Case Studies of Causal Discovery from IT Monitoring Time Series

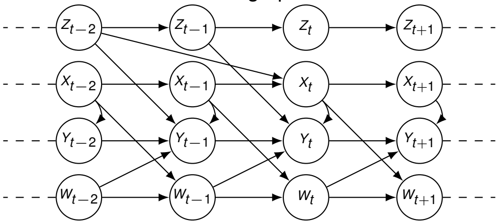
Ali Aït-Bachir, Charles K. Assaad, Christophe de Bignicourt, Emilie Devijver, **Simon Ferreira**, Eric Gaussier, Hosein Mohanna and Lei Zan

Outline

Outline

Temporal causal graphs and abstractions

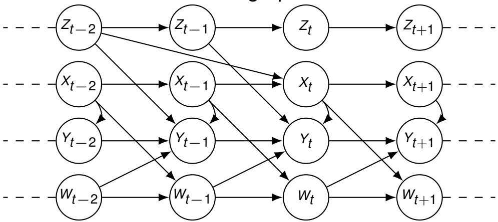
Full time graph: $\mathcal{G}^f$



C. K. Assaad et al. *Survey and evaluation of causal discovery methods for time series*. JAIR, 2022.

Temporal causal graphs and abstractions

Full time graph: $\mathcal{G}^f$



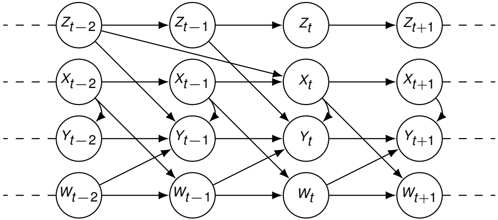
Additional assumption: consistency throughout time

$\gamma_{\max}$: maximal lag between causes and their effets

C. K. Assaad et al. *Survey and evaluation of causal discovery methods for time series*. JAIR, 2022.

Temporal causal graphs and abstractions

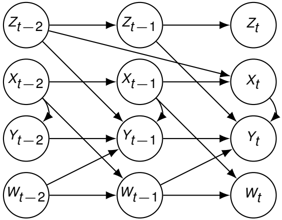
Full time graph: $\mathcal{G}^f$



Additional assumption: consistency throughout time

$\gamma_{\max}$: maximal lag between causes and their effets

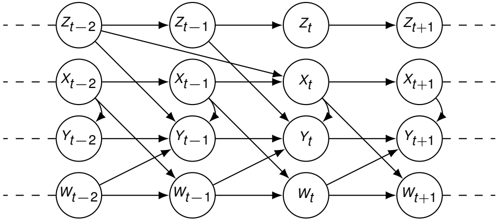
Window graph: $\mathcal{G}^w$



C. K. Assaad et al. *Survey and evaluation of causal discovery methods for time series*. JAIR, 2022.

Temporal causal graphs and abstractions

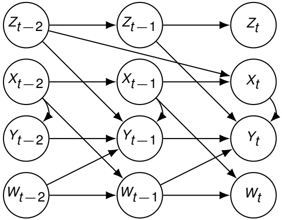
Full time graph: $\mathcal{G}^f$



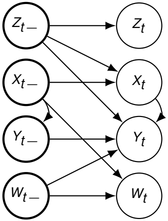
Additional assumption: consistency throughout time

$\gamma_{\max}$: maximal lag between causes and their effets

Window graph: $\mathcal{G}^w$



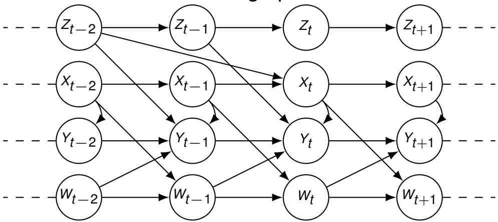
Extended summary graph: $\mathcal{G}^e$



C. K. Assaad et al. *Survey and evaluation of causal discovery methods for time series*. JAIR, 2022.

Temporal causal graphs and abstractions

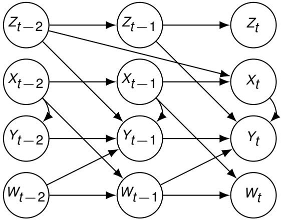
Full time graph: $\mathcal{G}^f$



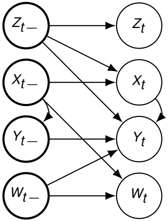
Additional assumption: consistency throughout time

$\gamma_{\max}$: maximal lag between causes and their effets

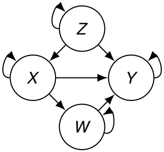
Window graph: $\mathcal{G}^w$



Extended summary graph: $\mathcal{G}^e$



Summary graph (SCG): $\mathcal{G}^s$



C. K. Assaad et al. *Survey and evaluation of causal discovery methods for time series*. JAIR, 2022.

Additional assumption on causal relations

- Stationary and linear dynamic structural causal model

$$\forall Y_t \in \mathcal{V}_w$$

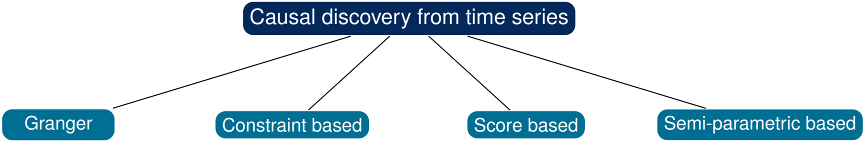
$$Y_t = \sum_{X_{t-\gamma_{xy}} \in Parents(Y_t, \mathcal{G}_w)} \alpha_{XY\gamma_{xy}} X_{t-\gamma_{xy}} + \xi_t^y,$$

where $\alpha_{XY\gamma_{xy}} \neq 0$ is the direct effect of $X_{t-\gamma_{xy}}$ on Y_t

Rajen D. Shah, Jonas Peters. *The hardness of conditional independence testing and the generalised covariance measure.* Ann. Statist. 2020.

Outline

Causal discovery from time series



A. Arnold et al. *Temporal causal modeling with graphical granger method*. KDD, 2007.

J. Runge. *Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets*. UAI, 2020.

C. K. Assaad et al. *Discovery of extended summary graphs in time series*. UAI. PMLR. 2022.

R. Pamfil et al. *Dynotears: Structure learning from time-series data*. AISTATS, 2020.

A. Hyvärinen et al. *Causal modelling combining instantaneous and lagged effects: An identifiable model based on non-gaussianity*. ICML, 2010.

C. K. Assaad et al. *Hybrids of constraint-based and noise-based algorithms for causal discovery from time series*. Arxiv, 2023.

Causal discovery from time series

- Causal discovery from time series
  - Granger
    - GCMVL (Arnold, 2007)
  - Constraint based
    - PCMCI^{+} (Runge, 2020)
    - PCGCE (Assaad, 2022)
  - Score based
    - Dynotears (Pamfil, 2020)
  - Semi-parametric based
    - VLiNGAM (Hyvärinen, 2010)
    - TiMINo (Peters, 2010)

A. Arnold et al. *Temporal causal modeling with graphical granger method*. KDD, 2007.

J. Runge. *Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets*. UAI, 2020.

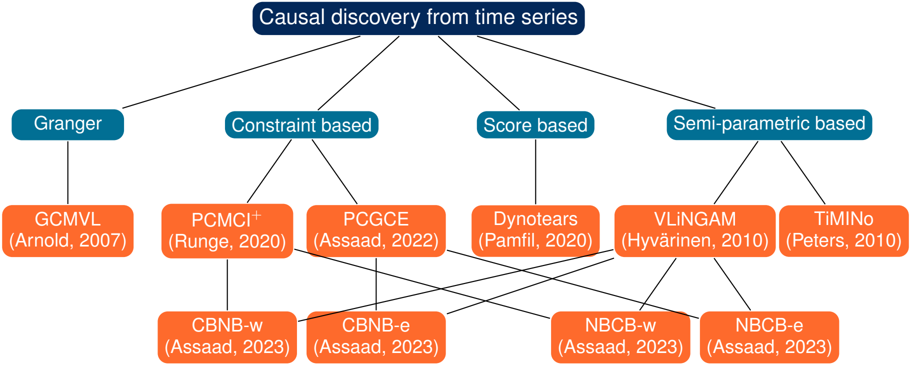
C. K. Assaad et al. *Discovery of extended summary graphs in time series*. UAI. PMLR. 2022.

R. Pamfil et al. *Dynotears: Structure learning from time-series data*. AISTATS, 2020.

A. Hyvärinen et al. *Causal modelling combining instantaneous and lagged effects: An identifiable model based on non-gaussianity*. ICML, 2010.

C. K. Assaad et al. *Hybrids of constraint-based and noise-based algorithms for causal discovery from time series*. Arxiv, 2023.

Causal discovery from time series



A. Arnold et al. *Temporal causal modeling with graphical granger method*. KDD, 2007.

J. Runge. *Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets*. UAI, 2020.

C. K. Assaad et al. *Discovery of extended summary graphs in time series*. UAI. PMLR. 2022.

R. Pamfil et al. *Dynotears: Structure learning from time-series data*. AISTATS, 2020.

A. Hyvärinen et al. *Causal modelling combining instantaneous and lagged effects: An identifiable model based on non-gaussianity*. ICML, 2010.

C. K. Assaad et al. *Hybrids of constraint-based and noise-based algorithms for causal discovery from time series*. Arxiv, 2023.

Summary of the main characteristics

Family	Algorithm	Causal graph	Instantaneous relations	Faithfulness	Non-Gaussianity	Equal noise variances
Granger	GCMVL	SCG	✗	✗	✗	✗
Constraint-based	PCMCI^{+}	WCG	✓	✓	✗	✗
	PCGCE	ECG	✓	✓	✗	✗
Score-based	DYNOTEARS	WCG	✓	✗	✗	✓
Semi-parametric-based	VarLiNGAM	WCG	✓	✗	✓	✗
	TiMINo	SCG*	✓	✗	✓	✗
Hybrid-based	CBNB-w	WCG	✓	✗	✓	✗
	CBNB-e	ECG	✓	✗	✓	✗
	NBCB-w	WCG	✓	✗	✓	✗
	NBCB-e	ECG	✓	✗	✓	✗

Outline

Pre-Processing

Alignment issues and missing values:

- Strategy 1: Re-sample according to the lowest sampling rate by taking the closest value.

Pre-Processing

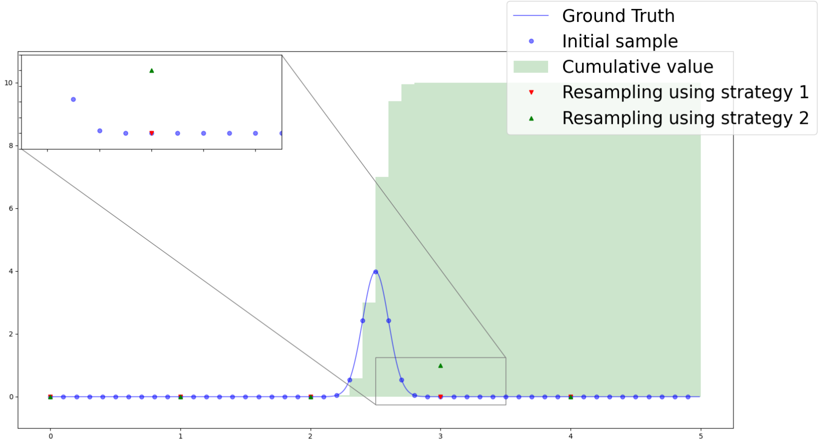
Alignment issues and missing values:

- Strategy 1: Re-sample according to the lowest sampling rate by taking the closest value.
- Strategy 2: Re-sample according to the lowest sampling rate by iteratively interpolating each value from the cumulative values.

Pre-Processing

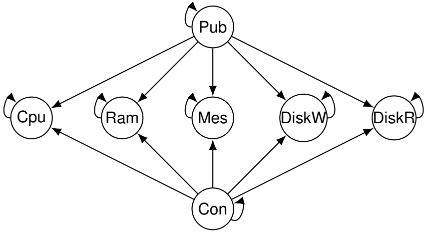
Alignment issues and missing values:

- Strategy 1: Re-sample according to the lowest sampling rate by taking the closest value.
- Strategy 2: Re-sample according to the lowest sampling rate by iteratively interpolating each value from the cumulative values.



The use case of MoM activity

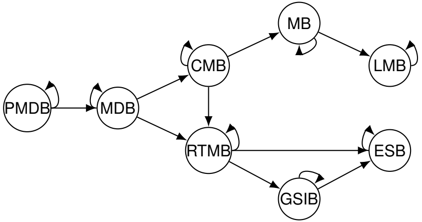
Message ingestion activity based on Publish/Subscribe architecture



Two datasets of 288 and 364 timestamps with a sampling rate of 1 seconde.

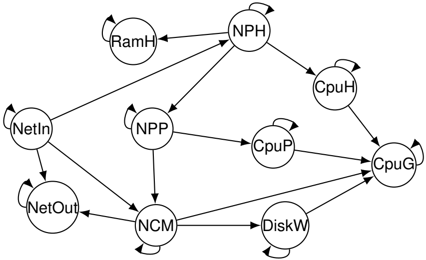
The use case of Ingestion activity

Message ingestion activity based on Storm architecture



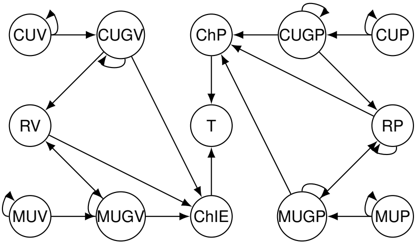
A dataset of 991 timestamps with a sampling rate of 1 minute.

The use case of Web activity



A dataset of 3000 timestamps with a sampling rate of 1 minute.
The data was misaligned so the two pre-processing strategies were used.

The use case of Antivirus activity



A dataset of 1321 timestamps with a sampling rate of 1 and 5 minutes.
The data was misaligned so the two pre-processing strategies were used.

IT monitoring case studies: results

Estimation

- Partial correlation
- ...

Hyper-parameters

- $\gamma_{max} = 15$ for MoM and Ingestion, $\gamma_{max} = 3$ for other datasets
- Significance level $= 0.05$
- ...

Evaluation measure

- F1-score on orientations in the SCG.

IT monitoring case studies: results

Estimation

- Partial correlation
- ...

Hyper-parameters

- $\gamma_{max} = 15$ for MoM and Ingestion, $\gamma_{max} = 3$ for other datasets
- Significance level $= 0.05$
- ...

Evaluation measure

- F1-score on orientations in the SCG.

Results

	MoM 1	MoM 2	Ingestion	Web 1	Web 2	Antivirus 1	Antivirus 2
GCMVL	0.0	0.0	0.2	0.2	0.0	0.08	0.0
Dynotears	0.26	0.2	0.14	0.23	0.3	0.18	0.19
PCMCI^{+}	**0.4**	0.0	0.0	0.23	0.3	0.04	0.11
PCGCE	0.0	0.12	0.12	0.22	0.15	0.3	**0.45**
VLiNGAM	0.0	0.0	0.19	**0.29**	0.18	0.15	0.22
TiMINo	0.0	0.17	0.18	0.0	0.0	0.0	0.0
NBCB-w	**0.4**	0.0	0.13	0.23	0.3	0.14	0.24
NBCB-e	0.13	**0.29**	**0.27**	0.19	**0.42**	**0.31**	**0.45**
CBNB-w	**0.4**	0.0	0.15	0.23	0.3	0.17	0.16
CBNB-e	0	0.24	0.13	0.22	0.29	**0.31**	0.38

IT monitoring case studies: results

Estimation

- Partial correlation
- ...

Hyper-parameters

- $\gamma_{max} = 15$ for MoM and Ingestion, $\gamma_{max} = 3$ for other datasets
- Significance level $= 0.05$
- ...

Evaluation measure

- F1-score on orientations in the SCG.

Results

	MoM 1	MoM 2	Ingestion	Web 1	Web 2	Antivirus 1	Antivirus 2
GCMVL	0.0	0.0	0.2	0.2	0.0	0.08	0.0
Dynotears	0.26	0.2	0.14	0.23	0.3	0.18	0.19
PCMCI$^+$	**0.4**	0.0	0.0	0.23	0.3	0.04	0.11
PCGCE	0.0	0.12	0.12	0.22	0.15	0.3	**0.45**
VLiNGAM	0.0	0.0	0.19	**0.29**	0.18	0.15	0.22
TiMINo	0.0	0.17	0.18	0.0	0.0	0.0	0.0
NBCB-w	**0.4**	0.0	0.13	0.23	0.3	0.14	0.24
NBCB-e	0.13	**0.29**	**0.27**	0.19	**0.42**	**0.31**	**0.45**
CBNB-w	**0.4**	0.0	0.15	0.23	0.3	0.17	0.16
CBNB-e	0	0.24	0.13	0.22	0.29	**0.31**	0.38

Not satisfactory for real applications!

Outline

Conclusion

Causal Discovery relies on strong assumptions which are not verified in real IT monitoring data. Therefore, there is a need for robustness against violation of some assumptions.

Conclusion

Causal Discovery relies on strong assumptions which are not verified in real IT monitoring data. Therefore, there is a need for robustness against violation of some assumptions.

- Faithfulness, NBCB algorithms work best.

Conclusion

Causal Discovery relies on strong assumptions which are not verified in real IT monitoring data. Therefore, there is a need for robustness against violation of some assumptions.

- Faithfulness, NBCB algorithms work best.

Conclusion

Causal Discovery relies on strong assumptions which are not verified in real IT monitoring data. Therefore, there is a need for robustness against violation of some assumptions.

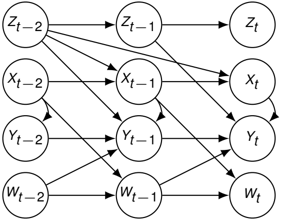
- Faithfulness, NBCB algorithms work best.
- Consistency throughout time. Algorithms looking for the ECG work best.

Conclusion

Causal Discovery relies on strong assumptions which are not verified in real IT monitoring data. Therefore, there is a need for robustness against violation of some assumptions.

- Faithfulness, NBCB algorithms work best.
- Consistency throughout time. Algorithms looking for the ECG work best.

Window graph: $\mathcal{G}^w$

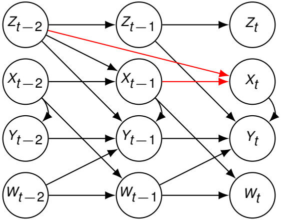


Conclusion

Causal Discovery relies on strong assumptions which are not verified in real IT monitoring data. Therefore, there is a need for robustness against violation of some assumptions.

- Faithfulness, NBCB algorithms work best.
- Consistency throughout time. Algorithms looking for the ECG work best.

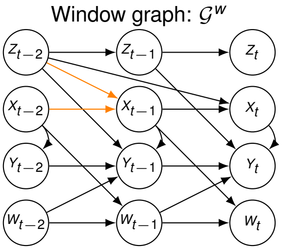
Window graph: $\mathcal{G}^w$



Conclusion

Causal Discovery relies on strong assumptions which are not verified in real IT monitoring data. Therefore, there is a need for robustness against violation of some assumptions.

- Faithfulness, — NBCB algorithms work best.
- Consistency throughout time. — Algorithms looking for the ECG work best.

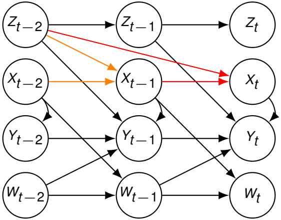


Conclusion

Causal Discovery relies on strong assumptions which are not verified in real IT monitoring data. Therefore, there is a need for robustness against violation of some assumptions.

- Faithfulness, NBCB algorithms work best.
- Consistency throughout time. Algorithms looking for the ECG work best.

Window graph: $\mathcal{G}^w$

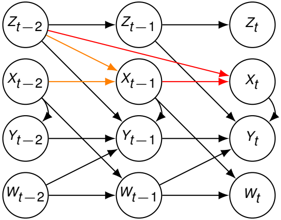


Conclusion

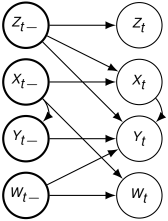
Causal Discovery relies on strong assumptions which are not verified in real IT monitoring data. Therefore, there is a need for robustness against violation of some assumptions.

- Faithfulness, — NBCB algorithms work best.
- Consistency throughout time. — Algorithms looking for the ECG work best.

Window graph: $\mathcal{G}^w$



Extended summary graph: $\mathcal{G}^e$

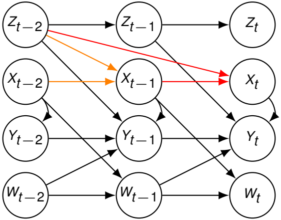


Conclusion

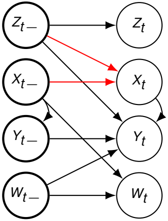
Causal Discovery relies on strong assumptions which are not verified in real IT monitoring data. Therefore, there is a need for robustness against violation of some assumptions.

- Faithfulness, — NBCB algorithms work best.
- Consistency throughout time. — Algorithms looking for the ECG work best.

Window graph: $\mathcal{G}^w$



Extended summary graph: $\mathcal{G}^e$

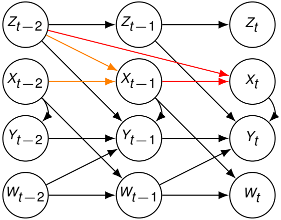


Conclusion

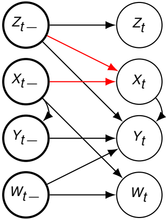
Causal Discovery relies on strong assumptions which are not verified in real IT monitoring data. Therefore, there is a need for robustness against violation of some assumptions.

- Faithfulness, NBCB algorithms work best.
- Consistency throughout time. Algorithms looking for the ECG work best.

Window graph: $\mathcal{G}^w$



Extended summary graph: $\mathcal{G}^e$



$\Longrightarrow$ NBCB-e performs better.

Prespectives

Many things to improve for causal discovery in time series:

- non-stationarity
- regime-change
- measurement errors
- instantaneous cycles

Prespectives

Many things to improve for causal discovery in time series:

- non-stationarity
- regime-change
- measurement errors
- instantaneous cycles
- **Robustness to violation of hypothesis**

Prespectives

Many things to improve for causal discovery in time series:

- non-stationarity
- regime-change
- measurement errors
- instantaneous cycles
- **Robustness to violation of hypothesis**

Thank you for your attention, feel free to ask questions.

References

- C. K. Assaad et al. Survey and evaluation of causal discovery methods for time series. 2022.
- Rajen D. Shah, Jonas Peters. The hardness of conditional independence testing and the generalised covariance measure. Ann. Statist. 2020.
- A. Arnold, Y. Liu, and N. Abe. Tmporal causal modeling with graphical granger method. KDD, 2007.
- Kailash Budhathoki, Dominik Janzing, Patrick Bloebaum, and Hoiyi Ng. Why did the distribution change? AISTATs, 2021.
- A. Hyvärinen, S. Shimizu, and P. Hoyer. Causal modelling combining instantaneous and lagged effects: An identifiable model based on non-gaussianity. ICML, 2010.
- R. Pamfil et al. Dynotears: Structure learning from time-series data. AISTATS, 2020.
- J. Pearl. Causality : Models, Reasoning and Inference. Cambridge University Press, 2009.
- J. Runge. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. UAI, 2020.
- P. Spirtes, C. Glymour, and R. Scheines. Causation, Prediction, and Search. Mit Press, 2000.
- C. K. Assaad et al. Hybrids of constraint-based and noise-based algorithms for causal discovery from time series. Arxiv, 2023.

Collaborators

Ali Aït-Bachir

Charles K. Assaad

Christophe de Bignicourt

Emilie Devijver

Eric Gaussier

Hosein Mohanna

Lei Zan

Thank you !

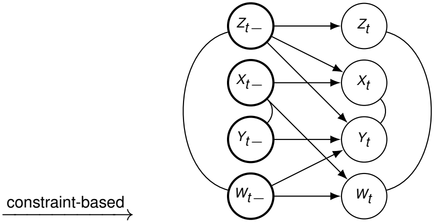
Hybrid-based Methods

CBNB:

Hybrid-based Methods

CBNB:

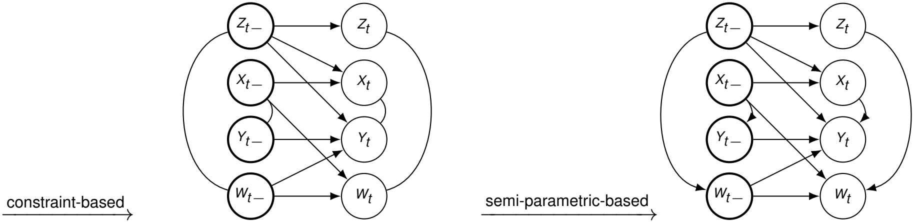
- Step 1: Use the first part of a constraint-based algorithm (PCMCI+ or PCGCE) to discover the skeleton.



Hybrid-based Methods

CBNB:

- Step 1: Use the first part of a constraint-based algorithm (PCMCI+ or PCGCE) to discover the skeleton.
- Step 2: Use a semi-parametric-based algorithm (VarLiNGAM) on relevant vertices to orient the edges.



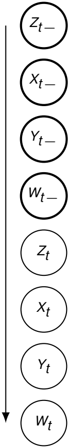
Hybrid-based Methods

NBCB:

Hybrid-based Methods

NBCB:

- Step 1: Use the first part of a semi-parametric-based algorithm (VarLiNGAM) an order in the vertices.



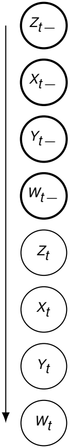
semi-parametric-based

Hybrid-based Methods

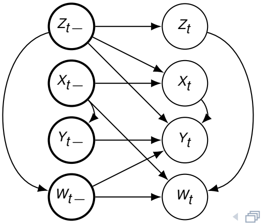
NBCB:

- Step 1: Use the first part of a semi-parametric-based algorithm (VarLiNGAM) an order in the vertices.
- Step 2: Use a constraint-based algorithm (PCMCI+ or PCGCE) with the knowledge of parents to eliminate some remaining edges.

semi-parametric-based $\longrightarrow$



constraint-based $\longrightarrow$



Hybrid-based Methods

CBNB:

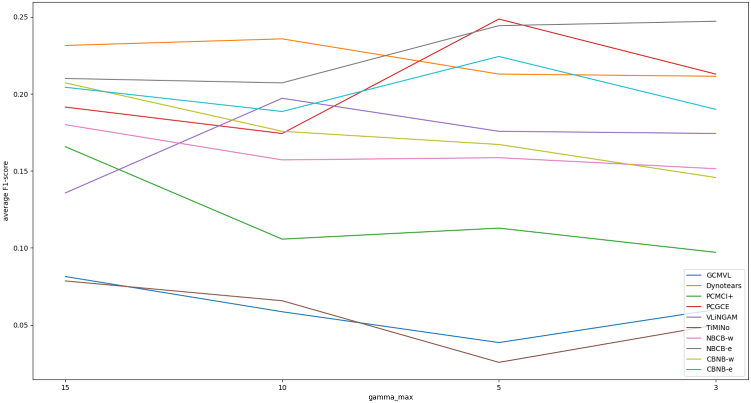
- Step 1: Use the first part of a constraint-based algorithm (PCMCI+ or PCGCE) to discover the skeleton.
- Step 2: Use a semi-parametric-based algorithm (VarLiNGAM) on relevant vertices to orient the edges.

NBCB:

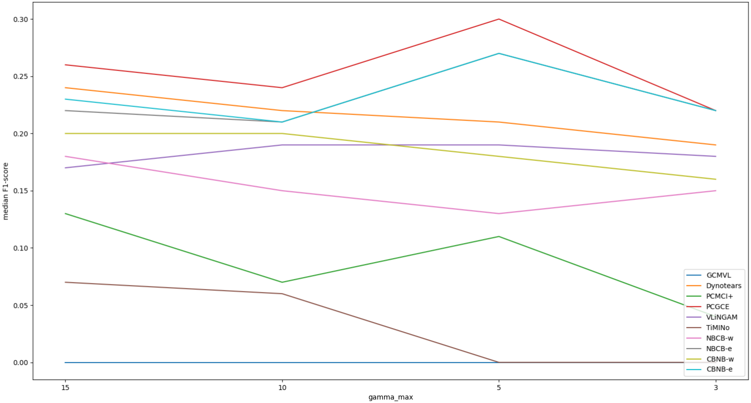
- Step 1: Use the first part of a semi-parametric-based algorithm (VarLiNGAM) an order in the vertices.
- Step 2: Use a constraint-based algorithm (PCMCI+ or PCGCE) with the knowledge of parents to eliminate some remaining edges.

Advantages:
In theory both need the same assumptions. However, CBNB algorithms are robust regarding the semi-parametric assumption and NBCB algorithms are robust regarding the faithfulness assumption.

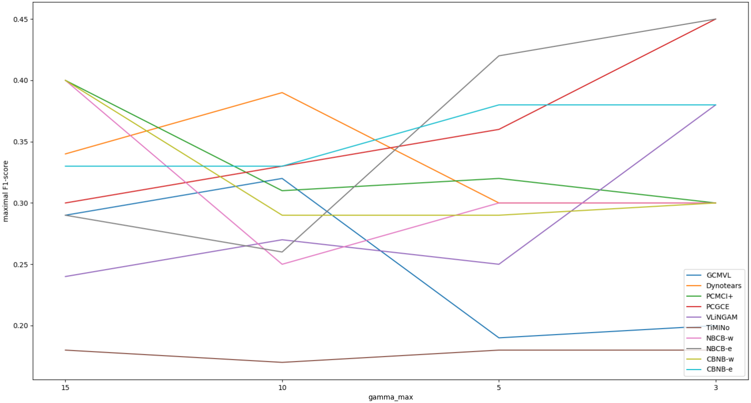
Average F1-Score



Median F1-Score



Maximal F1-Score



Minimal F1-Score