

Trojanje: 120 minuta. Rješenja zadataka na košuljici i po potrebi na zasebnim papirima. Zaokružiti na košuljici redni broj rješavanog zadatka (1-7).

1. (4 boda) Točke (1, 5), (2, 4), (2, 8) i (7, 7) algoritam CURE (engl. *clustering using representatives*) dodijelio je u jednu grupu. Odredite udaljenost točke (6, 4) od navedene grupe uz pretpostavku da grupa ima dvije reprezentativne točke koje se odabiru po principu najveće udaljenosti, a faktor približavanja reprezentativnih točaka iznosi 10%.

2. (4 boda) Graf društvene mreže zadan na slici 1 podijeljen je na dvije zajednice X i Y . Nacrtajte AGM (engl. *affiliation graph model*) za zadanu mrežu. Izračunajte vjerojatnost da su entiteti u zajednici Y povezani, ako je poznato da vjerojatnost povezanosti entiteta u zajednici X iznosi $p_x = 0.8$. $p(AGM) = 0.01$ $-3,92$

3. (4 boda) Odredite zadovoljava li graf zadan na slici 2 svojstva društvene mreže. Obrazložite odgovor. $\frac{5}{13} < \frac{3}{5}$
4. (4 boda) U sustavu oglašavanja na webu primjenjuje se *generalized balance* algoritam. Tri oglašivača se natječu za oglasni prostor:

- Oglašivač O_1 želi objaviti oglas koji se do sada prikazao 50 puta, od čega su korisnici odabrali oglas 40 puta, za oglas želi platiti 11 kuna, a do sada potrošio je 40 kuna od ukupnog budžeta koji iznosi 100 kuna.
- Oglašivač O_2 želi objaviti oglas koji se do sada prikazao 120 puta, od čega su korisnici odabrali oglas 72 puta, za oglas želi platiti 12 kuna, a do sada potrošio je 240 kuna od ukupnog budžeta koji iznosi 300 kuna.
- Oglašivač O_3 želi objaviti oglas koji se do sada prikazao 100 puta, od čega su korisnici odabrali oglas 55 puta, za oglas želi platiti 12 kuna, a do sada potrošio je 570 kuna od ukupnog budžeta koji iznosi 600 kuna.

U sustav dolaze tri uzastupna korisnička upita koja odgovaraju oglasnom profilu za koji se natječu stva tri oglašivača. Odredite redoslijed prikazivanja oglasa. 010201

5. (5 boda) U tablici ispod zadana je matrica ocjena korisnika za pojedine filmove (eng. *user-item matrix*). Prazna polja u matrici predstavljaju ocjene koje nedostaju. Korištenjem algoritama suradničkog filtriranja (eng. *Collaborative Filtering*) potrebno je izračunati preporuku/ocjenu za korisnika U_1 i film M_1 .

a.) Ako se koristi *Item-Item* pristup suradničkog filtriranja? $3,705$

b.) Ako se koristi *User-User* pristup suradničkog filtriranja? 2

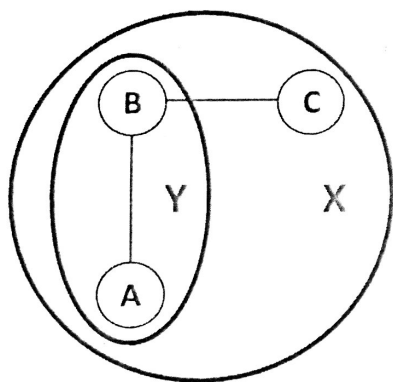
	U_1	U_2	U_3	U_4	U_5	U_6
M_1	?	1		2	2	
M_2	4	1	2	5	3	3
M_3	1		5	5	2	4
M_4	3		3		5	
M_5	4	3		2	4	
M_6	1	4	4	1	1	5

U oba slučaja kardinalni broj skupa sličnih filmova/korisnika koje sustav preporuke razmatra iznosi najviše 3. Nadalje, sustav ne razmatra korisnike i filmove koji nisu slični ($\text{sim}(A, B) \leq 0.0$). Tako npr. može biti i samo jedan (ili čak nijedan) sličan entitet, jer svi entiteti imaju negativnu sličnost s obzirom na entitet za kojega se procjenjuje preporuka. U slučaju da nema sličnih entiteta, napisati da sustav ne može izračunati preporuku. Za računanje sličnosti koristi se *PCC* (eng. *Pearson Correlation Coefficient*). Dakle, potrebno je od pojedinih ocjena oduzeti prosjek filma (za *Item-Item* pristup), odnosno prosjek korisnika (za *User-User* pristup) te nad normaliziranim ocjenama izračunati *cosine* mjeru sličnosti.

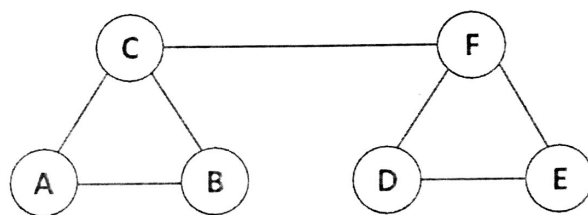
c.) Navesti prednosti i nedostatke suradničkog filtriranja u odnosu na ostale načine preporučivanja?

6. (4 boda) Ulaz u algoritam *DGIM* je binarni niz 11101110101011. a) Skicirati sadržaj pretinaca za svaki dolazak novog bita u ulaznom nizu za veličine prozora $N = 5$ i $N = 10$. b) Procijeniti koliko ima jedinica u zadnjih N bitova ($k=N$) za obe zadane vrijednosti parametra N . $3, 0$ $(2, 1, 1)$ $(4, 2, 2, 1, 1)$

7. (5 boda) a) Kratko opišite postupak te navedite izraz za procjenu broja različitih elemenata u toku podataka prema Flajolet-Martinu. b) Dokažite da je izraz iz a) dijela zadatka dobra aproksimacija stvarnog broja različitih elemenata u toku podataka. Napomena: Kratko objasniti korake dokaza!



Slika 1: Graf uz zadatak 2



Slika 2: Graf uz zadatak X 3