Predavanja

Friday, November 10, 2023 6:31 PM

Socijalne mreze mogu se posmatrati na vise nivoa:

- Mikro - ako se gleda na nivou pojedinacnih ljudi
- Mezo - na nivou institucije
- Makro na nivou drzave

Pojam i predstavljanje socijalne mreže

Teorija mreža - proučavanje usmerenih i neusmerenih grafova kojima se modeluju odnosi između entiteta.
Mreža se prestavlja pomoću skupova čvorova i grana, diskretni objekti (entiteti) predstavljaju čvorove. Odnosi se modeluju granama grafa. U graf se unose i dodatne osobine (težine grana, atribute grana, atribute čvorova i sl.)

Socijalna mreža je širi pojam od pojedinačnog servisa na internetu, tj. od ovih društvenih mreža. U najširem smislu, socijalna mreža modeluje odnose i interakcije između različitih entiteta.

Imamo usmerene, neusmerene i mešovite grafove, ali su mešoviti ređi i najčešće se konvertuju u usmereni graf.
Relacija koju mreže modeluju može biti eksplicitno ili implicitno.

- Direktna mreža - relacije su eksplicitno prisutne -(prijateljstvo na Fbu)
- Indirektna mreža - relacije su implicitno prisutne - (prisustvo aktera u istoj sceni)

Relacije u mreži mogu biti jednostruke odnosno višestruke, tj da dođu po jednoj osnovi ili po više

- Simplex - mreža predstavlja samo jedan tip relacija izmežu aktera u mreži
- Multiplex - višestruka relacija

Kod višestrukih može doći do problema vizuelizacije i gubitka informacija usled agregacije.
Relacija može nositi informaciju o njenom intezitetu.

- Binarni graf - netežinski
- Signed – mešoviti, postoje i usmerene i neusmerene grane, najčešće se konvertuju u težinske, pre su se koristili
- Valued - težinski

Bipartitne mreže - *two-mode networks*
Imaju dve vrste čvorova koji predstavljaju različite entitete, relacije moguće samo između entiteta različitog tipa.
Ego mreže - fokusiraju se na izabranog aktera (ego) i njegove neposredne veze (alters). Svi imaju veze sa ego čvorom, a dozvoljene su i međusobne veze alter-a. Obično se dobijaju iz drugih mreža.
MORA GLAVNI JUNAK DA IMA VEZU SA SVIM ČVOROVIMA

Predstavljanje mreže:

- Matrična reprezentacija
  - Matrica susednosti
  - Matrica incidentnosti -> sta je ono bese ovo?
- Reprezentacija listom
  - Lista susednosti -> ona klasika što znam sa ASP-a
  - Linearizovana lista susednosti ---------->
  -

Standardizovani formati: ***GML, GEXF -> dodaju se labele i slično za dodatne informacije nego samo ovo kako se predstavljaju grafovi***
Načini prikupljanja podataka o relacijama:

- ***Full Network***
  - Sakupljaju se potpune informacije o vezama svih aktera sa drugim akterima
- ***Snowball metod***
  - Polazi se od određenog skupa aktera u fokusu, a zatim se podaci o vezama prikupljaju na osnovu njihovih veza i novoidentifikovanih aktera
- ***Ego-centrični metod***
  - Prikupljaju se podaci o pojedinačnom akteru (egu)
  - Sa ili bez veza između njegovih aktera

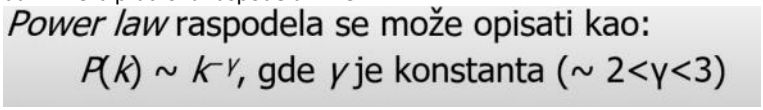
Web crawler - tragači za prikupljanje informacija - Scrapy u Pythonu

Osnovna svojstva socijalnih mreža

Često korišćene metrike: veličina čvora, gustina, fregmentavija, minimalni, maksimalni, stepen čvora, distribucija čvora. Metrike u vezi distanci: povezanost, geodezijska distanca, eksecnetričnost čvora, dijametar, protoci.
Nekad se težinski graf transformiše u binarni, zato što mogu da se vide bolji, jasniji rezultati primenom metrike na takav graf, ali jako često metrike mogu da se primene direktno na težinski graf
Dihotomizacija tj binarizacija mreže, podrazumeva na neki način pročišćivanje same mreže. U kontekstu da ostanu samo binarne relalcije. Zasnovana na srednjoj, medijalnoj ili nekoj drugoj vrednosti grane u mreži za odsecanje grane u mreži. Ovaj postupak nije uvek neophodan ali nekad može poboljšati same rezultate.
Stepen čvora - broj grana koje su incidentne (susedne) posmatranom čvoru. To je u direktnoj vezi sa centralnošću po stepenu. Kod neusmerenoj mreži govorimo samo o stepenom čvoru, dok kod usmerenih imamo ulazni i izlazni stepen čvora. Tražimo minimalan, maksimalan i prosečan stepen čvora i možemo da računamo i standardnu devijaciju. Čvorovi koji imaju veći stepen su bolji za saradnju.
Dostižnost čvora - da li se iz nekog čvora mreže mogu dosegnuti preostali čvorovi
Povezanost čvora - definiše broj čvorova koji treba ukloniti na putanji između dva posmatrana čvora da bi oni postali nepovezani. Ukoliko postoji veći broj puteva između dva aktera u mreži tada su oni povezaniji.
Veličina mreže - se odnosi na ukupan broj čvorova u mreži
Maksimalan mogući stepen svakog čvora - (n-1), gde je n broj čvorova u mreži
Broj mogućih konekcija u mreži:

- Usmerenih mreža n*(n-1)
- Neusmerenih mreža n*(n-1)/2

Kompleksnost mreže raste eksponencijalno sa njenom veličinom
Gustina mreže - računa se kao količnik ukupnog broja grana i maksimalnog mogućeg broja grana u mreži -> gruba jako procena, i retko se koristi pošto u realnom smislu neka mnogo uticaja
Fragmentacija mreže - se odnosi na proporciju parova čvorova u mreži koji nisu dostižni u odnosu na ukupan broj parova, koliko ima parova koji su međusobno dostižni / ukupan broj parova -> bolja
Alternativna metrika gustini za poređenje je *links per node,* predstavlja odnos broja konekcija i broja čvorova u mreži. **Gustinu treba koristiti pažljivo prilikom poređenja različitih mreža, različite veličine. Sa njom se stiče osećaj da li je mreža realna, ili ne, pošto realne mreže (socijalne, koje predstavljaju neku interakciju između ljudi) nisu guste**
Distribucija čvorova po stepenu - Random mreža liči na gausovu raaspodelu stepeni čvorova, tj pausonova je. U realnim mrežama, odnosno društvenim mrežama - većina čvorova ima nekoliko veza, ali postoje čvorovi sa veoma velikim brojem veza. Takve mreže prate power-law = scale-free raspodelu. Znači niajviše ih ima malo, blizu nula (skoro da eksponencijalno opada). Proučavamo da li mreža prati ovu raspodelu ili ne



Koncep distanci, govore o načinu širenja i dosega informacija.
Prost put (*walk)* u mreži se definiše kao niz aktera i relacija koji počinju i završavaju se aktterom. Na složenom putu su dozvoljena ponavljanja aktera.
Ciklus je prost put koji počinje i završava se istim čvorom.
Geodezijska distanca predstavlja broj grana na najkraćoj putanji između dva aktera u mreži. Najkraća putanja u smislu broja grana, kada pustimo BFS. Ponekad se ona transformiše u meru bliskosti.
Kod težinskih grafova, se ona može računati i na drugi način - Floyd-Warshall, Dijkstra itd. Moze i da se racuna cost na putu, ili minimum težina svih grana na putu između A i B (strength, maximum flow), ili proizvod težina na putu ako ona predstavlja na primer verovatnoću.
Ekcentričnog čvora - max geodezijska distanca čvora. Središte grafa je čvor sa min ekcentričnosti.
Dijametar mreže jednak je najdužoj od svih najkraćih putanja u mreži između parova čvorova. -> najčešće se koristi
Ili prosečna dužina putanje u mreži

Mere centralnosti

Topologija mreže: toplogija zvetde, linije i kruga
Otkrivaju koji čvorovi su bitni za održavanje povezanosti mreže. Daju uvid u položaj pojedinačnih aktera u mreži. Računaju se za pojedinačne čvorove ali i za celu mrežu.
Degree centraliry - broj direktnih suseda svakog čvora u mreži - **Centralnost po stepenu - komunikativan**

Za poredjenje sa drugim stepenom koristi se normalizovana vrednost -> podelim iznos metrike sa velicinom mreze - 1 (taj 1 je trenutno posmatrani cvor)

- Definiše kao recipročna vrednost sume najkraćih rastojanja od posmatranog čvora do ostalih sa kojima je povezan

$$C_C(p_i) = \frac{N-1}{\sum_{k=1}^{N} d(p_i, p_k)}$$

- $d(p_i, p_k)$ predstavlja najkraće rastojanje između posmatranog čvora i nekog od preostalih u mreži sa kojima je povezan
- N je ukupan broj čvorova u grafu

Clos **i, mostovi I slici**

Suma je normalizovana veličinom mreže umanjenom za jedan, zbog poređenja mreža različitih dimenzija

Interpretacija:

Čvorovi sa većom vrednošću su na centralnijim pozicijama u mreži

U proseku je potrebno manje koraka da dođu do preostalih čvorova u mreži

Odraža brzinu kojom informacije mogu da se šire

U slabo povezanim grafova koristi se varijanta **harmonijske centralnosti po bliskosti**

Betweenness centrality - relaciona centralnost - bitan u mrezi za povezivanje

- Proporcionalna je broju najkraćih puteva između svih ostalih parova čvorova na kojima se posmatrani čvor nalazi u odnosu na ukupan broj takvih najkraćih puteva

$$C_B(p_i) = \sum_{j=1}^{N} \sum_{k=1}^{j-1} \frac{g_{jk}(p_i)}{g_{jk}}$$

- g_{jk} predstavlja ukupan broj najkraćih puteva koji povezuju čvorove p_j i p_k
- $g_{jk}(p_i)$ predstavlja broj takvih puteva koji uključuju čvor p_i
- N je ukupan broj čvorova u grafu

- Algoritam za izračunavnanje za neki čvor v u grafu $G(V,E)$ sa N čvorova
  - Za svaki par čvorova (s,t), odrediti skup najkraćih puteva između njih
  - Za svaki par čvorova (s,t), odrediti frakciju (broj) puteva koji prolaze kroz čvor v
  - Odrediti ukupan zbir ovih frakcija za sve parove čvorova u mreži
- Normalizovana vrednost se dobija podelom dobijene sume sa ukupnim brojem parova čvorova (s,t) koji ne uključuje v
  - Za usmerene grafove: $(n-1)*(n-2)$
  - Za neusmerene grafove: $(n-1)*(n-2)/2$

Kvantifikuje kontrolu komunikacije unutar mreže od strane pojedinih aktera. Onaj sa većom se nalazi na većem delu najkraćih puteva i

Onda to dođe neki most/posrednik/boker

Interpretacija:

Čvorovi sa višim vrednostima relacione centralnosti se nalaze na većem broju najkraćih putanja

Ovakvi čvorovi predstavljaju mostove između ostalih čvorova u mreži

Označava broj čvorova koji su indirektno povezani u mreži preko direktnih grana

Mogu biti i tačka prekida u okviru mreže
Čvorovi sa visokom relacionom centralnošću su često na periferiji grupa unutar mreže

	Niska DC	Niska CC	Niska BC
Visoka DC	/	Čvor se nalazi u klasteru koji je daleko u odnosu na ostatak mreže	Konekcije čvora su redundantne, komunikacija ga zaobilazi
Visoka CC	Bitan čvor, povezan sa važnim drugim akterima	/	Veliki broj putanja u mreži, čvor je blizak drugima, ali su i drugi međusobno bliski
Visoka BC	Veze čvora izuzetno bitne za tok informacija u mreži	Čvor monopolizuje veze manjeg broja ljudi ka ostalim čvorovima mreže	/

Eigenvector centrality - centralnost po sopstvenom vektoru, modifikacija centralnosti po stepenu
Koja uzima u obir i sredstvo posmatranog čvora
Koristi koncepte uticaja i moći
Čvor je uticajniji ukoliko i njegovi susedi imaju veliki broj suseda
Čvor je moćniji ukoliko njegovi susedi nemaju veliki broj svojih suseda

- Računa se određivanjem svojstvenog vektora koji sadrži relativne centralnosti x_i za svaki čvor

$$x_i = \frac{1}{\lambda}\sum_{k\epsilon M(i)} x_k = \frac{1}{\lambda}\sum_{k\epsilon G} A(p_i, p_k)x_k$$

  - λ je neka konstanta
  - A matrica susednosti za posmatranu mrežu
  - $M(i)$ skup suseda čvora p_i
- Mera centralnosti se dobija pronalaženjem najveće odgovarajuće sopstvene vrednosti λ, što se vrši dalje iterativnim postupkom

Beta (Bonacich) centrality
Je varijanta centralnosti po sopstvenom vektoru, kojoj se dodaje beta parametar sa vrednosti [-1,1], pozitivne vrednosti parametra utiču na uticajne čvorove, dok negativni ističu moćne čvorove

Generalno, centralnost po stepenu, bliskosti i relaciona centralnost su pozitivno korelisane. Prve dve bas korelisu.
U slučaju kada je korelacija niska, to verovatno govori o nekom interesantnom svojstvu posmatrane mreže

Centralizacija na nivou mreže
Pokazuje varijansu izračunate mere centralnosti posmatrane mreže procentualno u odnosu na mrežu iste veličine koja ima topologiju zvezde (u takvoj mreži raspodela moći je najviše nejednaka)
Razlika svih sračunatih metrika, tražim maks u obe mreže:

- Centralizacije mreže po stepenu se računa na sledeći način

$$C_D(\mathrm{G}) = \frac{\sum_{i=1}^{|V|}[C_D(v_*) - C_D(v_i)]}{\sum_{j=1}^{|Y|}[C_D(y_*) - C_D(y_j)]}$$

  - *G(V,E)* predstavlja posmatranu mrežu
  - *X(Y,Z)* predstavlja mrežu sa istim brojem čvorova koja ima topologiju zvezde (*star network*)
  - Čvor označen sa * je onaj sa najvećim stepenom u mreži
- Veći procenat centralizacije mreže ukazuje na postojanje čvorova koji su boljoj poziciji u odnosu na ostale

Detekcije komuna

Stvaranje komuna u mrežama je posledica homofilije i asortivnosti.
Homofilija predstavlja tendeciju ljudi da stvaraju veze sa sličnim ljudima
U analizi socijalnih mreža homofilija se javlja u formi asortivnosti

Izračunava se u smislu stepena čvora, čvorovi imaju tendeciju da se poveziju sa čvorovima sličnog stepena, izračunava se kroz koeficijent asortivnosti
Karakteristično samo za socijalne mreže

Heterofilija predstavlja tendeciju ljudi da usvajaju nova iskustva u kontaktu sa drugačijim ljudima

Suprotnost homofiliji, postiče inovativnost i razmenu ideja

Osnovne jedinice u mreži su dijade (dyad) i trijade (triad)
Dijadu čini par aktera koji mogu biti povezani granom, trijadu čine tri aktera i njihove moguće veze
Od interesa su nam oni koji su povezani sa vezama
Triade se formiraju na osnovu dve dyade, preko poznavanje zajednicke osobe, i kazu da ako postoji jaka veza u te dve diade, nema sanse da ne dodje do spajanjem u triade

Komuna je jedna grupa čvorova koja je gušće povezana unutar sebe nego sa ostatkom mreže.
Komuna == klaster
Podela na komune se može raditi na osnovu:

Prisutnosti direktnih veza među članovima

Sa svim ostalim članovima
Sa određenim minimalnim broje članova

Povezanost članova u maksimalnom broju koraka

Polazeći od jednog člana do ostalih se može uraditi u k-koraka

Odnos gustina konekcije unutar grupe i prosečne gustine konekcija u mreži -> definicija na osnovu intra-klaster i inter-klaster gustin

C podgraf grafa G(V,E)
Intra-klaster gustina podgrafa C je odnos broja internih veza među članovima podgrafa, i svih mogućih internih veza
Inter-klaste gustina podgrafa C je odnos broja eksternih veza među članovima podgrafa, i svih mogućih eksternih veza
Da bi C podgraf bio komuna intra-klaster gustina treba da bude značajno veća od inter klase gustine

Klika (clique) je podgraf, koji se sastoji od najvećeg mogućeg skupa čvorova koji su međusobno povezani svaki sa svakim (mora postojati direkt grana između svakog čvora, predstavlja max kompletan graf, u okviru jednog grafa može postojati više klika, mogu se međusobno preklapai i jedan čvor može pripadati većem broju klika. Nije interesantna sa stanovništa analize socijalnih mreža, zato što imamo potpuno povezan graf (kompletan) i onda su svi čvorovi na istom i ni jedan se ne ističe u odnosu na ostale.

Onda pravimo neke relaksiranije definicije klike:
N-klika svi čvorovi su povezani direktno ili putevima dužine N(obično 2).

N-klan svi putevi <= N

K-pleks svi čvorovi imaju veze sa svima, osim nekih k (ako nam strči par koji nisu potpuno

povezani)
K-jezgra podrezumevaju postojanje najmanje K veza ka ostalim članovima grupe

Povezana komponenta – podgraf u kome postoji put između bilo koja dva čvora:
*ako postoje izolovani čvorovi, oni predstavljaju zasebne komponente
Kod usmerenih imamo jako i slabo povezane komponente. Kao svojstvo mreže često se koristi broj povezanih komponenti

Postupak dihotomizacije grafa se koristi da se od težinskog grafa napravi bitenarni, na osnovu težina i nekog zadatog praga. Polazimo od težine najveće grane i onda iteriramo tako, dobijamo stablo podele koje se naziva dedongram
Možemo da vršimo podelu na osnovu teorije jakih i slabih veza

Tačke prekida - čvorovi grafa čije bi uklanjanje dovelo do podele mreže
-> **Blokovi** – 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 betwenness skorom

Koeficient klasterizacije, clustering coefficient
Opisuje tendeciju čvora ka tome da stvori kliku
Lokalni koeficijent klasterizacije se računa kao gustina mreže koju čine posmatrani čvor, njegovi susedi i njihove međusobne mreže, tj. Ego mreža posmatranog čvora
Kompletne mreže se rečuna na osnovu proseka pojedinih čvorova, što je veći koeficijent klasterizacije to je veća šansa da se formiraju klike

Globalni koeficijent klasterizacije, posmatra zatvorene trijade unutar mreže u odnosu na sve ostale trijade u okviru mreže

Moludarnost – odnos broja grana u okviru nekog klastera i ukupnog broja grana umanjenog za isti takav odnos kada bi se grane rasporedile na slučajan način

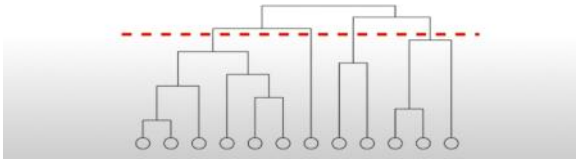
Algoritmi za detekciju komuna u mrežama

Za detekciju komuna imamo
Top-dow koji polazi od kompletne mreže i gleda povezanost komponenti – Girvan-Newman metod
Rezultat je dendogram koji prikazuje grupisanje manjih jedinica u veće
Bottom-up koji polazi od manjih gradivnih celina i ide ka sagledavanju cele mreže - Propagacija labela i Louvain metod, više se sada ovaj koristi

Particijsko grupisanje -> najstarija tehnika, oslanja se na podelu čvorova izabranog grafa u k grupa, predefinisanih veličina, tako da moramo da znamo ta dva podatka unapred, tj da imamo pretpostavku o topologiji mreže.
Ako ne zadamo broj grupa, pošto on traži minimalan broj grana između grupa, pa bi dobili samo jednu grupu jer je tad broj grana između grupa null, a ako ne zadamo željenu veličinu grupe, dobijemo podelu na dve grupe, gde minimalan broj grana dobijamo odvajanjem čvora sa najmanjim stepenom u posebnu grupu i od ostatka komune.
Najpre se odredi broj k, tacke prikazemo u prostoru, razdaljine predstavljaju sličnost, pa odradimo cenu neku toga.
Imamo metode koje ne koriste centroide, oni se ne koriste obično
A metode koji koriste centroide: k-center, k-median (koriste daljenost od centroida)
K-means clustering (minimizacija sume kvadrata udaljenosti), ali ove stvari se retko koriste u analizi socijalnih mreža zato što je teško definisati, broj grupa i veličinu grupe.

Hijerarhijske metode -> dele mrežu na komune bez ikakvih pretpostavki o strukturi mreže
Oslanjaju se na premisu da je moguće ikrementalnim grupisanjem jako povezanih čvorova podeliti mrežu na komune
Potrebno je odrediti meru za jačinu povezanosti ili sličnost čvorova na osnovu koje se vrši grupisanje
Formira se matrica sličnosti za sve parove čvorova u mreži
- Rezultati hijerarhijskog grupisanja mogu se

- Rezultati hijerarhijskog grupisanja mogu se prikazati pomoću dendrograma
  - Hijerarhijsko stablo
  - Idealno za ilustraciju rasporeda klastera
  - Analizom horizontalnih preseka stabla može se uočiti napredak podele grafa na komune u različitim fazama obrade



Moramo da uvedemo neku metriku da bi znala koja podela je najbolja
Algomerativni algoritmi pretpostavljaju da je svaki čvor komuna za sebe, pa ih grupišu na osnovu sličnosti, staje se dok nije jedan graf cela komuna
Razorni algoritmi posmatraju ceo gaf kao jednu komunu, pa razaraju dok ne dođu do pojedinačnog čvora
Naravno mi izaberemo neki drugi prag za zaustavljanje za oba tipa algoritma
Girwan-Newman metod – razorni algoritam, uklanja grane iz grafa i tako formira komune

1. Računanje centralnosti sveke grane u mreži, gleda najkraće rastojanje između svih parova čvorova, pa rastojanja koja prolaze preko te neke grane, odnos najkraćih puteva koji prolaze kroz tu granu i ukupnog broja puteva (opis relacione tj geodezijske centralnosti, ima ih još ali se ova najčešće koristi)
2. Uklanjanje grane koja ima najveću centralnost
3. Ponovo računanje u modifikovanom grafu
4. I sve u krug dokle god postoje grane u tekućem grafu

Problem je velika složenost ovog algoritma ~ O(mn^2) m broj grana, n čvorova
Modularnost -> mera kvaliteta podele, ista za sve logično
Girvan-Newman modularnost -> računa se u odnosu na prazan (null) model grafa
Graf sa slučajnim rasporedom grana koji ne sadrži komune
Poređenjem gustine grana u komuni sa gustinom grana u podgrafu istih čvorova u praznom modelu, dobiha se odstupanje komune od slučajnog rasporeda grana

Mera kvaliteta podele (2)

- Računa se po sledećoj formuli:

$$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(C_i, C_j)$$

  - Suma se računa za svaki par čvorova i i j
  - A_{ij} predstavlja broj grana između čvorova i i j
  - k_i i k_j predstavljaju stepene čvorova i i j u praznom modelu, a čitav proizvod verovatnoću postojanja grane
  - Ukoliko i i j pripadaju istoj komuni $\delta(C_i, C_j)$ je jednaka jedan, a u suprotnom nula
- U sumi učestvuju samo oni parovi čvorova koji pripadaju istim komunama:

$$Q = \sum_{c=1}^{n_c}\left[\frac{l_c}{m} - \left(\frac{d_c}{2m}\right)^2\right]$$

  - gde je n_c broj komuna, l_c je broj grana unutar komune
  - d_c zbir stepeni čvorova unutar komune

Uzima vrednosti od –0.5 do 1, kod realnih mreža od 0.3 do 0.7, 0.7 se smatra kao dobra podela.
Alternativna metrika – **konduktansa (conductance)**
Kvalitet procenjuje na osnovu spoljašnje povezanosti komune, predstavlja odnos broja grana koje izlaze iz komune u ostatak grafa i broja grana koje imaju bar jedan čvor unutar komune

Propagacija labele

1. Svakom čvoru se dodeli jedinstvena labela
2. Nasumičnim redosledom obilaze se svi čvorovi grafa i nad njima se vrši preračunavanje vrednosti labela
3. Svakom čvoru se dodeljuje vrednost labele koju dele najveći broj njegovih suseda
4. Ponavlja se do konvergencije
5. Konvergenicja je postignuta kada prestanu da se menjaju labele novim obilaskom

Čvorovi unutar komune su dosta bolje povezani međusobno nego ostatak grafa, većina čvorova iz iste komune će deliti labelu već nakon nekoliko iteracija
Kada bude stabilno stanje, ista labela znači jedna komuna
Sihrona propagacija-> koristi informacije iz prethodne iteracije
Moramo da imamo dodatne kriterijume, ako ne konvengira algoritam

Asihrona propagacija-> i prethodno stanje i do tada obradjeni u tekućoj iteraciji
Obrađuju se čvorovi svaki put po nasumičnom redosledu
Semisihrona propagacija -> zasniva se na bojenju grafa tako da susedi imaju različite boje, u svakoj fazi čvorovi iste boje ažuriraju svoje labele (faza ima koliko i boja) , stabilan i brz

Imamo nekad problem ako dve ili više labela imaju podjednaku zastupljenost kod suseda:

1. LPA-Random -> slučajan odabir
2. LPA-Prec -> daje prednost prethodnoj labeli, brže dovodi do konvergencije, ali podela u particije može da bude slabijeg kvaliteta
3. LPA-Max -> labela sa najvecim prioritetom, labele su obično int, pa uzimamo najmanju ili najveću vrednost
4. LPA-Prec-Max -> gleda se prioritet ali se prednost daje tekućoj labeli

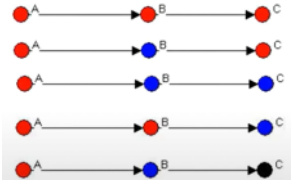
Kriterijum zaustavljanja, možemo da fiksiramo broj iteracija u koliko se labele ne menjaju, ne valja da kažemo samo idemo n iteracije, ako npr koristim LPA-Max ako ima istu labelu kao u poslednje dva stanja onda kažemo da je stigao do kraja
Louvain metod – zasnovan na maksimizaciji modularnosti particija u okviru mreže, pohlepan algoritam ~O(nlogn), daje rešenja bliska optimalnom, u dva koraka prvo sa LPA pronalaze manje komune i optimizuje se modularnost lokalno, ponavlja se sve dok postoji uvećanje modularnosti, zatim se agregiraju čvorovi koji pripadaju istim komunama i pravi mreža komuna, pa se ponavlja LPA na ovakvoj mreži komuna

Klasteri u realnim mrežama

Imamo tri vrste veze, jake, slabe i one između. E sad te između brišemo, prag slabe veze definišemo sami, dok se prag jake veze posle algoritamski računa
Fora trijada x,z,y gde ima jaka veza x,z i x,y to znaci da postoji slaba veza između z i y (osobina slabe tranzitivnosti)
I onda se jačina jake veze s (tj prag da veza bude jaka) je najmanja težina, takva da postoji slabo tranziitivna trijada takva da slaba veza ima težinu veću od w

Brokeri (mostovi) između klastera, omogućavaju protok informacija između različitih grupa od poverenja. Može se iskazati i kroz odgovarajuću metriku (brokerage)
Broj parova suseda u ego mreži posmatranog čvora koji nisu međusobno povezani
Svaki čvor može imati neku od sledećih uloga:

1. Koordinator – izrazit brokerage ima
2. Konsultant – A i C su u istom klasetreu ali B nije
3. Gatekeeper, ulazni predstavnik
4. Represenative, izlazni predstavnik
5. Veze



Mrežni modeli

Mrežni modeli daju odgovor na pitanje kako se mreža formira i menja kroz vreme, uvid u proces kreiranja mreže
Modeliraju ključne strukturne karakteristike mreže:

1. Distribucija čvorova po stepeu
2. Prosečne putanje u mreži
3. Dijametar mreže
4. Klastere u mreži
5. Koeficijent klasterizacije

Random network Model - Erdoš renijeve mreže
Mreža je neusmerena, u početku sadrži sve čvorove, a nijednu vezu, ali se veze formiraju sa uniformnom verovatnoćom, i potpuno nezavisno jedna od drugih

- Da bi se izgradila mreža:
  - Započne se sa *N* izolovanih čvorova
  - Izabere se jedan par čvorova
  - Generiše se slučajan broj u opsegu od 0 do 1
  - Ako je generisani broj veći od *p*, izabrani čvorovi se povezuju granom
  - Prethodni postupak se ponavlja za svih *N(N-1)/2* parova čvorova
- *G(N,p)* model (*Erdos-Renyi* mreža, *Gilbert*-ov model)
  - Prazan model, *null* model mreže
- Alternativa – *G(N,L)*
  - *L* predstavlja maksimalan broj na slučajan način raspoređenih veza

Broj veza između čvorova kod ovih mreža sledi binomnu raspodelu, dešava se nagomilavanje oko prosečnog stepena nezavisno od veličine mreže
Za određene vrednosti p dolazi do značajnih promena u strukturi ER mreže
Pragovi u mreži i faze prelaza (subkritični, super kritični i povezani režim)

- Značajne faze prelaza:
  - Prag za stvaranje veza unutar mreže
    - $p=1/n^2$, prosečan stepen čvora je tada ~ *1/n*
  - Prag za pojavu ciklične putanje i gigantske komponente (kritična tačka)
    - *p=1/n*, prosečan stepen čvora je tada ~ *1*
  - Prag za stvaranje povezane mreže
    - *p=ln(n)/n*, prosečan stepe čvora je tada ~ *ln(n)*

Realne mreže najčešće ne slede Pausonovu raspodelu, zato što postoji značajno veća varijacija u stepenu čvorova, postoje čvorovi koji su izolovani (sa malo veza), kao i oni koji prestavljaju habove (sa puno veza)
ER model se razmatra u kontekstu velikih mreža, kada n teži beskonačno, omogućava analizu na osnovu verovatnoće i statistike i najbolje opisuje realne mreže kada je p > 1/N
Koristi se za upoređivanje sa realnim mrežama, gde tražimo neka velika odstupanja
I slučajne i realne mreže karakteriše mali dijametar i mala prosečna dužina putanje
Varijacije:
Model upoznavanja – u jednom režimu radi nasumičan pristup, a u drugom upoznavanja preko prijatelja
Statički geo-model – uzme se u obzir i pozicija čvora u prostoru, pa se svaki čvor povezuje sa zadatim brojem najbližih suseda
Model slučajnog susreta – opet se stave u prostoru pa se uzme n body simulacija i ako se sudare formira se veza
Model rasta - počinje se od malo broja čvorova koji formiraju kliku pa se dodaju novi čvorovi i na random povezuju sa tim

Fenomen malog sveta -

Mala prosečna odaljenost
Visok stepen klasterizacije
Prosečna udaljenost dve osobe koje se ne poznaju iznosi 6 koraka, toliko im trebalo za neka pisma

Stepen razdvajanje je za jedan manji od prosečne dužine putanje
Teorema mrežne strukture:

Velika verovatnoća postojanja puta između dva čvora, mreža je uglavnom povezana
Svaki čvor nije direktno povezan sa svakim drugim čvorom mreže

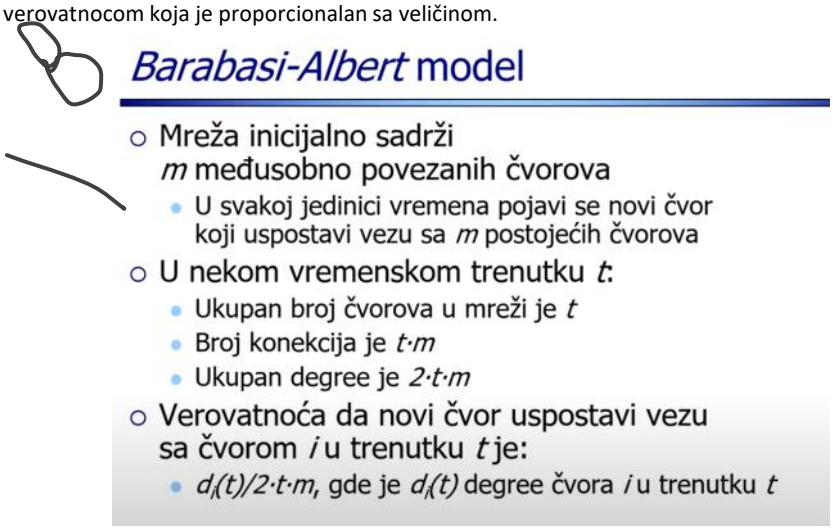
- U dovoljno velikoj *G(N,p)* mreži, za prosečnu dužinu puta l_{av} važi sledeći odnos:

$$l_{av} \sim ln(n) / ln(d)$$

gde je *d* prosečan stepen čvorova u mreži

Prethodno opisani modeli imaju problem opisivanja realnih mreža koje sadrže habove, pa se za to koriste modeli preferencijalnog vezivanja, tako što se novi čvorovi kače za one čvorove koji imaju visok stepen. Ovakve mreže imaju mali broj sa visokim stepenom, a veliki broj čvorova sa niskim stepenom – Barbasi-Albert model, power low raspodelu prati, a ovde se povezuje sa verovatnocom koja je proporcionalan sa veličinom.

imaju visok stepen. Ovakve mreže imaju mali broj sa visokim stepenom, a veliki broj čvorova sa niskim stepenom – Barbasi-Albert model, power low raspodelu prati, a ovde se povezuje sa verovatnocom koja je proporcionalan sa veličinom.



Identifikacija brokera(mostova) se vrši na osnovu:

1. Relacione centralnosti
2. Koeficijenta posredovanja (brokerage)
3. Mrežnih ograničenja (network constraint)

Mrežna ograničenja pokazuje koliko je član mreže povezan sa drugim članovima mreže koji su već međusobno povezani, što su susedni akteri bolje povezani, to više ograničava njegove mogućnosti za delovanje
Brokeri imaju visoku relacionu centralnost i relativno niska mrežna ograničenja
Habovi, članovi mreže sa najvećim uticajem, na osnovu ulazne centralnosti po stepenu

Čišćenje podataka

Validnost nam znači da podaci zadovoljavaju neka ograničenja tip, i slično ali to ne znači da su ti podaci tačni tj. Da postoje u realnosti.
Treba praviti i razliku između tačnosti i preciznosti
Proces čišćenja podataka se sastoji od:

1. Inspekcija, detekcija neočekivanih, nekorekthin i nekozistetnih podataka
2. Čišćenje, modifikacija ili uklanjanje problematičnih podataka
3. Verifikacija
4. Izveštavanje

Ako imam pogrešne podatke kao m, Male, femalE, felmal, f kako mogu da ih mapiram u samo dve vrednosti:

1. Rečnik za mapiranje, prvo identifikujem sve unique values pa ih mapiram posle na osnovu rečnika
2. Korišćenjem regularnih izraza npr sve na m ili f što počinje
3. Aproksimativno mapiranje, zasnovano na nekoj meri distance od korektne forme:
    a. Levenštajnova distanca – koliko slova treba da se promeni da bi došli do korektne vrednosti, pa uzimamo min

Skaliranje podataka

1. Scaling to a range
2. Feature clipping, odsecanje odlika, svi koji odstupaju se stave na jednu max vrednost
3. Log scaling, vrednost se zamenjuje sa sopstvenim algoritmom, za podatke koji prate power low raspodelu
4. Z-score prvo se odredi aritmetička sredina, pa standardna devijacija i onda se to stavi na z-score skalu gde gledam koliko standradnih devijacija odstupamo od srednje vrednosti

Bibliometrija I naukometrija – ako pisem master mogu ovo da pogledam opet

Naukometrija - je nauka koja se bavi metrikama naučne produkcije
Meri se produktivnost u nauci, tehnologiji i inovacijama

Od interesa je i uticaj koji se bazira na citiarnosti
Parametri zavise od oblasti
Bibliometrijska analiza – je statistička analiza publikacija, koristi se za kvantitativnu analizu naučne literature
Broj publikacija – problem kvantitet vs kvalitet, i postoje radovi sa multi-autora gde se nalazi potpis sa preko 1k autora
Impakt faktor IF – broj citata u tekućoj godini koji citiraju radove iz dve prethodne godine, podeli sa brojem objavljenih radova u te dve godine, uzimaju se citati iz ograničenog skupa časopisa
IF za datu godinu se objavljuje u junu sledeće godine, to se radi za časopise
Postoji i petogodišnji

H indeks - ukjljučuje dve dimenzije, produktivnost i citiranost. Naučnik ima H indeks vrednosti h, ako ima h radova, citiranih h puta
Bitno je gde su radovi citirani i favorizuje naučnike sa dužim radnim stažom, neki predlažu normalizaciju po godinama radnog staža

G indeks - . Naučnik ima G indeks vrednosti g, ako ima g radova, citiranih g*g puta
I10 – broj radova sa barem 10 citata
Izvori podataka: Web of Science

Transportne mreže

Mreža gradskog saobraćaja se može posmatrati kao jedna kompleksna mreža. Čvorovi su linije ili stajališta, tipično je neusmereni graf.

L – prostor model, čvorovi stajališta, grana postoji između čvorova ukoliko bar jedna linija sukcesivno prolazi kroz oba stajališta. Gubi se informacija o linijama, APL koliko se u proseku prolazi stanica na putu izmedju dva stajališta, stepen čvora, broj pravaca u kojima neko može da se kreće koristeći javni prevoz

B-prostor, bipartini grf, čvorovi predstavljaju stajališta i linije, svaka linija je povezana sa svim stajalištima kroz koje prolazi, nema direktnih veta između čvorova istog tipa, susedi neke linije su sva stajališta kroz koja ona prolazi

P-prostor čvorovi stajališta, grana postoji ukoliko bar jedna linija prolazi kroz oba čvora, svaka linija proizvodi po jedan kompletan podgraf. APL prosesčan broj presedanja koje neko treba da izvede da bi promenio svoju rutu i prešao na drugu. Stepen definiše do koliko različitih stajališta je moguće doći bet pomene linije.

C-prostor čvorovi linije, grane postoje ukoliko odgovarajuće linije imaju barem jednu zajedničku stanicu, APL koliko u proseku iznosi broj presedanja između različitih linija, na koliko drugih linija je moguće presedanje sa date linije

Scale free mreže, uklapaju se u model malog sveta,

Bipartne mreže

Bipartni graf se koristi za modelovanje entiteta iz stvarnog sveta, gde su čvorovi entiteti, a veze predstavljaju njihe međusobne veze.
Graf G = (U, V, E) je bipartitan (bimodalni) ako:

1. Čvorovi mogu biti podeljeni u dva odvojena skupa U i V
2. Svaka grana iz skupa E povezuje dva čvora koja pripadaju različitim skupovima
    a. Čvor iz U može biti povezan samo sa jednim ili više čvorova iz skupa V

Graf koji ne sadrži cikluse neparne dužine je po definiciji bipartitan, može se obojiti korišćenjem dve boje, bipartitivnost je mera koja kvantifikuje koliko je zadata mreža blizu toga da bude bipartitivna
Osobine:

1. Stepen čvora je ograničen brojem čvorova drugog skupa
2. Suma stepenova jednog skupa je jednaka sumi stepenova drugog skupa
3. Svi putevi između dva čvora iz istog skupa su parne dužine

Koeficijent klasterizacije se ne može direktno koristiti
Radi se projekcija iz bipartivnog u unipartivni graf
Projekcija treba da sadrži važne veze između izabranog skupa čvorova
Potrebna je ekstrakcija kičme mreže, izdvajanje najvažnijih relacija koje nose informacije

Ovo je problematično uraditi za (težinske) bipartitne grafove
Tehnike agregacije za unipartnu projekciju:

1. Broj deljenih (preklapajućih) čvorova drugog skupa
2. Suma minimalnih težina grana koje spajaju čvorove prvog skupa preko čvora drugog skupa
3. Dvoprolazna agregacija

Rezultujući graf treba da bude značajno ređi, potrebna dodatna obrada za eliminaciju grana koje nose malo informacija, redudantne grane
Dvoprolazna agregacija:

1. Svakom čvoru prvog skupa se dodeli jedna jedinica resursa
2. Resurs se rasšpdečo susedima iz drugog skupa proporcionalno težinama grana koje izlaze iz čvora
3. Dobijeni resursi se na isti način vrate nazad

Konfiguracioni model - metod generisanja slu;ajne mreže, na osnovu sekvence stepena čvorova, ne ograničava se na pausonovu raspodelu, referentni model za socijalne I kompleksne mreže
Jednostavan algoritam:

1. Pridružiti redom svakom čvoru jedan od zadatih stepenova u obliku polu-linkova
    a. Ukupan broj polu-linkova mora biti paran
2. Izabrati dva polu-linka i povezati ih, esencijalno formirajući granu
    a. Ponavljati postupak dok svi čvorovi ne budu povezani

Topologija dobijene mreže zavisi od redosleda izbora polu-linkova. Problem petlji i paralelnih grana