A Constraint Optimization Approach to Causal Discovery from Subsampled Time Series Data

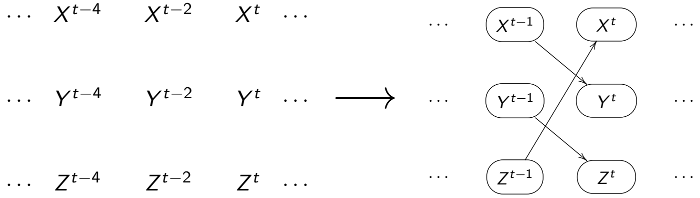
Antti Hyttinen
Joint work with Sergey Plis, Matti Järvisalo,
Frederick Eberhardt, David Danks

University of Helsinki, Helsinki Institute for Information Technology
Mind Research Network and University of New Mexico
California Institute of Technology
Carnegie Mellon University

AMBN 2017, Kyoto, Japan
20.9.2016

Problem Statement

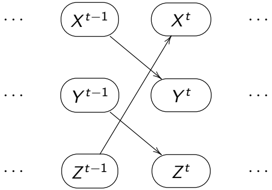
How to discover the causal structure at the **system timescale** from time series data obtained at a coarser **measurement timescale**?



Problem Statement

How to discover the causal structure at the **system timescale** from time series data obtained at a coarser **measurement timescale**?

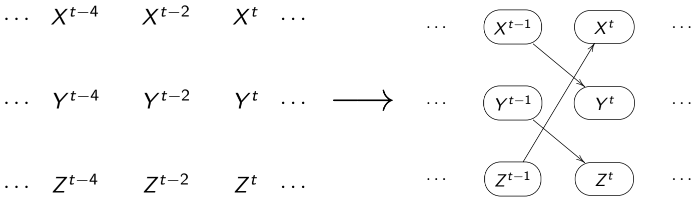
$$\begin{array}{cccccc} \cdots & X^{t-4} & X^{t-2} & X^{t} & \cdots & \\ \cdots & Y^{t-4} & Y^{t-2} & Y^{t} & \cdots & \longrightarrow \\ \cdots & Z^{t-4} & Z^{t-2} & Z^{t} & \cdots & \end{array}$$



- Only every u:th vector of values is observed (**subsampling rate** u)

Problem Statement

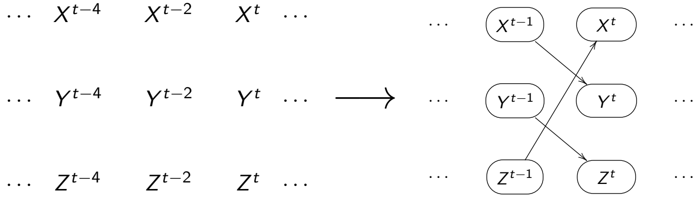
How to discover the causal structure at the **system timescale** from time series data obtained at a coarser **measurement timescale**?



- Only every u:th vector of values is observed (**subsampling rate** u)
- Subsamping induces confounding, and unidentifiability

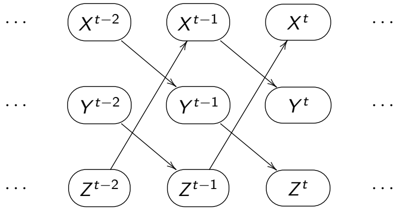
Problem Statement

How to discover the causal structure at the **system timescale** from time series data obtained at a coarser **measurement timescale**?

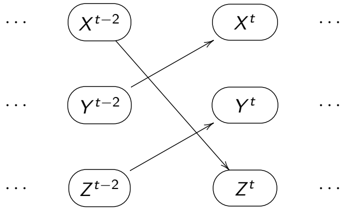


- Only every u:th vector of values is observed (**subsampling rate** u)
- Subsamping induces confounding, and unidentifiability
- Applications: e.g. fMRI.

Subsampling needs to be taken into account!

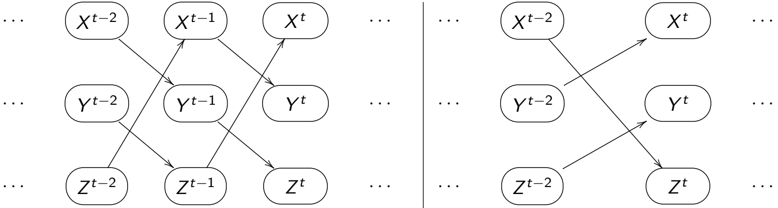


True structure at the system timescale



measurement time scale structure

Subsampling needs to be taken into account!



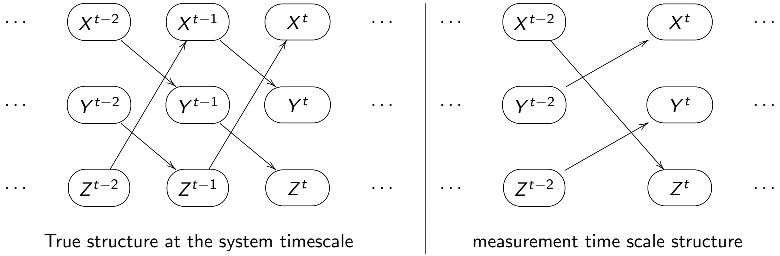
True structure at the system timescale

measurement time scale structure

When ignoring subsampling:

- All direct causal relationships misspecified.

Subsampling needs to be taken into account!



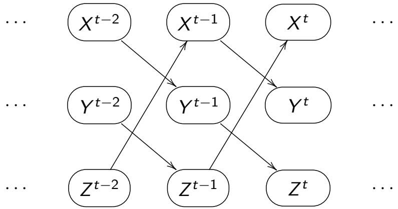
True structure at the system timescale

measurement time scale structure

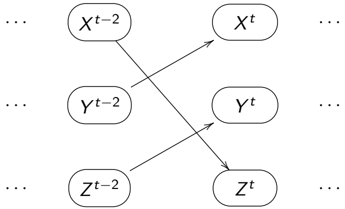
When ignoring subsampling:

- All direct causal relationships misspecified.
- Wrong result for interventions.

Subsampling needs to be taken into account!



True structure at the system timescale



measurement time scale structure

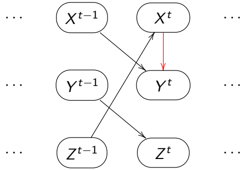
When ignoring subsampling:

- All direct causal relationships misspecified.
- Wrong result for interventions.
- Wrong interventions suggested.

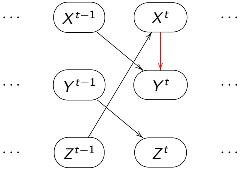
Overview

Previous Literature

- Adding instantaneous effects in a linear model (see for example Lütkepohl 2005 or Hyvärinen et al 2010).



- Adding instantaneous effects in a linear model
(see for example Lütkepohl 2005 or Hyvärinen et al 2010).



- Continuous time approaches, but some processes are inherently discrete time (e.g. salary payment).

Previous Literature 2

Recently Plis et al. (UAI2015,NIPS2015) considered modeling subsampling directly, assuming on the system timescale level:

- discrete time
- first order Markov: $\mathbf{V}^t \perp\!\!\!\perp \mathbf{V}^{t-k} | \mathbf{V}^{t-1}$
- no instantaneous effects, or unobserved common causes
- nonparametric (continuous or discrete values, SVAR processes, or dynamic BNs)
- Measurements from this at integer intervals (e.g. every second).

Previous Literature 2

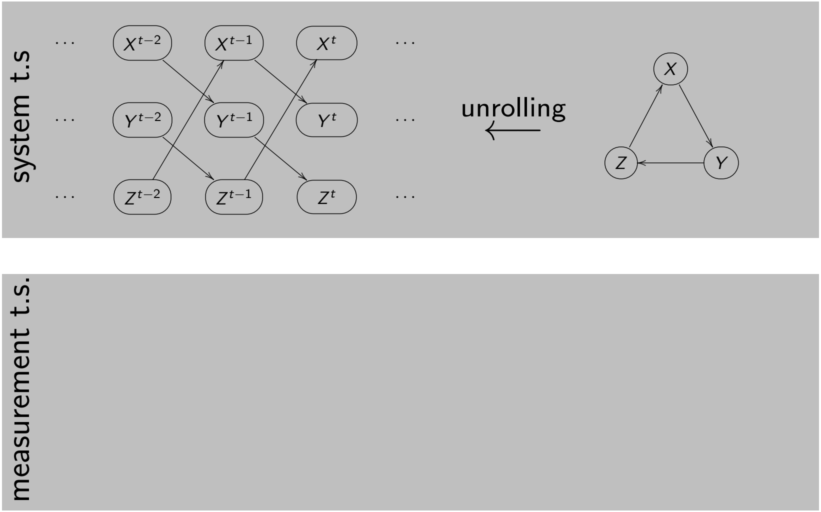
Recently Plis et al. (UAI2015,NIPS2015) considered modeling subsampling directly, assuming on the system timescale level:

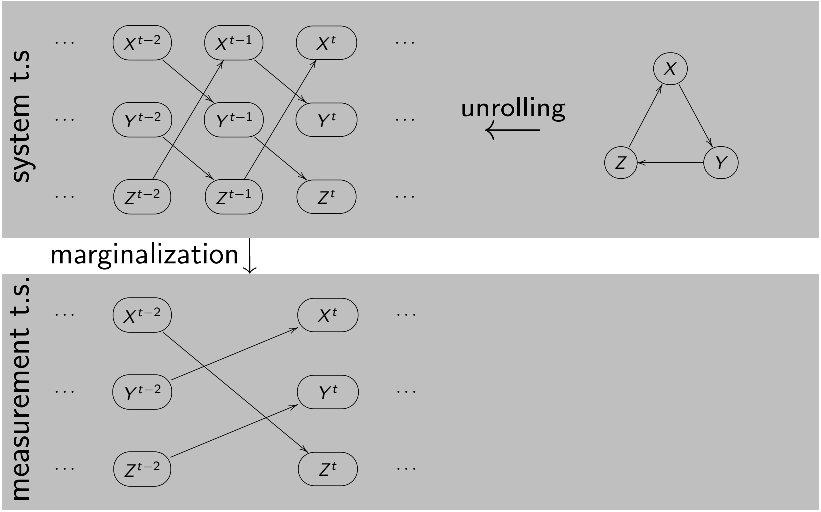
- discrete time
- first order Markov: $\mathbf{V}^t \perp\!\!\!\perp \mathbf{V}^{t-k} | \mathbf{V}^{t-1}$
- no instantaneous effects, or unobserved common causes
- nonparametric (continuous or discrete values, SVAR processes, or dynamic BNs)
- Measurements from this at integer intervals (e.g. every second).

Corresponding parametric method: Gong et al. (ICML2015) discovered linear models using non-Gaussianity.

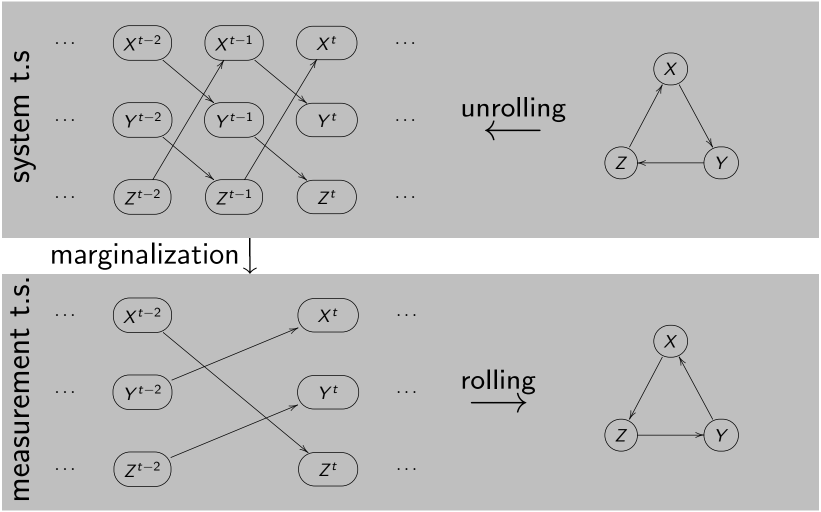
Graphical Representation

Rolled Representation

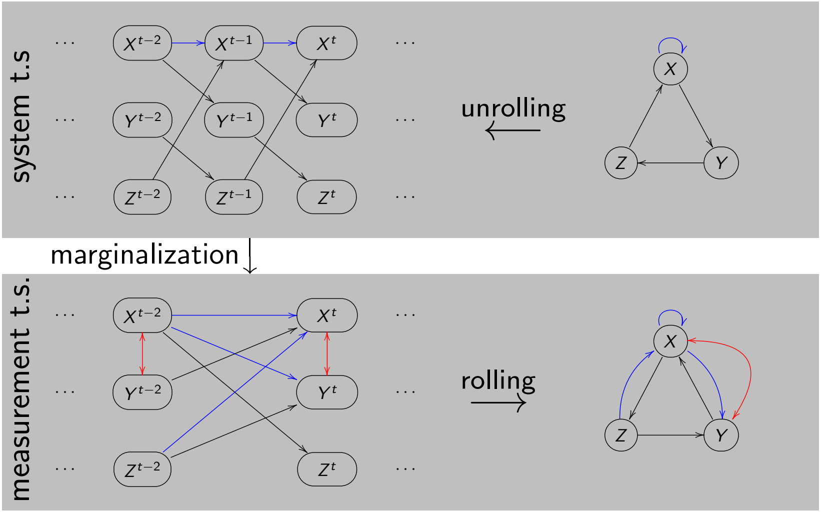



system t.s
...
X^{t-2}
X^{t-1}
X^t
Y^{t-2}
Y^{t-1}
Y^t
Z^{t-2}
Z^{t-1}
Z^t
unrolling
X
Z
Y
marginalization
measurement t.s.
X^{t-2}
X^t
Y^{t-2}
Y^t
Z^{t-2}
Z^t

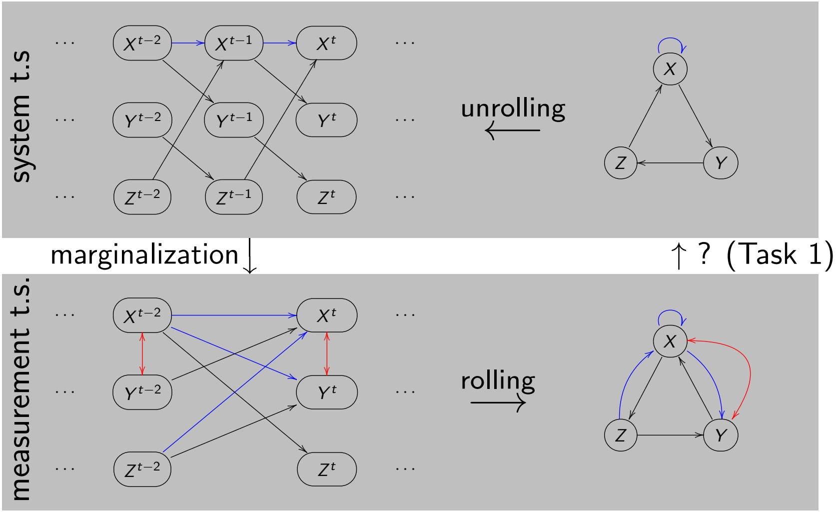
Rolled Representation



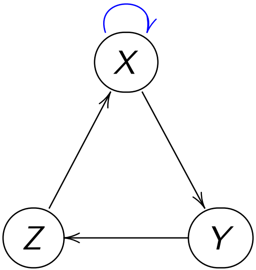
Induced confounding



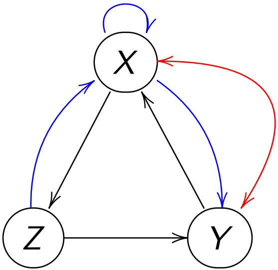
Induced confounding



Correspondence between System and Measurement T.S.



system timescale



measurement timescale

When subsampling by u:

- Measurement time scale edge $Y \to X$ corresponds to path of length u: $Y \to \cdots \to X$
- Measurement time scale edge $X \leftrightarrow Y$ corresponds to paths of length $k < u$: $W \to \cdots \to X$ and $W \to \cdots \to Y$.

A Constraint Satisfaction Solution

Partial Complexity Result

Result: Deciding whether there is a system t.s. structure compatible with the directed edges of a measurement t.s. structure is **NP-complete** for any fixed $u \geq 2$.

Proof: Binary matrix root.

A Constraint Satisfaction Solution by ASP

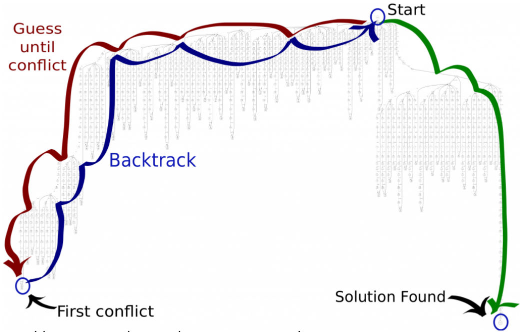
- You write a symbolic encoding.

A Constraint Satisfaction Solution by ASP

- You write a symbolic encoding.
- The symbolic encoding gets grounded.
- The encoding gets turned into conjunctive normal form.
- Backtracking DFS by `Clingo` (Gebser et al. 2011).

A Constraint Satisfaction Solution by ASP

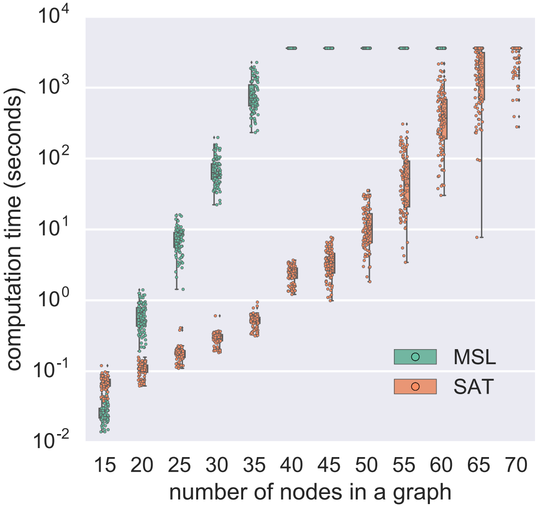
- You write a symbolic encoding.
- The symbolic encoding gets grounded.
- The encoding gets turned into conjunctive normal form.
- Backtracking DFS by `Clingo` (Gebser et al. 2011).
- Exact and complete solution.
- Subsampling rate *u*: fixed or free.



https://srlabs.de/bites/minisat-intro/

```
node(1..3). % Measurement timescale structure
edgeh(1,2).no_edgeh(1,3).confh(2,3).no_confh(1,2). %and so on

urange(1..5).  % Define a range of u:s
1 { u(U): urange(U) } 1. % u(U) is true for only one U

{ edge1(X,Y) } :- node(X), node(Y). %draw G1

% Derive all directed paths up to length U
path(X,Y,1) :- edge1(X,Y).
path(X,Y,L) :- path(X,Z,L-1), edge1(Z,Y), L <= U, u(U).

% Determine measurement t.s. for G1
edgeu(X,Y) :- path(X,Y,L), u(L).
confu(X,Y) :- path(Z,X,L), path(Z,Y,L), node(X;Y;Z),
                               X < Y, L < U, u(U).
% Check consistency
:- edgeh(X,Y), not edgeu(X,Y). :- no_edgeh(X,Y), edgeu(X,Y).
:- confh(X,Y), not confu(X,Y).  :- no_confh(X,Y), confu(X,Y).
```

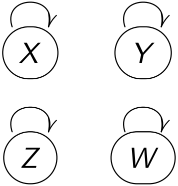
Scalability of Enumerating 1000 Solutions



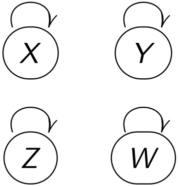
(fixed subsampling rate 2, SAT is our approach,
MSL is the previous state of art by Plis et al. (2015))

Identifiability: Underdetermination

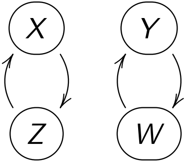
Measurement timescale structure:



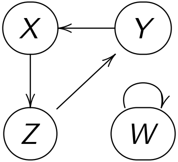
could be produced by system timescale structures:



$u = 1, 2, 3, \cdots$



$u = 2, 4, 6, \cdots$

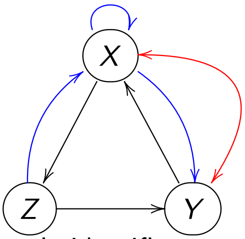


$u = 3, 6, 9, \cdots$

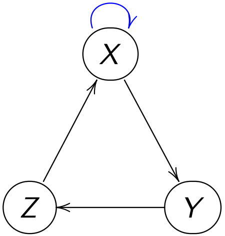
or a four cycle in either direction and symmetrically!

Identifiability: An Identified Case

But measurement timescale structure:



uniquely identifies system timescale structure

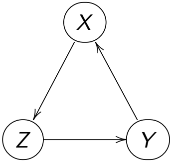


and the subsampling rate $u = 2$.

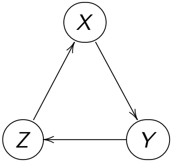
A Constraint Optimization Solution

A Constraint Optimization Solution

$$\begin{array}{cccccc} \cdots & X^{t-4} & X^{t-2} & X^{t} & \cdots \\ \cdots & Y^{t-4} & Y^{t-2} & Y^{t} & \cdots \\ \cdots & Z^{t-4} & Z^{t-2} & Z^{t} & \cdots \end{array} \longrightarrow$$



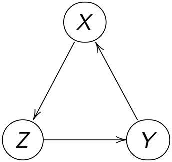
$\longrightarrow$



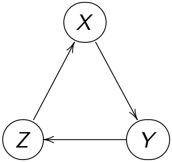
data | measurement t.s. | system t.s.

A Constraint Optimization Solution

$$\begin{array}{cccccc} \cdots & X^{t-4} & X^{t-2} & X^{t} & \cdots & \\ \cdots & Y^{t-4} & Y^{t-2} & Y^{t} & \cdots & \longrightarrow \\ \cdots & Z^{t-4} & Z^{t-2} & Z^{t} & \cdots & \end{array}$$



$\longrightarrow$



data — measurement t.s. — system t.s.

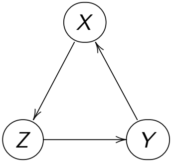
- Measurement t.s. structure can be consistently estimated from data under faithfulness: e.g.

$$X \rightarrow Z \quad \Leftrightarrow \quad X^{t-u} \not\perp\!\!\!\perp Z^{t} \mid \mathbf{V}^{t-u} \setminus X^{t-u}$$

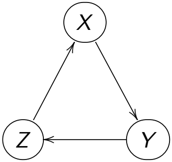
$$X \leftrightarrow Z \quad \Leftrightarrow \quad X^{t} \not\perp\!\!\!\perp Y^{t} \mid \mathbf{V}^{t-u}$$

A Constraint Optimization Solution

$$\begin{array}{cccccc} \cdots & X^{t-4} & X^{t-2} & X^{t} & \cdots \\ \cdots & Y^{t-4} & Y^{t-2} & Y^{t} & \cdots & \longrightarrow \\ \cdots & Z^{t-4} & Z^{t-2} & Z^{t} & \cdots \end{array}$$



$\longrightarrow$

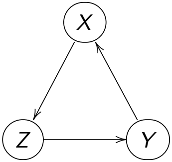


data | measurement t.s. | system t.s.

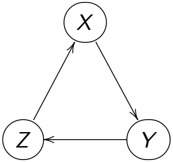
- Measurement t.s. structure can be consistently estimated from data under faithfulness: e.g.
  $X \rightarrow Z \quad \Leftrightarrow \quad X^{t-u} \not\perp\!\!\!\perp Z^{t} \mid \mathbf{V}^{t-u} \setminus X^{t-u}$
  $X \leftrightarrow Z \quad \Leftrightarrow \quad X^{t} \not\perp\!\!\!\perp Y^{t} \mid \mathbf{V}^{t-u}$
- Due to finite samplesize, the constraint satisfaction approach will often return UNSATISFIABLE.

A Constraint Optimization Solution

$$\begin{array}{cccccc} \cdots & X^{t-4} & X^{t-2} & X^{t} & \cdots \\ \cdots & Y^{t-4} & Y^{t-2} & Y^{t} & \cdots & \longrightarrow \\ \cdots & Z^{t-4} & Z^{t-2} & Z^{t} & \cdots \end{array}$$



$\longrightarrow$



data measurement t.s. system t.s.

- Measurement t.s. structure can be consistently estimated from data under faithfulness: e.g.
  $X \to Z \quad \Leftrightarrow \quad X^{t-u} \not\perp\!\!\!\perp Z^{t} \mid \mathbf{V}^{t-u} \setminus X^{t-u}$
  $X \leftrightarrow Z \quad \Leftrightarrow \quad X^{t} \not\perp\!\!\!\perp Y^{t} \mid \mathbf{V}^{t-u}$
- Due to finite samplesize, the constraint satisfaction approach will often return UNSATISFIABLE.
- Find the system t.s. structure such that its measurement t.s. structure is optimally close to the estimated (Task 2).

A Constraint Optimization Solution by ASP

Specifics:

- Penalize inconsistencies between absences and precences of edges in the measurement t.s.:
  - Either uniform weights, or
  - log Bayesian probabilities of the corresponding (in)dependence, obtained through Bayesian model selection (see Hyttinen et al. 2014)
  - Objective function is the sum of the penalities

A Constraint Optimization Solution by ASP

Specifics:

- Penalize inconsistencies between absences and precences of edges in the measurement t.s.:
  - Either uniform weights, or
  - log Bayesian probabilities of the corresponding (in)dependence, obtained through Bayesian model selection (see Hyttinen et al. 2014)
  - Objective function is the sum of the penalities
- `Clingo` uses Branch-and-Bound search to find the exact weighted Maximum Satisfiability solution.

A Constraint Optimization Solution by ASP

Specifics:

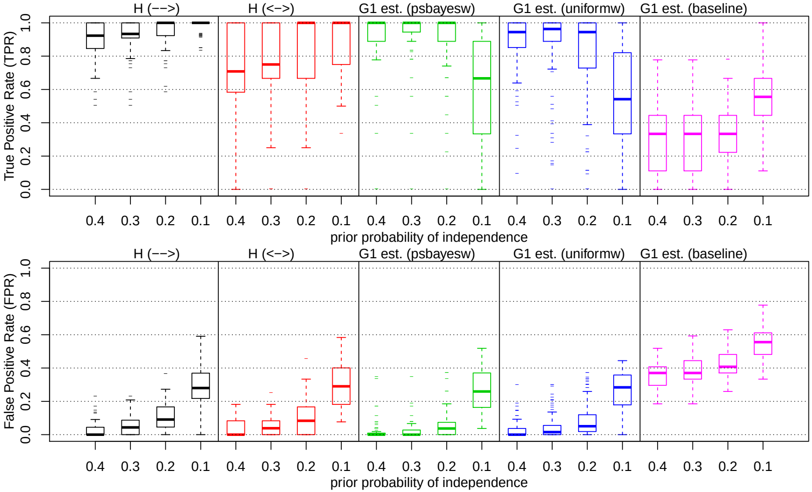
- Penalize inconsistencies between absences and precences of edges in the measurement t.s.:
  - Either uniform weights, or
  - log Bayesian probabilities of the corresponding (in)dependence, obtained through Bayesian model selection (see Hyttinen et al. 2014)
  - Objective function is the sum of the penalities
- `Clingo` uses Branch-and-Bound search to find the exact weighted Maximum Satisfiability solution.
- We scale to 11-12 within 10 minutes, depending on the sample size and other specifics

A Constraint Optimization Solution by ASP

Specifics:

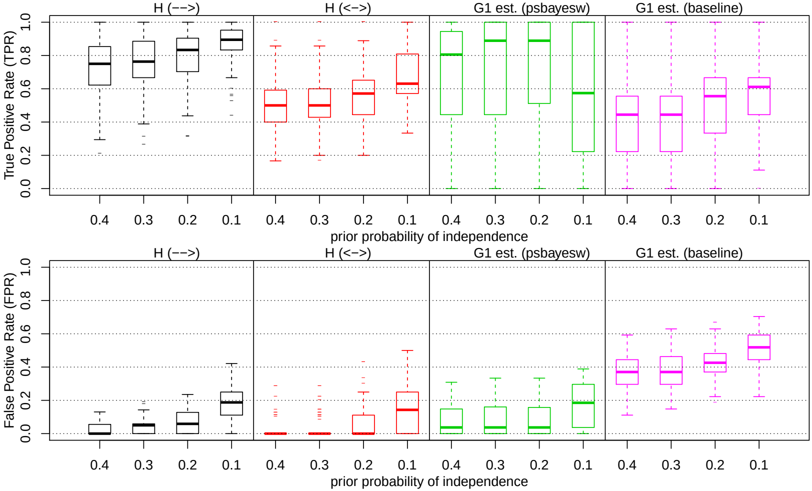
- Penalize inconsistencies between absences and precences of edges in the measurement t.s.:
  - Either uniform weights, or
  - log Bayesian probabilities of the corresponding (in)dependence, obtained through Bayesian model selection (see Hyttinen et al. 2014)
  - Objective function is the sum of the penalities
- `Clingo` uses Branch-and-Bound search to find the exact weighted Maximum Satisfiability solution.
- We scale to 11-12 within 10 minutes, depending on the sample size and other specifics
- Previous work by Plis et al. 2015.

Accuracy for fixed $u = 2$



(fixed subsampling rate 2, average result of the eq. class,
6 nodes, av. degree 3, 200 samples, 100 data sets, linear models)

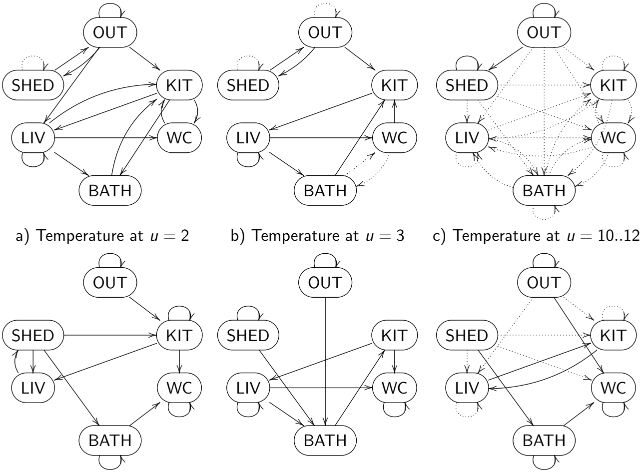
Accuracy for $u = 3$



Analysis of Temperature/Humidity data 1

- Hourly measurements of six sensors placed in a house.
- Temperature and humidity recorded.
- Removed trends.
- Handle undetermination: for each edge [Magliacane et al.]
  - run the inference procedure enforcing presence
  - and then enforcing absence
  - difference in objectives gives the support for the edge.

Analysis of Temperature/Humidity data 2



a) Temperature at $u = 2$ b) Temperature at $u = 3$ c) Temperature at $u = 10..12$

d) Humidity at $u = 2$ e) Humidity at $u = 3$ f) Humidity at $u = 10..12$

Edges with full lines are found to be present, absent edges are found to be absent, edges with dotted lines are present or absent.

Conclusion

Conclusion

Causal discovery from subsampled time series data:

Conclusion

Causal discovery from subsampled time series data:

- A non-parametric constraint satisfaction approach:
  Much better scalability than previous state-of-the-art.

Conclusion

Causal discovery from subsampled time series data:

- A non-parametric constraint satisfaction approach:
  Much better scalability than previous state-of-the-art.
- A (first) constraint optimization approach:
  More accurate than unweighted or unoptimal solutions.

Conclusion

Causal discovery from subsampled time series data:

- A non-parametric constraint satisfaction approach:
  Much better scalability than previous state-of-the-art.
- A (first) constraint optimization approach:
  More accurate than unweighted or unoptimal solutions.
- Future work: generalizing the model space, e.g. allowing for unobserved confounding time series.

Conclusion

Causal discovery from subsampled time series data:

- A non-parametric constraint satisfaction approach:
  Much better scalability than previous state-of-the-art.
- A (first) constraint optimization approach:
  More accurate than unweighted or unoptimal solutions.
- Future work: generalizing the model space, e.g. allowing for unobserved confounding time series.

Thanks!