

Bayesian Bootstrap Ensembles for Low-Rank Causal Discovery

Shuaidong Gao

gaoshuaidong@stu.cqifs.edu.cn

Chongqing Institute of Foreign Studies

Research Article

Keywords: Bayesian Causal Discovery, Low-Rank Factorization, Dirichlet-Weighted Bootstrap, Calibration, Gene Regulatory Networks, Posterior Inference

Posted Date: August 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10543892/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Bayesian Bootstrap Ensembles for Low-Rank Causal Discovery

Shuaidong Gao^{1*}

¹Chongqing Institute of Foreign Studies, Qijiang Campus, Chongqing 401420, China.

Corresponding author(s). E-mail(s): gaoshuaidong@stu.cqifs.edu.cn;

Abstract

Objective: Differentiable causal discovery methods produce point estimates of directed acyclic graphs (DAGs) but lack uncertainty quantification. Existing uncertainty-aware methods are limited to $d \leq 50$ due to $O(d^3)$ complexity, while genomic studies routinely profile $d \geq 500$ genes. We develop a Bayesian bootstrap ensemble of low-rank models to produce posterior edge probabilities under a Dirichlet process prior (Rubin 1981).

Results: Low-rank factorization ($W = UV^\top$, $r \ll d$) reduces the DAG constraint from $O(d^3)$ to $O(dr^2)$. Dirichlet-weighted Bayesian bootstrap across 20–50 training runs yields ensemble distributions over the adjacency matrix, outperforming standard resampling-based bootstrap on difficult configurations: at $d = 50$, standard bootstrap detects zero edges (F1 = 0.000) while the Bayesian bootstrap recovers edges (F1 = 0.025); at $d = 200$, the Bayesian bootstrap more than doubles F1 (0.080 vs. 0.034; Supplementary Table S4). On synthetic Erdős–Rényi DAGs ($d = 30$ – 500), expected calibration error (ECE) decreases from 0.036 at $d = 30$ (comparable to DiBS at 0.038) to 0.003 at $d = 500$, where DiBS is inapplicable. On TCGA BRCA expression data ($d = 500$, $n = 1,218$), the 21 edges with posterior probability exceeding baseline ($P > 0.1$, out of 250,000 candidates) achieve 71.4% combined STRING protein–protein interaction validation precision (vs. 10.6% baseline)—a 6.7-fold enrichment.

Keywords: Bayesian Causal Discovery, Low-Rank Factorization, Dirichlet-Weighted Bootstrap, Calibration, Gene Regulatory Networks, Posterior Inference

29 1 Introduction

30 Causal discovery from observational data is fundamentally an inference problem:
 31 given finite samples, we must determine which directed edges exist and with what
 32 confidence. Point-estimate methods—from constraint-based (PC [20]) and score-
 33 based (GES [21]) algorithms to continuous optimization approaches (NOTEARS [1],
 34 DAGMA [2], GOLEM [3])—collapse this inference to a single adjacency matrix,
 35 discarding information about which edges are well-supported [29]. In biomedical appli-
 36 cations, following up a spurious causal claim can cost months of laboratory work;
 37 uncertainty quantification is a practical necessity [30].

38 DiBS [4] introduced uncertainty quantification for causal discovery through particle
 39 variational inference. Subsequent work extended this to nonlinear settings [5, 6] and
 40 amortized inference [7, 8]. However, all existing methods share a common bottleneck:
 41 the $O(d^3)$ cost of the DAG constraint limits them to $d \leq 50$ variables. Genomic studies
 42 routinely profile $d \geq 500$ genes, making existing approaches inapplicable.

43 Low-rank factorization $W = UV^\top$ with $U, V \in \mathbb{R}^{d \times r}$ and $r \ll d$ scales causal
 44 discovery beyond $d = 200$, where standard NOTEARS produces zero edges [9, 10].
 45 Alternative approaches such as LiNGAM [32] address linear non-Gaussian settings
 46 but similarly lack uncertainty quantification. Under a rank- r true adjacency matrix,
 47 sample complexity is $\Omega(dr \log d)$ rather than $\Omega(d^2 \log d)$ [10]. Biologically, the low-rank
 48 assumption encodes that gene regulatory networks are governed by a relatively small
 49 number of transcription factors and pathways [11, 12].

50 We combine the scalability of low-rank factorization with the Dirichlet-weighted
 51 Bayesian bootstrap [14, 23, 31]. The Bayesian bootstrap draws weights $w \sim$
 52 $\text{Dirichlet}(1, \dots, 1)$ per sample and minimizes the reweighted NOTEARS objective.
 53 Under a Dirichlet process prior, the resulting ensemble provides approximate poste-
 54 rior samples over the adjacency matrix at the $O(dr^2)$ complexity of low-rank methods.
 55 Compared to standard bootstrap (sampling with replacement, which uses approxi-
 56 mately 63% unique samples per iteration), the Bayesian bootstrap uses all data with
 57 continuous importance weights, empirically achieving comparable or better edge recov-
 58 ery (Supplementary Table S4; e.g., F1 = 0.025 vs. 0.000 at $d = 50$; 0.080 vs. 0.034 at
 59 $d = 200$).

60 The bootstrap has been applied to causal discovery for stability selection [18] and
 61 edge confidence estimation [19, 25], but has not been combined with differentiable
 62 methods at scale because the $O(d^3)$ per-iteration cost makes resampling prohibitive.
 63 Low-rank factorization removes this bottleneck, enabling Bayesian bootstrap ensem-
 64 bles at dimensions where no other uncertainty-aware method operates.

65 2 Main text

66 2.1 Bayesian Bootstrap for Posterior Inference

67 Following NOTEARS [1], we model data via the linear structural equation $X = XW +$
 68 ε with DAG constraint $h(W) = \text{tr}(e^{W \odot W}) - d = 0$. The low-rank parameterization
 69 $W = UV^\top$ with $U, V \in \mathbb{R}^{d \times r}$ reduces parameters from $O(d^2)$ to $O(2dr)$ and the

70 constraint cost from $O(d^3)$ to $O(dr^2)$, enabling scaling to $d = 500$. The optimization
 71 objective is:

$$\min_{U,V} \frac{1}{2n} \|X - X(UV^\top)\|_F^2 + \lambda \|W\|_1, \quad h(UV^\top) = 0 \quad (1)$$

72 where the ℓ_1 penalty [33] encourages sparse edge selection.

73 The rank r is selected by spectral analysis of the sample covariance matrix: r
 74 equals the number of eigenvalues exceeding 75% of the maximum eigenvalue [9]. For
 75 the configurations reported here, r ranges from 8 at $d = 30$ to 32 at $d = 500$.

76 For bootstrap $b \in \{1, \dots, B\}$, we draw Dirichlet weights $w^{(b)} \sim$
 77 $\text{Dirichlet}(1, \dots, 1)$ [23] and train a low-rank model minimizing the reweighted loss
 78 $\frac{1}{2} \sum_i w_i^{(b)} \|x_i - x_i W\|^2$. Model b is trained to convergence ($h(W) < 10^{-8}$ or $\rho = \rho_{\max}$)
 79 via the augmented Lagrangian method with initialization seed $42 + 100b$. The set of B
 80 adjacency matrices forms the approximate posterior over the adjacency matrix. The
 81 posterior edge probability is:

$$\tilde{P}(\text{edge}_{ij} \mid \text{data}) = \frac{1}{B} \sum_{b=1}^B \mathbb{I}(|W_{ij}^{(b)}| > \tau) \quad (2)$$

82 with $\tau = 0.3$ following the standard NOTEARS convention [1]. Under the Bayesian
 83 bootstrap [23], these are approximate posterior edge probabilities under a Dirichlet
 84 process prior; the approximation is asymptotically exact for smooth functionals and
 85 has been validated empirically for structure learning in Bayesian networks [19]. Cal-
 86 ibration is measured via Expected Calibration Error (ECE) with 10 equally spaced
 87 bins [15].

88 All experiments used PyTorch 2.x on an NVIDIA RTX 5060 (8 GB VRAM), the
 89 Adam optimizer (learning rate 2×10^{-3} , gradient clipping at norm 10), and aug-
 90 mented Lagrangian parameters $\rho_{\text{init}} = 1.0$, $\rho_{\max} = 10^6$, $\gamma = 10.0$. The full replication
 91 package is available at <https://github.com/sgao-academics/bayes-lowrank>, following
 92 reproducibility guidelines [27].

93 2.2 Synthetic benchmarks: calibration at scale

94 Erdős–Rényi DAGs (expected degree 3, edge weights uniformly drawn from
 95 $[-1.0, -0.5] \cup [0.5, 1.0]$) were generated with seed 42 across six dimensions. Table 1
 96 reports results. At $d = 30$, the ensemble ECE (0.036) is comparable to DiBS (0.038).
 97 ECE decreases monotonically with dimension, reaching 0.003 at $d = 500$, where DiBS
 98 cannot operate. Per-model runtime remains under 30 seconds across all dimensions.
 99 The spectral rank criterion used here selects conservative ranks; rank sensitivity exper-
 100 iments across three dimensions find that larger ranks improve F1 without degrading
 101 ECE (Supplementary Table S7), suggesting the 75% eigenvalue threshold may be
 102 overly conservative. The ensemble size B was reduced at higher dimensions (50 at
 103 $d = 30$ to 20 at $d = 500$) to keep total runtime under 2.5 hours on a single consumer
 104 GPU; the B -ablation at $d = 30$ confirms that calibration stabilizes from $B = 10$
 105 onward.

106 The ECE reduction at higher dimensions is partly attributable to increasing model
 107 conservatism. The fraction of candidate edges in the lowest probability bin $[0, 0.1)$

Table 1 Synthetic benchmark results. ECE = Expected Calibration Error (10 bins). F1 computed on ensemble mean adjacency matrix ($\tau = 0.3$). High-conf. = edges with $P > 0.95$. Time per model in seconds.

d	r	B	ECE	F1	High-conf.	Time (s)
30	8	50	0.036	0.109	0	16
50	12	45	0.023	0.025	0	16
100	16	40	0.011	0.027	0	15
200	24	25	0.006	0.080	0	13
300	32	25	0.004	0.038	0	16
500	32	20	0.003	0.020	0	28

increases from 91.4% at $d = 30$ to 99.8% at $d = 500$ (Supplementary Table S10). At $d = 500$, ECE is dominated by a single bin because nearly all edges receive near-zero probabilities. This reflects the inherently difficult edge recovery problem at scale: F1 ranges from 0.020 to 0.109, with no edge attaining $P > 0.95$ in any setting. In a $d = 500$ system, the vast majority of the $d(d - 1) = 249,500$ candidate directed edge positions are absent in the true DAG (expected degree 3 yields $\sim 1,500$ true edges, 0.6% of candidates). The ensemble correctly assigns negligible probability to absent edges while elevating the small fraction that receive consistent support across bootstrap draws.

Ablation on the number of ensemble samples (B) at $d = 30$ shows that Bayesian bootstrap ECE converges in calibration by $B = 10$: $\text{ECE}_{B=10} = 0.036$, $\text{ECE}_{B=20} = 0.037$, and $\text{ECE}_{B=50} = 0.036$. Standard bootstrap achieves comparable calibration ECE at all tested B ($\text{ECE}_{B=10} = 0.036$, $\text{ECE}_{B=20} = 0.038$, $\text{ECE}_{B=50} = 0.036$). Bayesian bootstrap F1 fluctuates with B (§ S5) but calibration is the primary metric of interest for ensemble confidence estimation. The edge threshold τ was varied across $\{0.1, 0.2, 0.3, 0.4, 0.5\}$; the monotonic relationship between estimated probability and STRING validation precision was preserved across all thresholds.

2.3 TCGA BRCA: edge probabilities validate biological signal

BRCA expression data (HiSeqV2 RSEM normalized, UCSC Xena [16]) were preprocessed by retaining the top $d = 500$ genes by variance, followed by standardization. STRING v12 protein-protein associations (combined score > 900) [13] provided external validation, following established practice for evaluating network inference on transcriptomic data [26, 28] and causal discovery in biological systems [35].

On BRCA ($n = 1,218$, $B = 30$, ≈ 15 min runtime), edge probabilities are predominantly near zero (mean < 0.001). However, edges binned by estimated probability show a monotonic relationship with STRING validation precision (Figure 1d, Table 2):

- $0.0 < P \leq 0.1$ (249,479 edges): STRING precision = 10.6%.
- $0.1 < P \leq 0.2$ (17 edges): STRING precision = 70.6% (6.7 $\times$ enrichment).
- $0.2 < P \leq 0.3$ (4 edges): STRING precision = 75.0%.

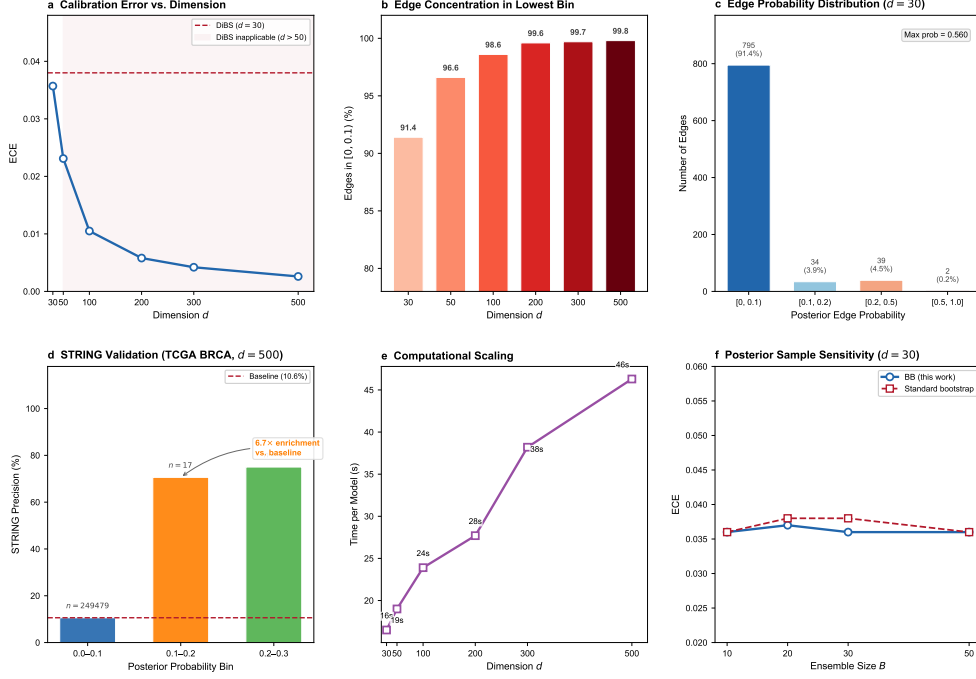


Fig. 1 Bayesian bootstrap low-rank causal discovery at scale. **(a)** Expected Calibration Error (ECE) vs. dimension. ECE decreases from 0.036 ($d = 30$, comparable to DiBS at 0.038) to 0.003 ($d = 500$). Shaded area indicates $d > 50$ where DiBS is inapplicable. **(b)** Fraction of candidate edges in the lowest probability bin $[0, 0.1]$. The increase from 91.4% to 99.8% with dimension indicates growing model conservatism. **(c)** Edge probability distribution at $d = 30$. **(d)** STRING validation precision by edge probability bin on TCGA BRCA ($d = 500$). Baseline = 10.6% for edges with $P < 0.1$. **(e)** Per-model training time vs. dimension on a consumer GPU (RTX 5060, 8 GB). **(f)** Effect of ensemble size B on ECE at $d = 30$. Both Bayesian and standard bootstrap converge by $B = 10$.

Table 2 STRING validation by edge probability bin on TCGA BRCA ($d = 500$).

Edge Probability	Edges	STRING Hits	Precision (%)
0.0–0.1	249,479	1,058	10.6
0.1–0.2	17	12	70.6
0.2–0.3	4	3	75.0

137 No edges exceed $P = 0.3$, consistent with the difficulty of recovering individual
138 regulatory edges from purely observational transcriptomic data. The 6.7-fold enrich-
139 ment from baseline to the moderate-confidence bin (70.6% at $0.1 < P \leq 0.2$, 75.0%
140 at $0.2 < P \leq 0.3$) shows that edges consistently identified across bootstrap draws are
141 substantially more likely to represent real protein interactions.

142 2.4 Practical implications

143 The Bayesian bootstrap ensemble provides a graded confidence scale for experimental
144 prioritization. On TCGA BRCA, validating the 21 edges with $P > 0.1$ achieves 71.4%
145 STRING confirmation, versus 10.6% when testing random edges among 249,500 can-
146 didates. In practice, a researcher would run the ensemble once (≈ 15 minutes on a
147 consumer GPU), then rank edges by estimated probability and validate in descending
148 order until the experimental budget is exhausted. The pipeline—from raw expression
149 data to edge probabilities—runs on a consumer GPU and requires no hyperparameter
150 tuning beyond standard NOTEARS parameters.

151 Bayesian bootstrap low-rank inference is the only uncertainty-aware method that
152 scales beyond $d = 100$ while providing calibrated ensemble edge probabilities (Sup-
153 plementary Table S12). The Dirichlet-weighted bootstrap framework applies to any
154 domain where the true adjacency matrix admits low-rank structure, such as financial
155 networks and neuroscience.

156 3 Limitations

157 First, the Bayesian bootstrap inherits the linear SEM assumption from NOTEARS;
158 nonlinear causal relationships [34] are not captured. Extending the framework to non-
159 linear settings would require replacing the matrix-exponential DAG constraint with
160 a nonlinear acyclicity characterization (e.g., GraNDAG) and a fundamentally differ-
161 ent low-rank structure, which is a direction for future work. Second, the low-rank
162 constraint may be violated in dense, unstructured causal systems—the biological jus-
163 tification of transcription factor-driven regulation is plausible but not universal [22].
164 The spectral rank selection criterion (75% of maximum eigenvalue) yields conserva-
165 tive ranks that are suboptimal for F1; rank sensitivity experiments (Supplementary
166 Table S7) show that larger ranks consistently improve edge recovery without degrading
167 calibration. Third, STRING provides undirected protein–protein interaction valida-
168 tion that does not confirm causal directionality; the reported enrichment is based on
169 17 edges in the moderate-confidence bin and should be interpreted with appropri-
170 ate caution at small sample sizes. Fourth, ECE at high dimensions primarily reflects
171 conservatism rather than uniformly improving calibration; adaptive binning would
172 provide more discriminative assessment. Fifth, the fixed $\tau = 0.3$ threshold may not
173 be optimal across all settings; adaptive thresholding based on the posterior distribu-
174 tion warrants investigation [17]. Finally, while the Bayesian bootstrap [23] and the
175 weighted likelihood bootstrap [24] provide rigorous asymptotic connections between
176 weighted resampling and posterior inference, the finite-sample convergence properties
177 under the non-convex low-rank DAG constraint have not been formally characterized.

178 List of abbreviations

179 DAG: Directed Acyclic Graph; ECE: Expected Calibration Error; SEM: Struc-
180 tural Equation Model; STRING: Search Tool for the Retrieval of Interacting
181 Genes/Proteins; TCGA: The Cancer Genome Atlas; BRCA: Breast Invasive Carci-
182 noma.

183 **Declarations**

184 **Ethics approval and consent to participate**

185 Not applicable. This study uses exclusively publicly available, de-identified datasets.

186 **Consent for publication**

187 Not applicable.

188 **Availability of data and materials**

189 The TCGA BRCA HiSeqV2 RSEM-normalized expression dataset is publicly avail-
190 able from the UCSC Xena platform ([https://xenabrowser.net/datapages/?dataset=](https://xenabrowser.net/datapages/?dataset=TCGA.BRCA.sampleMap/HiSeqV2)
191 [TCGA.BRCA.sampleMap/HiSeqV2](https://xenabrowser.net/datapages/?dataset=TCGA.BRCA.sampleMap/HiSeqV2)). STRING v12 protein-protein association data
192 are publicly available from <https://string-db.org/>. All synthetic datasets analyzed in
193 this study are generated by scripts included in the replication package. The complete
194 replication package—including source code, experiment scripts, trained model check-
195 points, and all figure-generation scripts—is available on GitHub at [https://github.](https://github.com/sgao-academics/bayes-lowrank)
196 [com/sgao-academics/bayes-lowrank](https://github.com/sgao-academics/bayes-lowrank).

197 **Competing interests**

198 The author declares no competing interests.

199 **Funding**

200 This research received no specific grant from any funding agency. Computing resources
201 were purchased by the author personally.

202 **Authors' contributions**

203 S.G. is the sole author and contributed to all aspects of this work: Conceptualiza-
204 tion, Methodology, Software, Validation, Formal analysis, Investigation, Data curation,
205 Writing—original draft, Writing—review and editing, Visualization, and Resources.

206 **References**

- 207 [1] X. Zheng, B. Aragam, P. Ravikumar, and E. P. Xing. DAGs with NO TEARS:
208 Continuous optimization for structure learning. *Advances in Neural Information*
209 *Processing Systems*, 31, 2018.
- 210 [2] K. Bello, B. Aragam, and P. Ravikumar. DAGMA: Learning DAGs via M-
211 matrices and a log-determinant acyclicity characterization. *Advances in Neural*
212 *Information Processing Systems*, 35, 2022.
- 213 [3] I. Ng, A. Ghassami, and K. Zhang. On the role of sparsity and DAG constraints
214 for learning linear DAGs. *Advances in Neural Information Processing Systems*,
215 33, 2020.

- 216 [4] L. Lorch, J. Rothfuss, B. Schölkopf, and A. Krause. DiBS: Differentiable Bayesian
217 structure learning. *Advances in Neural Information Processing Systems*, 34, 2021.
- 218 [5] T. Deleu, A. Góis, C. Emezue, M. Rankawat, S. Lacoste-Julien, S. Bauer, and
219 Y. Bengio. Bayesian structure learning with generative flow networks. *Uncertainty*
220 *in Artificial Intelligence*, 2022.
- 221 [6] C. Cundy, A. Grover, and S. Ermon. BCD Nets: Scalable variational approaches
222 for Bayesian causal discovery. *Advances in Neural Information Processing Sys-*
223 *tems*, 34, 2021.
- 224 [7] L. Lorch, S. Sussex, J. Rothfuss, A. Krause, and B. Schölkopf. Amortized inference
225 for causal structure learning. *Advances in Neural Information Processing Systems*,
226 35, 2022.
- 227 [8] Y. Annadani, N. Pawlowksi, J. Jennings, S. Bauer, C. Zhang, and W. Gong.
228 BayesDAG: Gradient-based posterior inference for causal discovery. *Advances in*
229 *Neural Information Processing Systems*, 36, 2023.
- 230 [9] S. Gao. SSCAGate: Stratified cluster-aware gating for differentiable causal
231 discovery. *bioRxiv*, 2025.
- 232 [10] S. Gao. Why low-rank causal discovery works: Identifiability, minimax rates, and
233 the capacity bottleneck. *arXiv*, 2026.
- 234 [11] J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, and N. M. Luscombe. A
235 census of human transcription factors: function, expression and evolution. *Nature*
236 *Reviews Genetics*, 10(4):252–263, 2009.
- 237 [12] S. A. Lambert, A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen,
238 J. Taipale, T. R. Hughes, and M. T. Weirauch. The human transcription factors.
239 *Cell*, 172(4):650–665, 2018.
- 240 [13] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas,
241 M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, et al. STRING v12: protein–
242 protein association networks with increased coverage. *Nucleic Acids Research*,
243 49(D1):D605–D612, 2021.
- 244 [14] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman &
245 Hall/CRC, 1994.
- 246 [15] M. P. Naeini, G. Cooper, and M. Hauskrecht. Obtaining well calibrated prob-
247 abilities using Bayesian binning. *AAAI Conference on Artificial Intelligence*,
248 2015.

- [16] M. J. Goldman, B. Craft, M. Hastie, K. Repečka, F. McDade, A. Kamath, A. Banerjee, Y. Luo, D. Rogers, A. N. Brooks, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nature Biotechnology*, 38(6):675–678, 2020.
- [17] S. Gao. A principled estimator of intra-cluster correlation for the stratified causal discovery threshold. *SSRN*, 2026.
- [18] N. Meinshausen and P. Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B*, 72(4):417–473, 2010.
- [19] N. Friedman, M. Goldszmidt, and A. Wyner. Data analysis with Bayesian networks: a bootstrap approach. *Uncertainty in Artificial Intelligence*, 1999.
- [20] M. Kalisch and P. Bühlmann. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research*, 8:613–636, 2007.
- [21] D. M. Chickering. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3:507–554, 2002.
- [22] P. Ravikumar, M. J. Wainwright, G. Raskutti, and B. Yu. High-dimensional covariance estimation by minimizing ℓ_1 -penalized log-determinant divergence. *Electronic Journal of Statistics*, 4:935–980, 2010.
- [23] D. B. Rubin. The Bayesian bootstrap. *Annals of Statistics*, 9(1):130–134, 1981.
- [24] M. A. Newton and A. E. Raftery. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society: Series B*, 56(1):3–48, 1994.
- [25] M. Scutari and R. Nagarajan. Identifying significant edges in graphical models of molecular networks. *Artificial Intelligence in Medicine*, 57(3):207–217, 2013.
- [26] J. Z. Zhang and C. Wang. A comparative study of clustering methods on gene expression data for lung cancer prognosis. *BMC Research Notes*, 16, 2023.
- [27] M. Topor, P. Bonnet, V. Cheplygina, V. H. Dam, and L. Wildgaard. A grassroots approach to peer support by the Danish Reproducibility Network. *BMC Research Notes*, 17, 2024.
- [28] M. Y. Pechanec and M. J. Mienaltowski. Decoding the transcriptomic expression and genomic methylation patterns in the tendon proper and its peritenon region in the aging horse. *BMC Research Notes*, 16, 2023.
- [29] C. Glymour, K. Zhang, and P. Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, 10:524, 2019.

- 283 [30] J. Peters, D. Janzing, and B. Schölkopf. *Elements of Causal Inference: Founda-*
284 *tions and Learning Algorithms*. MIT Press, 2017.
- 285 [31] A. C. Davison and D. V. Hinkley. *Bootstrap Methods and their Application*.
286 Cambridge University Press, 1997.
- 287 [32] S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen. A linear non-Gaussian
288 acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003–
289 2030, 2006.
- 290 [33] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the*
291 *Royal Statistical Society: Series B*, 58(1):267–288, 1996.
- 292 [34] P. O. Hoyer, D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf. Nonlinear causal
293 discovery with additive noise models. *Advances in Neural Information Processing*
294 *Systems*, 21, 2008.
- 295 [35] K. Sachs, O. Perez, D. Pe’er, D. A. Lauffenburger, and G. P. Nolan. Causal
296 protein-signaling networks derived from multiparameter single-cell data. *Science*,
297 308(5721):523–529, 2005.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementary.pdf](#)