SVEUČILIŠTE U ZAGREBU

FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

Određivanje središnjih čvorova u kompleksnim mrežama i primjene

Katarina Kušević

Voditelj: *Doc.dr.sc. Mile Šikić*

Zagreb, travanj, 2010.

Sadržaj

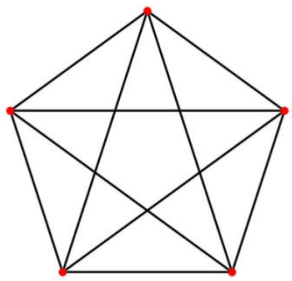
1. Uvod

Kompleksne mreže su sve mreže koje nisu klasične mreže. To znači da nemaju distribuciju Poissonovog tipa s nasumično distribuiranim vezama između čvorova, nego su te distribucije kompleksne. Kompleksne se mreže koriste za objašnjavanje podataka iz stvarnog svijeta koji su dobiveni empirijski. Pri istraživanju kompleksnih mreža važno je odrediti središnje čvorove mreže, jer oni uvelike određuju svojstva mreže i objašnjavaju razne pojave, kao što su širenje zaraza u kompleksnim mrežama. U svom ću radu objasniti osnove teorije grafova, pojasniti što su kompleksne mreže i koje su vrste kompleksnih mreža, razraditi kako se može odrediti koji su čvorovi središnji i istražiti u kojim slučajevima je važno znati središnje čvorove mreže, koje su poveznice čvorova i koje su primjene kompleksnih mreža i određivanja središnjih čvorova. Neke od tih primjena su poznate cijelom svijetu i služimo se njima svakodnevno.

2. Osnove teorije grafova

Za lakše razumijevanje kompleksnih mreža, koje se modeliraju grafovima, objasnit ću nekoliko pojmova vezanih za grafove. [2]

Jednostavan graf G sastoji se od nepraznog konačnog skupa V(G), čiji elementi su vrhovi (*čvorovi*) grafa G i konačnog skupa E(G) različitih dvočlanih podskupova skupa V(G) koji zovemo bridovi (*veze*). Definicija jednostavnog grafa osigurava da su dva čvora povezana samo jednim bridom. Ako brid *e* spaja vrhove *v* i *w*, kažemo da su i vrh *v* i vrh *w* incidentni s bridom *e*, te da su vrhovi *v* i *w* susjedni. Neke od vrsta grafova su potpuni graf (Slika 2.1), u kojem je svaki vrh povezan sa svim ostalim vrhovima, i zvijezda graf u kojem je jedan čvor povezan sa svim ostalim čvorovima, a oni niti sa jednim drugim čvorom.



Slika 2.1 Potpuni graf K_5

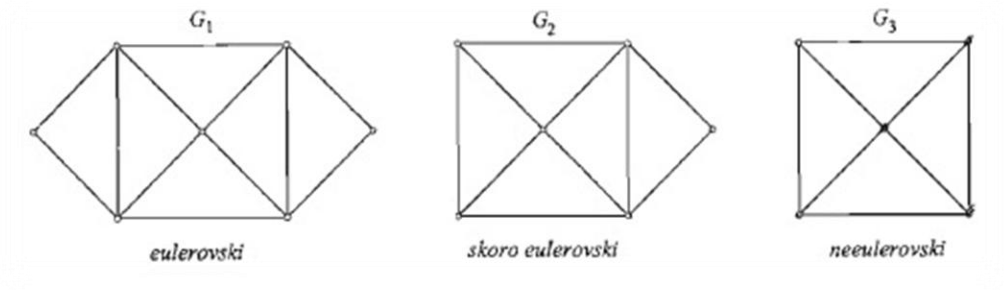
Stupanj vrha *v* (*deg(v)*) grafa *G* je broj bridova koji su incidentni s *v*. Svakom grafu *G* možemo pridružiti niz stupnjeva. Ako graf ima *n* vrhova, niz stupnjeva je niz od *n* članova koji se sastoji od rastućeg niza cijelih brojeva koji predstavljaju stupnjeve svih vrhova grafa *G*. Leonhard Euler je 1736. godine dokazao da je zbroj stupnjeva svih vrhova paran broj (*lema o rukovanju*). Graf u kojem su svi vrhovi istog stupnja *r*, kažemo da je *r*-regularan. U svakom grafu je broj vrhova neparnog stupnja paran.

Za svaki graf možemo konstruirati matricu susjedstva i matricu incidencije. Ako označimo vrhove grafa *G* s $V=\{1,2,\ldots,n\}$, tada definiramo matricu susjedstva $A=[a_{ij}]$ kao $n*n$ matricu čiji je element a_{ij} jednak broju bridova koji spajaju vrh *i* s vrhom *j*. U matrici susjedstva zbroj elemenata svakog stupca (ili retka) odgovara stupnju odgovarajućeg vrha. Ukoliko označimo i bridove grafa *G* kao *E={1,2,…,m}*, definiramo matricu incidencije $B=[b_{ij}]$ kao $n*m$ matricu čiji su elementi b_{ij} jednaki 1 ako je vrh *i* incidentan s vrhom *j*, ili su jednaki 0 ako nisu incidentni. Matrica incidencije u svakom stupcu ima na samo dva mjesta 1, i te jedinice pokazuju koja su dva vrha spojena tim bridom. Matrica incidencije ovisi o numeraciji bridova, ali ona jednoznačno određuje graf.

Šetnju u grafu *G* je konačan slijed bridova u kojem su svaka dva uzastopna brida ili jednaka ili susjedna. Staza je šetnja u kojoj su svi bridovi različiti. Put je staza u kojoj su i svi vrhovi različiti (osim eventualno početnog i krajnjeg). Stazu ili put kojem je početni vrh jednak krajnjem nazivamo zatvorenim. Ako je put zatvoren i sadrži barem jedan brid, nazivamo ga ciklusom. Struk grafa *G* je duljina njegova najkraćeg ciklusa.

Ukoliko graf nema ciklusa, zovemo ga šumom, a povezanu šumu nazivamo stablo.

Povezan graf je eulerovski ukoliko postoji zatvorena staza koja sadrži sve bridove grafa. U takvom grafu stupanj svakog vrha mora biti paran. (Slika 2.2)



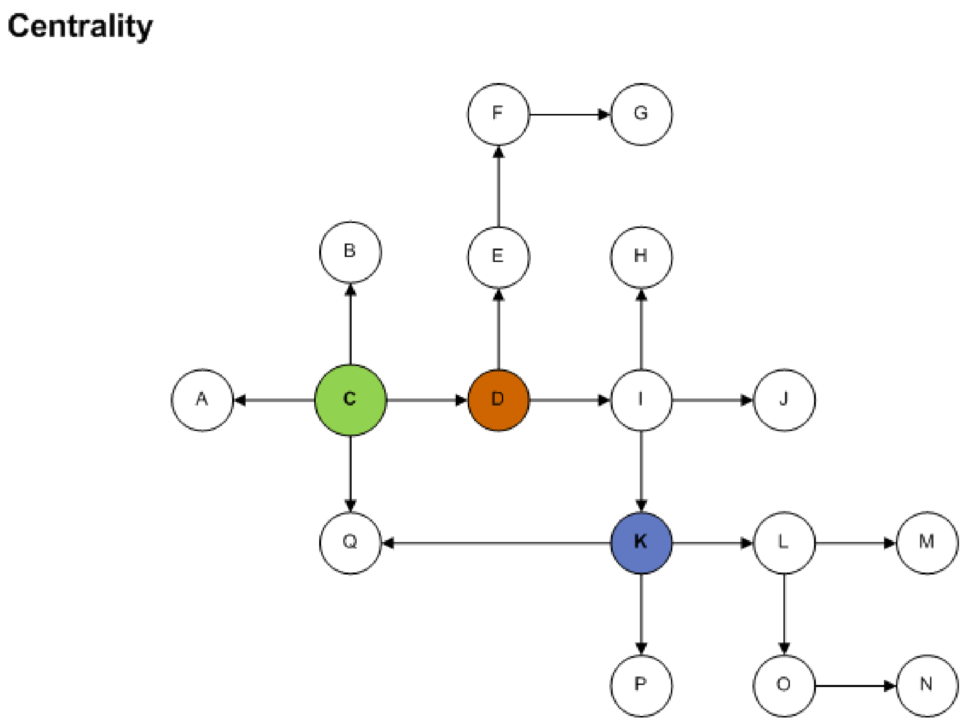
Slika 2.2 Eulerovski, skoro eulerovski i neeulerovski graf

Graf koji posjeduje ciklus koji prolazi svim vrhovima grafa zovemo hamiltonovski graf.

Za razliku od neusmjerenih grafova, kod kojih bridovi povezuju dva vrha u oba smjera, brid u usmjerenom grafu povezuje dva vrha, ali se točno zna koji je izvorišni, a koji odredišni, tj. brid vodi samo u jednom smjeru. Grafovi se dijele i na težinske i bestežinske; bridu težinskog grafa pridajemo neku vrijednost, npr. duljinu. U bestežinskom grafu svaki brid vrijedi jednako.

3. Kompleksne mreže

Kompleksne mreže su grafovi (mreže) sa topološkim značajkama koje se ne mogu naći u jednostavnim, slučajnim grafovima, kao što su komplicirane distribucije stupnjeva čvorova, hijerarhijska struktura grafa, visok koeficijent grupiranja čvorova i/ili podudaranje čvorova (Slika 3.1). Jednostavne, klasične mreže imaju Poissonovu distribuciju stupnjeva čvorova, dok kompleksne mreže većinom imaju Pareto distribuciju stupnjeva vrhova. Najvažnije karakteristike kompleksnih mreža su grupiranje čvorova (eng. *clustering*), distribucija stupnjeva čvorova i prosječna duljina puta. Pokušat ću objasniti ta obilježja. [3]



Slika 3.1 Socijalna mreža prikazana grafom: čvorovi C i K imaju najveći stupanj, tj. povezani su sa najviše čvorova, dok je čvor D most (nije jedini) ovog grafa

3.1. Koeficijent grupiranja

Koeficijent grupiranja, C, grafa je vjerojatnost da su dva slučajno odabrana vrha povezana uz uvjet da su oba povezana sa zajedničkim vrhom. Ako je k_i broj susjeda i-tog vrha, te l_i broj susjeda i-tog vrha koji su međusobno povezani, tada je koeficijent grupiranja dan formulom (3.1).

$$C = \frac{1}{n}\sum_{i=1}^{n}\frac{2l_i}{k_i(k_i-1)} \tag{3.1}$$

Grupiranje podrazumijeva okupljanje više čvorova oko jednog čvora, koji onda ima sve više veza s drugim čvorovima i takve čvorove nazivamo *hubovi.*

3.2. Distribucija stupnjeva čvorova

Distribucija stupnjeva čvorova grafa *P(k)* je vjerojatnost da slučajno izabran vrh ima k susjeda. Ako je *k(i)* stupanj vrha *i*, te *n(k)* broj vrhova stupnja *k*, tada vrijedi (3.2).

$$P(k) = \frac{n(k)}{n} \tag{3.2}$$

Pareto distribucija je distribucija bez skale, koja opisuje distribuciju stupnjeva u mrežama koje imaju *hubove* (γ je parametar mreže). Distribucija stupnjeva (3.3) prisutna u mrežama bez skale.

$$P(k) = k^{-\gamma} \tag{3.3}$$

3.3. Prosječna duljina puta

Prosječna duljina puta je prosječni najkraći put između dva slučajno odabrana vrha (ukupni broj vrhova je *n*). Ako je graf G dan s *n* vrhova, d_{ij} je najkraći put između vrhova v_i i v_j, prosječna duljina puta u mreži dana je formulom (3.4). Što je prosječna duljina puta kraća, to je brže širenje informacija kroz mrežu.

$$D = \frac{2}{n(n-1)}\sum_{i\geq j} d_{ij} \tag{3.4}$$

3.4. Primjeri kompleksnih mreža

Najpoznatiji primjeri kompleksnih mreža su mreže bez skale (eng. *scale-free networks*) i mreže malog svijeta (eng. *small-world networks*).

3.4.1. Mreže malog svijeta

Primjer mreže malog svijeta je pokus Stanleya Milgrama, socijalnog psihologa iz 1960-tih, kojim pokazuje da su sve osobe na svijetu povezane u prosjeku preko 6 drugih ljudi (engl. *Six Degrees of Separation*). Ukoliko prosječna duljina puta kompleksne mreže (D) raste sporije od nego pozitivna potencija broja najbližih susjeda (N) tada za mrežu kažemo da je mreža malog svijeta.

3.4.2. Mreže bez skale

Mreže bez skale imaju polinomnu distribuciju stupnjeva, najčešće Pareto distribuciju, i takve su mreže na primjer World Wide Web, mreže proteinskih interakcija u metabolizmu itd. Ove mreže imaju prisutne *hubove*, čvorove koji imaju stupanj puno veći od prosječnog stupnja ostalih čvorova u mreži.

4. Određivanje središnjih čvorova u kompleksnim mrežama

U analizi kompleksnih mreža često se postavlja pitanje koji je čvor središnji. Odgovor na to pitanje ovisi o kakvom se središtu radi, o kakvoj je mreži riječ. Najčešće je čvor koji ima najviše veza središnji, ali za određivanje središnjih čvorova u kompleksnim mrežama, tj. analizu mreža, postoji nekoliko mjera koje ću opisati.[1]

3.5. Stupanj čvorova

Stupanj čvorova (engl. *degree centrality*) je mjera definirana kao broj bridova koji su incidentnim s čvorom, tj. broj direktnih veza koje neki čvor ima . Za svaki usmjereni graf postoje tri mjere utvrđivanja stupnjeva čvorova; prema ulaznom stupanju (eng. *in-degree*) koji broji broj veza usmjerenih u čvor, prema izlaznom stupnju (eng. out-*degree*) koji daje broj veza usmjerenih iz čvora i prema ukupnom stupnju (eng. *total degree)* koji broji ukupni broj veza u čvor i iz čvora. Npr. za pozitivni odnos kao što je prijateljstvo, ulazni stupanj je popularnost, a izlazni društvenost.

Za čvor *v* grafa *G* sa *n* čvorova stupanj centralnosti je omjer stupnja čvora i broja ostalih vrhova grafa (4.1). Taj broj možemo očitati i iz matrice susjedstva grafa, kao zbroj elemenata matrice u pojedinom stupcu (4.2). Složenost računanja stupnja centralnosti (C_D) za sve čvorove (V) ima složenost $\Theta(V^2)$.

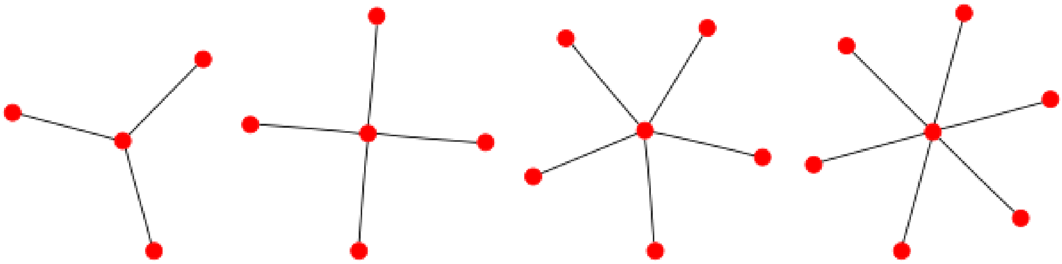
$$C_D(v) = \frac{\deg(v)}{n-1} \tag{4.1}$$

$$\mathrm{c}_{\mathrm{i}}^{\mathrm{DEG}} = \sum_{\mathrm{j}} a_{ij} \tag{4.2}$$

Definiciju centralnosti možemo proširiti i na cijeli graf, ako čvor s najvećim stupnjem centralnosti u grafu *G* označimo s *v**, neka je graf *G* s *n* čvorova povezan graf u kojem je razlika između stupnja centralnosti vrha *v** i svakog drugog vrha tog grafa najveća (4.3). Tada je stupanj centralnosti grafa *G* definiran sa (4.4).

$$H = \sum_{j=1}^{|V'|} [C_D(v'*) - C_D(v'_j)] \tag{4.3}$$

$$C_D(G) = \frac{\sum_{i=1}^{|V|}[C_D(v*) - C_D(v_i)]}{H} \quad (4.4)$$



Slika 4.1 Grafovi zvijezde s 4,5,6 i 7 vrhova

Kod zvijezda grafa (Slika 4.1) H je maksimalan jer graf *G* sadrži jedan čvor povezan sa svim ostalim čvorovima, a ostali čvorovi su povezani samo na taj centralni čvor. Tada vrijedi (4.5).

$$H = (n-1)\left(1 - \frac{1}{n-1}\right) = n - 2 \quad (4.5)$$

3.6. Blizina čvorova

Udaljenost d_G između dva vrha *v* i *t* je minimalna duljina putova koji ih povezuju. Blizina čvorova (engl. *closeness centrality*) se definira kao prosječan broj koraka koji nam je potreban da od jednog vrha grafa dođemo do svih ostalih čvorova koji su mu u dometu (4.6). Ukoliko imamo matricu udaljenosti nekog grafa, možemo zbrojiti udaljenosti među vrhova i na taj način dobiti ovu mjeru udaljenosti (4.7). Ako na primjer, neki čvor ima manje veza sa ostalim čvorovima, a ipak mu je prosječan broj do svih ostalih čvorova malen, taj čvor je blizu svih čvorova, informacija brzo putuje do svih čvorova i ima veliku preglednost mreže.

$$C_C(v) = \frac{1}{\sum_{t \in V \setminus v} d_G(v,t)} \quad (4.6)$$

$$c_i^{CLO} = \sum_j d_{ij} \quad (4.7)$$

3.7. Međupoloženost čvorova

U svakom grafu važno je znati kroz koji čvor prolazi najviše najkraćih putova u grafu. Takvi bridovi su mostovi grafa i ukoliko maknemo čvor kojim prolazi most, graf će se podijeliti na nepovezane podgrafove. Središnji čvorovi su dijelovi velikog broja najkraćih putova grafa. Za graf *G:=(V,E)* sa *n* čvorova, međupoloženost čvora (engl. *betweenness centrality*) *v* dana je formulom (4.8), gdje je σ_{st} broj najkraćih putova iz *s* do i *t*, a $\sigma_{st}(v)$ je broj najkraćih putova od *s* do *t* koji prolaze vrhom *v*.

$$C_V(v) = \sum_{\substack{s \neq v \neq t \in V \\ s \neq t}} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{4.8}$$

Izračun međupoloženosti i blizine svih čvorova u grafu uključuje računanje najkraće staze između svih parova vrhova u grafu. To računanje je složenosti $\Theta(V^3)$.

3.8. Svojstveni vektor čvorova

Određivanje središnjih čvorova svojstvenim vektorom (engl. *eigenvector centrality*) je mjera važnosti čvora u mreži. Ova se mjera čvoru dodjeljuje u odnosu na rezultate svih čvorova u mreži tako da se veza do čvorova veće važnosti pridonosi rezultatu čvora više nego jednaka veza do čvora manje važnosti. Ovakva se mjera određivanja središnjih čvorova koristi u Google Page Rank –u. Recimo da nam je matrica susjedstva grafa *G* označena sa $A_{i,j}$, *n* je ukupan broj čvorova, *M(i)* je skup čvorova povezanih sa *i*-tim čvorom, λ označava konstantu, tada je svojstveni vektor *i*-tog vrha x_i proporcionalan zbroju bodova svih čvorova (4.9).

$$x_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j = \frac{1}{\lambda} \sum_{j=1}^{n} A_{i,j} x_j \tag{4.9}$$

Ovaj vektor možemo zapisati i kao (4.10).

$$x = \frac{1}{\lambda} Ax \tag{4.10}$$

Postoji mnogo različitih svojstvenih vrijednosti λ za koje postoje svojstveni vektori. No, zahtjeva se da svojstveni vektor mora biti pozitivna vrijednost. Postoji jednostavan algoritam "Power iteration" za određivanje dominantne svojstvene vrijednosti λ i vektora *v* različitog od nule.

5. Primjene

Kompleksne mreže su rasprostranjene svud oko nas i primjenjuju se u objašnjavanju raznih pojava, kao što su proteinske interakcije u metabolizmu, širenje zaraza, socijalne mreže, Internet i bilo koje mreže koje možemo modelirati grafovima.

Primjer gdje se koriste metode određivanja središnjih čvorova kompleksne mreže je i "Google Page Rank" koji koristi "Power iteration" algoritam za računanje ranga stranice u tražilici (Slika 5.1)[16]. Google indeksira linkove sa neke internet stranice i linkove na tu internet stranicu i što veći Page Rank stranica ima, to će je tražilica postaviti bliže vrhu rezultatima pretrage. Tako Google radi istraživanje cijele strukture weba. Page Rank objektivno mjeri važnosti web-stranica rješavanjem jednadžbe s više od 500 milijuna varijabli i 2 milijarde pojmova. Tumače se veze između stranica i procjenjuje se važnost stranice prema broju glasova koje dobije. Tako se link sa stranice A na stranicu B boduje kao glas za stranicu B, od stranice A. U obzir se uzimaju i važnosti stranice koja daje glas, jer važnije stranice dobivaju viši Page Rank i pojavljuju se na vrhu rezultata pretraživanja.

Slika 5.1 Google Page Rank

Algoritam Google Page Rank-a osmislili su, i na njemu se obogatili, Sergey Brin i Larry Page (5.1). Page Rank je razdioba vjerojatnosti korištena da predstavi vjerojatnost da će osoba slučajnim klikom na linkove stići na bilo koju stranicu.

$$PR(A) = (1-d) + d * \left[\frac{PR(T_1)}{C(T_1)} + \cdots + \frac{PR(T_n)}{C(T_n)}\right] \tag{5.1}$$

U ovome algoritmu *PR(A)* je Page Rank stranice A, $PR(T_1)$ Page Rank stranice T_1, $C(T_1)$ je broj linkova sa stranice T_1, *d* je faktor prigušenja. Faktor prigušenja *d* je vjerojatnost da će osoba koja pretražuje Internet nasumičnim pritiskom na linkove, u bilo kojem trenutku, nastaviti dalje tražiti. Vrijednost tog faktora je oko 0,85 (a kao vjerojatnost uvijek je između 0 i 1). Algoritam iterativno prolazi kroz sve otvorene stranice i linkove, i tako se računa Page Rank svake web-stranice, te se stranica sa većim Page Rank-om postavlja više u rezultatima tražilice.

"KCores" je metoda kojom se računaju sve k-jezgre u grafu, tj. to je alat za identificiranje dobro povezanih struktura unutar grafova. K-jezgra se računa rekurzivno odbacivanjem čvorova grafa koji imaju stupanj čvora manji od *k*. Primjena k-jezgri prisutna je u biologiji, na primjer u predviđanjima funkcije proteina u interakcijama.

Metode za određivanje središnjih čvorova u kompleksnim mrežama koriste se i za razbijanje terorističkih organizacija, objašnjavanje interakcija imunizacija kod širenja zaraza, objašnjavanju marketinških pojava i sl.

6. Zaključak

Istraživanje kompleksnih mreža zanimljivo je područje u istraživanju teorija mreža. Određivanjem središnjih čvorova kompleksnih mreža dobivaju se informacije o mreži, čvorovima, poveznicama. Te se informacije mogu iskoristiti u razne svrhe. Kompleksne mreže je primjerice i mreža teorističke organizacije. Svi znamo da razbijanjem glavnog ili više glavnih čvorova takve organizacije počinje njezino uništavanje. Kompleksna mreža je svaka socijalna mreža, kao na primjer "Facebook" i "Twitter". Širenje epidemija se može objasniti kompleksnim mrežama, pogotovo ako se znaju središnji čvorovi tih mreža. Web-stranice se na Google-ovoj tražilici rangiraju prema svojstvenim vektorima čvorova. Za određivanje središnjih čvorova postoji nekoliko metoda, ovise o stupnjevima čvorova, bridovima koji ih povezuju i putovima koji idu kroz čvorove. U velikom broji slučajeva, poznavanje središnjih čvorova kompleksnih mreža, uvelike nam olakšava objašnjavanje raznih pojava i pojednostavnjuje ih.

7. Literatura

[1] Borgatti S.P., Everett M.G., A Graph-theoretic perspective on centrality, Social Networks, 28/ 4, Listopad 2006, str. 466-484; Science direct - *doi:10.1016/j.socnet.2005.11.005*

[2] Knjižica FER-a; ZPM; Mario-Osvin Pavčević: Uvod u teoriju grafova

[3] Kostanjčar Zvonko, Modeliranje dinamike kompleksnih sustava na mikro razini - primjer financijska tržišta, Fakultet elektrotehnike i računarstva

[4] Kvastek Ivan, Condor, Seminarski rad, Fakultet elektrotehnike i računarstva, 2008.

[5] Internet: Wikipedia – Centrality, *http://en.wikipedia.org/wiki/Centrality*

[6] Glavinić Đani: Osnovna svojstva kompleksnih mreža i njihova promjena, Seminarski rad, FER, 2007.

[7] Piškorec Matija, Modeli rasta kompleksnih mreža(seminar), 2007.

[8] Internet: *http://www.orgnet.com/sna.html*

[9] Internet: *http://www.faculty.ucr.edu/~hanneman/nettext/C10_Centrality.html*

[10] Internet: *http://geza.kzoo.edu/~csardi/module/html/centrality.html*

[11] Antulov-Fantulin Nino, Utjecaj zaraze na svojstva kompleksne mreže, Završni rad br.243, FER, 2008.

[12] Borgatti S.P., White D.G., Betweenness centrality measures for directed graphs, Social Networks, 16/ 4, Listopad 1994, str. 335-346; Science direct - *doi:10.1016/0378-8733(94)90015-9*

[13] Everett M., Borgatti S.P., Ego network betwenness, Social Networks, 27/1, Siječanj 2005, str. 31-38; Science direct - *doi:10.1016/j.socnet.2004.11.007*

[14] Perra N., Fortunato S., Spectral centrality measures in complex networks, 23.09.2008., *arXiv:0805.3322v2*

[15] Newman M.E.J., The structure and function of complex networks, *arXiv:cond-mat/0303516*

[16] Internet: *http://www.google.com/intl/hr/corporate/tech.html*

8. Sažetak

Kompleksne mreže su grafovi (mreže) sa topološkim značajkama koje se ne mogu naći u jednostavnim, slučajnim grafovima, kao što su komplicirane distribucije stupnjeva čvorova, hijerarhijska struktura grafa, koja je ujedno i struktura zajednice, visok koeficijent grupiranja čvorova i/ili podudaranje čvorova. Središnji se čvorovi određuju raznim metodama. Najvažnije su stupanj čvora, blizina čvorova, međupoloženost čvorova i svojstveni vektor čvora. Kada saznamo koji su središnji čvorovi, možemo si puno jednostavnije objasniti razne efekte te mreže. Sve te metode imaju matematičku podlogu, zasnovanu na teoriji grafova. Nadam se da je ovaj rad objasnio metode i poneke primjene tih metoda na kompleksne mreže i na pojave oko nas.