Abstract

Diabetes is a serious disease that leads to the deterioration of many of the body's systems over time. This project aims to analyse diabetes related complications and their correlations, in order to create statistics and rules of thumb which clinicians and patients can use to predict outcomes. By modelling diagnosis codes as nodes and their co-occurrences as relationships, we represented the cumulative diagnosis statistics of over 100,000 patients, analysing them by applying algorithms from network science on the Neo4j graph data platform. Our results centre around 2 main metrics: centrality and clustering. We found that diagnosis of neoplasms, circulatory disease, and respiratory diseases are the most common triggers for hospitalisations among diabetes patients. The former two, along with diabetes, are the diagnoses with highest betweenness centrality, suggesting that they have high connectivity. We found that degree and betweenness centrality are highly intercorrelated, with a few notable distinctions. By applying weighted clustering, we obtained a high clustering coefficient of diagnosis nodes, suggesting that the graph is highly modular and that symptoms of diabetes come in clusters. Ischemic heart disease, in particular, is well-associated with other secondary ailments. Furthermore, we found that physical injuries are associated with high rates of being referred to specialised healthcare.