SREDIŠČNOST IN POMEMBNOST

Mere središčnosti in pomembnosti

Ena od zelo pogostih uporab analize omrežij je poiskati ’najbolj središčne’ enote v omrežju.

Mere središčnosti in pomembnosti lahko računamo

- za vsako enoto omrežja posebej ali
- za celotno omrežje. Mere za celotno omrežje se imenujejo mere **usredinjenosti omrežja**.

Mere središčnosti in pomembnosti posamezne enote

Mere središčnosti lahko računamo tako za neusmerjena kot tudi za usmerjena omrežja. Vse mere središčnosti, ki jih lahko izračunamo za neusmerjena omrežja lahko izračunamo tudi za usmerjena, obratno pa ni res. Pri usmerjenih omrežjih imamo še dodaten kriterij: lahko se omejimo na povezave, ki vstopajo v točko, ali pa na povezave, ki iz točke izstopajo.

Mere središčnosti in pomembnosti posamezne enote v omrežju imenujemo:

- **mere središčnosti** (tudi centralnosti), če analiziramo **neusmerjeno** omrežje;
  Primer: Neko mesto je *središčno*, če leži na križišču številnih poti.
  Te mere lahko posplošimo tudi na usmerjena omrežja.

- **mere pomembnosti**, če analiziramo **usmerjeno** omrežje.
  V tem primeru lahko računamo mere pomembnosti glede na to, ali je enota izhodišče povezav (**mere vplivnosti**), ali pa je enota konec povezav (**mere podpore**). Te mere torej ne moremo definirati za neusmerjena omrežja.

  Poimenovanje mer pomembnosti kot vplivnosti in podpore je odvisno tudi od vsebine obravnavane relacije.

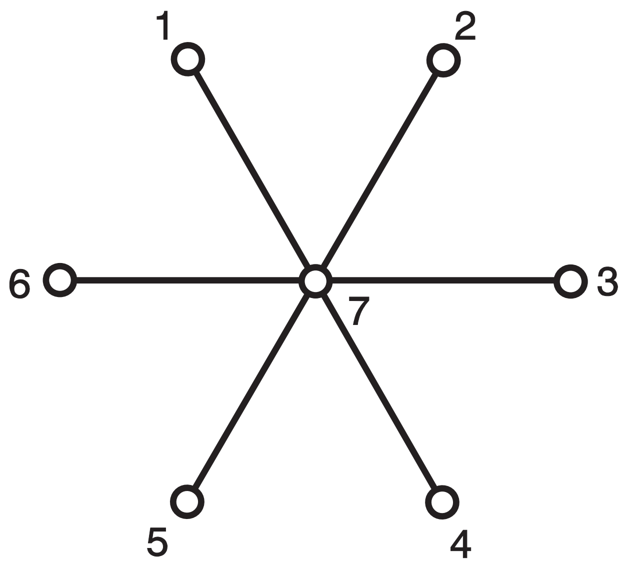
  Primera:
  Neka oseba je *vplivna*, če ukazuje veliko ostalim.
  Neka oseba ima veliko *podporo*, če zanjo glasuje veliko ljudi.

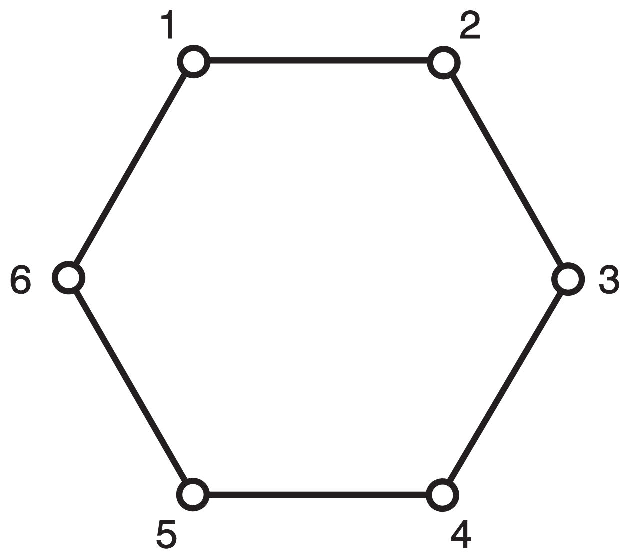
Mere središčnosti posamezne enote

Enota je tem bolj središčna

- čim večjo **stopnjo** ima,
- čim bolj je **dostopna** od vseh ostalih enot,
- se nahaja na največjem možnem številu najkrajših (geodezičnih) poti **vmes** med drugimi enotami.

Zvezda in cikel





Po vseh treh kriterijih je enota 7 v zvezdi najbolj središčna in ostale enote v njej enako središčne. Vse enote v ciklu so enako središčne po vseh treh kriterijih.

Mere središčnosti glede na stopnjo

Degree Centrality

Najbolj preprosta mera – enota je središčna v omrežju, če je ’dovolj aktivna’ v smislu, da ima veliko povezav do ostalih enot v omrežju (srednja točka v zvezdi). V primeru cikla pa so vse enote enako središčne.

Ta mera središčnosti je definirana s stopnjo enote x

$$c_D(x) = \text{stopnja enote } x$$

To je **absolutna mera središčnosti** glede na stopnjo. Absolutnih mer ne moremo uporabiti za primerjavo središčnosti po omrežjih z različnim številom točk, zato te mere ponavadi normaliziramo, tako da dobimo mero iz intervala med 0 in 1, kjer 0 pomeni, najmanjšo možno, vrednost 1 pa najvišjo možno središčnost. Te mere imenujemo **relativne mere središčnosti**.

Relativna mera središčnosti glede na stopnjo je

$$C_D(x) = \frac{c_D(x)}{\text{največja stopnja}} = \frac{c_D(x)}{n-1}$$

če je n število enot v omrežju, je največja možna stopnja enote v omrežju brez zank $n - 1$.

Omenjeno mero središčnosti lahko uporabimo tudi za mero pomembnosti za usmerjena omrežja, le da pri teh ločimo dve možnosti

- izbirati (*vplivnost* – izhodna stopnja: število puščic ven)
- biti izbran (*podpora* – vhodna stopnja: število puščic noter).

Mere središčnosti glede na dostopnost

Closeness Centrality

Za povezano omrežje je Sabidussi (1966) predlagal naslednjo mero središčnosti enote x glede na dostopnost:

$$c_C(x) = \frac{1}{\sum_{y \in E} d(x, y)}$$

kjer je $d(x, y)$ najkrajša razdalja med enotama x in y in E množica vseh enot.

Če omrežje ni krepko povezano, upoštevamo samo dosegljive točke, s tem da mero utežimo z močjo množice dosegljivih točk.

Po meri središčnosti glede na dostopnost so najbolj središčne tiste enote, ki so 'dovolj blizu' vsem ostalim – take enote lahko hitro komunicirajo z vsemi ostalimi. Ta mera središčnosti je boljša od središčnosti glede na stopnjo, ker ne upošteva samo neposrednih sosedov neke enote, ampak vse – tudi vse posredne sosede.

Relativna mera središčnosti glede na dostopnost je

$$C_C(x) = (n-1) \cdot c_C(x)$$

Razlaga: najmanjšo skupno oddaljenost neke enote od vseh drugih enot v omrežju dobimo, če ima enota vse druge enote za sosede. V tem primeru je absolutna mera središčnosti $\frac{1}{n-1}$.

Dostopnost lahko računamo za neusmerjena omrežja (mera središčnosti) in za usmerjena omrežja (mera pomembnosti).

Pri usmerjenih omrežjih pa imamo dve možnosti: dostopnost lahko izračunamo glede na

- izhodne povezave (kako blizu so vse ostale točke izbrani točki: v koliko korakih dosežemo iz dane točke vse ostale)
- vhodne povezave (kako blizu je izbrana točka vsem drugim: v koliko korakih iz vseh ostalih točk dosežemo izbrano).

Mere središčnosti glede na vmesnost

Betweenness Centrality

Pri komunikacijskih omrežjih ni pomembna le oddaljenost enote od vseh ostalih, ampak tudi katere enote ležijo na najkrajših poteh med pari enot – te enote imajo nadzor nad pretokom informacij med pari enot.

Ideja mer središčnosti glede na vmesnost: neke enota je središčna, če leži na veliko najkrajših poteh med drugimi pari enot.

Freeman (1977) je definiral mero središčnosti enote x glede na vmesnost za povezano omrežje takole

$$c_B(x) = \sum_{y<z} \frac{\text{število najkrajših poti med } y \text{ in } z \text{ skozi enoto } x}{\text{število vseh najkrajših poti med } y \text{ in } z}$$

Predpostavimo, da pri komunikacijskih omrežjih poteka komunikacija po najkrajših možnih poteh:

Središčnost enote x glede na vmesnost je vsota verjetnosti preko vseh možnih parov točk, da bo najkrajša pot med y in z potekala skozi točko x.

Relativno mero središčnosti glede na vmesnost je potrebno definirati posebej za

- neusmerjena omrežja:

$$C_B(x) = \frac{c_B(x)}{(n-1)(n-2)/2}$$

- usmerjena omrežja:

$$C_B(x) = \frac{c_B(x)}{(n-1)(n-2)}$$

Razlaga: če ima omrežje n točk, je za izbrano točko x število drugih različnih parov točk y in z

- $C^2_{n-1} = \binom{n-1}{2} = \frac{(n-1)(n-2)}{2}$
  v primeru, da vrstni red izbire ni pomemben (neusmerjeno omrežje)

- $V^2_{n-1} = (n-1)(n-2)$
  v primeru, da je vrstni red izbire pomemben (usmerjeno omrežje)

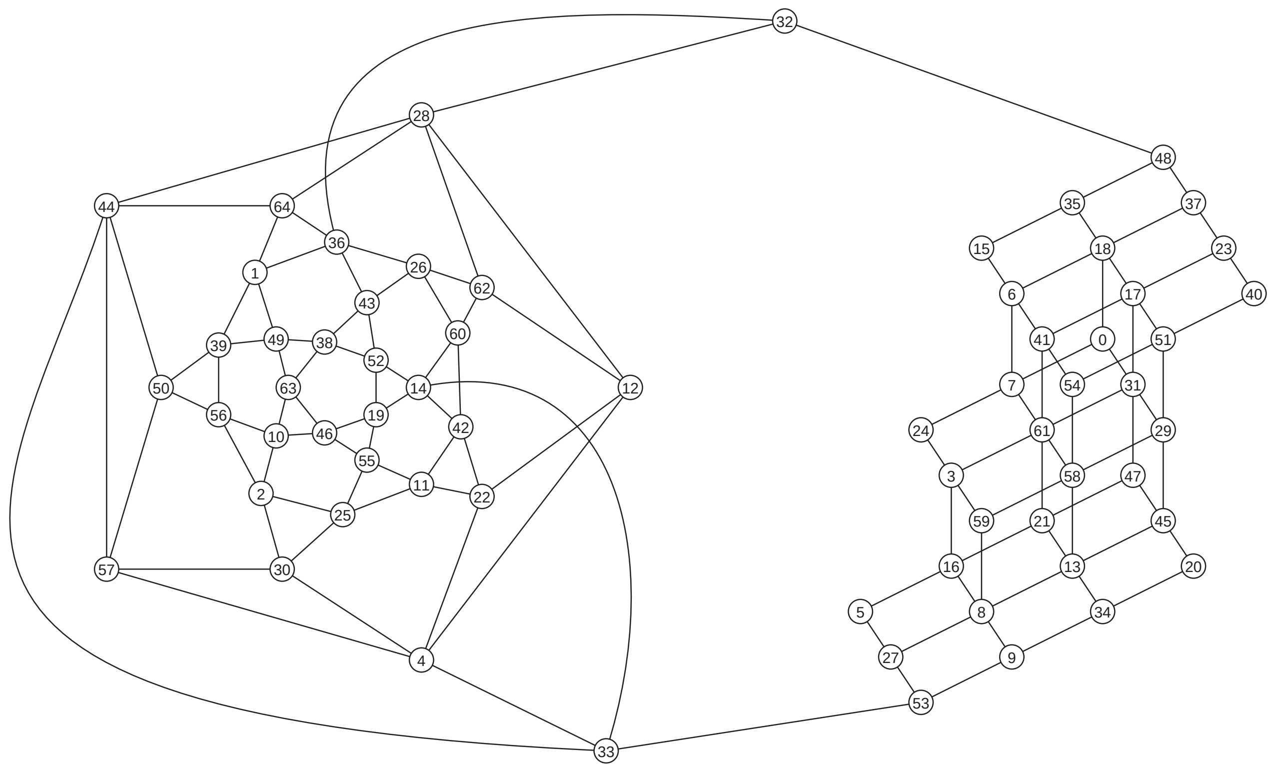
Obstajajo še številne druge mere središčnosti enote.

Izbira ustrezne mere središčnosti

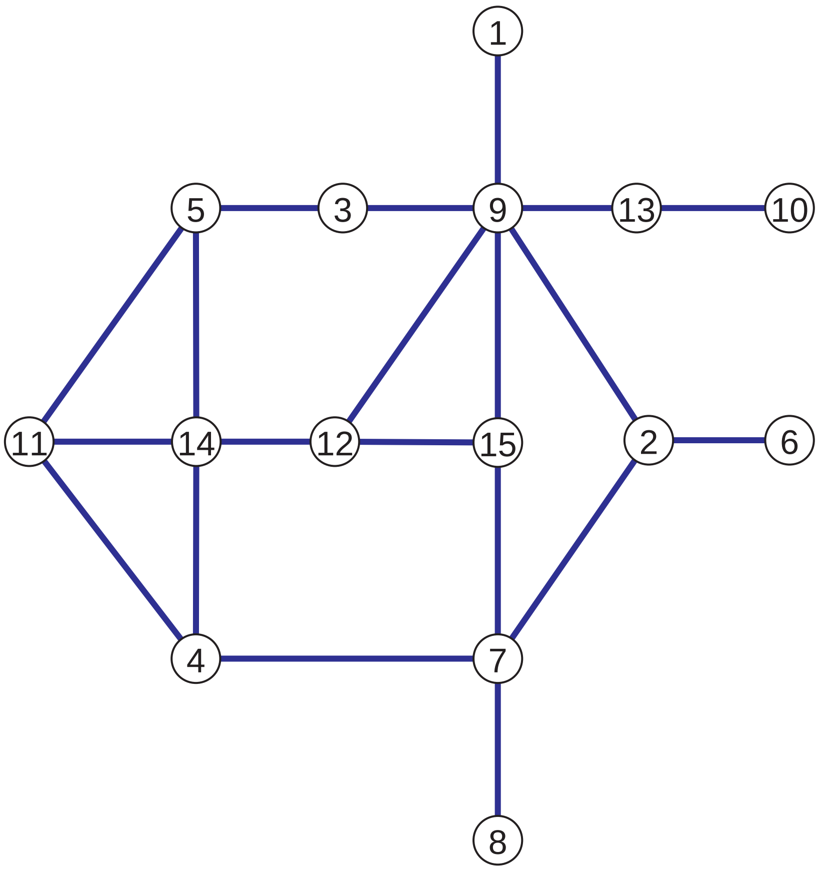
Omenjene mere središčnosti lahko v določenih primerih vrnejo zelo različne rezultate. Zato moramo biti pri izbiri ustrezne mere središčnosti zelo previdni:

Lahko se zgodi, da imajo enote sorazmerno nizke stopnje, a imajo visoko središčnost glede na vmesnost.

Npr. v omrežju na sliki po vmesnosti zelo izstopajo točke, 33, 53, 32 in 48, čeprav obstaja precej točk z višjimi stopnjami od naštetih (npr, točka 61 ima stopnjo 6, obstaja precej točk s stopnjami 5). Omenjene štiri točke pa imajo stopnje samo 3 ali 4.



Primer: Zakonske vezi med florentinskimi rodbinami



1.	Acciaiuoli	9.	Medici
2.	Albizzi	10.	Pazzi
3.	Barbadori	11.	Peruzzi
4.	Bischeri	12.	Ridolfi
5.	Castellani	13.	Salviati
6.	Ginori	14.	Strozzi
7.	Guadagni	15.	Tornabuoni
8.	Lamberteschi		

Relativne mere središčnosti florentinskih rodbin

Št.	Družina	C_D	C_C	C_B
1.	Acciaiuoli	0.071	0.368	0.000
2.	Albizzi	0.214	0.483	0.212
3.	Barbadori	0.143	0.438	0.093
4.	Bischeri	0.214	0.400	0.104
5.	Castellani	0.214	0.389	0.055
6.	Ginori	0.071	0.333	0.000
7.	Guadagni	0.286	0.467	0.255
8.	Lamberteschi	0.071	0.326	0.000
9.	Medici	*0.429*	*0.560*	*0.522*
10.	Pazzi	0.071	0.286	0.000
11.	Peruzzi	0.214	0.368	0.022
12.	Ridolfi	0.214	0.500	0.114
13.	Salviati	0.143	0.389	0.143
14.	Strozzi	0.286	0.438	0.103
15.	Tornabuoni	0.214	0.483	0.092

Primer
Sampsonovi podatki – naklonjenost

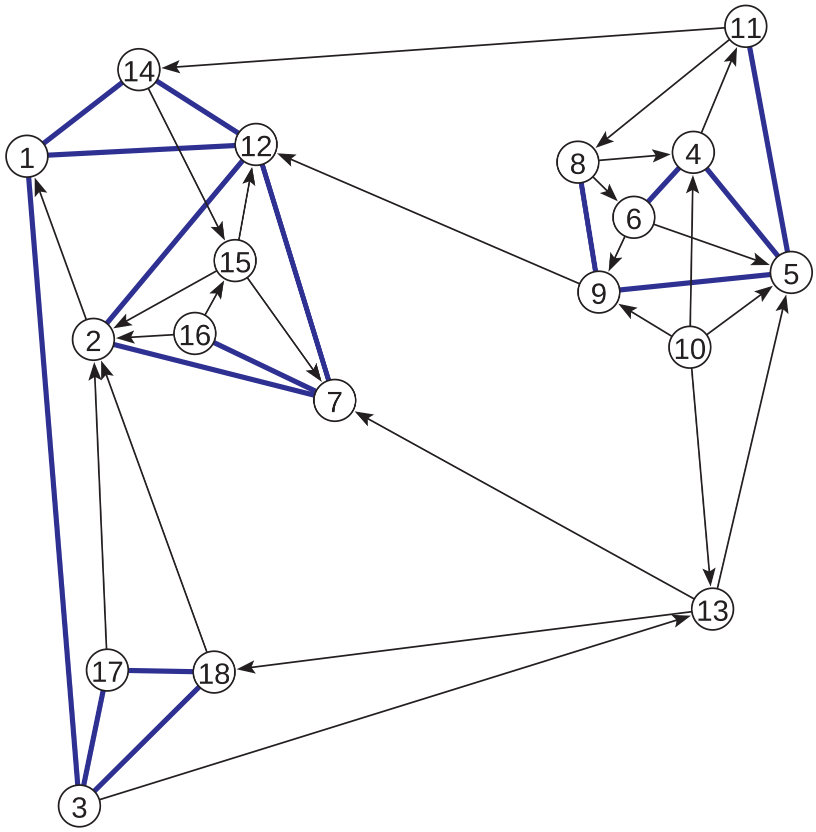
Sampson je preučeval odnose med 18 menihi v samostanu New England. Izmeril je več relacij med njimi

- prijateljstvo, naklonjenost (affect)
- spoštovanje (esteem)
- vplivnost (influence)
- odobravanje (sanction)

Za relacijo prijateljstva imamo podatke za tri časovne točke T_2, T_3 in T_4, za ostale tri relacije pa samo v časovni točki T_4.

Sampson, S (1969): *Crysis in a cloister.* Unpublished doctoral dissertation. Cornell University.

Omrežje prijateljstva v četrti časovni točki



1.	John Bosco	10.	Romuald
2.	Gregory	11.	Louis
3.	Basil	12.	Winfrid
4.	Peter	13.	Amand
5.	Bonaventura	14.	Hugh
6.	Berthold	15.	Boniface
7.	Mark	16.	Albert
8.	Victor	17.	Elias
9.	Ambrose	18.	Simplicius

Mere središčnosti za Sampsonove menihe (vhodne)

Št.	Menih	C_D	C_B	C_C
1.	John Bosco	0.235	0.332	0.472
2.	Gregory	0.353	0.122	0.486
3.	Basil	0.176	0.358	0.378
4.	Peter	0.235	0.065	0.293
5.	Bonaventura	0.353	0.227	0.370
6.	Berthold	0.118	0.007	0.239
7.	Mark	0.294	0.079	0.486
8.	Victor	0.118	0.013	0.239
9.	Ambrose	0.235	0.146	0.293
10.	Romuald	0.000	0.000	0.000
11.	Louis	0.118	0.089	0.283
12.	Winfrid	0.353	0.220	0.586
13.	Amand	0.118	0.250	0.298
14.	Hugh	0.177	0.111	0.447
15.	Boniface	0.118	0.016	0.354
16.	Albert	0.059	0.011	0.333
17.	Elias	0.118	0.002	0.293
18.	Simplicius	0.176	0.021	0.304

Usredinjenost omrežja

Izračunane središčnosti ali pomembnosti posameznih enot v omrežju variirajo. Omrežje, ki ima enoto z izstopajočo središčnostjo ali pomemnostjo glede na druge enote omrežja, je bolj usredinjeno (primer je zvezda). Če so izračunane središčnosti ali pomembnosti enot v omrežju zelo izenačene (primer je cikel), je usredinjenost omrežja zelo majhna.

Freeman (1979) je definiral splošno mero usredinjenosti takole

$$C_A = \frac{\sum_{x \in E}(C_A^\star - C_A(x))}{\max \sum_{x \in E}(C_A^\star - C_A(x))}$$

kjer je $C_A^\star$ največja vrednost izbrane mere središčnosti ali pomembnosti $C_A(x)$ v množici enot omrežja. Mera usredinjenosti zavzema vrednosti od 0 do 1: vrednost 0 ima, ko so vse vrednosti mere posameznih enot enake (cikel) in vrednost 1, ko ena enota popolnoma dominira ostale enote (zvezda).

Mere usredinjenosti omrežja

Mera usredinjenosti omrežja glede na stopnjo

$$C_D = \frac{\sum_{x \in E}(C_D^\star - C_D(x))}{n-2}$$

Mera usredinjenosti omrežja glede na dostopnost

$$C_C = \frac{\sum_{x \in E}(C_C^\star - C_C(x))}{(n-1)(n-2)/(2n-3)}$$

Mera usredinjenosti omrežja glede na vmesnost

$$C_B = \frac{\sum_{x \in E}(C_B^\star - C_B(x))}{n-1}$$

Pajek vrne mero usredinjenosti omrežja glede na dostopnost samo za krepo povezana omrežja.

Primer: Florentinske rodbine

Mere usredinjenosti omrežja florentinskih rodbin glede na njihove zakonske povezave so:

$$C_D = 0.275$$

$$C_C = 0.322$$

$$C_B = 0.437$$

Mere središčnosti v Pajku

Pajek vedno računa relativne mere središčnosti. Izjema so stopnje, kjer vrne absolutno mero. Mere središčnosti se nahajajo v izbiri Network/Create Vector/Centrality. Pri računanju središčnosti glede na stopnjo in dostopnost se moramo odločiti še med Input, Output ali All. Pri neusmerjenih omrežjih izberemo Input ali Output (obe izbiri vrneta enak rezultat).

Pri vseh treh merah dobimo v posebnem oknu izpisano tudi ustrezno usredinjenost celotnega omrežja.

Izpis najbolj središčnih točk dobimo z Vector/Info – izpiše se toliko najbolj središčnih točk, kot zahtevamo.

Izračunano mero središčnosti vsake točke (realno število med 0 in 1) lahko pri risanju prikažemo z velikostjo točke, če zahtevamo risanje Draw/Network + First Vector. Lahko pa izberemo tudi dve lastnosti (npr. vhodno in izhodno stopnjo) in uporabimo Draw/Network + First Vector + Second Vector. V tem primeru predstavlja velikost točke v smeri x prvo lastnost (vhodno stopnjo), velikost v smeri y pa drugo lastnost (izhodno stopnjo).

Vaje

1. V neusmerjenem omrežju flor.net (florentinske rodbine) izračunajte vse mere središčnosti. Na sliki omrežja predstavite središčnosti posameznih družin z ustreznimi velikostmi točk.
2. V usmerjenem omrežju sampson.net (naklonjenost med Sampsonovimi menihi) izračunajte vse mere središčnosti.
3. V neusmerjenem omrežju write.net izračunajte vse mere središčnosti. Glede na katero mero središčnosti so si točke najbolj različne med sabo (standardni odklon)?
4. Katera so najpomembnejša letališča v omrežju letalskih povezav v ZDA (usair97.net) glede na stopnjo, dostopnost in vmesnost?
5. V datoteki sk_trade.net so podani podatki o uvozu in izvozu med 118 državami, v datoteki sk_trade.clu pa je podano razbitje držav po kontinentih. Katere države so najpomembnejše glede na stopnjo, dostopnost in vmesnost?

Kazala in opisi

Kazala in opise izračunamo v izbiri
Network/Create Vector/Centrality/Hubs-Authorities
Dobimo še vprašanje koliko kazal in koliko opisov naj se označi v razbitju. Rezultati operacije so trije:

- Razbitje, kjer vrednost 1 (rumena) pomeni, da je točko dober opis, vrednost 2 (zelena) pomeni, da je točka dober opis in dobro kazalo, vrednost 3 (rdeča) pa pomeni, da je točka dobro kazalo.
- Vektor z utežmi za kazala (*Hubs Weights*), večja vrednost pomeni boljše kazalo.
- Vektor z utežmi za opise (*Authority Weights*), večja vrednost pomeni boljši opis.

Pozor: algoritem predpostavlja, da vrednosti na povezavah predstavljajo podobnosti (večja vrednost pomeni bolj pomembno izbiro). V primeru, da je ravno obratno, moramo vse vrednosti postaviti na 1, sicer bodo rezultati napačni. To lahko storimo z ukazom: Network/Create New Network/Transform/.. ../Line Values/Set All Line Values to 1.

Primerjava mer pomembnosti

Obravnavali smo naslednje mere pomembnosti:

- vhodna stopnja (*input degree*);
- vhodna dostopnost (*input closeness*);
- vmesnost (*betweenness*);
- opisi (*authorities*).

Nekatere od teh količin merijo pomembnost na zelo podoben način, nekatere pa so si precej različne. Za ugotavljanje podobnosti rezultatov lahko uporabimo *Pearsonov korelacijski koeficient*, ki ga v Pajku najdemo v Vectors/Info, še prej pa moramo izbrati oba vektorja, ki ju želimo primerjati. Postopek lahko pohitrimo, če vektorje izvozimo v R ali SPSS in korelacijo izračunamo tam.

Na splošno ugotovimo, da so si mere zelo podobne, razen vmesnosti, ki meri pomembnost (oziroma bolj središčnost) na precej drugačen način kot ostale mere.

Vaje

1. V omrežju football.net (Pariz 1998) in/ali football2002.net (Koreja/Japonska 2002) izračunajte vhodne in izhodne stopnje (brez upoštevanja vrednosi in z upoštevanjem vrednosti) ter kazala in opise.