



UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
DEPARTMAN ZA MATEMATIKU I
INFORMATIKU



Implementacija, testiranje i primena Batagelj-Zaveršnik algoritma

-Praktični projekat iz predmeta Socijalne mreže-

Marko Mirković, 351/21

Novi Sad, jun 2024.

Sadržaj

1. Uvod.....	3
2. Poređenje i validacija algoritama dekompozicije.....	4
2.1 Ručno napravljena mreža od 15 čvorova.....	4
2.2 Ručno napravljena mreža od 17 čvorova.....	5
2.3 Ručno napravljena mreža od 20 čvorova.....	6
3. Veštački generisane mreže.....	7
3.1 Erdos-Renyi model.....	7
3.1.1 Karakteristične metrike.....	7
3.2 Barabasi-Albert model.....	8
3.2.1 Karakteristične metrike.....	8
3.3 Jezgro-periferija model.....	8
3.3.1 Karakteristične metrike.....	8
3.4. Poređenje algoritama.....	9
Za poređenje nad veštačkim mrežama koristićemo sledeći pristup:.....	9
4. Realne kompleksne mreže.....	10
4.1 Astro Physics collaboration mreža.....	10
4.1.1 Raspodela stepeni.....	10
4.1.2. Korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora.....	10
4.1.3. Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora.....	11
4.1.4. Korelacija k-core i makroskopskih strukturnih karakteristika.....	11
4.2 Facebook pages (Public figure) mreža.....	12
4.2.1 Raspodela stepeni.....	12
4.2.2. Korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora.....	12
4.2.3. Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora.....	13
4.2.4. Korelacija k-core i makroskopskih strukturnih karakteristika.....	13
4.3 High Energy Physics – Phenomenology collaboration mreža.....	14
4.3.1 Raspodela stepeni.....	14
4.3.2. Korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora.....	14
4.3.3. Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora.....	15
4.3.4. Korelacija k-core i makroskopskih strukturnih karakteristika.....	15
4.4 Gemsec-RO mreža.....	16
4.4.1 Raspodela stepeni.....	16
4.4.2. Korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora.....	16
4.4.3. Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora.....	17
4.4.4. Korelacija k-core i makroskopskih strukturnih karakteristika.....	17
5. Zaključak.....	18
6. Linkovi do mreža.....	18

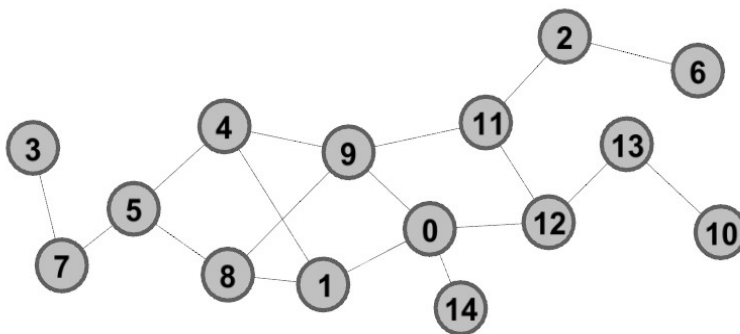
1. Uvod

Cilj ovog projekta je implementacija Batagelj-Zaveršnik algoritma za dekompoziciju grafa i poređenje ispravnosti naspram Straightforward algoritma. Nakon validacije ispravnosti koristimo Batagelj-Zaveršnik da pratimo promenu metrika kompleksnih mreža po k-jezgrima. Pored metrika nad k-jezgrima ćemo izračunati i neke globalne metrike kompleksnih mreža.

2. Poređenje i validacija algoritama dekompozicije

Algoritam ćemo validirati jednostavnim upoređivanjem vrednosti sa Straightforward algoritmom i ručnim računanjem. Nakon toga ćemo porediti vrednosti nad veštački generisanim mrežama. Mreže nad kojima vršimo validaciju su ručno osmišljene, bez neke konkretne strategije. Za vizualizaciju grafa ćemo koristiti alat „Gephi“.

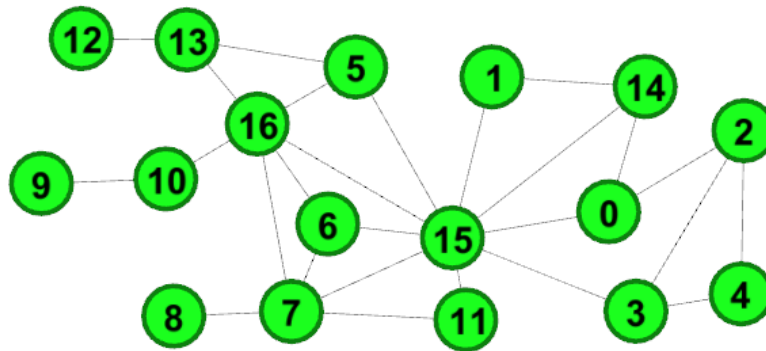
2.1 Ručno napravljena mreža od 15 čvorova



Slika 2.1.1: Ručno napravljena mreža od 15 čvorova prikazana u alatu Gephi

Čvor	Batagelj-Zaveršnik	Straightforward	Ručno
0	2	2	2
1	2	2	2
2	1	1	1
3	1	1	1
4	2	2	2
5	2	2	2
6	1	1	1
7	1	1	1
8	2	2	2
9	2	2	2
10	1	1	1
11	2	2	2
12	2	2	2
13	1	1	1
14	1	1	1

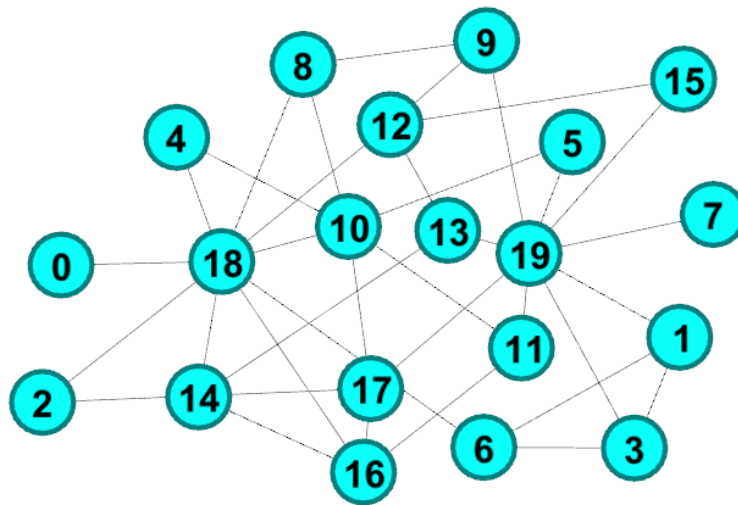
2.2 Ručno napravljena mreža od 17 čvorova



Slika 2.2.1: Ručno napravljena mreža od 17 čvorova prikazana u alatu Gephi

Čvor	Batagelj-Zaveršnik	Straightforward	Ručno
0	2	2	2
1	2	2	2
2	2	2	2
3	2	2	2
4	2	2	2
5	2	2	2
6	3	3	3
7	3	3	3
8	1	1	1
9	1	1	1
10	1	1	1
11	2	2	2
12	1	1	1
13	2	2	2
14	2	2	2
15	3	3	3
16	3	3	3

2.3 Ručno napravljena mreža od 20 čvorova



Slika 2.3.1: Ručno napravljena mreža od 20 čvorova prikazana u alatu Gephi

Čvor	Batagelj-Zaveršnik	Straightforward	Ručno
0	1	1	1
1	3	3	3
2	2	2	2
3	3	3	3
4	2	2	2
5	2	2	2
6	3	3	3
7	1	1	1
8	3	3	3
9	3	3	3
10	3	3	3
11	3	3	3
12	3	3	3
13	3	3	3
14	3	3	3
15	2	2	2
16	3	3	3
17	3	3	3
18	3	3	3
19	3	3	3

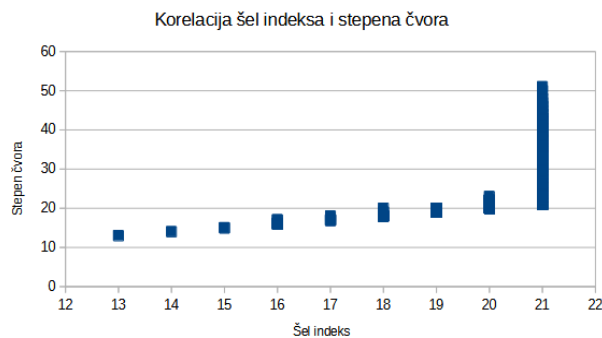
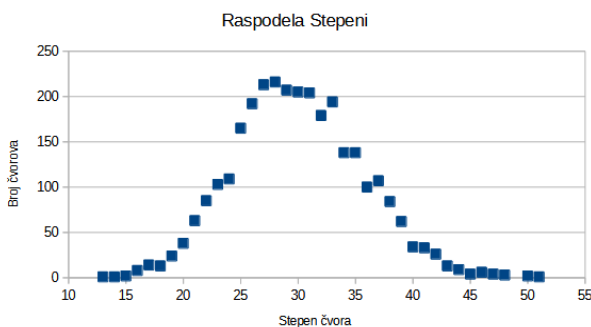
3. Veštački generisane mreže

U ovom delu ćemo porediti Batagelj-Zaveršnik sa Straightforward algoritmom nad veštački generisanim mrežama. Poređenje ćemo vršiti pokretanjem oba algoritma nad istim grafovima i poređenjem vrednosti. Pomoću Erdos-Renyi modela, Barabasi-Albert modela i modifikovanog Planted partition modela, koji kreira mreže sa strukturom jezgro-periferija, ćemo generisati mreže. Pored samih poređenja algoritama, pokazaćemo i neke karakteristike mreža generisanih tim modelima, koje nam mogu pomoći da razumemo metrike nad realnim mrežama.

3.1 Erdos-Renyi model

Erdos-Renyi model ima više različitih varijanti, u ovom slučaju implementiran je Gilbertov model. Parametri mreže su: 3000 čvorova i verovatnoća spajanja dva čvora je 0.01.

3.1.1 Karakteristične metrike

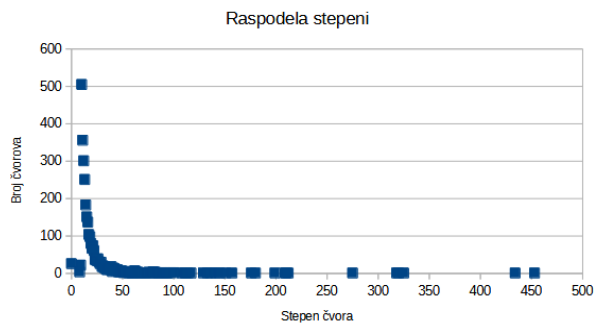


Na slici vidimo da ovaj model ima tipičnu zvonastu (bell, tj. Poasonovu) raspodelu. Takođe i na drugom grafikonu vidimo jednu od korelacija ER mreže koja će nam biti korisna da razumemo model jezgro-periferija.

3.2 Barabasi-Albert model

Barabasi-Albert model generiše scale-free mreže po principu preferencijalnog vezivanja. Parametri mreže su: 3000 čvorova, 30 početnih čvorova, verovatnoća za početne je 0.01 i svaki nov čvor se vezuje sa 10 drugih.

3.2.1 Karakteristične metrike

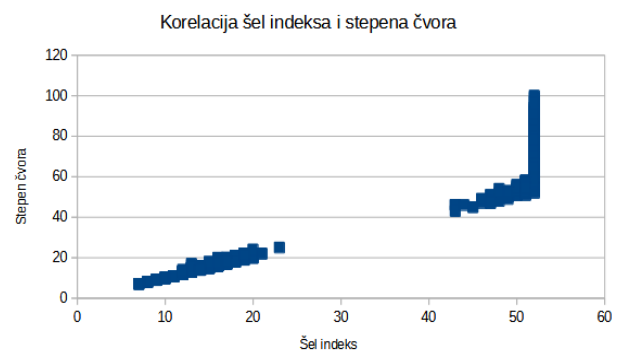
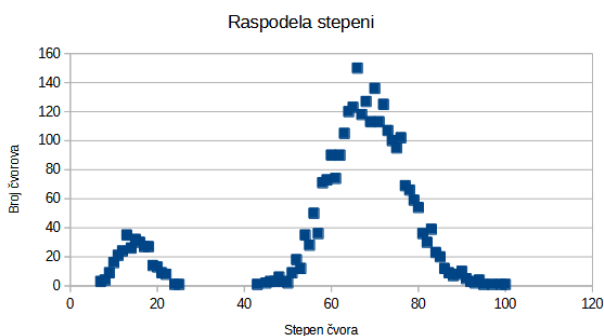


Na slici vidimo da ovaj model ima tipičnu raspodelu stepenog zakona (Power Law). Videćemo da neke realne mreže imaju ovakvu raspodelu stepeni.

3.3 Jezgro-periferija model

Jezgro-periferija model je modifikacija Planted partition modela (koji je specijalni slučaj Stohastičkog blok modela). Omogućava nam da dobijemo grafove koji imaju jezgro i periferiju. Parametri mreže su: 3000 čvorova, 90% čvorova je u jezgri, verovatnoća vezivanja dva čvora iz jezgra je 0.02, periferije 0.001, jednog iz jezgra drugog iz periferije 0.005.

3.3.1 Karakteristične metrike



Na slici vidimo da ovaj model ima nešto što liči na dve Poasonove raspodele, tj. kao da imamo dve Erdos-Renyi mreže sa različitim karakteristikama. Takođe, u globalnim metrikama centralnosti možemo jasno videti tu razliku.

3.4. Poređenje algoritama

Za poređenje nad veštačkim mrežama korišćemo sledeći pristup:

1. Generisaćemo po jednu mrežu sa Erdos-Renyi, Barabasi-Albert i Jezgro-periferija modelima;
2. Izračunaćemo šel indekse nad tim mrežama sa oba algoritma;
3. Proćićemo kroz svaki čvor i porediti rezultate algoritama, ukoliko dodje do razlike, ispisaće se na kom čvoru se nalazi razlika.

```
//comparing Batagelj-Zaversnik and Straightforward on artificial networks
ErdosRenyi<Integer, String> er = new ErdosRenyi<>();
BarabasiAlbert<Integer, String> ba = new BarabasiAlbert<>();
PlantedPartition<Integer, String> pp = new PlantedPartition<>();
UndirectedSparseGraph<Integer, String> graphErdos = er.generateER( n: 3000, p: 0.01, i -> i, s -> s);
UndirectedSparseGraph<Integer, String> graphBarabasi = ba.generateBA( n: 3000, m0: 30, p: 0.01, m: 10, i -> i, e -> e);
UndirectedSparseGraph<Integer, String> graphPlanted = pp.generatePP( n: 3000, coreNodes: 0.9, coreProbability: 0.02,
    peripheryProbability: 0.001, corePeripheryProbability: 0.005, i -> i, s -> s);
dm.exportDecompositionAsCSV(graphErdos, fileName: "Erdos-Renyi");
System.out.println("Erdos-Renyi done.");
dm.exportDecompositionAsCSV(graphBarabasi, fileName: "Barabasi-Albert");
System.out.println("Barabasi-Albert done.");
dm.exportDecompositionAsCSV(graphPlanted, fileName: "PlantedPartition");
System.out.println("PlantedPartition done.");
```

Slika 3.4.1: Pokretanje koda za poređenje

```
public void exportDecompositionAsCSV(UndirectedSparseGraph<Integer, String> graph, String fileName) { 3 usages 1 MarkbKillin
    BatageljZaversnik<Integer, String> bz = new BatageljZaversnik<>(graph);
    Straightforward<Integer, String> sf = new Straightforward<>(graph);
    Map<Integer, Integer> shellIndicesBZ = bz.getShellIndices();
    Map<Integer, Integer> shellIndicesSF = sf.getShellIndices();
    try (PrintWriter pw = new PrintWriter(new BufferedWriter(new FileWriter(fileName: "./exportedComparison/" + fileName + ".csv")))) {
        for (Integer vertex : graph.getVertices()) {
            if (shellIndicesBZ.get(vertex).intValue() != shellIndicesSF.get(vertex).intValue()) {
                System.out.println("Vertex " + vertex + " has different shell indices.");
            }
            pw.println(vertex + "," + shellIndicesBZ.get(vertex) + "," + shellIndicesSF.get(vertex));
        }
    } catch (IOException e) {
        System.out.println("Error writing to file " + fileName + ".csv");
    }
}
```

Slika 3.4.2: Metoda za poređenje i izvoz rezultata

```
Erdos-Renyi done.
Barabasi-Albert done.
PlantedPartition done.

Process finished with exit code 0
```

Slika 3.4.3: Rezultat izvršavanja koda sa slike 3.4.1

Kao što se vidi sa slike 3.4.3, rezultati oba algoritma su identični.

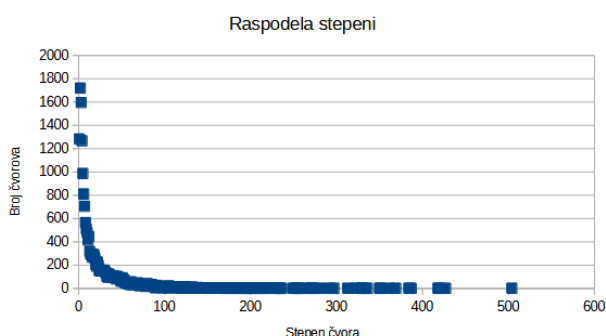
4. Realne kompleksne mreže

U ovom delu ćemo prikazati metrike nad realnim neusmerenim mrežama preuzetih sa repozitorijuma grafovskih podataka. Sve mreže su predstavljene preko grana, što znači da nećemo imati izolovanih čvorova, ali i dalje je moguće imati više komponenti povezanosti.

4.1 Astro Physics collaboration mreža

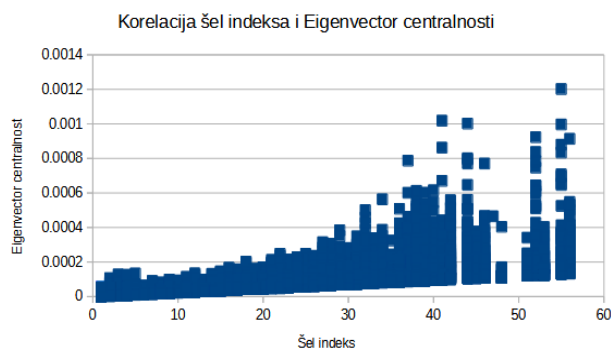
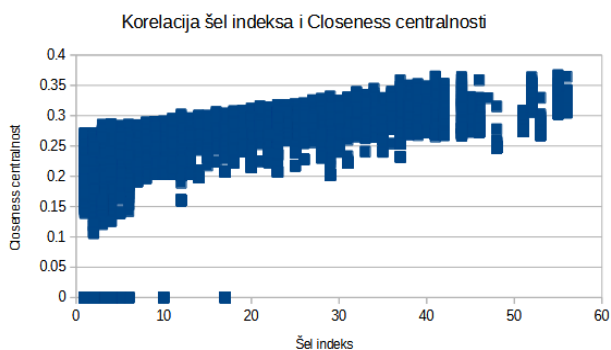
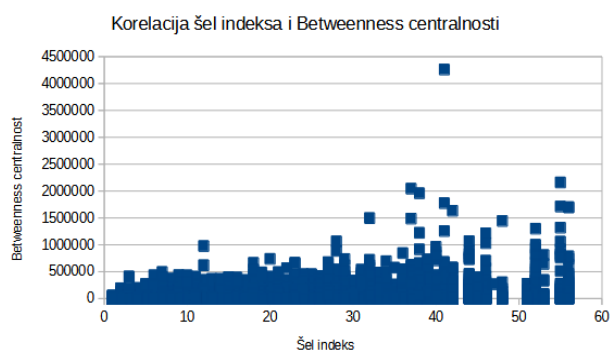
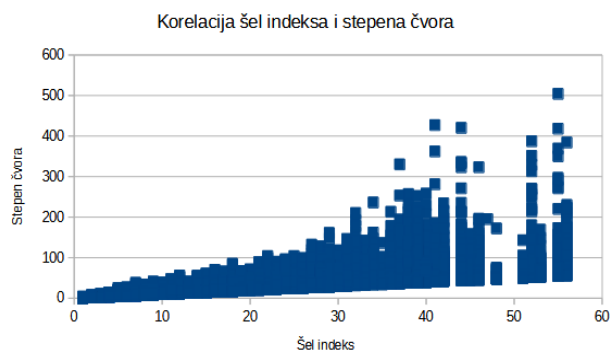
Astro-ph je mreža saradnje naučnika. Čvorovi su naučnici, grane predstavljaju saradnju dva naučnika na jednom radu. Mreža ima 18.8k čvorova i 198k grana.

4.1.1 Raspodela stepeni



Na slici vidimo da ovaj model ima raspodelu stepenog zakona, koju smo prethodno videli kod Barabasi-Albert modela.

4.1.2. Korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora



4.1.3. Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora

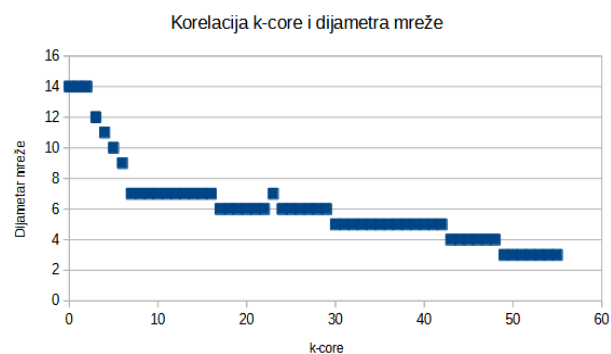
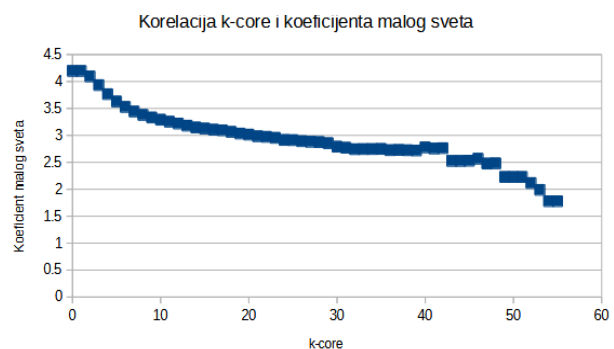
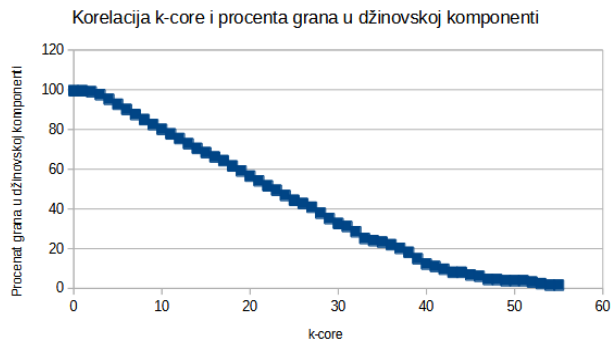
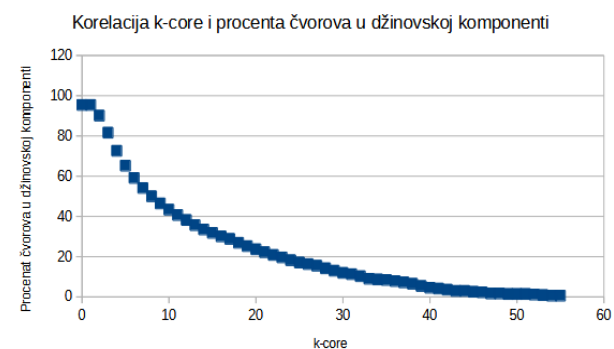
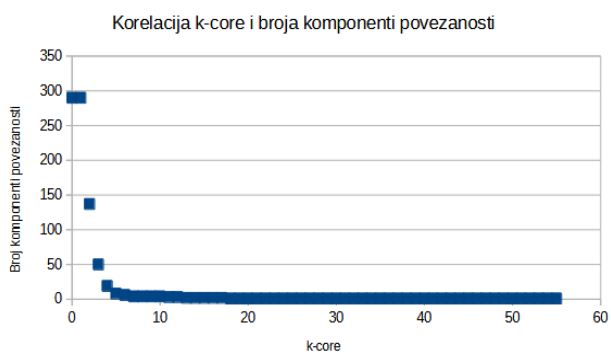
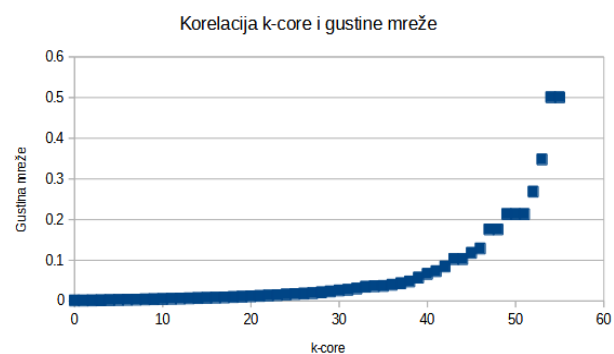
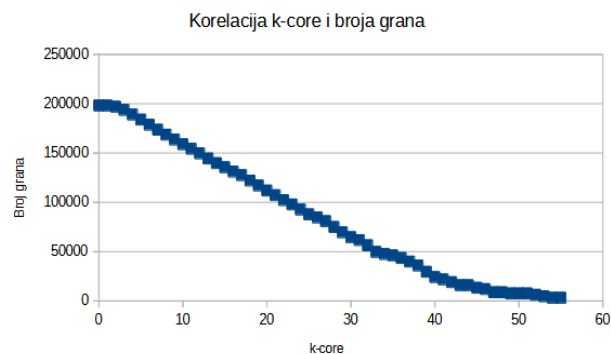
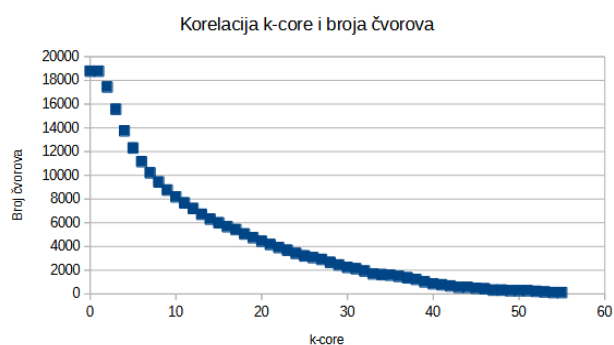
Šel-Stepen: 0.976933249543122

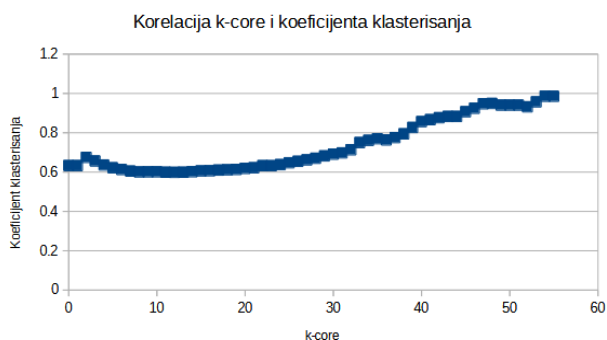
Šel-Betweenness: 0.581828584597045

Šel-Closeness: 0.615510356545612

Šel-Eigenvector: 0.878294669530764

4.1.4. Korelacija k-core i makroskopskih strukturnih karakteristika

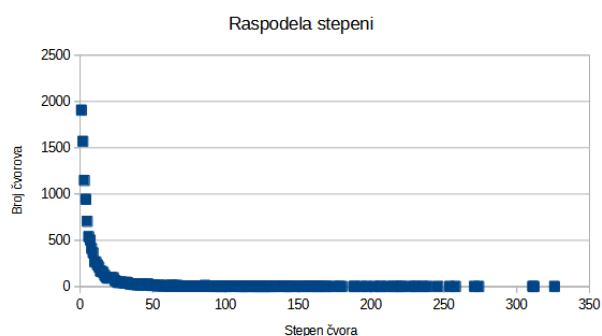




4.2 Facebook pages (Public figure) mreža

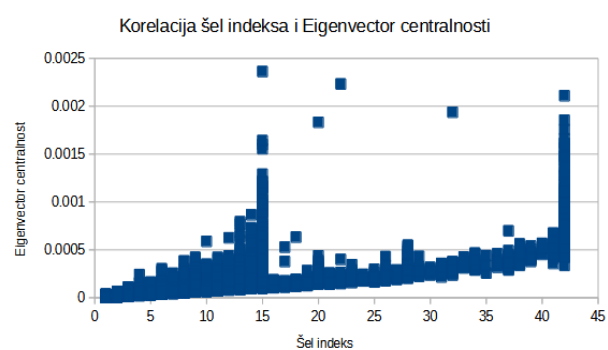
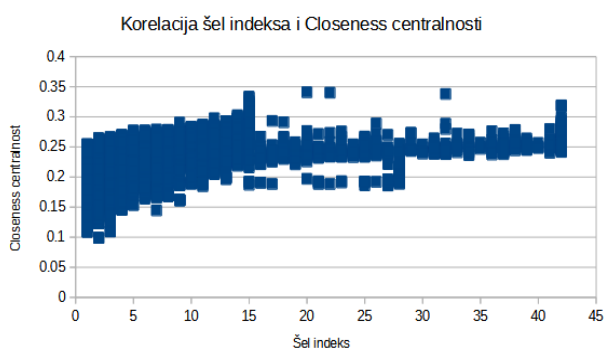
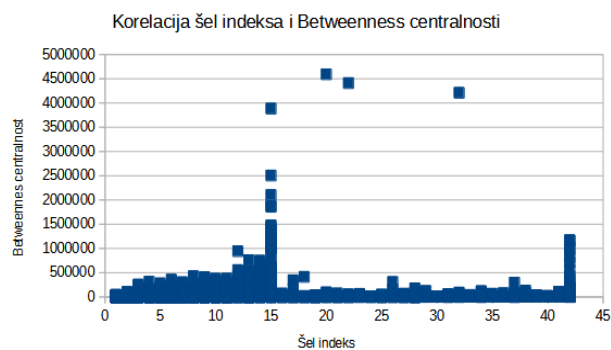
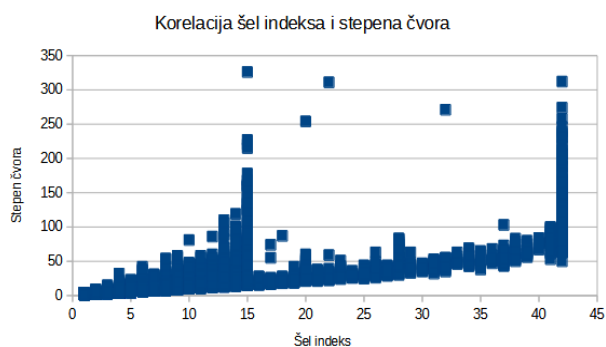
Predstavlja mrežu zajedničkih lajkova stranica poznatih ličnosti na Facebook-u. Čvorovi su stranice, a grane su zajednički lajkovi između njih. Mreža ima 11.6k čvorova i 67k grana.

4.2.1 Raspodela stepeni



Na slici vidimo da ovaj model takođe ima raspodelu stepenog zakona.

4.2.2. Korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora



4.2.3. Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora

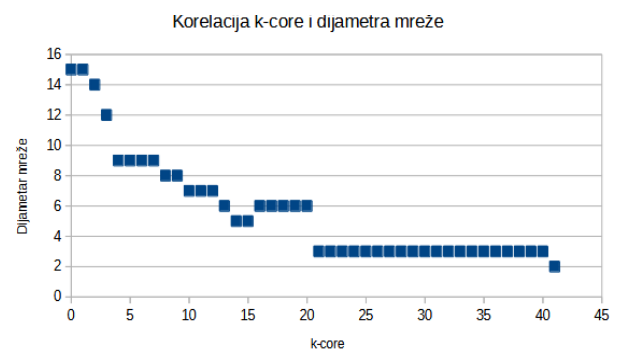
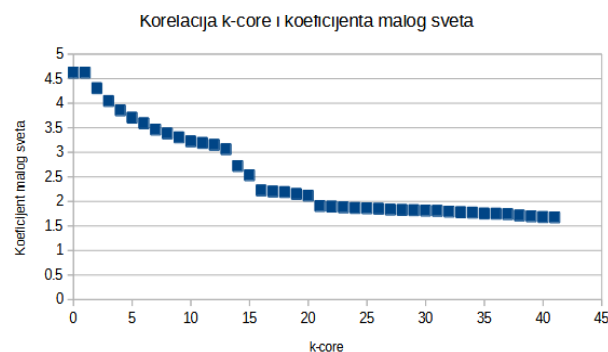
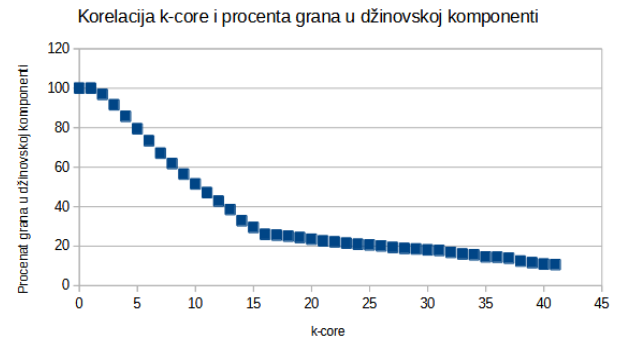
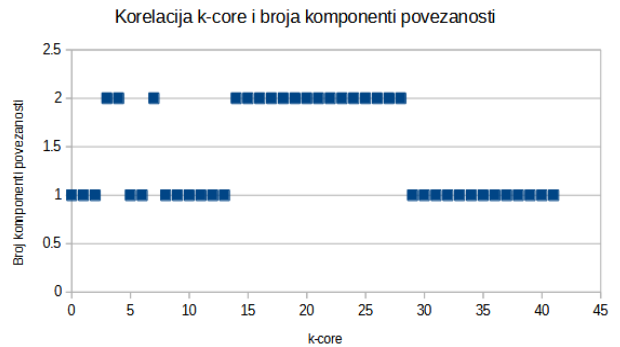
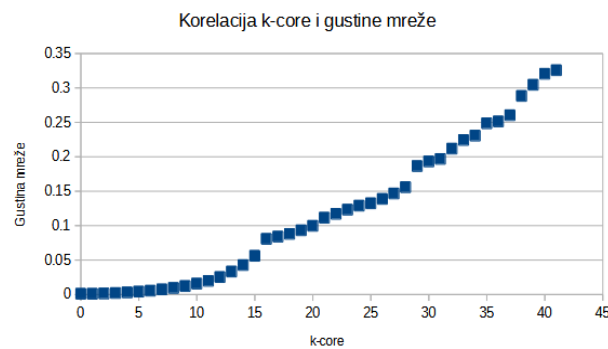
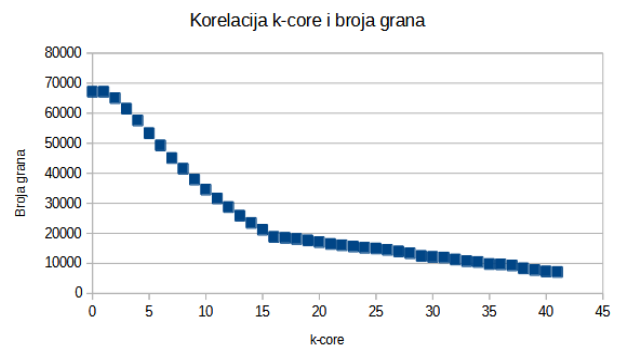
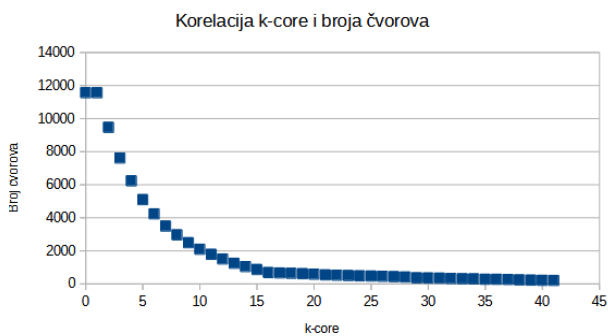
Šel-Stepen: 0.971099429844222

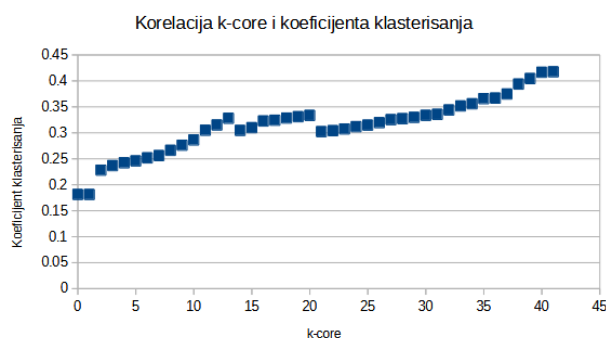
Šel-Betweenness: 0.680354330310272

Šel-Closeness: 0.702134839181497

Šel-Eigenvector: 0.963199361351083

4.2.4. Korelacija k-core i makroskopskih strukturnih karakteristika

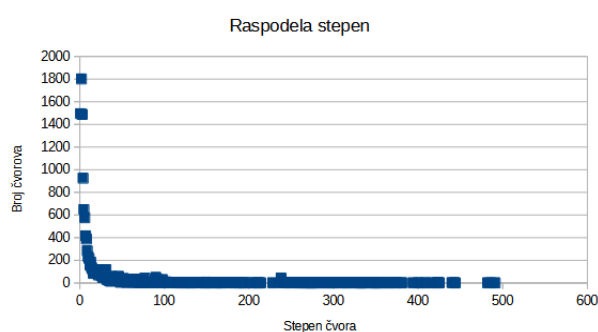




4.3 High Energy Physics – Phenomenology collaboration mreža

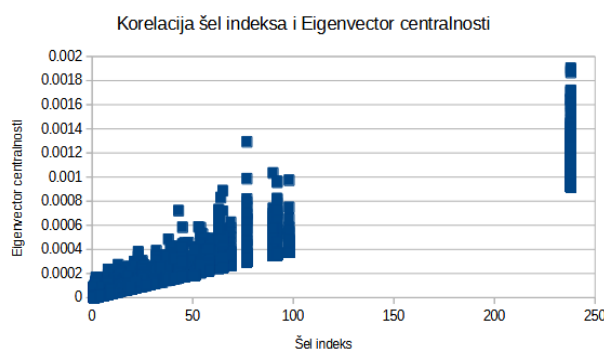
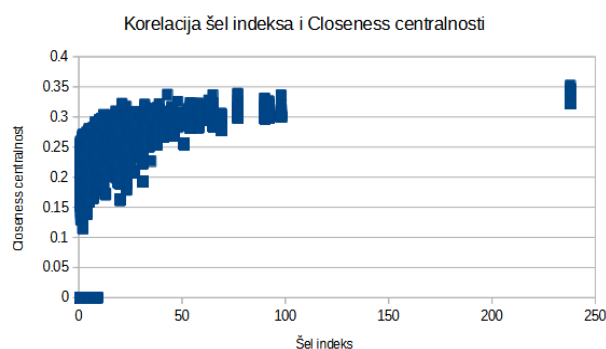
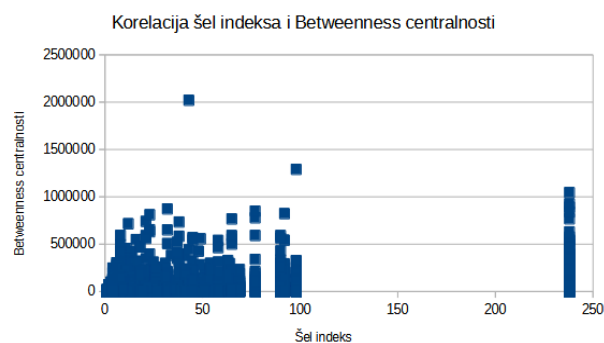
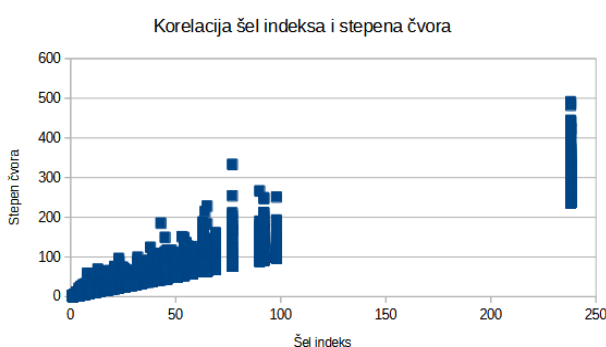
HepPh je mreža saradnje naučnog istraživanja. Čvorovi su autori radova, a grane predstavljaju saradnju na nekom radu između dva autora. Mreža ima 12k čvorova i 118.5k grana. Ova mreža nam daje specifične rezultate. Na prvi pogled izgleda kao obična mreža kolaboracije, ali zapravo ima nešto slično modelu jezgro-periferija. Očigledno postoji bolje povezan deo mreže.

4.3.1 Raspodela stepeni



Na slici vidimo da ovaj model ima raspodelu stepenog zakona, koju smo prethodno videli kod Barabasi-Albert modela i drugih realnih mreža.

4.3.2. Korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora



4.3.3. Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora

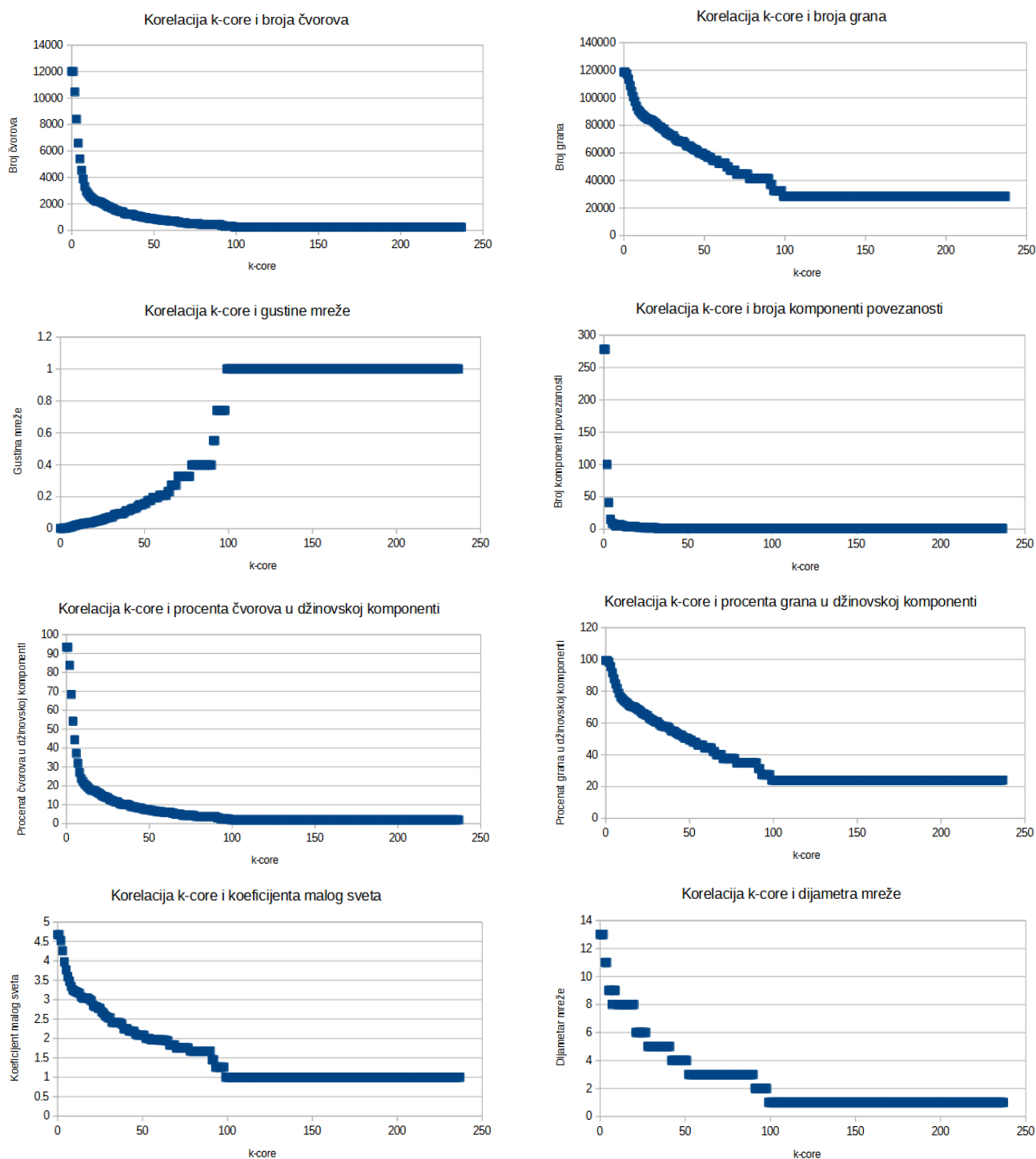
Šel-Stepen: 0.974901842330749

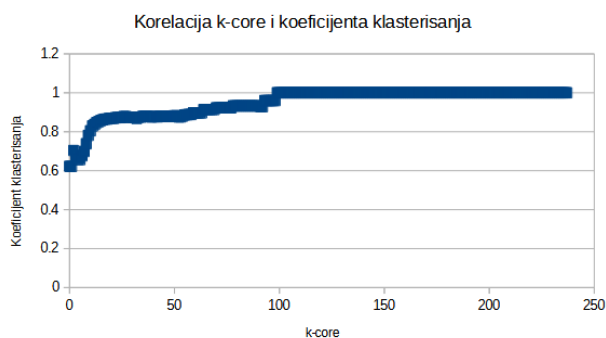
Šel-Betweenness: 0.49906502499704

Šel-Closeness: 0.467528711092914

Šel-Eigenvector: 0.821733232585556

4.3.4. Korelacija k-core i makroskopskih strukturnih karakteristika

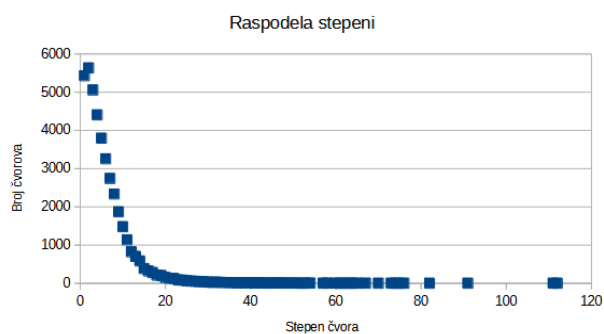




4.4 Gemsec-RO mreža

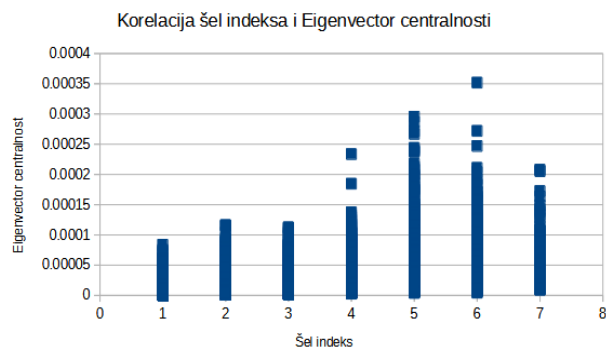
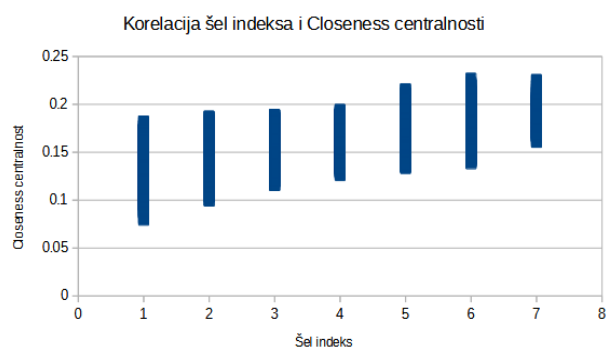
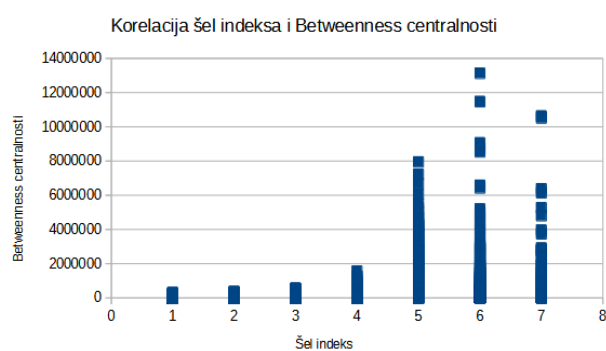
Gemsec je socijalna mreža prikupljena sa platforme Deezer. Čvorovi predstavljaju korisnike, a grane zajedničke prijatelje. Mreža ima 41.8k čvorova i 125.8k grana.

4.4.1 Raspodela stepeni



Na slici vidimo da ovaj model ima raspodelu stepenog zakona, koju smo prethodno videli kod Barabasi-Albert modela i drugih realnih mreža.

4.4.2. Korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora



4.4.3. Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora

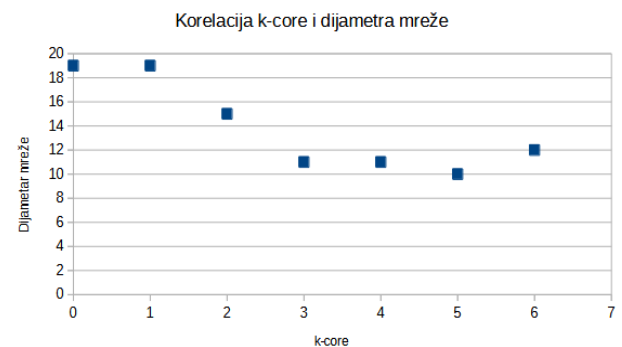
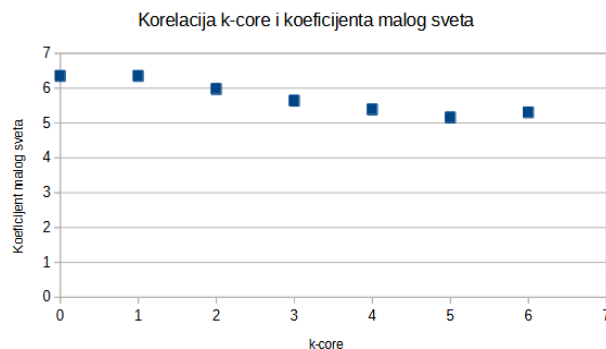
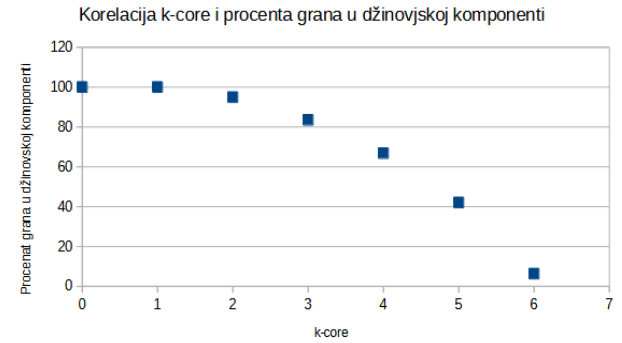
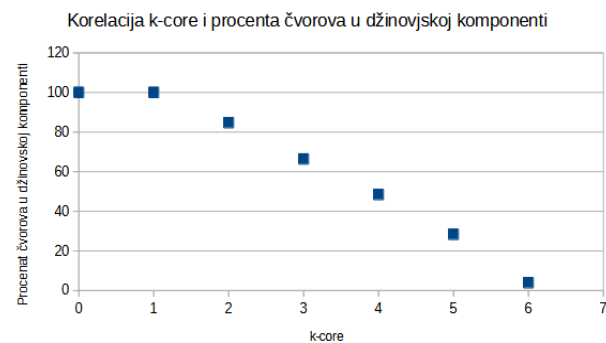
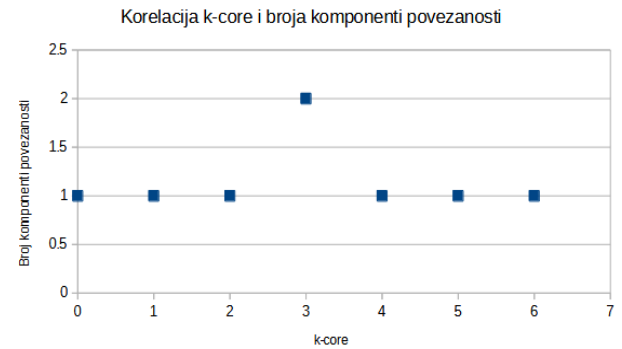
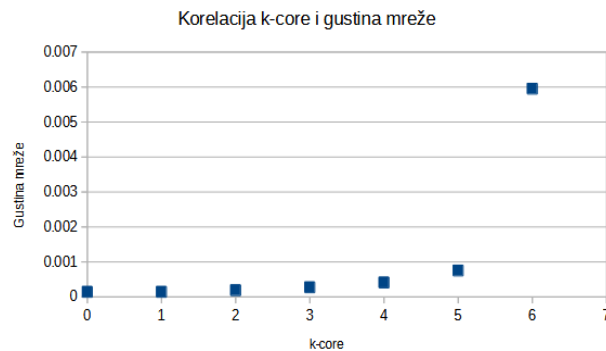
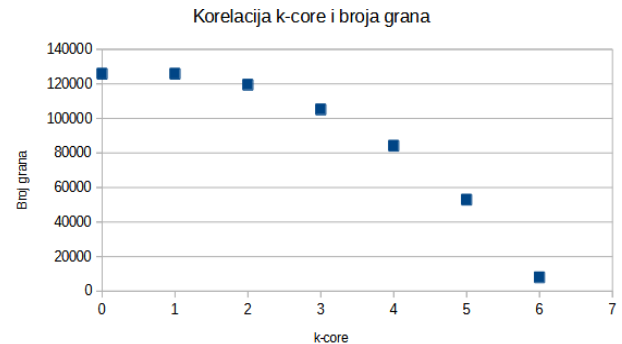
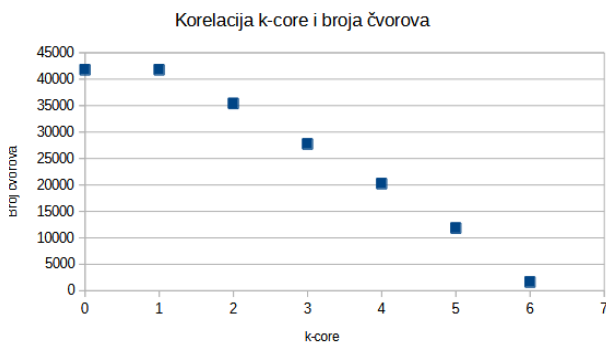
Šel-Stepen: 0.926700940300059

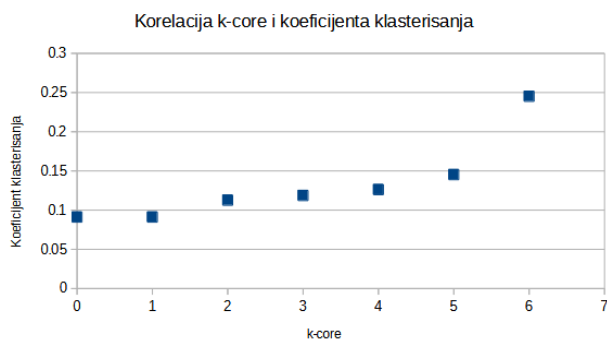
Šel-Betweenness: 0.728906611820194

Šel-Closeness: 0.791381856669232

Šel-Eigenvector: 0.578312530572922

4.4.4. Korelacija k-core i makroskopskih strukturnih karakteristika





5. Zaključak

Primećujemo da na svim realnim mrežama makroskopske strukturne karakteristike imaju sličan trend. Sa povećanjem k-core koeficijent klasterisanja raste, gustina raste, a skoro sve ostale opadaju. Takođe kod svih realnih mreža primećujemo da je raspodela stepeni zapravo raspodela stepenog zakona i ukazuje na postojanje habova. Vidimo da čak i velike mreže mogu imati mali maksimalni šel indeks kao na primer mreža 4.4. Takođe i globalne metrike centralnosti imaju rastući trend sa rastom šel indeksa. Korelacija između globalnih metrika centralnosti i šel indeksa je u velikoj većini slučajeva veća od 0.5.

6. Linkovi do mreža

4.1 – AstroPH - <https://snap.stanford.edu/data/ca-AstroPh.html>

4.2 – Facebook - <https://networkrepository.com/fb-pages-public-figure.php>

4.3 – High energy physics - <https://snap.stanford.edu/data/ca-HepPh.html>

4.4 – Gemsec - <https://networkrepository.com/soc-gemsec-RO.php>