

K. C. Claffy: Skitter data

Univerza v Ljubljani
podiplomski študij statistike

Analiza omrežij
9. Pa še to

Vladimir Batagelj

Univerza v Ljubljani

Ljubljana, 10. december 2006 / 12. januar 2004

Kazalo

1	Analiza časovnih omrežij	1
4	Časovna analiza omrežja US Patents	4
12	PajektoSVGanim	12
13	Omrežja ER	13
15	Prevezovanje in mali svetovi	15
16	Brezlestvična omrežja	16
20	Razlike med brezlestvičnimi in omrežji ER	20
21	Slučajna omrežja izbranih vrst	21
22	Prikazi omrežij	22
23	Kaj še?	23

Analiza časovnih omrežij

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 anlizi in prikazu razvoja časovnih omrežij je razvil U. Brandes s sodelavci (*članek, prikazi*).

NAS: *Dynamic Social Network Modeling and Analysis*.

KEDS

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)

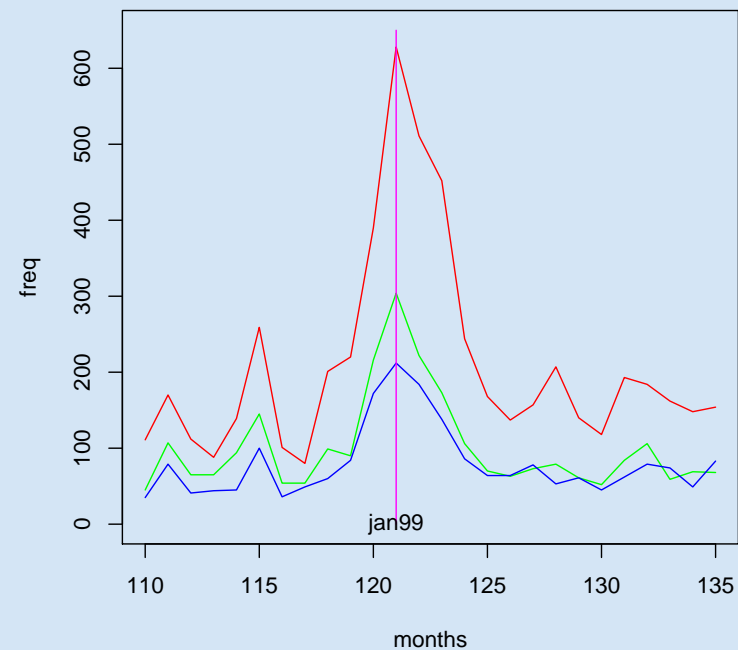
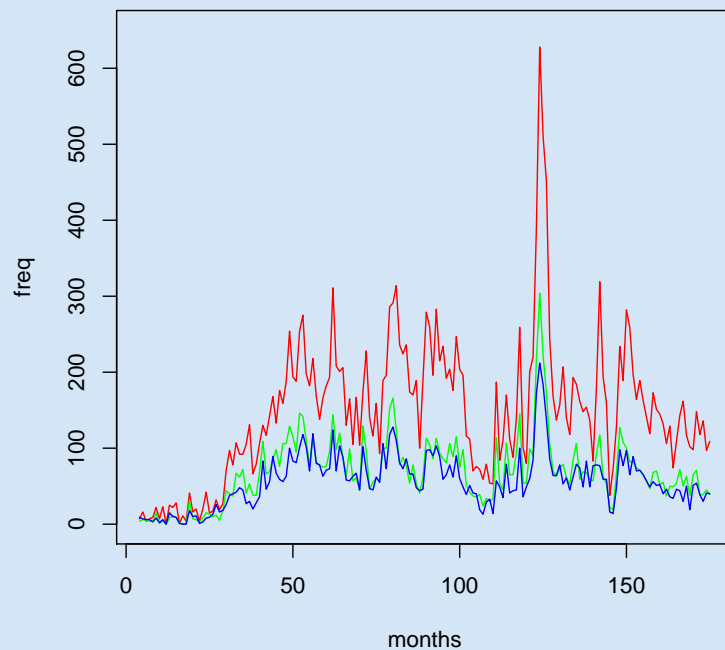
08 Agree
09 Request
10 Propose
11 Reject
12 Accuse
13 Protest
14 Deny

Negativne
(rdeča)

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

KEDS – statistika

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

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 točk
- 16522438 sklicevanj (usmerjenih povezav)

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

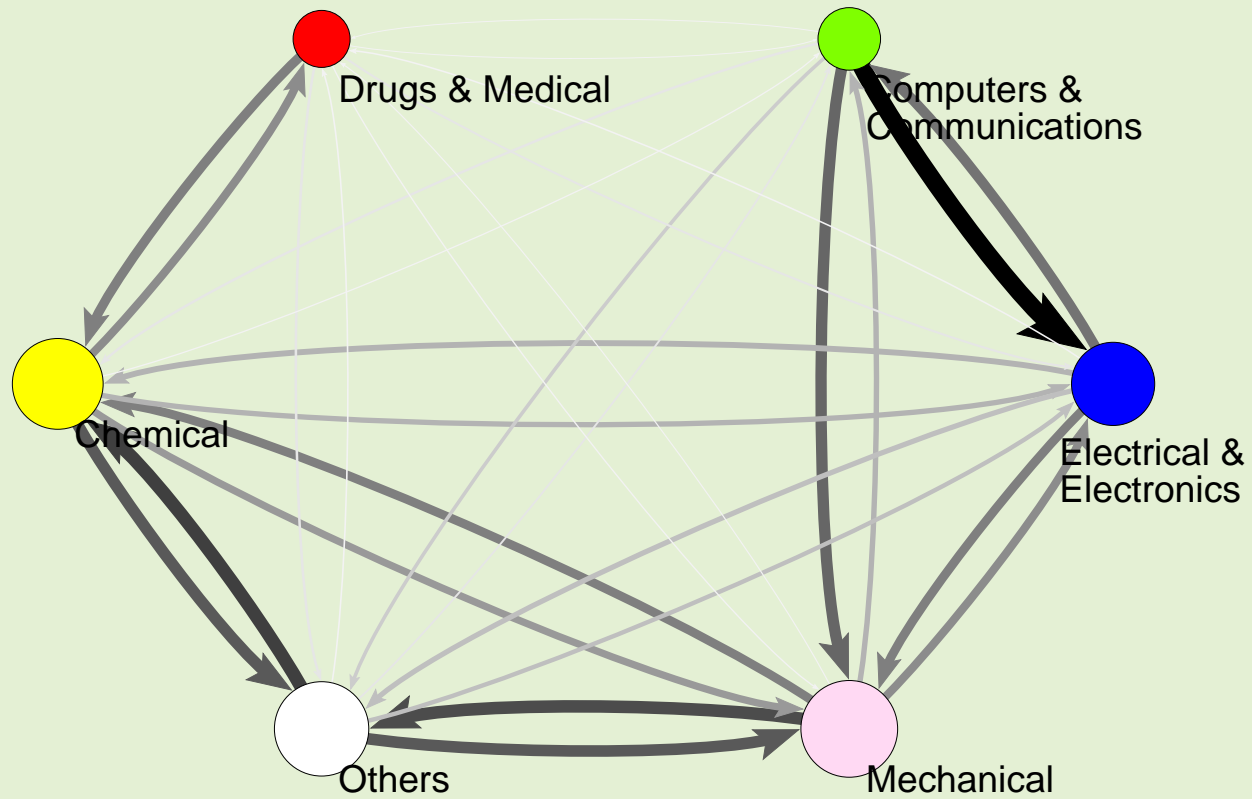
Vse točke iz iste skupine in iste časovne rezine so skrčene v eno točko.

Dobljena skrčena omrežja so izhodišče za nadaljnje analize.

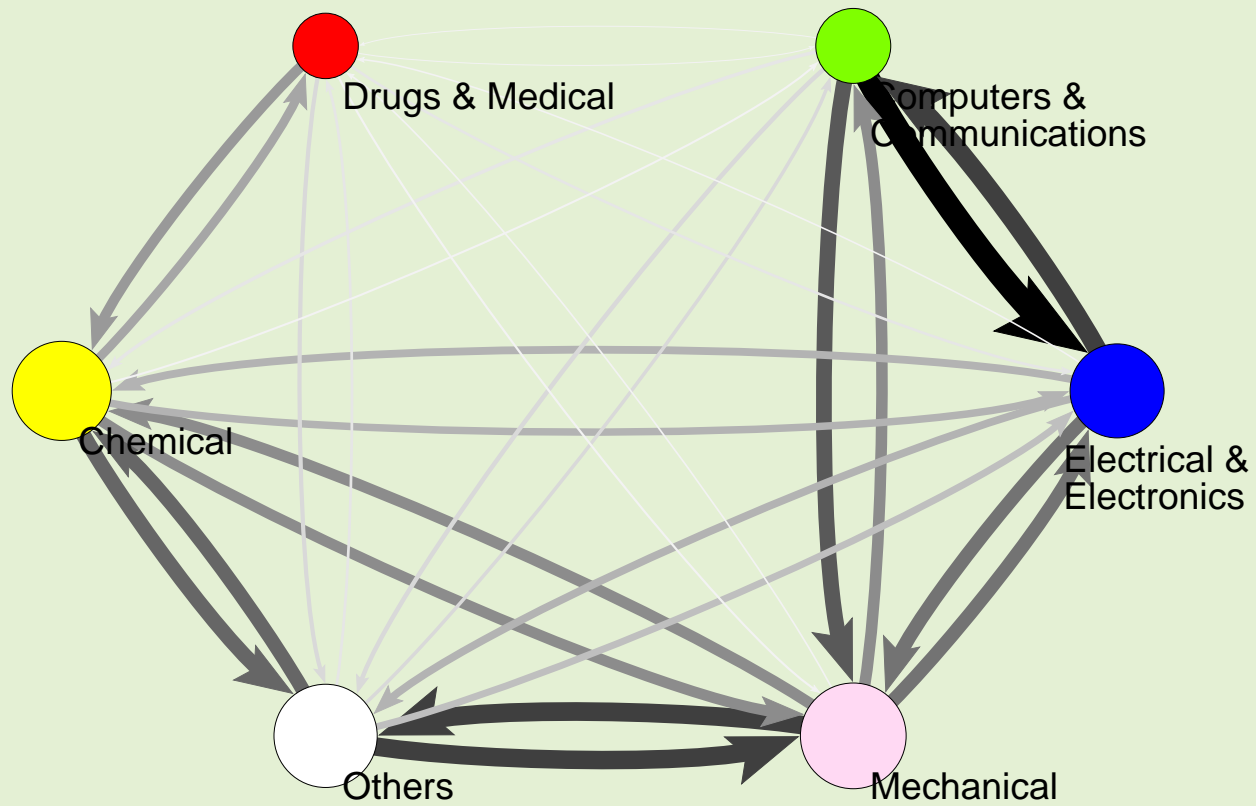
Izbira okna

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.

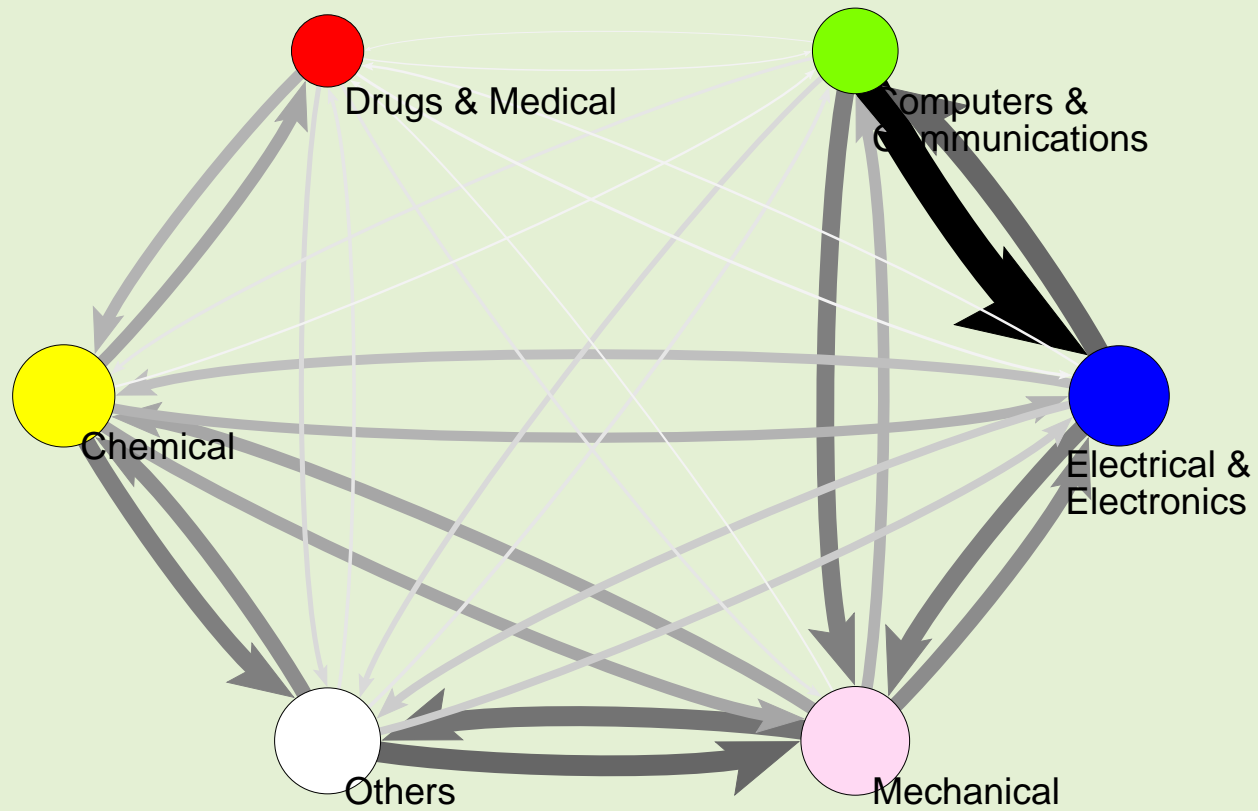
Patenti (1984 – 1987)



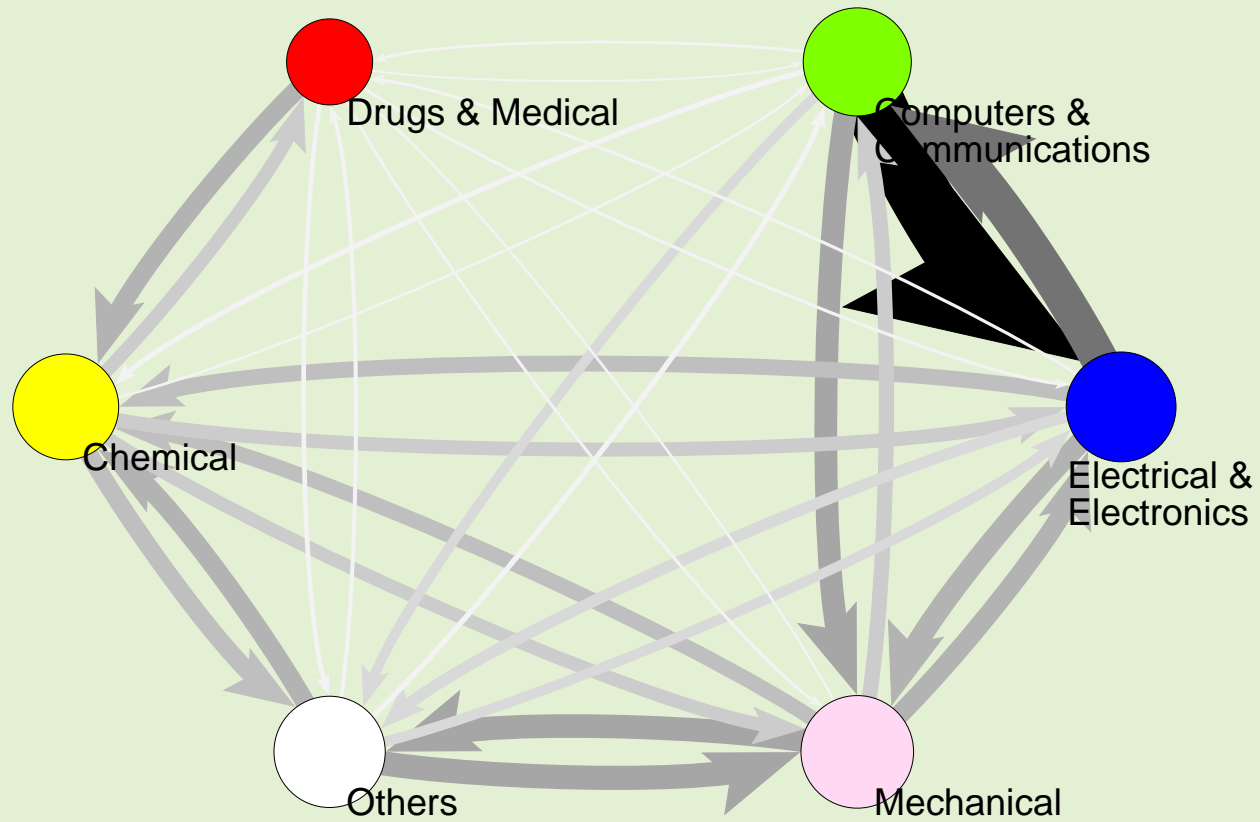
Patenti (1987 – 1990)



Patenti (1990 – 1993)

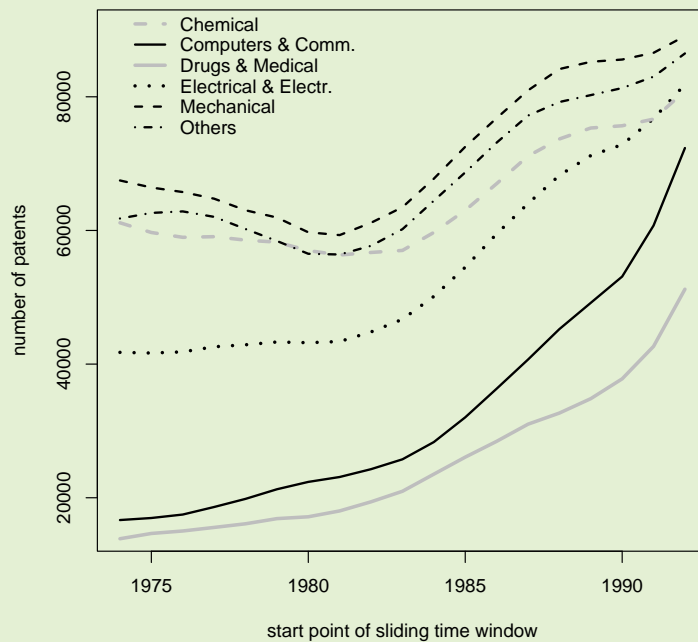


Patenti (1993 – 1996)

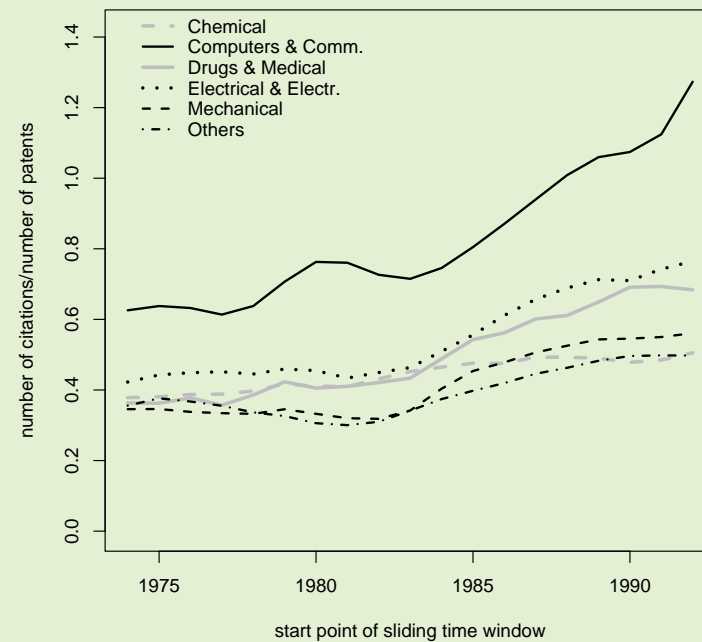


Rast števila patentov in števila sklicevanj v skupinah

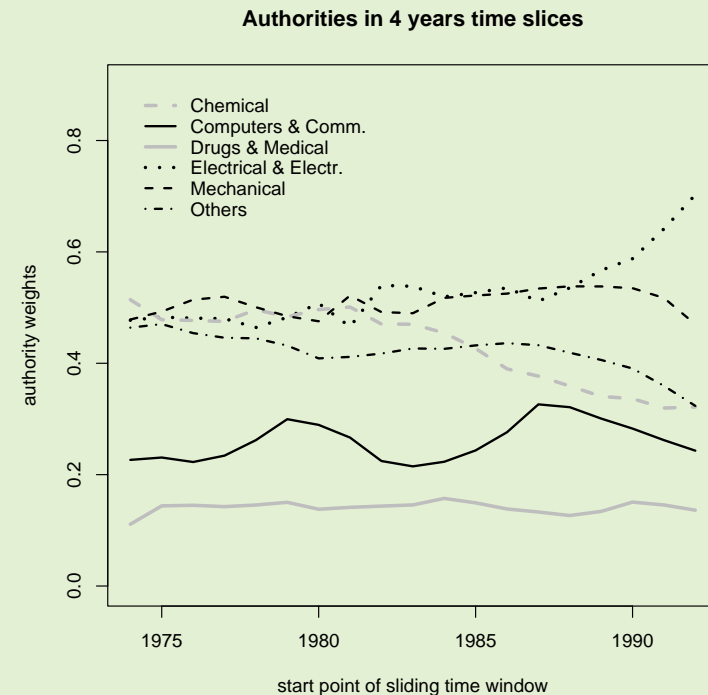
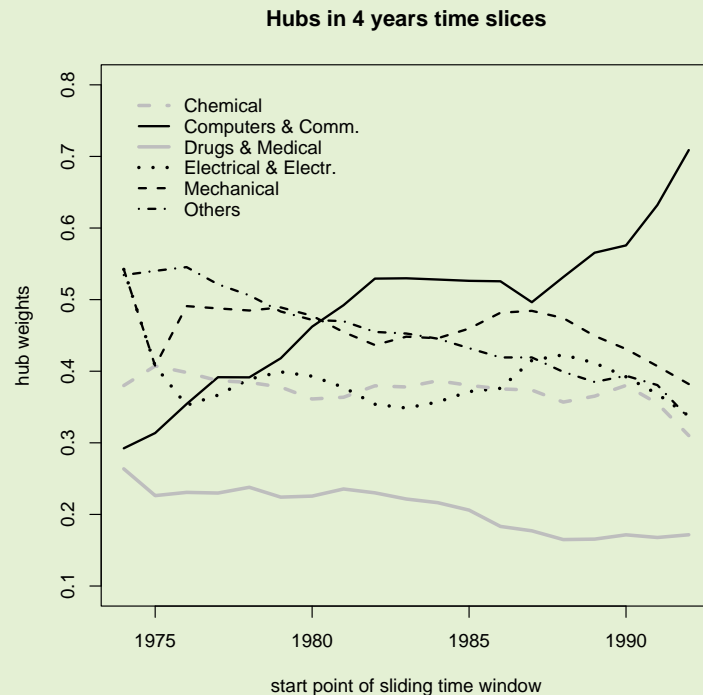
Number of patents in 4 years time slices



Relative growth of citations in 4 years time slices



Pomembnost skupin



Za določitev pomembnosti skupin je bil uporabljen postopek *kazala in vsebine*.

Pomembna kazala (Computers & Communication and Mechanical, Others) **uporabljajo znanje** iz drugih skupin.

Pomembne vsebine (Mechanical and Electrical & Electronic) predstavljajo skupine, ki nudijo **osnovna znanja**.

PajektoSVGanim

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

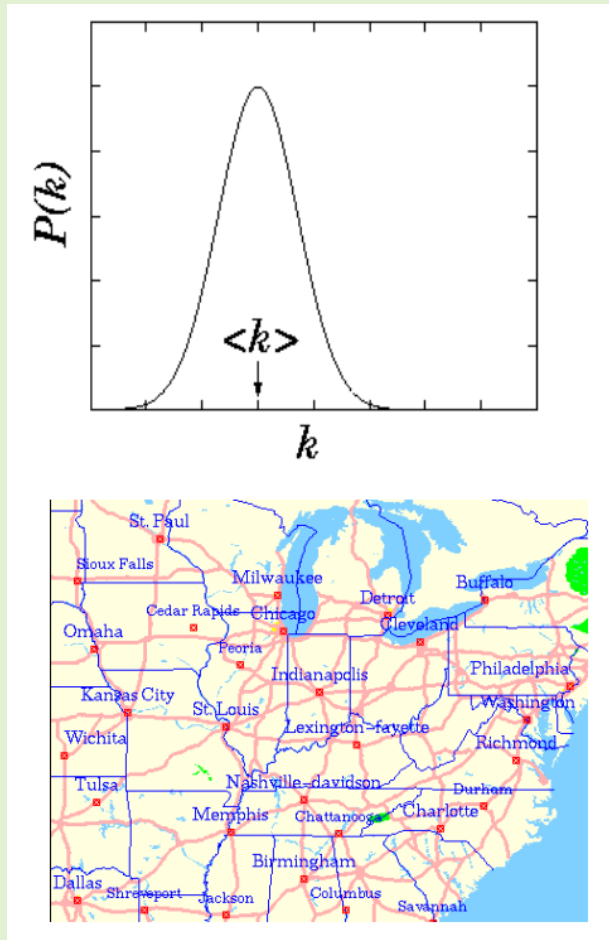
Obstajajo še drugi podobni programi:

Skye Bender-deMoll, Daniel A. McFarland, James Moody: **SONIA** ([prikazi](#), [program](#), [članek](#)).

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

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

Omrežja ER



Barabasi: The Architecture of Complexity

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

Nekaj spoznanj:

- porazdelitev števila točk z dano stopnjo je Poissonova in večina točk 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

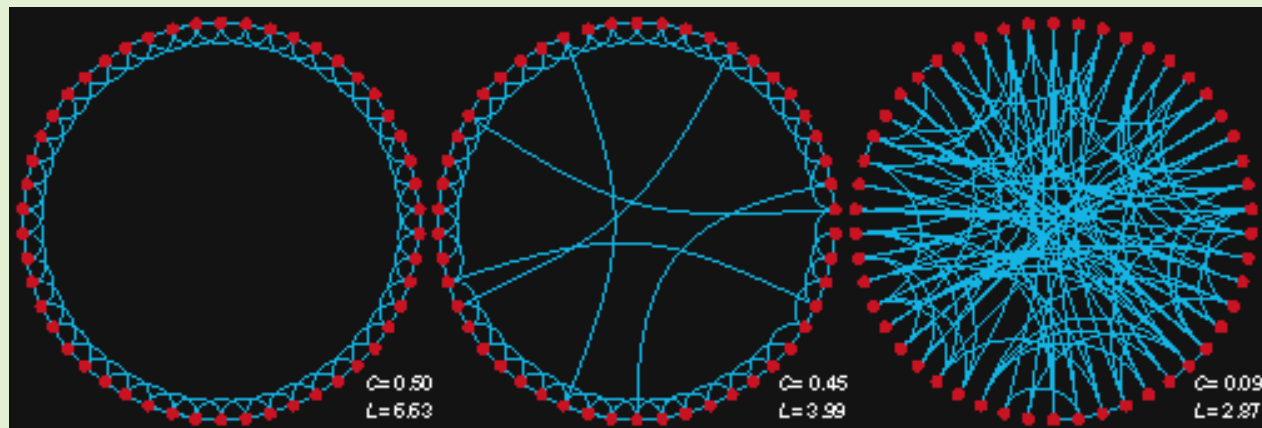
Leta 1967 je psiholog **Stanley Milgram** naredil poskus s pismi. Izbrani posamezniki naj bi s pismi v čim manj korakih dosegli izbrano osebo, pri čemer naj bi vsak posameznik poslal pismo (s temi navodili) svojim znancem, za katere domneva, da so po poznanstvih lahko bliže izbrani osebi, ... Izkazalo se je, da je bilo za dosego 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 točkami majhna, pravimo *mali svetovi* (*small worlds*).

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

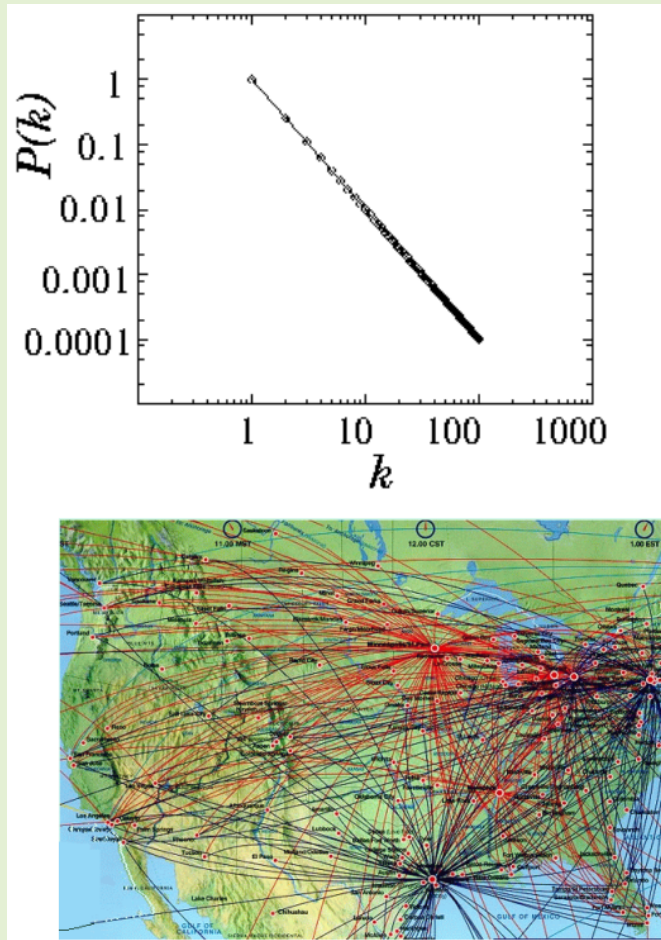
Prevezovanje in mali svetovi

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



Densmore: Power-Law Networks

Brezlestvična omrežja



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 točk (manj kot 0.01%) z zelo *veliko stopnjo* (, ki je v omrežjih ER skoraj neverjetna). Te točke držijo splet skupaj;
- porazdelitev stopenj točk je *potenčna* (premica v dvojno-logaritemskem prikazu)

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

... brezlestvična omrežja

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 nove točke povezujejo slučajnos starimi glede na *priljubljenost* slednjih (*preferential attachment*). Običajno je priljubljenost točke sorazmerna njeni stopnji.

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

... brezlestvična omrežja – določanje eksponenta

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

Graf porazdelitvene funkcije je tudi potenčna funkcija

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

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

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

Pokazati je uspelo tudi, da pri ustreznih pogojih,

- gruče (*modularnost*) in
- *gnezdno povezovanje* – postopek **R-MAT**

ustvarjata brezlestvična omrežja.

Pri izpopolnjevanju modelov so vpeljali še več dodatnih prijemov: nelinearna priljubljenost, začetna priljubljenost, staranje, prevezovanje, zasičenje točk, ...

Razlike med brezlestvičnimi in omrežji ER

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

To spoznanje je pomembno za zaščito omrežij (virusi, vdori, ...). V omrežjih ER se epidemija razširi, ko število okuženih točk 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

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

pregled

Prikazi omrežij

Lepotne zahteve na prikaz omrežja – v oknu za prikaz omrežja izberi možnost Info.

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)
- strukturne luknje
- ...
- ...
- ...