

Društvene mreže ak. god. 2021./22. – 2. laboratorijska vježba

Mreža riječi izrađena je pomoću R skripte napravljene modifikacijom skripte *01_Twitter_core_addon – sent.R* s predavanja za 2. laboratorijsku vježbu. Skripta dohvaća 100 *tweetova* korisnika @AP i stvara graf u .csv obliku. *Tweetovi* su dohvaćeni dana 27.12.2021.

ZADATAK 1

Analiza mreže

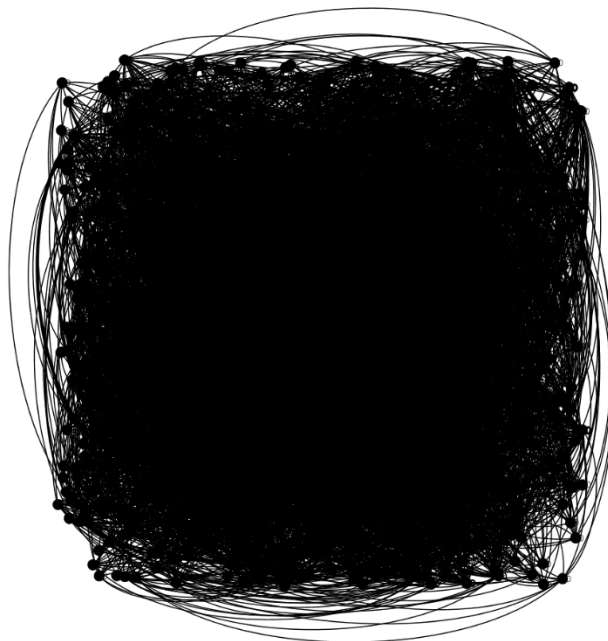
U kartici *Context* prikazane su informacije o čvorovima, granama i usmjerenosti. Javljaju se vrijednosti:

Broj čvorova: 798

Broj grana: 6365

Čvorovi predstavljaju riječi u objavama. Dva čvora su povezana granama ako se riječi nalaze u istoj objavi. U 100 dohvaćenih objava korisnika @AP koristi se 798 različitih riječi.

Pomoću kartice Graph vidljiva je vizualizacija grafa. Početni izgled mreže prikazan je slikom 1.

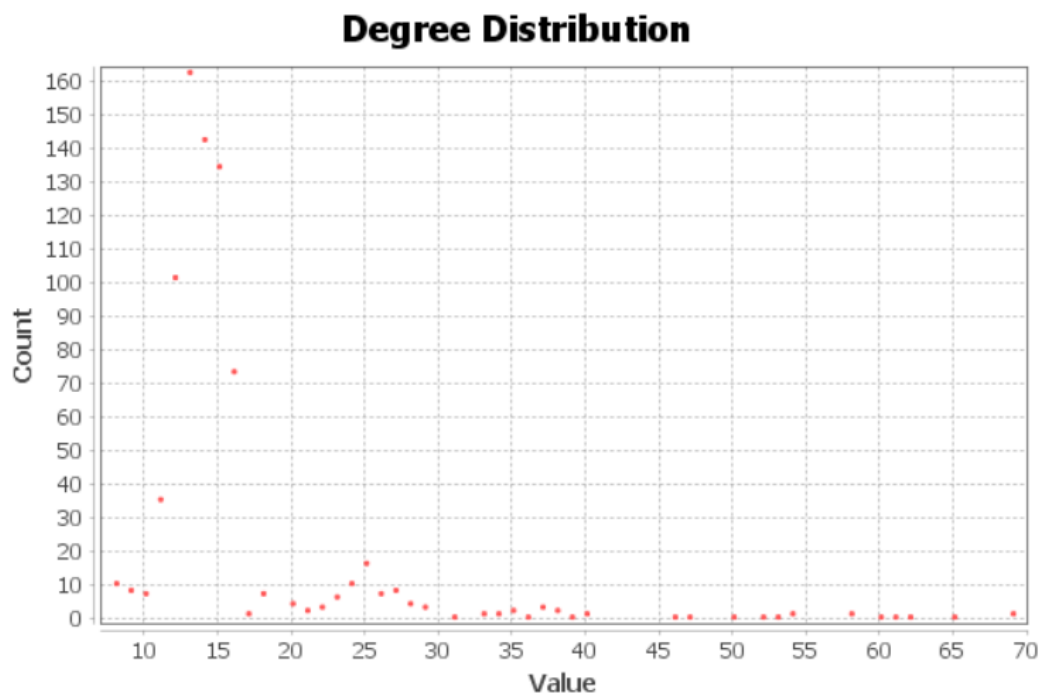


Slika 1. Slika početne mreže

U kartici *Statistics* nudi se izračun prosječnog stupnja čvora te prikaz distribucije čvorova (*Average Degree*). Prosječni stupanj čvora je 15.952. Distribucija čvorova u mreži prikazana je na slici 2.

Results:

Average Degree: 15,952



Slika 2. Distribucija stupnja čvorova u mreži

U dobivenoj distribuciji vidljivo je da se javljaju mnoge vrijednosti stupnja čvora, od stupnjeva manjih od 10 do stupnjeva vrijednosti gotovo 70. Čvorovi u najvećem broju imaju stupanj između 10 i 20. Za čvor s velikim stupnjem može se zaključiti da se riječ često pojavljuje u *tweetu* ili se koristi u dugačkim *tweetovima* jer je povezana s mnogo različitih čvorova. Analogno, riječ s malim stupnjem koristi se rijetko ili se koristi često, ali svaki put s istim riječima.

Gustoća mreže je mjera koja pokazuje koliko je mreža cjelovita. U cjelovitoj mreži svaki čvor je povezan sa svakim te je gustoća jednaka 1. U kartici *Statistics* nudi se računanje gustoće mreže (*Graph Density*) i dobiveni rezultat je 0.02.

Broj grana da bi gustoća mreže D bila 1 (za neusmjereni graf G , broj grana $|E|$, broj čvorova $|V|$):

$$D = \frac{2 * |E|}{|V| * (|V| - 1)}$$

$$1 = \frac{2 * |E|}{798 * (798 - 1)} = \frac{2 * |E|}{798 * 797}$$

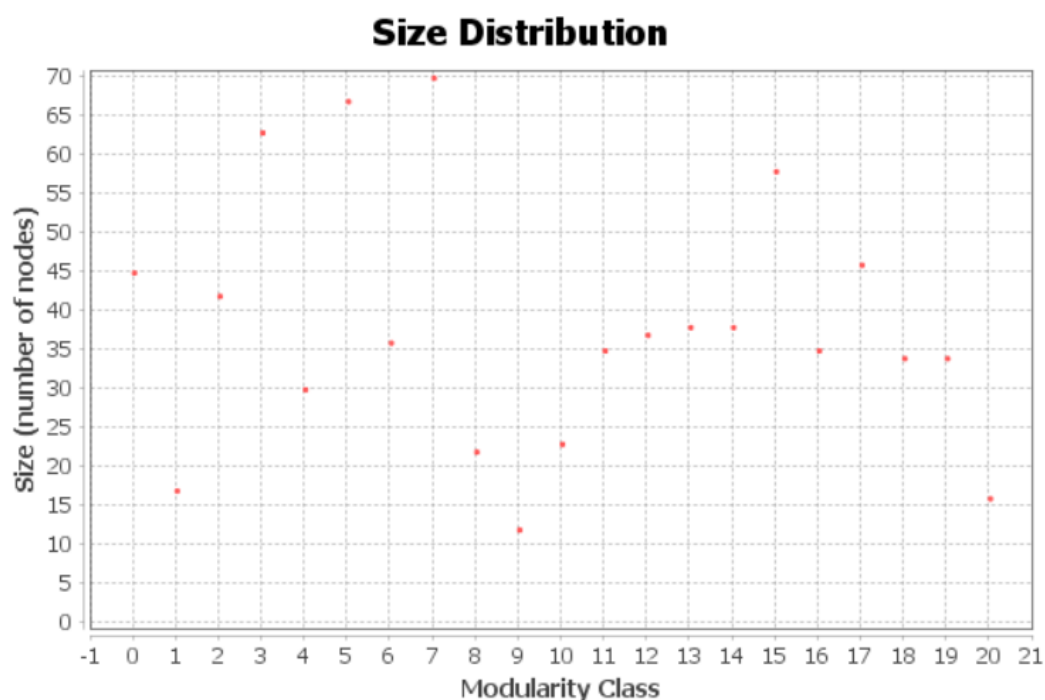
$$|E| = \frac{798 * 797}{2} = 318003$$

Promjer mreže prikazuje najveću udaljenost između bilo koja dva čvora u mreži. Povezani čvorovi imaju udaljenost 1. U kartici *Statistics* odabran je izračun promjera mreže (*Network Diameter*) i dobiven je promjer 7.

Koeficijent modularnosti mreže detektira zajednice u mreži. U kartici *Statistics* odabran je izračun modularnosti (*Modularity*). Koeficijent modularnosti mreže je 0.797. Detektirana je 21 zajednica. Distribucija veličine zajednica prikazana je na slici 3.

Results:

Modularity: 0,797
Modularity with resolution: 0,797
Number of Communities: 21



Slika 3. Distribucija veličine zajednica u mreži

Distribucija prikazuje zajednice mnogih različitih veličina. Najmanja zajednica (*Modularity Class* = 9) se sastoji od 12 čvorova. Najveća zajednica (*Modularity Class* = 7) sastoji se od 70 čvorova.

Filtriranje mreže omogućeno je kroz karticu *Filters*, opcijom *Topology* → *Degree Range*. Kroz klizač se može promijeniti prikazani dio mreže. *Degree Range* se smanjuje s 8-69 na 12-69 kako bi se postigla eliminacija otprilike 10% čvorova (8.02%). Utjecaj filtera na broj čvorova i grana je vidljiv u tablici 1.

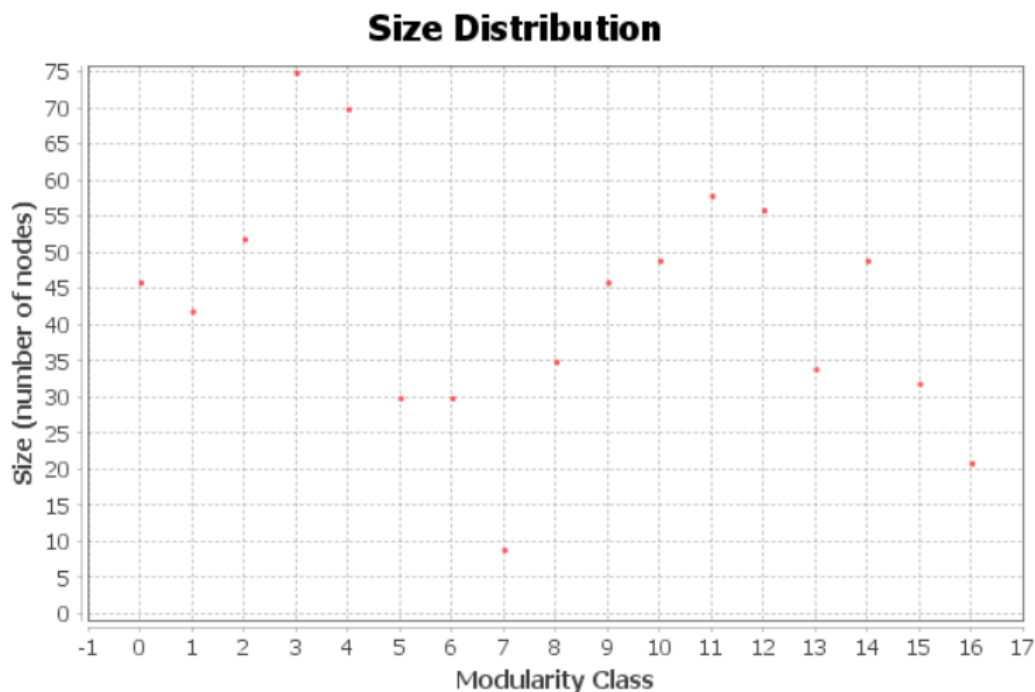
	Prije	Poslije	Postotak (Visible)
Broj čvorova	798	734	91.98%
Broj grana	6365	5962	93.67%

Tablica 1. Utjecaj filtera *Degree Range* na vidljivost mreže

Nakon filtriranja modularnost se računa na isti način i dobiva se koeficijent modularnosti 0.795. Broj zajednica je 17. Distribucija veličine zajednica prikazana je na slici 4.

Results:

Modularity: 0,795
Modularity with resolution: 0,795
Number of Communities: 17



Slika 4. Distribucija veličine zajednica u filtriranoj mreži

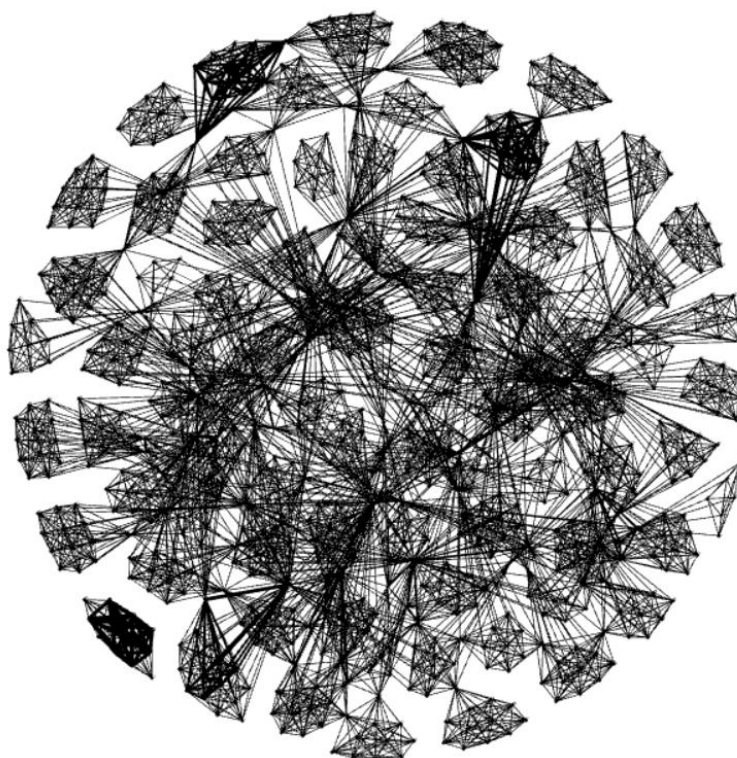
Nakon filtriranja broj zajednica se smanjio s 21 na 17. Čvorovi su sada drukčije raspoređeni u zajednice. Najmanja zajednica (*Modularity Class* = 7) ima 8 čvorova, a najveća (*Modularity Class* = 3) 75. Minimalni broj čvorova u zajednici smanjio se, a maksimalni povećao.

Nova gustoća mreže je iznosa 0.022. Gustoća se povećala za 0.002. U novom izgledu grafa čvorovi su povezaniji nego prije pa je gustoća veća. Maknuti čvorovi nisu bili gusto povezani s drugim čvorovima.

Promjer mreže ostao je jednak, duljine 7. Uklanjanje 10% čvorova nije imalo utjecaj na promjer.

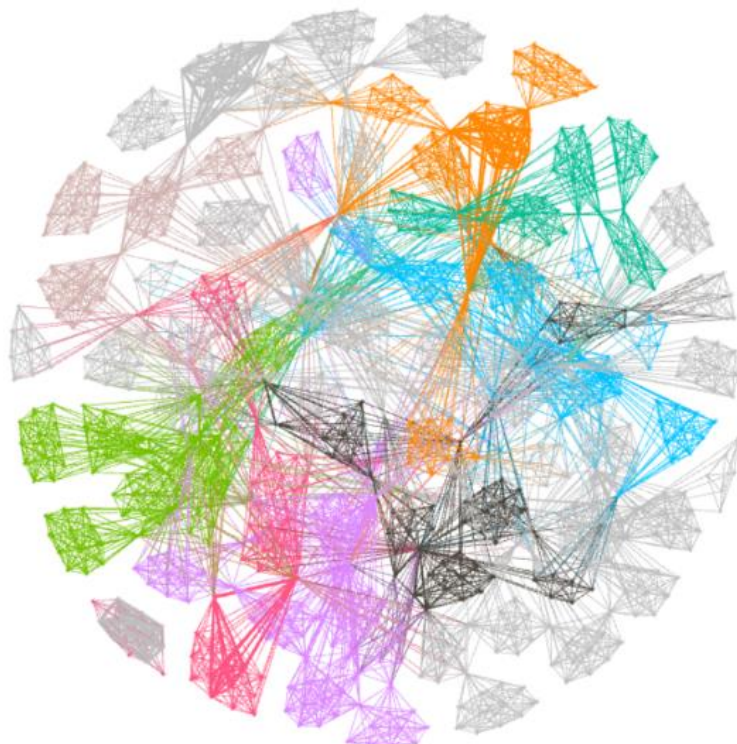
Komponente

- Prikaz grafa se može promijeniti u kartici *Layout* kroz padajući izbornik. Slika 5 prikazuje mrežu u *layout-u Fruchterman Reingold*.



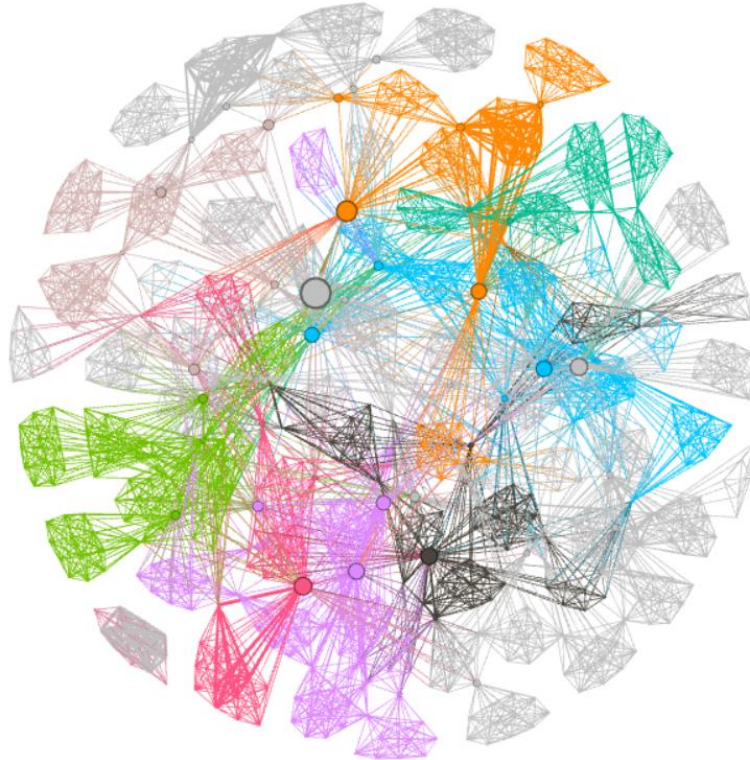
Slika 5. Fruchterman Reingold prikaz mreže

- Odabir boja po zajednicama moguće je u kartici *Appearance* → *Nodes* → *Partition* i odabirom *Modularity Class* u padajućem izborniku. Rezultat je prikazan slikom 6.



Slika 6. Mreža u kojoj su zajednice obojane različitim bojama

- Različita veličina čvorova moguća je kroz izbornik u kartici *Appearance* (*Appearance* → *Nodes* → *Size* → *Ranking*). Odabire se računanje veličine po atributu *Betweenness Centrality* (koliko se često čvor pojavljuje u najkraćem putu u mreži) i veličina čvorova. Rezultat je prikazan slikom 7.



Slika 7. Mreža u kojoj je veličina čvora određena njegovom *Betweenness Centrality* vrijednošću

- Prikaz imena čvorova može se dobiti u kartici *Preview* označavanjem parametra *Show Labels*. Rezultat je prikazan slikom 8. Najistaknutiji čvor ima najveću labelu i predstavlja riječ *two*. Neki preostali veliki čvorovi su riječi *years*, *new*, *death*, *covid* i *christmas*. Različite zajednice su različito obojane i sadrže riječi koje se često koriste zajedno u *tweetovima* (npr. plava zajednica i povezanost *covid* s riječima *death* i *died*).

Za Closeness Centrality top 5 najvećih vrijednosti imaju čvorovi *anniversary*, *ceremony*, *commemorate*, *engraved* i *firefighter*. Imaju vrijednost 1.0 kao i iduća 4 čvora koji zbog abecednog poretka labela nisu ušli u top 5. Top 5 najmanjih vrijednosti imaju *britishborn*, *chinese*, *historian*, *jonathan* i *longtime*. Imaju vrijednost 0.201223, kao i idućih 6 čvorova. Ove vrijednosti opisuju koliko brzo čvor može dosegnuti bilo koji drugi čvor u mreži. Maksimalan dobiven broj skokova je 1, mreža je dobro centralizirana.

Za Betweenness Centrality najviše vrijednosti imaju:

- 1.) two (0.173115)
- 2.) years (0.113482)
- 3.) first (0.096704)
- 4.) new (0.094908)
- 5.) christmas (0.090748)

Betweenness Centrality opisuje koliko često se čvor pojavljuje u najkraćem putu u mreži. 3 riječi se pojavljuju i ovdje i u top listi za *Degree*. Čvorovi s puno veza se često pojavljuju u najkraćem putu u mreži, mreža je dobro povezana.

Za PageRank najviše vrijednosti imaju

- 1.) two (0.003797)
- 2.) christmas (0.00356)
- 3.) first (0.003335)
- 4.) new (0.003305)
- 5.) says (0.003207)

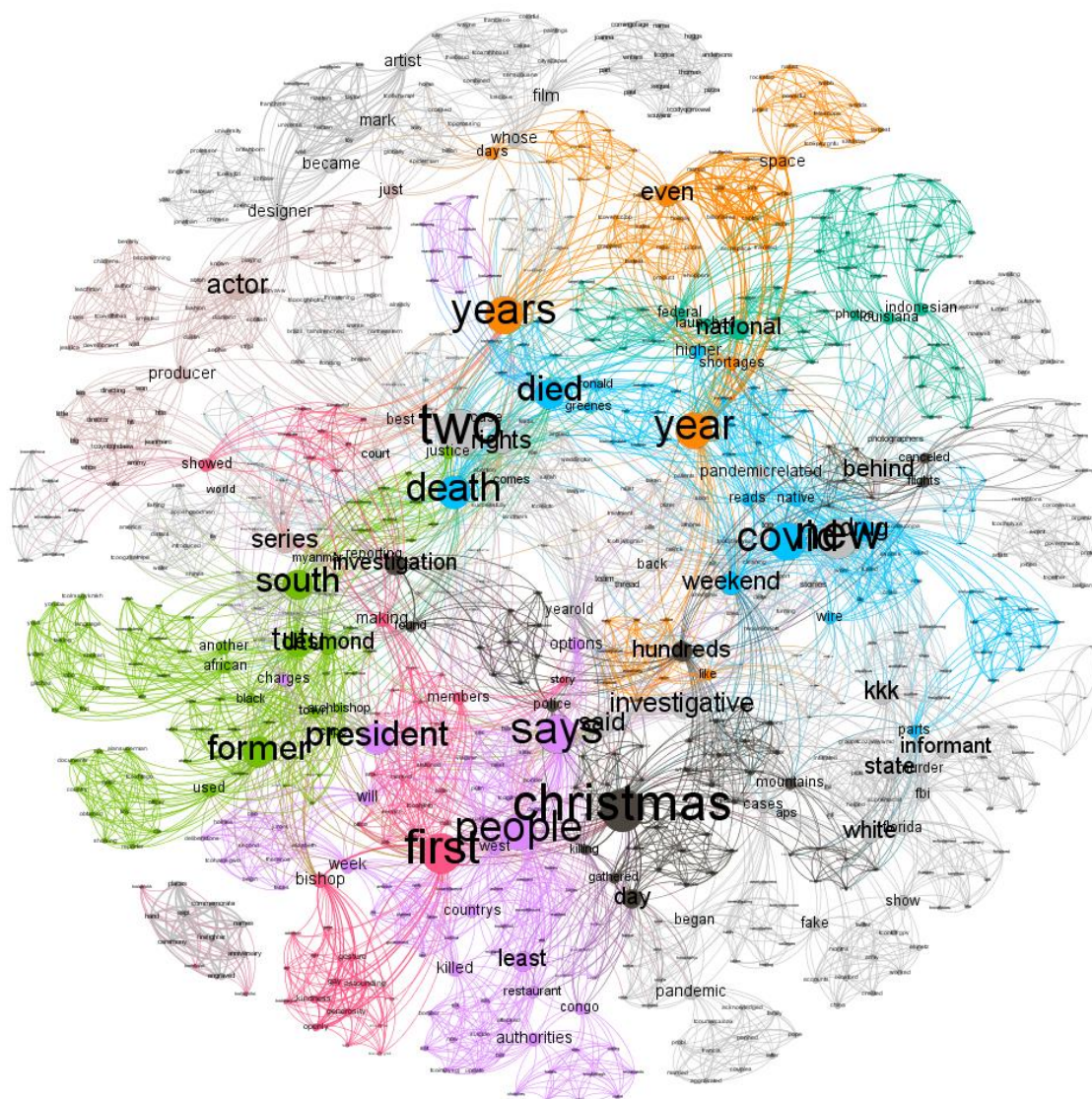
PageRank i *Eigenvector Centrality* prikazuju koliko dobro su čvorovi povezani s ostalim čvorovima koji su jako dobro povezani. U slučaju *PageRank*-a 4 riječi se poklapaju s *Degree* listom kao i *Betweenness Centrality* listom. Dobro povezani čvorovi su povezani s drugim dobro povezanim čvorovima.

Za Eigenvector Centrality najviše vrijednosti imaju

- 1.) president (1.0)
- 2.) christmas (0.954831)
- 3.) first (0.937866)
- 4.) says (0.838183)
- 5.) covid (0.822965)

Iz ove liste 2 se riječi pojavljuju na listi *Betweenness Centrality* i 4 na listi *Degree*. Kao i *PageRank* preklapanjem navedenih rezultata ukazuje na dobro povezanu mrežu.

Slike 9 i 10 prikazuju mreže gdje je veličina čvorova ovisna o vrijednostima *PageRank* i *Eigenvector Centrality*.



Slika 9. Mreža u kojoj je veličina čvora određena njegovom *PageRank* vrijednošću

ZADATAK 2

Tijekom izvršavanja *R* skripte *01_Twitter_core_addon – sent.R* stvaraju se varijable kroz koje se može proučiti sentiment *tweetova*. Za analizu sentimenta nrc leksikon koristi binarnu vrijednost sentimenta (pozitivni/negativni) kao i emocije (*anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise*, *trust*).

Varijabla *sentiment_val* sadrži riječi iz *tweetova* uz njihove sentimente i emocije, rangirane po frekvenciji pojavljivanja.

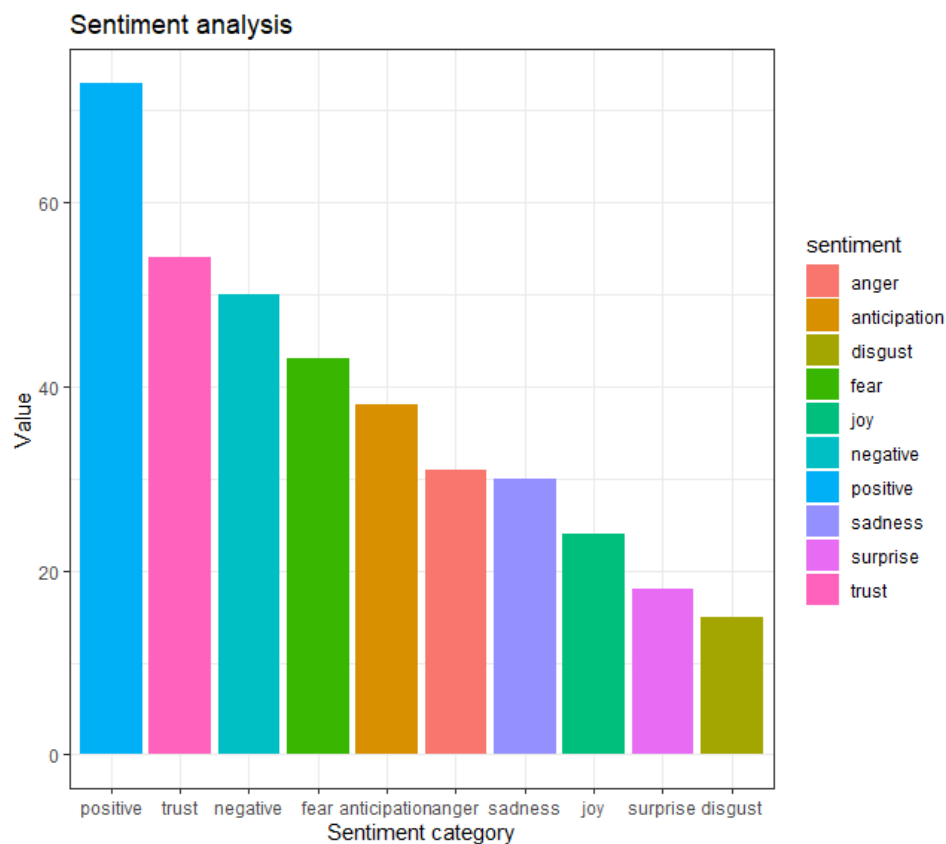
Top 3 riječi pozitivnog sentimenta:

- 1.) president
- 2.) ceremony
- 3.) commemorate

Top 3 riječi negativnog sentimenta:

- 1.) death
- 2.) case
- 3.) killing

Skripta računa i frekvenciju pojavljivanja svih sentimenta u *tweetovima*. U programu *RStudio* u kartici *Plots* moguće je vidjeti grafički prikaz izračunatih podataka. Ukupan ranking sentimenta prikazan je na slici 11.



Slika 11. Ukupan ranking sentimenta *tweetova*

Pozitivni sentiment (*positive*) pojavio se 73 puta, a negativni (*negative*) 50 puta. Od emocija najčešće se pojavljuje povjerenje (*trust*), 54 puta, a najmanje gađenje (*disgust*), 15 puta.

Uzorak podataka je malen pa se generalizirani zaključci ne mogu izvesti, ali osobno ovakve rezultate smatram očekivanim. Kako je autor *tweetova* novinarska agencija Associated Press, ton objava treba biti profesionalan. Iznenađenje i gađenje stoga nisu toliko zastupljeni, kao ni radost i tuga. Povjerenje, strah i iščekivanje češće su prihvaćene emocije u „objektivnom“ izvještavanju.