Big data – Graph Mining
A.A. 2019/2020

Appello del 6 febbraio 2020

Esercizio 1 [9 punti]

- **[7 punti**] Si calcoli, applicando il metodo visto a lezione ed evidenziando tutti i passaggi fatti riga per riga, la matrice delle firme minhash dei seguenti insiemi S1, S2, S3, S4 e S5 secondo le permutazioni indotte dalle 3 funzioni hash evidenziate in tabella.
- **[2 punti]** Qual è la proprietà fondamentale che mette in relazione le funzioni minhash con la Jaccard Similarity degli insiemi? Discutere se nell'esercizio svolto tale proprietà risulta empiricamente rispettata.

Riga	S_1	S_2	S_3	S_4	S_5	(x+3) mod 11	(2x+1) mod 11	3x mod 11
0	**0**	**1**	**0**	**1**	**1**	*3*		
1	**1**	**0**	**0**	**0**	**1**	*4*		
2	**1**	**1**	**0**	**0**	**1**	*5*		
3	**1**	**1**	**0**	**0**	**1**	*6*		
4	**0**	**0**	**1**	**1**	**1**	*7*		
5	**0**	**0**	**1**	**1**	**0**	*8*		
6	**0**	**0**	**1**	**1**	**0**	*9*		
7	**0**	**1**	**1**	**1**	**0**	*10*		
8	**1**	**0**	**0**	**0**	**0**	*0*		
9	**1**	**1**	**0**	**0**	**1**	*1*		
10	**1**	**0**	**0**	**0**	**1**	*2*		

Esercizio 2 [9 punti]

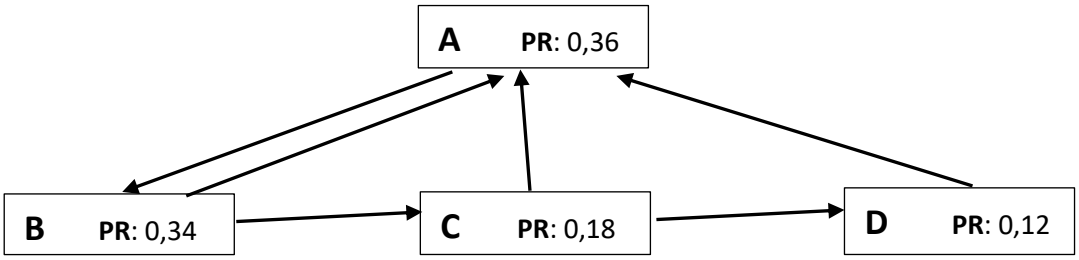
Si considerino i seguenti basket:

basket	item
1	pasta, formaggio, pomodoro, legumi
2	nutella, pasta, pomodoro, legumi
3	riso, latte, succo di frutta, zucchero, sale
4	sale, pasta, pomodoro, nutella
5	legumi, pomodoro, sale, latte
6	formaggio, sale, pasta, pomodoro

- **[6 punti]** Far vedere l'esecuzione passo passo dell'algoritmo A-priori per individuare tutti gli insiemi di item che abbiano supporto almeno uguale a 2.
- **[3 punti]** Individuare una regola di associazione con *interesse* positivo, calcolando la sua *confidenza* e il suo *interesse*.

Esercizio 3 [8 punti]

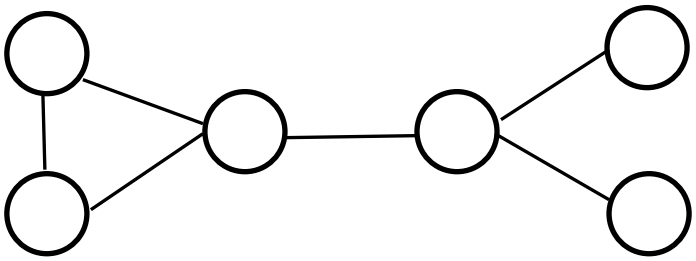
Si consideri il grafo del web in figura, composto da quattro pagine, con indicati dei valori di pagerank.



- **[5 punti]** Si dica, giustificando la risposta, se i valori di pagerank indicati corrispondono ai corretti valori per un certo parametro **β** di teleport. In caso affermativo, individuare tale parametro **β.**
- **[3 punti]** Si motivi perché, nel calcolo del pagerank, si introduce il parametro di teleport, indicando in particolare quali problemi esso risolve e come li risolve.

Esercizio 4 [8 punti]

Si consideri la stessa rete sociale in figura.



- **[6 punti]** Si calcoli la **edge betweeness** (con l'algoritmo di Girvan–Newman) per ogni arco della rete.
- **[2 punti]** Si discuta se tale calcolo può portare a dividere il grafo in due cluster, giustificando la risposta.

Regole per lo svolgimento della prova scritta:

- Per svolgere il compito si hanno a disposizione **110** minuti
- Scrivere **subito** nome, cognome, matricola su OGNI FOGLIO.
- Durante la prova scritta **non** è possibile abbandonare l'aula.
- Non è ammesso **per nessun motivo** comunicare in qualsiasi modo con altre persone
- Non è possibile consultare appunti, libri e dispense.
- Qualsiasi strumento elettronico di calcolo o comunicazione (telefoni cellulari, calcolatrici, palmari, computer, etc…) deve essere **completamente disattivato** e **depositato in vista sulla cattedra**
- Mettere in vista sul banco un valido documento di identità.