



## RECUPERACIÓ DE LA INFORMACIÓ

## Control 2

Data: 8 de gener de 2019

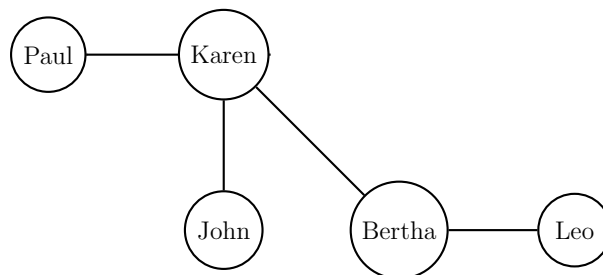
Temps: 2 hores

### Problema 1 [3 punts]

- (a) 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.
- (b) Digues si la xarxa que has construït té les propietats d'una xarxa de món petit, o *small world*, i justifica la teva resposta.
- (c) 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:

- (a) Quin és el diàmetre de la xarxa?
- (b) 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.



**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`...”.

Describeu 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) Escribeu i explica la fórmula de la matriu de Google  $G$  pel càlcul del PageRank amb un factor d'amortiment  $\lambda$ . Podries dir com afecta el valor de  $\lambda$  al càlcul del PageRank i a la seva convergència?
- (c) Escribeu 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.