SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Tomislav Horina

MATEMATIČKA ANALIZA SOCIJALNIH MREŽA

Diplomski rad

Voditelj rada:
Ivica Nakić, prof. dr. sc

Zagreb, rujan,2018

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

Uvod

Mreža ili graf je pojednostavljeni prikaz sustava koji se sastoji od međusobno povezanih elemenata te se može promatrati kao povezanost gradova, atoma u molekulama, odnos između ljudi/organizacija i slično. One virtulano okupljaju ljude i organizacije ovisno o tome za što su specijalizirane te tako postoje poslovne, organizacijske i druge socijalne mreže. Socijalne mreže su pojednci i grupe koje povezuje zajednički status, slične funkcije, kulturne i druge sličnosti ili zajednički interes. Najpoznatije socijalne mreže su: Facebook, YouTube, WhatsUp i Instagram.

Prikupljanje podataka i provođenje istraživanja o socijalnim mrežma u prošlosti provodilo se otežano zbog velike količine podataka socijalnih mreža i primjene zastarjelih klasičnih metoda prikupljanja podataka (upitnici, promatranje, intervjui i tako dalje). Razvojem tehnologije i interneta, odnosno razvojem socijalnih mreža, olakšalo se prikupljanje podataka o socijalnim mrežama. Prikupljanje podataka o socijalnim mrežama provodi se putem API-ija (aplikacijskog programskog sučelja), web crawlers (pauka) i tako dalje.

Cilj ovog rada je prikazati mrežu kao matematički objekt i istražiti mogućnost kreiranja slučajne mreže sličnoj socijalnoj mreži. Koristite ćemo sljedeće modele slučajnih mreža: Erdös–Rényijev, Watts–Strogatzov, Barabási–Albert i eksponencijalne slučajne mreže.

U prvom poglavlju navedene su osnovne definicije mreža, prikaz vrste mreža te način zapisa mreže u matričnom obliku.

Drugo poglavlje prikazuje opis mreža pomoću parametara.

U trećem poglavlju napravljen je osvrt na socijalne mreže, načini prikupljanja podataka u socijalnim mrežama te karakteristike socijalnih mreža.

Četvrto poglavlje navodi modele slučajnih mreža, ulazne parametre slučajnih mreža i načini konstrukcije slučajnih mreža putem parametara.

Peto poglavlje prikazuje dio podataka na temu socijalnih mreža iz baze podataka mreže sveučilišta Stanford ([9]). Nakon prikaza podataka socijalnih mreža provedeno je istraživanje o mogućnosti stvaranja slučajne mreže sličnoj socijalnoj mreži.

Poglavlje 1

Osnovni pojmovi u teoriji mreža

U ovom poglavlju navest ćemo pojmove iz teorije mreža potrebne za izradu ovog rada. Prvi rad vezan uz mreže objavljen je 1736. godine kada je Euler riješio problem Koenigsberških mostova. Problem glasi: u gradu Koenigsberg (današnji Kalinjgrad) postojalo je sedam mostova na rijeci Pregel, te je građane zanimalo mogu li prijeći svih sedam mostova jednom tako da završe šetnju u početnoj točci. Prva knjiga iz teorije mreža napisao je D. Koenig otprilike 200 godina kasnije, što se smatra početkom razvoja teorije mreža. Šesdestih godinama dvadesetog stoljeća počinje veliki razvoj teorije mreža te se broj objavljenih radova više nego udeseterostručio. Razvojem informacijskih tehnologija, teorija mreža se intenzivno razvoja i ima veliku popularnost.

Mreža je uređen par $G = (V, E)$, gdje je V konačni skup čvorova i E je simetrična i irefleksivna relacija na V. Elemente relacije E nazivamo bridovi i označavat ćemo ih s (u, v) za brid između čvorova u i v. Kod usmjerene mreže relacija E ne mora biti simetrična. Definicija se može proširiti tako da dopustimo višestruke bridove (više bridova između para čvorova) i petlje (bridove koje spajaju čvor sa samim sobom). Mrežu koja ima višestruke bridove nazivamo multimreža. Mreža u kojoj bridovi imaju dodatne informacije (naprimjer, čvorovi su gradovi, bridovi su cestovni putovi, informacija je cestovna udaljenost između gradova) nazivamo težinska mreža, a dodatne informacije težine.

Dva čvora u i v susjedni su ako postoji brid $e = (u, v) \in E$. Kažemo da su čvorovi u i v incidentni s bridom e te da je brid e incidentan s čvorovima u i v. Dva brida $e_1 = (u, v)$ i $e_2 = (v, w)$ susjedni ako su oba incidentni s istim čvorom. Stupanj čvora je broj bridova s kojima je incidentan dani čvor. U usmjerenoj mreži, kažemo da je čvor u susjedan čvoru v ako postoji brid od u do v $e = (u, v)$. Brid iz u prema v je incidentan prema v; u je incidentan prema e i v je incidentan iz e. Ulazni stupanj je broj incidentnih bridova prema njemu, a izlazni stupanj je broj incidentnih bridova iz njega.

Čvorove mreže grafički označavamo s kružićima, a bridove spojnicama između vrhova. Jednu te istu mrežu možemo prikazati na različite načine, pa iz toga razloga navodimo kada su dvije mreže jednake (izomorfne). Dvije mreže G_1 i G_2 su izomorfne ukoliko je moguće označiti čvorove obje mreže na isti način i to tako da za svaki označeni par u i v čvorova, broj bridova koji spajaju u i v u G_1 je jednak broju bridova koji spajaju u i v u G_2. Ukoliko je mreža usmjerena, bridovi trebaju odgovarati i po smjeru.

Mreža $S = (V', E')$ je podmreža mreže $G = (V, E)$ ako je $V' \subseteq V$ i $E' \subseteq E$ te za svaki brid $e \in E$ vrijedi da su oba njegova vrha u V'. Podmreže dobivamo iz dane mreže brisanjem čvorova ili bridova. Ako je e neki brid mreže G, onda s $G - e$ označavamo mrežu G bez brida e. Ako je v vrh mreže G, onda je $G - v$ podmreža bez vrha v i svih bridova incidentnih sa v.

Ovdje navodimo načine kretanja (obilaska čvorova) po mreži. Šetnja u mreži dužine l je niz od (ne nužno različitih) čvorova $v_1, v_2, \ldots, v_l, v_{l+1}$ tako da za svaki $i = 1, 2, \ldots, l$ postoji brid između v_i do v_{i+1}. Zatvorena šetnja dužine l je šetnja $v_1, v_2, \ldots, v_l, v_{l+1}$ u kojoj je $v_{l+1} = v_1$. Put dužine l je šetnja dužine l u kojoj su svi čvorovi jedinstveni. Ciklus je zatvorena šetnja u kojoj su svi bridovi i svi čvorovi jedinstveni (osim prvog i zadnjeg čvora).

Mreža je povezana ako postoji put između bilo kojeg para čvorova u mreži, a u suprotnom je nepovezan. Svaka povezana podmreža je komponenta poveznaosti u mreži. Poveznanost čvorova u povezanoj mreži je najmanji broj čvorova čije uklanjanje čini mrežu nepovezanom. Povezanost bridova u povezanoj mreži je najmanji broj bridova čije uklanjanje čini mrežu nepovezanom. Usmjerena mreža je jako povezana ako postoji put između svakog para čvorova. Jako povezane komponente u usmjerenoj mreži su maksimalno jako povezane podmreže.

Navedimo sad nekoliko posebnih vrsta mreža. Stablo je povezana mreža bez ciklusa. Stablo od n čvorova ima $n-1$ brid te uklananjem jednog od njih mreža postaje nepovezana. Razapinjuće stablo mreže je podmreža koja je stablo i sadrži sve čvorove. Mreža u kojoj svi čvorovi imaju stupanj k nazivamo k-regularna mreža. Potpuna mreža, oznake K_n, je mreža s n čvorova gdje je svaki par čvorova povezan bridom. Prazna mreža, oznake $\overline{K}_n$, je mreža s n čvorova bez bridova.

K-bojanje je mreža gdje je svakom čvoru pridružena boja, tako da je svaki susjedni čvor drugčije obojan. Onda se kaže da je mreža k-obojiva. Kromatski broj je najmanji broj k za koji je mreža k-obojiva.

Mreža je bipartitna ako čvorove možemo podjeliti u dva disjunktna podskupa $V_1 \subset V(V_1 \neq \emptyset)$ i $V_2 \subset V(V_2 \neq \emptyset)$,te vrijedi $V_1 \cup V_2 = V$, tako da svaki brid povezuje čvor iz V_1 s čvorom iz V_2. Bipartitne mreže ne sadrže cikluse neparne duljine. Ako su svi čvorovi iz V_1 spojeni sa svima čvorovima iz V_2, mrežu nazivamo potpuna bipartitna mreža i označujemo ga s K_{n_1,n_2}, gdje su $n_1 = |V_1|$ i $n_2 = |V_2|$ ukupni brojevi čvorova. Bipartitna mreža je 2-obojiva.

Mreža je k-partitna ako se može podjeliti u k disjunktnih podskupova $V_1 \cup V_2 \cup \cdots \cup V_k = V$ $(V_i \neq \emptyset,\ \text{za } i = 1, 2, \ldots, k)$ tako da svaki brid povezuje čvorove iz različitih podskupova.

Ako su svi čvorovi svih podskupova povezani u druge podskupove, mreža se naziva potpuna k-partitina mreža, oznake $K_{n_1,n_2,\ldots,n_k}$, gdje je $n_i = |V_i|$ ukupan broj čvorova u V_i.

Zapisivanje svih bridova mreže pomoću čvorova nije praktično, stoga navodimo jedan način zapisivanja mrežu pomoću matrice. Neka je $G = (V, E)$ mreža. Matrica susjedstva A dimenzija $|V| \times |V|$ dana je formulom:

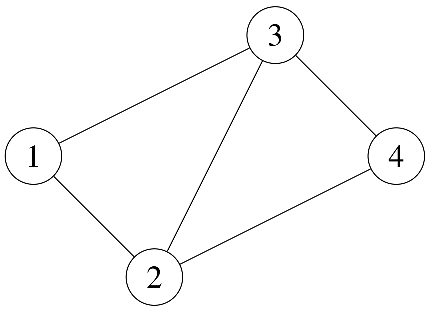
$$A_{ij} = \begin{cases} 1 & \text{ako } (i, j) \in E \\ 0 & \text{inače} \end{cases}$$

Ako mreža ne sadrži petlje, vrijednosti glavne dijagonale od A su nula. U slučaju težinskih mreža, vrijednost A_{ij} je jednaka w_{ij} ako su čvorovi i i j povezani i brid $\{i, j\}$ ima težinu w_{ij}. Matrica susjedstva za bipartitne mreže imaju dva skupa međusobno nepovezanih čvorova V_1 i V_2, gdje je $|V_1| = n_1$ i $|V_2| = n_2$, može se napisati kao:

$$A = \begin{bmatrix} 0 & R^T \\ R & 0 \end{bmatrix}$$

gdje je R $n_1 \times n_2$ matrica, a 0 matrica nula.

Primjer 1.1. Za danu mrežu navodimo matricu susjedstva.



$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

Označimo s k_u stupanj čvora u. U neusmjerenoj mreži, suma vrijednosti u-tog retka/stupaca matrice A je upravo k_u. Označimo s $\mathbf{1}$ vektor jedninica dimenzije $|V| \times 1$. Vektor stupnjeva čvorova k dan je s

$$k = (\mathbf{1}^T A)^T = A\mathbf{1} \tag{1.1}$$

Ulazni stupanj čvora je broj bridova koji su usmjereni prema danom čvoru. Označimo s k^{in} vektor ulaznih stupnjeva čvorova. Izlazni stupanj čvora je broj bridova koji izlazi iz čvora. Označimo s k^{out} vektor izlaznih stupnjeva čvorova. Ulazni i izlazni vektor stupnjeva se mogu dobiti:

$$k^{in} = (\mathbf{1}^T A)^T \tag{1.2}$$

$$k^{out} = A\mathbf{1}. \tag{1.3}$$

Drugi način prikazivanja stupnjeva čvorova je dijagonalna matrica K, koju nazivamo matrica stupnjeva čvorova mreže. Matrica susjedstva se može dobiti iz matrice incidencije B i matrice stupnjeva mreže. Matrica incidencije je $|V| \times |E|$ matrica čija vrijednost B_{ue} je 1 ako je čvor u incidentan s bridom e. Tada vrijedi $A = BB^T - K$.

Poglavlje 2

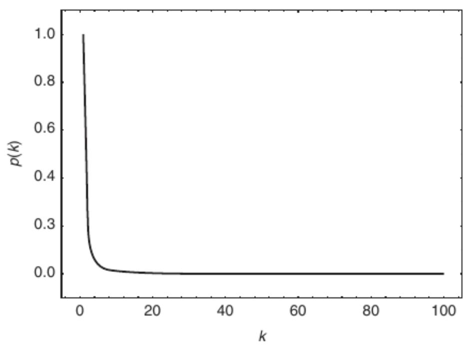
Parametri mreže

2.1 Parametri vezani uz susjedstvo

Stupanj–stupanj korelacija

Neka je $p(k) = n(k)/n$, gdje je $n(k)$ broj čvorova sa stupnjem k u mreži veličine n. Tada $p(k)$ predstavlja vjerojatnost da nasumični uniformno odabrani čvor ima stupanj k. Graf $p(k)$ naspram k predstavlja distribuciju stupnjeva za mrežu.

Od svih mogućih distribucija stupnjeva najčešće se koristi distribucija potencijskog oblika stupnjeva (slika 2.1). U toj distribuciji vjerojatnost ponalaska čvora sa stupnjem k iznosi otprilike $k^{-\gamma}$. To znači da je vjerojatnost pronalaska čvora s visokim stupnjem relativno mala u usporedbi s vjerojatnosti pronalaska čvora s niskim stupnjem.



$p(k) \sim k^{-\gamma}$

Slika 2.1: Distribucija potencijskog oblika

Stupanj–stupanj korelacija prikazuje kako je čvor s danim stupnjem spojen u mreži.

Naprimjer, mreža u kojoj su čvorovi s visokim stupnjem češće spojeni s čvorovima s visokim stupnjem pokazuje pozitivnu stupanj–stupanj korelaciju i tada se naziva međupoložena mreža (eng. assortative network). Mreže u kojoj čvorovi s visokim stupnjem teže da budu spojeni s čvorovima s niskim stupnjem pokazuju negativnu stupanj–stupanj korelaciju i naziva se nemeđupoložena (eng. diassortative) mreža.

Je li mreža međupoložena možemo odrediti tako što izmjerimo koeficijent korelacije od stupanj-stupanj korelacije. Taj koeficijent se naziva koeficijent međupoloženosti. Neka je $e(k_i, k_j)$ broj bridova koji spajaju čvor stupnja k_i s čvorom stupnja k_j. Ovdje se uzimaju stupnjevi čvorova za jedan manji nego stupanj toga čvora i nazivamo ih pretjerani stupnjevi (eng. excess degrees). Neka je $p(k_j)$ vjerojatnost da nasumično odabran čvor u mreži ima stupanj k_j. Tada je distribucija pretjeranog stupnja od čvora na kraju nasumično odabranog brida

$$q(k_j) = \frac{(k_j + 1)p(k_j + 1)}{\sum_i k_i p(k_i)}. \tag{2.1}$$

Koeficijent međupoloženosti je definiran kao:

$$r = \frac{\sum_{k_i k_j} k_i k_j [e(k_i k_j) - q(k_i)q(k_j)]}{\sigma_q^2}, \tag{2.2}$$

gdje je σ_q^2 standardna devijacija distribucije $q(k_j)$. Neka su k_i, k_j vektori stupnjeva čvorova brida u mreži. Ako je $k_i = k\mathbf{1}$ ili $k_j = k\mathbf{1}$ tada koeficijent međupoloženosti nije definiran jer imamo dijeljenje s nulom u (2.2).

U slučaju usmjerene mreže, koeficijent međupoloženosti uzima u obzir distribuciju za oba kraja brida i njihovu standarnu devijaciju. Zapisuje se s

$$r = \frac{m^{-1} \sum_e k_i(e)k_j(e) - \left[m^{-1} \sum_e \frac{1}{2}[k_i(e) + k_j(e)]\right]^2}{m^{-1} \sum_e \frac{1}{2}[k_i^2(e) + k_j^2(e)] - \left[m^{-1} \sum_e \frac{1}{2}[k_i(e) + k_j(e)]\right]^2}, \tag{2.3}$$

gdje su $k_i(e)$ i $k_j(e)$ stupnjevi oba kraja brida e i $m = |E|$.

Susjedstvo bridova

Označimo s P matricu susjedstva bridova neusmjerene mreže. Matrica je kvadratna i simetrična čije vrijednosti su jednice ili nule ovisno jesu li bridovi susjedni ili ne. Matrica se može dobiti iz:

$$P = B^T B - 2I, \tag{2.4}$$

gdje je I matrica identiteta. Kao kod susjedstva čvorova, može se promatrati stupanj brida κ_i. Neka su u i v krajne točke brida i, $i = (u, v)$, tada je

$$\kappa_i = k_u + k_v - 2. \tag{2.5}$$

Vektor stupnja bridova κ možemo dobiti iz:

$$\kappa = (\mathbf{1}^T P)^T = P\mathbf{1}. \tag{2.6}$$

Laplaceov matrica

Za danu mrežu G, Lalpaceova matrica se definira se:

$$L = K - A, \tag{2.7}$$

gdje je A matrica susjedstva, a K matrica stupnjeva.

U slučaju povezanih mreža korisno je definirati normaliziranu Laplaceovu matricu kao:

$$\mathcal{L} = K^{-1/2} L K^{-1/2} = I - K^{-1/2} A K^{-1/2}. \tag{2.8}$$

Spektralna svojstva

Spektar čvorova matrice susjedstva mreže je skup svojstvenih vrijednosti od A i njihove kratnosti. Neka su $\lambda_1(A) > \lambda_2(A) > \ldots > \lambda_n(A)$ jedinstvene svojstvene vrijednosti od A i neka su $m(\lambda_1(A)), m(\lambda_2(A)), \ldots, m(\lambda_n(A))$ njihove kratosti. Tada se spektar od A može zapisati kao

$$\operatorname{Sp} A = \begin{bmatrix} \lambda_1(A) & \lambda_2(A) & \ldots & \lambda_n(A) \\ m(\lambda_1(A)) & m(\lambda_2(A)) & \ldots & m(\lambda_n(A)) \end{bmatrix}. \tag{2.9}$$

Najveća svojstvena vrijednost matrice susjedstva A naziva se indeks mreže ili spektralni radijus od A.

Neka je Λ diagonalna matrica svojstvenih vrijednosti matrice $\Lambda = \operatorname{diag}(\lambda_1(A), \lambda_2(A), \ldots, \lambda_n(A))$, te neka je Φ matrica čiji su stupci ortonomirani svojstveni vektori $\overrightarrow{\varphi}_1, \overrightarrow{\varphi}_2, \ldots, \overrightarrow{\varphi}_n$. Tada je spektrala dekompozicija matrice susjedstva dana s:

$$A = \Phi \Lambda \Phi^T \tag{2.10}$$

Korištenjem spektralne dekompozicije matrice susjedstva, potencije matrice susjedstva može se napisati:

$$(A^k)_{uv} = \sum_{j=1}^{n} \varphi_j^u(A) \varphi_j^v(A) [\lambda_j(A)]^k \tag{2.11}$$

gdje je $\varphi_j^u(A)$ u-ta vrijednost svojstvenog vektora povezanog s j-tom svojstvenom vrijednosti. $(A^k)_{uv}$ broji koliko ima šetnji dužine k između čvorova u i v.

Laplaceov spektar je skup svojstvenih vrijednosti od L s njihovim kratnosti. Neka su $\lambda_1(L) < \lambda_2(L) < \ldots < \lambda_n(L)$ jedinstvene svojstvene vrijednosti od L i neka su $m(\lambda_1(L))$, $m(\lambda_2(L)), \ldots, m(\lambda_n(L))$ njihove kratnosti. Tada se spektar od L može zapisati kao

$$\operatorname{Sp} L = \begin{bmatrix} \lambda_1(L) & \lambda_2(L) & \ldots & \lambda_n(L) \\ m(\lambda_1(L)) & m(\lambda_2(L)) & \ldots & m(\lambda_n(L)) \end{bmatrix}. \tag{2.12}$$

Kratnost $\lambda_1(L) = 0$ jednaka je broju povezanih komponeti u mreži. Druga najmanja svojstvena vrijednost od L, $\lambda_2(L)$, najćešče se naziva algebarska povezanost mreže.

Za regularnu matricu M, elementi inverza mogu se dobiti iz

$$(M^{-1})_{ij} = \sum_{k=1}^{n} \frac{1}{\tau_k} c_k(i) c_k(j), \tag{2.13}$$

gdje su τ_k svojstvene vrijednosti od M i $c_k(i)$ je i-ta vrijednost svojstvenog vektora povezanog s τ_k. Za singularne matrice koristi se generalni pseudoinverz, oznake M^+. Generalni pseudoinverz Laplaceove matrice L^+ ima važne primjene u teoriji grafova. Za povezani graf, matrica L^+ je realna i simeterična, te se može zapisati

$$U^T L^+ U = \mathrm{diag}(1/\mu_1, 1/\mu_2, \ldots, 1/\mu_{n-1}, 0), \tag{2.14}$$

gdje je U matrica ortonomiranih svojstvenih vektora od L^+.

2.2 Metrička i topološka struktura mreže

Najkraća udaljenost puta

U neusmjerenoj mreži, duljina puta $d(u, v)$ je broj bridova u najkraćem putu između čvorova u i v u mreži. U slučaju kada u i v su u različitim komponetama povezanosti, duljina je beskonačna, $d(u, v) = \infty$. U teoriji mreža, direktna duljina $\overrightarrow{d}(u, v)$ između para čvorova u i v u usmjerenoj mreži je duljina usmjerenog najkraćeg puta iz u do v. U slučaju da ne postoji usmjereni put između dva čvorova, duljina je beskonačna.

U neusmjernoj povezanoj mreži, duljina $d(u, v)$ među svim parovima čvorova u mreži možemo zapisati matrično. Ova kvadratna i simetrična matrica poznata je kao matrica udaljenosti mreže D. U usmjerenoj mreži ova matrica nije nužno simetrična i može sadržavati beskonačne vrijednosti.

Maksimalna vrijednost za dani redak/stupac matrice udaljenosti od neusmjerene (jako povezane usmjerene) mreže je poznata kao ekscentritet (eng. eccentricity) $e(v)$ od čvora v s

$$e(v) = \max_{x \in V(G)} d(v, x). \tag{2.15}$$

Maksimum ekscentriteta između svih čvorova mreže poznat je kao promjer mreže. Radijus mreže je minimalna ekscentritet čvorova. Čvor se naziva centar ako je ekscentritet jednak radijusu mreže.

Suma svih vrijednosti retka/stupca u matrici udaljenosti je poznata kao suma udaljenosti $s(u)$ od danog čvora:

$$s(u) = \sum_{v \in V(G)} d(u, v). \tag{2.16}$$

S $W(G)$ označavamo Wienerov indeks kojeg možemo izraziti s:

$$W(G) = \frac{1}{2}\mathbf{1}^T D\mathbf{1} = \sum_{u=1}^{n} s(u) = \frac{1}{2}\sum_{u}\sum_{v} d(u,v). \tag{2.17}$$

Prosječna dužina puta $\bar{l}$ može se dobiti preko Weinerovog indeksa:

$$\bar{l} = \frac{2W(G)}{n(n-1)}. \tag{2.18}$$

Udaljenost otpora

Kako bismo definirali ovu udaljenost između para čvorova u i v, pretpostavimo da je postavljen fiksni električni otpornik na svaki brid mreže i da je spojena baterija na čvorove u i v. Radi jednostavnosti uzima se vrijednost otpornika $1 Ohm$. Tada se računa efektivan otpor $\Omega(u,v)$ između čvorova pomoću Kirchhoffovih i Ohmovih zakona. Karakteristično je da se efektivan otpor $\Omega(u,v)$ smanjuje kako se povećava broj putova između u i v. Zato generalno vrijedi $\Omega(u,v) \leq d(u,v)$. Klein i Randić ([7]) pokazali su da je efektivan otpor funkcija udaljenosti, koju nazivamo udaljenost otpora. Ako su u i v u različitim komponetama povezanosti mreže tada je $\Omega(u,v) = \infty$.

Udaljenost otpora $\Omega(u,v)$ između čvorova unutar istih komponenti povezanosti mreže može se izračunati korištenjem generaliziranog inverza L^+ Laplaceove matrice grafa L. Udaljenost otpora između čvorova u i v se računa s

$$\Omega(u,v) = L^+(u,u) + L^+(v,v) - 2L^+(u,v) \tag{2.19}$$

za $u \neq v$.

Neka je $L(G-u)$ matrica kojoj je maknut u-ti redak i stupac iz Laplaceove matrice i neka je $L(G-u-v)$ matrica kojoj je maknut u i v redak i stupac iz L. Tada se udaljenost otpora između čvorova može izračunati s

$$\Omega(u,v) = \frac{\det L(G-u-v)}{\det L(G-u)}. \tag{2.20}$$

Još jedan način računanja udaljenosti otpora između čvorova koji koristi Laplaceov spektar je dan s:

$$\Omega(u,v) = \sum_{k=2}^{n} \frac{1}{\mu_k}\left[U_k(u) - U_k(v)\right]^2, \tag{2.21}$$

gdje je $U_k(u)$ u-ta vrijednost k-tog ortonomiranog svojstvenog vektora povezanog s Laplaceovom svojstvenom vrijednošću μ_k, $0 = \mu_1 < \mu_2 \leq \ldots \leq \mu_n$.

Udaljenost otpora između svih parova čvorova u mreži može se zapisati matrično i naziva se matrica otpora Ω mreže. Matrica se može dobiti iz:

$$\Omega = \mathbf{1}\,\mathrm{diag}\left[\left(L + \frac{1}{n}J\right)^{-1}\right]^T + \mathrm{diag}\left(L + \frac{1}{n}J\right)^{-1}\mathbf{1}^T - 2\left(L + \frac{1}{n}J\right)^{-1} \tag{2.22}$$

gdje je $J = \mathbf{1}\mathbf{1}^T$ matrica s jedinicama na svim mjestima. Suma udaljenosti otpora $R(u)$ je suma svih vrijednosti u u-tom retku/stupcu matrice otpora:

$$R(u) = \sum_{v \in V(G)} \Omega(u, v). \tag{2.23}$$

Analogon Wienerovon indeksu za udaljenost otpora naziva se Kirchhoffov indeks mreže i dan je s:

$$Kf(G) = \frac{1}{2}\mathbf{1}^T\Omega\mathbf{1} = \sum_u R(u) = \frac{1}{2}\sum_u \sum_v \Omega(u, v). \tag{2.24}$$

Omjer fragmenata

Koeficijenti klasteriranja

Podmreža koja se sastoji od tri čvora naziva se trokut. Postoje dvije vrste trokuta: otvoreni i zatvoreni. Trokut, koji se sastoji od čvorova u, v i z, je otvoren ako postoje bridovi (u, v) i (v, z) i ne postoji brid (u, z), dok je zatvoren ako postoji brid (u, z).

Tranzitivnost mreže je parametar koji se definira kao omjer broja zatvorenih trokuta (C_3) i svih trokuta(C_t):

$$T = \frac{3|C_3|}{|C_t|}. \tag{2.25}$$

Blizak parametar tranzitivnosti mreže je koeficijent klasteriranja. On se definira kao omjer mogućih zatvorenih trokuta kroz zadani čvor i :

$$C_i = \frac{2|C_3(i)|}{k_i(k_i - 1)}, \tag{2.26}$$

gdje je k_i stupanj čvora. Prosječna vrijednost klasteriranja svih čvorova u mreži $\overline{C}$ koristi se u analizi kompleksnih mreža i dana je s

$$\overline{C} = \frac{1}{n}\sum_{i=1}^{n} C_i. \tag{2.27}$$

Mrežni reciprocitet

Još jedan parametar baziran na omjeru broja podmreža u mreži je reciprocitet, definiran za usmjerenu mrežu kao:

$$r = \frac{L^{\leftrightarrow}}{L}, \tag{2.28}$$

gdje je $L^{\leftrightarrow}$ broj dvosmjernih bridova, a L ukupan broj usmjerenih bridova. Tako, ako postoji brid od A do B, reciprocitet mjeri vjerojatnost da postji brid od B do A.

Estrada indeks

Neka je M_k broj zatvorenih šetnji dužine k. Estrada indeks $EE(G)$ definiramo kao:

$$EE(G) = \sum_{k=0}^{\infty} \frac{M_k}{k!} = \sum_{k=0}^{\infty} \sum_{j=1}^{n} \lambda_j^k = \sum_{j=1}^{n} e^{\lambda_j} = \mathrm{tr}(e^A), \tag{2.29}$$

gdje su λ_j svojstvene vrijednosti matrice susjedstva A. Koeficijent $\frac{1}{k!}$ osigurava da red konvergira. $EE(G)$ daje veću težinu manjim podmrežama.

Promotrimo slučaj kada svaki brid mreže ima težinu s parametrom β. Neka je W matrica susjedstva te homogene težinske mreže. Tada je $W = \beta A$ i spektar matrice susjedstva je $M_r(W) = \mathrm{tr}\, W^r = \beta^r \,\mathrm{tr}\, A^r = \beta^r M_r$. Tada se Estrada indeks može napisati kao:

$$EE(G, \beta) = \sum_{r=0}^{\infty} \frac{\beta^r M_r}{r!} = \sum_{j=1}^{N} e^{\beta\lambda_j} = \mathrm{tr}(e^{\beta A}). \tag{2.30}$$

2.3 Komunikabilnost u mreži

Mrežna komunikabilnost

Pretpostavimo da komunikabilnost između para čvorova u mreži zavise o svima putovima što spajaju ta dva čvora. Između svih tih putova, najkraći put je onaj koji daje najviše doprinosa. Neka je $P_{rs}^{(l)}$ broj najkraćih putova između čvorova r i s koji su dužine l i neka je$W_{rs}^{(k)}$ broj šetnji koje spajaju r i s dužine $k > l$. Komunikabilnost možemo izraziti s

$$G_{rs} = C_l P_{rs} + \sum_{k>l} C_k W_{rs}^{(k)}, \tag{2.31}$$

gdje su koeficijenti C_k takvi da red konvergira. Za koeficijente najčešće se koristi $C_k = \frac{1}{k!}$. Korištenjem veze između potencija matrice susjedstva i broja šetnji u mreži, G_{rs} možemo zapisatis

$$G_{rs} = \sum_{k=0}^{\infty} \frac{(A^k)_{rs}}{k!} = \left(e^A\right)_{rs}. \tag{2.32}$$

Korištenjem parametra β, G_{rs} može se zapisati s

$$G_{rs}(\beta) = \sum_{k=0}^{\infty} \frac{(\beta A^k)_{rs}}{k!} = \left(e^{\beta A}\right)_{rs}. \tag{2.33}$$

Komunikabilnost u usmjerenim i težinskim mrežama

Komunikabilnost u usmjerenim mrežama se može izračunati za bilo koji par čvorova u usmjerenoj mreži korištenjem izraza (2.31). U slučaju težinske mreže preporuča se korištenje normirane matrice susjedstva. Koristi se sljedeći izraz za komunikabilnost u težinskim mrežama:

$$\tilde{G}_{pq} = \left[\exp(W^{-1/2} A W^{-1/2})\right]_{pq}, \tag{2.34}$$

gdje je A težinska matrica susjedstva, a W je diagonalna matrica težina stupnjeva čvorova.

2.4 Mjere centralnosti

Centralnost stupnja

Ideja korištenja stupnja čvora kao mjere centralnosti je sljedeća čvor je više centralan ili ima više značaja nego drugi ako je stupanj prvog čvora veći od stupnja drugog čvora. Stupanj čvora broji šetnje duljine 1 za dani čvor ili zatvorene šetnje duljine 2 danog čvora:

$$k_i = (A^2)_{ii}. \tag{2.35}$$

Naprimjer, ako se promatra proces u kojem je informacija proslijeđena s jednog čvora na drugi u mreži, mogućnost primitka takve informacije kroz mrežu proporcionalna je stupnju danog čvora.

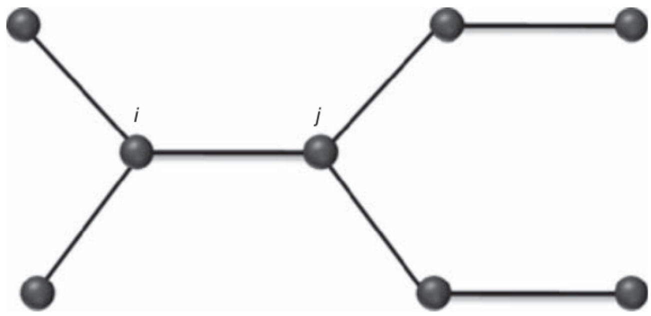
Kod usmjerenih mreža potrebno je korisiti ulazni i izlazni stupanj. Korelacija između ulaznog i izlaznog stupnja može se izmjeriti Pearsonovim koeficijentom korelacije Ω:

$$\Omega = \frac{\sum_{i=1}^{n}\left(k_i^{in} - \overline{k^{in}}\right)\left(k_i^{out} - \overline{k^{out}}\right)}{\sqrt{\sum_{i=1}^{n}\left(k_i^{in} - \overline{k^{in}}\right)^2}\sqrt{\sum_{i=1}^{n}\left(k_i^{out} - \overline{k^{out}}\right)^2}}, \tag{2.36}$$

gdje su $\overline{k^{in}}$ i $\overline{k^{out}}$ prosječni ulazni i izlazni stupanjevi za mrežu. Prema koeficijentu, mreža može biti pozitivno ($\Omega > 0$), negativno ($\Omega < 0$) ili neutralno ($\Omega = 0$) korelirana.

Centralnost iznad najbližeg susjeda

Problem kod centralnosti stupnja je da je dobivena na osnovu utjecaja najbližeg susjeda za dani čvor. U mnogim stvarnim situacijama neophodno je promatrati i utjecaj drugih čvorova koji nisu susjedni. Na slici 2.2 prikazana je jednostavna mreža s naznačenim



Slika 2.2: Jednostavan graf s naznačenim čvorovima i i j

čvorovima i i j oba stupnja 3. Susjedi čvora i nemaju drugih susjeda te utjecaj dobiva samo od susjednih čvorova. Susjedni čvorovi od j imaju svoje susjedne čvorove koji indirektno utječu na čvor j.

Stupanj čvora može se izraziti kao:

$$k_i = \left[\left(A^0 + A^1\right) - I\right] \mathbf{1}_i, \tag{2.37}$$

gdje je I matrica identiteta. A^0 broji šetnje dužine nula, što je ekvivalentno ostajanju u istom čvoru, i A^1 broji šetnje dužine jedan, što odgovara posjećivanju najbližeg susjeda.

Katz centralnost

Kada se promatra proširenje jednadžbe (2.37) na bazi drugih potencija matrice susjedstva, treba uzeti u obzir da red $A^0 + A^1 + A^2 + \ldots + A^k + \ldots$ divergira. Stoga uvodimo faktor $\alpha > 1$ tako da red konvergira i Katz centralnost se izražava kao:

$$K_i = \left[\left(\sum_{k=0}^{\infty} \alpha^{-k} A^k\right) \mathbf{1}\right]_i = \left[\left(I - \frac{1}{\alpha} A\right)^{-1} \mathbf{1}\right]_i. \tag{2.38}$$

Faktor treba biti odabran tako da $\alpha \neq \lambda_1$, gdje je λ_1 je prva svojstvena vrijednost matrice susjedstva, kako bi se izbjegla divergencija u (2.38). Faktor se bira tako da bude manji od λ_1 kako bi se dobili značajniji centralni indeksi.

Centralnost svojstvenih vektora

Promotrimo vektor centralnosti za čvorove mreže korištenjem slijedeće varijante Katz indeksa:

$$c = \left(\sum_{k=1}^{\infty} \alpha^{1-k} A^k \right) \mathbf{1}. \tag{2.39}$$

Korištenjem spektralne dekompozicije matrice susjedstva:

$$A^k = \sum_{j=1}^{n} \varphi_j \varphi_j^T \lambda_j^k \tag{2.40}$$

indeks se može napisati kao:

$$c = \alpha \left(\sum_{j=1}^{n} \frac{\lambda_j}{\alpha - \lambda_j} \varphi_j \varphi_j^T \right) \mathbf{1}. \tag{2.41}$$

i-ta glavna vrijednost svojstvenog vektora matrice susjedstva poznata je kao svojstveni vektor centralnosti čvora i:

$$\varphi_i(i) = \left(\frac{1}{\lambda_1} A \overrightarrow{\varphi_i} \right)_i . \tag{2.42}$$

Kod korištenja na usmjerenoj mreži, dva tipa centralnosti svojstvenih vektora mogu se definirati korištenjem desnog i lijevog glavnog svojstvenog vektora matrice susjedstva:

$$\varphi_1^R(i) = \left(\frac{1}{\lambda_1} A \varphi_1^R \right)_i , \tag{2.43}$$

$$\varphi_1^L(i) = \left(\frac{1}{\lambda_1} A^T \varphi_1^L \right)_i . \tag{2.44}$$

Desna centralnost svojstvenog vektora računa značajnost čvora tako da uzima u obzir značajnost čvora u kojeg je usmjeren. Lijeva centralnost svojstvenog vektora mjeri značajnost čvora uzimajući u obzir čvorove koji su usmjereni prema danom čvoru .

Centralnost podmreže

Ideja iza centralnosti podmreže je da se karakterizira značajnost čvora uzimajući u obzir sve zatvorene šetnje koje počinju u njemu. Centralnost podmreže može se definirati kao:

$$f_i(A) = \left(\sum_{l=0}^{\infty} c_l A^l \right)_{ii} , \tag{2.45}$$

gdje su koeficijenti c_l odabrani tako da red konvergira. Za koeficijent može se koristiti isti onaj koji je korišten kod Estradinog indeksa. Tada je centralnost dana s:

$$EE(i) = \left(\sum_{l=0}^{\infty} \frac{A^l}{l!} \right)_{ii} = \left(e^A \right)_{ii}. \tag{2.46}$$

Centralnost bliskosti

U nekim slučajevima promatra se koliko je čvor relativno blizu svim drugim čvorovima u mreži. U tom slučaju koristi se mjera centralnosti čvora koja se dobiva iz inverza sume najkraćih putova od danog čvora do svih ostalih čvorova u mreži. Ta mjera se naziva centralnost bliskosti (eng. closeness centrality) koja se može izraziti kao:

$$CC(u) = \frac{n-1}{s(u)}, \tag{2.47}$$

gdje je $s(u)$ definiran u (2.16).

Centralnost između čvorova

Centralnost između čvorova (eng. betweenness centrality) uzima u obzir omjer informacija koje prolaze kroz dani čvor u komunikaciji drugih parova čvorova u mreži. Ako je $\rho(i, j)$ broj nakraćih putova između i i j i $\rho(i, k, j)$ broj nakraćih putova kroz čvor k , tada je centralnost između čvorova za zadani čvora k dana s:

$$BC(k) = \sum_i \sum_j \frac{\rho(i, k, j)}{\rho(i, j)}, i \neq j \neq k. \tag{2.48}$$

Nekoliko mjera je predloženo za računanje centralnosti čvora kada se promatra komunikacija korištenjem alternativnih putova.

Jedna od tih mjera koristi tok informacija između čvorova ne samo korištenjem najkraćih putova. Mjera se naziva centralnost toka između čvorova i za čvor k te se definira kao:

$$BC_f(k) = \sum_i \sum_j \frac{f(i, k, j)}{f(i, j)}, i \neq j \neq k, \tag{2.49}$$

gdje je $f(i, j)$ maksimalan tok između čvorova i i j i $f(i, k, j)$ maksimalan tok između danih čvorova kroz čvor k. Maksimalni tok između dva čvora je minimalni kapacitet prereza. Neka su i i j povezani barem jednim putom te neka je $E_{ij} \subseteq E$ podskup bridova koji sadrži jednom brid iz svakog puta između i i j. Podskup E_{ij} prestavlja $i - -j$ prerez i njegov kapacitet je suma kapaciteta (težine) bridova. Minimalni kapacitet prereza je najmanji kapacitet od svih prereza.

Druga varijanta je centralnost komunikabilnosti između čvorova. Za dani čvor r mjera je definirana kao:

$$BC_r = \frac{1}{C}\sum_p \sum_q \frac{G_{prq}}{G_{pq}}, \; p \neq q, p \neq r, q \neq r, \tag{2.50}$$

gdje je G_{prq} jefunkcija komunikabilnosti između čvorova p i q gdje promatramo samo šetnje koje uključuju r, G_{pq} funkcija komunikabilnosti između čvorova p i q i $C = (n-1)^2 - (n-1)$ je faktor normiranja koji je jednak broju članova sume tako da BC_r poprima vrijednosti između nula i jedan.

2.5 Globalne invarijante mreže

Prethodno su prikazane neke lokalne invarijante za karakerizaciju čvorova, a sada želimo želimo promotriti mrežu kao cjelinu kroz invarijate strukture. Jedan primjer globalne invarijante je indeks centralosti (mreža koja pokazuje najveću varijabilnost dane mjere centralnosti smatra se da ima najveću centralnost). Za mrežu koja za danu mjeru centralosti ima iste vrijednosti za sve čvorove očekuje se da ima najnižu centralnost. Generalni indeks centralosti C^T za centralost vrste T može se izraziti s:

$$C^T = \frac{\sum_{i=1}^{n}(C^T_{\max} - C^T_i)}{\max \sum_{i=1}^{n}(C^T_{\max} - C^T_i)}, \tag{2.51}$$

gdje je C^T_i centralnost čvora i u mreži, te $C^T_{\max} = \max C^T_i$.

Promotrimo centralnost stupnja čvora, $C^T = C^D$. Ako je mreža povezana onda znamo da je maksimalni stupanj koji čvor može imati $C^T_{\max} = n-1$, a minimalni $C^T_i = 1$. Izraz za centralnost stupnja čvora je:

$$C^D = \frac{\sum_{i=1}^{n}(k_{\max} - k_i)}{(n-1)(n-2)}. \tag{2.52}$$

Ovaj indeks poprima vrijednosti između nula i jedan, gdje se minimum dobiva za regularne mreže, a maksimum za zvjezdastu mrežu.

Gustoća je globalna mjera mreže i možemo ju izraziti s:

$$\delta(G) = \frac{2m}{n(n-1)} = \frac{\overline{k}}{n-1}, \tag{2.53}$$

gdje je $m = |E|$ broj bridova u mreži, a $\overline{k}$ prosjek stupnjeva. Mreža se smatra raspršenom ako je $\overline{k} << n$, što je ekvivaletno $\delta(G) << 1$.

Raspršenost se može gledati i lokalno, gledajući susjedstvo zadanog čvora. Neka je $N_\alpha(i) \subseteq V$ α-susjedstvo od čvora i , to jest, $v_j \in N_\alpha(i)$ ako i samo ako $d(v_i, v_j) \leq \alpha$, $\forall v_j \in V$. Neka je $n_i = |N_\alpha(i)|$ broj čvorova, a m_i broj bridova u $N_\alpha(i)$. Tada je lokalna raspršenost za čvor i dana s

$$d_\alpha(i) = \frac{2m_i}{n_i(n_i - 1)}. \tag{2.54}$$

Promotrimo indekse bazirane na centralnosti bliskosti i centralosti između čvorova čvora. Njih možemo zapisati s:

$$C^C = \frac{(2n-3)\sum_{i=1}^{n}(C^C_{\max} - C^C_i)}{(n-1)(n-2)}, \tag{2.55}$$

$$C^B = \frac{2\sum_{i=1}^{n}(C^B_{\max} - C^B_i)}{(n-1)^2(n-2)}. \tag{2.56}$$

Wienerov indeks smo definirali u (2.17) i njegov normalizirani oblik možemo zapisati kao

$$\zeta(G) = \frac{n(n^2-1) - 6W}{n(n-1)(n-2)}. \tag{2.57}$$

Ovdje dobivamo vrijednost nula za put, te vrijednost jedan za potpuni graf.

Globalni indeks efikasnosti mreže dan je s:

$$E_{Global} = \frac{2}{n(n-1)} \sum_{i<j} \frac{1}{d(i,j)}. \tag{2.58}$$

Poglavlje 3

Socijalne mreže

Socijalne mreže sastoje se od osoba ili grupa osoba (aktera) koji su u nekom odnosu. Socijalna mreža može se modelirati mrežom gdje su čvorovi osobe ili grupe, a brid između dva čvora predstavlja inerakciju između osoba ili grupa.

Postoje tri vrste socijalnih mreža koje sociolozi proučavaju: egocentrične mreže, sociocentrične mreže i otvorene mreže. Kod egocentričnih mreža postoji jedan čvor koji je spojen sa svim ostalima. Primjer takve mreže je učitelj koji je povezan sa svojim učenicima. Sociocentrične mreže su zatvorene mreže u kojoj su čvorovi povezani međusobno. Primjer takve mreže je mreža radnika u tvornici. U otvorenim mrežama veze nisu najjasnije definirane. Primjer takve mreže je mreža korporacija.

Razvojem tehnologije razvile su se socijalne mreže na internetu. Internetske socijalne mreže dopuštaju korisnicima dijeljenje teksta, zvuka, fotografija , videa, . . . Prema stranici Statista, trenutno najpopularnije mreže su Facebook, YouTube, WhatsApp i Instagram. Postoje četri osnovne karakteristike pomoću kojih se razlikuju internetske socijalne mreže od ostalih stranica:

- bazirane su na korisnicima: prije je na stranicama bio postavljen sadržaj od strane administratora stranice te je bio čitan od strane posjetitelja stranice. Tok informacija je bio u jednom smjeru. Sada je sadržaj stranice određen od strane svih osoba koje sudjeluju u raspravi.

- interaktivne: to znači da sadržaj stranice nije samo napravljen od foruma ili privatnih razgovora, nego ima i raznih igra, videa i slično.

- upravljane su od strane zajednica: članovi socijalnih mreža su povezani ovisno o zajedničkim hobijima i uvjerenjima.

- odnosi između članova: članovi socijalnih mreža koji su povezani s više članova teže centru mreže.

3.1 Podaci socijalnih mreža

Jedna od karakteristika socijalnih mreža je da su podaci utemeljeni na simbolima i kulturnim vrijednostima i oblikovani su motivima i značenjima. Može ih se podjeliti u dva tipa ovisno o organizaciji podataka. U prvom načinu, svi entiteti reprezentirani čvorovima pripadaju istoj klasi, kao naprimjer pojednici, korporacije, nogometne momčadi. U drugom načinu, entiteti su dvije različite prirode, kao naprimjer banke i korporacije.

Implemetacija analize mreže zahtjeva potpuni skup podataka ili uzorak iz potpunog skupa podataka. Bitan aspekt prikupljanja podataka za gradnju socijalnih mreža je u prirodi odnosa koji se proučavaju. To određuje samu prirodu mreže koja se proučava, a svi analizirani rezultati trebaju biti u ovom konkretnom kontekstu. Među tim tipovima odnosa su oni temeljeni na individualnim procjenama, kao što su prijateljstvo, povezivanje, poštovanje, uzajamna pomoć i tako dalje. Još jedno važno područje obuhvaća one koje uključuju transakcije ili prijenos materijalnih resursa, kao što su pozajmljivanje novca, komercijalne transakcije, transakcije kupca i prodavatelja i druge. U našem svijetu komunikacijske razmjene, prijenos važnih resursa, kao što su informacije, sve je važniji. Druge vrste odnosa mogu uključivati fizičke pokrete, formalne uloge, srodstvo, fizičke interakcije i drugo.

Postoji nekoliko metoda prikupljanja podataka, podijeljene na trandicionalne metode prikupljanja i prikupljanje putem interneta. Tradicionalne metode prikupljanja podataka su:

- upitnici – najčešći način prikupljanja podataka. Pogodni su zato što se može postaviti pitanje koje se direktno odnosi na cilj istraživanja.
- intervjuji – koristi se kada se podaci žele saznati komunikacijom (licem u lice, telefonom, ...). Koriste se kada upitnici ne daju željeni rezultat.
- promatranje – promatraju se interakcije između aktera. Koristi se kod malog broja aktera mreže, kad intervjuji nisu mogući.
- arhiva – proučavaju se iz starih zapisa. Naprimjer, odnos predsjednika i nacije.

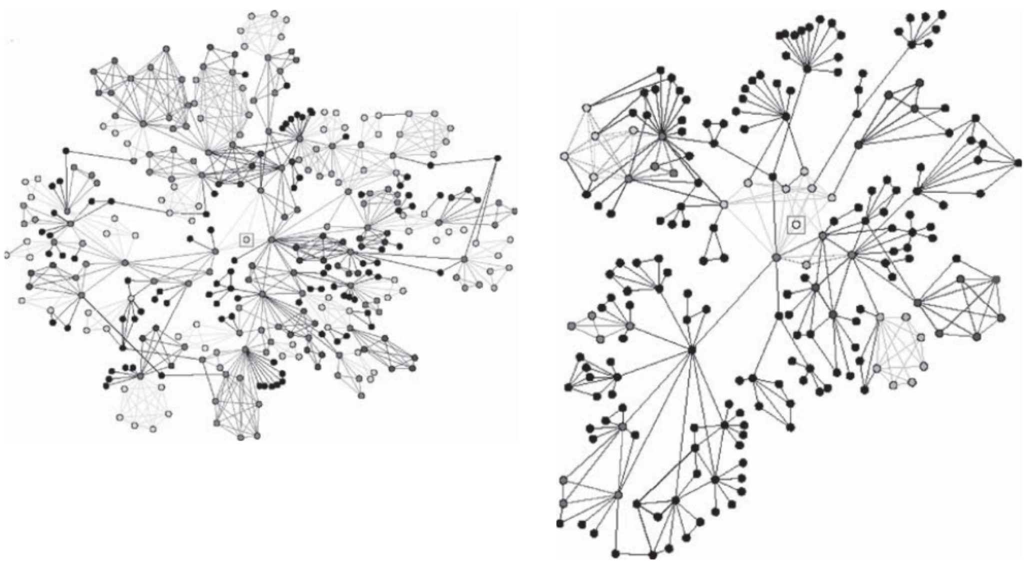
Metode prikupljanja podataka putem interneta su:

- API-ji (aplikacijsko programsko sučelje) – korištenje API-ja je usluga od internetske socijalne mreže. Šalje se zahtjev stranici da se uz pomoć API-ja prikupe podatci.
- pauk (eng. web crawlers) – metoda koja zahtijeva gradnju pauka za pretraživanje web-stranice. Pauk koristi HTTP zahtjeve/odgovore za prikupljanje podataka.
- internetski upitnici – prikupljaju se aktivnosti korisnika pitajući njih kako, kad i s kim se dijeli sadržaj.

- implemntirana aplikacija – aplikacija je postavljena na web-stranicu i tamo prikuplja podatke promatranjem ponašanja korisnika. Tako dobivamo podatke koje korisnik dijeli na internetu.

3.2 Zajednice socijalnih mreža

Zajednica unutar mreže postoji ako se čvorovi mogu grupirati u skupove (moguće je preklapanje) tako da su čvorovi unutar skupa gusto povezani međusobno. Stoga je otkrivanje zajednica u socijalnim mrežama jedan od važnijih zadataka.



Slika 3.1: Evolucija socijalne mreže na primjeru surađivanja autora (lijevo) i telefonskih poziva (desno)

Jedna od karakteristika zajednica je mogućnost evolucije kroz vrijeme, to jest one mogu rasti ili smanjivati. Grupe se mogu spajati ili razdvajati, nove zajednice mogu nastati te neke postojeće mogu nestati. Otkriveno je da su veličina i starost zajednice pozitivno korelirani što nam pokazuje da su veće zajednice u prosjeku starije. Najzanimljivije je to što ustrajnost grupa tijekom vremena jako ovisi o veličini grupe. Velike grupe traju dulje ako su sposobne mijenjati svoje članstvo na dinamički način. Međutim, male skupine imaju tendenciju da budu stabilnije ako njihov sastav ostaje nepromijenjen. Drugim riječima, manje grupe su destabilizirane i mogu nestati ako drastično mijenjaju svoj sastav s vremenom.

3.3 Centralnost u socijalnim mrežama

Postoje dvije glavne teme interesa u analizi globalnih svojstava mjera centralnosti u društvenim znanostima i šire. Prva se odnosi na robustnost različitih mjera centralnosti na pogrešne ili nepotpune podatke. U istraživanju koje je proveo Bolland ([2]), promatrana je stabilnost sljedećih centralnosti na slučajne i sistemske varijante strukture mreže: stupanjske, bliskosti, između čvorova i svojstvenih vrijednosti. Utvrđeno je da se centralnost između čvorova jako mijenja ovisno o varijanti strukture mreže. Stupanjska centralnost i centralnost bliskosti bile su najčešće stabilne, dok je centralost svojstvenih vektora bila najstabilnija za promatrane indekse.

Drugo globalno svojstvo koje se promatra za mjere centralnosti je njihova zajednička korelacija. To znači da je za različite mjere centralnosti rang čvorova znatno drugačiji. U istraživanju koje je proveo Lee ([8]) na mrežama glumaca u filmovima i suradnje između znastvenika, pokazano je da je u oba slučaja centralnost između čvorova značajno korelirana sa stupanjskom centralnošću (koeficijent korelacije je veći od 0.7), dok je centralnost svojstvenih vrijednosti i centralnost bliskosti slabo korelirana. Stupanjska centralnost je slabo korelirana s centralnosti bliskosti i centralnost svojstvenih vrijednosti su slabo korelirane s centralnosti bliskosti.

3.4 Tranzitivnost u socijalnim mrežama

Veoma važno svojstvo u socijalnim mrežama je tranzitivnost. Postoje dvije vrste tranzitivnosti koje promatramo u mrežama: savršena i parcijalna. Savršena tranzitivnost postoji unutar mreže ako vrijedi: ako je x povezan bridom s y i y povezan sa z onda je x povezan sa z. To je vrlo rijetko u stvarnim mrežama pošto to znači da je ta komponeta 3–klika. Kod parcijalne tranzitivnosti, ako x je povezan bridom s y i y je povezan sa z, nije nužno da je x povezan sa z, ali je vjerojatnije da su povezani nego s nekim nasumičnim čvorom. Naprimjer, ako x je prijatelj s y i y je prijatelj sa z, nije nužno da je x prijatelj sa z, ali je vjerojatnije da su prijatelji nego s nasumičnim članom populacije.

Socijalne mreže imaju visoku vrijednost koeficijenta tranzitivnost. Naprimjer, vrijednost tranzitivnosti mreže suradnji u izradi radova vezanih uz računalne znanosti je 0.24. To znači da dva znanstvenika koja imaju zajedničkog suradnika s kojim su napisali rad, vjerojatnost da imaju zajednički rad je skoro jedna četvrtina. Mreže koje nisu socijalne imaju malu vrijednost koeficijenta tranzitivnosti. Naprimjer, Internet ima tranzitivnost 0.012.

Poglavlje 4

Modeli slučajnih mreža

4.1 Klasične slučajne mreže

Ovaj model mreže poznat je kao Erdös–Rényijev model. U ovom modelu krećemo s n izoliranih čvorova. Nakon toga, svaki par čvorova spojen je s vjerojatnošću $p > 0$. Ovaj je model određen samo s brojem čvorova i bridova pa se označava s $G(n, m)$ ili $G(n, p)$.

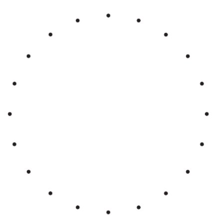
Ako se ponavlja proces više puta s istim brojem čvorova i istom vjerojatnošću, ne moramo nužno dobiti istu mrežu. Stoga te mreže promatramo kao cjelinu za dani broj čvorova i danu vjerojatnost te ih označavamo s $G_{ER}(n, m)$ ili $G_{ER}(n, p)$. U tom slučaju prosječan broj bridova $\overline{m}$ se može dobiti iz

$$\overline{m} = \frac{n(n-1)p}{2}. \tag{4.1}$$

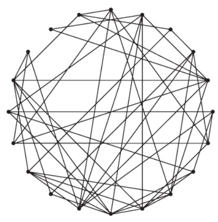
Očekivana vrijednost stupnja čvora dana je s

$$\overline{k} = (n-1)p. \tag{4.2}$$

Gustoću mreže smo definirali u (2.53), iz čega je vidljivo da je $p = \delta(G)$.



(a) p=0.0



(b) p=0.265

Slika 4.1: Primjeri Erdös–Rényijeve mreže s 20 čvorova i danim vjerojatnostima

U slučaju velikih slučajnih mreža $H \in G_{ER}(n, p)$, prosječna duljina puta dana je s

$$\bar{l}(H) = \frac{\ln n - \gamma}{\ln(pn)} + \frac{1}{2}, \tag{4.3}$$

gdje je $\gamma \approx 0.577$ Euler–Mascheronijeva konstanta. To znači da Erdös–Rényijeva mreža ima veoma malu prosječnu dužinu puta. Prosječni koeficijent klasteriranja dan je s $\overline{C} = p = \delta(G)$, što znači da je za raspršen Erdös–Rényijev model mreže koeficijent klasteriranja vrlo mali.

Kada se p povećava, većina čvorova će biti grupirana u jednu veliku komponetu, dok su ostali čvorovi izolirani u malim komponetama. Struktura $G_{ER}(n, p)$ mijenja se kao funkcija od $p = \overline{k}/(n-1)$, te postoje tri faze:

1. Podkritična $\overline{k} < 1$, gdje su sve komponete jednostavne i vrlo male. Veličina najveće komponete je $s = O(\ln n)$.
2. Kritična $\overline{k} = 1$, gdje je najveća komponeta $S = \Theta(n^{2/3})$
3. Superkritična $\overline{k} > 1$, gdje je vjerojatnost da je $(f - \varepsilon)n < S < (f + \varepsilon)n$ 1 kada $n \to \infty$ za $\varepsilon > 0$, gdje je $f = f(\overline{k})$ pozitivno rješenje jednadžbe $e^{-\overline{k}f} = 1 - f$. Ostale komponetne su vrlo male, sljedeća najveća komponeta je veličine $\ln n$.

Najveća komponeta u superkritičnoj fazi naziva se gigantska komponeta (eng. giant component).

Najveća svojstvena vrijednost matrice susjedstva u *ER* mrežama raste proporcionalno s n: $\lim_{n\to\infty}(\lambda_1(A)/n) = p$, gdje je p vjerojatnost da je svaki par čvorova spojen bridom. Druga najveća svojstvena vrijednost raste sporije od λ_1 : $\lim_{n\to\infty}(\lambda_2(A)/n^{\varepsilon}) = 0$ za svaki $\varepsilon > 0.5$ i najmanja svojstvena vrijednost raste slično kao i λ_2 : $\lim_{n\to\infty}(\lambda_n(A)/n^{\varepsilon}) = 0$ za svaki $\varepsilon > 0.5$.

U slučaju slučajnih mreža, definira se sljedeći faktor normalizacije:

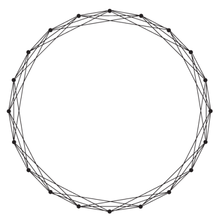
$$r = \sqrt{np(1-p)}. \tag{4.4}$$

Spektralna gustoća od *ER* mreža sljedi Wignerov polukružni zakon, koji koristi prethodno definirani faktor :

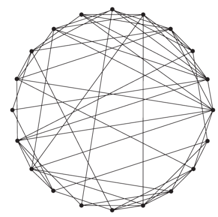
$$\rho(\lambda) = \begin{cases} \frac{\sqrt{4-\lambda^2}}{2\pi} & -2 \leq \frac{\lambda}{r} \leq 2, \\ 0 & \text{inače.} \end{cases} \tag{4.5}$$

4.2 Slučajne mreže malog svijeta

Ovaj model mreže poznat je kao Watts–Strogatzov model. Za ovaj model karakteristična su sljedeća dva svojstva: mala prosječna dužina puta i veliki koeficijent klasteriranja. Neka



(a) p=0.0



(b) p=0.5

Slika 4.2: Primjeri Watts–Strogatzov mreže s $n = 20$, $k = 6$ i danom vjerojatnosti p

je n broj čvorova i neka je k paran broj. Mreža se konstruira na sljedeći način. Svi čvorovi se postave u krug i spoji se svaki čvor s prvih $k/2$ najbližih susjeda. Ovako se dobiva prsten koji je za $k > 2$ pun trokuta i stoga je koeficijent klasteriranja velik. Prosječan koeficijent klasteriranja za ovaj model je dan s:

$$\overline{C} = \frac{3(k-2)}{4(k-1)}, \tag{4.6}$$

što znači da koeficijent klasteriranja ne ovisi o veličini mreže.

Najkraća udaljenost puta između nasuprotnih čvorova u mreži je relativno velika, te iznosi $\lceil \frac{n}{k} \rceil$. Tada je prosječna dužina puta reprezentirana s:

$$\bar{l} \approx \frac{(n-1)(n+k-1)}{2kn}. \tag{4.7}$$

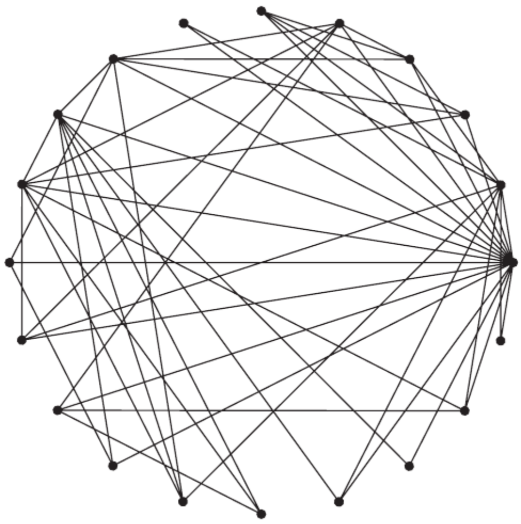
To znači da je mreža raspršena, $n >> k$. Kako raste veličina mreže prosječna dužina puta teži $n/2$, što je mnogo veća vrijednost nego što je potrebna vrijednost da je mreža malog svijeta ($\bar{l} \sim \ln n$). Watts i Strogatz su riješili ovu situaciju uvođenjem vjerojatnosti za preusmjeravanje bridova u prstenu tako da se prosječna duljina puta smanjuje, a koeficijent klasteriranja ostaje visok. Proces preusmjeravanja promatra se kao proces kojim mijenjamo brid r, s s bridom r, t s vjerojatnosti p, gdje je t nasumični čvor različit od r i s. Ako brid r, t već postoji, onda se ništa ne mijenja. Ako $p \to 1$, onda mreža teži potpunoj. Watts–Strogatzovu mrežu označavamo s $WS(n, k, p)$.

4.3 Mreže slobodne skale

Najpopularniji model mreže slobodne skale (eng. Scale-free networks) koja koristi distribuciju potencijskog oblika stupnjeva je Barabási–Albertova mreža. Konstrukciju počinjemo s malim brojem povezanih čvorova m_0. Svaki korak dodaje novi čvor u u mrežu i spaja ga s $m \leq m_0$ postojećih čvorova $v \in V$ s vjerojatnosti

$$u = \frac{k_v}{\sum\limits_{w} k_w}. \tag{4.8}$$

Mrežu se označava s $BA(n, n_0, d)$.



Slika 4.3: Primjer Barabási–Albertove mreže s $n = 20$ i $d = 4$

Karakteristika BA mreže je da je vjerojatnost da čvor ima stupanj $k \leq d$ dana s

$$p(k) = \frac{2d(d-1)}{k(k+1)(k+2)} \sim k^{-3}, \tag{4.9}$$

što daje da je kumulativna distribucija stupnja dana s $P(k) \sim k^{-2}$. Za fiksnu vrijednost $d \leq 1$, očekivana vrijednost za koeficijent klasteriranja $\overline{C}$ dana je s

$$\overline{C} \sim \frac{(d-1)\log^2 n}{8n}. \tag{4.10}$$

Prosječna dužina puta procijenjena je s

$$\bar{l} = \frac{\ln n - \ln(d/2) - 1 - \gamma}{\ln\ln n + ln(d/2)} + \frac{3}{2}, \tag{4.11}$$

gdje je γ Euler–Mascheronijeva konstanta.

Karakteristika *BA* mreže i drugih mreža koje imaju distribuciju potencijskog oblika čvorova je da spektralna gustoća slijedi distribuciju trokuta, koja se može napisati koristeći faktor normalizacije (4.4) za *ER* mreže:

$$p(\lambda) = \begin{cases} \frac{(\lambda+2)}{4} & -2 \leq \frac{\lambda}{r} \leq 0, \\ \frac{2-\lambda}{4} & 0 \leq \frac{\lambda}{r} \leq 2, \\ 0 & \text{inače.} \end{cases} \tag{4.12}$$

4.4 Slučajne geometrijske mreže

Slučajna geometrijska mreža veličine n konstruira se postavljanjem n čvorova u jednični kvadrat u kojem čvor i ima koordinate (x_i, y_i). Koordinate su nezavisne i raspoređene po uniformnoj distribuciji. Drugi korak konstrukcije je spajanje parova čvorova čija je udaljenost manja od zadane euklidske udaljenosti. Neka je r zadani radijus, tada su čvorovi i i j povezani ako vrijedi

$$(x_i - x_j)^2 + (y_i - y_j)^2 \leq r^2. \tag{4.13}$$

(a) Prvi korak

(b) Drugi korak

Slika 4.4: Primjer konstrukcije slučajne geometrijske mreže s 100 čvorova i radijusom $r = 0.15$

Karakteristika ove mreže je postojanje prijelaza iz nepovezane mreže u povezanu za neku vrijednost radijusa. Vrijednost radijusa za koju je mreža povezna s velikom vjerojatnošću naziva se kritični radijus r_C te se može izraziti kao

$$r_C = \frac{\sqrt{\ln n + O(1)}}{\pi n}. \tag{4.14}$$

Generalizacija ovog modela je slučajno udaljena mreža. U ovom slučaju, n čvorova je postavljeno uniformno nasumično na jedinicu diska. Nakon toga, brid postoji između čvorova i i j ako su udaljeni d_{ij}. Takvi bridovi su postavljeni neovisno o drugim s vjerojatnošću $p_{ij} = g(d_{ij})$. To se može promatrati kao step funkciju s granicom r, koja kreira brid s vjerojatnošću $\alpha = \alpha(n)$ za čvorove na udaljenosti $d_{ij} \leq r$, i s vjerojatnosti $\beta = \beta(n)$ za čvorove na udaljenosti $d_{ij} > r$. Poželjno je birati $\beta(\alpha)$ kao funkciju od α i

$$(1-\alpha)\pi r^2 = \beta(1-\pi r^2) \tag{4.15}$$

kako bi prosječan stupanj ostao u vezi s α. Tada, za $\pi r^2 \leq \alpha \leq 1$ imamo

$$g_r^\alpha = \begin{cases} \alpha & \text{ako } d \leq r, \\ \beta = \frac{(1-\alpha)\pi r^2}{1-\pi r^2} & \text{ako } d > r. \end{cases} \tag{4.16}$$

Slučajnu udaljenu mrežu označavamo s $D(n, g_r^\alpha)$. Ona ima sljedeću karakteristiku povezanosti. Neka je

$$\pi r^2 = \frac{\log n + c_n}{n}. \tag{4.17}$$

Mreža je povezana s velikom vjerovatnosti ako $c_n \to +\infty$ i nepovezana ako $c_n \to -\infty$.

Za povezane mreže, pronađena su sljedeća svojstva:

$$\overline{k} = \Theta(\log n) \tag{4.18}$$

$$\overline{C} = 0.5865\alpha + o(1) \text{ za } \pi r^2 \leq \alpha \leq 1 \tag{4.19}$$

$$l_{\max} = \theta\left(\frac{\log n}{\log\log n}\right) \text{ za } \pi r^2 \leq \alpha \leq 1-\varepsilon, \varepsilon > 0. \tag{4.20}$$

4.5 Modeli eksponencijalne slučajne mreže

Modeli eksponencijalne slučajne mreže (eng. Exponential random graph models, ERGM) familija su statističkih modela za analizu podataka vezanih uz mreže. Neka je slučajna mreža Y skup od n čvorova i m bridova $\{Y_{ij} : i = 1, \ldots, n; j = 1, \ldots, n\}$ gdje je $Y_{ij} = 1$ ako postoji brid između i i j, a inače je $Y_{ij} = 0$. Modele eksponencijalne slučajne mreže možemo zapisati u sljedećem obliku:

$$P(Y = y) = \frac{1}{\kappa}\exp\left\{\sum_A \eta_A g_A(y)\right\}, \tag{4.21}$$

gdje je A skup svih konfiguracija, η_A je parametar koji se povezan s konfiguracijama A (nije nula ako je za sve parove varijabli iz A pretpostavljeno da su uvjetno zavisne), $g_A(y) = \prod_{y_{ij}\in A} y_{ij}$ je mreža koja statistički odgovara konfiguraciji A ($g_A(y) = 1$ ako je konfiguracija promatrana u mreži y, a 0 inače) i κ je parametar normaliziranja koji osigurava da je (4.21) valjana distribucija. Svi modeli eksponencijalne slučajne mreže su oblika (4.21) koji opisuje generalnu distribuciju mreže s n čvorova. Vjerojatnost bilo koje promatrane mreže y u ovoj distribuciji dana je s tom jednadžbom i vjerojatnost ovisi o $g_A(y)$ u mreži y i raznovrsnosti ne–nula parametra η_A za sve konfiguracije od A u modelu.

Konstrukcija modela se generalno provodi kroz sljedećih pet koraka:

1. svaki brid mreže promatra se kao slučajna varijabla,
2. hipoteze zavisnosti su predložene, definiraju se nepredviđenosti između varijabli mreže,
3. hipoteza zavisnosti daje poseban oblik modela,
4. pojednostavljuje se model kroz homogenost ili drugu konstantu,
5. procjenjuju se i interpretiraju parametri modela.

Za procjenu parametara modela koriste se procjene pseudo–vjerojatnosti, maksimalne vjerojatnost ili Monte Carlova maksimalne vjerojatnost.

Vrste modela obzirom na pretpostavke o ovisnosti

Postoji nekoliko vrsta modela eksponencijalne slučajne mreže obzirom na pretpostavke o ovisnosti.

Bernulijeva mreža

Distribucija Bernulijeve slučajne mreže je generirana pod pretpostavkom da su bridovi nezavisni, te da su svi mogući jedinstveni bridovi nezavisni jedan od drugog. Tada je jedina moguća konfiguracija povezana s bridovima $\{Y_{ij}\}$ te se generalni model može zapisati kao:

$$P(Y=y)=\frac{1}{\kappa}\exp\left(\sum_{i,j}\eta_{ij}g_{ij}\right). \tag{4.22}$$

Uzme li se u obzir pretpostavka homogenosti tako da se izjednači parametar kao $\eta_{ij}=\Theta$ za sve i i j, tada je

$$P(Y=y)=\frac{1}{\kappa}\exp\left(\Theta L(y)\right), \tag{4.23}$$

gdje je $L(y)=\sum_{ij}y_{ij}$ broj bridova u mreži y, a parametar Θ je povezan s vjerojatnosti da se brid promatra. Parametar Θ se naziva parametar brida ili parametar gustoće.

Binarni model

U ovom modelu za usmjerene mreže pretpostavlja se da su parovi bridova nezavisni jedni od drugih. S pretpostavkom homogenosti, model se može zapisati kao:

$$P(Y=y)=\frac{1}{\kappa}\exp\left(\Theta\sum_{ij}y_{ij}+\rho\sum_{ij}y_{ij}y_{ji}\right)=\frac{1}{\kappa}\exp(\theta L(y)+\rho M(y)), \tag{4.24}$$

gdje je $L(y)$ broj jednosmjernih bridova u y i $M(y)$ je broj dvosmjernih bridova u y.

Markovljeva slučajna mreža

Kod ovog modela koristi se Markovljeva ovisnost. U njoj se pretpostavlja da postojanje brida između i i j ovisi o postojanju svih bridova koji uključuju i ili j. Tada su bridovi uvjetno ovisni.

Generalni model ovisi što se promatra u danom modelu (brid, 2–zvijezde , 3–zvijezde, trokute, ...), te jesu li mreže usmjerene ili ne. Naprimjer, za neusmjerenu mrežu gdje se promatraju bridovi, 2–zvijezde, 3–zvijezde i trokute, formula je dana s

$$P(Y = y) = \frac{1}{\kappa} \exp(\theta L(y) + \sigma_2 S_2(y) + \sigma_3 S_3(y) + \tau T(y)), \tag{4.25}$$

gdje su $S_2(y)$ i $S_3(y)$ broj 2–zvijezdi i 3–zvijezdi u mreži y, a $T(y)$ je broj trokuta u y.

Poglavlje 5

Eksperimentalni rezultati

U ovom poglavlju provedeno je istraživanje o mogućnosti kreiranja slučajne mreže koja bi bila slična socijalnoj mreži. Za potrebe istraživanja korišten je paket NetworkX, biblioteka pisana u programskom jeziku Python.

5.1 NetworkX

NetworkX je Pythonov paket za stvaranje, manipuliranje i proučavanje strukture, dinamike i funkcionalnosti složenih mreža . Prva verzija je izašla 2005. godine, a trenutna verzija je 2.2. Ovaj paket sadrži:

- strukture podataka za grafove, usmjeren grafove i multigrafove,
- standardne algoritme za grafove,
- strukture mreže i mjere analize,
- generatore za klasične grafove, slučajne grafove i sintetske grafove,
- čvorovi mogu biti bilo što (tekst, slike, XML zapisi),
- bridovi mogu sadržavati podatke (težine, vremeski zapis).

NetworkX omogućuje pohranu, učitavanje standardnih i nestandarnih tipova podataka, analizu strukture mreže, izradu modela mreže te dizajn novih algoritama vezanih za mreže. Pogodan je za operacije na velikim mrežama (do 10 milijuna čvorova i 100 milijuna bridova).

Paket učitavamo s `import networkx as nx`. Sada navodimo funkcije koje ćemo koristiti u eksperimentu:

- `nx.adjacency_matrix(G)` – funkcija koja vraća matricu susjedstva za mrežu G,
- `nx.incidence_matrix(G)` – funkcija koja vraća matricu incidencije za mrežu G,
- `nx.degree_assortativity_coefficient(G)` – funkcija koja vraća koeficijent međupoloženosti za mrežu G,
- `nx.laplacian_matrix(G)` – funkcija koja vraća Laplaceovu matricu za mrežu G,
- `nx.eccentricity(G)` – funkcija koja vraća ekscentritet za svaki čvor mreže G,
- `nx.diameter(G)` – funkcija koja vraća radijus mreže G,
- `nx.radius(G)` – funkcija koja vraća radijus mreže G,
- `nx.center(G)` – funkcija koja vraća listu čvorova koji su centar mreže G,
- `nx.average_shortest_path_length(G)` – funkcija koja vraća najkraću prosječnu dužinu puta u mreži G,
- `nx.clustering(G)` – funkcija koja vraća koeficijent klasteriranja mreže G,
- `nx.average_clustering(G)` – funkcija koja vraća prosječnu vrijednost klasteriranja mreže G,
- `nx.transitivity(G)` – funkcija koja vraća tranzitivnost mreže G,
- `nx.degree_centrality(G)` – funkcija koja vraća vrijednost stupanjske centralnosti za svaki čvor mreže G,
- `nx.katz_centrality_numpy(G)` – funkcija koja vraća vrijednost Katz centralnosti za svaki čvor mreže G,
- `nx.eigenvector_centrality(G)` – funkcija koja vraća vrijednost centralnost svojstvenih vrijednosti za svaki čvor mreže G,
- `nx.subgraph_centrality(G)` – funkcija koja vraća vrijednost centralnosti podmreže za svaki čvor mreže G,
- `nx.closeness_centrality(G)` – funkcija koja vraća vrijednost centralnosti bliskosti za svaki čvor mreže G,
- `nx.betweenness_centrality(G)` – funkcija koja vraća vrijednost stupanjske centralnosti za svaki čvor mreže G,
- `nx.density(G)` – funkcija koja vraća gustoću mreže G,

- `nx.global_efficiency(G)` – funkcija koja vraća globalni indeks efikasnosti mreže G,
- `nx.erdos_renyi_graph(n,p)` – funkcija koja vraća Erdös–Rénijev graf s n čvorova gdje je vjerojatnosti postojanja brida p,
- `nx.watts_strogatz_graph(n,k,p)` – funkcija koja vraća Watts–Strogatzov graf s n čvorova s početno spojenih $k/2$ susjedni čvorovima gdje je vjerojatnosti preusmjeravanja brida p,
- `nx.barabasi_albert_graph(n,k,p)` – funkcija koja vraća Barabaási–Albertov graf s n čvorova i m početnih čvorova na koje se novi čvorovi povezuju.

5.2 Podaci socijalnih mreža

Za potrebe ovog rada, podaci socijalnih mreža preuzeti su sa stranice sveučilišta Stanford ([9]). Podaci su o interakcijama ljudi unutar socijalnih krugova koji se nalaze na Facebooku, Twitteru ili Google+. Podaci su zapisani u datotekama tipa *edges*, to jest u svatkom retku nalaze se dva čvora koji su prikazani brojevima. Čvorovi koji se nalaze u istom redu povezani su u grafu.

U tablicama 5.1, 5.2 i 5.3 navodi se nekoliko mreža te njhove karakteristke (broj čvorova i bridova) te vrijednosti parametara gustoće, radiusa, globalnog indeksa efikasnosti, prosječne duljine puta, prosječnog koeficijenta klasteriranja i tranzitivnost.

Naziv	broj čvorova	broj bridova	gustoća	radius	globalni indeks efikasnosti	prosječna duljina puta	prosječni koeficijent klasteriranja	tranzitivnost
fb-0	324	2514	0.04804	6	0.33847	3.75274	0.52236	0.42587
fb-348	224	3192	0.1278	5	0.47835	2.52346	0.54428	0.49027
fb-414	148	1692	0.15554	4	0.46584	2.69157	0.67935	0.64579
fb-686	168	1656	0.11805	3	0.4795	2.42507	0.53379	0.45356
fb-1684	775	14006	0.04669	5	0.37924	3.04226	0.47138	0.45226
fb-1912	744	30023	0.10862	4	0.45631	2.55835	0.63797	0.70002
fb-3437	532	4812	0.03406	5	0.33525	3.44745	0.54576	0.44889

Tablica 5.1: Vrijednosti parametara socijalnih mreža s Facebooka

U tablicama 5.4, 5.5 i 5.6 nalaze se podaci o centralnosti mreža. Centralnost je zapisana na način da je uzeta maksimalna vrijednost od dane centralosti (koliko je značajan najznačajniji čvor). Promatrana je stupanjska centralnost, Katz centralnost, centralnost svoj-

Naziv	broj čvorova	broj bridova	gustoća	radius	globalni indeks efikasnosti	prosječna duljina puta	prosjenčni koeficijent klasteriranja	tranzitivnost
gp-3	94	1005	0.22992	3	0.59199	1.9135	0.6111	0.51529
gp-6	486	15163	0.12865	3	0.5353	2.0545	0.54797	0.39317
gp-7	268	5398	0.15087	3	0.56213	1.93026	0.63864	0.3531
gp-10	1052	62088	0.11231	3	0.50282	2.26334	0.58172	0.39112
gp-11	780	21662	0.0713	3	0.46862	2.35395	0.59308	0.37286
gp-13	559	26755	0.17154	3	0.55176	2.04262	0.6389	0.55384
gp-15	342	3361	0.05763	3	0.39683	2.87756	0.54942	0.44664

Tablica 5.2: Vrijednosti parametara socijalnih mreža s Google Plusa

Naziv	broj čvorova	broj bridova	gustoća	radius	globalni indeks efikasnosti	prosječna duljina puta	prosječni koeficijent klasteriranja	tranzitivnost
tw-1	88	704	0.1839	3	0.54786	2.10815	0.53595	0.46186
tw-2	101	430	0.08514	3	0.45137	2.51326	0.39347	0.31589
tw-3	159	796	0.06337	4	0.39576	2.93567	0.46635	0.41733
tw-4	57	577	0.36152	2	0.68065	1.63909	0.67071	0.53258
tw-5	247	5629	0.18528	3	0.55843	2.02916	0.50849	0.48288
tw-6	240	3501	0.12207	3	0.51248	2.17343	0.49471	0.44961
tw-7	242	4372	0.14992	4	0.50223	2.35088	0.47984	0.53877
tw-8	235	2612	0.09499	3	0.47411	2.37217	0.43814	0.42256

Tablica 5.3: Vrijednosti parametara socijalnih mreža s Twittera

stvenih vektora, centralnost bliskosti, centralnost između čvorova, centralnost toka između čvorova i centralnost komunikabilnosti.

5.3 Kreiranje slučajnih mreža

Iz tablica je vidljivo da socijalne mreže imaju slične vrijednosti parametara u slučaju gustoće, prosječne udaljenosti puta, koeficijenta klasteriranja, tranzitivnosti, globalnog indeksa efikasnosti, centralnosti svojstvenih vektora, Katz centralnosti i centralnosti između čvorova. U tablicama 5.7 i 5.8 prikazane su prosječne vrijednosti i standardne devijacije parametara koje su promatrane. Pri kreiranju slučajnih mreža promatra se gustoća, prosječna udaljenost puta, tranzitivnost, globalni indeks efikasnosti i Katz centralnost. Koeficijent

Naziv	stupanjaska	Katz	svojstvenih vrijednosti	bliskosti	između čvorova	tok između čvorova	komunikabilnosti
fb-0	0.23839	0.2114	0.19814	0.425	0.2809	0.30517	0.72328
fb-348	0.44394	0.23102	0.16411	0.57179	0.1012	0.09383	0.73704
fb-414	0.38775	0.30879	0.19335	0.52127	0.23233	0.17965	0.70221
fb-686	0.46107	0.28983	0.21442	0.60071	0.13741	0.12338	0.78949
fb-1684	0.17571	0.21209	0.14463	0.45	0.08296	0.1037	0.51755
fb-1912	0.39434	0.13354	0.08751	0.55118	0.09103	0.07613	0.72187
fb-3437	0.20150	0.15144	0.19911	0.42142	0.21501	0.15388	0.75671

Tablica 5.4: Vrijednosti centralosti socijalnih mreža s Facebooka

,

Naziv	stupanjaska	Katz	svojstvenih vrijednosti	bliskosti	između čvorova	tok između čvorova	komunikabilnosti
gp-3	0.78494	0.33104	0.22091	0.80869	0.14615	0.13918	0.81899
gp-6	0.68659	0.14246	0.13292	0.74615	0.0495	0.03723	0.86858
gp-7	0.88014	0.2227	0.16821	0.88704	0.13651	0.10841	0.88784
gp-10	0.62511	0.15942	0.09118	0.65482	0.10698	0.08381	0.85657
gp-11	0.41848	0.13192	0.15089	0.59693	0.07749	0.04391	0.87645
gp-13	0.72401	0.1593	0.09347	0.77178	0.06361	0.05486	0.79279
gp-15	0.29325	0.24505	0.18393	0.49926	0.16678	0.16943	0.77735

Tablica 5.5: Vrijednosti centralosti socijalnih mreža s Google Plusa

Naziv	stupanjaska	Katz	svojstvenih vrijednosti	bliskosti	između čvorova	tok između čvorova	komunikabilnosti
tw-1	0.60919	0.3014	0.24495	0.67968	0.14991	0.13438	0.77535
tw-2	0.38	0.33096	0.29854	0.58479	0.17561	0.18011	0.73523
tw-3	0.23417	0.25972	0.262	0.51633	0.17141	0.19009	0.51373
tw-4	0.98214	0.2803	0.26242	0.98245	0.1358	0.12335	0.84067
tw-5	0.61788	0.18566	0.14234	0.71098	0.07157	0.06828	0.76468
tw-6	0.41004	0.17923	0.1665	0.62077	0.09811	0.07755	0.71964
tw-7	0.53112	0.22694	0.14633	0.6025	0.0486	0.0668	0.76237
tw-8	0.39316	0.14553	0.18466	0.60465	0.16985	0.12692	0.74945

Tablica 5.6: Vrijednosti centralosti socijalnih mreža s Twittera

klasteriranja ne promatra se pošto je on sličan tranzitivnosti. Katz centralnost je odabrana zato što je stavljen fokus istraživanja na procjenu utjecaja aktera u socijalnoj mreži.

Parametar	Prosjek	Standardna deviacija
gustoća	0.138853738110286	0.078766301523065
prosječna duljina puta	2.33576568030419	0.405966774875674
prosječni koeficijent klasteriranja	0.538951415263339	0.084097392387576
tranzitivnost	0.44211462618335	0.076794975082009
globalni indeks efikasnosti	0.497375562510638	0.08473982308165

Tablica 5.7: Prosjeci i standardne devijacije dijela parametara socijalnih mreža

Centralnost	Prosjek	Standardna devijacija
Katz	0.21240907848737	0.05877855262672
Svojstvenih vektora	0.17269646368758	0.051234007425969
Bliskosti	0.651269623781551	0.142124853120923
Između čvorova	0.119369798054343	0.057507234813576
Toka između čvorova	0.107565654061034	0.058450418775531
Komunikabilnosi između čvorova	0.761126241151368	0.091685195776827

Tablica 5.8: Prosjeci i standardne devijacije mjera centralosti socijalnih mreža

Dio podataka koji nije iskorišten za određivanje parametra koji se koriste za kreiranje slučajnih mreža, poslužiti će kao testni podaci za usporedbu socijalne mreže sa slučajnom mrežom. U tablici 5.9 prikazane su vrijednosti parametara testnih mreža. Broj čvorova i broj bridova testne mreže iskoristit će se za kreiranje slučajne mreže. Vjerojatnost p, koju koriste neki modeli slučajnih mreža, zadavat ćemo kao omjer broja bridova testne mreže i broja bridova potpune mreže zadane s brojem čvorova testne mreže. Kako bi se prikazala sličnost slučajne mreže socijalnoj mreži, izračunat će se parametri slučajne mreže i uvidjeti jesu li oni unutar intervala određenog s vrijednostima parametara testne mreže i standardne devijacije iz tablica 5.7 i 5.8.

U slučaju Erdös–Renijevog i Barabási–Albertovog modela nije bilo uspjeha. Kod ovih modela mreže, problem je u tome što novonastale mreže imaju malu tranzitivnost, a za socijalne mreže je karakteristično da imaju visoku tranzitivnost. Istraživanje nije dalo uspjeha niti kod modela eksponencijalne slučajne mreže. Ovaj model nije implementiran u paketu NetworkX, pa je korištena implementacija koja se može pronaći na [1]. Ulazna

Red.br.	broj čvorova	broj bridova	gustoća	prosječna duljina puta	tranzitivnost	globalni indeks efikasnosti	Katz centralnost
1	136	2105	0.2293	1.8775	0.52596	0.597	0.26052
2	197	2632	0.13633	2.24551	0.49599	0.50807	0.19362
3	149	1590	0.1442	2.15182	0.46271	0.5249	0.1707
4	117	593	0.08738	2.38829	0.31169	0.46863	0.34864
5	128	1822	0.22416	1.88631	0.5974	0.59368	0.27443
6	184	2126	0.12627	2.32186	0.52209	0.49377	0.214446
7	133	1201	0.13681	2.38915	0.45710	0.49319	0.249064
8	134	1170	0.13129	2.43272	0.52525	0.48697	0.21183
9	102	664	0.1289	2.09551	0.3337	0.52769	0.185271

Tablica 5.9: Vrijednosti parametara testnih socijalnih mreža

vrijednost je broj čvorova n te se promatra kroz zadane parametre provjeravajući njihovu konvergenciju. Kada zadani parametri prestanu konvergirati program se završava. Promatran je Markovljev model slučajne mreže modela s 2–zvijezde i trokuta. U tablici 5.10 prikazane su vrijednosti parametara kreiranih mreža. Stupac i označava broj iteracija za koje je računanje parametara konvergiralo. Ovdje je uočen problem male tranzitivnosti i velike prosječne dužine puta.

naziv	i	broj čvorova	broj bridova	gustoća	prosječna duljina puta	tranzitivnost	globalni indeks efikasnosti	Katz centralnost
ergm_80	149	77	147	0.05023	3.33185	0.05222	0.34871	0.26026
ergm_100	165	99	163	0.0336	4.1746	0.03734	0.2868	0.22802
ergm_120	177	111	175	0.02866	4.26732	0.02357	0.27663	0.24536
ergm_140	182	118	177	0.02564	4.47921	0.02268	0.26517	0.23661
ergm_160	201	139	198	0.02064	5.07183	0.006	0.23631	0.25717
ergm_180	212	157	209	0.01706	5.69622	0.01829	0.21326	0.243991

Tablica 5.10: Vrijednosti parametara kreiranih ERGM mreža

Watts–Strogatzov model mreže

Ovaj model zahtijeva dodatni parametar k ($k/2$ najbližih susjeda je spojeno). Vrijednost parametara k nije zadana, nego je itrerirano po njemu i provjeravani su parametri novonastale mreže za dani k u unaprijed definiranim granicama. Zapisan je prvi k za koji slučajna mreža prikazuje sličnosti socijalne mreže. Primjenom ovog modela ostvareni su pozitivni rezultati istraživanja. Model je rezultirao stvaranjem slučajne mreže slične socijalnoj mreži. Pozitivan rezultat istraživanja prikazan je u tablici 5.11 gdje su prikazane dobivene vrijednosti parametara.

Red. br.	broj čvorova	k	vjerojatnost	broj bridova	gustoća	prosječna duljina puta	tranzitivnost	globalni indeks efikasnosti	Katz centralnost
1	136	14	0.21949	952	0.1037	2.33671	0.37589	0.47867	0.25581
2	197	14	0.13633	1379	0.07142	2.6542	0.45748	0.42018	0.19334
3	149	12	0.1442	894	0.08108	2.62298	0.4302	0.42936	0.20632
4	117	14	0.08738	819	0.12068	2.42219	0.51347	0.47105	0.20311
5	128	17	0.22416	1024	0.12598	2.19611	0.37533	0.50931	0.23341
6	184	14	0.12627	1288	0.0765	2.62782	0.47472	0.4255	0.25278
7	133	12	0.13681	798	0.0909	2.57883	0.46195	0.4384	0.22991
8	134	12	0.13129	804	0.09022	2.59824	0.46205	0.43625	0.25596
9	102	15	0.1289	714	0.13861	2.29256	0.46478	0.49779	0.24346

Tablica 5.11: Vrijednosti parametara kreiranih Watts–Strogazovih mreža

Bibliografija

[1] J. Atwood, *Exponential random graph model*, (2014), https://github.com/jcatw/ergm.

[2] J. M. Bolland, *Sorting out centrality: An analysis of the performance of four centrality models in real and simulated networks*, (1988), https://www.sciencedirect.com/science/article/pii/0378873388900147.

[3] K. Erciyes, *Complex Networks, An Algorithmic Perspective*, CRC Press, 2015.

[4] E. Estrada, *The Structure of Complex Networks, Theory and Applications*, Oxford University Press, 2011.

[5] M. Franceschet, *Transitivity*, (2014), https://www.sci.unich.it/~francesc/teaching/network/transitivity.html.

[6] A. Golemac, A. Mimica i T. Vučičić, *Od koenigsberških mostova do kineskog poštara*, (2012), http://e.math.hr/category/klju-ne-rije-i/teorija-grafova.

[7] D. Klein i M. Randić, *Centrality Measures for Bibliometric Network Analysis*, (1993), https://www.researchgate.net/publication/226420782_Resistance_Distance.

[8] J. Y. Lee, *Centrality Measures for Bibliometric Network Analysis*, (2006), https://www.researchgate.net/publication/263644452_Centrality_Measures_for_Bibliometric_Network_Analysis.

[9] J. Leskovec, *SNAP Datasets: Stanford Large Network Dataset Collection*, (2014), http://snap.stanford.edu/data.

[10] I. Nakić, *Diskretna matematika*, 2011, https://web.math.pmf.unizg.hr/nastava/komb/predavanja/predavanja.pdf.

[11] M.E.J. Newman, *The Structure and Function of Complex Networks*, Oxford University Press, 2003.

[12] ———, *Networks, An Introduction*, Oxford University Press, 2010.

[13] G. Robins, P. Pattison, Y. Kalish i D. Lusher, *An introduction to exponential random graph (p^*) models for social networks*, (2017), https://www.sciencedirect.com/science/article/pii/S0378873306000372.

[14] M. Zuhair Al-Taie i S. Kadry, *Python for Graph and Network Analysis*, Springer, 2017.

Sažetak

Rad prikazuje osnovne teorije mreža i parametre koji opisuju mreže, opis socijalnih mreža, načine prikupljanja i obrade podataka socijalnih mreža, modele slučajnih mreža te istraživanje o mogućnosti stvaranja sličnosti slučajne mreže slične socijalne mreže putem raznih modela mreže slučaja.

Mreža je pojednostavljeni prikaz sustava koji se sastoji od međusobno povezanih elemenata. Naprimjer, mreža može simulirati povezanost gradova, atome u molekulama, odnos između ljudi/organizacija i slično. U socijalnim mrežama, čvorovi predstavljaju ljude ili grupe ljudi, a bridovi neki oblik interakcije između njih. Za prikupljanja podataka socijalnih mreža koriste se API–ji (aplikacijsko programsko sučelje), web crawleri (pauk), …

U radu je provedeno istraživanje pomoću paketa NetworkX napisanog u programskom jeziku Python gdje su proučavani modeli kreiranja slučajne mreže sa sličnostima socijalne mreže. Ispitivani su sljedeći modeli slučajne mreže: Erdös–Renijev, Barabási–Albertov, Watts–Strogatzov i eksponecijalne slučajne mreže. Provjera sličnosti između mreža napravljena je provjerom jesu li paramteri slučajne mreže u zadnim vrijednostima parametara koji su dobiveni iz podataka socijalnih mreža (Facebook, Google Plus, Twitter) sa stranice sveučilišta Stanford ([9]). Korišteni parametri su gustoća, prosječna udaljenost puta, tranzitivnost, globalni indeks efikasnosti i Katz centralnost. Negativan rezultat istaživanja ostvarili su Erdös–Renijevi i Barabási–Albertovi modeli slučajne mreže zbog male tranzitivnosti, dok je kod eksponecijalne slučajne mreže bio problem i prosječna duljina puta. Pozitivan rezultat istraživanja je dobiven kod Watts–Strogatzove slučajne mreže.

Summary

The paper presents basic network theory as well as parameters describing networks, social networks, ways of collecting and processing social network data, random network models, and exploring the probability of creating a random network similar to a social network through various network case models.

The network is a simplified representation of a system consisting of interconnected elements. For example, the network can simulate the association of cities, atoms in molecules, relationship between people/organizations and similar. In social networks, nodes represent people or groups of people and bridges are some form of interaction between them. To collect social network data, use APIs (application program interface), web crawlers, ...

In this paper, research was conducted using NetworkX package written in Python programming language, where models of random networking with similarities in the social network were studied. The following random network models were investigated: Erdös–Renije, Barabási–Albert, Watts–Strogatz and exponential random graph. Verification of network similarities was connected by verifying whether random network parameters have the default values of the parameters obtained from the social network data (Facebook, Google Plus, Twitter) from the Stanford University website ([9]). The parameters used are density, average shortest path length, transitivity, global efficiency index and Katz centrality. Negative result of the research was achieved by Erdös–Renije and Barabási–Albert models of random network due to low transitivity, while in exponential random network the problem was the average length of the path. A positive result of the research was obtained for Watts–Strogatz's random network.

Životopis

Rođen sam 5. travnja 1993. godine u Bjelovaru. Pohađao sam Osnovnu školu Garešnica u Garešnici. Nakon toga upisujem Tehničku školu Bjelovar, smjer tehničar za računarstvo. Godine 2011. upisujem preddiplomski sveučilišni studij Matematike na Prirodoslovno–matematičkom fakultetu u Zagrebu. Njega završavam 2015. godine te iste godine upisujem smjer Računarstvo i matematika.