



# Analiza omrežij

11. Pa še to

Vladimir Batagelj

Univerza v Ljubljani, FMF, matematika

Interdisciplinarni doktorski študijski program Statistika  
Ljubljana, maj 2014



# Kazalo

## Analiza omrežij

V. Batagelj

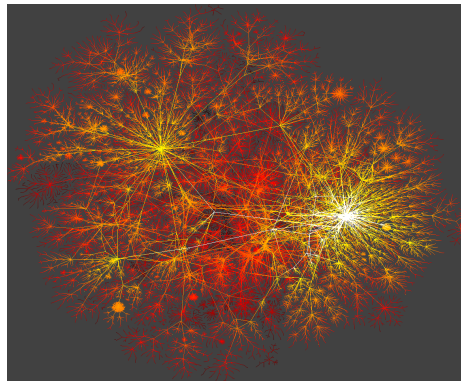
Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?

- 1 Analiza časovnih omrežij
- 2 Brezlestvična omrežja
- 3 Skeleti
- 4 Kaj še?



K. C. Claffy: Skitter data

wiki: <http://pajek.imfm.si/doku.php?id=event:pd>  
April 24, 2014/ junij 2013



# Analiza časovnih omrežij

## Analiza omrežij

V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?

Pajek podpira opise časovnih omrežij od leta 1999.

Omrežje lahko analiziramo kot celoto (vse časovne točke skupaj) ali pa po posameznih *časovnih rezinah*. V programu Pajek za posamezne časovne rezine določimo vrednosti izbranih značilnosti rezin. Te nato izvozimo v nek statistični program (R ali SPSS), kjer jih podrobneje analiziramo.

V programu Pajek še ni posebnih orodij za analizo časovnih omrežij – potrebno jih je še razviti.

Zanimiv pristop k analizi in prikazu razvoja časovnih omrežij je razvil U. Brandes s sodelavci ([članek](#), [prikazi](#)).

NAS: [Dynamic Social Network Modeling and Analysis](#).



Običajni pristop bi bil:

- prikaz celotnega omrežja z vzmetnim postopkom;
- prikaz zaporedja rezin;
- prikaz izbrane relacije.

Omrežja KEDS vsebuje veliko relacij in dobimo 'mavrične' slike, iz katerih je težko kaj razbrati. Za to, da bi dobili preglednejšo sliko smo se odločili združiti relacije v tri skupine:

Pozitivne  
(modra)

01 Yield  
02 Comment  
03 Consult  
04 Approve  
05 Promise  
06 Grant  
07 Reward

Nevtralne  
(zelena)

08 Agree  
09 Request  
10 Propose  
11 Reject  
12 Accuse  
13 Protest  
14 Deny

Negativne  
(rdeča)

15 Demand  
16 Warn  
17 Threaten  
18 Demonstrate  
19 Reduce Relationship  
20 Expel  
21 Seize  
22 Force



# KEDS – statistika

Analiza  
omrežij

V. Batagelj

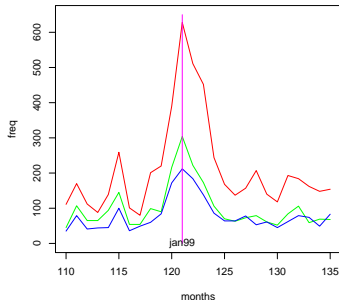
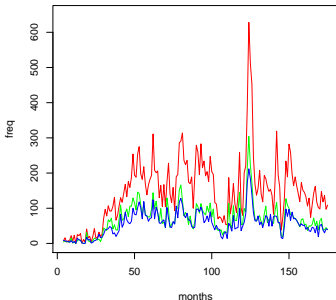
Analiza  
časovnih  
omrežij

Brezlestvična  
omrežja

Skeleti

Kaj še?

Pogostost povezav skozi čas. Uporaba operacij na zaporedjih omrežij !!!



```
months <- 4:175
plot(months,v3,type="l",ylim=c(0,650),ylab="freq",xlab="months",col="red")
lines(months,v2,col="green"); lines(months,v1,col="blue")

m <- 110:135
plot(m,v3[m],type="l",ylim=c(0,650),ylab="freq",xlab="months",col="red")
lines(m,v2[m],col="green"); lines(m,v1[m],col="blue")
t <- 121; lines(c(t,t),c(0,650),col="magenta"); text(t,0,"jan99")
```



# Časovna analiza omrežja US Patents

## Nataša Kejžar

Analiza  
omrežij

V. Batagelj

Analiza  
časovnih  
omrežij

Brezlestvična  
omrežja

Skeleti

Kaj še?

Hall, B.H., Jaffe, A.B. and Tratjenberg M.: The NBER U.S. Patent Citations Data File. NBER Working Paper 8498 (2001).

<http://www.nber.org/patents/>

- omrežje zbrano v obdobju 1975 – 1999  
vsebuje patente podeljene v obdobju januar 1963 – december 1999
- 2923922 patentov z znakovnim opisom, 850846 slikovni opis  
3774768 vozlišč
- 16522438 sklicevanj (usmerjenih povezav)

Poleg omrežja je zbranih še več vozliščnih lastnosti: leto prijave, oznaka prijavitelja, tehnološka (pod)skupina, ...

Vsa vozlišča iz iste skupine in iste časovne rezine so skrčene v eno vozlišče. Dobljena skrčena omrežja so izhodišče za nadaljnje analize.



# Izbira okna

Analiza  
omrežij

V. Batagelj

Analiza  
časovnih  
omrežij

Brezlestvična  
omrežja

Skeleti

Kaj še?

Uporabljen je bil izsledek o *sklicnem zamiku* (Hall, Jaffe, Trajtenberg): razlika med letom podelitve patenta in leti podelitev patentov, na katere se sklicuje, je najpogosteje 3 ali 4 leta. Za starejše patente število zelo hitro upada. Zato je bil za dolžino rezine izbran čas 4 let.



# Patenti (1984 – 1987)

Analiza omrežij

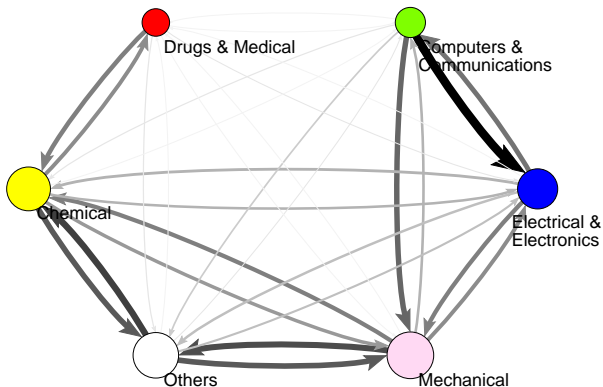
V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?







# Patenti (1987 – 1990)

Analiza omrežij

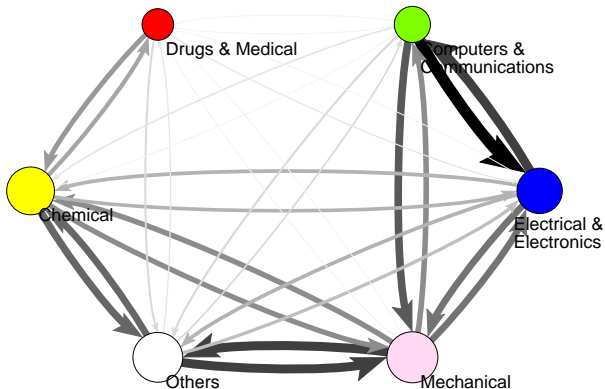
V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?





# Patenti (1990 – 1993)

Analiza omrežij

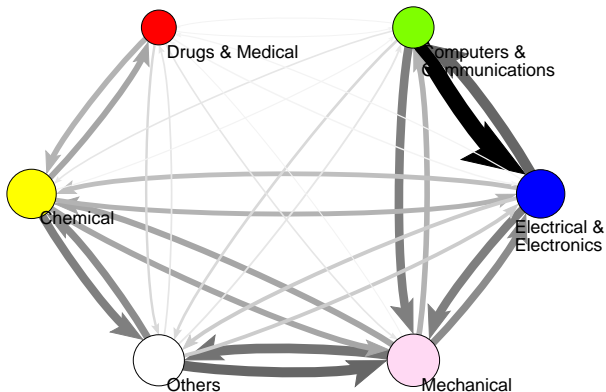
V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?





# Patenti (1993 – 1996)

Analiza omrežij

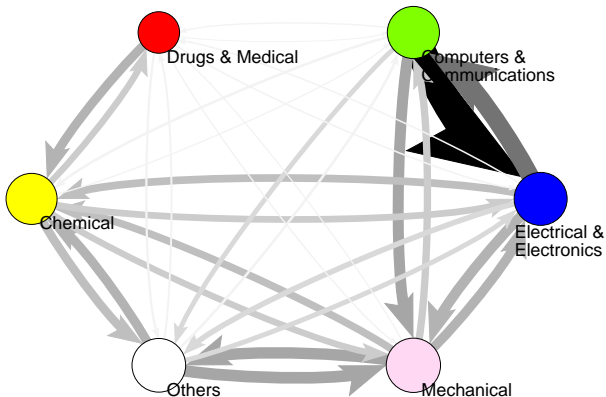
V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?





# Rast števila patentov in števila sklicevanj v skupinah

Analiza omrežij

V. Batagelj

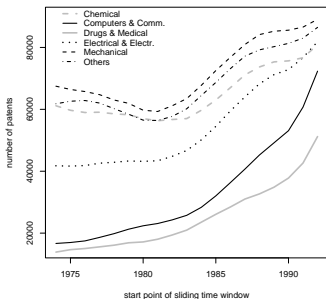
Analiza časovnih omrežij

Brezlestvična omrežja

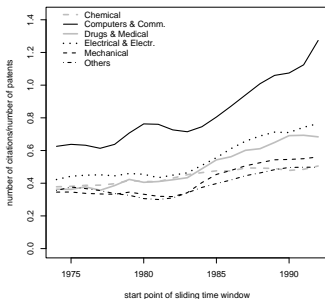
Skeleti

Kaj še?

Number of patents in 4 years time slices



Relative growth of citations in 4 years time slices





# Pomembnost skupin

Analiza omrežij

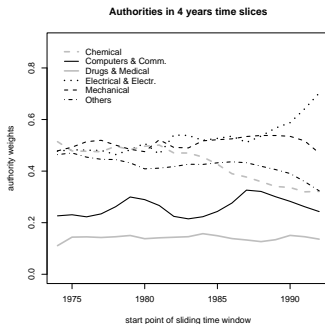
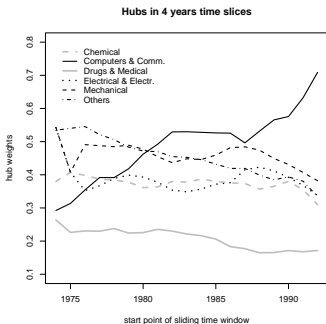
V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?



Za določitev pomembnosti skupin je bil uporabljen postopek *kazala in viri*. Pomembna kazala (Computers & Communication and Mechanical, Others) **uporabljajo znanje** iz drugih skupin. Pomembni viri (Mechanical and Electrical & Electronic) predstavljajo skupine, ki nudijo **osnovna znanja**.



# PajektoSVGanim

Analiza  
omrežij

V. Batagelj

Analiza  
časovnih  
omrežij

Brezlestvična  
omrežja

Skeleti

Kaj še?

Za gladko prikazovanje (film) razvoja časovnega omrežja je bil razvit poseben program **PajektoSVGanim** (sprogramiral ga je Darko Brvar), ki predela zaporedje Pajekovih prikazov v 'risanko' v SVGju.

Obstajajo še drugi podobni programi:

Skye Bender-deMoll, Daniel A. McFarland, James Moody: SoNIA (**prikazi, program, članek**).

Peter A. Gloor: TeCFlow (**primeri, program, članek**).

**Franzosi**

Zanimiv pristop k analizi časovnih omrežij bi bila razširitev postopka iskanja vzorcev na časovna omrežja – odkrivanje značilnih zgodb.



# Omrežja ER

## Analiza omrežij

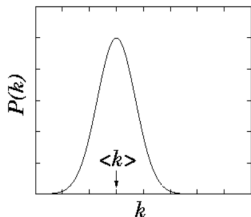
V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?



Barabasi: The Architecture of Complexity

Paul Erdős in Alfréd Rényi, 1959 sta vpeljala slučajne grafe/omrežja, kjer sta vozlišči grafa povezani z dano verjetnostjo  $p$ . Pravimo jim **omrežja ER** in so bila teoretično temeljito obdelana (npr. **B. Bollobás**). Nekaj spoznanj:

- porazdelitev števila vozlišč z dano stopnjo je Poissonova in večina vozlišč ima stopnjo (zelo) blizu povprečne  $\bar{d}$

$$P(d) = \binom{n}{d} p^d (1-p)^{n-d} \approx \frac{1}{d!} \bar{d}^d e^{-\bar{d}}$$

- za  $p \geq \frac{1}{n}$  se v grafu pojavijo cikli; kmalu zatem **velika komponenta**;
- za  $p \geq \frac{\log_2 n}{n}$  so skoraj vsi grafi povezani;
- skoraj vsi grafi imajo premer 2.



# 6 korakov

## Analiza omrežij

V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?

Leta 1967 je psiholog **Stanley Milgram** naredil poskus s pismi. Izbrani posamezniki naj bi s pismi v čim manj korakih dosegli izbrano osebo, pri čemer naj bi vsak posameznik poslal pismo (s temi navodili) svojim znancem, za katere domneva, da so po poznanstvih lahko bližje izbrani osebi, . . . Izkazalo se je, da je bilo za doseg izbrane osebe potrebnih 6 korakov (*six degrees of separation*). Podobne lastnosti so opazili tudi v drugih omrežjih: 19 klikov na spletu, 3 reakcije med molekulami v celici, . . . , v Sloveniji smo skoraj vsi 'v žlahti'.

Omrežjem, v katerih je povprečna pot med vozlišči majhna, pravimo **mali svetovi** (*small worlds*).

Podobno je **Mark Granovetter**, 1973 opazil, da se v družbenih omrežjih pojavljajo gruče vozlišč povezanih s **krepkimi** povezavami (*strong ties*), ki se med seboj povezujejo s **šibkimi** povezavami (*weak ties*) v večja omrežja.





# Prevezovanje in mali svetovi

## Analiza omrežij

V. Batagelj

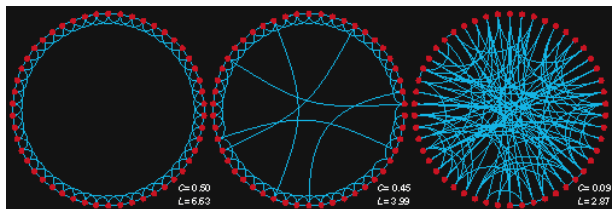
Analiza časovnih omrežij

Brezlestvna omrežja

Skeleti

Kaj še?

**Duncan Watts** in **Steven Strogatz** sta razvila postopek pridobivanja malih svetov iz pravilnih omrežij s *prevezovanjem* – povezavo prestavimo na neko drugo vozlišče. Ta postopek povzroči, da se dolžine najkrajših poti med vozlišči krajšajo, ker prevezovanje ustvarja bližnjice.



Densmore: Power-Law Networks



# Brezlestvična omrežja

## Analiza omrežij

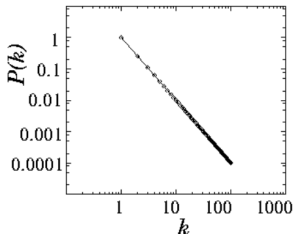
V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?



Barabasi: The Architecture of Complexity

**Albert-László Barabási** s sodelavci z Univerze Notre Dame se je leta 1998 lotil proučevanja značilnosti spleta kot omrežja. Pričakovali so, da se bo obnašal kot slučajni graf, a so naleteli na presenečenje:

- na upoštevanem delu spleta je manjše število vozlišč (manj kot 0.01%) z zelo **veliko stopnjo** ( $k$ , ki je v omrežjih ER skoraj neverjetna). Ta vozlišča držijo splet skupaj;
- porazdelitev stopenj vozlišč je **potenčna** (premica v dvojno-logaritemskem prikazu)

$$P(d) = Cd^{-\tau}$$



# ... brezlestvična omrežja

## Analiza omrežij

V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?

Ker v teh omrežjih povprečna stopnja ni 'značilna' so jih poimenovali *brezlestvična* (*scale free*).

Nadaljnje raziskave so pokazale, da se tovrstna omrežja pojavljajo na različnih področjih: osebe – elektronska pošta, spolni odnosi (mamila, AIDS), telefonski klici, soavtorstvo; filmski igralci – igranje v istem filmu; proteini – medsebojno delovanje; besede – pomenske zveze; ...

Prva razlaga pojave brezlestvičnih omrežij je bila, da so to omrežja, ki:

- (z)*rastejo*
- pri rasti se nova vozlišča povezujejo slučajno s starimi glede na *priljubljenost* slednjih (*preferential attachment*). Običajno je priljubljenost vozlišča sorazmerna njeni stopnji.

Ti lastnosti omogočata razložiti pojavitev vozlišč z zelo veliko stopnjo in potenčne porazdelitve.

Kasneje so opazili tudi, da se v tovrstnih omrežjih pojavljajo *gruče* (moji prijatelji se najbrž poznajo) – nakopičenost je večja kot v omrežjih RE.



# ... brezlestvična omrežja – določanje eksponenta

## Analiza omrežij

V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?

Problemi: razpršenost na koncu, premica le na podintervalu, neenakomerna gostota, ...

Graf porazdelitvene funkcije je tudi potenčna funkcija

$$\int Cx^{-\tau} = C \frac{x^{1-\tau}}{1-\tau}$$

Ocena po metodi največjega verjetja

$$\tau = 1 + n \left( \sum_{i=1}^n \ln \frac{x_i}{x_{min}} \right)^{-1}$$

power, Pareto



# ... brezlestvična omrežja

Based on this model it can be shown that:

- the degree distribution is the power law;
- the average length of geodesics is  $O(\log n)$  ;
- these networks are resilient against random vertex or edge removals (random attacks), but quickly become disconnected when large degree nodes (Achilles' heel) are removed (targeted attacks).

In real-life networks vertices often also form groups – clustering.

Several improvements and alternative models were proposed that also produce scale-free networks with some additional properties characteristic for real-life networks: copying (Kleinberg 1999), combining random and preferential attachment (Pennock et al. 2002), R-mat (Chakrabarti et al. 2004), forest fire (Leskovec et al. 2005), aging, fitness, nonlinear preferences, ...



# O brezlestvičnih omrežjih

Analiza  
omrežij

V. Batagelj

Analiza  
časovnih  
omrežij

Brezlestvična  
omrežja

Skeleti

Kaj še?

Matthew Effect: [Wikipedia](#), [When Do Matthew Effects Occur?](#)  
Epidemics: [Barthélemy](#), [Barrat](#), [Pastor-Sattoras](#), [Vespignani](#), [Complex Networks Collaboratory](#).  
Searching: [Adamic et al.](#)  
General: [Center for Complex Network Research](#), [Albert](#), [Barabási](#), [Newman](#), [Borner](#), [Sanyal](#), [Vespignani](#).  
Determining the power law: [M. E. J. Newman](#): [Power laws, Pareto distributions and Zipf's law](#) and [Power-law distributions in empirical data](#).  
Packages in R: [igraph](#), [plfit](#) / [Santa Fe](#).



# Razlike med brezlestvičnimi in omrežji ER

## Analiza omrežij

V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?

Brezlestvična omrežja so veliko odpornejša kot omrežja ER na *slučajno izločanje* vozlišč; zelo občutljiva pa so na *ciljane napade* na vozlišča z veliko stopnjo (*Ahilova peta*).

To spoznanje je pomembno za zaščito omrežij (virusi, vdori, ...). V omrežjih ER se epidemija razširi, ko število okuženih vozlišč doseže nek prag, sicer okužba izgine. V brezlestvičnih omrežjih pa se virusi vselej razširijo in ostanejo v omrežju (*Pastor-Sattoras, Vespignani*). To spoznanje postavlja 'na glavo' dosedanje poglede in ukrepe pri preprečevanju epidemij. Za preprečevanje okužb slučajna cepljenja, čeprav obsežna, niso učinkovita; učinkoviteje je cepiti osebe, ki imajo zelo veliko stikov. Podobno velja pri trženju – farmacevtska podjetja se usmerjajo na pomembne zdravnike.

Pregledi: 1, 2; 3/ L3; iskanja.

4, 5



# Slučajna omrežja izbranih vrst

Analiza  
omrežij

V. Batagelj

Analiza  
časovnih  
omrežij

Brezlestvična  
omrežja

Skeleti

Kaj še?

Batagelj V., Brandes U.: *Efficient Generation of Large Random Networks*.  
Physical Review E 71, 036113, 2005 **stran**  
**pregled**





# Dense networks

Analiza  
omrežij

V. Batagelj

Analiza  
časovnih  
omrežij

Brezlestvična  
omrežja

Skeleti

Kaj še?

The important parts of networks are smaller, but relatively dense. For such networks the standard "nodes and links" visualization is not readable. Much better visualization can be produced using the matrix representation for an appropriate ordering (determined for example by clustering or blockmodeling).

Another approach is to display only the *skeleton* of the network obtained by removing less important links. The standard skeleton is a minimal spanning tree; often also Pathfinder skeletons are used.



# Pathfinder / Dissimilarities

Analiza  
omrežij

V. Batagelj

Analiza  
časovnih  
omrežij

Brezlestvična  
omrežja

Skeleti

Kaj še?

Joly and Le Calvé theorem:

For any even dissimilarity measure  $d$  there is a unique number  $p \geq 0$ , called its *metric index*, such that:  $d^r$  is metric for all  $r \leq p$ , and  $d^r$  is not metric for all  $r > p$ .

In the opposite direction we can say: Let  $d$  be a dissimilarity and for  $x, y$  and  $z$  we have  $d(x, z) + d(z, y) \geq d(x, y)$  and  $d(x, y) > \max(d(x, z), d(z, y))$  then there exists a unique number  $p \geq 0$  such that for all  $r > p$

$$d^r(x, z) + d^r(z, y) < d^r(x, y)$$

or equivalently

$$d(x, z) \boxplus d(z, y) < d(x, y)$$

where  $a \boxplus b = \sqrt[r]{a^r + b^r}$ .



# Pathfinder

Analiza  
omrežij

V. Batagelj

Analiza  
časovnih  
omrežij

Brezlestvična  
omrežja

Skeleti

Kaj še?

The Pathfinder algorithm was proposed in eighties (Schvaneveldt 1981, Schvaneveldt et al. 1989; Schvaneveldt, 1990) for simplification of weighted networks – it removes from the network all lines that do not satisfy the triangle inequality – if for a line a shorter path exists connecting its endpoints then the line is removed. The basic idea of the Pathfinder algorithm is simple. It produces a network  $PFnet(\mathbf{W}, r, q) = (\mathcal{V}, \mathcal{L}_{PF})$

```
compute  $\mathbf{W}^{(q)}$ ;  
 $\mathcal{L}_{PF} := \emptyset$ ;  
for  $e(u, v) \in \mathcal{L}$  do begin  
  if  $\mathbf{W}^{(q)}[u, v] = \mathbf{W}[u, v]$  then  $\mathcal{L}_{PF} := \mathcal{L}_{PF} \cup \{e\}$   
end;
```

where  $\mathbf{W}$  is a network dissimilarity matrix and  $\mathbf{W}^{(q)}$  the matrix of values of all walks of length at most  $q$  computed over the semiring  $(\mathbb{R}_0^+, \oplus, \boxtimes, \infty, 0)$  where  $a \oplus b = \min(a, b)$ .



# Pathfinder

## Analiza omrežij

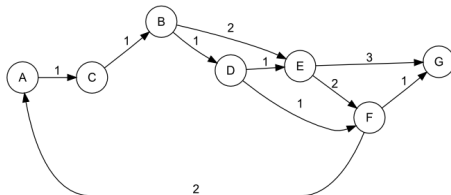
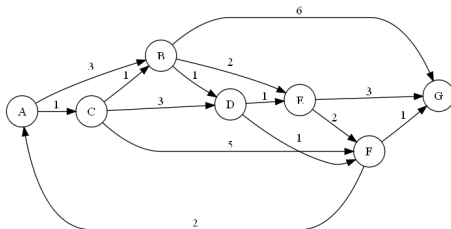
V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?





# Pathfinder

Analiza omrežij

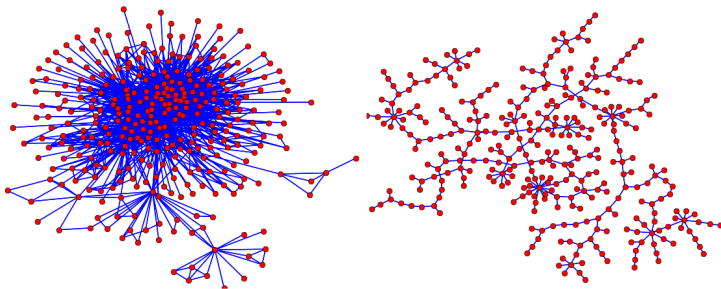
V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?





# PathFinderR

Analiza  
omrežij

V. Batagelj

Analiza  
časovnih  
omrežij

Brezlestvična  
omrežja

Skeleti

Kaj še?

We developed a fast Pathfinder algorithm for large sparse networks. For smaller (up to 1000) networks we wrote a program in R (based on the Fletcher's algorithm) to do the job.

```
# PathFinder
# http://pajek.imfm.si/lib/exe/fetch.php?media=slides:pfxxx.pdf
# by Vladimir Batagelj, December 24-28, 2011
#
# PathFinder(D,r,q) - determines the skeleton of network represented by
# matrix D . The weights in D should be dissimilarities; the value 0
# denotes nonlinked nodes.
# r - is the parameter in Minkowski operation
# q - is the limit on the length of considered paths; if q >= n-1
#   all paths are considered.
#
# PathFinderSim(S,r,q,s) - is a version of PathFinder for the case
# when the weights are similarities.
# s - determines how the similarity is transformed into dissimilarity
#   s = 1 - D = 1+max S - S
#   s = 2 - D = 1/S
# In the resulting skeleton the weights are the original similarities.

Multiply <- function(A,B,r){
  n <- nrow(A); C <- matrix(Inf,nrow=n,ncol=n)
  if(is.infinite(r)){
    for(i in 1:n) for(j in 1:n) C[i,j] <- min(pmax(A[i,],B[,j]))
  } else if (r==1){
    for(i in 1:n) for(j in 1:n) C[i,j] <- min(A[i,]+B[,j])
  } else {
    for(i in 1:n) for(j in 1:n) C[i,j] <- min((A[i,]^r+B[,j]^r)^(1/r))
  }
  C
}
```



# PathFinderR

## Analiza omrežij

### V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?

```
Closure <- function(W,r){
  n <- nrow(W); W[W==0] <- Inf; diag(W) <- 0
  if(is.infinite(r)){for(k in 1:n) for(i in 1:n) W[i,] <- pmin(W[i,],pmax(W[i,k],W[k,]))
  } else if (r==1){for(k in 1:n) for(i in 1:n) W[i,] <- pmin(W[i,],(W[i,k]+W[k,]))
  } else {for(k in 1:n) for(i in 1:n) W[i,] <- pmin(W[i,],(W[i,k]^r+W[k,]^r)^(1/r)) }
  W
}

Power <- function(W,r,q){
  n <- nrow(W); W[W==0] <- Inf; diag(W) <- 0
  T <- matrix(Inf,nrow=n,ncol=n); diag(T) <- 0
  if (q > 0) {
    i <- q; S <- W
    repeat{
      if ((i %% 2) == 1) { T <- Multiply(T,S,r) }
      i <- i %% 2; if (i == 0) break
      S <- Multiply(S,S,r)
    }
  }
  rownames(T) <- colnames(T) <- rownames(W)
  T
}
```



# PathFinderR

Analiza  
omrežij

V. Batagelj

Analiza  
časovnih  
omrežij

Brezlestvična  
omrežja

Skeleti

Kaj še?

```
PathFinder <- function(D,r=Inf,q=Inf,eps=0.0000001){
  if(r<1) stop("Error: r < 1")
  if(q>=nrow(D)-1) {D[(D>0)&(abs(D-Closure(D,r))>eps)] <- 0
  } else {D[(D>0)&(abs(D-Power(D,r,q))>eps)] <- 0}
  D
}

PathFinderSim <- function(S,r=Inf,q=Inf,s=1,eps=0.0000001){
  if(r<1) stop("Error: r < 1")
  n <- nrow(S); D <- S
  if(s==1) {D[S>0] <- 1+max(S)-S[S>0]} else {D[S>0] <- 1/S[S>0]};
  if(q>=n-1) {S[(S>0)&(abs(D-Closure(D,r))>eps)] <- 0
  } else {S[(S>0)&(abs(D-Power(D,r,q))>eps)] <- 0}
  S
}

# setwd("C:/Users/Batagelj/work/R/pf")
# PF <- PathFinder(n1,1,Inf)
# savenetwork(PF,'PFtest.net')

# cat(date(),"\n"); PF2 <- PathFinderSim(n2,1,Inf,2); cat(date(),"\n");
# savenetwork(PF2,'PF2500.net'); cat(date(),"\n")
```

PathfinderR / code

Cosine dissimilarity  $d_2 = 1 - \frac{x \cdot y}{|x| \cdot |y|}$  to be included.





# Kaj še?

## Analiza omrežij

V. Batagelj

Analiza časovnih omrežij

Brezlestvična omrežja

Skeleti

Kaj še?

Več tem iz analize omrežij se nismo dotaknili:

- posredništvo (brokerage) (glej ESNA, ch. 7)
- širjenje (difussion, epidemics) (glej ESNA, ch. 8)
- verjetnostni modeli in analize ( $p^*$ , SIENA)
- homofilija
- strukturne luknje (Burt)
- prostorska omrežja
- how to?
- naslovi URL v vozliščih, nova vrsta, Unicode
- delo z zelo velikimi omrežji
- ...