



**Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva
Zavod za telekomunikacije**

**DRUŠTVENE MREŽE
2020./2021.**

Upute za izradu

Laboratorijska vježba 2

***Praktične primjene analize
društvenih mreža***

Cilj ove laboratorijske vježbe je naučiti osnove analize društvenih mreža i upoznati se s alatom za analizu i vizualizaciju mreža poput Gephija (gephi.org).

Napomena za izradu: Za određivanje traženih vrijednosti i slika potrebno je **navesti njihov postupak dobivanja te ih odgovarajuće opisati i po potrebi interpretirati.**

Zadatak 1. Analiza mreže riječi s Twittera

U ovome zadatku potrebno je analizirati mrežu riječi s Twittera. Prije analize potrebno je izraditi mrežu prema sljedećim specifikacijama:

- Koristeći svoj *Twitter account*, najprije treba dohvatiti 100 *tweetova* sa *timelinea* zadanog Twitter ID-a (svaki student ima dodijeljen jedan Twitter ID za dohvat podataka - popis je u posebnom dokumentu).
- Iz dohvaćenih *tweetova* treba izbaciti brojeve, URL-ove, rečeničnu interpunkciju, te ostale engleske, tzv. *stopwords* (članovi, prijedlozi, veznici, pomoćni glagoli i slične riječi kraće od tri znaka).
- Iz ovako "pročišćenih" *tweetova* treba izraditi mrežu/graf riječi na temelju njihova pojavljivanja u *tweetovima*, tj. riječ u *tweetu* predstavlja čvor grafa, a veza između dva čvora/riječi postoji samo ako se te dvije riječi zajedno pojavljuju u istom *tweetu*.
- Dobiveni graf može biti u .csv ili drugom *Gephi*u prihvatljivom obliku.

Za navedenu specifikaciju može se kao predložak koristiti R skripta kako je objašnjeno na predavanju za 2 laboratorijsku vježbu (7-mo predavanje) ili izraditi vlastiti program u proizvoljnom programskom jeziku koji automatizira gornji postupak.

Analiza mreže

Odredite broj čvorova i grana takve mreže.

| | |
|---------------|----------------------|
| Broj čvorova: | <input type="text"/> |
| Broj grana: | <input type="text"/> |

Prije početka analize prikažite početnu sliku mreže.

| | |
|----------------------|----------------------|
| Slika početne mreže: | <input type="text"/> |
|----------------------|----------------------|

Nakon toga odredite prosječni stupanj čvora u mreži i prikažite distribuciju stupnjeva čvorova u mreži. Komentirajte dobivenu distribuciju.

| | |
|-----------------------------|----------------------|
| Prosječni stupanj čvora: | <input type="text"/> |
| Slika distribucije čvorova: | <input type="text"/> |
| Komentar: | <input type="text"/> |

Odredite:

- **gustoću mreže** i koliko mreža mora imati grana da bi gustoća mreže bila 1 (izračunajte, potrebno je objasniti izračun, tj. postupak);
- **promjer mreže** (engl. *diameter*);
- **koeficijent modularnosti mreže** (engl. *modularity*) i navedite koliko je detektirano zajednica (engl. *communities*) u mreži, priložite sliku te komentirajte.

Početnom detekcijom zajednica postoji određeni broj zajednica koje imaju premali broj članova te je potrebno „filtrirati šum“. Filtrirajte mrežu tako da eliminirate oko 10% (procijenite postotak proizvoljno) čvorova po parametru „*Degree Range*“. Opišite dobivene vrijednosti i postupak.

| | Prije | Poslije | Postotak (Visible) |
|---------------|-------|---------|--------------------|
| Broj čvorova: | | | |
| Broj grana: | | | |

Ponovno odredite koeficijent modularnosti mreže (engl. *modularity*) i navedite koliko je detektirano zajednica (engl. *communities*) u mreži, priložite sliku te komentirajte. Također ponovno odredite i komentirajte:

- gustoću mreže;
- promjer mreže (engl. *diameter*);

Komponente:

- Za prikaz mreže definirajte jedan od željenih *layout*-a *Fruchterman Reingold* ili *Force Atlas*, te priložite dobivenu sliku.
- Odaberite karticu *Appearance->Nodes->Attribute* odabir boja te postavite parametar na *Modularity Class* za bojanje zajednica u mreži. Priložite dobivenu sliku.
- Nadalje, kliknite na *Size* ikonicu, odaberite parametar *Betweenness Centrality* te postavite parametre *min-max* u rasponu od 1-70 (možete i samostalno odrediti raspon po želji), pa priložite dobivenu sliku.
- Nakon toga odaberite karticu *Preview* te označite parametar *Show Labels* i kliknite na *Refresh*. Priložite dobivenu sliku. Također na slici označite zajednice koje možete detektirati i komentirajte dobivene rezultate.

Odaberite karticu *Data Laboratory* te odredite, usporedite, komentirajte i interpretirajte dobivene rezultate po parametrima:

- *Degree* top 5 s najvećim vrijednostima;
- *Closeness Centrality* top 5 s najvećim vrijednostima i top 5 s najmanjim vrijednostima;
- *Betweenness Centrality* top 5 s najvećim vrijednostima;
- *PageRank* top 5 s najvećim vrijednostima;
- *Eigenvector Centrality* top 5 s najvećim vrijednostima.

Kreirajte i usporedite slike mreža kada veličinu čvorova definirate korištenjem *PageRank* i *Eigenvector Centrality*, te priložite slike mreža. Odredite i komentirajte koja od analiziranih mjera najbolje određuje najvažnije čvorove u mreži i zašto.

Zadatak 2. Kreiranje sentiment analize prikupljenih poruka s Twittera

Zadatak je kroz R napraviti sentiment analizu prikupljenih Twitter poruka i na taj način dobiti relevantne procjene afiniteta Twitter poruka (numerički), te na taj način također dobiti ukupni utjecaj sentimenta za zadnjih 100 Tweetova (povući broj tweetova da aproksimativno ostane 100 Twitter poruka nakon procesiranja), te paralelno izvući najčešće riječi koje su korištene unutar pozitivnog ili negativnog sentimenta. Analiza se radi na agregatnim podacima (tj. na svih 100 poruka).

Za sentiment koristi se **nrc sentiment leksikon**.

Cilj ove vježbe upoznavanje je s kontekstualnim sentimentom i procesiranjem teksta kao i njihova primjena unutar analize društvenih mreža.

Kao rezultat ispisati:

Top 3 riječi kod pozitivnog segmenta:

| |
|--|
| |
|--|

Top 3 riječi kod negativnog segmenta:

| |
|--|
| |
|--|

Ukupan ranking sentimenta koristeći *word frequency* (svih emocija) – graf slika (objasniti ukratko rezultate)