Members: Fabrizio Cerreto **University:** Technical University of Denmark

Key Characteristics: Systematic use of automatic data sources in transportation • Identify recurrent delay patterns to concentrate corrective actions • Fits in follow-up phases of timetabling to improve operations reliability • Free from subjective interpretation • Separate different delay patterns overlapping on location • Transferable to other transport systems •

Railway delay patterns recognition using data clustering

In a Plan-Do-Check-Act framework, operation analysis is a key step of the process to improve transport services. A better understanding of the development and propagation of delays provides the opportunity to plan appropriate slack in timetables and identify structural conflicts that require mitigation actions.

Several methods have been proposed in the past and are currently used for operation analysis. They can be divided into traditional statistical methods and big-data techniques.

The first tend to aggregate and summarise information, so these can provide a general picture or detailed information on specific stations or trains. The latter can be used to investigate recurring patterns, or internal structure in operation. Several techniques were deployed to determine the recurrences of delays and describe predict delays, but nothing has been presented to identify different ways a train develops its delay over the line.

This project presents a big-data technique to identify recurring delay patterns in railway operations. The absolute delay and delay change are tracked for individual trains along a railway line building absolute delay and delay change profiles. This data format is not new, and was used previously associated to traditional statistics. In general, recurrent delay patterns were identified through visual inspection, which suffers from subjective interpretation. Our project uses k-means to find recurrent ways the train delay develops along the path and investigates the influence of several factors to determine what makes a train belong to a specific cluster of delay. In this way, it is possible to address resources to improve service quality, aiming at the mitigation of specific delay patterns •