

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

**SEMINAR**

**Osnovna svojstva kompleksnih mreža i njihova  
primjena**

*Đani Glavinić 1.02*

*Voditelj: Mr.sc. Mile Šikić*

Zagreb, 05, 2007.

## Sadržaj

1. Uvod .....	1
2. Uvod u teoriju grafova .....	2
2.1 Osnovni pojmovi teorije grafova.....	2
2.2 Postupak rješavanja grafova.....	5
3. Clustering .....	8
4. Distribucija stupnjeva čvorova ili raspodjela veza .....	9
5. Preferirano povezivanje.....	<b>Error! Bookmark not defined.</b> 11
6. Objašnjenje srednjeg najkraćeg puta .....	<b>Error! Bookmark not defined.</b> 14
7. Metode .....	16
8. Procjena maksimalne vjerojatnosti .....	17
9. Metoda najmanjih kvadrata .....	18
10. Primjena kompleksnih mreža.....	19
11. Zaključak.....	20
12. Sadržaj.....	21
13. Literatura.....	22

## 1. Uvod

Zbog različitih struktura koje se pojavljuju u kompleksnim mrežama, javljaju se i različite raspodjele veza između čvorova. Proučavanjem realnih kompleksnih mreža otkriveno je da u većini slučajeva one imaju Pareto ili Pareto raspodjelu s eksponencijalnim repom, dok uniformno slučajne mreže imaju Poissonovu ili eksponencijalnu raspodjelu. Da bi se izvršila analiza kompleksne mreže potrebno je razviti alat kojim bi se prvo utvrdila raspodjela veza između čvorova, a na temelju koje se kasnije određuju koja osnovna svojstva ima promatrana mreža.

U prvom poglavlju se govori o uvodu u teoriju grafova i njihovim svojstvima. Dalje govorimo o nekim osnovnim metodama i pojmovima grafova te pregledu osnovnih svojstava koja se pojavljuju u svim kompleksnim mrežama. U poglavlju metode opisane su metode korištene za određivanje parametara pretpostavljenih modela raspodjela veza.

## 2. Uvod u teoriju grafova

### 2.1 Osnovni pojmovi teorije grafova

Grafovi su jedno od osnovnih matematičkih struktura. Stoga se pojavljuju u raznim oblicima i raznim situacijama. Mnoge se pojave modeliraju grafovima (dijagramima) koji se sastoje od točkica i njihovih spojnica. Na primjer, točke (vrhovi ili čvorovi) mogu predstavljati ljude iz neke skupine, a spojnice (bridovi) parove prijatelja, ili točke mogu predstavljati komunikacijske centre, a spojnice komunikacijske veze. Graf može predstavljati električnu mrežu, čiji su vrhovi električne komponente, a spojnice električne veze. Cestovne, željezničke, zrakoplovne veze itd. daljnji su primjeri modela sa grafovima. U računarstvu se često dijagram toka nekog algoritma pokazuje grafom kojem se čvorovi naredbe (instrukcije), a lukovi iz jedne u drugu naredbu su bridovi. Isto se tako grafovima prezentiraju i razne kompjutorske strukture podataka, umrežavanje i paralelizam računala i njihov sekvencijalni rad, evolucijska ili porodična stabla u biologiji, kemijske veze među atomima ili molekulama, raspored poslova u velikim gospodarskim projektima itd.

Neki od osnovnih pojmova u vezi grafova:

**Graf** je uređeni par  $G = (V, E)$ , gdje je  $0 \neq V = V(G)$  skup **vrhova** (eng. vertex),  $E = E(G)$  skup **bridova** (eng. edge) disjunktin s  $V$ , a svaki brid  $e \in E$  spaja dva vrha  $u, v \in V$  koji se zovu krajevi od  $e$ . kažemo još tada da su vrhovi  $u$  i  $v$  **susjedni** i pišemo  $e = uv$  (ili pravilno  $e = \{u, v\}$ ). Bridove sa barem jednim zajedničkim krajem također zovemo incidentnim. Grafovi  $G$  i  $H$  su **izomorfni**, i pišemo  $G \approx H$  ako postoje bijekcije  $\Theta : V(G) \rightarrow V(H)$  i  $\varphi : E(G) \rightarrow E(H)$  tako da je vrh „ $v$ “ incidentan s bridom „ $e$ “ u „ $G$ “ ako i samo ako je  $\Theta(v)$  incidentan sa  $\varphi(e)$  u „ $H$ “. Uređeni par  $f = (\Theta, \varphi) : G \rightarrow H$  se tada zove **izomorfizam** iz „ $G$ “ u „ $H$ “.

Izomorfizam, dakle, čuva incidenciju i susjednost. Brid čiji se krajevi podudaraju zove se **petlja**, a ako su krajevi različiti – **pravi brid** ili **karika**. Dva brida ili više njih sa istim parom krajeva zovu se **višestruki bridovi**. Graf  $G$  je jednostavan ako nema ni petlja ni višestrukih bridova. Graf sa samo jednim vrhom zove se **trivijalan**, a inače **netrivijalan**.  $G$  je prazan graf ako je  $E(G) = 0$ . Ako se ne

kažemo drukčije, mi isključivo proučavamo konačne grafove. Dva osnovna parametra vezana uz osnovni graf su:

$$v(G) = |V(G)| = \text{red od } G = \text{broj vrhova od } G,$$

$$e(G) = |E(G)| = \text{veličina od } G = \text{broj bridova od } G.$$



Slika 2.1. Graf sa 6 vrhova

Graf G na ovoj slici ima 6 vrhova, tj.  $v(G)=6$ , 10 bridova, tj.  $e(G)=10$ , ima jednu petlju, jedan dvostruki i jedan trostruki brid. Prilikom crtanja većinom izostavljamo oznake vrhova i bridova te crtežom reprezentiramo klasu ekvivalencije izomorfnih grafova.

Opisi grafova koji se najčešće rabe:

Jednostavan graf u kojem je svaki par vrhova spojen bridom zove se **potpun graf**

.Do na izomorfizam postoji jedinstveni potpun graf s n vrhova i  $\binom{n}{2}$  bridova

koje označavamo sa  $K_n$ . Ustvari ako je  $V(K_n) = \{1, 2, 3, \dots, n\} = [n]$ , onda je

$E(K_n) = \binom{[n]}{2}$ . Graf G je bipartitan ili dvodjelni ako mu se skup vrhova može

particionirati u dva skupa X i Y tako da svaki brid ima jedan kraj u X, a drugi u Y.

Particija (X,Y) zove se tada biparticija grafa. Bipartitni graf sa biparticijom (X,Y)

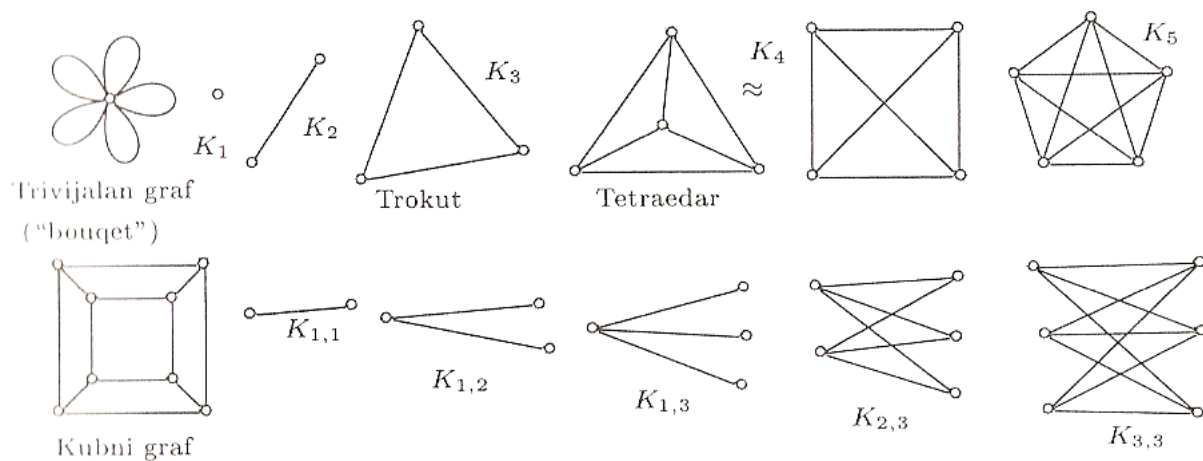
označavamo sa  $G(X,Y)$ . Potpun bipartitni graf jednostavan je bipartitni graf sa

biparticijom (X,Y) u kojem je svaki vrh iz X spojen sa svakim vrhom u Y. Ako je

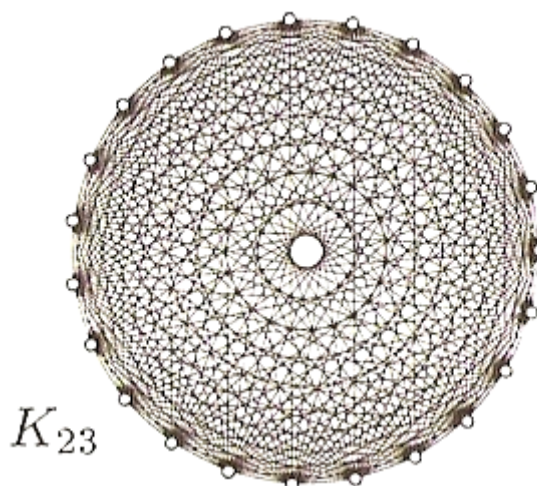
$|X|=m$  i  $|Y|=n$ , takav je graf jedinstven do na izomorfizam i označava se s  $K_{m,n}$ ;

$v(K_{m,n})=m+n$ ,  $e(K_{m,n})=mn$ . Graf određen vrhovima i bridovima kocke zove se kubni

graf. Evo nekih primjera:



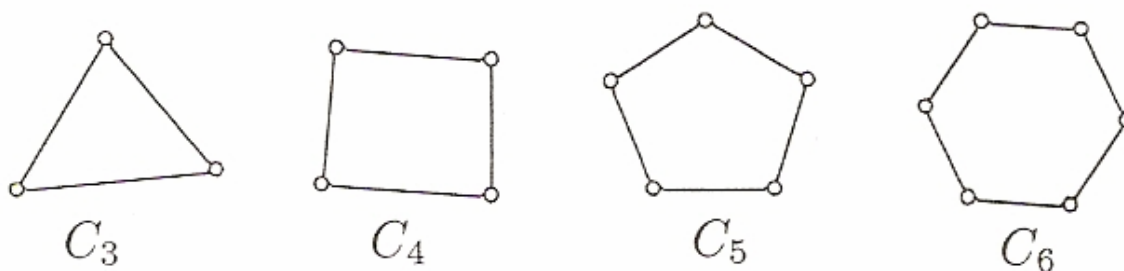
Slika 2.2. *Primjeri grafova*



Slika 2.3. *Crtež grafa  $K_{23}$  s 253 bridova (ornament, pleter, čipka)*

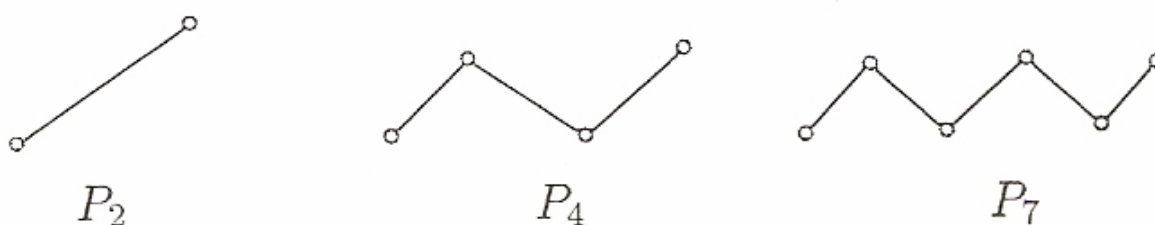
Daljnji važni primjeri (jednostavnih ) grafova su ciklusi i putovi. Ciklus  $C_n$  na  $n$  vrhova možemo definirati skupom vrhova  $V=\{1,2,\dots,n\}$  i skupom bridova

$$E=\{\{i,i+1\} | i < n\} \cup \{1,n\}$$



Slika 2.4. Ciklusi

Put  $P_n$  na  $n$  vrhova  $V=\{1,2,\dots,n\}$ ,  $E=\{\{i,i+1\} | i < n\}$



Slika 2.5. Put

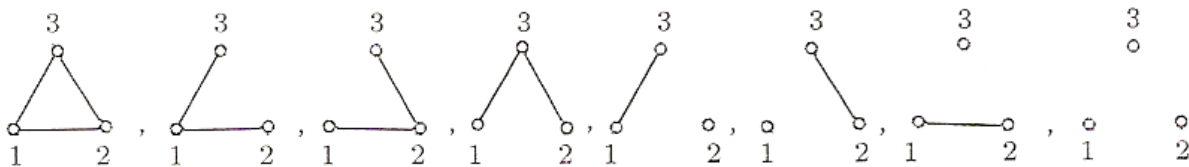
## 2.2 Postupak rješavanja grafova

Jedan od velikih otvorenih problema u teoriji grafova pita dali postoji neki efikasni postupak ( algoritam) kojim bi se moglo odlučiti jesu li dva dana grafa izomorfna ili ne. Postoje opravdane sumnje da takav algoritam ne postoji. Poteškoća je u tome da se provjeri jesu li dva grafa s  $n$  vrhova neizomorfna, mi trebamo prema definiciji provjeriti da ni jedan od  $n!$  Bijekcija na skupovima vrhova ne daje izomorfizam grafova. Često te provjere možemo izbjeći. Npr., ako su brojevi brida grafa različiti onda je jasno da su takvi grafovi neizomorfni. Općenito nije poznata neka efikasna metoda koja će uvijek razlikovati neizomorfne grafove.

Ako je  $V =\{1,2,\dots,n\}$  skup vrhova , onda izbor jednostavnog grafa sa skupom vrhova  $V$  znači da treba odabrati podskup  $E \subseteq \binom{V}{2}$ . Kako je  $\binom{V}{2} = \binom{n}{2}$ , slijedi da svih grafova

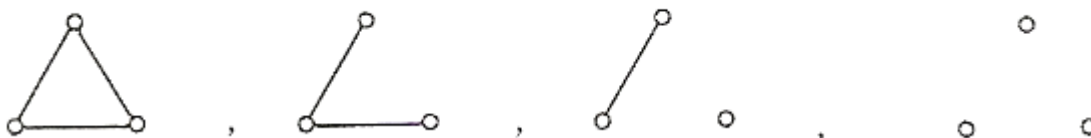
na  $V$  ima  $2^{\binom{n}{2}}$ . No među njima je znatno manje međusobno neizomorfni grafova. Npr., za  $n =3$ , imamo 8 grafova čiji je skup vrhova

$V = \{1, 2, 3\}$ . To su :



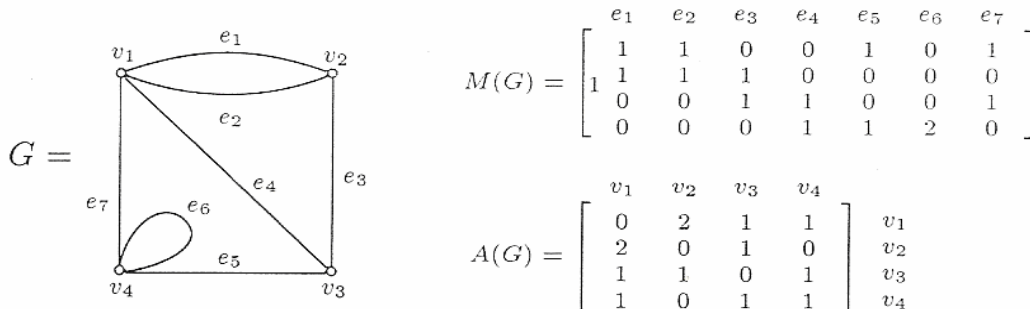
Slika 2.6. grafovi iz primjera

Među njima možemo naći samo 4 vrste međusobno neizomorfni. To su:



Slika 2.7. neizomorfni grafovi

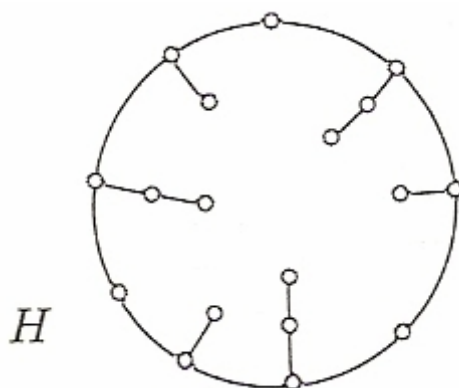
Često je vrlo korisno relacije incidencije i susjedstva u grafu prikazati matricama. Neka je  $G$  graf sa vrhovima  $v_1, v_2, \dots, v_n$ , u nekom poretku i bridovima  $e_1, e_2, \dots, e_m$  u nekom poretku. Matrica incidencije grafa  $G$  je (pravokutna) je  $n \times m$ -matrica  $M = M(G) = [m_{ij}]$ , gdje je  $m_{ij}$  broj (0, 1 ili 2) koliko su putav  $v_i$  i  $e_j$  incidentni. Matrica incidencije potpuno određuje graf. Matrica susjedstva (eng. adjacency matrix) grafa  $G$  je (kvadratna)  $n \times n$  matrica. Svaka takva matrica reprezentira neki graf.



Primjer 1. uspordeba grafova sa matricama



Matrica susjedstva je u pravilu mnogo manja od matrice incidencije. Napomenimo da obje ovise o odabranim poretcima, odnosno označavanju vrhova i bridova.



Slika 2.8 primjer grafova

Operacijom nad bridovima (osim odstranjenja i dodavanja) ima značajnu ulogu. To je kontrakcija bridova kažemo da je brid  $e \in E(G)$  kontraktiran ako je odstranjen, a njegovi vrhovi identificirani.



Slika 2.9 primjer kontraktiranog grafa

### 3. Clustering

**Clustering** ([ugroždnjavanje](#)) Petlje su specifične veze u mrežama. Sam pojam ugroždnjavanja je povezan

s petljama dužine 3 (rubovi trokuta). *Lokalno ugroždnjavanje* predstavlja relativni broj veza između najbližeg susjeda tjemena  $i$

$$C_i = \frac{n_i}{k_i(k_i - 1)/2}$$

$k_i$  je stupanj tjemena,  $n_i$  je ukupni broj veza između najbližih tjemena. Srednja vrijednost  $C_i$  tjemena stupnja  $k$  daje stupanj ovisnosti lokalnog ugroždnjavanja  $C(k)$ , koji pokazuje vjerojatnost da će se dva susjedna tjemena stupnja  $k$  spojiti.

Značenje samog ugroždnjavanja definirano je:

$$\langle C \rangle \equiv \langle C_i \rangle = \sum_i P(k)C(k)$$

Koeficijent ugroždnjavanja definiran je

$$C \equiv \frac{\langle n_i \rangle}{\langle k_i(k_i - 1)/2 \rangle} = \frac{\sum_k P(k)\langle n(k) \rangle}{(\langle k^2 \rangle - \langle k \rangle)/2}$$

Koeficijent ugroždnjavanja je tri puta proporcionalan omjeru svih stranica kutova trokuta i broju spojenih trostrukih tjemena.

Ako je prisutna neograničeno velika uzajamna veza, ugroždnjavanje je odsutno.

Tako, u uzajamnim vezama ugroždnjavanje je konačno velikog efekta. Na primjer, u klasičnim grafovima vrijedi

$$C(k) = C = \langle C \rangle \cong \frac{\langle k \rangle}{N}$$

#### 4. Distribucija stupnjeva čvorova ili raspodjela veza

Stupnjevi čvorova u slučajnim mrežama su statistički raspodijeljeni, gdje je stupanj čvora broj veza koje čvor ima sa susjednim čvorovima. U neusmjerenoj mreži, ako se čvorovi međusobno razlikuju, svaki čvor ima stupanj distribucije  $p(k,s,N)$  **Error! Reference source not found.** Drugim riječima, kažemo da je to vjerojatnost da čvor „s“ u mreži veličine N ima „k“ veza (k susjeda). Na osnovu distribucije stupnjeva za svaki čvor u mreži, moguće je onda pronaći ukupnu distribuciju stupnjeva :

$$P(k,N) = \frac{1}{N} \sum_{s=1}^N p(s,k,N)$$

Ako su svi čvorovi u slučajnim mrežama statistički jednaki, tada svaki od njih ima jednaku raspodjelu stupnjeva  $P(k,N)$ . Prvi moment te raspodjele daje nam srednji stupanj mreže:

$$\bar{k} = \sum_k kP(k)$$

Tada ukupan broj veza u mreži L možemo izračunati preko srednjeg stupnja mreže:

$$L = \bar{k} \frac{N}{2}$$

Distribucija stupnjeva čvorova opisuje samo lokalna svojstva mreže, iako je to mala količina informacija o mreži , obično bude dovoljno da se odrede neka osnovna svojstva mreže.

Neke od tipičnih distribucija stupnjeva čvorova:

- Poissonova distribucija  $P(k) = \frac{e^{-\bar{k}} \bar{k}^k}{k!}$

Klasični slučajni grafovi imaju ovakvu distribuciju ako im se broj čvorova približava beskonačnosti uz ograničenje da je srednji stupanj grafa konstantan.

- Eksponencijalna distribucija  $P(k) = e^{-\frac{k}{\bar{k}}}$

Ovo je distribucija rastućeg grafa

- Zakon potencije (*eng. power-law*)  $P(k) = k^{-\alpha}$

Za razliku od prve dvije distribucije power-law nema prirodni opseg te ju se još naziva distribucija bez skale, a mreže sa takvom distribucijom su mreže bez skale. U beskonačno velikim mrežama, svi viši momenti power-law distribucije reda  $m \geq \alpha - 1$  divergiraju. Iz toga se može vidjeti koje vrijednosti može poprimiti eksponent  $\alpha$  za mreže bez skale:

- ako je prosječan stupanj mreže bez skale (prvi moment distribucije) konačan onda je eksponent  $\alpha$  veći od 2
- ako prosječan stupanj mreže bez skale varira, što je slučaj za većinu realnih mreža, a distribucija stupnja čvorova je konstantna onda je  $1 < \alpha < 2$

Konačne mreže, a to znači sve realne mreže, imaju distribucije koje imaju tzv. rep i one su prirodno odrezane.

## 5. Preferirano povezivanje

Ovo svojstvo kompleksnih mreža nam govori što su čvorovi višeg stupnja imaju veću vjerojatnost, u odnosu na ostale čvorove, da svoj stupanj još više povećaju.

To je već odavno poznati fenomen u socijalnim mrežama, a poznat je kao Matijin efekt, dobio je ime po ulomku iz Matejevog evanđelja. Matijin efekt primijenjen u mrežama u biti kaže da čvorovi s puno veza će na sebe privući nove veze, dok slabije povezani čvorovi će vjerojatno takvi i ostati.

Slično ponašanje javlja se i u ekonomiji, koje je još u 19. st. Primijetio V. Pareto, a danas je poznato kao Paretov zakon ili 80/20 koji ukratko kaže da se bogati još više bogate odnosno da 20% populacije posjeduje 80% dobara. Taj je zakon u statistici poznat kao zakon potencije (eng. power-law).

Kako je primijećeno takvo ponašanje i u realnim mrežama, možemo reći da je proces preferiranog povezivanja glavni sastojak pri kreiranju mreže bez skale. Model mreža bez skale pretpostavlja da je vjerojatnost  $p(k)$  povezivanja čvora na neki čvor „i“ proporcionalna stupnju čvora „i“.

$$p(k_i) = \frac{k_i}{\sum_j k_j}$$

Ta pretpostavka uključuje dvije hipoteze: prvo da  $p(k)$  ovisi o stupnju „k“ čvora, a druga da je  $p(k)$  linearno ovisna o  $k$ . Funkcijski oblik  $p(k)$  može biti određen za mreže za koje znamo u kojem će se trenutku pojaviti novi čvor i biti povezan s mrežom. Primjer takve mreže je mreža koautorstva među znanstvenicima ili mreža citata članaka.

Pretpostavimo neko trenutno stanje mreže, zapamtimo njen broj čvorova i njihov stupanj. Nakon nekog intervala  $\Delta T$ , koji je puno kraći od starosti mreže, dolazi do povećanja stupnja čvorova. Ako prema relaciji (7) prikazujemo relativnu promjenu stupnja čvorova  $\frac{\Delta k_i}{\Delta k}$  u ovisnosti o prijašnjem stupnju pojedinog čvora, dobivamo funkciju  $p(k)$ .  $\Delta k$  je broj veza koje su otvorene tokom vremena  $\Delta T$ .

Da bi prikazali tu ovisnost i smanjili oscilacije podataka tokom statističke obrade često se prikazuje kumulativna distribucija

$$P(k) = \sum_{k_i=0}^k p(k_i)$$

Iz empirijskih zapažanja, kao u slučaju mreže koautorstva i mreži citiranih članaka, primijećeno je da  $p(k)=k^\alpha$ .

Efekt nelinearne funkcije  $p(k)$  u dinamici mreže i njenoj topologiji gdje je zamjenom linearnog preferiranog povezivanja sa nelinearnim  $p(k)=k^\alpha$  u usmjerenj mreži izračunat srednji broj  $N_k(t)$  čvorova sa  $k-1$  ulaznim vezama u vremenu  $t$ . Iz proračuna proizlazi da priroda mreže bez skale se uništava za nelinearno preferirano povezivanje. Jedini slučaj u kojem je topologija mreže bez skale sačuvana kada je preferirano spajanje asimptotski linearno,  $p(k_i) \sim ak_i$  kada  $k_i \rightarrow \infty$ . U tom slučaju dobivamo

$$P(k) = k^{-\alpha}$$

eksponent  $\alpha$  može biti podešen na bilo koju vrijednost,  $2 < \alpha < \infty$ .

Još jedno svojstvo  $p(k)$  u realnim mrežama da je  $p(0) \neq 0$ , što znači da postoji vjerojatnost da se novi čvor poveže sa nekim izoliranim čvorom, otuda proizlazi

$$p(k) = A + k^\alpha$$

gdje je  $A$  početna atraktivnost čvora „i“ **Error! Reference source not found.** Ako je  $A = 0$  čvor koji ima stupanj povezanosti  $k = 0$  nikada neće moći povećati svoj stupanj (povezanost) što proizlazi iz (7). No u realnim mrežama svaki čvor ima neku konačnu šansu da bude otkriven i povezan. Zato parametar  $A$  označava vjerojatnost da je novi čvor otkriven, kao u slučaju kad je novi članak po prvi puta citiran. U modelu iz jedinice vremena mreži se dodaje novi čvor što je popraćeno i dodavanjem novih veza iz bilo kojih čvorova u mreži prema nekom od preferiranih čvorova. Vjerojatnost

da neki čvor dobije neku od tih  $m$  novih veza proporcionalna je sumi početne atraktivnosti i broju novih veza.

$$p(k_{in}) = A + k_{in}$$

gdje  $k_{in}$  označava ulazni stupanj čvora (stupanj čvora po broju veza koje su usmjerene prema njemu). Proračun iz **Error! Reference source not found.** ukazuje da distribucija stupnjeva čvorova prati zakon potencije  $P(k) = k^{-\alpha}$  uz iznos potencije  $\alpha = 2 + A / m$ . Posljedice su da početna atraktivnost ne uništava prirodu mreža bez skale već im samo mijenja potenciju.

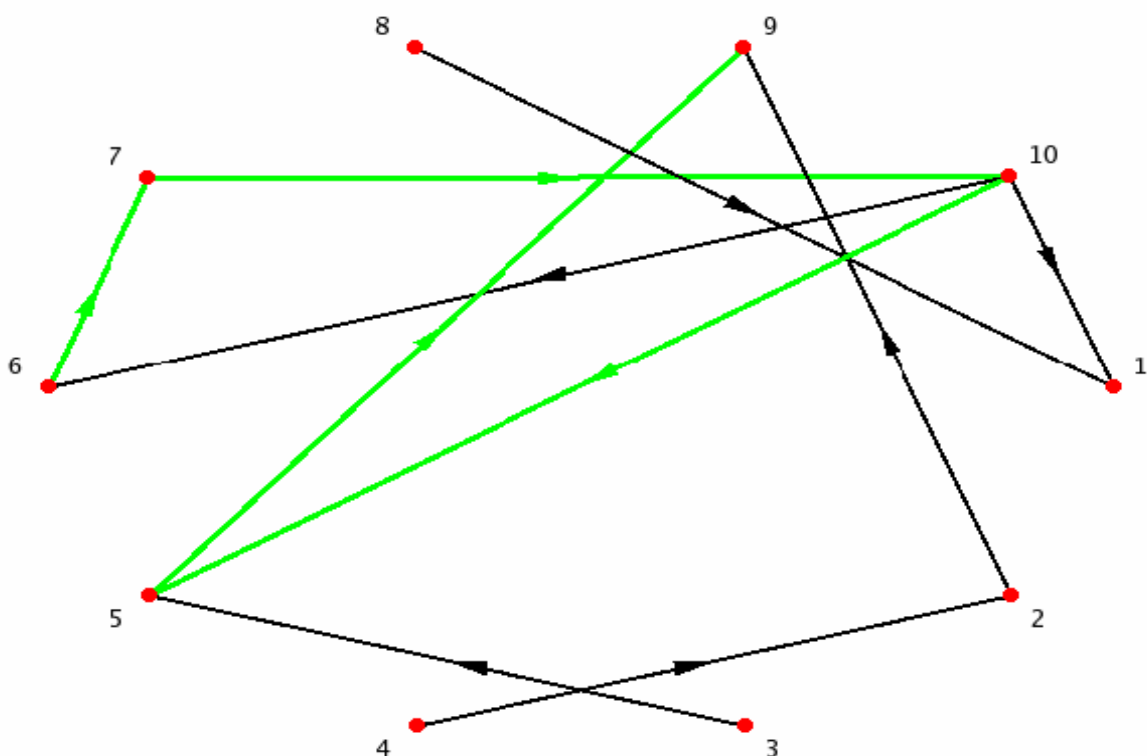
## 6. Objašnjenje srednjeg najkraćeg puta

Najkraći put (engl. shortest path) se često koristi za optimiziranje raznih ruta u Internetu, transportu itd. Pokazalo se kako je srednji najkraći put iznimno mali s obzirom na veličinu mreže te se taj efekt zove efekt malog svijeta.

Izračunavanje najkraćeg puta možemo raditi pomoću više algoritama, ovisno kakvog je tipa mreža. U ovom radu najkraći put za bestežinske mreže izračunat je pomoću pretraživanja u širinu (engl. *breadth first search*)[1] dok za težinske grafove pomoću Dijkstrinog algoritma[1].

Složenost pretraživanja u širinu je  $O(V + E)$  dok je složenost Dijkstrinog algoritma ovisi o tome kako je implementiran prioritetni red. U ovom radu red je implementiran pomoću hrpe te je složenost  $O((V + E) * \log V)$ .

Slika 5-1 prikazuje usmjereni bestežinski graf te je zelenom bojom istaknut najkraći put između vrhova 6 i 9.



Slika 6.1 Bestežinski usmjereni graf. Zelenom bojom je istaknut najkraći put između vrha 6 i 9



## 7. Metode

U ovome poglavlju bit će obrazložene metode korištene za procjenu parametara odabranih modela distribucije.

Da bi se mogla procijeniti svojstva kompleksnih mreža potrebno je statistički obraditi podatke koji se mogu dobiti iz mreže. Iako količina informacija koju možemo izvući je mala, obično je to distribucija stupnjeva čvorova, ona ponekad može biti i sasvim dovoljna za procjenu nekih osnovnih svojstava. Ovdje se spominju razne vrste realnih mreža, te se napominje kako je tek empirijski utvrđeno da se velik broj realnih mreža ponaša kao mreže bez skale ili ako se njihova topologija može usporediti sa topologijom slučajnih grafova tada vjerojatno distribucija stupnjeva čvorova ima binomnu, eksponencijalnu ili Poissonovu raspodjelu.

Nakon što se prikupe podaci problem je što nikad ne znamo koliko naš odabrani model dobro prati prikupljene podatke. Da bi mogli procijeniti koliko je dobar odabrani model potrebno je pronaći vrijednosti njegovih parametara, odnosno estimirati ih na temelju podataka kojima raspolažemo. No i nakon estimacije parametara kako biti siguran da je model s dobivenim parametrima dobar. Jedan od načina da grafički provjerimo kako za dobiveni parametar pretpostavljeni model prati podatke. Drugi način je numerički, korištenjem statističkog testa, Kolmogorov-Smirnov test.

No prvo je potrebno odrediti parametre. Dvije su glavne metode estimacije parametara. Metoda najmanjih kvadrata (LSE) i metoda maksimalne vjerojatnosti (eng. *maximum likelihood estimation* – MLE). Naglasak u ovom radu je na MLE metodi, zbog više razloga koji su navedeni u narednom tekstu, dok se metoda najmanjih kvadrata koristi u slučaju kad MLE metodom nismo mogli dobiti traženu ovisnost parametara.

## 8. Procjena maksimalne vjerojatnosti

MLE je standardna metoda estimacije u statistici. Neka svojstva MLE estimacije: dostatnost - kompletne informacije o traženom parametru sadržane su u MLE procjeni, konzistentnost, efikasnost – najmanja moguća varijanca estimiranog parametra i parametarska nezavisnost – model dobivanja MLE neovisan je o parametru koji se traži.

Statistički gledano,  $x = (x_1, x_2, \dots, x_n)$  je vektor podataka slučajnih uzoraka nekog skupa. Cilj analize podataka da se identificira skup za koji je najvjerojatnije da je producirao te uzorke. U statistici svaki skup je identificiran sa odgovarajućom distribucijom vjerojatnosti. Odnosno svaka vrijednost parametra nekog modela vezana je za distribuciju vjerojatnosti. Ako parametar mijenja svoju vrijednost tada se dobiva i drugačija distribucija vjerojatnosti. Po definiciji, model je definiran kao familija distribucija vjerojatnosti označen parametrom modela.

Neka je  $p(x|\alpha)$  funkcija gustoće vjerojatnosti koja nam određuje vjerojatnost pojavljivanja promatranih podataka vektora  $x$  sa pripadajućim parametrom  $\alpha$ . Ako su elementi vektora  $x_i$  statistički neovisni jedni o drugima tada prema teoriji vjerojatnosti funkcija gustoće vjerojatnosti za podatke  $x=(x_1, x_2, \dots, x_n)$  sa pripadajućim parametrom  $\alpha$  se može izraziti kao umnožak funkcija gustoća vjerojatnosti svakog elementa vektora podataka:

$$p(x=(x_1, x_2, \dots, x_n) | \alpha) = p_1(x_1 | \alpha) p_2(x_2 | \alpha) \dots p_n(x_n | \alpha)$$

$$p(x=(x_1, x_2, \dots, x_n) | \alpha) = \prod_{i=1}^n p(x_i | \alpha)$$

Dakle, za skup vrijednosti parametara odgovarajuća funkcija gustoće vjerojatnosti nam govori koliko su neki podaci vjerojatniji od drugih te na taj način možemo odrediti skupinu za koju možemo reći da je najvjerojatnije da je generirala podatke. Međutim u praksi mi ne znamo kolika nam je vrijednost našeg parametra odnosno

situacija je obrnuta, potrebno je odrediti vrijednost parametra za koje će nam naš model (funkcija gustoće vjerojatnosti) reći da su naši podaci za taj model najvjerojatniji. Da riješimo ovaj problem definiramo *likelihood* funkciju koja okreće uloge vektora modela i traženog parametra.

Likelihood funkcija je dana kao  $L(\alpha|x_i)$ , odnosno  $L(\alpha|x_i) = p(x_i|\alpha)$  i predstavlja vjerojatnost parametra  $\alpha$  ako nam je poznat vektor podataka.

Cilj MLE estimacije je pronaći vrijednost parametra modela za koje su dobiveni podaci najvjerojatniji, odnosno što znači da je potrebno pronaći vrijednosti u vektoru parametara za koje će likelihood funkcija imati svoj maksimum. Dobiveni parametar vektor nazivamo MLE procjena. U općenitom slučaju MLE procjena ne mora postojati niti može biti jedinstvena. No za naše predložene modele moguće je odrediti MLE. Često se zbog lakšeg računanja, odnosno zbog nelinearnih funkcija gustoće vjerojatnosti za proračun MLE procjene koristi logaritmiran oblik likelihood funkcije, tzv. log-likelihood  $\ln L(\alpha|x_i)$ , te tako iz nelinearne dobivamo linearnu likelihood funkciju.

Ako je log-likelihood derivirabilna i ako postoji  $\alpha_{MLE}$  tada  $\ln L(\alpha|x_i)$  mora zadovoljavati parcijalnu jednadžbu:

$$\frac{\partial \ln L(\alpha | x_i)}{\partial \alpha_{MLE}} = 0$$

Ova jednadžba predstavlja nužan uvjet da bi MLE procjena postojala. Dodatni uvjet je svakako da je  $\ln L(\alpha|x_i)$  maksimum, a ne minimum što se provjerava drugom parcijalnom derivacijom:

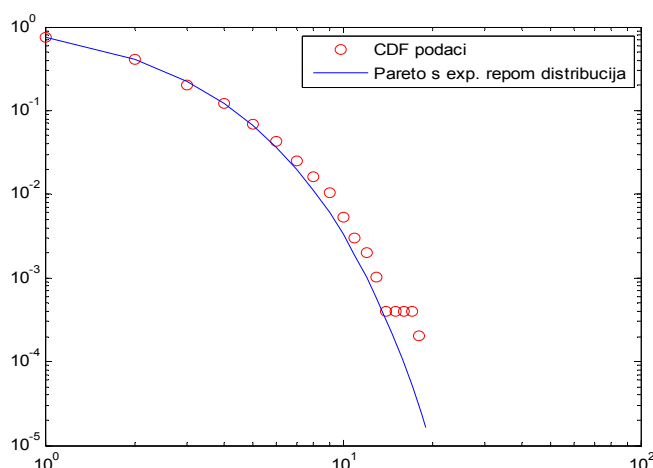
$$\frac{\partial^2 \ln L(\alpha | x_i)}{\partial \alpha_{MLE}^2} = 0$$

## 9. Metoda najmanjih kvadrata

Druga korištena metoda za procjenu parametara distribucije je metoda najmanjih kvadrata. Za razliku od MLE metode, pomoću ove tražimo vrijednost parametra koji će nam dati najtočniji opis podataka u kontekstu mjerenja – koliko predviđeni model je blizu podacima. U ovoj metodi suma kvadrata pogrešaka (SSE) između promatranih i predviđenih podataka je minimalna.

Ovu metodu korištena je za model Pareto distribucije s eksponencijalnim repom jer MLE metoda nije davala ovisnost parametra  $\alpha$  o parametru  $\kappa$ . Prva derivacija log-likelihood funkcije ovog modela anulira parametar  $\kappa$ .

Estimaciju parametra  $\alpha$  smo dobili koristeći funkciju `nlinfit()` iz programskog paketa `matlab`. Ova funkcija vraća parametar estimiran metodom najmanjih kvadrata. Parametar određuje tako da minimizira sumu kvadratnih razlika između predviđenih i promatranih vrijednosti. Oblik funkcije je `nlinfit(X,Y, fun, beta0)` gdje je `X` matrica predviđenih vrijednosti (neovisna varijabla) gdje je svaki redak predviđen za pojedinu vrijednost u `Y` vektoru. `Y` predstavlja vektor promatranih vrijednosti (zavisna varijabla). `Fun` je definirana funkcija, odnosno model koji vraća vektor predviđenih vrijednosti definirane funkcije i koji se zatim uspoređuju s vektorom `Y`. `Beta0` je vektor početnih vrijednosti parametara koje se estimira.



*Slika 8.1 loglog graf Pareto s exp. repom distr. s parametrom estimiranim LSE metodom*

## 10. Primjena kompleksnih mreža

Primjena kompleksnih mreža jako je opširna jer se koriste u gotovo svim područjima znanosti pa tako naprimjer kompleksne mreže koristimo u elektrotehnici, akustici, molekularnoj biologiji, i mnogi sustavima gdje god nešto radimo sa nekim grafovima bilo koje konstrukcije najbolji primjer nam je primjena kompleksnih mreža u biološkom sustavu gdje promatranjem nekih od stanica bića ubiti vidimo jako lijepe konstrukcije grafova i mnogo zakona koje upravo na tome principu i mi danas koristimo.

## 11. Zaključak

Grafičkim prikazom dobivenih raspodjela s estimiranim parametrima, vidi se da je algoritam uspio pronaći model koji se najbolje poklapa s distribucijom veza između čvorova. Što samo potvrđuje pretpostavku da kompleksne mreže imaju jedan od pretpostavljenih modela distribucije veza. Za razliku od grafičkih rezultata iz numeričkih bi se mogao donijeti zaključak da niti jedan od modela ne odgovara distribuciji podataka. No uspoređujući vrijednosti maksimalnih udaljenosti između kumulativnih funkcija  $D$ , odluku koji od modela najbolje prati podatke mogli bi temeljiti na najmanjoj vrijednosti  $D$  među predloženim modelima.

Da kompleksne mreže često imaju i kompleksniju raspodjelu veza među čvorovima, odnosno ta je to nerijetko kombinacija raspodjela, potvrđuju nam i dobiveni rezultati, gdje se vidi da se model sačinjen od Paretove i eksponencijalne raspodjele najbolje poklapa s distribucijom podataka. To se može potkrijepiti s još jednim primjerom, a to je Pareto raspodjela sa skraćenim repom. Vidi se da ona jednim dijelom prati raspodjelu podataka, no da bi je pratila i u repu raspodjele potreban je promijenjeni model za estimaciju parametra samo za rep.

## **12. Sadržaj**

Iskustvene studije mreža u stvarnom svijetu kao što su Internet, WWW, socijalne i suradničke te razne biološke mreže potakle su na daljnje istraživanje i razvijanje u tom smjeru. Mnogi istraživači predložili su razne modele koji su objašnjavali nastajanje mreža takvih struktura ili očekivane efekte istih. Kako je napredovalo razvijanje ovog područja tako se pojavila i potreba za adekvatnim alatima koji će omogućiti daljnji napredak u istraživanju. Ti alati imaju mogućnost simulacije i analize kompleksnih mreža. I u ovom radu opisujemo neke od tih alata i neka pravila kod kompleksnih mreža.

### 13. Literatura

- [1] S. N. Dorogovtsev, J. F. F. Mendes: The shortest path to complex networks, 24. srpnja 2004., arXiv:cond-mat/0404593 v4, 27. travnja 2007.
- [2] A. L. Barabási and R. Albert: Emergence of scaling in random networks, Science 286, 509.
- [3] M. E. J. Newman: The structure and function of complex networks, 25. ožujka 2003., arXiv:cond-mat/0303516v1, 27. travnja 2007.
- [4] Darko Veljan, Kombinatorna i diskretna matematika –Zagreb : Algoritam, 2001.
- [5] Internet, <http://www.irb.hr/hr/research/initiatives/bioinf/popis/>
- [6] Internet, <http://www.politehnikapula.hr/kolegiji/elektrotehnika/upload/>
- [7] Internet, <http://www.tel.fer.hr>