

Analiza socijalnih mreža

Projektni zadatak
2018-2019.

Marko Mišić, Jelica Protić

13M111ASM

2018/2019.

Uvod

- Projektni zadatak na predmetu
Analiza socijalnih mreža se sastoji od dva zadatka
- Prvi projektni zadatak nosi 20 poena
 - Analiza i prezentovanje osnovnih rezultata jednog naučno-istraživačkog rada
- Drugi projektni zadatak nosi 40 poena
 - Analiza konkretnog istraživačkog problema
 - Realan skup podataka

Cilj projektnog zadatka

- Praktična primena stečenog teorijskog znanja na primeru konkretnog istraživačkog problema
- Prikupljanje, obrada i preliminarna analiza primarnog (sirovog) skupa podataka
- Modeliranje problema mrežom odgovarajućeg tipa
- Analiza modelirane mreže alatima za obradu socijalnih mreža po izboru
- Vizuelizaciju mreže
- Interpretacija rezultata u skladu sa postavljenim istraživačkim pitanjima

Prvi projektni zadatak

- Radovi za analizu su dostupni u arhivi na sajtu predmeta
 - Lozinka za pristup je saopštena na listi elektronske pošte
- Prilikom obrade konkretnog rada, odnosno problema treba obratiti pažnju na sledeće:
 - Oblast problema koji se rešava
 - Motivaciju za upotrebu tehnika za analizu socijalnih mreža i modeliranje problema mrežom
 - Način modeliranja problema mrežom
 - Tip grafa, način uspostavljanja veza i sl.
 - Primenjene tehnike i metode za analizu
 - Osnovna svojstva, centralnosti, klasterizacija...
 - Glavne rezultate analize
 - Vizuelizaciju (ako postoji)

Drugi projektni zadatak

- Analiza podataka o 1000 najpopularnijih filmova
 - Filmovi iz *Internet Movie Database* (IMDb)
 - Period od 2006. do 2016. godine
- IMDb baza sadrži podatke o filmovima, televizijskim serijama i drugim vidovima zabave:
 - Glumačke postave
 - Produkciono osoblje
 - Opisi filmova
 - Podatke o budžetu i zaradi, Ocene kritike i gledalaca i sl.

Skup podataka za analizu

- Primarni skup podataka čine podaci o filmovima:
 - 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 proseki ocena kritike. Vrednosti su u opsegu od 0 do 100, a više vrednosti predstavljaju pozitivne ocene.

Problemi primarnog skupa podataka

- CSV dokument
 - Problem parsiranja polja
 - Žanr, lista glumaca
 - Određena polja sama predstavljaju listu vrednosti
- Nepotpuni podaci za određene kolone
- Podaci samo o glavnim glumcima
 - Kako to utiče na rezultate?
- Jednoznačnost imena u dokumentu
 - Da li glumci i režiseri imaju jednoznačna imena?

Obrada primarnog skupa podataka

- Potrebna obrada i čišćenje primarnog skupa podataka
 - Excel za osnovna filtriranja
 - Skripte (Python, Java, R...) za formiranje mreže
 - Upotreba *hash* mapa ili *dictionary*-ja
 - Ručno razrešavanje kod dvosmislenosti
- Izbor međuformata i izlaznog formata
 - CSV ili XLS kao međuformati
 - Grafovski formati koji su podržanu u alatima za analizu i vizuelizaciju
 - GML, GEXF, UCINET DL
 - Varijante *edge* liste

Modelovanje mreže

- Koautorska težinska mreža glumaca
 - Dva glumca će biti povezana ukoliko su zajedno glumili u bar jednom filmu
 - Težinu grane će predstavljati broj zajedničkih pojavljivanja
 - Jedinična težina ili frakciono brojanje
 - Projekcija bipartitne mreže glumac-film
- Mreža žanrova
 - Dva žanra će biti povezana ukoliko postoji bar jedan film koji je klasifikovan u oba žanra
- Mreža filmova
 - Dva filma će biti povezana ukoliko postoji bar jedan glumac koji je glumio u oba filma
 - Usmerena mreža na osnovu godine izdavanja

Istraživačka pitanja (1)

- Ko su glumci koji su glumili sa najviše drugih glumaca?
- Koliki je prosečan broj glumaca sa kojima je jedan glumac igrao?
- Ko su najproduktivniji glumici i u kojim žanrovima su najviše igrali?
- Koje zajednice glumaca se mogu uočiti prilikom analize mreže?
- Da li se glumci u mreži grupišu na osnovu filmskog žanra u kome najčešće glume?
- Ko su glumci koji povezuju različite zajednice glumaca?

Istraživačka pitanja (2)

- Kolika je gustina mreže?
- U kojoj meri je mreža povezana i centralizovana?
- Kolike su prosečne distance u okviru mreže i dijametar mreže?
- Koliki je koeficijent klasterizacije mreže i njenih čvorova?
- Kakva je distribucija čvorova po stepenu i da li prati neku zakonomernost?

Istraživačka pitanja (3)

- Da li mreža iskazuje osobine malog sveta?
- 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
- Koji glumci predstavljaju jezgro mreže?

Istraživačka pitanja (4)

- Koji filmski žanrovi su najpopularniji? U kojim kombinacijama se najčešće javljaju?
- Koji filmovi su najviše uticali da njihovi glumci igraju u kasnijim filmovima?
- Kako se svojstva mreže menjaju ukoliko se pre formiranja filmovi filtriraju po zaradi?
- Koji režiser je režirao najveći broj filmova?
- Da li režiseri imaju omiljene glumce koje često angažuju u svojim filmovima?
- Koje godine je filmska produkcija bila najveća?

Metode i alati za analizu

- Alati za analizu modelirane socijalne mreže
 - Programski jezik Python
 - *NetworkX* biblioteka
 - UCINET, Gephi
 - Programski jezik R
 - *sna* i *igraph* paketi
- Vizuelizacija mreže
 - Gephi, NodeXL
 - Programski jezik Python
 - *matplotlib* biblioteka
 - Programski jezik R
 - *igraph* paket

Tehnički detalji

- Radi se samostalno ili u paru
- Predaja pred ispit elektronskom poštom
- Rezultati u obliku pisanog izveštaja
 - Preporučeni obim do 10 stranica teksta
 - Dopunske datoteke sa rezultatima rada, kodom i sl.
 - Šablon dostupan na sajtu predmeta
- Poeni važe jednu školsku godinu

Literatura

- D. Hansen, B. Shneiderman, M. Smith, Analyzing Social Media Networks with NodeXL: Insights from a Connected World, Morgan Kaufmann, 2010.
- Charles Kadushin, Understanding Social Networks: Theories, Concepts and Findings, Oxford University Press, 2012
- Christina Prell, Social Network Analysis: History, Theory and Methodology, SAGE Publications Ltd, 2012.
- Hanneman, Robert A. and Mark Riddle, Introduction to social network methods, University of California, Riverside, 2005.