

Analiza socijalnih mreža

Detekcija komuna u mreži

Marko Mišić, Jelica Protić

13M111ASM

2017/2018.

Proučavanje strukture mreže (1)

- U realnim mrežama, često se mogu uočiti gušće povezani delovi mreže
 - Određeni broj aktera koji su bolje međusobno povezani, nego sa ostalim članovima mreže
 - To su grupe ili zajednice (komune, *communities*)
 - Mogu biti odvojene ili delimično se preklapati
- Često je od interesa proučavati podstrukturu mreže
 - Može pomoći u razumevanju ponašanja mreže kao celine
 - Grupe koje se preklapaju imaju mnogo manje šanse da razviju konflikte
 - Način širenja ideja i informacija je takođe zavistan od postojanja grupa u okviru mreže

Proučavanje strukture mreže (2)

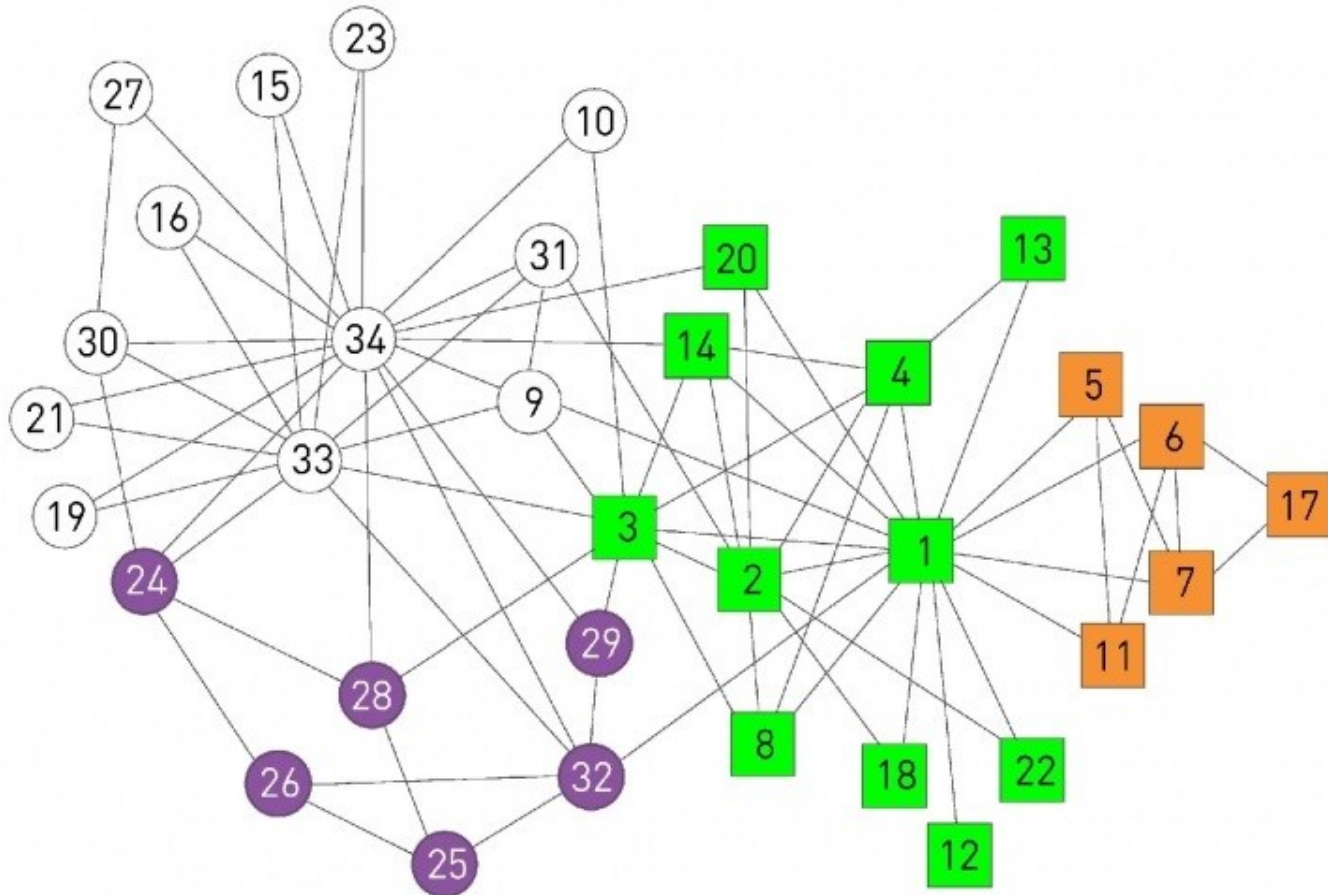
- Pojava podstruktura u okviru mreža je posledica mreža afilijacija (*affiliation networks*)
- U realnom svetu, pojedinci su povezani na mnogo načina
 - Mestom ili naseljem u kojem žive
 - U okviru organizacija u kojima rade
 - Putem događaja kojima su prisustvovali
 - Članstvom u organizacijama i klubovima
 - Deljenjem zajedničkih vrednosti, bavljenje istom aktivnošću
 - Pripadnošću rasnim, etničkim, verskim ili drugim skupinama
- Ovakve veze se mogu predstaviti bipartitnim grafovima
 - Transformacijom ovakvih grafova u unipartitne, pojavljuju se grupe

Zachary's Karate Club (1)

- Klasičan primer u detekciji komuna
 - 34 člana koji su se svi međusobno poznavali
 - Detektovano 78 međusobnih regularnih interakcija članova kluba van kluba
- Nakon konflikta između predsednika i glavnog instruktora, klub se raspao na dva dela
 - Sledeći podelu koja je teorijski predvidljiva
- Često korišćen skup podataka za algoritme koji rade detekciju komuna

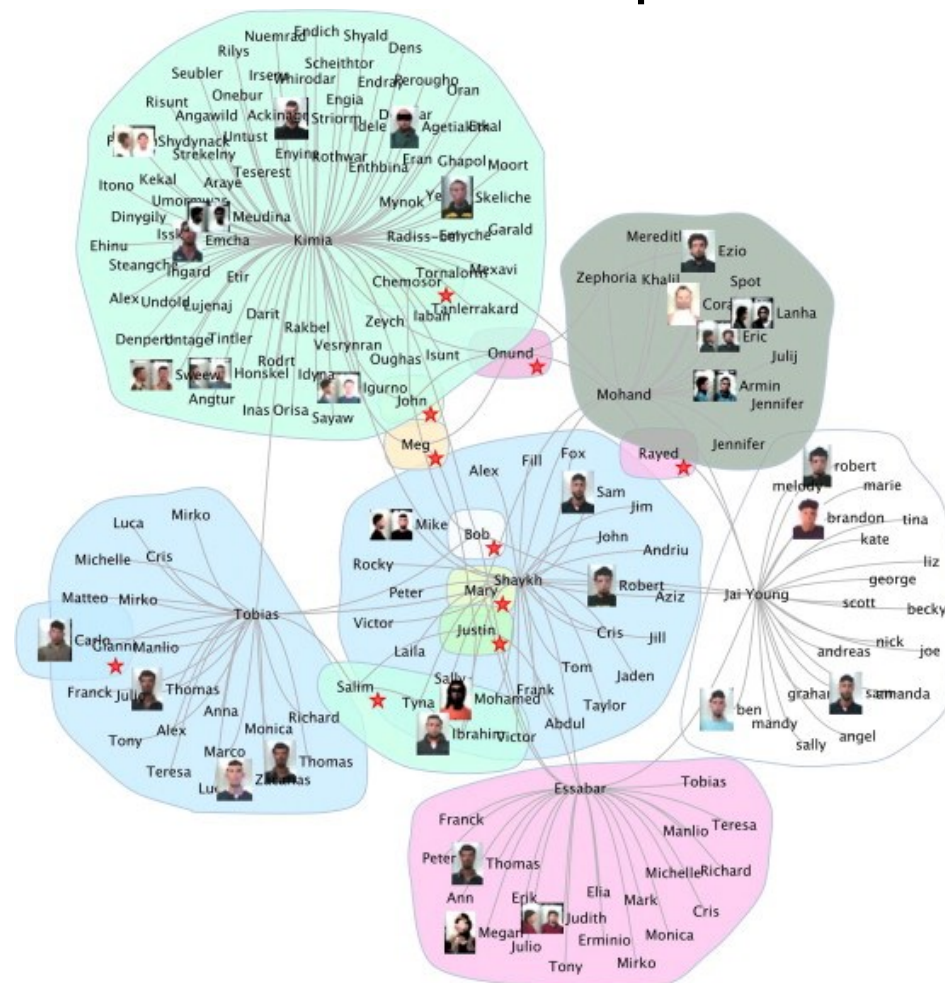
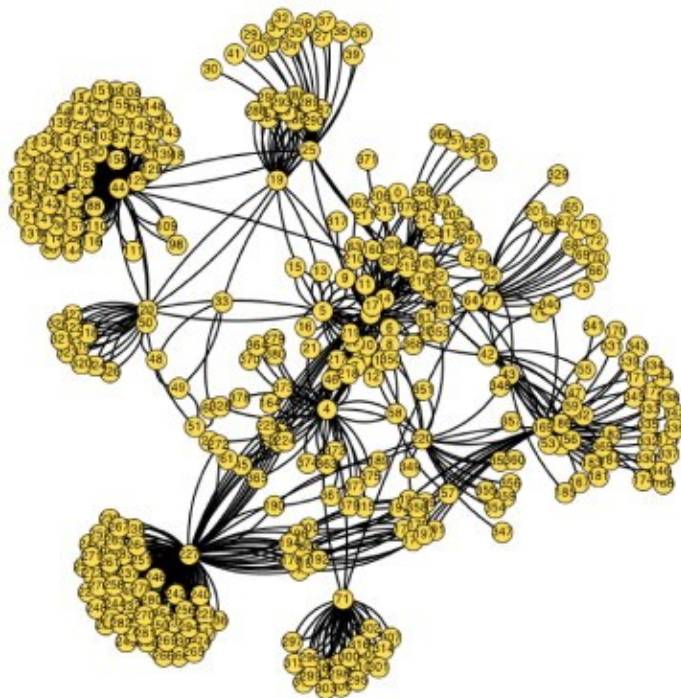
Zachary's Karate Club (2)

- Dve komune – zelena i ostali



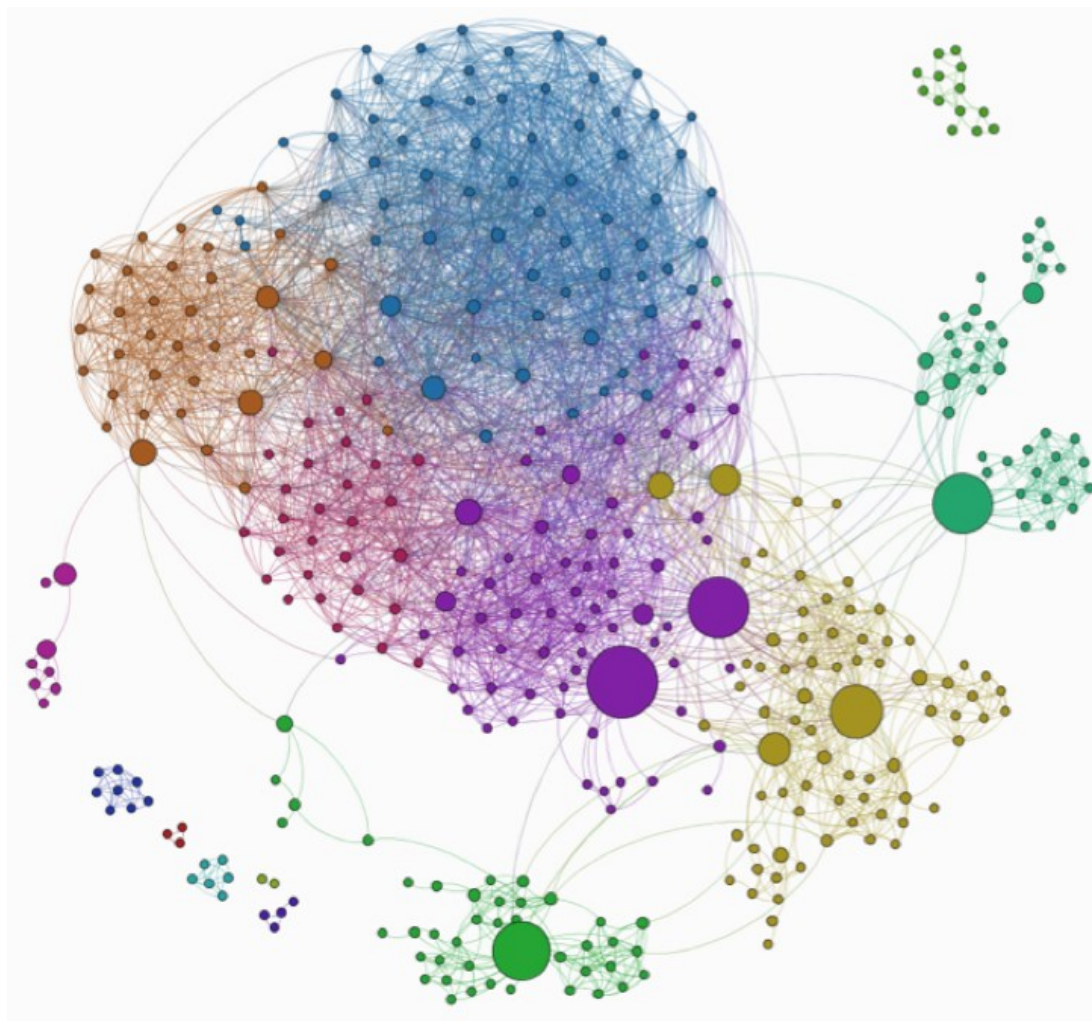
Mreža kriminalaca

- Mreža kriminalaca zasnovana na telefonskim pozivima



Ego mreža korisnika Facebook-a

- The blue community is made up of people I met at school, all around my age
- The brown community is people I went to school with, but also lived close to where I grew up
- The maroon community also went to school with me, but all at least a year older than me
- The purple community is people I attended college with directly after finishing school



Homofilija i asortativnost (1)

- Stvaranje komuna u mrežama je posledica homofilije (*homophily*) i asortativnosti (*assortativity*)
- Homofilija predstavlja tendenciju ljudi da stvaraju veze sa sličnim ljudima
 - Sličnost uključuje socijalni status i karakteristike kao što su rasa, nacija, starost, pol, klasa i sl.
 - U evolutivnoj biologiji poznata pojava *assortative mating*
- U analizi socijalnih mreža homofilija se javlja u formi asortativnosti
 - Izučava se u smislu stepena čvora
 - Čvorovi imaju tendenciju da se povezuju sa čvorovima sličnog stepena
 - Izražava se kroz koeficijent asortativnosti
 - Karakteristično samo za socijalne mreže

Homofilija i asortativnost (2)

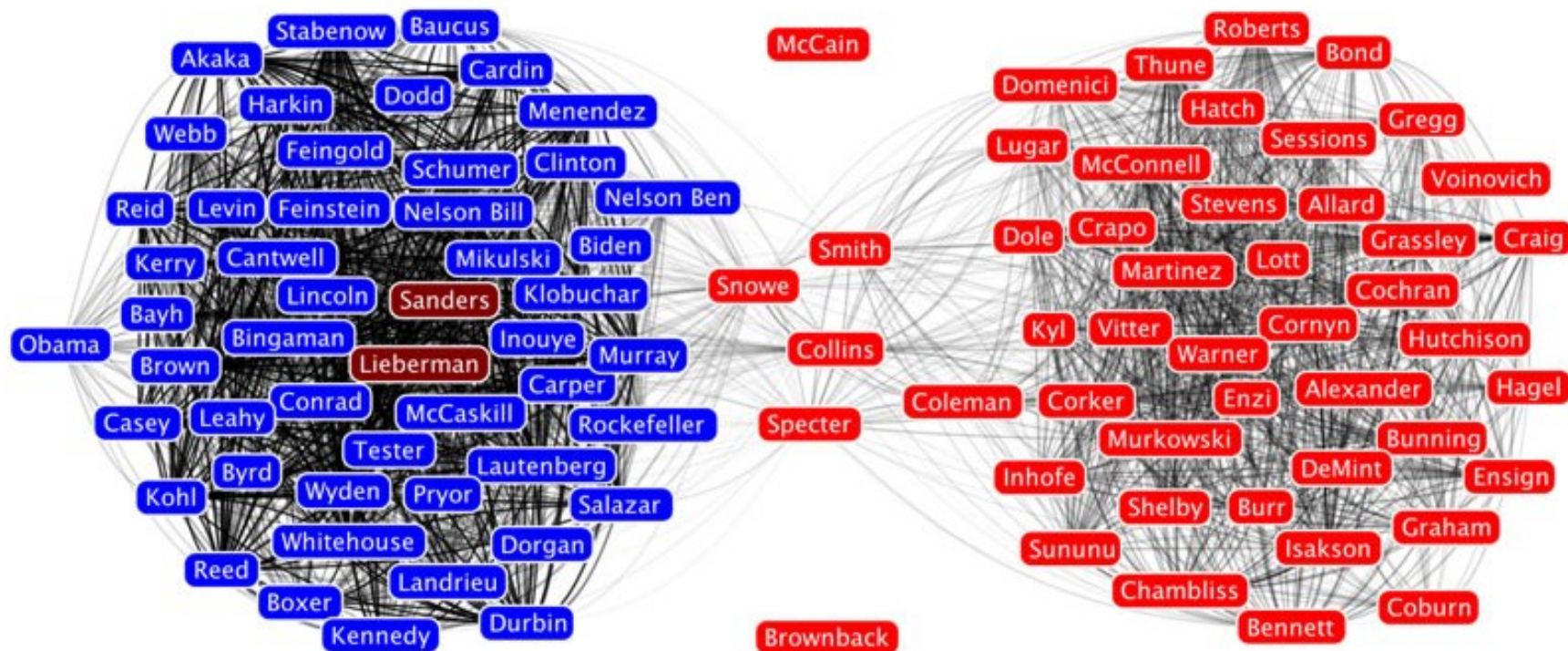
- Brojni razlozi za grupisanje po sličnostima
- Geografski razlozi
 - Ljudi koji žive zajedno lakše stupaju u kontakt
- Rodbinske veze
 - Ljudi ih održavaju čak i na velikim udaljenostima
- Pripadnost organizacijama
 - Škole, posao, članstvo u različitim organizacijama
- Izomorfni izvori
 - Često se povezuju ljudi koji su na ekvivalentnim strukturnim pozicijama
 - Profesori fakulteta, dekani, direktori, majke, HR menadžeri
- Kulturni razlozi, uverenja i mišljenja
 - Ljudi iz sličnog demografskog i kulturnog miljea i zajedničkog nasleđa se lakše povezuju i komuniciraju

Homofilija i asortativnost (3)

- Glavni efekti homofilije
 - Povećava toleranciju i saradnju
 - Pomaže ljudima da pristupe informacijama
 - Utiče na usvajanje socijalnih normi i ponašanja
 - Vremenom postajemo sličniji ljudima sa kojima smo povezani
 - Omogućava se lakše ostvarivanje društvenog uticaja
- Heterofilija predstavlja tendenciju ljudi da usvajaju nova iskustva u kontaktu sa drugačijim ljudima
 - Suprotnost homofiliji
 - Podstiče inovativnost i razmenu ideja

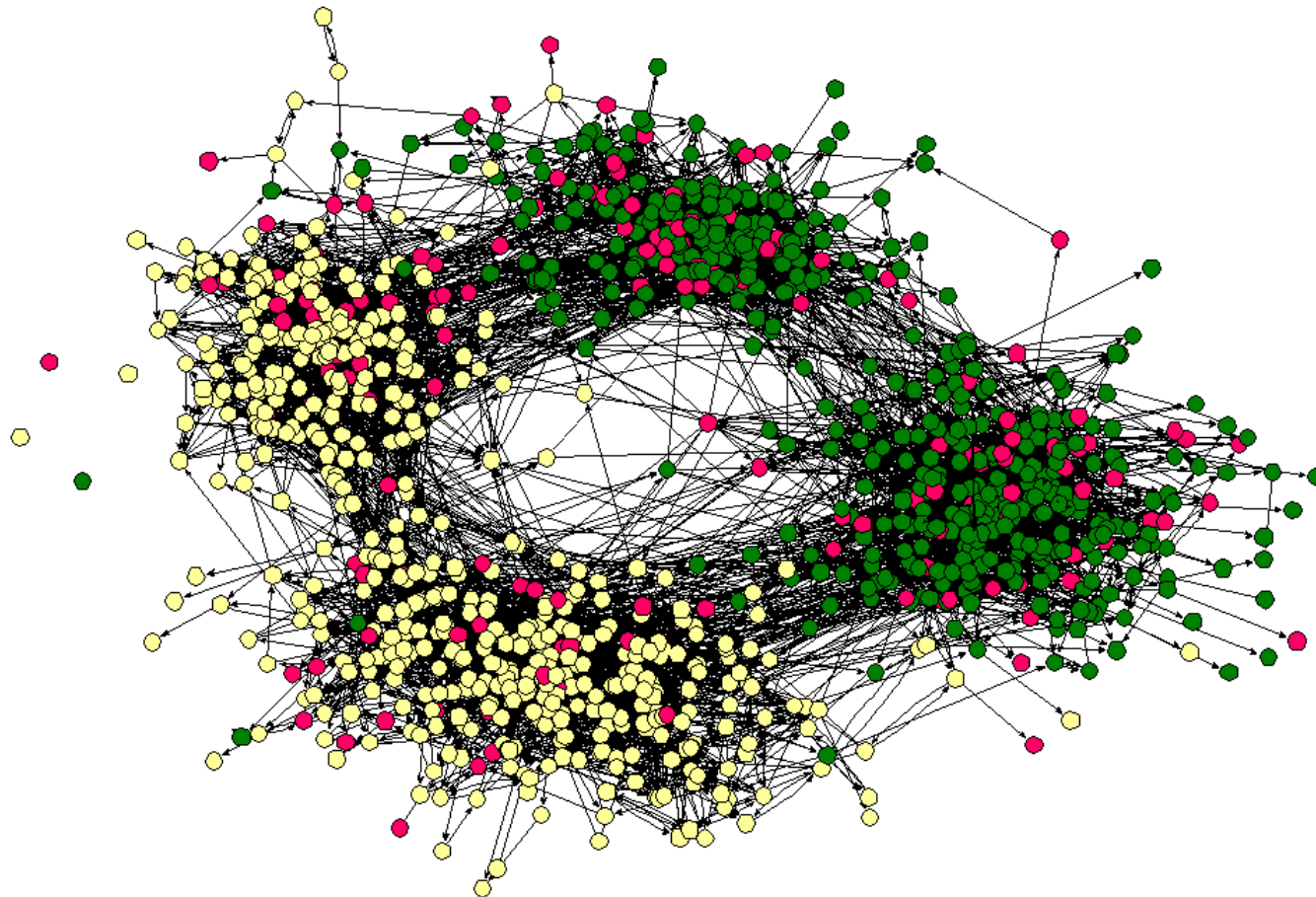
Homofilija i asortativnost (4)

- Glasanja u američkom senatu 2007. godine



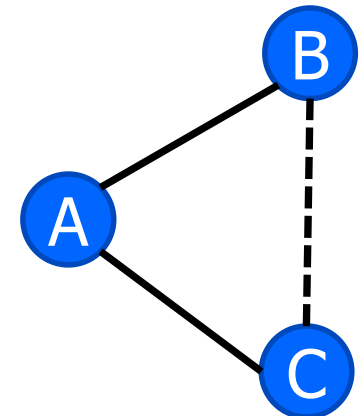
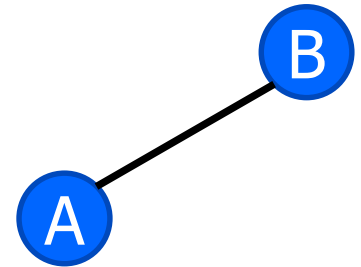
Homofilija i asortativnost (5)

- High school students by race



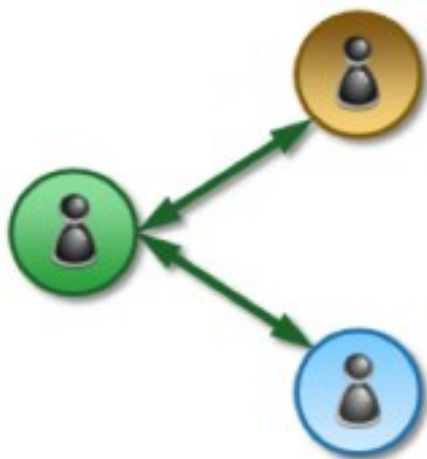
Osnovne mrežne jedinice (1)

- Osnovne jedinice u okviru mreže su dijade (*dyad*) i trijade (triad)
- Dijadu čini par aktera koji mogu biti povezani granom
- Trijadu čine tri aktera i njihove moguće veze
- Za proučavanje strukture mreže su od interesa dijade i trijade kod kojih postoji povezanost među čvorovima

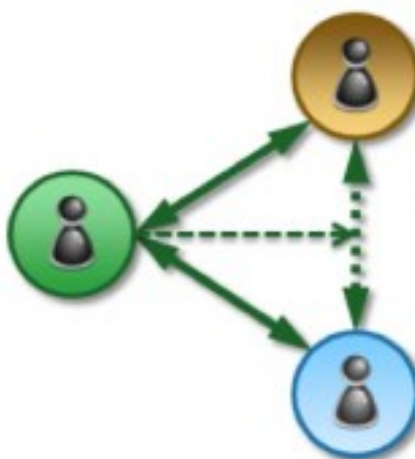


Osnovne mrežne jedinice (2)

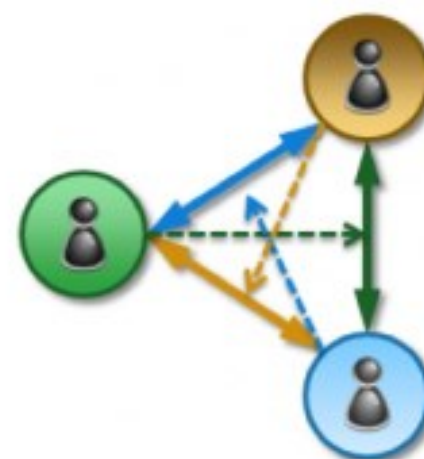
- Primer formiranja trijade od dve dijade



Moe knows Larry and Curly. He also knows that Larry could help Curly prepare a comedy routine.



Moe introduces Larry and Curly and helps them build their relationship



The relationship between Larry and Curly stabilizes, everyone becomes responsible for the triad.

Pojam komune (1)

- Grupa čvorova koja je gušće povezana unutar sebe nego sa ostatkom mreže
 - Termini klaster i komuna se u analizi socijalnih mreža mogu smatrati sinonimima
- Podela na komune se može raditi na osnovu:
 - Prisutnosti direktnih veza među članovima
 - Direktna povezanost sa svim ostalim članovima
 - Direktna povezanost sa određenim minimalnim brojem članova
 - Povezanosti članova u maksimalnom broju koraka
 - Polazeći od jednog člana do ostalih članova grupe se može stići u garantovanom broju koraka
 - Odnosa gustina konekcija unutar grupe i prosečne gustine konekcija u mreži

Pojam komune (2)

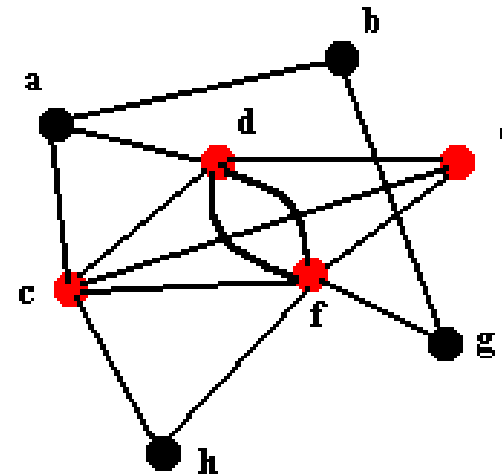
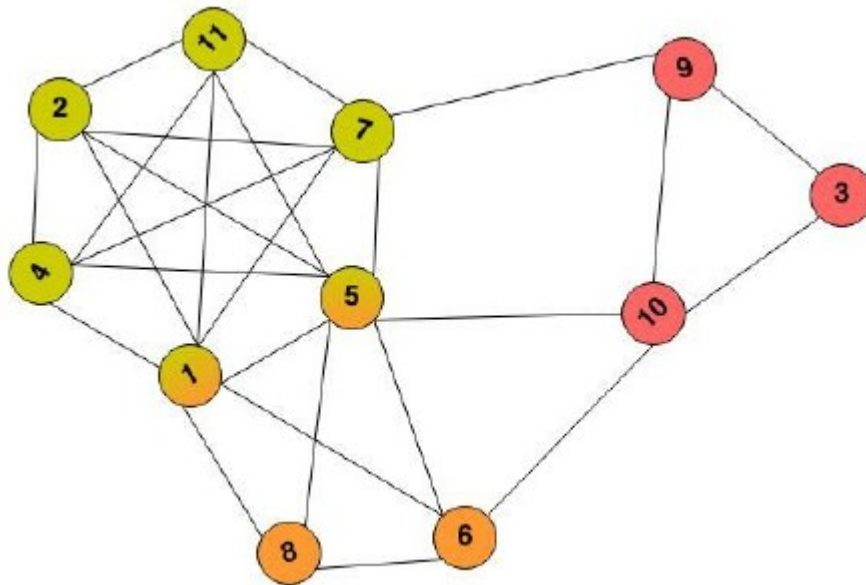
- Definicija na osnovu intra-klaster i inter-klaster gustina
 - Neka je C podgraf grafa $G(V,E)$
 - Intra-klaster gustina podgraфа C je odnos broja internih veza među čvorovima u C i broja svih mogućih internih veza
 - Inter-klaster gustina podgraфа C je odnos broja eksternih veza čvorova u C ka ostatku grafa i maksimalnog broja eksternih veza
 - Da bi C bio komuna, intra-klaster gustina je očekivano da bude značajno veća od prosečne gustine mreže
- Veći broj specifičnih definicija
 - Klike, pleksovi, klanovi, povezane komponente

Klika i varijante (1)

- Klika (*clique*) je podgraf posmatranog grafa koji se sastoji od najvećeg mogućeg skupa čvorova koji su međusobno povezani svaki sa svakim
 - Mora postojati direktna grana između svih članova
 - Predstavljaju maksimalan kompletan podgraf posmatranog grafa
 - U okviru jednog grafa može biti identifikovano više klika
 - Klike se mogu međusobno preklapati
 - Jedan čvor može pripadati većem broju klika

Klika i varijante (2)

○ Primer klika



Cliques

a,b
a,c,d
b,g
c,d,e,f
c,f,h
f,g

Klika i varijante (3)

- Previše stroga definicija klike
 - Ukoliko nedostaje samo jedna veza između dva člana neke grupe, onda više nije klika
 - Sa stanovišta analize, svi čvorovi imaju istu poziciju unutar grupe
 - Niz alternativnih, relaksiranih definicija
- N -klika je skup čvorova koji su međusobno povezani bilo direktno, bilo putevima dužine N
 - N obično uzima vrednost dva
 - N -klan, ukoliko svi putevi među akterima moraju biti dužine $\leq N$
- K -pleks se definiše kao skup čvorova kod kojeg pojedinačni čvorovi imaju veze sa svim čvorovima, osim nekih K čvorova
- K -jezgra podrazumevaju postojanje najmanje K veza ka ostalim članovima grupe

Povezane komponente (1)

- Povezana komponenta (*connected component*) mreže predstavlja maksimalni skup čvorova (podgraf) takvih da postoji put između bilo koja dva čvora u mreži
 - Slabo povezane komponente
 - Ukoliko postoje izolovani čvorovi, oni takođe predstavljaju zasebne komponente
- Kod usmerenih grafova se razlikuju termini slabo i jako povezanih komponenti
 - Jako povezanu komponentu čine čvorovi kod kojih postoji dostižnost između bilo koja dva čvora komponente
- Broj povezanih komponenti se odnosi na ukupan broj slabo povezanih komponenti
 - Ponekad se koristi kao metrika za opis mreže

Povezane komponente (2)

- Mogu se odrediti i u težinskim mrežama
 - Podaci o težini grana se koriste za dihotomizaciju grafa
 - Polazi se od težine najveće grane
 - U svakom koraku se dobija po jedna podela na komponente
 - Algoritam iterira dok se ne dobije jedna komponenta
- Rezultat je dendrogram sa podelom
- Podela na osnovu teorije jakih i slabih veza
 - Granoveter, 1973.
 - Biće obrađeno kasnije

Povezane komponente (3)

- Tačke prekida (*cut points*)
 - Čvorovi grafa čije bi uklanjanje dovelo do podele mreže na dve ili više komponenti
- Blokovi (*blocks, bi-components*)
 - Grupe koje bi se formirale uklanjanjem tačaka prekida iz grafa
- Lambda skup
 - Skup veza čije bi uklanjanje dovelo do podele mreže na manje nepovezane komponente
 - Grane sa visokim *edge betweenness* skorom

Metrike za procenu strukture mreže (1)

- *Clustering coefficient*
 - Lokalni i globalni
- Lokalni koeficijent klasterizacije se računa kao gustina mreže koju čine posmatrani čvor, njegovi susedi i njihove međusobne veze
 - *Ego* mreža posmatranog čvora
 - Stepen klasterizacije kompletne mreže se računa na osnovu proseka pojedinačnih čvorova
 - Ispituje tendenciju formiranja klika
- Globalni koeficijent klasterizacije (alternativa)
 - Posmatra zatvorene trijade unutar mreže u odnosu na sve moguće povezane trijade

Metrike za procenu strukture mreže (2)

- Modularnost (*modularity*) je mera kvaliteta particionisanja čvorova mreže u odgovarajuće klastere
 - Mreže sa većom modularnošću su podeljene na klastere koji su, unutar sebe, gusto povezani, a nemaju veliki broj veza sa ostatkom mreže
- Definiše se kao odnos broja grana u određenim klasterima u odnosu na ukupan broj grana umanjen za isti takav odnos koji bi se dobio kada bi grane između čvorova bile raspoređene na slučajan način

Tehnike za detekciju komuna (1)

○ *Top-down* pristupi

- Polaze od kompletne mreže i pokušavaju da izvrše hijerarhijsku podelu odozgo na dole
- Najčešće kao rezultat daju dendogram koji prikazuje grupisanje manjih jedinica u veće
- Analiza povezanih komponenti, *Girvan-Newman* metod

○ *Bottom-up* pristupi

- Grade veće grupe polazeći od manjih gradivnih jedinica odozdo na gore
- Propagacija labela, *Louvain* metod i sl.

Tehnike za detekciju komuna (2)

- Particijsko grupisanje
 - *k-means clustering* i slične metode
- Hijerarhijska klasterizacija
 - Aglomerativni algoritmi
 - *Ravasz* algoritam
 - Razorni algoritmi
 - *Girwan-Newman* metoda
- Propagacija labela
 - Sinhrona, semisingrona i asinhrona propagacija
 - Balansirana propagacija

Literatura

- Hanneman, Robert A. and Mark Riddle, Introduction to social network methods, University of California, Riverside, 2005.
- J. Jovanović, Softverska analiza društvenih mreža, FON, 2017.
- M. Mišić, Unapređenja sistema za detekciju plagijarizma u izvornom programskom kodu – Primena metoda za analizu socijalnih mreža u detekciji plagijarizma, PhD disertacija, ETF 2017.
- B. Hoppe, C. Reinelt, Social Network Analysis and the Evaluation of Leadership Networks, Leadership Quarterly, 21, 2010. pp. 600-619.
- <http://www.network-science.org/>
- <http://social-dynamics.org/homophily/>
- <https://www.pulsarplatform.com/blog/2014/detecting-communities-using-social-network-analysis/>