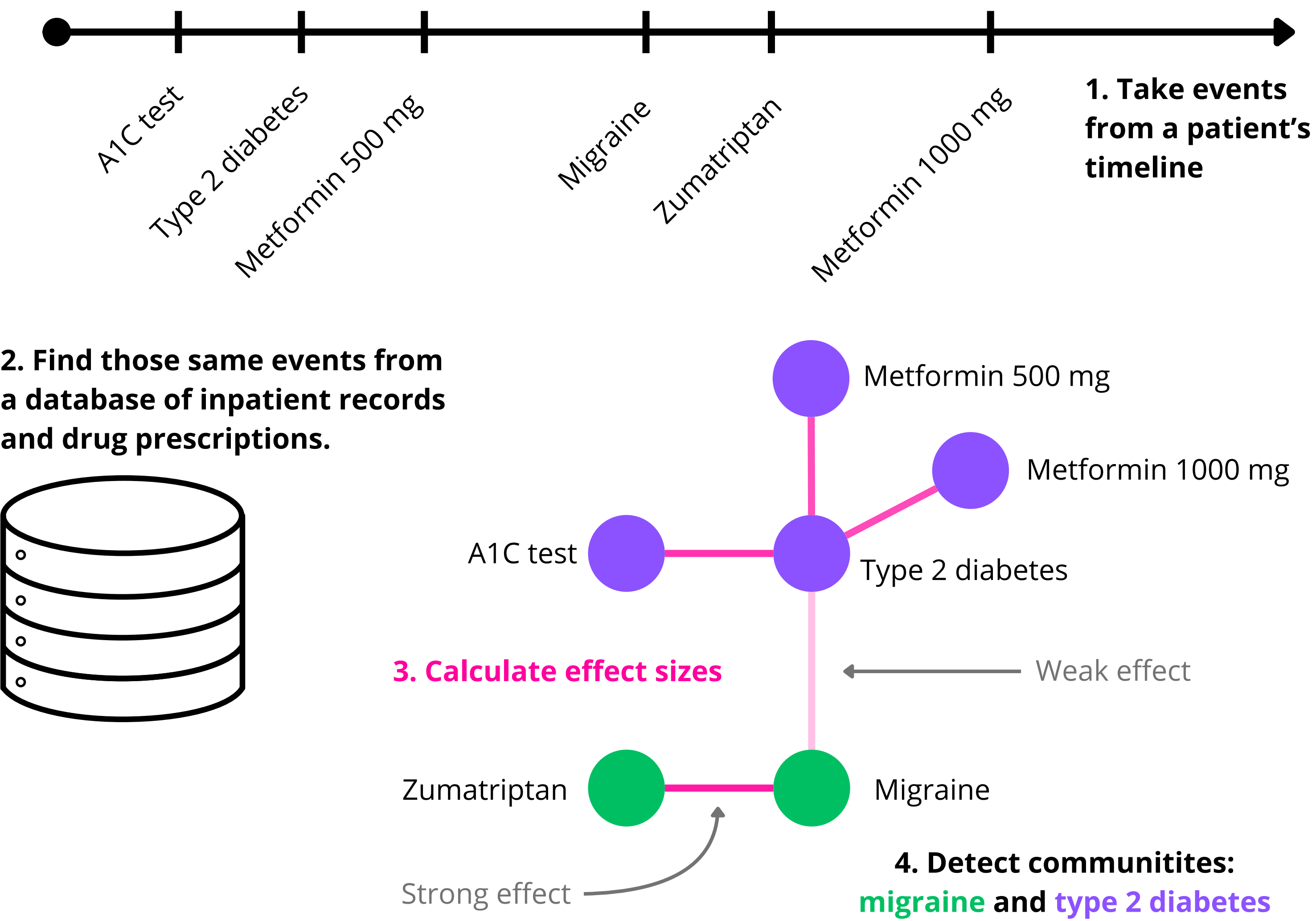


Community detection algorithms reduce noise in medical history



Introduction

Every medical condition creates a large number of data points (e.g **events** like lab measurements, diagnoses and drug prescriptions) in an EHR system. Such level of detail makes it difficult to get a high level overview and similarities between patients are difficult to see on the surface level. This work explores the possibility of reducing noise in medical history by grouping related data points together.

Table 1: average number of clusters per document. Ideal value is 1. The value closest to 1 is highlighted in red.

Community detection	Effect size: OR	Effect size: PMI
Edge betweenness	1.8	1.722
Infomap	1.285	1.205
Louvain	1.17	1.258

Methods

An **effect size** was calculated between every unique pair of events based on their co-occurrence in the same **document** (inpatient record or drug prescription). These pairwise weights were then applied on an individual patient's data to form a **network**. Community detection algorithms were applied on this network to detect strongly linked **clusters** of events of this patient.

Table 2: average number of different ICD10 chapters per cluster. Smaller is better and the smallest value is highlighted in red.

Community detection	Effect size: OR	Effect size: PMI
Edge betweenness	1.383	1.508
Infomap	1.887	2.138
Louvain	2.35	2.165

Three different community detection algorithms (edge betweenness, infomap and Louvain) and two effect size metrics (OR - odds ratio, PMI - pointwise mutual information) were tested as parameters. Two loss metrics were calculated: average number of clusters per document (table 1) and number of different ICD10 chapters (for diagnosis events) per cluster (table 2).

Grouping related events on patient's timeline using graph community detection algorithms

Kermo Saarse¹, Sulev Reisberg^{1,2}, Jaak Vilo^{1,2}

¹University of Tartu, Institute of Computer Science, Estonia

²STACC, Estonia