

Analiza velikih omrežij

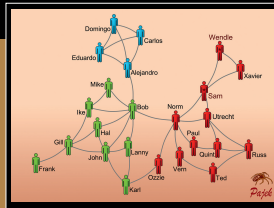
STRUCTURAL ANALYSIS IN THE SOCIAL SCIENCES

46

Exploratory Social Network Analysis with Pajek

REVISED AND EXPANDED EDITION
FOR UPDATED SOFTWARE

THIRD EDITION



Wouter de Nooy, Andrej Mrvar
and Vladimir Batagelj

Andrej Mrvar



Projektna datoteka

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Vse vrste podatkov o omrežju lahko združimo v eno samo datoteko - Pajekovo projektno datoteko - **Pajek Project File**. Najenostavneje to naredimo tako:

- V program Pajek preberemo vse podatkovne datoteke,
- Izračunamo še morebitne dodatne podatke,
- Odstranimo (dispose) morebitne odvečne podatke,
- Shranimo vse skupaj na projektno datoteko:
File / Pajek Project File / Save
- Naslednjič lahko obnovite stanje z eno samo zahtevo
File / Pajek Project File / Read



Pajek

Unicode...

Uporaba Unicode za oznake

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

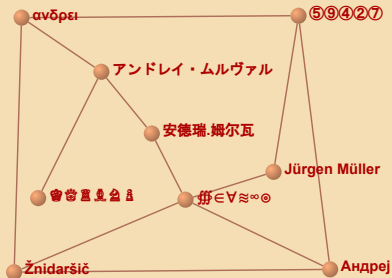
Druge omrežja

Za shranjevanje neangleških oznak (npr. oznake, ki vsebujejo č, š, ž, kitajske, črke...) lahko uporabimo **Unicode**. Pajek uporablja **UTF-8**, za prepoznavanje te vrste datotek pa zahteva tudi **BOM (Byte Order Mark)**. Pajek shranjuje datoteke v formatu UTF-8 with BOM, če v **Options / Read - Write**, izberemo:

Save Files as Unicode UTF-8 in with BOM

Če tega ne izberemo, se datoteke shranjujejo v formatu ASCII, kjer pa se posebni znaki shranijo z uporabo številskih kod: (npr. Ž za ž).

Če č, š, ž vnesemo v načinu cp1250, jih Windows10 avtomatsko pretvori v Unicode.





...Unicode

Uporaba simbolov Unicode za vizualizacijo

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druga omrežja

Simbole Unicode lahko narišemo v sredini točk in tako predstavimo neko lastnost shranjeno v razbitju (npr. spol).

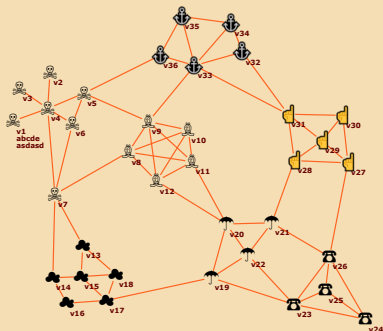
Najprej izberemo kateri simbol Unicode predstavlja kateri razred:

Draw / Options / Symbols for Partition Clusters

potem izberemo neko razbitje kot *Second Partition* in na koncu še:

Options / Mark Vertices Using / Cluster Symbols of Second Partition

Več o uporabi simbolov v prikazih omrežij: <http://mrvar.fdv.uni-lj.si/pajek/Symbols/symbols-examples.htm>





Excel2Pajek, Txt2Pajek

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Povezave podamo kot pare točk, ki niso zaporedna cela števila $1..n$ ampak poljubna števila ali oznake, ki označujejo točke.

Vnesemo jih v dva stolpca v Excelovo datoteko in uporabimo program **ExcelToPajek** za pretvorbo v Pajkovo datoteko.

Če pa uporabimo tekstovno datoteko namesto Excelove, pa oznake ločimo s tabulatorjem in za pretvorbo uporabimo program **Txt2Pajek**.



Pajek

Večrelacijska omrežja...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Omrežje je lahko sestavljeno iz večih relacij na isti množici točk (*povabilo na zabavo, izposoja zapiskov, pogovor o osebnih zadevah...*), ki jih lahko vse zapišemo na eno datoteko. To lahko naredimo na dva načina:

- Za geslom, ki napoveduje opis povezav:
`*arcs, *edges, *arcslist, *edgeslist, *matrix`
dodamo številko relacije in lahko tudi njeno ime. Npr.
`*Arcslist :3 "posojanje gradiv"`
Vse geslu podrejene povezave pripadajo navedeni relaciji (primer: `sampsonmul.net` v 'Testdata').
- Med povezavami, podrejenimi gesloma `*arcs` ali `*edges`, lahko posamezno povezavo pripišemo izbrani relaciji, tako da njen opis začnemo s številko relacije
`3: 47 14 5`
Razlaga: Povezava s krajiščema 47 in 14 ter utežjo 5 pripada relaciji 3.



...Večrelacijska omrežja

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Nekaj Pajkovih ukazov:

Network / Multiple Relations Network / Info – število usmerjenih in neusmerjenih povezav v vsaki relaciji.

Network / Multiple Relations Network / Extract Relation(s) into Separate Networks – izloči samo izbrane relacije iz večrelacijskega omrežja.

Prikaz večrelacijskih omrežij:

Options / Colors / Edges / Relation Number in **Options / Colors / Arcs / Relation Number** - barve povezav bodo določene s številko relacije

Options / Colors / Relation Colors - s katero barvo bo prikazana posamezna relacija.

Options / Lines / Draw Lines / Relations - prikaži samo povezave, ki pripadajo izbrani relaciji, npr. 1-3,6,10-15.



Časovna omrežja...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Redovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Časovna omrežja so omrežja, ki se spreminjajo s časom. V časovnem omrežju določene točke in povezave niso nujno vedno prisotne v vseh časovnih točkah. V časovnem omrežju morajo biti podani časi kdaj so določene točke in povezave aktivne. Te čase podamo med znakoma [in] takole:

- znak – pomeni interval;
- znak , je ločilo med intervali;
- * pomeni neskončno (od danega trenutka dalje).

*Vertices 3

1 "a" [5-10, 12-14]

2 "b" [1-3, 7]

3 "c" [4-*

*Edges

1 2 1 [7]

1 3 1 [6-8]

Točka *a* je aktivna od časa 5 do 10 in od 12 do 14, točka *b* od časa 1 do 3 in v času 7, točka *c* pa od časa 4 naprej. Povezava od 1 do 2 je aktivna samo v času 7, povezava 1 do 3 od časa 6 do 8.



...Časovna omrežja...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj
Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Dodatna omejitev: ko gradimo trenutna omrežja, se določena povezava ne ustvari (tudi če je aktivna), če nista aktivni tudi krajišči povezave. Časovne oznake se morajo vedno nahajati na koncu vrstice, kjer so definirane točke oziroma povezave.

Primer 1: **Sampsonovi menihi** so časovno omrežje.

Sampson je opazoval relacije med 18 menihi v samostanu New England. Opazoval je številne relacije, med drugim prijateljstvo, spoštovanje, vplivnost, odobravanje. Relacije je meril v več časovnih točkah.

Vir: Sampson, S (1969): Crysis in a cloister. Unpublished doctoral dissertation. Cornell University



...Časovna omrežja...

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Primer 2: **LindenStrasse** – igralci in razmerja med njimi v nemški nadaljevanki.

Za vsakega igralca poznamo njegovo/njeno ime, rojstni datum in številne druge podatke. Poznamo tudi številko epozode v kateri je igralec aktivno prisoten. Za vsako povezavo poznamo njen pomen: družinska relacija, poslovna relacija, sovražna relacija, . . .

Lastnosti točk so podane z različnimi oblikami, barvami in velikostmi točk: npr. trikotniki za moške, krožci za ženske; lastnosti povezav so podane z barvami: zelena za sorodstvene povezave, modre za poslovne povezave,

LindenStrasse je časovno pa tudi večrelacijsko omrežje:

- 1 "family relation" (arcs and edges)
- 2 "unfriendly relationships" (edges)
- 3 "business relations" (edges)
- 4 "friendships" (edges)
- 5 "partner relations"



...Časovna omrežja...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj
Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Network / Temporal Network / Generate in Time

Iz časovnega omrežja zgradimo omrežja v danih časovnih trenutkih ali omrežje v danem intervalu. Vnesti moramo prvi čas, zadnji čas in korak (cela števila).

- **All** – Zgradijo se vsa omrežja v izbranih trenutkih.
- **Only Different** – V danem trenutku se zgradi omrežje le, če se bo novo zgrajeno omrežje razlikovalo od zadnjega zgrajenega vsaj v eni točki ali povezavi.
- **Interval** – Zgradimo omrežje v danem časovnem intervalu.



Pajek

...Časovna omrežja...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

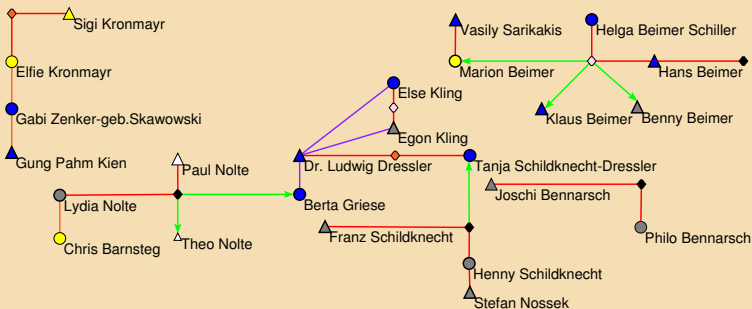
Blogi

Modeli omrežij

Simulacije

Druga omrežja

5. del





...Časovna omrežja...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

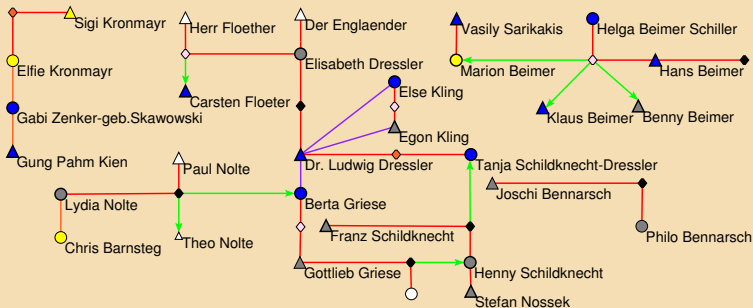
Blogi

Modeli omrežij

Simulacije

Druge omrežja

6. del





...Časovna omrežja...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

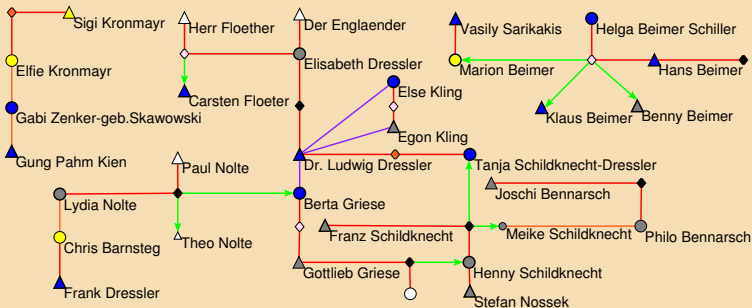
Blogi

Modeli omrežij

Simulacije

Druge omrežja

7. del





Pajek

...Časovna omrežja...

Animacije...

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Z uporabo programa **PajekToSVGAnim** lahko spremembe v časovnih omrežjih prikažemo z animacijami:

- **Sampsonovi menihi:**

<http://mrvar.fdv.uni-lj.si/sola/info4/andrej/Anim/Sampson/sampson-anim.htm>

- **Sampsonovi menihi kot večrelacijsko omrežje:**

<http://mrvar.fdv.uni-lj.si/sola/info4/andrej/Anim/Sampson/sampson-multi-anim.htm>

- **Lindenstrasse:**

<http://mrvar.fdv.uni-lj.si/sola/info4/andrej/Anim/Linden/linden-anim.htm>



Pajek

...Časovna omrežja...

...Animacije...

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Najprej pripravimo vhodno datoteko, pri tem si lahko pomagamo s programom *Pajek*:

- V program *Pajek* naložimo časovno omrežje (npr. `Sampson.net`, `Sampson-Multi.net`, ali `linden.net`).
- Zgeneriramo omrežje v izbranih časovnih točkah: **Network / Temporal Network / Generate in Time / Only Different.**
- V oknu *Draw* poženemo **Export / Append to Pajek Project File / Select File** in izberemo ime za zgenerirano projektno datoteko.
- Najprej izberemo originalno omrežje in ga dodamo v projektno datoteko: **Export / Append to Pajek Project File / Append - F3**
- Prestavimo se na omrežje v prvi časovni točki, po potrebi točke malo premaknemo, in dodamo to omrežje v projektno datoteko (**F3**).
- Prestavimo se na omrežje v naslednji časovni točki, po potrebi točke malo premaknemo, in dodamo to omrežje v projektno datoteko (**F3**).
- Ko dodamo tudi omrežje v zadnji časovni točki, je priprava podatkov za animacije končana.



Pajek

...Časovna omrežja

...Animacije

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Redovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Uporabimo program **PajekToSVGAnim** s katerim skreiramo animacijo:

- S klikom na **Browse** pri polju **Source** izberemo projektno datoteko, ki smo jo ravnokar zgenerirali s Pajkom (npr. `Sampson-Anim.paj`, `Sampson-Multi-Anim.paj`, ali `Linden-Anim.paj`).
- Po potrebi nastavimo nekatere parametre.
- Poženemo: **SVG / Generate - Firefox / Chrome**.

Vse datoteke (začetno časovno omrežje, vhodna projektna datoteka zgenerirana s programom Pajek in animacije) so na voljo tule:

- Sampson:
`http://mrvar.fdv.uni-lj.si/sola/info4/andrej/Anim/Sampson/Sampson.zip`
- Lindenstrasse:
`http://mrvar.fdv.uni-lj.si/sola/info4/andrej/Anim/Linden/Linden.zip`

Pozor: nekatere izbire v HTMLju delujejo pravilno le, če ga pogledamo na strežniku, medtem ko lokalno ne delujejo.



Pajek

Omrežja pridobljena iz besedil...

Slovarji...

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

V slovarskem omrežju so točke v slovarju opisani pojmi; iz pojma *u* vodi povezava do pojma *v* če pojem *v* nastopa v opisu pojma *u*.

ODLIS – Online Dictionary of Library and Information Science

library

From the Latin *liber*, meaning "book." In Greek and the Romance languages, the corresponding term is *bibliotheca*. A collection or group of collections of books and/or other print or nonprint materials organized and maintained for use (reading, consultation, study, research, etc.). Institutional libraries, organized to facilitate access by a specific clientele, are staffed by librarians and other personnel trained to provide services to meet user needs. By extension, the room, building, or facility that houses such a collection, usually but not necessarily built for that purpose. Directory information on libraries is available alphabetically by country in *World Guide to Libraries*, a serial published by K.G. Saur. Two comprehensive worldwide online directories of library homepages are *Libdex* and *Libweb*. See also the UNESCO Libraries Portal. Abbreviated *lib*. **See also:** academic library, government library, monastic library, new library, proto-library, public library, special library, and subscription library.

Also, a collective noun used by publishers, particularly during the Victorian period, for certain books published in series (example: *Everyman's Library*).



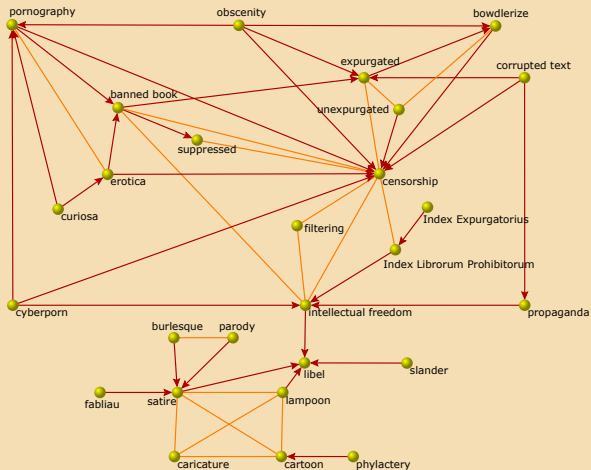
Pajek

...Omrežja pridobljena iz besedil...

...Slovarji...

Omrežja

ODLIS



Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja



Pajek

...Omrežja pridobljena iz besedil...

...Slovarji...

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj
Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

FOLDOC – Free On-Line Dictionary Of Computing

computer

<computer>

A machine that can be programmed to manipulate symbols. Computers can perform complex and repetitive procedures quickly, precisely and reliably and can store and retrieve large amounts of data. Most computers in use today are electronic **digital computers** (as opposed to **analogue computers**).

The physical components from which a computer is constructed are known as **hardware**, which can be of four types: **CPU**, **memory**, **input devices** and **output devices**.



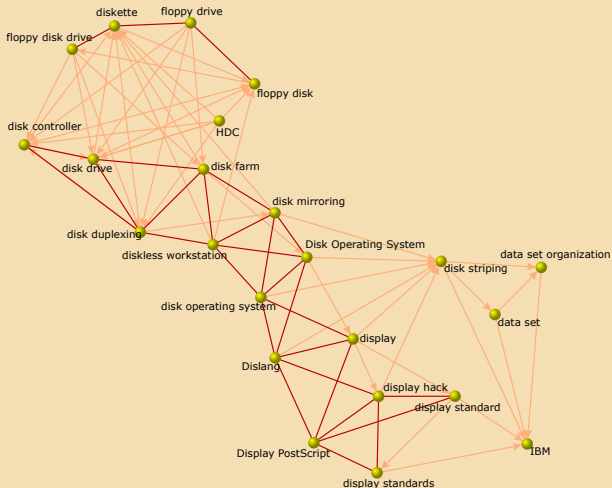
Pajek

...Omrežja pridobljena iz besedil...

...Slovarji

Omrežja

FOLDOC



Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

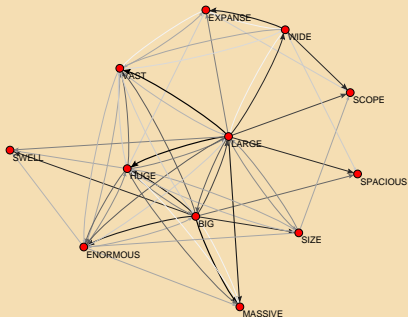
Simulacije

Druge omrežja



Edinburgh Associative Thesaurus - EAT

To je omrežje asociacij med besedami zbranih na študentski populaciji. Točke so besede. Povezave $u \rightarrow v$ pa so določene z vprašanjem: Katera beseda v vam pride prva na misel, ko slišite besedo u ? Utež povezave pove, kolikokrat je bila ta beseda izbrana.





Pajek

...Omrežja pridobljena iz besedil...

Semantična omrežja...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj
Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Semantično omrežje - **WordNet** – A Lexical Database for English

To je večrelacijsko omrežje z naslednjimi relacijami:

- 1 **hypernym pointers - nadpomenke** (`maple.tree`, `tree.plant`, `fire.attack`)
- 2 **entailment pointers - vključenost** (`drive.ride`, `push.press`)
- 3 **similar pointers - sinonimi/sopomenke** (`boiling.hot`)
- 4 **member meronym pointers - sestavni del/je član** (`Luxemburg.Benelux`), (`the Netherlands.Benelux`)
- 5 **substance meronym pointers - je del** (`pavement.street`)
- 6 **part meronym pointers - je del** (`medical-diagnosis.medical-care`)
- 7 **cause pointers - povzročča** (`anesthetize.sleep`, `pasteurize.condense`)
- 8 **grouped verb pointers - grupirani glagoli** (`exit.leave`)
- 9 **attribute pointers - lastnosti** (`quality.bad`)



Pajek

...Omrežja pridobljena iz besedil...

...Semantična omrežja...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

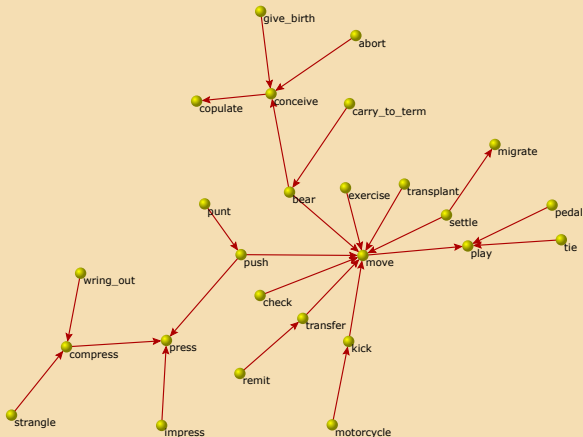
Blogi

Modeli omrežij

Simulacije

Druge omrežja

WordNet - Entailment pointers / vključenost





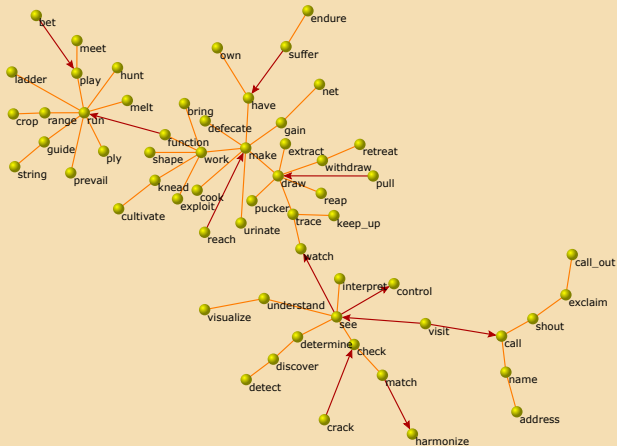
Pajek

...Omrežja pridobljena iz besedil...

...Semantična omrežja...

Omrežja

WordNet - Grouped verb pointers / grupirani glagoli



Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blagi

Modeli omrežij

Simulacije

Druge omrežja



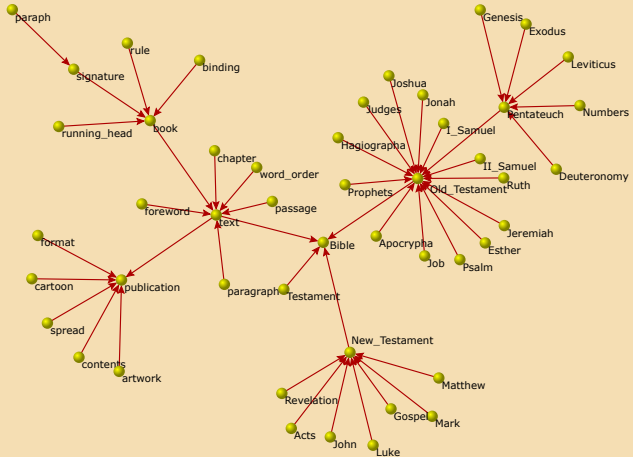
Pajek

...Omrežja pridobljena iz besedil...

...Semantična omrežja...

Omrežja

WordNet - Part meronym pointers / je del



Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja



Pajek

...Omrežja pridobljena iz besedil...

...Semantična omrežja...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

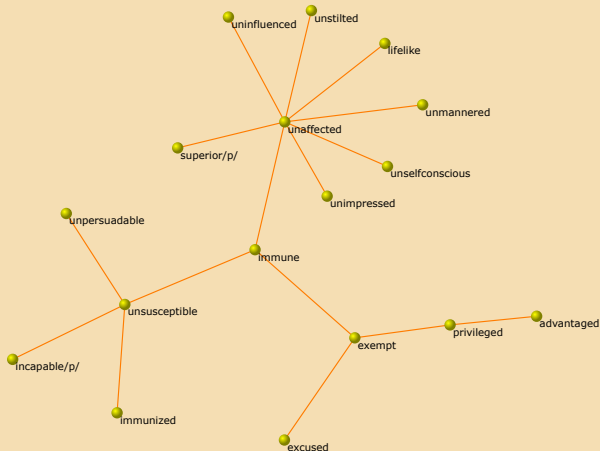
Blogi

Modeli omrežij

Simulacije

Druge omrežja

WordNet - Similar pointers / sinonimi-sopomenke





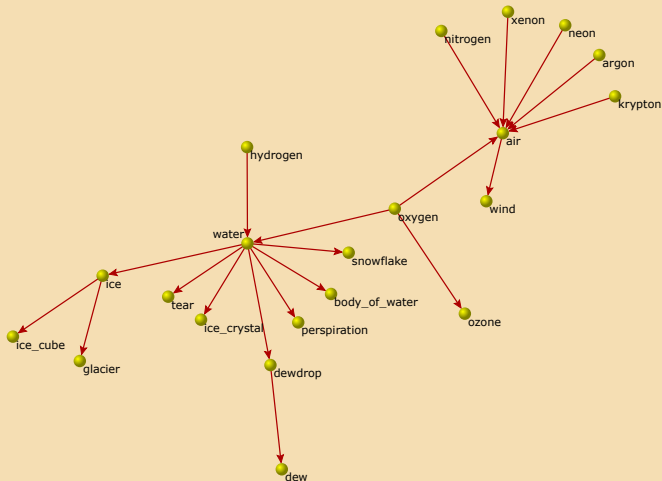
Pajek

...Omrežja pridobljena iz besedil...

...Semantična omrežja

Omrežja

WordNet - Substance meronym pointers / je del



Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja



...Omrežja pridobljena iz besedil

Zamenjava, brisanje, dodajane črk

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

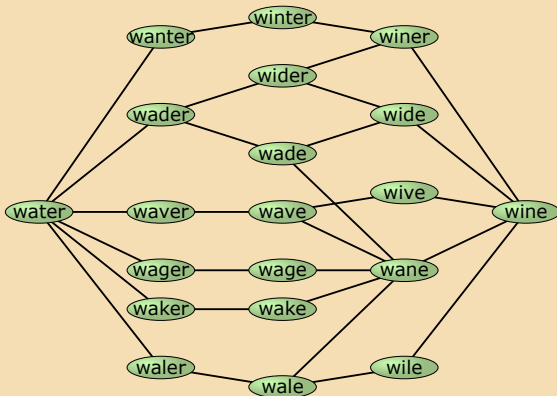
Modeli omrežij

Simulacije

Druge omrežja

Povezave besed: Dve besedi sta povezani, če lahko naredimo eno iz druge z brisanjem / dodajanjem (e. g., kaos — kos) ali zamenjavo (e. g., gora — zora) ene črke.

Primeri: **Dic28**, **SI5**





Aciklična omrežja...

Omrežja citiranj...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

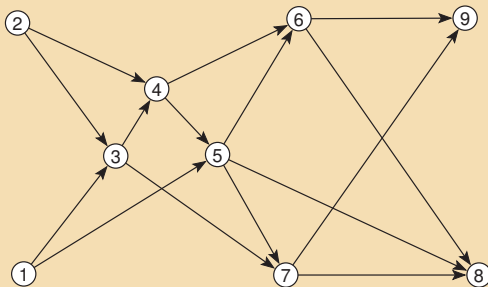
Simulacije

Druga omrežja

Vhodni podatki: Aciklično omrežje, ki predstavlja citiranje med članki, knjigami ali drugimi publikacijami.

Rezultat: Relativne pomembnosti člankov in relativne pomembnosti citiranj.

Primer omrežja citiranj:



V omrežju citiranj predstavljajo točke članke, usmerjena povezava od članka x do članka y pomeni, da y citira x (x je citiran v y).



Pajek

...Aciklična omrežja...

...Omrežja citiranj...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Ugotoviti želimo, kateri članek je najpomembnejši in katera povezava oz. citiranje je bistveno za celotno omrežje.

Najpreprostejša mera pomembnosti je, da preštejemo kolikokrat je dani članek citiran. V naši obliki predstavitve to pomeni, da preštejemo koliko puščic vodi iz dane točke.

Število citiranj (število puščic iz dane točke) dobimo z:

Network / Create Partition / Degree / Output.

Glede na ta kriterij je najpomembnejši članek 5 s tremi citiranjmi.

Boljša mera za pomembnost točke v omrežju je zgrajena na osnovi števila poti, ki vodijo skozi dano točko.



Pajek

...Aciklična omrežja...

...Omrežja citiranj...

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Algoritem za štetje vseh poti v omrežju

- Omrežju dodamo enolični začetek in enolični konec.
- Opravimo dva prehoda skozi omrežje:
 - 1 V prvem prehodu gremo od začetne do končne točke in si za vsako točko zapomnimo, koliko poti vodi vanjo iz začetne točke (to ugotovimo z zaporednim prištevanjem poti glede na točke iz katerih pridemo v dano točko).
 - 2 V drugem prehodu gremo od končne točke nazaj proti začetni in si za vsako točko zapomnimo, koliko poti vodi iz te točke proti končni točki.



Pajek

...Aciklična omrežja...

...Omrežja citiranj...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Redovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Iz dobljenih podatkov lahko za vsako točko v omrežju ugotovimo število vseh poti od začetne do končne točke, ki vodijo skozi to točko. Uporabimo osnovni izrek kombinatorike:

Če iz začetne do izbrane točke vodi M poti, od izbrane do končne pa N poti, potem je število vseh poti, od začetne do končne, ki gredo skozi izbrano točko, enako MN .

Podobno ugotovimo koliko poti uporablja dano povezavo. Naj vodi izbrana povezava od točke i do točke j .

Če iz začetne točke vodi v točko i M različnih poti, iz točke j pa vodi v končno točko N različnih poti, potem je število poti, ki uporablja povezavo (i, j) , enako MN .

Za lažje primerjanje rezultatov iz različnih omrežij ponavadi omenjena števila delimo s številom vseh poti v omrežju, tako da dobimo pomembnosti člankov/citiranj v intervalu $[0, 1]$.



Pajek

...Aciklična omrežja...

...Omrežja citiranj...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj
Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Obstajajo tri variante algoritma:

- metoda **Paths Count (PC)** – to je ravnokar opisana metoda (štejemo poti od začetka do konca);
- metoda **Search Path Link Count (SPLC)** – ta metoda se od prejšnje loči po tem, da vsako točko v omrežju obravnavamo kot začetek, zato iz enoličnega začetka dodamo še povezave do vseh točk omrežja – število vseh poti v omrežju se poveča. Po tej metodi dobijo točke na koncu omrežja višje uteži.
- metoda **Search Path Node Pair (SPNP)** – vsako točko v omrežju obravnavamo kot začetek in kot konec, zato iz enoličnega začetka dodamo povezave do vseh točk omrežja in iz vsake točke omrežja dodamo še povezave do enoličnega konca. Zaradi tega dobijo točke in povezave v sredini omrežja višje uteži.



Pajek

...Aciklična omrežja...

...Omrežja citiranj...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Redovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Absolutne in relativne pomembnosti člankov po metodah Paths Count (PC) in Search Path Link Count (SPLC)

članek	PC	PC/30	SPLC	SPLC/58
1	14	0.4667	14	0.2414
2	16	0.5333	16	0.2759
3	18	0.6000	27	0.4655
4	21	0.7000	35	0.6034
5	20	0.6667	35	0.6034
6	14	0.4667	26	0.4483
7	12	0.4000	22	0.3793
8	17	0.5667	32	0.5517
9	13	0.4333	25	0.4310



Pajek

...Aciklična omrežja...

...Omrežja citiranj...

Omrežja

Absolutne in relativne pomembnosti citiranj po metodah Paths Count (PC) in Search Path Link Count (SPLC)

od	do	PC	PC/30	SPLC	SPLC/58
1	3	9	0.3000	9	0.1551
2	3	9	0.3000	9	0.1551
1	5	5	0.1667	5	0.0862
2	4	7	0.2333	7	0.1207
3	7	4	0.1333	6	0.1034
3	4	14	0.4667	21	0.3621
4	5	15	0.5000	25	0.4310
4	6	6	0.2000	10	0.1724
5	7	8	0.2667	14	0.2414
5	6	8	0.2667	14	0.2414
5	8	4	0.1333	7	0.1207
6	8	7	0.2333	13	0.2241
6	9	7	0.2333	13	0.2241
7	8	6	0.2000	11	0.1897
7	9	6	0.2000	11	0.1897

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Redovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja



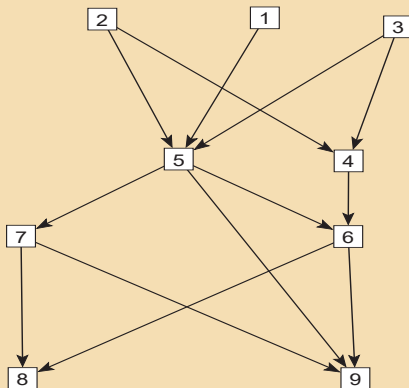
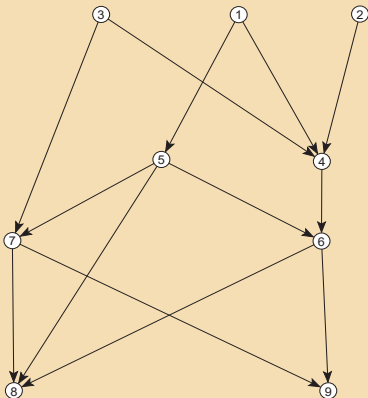
Pajek

...Aciklična omrežja...

...Omrežja citiranj

Omrežja

Še dva manjša primera: citem.net in Citem2a.net



Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja



Pajek

...Aciklična omrežja...

Omrežja citiranj v Pajku...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Pomembnosti člankov in citiranj izračunamo z ukazom:

Network / Acyclic Network / Create Weighted Network + Vector / Traversal Weights / SPC ali **SPLC**

Pajek vrne naslednje rezultate:

- 1 Vektor s številom poti, ki vodijo iz začetne točke v vsako točko (*Number of Different Incoming Paths*).
- 2 Vektor s številom poti, ki vodijo iz vsake točke v končno točko (*Number of Different Outgoing Paths*).
- 3 Vektor s pomembnostmi člankov.
- 4 Omrežje, kjer vrednosti na povezavah predstavljajo pomembnosti citiranj (*Citation weights*). Rezultate pogledamo v oknu **Draw**, kjer povezave označimo z vrednostmi: **Options / Lines / Mark Lines / with Values**. Če želimo ugotoviti, katera citiranja so najpomembnejša, poženemo **Network / Info / General** in odtipkamo koliko najvišjih vrednosti želimo videti (npr. če želimo pogledati 10 najvišjih, vnesemo 10).



Pajek

...Aciklična omrežja...

...Omrežja citiranj v Pajku...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Določanje glavne poti od začetne do končne točke

Network / Acyclic Network / Create (Sub)Network / Main Paths

Lokalne glavne poti: v vsakem koraku iskanja izberemo povezave z največjo utežjo, ki so sosednje trenutni povezavi. Lahko uporabimo tudi toleranco (število med 0 in 1). Toleranca 0 pomeni, da v vsakem koraku upoštevamo samo povezave z največjo utežjo, medtem ko večje tolerance pomenijo, da upoštevamo tudi povezave z malenkost manjšimi utežmi. Dve varianti lokalnega iskanja:

- 1 Local / Forward** – poiščemo lokalne glavne poti od začetnih do končnih točk.
- 2 Local / Backward** – poiščemo lokalne glavne poti v nasprotni smeri (od končnih do začetnih točk).



Pajek

...Aciklična omrežja...

...Omrežja citiranj v Pajku

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Redovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Globalne glavne poti: poiščemo glavne poti z največjo skupno vsoto povezav na poti.

- 1 Global / Standard** – poiščemo glavne poti od začetnih do končnih točk z največjo skupno vsoto povezav na poti. Ta postopek se imenuje tudi Metoda kritične poti.

Če izberemo tudi **Through Vertices in Cluster**, poiščemo lokalne oz. globalne glavne poti, ki vodijo skozi točke, ki se nahajajo v izbrani skupini (Cluster).



Pajek

...Aciklična omrežja...

Rodovniki...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Ljudje zbirajo rodoslovne podatke za različne namene:

- Raziskave različnih kultur v sociologiji, antropologiji in zgodovini – sorodstvo kot osnovna socialna relacija.
- Rodovniki družin in ali teritorialnih enot, npr.:
 - rodovnik dubrovniških plemiških rodbin (Ragusa)
 - rodovniki Mormonov:
<http://www.familytreemaker.com/>
 - rodovnik ljudi živečih na področju Škofje Loke:
 - rodovnik ameriških predsednikov
- Posebni rodovniki
 - Študenti in njihovi mentorji pri izdelavi doktoratov:
 - Theoretical Computer Science Genealogy:
<http://sigact.acm.org/genealogy/>
 - Mathematics



Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druga omrežja

GEDCOM je standard za shranjevanje rodoslovnih podatkov, ki se uporablja tudi za izmenjavo podatkov pridobljenih z različnimi programi za vnos. Naslednje vrstice so izločene iz datoteke GEDCOM Evropskih kraljevih družin:

```

0 HEAD
1 FILE ROYALS.GED
...
0 @I58@ INDI
1 NAME Charles Philip Arthur/Windsor/
1 TITL Prince
1 SEX M
1 BIRT
2 DATE 14 NOV 1948
2 PLAC Buckingham Palace, London
1 CHR
2 DATE 15 DEC 1948
2 PLAC Buckingham Palace, Music Room
1 FAMS @F16@
1 FAMC @F14@
...
...
0 @I65@ INDI
1 NAME Diana Frances /Spencer/
1 TITL Lady
1 SEX F
1 BIRT
2 DATE 1 JUL 1961
2 PLAC Park House, Sandringham
1 CHR
2 PLAC Sandringham, Church
1 FAMS @F16@
1 FAMC @F78@
...
...
0 @I115@ INDI
1 NAME William Arthur Philip/Windsor/
1 TITL Prince
1 SEX M
1 BIRT
2 DATE 21 JUN 1982
2 PLAC St.Mary's Hospital, Paddington
1 CHR
2 DATE 4 AUG 1982
2 PLAC Music Room, Buckingham Palace
1 FAMC @F16@
...
0 @I116@ INDI
1 NAME Henry Charles Albert/Windsor/
1 TITL Prince
1 SEX M
1 BIRT
2 DATE 15 SEP 1984
2 PLAC St.Mary's Hosp., Paddington
1 FAMC @F16@
...
0 @F16@ FAM
1 HUSB @I58@
1 WIFE @I65@
1 CHIL @I115@
1 CHIL @I116@
1 DIV N
1 MARR
2 DATE 29 JUL 1981
2 PLAC St.Paul's Cathedral, London

```



Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Predstavitev rodovnikov z omrežji

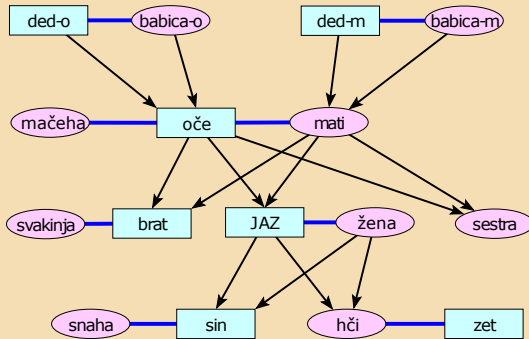
Rodovnike lahko predstavimo z omrežji na več različnih načinov:

- kot navadne rodovnike (Ore-graph),
- kot parne rodovnike (p-graph),
- kot dvodelne parne rodovnike (bipartite p-graph).



Navadni rodovnik:

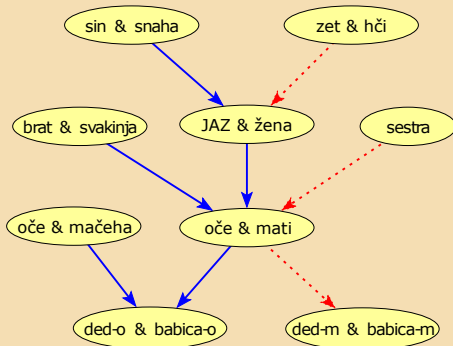
V navadnem rodovniku je vsaka oseba predstavljena s svojo točko, poroke so predstavljene z neusmerjenimi povezavami, relacija *je eden od staršev od* pa z usmerjenimi povezavami, ki kažejo od vsakega od staršev do njihovih otrok.





Parni rodovnik:

V parnem rodovniku predstavljajo točke posameznike ali pare. Če neka oseba še ni poročena, je predstavljena s svojo točko, sicer je predstavljena s svojim partnerjem v skupni točki. V parnem rodovniku obstajajo samo usmerjene povezave, ki kažejo od otrok na njihove starše.





Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

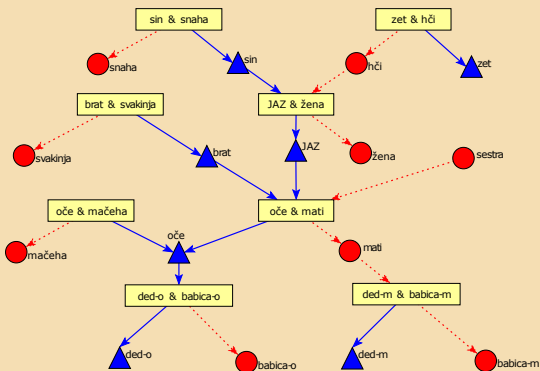
Modeli omrežij

Simulacije

Druge omrežja

Dvodelni parni rodovnik:

ima dve vrsti točk – točke, ki predstavljajo poročene pare (pravokotniki) in točke, ki predstavljajo posameznike (okrogle točke za ženske in trikotniki za moške). Usmerjene povezave tudi tu kažejo od otrok na njihove starše.





Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Rodovniki so redka omrežja:

podatki	navadni rodovnik				parni rodovnik			
	$ V $	$ E $	$ A $	$\frac{ L }{ V }$	$ V_{ip} $	$ V_{cp} $	$ A_p $	$\frac{ A_p }{ V_p }$
Bruno	15512	4841	18664	1.52	6000	5289	10053	0.89
Combo	20350	7248	26199	1.64	6931	7945	14845	1.00
Dodderer	16761	5650	22425	1.68	6029	5652	11765	1.01
Drame	29606	8256	41814	1.69	13254	8939	21862	0.99
Little	25968	8778	34640	1.67	9212	8850	18233	1.01
President	2145	978	2223	1.49	218	1042	1222	0.97
Tillotsn	42559	12796	54043	1.57	15177	15959	31234	1.00
Loka	47956	14154	68052	1.71	19189	16039	36192	1.03
Silba	6427	2217	9627	1.84	2001	2479	5281	1.18
Ragusa	5999	2002	9315	1.88	2066	2310	5336	1.22
Tur	1269	407	1987	1.89	0	956	1114	1.17
Royal	3010	1138	3724	1.62	719	1422	2259	1.06



Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

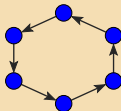
Modeli omrežij

Simulacije

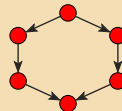
Druge omrežja

Prednosti parnih rodovnikov

- V parnem rodovniku je manj točk in povezav.
- Parni rodovniki so usmerjena aciklična omrežja.



Cikel:



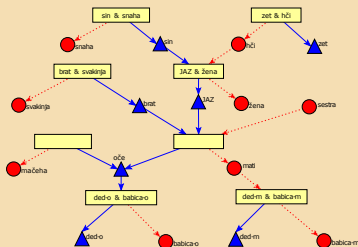
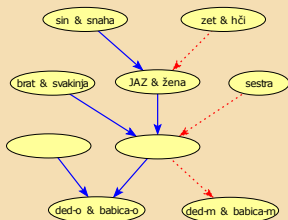
Sklenjena veriga:

- V parnem rodovniku pomeni vsaka sklenjena veriga (semi-cycle) *prepletenost porok*. Obstajata dve vrsti prepletenosti:
 - krvne poroke: npr., poroka med bratom in sestro;
 - ne-krvne poroke: npr., dva brata se poročita z dvema sestrama iz druge družine.
- Parni rodovniki so bolj primerni za analize.



Prednosti dvodelnih parnih rodovnikov

Pri dvodelnih parnih rodovnikih lahko dodatno razlikujemo še med *poročenim stricem in ponovno poročenim očetom* ter tako poiščemo npr. *poroke med polbrati in polsestrami*.





Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Indeks prepletenosti (*relinking index*) meri, kako pogosto se pripadniki istih družin poročajo med sabo. Poseben primer prepletenosti so krvne poroke.

Naj bo n število točk v parnem rodovniku, m število povezav in M število končnih točk (točk z izhodno stopnjo 0, $M \geq 1$).

V povezanem rodovniku velja

$$RI = \frac{m - n + 1}{n - 2M + 1}$$

Za omrežje z eno samo točko postavimo $RI = 0$.

- $0 \leq RI \leq 1$
- Če je rodovnik gozd/drevo, potem je $RI = 0$ (ni prepletenosti).
- Obstajajo rodovniki z $RI = 1$ (največja možna prepletenost).



Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

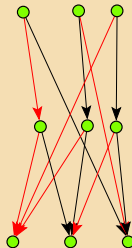
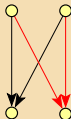
Blogi

Modeli omrežij

Simulacije

Druge omrežja

Vzorci z indeksom prepletenosti enakim 1





Pajek

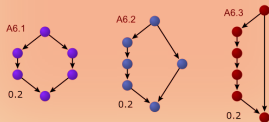
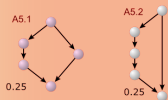
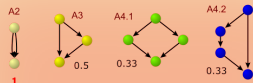
...Aciklična omrežja...

...Rodovniki...

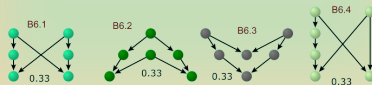
Omrežja

Prepletene poroke (parni rodovniki z 2 do 6 točkami)

Krvne poroke



Prepletene poroke



Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja



Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Še nekaj primerov krvnih porok z $RI=1$

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

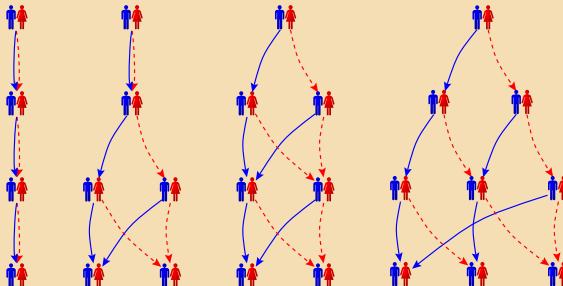
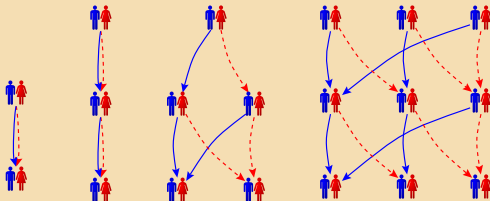
Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja





Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

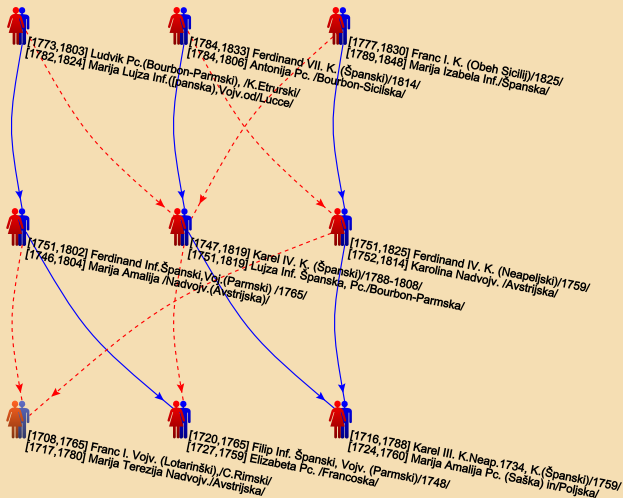
Blogi

Modeli omrežij

Simulacije

Druge omrežja

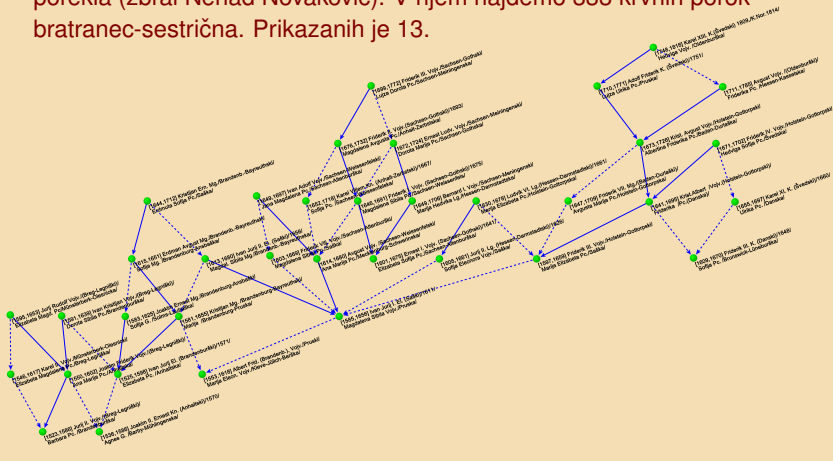
Številne izmenjave partnerjev v samo dveh generacijah





Evropsko plemstvo

Rodovnik sicer vsebuje podatke za okrog 60 tisoč oseb plemiškega porekla (zbral Nenad Novaković). V njem najdemo 333 krvnih porok bratranec-sestrična. Prikazanih je 13.



- Projektna datoteka
- Unicode
- Excel2Pajek
- Večrelacijska o.
- Časovna omrežja
- Besedilna omr.
- Aciklična omrežja
- Omrežja citiranj
- Rodovniki
- Nakopičenost
- Otoki
- Skupnosti
- E-I Index
- Razbitja
- Kratki cikli
- Blogi
- Modeli omrežij
- Simulacije
- Druga omrežja



Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Rodovnike lahko primerjamo po tem kolikokrat se v njih pojavljajo karakteristični vzorci. Kot primer smo vzeli pet rodovnikov:

- **Loka.ged** – rodovnik s področja škofje Loke (P. Hawlina).
- **Silba.ged** – rodovnik hrvaškega otoka Silba (P. Hawlina).
Poseben geografski položaj.
- **Ragusa.ged** – poroke med dubrovniškimi plemiškimi družinami med 12 in 16 stoletjem. Podatke je zbrala I. Mahnken (1960); v elektronsko obliko pa jih je vnesla P. Dremelj (1999).
Zelo omejena pravila dovoljenih porok – pripadnik plemiške družine naj bi se poročil s pripadnikom druge plemiške družine.
- **Tur.ged** – rodovnik turškega nomadskega plemena Yörük. Podatke sta zbrala Ulla C. Johansen in D.R. White (2001)
Prepletena poroka je znak pripravljenosti ostati v nomadski skupini.
- **Royal.ged** – rodovnik evropskih kraljevih družin.



Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Frekvenčna porazdelitev vzorcev

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

















Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

vzorec	Loka	Silba	Ragusa	Tur	Royal	Σ
 A2	1	0	0	0	0	1
 A3	1	0	0	0	3	4
 A4.1	12	5	3	65	21	106
 B4	54	25	21	40	7	147
 A4.2	0	0	0	0	0	0
 A5.1	9	7	4	15	13	48
 A5.2	0	0	0	0	0	0
 B5	19	11	47	19	8	104
 A6.1	28	28	2	65	13	140
 A6.2	0	2	0	0	1	3
 A6.3	0	0	0	0	0	0
 C6	10	12	19	15	5	61
 B6.1	0	1	2	0	0	3
 B6.2	27	39	63	54	12	194
 B6.3	47	30	82	46	13	218
 B6.4	0	0	5	3	0	8
Št. pos.	47956	6427	5999	1269	3010	
Naj. dv. komp.	4095	1340	1446	250	435	
RI	0.55	0.78	0.74	0.75	0.37	



Opazanja

- Generacijski preskoki za več kot eno generacijo so zelo malo verjetni.
- Obstaja veliko porok tipa B6.3 (dva vnuka se poročita v isto družino) in B6.2 (dve družini sta povezani s poroko med otrokoma in kasneje še enkrat s poroko med vnukoma)
- V rodovniku Tur obstaja veliko porok A4.1 in A6.1.
- V vseh rodovnikih je število 'ne-krvnih' porok veliko večje od števila krvnih porok (kar še posebej velja za rodovnik Ragusa, izjema je Royal). Za tako vrsto porok obstajajo ekonomski razlogi: ohraniti bogastvo in moč znotraj izbranih družin.

vrsta poroke	Loka	Silba	Ragusa	Tur	Royal
krvne poroke	51	42	9	149	51
ostale prepletene poroke	157	118	239	176	45

Število posameznikov v rodovniku Tur je veliko manjše kot v ostalih rodovnikih, Silba in Ragusa sta približno enako velika, medtem je Loka veliko večji rodovnik, kar moramo pri primerjavah pravtako upoštevati.



Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

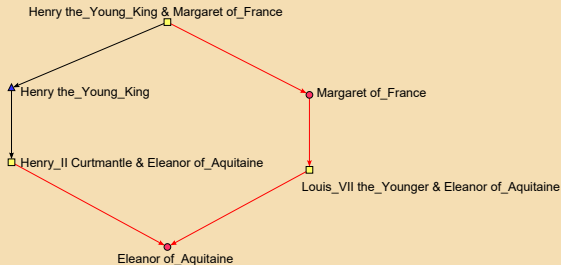
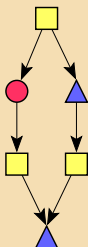
Frekvence normalizirane s številom parov v rodovniku $\times 1000$

	vzorec	Loka	Silba	Ragusa	Tur	Royal
Projektna datoteka	 A2	0.07	0.00	0.00	0.00	0.00
Unicode	 A3	0.07	0.00	0.00	0.00	2.64
Excel2Pajek	 A4.1	0.85	2.26	1.50	159.71	18.45
Večrelacijska o.	 B4	3.82	11.28	10.49	98.28	6.15
Časovna omrežja	 A4.2	0.00	0.00	0.00	0.00	0.00
Besedilna omr.	 A5.1	0.64	3.16	2.00	36.86	11.42
Aciklična omrežja	 A5.2	0.00	0.00	0.00	0.00	0.00
Omrežja citiranj	 B5	1.34	4.96	23.48	46.68	7.03
Rodovniki	 A6.1	1.98	12.63	1.00	169.53	11.42
Nakopičenost	 A6.2	0.00	0.90	0.00	0.00	0.88
Otoki	 A6.3	0.00	0.00	0.00	0.00	0.00
Skupnosti	 C6	0.71	5.41	9.49	36.86	4.39
E-I Index	 B6.1	0.00	0.45	1.00	0.00	0.00
Razbitja	 B6.2	1.91	17.59	31.47	130.22	10.54
Kratki cikli	 B6.3	3.32	13.53	40.96	113.02	11.42
Blogi	 B6.4	0.00	0.00	2.50	7.37	0.00
Modeli omrežij	Σ	14.70	72.17	123.88	798.53	84.36
Simulacije						
Druge omrežja						



Dvodelni parni rodovniki: Poroka med polbratom in polsestro

Pri parnih rodovnikih iz same strukture omrežja ne moremo prepoznati večkrat poročenih oseb. V tem primeru moramo uporabiti dvodelne parne rodovnike. V njih lahko najdemo poroke med polbrati in polsestrami.





Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

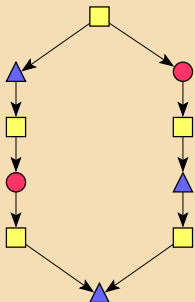
Blogi

Modeli omrežij

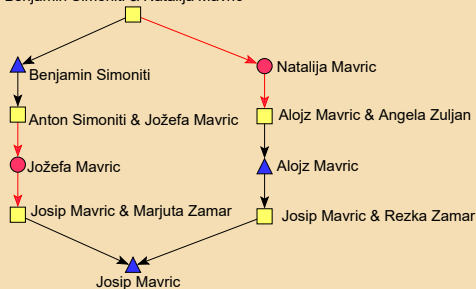
Simulacije

Druge omrežja

Dvodelni parni rodovniki: Poroka med polbratrancem - polsestrično



Benjamin Simoniti & Natalija Mavric





Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Druge analize

Posameznike, ki zbirajo podatke o svojih družinah, zanimajo še številne druge podrobnosti:

- posebnosti: velikokrat poročene osebe, posamezniki z velikim številom otrok;
- preverjanje ali sta dve osebi v sorodstveni zvezi in iskanje najkrajše sorodstvene vezi med njima;
- iskanje vseh prednikov/potomcev izbrane osebe in iskanje osebe z največjim znanim številom prednikov ali potomcev;
- največja razlika v starosti med možem in ženo, najstarejša/najmlajša oseba ob času poroke, najstarejša/najmlajša oseba ob rojstvu otroka;
- spreminjanje vzorcev porok skozi čas;
- iskanje najdaljše moške/ženske verige;
- posebni primeri → napake storjene pri vnosu podatkov (konsistentnost).



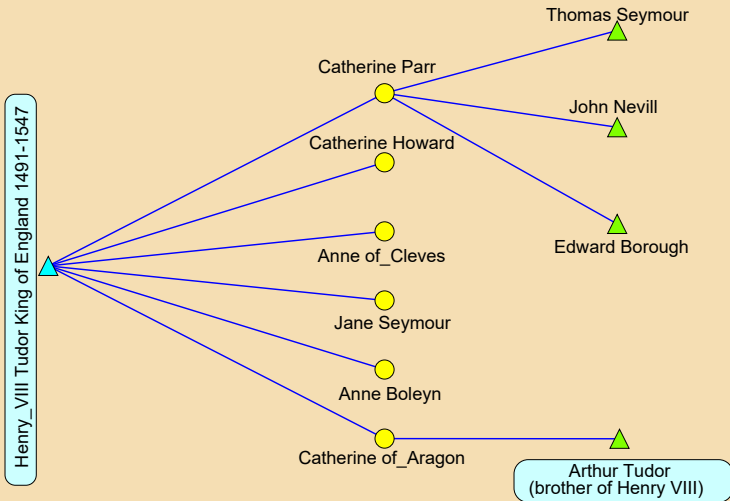
Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Največkrat poročena oseba



Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

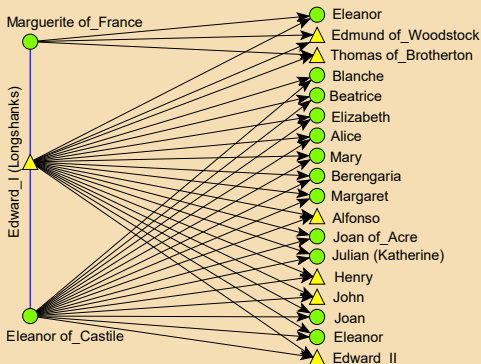
Modeli omrežij

Simulacije

Druge omrežja



Največje število otrok



Angleški kralj Edward I (1237-1307) in njegova žena kraljica Eleanor (1241-1290) sta imela 16 otrok rojenih med 1255-1284 (na sliki manjka hčerka brez imena). Najmlajši sin (Edward) je bil prvi od sinov, ki je preživel nevarna otroška leta. Eleanor je morala poskusiti šestnajstkrat, da je izpolnila najpomembnejšo nalogo angleške kraljice: svojemu možu zagotoviti naslednika moškega spola. Do desetega leta je umrlo 10 od 16 otrok, več kot 40 let so dočakali le trije.



Pajek

...Aciklična omrežja...

...Rodovniki...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

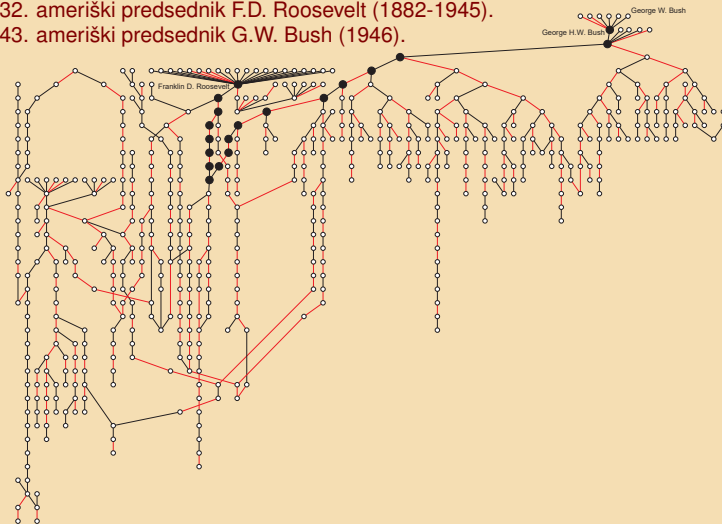
Simulacije

Druga omrežja

Iskanje sorodstvenih vezi

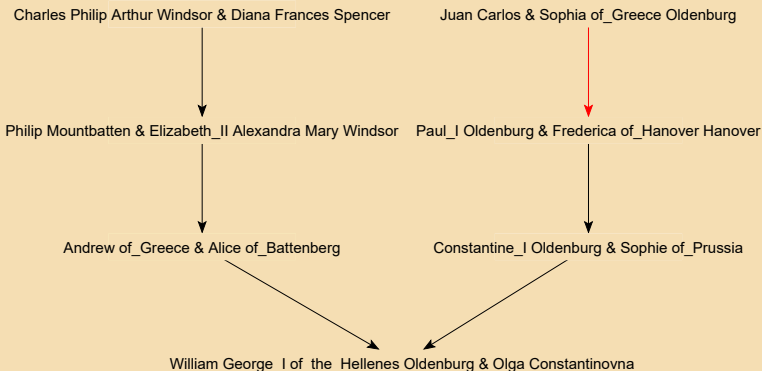
32. ameriški predsednik F.D. Roosevelt (1882-1945).

43. ameriški predsednik G.W. Bush (1946).





Najkrajša sorodstvena zveza med angleškim princem *Charlesom* in bivšim španskim kraljem *Juanom Carlosom*





Nakopičenost - Clustering Coefficient...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Nakopičenost v točki omrežja je določena kot razmerje med številom vseh povezav med sosedi dane točke in največjim možnim številom povezav med sosedi. Največje možno število povezav med vsemi sosedi dobimo, če so vsi sosedi povezani med sabo.

Naj bo deg_v stopnja točke v , $|E(G^1(v))|$ število povezav med točkmi v 1-okolici točke v , Δ največja stopnja točke v omrežju in $|E(G^2(v))|$ število povezav med točkami v 2-okolici točke v .

- **CC1** – mere, ki upošteevajo le 1-okolico točke:

$$CC_1(v) = \frac{2|E(G^1(v))|}{deg_v(deg_v - 1)}$$

$$CC'_1(v) = \frac{deg_v}{\Delta} CC_1(v)$$

$CC_1(v)$ je *Clustering Coefficient*, kot sta ga definirala Watts in Strogatz, 1998.



...Nakopičenost - Clustering Coefficient

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

- **CC2** – mere, ki upoštevajo 2-okolico točke:

$$CC_2(v) = \frac{|E(G^1(v))|}{|E(G^2(v))|}$$

$$CC'_2(v) = \frac{deg_v}{\Delta} CC_2(v)$$

Če je $deg_v \leq 1$, zavzamejo vsi koeficienti za točko v vrednost 999999998 (manjkajoča vrednost).

V oknu Report se izpišeta tudi Watts-Strogatz Clustering Coefficient (Transitivity) in Network Clustering Coefficient. Nakopičenost ali tranzitivnost omrežja je delež vseh poti na dveh točkah v omrežju, ki je zaprt s tretjo povezavo.

Ukaz v Pajku:

Network / Create Vector / Clustering Coefficients



Pajek

Otoki...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

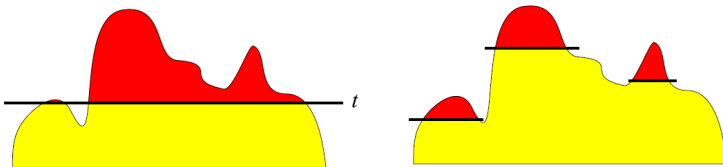
Blogi

Modeli omrežij

Simulacije

Druge omrežja

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





Pajek

...Otoki

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Otoki na povezavah:

Network / Create Partition / Islands / Line Weights

Otoki na točkah (kot vhod rabimo poleg omrežja še vektor, ki določa vrednosti v točkah):

Operations / Network + Vector / Islands / Vertex Weights



Iskanje skupnosti - Community Detection...

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Poiskati želimo take skupine, da bodo povezave znotraj skupin gostejše (in z večjimi vrednostmi), povezave med skupinami pa redkejša (in z manjšimi vrednostmi).

V Pajku sta na voljo dve metodi za iskanje skupin po metodah 'community detection' in sicer metoda **Louvain** in metoda **VOS Clustering**.

Po metodi **Louvain** iščemo razvrstitev v skupine, ki ima največjo **modularnost (modularity - Q)**:

$$Q = \frac{1}{2m} \sum_s (e_s - r * \frac{K_s^2}{2m})$$

- m – skupno "stevilo povezav,
- s – skupina,
- $e_s = \sum_{ij \in s} A_{ij}$ – dvakratno "stevilo povezav znotraj skupine s
- $K_s = \sum_{i \in s} k_i$ – vsota stopenj za to "cke v skupini s
- r – *resolution parameter*, privzeta vrednost je 1, ki ustreza osnovni definiciji modularnosti



...Iskanje skupnosti - Community Detection

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Podobna metoda je **VOS Clustering**, le da se namesto modularnosti vzame *VOS quality function*.

Obe metodi sta na voljo v izbiri:

Network / Create Partition / Communities

Na voljo je precej parametrov s katerimi usmerjamo iskanje skupin, ki pa jih za naše potrebe ni potrebno spreminjati (v poštev pridejo pri zelo velikih omrežjih).

Smiselno pa je preizkusiti različne vrednosti parametra **resolution**. Ta je ponavadi nastavljen na vrednost 1. Vrednost večja od 1 pomeni iskanje večjega števila (manjših) skupin, vrednost manjša od 1 pa iskanje manjšega števila (večjih) skupin.



E-I Index

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Obstaja enostavna mera za preverjanje kako dobro izračunane skupine razdelijo omrežje na kohezivne skupine. Imenuje se **E-I Index: External-Internal Index**. Kot je razvidno iz imena, ta indeks izračunamo tako, da od števila povezav med skupinami odštejemo število povezav znotraj skupin in razliko delimo s skupnim številom povezav. V poštev lahko vzamemo tudi vrednosti na povezavah.

Rezultat je število med -1 in 1 . Če je E-I Index enak -1 , so vse povezave znotraj skupin, če je 1 pa so vse med skupinami. Vrednost 0 dobimo če je število povezav (ali vsota vrednosti na povezavah) med skupinami enaka številu povezav (ali vsoti vrednosti na povezavah) znotraj skupin. Če razbitje dobro razdeli točke v kohezivne skupine, mora biti večina povezav znotraj skupin, torej bo E-I Index negativen in blizu -1 .

Operations / Network + Partition / Info / E-I Index



Pajek

Primerjave razbitij...

Cramerjev koeficient

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Ponavadi so si skupine dobljene po obeh metodah zelo podobne. Podobnost lahko preverimo s **Partitions / Info**, ki nam vrne **Cramerjev** in še nekatere druge koeficiente za primerjavo dveh razbitij.

Obstajajo številne mere za primerjavo dveh razbitij. Razbitja ponavadi predstavljajo nominalne spremenljivke, ki jih lahko predstavimo s kontingentnimi tabelami

Partitions / Info / Cramer's V, Rajski, Adjusted Rand Index

Za merjenje povezanosti poljubnih dveh spremenljivk (lahko tudi nominalnih) lahko uporabljamo Cramerjev koeficient (*Cramer's V*).

$$V = \sqrt{\frac{\chi^2}{n(k-1)}}$$



Pajek

...Primerjave razbitij...

Koeficienti Rajskega...

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Koeficient Rajskega (1964) je zgrajen na osnovi entropije: Imejmo dve spremenljivki X in Y . Spremenljivka X naj zavzame n različnih vrednosti, spremenljivka Y pa m različnih vrednosti.

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i)$$

$$H(Y) = - \sum_{i=1}^m p(y_i) \log_2 p(y_i)$$

in

$$H(XY) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(x_i, y_j)$$



Informacija med spremenljivkama X in Y je definirana takole:

$$I(X, Y) = H(X) + H(Y) - H(XY)$$

Informacija $I(X, Y)$ doseže vrednost 0, natanko takrat ko za vsak par x_i in y_j velja $p(x_i, y_j) = p(x_i)p(y_j)$, kar pomeni, da sta spremenljivki neodvisni.

Informacija $I(X, Y)$ doseže največjo vrednost, natanko takrat ko med spremenljivkama obstaja funkcijska zveza – v vsakem stolpcu in vsaki vrstici ustrezne kontingenčne tabele je največ en od nič različen element. Tedaj velja:

$$H(X) = H(Y) = H(XY) = I(X, Y)$$

Torej je informacija $I(X, Y)$ mera funkcijske odvisnosti (določenosti) med spremenljivkama X in Y .



Pajek

...Primerjave razbitij...

...Koeficienti Rajskega...

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Koeficienti Rajskega:

$$R(X \leftrightarrow Y) = \frac{I(X, Y)}{H(XY)}$$

$$R(X \rightarrow Y) = \frac{I(X, Y)}{H(Y)}$$

$$R(X \leftarrow Y) = \frac{I(X, Y)}{H(X)}$$

Vsi koeficienti zavzamejo vrednosti med 0 in 1. Vrednost 0 zavzamejo, ko sta spremenljivki neodvisni.

$R(X \rightarrow Y) = 1$, ko je Y funkcija X ,

$R(X \leftarrow Y) = 1$, ko je X funkcija Y in

$R(X \leftrightarrow Y) = 1$, ko obe spremenljivki ena drugo natanko določata.



Primer 1:

	y_1	y_2	y_3	Sum
x_1	2	2	1	5
x_2	2	1	2	5
Sum	4	3	3	10

$p(x_i, y_j)$	y_1	y_2	y_3	$p(x_i)$
x_1	0.2	0.2	0.1	0.5
x_2	0.2	0.1	0.2	0.5
$p(y_j)$	0.4	0.3	0.3	1

$$R(X \leftrightarrow Y) = 0.0194$$

$$R(X \rightarrow Y) = 0.0312$$

$$R(X \leftarrow Y) = 0.0490$$

Vrednosti vseh treh koeficientov so majhne, na osnovi vrednosti ene spremenljivke ne moremo napovedati vrednosti druge spremenljivke.



Primer 2:

	y_1	y_2	y_3	Sum
x_1	0	3	0	3
x_2	4	0	3	7
Sum	4	3	3	10

$p(x_i, y_j)$	y_1	y_2	y_3	$p(x_i)$
x_1	0	0.3	0	0.3
x_2	0.4	0	0.3	0.7
$p(y_j)$	0.4	0.3	0.3	1

$$R(X \leftrightarrow Y) = 0.5610$$

$$R(X \rightarrow Y) = 0.5610$$

$$R(X \leftarrow Y) = 1$$

Spremenljivka X je funkcija Y : če poznamo vrednost spremenljivke Y lahko natanko napovemo vrednost spremenljivke X , obratno pa ni res.



Pajek

...Primerjave razbitij...

...Koefficienti Rajskega...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Redovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Primer 3:

	y_1	y_2	y_3	Sum
x_1	4	0	0	4
x_2	0	0	3	3
x_3	0	3	0	3
Sum	4	3	3	10

$$R(X \leftrightarrow Y) = 1$$

$$R(X \rightarrow Y) = 1$$

$$R(X \leftarrow Y) = 1$$

Spremenljivki X in Y ena drugo natanko določata.



Pajek

...Primerjave razbitij...

...Koeficienti Rajskega

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Primer 4:

	y_1	y_2	Sum
x_1	2	2	4
x_2	2	2	4
Sum	4	4	8

$$R(X \leftrightarrow Y) = 0$$

$$R(X \rightarrow Y) = 0$$

$$R(X \leftarrow Y) = 0$$

Spremenljivki X in Y sta neodvisni



Pajek

...Primerjave razbitij

Adjusted Rand Index

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druga omrežja

Adjusted Rand index [\[edit\]](#)

The adjusted Rand index is the corrected-for-chance version of the Rand index.^{[1][2][3]} Such a correction for chance establishes a baseline by using the expected similarity of all pair-wise comparisons between clusterings specified by a random model. Traditionally, the Rand Index was corrected using the Permutation Model for clusterings (the number and size of clusters within a clustering are fixed, and all random clusterings are generated by shuffling the elements between the fixed clusters). However, the premises of the permutation model are frequently violated; in many clustering scenarios, either the number of clusters or the size distribution of those clusters vary drastically. For example, consider that in **K-means** the number of clusters is fixed by the practitioner, but the sizes of those clusters are inferred from the data. Variations of the adjusted Rand Index account for different models of random clusterings.^[4]

Though the Rand Index may only yield a value between 0 and +1, the adjusted Rand index can yield negative values if the index is less than the expected index.^[5]

The contingency table [\[edit\]](#)

Given a set S of n elements, and two groupings or partitions (e.g. clusterings) of these elements, namely $X = \{X_1, X_2, \dots, X_r\}$ and $Y = \{Y_1, Y_2, \dots, Y_s\}$, the overlap between X and Y can be summarized in a contingency table $[n_{ij}]$ where each entry n_{ij} denotes the number of objects in common between X_i and Y_j : $n_{ij} = |X_i \cap Y_j|$.

$X \setminus Y$	Y_1	Y_2	\dots	Y_s	sums
X_1	n_{11}	n_{12}	\dots	n_{1s}	a_1
X_2	n_{21}	n_{22}	\dots	n_{2s}	a_2
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
X_r	n_{r1}	n_{r2}	\dots	n_{rs}	a_r
sums	b_1	b_2	\dots	b_s	

Definition [\[edit\]](#)

The original Adjusted Rand Index using the Permutation Model is

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}$$

where n_{ij} , a_i , b_j are values from the contingency table.

https://en.wikipedia.org/wiki/Rand_index



Pajek

Kratki cikli...

Trikotniki (3-obroči)

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

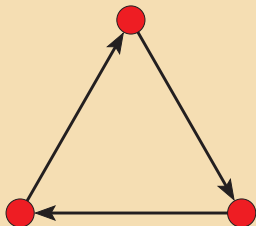
Druge omrežja

Za vsako povezavo preštejemo koliko trikotnikom pripada. Števila se shranijo kot vrednosti na povezavah.

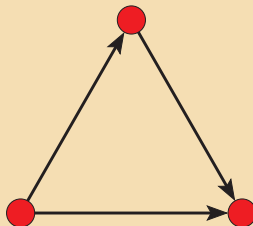
Network / Create New Network / with Ring Counts stored as Line Values / 3-Rings

- **Undirected** – preštejemo število trikotnikov v neusmerjenem omrežju.
- **Directed** – preštejemo ciklične, tranzitivne ali vse trikotnike v usmerjenem omrežju, ali pa kolikokrat neka povezava nastopa kot bližnjica (shortcut).

....



ciklični



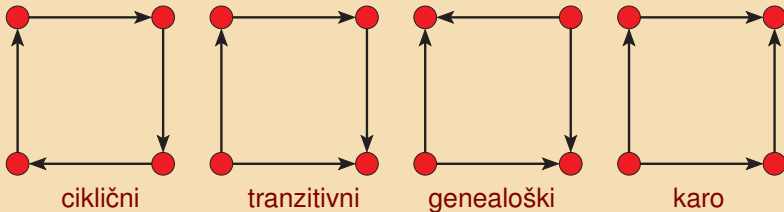
tranzitivni



Za vsako povezavo preštejemo koliko štirikotnikom pripada. Števila se shranijo kot vrednosti na povezavah.

Network / Create New Network / with Ring Counts stored as Line Values / 4-Rings

- **Undirected** – preštejemo število štirikotnikov v neusmerjenem omrežju.
- **Directed** – preštejemo ciklične, 'karo', genealoške, tranzitivne, ali vse štirikotnike v usmerjenem omrežju ali pa kolikokrat neka povezava nastopa kot bližnjica (shortcut).





Pajek

Primer omrežja...

Politična blogosfera, ZDA, 8. februar, 2005

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Omrežje povezav med 1,490 političnimi blogi februarja 2005, ki razpravljajo o volitvah leta 2004 v ZDA. Podatke sta zbrala L. A. Adamic in N. Glance. Usmerjena povezava med dvema blogoma predstavlja sklicevanje v blogu.



Uporabili bomo poenostavljeno - neusmerjeno verzijo omrežja in razbitje avtorjev na bolj liberalne in bolj konzervativne (**Political-blogs.paj**).



Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Nekaj lastnosti tega omrežja:

- število šibko povezanih komponent (WCC) = 268
- velikost največje WCC = 1222 vertices (82%)
- povprečna stopnja (average degree) = 22.436
- premer (diameter) = 8
- povprečna oddaljenost (average distance) = 2.738
- nakopičenost (clustering coefficient - transitivity) = 0.226
- usredinjenost glede na stopnjo = 0.221
- usredinjenost glede na vmesnost = 0.065
- E-I Index = -0.81

Povprečno oddaljenost izračunamo z ukazom:

Network / Create Vector / Distribution of Distances*

Ukaz vrne porazdelitev dolžin najkrajših poti in povprečno dolžino poti v omrežju (izračunano samo na parih dosegljivih točk).



Pajek

...Primer omrežja

Politična blogosfera, ZDA, 8. februar, 2005

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

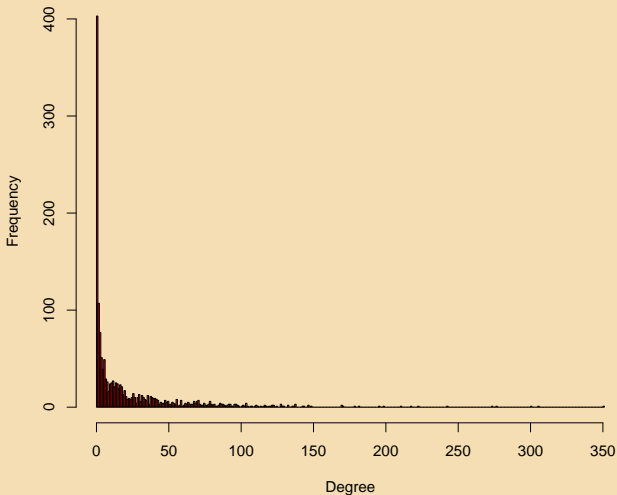
Blogi

Modeli omrežij

Simulacije

Druge omrežja

Degree Distribution of Blogosphere network





Pajek

Standardni modeli omrežij...

Bernoulli in Erdős–Rényi model...

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj
Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druga omrežja

Bernoulli-jev model

Najbolj enostaven model slučajnega omrežja priredi povezavo med dvema točkama neodvisno, z naprej predpisano verjetnostjo. Neodvisnost pomeni, da je verjetnost, da bo par točk povezan s povezavo neodvisna od prisotnosti oz. odsotnosti povezav med drugimi pari.

Network / Create Random Network / Bernoulli/Poisson

Erdős–Rényi-jev model je podoben **Bernoulli-jevemu modelu**, namesto verjetnosti posamezne povezave podamo skupno število povezav. Vedno zgenerira usmerjeno omrežje.

Network / Create Random Network / Total No. of Arcs

Ta dva modela sta v socialnih omrežjih malo verjetna: Posameznikom v tem primeru je vseeno s kom se povežejo - prijatelje izberejo slučajno.



Pajek

...Standardni modeli omrežij...

...Bernoulli in Erdős–Rényi model...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Lastnost Bernoulli-jevih omrežij

Če je povprečna stopnja točk v takih omrežjih večja od 1, bomo z veliko verjetnostjo dobili eno zelo veliko komponento in nekaj manjših komponent.

Če je povprečna stopnja večja od 1.5, lahko pričakujemo, da bo velika komponenta vsebovala vsaj 50% vseh točk (giant component).

Če pa je povprečna stopnja manjša od 1, pa bo dobljeno omrežje vsebovalo več manjših komponent.

Premer v takih omrežjih je relativno majhen, okvirna ocena je

$$Diameter_{expected} = \frac{\ln(n)}{\ln(c)}$$

kjer je n število točk in c povprečna stopnja,



Pričakovana nakopičenost (število tranzitivnih triad v primerjavi z vsemi povezanimi triadami) je v Bernoulli-jevih omrežjih

$$CC_{expected} = \frac{c}{n-1}$$

kar v primeru velikih redkih omrežij konvergira k 0.

Očitna slabost omenjenih dveh modelov je nizka *nakopičenost*.

Potrebujemo modele, ki bodo zgenerirali omrežja z nakopičenostjo, ki se ponavadi pojavlja v socialnih omrežjih, to je med .05 and .50.

Naslednji modeli bodo rešili ta problem.



...Standardni modeli omrežij...

...Bernoulli in Erdős–Rényi model...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Če vzamemo neusmerjeno Bernoulli-jevo omrežje z istim številom točk (1490) in povprečno stopnjo (22.436) kot ju dobimo v omrežju blogosfere, dobimo naslednje pričakovane vrednosti:

$$Diameter_{expected} = \frac{\ln(n)}{\ln(c)} = \frac{\ln(1490n)}{\ln(22.436)} = 2.4$$

$$CC_{expected} = \frac{c}{n-1} = \frac{22.436}{1490-1} = 0.015$$

Premer omrežja blogosfere (8) je približno trikrat večji od pričakovanega premera Bernoullijevega omrežja (2.4).

Nakopičenost v omrežju blogosfere (0.226) je približno deset krat večja od pričakovane nakopičenosti v Ber. omrežju (0.015).

Zgenerirajte nekaj Bernoulli-jevih slučajnih omrežij z enakim številom točk (1490) in povprečno stopnjo (22.436) kot je v omrežju blogosfere in preverite premer in nakopičenost, ki ju dobite. Ali so dobljene vrednosti premera in nakopičenosti blizu pričakovanim? Kolikšno je število in velikost največje šibko povezane komponente?



Pajek

...Standardni modeli omrežij...

...Bernoulli in Erdős–Rényi model

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

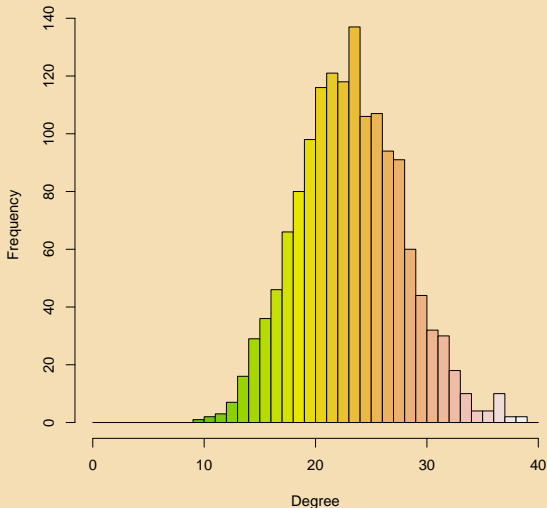
Blogi

Modeli omrežij

Simulacije

Druge omrežja

Degree Distribution of Random Bernoulli network





Pajek

...Standardni modeli omrežij...

Bernoulli-jev model s podanimi stopnjami...

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj
Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Stopnje točk lahko shranimo v razbitje in zgeneriramo slučajno Bernoulli-jevo omrežje, kjer bodo imele točke take stopnje, kot jih zahtevamo v razbitju.

Razbitje določa tudi število točk v omrežju, zato nam tega ni treba podajati.

Partition / Make Network / Random Network

Dobljeno omrežje lahko vsebuje nekaj zank in večkratnih povezav, ki jih ponavadi izločimo:

Network / Create New Network / Transform / Remove / Loops

Network / Create New Network / Transform / Remove / Multiple Lines / Single Line



Pajek

...Standardni modeli omrežij...

...Bernoulli-jev model s podanimi stopnjami...

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Stopnje lahko podamo poljubno ali pa jih izračunamo iz nekega omrežja:

Network / Create Partition / Degree

in zgeneriramo slučajno omrežje z enako porazdelitvijo stopenj:

Partition / Make Network / Random Network

V dobljenem slučajnem omrežju bodo imele vse točke enake stopnje kot točke v začetnem omrežju, toda povezave bodo drugače razporejene.

Zgenerirajte Bernoullijev model z enako porazdelitvijo stopenj kot jo ima omrežje blogosfere. Ne pozabite odstraniti zank in večkratnih povezav. Primerjajte strukturne lastnosti dobljenega omrežja z lastnostmi omrežja blogosfere. Katere lastnosti obeh omrežij so enake, podobne in katere različne?



Pajek

...Standardni modeli omrežij...

...Bernoulli-jev model s podanimi stopnjami

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Redovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

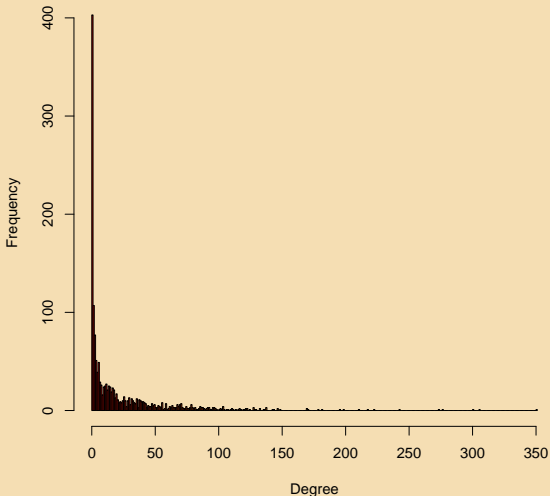
Blogi

Modeli omrežij

Simulacije

Druga omrežja

Degree Distribution of Degree Conditional Random Bernoulli netw.





Pajek

...Standardni modeli omrežij...

Poskus: Mali svetovi

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Leta 1967 je psiholog Stanley Milgram izpeljal naslednji poskus s pismi: Pismo je poslal iz Bostona (Massachusetts), pismo pa naj bi dobila izbrana končna oseba v Omahi (Nebraska). Osebe, ki so bile izbrane za poskus so bile naprošene, naj pismo pošljejo direktno končni osebi (če to osebo poznajo), sicer pa naj pismo s temi navodili pošljejo tistemu svojemu stiku, ki po njegovi/njeni oceni končno osebo pozna. Povprečna dolžina potovanja pisem, ki so dospeli do končne osebe je bila 6 – *six degrees of separation*.

https://en.wikipedia.org/wiki/Small-world_experiment

Povprečna dolžina poti med dvema poljubnima stranema na WWW je 19 klikov.

Omrežja, za katere velja, da je povprečna dolžina najkrajše poti med dvema točkama kratka, se imenujejo *mali svetovi* **Small Worlds**.



Pajek

...Standardni modeli omrežij...

Mali svetovi...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Mali svetovi (Watts-Strogatz model): Posamezniki so povezani s tistimi, ki živijo v bližini, posamezne povezave pa zamenjamo s povezavami do bolj oddaljenih posameznikov - *preklapljanje (rewiring)*.

Originalni model *malih svetov* postavi točke na krožnico in poveže vsako točko s fiksnim številom svojih najbližjih sosedov na krožnici. Če število sosedov preseže 2 se pojavijo tudi trikotniki, ker je vsaka točka povezana s svojim sosedom in s sosedovim sosedom.

Pričakovana nakopičenost (*clustering coefficient - network transitivity*) je določena samo s številom sosedov s katerim je točka povezana na vsaki strani (r):

$$CC_{expected} = \frac{3r - 3}{4r - 2}$$

Če je število sosedov na vsaki strani 1, bo nakopičenost dosegla svoj minimum (0), medtem ko se nakopičenost z večanjem številu sosedov približuje 0.75.



Pajek

...Standardni modeli omrežij...

...Mali svetovi...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Tak pristop pa na žalost še ne zadošča: Veliko omrežje, ki vsebuje samo lokalne povezave ima povprečno oddaljenost veliko večjo kot jo zasledimo v realnih socialnih omrežjih. V takih omrežjih fenomen *Small world* zagotavlja, da celo na celotni populacijo na svetu, pridejo ljudje po poznanstvih od enega do drugega v povprečju v šestih korakih.

Ta problem rešimo tako, da pri majhnem deležu povezav eno krajišče zamenjamo s slučajno izbrano točko (preklapljanje *rewiring*).

S preklapljanjem od 1 do 10 odstotkov povezav že dosežemo efekt *six-degrees* - *povprečna razdalja* med točkami bo nizka.

Nizko število preklopljenih povezav ne spremeni *gostote* in *povprečne stopnje* v omrežju. Samo malenkost lahko spremeni nakopičenost, tako da je tudi visoka nakopičenost *clustering* takih omrežij zagotovljena.



Pajek

...Standardni modeli omrežij...

...Mali svetovi...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

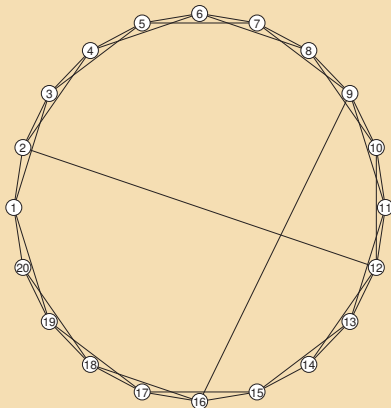
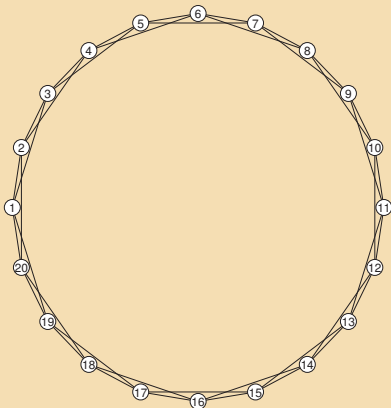
Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja



Omrežje malih svetov zgradimo z:

Network / Create Random Network / Small World



Pajek

...Standardni modeli omrežij...

...Mali svetovi...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Povprečna stopnja v omrežju blogosfere je 22.4, zato to omrežje lahko primerjamo z malim svetom, kjer je vsaka točka povezana z **enajstimi** najbližjimi sosedi na vsako stran ($r = 22.436/2$).

Pričakovana nakopičenost v takem malem svetu je:

$$CC_{expected} = \frac{3r - 3}{4r - 2} = \frac{3 * 11 - 3}{4 * 11 - 2} = 0.71$$

To je skoraj na robu nakopičenosti, ki ko dobimo v omrežjih malih svetov, če ne naredimo preklapljanj (0.75). Dobljena vrednost v omrežju blogosfere (0.226) je precej nižja. Razlog je verjetno v tem, da imajo v malem svetu vse točke približno enake stopnje, v blogosferi pa je porazdelitev stopenj precej asimetrična.

Zgenerirajte nekaj slučajnih malih svetov z enakim številom točk (1490) in **enajstimi** najbližjimi sosedi na vsaki strani, kot jih imamo v omrežju blogosfere. Za verjetnost preklapljanja vzemite vrednosti iz intervala 0.01 do 0.1. Preverite nakopičenosti, ki jih dobite. Ali so dobljene vrednosti nakopičenosti blizu pričakovanim?



Pajek

...Standardni modeli omrežij...

...Mali svetovi

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

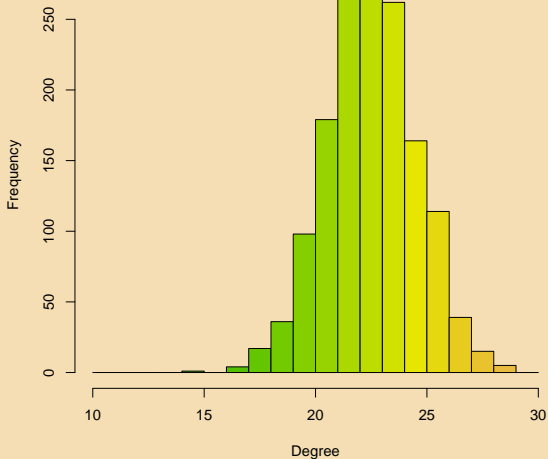
Blogi

Modeli omrežij

Simulacije

Druge omrežja

Degree Distribution of Random Small World network





Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Do zdaj omenjeni modeli še vedno ne posnemajo realnih socialnih omrežij, ki imajo ponavadi nekaj točk z zelo visokimi stopnjami, večina točk pa ima nizke stopnje
Verjetnost, da bo nekdo izbral drugega za prijatelja je proporcionalna stopnji tega posameznika - posamezniki z visokimi stopnjami so bolj zaželeni (*the rich get richer*).
Pri brezlestvičnih omrežjih (scale free) je porazdelitev stopenj močno asimetrična v desno.

Network / Create Random Network / Scale Free

Verjetnost da izberemo točko v za drugi konec povezave je določena z:

$$\Pr(v) = \alpha \frac{\text{indeg}(v)}{|E|} + \beta \frac{\text{outdeg}(v)}{|E|} + \gamma \frac{1}{|V|}$$

kjer je $\alpha + \beta + \gamma = 1$.



...Standardni modeli omrežij...

...Preferential attachment - Scale free (brezlestvična omrežja)...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

V vsakem koraku dodamo omrežju novo točko in nekaj (povprečna stopnja) povezav.

Če je *Adding=Free* izbrano (prosto izbiranje kje bomo dodali povezavo), potem v vsakem koraku dodamo novo povezavo med dvema poljubnima točkama (novo dodana točka ne bo nujno povezana s preostankom omrežja in zato lahko ostane nekaj točk izoliranih).

Če pa je *Adding=Free* neizbrano, potem v vsakem koraku novo točko povežemo z eno od že obstoječih točk.

Za neusmerjena omrežja izberemo samo α in postavimo $\beta = \alpha$.

Če izberemo $\alpha = 0.5$, potem je $\alpha + \beta = 1$ in v neusmerjenem omrežju vedno izbiramo po načelu: večja kot je stopnja točke, bolj verjetno bo ta točka izbrana za drugo krajišče povezave.

Če pa v neusmerjenem omrežju izberemo $\alpha = 0.25$, potem je $\alpha + \beta = 0.5$ in z enako verjetnostjo drugo krajišče povezave izbiramo glede na stopnjo in povsem prosto (kot pri Bernoulli-jevem modelu).



Pajek

...Standardni modeli omrežij...

...Preferential attachment - Scale free (brezlestvična omrežja)...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druga omrežja

Slučajna brezlestvična omrežja, ki jih zgeneriramo na ta način lahko vsebujejo večkratne povezave, kar povzroča težave (npr. nakopičenosti za omrežja z večkratnimi povezavami ne moremo izračunati). Zato večkratne povezave odstranimo:

Network / Create New Network / Transform / Remove / Multiple Lines / Single Line.

Zgenerirajte nekaj brezlestvičnih omrežij (in odstranite večkratne povezave) z enakim številom točk (1490) in enako povprečno stopnjo (22.436) kot velja v omrežju blogosfere.

Za α (verjetnost da bo drugi konec povezave izbran sorazmerno s stopnjami točk) poskusite različne vrednosti iz intervala 0..0.5 (ne pozabite, da se ta verjetnost pri neusmerjenih omrežjih pomnoži z 2).

Izbiro: **Network / Create Random Network / Scale Free / Adding / Free**

enkrat označite, drugič pa ne. Kakšna je razlika (npr. v številu šibko povezanih komponent)?



Pajek

...Standardni modeli omrežij...

...Preferential attachment - Scale free (brezlestvična omrežja)

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

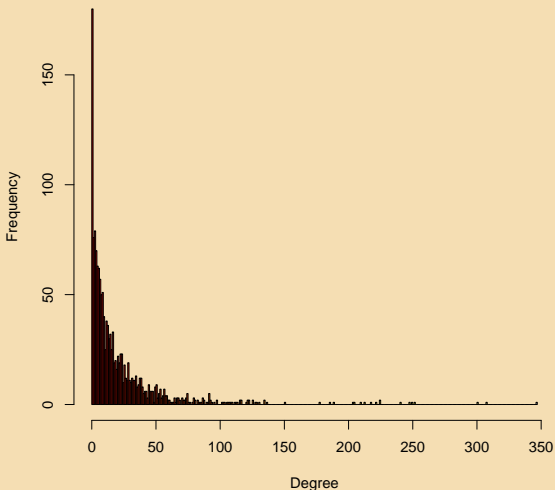
Blogi

Modeli omrežij

Simulacije

Druge omrežja

Degree Distribution of Random Scale Free network





Za naslednja omrežja:

- omrežje blogosfere
- Bernoulli-jevo omrežje ($n = 1490, \overline{deg} = 2.436$)
- omrežje mali svetovi ($n = 1490, r = 11, p_{rew} = 0.1$)
- omrežje mali svetovi ($n = 1490, r = 11, p_{rew} = 0.2$)
- brezlestvično ($n = 1490, \overline{deg} = 2.436, \alpha = 0.5$)
- brezlestvično ($n = 1490, \overline{deg} = 2.436, \alpha = 0.5, Add = Free$)

izračunajte naslednje lastnosti:

- število šibko povezanih komponent
- velikost največje šibko povezane komponente
- premer
- povprečna oddaljenost
- nakopičenost - tranzitivnost
- usredinjenost glede na stopnjo
- usredinjenost glede na vmesnost

Rezultate vpišite v tabelo in jih primerjajte. Za vsako lastnost ugotovite, katero slučajno omrežje je najbolj podobno omrežju blogosfere.



Pajek

...Standardni modeli omrežij

...Pregled

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Za vsako od zgornjih omrežij izračunajte stopnje:

Network / Create Vector / Centrality / Degree

in narišite porazdelitev stopenj v Ru:

Tools / R / Send to R / Current Vector

R: hist(v??, breaks= 0:???)

Podrobnosti: Slika porazdelitve stopenj za male svetove je bila dobljena z uporabo naslednjih parametrov za risanje histogramov v Ru:

```
par (bg="wheat ")  
hist (v1,breaks=10:30,col=terrain.colors(30),  
main="Degree Distribution of Small World network",  
xlab='Degree', right=FALSE)
```



Simulacije Monte Carlo...

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Redovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Če želimo biti prepričani, da so rezultati reprezentativni moramo zgenerirati več (npr. 100) slučajnih omrežij in vzeti povprečje.

Network / Create New Network / Random Network / ...

Zgeneriramo še 99 takih omrežij (da dobimo skupaj 100 omrežij):
Macro / Repeat Last Command (F10) 99

Pogledamo info za prvo omrežje:

Network / Info / General

Pogledamo še za ostalih 99:

Macro / Repeat Last Command (F10) 99

Rezultat je tudi vektor, ki prikaže agregirane rezultate za vseh 100 omrežij, npr. vektor dimenzije 100, kjer posamezna vrednost prikazuje posamezne povprečne stopnje za posamezna omrežja. Z **Vector / Info** pogledamo deskriptivne statistike (*povprečje, mediana, standardni odklon, kvantili*) za izbrano lastnost (povprečna stopnja). Podobo lahko naredimo povzetek tudi za druge lastnosti omrežja (povprečna oddaljenost, premer...).



...Simulacije Monte Carlo

Omrežja

Projektna datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Redovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druge omrežja

Ponovite nalogo iz pregleda, samo da ne zgenerirate samo enega omrežja določenega tipa, ampak jih zgenerirate 100. Namesto rezultatov za 100 omrežij podajte samo agregirane rezultate, ki jih dobite z **Macro / Repeat Last Command**.



Druga velika omrežja

Omrežja

Projektna
datoteka

Unicode

Excel2Pajek

Večrelacijska o.

Časovna omrežja

Besedilna omr.

Aciklična omrežja

Omrežja citiranj

Rodovniki

Nakopičenost

Otoki

Skupnosti

E-I Index

Razbitja

Kratki cikli

Blogi

Modeli omrežij

Simulacije

Druga omrežja

Označena omrežja

Posplošena uravnoveženost

Omrežja sodelovanj (npr. Erdos).

Dvovrstna omrežja - direktna analiza dvovrstnih omrežij -

Important Vertices