

# CAIM – FIB

## Pràctica 4

### Page Rank



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH

Pau Bosch Coll

Ramon Ribas Domingo

# **Index**

<b>Index</b>	<b>2</b>
<b>Resum Pràctica:</b>	<b>3</b>
<b>Experimentació</b>	<b>3</b>
Valor d'Epsilon	3
Damping Factor	3
<b>Conclusions</b>	<b>4</b>

# Resum Pràctica:

En aquesta pràctica hem implementat l'algorisme per calcular el pagerank, per tal de fer-ho hem seguit l'esquelet donat, on hem completat les funcions de readRoutes, computePageRanks i outPutPageRanks, tot modificant les parts necessàries del codi per què tot funcionés correctament.

El principal problema que hem trobat en la implementació ha estat que en el computePageRanks inicialment no aconseguim el resultat esperat i la suma dels pageranks no ens donava 1, això ho hem solucionant normalitzant a cada iteració tots els valors dels pageranks de manera que continuessin sent coherents entre ells i la suma d'aquests dones un, tal i com havien explicat a teoria.

## Experimentació

### Valor d'Epsilon

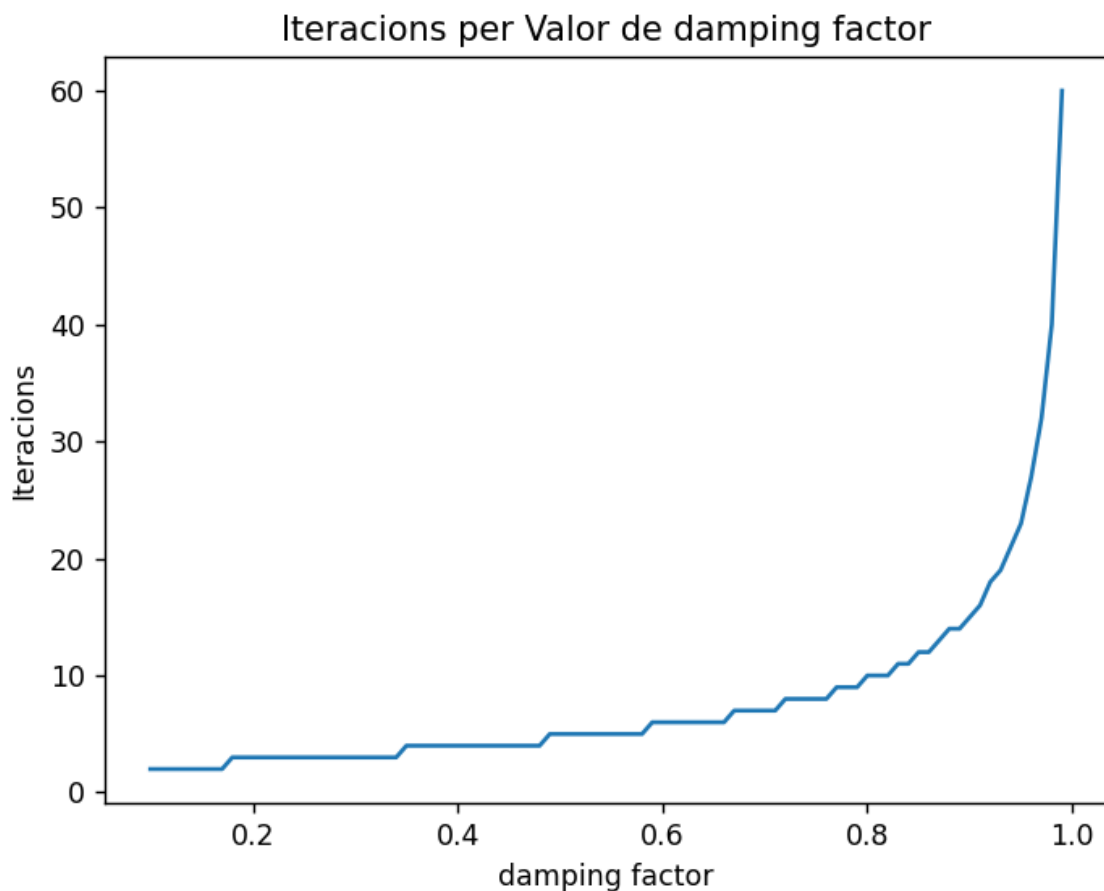
Un valor adequat d'epsilon ha de dependre de l'entrada, ja que els valors de pagerank sumen 1 i per tant, fluctuen en relació a n.

Per tal de calcular els primers dígit rellevants de cada pagerank, hem pensat que com el valor mínim és  $(1-L)/n$ , llavors és més adequat fer-lo dependent a aquest. Finalment, hem decidit que dos dígit de rellevància son suficients així que hem acabat seleccionant:

$$\text{Epsilon} = ((1 - L) / n) / 100$$

### Damping Factor

Per trobar el factor de damping indicat, hem decidit experimentar amb diferents valors entre 0 i 1. Els resultats han estat els següents:



A mesura que decrementa el factor de damping, els pageranks convergeixen més ràpid. Això ho fan a costa de precisió del càlcul del pagerank, ja que aquests es “normalitzant”, o sigui; que es van apropant cada cop més a  $1/n$ .

Per tal d'obtenir valors de pagerank prou diferents mantenint eficiència hem acabat seleccionant un damping factor de 0.85.

## Conclusions

Al llarg de la pràctica hem vist com diferents paramentres com l'epsilon per la condició d'aturada i el damping factor influeixen en el resultats. On per una banda és important tenir un epsilon que s'ajusti als possibles valors de pagerank per tal de no aturar massa prest ni tenir una precisió massa elevada que faci que tardi molt més per digits poc significatius.

Per altra banda tal i com hem esmentat en l'apartat anterior, triar un bon damping factor no molt allunyat de 1 ja que si no es dóna més pes a la part aleatoria de la google matrix que al enllaços reals que hi ha, i alhora poder accedir als nodes que no tenen cap aresta incident.