

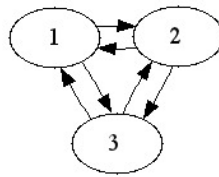


RECUPERACIÓ DE LA INFORMACIÓ

EXERCICIS DEL TEMA 5: Cerca a internet

Abans de començar a resoldre els exercicis, comprova si saps respondre aquestes preguntes:

1. Calcula, a ull, el PageRank de tots els nodes del graf següent:



2. Cert o fals: el PageRank d'una pàgina web depèn de la consulta. ni ha una sensible al context, pero en general no
 3. Cert o fals: els valors de *hub* i *authority* d'una pàgina web si depenen de la consulta.
 4. Cert o fals: l'algorisme del PageRank no té en compte el contingut d'una pàgina web. No el te en compte
 5. Cert o fals: l'algorisme HITS no té en compte el contingut d'una pàgina web. Si que el te en compte (només de les inicials)
-

Exercici 1

Considereu un graf $G=(V,E)$ amb $V = \{A, B, C, D, E\}$ i $E =$

$$\{(A, B), (A, C), (A, D), (B, C), (C, B), (C, D), (D, C), (E, B), (E, C), (E, D)\}.$$

1. Doneu els valors de PageRank d' A i d' E com a funció del factor d'amortiment.
2. Justifiqueu que B i D tenen el mateix PageRank, independentment del factor d'amortiment.
3. Fixeu el factor d'amortiment a 0.9. Doneu la matriu de Google i les equacions associades al PageRank. Després, calculeu el PageRank de cada node.

Exercici 2

Considereu una xarxa reduïda de només tres pàgines, A , B i C , on A té enllaços a B i a C , B té un enllaç a C , i C té un enllaç a B .

1. Doneu les equacions inicials per aquest sistema (sense factor d'amortiment), la matriu de probabilitats de transició associada, i els valors de PageRank resultants per a tots els nodes.



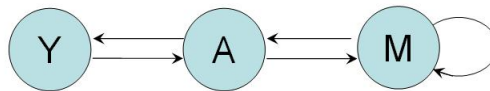
2. Doneu la matriu de Google usant un factor d'amortiment de 0.85, els sistema d'equacions associat pel PageRank, i els valors de PageRank resultants per a tots els nodes.
3. Doneu les equacions pels valors *authority* i *hub* de HITS. Resoleu les equacions usant el mètode iteratiu de la potència o usant un paquet de càlcul numèric.

Exercici 3

Doneu un exemple de graf fortament connectat amb tres nodes tals que 1) cada node té exactament dues arestes que arriben a ell, 2) no tots tres nodes tenen el mateix valor de PageRank. Construïu les equacions pel graf que doneu, solucioneu el sistema i proveu, per substitució directa, que la solució que doneu satisfà les equacions.

Exercici 4

Considereu la web reduïda següent:



Calculeu les equacions de PageRank sense factor d'amortiment i el PageRank de cada node. Repetiu els càlculs amb un factor d'amortiment de 0.85. Després, calculeu els valors de *hub* i *authority*, de HITS, per cada node.

Exercici 5

Volem construir un rastrejador dirigit (*focused crawler*). Expliqueu maneres diferents per esbiaixar el vostre rastrejador cap a determinats temes o continguts. Raoneu la vostra resposta de forma genèrica (sense suposar que esteu interessats en un tema concret) però de forma precisa (que sigui factible de ser implementada).