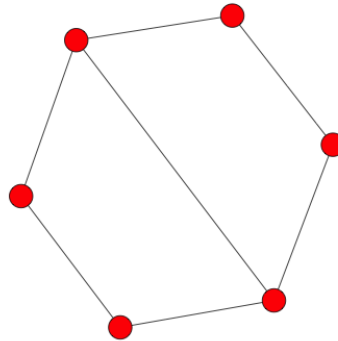


CAIM, segon examen parcial

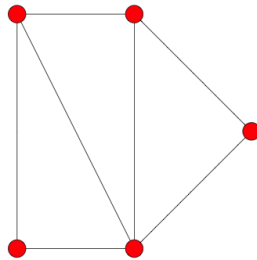
10 de gener de 2014. Temps: 1 hora 50 minuts

**Exercici 1** (2.2 punts) Considera el següent graf no dirigit amb 6 nodes i 7 arestes.



Calcula la centralitat *closeness*, *betweenness*, i de *grau* de cada un dels nodes.

**Exercici 2** (2.2 punts) Considera el següent graf no dirigit amb 5 nodes i 7 arestes.



1. Calcula la similaritat del cosinus per a cada un dels 10 parells de nodes.
2. Basant-te en les similaritats calculades a l'apartat anterior, construeix el dendograma resultant d'aplicar clustering jeràrquic utilitzant *single-linkage*.

**Exercici 3** (2.2 punts) Volem fer un cercador especialitzat en un tema (per concretar, posem literatura mèdica). Digueu totes les maneres que se t’acuden de fer servir machine learning i data mining (en les diferents parts del cercador: crawler, preprocessador, indexador, calculador de “ranks”, el recuperador de documents, el formatejador de la resposta...). Posa-hi imaginació. (Longitud orientativa: com a molt 1 pàgina).

**Exercici 4** (2.2 punts) Diem que dues paraules consecutives en un document formen una 2-frase; així doncs, en un document amb  $L$  paraules poden haver-hi com a màxim  $L - 1$  2-frases diferents.

1. Fes un programa en map-reduce que, donat un conjunt de documents i un paràmetre  $k$  retorni el conjunt de 2-frases que apareixen com a mínim  $k$  vegades en tot el conjunt.
2. Fes un programa map-reduce que, donada una paraula  $w$ , doni una llista de tots els parells  $(d, v)$  per a cada  $d$  i  $v$  tals que “ $v w$ ” és una 2-frase en el document de docid  $d$ .

En els dos casos, dóna pseudocodi de les funcions map i reduce i, si fan falta, de les funcions combine i partition.

**Exercici 5** (2.2 punts) (Aquesta pregunta serveix per avaluar la competència transversal “aprenentatge autònom”.) Descriu dos exemples (amb 5-6 línies per cadascun) d’escenaris concrets on seria preferible usar Storm en comptes de Hadoop, i expliqueu per què.