Typing Assumptions Improve Identification in Causal Discovery

Philippe Brouillard [1,2]

Perouz Taslakian [1]

Alexandre Lacoste [1]

Sébastien Lachapelle [2]

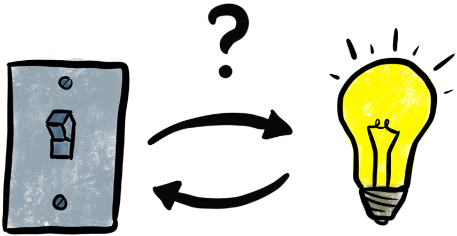
Alexandre Drouin [1]

[1] Element AI, a ServiceNow company
[2] Mila & DIRO, Université de Montréal

ELEMENT AI
a ServiceNow Company

Motivation

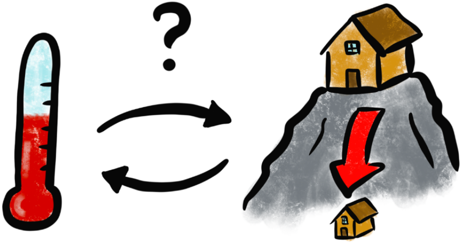
Can a light bulb change the state of a switch?



ELEMENT AI
a ServiceNow Company

Motivation

Can the temperature of a city alter its altitude?



ELEMENT AI
a ServiceNow Company

Motivation

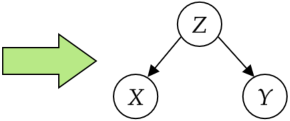
- As humans, we understand the implausibility of causal relationships between certain types of entities.
- In fact, most of the time, we use prior knowledge to generalize causal relation between similar entities.
  [Griffiths et al., 2011, Schulz & Gopnik, 2004, Gopnik & Sobel, 2000]

ELEMENT AI
a ServiceNow Company

Causal Discovery

The task of causal discovery consists of learning 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

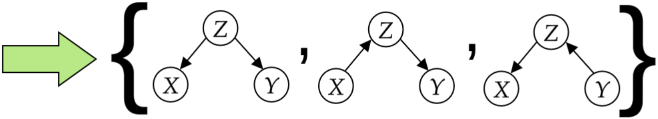


$$p(z)\, p(x|z)\, p(y|z)$$

Markov Equivalence Class

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

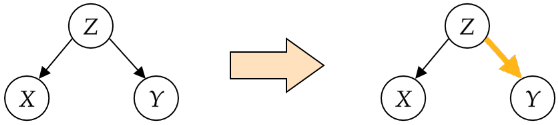
X	Y	Z
1.21	1.58	0.33
1.50	1.84	0.51
⋮	⋮	⋮
0.96	1.07	0.11



Why identification matters?

The correct model is "robust" to intervention.

With the correct model, **causal mechanisms are independent**:
only one (or a few) conditional changes when an intervention is applied.



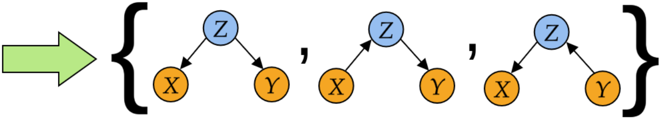
$p(z)\, p(x|z)\, p(y|z)$ $\quad\Rightarrow\quad$ $p(z)\, p(x|z)\, \tilde{p}(y|z)$

ELEMENT AI
a ServiceNow Company

Variable types

In our new setting, each vertex has an associated type (that might have been given by an expert). We call these graphs **t-DAGs**.

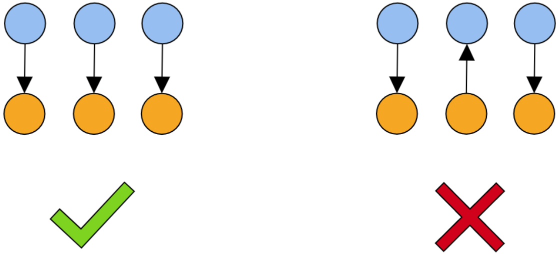
X	Y	Z
1.21	1.58	0.33
1.50	1.84	0.51
⋮	⋮	⋮
0.96	1.07	0.11



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**.

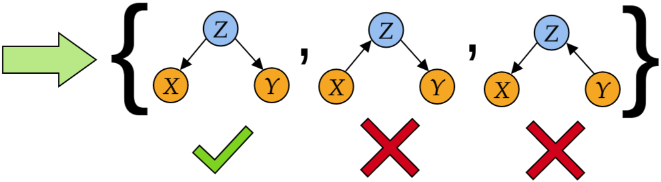


ELEMENT AI
a ServiceNow Company

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

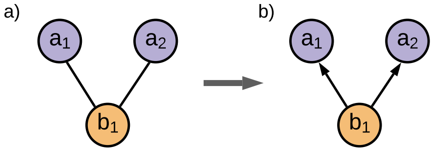
With this assumption, the size of the MEC can be greatly reduced by removing t-DAGs that violate type consistency. We call this equivalence class a **t-MEC**.

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



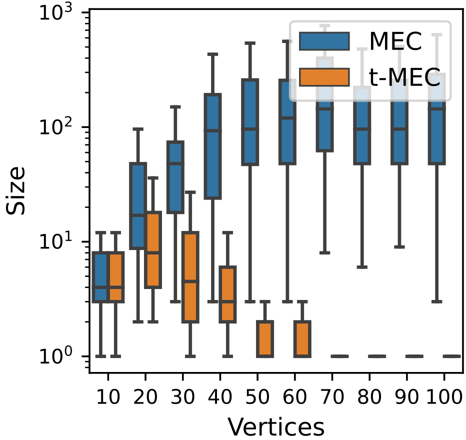
Identification guarantees for random graphs

- We show that there exists conditions under which our assumptions lead to benefits in identification.
- For a random sequence of t-DAGs with a fixed number of types, the size of the t-MEC **converge to a singleton** exponentially fast with **the number of vertices** (measured variables).
- **Proof sketch:** There exists a structure, called a two-type fork, that forces the orientation of edges 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.



Two-type fork

Empirical validation



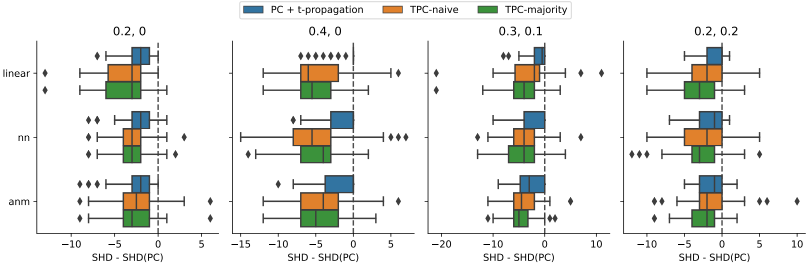
ELEMENT AI
a ServiceNow Company

Causal discovery algorithms for t-DAGs

- Since the t-MEC is often much smaller than the MEC, it is important to find algorithms that are consistent to it.
- Standard causal discovery algorithms, like PC [Spirtes et al., 2000], are usually consistent w.r.t. the MEC
- We propose three t-MEC-consistent algorithms, which are extensions of the classical PC algorithm.

ELEMENT AI
a ServiceNow Company

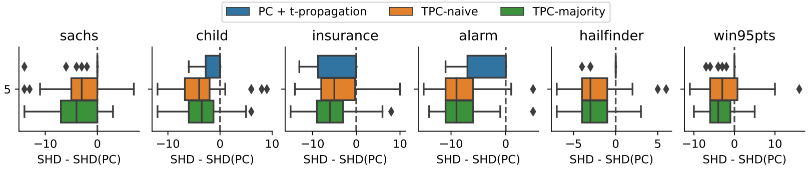
Structure learning experiments: simulated data sets



- Report the SHD improvement (lower is better) on 20-node graphs.
- All t-MEC-consitent algorithms outperform the MEC-consistent PC algorithm.

ELEMENT AI
a ServiceNow Company

Structure learning experiments: pseudo-real datasets



- Use data sets from the Bayesian Network Repository.
- To assign types to variables, we randomly partition their topological ordering into groups of expected size 5.

Conclusion

This work shows that our typing assumptions can help reduce the size of the MEC and thus help in causal discovery when our assumptions hold.

Future work:

- Explore practical applications (e.g., Alzheimer's disease data [Shen et al., 2020])
- Learn types automatically from metadata related to each variable
- Propagate knowledge of type interactions across multiple tasks

ELEMENT AI
a ServiceNow Company

Thank you!

Philippe Brouillard [1,2]

Perouz Taslakian [1]

Alexandre Lacoste [1]

Sébastien Lachapelle [2]

Alexandre Drouin [1]

Come to the **poster 36** to learn more about:

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

ELEMENT AI
a ServiceNow Company

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.

Shen, X., Ma, S., Vemuri, P., & Simon, G. (2020).
Challenges and opportunities with causal discovery algorithms: application to alzheimer's pathophysiology.
Scientific reports, 10(1), 1–12.

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

ELEMENT AI
a ServiceNow Company