

# Detekcija zajednica u socijalnim mrežama

Seminarski rad u okviru kursa

Računarska inteligencija

profesor: Aleksandar Kartelj

asistenti: Stefan Mišković, Denis Aličić

Aleksandra Nikšić  
mi16072

Anđelka Milovanović  
mi15145

Matematički fakultet

25. april 2020.

# Sadržaj

## 1. Uvod

### 1.1 Motivacija

## 2. Algoritmi

### 2.1 Girvan-Newman

### 2.2 Meme-Net

## 3. Rezultati

### 3.1 Karate Club

### 3.2 Dolphins

### 3.3 Jazz

## 4. Diskusija i zaključak

## 5. Pitanja?

# Motivacija

- Zašto je detekcija zajednica važna?
  - uočavanje strukture mreže
- Koji problemi postoje?
  - posmatramo samo eksplisitne veze između čvorova
  - pronađazak jače i slabije povezanih delova mreže
- Gde se može primeniti? U disciplinama u kojima su sistemi predstavljeni kao grafovi
  - sociologija
  - biologija
  - računarstvo

# Različiti pristupi

- Pristupi
  - detektovanje i uklanjanje grana sa ciljem razdvajanja klastera
  - genetski algoritmi, gde su jedinke nizovi čvorova sa određenim pripadnostima
- Naš pristup
  - GN sa gustinskom modularnošću
  - MemeNet sa heuristikom simuliranog kaljenja umesto lokalne pretrage

## Girvan-Newman Q

- klasična modularnost za GN:

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - P_{ij}) \delta(C_i, C_j) \quad (1)$$

- **naš pristup** za GN sa gustinskom modularnošću:

$$D = \sum_{i=1}^m \frac{L(V_i, V_i) - L(V_i, \bar{V}_i)}{|V_i|} \quad (2)$$

## Girvan-Newman

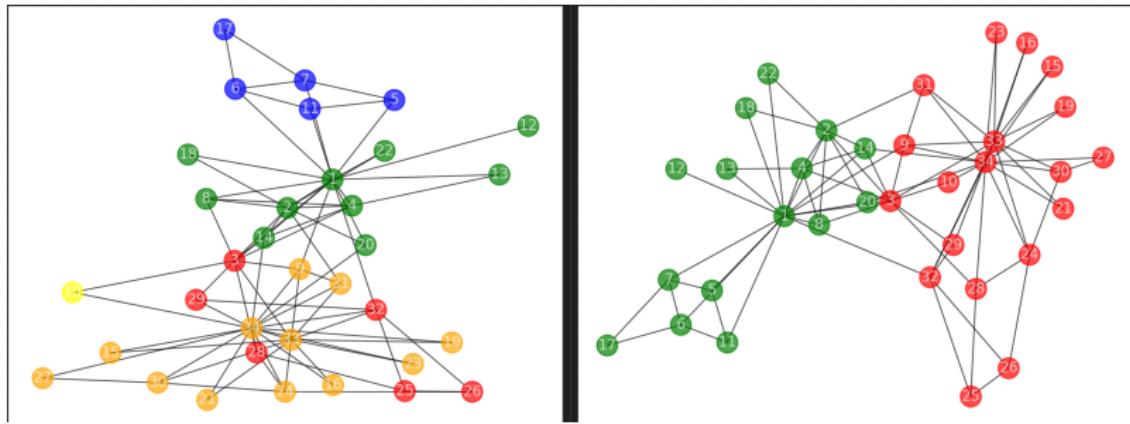
- ① postavi se najbolja modularnost na BestQ=0
- ② učita se graf i izračunaju se njegove komponente
- ③ izračuna se edge-betweenness za sve ivice i obrišu se sve one sa maksimalnom vrednošću (one su most između zajednica)
- ④ izračuna se novi broj komponenti grafa
- ⑤ ako je novi broj komponenti  $\leq$  od početnog onda se ponavlja korak 3
- ⑥ izračuna se modularnost i sačuva se u Q
- ⑦ ako je  $Q > \text{BestQ}$  onda se ažurira najbolja modularnost i sačuva se ta podela grafa kao najbolja u BestComps
- ⑧ ako nema više ivica u grafu vraća se BestComps, u suprotnom se ponavlja proces od koraka 3 na dalje

# Meme-Net

- rekombinacija gena + pojedinačne optimizacije
- optimizacija gustinske modularnosti
- podesivi parametar  $\lambda$
- simbioza genetskog algoritma i strategije lokalne pretrage
  - modularnost je fitnes funkcija
  - hromozom predstavlja particiju
- **naš pristup:** simulirano kaljenje umesto lokalne pretrage, gde se uzimaju komšijske particije sa nasumičnom vrednošću fitnesa (ne sa najboljom)

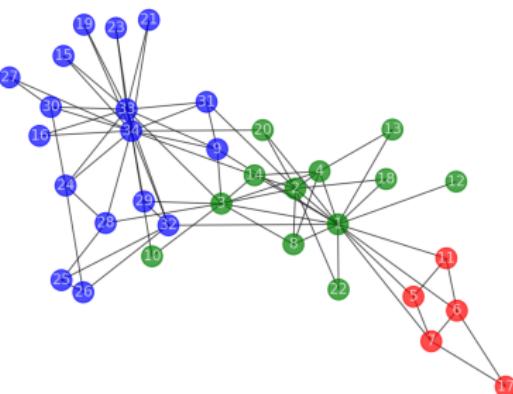
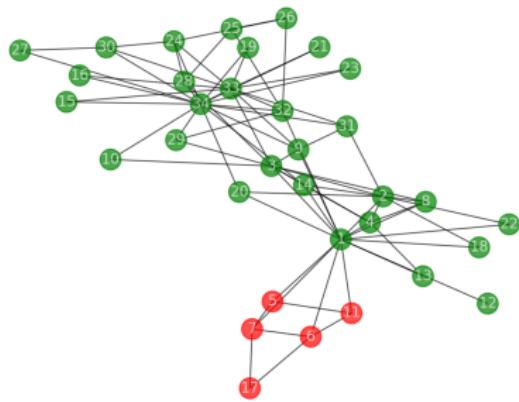
- normalizovana zajednička informacija
- mera sličnosti između stvarnih i algoritmom dobijenih particija
- particije A i B, matrica konfuzije
- $C_{ij}$  - broj čvorova u zajednici  $i$  particije A koji su takođe u zajednici  $j$  particije B

# Karate Club



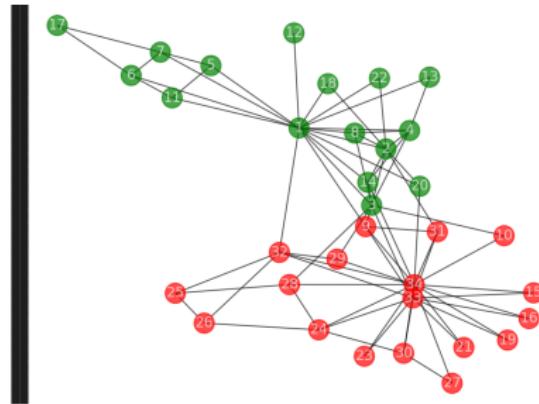
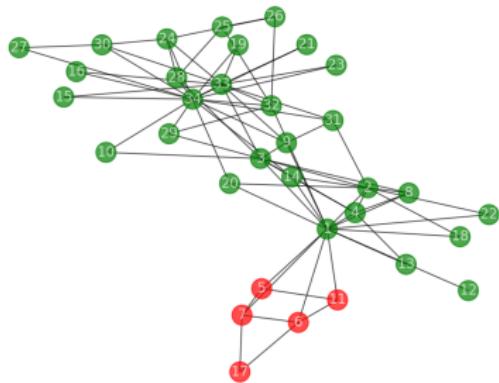
**Slika:** Sa leve strane slike je podrazumevani GN, dok je sa desne strane GN sa gustinskom modularnošću sa parametrom  $\lambda=0.3$

# Karate Club



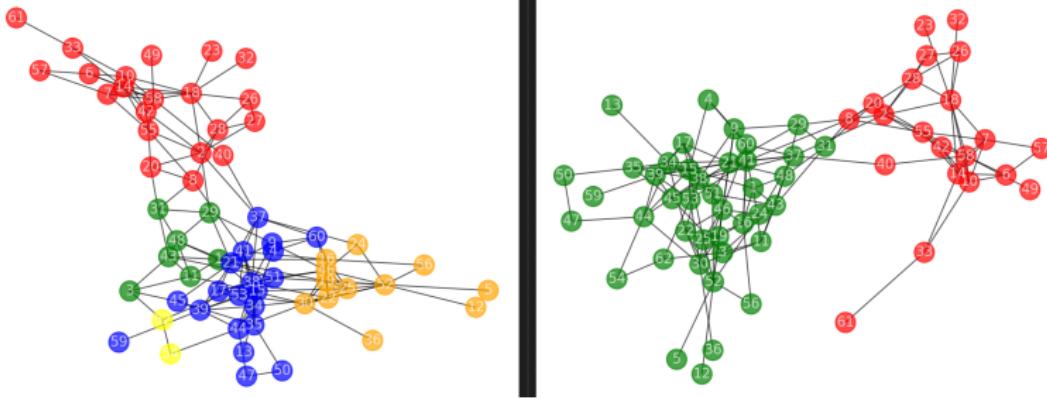
**Slika:** Sa leve strane slike je Meme-Net, dok je sa desne strane Meme-Net sa simuliranim kaljenjem (oba za 20 generacija) i  $\lambda=0.5$

# Karate Club



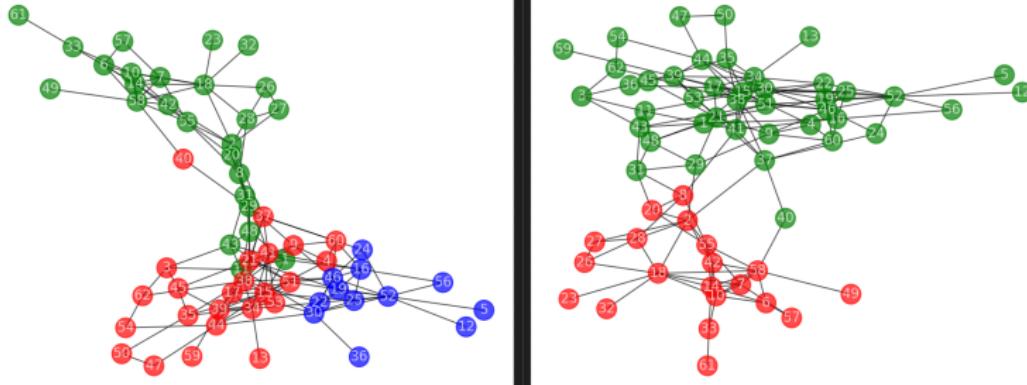
**Slika:** Sa leve strane slike je Meme-Net, dok je sa desne strane Meme-Net sa simuliranim kaljenjem (oba za 20 generacija) i  $\lambda=0.3$

# Dolphins



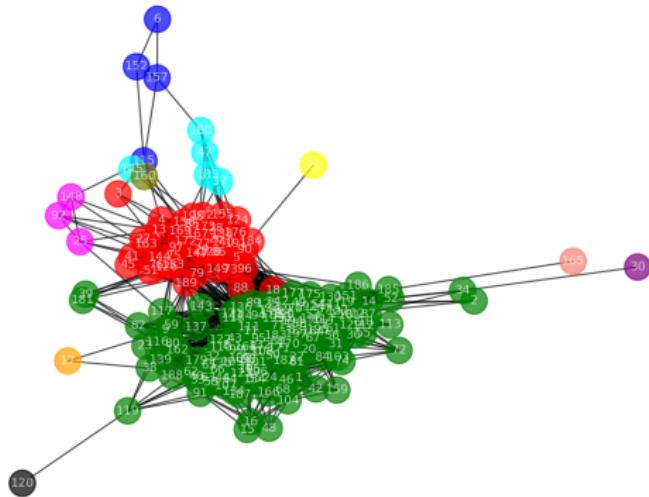
**Slika:** Sa leve strane slike je podrazumevani GN, dok je sa desne strane GN sa gustinskom modularnošću sa parametrom  $\lambda=0.3$ .

# Dolphins



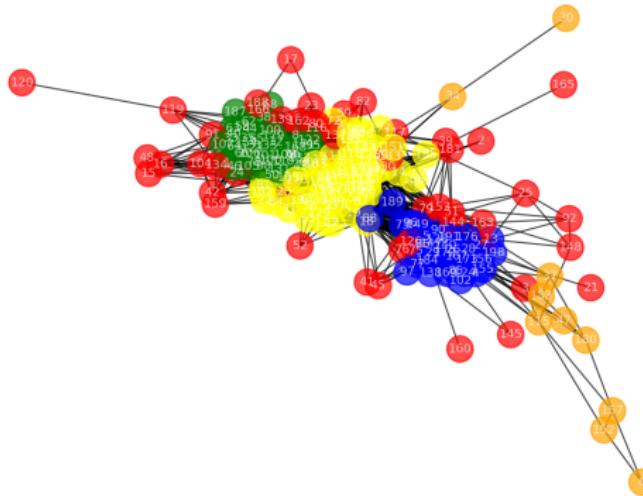
**Slika:** Sa leve strane slike je Meme-Net, dok je sa desne strane Meme-Net sa simuliranim kaljenjem (oba za 50 generacija) i  $\lambda=0.3$

# Jazz



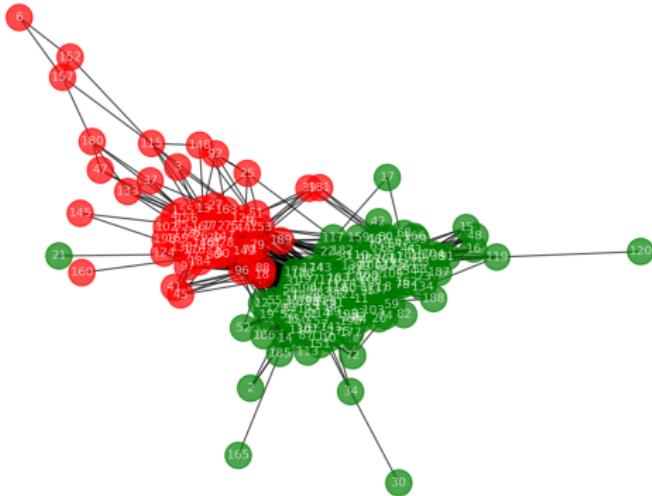
Slika: GN sa gustinom  $\lambda=0.7$

# Jazz



Slika: Meme-Net lokalna, 100 generacija,  $\lambda=0.7$

# Jazz



Slika: Meme-Net simulirano, 100 generacija,  $\lambda=0.7$

## Diskusija i zaključak

- ponašanje algoritma zavisi od problema
- rezlutati istraživanja pokazuju prednost gustinske modularnosti u odnosu na klasičnu (kod Girvan-Newman algoritma)
- za ispitane slučajeve konvergencija je brža ukoliko se koristi simulirano kaljenje, što ne mora biti opšti obrazac ponašanja (kod Meme-Net algoritma)
- dalji fokus istraživanja može biti primena ovih algoritama na veće zajednice, kao i testiranje u više istih iteracija kako bi se izvukao prosek

Hvala na pažnji.  
Pitanja?