

## Internship Proposal (M2)

### Causal discovery in complex data

- Research Institute: Sorbonne University, LIP6, LCQB
- Location: 4, place Jussieu, Paris, France
- Supervision: Pierre-Henri Wuillemin (LIP6, SU), Nataliya Sokolovska (LCQB, SU)
- How to apply? Contact (attach your CV) [pierre-henri.wuillemin@lip6.fr](mailto:pierre-henri.wuillemin@lip6.fr) and [nataliya.sokolovska@sorbonne-universite.fr](mailto:nataliya.sokolovska@sorbonne-universite.fr)

**Context** The decision what is a “cause” and what is an “effect” is generally made taking into account several principles based on moral aspects, background knowledge about normal cases to identify deviations, temporal relations. All these premises may not be found in observed data, making them insufficient (*Pearl, 2000*). In real-life applications, we are often challenged to infer what is a cause and what is an effect from static non-temporal data.

**Objectives** The 3off2 is a causal discovery algorithm (*Affeldt et al., 2016*) which merges the principles of constraint-based and score-based approaches to reconstruct the causal graphical model from data. MIIC (Multivariate Information based Inductive Causation, *Sella, 2018*), its extension, disentangles direct from indirect effects amongst correlated variables, including cause-effect relationships and the effect of unobserved latent causes.

In this project, we would like to go further, and to propose an efficient generalisation of the 3off2 and the MIIC methods for complex data. There are several challenges to tackle: identification of latent (hidden, non-observed) variables and their characterisation (e.g., identification of their domain size), reconstruction of networks taking the underlying communities (or clusters) into consideration. Another direction is a personalised causal inference, considered, e.g., in (*Nilforoshan et al., 2023*).

**Background** Master in Computer Science/Statistics/Mathematics/Bioinformatics. The student will test the existing methods, develop and implement novel approaches. Knowledge of a programming language (Python) is required.

**Supervision** The student will be supervised by Pierre-Henri Wuillemin (LIP6, SU) and Nataliya Sokolovska (LCQB, SU).

#### References

1. Pearl, Judea. *The art and science of cause and effect. Causality: models, reasoning and inference*, pp. 331–358, 2000.
2. Affeldt, S., Verny, L. and Isambert, H. *3off2: A network reconstruction algorithm based on 2-point and 3-point information statistics*. BMC Bioinformatics 17, S12, 2016.
3. Nadir Sella, Louis Verny, Guido Uguzzoni, Séverine Affeldt, Hervé Isambert: *MIIC online: a web server to reconstruct causal or non-causal networks from non-perturbative data*. Bioinform. 34(13): 2311-2313, 2018.
4. Hamed Nilforoshan et al., *Zero-shot causal learning*, arXiv:2301.12292, 2023.