

## **Izveštaj projekta iz predmeta Socijalne Mreže**

Tema: Implementacija, testiranje i primena Batagelj-Zaveršnik algoritma

Lazić Veselin 162/23

## **0. Uvod**

Ovaj projekat je urađen u okviru kursa Analiza socijalnih mreža i ima za cilj implementaciju, testiranje i primenu Batagelj–Zaveršnik algoritma za k-core dekompoziciju i računanje šel indeksa u neusmerenim mrežama. Projekat je razvijen u programskom jeziku Java, koristeći biblioteku JUNG za rad sa grafovima.

Osnovne klase koje definišu strukturu mreže su:

- Node – predstavlja čvor u mreži. Svaki čvor ima jedinstveni identifikator id. Override-ovane su metode equals() i hashCode() kako bi se omogućilo pravilno poređenje i korišćenje u mapama i skupovima (HashMap, HashSet). Metoda toString() omogućava vizualizaciju čvora prilikom ispisivanja, po id-u čvora.
- Edge – predstavlja granu između dva čvora u grafu. Svaka grana može imati svoju oznaku, što omogućava razlikovanje i eventualno dodavanje dodatnih informacija.

Za manipulaciju grafovima i računanje osnovnih metriki, implementiran je paket GraphUtils, koji sadrži sledeće funkcionalnosti:

- avgDegree() – računa prosečan stepen čvorova u grafu.
- deepCopy() – kreira kopiju priloženog grafa, čime se omogućava manipulacija kopija bez uticaja na originalni graf.
- edgesToString() – konvertuje sve grane grafa u tekstualni oblik, u formatu čvor1-čvor2, što olakšava ispis i analizu.
- nodeDegrees() – vraća mapu čvor → stepen, što je korisno za dalje analize centralnosti i k-core dekompozicije.
- graphDensity() – računa gustinu mreže.
- computeDiameter() – izračunava prečnik mreže, tj. najdužu najkraću putanju između bilo koja dva čvora, koristeći Dijkstra shortest path algoritam.

## **1. Implementacija Batagelj–Zaveršnik algoritma**

Batagelj–Zaveršnik algoritam se koristi za k-core dekompoziciju neusmerenih mreža, tj. za identifikaciju k-jezgra, odnosno podmreža gde svaki čvor ima najmanje k veza sa drugim čvorovima iz te podmreže. Šel indeks čvora predstavlja najveće k-jezgro kojem taj čvor pripada.

Algoritam funkcioniše tako što:

1. Počinje od čvorova sa najmanjim stepenom.

2. Postepeno uklanja čvorove koji nemaju dovoljno veza da zadovolje trenutno k-jezgro.
3. Postavlja šel indeks svakom čvoru dok se cela mreža ne obradi.

Ovaj pristup omogućava linearnu vremensku složenost u odnosu na broj čvorova i grana, što je čini efikasnom za velike mreže, za razliku od Straight forward algoritma, koji će biti pomenut malo kasnije.

### *Implementacija u Javi*

Apstraktna klasa KCoreDecomposition definiše osnovu algoritma:

- Atribut graph tipa UndirectedSparseGraph<Node, Edge> predstavlja mrežu nad kojom se vrši analiza.
- Atribut shellIndices čuva mapu čvor → šel indeks tog čvoar.
- Metoda run() pokreće algoritam pozivom metode computeKCoress(), koja je apstraktna i implementira se u podklasama.
- Metoda getKcoreNodes(int k) vraća skup čvorova koji pripadaju k-jezgru, a getKcoreNetwork(int k) vraća k-jezgro za dato k.

Klasa BatageljZaversnik nasleđuje KCoreDecomposition i implementira konkretan algoritam:

1. computeKCoress() - algoritam:
  - Inicijalizuje čvorove u bucket-e po stepenu, tj. kcoreBuckets (LinkedHashSet za svaki stepen).
  - Iterativno prolazi kroz stepene od 0 do maksimalnog stepena.
  - Svaki čvor iz trenutnog bucketa dobija šel indeks k.
  - Ažurira stepen susednih čvorova i premešta ih u odgovarajući bucket ako im je smanjen stepen.
2. initializeBuckets(int maxDegree) – inicijalizuje početne skupove čvorova po stepenu. Vraća mapu čvor → trenutni stepen.
3. getMaxDegree() – pronađi maksimalni stepen čvora u mreži, potrebno za petlju u metodi computeKCoress().
4. getKcoreNetwork(int k) – kreira novi graf koji sadrži samo čvorove pripadajuće k-jezgru i grane između thi čvorova.
5. getShellIndex(Node node) – vraća šel indeks traženog čvora.
6. maxShellIndex() – vraća najveći šel indeks u mreži, što označava najveće k-jezgro.

Ova implementacija omogućava generičku primenu na bilo koju neusmerenu mrežu, kao i dodatnu funkcionalnost: izdvajanje bilo kog k-jezgra kao zasebne mreže za dalju analizu.

## **2. Mali test primeri**

U ovom delu se nalaze rezultati testiranja Batagelj–Zaveršnik algoritma na nekoliko malih mreža, i na mreži Zachary Karate Club. Za sve mreže su ručno provereni šel indeksi čvorova, kako bi se testirala tačnost algoritma. Metoda testSmallNetwork kao parametar uzima jednu mrežu, crta je(drawNetwork()), i zatim iterira po k-jezgrima mreže i štampa osnovne informacije o toj podmreži. Slike priložene ovde su generisane pomoću metode drawNetwork(), ova metoda služi za vizuelno prikazivanje neusmerene mreže (UndirectedSparseGraph). Koristi FRLayout za automatsko pozicioniranje čvorova u prostoru, a zatim kreira JFrame prozor, iz biblioteke Swing, u kojem se mreža prikazuje. Čvorovi su označeni svojim ID-evima, a veličina prikaza je podešena na  $600 \times 600$  piksela. Korišćenjem ovih metoda su testirane sledeće mreže:

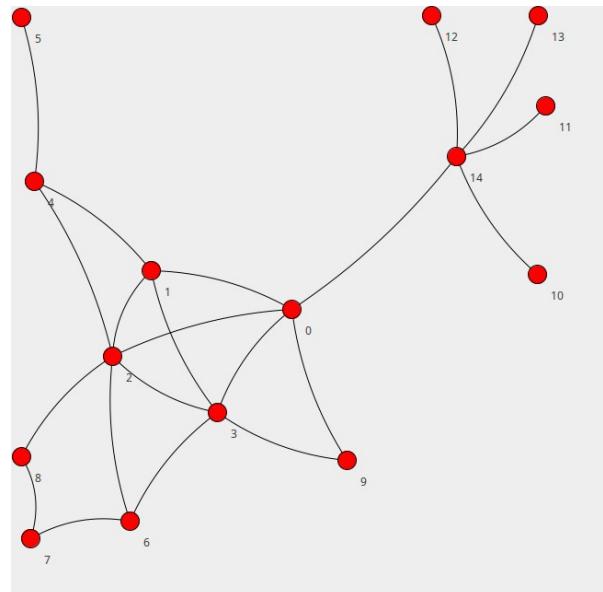
### 2.1 Graf sa prezentacije

#### **Opis mreže:**

Mreža predstavlja graf sa prezentacije 4, slajda 24. Mreža se sastoji od 15 čvorova i 21 grane. Struktura uključuje centralne čvorove sa višim stepenom povezivanja i nekoliko periferijskih čvorova sa nižim stepenom.

#### **Karakteristike:**

- Mreža pokazuje klasičnu strukturu sa centralnim “hub” čvorovima.
- Razlika između 2-core i 3-core je značajna, jer se periferni čvorovi brzo uklanjuju.
- Šel indeksi se povećavaju od periferije ka centru.



#### **Rezultati testiranja:**

#### **Šel indeksi:**

{0=3, 1=3, 2=3, 3=3, 4=2, 5=1, 6=2, 7=2, 8=2, 9=2, 10=1, 11=1, 12=1, 13=1, 14=1}

#### **Stepeni:**

{0=5, 1=4, 2=6, 3=5, 4=3, 5=1, 6=3, 7=2, 8=2, 9=2, 10=1, 11=1, 12=1, 13=1, 14=5}

#### **0-Jezgro:**

Broj čvorova: 15

Broj grana: 21

Čvorovi: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]

Grane: 3-9, 4-5, 10-14, 2-4, 11-14, 13-14, 3-6, 0-2, 0-1, 12-14, 0-9, 1-2, 7-8, 1-4, 0-14, 0-3, 1-3, 2-6, 2-3, 2-8, 6-7

### **1-Jezgro:**

Broj čvorova: 15

Broj grana: 21

Čvorovi: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]

Grane: 11-14, 1-4, 6-7, 0-9, 0-1, 3-9, 13-14, 10-14, 1-3, 2-6, 12-14, 4-5, 2-4, 1-2, 3-6, 2-8, 0-3, 0-14, 2-3, 0-2, 7-8

### **2-Jeygro:**

Broj čvorova: 9

Broj grana: 15

Čvorovi: [0, 1, 2, 3, 4, 6, 7, 8, 9]

Grane: 0-9, 2-3, 1-2, 3-9, 0-1, 3-6, 1-4, 2-8, 0-3, 6-7, 1-3, 2-4, 7-8, 2-6, 0-2

### **3-Jezgro:**

Broj čvorova: 4

Broj grana: 6

Čvorovi: [0, 1, 2, 3]

Grane: 1-3, 0-2, 2-3, 0-3, 0-1, 1-2

Pearsonov koeficijent za šel indeks i stepen čvora iznosi 0.750, a spearmanov 0.780. Ovo nam govori da postoji jaka korelacija između ove 2 metrike, međutim, za primer možemo uzeti čvor 14, koji iako ima velik stepen, njegov šel indeks je jednak 1. Razlog toga je jer nije povezan sa hubovima te mreže, već je većinski vezan za viseće čvorove, koji otpadaju u 2-jezgru, pa samim tim i čvor 14 otpada te je njegov šel indeks jednak 1.

## **2.2 Graf sa 2 klike**

### **Opis mreže:**

Mreža predstavlja graf sa dve klike, sa nekoliko čvorova koji spajaju klike. Mreža se sastoji od 20 čvorova i 32 grane. Struktura uključuje čvorove sa visokim stepenom u kliki i mostove koji povezuju različite delove mreže.

### **Karakteristike:**

- Graf pokazuje jasno izdvojene klike.
- Razlika između 3-core i 4-core nije velika jer su čvorovi u klikama svi povezani.
- Šel indeksi rastu ka unutrašnjim čvorovima klika.

### Rezultati testiranja:

#### Šel indeksi:

$\{0=4, 1=4, 2=4, 3=4, 4=4, 5=4, 6=4, 7=4, 8=4, 9=4, 10=2, 11=2, 12=2, 13=2, 14=2, 15=2, 16=2, 17=2, 18=2, 19=2\}$

#### Stepeni:

$\{0=7, 1=7, 2=8, 3=7, 4=6, 5=7, 6=7, 7=7, 8=6, 9=6, 10=2, 11=2, 12=2, 13=2, 14=2, 15=2, 16=2, 17=2, 18=3, 19=3\}$

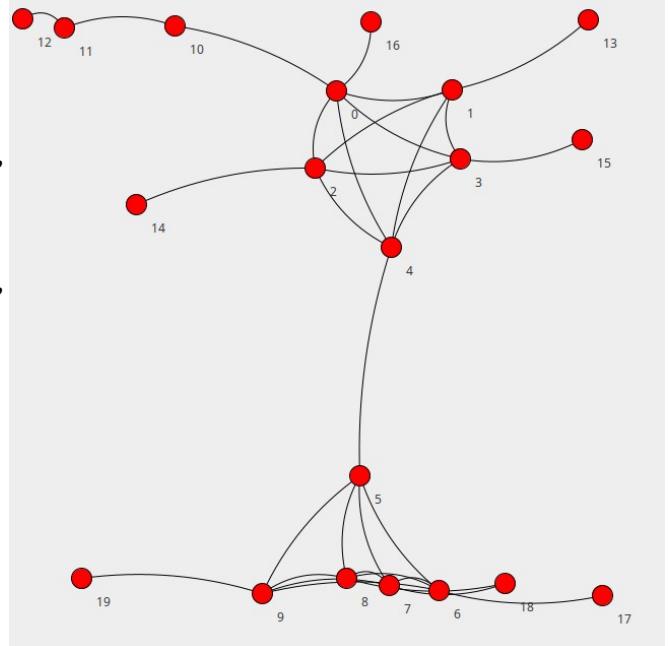
#### 0-Jezgro:

Broj čvorova: 20

Broj grana: 32

Čvorovi: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]

Grane: 11-12, 5-7, 6-17, 3-4, 1-2, 4-5, 5-6, 3-15, 8-9, 0-3, 7-18, 1-4, 2-14, 6-9, 0-4, 2-4, 0-1, 7-8, 9-19, 6-7, 5-8, 8-18, 2-3, 0-2, 1-13, 1-3, 0-10, 5-9, 6-8, 0-16, 7-9, 10-11



#### 1-Jezgro:

Broj čvorova: 20

Broj grana: 32

Čvorovi: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]

Grane: 1-2, 6-7, 7-9, 0-10, 0-16, 2-4, 4-5, 3-4, 7-18, 8-9, 6-9, 8-18, 0-2, 0-3, 6-17, 1-3, 11-12, 9-19, 6-8, 5-6, 2-3, 0-1, 1-4, 3-15, 10-11, 5-7, 5-8, 7-8, 1-13, 0-4, 2-14, 5-9

#### 2-Jezgro:

Broj čvorova: 11

Broj grana: 23

Čvorovi: [0, 1, 2, 18, 3, 4, 5, 6, 7, 8, 9]

Grane: 0-3, 1-2, 6-9, 18-8, 6-8, 0-1, 0-4, 1-4, 2-3, 4-5, 8-9, 5-6, 2-4, 5-7, 3-4, 6-7, 5-9, 0-2, 7-9, 5-8, 18-7, 7-8, 1-3

#### 3-Jezgro:

Broj čvorova: 10

Broj grana: 21

Čvorovi: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]

Grane: 7-8, 5-8, 2-3, 4-5, 0-4, 5-7, 6-7, 0-2, 0-1, 1-3, 1-4, 2-4, 5-9, 7-9, 3-4, 5-6, 8-9, 6-8, 6-9, 0-3, 1-2

#### 4-Jezgro:

Broj čvorova: 10

Broj grana: 21

Čvorovi: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]

Grane: 2-3, 5-7, 0-2, 7-9, 0-1, 5-9, 8-9, 1-2, 0-3, 6-7, 1-3, 4-5, 6-8, 6-9, 7-8, 1-4, 2-4, 0-4, 5-6, 3-4, 5-8

Pearsonov koeficijent za šel indeks i stepen čvora iznosi 0.982, a Spearmanov 0.944.

Graf pokazuje vrlo jaku korelaciju između šel indeksa i stepena čvorova, što znači da u ovom grafu prosečno važi visok stepen → visok šel indeks.

### 2.3 Graf sa 4 particije

#### Opis mreže:

Mreža predstavlja graf sa četiri particije čvorova koje su međusobno povezane manjim brojem grana.

Mreža se sastoji od 18 čvorova i 16 grana. Struktura uključuje čvorove koji povezuju različite grupe, dok većina čvorova u grupama ima manji stepen.

#### Karakteristike:

- Graf je raspodeljen u četiri particije.
- Razlika između 1-core i 2-core pokazuje da periferni čvorovi brzo otpadaju.
- Šel indeksi su niži i čvorovi centralnih veza imaju veći indeks.

#### Rezultati testiranja:

##### Šel indeksi:

{0=2, 1=2, 2=2, 3=2, 4=2, 5=2, 6=2, 7=1, 8=1, 9=1, 10=1,

11=2, 12=2, 13=2, 14=1, 15=1, 16=1, 17=1}

##### Stepeni:

{0=3, 1=3, 2=4, 3=3, 4=3, 5=3, 6=5, 7=1, 8=1, 9=1, 10=1,

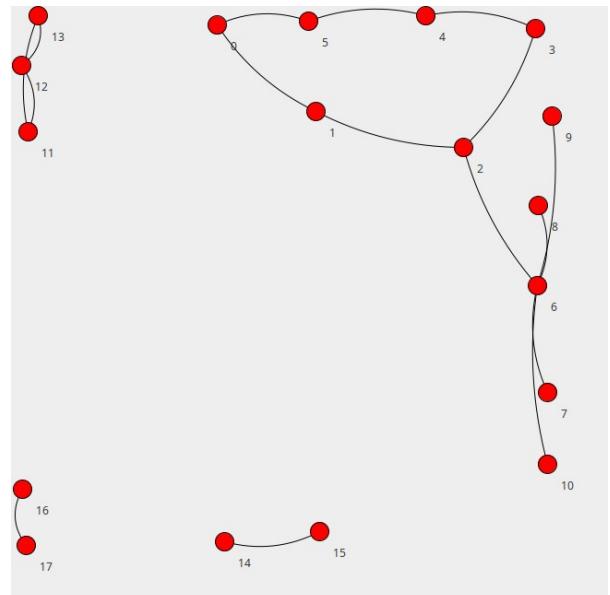
11=3, 12=3, 13=3, 14=1, 15=1, 16=1, 17=1}

#### 0-Jezgro:

Broj čvorova: 18

Broj grana: 16

Čvorovi: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,



17]

Grane: 14-15, 0-1, 11-13, 6-7, 12-13, 2-3, 4-5, 0-5, 11-12, 16-17, 1-2, 2-6, 6-9, 6-10, 3-4, 6-8

### **1-Jezgro:**

Broj čvorova: 18

Broj grana: 16

Čvorovi: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]

Grane: 1-2, 12-13, 3-4, 0-5, 2-6, 16-17, 6-8, 11-12, 0-1, 2-3, 6-7, 4-5, 6-10, 6-9, 11-13, 14-15

### **2-Jezgro:**

Broj čvorova: 9

Broj grana: 9

Čvorovi: [0, 1, 2, 3, 4, 5, 11, 12, 13]

Grane: 0-5, 3-4, 4-5, 0-1, 2-3, 11-12, 12-13, 1-2, 11-13

Pearsonov koeficijent za šel indeks i stepen čvora iznosi 0.342, a Spearmanov 0.742.

Graf pokazuje slabiju Pearsonovu korelaciju, ali Spearmanov rezultat sugerira da postoji umeren odnos između rankova šel indeksa i stepena čvorova.

## 2.4 Zachary Karate Klub

### **Opis mreže:**

Zachary Karate Club mreža je poznata realna društvena mreža sa 34 čvora i 34 gerane, predstavljajući veze između članova karate kluba. Mreža je često korišćena u istraživanjima društvenih mreža zbog jasno definisanih grupa i interakcija između članova.

### **Karakteristike:**

- Većina čvorova ima nisku konektivnost i brzo se uklanja iz viših k-jezgra.
- Jezgro mreže (2-jezgro) je veoma malo, što ukazuje na postojanje nekoliko centralnih članova sa višim stepenom povezanosti.
- Šel indeksi jasno pokazuju koje čvorove čine ključne hub-čvorove u mreži.

### **Rezultati testiranja:**

#### **Šel indeksi:**

{0=2, 1=2, 2=3, 3=3, 4=3, 5=2, 6=3, 7=2, 8=2, 9=2, 10=2, 11=2, 12=2, 13=2, 14=2, 15=2, 16=2, 17=2, 18=1, 19=1, 20=1, 21=1, 22=1, 23=1, 24=1, 25=1, 26=1, 27=1, 28=1, 29=1, 30=1, 31=1, 32=1, 33=1}

#### **Stepeni:**

{0=3, 1=3, 2=4, 3=4, 4=4, 5=3, 6=3, 7=2, 8=2, 9=2, 10=2, 11=2, 12=2, 13=2, 14=2, 15=2, 16=2, 17=2, 18=1, 19=1, 20=1, 21=1, 22=1, 23=1, 24=1, 25=1, 26=1, 27=1, 28=1, 29=1, 30=1, 31=1, 32=1, 33=1}

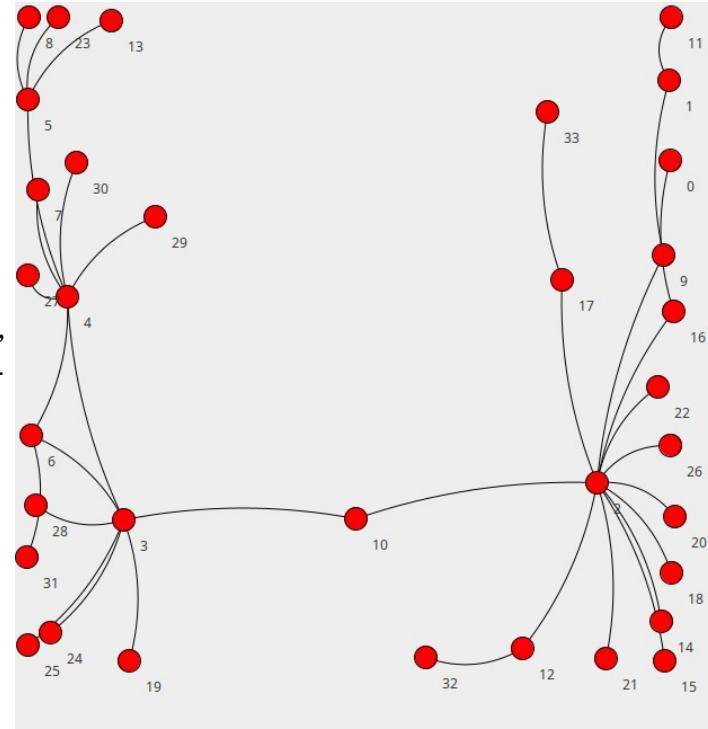
### 0-Jezgro:

Broj čvorova: 34

Broj grana: 34

Čvorovi: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33]

Grane: 2-17, 5-23, 3-10, 6-31, 2-20, 3-24, 4-7, 0-16, 2-26, 2-12, 4-5, 2-10, 4-29, 2-15, 2-9, 4-30, 2-21, 2-14, 3-25, 5-13, 3-28, 2-18, 1-11, 2-22, 3-4, 17-33, 2-16, 3-19, 3-6, 5-8, 12-32, 4-27, 1-9, 4-6



Broj čvorova: 34

Broj grana: 34

Čvorovi: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33]

Grane: 2-17, 2-16, 2-20, 4-27, 5-8, 3-25, 3-10, 3-28, 2-10, 3-4, 1-11, 4-5, 0-16, 2-21, 3-19, 12-32, 4-30, 5-23, 4-7, 4-6, 3-6, 2-18, 6-31, 5-13, 3-24, 1-9, 2-15, 2-14, 4-29, 2-26, 2-12, 2-9, 17-33, 2-22

### 2-Jezgro:

Broj čvorova: 3

Broj grana: 3

Čvorovi: [3, 4, 6]

Grane: 4-6, 3-6, 3-4

Pearsonov koeficijent za šel indeks i stepen čvora iznosi 0.499, a Spearmanov 0.543.

Graf/mreža pokazuje umerenu korelaciju između šel indeksa i stepena čvorova.

### Zaključak

Struktura mreže direktno utiče na maksimalni k-core i veličinu pojedinačnih k-jezgara.

- Mreže sa zajednicama ili klikama imaju veće maksimalne k-jezgra, razdvojene mreže imaju manja jezgra.
- Batagelj-Zaveršnik algoritam precizno identificuje k-jezgra u svim test mrežama.
- Šel indeksi omogućavaju kvantifikaciju važnosti čvorova i vizualizaciju hijerarhije čvorova u mreži.

### **3. Sintetički test primeri**

Opis testova:

Za testiranje implementacije šel indeksa korišćeni su veštački generisani sparse random grafovi sa više od 1000 čvorova. Šel indeksi su izračunati pomoću jednostavnog “straightforward” algoritma koji kopira mrežu i uklanja čvorove po stepenu dok se ne izračuna coreness. Ovaj algoritam je poslužio kao referentna tačka za poređenje sa Batagelj-Zaversnik algoritmom. Testovi su izvedeni na tri tipa mreža:

1. **Erdos-Renyi model (ER)** – random mreža gde je svaka moguća grana između čvorova prisutna sa verovatnoćom  $p$ .
2. **Barabasi-Albert model (BA)** – mreža koja se gradi dodavanjem novih čvorova sa  $m$  granama i preferencijalnim povezivanjem prema već postojećim čvorovima proporcionalno stepenu, prati stepeni zakon. Početni graf u testiranju je bio po Erdos-Renyi modelu.
3. **Core-Periphery model (CP)** – mreža sa izraženom strukturom jezgro-periferija. Testirani su i diskretni i kontinuirani modeli, gde diskretni ima definisane core i periphery čvorove sa različitim verovatnoćama povezivanja, dok kontinuirani model dodeljuje random coreness vrednosti čvorovima i povezuje ih prema proizvodu tih coreness vrednosti.

Straight forward algoritam predstavlja u teoriji vrlo jednostavan algoritam za računanje šel indeksa u mreži. Zasniva se na odstranjuvanju čvorova čiji je stepen manji od  $k$ , sve dok takvih čvorova ima u mreži. U okviru projekta implementiran je tako da je  $k$  uvek jednak maksimalnom šel indeksu, odnosno, implementiran je tako da izračuna šel indeks za sve čvorove odjednom. Koristi pomoćne skupove u kojima se čuvaju čvorovi koji treba da budu odstranjeni, zbog prirode metode `.getVertices()`. Vremenska složenost (time complexity) ovog algoritma u najgorem slučaju je  $O(n^*(n+m))$ , velika razlika u poređenju sa Batagelj-Zaversnik algoritmom, koji ima vremensku složenost  $O(n+m)$  u najgorem slučaju. Ovo se da primetiti u testovima koji su izvedeni, gde su oba algoritma proizvela identične šel indekse, međutim, Batagelj-Zaversnik je to uradio mnogo brže.

#### **Erdos-Renyi model (ER)**

Erdos-Renyi model, često označen kao  $G(n,p)$ , je jedan od najjednostavnijih i najpoznatijih modela za generisanje slučajnih grafova. U ovom modelu, graf se sastoji od  $n$  čvorova, a verovatnoća  $p$  određuje verovatnoću postojanja grane/veze između bilo koja dva čvora. Ovaj model se često koristi kao osnovna polazna tačka za proučavanje teorije grafova.

Graf se generiše dodavanjem  $n$  čvorova. Zatim, za svaki mogući par čvorova, nasumično se odlučuje da li će postojati veza između njih, pri čemu je verovatnoća te grane fiksirana na  $p$ . To znači da svaki par čvorova ima istu šansu da bude povezan, nezavisno od bilo kojih drugih veza.

**Karakteristike:** Glavna karakteristika Erdos-Renyi grafova je da su svi čvorovi uniformno povezani. Distribucija stepena čvorova (broja veza koje čvor ima) sledi binomnu distribuciju, a prosečan stepen je

približno  $p \cdot (n-1)$ . Ovi grafovi ne modeluju dobro fenomene iz stvarnog sveta kao što su društvene mreže, gde distribucija stepena obično sledi pravilo snage (power law).

Rezultati testiranja:

- ER mreže pokazuju uniformnu raspodelu šel indeksa sa maksimalnim k-core vrednostima koje rastu sa gustinom mreže, velika većina čvorova se nalazi u maksimalnom jezgru mreže
- Batagelj-Zaveršnik algoritam je 10–20 puta brži od straight forward algoritma na mrežama sa 10.000–20.000 čvorova.
- Idenični rezultati oba algoritma potvrđuju ispravnost implementacije Batagelj-Zaveršnik algoritma.

### **Barabasi-Albert model (BA)**

Barabasi-Albert (BA) model je model za generisanje slučajnih, bez-skalnih (scale-free) grafova. Glavna ideja je preferencijalno vezivanje (preferential attachment), princip da čvorovi sa većim stepenom privlače više novih veza. Ovaj model mnogo bolje od Erdos-Renyi modela opisuje mnoge mreže iz stvarnog sveta, poput interneta i društvenih mreža, gde postoji veliki broj slabo povezanih čvorova i mali broj hubova sa mnogo veza.

Algoritam počinje sa malim, početnim grafom od  $m_0$  čvorova, u ovom testiranju je to Erdos-Renyi graf. U svakom koraku, dodaje se novi čvor sa  $m$  grana. Svaka od tih  $m$  grana se povezuje sa postojećim čvorovima u grafu. Verovatnoća da se novi čvor poveže sa nekim postojećim čvorom je proporcionalna stepenu tog postojećeg čvora. To znači da je veća verovatnoća da se novi čvor veže za čvor koji već ima mnogo veza.

**Karakteristike:** Ključna karakteristika BA modela je distribucija stepena koja sledi power law (stepeni zakon). Rezultat su grafovi sa čvorovima koji imaju jako malo veza i manjim brojem čvorova sa jako puno veza ("hubova"), što je tipična struktura mnogih stvarnih mreža.

Rezultati testiranja:

- BA mreže pokazuju snažnu strukturu jedyro-periferija sa visokim maksimalnim k-core vrednostima.
- Batagelj-Zaveršnik algoritam je 5–7 puta brži od straightforward algoritma na mrežama sa 10.000 čvorova.
- Idenični rezultati oba algoritma potvrđuju ispravnost implementacije Batagelj-Zaveršnik algoritma.

## Core-Periphery model (CP)

Core-Periphery model opisuje strukturu grafova koja se sastoji od dva dela: jezgra (core) i periferije (periphery). Jezgro je gusto povezan skup čvorova, dok je periferija skup čvorova koji su labavo povezani, uglavnom sa čvorovima u jezgru, a manje međusobno. Ovaj model se često koristi za analizu ekonomskih mreža, mreža transporta i društvenih grupa.

Implementacija ovog modela pomoću klase CorePeripheryModel pokazuje dva pristupa:

1. **Diskretni model:** U ovom modelu se čvorovi eksplicitno dele u dve grupe, jezgro i periferiju.

Verovatnoća stvaranja grane je najveća unutar jezgra, manja između jezgra i periferije, a najmanja unutar same periferije (u ovom testiranju ta verovatnoća ne postoji-vrednost je 0). Ovu strukturu simuliraju parametri cc (verovatnoća veza unutar jezgra), pc (verovatnoća veza između jezgra i periferije) i pp (verovatnoća veze između periferije i periferije).

2. **Kontinualni model:** Ovde svaki čvor ima vrednost "jezgrovitosti" (coreness), broj između 0 i 1.

Verovatnoća da se dva čvora povežu je proizvod njihovih vrednosti "jezgrovitosti". Što je veća vrednost, to je veća verovatnoća da čvor bude povezan, i to na način koji stvara klaster jezgra.

**Karakteristike:** Osnovna karakteristika Core-Periphery modela je njegova jasna, dvoslojna struktura. Graf ima gusto povezan centralni deo (jezgro) koji služi kao "hub" za prenos informacija, dok je ostatak mreže (periferija) mnogo ređe povezan i zavisi od veza sa jezgrom. Ovaj model se razlikuje od BA modela po tome što struktura nije samo posledica distribucije stepena, već je definisana eksplicitnom podelom na uloge.

Rezultati testiranja:

- Jezgro mreže formira gustu klasu čvorova sa visokim k-core vrednostima, dok periferija brzo otpada kod viših k-jezgra. Većina "core" čvorova završava u maksimalnom jezgru mreže.
- Batagelj-Zaveršnik algoritam je znatno efikasniji od straightforward algoritma, naročito za velike mreže (više od 10.000 čvorova).
- Oba algoritma daju identične rezultate, što potvrđuje ispravnost implementacije Batagelj-Zaveršnik algoritma.

## 4. Primena i empirijske analize:

Proučavanje **k-core dekompozicije** i **šel indeksa** pruža dubok uvid u strukturalnu organizaciju kompleksnih mreža, naročito velikih društvenih i tehnoloških mreža.

U ovoj empirijskoj analizi fokusiraćemo se na 5 realnih, velikih neusmerenih mreža preuzete sa javno dostupnih repozitorijuma grafovskih podataka. Za svaku mrežu izvršićemo k-core dekompoziciju i analizirati distribuciju šel indeksa među čvorovima. Takođe, ispitaćemo kako šel indeks čvora korelira sa drugim ključnim metrikama mreže, uključujući:

- Stepen čvora, koji meri lokalnu povezanost.
- Closeness centralnost, koja pokazuje koliko je čvor „blizu“ svim ostalim čvorovima u mreži.
- Eigenvector centralnost, koja se određuje na osnovu povezanosti čvora sa “bitnim” čvorovima u mreži.
- Betweenness centralnost, koja pokazuje ulogu čvora u posredovanju protoka informacija.

Korelacije između šel indeksa i ovih centralnosti kvantifikovane su korišćenjem Pearsonovog, koji prikazuje direktnu povezanost između podataka i Spearmanovog koeficijenta rang-korelacije, koji prikazuje monotonu povezanost između podataka.

Na kraju, ispitaćemo makroskopske strukturalne karakteristike k-core-ova kako se k povećava, uključujući:

- Broj čvorova i grana unutar k-core-a.
- Gustinu grana i broj povezanih komponenti.
- Udeo čvorova i grana u najvećoj povezanoj komponenti.
- Prečnik najveće povezane komponente

Cilj ove analize je da se otkriju obrasci otpornosti mreže, identifikuju uticajni čvorovi i razume kako lokalna svojstva čvorova (stepen i centralnosti) utiču na njihovu poziciju u hijerarhijskoj k-core strukturi. Grafički prikazi dopunjuju numeričke analize, sa scatter plotovima koji prikazuju šel indeks u odnosu na stepen i centralnosti čvorova, ilustrujući vezu između coreness-a čvora i njegove uloge u mreži.

#### **4.0.1 Klasa Network**

Klasa Network je apstraktna klasa koja predstavlja generičku neusmerenu mrežu i pruža metode za analizu k-core struktura i centralnosti čvorova. Glavne karakteristike:

- Čuvanje grafa: koristi UndirectedSparseGraph<Node, Edge> iz JUNG biblioteke.
- k-core analiza:
  - Koristi implementaciju Batagelj-Zaveršnik algoritma (BatageljZaversnik) da odredi šel indekse čvorova.
  - Za svaki k-core računa: broj čvorova i grana, gustinu, broj povezanih komponenti, udeo čvorova i grana u najvećoj komponenti, prečnik najveće komponente (ako nije prevelika)

Interaktivna centralnost:

- Pitanja korisniku da li želi da generiše grafove i računa centralnosti.

- Računa:
  - Stepen čvorova
  - Closeness, betweenness, eigenvector centralnost
- Koristi Spearmanovu i Pearsonovu korelaciju za poređenje sa šel indeksom.
- Može pisati metrike u CSV fajl i prikazivati regresione grafikone.

#### **4.0.2 Klasa BatchNetworkAnalysis**

Omogućava paralelnu obradu više mreža koristeći ExecutorService i thread pool.

- Korisno kada se radi računanje metrika za više ogromnih mreža(gde je broj čvorova > 100\_000)
- Svaka mreža se analizira pozivanjem Network.runCentralityAnalysisBatchMode().
- Podržava timeout i izveštavanje o neuspelim zadacima.

Ova klasa i njena metoda su korišćene kako bi se rezultati računanja metrika ispisali u fajlove, pomoću pomoćne klase GraphFileWriter, koji ove metrike formatira u CSV fajl.

#### **4.0.3 Klasa ApproximationUtils**

Izračunavanja closeness centralnosti je vrlo neefikasno( $O(n*(n+m))$ ), i u velikim mrežama ovaj proces tako dugo traje, jednako neefikasan je i klasičan metod za izračunavanje betweenness centralnosti( $O(n*(n+m))$ ), dok je metod za eigenvector centralnost malo bolji te u najgorem slučaju ima složenost  $O(n^2)$ . Kao neko rešenje rade se aproksimacije ovih centralnosti, tako što se za određen, random odabran “uzorak” čvorova izračunava centralnost(BFS-breadth first search, u closeness i betweenness), a za ostale čvorove se te vrednosti aproksimiraju.

ApproximationUtils omogućava aproksimaciju centralnosti za velike mreže kako bi se smanjilo vreme računanja:

- Harmonic closeness centrality (sample):
  - Izračunava reciprocne udaljenosti od nasumičnog uzorka čvorova.
  - Time se izbegava BFS od svih čvorova, radi se BFS samo zta te random odabrane čvorove.
- Betweenness centrality (sample):
  - Koristi uzorak čvorova i Brandes algoritam.
  - Akumulira doprinose samo za uzorkovane BFS pretrage.
- Eigenvector centrality (iteracije):

- Power iteration metoda sa zadatim brojem iteracija.
- Normalizuje rezultate nakon svake iteracije.
- Sve aproksimacije vraćaju `HashMap<Node, Double>` sa rezultatima po čvoru.(čvor → vrednost centralnosti)
- BFS metoda (`bfsDistances`) je pomoćna metoda koja računa udaljenosti od jednog čvora do svih drugih.

## **4.1 Protein Pathways**

### **4.1.1 Uvod**

Protein Pathways je protein-protein interakciona mreža koja predstavlja fizičke interakcije između proteina u ljudima.

- Čvorovi: 21.557 proteina
- Grane: 338.636 fizičkih interakcija (nakon uklanjanja self-loops)
- Petlje: 3.717
- Tip mreže: neusmerena, bez težina
- Podaci: eksperimentalno potvrđene metabolističke i signalne interakcije

Mreža je učitana iz CSV fajla `PP-Pathways_ppi.csv` koristeći statičku metodu `readUndirectedSparseGraph`.

### **4.1.2 Osnovne karakteristike mreže**

Nakon učitavanja mreže, izvršena je analiza osnovnih statističkih karakteristika:

- Prosečan stepen čvora: 31,417
- Maksimalni stepen čvora: 2130 (najviše povezan protein)
- Minimalni stepen čvora: 0 (izolovani protein)
- Gustina mreže: 0.146

Ove vrednosti pokazuju da je mreža relativno retka, sa velikim brojem čvorova koji imaju mali stepen, dok postoji nekoliko hub proteina sa velikim brojem interakcija, što je karakteristično za biološke mreže.

#### **4.1.3 Stepen distribucije**

Analiza distribucije stepena čvorova pokazuje tipičnu "scale-free" strukturu: većina čvorova ima mali broj interakcija, dok mali broj hub proteina ima veoma veliki stepen, što je karakteristično za biološke mreže.

#### **4.1.4 K-core analiza**

Za svaki k-core analizirani su prosečan stepen i prosečna dužina puta tamo gde je moguće:

- Za niske k-coreove (0–77) mreža je prevelika, pa je izračunavanje prosečne dužine puta preskočeno.
- Kod visokih k-coreova (105+) mreža je gotovo potpuna, pa je prosečna dužina puta 1.

Trendovi k-coreova:

1. Smanjenje čvorova i grana sa rastom k:

- Od 64-core do ~104-core, broj čvorova polako opada ( $615 \rightarrow 309$ ), dok se broj grana smanjuje ( $45.258 \rightarrow 22.943$ ).
- Posle ~105-core, broj čvorova stabilizuje se na 130, a broj grana na 8.385.

2. Trend gustine:

- Gustina raste kako k raste:
  - 64-core: ~23,97%
  - 100-core: ~47,62%
  - 104-core: ~48,21%
  - 105-core i dalje: 100% (svi čvorovi povezani-klika)

3. Povezane komponente:

- Svaki k-core ostaje potpuno povezan (1 komponenta), a najveća komponenta uključuje 100% čvorova i grana.

4. Dijametar:

- Dijametar je 3 za većinu coreova, što znači da su bilo koja dva čvora udaljena najviše 3 koraka.
- Od 95-core dijametar se smanjuje na 2, a u kliki (105-core i dalje) postaje 1 (svi čvorovi direktno povezani).

5. Prosečna dužina puta:

- K-coreovi srednjeg reda (78–104) imaju veoma nisku prosečnu dužinu puta (1,524–1,635), što ukazuje na efikasnu signalnu transmisiju među proteinima.

#### **4.1.5 Spearmanovi koeficijenti korelaciјe**

Šel indeks vs stepen: 0.996

Šel indeks vs closeness: 0.865

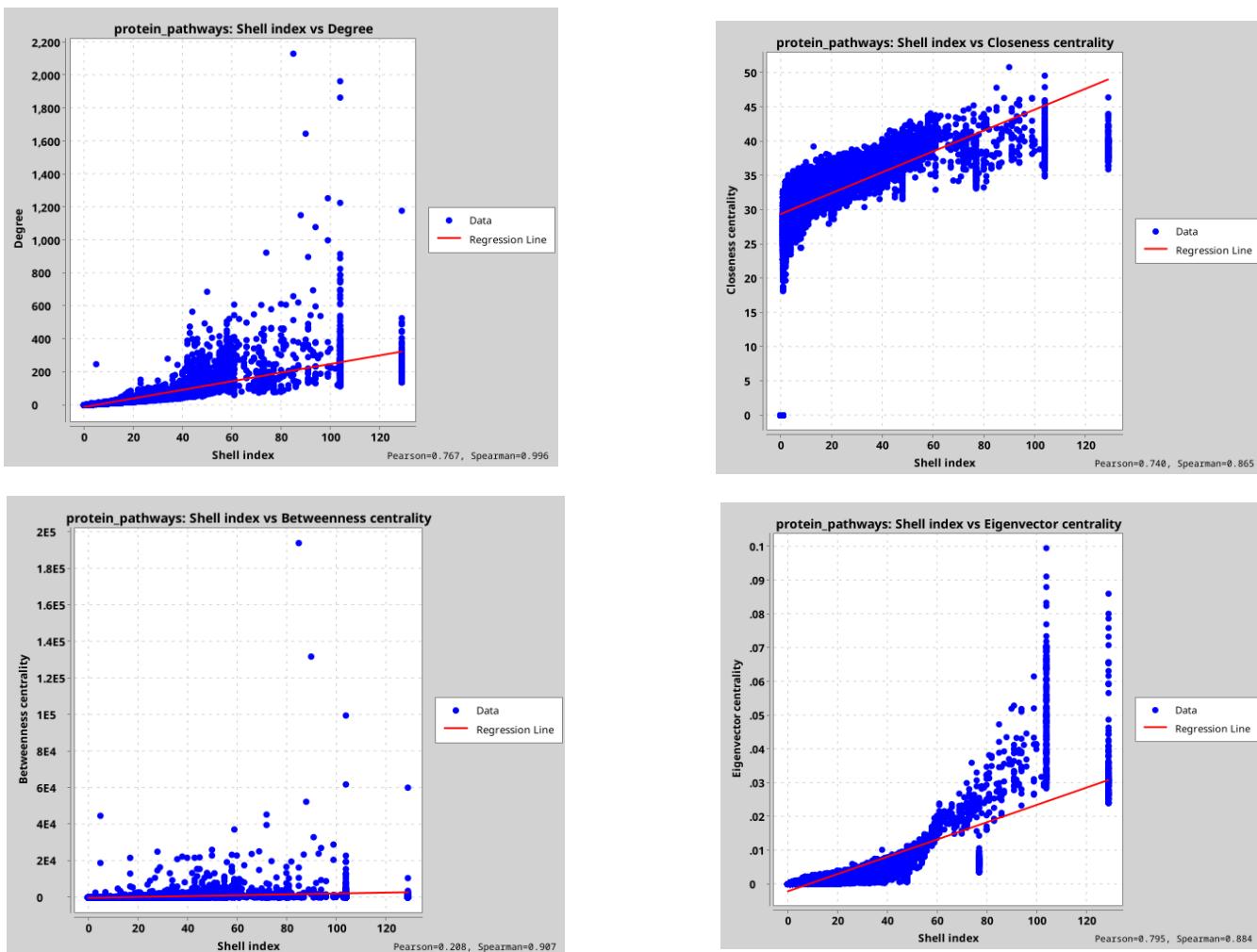
Šel indeks vs betweenness: 0.907

Šel indeks vs eigenvector: 0.884

Iz ovih rezultata možemo zaključiti da je šel indeks u visokoj korelaciji sa stepenom i metrikama centralnosti, odnosno visok šel indeks → visoka metrika.

#### **4.1.6 Grafički prikaz**

Grafički su prikazani čvorovi, tako da: jedan čvor je jedna tačka na grafikonu, na x osi šel indeks, na y osi su stepen/centralnost, redom: stepen, closeness, betweenness, eigenvector.



## **4.2 EmailEuCore**

### **4.2.1 Uvod**

Mreža je generisana na osnovu email komunikacije unutar velike evropske istraživačke institucije. Sadrži informacije o svim dolaznim i odlaznim email-ovima između članova institucije.

- **Broj čvorova:** 1005
- **Broj grana (edges):** 16 064
- **Broj petlji:** 642
- **Broj duplih greana:** 8,865

### **4.2.2 Osnovne karakteristike mreže**

- Prosečan stepen čvora: 31,97
- Maksimalni stepen čvora: 345
- Minimalni stepen čvora: 0
- Gustoća mreže: 3,184%

#### **Zapažanja:**

Mreža je generalno retka, sa nekoliko centralnih hub čvorova koji drže mrežu povezану. Većina čvorova ima mali broj veza, što je tipično za biološke ili društvene mreže sa hubovima.

### **4.2.3 K-Core analiza**

#### **Opšti trendovi:**

- Gustina mreže raste u višim k-core-ovima: od ~3,18% u 0-core do 61,15% u 34-core.
- Sve komponente od 1-core nadalje ostaju povezane u jednu komponentu, što pokazuje postojanje jednog velikog centralnog klastera.

k-core	Čvorovi	Grane	Prosečan stepen	Gustoća	Komponente	Prečnik
0	1005	16,064	31,97	3,18%	20	7
5	784	15,655	39,94	5,10%	1	4
10	671	14,881	44,35	6,62%	1	4
20	461	11,731	50,89	11,06%	1	3
30	166	4,542	54,72	33,16%	1	3
34	79	1,884	47,70	61,15%	1	2

#### 4.2.4 Spearmanovi koeficijenti korelaciјe

Šel indeks vs stepen: 0.985

Šel indeks vs closeness: 0.944

Šel indeks vs betweenness: 0.896

Šel indeks vs eigenvector: 0.972

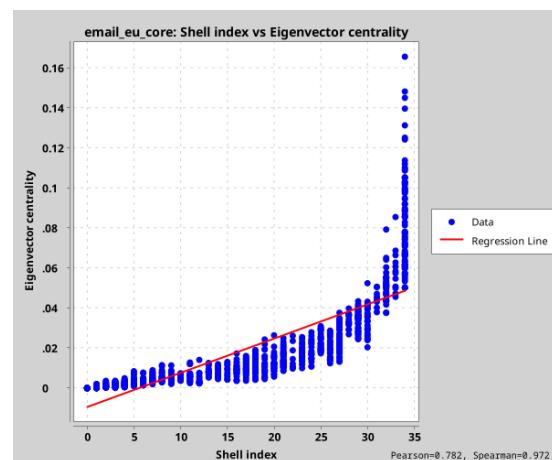
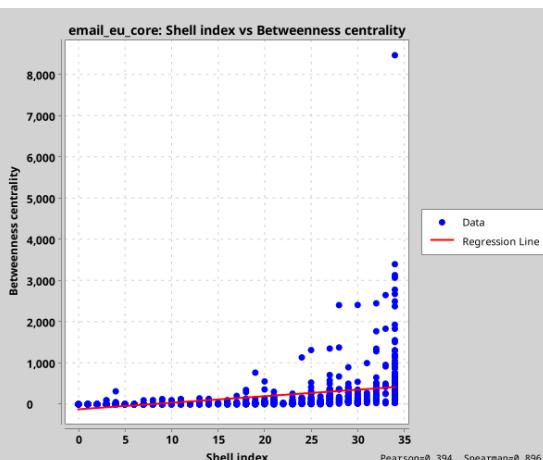
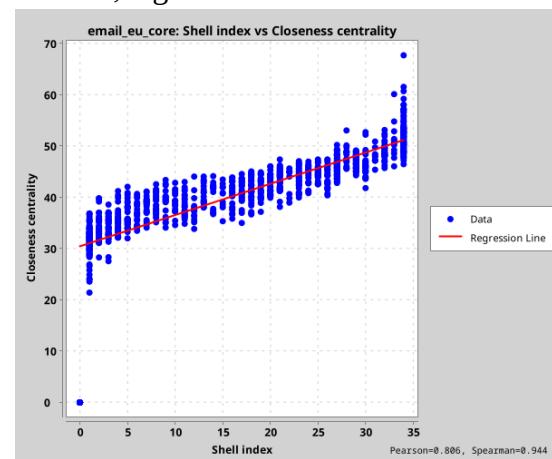
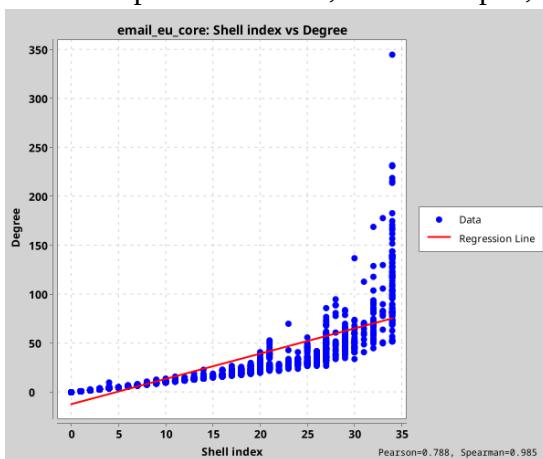
Iz ovih rezultata možemo zaključiti da je šel indeks u visokoj korelaciji sa stepenom i metrikama centralnosti, odnosno visok šel indeks → visoka metrika.

#### 4.2.5 Prosečna dužina puta

- Prosečna dužina puta u višim k-core-ovima se kreće oko 1,5–2.
- Mala vrednost prečnika i kratki putevi potvrđuju small-world karakteristike jezgra mreže: centralni čvorovi su vrlo blizu jedni drugima, olakšavajući brzu distribuciju informacija.

#### 4.2.6 Grafički prikaz

Grafički su prikazani čvorovi, tako da: jedan čvor je jedna tačka na grafikonu, na x osi šel indeks, na y osi su stepen/centralnost, redom: stepen, closeness, betweenness, eigenvector.



## Zaključak:

- Mreža **EmailEuCore** ima jedno jasno jezgro visokog k-core-a, gde je centralna grupa čvorova vrlo gusto povezana.
- Niske k-core-ove čine periferiju mreže sa retkim vezama, dok centralno jezgro predstavlja intenzivnu komunikacionu mrežu unutar institucije.
- Struktura je tipična za organizacione i društvene mreže: mali broj hub čvorova povezuje veliki broj perifernih čvorova u jedinstvenu, efikasno povezanu komponentu.

## 4.3 Facebook Athletes

### 4.3.1 Uvod

Mreža **FacebookAthletes** predstavlja verifikovane Facebook stranice sportista.

- **Čvorovi:** 13,866 → svaka stranica je jedan čvor.
- **Grane:** 86,811 → međusobne veze između stranica (mutual likes).
- **Petlje:** 47 - nisu značajne za analizu centralnosti i k-core, prilikom čitanja iz fajla se zanemaruju

### 4.3.2. Osnovne karakteristike mreže

- Prosečan stepen čvora: 12.52
- Gustina (density): 0.09% - mreža je vrlo retko povezana, što je očekivano za društvene mreže velikih razmara.
- Broj povezanih komponenti: 1 – povezani graf
- Maksimalni stepen čvpora: 31
- Minimalni stepen: 1(logično, inače ne bi bio povezatan graf)

### 4.3.3 K-Core analiza

#### **Niski k-core (0–6)**

- Broj čvorova se lagano smanjuje, dok gustina i prosečan stepen rastu.
- Gustina se povećava od 0.09% (0-core) do 0.25% (6-core).
- Prečnik nije izračunavan (graf je prevelik).

**Zaključak:** Niski k-core sadrži većinu čvorova, mreža je retko povezana, ali već povezana.

### Srednji k-core (7–15)

- Broj čvorova značajno opada: 6266 čvorova u 7-core do 812 u 15-core.
- Gustina raste do 3.419% u 15-core → mreža postaje gušće povezana.
- Prosečan stepen se povećava: od 19.7 do 27.7.
- **Prečnik 15-core:** 6 - mreža postaje dovoljno mala da se prečnik izračuna.

**Zaključak:** U ovom opsegu k-core počinje se formirati jezgro najpovezanih čvorova.

### Visoki k-core (16–22)

- Broj čvorova drastično opada, gustina raste:
  - 16-core: 446 čvorova, gustina 6.25%
  - 17-core: 245 čvorova, gustina 11.75% → počinje fragmentacija, 2 komponente povezanosit
  - 18-core: 150 čvorova, gustina 19.61% → 5 komponenti
  - 22-core: 106 čvorova, gustina 31.2%, 3 komponente
- **Prosečna dužina puta:** od ~3 (16-core) do ~1.1 (18–22-core)
- **Prečnik najveće komponente:** od 2 do 8

**Zaključak:** Visoki k-core pokazuje jezgro visoko povezanih stranica sportista, ali mreža počinje da se fragmentira. Jezgro nije više jedno povezano telo.

### Najviši k-core (23–31)

- Broj čvorova vrlo mali, gustina ekstremno visoka:
  - 28-core: 42 čvora, gustina 92%
  - 31-core: 39 čvorova, gustina 95% → skoro klika
- **Prosečna dužina puta:** ~1.05–1.08
- **Prečnik:** 2
- Mreža postaje praktično **kompletna među najvažnijim čvorovima**

**Zaključak:** Jezgro najpovezanih stranica sportista je gotovo klika, što ukazuje na malu grupu izrazito međusobno povezanih stranica.

#### **4.3.4 Spearmanovi koeficijenti korelacije**

Šel indeks vs stepen: 0.969

Šel indeks vs closeness: 0.751

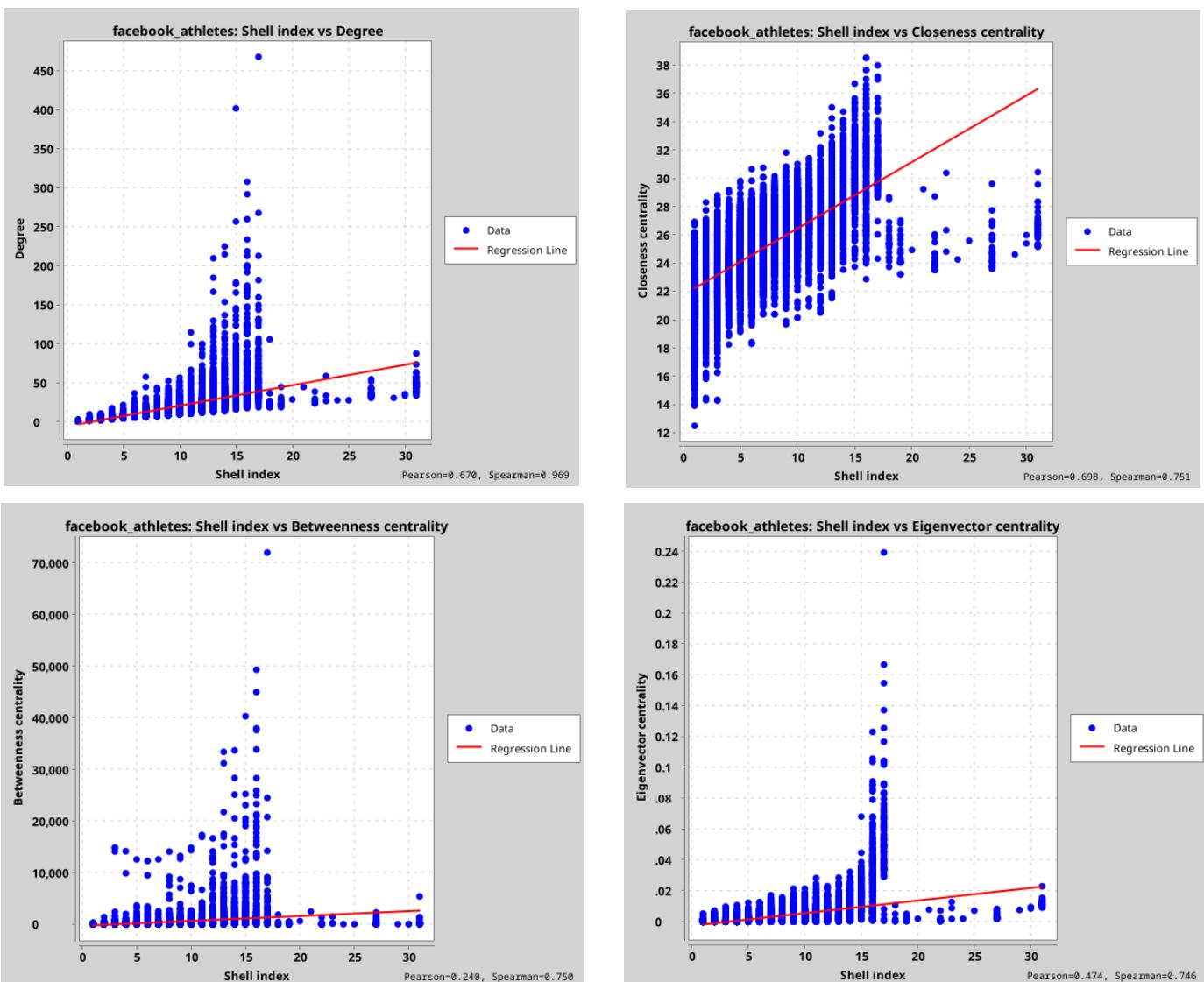
Šel indeks vs betweenness: 0.750

Šel indeks vs eigenvector: 0.746

Najjača veza je između šel indeksa i stepena (0.969) → mreža je skoro ekvivalentna stepen mreži (hubovi se nalaze u maksimalnom jezgru). Sve ostale korelacije (closeness, betweenness, eigenvector) su oko 0.75 → visoke, ali ne previsoke. Jezgro mreže nije samo numerički definisano stepenom, već i globalnim merama centralnosti – što znači da najvažniji čvorovi po različitim merama uglavnom pripadaju visokom k-core-u.

#### 4.3.5 Grafički prikaz

Grafički su prikazani čvorovi, tako da: jedan čvor je jedna tačka na grafikonu, na x osi šel indeks, na y osi su stepen/centralnost, redom: stepen, closeness, betweenness, eigenvector.



## Zaključak:

Mreža FacebookAthletes je retko povezana u većini čvorova, ali jezgro visoko povezanih stranica je gotovo kompletno.

- k-core analiza pokazuje jasno jezgro najvažnijih čvorova (stranica sportista) koje formira vrlo gusto povezanu komponentu.
- Šel indeksi i stepen čvorova su snažno povezani → mreža se ponaša tipično za društvene mreže: nekoliko hub čvorova sa mnogo veza, veliki rep čvorova sa malim stepenom.
- Visoko povezana jezgra bi mogla biti interesantna za studije informacione propagacije i identifikaciju uticajnih stranica u sportskom segmentu Facebooka.

Rađene su analize na još par mreža: FacebookArtists(mreža umetničkih stranica na facebooku), FacebookEgoSample, ZacharyKarateClub, CaliforniaRoads, male mreže iz poglavlja 2.

## Zaključak projekta:

U ovom projektu uspešno je implementiran Batagelj–Zaveršnik algoritam za k-core dekompoziciju u programskom jeziku Java, uz korišćenje JUNG biblioteke za rad sa grafovima. Testiranjem na različitim primerima, od veštački konstruisanih mreža do realnih velikih socijalnih mreža, potvrđena je tačnost i efikasnost algoritma. Rezultati pokazuju da šel indeksi pružaju korisnu meru strukturalne uloge čvorova u mreži i omogućavaju jasnu identifikaciju centralnih podmreža.

Projekat je napravljen tako da efikasno izračunava sve vrednosti vrlo brzo za srednje-velike socijalne mreže(500 – 5000 čvorova), za velike socijalne mreže se izračunavaju aproksimacije metrika centralnosti i preskaču se neka izračunavanja metrika. Izračunavanja metrika centralnosti na velikim socijalnim mrežama preko normalne metode bi trajala od nekoliko minuta do čak nekoliko sati, i iz tog razloga se preskače ovakav pristup.

Pored same implementacije, sprovedena je i analiza korelacije između stepena čvorova i njihovih šel indeksa, čime je dodatno naglašena razlika između lokalne mere (stepen) i globalnije strukturalne pozicije čvora (k-core). Dobijeni rezultati pokazuju da, iako često postoji jaka pozitivna povezanost između ove dve metrike, u određenim mrežama visok stepen ne garantuje i visok šel indeks, što naglašava prednost k-core analize.

Kao celina, projekat je pokazao praktičnu vrednost Batagelj–Zaveršnik algoritma za analizu socijalnih mreža i potvrdio da k-core dekompozicija predstavlja moćan alat za izdvajanje stabilnih i strukturno značajnih podgrupa u složenim mrežama.