

Assessing the Existence of a Function in a Dataset with the g_3 Indicator

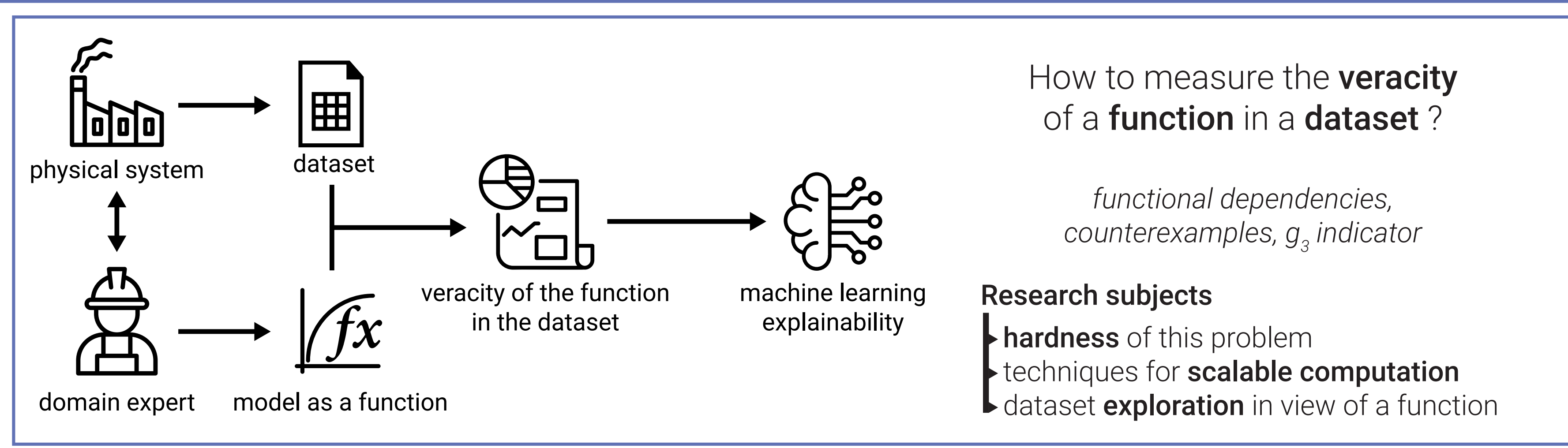
hardness, algorithmics and visualization

Pierre Faure-Giovagnoli^{1,2}, Simon Vilmin¹, Marie Le Guilly¹,
Jean-Marc Petit¹, Vasile-Marian Scuturici¹

¹Univ Lyon, INSA Lyon, CNRS, UCBL, LIRIS UMR 5205, Villeurbanne, France | ²Compagnie Nationale du Rhône, Lyon, France



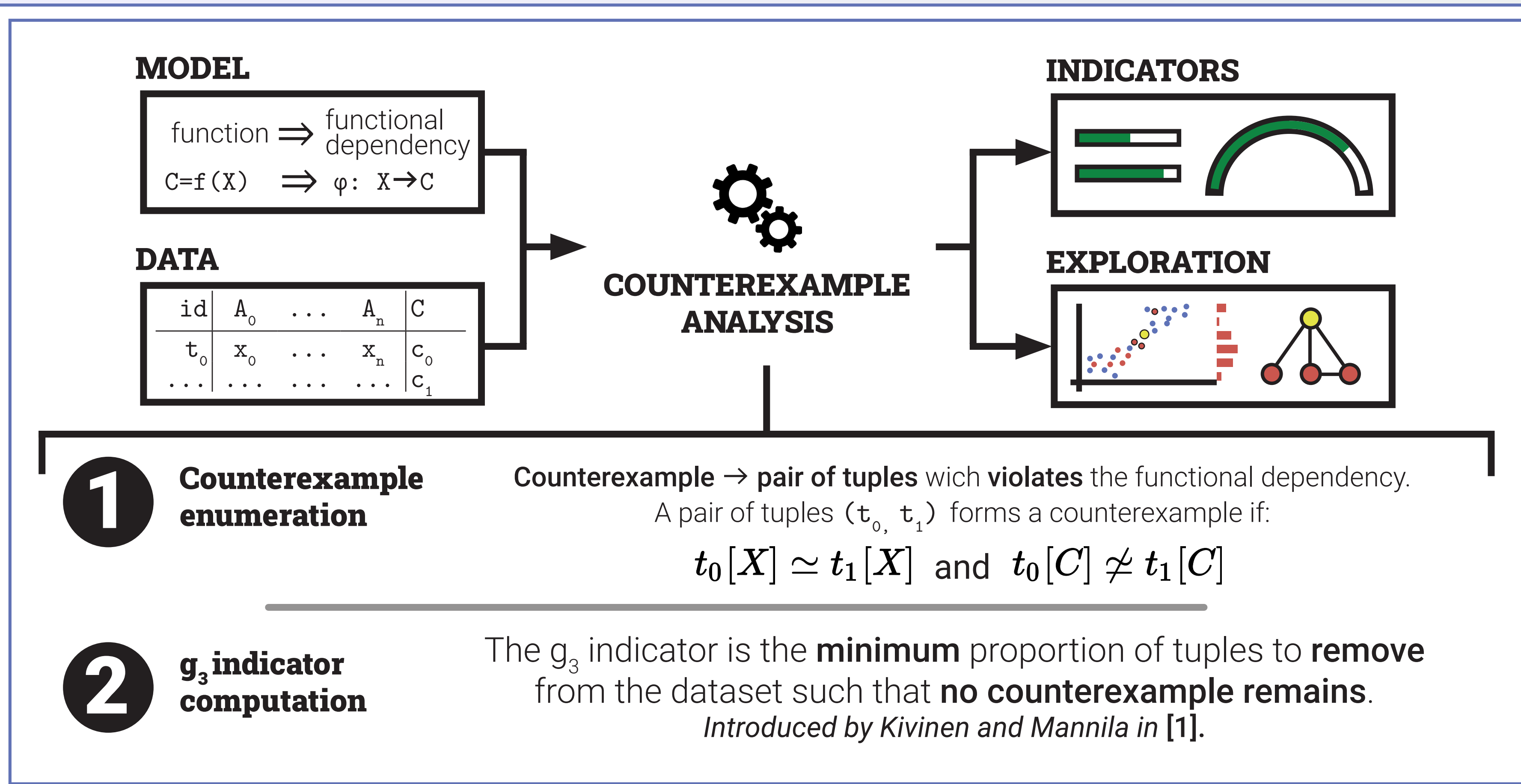
INTRODUCTION



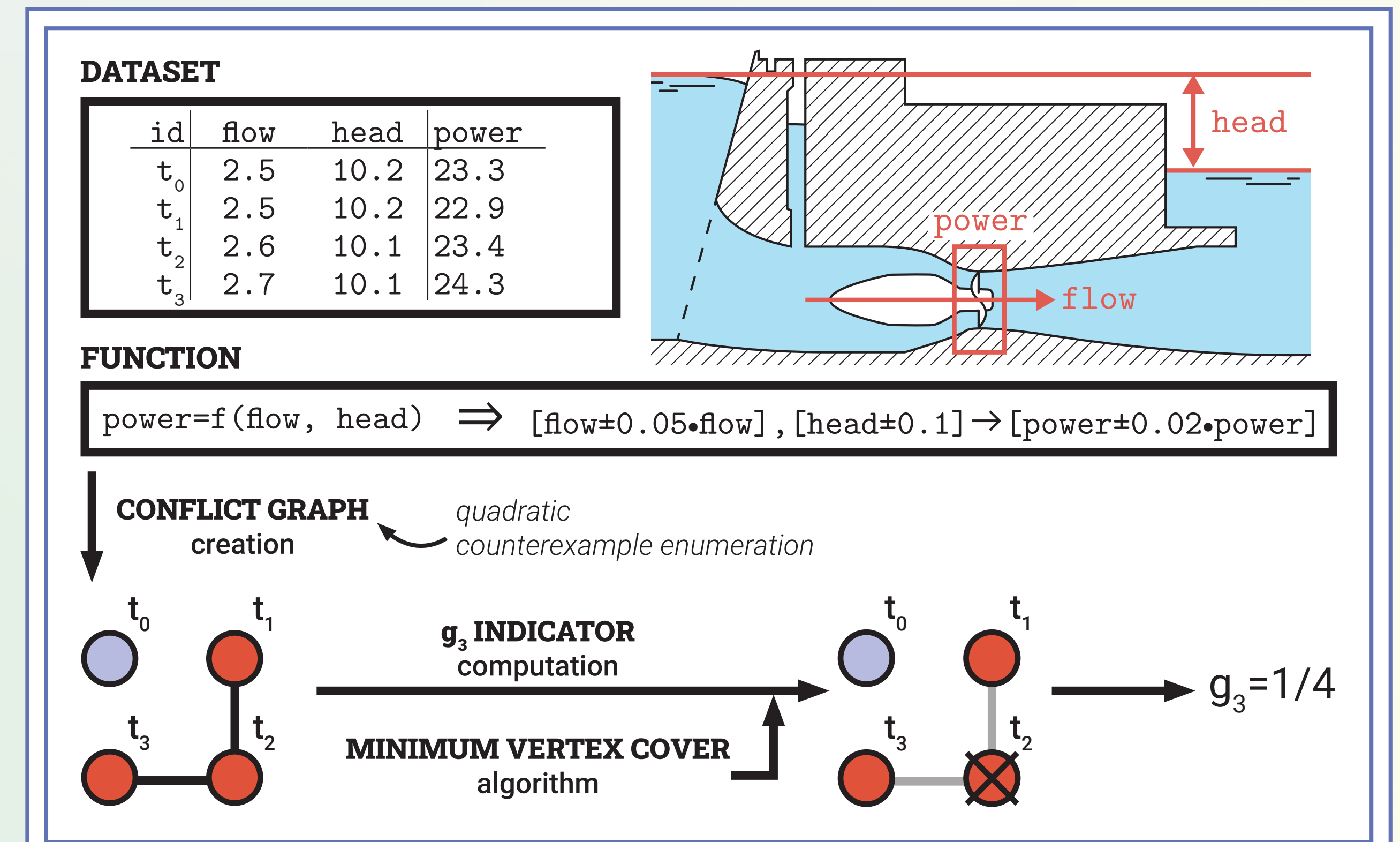
REFERENCES

- [1] Kivinen et al. 1995. *Approximate inference of functional dependencies from relations*. Theoretical Computer Science.
- [2] Delbot et al. 2010. *Analytical and experimental comparison of six algorithms for the vertex cover problem*. Journal of Experimental Algorithmics.
- [3] Onak et al. 2012. *A near-optimal sublinear-time algorithm for approximating the minimum vertex cover size*. ACM-SIAM symposium on Discrete Algorithms.
- [4] Song et al. 2013. *Comparable dependencies over heterogeneous data*. International Journal on Very Large Data Bases.
- [5] Cai et al. 2013. *NuMVC: An efficient local search algorithm for minimum vertex cover*. Journal of Artificial Intelligence Research.
- [6] Hespe et al. 2020. *WeGotYouCovered: The Winning Solver from the PACE 2019 Challenge*. SIAM Workshop on Combinatorial Scientific Computing.
- [7] Le Guilly et al. 2020. *Evaluating Classification Feasibility Using Functional Dependencies*. Transaction on Large-Scale Data- and Knowledge-Centered Systems.
- [8] Faure-Giovagnoli et al. 2021. *ADESIT: Visualize the Limits of your Data in a Machine Learning Process*. International Conference on Very Large Data Bases.
- [9] Faure-Giovagnoli et al. 2022. *Assessing the Existence of a Function in your Dataset with the g_3 Indicator*. IEEE International Conference on Data Engineering.
- [10] Vilmin et al. 2022. *Functional dependencies with predicates: what makes the g_3 -error easy to compute?* BDA conference.

COUNTEREXAMPLE ANALYSIS

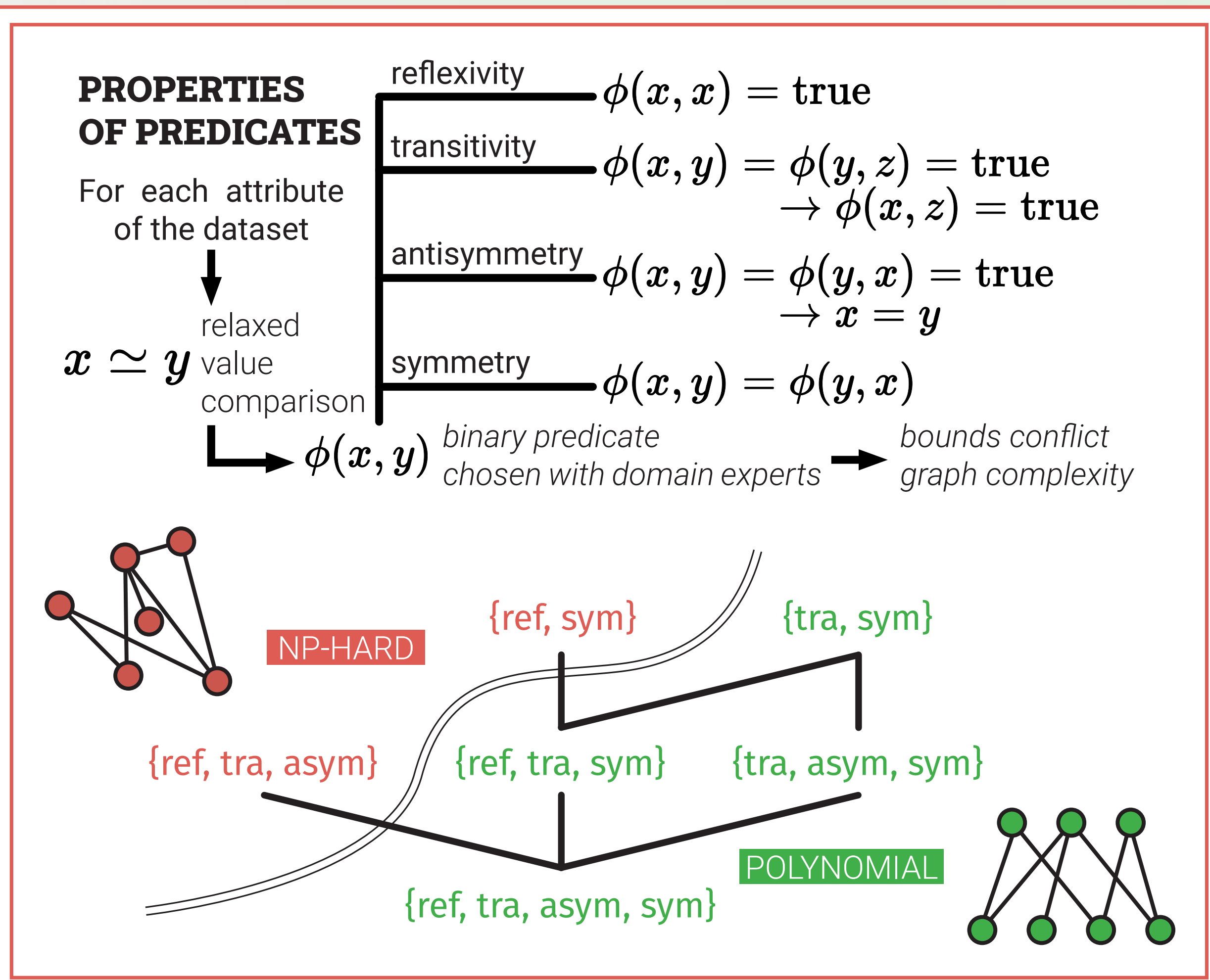


EXAMPLE

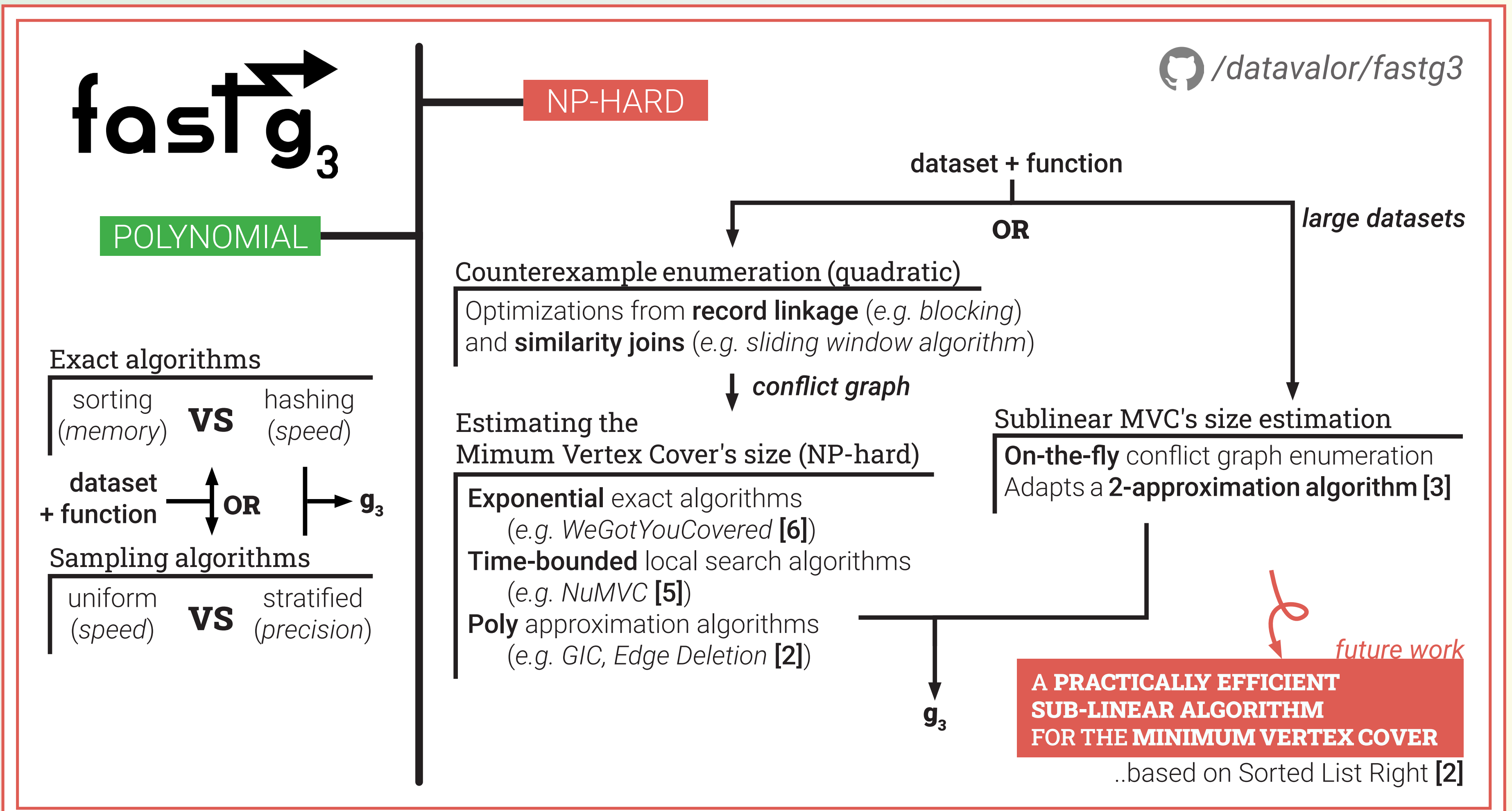


$1-g_3$ is also the maximum accuracy of any model in the case of classical functional dependencies [7].

HARDNESS ANALYSIS [10]



COMPUTE WITH THE FASTG3 PYTHON LIBRARY [9]



VISUALIZE WITH THE ADESIT WEB APP [8]

adesit.liris.cnrs.fr

ADESIT offers a way to visually explore the existence of a function in a dataset for machine learning and much more!