ETH-UCPH-TUM Workshop

2022-10

SORT'N'REGRESS

Beware of the DAG!

Alexander Reisach
Université Paris Cité

Christof Seiler
Maastricht University

Sebastian Weichwald
Copenhagen University

ETH-UCPH-TUM Workshop
2022-10

SORT'N'REGRESS

Beware of the Simulated DAG!



Alexander Reisach
Université Paris Cité

Christof Seiler
Maastricht University

Sebastian Weichwald
Copenhagen University

$$X = W^\top X + N$$

$$X = W^\top X + N$$

- **MEC:** faithfulness $+$ Markov

Chickering, 1995; Spirtes et al., 2000

$$X = W^\top X + N$$

- **MEC:** faithfulness + Markov

  Chickering, 1995; Spirtes et al., 2000

- **DAG:** LiNGAM + faithfulness + sufficiency

  Shimizu et al., 2006

$$X = W^\top X + N$$

- **MEC:** faithfulness + Markov

  Chickering, 1995; Spirtes et al., 2000

- **DAG:** LiNGAM + faithfulness + sufficiency

  Shimizu et al., 2006

- **DAG:** equal error variance

  Peters & Bühlmann, 2014; Loh & Bühlmann, 2014

$$X = W^\top X + N$$

- **MEC:** faithfulness + Markov

  Chickering, 1995; Spirtes et al., 2000
- **DAG:** LiNGAM + faithfulness + sufficiency

  Shimizu et al., 2006
- **DAG:** equal error variance

  Peters & Bühlmann, 2014; Loh & Bühlmann, 2014
- **DAG:** regularity conditions on error variances along causal order

  Hoyer et al., 2009; Ghoshal and Honorio, 2017, 2018; Chen et al., 2019; Park 2020

$$X = W^\top X + N$$

- **MEC:** faithfulness + Markov

  Chickering, 1995; Spirtes et al., 2000
- **DAG:** LiNGAM + faithfulness + sufficiency

  Shimizu et al., 2006
- **DAG:** equal error variance

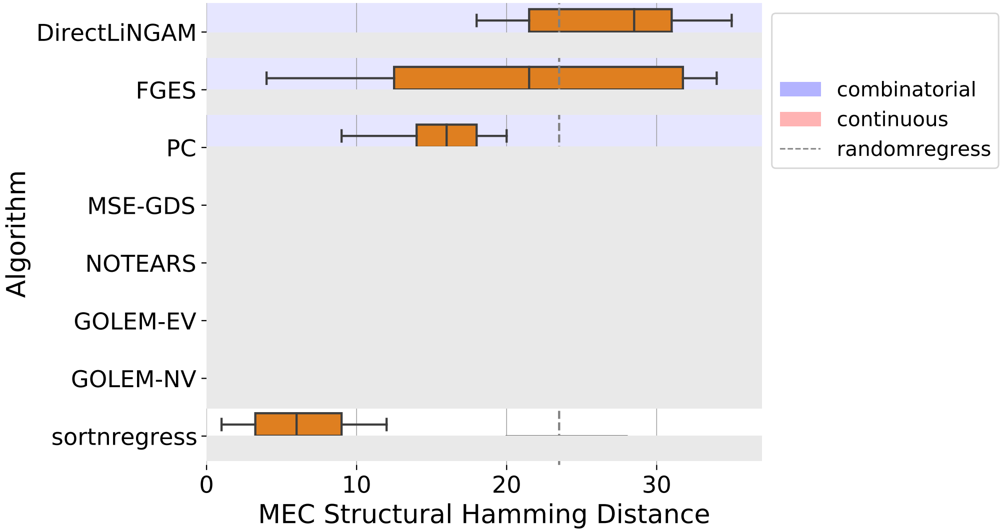
  Peters & Bühlmann, 2014; Loh & Bühlmann, 2014
- **DAG:** regularity conditions on error variances along causal order

  Hoyer et al., 2009; Ghoshal and Honorio, 2017, 2018; Chen et al., 2019; Park 2020
- ... ...

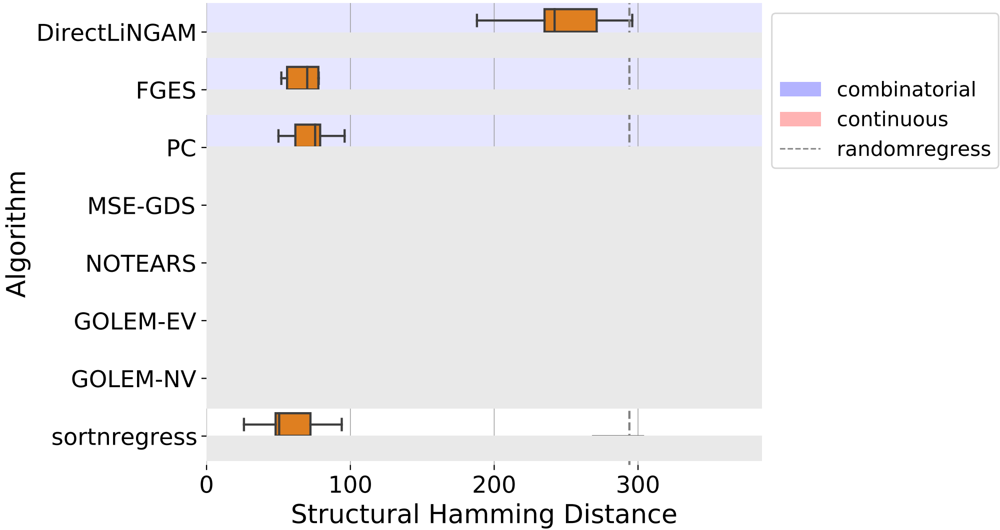
sort variables by increasing marginal variance

’n’

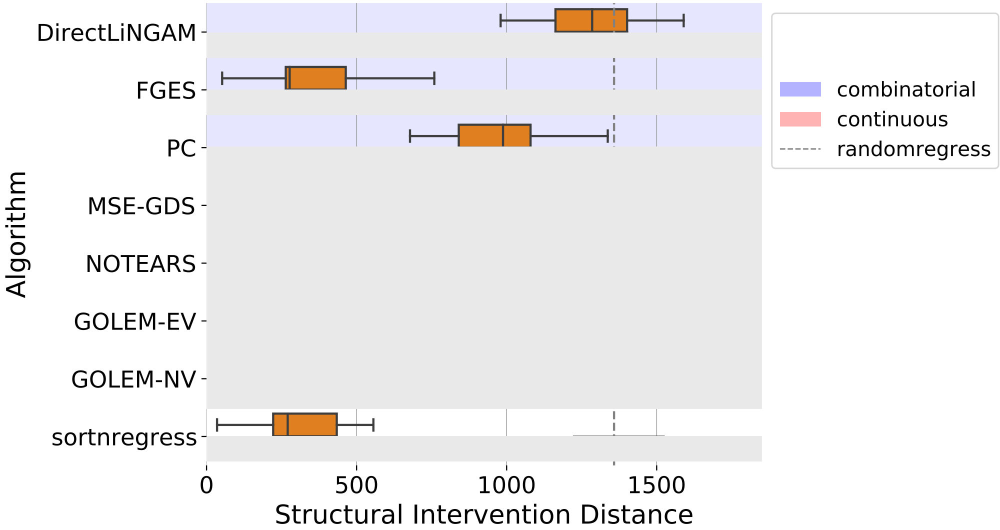
regress, sparsely, each variable onto its predecessors



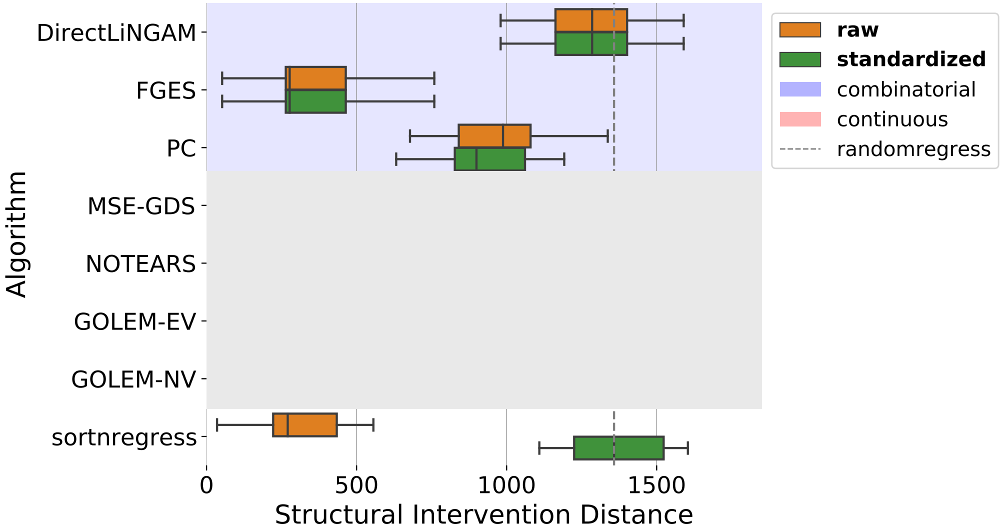
$d = 10$, ER-2, Gauss-NV



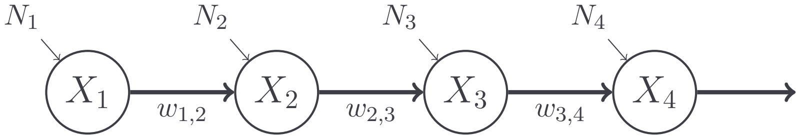
$d = 50,$ ER-2, Gauss-NV

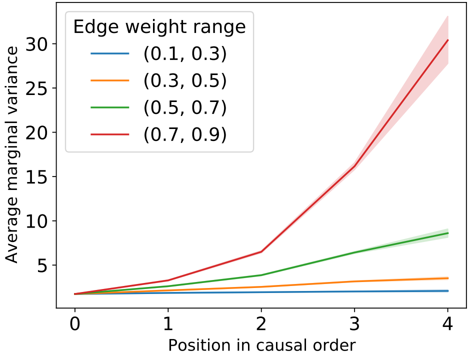


$d = 50,$ ER-2, Gauss-NV



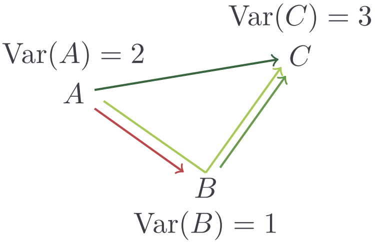
$d = 50,$ ER-2, Gauss-NV


N_1
N_2
N_3
N_4
X_1
X_2
X_3
X_4
$w_{1,2}$
$w_{2,3}$
$w_{3,4}$


Edge weight range
(0.1, 0.3)
(0.3, 0.5)
(0.5, 0.7)
(0.7, 0.9)
Average marginal variance
5
10
15
20
25
30
0
1
2
3
4
Position in causal order

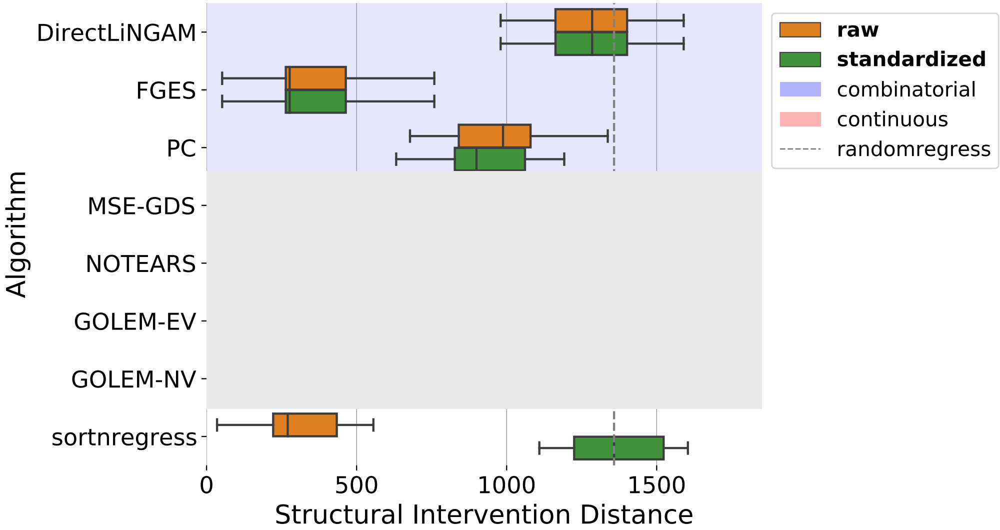
$$\text{varsortability} := \frac{\#\ \text{directed paths from lower to higher variance node}}{\#\ \text{directed paths}}$$

$$v = \frac{1+1+1}{1+1+1+1} = \frac{3}{4}$$



graph	noise	varsortability min	mean	max
ER-1	Gauss-EV	0.94	0.97	0.99
	exponential	0.94	0.97	0.99
	gumbel	0.94	0.97	1.00
ER-2	Gauss-EV	0.97	0.99	1.00
	exponential	0.97	0.99	1.00
	gumbel	0.98	0.99	0.99
ER-4	Gauss-EV	0.98	0.99	0.99
	exponential	0.98	0.99	0.99
	gumbel	0.98	0.99	0.99
SF-4	Gauss-EV	0.98	1.00	1.00
	exponential	0.98	1.00	1.00
	gumbel	0.98	1.00	1.00

graph	ANM-type	varsortability min	mean	max
ER-1	Additive GP	0.81	0.91	1.00
	GP	0.72	0.86	0.96
	MLP	0.55	0.79	0.96
	Multi Index Model	0.62	0.82	1.00
ER-2	Additive GP	0.79	0.91	0.98
	GP	0.82	0.89	0.97
	MLP	0.46	0.71	0.87
	Multi Index Model	0.65	0.79	0.89
ER-4	Additive GP	0.90	0.95	0.98
	GP	0.74	0.88	0.93
	MLP	0.59	0.72	0.85
	Multi Index Model	0.57	0.73	0.85
SF-4	Additive GP	0.95	0.97	0.99
	GP	0.88	0.94	0.97
	MLP	0.75	0.83	0.93
	Multi Index Model	0.77	0.84	0.97



$d = 50,$ ER-2, Gauss-NV

$$\min_{W} \ \ell(W;X) \qquad \overset{?}{\Longleftrightarrow} \qquad \min_{W} \ \ell(W;X)$$

$$\text{s.t. } G(W) \in DAG \qquad\qquad \text{s.t. } h(W) = 0$$

(combinatorial 😱) (smooth 😎)

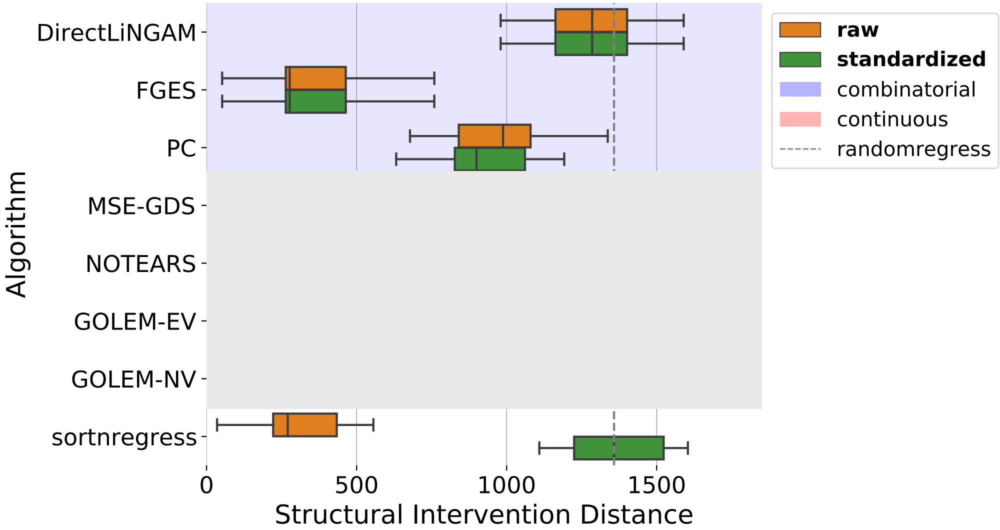
$$h(W) = \mathrm{tr}(\exp(W \odot W)) - d = 0$$

NOTEARS: MSE + sparsity

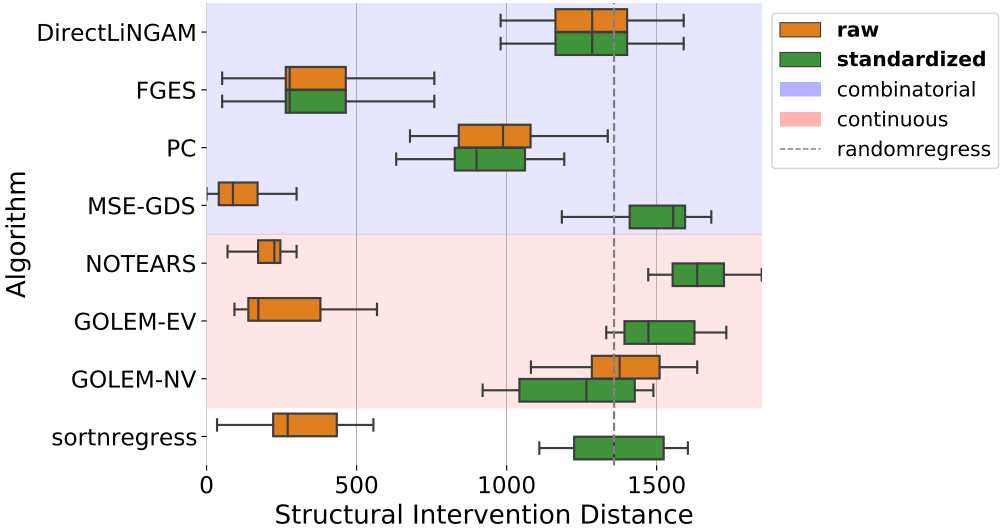
GOLEM: Gaussian likelihood + sparsity

https://blog.ml.cmu.edu/2020/04/10/learning-dags-with-continuous-optimization/

Method	Year	Data	Acycl.	Interv.	Output
CMS [152]	2014	low	-	no	Bi
NO TEARS [267]	2018	low	yes	no	DAG
CGNN [75]	2018	low	yes	no	DAG
Graphite [83]	2019	low/medium	no	no	UG
SAM [122]	2019	low/medium	yes	no	DAG
DAG-GNN [262]	2019	low	yes	no	DAG
GAE [177]	2019	low	yes	no	DAG
NO BEARS [142]	2019	low/medium/high	yes	no	DAG
Meta-Transfer [19]	2019	Bi	yes	yes	Bi
DEAR [214]	2020	high	yes	no	-
CAN [167]	2020	low/medium/high	yes	no	DAG
NO FEARS [251]	2020	low	yes	no	DAG
GOLEM [176]	2020	low	yes	no	DAG
ABIC [20]	2020	low	yes	no	ADMG/PAG
DYNOTEARS [178]	2020	low	yes	no	SVAR
SDI [124]	2020	low	yes	yes	DAG
AEQ [64]	2020	Bi	-	no	direction
RL-BIC [272]	2020	low	yes	no	DAG
CRN [125]	2020	low	yes	yes	DAG
ACD [151]	2020	low	Granger	no	time-series DAG
V-CDN [145]	2020	high	Granger	no	time-series DAG
CASTLE (reg.) [138]	2020	low/medium	yes	no	DAG
GranDAG [139]	2020	low	yes	no	DAG
MaskedNN [175]	2020	low	yes	no	DAG
CausalVAE [257]	2020	high	yes	yes	DAG
CAREFL [126]	2020	low	yes	no	DAG / Bi
Varando [244]	2020	low	yes	no	DAG
NO TEARS+ [268]	2020	low	yes	no	DAG
ICL [250]	2020	low	yes	no	DAG
LEAST [271]	2020	low/medium/high	yes	no	DAG



$d = 50$, ER-2, Gauss-NV



$d = 50$, ER-2, Gauss-NV

standardization is not enough

raw ground-truth model

$$A := N_A$$
$$B := \beta_{A\to B} A + N_B$$
$$C := \beta_{B\to C} B + N_C$$

standardized model

$$A_s := {}^{A}\!/\!\sqrt{\mathrm{Var}(A)}$$
$$B_s := {}^{B}\!/\!\sqrt{\mathrm{Var}(B)}$$
$$C_s := {}^{C}\!/\!\sqrt{\mathrm{Var}(C)}$$

---

observe: either $(X_1, X_2, X_3) = (A, B, C)$ or $(X_1, X_2, X_3) = (C, B, A)$

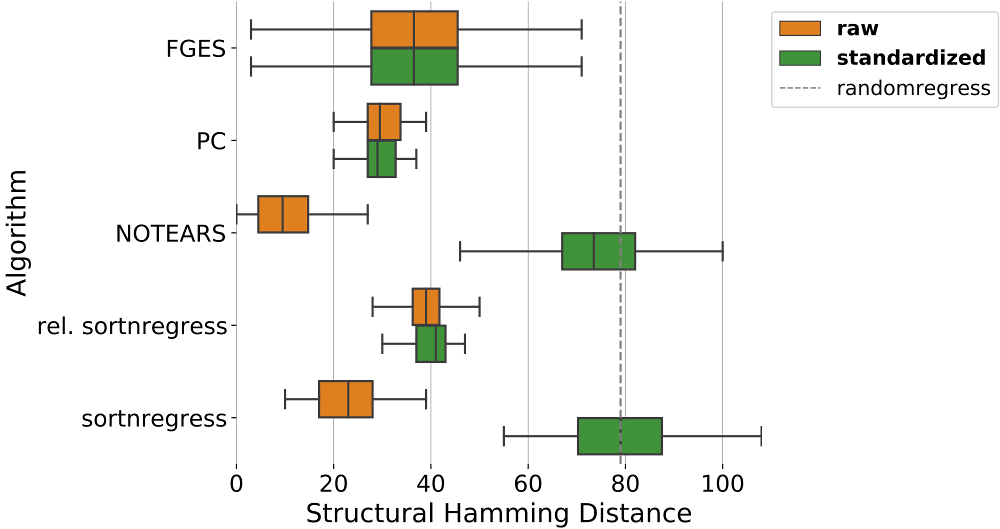
task: $X_1 \to X_2 \to X_3$ or $X_1 \leftarrow X_2 \leftarrow X_3$?

"left-to-right regression coefficients" $|\hat{\beta}_{1\to 2}| \leq |\hat{\beta}_{2\to 3}|$?

"right-to-left regression coefficients" $|\hat{\beta}_{3\to 2}| \leq |\hat{\beta}_{2\to 1}|$?

chain orientation task

d	edge range	accuracy by variance-sorting			accuracy by coefficient-sorting		
		raw	standardized	harmonized	raw	standardized	harmonized
3	$\pm(0.5, 2.0)$	97.50%	50.05%	84.70%	62.58%	73.03%	57.30%
	$\pm(0.5, 0.9)$	80.38%	50.05%	69.62%	57.15%	62.38%	55.65%
	$\pm(0.1, 0.9)$	65.65%	50.30%	60.08%	54.17%	55.88%	53.45%
5	$\pm(0.5, 2.0)$	98.67%	50.15%	82.17%	78.60%	86.58%	64.20%
	$\pm(0.5, 0.9)$	77.65%	49.27%	66.30%	61.83%	68.65%	57.50%
	$\pm(0.1, 0.9)$	63.08%	50.38%	57.65%	58.17%	57.33%	56.35%
10	$\pm(0.5, 2.0)$	99.38%	50.02%	79.30%	93.72%	96.97%	69.08%
	$\pm(0.5, 0.9)$	73.75%	50.25%	62.00%	64.97%	70.70%	58.50%
	$\pm(0.1, 0.9)$	62.55%	51.23%	58.25%	55.85%	56.05%	54.40%



$d = 20,$ ER-2, Gauss-NV

"True" data scale?

iid edge weights?

How to benchmark?

Additive noise models?

Reisach, Seiler, Weichwald (2021). Beware of the Simulated DAG! Causal Discovery Benchmarks May Be Easy To Game. NeurIPS.