Optimal-size problem kernels for ***d*-Hitting Set** in linear time and space

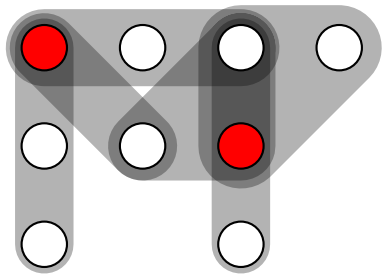
R. v. Bevern and P. V. Smirnov, *Information Processing Letters* 163 (2020).

d-Hitting Set Problem

Input: Sets of of size at most d and a natural number k.

Question: Is there a set of size at most k intersecting all input sets?

Example: Solution for $k = 2$ in red:



Applications

- combinatorial feature selection,
- optimal multi-drug treatment,
- genome association studies,
- graph-based data clustering,
- software engineering,
- public transport optimization,
- radio frequency assignment.

Theoretical results

Data reduction / kernelization:

Theorem. Every d-Hitting Set instance can be reduced to an equivalent instance with $O(k^d)$ sets in linear time and space.

Previously known linear-time kernelizations relied on (partially initialized) quadratic-size data structures.

Running **linear-time kernelization** first can **speed up** Weihe's algorithm and increase its **data reduction** effect.

Mathematical Center in Akademgorodok

РФФИ

Experimental results

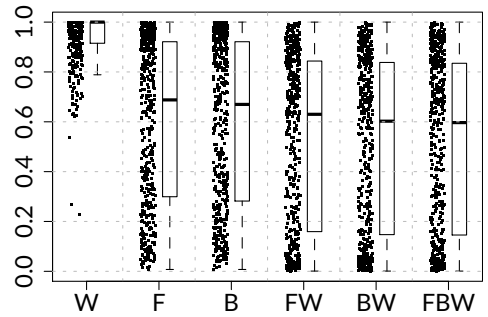
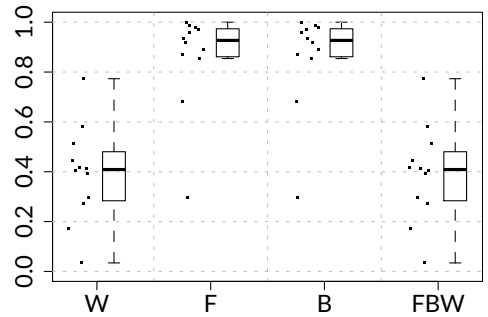
Comparison of three data reduction algorithms for Hitting Set:

- **W**eihe's algorithm (1998),
- van **B**evern's linear-time kernelization (2014),
- **F**afianie and Kratsch's linear-time kernelization (2015),

and combinations: **FW**, **BW**, **FBW**.

Fraction of remaining sets

In instances induced by the task of clustering protein-protein interaction networks (small d):



In instances induced by the station cover problem in European transportation networks (large d):