

Association Rule and Frequent Pattern Mining using the OMOP- CDM

PRESENTER: Solomon Ioannou

INTRODUCTION

To better understand the cooccurrence of data elements and their sequence, association rules analysis and frequent pattern analysis are powerful tools.

An Association Rule analysis answers the question “Given a cohort of patients, what are the most associated concepts that occur together?”

A Frequent Pattern analysis answers the question “What are the most common sequences of concepts observed in a cohort of patients”?

Potentially, they are also promising tools to improve other data mining tasks such as patient-level prediction.

We introduce here an open-source analytics framework, an R package, for performing Association Rule and Frequent Pattern mining using data in the OMOP- CDM.

METHODS

1. The AssociationRuleMining R package makes use of the open source SPMF Java library by Phillippe Fournier-Viger that implements a large collection of association rule and frequent pattern mining algorithms.
2. Using standard HADES packages the user can connect to a database, create the cohort/s of interest and extract relevant covariates.
3. Functionalities within the package allow efficient preparation of the input datasets and analysis using the algorithm of choice.

Workflow Description

1 Create a cohort using one of OHDSI’s tool of choice

Extract covariates using the FeatureExtraction package.

- For Association Rule Mining, extracting the first occurrence of an event (diagnosis, drug subscription, etc) will suffice to perform the analysis.
- For Frequent Pattern Mining, the order of events matters, therefore extracting temporal covariates is essential.

2

Choose an algorithm for the relevant analysis and set its parameters.

- A required parameter to extract highly occurring itemsets or frequent patterns is minimum support, which acts as the threshold for the minimum number of patients that should have the concept set in their medical history, e.g., {obesity, diabetes}
- Algorithms that extract either association rules or frequent patterns require also to specify minimum confidence, which is the threshold for determining how often the left side of the rule occurs together with the right side, e.g., {obesity, diabetes} -> {heart failure}

3

Prepare input datasets and run the analysis.

- The package provides specific functionalities to prepare the input datasets to the necessary format and execute the algorithm.
- Based on the size of the cohort, an iterative procedure to select the optimal value for minimum support and minimum confidence may be applied.

4

Viewing and exploring the results through interactive plots.

5

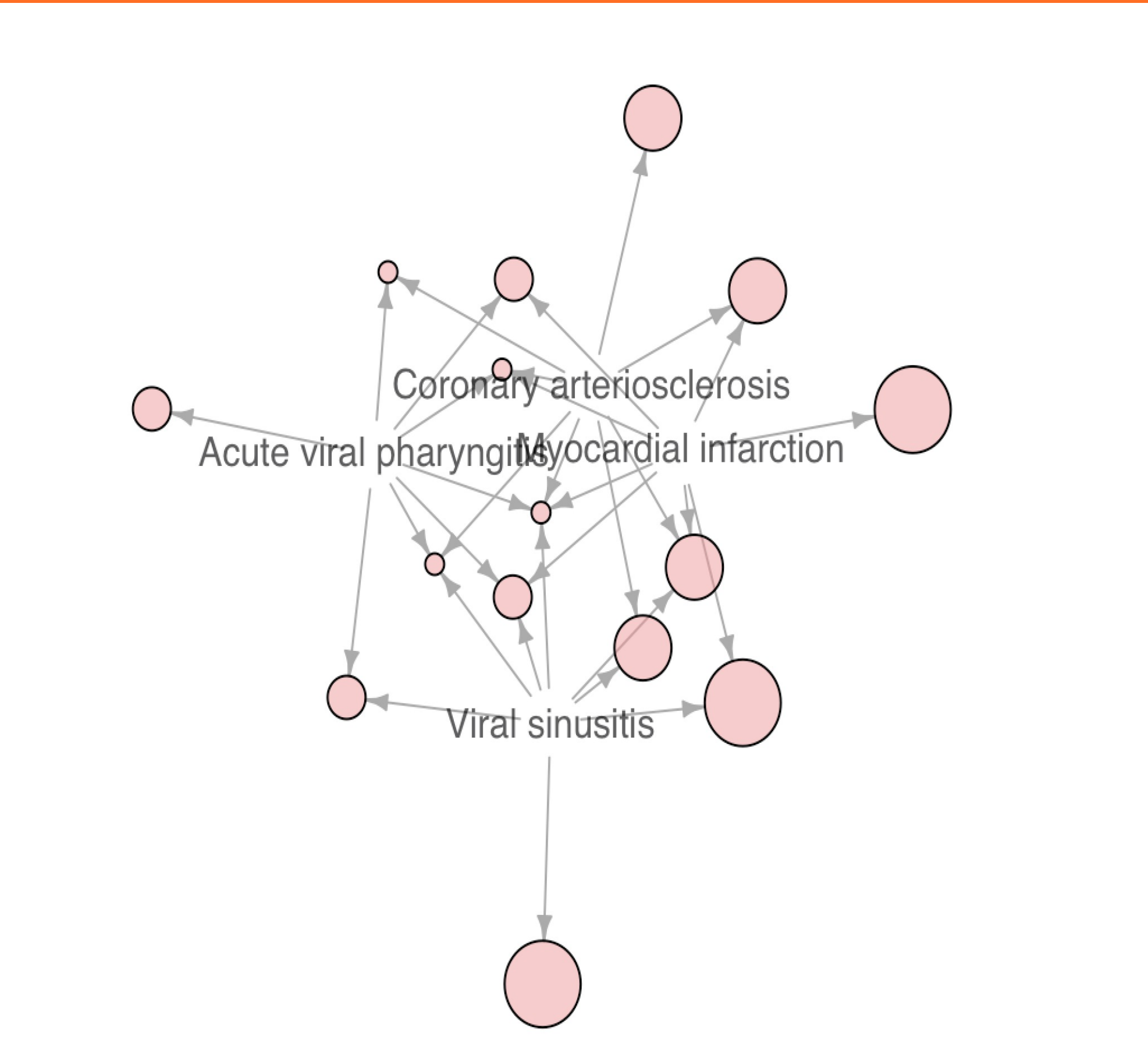


Figure 1: Highly associated concept sets.

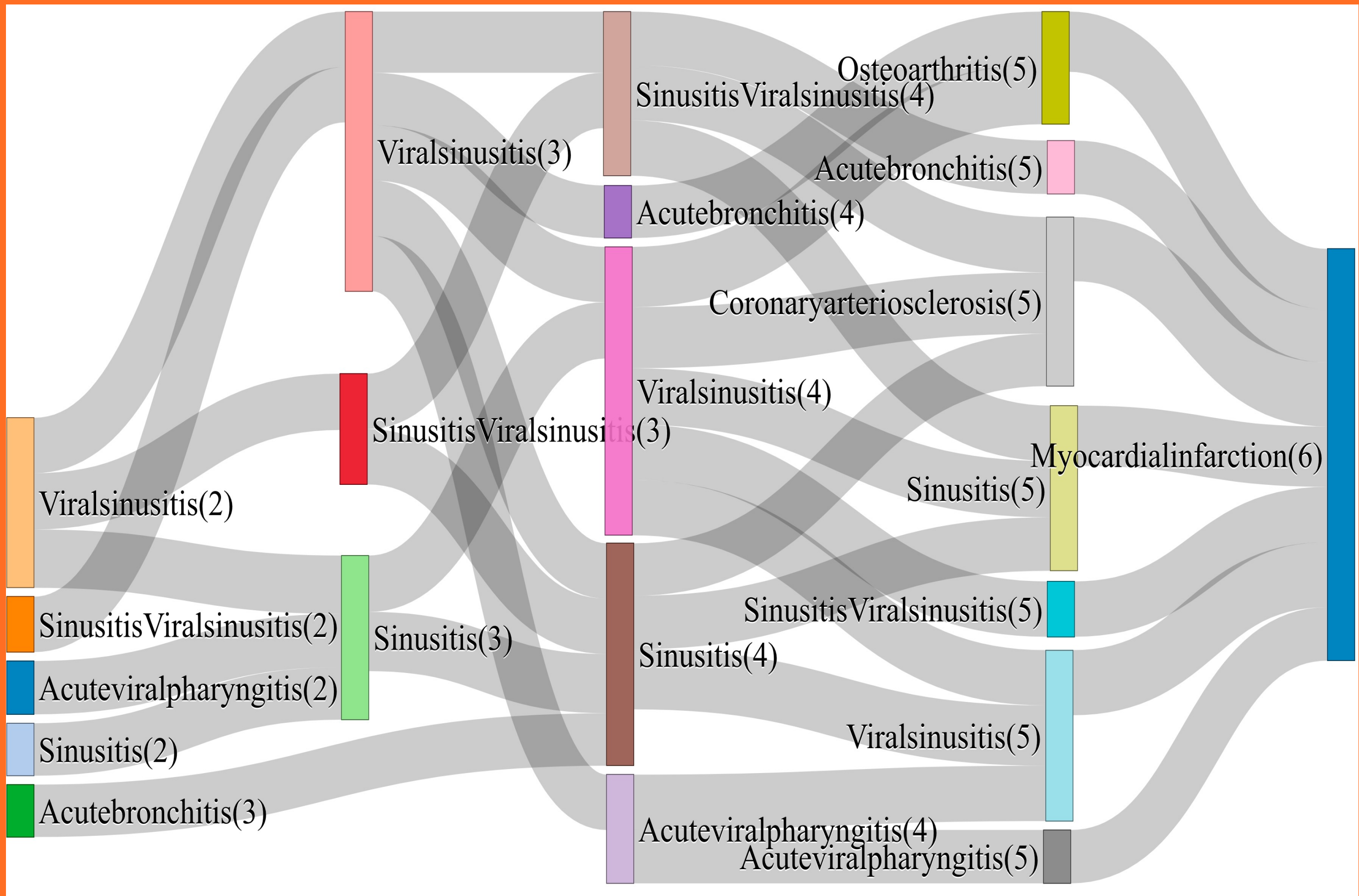


Figure 2: Frequent patterns indicating the chronological ordering of events.

RESULTS

1. Depending on the size of the cohort to be analyzed, the number of concepts included, and the values of predefined parameters of minimum support and minimum confidence, a huge number of rules or patterns can be revealed.
2. Currently results are presented in lists for further processing and use, such as, covariates in prediction problems.
3. Interactive visualizations are also implemented to explore the results graphically.

How can this tool be used?

1. We are exploring the possibilities of using these methods for characterization purposes.
2. Another research direction is the added predictive value, especially of frequent patterns, in clinical prediction problems.

Clinical relevance

1. Characterising frequent patterns and associations in health data can help to identify different types of patients that may need different types of treatment.
2. Frequent pattern analyses could help to generate new hypothesis for the pathogenesis of diseases.

The European Health Data & Evidence Network has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union’s Horizon 2020 research and innovation programme and EFPIA.

Solomon Ioannou,  
Egill Fridgeirsson,  
Jan Kors,  
Peter Rijnbeek

