

Introduction

For this code file, given a known rumor, we are looking for the unknown tweets that have high relevance to the known rumor using random walk method. We are trying to find the stationary probability of all the other tweets if we start from a rumor tweet to do the random walk. Tweets with high stationary probability means that it has higher chance to be landed in random walk, also means it is highly relevant to the rumor tweet. The stationary probability will be presented as a vector with length of the number of tweets. By updating the vector using adjacency matrix till it converges, we can get the stationary probability.

Algorithm

First initialize a random vector, m_u , of length $(n + k)$ (n is the size of the vocabulary and k is the number of tweets) with values between 0 and 1. Then update the probabilities in m_u .

Let M be a weighted bipartite matrix with tweets along rows and vocabulary along columns. Then create an adjacency matrix:

$$M_{adj} = \begin{bmatrix} 0 & M \\ M^\top & 0 \end{bmatrix}$$

The inputs to the algorithm consist of the bipartite matrix M , the target rumor keyword(s) and a tolerance threshold ϵ . The method iteratively updates the probability vector, m_u , until convergence. The first k entries of the stationary probability vector, m_u contain the relevance scores of each tweet to the target rumor keyword.