

KartOO network

# Analiza velikih omrežij

Vladimir Batagelj

Univerza v Ljubljani  
FMF, matematika

**net konferenca 2007**  
Ljubljana, 31. januar 2007

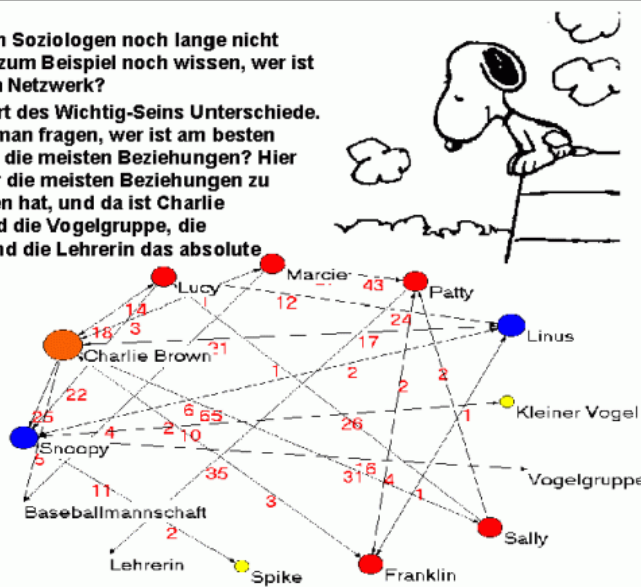
# Kazalo

1	Omrežja – osnovni pojmi . . . . .	1
6	<b>Pajek</b> . . . . .	6
7	<b>Pajek</b> in velika omrežja . . . . .	7
9	Pristopi k analizi velikih omrežij . . . . .	9
10	Vmesnost . . . . .	10
12	Kazala in vsebine . . . . .	12
14	Otoki . . . . .	14
20	Sredice in posplošene sredice . . . . .	20
23	Dvovrstna omrežja . . . . .	23
24	Omrežje z <b>They Rule</b> . . . . .	24
25	Omrežja iz podatkovnih tabel . . . . .	25
27	Primer: Evropski projekti na temo simulacij . . . . .	27
34	Pretvorba dvovrstnih omrežij na enovrstna . . . . .	34
38	Dvovrstne sredice . . . . .	38
42	Izbrane knjige o analizi omrežij . . . . .	42

## Omrežja – osnovni pojmi

Aber damit geben sich Soziologen noch lange nicht zufrieden. Sie wollen zum Beispiel noch wissen, wer ist hier wichtig in diesem Netzwerk?

Dabei gibt es in der Art des Wichtig-Seins Unterschiede. Beispielsweise kann man fragen, wer ist am besten eingebunden, wer hat die meisten Beziehungen? Hier wurde gemessen, wer die meisten Beziehungen zu anderen aufgenommen hat, und da ist Charlie Spitzenreiter, während die Vogelgruppe, die Baseballmannschaft und die Lehrerin das absolute Schlußlicht bilden.



Alexandra Schuler/ Marion Laging-Glaser:

Analyse von Snoopy Comics

Osnovni sestavini *omrežja* sta množica *točk*, ki predstavljajo izbrane *enote*, in množica *povezav*, ki predstavljajo *odnose* (relacije) med enotami. Točke in povezave določajo *graf*.

Pri povezavi je lahko smer pomembna – *usmerjena* povezava, ali pa ni *neusmerjena* povezava.

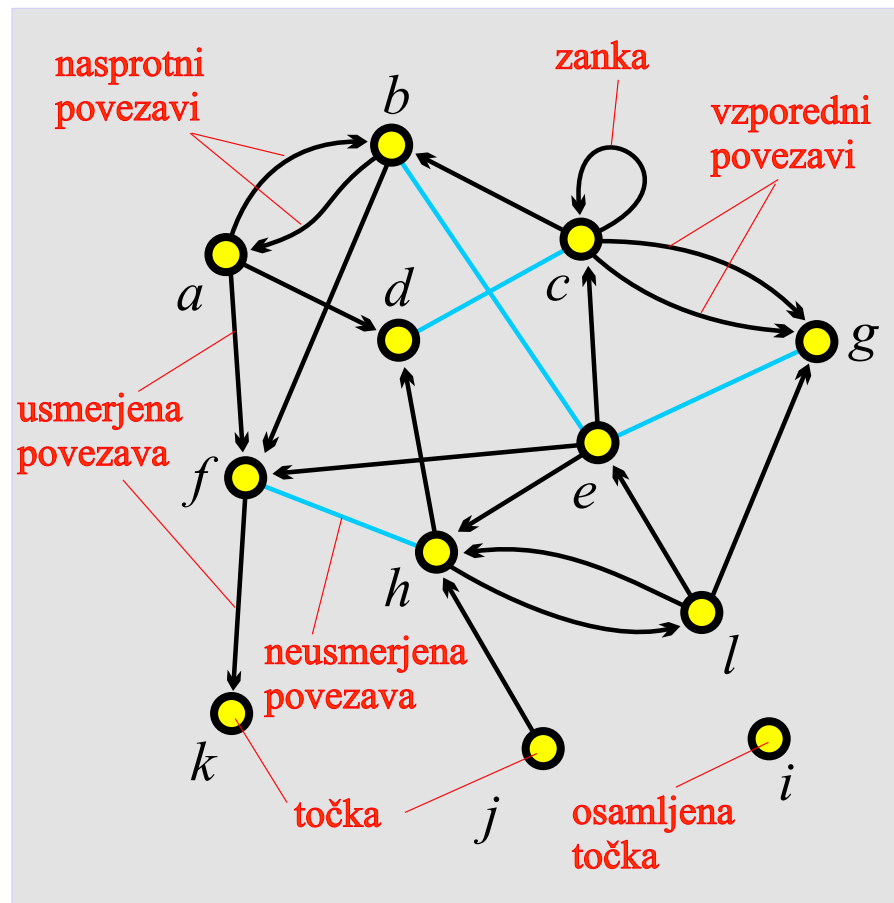
O točkah in povezavah lahko poznamo dodatne podatke – njihove *lastnosti*.

Npr. oznaka, vrsta, vrednost, ...

## Omrežje = Graf + Podatki

Ti podatki so lahko izmerjeni ali izračunani.

## Graf



enota – *točka*, vozlišče  
relacija – *povezava*

*usmerjena povezava* ( $a, d$ )

$a$  je njen *začetek*

$d$  pa njen *konec*.

*neusmerjena povezava* ( $c: d$ )

$c$  in  $d$  sta njeni *krajišči*.

## Omrežja

*Omrežje*  $\mathcal{N} = (\mathcal{V}, \mathcal{L}, \mathcal{P}, \mathcal{W})$  je določeno z:

- *grafom*  $\mathcal{G} = (\mathcal{V}, \mathcal{L})$ , kjer je  $\mathcal{V}$  množica točk,  $\mathcal{A}$  je množica usmerjenih povezav, in  $\mathcal{E}$  je množica neusmerjenih povezav. Z  $\mathcal{L} = \mathcal{E} \cup \mathcal{A}$  označimo množico vseh povezav.

$$n = |\mathcal{V}|, m = |\mathcal{L}|$$

- $\mathcal{P}$  *lastnosti* / točkovne vrednosti:  $p: \mathcal{V} \rightarrow A$
- $\mathcal{W}$  *uteži* / povezavne vrednosti:  $w: \mathcal{L} \rightarrow B$

Velikost omrežja/grafa je določena z dvema številoma: številom točk  $n = |\mathcal{V}|$  in številom povezav  $m = |\mathcal{L}|$ .

Omrežja na nekaj deset točkah so *majhna* – lahko jih narišemo in za njihovo analizo uporabimo veliko algoritmov (*UCINET*, *NetMiner*). Tudi *srednje* velika omrežja (nekaj sto točk) še lahko narišemo (preglednost !?), nekateri postopki že odpovejo.

## Velika omrežja

*Veliko* omrežje – več tisoč ali celo milijonov točk. V celoti ga lahko shranimo v pomnilnik računalnika – sicer *ogromno* omrežje. Velika omrežja so postala dejstvo sredi devetdesetih let.

Običajno so redka  $m \ll n^2$ ; pogosto:  $m = O(n)$  ali  $m = O(n \log n)$ .

Primeri:

omrežje	velikost	$n =  \mathcal{V} $	$m =  \mathcal{L} $	vir
slovar ODLIS	61K	2909	18419	ODLIS online
Sklicevanja SOM	168K	4470	12731	Garfieldova zbirka
Molekula 1ATN	74K	5020	5128	Brookhaven PDB
Comput. geometry	140K	7343	11898	BiBTEX bibliographies
angleške besede 2-8	520K	52652	89038	Knuth's English words
Internetske iskalne poti	1.7M	124651	207214	Internet Mapping Project
Rodovnik Franklin	12M	203909	195650	RoperId.com gedcoms
World-Wide-Web	3.6M	325729	1497135	Notre Dame Networks
Igralci	3.9M	392400	1342595	Notre Dame Networks
Patenti ZDA	82M	3774768	16522438	Nber
SI internet	38M	5547916	62259968	Najdi Si

## Zahtevnost algoritmov

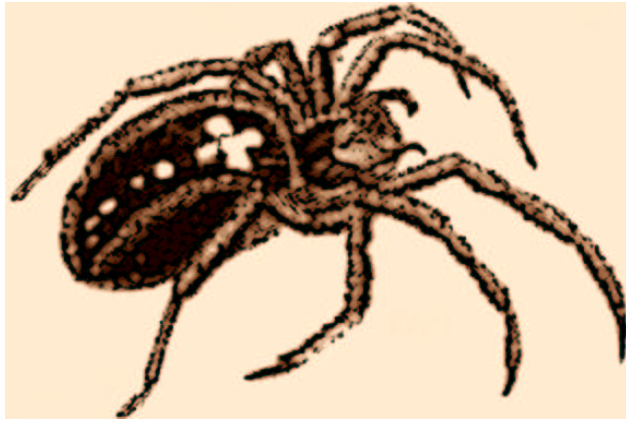
Poglejmo časovne zahtevnosti nekaj značilnih algoritmov  
(1999, Pentium/64M/90MHz):

	$T(n)$	1.000	10.000	100.000	1.000.000	10.000.000
Shuffle	$O(n)$	0.00 s	0.015 s	0.17 s	2.22 s	22.2 s
HeapSort	$O(n \log n)$	0.00 s	0.06 s	0.98 s	14.4 s	2.8 m
AB	$O(n\sqrt{n})$	0.01 s	0.32 s	10.0 s	5.27 m	2.78 h
Insert sort	$O(n^2)$	0.07 s	7.50 s	12.5 m	20.8 h	86.8 d
XY	$O(n^3)$	0.10 s	1.67 m	1.16 d	3.17 L	3.17 kL

Za interaktivno uporabo na velikih podatkih so že kvadratični algoritmi,  $O(n^2)$ , prezahtevni.

Pri algoritmih eksponentne zahtevnosti, npr.  $O(2^n)$ , časovne zahteve naraščajo izjemno hitro. Recimo, da za  $n = 20$  potrebujemo 1 s, tedaj za  $n = 30$  potrebujemo  $2^{10}$  s = 1024 s = 17 m; za  $n = 40$  dobimo 1048576 s = 12.14 d; za  $n = 50$  že 1073741824 s = 34.05 L; ... za  $n = 80$  pa z  $36.56 \cdot 10^9$  L presežemo domnevno starost Zemlje.

## Pajek

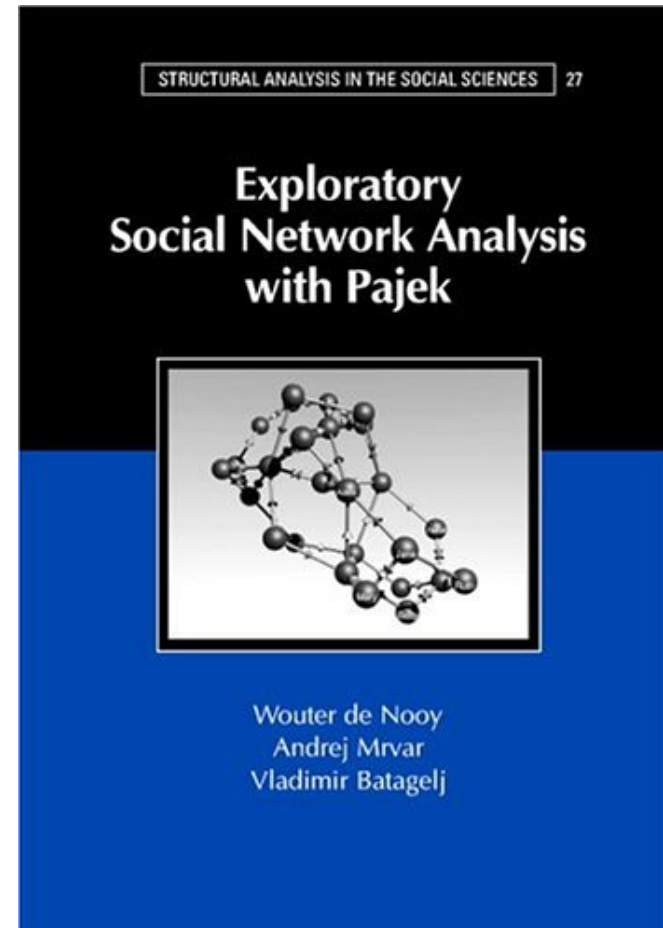


**Pajek** je program za okolje Windows za analizo in prikaz *velikih omrežij*. Z Andrejem Mrvarjem sva ga začela razvijati leta 1996.

Knjiga: W. de Nooy, A. Mrvar, V. Batagelj: *Exploratory Social Network Analysis with Pajek*, CUP, 2005. [Amazon](#). [stran ESNA](#).

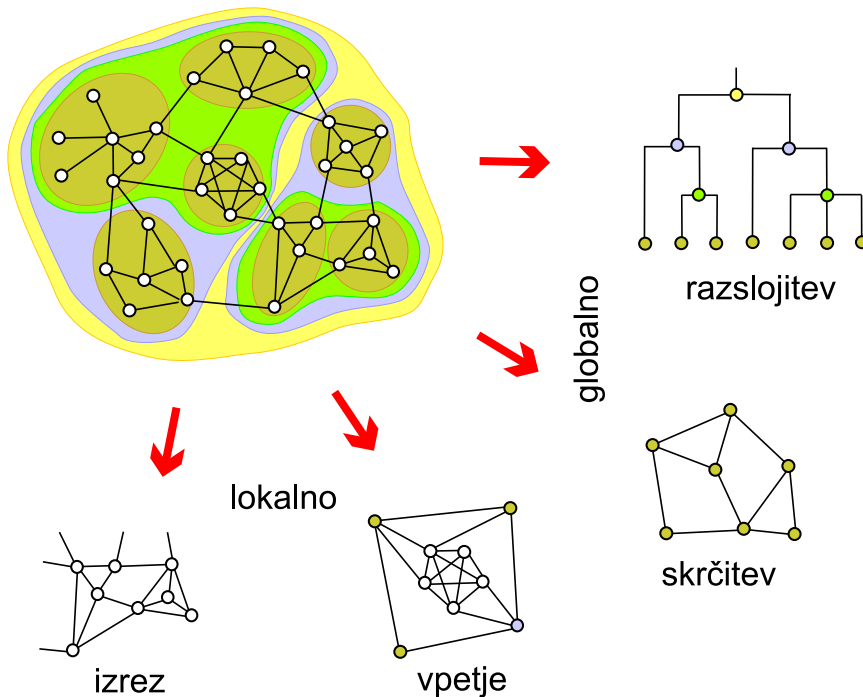
Zadnja izdaja programa **Pajek** je prosto dostopna (za netržne uporabe) na:

<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>





## Pa jek in velika omrežja



Glavni cilji v zasnovi programa **Pa jek** so:

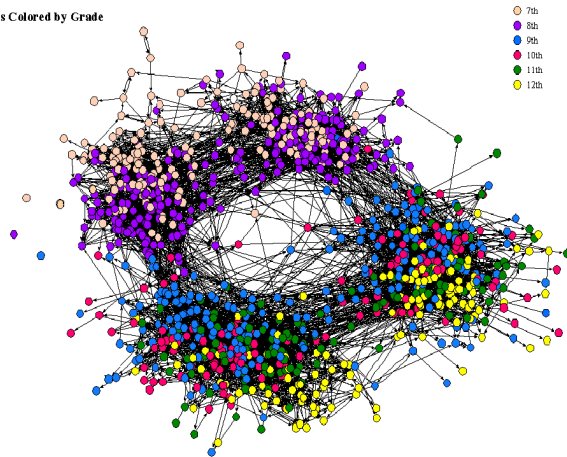
- podpora abstrakciji z (rekurzivnim) *razčlenjevanjem* velikega omrežja v več manjših, ki jih lahko v nadaljevanju posamično obdelamo z zahtevnejšimi postopki;
- ponuditi uporabniku izbor orodij za *prikaz* omrežij;
- izbrati/razviti nabor učinkovitih *podkvadratičnih* algoritmov za analizo velikih omrežij.

S **Pa jekom** lahko: *določimo* zanimive skupine (komponente, soseščine 'pomembnih' točk, sredice, ...) v omrežju, *izrežemo* del omrežja, ki pripada izbranim skupinam in ga ločeno *prikažemo*; lahko s 'povzetkom vpetosti' v celotno omrežje (lokalni pogled), *skrčimo* točke, ki pripadajo isti skupini, in prikažemo odnose med skupinami (globalni pogled).

## Prikaz omrežij

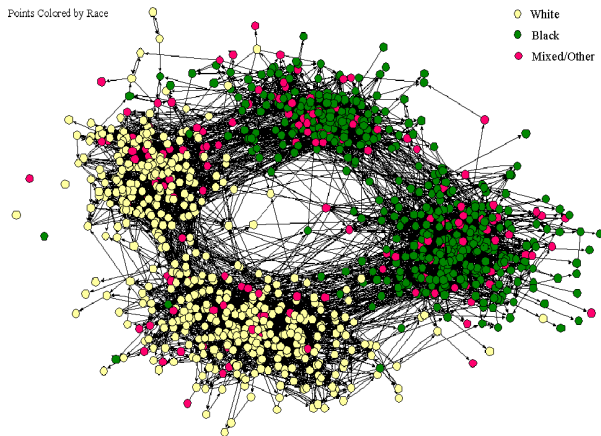
The Social Structure of "Countryside" School District

Points Colored by Grade



The Social Structure of "Countryside" School District

Points Colored by Race



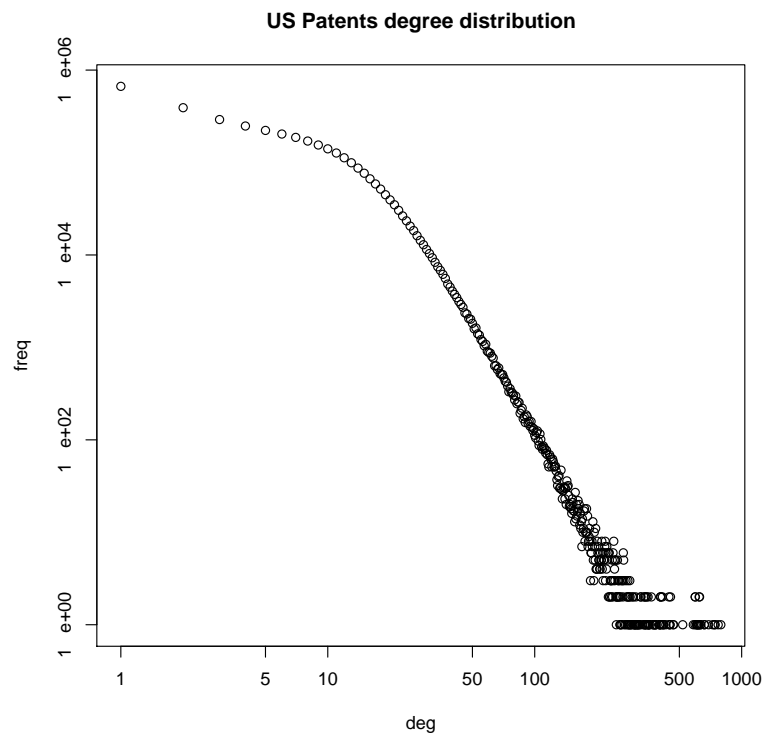
Pomembna sestavina analize omrežij so prikazi (delov) omrežij.

Prikaze lahko ustvarimo programsko (vzmetne vložitve) ali ročno.

Za posebne vrste omrežij obstajajo posebni načini prikazov. Npr. aciklična omrežja lahko prikažemo po ravneh.

Prikažemo lahko tudi izbrane lastnosti točk in povezav. Primer: šolarji (**Moody**).

## Pristopi k analizi velikih omrežij



Običajni pristop k analizi velikih podatkovij je *statistika*. V zadnjih desetih letih deležni velike pozornosti brezlestvični (scale free) verjetnostni modeli omrežij.

Druga pot so postopki za določanje *zanimivih delov* omrežij in razkrivanje *zgradbe* omrežja. V tem sestavku bomo prikazali pomembnejše izmed njih.

## Vmesnost

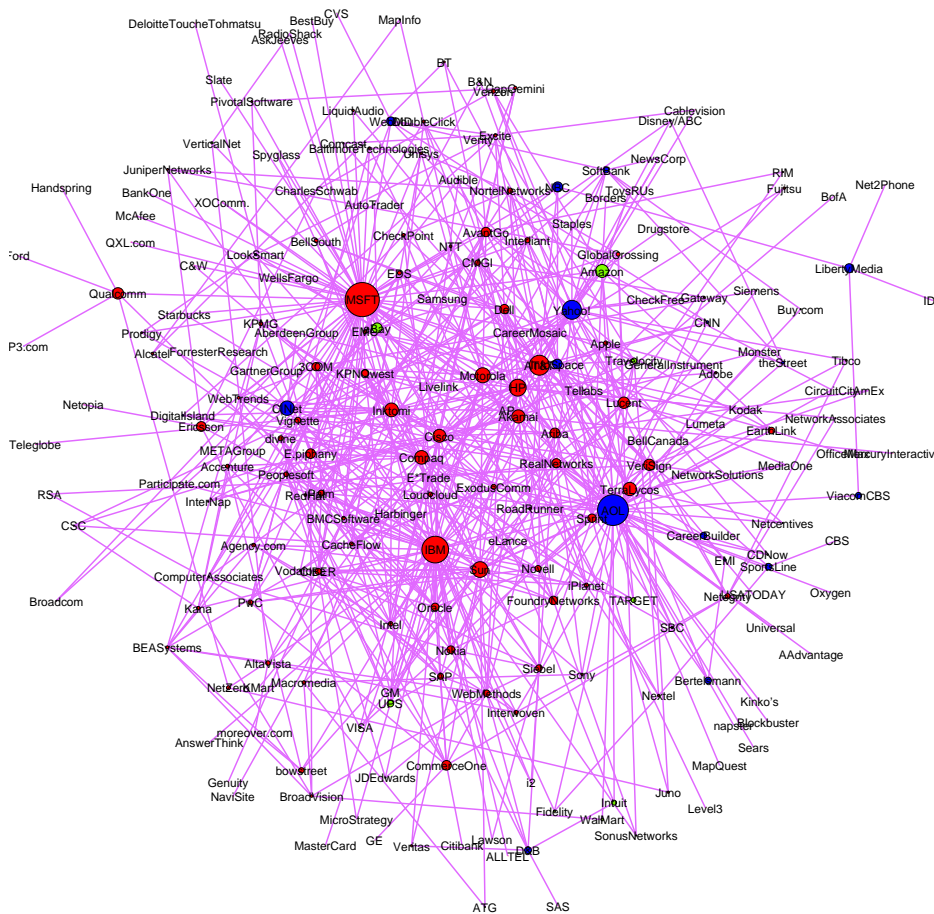
Obstaja cela vrsta mer pomembnosti točk, ki številsko izrazijo vlogo točke v omrežju. Kot primer take mere si pogledjmo vmesnost.

Pomembne so točke, ki lahko nadzirajo pretok podatkov po omrežju. Če privzamemo, da so za prenos pomembne le najkrajše poti, dobimo kot mero *vmesnosti* (Anthonisse 1971, Freeman 1977)

$$b(v) = \frac{1}{(n-1)(n-2)} \sum_{\substack{u,t \in \mathcal{V}: g_{u,t} > 0 \\ u \neq v, t \neq v, u \neq t}} \frac{g_{u,t}(v)}{g_{u,t}}$$

kjer je  $g_{u,t}$  število najkrajših poti iz  $u$  v  $t$ ; in  $g_{u,t}(v)$  število takih med njimi, ki gredo skozi točko  $v$ .

## Krebsova Internetska podjetja – vmesnost



Omrežje sestavljajo izbrana  
Internetska podjetja v obdobju  
1998 do 2001.

$n = 219$ ,  $m = 631$ .

rdeča – vsebina,  
modra – podpora,  
zelena – trgovina.

Podjetji sta povezani, če sta  
objavili skupni posel ali sode-  
lovanje.

Spletni naslov: [orgnet](http://orgnet)

## Kazala in vsebine

Točkam povezanega usmerjenega omrežja  $\mathcal{N} = (\mathcal{V}, \mathcal{L})$  priredimo dve vrednosti: kakovost vsebine  $x_v$  in kakovost kazala  $y_v$  (Kleinberg, 1998).

Na dobro vsebino kažejo dobra kazala in dobro kazalo kaže na dobre vsebine

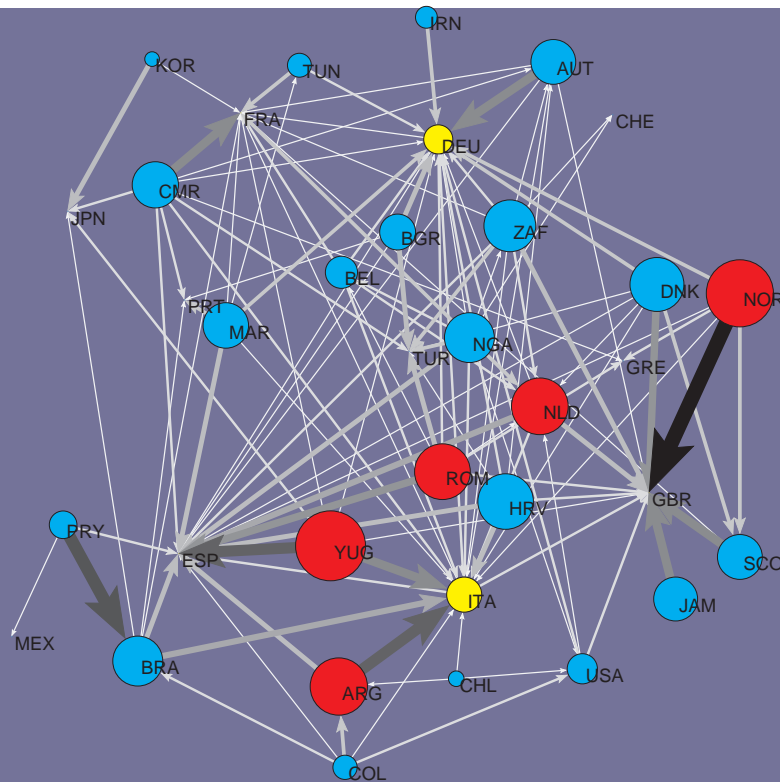
$$x_v = \sum_{u:(u,v) \in \mathcal{L}} y_u \quad \text{in} \quad y_v = \sum_{u:(v,u) \in \mathcal{L}} x_u$$

Naj bo  $\mathbf{W}$  matrika omrežja  $\mathcal{N}$  in  $\mathbf{x}$  ter  $\mathbf{y}$  vektorja obeh lastnosti. Tedaj lahko zvezi zapišemo  $\mathbf{x} = \mathbf{W}^T \mathbf{y}$  oziroma  $\mathbf{y} = \mathbf{W} \mathbf{x}$ .

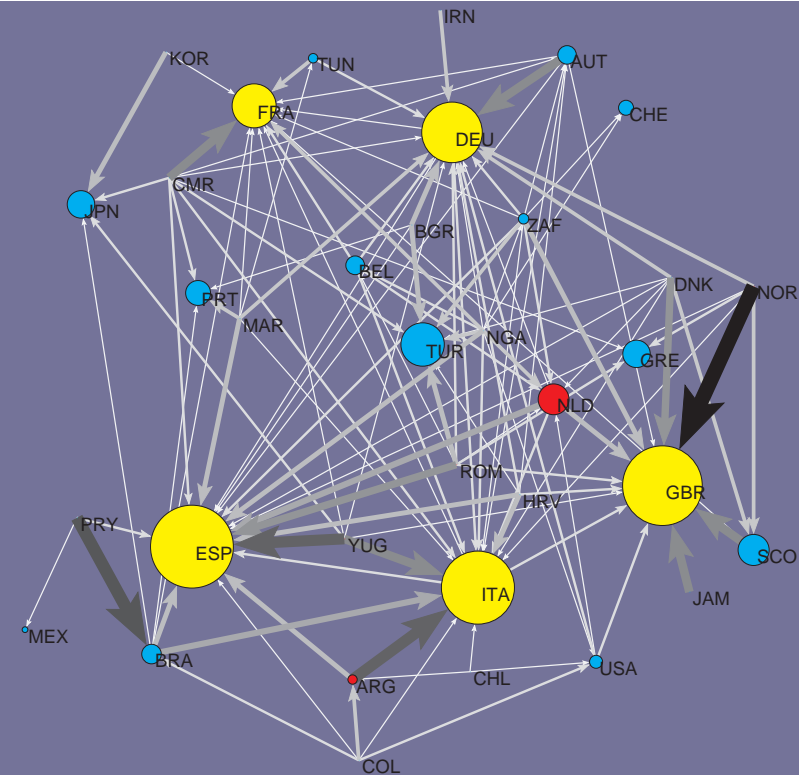
Začnimo z  $\mathbf{y} = [1, 1, \dots, 1]$  in nato zaporedoma izračunamo po obeh zvezah nove približke za  $\mathbf{x}$  in  $\mathbf{y}$ . Oba vektorja po vsakem koraku normaliziramo. To ponavljamo dokler se vektorja ne ustalita. Pokazati je mogoče, da opisani postopek konvergira. Limitni vektor  $\mathbf{x}^*$  je glavni lastni vektor matrike  $\mathbf{W}^T \mathbf{W}$ ;  $\mathbf{y}^*$  pa matrike  $\mathbf{W} \mathbf{W}^T$ .

PageRank, PageRank / Google, HITS / AltaVista, SALSA, teorija.

## ... Kazala in vsebine: nogometaši



Izvozniki (kazala/hubs)

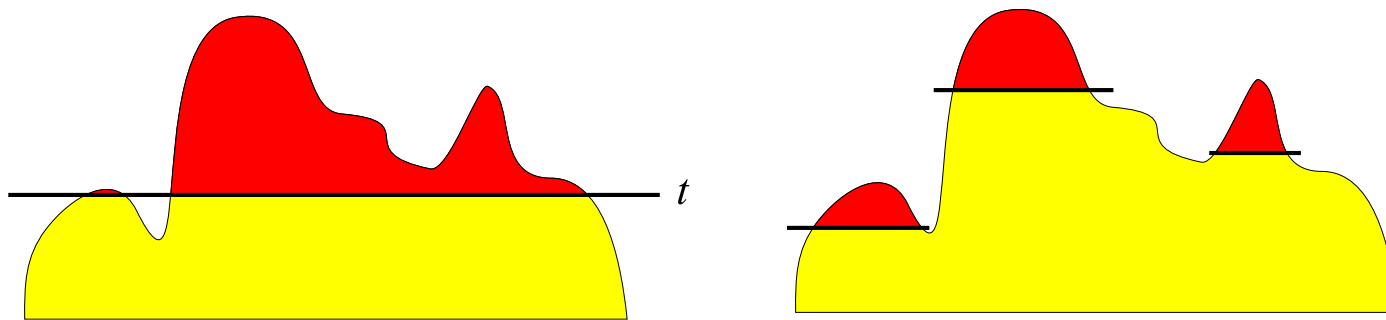


Uvozniki (vsebine/authorities)

Na svetovnem nogometnem prvenstvu v Parizu leta 1998 je sodelovalo 22 nogometnih reprezentanc. V omrežju so vse države, iz katerih so nogometaši igrali v ligah teh 22 držav, in vse države, v katerih ligah so igrali nogometaši iz teh 22 držav. Relacija je *igralec iz države x igra v državi y*; utež je število takih igralcev. Podatke je zbral Lothar Krempel. [football.net](http://football.net)

## Otoki

Če dano lastnost/utež točk/povezav predstavimo kot njihovo višino, nam naše omrežje določa nekakšno pokrajino s hribi in dolinami. Če to pokrajino potopimo v vodo do izbrane višine, dobimo *otoke*. S spreminjanjem višine vode dobivamo različne otoke. V uprabah nas običajno zanimajo ne preveliki in ne premajhni otoki – le otoki velikosti med izbranima  $k$  in  $K$ . Postopek 'otoki' začne s pokrajino popolnoma potopljeno v vodo. Nato znižujemo višino, dokler se ne pojavi otok prave velikosti ...

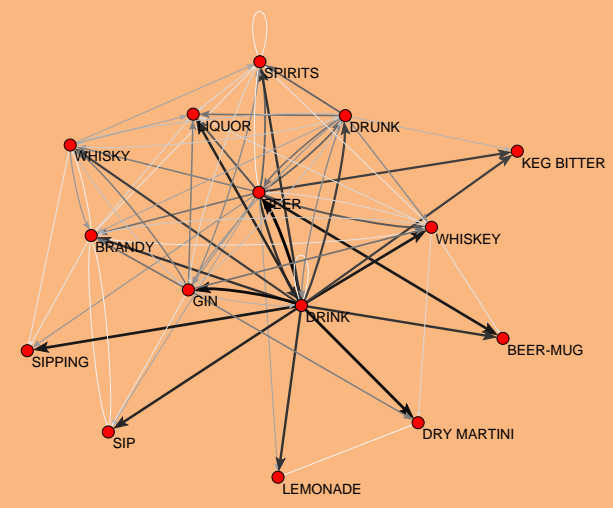
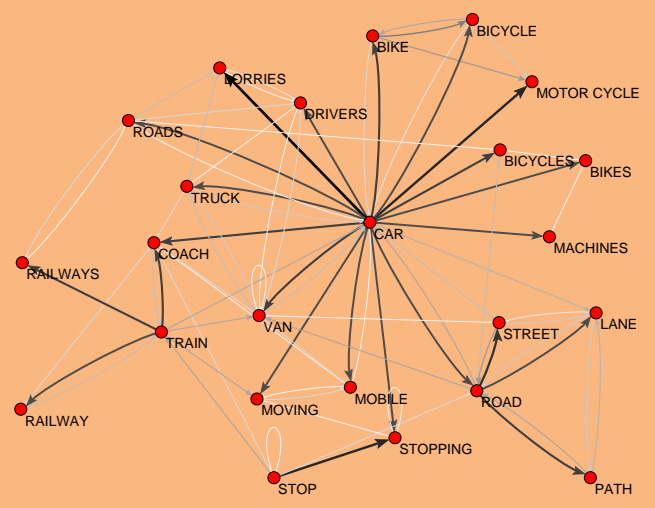
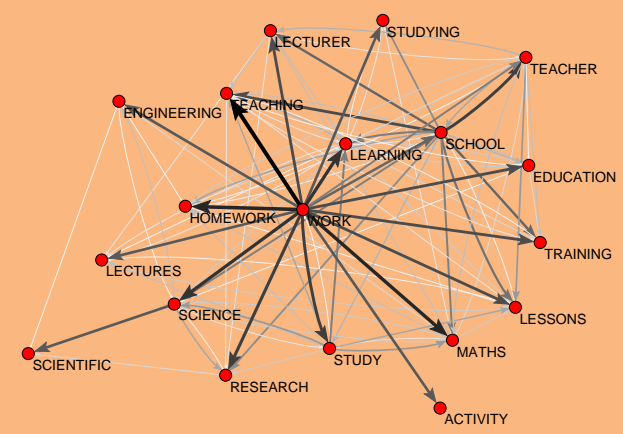
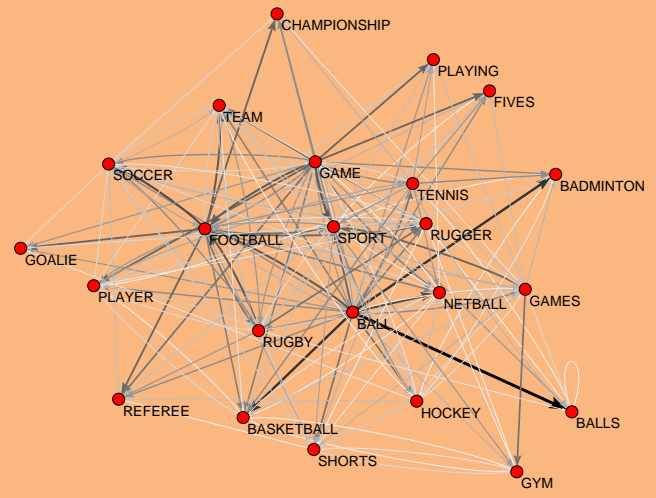


Otoki so zelo splošen in učinkovit postopek za določanje 'pomembnih' podomrežij v danem omrežju.

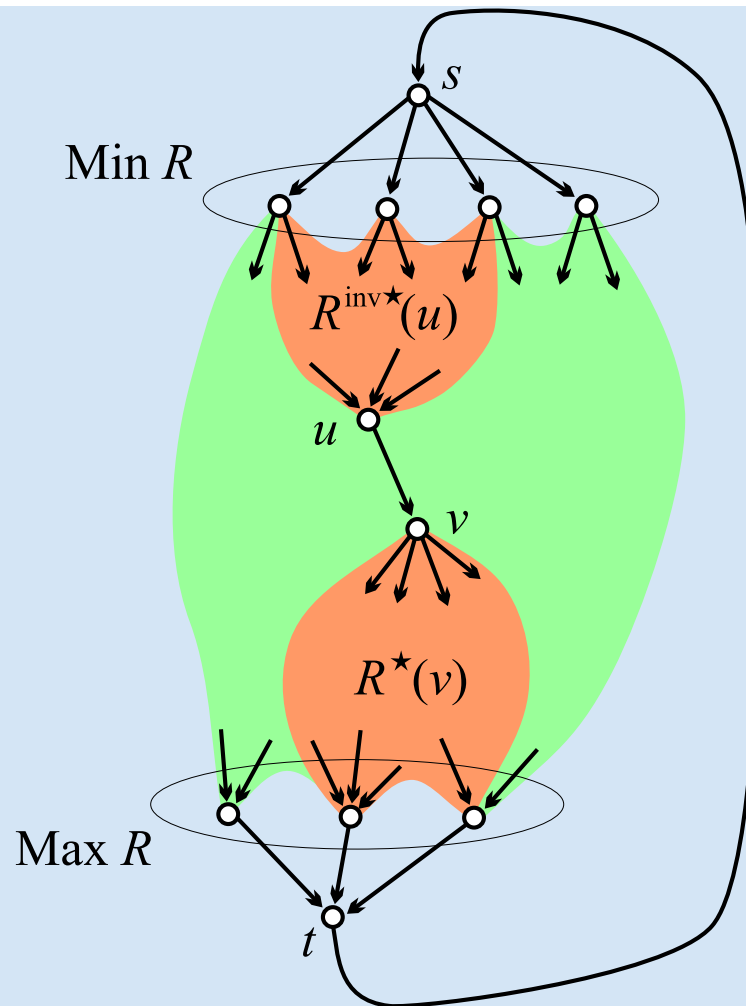


# Otoki – The Edinburgh Associative Thesaurus

$n = 23219, m = 325624$ , transitivity weight



## Štetje poti – Search path count method



Metoda *štetja poti* – *search path count method* (SPC) temelji na števcih  $n(u, v)$ , ki štejejo število različnih poti iz izvora  $s$  v ponor  $t$ , ki gredo čez povezavo  $(u, v)$ .

Hummon in Doreian uteži so takole določene:

- *search path link count* (SPLC) method: utež  $w_l(u, v)$  je enaka številu “vseh možnih poti po omrežju, ki imajo začetek v izvoru” in gredo po povezavi  $(u, v) \in \mathbf{R}$ .
- *search path node pair* (SPNP) method:  $w_p(u, v)$  “upoštevava vse povezanosti parov točk  $s$  potmi, ki gredo po povezavi  $(u, v) \in \mathbf{R}$ ”.

## Lastnosti uteži SPC

Vrednosti števecov  $n(u, v)$  določajo tok po omrežju sklicevanj – zanj velja *Kirchoffov točkovni zakon*: v vsaki točki  $u$  standardnega omrežja sklicevanj velja *vstopni tok = izstopni tok*:

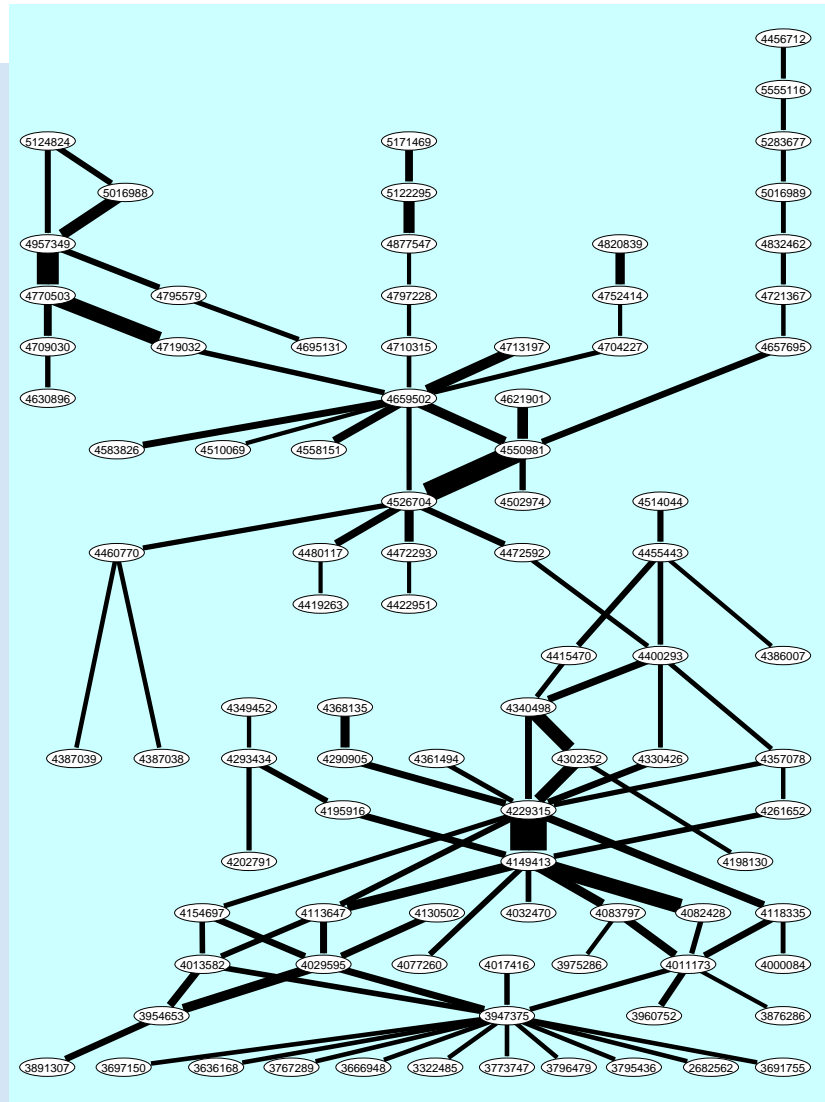
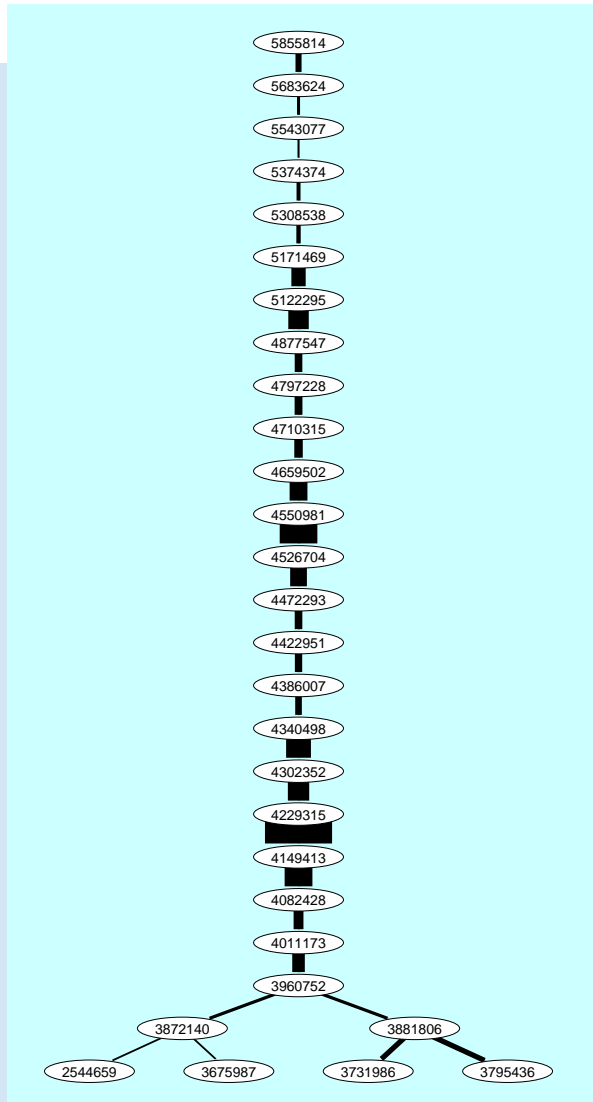
$$\sum_{v: v \mathbf{R} u} n(v, u) = \sum_{v: u \mathbf{R} v} n(u, v) = n^-(u) \cdot n^+(u)$$

Utež  $n(t, s)$  je enaka celotnemu toku skozi omrežje in ponuja naravni način normalizacije uteži

$$w(u, v) = \frac{n(u, v)}{n(t, s)} \Rightarrow 0 \leq w(u, v) \leq 1$$

in, če je  $C$  minimalni povezavni prerez, velja še  $\sum_{(u,v) \in C} w(u, v) = 1$ .

# Glavna pot in glavni otok v US Patents



# Glavni otok – Liquid crystal display

Table 1: Patents on the liquid-crystal display

patent	date	author(s) and title
2544659	Mar 13, 1951	Dreyer. Dichroic light-polarizing sheet and the like and the formation and use thereof
2682562	Jun 29, 1954	Wender, et al. Reduction of aromatic carbinols
3322485	May 30, 1967	Williams. Electro-optical elements utilizing an organic nematic compound
3636168	Jan 18, 1972	Josephson. Preparation of polynuclear aromatic compounds
3666948	May 30, 1972	Mechlowitz, et al. Liquid crystal thermal imaging system having an undisturbed image on a disturbed background
3675987	Jul 11, 1972	Rafuse. Liquid crystal compositions and devices
3691755	Sep 19, 1972	Girard. Clock with digital display
3697150	Oct 10, 1972	Wysocki. Electro-optic systems in which an electrophoretic-like or dipolar material is dispersed throughout a liquid crystal to reduce the turn-off time
3731986	May 8, 1973	Ferguson. Display devices utilizing liquid crystal light modulation
3767289	Oct 23, 1973	Aviram, et al. Class of stable trans-stilbene compounds, some displaying nematic mesophases at or near room temperature and others in a range up to 100°C
3773747	Nov 20, 1973	Steinstrasser. Substituted azoxy benzene compounds
3795436	Mar 5, 1974	Boller, et al. Nematogenic material which exhibit the Kerr effect at isotropic temperatures
3796479	Mar 12, 1974	Helfrich, et al. Electro-optical light-modulation cell utilizing a nematogenic material which exhibits the Kerr effect at isotropic temperatures
3872140	Mar 18, 1975	Klanderaman, et al. Liquid crystalline compositions and method
3876286	Apr 8, 1975	Deutscher, et al. Use of nematic liquid crystalline substances
3881806	May 6, 1975	Suzuki. Electro-optical display device
3891307	Jun 24, 1975	Tsakamoto, et al. Phase control of the voltages applied to opposite electrodes for a cholesteric to nematic phase transition display
3947375	Mar 30, 1976	Gray, et al. Liquid crystal materials and devices
3954653	May 4, 1976	Yamazaki. Liquid crystal composition having high dielectric anisotropy and display device incorporating same
3960752	Jun 1, 1976	Klanderaman, et al. Liquid crystal compositions
3975286	Aug 17, 1976	Oh. Low voltage actuated field effect liquid crystals compositions and method of synthesis
4000084	Dec 28, 1976	Hsieh, et al. Liquid crystal mixtures for electro-optical display devices
4011173	Mar 8, 1977	Steinstrasser. Modified nematic mixtures with positive dielectric anisotropy
4013582	Mar 22, 1977	Gavrilovic. Liquid crystal compounds and electro-optic devices incorporating them
4017416	Apr 12, 1977	Inukai, et al. P-cyanophenyl 4-alkyl-4'-biphenylcarboxylate, method for preparing same and liquid crystal compositions using same
4029595	Jun 14, 1977	Rees, et al. Novel liquid crystal compounds and electro-optic devices incorporating them
4032470	Jun 28, 1977	Bloom, et al. Electro-optic device
4077260	Mar 7, 1978	Gray, et al. Optically active cyano-biphenyl compounds and liquid crystal materials containing them
4082428	Apr 4, 1978	Hsu. Liquid crystal composition and method

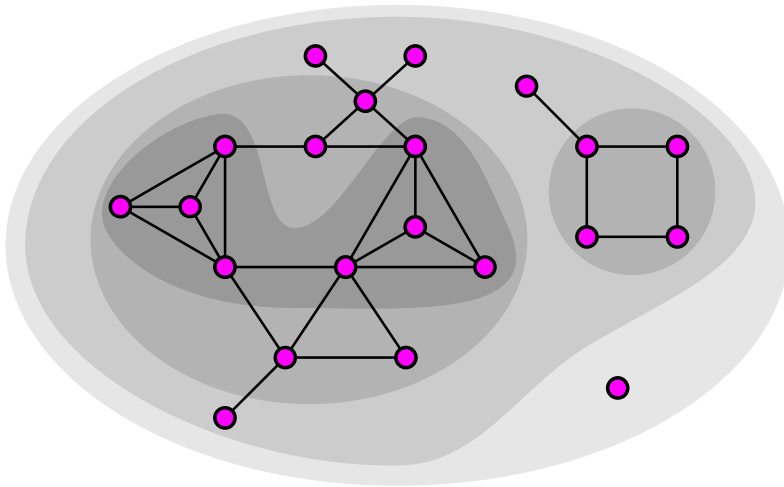
Table 2: Patents on the liquid-crystal display

patent	date	author(s) and title
4083797	Apr 11, 1978	Oh. Nematic liquid crystal compositions
4113647	Sep 12, 1978	Coates, et al. Liquid crystalline materials
4118335	Oct 3, 1978	Krause, et al. Liquid crystalline materials of reduced viscosity
4130502	Dec 19, 1978	Eidenschink, et al. Liquid crystalline cyclohexane derivatives
4149413	Apr 17, 1979	Gray, et al. Optically active liquid crystal mixtures and liquid crystal devices containing them
4154697	May 15, 1979	Eidenschink, et al. Liquid crystalline hexahydroterphenyl derivatives
4195916	Apr 1, 1980	Coates, et al. Liquid crystal compounds
4198130	Apr 15, 1980	Boller, et al. Liquid crystal mixtures
4202791	May 13, 1980	Sato, et al. Nematic liquid crystalline materials
4229315	Oct 21, 1980	Krause, et al. Liquid crystalline cyclohexane derivatives
4261652	Apr 14, 1981	Gray, et al. Liquid crystal compounds and materials and devices containing them
4290905	Sep 22, 1981	Kanbe. Ester compound
4293434	Oct 6, 1981	Deutscher, et al. Liquid crystal compounds
4302352	Nov 24, 1981	Eidenschink, et al. Fluorophenylcyclohexanes, the preparation thereof and their use as components of liquid crystal dielectrics
4330426	May 18, 1982	Eidenschink, et al. Cyclohexylbiphenyls, their preparation and use in dielectrics and electrooptical display elements
4340498	Jul 20, 1982	Suginori. Halogenated ester derivatives
4349452	Sep 14, 1982	Osman, et al. Cyclohexylcyclohexanoates
4357078	Nov 2, 1982	Carr, et al. Liquid crystal compounds containing an alicyclic ring and exhibiting a low dielectric anisotropy and liquid crystal materials and devices incorporating such compounds
4361494	Nov 30, 1982	Osman, et al. Anisotropic cyclohexyl cyclohexylmethyl ethers
4368135	Jan 11, 1983	Osman. Anisotropic compounds with negative or positive DC-anisotropy and low optical anisotropy
4386007	May 31, 1983	Krause, et al. Liquid crystalline naphthalene derivatives
4387038	Jun 7, 1983	Fukui, et al. 4-(Trans-4'-alkylcyclohexyl) benzoic acid 4"-cyano-4"-biphenyl esters
4387039	Jun 7, 1983	Suginori, et al. Trans-4-(trans-4'-alkylcyclohexyl)-cyclohexane carboxylic acid 4"-cyanobiphenyl ester
4400293	Aug 23, 1983	Romer, et al. Liquid crystalline cyclohexylphenyl derivatives
4415470	Nov 15, 1983	Eidenschink, et al. Liquid crystalline fluorine-containing cyclohexylbiphenyls and dielectrics and electro-optical display elements based thereon
4419263	Dec 6, 1983	Praefcke, et al. Liquid crystalline cyclohexylcarbonitrile derivatives
4422951	Dec 27, 1983	Suginori, et al. Liquid crystal benzene derivatives
4455443	Jun 19, 1984	Takatsui, et al. Nematic halogen Compound
4456712	Jun 26, 1984	Christie, et al. Bismaleimide triazine composition
4460770	Jul 17, 1984	Petrzalka, et al. Liquid crystal mixture
4472293	Sep 18, 1984	Suginori, et al. High temperature liquid crystal substances of four rings and liquid crystal compositions containing the same
4472592	Sep 18, 1984	Takatsui, et al. Nematic liquid crystalline compounds
4480117	Oct 30, 1984	Takatsui, et al. Nematic liquid crystalline compounds
4502974	Mar 5, 1985	Suginori, et al. High temperature liquid-crystalline ester compounds
4510069	Apr 9, 1985	Eidenschink, et al. Cyclohexane derivatives

Table 3: Patents on the liquid-crystal display

patent	date	author(s) and title
4514044	Apr 30, 1985	Gunjima, et al. 1-(Trans-4-alkylcyclohexyl)-2-(trans-4'-(p-substituted phenyl) cyclohexyl)ethane and liquid crystal mixture
4526704	Jul 2, 1985	Petrzalka, et al. Multiring liquid crystal esters
4550981	Nov 5, 1985	Petrzalka, et al. Liquid crystalline esters and mixtures
4558151	Dec 10, 1985	Takatsui, et al. Nematic liquid crystalline compounds
4583826	Apr 22, 1986	Petrzalka, et al. Phenylethanes
4621901	Nov 11, 1986	Petrzalka, et al. Novel liquid crystal mixtures
4630896	Dec 23, 1986	Petrzalka, et al. Benzotrioles
4657695	Apr 14, 1987	Saito, et al. Substituted pyridazines
4659502	Apr 21, 1987	Fearon, et al. Ethane derivatives
4695131	Sep 22, 1987	Balkwill, et al. Disubstituted ethanes and their use in liquid crystal materials and devices
4704227	Nov 3, 1987	Krause, et al. Liquid crystal compounds
4709030	Nov 24, 1987	Petrzalka, et al. Novel liquid crystal mixtures
4710315	Dec 1, 1987	Schad, et al. Anisotropic compounds and liquid crystal mixtures therewith
4713197	Dec 15, 1987	Eidenschink, et al. Nitrogen-containing heterocyclic compounds
4719032	Jan 12, 1988	Wachtler, et al. Cyclohexane derivatives
4721367	Jan 26, 1988	Yoshinaga, et al. Liquid crystal device
4752414	Jun 21, 1988	Eidenschink, et al. Nitrogen-containing heterocyclic compounds
4770503	Sep 13, 1988	Buechecker, et al. Liquid crystalline compounds
4795579	Jan 3, 1989	Vauchier, et al. 2,2'-difluoro-4-alkoxy-4'-hydroxydiphenyls and their derivatives, their production process and their use in liquid crystal display devices
4797228	Jan 10, 1989	Goto, et al. Cyclohexane derivative and liquid crystal composition containing same
4820839	Apr 11, 1989	Krause, et al. Nitrogen-containing heterocyclic esters
4832462	May 23, 1989	Clark, et al. Liquid crystal devices
4877547	Oct 31, 1989	Weber, et al. Liquid crystal display element
4957349	Sep 18, 1990	Clerc, et al. Active matrix screen for the color display of television pictures, control system and process for producing said screen
5016988	May 21, 1991	Imura. Liquid crystal display device with a birefringent compensator
5016989	May 21, 1991	Okada. Liquid crystal element with improved contrast and brightness
5122295	Jun 16, 1992	Weber, et al. Matrix liquid crystal display
5124824	Jun 23, 1992	Kozaki, et al. Liquid crystal display device comprising a retardation compensation layer having a maximum principal refractive index in the thickness direction
5171469	Dec 15, 1992	Hittich, et al. Liquid-crystal matrix display
5283677	Feb 1, 1994	Sagawa, et al. Liquid crystal display with ground regions between terminal groups
5308538	May 3, 1994	Weber, et al. Supertwist liquid-crystal display
5374374	Dec 20, 1994	Weber, et al. Supertwist liquid-crystal display
5543077	Aug 6, 1996	Rieger, et al. Nematic liquid-crystal composition
5551116	Sep 10, 1996	Ishikawa, et al. Liquid crystal display having adjacent electrode terminals set equal in length
5683624	Nov 4, 1997	Sekiguchi, et al. Liquid crystal composition
5855814	Jan 5, 1999	Matsui, et al. Liquid crystal compositions and liquid crystal display elements

## Sredice in posplošene sredice

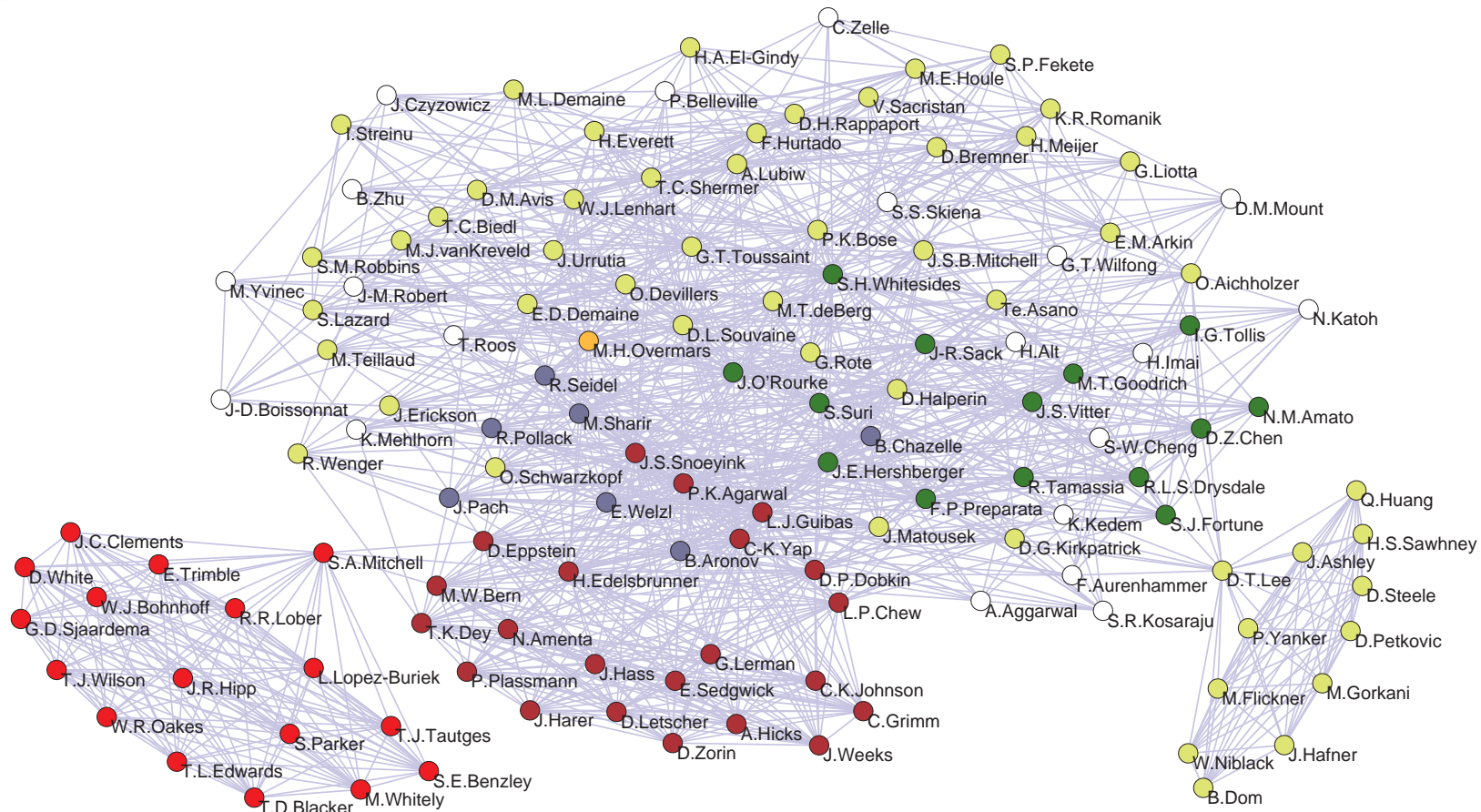


Pojem sredice je vpeljal Seidman leta 1983. Podgraf  $\mathcal{H} = (\mathcal{W}, \mathcal{L}(\mathcal{W}))$  grafa  $\mathcal{G} = (\mathcal{V}, \mathcal{L})$  porojen z množico  $\mathcal{W}$  je *k-sredica* ali *sredica reda k* ntk.  $\forall v \in \mathcal{W} : \deg_{\mathcal{H}}(v) \geq k$  in je  $\mathcal{H}$  maksimalni podgraf s to lastnostjo.

Sredici največjega reda v grafu pravimo tudi *glavna sredica*. *Sredičnost* točke  $v$  je enaka največjemu redu sredice, ki še vsebuje to točko.

Privzemimo, da imamo na  $\mathcal{N}$  določeno funkcijo  $p(v, \mathcal{U})$ ,  $v \in \mathcal{V}$ ,  $\mathcal{U} \subseteq \mathcal{V}$  z realnimi vrednostmi – *p-funkcijo*. Podgraf  $\mathcal{H} = (\mathcal{C}, \mathcal{L}(\mathcal{C}))$  porojen z množico  $\mathcal{C} \subseteq \mathcal{V}$  je *p-sredica* na ravni  $t \in \mathbb{R}$  ntk.  $\forall v \in \mathcal{C} : t \leq p(v, \mathcal{C})$  in je  $\mathcal{C}$  maksimalna taka množica.

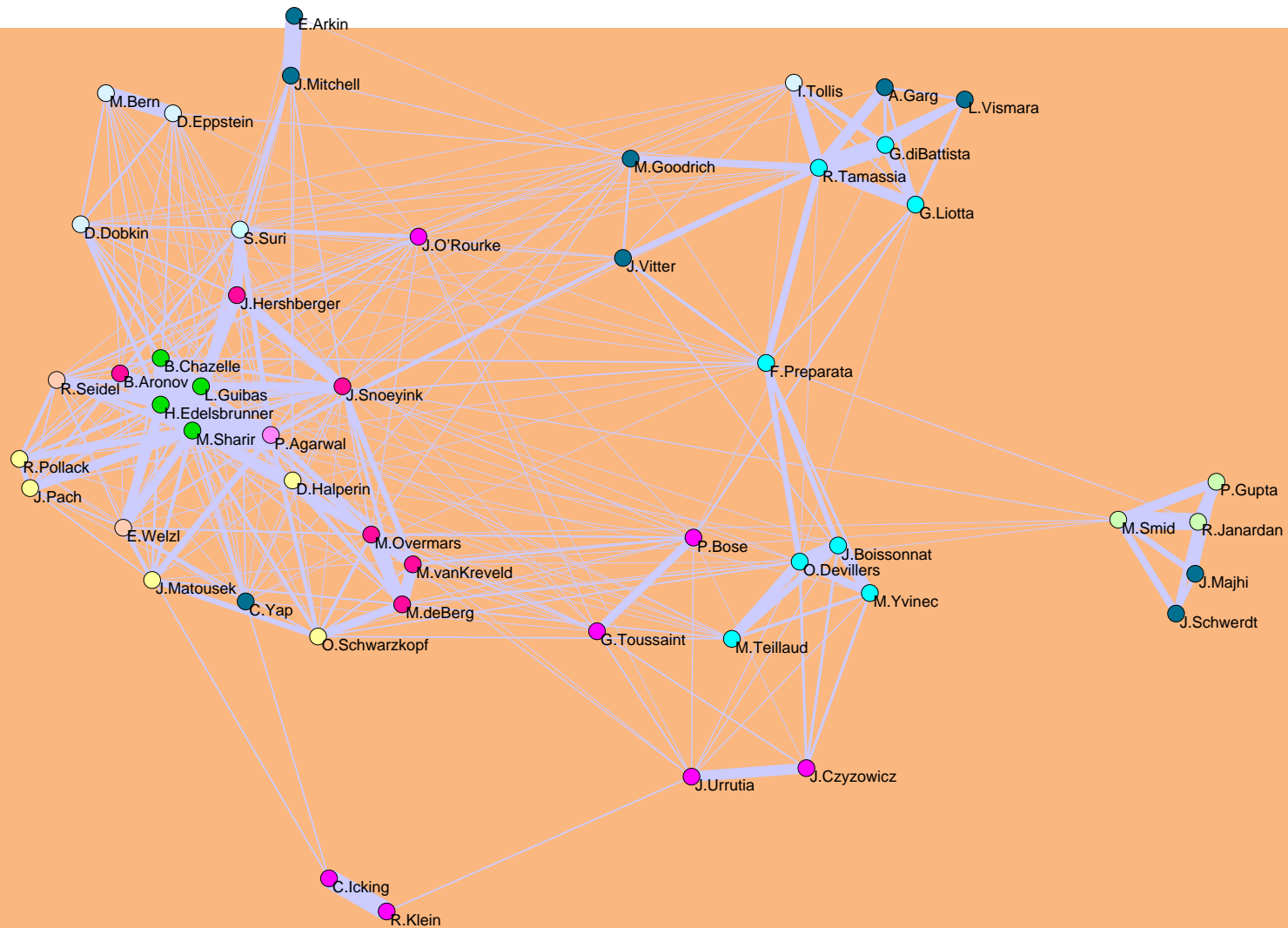
## GeomBib – navadne sredice reda 10



Iz B. Jones-ove bibliografija računalniške geometrije *Computational geometry database* (2002), *FTP*, *Geom.net*. smo sestavili omrežje sodelovanj. Avtorja določata povezavo, če sta napisala skupno delo. Utež povezave je enaka številu skupnih del.

$$n = 9072, m = 13567/22577 \rightarrow n' = 7343, m' = 11898.$$

# $p_S$ -sredica omrežja GeomBib na ravni 46





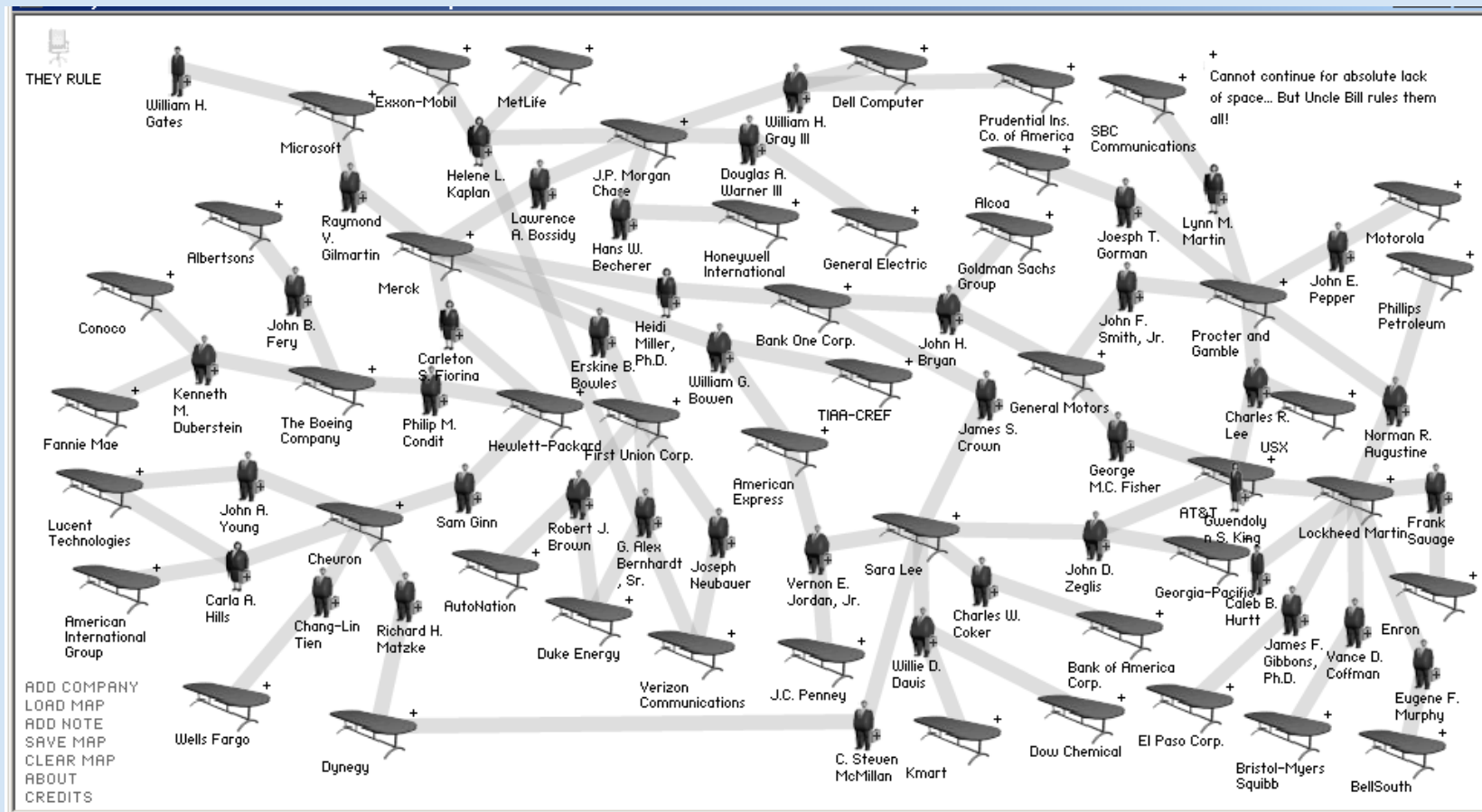
## Dvovrstna omrežja

*Dvovrstno omrežje* je omrežje  $\mathcal{N} = (\mathcal{U}, \mathcal{V}, \mathcal{A}, w)$ , v katerem je množica točk sestavljena iz dveh ločenih podmnožic  $\mathcal{U}$  in  $\mathcal{V}$  ter imajo povezave iz množice  $\mathcal{A}$  svoj začetek v množici  $\mathcal{U}$  in svoj konec v množici  $\mathcal{V}$ .  $w : \mathcal{A} \rightarrow \mathbb{R}$  je utež povezav. Če utež ni podana, postavimo  $w(u, v) = 1$  za vse povezave  $(u, v) \in \mathcal{A}$ . Množica povezav  $\mathcal{A}$  določa relacijo  $\mathcal{A} \subseteq \mathcal{U} \times \mathcal{V}$ .

Dvovrstno omrežje lahko predstavimo tudi s pravokotno matriko  $\mathbf{A} = [a_{uv}]_{\mathcal{U} \times \mathcal{V}}$ .

$$a_{uv} = \begin{cases} w(u, v) & (u, v) \in \mathcal{A} \\ 0 & \text{sicer} \end{cases}$$

## Omrežje z They Rule



# Omrežja iz podatkovnih tabel

RuthDELmain.csv																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N		
1	Ident	Num	File	ORGANISATION OR	ORG	Org	Contact Name	Street	ZIP	Project	City	Country	coun	EU Region		
2	1	1480	613.html	3D PLUS SA	3D F3D	LIGNIER, Olivier	641 Ru	78530	IST-2001-3440	Buc	FRANCE	20	2	ÎLE DE FF		
3	2	1481	613.html	3D PLUS SA	3D PLUS	LIGNIER, Olivier	641 Ru	78530	IST-2001-3440	Buc	FRANCE	20	2	ÎLE DE FF		
4	3	4001	924.html	3D VISION	3D V\3D	MARIAT, Jacques	Savoie	73375		502909	Le Bc	FRANCE	20	2	CENTRE-I	
5	4	1648	160.html	3D Web Technologies	3D WEB	DENNISON, Andrew		M31 4XL	BMH4989519		Carrir	UNITED KI	60	2	NORTH W	
6	5	1406	442.html	3E	3E	PALMERS, Geer	Eredier	1000	NNE5/51/1999		Bruxe	BELGIQUE	8	2	REG.BRU	
7	6	1007	884.html	4M2C PATRIC SALON	4M2C PA	N/A		CRANA	12157		507255	Berlin	DEUTSCH	15	2	BERLIN E
8	7	7914	991.html	5T S.c.r.l.	5T S.C.	RN/A		C.so B	10126	Road2/506716	Torinc	ITALIA	26	2	NORD OV	
9	8	6880	588.html	A & C 2000 S.R.L.	A & C	SAN CARLUCCI, Renz	VIALE		148	IST-2001-3454	Roma	ITALIA	26	2	LAZIO Rc	
10	9	6881	588.html	A & C 2000 S.R.L.	A & C 20	CARLUCCI, Renz	Viale C		148	IST-2001-3454	Roma	ITALIA	26	2	LAZIO Rc	
11	10	1647	176.html	A. BENETTI MACCHIA	A. BENE	Federico BENETTI	Via Pro		54033	BRST985466	Carra	ITALIA	26	2	CENTRO	
12	11	6605	984.html	A. Mickiewicz Univers	A. MInst	PATKOWSKI, Ad	Ul. H.	161-712		502235	Pozn	POLSKA	45	2		
13	12	6571	135.html	A. BRITO - INDUSTRIA	A. BRITO	VIEIRA DE BRITC	5109, E	4350-119	BRST985263		Porto	PORTUGA	46	2	CONTINEI	
14	13	1813	409.html	A.L. DIGITAL LIMITED	A.L. A.L.	LAURIE, Ben	VOYSE	W4 4GB	IST-2000-2633		Chisv	UNITED KI	60	2	SOUTH E.	
15	14	1814	409.html	A.L. Digital Limited	A.L. DIG	LAURIE, Ben	Voysey	W4 4GB	IST-2000-2633		Chisv	UNITED KI	60	2	SOUTH E.	
16	15	1885	960.html	A.P. MOLLER-MAER	A.P. TEC	DRAGSTED, Jorr	Esplan	1098		506676	Kope	DANMARK	14	2	Københavi	
17	16	6731	537.html	A.S.M. S.A.	A.S.M.	SMOYA GARCIA, J	Carrete	43206	IST-2000-3008		Reus	ESPAÑA	19	2	ESTE CA	
18	17	8150	232.html	AABO AKADEMI UNIV	AAB CO	NYBACKA-WILLM	14-18B	20500	ERK5-CT-1999		Turku	SUOMI/FIN	53	2	MANNER-	
19	18	8152	662.html	AABO AKADEMI UNIV	AAB DEF	BJORKSTRAND, J	3,Tykis	20521	EVK1-CT-2002		Turku	SUOMI/FIN	53	2		
20	19	8148	959.html	AABO AKADEMI UNIV	AAB Dep	HUPA, Mikko	Domky	20500		502679	Turku	SUOMI/FIN	53	2	MANNER-	
21	20	8151	233.html	AABO AKADEMI UNIV	AAB DEF	NYBACKA-WILLM	Lemmi	20500	ERK6-CT-1999		Turku	SUOMI/FIN	53	2	MANNER-	
22	21	125	116.html	AACHEN UNIVERSIT	AAC GIE	E. NEUSSL	Intzest	52072	BRPR980663		Aach	DEUTSCH	15	2	NORDRHI	
23	22	123	104.html	AACHEN UNIVERSIT	AAC GIE	MEISER, Lukas	Intzest	52072	BRPR980695		Aach	DEUTSCH	15	2	NORDRHI	
24	23	155	364.html	AACHEN UNIVERSIT	AAC INS'	RAUHUT, Burkha	18,Eilfs	52062	G1RD-CT-2000		Aach	DEUTSCH	15	2	NORDRHI	

*Podatkovna tabela*  $\mathcal{T}$  je sestavljena iz množice *zapisov*  $\mathcal{T} = \{T_k : k \in \mathcal{K}\}$ , kjer je  $\mathcal{K}$  množica *ključev*. Posamezni zapis ima obliko  $T_k = (k, q_1(k), q_2(k), \dots, q_r(k))$  kjer je  $q_i(k)$  vrednost *lastnosti*  $q_i$  za ključ  $k$ .

## ... Omrežja iz podatkovnih tabel

Naj ima lastnost  $\mathbf{q}$  zalogo vrednosti  $\mathcal{Q}$ . Če je ta končna (to lahko vselej dosežemo z razbitjem na razrede), lahko lastnosti  $\mathbf{q}$  priredimo dvovrstno omrežje  $\mathcal{K} \times \mathbf{q} = (\mathcal{K}, \mathcal{Q}, \mathcal{E}, w)$  določeno z usmerjenimi povezavami  $(k, v) \in \mathcal{E}$  ntk.  $q(k) = v$ , z utežmi  $w(k, v) = 1$ . **txt2pajek** (Jürgen Pfeffer)

Dvovrstno omrežje  $\mathbf{q}_i \times \mathbf{q}_j = (\mathcal{Q}_i, \mathcal{Q}_j, \mathcal{E}, w)$  lahko definiramo tudi za lastnosti  $\mathbf{q}_i$  in  $\mathbf{q}_j$ , kjer je  $(u, v) \in \mathcal{E}$  natanko takrat, ko  $\exists k \in \mathcal{K} : (q_i(k) = u \wedge q_j(k) = v)$ , in je  $w(u, v) = \text{card} \{k \in \mathcal{K} : (q_i(k) = u \wedge q_j(k) = v)\}$ .

Naj bo še  $[\mathbf{q}_i \times \mathbf{q}_j]^T = \mathbf{q}_j \times \mathbf{q}_i$ .

Potem velja  $\mathbf{q}_i \times \mathbf{q}_j = [\mathcal{K} \times \mathbf{q}_i]^T * [\mathcal{K} \times \mathbf{q}_j] = [\mathbf{q}_i \times \mathcal{K}] * [\mathcal{K} \times \mathbf{q}_j]$ .

Par lastnosti  $\mathbf{q}_i$  in  $\mathbf{q}_j$  lahko združimo tudi glede na neko tretjo lastnost  $\mathbf{q}_s$ : dobimo dvovrstno omrežje  $[\mathbf{q}_i \times \mathbf{q}_j] / \mathbf{q}_s = [\mathbf{q}_i \times \mathbf{q}_s] * [\mathbf{q}_s \times \mathbf{q}_j]$ .

## Primer: Evropski projekti na temo simulacij

Za srečanje *The Age of Simulation* januarja 2006 na Ars Electronica v Linzu smo skupaj s sodelavci podjetja FAS z Dunaja analizirali podatke o evropskih projektih na temo simulacij. Podatke so zbrali sodelavci FASa s **spletišča projektov** in jih uredili v obliki velike tabele v Excelu. Posamezni zapis sestavljajo različni podatki o posamezni sodelujoči ustanovi na posameznem projektu. Tabelo smo najprej shranili v obliki CSV in nato s programom **Text2Pajek** iz nje ustvarili tri dvovrstna omrežja:

- `project.net` – `idents` × `projects` = **P**
- `country.net` – `idents` × `countries` = **C**
- `institution.net` – `idents` × `institutions` = **U**

Velikosti posameznih množic so naslednje:

$|idents| = 8869$ ,    $|projects| = 933$ ,    $|institutions| = 3438$ ,  
 $|countries| = 60$ .

## Evropski projekti – množenje omrežij

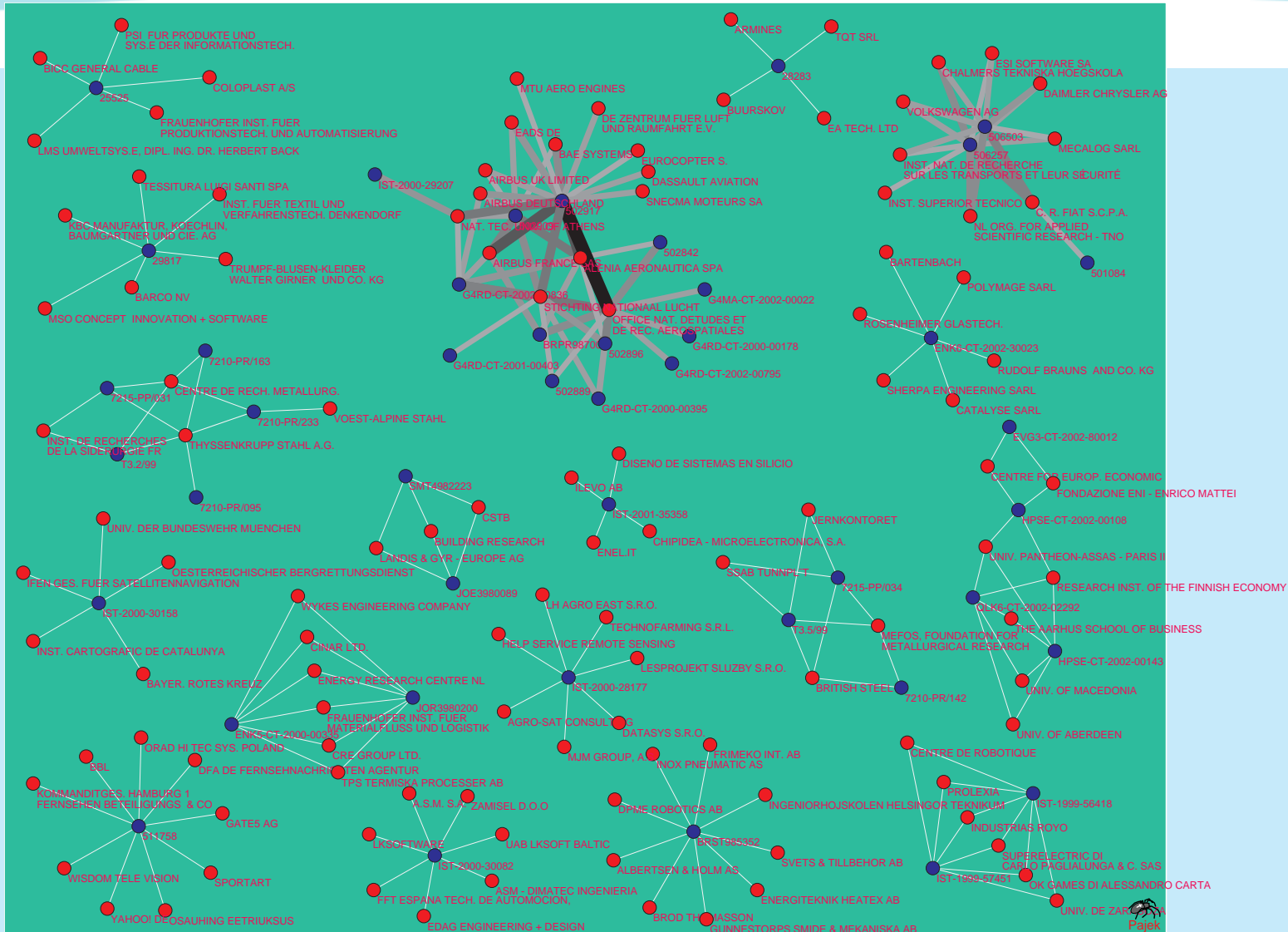
Ker imajo vsa tri omrežja skupno množico  $\mathcal{K}$  =idents, lahko na prej opisani način iz njih z množenjem pridobimo različna omrežja:

- ProjInst.net – projects  $\times$  institutions  $\mathbf{W} = \mathbf{P}^T \star \mathbf{U}$
- Countries.net – countries  $\times$  countries  $\mathbf{S} = \mathbf{C}^T \star \mathbf{C}$
- Institutions.net – institutions  $\times$  institutions  $\mathbf{Q} = \mathbf{W}^T \star \mathbf{W}$

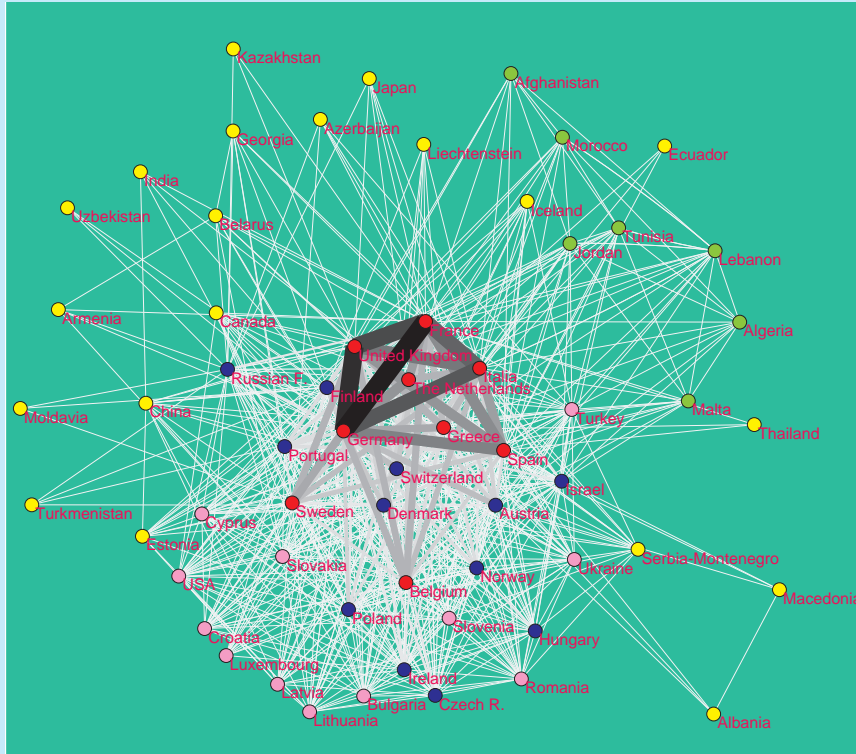
Za določitev pomembnih delov omrežja ProjInst.net smo najprej določili omrežje 4-obročnih uteži in na tem omrežju določili povezavne otoke:

Dobili smo 101 otok, 18 med njimi ima velikost vsaj 5 (točk). Najpomembnejša otoka sestavljajo letalske ustanove in avtomobilske ustanove.

# Analiza omrežja ProjInst.net



## Analiza omrežja Countries.net



Omrežje Countries.net ima le 60 točk, je pa gosto. Uteži povezav predstavljajo število projektov, pri katerih hkrati sodelujeta krajiščni državi. Za preglednejšo sliko moramo povezave urediti glede na uteži:

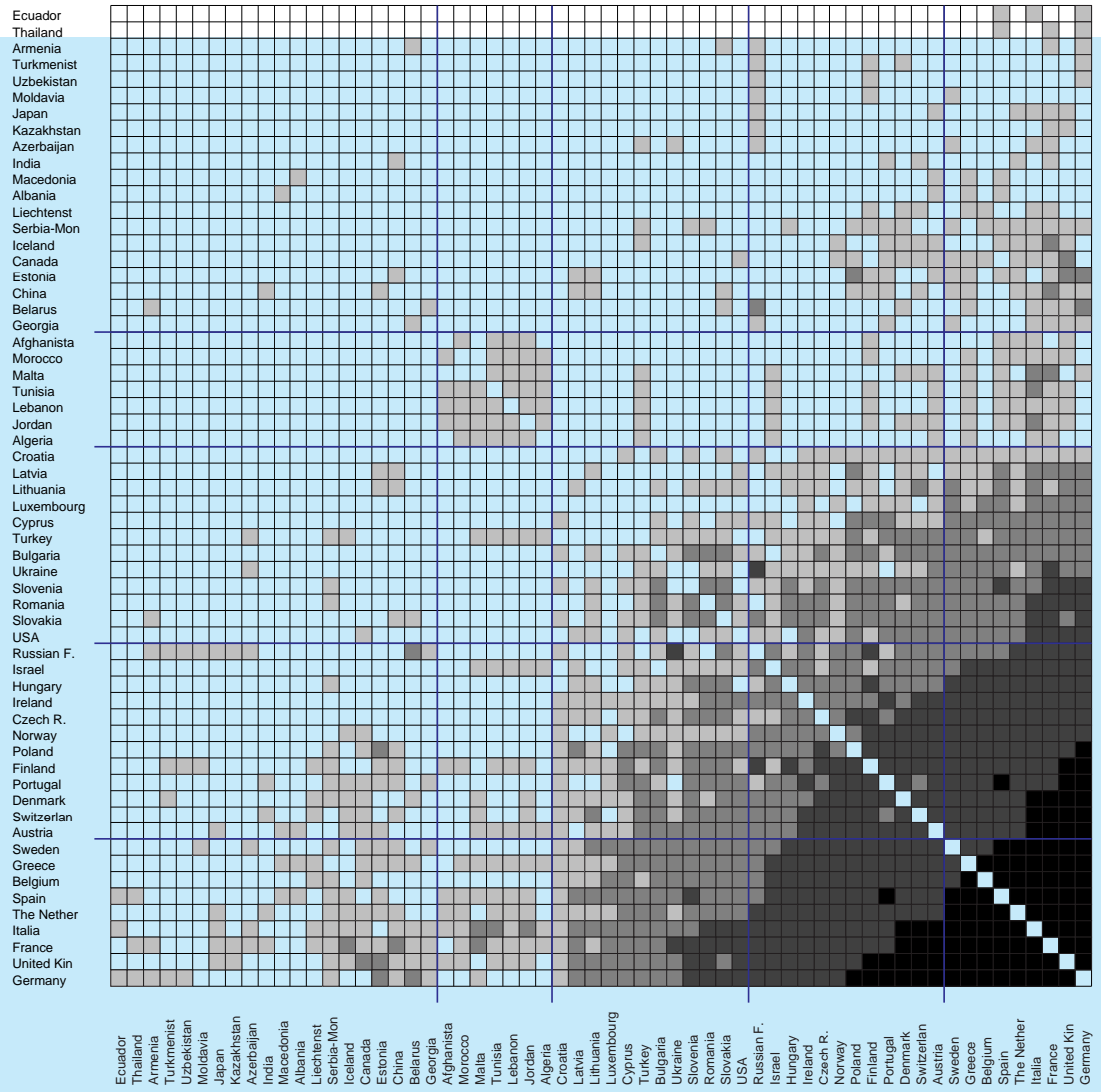
Še preglednejši vpogled v sodelovanje med državami dobimo z matričnim prikazom. Ustrezno urejenost dobimo iz hierarhične razvrstitve dobljene z Wardovim postopkom uporabljenim nad različnostjo  $d_5$ . Uteži so prekodirane glede na prage (2,10,50).

Urejenost lahko še izboljšamo s preurejanjem poddreves v hierarhiji. Kakor vidimo na sliki, dobimo značilno (večslojno) zgradbo središče – obrobje.

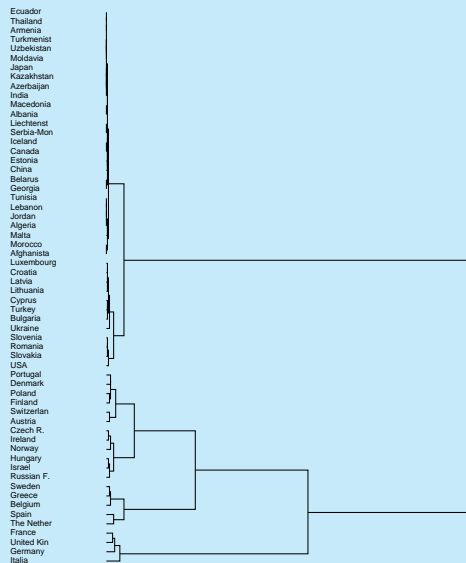


# Analiza omrežja Countries.net

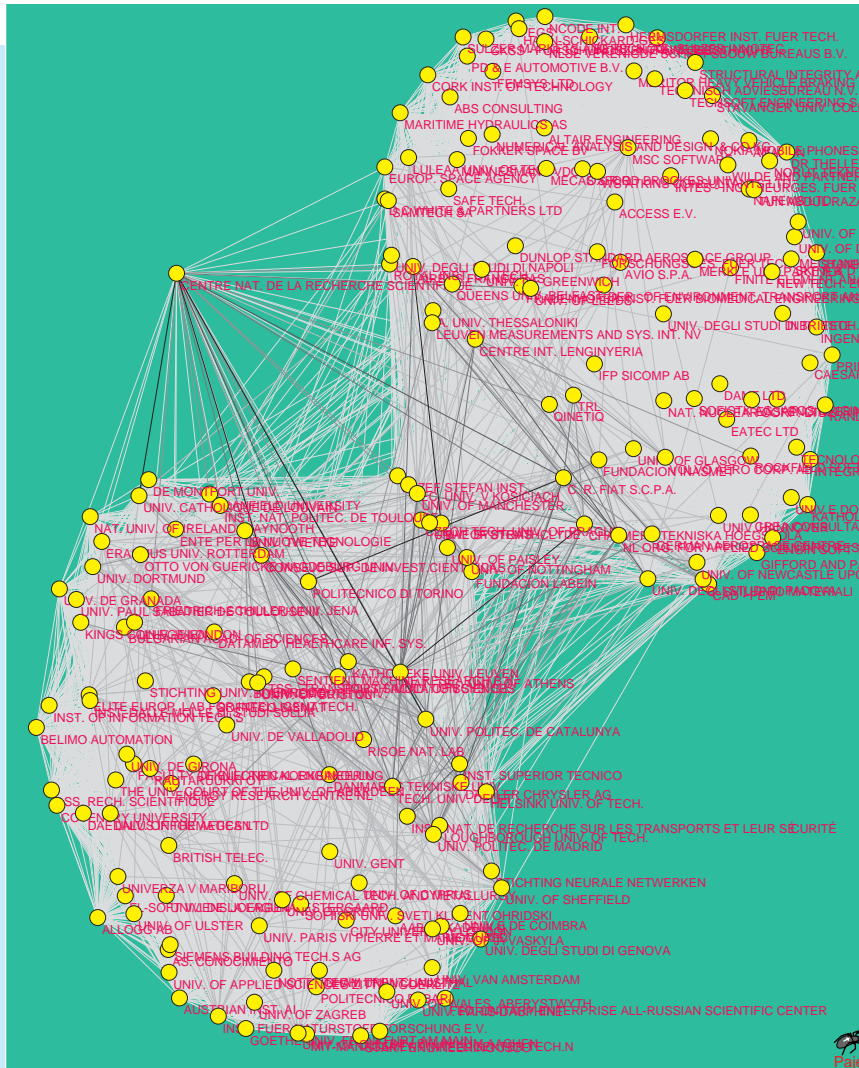
Pajek - shadow [0.00,4.00]



Pajek - Ward [0.00,4785.14]



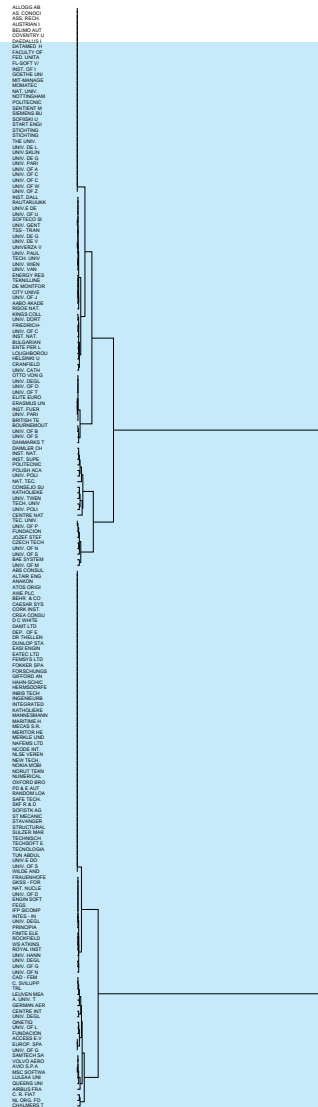
## Analiza omrežja Institutions.net



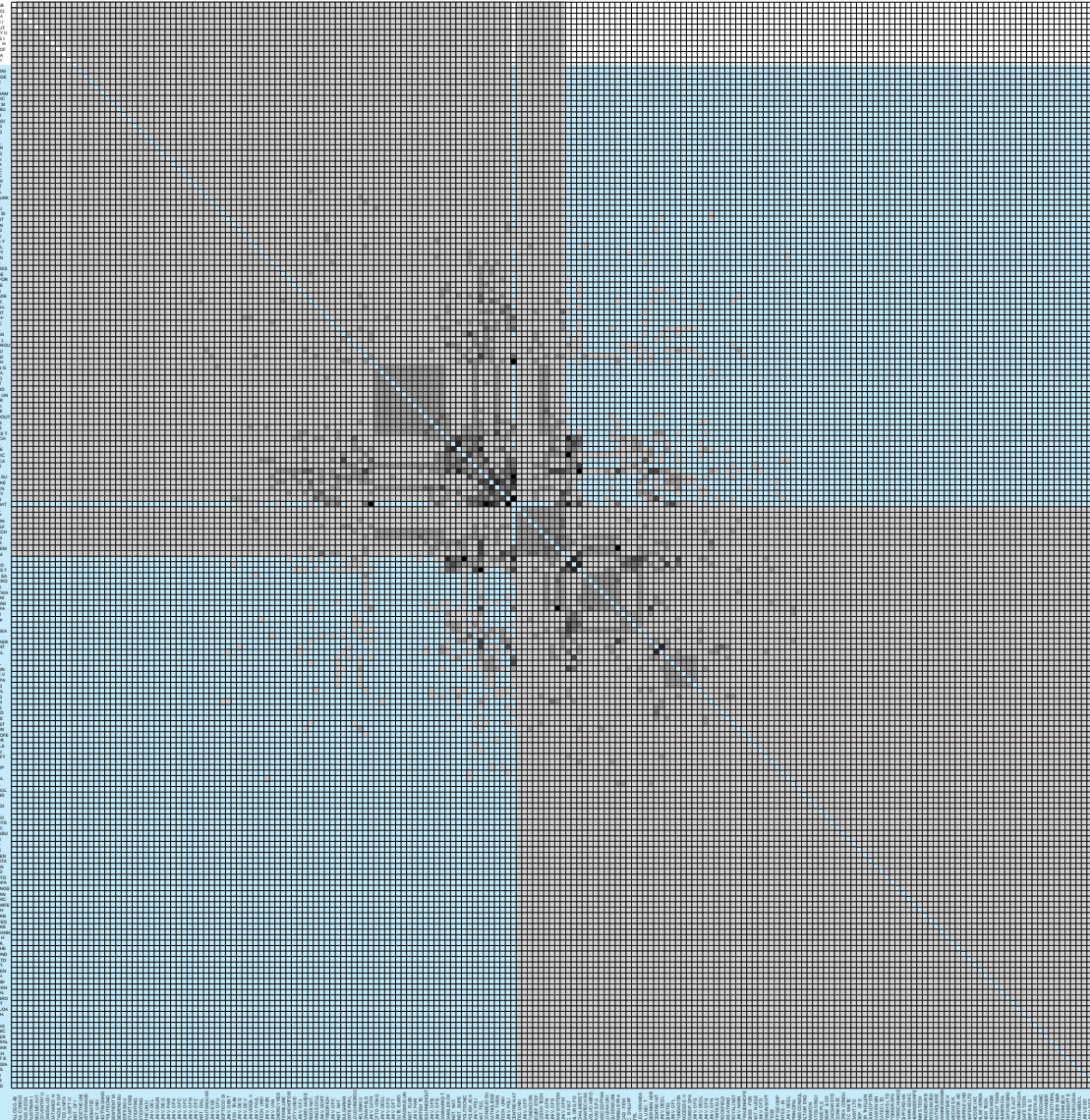
Za določitev najpomembnejših ustanov smo najprej določili sredice vrste  $p_S$  in za tako določen vektor določili točkovne otoke. V bistvu smo dobili en sam velik otok prikazan na sliki. Ta je sestavljen iz dveh večjih skupin povezanih čez posredniške ustanove (med njimi je tudi Inštitut Jožef Stefan). Posebej odstopajoča točka je Centre national de la recherche scientifique (CNRS). Še boljše je medsebojna povezanost med ustanovami razvidna iz matričnega prikaza.

# Analiza omrežja Institutions.net

Pajek - Ward [0.00,1376.93]



Pajek - shadow [0.00,6.00]



## Pretvorba dvovrstnih omrežij na enovrstna

Z uporabo množenja omrežij lahko dvovrstno omrežje  $\mathcal{N} = (\mathcal{I}, \mathcal{J}, \mathcal{E}, w)$  pretvorimo v dve enovrstni omrežji  $\mathcal{N}_1 = \mathcal{N} * \mathcal{N}^T$  in/ali  $\mathcal{N}_2 = \mathcal{N}^T * \mathcal{N}$ . Naj bo  $\mathbf{A}$  matrika omrežja  $\mathcal{N}$  in  $\mathbf{B}$  matrika omrežja  $\mathcal{N}_1$ . Tedaj velja  $\mathbf{B} = \mathbf{A}\mathbf{A}^T$ ,  $b_{uv} = \sum_{z \in \mathcal{J}} a_{uz} \cdot a_{zv}^T = \sum_{z \in \mathcal{J}} a_{uz} \cdot a_{vz}$ . Očitno velja  $b_{uv} = b_{vu}$  – matrika  $\mathbf{B}$  je simetrična.

Tudi omrežju  $\mathcal{N}_2$  pripadajoča matrika  $\mathbf{C} = \mathbf{A}^T \mathbf{A}$  je simetrična.

Enovrstni omrežji  $\mathcal{N}_1$  in  $\mathcal{N}_2$  lahko analiziramo z običajnimi postopki analize enovrstnih omrežij. Osnovna težava je, da sta lahko eno ali celo obe preveliki.

V enovrstnih omrežjih dobljenih iz velikih dvovrstnih omrežij so pogosto razlike v utežeh zelo velike. Zato ni mogoče primerjati točk glede na te vrednosti. Pred primerjavo jih moramo normalizirati – poskrbeti za primerljivost uteži.

## ...Normalizacije

Obstaja več načinov normalizacije. Uporabimo jih lahko tudi na drugih enovrstnih uteženih omrežjih.

$$\text{Geo}_{uv} = \frac{w_{uv}}{\sqrt{w_{uu}w_{vv}}}$$

$$\text{GeoDeg}_{uv} = \frac{w_{uv}}{\sqrt{\text{deg}_u \text{deg}_v}}$$

$$\text{Input}_{uv} = \frac{w_{uv}}{w_{vv}}$$

$$\text{Output}_{uv} = \frac{w_{uv}}{w_{uu}}$$

$$\text{Min}_{uv} = \frac{w_{uv}}{\min(w_{uu}, w_{vv})}$$

$$\text{Max}_{uv} = \frac{w_{uv}}{\max(w_{uu}, w_{vv})}$$

$$\text{MinDir}_{uv} = \begin{cases} \frac{w_{uv}}{w_{uu}} & w_{uu} \leq w_{vv} \\ 0 & \text{sicer} \end{cases}$$

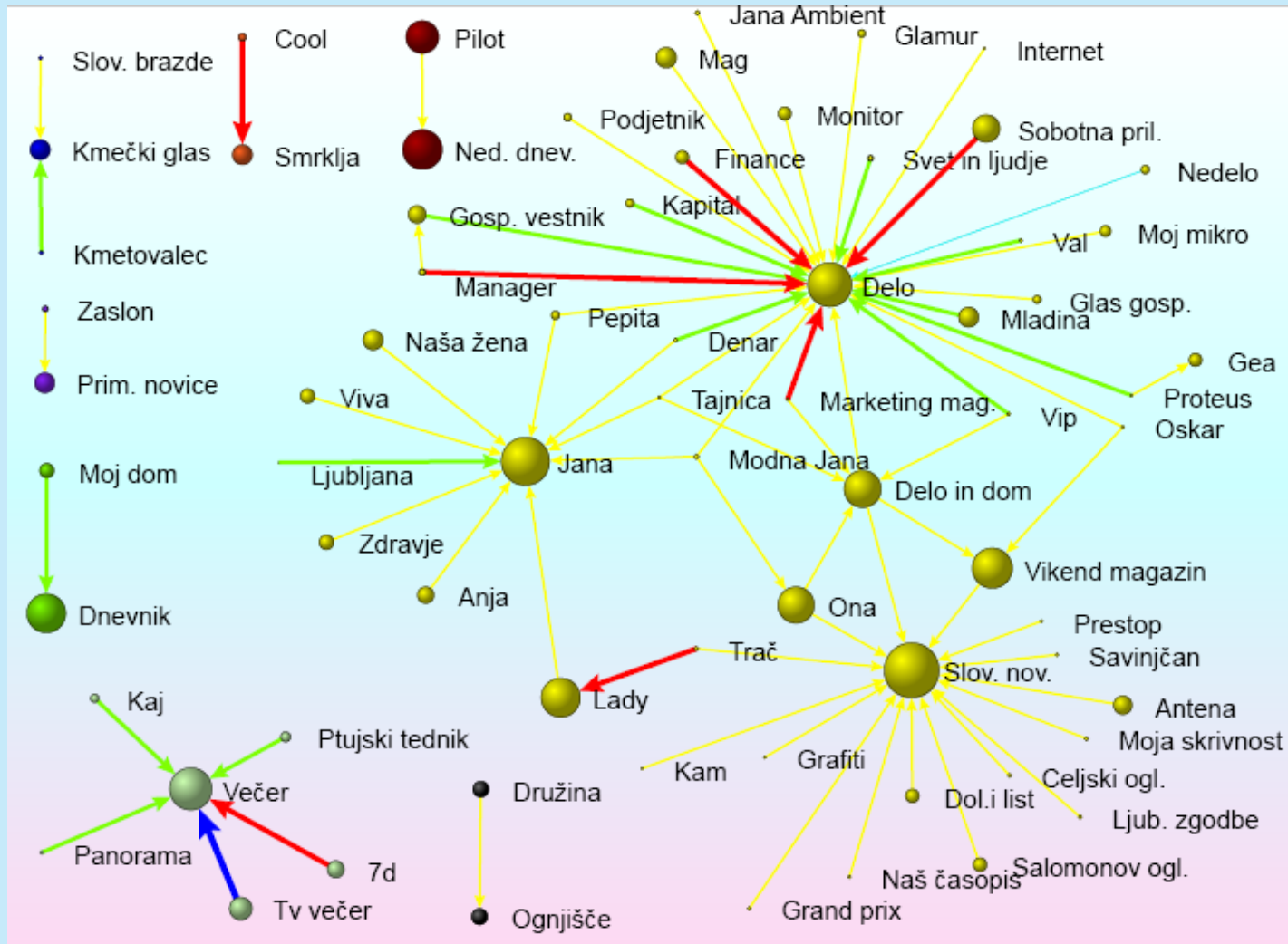
$$\text{MaxDir}_{uv} = \begin{cases} \frac{w_{uv}}{w_{vv}} & w_{uu} \leq w_{vv} \\ 0 & \text{sicer} \end{cases}$$

Normalizirano omrežje analiziramo z uporabo povezavnih prerezov ali otokov.

Reuters Terror News: **GeoDeg**, **MaxDir**, **MinDir**.

# MinDir na omrežju Slovenski časopisi 2000

Čez 100000 oseb je bilo v letih 1999 in 2000 vprašanih, katere časopise berejo. Navedli so 124 različnih časopisov. (vir Cati) **SVG**



**IMDb**  
Earth's Biggest Movie Database™

Home | Top Movies | Photos | Independent Film | Browse | Help | Login | Register to personalize

**The Internet Movie Database**  
Visited by over **30 million** movie lovers each month!

Welcome to the Internet Movie Database, the biggest, best, most award-winning movie site on the planet. Want to make IMDb your home page? Drag [this link](#) onto your Home button.

**Honda Civic** and IMDb Want You to **"Pitch Your Picture"** Today!

**PITCH YOUR PICTURE.**

You have the idea for your movie. You even have the poster. Now, [Honda Civic](#) and IMDb want you to "Pitch Your Picture." Submit your poster for your made-up movie, along with the tagline, and you may be eligible to be [entered into](#) our "Pitch Your Picture" [competition](#) (please note [game rules and restrictions](#)). We are now accepting submissions (voting will commence on the 14th). Use only your original ideas and your original images. Do not use existing screen captures, posters, or stills from other

**Movie and TV News**  
**Wed 19 October 2005:**  
Celebrity News  

- [Kidman Photographer Wins DNA Appeal](#)
- [Sizemore Has His Probation Reinstated](#)
- [Madonna Thanks ABBA for the Music](#)

 Studio Briefing  

- ['Fog' Obscures Box Office](#)
- [Schwarzenegger Wants To Terminate Video Game Lawsuit](#)
- [Jackson Dumps 'King Kong' Music](#)

**Born Today**  
Wednesday, 19 October 2005:

**Search the IMDb**  
All    
[More searches](#) | [Tips](#)  
[IMDbPro.com free trial](#)

**WEB SEARCH**  
   
Powered by A9.com

**Tops at the Box Office**  

- 1 [The Fog](#)
- 2 [Wallace & Gromit in The Curse of the Were-Rabbit](#)
- 3 [Elizabethtown](#)
- 4 [Flightplan](#)
- 5 [In Her Shoes](#)

[more](#)

**Opening this Week**  

- [Doom](#)
- [Where the Truth Lies](#)
- [Kiss Kiss, Bang Bang](#)
- [Shopgirl](#)

Omrežje IMDB je bilo pripravljeno za **12th Annual Graph Drawing Contest, 2005**. Je dvovrstno in ima  $1324748 = 428440 + 896308$  točk ter 3792390 povezav.

## Dvovrstne sredice

Podmnožica točk  $C \subseteq \mathcal{V}$  je  $(p, q)$ -sredica dvovrstnega omrežja  $\mathcal{N} = (\mathcal{V}_1, \mathcal{V}_2; \mathcal{L})$ ,  $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$  natanko takrat, ko

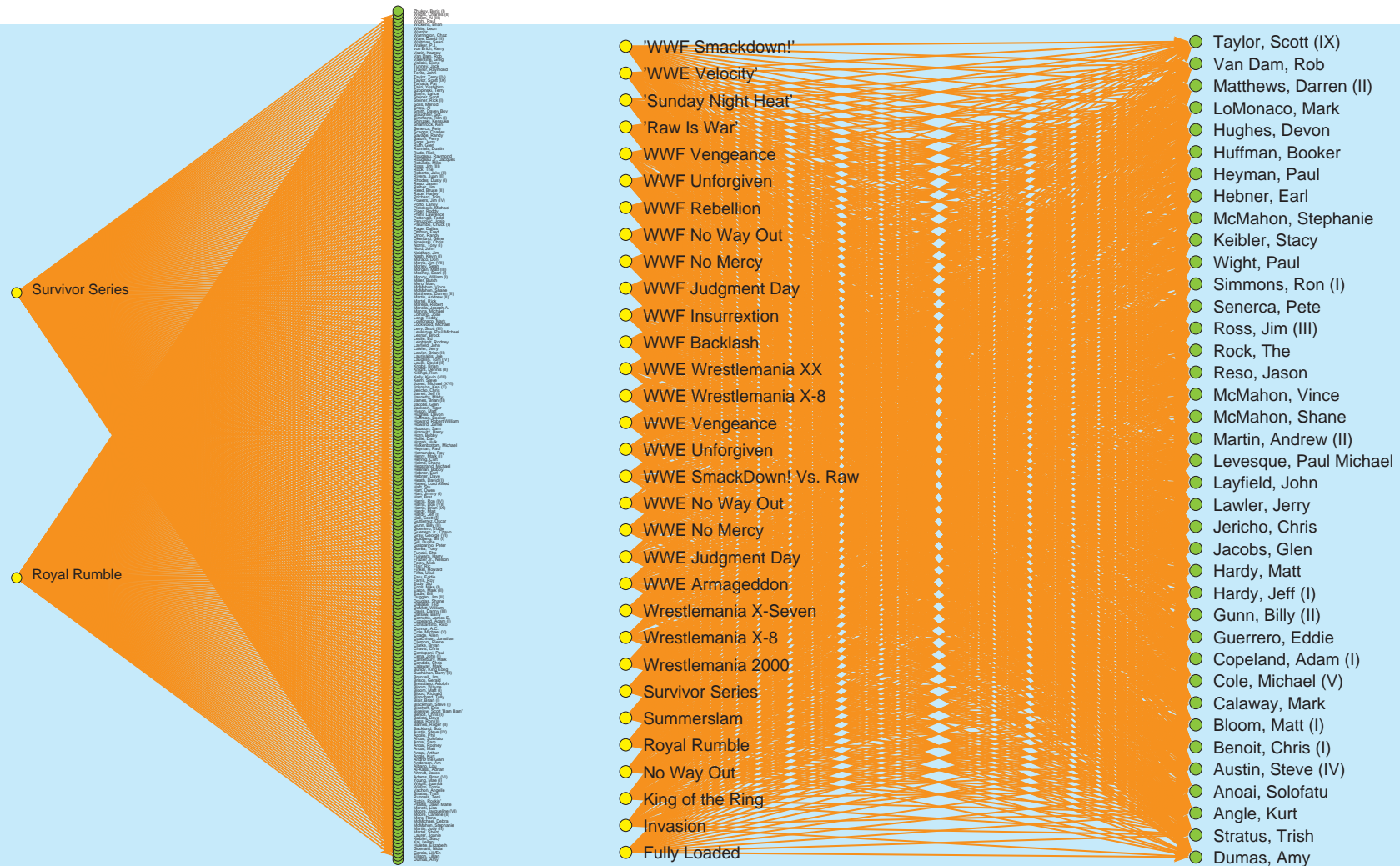
- a. v s  $C$  porojenem podomrežju  $\mathcal{K} = (C_1, C_2; \mathcal{L}(C))$ ,  $C_1 = C \cap \mathcal{V}_1$ ,  $C_2 = C \cap \mathcal{V}_2$  velja  $\forall v \in C_1 : \deg_{\mathcal{K}}(v) \geq p$  in  $\forall v \in C_2 : \deg_{\mathcal{K}}(v) \geq q$ ;
- b.  $C$  je maksimalna podmnožica  $\mathcal{V}$ , ki zadošča pogoju a.

Za določitev  $(p, q)$ -sredice lahko uporabimo postopek podoben postopku za določanje običajnih sredic.

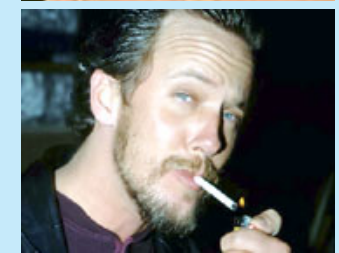
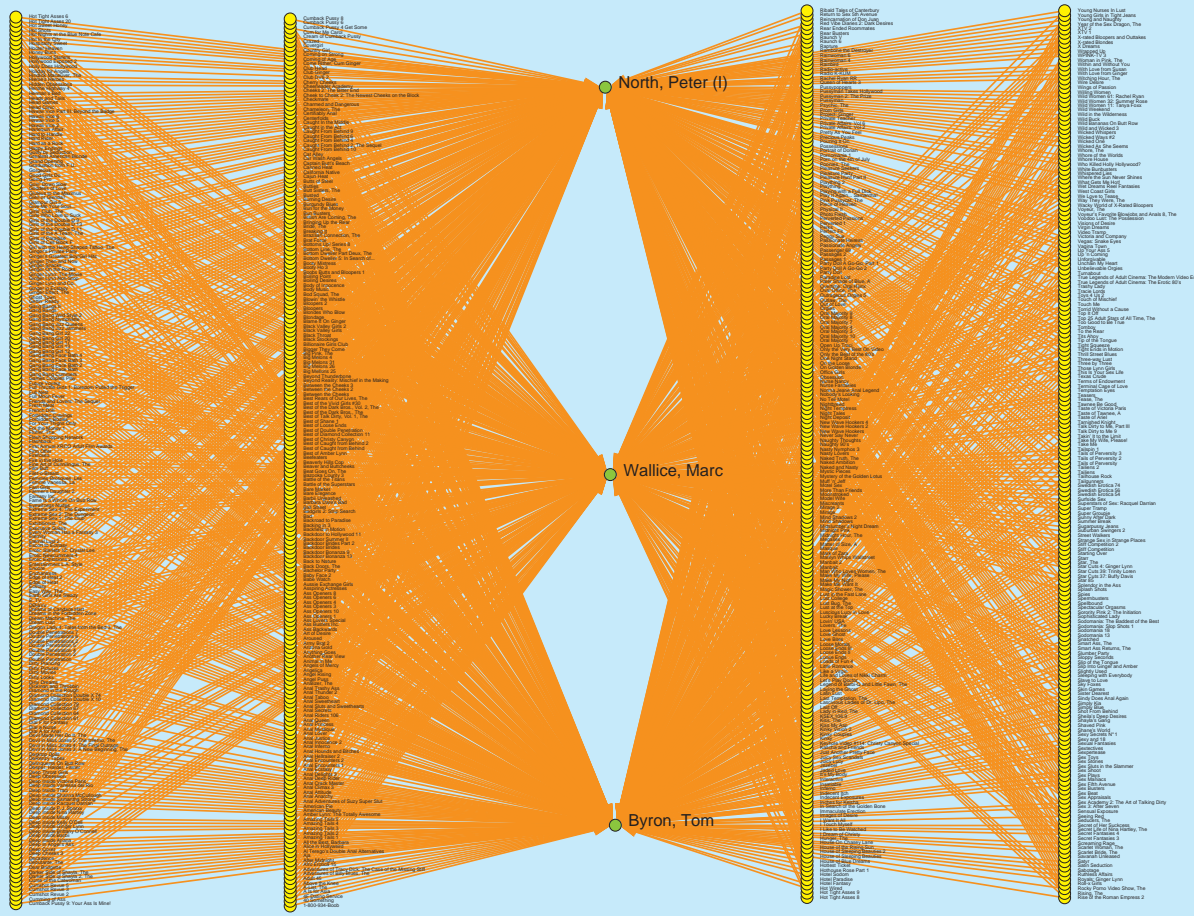
Težava je v tem, da je lahko veliko dvovrstnih sredic. Kako izbrati zanimive? Pomagamo si lahko s tabelo značilnosti sredic.



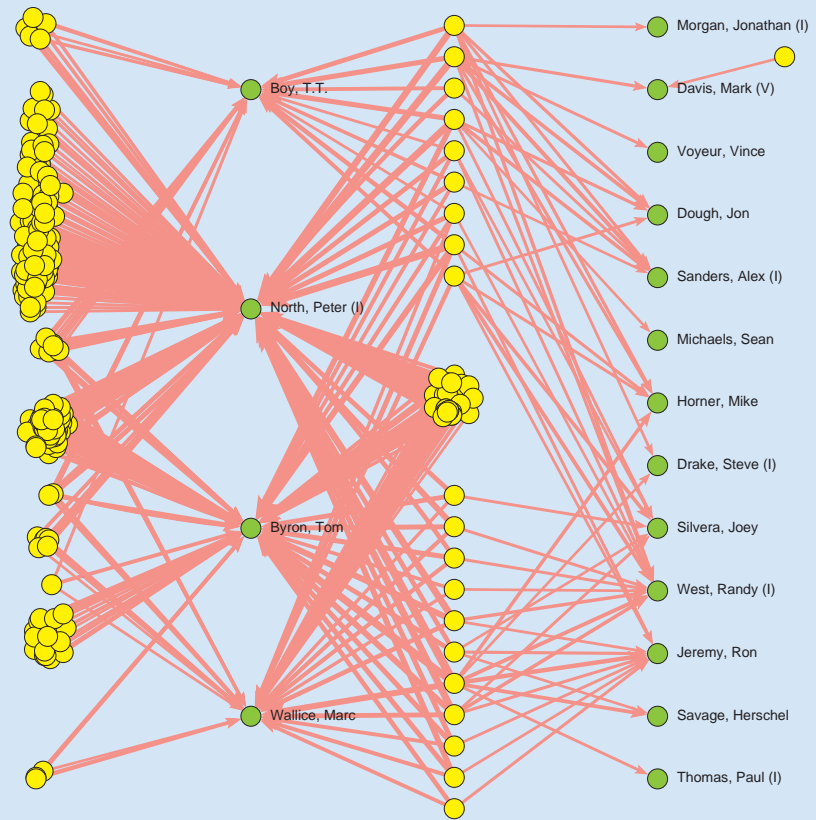
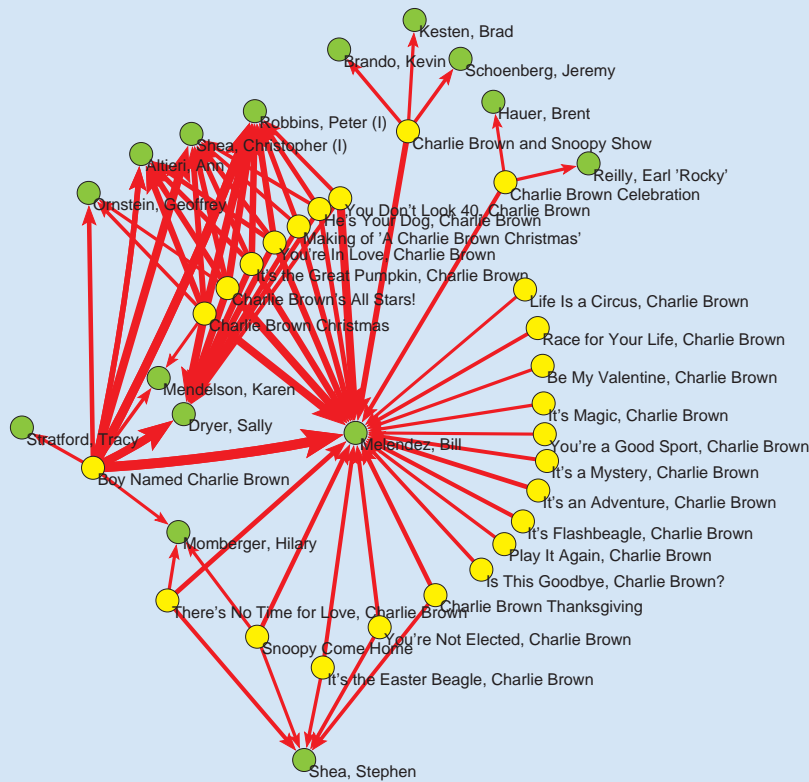
## (247,2)-sredica in (27,22)-sredica



# (2,516)-Hard core



## Primer: 4-obroči – otoka za $w_4$ / Charlie Brown in Adult



## Izbrane knjige o analizi omrežij

- J. P. Scott: *Social Network Analysis: A Handbook*. SAGE Publications, 2000. Amazon.
- A. Degenne, M. Forsé: *Introducing Social Networks*. SAGE Publications, 1999. Amazon.
- S. Wasserman, K. Faust: *Social Network Analysis: Methods and Applications*. CUP, 1994. Amazon.
- W. de Nooy, A. Mrvar, V. Batagelj: *Exploratory Social Network Analysis with Pajek*, CUP, 2005. Amazon. ESNA page.
- P. Doreian, V. Batagelj, A. Ferligoj: *Generalized Blockmodeling*, CUP, 2004. Amazon.
- E. Lazega: *The Collegial Phenomenon: The Social Mechanisms of Cooperation among Peers in a Corporate Law Partnership*. OUP, 2001. Amazon.
- P.J. Carrington, J. Scott, S. Wasserman (Eds.): *Models and Methods in Social Network Analysis*. CUP, 2005. Amazon.
- U. Brandes, T. Erlebach (Eds.): *Network Analysis: Methodological Foundations*. LNCS, Springer, Berlin 2005. Amazon.