



Analiza omrežij

8. Dvovrstna omrežja in množenje omrežij

Vladimir Batagelj

Magistrski program Uporabna statistika
Ljubljana, maj 2024



Kazalo

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

- 1 Dvovrstna omrežja
- 2 IMDB
- 3 Množenje omrežij
- 4 Evropski projekti
- 5 Bibliografska omrežja

NAMES OF PATRIZIANTS OF GROUP I	CODE NUMBERS AND DATES OF SOCIAL EVENTS REPORTED IN Old City Herald													
	(1) 4/75	(2) 4/75	(3) 4/75	(4) 9/76	(5) 3/75	(6) 8/75	(7) 8/75	(8) 9/76	(9) 4/76	(10) 4/76	(11) 3/75	(12) 4/76	(13) 11/75	(14) 4/75
1. Mrs. Evelyn Jefferson.....	X	X	X	X	X	X	X	X	X					
2. Miss Laura Manderville.....	X	X	X	X	X	X	X	X	X					
3. Miss Thelma Anderson.....		X	X	X	X	X	X	X	X					
4. Miss Brenda Rogers.....	X		X	X	X	X	X	X	X					
5. Miss Charlotte McDowd.....			X	X	X	X	X	X	X					
6. Miss Frances Anderson.....			X	X	X	X	X	X	X					
7. Miss Eleanor Nye.....				X	X	X	X	X	X					
8. Miss Paul Ogden.....					X	X	X	X	X					
9. Miss Ruth DeSaut.....				X	X	X	X	X	X					
10. Miss Verne Sanderson.....					X	X	X	X	X			X	X	X
11. Miss Myra Liddell.....						X	X	X	X			X	X	X
12. Miss Katherine Rogers.....						X	X	X	X			X	X	X
13. Mrs. Sylvia Avondale.....						X	X	X	X			X	X	X
14. Mrs. Nora Fayette.....					X	X	X	X	X			X	X	X
15. Mrs. Helen Lloyd.....						X	X	X	X			X	X	X
16. Mrs. Dorothy Marchion.....						X	X	X	X			X	X	X
17. Mrs. Olivia Carleton.....						X	X	X	X			X	X	X
18. Mrs. Flora Price.....						X	X	X	X			X	X	X

prof. Vladimir Batagelj: vladimir.batagelj@mf.uni-lj.si
prosojnice (PDF)

21. maj 2024 ob 01:39/ maj 2013



Dvovrstna omrežja

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Dvovrstno omrežje je omrežje $\mathcal{N} = (\mathcal{U}, \mathcal{V}, \mathcal{A}, w)$, v katerem je množica vozlišč 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 $\mathbf{A} = [a_{uv}]_{\mathcal{U} \times \mathcal{V}}$.

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

Primeri: (osebe, društva, leta članstva),
(kupci, dobrine, vrednost),
(poslanci, zakoni, glasoval za),
(osebe, časopisi, bere),
(članki, gesla, je opisan z), itd.



Pristopi k analizi dvovrstnih omrežij

Za analizo dvovrstnih omrežij lahko prilagodimo Kleinbergov postopek (hubs and authorities) z *lastnimi vektorji*. Vpeljemo vektorja pomembnosti (\mathbf{x}, \mathbf{y}) na $\mathcal{U} \cup \mathcal{V}$ določena z zvezama $\mathbf{y} = \mathbf{A}\mathbf{x}$ in $\mathbf{x} = \mathbf{A}^T\mathbf{y}$.

Network/2-Mode Network/Important Vertices

Novejša pristopa sta: *dvovrstne sredice* in *4-obroči*. V naslednjem predavanju bomo spoznali še uporabo *razvrščanja* in *bločnih modelov* pri analizi dvovrstnih omrežij.



Internet Movie Database <http://www.imdb.com/>

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

The screenshot shows the IMDb website interface. At the top, there's a navigation bar with links like 'NOW PLAYING', 'MOVIE / TV NEWS', 'MY MOVIES', 'DVD / VIDEO', 'IMDb TV', 'MESSAGE BOARDS', 'SHOWTIMES & TICKETS', 'GAME BASE', and a 'FREE TRIAL! IMDb pro' button. Below this is a search bar with the text 'Search the IMDb' and a 'go' button. To the right of the search bar, there's a section titled 'The Internet Movie Database' with a welcome message and a link to 'Pitch Your Picture'. Below this, there's a section titled 'Pitch Your Picture' with a Honda logo and a link to 'Pitch Your Picture'. To the right of the 'Pitch Your Picture' section, there's a section titled 'Movie and TV News' with a list of news items. At the bottom, there's a section titled 'Born Today' with a list of actors born on Wednesday, 19 October 2005.

IMDb
Earth's Biggest Movie Database™

[Home](#) | [Top Movies](#) | [Photos](#) | [Independent Film](#) | [Browse](#) | [Help](#) | [Login](#) | [Register](#) to personalize

Search the IMDb

All

[More searches](#) | [Tips](#) | [IMDbPro.com free trial](#)

WEB SEARCH

Powered by A9.com

Top's 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](#)

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

Born Today

Wednesday, 19 October 2005:

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



Dvovrstne sredice

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Podmnožica vozlišč $C \subseteq \mathcal{V}$ 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.

Lastnosti dvovrstnih sredic:

- $C(0, 0) = \mathcal{V}$
- $\mathcal{K}(p, q)$ ni vedno povezano
- $(p_1 \leq p_2) \wedge (q_1 \leq q_2) \Rightarrow C(p_1, q_1) \subseteq C(p_2, q_2)$
- $\mathcal{C} = \{C(p, q) : p, q \in \mathbb{N}\}$. Če so vse neprazne podmnožice iz \mathcal{C} različne, je mreža.



Postopek za določitev dvovrstnih sredic

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

repeat

iz množice \mathcal{V}_1 odstrani vsa vozlišča s trenutno stopnjo manjšo od p , in

iz množice \mathcal{V}_2 odstrani vsa vozlišča s trenutno stopnjo manjšo od q

until nobeno vozlišče ni bilo odstranjeno;

Postopek je mogoče sprogramirati tako, da ima zahtevnost $O(m)$.

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

$n_1 = |C_1(p, q)|$, $n_2 = |C_2(p, q)|$ in k – je število komponent v $\mathcal{K}(p, q)$:

- $n_1 + n_2 \leq$ izbrano zgornjo mejo
- večji skoki z $C(p - 1, q)$ ali $C(p, q - 1)$ na $C(p, q)$.

Network/2-Mode Network/Core/2-Mode Border



Tabela $(p, q : n_1, n_2)$ za Internet Movie Database

Analiza omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

1	1590:	1590	1	22	24:	1854	1153	43	14:	29	83
2	516:	788	3	23	23:	47	56	44	14:	29	83
3	212:	1705	18	24	23:	34	39	45	13:	30	95
4	151:	4330	154	25	22:	42	53	46	13:	29	94
5	131:	4282	209	26	22:	31	38	47	12:	29	101
6	115:	3635	223	27	22:	31	38	48	12:	28	100
7	101:	3224	244	28	20:	36	53	49	12:	26	95
8	88:	2860	263	29	20:	35	52	50	11:	27	111
9	77:	3467	393	30	19:	35	59	51	11:	26	110
10	69:	3150	428	31	19:	35	59	52	11:	16	79
11	63:	2442	382	32	19:	34	57	53	10:	35	162
12	56:	2479	454	33	18:	34	62	54	10:	35	162
13	50:	3330	716	34	18:	34	62	55	10:	34	162
14	46:	2460	596	35	18:	33	61	56	10:	34	162
15	42:	2663	739	36	17:	33	65	57	9:	35	187
16	39:	2173	678	37	16:	33	75	58	9:	33	180
17	35:	2791	995	38	16:	30	73	59	9:	33	180
18	32:	2684	1080	39	16:	29	70	60	9:	32	178
19	30:	2395	1063	40	15:	29	77	61	9:	31	177
20	28:	2216	1087	41	15:	28	76	62	9:	31	177
21	26:	1988	1087	42	15:	28	76	63	8:	31	202



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

Analiza omrežij

V. Batagelj

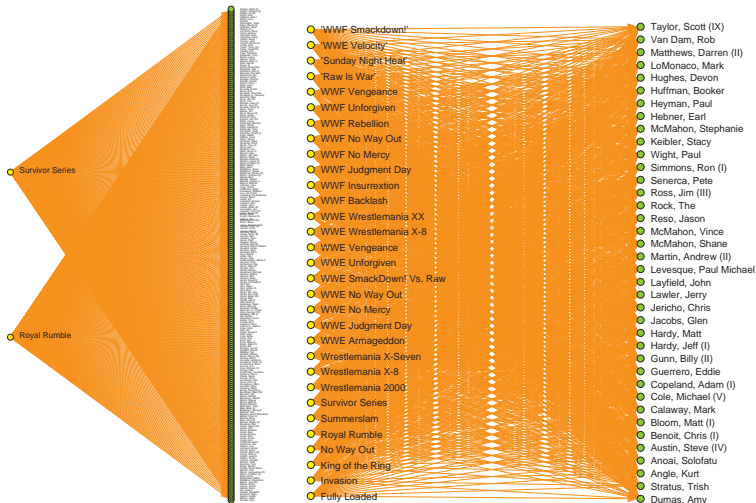
Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja





(2,516)-Hard core

Analiza
omrežij

V. Batagelj

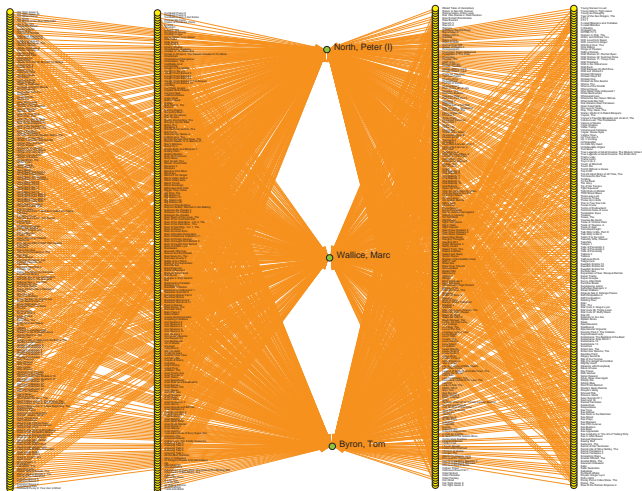
Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja





4-obroči in analiza dvovrstnih omrežij

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

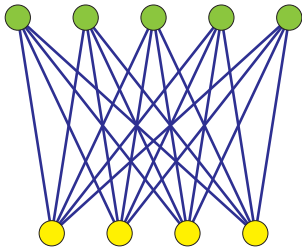
IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

V dvovrstnih omrežjih ni 3-obročev. Najkrajše sklenjene verige imajo dolžino 4 – so 4-obroči. Najgostejše podstrukture v dvovrstnih omrežjih so polni dvodelni podgrafi $K_{p,q}$. Ti vsebujejo veliko 4-obročev.



$$w_4(K_{p,q}) = (p-1)(q-1)$$

Določanje 4-obročnih uteži je bilo vključeno v program Pajek avgusta 2005.

Network/Create New Network/with Ring Counts.../4-Rings/Undirected



4-obroči v usmerjenih omrežjih

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

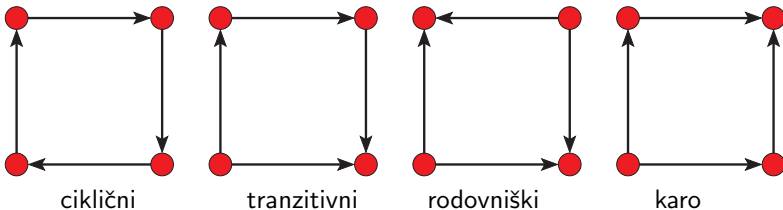
IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Obstajajo 4 vrste usmerjenih 4-obročev:



Za tranzitivne obroče Pajek določi še posebno utež, ki pove, kolikokrat je povezava *bližnjica* v tranzitivnem 4-obroču.

Network/Create New Network/with Ring Counts.../4-Rings/Directed



Enostavni povezavni otoki v IMDB za w_4

Analiza
omrežij

V. Batagelj

Dobili smo 12465 enostavnih povezavnih otokov na 56086 vozliščih.
Tu je porazdelitev njihovih velikosti.

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Size	Freq	Size	Freq	Size	Freq	Size	Freq
2	5512	20	19	38	4	59	2
3	1978	21	18	39	3	61	1
4	1639	22	15	40	2	64	1
5	968	23	9	42	2	67	1
6	666	24	13	43	3	70	1
7	394	25	12	45	3	73	1
8	257	26	6	46	4	76	1
9	209	27	6	47	5	82	1
10	148	28	5	48	1	86	1
11	118	29	6	49	2	106	1
12	87	30	3	50	2	122	1
13	55	31	6	51	1	135	1
14	62	32	5	52	2	144	1
15	46	33	3	53	1	163	1
16	39	34	1	54	2	269	1
17	27	35	5	55	1	301	1
18	28	36	4	57	1	332	2
19	29	37	7	58	1	673	1



Primer: Otoka za w_4 / Charlie Brown in Adult

Analiza
omrežij

V. Batagelj

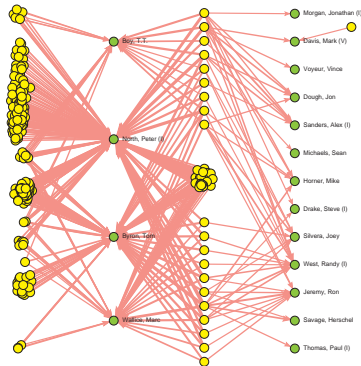
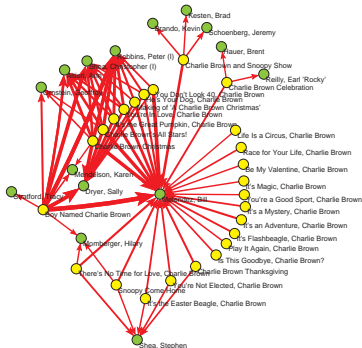
Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja





Primer: Otok za w_4 / Polizeiruf 110 in Starkes Team

Analiza
omrežij

V. Batagelj

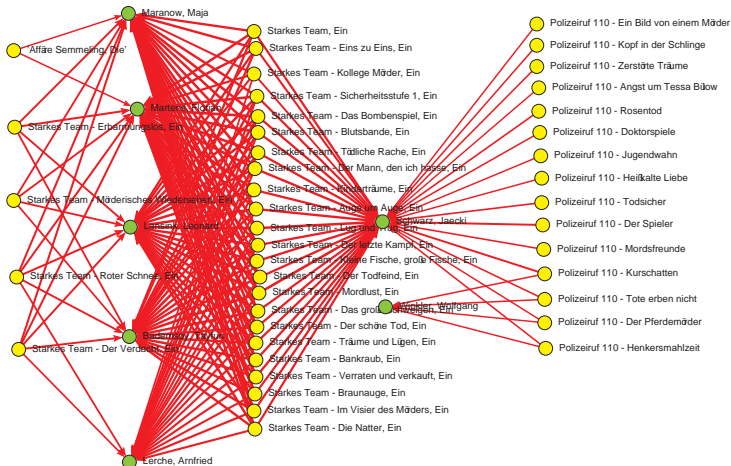
Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja





5-obroči

Analiza
omrežij

V. Batagelj

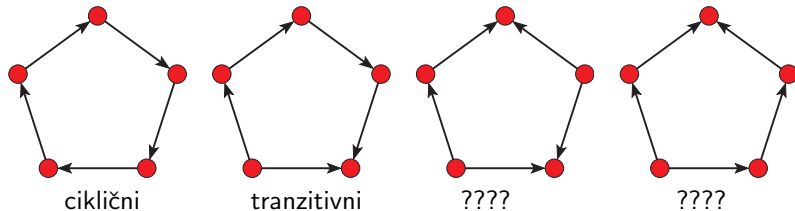
Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja





Množenje omrežij

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Naj bo $\mathcal{N} = (\mathcal{I}, \mathcal{J}, \mathcal{E}, w)$ enostavno dvovrstno omrežje nad množicama vozlišč \mathcal{I} in \mathcal{J} ter množico povezav \mathcal{E} , ki imajo krajišča v obeh množicah \mathcal{I} in \mathcal{J} . $w : \mathcal{E} \rightarrow \mathbb{R}$ (ali kak drug polkolobar) je utež. Omrežju priredimo matriko omrežja $\mathbf{W} = [w_{i,j}]$ z vrednostmi: $w_{i,j} = w(i,j)$ za $(i,j) \in \mathcal{E}$ in $w_{i,j} = 0$ sicer.

Imejmo usklajeni omrežji $\mathcal{N}_A = (\mathcal{I}, \mathcal{K}, \mathcal{E}_A, w_A)$ in $\mathcal{N}_B = (\mathcal{K}, \mathcal{J}, \mathcal{E}_B, w_B)$ s pripadajočima matrikama $\mathbf{A}_{\mathcal{I} \times \mathcal{K}}$ in $\mathbf{B}_{\mathcal{K} \times \mathcal{J}}$. **Produkt omrežij** \mathcal{N}_A in \mathcal{N}_B imenujemo omrežje $\mathcal{N}_C = (\mathcal{I}, \mathcal{J}, \mathcal{E}_C, w_C)$, kjer so $\mathcal{E}_C = \{(i,j) : i \in \mathcal{I}, j \in \mathcal{J}, c_{i,j} \neq 0\}$ in $w_C(i,j) = c_{i,j}$ za $(i,j) \in \mathcal{E}_C$. Matrika produkta $\mathbf{C} = [c_{i,j}]_{\mathcal{I} \times \mathcal{J}} = \mathbf{A} * \mathbf{B}$ je določena na običajen način

$$c_{i,j} = \sum_{k \in \mathcal{K}} a_{i,k} \cdot b_{k,j}$$

V primeru, ko so $\mathcal{I} = \mathcal{K} = \mathcal{J}$ imamo opravka z navadnimi enovrstnimi omrežji (s kvadratnimi matrikami).



Množenje omrežij

Analiza omrežij

V. Batagelj

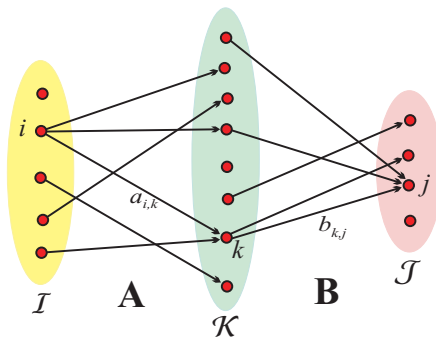
Dvovrstna omrežja

IMDB

Množenje omrežij

Evropski projekti

Bibliografska omrežja



$$c_{i,j} = \sum_{k \in N_A(i) \cap N_B^-(j)} a_{i,k} \cdot b_{k,j}$$

Če so vse uteži v omrežjih \mathcal{N}_A in \mathcal{N}_B enake 1, vrednost $c_{i,j}$ šteje na koliko načinov lahko pridemo iz vozlišča $i \in \mathcal{I}$ v vozlišče $j \in \mathcal{J}$ čez neko vozlišče iz \mathcal{K} .



Običajno množenje matrik

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Običajni postopek za izračun produkta matrik \mathbf{C}

$$c_{i,j} = \sum_{k \in \mathcal{K}} a_{i,k} \cdot b_{k,j}$$

je naslednji

```
for  $i$  in  $\mathcal{I}$  do  
  for  $j$  in  $\mathcal{J}$  do begin  
     $s := 0$ ;  
    for  $k$  in  $\mathcal{K}$  do  $s := s + a_{i,k} * b_{k,j}$ ;  
     $c_{i,j} := s$ ;  
  end;
```

Njegova zahtevnost je reda $O(|\mathcal{I}| \cdot |\mathcal{K}| \cdot |\mathcal{J}|)$ – in je zato prepočasno za uporabo na velikih omrežjih.



Hitro množenje redkih matrik

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Za velika redka omrežja lahko izračunamo produkt veliko hitreje, če upoštevamo le neničelne vrednosti:

```
for  $k$  in  $\mathcal{K}$  do
  for  $i$  in  $N_A^-(k)$  do
    for  $j$  in  $N_B(k)$  do
      if  $\exists c_{i,j}$  then  $c_{i,j} := c_{i,j} + a_{i,k} * b_{k,j}$ 
      else new  $c_{i,j} := a_{i,k} * b_{k,j}$ 
```

$N_A^-(k)$: sosedi vozlišča k v omrežju \mathcal{N}_A

$N_B(k)$: sosedi vozlišča k v omrežju \mathcal{N}_B

V splošnem je množenje velikih redkih omrežij nevarna operacija, ker se lahko izid 'razpoči' – ni redek.

Networks/Multiply Networks



Zahtevnost hitrega množenja

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Iz postopka množenja omrežij vidimo, da vsako vmesno vozlišče $k \in \mathcal{K}$ doda produktu polno dvovrstno podomrežje $K_{N_A^-(k), N_B(k)}$ (ali, v primeru $\mathcal{I} = \mathcal{J}$, polno podomrežje $K_{N(k)}$). Če sta obe stopnji $\deg_A(k) = |N_A^-(k)|$ in $\deg_B(k) = |N_B(k)|$ veliki, potem je že izračun tega podomrežja kvadratičen (prostor in čas) – izid se razpoči.

Enostavno je videti, da če ima vsaj eno od redkih omrežij \mathcal{N}_A in \mathcal{N}_B majhno največjo stopnjo Δ na vmesni množici \mathcal{K} , potem je tudi produktno omrežje \mathcal{N}_C redko.

Pokazali bomo močnejše trditev: Če za redki omrežji \mathcal{N}_A in \mathcal{N}_B obstaja v \mathcal{K} le nekaj vozlišč velike stopnje in ni nobeno tako, da bi imelo veliko stopnjo v obeh omrežjih, potem je tudi produktno omrežje \mathcal{N}_C redko.



Podrobnejša analiza zahtevnosti

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Bodita **A** in **B** matriki omrežij $\mathcal{N}_A = (\mathcal{I}, \mathcal{K}, \mathcal{E}_A, w_A)$ in $\mathcal{N}_B = (\mathcal{K}, \mathcal{J}, \mathcal{E}_B, w_B)$. Privzemimo, da lahko telo zank izračunamo v konstantnem času c . Označimo $d_{\min}(k) = \min(\deg_A(k), \deg_B(k))$, $\Delta_{\min} = \max_{k \in \mathcal{K}} d_{\min}(k)$, $d_{\max}(k) = \max(\deg_A(k), \deg_B(k))$, $\mathcal{K}(d) = \{k \in \mathcal{K} : d_{\max}(k) \geq d\}$, $d^* = \operatorname{argmin}_d (|\mathcal{K}(d)| \leq d)$ in $\mathcal{K}^* = \mathcal{K}(d^*)$. Velja $|\mathcal{K}^*| \leq d^*$. Tedaj je zahtevnost izračuna produkta enaka

$$\begin{aligned} C &= \sum_{k \in \mathcal{K}} \sum_{i \in N_A(k)} \sum_{j \in N_B(k)} c = c \cdot \sum_{k \in \mathcal{K}} \deg_A(k) \cdot \deg_B(k) = c \cdot \sum_{k \in \mathcal{K}} d_{\min}(k) \cdot d_{\max}(k) \\ &= c \cdot \left(\sum_{k \in \mathcal{K}^*} d_{\min}(k) \cdot d_{\max}(k) + \sum_{k \in \mathcal{K} \setminus \mathcal{K}^*} d_{\min}(k) \cdot d_{\max}(k) \right) \\ &\leq c \cdot \left(\Delta_{\min} \cdot \sum_{k \in \mathcal{K}^*} d_{\max}(k) + d^* \cdot \sum_{k \in \mathcal{K} \setminus \mathcal{K}^*} d_{\min}(k) \right) \\ &\leq c \cdot d^* \cdot (\Delta_{\min} \cdot \max(|\mathcal{I}|, |\mathcal{J}|) + \min(|\mathcal{E}_A|, |\mathcal{E}_B|)) \end{aligned}$$

Če sta za omrežji \mathcal{N}_A in \mathcal{N}_B količini Δ_{\min} in d^* majhni, tedaj je tudi produktno omrežje \mathcal{N}_C redko.



Primer: Sorodstvene vezi

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

V antropologiji štejejo za osnovne sorodstvene relacije:

Kin	Type	English	Type	Slovensko
P		Parent		starši
	F	Father		oče
	M	Mother		mati
C		Child		otrok
	D	Daughter		hči
	S	Son		sin
G		Sibling		brat sestra
	Z	Sister		sestra
	B	Brother		brat
E		Spouse		zakonec
	W	Wife		žena
	H	Husband		mož

Rodovniki so običajno opisani v obliki **GEDCOM**. Primeri family, Bouchards.



Orejev graf

Analiza omrežij

V. Batagelj

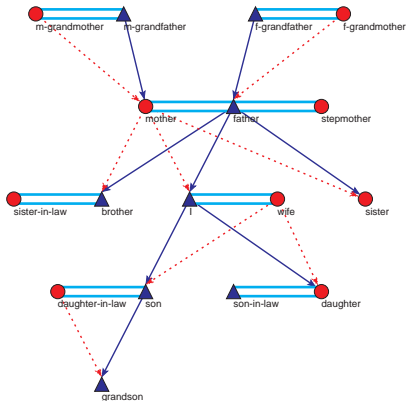
Dvovrstna omrežja

IMDB

Množenje omrežij

Evropski projekti

Bibliografska omrežja



V Orejevem grafu
je vsaka oseba predstavljena z
vozliščem,
poroke, odnos
_ is a spouse of _,
je predstavljen z neusmerjeno
povezavo,
odnosa
_ is a mother of _
in
_ is a father of _
pa z usmerjeno povezavo s
starša na otroka.



Izračuni sorodstvenih vezi

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Ko prebere rodovnik kot Orejev graf, Pajek ustvari tri relacije:

F: _ is a father of _

M: _ is a mother of _

E: _ is a spouse of _

Za nadaljnje potrebe moramo ustvariti še dve (diagonalni) relaciji, ki nam omogočata razlikovanje med moškimi in ženskami:

L: _ is a male _ / 1-moški, 0-ženska

J: _ is a female _ / 1-ženska, 0-moški

$$\mathbf{F} \cap \mathbf{M} = \emptyset, \quad \mathbf{L} \cup \mathbf{J} \subseteq \mathbf{I}, \quad \mathbf{L} \cap \mathbf{J} = \emptyset$$



Izpeljane sorodstvene relacije

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Druge osnovne sorodstvene relacije lahko določimo z uporabo **makrojev**, ki temeljijo na naslednjih zvezah:

_ is a parent of _	$P = F \cup M$
_ is a child of _	$C = P^T$
_ is a son of _	$S = L * C$
_ is a daughter of _	$D = J * C$
_ is a husband of _	$H = L * E$
_ is a wife of _	$W = J * E$
_ is a sibling of _	$G = ((F^T * F) \cap (M^T * M)) \setminus I$
_ is a brother of _	$B = L * G$
_ is a sister of _	$Z = J * G$
_ is an uncle of _	$U = B * P$
_ is an aunt of _	$A = Z * P$
_ is a semi-sibling of _	$G_e = (P^T * P) \setminus I$

Z njihovo uporabo lahko določimo še vrsto drugih relacij:

_ is a grand mother of _	$M_2 = M * P$
_ is a niece of _	$Ni = D * G$



Razmerja med velikostmi sorodstvenih relacij v rodovnikih

Analiza omrežij

V. Batagelj

Dvovrstna omrežja

IMDB

Množenje omrežij

Evropski projekti

Bibliografska omrežja

Kin Type	Turks	Ragusa	Loka	Silba	Royal
P-Parent	1.000	1.000	1.000	1.000	1.000
F-Father	0.514	0.532	0.504	0.519	0.540
M-Mother	0.486	0.468	0.496	0.481	0.460
C-Child	1.000	1.000	1.000	1.000	1.000
D-Daughter	0.431	0.384	0.480	0.469	0.427
S-Son	0.569	0.616	0.520	0.531	0.573
G-Sibling	1.250	0.943	1.019	0.811	0.767
Z-Sister	1.135	0.746	0.983	0.760	0.707
B-Brother	1.366	1.140	1.055	0.861	0.828
E-Spouse	0.205	0.215	0.208	0.230	0.306
H-Husband	0.205	0.215	0.208	0.230	0.306
W-Wife	0.205	0.215	0.208	0.230	0.306
U-Uncle	1.920	1.789	1.200	1.181	0.927
A-Aunt	1.750	1.143	1.190	1.097	0.798
Ge-Semi-sibling	1.473	1.155	1.128	0.932	0.905
n	1269	5999	47956	6427	3010
mE = Spouse	407	2002	14154	2217	1138
mA = Parent	1987	9315	68052	9627	3724



Omrežja iz podatkovnih tabel in baz

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

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	V3D	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	Carri	UNITED KI	60	2	NORTH V		
6	5	1406	442.html	3E	3E		PALMERS, Geert	Erediet	1000	NNE5/51/1999	Brux	BELGIQUE	8	2	REG. BRU	
7	6	1007	884.html	4M2C PATRIC SALOM	4M2C	P	N/A	CRANA	12157	507255	Berlin	DEUTSCH	15	2	BERLIN	
8	7	7914	991.html	5T S.c.r.l.	5T	S.C.	N/A	C.so B	10126	Road2/506716	Torin	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	Rome	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	Rome	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 Univer	A. MInst	PATKOWSKI, Ad	UL H. V61-712		502235	Pozn	POLSKA	45	2			
13	12	6571	135.html	A. BRITO - INDUSTRIA	A. BRITO	VIEIRA DE BRIT	5109_E4350-115	BRST985263	Porto	PORTUGA	46	2	CONTINEI			
14	13	1813	409.html	A.L. DIGITAL LIMITED	A.L.	A.L.	LAURIE, Ben	VOYSEW4	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	Voysew4	4GB	IST-2000-2633	Chisv	UNITED KI	60	2	SOUTH E	
16	15	1885	960.html	A.P. MOLLER-MAER	A.P.	TEC	DRASTED, Jorr	Esplan	1098	506676	Kope	DANMARK	14	2	København	
17	16	6731	537.html	A.S.M. S.A.	A.S.M.	S.MOYA GARCIA,	Carrete	43206	IST-2000-3008	Reus	ESPAÑA	19	2	ESTE CA		
18	17	8150	232.html	AABO AKADEMI UNIF	AAB	CO	NYBACKA-WILL	14-18B	20500	ERK5-CT-1995	Turku	SUOMI/FIN	53	2	MANNER-	
19	18	8152	662.html	AABO AKADEMI UNIF	AAB	DEF	BJORKSTRAND,	3,Tykie	20521	EVK1-CT-2002	Turku	SUOMI/FIN	53	2		
20	19	8148	959.html	AABO AKADEMI UNIF	AAB	Dep	HUPA, Mikko	Domky	20500	502679	Turku	SUOMI/FIN	53	2	MANNER-	
21	20	8151	233.html	AABO AKADEMI UNIF	AAB	DEF	NYBACKA-WILL	Lemmi	20500	ERK5-CT-1995	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 ali baza \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 .



Zapis z Web of Science

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

PT J
AU Dipple, H
Evans, B
TI The Leicestershire Huntington's disease support group: a social network
analysis
SO HEALTH & SOCIAL CARE IN THE COMMUNITY
LA English
DT Article
C1 Rehabil Serv, Troon Way Business Ctr, Leicester LE4 9HA, Leics, England.
RP Dipple, H, Rehabil Serv, Troon Way Business Ctr, Sandringham
Suite, Humberstone Lane, Leicester LE4 9HA, Leics, England.
CR BORGATTI SP, 1992, UCINET 4 VERSION 1 0
FOLSTEIN S, 1989, HUNTINGTONS DIS DISO
SCOTT J, 1991, SOCIAL NETWORK ANAL
NR 3
TC 3
PU BLACKWELL SCIENCE LTD
PI OXFORD
PA P O BOX 88, OSNEY MEAD, OXFORD OX2 ONE, OXON, ENGLAND
SN 0966-0410
J9 HEALTH SOC CARE COMMUNITY
JI Health Soc. Care Community
PD JUL
PY 1998
VL 6
IS 4
BP 286
EP 289
PG 4
SC Public, Environmental & Occupational Health; Social Work
GA 105UP
UT ISI:000075092200008
ER

WoS2Pajek



Zapisi BiBTeX

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

```
@Article{int:Mizuno1,
  author =      "S. Mizuno",
  title =       "An  $O(n^3L)$  algorithm using a sequence for
                linear complementarity problems",
  journal =     "Journal of the Operations Research Society of Japan",
  volume =     "33",
  year =       "1990",
  pages =      "66--75",
}

@InCollection{int:Vorst1,
  author =      "{J. G. G. van de} Vorst",
  title =       "An attempt to use parallel computing in large scale
                optimisation",
  booktitle =   "Logistics, Where Ends Have to Meet~: Proceedings of
                the Shell Conference on Logistics in Apeldoorn, The
                Netherlands, November 1988",
  editor =      "{C. F. H. van} Rijn",
  year =       "1989",
  pages =      "112--119",
  publisher =   "Pergamon Press",
  address =     "Oxford, United Kingdom",
}
```

Bib2Pajek.py



... Omrežja iz podatkovnih tabel

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

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 $[\mathbf{q}_i \times \mathbf{q}_j] / \mathbf{q}_s = [\mathbf{q}_i \times \mathbf{q}_s] * [\mathbf{q}_s \times \mathbf{q}_j]$.



Dvovrstna omrežja iz podatkovnih baz

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Za podatke iz **Web of Science** (Knowledge) lahko ustvarimo ustrezna omrežja s programom **WoS2Pajek**:

- omrežje sklicevanj **Ci**: dela \times dela;
- omrežje avtorstev **WA**: dela \times avtorji; za dela brez polnih opisov je znan le prvi avtor;
- omrežje gesel (ključnih besed) **WK**: dela \times gesla (samo za dela s polnimi opisi);
- omrežje revij **WJ**: dela \times revije;
- razbitje del glede na leto objave;
- razbitje del – polni opis (1) / samo ime ISI (0);

Podobni programi obstajajo tudi za druge vire: Scopus, BibTeX, Zentralblatt Math, Google Scholar, DBLP, IMDB, itd.



Evropski projekti na temo simulacij

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

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:

- $\text{project.net} - \text{idents} \times \text{projects} = \mathbf{P}$
- $\text{country.net} - \text{idents} \times \text{countries} = \mathbf{C}$
- $\text{institution.net} - \text{idents} \times \text{institutions} = \mathbf{U}$

Velikosti posameznih množic so naslednje:

$$\begin{aligned} |\text{idents}| &= 8869, & |\text{projects}| &= 933, & |\text{institutions}| &= 3438, \\ |\text{countries}| &= 60. \end{aligned}$$



Evropski projekti – izpeljana omrežij

Analiza omrežij

V. Batagelj

Dvovrstna omrežja

IMDB

Množenje omrežij

Evropski projekti

Bibliografska omrežja

Ker imajo vsa tri omrežja skupno množico \mathcal{K} =idents, lahko na prej opisani način iz njih z množenjem pridobimo različna *izpeljana* 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 $\mathbf{Q} = \mathbf{W}^T \star \mathbf{W}$
- ...

Obstajata posebna ukaza za $\mathbf{A}^T \star \mathbf{A}$ in $\mathbf{A} \star \mathbf{A}^T$

Network/2-Mode Network/2-Mode to 1-Mode/Rows

Network/2-Mode Network/2-Mode to 1-Mode/Columns



Evropski projekti – izločeni projekti

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Za 27 zapisov v tabeli podatki niso popolni. Običajno problem rešimo tako, da ustvarimo novo tabelo, iz katere te zapise izločimo. No, mogoča je tudi druga pot: ustvarimo skupino C_D nepopolnih zapisov in iz nje matriko \mathbf{D} – $\text{idents} \times \text{idents}$. Matrika \mathbf{D} je diagonalna matrika z vrednostjo 1 za popolne zapise in vrednostjo 0 za zapise iz C_D . Z matriko \mathbf{D} lahko na primer določimo omrežje `ProjInst.net` iz celotne tabele kot $\mathbf{W} = \mathbf{P}^T \star \mathbf{D} \star \mathbf{U}$ – nepopolni zapisi ne prispevajo k omrežju.



Analiza omrežja ProjInst.net

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

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

```
Network/Create New Network/With Ring Counts .../4-Rings/Undirected  
Network/Create Partition/Islands/Line Weights[Simple] [2,200]
```

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

V zapisu nekaterih oznak vozlišč smo uporabili $\backslash n$, ki oznako prelomi. Za analizo bi lahko uporabili tudi (p, q) —sredice.



Analiza omrežja ProjInst.net

Analiza
omrežij

V. Batagelj

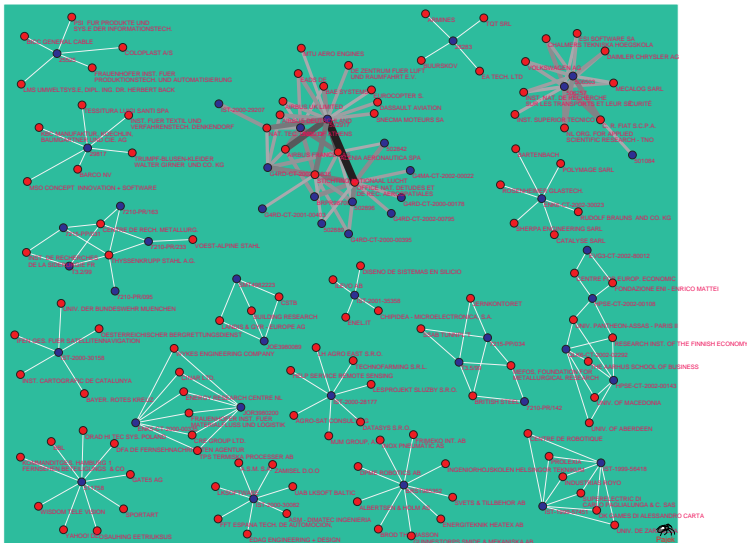
Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja



V. Batagelj

Analiza omrežij



Analiza omrežja Countries.net

Analiza omrežij

V. Batagelj

Dvovrstna omrežja

IMDB

Množenje omrežij

Evropski projekti

Bibliografska omrežja



Omrežje Countries.net ima le 60 vozlišč, je pa gosto. Uteži povezav predstavljajo število projektov, pri katerih hkrati sodelujeta krajiščni državi. Za preglednejšo sliko moramo povezave urediti glede na uteži: Network/Create New Network/Transform/Sort lines/Line values/Ascending

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

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

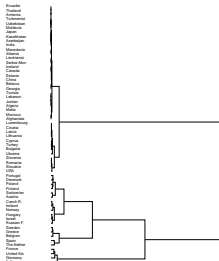
IMDB

Množenje
omrežij

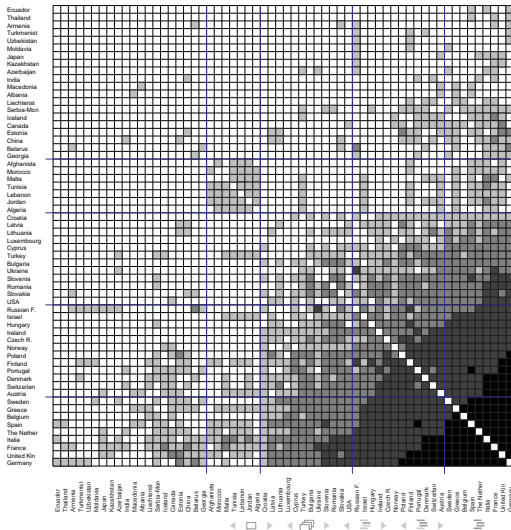
Evropski
projekti

Bibliografska
omrežja

Pajek - Ward [0.00,4785.14]

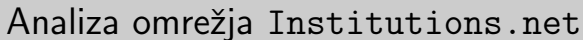


Pajek - shadow [0.00,4.00]



V. Batagelj

Analiza omrežij



Analiza omrežij

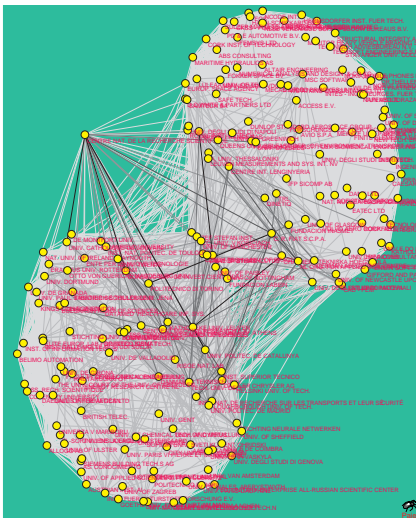
V. Batagelj

Dvovrstna omrežja

IMDB

Množenje omrežij

Evropski
projekti

Bibliografska
omrežja

Za določitev najpomembnejših ustanov smo najprej določili sredice vrste p_5 in za tako določen vektor določili vozliščne 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 odstopajoče vozlišče je Centre national de la recherche scientifique (CNRS). Še boljše je medsebojna povezanost med ustanovami razvidna iz matričnega prikaza.



Analiza omrežja Institutions.net

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

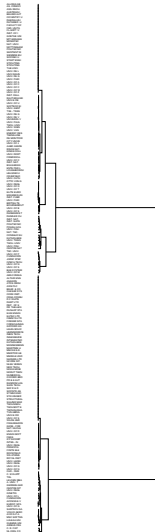
IMDB

Množenje
omrežij

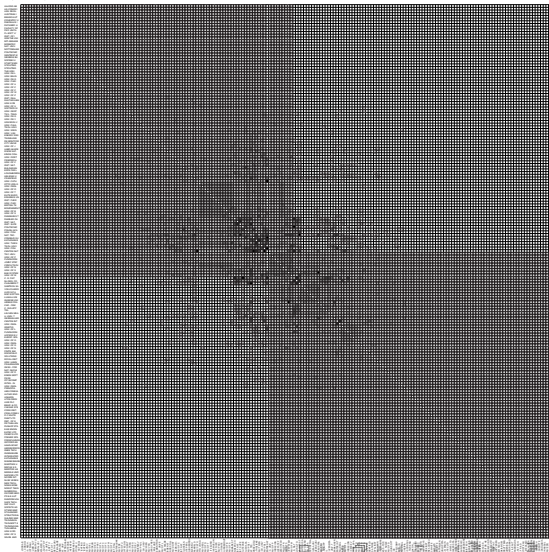
Evropski
projekti

Bibliografska
omrežja

Pajek - Ward [0.00,1376.93]



Pajek - shade [0.00,6.00]



V. Batagelj

Analiza omrežij



Pretvorba dvovrstnih omrežij na enovrstna

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

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.



Normalizacije

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Normalizacije naj bi omogočile hiter pregled enovrstnih omrežij, ki jih dobimo iz dvovrstnih.

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

Obstaja več načinov, kako lahko to naredimo. Nekaj izmed njih je prikazanih v tabeli na naslednji prosojnici. Uporabimo jih lahko tudi na drugih enovrstnih uteženih omrežjih.

V primeru omrežij brez zank postavimo za neusmerjena omrežja diagonalne vrednosti na vsoto izvendiagonalnih elementov v pripadajoči vrstici (ali stolpcu) $w_{vv} = \sum_{u \neq v} w_{vu}$; za usmerjena omrežja pa neko srednjo vrednost iz vrstične in stolpčne vsote – npr. $w_{vv} = \frac{1}{2}(\sum_{u \neq v} w_{vu} + \sum_{u \neq v} w_{uv})$. Običajno privzamemo, da omrežje nima osamljenih vozlišč.



... Normalizacije

Analiza omrežij

V. Batagelj

Dvovrstna omrežja

IMDB

Množenje omrežij

Evropski projekti

Bibliografska omrežja

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

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

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

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

$$\text{GeoDeg}_{uv} = \frac{w_{uv}}{\sqrt{\deg_u \deg_v}}$$

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

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

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

Normalizirano omrežje analiziramo z uporabo povezavnih prerezov ali otokov.

Network/2-Mode Network/2-Mode to 1-Mode/Normalize 1-Mode/

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



MinDir na omrežju Slovenski časopisi 2000

Analiza
omrežij

V. Batagelj

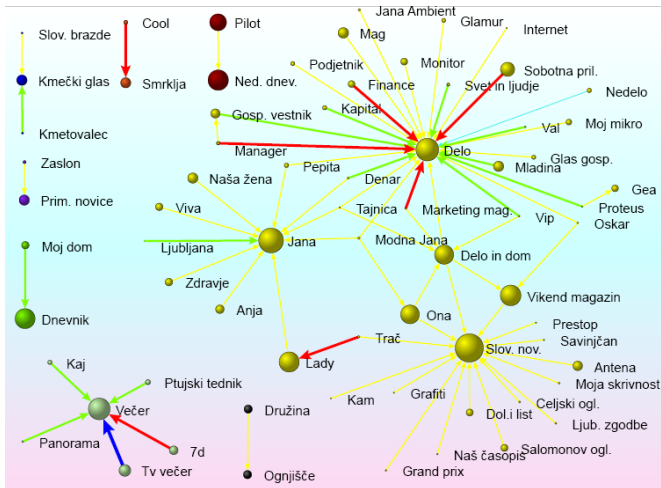
Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja



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



GeoDeg na omrežju Reuters terror news

Analiza
omrežij

V. Batagelj

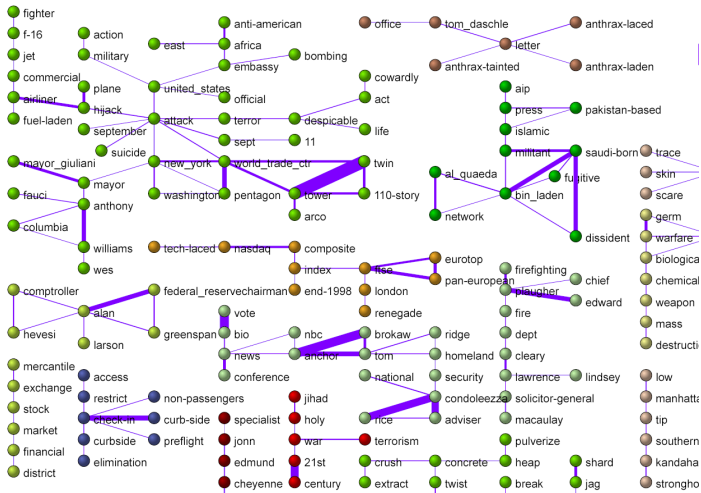
Dvovrstna
omrežja

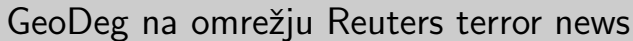
IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja







Omrežje avtorstev

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

V dvovrstnem **omrežju avtorstev** **WA** na dela \times avtorji imajo povezave (p, i) utež $wa_{pi} = 1$ in povedo, da je oseba i (so)avtor dela p .

$$\forall p \in W : \sum_{i \in A} wa_{pi} = \text{outdeg}_{WA}(p) = \# \text{ avtorjev dela } p$$

Označimo z **N** njegovo normalizirano različico

$$\forall p \in W : \sum_{i \in A} n_{pi} \in \{0, 1\}$$

ki jo dobimo iz **WA** z $n_{pi} = wa_{pi} / \max(1, \text{outdeg}_{WA}(p))$ ali pa po kakem drugem pravilu, ki določa delež posameznega avtorja.

$$i \sim p \Leftrightarrow i \in N(p) \Leftrightarrow p \in N^-(i) \Leftrightarrow wa_{pi} > 0$$



Nekaj transformacij omrežij

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Graf omrežja $b(\mathcal{N})$ je omrežje, ki ga dobimo iz danega omrežja \mathcal{N} tako, da vse njegove uteži postavimo na 1.

Transpozicija ali preusmeritev \mathcal{N}^T ali $t(\mathcal{N})$ je omrežje, ki ga dobimo iz danega omrežja \mathcal{N} tako, da obrnemo smeri vseh njegovih usmerjenih povezav. $\mathbf{AW} = \mathbf{WA}^T$, $\mathbf{KW} = \mathbf{WK}^T$, ...

(Izhodna) normalizacija $n(\mathcal{N})$ je omrežje, ki ga dobimo iz danega omrežja \mathcal{N} tako, da utež vsake povezave a delimo z vsoto uteži vseh povezav, ki imajo isto začetno vozlišče kot povezava a . Za omrežja s konstantno utežjo velja

$$n(\mathbf{A}) = \text{diag}\left(\frac{1}{\max(1, \text{outdeg}_{\mathbf{WA}}(i))}\right)_{i \in \mathcal{I}} * \mathbf{A}$$

$$\mathbf{N} = n(\mathbf{WA}), \mathbf{WA} = b(\mathbf{N})$$



Prvo omrežje sodelovanj

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

$$\mathbf{Co} = \mathbf{AW} * \mathbf{WA}$$

$$co_{ij} = \sum_{p \in W} wa_{pi} wa_{pj} = \sum_{p \in N^-(i) \cap N^-(j)} 1$$

co_{ij} = število del, ki sta jih avtorja i in j napisala skupaj

Velja: $co_{ij} = co_{ji}$.

Z uporabo uteži co_{ij} lahko izrazimo Saltonovo kosinusno podobnost imenovano tudi Ochiai-jev koeficient med avtorjema i in j

$$\cos(i, j) = \frac{co_{ij}}{\sqrt{co_{ii} co_{jj}}}, \quad \text{for } co_{ij} > 0$$

Težava: Kot smo že povedali, je omrežje **Co** lepljenka polnih podgrafov. Dela z velikim številom avtorjev ustvarijo velike polne podgrafe in zameglijo sliko sodelovanj.



Sredice redov 10–21 v Computational Geometry

Analiza
omrežij

V. Batagelj

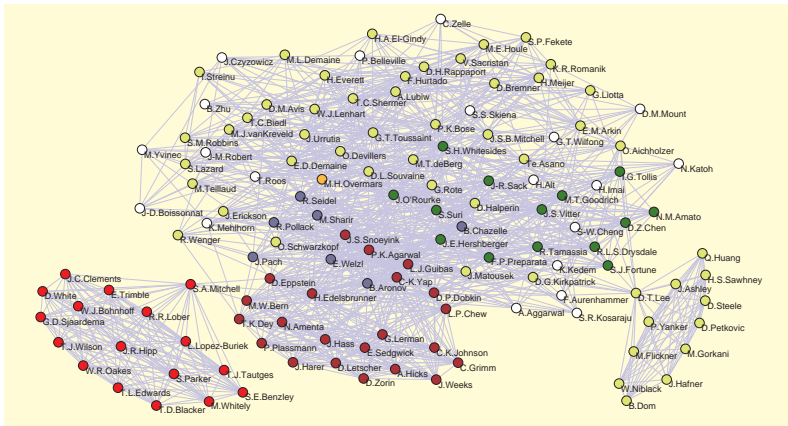
Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja





p_5 -sredica za prag 46 v Computational Geometry

Analiza
omrežij

V. Batagelj

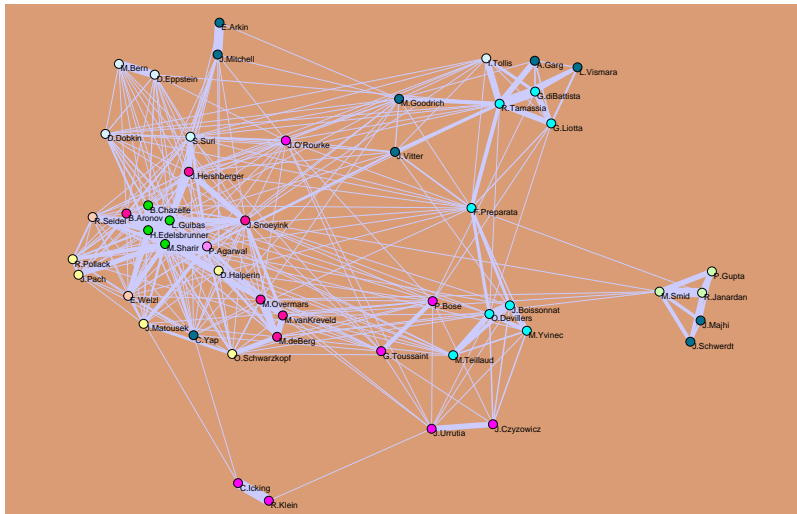
Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja





Sredice redov 20–47 v **Co(SN5)**

Analiza
omrežij

V. Batagelj

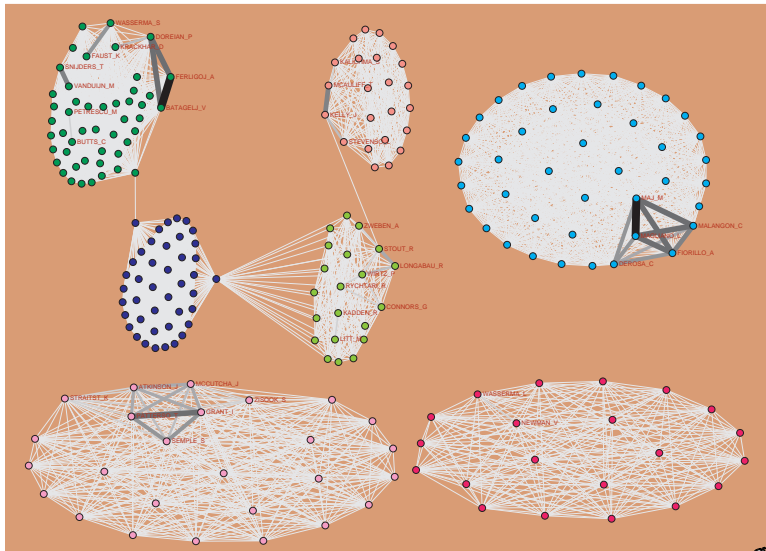
Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja



V. Batagelj

Analiza omrežij



Članki glede na število avtorjev v SN5

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

outdeg	frequency	outdeg	frequency	paper
1	2637	12	8	
2	2143	13	4	
3	1333	14	3	
4	713	15	2	
5	396	21	1	Pierce et al. (2007)
6	206	22	1	Allen et al. (1998)
7	114	23	1	Kelly et al. (1997)
8	65	26	1	Semple et al. (1993)
9	43	41	1	Magliano et al. (2006)
10	24	42	1	Doll et al. (1992)
11	10	48	1	Snijders et al. (2007)



Snijders et al. (2007)

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Snijders et al.(2007): Snijders, T.A.B., Robinson, T., Atkinson, A.C., Riani, M., Gormley, I.C., Murphy, T.B., Sweeting, T., Leslie, D.S., Longford, N.T., Kent, J.T., Lawrance, T., Airolidi, E.M., Besag, J., Blei, D., Fienberg, S.E., Breiger, R., Butts, C.T., Doreian, P., Batagelj, V., Ferligoj, A., Draper, D., van Duijn, M.A.J., Faust, K., Petrescu-Prahova, M., Forster, J.J., Gelman, A., Goodreau, S. M., Greenwood, P.E., Gruenberg, K., Francis, B., Hennig, C., Hoff, P.D., Hunter, D.R., Husmeier, D., Glasbey, C., Krackhardt, D., Kuha, J., Skrandal, A., Lawson, A., Liao, T. F., Mendes, B., Reinert, G., Richardson, S., Lewin, A., Titterington, D.M., Wasserman, S., Werhli, A.V. and Ghazal, P.. *Discussion on the paper by Handcock, Raftery and Tantrum*. Journal of the Royal Statistical Society: Series A - Statistics in Society, 170 (2007), pp. 322-354.



Drugo omrežje sodelovanj

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

$$\mathbf{Cn} = \mathbf{AW} * \mathbf{N}$$

$$cn_{ij} = \sum_{p \in W} wa_{pi} n_{pj} = \sum_{p \in N^-(i) \cap N^-(j)} n_{pj}$$

cn_{ij} = prispevek avtorja j k delom, ki jih napisal skupaj z avtorjem i .

Velja $\sum_{j \in A} \sum_{j \in A} wa_{pi} n_{pj} = \text{outdeg}_{WA}(p)$ in $\sum_{j \in A} cn_{ij} = \text{indeg}_{WA}(i)$

Če je $n_{pi} = \frac{wa_{pi}}{\text{outdeg}(p)}$, velja še $cn_{ij} = cn_{ji}$.

$cn_{ii} = \sum_{p \in N(i)} n_{pi}$ je prispevek avtorja i svojim delom.

Samozadostnost: $S_i = \frac{cn_{ii}}{\text{outdeg}_{WA}(i)}$

Sodelovalnost: $K_i = 1 - S_i$

$$\sum_{i \in A} \sum_{j \in A} cn_{ij} = \sum_{i \in A} \text{indeg}_{WA}(i) = m_{WA}$$

Za izračun tabele smo v programu Pajek pripravili makro.



"Najboljši" avtorji na področju Social Networks

Analiza omrežij

V. Batagelj

Dvovrstna omrežja

IMDB

Množenje omrežij

Evropski projekti

Bibliografska omrežja

i	author	cn_{ij}	total	K_i	i	author	cn_{ij}	total	K_i
1	Burt,R	43.83	53	0.173	26	Latkin,C	10.14	37	0.726
2	Newman,M	36.77	60	0.387	27	Morris,M	9.98	20	0.501
3	Doreian,P	34.44	47	0.267	28	Rothenberg,R	9.82	28	0.649
4	Bonacich,P	30.17	41	0.264	29	Kadushin,C	9.75	11	0.114
5	Marsden,P	29.42	37	0.205	30	Faust,K	9.72	18	0.460
6	Wellman,B	26.87	41	0.345	31	Batagelj,V	9.69	20	0.516
7	Leydesdorf,L	24.37	35	0.304	32	Mizruchi,M	9.67	15	0.356
8	White,H	23.50	33	0.288	33	[Anon]	9.00	9	0.000
9	Friedkin,N	20.00	23	0.130	34	Johnson,J	8.89	21	0.577
10	Borgatti,S	19.20	41	0.532	35	Fararo,T	8.83	16	0.448
11	Everett,M	16.92	31	0.454	36	Lazega,E	8.50	12	0.292
12	Litwin,H	16.00	21	0.238	37	Knoke,D	8.33	11	0.242
13	Freeman,L	15.53	20	0.223	38	Ferligoj,A	8.19	19	0.569
14	Barabasi,A	14.99	35	0.572	39	Brewer,D	8.03	11	0.270
15	Snijders,T	14.99	30	0.500	40	Klov Dahl,A	7.96	17	0.532
16	Valente,T	14.80	34	0.565	41	Hammer,M	7.92	10	0.208
17	Breiger,R	14.44	20	0.278	42	White,D	7.83	15	0.478
18	Skvoretz,J	14.43	27	0.466	43	Holme,P	7.42	14	0.470
19	Krackhardt,D	13.65	25	0.454	44	Boyd,J	7.37	13	0.433
20	Carley,K	12.93	28	0.538	45	Kilduff,M	7.25	16	0.547
21	Pattison,P	12.10	27	0.552	46	Small,H	7.00	7	0.000
22	Wasserman,S	11.72	26	0.549	47	Iacobucci,D	7.00	12	0.417
23	Berkman,L	11.21	30	0.626	48	Pappi,F	6.83	10	0.317
24	Moody,J	10.83	15	0.278	49	Chen,C	6.78	12	0.435
25	Scott,J	10.47	15	0.302	50	Seidman,S	6.75	9	0.250



Tretje omrežje sodelovanj

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

$$\mathbf{Ct} = \mathbf{N}^T * \mathbf{N}$$

ct_{ij} = skupni prispevek sodelovanja avtorjev i in j delom.

Velja $ct_{ij} = ct_{ji}$ in

$$\sum_{i \in A} \sum_{j \in A} n_{pi} n_{pj} = 1$$

Skupni prispevek polnega podomrežja, ki pripada posameznemu članku p , je enak 1.

$\sum_{j \in A} ct_{ij} = \sum_{p \in W} n_{pi} =$ skupni prispevek avtorja i delom iz W .

$$\sum_{i \in A} \sum_{j \in A} ct_{ij} = |W_{\mathbf{WA}}^+|$$

Prispevek avtorja i : $ac_i = \frac{|A|}{|W_{\mathbf{WA}}^+|} \sum_{p \in W} n_{pi}$ in $\sum_{i \in A} ac_i = |A|$.

Torej je povprečni avtorjev prispevek ac enak 1.

Vsa tri omrežja sodelovanj imajo isti graf: $b(\mathbf{Co}) = b(\mathbf{Cn}) = b(\mathbf{Ct})$.

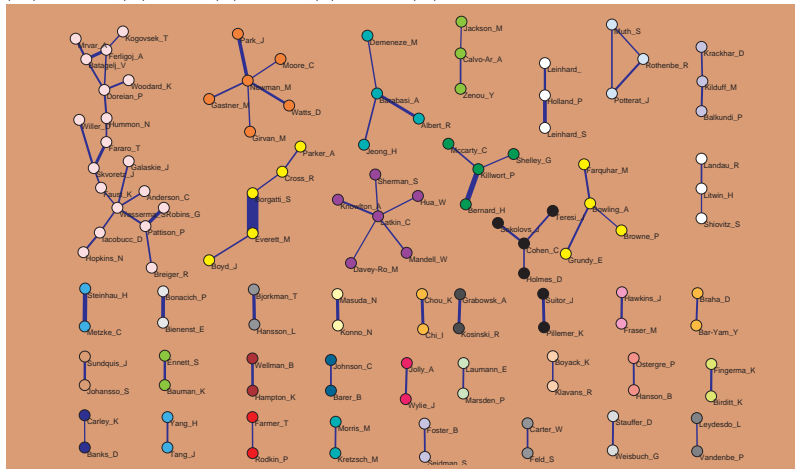


Komponente **Ct(SN5)** za izrez pri pragu 0.5

Analiza
omrežij

V. Batagelj

Omrežje SN5 (2008): za "social network*" + članki z največ sklici nanje + okrog 100 raziskovalcev;
 $|W| = 193376$, $|C| = 7950$, $|A| = 75930$, $|J| = 14651$, $|K| = 29267$





p_S -sredica za prag 0.75 v Ct(SN5)

Analiza
omrežij

V. Batagelj

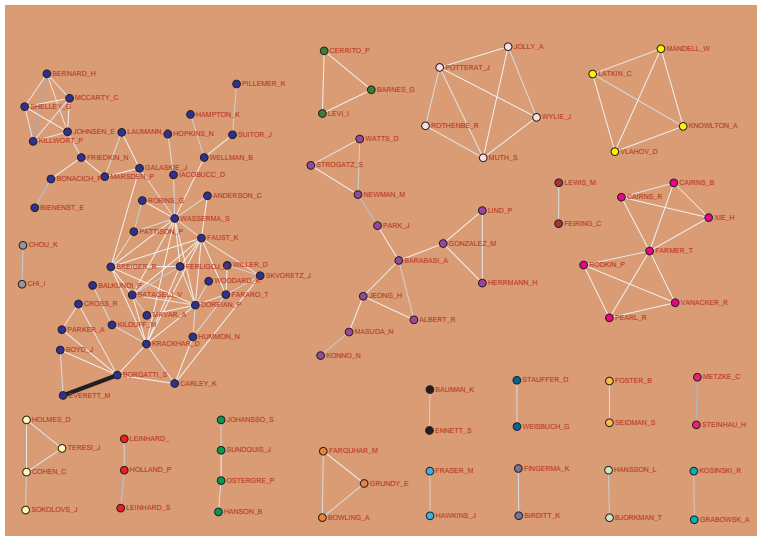
Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja



V. Batagelj

Analiza omrežij



Omrežje sklicevanj med avtorji

Analiza
omrežij

V. Batagelj

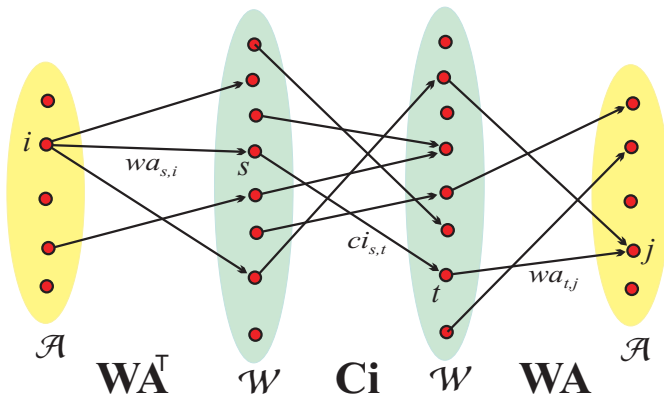
Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja



$Ca = AW * Ci * WA$ je **omrežje sklicevanj med avtorji**. Utež ca_{ij} šteje kolikokrat se kako delo, ki ga je soustvaril avtor i , sklicuje na delo, ki ga je soustvaril avtor j .



Otoki v $\text{Ca}(\text{SN5})$

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

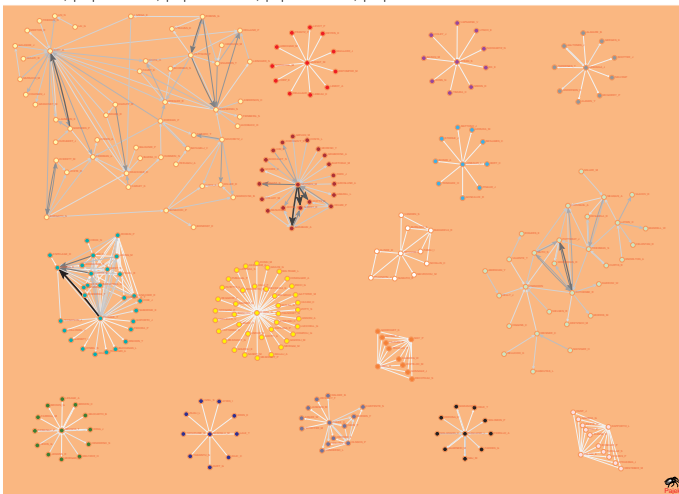
IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Omrežje SN5 (2008): za "social network*" + članki z največ sklici nanje + okrog 100 raziskovalcev;
 $|W| = 193376$, $|C| = 7950$, $|A| = 75930$, $|J| = 14651$, $|K| = 29267$





(Sklicna) sklopljenost

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

V programu WoS2Pajek ima relacija sklicovanja **Ci** naslednji pomen
 $p\mathbf{Ci}q \equiv$ delo p se sklicuje na delo q . Zato je omrežje (*sklicne*)
sklopljenosti **biCo** določeno z

$$\mathbf{biCo} = \mathbf{Ci} * \mathbf{Ci}^T$$

$bico_{pq} = \#$ del, na katera se sklicujeta obe deli p in q . $bico_{pq} = bico_{qp}$.
Zopet so težave z deli z dolgimi seznamami virov – posebej s preglednimi
članki. Da bi omilili njihov vpliv, lahko uvedemo normalizirane mere kot je

$$\mathbf{biCon} = \frac{1}{2}(n(\mathbf{Ci}) * \mathbf{Ci}^T + \mathbf{Ci} * n(\mathbf{Ci})^T)$$

Velja $bicon_{pq} \in [0, 1]$ in $bicon_{pq} = bicon_{qp}$ (simetričnost). Velja še:
 $bicon_{pq} = 1$ ntk deli p in q se sklicujeta na ista dela.



Omrežje sosclicov in druga

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Omrežje *sosclicov* **coCi** je določeno z

$$\mathbf{coCi} = \mathbf{Ci}^T * \mathbf{Ci}$$

$coci_{pq} = \#$ del, ki se sklicujejo hkrati na deli p in q . $coci_{pq} = coci_{qp}$.

Utež $w(a, p)$ v omrežju *avtorskih sklicev*

$$\mathbf{ACi} = \mathbf{AW} * \mathbf{Ci}$$

šteje kolikokrat se je avtor a sklical na delo p .

Omrežje *avtorskih sosclicov* dobimo z

$$\mathbf{ACo} = b(\mathbf{ACi}) * b(\mathbf{ACi})^T$$

Omrežje *Avtorske rabe gesel* pa z $\mathbf{AK} = \mathbf{AW} * \mathbf{WK}$.



Druga izpeljana omrežja

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

Sklicevanja med revijami

$$\mathbf{CJt} = \mathbf{WJ}^T * n(\mathbf{Ci}) * \mathbf{WJ}$$

Avtorjev prispevek reviji

$$\mathbf{AJn} = \mathbf{N}^T * \mathbf{WJ}$$

Prispevek avtorjev prve revije drugi reviji

$$\mathbf{JJr} = b(\mathbf{AJn})^T * \mathbf{AJn}$$



Prekrivanje in ohranjanje vsebine

Analiza
omrežij

V. Batagelj

Dvovrstna
omrežja

IMDB

Množenje
omrežij

Evropski
projekti

Bibliografska
omrežja

V analizi omrežij sklicevanj se izkaže za koristno operacija $\mathcal{N}_A \triangleright \mathcal{N}_B$ **prekrivanja** omrežja \mathcal{N}_B z omrežjem \mathcal{N}_A .

Za omrežji $\mathcal{N}_A = (\mathcal{V}, \mathcal{A}_A, w_A)$ in $\mathcal{N}_B = (\mathcal{V}, \mathcal{A}_B)$ je $\mathcal{N}_A \triangleright \mathcal{N}_B = \mathcal{N}_C = (\mathcal{V}, \mathcal{A}_C, w_C)$, kjer je $\mathcal{A}_C = \mathcal{A}_A \cap \mathcal{A}_B$ in $(u, v) \in \mathcal{A}_C \Rightarrow w_C(u, v) = w_A(u, v)$.

Networks/Cross-Intersection/First

Z uporabo prekrivanja lahko vpeljamo omrežje **ohranjanja vsebine** $\mathbf{CiK} = (\mathbf{WK} * \mathbf{WK}^T) \triangleright \mathbf{Ci}$. Utež cik_{pq} = število skupnih gesel del p in q .