



# Analiza omrežij

## 4. Zgradba omrežij podomrežja

Vladimir Batagelj

Univerza v Ljubljani, FMF, matematika

Interdisciplinarni doktorski študijski program Statistika  
Ljubljana, april 2014



# Kazalo

Analiza  
omrežij

V. Batagelj

Statistika

Homomorfizmi

Razbitja

Skrčitev

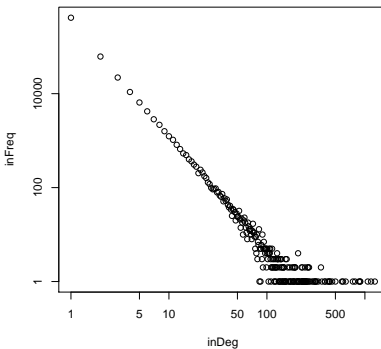
Izrezi

Prerezi

- 1 Statistika
- 2 Homomorfizmi
- 3 Razbitja
- 4 Skrčitev
- 5 Izrezi
- 6 Prerezi

wiki: <http://pajek.imfm.si/doku.php?id=event:pd>  
April 14, 2014/ marec 2013

in-degree distribution





# Pristopi k velikim omrežjem

## Analiza omrežij

V. Batagelj

Statistika

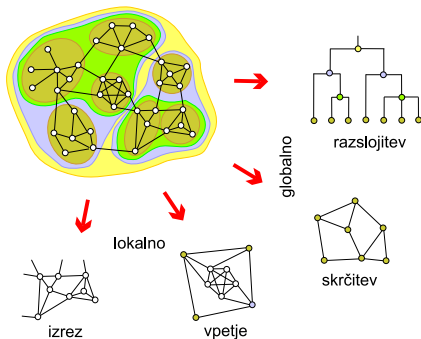
Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi

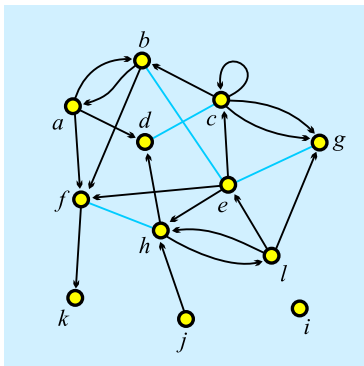


Pri *velikih* omrežjih (več tisoč ali milijonov vozlišč, omrežje je mogoče shraniti v pomnilniku) se moramo odpovedati celoviti sliki, uporabni so le redki postopki.

Za analizo velikih omrežij lahko uporabimo statistiko ali pa zanimiva mala in srednja podomrežja.



# Stopnje vozlišč grafa



**stopnja** vozlišča  $v$ ,  $\deg(v)$  = je število povezav, ki imajo vozlišče  $v$  za krajišče;

**vhodna stopnja** vozlišča  $v$ ,  $\text{indeg}(v)$  = je število povezav, ki imajo vozlišče  $v$  za konec (krajišče neusmerjene povezave je hkrati njen začetek in konec);

**izhodna stopnja** vozlišča  $v$ ,  $\text{outdeg}(v)$  = je število povezav, ki imajo vozlišče  $v$  za začetek.

$$n = 12, m = 23, \text{indeg}(e) = 3, \text{outdeg}(e) = 5, \deg(e) = 6$$

$$\sum_{v \in \mathcal{V}} \text{indeg}(v) = \sum_{v \in \mathcal{V}} \text{outdeg}(v) = |\mathcal{A}| + 2|\mathcal{E}|, \sum_{v \in \mathcal{V}} \deg(v) = 2|\mathcal{L}| - |\mathcal{L}_0|$$



## Vhodni podatki o vozliščih

- številski  $\rightarrow$  vector
- urejenostni  $\rightarrow$  permutation
- imenski  $\rightarrow$  clustering (razbitje)

## Izračunane lastnosti vozlišč

*globalne*: število vozlišč, usmerjenih/neusmerjenih povezav, komponent; največje sredično število, ...

*lokalne*: stopnje, sredična števila, indeksi (vmesnost, dostopnost, kazala in viri, ...)

*pregledi*: razbitja, vektorji, vrednosti povezav, ...

Analiza povezanosti med izračunanimi (strukturnimi) lastnostmi in vhodnimi (izmerjenimi) lastnostmi.



# ... Statistika

Analiza  
omrežij

V. Batagelj

Statistika

Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi

Globalne lastnosti izpišejo Pajekovi ukazi v poročilo; največ jih je dosegljivih v izbiri Info. Pri uporabi *ponavljajočih* ukazov se shranijo v vektorje.

Lokalne lastnosti izračunajo razni Pajekovi ukazi in jih shranijo v vektorje ali razbitja. Njihove vrednosti / porazdelitev si lahko ogledamo v izbiri Info.

Za primer si oglejmo omrežje [The Edinburgh Associative Thesaurus](#). EAT je omrežje asociacij med besedami zbranih na študentski populaciji. Točke so besede. Povezave  $(X, Y)$  pa so določene z vprašanjem: Katera beseda  $Y$  vam pride prva na misel, ko slišite besedo  $X$ ? Utež povezave pove, kolikokrat je bila izbrana.

```
File/Network/Read eatRS.net  
Info/Network/General
```

Ima 23219 vozlišč in 325624 usmerjenih povezav (564 zank); 227481 povezav ima utež 1.



# ... Statistika

Analiza  
omrežij

V. Batagelj

Statistika

Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi

Točke z največjimi stopnjami dobimo takole:

```
Network/Create Partition/Degree/All  
Partition/Copy to Vector  
Info/Vector +10
```

V EAT so to:

	vertex	deg	label
1	12720	1108	ME
2	12459	1074	MAN
3	8878	878	GOOD
4	18122	875	SEX
5	13793	803	NO
6	13181	799	MONEY
7	23136	732	YES
8	15080	723	PEOPLE
9	13948	720	NOTHING
10	22973	716	WORK



# Pajek in R

Analiza  
omrežij

V. Batagelj

Statistika

Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi

Pajek 0.89 (in kasnejši) omogoča uporabo statističnega programa R in tudi drugih programov kot orodij (izbira Tools).  
V programu Pajek določimo stopnje vozlišč in jih 'podtaknemo' R-ju

```
Network/Info/General  
Network/Create Vector/Centrality/Degree/All  
Tools/R/Send to R/Current Vector
```

Tu določimo porazdelitev stopenj in jo narišemo

```
summary(v2)  
t <- table(v2)  
x<-as.numeric(names(t))  
plot(x,t,log='xy',main='degree distribution',  
      xlab='deg',ylab='freq')
```

Dobljeno sliko lahko s File/Save as shranimo v izbrani obliki (PDF ali PS za  $\LaTeX$ ; Windows metafile za Word).

Pozor! Vozlišča stopnje 0 delajo težave pri log='xy'.





# EAT – porazdelitev stopenj

Analiza omrežij

V. Batagelj

Statistika

Homomorfizmi

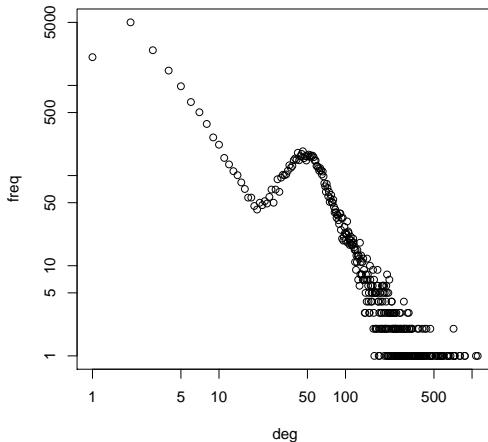
Razbitja

Skrčitev

Izrezi

Prerezi

EAT all-degree distribution





# Slučajni grafi

## Analiza omrežij

V. Batagelj

Statistika

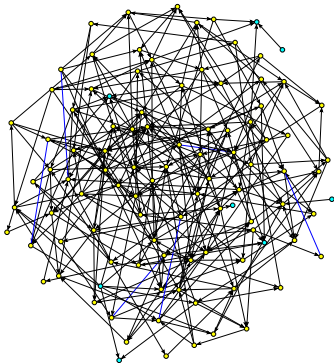
Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi



Erdős in Rényi sta definirala *slučajni graf* takole: vsako mogočo povezavo vključimo v slučajni graf z dano verjetnostjo  $p$ .

V programu Pajek v Network/Create Random Network/Bernoulli/Poisson/Undirected/General [100] [2.5] uporabljamo namesto verjetnosti  $p$  povprečno stopnjo

$$\overline{\deg} = \frac{1}{n} \sum_{v \in V} \deg(v)$$

Velja  $p = \frac{m}{m_{\max}}$  in, za grafe brez zank, še  $\overline{\deg} = \frac{2m}{n}$ .

Na sliki je prikazan slučajni graf na 100 vozliščih z  $\overline{\deg} = 3$ .



# Porazdelitve stopenj

Analiza omrežij

V. Batagelj

Statistika

Homomorfizmi

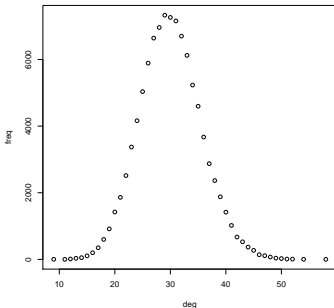
Razbitja

Skrčitev

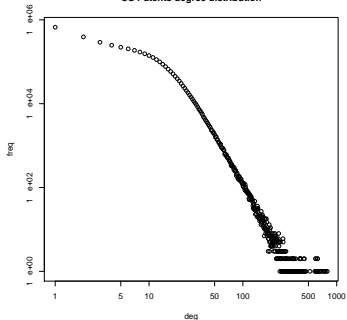
Izrezi

Prerezi

Random graph degree distribution,  $n=100000$ ,  $\text{degav}=30$



US Patents degree distribution



Dejanska omrežja so vse prej kot slučajna. Analiza porazdelitev je dala nov pogled na zgradbo dejanskih omrežij – Watts (**Small worlds**), Barabási (**nd/networks, Linked**).



# Porazdelitve v/iz-hodnih stopenj

Analiza  
omrežij

V. Batagelj

Statistika

Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi

V Pajek-a preberemo omrežje sklicevanj cite.net in odstranimo zanke ter večkratne povezave. Nato določimo vhodne in izhodne stopnje ter iz Pajek-a pokličemo R in mu 'podtaknemo' vse vektorje.

```
#####  
R called from Pajek  
The following vectors read:  
v3 : From partition 1 (548600)  
v4 : From partition 2 (548600)  
-----  
> inTab <- table(v3)  
> indeg <- as.integer(names(inTab))  
> inDeg <- indeg[indeg>0]  
> inFreq <- as.vector(inTab[indeg>0])  
> plot(inDeg,inFreq,log='xy',main="in-degree distribution")  
> ouTab <- table(v4)  
> outdeg <- as.integer(names(ouTab))  
> outDeg <- outdeg[outdeg>0]  
> outFreq <- as.vector(ouTab[outdeg>0])  
> plot(outDeg,outFreq,log='xy',main="out-degree distribution")
```



# Porazdelitve v/iz-hodnih stopenj

## Analiza omrežij

V. Batagelj

Statistika

Homomorfizmi

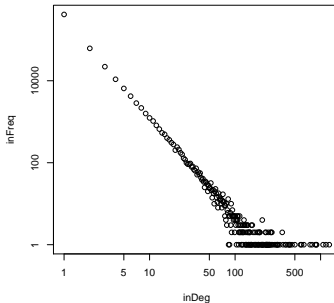
Razbitja

Skrčitev

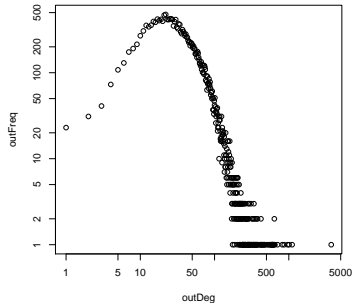
Izrezi

Prerezi

in-degree distribution



out-degree distribution



Porazdelitev vhodnih stopenj kaže na brezlestvičnost (scale-free).  
Parametre lahko ocenimo s paketom, ki so ga pripravili [Clauset, Shalizi and Newman](#). Glejte še [Stumpf, et al.: Critical Truths About Power Laws](#).



# Števila objav člankov po letih

omrežja na temo središčnosti

Analiza omrežij

V. Batagelj

Statistika

Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi

Iz datoteke `Year.clu`, ki vsebuje letnice objav člankov, lahko dobimo porazdelitve *števila objav člankov po letih*. Za omrežja na temo središčnosti smo dobili:

```
> setwd("C:/Users/Batagelj/work/Python/WoS/Central")
> years <- read.table(file="Year.clu",header=FALSE,skip=2)$V1
> t <- table(years)
> year <- as.integer(names(t))
> freq <- as.vector(t[1950<=year & year<=2009])
> y <- 1950:2009
> plot(y,freq)
> model <- nls(freq~c*dlnorm(2010-y,a,b),start=list(c=350000,a=2,b=0.7))
> model
Nonlinear regression model
  model: freq ~ c * dlnorm(2010 - y, a, b)
  data: parent.frame()
5.427e+05 2.491e+00 6.624e-01
residual sum-of-squares: 20474181

Number of iterations to convergence: 7
Achieved convergence tolerance: 3.978e-06
> lines(y,predict(model,list(x=2010-y)),col='red')
```

Porazdelitev lahko dobro povzamemo *logaritemsko normalno porazdelitvijo*, toda tudi s funkcijo  $c * (x + d)^{\frac{a}{b+x}}$ .





# Števila objav člankov po letih omrežja na temo središčnosti

Analiza  
omrežij

V. Batagelj

Statistika

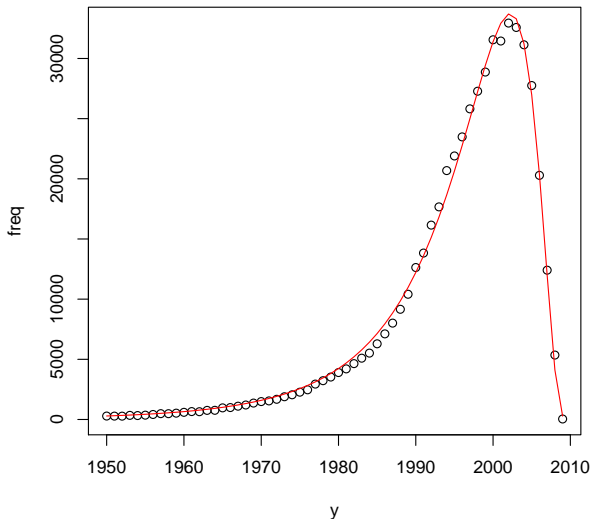
Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi





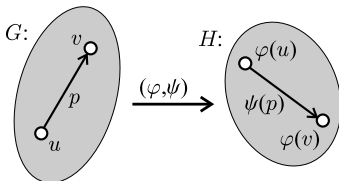
# Povezanosti med grafi

Preslikavi  $(\varphi, \psi)$ ,  $\varphi: \mathcal{V} \rightarrow \mathcal{V}'$  in  $\psi: \mathcal{L} \rightarrow \mathcal{L}'$  določata **šibki homomorfizem** grafa  $\mathcal{G} = (\mathcal{V}, \mathcal{L})$  v graf  $\mathcal{H} = (\mathcal{V}', \mathcal{L}')$  ntk velja:

$$\forall u, v \in \mathcal{V} \forall p \in \mathcal{L} : (p(u : v) \Rightarrow \psi(p)(\varphi(u) : \varphi(v)))$$

in določa **(krepki) homomorfizem** grafa  $\mathcal{G}$  v graf  $\mathcal{H}$  ntk velja:

$$\forall u, v \in \mathcal{V} \forall p \in \mathcal{L} : (p(u, v) \Rightarrow \psi(p)(\varphi(u), \varphi(v)))$$



Ko sta  $\varphi$  in  $\psi$  bijekciji in ustrezní pogoj velja v obe smeri, govorimo o **izomorfizmu** grafov  $\mathcal{G}$  in  $\mathcal{H}$ . Da sta grafa šibko izomorfna zapišemo  $\mathcal{G} \sim \mathcal{H}$ ; da sta (krepko) izomorfna pa  $\mathcal{G} \approx \mathcal{H}$ . Velja  $\approx \subset \sim$ .

**Stalnica** ali **invarianta** grafa imenujemo vsako grafu prirejeno število, ki je enako za vse med seboj izomorfne grafe. .





# Homomorfizem

Analiza omrežij

V. Batagelj

Statistika

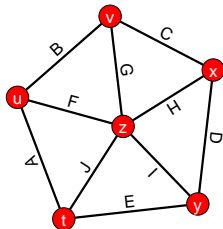
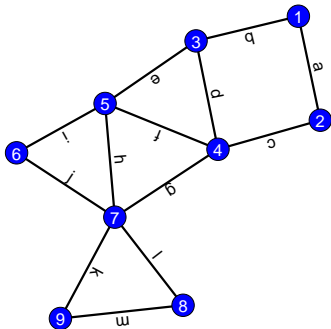
Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi



$\varphi$	1	2	3	4	5	6	7	8	9				
	x	y	t	z	x	z	t	z	x				
$\psi$	a	b	c	d	e	f	g	h	i	j	k	l	m
	D	E	H	J	E	I	J	E	I	J	E	J	I



# Izomorfna grafa

Analiza omrežij

V. Batagelj

Statistika

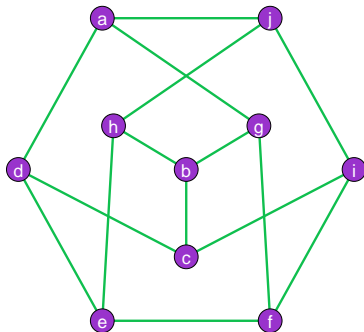
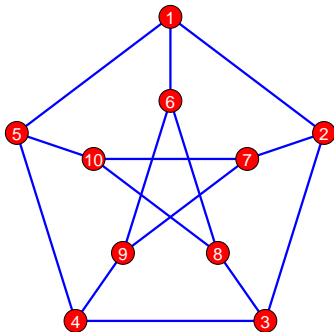
Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi



$\varphi$	1	2	3	4	5	6	7	8	9	10
	b	h	j	a	g	c	e	i	d	f



# Skupine, razvrstitve, razbitja, razslojitve

Analiza  
omrežij

V. Batagelj

Statistika

Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi

Neprazno podmnožico  $C \subseteq \mathcal{V}$  imenujemo *skupina*. Neprazna množica skupin  $\mathbf{C} = \{C_i\}$  je *razvrstitev*.

Razvrstitev  $\mathbf{C} = \{C_i\}$  je *razbitje* ntk

$$\cup \mathbf{C} = \bigcup_i C_i = \mathcal{V} \quad \text{in} \quad i \neq j \Rightarrow C_i \cap C_j = \emptyset$$

Razvrstitev  $\mathbf{C} = \{C_i\}$  je *razslojitev* ali *hierarhija* ntk

$$C_i \cap C_j \in \{\emptyset, C_i, C_j\}$$

Razslojitev  $\mathbf{C} = \{C_i\}$  je *polna*, če je  $\cup \mathbf{C} = \mathcal{V}$ ; in je *osnovna*, če je za vsak  $v \in \cup \mathbf{C}$  tudi  $\{v\} \in \mathbf{C}$ .



# Primer razbitja in razslojitve

Množica vozlišč:

$$\mathcal{V} = \{a, b, c, d, e, f, g\}$$

Razbitje:

$$\mathbf{C} = \{\{a, b, e\}, \{c, g\}, \{d, f\}\}$$

Skupina:

$$C_2 = \{c, g\}$$

Razslojitev:

$$\mathbf{H} = \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\}, \{g\}, \\ \{a, e\}, \{c, g\}, \{d, f\}, \{a, b, e\}, \\ \{c, d, f, g\}, \{a, b, c, d, e, f, g\}\}$$



# Skrčitev skupine

## Analiza omrežij

V. Batagelj

Statistika

Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi

*Skrčitev skupine*  $C$  imenujemo graf  $\mathcal{G}/C$ , ki ga dobimo tako da vsa vozlišča skupine  $C$  zamenjamo z enim vozliščem, recimo  $c$ .

Natančneje  $\mathcal{G}/C = (\mathcal{V}', \mathcal{L}')$ , kjer je  $\mathcal{V}' = (\mathcal{V} \setminus C) \cup \{c\}$  in  $\mathcal{L}'$  sestavljajo povezave iz  $\mathcal{L}$ , ki imajo obe krajišči v  $\mathcal{V} \setminus C$ . Večkratne vzporedne povezave združimo v eno. Poleg teh pa še 'zvezda' z vrhom  $c$  in krakom  $(v, c)$ , če  $\exists p \in \mathcal{L}, u \in C : p(v, u)$ , oziroma krakom  $(c, v)$ , če  $\exists p \in \mathcal{L}, u \in C : p(u, v)$ . V vozlišču  $c$  je zanka  $(c, c)$ , če  $\exists p \in \mathcal{L}, u, v \in C : p(u, v)$ .

V omrežju nad grafom  $\mathcal{G}$  moramo povedati še, kako so določene vrednosti/uteži v skrčenem delu. Običajno kar kot vsota ali maksimum/minimum izvornih vrednosti.

Operations/Network + Partition/Shrink Network





# Izračun uteži w

Analiza  
omrežij

V. Batagelj

Statistika

Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi

```
File/Pajek Project File/Read [SKtrade.paj]
Network/Create New Network/Transform/Remove/Loops [No]
Network/Create New Network/Transform/Edges -> Arcs [No]
Operations/Network+Partition/Shrink Network [1 0]
```

	1	2	3	4	5	6	7	Label
1.	2	30	13	56	42	45	4	#usa
2.	30	74	25	196	20	37	12	#cub
3.	12	28	33	124	16	36	5	#per
4.	55	217	130	694	427	483	41	#uki
5.	42	8	14	406	122	117	11	#mli
6.	43	37	43	444	142	307	30	#irn
7.	4	4	5	39	9	30	2	#aut

```
Partition/Make Permutation
[select partition (Sub)continents]
Operations/Partition+Permutation/
  Functional Composition Partition*Permutation
Partition/Count
```

```
count      2  15   7  29  33  30   2
```



# ... Izračun uteži $w$

Analiza omrežij

V. Batagelj

Statistika

Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi

```
Partition/Copy to Vector
Vector/Create Constant Vector [7 1.0]
[select as second vector Copy of partition ...]
Vectors/Divide (First/Second)
Network/Create Vector/Get Loops
Vectors/Add (First+Second)
Operations/Network+Vector/Transform/Put Loops/as Arcs
[select vector Divide V? by ...]
Operations/Network+Vector/Vector#Network/input
Operations/Network+Vector/Vector#Network/output
```

	1	2	3	4	5	6	7	
#usa	1.	0.50	1.00	0.93	0.97	0.64	0.75	1.00
#cub	2.	1.00	0.33	0.24	0.45	0.04	0.08	0.40
#per	3.	0.86	0.27	0.67	0.61	0.07	0.17	0.36
#uki	4.	0.95	0.50	0.64	0.83	0.45	0.56	0.71
#mli	5.	0.64	0.02	0.06	0.42	0.11	0.12	0.17
#irn	6.	0.72	0.08	0.20	0.51	0.14	0.34	0.50
#aut	7.	1.00	0.13	0.36	0.67	0.14	0.50	0.50

Ker so na diagonalni 0, bi jih bilo smiselno pred izračunom postaviti na 1 ali pa diagonalne vrednosti deliti z  $n(C_i)(n(C_i) - 1)$ , če je  $n(C_i) > 1$ .

Macro **weights**.





# Podgraf

Analiza omrežij

V. Batagelj

Statistika

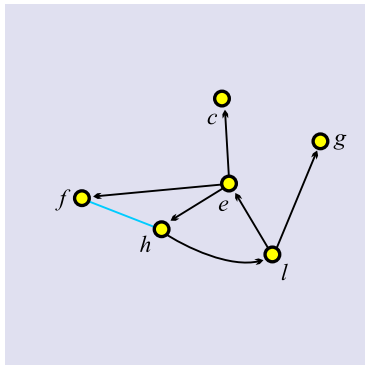
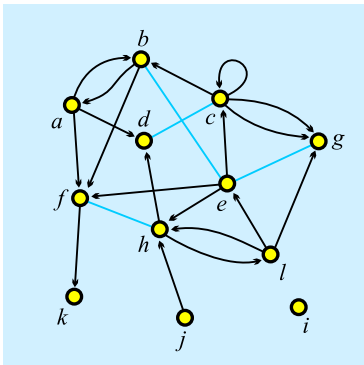
Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi



**Podgraf**  $\mathcal{H} = (\mathcal{V}', \mathcal{L}')$  danega grafa  $\mathcal{G} = (\mathcal{V}, \mathcal{L})$  je graf, katerega povezave  $\mathcal{L}'$  so vsebovane v povezavah grafa  $\mathcal{G}$ ,  $\mathcal{L}' \subseteq \mathcal{L}$ , vozlišča  $\mathcal{V}'$  pa v vozliščih grafa  $\mathcal{G}$ ,  $\mathcal{V}' \subseteq \mathcal{V}$ , in vsebujejo tudi vsa krajišča povezav  $\mathcal{L}'$ .

Podgraf je lahko **porojen** z dano podmnožico vozlišč ali povezav. Podgraf je **vpet**, če je  $\mathcal{V}' = \mathcal{V}$ .





# Izrez: Snyder in Kick

## Latinska Amerika : Južna Amerika

### Analiza omrežij

V. Batagelj

Statistika

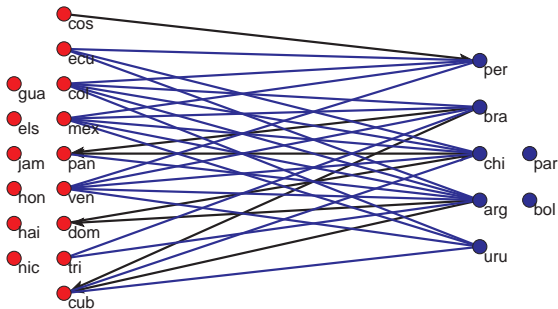
Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi



Operations/Network + Partition/Extract Subnetwork [3,4]  
Operations/Network + Partition/Transform/Remove lines/  
Inside clusters [3,4]

Vozlišča lahko ročno razmestimo po pravokotni mreži ustvarjeni z

[Draw] Move/Grid



# Prerezi

Analiza  
omrežij

V. Batagelj

Statistika

Homomorfizmi

Razbitja

Skraćitev

Izrezi

Prerezi

*Vozliščni prerez* omrežja  $\mathcal{N} = (\mathcal{V}, \mathcal{L}, p)$ ,  $p : \mathcal{V} \rightarrow \mathbb{R}$ , na *ravni*  $t$  je podomrežje  $\mathcal{N}(t) = (\mathcal{V}', \mathcal{L}(\mathcal{V}'), p)$ , določeno z množico vozlišč

$$\mathcal{V}' = \{v \in \mathcal{V} : p(v) \geq t\}$$

kjer je  $\mathcal{L}(\mathcal{V}')$  množica vseh povezav iz  $\mathcal{L}$ , ki imajo obe krajišči v  $\mathcal{V}'$ .

*Povezavni prerez* omrežja  $\mathcal{N} = (\mathcal{V}, \mathcal{L}, w)$ ,  $w : \mathcal{L} \rightarrow \mathbb{R}$ , na *ravni*  $t$  je določeno z množico povezav

$$\mathcal{L}' = \{e \in \mathcal{L} : w(e) \geq t\}$$

To je podomrežje  $\mathcal{N}(t) = (\mathcal{V}(\mathcal{L}'), \mathcal{L}', w)$ , kjer je  $\mathcal{V}(\mathcal{L}')$  množica vseh krajišč povezav iz  $\mathcal{L}'$ .

Vrednosti ravni  $t$  določimo na osnovi porazdelitve vrednosti funkcij  $w$  oziroma  $p$ . Običajno nas zanimajo komponente prereza, ki niso niti prevelike, niti premajhne.





# Povezavni prerez: Krebsova spletna podjetja,

$$w_3 \geq 5$$

## Analiza omrežij

V. Batagelj

Statistika

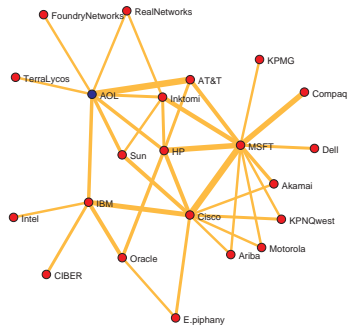
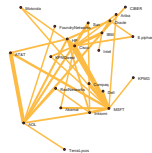
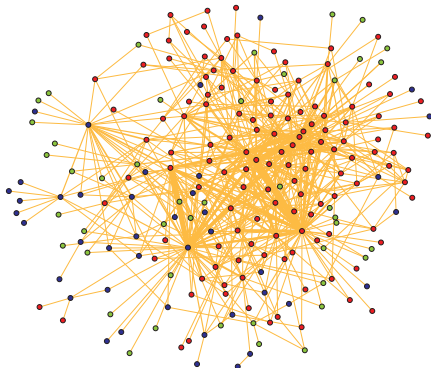
Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi





# Analiza omrežja s prerezi

## Analiza omrežij

V. Batagelj

Statistika

Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prezezi

Vrednosti ravni  $t$  določimo na osnovi porazdelitve vrednosti funkcij  $w$  oziroma  $p$ . Običajno nas zanimajo komponente prereza, ki niso niti prevelike, niti premajhne.

Vozliščni prerez:  $p$  shranjena v vektorju

```
Vector/Info [+10] [#10]
Vector/Make Partition/by Intervals/Selected Thresholds [t]
Operations/Network + Partition/Extract Subnetwork [2]
```

Povezavni prerez: omrežje z utežmi

```
Network/Info/Line values [#10]
Network/Create New Network/Transform/Remove/Lines with Value/
  lower than [t]
Network/Create Partition/Degree/All
Operations/Network + Partition/Extract Subnetwork [1-*]
```



# Analiza omrežja s prerezi

## Analiza omrežij

V. Batagelj

Statistika

Homomorfizmi

Razbitja

Skrčitev

Izrezi

Prerezi

Prerezi ponujajo enostaven pristop k analizi omrežij. Za izbrano lastnost/utež in raven  $t$  določimo pripadajoči prerez  $\mathcal{N}(t)$ . Pozornost posvetimo njegovim komponentam.

Število in velikost komponent je odvisna od ravni  $t$ . Pogosto se pojavi več majhnih komponent. Pri analizi nas običajno zanimajo le komponente 'prave' velikosti – vsaj  $k$  in ne večje kot  $K$ . Premajhne komponente zavržemo kot 'nezanimive'; prevelike komponente pa ponovno prererežemo na neki višji ravni.

Vrednost  $t$ ,  $k$  in  $K$  določimo s pregledom porazdelitve vrednosti lastnosti/uteži in z upoštevanjem dodatnega vedenja o značilnostih omrežja in ciljnih raziskave.

V program Pajek je vgrajenih nekaj novih, učinkovito izračunljivih lastnosti/uteži.