

## Upravljanje znanjem – Teme za projektni zadatak 1 - 2019/20

**Zadatak:** Primijeniti metode iz područja analize kompleksnih mreža na zadane mreže iz odabrane domene. Zadano je pet domena i za svaku domenu ukupno pet različitih varijanti zadataka.

### Domene

1. Analiza mreža suradništva u znanstvenim publikacijama
2. Analiza mreža jezika (identifikacija i rangiranje ključnih pojmova u mreži)
3. Analiza mreža životinjskih interakcija
4. Analiza ekonomskih mreža
5. Analiza infrastrukturnih mreža

### OPĆE UPUTE ZA SVE ZADATKE

- Vodite računa o tome kakva mreža najbolje odgovara za podatke s kojima radite: usmjerena/neusmjerena; težinska/netežinska
- U zadacima u kojima student sam bira podatke za mrežu (od ponuđenih mreže za domene 3,4 i 5), pazite da ne birate mreže s prevelikim brojem čvorova/veza jer će se sporo procesirati (ne više od 10000 čvorova)
- Za neke zadatke nije navedeno koje sve mjere je potrebno koristiti u prvom dijelu analize. Za te zadatke bit će naknadno (kroz par dana) objavljen detaljni popis globalnih, središnjih i lokalnih mjera.

### Zadaci

#### DOMENA 1 - analiza mreža kolaboracija na znanstvenim radovima

##### Upute za generiranje mreže za sve zadatke iz domene 1.

Podaci se mogu naći na mrežnim stranicama CROSBİ znanstvene bibliografije:

<https://www.bib.irb.hr/>

Potrebno je prikupiti podatke iz CROSBİ baze podataka o suradnji na znanstvenim publikacijama zadane sastavnice. Na početnoj web stranici nudi se mogućnost odabira institucije/sastavnice.

Iz danih podataka potrebno je konstruirati mrežu kolaboracija na znanstvenim radovima u kojoj su **čvorovi znanstvenici, a veza između dva čvora uspostavljena je ukoliko su ta dva znanstvenika objavila zajedničku publikaciju**. Publikacije koje je pisao samo 1 autor ne uzimaju se u obzir.

Osnovni podaci koje želimo pohraniti dani su u sljedećem primjeru.

Konkretno, za rad koji je u CROSBİ bazi podataka zapisan na sljedeći način:

18. Uchytıl, Andrej; Barišić Marenić, Zrinka; Kahrović, Emir.

[Lexicon of Architects - Atlas of 20th Century Croatian Architecture](#) .

Zagreb : Arhitektonski fakultet Sveučilišta u Zagrebu, 2011 (leksikon arhitekata).

potrebno je ekstrahirati podatke potrebne za konstrukciju mreže prema prethodno prikazanoj formi, a to će izgledati kao što je prikazano u nastavku (ukoliko se dvoje autora pojavljuje više puta, težina im se povećava - odnosno 3 komponenta trojke mora biti točan broj radova na kojima su autori surađivali):

"Uchytíl, Andrej";"Barisic Marenic, Zrinka";1

"Uchytíl, Andrej";"Kahrović, Emir";1

"Barisic Marenic, Zrinka";"Kahrović, Emir";1

#### **ZAD 1.1.**

Analiza i usporedba mreža kolaboracija za: **Odjel za informatiku i Tehnički fakultet**

Za navedene sastavnice potrebno je pohraniti podatke u mrežu na način da se osim osnovnih podataka pohrani i godina kada je publikacija objavljena (kao dodatni atribut veze).

U prvome dijelu analize potrebno je izračunati sve globalne, središnje i lokalne mjere mreže i u tablici prikazati rezultate, te napraviti vizualizaciju mreže na više načina: vizualizacija čvorova na način da ih se poveća prema svim mjerama centralnosti koje su dostupne u alatu Gephi.

Posebni dio zadatka: PREDVIĐANJE BUDUĆIH LINKOVA. Proučiti i isprobati algoritme za predviđanje budućih veza u mreži koji su dostupni u paketu NetworkX a mogu se koristiti za odabrani skup podataka primjenom sljedećih algoritama: Resource allocation index, Jaccard coefficient.

Potrebno je primijeniti i evaluirati zadane algoritme za predviđanje poveznica u jednoj od mreža (odabrati mrežu koja je veća) na način da se izbace sve veze koje su nastale u zadnje tri godine te da se primjenom zadanih algoritama predvidi koje veze nastaju (isti broj veza koji je izbačen) i na temelju toga izračunati preciznost za svaki od algoritama.

#### **ZAD 1.2.**

Analiza i usporedba mreža kolaboracija za: **Odjel za matematiku i Ekonomski fakultet**

Za navedene sastavnice potrebno je pohraniti podatke u mrežu na način da se osim osnovnih podataka pohrani i godina kada je publikacija objavljena (kao dodatni atribut veze).

U prvome dijelu analize potrebno je izračunati sve globalne, središnje i lokalne mjere mreže i u tablici prikazati rezultate, te napraviti vizualizaciju mreže na više načina: vizualizacija čvorova na način da ih se poveća prema svim mjerama centralnosti koje su dostupne u alatu Gephi.

Posebni dio zadatka: PREDVIĐANJE BUDUĆIH LINKOVA. Proučiti i isprobati algoritme za predviđanje budućih veza u mreži koji su dostupni u paketu NetworkX a mogu se koristiti za odabrani skup podataka primjenom sljedećih algoritama: adamic\_adar\_index, preferential\_attachment.

Potrebno je primijeniti i evaluirati zadane algoritme za predviđanje poveznica u jednoj od mreža (odabrati mrežu koja je veća) na način da se izbace sve veze koje su nastale u zadnje tri godine te da se primjenom zadanih algoritama predvidi koje veze nastaju (isti broj veza koji je izbačen) i na temelju toga izračunati preciznost za svaki od algoritama.

### ZAD 1.3.

Analiza i usporedba mreža kolaboracija za: **Odjel za biotehnologiju i Medicinski fakultet**

Za navedene sastavnice potrebno je pohraniti podatke u mrežu na način da se osim osnovnih podataka pohrani i godina kada je publikacija objavljena (kao dodatni atribut veze).

U prvome dijelu analize potrebno je izračunati sve globalne, središnje i lokalne mjere mreže i u tablici prikazati rezultate, te napraviti vizualizaciju mreže na više načina: vizualizacija čvorova na način da ih se poveća prema svim mjerama centralnosti koje su dostupne u alatu Gephi.

Posebni dio zadatka: PREDVIĐANJE BUDUĆIH LINKOVA. Proučiti i isprobati algoritme za predviđanje budućih veza u mreži koji su dostupni u paketu NetworkX a mogu se koristiti za odabrani skup podataka primjenom sljedećih algoritama: Resource allocation index, preferential\_attachment. Potrebno je primijeniti i evaluirati zadane algoritme za predviđanje poveznica u jednoj od mreža (odabrati mrežu koja je veća) na način da se izbacе sve veze koje su nastale u zadnje tri godine te da se primjenom zadanih algoritama predvidi koje veze nastaju (isti broj veza koji je izbačen) i na temelju toga izračunati preciznost za svaki od algoritama.

### ZAD 1.4.

Analiza i usporedba mreža kolaboracija za: **Odjel za informatiku i Filozofski fakultet**

Za navedene sastavnice potrebno je pohraniti podatke u mrežu.

U prvome dijelu analize potrebno je izračunati sve globalne, središnje i lokalne mjere mreže i u tablici prikazati rezultate, te napraviti vizualizaciju mreže na više načina: vizualizacija čvorova na način da ih se poveća prema svim mjerama centralnosti koje su dostupne u alatu Gephi.

Posebni dio zadatka: U drugom dijelu zadatka potrebno je analizirati mreže na središnjoj razini (odreditи k-core, centar mreže i periferiju mreže - za svaku mrežu). Analizirati i usporediti dva algoritma za otkrivanje zajednica, (Luvain algoritam implementiran u Pythonu i GN algoritam - <https://github.com/kjahan/community>).

### ZAD 1.5.

Analiza i usporedba mreža kolaboracija za: **Odjel za informatiku i Građevinski fakultet**

Za navedene sastavnice potrebno je pohraniti podatke u mrežu na način da se osim osnovnih podataka pohrani i godina kada je publikacija objavljena (kao dodatni atribut veze).

U prvome dijelu analize potrebno je izračunati sve globalne, središnje i lokalne mjere mreže i u tablici prikazati rezultate, te napraviti vizualizaciju mreže na više načina: vizualizacija čvorova na način da ih se poveća prema svim mjerama centralnosti koje su dostupne u alatu Gephi.

Posebni dio zadatka: U drugom dijelu zadatka potrebno je pratiti rast mreža kroz godine (2007-danas; ukupno 10 mreža po sastavnici). Za 5 odabranih globalnih mjera mreže generirati grafove koji prikazuje kako se mjera mijenja kroz vrijeme (na jednom grafu neka se paralelno prikazuju podaci za sve 3 sastavnice u različitim bojama).

## DOMENA 2 - Analiza jezičnih mreža

Konačan cilj projekta je konstruiranje jezične mreže iz zadanih tekstualnih podataka i analiziranje riječi koje su ključne u mreži (nalaze se na „centralnim“ pozicijama u mreži), tj. koje su ključne za pojedini tekst. Ključne riječi sažeto i jezgrovito opisuju tematiku teksta. Na temelju ključnih riječi možemo kasnije razviti algoritam za automatsko sumarizaciju teksta, tj. kreiranje sažetaka teksta (kratkog sadržaja).

### ZAD 2.1.

**PRIKUPLJANJE PODATAKA.** Potrebno je prikupiti podatke (tekstove) koji će se analizirati. Preporuča se koristiti automatsku tehniku prikupljanja podataka s Weba - engl. *Web scraping*. Analizirat će se tekstovi iz sportske domene. Potrebno je na web portalu 24sata.hr prikupiti najmanje 10 tekstova iz kategorije sport (adresa: <https://www.24sata.hr/sport>). Mogu se preuzeti tekstovi iz bilo koje podkategorije Nogomet, Rukomet, Košarka, Vaterpolo, itd. Potrebno je preuzeti tekstualni dio vijesti uključujući sam tekst vijesti zajedno s naslovima i podnaslovima, ali bez slika. Za svaku vijest u zasebni dokument potrebno je preuzeti i #tagove (npr. #dinamo, #manchester city, #liga prvaka). Iz tagova je potrebno očistiti znak # te ostaviti samo riječi. Preuzeti smo one tekstove koji nemaju manje od 3 #taga.

**PRIPREMA PODATAKA.** Potrebno je na temelju svakog preuzetog teksta generirati usmjerenu, težinsku mrežu. Riječi iz teksta predstavljaju čvorove, a veze između čvorova postoje samo ako su riječi u tekstu susjedne. Smjer veza u mreži određen je redoslijedom čitanja rečenice, tj. riječi u rečenici (s lijeva na desno). Težine veza u grafu predstavljene su brojem frekvencija pojavljivanja pojedinog para riječi u samome tekstu. Prije gradnje mreže, tekstove je potrebno lematizirati. (Za algoritam lematizacije javiti se na [sbeliga@uniri.hr](mailto:sbeliga@uniri.hr)).

Primjer: za rečenice „Danas je sunčan dan. Danas je lijep sunčan dan.“ egde lista izgledala bi ovako:

danas je 2  
je sunčan 1  
sunčan dan 2  
je lijep 1  
lijep sunčan 1

**ANALIZA MREŽE.** Konstruiranoj mreži potrebno je analizirati na lokalnoj razini. Za konstruiranu mrežu svakom čvoru je potrebno izračunati određene mjere centralnosti te rangirati čvorove prema dobivenim vrijednostima centralnosti. Analizirati degree centrality, closeness i eigenvector centrality. Čvorove iz mreže je potrebno rangirati prema dobivenim vrijednostima pojedine centralnosti te za svaku centralnost izdvojiti po 5 ključnih riječi koje su postigle najbolju vrijednost. Dakle za svaki tekst odvojiti zasebnih 5 ključnih riječi koje su najbolje za pojedinu centralnost. Ukupno 15 riječi za svaki tekst.

**ANALIZA REZULTATA.** Rezultate prikazati u tablicama. U stupcima prikazati vrijednosti izmjerenih rezultata (koristiti standardne mjere Precision, Recall i F1-score) za pojedinu mjeru centralnosti. Mjere se mogu izračunati u Pythonu (npr. scikit-learn biblioteka). P, R i F1 računati na način da riječi dobivene iz mreže na temelju pojedine mjere centralnosti usporedite s riječima (#tagovima) koje ste prikupili za pojedini tekst.

## ZAD 2.2.

**PRIKUPLJANJE PODATAKA.** Potrebno je prikupiti podatke (tekstove) koji će se. Preporuča se koristiti automatsku tehniku prikupljanja podataka s Weba - engl. *Web scraping*. Analizirat će se tekstovi iz domene *show businessa*. Potrebno je na web portalu 24sata.hr prikupiti najmanje 10 tekstova iz kategorije sport (adresa: <https://www.24sata.hr/show>). Mogu se preuzeti tekstovi iz bilo koje podkategorije Muzika, filmovi i serijem domaće zvijezde, itd. Potrebno je preuzeti tekstualni dio vijesti uključujući sam tekst vijesti zajedno s naslovima i podnaslovima, ali bez slika. Za svaku vijest u zasebni dokument potrebno je preuzeti i #tagove (npr. #anne hathaway, #glumica, #film). Iz tagova je potrebno očistiti znak # te ostaviti samo riječi. Preuzeti smo one tekstove koji nemaju manje od 3 #taga.

**PRIPREMA PODATAKA.** Potrebno je na temelju svakog preuzetog teksta generirati usmjerenu, težinsku mrežu. Riječi iz teksta predstavljaju čvorove, a veze između čvorova postoje samo ako su riječi u tekstu susjedne. Smjer veza u mreži određen je redoslijedom čitanja rečenice, tj. riječi u rečenici (s lijeva na desno). Težine veza u grafu predstavljene su brojem frekvencija pojavljivanja pojedinog para riječi u samome tekstu. Prije gradnje mreže, tekstove je potrebno lematizirati. (Za algoritam lematizacije javiti se na [sbeliga@uniri.hr](mailto:sbeliga@uniri.hr)).

Primjer: za rečenice „Danas je sunčan dan. Danas je lijep sunčan dan.“ egde lista izgledala bi ovako:

danas je 2  
je sunčan 1  
sunčan dan 2  
je lijep 1  
lijep sunčan 1

**ANALIZA MREŽE.** Konstruiranoj mreži potrebno je analizirati na lokalnoj razini. Za konstruiranu mrežu svakom čvoru je potrebno izračunati određene mjere centralnosti te rangirati čvorove prema dobivenim vrijednostima centralnosti. Analizirati degree centrality, betweenness centrality i PageRank. Čvorove iz mreže je potrebno rangirati prema dobivenim vrijednostima pojedine centralnosti te za svaku centralnost izdvojiti po 5 ključnih riječi koje su postigle najbolju vrijednost. Dakle za svaki tekst odvojiti zasebnih 5 ključnih riječi koje su najbolje za pojedinu centralnost. Ukupno 15 riječi za svaki tekst.

**ANALIZA REZULTATA.** Rezultate prikazati u tablicama. U stupcima prikazati vrijednosti izmjerenih rezultata (koristiti standardne mjere Precision, Recall i F1-score) za pojedinu mjeru centralnosti. Mjere se mogu izračunati u Pythonu (npr. scikit-learn biblioteka). P, R i F1 računati na način da riječi dobivene iz mreže na temelju pojedine mjere centralnosti usporedite s riječima (#tagovima) koje ste prikupili za pojedini tekst.

## ZAD 2.3.

**PRIKUPLJANJE PODATAKA.** Potrebno je prikupiti podatke (tekstove) koji će se. Preporuča se koristiti automatsku tehniku prikupljanja podataka s Weba - engl. *Web scraping*. Analizirat će se tekstovi općenitih Vijesti dana. Potrebno je na web portalu 24sata.hr prikupiti najmanje 10 tekstova iz kategorije News (adresa: <https://www.24sata.hr/news>). Mogu se preuzeti tekstovi iz bilo koje podkategorije politika, svijet, crna kronika, itd. Potrebno je preuzeti tekstualni dio vijesti uključujući sam tekst vijesti zajedno s naslovima i podnaslovima, ali bez slika. Za svaku vijest u zasebni dokument potrebno je preuzeti i #tagove (npr. #vladimir putin, #bijela kuća, #predsjednik, #izbori).

Iz tagova je potrebno očistiti znak # te ostaviti samo riječi. Preuzeti smo one tekstove koji nemaju manje od 3 #taga.

**PRIPREMA PODATAKA.** Potrebno je na temelju svakog preuzetog teksta generirati usmjerenu, težinsku mrežu. Riječi iz teksta predstavljaju čvorove, a veze između čvorova postoje samo ako su riječi u tekstu susjedne. Smjer veza u mreži određen je redoslijedom čitanja rečenice, tj. riječi u rečenici (s lijeva na desno). Težine veza u grafu predstavljene su brojem frekvencija pojavljivanja pojedinog para riječi u samome tekstu. Dodatno, iz mreže je potrebno izostaviti stop riječi kao što su veznici, čestice i sl. To su riječi koje se u tekstu javljaju često, a ne nose nikakvo značenje, npr. riječi i, ili, ni, niti, a, ali, nego, već, no, ... (za konačnu listu takvih riječi javiti se na sbeliga@uniri.hr)

Primjer: za rečenice „Danas je sunčan dan. Danas je lijep sunčan dan.“ egde lista izgledala bi ovako:

danas je 2  
je sunčan 1  
sunčan dan 2  
je lijep 1  
lijep sunčan 1

**ANALIZA MREŽE.** Konstruiranoj mreži potrebno je analizirati na lokalnoj razini. Za konstruiranu mrežu svakom čvoru je potrebno izračunati određene mjere centralnosti te rangirati čvorove prema dobivenim vrijednostima centralnosti. Analizirati degree centrality, Katz centrality i selectivity. Čvorove iz mreže je potrebno rangirati prema dobivenim vrijednostima pojedine centralnosti te za svaku centralnost izdvojiti po 5 ključnih riječi koje su postigle najbolju vrijednost. Dakle za svaki tekst odvojiti zasebnih 5 ključnih riječi koje su najbolje za pojedinu centralnost. Ukupno 15 riječi za svaki tekst.

**ANALIZA REZULTATA.** Rezultate prikazati u tablicama. U stupcima prikazati vrijednosti izmjerenih rezultata (koristiti standardne mjere Precision, Recall i F1-score) za pojedinu mjeru centralnosti. Mjere se mogu izračunati u Pythonu (npr. scikit-learn biblioteka). P, R i F1 računati na način da riječi dobivene iz mreže na temelju pojedine mjere centralnosti usporedite s riječima (#tagovima) koje ste prikupili za pojedini tekst.

## **ZAD 2.4.**

**PRIKUPLJANJE PODATAKA.** Potrebno je prikupiti podatke (tekstove) koji će se. Preporuča se koristiti automatsku tehniku prikupljanja podataka s Web - engl. *Web scraping*. Analizirat će se tekstovi iz kategorije vijesti o životnom stilu. Potrebno je na web portalu 24sata.hr prikupiti najmanje 10 tekstova iz kategorije Lifestyle (adresa: <https://www.24sata.hr/lifestyle>). Mogu se preuzeti tekstovi iz bilo koje podkategorije zdravlje, moda, kuhanje, itd. Potrebno je preuzeti tekstualni dio vijesti uključujući sam tekst vijesti zajedno s naslovima i podnaslovima, ali bez slika. Za svaku vijest u zasebni dokument potrebno je preuzeti i #tagove (npr. #čišćenje stakala, #potrošač, #staklo, #automobil). Iz tagova je potrebno očistiti znak # te ostaviti samo riječi. Preuzeti smo one tekstove koji nemaju manje od 3 #taga.

**PRIPREMA PODATAKA.** Potrebno je na temelju svakog preuzetog teksta generirati usmjerenu, težinsku mrežu. Riječi iz teksta predstavljaju čvorove, a veze između čvorova postoje samo ako su riječi u tekstu susjedne. Smjer veza u mreži određen je redoslijedom čitanja rečenice, tj. riječi u

rečenici (s lijeva na desno). Težine veza u grafu predstavljene su brojem frekvencija pojavljivanja pojedinog para riječi u samome tekstu. Dodatno, iz mreže je potrebno izostaviti stop riječi kao što su veznici, čestice i sl. To su riječi koje se u tekstu javljaju često, a ne nose nikakvo značenje, npr. riječi i, ili, ni, niti, a, ali, nego, već, no, ... (za konačnu listu takvih riječi javiti se na sbeliga@uniri.hr)

Primjer: za rečenice „Danas je sunčan dan. Danas je lijep sunčan dan.“ egde lista izgledala bi ovako:

danas je 2  
je sunčan 1  
sunčan dan 2  
je lijep 1  
lijep sunčan 1

**ANALIZA MREŽE.** Konstruiranoj mreži potrebno je analizirati na lokalnoj razini. Za konstruiranu mrežu svakom čvoru je potrebno izračunati određene mjere centralnosti te rangirati čvorove prema dobivenim vrijednostima centralnosti. Analizirati degree centrality, load centrality i eccentricity. Čvorove iz mreže je potrebno rangirati prema dobivenim vrijednostima pojedine centralnosti te za svaku centralnost izdvojiti po 5 ključnih riječi koje su postigle najbolju vrijednost. Dakle za svaki tekst odvojiti zasebnih 5 ključnih riječi koje su najbolje za pojedinu centralnost. Ukupno 15 riječi za svaki tekst.

**ANALIZA REZULTATA.** Rezultate prikazati u tablicama. U stupcima prikazati vrijednosti izmjerenih rezultata (koristiti standardne mjere Precision, Recall i F1-score) za pojedinu mjeru centralnosti. Mjere se mogu izračunati u Pythonu (npr. scikit-learn biblioteka). P, R i F1 računati na način da riječi dobivene iz mreže na temelju pojedine mjere centralnosti usporedite s riječima (#tagovima) koje ste prikupili za pojedini tekst.

## **ZAD 2.5.**

**PRIKUPLJANJE PODATAKA.** Potrebno je prikupiti podatke (tekstove) koji će se. Preporuča se koristiti automatsku tehniku prikupljanja podataka s Web-a - engl. *Web scraping*. Analizirat će se tekstovi iz kategorije tehnologije i znanosti. Potrebno je na web portalu 24sata.hr prikupiti najmanje 10 tekstova iz kategorije Tech (adresa: <https://www.24sata.hr/tech>). Mogu se preuzeti tekstovi iz bilo koje podkategorije gaming, auto, znanost, itd. Potrebno je preuzeti tekstualni dio vijesti uključujući sam tekst vijesti zajedno s naslovima i podnaslovima, ali bez slika. Za svaku vijest u zasebni dokument potrebno je preuzeti i #tagove (npr. #mercedes glb, #vozili smo, #suv, #terenac). Iz tagova je potrebno očistiti znak # te ostaviti samo riječi. Preuzeti smo one tekstove koji nemaju manje od 3 #taga.

**PRIPREMA PODATAKA.** Potrebno je na temelju svakog preuzetog teksta generirati usmjerenu, težinsku mrežu. Riječi iz teksta predstavljaju čvorove, a veze između čvorova postoje samo ako su riječi u tekstu susjedne. Smjer veza u mreži određen je redoslijedom čitanja rečenice, tj. riječi u rečenici (s lijeva na desno). Težine veza u grafu predstavljene su brojem frekvencija pojavljivanja pojedinog para riječi u samome tekstu. Dodatno, mrežu je potrebno modelirati tako da se uključe samo promjenjive vrste riječi (za konačnu listu takvih riječi javiti se na sbeliga@uniri.hr).

Primjer: za rečenice „Danas je sunčan dan. Danas je lijep sunčan dan.“ egde lista izgledala bi ovako:

danas je 2  
je sunčan 1

sunčan dan 2  
je lijep 1  
lijep sunčan 1

**ANALIZA MREŽE.** Konstruiranoj mreži potrebno je analizirati na lokalnoj razini. Za konstruiranu mrežu svakom čvoru je potrebno izračunati određene mjere centralnosti te rangirati čvorove prema dobivenim vrijednostima centralnosti. Analizirati degree centrality, selectivity i current flow closeness centrality. Čvorove iz mreže je potrebno rangirati prema dobivenim vrijednostima pojedine centralnosti te za svaku centralnost izdvojiti po 5 ključnih riječi koje su postigle najbolju vrijednost. Dakle za svaki tekst odvojiti zasebni 5 ključnih riječi koje su najbolje za pojedinu centralnost. Ukupno 15 riječi za svaki tekst.

**ANALIZA REZULTATA.** Rezultate prikazati u tablicama. U stupcima prikazati vrijednosti izmjerenih rezultata (koristiti standardne mjere Precision, Recall i F1-score) za pojedinu mjeru centralnosti. Mjere se mogu izračunati u Pythonu (npr. scikit-learn biblioteka). P, R i F1 računati na način da riječi dobivene iz mreže na temelju pojedine mjere centralnosti usporedite s riječima (#tagovima) koje ste prikupili za pojedini tekst.

### **DOMENA 3 - analiza mreža interakcija životinja**

#### **ZAD 3.1.**

Odabrati pet različitih mreža interakcija insekata sa zadane web stranice:

<https://github.com/bansallab/asnr/tree/master/Networks/Insecta>

Proučiti i opisati podatke.

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi.

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).

Posebni dio zadatka: MJERE CENTRALNOSTI. Odabrati, proučiti i isprobati 7 različitih mjera centralnosti koje su dostupne u paketu NetworkX

(<https://networkx.github.io/documentation/stable/reference/algorithms/centrality.html>) a mogu se koristiti za odabrani skup podataka. Potrebno je primijeniti mjere za rangiranje čvorova u mreži, kao rezultat napraviti tablicu s popisom top 10 rangiranih čvorova. Dodatno je potrebno izračunati Spearmanov koeficijent korelacije između svake dvije mjere centralnosti i prikazati u tablici.

#### **ZAD 3.2.**

Odabrati pet različitih mreža interakcija sisavaca sa zadane web stranice:

<https://github.com/bansallab/asnr/tree/master/Networks/Mammalia>

Proučiti i opisati podatke.

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi.

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).



Posebni dio zadatka: PREDVIĐANJE BUDUĆIH LINKOVA. Proučiti i isprobati sve algoritme za predviđanje budućih veza u mreži koji su dostupni u paketu NetworkX a mogu se koristiti za odabrani skup podataka

([https://networkx.github.io/documentation/stable/reference/algorithms/link\\_prediction.html](https://networkx.github.io/documentation/stable/reference/algorithms/link_prediction.html)).

Potrebno je primijeniti i evaluirati algoritme na način da se koristi odabrana mjera točnosti (može postotak točno predviđenih veza) i 10-cross-validation metoda evaluacije.

### **ZAD 3.3.**

Odbрати pet različitih mreža interakcija ptica sa zadane web stranice:

<https://github.com/bansallab/asnr/tree/master/Networks/Aves>

Proučiti i opisati podatke.

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi.

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).

Posebni dio zadatka: ZAJEDNICE U MREŽI. Proučiti i objasniti sve algoritme za otkrivanje zajednica u mreži dostupne u NetworkXu

(<https://networkx.github.io/documentation/stable/reference/algorithms/community.html>). Prikazati različite načine otkrivanja zajednica i sličnih pravilnih struktura u mreži, te prezentirati podatke za mreže iz zadatka.

### **ZAD 3.4.**

Odbрати pet različitih mreža interakcija gmazova sa zadane web stranice:

<https://github.com/bansallab/asnr/tree/master/Networks/Reptilia>

Proučiti i opisati podatke.

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi.

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).

Posebni dio zadatka: ANALIZA JEZGRE GRAFA. Proučiti, opisati i objasniti pojam omotača grafa (k-core), te proučiti i primijeniti sve algoritme za pronalaženje omotača grafa i sličnih struktura

(<https://networkx.github.io/documentation/stable/reference/algorithms/core.html>).

### **ZAD 3.5.**

Odbрати pet različitih mreža interakcija riba sa zadane web stranice:

<https://github.com/bansallab/asnr/tree/master/Networks/Actinopterygii>

Proučiti i opisati podatke.

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi.

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).

Posebni dio zadatka: ANALIZA MOTIFA. Proučiti postupke analize motifa, te primijeniti alat acc-Motif (<https://www.ft.unicamp.br/docentes/meira/accmotifs/>) ili FANMOD (<http://theinf1.informatik.uni-jena.de/~wernicke/motifs/>) za pronalaženje motifa u odabranim mrežama.

#### **DOMENA 4 - analiza ekonomskih mreža**

##### **ZAD 4.1.**

Odbrati pet različitih ekonomskih mreža sa zadane web stranice:

<http://networkrepository.com/econ.php>

Proučiti i opisati podatke

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).

Posebni dio zadatka: MJERE CENTRALNOSTI. Odabrati, proučiti i isprobati 7 različitih mjera centralnosti koje su dostupne u paketu NetworkX

(<https://networkx.github.io/documentation/stable/reference/algorithms/centrality.html>) a mogu se koristiti za odabrani skup podataka. Potrebno je primijeniti mjere za rangiranje čvorova u mreži, kao rezultat napraviti tablicu s popisom top 10 rangiranih čvorova. Dodatno je potrebno izračunati spearmanove korelacije između svake dvije mjere centralnosti i prikazati u tablici.

##### **ZAD 4.2.**

Odbrati pet različitih ekonomskih mreža sa zadane web stranice:

<http://networkrepository.com/econ.php>

Proučiti i opisati podatke

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).

Posebni dio zadatka: PREDVIĐANJE BUDUĆIH LINKOVA. Proučiti i isprobati sve algoritme za predviđanje budućih veza u mreži koji su dostupni u paketu NetworkX a mogu se koristiti za odabrani skup podataka

([https://networkx.github.io/documentation/stable/reference/algorithms/link\\_prediction.html](https://networkx.github.io/documentation/stable/reference/algorithms/link_prediction.html)).

Potrebno je primijeniti i evaluirati algoritme na način da se koristi odabrana mjera točnosti (može postotak točno predviđenih veza) i 10-cross-validation metoda evaluacije.

##### **ZAD 4.3.**

Odbrati pet različitih ekonomskih mreža sa zadane web stranice:

<http://networkrepository.com/econ.php>

Proučiti i opisati podatke

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).

Posebni dio zadatka: ZAJEDNICE U MREŽI. Proučiti i objasniti sve algoritme za otkrivanje zajednica u mreži dostupne u NetworkXu (<https://networkx.github.io/documentation/stable/reference/algorithms/community.html>). Prikazati različite načine otkrivanja zajednica i sličnih pravilnih struktura u mreži, te prezentirati podatke za mreže iz zadatka.

#### **ZAD 4.4.**

Odbrati pet različitih ekonomskih mreža sa zadane web stranice:

<http://networkrepository.com/econ.php>

Proučiti i opisati podatke

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).

Posebni dio zadatka: ANALIZA JEZGRE GRAFA. Proučiti, opisati i objasniti pojam omotača grafa (k-core), te proučiti i primijeniti sve algoritme za pronalaženje omotača grafa i sličnih struktura (<https://networkx.github.io/documentation/stable/reference/algorithms/core.html>).

#### **ZAD 4.5.**

Odbrati pet različitih ekonomskih mreža sa zadane web stranice:

<http://networkrepository.com/econ.php>

Proučiti i opisati podatke

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).

Posebni dio zadatka: ANALIZA MOTIFA. Proučiti postupke analize motifa, te primijeniti alat acc-Motif (<https://www.ft.unicamp.br/docentes/meira/accmotifs/>) ili FANMOD (<http://theinf1.informatik.uni-jena.de/~wernicke/motifs/>) za pronalaženje motifa u odabranim mrežama.

### **DOMENA 5 - analiza infrastrukturnih mreža**

#### **ZAD 5.1.**

Odbrati pet različitih infrastrukturnih mreža sa zadane web stranice:

<http://networkrepository.com/inf.php>

Proučiti i opisati podatke

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).

Posebni dio zadatka: MJERE CENTRALNOSTI. Odabrati, proučiti i isprobati 7 različitih mjera centralnosti koje su dostupne u paketu NetworkX (<https://networkx.github.io/documentation/stable/reference/algorithms/centrality.html>) a mogu se koristiti za odabrani skup podataka. Potrebno je primijeniti mjere za rangiranje čvorova u mreži, kao rezultat napraviti tablicu s popisom top 10 rangiranih čvorova. Dodatno je potrebno izračunati Spearmanov koeficijent korelacije između svake dvije mjere centralnosti i prikazati u tablici.

#### **ZAD 5.2.**

Odobri pet različitih infrastrukturnih mreža sa zadane web stranice:

<http://networkrepository.com/inf.php>

Proučiti i opisati podatke.

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi.

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).

Posebni dio zadatka: PREDVIĐANJE BUDUĆIH LINKOVA. Proučiti i isprobati sve algoritme za predviđanje budućih veza u mreži koji su dostupni u paketu NetworkX a mogu se koristiti za odabrani skup podataka

([https://networkx.github.io/documentation/stable/reference/algorithms/link\\_prediction.html](https://networkx.github.io/documentation/stable/reference/algorithms/link_prediction.html)).

Potrebno je primijeniti i evaluirati algoritme na način da se koristi odabrana mjera točnosti (može postotak točno predviđenih veza) i 10-cross-validation metoda evaluacije.

#### **ZAD 5.3.**

Odobri pet različitih infrastrukturnih mreža sa zadane web stranice:

<http://networkrepository.com/inf.php>

Proučiti i opisati podatke.

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi.

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).

Posebni dio zadatka: ZAJEDNICE U MREŽI. Proučiti i objasniti sve algoritme za otkrivanje zajednica u mreži dostupne u NetworkXu

(<https://networkx.github.io/documentation/stable/reference/algorithms/community.html>). Prikazati različite načine otkrivanja zajednica i sličnih pravilnih struktura u mreži, te prezentirati podatke za mreže iz zadatka.

#### **ZAD 5.4.**

Odobri pet različitih infrastrukturnih mreža sa zadane web stranice:

<http://networkrepository.com/inf.php>

Proučiti i opisati podatke.

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi.

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).

Posebni dio zadatka: ANALIZA JEZGRE GRAFA. Proučiti, opisati i objasniti pojam omotača grafa (k-core), te proučiti i primijeniti sve algoritme za pronalaženje omotača grafa i sličnih struktura (<https://networkx.github.io/documentation/stable/reference/algorithms/core.html>).

#### **ZAD 5.5.**

Odbрати pet različitih infrastrukturnih mreža sa zadane web stranice:

<http://networkrepository.com/inf.php>

Proučiti i opisati podatke.

Generirati mreže u odgovarajućem format za učitavanje u NetworkX i u Gephi.

Napraviti analizu na lokalnoj, središnjoj i globalnoj razini (mjere koje su se radile na predavanjima i vježbama – popis obaveznih mjera bit će objavljen na Merlinu).

Posebni dio zadatka: ANALIZA MOTIFA. Proučiti postupke analize motifa, te primijeniti alat acc-Motif (<https://www.ft.unicamp.br/docentes/meira/accmotifs/>) ili FANMOD (<http://theinf1.informatik.uni-jena.de/~wernicke/motifs/>) za pronalaženje motifa u odabranim mrežama.