

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

Zajednice u mrežama

Mario Komljenović

Voditelj: *Dr.sc. Mile Šikić*

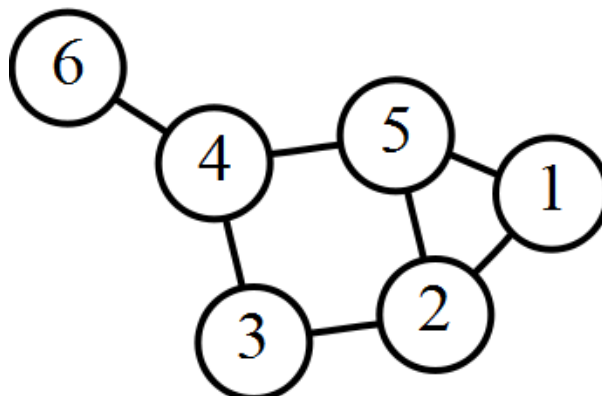
Zagreb, svibanj 2009.

Sadržaj

1. Uvod – Mreže i zajednice	1
2. Algoritmi za pronalazak zajednica u mrežama	2
2.1. Standardne metode clusteringa	2
2.2. Kernighan-Linov algoritam	3
2.3. Pronalazak zajednica baziran na centralnosti	4
2.4. k-grupno filtriranje i ostale lokalne metode	4
2.5. Optimizacija modularnosti	5
2.6. Spektralna podjela.....	6
2.7. Pottsova metoda	7
2.8. Parametar razlučivosti	8
3. Primjena	10
3.1. Znanstvena suradnja.....	10
3.2. Zajednice korisnika mobitela	10
3.3. <i>Online</i> društvene mreže	11
4. Zaključak.....	13
5. Literatura	14
6. Sažetak	15

1. Uvod – Mreže i zajednice

Pojam *mreže* intuitivno je lako shvatljiv, no teško ga je formalno definirati. S apstraktnog gledišta, mogli bismo reći da je mreža ništa drugo doli *matematički graf* – objekt koji se sastoji od skupa čvorova i skupa veza koji povezuju te čvorove [4]. Grafovima možemo prikazati različite prirodne i društvene pojave pa tako i zajednice. *Zajednica* je skupina zavisnih entiteta koji međusobno dolaze u kontakt, a grafički ju možemo prikazati skupinom čvorova koji su relativno gusto povezani jedan s drugim, ali su labavo povezani s drugom skupinom čvorova – drugom zajednicom [1].



Slika 1.1. Graf sa 6 čvorova i 7 veza [4]

Skupine primitivnih lovaca-sakupljača, feudalne strukture, kraljevske obitelji, političke i poslovne organizacije, obitelji, sela, gradovi, države, pa čak i Facebook grupe – sve su to primjeri društvenih zajednica. Već navođenjem ovih, međusobno potpuno različitih primjera daje se naslutiti da je proučavanje zajednica složen proces koji zadire u područje raznih znanosti. Cilj takvog proučavanja je shvatiti kako, gdje i zašto se zajednice formiraju te na temelju interakcija pojedinih članova (mikroskopska razina) pokušati predvidjeti interakcije kompletne zajednice (makroskopska razina).

No, da bismo zajednice mogli proučavati, moramo ih prvo otkriti, pronaći. To je poprilično složen problem koji od svakog sudionika koji ga rješava zahtijeva interdisciplinarnu suradnju, ali i vlastito bogato znanje o specifičnom području.

U ovome ću se seminarskome radu prvenstveno baviti upravo problemom pronalaska zajednica u mrežama, s naglaskom na već postojeće algoritme koji rješavaju taj problem.

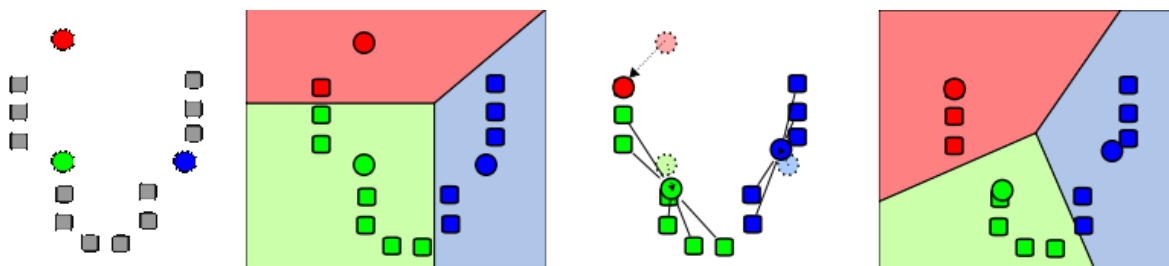
2. Algoritmi za pronalazak zajednica u mrežama

Zajednica je kohezivna grupa čvorova koji su gušće povezani jedni sa drugima nego s čvorovima u ostalim zajednicama. Razlike između algoritama za pronalazak zajednica u mrežama uglavnom se svode na preciznu definiciju „gušće povezanosti” i korištenje algoritamskih metoda temeljenih na toj definiciji. Do danas je otkriveno mnogo ovakvih algoritama, od kojih ću pomnije analizirati nekoliko najkorištenijih.

2.1. Standardne metode clusteringa

Zamisao o organizaciji podataka metodom grube podjele na veće skupine po principu zajedničkih svojstava poprilično je stara. Prve računalne pokušaje pronalaska *cluster*a (grozdova, skupina) sličnih objekata pronalazimo u statistici i inteligentnoj analizi podataka (engl. *data mining*).

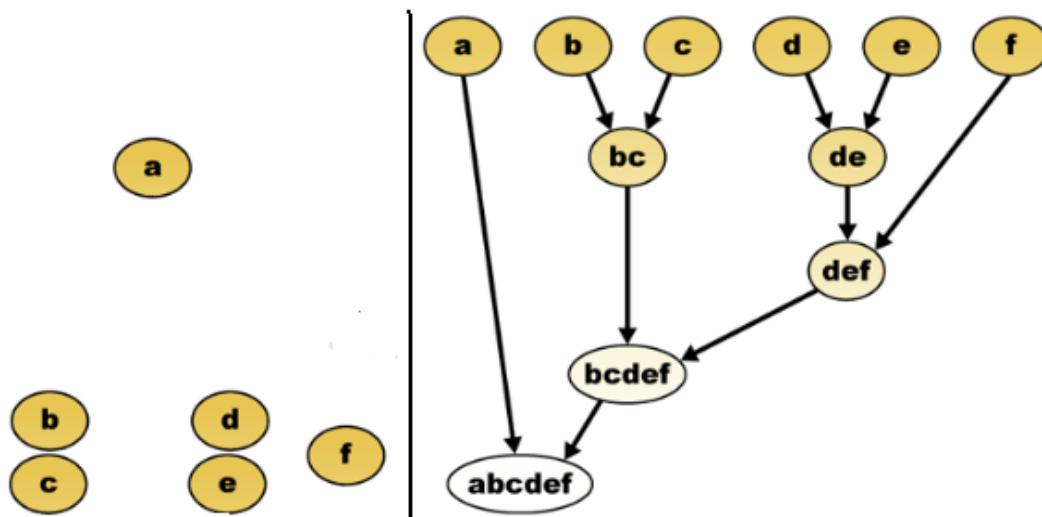
Među značajnije metode ubrajamo *metode partitivnog clusteringa* (engl. *partitional clustering techniques*), npr. *k-stožerni clustering* (engl. *k-means clustering*, podjela n podataka na k particija tako da svaki podatak pripada particiji u kojoj se nalazi njemu najbliži stožer (unaprijed odabran referentni podatak)) te *metode višedimenzionalnog skaliranja* (engl. MDS – *multi-dimensional scaling*), npr. *razgradnja pojedinačnih vrijednosti* (engl. SVD – *singular value decomposition*) i *analiza glavnih komponenti* (engl. PCA – *principal component analysis*) [1]. Primjera radi, MDS algoritmi pokazali su se iznimno uspješnima u pronalaženju clustera sličnih podataka te se koriste u mnogim područjima – npr. kod izglasavanja zakona. Za ovaj konkretan primjer postupak je slijedeći: stvara se tablica koja pokazuje kako je svaki pojedini glasač glasao za svaki pojedini zakon, ta se tablica zatim pojednostavljuje, a zatim se iz nje iščitava kakva su razmišljanja pojedine političke struje i kako se svaki pojedinačni glasač slaže s ostalim glasačima.



Slika 2.1. Postupak *k-stožernog clusteringa*: izabiremo k (u ovome slučaju 3) nasumičnih stožera (prva slika), formiramo *cluster*e (druga slika), središnji podatak u svakom od *cluster*a postaje novi stožer (treća slika), postupak ponavljamo (četvrta slika) sve do pojave konvergencije [5]

Svakako treba spomenuti i *algoritme hijerarhijskog clusteringa* (engl. *hierarchical clustering algorithms*), npr. *metode clusteringa vezama* (engl. *linkage clustering methods*) koje se koriste u filogenetičkoj biologiji [1]. Kod ovog se algoritma započinje s mrežom koja se sastoji od N zasebnih stupnjevanih čvorova. Ta je mreža opisana matricom čiji članovi A_{ij} označavaju koliko su čvorovi i i j blisko

povezani. Nakon toga, čvorovi se sekvencijalno udužuju u veće clusterse, počevši s parom koji ima najveći iznos A_{ij} (par s najjačom vezom). U svakom slijedećem koraku uspoređuju se novi clusteri s već stvorenima te se ponovno spajaju najbliži. Postupak se provodi sve dok svi clusteri čija je sličnost različita od nule ne postanu povezani. Način na koji se mjeri sličnost clustera različit je za svaku pojedinačnu metodu. Kod clusteringa jednostrukim vezama sličnost clustera X i Y je definirana kao najveća sličnost između bilo kojeg para čvorova $x \in X$ i $y \in Y$. Kod clusteringa bitan je i raspored clustera, a njega najlakše možemo prikazati *dendrogramom* (prikaz u obliku stabla) čija dubina označava korake u kojima su neka dva clustera spojena.



Slika 2.2. Hijerarhijski clustering: prije (lijeva slika) i poslije (desna slika) postupka [6]

Postoji i nekoliko klasičnih tehnika divizije kod kojih se počinje s potpunim grafom koji se zatim „lomi” na manje dijelove s ciljem pronalaženja zajednica.

Sve navedene metode i dalje se uporno unapređuju te samim time iz dana u dan daju sve preciznije rezultate.

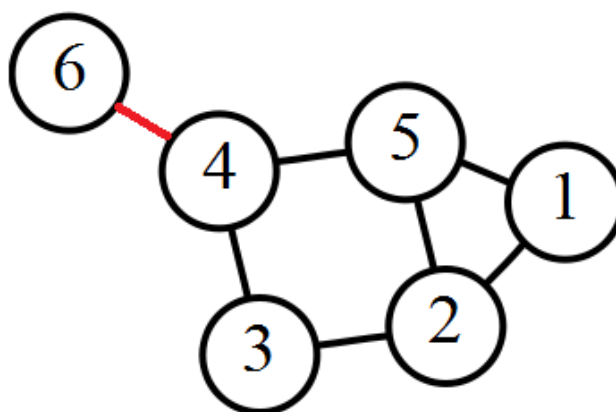
2.2. Kernighan-Linov algoritam

Ovaj je algoritam nastao na području računalne znanosti, a stvorili su ga Brian Kernighan i Shen Lin 1970. u okviru istraživanja kojim su željeli otkriti kako rasporediti električne sklopove na ploče tako da čvorovi na različitim pločama mogu biti spojeni jedan s drugim uporabom najmanjeg mogućeg broja veza [1]. Da bi to ostvarili, maksimizirali su kvalitativnu funkciju Q koja dovodi u vezu broj veza u svakom skupu čvorova i broj veza između različitih skupova. Počevši s inicijalnom podjelom grafa na dvije grupe unaprijed određene veličine, Kernighan-Linov algoritam u svakom svom koraku zamjenjuje podskupove koji sadrže jednak broj veza između dvije grupe. Da bi se smanjila šansa zapinjanja na lokalnom maksimumu, Kernighan-Linov algoritam dopušta zamjene koje umanjuju iznos funkcije Q . Nakon određenog broja zamjena particija s najvećim iznosom funkcije Q koristi se kao početni uvjet za novi niz koraka. Kada broj i veličina zajednica nisu određeni, najčešće korišten princip je pomicanje za jedan čvor u jednome koraku. Kvaliteta podjele pomoću Kernighan-Linovog algoritma ovisna je o početnoj particiji i

upravo iz tog razloga ovaj je algoritam najbolje koristiti kao dodatak nekoj drugoj metodi koja daje visoko kvalitetne particije.

2.3. Pronalazak zajednica baziran na centralnosti

Michelle Girvan i Mark Newman proslavili su se na području matematike i statističke fizike kada su osmislili algoritam za pronalazak zajednica baziran na principu *međupoložene centralnosti* (engl. *betweenness centrality*) [1]. Veza ima veliku *međupoloženost* (engl. *betweenness*) ako leži na velikom broju puteva između čvorova [1]. Počnemo li od nekog čvora i poželimo li stići do nekog drugog čvora u mreži, očito je da će neke veze biti prometnije od drugih. Međupoloženost veze je mjera te prometnosti, uzimajući u obzir isključivo najkraće puteve (*geodezijska međupoloženost*) ili pak gustoću nasumičnih puteva (*međupoloženost nasumičnih puteva*) između svakog para čvorova ili prosjeka svih mogućih parova čvorova [1]. Zajednice se mogu identificirati procesom rangiranja svake veze na temelju njezine međupoloženosti, zatim uklanjanja veze s najvećom vrijednošću te ponovnog izračunavanja međupoloženosti preostalih veza. Ponovno izračunavanje je obavezno jer uklanjanje veze može uzrokovati da ranije slabo prometna veza postane bitno prometnija. Iterativna implementacija ovih koraka predstavlja djelišni algoritam za pronalazak zajednica koji rastavlja početni graf na sve manje komade koji su međusobno povezani sve dok se ne dobije skup izoliranih čvorova.



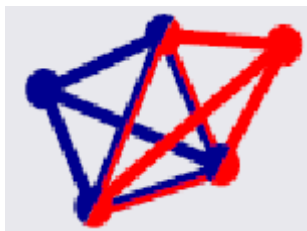
Slika 2.3. Veza između čvorova 4 i 6 ima najveću međupoloženost u ovome grafu; najprometnija je jer njom prolaze svi putevi do čvora 6 [4]

Glavni nedostaci ovakvog pristupa pronalasku zajednica su sporst kod analize velikih mreža (osim ako su izuzetno raštrkane) i poprilično loši rezultati za guste, zbijene mreže.

2.4. k-grupno filtriranje i ostale lokalne metode

Metoda *k-grupnog filtriranja* (engl. *k-clique percolation*) temeljena je na konceptu *k-grupe* (engl. *k-clique*). Riječ je o potpunom podgrafu s k čvorova koji su povezani svim mogućim vezama (ima ih $k(k-1)/2$) [1]. Pretpostavka od koje se polazi je ta da se zajednice sastoje od nekoliko manjih grupa koje dijele mnoge svoje čvorove s drugim grupama u istoj zajednici. *K-grupna zajednica* (engl. *k-clique community*) definirana je kao unija svih „susjednih“ k -grupa, koje, po definiciji, dijele $k-1$ čvor [1].

U takvoj zajednici moguće je izvesti postupak koji se naziva „*kotrljanje*“ *k-grupnog predloška* (engl. *k-clique template rolling*). *K-grupni predložak* (engl. *k-clique template*) je objekt koji je izomorfan potpunom grafu s k čvorova [1]. Takav predložak može se postaviti na bilo koju k -grupu u mreži, a zatim „*kotrljati*“ do susjedne k -grupe tako da mu se promjeni pozicija jednog čvora, a ostalih $k-1$ ostane netaknuto [3]. Unija svih podgrafova koje možemo obići na ovaj način naziva se k -grupna zajednica. Što je k veći, uočavanje zajednice postaje očitije. Najprikladniji iznosi za k su 3, 4, 5 i 6 jer veće vrijednosti dovode do nezgrapnosti u postupku. Kada je $k=2$ govorimo o filtriranju veza, a kada je $k=1$ govorimo o filtriranju čvorova.



Slika 2.4. Susjedne k -grupe [7]

Algoritam k -grupnog filtriranja primjer je lokalne metode za pronalazak zajednica. Uvid u globalnu strukturu zajednica neke mreže dobiva se promatrajući skup zajednica dobivenih postupkom kotrljanja k -grupnog predloška. Neki čvorovi neće pripadati niti jednoj zajednici (jer nisu dio niti jedne grupe), dok će neki pripadati nekolicini zajednica (ako se nalaze na granici dviju ili više zajednica). Ugniježdjena struktura zajednica uočava se ako se u obzir uzmu različite vrijednosti za k .

Važno je napomenuti da postupak k -grupne filtracije često može dati iskrivljen, nedovoljno precizan uvid u strukturu mreže jer dovodi do previda ostalih gusto smještenih struktura koje nisu tako dobro povezane. S druge strane, glavna prednost ove i ostalih lokalnih metoda je to da izuzetno uspješno uzimaju u obzir preklapanje zajednica.

Uočavanje zajednica između kojih se javlja preklapanje posebice je važno u društvenim znanostima, jer ljudi istovremeno pripadaju nekolicini zajednica (poduzeće u kojem rade, obitelj, sportska udruženja itd.). Čisto aglomerativne i djelišne metode ne uzimaju u obzir preklapanje zajednica i upravo je zbog toga potrebno koristiti i lokalne metode.

2.5. Optimizacija modularnosti

Jedna od najčešće korištenih kvalitativnih funkcija je *modularnost* (engl. *modularity*). Ona mjeri koliko dobro neka particija mreže dijeli njezine zajednice [2]. Postoji mnogo algoritama koji pokušavaju optimizirati modularnost i njoj slične kvalitativne funkcije. Modularnost netežinske i neusmjerene mreže koja je podijeljena na zajednice računa se pomoću formule [2]:

$$Q = \sum_i (e_{ii} - b_i^2) \quad (1)$$

Član e_{ij} označava broj veza u mreži koji povezuju čvorove grupe i s čvorovima grupe j , a b_i se računa po formuli $b_i = \sum_j e_{ij}$ i označava broj veza kojima je jedan kraj u grupi i . Iz toga slijedi da je modularnost Q broj veza koje se nalaze unutar zajednica umanjen za očekivanu vrijednost jednakog broja veza koje su odabrane nasumično, bez obzira na pripadnost zajednici [2]. Velik iznos modularnosti Q pokazatelj je podjele mreže u kojoj veze pripadaju skupinama više nego što je očekivano. To je pak dobar pokazatelj funkcionalne podjele mreža.

Kod težinskih mreža u obzir se uzima zbroj težina svake veze, pa tako „teže“ veze doprinose formuli više nego „lakše“. Modularnost se i u tom slučaju računa pomoću gore navedene formule i njezino je značenje jednako kao i kod netežinskih mreža.

Sve kvalitativne funkcije, pa tako i modularnost, nude precizan način računanja jakosti veza unutar zajednica u odnosu na jakost veza među zajednicama.

2.6. Spektralna podjela

Postupak *spektralne podjele* (engl. *spectral partitioning*) intenzivnije se počeo koristiti u razvoju algoritama za paralelno računanje. Kod standardne spektralne podjele svojstva mreže ovise o spektru pripadajuće Laplaceove matrice L_i čiji su elementi $L_{ij} = k_i \delta(i, j) - A_{ij}$. Član k_i je stupanj čvora i (ili, u težinskom grafu, njegova jakost), a $\delta(i, j)$ je Kroneckerova delta funkcija (iznosi 1 ako je $i=j$, inače iznosi 0) [1].

Najjednostavnija metoda započinje s podjelom mreže na dvije komponente. Nakon toga, svaka se komponenta nastavlja razdvajati na dvije grupe sve dok za time ima potrebe (ovisno o namjeri osobe koja postupak provodi). Za svaki korak podjele definira se vektor s čiji elementi poprimaju vrijednost +1 ako pripadaju prvoj grupi ili -1 ako pripadaju drugoj grupi. Matrica veza između svakog para čvorova

može se prikazati kao $R = \frac{1}{4} s^T L s$. Najbolja podjela mreže nastat će kada se

izabere takav s da se R , a samim time i ukupna jakost veza između dvije grupe, umanjuje. To se uglavnom postiže tako da se unaprijed utvrdi veličina grupa, a zatim se taj podatak koristi u daljnjem postupku. No, iako je uspješan prilikom izrade algoritama za paralelno računanje, ovaj je pristup slabo iskoristiv za pronalazak zajednica, prvenstveno zato jer zajednicama ne možemo unaprijed znati veličinu.

Kombinacijom ideje modularnosti i postupka spektralne podjele dobivamo algoritme koji su primjenjivi na veći broj problema. Za početak, definiramo *matricu modularnosti* B s elementima

$$B_{ij} = A_{ij} - P_{ij} \quad (2)$$

Član P_{ij} označava elemente *null-modelne matrice* (engl. *null model matrix*) koja opisuje relativnu vrijednost veza unutar zajednice [1]. Način na koji ćemo specificirati null-modelnu matricu proizvoljan je i ovisi o namjeri, no najčešće korišten izbor predložili su M. E. J. Newman i M. Girvan i on se može opisati formulom [2]:

$$P_{ij} = \frac{k_i k_j}{2W} \quad (3)$$

Pomoću matrice modularnosti možemo optimizirati formulu modularnosti:

$$Q = \frac{1}{2W} \sum_{i,j} B_{ij} \delta(C_i, C_j) \quad (4)$$

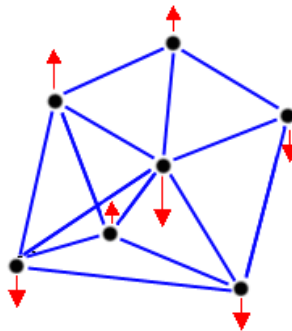
U ovoj formuli član $\delta(C_i, C_j)$ pokazuje da se komponente B_{ij} zbrajaju samo ako se čvorovi i i j nalaze u istoj zajednici. Faktor $W = \frac{1}{2} \sum_{ij} A_{ij}$ je ukupna jakost veza u mreži (kod netežinskih mreža to je ukupan broj veza u mreži), gdje je k_i jakost čvora i .

Kod spektralne podjele može se koristiti onoliko svojstvenih vektora (engl. *eigenvectors*) B koliko je pozitivnih svojstvenih vrijednosti (engl. *eigenvalues*), ali najjednostavnije je rekurzivno razdijeliti mrežu koristeći samo „vodeći“ svojstveni vektor v , koji je sparen s najvećom pozitivnom svojstvenom vrijednošću od B . Nakon toga mreža se može razdvojiti u dvije zajednice prema vrijednostima varijable $s_i = \text{sgn}(v_i)$ [1]. Duljina vektora v_i označava koliko jako i -ti čvor pripada svojoj zajednici. Za $v_i=0$ vrši se izbor zajednice u koju će se uvrstiti čvor i , i to na temelju toga u kojem će slučaju modularnost biti veća, s time da je u tom slučaju dozvoljeno promijeniti i s_i .

Modularnost rezultatne podjele mreže na dvije grupe je $Q = \frac{1}{4W} s^T B s$. Nakon ove podjele, postupak se rekurzivno provodi za svaku novodobivenu komponentu grafa sve dok se modularnost više ne može povećati. Konačna podjela predstavlja strukturu zajednica te mreže.

2.7. Pottsova metoda

Čestice koje imaju magnetski moment često se nazivaju *spinovi*. Međudjelovanje spinova može biti *feromagnetsko* (istog su usmjerenja) ili *antiferomagnetsko* (suprotnog su usmjerenja). *Staklo spinova* (engl. *spin glass*) je sustav koji obuhvaća i nered i uređenost feromagnetskih i antiferomagnetskih međudjelovanja [1]. To dovodi do velikog broja metastabilnih konfiguracija spinova koje su razdvojene energetskim barijerama s dugim vremenom opuštanja (vrijeme iz kojeg sustav prijeđe iz ravnotežnog u neravnotežno stanje).



Slika 2.5. Model stakla spinova [8]

Kod *Pottsovog stakla spinova s q stanja (q-state Potts spin glass)*, svaki se spin može nalaziti u jednom od q stanja [1]. Energija međudjelovanja između spinova i i j je $-J_{ij}$ ako se spinovi nalaze u istom, a nula ako se nalaze u različitom stanju. Hamiltonova funkcija sustava računa se kao zbroj energija svih interakcija:

$$H(\{\sigma\}) = -\sum_{ij} J_{ij} \delta(\sigma_i, \sigma_j) \quad (5)$$

U ovoj formuli član σ_i predstavlja stanje u kojem se nalazi spin i , a $\{\sigma\}$ konfiguraciju spinova (stanje svakog od N spinova). Takvih je konfiguracija ukupno q^N .

Ovaj pristup primijenjujemo na problem pronalaženja zajednica tako da svakom čvoru pridružimo jedan spin te stavimo $q=N$. Energiju međudjelovanja $-J_{ij}$ zbrajamo ako i samo ako se čvorovi i i j nalaze u istoj zajednici. Čvorovi koji su spojeni vezom međudjeluju feromagnetski ($J_{ij} > 0$) ako je težina veze veća od očekivane, a antiferomagnetski ($J_{ij} < 0$) ako je težina veze manja od očekivane. Dakle, dva bi čvora trebala biti u istoj zajednici ako međudjeluju feromagnetski, a u različitim zajednicama ako međudjeluju antiferomagnetski. Nije moguće pronaći takav raspored spinova (tj. podjelu čvorova u zajednice) koji istovremeno minimizira energije međudjelovanja svih parova. Bez obzira na nemogućnost zadovoljenja svih veza istovremeno (taj se fenomen često naziva „*frustracija*“), treba težiti minimiziranju Hamiltonove funkcije H kako bi se pronašlo stacionarno stanje sustava. Matrica međudjelovanja s elementima

$$J_{ij} = \frac{A_{ij} - P_{ij}}{W} \quad (6)$$

implicira da je $H = -Q$ i obnavlja maksimiziranje modularnosti.

2.8. Parametar razlučivosti

Santo Fortunato i Marc Barthélemy su 2007. g. koristeći primjere i stvarnih i računalno generiranih mreža pokazali da modularnost u svome osnovnome obliku rezultira problematičnom granicom razlučivosti [1]. Naime, zajednice koje su manje od donje granične veličine (ona ovisi o veličini mreže i mjeri povezanosti njezinih zajednica) nisu uočljive kao zasebni entiteti nego kao dio većih zajednica, što je značajan propust.

Jedan od načina na koji se ovaj problem može riješiti je uvrštavanje *parametra razlučivosti* λ (engl. *resolution parameter*) u formule kojima smo do sada koristili, npr. u formulu prema kojoj se računaju elementi matrice međudjelovanja:

$$J_{ij} = \frac{A_{ij} - \lambda P_{ij}}{W} \quad (7)$$

Parametar razlučivosti može se ukomponirati i u druge formule koje definiraju kvalitativne funkcije. Njegovo uvrštavanje omogućuje nam bliži ili udaljeniji pogled na mrežu s ciljem pronalaženja zajednica različite veličine. Što su vrijednosti parametra veće, to je uvid u manje zajednice bolji (i obratno).

Uvrštavanje parametra razlučivosti rezultira zanimljivim, novim uvidima. Npr., formula $J_{ij} = \frac{A_{ij} - \lambda}{W}$ daje elemente uniformne null-modelne matrice kod koje je svaki par čvorova povezan vezama sa zadanom srednjom težinom. Takav model može biti koristan kod korelacijskih mreža i mreža sličnosti, npr. onih koje su stvorene iz matrica koje sadrže podatke o jednostavnim da – ne izborima. Čvorovi i i j nalaze se u istoj zajednici ako i samo ako su jednako glasali barem granični broj puta (granica je određena vrijednošću parametra razlučivosti).

3. Primjena

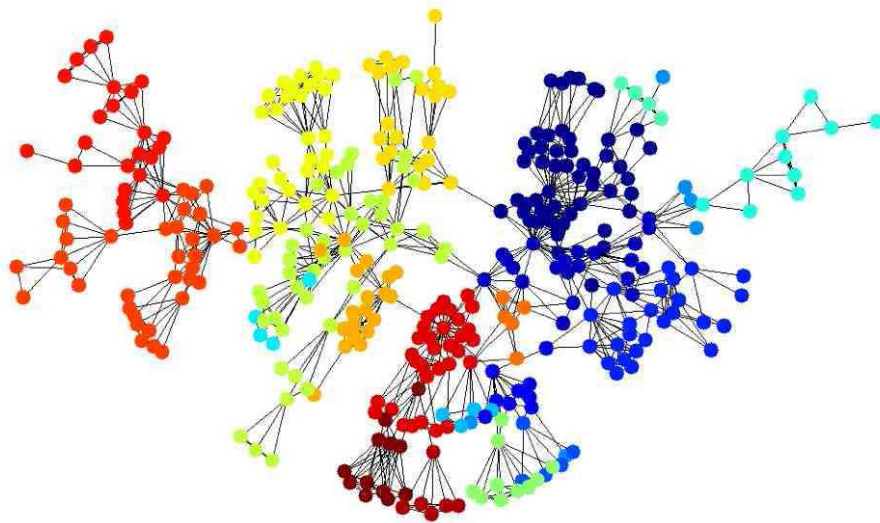
Kroz nekoliko različitih primjera koje ću opisati u ovome poglavlju pružit ću uvid u širok spektar primjene algoritama i ideja opisanih u prethodnome poglavlju.

3.1. Znanstvena suradnja

Bipartitna mreža znanstvenog koautorstva, u kojoj su znanstvenici povezani s radovima kojima su autori ili koautori, može se definirati tako da vrijednost izraza δ_i^p poprima vrijednost 1 ako je znanstvenik i koautor rada p , a u suprotnome poprima vrijednost 0 [1]. Da bismo prikazali jakost suradnje između znanstvenika i i znanstvenika j definiramo:

$$A_{ij} = \sum_p \frac{\delta_i^p \delta_j^p}{n_p - 1}$$

kao komponente težinske unipartitne mreže u kojima n_p predstavlja broj autora rada p , a suma se računa samo za radove koji imaju više od jednog autora.



Slika 3.1. Jedna od komponenata mreže znanstvenika nastale na temelju (ko)autorstva radova; grupe znanstvenika označene su različitim bojama [1]

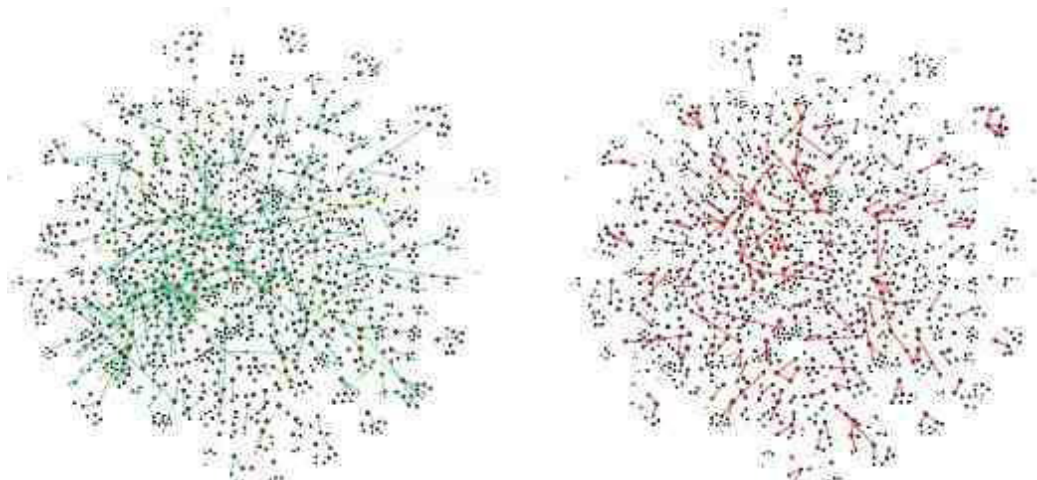
Grupiranje znanstvenika na ovakav način korisno je jer osim što omogućava lakše povezivanje znanstvenika koji su u prošlosti već surađivali, otvara mogućnost suradnje znanstvenika koji nisu u prošlosti surađivali, ali proučavaju isto područje pa su samim time moguće njihove interakcije u budućnosti.

3.2. Zajednice korisnika mobitela

Zajednice korisnika mobitela su longitudinalne – vremenski ovisne. Veze u telefonskim mrežama nastaju na temelju trenutnih uspostava komunikacije. To znači

da se u svakom trenutku mreža sastoji od velikog broja veza koje povezuju ljude koji u tom trenutku razgovaraju telefonom.

2007. g. provedeno je istraživanje [1] na velikoj komunikacijskoj mreži koja se sastojala od nekoliko milijuna povezanih korisnika mobitela u neimenovanoj euopskoj zemlji. Cilj istraživanja bio je uočavanje odnosa između mikroskopske, mezoskopske i makroskopske strukture mreže i jakosti veza između pojedinaca na društvenoj razini. Dobiven je zanimljiv uvid u tzv. *hipotezu o slaboj vezi* (engl. *weak tie hypothesis*, prvi ju je postavio američki sociolog Mark Granovetter) koja tvrdi da preklapanje krugova prijatelja dvaju pojedinaca raste s jakošću veze koja ih povezuje. Na mezoskopskoj razini to rezultira strukturom u kojoj su pojedinci unutar zajednica povezani jakim vezama, a same zajednice su međusobno povezane slabim vezama. Slabe veze su najzaslužnije za strukturalni integritet komunikacijskih mreža: otporne su na uklanjanje jakih veza, ali se raspadaju ako se uklone slabe veze. Može se pokazati da uklanjanje slabih veza dovodi do prijelaza iz mreže koja je globalno dobro povezana u mrežu koja se sastoji od izoliranih zajednica. Ako se pak uklone jake veze, mreža ostaje globalno dobro povezana. Ovakva mezoskopska organizacija mreža ima značajne posljedice na tok informacija. Pretpostavimo li da je svaka veza (bez obzira na njezinu jakost) jednako učinkovita u prijenosu informacija, zaključujemo (na temelju Granovetterove hipoteze) da su slabe veze najzaslužnije za širenje informacija. No, ako se pretpostavi da je intenzitet prijenosa informacija proporcionalan s jakošću veze, tada su jake i slabe veze podjednako zaslužne za širenje informacija.



Slika 3.2. Mreža korisnika mobitela korištena u gore navedenom istraživanju; uklanjanjem jakih veza zadržana je globalna povezanost mreže (lijeva slika), a uklanjanje slabih veza dovodi do raspada mreže (desna slika) [1]

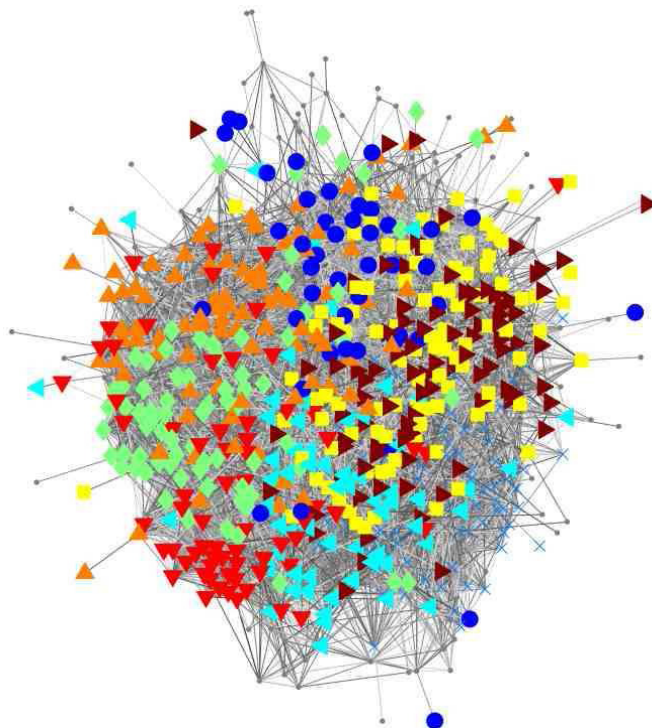
3.3. Online društvene mreže

Web-stranice koje omogućuju povezivanje i interakciju među pojedincima (engl. *social networking sites*) postale su nezaobilazan dio svakodnevice. One korisniku nude otvaranje vlastitog *online* profila, kreiranje liste ostalih korisnika (engl. *friends list*) s kojima je povezan te uvid u njihove profile i *friends* liste. Otkako su se pojavile,

stranice kao što su Facebook, MySpace, LinkedIn i stotine ostalih zajedno su privukle više od milijardu korisnika.

Pojava i nagli rast popularnosti tih stranica pobudili su interes među sociolozima, ali i matematičarima, fizičarima, inženjerima računarstva i ostalim znanstvenicima, prvenstveno zato jer je u kratkome roku javnosti postala dostupna ogromna količina društvenih i demografskih podataka. Ti podaci pružaju uvid u nastanak *online* društvenih zajednica te uvid u (ne)prijateljstva i interakcije njihovih članova.

Jedno takvo istraživanje [1] koristilo je anonimne podatke s Facebooka kako bi se usporedile zajednice prijatelja nekolicine američkih sveučilišta. Čvorovi mreža predstavljali su osobe, a veze između čvorova predstavljale su *online* prijateljstva (potvrđena od obje strane). Korištenjem različitih metoda, posebice algoritama za pronalazak zajednica, omogućena je usporedba različitih skupina unutar mreže. Navest ću samo jedan od mnogobrojnih rezultata i zaključaka ovog istraživanja: pokazalo se da se zajednice na Sveučilištu Princeton i na Sveučilištu u Sjevernoj Karolini formiraju uglavnom na temelju godine studija kojoj studenti pripadaju, dok se na Sveučilištu Caltech formiraju uglavnom na temelju zgrade kampusa u kojoj studenti obitavaju. Iako naizgled beskorisni, ovi jednostavni zaključci mogu se međusobno kombinirati i koristiti u daljnjim istraživanjima ili pak odmah naći svoju upotrebu, npr. u marketingu (želimo li proširiti vijest o nekom novom proizvodu, na Sveučilištu Princeton potrudit ćemo se da ju prvo sazna nekoliko studenata s različitih godina – koji će ju zatim prenijeti ostalima u svojoj zajednici, dok ćemo na Sveučilištu Caltech prvo iznijeti vijest studentima iz različitih zgrada kampusa – koji će ju isto tako prenijeti ostalima u svojoj zajednici).



Slika 3.3. Mreža prijateljstava Sveučilišta Caltech (pripadnosti pojedinoj zgradi kampusa označene su različitim bojama i oblicima) [1]

4. Zaključak

Proučavanje mreža u početku se provodilo samo na području teorije grafova te je davalo korektne, ali preopćenite rezultate. Tek kad je uočena velika važnost pojma mreže i njemu srodnih pojmova u specifičnim granama znanosti započela su efikasnija istraživanja. U koštac s ovim problemom posebice su se uhvatili sociolozi, koji su proučavali zajednice ljudi, način na koji nastaju i opstaju.

Kako su zajednice postajale sve veće i složenije problem njihovog otkrivanja, a zatim i proučavanja postao je teško rješiv u okviru samo jedne znanosti, javila se potreba za znanstvenom suradnjom. Tada u pomoć sociolozima pristižu matematičari, fizičari i računalni inženjeri koji razvijaju matematičke i računalne metode i algoritme koji olakšavaju proces uočavanja zajednica. Rezultati koje ostvaruju zajedno znatno su bolji od onih koje su ostvarivali pojedinačno.

Bez obzira na zadovoljavajuće rezultate, područje proučavanja mreža i dalje se intenzivno razvija, ideje se marljivo skupljaju, svakim danom nastaju sve bolje metode i algoritmi.

5. Literatura

- [1] M. A. Porter, J. Onnela, P. J. Mucha, 22.02.2009., *Communities in networks*, http://arxiv.org/PS_cache/arxiv/pdf/0902/0902.3788v1.pdf, 02.03.2009.
- [2] M. E. J. Newman, M. Girvan, 11.08.2003., *Finding and evaluating community structure in networks*, http://arxiv.org/PS_cache/cond-mat/pdf/0308/0308217v1.pdf, 09.03.2009.
- [3] M. E. J. Newman, M. Girvan, 11.06.2002., *Community structure in social and biological networks*, <http://www.pnas.org/content/99/12/7821.full.pdf>, 09.03.2009.
- [4] *Graph theory*, http://en.wikipedia.org/wiki/Graph_theory, 04.05.2009
- [5] *k-means clustering*, http://en.wikipedia.org/wiki/K-means_clustering, 04.05.2009.
- [6] *Cluster analysis*, http://en.wikipedia.org/wiki/Cluster_analysis, 04.05.2009.
- [7] G. Palla, D. Ábel, I. Derény, I. Farkas, P. Pollner, T. Vicsek, ožujak 2008., *k-clique percolation and clustering in directed and weighted networks*, <http://www.inma.ucl.ac.be/~blondel/workshops/2008/files/palla.pdf>, 04.05.2009.
- [8] I. Yevin, *Attractor network model of pattern recognition and structure of artworks*, <http://yevin.fupm.mipt.ru/vismath/>, 04.05.2009.

6. Sažetak

U seminaru se objašnjava pojam mreže i zajednice, pruža se kratak uvod u teoriju grafova. Iznose se najpoznatiji i najčešće korišteni algoritmi i ideje za pronalazak zajednica u mrežama. Algoritmi su popraćeni matematičkim objašnjenjem i formulama te slikovnim prikazima. Naposljetku se pokazuju primjene tih algoritama i ideja kroz nekoliko jednostavnih primjera iz različitih područja ljudskog djelovanja.