Typing Assumptions Improve Identification in Causal Discovery



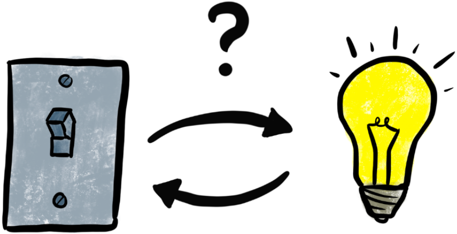


Philippe Brouillard [1,2]
Perouz Taslakian [3]
Alexandre Lacoste [1]
Sébastien Lachapelle [2]
Alexandre Drouin [1]

[1] ServiceNow Research
[2] Mila & DIRO, Université de Montréal
[3] Samsung AI Center Montreal

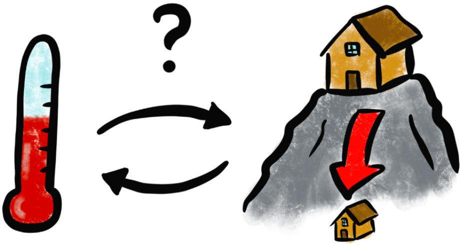
Motivation

Can a light bulb change the state of a switch?



Motivation

Can the temperature of a city alter its altitude?



Motivation

- As humans, we understand the implausibility of causal relationships between certain types of entities.
- We use prior knowledge to generalize causal relation between similar entities. [Griffiths et al., 2011, Schulz & Gopnik, 2004, Gopnik & Sobel, 2000]
- We believe that this is a key property of intelligent agents.

Motivation

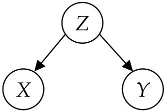
- As humans, we understand the implausibility of causal relationships between certain types of entities.
- We use prior knowledge to generalize causal relation between similar entities. [Griffiths et al., 2011, Schulz & Gopnik, 2004, Gopnik & Sobel, 2000]
- We believe that this is a key property of intelligent agents.

➔ **Goal**: Integrate this “common sense” to causal discovery.

Causal Discovery

Task: learn the structure of G based on observations from P_X.

X	Y	Z
1.21	1.58	0.33
1.50	1.84	0.51
⋮	⋮	⋮
0.96	1.07	0.11

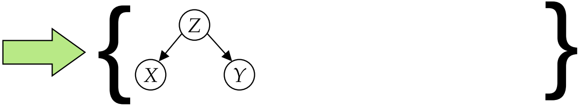


$$p(z)\, p(x|z)\, p(y|z)$$

Markov Equivalence Class

Unfortunately, from observational data, one can only retrieve a set of equivalent DAGs called the **Markov Equivalence Class** (MEC).

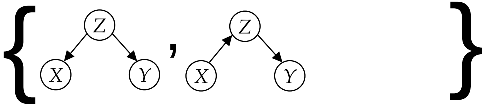
X	Y	Z
1.21	1.58	0.33
1.50	1.84	0.51
⋮	⋮	⋮
0.96	1.07	0.11



Markov Equivalence Class

Unfortunately, from observational data, one can only retrieve a set of equivalent DAGs called the **Markov Equivalence Class** (MEC).

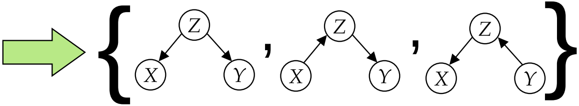
X	Y	Z
1.21	1.58	0.33
1.50	1.84	0.51
⋮	⋮	⋮
0.96	1.07	0.11



Markov Equivalence Class

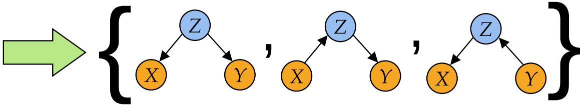
Unfortunately, from observational data, one can only retrieve a set of equivalent DAGs called the **Markov Equivalence Class** (MEC).

X	Y	Z
1.21	1.58	0.33
1.50	1.84	0.51
⋮	⋮	⋮
0.96	1.07	0.11



In our proposed setting, each vertex has an associated type (possibly given by an expert). We call these graphs **t-DAGs**.

X	Y	Z
1.21	1.58	0.33
1.50	1.84	0.51
⋮	⋮	⋮
0.96	1.07	0.11

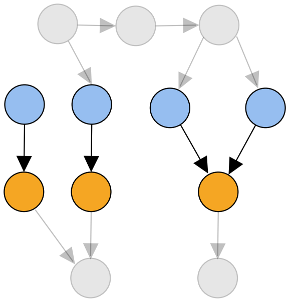


Types are represented by node's color.

Type consistency: constraints on type interactions

Assumption: all edges between a pair of types are oriented in the same direction.

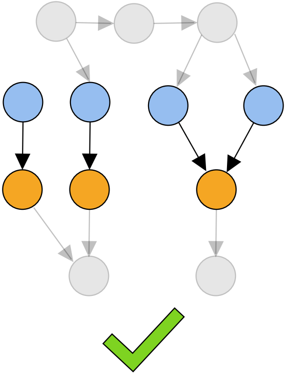
We call t-DAGs that satisfy this condition **consistent t-DAGs**.



Type consistency: constraints on type interactions

Assumption: all edges between a pair of types are oriented in the same direction.

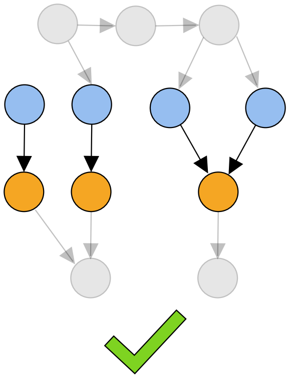
We call t-DAGs that satisfy this condition **consistent t-DAGs**.

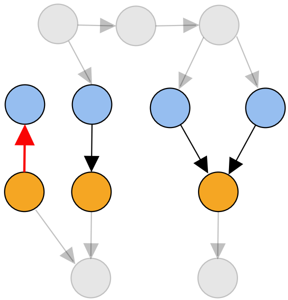


Type consistency: constraints on type interactions

Assumption: all edges between a pair of types are oriented in the same direction.

We call t-DAGs that satisfy this condition **consistent t-DAGs**.

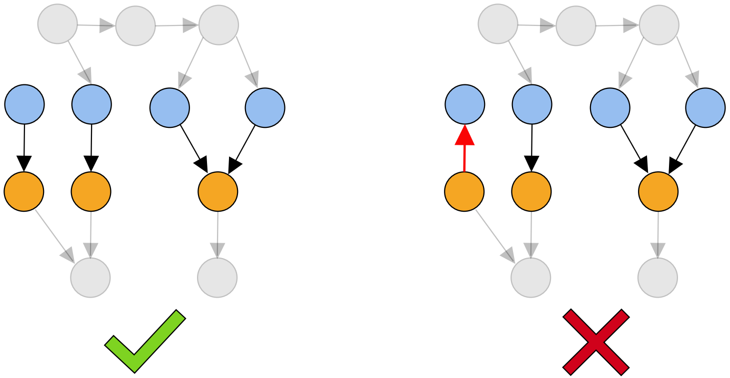




Type consistency: constraints on type interactions

Assumption: all edges between a pair of types are oriented in the same direction.

We call t-DAGs that satisfy this condition **consistent t-DAGs**.

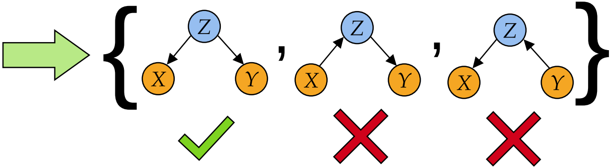


We call the set of edges between a given pair of types a **t-edge**.

t-MEC: an equivalence class for consistent t-DAGs

With this assumption, the size of the MEC can be greatly reduced by removing t-DAGs that violate type consistency. We call this equivalence class a **t-MEC**.

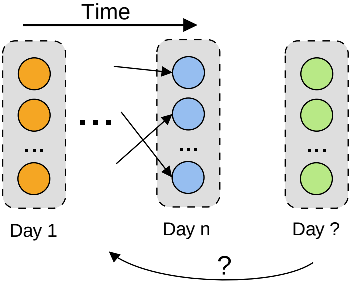
X	Y	Z
1.21	1.58	0.33
1.50	1.84	0.51
⋮	⋮	⋮
0.96	1.07	0.11



Examples: beyond tiered background knowledge

Types given by an expert. More general than tiered background knowledge.

Particularly useful when it is reasonable to assume that entities interact in a directional manner.

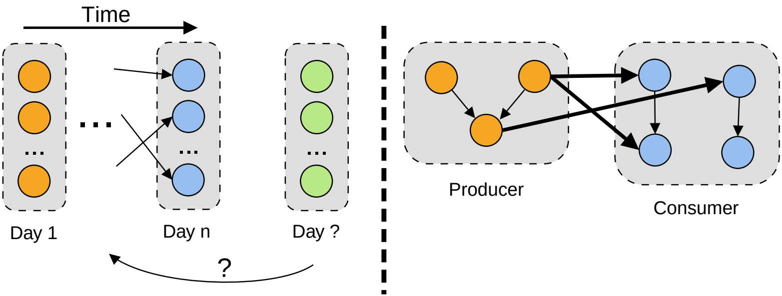


Ambiguous tiered background knowledge

Examples: beyond tiered background knowledge

Types given by an expert. More general than tiered background knowledge.

Particularly useful when it is reasonable to assume that entities interact in a directional manner.



Ambiguous tiered background knowledge

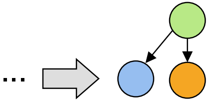
Interacting entities

Typing assumptions improve identification in causal discovery

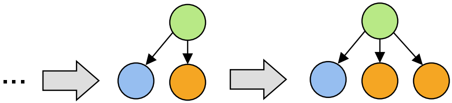
Theorem

For a random sequence of t-DAGs with a growing number of vertices but a fixed set of types, the ***number of unoriented t-edges converges to 0*** *exponentially fast.*

Random sequence of growing t-DAGs



Random sequence of growing t-DAGs

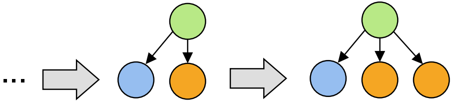


At each step:

- add a node with a random type

$$\begin{bmatrix} 0.2 & 0.4 & 0.4 \end{bmatrix}$$

Random sequence of growing t-DAGs

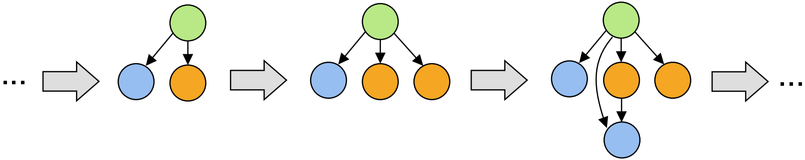


At each step:

- add a node with a random type
- add edges to the existing nodes following probabilities based on the types

$$\begin{bmatrix} 0 & 0.8 & 0.7 \\ 0 & 0 & 0.3 \\ 0 & 0 & 0 \end{bmatrix}$$

Random sequence of growing t-DAGs



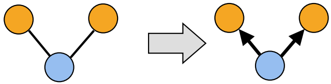
At each step:

- add a node with a random type
- add edges to the existing nodes following probabilities based on the types

$$\begin{bmatrix} 0 & 0.8 & 0.7 \\ 0 & 0 & 0.3 \\ 0 & 0 & 0 \end{bmatrix}$$

Proof sketch

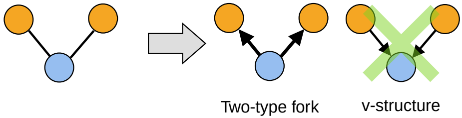
- There exists a structure, called a two-type fork, that has to be oriented due to type consistency.



Two-type fork

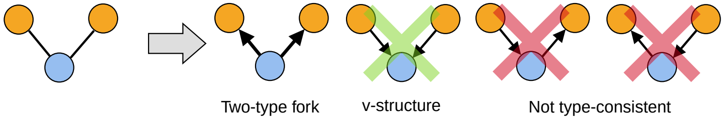
Proof sketch

- There exists a structure, called a two-type fork, that has to be oriented due to type consistency.



Proof sketch

- There exists a structure, called a two-type fork, that has to be oriented due to type consistency.



- As the graph grows, the probability of observing it converges to 1 and thus all edges of the t-DAG are oriented.

Typing assumptions improve identification in causal discovery

Theorem

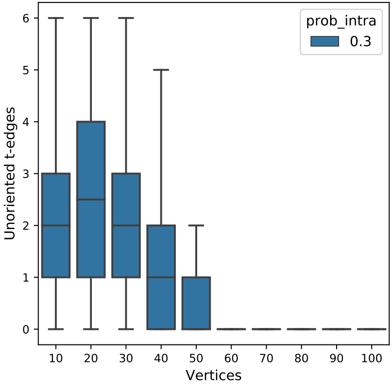
For a random sequence of t-DAGs with a growing number of vertices but a fixed set of types, the ***number of unoriented t-edges converges to 0*** *exponentially fast.*

Typing assumptions improve identification in causal discovery

Theorem

For a random sequence of t-DAGs with a growing number of vertices but a fixed set of types, the ***number of unoriented t-edges converges to 0*** *exponentially fast.*

For 10 types, we observe this with medium-sized graph ($|V| > 60$).

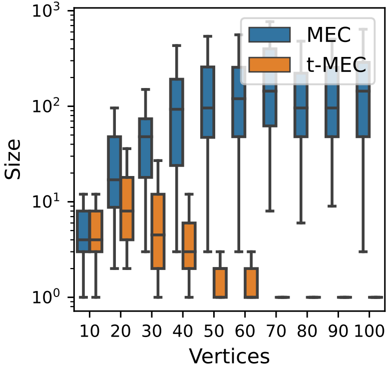


Typing assumptions improve identification in causal discovery

Corollary

If there are no edge between variables of the same type, then the ***size of the t-MEC converges to 1*** *exponentially fast.*

The size of the t-MEC converges to 1 for medium-sized graphs.

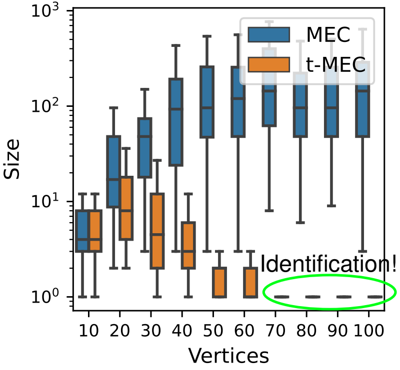


Typing assumptions improve identification in causal discovery

Corollary

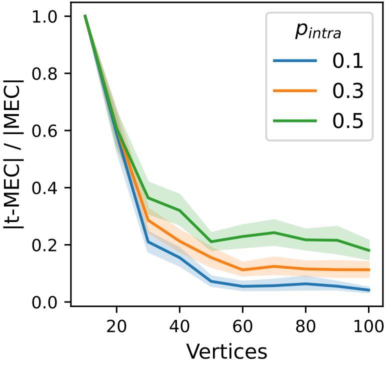
If there are no edge between variables of the same type, then the ***size of the t-MEC converges to 1*** *exponentially fast.*

The size of the t-MEC converges to 1 for medium-sized graphs.



Open question

When intra-type edges are allowed, we observed that the ratio |t-MEC| / |MEC| seems to converge to a value that depends on the density of such edges.



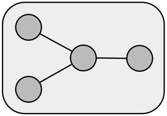
Causal discovery algorithms for t-DAGs

- Since |t-MEC| < |MEC|, we want algorithms that are **consistent** to it, i.e. will recover the equivalence class in the population case.
- PC [Spirtes et al., 2000] is consistent w.r.t. the MEC.
- We propose three extensions of PC that are t-MEC-consistent.

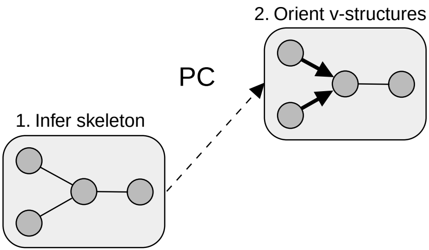
A t-MEC consistent extension of PC

PC

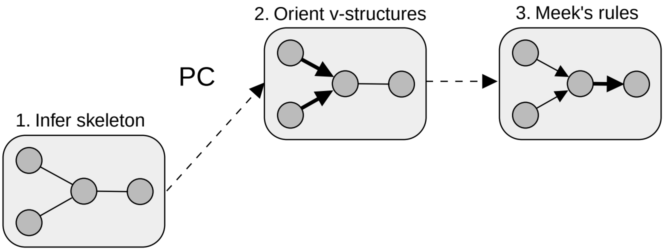
1. Infer skeleton



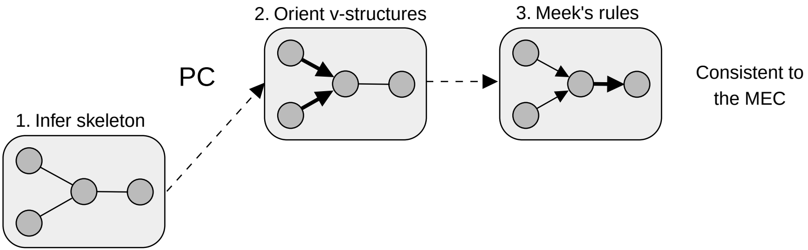
A t-MEC consistent extension of PC



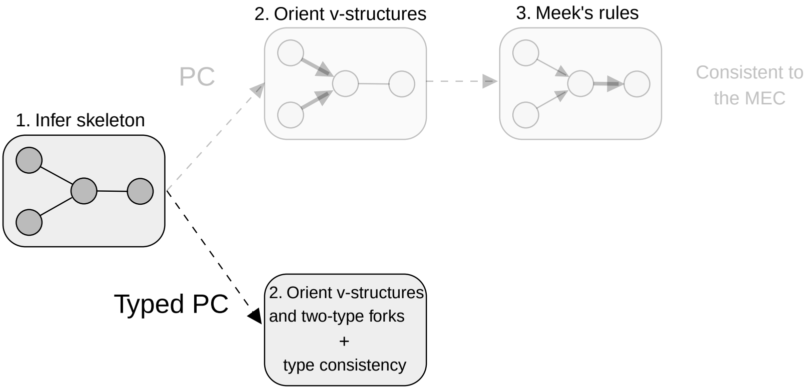
A t-MEC consistent extension of PC



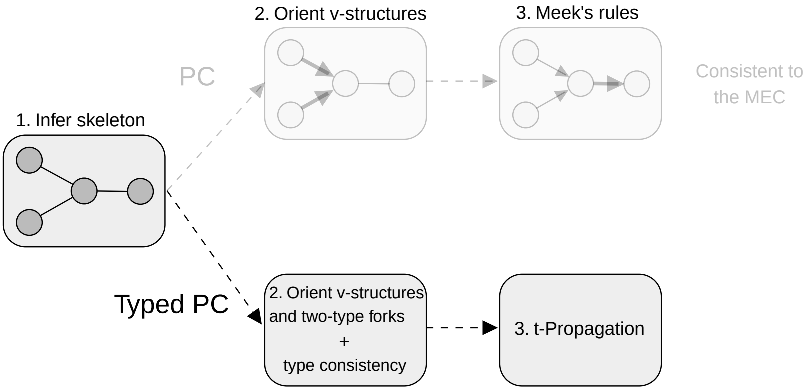
A t-MEC consistent extension of PC



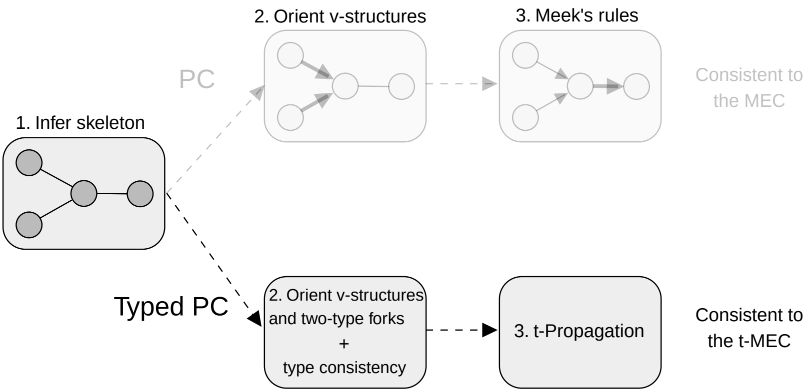
A t-MEC consistent extension of PC



A t-MEC consistent extension of PC



A t-MEC consistent extension of PC



A t-MEC consistent extension of PC

Could we apply t-Propagation to the output of PC?

A t-MEC consistent extension of PC

Could we apply t-Propagation to the output of PC?

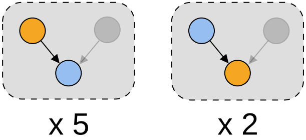
➔ With finite sample size, it may lead to inconsistent t-DAGs!

A t-MEC consistent extension of PC

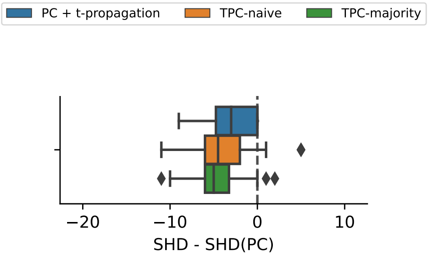
Could we apply t-Propagation to the output of PC?

➔ With finite sample size, it may lead to inconsistent t-DAGs!

TPC-majority:
Orient t-edges based on the most frequent orientation in v-structures and two-typed forks

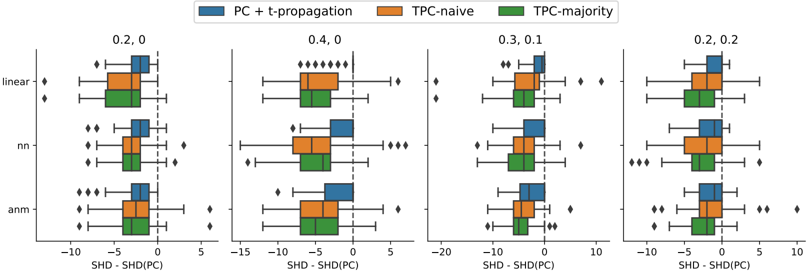


Typed methods improve over PC (simulated data)



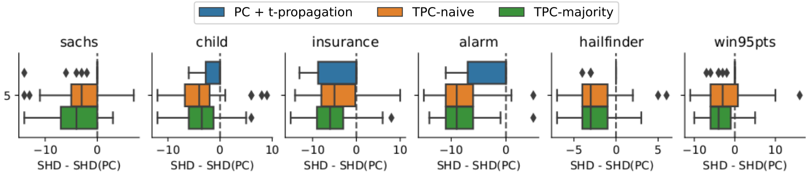
Improvement in SHD over PC (lower is better) on 20-node graphs.

Typed methods improve over PC (simulated data)



Improvement in SHD over PC (lower is better) on 20-node graphs.

Typed methods improve over PC (pseudo-real data)



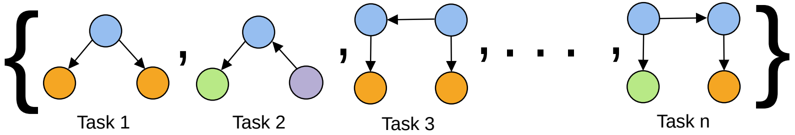
We artificially partition data sets from the Bayesian Network Repository into types.

Conclusion

- This work shows that our typing assumptions can help reduce the size of equivalence classes and thus improve identification in causal discovery.
- Typing assumptions are likely to be a key component of causal reasoning in intelligent agents.
- Lead to several interesting future directions of research.

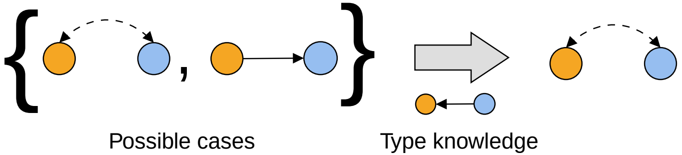
Future directions: multi-task causal discovery

Multi-task causal discovery: Can we deal with multiple environments where types overlap?



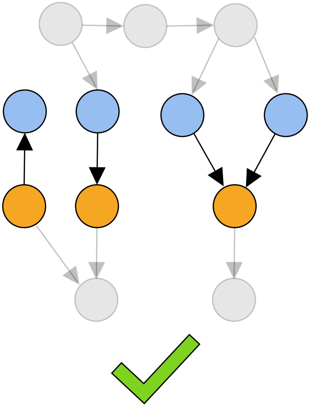
Future directions: relaxing our assumptions

Latent confounders: what can we gain? Can we adapt FCI [Spirtes et al., 1995].



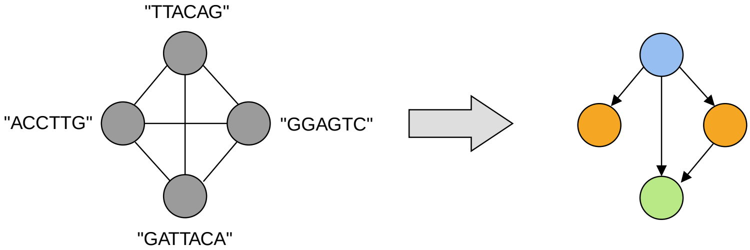
Future directions: relaxing our assumptions

Soft type-consistency: direction of edges between pairs of type is probabilistic.



Future directions: learning the types

Can we learn the types and the graph simultaneously from metadata?



Thank you!

Philippe Brouillard [1,2]

Perouz Taslakian [3]

Alexandre Lacoste [1]

Sébastien Lachapelle [2]

Alexandre Drouin [1]

Come to our **poster** to learn more about:

- our assumptions based on variable types,
- the theoretical results,
- algorithms to incorporate these assumptions in causal discovery.

References

Gopnik, A. & Sobel, D. M. (2000).
Detecting blickets: How young children use information about novel causal powers in categorization and induction.
Child development, 71(5), 1205–1222.

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011).
Bayes and blickets: Effects of knowledge on causal induction in children and adults.
Cognitive Science, 35(8), 1407–1455.

Schulz, L. E. & Gopnik, A. (2004).
Causal learning across domains.
Developmental psychology, 40(2), 162.

Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000).
Causation, prediction, and search.
MIT press.

Spirtes, P. L., Meek, C., & Richardson, T. S. (1995).
Causal inference in the presence of latent variables and selection bias.
arXiv preprint arXiv:1302.4983.