



RECUPERACIÓ DE LA INFORMACIÓ

Data: 8 de gener de 2021

Control 2

Temps: 2 hores

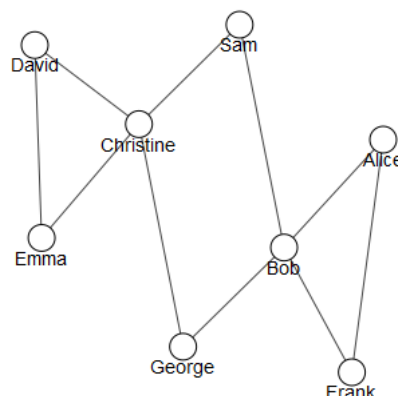
Problema 1 [3 punts]

Respon les preguntes següents i justifica les teves respostes.

- Representa gràficament una xarxa en què un mateix node presenta els valors màxims per les tres mesures de centralitat: de grau, de proximitat i d'intermediació.
- Et donen un xarxa amb 6 nodes i 8 arestes. Si la xarxa fos un graf Erdős-Rényi, quin seria el coeficient de clustering esperat?
- Quina de les mesures de xarxes vistes a classe creus que es veu més afectada pel missatge de Twitter “Qui seguir” quan et recomana comptes que segueixen els teus amics (aquells a qui tu segueixes).
- Quina és la idea central rere la noció de “sis graus de separació”?
 - Les xarxes socials tenen coeficients de clustering elevats.
 - Les xarxes socials són esparses.
 - Les xarxes socials presenten *hubs*.
 - Les xarxes socials tenen un valor petit de mitjana de la longitud dels camins més curts.

Problema 2 [3,5 punts]

Donada la xarxa:



- Calcula, per cadascun dels nodes de la xarxa, els valors de centralitat de grau, centralitat de proximitat i centralitat d'intermediació. Escriu els valors (3 columnes) en una taula ordenada alfabèticament pel nom del node (8 files).



- (b) La Paradoxa de l'amistat (*Friendship Paradox*), descrita per Scott L. Feld el 1991, diu que en una xarxa social, els teus amics tenen més amics que tu, de mitjana. Aquesta paradoxa, es compleix en aquesta xarxa? És a dir, el grau mitjà dels veïns més propers és major que el grau mitjà dels nodes? Justifica la teva resposta.

Problema 3 [3,5 punts]

Hem rastrejat la pàgina web d'un aeroport que conté les dades dels vols que surten d'aquest aeroport. A partir d'aquestes dades, hem generat un llarg fitxer de text amb el format següent:

```
2021-01-01,06:00:00,08:10:00,Barcelona,Lisboa
```

Cada línia conté la data, l'hora prevista d'enlairament, l'hora real d'enlairament, la ciutat de l'aeroport d'origen i la ciutat de l'aeroport destinació. Els 5 camps estan separats pel caràcter ','.

Volem obtenir el percentatge de vols que han sortit amb retard. Es considera que un vol que s'ha enlairat més de 2 hores després de la seva hora prevista, ha sortit amb retard.

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

Es valorarà l'eficiència de la solució proposada.