NOKIA Bell Labs **PhD Study** Inria

Inference of causal models for networks from single observations

Context. Networks such as modern telecommunications networks or distributed embedded systems are permanently monitored to allow identification of failure situations; thousands of new data points reflecting the system state changes are generated every minute. Even if faults are rare events, they can easily propagate driven by local and remote dependencies, which makes it challenging to distinguish causes from effects among the thousands of highly correlated alerts.

A timely automated identification and root cause analysis (RCA) of the origins of performance issues allows executing the most adequate corrective actions and preventing their further propagation. In general, RCA is a hard problem, because it requires a deep knowledge of cause-effect dependencies among many features, physical and logical components the network nodes. In a data driven approach, where most of this knowledge is unavailable a priori, a major difficulty emanates from hidden or unknown variables. Furthermore, even in a fully observable system we are faced with the combinatorial explosion of potential cause-effect dependencies and the difficulty to collect enough information for distinguishing causality from spurious correlations.

Goals. The objective of this project is to develop techniques to infer a causal model that represents the dependencies between components (or nodes) of the network, given a set of event logs and possibly sampled KPI of these components. The vector of event logs can be seen as a single data point, hence in absence of prior knowledge - about, e.g., distributions of events -, well-known statistical inference approaches are not applicable. The PhD candidate will benefit from a large degree of autonomy regarding the evaluation and interpretation of results as well as the tuning of the algorithm.

Approach. We will explore the use of non-reversibility to infer direction of causation. The rationale is that the complexity of the "true" causal process is expected to be in a lower class than the complexity of reconstructing a cause by only knowing its effect. This principle has been studied in the literature in statistical settings [1][2]. However, these results have two shortcomings: they require the probability distributions to be known, and [2] is based on Kolmogorov complexity, which is not computable.

1. To apply complexity-based causal discovery to a single observation or when the distribution is complex or unknown, the first goal of the project is therefore to formalize this principle in a deterministic and decidable setting. Similarly, there is an extensive body of work on process discovery from logs [3], which guesses causal dependencies between events. However, this work does not provide any information about causal dependencies on the level of components.
2. To cope with the complexity of causal discovery, our second goal is to (1) study whether it can be decomposed, in a multi-variable setting, into local analyses like the decomposition into Markov kernels in a statistical setting, and (2) identify tractable complexity classes that match typical behavior of basic network equipment and services.
3. Our third goal of the project is to study applications of the proposed causal inference to the construction of causal explanations for network failures, and/or the detection of change of behavior in terms of altered causal dependencies.

To apply Interested candidates should send a complete CV to Armen Aghasaryan armen.aghasaryan@nokia-bell-labs.com, Emilie Devijver emilie.devijver@univ-grenoble-alpes.fr, Eric Gaussier eric.gaussier@imag.fr and Gregor Goessler gregor.goessler@inria.fr. Candidates should be pursuing internationally recognized research in ML/AI, information theory, or formal methods, with a strong interest in causal inference and causal reasoning.

Starting date / duration 3 years, starting as soon as possible and no later than October 2023.

Location The work will take place either at Nokia, Massy, France, or at Inria Univ. Grenoble Alpes, Montbonnot, France, with frequent travels between the two locations.

References

[1] S. Shimizu et al. A Linear Non-Gaussian Acyclic Model for Causal Discovery. Journal of ML Research 7, 2006.

[2] D. Janzing et al. Causal inference using the algorithmic Markov condition. IEEE Trans. Inf. Theory 56(10), 2010.

[3] W.M.P. van der Aalst, Josep Carmona (Eds.). Process Mining Handbook. LNBIP 448, Springer, 2022.