Mjere centralnosti u kompleksnim mrežama

Marinović, Marko

Sveučilište J.J. Strossmayera u Osijeku
Odjel za matematiku
Sveučilišni nastavnički studij matematike i informatike

Marko Marinović

Mjere centralnosti u kompleksnim mrežama

Diplomski rad

Osijek, 2018.

Sveučilište J.J. Strossmayera u Osijeku
Odjel za matematiku
Sveučilišni nastavnički studij matematike i informatike

Marko Marinović

Mjere centralnosti u kompleksnim mrežama

Diplomski rad

Mentor: doc. dr. sc. Snježana Majstorović

Osijek, 2018.

Sadržaj

Uvod

Modeliranje društvenih, bioloških i informacijsko-tehnoloških sustava pomoću kompleksnih mreža pokazalo se kao uspješan pristup za razumijevanje njihovih funkcija. Među različitim aspektima mreža koje su do sada istraživane, pitanje centralnosti i s njime povezanog problema identificiranja centralnih elemenata u mreži ostalo je ključno od trenutka svojeg prvog uvođenja. Ideja o centralnosti je u početku predložena u kontekstu društvenih sustava u kojima se pretpostavlja odnos između položaja pojedinca u mreži i njegovog utjecaja i moći unutar sudjelovanja u nekoj grupi. Od tada su tijekom godina uvedene različite mjere centralnosti za rangiranje vrhova unutar mreže prema njihovoj strukturnoj važnosti. Problem centralnosti se najčešće primjenjuje u društvenim i biološkim sustavima.

Kompleksna mreža je mreža ili graf s netrivijalnim strukturnim svojstvima kao što su npr. visoki koeficijent grupiranja, relativno mali dijametar (proporcionalan je s $\log n$, gdje je n broj vrhova u mreži), pojavljivanje *hubova* (vrhova s visokim stupnjem), hijerarhijska struktura i pojavljivanje zajednica, tj. istaknutih grupa vrhova među kojima postoji veliki broj bridova. Teorija kompleksnih mreža se intenzivno počela razvijati devedesetih godina prošlog stoljeća kada se pokazalo da različite mreže, koje opisuju realne sustave (npr. električna mreža, internet, mreža neurona jednostavnih organizama), imaju neka zajednička svojstva. D. Watts i S. Strogatz 1998. godine pokazali su da takve mreže imaju visok koeficijent grupiranja i male udaljenosti između vrhova i time definirali model mreže malog svijeta. A. Barabási i R. Albert su 1999. godine objavili rad u kojem opisuju preferencijalno povezivanje kao pravilo rasta mreža, a takav model nazivaju model mreža bez skale.

Nakon tih otkrića principi analize strukture mreža se na globalnoj i lokalnoj razini počinju primjenjivati u različitim znanstvenim područjima. Može se reći da na globalnoj razini različite mreže iz realnog svijeta (npr. internet, društvene mreže, mreže prometnica, terorističke mreže, mreže širenja zaraza) imaju neka univerzalna strukturna svojstva. Uobičajeno je da su to modeli mreže malog svijeta i/ili modeli mreže bez skale. Međutim, gledano na lokalnoj razini, strukture različitih mreža s istim globalnim svojstvima se značajno razlikuju.

Primjer mreže malog svijeta (eng. *small world effect*) je pokus Stanleya Milgrama [12], socijalnog psihologa iz 1960-tih, kojim je ustanovio da su sve osobe na svijetu povezane u prosjeku preko 6 drugih ljudi. Ukoliko udaljenost između dva proizvoljno odabrana vrha mreže raste proporcionalno logaritamskoj vrijednosti broja vrhova te mreže, onda govorimo o mreži malog svijeta. Mreže bez skale (eng. *power law degree distribution*) imaju polinomnu distribuciju stupnjeva, najčešće Pareto distribuciju. Takve su mreže na primjer World Wide Web, mreže proteinskih interakcija u metabolizmu itd. Ove mreže sadrže *hubove*, vrhove koji imaju stupanj puno veći od prosječnog stupnja ostalih vrhova u mreži. Još jedan primjer kompleksne mreže je mreža preferencijalne privrženosti (eng. *preferential attachment*) autora Udnyja Yule [15] iz 1925. godine, koju koristi za objašnjenje distribucije moći i broja vrsta po rodu cvjetnih biljaka. Postupak se ponekad naziva *"procesom jula"* u njegovu čast. Yule je pokazao da proces dovodi do raspodjele snage, ali pojedinosti njegovog dokaza su, prema današnjim standardima, proturječne i teške, budući da moderni alati stohastičke teorije procesa još nisu postojali i on bio je prisiljen koristiti tzv. "labave metode" dokazivanja.

U ovom radu bavit ćemo se analizom kompleksnih mreža na globalnoj razini, i to na način da upoznamo metode koje pomažu u identifikaciji najvažnijeg ili najutjecajnijeg vrha u mreži. Najutjecajniji vrh u mreži ima svojstvo da se njegovom eliminacijom poremeti, odnosno promijeni struktura cijele mreže.

Prije nego što se upoznamo s tim metodama − mjerama centralnosti − u prvom poglavlju ćemo se prisjetiti najvažnijih pojmova iz teorije grafova koji su nam potrebni za lakše čitanje i shvaćanje ovoga rada. U drugom poglavlju iznosimo mjere centralnosti koje identificiraju najvažnije vrhove i prikazujemo njihova svojstva, prednosti i nedostatke. Posljednje, treće poglavlje sadrži usporedbu stupnja vrha kao mjere centralnosti zajedno sa centralnostima koje su "dalje" od najbližih susjeda.

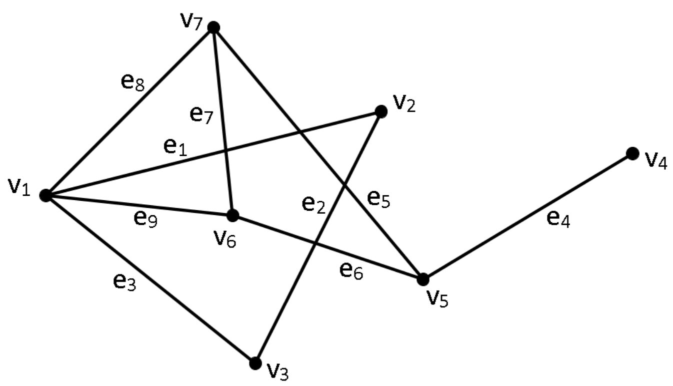
1 Osnovni pojmovi i tvrdnje iz teorije grafova

Kako tema ovoga diplomskoga rada dolazi iz područja teorije grafova, u ovom poglavlju ćemo navesti i objasniti pojmove iz teorije grafova koji su neophodni za razumijevanje rada.

1.1 Neusmjereni grafovi

Definicija 1. ***Graf*** *(mreža) G je uređena trojka $G = (V(G), E(G), \psi_G)$ koja se sastoji od nepraznog skupa $V(G)$, čiji su elementi* ***vrhovi*** *od G, skupa $E(G)$ disjunktnog s $V(G)$, čiji su elementi* ***bridovi*** *od G i* ***funkcije incidencije*** *ψ_G koja svakom bridu od G pridružuje neuređeni par (ne nužno različitih) vrhova od G.*

Primjer 1. *Odredimo skup vrhova $V(G)$, skup bridova $E(G)$ i funkciju incidencije ψ_G grafa (mreže) na slici 1.*



Slika 1: Primjer grafa (mreže).

Skup vrhova: $V(G) = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$
Skup bridova: $E(G) = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9\}$
Funkcija incidencije ψ_G: $\psi_G(e_1) = v_1v_2$, $\psi_G(e_2) = v_2v_3$, $\psi_G(e_3) = v_1v_3$, $\psi_G(e_4) = v_4v_5$, $\psi_G(e_5) = v_5v_7$, $\psi_G(e_6) = v_5v_6$, $\psi_G(e_7) = v_6v_7$, $\psi_G(e_8) = v_1v_7$, $\psi_G(e_9) = v_1v_6$.

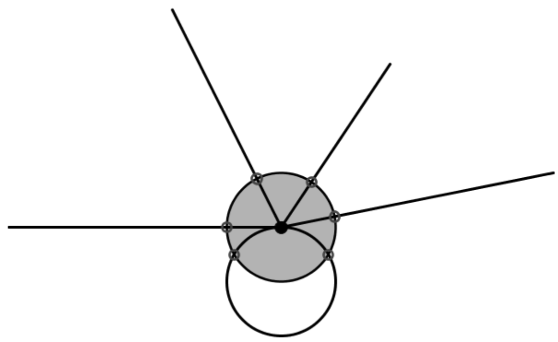
Ukoliko funkcija incidencije ψ_G nije injekcija, tj. ako su dva različita vrha spojena s više bridova, onda kažemo da promatrani graf sadrži *višestruke bridove* i zovemo ga *multigraf*. Graf može sadržavati brid koji spaja neki vrh sa samim sobom te se takav brid naziva *petlja*. Stoga postoji još jedna vrsta grafova, zovemo ju *pseudografovi*, a to su (multi)grafovi koji sadržavaju (višestruke) petlje. Skup vrhova i skup bridova općenito ne moraju imati konačan broj elemenata, a u slučaju kada imaju, promatrani graf se naziva *konačan graf*.

Radi jednostavnosti, pretpostavit ćemo da imamo konačan graf (mrežu) koji nema višestrukih bridova ni petlji.

Za vrhove u i v kažemo da su *incidentni* s bridom e ako je vrh u jedan kraj brida e, a vrh v drugi kraj brida e. Također kažemo da su proizvoljna dva vrha *susjedna* ako su oba incidentna s nekim bridom.

Sljedeći pojam koji se prirodno nameće u teoriji grafova jest stupanj vrha u grafu. Intuitivno, stupanj vrha je broj sjecišta male kružnice oko promatranog vrha s bridovima koji izlaze iz toga vrha (vidi sliku 2).

Definicija 2. ***Stupanj*** *ili* ***valencija vrha*** *v grafa G je broj $d_G(v)$ bridova grafa G koji su incidentni s vrhom v.*



Slika 2: Vrh stupnja - intuitivna ideja.

Važno je napomenuti da, ukoliko imamo petlju u grafu, moramo ju računati kao dva brida. Ukoliko u grafu imamo vrh čiji je stupanj jednak $d_G(v) = 0$, tada za taj vrh v kažemo da je *izolirani vrh* grafa G. Ovdje ćemo navesti i dva vrlo važna svojstva stupnja vrha: vezu stupnja vrha i broja bridova u grafu te poznatu lemu o rukovanju.

Propozicija 1. *Za svaki graf G vrijedi*

$$\sum_{v \in V(G)} d(v) = 2 \cdot |E(G)|.$$

Dokaz. Znamo da se stupanj svakog vrha u proizvoljnom grafu definira kao broj bridova incidentnih tome vrhu. Međutim, kako svaki brid ima dva kraja, slijedi da zbroj stupnjeva u promatranom grafu mora biti jednak dvostrukom broju bridova toga grafa. □

Korolar 1 (Lema o rukovanju)**.** *U svakom grafu G je broj vrhova neparnog stupnja paran broj.*

Dokaz. Znamo da propozicija 1 vrijedi za svaki graf G. Kako bismo dokazali ovaj korolar, napravimo particiju skupa vrhova $V(G)$ grafa G u dva skupa. Neka je V_1 skup svih vrhova koji imaju neparan stupanj, te neka je V_2 skup svih vrhova koji imaju paran stupanj. Tada prema propoziciji 1 mora vrijediti

$$\sum_{v \in V_1} d(v) + \sum_{v \in V_2} d(v) = 2 \cdot |E(G)|.$$

Kako su $2 \cdot |E(G)|$ i drugi pribrojnik prethodnog zbroja uvijek parni brojevi, slijedi da i prvi pribrojnik, $\sum\limits_{v \in V_1} d(v)$, mora biti paran broj.
Skup V_1 možemo zapisati kao $V_1 = \{u_1, u_2, \ldots, u_{|V_1|}\}$ gdje je $d(u_i) = 2k_i + 1$ za neki $k_i \in \mathbb{N}_0$, $\forall i = 1, 2, \ldots, |V_1|$. Tada je

$$\underbrace{\sum_{i=1}^{|V_1|} d(u_i)}_{paran\ broj} = \sum_{i=1}^{|V_1|} (2k_i + 1) = \underbrace{2 \cdot \sum_{i=1}^{|V_1|} k_i}_{paran\ broj} + |V_1|,$$

iz čega slijedi da je i $|V_1|$ paran broj, čime je dokaz završen. □

Definicija 3. *Neka je zadan graf G sa skupom vrhova $V = \{v_1, v_2, \ldots, v_n\}$. Niz $(d(v_1), d(v_2), \ldots, d(v_n))$ zove se* ***niz stupnjeva*** *od G.*

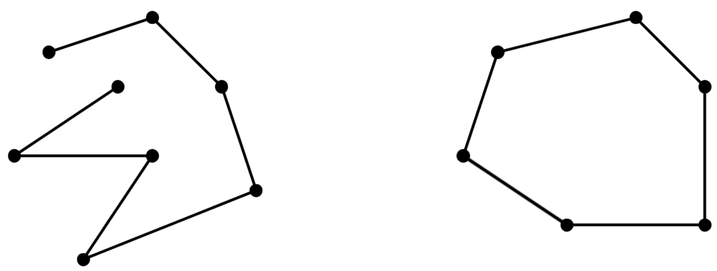
Niz stupnjeva nekog grafa je jednostavno odrediti, no obratna situacija je nešto složenija.

Definicija 4. *Niz brojeva $d = (d_1, d_2, \ldots, d_n) \in \mathbb{N}_0^n$ je* ***grafički*** *ako postoji jednostavan graf čiji je niz stupnjeva d.*

Navedimo neke specijalne vrste grafova.

- **Put** s n vrhova P_n je jednostavan graf definiran sa skupom vrhova $V(P_n) = \{v_1, v_2, \ldots, v_n\}$ i skupom bridova $E(P_n) = \{v_i v_{i+1} : i \in \{1, \ldots, n-1\}\}$.

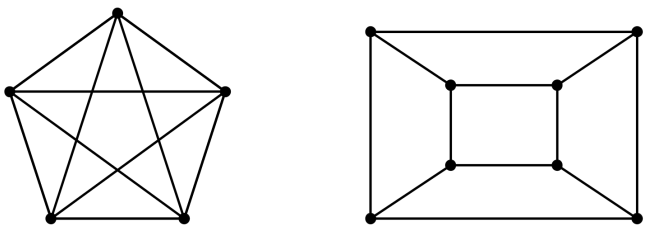
- **Ciklus** s n vrhova C_n je jednostavan graf definiran skupom vrhova $V(C_n) = \{v_1, v_2, \ldots, v_n\}$ i skupom bridova $E(C_n) = \{v_1v_2, v_2v_3, \ldots, v_{n-1}v_n, v_nv_1\}$.
- **Potpun graf** s n vrhova K_n je jednostavan graf u kojemu je svaki par vrhova spojen bridom.
- **Regularan graf** je jednostavan graf koji je r-regularan za neki r (graf je r-regularan ako mu je svaki vrh stupnja r).
- **Bipartitan graf** je jednostavan graf čiji se skup vrhova može particionirati u dva međusobno disjunktna skupa X i Y tako da svaki brid ima jedan kraj u X, a drugi u Y.
- **Potpun bipartitan graf** je jednostavan bipartitan graf s biparticijom (X, Y) u kojemu je svaki vrh iz X spojen sa svakim vrhom iz Y; označava se s $K_{m,n}$ gdje je $m = |X|$ i $n = |Y|$.
- **Zvijezda** S_n je potpun bipartitan graf $K_{1,n-1}$ (vidi sliku 27).



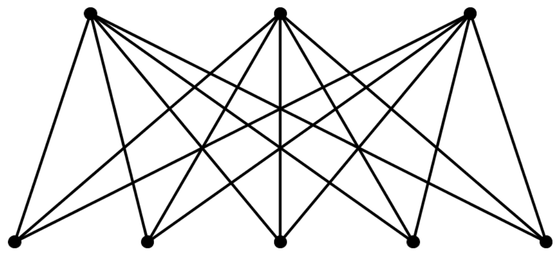
Slika 3: Put P_8 (lijevo) i ciklus C_6 (desno).

Definicija 5. *Graf H je* ***podgraf*** *grafa G, u oznaci $H \subseteq G$, ako je $V(H) \subseteq V(G)$ i $E(H) \subseteq E(G)$, a za funkcije incidencije vrijedi $\psi_H = \psi_{G_{|E(H)}}$, odnosno ψ_H je restrikcija od ψ_G na $E(H)$.*

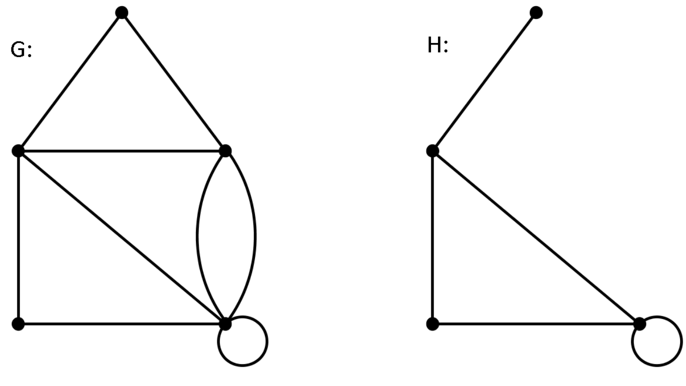
Ukoliko imamo da je $H \subseteq G$ te da je $H \neq G$, tada pišemo $H \subset G$ i graf H nazivamo *pravi podgraf* grafa G. Na slici 6 možemo vidjeti primjer grafa i jednog njegovog podgrafa.



Slika 4: Potpun graf K_5 (lijevo) i 3-regularan graf (desno).



Slika 5: Potpun bipartitan graf $K_{3,5}$.



Slika 6: Graf H je podgraf grafa G.

Definicija 6. *Neka je $\emptyset \neq V' \subseteq V(G)$. Podgraf od G čiji je skup vrhova V', a skup bridova je podskup skupa svih bridova od G čija su oba kraja u V' zove se* ***podgraf***

inducirani s V *i označava s* $G[V_1]$. *Kažemo još da je* $G[V_1]$ *inducirani podgraf od* G.

1.2 Povezanost grafa

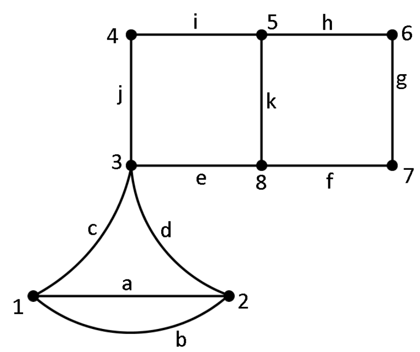
Česti problemi u teoriji grafova vezani su za određivanje najkraćeg puta između neka dva ili svih vrhova, pri čemu se može dogoditi da takav put ne postoji. U ovome odjeljku definirat ćemo pojmove koji su vrlo važni za analizu promatranog grafa, a to su šetnja, put, ciklus, povezan graf, udaljenost dvaju vrhova, komponente povezanosti promatranog grafa, te rezni vrh i most.

Definicija 7. ***Šetnja*** *u grafu* G *je netrivijalan konačan niz* $W = v_0e_1v_1e_2v_2 \dots e_kv_k$ *čiji su članovi naizmjenično vrhovi* v_i *i bridovi* e_i *tako da su krajevi od* e_i *vrhovi* v_{i-1} *i* v_i, $\forall i = 1, 2, \dots, k$.

Kažemo da je W šetnja od v_0 do v_k ili (v_0, v_k)-šetnja. Vrhove v_0 i v_k redom zovemo *početak* i *kraj* šetnje, a sve ostale vrhove *unutarnji vrhovi* šetnje. Broj $k \in \mathbb{N}$ nazivamo *duljina* šetnje W. Ukoliko imamo da je $v_0 = v_k$, tada kažemo da je šetnja W *zatvorena*.

Ako su svi bridovi u šetnji W međusobno različiti, onda se W naziva ***staza*** duljine k. Ako se šetnja W sastoji od međusobno različitih vrhova (a onda i bridova), tada se W zove ***put*** duljine k. Specijalno se može dogoditi da put ima duljinu 0, a onda za takav put kažemo da je *trivijalan*. Zatvorena šetnja u kojoj su svi vrhovi (osim početnog i krajnjeg) međusobno različiti naziva se ***ciklus***.

Primjer 2. *Odredimo jednu šetnju, stazu, put i ciklus grafa (mreže) sa slike 7.*



Slika 7

Šetnja: $3e8k5i4j3e8k5i4j3$ *; Staza:* $1a2b1c3e8k5i4j3$
Put: $1c3j4i5k8f7$ *; Ciklus:* $6h5i4j3e8f7g6$

Definicija 8. ***Udaljenost*** $d_G(u,v)$ *dvaju vrhova* $u, v \in V(G)$ *je duljina najkraćeg* (u,v)*-puta u grafu* G*. Ako ne postoji takav put u* G*, onda stavljamo* $d_G(u,v) = \infty$*.*

Za dva vrha u grafu G ćemo reći da su *povezana* ako postoji barem jedan (u,v)-put u grafu G. Iz ovoga kratkog razmatranja direktno slijedi definicija povezanosti grafa.

Definicija 9. *Graf* G *je* **povezan** *ako je* $d_G(u,v) < \infty \ \forall u, v \in V(G)$*. U suprotnom kažemo da je graf* G *nepovezan.*

Uočimo da svakom grafu G možemo particionirati skup vrhova V na neprazne podskupove $V_1, V_2, \ldots, V_k$ tako da su dva vrha $u, v \in V(G)$ povezana u G ako i samo ako postoji $i \in \{1, 2, \ldots, k\}$ takav da je $u, v \in V_i$. Tada grafove $G[V_1], G[V_2], \ldots, G[V_k]$ nazivamo ***komponente povezanosti*** grafa G. Broj komponenti povezanosti grafa G označavat ćemo s $c(G)$.

Vrh v grafa G za kojeg vrijedi $c(G - v) > c(G)$ nazivamo **rezni vrh**. Brid e grafa G za kojeg vrijedi $c(G - e) > c(G)$ nazivamo **most**.

1.3 Specijalne matrice grafa G

Definirat ćemo matricu susjedstva i matricu udaljenosti zadanog grafa i navesti njihova osnovna svojstva.

Definicija 10. *Neka je* G *proizvoljan graf sa skupom vrhova* $V(G) = \{v_1, v_2, \ldots, v_n\}$*.* ***Matrica susjedstva*** $A(G)$ *grafa* G *je kvadratna* $n \times n$ *matrica s elementima* $a_{ij},\ i, j \in \{1, 2, \ldots, n\}$*, pri čemu je* a_{ij} *broj bridova koji spajaju vrhove* v_i *i* v_j*.*

Iz definicije 10 vidimo da je matrica susjedstva $A(G)$ neusmjerenog grafa G simetrična matrica, tj. $a_{ij} = a_{ji}$. Također uočimo da su svi elementi matrice susjedstva nenegativni cijeli brojevi. Zbog svojstva simetričnosti, matrica susjedstva se može dijagonalizirati, a njezini pripadni svojstveni vektori, dobiveni uz pomoć odgovarajuće karakteristične jednadžbe, čine bazu vektorskog prostora $\mathbb{R}^n$ (dokaz ove tvrdnje se može vidjeti u [4]). Ako vrhovi v_i i v_j nisu spojeni bridom, onda je $a_{ij} = a_{ji} = 0$. Za jednostavan graf G matrica susjedstva je $0,1$-matrica takva da su svi elementi na glavnoj dijagonali jednaki 0, a svaki od preostalih elemenata je ili

1 ili 0. Za multigrafove također vrijedi $a_{ii} = 0$, $\forall i = 1, 2, \ldots, n$, jer ne postoji brid koji spaja neki vrh sa samim sobom, odnosno nemamo petlje u grafu.

Za multigraf G suma elemenata u i-tom retku (ili i-tom stupcu) jednaka je stupnju vrha $v_i \in V(G)$, a suma svih elemenata u matrici susjedstva $A(G)$ jednaka je dvostrukom broju bridova u grafu G (vidi propoziciju 1). Ako je i-ti redak od $A(G)$ nul-redak, onda je pripadni vrh v_i *izolirani vrh* u grafu G.

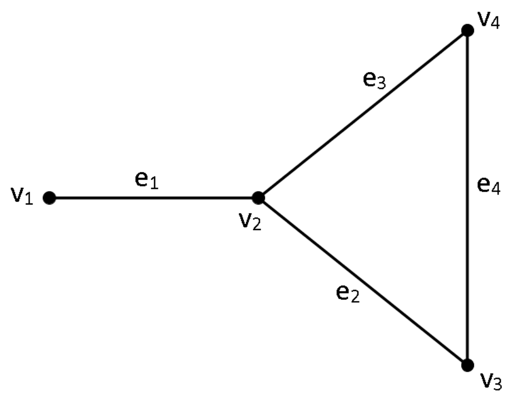
Sljedeći teorem smatra se jednim od važnijih teorema teorije grafova jer uspostavlja vezu između broja šetnji duljine k i potencije matrice susjedstva. Ovaj teorem je poseban i po tome što vrijedi i za usmjerene grafove. Dokaz se može pronaći u [10].

Teorem 1. *Neka je $A(G) = [a_{ij}]$ matrica susjedstva grafa G i $V(G) = \{v_1, v_2, \ldots, v_n\}$. Tada je (i, j)-ti član k-te potencije A^k jednak broju (v_i, v_j)-šetnji duljine k. Broj svih šetnji na grafu G duljine k jednak je sumi svih članova od A^k.*

Definicija 11. *Neka je G povezan graf sa skupom vrhova $V(G) = \{v_1, v_2, \ldots, v_n\}$. **Matrica udaljenosti** $D(G)$ grafa G je kvadratna $n \times n$ matrica s elementima $d(v_i, v_j)$ za $i, j = 1, 2, \ldots, n$.*

Uočimo da je matrica udaljenosti nenegativna simetrična matrica kojoj su svi elementi na glavnoj dijagonali jednaki 0.
Sada, kada smo definirali neke specifične matrice zadanog grafa, navedimo i primjer.

Primjer 3. *Za dani graf G na slici 8 odredimo odgovarajuću matricu susjedstva $A(G)$ i matricu udaljenosti $D(G)$.*



Slika 8

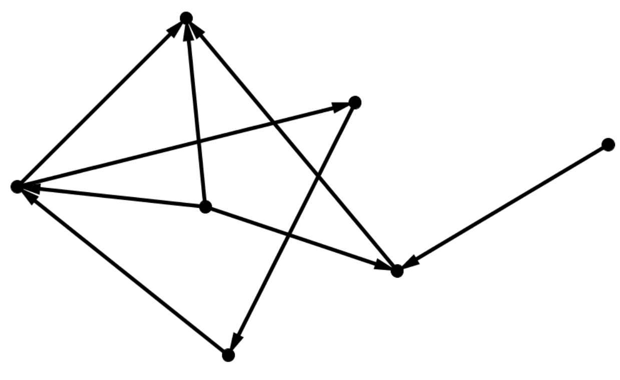
$$A(G) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}, \qquad D(G) = \begin{bmatrix} 0 & 1 & 2 & 2 \\ 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{bmatrix}.$$

1.4 Usmjereni grafovi

Preostalo nam je još definirati usmjerene grafove (mreže). Usmjereni grafovi se razlikuju od neusmjerenih u tome što svaki brid usmjerenog grafa ima specificiran početak i kraj.

Definicija 12. ***Usmjereni graf*** *ili* ***digraf*** *D je uređena trojka $(V(D), A(D), \psi_D)$ koja se sastoji od nepraznog skupa vrhova $V(D)$, skupa lukova ili usmjerenih bridova $A(D)$ i funkcije incidencije ψ_D koja svakom luku a pridružuje uređeni par (ne nužno različitih) vrhova u i v. Kažemo još da je vrh u početni, a vrh v krajnji vrh od a.*

Svakom usmjerenom grafu D možemo pridružiti pripadni neusmjereni graf koji će imati isti skup vrhova kao i D, a svaki luk u D ćemo zamijeniti sa neusmjerenim bridom. Obratno, svakom neusmjerenom grafu G možemo pridružiti odgovarajući usmjereni graf D tako da specificiramo početak i kraj svakog brida. Još se kaže da je takav usmjereni graf *orijentacija* na G. Na slici 9 možemo vidjeti primjer usmjerenog grafa čiji je pripadni graf prikazan na slici 1.



Slika 9: Usmjereni graf D.

Većina pojmova koji vrijede za neusmjerene grafove mogu se odnositi i na usmjerene grafove, ali pri tome moramo paziti na pojmove vezane za orijentaciju bridova. Usmjereni podgraf usmjerenog grafa definira se slično kao i podgraf grafa. Također, usmjerena šetnja, staza, put i ciklus definiraju se slično kao i za grafove te ih, redom, zovemo *dišetnja*, *distaza*, *diput* i *diciklus*. No, treba paziti na to da se kod tih pojmova šećemo usmjerenim grafom kao što bi vozili jednosmjernim ulicama.

Definicija 13. ***Acikličkim usmjerenim grafom*** *(mrežom) nazivamo onaj graf koji u sebi ne sadrži niti jedan diciklus.*

Kao što je ranije rečeno, usmjereni grafovi sadrže usmjerene bridove. Ti bridovi "izlaze" iz jednog vrha i "ulaze" u drugi vrh. Stoga se prirodno nameće zaključak da će definicija stupnja vrha u usmjerenim grafovima (mrežama) biti nešto drugačija od one definicije za stupanj vrha neusmjerenog grafa (vidi definiciju 2).

Definicija 14. ***Ulazni stupanj*** *vrha v usmjerenog grafa D je broj $d_D^-(v)$ lukova u D kojima je kraj vrh v.* ***Izlazni stupanj*** *vrha v usmjerenog grafa D je broj $d_D^+(v)$ lukova u D kojima je početak vrh v.*

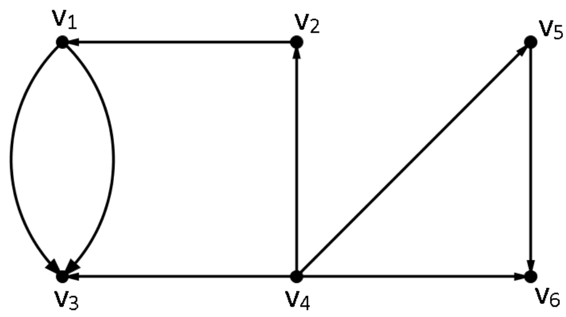
Sada ćemo definirati čvrsto povezane (ili dipovezane) komponente usmjerenog grafa. Uočimo da se ovaj pojam ne može usporediti s komponentama povezanosti neusmjerenog grafa.

Definicija 15. *Usmjereni graf je* ***čvrsto povezan*** *ukoliko za njega vrijedi da se iz svakog vrha toga grafa može doći u svaki drugi vrh istog grafa.*
Čvrsto *(jako)* ***povezana komponenta*** *usmjerenog grafa (eng.* strongly connected component*) je najveći čvrsto povezan podgraf zadanog grafa.*

Budući da je sama definicija stupnja vrha u usmjerenim grafovima različita od one u slučaju neusmjerenih grafova, za očekivati je da će se međusobno razlikovati i definicija matrice susjedstva za usmjerene grafove od one definicije koju smo naveli za neusmjereni slučaj.

Definicija 16. *Neka je D usmjereni graf sa skupom vrhova $V(D) = \{v_1, v_2, \ldots, v_n\}$.* ***Matrica susjedstva*** *$A(D)$ usmjerenog grafa D je kvadratna $n \times n$ matrica s elementima a_{ij}, $i, j \in \{1, 2, \ldots, n\}$, pri čemu je a_{ij} broj lukova u D kojima je početak vrh v_i, a kraj vrh v_j.*

Primjer 4. *Neka je zadan usmjereni graf D kao na slici 10. Odredimo mu matricu susjedstva.*



Slika 10

$$A(D) = \begin{bmatrix} 0 & 0 & 2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Uočimo da je matrica susjedstva usmjerenog grafa asimetrična, a kako imamo dvije vrste stupnjeva kod usmjerenih grafova, onda imamo i dvije vrste svojstvenih vektora koje možemo pridružiti svojstvenoj vrijednosti matrice $A(D)$.

Definicija 17. *Vektor $x \in \mathbb{R}^n$, $x \neq 0$, koji zadovoljava jednakost $Ax = \lambda x$, pri čemu je A kvadratna matrica reda n, zovemo* ***desni svojstveni vektor*** *matrice A za svojstvenu vrijednost λ, a vektor $y \in \mathbb{R}^n$, $y \neq 0$, shvaćen kao vektor-redak, koji zadovoljava $A^\tau y^\tau = \lambda y^\tau$ zovemo* ***lijevi svojstveni vektor*** *matrice A za svojstvenu vrijednost λ.*

Usmjereni težinski grafovi se uglavnom primjenjuju u modeliranju analize protoka robe od proizvođača (izvora) do nekih unaprijed određenih potrošača, odnosno odredišta (ponora). Tako npr. možemo imati protok nafte u naftovodu, transport proizvedene robe do trgovačkih centara, prijenos poruka ili informacija u nekoj društvenoj mreži, itd. Težina luka takvog usmjerenog grafa je količina nafte koja može proteći nekim dijelom naftovoda, maksimalna brzina prijevoza artikla itd.

Definicija 18. ***Transportna mreža*** *N je povezan usmjereni graf bez petlji za koji vrijedi:*

(i) postoji jedinstven vrh i sa ulaznim stupnjem $d^-(i) = 0$ koji se zove ***izvor****;*

(ii) postoji jedinstven vrh p sa izlaznim stupnjem $d^+(p) = 0$ koji se zove ***ponor****;*

(iii) svakom luku $a = (u, v) \in A(N)$ pridružen je realan broj $c(a) = c(u, v) \geq 0$ koji se zove ***kapacitet*** *od a (ako takav luk (u, v) ne postoji, stavljamo $c(u, v) = 0$).*

Definicija 19. ***Protok*** *mreže N je funkcija $f : A(N) \to \mathbb{R}^+$ koja svakom luku $a = (u, v)$ pridružuje broj $f(a) = f(u, v)$ tako da vrijedi*

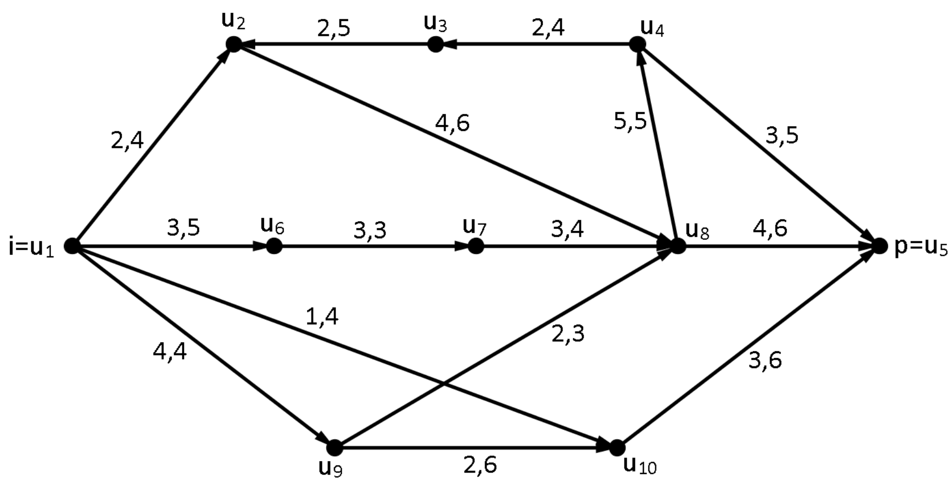
$$f(u, v) \leq c(u, v),$$

*(**ograničenje kapacitetom**) te*

$$\sum_{v \in V(N)} f(u, v) = \sum_{v \in V(N)} f(v, u) \qquad \forall u \neq i, p$$

*(**uvjet očuvanja** – količina robe pristigle u vrh u jednaka je količini robe koja izlazi iz vrha u, osim ako se radi o izvoru i ponoru).*

Na slici 11 možemo vidjeti primjer transportne mreže N. Ako ne postoji luk od



Slika 11: Transportna mreža N.

u do v, onda stavljamo da je $f(u, v) = 0$. Uočimo da na transportnoj mreži uvijek imamo protok jer uvijek možemo staviti $f(a) = 0 \ \forall a \in A(N)$.

Definicija 20. ***Vrijednost protoka*** *f na mreži N je*

$$val(f) = \sum_{v \in V(N)} f(i, v).$$

Kažemo da je protok f^* ***maksimalan protok*** *na mreži N ako ne postoji protok f u N za koji je $val(f) > val(f^*)$.*

Razmatranje o usmjerenim grafovima završit ćemo važnim teoremom vezanim za protok u transportnoj mreži. Dokaz teorema nećemo provoditi, nego zainteresiranog čitatelja upućujemo na [3]. Uvest ćemo još nekoliko pojmova koji nam trebaju za iskaz tog teorema. Za $S, T \subseteq V(N)$, sa (S, T) označimo skup svih lukova s početnim vrhom u skupu S, a krajnjim u skupu T. *Vršni rez* u N je skup oblika $R = (S, S^C)$. Kažemo da rez R *separira* izvor i od ponora p ako je $i \in S,\ p \in S^C$, odnosno kažemo da je takav rez $i - p$ rez. Kapacitet reza R, $c(R) = c(S, S^C)$, definira se kao

$$\sum_{(u,v) \in R} c(u, v).$$

Kažemo da je $i - p$ rez R mreže N *minimalan* ako ne postoji $i - p$ rez R' u N za koji je $c(R') < c(R)$.

Korolar 2. *Vrijednost protoka koji izlazi iz izvora mreže jednak je vrijednosti protoka koji ulazi u ponor mreže.*

Korolar 3. *Neka je f protok na N, a R $i - p$ rez za koji je $val(f) = c(R)$. Tada je f maksimalan protok, a R minimalan rez.*

Teorem 2 (Max flow − min cut)**.** *U transportnoj mreži je vrijednost maksimalnog protoka jednaka kapacitetu minimalnog reza.*

2 Mjere centralnosti

2.1 Stupanj vrha

Stupanjska centralnost, centralnost stupnja vrha ili samo stupanj vrha je povijesno prva i konceptualno najjednostavnija mjera centralnosti koja se koristi za pronalaženje najvažnijeg vrha u kompleksnoj mreži. Prema ovoj mjeri, važnost nekog vrha u mreži ovisi isključivo o broju bridova s kojima je taj vrh incidentan.

Koristeći matricu susjedstva A promatrane neusmjerene mreže s n vrhova, stupanj d_i pojedinog vrha v_i u mreži možemo izraziti kao:

$$d_i = (A \cdot \mathbf{1})_i, \quad i = 1, \ldots, n,$$

gdje je $\mathbf{1}$ vektor-stupac jedinica, odnosno $\mathbf{1} = (1, 1, 1, \ldots)^{\tau}$. Stupanj vrha kao mjeru centralnosti u mrežama prvi je koristio Freeman [8] koji ga je u modelni oblik stavio 1979. godine. Zbog jednostavnosti i intuitivne prirode, ovu je mjeru centralnosti predlagalo nekoliko autora od 1950-ih do 1970-ih, dok ju Freeman nije usavršio i počeo koristiti. Ideja za korištenje stupnja vrha kao mjere centralnosti polazi od intuicije da je jedan vrh u mreži centralniji ili utjecajniji od drugog vrha te mreže ako je stupanj prvoga vrha veći od stupnja drugoga vrha. Također znamo da stupanj vrha v odgovara broju šetnji duljine 1 koje počinju u v ili broju zatvorenih šetnji duljine 2 koje počinju i završavaju u v. Zato stupanj vrha možemo zapisati i ovako:

$$d_i = (A^2)_i, \quad i = 1, \ldots, n.$$

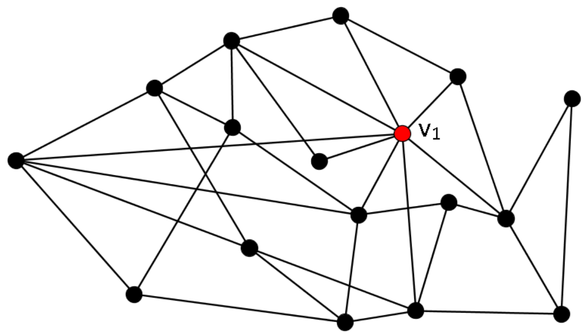
Mjera stupanjske centralnosti naročito je popularna u analizi društvenih mreža kada je cilj pronaći najpopularnije, odnosno najutjecajnije ljude, tj. ljude koji imaju najviše izravnih kontakata sa ostalim ljudima u mreži. Nadalje, ako promatramo širenje neke informacije ili virusa u mreži, onda je vjerojatnost da neki vrh primi tu informaciju ili da se zarazi virusom proporcionalna njegovom stupnju.

Na slici 12 prikazana je mreža N sa 17 vrhova i s nizom stupnjeva

$$(8, 5, 5, 5, 5, 5, 4, 4, 4, 4, 3, 3, 3, 3, 3, 2, 2).$$

Prema stupanjskoj centralnosti vrh v_1 je najutjecajniji vrh u mreži jer ima najveći mogući stupanj, odnosno $d_N(v_1) = 8$.

Stupanj vrha kao mjera centralnosti može se koristiti i u analizi usmjerenih mreža. S obzirom da u usmjerenim mrežama imamo usmjerene bridove, to ćemo



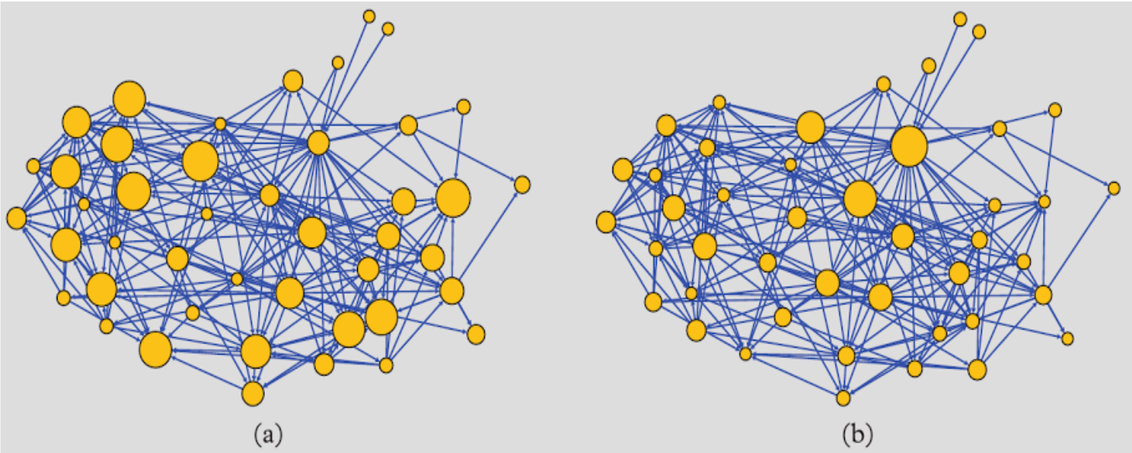
Slika 12: Mreža N s najutjecajnijim vrhom v_1.

razlikovati ulazni stupanj i izlazni stupanj, te tako definirati dvije mjere centralnosti promatranog vrha. Za $i = 1, \ldots, n$ s d_i^{in} ćemo označiti ulazni, a s d_i^{out} izlazni stupanj vrha v_i. Kao i kod neusmjerenih mreža, ove veličine možemo izraziti pomoću matrice susjedstva A usmjerene mreže. Imamo:

$$d_i^{in} = (\mathbf{1}^\tau \cdot A)_i,$$

$$d_i^{out} = (A \cdot \mathbf{1})_i.$$

Moguće je da obje mjere centralnosti, dakle i ulazni i izlazni stupanj, identificiraju jedan vrh kao najutjecajniji. No, to općenito ne mora vrijediti. Vrh koji je najutjecajniji prema ulaznom stupnju može biti potpuno nebitan vrh u centralnosti izlaznog stupnja, tj. može se dogoditi da promatrani vrh u usmjerenoj mreži ima vrlo visok ulazni stupanj, a istovremeno vrlo nizak izlazni stupanj, ili pak obrnuto. Jedan od najpoznatijih primjera usmjerenih mreža gdje učinkovitost mjera centralnosti ulaznog i izlaznog stupnja dolazi do velikog izražaja jesu mreže koje prikazuju odnos grabežljivca i plijena u nekom ekosustavu. Slika 13 prikazuje hranidbenu mrežu na otoku Sveti Martin na Karibima. Vrhovi mreže predstavljaju određenu vrstu, odnosno izvor hrane, a usmjereni bridovi opisuju što jede što u mreži. Na slici 13(a) vrhovi mreže su prikazani krugom čija je veličina proporcionalna veličini njihovog ulaznog stupnja, a na slici 13(b) veličina kruga - vrha mreže odgovara veličini njegovog izlaznog stupnja.



Slika 13: Hranidbena mreža na otoku Sveti Martin na Karibima. Slika je preuzeta iz [5].

Vrhovi sa velikim izlaznim stupnjem predstavljaju grabežljivce s velikim izborom plijena. Primjeri takvih grabežljivaca su gušteri *Anolis gingivinus* i *Anolis pogus*, te ptice *Margarops fuscatus* i *Setophaga petechia*. Vrhovi s velikim ulaznim stupnjem predstavljaju vrste i organske tvari kojih jede najviše grabežljivaca u ekosustavu. To je, primjerice lišće, zatim detritus i lisna uš. Općenito, grabežljivci koji su pri vrhu hranidbenog lanca nisu ugroženi od nekih drugih vrsta pa imaju visok izlazni stupanj i nizak ulazni stupanj. No, one vrste koje su plijen mnogim grabežljivcima općenito nemaju mnogo plijena te stoga imaju visok ulazni stupanj, a nizak izlazni stupanj. Vrhovi čiji je ulazni stupanj jednak nuli najčešće su ptice, dok su vrhovi čiji je izlazni stupanj jednak nuli neke vrste biljaka i detritus.

Iz svega navedenog možemo zaključiti da u hranidbenoj mreži mora postojati određena antikorelacija između ulaznih i izlaznih stupnjeva svih vrhova. Korelaciju između ulaznog i izlaznog stupnja možemo mjeriti Pearsonovim koeficijentom korelacije. Prisjetimo se definicije i značenja tog pojma.

Definicija 21. *Neka je zadan dvodimenzionalni slučajni vektor* (X, Y) *te neka su* $(x_1, y_1), \ldots, (x_n, y_n)$ *parovi nezavisnih realizacija tog vektora. Tada se* ***Pearsonov koeficijent korelacije*** *računa prema formuli:*

$$\Omega = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2}},$$

pri čemu je $\bar{x}_n$ prosječna vrijednost varijable X i $\bar{y}_n$ prosječna vrijednost varijable Y.

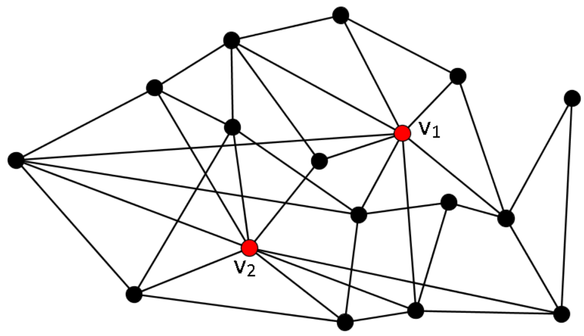
Pearsonov koeficijent korelacije koristi se u slučajevima kada između varijabli promatranog modela postoji linearna povezanost i neprekidna normalna distribucija. Vrijednost Pearsonovog koeficijenta korelacije kreće se od $+1$ (savršena pozitivna korelacija) do -1 (savršena negativna korelacija). Predznak koeficijenta upućuje na smjer korelacije – je li korelacija pozitivna ($\Omega > 0$), negativna ($\Omega < 0$) ili neutralna ($\Omega = 0$), ali nas ne upućuje na snagu korelacije. Ako je korelacija između dviju varijabli pozitivna, to znači da se povećanjem vrijednosti prve varijable istovremeno povećava vrijednost druge varijable i obrnuto. Ukoliko je korelacija između dviju varijabli negativna, to znači da se povećanjem vrijednosti prve varijable istovremeno smanjuju vrijednosti druge varijable i obrnuto. Pearsonov koeficijent korelacije bazira se na usporedbi stvarnog utjecaja promatranih varijabli jedne na drugu u odnosu na maksimalni mogući utjecaj dviju varijabli.

Primijenimo li Pearsonov koeficijent korelacije na ulazne i izlazne stupnjeve promatrane mreže, to će značiti da mreža može prikazivati pozitivnu ($\Omega > 0$), negativnu ($\Omega < 0$) ili neutralnu ($\Omega = 0$) korelaciju između ulaznih i izlaznih stupnjeva. Formula za Pearsonov koeficijent korelacije između ulaznog i izlaznog stupnja glasi:

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

gdje je sa $\overline{d^{in}}$ i $\overline{d^{out}}$ označen prosječan ulazni, odnosno prosječan izlazni stupanj vrha u mreži. Provedemo li izračun Pearsonovog koeficijenta korelacije Ω za hranidbenu mrežu na otoku Sveti Martin, dobivamo $\Omega = -0.308$. Ovaj rezultat pokazuje negativnu korelaciju između ulaznih i izlaznih stupnjeva te mreže. Takve mreže pripadaju skupini takozvanih antirecipročnih mreža, tj. onih koje sadrže vrhove s više ulaznih nego izlaznih bridova te vrhove koji imaju više izlaznih nego ulaznih bridova. Slika 13(a) ističe grupu vrhova koji imaju visoki ulazni stupanj i nizak izlazni stupanj, a slika 13(b) prikazuje grupu vrhova koja ima visok izlazni stupanj i nizak ulazni stupanj, što objašnjava antikorelaciju između ulaznog i izlaznog stupnja u ovoj mreži, tj. $\Omega < 0$.

Stupanjska centralnost, kao najjednostavnija mjera centralnosti u kompleksnim mrežama, ima bitnih nedostataka. Ako preuredimo mrežu N prikazanu na slici 12 tako da joj dodamo neke bridove, dobivamo mrežu N' prikazanu na slici 14. Mreža N' je i dalje jednostavna neusmjerena mreža kao što je i N, ali u njoj više nemamo jedinstven najutjecajniji vrh. Vrh v_2 ima isti stupanj kao i vrh v_1. Kao što smo ranije

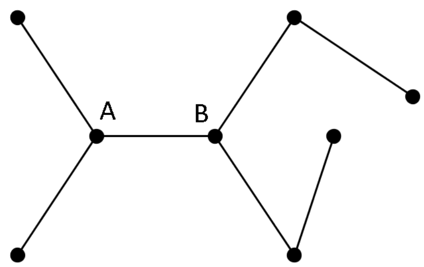


Slika 14: Mreža N' s dva najutjecajnija vrha.

spomenuli, stupanj vrha kao mjera centralnosti pronaći će vrh(ove) sa najvećim stupnjem u mreži, ali ako više vrhova imaju jednak stupanj, tada nastaje problem jer ne znamo koji od njih treba proglasiti najutjecajnijim vrhom u mreži. Ta odluka može biti od velike važnosti ukoliko želimo riješiti problem u nekoj terorističkoj mreži ili u mreži koja opisuje širenje epidemije. Ova mjera nam neće pomoći da ustanovimo koji vrh treba "eliminirati" kako bi uklonili moguću opasnost.

2.2 Centralnosti koje su "dalje" od najbližih susjeda

Jedna od najvažnijih značajki stupnja vrha kao mjere centralnosti jest uzimanje u obzir samo utjecaj najbližih susjeda promatranog vrha. U realnim situacijama nije dovoljno promatrati samo utjecaj najbližih susjeda, nego i utjecaj onih vrhova koji su na određenoj udaljenosti od promatranog vrha. Na slici 15 prikazana je jednostavna mreža u kojoj vrhovi, označeni s A i B, imaju stupanj 3. Susjedi vrha



Slika 15: Dva vrha imaju isti stupanj, ali različit broj vrhova koji su od njih udaljeni za dva.

A nemaju drugih susjeda tako da na A utječu jedino oni vrhovi koji su mu najbliži, tj. susjedni vrhovi. Međutim, svaki susjed vrha B ima po jednog susjeda koji može posredno utjecati na B. Ova situacija nam sugerira da bi vrh B mogao biti proglašen najutjecajnijim vrhom u ovoj mreži.

Na vrlo "umjetan" način stupanj vrha možemo izraziti kao zbroj potencija matrice susjedstva. Imamo

$$d_i = \left(\left[\left(A^0 + A^1 \right) - I \right] \cdot \mathbf{1} \right)_i, \tag{2}$$

gdje je I jedinična matrica, $(A^0)_{jl}$ računa broj šetnji duljine 0 od vrha j do vrha l, što je ekvivalentno ostanku u istom vrhu, dok $(A^1)_{jl}$ računa broj šetnji duljine 1 od j do l, što odgovara posjetu samo najbližem susjedu. Jedna mogućnost da se u obzir ne uzme samo utjecaj najbližih susjeda jest promatranje šetnji duljine veće od 1 koje počinju u promatranom vrhu. Sljedeće mjere centralnosti ubrajaju se u strategije koje rješavaju ovaj problem.

2.2.1 Svojstvena centralnost

Prirodni nastavak jednostavne centralnosti stupnja vrha jest svojstvena centralnost ili centralnost svojstvenog vektora. Znamo da susjedi nekog vrha u mreži općenito nisu ekvivalentni. Stoga je u mnogim okolnostima važnost vrha veća što je veća povezanost (putovima) s nekim drugim vrhovima koji su sami po sebi također važni. Upravo je to koncept svojstvene centralnosti. Umjesto da vrhovima dodjeljujemo po jedan *bod važnosti* za svakog susjeda, svojstvena centralnost svakom vrhu dodjeljuje broj bodova koji je proporcionalan sumi bodova njegovih susjeda.

Pretpostavimo da imamo početne vrijednosti x_i centralnosti vrhova i, $i = 1, \dots, n$, u neusmjerenoj mreži s n vrhova. Na primjer, pretpostavimo da je $x_i = 1$ za sve i. Jasno je da ovo nije korisna mjera centralnosti, ali nam može pomoći za izračunavanje bolje procjene x'_i, koju definiramo kao zbroj centralnosti svih susjeda vrha i na sljedeći način:

$$x'_i = \sum_{j=1}^{n} A_{ij} x_j,$$

gdje je A_{ij} odgovarajući element matrice susjedstva. Ovaj izraz možemo zapisati u matričnom obliku kao

$$x' = Ax,$$

gdje je $x = (x_1, \dots, x_n)^\tau$ i $x' = (x'_1, \dots, x'_n)^\tau$. Ponavljajući ovaj proces dobivanja "bolje" procjene za centralnost vrha, nakon k koraka dobivamo vektor centralnosti $x(k)$ u obliku:

$$x(k) = A^k x(0). \tag{3}$$

Sljedeći korak je zapisati $x(0)$ kao linearnu kombinaciju svojstvenih vektora v_i matrice susjedstva, te dobivamo

$$x(0) = \sum_{i=1}^{n} c_i v_i$$

za odgovarajući izbor konstanti $c_i \in \mathbb{R}$. Uvrstimo li to u jednadžbu (3), imamo:

$$x(k) = A^k \sum_{i=1}^{n} c_i v_i = \sum_{i=1}^{n} c_i \lambda_i^k v_i = \lambda_1^k \sum_{i=1}^{n} c_i \left[\frac{\lambda_i}{\lambda_1}\right]^k v_i,$$

gdje su λ_i svojstvene vrijednosti matrice A, a λ_1 je najveća od njih. Uočimo da vrijedi

$$\frac{\lambda_i}{\lambda_1} < 1, \quad \forall i \neq 1,$$

pa svi članovi sume, osim prvoga, eksponencijalno padaju kako k raste, te stoga u graničnom slučaju dobivamo $x(k) \to c_1\lambda_1^k v_1$. Drugim riječima, granični vektor centralnosti je proporcionalan svojstvenom vektoru koji pripada najvećoj svojstvenoj vrijednosti matrice susjedstva. Želimo li ovo ljepše zapisati, možemo reći da centralnost x zadovoljava jednadžbu

$$Ax = \lambda_1 x, \tag{4}$$

odnosno da je centralnost vrha i jednaka i-toj komponenti svojstvenog vektora pridruženog najvećoj svojstvenoj vrijednosti matrice A. Jednadžba (4) predstavlja svojstvenu centralnost koju je prvi predložio Bonacich [1] 1987. godine. Kao što smo ranije spomenuli, centralnost x_i vrha i proporcionalna je sumi centralnosti njegovih susjeda pa dobivamo

$$x_i = \lambda_1^{-1} \sum_j A_{ij} x_j. \tag{5}$$

Ovim izrazom možemo eksplicitno računati svojstvene centralnosti za pojedini vrh u mreži. Neki autori, kao na primjer [6], u svojim radovima za oznaku svojstvene centralnosti vrha koriste oznaku φ, te je stoga (5) ekvivalentno s

$$\varphi_1(i) = \left(\frac{1}{\lambda_1} A\varphi_1\right)_i. \tag{6}$$

Zbog jednostavnosti zapisa i daljnje analize, za svojstvenu centralnost koristit ćemo prethodni izraz.

Svojstvenu centralnost su, kao i centralnost stupnja vrha, proučavali i drugi matematičari dobivajući rezultate koji su svojstvenoj centralnosti davali prednost nasuprot stupnju vrha. U proučavanju svojstvene centralnosti daleko su otišli Cvetković i suradnici 1997. godine. Oni su pokazali da se svojstvena centralnost $\varphi_1(i)$ vrha i u mreži može shvatiti kao omjer broja šetnji duljine k koje počinju u vrhu i te ukupnog broja šetnji duljine k u nebipartitnoj povezanoj mreži, kada je duljina šetnje dovoljno velika. Mi ovu tvrdnju nećemo dokazivati, nego ćemo uputiti čitatelja na [2].

Vratimo se nakratko na mrežu prikazanu slikom 15. Računanjem svojstvene centralnosti vrhova A i B, dobivamo $\varphi_1(A) = 0.5$ i $\varphi_1(B) = 0.5745$. Ovo je i očekivano budući da na vrh B ne utječu samo susjedni vrhovi, nego i oni koji su od njega udaljeni za 2.

Primijetimo da izraz (5) svojstvenoj centralnosti daje lijepo svojstvo, a to je da centralnost vrha može biti jako velika zato što vrh u mreži može ili imati mnogo (možda i ne previše važnih) susjeda ili imati manje susjeda koji su vrlo važni ili pak oboje. Pojedinac u nekoj društvenoj mreži ovom mjerom centralnosti može biti vrlo važan najviše zbog toga što on ili ona poznaje mnogo ljudi (iako oni možda i nisu previše utjecajni) ili poznaje nekoliko ljudi na visokim i uglednim položajima. Također primijetimo da su sve vrijednosti svojstvene centralnosti nenegativne. Kako bismo se uvjerili u to, promotrimo što će se dogoditi ukoliko početni vektor $x(0)$ ima samo nenegativne elemente. Budući da su svi elementi matrice susjedstva također nenegativni, množenjem s A nikad ne možemo imati neki negativan element u promatranom vektoru. Stoga zaključujemo kako $x(k)$ iz jednadžbe (3) mora imati sve nenegativne elemente.

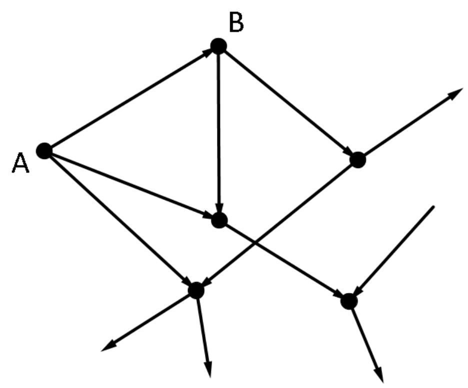
Teoretski se svojstvena centralnost može računati i za usmjerene mreže. Međutim, svojstvena centralnost je ipak najbolja mjera za neusmjerene mreže. U slučaju usmjerene mreže nastaju određene komplikacije. Prije svega, matrica susjedstva usmjerene mreže nije simetrična. Kao posljedica toga, postoje dva skupa svojstvenih vektora: skup lijevih i skup desnih svojstvenih vektora, a to znači da postoje i dva svojstvena vektora koji pripadaju najvećoj svojstvenoj vrijednosti. Stoga moramo definirati dvije svojstvene centralnosti, onu koja koristi desne i onu koja koristi lijeve svojstvene vektore matrice susjedstva:

$$\varphi_1^R(i) = \left(\frac{1}{\lambda_1} A \varphi_1^R\right)_i, \tag{7}$$

$$\varphi_1^L(i) = \left(\frac{1}{\lambda_1} A^T \varphi_1^L\right)_i. \tag{8}$$

Desna svojstvena centralnost (7) računa važnost vrha uzimajući u obzir važnosti vrhova na koje taj vrh pokazuje. Desnu svojstvenu centralnost možemo shvatiti kao širi koncept mjere izlaznog stupnja koja ne uzima u obzir samo najbliže susjede. S druge strane, lijeva svojstvena centralnost (8) mjeri važnost vrha uzimajući u obzir one vrhove koji pokazuju na njega. Lijevu svojstvenu centralnost tako možemo shvatiti kao proširenje mjere centralnosti ulaznog stupnja. Postavlja se pitanje koji je od ova dva vektora najprimjereniji za mjerenje svojstvene centralnosti u usmjerenim mrežama. U većini slučajeva prikladan je lijevi svojstveni vektor. Razlog je taj što centralnosti vrha u usmjerenim mrežama obično doprinose oni vrhovi koji pokazuju na njega, a ne oni vrhovi na koje pokazuje taj određeni vrh. Kako bismo ovo objašnjenje bolje shvatili, promotrimo, na primjer, jednu od najvećih svjetskih mreža – World Wide Web. Na WWW-u broj i ugled stranica koje pokazuju na našu web stranicu može dati razumnu indikaciju o tome koliko je važna ili korisna naša web stranica. S druge strane, činjenica da naša web stranica može pokazivati na druge važne stranice nije toliko bitna. Razlog za to je da bilo tko može postaviti da njegova stranica pokazuje na tisuće drugih, no to stranicu ne čini bitnom. Stoga se nameće zaključak da je ispravna definicija svojstvene centralnosti za vrh i u usmjerenoj mreži proporcionalna centralnostima vrhova koji pokazuju na fiksirani vrh i, odnosno možemo bez ikakve sumnje koristiti izraz (8) za računanje svojstvene centralnosti u usmjerenim mrežama.

Ipak, postoje i neki drugi problemi vezani za svojstvenu centralnost u usmjerenim mrežama. Promotrimo sliku 16. Vrh A je povezan sa ostatkom mreže samo sa izlaz-



Slika 16: Mreža s vrhovima A i B kojima je svojstvena centralnost jednaka nuli.

nim bridovima, odnosno, nema niti jedan ulazni brid. Svojstvena centralnost takvog vrha je jednaka nuli s obzirom da nema članova u sumi (5). Ali, ako promotrimo vrh B koji ima jedan ulazni brid, tada je i njegova svojstvena centralnost jednaka 0 iz razloga što taj ulazni brid dolazi iz vrha A. To znači da je za vrh B jedinstveni član sume (5) jednak 0. Zaključujemo da na promatrani vrh mogu pokazivati neki vrhovi na koje pokazuju neki drugi vrhovi itd. sve dok ne završimo s vrhom ili vrhovima koji imaju ulazni stupanj 0. Time će konačna vrijednost svojstvene centralnosti opet biti jednaka nuli. Ako bismo ovo htjeli objasniti jezikom teorije grafova, onda možemo reći da jedino vrhovi koji se nalaze u čvrsto povezanim komponentama mreže koje sadrže dva ili više vrhova, mogu imati svojstvenu centralnost različitu od nule. Međutim, u mnogim slučajevima je prikladno da vrhovi s velikim ulaznim stupnjem imaju veliku centralnost čak i ako se ne nalaze u čvrsto povezanoj komponenti. Na primjer, web stranice s mnogo linkova mogu se smatrati važnima čak i ako se ne nalaze u čvrsto povezanim komponentama. Sjetimo se acikličkih mreža. One ne sadrže čvrsto povezane komponente s više od jednog vrha. U takvim mrežama je svojstvena centralnost svakog vrha jednaka nuli. Iz ovog razmatranja je jasno da se svojstvena centralnost čini potpuno beskorisna za acikličke mreže. Stoga, ovakav slučaj možemo shvatiti kao nedostatak svojstvene centralnosti. Sljedeća mjera centralnosti predstavlja varijaciju svojstvene centralnosti te rješava gore navedeni problem.

2.2.2 Katzova centralnost

Prethodnu mjeru centralnosti završili smo s problemom koji predstavlja nedostatak spomenute mjere centralnosti. Jedno od rješenja za taj problem jest da svakom vrhu i, $i = 1, \ldots, n$, promatrane mreže dodamo neku konstantu bez obzira na njegov položaj u mreži ili centralnost njegovih susjeda. U tu svrhu definiramo:

$$x_i = \alpha \sum_{j=1}^{n} A_{ij} x_j + \beta, \tag{9}$$

gdje su α i β pozitivne konstante. Prvi član prethodnog izraza je svojstvena centralnost (u kojoj su zbrojene centralnosti onih vrhova koji pokazuju na vrh i) pomnožena konstantom α, a drugi član je slobodni član kojeg dobivaju svi vrhovi. Taj drugi član u izrazu (9) pridružuje vrijednost β svakom vrhu čiji je ulazni stupanj jednak nuli. Sada centralnost niti jednog vrha mreže nije jednaka nuli, te je ovo već svojevrsna prednost u odnosu na svojstvenu centralnost.

Izraz (9) možemo zapisati u matričnom obliku:

$$x = \alpha A x + \beta \cdot \mathbf{1}, \tag{10}$$

gdje je $\mathbf{1}$ vektor jedinica, odnosno $\mathbf{1} = (1, 1, 1, \ldots)^\tau$. Slijedi

$$x = \beta \left(I - \alpha A\right)^{-1} \cdot \mathbf{1}.$$

Kako nas točan iznos centralnosti vrha ne zanima, već nas zanimaju odnosi tih veličina za pojedine vrhove, vrijednost konstante β je nebitna pa nećemo pogriješiti ukoliko uzmemo da je $\beta = 1$. Dobivamo

$$x = \left(I - \alpha A\right)^{-1} \cdot \mathbf{1}. \tag{11}$$

Izraz (11) predstavlja mjeru centralnosti koju je predložio Katz 1953. godine [11] te je stoga poznata kao Katzova centralnost.

Katzova centralnost se razlikuje od obične svojstvene centralnosti u tome što sadrži parametar α koji regulira ravnotežu između promatrane komponente svojstvenog vektora i slobodnog člana jednadžbe (9). Dakle, ako želimo koristiti Katzovu centralnost kao mjeru centralnosti kojom ćemo pronaći najutjecajniji vrh u mreži, moramo prvo izabrati neku vrijednost parametra α. Pri tome je važno shvatiti da α ne može poprimiti bilo koju vrijednost. Ako stavimo da $\alpha \to 0$, tada će u (9) opstati jedino konstantni član pa će svi vrhovi imati istu centralnost β (za koju smo odabrali $\beta = 1$), što nikako nije korisno. Ako povećavamo vrijednost α, centralnosti će neograničeno rasti i eventualno divergirati. To će se dogoditi onda kada je $det(I - \alpha A)$ vrlo blizu nule. Možemo pisati

$$det(A - \alpha^{-1} I) = 0,$$

te vidimo da se radi o karakterističnoj jednadžbi čija su rješenja α^{-1} jednaka svojstvenim vrijednostima matrice susjedstva A. Kako se α povećava, determinanta se približava nuli kada je $\alpha^{-1} = \lambda_1$, odnosno kada je $\alpha = \frac{1}{\lambda_1}$. Dakle, ako želimo da Katzova centralnost konvergira, za α trebamo izabrati vrijednost koja je manja od recipročne vrijednosti prve svojstvene vrijednosti matrice susjedstva A.

Katzova centralnost se može direktno računati pomoću jednadžbe (11) tako da se računa inverz matrice na njenoj desnoj strani, ali to najčešće nije najbolji način za izračun Katzove centralnosti. Znamo da je računanje inverza matrice na računalu proporcionalno s n^3, gdje je n broj vrhova u mreži. Za jako velike mreže (npr. s više od tisuću vrhova), takav je izračun Katzove centralnosti užasno spor. U mnogim slučajevima je bolje koristiti jednadžbe (9) ili (10) i to tako da se uzme neka početna procjena za x (npr. $x = 0$) pa se takva procjena iskoristi za računanje iduće, bolje procjene:

$$x' = \alpha A x + \beta \cdot \mathbf{1}.$$

Ponavljajući ovaj postupak određeni broj puta, x će konvergirati prema veličini koja je blizu stvarne vrijednosti centralnosti. Kako matrica A ima m elemenata koji su različiti od nule, gdje je m broj bridova, svaka iteracija zahtjeva m operacija množenja te je ukupno vrijeme potrebno za računanje proporcionalno s $r \cdot m$, pri čemu je r broj iteracija koji je potreban za konvergenciju. Međutim, broj iteracija r ovisi o parametrima mreže i izboru konstante α, tako da ne postoji opći odgovor na pitanje koliko je iteracija potrebno za dovoljno precizan izračun. Umjesto toga, nužno je promatrati vrijednosti x_i i vidjeti konvergiraju li nekim konstantnim vrijednostima.

Iako Katzova centralnost rješava probleme koji se javljaju kod svojstvene centralnosti u usmjerenim mrežama, nema razloga da se takva mjera ne koristi i u neusmjerenim mrežama. Sama ideja dodavanja konstantnog člana u mjeru centralnosti kojom svaki vrh dobiva određenu težinu na temelju postojeće centralnosti, sasvim je prirodna. Ona omogućava vrhu koji ima mnogo susjeda veću vrijednost centralnosti bez obzira imaju li njegovi susjedi visoke centralnosti ili ne, a to može biti poželjno u nekim primjenama.

Katzovu centralnost možemo definirati i na drugačiji način. Možemo konstantni član β u jednadžbi (9) definirati posebno za svaki vrh. Time dobivamo poopćenu mjeru centralnosti

$$x_i = \alpha \sum_j A_{ij} x_j + \beta_i, \tag{12}$$

gdje je β_i neki unutarnji doprinos centralnosti vrha i, tj. onaj koji ne ovisi o strukturi mreže. Na primjer, u društvenoj mreži važnost pojedinca može ovisiti o nemrežnim faktorima kao što su njegove godine ili mjesečni prihod (ukoliko imamo informacije o tim faktorima) te ih možemo uključiti u vrijednosti β_i. Tada vektor centralnosti x izgleda ovako

$$x = (I - \alpha A)^{-1} \beta,$$

pri čemu je β vektor s komponentama β_i. Jedna zgodna značajka ovog pristupa je da se račun za inverz matrice A mora napraviti samo jednom za promatranu mrežu i vrijednost α. Stoga za različite odabire β_i ne moramo iznova računati inverz matrice, nego postojeći inverz moramo pomnožiti s različitim vektorima β.

2.2.3 PageRank centralnost

Jedna od najvažnijih primjena mjera centralnosti je rangiranje navoda web stranica na World Wide Webu. Važnost ovog rangiranja je očita kada se uzme u obzir da kvaliteta našeg pretraživanja kroz Web ovisi o tome. Ako, na primjer, pretražujemo izraz "*mjera centralnosti*" na Google tražilici, očekujemo da ćemo za ispis dobiti rangirane web stranice koje su usko vezane uz naš traženi izraz, pri čemu je najposjećenija stranica na vrhu dane liste. Ili, na primjer, ako posjedujemo vlastitu web stranicu, tada poznati web imenik *Yahoo!* može sadržavati link na našu web stranicu iako ima linkove na još milijune drugih stranica. Znamo da je *Yahoo!* važna web stranica i mora imati veliku centralnost po svim prethodno spomenutim mjerama centralnosti. Treba li zbog toga moja web stranica biti također važna samim time što na nju upućuje *Yahoo!*? Mnogi se stručnjaci slažu da ne treba − visoka centralnost *Yahoo!*-a će se na neki način razdijeliti po svim onim stranicama čiji se link nalazi na *Yahoo!*.

Upravo gore opisan nedostatak Katzove centralnosti možemo ukloniti tako da definiramo varijaciju Katzove centralnosti u kojoj je centralnost koju vrh dobiva obzirom na svoje susjede proporcionalna centralnošću tih susjeda podijeljenom sa njihovim izlaznim stupnjem. Dakle, iako neki vrh može imati veliku centralnost, ako mu je izlazni stupanj vrlo visok, samo će mali dio te centralnosti biti prenesen na susjedne vrhove. Ovakvu varijaciju Katzove centralnosti za vrh i definiramo s

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{d_j^{out}} + \beta. \tag{13}$$

Međutim, tu nastaje problem ako u mreži postoje vrhovi čiji je izlazni stupanj jednak nuli. Ako je takav vrh, npr. vrh k, onda je prvi član izraza (13) neodređen, zato što je $d_k^{out} = 0$ i $A_{ik} = 0,\ \forall i$. Ovaj problem možemo lako riješiti tako da postavimo $d_j^{out} = 1$ za sve takve vrhove jer oni ionako ne doprinose centralnosti ostalih vrhova. Zapravo, možemo staviti da je $d_j^{out} > 0$ i to opet neće utjecati na konačno rješenje. Ukoliko želimo jednadžbu (13) zapisati u matričnom obliku, tada imamo

$$x = \alpha A D^{-1} x + \beta \cdot \mathbf{1},$$

gdje je $\mathbf{1}$ vektor jedinica, odnosno $\mathbf{1} = (1, 1, 1, \ldots)^\tau$, a D je dijagonalna matrica čiji su dijagonalni elementi dani s $D_{ii} = \max\{d_i^{out}, 1\}$. Postupimo li sada kao i kod Katzove centralnosti, dobivamo

$$x = \beta \left(I - \alpha A D^{-1}\right)^{-1} \cdot \mathbf{1}.$$

I ovdje β nema pretjerano važnu ulogu pri množenju pa možemo staviti da je $\beta = 1$. Time dobivamo

$$x = \left(I - \alpha A D^{-1}\right)^{-1} \cdot \mathbf{1},$$

odnosno

$$\pi = D \cdot (D - \alpha A)^{-1} \cdot \mathbf{1}, \tag{14}$$

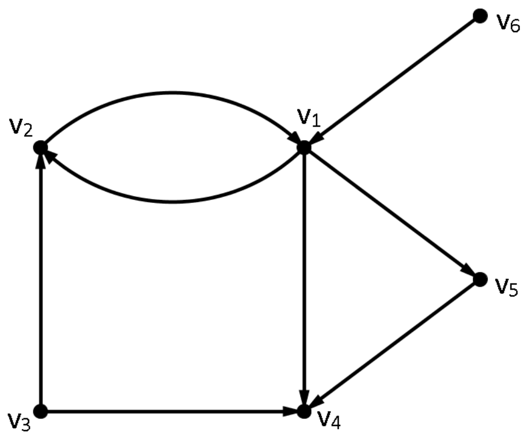
pri čemu smo uveli novu oznaku "π" za vektor PageRank centralnosti. Izraz (14) predstavlja mjeru centralnosti koja je obično poznata pod nazivom PageRank, što je brand Google-ove korporacije za web pretraživanje i središnji je dio njihove tehnologije rangiranja web stranica. Osnovna ideja koja se krije iza PageRank-a je da se važnost web stranice oslanja na važnosti web stranica koje pokazuju na tu web stranicu. Dodatno svojstvo ove mjere, kojim se centralnosti vrhova dijele sa pripadnim izlaznim stupnjem, osigurava da ona stranica koja pokazuje na ogroman broj drugih stranica ne prenosi svoju eventualno visoku centralnost tim stranicama. Drugačije rečeno, PageRank jedne stranice jednak je zbroju PageRank-ova svih stranica koje pokazuju na promatranu web stranicu. Druga intuitivna ideja utjelovljena u PageRank-u je važnost činjenice da jedna relevantna web stranica pokazuje na druge web stranice koje ovise o početnom broju preporuka na početnoj stranici. Ovo je sasvim prirodno jer niti jedna ozbiljna web stranica neće postaviti link na neku neprovjerenu i nekvalitetnu web stranicu. Stoga PageRank algoritam relevantnost web stranice određuje na temelju relevantnosti web stranica koje imaju linkove na promatranu stranicu.

Vratimo se nakratko na web tražilicu Google koja koristi PageRank. Cilj Google-ove web tražilice je generiranje liste korisnih web stranica iz unaprijed sastavljenog indeksa stranica u rasponu tekstualnih upita. Prvo se pretražuju indeksi stranica koji odgovaraju zadanom upitu koristeći jednostavne kriterije kao što je odgovarajući tekst, a zatim se rangiraju odgovori prema rezultatima temeljenima na kombinaciji različitih vrsta rangiranja (jedan od njih je PageRank). Google vraća korisne odgovore na postavljene upite ne zato što je bolji u pronalaženju relevantnih stranica, nego zato što je bolji u odlučivanju kako predstaviti dobivene rezultate. Točnost percipiranja rezultata nastaje zato što se na vrhu dobivene liste nalaze oni odgovori koji su najčešći rezultati za zadani upit, te je isto tako (i vrlo vjerojatno) moguće da se mnogi rijetki odgovori nađu na toj listi, ali na dnu popisa.

Primijetimo da formula za PageRank, jednadžba (14), sadrži jedan slobodan parametar α čija se vrijednost mora odabrati prije nego se iskoristi formula za izračunavanje vrijednosti centralnosti. Iskoristimo li analogiju sa Katzovom centralnosti, možemo vidjeti da vrijednost parametra α mora biti manja od inverza najveće svojstvene vrijednosti matrice AD^{-1} koja ima svojstvo da je stohastička po stupcima. Za kvadratnu matricu kažemo da je *stohastička po stupcima* ako su svi njezini elementi nenegativni i suma elemenata u svakom stupcu je jednaka 1. U slučaju neusmjerene mreže vrijedi da je $\lambda_1 = 1$, te je odgovarajući svojstveni vektor jednak $(d_1, d_2, d_3, \ldots, d_n)^\tau$, gdje je d_i stupanj i-tog vrha, $i = 1, \ldots, n$. Do ovoga rezultata lako možemo doći ako promotrimo matricu $(AD^{-1})^\tau$. Ona ima svojstvo da je suma elemenata u svakom retku jednaka 1. To implicira da je jedna svojstvena vrijednost te matrice jednaka 1 s pripadnim svojstvenim vektorom $\mathbf{1} = (1, 1, 1, \ldots)^\tau$, a prema Gerschgorinovom teoremu [9] to je ujedno i najveća svojstvena vrijednost matrice $(AD^{-1})^\tau$. Zbog činjenice da se spektar matrice ne mijenja ako ju transponiramo, slijedi da je $\lambda_1(AD^{-1}) = 1$. Dokaz da je svojstveni vektor od AD^{-1} pridružen $\lambda_1 = 1$ jednak $(d_1, d_2, d_3, \ldots, d_n)^\tau$ prepuštamo čitatelju. U slučaju usmjerene mreže, ovaj rezultat ne vrijedi jer je općenito najveća svojstvena vrijednost različita od 1.

Klasični izbor vrijednosti parametra α kojeg koristi Google tražilica jest $\alpha = 0.85$ premda nije jasno stoji li iza te vrijednosti neka velika i teška teorija ili je to rezultat nekog eksperimentiranja koje je provedeno kako bi se saznalo uz koju vrijednost parametra je formula za PageRank najučinkovitija.

Pogledajmo kako funkcionira PageRank na nekom konkretnom primjeru. Promotrimo usmjerenu mrežu web stranica prikazanu na slici 17. PageRank centralnosti



Slika 17

vrhova v_i, $i = 1, \ldots, 6$, te mreže dani su u tablici 1. Prema PageRank-u, najvažnija

vrh	v_1	v_2	v_3	v_4	v_5	v_6
π	0.2665	0.1697	0.0611	0.2900	0.1416	0.0611

Tablica 1: PageRank centralnosti vrhova mreže prikazane na slici 17.

web stranica je ona označena s v_4 jer na nju pokazuje najveći broj linkova. Najzanimljiviju situaciju dobivamo iz analize vrhova v_1 i v_2. Vidimo da svaka od tih stranica ima dva dolazna linka, ali PageRank veću važnost daje stranici v_1 u odnosu na stranicu v_2. Zašto? Najvažnija stranica koja pokazuje na v_2 je stranica v_1. Ali, v_1 pokazuje na još dvije stranice osim v_2 pa je relevantnost stranice v_1 na neki način podijeljena na te tri stranice. Međutim, najvažnija stranica koja pokazuje na v_1 je stranica v_2, koja preporučuje samo tu stranicu. To znači da se njezina preporuka ne dijeli na ostale stranice te iz toga razloga stranica v_1 dobiva veći rang u odnosu na stranicu v_2.

Kao i Katzovu centralnost, tako i PageRank centralnost možemo poopćiti tako da konstantni član β u jednadžbi (9) ovisi o pojedinom vrhu:

$$x_i = \alpha \sum_{j=1}^{n} A_{ij} \frac{x_j}{d_j^{out}} + \beta_i.$$

Matrični oblik daje vektor centralnosti definiran s

$$x = D(D - \alpha A)^{-1} \beta.$$

Ovakvu varijantu PageRank centralnosti možemo koristiti za rangiranje web stranica uzimajući za β_i vrijednost temeljenu na tekstualnoj relevantnosti upita za pretraživanje. To znači da stranica koja sadržava riječ ili riječi koja je tražena na mnogo češćim ili istaknutijim mjestima može dobiti višu unutarnju centralnost od ostalih stranica, pa će time zauzeti više mjesto na rangiranoj listi.

2.2.4 Podgrafovska centralnost

Ideja podgrafovske centralnosti jest da važnost promatranog vrha možemo ocijeniti s obzirom na njegovu pripadnost zatvorenim šetnjama koje počinju i završavaju u tom vrhu. Prisjetimo se da se zatvorena šetnja duljine l, koja počinje u vrhu i, može dobiti kao i-ti dijagonalni element l-te potencije matrice susjedstva A zadane mreže. To znači da podgrafovsku centralnost vrha i možemo definirati sljedećom funkcijom:

$$f_i(A) = \left(\sum_{l=0}^{\infty} c_l A^l \right)_{ii}, \tag{15}$$

gdje se koeficijenti c_l biraju tako da navedena beskonačna suma konvergira. Znamo da je broj zatvorenih šetnji duljine $l = 0$ koje uključuju promatrani vrh jednak 1. Također, broj zatvorenih šetnji duljine $l = 2$ koje počinju u vrhu i jednak je njegovu stupnju d_i. Ukoliko mreža ne sadrži petlje, broj zatvorenih šetnji duljine $l = 1$ je jednak 0, što nam omogućuje da izraz (15) zapišemo kao:

$$f_i(A) - 1 = c_2 d_i + \left(\sum_{l=3}^{\infty} c_l A^l \right)_{ii}.$$

U tom slučaju koeficijente c_l možemo odabrati tako da stupanj vrha daje najveći doprinos funkciji $f_i(A)$, a kako duljina zatvorene šetnje raste, njihov doprinos ukupnoj centralnosti se smanjuje. Drugim riječima, koeficijente c_l možemo birati tako da kraće zatvorene šetnje, koje počinju u promatranom vrhu, daju veći doprinos centralnosti nego što ga daju dulje zatvorene šetnje.

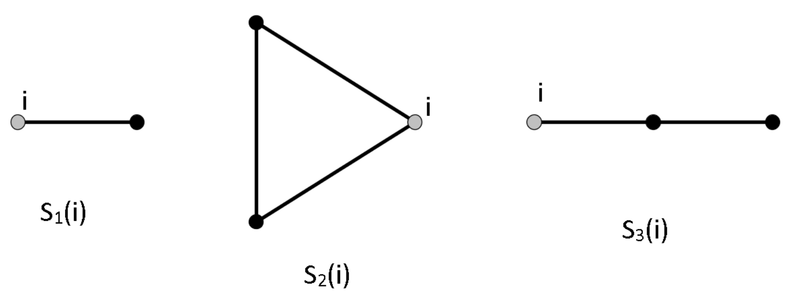
Objasnimo sada zašto se ovaj indeks centralnosti naziva *podgrafovska* centralnost. Zapišimo indeks u terminima prvih nekoliko potencija matrice susjedstva:

$$f_i(A) - 1 = c_2(A^2)_{ii} + c_3(A^3)_{ii} + c_4(A^4)_{ii} + c_5(A^5)_{ii} + \cdots$$

Svaki dijagonalni član potencije matrice susjedstva se može zapisati u terminu broja pojavljivanja odgovarajućeg vrha u korijenskom podgrafu određenog tipa, kao npr.:

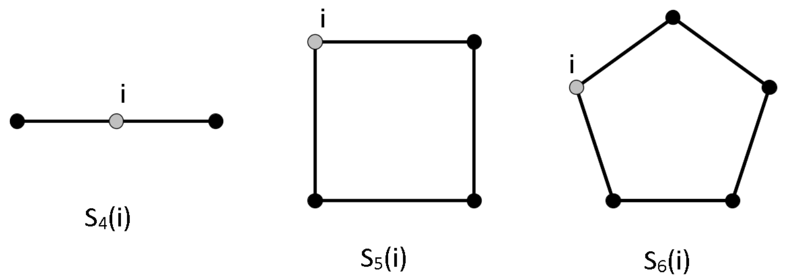
$$
\begin{aligned}
(A^2)_{ii} &= |S_1(i)|, && (16)\\
(A^3)_{ii} &= 2\,|S_2(i)|, && (17)\\
(A^4)_{ii} &= |S_1(i)| + |S_3(i)| + 2\,|S_4(i)| + 2\,|S_5(i)|, && (18)\\
(A^5)_{ii} &= 10\,|S_2(i)| + 2\,|S_6(i)| + 2\,|S_7(i)| + 4\,|S_8(i)| + 2\,|S_9(i)|, && (19)
\end{aligned}
$$

pri čemu su odgovarajući korijenski podgrafovi $S_k(i)$, $k = 1, \ldots, 9$, prikazani na slikama 18, 19 i 20, a sa $|S_k(i)|$ je označen broj pojavljivanja vrha i u korijenskom podgrafu $S_k(i)$.

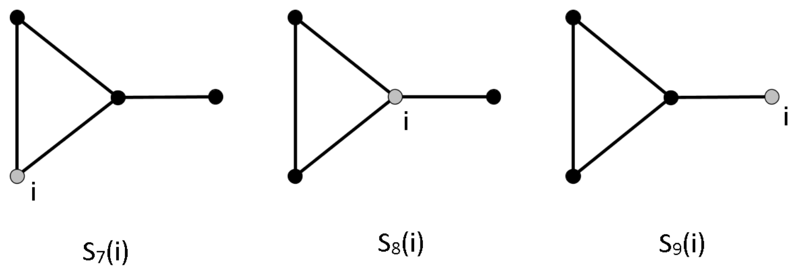


Slika 18: Korijenski podgrafovi I.

Sada $f_i(A)$ možemo zapisati kao težinsku sumu pripadnosti vrha i navedenim



Slika 19: Korijenski podgrafovi II.



Slika 20: Korijenski podgrafovi III.

korijenskim podgrafovima:

$$\begin{aligned} f_i(A) \;=\; & (c_2 + c_4 + \cdots)\,|S_1(i)| + (2c_3 + 10c_5 + \cdots)\,|S_2(i)| + \\ & +(c_4 + \cdots)\,|S_3(i)| + (c_4 + \cdots)\,|S_4(i)| + (2c_4 + \cdots)\,|S_5(i)| + \\ & +(2c_5 + \cdots)\,|S_6(i)| + (2c_5 + \cdots)\,|S_7(i)| + (4c_5 + \cdots)\,|S_8(i)| + \\ & +(2c_5 + \cdots)\,|S_9(i)| + \cdots \end{aligned} \tag{20}$$

Postoji mnogo načina za odabir koeficijenata ukoliko je glavna ideja da oni ispunjavaju oba zadana uvjeta − konvergencijski i težinski. Jedan od posebno korisnih načina je onaj koji se koristi za definiranje Estradinog indeksa, a koji konvergira prema eksponencijalnoj matričnoj funkciji e^A. Mi se ovdje nećemo baviti Estradinim indeksom (čitatelj se o njemu može informirati u [6]), budući da je to veličina koja opisuje globalno svojstvo mreže, ali ćemo iz njega izdvojiti doprinos $EE(i)$ vrha i:

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

koji je u našem kontekstu upravo podgrafovska centralnost vrha i. Na isti način možemo definirati podgrafovske centralnosti koje uzimaju u obzir posebno doprinose neparnih, a posebno doprinose parnih šetnji u mreži:

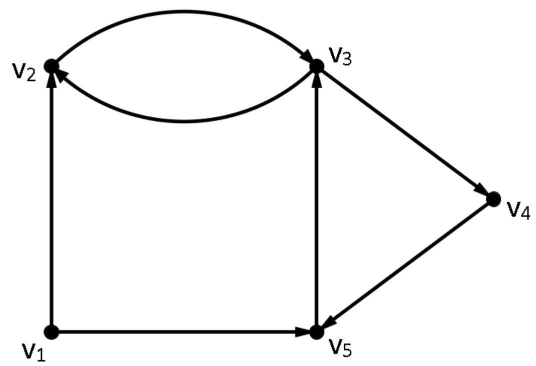
$$EE_{par}(i) = (\mathrm{sh}A)_{ii},$$

$$EE_{nepar}(i) = (\mathrm{ch}A)_{ii}.$$

Stoga se u nekim slučajevima prethodni izrazi koriste za lakše izračunavanje podgrafovske centralnosti neke neusmjerene mreže.

Karakteristika podgrafovske centralnosti u usmjerenim mrežama jest da uzima u obzir sudjelovanje promatranog vrha u usmjerenim šetnjama. Prisjetimo se da je usmjerena šetnja zadana nizom usmjerenih bridova oblika $uv, vw, wx, \ldots, yz, zu$. To znači da je podgrafovska centralnost $EE(i)$ vrha i u usmjerenoj mreži veća od jedan samo ako postoji barem jedna zatvorena šetnja koja počinje i završava u vrhu i. Inače je $EE(i) = 1$. Prema tome, podgrafovska centralnost vrha u usmjerenoj mreži ukazuje na povratak informacije, koja je poslana iz promatranog vrha, opet u taj isti vrh.

Promotrimo mrežu na slici 21. Uočimo da se informacija poslana iz vrha v_1 ne



Slika 21

može vratiti u taj isti vrh jer ne postoji niti jedna usmjerena šetnja koja počinje i završava u v_1. S druge strane, vrh v_2 sudjeluje u zatvorenim šetnjama $v_2 - v_3 - v_2$ i $v_2 - v_3 - v_4 - v_5 - v_3 - v_2$. Međutim, najutjecajniji vrh u zadanoj mreži je vrh v_3 koji sudjeluje u zatvorenim šetnjama $v_3 - v_2 - v_3$ i $v_3 - v_4 - v_5 - v_3$. U tablici 2 dani su podaci o podgrafovskim centralnostima vrhova mreže sa slike 21.

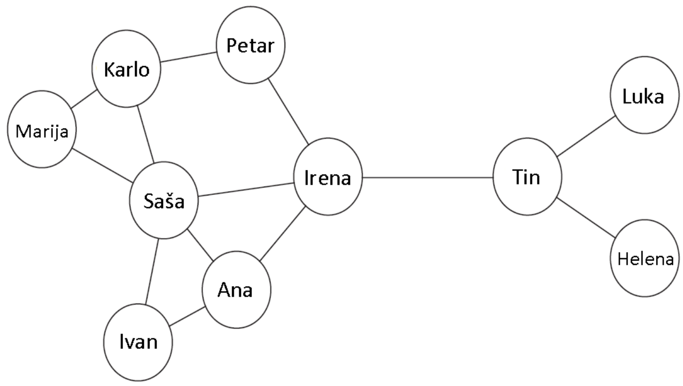
vrh	v_1	v_2	v_3	v_4	v_5
EE	1.000	1.552	1.728	1.177	1.177

Tablica 2: Podgrafovske centralnosti vrhova mreže prikazane na slici 21.

Korisno je napomenuti da se informacije sadržane u podgrafovskoj centralnosti razlikuju kada je riječ o usmjerenom i neusmjerenom slučaju iste mreže. Drugim riječima, oni vrhovi mreže koji su proglašeni najcentralnijima u podgrafovskoj centralnosti usmjerene mreže, mogu se i ne moraju pojaviti kao najcentralniji vrhovi u podgrafovskoj centralnosti neusmjerene verzije iste mreže. Kao glavni zaključak vezan uz analizu podgrafovske centralnosti, nameće se tvrdnja da moramo biti jako oprezni sa korištenjem podgrafovske centralnosti s obzirom na problem koji promatramo.

2.3 Udaljenost vrhova

U nekim situacijama nas neće zanimati oni vrhovi koji su centralni zbog broja veza ili odnosa sa drugim dobro povezanim vrhovima, nego će nas zanimati oni vrhovi koji su relativno blizu, odnosno daleko u odnosu na ostale vrhove u mreži. Pretpostavimo da želimo dati neku informaciju pojedincu u društvenoj mreži tako da ju on podijeli sa ostalim članovima mreže kroz najmanji mogući broj posrednika. Ako pogledamo, na primjer, malu društvenu mrežu prikazanu na slici 22, mislimo da bi davanje takve informacije Saši bilo dobra polazna točka budući da je on osoba u mreži koja ima najviše veza. Međutim, iako Saša može prenijeti informaciju pet drugih osoba



Slika 22: Blizina vrhova u maloj društvenoj mreži

u samo jednom koraku, ta ista informacija će trebati tri koraka kako bi došla do Helene i Luke. S druge strane, ukoliko tu informaciju damo Ireni kao početnu točku, tada će informacija doći do svakog člana promatrane mreže u najviše dva koraka, što je dosta ekonomičnije nego u slučaju sa Sašom. Zaključujemo kako bi bilo korisno definirati mjeru centralnosti koja se odnosi na međusobnu udaljenost vrhova u mreži.

U ovakvim slučajevima koristi se mjera centralnosti poznata pod nazivom *centralnost blizine* koja je definirana s:

$$CC(u) = \frac{n-1}{s(u)}, \tag{21}$$

gdje je n broj vrhova u mreži G, a $s(u)$ je suma udaljenosti vrha u do svih ostalih vrhova u G, odnosno:

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

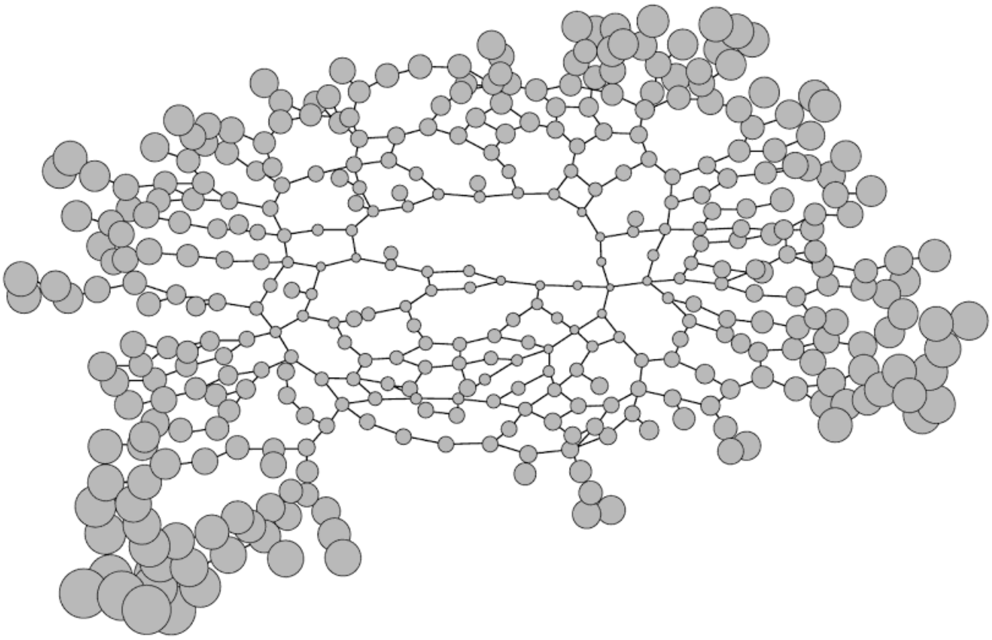
U tablici 3 prikazane su vrijednosti za centralnost blizine društvene mreže sa slike 22. Kao što možemo vidjeti, najveću centralnost blizine imaju Irena i Saša, dok najmanju centralnost blizine imaju Helena i Luka.

vrh v	$CC(v)$	vrh v	$CC(v)$
Irena	0.6428	Karlo	0.4500
Saša	0.6000	Marija	0.4286
Ana	0.5294	Ivan	0.4090
Tin	0.5000	Helena	0.3462
Petar	0.4737	Luka	0.3462

Tablica 3: Vrijednosti centralnosti blizine za vrhove u mreži prikazanoj na slici 22.

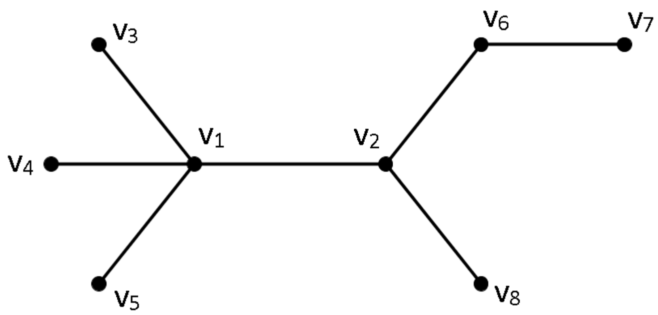
Kao što nas može zanimati koliko su vrhovi u mreži relativno blizu jedan drugome, tako nas može zanimati i informacija koliko su vrhovi u mreži relativno daleko jedan od drugoga. Recipročna vrijednost centralnosti blizine poznata je pod nazivom *centralnost daljine.* Slika 23 prikazuje gradsku mrežu starog dijela grada Cordobe u Španjolskoj. Bridovi ovakvog tipa mreže predstavljaju dijelove ulica, a vrhovi predstavljaju križanja ulica. Vrhovi mreže na slici 23 prikazani su u obliku kruga čija je veličina proporcionalna centralnosti daljine. Možemo vidjeti da centralnost daljine jako dobro definira strukturu centra i periferije grada. Gradski centar ili jezgra nalazi se u geometrijskom centru gdje većina mjesta ima malu centralnost daljine, a periferija grada karakterizirana je mjestima koja imaju velike centralnosti daljine.

Centralnost blizine je vrlo prirodna mjera centralnosti i često se koristi u istraživanjima društvenih i ostalih mreža. Međutim, kao i kod ostalih mjera centralnosti, tako i kod ove mjere možemo naići na neke probleme. Jedan od takvih problema jest da vrijednosti ovakve centralnosti vrhova imaju vrlo malen raspon, odnosno razlika između najmanje i najveće moguće vrijednosti koju ova mjera može poprimiti u nekoj zadanoj mreži je vrlo mala. Tipično svojstvo mnogih kompleksnih mreža jest da je udaljenost između vrhova obično malena, te se povećava logaritamski u ovisnosti o broju vrhova mreže. Stoga je u praksi, rabeći ovu mjeru, teško razlikovati centralne i nešto manje centralne vrhove. Vrijednosti koje dobivamo imaju tendenciju da budu zgusnute, a razlike između susjednih vrijednosti vidljive su samo onda kada se detaljno pregledaju dobivene brojčane vrijednosti. To znači da bilo kakva mala promjena strukture mreže može znatno promijeniti redoslijed vrijednosti, pa čak i dati posve drugačiji centralni vrh.



Slika 23: Gradska mreža starog dijela grada Cordobe, Španjolska. Slika je preuzeta iz [6].

Postoji još jedan problem sa centralnosti blizine/daljine. Promotrimo mrežu na slici 24. Vrhovi v_1 i v_2 imaju isti zbroj udaljenosti do ostalih vrhova mreže



Slika 24: Vrhovi v_1 i v_2 imaju istu centralnost blizine/daljine.

$(s(v_1) = s(v_2) = 11)$, te prema tome imaju i istu vrijednost centralnosti blizine, iako se intuitivno vrh v_2 čini više centralnijim vrhom nego vrh v_1. Možemo vidjeti da je vrh v_1 udaljen za 3 od vrha v_7, dok je udaljenost vrha v_2 do svih ostalih vrhova najviše 2. Iz toga razloga nam se vrh v_2 čini više centralnim od vrha v_1.

Ovakav problem možemo riješiti ako u obzir uzmemo varijantu svojstvene centralnosti za duljine najkraćih puteva između vrhova u mreži. Umjesto matrice susjedstva, promatramo matricu udaljenosti U te koristimo njena spektralna svojstva iz kojih proizlazi da je i-ti član $\delta_1(i)$ svojstvenog vektora δ_1 pridruženog najvećoj svojstvenoj vrijednosti matrice udaljenosti jednak *spektralnoj centralnosti daljine* vrha i, odnosno

$$\delta_1(i) = \left(\frac{1}{\eta_1} U \delta_1\right)_i, \tag{22}$$

gdje je η_1 najveća svojstvena vrijednost matrice U, a δ_1 odgovarajući svojstveni vektor. Detalji o spektralnim svojstvima matrice udaljenosti nalaze se npr. u [16]. Odnosi između centralnosti daljine i spektralne centralnosti daljine vrhova mreže sa slike 24 prikazani su u tablici 4. Iz dane tablice možemo vidjeti da su centralnost

vrh	FC	δ_1	$(\delta_1)^{-1}$
v_1	1.5714	0.2527	3.9573
v_2	1.5714	0.2518	3.9714
v_3	2.4286	0.3771	2.6518
v_4	2.4286	0.3771	2.6518
v_5	2.4286	0.3771	2.6518
v_6	2.1426	0.3278	3.0506
v_7	3.0000	0.4440	2.2522
v_8	2.4286	0.3763	2.6575

Tablica 4: Centralnosti daljine FC, spektralne daljine δ_1 i spektralne blizine $(\delta_1)^{-1}$ vrhova mreže prikazane na slici 24.

daljine i spektralna centralnost daljine proporcionalne veličine.

Na sličan način kao i za daljinu, možemo definirati spektralnu centralnost blizine tako da računamo inverz od δ_1. Blizina/daljina vrhova i njihovi spektralni analozi u strogoj su korelaciji za većinu mreža (vidi tablicu 4), a to znači da korištenje spektralnih mjera ima mali utjecaj na rangiranje dobiveno pomoću nespektralnih mjera. Međutim, spektralne mjere ipak mogu biti korisne ukoliko se u nespektralnim mjerama pojave situacije kada neka dva vrha imaju istu centralnost blizine/daljine, a obzirom na strukturu mreže, to ne bi trebalo biti tako.

2.4 Međupoloženost vrhova

Vrlo drugačiji koncept centralnosti predstavlja centralnost međupoloženosti vrha, koja daje informaciju o tome u kojoj se mjeri promatrani vrh nalazi na putovima između drugih vrhova. Ideja međupoloženosti se obično pripisuje Freemanu, [7] 1977. godine, iako postoje i raniji radovi koji su se bavili ovom problematikom.

Pretpostavimo da imamo mrežu sa zadanim protokom. Na primjer, u društvenoj mreži razne poruke, novosti, informacije ili glasine mogu se prenositi od jedne osobe do druge. Pretpostavimo da svaki par vrhova u mreži, između kojih postoji najmanje jedan put, razmjenjuje poruku s jednakom vjerojatnošću po jedinici vremena i da poruke uvijek biraju najkraći put kroz mrežu (slučajno odabran, ako imamo više najkraćih putova). Zanima nas sljedeće: ako čekamo dovoljno dugo, odnosno ako dovoljno velik broj poruka prođe između proizvoljna dva vrha mreže, koliki je prosječan broj poruka koji prolazi kroz promatrani vrh mreže na putu do njihovog odredišta? Budući da poruke prolaze kroz mrežu jednakom brzinom neovisno o izboru puta, to je broj onih poruka koji prolaze kroz svaki vrh proporcionalan ukupnom broju najkraćih putova kojima taj vrh pripada. Takav broj putova je upravo ono što nazivamo *centralnost međupoloženosti*.

Vrhovi čija je centralnost međupoloženosti prilično velika, mogu imati značajan utjecaj unutar promatrane mreže, najviše zbog toga što imaju kontrolu nad informacijama koje prolaze kroz njih. Ako uzmemo u obzir društvenu mrežu kroz koju putuju određene informacije, tada su vrhovi s najvećom međupoloženosti oni vrhovi kroz koje prolazi najveći broj informacija pa mogu izvući puno koristi iz svojeg položaja unutar mreže (npr. mogu tražiti plaćanje prolaska informacije kroz njih i sl.). Vrhovi s najvećom međupoloženosti su također oni čije će uklanjanje iz mreže najviše poremetiti komunikaciju između ostalih vrhova u mreži zbog toga što se oni nalaze na najvećem broju putova kroz koje prolaze informacije. U realnim situacijama je normalno da svi vrhovi mreže ne razmjenjuju podatke jednakom frekvencijom i u većini slučajeva komunikacija ne ide uvijek najkraćim putom. Ipak, centralnost međupoloženosti može ukazivati na utjecaj vrhova mreže na protok informacija putem drugih vrhova u istoj mreži.

Nakon što smo se upoznali sa osnovnom idejom međupoloženosti, sada ju možemo precizno definirati. Zbog jednostavnosti ćemo pretpostaviti da imamo neusmjerenu mrežu u kojoj postoji najviše jedan najkraći put između bilo koja dva vrha u mreži (dakle, može se dogoditi da ne postoji niti jedan najkraći put između neka dva vrha, odnosno da je mreža nepovezana). Centralnost međupoloženosti vrha i definira se kao broj svih najkraćih putova koji prolaze kroz i.

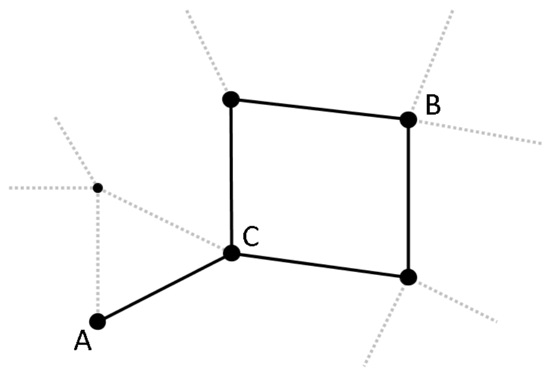
Neka je $n_{st}^i = 1$ ako vrh i leži na putu od vrha s do vrha t, a $n_{st}^i = 0$ ukoliko takav put ne postoji. Tada je centralnost međupoloženosti vrha i definirana s

$$BC(i) = \sum_{s,t} n_{st}^i, \tag{23}$$

pri čemu sumiramo po svim uređenim parovima vrhova u mreži. Imajmo na umu da definicija (23) posebno bilježi putove u bilo kojem smjeru između svakog para vrhova u neusmjerenoj mreži. Kako su ti putovi isti, svaki će se put bilježiti dvaput. To se može korigirati dijeljenjem s 2, ali nije potrebno. Navest ćemo dva razloga u korist prethodne tvrdnje. Prvo, u praksi nema nikakve razlike dijeli li se centralnost s 2 ili ne zbog toga što nas uglavnom zanimaju samo relativne vrijednosti centralnosti. Drugi je razlog što se izraz (23) može neizmijenjen koristiti u slučaju usmjerene mreže u kojoj postoji jasna razlika između puta od vrha s do vrha t, te puta od vrha t do vrha s. Također je vidljivo da definicija međupoloženosti uključuje putove od svakog vrha do sebe samoga. Svaki vrh leži na putu od sebe do sebe, tako da uključivanje ovog člana u sumu povećava vrijednost centralnosti za 1, ali to ne utječe na rangiranje vrhova nakon što se izračunaju vrijednosti centralnosti za svaki vrh.

Također je upitno treba li put od vrha s do vrha t računati kao put koji prolazi vrhom s, odnosno vrhom t. Čini se razumno definirati situaciju da vrh bude na putu između sebe i nekog drugog vrha, zato što vrh ima kontrolu nad informacijama koje teku od njega prema nekom drugom vrhu i obrnuto. Ako se isključe krajnje točke puta (što je čest slučaj u društvenim mrežama), onda se broj putova kroz svaki vrh smanjuje za dvostruki broj vrhova komponente povezanosti mreže u kojoj se vrh nalazi. Na taj način se međupoloženost svih vrhova unutar jedne komponente smanjuje za neku aditivnu konstantu, a poredak vrhova, dobiven vrijednostima ove centralnosti, ostaje nepromijenjen. Poredak vrhova u različitim komponentama može se promijeniti, no to je rijedak problem jer se centralnost međupoloženosti obično ne koristi za usporedbu vrhova koji dolaze iz različitih komponenata povezanosti mreže, zato što se ti vrhovi ne natječu za najcentralnijeg u istom "području" promatrane mreže.

Sve što smo prethodno analizirali, odnosi se na slučaj u kojemu postoji najviše jedan najkraći put između svaka dva vrha. Općenito, broj takvih putova može biti proizvoljan. To možemo vidjeti iz slike 25. Vrhovi A i B povezani su s dva najkraća puta, a vrh C leži na oba puta. Stoga standardno proširenje međupoloženosti u



Slika 25: Dio mreže u kojoj je broj najkraćih putova između vrhova A i B veći od jedan.

ovom slučaju pridružuje svakom putu težinu jednaku inverznom broju ukupnog broja svih najkraćih putova između promatrana dva vrha. Na primjer, ako postoje dva najkraća puta između danog para vrhova (kao na slici 25), tada svaki od putova dobiva težinu $\frac{1}{2}$. Tada je međupoloženost vrha definirana kao suma težina svih putova koji prolaze kroz taj vrh. Ako dva ili više putova prolaze kroz isti vrh, tada zbroj međupoloženosti uključuje doprinos svakog od njih.

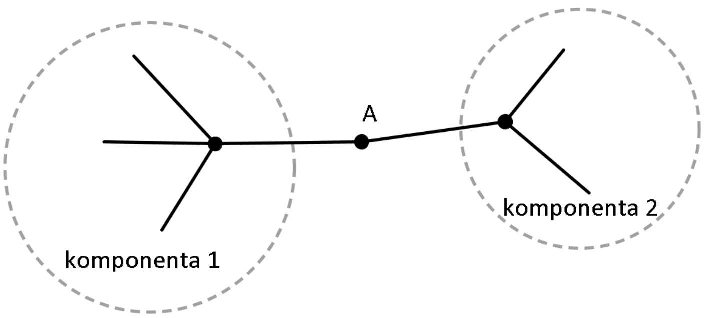
Formalno, međupoloženost možemo izraziti za općenitu mrežu tako da redefiniramo veličinu n_{st}^i tako da bude jednaka broju najkraćih putova od s do t koji prolaze kroz vrh i. Definirajmo g_{st} kao ukupan broj najkraćih putova od vrha s do vrha t. Tada je centralnost međupoloženosti vrha i definirana ovako:

$$BC(i) = \sum_{s,t} \frac{n_{st}^i}{g_{st}}, \tag{24}$$

pri čemu je usvojen dogovor da je $\frac{n_{st}^i}{g_{st}} = 0$ ako je $g_{st} = 0$. Ova definicija je ekvivalentna gore spomenutom eksperimentu prolaska poruka, u kojemu poruke prolaze između svih parova vrhova u mreži istom prosječnom brzinom putujući najkraćim putovima, a u slučaju da postoji nekoliko najkraćih putova između zadanog para vrhova oni slučajno odabiru jedan od njih. U tom slučaju je centralnost međupoloženosti vrha i proporcionalna prosječnoj brzini kojom sav promet prolazi kroz vrh i.

Kao što smo ranije spomenuli, centralnost međupoloženosti možemo primijeniti i na usmjerene mreže. U usmjerenim mrežama najkraći put između dva vrha općenito ovisi o smjeru u kojem putujemo, što znači da je najkraći put od vrha A do vrha B znatno različit od najkraćeg puta od vrha B do vrha A. Upravo to uvažava definicija (24), odnosno, posebno broji putove od vrha s do vrha t, a posebno putove od t do s. Ipak, treba napomenuti da se centralnost međupoloženosti rijetko ili gotovo nikad koristi u slučaju usmjerenih mreža.

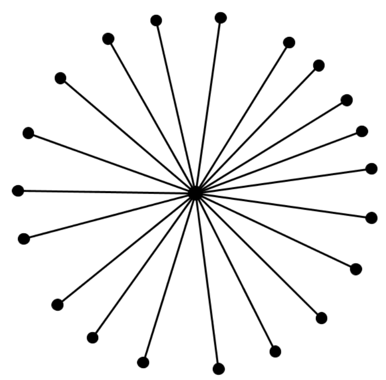
Centralnost međupoloženosti razlikuje se od ostalih mjera centralnosti u tome da ne mjeri koliko je vrh dobro povezan, nego mjeri koliko vrh sudjeluje na putovima između drugih vrhova. To znači da vrh može imati prilično nizak stupanj, biti povezan s vrhovima koji imaju jednak stupanj kao i on, čak i biti relativno daleko od drugih vrhova u mreži, a u istom trenutku imati visoku međupoloženost te biti proglašen za najcentralniji vrh u mreži. Promotrimo situaciju prikazanu na slici 26. Vrh A je stupnja dva i rezni je vrh u mreži. To znači da se njegovim uklanjanjem



Slika 26: Vrh velike međupoloženosti, a niskog stupnja.

mreža raspada na dvije komponente povezanosti. Kako bilo koji put (ne nužno najkraći) koji povezuje proizvoljan vrh jedne komponente s proizvoljnim vrhom druge komponente mora proći kroz vrh A, to je međupoloženost vrha A iznimno velika iako se A nalazi na rubovima dviju komponenti, a zbog izrazito malog stupnja nije dobro povezan s ostalim vrhovima mreže. Vrh A nema osobito impresivne vrijednosti svojstvene centralnosti ili centralnosti blizine, stupanjska centralnost mu je tek 2, ali ipak može imati velik utjecaj na mrežu kao rezultat njegove kontrole nad protokom informacija između drugih vrhova mreže. Vrhovi sa ovakvim ulogama često se u društvenim mrežama nazivaju *brokeri* ili *posrednici*.

Centralnost međupoloženosti ima još jedno zanimljivo svojstvo: njene vrijednosti su općenito velikog raspona (za razliku od npr. centralnosti blizine/daljine). Najveća moguća vrijednost za međupoloženost postiže se onda kada se vrh nalazi na najkraćem putu između svaka dva vrha u mreži. To se događa, na primjer, sa središnjim vrhom u *zvijezdi* (vidi sliku 27), mreži sastavljenoj od vrha koji je spojen bridom sa svim ostalim vrhovima mreže te nema drugih bridova u mreži. U ovak-



Slika 27: Mreža *zvijezda*.

voj mreži središnji vrh leži na svim n^2 najkraćim putovima između vrhova osim na $n-1$ putova koji su duljine nula, a počinju i završavaju u perifernom vrhu. Tada je centralnost međupoloženosti središnjeg vrha jednaka $n^2 - n + 1$. S druge strane, najmanja moguća vrijednost međupoloženosti vrha u mreži s jednom komponentom je $2n-1$, zato što se svaki vrh nalazi na svakom putu koji počinje ili završava samim sobom. To je zato što postoji $n-1$ putova od promatranog vrha do drugih vrhova, $n-1$ putova od drugih vrhova do tog promatranog vrha i jedan put od promatranog vrha do samog sebe, što je ukupno $2(n-1)+1 = 2n-1$. Ta se situacija događa kada mreža sadrži *list*, odnosno vrh stupnja jedan. Zaključujemo da je približna vrijednost omjera najvećih i najmanjih mogućih vrijednosti dana s

$$\frac{n^2 - n + 1}{2n - 1} \simeq \frac{1}{2}n,$$

pri čemu se jednakost postiže jedino za dovoljno velik n.

Ponekad je korisno normalizirati vrijednosti međupoloženosti. Nekoliko standardnih računalnih programa koji se koriste za analizu kompleksnih mreža, kao npr.

Pajek i UCINET, obavljaju takve normalizacije. Jedan od prirodnih izbora je normaliziranje dijeljenjem sa ukupnim brojem (uređenih) parova vrhova kojih je n^2, tako da centralnost međupoloženosti postaje

$$BC(i) = \frac{1}{n^2} \sum_{s,t} \frac{n^i_{st}}{g_{st}}.$$

Time se vrijednosti međupoloženosti vrhova smještaju strogo unutar intervala $[0,1]$.

Postoje još tri varijacije centralnosti međupoloženosti. Te varijacije uvjetovane su posebnim zahtjevima na mrežu kao što su komunikacija među vrhovima, protok informacija kroz mrežu te kretanje informacija po proizvoljnim šetnjama. Stoga razlikujemo:

- centralnost protočne međupoloženosti $BC_f(i)$,
- centralnost komunikacijske međupoloženosti $BC_C(i)$,
- centralnost međupoloženosti slučajnom šetnjom $BC_{RW}(i)$.

Ukratko ćemo objasniti svaku od njih i ilustrirati u kakvom su odnosu sa originalnom centralnosti međupoloženosti.

Centralnost protočne međupoloženosti predložili su White i Smith 1988. godine, a definirana je s:

$$BC_f(i) = \sum_{s,t} \frac{f(i;s,t)}{f(s,t)}, \tag{25}$$

gdje je $f(s,t)$ maksimalan protok od vrha s do vrha t, a $f(i;s,t)$ maksimalan protok između vrhova s i t koji prolazi vrhom i. Zamislimo da je svaki brid u mreži jedna cijev koja može prenositi jedinstveni protok neke tekućine ili fluida. Pitamo se koliki je najveći mogući protok između izvora - vrha s i destinacijskog vrha t koji prolazi kroz ove cijevi. Općeniti je odgovor da se više od jedne jedinice protoka može provesti od izvora do destinacije tako da se istovremeno može koristiti nekoliko različitih putova kroz mrežu. Uočimo da je protočna međupoloženost vrha i definirana jednadžbom (23), no u njoj veličina n^i_{st} predstavlja količinu protoka kroz vrh i kada se najveći protok prenosi iz vrha s u vrh t. Znamo da je najveći protok između vrhova s i t jednak broju bridno nezavisnih putova između ta dva vrha. To znači da bismo na protočnu međupoloženost mogli gledati tako da n^i_{st} promatramo kao broj nezavisnih putova između s i t koji prolaze vrhom i. Tu, međutim, može nastati mali problem jer nezavisni putovi između zadanog para vrhova u mreži općenito ne moraju biti jedinstveni. Kako bi se ovaj problem riješio, Freeman je sa suradnicima zamislio protok kroz vrh tako da on bude najveći mogući protok obzirom na sve moguće izbore putova (ili ekvivalentno, da bude najveći broj nezavisnih putova).

Sljedeća varijacija centralnosti međupoloženosti je *centralnost komunikacijske međupoloženosti*. Ovu mjeru centralnosti predložio je Estrada zajedno sa suradnicima 2009. godine, a definira se na sljedeći način:

$$BC_C(i) = \frac{1}{C} \sum_{s,t} \frac{G^i_{s,t}}{G_{s,t}}, \quad s \neq t,\, s \neq i,\, t \neq i, \tag{26}$$

gdje je $G^i_{s,t}$ odgovarajuća težinska suma u kojoj se promatraju one šetnje koje uključuju vrh i te $C = n^2 - 3n + 2$ normalizirajući faktor koji je jednak broju članova sume. Zbog normalizirajućeg faktora, $BC_C(i)$ uvijek proprima vrijednosti iz segmenta $[0, 1]$. Preostaje još vidjeti kako ćemo izračunati $G^i_{s,t}$.

Neka je $A + E(i)$ matrica susjedstva mreže $G(i) = (V, E')$ nastale uklanjanjem svih bridova povezanih s vrhom $i \in V$, ali vrh i nije uklonjen. Matrica $E(i)$ ima nenul elemente samo u i-tom retku i i-tom stupcu, a u tom retku i stupcu je -1 na onoj poziciji na kojoj matrica A ima element 1. Tada je

$$G^i_{s,t} = \left(e^A\right)_{st} - \left(e^{A+E(i)}\right)_{st},$$

pa uz

$$G_{s,t} = (e^A)_{st}$$

lako izračunamo traženu centralnost.

Još treba objasniti *centralnost međupoloženosti slučajnom šetnjom.*

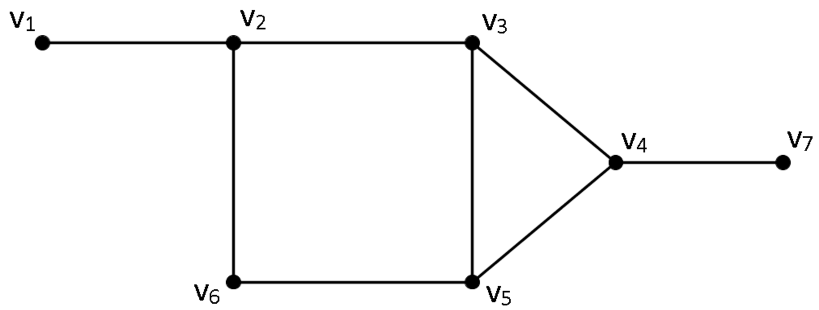
Definicija 22. ***Slučajna šetnja*** *duljine k na grafu G s korijenom O je stohastički proces sa slučajnim varijablama $X_1, X_2 \ldots, X_k$ tako da je $X_1 = O$, a X_{i+1} je slučajno odabran susjed vrha i.*

Pojednostavljeno rečeno, slučajna šetnja u proizvoljnom grafu je stohastički proces u kojem vrhove grafa obilazimo na slučajan način.

Ovu mjeru centralnosti uveo je Newman 2005. godine, a temelji se na izvođenju (apsorpcijske) slučajne šetnje koja počinje u vrhu s i nastavlja se sve dok ne završi u vrhu t. Ova je mjera centralnosti definirana s (23), ali je sada n^i_{st} broj proizvoljnih šetnji od s do t koji prolaze vrhom i. Imajmo na umu da vrijedi

$$n^i_{st} \neq n^i_{ts},$$

čak i u neusmjerenim mrežama. Kako se svaki slučajni put između dva vrha pojavljuje na slučajnoj šetnji između ta dva ista vrha s određenom vjerojatnošću, međupoloženost slučajnom šetnjom uključuje doprinose svih putova. Međutim, proizvoljni putovi se općenito pojavljuju s različitim vjerojatnostima, tako da putovi ne daju jednak doprinos međupoloženosti. Doprinos duljih putova je obično manji nego doprinos kraćih, a to je pristranost koja je prisutna u nekim, ali ne i u svim slučajevima.



Slika 28

Slika 28 prikazuje jednostavnu mrežu na kojoj ćemo pogledati rezultate prethodno navedenih verzija centralnosti međupoloženosti. Rezultati su dani tablicom 5. Kao što vidimo, originalna centralnost međupoloženosti identificira vrh v_2 kao najcentralniji, a vrhovima v_3 i v_4 dodjeljuje isti značaj u mreži. S druge strane, protočna međupoloženost i međupoloženost slučajne šetnje identificiraju vrh v_4 kao drugi najutjecajniji vrh u zadanoj mreži, odmah nakon vrha v_2. Komunikacijska međupoloženost identificira vrh v_3 kao najcentralniji vrh, što pokazuje da vrh v_3 igra središnju ulogu u distribuciji informacija između ostalih vrhova kada se sve vrste šetnji koriste kao komunikacijski kanali.

v_i	$BC(v_i)$	$BC_f(v_i)$	$BC_{RW}(v_i)$	$BC_C(v_i)$
1	0.0	0.000	0.286	0.882
2	5.5	13.667	0.619	13.535
3	5.0	6.000	0.576	15.162
4	5.0	10.667	0.584	13.385
5	2.5	6.000	0.545	11.909
6	1.0	3.667	0.450	5.801
7	0.0	0.000	0.286	0.821

Tablica 5: Varijacije centralnosti međupoloženosti vrhova mreže prikazane na slici 28.

3 Usporedba mjera centralnosti

U prethodnom poglavlju predstavili smo mjere centralnosti koje služe kao strategije za pronalazak najutjecajnijeg vrha u nekoj kompleksnoj mreži. Stupanj vrha, svojstvena centralnost, Katzova centralnost i PageRank centralnost imaju lijepo svojstvo da se mogu međusobno usporediti te da se lako može iz jedne mjere centralnosti prijeći u drugu. Stoga ćemo na samom kraju rada napraviti usporedbu tih mjera.

Pogledajmo u kakvom je odnosu PageRank centralnost sa ostale tri mjere centralnosti koje smo gore naveli. Ako zamislimo konačnu verziju PageRank-a koja uopće nema dodatnog konstantnog člana

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{d_j},$$

vidimo da je to slično izvornoj svojstvenoj centralnosti, ali sada imamo dodatno dijeljenje sa d_j. Međutim, za neusmjerene mreže ta je mjera trivijalna: lako se vidi da dobivamo $x_i = d_i$ te je stoga to isto kao i stupanj vrha. U slučaju usmjerenih mreža dobivamo nenul rezultate samo za vrhove koji pripadaju čvrsto povezanoj komponenti sa dva ili više vrhova, što je ekvivalentno problemu originalne svojstvene centralnosti. U tablici 6 je prikazan sažetak različitih matričnih mjera centralnosti o kojima je do sada bilo riječi, a organizirane su prema njihovim definicijama i svojstvima. Ako upotrijebimo jednu od tih mjera u nekoj proizvoljnoj mreži i dobijemo

	s konstantnim članom	bez konstantnog člana
dijeljenje sa izlaznim stupnjem	$x = D(D - \alpha A)^{-1} \cdot \mathbf{1}$ PageRank	$x = AD^{-1}x$ stupanj vrha
bez dijeljenja	$x = (I - \alpha A)^{-1} \cdot \mathbf{1}$ Katzova	$x = \lambda_1^{-1} A x$ svojstvena

Tablica 6: Mjere centralnosti i njihov odnos

zbunjujuće vrijednosti, velika je vjerojatnost da smo za mjeru centralnosti izabrali PageRank ili svojstvenu centralnost. To i nije prevelika greška, zato što se te dvije mjere centralnosti ubrajaju među najčešće korištene. Katzova centralnost je pronašla široku primjenu u prošlosti (prije otkrića drugih mjera centralnosti), ali se sve manje spominje u znanstvenim radovima. PageRank mjera centralnosti bez konstantnog člana je jednaka stupnju vrha u neusmjerenim mrežama i nije u zajedničkoj upotrebi za usmjerene mreže.

Literatura

[1] P.F. Bonacich, *Power and centrality: A family of measures*, American Journal of Sociology 92, 1170–1182 (1987).

[2] D. Cvetković, P. Rowlinson, S. Simić, *Eigenspaces of Graphs*, Cambridge University Press, Cambridge, 1997.

[3] R. Diestel, *Graph Theory*, Electronic Edition, 2000.

[4] N. Elezović, *Linearna algebra*, Element, Zagreb, 2006.

[5] E. Estrada, P.A. Knight, *A First Course in Network Theory*, Oxford University Press, 2015.

[6] E. Estrada, *The Structure of Complex Networks, Theory and Applications*, Oxford University Press, 2012.

[7] L.C. Freeman, *A set of measures of centrality based upon betweenness*, Sociometry 40, 35–41 (1977).

[8] L.C. Freeman, *Centrality in Social Networks - Conceptual Clarification*, Social Networks 1, 215–239 (1978/1979).

[9] G.H. Golub, C.F. Van Loan, *Matrix Computations*, Baltimore: Johns Hopkins University Press, p. 320, 1996.

[10] T. Harju, *Lecture Notes on Graph Theory*, Department of Mathematics, University of Turku, 2011.

[11] L. Katz, *A new status index derived from sociometric analysis*, Psychometrika 18, 39–43 (1953).

[12] S. Milgram, *The small world problem*, Psychology Today 1, 60–67 (1967).

[13] M.E.J. Newman, *Networks, An Introduction*, Oxford University Press, 2010.

[14] D. Veljan, *Kombinatorika s teorijom grafova*, Školska knjiga, Zagreb, 1989.

[15] G.U. Yule, *A mathematical theory of evolution, based on the conclusions of dr. J. C. Willis*, Philosophical Transactions of The Royal Society B. 213, 21–87 (1925).

[16] B. Zhou, N. Trinajstić, *On the largest eigenvalue of the distance matrix of a connected graph*, Chemical Physics Letters 447, 384—387 (2007).

Sažetak i ključne riječi

Sažetak. Prilikom proučavanja kompleksnih mreža često se javlja problem identifikacije "najvažnijih" vrhova. U tu svrhu definirali smo razne mjere centralnosti, a odabir one koja je najprikladnija ovisi ponajprije o tipu kompleksne mreže. Podaci o ključnim vrhovima u mreži mogu ubrzati diseminaciju informacija u informacijskim mrežama, zaustaviti širenje epidemije u društvenim mrežama, razbiti mrežu terorista u cilju sprečavanja mogućih terorističkih napada itd.

Stupanj vrha je mjera definirana kao broj bridova koji su incidentni s tim vrhom. Za svaki usmjereni graf ulazni stupanj broji veze usmjerene u vrh, a izlazni stupanj veze usmjerene iz vrha. Svojstvena centralnost se dodjeljuje vrhu u odnosu na važnosti svih vrhova u mreži. Katzova centralnost ima ideju dodavanja konstantnog člana u mjeru centralnosti i tako omogućava vrhu koji ima mnogo susjeda da ima veću vrijednost centralnosti bez obzira imaju li njegovi susjedi visoku centralnost ili ne. Osnovna ideja PageRank centralnosti jest da se važnost vrha oslanja na važnost onih vrhova koji pokazuju na taj vrh. Ideja podgrafovske centralnosti je da važnost promatranog vrha možemo ocijeniti s obzirom na njegovu pripadnost zatvorenim šetnjama koje počinju i završavaju u tom vrhu. Udaljenost vrhova kao mjera centralnosti definira prosječan broj koraka koji je potreban da od jednog vrha grafa dođemo do svih ostalih vrhova koji su mu u dosegu. Centralnost međupoloženosti govori koliko je vjerojatno da se vrh nalazi na putu između neka dva vrha, a koristi se i kao pokazatelj gdje bi se mreža mogla raspasti.

Ključne riječi: kompleksna mreža, centralnost, stupanj vrha, PageRank, međupoloženost

Centrality measures in complex networks

Summary. In the study of complex networks very often arises a problem of identifying the "most important" nodes. For this purpose, we have defined various centrality measures, and selecting the one that is best suited depends primarily on the type of complex network. Network key data about most important nodes

can accelerate dissemination of information in information networks, stop spreading social network epidemics, break a network of terrorists to prevent possible terrorist attacks, etc.

Degree centrality is a measure defined as the number of edges that are incident with the node. For each directed network the in-degree counts the links pointing to the node, and the out-degree counts the links pointing from the node. Eigenvector centrality is assigned to the node in relation to the results of all nodes in the network. Katz's centrality has the idea of adding a constant term to the measure of centrality and thus allowing a node with many neighbors to have greater value of centrality regardless of whether his neighbors have high centrality or not. The central idea of PageRank centrality is that the importance of the node relies on the importance of those nodes that point to that node. The idea of subgraph centrality is that the importance of the observed node can be evaluated in view of its belonging to the closed walks that begin and end in that node. Closeness centrality defines the average number of steps we need to reach one of the other nodes in the range from one network node. The betweenness centrality indicates how likely the node is located on the path between two nodes, and is also used as an indicator of where the network could collapse.

Keywords: complex network, centrality, vertex degree, PageRank, betweenness

Životopis

Rođen sam 24. listopada 1993. godine u Osijeku. Živim sa svojom obitelji u Beketincima, mjestu koje je 25 km udaljeno od Osijeka. U rodnom selu pohađao sam prva tri razreda osnovne škole u PŠ Milka Cepelića Beketinci, a preostalih pet razreda u OŠ Milka Cepelića u Vuki. 2008. godine sudjelovao sam na državnoj smotri *LiDraNo 2008.* u Dubrovniku u kategoriji dramskog izričaja.

Srednju školu pohađao sam u I. gimnaziji u Osijeku, a završio sam ju 2012. godine uspješnim polaganjem državne mature. Akademske 2012./2013. godine upisao sam prvu godinu Sveučilišnog nastavničkog studija matematike i informatike na Odjelu za matematiku Sveučilišta Josipa Jurja Strossmayera u Osijeku. Stručnu praksu obavljao sam u OŠ Retfala u Osijeku te u Ugostiteljsko-turističkoj školi u Osijeku.

U slobodno vrijeme vozim bicikl te povremeno rekreacijski igram stolni tenis. Pomažem oko poslova kod kuće, a najviše me opušta čitanje knjiga, posebno krimića te povijesnih i pustolovnih romana.