

UNIVERZITET U BEOGRADU  
ELEKTROTEHNIČKI FAKULTET



# ANALIZA SOCIJALNIH MREŽA

Projektni zadatak

Verzija 1.0

**Predmetni nastavnici:**

dr Marko Mišić, docent

dr Jelica Protić, redovni profesor

**Školska godina:**

2018/2019.

Beograd, decembar 2018.

# SADRŽAJ

<b>SADRŽAJ .....</b>	<b>2</b>
<b>1. UVOD.....</b>	<b>3</b>
<b>2. CILJ .....</b>	<b>3</b>
<b>3. POSTAVLJENI PROBLEM.....</b>	<b>3</b>
3.1. ANALIZA PODATAKA O NAJPOPULARNIJIM FILMOVIMA IZ IMDB BAZE.....	3
3.2. SKUP PODATAKA ZA ANALIZU .....	4
3.3. MODELOVANJE MREŽE .....	5
3.4. ISTRAŽIVAČKA PITANJA I CILJEVI .....	5
3.5. PREPORUČENE METODE I ALATI.....	7
<b>4. REZULTATI.....</b>	<b>7</b>
<b>5. PREDAJA, ODBRANA I VREDNOVANJE .....</b>	<b>7</b>

# 1. UVOD

U okviru ovog dokumenta su data uputstva za izradu projektnog zadatka na predmetu Analiza socijalnih mreža (13M111ASM) u školskoj 2018/2019. godini. Studenti treba da pažljivo pročitaju ovo uputstvo pre izrade projektnog zadatka. Studenti projektni zadatak rade **samostalno ili u paru**.

## 2. CILJ

Cilj projektnog zadatka na predmetu Analiza socijalnih mreža je praktična primena stečenog teorijskog znanja iz predmeta na primeru jednog konkretnog istraživačkog problema. Kroz zadati istraživački problem, studenti treba da izvrše prikupljanje, obradu i preliminarnu analizu primarnog (sirovog) skupa podataka, izdvoje neophodne podatke i modeliraju problem mrežom odgovarajućeg tipa. Modeliranu mrežu treba da analiziraju alatima za obradu socijalnih mreža po izboru i izvrše vizuelizaciju mreže. Dobijene rezultate analize treba na odgovarajući način interpretirati u skladu sa postavljenim istraživačkim pitanjima.

## 3. POSTAVLJENI PROBLEM

U okviru ove sekcije je dat predlog projektnog zadatka za tekuću školsku godinu. Studenti mogu predložiti predmetnom nastavniku drugu temu. U tom slučaju, poželjno je priložiti i deo skupa podataka koji bi se analizirao, kako bi student na adekvatan način u saradnji sa nastavnikom postavio ciljeve istraživanja i istraživačka pitanja.

### 3.1. Analiza podataka o najpopularnijim filmovima iz IMDb baze

Tema projektnog zadatka u tekućoj školskoj godini je analiza podataka o 1000 najpopularnijih filmova iz *Internet Movie Database* (IMDb) u periodu od 2006. do 2016. godine. IMDb je *online* internet baza podataka o filmovima, televizijskim serijama i drugim vidovima kućne zabave koja obuhvata podatke o glumačkim postavama, produkcionom osoblju, opisima filmova, kao i podatke o zaradi, ocenama kritike i gledalaca i sl.

Na popularnost i uspešnost jednog filma često utiču žanr, scenario, budžet, a svakako odlučujući doprinos daju i režiser i glumci. Određeni filmovi su neke glumce učinili veoma popularnim, dok sa druge strane, neki filmovi imaju visoke ocene upravo zbog popularne glumačke postave i dobrog režisera. Takođe, određeni glumci se često zajedno pojavljuju u filmovima, bilo generalno, bilo određenog žanra. Isto tako, filmovi određenog žanra često imaju značajno bolji komercijalni uspeh.

### 3.2. Skup podataka za analizu

U okviru ovog projektnog zadatka je potrebno analizirati podatke o 1000 najpopularnijih filmova iz IMDb baze u periodu od 2006. do 2016. godine. Podaci za analizu (primarni skup podataka) su dostupni u vidu odgovarajuće *Comma Separated Value* (CSV) datoteke *IMDB-Movie-Data.csv* koja se nalazi u arhivi koja je priložena uz tekst projektnog zadatka.

Primarni skup podataka (eng. *primary dataset*) je javno dostupan na *Kaggle* stranici:

<https://www.kaggle.com/PromptCloudHQ/imdb-data>

Primarni skup podataka je dobijen upitom u odgovarajuću IMDb bazu kroz API koji je javno dostupan. Podaci su u određenoj meri prečišćeni, ali određeni podaci mogu biti nepotpuni ili nisu javno dostupni za sve filmove. Podaci se sastoje od sledećih jedanaest kolona:

- Rank – rang filma u skupu podataka
- Title – naslov filma
- Genre – lista žanrova kojima film pripada, odvojena zarezima
- Description – kratak opis filma u jednoj rečenici
- Director – ime režisera filma
- Actors – lista glavnih glumaca filma, odvojena zarezima
- Year – godina izdavanja filma
- Runtime – trajanje filma u minutima
- Rating – ocena filma od strane korisnika u opsegu od 0 do 10
- Votes – broj korisničkih glasova
- Revenue – zarada filma u milionima dolara

- Metascore – agregirani prosek ocena kritike. Vrednosti su u opsegu od 0 do 100, a više vrednosti predstavljaju pozitivne ocene.

Na osnovu primarnog skupa podataka treba formirati sekundarni skup podataka (eng. *secondary dataset*) koji predstavlja prečišćenu verziju podataka za analizu. Prečišćavanje izvršiti prema potrebama zadatka i ciljevima istraživanja. Prilikom prečišćavanja se mogu izostaviti svi nepotrebni podaci.

### 3.3. Modelovanje mreže

Sekundarni skup podataka je potrebno iskoristiti za modelovanje odgovarajućih socijalnih mreža. Potrebno je modelovati mrežu glumaca, mrežu žanrova i mrežu filmova. Prilikom modelovanja mreže implementirati odgovarajući tip mreže (usmerena, neusmerena, težinska i sl.) u skladu sa postavljenim istraživačkim pitanjima i ciljevima. Primarna mreža za analizu treba da bude mreža glumaca, a druge dve treba iskoristiti u funkciji odgovaranja na određena istraživačka pitanja. Po potrebi se mogu napraviti i analizirati i druge mreže na osnovu zadatog skupa podataka.

U okviru mreže glumaca, glumci treba da predstavljaju čvorove mreže, a vezu između dva čvora treba uspostaviti ukoliko su dva glumca igrala zajedno u bar jednom filmu. Više informacija o ovakvom tipu mreže se može videti na [https://en.wikipedia.org/wiki/Co-stardom\\_network](https://en.wikipedia.org/wiki/Co-stardom_network). Obratiti pažnju da se u zadatom skupu podataka nalaze samo podaci o glavnim glumcima u nekom filmu, što treba uzeti u obzir kod tumačenja dobijenih rezultata.

U okviru mreže žanrova, žanr filma treba da predstavlja čvor mreže, a vezu između dva čvora treba uspostaviti ukoliko postoji bar jedan film koji je klasifikovan u oba žanra. U okviru mreže filmova, film treba da predstavlja čvor mreže, a vezu između dva čvora treba uspostaviti ukoliko postoji bar jedan glumac koji je glumio u oba filma. Veze u okviru mreže filmova treba da budu usmerene na osnovu godine izdavanja filma, od starijeg ka novijem.

### 3.4. Istraživačka pitanja i ciljevi

Prilikom obrade primarnog i sekundarnog skupa podataka pogodno je kao smernice koristiti prethodno definisana istraživačka pitanja. U okviru ove sekcije je postavljen jedan broj takvih pitanja, a studenti treba da na osnovu analize problema i samih podataka definišu dodatna pitanja ili specijalizuju navedena čime mogu bliže usmeriti samu analizu.

- 1) Ko su glumci koji su glumili sa najviše drugih glumaca?
- 2) Koliki je prosečan broj glumaca sa kojima je jedan glumac igrao?
- 3) Ko su najproduktivniji glumici i u kojim žanrovima su najviše igrali?
- 4) Koje zajednice glumaca se mogu uočiti prilikom analize mreže?
- 5) Da li se glumci u mreži grupišu na osnovu filmskog žanra u kome najčešće glume?
- 6) Ko su glumci koji povezuju različite zajednice glumaca?
- 7) Kolika je gustina mreže?
- 8) U kojoj meri je mreža povezana i centralizovana?
- 9) Kolike su prosečne distance u okviru mreže i dijametar mreže?
- 10) Koliki je koeficijent klasterizacije mreže i njenih čvorova?
- 11) Kakva je distribucija čvorova po stepenu i da li prati neku zakonomernost?
- 12) Da li mreža iskazuje osobine malog sveta?
- 13) Kolika je prosečna udaljenost, a kolika maksimalna udaljenost nekog glumca od Kevina Bejkona (*Kevin Bacon*)? Više detalja se može videti na: [https://en.wikipedia.org/wiki/Six\\_Degrees\\_of\\_Kevin\\_Bacon](https://en.wikipedia.org/wiki/Six_Degrees_of_Kevin_Bacon)
- 14) Koji glumci predstavljaju jezgro mreže?
- 15) Koji filmski žanrovi su najpopularniji? U kojim kombinacijama se najčešće javljaju?
- 16) Koji filmovi su najviše uticali da njihovi glumci igraju u kasnijim filmovima?
- 17) Kako se svojstva mreže menjaju ukoliko se pre formiranja filmovi filtriraju po zaradi?
- 18) Koji režiser je režirao najveći broj filmova?
- 19) Da li režiseri imaju omiljene glumce koje često angažuju u svojim filmovima?
- 20) Koje godine je filmska produkcija bila najveća?

Da bi se odgovorilo na postavljena pitanja, potrebno je primeniti odgovarajuće mere i metode za analizu mreže ili statističke metode. Mrežu bi trebalo karakterisati kako kroz osnovna svojstva mreže, tako i kroz složenije mere centralnosti i metode za detekciju komuna. Mere i metode izabrati prema adekvatnosti spram postavljenog problema.

### 3.5. Preporučene metode i alati

Za analizu modelirane socijalne mreže se preporučuje korišćenje programskih jezika Python (NetworkX biblioteka) i R (*sna* i *igraph* paketi) ili softverskih alata Gephi, UCINET, ili Pajek. Obrada primarnog skupa podataka se može obaviti pomoću MS Excel alata ili pisanjem odgovarajućih skripti u programskom jeziku po izboru. Ukoliko nije moguće drugačije, razrešavanje eventualnih dvosmislenosti u primarnom skupu podataka izvršiti ručno.

Vizuelizacija mreže se može obaviti korišćenjem alata Gephi, NodeXL ili kroz podršku u okviru programskih jezika Python (*matplotlib* biblioteka) i R (*igraph* paket).

## 4. REZULTATI

Projektni zadatak se predaje u vidu pisanog izveštaja koji sadrži rezultate sprovedene analize i pisana objašnjenja uočenih fenomena. Uz izveštaj se dostavljaju i odgovarajuće dopunske datoteke, kao što su tabele sa rezultatima analize, izvorni programski kod skripti ili programa korišćenih u analizi, datoteke koje sadrže produkovane vizuelizacije i sl. Potpuno odsustvo dopunskih datoteke koje predstavljaju rezultate rada može povući umanjeње broja poena na projektnom zadatku. Za pisanje izveštaja se može koristiti šablon koji se nalazi u odgovarajućoj sekciji na sajtu predmeta. Preporučeni obim izveštaja je do 10 stranica teksta.

## 5. PREDAJA, ODBRANA I VREDNOVANJE

Projektni zadatak se predaje elektronskom poštom najkasnije do termina ispita u odgovarajućem ispitnom roku. Na odbranu je potrebno doneti štampanu verziju izveštaja. Po pravilu, projektni zadatak se brani pred predmetnim nastavnikom u ispitnom roku u kome student želi da polaže ispit. Ukoliko student želi da brani zadatak u nekom drugom terminu, treba o tome da blagovremeno obavesti predmetnog nastavnika, radi eventualnog dogovora. Ukoliko se projektni zadatak radi u paru, studenti zajedno brane projektni zadatak.

Projektni zadatak nosi 40 poena. Poeni sa jednom odbranjenog projektnog zadatka važe jednu školsku godinu. Postoji mogućnost da se dobro urađeni projektni zadaci prošire u završni, master rad.