Learning the Structure of Causal Models with Relational and Temporal Dependence

Katerina Marazopoulou, Marc Maier, David Jensen

College of Information and Computer Sciences
University of Massachusetts Amherst

UAI 2015: Advances in Causal Inference Workshop

Learning the Structure of Causal Models with Relational and Temporal Dependence

Learning the Structure of Causal Models with Relational and Temporal Dependence

1. Representation

Learning the Structure of Causal Models with Relational and Temporal Dependence

1. Representation
   - Relational (multiple types of interacting entities)

Learning the Structure of Causal Models with Relational and Temporal Dependence

1. Representation
    - Relational (multiple types of interacting entities)
    - Temporal (values of variables change over time)

Learning the Structure of Causal Models with Relational and Temporal Dependence

1. Representation
   - Relational (multiple types of interacting entities)
   - Temporal (values of variables change over time)
2. Structure learning algorithm

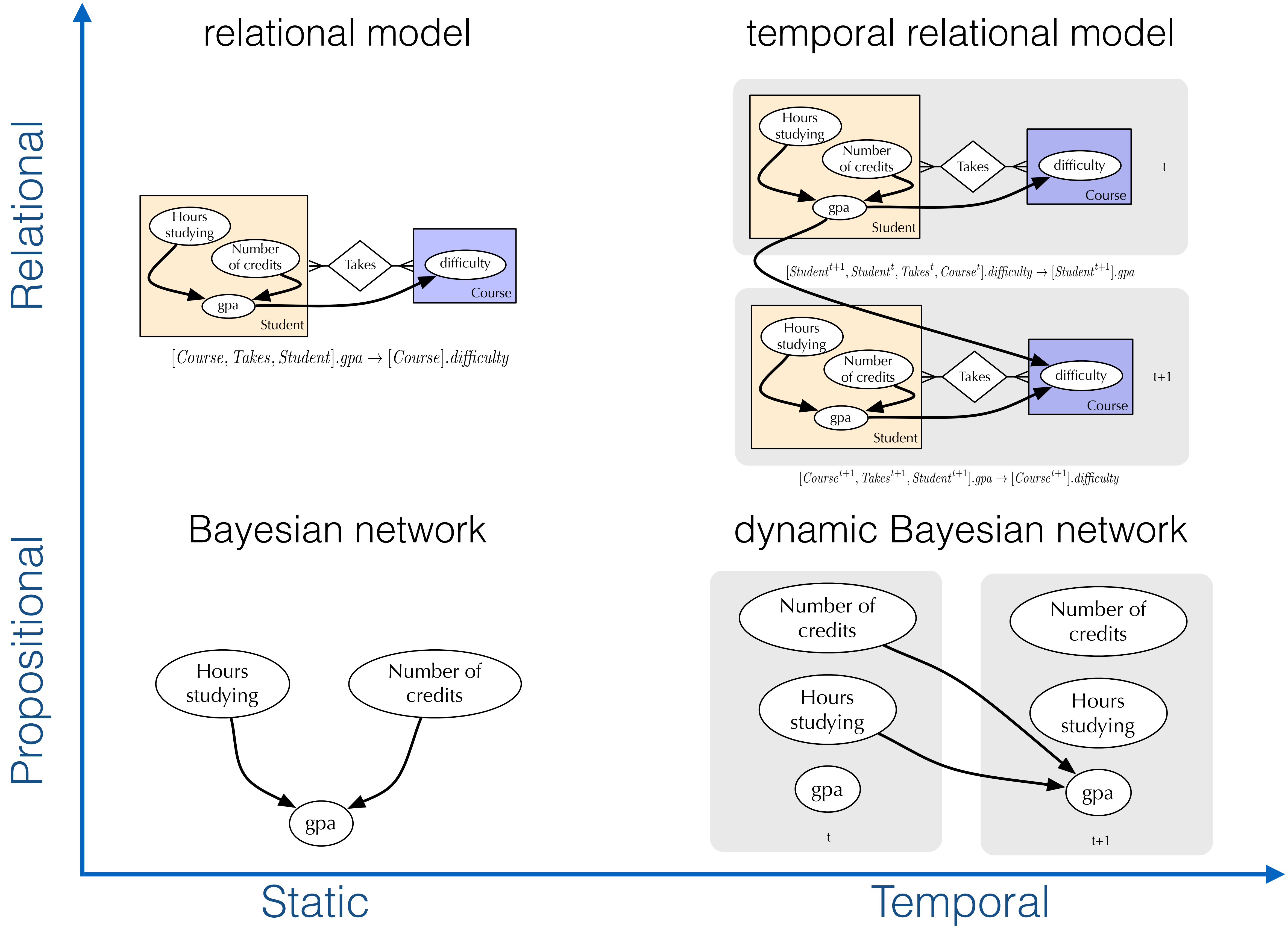
Learning the Structure of Causal Models with Relational and Temporal Dependence

1. Representation
   - Relational (multiple types of interacting entities)
   - Temporal (values of variables change over time)
2. d-separation for temporal relational models
3. Structure learning algorithm

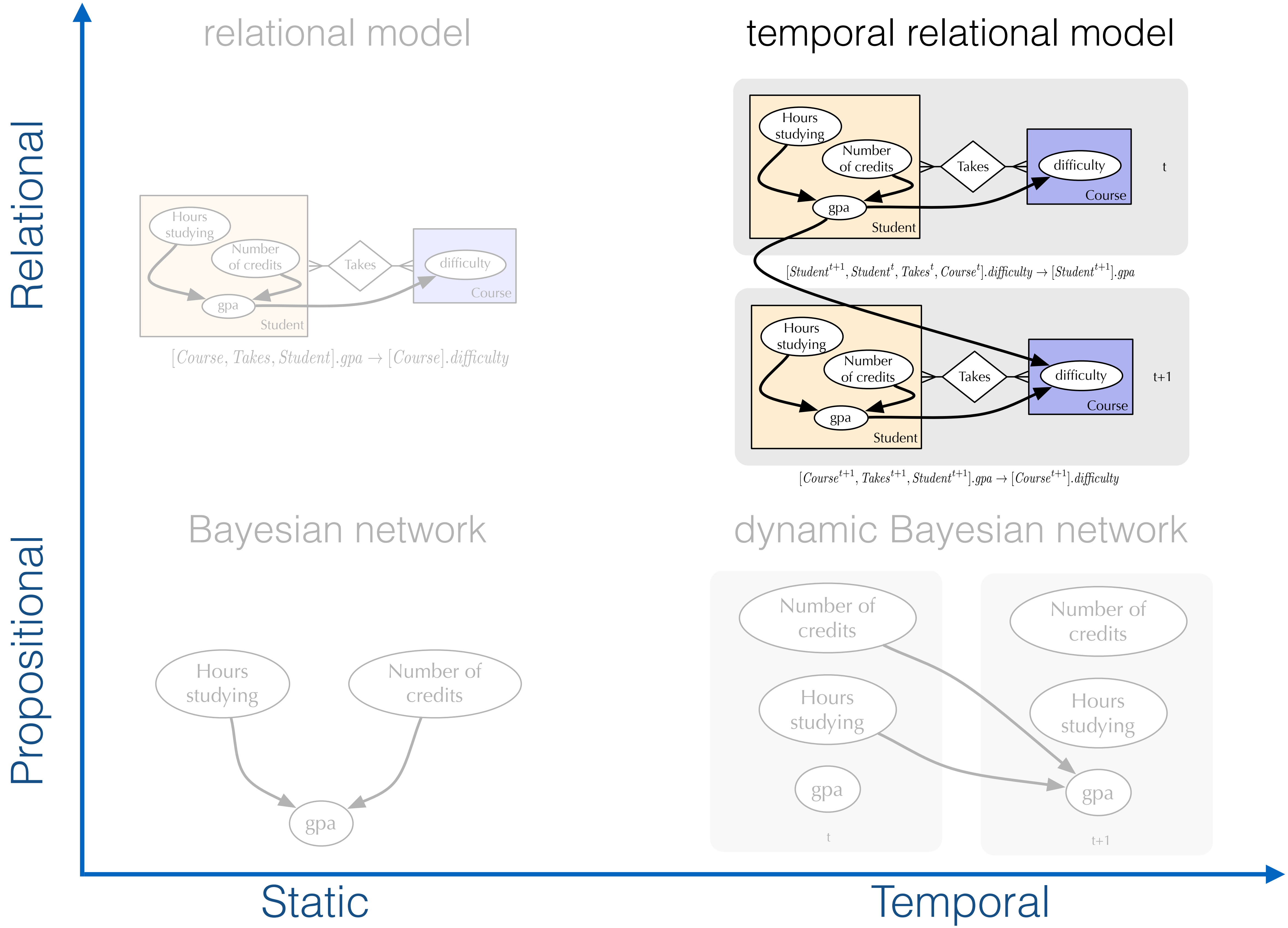
Why?

- Many real-world systems consist of heterogeneous entities that interact with each other (relational) over time (temporal)
  - Authors and citations
  - Social networks
  - Epidemiology
  - Education
- More expressive models can represent such domains more accurately.
- Expressiveness is particularly important for causality.

Representation

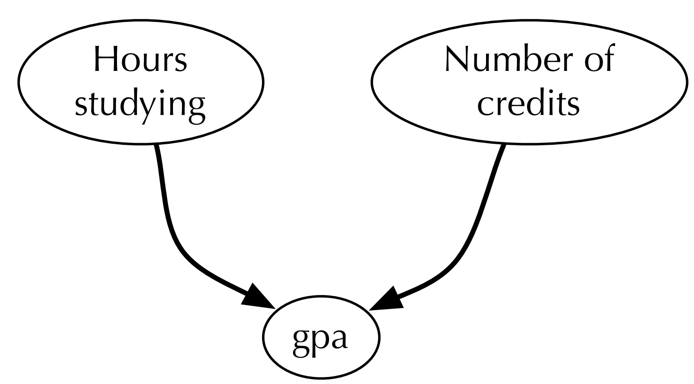


Representation

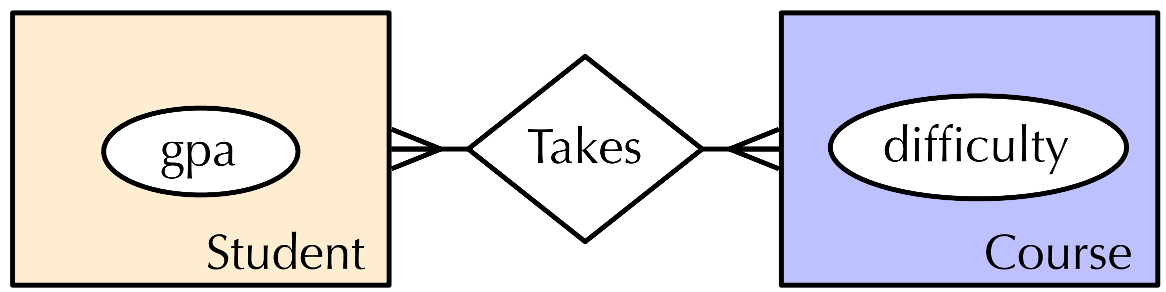


Bayesian networks

- Directed acyclic graphs
- Nodes are random variables
- Model dependencies between variables of a single entity type

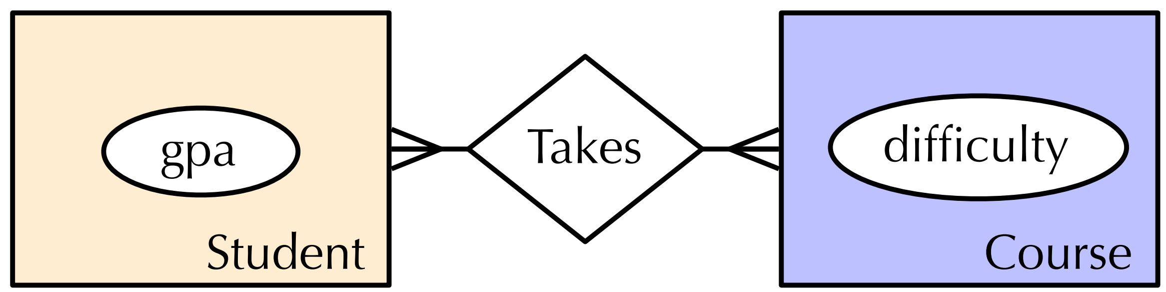


Background on Relational Models



Relational schema

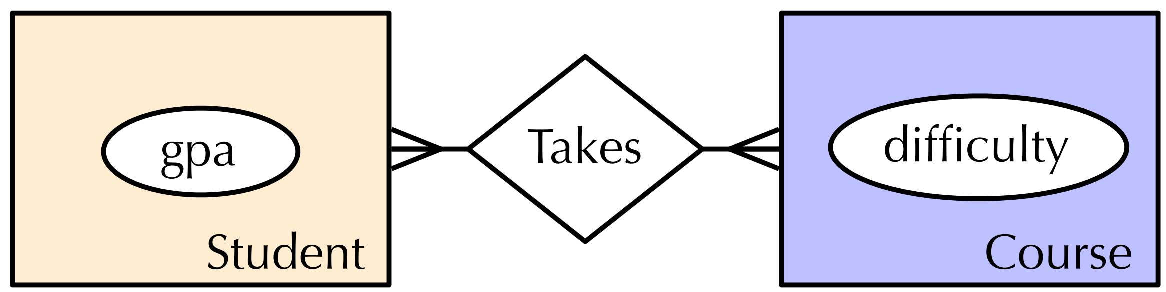
Background on Relational Models



Relational paths: [Course, Takes, Student]

"The students that take a course"

Background on Relational Models

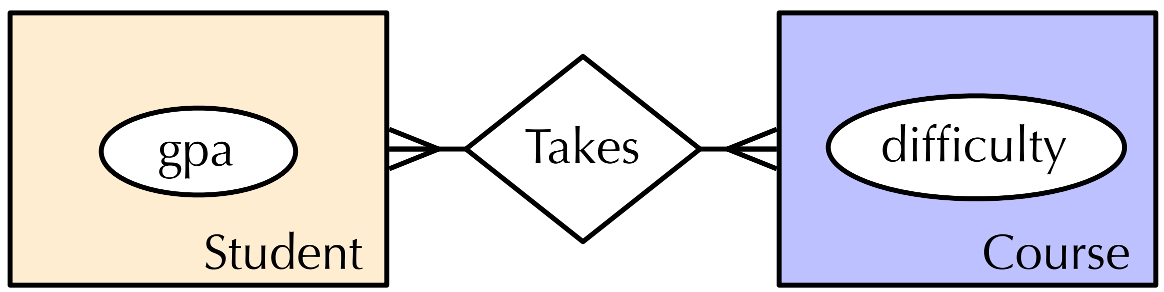


Relational paths: [Course, Takes, Student]

Relational variables: [Course, Takes, Student].gpa

"The GPA of the students that take a course"

Background on Relational Models



Relational paths: [Course, Takes, Student]

Relational variables: [Course, Takes, Student].gpa

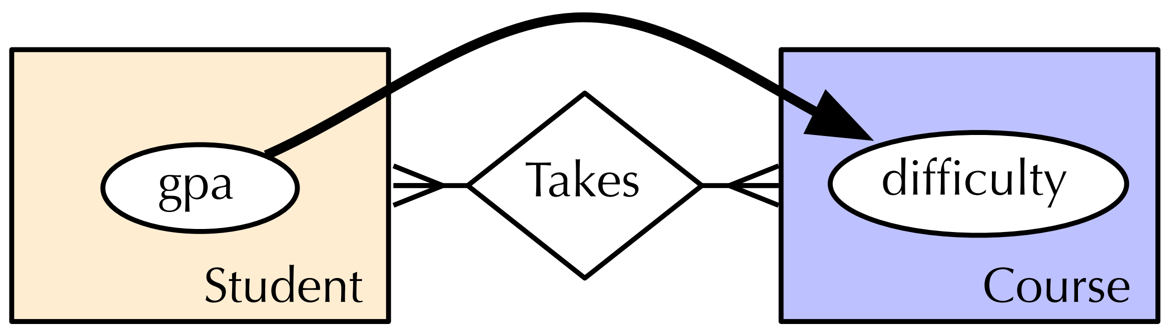
Relational dependencies:

[Course, Takes, Student].gpa → [Course].difficulty

"The difficulty of a course depends on the GPA of the students that take that course."

Background on Relational Models

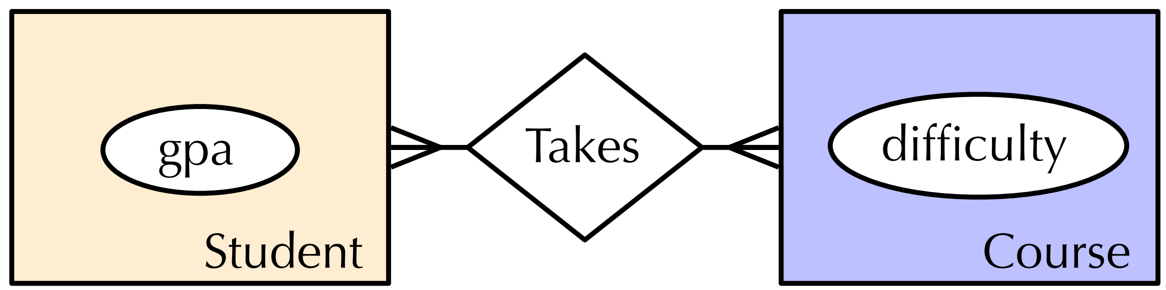
$[Course, Takes, Student].gpa \rightarrow [Course].difficulty$



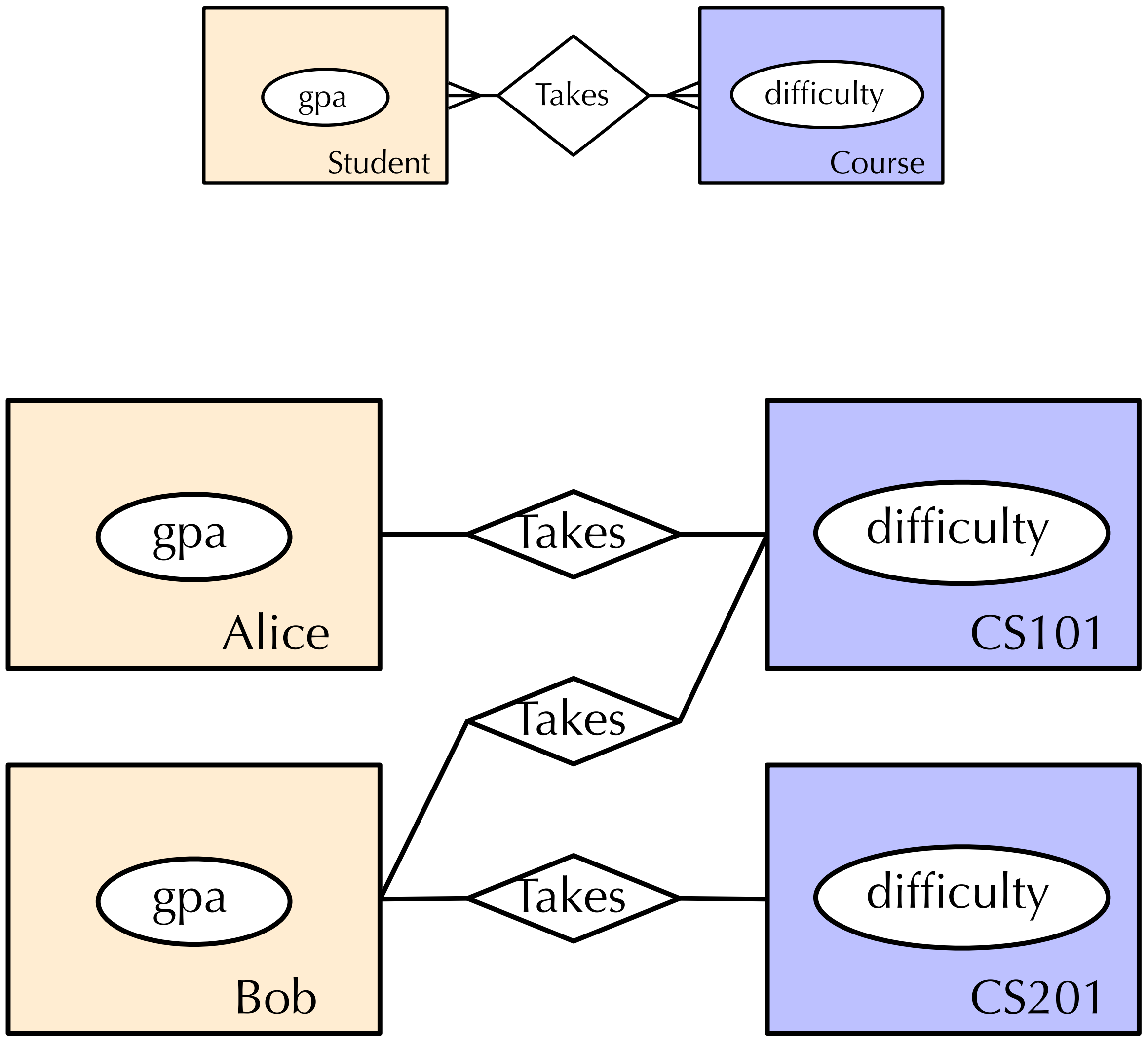
Relational model:

relational schema + set of relational dependencies

Background on Relational Models



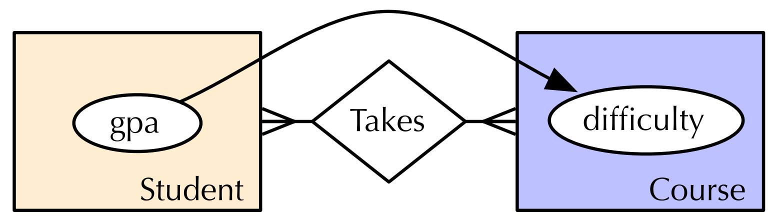
Background on Relational Models

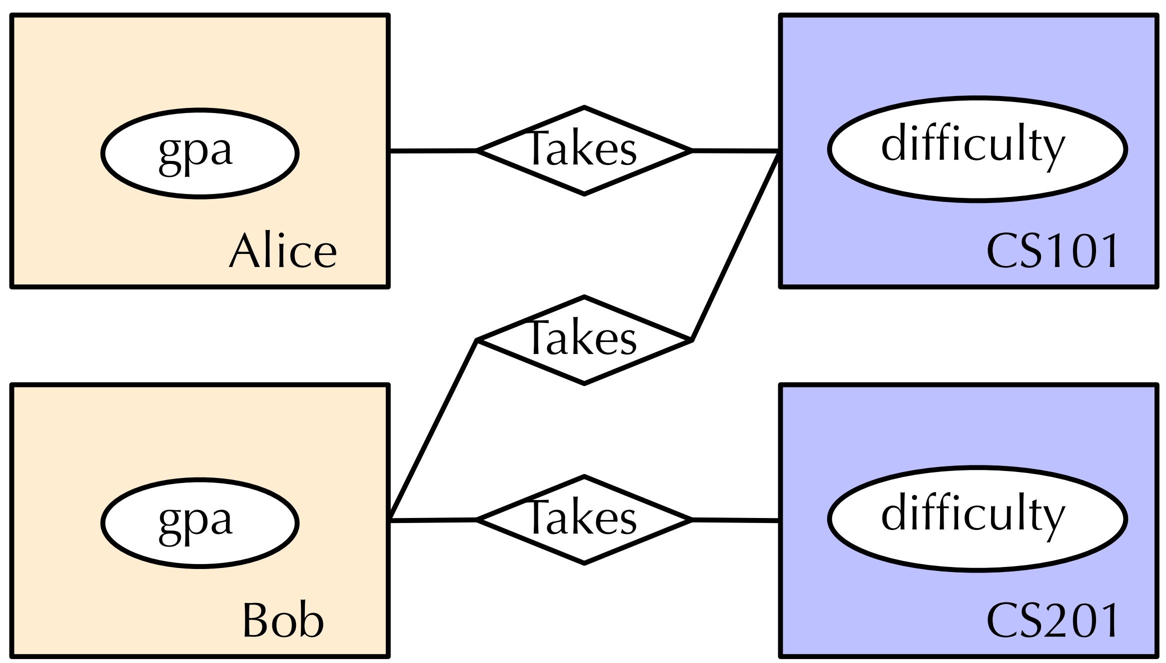


Relational skeleton: Set of entity and relationship instances

Background on Relational Models

$$[Course, Takes, Student].gpa \rightarrow [Course].difficulty$$

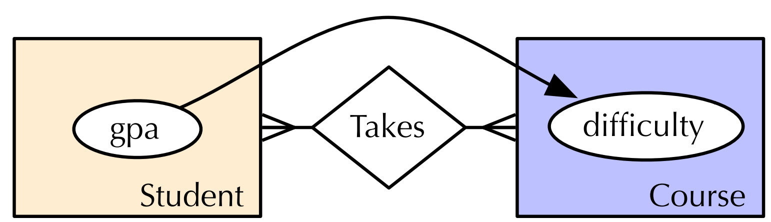


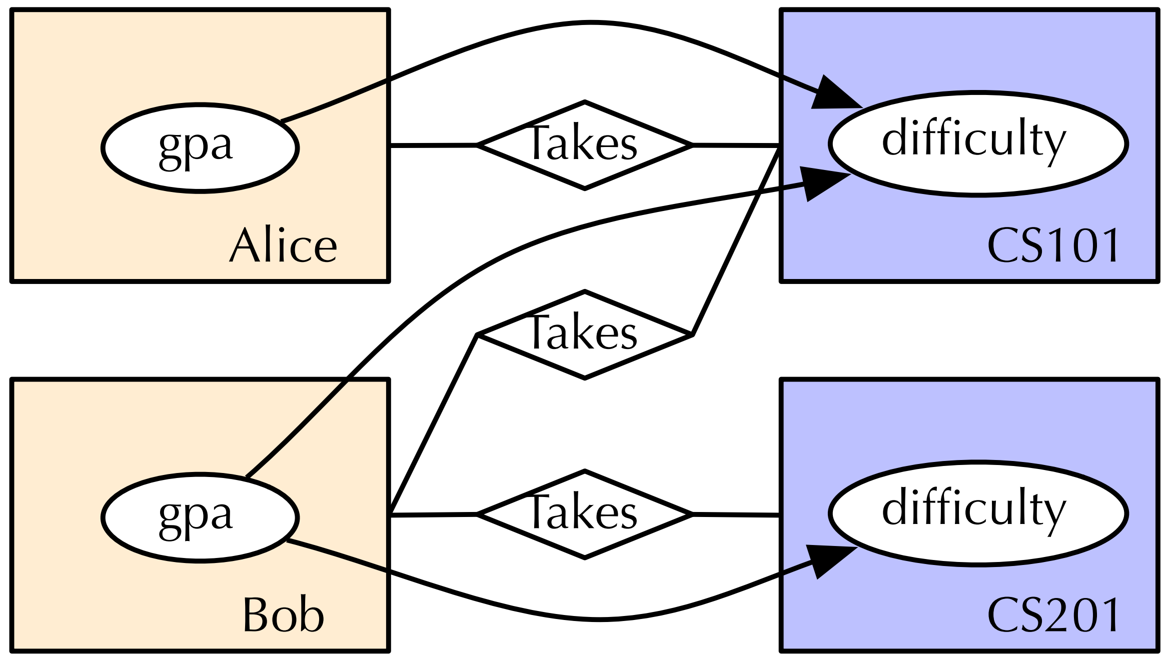


Model + relational skeleton → ground graph

Background on Relational Models

$$[Course, Takes, Student].gpa \rightarrow [Course].difficulty$$

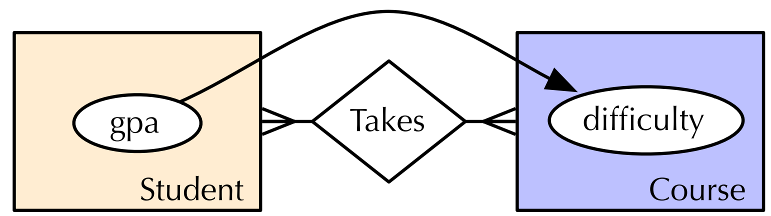


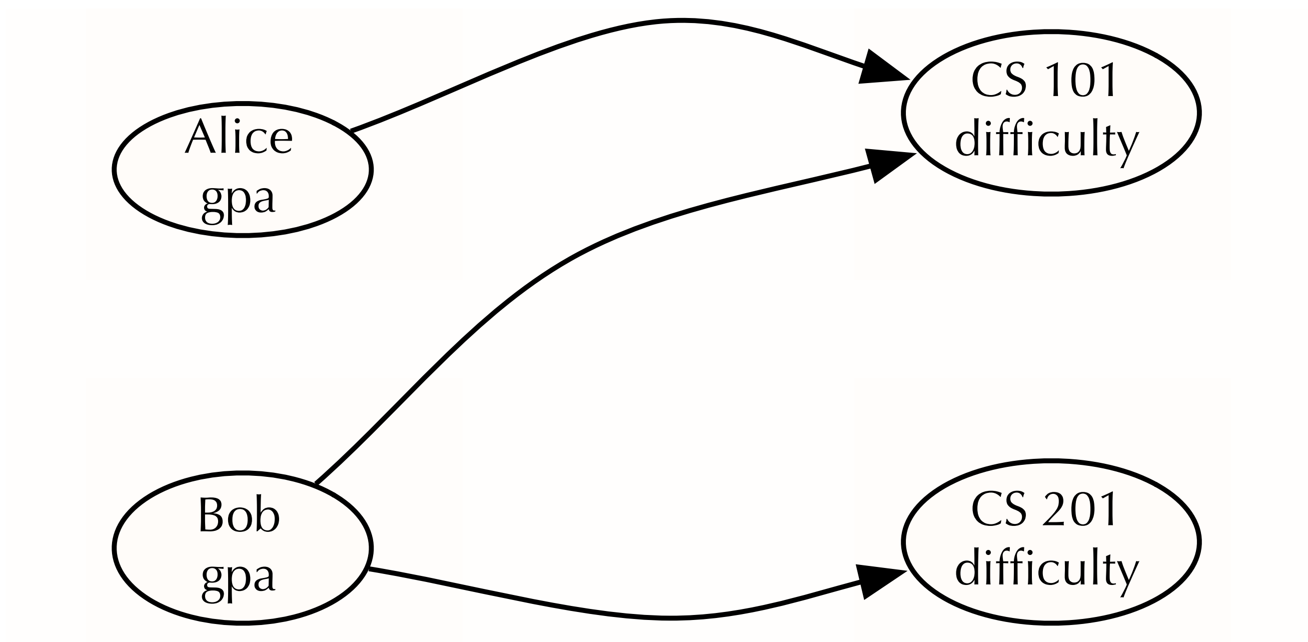


Ground graph: Apply the model to the relational skeleton

Background on Relational Models

$[Course, Takes, Student].gpa \rightarrow [Course].difficulty$





Ground graph: Apply the model to the relational skeleton

Outline

1. Relational concepts
2. Temporal relational models
3. Temporal relational d-separation
4. TRCD algorithm

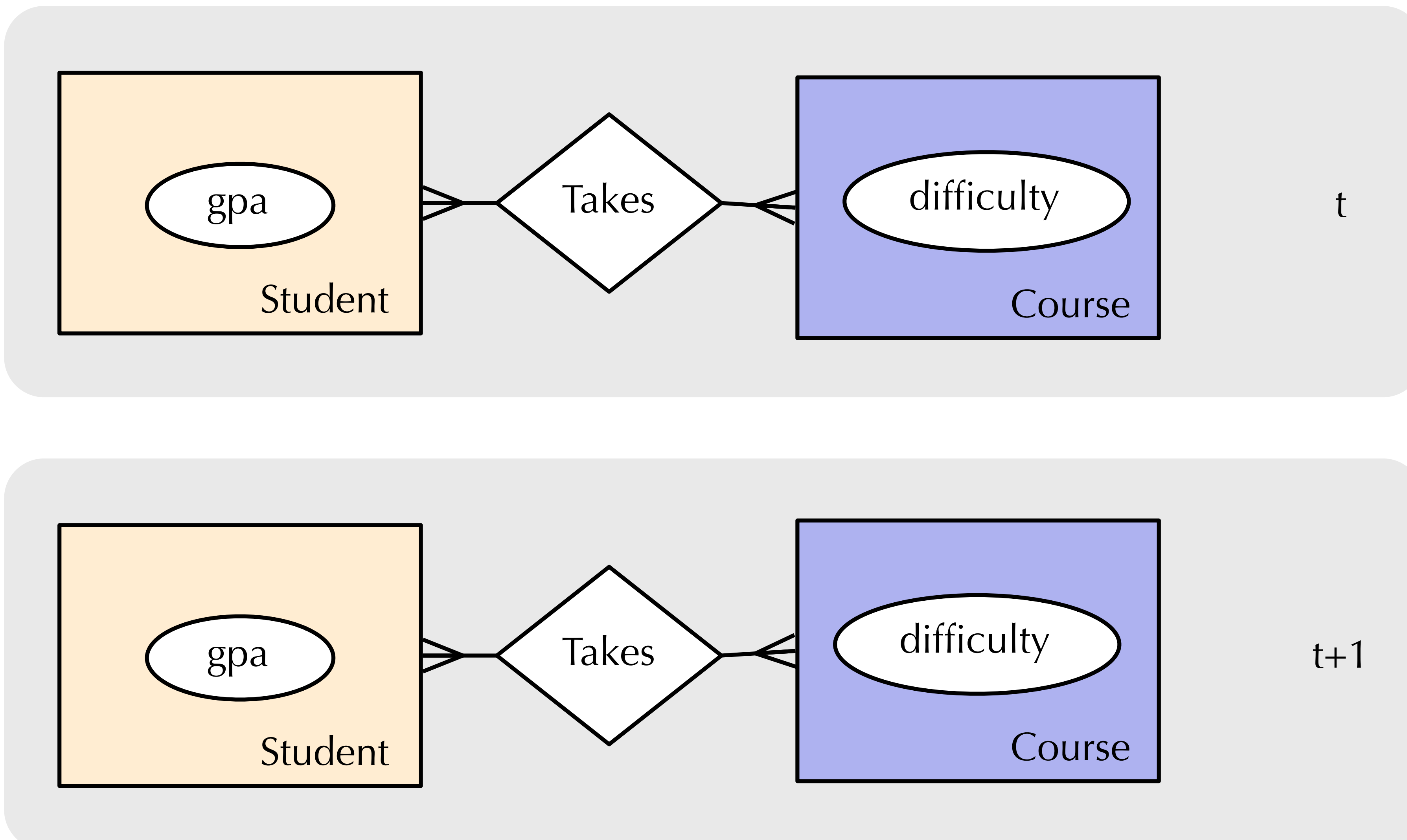
Temporal relational models

Assumptions for the representation

1. Discrete time
2. Model is stationary
3. Temporal skeleton is known *a priori*
4. Dependencies follow first-order Markov assumption

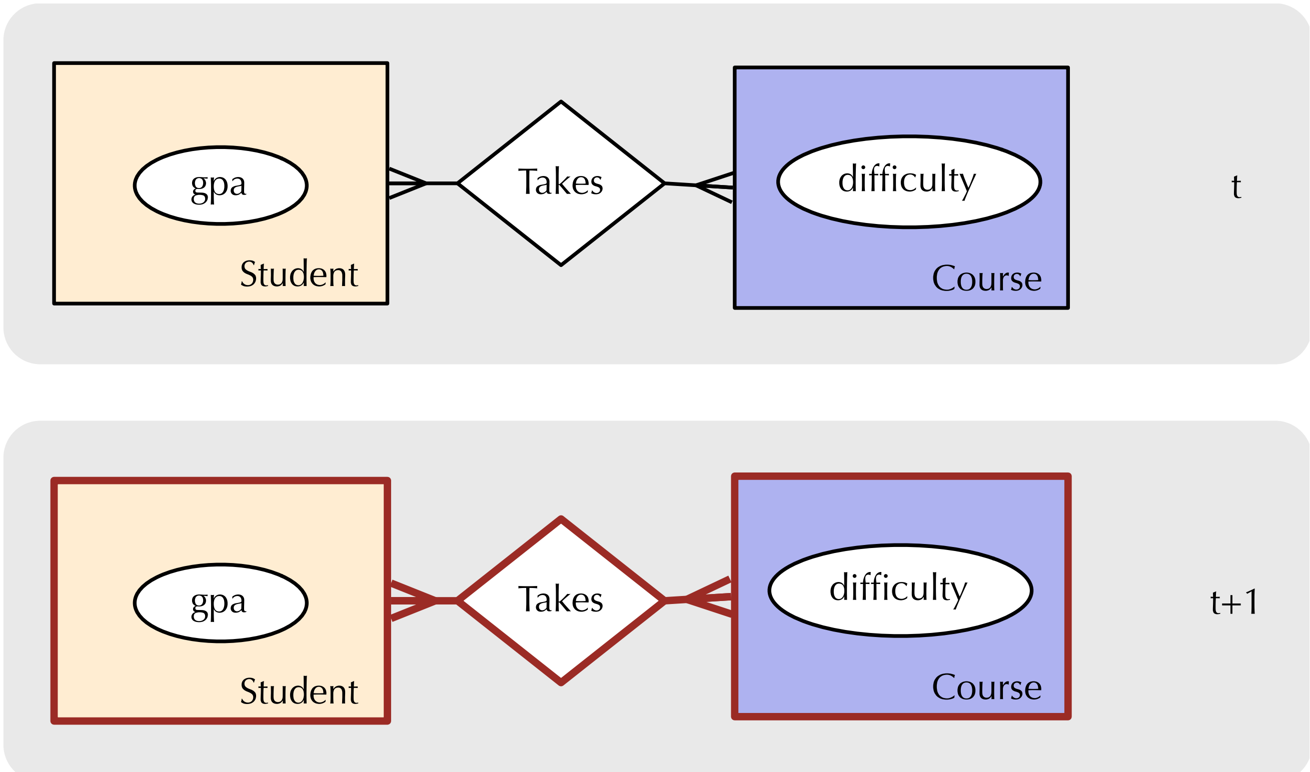
Represent only two consecutive time points
(2-slice models)

Temporal relational models



Temporal relational schema

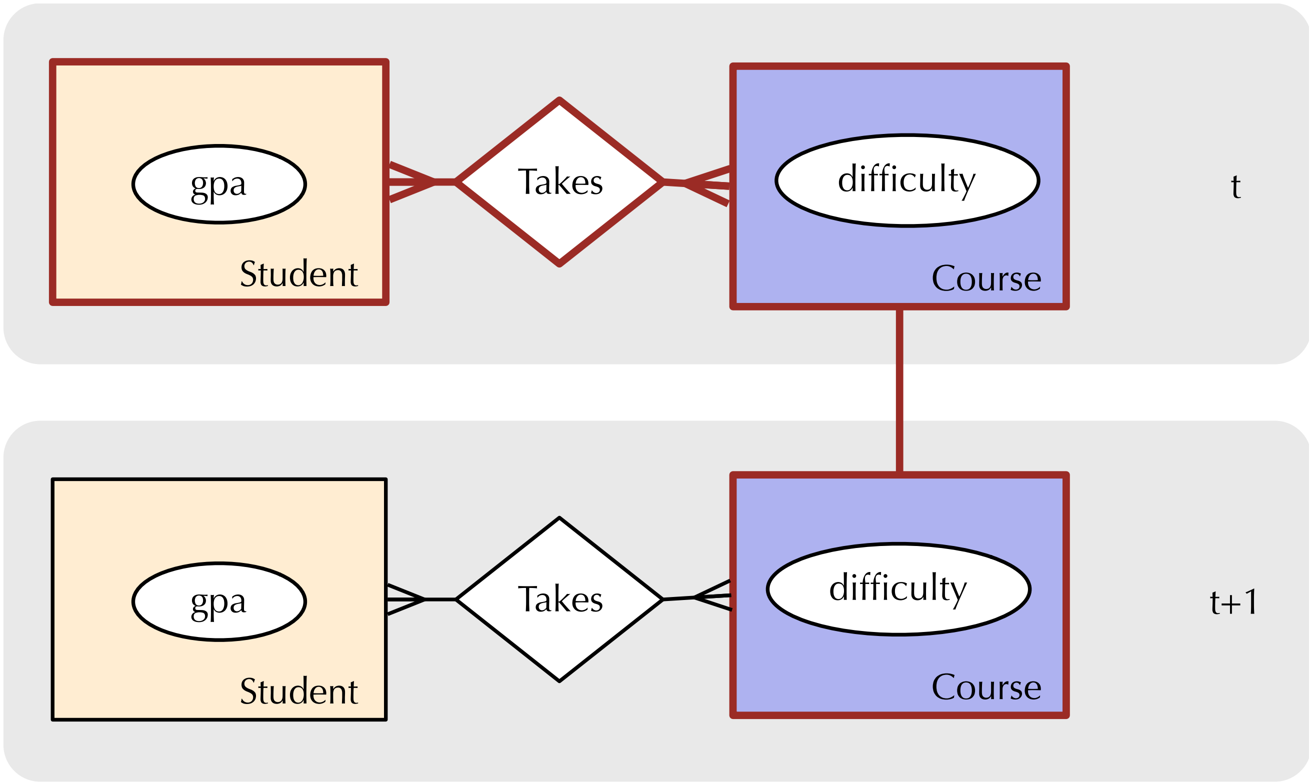
Temporal relational paths



$[\text{Course}^{t+1}, \text{Takes}^{t+1}, \text{Student}^{t+1}]$

"Students that take a course in the current semester"

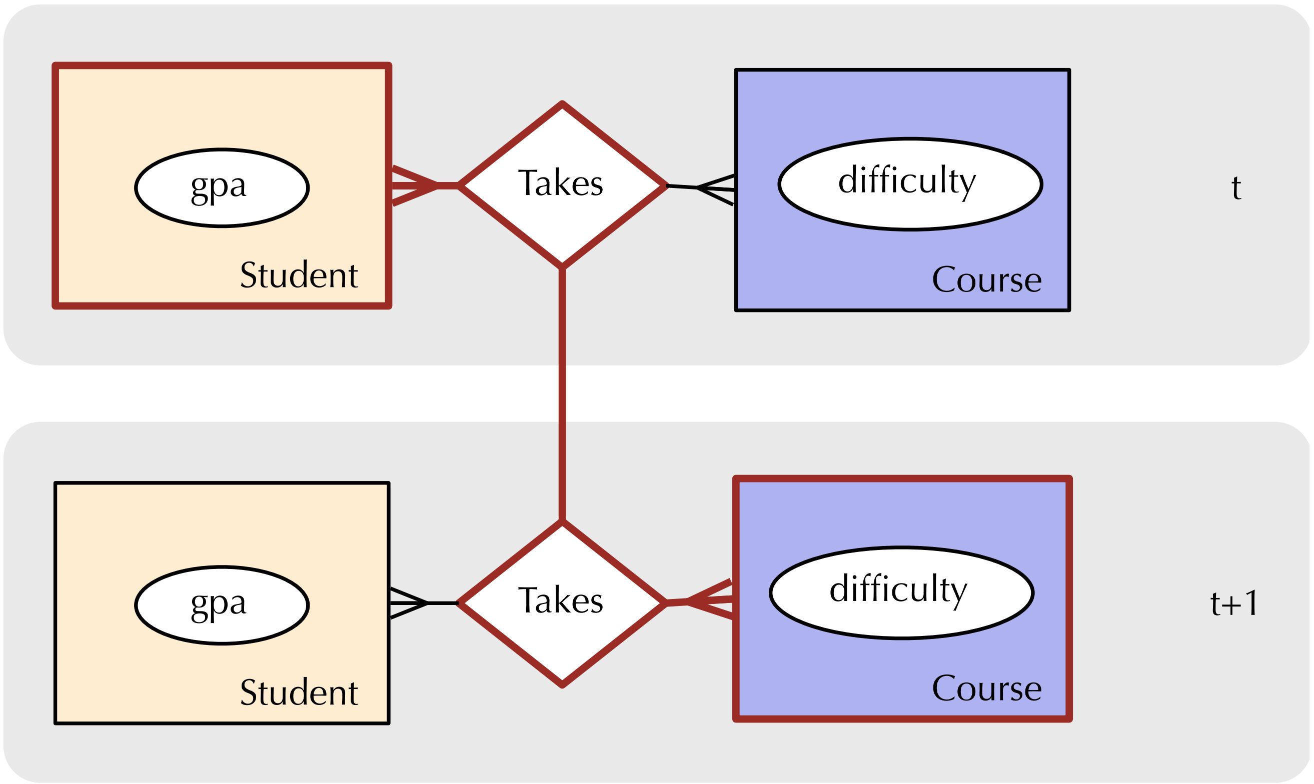
Temporal relational paths



[Course^{t+1}, Course^{t},Takes^{t}, Student^{t}]

"Students that took the course in the previous semester"

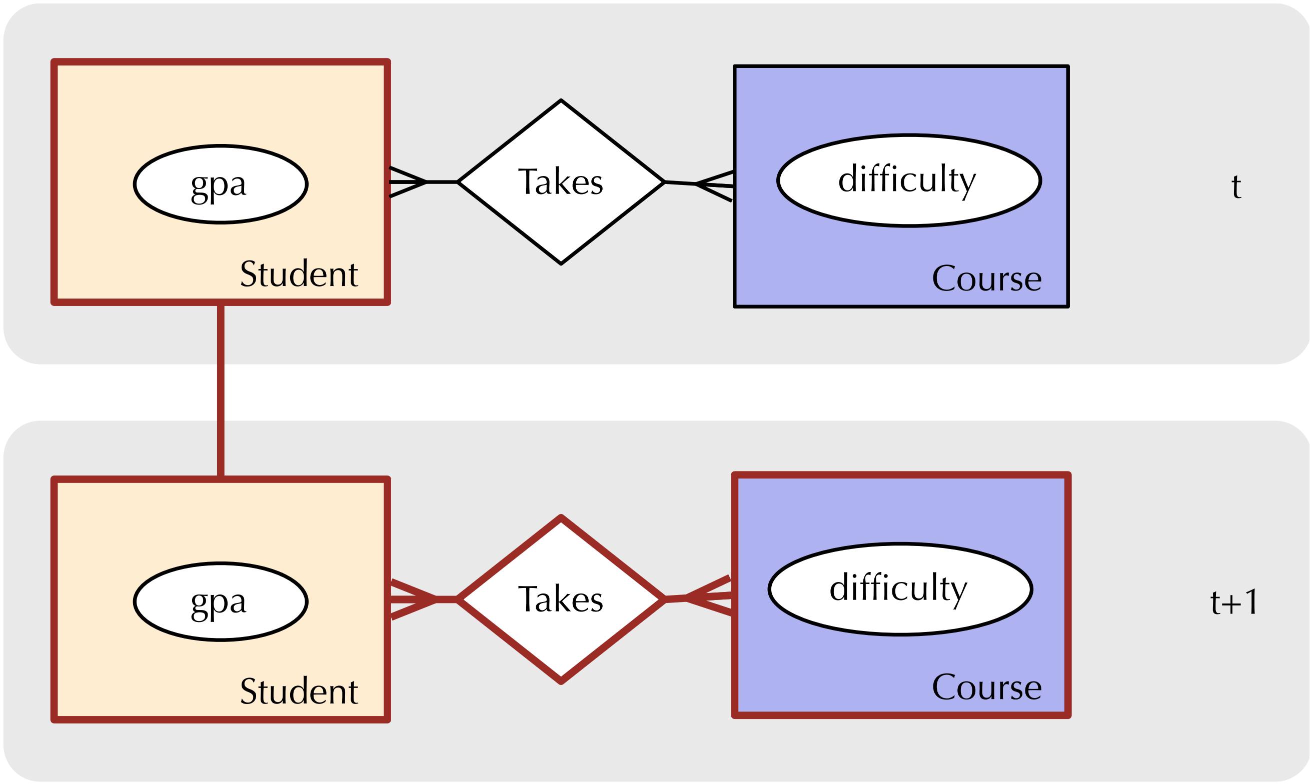
Temporal relational paths



[Course^{t+1}, Takes^{t+1},Takes^{t}, Student^{t}]

“Students that took the course both in the current and in the previous semester”

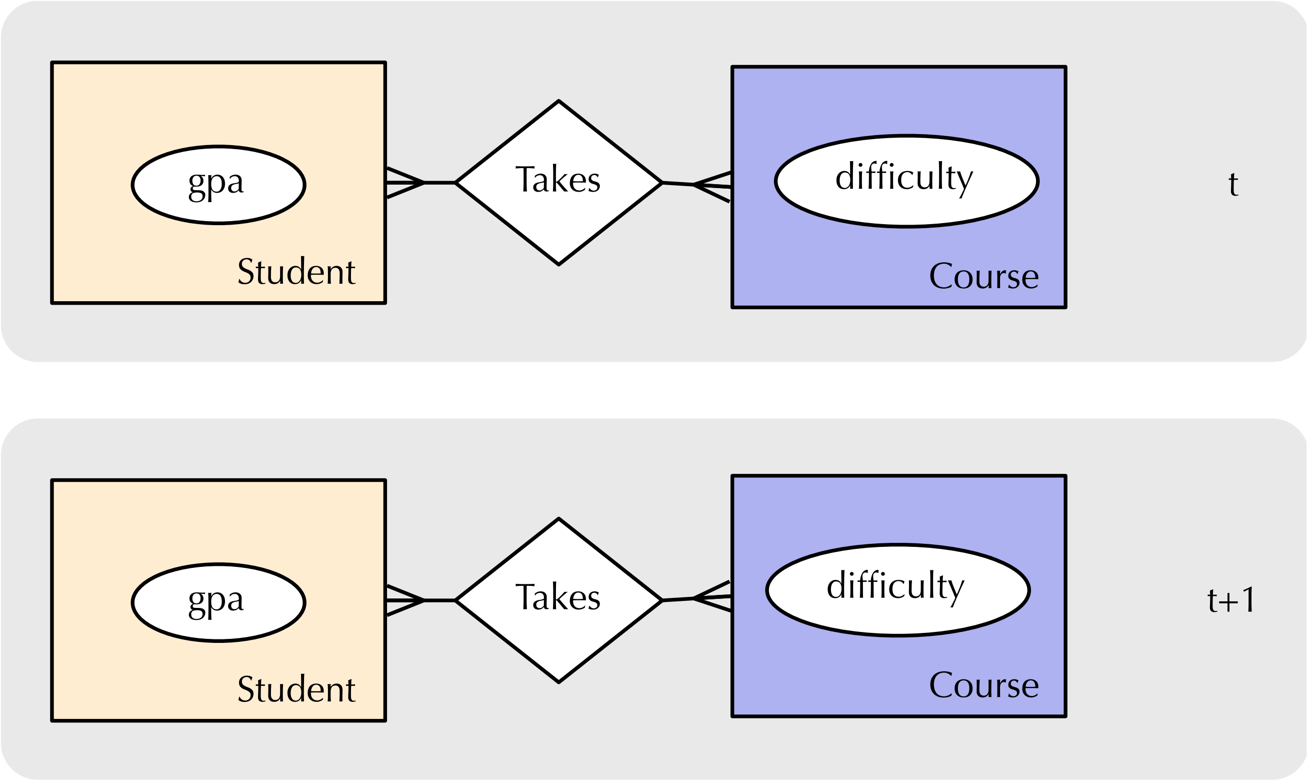
Temporal relational paths



[Course^{t+1}, Takes^{t+1}, Student^{t+1}, Student^{t}]

"The previous state of the students that took the course in the current semester"

Temporal relational paths



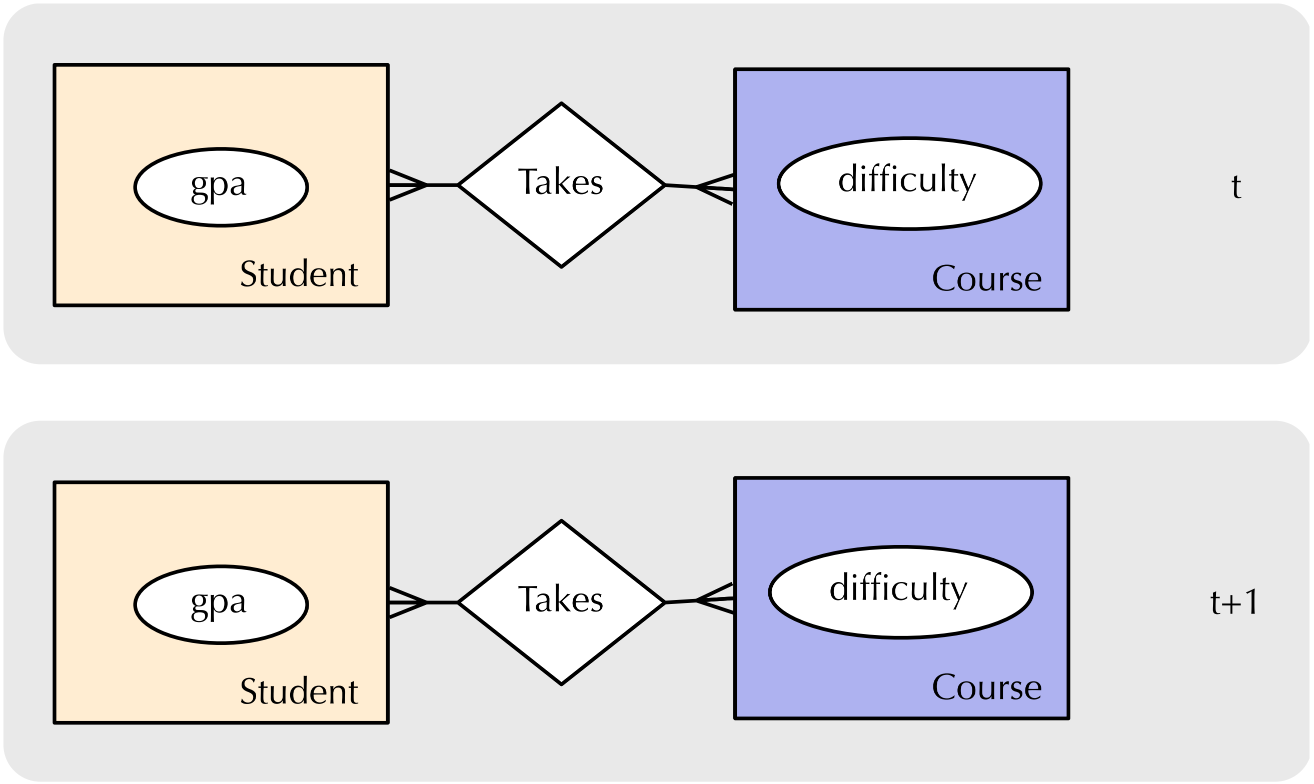
[Course^{t+1}, Takes^{t+1}, Student^{t+1}]

[Course^{t+1}, Course^{t}, Takes^{t}, Student^{t}]

[Course^{t+1}, Takes^{t+1}, Takes^{t}, Student^{t}]

[Course^{t+1}, Takes^{t+1}, Student^{t+1}, Student^{t}]

Temporal relational variables



$[\text{Course}^{t+1}, \text{Takes}^{t+1}, \text{Student}^{t+1}].\text{gpa}$

$[\text{Course}^{t+1}, \text{Course}^{t}, \text{Takes}^{t}, \text{Student}^{t}].\text{gpa}$

$[\text{Course}^{t+1}, \text{Takes}^{t+1}, \text{Takes}^{t}, \text{Student}^{t}].\text{gpa}$

$[\text{Course}^{t+1}, \text{Takes}^{t+1}, \text{Student}^{t+1}, \text{Student}^{t}].\text{gpa}$

Temporal relational dependencies

$[\text{Course}^{t+1}, \text{Takes}^{t+1}, \text{Student}^{t+1}].\text{gpa} \rightarrow [\text{Course}^{t+1}].\text{difficulty}$

"The difficulty of a course in the current semester depends on the GPA of the students that take this course in the current semester."

Temporal relational dependencies

$[\text{Course}^{t+1}, \text{Takes}^{t+1}, \text{Student}^{t+1}].\text{gpa} \rightarrow [\text{Course}^{t+1}].\text{difficulty}$

"The difficulty of a course in the current semester depends on the GPA of the students that take this course in the current semester."

$[\text{Student}^{t+1}, \text{Student}^{t}, \text{Takes}^{t}, \text{Course}^{t}].\text{difficulty} \rightarrow [\text{Student}^{t+1}].\text{gpa}$

"The GPA of a student in the current semester depends on the difficulty of the courses that student took in the previous semester."

Constraints on temporal relational dependencies

[Course^{t+1}, Takes^{t+1}, **Student$^{\mathbf{t+1}}$**].gpa → [**Course$^{\mathbf{t+1}}$**].difficulty

[Student^{t+1}, Student^{t}, Takes^{t}, **Course$^{\mathbf{t}}$**].difficulty → [**Student$^{\mathbf{t+1}}$**].gpa

- Never from the future to the past

Constraints on temporal relational dependencies

$[\text{Course}^{t+1}, \text{Takes}^{t+1}, \textbf{Student}^{\mathbf{t+1}}].\text{gpa} \rightarrow [\textbf{Course}^{\mathbf{t+1}}].\text{difficulty}$

$[\text{Student}^{t+1}, \text{Student}^{t}, \text{Takes}^{t}, \textbf{Course}^{\mathbf{t}}].\text{difficulty} \rightarrow [\textbf{Student}^{\mathbf{t+1}}].\text{gpa}$

- First order Markov assumption
  - Cause and effect at most one time step apart

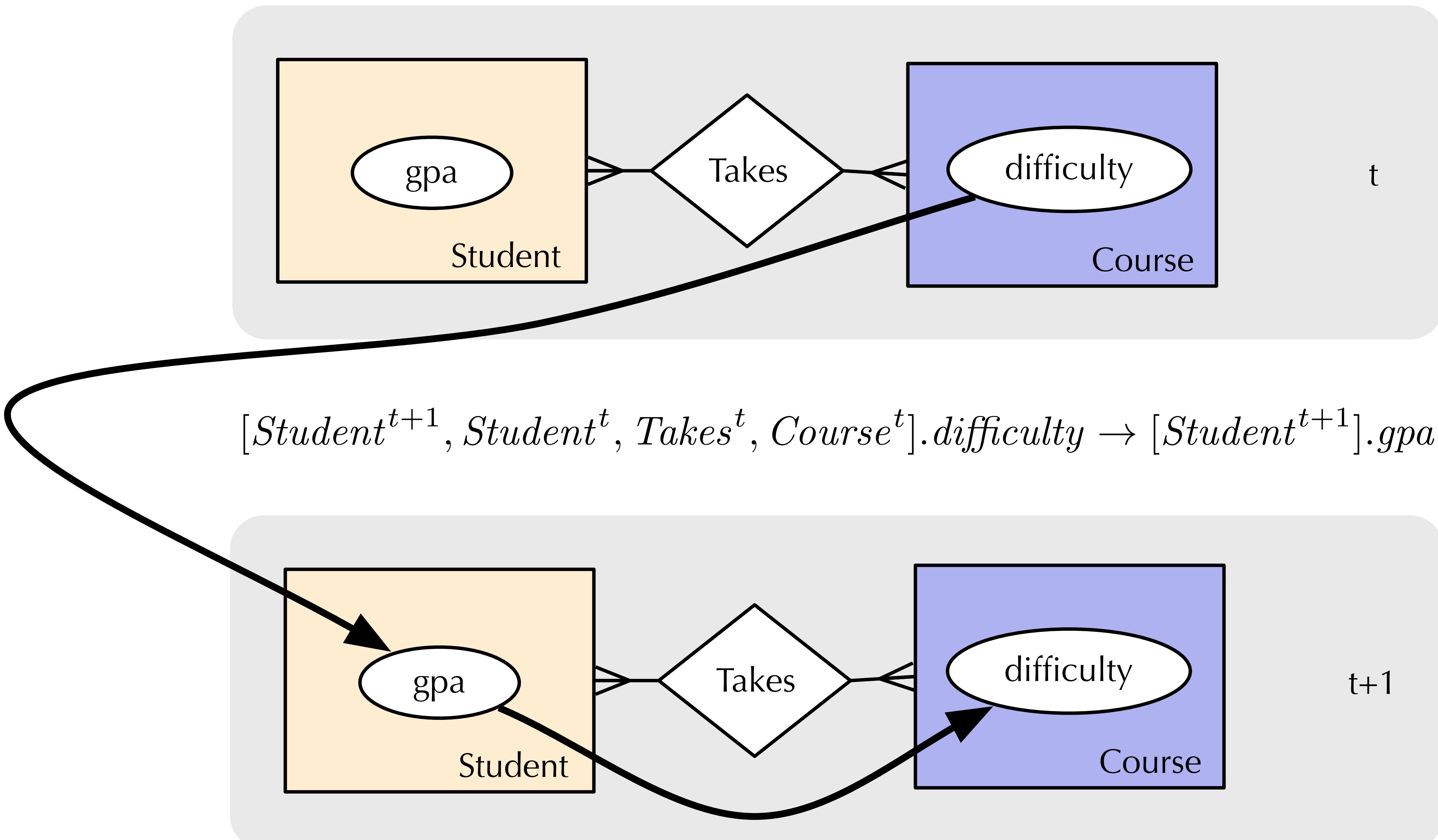
Constraints on temporal relational dependencies

[Course$^{\mathbf{t+1}}$, Takes$^{\mathbf{t+1}}$, Student$^{\mathbf{t+1}}$].gpa → [Student^{t+1}].gpa

[Student$^{\mathbf{t+1}}$, Student$^{\mathbf{t}}$, Takes$^{\mathbf{t}}$, Course$^{\mathbf{t}}$].gpa → [Student^{t+1}].gpa

- First order Markov assumption
  - Cause and effect at most one time step apart
  - The temporal path of the cause contains only two time points.

Temporal relational model

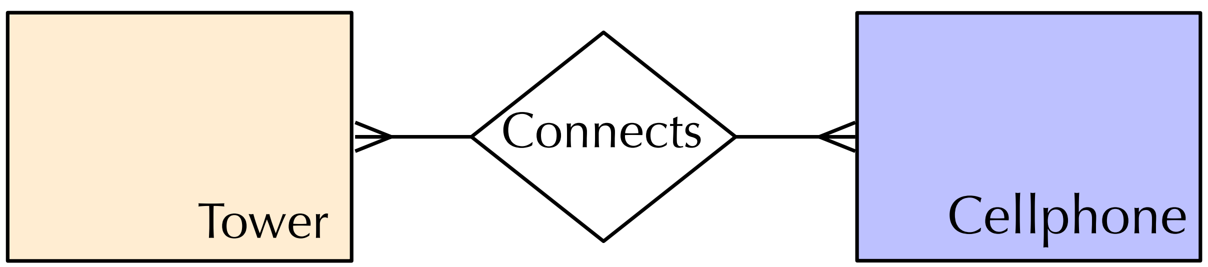


$$[Student^{t+1}, Student^{t}, Takes^{t}, Course^{t}].difficulty \rightarrow [Student^{t+1}].gpa$$

$$[Course^{t+1}, Takes^{t+1}, Student^{t+1}].gpa \rightarrow [Course^{t+1}].difficulty$$

What is the value
of this added expressivity?

Expressivity

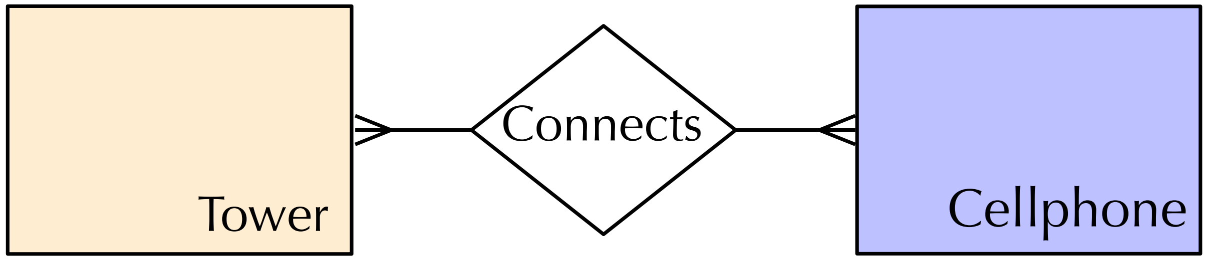


Reality mining domain [1]

- Consider different paths starting at the same item class and ending at the same item class.
- How different are the sets of items reached through these paths?

[1] Eagle, N., Pentland, A., *Reality mining: sensing complex social systems*. Journal of Personal and Ubiquitous Computing, Vol. 10 Issue 4, 2006

Real-data example: Value of expressivity



Path 1: [Tower^{t+1}, Tower^{t}, Connects^{t}, Cellphone^{t}]

"Cellphones that were connected to a tower yesterday"

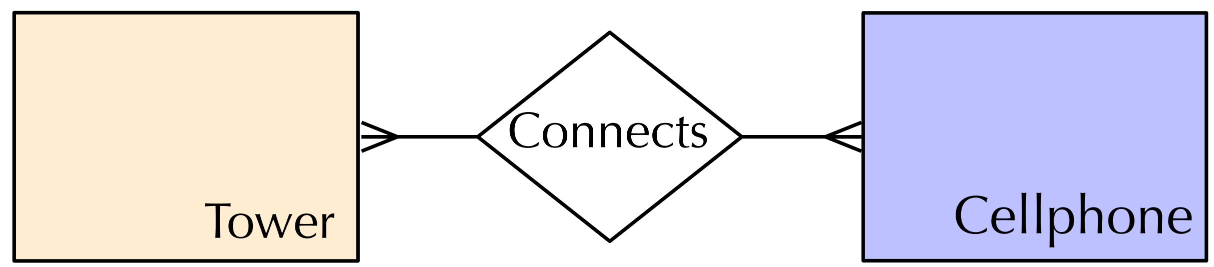
Path 2: [Tower^{t+1}, Connects^{t+1},Connects^{t}, Cellphone^{t}]

"Cellphones that connected to a tower both today and yesterday"

Path 3: [Tower^{t+1}, Connects^{t+1}, Cellphone^{t+1}, Cellphone^{t}]

"The previous state of the phones that are connected to a tower today"

Real-data example: Value of expressivity



Path 1: [Tower^{t+1}, Tower^{t}, Connects^{t}, Cellphone^{t}]

Path 2: [Tower^{t+1}, Connects^{t+1},Connects^{t}, Cellphone^{t}]

Path 3: [Tower^{t+1}, Connects^{t+1}, Cellphone^{t+1}, Cellphone^{t}]

Jaccard distance	**Path 1 vs. Path 3**	**Path 1 vs. Path 2**	**Path 2 vs. Path 3**
mean	0.47	0.31	0.31
median	0.5	0	0

Outline

1. Relational concepts
2. Representation for temporal and relational directed graphical models
3. Temporal relational d-separation
4. TRCD algorithm

d-separation for relational models

- d-separation: a graphical criterion that ties the structure of a Bayesian network to a set of conditional independence facts that hold in the underlying distribution
- d-separation cannot be applied directly to the structure of relational models [2]
- It can be applied to the ground graph, but the size of the ground graph scales with the number of instances
- Solution: Abstract Ground Graphs

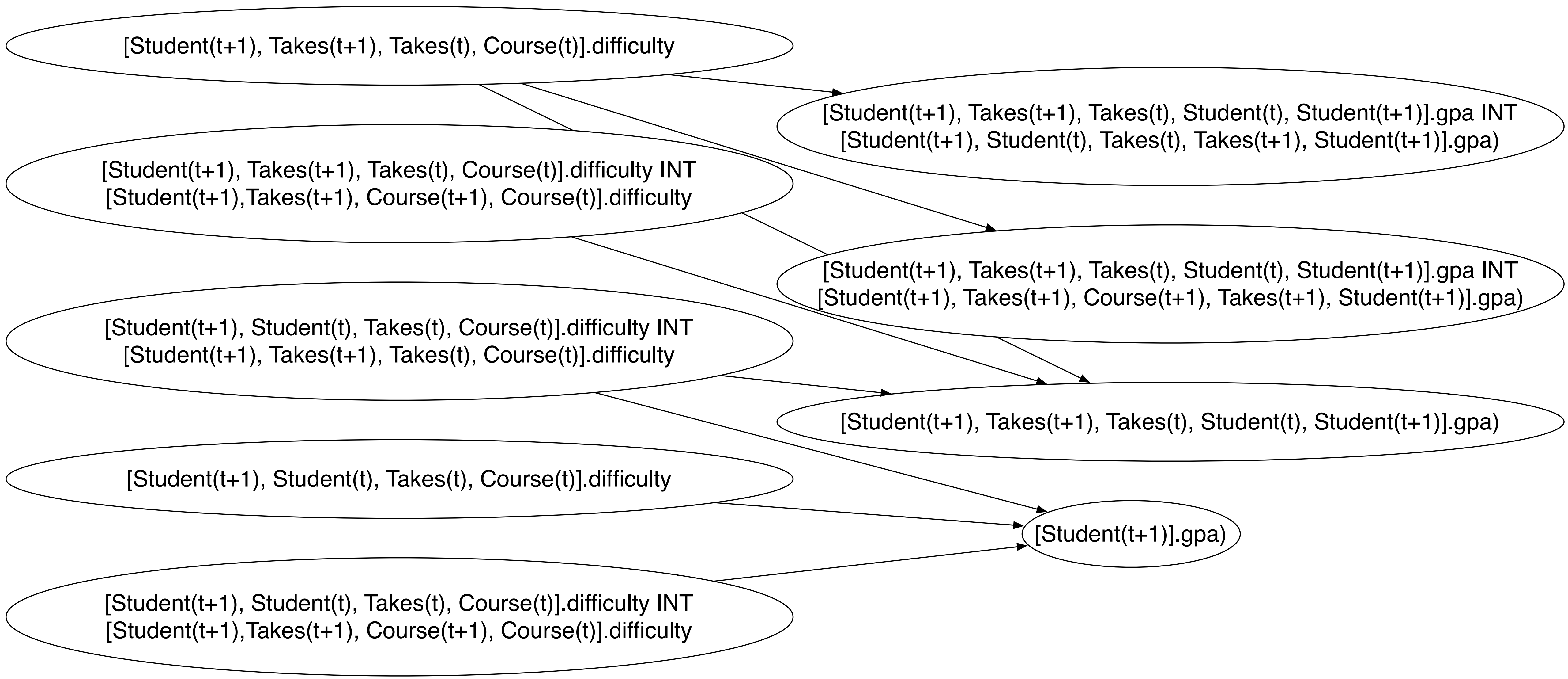
[2] Maier, M., Marazopoulou, K., and Jensen, D. *Reasoning about Independence in Probabilistic Models of Relational Data, arXiv preprint arXiv:1302.4381, 2013.*

Temporal abstract ground graphs

Abstract ground graph: lifted representation that abstracts paths of dependence over all possible ground graphs for a given relational model.

Temporal abstract ground graph: generalizes abstract ground graphs for 2-slice temporal relational models.

Temporal abstract ground graphs



Temporal relational d-separation

The rules of relational d-separation can be applied to abstract ground graphs in order to infer conditional independencies [Maier et al. 2013].

The rules of relational d-separation can be applied to temporal abstract ground graphs in order to infer conditional independencies.

Temporal relational d-separation allows us to derive the set of conditional independence facts that are consistent with the structure of a given temporal relational model.

Outline

1. Relational concepts
2. Temporal relational models
3. Temporal relational d-separation
4. TRCD algorithm

Structure learning

PC algorithm [3]

Entity 1				
X	Y	Z	V	W

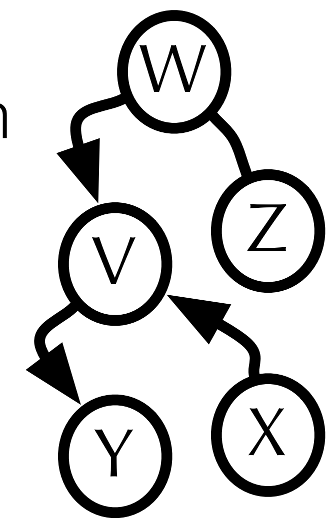
propositional data

Hypothesis tests

$$X \perp\!\!\!\perp W$$
$$W \perp\!\!\!\perp Y \mid \{V\}$$
$$W \not\!\perp\!\!\!\perp V \mid \{X, Y, Z\}$$
$$\cdots$$

conditional independence facts

d-separation + edge orientation

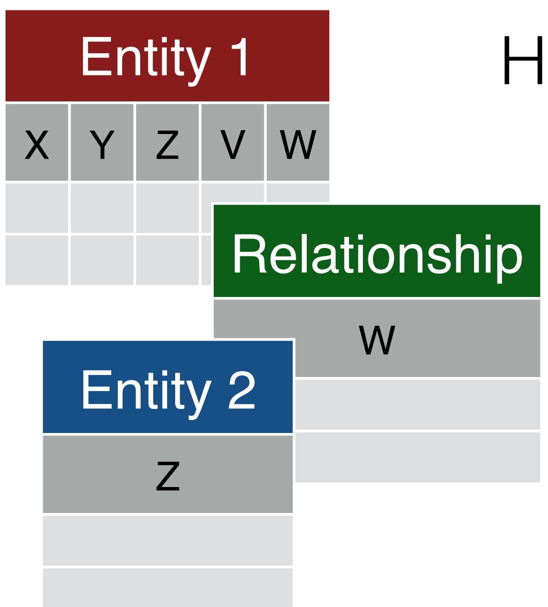


Markov equivalence class

[3] Spirtes, P., Glymour, C., and Scheines, R. *Causation, Prediction, and Search*. MIT Press, 2nd edition, 2000.

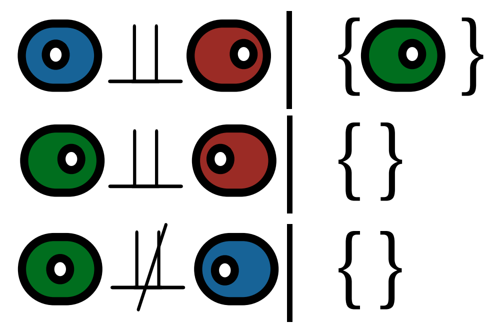
Structure learning

RCD algorithm [4]



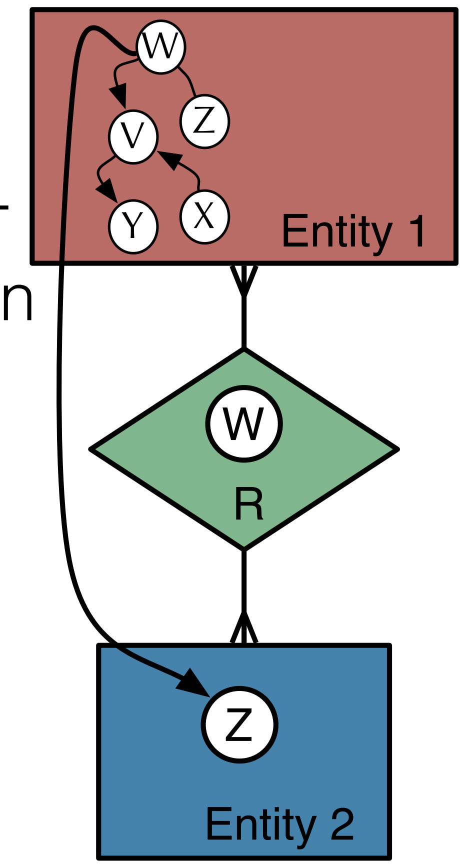
relational
data

Hypothesis tests



conditional
independence facts

relational
d-separation +
edge orientation

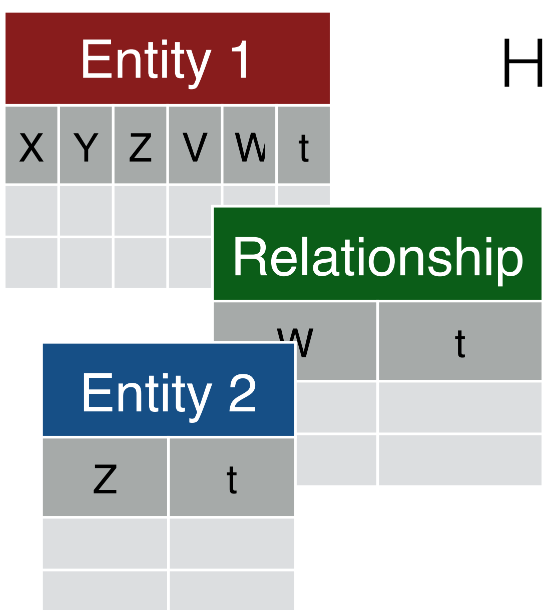


Markov
equivalence class

[4] Maier, M., Marazopoulou, K, Arbour, D, and Jensen, D. *A sound and completete algorithm for learning causal models from relational data. UAI 2013.*

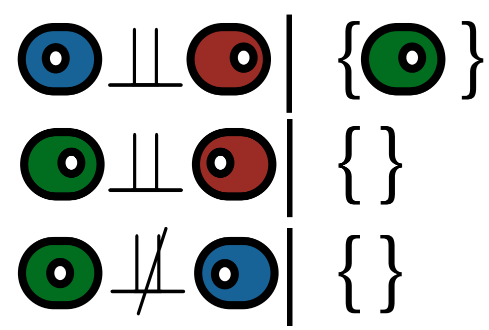
Structure learning

TRCD algorithm



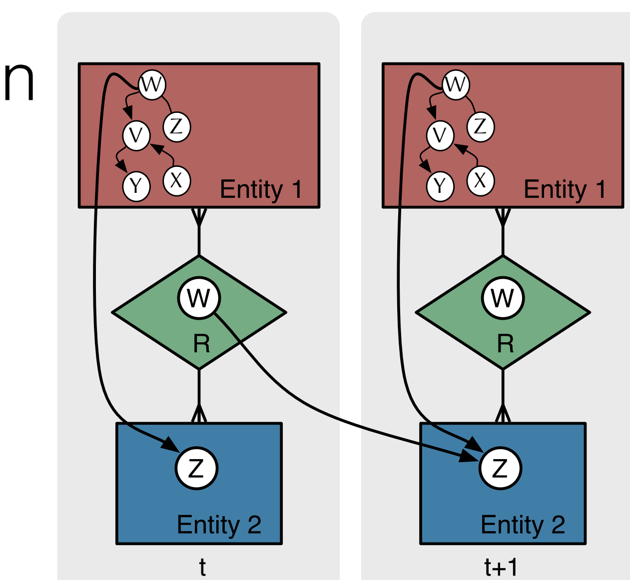
temporal
relational
data

Hypothesis tests



conditional
independence facts

temporal relational
d-separation +
edge orientation



Markov
equivalence class

Temporal relational causal discovery (TRCD)

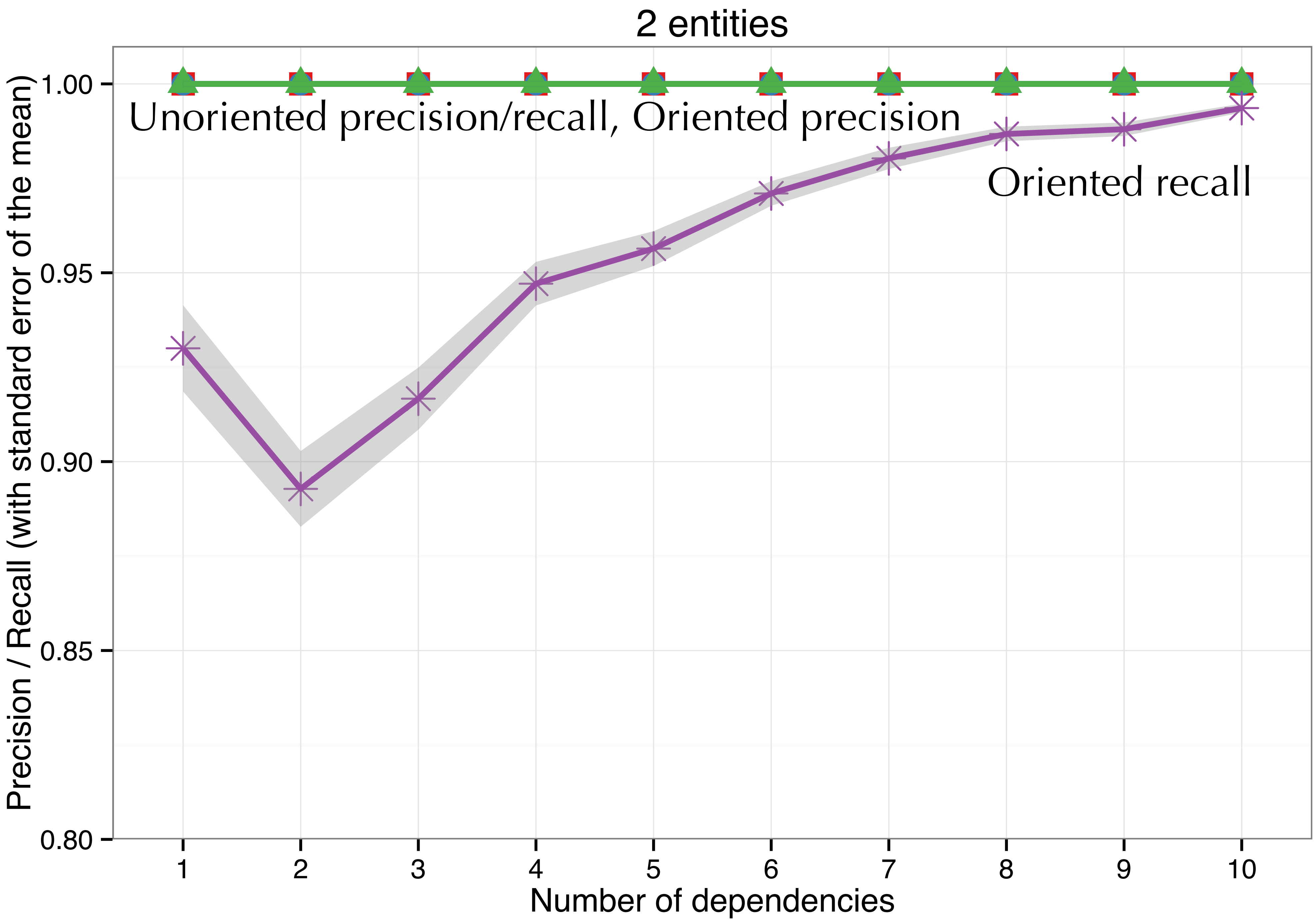
- Constraint-based algorithm to learn the structure of temporal relational causal models from data.

- Extends RCD to operate over a temporal relational model.
  - Phase I learns a set of undirected dependencies.
  - Phase II employs a set of orientation rules to orient the dependencies.

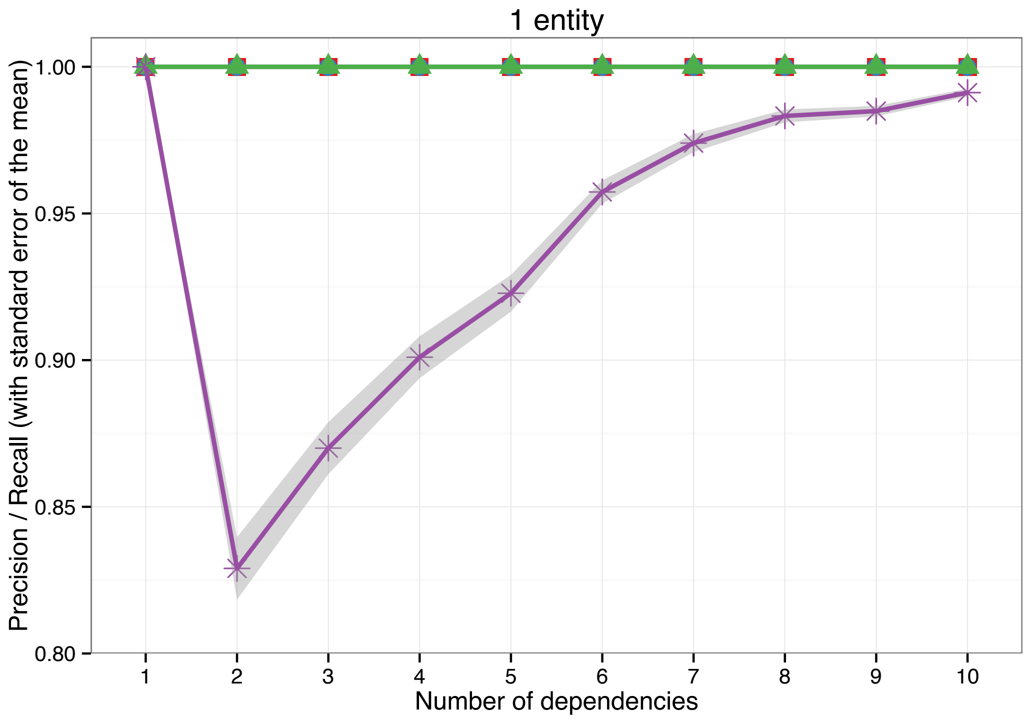
Experiments with an oracle

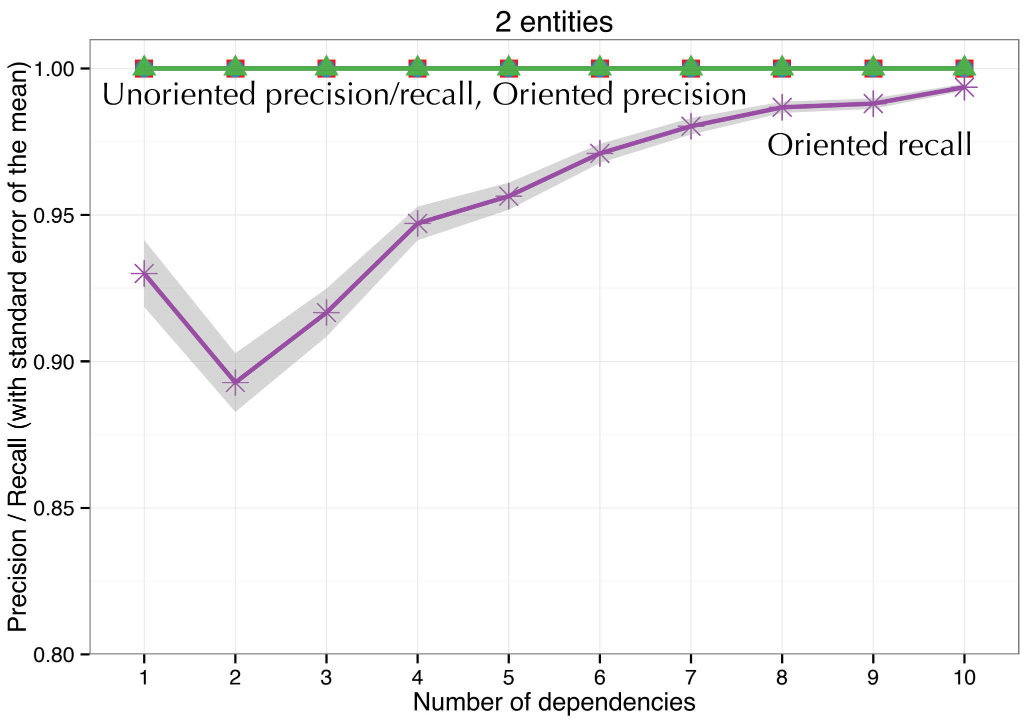
- Generated random schemas (1-3 entities)
- Generated random models (1-10 dependencies)
- 15,000 models in total
- TRCD with a d-separation oracle instead of conditional independence tests

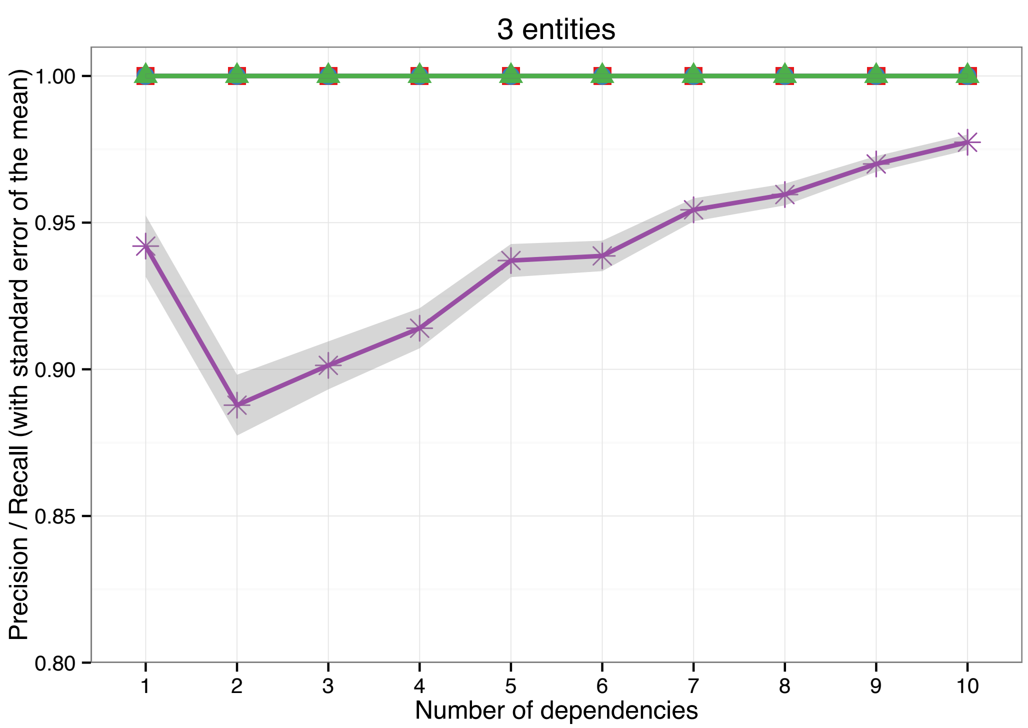
Experiments with an oracle



Experiments with an oracle



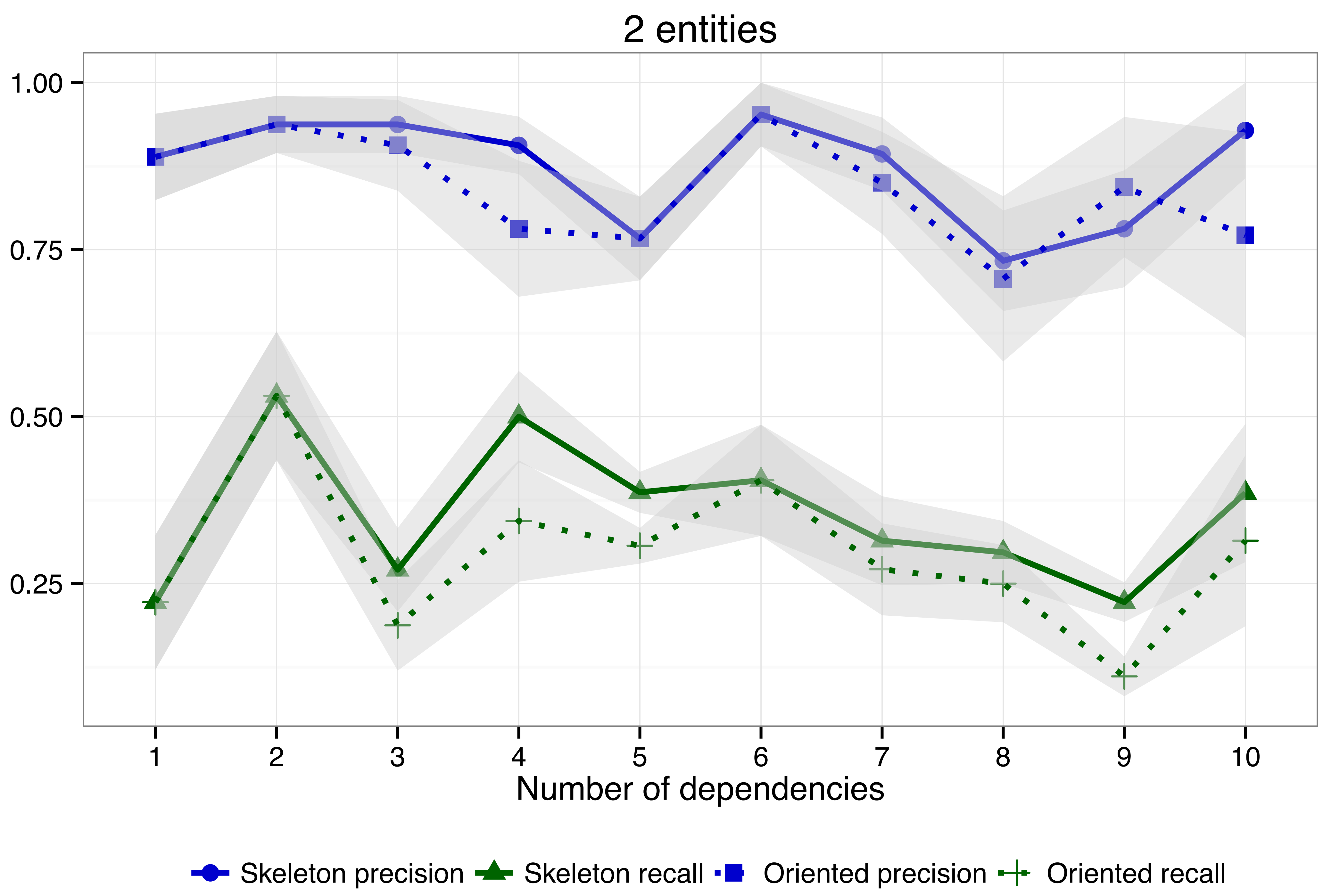




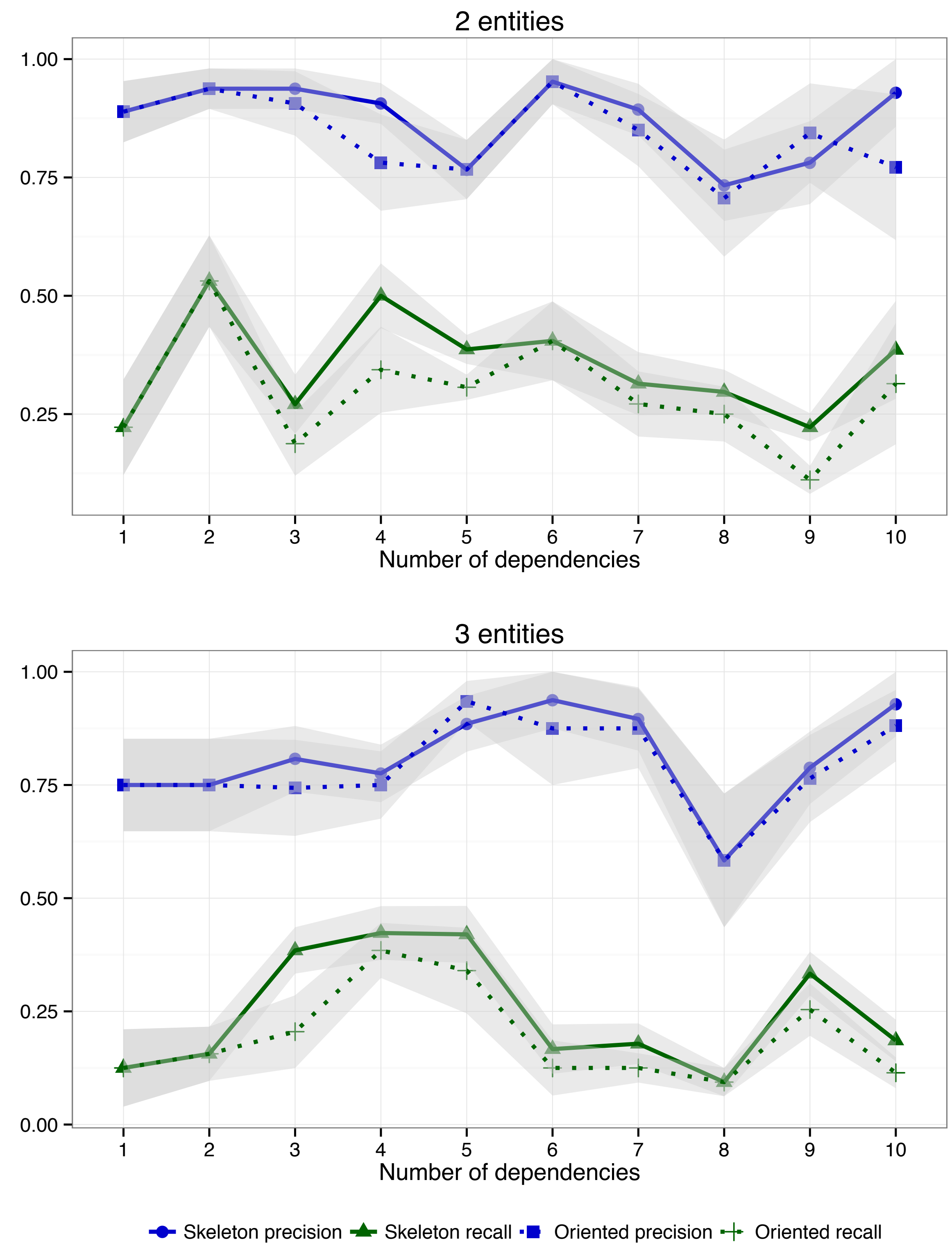
Experiments with synthetic data

- Generated random models
- Generated random temporal relational skeletons with 300 time points
- Generated synthetic data on top of the skeletons
- Ran TRCD using conditional independence tests

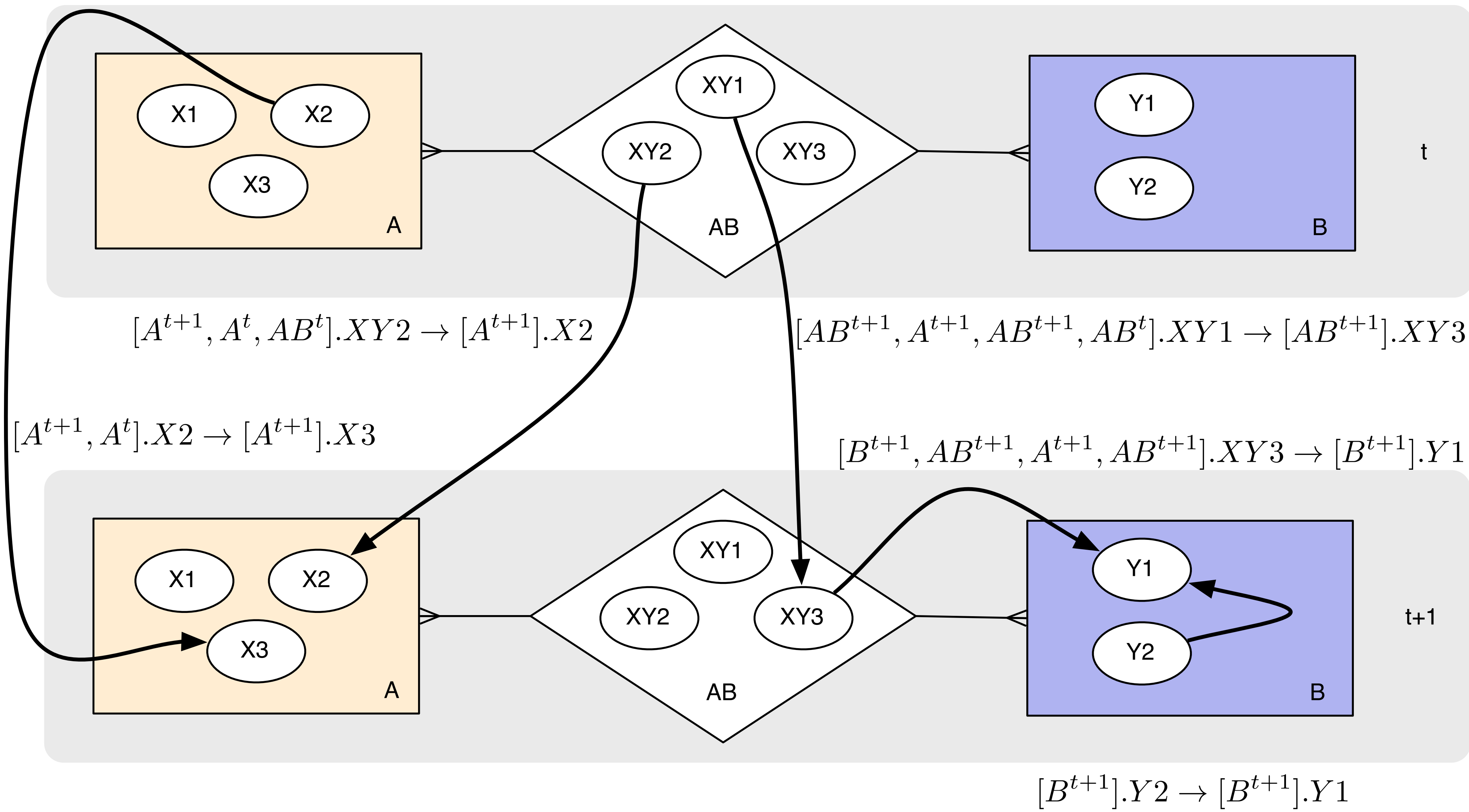
Experiments with synthetic data



Experiments with synthetic data

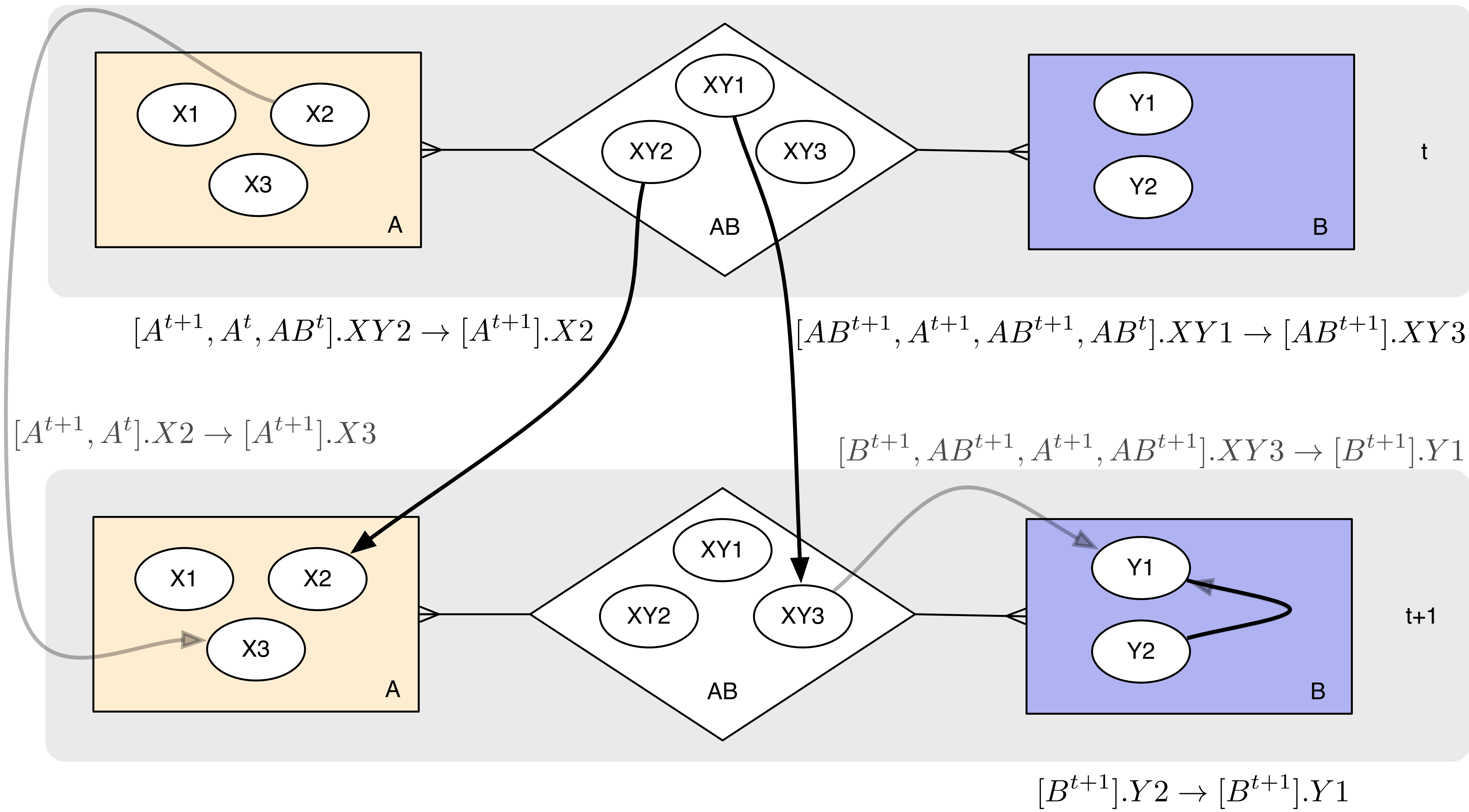


Comparison with RCD



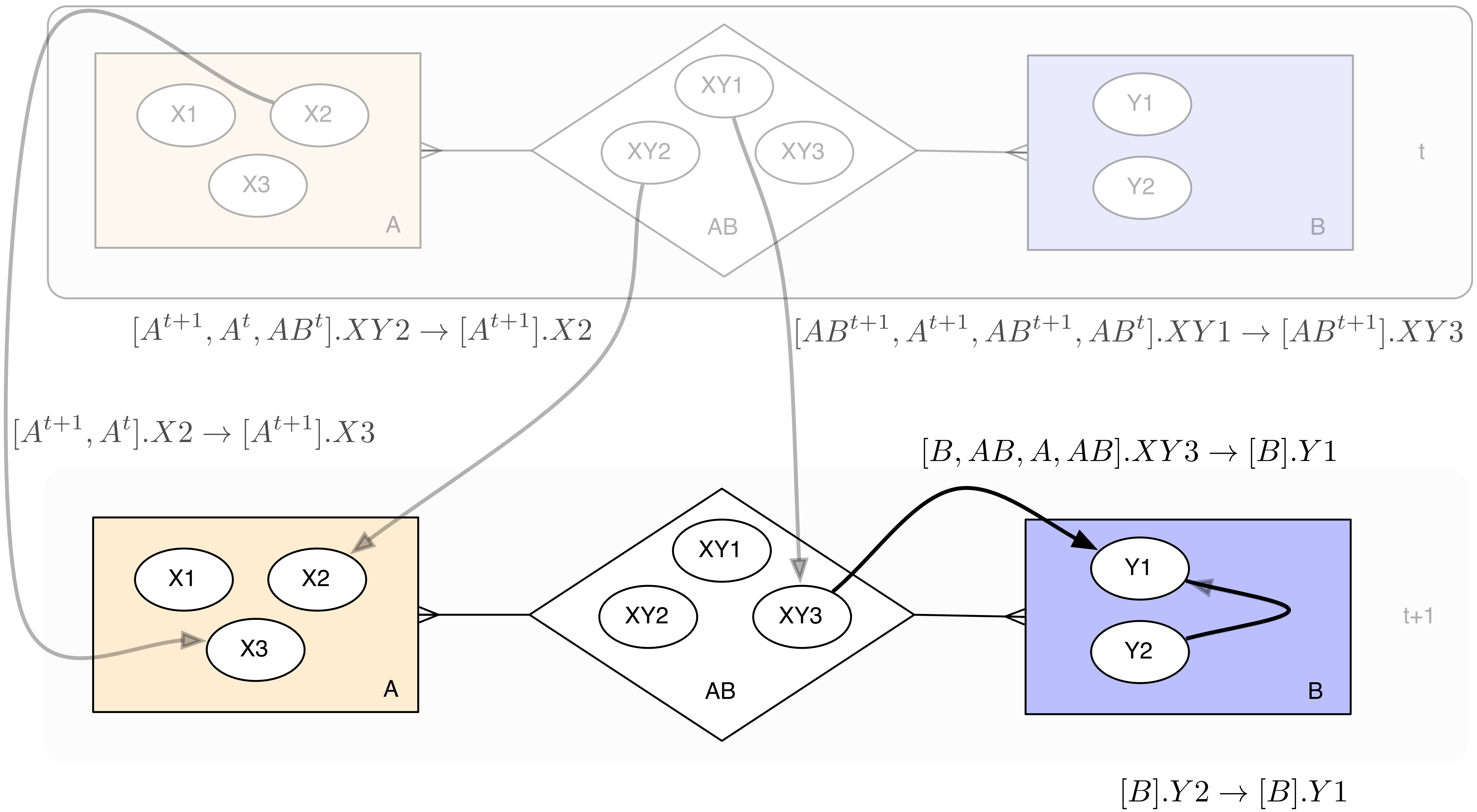
True model

Comparison with RCD



Model learned by TRCD

Comparison with RCD



Model learned by RCD

Summary

- Representation: Directed graphical model that supports time and relational information
- Temporal relational d-separation
- TRCD algorithm: constraint-based algorithm to learn the structure of a directed temporal relational model

Future work

- Further experiments with more realistic/real data.
- Relax assumptions

Questions?

Thank you!