



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



# Dekompozicija grafa na K-jezgra i analiza menjanja globalnih metrika po jezgrima

Projekat iz predmeta Socijalne mreže

Stefan Obradović 287/21

# Sadržaj

<b>Sadržaj</b>	<b>1</b>
<b>1 Uvod</b>	<b>2</b>
<b>2 Poređenje rezultata algoritama</b>	<b>2</b>
2.1 Mreža 1 . . . . .	2
2.2 Mreža 2 . . . . .	3
2.3 Mreža 3 . . . . .	4
<b>3 Modeli za generisanje nasumičnih mreža</b>	<b>4</b>
3.1 Model Jezgro-Periferija . . . . .	4
<b>4 Prikupljeni podaci</b>	<b>5</b>
4.1 Erdos-Renyi model . . . . .	5
4.2 Barabas-Albert model . . . . .	6
4.3 Jezgro-periferija model . . . . .	8
4.3.1 Mreža 1 . . . . .	8
4.3.2 Mreža 2 . . . . .	10
4.3.3 Mreža 3 . . . . .	11
4.4 Condensed matter collaboration 2003 . . . . .	13
4.5 Condensed matter collaborations 2005 . . . . .	15
4.6 Enron email mreža . . . . .	16
4.7 Github mreža . . . . .	18
<b>5 Zaključak</b>	<b>20</b>
<b>6 Literatura</b>	<b>21</b>

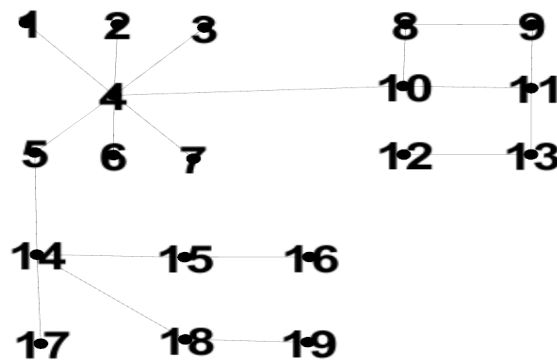
## 1 Uvod

Implementiran je Batagelj-Zaverštnik algoritam za dekompoziciju mreže na K-jezgra. Poredi se tačnost ovog algoritma upoređujući njegov rezultat sa rezultatima „Straight-forward” algoritma. Posmatra se promena globalnih metrika mreže po K-jezgrima nad veštački generisanim i realnim mrežama.

## 2 Poređenje rezultata algoritama

Mreže u ovoj oblasti su osmišljeni bez neke posebne strukture, već su napravljeni nasumičnim crtanjem čvorova i veza na papiru. Napravljeni su samo za testiranje tačnosti Batagelj-Zaverštnik algoritma.

### 2.1 Mreža 1

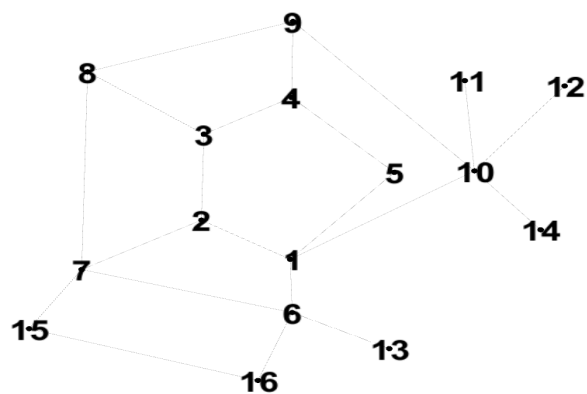


Slika 1: Izgled malog grafa za testiranje algoritma

Čvor	Ručno računanje	Batagelj-Zaverštnik	„Straight forward”
1	1	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	1	1	1
6	1	1	1
7	1	1	1
8	2	2	2
9	2	2	2
10	2	2	2
11	2	2	2
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	1
16	1	1	1
17	1	1	1
18	1	1	1
19	1	1	1

Tabela 1: Šel indeksi za čvorove dobijeni različitim metodama

## 2.2 Mreža 2

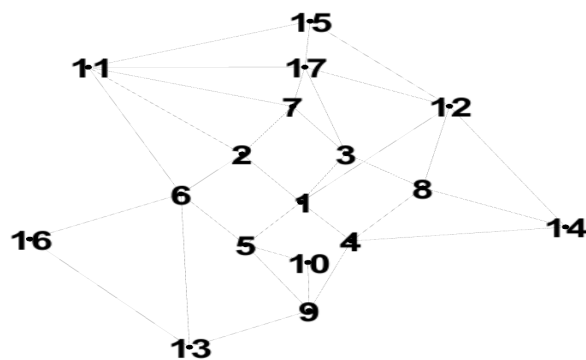


Slika 2: Izgled malog grafa za testiranje algoritma

Čvor	Ručno računanje	Batagelj-Zaverštnik	„Straight forward”
1	2	2	2
2	2	2	2
3	2	2	2
4	2	2	2
5	2	2	2
6	2	2	2
7	2	2	2
8	2	2	2
9	2	2	2
10	2	2	2
11	1	1	1
12	1	1	1
13	1	1	1
14	1	1	1
15	2	2	2
16	2	2	2

Tabela 2: Šel indeksi za čvorove dobijeni različitim metodama

## 2.3 Mreža 3



Slika 3: Izgled malog grafa za testiranje algoritma

Čvor	Ručno računanje	Batagelj-Zaverštnik	„Straight forward”
1	3	3	3
2	3	3	3
3	3	3	3
4	3	3	3
5	2	2	2
6	2	2	2
7	3	3	3
8	3	3	3
9	2	2	2
10	2	2	2
11	3	3	3
12	3	3	3
13	2	2	2
14	3	3	3
15	3	3	3
16	2	2	2
17	3	3	3

Tabela 3: Šel indeksi za čvorove dobijeni različitim metodama

## 3 Modeli za generisanje nasumičnih mreža

Svaki model je modifikovan tako da svaki čvor ima sigurno barem jednog suseda. „Erdos-Renyi” model je modifikovan tako da za svaki čvor, ukoliko ima stepen 0 pre nasumičnog povezivanja sa drugim čvorovima, nasumično izabere 1 čvor (svaki čvor ima jednaku šansu da bude izabran) i poveže se sa njim.

„Barabasi-Albert” model se oslanja na prethodni model, zbog čega početna konfiguracija mreže ne može da ima čvor koji nema suseda. Svaki sledeći čvor koji se ne nalazi u početnoj konfiguraciji unosi novu granu u mrežu, samim tim imaće barem 1 suseda.

### 3.1 Model Jezgro-Periferija

Model je baziran na ideji stohastičkog blok modela za mreže koje imaju strukturu zajednica. Model umesto proizvoljnog broja zajednica ima već 2 ugrađene zajednice, jezgro i periferiju. Model ima 3 parametra:

- Broj čvorova u mreži
- Procenat čvorova u jezgru mreže
- Niz verovatnoća po kojima se čvorovi povezuju

Niz verovatnoća očekuje 3 vrednosti koje predstavljaju verovatnoću da su neka 2 čvora povezana u zavisnosti od toga u kom delu mreže se nalaze (jezgro-jezgro, periferija-periferija, jezgro-periferija).

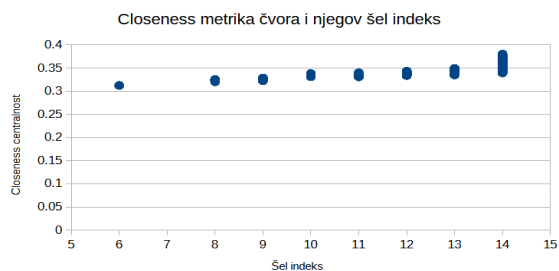
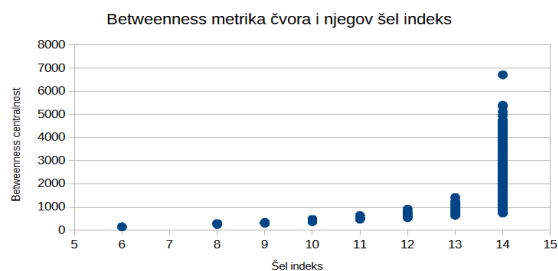
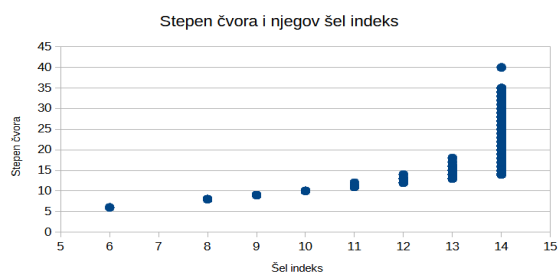
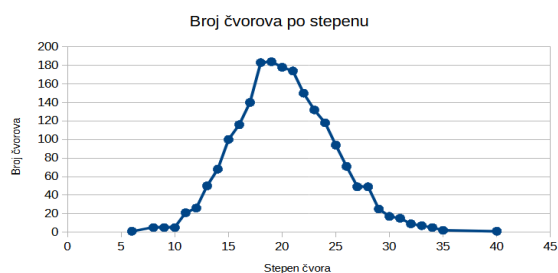
Čvorovi se povezuju slično kao u „Erdos-Renyi” modelu. Prvo se za svaki čvor proveriti da li čvor ima suseda, ako nema poveže se sa nasumično izabranim čvorom iz iste oblasti u kojoj se on nalazi (jezgro ili periferija). U suprotnom čvor se povezuje sa ostalim čvorovima po odgovarajućoj verovatnoći.

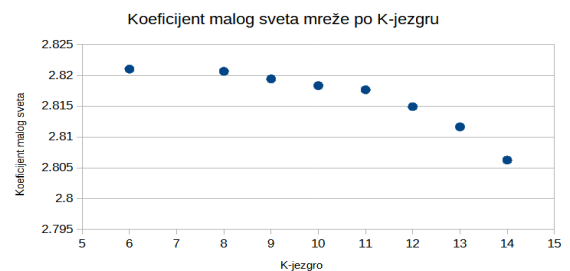
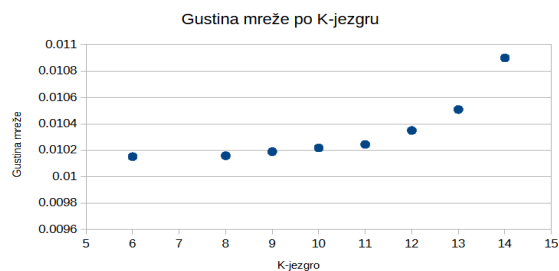
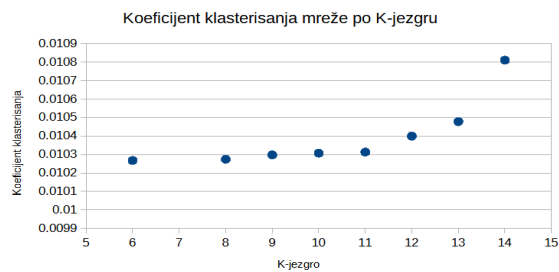
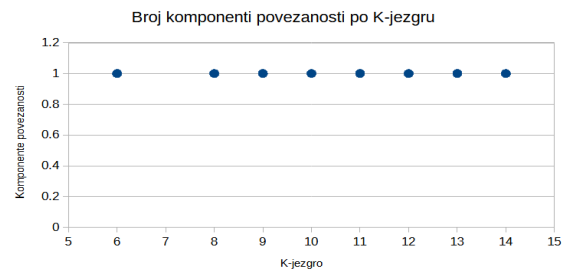
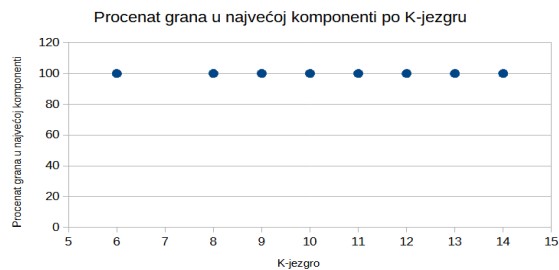
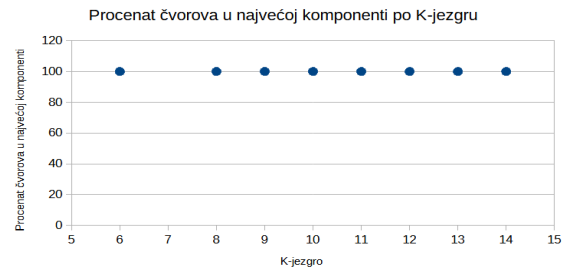
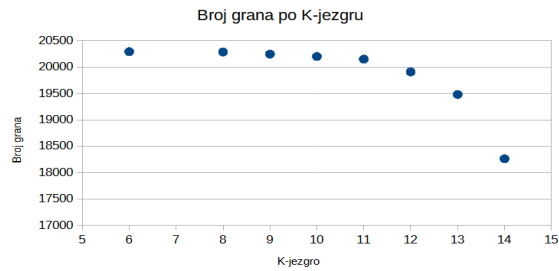
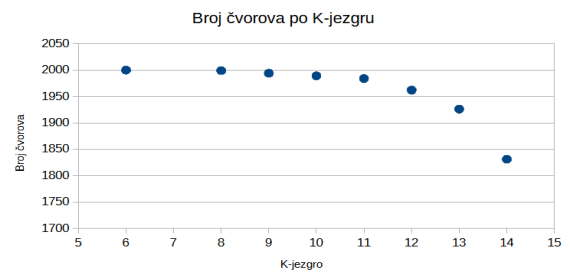
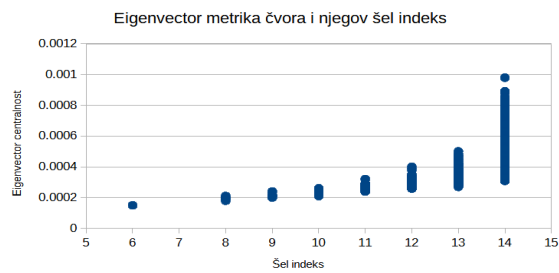
## 4 Prikupljeni podaci

Podaci koji su prikazani za neku od mreža koja je generisana po nekom modelu je nasumično izabrana od generisanih.

### 4.1 Erdos-Renyi model

- Broj čvorova u mreži: 2000
- Broj grana u mreži: 20292
- Procenat čvorova u najvećoj komponenti: 100%
- Procenat grana u najvećoj komponenti: 100%
- Gustina mreže: 0,01
- Broj komponenti povezanosti: 1
- Dijametar mreže: 4
- Koeficijent klasterisanja mreže: 0,010
- Koeficijent malog sveta: 2,821
- Spermanov indeks korelacije za šel indeks i stepen povezanosti: 0,473
- Spermanov indeks korelacije za šel indeks i „Betweenness” metriku: 0,468
- Spermanov indeks korelacije za šel indeks i „Closeness” metriku: 0,473
- Spermanov indeks korelacije za šel indeks i „Eigenvector” metriku: 0,450

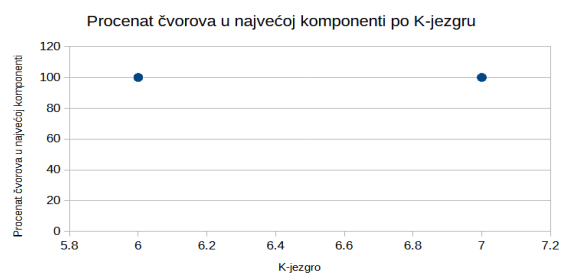
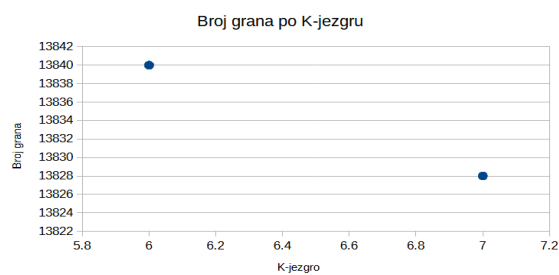
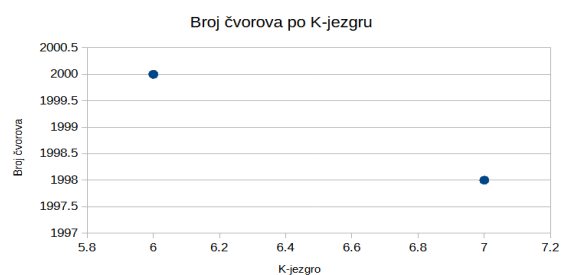
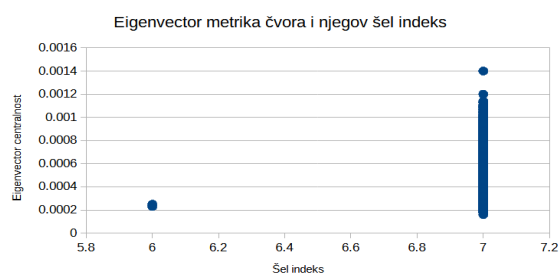
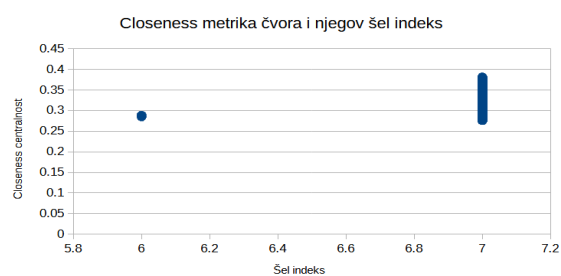
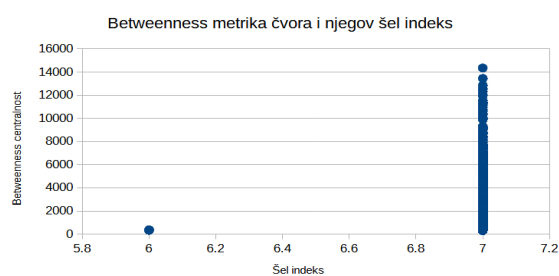
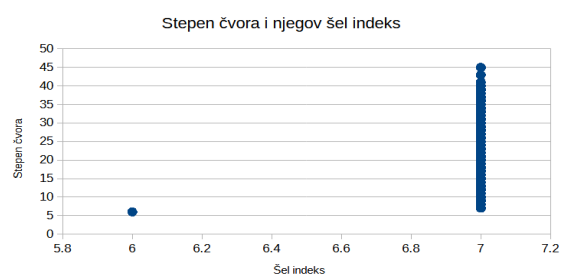
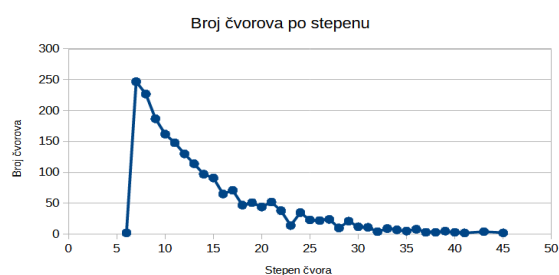




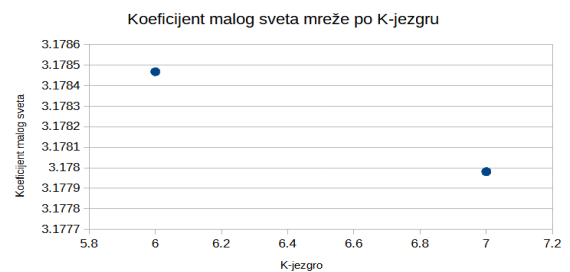
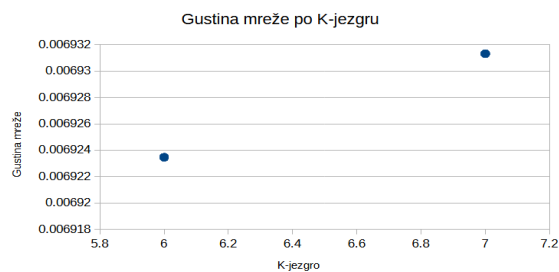
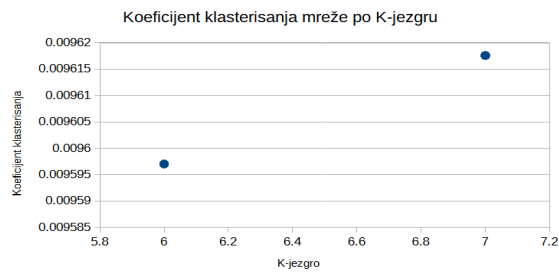
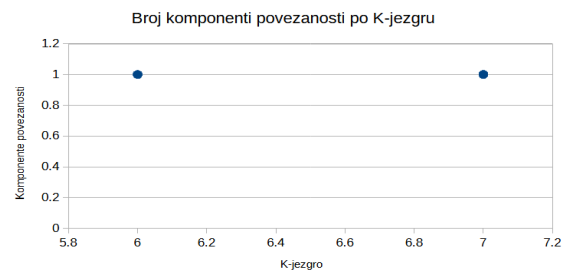
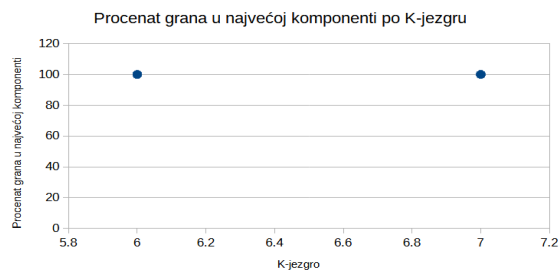
## 4.2 Barabas-Albert model

- Broj čvorova u mreži: 2000
- Broj grana u mreži: 13840
- Procenat čvorova u najvećoj komponenti: 100%
- Procenat grana u najvećoj komponenti: 100%
- Gustina mreže: 0,01

- Broj komponenti povezanosti : 1
- Dijametar mreže: 5
- Koeficijent klasterisanja mreže: 0,009
- Koeficijent malog sveta: 3,178
- Spermanov indeks korelacije za šel indeks i stepen povezanosti: 0,054
- Spermanov indeks korelacije za šel indeks i „Betweenness” metriku: 0,053
- Spermanov indeks korelacije za šel indeks i „Closeness” metriku: 0,051
- Spermanov indeks korelacije za šel indeks i „Eigenvector” metriku: 0,050



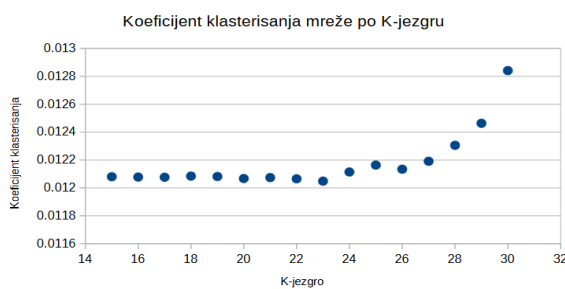
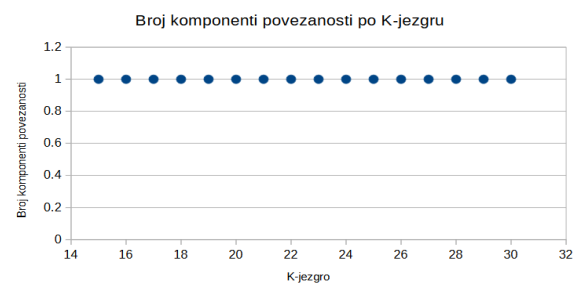
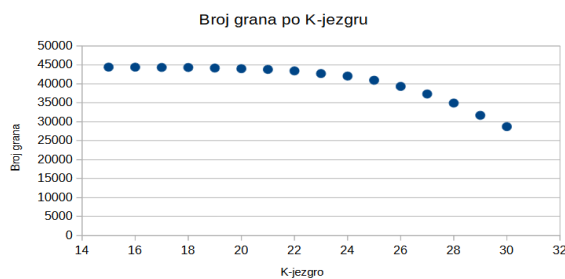
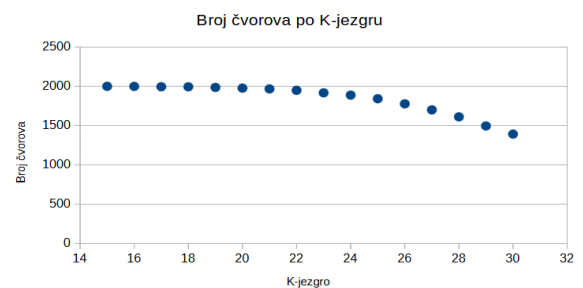
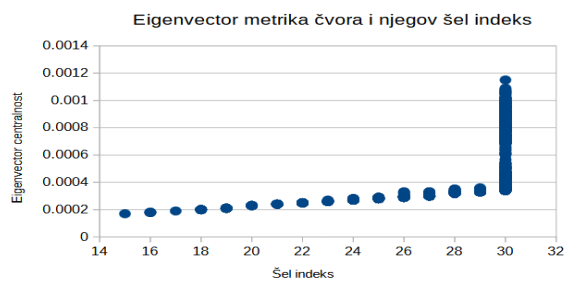
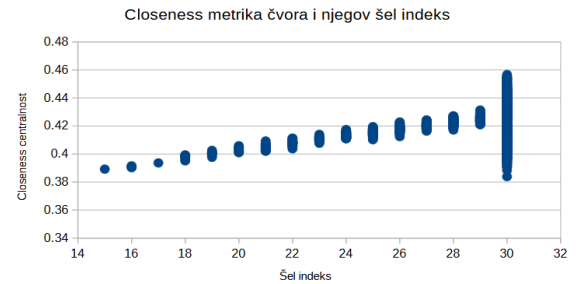
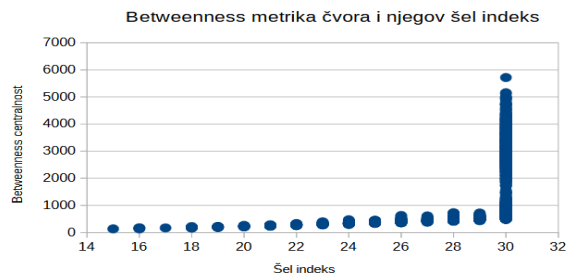
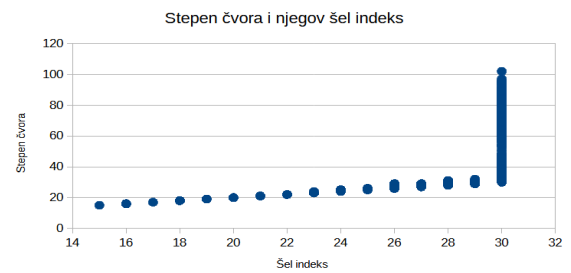
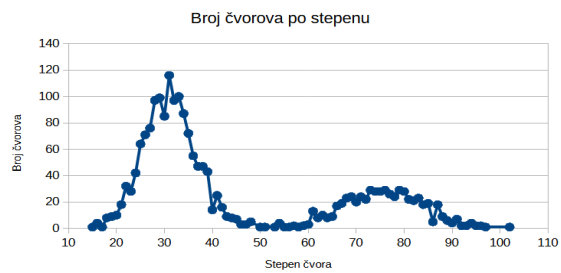


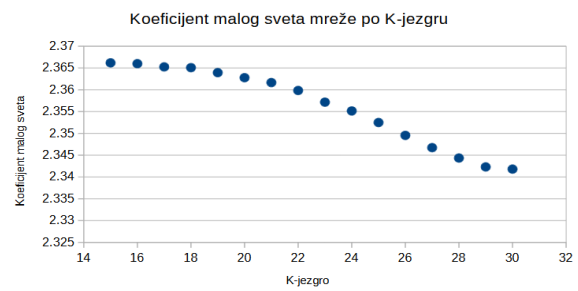
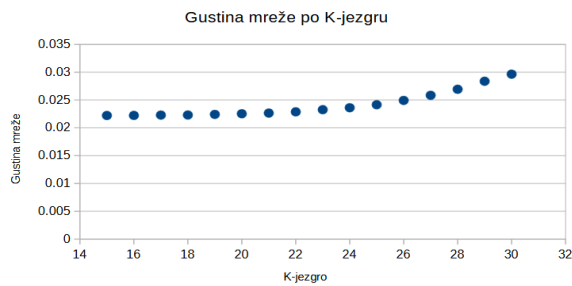


## 4.3 Jezgro-periferija model

### 4.3.1 Mreža 1

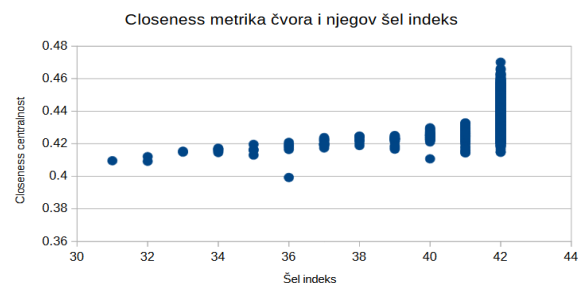
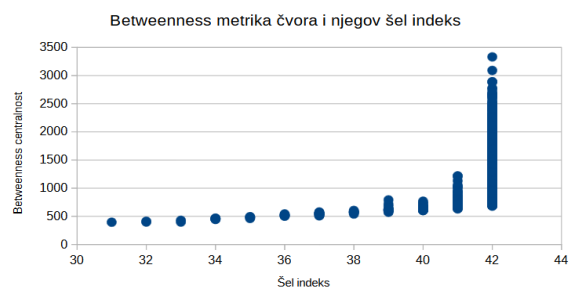
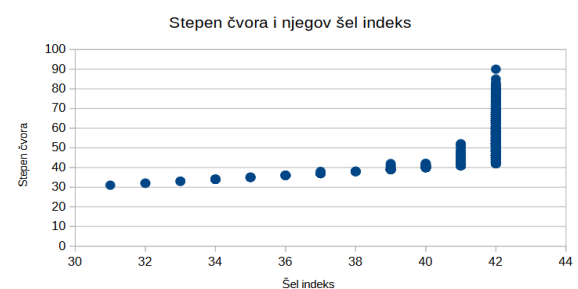
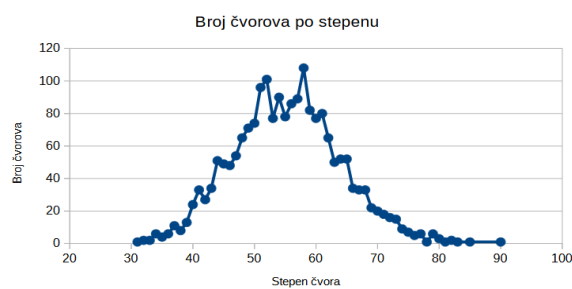
- Broj čvorova u mreži: 2000
- Procentat čvorova u jezgru: 30%
- Broj grana u mreži: 44412
- Procentat čvorova u najvećoj komponenti: 100%
- Procentat grana u najvećoj komponenti: 100%
- Gustina mreže: 0,022
- Broj komponenti povezanosti : 1
- Dijametar mreže: 4
- Koeficijent klasterisanja mreže: 0,012
- Koeficijent malog sveta: 2,366
- Spermanov indeks korelacije za šel indeks i stepen povezanosti: 0,809
- Spermanov indeks korelacije za šel indeks i „Betweenness” metriku: 0,791
- Spermanov indeks korelacije za šel indeks i „Closeness” metriku: 0,374
- Spermanov indeks korelacije za šel indeks i „Eigenvector” metriku: 0,811

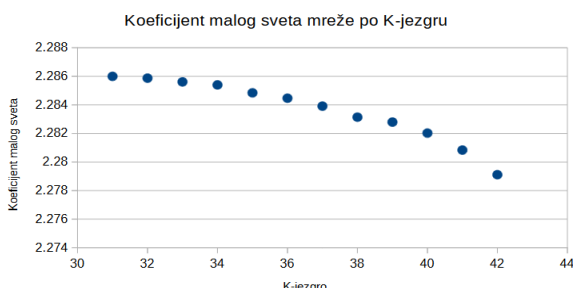
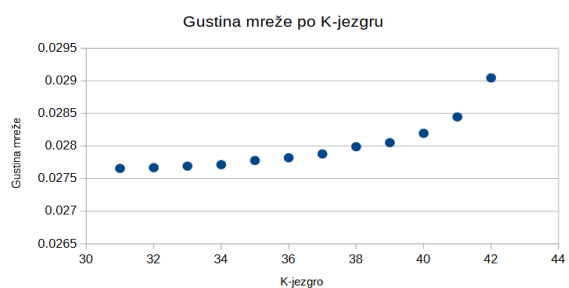
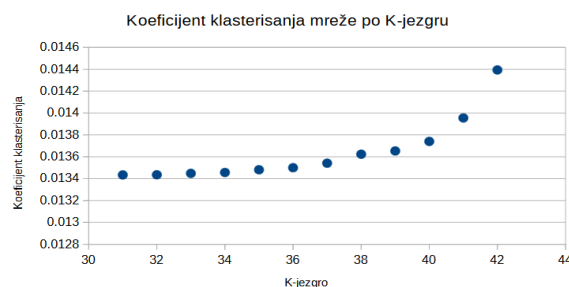
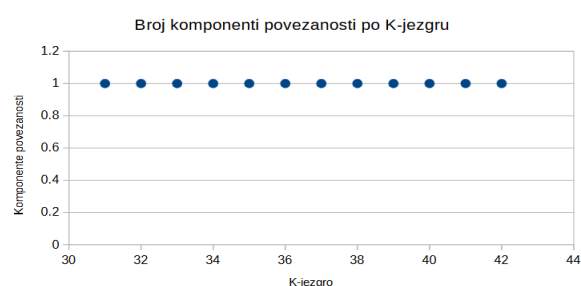
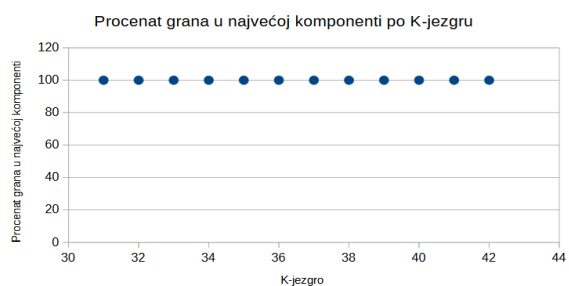
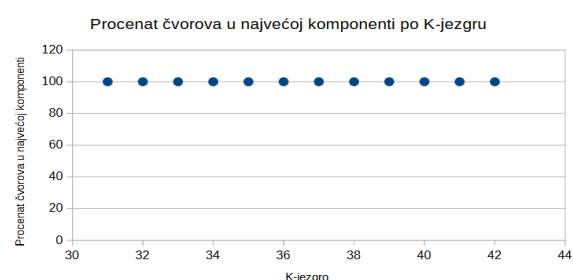
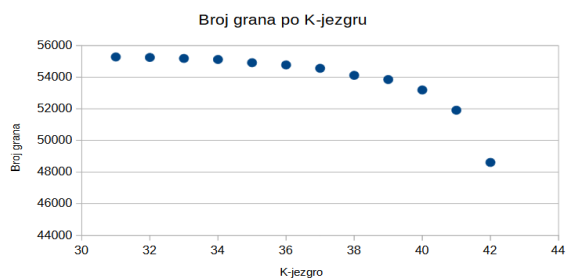
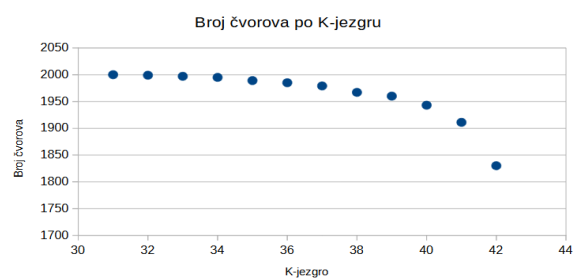
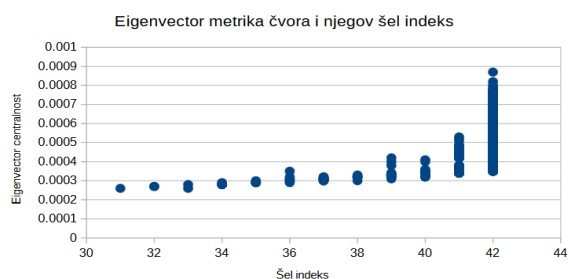




#### 4.3.2 Mreža 2

- Broj čvorova u mreži: 2000
- Procenat čvorova u jezgru: 50%
- Broj grana u mreži: 55285
- Procenat čvorova u najvećoj komponenti: 100%
- Procenat grana u najvećoj komponenti: 100%
- Gustina mreže: 0,027
- Broj komponenti povezanosti : 1
- Dijametar mreže: 3
- Koeficijent klasterisanja mreže: 0,013
- Koeficijent malog sveta: 2,285
- Spermanov indeks korelacije za šel indeks i stepen povezanosti: 0,468
- Spermanov indeks korelacije za šel indeks i „Betweenness” metriku: 0,460
- Spermanov indeks korelacije za šel indeks i „Closeness” metriku: 0,443
- Spermanov indeks korelacije za šel indeks i „Eigenvector” metriku: 0,433

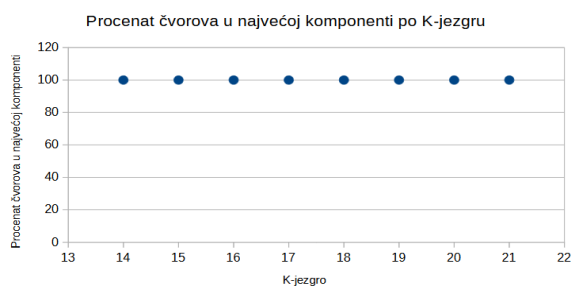
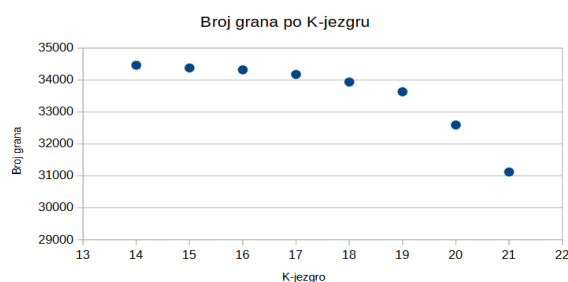
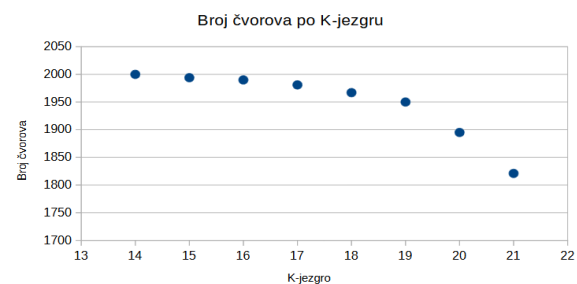
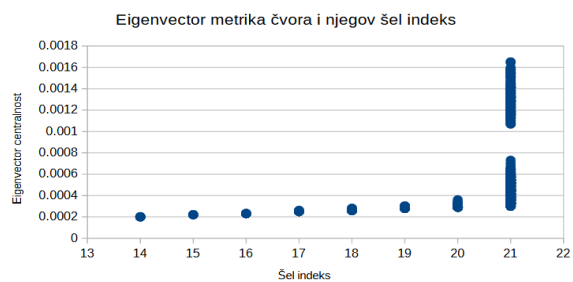
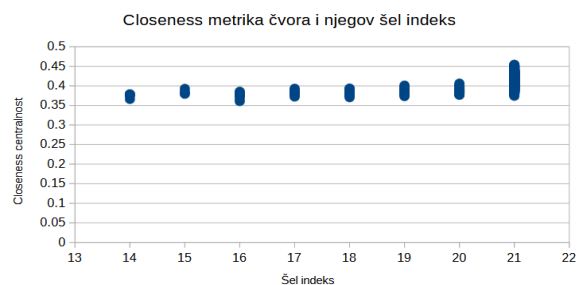
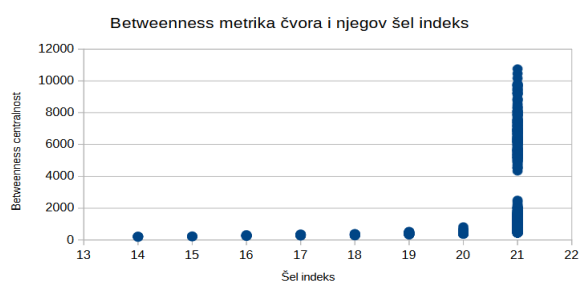
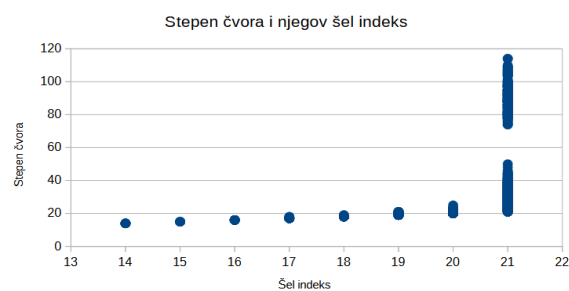
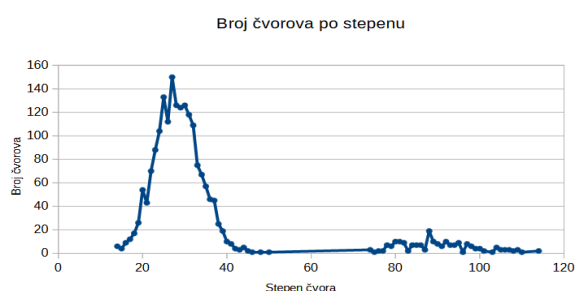


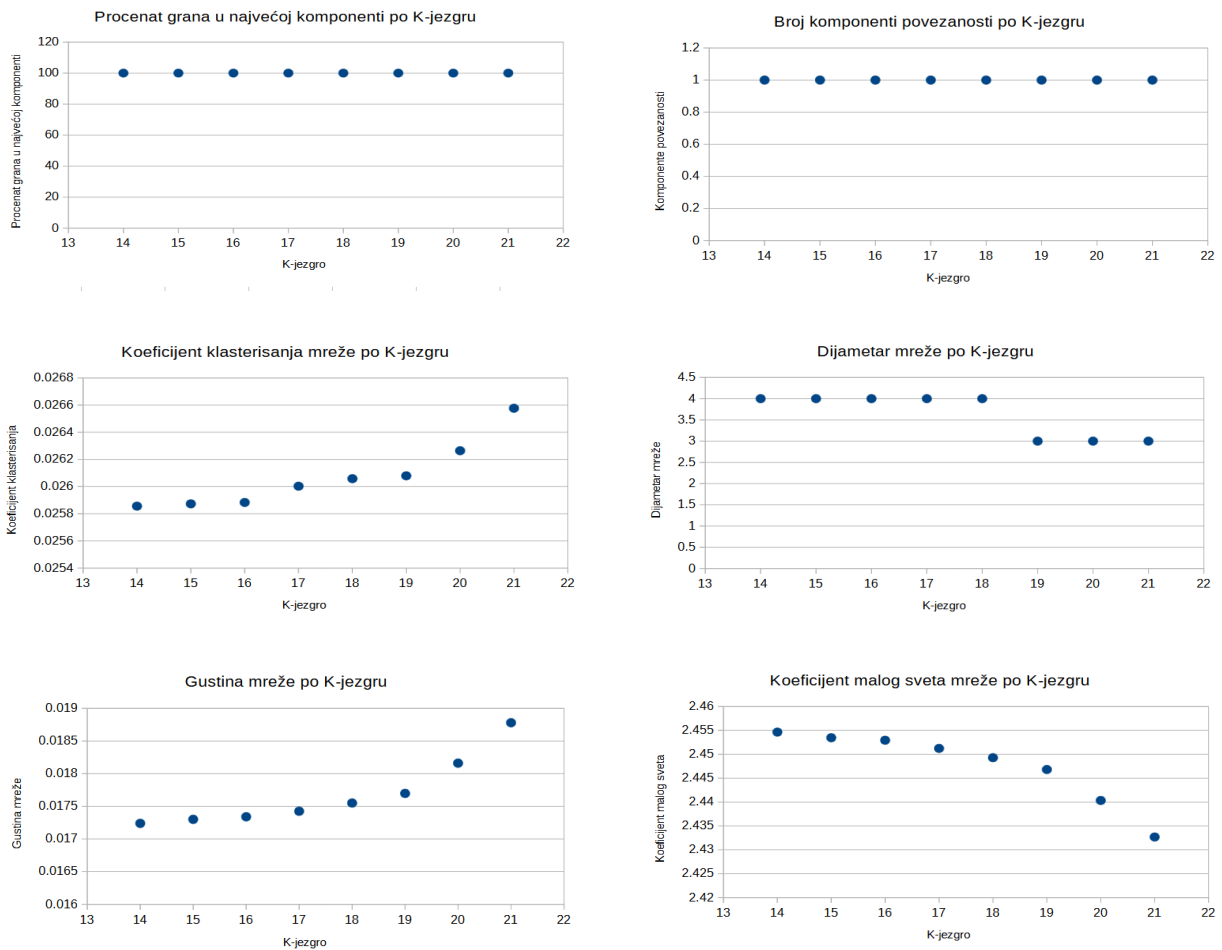


#### 4.3.3 Mreža 3

- Broj čvorova u mreži: 2000
- Procenat čvorova u jezgru: 90%
- Broj grana u mreži: 34461
- Procenat čvorova u najvećoj komponenti: 100%
- Procenat grana u najvećoj komponenti: 100%

- Gustina mreže: 0,017
- Broj komponenti povezanosti : 1
- Dijametar mreže: 3
- Koeficijent klasterisanja mreže: 0,025
- Koeficijent malog sveta: 2,454
- Spermanov indeks korelacije za šel indeks i stepen povezanosti: 0,490
- Spermanov indeks korelacije za šel indeks i „Betweenness” metriku: 0,482
- Spermanov indeks korelacije za šel indeks i „Closeness” metriku: 0,447
- Spermanov indeks korelacije za šel indeks i „Eigenvector” metriku: 0,490

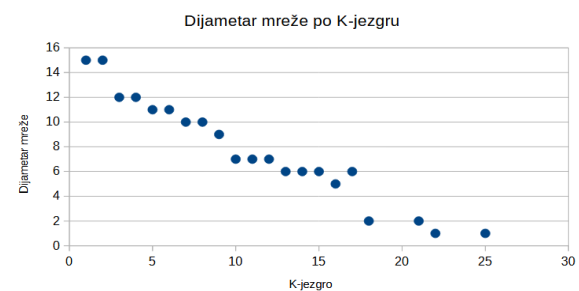
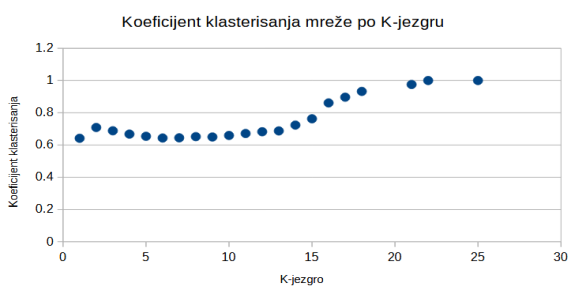
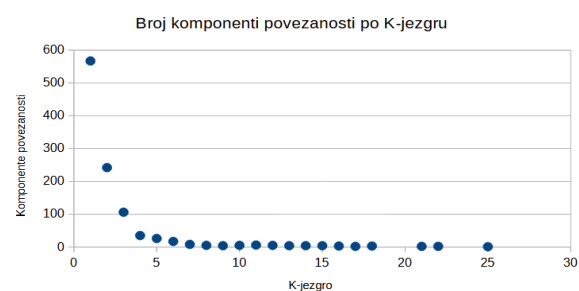
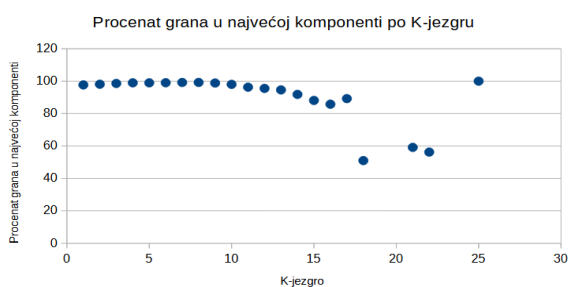
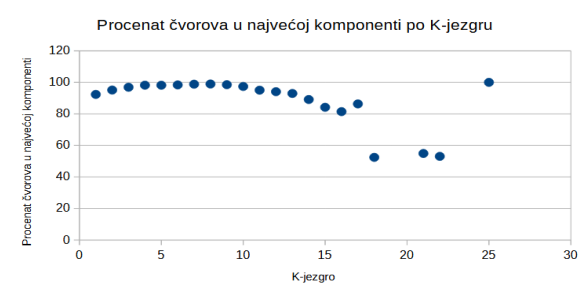
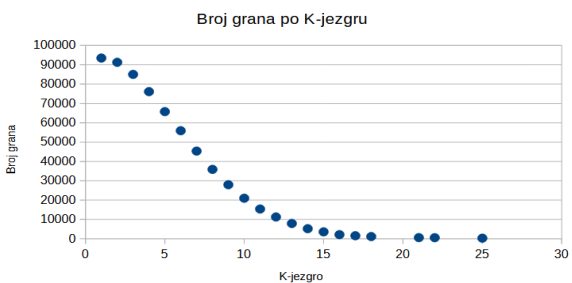
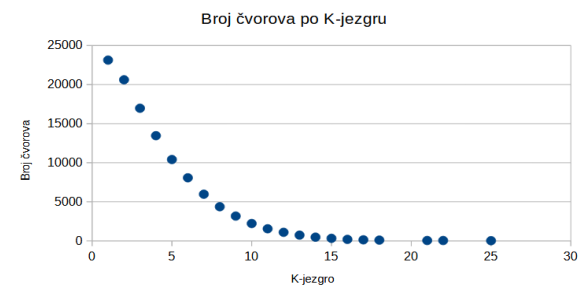
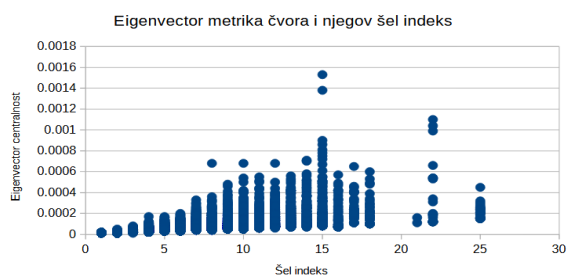
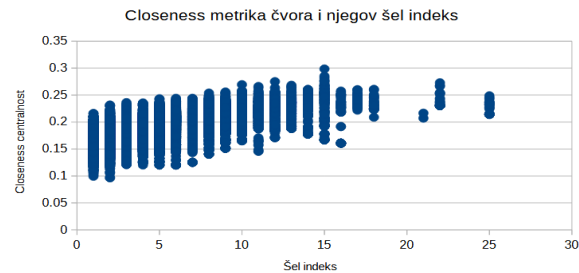
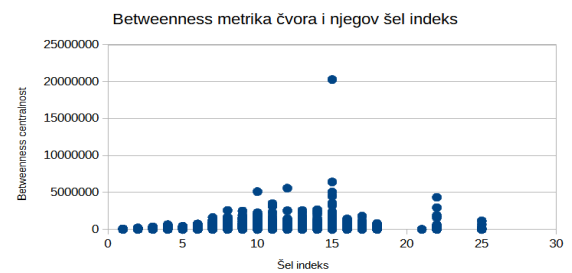
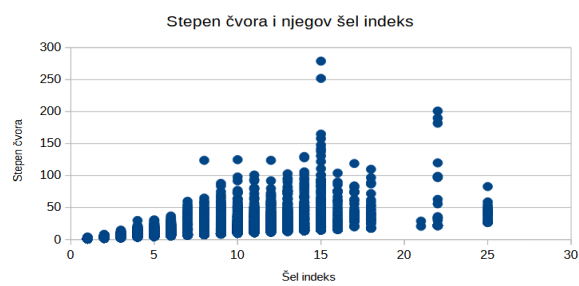
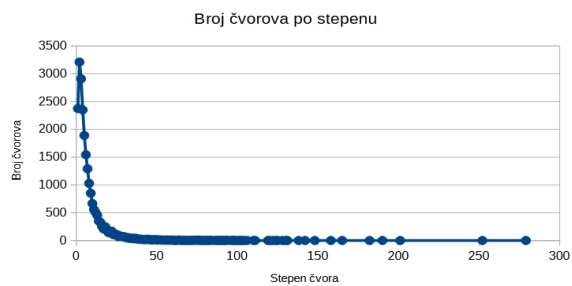


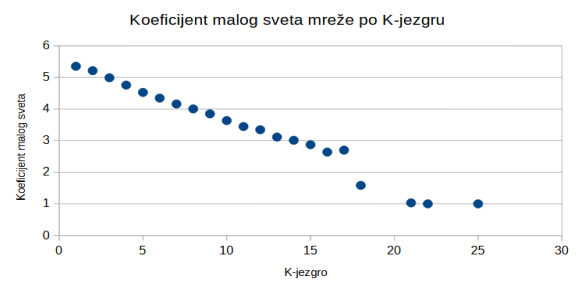
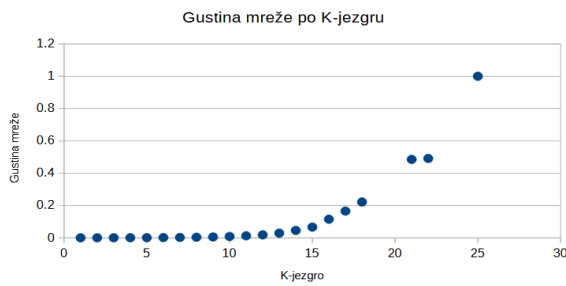


#### 4.4 Condensed matter collaboration 2003

Mreža koautorstva naučnika koji su postavili preprint na temu: Kondenzovana materija, u periodu od januara 1993. godine do aprila 2003. godine.

- Broj čvorova u mreži: 23133
- Broj grana u mreži: 93439
- Procentat čvorova u najvećoj komponenti: 92,348%
- Procentat grana u najvećoj komponenti: 97,695%
- Gustina mreže:  $3,49 \cdot 10^{-4}$
- Broj komponenti povezanosti : 567
- Dijametar mreže: 15
- Koeficijent klasterisanja mreže: 0,641
- Koeficijent malog sveta: 5,352
- Spermanov indeks korelacije za šel indeks i stepen povezanosti: 0,932
- Spermanov indeks korelacije za šel indeks i „Betweenness” metriku: 0,482
- Spermanov indeks korelacije za šel indeks i „Closeness” metriku: 0,649
- Spermanov indeks korelacije za šel indeks i „Eigenvector” metriku: 0,923

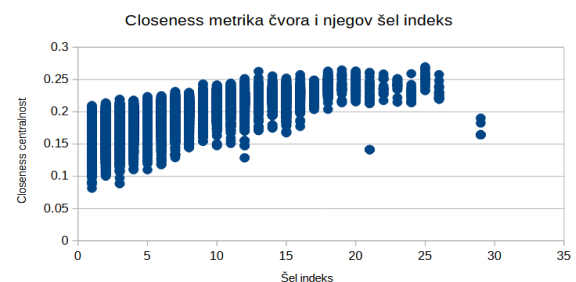
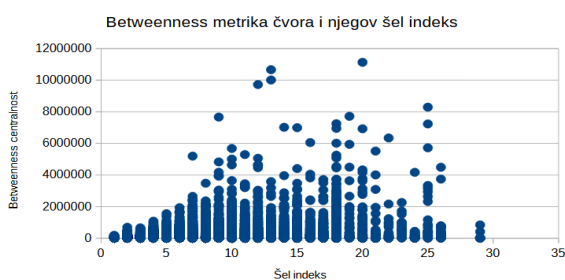
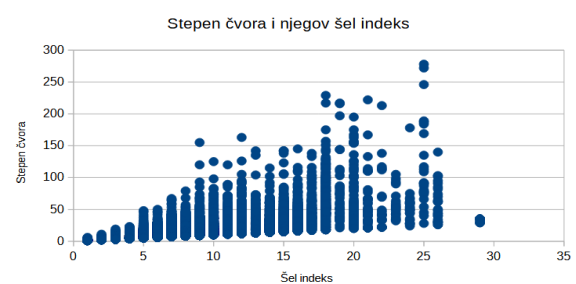
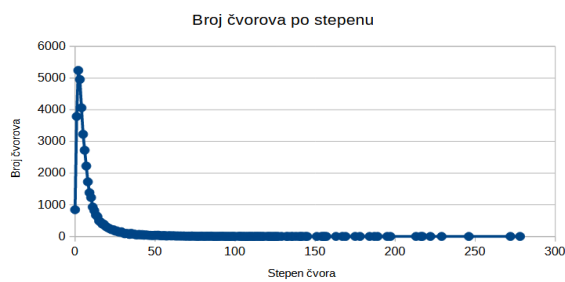




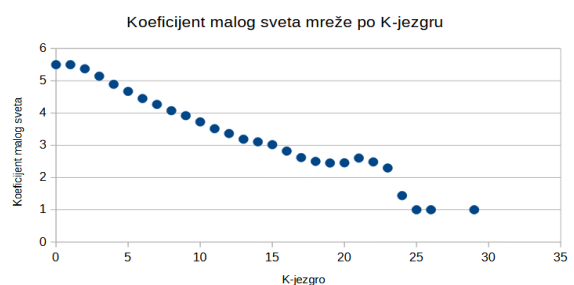
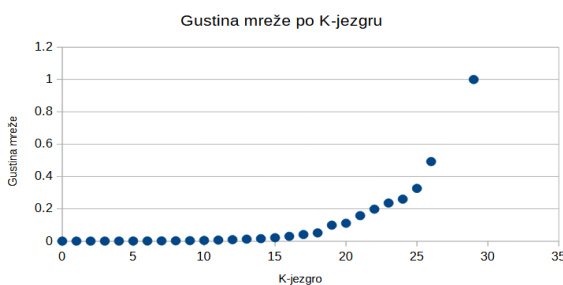
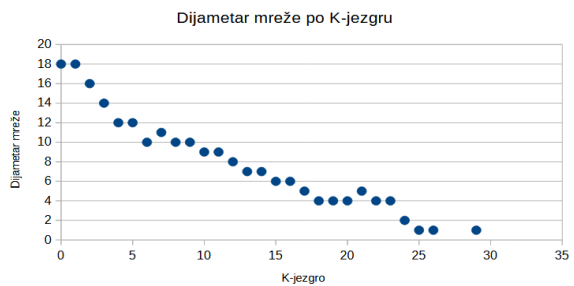
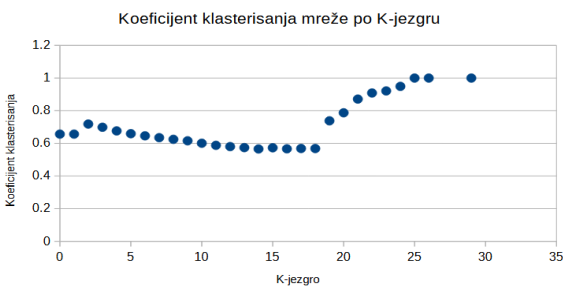
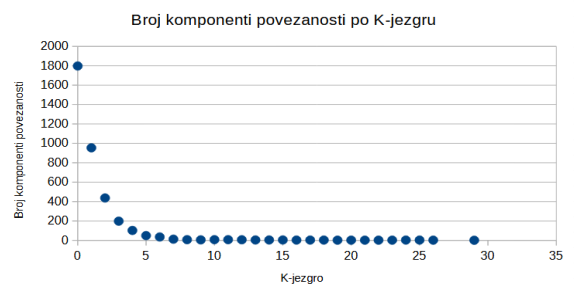
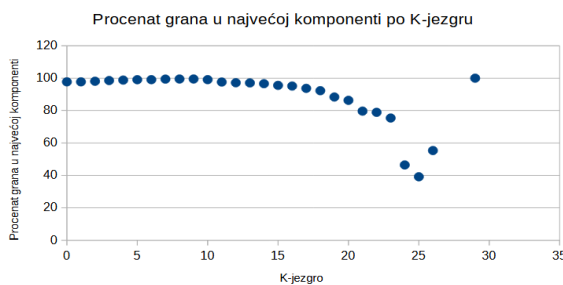
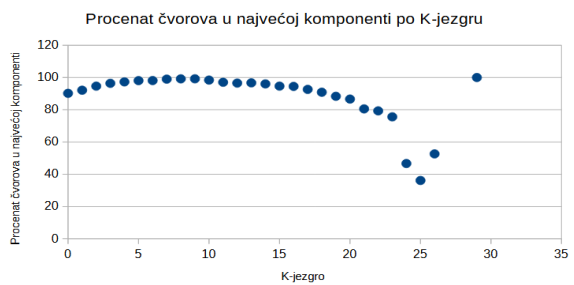
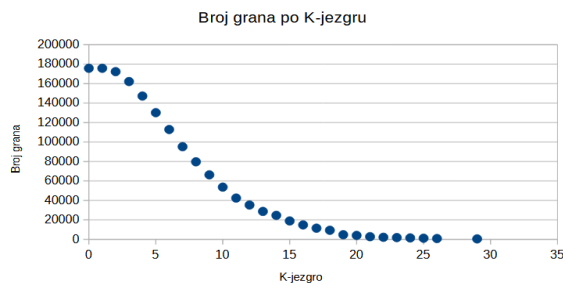
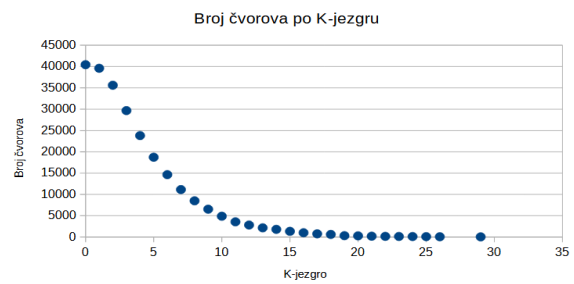
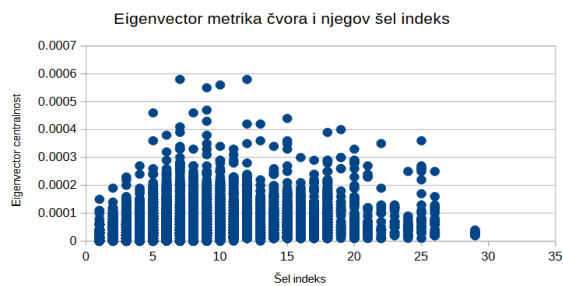
## 4.5 Condensed matter collaborations 2005

Mreža koautorstva naučnika koji su postavili preprint na temu: Kondenzovana materija, u periodu od januara 1995. godine do marta 2005. godine.

- Broj čvorova u mreži: 40421
- Broj grana u mreži: 175692
- Procenat čvorova u najvećoj komponenti: 90,195%
- Procenat grana u najvećoj komponenti: 97,747%
- Gustina mreže:  $2,15 \cdot 10^{-4}$
- Broj komponenti povezanosti : 1798
- Dijametar mreže: 18
- Koeficijent klasterisanja mreže: 0,656
- Koeficijent malog sveta: 5,499
- Spermanov indeks korelacije za šel indeks i stepen povezanosti: 0,937
- Spermanov indeks korelacije za šel indeks i „Betweenness” metriku: 0,497
- Spermanov indeks korelacije za šel indeks i „Closeness” metriku: 0,679
- Spermanov indeks korelacije za šel indeks i „Eigenvector” metriku: 0,558





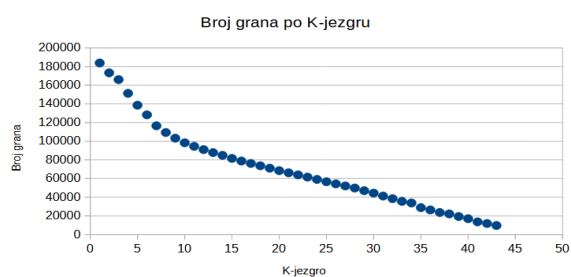
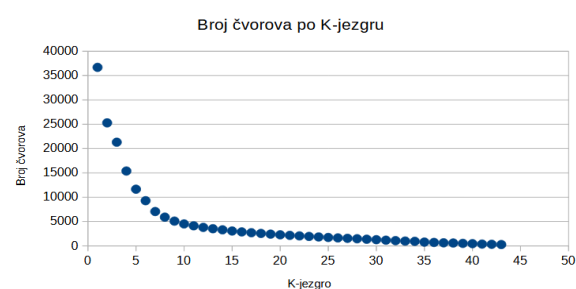
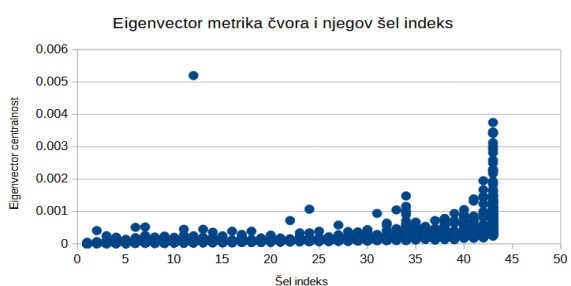
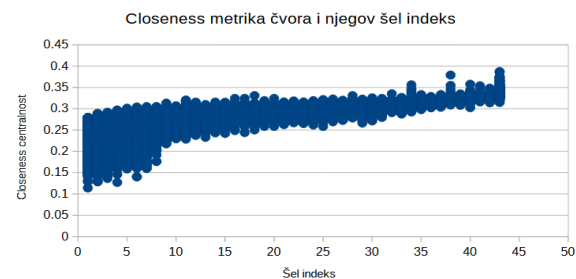
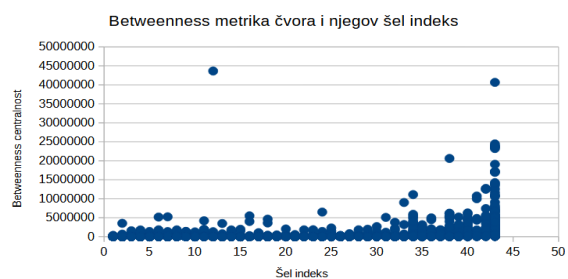
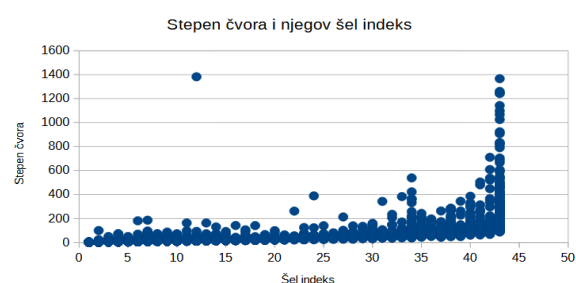
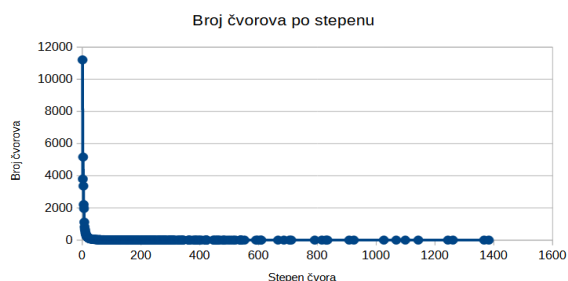


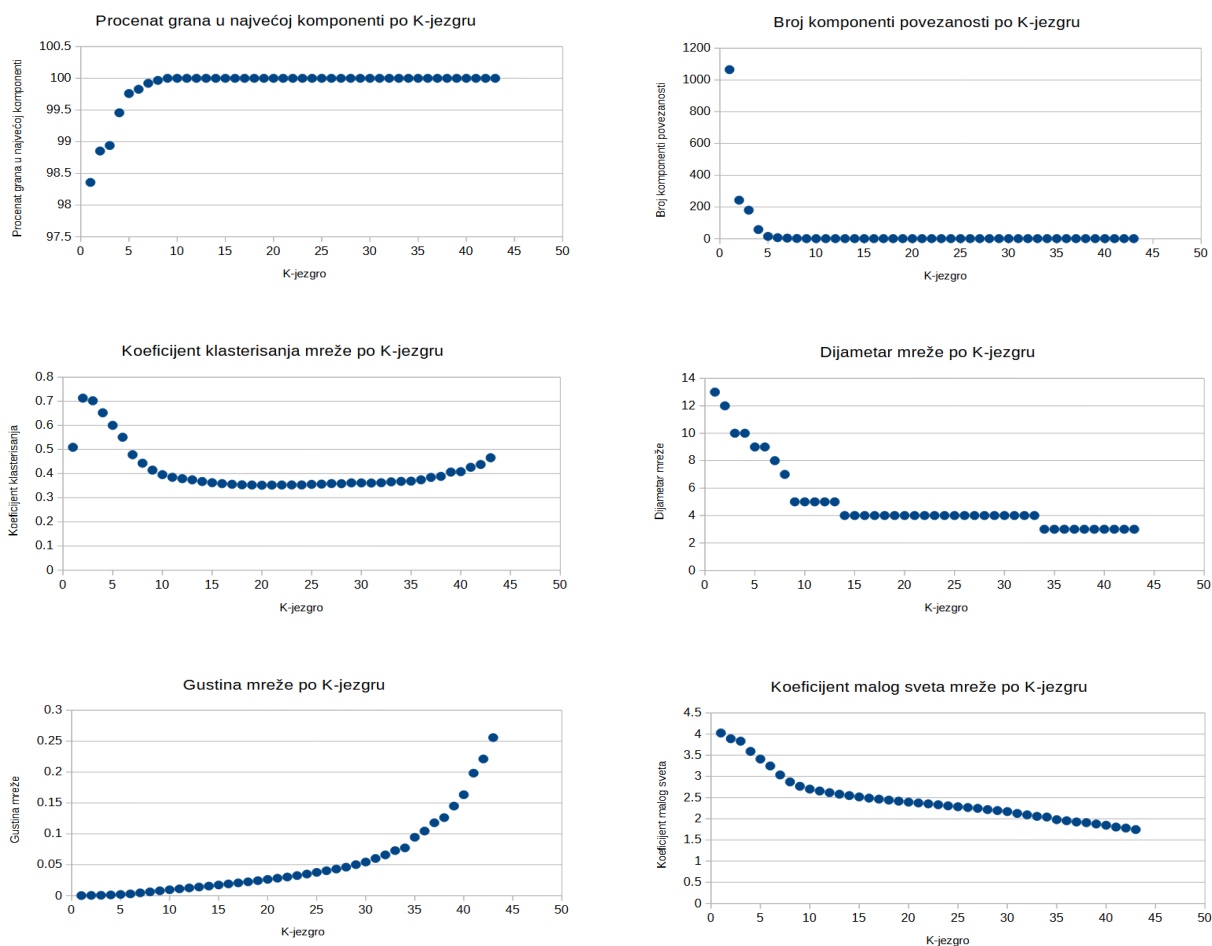
## 4.6 Enron email mreža

Mreža komunikacije između Enron mejlova. Jedan čvor predstavlja jednu mejl adresu, a čvorovi A i B su povezani ako je bilo koji od njih poslao mejl drugom.

- Broj čvorova u mreži: 36692
- Broj grana u mreži: 183831
- Procenat čvorova u najvećoj komponenti: 91,834%

- Procenat grana u najvećoj komponenti: 98,357%
- Gustina mreže:  $2,73 \cdot 10^{-4}$
- Broj komponenti povezanosti : 1065
- Dijametar mreže: 13
- Koeficijent klasterisanja mreže: 0,509
- Koeficijent malog sveta: 4,025
- Spermanov indeks korelacije za šel indeks i stepen povezanosti: 0,983
- Spermanov indeks korelacije za šel indeks i „Betweenness” metriku: 0,693
- Spermanov indeks korelacije za šel indeks i „Closeness” metriku: 0,495
- Spermanov indeks korelacije za šel indeks i „Eigenvector” metriku: 0,964

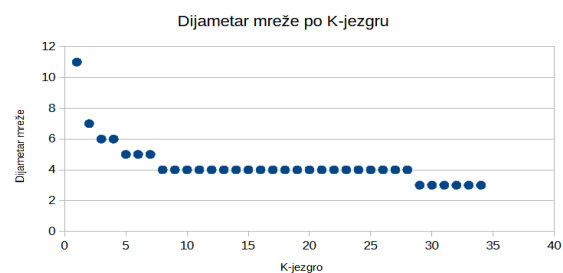
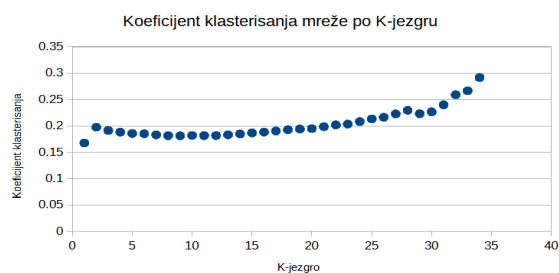
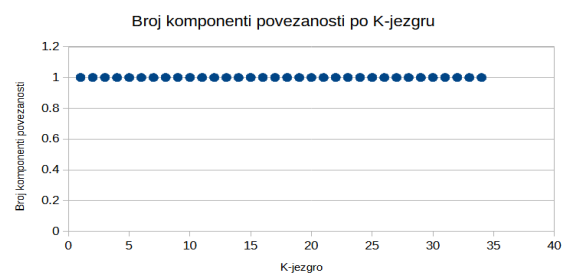
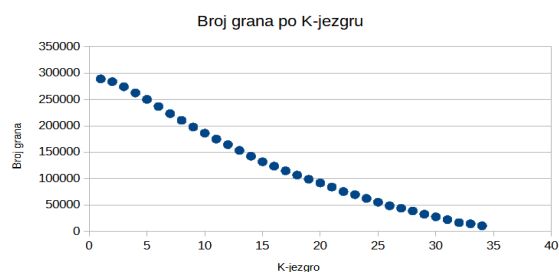
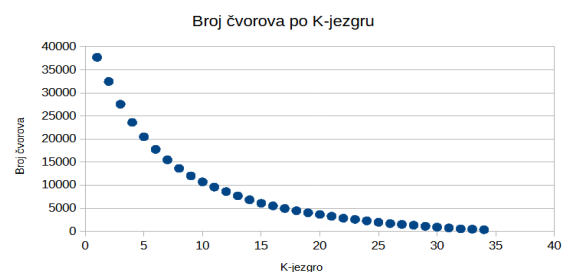
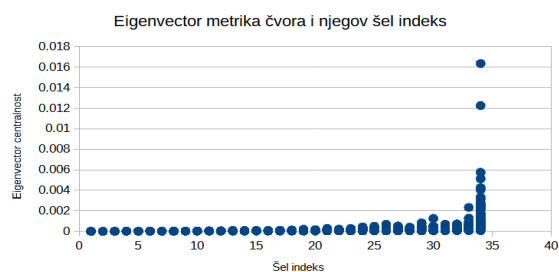
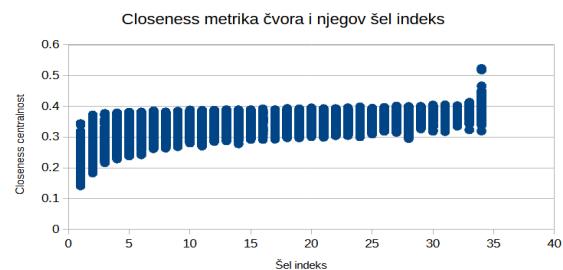
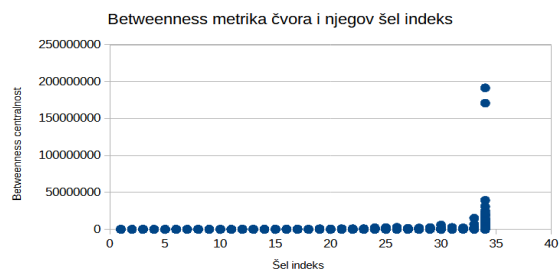
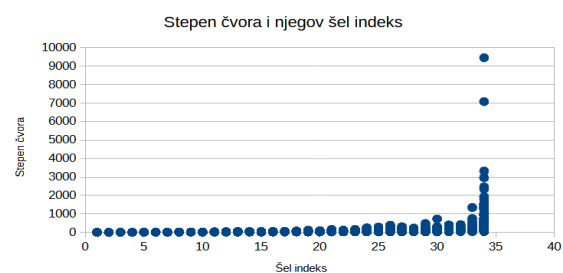
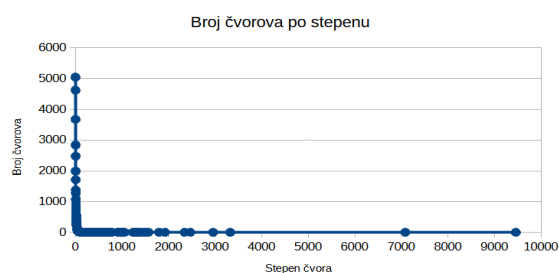


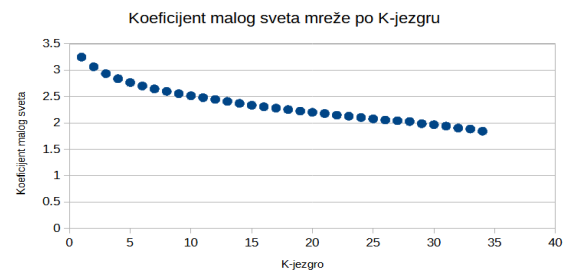
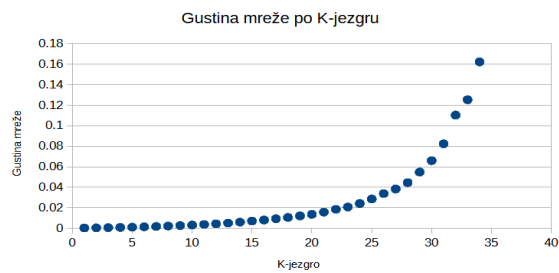


## 4.7 Github mreža

Mreža programera sa Github mreže. Čvorovi su programeri koji imaju barem 10 repozitorijuma, a veza između programera postoji ako se uzajamno prate.

- Broj čvorova u mreži: 37700
- Broj grana u mreži: 289003
- Procenat čvorova u najvećoj komponenti: 100%
- Procenat grana u najvećoj komponenti: 100%
- Gustina mreže:  $4,06 \cdot 10^{-4}$
- Broj komponenti povezanosti : 1
- Dijametar mreže: 11
- Koeficijent klasterisanja mreže: 0,167
- Koeficijent malog sveta: 3,246
- Spermanov indeks korelacije za šel indeks i stepen povezanosti: 0,990
- Spermanov indeks korelacije za šel indeks i „Betweenness” metriku: 0,818
- Spermanov indeks korelacije za šel indeks i „Closeness” metriku: 0,729
- Spermanov indeks korelacije za šel indeks i „Eigenvector” metriku: 0,988





## 5 Zaključak

U veštački generisanim mrežama korelacija između metrika centralnosti i šel indeksa čvora je uglavnom manja od 0,5, osim u slučaju mreže generisane po modelu jezgro-periferija gde je procenat čvorova u jezgru 30%, gde imamo visoke korelacije. U realnim mrežama uglavnom imamo visoke korelacije sa retkim primerima gde je manja od 0,6.

U svim mrežam vidimo da promena metrika koje pratimo kroz k-jezgra je uglavnom ista, gustina mreže raste, koeficijent malog sveta opada, koeficijent klasterisanja uglavnom raste...

## 6 Literatura

- [1] Leskovec, J., Kleinberg, J., Faloutsos, C. *Graph evolution: Densification and shrinking diameters*. *ACM Transactions on Knowledge Discovery from Data*, 2007.
- [2] M. E. J. Newman, *Proc. Natl. Acad. Sci. USA*, *The structure of scientific collaboration networks*, 2001.
- [3] J. Leskovec, K. Lang, A. Dasgupta, M. Mahoney, *Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters*, 2009.
- [4] B. Klimmt, Y. Yang, *Introducing the Enron corpus*, 2004.
- [5] B. Rozemberczki, C. Allen and R. Sarkar, *Multi-scale Attributed Node Embedding*, 2019.