Hamilton Institute Student Seminar Series

# Bruna Wundervald, Maths & Stats / Hamilton Institute

# Title: Regularization in Random Forests

# Abstract: In real-life problems, predictor features can sometimes be hard or even economically expensive to obtain, which motivates the study of regularisation methods for machine learning models. However, for tree-based methods, there is not yet a standard regularization procedure well established in the literature. In this work, we evaluate and extend the currently available methods of Deng, H. and Runger, G. C., aiming to provide a generalized framework for regularization in tree-based methods, especially Random Forests.

Along with that, we also propose a faster implementation as an extension to the ranger package for R (Wright, M. N. and Ziegler, A), since high-dimensional problems need implementations that scale well. Our new methods demonstrate to produce better results in terms of the (number of variables) x (prediction error) trade-off, for both regression and classification problems, as well as significantly improves the computation time of the models.