



Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva
Zavod za telekomunikacije

DRUŠTVENE MREŽE
2021./2022.

Upute za izradu s odgovorenim pitanjima

Laboratorijska vježba 2

*Praktične primjene analize
društvenih mreža*

Nikola Kešćec

14.12.2021.

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.

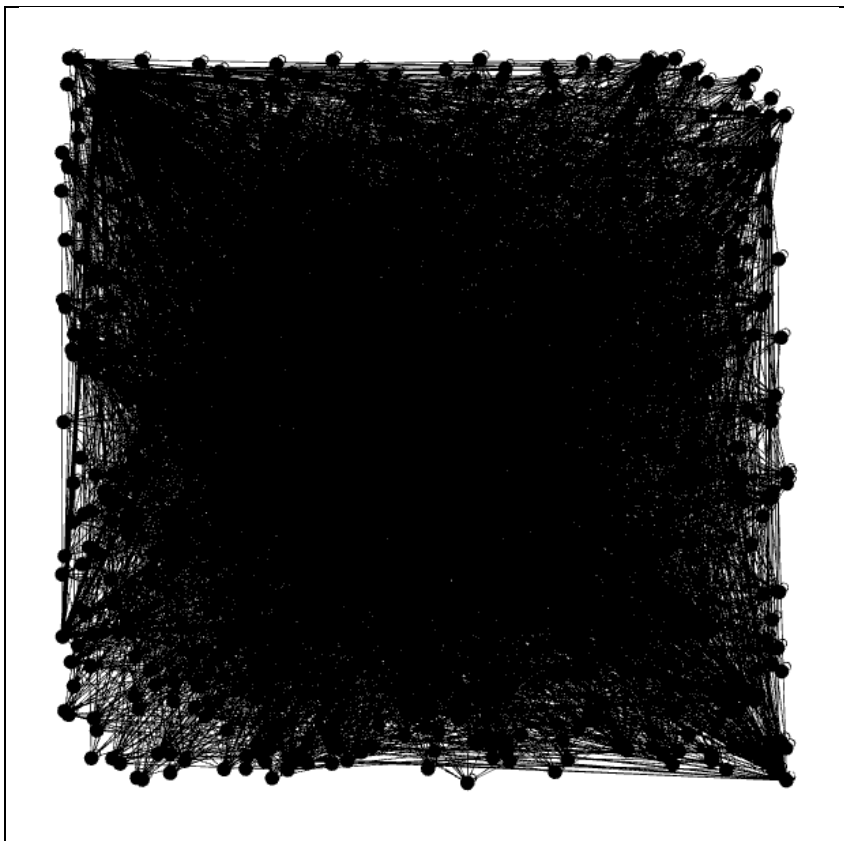
Analiza mreže

Odredite broj čvorova i grana takve mreže.

| | |
|---------------|------|
| Broj čvorova: | 924 |
| Broj grana: | 7940 |

Prije početka analize prikažite početnu sliku mreže.

Slika početne mreže:



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

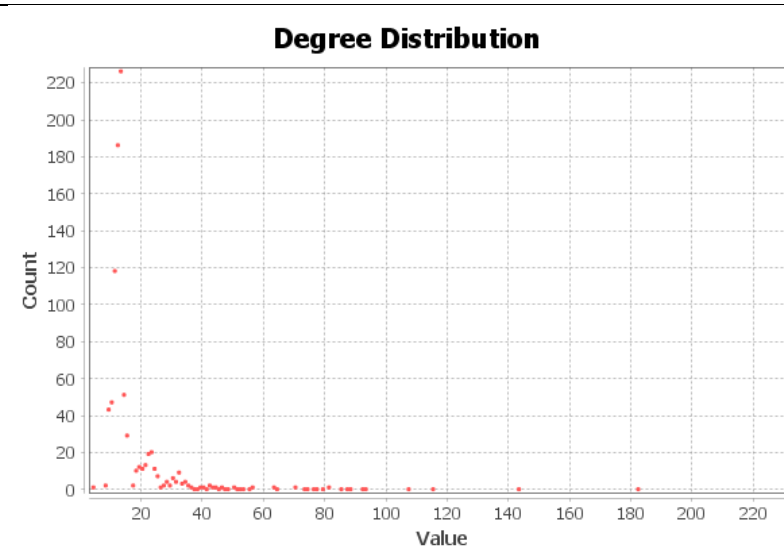
Prosječni stupanj

17.186

čvora:

Slika distribucije

čvorova:



Komentar:

U ovom grafu os *value* predstavlja **stupanj**, a os *count* **broj čvorova** koji imaju taj stupanj. Vidljivo je da distribucija stupnja čvorova grafa zoran primjer „power-law“ distribucije. Ta distribucija predstavlja da

20% prvih vrijednosti vrijednosno „pokriva“ ostalih 80%. Dakle, čvorove koji imaju veći stupanj ima znatno manje nego onih koji imaju manji stupanj.

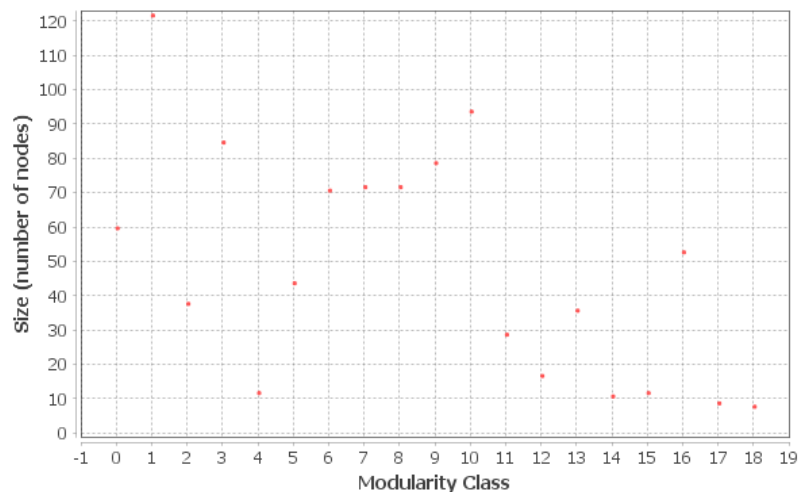
Odredite:

- **gustoću mreže** i koliko mreža mora imati grana da bi gustoća mreže bila 1 (izračunajte, potrebno je objasniti izračun, tj. postupak);
 - **ODGOVOR I POSTUPAK:**
 - gustoća mreže iznosi 0.019 za promatrani graf
 - graf koji ima gustoću mreže jedan je potpuno povezani graf, tj. svaki čvor je povezan sa svakim drugim čvorom
 - formula kojom se izračunava gustoća mreže jest broj svih bridova kroz broj svih mogućih bridova (predstavljeno matematičkom formulom $D = \frac{|E|}{\binom{|V|}{2}}$)
 - iz dane matematičke formule, moguće je izračunati da je broj potrebnih bridova jednak maksimalnom broju bridova jer D treba iznositi jedan iz čega slijedi $|E| = \binom{|V|}{2}$
 - $\binom{|V|}{2} \Rightarrow \binom{924}{2} = 426426$
- **promjer mreže** (engl. *diameter*);
 - **ODGOVOR:**
 - promjer mreže iznosi 5
- **koeficijent modularnosti mreže** (engl. *modularity*) i navedite koliko je detektirano zajednica (engl. *communities*) u mreži, priložite sliku te komentirajte.
 - **ODGOVOR:**
 - Koeficijent modularnosti mreže iznosi 0.630 te je detektirano 19 zajednica u mreži.

Results:

Modularity: 0.630
Modularity with resolution: 0.630
Number of Communities: 19

Size Distribution



- Slika predstavlja distribuciju čvorova po dobivenim zajednicama. Vidljivo je da je najveći broj čvorova u zajednici broj jedan (čak 120). Druga najveća zajednica je zajednica broj 10 s otprilike 90 članova (nije dovoljno granulirana os).

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.

POSTUPAK:

Postupak filtriranja pomoću Gephija dobiven je korištenjem mogućnosti filtriranja po rasponu stupnjeva čvorova. Filtrirao sam svoj graf tako da svi čvorovi koji imaju stupanj manji od 11 ne budu prikazani te s njima ni njihovi vrhovi. Vrijednosti dobivene tim filtriranjem navedene su u tablici.

| | Prije | Poslije | Postotak (Visible) |
|---------------|-------|---------|--------------------|
| Broj čvorova: | 924 | 827 | 89.5 |
| Broj grana: | 7940 | 7290 | 91.81 |

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:

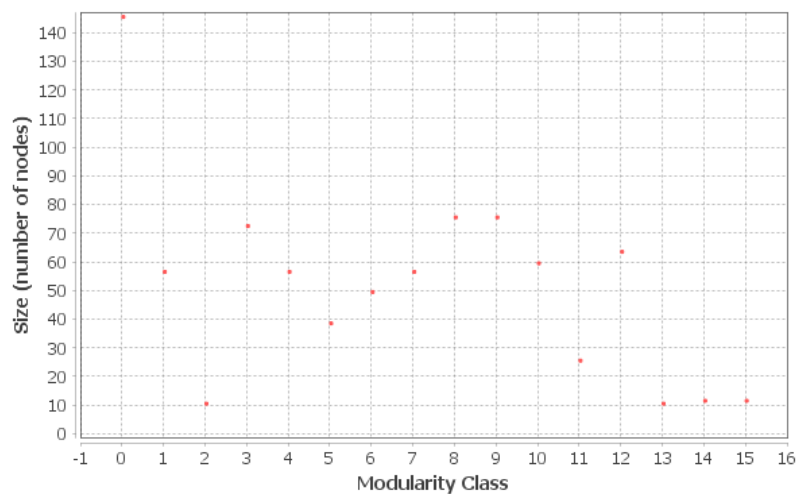
- gustoću mreže;
 - **ODGOVOR:**
 - gustoća filtrirane mreže iznosi 0.021
- promjer mreže (engl. *diameter*);
 - **ODGOVOR:**
 - promjer mreže iznosi i dalje 5

ODGOVOR I KOMENTAR KOEFICIJENTA MODULARNOSTI:

Koeficijent modularnosti nakon filtriranja šuma grafa iznosi 0.612 te je sveukupno detektirano 16 zajednica (tri manje nego prije).

Results:

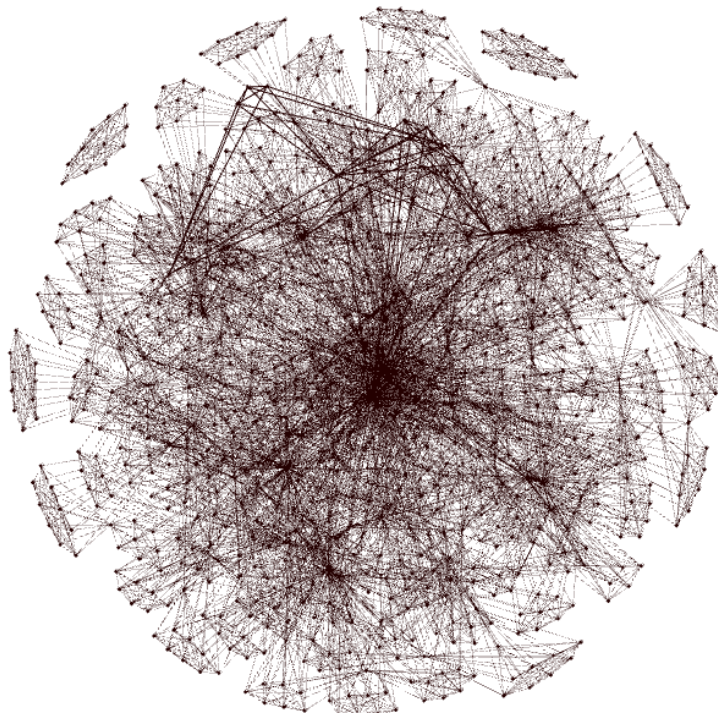
Modularity: 0.612
 Modularity with resolution: 0.612
 Number of Communities: 16

Size Distribution

Sa slike je vidljivo da je došlo do promjene distribucije čvorova po njihovim pripadajućim zajednicama. Sada je najveća zajednica broj 0 s otprilike 140 članova, a druga je zajednica broj 9.

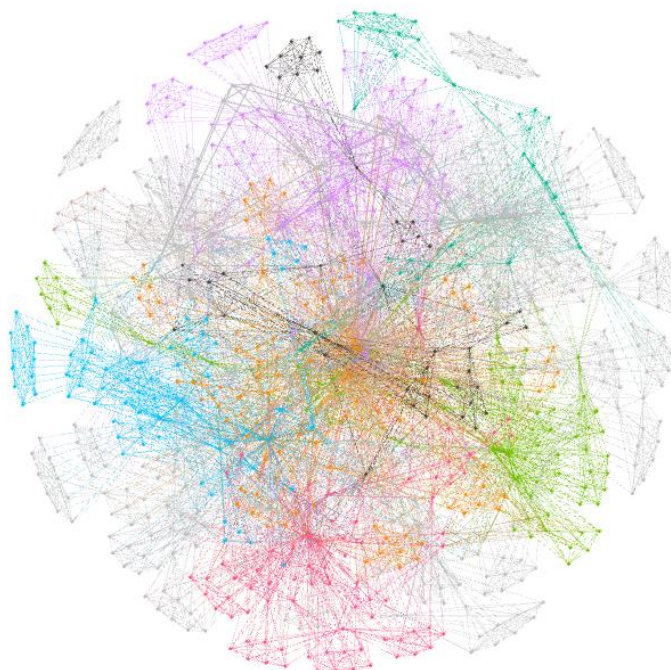
Komponente:

- Za prikaz mreže definirajte jedan od željenih *layout*-a *Fruchterman Reingold* ili *Force Atlas*, te priložite dobivenu sliku.
 - **ODGOVOR:**
 - Za prikaz grafa korišten je *Fruchterman Reingold*

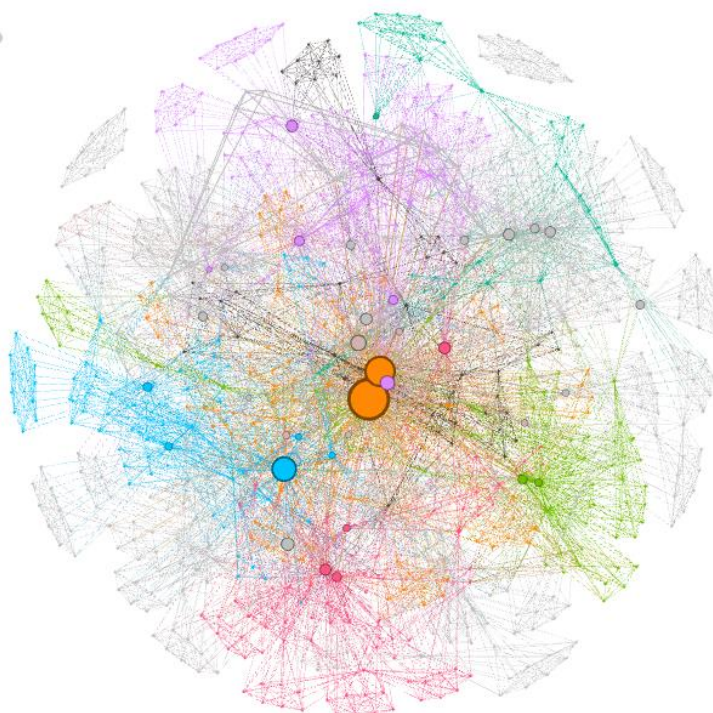


○

- Odaberite karticu *Appearance->Nodes->Attribute* odabir boja te postavite parametar na *Modularity Class* za bojanje zajednica u mreži. Priložite dobivenu sliku.
 - **ODGOVOR:**

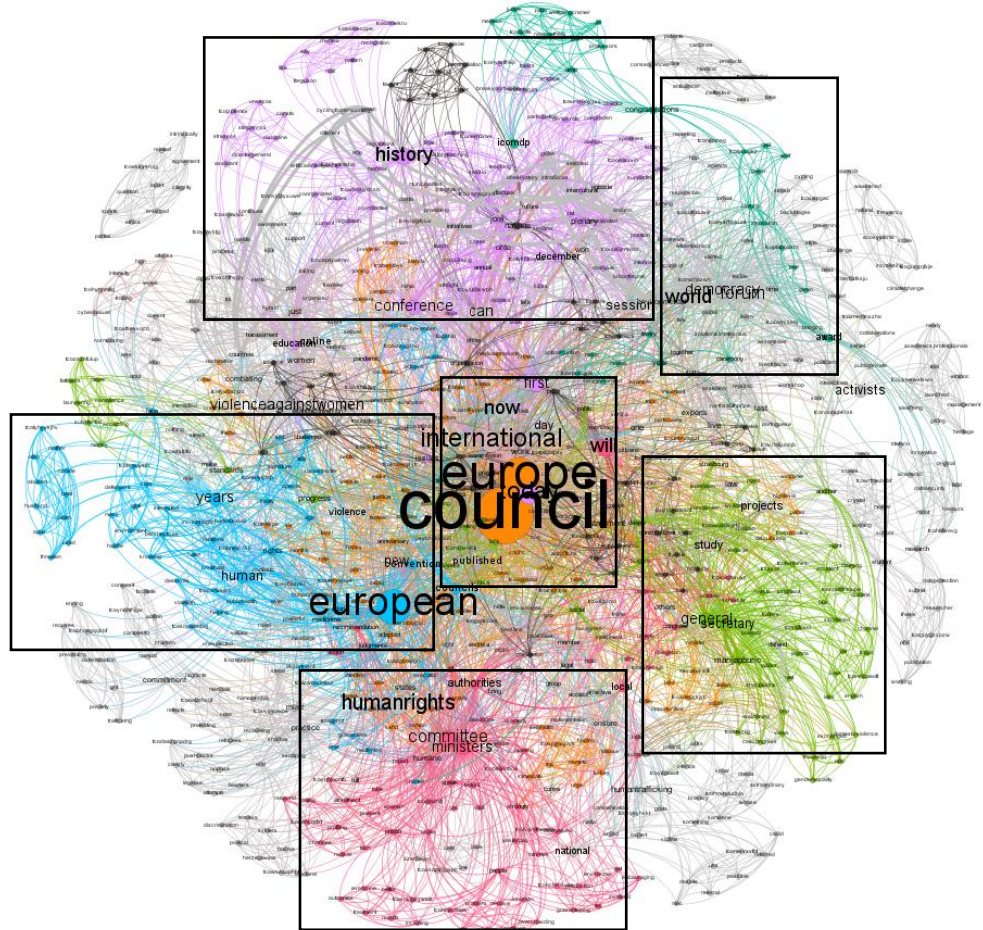


- - Nadalje, kliknite na *Size* ikonicu, odaberite parametar *Betweenness Centrality* te postavite parametre *min-max* u rasponu od 1-70 (možete i samostalno odrediti raspon po želji), pa priložite dobivenu sliku.
 - **ODGOVOR:**
 - parametri su postavljeni na raspon od 8 do 100



- 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.

○ **ODGOVOR:**



-
- Na slici su vidljive 6 dominantne zajednice. Svaka od njih pokriva jednu od grupa riječi koje se često zajedno pojavljuju u istim „tweetovima“ te kao takve mogu poslužiti kao dobar temelj za analizu skupa riječi koje se koriste u određenom kontekstu, a i potencijalno za okvirno otkrivanje tematike „tweeta“ iz same analize zajednice.

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

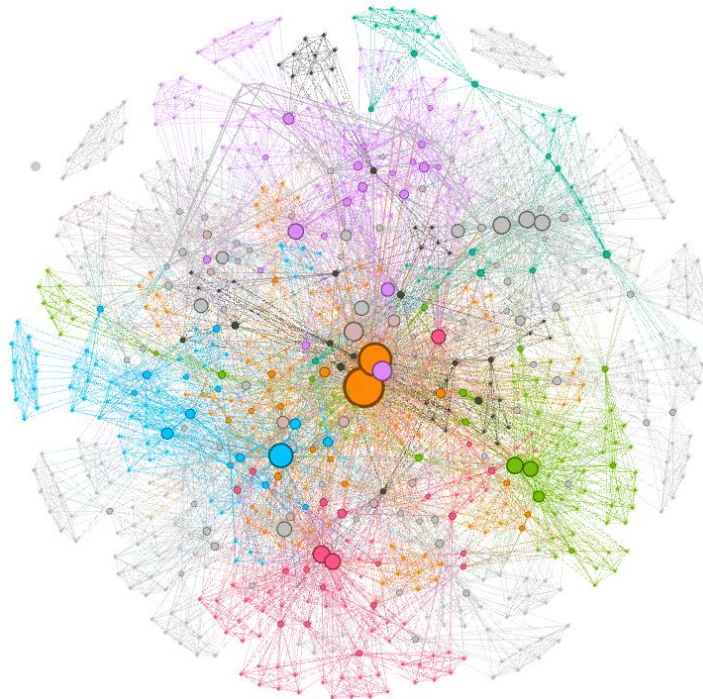
- *Degree* top 5 s najvećim vrijednostima;
 - **ODGOVOR:**
 - *Degree* kao mjera predstavlja s koliko je drugih čvorova, tj. riječi u istim tweetovima pronađena riječ
 - Top 5 riječi su:
 1. council
 2. europe
 3. european
 4. today
 5. international
 - za očekivati je da će te riječi nalaziti se u velikom broju tweetova jer dobar je dio njih i u samom nazivu promatranog profila
- *Closeness Centrality* top 5 s najvećim vrijednostima i top 5 s najmanjim vrijednostima;
 - **ODGOVOR:**
 - *Closeness Centrality (ili Ranking)* označava koliko brzo čvor može dosegnuti bilo koji drugi čvor u mreži
 - Top 5 riječi s najvišim vrijednostima su:
 1. anticancer
 2. consequences
 3. dire
 4. easy
 5. fake
 - Top 5 riječi s najmanjim vrijednostima su:
 1. academics
 2. elearning
 3. giving
 4. heritage
 5. launched
- *Betweenness Centrality* top 5 s najvećim vrijednostima;
 - **ODGOVOR:**
 - *Betweenness Centrality (ili Ranking)* označava koliko se često čvor pojavljuje na najkraćem putu u mreži
 - Top 5 riječi s najvišim vrijednostima su iste kao i kod Degree, ali djelomično drugog poretka:
 1. council
 2. europe
 3. european
 4. international
 5. today
- *PageRank* top 5 s najvećim vrijednostima;
 - **ODGOVOR:**

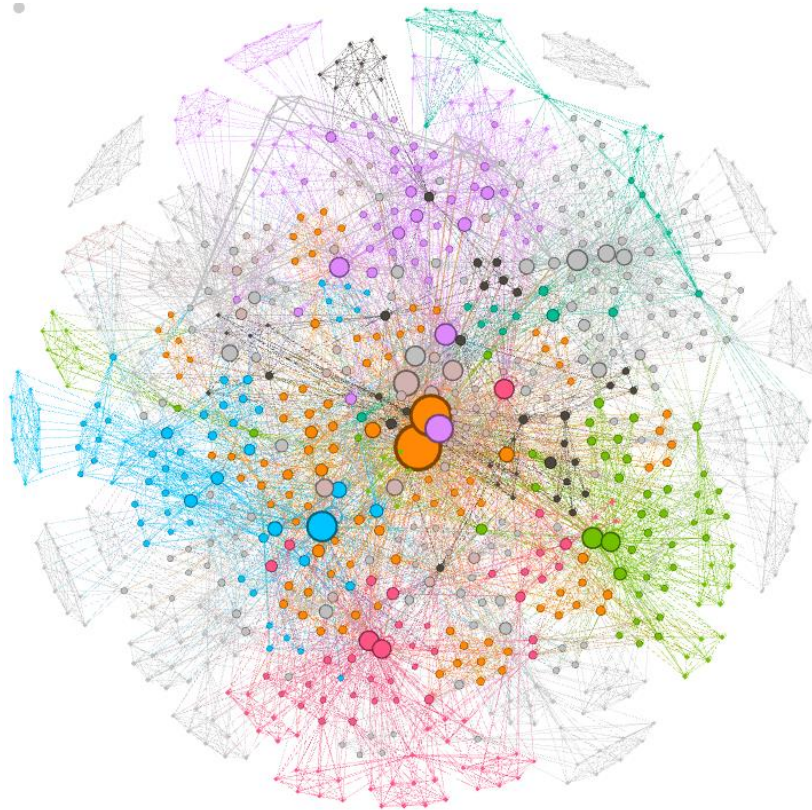
- *PageRank* je algoritam koji interpretira čvorove kao web-stranice te simulira kliktanje korisnika po poveznicama na web-stranici.
- Top 5 riječi s najvišim vrijednostima iste su kao i kod prijašnjih mjerila:
 1. council
 2. europe
 3. european
 4. today
 5. international
- *Eigenvector Centrality* top 5 s najvećim vrijednostima.
 - **ODGOVOR:**
 - *Eigenvector Centrality* (ili *Ranking*) određuje važnost čvora u mreži na temelju konekcija između čvorova (sličan *PageRank* algoritmu)
 - Top 5 riječi s najvišim vrijednostima su iste kao i svih kod ostalih mjerila:
 1. council
 2. europe
 3. european
 4. today
 5. international

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.

ODGOVOR I USPOREDBA:

PageRank:



Eigenvector Centrality:**Usporedba:**

Kao što je vidljivo sa slika, veličina čvorova s Eigenvector Centrality i PageRank su slične, ali ne iste. PageRank prikazuje malo manje čvorova od Eigenvector Centrality mjere.

Najbolje određuje prikazivanje čvorova PageRank mjera jer ona prikazuje manje čvorova, što smanjuje informacijski šum te daje bolji fokus na zaista relevantne riječi. Samo značenje slično mu je kao i kod Eigenvector Centrality koji određuje važnost čvora u mreži na temelju konekcija između čvorova, dok PageRank gleda na čvorove kao web-stranice te simuliranje kliktnje na njih poput poveznica.

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**.

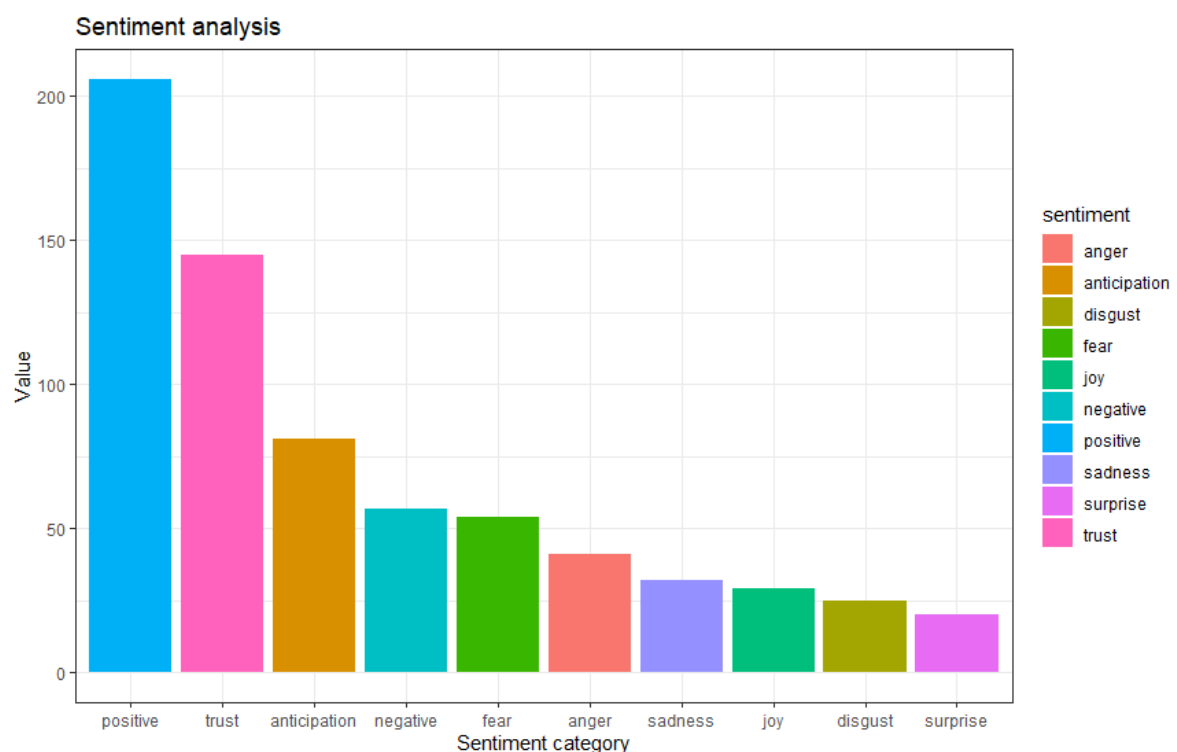
Cilj ove vježbe upoznavanje je s kontekstualnim sentimentom i procesiranjem teksta kao i njihova primjena unutar analize društvenih mreža.

Kao rezultat ispisati:

| | |
|---------------------------------------|--------------------------------------|
| Top 3 riječi kod pozitivnog segmenta: | 1. council, 2. democracy, 3. general |
| Top 3 riječi kod negativnog segmenta: | 1. violence, 2. abuse, 3. pandemic |

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

ODGOVOR:



Rezultati pokazuju da je velikim dijelom skup riječi koje se koriste unutar promatranih „tweetova“ pozitivnog sadržaja. To se poklapa s ukupnim dojmom promatranog profila koji promiče razvoj, slobodu i ravnopravnost, što su svakako pozitivne stvari. Osim što korištene riječi su povezane s pozitivnim, uvelike probuđuju i sentiment povjerenja. Gledajući analizu provedenu u prvom zadatku, često se koriste riječi povezane uz pravdu, zakon, sastanke, vijeća i dogovore, što sve budi prirodno budi povjerenje. Riječi povezane uz negativne sentimente, sentimente straha i ljutnje su riječi koje se koriste kod promocije borbe protiv nasilja, ali i političkih tenzija.