

Graph analysis and similarity detection: an application on Italian medical prescriptions

Paolo Mariani¹, Ilaria Giordani², Andrea Marletta¹, Mauro Mussini¹, Mariangela Zenga¹

Abstract

This work performs an analysis on antibiotic patterns through a graph database, applying graph algorithms to identify similarities between prescriptions and popularity among prescriptions. Graph databases are data management systems allowing persistent representation of entity and relationship in a graph structure, implementing the Property Graph Model efficiently down to the storage level. A graph is an abstract data type showing connections between pairs of vertices. Nodes identify entities and their properties, while relationships are joining attributes between tables with eventual additional characteristics. Unlike other databases, relationships take first priority. A graph database is purpose-built to handle highly connected data, providing great performance, flexibility and frictionless development. This technique has been implemented on a wide datasets comprehending patients, general practitioners, and their prescriptions in the time span from 2000 to 2018, located in Campania. An excessive usage of antibiotics causes death of microorganisms in the human body which provide to maintaining immune cells and killing certain oral infections. To equilibrate the intestinal flora, lactic ferments are often taken together with antibiotics, so that new “good” bacteria can restore the probiotic action. If this hypothesis is correct, the dataset will show antibiotic prescriptions paired with other drugs, on the same date or it will highlight potential linkings between infections and other pathologies receiving a specific prescription.

Goals of analytics through graphs is completion of antibiotic patterns changes and patient journey, providing a different point of view on those two important aspects altogether. This part of the research aims to focus on: co-prescriptions, understanding whether specified couples of drugs are often prescribed together; clustering in communities, to identify similar kinds of doctors according to their prescription history; centrality measures of nodes, to highlight particularly important entities in the graph.

¹University of Milano-Bicocca

²Consorzio Milano Ricerche