

CAIM, segundo parcial

11 de gener de 2019. Temps: 2 hores 30 minuts

Exercici 1 (2.5 puntos). Considera el siguiente grafo no dirigido con la siguiente matriz de adyacencia:

$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Calcula el coeficiente de clustering local y la “betweenness” de cada nodo.

Exercici 2 (2.5 puntos) Somos los propietarios de una red social similar a “Facebook” y queremos calcular, para cada par de usuarios, el número de amigos comunes que tienen. La información sobre la red viene dada en ficheros que contienen las listas de amigos de cada usuario, por ejemplo:

```
marta: balqui, ricard, ramon, larri, javier
ricard: ramon, marta, toni
...
```

Lo cual quiere decir que Marta está conectada con Balqui, Ricard, Ramon, Larri y Javier, y que Ricard está conectado con Ramon, Marta y Toni. Las relaciones en esta red son simétricas, por lo cual si una persona x aparece en la lista de amigos de y , entonces y aparece también en la lista de x .

Explica como calcular el número de amigos comunes para aquellas parejas de usuarios que tienen amigos comunes en el formalismo de mapreduce. Indica cuántas fases de mapreduce necesitas y da pseudocódigo las funciones de map y reduce de cada fase.

Exercici 3 (2.5 puntos). Describe brevemente un par de posibles aplicaciones de locality sensitive hashing en el contexto de information retrieval.

Exercici 4 (2.5 puntos). Indica un escenario donde Spark sería preferible a Hadoop para procesar tus datos. Justifica brevemente tu respuesta.

Nota: “Machine learning” o “real-time processing” *no* son escenarios. Un escenario es algo así como “somos fabricantes de paraguas y necesitamos incrementar nuestra producción y distribución de forma rápida si el pronóstico del tiempo es de lluvia”.

Nota 2: Este ejercicio será utilizado para calcular la nota de la “Competència Transversal Aprenentatge Autònom”.