

# Caim Lab 5: Page Rank

Pol Renau Larrodé

20th November 2019

## 1 Implementació

He fet una petita variació en el codi que venia donat per tal d'evitar tenir les edges replicades en cada aeroport. De tal manera que el que he fet és usar el routeHash amb clau Codi origen, i valor el índex de la llista routes, del propi aeroport, on està aquesta nova Edge.

Per cada Edge llegida, el que fa el codi és augmentar en 1, el pes del out-weight del aeroport origen, i en l'aeroport destí afegir l'aresta a la llista routes i en el routeHash, com he explicat anteriorment.

Per la funció page rank, la implementació bàsicament és la que ve donada en el pdf, però amb la variació de recalculat el pes dels nodes no connexos a cada iteració, i fent que els nodes no connexos estiguin connectats a tots els nodes de forma virtual, però que aquest increment de pés també es veu en els nodes que no són connexos, com està explicat en la següent secció.

Per últim hi ha una variable que hem anomenat threshold que és el límit superior establert per mi per tal de decidir quan aturar l'algorisme. Aquest threshold havia de ser inferior a la suma absoluta de les diferències dels Vectors en cada iteració respecte a l'obtingut en l'anterior. En aquest cas he definit un threshold de  $9e-17$ , per tal de tenir un valor de convergència molt petit. Com hem pogut apreciar, tot i tenir un valor de convergència molt reduït, ha estat bastant ràpid de trobar, ja que al llarg de 192 iteracions hem trobat un valor que convergia.

## 2 Problemes al llarg de la pràctica

El primer problema que vaig veure al moment de realitzar la pràctica, va ser tractar aquells nodes que són independents, aquells que no són connexos a la resta del graf.

En la primera implementació del PageRank, no vaig tenir en compte aquells nodes que no eren connexos, la qual cosa provocava que aquests tinguessin un pes bastant baix, i aleshores la suma total de tots els nodes era molt distant de l'objectiu (1), em donava un valor proper a 0.5, ja que tots els nodes no connexos tenien un valor de  $(1/n)$ .

Un cop vaig detectar aquest error, vaig pensar a incrementar el pes d'aquells nodes que no són connexos a cada iteració, i sumar el pes d'aquells nodes que no són connexos al pes de la resta, de tal manera donem importància als nodes no connexos, però també retribuïm a la resta.

El segon problema va ser provant diferents threshold, vaig provar amb un threshold = 0, però l'algorisme no trigava un temps acceptable per tant vaig decidir posar un valor molt proper a 0.