Analiza omrežij

Andrej Mrvar (FDV)

Doktorski študij

Omrežje

’Običajna’ analiza podatkov (statistika): proučujemo lastnosti enot (dohodek, starost,...).

Analiza omrežij: proučujemo relacije med enotami – poleg enot obravnavamo še relacije (povezave).

Omrežje, v katerem nastopa relacija R, lahko predstavimo na več načinov:

- Predstavitev s pripadajočo **dvojiško matriko** $\mathbf{R} = [r_{ij}]_{n\times n}$, kjer je

$$r_{ij} = \begin{cases} 1 & X_i R X_j \\ 0 & \text{sicer} \end{cases}$$

  Včasih je lahko r_{ij} tudi realno število, ki izraža moč relacije R med enotama X_i in X_j.

- **seznam sosedov**

  Omrežje lahko opišemo tako, da za vsako enoto navedemo seznam vseh enot, s katerimi je enota v relaciji. Kot bomo videli kasneje, je taka predsatvitev veliko krajša. ker so velika omrežja tipično redka in je večina vrednosti v matriki 0.

- opis z **grafom** $G = (V, L)$ kjer je V množica točk (*Vertices*) in L množica (usmerjenih ali neusmerjenih) povezav (*Lines*). Točke grafa predstavljajo enote v omrežju, povezave pa relacijo.

  Graf predstavimo s sliko, kjer točke grafa predstavimo s krožci, usmerjene povezave s puščicami, neusmerjene povezave pa z daljicami, ki povezujejo ustrezne točke.

  $X_i R X_j \Rightarrow$ v grafu obstaja usmerjena povezava, ki vodi od točke X_i do točke X_j. Točka X_i se imenuje *začetna*, točka X_j pa *končna* točka povezave. Povezavo, kjer je začetna točka enaka končni, imenujemo *zanka*.

  Število točk v grafu označimo z n, število usmerjenih povezav pa z m.

	1	2	3	4	5
1	0	1	0	1	1
2	0	0	1	0	0
3	1	0	0	1	0
4	0	0	0	0	1
5	1	0	0	0	0

Arcslist

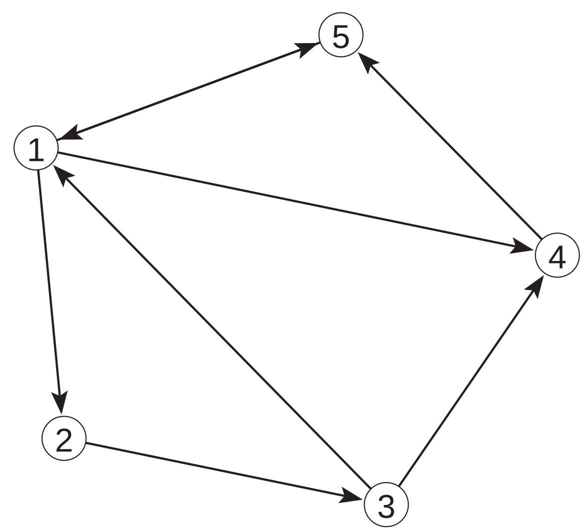
1: 2 4 5

2: 3

3: 1 4

4: 5

5: 1



Usmerjena in neusmerjena omrežja

- *neusmerjeno omrežje* – relacija je simetrična, oziroma povezave v pripadajočem grafu so neusmerjene (*edges*). Neusmerjene povezave rišemo brez puščic. Primer neusmerjene relacije: *poroka*.
- *usmerjeno omrežje* – relacija ni simetrična, oziroma vse povezave v pripadajočem grafu so usmerjene (*arcs*). Primer usmerjene relacije: *je otrok od*.

Mala in velika omrežja

Omrežja z nekaj 100 enotami in povezavami bomo imenovali *mala omrežja*, omrežja z nekaj 10000 enotami in povezavami pa *velika omrežja*.

Redka in gosta omrežja

Omrežje je *redko*, če je v pripadajočem grafu število povezav istega velikostnega reda kot število točk ($n \approx km$). Če je omrežje redko, lahko z nekaterimi algoritmi hitro analiziramo tudi zelo velika omrežja. V praksi pogosto naletimo na zelo velika, a redka omrežja.

V splošnem je število povezav lahko precej večje kot število točk. Tako omrežje imenujemo *gosto*.

Če je vsaka enota omrežja povezana z vsako drugo enoto, je število povezav n^2 (število elementov v ustrezni matriki).

Če pa je vsaka enota omrežja povezana z vsako drugo enoto, razen sama s sabo (graf brez zank), je število povezav $n(n-1)$ (število elementov v matriki brez diagonale).

Na tej osnovi lahko definiramo gostoto omrežja:

Za omrežja z zankami:

$$Density1 = \frac{m}{n^2}$$

Za omrežja brez zank:

$$Density2 = \frac{m}{n(n-1)}$$

Če predpostavimo, da med dvema točkama ne more biti večkratnih povezav, (točki sta torej lahko nepovezani ali povezani z eno samo povezavo), je gostota število med 0 in 1.

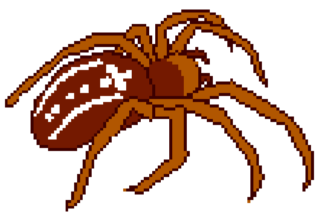
Gostota omrežja je ena od mer, s katerimi lahko primerjamo različna omrežja med sabo. Vendar se izkaže, da večja kot so omrežja, bolj redka so.

Velika omrežja so zelo, zelo redka, zato se kot boljša mera ponavadi uporablja povprečna stopnja (povprečno število sosedov).

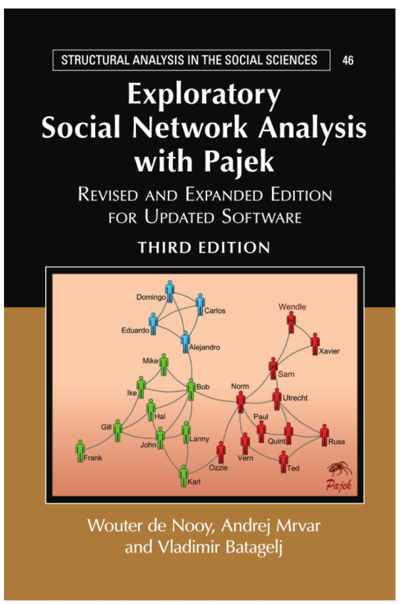
Primeri velikih omrežij

- socialna omrežja
  - rodovniki in druga sorodstvena omrežja (flor.net);
  - omrežja razširjanja inovacij ali bolezni (*diffusion networks*, HIV);
  - omrežje telefonskih klicev znotraj izbrane množice številk (raziskave kriminala, primer b.net);
  - članstva v nadzornih svetih, omrežja lastništev delnic;
  - trgovanje med organizacijami, državami (import.net + cont.clu);
  - uvoz in izvoz nogometašev (football.net, co.htm, footpart.htm),
  - omrežja citiranj, omrežja soavtorstev (*Web of Science, Cobiss*);
  - računalniška omrežja (lokalna omrežja, Internet, povezave med predstavitvenimi stranmi);
- organske molekule v kemiji (dna.net, *protein interaction networks*);
- omrežja pridobljena iz slovarjev in drugih besedil (write.net);
- transportna omrežja (letalska USAir.net+USAir.clu, vodovodna omrežja. . .).

Pajek



`Pajek` je programski paket za Windows 32 in 64, ki omogoča analizo *velikih omrežij*. Program je prosto dostopen na naslovu: http://mrvar.fdv.uni-lj.si/pajek/

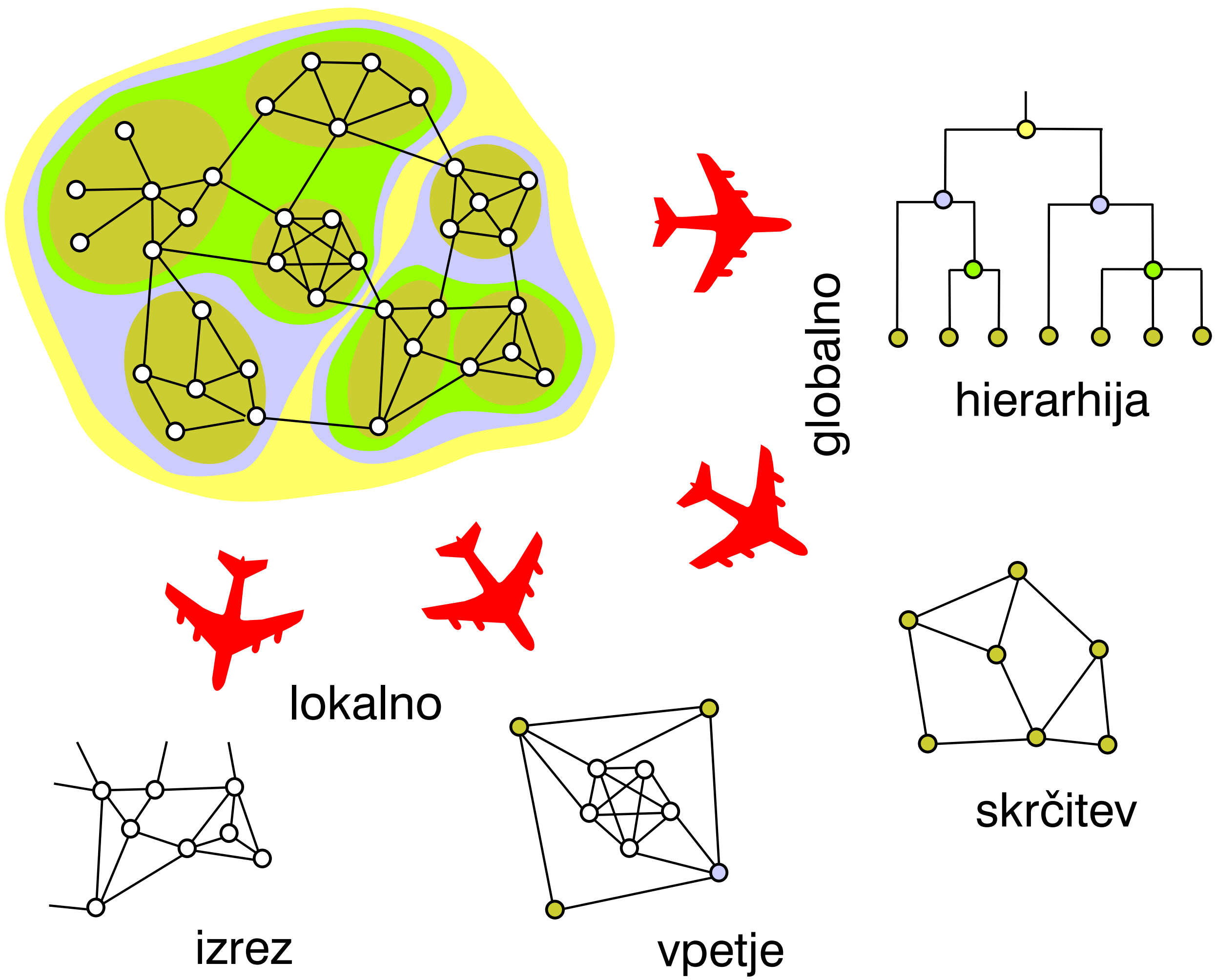


de Nooy, Mrvar, Batagelj (2018): *Exploratory Social Network Analysis with Pajek Revised and Expanded Edition for Updated Software.* Cambridge University Press, New York.

Cilji pri zasnovi programa

Glavni cilji pri zasnovi programa **`Pajek`** so:

- podpreti *abstrakcijo* z (rekurzivno) razčlenitvijo velikega omrežja na več manjših omrežij, ki jih lahko nadalje analiziramo z uporabo običajnih metod;
- ponuditi uporabniku močna orodja za *vizualizacijo* omrežij;
- vgraditi večje število *učinkovitih* algoritmov za analizo obsežnih omrežij.


globalno
hierarhija
lokalno
skrčitev
izrez
vpetje

Dvovrstna omrežja

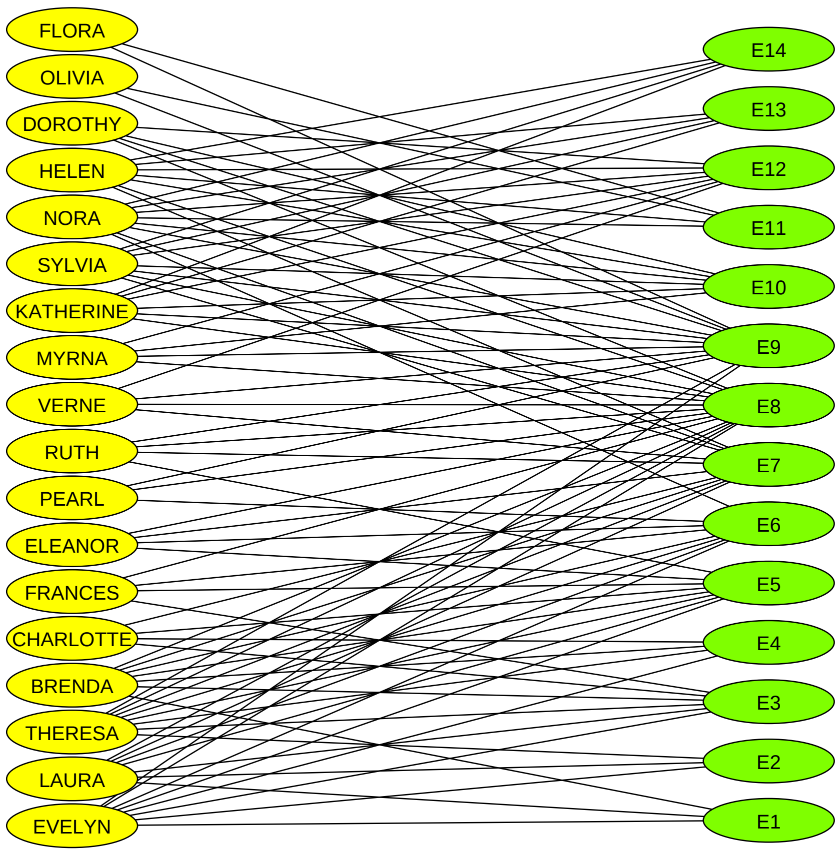
Običajna omrežja: množica enot in povezave med njimi.

Dvovrstno omrežje pa sestavljata dve množici enot (npr. osebe in dogodki), relacija pa ti dve množici povezuje, npr. vključenost oseb v družabne dogodke. Primerov takih omrežij je še ogromno:

- *Omrežje citiranj*, katerega prva množica točk so avtorji, druga množica točk so članki, povezave med avtorji in članki pa povedo, kateri avtor je citiral kateri članek.
- *Omrežje nakupov*, v katerem so prva množica kupci, druga množica artikli, povezava pa pove, kateri artikel je kupec kupil.
- *Bralci in revije, ki jih osebe berejo* – primer 124 slovenskih revij (slike SVG)

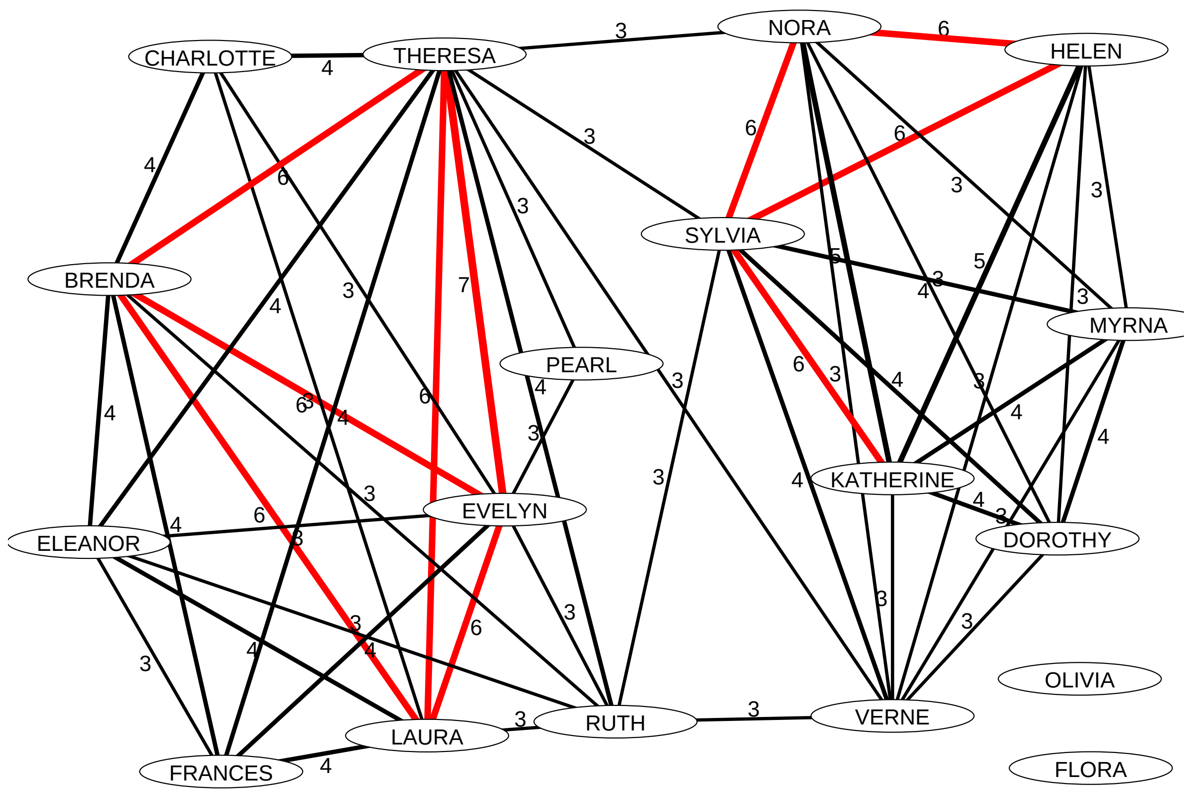
Tu je analiza omrežij najbližje 'rudarjenju' (data mining).

Primer dvovrstnega omrežja



Davis.net

Dvovrstno omrežje predelano v običajno omrežje



Vrednost na povezavi predstavlja število skupnih dogodkov, pri katerih sta bili prisotni obe osebi.

Primer: Slovenski dnevniki in revije

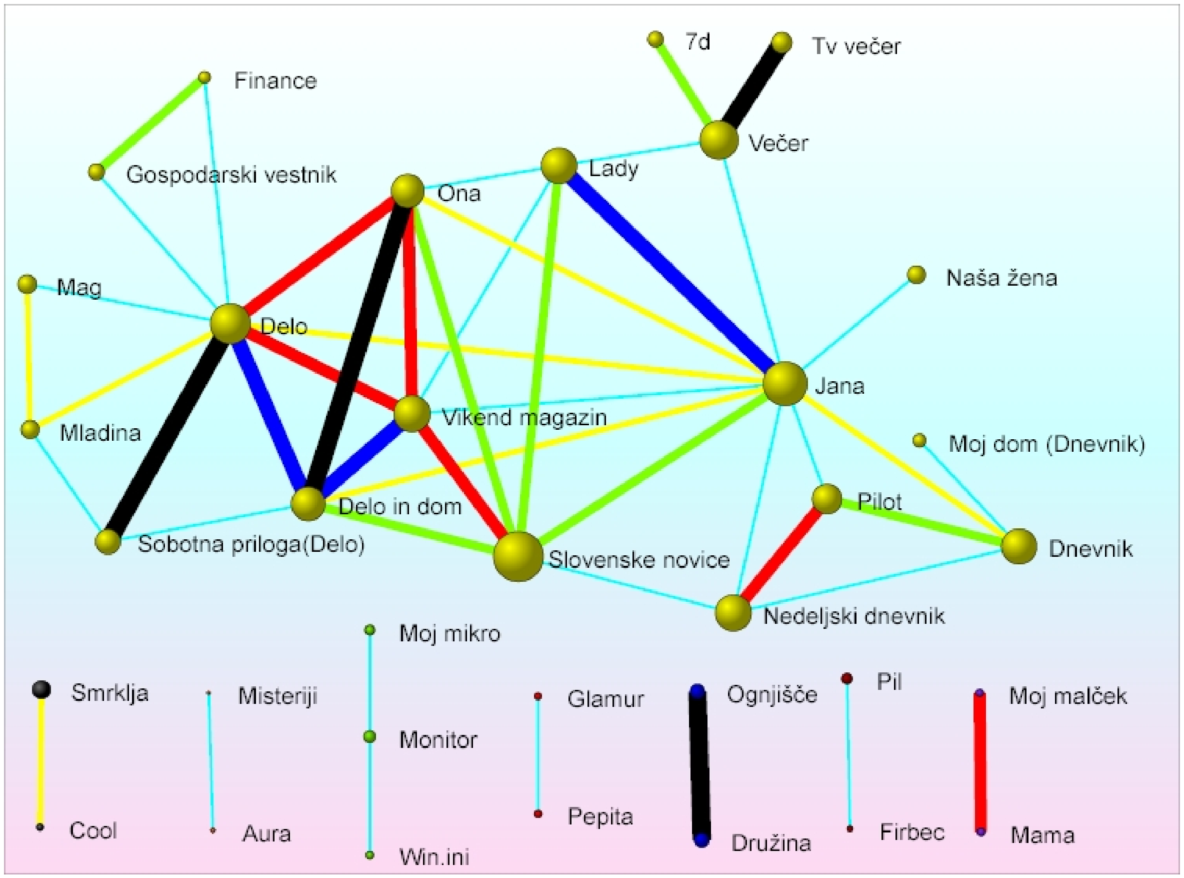
V letih 1999 in 2000 je v anketah Cati Centra več kot 100000 anketirancev odgovarjalo na vprašanje, katere dnevnike in revije berejo. Našteli so skupaj 124 različnih revij. Dobljene podatke lahko predstavimo kot dvovrstno omrežje:

	Delo	Dnevnik	Sl.novice	...
Anketiranec1	X		X	...
Anketiranec2		X		...
Anketiranec3	X			...
....	.....	....	....	...

Iz dvovrstnega omrežja bralec / revija smo zgenerirali običajno omrežje, kjer so enote revije. Neusmerjena povezava z vrednostjo a med revijama pomeni, da obe reviji hkrati bere a anketirancev. Zanka pri posamezni reviji pa pomeni skupno število anketirancev, ki berejo to revijo.

Dobimo matriko A:

	Delo	Dnevnik	Sl.novice	...
Delo	20714	3219	4214	...
Dnevnik		15992	3642	...
Sl. novice			31997	...
....	......	.....	.....	...

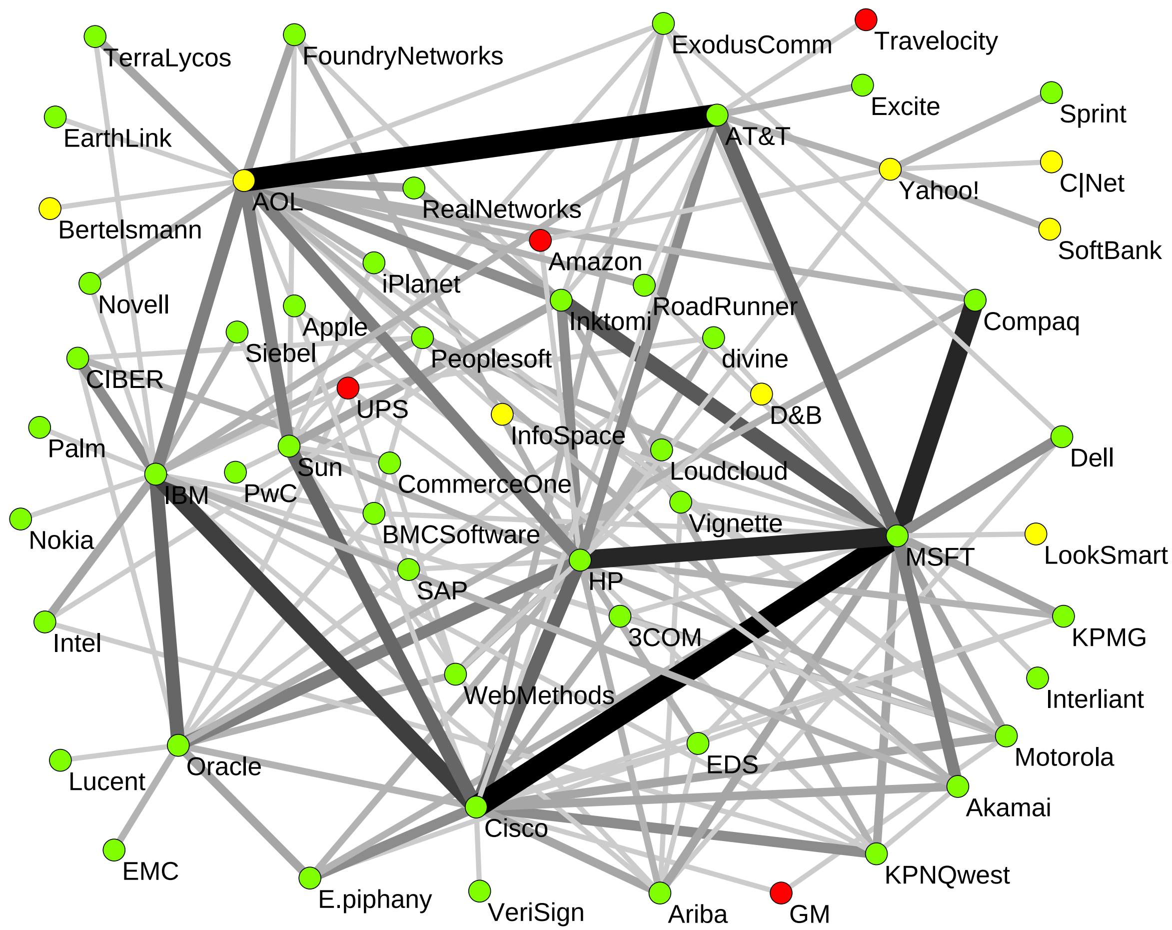

7d
Tv večer
Finance
Večer
Gospodarski vestnik
Lady
Ona
Naša žena
Mag
Delo
Jana
Vikend magazin
Mladina
Moj dom (Dnevnik)
Pilot
Delo in dom
Sobotna priloga(Delo)
Dnevnik
Slovenske novice
Nedeljski dnevnik
Moj mikro
Smrklja
Misteriji
Glamur
Ognjišče
Pil
Moj malček
Monitor
Cool
Aura
Pepita
Družina
Firbec
Mama
Win.ini

Revije3.htm.

Primer: Internet Industry Partnerships

- Vsaka točka v omrežju predstavlja neko podjetje, ki se ukvarja z internetom. Podjetji sta povezani, če sta objavili skupna vlaganja, strateški plan ali drugo vrsto partnerstva.
- Omrežje predstavlja samo podmnožico celotne industrije na področju interneta v obdobju od 1998 do 2001. Vsebuje 219 točk (podjetij) in 631 povezav.
- Podjetja so razdeljena v tri skupine glede na njihovo vrsto: rumena = vsebina, rdeča = trgovina, zelena = infrastruktura.
- http://www.orgnet.com/netindustry.html

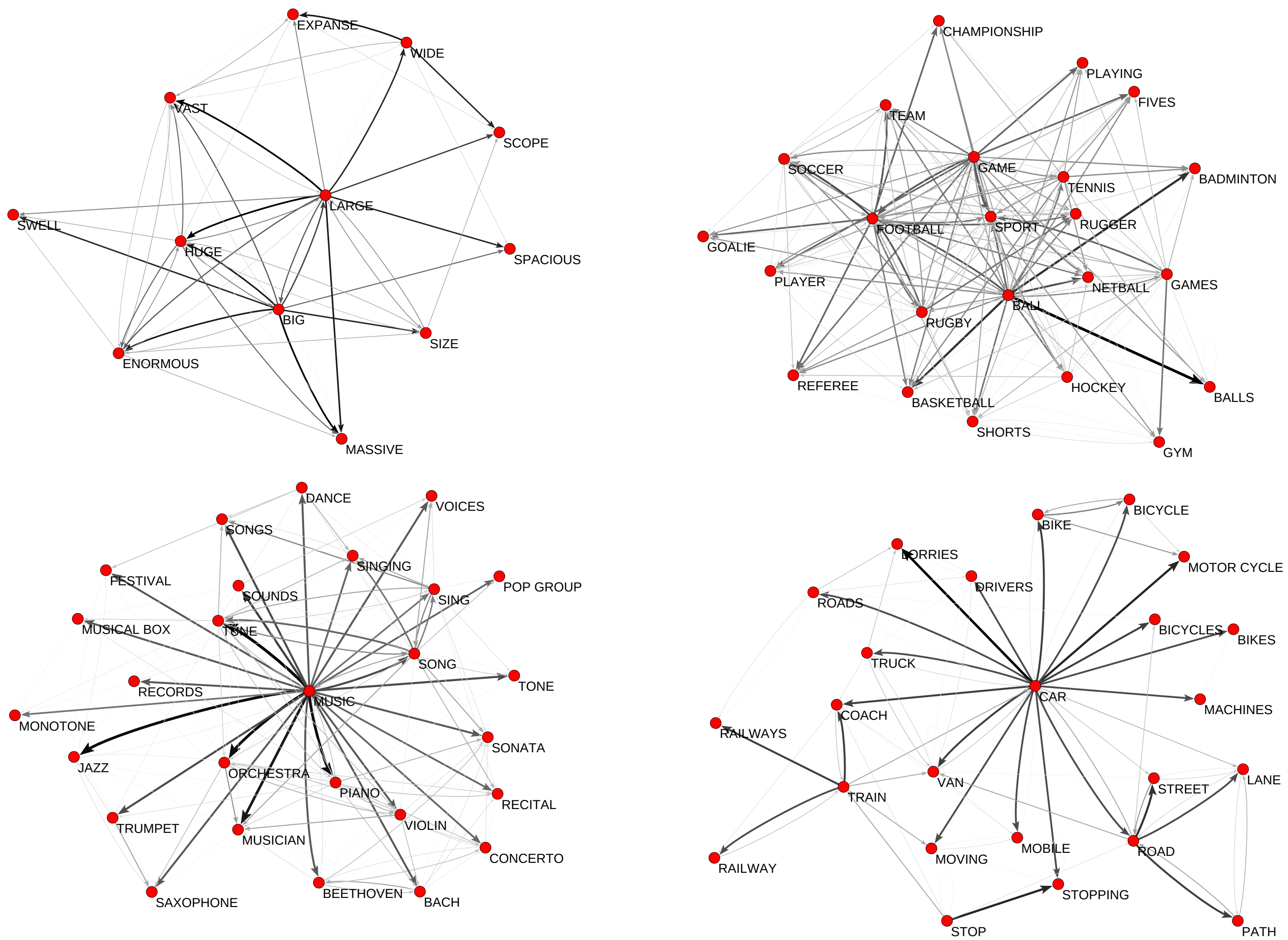
Najpomembnejše povezave



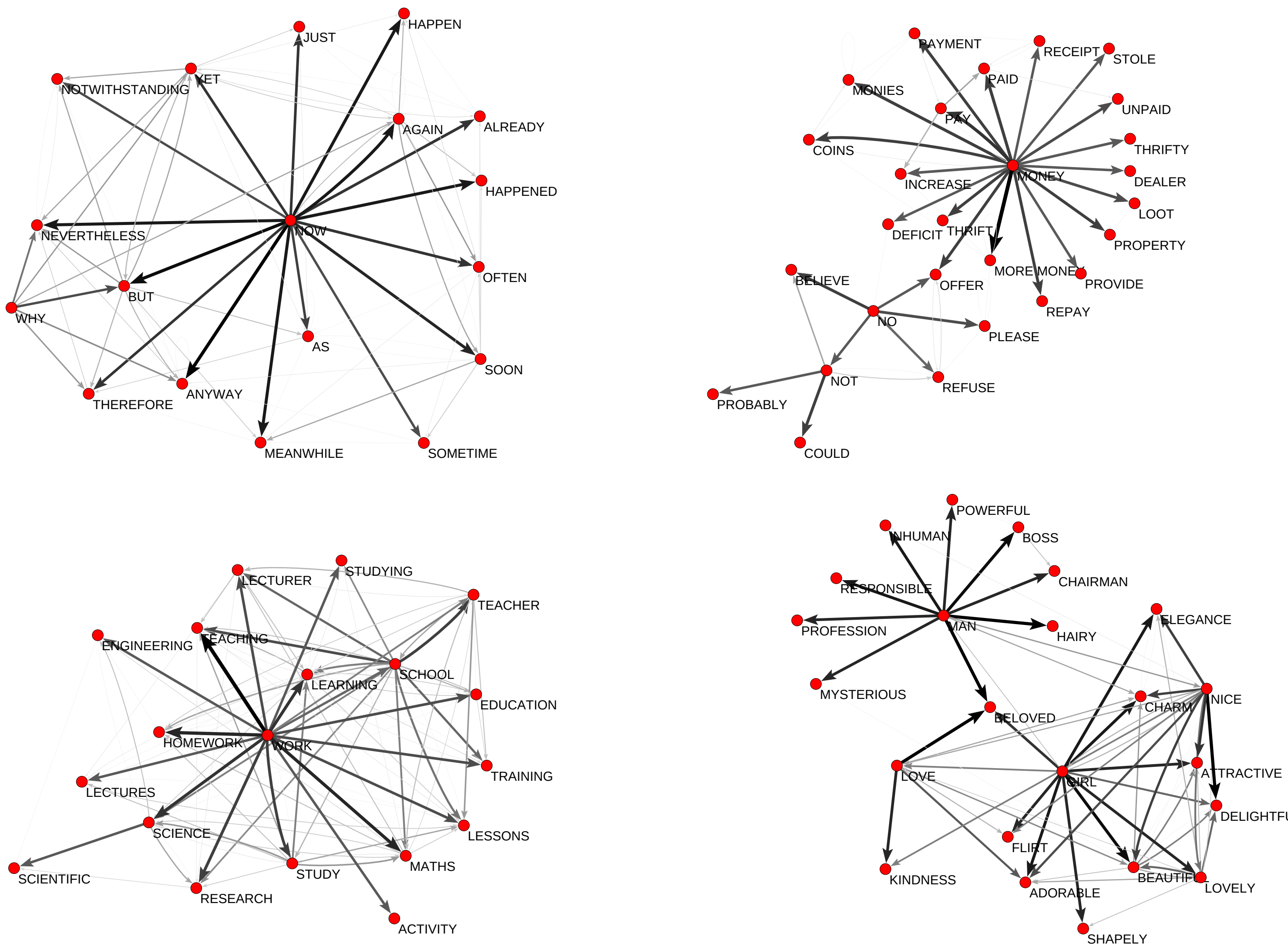
Primer: The Edinburgh Associative Thesaurus

- Na britanskih univerzah so od junija 1968 do maja 1971 anketirali tamkajšnje študente. Za vsako od besed iz danega seznama so morali napisati besedo, ki jim prva pride na misel.
- Dobili so veliko omrežje, sestavljeno iz 23219 točk (besede) in 325624 usmerjenih povezav, med katerimi je tudi 564 zank.

Izbrane teme v EAT



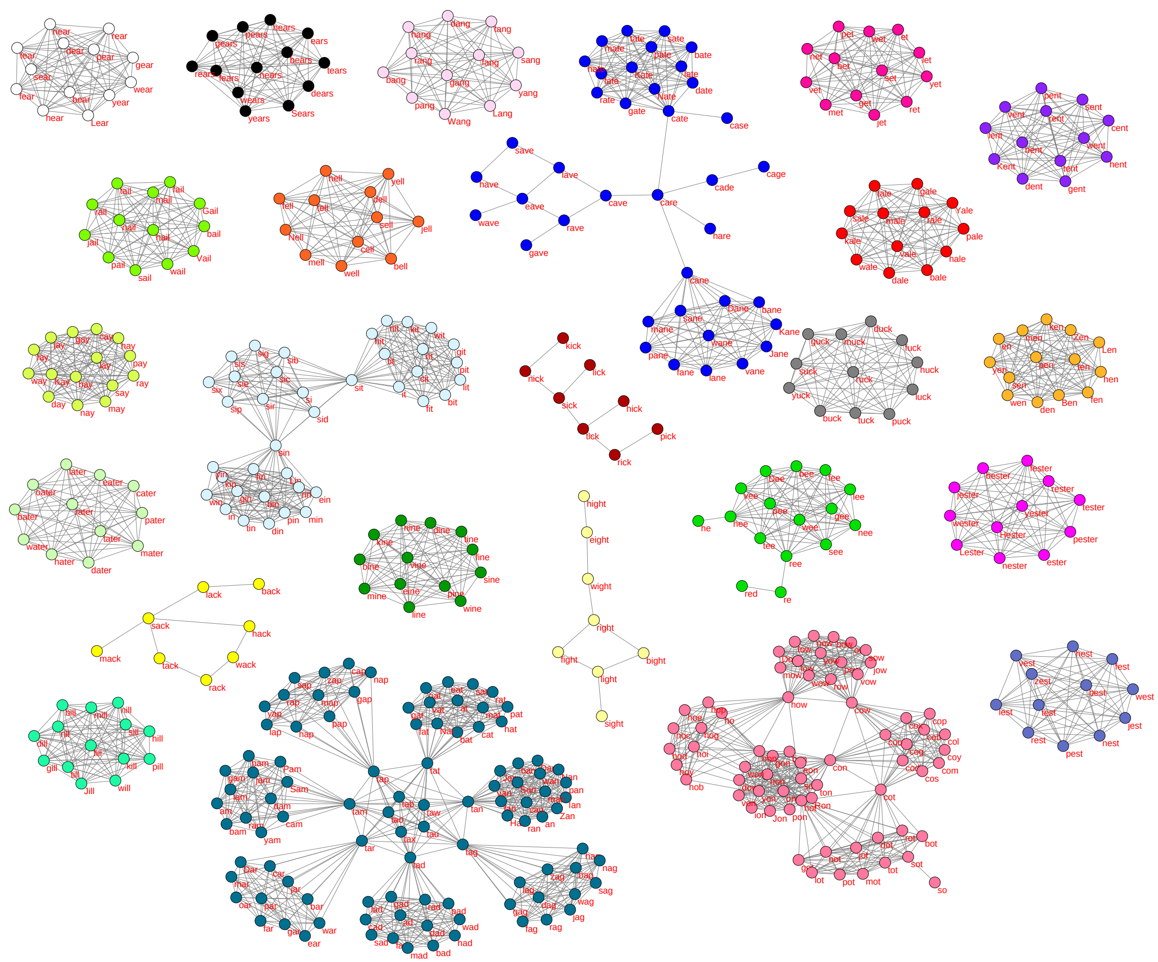
Izbrane teme v EAT



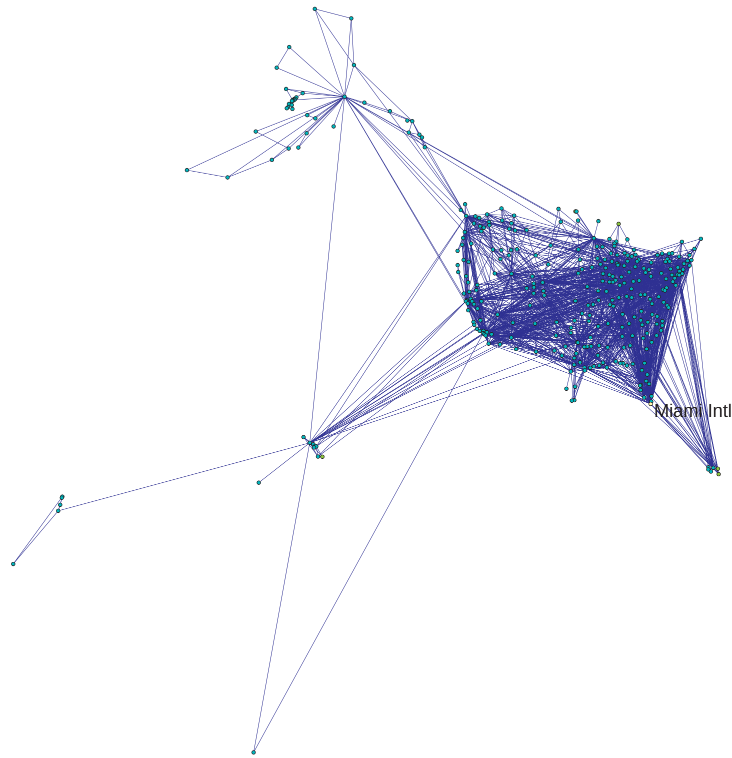
Primer: The Knuth's English Dictionary

- Knuthov slovar angleških besed je omrežje, ki ima:
  - 52652 točk – angleške besede dolžine od 2 do 8 črk
  - 92307 povezav – besedi sta sosedni, če lahko dobimo eno iz druge, tako da zamenjamo, odstranimo ali dodamo eno črko

Izbrane skupine v Knuthovem slovarju



Letalske povezave med 332 ameriškimi letališči (332 točk, 2116 povezav)



Poti v grafu

V *usmerjenem grafu* velja:

- Zaporedje točk grafa $(v_1, v_2, \ldots, v_k)$ imenujemo ***sprehod***, če velja

$$(v_i, v_{i+1}) \in A, i = 1 \ldots k-1$$

  (zaporedne točke morajo biti povezane z usmerjenimi povezavami).

- Zaporedje točk grafa $(v_1, v_2, \ldots, v_k)$ imenujemo ***veriga***, če velja

$$(v_i, v_{i+1}) \in A \text{ ali } (v_{i+1}, v_i) \in A, i = 1 \ldots k-1$$

  (zaporedne točke morajo biti povezane, smer povezav ni pomembna).

- Sprehod je ***osnoven***, če so vse točke, razen morda začetne in končne, različne. Osnoven sprehod bomo imenovali tudi *pot*.

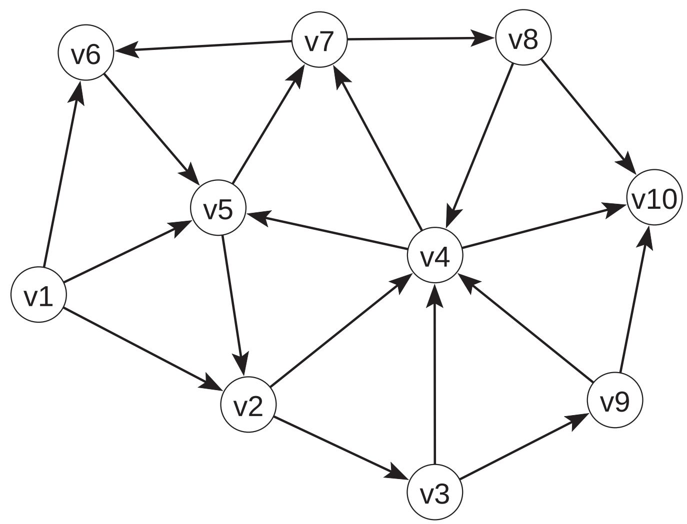
- Sprehod je ***enostaven***, če so vse povezave različne.

- Če je $v_1 = v_k$, se sprehod imenuje ***cikel*** (sprehod se začne in konča v isti točki).

- Veriga, ki se začne in konča v isti točki, se imenuje ***sklenjena veriga***.
- Če je $v_1 = v_k$ in $k = 1$, se cikel imenuje ***zanka*** (točka je povezana sama s sabo).
- ***Dolžina poti*** je $k - 1$ (število povezav na poti med dvema točkama).
- Med dvema točkama lahko obstaja več poti. Posebej zanimiva je ***najkrajša pot*** (oziroma vse najkrajše poti).

  V primeru relacije *sporočanje novic* predstavlja najkrajša pot potovanje informacije med osebami – npr. čimkrajše so najkrajše poti od vseh oseb do izbrane osebe, prej bo oseba seznanjena z novicami.

  V primeru relacije *je (bil) v stiku z* predstavlja najkrajša pot najverjetnejšo možnost prenosa okužbe. Za razliko od prejšnjega primera pa v tem primeru za neko osebo ni ugodno, da so poti čimkrajše.
- Dolžina najdaljše najkrajše poti v grafu se imenuje ***premer*** (*diameter*) grafa.



$\boxed{v1 - v6 - v4 - v8}$ ni niti veriga

$\boxed{v1 - v6 - v7 - v8}$ je veriga, ni pa sprehod

$\boxed{v8 - v4 - v5 - v2 - v4 - v10}$ je enostaven, ni pa osnoven sprehod

$\boxed{v8 - v4 - v5 - v2 - v3 - v9}$ je osnoven sprehod (pot)

$\boxed{v4 - v10 - v8 - v4}$ je sklenjena veriga

$\boxed{v6 - v5 - v7 - v6}$ je cikel

$\boxed{v1 - v2 - v3 - v9 - v10}$
je pot med točkama $v1$ in $v10$, ni pa najkrajša pot (dolžina 4).

Najkrajša pot med $v1$ in $v10$: $\boxed{v1 - v2 - v4 - v10}$ (dolž. 3).

Premer je 5: $\boxed{v7 - v6 - v5 - v2 - v3 - v9}$

Nekaj zanimivih rezultatov

Omrežja z zelo velikim številom točk imajo velikokrat kratke najkrajše poti:

- Povprečna dolžina najkrajše poti omrežja WWW, z več kot 800 milijoni točk, je okrog 19. Albert, R., Jeong, H., and Barabasi, A.-L. (1999): Diameter of the World-Wide Web. *Nature*, **401**, 130-131.

- Ocenjuje se, da je v socialnih omrežjih (koga poznaš) s preko 6 milijardami posameznikov, povprečna dolžina najkrajše poti med dvema posameznikoma okrog 6. Milgram, S. (1967): The small-world problem. *Psychol. Today*, **2**, 60-67.

  Small world experiment: Leta 1967 je psiholog Stanley Milgram izpeljal naslednji poskus s pismi: Pismo poslano iz Omahe (Nebraska), naj bi dobila izbrana končna oseba v Bostonu (Massachusetts). Osebe, ki so bile izbrane za poskus so bile naprošene, naj pismo pošljejo direktno končni osebi (če to osebo poznajo), sicer pa naj pismo s temi navodili pošljejo tistemu svojemu stiku, ki po njegovi/njeni oceni končno osebo pozna. Povprečna dolžina potovanja pisem, ki so dospela do končne osebe je bila 6 – *six degrees of separation*.

Komponente omrežja

Obravnavali bomo tri vrste komponent: *krepke*, *šibke* in *dvopovezane*.

Skupina točk omrežja se imenuje ***krepko povezana komponenta***, če lahko ob upoštevanju smeri povezav pridemo iz vsake točke te skupine v vsako drugo točko iz iste skupine.

Če smer ni pomembna (omrežje obravnavamo kot neusmerjeno), se taka skupina imenuje ***šibko povezana komponenta***.

Primer: Točke omrežja naj bodo zgradbe v mestu, povezave pa ulice, ki zgradbe povezujejo. Nekatere ulice so dvosmerne (neusmerjene povezave), nekatere pa enosmerne (usmerjene povezave).

Vse zgradbe, ki jih lahko medsebojno dosežemo s pomočjo avtomobila, se nahajajo v isti krepko povezani komponenti (upoštevati moramo enosmerne ceste).

Vse zgradbe, ki jih lahko medsebojno dosežemo peš, pa se nahajajo v isti šibko povezani komponeni (enosmerne ceste niso ovira).

Vzemimo primer relacije *koga bi povabil na zabavo*. Za vse osebe, ki se nahajajo v isti krepko povezani komponenti, velja:

- vsak posameznik v izbrani komponenti bi (vsaj posredno) povabil vsakega drugega iz iste komponente
- vsak posameznik v izbrani komponenti bi bil (vsaj posredno) povabljen od vsakega drugega iz iste komponente

Če je omrežje neusmerjeno, so krepke komponente kar enake šibkim. Te imenujemo kar komponente.

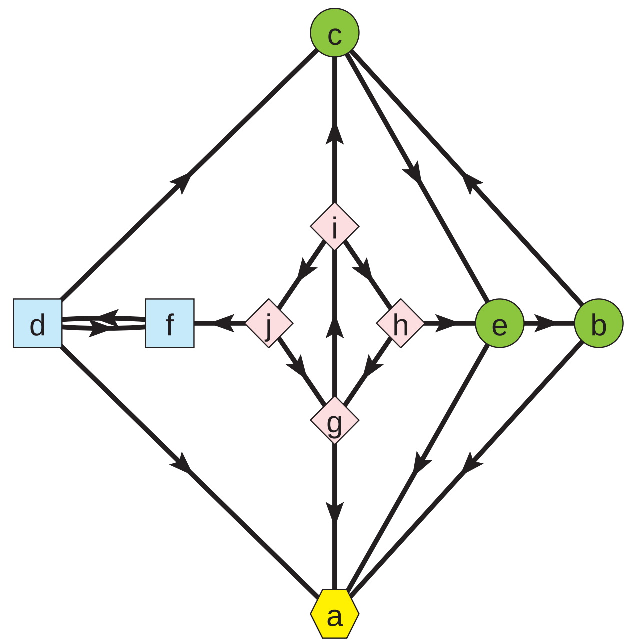
Povezava ***sprehodov*** v omrežju in ***komponent***

Med dvema točkama obstaja *sprehod*, če se točki nahajata v isti *krepko* povezani komponenti (rečemo tudi, da sta točki krepko povezani).

Med dvema točkama obstaja *veriga*, če se točki nahajata v isti *šibko* povezani komponenti (rečemo tudi, da sta točki šibko povezani).

Omrežje je ***krepko/šibko povezano***, če je vsak par točk krepko/šibko povezan.

Primer

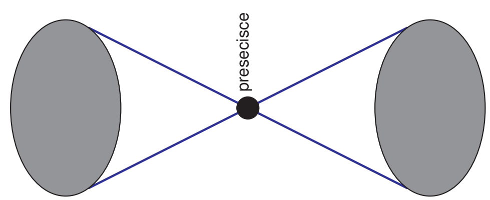


Vse točke v omrežju na sliki so šibko povezane.
Obstajajo pa štiri krepko povezane komponente:
1. (a), 2. (b, c, e), 3. (d, f), 4. (g, h, i, j)

Točke, ki se nahajajo v istih krepko povezanih komponentah, so predstavjene z istimi oblikami.

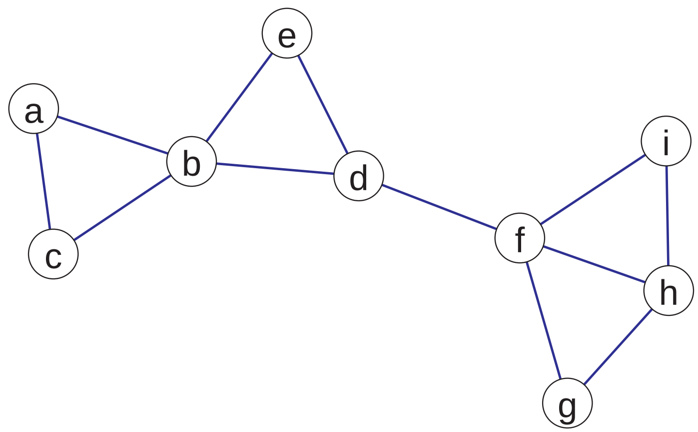
Dvopovezane komponente omrežja

Vzemimo povezano neusmerjeno omrežje. Točka a tega omrežja je *presečišče (artikulacijska točka)* omrežja, če obstajata dve drugi, različni točki v in w, taki, da vsaka pot med tema točkama vsebuje tudi točko a. Preprosteje rečeno je točka a presečišče omrežja, če odstranitev te točke iz omrežja povzroči razpad omrežja na dve ali več komponent.



Omrežje je *dvopovezano*, če za vsako trojico točk a, v in w obstaja pot med točkama v in w, ki ne vsebuje tudi točke a. Enostavneje rečeno: omrežje je (po točkah) dvopovezano, če pri odstranitvi katerekoli točke iz njega ostane povezano. Dvopovezano omrežje torej ne vsebuje nobenega presečišča.

Primer



Omrežje na sliki (aho1.net) je sestavljeno iz 4 dvopovezanih komponent:

- (a, b, c)
- (b, d, e)
- (d, f)
- (f, g, h, i)

Presečišča so b, d in f.

Primer: Ameriška letališča.

Jedra

Skupina točk je k**-jedro**, če je vsaka točka iz te skupine povezana vsaj s k točkami iz iste skupine.

Oziroma formalno:

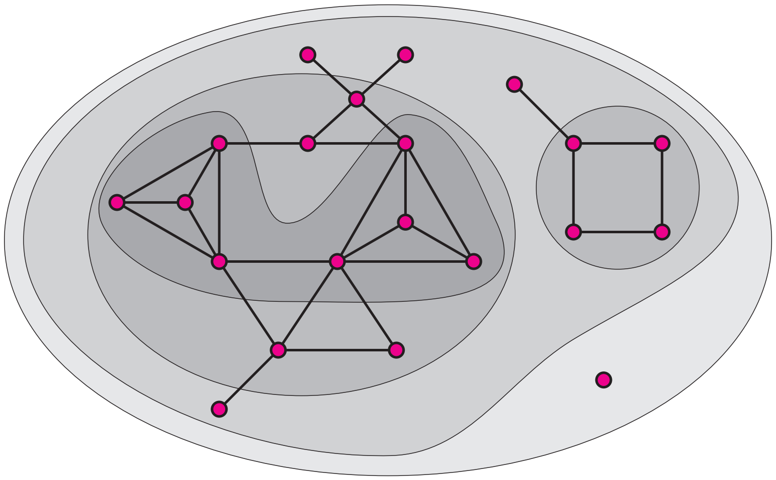
Naj bo $N = (V, L)$, $L \subseteq V \times V$ omrežje.

Največje podomrežje $J = (W, R \cap W \times W)$ porojeno s podmnožico $W \subseteq V$ imenujemo k-*jedro* (glede na stopnje točk) natanko takrat, ko velja: $\forall v \in W : stopnja_J(v) \geq k$.

Poseben primer:

Skupina točk se imenuje ***klika***, če je vsaka točka iz skupine povezana z vsemi drugimi točkami iz iste skupine.

0, 1, 2 in 3 jedro:



Lastnosti jeder:

- Jedra so gnezdena: $i < j \implies H_j \subseteq H_i$
- Jedra niso nujno povezani podgrafi.
- Jedra lahko posplošimo na omrežja z vrednostmi na povezavah.

Primer: Ameriška letališča.

Iskanje skupin - metode community detection

Poiskati želimo take skupine, da bodo povezave znotraj skupin gostejše (in z večjimi vrednostmi), povezave med skupinami pa redkejše (in z manjšimi vrednostmi).

V Pajku sta na voljo dve metodi za iskanje skupin po metodah 'community detection' in sicer metoda *Louvain* in metoda *VOS Clustering*.

Po metodi Louvain iščemo razvrstitev v skupine, ki ima največjo *modularnost (modularity - Q)*:

$$Q = \frac{1}{2m} \sum_s (e_s - r * \frac{K_s^2}{2m})$$

- m – skupno število povezav,
- s – skupina,
- $e_s = \sum_{ij \in s} A_{ij}$ – dvakratno število povezav znotraj skupine s

- $K_s = \sum_{i \in s} k_i$ – vsota stopenj za točke v skupini s
- r – *resolution parameter*, privzeta vrednost je 1, ki ustreza osnovni definiciji modularnosti

Podobna metoda je *VOS Clustering*, le da se namesto modularnosti vzame *VOS quality function*.

Na voljo je precej parametrov s katerimi usmerjamo iskanje skupin, ki pridejo v poštev pri večjih omrežjih.

Smiselno pa je preizkusiti različne vrednosti parametra *resolution*. Ta je ponavadi nastavljen na vrednost 1. Vrednost večja od 1 pomeni iskanje večjega števila (manjših) skupin, vrednost manjša od 1 pa iskanje manjšega števila (večjih) skupin.

Primeri: football.net, import.net.

Še nekaj primerov

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** (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.

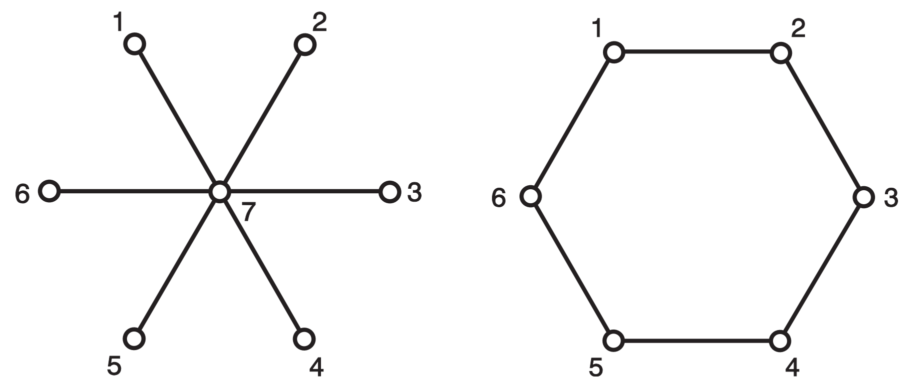
  Primera:
  Neka oseba je *vplivna*, če ukazuje veliko ostalim.
  Neka oseba ima veliko *podporo*, če zanjo glasuje veliko ljudi.

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



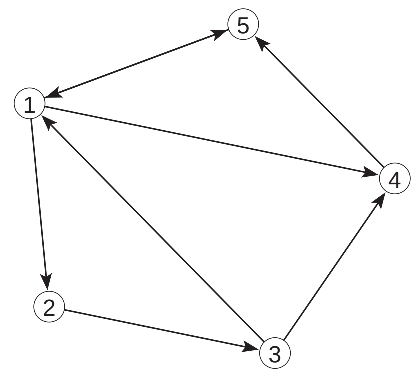
Po vseh treh kriterijih je enota 7 v zvezdi najbolj središčna in ostale enote enako ne-središčne. Vse enote v ciklu so enako središčne po vseh kriterijih.

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.

Vhodna stopnja točke **(input degree)**: število povezav, ki vstopa v točko.
Izhodna stopnja **(output degree)**: število povezav, ki izstopa iz točke.
Skupna stopnja točke **(all degree)**: skupno število povezav, ki ima eno krajišče v točki.

Primer



Točka	Vhodna st.	Izhodna st.	Skupna st.
1	2	3	5
2	1	1	2
3	1	2	3
4	2	1	3
5	2	1	3

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 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.

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.

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

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

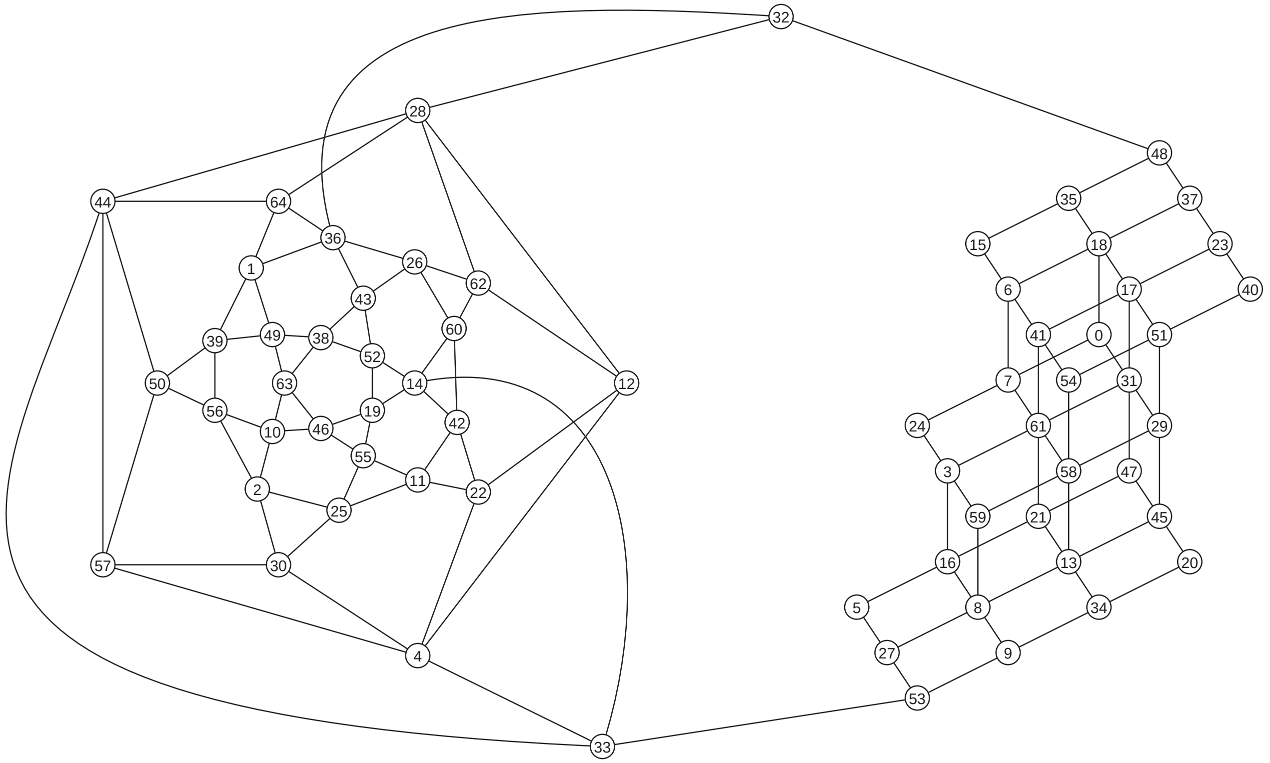
Središčnost enote x glede na vmesnost je vsota verjetnosti preko vseh možnih parov točk, da bo najkrajša pot med y in z potekala skozi točko x.

Izbira ustrezne mere središčnosti

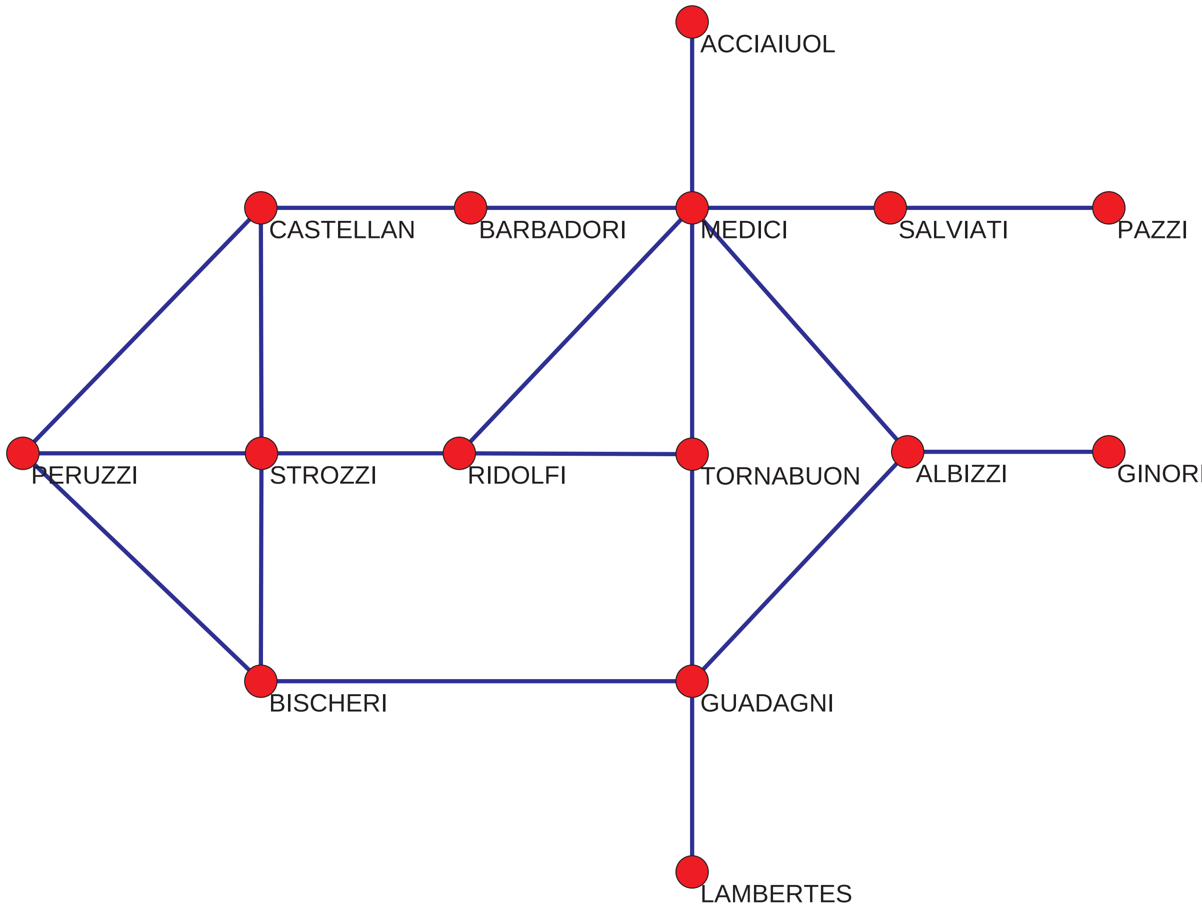
Omenjene mere središčnosti lahko v določenih primerih vrnejo zelo različne rezultate. Zato moramo biti pri izbiri ustrezne mere središčnosti zelo previdni:

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.


32
28
48
44
64
35
37
36
15
18
23
1
26
62
6
17
40
43
60
41
0
51
39
49
38
52
50
63
14
12
7
54
31
56
19
10
46
42
24
61
29
55
11
2
22
3
58
47
25
59
21
45
57
30
16
13
20
5
8
34
4
27
9
53
33

Primer: Zakonske vezi med florentinskimi rodbinami



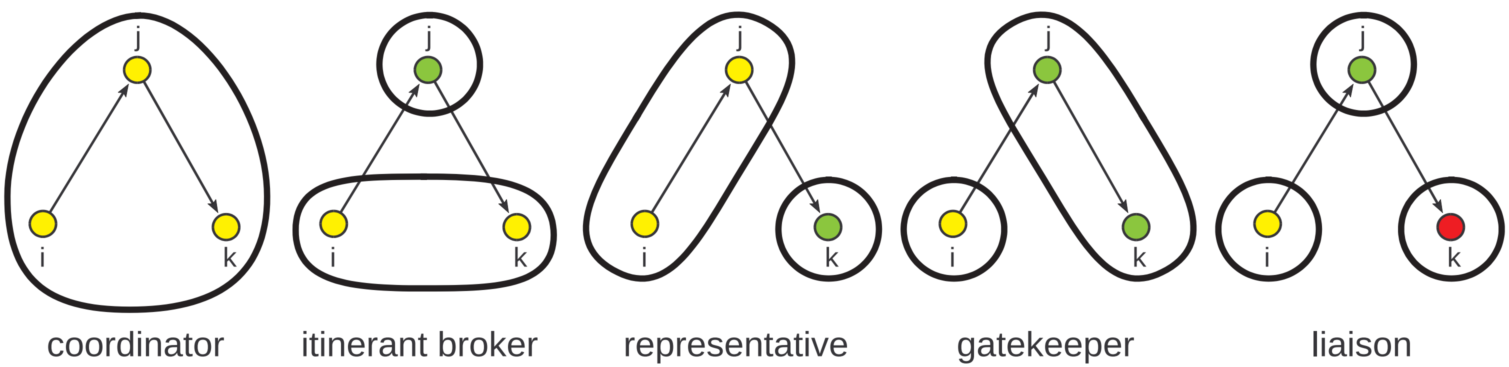
Relativne mere središčnosti florentinskih rodbin

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

Posredniške vloge

Enota v omrežju, kjer so na nek način določene skupine (npr. osebe ženskega spola in osebe moškega spola) lahko nastopa v naslednjih posredniških vlogah (brokerage roles):

- notranji posrednik (coordinator),
- zunanji posrednik (itinerant broker),
- predstavnik (representative),
- vratar (gatekeeper),
- zveza (liaison).



Rodovniki kot velika omrežja

```
0 HEAD
1 FILE ROYALS.GED
...
0 @I58@ INDI
1 NAME Charles Philip Arthur/Windsor/
1 TITL Prince
1 SEX M
1 BIRT
2 DATE 14 NOV 1948
2 PLAC Buckingham Palace, London
1 CHR
2 DATE 15 DEC 1948
2 PLAC Buckingham Palace, Music Room
1 FAMS @F16@
1 FAMC @F14@
...
...
0 @I65@ INDI
1 NAME Diana Frances /Spencer/
1 TITL Lady
1 SEX F
1 BIRT
2 DATE 1 JUL 1961
2 PLAC Park House, Sandringham
1 CHR
2 PLAC Sandringham, Church
1 FAMS @F16@
1 FAMC @F78@
...
...
```

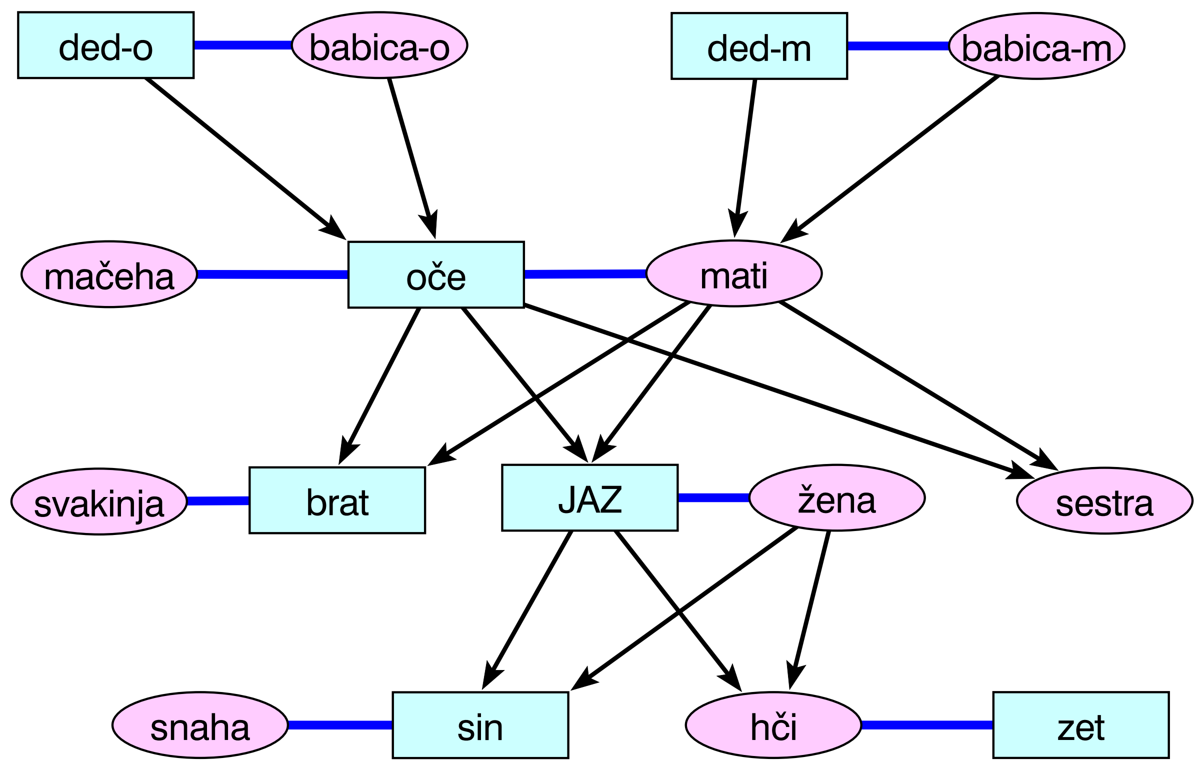
```
0 @I115@ INDI
1 NAME William Arthur Philip/Windsor/
1 TITL Prince
1 SEX M
1 BIRT
2 DATE 21 JUN 1982
2 PLAC St.Mary's Hospital, Paddington
1 CHR
2 DATE 4 AUG 1982
2 PLAC Music Room, Buckingham Palace
1 FAMC @F16@
...
0 @I116@ INDI
1 NAME Henry Charles Albert/Windsor/
1 TITL Prince
1 SEX M
1 BIRT
2 DATE 15 SEP 1984
2 PLAC St.Mary's Hosp., Paddington
1 FAMC @F16@
...
0 @F16@ FAM
1 HUSB @I58@
1 WIFE @I65@
1 CHIL @I115@
1 CHIL @I116@
1 DIV N
1 MARR
2 DATE 29 JUL 1981
2 PLAC St.Paul's Cathedral, London
```

Predstavitev rodovnikov z omrežji

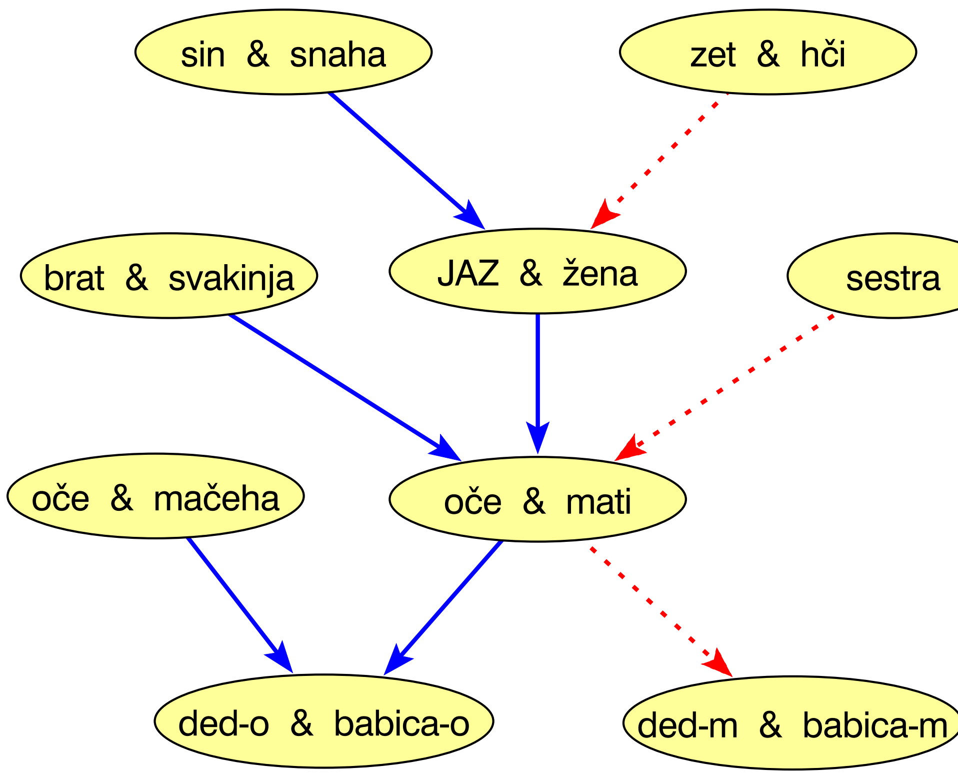
Rodovnike lahko predstavimo z omrežji na več različnih načinov:

- kot navadne rodovnike (Ore-graph),
- kot parne rodovnike (p-graph),
- kot dvodelne parne rodovnike (bipartite p-graph).

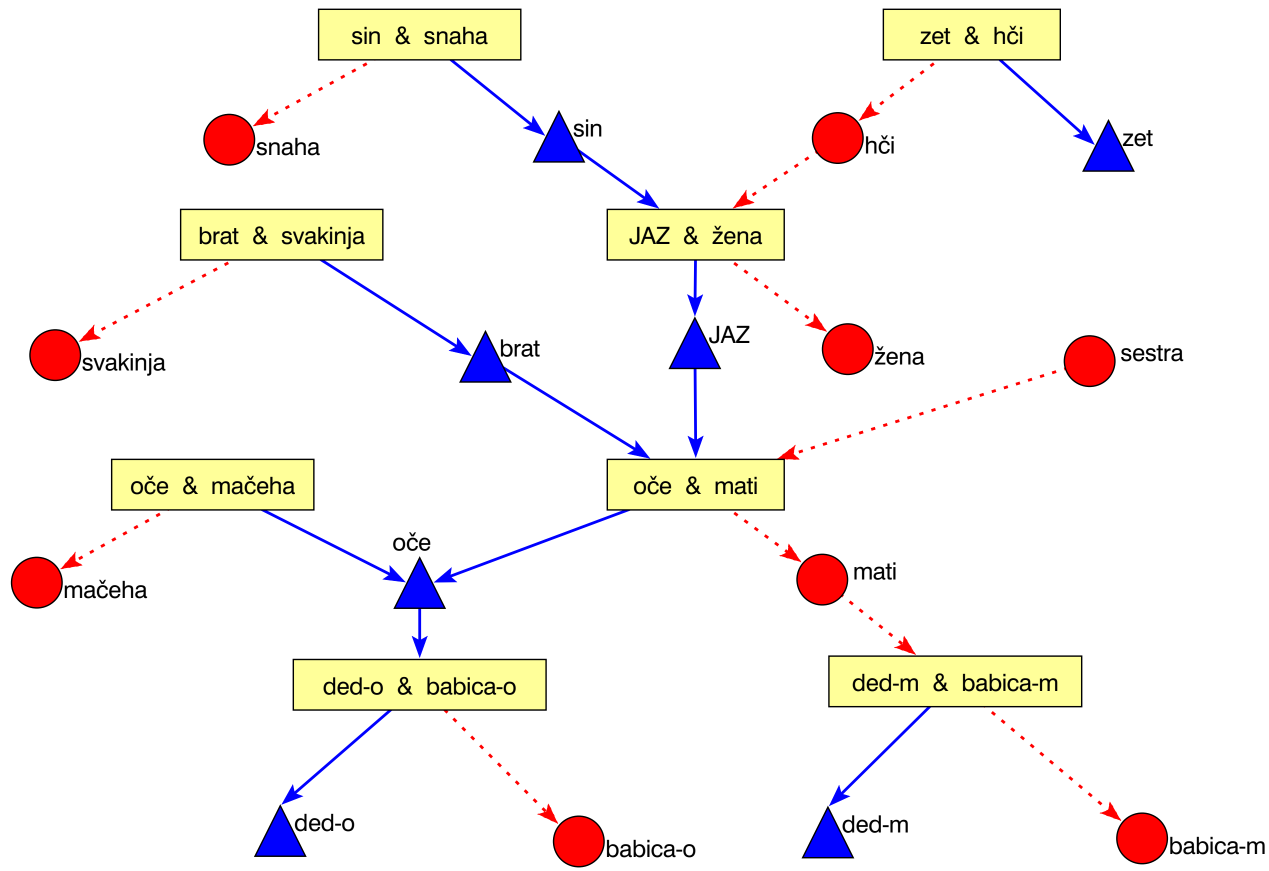
Navadni rodovnik: V navadnem rodovniku je vsaka oseba predstavljena s svojo točko, poroke so predstavljene z neusmerjenimi povezavami, relacija *je eden od staršev od* pa z usmerjenimi povezavami, ki kažejo od vsakega od staršev do njunih otrok.



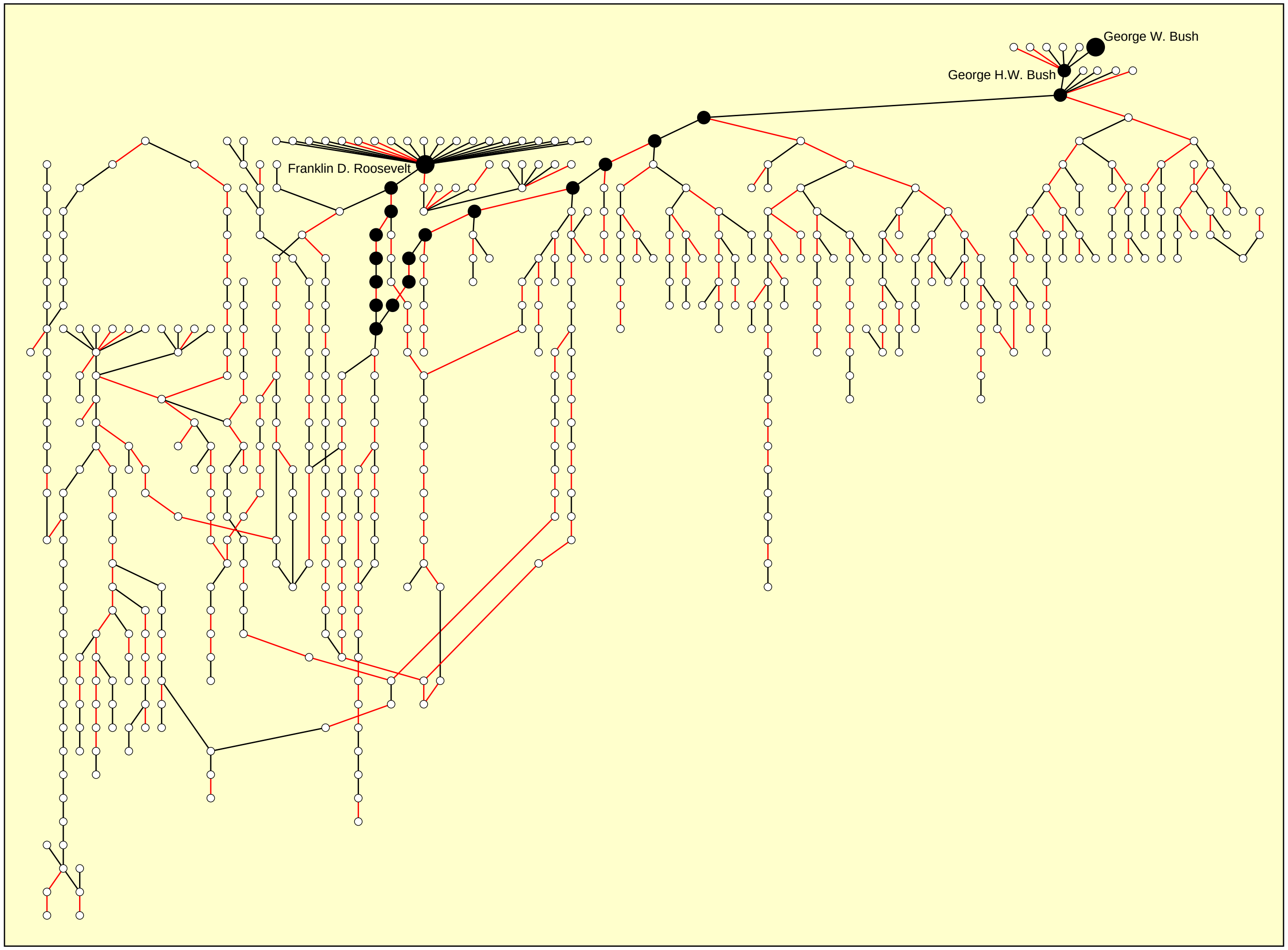
Parni rodovnik: V parnem rodovniku predstavljajo točke posameznike ali pare. Če neka oseba še ni poročena, je predstavljena s svojo točko, sicer je predstavljena s svojim partnerjem v skupni točki. V parnem rodovniku obstajajo samo usmerjene povezave, ki kažejo od otrok na njihove starše.



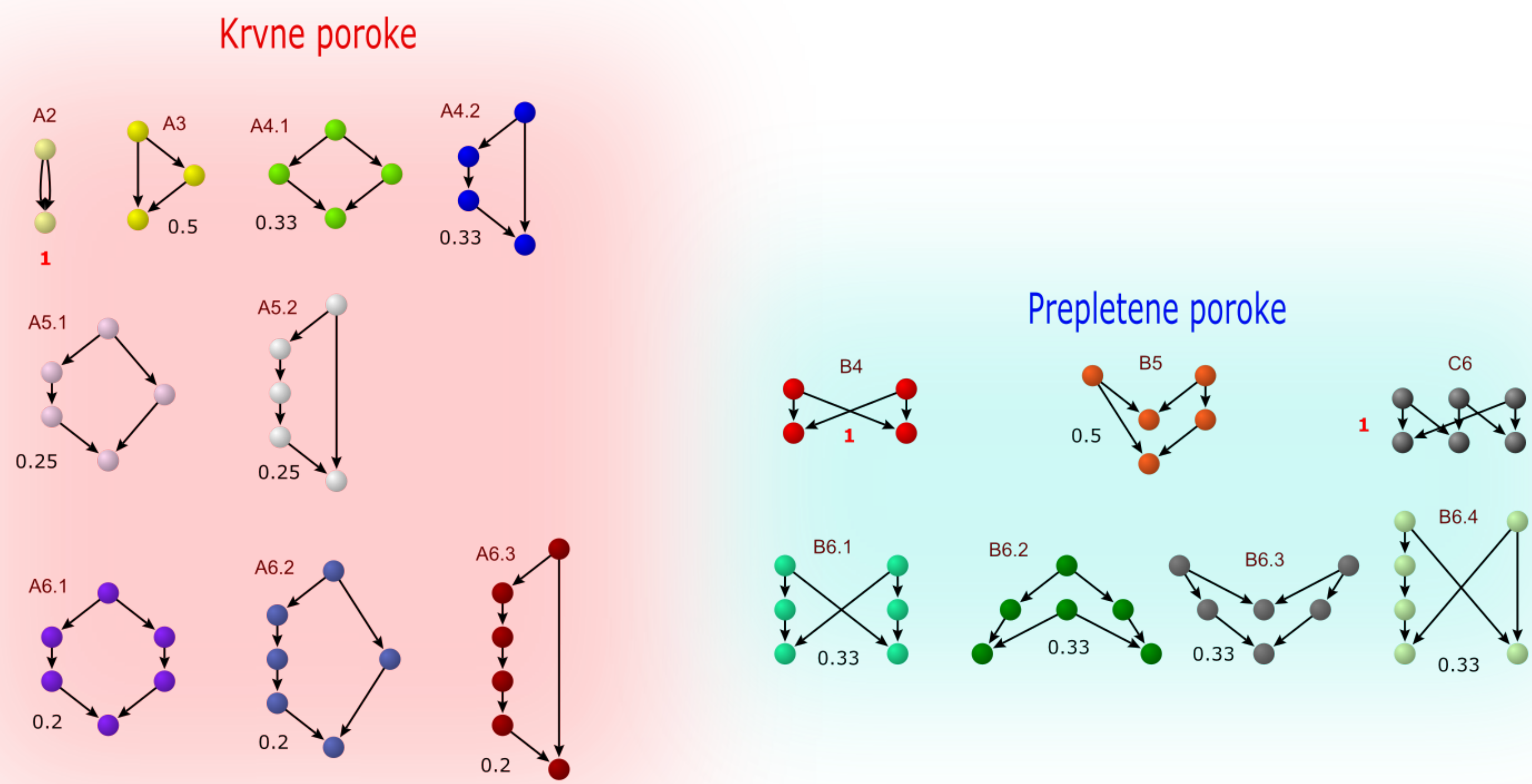
Dvodelni parni rodovnik: ima dve vrsti točk – točke, ki predstavljajo poročene pare (pravokotniki) in točke, ki predstavljajo posameznike (okrogle točke za ženske in trikotniki za moške). Usmerjene povezave tudi tu kažejo od otrok na njihove starše.



Največja povezana komponenta v rodovniku ameriških predsednikov

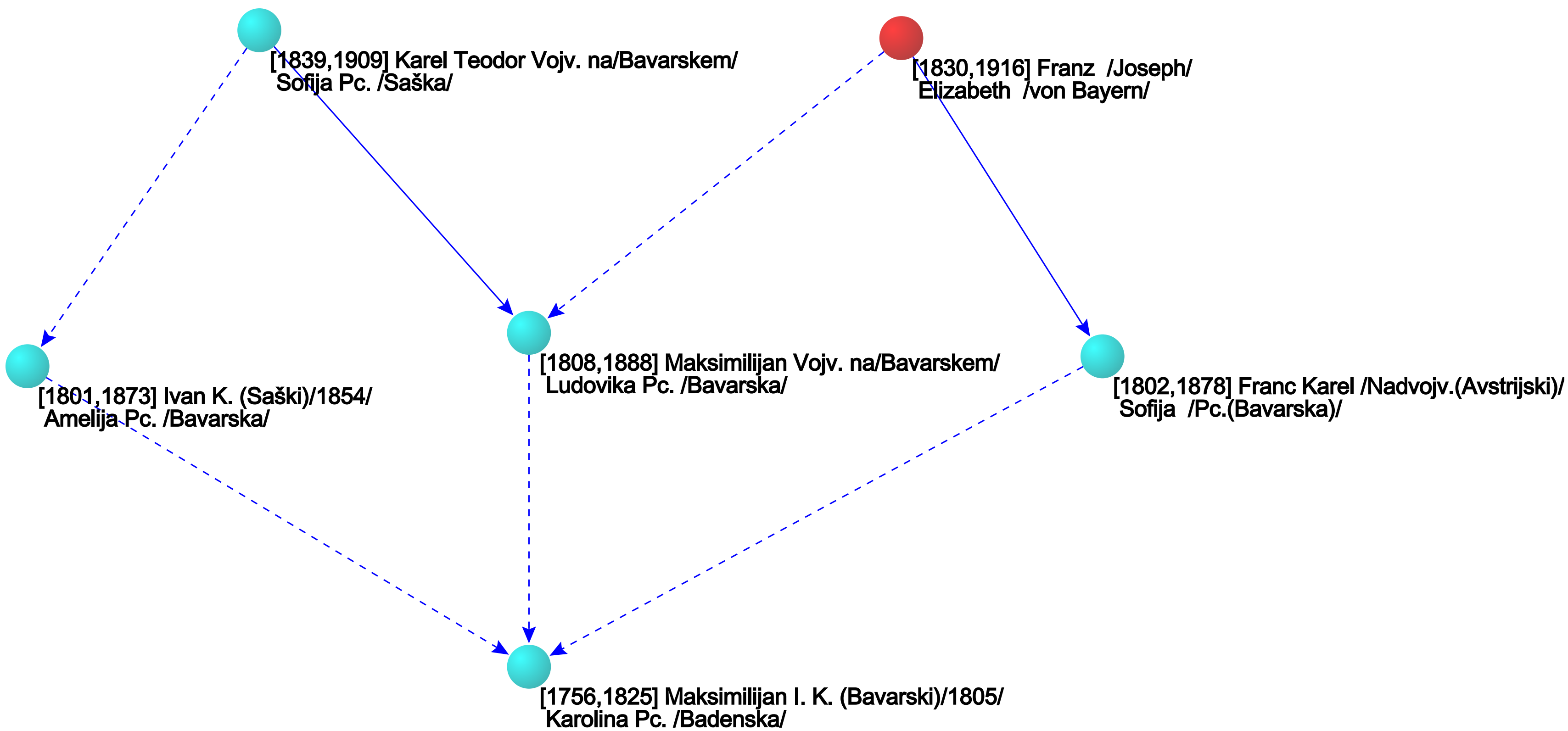


Prepletene poroke (parni rodovniki z 2 do 6 točkami)

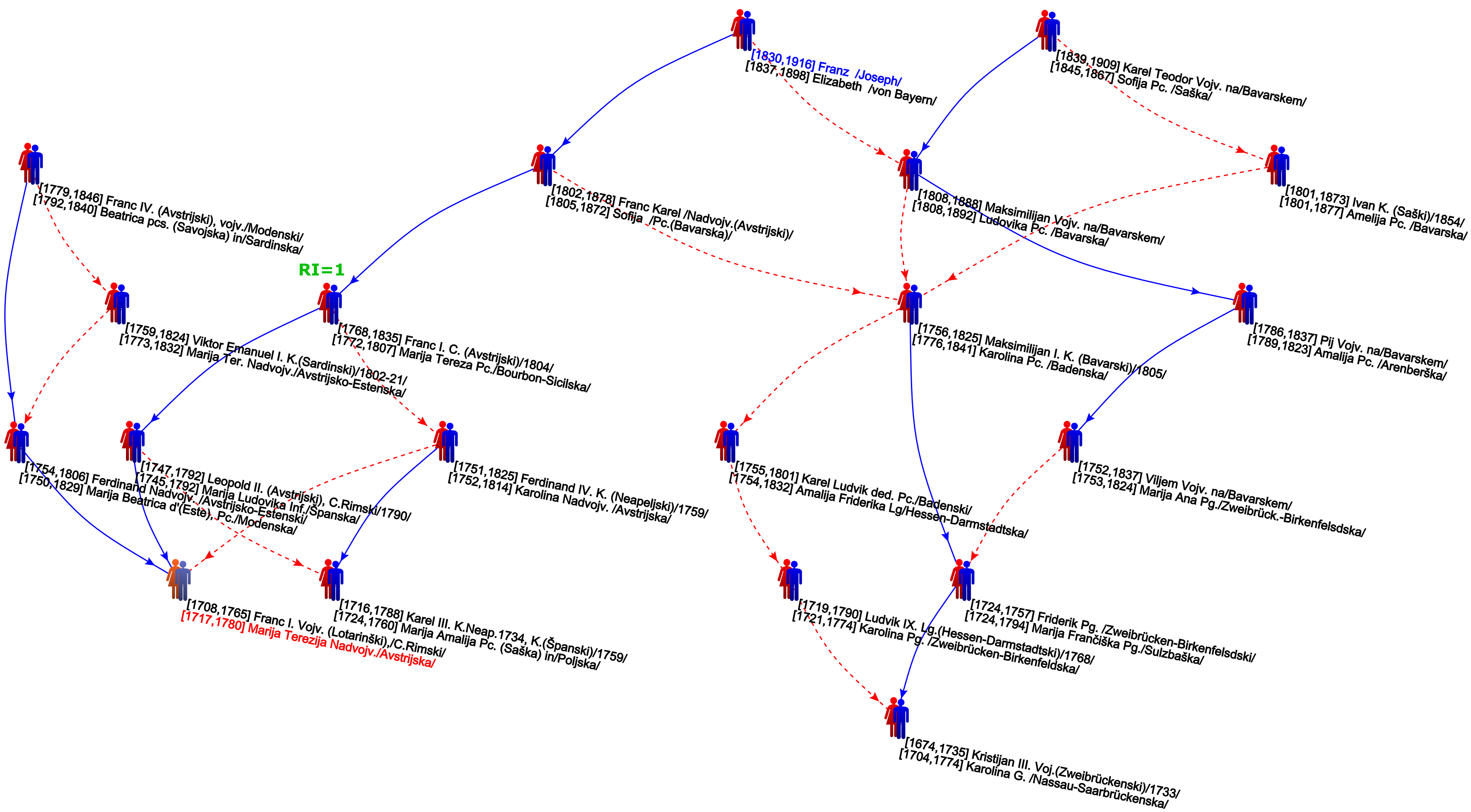


Krvni poroki vnuk-vnukinja v rodovniku evropskih plemiških družin

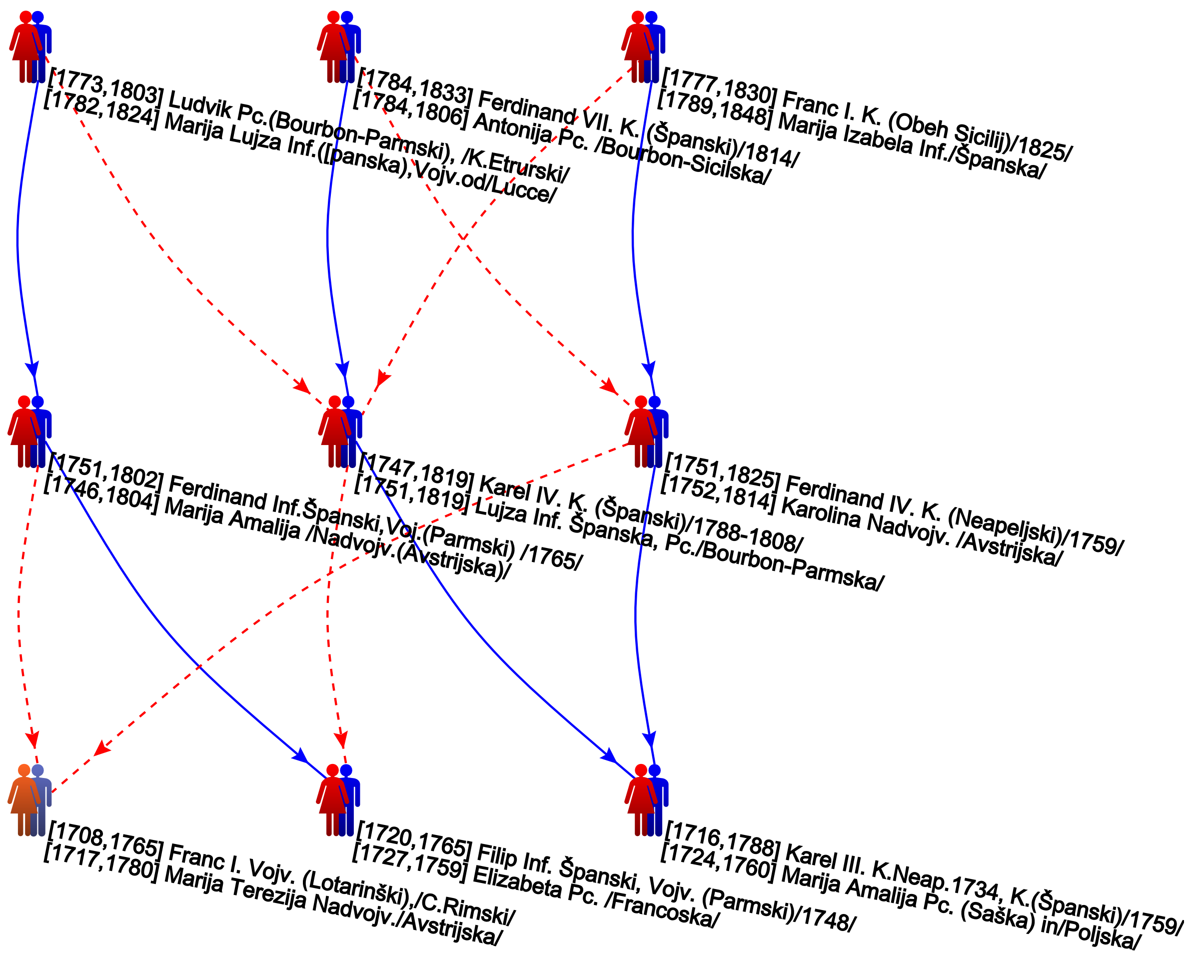
Plemiški rodovi so med seboj veliko bolj sorodstveno povezani kot so 'navadni'. Leta 2016 je bila 100 letnica smrti cesarja Franca Jožefa. Poročil se je s svojo sestično 'Sisi'.



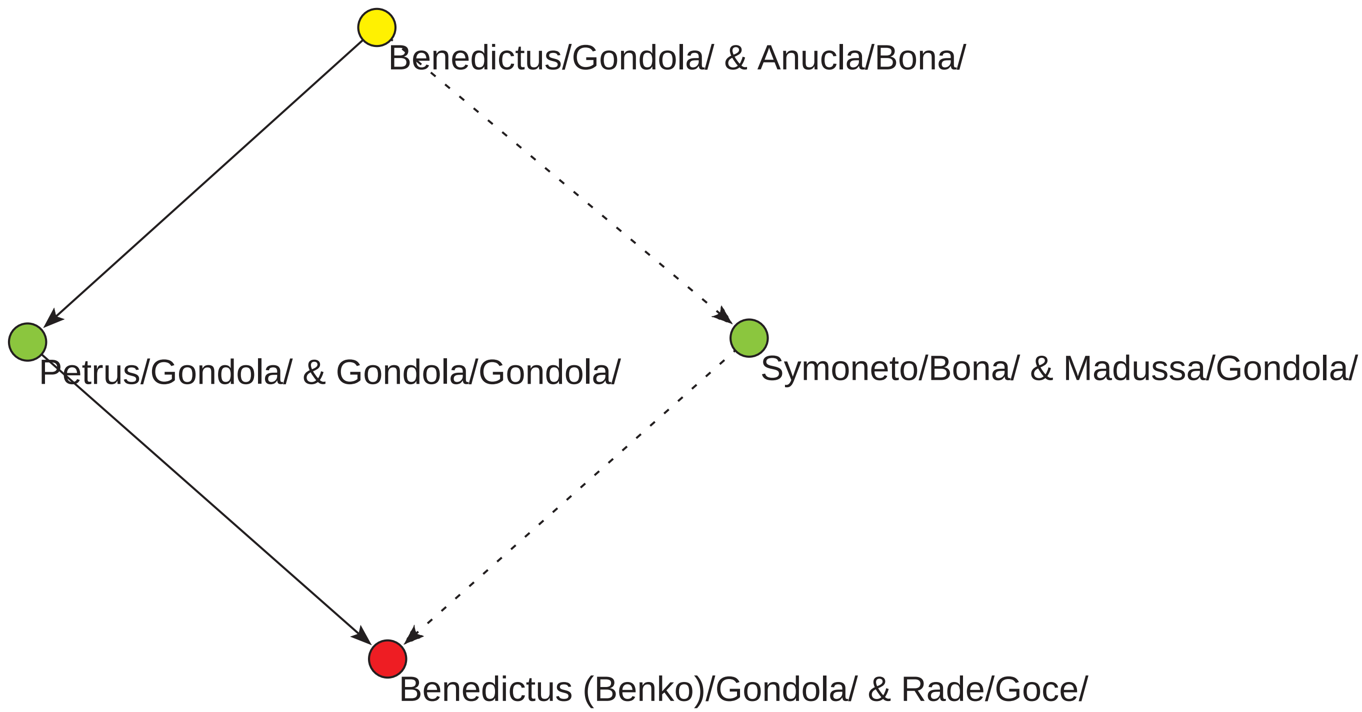
...Prepletene poroke



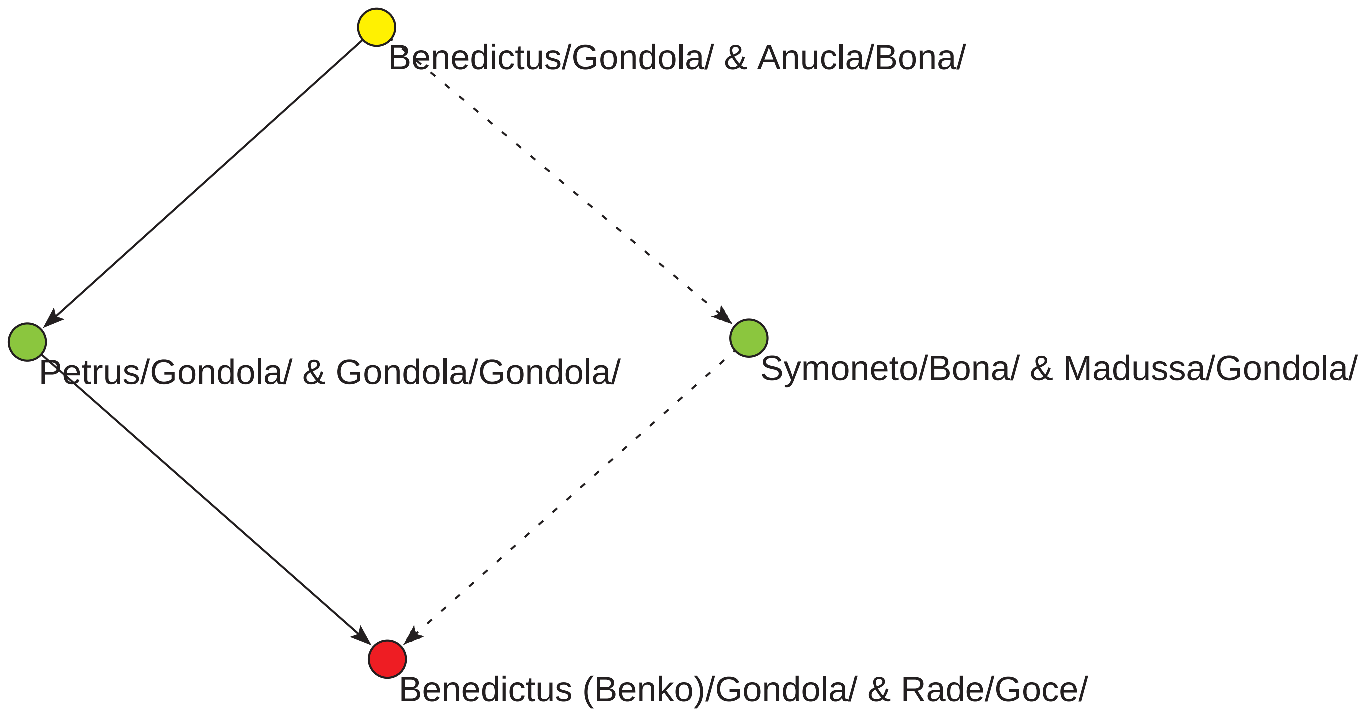
...Prepletene poroke



Krvna poroka vnuk-vnukinja v rodovniku dubrovniških plemiških družin



Krvna poroka vnuk-vnukinja v rodovniku dubrovniških plemiških družin



Izmanjave bratov sester v rodovniku dubrovniških plemiških družin