In this part, a Scala script is used to calculate the similarity between nodes according to the characters they have related to the ego node in the ego network.

Considering each node has 224 characters value compared to the ego node, these 224 values are integrated as a vector owned by each node. In this script cosine similarity is used to describe the similarity between nodes. Euclidean norm is used to determine the distance between vectors. The equation of similarity S is defined as following.

According to the Similarity between nodes, we exclude the links already existed in edges file and selected top 5 most similar nodes of each node and made recommendation.