



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



# **Implementacija i analiza K-core dekompozicije u neusmerenim socijalnim mrežama**

Praktični projekat iz predmeta Socijalne mreže

Nikolina Puač 267/21

Novi Sad, 2024.

# Sadržaj

|   |           |
|---|-----------|
| <b>1 Uvod.....</b>  | <b>2</b>  |
| <b>2 Ručno generisane mreže.....</b>  | <b>3</b>  |
| 2.1 Mreža 1.....  | 3         |
| 2.2 Mreža 2.....  | 4         |
| 2.3 Mreža 3.....  | 5         |
| <b>3 Veštački generisane mreže.....</b>   | <b>6</b>  |
| <b>3.1 Erdos-Renyi Model.....</b>   | <b>6</b>  |
| 3.1.1 Korelacija šel indeksa i stepena čvora i globalnih metrika centralnosti.....            | 7         |
| 3.1.2 Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora..... | 7         |
| <b>3.2 Barabasi-Albert Model.....</b>   | <b>8</b>  |
| 3.2.1 Korelacija šel indeksa i stepena čvora i globalnih metrika centralnosti.....            | 8         |
| 3.2.2 Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora..... | 9         |
| <b>3.3 Core-PeripheryModel.....</b>   | <b>9</b>  |
| 3.3.1 Korelacija šel indeksa i stepena čvora i globalnih metrika centralnosti.....            | 10        |
| 3.3.2 Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora..... | 10        |
| <b>4 Realne kompleksne mreže.....</b>   | <b>11</b> |
| <b>4.1 CA-HepPH mreža.....</b>  | <b>11</b> |
| 4.1.1 Korelacija šel indeksa i stepena čvora i globalnih metrika centralnosti.....            | 11        |
| 4.1.2 Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora..... | 12        |
| <b>4.2 CA-CondMat mreža.....</b>  | <b>12</b> |
| 4.2.1 Korelacija šel indeksa i stepena čvora i globalnih metrika centralnosti.....            | 12        |
| 4.2.2 Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora..... | 13        |
| <b>4.3 Email-Enron mreža.....</b>   | <b>13</b> |
| 4.3.1 Korelacija šel indeksa i stepena čvora i globalnih metrika centralnosti.....            | 13        |
| 4.3.2 Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora..... | 14        |
| <b>5 Zaključak.....</b>   | <b>14</b> |
| <b>6 Literatura.....</b>  | <b>15</b> |

# 1 Uvod

Cilj ovog projekta je implementacija Batagelj-Zaveršnik algoritma za dekompoziciju grafa i poređenje njegove tačnosti sa Straightforward algoritmom.

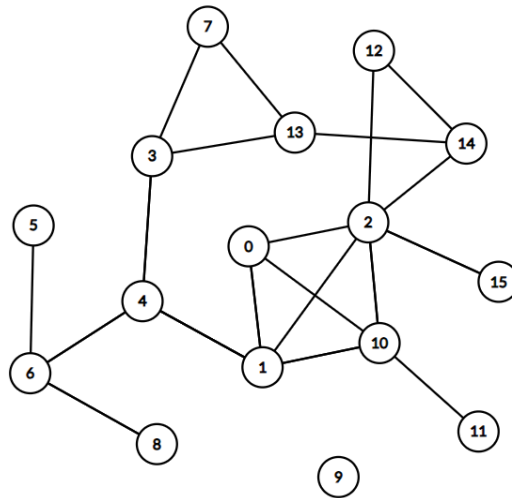
Nakon što je potvrđeno da su oba algoritma tačna i da im se rezultati poklapaju, upotrebljeni su za analizu tri mreže koje su generisane po modelu i za tri realne mreže.

Takođe, izračunata je korelacija između shell indexa i metrika betweenness, closeness i eigenvector. Rezultati korelacije su grafički prikazani za sve mreže.

## 2 Ručno generisane mreže

Za potrebe provere tačnosti implementiranog Batagelj-Zaveršnik algoritma, kreirane su tri male proizvoljne mreže koje su generisane pomoću *SmallGraphs* klase. Rezultati će dalje u tekstu biti prikazani kao šel indeks - svi čvorovi sa tim indeksom.

### 2.1 Mreža 1



Slika 2.1.1: Ručno napravljena mreža od 16 čvorova

#### **Batagelj-Zaveršnik**

Šel indeks 3 - 0, 1, 2, 10

Šel indeks 2 - 3, 4, 7, 12, 13, 14

Šel indeks 1 - 5, 6, 8, 11, 15

Šel indeks 0 - 9

#### **StraightForward**

Šel indeks 3 - 0, 1, 2, 10

Šel indeks 2 - 3, 4, 7, 12, 13, 14

Šel indeks 1 - 5, 6, 8, 11, 15

Šel indeks 0 - 9

#### **Ručno računanje**

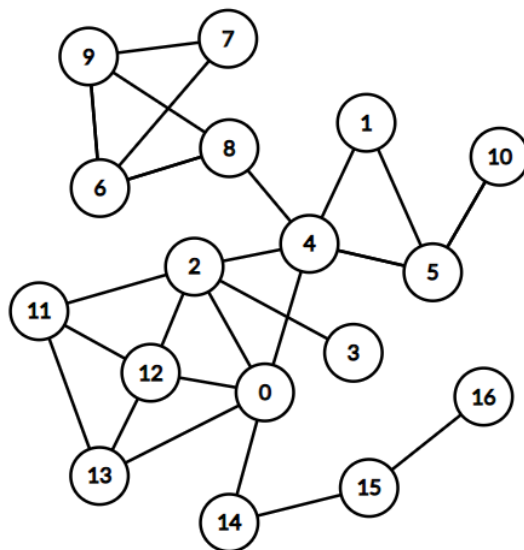
Šel indeks 3 - 0, 1, 2, 10

Šel indeks 2 - 3, 4, 7, 12, 13, 14

Šel indeks 1 - 5, 6, 8, 11, 15

Šel indeks 0 - 9

## 2.2 Mreža 2



Slika 2.2.1: Ručno napravljena mreža od 17 čvorova

### **Batagelj-Zaveršnik**

Šel indeks 3 - 0, 2, 11, 12, 13

Šel indeks 2 - 1, 4, 5, 6, 7, 8, 9

Šel indeks 1 - 3, 10, 14, 15, 16

### **StraightForward**

Šel indeks 3 - 0, 2, 11, 12, 13

Šel indeks 2 - 1, 4, 5, 6, 7, 8, 9

Šel indeks 1 - 3, 10, 14, 15, 16

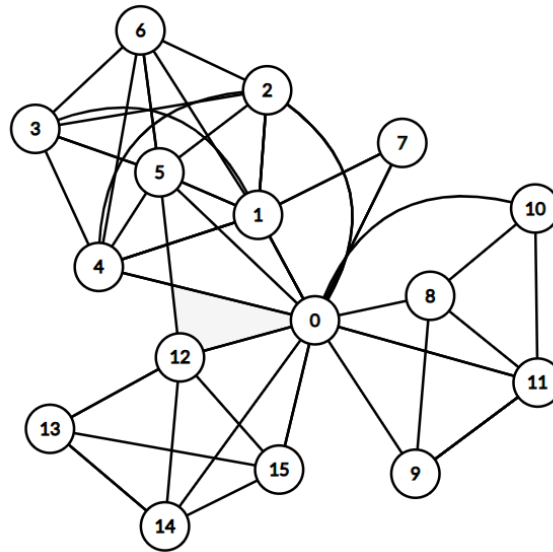
### **Ručno računanje**

Šel indeks 3 - 0, 2, 11, 12, 13

Šel indeks 2 - 1, 4, 5, 6, 7, 8, 9

Šel indeks 1 - 3, 10, 14, 15, 16

## 2.3 Mreža 3



*Slika 2.3.1: Ručno napravljena mreža od 16 čvorova*

### **Batagelj-Zaveršnik**

Šel indeks 5 - 1, 2, 3, 4, 5, 6

Šel indeks 4 - 0

Šel indeks 3 - 8, 9, 10, 11, 12, 13, 14, 15

Šel indeks 2 - 7

### **StraightForward**

Šel indeks 5 - 1, 2, 3, 4, 5, 6

Šel indeks 4 - 0

Šel indeks 3 - 8, 9, 10, 11, 12, 13, 14, 15

Šel indeks 2 - 7

### **Ručno računanje**

Šel indeks 5 - 1, 2, 3, 4, 5, 6

Šel indeks 4 - 0

Šel indeks 3 - 8, 9, 10, 11, 12, 13, 14, 15

Šel indeks 2 - 7

U implementaciji Batagelj-Zaveršnik algoritma, dodata je funkcija koja omogućava korisniku da generiše k-core podgraf za bilo koji k iz originalnog grafa. Funkcija se zove kCore i kao parametar prima k.

### 3 Veštački generisane mreže

Podaci koji će biti prikazani odnose se na mrežu koja je nasumično odabrana iz skupa mreža generisanih za svaki model. Nakon što je mreža generisana pozvana je metoda *exportGraph* koja ju je sačuvala u poseban csv fajl radi lakše provere rezultata i grafičkog prikaza, zato što svaki put kada se mreža generiše rezultati su statistički različiti zbog nasumične prirode modela.

Metrike su dobijene korišćenjem klase *GraphMetrics*.

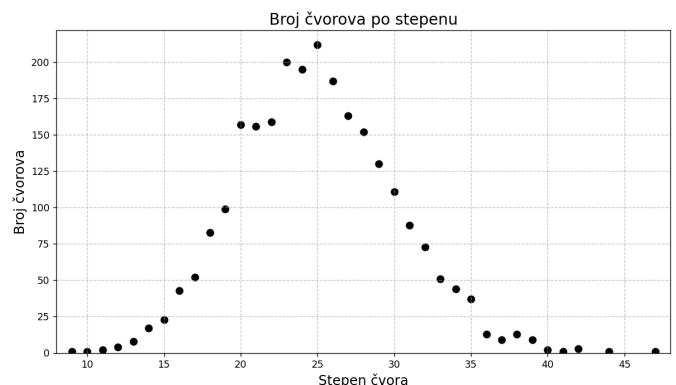
Kako bi se lakše proverilo da li su rezultati algoritama identični napravljena je metoda *compareAlgorithms* koja proverava da li su dobijeni shell indeksi za svaki čvor isti, ukoliko jesu ispisuje da je rezultat isti, ukoliko ne ispisuje gde se nalazi razlika.

#### 3.1 Erdos-Renyi Model

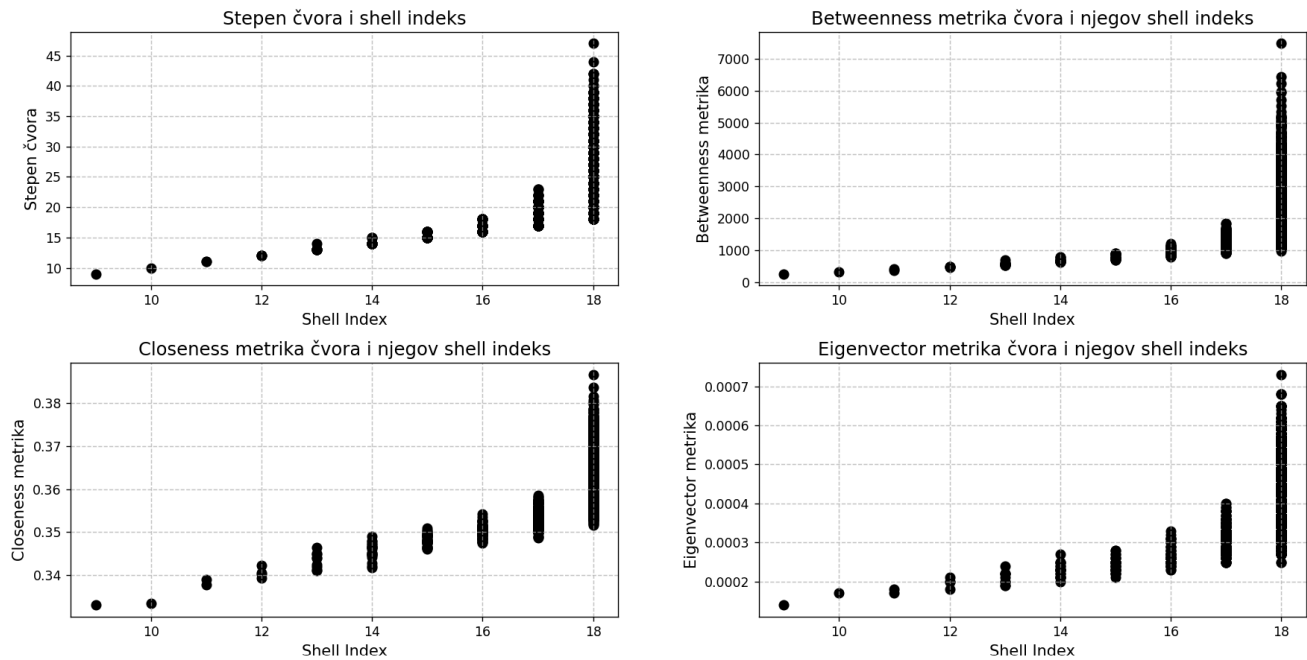
Erdos-Renyi model je model za generisanje slučajnih grafova. U ovom modelu, graf se gradi tako što se za svaki mogući par čvorova u mreži sa verovatnoćom  $p$  odlučuje da li će između njih postojati grana ili ne. To znači da se na osnovu ove verovatnoće, grane nasumično dodaju između čvorova.

Parametri mreže su: 2500 čvorova i 0.01 verovatnoća

- Broj čvorova: 2500
- Broj grana: 31179
- Gustina mreže: 0.0099812725090036
- Minimalni stepen: 9
- Maksimalni stepen: 47
- Broj komponenti: 1
- Dijametar: 4
- Koeficijent malog sveta: 2.767269467787115
- Koeficijent klasterisanja: 0.010094496809065419



### 3.1.1 Korelacija šel indeksa i stepena čvora i globalnih metrika centralnosti



### 3.1.2 Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora

Shell indeks i stepena čvora: 0.569418130700864

Shell indeks i betweenness centralnosti: 0.5622441566136488

Shell indeks i closeness centralnosti: 0.5760740390648802

Shell indeks i eigenvector centralnosti: 0.5380484035736429

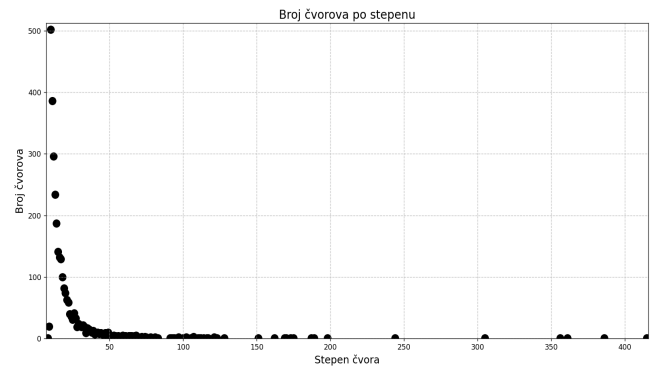


## 3.2 Barabasi-Albert Model

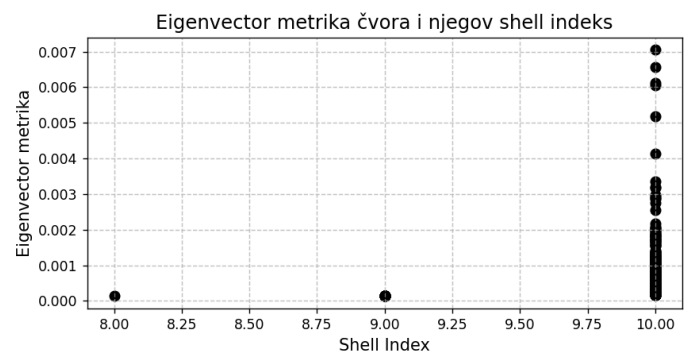
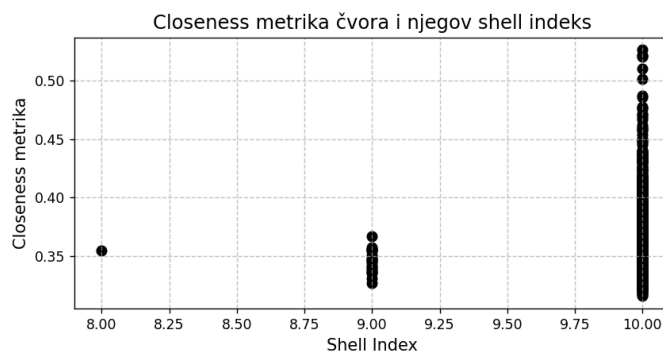
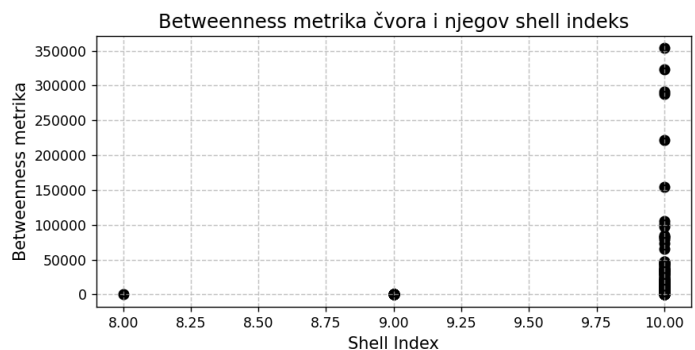
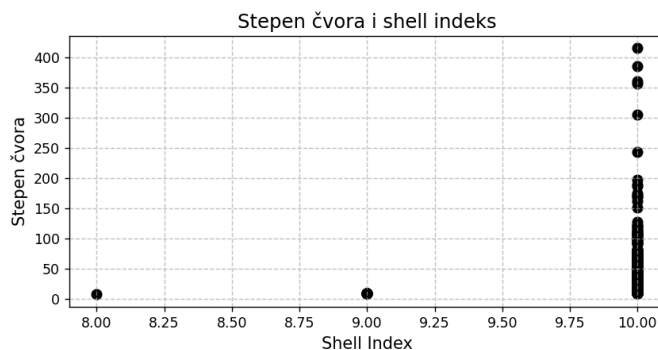
Barabási-Albert model počinje sa malim brojem čvorova i dodaje nove čvorove tako što svaki novi čvor preferencijalno povezuje sa već povezanim čvorovima.

Parametri mreže su: 3000 čvorova, 30 početnih čvorova, verovatnoća je 0.01 i svaki novi čvor se vezuje sa 10 drugih.

- Broj čvorova: 3000
- Broj grana: 29443
- Gustina mreže: 0.006575726606288112
- Minimalni stepen: 8
- Maksimalni stepen: 415
- Broj komponenti: 1
- Dijametar: 4
- Koeficijent malog sveta: 2.789719461274167
- Koeficijent klasterisanja: 0.030382583466189633



### 3.2.1 Korelacija šel indeksa i stepena čvora i globalnih metrika centralnosti



### 3.2.2 Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora

Shell indeks i stepena čvora: 0.14522294150153298

Shell indeks i betweenness centralnosti: 0.1184710455252832

Shell indeks i closeness centralnosti: 0.06420750088683404

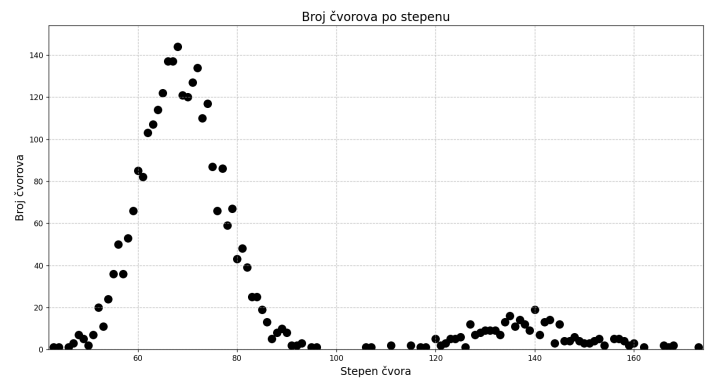
Shell indeks i eigenvector centralnosti: 0.1445825719575981

### 3.3 Core-PeripheryModel

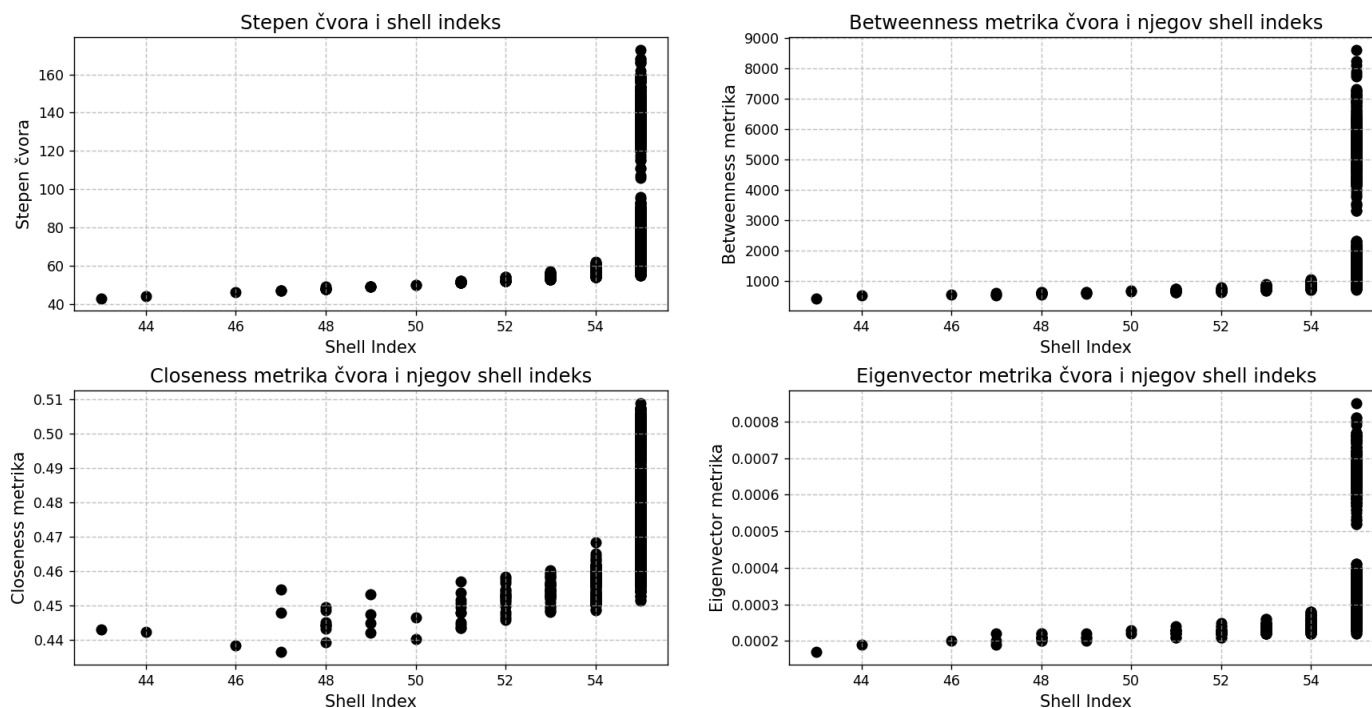
Jezgro-periferija model je modifikacija Planted partition modela i generiše mreže prosleđenog broja čvorova koje imaju strukturu jezgro-periferija.

Parametri mreže su: 3000 čvorova, 90% čvorova je u jezgru, verovatnoća vezivanja dva čvora iz jezgra je 0.02, iz periferije 0.01 i verovatnoća vezivanja jednog čvora iz jezgra a drugog čvora iz periferije 0.05.

- Broj čvorova: 3000
- Broj grana: 113556
- Gustina mreže: 0.025243081027009003
- Minimalni stepen: 43
- Maksimalni stepen: 173
- Broj komponenti: 1
- Dijametar: 3
- Koeficijent malog sveta: 2.1129434255863067
- Koeficijent klasterisanja: 0.02868543091774911



### 3.3.1 Korelacija šel indeksa i stepena čvora i globalnih metrika centralnosti



### 3.3.2 Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora

Shell indeks i stepena čvora: 0.42406845179414904

Shell indeks i betweenness centralnosti: 0.416081406297829

Shell indeks i closeness centralnosti: 0.4148203009009937

Shell indeks i eigenvector centralnosti: 0.4083840874177096

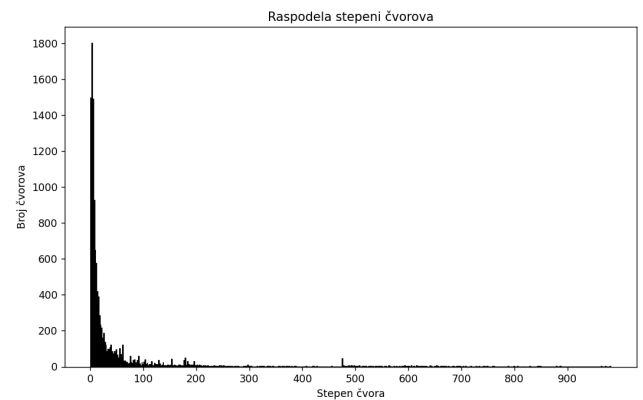
## 4 Realne kompleksne mreže

U ovom delu će biti prikazane metrike nad realnim neusmerenim mrežama preuzetih sa repozitorijuma grafovskih podataka.

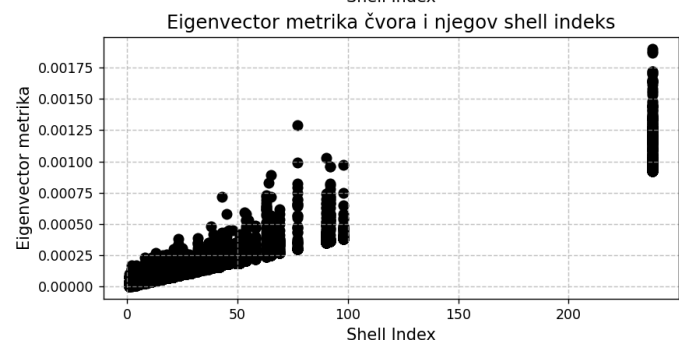
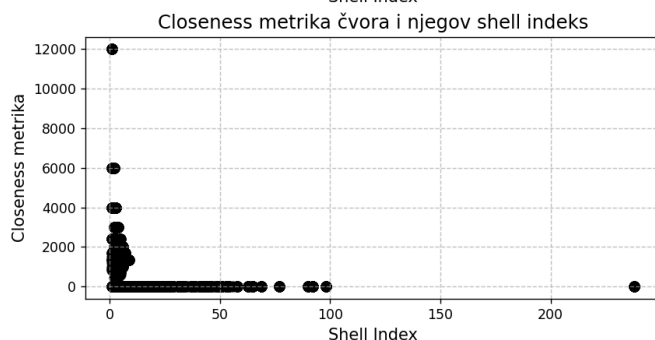
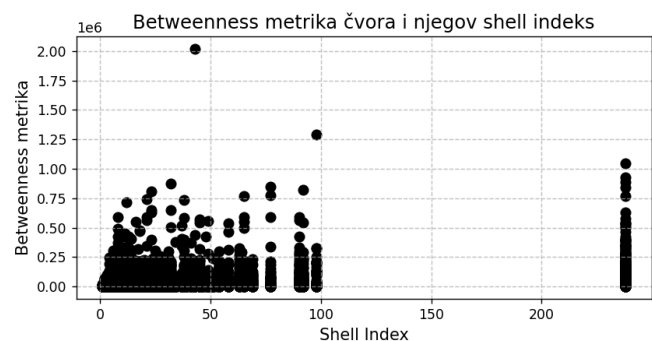
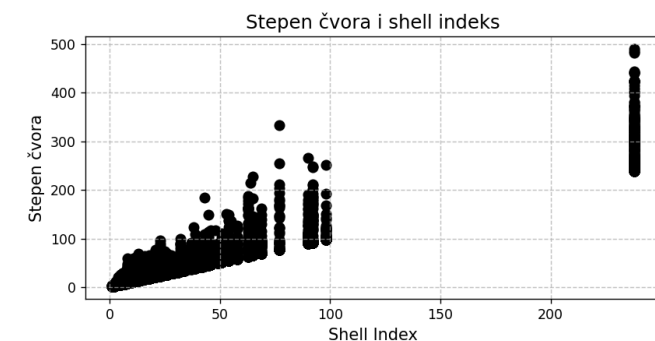
### 4.1 High Energy Physics - Phenomenology mreža

Mreža Arxiv HEP-PH predstavlja naučne saradnje između autora čija su istraživanja objavljena u kategoriji High Energy Physics - Phenomenology na arXiv-u. Čvorovi su autori, a grane predstavljaju naučne saradnje između autora.

- Broj čvorova: 12008
- Broj grana: 118521
- Gustina mreže: 0.0016440692740471677
- Minimalni stepen: 1
- Maksimalni stepen: 491
- Broj komponenti: 278
- Dijametar: 13
- Koeficijent malog sveta: 4.06790
- Koeficijent klasterisanja: 0.6117050103543644



#### 4.1.1 Korelacija šel indeksa i stepena čvora i globalnih metrika centralnosti



## 4.1.2 Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora

Shell indeks i stepena čvora: 0.9749018423307486

Shell indeks i betweenness centralnosti: 0.49906198053698486

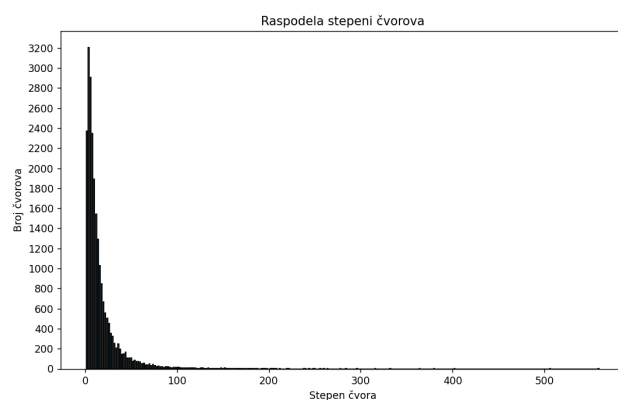
Shell indeks i closeness centralnosti: 0.46752871109291394

Shell indeks i eigenvector centralnosti: 0.8217332521890978

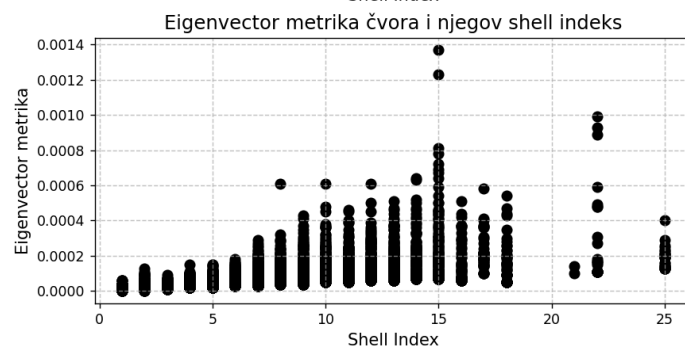
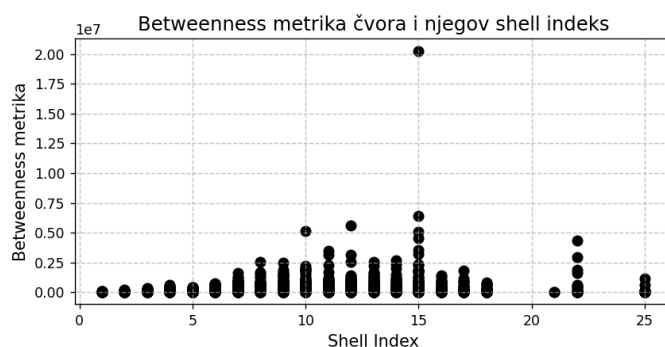
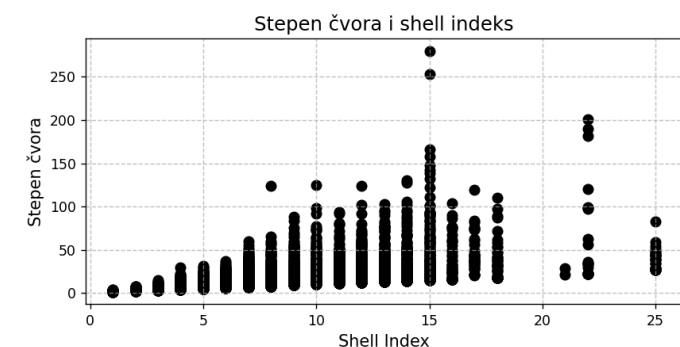
## 4.2 Condensed Matter Physics mreža

Mreža Arxiv COND-MAT predstavlja saradnju između autora koji su objavili radove u oblasti kondenzovane materije. Čvorovi su autori, a grane predstavljaju saradnju između autora.

- Broj čvorova: 23133
- Broj grana: 93497
- Gustina mreže: 3.494479766981958
- Minimalni stepen: 1
- Maksimalni stepen: 280
- Broj komponenti: 567
- Dijametar: 15
- Koeficijent malog sveta: 4.56446
- Koeficijent klasterisanja: 0.6335646052145414



### 4.2.1 Korelacija šel indeksa i stepena čvora i globalnih metrika centralnosti



## 4.2.2 Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora

Shell indeks i stepena čvora: 0.9420516577772977

Shell indeks i betweenness centralnosti: 0.5029873574089361

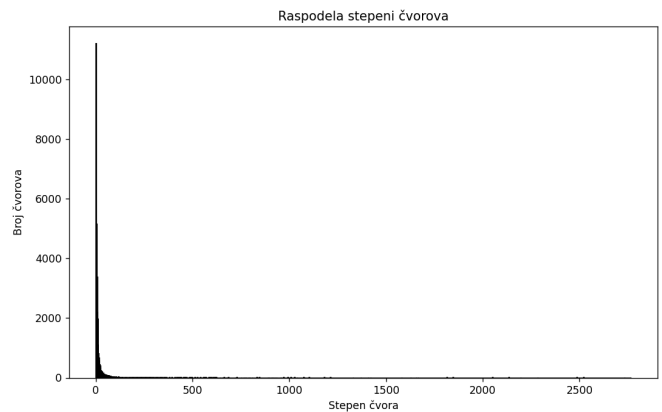
Shell indeks i closeness centralnosti: 0.40434749091777694

Shell indeks i eigenvector centralnosti: 0.8013769726381402

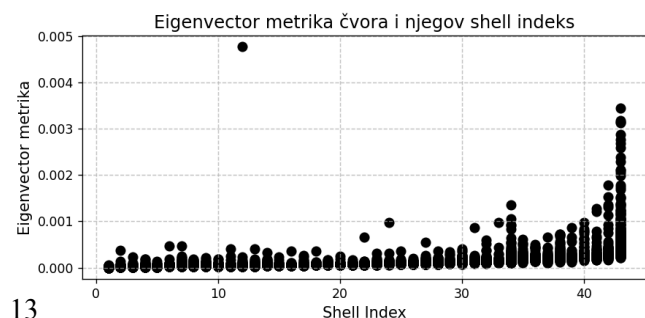
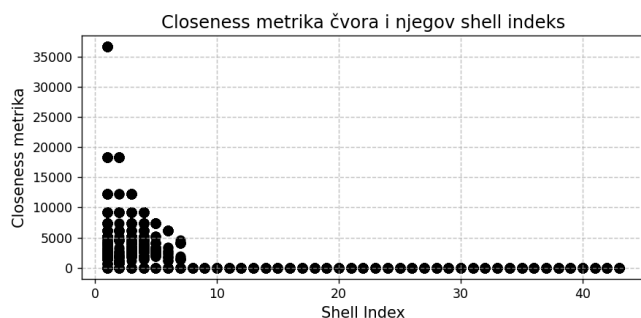
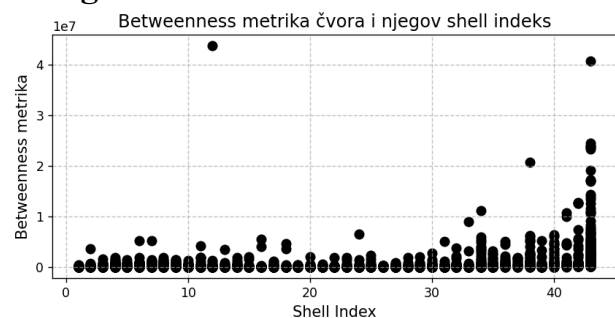
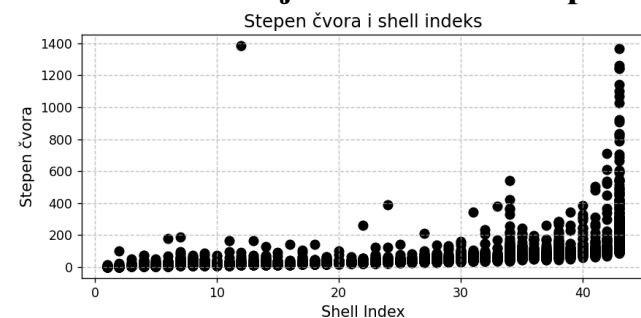
## 4.3 Email-Enron mreža

Mreža komunikacije putem e-maila Enron pokriva sve e-mail komunikacije unutar skupa podataka od oko pola miliona e-mailova. Čvorovi su pojedinačne e-mail adrese, a grane predstavljaju komunikaciju između e-mail adresa.

- Broj čvorova: 36692
- Broj grana: 183831
- Gustina mreže: 2.7309755503535
- Minimalni stepen: 1
- Maksimalni stepen: 1383
- Broj komponenti: 1065
- Dijametar: 13
- Koeficijent malog sveta: 4.025164
- Koeficijent klasterisanja: 0.4969825595995024



### 4.3.1 Korelacija šel indeksa i stepena čvora i globalnih metrika centralnosti



### **4.3.2 Spermanova korelacija šel indeksa i globalnih metrika centralnosti i stepena čvora**

Shell indeks i stepena čvora: 0.9838888348604387

Shell indeks i betweenness centralnosti: 0.6862312576427284

Shell indeks i closeness centralnosti: 0.3054598961801057

Shell indeks i eigenvector centralnosti: 0.8079525693718086

## **5 Zaključak**

Korektnost implementiranih algoritama je potvrđena kroz različite testove, tako što su svi dali očekivane rezultate. Ovo potvrđuje da su algoritmi za izračunavanje šel indeksa i globalnih metrika centralnosti tačni i pouzdani.

U realnim mrežama, shell indeks pokazuje visoku korelaciju sa stepenom čvora i eigenvector centralnošću, dok je korelacija sa closeness centralnošću nešto slabija.

U veštački generisanim mrežama, shell indeks pokazuje slabiju korelaciju sa metrikama koja je uglavnom manja od 0.5.

Vreme potrebno da bi se mreže analizirale značajno raste sa veličinom mreže.

## 6 Literatura

1. J. Leskovec, J. Kleinberg and C. Faloutsos. *Graph Evolution: Densification and Shrinking Diameters*. *ACM Transactions on Knowledge Discovery from Data* (ACM TKDD), 1(1), 2007.
2. J. Leskovec, J. Kleinberg and C. Faloutsos. *Graph Evolution: Densification and Shrinking Diameters*. *ACM Transactions on Knowledge Discovery from Data* (ACM TKDD), 1(1), 2007.
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*. *Internet Mathematics* 6(1) 29--123, 2009.
4. B. Klimmt, Y. Yang. *Introducing the Enron corpus*. CEAS conference, 2004.

### Linkovi za mreže

1. [CA-HepPh mreža](#)
2. [CA-CondMat mreža](#)
3. [Email-Enron mreža](#)