Sveučilište u Zagrebu
Fakultet Elektrotehnike i Računarstva

Seminar

Modeli rasta kompleksnih mreža

Matija Piškorec

voditelj: *Mr. sc. Mile Šikić*

15. svibnja 2007.

Sadržaj

Uvod

Teorija mreža[1] dugo je vremena bila predmet istraživanja uglavnom matematičara. Sredinom 20. stoljeća prodire u znanosti kao što su računarstvo, komunikacijske znanosti, biologija, sociologija, ekonomija... U kasnim devedesetima pitanje evolucije i strukture mreža praktički postaje novo područje fizike [1].

Ovaj prijelaz omogućen je brojnim empirijskim opažanjima mreža sa polinomnom (eng. *power-law*) distribucijom stupnjeva, prije svega World Wide Weba i Interneta. Klasična teorija mreža do tada se bavila samo mrežama s distribucijama Poissonovog tipa gdje su veze distribuirane nasumično između konačnog broja čvorova. Upravo je zbog toga koncept *kompleksnih mreža* iznikao kao važan model za razumijevanje rasta i evolucije velikih mreža.

U ovom seminaru ukratko ću iznijeti glavne ideje u svezi s rastom i evolucijom kompleksnih mreža. Cilj mi nije ulaziti detaljno u matematičku analizu nego samo ukratko predstaviti osnovne modele klasičnih i kompleksnih mreža te njihova svojstva i način konstrukcije.

U prvom poglavlju "Što su kompleksne mreže?" ukratko se objašnjava sam pojam *kompleksne mreže* i njihovo značenje u današnjim istraživanjima. Također se definiraju važniji pojmovi iz teorije mreža, nužni za razumijevanje ostatka seminara. Konstrukcije važnijih tipova mreža iznose se u poglavlju "Ravnotežne i neravnotežne mreže". Nadalje, u poglavlju "Neki važniji primjeri kompleksnih mreža" definiraju se osnovni tipovi kompleksnih mreža i metode njihove konstrukcije. Napokon, u poglavlju "Barabási-Albert model" iznosi se model nastanka kompleksnih mreža prema Barabási-Albertu i opisuje se programska implementacija. Poglavlje "Preferencijalno povezivanje" uz kratko poopćenje Barabási-Albert modela argumentira i važnost preferencijalnog povezivanja u nastanku kompleksnih mreža. Na kraju, u poglavlju "Alternativni modeli konstrukcije mreža bez skale", predstavljaju se još neke nestandardne metode konstrukcije koje također vode do mreža bez skale.

[1] U literaturi prevladavaju dvije terminologije kada se govori o ovoj tematici - matematičarska sa izrazima graf, vrh i brid (eng. *graph, vertice, edge*) i fizičarska sa izrazima mreža, čvor i veza (eng. *network, node, connection*). Kako je riječ o ekvivalentnim terminima, u radu ću se koristiti fizičarskom terminologijom.

1 Što su kompleksne mreže?

U suštini, kompleksne mreže su mreže koje su u svojim karakteristikama kompleksnije od klasičnih slučajnih mreža. Pri tome se misli da im je organizacija veza kompleksnija, na primjer distribucija stupnjeva može biti kompleksnija od standardne Poissonove distribucije ili imati efekt "debelog repa" (eng. *fat tail*).

Zašto su kompleksne mreže važne? Pokazuje se da klasične slučajne mreže nemaju mnogo sličnosti sa stvarnim mrežama, zbog čega se ne mogu uspješno koristiti kao modeli za objašnjavanje empirijski dobivenih podataka iz stvarnog svijeta. Čak i danas najrazvijeniji modeli rasta i evolucije kompleksnih mreža, kao na primjer Barabási-Albert model, imaju ograničene mogućnosti predviđanja karakteristika stvarnih mreža.

U nastavku ću ukratko definirati važnije pojmove vezane uz analizu mreža - *distribucija stupnjeva* i *grupiranje*. Navedene karakteristike pomažu nam da identificiramo svojstva i tip mreže i s njima ćemo se često susretati u nastavku seminara.

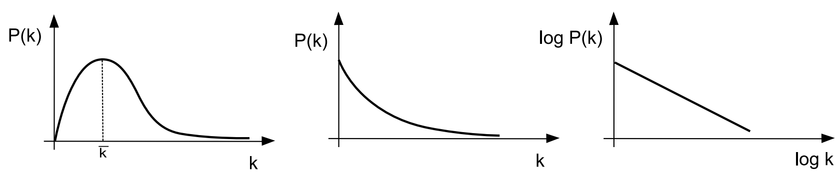
Distribucija stupnjeva (*degree distribution*)

Jedna od najvažnijih karakteristika mreže je njena distribucija stupnjeva. Nama najvažniji oblici distribucije (slika 1) su Poissonova, eksponencijalna i polinomna distribucija. Detaljnije ću ih objasniti kada se susretnemo s njima. Za početak izvedimo opću definiciju.

Za svaki čvor možemo definirati njegovu distribuciju stupnjeva $p(k, s, N)$. To je vjerojatnost da će čvor s u mreži veličine N čvorova imati stupanj k. Poznavajući distribuciju svakog stupnja možemo naći *ukupnu distribuciju stupnjeva* mreže:

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

Pretpostavimo li da su svi čvorovi statistički ekvivalentni, kao što je slučaj u klasičnim slučajnim mrežama, onda svaki od njih ima jednaku distribuciju stupnjeva $P(k, N)$. Prvi moment distribucije je zapravo prosječni stupanj cijele mreže $\overline{k} = \sum_k kP(k)$. Moguće je još definirati *ulazne* i *izlazne* distribucije stupnjeva, ali one nam ovdje neće biti od pretjerane važnosti.



Slika 1: Poissonova, eksponencijalna i polinomna distribucija stupnjeva.

Grupiranje (*clustering*)

Koeficijent grupiranja određuje "gustoću" veza u okolini bliskoj čvoru. Watts i Strogatz u [5] definiraju koeficijent grupiranja C kao omjer između svih y veza koje spajaju najbližih z susjeda čvora i svih *mogućih* veza koje bi ih mogle spajati. Dobiveni omjer je

$$C = \frac{2y}{z(z-1)}$$

Iako bi se mogla uvesti nekakva distribucija od C, najčešće se koristi samo prosječna vrijednost grupiranja $\overline{C}$. Koeficijent grupiranja direktno je vezan za prisutnost "trokutova" (tj. ciklusa duljine tri) u mreži. Ujedno je i vrlo dobar pokazatelj toga postoje li korelacije između čvorova mreže.

Za klasične slučajne mreže vrijedi omjer $\overline{C} = \frac{\overline{k}}{N}$ gdje je N broj čvorova mreže a $\overline{k}$ prosječni stupanj čvorova. Kako je $\overline{k}$ fiksna vrijednost (u svakom vremenskom koraku dodaje se jednak omjer čvorova i veza) on se približava nuli kako veličina mreže raste. Empirijska istraživanja pokazuju da to nije slučaj s mnogim stvarnim mrežama.

2 Ravnotežne i neravnotežne mreže

Ovdje uvodim vrlo važne pojmove - ravnotežne (eng. *equilibrium*) i neravnotežne (eng. *non-equilibrium*) mreže. U čemu je razlika?

Fizikalno gledano, za proučavanje evolucije i rasta mreža od ključne je važnosti znati mijenjaju li se one s vremenom ili su "statične". Potonje imaju fiksan broj čvorova te nasumičnim dodavanjem veza s vremenom dolaze u ravnotežno stanje. Kod neravnotežnih mreža novi čvorovi i veze se neprestano dodavaju u mrežu. Takve mreže očito su daleko od statičnih i ne postižu ravnotežno stanje.

U nastavku se demonstriraju konstrukcije najvažnijih primjera ravnotežnih i neravnotežnih mreža.

Konstrukcija klasične slučajne mreže

Prvi jednostavni model konstrukcije klasične slučajne mreže predstavili su Erdös i Rény 1959. i 1960. u [6] i [7]. Takva mreža definirana je dvijema jednostavnim pravilima:

(a) Ukupni broj čvorova N je fiksan.

(b) Nasumično odabrana dva čvora spojena su neusmjerenim vezama[2]. Isto tako može se reći da za svaka dva čvora mreže postoji vjerojatnost p da su spojeni (neusmjerenom vezom).

Izvedimo ukratko formulu za distribuciju stupnjeva takve mreže. Svaki čvor u mreži može imati stupanj u rasponu od 0 do $N-1$. Ako je čvor stupnja k svaka od tih k veza može biti spojena na bilo koji od preostalih $N-1$ čvorova. Standardna kombinatorika vodi nas na formulu za distribuciju stupnjeva u klasičnoj slučajnoj mreži

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

Vidimo da je riječ o binomnoj distribuciji. Prosječan stupanj čvorova $\overline{k}$ je $\overline{k} = p(N-1)$ i mreža u prosjeku sadrži $\frac{pN(N-1)}{2}$ veza. Za velike N i fiksne $\overline{k}$ binomna distribucija se može aproksimirati Poissonovom:

$$P(k) = \frac{\overline{k}^k}{k!} e^{-\overline{k}}$$

Poissonova distribucija je, kako vidimo na slici 1, brzo opadajuća distribucija sa prirodnom skalom $k \sim \langle k \rangle$.

Konstrukcija neravnotežne slučajne mreže

Ovakva mreža konstruira se jednostavnim dodavanjem novih čvorova i povezivanjem slučajno odabranog para čvorova. Postupak je:

(a) U svakom vremenskom koraku u mrežu se dodaje novi čvor.

(b) Istovremeno, jedan ili više nasumično odabranih parova čvorova se povezuju vezom.

[2] Ili *vezom*, ako zabranimo višestruko spajanje čvorova, ali to ionako nije važno za velike mreže gdje su takve veze gotovo nepostojeće.

Statistički gledano, najstariji čvorovi su i najbolje povezani (tj. većeg stupnja). Ako u nekom trenutku prestanemo dodavati nove čvorove i nastavimo s povezivanjem mreža će težiti ravnotežnom stanju[3].

Ovakva mreža ima eksponencijalnu distribuciju stupnjeva

$$P(k) \propto e^{\frac{-k}{\overline{k}}}$$

gdje je $\overline{k}$ prosječan stupanj mreže, $\overline{k} = \sum_{k=0}^{\infty} kP(k)$. I ovakva distribucija, poput Poissonove, je ekstremno opadajuća.

Značajniji pomak učinit će tek Barabási i Albert 1999. u [3]. Oni predlažu model u kojem se kombinira rast i preferencijalno spajanje novih vrhova. Rezultat je *mreža citata* u kojem se novi čvorovi spajaju na stare sa vjerojatnošću proporcionalnom njihovom stupnju.

Na Barabási-Albert model vratiti ću se kasnije. Definirajmo prvo mrežu citata i metodu njezine konstrukcije.

Konstrukcija mreže citata

Poseban slučaj neravnotežne slučajne mreže je mreža citata. U najjednostavnijoj varijanti konstruira se na sljedeći način:

(a) U svakom vremenskom koraku u mrežu se dodaje novi čvor.

(b) Novopristigli čvor spaja se s nekim od starih čvorova.

Konfiguracija tako stvorene mreže snažno ovisi o načinu na koji se bira stari čvor. Kao što ćemo vidjeti u nastavku seminara, od interesa će nam biti *preferencijalno spajanje* u kojem se čvor bira s vjerojatnošću proporcionalnom stupnju koji posjeduje. Općenito, vjerojatnost da će se stari čvor spojiti s novim proporcionalna je s $f(k)$ - nekom funkcijom od stupnja k starog čvora.

Naziv *mreža citata* dolazi od modela iz stvarnog svijeta - mreža citiranja iz znanstvenih časopisa u kojoj su dva rada povezana ako jedan sadrži citat (referencu) na drugi. Činjenica da u ovom modelu mreža po definiciji nije usmjerena (tj. veze nemaju definiran smjer) a u stvarnosti je nije nam sada od krucijalne važnosti. Najvažnija sličnost je da u tako definiranom modelu nije moguće pojavljivanje veza između starih čvorova, kao što ni u stvarnosti stvarni članci ne mogu mijenjati svoje citate (tj. stvarati nove veze). Ta mogućnost je omogućena isključivo novim čvorovima.

[3]Ali ga, prema definiciji, nikad neće dostići. Jedini izlaz je da omogućimo da stari čvorovi s vremena na vrijeme odumru.

3 Neki važniji primjeri kompleksnih mreža

Sada, nakon kratkog pregleda osnovnih tipova mreža i njihovih konstrukcija, navest ću i najvažnije primjere kompleksnih mreža - mreže bez skale (eng. *scale-free*) i mali svijetovi (eng. *small-world*). Prve su važne zbog svoje polinomne distribucije stupnjeva, što je svojstvo koje je zamijećeno u brojnim realnim mrežama, od Interneta i WWW-a do proteinskih interakcija u metabolizmu. Potonje iskazuju takozvani *efekt malog svijeta* koji je zapravo vrlo uobičajen u većini mreža (pa čak i u klasičnim slučajnim mrežama). Bez obzira na to, ovdje se ipak misli na vrlo specifičnu vrstu mreža koje samo ispoljavaju efekt malog svijeta.

Mreže bez skale (*scale-free networks*)

Karakteristično za mreže bez skale je da se u njima pojavljuju čvorovi (takozvani *hubovi*) koji imaju stupnjeve za red veličine veće od prosječnog stupnja svih čvorova. Pri tome karakteristike mreža bez skale ne divergiraju bitno bez obzira na veličinu mreže, tj. broj čvorova N. Najvažnija karakteristika takvih mreža je njihova polinomna distribucija stupnjeva

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

Ključni parametar je γ koji se još naziva i *eksponent distribucije*.

Mreže sa polinomnom distribucijom nemaju prirodnu skalu[4] zbog čega i nose naziv bez skale.

Mali svijetovi (*small-world networks*)

Fenomen "malog svijeta" poznat je već neko vrijeme iz stvarnog svijeta. Najpoznatiji primjer je slavni pokus Stanleya Milgrama iz 1960-tih [4] i njegova demonstracija da su svaka dva čovjeka na svijetu povezana u prosjeku preko najviše šest drugih ljudi. Formalno, efekt malog svijeta zahtjeva preciznu definiciju *prosječne najkraće udaljenosti između dva čvora mreže.* Očito, prosječni broj n najbližih susjeda jednog čvora raste usporedivo sa $\langle k \rangle^n$. Znači, prosječan najkraći put $\overline{\ell}$ ugrubo se aproksimira relacijom $\langle k \rangle^{\overline{\ell}} \sim N$. Stoga vrijedi

$$\overline{\ell} \approx \frac{\ln N}{\ln \langle k \rangle}$$

Uglavnom, ako $\overline{\ell}(N)$ raste sporije nego ijedna pozitivna potencija od N za mrežu možemo reći da ima efekt malog svijeta.

Mali svijet - model Wattsa i Strogatza

Jedni od prvih koji su predložili male svijetove kao specifičnu klasu kompleksnih mreža su Watts i Strogatz u [5]. U svojem radu oni predlažu superpoziciju između rešetke i klasične slučajne mreže kao optimalnu metodu konstrukcije malih svijetova.

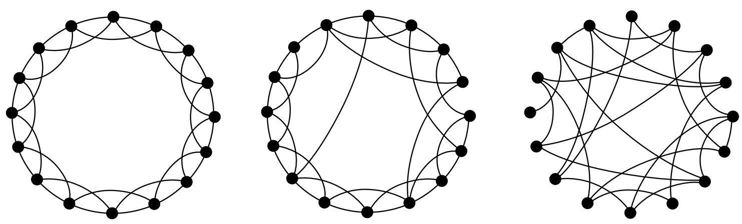
---

[4]Podsjetimo se da, za razliku od njih, prirodna skala kod mreža sa Poissonovom ili eksponencijalnom distribucijom postoji i reda je veličine prosječnog stupnja svih čvorova $\overline{k}$.

U tom modelu počinje se od pravilne rešetke i potom se nasumičnim prespajanjem postojećih veza stvaraju "prečaci". Rezultat je mreža sa visokim koeficijentom grupiranja i kompaktnošću klasične slučajne mreže. Metoda koja se provodi je:

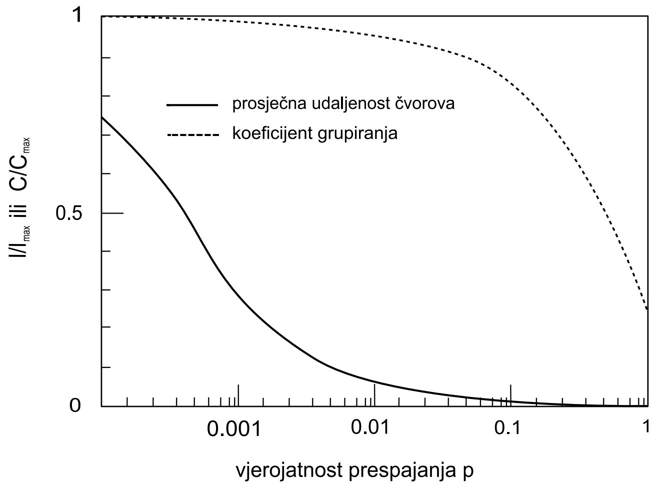
(a) Počinje se sa 1-dimenzionalnom rešetkom od $\mathcal{L}$ čvorova od kojih svaki čvor ima $z \geq 4$ najbližih susjeda.

(b) Svaka veza ima vjerojatnost p da će se odspojiti i spojiti na dva nasumično odabrana čvora.



Slika 2: Model Watts-Strogatza sa prespajanjem veza.

Općenito, prosječni najkraći put $\bar{\ell}$ je vrlo osjetljiv na prečace. To znači da će se karakteristike mreže značajno promijeniti već ako vrijedi $pz\mathcal{L} \sim 1$ tj. ako je u prosjeku tek jedan prečac u mreži. Pri tome se ostale karakteristike mreže, prije svega koeficijent grupiranja, ne mijenjaju bitno.



Slika 3: Raspodijela prosječne najkraće udaljenosti $\bar{\ell}$ i koeficijenta grupiranja.

Usporedimo li raspodijelu prosječne najkraće udaljenosti $\bar{\ell}$ i koeficijenta grupiranja sa slike 3 primjećujemo da postoji stanoviti interval vjerojatnosti prespajanja p gdje je prosječna najkraća udaljenost višestruko smanjena dok ostale značajke mreže (prije svega koeficijent grupiranja) nisu bitno promijenjene.

Naravno da slični zaključi vrijede i za općenite n-dimenzionalne rešetke. Naknadno su razvijeni i modeli koji umjesto prespajanja starih veza stvaraju nove. Glavne karakteristike mreža dobivenih u jednom ili drugom modelu ne razlikuju se bitno.

4 Barabási-Albert model

Barabási-Albert model je danas jedan od najproučavanijih modela generiranja mreža bez skale. Ono što su oni napravili zapravo se nastavlja na rad D. S. Pricea koji je 1965. [8] prvi opisao ono što bi se danas nazvalo mrežom bez skale.

Price je proučavao mrežu citiranja znanstvenih članaka i pokazao da njene ulazne kao i izlazne distribucije stupnjeva imaju polinomnu distribuciju. To je jedan od glavnih razloga zašto se njegov i modeli zasnovani na njegovom danas nazivaju mreže citata. U takvim modelima čvorovi se dodaju u mrežu sa početnim stupnjem m te se spajaju na m starih čvorova[5]. Spajanje je preferencijalno.

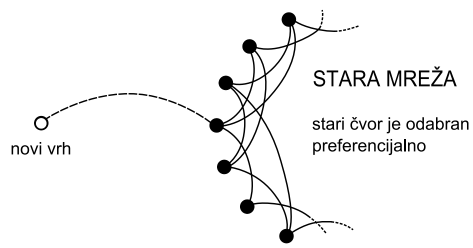
Razlika između modela je ta što Priceov model opisuje *usmjerene* mreže čije veze imaju smjer pa mogu ulaziti ili izlaziti iz čvor, a Barabási-Albertov model *neusmjerene* mreže gdje ta distinkcija ne postoji. Ta razlika je ključna promatraju li se stvarne mreže poput Interneta ili WWW-a u kojima se takvo usmjerenje veza pojavljuje - jedna web stranica može imati link prema drugoj bez obzira ima li ova druga link prema njoj. Čak je i mreža citata, primjer koji je proučavao Price, ustvari usmjerena mreža. Ipak, neke su se aproksimacije trebale učiniti radi pojednostavljenja modela.

U nastavku rada neću detaljnije razrađivati Priceov model. Umjesto toga koncentriram se isključivo na Barabási-Albert model.

Algoritam za generiranje mreže prema Barabási-Albert modelu glasi:

(a) U svakom koraku novi čvor dolazi u mrežu.

(b) Spaja se na stari čvor s vjerojatnošću proporcionalnom stupnju čvora.



Slika 4: Dodavanje novog čvora u Barabási-Albert modelu.

Kako se u svakom vremenskom koraku u mrežu dodaje samo jedan čvor i jedna veza prosječan stupanj mreže je $\overline{k} = 2$. Stoga će suma svih stupnjeva u vremenskom trenutku t biti $t\overline{k} = 2t$. Vjerojatnost da čvor stupnja k dobije novu vezu je $\frac{k}{2t}$, vjerojatnost da ostane istog stupnja je $1 - \frac{k}{2t}$. Jednadžba za distribuciju stupnjeva (pojedinačnih čvorova!) glasi

$$p(k, s, t+1) = \frac{k-1}{2t} p(k-1, s, t) + \left(1 - \frac{k}{2t}\right) p(k, s, t)$$

Objasniti ću ukratko značenje jednadžbe. Postoje dvije mogućnosti da čvor u koraku $t+1$ (tj. novom koraku) bude stupnja k:

[5] Zamijetite da se u takvom modelu ne mijenjaju već postojeće veze između starih čvorova. To možda dobro opisuje mreže citata (gdje već postojeći članci ne mogu mijenjati svoje reference) ali ne i moderne stvarne mreže poput Interneta i WWW-a

(1) Prva mogućnost je da je u prošlom koraku bio stupnja $k-1$ i dobio novu vezu. To je prvi član desne strane jednadžbe.

(2) Druga mogućnost je da je u prošlom koraku bio stupnja k i da *nije* dobio novu vezu. To je drugi član desne strane jednadžbe.

Ne upuštajući se u formalni dokaz (koji se može pogledati u [1]) odmah navodim konačnu jednadžbu za distribuciju stupnjeva:

$$P(k) = \frac{4}{k(k+1)(k+2)}$$

Za velike stupnjeve to daje polinomnu distribuciju s eksponentom $\gamma = 3$. Znači

$$P(k) \sim k^{-3}$$

Programsko ostvarenje Barabási-Albert modela

U teoriji, za generiranje mreže po Barabási-Albert modelu potrebna nam je samo neka početna mreža od N_0 čvorova na koju ćemo dodavati nove čvorove. Ipak, za programsko ostvarenje potrebno je odrediti i konačan broj čvorova N koji će sačinjavati našu mrežu. U svakom koraku (iteraciji) ćemo dodavati jedan novi čvor i vezu pa će broj čvorova koje moramo dodati biti $N - N_0$.

Za reprezentaciju mreže možemo koristiti *matricu susjedstva*[6], no kako nas zanima samo distribucija stupnjeva a ne i međusobne veze između čvorova ovdje ćemo je izostaviti i umjesto nje koristiti samo jedno 1-dimenzionalno polje k u koje ćemo pohranjivati stupanj svakog vrha.

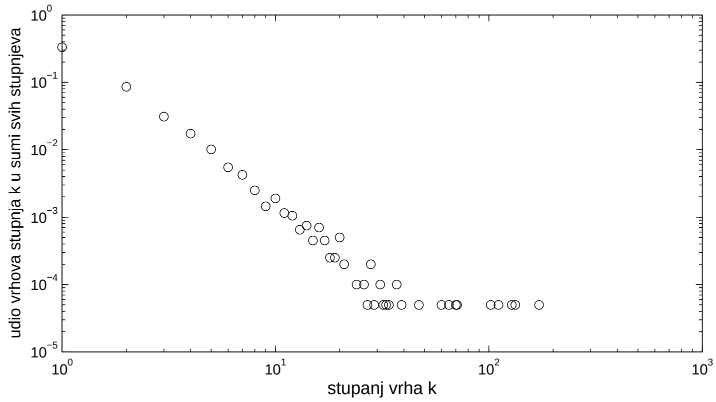
Pseudokod za programsko ostvarenje glasi:

```
učitaj početni broj čvorova N0
učitaj konačni broj čvorova N
početnih N0 čvorova spoji u potpunu[7] mrežu
ponavljaj od 1 do N − N0

    odaberi preferencijalno jedan od starih čvorova
    novom čvoru dodijeli stupanj 1
    odabranom čvoru povečaj stupanj za 1

ponavljaj od i=1 do i=N

    izbroji koliko ima čvorova stupnja i i spremi podatak
    u distribucija[i]

polje distribucija podijeli sa sumom svih stupnjeva mreže
```

Polje *distribucija* sadrži podatke za crtanje distribucije stupnjeva. Na k-tom elementu nalazi se podatak o tome koliki je udio čvorova stupnja k u sumi $\sum_N k_i$ svih stupnjeva mreže. Ovako koncipirani program može se implementirati u bilo kojem programskom jeziku sa mogućnošću grafičkog prikaza podataka (na primjer Matlab).

[6] Matrica susjedstva jednoznačno određuje mrežu. Reda je $N \times N$ gdje je N ukupni broj čvorova mreže. Element a_{ij} označava broj veza između čvorova s_i i s_j

[7] Mrežu u kojoj je svaki čvor spojen sa svim ostalim čvorovima.



Slika 5: Primjena Barabási-Albert algoritma na početnu mrežu od $N_0 = 4$ čvorova. Konačni broj čvorova N je 10000.

Distribucija stupnjeva takve mreže vidi se na slici 5. Dobivena mreža je bez skale i distribucija stupnjeva raspodijeljena je po polinomnom zakonu. Na dnu grafikona primjećuje se *cut-off efekt*. On je posljedica činjenice da su stvarne mreže (pa tako i naš simulirani model) ipak mreže konačne veličine. Tako će uvijek postojati određeni (mali) broj čvorova različitog stupnja kojih će u mreži biti jednako.

Kod nas su za to krivi čvorovi velikog stupnja ($k \sim 10^2$) kojih je u mreži jako malo - samo jedan ili dva. Kad bi im se omogučilo da dalje rastu, na primjer da se poveća broj iteracija, cut-off efekt bi se pomaknuo na čvorove većeg stupnja.

5 Preferencijalno povezivanje

Preferencijalno povezivanje opisano ovdje je zapravo generalizacija Barabási-Albert modela. Osnovna premisa preferencijalnog povezivanja je da čvorovi većeg stupnja imaju veću vjerojatnost da se spoje na novopristigle veze. Često se za ovaj princip kaže *popularnost je privlačna*[8] - čvorovi s više veza su preferirani.

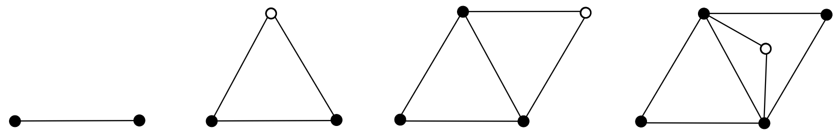
Općenito, preferencijalno povezivanje se definira kao povezivanje gdje je vjerojatnost spajanja na novi vezu određena nekom funkcijom od k, tj. preferencijalnom funkcijom $f(k)$[9]. Distribucija bez skale pojavljuje se samo ako je funkcija $f(k)$ *linearna*, tj. oblika $\frac{k+A}{\langle k\rangle+A}$ gdje je A neka konstanta a $\langle k\rangle$ suma svih stupnjeva u mreži. To se čini ispravnim za većinu stvarnih mreža.

Kako nastaje preferencijalno povezivanje?

U prethodnim poglavljima preferencijalno povezivanje u modelima rasta podrazumijevalo se kao dobitna metoda za dobivanje mreža bez skale. Ad hoc objašnjenje da se ono dobro slaže sa uobičajenim predodžbama nastanka realnih mreža (popularnost je privlačna) nadomjestilo je detaljnije opravdanje ovakvog načina povezivanja. Ukratko ću izložiti osnovnu ideju...

Osnovno pitanje koje se ovdje postavlja je - zašto baš stupanj čvora, a ne neko drugo svojstvo, igra bitnu ulogu u preferencijalnom povezivanju? Odgovor na njega leži u činjenici da su zapravo *veze*, a ne čvorovi, te koje su ključne u stvaranju novih veza, tj. u privlačenju novih čvorova.

Tako jednostavnim prihvaćanjem pretpostavke da novi čvorovi traže krajeve veza na koje bi se spojili lako objašnjavamo preferencijalno povezivanje. Stupanj čvora jednak je broju veza koje su na njega spojene. Kako se novi čvorovi zapravo spajaju na veze, što neki čvora ima više veza veće su šanse da će (nasumičnim odabirom) odabrati baš jednu od njegovih veza, što ujedno rezultira i nihovim povezivanjem. Ukratko, *nasumično* spajanje na veze ekvivalentno je *preferencijalnom* spajanju na čvorove.



Slika 6: Mreža raste pripajajući nove čvorove na oba kraja nasumično odabrane veze.

Koristeći ovu ideju, možemo konstruirati mrežu bez skale sa sličnim svojstvima kao u Barabási-Albert modelu:

(1) U početku postoje samo dva čvora povezana jednom vezom.

(2) U svakom vremenskom koraku dodaje se jedan novi čvor.

(3) Spaja se na oba kraj nasumično odabrane veze s dvije nove veze.

[8]Postoje i druge interpretacije, na primjer *bogatiji postaju bogatiji* ili *pobjednici dobivaju sve*.

[9]To vrijedi samo za homogene mreže. U nehomogenim mrežama preferencijalna funkcija $f(k)$ varira od čvora do čvora.

Dodatno svojstvo ovako konstruirane mreže je što ima veliki koefeicijent grupiranja. Kako se svaki novi čvor spaja na staru vezu sa dvije nove veze, mreža se praktički sastoji od trokutova (tripleta čvorova). Podsjetimo se da je koeficijent grupiranja kod Barabási-Albert mreže jako mali[10], što znatno odudara od realnih mreža. Dapače, koeficijent grupiranja Barabási-Albert mreže teži k nuli kako se povećava veličina mreže.

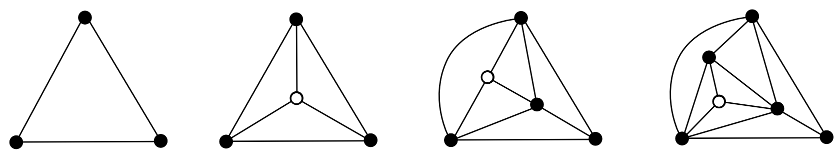
Jedan od načina da se zaobiđe ova poteškoća, tj. da se poveća koeficijent grupiranja, prikazan je u sljedećem poglavlju.

Postizanje velikog koeficijenta grupiranja

Grupiranje je, uz stupanj-stupanj korelaciju (*degree-degree correlation*), najvažnija vrsta korelacije među čvorovima mreža. Nastojanje da se postigne što veći koeficijent grupiranja ima svoje opravdanje u činjenici da je on relativno velik i kod mnogih stvarnih mreža. Postoje brojni načini da se konstruiraju mreže sa takvim svojstvom. Prijašnja ideja, ona sa spajanjem na veze umjesto na čvorove, može se varirati na još načina, koji svi rezultiraju mrežama bez skale. Jedan od mogućih načina je sljedeći:

(1) Početna konfiguracija je triplet čvorova.

(2) Svaki vremenski korak dodaje se jedan čvor koji se sa tri veze spaja na nasumično odabran triplet čvorova.



Slika 7: Rastuća mreža bez skale sa velikim koeficijentom grupiranja. Novi čvorovi se spajaju na nasumično odabran triplet čvorova.

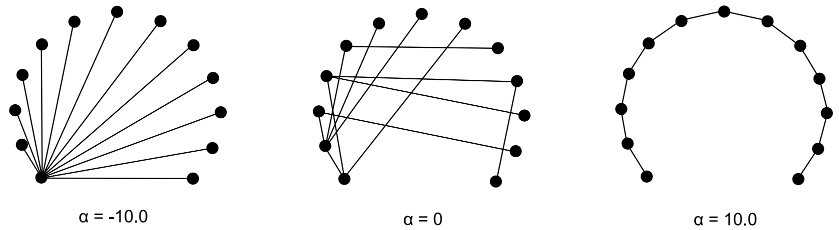
Ovakva mreža ima polinomnu distribuciju stupnjeva sa koeficijentom $\gamma = 3$. Koeficijent grupiranja je relativno velik i iznosi $C = \frac{1}{2}$.

Još jedan primjer linearne preferencije

Iako samo linearna preferencija dovodi do distribucija bez skale, postoji mnogo vrsta linearnih preferencija. U prijašnjim poglavljima kao mjeru privlačnosti uzimali smo stupanj čvora k. No, kao što smo već naslutili, ne postoji razlog da se kao alternativa ne uzme i neki drugi parametar.

Na slici 8 razmotren je slučaj kada se za parametar privlačnosti uzima starost čvora τ, čineći starije ili mlađe čvorove više ili manje poželjnima. Ovakav pristup dobro korespondira s realnim mrežama gdje starost čvorova igra veliku ulogu u stvaranju novih veza. Iako općenito nije pravilo, vjerojatnost da će netko stvoriti link prema nekoj web stranici ili citirati znanstveni članak svakako opada sa starošću te stranice i članka.

[10] Uostalom, kao i kod svih mreža citata. Teoretski, koeficijent grupiranja za svaku mrežu citata u kojem se novi čvor spaja samo s jednim starim čvorom jednaka je nuli.



Slika 8: Struktura mreže sa starenjem čvorova. Vjerojatnost da će se novi čvor spojiti na stari proporcionalna je sa $\tau^{-\alpha}$, gdje je τ starost čvora. Prikazane su strukture mreža s različitim parametrom α. Za $\alpha = -10.0$ svi novi čvorovi spajaju se na najstariji čvor. Za $\alpha = 10.0$ svi novi čvorovi spajaju se na najmlađi čvor - dobivena mreža je linearna.

6 Alternativni modeli konstrukcije mreža bez skale

Dosad sam razmatrao isključivo *stohastičke* modele nastanka kompleksnih mreža. Pri tome se misli da su takvi modeli bili ovisni o nekim slučajnim procesima - na primjer nasumičnim spajanjem novih veza. Tako stvorene mreže zapravo su statističke skupine (eng. *statistical ensembles*) mreža koje, iako dobivene istim postupkom, nisu identične u svakom pogledu. Nadalje, nehotice sam dopustio da se dobije dojam da su samo rastući modeli dobri kandidati za konstrukciju mreža bez skale.

U nastavku ću razmotriti jedan primjer kada je rast mreže strogo *deterministički*. Prednost takvih modela je što se mnogo preciznije mogu definirati neke karakteristike koje u slučajnim mrežama (bilo ravnotežnim ili neravnotežnim) izbjegavaju podrobniju analizu.

U drugom dijelu ću iznijeti primjer mreže koja se dobiva "statičnom" metodom, no koja je bez obzira na to bez skale.

Deterministički model rasta

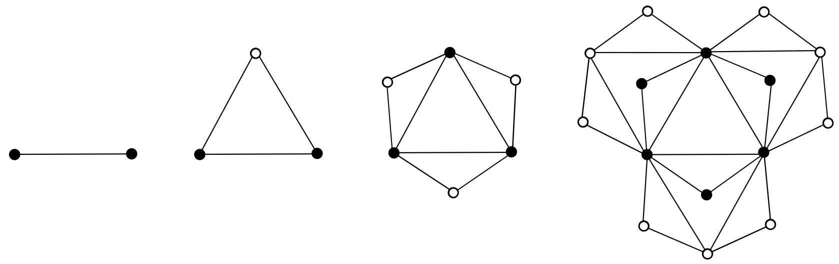
Definirajmo za početak pravilo za konstrukciju *determinističkih mreža*:

> Svaki vremenski korak svaka veza transformira se na isti način s vjerojatnošću $p = 1$.

Prisjetimo se da je u prijašnjim primjerima kompleksnih mreža svaki čvor na neki način "proizveo" nove veze. Ovdje se stvar okreće. Kao što smo i vidjeli iz gore navedenog pravila, u slučaju determinističkih mreža svaka veza na neki način "proizvodi" nove čvorove. Početnih par koraka mogu se vidjeti na slici 9. Rezultat nalikuje na neku vrstu fraktala, ali to nije pa otuda i naziv *pseudofraktal*.

Zašto to nije fraktal? Iako na prvi pogled nalikuje na klasične primjere fraktala, na primjer Kochovu pahuljicu, osnovna razlika je u tome što se u ovoj strukturi prosječne najkraće udaljenosti između starih čvorova ne mijenjaju. Struktura mreže nema fiksnu konačnu fraktalnu dimenziju i prema tome uopće nije fraktal!

Ako pažljivije pogledamo naš model, zamijetiti ćemo da i ovdje vrijedi pravilo preferencijalnog spajanja - čvorovi većeg stupnja dobivaju više novih veza. Time



Slika 9: Rast pseudofraktala bez skale. Svaki vremenski korak svaka veza dobiva novi čvor koji se na njega spaja sa dvije veze.

je ispunjen osnovni uvjet za stvaranje mreže bez skale, što ovakva mreža uistinu i je.

Kao što je već napomenuto, prednost ovako stvorenih mreža je što se neke njihove karakteristike mogu izračunati vrlo precizno. Na primjer, za pseudofraktal na slici 9 izračunata je distribucija najkraćih puteva između čvorova. Asimptota te distribucije za mreže velike veličine je

$$P(\ell, t) \simeq \frac{1}{\sqrt{2\pi(2^2/3^3)t}} exp\left[-\frac{(\ell - \overline{\ell}(t))^2}{2(2^2/3^3)t}\right]$$

Statične neravnotežne mreže

Za kraj navodim primjer konstrukcije jednostavne mreže koja je bez skale ali za čije stvaranje se ne koristi model rasta. Umjesto toga mreža je statična - ima fiksni broj čvorova N i neusmjerenih veza L. Prosječan stupanj čvora je $\overline{k} = \frac{2L}{N}$. Pravila za evoluciju takve mreže su:

(1) U svakom koraku nasumično odabran čvor gubi sve svoje veze.

(2) Te veze se prespajaju na nove preferencijalno odabrane čvorove. Preferencija je linearna: vjerojatnost da se veza prespoji na čvor stupnja k je proporcijalna $k + A$, gdje je A neka konstanta.

Ovako stvorena mreža ekivalentna je onima dobivenim modelom rasta. Jedina razlika je u tome što se parametar m - početni stupanj svakog novopristiglog čvora zamjenjuje s $\overline{k}$ - prosječnim stupnjem cijele mreže. Distribucija stupnjeva u tom slučaju je polinomna sa eksponentom $\gamma = 2 + \frac{A}{\overline{k}}$.

Zaključak

Pojam *mreže* danas je vrlo uobičajen u gotovo svim granama znanosti i tehnike, ali i u svakodnevnom žargonu. Kao iznimno generalizirani model njime se uspješno modeliraju brojne realne strukture - Internet i WWW, socijalne interakcije između ljudi, metaboličke interakcije u organizmima, suodnosi i utjecaji u pisanim djelima...

Proučavanje mreža zadnjih par stotina godina prije svega je bila domena teorije grafova koja je tamo napravila brojne pomake. Ipak, krajem 20. stoljeća metode se mijenjaju. Od Priceovih skromnih pokušaja indeksiranja citata u znanstvenim člancima u [8] informacijska tehnologija uznapredovala je dovoljno da omogući pojedincima da brzo i pouzdano dođu do empirijskih podataka za realne mreže velikih veličina. Usporedo s njima razvijaju se nove teorije koje objašnjavaju rast i evoluciju takvih mreža.

Takvi modeli se u mnogočemu razlikuju od onih koje je razvijala klasična teorija grafova. Nova znanost je od stare teorije preuzela terminologiju, no izmijenila je pristup i metode istraživanja. Rezultat je moderna teorija kompleksnih mreža koja se intenzivno razvija zadnjih desetak godina, dovoljno interdisciplinarna da ju sa jednakim žarom prihvate fizičari, kemičari, biolozi, sociolozi...

Iako danas postoje brojni popularni pravci istraživanja, prije svega istraživanja o načinima širenja zaraza u mrežama, koji imaju potencijalno unosnu praktičnu primjenu (na primjer antivirusni programi), još uvjek nije u potpunosti razjašnjen nastanak tih kompleksnih struktura. Uočavanjem i sistematiziranjem pravila pomoću kojih realne mreže vrše samoorganizaciju dopridonijelo bi razvitku cijelog područja.

Uostalom, kao što Dorogovtsev i Mendes kažu u uvodnom poglavlju svoje knjige "Evolution of Networks" izdanu prije samo koju godinu:

> *The field is open, and there are future challenges ahead.*

Literatura

[1] S. N. Dorogovtsev, J. F. F. Mendes: *Evolution of Networks*, Oxford, 2003.

[2] S. N. Dorogovtsev, J. F. F. Mendes: *The shortest path to complex networks*, 24. srpnja 2004., arXiv:cond-mat/0404593 v4, 27. travnja 2007.

[3] A. L. Barabási and R. Albert: *Emergence of scaling in random networks*, Science 286, 509.

[4] M. E. J. Newman: *The structure and function of complex networks*, 25. ožujka 2003., arXiv:cond-mat/0303516v1, 27. travnja 2007.

[5] D. J. Watts and S. H. Strogatz: *Collective dynamics of small-world networks*, Nature 393, 1998.

[6] P. Erdös and A. Rényi: *On random graphs*, Publ. Math. Debrecen 6, 290., 1959.

[7] P. Erdös and A. Rényi: *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci. 5, 17., 1960.

[8] D. J. de S. Price: *Networks of scientific papers*, Science 149, 510-515, 1965.

Sažetak

U seminaru se definiraju najvažniji pojmovi vezani uz analizu kompleksnih mreža - distribucija stupnjeva i grupiranje. Ukratko se, bez ulaženja u detaljniju matematičku analizu predstavljaju osnovni modeli klasičnih i kompleksnih mreža, njihove karakteristike i svojstva te metode konstrukcije. Detaljno se razrađuju modeli ravnotežnih i neravnotežnih mreža te mreža bez skale i malih svijetova. Opisuje se Barabási-Albert model te njegova programska implementacija. Uvodi se pojam preferencijalnog povezivanja i objašnjava se njegova važnost u konstrukciji mreža bez skale. Za kraj se predstavljaju i neki alternativni modeli konstrukcije kompleksnih mreža - deterministički model rasta i statične (nerastuće) neravnotežne mreže.

Kao opravdanje apstraktnih modela pružaju se i brojni primjeri iz stvarnog svijeta, prije svega empirijska istraživanja svojstava realnih mreža kao što su Internet i WWW te citiranje u znanstvenim časopisima.