

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI SEMINAR
ANALIZA DRUŠTVENIH MREŽA

Mirjam Šitum

Mentor: Doc. dr. sc. Vedran Podobnik

Zagreb, svibanj 2013.

Sadržaj

| | |
|--|----|
| Uvod | 1 |
| 1. Važnost analize društvenih mreža | 2 |
| 2. Komponente mreže..... | 3 |
| 2.1. Čvorovi | 4 |
| 2.2. Veze | 4 |
| 2.3. Djelovi mreže | 5 |
| 2.3.1. Razina povezanosti čvorova u komponenti | 5 |
| 2.3.2. Komponente u mreži | 5 |
| 2.3.3. Odnosi čvorova u mreži..... | 5 |
| 2.4. Podaci i zapis o mreži..... | 6 |
| 2.5. Pogled na mrežu | 6 |
| 3. Mjere za mreže | 7 |
| 3.1. Razina analize..... | 7 |
| 3.2. Uzorkovanje veza | 7 |
| 3.3. Mjerenje veza | 10 |
| 3.4. Statistika u društvenim mrežama..... | 12 |
| 3.5. Položaj čvora u mreži - Centralnost | 13 |
| 4. Model mreže | 15 |
| 4.1. Mala svjetska mreža | 15 |
| 4.2. Erdos Reny model mreže..... | 15 |
| 4.3. Scale free model mreže..... | 16 |
| 4.4. Procesi u mreži | 16 |
| 4.5. Otpornost mreže | 17 |
| 5. Analiza studijskog slučaja | 19 |
| 5.1. Mreža za analizu..... | 19 |

| | |
|------------------------------------|----|
| 5.2. Mjere analizirane mreže | 24 |
| Zaključak | 27 |
| Literatura | 28 |

Uvod

Društvene mreže su jedan od najčešćih pojmova u svijetu Interneta. One virtualno okupljaju ljudе i organizacije ovisno o tome za što su specijalizirane. Tako postoje poslovne, organizacijske i druge društvene mreže. Društvenu mrežu može se definirati kao društvenu strukturu koju sačinjavaju skupine društvenih aktora (osobe ili organizacije) i kompleksni skup između aktora. Cilj ovoga rada je pokušati društvene mreže prebaciti u matematički model pogodniji za analizu i upoznati se sa mogućnostima koje takav pristup omogućava. Analiza društvenih mreža temelji se na pretpostavci o važnosti veza između djelova koji komuniciraju.

U prvom poglavlju navedena je definicija i važnost analize društvenih mreža.

Drugo poglavlje baviti će se matematičkim terminima iz teorije grafova koji će predstavljati podlogu za obradu mreže kao grafa.

Treće poglavlje bavit će se mjerama za mreže. Naglasak će biti na razinama iz kojeg se gleda graf, način dobivanja grafa uzorkovanjem veza te mjerama grafa koje se odnose na veze i kada je najbolje koristiti koju. Bit će riječi i o ulozi koju statistika ima u analizi društvenih mreža.

U četvrtom poglavlju nabrojeni su opisani najvažniji modeli mreže id an osvrt na otpornost mreže.

U petom poglavlju analizira se mreža prijatelja sa društvene mreže Facebook u programskom alatu Gephi.

1. Važnost analize društvenih mreža

Pitanje koje se prirodno postavlja je zašto je uopće važna analiza društvenih mreža. Općenito, mreže se analiziraju prema teoriji grafova. To znači da se sama mreža analizira kao graf. Najčešće analizirane mreže su financijske, mreže električne energije, političke mreže, mreže organizacija, prometne mreže itd. Prebacivanje mreže u graf ne mora nužno značiti i njeno pojednostavljenje, ali iz njega se dobivaju važne mogućnosti za obradu. Ovakav pogled omogućava uočavanje veza koje prije nisu mogle biti uočene, pruža vizualizaciju, daje uvid u strukturu, omogućava mjerjenje, ispitivanje povezanosti i optimizaciju. Može se također promatrati kako struktura utječe na procese u njoj te što je bitno za društvene mreže, mogu se uspoređivati različiti parametri ovisno o povezanosti čvorova. Društvene mreže, čak ni one napravljene primarno za zabavu, obuhvaćaju puno veći spektar aktivnosti kao što je marketing, reklamiranje, širenje poslova i slično. Za takve aktivnosti izuzetno je važna analiza zbog izraženog ekonomskog faktora. Teorija grafova omogućava predikciju, organizaciju, drugačiji pogled na sigurnost mreže ili njenih djelova, prati širenje korisnih informacija, ali i štetnih komponenata kao što su virusi. Sve ovo omogućava, ovisno o interesima, odluku o pristupu i primjeni strategije na društvenu mrežu, ali i mreže općenito. Činjenice na kojima se temelji analiza:

- aktori i njihove akcije su međusobno ovisne
- linkovi između aktora su kanali između koji se prenose resursi (materijalni i nematerijalni)
- modeli mreža se fokusiraju na davanje pogleda individualcu na strukturalni pogled na mrežu kao okolinu koja mu daje mogućnosti za djelovanje ali i ukazuje na moguća ograničenja

2. Komponente mreže

Kao što je već spomenuto prikaz i analiza društvenih mreža se oslanjaju na teoriju grafova.

Tri osnovna cilja prilikom analize društvenih mreža su [2]:

1. karakterizirati strukturu mreže :

- jesu li čvorovi povezani u mreži?
- koliko su čvorovi međusobno udaljeni?
- jesu li neki čvorovi važniji zbog svoje pozicije u mreži?
- sastoji li se mreža od manjih zajednica?

2. formiranje modela mreže :

- nasumično generirane mreže,
- povezivanje novih čvorova s povlaštenim čvorovima (engl. Preferential attachment),
- male svjetske mreže,
- optimizacija formacija strateških mreža.

3. razumijevanje kako struktura utječe na procese :

- difuzija informacija,
- formiranje mišljenja,
- koordinacija/kooperativnost,
- otpornost na ispadne.

2.1. Čvorovi

Čvorovi predstavljaju entitete između kojih promatramo veze. U slučaju analize društvenih mreža ti entiteti će biti osobe koje su na neki način povezane s drugim osobama i kao takve dio društvene mreže. Stupanj čvora (engl. *degree*) označava koliko je veza incidentno s određenim čvorom.

Čvorove možemo promatrati na dva načina[2]:

- Iz izravne povezanosti
 - ulazni stupanj čvora (engl. *indegree*) - označava broj ulaznih veza u uzeti čvor
 - izlazni stupanj čvora (engl. *outdegree*) – označava izlazni broj veza koji izlazi iz uzetog čvora
- Iz povezanosti s ostatom grafa (bliskost s ostalim čvorovima u grafu) – vise u poglavljju centralnost.

Stupanj, ulazni stupanj i izlazni stupanj čvora mogu se iščitati iz matrice susjedstva. Ako su poznate sve ove vrijednosti može se kreirati niz stupnjeva grafa, odnosno mreže.

2.2. Veze

Veze dijelimo prema usmjerenosti na [2] :

- usmjerene veze (engl. *directed*) – veza postoji od čvora A do čvora B, ali ne i obrnuto
- neusmjerene veze (engl. *undirected*) – veza postoji između oba dva čvora

Veze se mogu opisati sljedećim atributima :

- težina (engl. *weight*) – npr. učestalost komunikacije. Težine mogu biti pozitivne i negativne.
- rang (engl. *rank*) – npr. rangiranje prijatelja (najbolji prijatelj, drugi najbolji prijatelj itd.)
- tip (engl. *type*) – npr. rođak, prijatelj, suradnik...
- svojstva koja ovise o strukturi preostalog dijela grafa – npr. betweenness

2.3. Djelovi mreže

2.3.1. Razina povezanosti čvorova u komponenti

Prema povezanosti komponente se dijele na [2]:

- jako povezane komponente (engl. *strongly connected component*) - svaki čvor unutar komponente mreže može se dohvatiti iz svakog drugog čvora unutar te komponente slijedeći usmjereni veze između čvorova mreže
- slabo povezane komponente (engl. *weakly connected component*) - svaki čvor može biti dohvaćen iz svakog drugog čvora slijedeći veze između čvorova u bilo kojem smjeru.

Ovakva podjela ima smisla samo za grafove s usmjerenim vezama. U slučaju grafa neusmjerenim vezama postoji samo jedna povezana komponenta.

2.3.2. Komponente u mreži

Komponente se dijele na[2]:

- divovska komponenta (engl. *giant component*) – najveća komponenta u grafu koja obuhvaća veći ili važniji dio grafa
- ostale komponente – djelovi grafa koji nisu obuhvaćeni divovskom komponentom. Između njih međusobno i između njih i divovske komponente nema veza ni komunikacije.

2.3.3. Odnosi čvorova u mreži

- homofilija (engl. *homophily*) – tendencija da se čvorovi (u analizi društvenih mreža osobe) spajaju sa čvorovima sličnih karakteristika (npr. ukus u glazbi, status, uvjerenja...). Homofilija vodi do formacije homogenih grupa koje se nazivaju klasteri (engl. *cluster*). Ekstremna homogenizacija grupa može dovesti do zastoja u kreaciji novih ideja, informacija i pogleda. Homofilija može biti jaka ili slaba.
- tranzitivnost (engl. *transitivity*) – je svojstvo veza. Ako postoji veza između A i B i između B i C, onda postoji tranzitivna povezanost čvorova A i C. Jake veze su češće povezane tranzitivno od slabih. Tranzitivnost i homofilija zajedno dovode do stvaranja kliksa (engl. *clique*), tj. potpuno povezanih klastera.

- Mostovi (engl. *bridges*) – su čvorovi i veze koje povezuju dvije podgrupe. Mostovi omogućavaju i olakšavaju komunikaciju između podgrupa i daju podlogu za stvaranje novih ideja, inovativnost i razmjenu informacija. Uglavnom su mostovi slabe veze, ne i nužno.

2.4. Podaci i zapis o mreži

- Ako postoji potreba za korištenjem nabranja po brojevima ili slovima u drugoj dubini, autor može koristiti sljedeće stilove: lista veza (engl. *edge list*) – sastoji se od dva stupca čiji su članovi vrhovi. Par vrhova u istom redu znači postojanje veze između njih u grafu [2].
- lista susjedstva (engl. *adjacency list*) – za svaki vrh su nabrojeni vrhovi s kojima je on u grafu povezan. Ovakav prikaz se koristi kod velikih mreža gdje je čvor povezan sa malim podskupom čvorova [2].
- matrica susjedstva (engl. *adjacency matrix*) – u ovoj matrici na mjestu $a_{i,j}$ je zapisana ili 0 ili 1. Ukoliko mjesto ima vrijednost 0 to znači da vrhovi i i j nisu povezani, a ako mjesto ima vrijednost 1 ta dva čvora u grafu su povezana. Ukoliko veze imaju atribut težine onda umjesto jedinica u matrici susjedstva navedene su težine veza između odgovarajućih čvorova [2].

2.5. Pogled na mrežu

- cijela mreža (engl. *whole network*) – ovakav pogled uzima u analizu sve postojeće čvorove i veze u mreži. U praksi se ovaj pojam odnosi na analizu dijela mreže uzimajući u obzir sve čvorove i veze u analiziranom dijelu mreže.
- ego mreža (engl. *ego network*) – ovaj pojam se odnosi na pogled na mrežu s obzirom na jedan čvor, tj. na sve čvorove s kojima je on povezan [1].

3. Mjere za mreže

3.1. Razina analize

Mikro razina – najmanja razina analize u društvenim i informacijskim znanostima. Odnosi se na pojedinca i njegovu socijalnu okolinu. Često se naziva i lokalna razina. Predmet istraživanja je mali dio mreže, pojedinac ili mala grupa pojedinaca u određenom drušvenom kontekstu. Primjeri analize na ovoj razini obuhvaćaju osobu, partnerstvo, vezu, obitelj, susjedstvo i slično. U svijetu mreža ovakva analiza se vrši nad egocentričnim mrežama .

Mezo razina – razina analize koja obuhvaća analizirane grupe između mikro i makro razine analize. Obuhvaća promatranje društva, dijela društva, organizacije ili države. Može se ciljano koristiti za otkrivanje konekcija između mikro i makro razina. U svijetu mreža ovakva analiza se koristi u specijaliziranim mrežama koje obuhvaćaju različite interesne skupine, npr. mreža LinkedIn .

Makro razina – razina na kojoj se analiziraju rezultati interakcija kao što su ekonomski i druge transakcije u velikoj populaciji. Često se naziva i analiza na globalnom nivou. Obuhvaća promatranje grupa kao što su nacije, društvo, civilizacija, internacionalne skupine, populaciju u globalu. U svijetu mreža ovakva analiza obuhvaća društvene mreže s velikim i heterogenim skupinama ljudi kao što je npr. Facebook .

3.2. Uzorkovanje veza

U mreži sudionici, odnosno čvorovi mogu biti povezani sa različitim vezama (npr. učenici u učionicama mogu se igrati ili ne igrati zajedno, mogu voljeti ili ne voljeti druge itd.). Nekad je pristup takav da će se sve vrste veza uzeti u obzir, a nekad (zbog manjih troškova, zbog relevantnosti u istraživanju itd.) će se veze uzorkovati i uzeti samo one relevantne.

Pristupi uzorkovanju mreža [1]:

- *Metode uzorkovanja cijele mreže*(engl. *Full network methods*) - ovaj pristup je temeljen na skupljanju informacija o svakom čvoru, odnosno relacijama sa svim njegovim susjedima. Ovo bi se moglo više poistovjetiti sa uzimanjem popisa svih veza nego sa uzorkovanjem. Npr. skupljaju se podaci o broju vozila koji se kreću između dva grada, podaci o toku poruka elektroničke

pošte između svih parova zaposlenika u tvrtki, informaciju o prijateljstvu između grupe ljudi itd. Zbog toga što se sakupljaju informacije o vezama između svakog para dioda, potpuni skup informacija daje kompletну sliku relacija u populaciji. U realnom svijetu istraživanja, često se izvodi ovakva vrsta uzorkovanja gdje je potrebno točno definirati i mjeriti mnogo strukturalnih koncepata u analizi mreže (npr. mjerjenje centralnosti). Nažalost, ovaj način uzorkovanja može biti veoma skup i komplikiran, čak i neostvariv. Održavanje podataka o svakom članu populacije, rangu ili težini svih veza može biti težak posao čak i za manje grupe. U mnogo slučajeva problemi nisu tako strogo postavljeni. Većina ljudi, organizacija i grupa imaju ipak ograničen broj relacija i čvrstih veza zbog limitiranih resursa, energije, vremena i kognitivnog kapaciteta te ne mogu održavati veliki broj čvrstih veza.

- *Metode "sniježne kugle"* (engl. *Snowball methods*) – u ovom pristupu počinje se s promatranjem fokusne grupe čvorova. Ispitivaju se samo neki njegovi susjedi. Zatim se nad novodobivenim čvorovima napravi ista metoda. Proces staje ili kada se više ne dobivaju novi čvorovi ili kada se odluči stati zbog vremena, resursa, nevažnosti marginalnih podataka i sličnih razloga. Snowball metoda može biti korisna posebno u slučajevima gdje se prati "posebna" populacija (najčešće brojčano mala grupa ljudi pomiješana s velikim brojem onih koji se ne žele promatrati). Poslovne mreže, društvena elita, devijantne subkulture i druge mreže mogu se učinkovito analizirati i opisati snowball metodama. Iako bi se možda činilo da je teško doći do kraja algoritma pronalaženja novih čvorova, zbog prirode grupe za koju se koristi ovaj pristup najčešće postoje ograničenja broja čvrstih veza i velika je tendencija stvaranja recipročnih veza što uvelike olakšava posao nalaženja krajnje granice grafa. Dvije su glavne slabosti, odnosno ograničenja ove metode. Prvo je to što izolirani čvorovi koji nisu povezani ne mogu biti locirani prema ovom algoritmu. Pojave izoliranih čvorova dakako mogu biti veoma važni pokazatelj karakteristika populacije za neke analize. Metoda može precijeniti povezanost čvorova. Uz to, ne garantira se pronađak ni svih povezanih važnih čvorova. Ovo ovisi o mjestu s kojeg algoritam počinje. Po ovome problemu je metoda i dobila ime (*Where does one start the snowball*

rolling?). Ako se za početak odredi pogrešni čvor, postoji mogućnost promašaja cijelog podskupa čvorova koji su povezani, ali ne preko odabranog početnog čvora. Povećanje efektivnosti ove metode se može povećati naknadnom analizom za pronalaženje optimalnog početnog čvora. Nekada naime to nije potrebno jer u velikom broju istraživanja postoji "prirodna" početna točka. Npr. u analiziranju društvene moći, uobičajeno se uzimaju za početnu točku ljudi s najvišim pozicijama iz političkih, kulturnih i ekonomskih organizacija. Iako će ovakav pristup promašiti izolirane elitiste, s velikom će točnošću uhvatiti sliku elitne mreže.

- *Ego-centric networks (with alter connections)* – alternativni pristup je započeti sa skupom fokusnih čvorova (engl. *egos*) i identificirati čvorove s kojima su povezani. Zatim se određuje koji od tih čvorova su međusobno spojeni. Ovaj se pristup koristi pri prikupljanju informacija o relacijama velike grupe čvorova, npr. ispitivanje prijateljstva između ljudi. Ovakva će mreža biti mreža s bliskim čvorovima(engl. *close – knit group*). Analize mogu proučavati kakve prilike, koristi, mogućnosti i ograničenja ima ego s obzirom na svoj položaj u mreži i način na koji su čvorovi povezani. Stvorene mreže mogu biti predložak za predikciju izgleda i karakteristika cijele mreže ali ne u tolikoj mjeri koliko to mogu dvije prethodno opisane. Ovi podaci se odnose na micro razinu pogleda na mrežu – uzorkovanje lokalnih podataka iz veće mreže. Neke osobine mreže kao što su udaljenost i centralnost ne mogu se točno ocijeniti iz ovakve mreže. Neke se pak mogu dobro procijeniti, npr. gustoća mreže, a neke se mogu izravno ili skoro izravno iščitati, npr. rasprostranjenost recipročnih veza.
- *Ego-centric networks (ego only)* – ove metode fokusiraju se na individualca, tj. jedan čvor, a ne na pogled na cijelu mrežu. Skupljaju se informacije o vezama između onih čvorova koji su spojeni sa fokusnim čvorom. Ovaj je pristupa dobar za promatranje slike "lokalnih" mreža i mreža "susjedstva" člana kojeg se promatra. Koristan je u slučajevima kada se želi razumjeti utjecaj mreže na pojedinca i ne služe kao prezentacija stanja cijele mreže.

3.3. Mjerenje veza

Kao i bilo koji drugi podaci, skupljene informacije o vezama između čvorova mogu se mjeriti prema različitim razinama mjerenja. Razine su važne jer ograničavaju broj pitanja koja se mogu postaviti pri analizi mreže. Mjere su važne jer svaka ima različita matematička svojstva, mogućnost izvođenja različitih algoritama u opisivanju uzoraka i testiranju zaključaka o vezama. Uobičajeno je razlikovati nominalne, intervalne i ordinalne razine mjerenja. Nominalne se dijele na binarne i varijacije više kategorija (engl. *multi – category variations*). Ordinalne se dijele na mjere punog ranga i grupirane ordinalne mjere [1].

- *Binarno mjerjenje relacija* (engl. *Binary measures of relations*) – najuobičajeniji pristup označavanju relacija. Pravi se razlika između veza koje ne postoje (pridružuje im se broj 0) i veza koje postoje, odnosno pristune su (označavaju se sa 1). Ako se naprimjer gleda mreža prijatelja u društvenoj mreži onda će između dva čvora koja su prijatelji veza dobiti oznaku 1, a između dva čvora koja nisu prijatelji, veza će dobiti oznaku 0. Velik dio razvoja teorije grafova u matematici i velik dio algoritama za mjerjenje svojstava čvorova i mreža su razvijani upravo za binarne podatke. Binarni podaci su toliko rasprostranjeni i korišteni u analizi mreža da je neuobičajeno naići na podatke koji su višeg stupnja od 0 i 1. Ako se to i dogodi, u većini slučajeva se viši stupnjevi tokom analize svedu na binarne. Ovo se radi tako da se izabere rez (cut point) i svi slučajevi iznad reza se svedu na 1, a oni ispod na 0. Ovakvo tretiranje podataka rezultira naravno gubitkom informacija. Prije toga se zato treba uzeti u obzir relevantnost informacija (ovisno o temi istraživanja) i pitati se je li se uopće isplati izgubiti informaciju. U praksi se pokazalo da se najčešće isplati žrtvovati informacije u svrhu pojednostavljenja analize binarnim mjerjenjem [1].
- Mjerenje veza prema kategorijama (engl. *Multiple – category nominal measures of relations*) – pri prikupljanju podataka u svrhu analize mreže mogu se postaviti kategorije kojima veza može pripadati. Na primjer, u društvenoj mreži veze između čvorova mogu biti kategorije "prijatelj", "poslovni suradnik", "rođak" ili "supružnik". Veze redom dobivaju oznake 0, 1, 2 i 3. Ovakvo mjerjenje je nominalno, odnosno kvalitativno – svaka veza između dvije osobe je svrstana po tipu i ne dobiva oznaku težine veze. Za razliku od binarnog mjerjenja (0 ili 1) ovo mjerjenje nudi više opcija za označavanje veze. Uobičajeni pristup ovoj analizi je serijska analiza prema pojedinoj kategoriji. To znači da se odvajaju slučajevi. U jednom slučaju promatra se mreža koja nastaje povezivanjem

vezama iz samo jedne kategorije, npr. iz kategorije "poslovni suradnik". Ovo također znači da postoji ograničenje – svaka dva čvora mogu biti povezana samo jednom vezom iz jedne kategorije, primjerice, čvorovi ne mogu biti povezani i sa kategorijom "prijatelj" i sa kategorijom "poslovni suradnik". Ovaj slučaj se može svesti na binarni tako da se zanemare kategorije veza te sve veze koje postoje dobivaju oznaku 1, a one koje ne postoje dobivaju oznaku 0. Drugi način je da se nekoliko relevantnih grupa poveže u jednu, npr. "obitelj" u koju će spadati kategorije "rođak" i "supružnik". Samo čvorovi koji su povezani u bilo koju od kategorija u nadkategoriji "obitelj" imati će povezanost označenu s 1, a ostali 0 bez obzira jesu li spojeni nekom kategorijom koja ne spada u određe u nadkategoriju ili nikakva veza uopće ne postoji [1].

- Mjerenje težinskom grupacijom (engl. *Grouped ordinal measures of relations*) – jedna od tradicionalnih metoda pri analizi društvenim mreža je pridjeljivanje svakom paru jednu od kategorija "liked", "disliked" ili "neutral". Rezultat je brojevna skala – najčešće 1 za pozitivan stav prema osobi, 0 za neutralan stav prema osobi i -1 za negativan stav prema osobi. Ovakva skala je pogodna za pisanje algoritama za računanje i opisivanje različitih osobina mreže (npr. strukturalni balans grafa). Skala je zapravo pridjeljivanje težine bridovima mreže. Težina mreže može se odnositi na različite stvari, grubo podijeljene u dvije grupe :
 - **učestalost interakcije** (npr. učestalost komunikacije putem elektroničke pošte, trenutnog poručivanja i slično), s mogućom vremenskom kvantizacijom (na dnevnoj, mjesečnoj, godišnjoj bazi)
 - **intenzitet povezanosti** (npr. stupanj emocionalne, društvene, poslovne ili ekonomske povezanosti i slično) [1].

Moguće je i kombinirati elemente ove dvije grupe čime se veze mogu još više pojačati ako je prisutno više elemenata iz više navedenih kategorija.

Ordinalno mjerenje sadrži više informacija od nominalnog. Rezultati mjerenja daju finije kvantiziran opis veze nego jednostavno binarno određivanje prisutnosti ili neprisutnosti mreže. Iako se ovo može učiniti kao pozitivna strana, teško je baratati ovakvom količinom i vrstom podataka. Također kao što je već prije spomenuto, većina algoritama je rađena za binarno mjerenje. Zato se ordinalni podaci često svode na binarne ili intervalne.

- Težinsko mjerenje određivanjem ranga (engl. *full – rank ordinal measures of relations*) – nekad je moguće rangirati težinu relacije u rangu od najjače do najslabije. Npr. u mreži

poznanika svaki čvor može odrediti s kojim je čvorom najpovezaniji (npr. po jačini prijateljstva, po interakciji) od 1 do n, gdje je n broj čvorova mreže, a čvorovi predstavljaju poznanike. Skala koja se dobije (engl. *full rank order scale*). Ova skala reflektira različitost u smislu intenziteta veze, ali ne i jednakost u razlici. Npr. razlika u prijateljstvu između poznanika koje je čvor rangirao na prvo i drugo mjesto ne mora biti ista kao kod onih koje je rangirao na treće i četvrto mjesto. Svaka veza ima jedinstven rang. U analizi društvenih mreža ovakvo mjerjenje nije često zastupljeno. Zato nije razvijeno ni puno algoritama i metoda prilagođeno težinskom mjerenuju određivanjem ranga te se ono najčešće svodi na intervalno ili mjerjenje veza prema kategorijama [1].

- *Intervalno mjerjenje relacija* (engl. *interval measures of relations*) – najnaprednija je razina mjerjenja koja omogućava ujednačavanje odnosa između rangiranih veza. To znači da je npr. razlika između ranga 1 i 2 ista kao i razlika između rangova 10 i 11. Ovakav pristup dopušta korištenje šireg spektra matematičkih i alata za statistiku za analizu društvenih mreža. Analizatori društvenih mreža su unaprijedili velik dio algoritama koji su primarno razvijeni za baratanje binarnim podacima i prilagodili radu s intervalnim. Kad god za to postoji mogućnost, najbolje je veze mjeriti na intervalnom nivou. Po potrebi se one onda mogu svesti na binarni nivo [1].

3.4. Statistika u društvenim mrežama

Analiza društvenih mreža više spada pod granu "matematičke" sociologije nego statističke ili kvantitativne analize, ipak, analizatori društvenih mreža koriste oba pristupa. Teško je naći točnu crtlu koja razdjeljuje ta dva područja. Matematički principi u analizi mreža tretiraju podatke kao "determinističke". To znači da su težine relacija i veza "točno", "ravnotežno" i "konačno" stanje mreže. Matematika također prepostavlja da se za opservaciju ne uzima uzorak od neke veće mreže nego se na promatrane mreže gleda kao na populaciju od interesa. Statističari gledaju na mrežu i njena mjerena kao na stohastičke i vjerojatnosne podatke. Oni smatraju dio mreže uzorkom uzetim iz veće mreže, odnosno populacije i očekivaju da će se zaključci izvučeni iz nje reproducirati i na sljedećem sličnom uzorku. Podaci iz društvene mreže lako se prikazuju u obliku polja brojeva – kao i bilo koji drugi sociološki podaci. Iz tog razloga se sve operacije koje se mogu izvoditi nad drugim vrstama podataka mogu izvoditi i nad podacima iz društvene mreže. Algoritmi iz statistike se često koriste za opis karakteristika za promatranje pojedinca i cijele mreže. Statistički algoritmi su korisni za postizanje stupnja jednakosti čvorova i za nalaženje uzoraka u

podacima iz mreže. Statistika je najvažnije u dijelu kada se postavlja pitanje pri predviđanju reprodukcije očekivanja da se pojavi uzorak koji se opisuje. Statistika daje odgovore na pitanja stabilnosti, reproducibilnosti i generalizacije rezultata koji nastaju promatranjem samo jednog uzorka. Pitanje je koliko je vjerojatno da će se i na sljedećem uzorku dobiti isti odgovor o tome što se događa u cijeloj mreži iz koje su uzeti uzorci. Isto pitanje odnosi se i na mogućnost generalizacije, odnosno proširenja teorije na druge iste ili slične mreže. Ovo pitanje nekada nije toliko bitno analitičarima društvenih mreža zbog više mogućih razloga – možda druga takva mreža ne postoji ili rezultati dobiveni za drugu mrežu nemaju nikakvu važnost. U nekim slučajevima ta generalizacija će biti od interesa, ali s razlikom u tome što se uzorak neće birati vjerojatnosnim metodama, već će uzorci biti dokumenti sa stvarnim podacima i ne postoji dobar način dobivanja uzorka vjerojatnosnim metodama. Statističke metode u mrežama se koriste i za provjeravanje točnosti hipoteza. Ako dobiveni podaci iz uzorka bitno odudaraju od pretpostavki hipoteze, hipoteza vjerojatno nije točna. Treba naravno uzeti u obzir i skokove u rezultatima koji mogu proizaći iz karakteristika odabiranog slučajnog uzorka [1].

3.5. Položaj čvora u mreži - Centralnost

Mjera centralnost služi da bi se lakše primjetili ključni i važniji čvorovi u mreži.

Centralnost (engl. *centrality*) je moguće izraziti kroz nekoliko mjera [5]:

- stupanj čvora (engl. *degree centrality*) – ulazni stupanj čvora (engl. *indegree centrality*)
 - izlazni stupanj čvora (engl. *outdegree centrality*)

Stupanj čvora u analizi društvenih mreža označava koliko čvorova izravno može dohvatiti zadani čvor. On u usmjerenoj mreži odgovara izlaznom stupnju čvora, a u neusmjerenoj mreži stupnju čvora. Većinom veći stupanj čvora označava i veću važnost ili popularnost čvora u mreži.

- blizina dva čvora (engl. *closeness centrality*) – koliko brzo čvor može dosegnuti bilo koji drugi čvor u mreži, tj. koliko u prosjeku iznosi najkraći put do svih ostalih čvorova u mreži. Ova mjera je važna za slučajeve kada se zahtjeva velika brzina

prijenosu informacije. Što je manja vrijednost, to je optimalnije za prijenos informacije.

- broj čvorova između koji se čvor nalazi (engl. *betweenness centrality*) – koliko je vjerojatno da se čvor nalazi na putu između neka dva čvora. Jednaka je broju najkraćih puteva koji prolaze kroz čvor podijeljenom sa svim najkraćim putevima u mreži. Normalizira se tako da je najveća vrijednost 1. Pokazuje koji čvorovi su više vjerojatni za prijenos komunikacije između dva druga čvora. Koristi se i kao pokazatelj gdje bi se mreža raspala, tj koji bi čvorovi bili otkinuti ako nestane dio čvorova.
- Svojstvena centralnost (engl. *eigenvector centrality*) – koliko dobro je čvor povezan za ostalim čvorovima koji su tako dobro povezani. Ova vrijednost je proporcionalna sumi svojstvenih centralnosti svih čvorova spojenih direktvo s tim čvorom.

4. Model mreže

4.1. Mala svjetska mreža

Mala svjetska mreža (engl. *small – world network*) je tip matematičkog grafa u kojemu većina čvorova nije susjedna ali gotovo svi čvorovi mogu biti dohvatljivi iz svakog drugog u malom broju koraka. Tipična distanca L između dva slučajno odabrana čvora raste proporcionalno logaritmu broja čvorova N u mreži :

$$L \sim N$$

u kontekstu društvene mreže, ovo rezultira "fenomenonom malog svijeta" – stranci su povezani zajedničkim poznanicima. Mnogi grafovi se mogu modelirati malom svjetskom mrežom. Društvene mreže koje su dio Internet mreže imaju karakteristike male svjetske mreže.

4.2. Erdos Reny model mreže

U teoriji grafova Erdos – Renyi model odnosi se na jedan od dva modela za generiranje slučajnih grafova. Nazvani su po Paulu Erdosu i Alfredu Renyiju koji su prvi predstavili ovaj model 1959. godine. Drugi model predstavio je Edgar Gilbert. Ovi se modeli mogu koristiti u probabilističkim metodama za dokazivanje postojanja grafova koji zadovoljavaju razna svojstva.

U prvom modelu $G(n, M)$, graf se izabire slučajno i jednoznačno iz kolekcije grafova koji imaju n čvorova i M bridova. Na primjer, u $G(3, 2)$ modeu svaki od 3 moguća grafa je uključen s vjerojatnošću od $1/3$.

u drugom $G(n, p)$ modelu, graf se konstruira tako da se čvorovi povezuju slučajno. Svaki brid je uključen u graf s vjerojatnosti p , neovisnom od svih ostalih bridova. Dakle, svi grafovi s n čvorova i M bridova imaju jednaku vjerojatnost poja iznosi :

$$p^M (1 - p)^{\binom{n}{2} - M}$$

Parametar p u modelu može se predočiti na način da označava težinsku funkciju. Kako se p povećava od 0 do 1 model će proporcionalno vjerojatno uključiti grafove s više bridova i sve

je manja vjerojatnost da će uključiti grafove s manje bridova. U slučaju p=0.5 model uključuje slučaj u kojem je svih toliko grafova izabrano s jednakom vjerojatnošću.

4.3. Scale free model mreže

Scale free model mreže u kojemu je vjerojatnost dodavanja čvora na već postojeći u mreži proporcionalan broju veza koje postojeći čvor već ima. Ovo rezultira mrežom koja se sastoji od malog rboja važnijih čvorova koji imaju puno veza te velikog broja manjih čvorova s nekolicinom veza. Ovakva mreža je izložena velikom riziku od ispada čvorova jer ako ispadnu čvorovi koji imaju puno veza dolazi do velikog zastoja toka komunikacije i informacija.

Scale free modeli mreža pojavljuju se u mnogim područjima znanosti i inženjerstva, uključujući topologiju stranica weba (gdje su čvorovi individualne stranice weba a veze su hiperveze, odnosno poveznice na druge stranice). [10]

4.4. Procesi u mreži

Procese u mreži dijelimo na [7]:

-difuziju(jednostavnu):

- Erdos Reny grafovi
- Grafovi bez mjerila
- Topologije malog svijeta

-kompleksnu „zarazu“ (engl. contagion)/ pragovi (engl. thresholds):

- Jednostavna „zaraza“ (prelijevanje) – svaki prijatelj inicira promatrani čvor s nekom vjerojatnošću za svaku jedinicu vremena
- Kompleksna „zaraza“ – čvor će poduzeti akciju samo ako ju poduzme određeni broj njegovih susjeda

-kolektivne akcije:

- U mreži postoje dvije ravnoteže; svi prihvataju A ili svi prihvataju B
- Dvojezični čvorovi; mogu prihvatiti i A i B, ali moraju plaćati dodatni trošak C

-inovacije:

- Topologija mreže utječe na to tko „razgovara“ s kim u mreži
- Međusobni razgovori čvorova u mreži utječu na inovacije i učenje u mreži

-rješavanje problema:

- Kauffmanov NK model
- N dimenzionalni prostor problema (N bitova, svaki može biti 0 ili 1)
- K opisuje glatkoću pristajanja (engl. fitness) ($K=0$ – nema sličnosti, K velik – glatkoća pristajanja)

4.5. Otpornost mreže

Otpornost mreže je sposobnost pružanja i održavanja prihvatljive razine usluge naspram kvarova i izazova u normalnom radu [8]. Najveći problem mreže je kvar ili uklanjanje čvorova i veza. Time ona gubi određeni kapacitet i nastaje opterećenje koje se mora raspodijeliti na ostale čvorove i veze koje su u mreži. Ukoliko svaki čvor ne može podnijeti opterećenje koje mu se nameće dolazi do greške ili ispada (engl. fail). Cilj mreže je održati prihvatljivu razinu davanja usluge čak i pri ispadima. Ispadi mogu biti uzrokovani namjernim napadima na sustav, slučajnim neispravnostima hardvera ili softvera, prirodnim nepogodama... Otpornost mreže ovisi o topologiji mreže i o reakciji mreže na greške i ispade pojedinih čvorova uslijed preopterećenja kapaciteta čvora (npr. kod električnih mreža opterećenje se preraspodjeljuje).

5. Analiza studijskog slučaja

5.1. Mreža za analizu

Mreža koja će se analizirati predstavlja mrežu prijatelja sa društvene mreže Facebook. Da bi se mreža mogla analizirati u programskom alatu Gephi (otvoreni alat za vizualizaciju i manipulaciju velikih mreža) potrebno je dobiti popis prijatelja koji će predstavljati čvorove grafa te matricu bridova u kojoj će biti povezani svi prijatelji [9]. Svaki član koji će biti čvor biti će jednoznačno određen svojim identifikacijskim brojem sa mreže Facebook, a kao labela za vizualni prikaz će služiti ime osobe. Na slici 1 prikazana je lista čvorova sa pripadajućim identifikacijskim brojevima i labelama. Na slici 2 nalazi se lista bridova (veza). Lista brdova se sastoji od para čvorova koje spaja, tipa (usmjereni veza) i identifikacijskog broja veze.

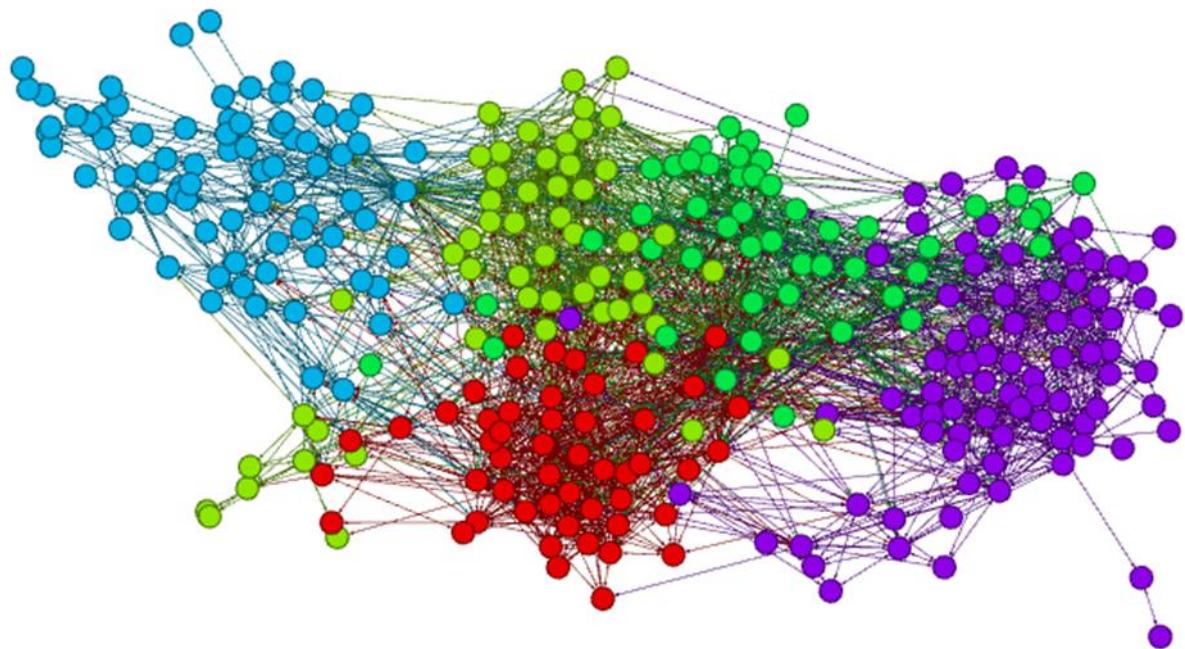
| Nodes | Id | Label |
|--------------|-----------|------------|
| ● Marijana | 525367942 | Marijana |
| ● Petra | 559218098 | Petra |
| ● Drago | 559942527 | Drago |
| ● Marija | 593591321 | Marija |
| ● Zrinka | 614717043 | Zrinka |
| ● Irène | 617224252 | Irène |
| ● Filip | 620430002 | Filip |
| ● Anja | 622148853 | Anja |
| ● Teo | 623784563 | Teo |
| ● Ana | 624673916 | Ana |
| ● Antonio | 631482175 | Antonio |
| ● Elsie | 638915856 | Elsie |
| ● Mirna | 640675849 | Mirna |
| ● Jura | 645394741 | Jura |
| ● Toni | 645981895 | Toni |
| ● Dora Petra | 646099234 | Dora Petra |
| ● Looktarn | 647871624 | Looktarn |
| ● Ivana | 681039392 | Ivana |
| ● Marino | 684571997 | Marino |
| ● Ivana | 703037713 | Ivana |

Slika 1. Dio liste čvorova

| Nodes | Edges | Configuration | Add node | Add edge | Search/Replace | Import Spreadsheet | Export table | More actions |
|----------------------|---------------------------|---------------|----------|----------|----------------|--------------------|--------------|--------------|
| Source | Target | Type | | | | | | Id |
| 525367942 - Marijana | 719410417 - Ana | Directed | | | | | | 16280 |
| 525367942 - Marijana | 757769808 - Maja | Directed | | | | | | 16281 |
| 525367942 - Marijana | 1079846669 - Mladen | Directed | | | | | | 16282 |
| 525367942 - Marijana | 1518053847 - Jasmina | Directed | | | | | | 16283 |
| 525367942 - Marijana | 100002115535245 - Mihaela | Directed | | | | | | 16284 |
| 559218098 - Petra | 593591321 - Marija | Directed | | | | | | 16285 |
| 559218098 - Petra | 617224252 - Irène | Directed | | | | | | 16286 |
| 559218098 - Petra | 646099234 - Dora Petra | Directed | | | | | | 16287 |
| 559218098 - Petra | 717994527 - Andrija | Directed | | | | | | 16288 |
| 559218098 - Petra | 719459635 - Luka | Directed | | | | | | 16289 |
| 559218098 - Petra | 724468356 - Goran | Directed | | | | | | 16290 |
| 559218098 - Petra | 743018711 - Sara | Directed | | | | | | 16291 |
| 559218098 - Petra | 840365255 - Viktor | Directed | | | | | | 16292 |
| 559218098 - Petra | 1005917977 - Josip | Directed | | | | | | 16293 |
| 559218098 - Petra | 1021987736 - Ivan | Directed | | | | | | 16294 |
| 559218098 - Petra | 1024759234 - Miranda | Directed | | | | | | 16295 |
| 559218098 - Petra | 1038324457 - Lukica | Directed | | | | | | 16296 |
| 559218098 - Petra | 1059578653 - Ivan | Directed | | | | | | 16297 |
| 559218098 - Petra | 1064822083 - Franjo | Directed | | | | | | 16298 |
| 559218098 - Petra | 1110430983 - Tina Z | Directed | | | | | | 16299 |
| 559218098 - Petra | 1116363688 - Filip Š | Directed | | | | | | 16300 |
| 559218098 - Petra | 1124091994 - Tamara | Directed | | | | | | 16301 |
| 559218098 - Petra | 1376352534 - Jasmina | Directed | | | | | | 16302 |
| 559218098 - Petra | 1467884127 - Adriano | Directed | | | | | | 16303 |
| 559218098 - Petra | 1508636590 - Goran | Directed | | | | | | 16304 |

Slika 2. Dio liste bridova (veza)

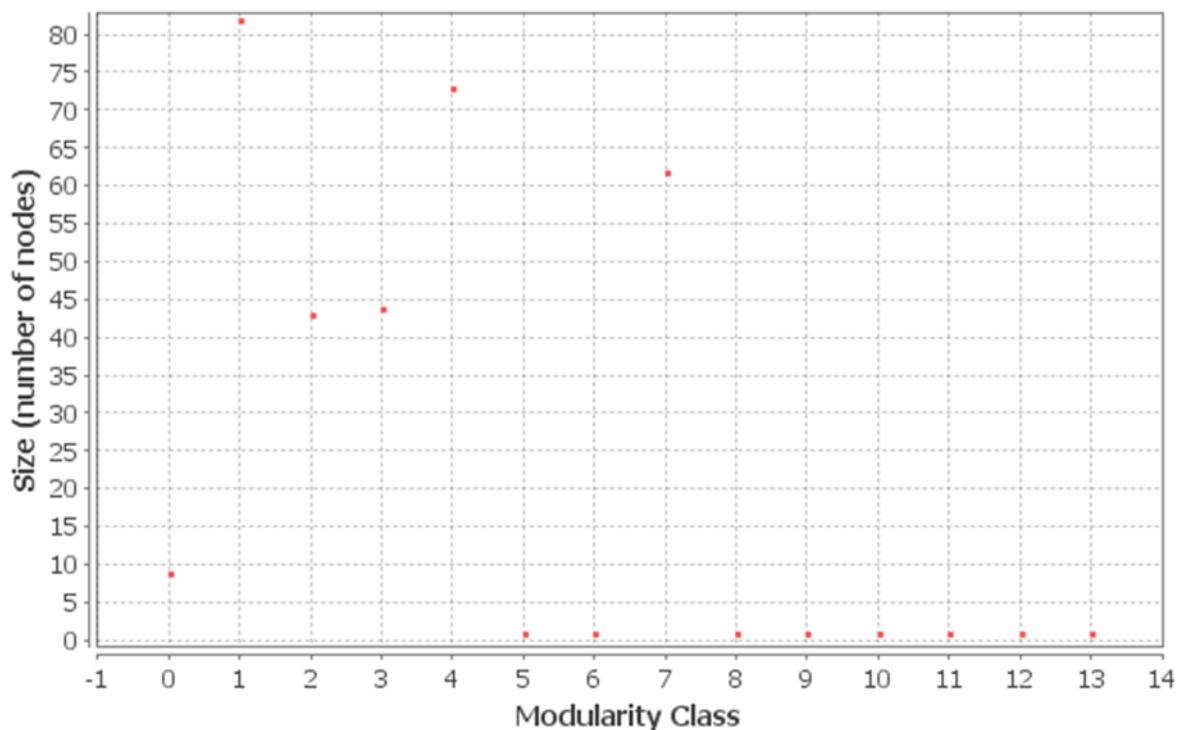
Nakon što je kreirana mreža, alat Gephi nudi mogućnost klasteriranja mreže. Na slici 3 prikazana je mreža na kojoj su različitim bojama obojani klasteri. Slika prikazuje divovsku komponentu grafa. Osim nje postoji još nekoliko čvorova koji su svaki komponenta sama za sebe. Osoba čije mreže se vrši analiza nema zajedničkih prijatelja sa ovim osobama. Ovakva mreža je egocentrična, jer se promatraju prijatelji jedne osobe. Međutim, na grafu nisu prikazane veze s tom osobom, pa je možemo promatrati i kao „cijelu“ mrežu.



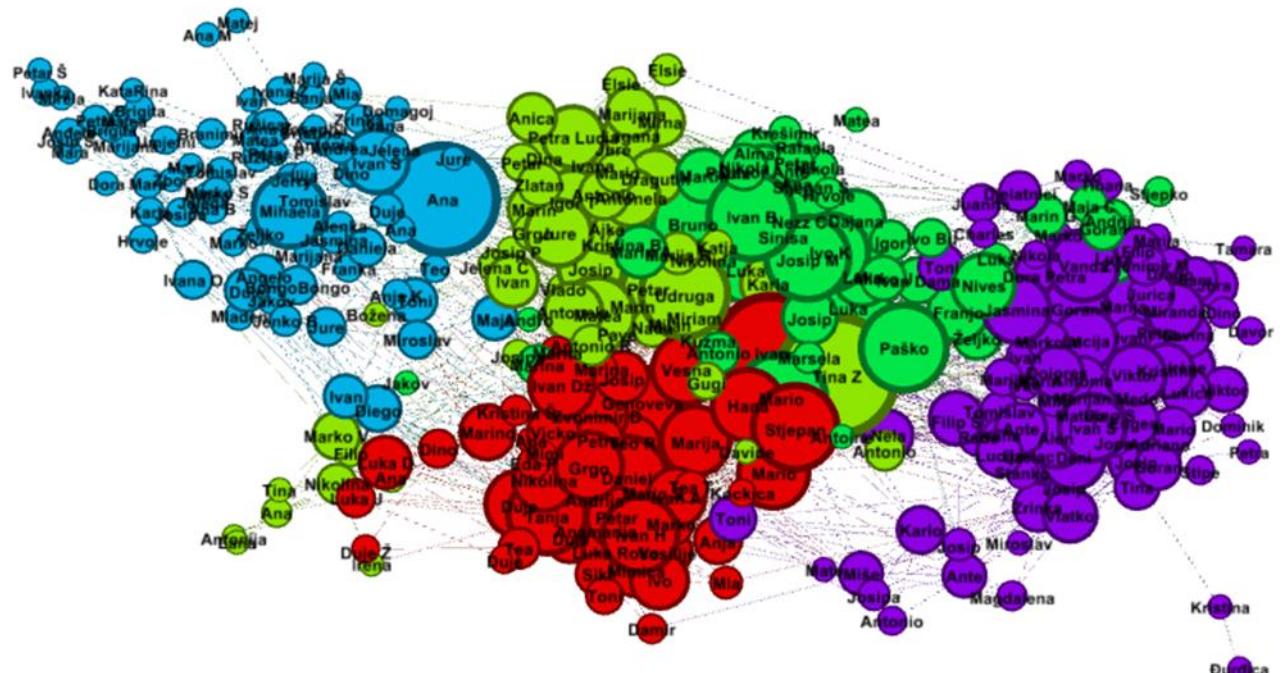
Slika 3. Mreža sa označenim klasterima

Osim klastera, alat nudi nalaženje autoriteta. Slika 5 prikazuje važnost čvorova. Čvorovi su označni labelama. Važnost čvora je prikazana proporcionalno njegovoj veličini. Osobe s labelama Ana i Tina imaju najveći autoritet. To znači da im je izlazni stupanj najveći. Pošto je prijateljstvo obostrano, ovi čvorovi su također i hubovi jer im je ulazni stupanj isti kao i izlazni. Da bi alat mogao odrediti klastera potrebno je algoritmom brute force naći modularnost grafa. Graf se sastoji od 13 klastera. Na slici 4 je prikazano koliko čvorova pripada kojem klasteru.

Size Distribution



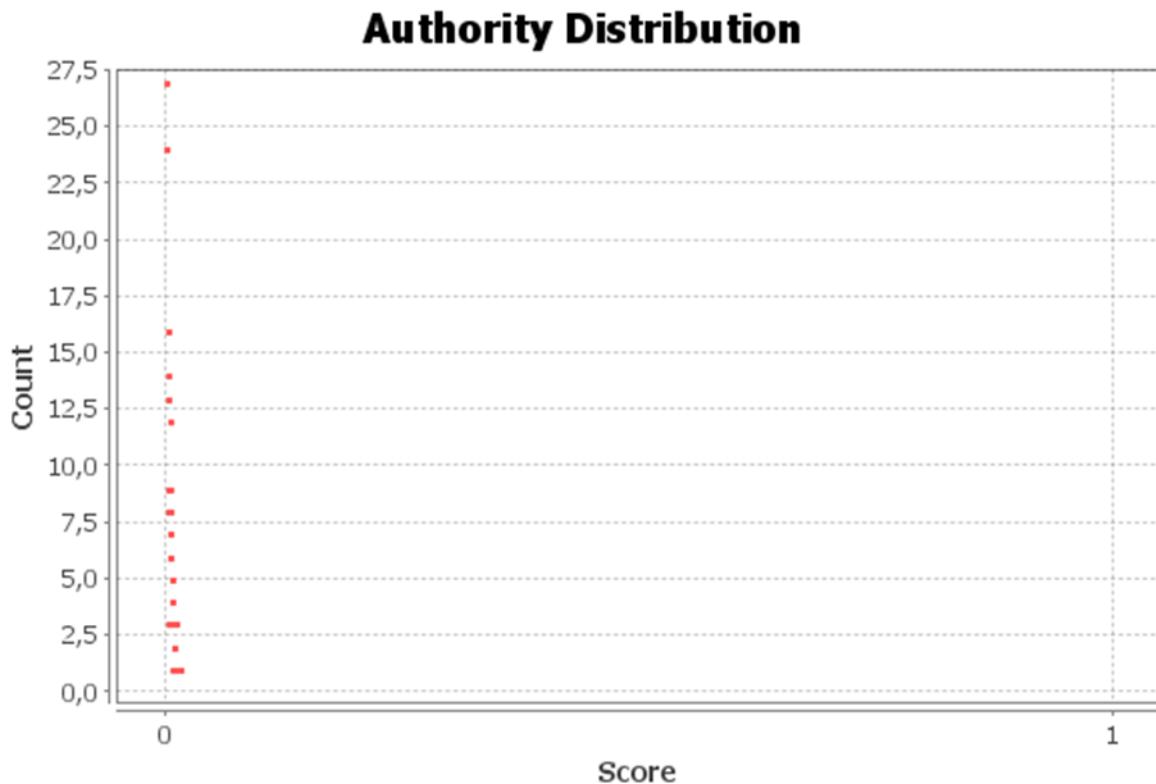
Slika 4. Modularizacija grafa



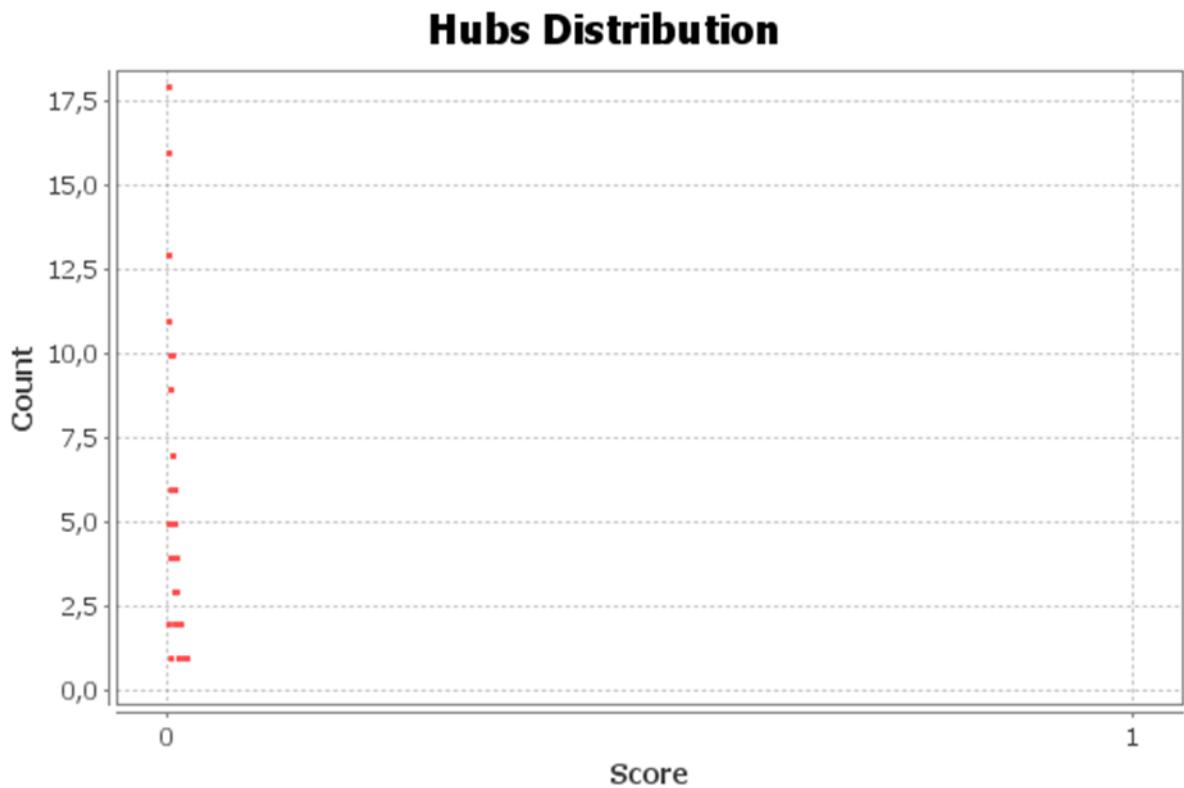
Slika 5. Važnost čvorova

Slike 6 i 7 prikazuju raspodjelu hubova i čvorova autoriteta.Grafovi su jednaki jer su prijateljstva obostrana.

Pošto je broj čvorova malen s obzirom na velike mreže i razlika između broja ulaznih, odnosno izlaznih čvorova je mala, autoriteti i hubovi ne ističu se previše, te je vrijednost važnosti svih čvorova organizirana oko 0,1 i slična za sve čvorove



Slika 6. Autoritet čvorova



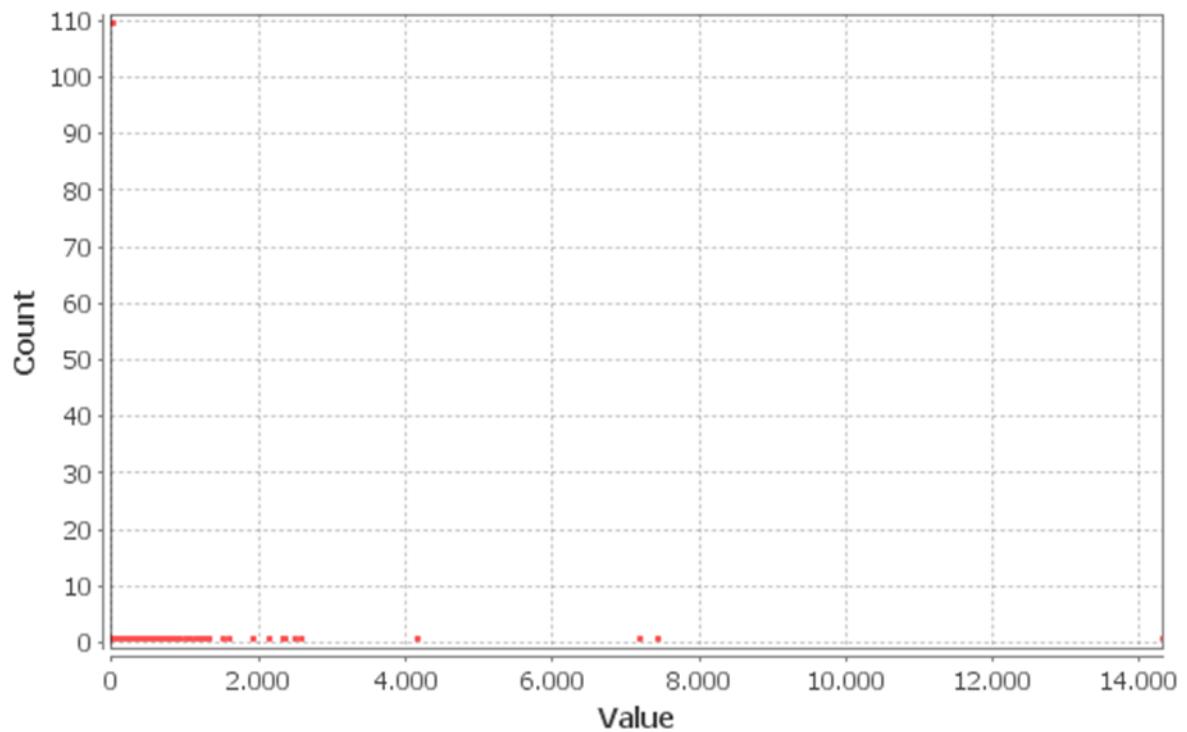
Slika 7. Raspodijela hubova

5.2. Mjere analizirane mreže

Dijametar je najveća udaljenost između dva čvora. On za promatanu mrežu iznosi 8. Znači da informacija mora prijeći najviše 8 skokova da bi došla od jednog do drugog čvora. Prosječna duljina puta je 2,664, a broj najkraćih puteva između svih parova čvorova je 68202. Sljedeća mjera koja se može promatrati je centralnost. Razdioba za betweenes centralnost je prikazana na slici 5.

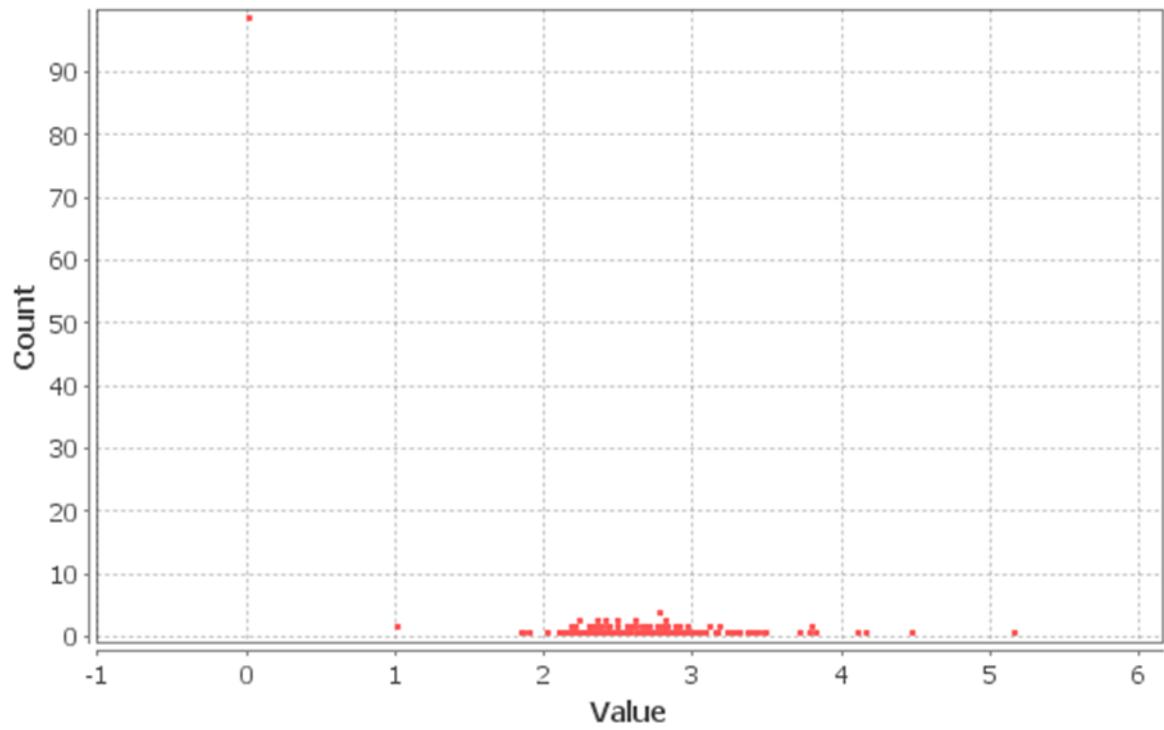
Gdje je ost ukupan broj najkraćih puteva od čvora s do čvora t a ostv broj tih puteva koji prolaze vrhom v. Na grafu se vidi da velik broj čvorova ima centralnost od 0 do dva, tek nekoliko njih ima veću centralnost. Zaključak je da je mreža poprilično decentralizirana. Na slici 6 prikazan je graf closeness centralnosti. Ova se mjera može intuitivno shvatiti kao blizina dva čvora, odnosno koliko dugo treba da se informacija proširi od jednog do drugog čvora. Velik broj čvorova je vrlo „blizu“, udaljen samo u otprilike 2 do tri skoka. Na grafu se može vidjeti nakupina parova čvorova između tih vrijednosti.

Betweenness Centrality Distribution



Slika 5. Betweenness centralnost

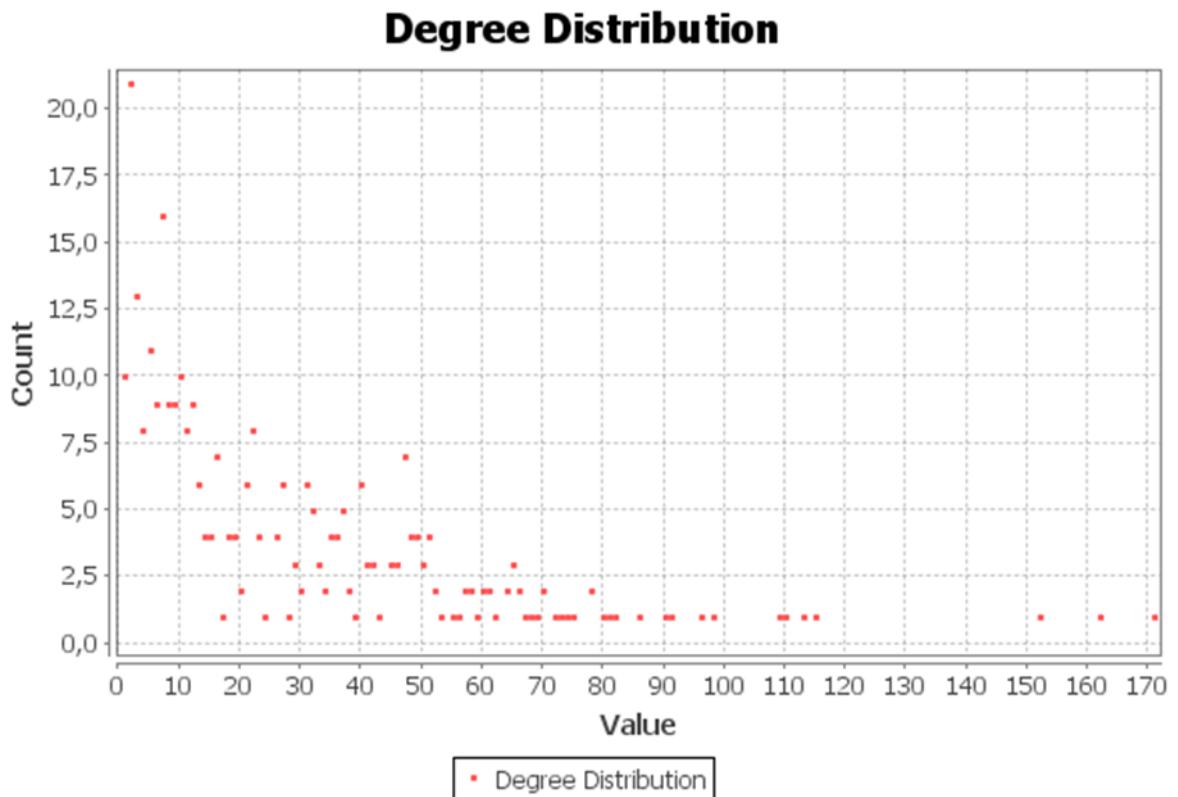
Closeness Centrality Distribution



Slika 6. Closeness centralnost

Na slici 8 može se vidjeti graf koji prikazuje broj slabih i broj jakih komponenata. Broj slabih komponenata je 5, a jako povezanih je 102.

Slika 9 prikazuje raspodijelu stupnjeva po čvorovima. Prosječan stupanj čvora je 13,835. Raspodjela stupnjeva prati logaritamsku funkciju.



Slika 9. Raspodjela stupnjeva

Zaključak

Društvene mreže virtualno okupljaju ljude i organizacije ovisno o tome za što su specijalizirane. Kao takve postale su dio svakodnevnice jer olakšavaju komunikaciju, interakciju i poslovne procese. Njihovom analizom može se doći do zaključaka koji su jako važni za donošenje odluka i bolje razumijevanje toka informacija. Ova analiza se temelji na teoriji grafova gdje su dva glavna elementa promatranja čvorovi i veze između njih. Postavljaju se pitanja o udaljenosti čvorova, postojanja veze i važnosti pojedinih čvorova. Prije odgovora na takva pitanja najprije se definiraju neke mjere kao što su stupanj čvora, usmjerenost veza, težina, rang mreže te razine povezanosti. Odgovore na postavljena pitanja daju mjere kao što je centralnost (stupanj čvora, blizina čvora, broj parova između kojih se čvor nalazi te svojstvena centralnost). Važnost čvora predstavlja mjeru autoritet. Uglavnom su važniji čvorovi koji se mostovi ili koji imaju visok stupanj izlaznih ili ulaznih veza. Ovo svojstvo čvorova se može iskoristiti pri pronašlasku ljudi traženih sposobnosti (npr. LinkedIn) ili u marketingu kada je potrebno što brže i jeftinije proširiti informaciju na što više čvorova.

Literatura

Svaki autor piše popis literature na kraju svog poglavlja. Popis literature se piše stilom literatura.

- [1] Hanneman, A.,Riddle M., Introduction to social network methods, Social and network data ,2008., Available:
http://faculty.ucr.edu/~hanneman/nettext/C1_Social_Network_Data.html
- [2] L. Adamic, "Lecture1C_Intro & Exploratory Network Analysis with Gephi.pdf", predavanje, "Social Network Analysis", University of Michigan, [20.4.2013.], Available:
<https://class.coursera.org/sna-2012-001/lecture/index>
- [5]L. Adamic, "Lecture3A_Centrality.pdf", predavanje, "Social Network Analysis", University of Michigan, [16.11.2012.], Available: <https://class.coursera.org/sna-2012-001/lecture/index>
- [6][Programski alat Gephi, [20.4.2013.], Available: <http://gephi.org>
- [7] L. Adamic, "Lecture6_ProcessesOnNetworks.pdf", predavanje, "Social Network Analysis", University of Michigan, [20.4.2013.], Available: <https://class.coursera.org/sna-2012-001/lecture/index>
- [8] L. Adamic, "Lecture8_NetworkResilience.pdf", predavanje, "Social Network Analysis", University of Michigan, [20.4.2013.], Available: <https://class.coursera.org/sna-2012-001/lecture/index>
- [9] Programske alat Gephi, [20.4.2013.], Available: <http://gephi.org>
- [10] Wolfram MathWorld, Scale Free Network, <http://mathworld.wolfram.com/Scale-FreeNetwork.html>