

UNIVERZITET U BEOGRADU
ELEKTROTEHNIČKI FAKULTET



ANALIZA SOCIJALNIH MREŽA

Projektni zadatak

Verzija 1.0

Predmetni nastavnici:

dr Marko Mišić, vanredni profesor

dr Jelica Protić, redovni profesor

Školska godina:

2023/2024.

Predmetni saradnik:

Predrag Obradović, asistent

Beograd, decembar 2023.

SADRŽAJ

SADRŽAJ	2
1. UVOD	3
2. CILJ	3
3. POSTAVLJENI PROBLEM	3
3.1. ANALIZA NAUČNE PRODUKCIJE IZ OBLASTI EPIDEMIOLOGIJE I ZARAZNIH BOLESTI NA MEDICINSKOM FAKULTETU UNIVERZITETU U BEOGRADU	3
3.2. SKUP PODATAKA ZA ANALIZU	4
3.3. MODELOVANJE MREŽE	6
3.4. ISTRAŽIVAČKA PITANJA I CILJEVI	6
3.4.1. Statistička obrada podataka [5 poena].....	7
3.4.2. Osnovna karakterizacija modelovanih mreža [10 poena]	7
3.4.3. Analiza mera centralnosti [5 poena]	8
3.4.4. Detekcija komuna Luvenskom metodom [5 poena]	8
3.4.5. Detekcija komuna spektralnim klasterisanjem [10 poena].....	8
3.4.6. Analiza mreže časopisa [5 poena].....	9
3.5. PREPORUČENE METODE I ALATI.....	9
4. REZULTATI	9
5. PREDAJA, ODBRANA I VREDNOVANJE	9
LITERATURA	10

1. UVOD

U okviru ovog dokumenta su data uputstva za izradu projektnog zadatka na predmetu Analiza socijalnih mreža (13M111ASM) u školskoj 2023/2024. 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 i predstaviti u obliku izveštaja.

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 naučne produkcije iz oblasti epidemiologije i zaraznih bolesti na Medicinskom fakultetu Univerzitetu u Beogradu

Pojava neočekivano velikog broja slučajeva određene bolesti u nekoj populaciji ili regionu u toku određenog vremenskog perioda, poznatija kao epidemija, vekovima je poznata čovečanstvu. Kuga u srednjem veku, kolera u XIX veku i španska influenza nakon Prvog svetskog rata su samo neke od zaraznih bolesti koje su za kratko vreme odnele više desetina hiljada života. Globalizacija u XX veku je olakšala međunarodnu razmenu ljudi i dobara. Međutim ista pojava je omogućila i da se zarazne bolesti još brže šire. Pandemija koronavirusa *SARS-CoV-2* (*Severe Acute Respiratory*

Syndrome Coronavirus 2) koja je zadesila svet početkom 2020. godine je upravo primer koji dokazuje prethodnu tezu, ali i primer koji ukazuje na značaj naučnih disciplina iz oblasti medicine čiji je predmet proučavanja povezan sa epidemijama u današnjem društvu.

Predmet projektnog zadatka u tekućoj školskoj godini je kvantitativna i kvalitativna analiza naučne produkcije Medicinskog fakulteta Univerziteta u Beogradu, uz razmatranje samo onih naučnih oblasti čiji je predmet proučavanja povezan sa epidemijama zaraznih bolesti. U tom smislu, potrebno je analizirati naučne radove i kolaboracije zaposlenih na određenim katedrama Medicinskog fakulteta Univerziteta u Beogradu. Katedre koje će biti posmatrane u okviru ovog projekta su Katedra za imunologiju, Katedra za epidemiologiju, Katedra za infektivne bolesti i Katedra za mikrobiologiju. Cilj projektnog zadatka je prikupljanje, obrada i analiza naučne produkcije, kao i saradnje zaposlenih unutar fakulteta, koristeći izdvajanje podataka, modeliranje i vizuelizaciju istih mrežama odgovarajućeg tipa. Rezultati treba da pokažu stanje naučne produkcije i kvaliteta naučnih istraživanja koja se obavljaju na pomenutim katedrama, meru saradnje u okviru posmatrane naučnoistraživačke zajednice, poziciju istraživanja iz Srbije u svetu i ukažu na istaknute pojedince.

Najveći deo naučne produkcije se odvija kroz publikacije kao što su knjige, disertacije, radovi u naučnim časopisima i radovi na međunarodnim i domaćim naučnim konferencijama. Međutim, u smislu evaluacije, najviše se cene radovi u naučnim časopisima i radovi na naučno-stručnim konferencijama. U okviru ovog projektnog zadatka stavljen je fokus na produkciju radova u naučnim časopisima sa *impact* faktorom.

Mreže naučne kolaboracije se uobičajeno modeluju putem grafa. Postoji veći broj načina da se to učini, a primeri sličnih analiza se mogu videti u [1][2][3].

3.2. Skup podataka za analizu

U okviru predloženog projektnog zadatka je potrebno analizirati naučnu produkciju iz oblasti epidemiologije i zaraznih bolesti sa četiri katedre Medicinskog fakulteta Univerziteta u Beogradu: Katedra za imunologiju, Katedra za epidemiologiju, Katedra za infektivne bolesti i Katedra za mikrobiologiju. Potrebno je analizirati naučnu produkciju koja se odnosi na radove u časopisima sa *impact* faktorom formirane na osnovu upita u bazu *Scopus*. Podaci za analizu (primarni skup podataka) su dostupni u vidu odgovarajućih *Excel* tabela u arhivi koja je priložena uz tekst projektnog zadatka.

Za prikupljanje podataka o naučnim publikacijama su korišćeni podaci o autorima koji su se vodili kao zaposleni u nastavnim zvanjima na prethodno pomenutim katedrama u oktobru 2020., što je prikupljeno sa internet prezentacija ovih fakulteta. Podaci su dostupni u okviru *Excel* datoteke *autori.xlsx*. Ostali autori naučnih publikacija koji se javljaju u okviru skupa podataka za analizu nisu od interesa za analizu.

Primarni skup podataka za analizu (eng. *primary dataset*) je dobijen prikupljanjem podataka iz indeksne baze naučnih radova *Scopus* zaključno sa oktobrom 2020. U skupu podataka se nalazi određeni broj radova prikupljenih na osnovu dostupnih podataka o zaposlenima na ranije pomenute četiri katedre Medicinskog fakulteta UB. Podaci su dostupni u okviru *Excel* datoteka *epidemiologija.xlsx*, *imunologija.xlsx*, *infektivne_bolesti.xlsx*, *mikrobiologija.xlsx*. Podaci su u određenoj meri prečišćeni, ali određeni podaci mogu biti nepotpuni ili u neodgovarajućem obliku. Podaci se sastoje od sledećih kolona:

- *Author* – ime zaposlenog kome je pridružena odgovarajuća publikacija u *Scopus* bazi,
- *Authors* – svi autori rada,
- *Title* – naslov publikacije,
- *Year* – godina izdavanja,
- *Source title* – ime časopisa u kome je rad objavljen,
- *Volume* – volumen časopisa u kome je rad objavljen,
- *Issue* – redni broj sveske časopisa u kome je rad objavljen,
- *Art. No.* – identifikacioni broj rada,
- *Page start* – početna stranica rada u časopisu,
- *Page end* – poslednja stranica rada u časopisu,
- *Page count* – ukupan broj stranica rada,
- *Cited by* – broj citata drugih autora prema *Scopus* bazi,
- *Link* – link ka izvoru,
- *Document Type* – tip rada,
- *Source* – indeksna baza koja predstavlja izvor podataka (*Scopus*).

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. U analizi su od interesa samo publikacije tipa *Article, Article in Press, Review, Book Chapter, Letter, Note*, dok se ostale mogu zanemariti. 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 koautora (na nivou katedri i fakulteta) i mrežu časopisa. 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 koautora na nivou fakulteta, a preostale treba iskoristiti u funkciji odgovaranja na određena istraživačka pitanja. Po potrebi se mogu napraviti i analizirati i druge mreže na osnovu zadanog skupa podataka.

U okviru mreže koautora, zaposleni na Medicinskom fakultetu UB koji se bave proučavanjem epidemija i zaraznih bolesti treba da predstavljaju čvorove mreže, a vezu između dva čvora treba uspostaviti ukoliko su dva zaposlena napisala bar jedan naučni rad zajedno. Težinu grane (veze) modelovati celovitim ili frakcionim brojanjem.

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, nakon što odgovore na ova pitanja, na osnovu analize problema i samih podataka definišu dodatna pitanja ili specijalizuju navedena čime mogu bliže usmeriti samu analizu. Odgovore na pitanja treba dati u formi specificiranoj u poglavlju 4.

Pitanja su grupisana u kategorije po tematici koju obrađuju i tehnikama analize koje se u njima sprovode. Ako nije drugačije navedeno, svako od pitanja se odnosi na svaku od konstruisanih mreže, ukoliko je za nju smisljeno i primenljivo. Studentima se predlaže da prvo sve analize sprovedu za jednu od mreža, pa zatim pređu na sledeću i da svoje odgovore na taj način strukturiraju i izlože u izveštaju o projektu.

Da bi se odgovorilo na postavljena pitanja, potrebno je primeniti odgovarajuće mere i metode za analizu mreže ili statističke metode. Mreže 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. Tamo gde se očekuje odgovor u obliku neke vrste rangiranja, navesti listu od 5 do 10 najrelevantnijih rezultata.

3.4.1. Statistička obrada podataka [5 poena]

- 1) Koliki je broj radova po svakom autoru? Koristiti i celovito i frakciono brojanje. Ko su najproduktivniji naučnici iz oblasti istraživanja i kojoj katedri pripadaju?
- 2) Koliki je prosečan broj koautora po svakom autoru?
- 3) Na osnovu dostupnih podataka, odrediti H-indeks svakog od naučnika i uporediti ga sa dostupnim H-indeksom u datoteci *autori.xlsx*.
- 4) Koje katedre su najproduktivnije posmatrajući naučnu produkciju i citiranost u časopisima na osnovu dostupnih podataka?
- 5) U kojim godinama su autori bili najproduktivniji na nivou fakulteta i pojedinačnim katedrama?
- 6) U kojim časopisima se u proseku najviše objavljuje?
- 7) Da li postoje razlike između katedri u smislu obima i učestanosti publikovanja u časopisima?
- 8) Da li postoji razlika u prosečnom broju autora po radovima u časopisima po katedrama?
- 9) Kakav je odnos broja koautora sa fakulteta u odnosu na broj autora van fakulteta po katedrama i na nivou celog fakulteta?

3.4.2. Osnovna karakterizacija modelovanih mreža [10 poena]

- 10) Kolika je gustina mreže?
- 11) Kolike su prosečne distance u okviru mreže i dijametar mreže?
- 12) U kojoj meri je mreža povezana i centralizovana? Navesti broj i veličine povezanih komponenata i proceniti da li postoji gigantska komponenta.
- 13) Koliki je prosečni, a koliki globalni koeficijent klasterizacije mreže? Kakva je raspodela lokalnog koeficijenta klasterizacije njenih čvorova? Da li je klasterisanje izraženo ili ne? Odgovor dati upoređivanjem sa slučajno generisanim *Erdos-Renyi* i *scale free* mrežama istih dimenzija.
- 14) U kojoj meri autori imaju tendenciju da pišu publikacije sa istim koautorima?
- 15) Na osnovu odgovora na pitanja 10 i 12, proceniti da li mreža iskazuje osobine malog sveta.

- 16) Izvršiti asortativnu analizu po stepenu čvora i dati odgovor da li je i koliko izraženo asortativno mešanje. Priložiti i vizuelizaciju.
- 17) Da li mreža ispoljava fenomen kluba bogatih (eng. *rich club phenomenon*)? Analizu sprovesti poređenjem sa adekvatnom mrežom dobijenom Havel-Hakimi algoritmom.
- 18) Kakva je distribucija čvorova po stepenu i da li prati *power law* raspodelu?

3.4.3. Analiza mera centralnosti [5 poena]

- 19) Sprovesti analize centralnosti po stepenu, bliskosti i relacionoj centralnosti. Dati pregled najvažnijih aktera po svakoj od njih. Koji autori predstavljaju centre okupljanja u okviru svojih katedri i na nivou fakulteta?
- 20) Ko su najvažniji akteri po centralnosti po sopstvenom vektoru? Šta nam to govori o njima?
- 21) Na osnovu prethodna dva pitanja predložiti i konstruisati heuristiku (kompozitnu meru centralnosti) za pronalaženje najvažnijih aktera i pronaći ih. Obratiti pažnju na tip mreže koji se analizira (usmerena ili neusmerena) i, shodno tome, prilagoditi koliko različite mrežne metrike utiču na heuristiku.
- 22) Da li autori sa različitih katedri međusobno sarađuju i u kojoj meri? Ko su autori koji povezuju različite grupe u okviru mreže? Da li na osnovu ove analize postoji spoljni autor koga smatrate da bi bilo veoma korisno zaposliti na Medicinskom fakultetu?

3.4.4. Detekcija komuna Luvenskom metodom [5 poena]

- 23) Sprovesti klasterisanje Luvenskom metodom (maksimizacijom modularnosti) u alatu *Gephi* za tri različite vrednosti parametra rezolucije. Konstruisati vizuelizacije i diskutovati izbor parametra rezolucije na dobijeno klasterisanje (broj i veličina klastera).
- 24) Koje zajednice (komune) se mogu uočiti prilikom analize mreže, a koji akteri su ključni brokeri? Da li postoji neko objašnjenje za detektovane komune?

3.4.5. Detekcija komuna spektralnim klasterisanjem [10 poena]

- 25) Sprovesti spektralnu analizu i proceniti potencijalne kandidate za broj komuna u mreži. Uporediti rezultat sa dendogramom konstruisanim *Girvan-Newman* metodom, ukoliko je primenjivo na nivou fakulteta ili katedri.
- 26) Ko su akteri koji se mogu okarakterisati kao ključni brokeri (mostovi) u mreži? Šta ih čini brokerima? Porediti odgovor sa brokerima dobijenim u pitanju 22 i 24.

3.4.6. Analiza mreže časopisa [5 poena]

27) Ponoviti sve analize koje imaju smisla za mrežu časopisa.

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, Cytoscape ili kroz podršku u okviru programskih jezika *Python* (*matplotlib*, *graphviz* i *graph-tool* biblioteke) 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 elektronskim putem najkasnije do termina ispita u odgovarajućem ispitnom roku na način kako to bude specificirao predmetni nastavnik. Na odbranu je potrebno doneti štampanu verziju izveštaja. Po pravilu, projektni zadatak se brani pred predmetnim nastavnikom ili saradnikom 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. Raspodela poena po tematskim oblastima je prikazana u sekciji 3.4. Studenti ne moraju realizovati sve zahteve u okviru projekta pre odbrane. Ne postoji minimalan broj poena koji je potrebno osvojiti na projektnom zadatku da bi se položio ispit.

Poeni sa jednom odbranjenog projektnog zadatka važe jednu školsku godinu. Postoji mogućnost da se dobro urađeni projektni zadaci prošire u završni, master rad. Upit u vezi sa takvom mogućnošću studenti mogu uputiti predmetnom nastavniku ili saradniku.

LITERATURA

- [1] Milovančević, D., Mišić, M., Protić, J., *Mrežna analiza naučne kolaboracije zaposlenih na Elektrotehničkom fakultetu u Beogradu na osnovu institucionalne evidencije objavljenih radova*, XXIV Skup TRENDovi RAZVOJA, Kopaonik, Serbia, 2019. http://www.trend.uns.ac.rs/stskup/trend_2018/radovi/T2-2/T2.2-4.pdf
- [2] Živanović, Đ., Mišić, M., Milovančević, D., Protić, J., *Uporedna analiza naučne produkcije iz oblasti računarstva na Univerzitetu u Beogradu*, XXV Skup TRENDovi RAZVOJA, Kopaonik, Serbia, 2019. http://www.trend.uns.ac.rs/stskup/trend_2019/radovi/T3.1/T3.1-3.pdf
- [3] Mitrović, I., Mišić, M., Protić, J., *Exploring high scientific productivity in international co-authorship of a small developing country based on collaboration patterns*, Journal of big Data, vol. 10, no. 1, 64, 2023. <https://link.springer.com/article/10.1186/s40537-023-00744-1>