



RECUPERACIÓ DE LA INFORMACIÓ

Control 2

Data: 8 de gener de 2019

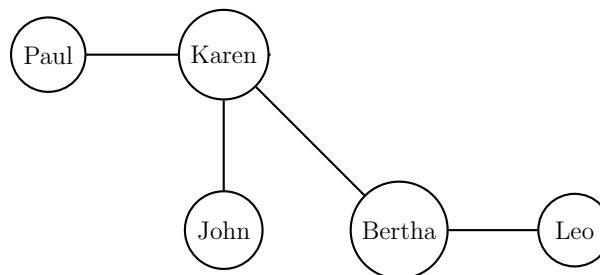
Temps: 2 hores

Problema 1 [3 punts]

- Construeix una xarxa de 4 nodes que tingui un coeficient de clustering global de valor 1. Escribe els càlculs que hagi fet per comprovar-ho.
- Digues si la xarxa que has construït té les propietats d'una xarxa de món petit, o *small world*, i justifica la teva resposta.
- Considera una xarxa de n Republicans i n Demòcrates. La probabilitat de que hi hagi un enllaç entre dos membres del mateix partit és p , mentre que la probabilitat de que hi hagi un enllaç entre membres de partits diferents és q . Quin és el grau mig d'un membre dels Republicans dins la subxarxa dels Republicans? I dins la xarxa sencera?

Problema 2 [2 punts]

Donada la xarxa:



Respon les preguntes següents:

- Quin és el diàmetre de la xarxa?
- Calcula els valors de centralitat d'intermediació (*betweenness*) dels nodes corresponents a la Karen, la Bertha i en Leo, i explica què representen els valors que has obtingut.

$$\text{betweenness_centrality}(i) \stackrel{\text{def}}{=} \sum_{j < k} \frac{g_{jk}(i)}{g_{jk}}$$

$$\left. \begin{array}{l} d_{PK} = 1 \\ d_{PJ} = 2 \\ d_{PB} = 2 \\ d_{PL} = 3 \end{array} \right\} \left. \begin{array}{l} d_{KJ} = 1 \\ d_{KB} = 1 \\ d_{KL} = 2 \end{array} \right\} \left. \begin{array}{l} d_{JB} = 2 \\ d_{JL} = 3 \end{array} \right\}$$

$$d_{BL} = 1 \quad d = 3$$

$$\left. \begin{array}{l} PK: \emptyset \\ PJ: K \\ PB: K \\ PL: KB \end{array} \right\} \left. \begin{array}{l} KJ: \emptyset \\ KB: \emptyset \\ KL: B \end{array} \right\} \left. \begin{array}{l} JB: K \\ JL: KB \end{array} \right\} \left. \begin{array}{l} BL: \emptyset \end{array} \right\}$$

$$\left(\frac{n-1}{2} \right) = \frac{n!}{(n-2)! 2!}$$

$$\left(\frac{4}{2} \right) = \frac{4!}{(4-2)! 2!} = \frac{4 \times 3 \times 2 \times 1}{2! 2!} = 6$$

$$b(P) = 0, b(K) = 5, b(J) = 0, b(B) = 3, b(L) = 0$$

$$b_m(K) = 5/6, b_m(B) = 3/6$$



Problema 3 [3 punts]

Ens donen una llarga llista de documents. Alguns d'aquests documents contenen enllaços a d'altres documents. Es vol obtenir el document que rep més enllaços d'entrada. El fitxer d'entrada té el format següent (sense la primera línia):

```
"font", "href"  
d1.html, d2.html  
d2.html, d4.html, d3.html  
d3.html, d4.html  
d4.html,  
d5.html, d1.html, d2.html, d3.html  
...
```

que es llegeix com “el document `d1.html` conté un enllaç al document `d2.html`, el document `d2.html` conté enllaços als documents `d4.html` i `d3.html`...”.

Descriu com resoldries aquest problema usant el model MapReduce. Pots usar més d'una tasca.

Problema 4 [2 punts]

Una xarxa de pàgines web presenta l'estructura d'enllaços següent:

- La pàgina *A* té un enllaç a la pàgina *C*.
 - La pàgina *B* té un enllaç a la pàgina *A*.
 - La pàgina *C* té enllaços a les pàgines *A* i *D*.
 - La pàgina *D* té un enllaç a la pàgina *B*.
- (a) Dibuixa la xarxa, dóna la matriu de probabilitats de transició associada i els valors de PageRank per a totes les pàgines (nodes).
- (b) Escribe i explica la fórmula de la matriu de Google G pel càlcul del PageRank amb un factor d'amortiment λ . Podries dir com afecta el valor de λ al càlcul del PageRank i a la seva convergència?
- (c) Escribe la matriu de Google usant un factor d'amortiment de 0.85 i el sistema d'equacions resultant per calcular els valors de PageRank de totes les pàgines de la xarxa. No cal que resolguis el sistema.