery lies in uncovering the footprints of causality embedded in data. One of the most important sources of such information is dependency relations. By analyzing conditional independence among different components of an observed system, we can infer causal relationships between pairs of variables. This allows us to construct a graph that encodes the results of essential conditional independence tests, revealing which variables *cause* others under appropriate conditions. Ideally, the output is a Directed Acyclic Graph (DAG) for a unique solution or a Completed Partially Directed Acyclic Graph (CPDAG) for a Markov equivalence class. However, when some variables remain unmeasured, certain causal relationships may be undetermined, leading to a Partial Ancestral Graph (PAG). For further reading on causal discovery and causality, see *Causation, Prediction, and Search* (Spirtes et al., 2001), *Causality* (Pearl, 2009), or the review in Jin et al. (2025b,

§2).¹

2.3. Causal Structures with Latent Variables

Since it is impossible to measure all variables in the Universe, latent variables are always present. These unmeasured variables can significantly impact the correctness of the causal structure discovered. For example, suppose that X and Y are independent in the general population, but a sample is selected based on a variable Z that influences both X and Y . In that case, X and Y may exhibit statistical dependence in the sample, even though no such relationship exists in the population. This can lead to spurious causal conclusions, falsely suggesting a direct causal relationship between X and Y .

To address this challenge, we employ a principled approach capable of uncovering causal relationships even in the presence of latent variables. A widely used method for this purpose is Fast Causal Inference (FCI; Spirtes et al., 1995; Zhang, 2008), a constraint-based algorithm that has been proven to provide sound causal conclusions despite unmeasured variables. FCI has been applied across various scientific domains, including biology, economics, and climate science. For our analysis of TNO orbits, we use the FCI implementation in the Python package `causal-learn` (Zheng et al., 2024) to infer the underlying causal structure.

FCI discovers causal relationships by performing a series of conditional independence (CI) tests. These tests examine whether the statistical dependence between two variables disappears when controlling for other variables. If two variables become independent when conditioning on a third, this suggests that the third variable may be an intermediary or a common cause.

Unlike many causal discovery methods that assume that all relevant variables are measured (such as those producing DAGs or CPDAGs), FCI accounts for the possibility of unobserved variables. As a result, its output is a PAG, which provides more nuanced causal information. The edges in a PAG have different interpretations:

- $X \longrightarrow Y$: X is a *cause* of Y .
- $X \circ \longrightarrow Y$: Y is not an *ancestor* of X . Intuitively, this implies Y cannot be a cause of X , whether directly or indirectly.
- $X \circ - \circ Y$: No set d -separates X and Y . In other words, they may be causally adjacent or share a latent common cause.
- $X \longleftrightarrow Y$: There is a latent common cause of X and

¹In addition to Jin et al. (2025b), readers may be interested in further applications of causal discovery to astrophysical data (Pasquato et al., 2023; Pasquato, 2024; Jin et al., 2024; 2025a).

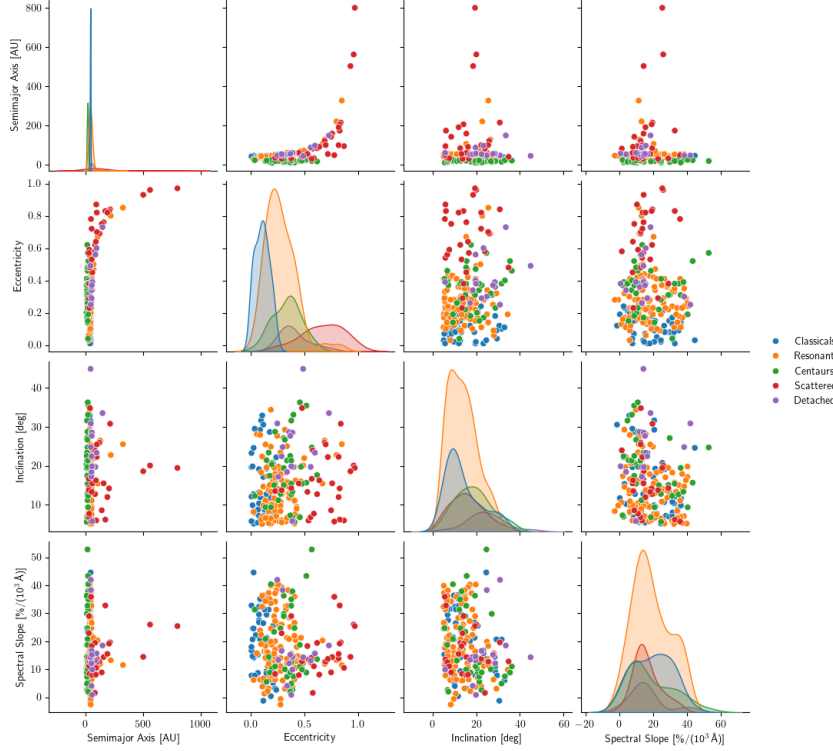


Figure 1. Pairplot of all 229 TNOs in our study. The TNOs are further divided into their individual populations: 48 Classicals (●), 102 Resonant (●), 36 Centaurs (●), 28 Scattered (●), and 15 Detached (●).

Y.

Therefore, by accounting for latent variables in the discovery process, we can uncover causal relations among measured variables while acknowledging uncertainties introduced by unmeasured factors. More importantly, when the algorithm cannot determine a definitive causal direction due to latent variables, it explicitly represents this uncertainty rather than arbitrarily assigning a direction. This principled approach distinguishes causal analysis from correlation-based techniques, ensuring that conclusions are drawn with a clear acknowledgment of underlying assumptions and limitations.

3. Results

3.1. Data-driven Results

Our primary findings are summarized in Fig. 2 showing the statistically most likely PAG fitting our data, at 98.7% confidence. This main result utilizes the FCI algorithm (Spirtes, 2001; Spirtes et al., 2013; Zheng et al., 2024), with linear Fisher-Z conditional independence tests (Fisher, 1921), and the threshold for each conditional independence test is $\alpha = 0.013$ (i.e., all tests must pass at the 98.7% level) on transformed data via Gaussianization. Specifically, we employ the Yeo-Johnson transformations (Yeo & Johnson, 2000) for our primary preprocessing. We still get the same

PAG using a linear Fisher-Z test without any transformation for $\alpha = 0.02$. It is also possible to directly use a non-linear conditional independence test. Here, we adopt a Kernel-based conditional independence (KCI) test (Zhang et al., 2012), with a polynomial kernel and reproduce the same PAG as in Fig. 2 at $\alpha = 0.09$.²

We emphasize that this PAG was obtained with a purely data-driven approach, without astrophysical insights. Moreover, we consistently reproduce the same PAG as in Fig. 2 by jackknifing our data by sequentially leaving out each subpopulation of TNOs. Thus, removing any subsample of 48 Classicals, 102 Resonant, 36 Centaurs, 28 Scattered, or 15 Detached TNOs results in no change to our discovered PAG. Therefore, we demonstrate that no single subpopulation is dominating the PAG and that our results are robust to outliers.

Alternatively, if we are to generate PAGs for the individual populations separately (i.e., analyzing only one subpopulation at a time), we find a large diversity in the results. Many of these PAGs however are based on very few datapoints. Taking this result at face value hints that our overall PAG represents that main-line dynamics dominate over the entire

²We did not apply non-linear tests in the first place because a non-linear method is prone to overfitting for the relatively small size of our data (229 TNOs).

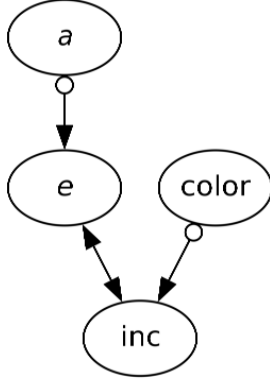


Figure 2. Partial Ancestral Graph (PAG) for 229 TNOs, calculated with the Fast Causal Inference (FCI) algorithm (Spirtes, 2001; Spirtes et al., 2013; Zheng et al., 2024), for linear Fisher-Z conditional independence tests (Fisher, 1921) on transformed data, with $\alpha = 0.013$ (significance level of individual partial correlation tests). On untransformed data, we recover the identical PAG with linear Fisher-Z tests and $\alpha = 0.02$, while the same PAG is produced with $\alpha = 0.09$ when we run Kernel-based conditional independence (KCI) tests (Zhang et al., 2012), with a polynomial kernel. This PAG has three causal edges, which can be described as follows: (i) eccentricity is not an ancestor of the semimajor axis, (ii) there is a latent common cause of eccentricity and inclination, and (iii) inclination is not an ancestor of color.

sample.

3.2. Astrophysical Interpretation

The first link we investigate is the one-way causal direction of the *current* semimajor axis causing the *current* eccentricity. While the correlation between a and e in TNOs is well established, the direction of the causality we find here is not surprising either, as its root physical causes are:

- Scattering by Neptune, where objects have to close-encounter Neptune first in order to get scattered into high eccentricity orbits. Moreover, objects usually cannot be both close to Neptune (today) and have a high eccentricity. It is the current semimajor axis of the objects that dictates what eccentricity they can have, and not the other way around.
- Mean motion resonances (MMRs), where the period (and thus current semimajor axis) of the objects dictates whether they are inside an eccentricity-raising resonance.

The connection $a \circ \rightarrow e$ rules out the possibility of $a \leftarrow e$. Clearly, $a \rightarrow e$ is possible, but also $a \leftrightarrow e$. The latter might imply that an unobserved confounder causes both a and e .

The second link in the PAG is the two-way dependency

between the eccentricity and the inclination, which is consistent with the von Zeipel-Lidov-Kozai (von Zeipel, 1910; Lidov, 1962; Kozai, 1962) anti-correlated oscillations between these two quantities (both inside and outside of MMRs), that plays a central role in the dynamics of TNOs. Here, $e \leftrightarrow i$ implies that there is an unobserved confounder. Indeed, the von Zeipel-Lidov-Kozai mechanism involves perturbations from a third body, here being Neptune. Moreover, if Neptune had not already been discovered in 1846, our result here would strongly suggest the presence of an unknown perturbing body. Together, the first two links successfully reestablish the main dynamical processes shaping the Kuiper belt (scattering, MMRs, and von Zeipel-Lidov-Kozai oscillations) without any physical inputs.

Finally, **the third piece** of the puzzle is the connection $\text{color} \circ \rightarrow i$ ruling out the possibility of $\text{color} \leftarrow i$. The “color” (i.e., a proxy for the formation location in our null hypothesis) is hence causing the inclination. This is again dynamically expected, as the formation location relative to inclination-raising secular resonances such as f_7 and f_8 will strongly affect the inclination distribution of TNOs (Murray & Dermott, 1999). Note that this link, however, leaves open the possibility of an unobserved confounder causing both color and the inclination. This confounder can be the formation location itself, if we were to assume the color and initial location to be two distinct variables instead of the color being a proxy for location.

Our result, that $\text{color} \leftarrow i$ is not allowed, rules out the model of Luu & Jewitt (1996) and Stern (2002), where collisional evolution shapes the colors of TNOs. Moreover, our result that $\text{color} \leftarrow a$ is not allowed either, rules out the model of Brown et al. (2011) and Wong & Brown (2017), where a would control the amount of irradiation a TNO is subjected to. We are hence left only with the “primordial origins” model where **the color is set entirely by the chemical composition of the formation location**.

4. Discussion & Conclusions

This work endeavors to resolve the tension between theories of primordial origins vs. subsequent evolution to account for the observed dispersion and correlations in TNO colors, a subject of a long debate. Our causal graph analysis, derived from a model-agnostic causal discovery framework, strongly favors the primordial origin hypothesis, with 98.7% certainty that *TNO color is causally antecedent to inclination, not a consequence of it*. Moreover, our model seems to exclude any effects from the current semimajor axis on the color of TNOs, disfavoring models where continuous irradiation plays a large role in shaping the colors. This will be explored further in the future. Finally, this work is a proof of principle for the use of causality models in planetary sciences. In our case, it even “rediscovered” the

existence of Neptune. With large datasets ranging from asteroids to exoplanets, many discoveries await.

Acknowledgements

This material is based on work supported by Tamkeen under the NYU Abu Dhabi Research Institute grant CASS. YZ and KZ are supported by NSF Award No. 2229881, AI Institute for Societal Decision Making, NIH R01HL159805, and grants from Quris AI, Florin Court Capital, and MBZUAI-WIS Joint Program. This research was carried out on the high-performance computing resources at New York University Abu Dhabi. This research has made use of NASA's Astrophysics Data System Bibliographic Services. The data and code used for this work are available for download from the following GitHub repository: <https://github.com/ZehaoJin/causalTNOs>.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

- Ali-Dib, M., Marsset, M., Wong, W.-C., and Dbouk, R. The Rarity of Very Red Trans-Neptunian Objects in the Scattered Disk. *The Astronomical Journal*, 162(1):19, July 2021. doi: 10.3847/1538-3881/abf6ca.
- Brown, M. E., Schaller, E. L., and Fraser, W. C. A Hypothesis for the Color Diversity of the Kuiper Belt. *The Astrophysical Journal Letters*, 739(2):L60, October 2011. doi: 10.1088/2041-8205/739/2/L60.
- Fisher, R. A. On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample. *Metron*, 1: 3–32, 1921.
- Jewitt, D. C. and Luu, J. X. Colors and Spectra of Kuiper Belt Objects. *The Astronomical Journal*, 122(4):2099–2114, October 2001. doi: 10.1086/323304.
- Jin, Z., Pasquato, M., Davis, B. L., Macciò, A. V., and Hezaveh, Y. Beyond Causal Discovery for Astronomy: Learning Meaningful Representations with Independent Component Analysis. *arXiv e-prints*, art. arXiv:2410.14775, October 2024. doi: 10.48550/arXiv.2410.14775.
- Jin, Z., Pasquato, M., Davis, B., Maccio, A., and Hezaveh, Y. Causal Discovery for Galaxy Evolution. In *American Astronomical Society Meeting Abstracts*, volume 245 of *American Astronomical Society Meeting Abstracts*, pp. 120.03D, January 2025a.
- Jin, Z., Pasquato, M., Davis, B. L., Deleu, T., Luo, Y., Cho, C., Lemos, P., Perreault-Levasseur, L., Bengio, Y., Kang, X., Macciò, A. V., and Hezaveh, Y. Causal Discovery in Astrophysics: Unraveling Supermassive Black Hole and Galaxy Coevolution. *The Astrophysical Journal*, 979(2): 212, February 2025b. doi: 10.3847/1538-4357/ad9ded.
- Kozai, Y. Secular perturbations of asteroids with high inclination and eccentricity. *The Astronomical Journal*, 67: 591–598, November 1962. doi: 10.1086/108790.
- Lidov, M. L. The evolution of orbits of artificial satellites of planets under the action of gravitational perturbations of external bodies. *Planetary Space Science*, 9(10):719–759, October 1962. doi: 10.1016/0032-0633(62)90129-0.
- Luu, J. X. and Jewitt, D. C. Color Diversity Among the Centaurs and Kuiper Belt Objects. *The Astronomical Journal*, 112:2310, November 1996. doi: 10.1086/118184.
- Marsset, M., Fraser, W. C., Pike, R. E., Bannister, M. T., Schwamb, M. E., Volk, K., Kavelaars, J. J., Alexandersen, M., Chen, Y.-T., Gladman, B. J., Gwyn, S. D. J., Lehner, M. J., Peixinho, N., Petit, J.-M., and Wang, S.-Y. Col-OSSOS: Color and Inclination Are Correlated throughout the Kuiper Belt. *The Astronomical Journal*, 157(3):94, March 2019. doi: 10.3847/1538-3881/aaf72e.
- Morbidelli, A. and Nesvorný, D. Kuiper belt: formation and evolution. In Prialnik, D., Barucci, M. A., and Young, L. (eds.), *The Trans-Neptunian Solar System*, pp. 25–59. 2020. doi: 10.1016/B978-0-12-816490-7.00002-3.
- Murray, C. D. and Dermott, S. F. *Solar System Dynamics*. 1999. doi: 10.1017/CBO9781139174817.
- Nesvorný, D., Vokrouhlický, D., Alexandersen, M., Bannister, M. T., Buchanan, L. E., Chen, Y.-T., Gladman, B. J., Gwyn, S. D. J., Kavelaars, J. J., Petit, J.-M., Schwamb, M. E., and Volk, K. OSSOS XX: The Meaning of Kuiper Belt Colors. *The Astronomical Journal*, 160(1):46, July 2020. doi: 10.3847/1538-3881/ab98fb.
- Pasquato, M. Bringing causality to astronomy. In *EAS2024, European Astronomical Society Annual Meeting*, pp. 362, July 2024.
- Pasquato, M., Jin, Z., Lemos, P., Davis, B. L., and Macciò, A. V. Causa prima: cosmology meets causal discovery for the first time. *arXiv e-prints*, art. arXiv:2311.15160, November 2023. doi: 10.48550/arXiv.2311.15160.
- Pearl, J. *Causality*. Cambridge university press, 2009.
- Schwamb, M. E., Fraser, W. C., Bannister, M. T., Marsset, M., Pike, R. E., Kavelaars, J. J., Benecchi, S. D., Lehner, M. J., Wang, S.-Y., Thirouin, A., Delsanti, A., Peixinho, N., Volk, K., Alexandersen, M., Chen, Y.-T., Gladman,

- B., Gwyn, S. D. J., and Petit, J.-M. Col-OSSOS: The Colors of the Outer Solar System Origins Survey. *The Astrophysical Journals*, 243(1):12, July 2019. doi: 10.3847/1538-4365/ab2194.
- Spirtes, P. An anytime algorithm for causal inference. In Richardson, T. S. and Jaakkola, T. S. (eds.), *Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics*, volume R3 of *Proceedings of Machine Learning Research*, pp. 278–285. PMLR, 04–07 Jan 2001. URL <https://proceedings.mlr.press/r3/spirtes01a.html>. Reissued by PMLR on 31 March 2021.
- Spirtes, P., Meek, C., and Richardson, T. Causal inference in the presence of latent variables and selection bias. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 499–506, August 1995.
- Spirtes, P., Glymour, C., and Scheines, R. *Causation, Prediction, and Search*. The MIT Press, 01 2001. ISBN 9780262284158. doi: 10.7551/mitpress/1754.001.0001. URL <https://doi.org/10.7551/mitpress/1754.001.0001>.
- Spirtes, P. L., Meek, C., and Richardson, T. S. Causal Inference in the Presence of Latent Variables and Selection Bias. *arXiv e-prints*, art. arXiv:1302.4983, February 2013. doi: 10.48550/arXiv.1302.4983.
- Stern, S. A. Evidence for a Collisional Mechanism Affecting Kuiper Belt Object Colors. *The Astronomical Journal*, 124(4):2297–2299, October 2002. doi: 10.1086/342863.
- von Zeipel, H. Sur l’application des séries de M. Lindstedt à l’étude du mouvement des comètes périodiques. *Astronomische Nachrichten*, 183(22):345, March 1910. doi: 10.1002/asna.19091832202.
- Wong, I. and Brown, M. E. The Bimodal Color Distribution of Small Kuiper Belt Objects. *The Astronomical Journal*, 153(4):145, April 2017. doi: 10.3847/1538-3881/aa60c3.
- Yeo, I. and Johnson, R. A. A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4):954–959, 12 2000. ISSN 0006-3444. doi: 10.1093/biomet/87.4.954. URL <https://doi.org/10.1093/biomet/87.4.954>.
- Zhang, J. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16-17): 1873–1896, 2008.
- Zhang, K., Peters, J., Janzing, D., and Schoelkopf, B. Kernel-based Conditional Independence Test and Application in Causal Discovery. *arXiv e-prints*, art. arXiv:1202.3775, February 2012. doi: 10.48550/arXiv.1202.3775.
- Zheng, Y., Huang, B., Chen, W., Ramsey, J., Gong, M., Cai, R., Shimizu, S., Spirtes, P., and Zhang, K. Causal-learn: Causal discovery in python. *Journal of Machine Learning Research*, 25(60):1–8, 2024.