SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO-MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Ivana Ciganović

MATEMATIČKI PRISTUP SOCIJALNOM UMREŽAVANJU

Diplomski rad

Voditelj rada:

Prof. dr. sc. Siniša Slijepčević

Zagreb, rujan 2014.

Ovaj diplomski rad obranjen je dana ________________ pred ispitnim povjerenstvom u sastavu:

1. _________________________________ , predsjednik
2. _________________________________ , član
3. _________________________________ , član

Povjerenstvo je rad ocijenilo ocjenom _____________________ .

Potpisi članova povjerenstva:

1. _________________________________
2. _________________________________
3. _________________________________

Sadržaj

Sadržaj

Uvod

U ovom radu promatrati ćemo stohastičke modele za društvene mreže podataka. Također ćemo uvesti posebne vrste statističkih distribucija za slučajno usmjerene grafove, za koje ćemo pokazati kako su poseban slučaj uniformne distribucije slučajno usmjerenog grafa.

Ovdje opisani modeli su modeli dijadičke interakcije koji koriste (prirodni) logaritam vjerojatnosti kao svoju osnovnu mjeru. Modeli prikazuju strukturalnu formu (prirodnog) logaritma vjerojatnosti da promatrani član i odabere promatranog člana j s nekom težinom, dok promatrani član j odabire promatranog člana i s ne nužno jednakom težinom.

Statistička mrežna analiza dozvoljava pristupanje modelu mjereći prilagodbu modela zadanim podacima. Također, statistički pristup dozvoljava fleksibilne vjerojatnosne modele koji mogu biti generalizirani koristeći slučajne usmjerene distribucije bazirane na karakteristikama mreže. Te distribucije dozvoljavaju usporedbe promatranih rezultata sa pretpostavljenim hipotezama, kao i testove značajnosti da bi došli do zaključka je li rezultat dobiven zbog varijabilnosti uzorkovanja.

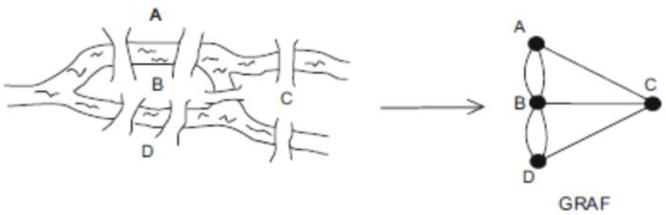
Prezentirat ćemo modele za mrežu s mjerenjima za jednostruke, direktne veze za jednu grupu promatranih podataka. Zatim ćemo opisati i prikazati interpretaciju i prilagodbu osnovnog statističkog modela za mrežu. Dodatne mjerene varijable mogu se uvrstiti u modele kako bi se dobila fleksibilnost u strukturi mrežnog modela između pojedinih podataka u situacijama u kojima su podskupovi podataka a priori bazirani na dodatnim varijablama. Također, možemo promatrati modele koji su fokusirani samo na veze među grupama ili pojedincima, do toga da možemo izbaciti pojedine podatke, ili skupove podataka. Rad se zasniva na petnaestom poglavlju knjige autora Stanely Wasserman i Katherine Faust: Social network analysis (vidi [10]).

Poglavlje 1

Osnovne definicije i pojmovi teorije grafova

U ovom poglavlju dane su osnovne definicije i pojmovi dijela teorije grafova.

Što je graf? Pojednostavljeno rečeno, graf je familija točaka, koje se zovu vrhovi, zajedno sa spojnicama među vrhovima, koje se zovu bridovi. Uz takvu ’definiciju’, graf je lako nacrtati:



Slika 1.1: Primjer grafa

• Skup vrhova $\mathcal{N} = \{A, B, C, D\}$

• Skup bridova $\mathcal{L} = \{\{A, B\}, \{A, B\}, \{B, D\}, \{B, D\}, \{A, C\}, \{B, C\}, \{C, D\}\}$

Definicija 1.1. *Graf* $\mathcal{G}$ *je uređena trojka* $\mathcal{G} = (\mathcal{N}(\mathcal{G}), \mathcal{L}(\mathcal{G}), \psi_{\mathcal{G}})$ *koja se sastoji od nepraznog skupa* $\mathcal{N} = \mathcal{N}(\mathcal{G}) = \{n_1, n_2, \dots, n_g\}$ *čiji su elementi vrhovi od* $\mathcal{G}$, *skupa* $\mathcal{L} = \mathcal{L}(\mathcal{G}) = \{l_1, l_2, \dots, l_L\}$ *disjunktnog sa* $\mathcal{N}$ *čiji su elementi bridovi od* $\mathcal{G}$ *i funkcije incidencije* $\psi_{\mathcal{G}}$ *koja svakom bridu od* $\mathcal{G}$ *pridružuje neuređeni par (ne nužno različitih) vrhova od* $\mathcal{G}$.

Iz definicije je jasno sljedeće:

• $|\mathcal{N}| \geq 1$, tj. graf mora sadržavati najmanje jedan vrh

• moguće je da $\mathcal{L} = \emptyset$, tj. graf ne mora sadržavati bridove

• ako $e, f \in \mathcal{L}$, $e \neq f$, i $u, v \in \mathcal{N}$, $u \neq v$, tada je moguće da $\psi(e) = \psi(f) = \{u, v\}$, tj. funkcija ψ ne mora biti injekcija (moguće je da dva različita vrha budu spojena sa više bridova, tj. da graf sadrži višestruke bridove)

• ako $e \in \mathcal{L}$ i $v \in \mathcal{N}$, tada je moguće da $\psi(e) = \{v, v\}$, tj. brid može spajati neki vrh sam sa sobom i zove se petlja

• skupovi $\mathcal{N}(\mathcal{G})$ i $\mathcal{L}(\mathcal{G})$ ne moraju biti konačni; ukoliko su oba skupa konačna, tada kažemo da je $\mathcal{G}$ konačan graf.

Ako $u, v \in \mathcal{N}(\mathcal{G})$ i $e \in \mathcal{L}(\mathcal{G})$ tako da $\psi_{\mathcal{G}}(e) = \{u, v\}$, kažemo da e spaja u i v, a u i v su krajevi od e. Za krajeve u, v brida e kažemo da su incidentni sa bridom e. Štoviše, reći ćemo da su v i e incidentni, ako je v jedan kraj brida e. Za dva vrha incidentna s nekim bridom kažemo da su susjedni. Za dva brida sa zajedničkim vrhom također kažemo da su susjedni.

Napomena 1.2. *Radi jednostavnosti pisati ćemo $\psi_{\mathcal{G}}(e) = uv$ (ili $\psi_{\mathcal{G}}(e) = vu$).*

• Jednostavan graf je graf koji ne sadrži petlje ni višestruke bridove

• Multigraf je graf koji ’dozvoljava’ višestruke bridove, ali ne ’dozvoljava’ petlje

• Pseudograf je (multi) graf koji ’dozvoljava’ (višestruke) petlje

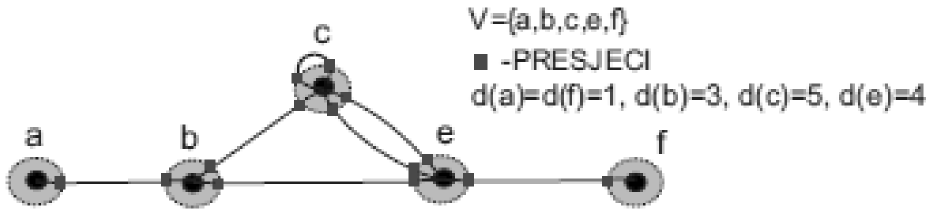
Definicija 1.3. *Hipergraf $\mathcal{H}$ je uređeni par $(\mathcal{N}(\mathcal{H}), \mathcal{L}(\mathcal{H}))$ koji se sastoji od nepraznog skupa vrhova $\mathcal{N}(\mathcal{H})$ i familije $\mathcal{L}(\mathcal{H})$ podskupova od $\mathcal{N}(\mathcal{H})$, tj. $\mathcal{L}(\mathcal{H}) \subseteq \mathcal{P}(\mathcal{N}(\mathcal{H}))$. Elemente od $\mathcal{L}(\mathcal{H})$ zovemo hiperbridovima.*

• Digraf možemo smatrati grafom čijim je bridovima pridružen smjer.

• Iz definicije hipergrafa jasno je da hiperbrid može spajati više od dva vrha. Ako u hipergrafu $\mathcal{H}$ skup hiperbridova $\mathcal{L}(\mathcal{H})$ sadrži isključivo dvočlane podskupove skupa $\mathcal{N}(\mathcal{H})$ tada dobivamo graf.

Intuitivno, stupanj vrha u grafu je broj sjecišta male kružnice oko vrha sa linijama koje izlaze iz vrha.

Definicija 1.4. *Stupanj vrha (ili valencija vrha) grafa $\mathcal{G}$ je broj $d_{\mathcal{G}}(v)$ bridova u $\mathcal{G}$ incidentnih sa v.*



Slika 1.2: Određivanje stupnja vrha

Napomena 1.5. *Pri računanju stupnja vrha, svaku petlju računamo kao dva brida. Također, za jednostavne grafove $\mathcal{G}$, stupanj vrha $v \in \mathcal{N}(\mathcal{G})$ se često definira kao kardinalni broj skupa $N_{\mathcal{G}}(v)$ pri čemu je $N_{\mathcal{G}}(v)$ skup svih susjeda od v. Najveći stupanj grafa $\mathcal{G}$ označavamo s $\Delta(\mathcal{G})$; $\Delta(\mathcal{G}) = max_{v\in\mathcal{N}(\mathcal{G})}\, d_{\mathcal{G}}\,(v)$, dok najmanji stupanj grafa $\mathcal{G}$ označavamo s $\delta(\mathcal{G})$; $\delta(\mathcal{G}) = min_{v\in\mathcal{N}(\mathcal{G})}\, d_{\mathcal{G}}(v)$. Prosječan stupanj grafa $\mathcal{G}$ označavamo s $d(\mathcal{G})$; $d(\mathcal{G}) = \frac{1}{|\mathcal{N}(\mathcal{G})|}\sum_{v\,\in\mathcal{N}(\mathcal{G})} d_{\mathcal{G}}(v)$. Jasno je da $\delta(\mathcal{G}) \leq d(\mathcal{G}) \leq \Delta(\mathcal{G})$. Ako je $d_{\mathcal{G}}(v) = 0$, onda za vrh v kažemo da je izolirani vrh grafa $\mathcal{G}$.*

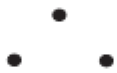
Propozicija 1.6. *Za svaki graf vrijedi $\sum_{v\in\mathcal{N}}\; d(v) = 2|\mathcal{L}|$.*

Dokaz: Stupanj svakog vrha definira se kao broj bridova incidentnih tom vrhu. Kako svaki brid ima dva kraja, broj stupnjeva u grafu mora biti jednak dvostrukom broju bridova. □

Korolar 1.7. (*Lema o rukovanju) U svakom je grafu broj vrhova neparnog stupnja paran broj.*

Dokaz: Prema prethodnoj propoziciji suma svih stupnjeva u grafu je paran broj. Neka su $\mathcal{N}_1$ i $\mathcal{N}_2$ redom skupovi vrhova neparnog i parnog stupnja u grafu, $\mathcal{N}_1 \cup \mathcal{N}_2 = \mathcal{N}$. Tada je $\sum_{v\in\mathcal{N}_1} d(v) + \sum_{v\in\mathcal{N}_2} d(v) = \sum_{v\in\mathcal{N}} d(v) = 2|\mathcal{L}|$. Kako je broj $\sum_{v\in\mathcal{N}_2} d(v)$ paran, slijedi da je i $\sum_{v\in\mathcal{N}_1} d(v)$ paran broj, pa je $|\mathcal{N}_1|$ paran broj. □

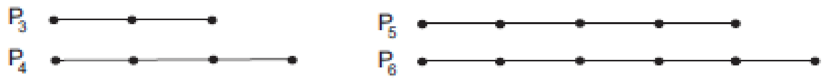
• Prazan graf je graf u kojem nema bridova.



Slika 1.3: Prazan graf sa tri vrha

• Put sa n vrhova P_n je jednostavan graf definiran sa:

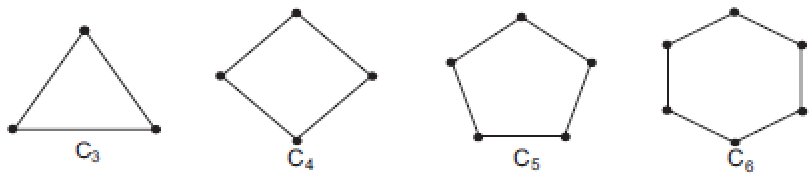
$$\mathcal{N}(P_n) = \{v_1, v_2, \dots, v_g\} \ , \ \mathcal{L}(P_n) = \{v_i \, v_{i+1} : i = 1, \dots, g-1\}$$



Slika 1.4: Primjeri puteva

• Ciklus sa g vrhova C_g je jednostavan graf definiran sa:

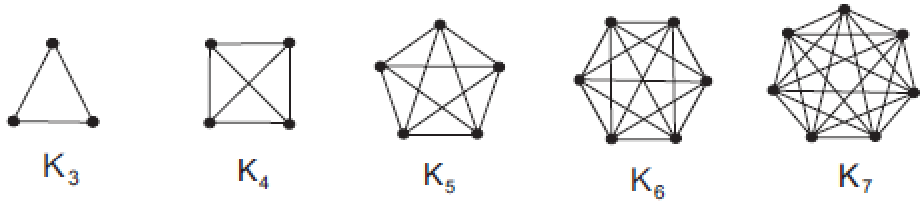
$$\mathcal{N}(C_g) = \{v_1, v_2, \dots, v_g\} \ , \ \mathcal{L}(C_g) = \{v_1 \, v_2, v_2 \, v_3, \dots, v_{g-1} \, v_g, v_g \, v_1\}$$



Slika 1.5: Primjeri ciklusa

• Potpun graf je jednostavan graf u kojem je svaki par vrhova spojen bridom.

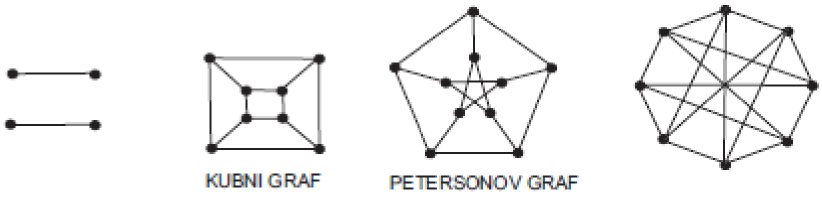
Oznaka za potpun graf s g vrhova je K_g.



Slika 1.6: Primjeri potpunih grafova

• Graf je r-regularan ako mu je svaki vrh stupnja r, tj. $d(v) = r, \forall v \in V$.

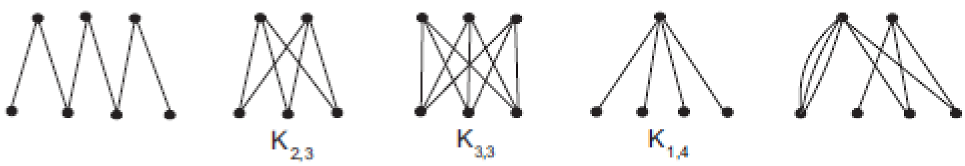
Graf je regularan ako je r-regularan za neko r.



Slika 1.7: Primjeri r-regularnih grafova

• Bipartitan graf je graf čiji se skup vrhova može particionirati u dva (međusobno disjunktna) skupa X i Y tako da svaki brid ima jedan kraj u X, a drugi u Y. Particija (X, Y) zove se biparticija grafa.

• Potpun bipartitan graf jednostavan je bipartitan graf s particijom (X, Y) u kojem je svaki vrh u X spojen sa svakim vrhom u Y. Uz $|X| = m$ i $|Y| = n$, oznaka takvog grafa je $K_{m,n}$.



Slika 1.8: Primjeri bipartitnih grafova

Definicija 1.8. *Graf $\mathcal{H}$ je podgraf od $\mathcal{G}$, u oznaci $\mathcal{H} \subseteq \mathcal{G}$, ako je $\mathcal{N}(\mathcal{H}) \subseteq \mathcal{N}(\mathcal{G})$, $\mathcal{L}(\mathcal{H}) \subseteq \mathcal{L}(\mathcal{G})$, a $\psi_{\mathcal{H}} = \psi_{\mathcal{G}}\big|_{\mathcal{L}(\mathcal{H})}$ (tj. $\mathcal{H}$ je restrikcija od $\mathcal{G}$ na $\mathcal{L}(\mathcal{H})$).*

• Ako je $\mathcal{H} \subseteq \mathcal{G}$ i $\mathcal{H} \neq \mathcal{G}$, pišemo $\mathcal{H} \subset \mathcal{G}$ i $\mathcal{H}$ zovemo pravi podgraf od $\mathcal{G}$.

• Ako je $\mathcal{H}$ podgraf od $\mathcal{G}$, onda je $\mathcal{G}$ nadgraf od $\mathcal{H}$.

Najjednostavniji tipovi podgrafova od $\mathcal{G}$ su oni koji su dobiveni izbacivanjem jednog vrha ili jednog brida iz $\mathcal{G}$:

• Ako je $v \in \mathcal{N}(\mathcal{G})$ i $|\mathcal{N}(\mathcal{G})| \geq 2$, tada $\mathcal{G} - v$ označava podgraf sa skupom vrhova $\mathcal{N}(\mathcal{G})\backslash\{v\}$, a čiji su bridovi svi oni iz $\mathcal{G}$ koji nisu incidentni sa v.

• Ako je $e \in \mathcal{L}(\mathcal{G})$, tada $\mathcal{G} - e$ označava podgraf sa skupom vrhova $\mathcal{N}(\mathcal{G})$ i skupom bridova $\mathcal{L}(\mathcal{G})\backslash\{e\}$.

Definicija 1.9. *Podgraf $\mathcal{H}$ od $\mathcal{G}$ za koji je $\mathcal{N}(\mathcal{G}) = \mathcal{N}(\mathcal{H})$ zove se podgraf koji razapinje $\mathcal{G}$ ili razapinjujući podgraf.*

Primijetimo:

• Podgraf $\mathcal{G} - e$ ima isti skup vrhova kao i $\mathcal{G}$ pa je razapinjujući podgraf.

• Ako sa $\mathcal{G} + f$ označimo graf dobiven iz grafa $\mathcal{G}$ dodavanjem brida f nekom paru ne nužno nesusjednih vrhova u $\mathcal{G}$, tada je $\mathcal{G}$ razapinjujući podgraf od $\mathcal{G} + f$.

Definicija 1.10. *Neka je $\emptyset \neq \mathcal{N}' \subseteq \mathcal{N}$. Podgraf od $\mathcal{G}$ čiji je skup vrhova $\mathcal{N}'$, a skup bridova je podskup bridova od $\mathcal{G}$ čija su oba kraja u $\mathcal{N}'$ zove se podgraf induciran sa $\mathcal{N}'$ i označava sa $\mathcal{G}[\mathcal{N}']$. $\mathcal{G}[\mathcal{N}']$ je inducirani podgraf od G.*

Definicija 1.11. *Za $\emptyset \neq \mathcal{L}' \subseteq \mathcal{L}$, slično se definira $\mathcal{G}[\mathcal{L}']$, tj. bridovima induciran podgraf od $\mathcal{G}$. Skup vrhova sačinjavati će svi oni vrhovi koji su krajevi bridova iz $\mathcal{L}'$.*

• Svaki inducirani podgraf od $\mathcal{G}$ može se dobiti uklanjanjem vrhova iz $\mathcal{G}$.

Definicija 1.12. *Šetnja u grafu $\mathcal{G}$ je netrivijalan konačan niz $W = n_0 l_1 n_1 l_2 n_2 \ldots l_g n_g$ čiji su članovi naizmjence vrhovi n_i i bridovi l_i, tako da su krajevi od l_i vrhovi n_{i-1} i n_i, za svako i,takvo da je $1 \leq i \leq g$.*

Kažemo da je W šetnja od n_0 do n_g ili $(n_0, n_g) -$ šetnja. Vrhovi n_0 i n_g zovu se redom početak i kraj šetnje W, a $n_1, n_2, \ldots, n_{g-1}$ unutarnji vrhovi od W. Broj g zove se duljina šetnje W.

Neka je W šetnja iz prethodne definicije. Inverzna šetnja W^{-1} od W je šetnja dobivena tako da se obrne redoslijed u W, tj. $W^{-1} = n_g l_g n_{g-1} \ldots l_1 n_0$. Ako je zadana šetnja $W' = n_g l_{g+1} n_{g+1} \ldots n_k$, tada je WW' šetnja dobivena povezivanjem W i W' kod vrha n_g i pišemo $WW' = n_0 l_1 n_1 l_2 n_2 \ldots l_g n_g l_{g+1} n_{g+1} \ldots l_k n_k$. Šetnja je zatvorena ako je $n_0 = n_g$. Ako su svi bridovi u šetnji W međusobno različiti, onda se W zove staza duljine g. Ako se šetnja W sastoji od međusobno različitih bridova i vrhova, onda se W zove put duljine k. Zatvorena šetnja u kojoj su svi vrhovi (osim, naravno, početnog i krajnjeg) i svi bridovi međusobno različiti zove se ciklus.

Napomena 1.13. *U jednostavnom grafu dovoljno je šetnju zadati samo nizom vrhova. Riječi "put" i "ciklus" označavaju specijalne šetnje, ali su ujedno i nazivi za specijalne grafove (ili podgrafove).*

Teorem 1.14. *Svaka (u, v) −šetnja u grafu $\mathcal{G}$ sadrži (u, v) −put.*

Dokaz: Neka je W (u, v) −šetnja u nekom grafu $\mathcal{G}$. Ako je W zatvorena, onda tvrdnja očito vrijedi. Pretpostavimo da W nije zatvorena i neka je $W \coloneqq u = u_0 e_1 u_1 e_2 u_2 e_3 u_3 \ldots e_k u_k = v$. Ako $u_i \neq u_j$ za svaki par $i, j \in \{0, 1, \ldots, k\}$, $i \neq j$, tada je W (u, v)-put. U suprotnom, pretpostavimo da $u_i = u_j$ za neke različite indekse i i j. Bez smanjenja općenitosti uzmimo da je $i < j$. Brisanjem vrhova $u_i, u_{i+1}, \ldots, u_{j-1}$ dobivamo novu (u, v) −šetnju W_1 koja je kraća od W. Ako u W_1 nema ponavljanja vrhova, tada je W_1 (u, v) −put. U suprotnom ponavljamo postupak skraćivanja sve dok ne dobijemo šetnju u kojoj su svi vrhovi međusobno različiti, tj. sve dok ne dobijemo put. □

Definicija 1.15. *Udaljenost $d_{\mathcal{G}}(u, v)$ dvaju vrhova $u, v \in \mathcal{N}(\mathcal{G})$ je duljina najkraćeg (u, v) −puta u $\mathcal{G}$. Ako ne postoji takav put u $\mathcal{G}$, stavljamo $d_{\mathcal{G}}(u, v) = \infty$.*

Za (u, v)-put kažemo da je trivijalan ako $u = v$.

Dva vrha u grafu $\mathcal{G}$ su povezana ako postoji (u, v) −put u $\mathcal{G}$.

Definicija 1.16. *Graf $\mathcal{G}$ je povezan ako $d_{\mathcal{G}}(u, v) < \infty$, $\forall u, v \in \mathcal{N}(\mathcal{G})$. U suprotnom, kažemo da je $\mathcal{G}$ nepovezan.*

Za svaki graf $\mathcal{G}$ možemo particionirati skup vrhova $\mathcal{N}$ na neprazne podskupove $\mathcal{N}_1, \mathcal{N}_2, \dots, \mathcal{N}_g$ tako da su dva vrha $u, v \in \mathcal{N}$ povezana u $\mathcal{G}$ ako i samo ako postoji $i \in \{1, \dots, g\}$ takav da je $u, v \in \mathcal{N}_i$.

Grafovi $\mathcal{G}[\mathcal{N}_1], \mathcal{G}[\mathcal{N}_2], \dots, \mathcal{G}[\mathcal{N}_g]$ zovu se komponente povezanosti od $\mathcal{G}$. Povezani grafovi imaju samo jednu komponentu povezanosti. Broj komponenata povezanosti označava se sa $c(\mathcal{G})$.

Napomena 1.17.

• *Povezanost je relacija ekvivalencije na skupu vrhova grafa.*

• *Komponente povezanosti nekog grafa $\mathcal{G}$ najveći su povezani podgrafovi tog grafa. Drugim riječima, ako je $\mathcal{H}$ komponenta povezanosti u $\mathcal{G}$, tada za svaki vrh $v \in \mathcal{N}(\mathcal{G}) \setminus \mathcal{N}(\mathcal{H})$ podgraf $\mathcal{G}[\mathcal{N}(\mathcal{H}) \cup \{v\}]$ nije povezan.*

• *Svaka komponenta povezanosti u $\mathcal{G}$ je inducirani podgraf od $\mathcal{G}$.*

Ekscentricitet $e(v)$ vrha v povezanog grafa $\mathcal{G}$ je $\max_{\mathrm{u} \in \mathcal{N}(\mathcal{G})} d(u, v)$.

Radijus $r(\mathcal{G})$ od $\mathcal{G}$ je $\min_{u \in \mathcal{N}(\mathcal{G})} e(v)$.

Dijametar $diam(\mathcal{G})$ od $\mathcal{G}$ je $\max_{u \in \mathcal{N}(\mathcal{G})} e(v)$.

Za vrh v kažemo da je centralni vrh grafa $\mathcal{G}$ ako vrijedi $e(v) = r(\mathcal{G})$. Centar grafa $\mathcal{G}$ $Cen(\mathcal{G})$ je podgraf od $\mathcal{G}$ induciran njegovim centralnim vrhovima.

Teorem 1.18. *Za svaki povezan graf* $\mathcal{G}$ *vrijedi* $r(\mathcal{G}) \leq diam(\mathcal{G}) \leq 2r(\mathcal{G})$.

Dokaz: Nejednakost $r(\mathcal{G}) \leq diam(\mathcal{G})$ je posljedica definicije radijusa i dijametra grafa $\mathcal{G}$. Dokažimo drugu nejednakost. Neka su u i v vrhovi u $\mathcal{G}$ takvi da $d(u,v) = diam(G)$. Nadalje, neka je w centralni vrh od $\mathcal{G}$. Obzirom da je $d : \mathcal{N}(\mathcal{G}) \times \mathcal{N}(\mathcal{G}) \to \mathbb{R}$ metrika na skupu svih vrhova $\mathcal{N}(\mathcal{G})$, imamo: $d(u,v) \leq d(u,w) + d(w,v) \leq 2e(w) = 2r(\mathcal{G})$. □

Teorem 1.19. *Svaki graf je centar nekog povezanog grafa.*

Dokaz: Neka je $\mathcal{G}$ proizvoljan graf. Konstruirajmo graf $\mathcal{H}$ tako da grafu $\mathcal{G}$ dodamo četiri nova vrha u_1, v_1, u_2, v_2 tako da je svaki vrh u $\mathcal{G}$ povezan sa vrhovima v_1 i v_2, a u_i i v_i su susjedi za $i = 1, 2$. Očito je da je $\mathcal{H}$ povezan. Tada je $e(u_i) = 4$ i $e(v_i) = 3$ za $i = 1, 2$. Također $e_{\mathcal{H}}(w) = 2$ $\forall w \in \mathcal{N}(\mathcal{G})$ pa slijedi da je $Cen(\mathcal{H}) = \mathcal{G}$. □

• Vrh v grafa $\mathcal{G}$ zovemo rezni vrh ako vrijedi $c(\mathcal{G} - v) > c(\mathcal{G})$.

• Brid e grafa $\mathcal{G}$ zovemo most ako vrijedi $c(\mathcal{G} - e) > c(\mathcal{G})$.

Sada ćemo dati karakterizaciju reznih vrhova.

Teorem 1.20. *Vrh* v *povezanog grafa* $\mathcal{G}$ *je rezni vrh ako i samo ako postoje vrhovi* u *i* w *(*$w \neq v$*) tako da* v *pripada svakom* (u,w) *−putu od* $\mathcal{G}$.

Dokaz:

⇒ Neka je v rezni vrh. Očito je $\mathcal{G} - v$ nepovezan graf. Ako su u i w vrhovi iz različitih komponenti grafa $\mathcal{G} - v$, tada nema (u,w) −puteva u $G - v$. Obzirom da je G povezan, postoje (u,w) −putevi u $\mathcal{G}$. Stoga, svaki (u,w) −put u $\mathcal{G}$ sadrži v.

⇐ Neka postoje vrhovi u i w u $\mathcal{G}$ takvi da v pripada svakom (u,w) −putu u G. Tada nema (u,w) −puteva u $\mathcal{G} - v$, a to znači da je $\mathcal{G} - v$ nepovezan pa je v rezni vrh. □

Na sličan ćemo način karakterizirati bridove-mostove.

Teorem 1.21. *Brid* e *povezanog grafa* $\mathcal{G}$ *je most ako i samo ako postoje vrhovi* u *i* w *takvi da* e *pripada svakom* (u,w) *−putu od* $\mathcal{G}$.

Teorem 1.22. *Svaki netrivijalan povezan graf sadrži najmanje dva vrha koji nisu rezni vrhovi.*

Dokaz: Pretpostavimo suprotno, tj. neka postoji netrivijalan povezan graf $\mathcal{G}$ takav da sadrži najviše jedan vrh koji nije rezni vrh, tj. svaki vrh grafa $\mathcal{G}$, osim eventualno jednog vrha, je rezni vrh. Neka su u i v vrhovi takvi da je $d(u,v) = diam(\mathcal{G})$. Slijedi da je najmanje jedan od tih vrhova rezni, npr. v. Neka je w vrh koji se nalazi u komponenti od $\mathcal{G} - v$ u kojoj nije u. Obzirom da svaki (u,w) −put u $\mathcal{G}$ sadrži v, zaključujemo da $d(u,w) > d(u,v) = diam(\mathcal{G})$ što ne može biti. □

Definicija 1.23. *Težinski graf $\mathcal{G}^{\alpha}$ je graf $\mathcal{G}$ čijim su bridovima pridruženi neki realni brojevi, tj. postoji težinska funkcija $\alpha : \mathcal{L}(\mathcal{G}) \to \mathbb{R}$ pri čemu broj $\alpha(e)$ zovemo težinom brida $e \in \mathcal{L}(\mathcal{G})$.*

Za podgraf $\mathcal{H}$ grafa $\mathcal{G}^{\alpha}$ broj $\alpha(\mathcal{H}) = \sum_{e\, \in\, \mathcal{L}(\mathcal{H})} \alpha(e)$ zovemo ukupnom težinom od $\mathcal{H}$. Specijalno, ako je $\mathcal{H}$ neki (u,v) −put u $\mathcal{G}^{\alpha}$, $u,v \in \mathcal{N}\,(\mathcal{G}^{\alpha})$, tada broj $d_{\mathcal{G}^{\alpha}}(u,v) = min\{\alpha(P){:}\, P\ je\ put\ od\ vrha\ u\ do\ vrha\ v\}$ zovemo minimalna težinska udaljenost između vrhova u i v.

Težinski grafovi se vrlo često pojavljuju u raznim primjenama. U takvom grafu potrebno je naći podgraf određenog tipa koji će imati najveću ili najmanju težinu.

Definicija 1.24. *Aciklički graf ili šuma je onaj koji ne sadrži cikluse. Stablo je povezan aciklički graf. Komponente povezanosti šume su stabla.*

Teorem 1.25. *Svaka dva vrha u stablu povezana su jednim jedinim putem.*

Dokaz: Neka je $\mathcal{G}$ stablo i neka su $\mathcal{P}_1$, $\mathcal{P}_2$ različiti (u,v) −putevi. Kako je $\mathcal{P}_1 \neq \mathcal{P}_2$, postoji brid $e = xy$ u $\mathcal{P}_1$ tako da e nije u $\mathcal{P}_2$. Očito $(\mathcal{P}_1 \cup \mathcal{P}_2) - e$ je povezan pa sadrži (x,y) −put $\mathcal{P}$. Slijedi $\mathcal{P} + e$ je ciklus što je kontradikcija. □

Korolar 1.26. *Povezan graf je stablo ako i samo ako su mu svi bridovi ujedno i mostovi.*

Dokaz:

$\Leftarrow$ Neka je svaki brid nekog povezanog grafa $\mathcal{G}$ most. Slijedi da graf ne sadrži cikluse pa smo dobili povezan aciklički graf, tj. stablo.

$\Rightarrow$ Neka je $\mathcal{G}$ stablo. Tada $\mathcal{G}$ ne sadrži cikluse pa mu svaki brid mora biti most. □

• Kažemo da je put $\mathcal{P} = (u, v)$ maksimalan put u grafu $\mathcal{G}$ ako ne postoji brid $e \in \mathcal{L}(\mathcal{G})$ za koji bi $\mathcal{P}e$ ili $e\mathcal{P}$ bio put.

Definicija 1.26. *Usmjereni graf ili digraf $\mathcal{D}$ je uređena trojka $(\mathcal{N}(\mathcal{D}), \mathcal{A}(\mathcal{D}), \psi_{\mathcal{D}})$ koja se sastoji od nepraznog skupa vrhova $\mathcal{N}(\mathcal{D})$, skupa lukova $\mathcal{A}(\mathcal{D})$ (ili usmjerenih bridova) i funkcije incidencije $\psi_{\mathcal{D}}$ koja svakom luku a pridružuje uređeni par (ne nužno različitih) vrhova u, v koje luk a spaja. Vrh u je početni, a v krajnji vrh od a.*

Svakom digrafu $\mathcal{D}$ možemo pridružiti pripadni graf s istim skupom vrhova, pri čemu svakom luku iz $\mathcal{D}$ pridružimo brid s istim krajevima. Obratno, svakom grafu $\mathcal{G}$ možemo pridružiti digraf tako da specificiramo početak i kraj svakog brida. Kažemo još da je takav usmjereni graf orijentacija na $\mathcal{G}$.

Većina pojmova koji vrijede za grafove, vrijede i za digrafove, ali pritom moramo paziti na pojmove vezane za orijentaciju jer se oni odnose samo na digrafove.

Na primjer, poddigraf digrafa se definira slično kao i podgraf grafa; usmjerena šetnja, staza, put, ciklus se definiraju slično kao i za grafove, a zovemo ih redom dišetnja, distaza, itd.

• Ako postoji usmjereni (u, v) $-$put u digrafu $\mathcal{D}$, kažemo da se vrh v može doseći u $\mathcal{D}$ iz vrha u.

• Dva vrha su dipovezana u $\mathcal{D}$ ako se svaki može doseći iz onog drugog.

• Digraf $\mathcal{D}$ je dipovezan ako su mu svaka dva međusobno različita vrha dipovezana.

• Za vrh v digrafa $\mathcal{D}$ definiramo ulazni stupanj ${d_{\mathcal{D}}}^{-}(v)$ i izlazni stupanj ${d_{\mathcal{D}}}^{+}(v)$ kao broj lukova u $\mathcal{D}$ s krajem, odnosno početkom v.

Slično definiramo minimalan ulazni stupanj $\rho^{-}(\mathcal{D})$ digrafa $\mathcal{D}$, minimalan izlazni stupanj $\rho^{+}(\mathcal{D})$,itd.

Napomena 1.27. *Primijetimo da digraf $\mathcal{D}$ ne mora biti dipovezan iako je njegov pripadni graf povezan.*

Notacija teorije grafova je jednostavan način prikazivanja članova mreže kao i veza među njima. Podatke prikazujemo u socio-matrici $\mathbb{X}$, pri čemu se retci i stupci odnose na veze među članovima. Dvije dimenzije socio-marice su indeksirane članovima pošiljateljima (retci) i članovima primateljima (stupci).

Definicija 1.28. *Pretpostavimo da postoji jednostruka veza mjerena na skupu koji sadrži g članova i označimo ih sa $\mathcal{N} = \{n_1, n_2, \dots, n_g\}$. Neka je $\mathcal{X}$ jednostruka direktna veza mjerena na uređenim parovima članova skupa $\mathcal{N}$. Elemente socio-matrice $\mathbb{X}$ definiramo sa : x_{ij}= vrijednost veze iz n_i u n_j za vezu $\mathcal{X}$, pri čemu $i, j = 1, \dots, g$ i $i \neq j$.*

Kako imamo g članova, dimenzija socio-matrice je $g \times g$. Ako je u pitanju dihotomna veza, tada su vrijednosti od x_{ij} jednake 0 ili 1. Uglavnom se na dijagonali nalaze nedefinirane vrijednosti, implicirajući da uglavnom član ne odabire samog sebe.

Primjer socio-matrice:

Promatrat ćemo šestero učenika drugog razreda i vezu prijateljstva među njima na početku školske godine i na kraju školske godine. Obadvije veze su dihotomne, dakle, $C_1 = C_2 = 2.$ U tablici ćemo prikazati socio-matrice za takve veze. Primijetimo da unos „1“ u socio-matrici implicira da postoji veza između člana n_i i člana n_j, dok „0“ govori da veza između članova ne postoji.

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	1	0	0	1	0
Drew	0	-	1	0	0	1
Eliot	0	1	-	0	0	0
Keith	0	0	0	-	1	0
Ross	0	0	0	0	-	1
Sarah	0	1	0	0	0	-

Tablica 1.1: Socio-matrica za vezu prijateljstva na početku školske godine

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	1	0	0	1	0
Drew	0	-	1	0	1	1
Eliot	0	0	-	0	1	0
Keith	0	1	0	-	1	0
Ross	0	0	0	1	-	1
Sarah	0	1	0	0	0	-

Tablica 1.2: Socio-matrica za vezu prijateljstva na kraju školske godine

Poglavlje 2

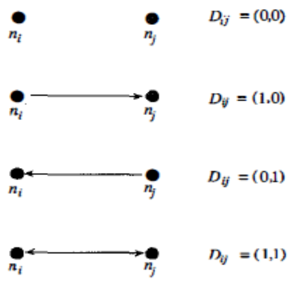
Dijade

2.1. Definicija dijade

Definicija 2.1.1. *Dijadu čine par članova i veza koja postoji među njima. Dijadu koja se sastoji od člana* i *i člana* j *označujemo sa* $D_{ij} = (X_{ij}, X_{ji})$. *Dijade su definirane za parove, gdje je indeks prvog člana manji od indeksa drugog člana (i< j). Ima točno* $\binom{g}{2} = g(g-1)/2$ *dijada.*

Postoje 3 moguća stanja u kojima se dijade mogu nalaziti:

- Uzajamna dijada između članova i i j postoji kada je $i \rightarrow j$ i kada $j \rightarrow i$. Označiti ćemo to stanje sa $i \leftrightarrow j$. Uzajamnu vezu primjećujemo u socio-matrici kada postoji simetrija s obzirom na dijagonalu. Dakle, $D_{ij} = (1,1)$.
- Sljedeće stanje je asimetrična dijada, koja se pojavljuje na dva načina; ili $i \rightarrow j$ ili $j \rightarrow i$, ali ne oboje. Dakle, $D_{ij} = (1,0)$ ili $D_{ij} = (0,1)$. Primijetimo da kod asimetričnih dijada veza nije uzajamna, tj. da ne postoji povratnost.
- Treće stanje dijade je nul-dijada, koje nastaje kada ne postoji veza između članova. Svakako, dijada koja nije uzajamna, ni asimetrična, mora nužno biti nul-dijada; $D_{ij} = (0,0)$.



Slika 2.1: Primjeri dijada

Vrijednosti socio-matrice $\mathbb{X}$ promatrat ćemo kao slučajne varijable, kao i dijade.

Ako su elementi matrice $\mathbb{X}$ binarni (promatramo dihotomnu vezu), pridružene dijade imaju bivarijatne slučajne vrijednosti (X_{ij}, X_{ji}). Ovaj par binarnih slučajnih varijabli može poprimiti 4 stanja, ovisno o vrsti veze u dijadi.

2.2. Dijadički cenzus

Dijada je zapravo primjer podgrafa - podskup vrhova uzetih iz $\mathcal{N}$, i svih bridova između njih. Postoji $\binom{g}{2}$ takvih podgrafova u usmjerenom grafu sa g vrhova. Kao što smo prije spomenuli, svaka od dijada mora biti uzajamna, asimetrična ili nul-dijada.

Definicija 2.2.1. *Definiramo M, A i N kao broj uzajamnih, asimetričnih i nul-dijada u skupu $\binom{g}{2}$ dijada. Trojka $\langle M, A, N\rangle$ se zove dijadički cenzus.*

Ta trojka se zove cenzus zato što je dobivena promatranjem svih dijada u mreži. Taj cenzus nam daje sveukupan pogled na dijade u mreži.

Možemo izračunati frekvencije $M, A, i\ N$ direktno preko elemenata socio-matrice $\mathbb{X}$ koja predstavlja digraf:

$$M = \sum_{i<j} X_{ij} X_{ji}$$

$$A = X_{++} - 2M$$

$$N = \binom{g}{2} - A - M\ ,$$

pri čemu je $X_{++} = L$, broj bridova u digrafu.

Možemo također izračunati $M, A, i\ N$ koristeći matrične operacije nad matricom $\mathbb{X}$:

$$M = \frac{1}{2} trag(\mathbb{X}\mathbb{X}),$$

$$A = trag(\mathbb{X}\mathbb{X}') - trag(\mathbb{X}\mathbb{X}),$$

$$N = \binom{g}{2} - trag(\mathbb{X}\mathbb{X}') + \frac{1}{2} trag(\mathbb{X}\mathbb{X}),$$

gdje je $\mathbb{X}'$ transponirana matrica $\mathbb{X}$. Transponirani element (i, j) matrice $\mathbb{X}$ je x_{ji}.

2.3. Indeksi uzajamnosti

Prvi indeks uzajamnosti

Katz i Powell (vidi [5]) su uveli ρ_{K_P}, da bi izmjerili tendenciju svakog člana grupe da uzvraća vezu češće nego što bi se dogodilo slučajno. Taj indeks nam daje „bolje" proučavanje nego dijadički cenzus, zato što ρ_{K_P} možemo koristiti za usporedbu grupe i veze među njima kada broj članova nije jednak.

Ako je $\rho_{K_P} = 0$ nema tendencije povratnosti, ako je $\rho_{K_P} = 1$ tendencija povratnosti je maksimalna, tj. sve veze su povratne. Ako je $\rho_{K_P} < 0$, promatramo premalo zajedničkih dijada. Dakle, ρ_{K_P} možemo koristiti da bi prikazali tendenciju povratnosti.

Ukratko, $-\infty < \rho_{K_P} \leq 1$, gdje 0 indicira da nema tendencije povratnosti, 1 je maksimalna tendencija povratnosti, dok negativne vrijednosti govore o premalo uzajamnih dijada, tj. prikazuju tendenciju prema asimetričnim i nul-dijadama.

Fiksan izbor

Prvo ćemo promatrati slučaj kada postoji fiksan izbor, tj. kada članovi imaju fiksan izbor mogućnosti. Pretpostavljamo da svaki član ima d izbora od $g-1$ članova koje može odabrati. Prikazati ćemo kako se računa ρ_{K_P}.

Jedan od prvih izračuna vjerojatnosti za zajednički izbor je napravljen od Paula Lazarsfelda i objavljen od strane Moreno i Jennigsa (vidi [8]). Ako je d fiksan broj izbora napravljen od svakog od g članova, tada je vjerojatnost za uzajaman izbor između bilo kojeg para članova $\frac{d^2}{(g-1)^2}$, u smislu da su izbori napravljeni posve slučajno i da članovi odgovaraju nezavisno.

Taj izračun koristi činjenicu da ako svaki član napravi d slučajnih izbora od $g-1$ ostalih članova, da je tada vjerojatnost da točno određeni član bude odabran $\frac{d}{g-1}$. Kako je odabir članova nezavisan, vjerojatnost da je dana dijada uzajamna je $(\frac{d}{g-1})^2$. Kako imamo $g(g-1)/2$ dijada, prosječan broj uzajamnih dijada mora biti broj dijada pomnožen s vjerojatnošću da bilo koja od njih bude uzajamna.

Očekivani broj zajedničkih dijada je

$$\mathbb{E}(M) = \frac{g(g-1)}{2}\frac{d^2}{(g-1)^2} = \frac{gd^2}{2(g-1)},$$

pretpostavljajući da svaki član odabire d drugih članova sasvim slučajno.

Vjerojatnost uzajamne dijade računamo kao produkt dviju vjerojatnosti:

$$P(i \to j, j \to i) = P(i \to j)P(j \to i | i \to j) \tag{2.1}$$

(pritom koristimo standardnu definiciju uvjetne vjerojatnosti).

Katz and Powell (vidi [5]) izražavaju tu uvjetnu vjerojatnost preko ρ_{K_P} :

$$P(j \to i | i \to j) = P(j \to i) + \rho_{K_P} P(j \nrightarrow i). \tag{2.2}$$

Ako je $\rho_{K_P} = 0$, tada je uvjetna vjerojatnost jednaka vjerojatnosti koja nije uvjetna i svi izbori su nezavisni, tj. nema tendencije povratnosti. Ako je $\rho_{K_P} = 1$, tada je desna strana gornje formule također jednaka 1. To nastaje kada izbor $i \to j$ povlači izbor $j \to i$, tj. te dvije povratne veze su zavisne.

Ako uvrstimo (2.2) u (2.1), dobijemo

$$P(i \to j, j \to i) = P(i \to j)P(j \to i | i \to j) = P(i \to j)\left[P(j \to i) + \rho_{K_P} P(j \nrightarrow i)\right] =$$
$$= \frac{d}{g-1}\left[\frac{d}{g-1} + \rho_{K_P} \frac{g-1-d}{g-1}\right].$$

Koristeći ovu vjerojatnost, očekivani broj uzajamnih dijada, uvjetno sa ρ_{K_P} je $\binom{g}{2}$ pomnožen s gornjom vjerojatnošću, tj. dobivamo

$$E(M|\rho) = \frac{gd^2}{2(g-1)}\left(1 - \rho_{K_P}\right) + \frac{gd}{2}\rho_{K_P} \tag{2.3}$$

Da bi procijenili ρ_{K_P} Katz i Powell (vidi [5]) su koristili procjenu očekivanja. Točnije, procjenjujemo očekivanu vrijednost uzajamnih dijada danu u formuli (2.3) koristeći broj uzajamnih dijada M. Ako desnu stranu formule (2.3) izjednačimo sa M, možemo riješiti takvu jednadžbu da bi dobili procjenu od ρ_{K_P}:

$$\hat{\rho}_{K_P} = \frac{2(g-1)M - gd^2}{gd(g-1-d)}. \tag{2.4}$$

Katz i Powell (vidi [5]) su pokazali da kod te procjene varijanca teži u 0 kako se g povećava, i da ima očekivanu vrijednost jednaku ρ_{K_P}. Jedna zanimljiva značajka ove procjene je da je to linearna funkcija od M. Kako se broj uzajamnih dijada povećava od 0 (što je minimum) do $\binom{g}{2}$ (što je maksimum), procjena se također linearno povećava. Primijetimo da kada je $M = 0$, procjena je negativna, što ne bi trebalo biti kada je izbor slučajan, što može biti posljedica toga da je d premalen.

Točnije, ako je $d > (g-1)/2$, tada minimalna vrijednost od $\hat{\rho}_{K_P}$ ne može biti negativna zato što broj uzajamnih dijada mora biti nenegativan. $\hat{\rho}_{K_P}$ postiže svoj maksimum 1, kada je $M = gd/2$ (što trivijalno proizlazi kada je $d = g - 1$).

Primjer

Da bi pokazali primjenu $\hat{\rho}_{K_P}$, promatrati ćemo ponovo Krackhardtovu mrežu prijateljstva, i pretpostaviti da svaki član odabire točno pet ostalih menadžera kao prijatelje. Imamo $g = 21$, $m = 23$, $d = 5$, pa iz jednadžbe (2.4) dobivamo

$$\hat{\rho}_{K_P} = \frac{2(20)(23)-(21)(5^2)}{21(5)(15)} = \frac{395}{1575} = 0.2508,$$

što ukazuje da takva mreža ima tendenciju nezavisnosti izbora napravljenog i primljenog u dijadi. Postoji mala tendencija da izbori budu uzajamni.

Slobodan izbor

Pretpostavimo sada da i −ti član odabire $d_0(n_i) = x_{i+}$ ostalih članova. Neka je $L = \sum x_{i+}$ ukupan broj izbora , i neka je $L_2 = \sum {x_{i+}}^2$ suma kvadrata izbora koji napravi svaki član. Kao u slučaju fiksnog izbora, vjerojatnost da članovi i i j imaju uzajamnu vezu je $x_{i+}x_{j+}/(g-1)^2$, ako su izbori napravljeni slučajno. Kako ta vjerojatnost ovisi o paru članova koje promatramo, očekivana vrijednost uzajamnih dijada je kompliciranija nego ranije spomenuto.

Katz i Powell (vidi [5]) su dali formulu:

$$E(M) = \frac{L^2-L_2}{2(g-1)^2}. \tag{2.5}$$

Koristeći (2.5) i (2.3) uvodimo ρ'_{KP} (analogno od ρ_{KP}) i računamo

$$E(M|\rho') = \frac{L^2-L_2}{2(g-1)}\left(1-\rho'_{KP}\right) + \frac{L}{2}\rho'_{KP}\ . \tag{2.6}$$

Kao i ranije, procjenjujemo ρ'_{KP} tako da desnu stranu formule (2.6) izjednačimo sa M i riješimo takvu jednadžbu:

$$\hat{\rho}'_{KP} = \frac{2(g-1)^2M-L^2+L_2}{L(g-1)^2-L^2+L_2}\ .$$

Primjer

Da bi to pokazali primjerom, promatrajmo broj bridova koji izlaze iz vrha usmjerenog grafa na Krackhardtovoj mreži, dakle izračunamo da je $L = 102$ (broj bridova) i da je $L_2 = 880$, suma kvadrata broja bridova koji izlaze iz vrha usmjerenog grafa. Iz tih vrijednosti, te koristeći da postoji $g = 21$ član i $M = 23$ dijada, imamo

$$\widehat{\rho}'_{KP} = \frac{2(20)^2(23) - 102^2 + 880}{102(20)^2 - 102^2 + 880} = \frac{8876}{31276} = 0.2838.$$

Tendencija prema uzajamnosti nije zanemariva.

Drugi indeks uzajamnosti

Achuthan, Rao i Rao (vidi [1]) promatraju broj uzajamnih dijada u usmjerenom grafu s određenim brojem bridova koji izlaze iz vrha usmjerenog grafa. Vrijednost M ovisi o broju napravljenih izbora. Koristeći neobjavljenu pretpostavku Bandyopadhyay ti autori su dali broj za uzajamnost koji se bazira na maksimalnoj i minimalnoj vrijednosti uzajamnih dijada M, koji se može pojaviti iz danog skupa bridova koji izlaze iz vrha usmjerenog grafa.

Definirat ćemo M_{min} i M_{max} kao minimalan i maksimalan broj uzajamnih dijada, pa je $M_{min} < M < M_{max}$. Te vrijednosti ovise o broju bridova koji izlaze iz vrha usmjerenog digrafa.

Achutan, Rao i Rao (vidi [1]) uvode

$$\rho_B = \frac{M - M_{min}}{M_{max} - M_{min}} \tag{2.7}$$

kao standardnu mjeru za uzajamnost u mreži.

Sada ćemo definirati minimum i maksimum uzajamnih dijada.

Promatrajmo prvo dvije funkcije

$$f(t) = \sum_{i=1}^{t} x_{i+} - t(g-1) - \binom{t}{2} \text{ i}$$

$$g(t) = \sum_{i=1}^{t} x_{i+} - t(t-1) - \sum_{i=t+1}^{g} \min(t, x_{i+}).$$

Tada je

$$M_{min} = \max_{0 \le t \le g} f(t) \text{ i} \tag{2.8}$$

$$M_{max} = \left\lfloor \frac{1}{2} \left\{ \sum_{i=1}^{g} x_{i+} - \max_{0 \le t \le g} g(t) \right\} \right\rfloor . \tag{2.9}$$

Sljedeće pitanje koje proizlazi iz prethodnih formula je unutar kojih uvjeta minimum uzajamnih dijada može biti jednak 0, tj. da li je moguće da nema uzajamnih dijada?

Pokazuje se da ako je $x_{min} = \min_i d_0(n_i)$ najmanji broj bridova koji izlaze iz vrha usmjerenog grafa, da je tada $M_{min} = 0$, ako je $x_{min} \le \lfloor (g-1)/2 \rfloor$. Dakle, ako svaki član ne odabere najmanje pola ostalih članova, moguće je da ne postoje uzajamne dijade u digrafu.

Također, postoje uvjeti pod kojima je maksimum uzajamnih dijada jednak $L/2$, pola broja bridova prikazanih u digrafu.

Primjer

Promatramo Krackhardtovu mrežu prijateljstva. Izračunato je da je $M_{min} = 0$ i $M_{max} = 51$. (koristeći program *Dijade* (vidi [9])).

Kako je $M = 23$, izračunamo $\hat{\rho}_B = \frac{23-0}{51-0} = 0.451$, što je prilično velika vrijednost, čak veća od $\hat{\rho}_{K_P}$. Općenito, $\hat{\rho}_B$ je veći od $\hat{\rho}_{K_P}$.

Krackhardtova mreža podataka sadrži i vezu za savjetovanje, mjerenu nad dvadeset jednim menadžerom. Gustoća ove druge veze je 190/420=0.452 i postoji 45 uzajamnih dijada. Obje vrijednosti su veće od onih za vezu prijateljstva.

Za tu vezu $\hat{\rho}_B = \frac{45-0}{95-0} = 0.474$ koja je samo malo veća od procjene za vezu prijateljstva. Dakle, iako ima skoro dvostruko više uzajamnih dijada za vezu savjetovanja, tendencija povratnosti je otprilike jednaka jednostavno zato što ima više veza prilikom savjetovanja (zbog činjenice da će menadžeri rjeđe biti prijatelji, dok češće traže ili daju savjet).

2.4. Jednostavne distribucije digrafova

Neka je $\mathcal{G}_d$ usmjereni graf (ili digraf) sa g vrhova. Sa $\mathcal{G}_d(\mathcal{N})$ označit ćemo sve moguće digrafove. Socio-matrica $\mathbb{X}$ sa elementima X_{ij} može biti korištena da bi detektirali sve bridove između vrhova grafa $\mathcal{G}_d$. Većina distribucija koje ćemo ovdje raspraviti biti će dana u terminima socio-matrice, ali znamo da postoji 1-1 korespodencija između digrafova i socio-matrica, pa te diskusije neće izgubiti na općenitosti. Digrafovi koje promatramo ovdje su slučajni, pa je tako i $\mathbb{X}$ slučajna socio-matrica. Jedna realizacija, ili jedna od mogućih vrijednosti ove slučajne socio-matrice biti će označena sa $\mathbb{x}$, sa elementima x_{ij}.

Jednostavne distribucije za slučajno usmjerene grafove su uniformna i Bernoullijeva. Pokazati ćemo kako se te distribucije koriste za proučavanje broja bridova prisutnih u digrafu.

Uniformna distribucija digrafa

Promatrajmo prvo koliko ima digrafova sadržanih u $\mathcal{G}_d(\mathcal{N})$, skupu svih mogućih usmjerenih grafova sa g vrhova. Kako nas zanimaju samo jednostruke dihotomne veze, svaki brid u digrafu ili postoji ili ne postoji, tj. svaka vrijednost socio-matrice je ili 0 ili 1.

Kako imamo g članova, od kojih svaki od njih može imati $g-1$ veza sa drugim članovima, postoji $g(g-1)$ mogućih bridova u digrafu. Zbog toga mora biti $2^{g(g-1)}$ različitih socio-matrica.

Na primjer, ako je $g = 3$, imamo $2^6 = 64$ različitih socio-matrica. Broj elemenata u $\mathcal{G}_d(\mathcal{N})$ raste eksponencijalno sa g, već kada imamo $g = 10$ članova, ima $2^{90} = 1.2379 \times 10^{27}$ mogućnosti. Dakle jasno je, da je broj mogućih realizacija slučajno usmjerenog grafa jako velik već za mali g.

Najjednostavnija distribucija od $\mathcal{G}_d(\mathcal{N})$ je uniformna distribucija. Oznaka $\mathbb{X} \sim U$. Svaka realizacija je jednako vjerojatna.

Vjerojatnosni prostor je $\mathcal{G}_d(\mathcal{N})$, koji sadrži $2^{g(g-1)}$digrafova, pa je uniformna vjerojatnosna funkcija dana sa

$$P(\mathbb{X} = \mathbb{x}) = \frac{1}{2^{g(g-1)}} \ . \tag{2.10}$$

Drugim riječima, vjerojatnost da usmjereni graf sa g članova bude jednak specijalnom broju izbora je $\frac{1}{2^{g(g-1)}}$. Zbog toga, svaki element vjerojatnosnog prostora ima jednaku vjerojatnost $\frac{1}{2^{g(g-1)}}$ da se dogodi. Jednostavnije je opisati bridove digrafa unutar ove distribucije kao statistički nezavisne, Bernoullijeve slučajne varijable sa vjerojatnošću izbora svakoga ½:

$$P\left(X_{ij} = 1\right) = \begin{cases} \frac{1}{2}, ako\ je\ i \neq j \\ 0, ako\ je\ i = j \end{cases}. \tag{2.11}$$

Nekad ćemo kraće zapisati $P\left(X_{ij} = 1\right)$ kao P_{ij}, vjerojatnost da je određeni brid prisutan u digrafu.

Ova distribucija ima najslabiju strukturu od svih distribucija digrafova. Svi elementi socio-matrice su nezavisni sa svim ostalim elementima, i vjerojatnosna distribucija bilo kojeg elementa je najjednostavnija moguća - Bernoullijeva distribucija sa jednakim vjerojatnostima.

To je isto kao i uniformna distribucija na istom prostoru sa dva elementa. Ova distribucija pretpostavlja da svi članovi odabiru oko polovice ostalih članova, pa je stupanj $(g-1)/2$ za sve članove, dok je očekivana gustoća 0.5.

Bernoullijeva distribucija

Općenita Bernoullijeva distribucija započinje sa jednadžbom (2.11) i dozvoljava da P_{ij} ne mora nužno biti ½. Bernoullijeva distribucija, koju označujemo sa B, pretpostavlja da su elementi od $\mathbb{X}$ statistički nezavisni. Dakle uniformnu distribuciju možemo gledati kao poseban slučaj Bernoullijeve distribucije, pri čemu su svi $P_{ij} = 1/2$.

Bridovi digrafa su Bernuollijeve slučajne varijable sa vjerojatnostima

$$P(X_{ij} = 1) = \begin{cases} P_{ij}, ako\ je\ i \neq j \\ 0, ako\ je\ i = j \end{cases} \tag{2.12}$$

gdje je $0 \leq P_{ij} \leq 1$.

P_{ij} se može razlikovati od elementa do elementa kako bi dozvolili da svaki član izabire druge članove sa različitim vjerojatnostima. Zbog toga, ova distribucija dozvoljava da neki bridovi u slučajnom digrafu imaju veće vjerojatnosti da budu prisutni, nego neki drugi bridovi. Ako slučajni usmjereni digraf ima B distribuciju i ako je $P_{ij} = \frac{1}{2}$ za sve $i \neq j$, tada je slučajan digraf uniformno distribuiran. Ako su svi P_{ij} jednaki, ali ne iznose ½ tada distribucija nije uniformna.

Poglavlje 3

Statistička analiza mreža s jednostrukom vezom

3.1. Jednostruke veze

Prvo ćemo opisati konstrukciju i modeliranje skupa $\mathbb{Y}$, kontingencijsku tablicu osnovnu za promatranje modela koja je dobivena iz podataka o vezama u skupu $\mathbb{X}$. Ti podaci se fokusiraju na dijade i opisuju na koji su način podaci povezani.

Prikazati ćemo ove metode na hipotetskom uzorku grupe učenika drugog razreda. Koristiti ćemo ovu društvenu mrežu kao lako zamisliv primjer zato što je u pitanju mala grupa, što omogućava lakše praćenje. Također, prikazati ćemo primjenu ovih metoda na Krackhardtovom odnosu prijateljstva koje je mjereno na menadžerima u high-tech organizaciji, te Padgetts-Florentine bračne i poslovne odnose. Analiza tih podataka prikazati će različite aspekte metodologije.

Skup $\mathbb{Y}$

Modeli za mreže s jednostrukom vezom ne prilagođavaju se dobro matrici $\mathbb{X}$ zbog čega je potrebno reorganizirati podatke za mrežu u drugačiju kontingencijsku tablicu kojoj se modeli bolje prilagođavaju. Prvo prikazujemo konstrukciju te nove tablice koristeći malu hipotetsku društvenu mrežu učenika drugog razreda.

Počinjemo opisujući model za jednostruku direktnu vezu mjerenu za jedan skup podataka $\mathcal{N}$. Dijade s podacima o mjerenju za direktnu vezu sastoje se od dva člana, i i j, te od veza između ta dva člana.

Veza između članova može se promatrati iz perspektive člana i ili iz perspektive člana j.

Prvo ćemo promatrati perspektivu n_i. Varijabla veze X_{ij} sadrži mogući odabir nekog n_j od strane člana n_i, dok varijabla X_{ji} sadrži mogući odabir nekog člana n_i od strane člana n_j. Promatrajući iz perspektive člana j varijabla veze X_{ji} sadrži moguće odabire i-ova za člana j, dok varijabla X_{ij} sadrži mogući odabir j-ova za član i. Obje perspektive su uključene u modeliranje.

Društvena mreža koja se sastoji od g članova, sadrži $\binom{g}{2}$ dijada. U statističkom modelu, svaka dijada se sastoji od informacija prikazanih s dvije slučajne varijable X_{ij} i X_{ji}. S D_{ij} označiti ćemo varijablu dijadi. S g članova i jednostrukom vezom imamo $g(g-1) = 2\binom{g}{2}$ dijadičkih slučajnih varijabli za promatranje. Modelirat ćemo istovremeno i što jednostavnije sve dijadičke veze u mreži. Promatrat ćemo par članova, jednostruku dihotomnu vezu i dijadu D_{ij}. Za oba člana veze u dijadi mogu biti prikazane sa 2×2 skupom. Prva varijabla sa dva nivoa, kojoj možemo pridružiti indeks k, i koja može biti samo 0 ili 1, sadrži vrijednost veze retka člana i i stupca člana j. Druga varijabla, također sa samo dva nivoa, koju možemo označiti s indeksom l, sadrži vrijednost veze stupca člana j s retkom člana i. Dakle, sve veze dijade D_{ij} mogu biti prikazane tim 2×2 skupom. Novi indeksi k i l iznose ili 0 ili 1 ovisno o stanju dijade.

Promatrajmo sada sve dijade i jednu dihotomnu vezu. Uzmemo li originalnu $g \times g$ binarnu matricu $\mathbb{X}$, i zamijenimo li svaku vrijednost unutar matrice s odgovarajućom 2×2 tablicom, dobivamo novu kontingencijsku tablicu. Kako ima $\binom{g}{2}$ dijada koje mogu biti indeksirane sa $g \times g$ parova promatranih članova, nova kontingencijska tablica će imati dimenziju $(g \times g) \times (2 \times 2)$.

Možemo promatrati vrijednosne i dihotomne veze. Da bi modelirali sve dijade za jednostruku vrijednosnu vezu istovremeno, napravit ćemo četverodimenzionalnu kontingencijsku tablicu dimenzija $(g \times g) \times (C \times C)$. Prve dvije dimenzije ove tablice su indeksirane članovima skupa $\mathcal{N}$. Veličina treće i četvrte dimenzije je C, broj cjelobrojnih vrijednosti koji mjerena relacija može poprimiti. Za dihotomne podatke indeksirane s k, l=0 ili 1, C iznosi 2. Za podatke o vezi indeksirane s k =0,1,2 i l =0,1,2 C iznosi 3.

Matricu dimenzija $(g \times g) \times (C \times C)$ zovemo matrica $\mathbb{Y}$ i definiramo s:

$$Y_{ijkl} = \begin{cases} 1, ako \;\; D_{ij} = (X_{ij} = k, X_{ji} = l) \\ 0, inače \end{cases}.$$

Vrijednosti skupa $\mathbb{Y}$ imaju četiri podskupa:

- članovi pošiljatelji (i),
- članovi primatelji (j),
- varijable relacije X_{ij}, (indeksirane po skupo k),
- varijable relacije X_{ji} (indeksirane po skupu l).

Struktura skupa $\mathbb{Y}$ je slična socio-matrici čiji retci predstavljaju članove pošiljatelje, a stupci članove primatelje. Vrijednost mjesta (i, j) u matrici je x_{ij}. Mjesto (i, j) u matrici nije broj, nego $C \times C$ podmatrica. U $C \times C$ podmatrici bit će samo jedna jedinica na mjestu (k, l). Preostalih $C^2 - 1$ mjesta će biti nule. Zbog toga te podmatrice možemo promatrati kao indikatorske matrice koje ce nam prikazati stanje svake dijade. Skup $\mathbb{Y}$ ima posebnu simetriju, Y_{ijkl}=Y_{jilk} za sve (i, j) i sve (k, l) parove, zbog činjenice da se dijade mogu promatrati iz različitih perspektiva, ili člana i ili člana j. Skup $\mathbb{Y}$ potreban je da bi promatrani model mogli prilagoditi relacijama sa diskretnim vrijednostima koristeći standardne log-linearne procedure modeliranja koje postoje u dostupnim statističkim paketima.

Primjer skupa $\mathbb{Y}$

		Allison	Drew	Eliot	Keith	Ross	Sarah
n_1	Allison	-	1	0	0	1	0
n_2	Drew	0	-	1	0	0	1
n_3	Eliot	0	1	-	0	0	0
n_4	Keith	0	0	0	-	1	0
n_5	Ross	0	0	0	0	-	1
n_6	Sarah	0	1	0	0	0	-

Tablica 3.1 Socio-matrica prijateljstva za učenike drugog razreda

U ovoj mreži, članovi su označeni na sljedeći način: n_1=Allison, n_2=Drew, n_3=Eliot, n_4=Keith, n_5=Ross, n_6=Sarah.

Kako bi se usredotočili na jednu dijadu, promatrat ćemo podatke za n_1=Allison i n_5=Ross za relaciju prijateljstva na početku školske godine. Podaci pokazuju da Ross ne navodi Allison kao učenicu koju simpatizira, ali Allison navodi Rossa. Iz njezine perspektive, varijabla relacije koju šalje je $x_{15} = 1$, implicirajući da smatra Rossa prijateljem, dok je varijabla koju prima $x_{51} = 0$ implicirajući da Allison nije navedena kao prijateljica od strane Rossa. Iz Rossove perspektive, poslana relacija je $x_{51} = 0$, dakle Ross ne bira Allison kao prijateljicu, a primljena relacija je $x_{15} = 1$, odnosno Ross je izabran od strane Allison. Podaci za članove 1 i 5 u paru $\langle n_1, n_5 \rangle$ su $D_{15} = (x_{15}, x_{51}) = (1,0)$, pa je $y_{1510} = 1$, dok $y_{1500} = y_{1501} = y_{1511} = 0$.

Neka je $\mathbb{Y}$ skup koji prikazuje mrežu izbora prijateljstava između tih šestero djece na početku školske godine.

Prikazat ćemo ove podatke kao socio-matricu u tablici (3.1). U tablici (3.2), prikazujemo skup $\mathbb{y}$ za ove podatke.

Veličina tog skupa je $(6 \times 6) \times (2 \times 2)$ zato što se kontingencijska tablica sastoji od članova $i = 1, \ldots, 6$, njihovih partnera $j = 1, \ldots, 6$ i težinom poslanih vrijednosti $x_{ij} = 0,1$, te težinom zaprimljenih vrijednosti $x_{ji} = 0,1$, dakle $C = 2$. U svakoj 2×2 podmatrici postoji jedna jedinica i $C^2 - 1 = 3$ nule. Podmatrice su na glavnoj dijagonali popunjene sa $-$, zato što povezanost ne može biti izmjerena sama sa sobom. I konačno, $\mathbb{y}$ je simetrična kako je i prije opisano, $y_{ijkl} = y_{jilk}$.

			j											
			Allison		Drew		Eliot		Keith		Ross		Sarah	
							$l = x_{ji}$							
i		k	0	1	0	1	0	1	0	1	0	1	0	1
n_1	Allison	$x_{ij} = 0$	-	-	0	0	1	0	1	0	0	0	1	0
		$x_{ij} = 1$	-	-	1	0	0	0	0	0	1	0	0	0
n_2	Drew	$x_{ij} = 0$	0	1	-	-	0	0	1	0	1	0	0	0
		$x_{ij} = 1$	0	0	-	-	0	1	0	0	0	0	0	1
n_3	Eliot	$x_{ij} = 0$	1	0	0	0	-	-	1	0	1	0	1	0
		$x_{ij} = 1$	0	0	0	1	-	-	0	0	0	0	0	0
n_4	Keith	$x_{ij} = 0$	1	0	1	0	1	0	-	-	0	0	1	0
		$x_{ij} = 1$	0	0	0	0	0	0	-	-	1	0	0	0
n_5	Ross	$x_{ij} = 0$	0	1	1	0	1	0	0	1	-	-	0	0
		$x_{ij} = 1$	0	0	0	0	0	0	0	0	-	-	1	0
n_6	Sarah	$x_{ij} = 0$	1	0	0	0	1	0	1	0	0	1	-	-
		$x_{ij} = 1$	0	0	0	1	0	0	0	0	0	0	-	-

Tablica 3.2: Skup $\mathbb{y}$ za učenike drugih razreda

Granične vrijednosti skupa $\mathbb{y}$ značajne su za procjenu parametara za različite modele. Te vrijednosti su sume elemenata od $\mathbb{y}$ i označene su podindeksom "+". Na primjer, y_{++k+} obilježava sumu svih vrijednosti $\mathbb{y}$ preko i, j, i l za svaki k. One tvore jednosmjernu tablicu s jednom ćelijom za svaki k. Granica $\{y_{++k+}\}$, daje broj veza za relaciju na različitim težinama $k = 0,1, \dots, C-1$. Tvore je članovi i, njihovi partneri j, i primljeni odabiri l.

Na primjeru mreže djece drugog razreda (skup $\mathbb{y}$ je prikazan u tablici (3.2)), granica $\{y_{++k+}\}$ je : $y_{++0+} = 22$ (zbroj jedinica za k=0), te $y_{++1+} = 8$ (zbroj jedinica za k=1). To nam govori da 22 veze imaju težinu 0, to jest da su 22 veze nepostojeće, i da ih 8 ima težinu k=1, odnosno da postoji 8 veza. Još jedan primjer je granica y_{i+k+}, koja daje broj postojećih veza (k=1) i nepostojećih veza (k=0) za svakog člana:

$$
\begin{array}{llll}
y_{1+0+} = 3 & y_{2+1+} = 2 & y_{4+0+} = 4 & y_{5+1+} = 1 \\
y_{1+1+} = 2 & y_{3+0+} = 4 & y_{4+1+} = 1 & y_{6+0+} = 4 \\
y_{2+0+} = 3 & y_{3+1+} = 1 & y_{5+0+} = 4 & y_{6+1+} = 1
\end{array}
$$

Na primjer, n_3, n_4, n_5 i n_6 imaju jednu vezu, tj. odabiru samo jedno dijete kao prijatelja.

3.2. Modeliranje skupa $\mathbb{Y}$

Prikazati ćemo statistički model za analizu jednostruke, direktne veze, čiji su podaci prikazani četverodimenzionalnim skupom $\mathbb{Y}$. Prije prikaza matematičkog modela, opisat ćemo model i njegovu korisnost. Za jednostruku, direktnu vezu usredotočit ćemo se na efekte koji prikazuju popularnost članova, njihovih partnera i odnos veza u dijadama.

Opis parametara modela

Osnovni model sastoji se od tri skupa parametara:

- skup parametara koji opisuje ponašanje člana pošiljatelja
- skup parametara koji prikazuje ponašanje člana primatelja
- skup parametara koji opisuje interakciju između parova članova u dijadi.

Prvi skup parametara se zove učinak širenja. U mreži dječjeg prijateljstva učinak širenja prikazuje sklonost da svako dijete odabere neko drugo kao prijatelja.

Drugi skup parametara je učinak popularnosti. U primjeru dječjeg prijateljstva, popularnost pokazuje tendenciju da dijete bude odabrano za prijatelja. Odabir je opisan širenjem i popularnošću svakog djeteta. Ovi se termini širenja i popularnosti mogu primijeniti jednako na neki drugi skup podataka, posebno kada su članovi ljudi i kada veze daju pozitivan efekt ili rast, dok se negdje drugdje ne mogu dobro primijeniti. Na primjer, ako mreža opisuje djecu koja uzimaju igračke od druge djece, kao najpopularnije dijete ne bi se moglo odabrati ono dijete čije se igračke najčešće uzimaju. Širenje i popularnost su parametri koji prikazuju sklonost članova da imaju veze sa drugim članovima. Pozitivne vrijednosti parametara povećavaju vjerojatnost da postoji veza.

Zadnji skup parametara su oni koji prikazuju povratnost ili uzajamnost između dva člana, nezavisno o širenju ili popularnošću bilo kojeg člana.

Parametri opisani ovdje nisu ograničeni na dihotomne podatke i vjerojatnosni su po svojoj prirodi. Nadalje, efekti povratnosti opisuju interaktivno ponašanje jedinstveno za dijade i to zajedno s vjerojatnosnim tendencijama za širenje i popularnost članova dijade. Povratnost je proširenje u kojem dijada prikazuje zajedničku, suprotnu ili asimetričnu vezu. Pozitivna povratnost parametara povećava vjerojatnost da je dijada zajednička.

Prikazani model jednostruke veze uključuje parametre koji mjere vjerojatnosnu tendenciju za sve nezavisne efekte: širenje, popularnost i povratnost.

Navedene parametre procjenjujemo log-linearnom tehnikom modeliranja. Log-linearni modeli su standardna statistička metoda za proučavanje diskretnih podataka. Velika većina podataka društvene mreže je diskretna i uglavnom je C malen. Društvena mreža koja nije diskretna može biti kategorizirana bez gubljenja bitnih podataka. Na primjer, možemo uzeti neprekidnu mjeru vremena provedenog pričajući i označiti je kao visoku, srednju ili slabu. Zato se usredotočujemo na veze sa diskretnim vrijednostima.

Osnovni model za dihotomnu povezanost

Proučavat ćemo modeliranje za dihotomne veze. Nakon prezentiranja modela za dihotomne relacije, proširit ćemo model na općenitiji slučaj diskretnih veza ($C > 2$).

Osnovni model, označen sa p_1, izražen je preko četiri faktora. Svaki od njih prezentira jedno od četiri moguća stanja za bilo koju dijadu:

- nul-dijadu

$$X_{ij} = X_{ji} = 0, \text{ ili } Y_{ij00} = 1,$$

- zajedničku dijadu

$$X_{ij} = X_{ji} = 1 \text{ ili } Y_{ij10} = 1$$

- dva slučaja asimetrične dijade

$$X_{ij} = 1, X_{ji} = 0 \text{ ili } Y_{ij10} = 1, \text{ i } X_{ij} = 0, X_{ji} = 1 \text{ ili } Y_{ij01} = 1.$$

Prirodni logaritam vjerojatnosti za svaki od četiri stanja dijada prikazan je kao funkcija od nekoliko parametara s ciljem kako bi se definirao p_1:

$$\log P(Y_{ij00} = 1) = \lambda_{ij},$$

$$\log P(Y_{ij10} = 1) = \lambda_{ij} + \theta + \alpha_i + \beta_j,$$

$$\log P(Y_{ij01} = 1) = \lambda_{ij} + \theta + \alpha_j + \beta_i,$$

$$\log P(Y_{ij11} = 1) = \lambda_{ij} + 2\theta + \alpha_i + \alpha_j + \beta_j + \beta_i + (\alpha\beta).$$

Navedeni model je log-linearan. Može biti analogan linearnom modelu koji proizlazi iz analize varijance. Log-linearni modeli u početku su rastući, logaritmiranjem varijable odaziva model postaje aditivan ili linearan u parametrima. Zbog toga je p_1 u početku vjerojatnost stanja dijade kao varijable odaziva, izjednačen s parametrom širenja (točnije, e na potenciju parametra širenja) pomnožen s parametrom popularnosti. Kada se model i vjerojatnosti odaziva transformiraju logaritmom, p_1 pokazuje parametar širenja dodan parametru popularnosti. Log-linearna forma modela je jednostavna za prilagodbu i razumijevanje. Logaritam vjerojatnosti u kojoj n_i ima veze iz i prema n_j postaje zbrajajuća funkcija koja uključuje širenje n_i i n_j, popularnost oba člana i veze između njih. Kada je p_1 pozitivan, on povećava logaritam vjerojatnosti da n_i ima vezu s n_j, a ukoliko je negativan, vjerojatnost se smanjuje.

Parametri λ_{ij} su matematička nužnost koja je uključena u model kako bi se osiguralo da se sve četiri vjerojatnosti zbroje za svaku dijadu. Zbog toga se ti parametri pojavljuju u sve četiri činjenice, bez obzira na stanje dijade.

Parametar θ se interpretira kao sveukupan izbor (analogan ukupnoj aritmetičkoj sredini), kako bi prikazali sve poslane i zaprimljene izbore. Ako je jedna veza prisutna u dijadi, pojavljuje se jedan θ, a kada je veza uzajamna, pojavljuju se dvije. Parametar θ se ne pojavljuje u dijelu modela kada ne postoji veza i $(\alpha\beta)$ je prisutno samo kada je dijada zajednička.

Zavisni parametri se ne pojavljuju u prvoj činjenici modela koja predstavlja nul-dijadu. Za asimetrične dijade, logaritam vjerojatnosti ovisi o parametrima koji prikazuju samo jednu od dvije moguće veze u dijadi: dijade u kojoj član i odabire člana j bez povratnosti (pa je α_i bitan, ali ne i α_j te je β_j uključen, dok β_i nije uključen) i dijade u kojima član j odabire člana i bez povratnosti (pa su bitni parametri α_j i β_i, ali ne α_i ili β_j ili $(\alpha\beta)$). Svi parametri se pojavljuju zajedno samo za zajedničke dijade.

Parametar $(\alpha\beta)$ se zove parametar zajedništva. Kada su odabiri veza u nekoj mreži uzajamni, kao što je na primjer odabir prijatelja, ti parametri će biti pozitivni i veliki. Tada taj parametar predstavlja mjeru uzajamnosti između veza koje su poslane i primljene (kao što je to koeficijent korelacije za neprekidne podatke). Za neke odnose kao na primjer prijateljstvo, pretpostavili bi da vrijedi uzajamnost. Ali ne bi očekivali uzajamnost za ostale odnose, kao što su davanje radnih zadataka ili savjetovanje. Iako bi nadređeni mogao pitati podređenog za savjet, pretpostavljamo da se to događa rjeđe nego slučaj kada podređeni pita nadređenog. Kada povratnost nije bitna značajka mreže, parametar povratnosti bi bio nula. Možemo promatrati $(\alpha\beta)$ kao mjeru povratnosti baziranu na modelu.

U ovom modelu ograničenja su nužna zbog procjene parametara. Koristimo standardnu analizu varijance kao ograničenja u kojima parametri i njihove procjene u zbroju daju nulu po njihovim podskupovima, dakle vrijedi $\sum_i \alpha_i$=0, i $\sum_j \beta_j$=0. Ta ograničenja određuju stupnjeve slobode (df) povezane sa svakim skupom parametara. Broj stupnjeva slobode određen sa bilo kojim skupom parametara je broj parametara koji su nezavisni ili slobodno variraju. Parametri širenja α_i imaju indeks i, koji ide od 1 do g, što označava broj članova. Imamo g α_i parametara, ali su takvi da u sumi daju nula. Broj stupnjeva slobode za ovaj skup parametara iznosi $g-1$ zato što možemo izračunati α_g iz ostalih $g-1$ parametra. Također, parametri popularnosti β_i zahtijevaju $g-1$ stupnjeva slobode. I zadnje, parametri povratnosti $(\alpha\beta)$ zahtijevaju jedan stupanj slobode.

U nastavku ćemo promatrati općenitiju formu p_1 modela, koja nam dozvoljava da proučavamo varijable sa jednom vezom koje su diskretne i ne nužno dihotomne. Pretpostavljamo da varijabla relacije može poprimiti vrijednosti $0,1,\dots,C-1$, i da je $Y_{ijkl}=1$, kada je $X_{ij}=k$ i $X_{ji}=l$.

Sljedeća stavka daje generalizaciju četiri faktora modela p_1:

$$\log P\left(Y_{ijkl}=1\right)=\lambda_j+\theta_k+\theta_l+\alpha_{i(k)}+\alpha_{j(l)}+\beta_{j(k)}+\beta_{i(l)}+(\alpha\beta)_{kl}\,. \quad (3.3)$$

Parametar $\alpha_{i(k)}$ mjeri tendenciju da član i šalje vezu s težinom k, dok $\beta_{j(l)}$ mjeri tendenciju da član j prima vezu s težinom l.

Parametri imaju sljedeća ograničenja:

$$\theta_0=0,$$

$$\alpha_{i(0)}=0, za\ sve\ i,$$

$$\sum_i \alpha_{i(k)}=0, za\ sve\ k,$$

$$\beta_{j(0)}=0, za\ sve\ j,$$

$$\sum_j \beta_{j(l)}=0, za\ sve\ l,$$

$$(\alpha\beta)_{k0}=0, za\ sve\ k,$$

$$(\alpha\beta)_{0l}=0, za\ sve\ l,$$

$$(\alpha\beta)_{kl}=(\alpha\beta)_{lk}.$$

Parametre ograničavamo, odnosno dajemo im vrijednost 0 kada je izbor s najmanjom težinom ($k=0$ ili $l=0$). Ova generalizacija p_1 modela zbog toga postaje ekvivalentna modelu p_1 kada je $C=2$. U modelu p_1, α_i je definirana samo kada je $k=1$, odnosno kada je izbor poslan. Parametar $\alpha_{i(k)}=0$ kada je $k=0$, ali $\alpha_{i(k)}$ može biti različit od nule, što najčešće i jest, kada postoji izbor bilo koje težine ($k=1,,2,\dots,C-1$).

Za svaki $k > 0$, $\alpha_{i(k)}$ ima sumu nula za bilo koje članove. Zato što procjena u sumi daje nula za svakog člana, relativne usporedbe (za svaku težinu) za članove se lako rade.

Prikazana ograničenja su konzistentna s činjenicom da parametri $\theta_{(k)}$, $\alpha_{i(k)}$, $\beta_{j(l)}$ i $(\alpha\beta)_{kl}$ zahtijevaju redom $(C-1)$, $(g-1)(C-1)$, $(g-1)(C-1)$ i $C(C-1)/2$ stupnjeva slobode.

Ograničenja za parametre α su dana tablicom (3.3).

	$k=0$	$k=1$	$\cdots$	$k=C-1$
$i=1$	0	$\alpha_{1(1)}$	$\cdots$	$\alpha_{1(C-1)}$
$i=2$	0	$\alpha_{2(1)}$	$\cdots$	$\alpha_{2(C-1)}$
$i=3$	0	$\alpha_{3(1)}$	$\cdots$	$\alpha_{3(C-1)}$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
$i=g$	0	$\alpha_{g(1)}$	$\cdots$	$\alpha_{g(C-1)}$
Ukupno	0	0	$\cdots$	0

Tablica 3.3: Ograničenja za parametre $\alpha_{i(k)}$

Kao i prije parametri $\alpha_{i(k)}$ su parametri širenja, a $\beta_{j(l)}$ parametri popularnosti. Parametri $\alpha_{i(k)}$ predstavljaju sklonost da član i šalje vezu sa težinom k. Isto tako, $\beta_{j(l)}$ predstavlja sklonost da član j prima vezu sa težinom l. Širenje članova i se prikazuje kao $\alpha_{i(k)}$, a njihova popularnost $\beta_{j(l)}$. Dok se širenje članova j prikazuje kao $\alpha_{j(l)}$, a njihova popularnost kao $\beta_{j(k)}$.

Parametri $(\alpha\beta)_{kl}$ su parametri povratnosti. Ti parametri ne ovise o točno jednom članu (nema indeksa i i j). Model pretpostavlja da su ti parametri konstanta za sve parove članova. Parametri povratnosti su simetrični sa svojim indeksima, $(\alpha\beta)_{kl} = (\alpha\beta)_{lk}$, pa imamo $C(C-1)/2$ stupnjeva slobode.

Model za dihotomne podatke ima samo jedan parametar povratnosti (zato što je $C(C-1)/2 = 2(2-1)/2 = 1$), kao što je određeno modelom p_1.

Parametar $(\alpha\beta)$ za modeliranje dihotomne veze je analogan mjeri asocijacije. Za slučaj kada je $C > 2$, $C \times C$ matrica parametara $(\alpha\beta)_{kl}$ je analogna cijeloj matrici takvih mjera.

Na primjer, $(\alpha\beta)_{15} = (\alpha\beta)_{51}$ mjeri pozitivnu ili negativnu asocijaciju između veza poslanih sa slabom težinom od $k = 1$ i veze primljene s mnogo većom težinom $l = 5$.

Bitno je naglasiti da kada dajemo vrijednosti vezama, procjene α i β ovise o broju mogućih vrijednosti. Za svaki nivo $k = 1, \dots, C-1$, imamo g α i g β. Za fiksan k ti parametri mjere koliko je vjerojatno da član ima vezu (poslanu ili primljenu) baš te težine.

Primjer - prilagodba p_1 za neku mrežu

Prilagodit ćemo model mreži učenika drugog razreda i proučavat ćemo njihove parametre. Procjene parametara koje smo dobili prilagodbom modela prikazane su tablicom (3.4).

	Član	$\hat{\alpha}_i$	$\hat{\beta}_i$
n_1	Allison	1.414	$-\infty$
n_2	Drew	0.817	0.867
n_3	Eliot	-0.474	-1.223
n_4	Keith	0.197	$-\infty$
n_5	Ross	-0.977	0.178
n_6	Sarah	-0.977	0.178

$$\widehat{(\alpha\beta)} = 3.077$$

$$\hat{\theta} = -1.437$$

Tablica 3.4: p_1 procjena parametara za učenike drugog razreda

Primijetimo da su ti parametri dani logaritamskom skalom. Zbog toga, ako povećamo α za jednu jedinicu, recimo sa 1 na 2, logaritam vjerojatnosti izbora se poveća za jednu jedinicu. Odnosno, stvarna vjerojatnost se poveća za exp(1)=2.718.

Parametar $\hat{\alpha}$ nam govori sljedeće: član 1 (Allison) ima najveći parametar širenja, to jest najveća je vjerojatnost da će na početku školske godine imati prijatelje s obzirom na bilo koje drugo dijete. Slijedi je član 2 (Drew), dok članovi 5 i 6 (Ross i Sarah) imaju najmanju vjerojatnost.

Parametar $\hat{\beta}$ pokazuje sklonost da je svako dijete odabrano kao prijatelj od nekog drugog djeteta unutar mreže. Te procjene su prilično različite od procjene parametara α, točnije dva procijenjena parametra su beskonačna. Primijetimo da Allison i Keith nisu odabrani za prijatelje od strane druge djece, dakle imaju 0 bridova koji ulaze u vrh usmjerenog grafa. To uzrokuje da parametri β budu $-\infty$ za to dvoje djece. Najveća vjerojatnost da će odabrati drugu djecu bila je za člana 1 (Allison), dok je najmanja vjerojatnost bila da će biti odabrana od ostalih, s obzirom na to da je njezin $\hat{\beta}$ najmanji (to jest njezin $\hat{\beta}$ je $-\infty$). Slično vrijedi i za člana 4 (Keith). Ostala djeca (n_2, n_5, n_6 i n_3) mogu biti odabrana za prijatelja, dok Drew, Ross i Sarah imaju pozitivne sklonosti.

Parametar povratnosti daje dodatne informacije o vezi. Za dihotomne podatke, analogija između parametara $(\alpha\beta)$ i mjere asocijacije je lako uočljiva. Procijenjena vrijednost je pozitivna i velika, pa time indicira sklonosti za pozitivnu asocijaciju ili međusobno prijateljstvo. Na primjer, ako dijete i nominira dijete j kao prijatelja, dijete j sklono je uzvratiti prijateljstvo. Slično, ako n_i ne nominira n_j, n_j također ima sklonost ne nominirati n_i. Ako bi procjene parametara bile negativne, to bi značilo da postoji mnogo neuzvraćenih prijateljstava, to jest asimetričnih dijada.

3.3. Procjena parametara- teorija i praksa

Distribucije i metoda maksimalne vjerodostojnosti

Model (3.3) je dio familije eksponencijalne distribucije, što znači da parametri mogu biti procijenjeni metodom maksimalne vjerodostojnosti.

Za model p_1 cilj metode maksimalne vjerodostojnosti je naći najbolju procjenu svih parametara u modelu (λ, θ, α, β i $(\alpha\beta)$) koji su mogli dati dane dijadičke interaktivne podatke prikazane u $\mathbb{y}$ skupu. Vjerojatnosna funkcija je zajednička vjerojatnosna distribucija podataka. Metoda najveće vjerodostojnosti teži pronalasku parametara koji maksimiziraju funkciju vjerodostojnosti. Log-vjerodostojnost nam govori koje granice skupa $\mathbb{y}$ su potrebne da bi procijenili parametre. Ograničenja moraju biti jasno naznačena.

Log-vjerodostojnost za model (3.3) pretpostavlja nezavisnost dijada kao što slijedi:

$$\sum_{i<j} \lambda_{ij} + \frac{1}{2}\sum_k \theta_k y_{++k+} + \frac{1}{2}\sum_l \theta_l y_{+++l} + \frac{1}{2}\sum_i \sum_k \alpha_{i(k)} y_{i+k+} + \frac{1}{2}\sum_j \sum_l \alpha_{j(l)} y_{+j+l} +$$
$$+ \frac{1}{2}\sum_j \sum_k \beta_{j(k)} y_{+jk+} + \frac{1}{2}\sum_i \sum_l \beta_{i(l)} y_{i++l} + \sum_k \sum_l (\alpha\beta)_{kl} \sum_{i<j} y_{ijkl}. \tag{3.4}$$

Izrazi u log-vjerodostojnosti koji ovise o podacima $\mathbb{y}$ sadrže indeks "$+$" .

Granice od $\mathbb{y}$ koje su potrebne da bi se procijenili parametri su one koje proizlaze iz log-vjerodostojnosti (3.4). Te granice su statistički dovoljne za parametre modela. To su jedini potrebni podaci da bi se odredila maksimalna vjerojatnost i odredila procjena parametra.

Procjena metodom maksimalne vjerodostojnosti ne daje samo procjenu parametara modela (kao p_1), nego i procjenu slučajnih varijabli modela. U ovom slučaju, modeliramo elemente skupa $\mathbb{Y}$ i označiti ćemo procjenitelja metodom maksimalne vjerodostojnosti sa $\hat{Y}_{ijkl}$.

Ti procjenitelji, koji se često nazivaju *prilagođene veličine*, uobičajeno se ne mogu jednostavno izračunati. Računaju se iterativno pomoću algoritma odabranog za rješavanje vjerojatnosnih jednadžbi koje proizlaze iz problema metode maksimalne vjerodostojnosti.

Statistička ocjena prilagodbe

Kada smo prilagodili model (3.3) podacima i dobili procjenitelje metodom maksimalne vjerodostojnosti, trebali bi moći ocijeniti koliko dobro je model prilagođen i koji parametri su statistički značajni (odnosno koji se statistički razlikuju od 0).

U procjeni metodom maksimalne vjerodostojnosti za ove potrebe koristi se statistika omjera vjerodostojnosti.

Za log-linearne modele koji su ovdje opisani statistika je G^2, koja za ovaj slučaj glasi:

$$G^2 = 2\sum_{i<j}\sum_{k,l} y_{ijkl}\log\left(y_{ijkl}/\hat{y}_{ijkl}\right). \tag{3.5}$$

Statistika omjera vjerodostojnosti je funkcija promatranih podataka y_{ijkl} i izračunatih predviđanja modela (prilagođenih vrijednosti) $\hat{y}_{ijkl}$ koje su predviđanja za podatke mreže dobivene iz modela p_1.

Procjene parametara se također mogu uključiti kao što je dano modelom; kako bi se dobile prilagođene vrijednosti ($\hat{y}$). Statistika je izračunata kao zbroj svih ćelija u kontingencijskoj tablici koja uspoređuje promatrane i prilagođene vrijednosti.

Prilagođene vrijednosti $\hat{y}_{ijkl}$ za odnos prijateljstva hipotetske mreže učenika drugog razreda prikazane su tablicom (3.5).

			j											
			Allison		Drew		Eliot		Keith		Ross		Sarah	
			$l = x_{ji}$											
i		k	0	1	0	1	0	1	0	1	0	1	0	1
n_1	Allison	$x_{ij} = 0$	-	-	0.30	0.00	0.78	0.00	1.00	0.00	0.46	0.00	0.46	0.00
		$x_{ij} = 1$	-	-	0.70	0.00	0.22	0.00	0.00	0.00	0.54	0.00	0.54	0.00
n_2	Drew	$x_{ij} = 0$	0.30	0.70	-	-	0.37	0.13	0.59	0.41	0.21	0.04	0.21	0.04
		$x_{ij} = 1$	0.00	0.00	-	-	0.06	0.45	0.00	0.00	0.13	0.62	0.13	0.62
n_3	Eliot	$x_{ij} = 0$	0.78	0.22	0.37	0.06	-	-	0.92	0.08	0.77	0.02	0.77	0.02
		$x_{ij} = 1$	0.00	0.00	0.13	0.45	-	-	0.00	0.00	0.14	0.07	0.14	0.08
n_4	Keith	$x_{ij} = 0$	1.00	0.00	0.59	0.00	0.92	0.00	-	-	0.74	0.00	0.74	0.00
		$x_{ij} = 1$	0.00	0.00	0.41	0.00	0.08	0.00	-	-	0.26	0.00	0.26	0.00
n_5	Ross	$x_{ij} = 0$	0.46	0.54	0.21	0.13	0.77	0.14	0.74	0.26	-	-	0.68	0.07
		$x_{ij} = 1$	0.00	0.00	0.04	0.62	0.02	0.08	0.00	0.00	-	-	0.07	0.17
n_6	Sarah	$x_{ij} = 0$	0.46	0.54	0.21	0.13	0.77	0.14	0.74	0.26	0.68	0.07	-	-
		$x_{ij} = 1$	0.00	0.00	0.04	0.62	0.02	0.08	0.00	0.00	0.07	0.17	-	-

Tablica 3.5: Prilagođene vrijednosti od y modela p_1 za učenike drugog razreda

Statistička ocjena prilagodbe za opisani statistički model za model podataka društvene mreže (u (3.5)) ima sličnu interpretaciju. Jedina razlika je u tome da suma nije uzeta preko svih ćelija $\mathbb{y}$ skupa, nego preko svih dijada $(i < j)$, s obzirom na to da je osnovna jedinica u ovim modelima dijada. Želimo samo jednom usporediti svaku dijadu s njegovom prilagođenom vrijednosti, pa sumiramo samo pola skupa $\mathbb{y}$ zbog njegove simetričnosti.

Statistika G^2 definirana jednadžbom (3.5) ima sljedeću karakteristiku: kako dodajemo parametre modelu i činimo ga kompliciranijim, ona ostaje jednaka ili se smanjuje, pokazujući bolju prilagođenost, ili veću zavisnost $\mathbb{y}$ i $\hat{y}$. Ova karakteristika je posebno bitna kada uspoređujemo dva hijerarhijski povezana modela. Primjerice, model s parametrima λ, θ, α i β je hijerarhijski povezan s modelom koji ima parametre λ, θ, α, β i $(\alpha\beta)$.

Modeli su identični osim što jedan model ima jedan skup parametara više $(\alpha\beta)_{kl}$. Širi model ima sve parametre prvog modela i dodatni skup parametara. Modeli koje ovdje razmatramo će biti hijerarhijski - ako jedan model može biti izveden iz drugoga, tako da neke parametre izjednačimo s nulom, onda možemo reći da je prvi model hijerarhijski povezan s drugim modelom.

Općenito, možemo testirati hipoteze parametara modela uspoređujući testne statistike hijerarhijski povezanih modela. Razlika između G^2 - statistike dva hijerarhijski povezana modela je približno asimptotski distribuirana kao slučajna varijabla χ^2. Navedena razlika može biti uspoređena preko tabličnih vrijednosti χ^2 (sa određenim stupnjevima slobode) da bi ustanovili da li je model s više parametara značajniji nego jednostavniji model. Ako je model značajniji, onda su dodatni parametri statistički značajni. Možemo usporediti prilagođenost modela (3.3) modelu koji je sličan, ali koji ne sadrži parametar α. Razlika između dvije G^2 statistike bila bi testirana na $(g-1)(C-1)$ stupnjeva slobode kako bi se odredila statistička značajnost parametra α_{ik} .

Treba napomenuti da G^2 i stupnjevi slobodne dobiveni iz standardnih paketa nisu točni i potrebne su im modifikacije. Prilagođene vrijednosti i reziduali također mogu poslužiti dijagnostički da bi se odredilo da li su ćelije posebno dobro ili posebno loše prilagođene. Kada model ne odgovara podacima, jedan način usklađivanja je dodavanje parametara modelu.

3.4. Praktični vodič za prilagodbu modela podacima

Jednom kada podaci tvore skup $\mathbb{y}$, model (3.3) može biti prilagođen (i parametri procijenjeni) koristeći log-linearne procedure modeliranja koje se nalaze u svim važnim statističkim računalnim programima.

Prilagodba modela

Svaki skup parametara povezan je s odgovarajućim statistikama, koje su granice skupa podataka koji se modelira. Jednadžba metode maksimalne vjerodostojnosti koje je potrebno riješiti da bi se dobio procjenitelj metodom maksimalne vjerodostojnosti za očekivane vrijednosti ćelija (prilagođene vrijednosti) imaju isti oblik: promatrane granice su jednake prilagođenim granicama. Dakle odgovarajući parametri modela biti će procijenjeni, a prilagođene vrijednosti dobivene onda kada ograničimo prilagođene granice tako da budu jednake promatranim granicama. Na primjer, ako 40 muškaraca i 60 žena čini uzorak za istraživanje, onda svako modeliranje koje sadrži spol kao varijablu mora dati rezultat varijable 40 i 60. Ili promatrajući podatke mreže, ako u model uključimo parametar β i ako je član 1 odabran od strane četvero članova od ukupno osam članova mreže, onda prilagođena vjerojatnost da je n_1 odabran mora iznositi $(4/8) = 0.50$. Ovaj izračun je ključan za dobivanje procjenitelja metodom maksimalne vjerodostojnosti očekivanih vrijednosti ćelija i parametara.

Modeli za jednostruku mrežu podataka mogu biti prilagođeni navedenoj teoriji; točnije, naglasak je na parametre modela i odgovarajuće statistike koje su granice skupa $\mathbb{Y}$. Svih šest dvodimenzionalnih granica četverodimenzionalnog skupa $\mathbb{Y}$ su odgovarajuće statistike za parametre u modelu (3.3). Isto vrijedi za model p_1, s obzirom na to da je ekvivalentan osnovnom modelu kada je $C = 2$.

Četiri varijable koje definiraju $\mathbb{Y}$ sadržavaju četiri dimenzije kontingencijske tablice koja se modelira. Na te varijable i njihove indekse smo se pozivali na sljedeći način:

- član pošiljatelj (i)
- član primatelj (j)
- težina poslanog izbora $(k = X_{ij})$
- težina primljenog izbora $(l = X_{ji})$

Te varijable označavamo 1,2,3,4; svaku zasebno.

Da bi prilagodili model (3.3), prilagođavamo sljedeći log-linearni model skupu $\mathbb{Y}$:

$$[12]\,[13]\,[24]\,[23]\,[14]\,[34] \tag{3.6}$$

Zagrade u relaciji (3.6) su granice skupa $\mathbb{Y}$, koje su odgovarajuće statistike za parametre osnovnog modela. Ovakva notacija je česta za log-linearne modele i naziva se *Fienbergova* notacija.

Skup varijabli u zagradama implicira da smo uključili parametar koji je povezan s odgovarajućim statistikama dobivenim preko varijabli modela. S obzirom na to da su log-linearni modeli ovdje promatrani hijerarhijski, svi uvjeti nižeg reda (primjerice glavni utjecaj) se također trebaju prilagoditi. Primjerice, $[12]$ je veza između varijabli 1 i 2 (indeks člana pošiljatelja i člana primatelja) i odgovara granici $\{Y_{ij++}\}$. Ova granica uključuje jednodimenzionalne granice za ove dvije varijable ($\{Y_{i+++}\}$ i $\{Y_{+j++}\}$).

Granica $[12]$ je prilagođena kako bi uključila parametre λ_{ij} u model. S obzirom da parametar λ mora biti uključen u sve modele kako bi se pravilno ograničila vjerojatnost, onda granica $[12]$ uvijek mora biti uključena. Granice $[13]$ i $[24]$ su odgovarajuće statistike za $\alpha_{i(k)}$ parametre. Postoji samo jedan skup od $(g-1)(C-1)$ parametara širenja $\alpha_{i(k)}$, ali dva skupa granica, $[13]$ i $[24]$. Te granice su zapravo jednake zbog simetričnosti skupa $\mathbb{Y}$ zbog čega ne postoje dodatni parametri. Obje granice $[13]$ i $[24]$ moraju biti uključene u svaki model za koji želimo procijeniti α. Slično, granice $[23]$ i $[14]$ su odgovarajuće statistike za parametre $\beta_{j(l)}$. Te granice su zapravo jednake zbog simetričnosti skupa $\mathbb{Y}$. Obje granice moraju biti uključene u svaki model za koji želimo procijeniti β. Posljednja granica $[34]$ je odgovarajuća statistika za povratne parametre $(\alpha\beta)_{kl}$. Jedini preostali parametri dani u modelu (3.3) su elementi skupova $\{\theta\}$. Odgovarajuće statistike za ove parametre su granice $[3]$ i $[4]$, ali s obzirom da je model hijerarhijski onda su one automatski prilagođene kad god su granice višeg reda $[13]$ i $[24]$, ili $[23]$ i $[14]$, ili $[34]$ prilagođene.

3.5. Beskonačni parametri

Problem kod ovog modeliranja je kad član ima nula ili $(g-1)$ bridova koji ulaze ili izlaze iz vrha usmjerenog grafa. U prvom slučaju član ne šalje ni ne prima nikakve veze, a u drugom ih ili prima od svih ili ih šalje svima. Kada veza ima vrijednost, onda se problem pojavljuje kada sve veze poslane ili primljene od strane određenog člana imaju jednaku težinu. Usredotočimo se na dihotomne veze.

Pretpostavimo da prvi član ima 0 bridova koji ulaze u vrh (ovaj član nema sklonost primanja veze). Dakle, sve prilagođene vjerojatnosti da član prima veze moraju biti jednake 0. Kada član ima 0 bridova koji izlaze iz vrha (član nema sklonost slanja veza) tada sve prilagođene vjerojatnosti da član šalje veze moraju biti jednake 0. Ako je prilagođena vjerojatnost jednaka 0, onda je logaritam ove vjerojatnosti $-\infty$. Da bi taj logaritam bio jednak $-\infty$, i odgovarajući parametar mora biti $-\infty$. Suma nultog reda u socio-matrici dovodi do toga da pripadajući parametar α ovog člana iznosi $-\infty$. Suma nultog stupca u socio-matrici dovodi do toga da pripadajući parametar β ovog člana iznosi $-\infty$. Na primjer, dvoje djece u mreži učenika drugog razreda nije odabrano za prijatelja. Dakle, dvije (β) prilagođene modelu p_1 iznose $-\infty$.

Pretpostavimo da član ima $(g-1)$ bridova koji ulaze u vrh (član prima veze od svih drugih članova). Dakle, sve prilagođene vjerojatnosti da ovaj član prima bilo koju vezu moraju iznositi 1. Kada član ima $(g-1)$ bridova koji izlaze iz vrha (član šalje veze prema svim drugim članovima) tada sve prilagođene vjerojatnosti da član šalje veze moraju biti jednake 1. Ako je prilagođena vjerojatnost jednaka 1, onda je logaritam ove vjerojatnosti ∞. Da bi taj logaritam bio beskonačan, onda i odgovarajući parametar također mora biti ∞. Suma reda socio-matrice jednaka $(g-1)$ dovodi do toga da pripadajući parametar α za tog člana iznosi ∞. Suma stupca socio-matrice jednaka $(g-1)$ dovodi do toga da pripadajući parametar β tog člana iznosi ∞.

Ovi beskonačni parametri nisu pribrojani kod određivanja df. Za učenike drugog razreda, imamo potpuni skup α i $df = 5$, ali procjenjujemo samo 4 β. Dakle, $df = 3$ za β. Nadalje, beskonačni parametri nisu uzeti u obzir kod postavljanja konačnih parametara na sumu 0. Primijetimo još jednom kod učenika drugog razreda da zbroj 4 procijenjene β iznosi 0.

3.6. Usporedba modela i statistički test parametara

Prikazali smo model (3.3) sa tri skupa značajnih parametara - pokazujući utjecaj širenja, popularnosti i povratnosti. Možemo provesti test hipoteze da bi otkrili pokazuju li članovi sve ili samo neke od navedenih utjecaja u odnosu. Možemo pretpostaviti i posljedično prilagoditi alternativne jednostavnije modele koji bi mogli prikazati podatke u mreži gotovo jednako dobro ili jednako dobro kao i kompliciraniji model (3.3).

Uspoređujući prilagođenost modela (3.3) s jednostavnijom verzijom modela možemo odrediti je li svaki skup parametara statistički različit od nule.

Koristimo testiranje hipoteza omjera vjerodostojnosti kako bi proučili svaki skup parametara. Testna statistika ovih hipoteza uspoređuje procjenu prilagodbe modela (3.3) s procjenom prilagodbe jednostavnijeg modela. Te sve statistike su statistike omjera vjerodostojnosti, gdje su prilagođene vrijednosti $\hat{y}_{ijkl}$ iz različitih usporednih modela.

U nastavku je navedeno sedam modela koji su jednostavniji od modela (3.3), a izvedeni su ispuštanjem iz modela jednog ili više skupova parametara $\alpha_{i(k)}$, $\beta_{j(l)}$ ili $(\alpha\beta)_{kl}$. Ako model sadrži manje parametara dati će lošije rezultate od modela (3.3).

Model (3.3) je ponovno napisan, a onda slijedi i sedam jednostavnijih modela:

$$\log P(Y_{ijkl} = 1) = \lambda_{ij} + \theta_k + \theta_l + \alpha_{i(k)} + \alpha_{j(l)} + \beta_{j(k)} + \beta_{i(l)} + (\alpha\beta)_{kl} \quad (3.3)$$

$$\log P(Y_{ijkl} = 1) = \lambda_{ij} + \theta_k + \theta_l + \beta_{j(k)} + \beta_{i(l)} + (\alpha\beta)_{kl}$$

$$\log P(Y_{ijkl} = 1) = \lambda_{ij} + \theta_k + \theta_l + \alpha_{i(k)} + \alpha_{j(l)} + (\alpha\beta)_{kl}$$

$$\log P(Y_{ijkl} = 1) = \lambda_{ij} + \theta_k + \theta_l + \alpha_{i(k)} + \alpha_{j(l)} + \beta_{j(k)} + \beta_{i(l)}$$

$$\log P(Y_{ijkl} = 1) = \lambda_{ij} + \theta_k + \theta_l + \alpha_{i(k)} + \alpha_{j(l)}$$

$$\log P(Y_{ijkl} = 1) = \lambda_{ij} + \theta_k + \theta_l + \beta_{j(k)} + \beta_{i(l)}$$

$$\log P(Y_{ijkl} = 1) = \lambda_{ij} + \theta_k + \theta_l + (\alpha\beta)_{kl}$$

$$\log P(Y_{ijkl} = 1) = \lambda_{ij} + \theta_k + \theta_l \quad (3.7)$$

U tablici (3.6) naveli smo te modele, (3.3) i (3.7a)-(3.7g), zajedno s listom granica koje određuju svaki model. Prisjetimo se iz definicije modela p_1 da ako je relacija koju promatramo dihotomna, onda se parametri u modelu pojavljuju samo ako su napravljeni izbori (primjerice $\alpha_i(0) \equiv 0$ i $\alpha_i \equiv \alpha_i$).

Model	Parametri uključeni u model	Granice
3.3	$\{\lambda_{ij}\}$, $\{\theta_k\}$, $\{\alpha_{i(k)}\}$,$\{\beta_{j(l)}\}$,$\{(\alpha\beta)_{kl}\}$	[12] [13] [24] [23] [14] [34]
3.7a	$\{\lambda_{ij}\}$, $\{\theta_k\}$, $\{\beta_{j(l)}\}$, $\{(\alpha\beta)_{kl}\}$	[12] [23] [14] [34]
3.7b	$\{\lambda_{ij}\}$, $\{\theta_k\}$, $\{\alpha_{i(k)}\}$, $\{(\alpha\beta)_{kl}\}$	[12] [13] [24] [34]
3.7c	$\{\lambda_{ij}\}$, $\{\theta_k\}$, $\{\alpha_{i(k)}\}$, $\{\beta_{j(l)}\}$	[12] [13] [24] [23] [14]
3.7d	$\{\lambda_{ij}\}$, $\{\theta_k\}$, $\{\alpha_{i(k)}\}$	[12] [13] [24]
3.7e	$\{\lambda_{ij}\}$, $\{\theta_k\}$, $\{\beta_{j(l)}\}$	[12] [23][14]
3.7f	$\{\lambda_{ij}\}$, $\{\theta_k\}$, $\{(\alpha\beta)_{kl}\}$	[12] [34]
3.7g	$\{\lambda_{ij}\}$, $\{\theta_k\}$	[12] [3] [4]

Tablica 3.6: Parametri modela p_1, modeli i pripadne granice

U modelu (3.7b) nema β. Ovaj model pretpostavlja da su sve β jednake nula što znači da nema utjecaja popularnosti između članova. Prema definiciji, očekujemo da ovaj model dobro odgovara jednostrukoj društvenoj mreži podataka samo onda kada je popularnost članova konstantna. Ako bi bilo točno da su sve β jednake nuli, onda bi očekivali da se model (3.7b) prilagođava promatranim dijadičkim odnosima jednako dobro kao i potpuni model (3.3).

Da bi testirali nultu hipotezu $H_0\colon \beta_{j(l)} = 0,\ za\ sve\ j$ i l, uspoređujemo procjenu prilagodbe za dva modela (3.7b) i (3.3). Slično, usporedba prilagođenih statistika za model (3.7a) ili model (3.7c) s modelom (3.3) bi testirala nultu hipotezu $H_0 : \alpha_{i(k)} = 0$ $za\ sve\ i$ i k, ili $H_0\colon (\alpha\beta)_{kl} = 0\ za\ sve\ k$ i l. Model (3.7b) je hijerarhijski usklađen u model (3.3), pa je prvi model posebni slučaj drugog modela. Dakle, prilagođena statistika za (3.7b) će biti veća ili jednaka nego prilagođena statistika za (3.3). Usporedba G^2 - statistike pokazuje da li su parametri statistički različiti od nule. Razlika između G^2 je testna statistika uvjetne vjerojatnosti, koja testira nultu hipotezu, odnosno jesu li parametri zaista jednaki nuli. Statistički postavljamo uvjete na kompliciranije modele i testiramo mogu li se pojednostaviti ispuštanjem uvjeta.

U praksi, da bi testirali uvjetnu hipotezu glede skupova parametara, uzimamo G^2 za model (3.3) i oduzimamo ga od G^2 za model koji ne sadrži testirane parametre. Primjerice, da bi ispitali da li su β nula, računamo dvije vrijednosti G^2, jednu za cijeli model, a drugu za (3.7b) .Dobivamo razliku rezultata G^2 (3.7b) manje G^2 (3.3), to je ΔG^2 koji određuje da li su dodatni parametri značajno različiti od nule. Ta nova statistika ΔG^2 se otprilike jednako asimptotski distribuira jednako kao slučajna varijabla χ^2 s određenim brojem stupnjeva slobode. Kao što je statistika ΔG^2 jednaka razlici G^2 dvaju modela, razlika stupnjeva slobode Δdf je jednaka razlici stupnjeva slobode za ta dva modela. Model koji sadrži više parametara je model alternative hipoteze (H_A) i model s manje uvjeta je model nulte hipoteze (H_0). Δdf je jednak broju neovisnih parametara koji se testiraju.

Primjerice, kod vrednovanja nulte hipoteze, $H_0 : (\alpha\beta)_{kl} = 0$ (za svaki k, l); odbacujemo H_0 ako je ΔG^2 veći od odgovarajuće tablične vrijednosti χ^2 distribucije sa $C(C-1)/2$ stupnjeva slobode.

Ako je ΔG^2 statistički velik u usporedbi sa χ^2 distribucijom sa $C(C-1)/2$ stupnjeva slobode, zaključiti ćemo da su $(\alpha\beta)_{kl}$ parametri statistički značajni i trebali bi biti uključeni u bilo koji model koji realno opisuje danu mrežu.

Treba napomenuti da je "asimptotska" distribucija G^2 ili ΔG^2 samo aproksimacija od χ^2.

Testovi koji mogu biti napravljeni za parametre kojima se modelira jednostruka veza su navedeni u tablici (3.7).

Prvi navedeni test uspoređuje prilagođenost modela (3.7a) s prilagođenošću modela (3.3). Model (3.7a) nema α, pa provodeći ovu usporedbu testiramo nultu hipotezu $H_0 : (\alpha\beta)_{kl} = 0$ (za svaki k, l). Broj stupnjeva slobode za ovaj test je $(g-1)(C-1)$. Ako ΔG^2 nije velik, onda ne možemo odbaciti nultu hipotezu, što znači da su članovi statistički jednaki s obzirom na njihovu proširenost.

Test	ΔG^2	Δdf	Nulta hipoteza
1	3.7a – 3.3	$(g-1)(C-1)$	$H_0 : \alpha_{i(k)} = 0$, za sve i i k
2	3.7b – 3.3	$(g-1)(C-1)$	$H_0: \beta_{j(l)} = 0$, za sve j i l
3	3.7c – 3.3	$C(C-1)/2$	$H_0 : (\alpha\beta)_{kl} = 0$, za sve k i l

Tablica 3.7: Testovi značajnosti za parametre u modelu (3.3)

Drugi test u tablici (3.7) uspoređuje usklađenost modela (3.7b) i (3.3). To je test nulte hipoteze, $H_0 : \beta_{j(l)} = 0$, za sve j i l. Ako je ΔG^2 velik, zaključujemo da članovi pokazuju različitu popularnost, stoga parametri $\beta_{j(l)}$ moraju biti uključeni u bilo koji model koji bi adekvatno opisivao podatke. Ako statistika nije dovoljno velika, zaključujemo da se podaci mogu adekvatno opisati i bez uključivanja utjecaja popularnosti. Parametri alfa, beta i parametri povratnosti su neovisan skup parametara u smislu da za bilo koji određeni skup članova koji se mjeri na jednostrukoj vezi, bilo koji od tri utjecaja može biti statistički velik.

Poznavanje karakteristika parametara α za neki skup mrežnih podataka neće nas primjerice informirati o karakteristikama parametra β ili parametara povratnosti prisutnih u istoj mreži.

3.7. Primjeri

Promatramo prilagođeni model (3.3) i s njime povezane modele (3.7a), (3.7b) i (3.7c) i različite skupove podataka. Prvo ćemo prikazati prilagodbu modela za mrežu djece drugog razreda, i onda raspraviti o Krackhardtovoj vezi prijateljstva. Mreža učenika drugih razreda i Krackhardtova mreža se sastoje od dihotomnih veza, pa možemo direktno prilagoditi p_1. Freemanova EIES komunikacijska mreža, koju bi također rado modelirali, ima vrijednosne veze, pa tome moramo prilagoditi općenitiji model.

Mreža učenika drugih razreda

Tablica (3.8) daje statistička ocjenu prilagodbe za svaki od četiri modela primijenjenih na dihotomne prijateljske veze među učenicima drugog razreda.

Model	Nulta hipoteza	G^2	ΔG^2	Δdf
3.3		20.630		
3.7a	$H_0 : \alpha_{i(k)} = 0$, za sve i i k	23.072	2.442	3
3.7b	$H_0 : \beta_{j(l)} = 0$, za sve j i l	31.956	11.326	5
3.7c	$H_0 : (\alpha\beta)_{kl} = 0$, za sve k i l	22.564	1.934	1

Tablica 3.8: Statistička ocjena prilagodbe za mrežu učenika drugog razreda na početku školske godine

Tablica prikazuje koji parametri su statistički značajni za ove podatke. Primijetimo da parametri širenja nisu toliko značajni (G^2 je prilično malen), pa ne trebamo proučavati procjene parametra α_i svakog individualnog člana pokušavajući ga razlikovati među djecom.

Međutim, parametar popularnosti β u ovoj mreži je jako zanimljiv jer je $\Delta G^2 = 31.956 - 20.630 = 11.326$, što je puno. Također treba napomenuti da parametar povratnosti ne izgleda kao bitan faktor u vezi prijateljstva.

Krackhardtova mreža

Pogledajmo analizu podataka Krackhardtove mreže pokazanih u tablici (3.9).

Krackhardtovi high-tech menadžeri - veza savjetovanja

Model	Nulta hipoteza	G^2	ΔG^2	Δdf
3.3		322.564		
3.7a	$H_0 : \alpha_{i(k)} = 0$, za sve i i k	506.778	184.214	20
3.7b	$H_0 : \beta_{j(l)} = 0$, za sve j i l	440.471	117.907	20
3.7c	$H_0 : (\alpha\beta)_{kl} = 0$, za sve k i l	339.632	17.068	1

Krackhardtovi high-tech menadžeri - veza prijateljstva

Model	Nulta hipoteza	G^2	ΔG^2	Δdf
3.3		288.303		
3.7a	$H_0 : \alpha_{i(k)} = 0$, za sve i i k	421.965	133.662	20
3.7b	$H_0 : \beta_{j(l)} = 0$, za sve j i l	337.068	48.765	20
3.7c	$H_0 : (\alpha\beta)_{kl} = 0$, za sve k i l	312.455	24.152	1

Tablica 3.9: Statistička ocjena prilagodbe za Krackhardtove mrežu

Zasebno smo analizirali vezu savjeta i vezu prijateljstva.

Procjene α i β su navedene u tablici (3.10). Parametri širenja pokazuju da će za vezu savjeta članovi 3, 5 i 18 vjerojatno dati savjet, a članovi 2, 6 i 12 vjerojatno neće. Također možemo primijetiti da član 15 daje savjete svim drugim članovima. Parametar popularnosti za ovu vezu pokazuje da će član 2 vjerojatno primiti savjet, a članovi 9 i 15 vjerojatno neće. Parametar širenja za prijateljstvo pokazuje da član 17 ima mnogo drugih članova za prijatelje, dok članovi 7 i 9 (koji ne biraju) nemaju. Član 2 je relativno popularan prijatelj, a član 10 nije. Primijetimo dvojnu ulogu člana 2, daje jako malo savjeta, ali je prijatelj mnogim drugim članovima.

Parametri povratnosti za savjet $((\widehat{\alpha\beta}) = 2.233)$ i prijateljstvo $((\widehat{\alpha\beta}) = 2.937)$ su pozitivni i veliki pokazujući ozbiljnu tendenciju uzajamnosti.

	Savjetovanje		Prijateljstvo	
Član	$\hat{\alpha}_i$	$\hat{\beta}_i$	$\hat{\alpha}_i$	$\hat{\beta}_i$
n_1	-0.98	1.75	-0.37	1.40
n_2	-2.64	4.54	-1.46	2.36
n_3	2.63	-2.09	-1.36	0.66
n_4	1.40	-0.63	0.41	0.03
n_5	2.63	-2.09	0.60	0.34
n_6	-3.75	1.10	0.86	-1.84
n_7	-0.28	1.56	$-\infty$	0.04
n_8	-0.01	0.42	-2.22	0.83
n_9	2.02	-2.44	$-\infty$	1.39
n_{10}	1.95	-0.44	1.29	-3.17
n_{11}	-2.20	1.25	2.39	-0.41
n_{12}	-2.47	-0.08	-0.76	1.53
n_{13}	-0.13	-1.79	-0.58	-2.30
n_{14}	-1.58	0.79	-1.36	0.66
n_{15}	$+\infty$	-2.91	1.16	-0.74
n_{16}	-1.38	0.07	-1.18	0.19
n_{17}	-1.03	0.33	4.59	-0.96
n_{18}	2.55	1.48	-2.04	0.38
n_{19}	1.42	-2.27	1.33	-0.36
n_{20}	1.40	-0.63	-0.99	0.38
n_{21}	0.46	2.07	-0.33	0.33

Tablica 3.10: Procjena parametara za Krackhardtove high-tech menadžere

Procjene α i β za vezu savjeta mogu biti promatrane koristeći informaciju o godinama službe unutar organizacije svakog čimbenika. Procjenitelji $\widehat{\alpha}$ su u negativnoj korelaciji sa stažem, dok su $\widehat{\beta}$ u pozitivnoj korelaciji, pokazujući da iskusniji radnici rjeđe traže savjet od drugih i da češće daju savjete nego drugi. Staž ima jako malu korelaciju s $\widehat{\alpha}$ i $\widehat{\beta}$ u vezi prijateljstva.

3.8. Da li je p_1 distribucija slučajno usmjerenog grafa?

Obzirom na diskusiju o distribucijama slučajno usmjerenih grafova u početku rada, važno je pitanje kako se p_1 uspoređuje sa tim distribucijama.

Primijetimo da je p_1 eksponencijalna familija distribucija, sa minimalno dovoljnih statistika koje se sastoje od bridova koji ulaze u vrhove usmjerenog grafa, bridova koji izlaze iz vrhova usmjerenog grafa parametara i broja uzajamnih dijada. Dakle, bilo koja dva slučajna digrafa sa jednakim vrijednostima statistika imaju jednako prilagođenu vrijednost od p_1.

Pokazano je da p_1 ima Bernoullijevu distribuciju digrafa (vidi [4]). Takve distribucije fokusiraju se na skup p_{ij} vjerojatnosti, dajući vjerojatnosti da i izabire j, za sve parove članova i i j. Obzirom da digrafovi predstavljaju dihotomne veze, θ_{ij} se može definirati kao logit za p_{ij} :

$$\theta_{ij} = \log\left(\frac{p_{ij}}{1-p_{ij}}\right).$$

Osnovna mjera modela p_1 je dijada, a ne individualna veza ili luk. Dakle, p_1 se može promatrati kao Bernoullijeva distribucija dijada gdje je pretpostavka da su sve dijade, a ne veze ili lukovi, nezavisne.

Bibliografija

[1] Achuthan, S.B., Rao, S.B., and Rao A.R. (1982). The number of symmetric edges in a digraph with prescribed out-degrees. In Vijayan, K.S., and Singhi, N.M., eds., Proceedings of the Seminar on Combinatorics and Applications in honour of Professor S. S. Shrikhande on his 65th Birthday, pages 8-20. Calcutta: Indian Statistical Institute.

[2] Holland, P.W., and Leinhardt, S. (1977a). Notes on the statistical analysis of social network data. Unpublished manuscript.

[3] Holland, P.W., and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. Journal of the American Statistical Association 76, 33-65 (with discussion) .

[4] Karonski, M. (1982). A rewiew of random graphs. Journal of Graph Theory. 6, 349-389.

[5] Katz, L., and Powell, J.H. (1955). Measurement of the tendency toward reciprocation of choice. Sociometry. 18, 659-665.

[6] Lazarsfeld, P.F., and Merton, R.K. (1954). Friendship as a social process: A substantive and methodological analysis. In Berger, M., Abel, T., and Page, C.H. (eds.), Freedom and Control in Modern Society, pages 18-66. Princeton, NJ: Van Nostrand.

[7] Dr. sc. doc. S. Majstorović, Grafovi,
http://www.mathos.unios.hr/~smajstor/teaching.html

[8] Moreno, J.L., and Jennings, H.H. (1938). Statistics of social configurations. Sociometry. 1, 342-374.

[9] Walker, M.E., and Wasserman, S. (1987). TRIADS: A Computer Program for Triadic Analyses. Urbana, IL: University of Illinois.

[10] Wasserman S., and Faust K. (1994) Social network analysis: Methods and applications. New York, NY: University of Cambridge.

Sažetak

U ovom radu promatrali smo stohastičke modele za društvene mreže podataka, te uveli posebne vrste statističkih distribucija za slučajno usmjerene grafove i pokazali da su poseban slučaj uniformne distribucije slučajno usmjerenog grafa.

Prezentirali smo modele za mrežu s mjerenjima za jednostruke, direktne veze za jednu grupu promatranih podataka. Zatim smo opisali i prikazati interpretaciju i prilagodbu osnovnog statističkog modela za mrežu.

Predstavili smo modele napravljene za proučavanje kako veze od slobodnih odnosa variraju preko članova. Ovdje opisani modeli su modeli dijadičke interakcije koji koriste (prirodni) logaritam vjerojatnosti kao svoju osnovnu mjeru.

Summary

In this diploma thesis we considered a class of statistical distributions for random directed graphs, which, as we have shown, is a special case of the uniform random directed graph distributions.

We have presented models for a network with measurements on a single, directional relation for one set of actors. We have then described and demonstrated the interpretation and fitting of a basic statistical network model. Also, we can study models that are focused only on ties between groups of actors or just single actor,

We have presented a class of models designed to study how ties from a single relation vary across actors. The models described here are dyadic interaction models, which use the (natural) log of probabilities as their basic modeling unit.

Životopis

Rođena sam 04.08.1985. u Zagrebu. Nakon završene osnovne škole, pohađala sam X. gimnaziju „Ivan Supek“.

Maturirala sam 2004. Godine, te iste godine upisala Preddiplomski sveučilišni studij-Matematika na Prirodoslovno-matematičkom fakultetu, Matematički odjel, te stekla titulu prvostupnika matematike. Nakon toga upisala sam Diplomski sveučilišni studij Matematička statistika na istom fakultetu.