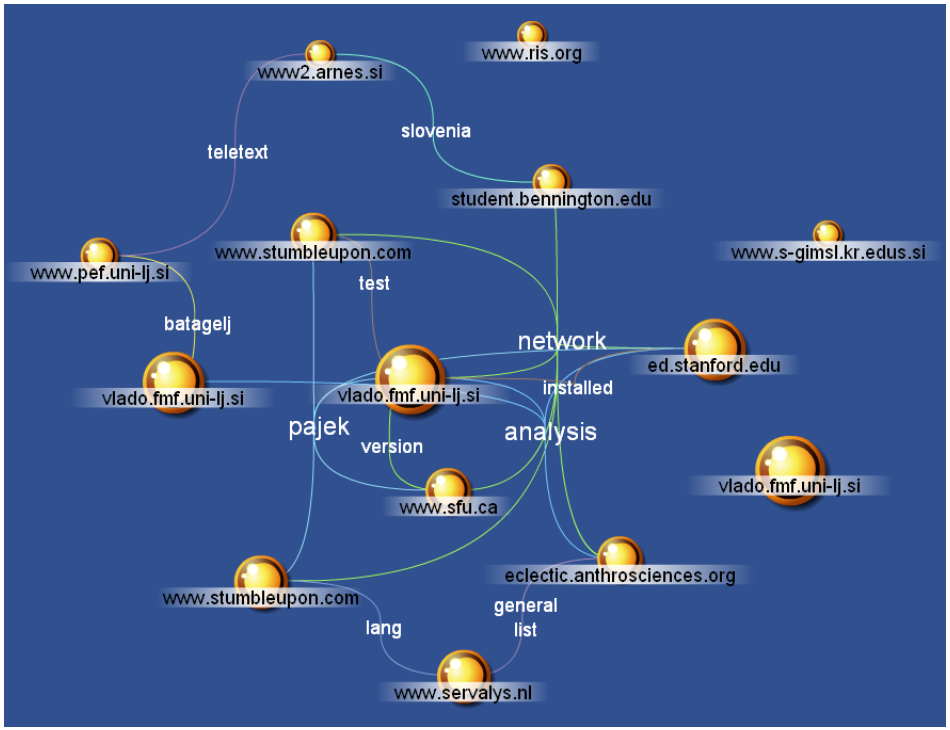


Univerza v Ljubljani
podiplomski študij statistike

Analiza omrežij
2. Omrežja vsepovsod

Vladimir Batagelj

Univerza v Ljubljani

Ljubljana, 29. oktober 2006 / 3. november 2003

Kazalo

Kako do omrežja?

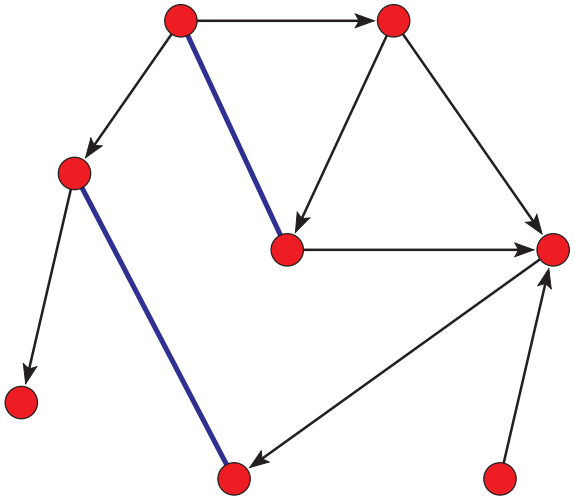
Pri zbiranju podatkov o omrežju $\mathcal{N} = (\mathcal{V}, \mathcal{L}, \mathcal{P}, \mathcal{W})$ se moramo odločiti, kaj je množica enot (točk) – *meje omrežja*, kdaj sta točki povezani – *polnost omrežja* in katere lastnosti točk/povezav bomo upoštevali.

Ta vprašanja so še posebej pereča pri merjenju družbenih omrežij (vprašalniki, pogovori, opazovanja, arhivski zapiski, poskusi, ...). Nekatere 'enote' nočeje sodelovati. Nekateri postopki merjenja, na primer, omejujejo število sosedov ...

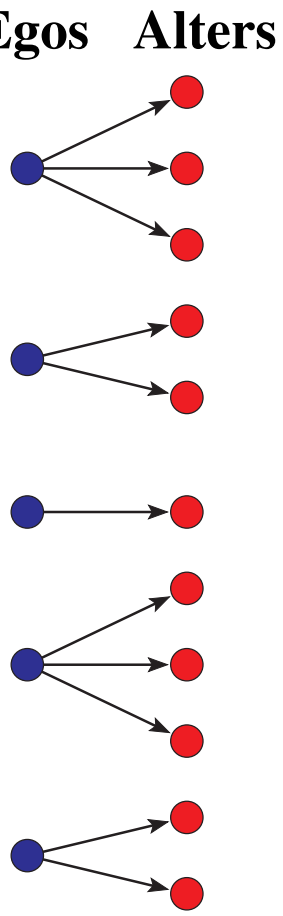
Pri velikih množicah enot si ne moremo privoščiti polnega opisa. Omrežje izmerimo samo za izbrane enote (in njihove sosede). Tako dobljena omrežja imenujemo *osebna omrežja*.

O merjenju socialnih omrežij si lahko preberete v knjigi V. Hlebec in T. Kogovšek (2006).

Cela in osebna omrežja



celo omrežje



osebna omrežja

Uporaba že zbranih omrežij

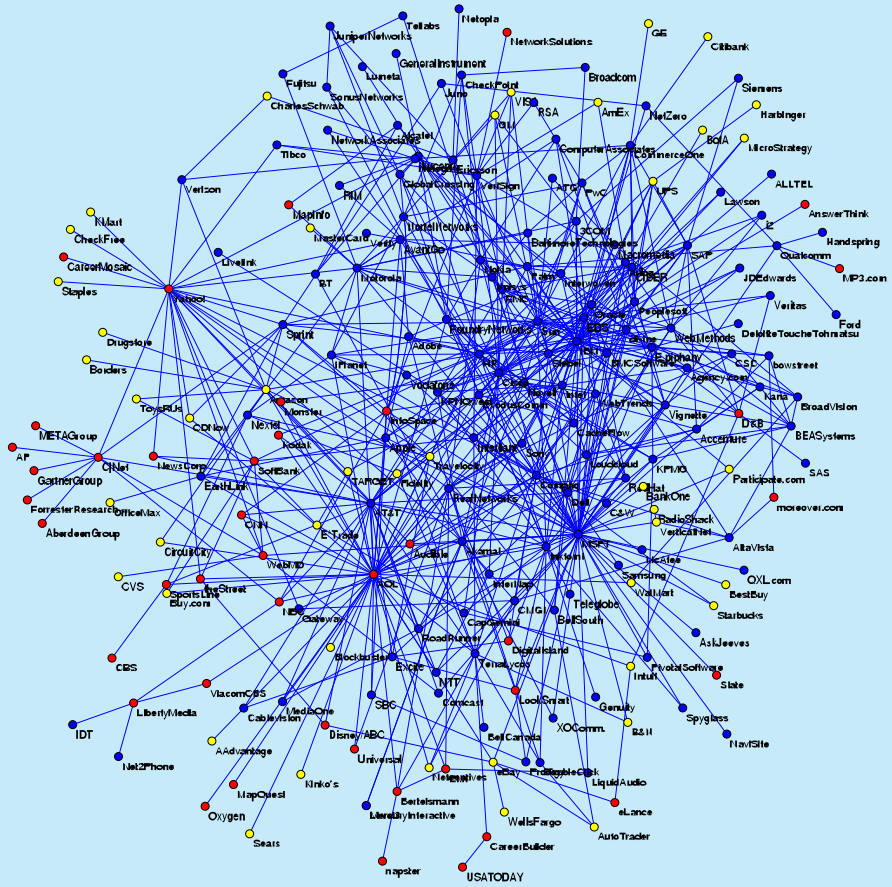
Veliko podatkov o omrežjih pa je že zbranih, pogosto so tudi v računalniški obliki.

Pajek omogoča branje več drugih oblik opisa omrežij: UCINETov datoteke DL, grafi projekta Vega, kemični opisi molekul MDLMOL, MAC in BS ter rodovniki v obliki GEDCOM.

Davis.DAT, C84N24.VGR, MDL, 1CRN.BS, DNA.BS, ADF073.MAC, Bouchard.GED.

Jürgen Pfeffer: txt2pajek.

Krebsova Internetska podjetja



Omrežje sestavljajo izbrana Internetska podjetja v obdobju 1998 do 2001.

$n = 219,\ m = 631.$

rdeča – vsebina,

modra – podpora,

rumena – trgovina.

Podjetji sta povezani, če sta objavili skupni posel ali sodelovanje.

Spletni naslov: **http://www.orgnet.com/netindustry.html**.

Recode, *InfoRapid*.

Rodovniki

Za opis rodovnikov se najpogostoje uporablja oblika zapisa GEDCOM (*GEDCOM standard 5.5*).

Veliko rodovnikov (datoteke `*.GED`) najdemo na spletu – na primer *Roper's GEDCOMs* ali *Isle-of-Man GEDCOMs*. `Family.GED`.

Za pripravo in vzdrževanje rodovnikov je na voljo več programov: prosti *GIM* in tržni *Brothers Keeper* (obstaja tudi slovenska različica – *SRD*).

Iz rodovnikov zbranih v doktoratu: Mahnken, Irmgard. 1960. Dubrovački patricijat u XIV veku. Beograd, Naučno delo. je bila ustvarjeno podatkovje *Ragusa*.

GEDCOM

GEDCOM je dogovor o zapisu rodoslovnih podatkov, ki se uporablja za izmenjavo in združevanje podatkov iz različnih programov uporabljenih za pripravo podatkov.

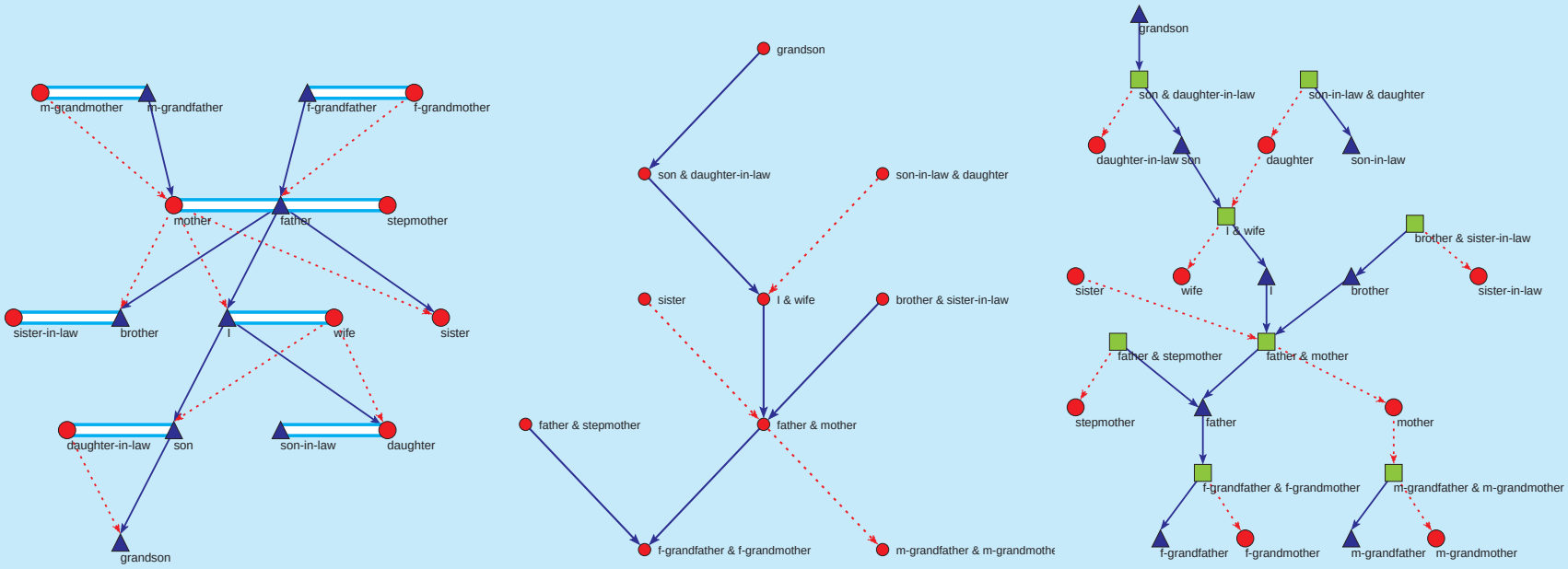
```
0 HEAD
1 FILE ROYALS.GED
...
0 @I58@ INDI
1 NAME Charles Philip Arthur/Windsor/
1 TITL Prince
1 SEX M
1 BIRT
2 DATE 14 NOV 1948
2 PLAC Buckingham Palace, London
1 CHR
2 DATE 15 DEC 1948
2 PLAC Buckingham Palace, Music Room
1 FAMS @F16@
1 FAMC @F14@
...
...
0 @I65@ INDI
1 NAME Diana Frances /Spencer/
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
```

Omrežne predstavitve rodovnikov

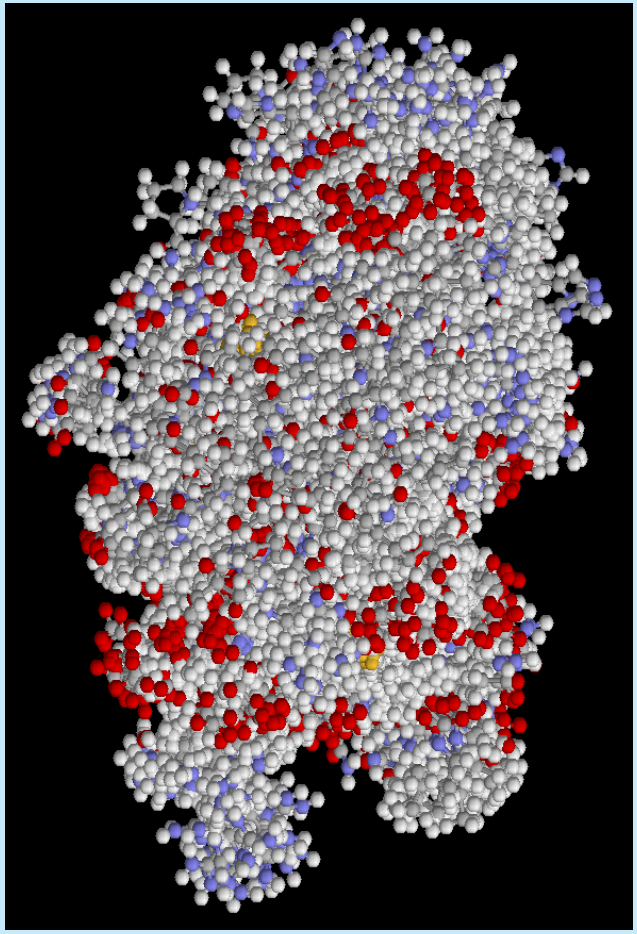
Običajna predstavitev rodovnika z grafom, *Ore*jev graf, ima za točke posameznike in združuje dve relaciji: *sta poročena* (modra neusmerjena) in *ima otroka* (črna usmerjena). V *parnem grafu* so točke poročeni pari ali neporočeni posamezniki, in relaciji *je sin* (polna modra) in *je hči* (črtkana rdeča).

Več o parnih grafih *D. White*.



Orejev graf, parni graf in dvodelni parni graf

Omrežja molekul



virus 1GDY: $n = 39865, m = 40358$

V zbirki Brookhaven Protein Data Bank lahko najdemo veliko velikih organskih molekul (npr. Simian / 1AZ5.pdb) opisanih v obliki PDB.

Molekulo si lahko ogledamo s programom Rasmol (*RasMol*, *program*, *RasWin*) ali *Protein Explorer*.

Molekulo v obliki PDB lahko predelamo v obliko BS, ki jo pozna **`Pajek`**, s programom *BabelWin* + *Babel16*.

GraphML

GraphML – zapis omrežja v obliki XML.

L'Institut de Linguistique et Phonétique Générales et Appliquées (ILPGA), Paris III; Traitement Automatique du Langage (TAL): BaO4 : Des Textes Aux Graphes Plurital

LibXML, `xsltproc` download, XSLT, Xalan, Python, Sxslt.

```
xsltproc GraphML2Pajek.xsl graph.xml > graph.net
java -jar saxon8.jar graph.xml GraphML2Pajek.xsl > graph.net
java org.apache.xalan.xslt.Process -IN p.xml -XSL m.xsl -OUT p.txt
```

XSLT/Zvon

GraphML → Pajek

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Title: 1. D:\vlado\docs\Books\SKRIPTA\Nets\nets\graph.net (12) -->
<!-- Creator: Pajek: http://vlado.fmf.uni-lj.si/pub/networks/pajek/ -->
<!-- CreationDate: 11-03-2006, 17:25:13 -->
<graphml>
  <key id="a1" for="node" attr.name="Label" attr.type="string">
    <desc>Label of the node</desc> <default>NoLabel</default>
  </key>
  <key id="b1" for="edge" attr.name="Weight" attr.type="double">
    <desc>Weight (value) of the edge</desc> <default>1</default>
  </key>
  <graph id="G" edgedefault="directed" parse.nodes="12" parse.edges="23">
    <node id="v1"><data key="a1">a</data></node>
    <node id="v2"><data key="a1">b</data></node>
    <node id="v3"><data key="a1">c</data></node>
    <node id="v4"><data key="a1">d</data></node>
    <node id="v5"><data key="a1">e</data></node>
    <node id="v6"><data key="a1">f</data></node>
    <node id="v7"><data key="a1">g</data></node>
    <node id="v8"><data key="a1">h</data></node>
    <node id="v9"><data key="a1">i</data></node>
    <node id="v10"><data key="a1">j</data></node>
    <node id="v11"><data key="a1">k</data></node>
    <node id="v12"><data key="a1">l</data></node>
    <edge source="v1" target="v2"/> <edge source="v2" target="v1"/>
    <edge source="v1" target="v4"/> <edge source="v1" target="v6"/>
    <edge source="v2" target="v6"/> <edge source="v3" target="v2"/>
    <edge source="v3" target="v3"/> <edge source="v3" target="v7"/>
    <edge source="v3" target="v7"/> <edge source="v5" target="v3"/>
    <edge source="v5" target="v6"/> <edge source="v5" target="v8"/>
    <edge source="v6" target="v11"/> <edge source="v8" target="v4"/>
    <edge source="v10" target="v8"/> <edge source="v12" target="v5"/>
    <edge source="v12" target="v7"/> <edge source="v8" target="v12"/>
    <edge source="v12" target="v8"/>
    <edge directed="false" source="v2" target="v5"/>
    <edge directed="false" source="v3" target="v4"/>
    <edge directed="false" source="v5" target="v7"/>
    <edge directed="false" source="v6" target="v8"/>
  </graph>
</graphml>
```

```
*Vertices 12
1 "a"
2 "b"
3 "c"
4 "d"
5 "e"
6 "f"
7 "g"
8 "h"
9 "i"
10 "j"
11 "k"
12 "l"
*Edges
2 5
3 4
5 7
6 8
*Arcs
1 2
2 1
1 4
1 6
2 6
3 2
3 3
3 7
3 7
5 3
5 6
5 8
6 11
8 4
10 8
12 5
12 7
8 12
12 8
```


GraphML → Pajek

```
<?xml version="1.0" encoding="iso-8859-1"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:output method="text" encoding="iso-8859-1"/>
  <xsl:template match="/">
    <xsl:text>*Vertices </xsl:text>
    <xsl:value-of select="count(graphml/graph/node)"/>
    <xsl:text>&#10;</xsl:text>
    <xsl:apply-templates select="graphml/graph/node"/>
    <xsl:text>*Edges&#10;</xsl:text>
    <xsl:apply-templates select="graphml/graph/edge" mode="edge"/>
    <xsl:text>*Arcs&#10;</xsl:text>
    <xsl:apply-templates select="graphml/graph/edge" mode="arc"/>
  </xsl:template>

  <xsl:template match="edge" mode="arc">
    <xsl:if test="not(./@directed='false')">
      <xsl:value-of select="substring(./@source,2)"/>
      <xsl:text> </xsl:text>
      <xsl:value-of select="substring(./@target,2)"/>
      <xsl:text> </xsl:text>
      <xsl:value-of select="./data"/>
      <xsl:text>&#10;</xsl:text>
    </xsl:if>
  </xsl:template>

  <xsl:template match="edge" mode="edge">
    <xsl:if test="./@directed='false'">
      <xsl:value-of select="substring(./@source,2)"/>
      <xsl:text> </xsl:text>
      <xsl:value-of select="substring(./@target,2)"/>
      <xsl:text> </xsl:text>
      <xsl:value-of select="./data"/>
      <xsl:text>&#10;</xsl:text>
    </xsl:if>
  </xsl:template>

  <xsl:template match="node">
    <xsl:value-of select="substring(./@id,2)"/>
    <xsl:text> "</xsl:text>
    <xsl:value-of select="./data"/>
    <xsl:text>"&#10;</xsl:text>
  </xsl:template>

</xsl:stylesheet>
```


Pristopi k računalniško podprti analizi besedil

R. Popping: Computer-Assisted Text Analysis (2000) razlikuje tri glavne pristope k RPAB: *tematska* AB, *pomenska* AB, in *omrežna* AB.

Pojmi (besede, besedne zveze, izrazi, ...) upoštevani pri AB so zbrani v *slovarju*. Ta je lahko določen vnaprej ali pa se gradi sproti. Osnovni vprašanji pri tem sta *enakovrednost zapisov* – različni zapisi, ki predstavljajo isti pojem; in *dvoumnost zapisov* – isti zapis lahko predstavlja več pojmov. Zato je *kodiranje* – pretvorba surovih podatkov v formalni *opis*, pogosto opravljeno pretežno ročno ali vsaj pod nadzorom uporabnika. Kot *enote* AB ponavadi vzamemo stavke, odstavke, novice, sporočila, ...

Dosedaj sta tematska in pomenska AB temeljili predvsem na statistični analizi kodiranih podatkov.

. . . pristopi k RPAB

Pri tematski AB so enote besedila kodirane s pravokotno matriko *Enote* $\times$ *Pojmi* – pojem p se pojavlja v enoti e. To matriko lahko obravnavamo kot dvovrstno omrežje.

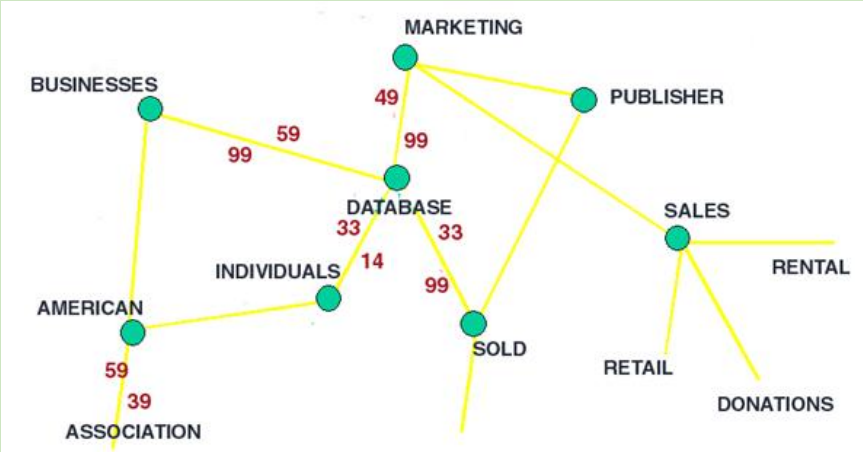
Primeri: M.M. Miller: VBPro, H. Klein: Text Analysis/ TextQuest.

Pri pomenski AB so enote (največkrat enostavni stavki) kodirane po shemi o-P-p (*osebek-Povedek-predmet*) ali njenih izpeljankah.

Primeri: Roberto Franzosi; *KEDS*, *Tabari*.

To kodiranje določa večrelacijsko omrežje na množici točk *Osebki* $\cup$ *Predmeti* s povezavami iz relacij *Povedki*.

Omrežna RPAB



TextAnalyst's 'semantic network'

Tako smo že v omrežni AB.

Primeri:

Carley: Cognitive maps,

J.A. de Ridder: CETA,

Megaputer: TextAnalyst.

Glejte še: W. Evans: Computer Environments for Content Analysis, K.A. Neuendorf: The Content Analysis Guidebook / Online and H.D. White: Publications.

Obstajajo tudi drugi načini, kako ustvariti omrežja iz besedil.

AB – slovarska omrežja

book

A collection of leaves of paper, parchment, vellum, cloth, or other material (written, printed, or blank) fastened together along one edge, with or without a protective case or cover. Also refers to a literary work or one of its volumes. Compare with monograph.

To qualify for the special parcel post rate known in the United States as media rate, a publication must consist of 24 or more pages, at least 22 of which bear printing consisting primarily of reading material or scholarly bibliography, with advertising limited to book announcements. UNESCO defines a book as a nonperiodical literary publication consisting of 49 or more pages, covers excluded. The ANSI standard includes publications of less than 49 pages which have hard covers. ***See also***: art book, board book, children's book, coffee table book, gift book, licensed book, managed book, new book, packaged book, picture book, premium book, professional book, promotional book, rare book, reference book, religious book, and reprint book.

Also, a major division of a longer work (usually of fiction) which is further subdivided into chapters. Usually numbered, such a division may or may not have its own title. Also refers to one of the divisions of the Christian ***Bible***, the first being *Genesis*.

opis pojma **book** v ODLIS

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

Online Dictionary of Library and Information Science *ODLIS*, *Odlis.net* (2909 / 18419).

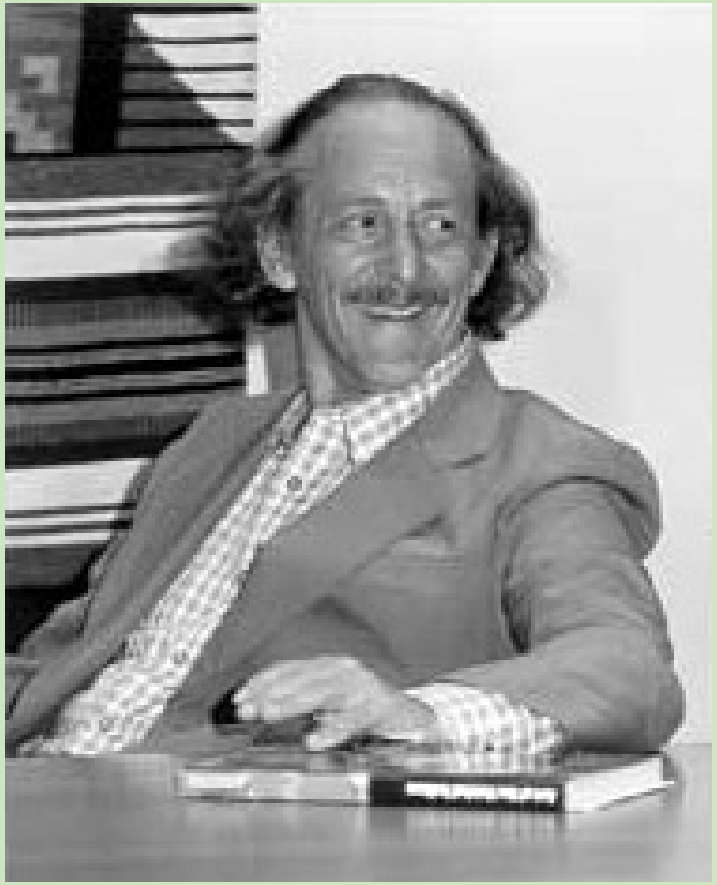
Free On-line Dictionary of Computing *FOLDOC*, *Foldoc2b.net* (133356 / 120238).

Artlex, *Wordnet*, *ConceptNet*, *OpenCyc*.

The Edinburgh Associative Thesaurus (*EAT*) / *net* je bil zbran s spraševanjem (študentov).

NASA Thesaurus. *Paper*.

AB – Omrežja sklicevanj



V *omrežju sklicevanj* so točke razna dela (članki, knjige, poročila, ...) iz izbranega področja; deli sta povezani z usmerjeno povezavo, če se prvo sklicuje na drugo. Omrežja sklicevanj so (skoraj) aciklična.

E. Garfield: HistCite / *Pajek*, *papers*.

Primer zelo velikega omrežja sklicevanj je US Patents / Nber,

$n = 3774768, m = 16522438.$

AB – Omrežja sodelovanj

V omrežjih sodelovanj so enote osebe ali ustanove. Enoti sta povezani, če sta sodelovali pri skupnem delu. Utež povezave je število skupnih del.

Najbolj poznano omrežje sodelovanja je *The Erdős Number Project*, *Erdos.net*. Določitev Erdősevega števila.

Bogat vir podatkov za izgradnjo omrežij sodelovanj so bibliografije v BibTEXu *Nelson H. F. Beebe's Bibliographies Page*.

Taka je na primer B. Jones-ova bibliografija računalniške geometrije *Computational geometry database* (2002), *FTP*, *Geom.net*.

Pri pripravi omrežja sodelovanj iz izvornih podatkov si lahko pomagamo z ustreznimi programi. Nato pa sledi mukotrpno čiščenje - ugotavljanje enot.

Zanimivo podatkovje *The Internet Movie Database*. , Trier DBLP: Digital Bibliography & Library Project, 1124. sredin seminar.

AB – Odnosi med državami

Paul Hensel's International Relations Data Site,

International Conflict and Cooperation Data,

Correlates of War,

Kansas Event Data System *KEDS*,

KEDSi na Pajkovih datotekah,

Prekodirni programi v R-ju.

Pretvorba podatkov KEDS/WEIS v Pajekovo obliko

```
% Recoded by WEISmonths, Sun Nov 28 21:57:00 2004
% from http://www.ku.edu/~keds/data.dir/balk.html
*vertices 325
1 "AFG" [1-*]
2 "AFR" [1-*]
3 "ALB" [1-*]
4 "ALBMED" [1-*]
5 "ALG" [1-*]
 ...
318 "YUGGOV" [1-*]
319 "YUGMAC" [1-*]
320 "YUGMED" [1-*]
321 "YUGMTN" [1-*]
322 "YUGSER" [1-*]
323 "ZAI" [1-*]
324 "ZAM" [1-*]
325 "ZIM" [1-*]
*arcs :0 "*** ABANDONED"
*arcs :10 "YIELD"
*arcs :11 "SURRENDER"
*arcs :12 "RETREAT"
 ...
*arcs :223 "MIL ENGAGEMENT"
*arcs :224 "RIOT"
*arcs :225 "ASSASSINATE TORTURE"
*arcs
224: 314 153 1 [4]              890402  YUG     KSV     224  (RIOT)  RIOT-TORN
212: 314 83 1 [4]               890404  YUG     ETHALB  212  (ARREST PERSON) ALB ETHNIC JAILED IN YUG
224: 3 83 1 [4]                 890407  ALB     ETHALB  224  (RIOT)  RIOTS
123: 83 153 1 [4]               890408  ETHALB  KSV     123  (INVESTIGATE)   PROBING
 ...
42: 105 63 1 [175]              030731  GER     CYP     042  (ENDORSE)       GAVE SUPPORT
212: 295 35 1 [175]             030731  UNWCT   BOSSER  212  (ARREST PERSON) SENTENCED TO PRISON
43: 306 87 1 [175]              030731  VAT     EUR     043  (RALLY) RALLIED
13: 295 35 1 [175]              030731  UNWCT   BOSSER  013  (RETRACT)       CLEARED
121: 295 22 1 [175]             030731  UNWCT   BAL     121  (CRITICIZE)     CHARGES
122: 246 295 1 [175]            030731  SER     UNWCT   122  (DENIGRATE)     TESTIFIED
121: 35 295 1 [175]             030731  BOSSER  UNWCT   121  (CRITICIZE)     ACCUSED
```


. . . Program v R-ju

Za pretvorbo podatkov KEDS/WEIS smo uporabili kratke programe v R-ju, kot je naslednji:

```
# WEISmonths
# recoding of WEIS files into Pajek's multirelational temporal files
# granularity is 1 month
# ------------------------------------------------------------------
# Vladimir Batagelj, 28. November 2004
# ------------------------------------------------------------------
# Usage:
#   WEISmonths(WEIS_file,Pajek_file)
# Examples:
#   WEISmonths('Balkan.dat','BalkanMonths.net')
# ------------------------------------------------------------------
# http://www.ku.edu/~keds/data.html
# ------------------------------------------------------------------

WEISmonths <- function(fdat,fnet){

  get.codes <- function(line){
    nlin <<- nlin + 1;
    z <- unlist(strsplit(line,"\t")); z <- z[z != ""]
    if (length(z)>4) {
      t <- as.numeric(z[1]); if (t < 500000) t <- t + 1000000
      if (t<t0) t0 <<- t; u <- z[2]; v <- z[3]; r <- z[4]
      if (is.na(as.numeric(r))) cat(nlin,'NA rel-code',r,'\n')
      h <- z[5]; h <- substr(h,2,nchar(h)-1)
      if (nchar(h) == 0) h <- '*** missing description'
      if (!exists(u,env=act,inherits=FALSE)){
        nver <<- nver + 1; assign(u,nver,env=act) }
      if (!exists(v,env=act,inherits=FALSE)){
        nver <<- nver + 1; assign(v,nver,env=act) }
      if (!exists(r,env=rel,inherits=FALSE)) assign(r,h,env=rel)
    }
  }
```


. . . Program v R-ju

```
recode <- function(line){
  nlin <<- nlin + 1;
  z <- unlist(strsplit(line,"\t")); z <- z[z != ""]
  if (length(z)>4) {
    t <- as.numeric(z[1]); if (t < 500000) t <- t + 1000000
    cat(as.numeric(z[4]),': ',get(z[2],env=act,inherits=FALSE),
      ' ',get(z[3],env=act,inherits=FALSE),' 1 [',
      12*(1900 + t %/% 10000) + (t %% 10000) %/% 100 - t0,
      ']\n',sep='',file=net)
  }
}

cat('WEISmonths: WEIS -> Pajek\n')
ts <- strsplit(as.character(Sys.time())," ")[[1]][2]
act <- new.env(TRUE,NULL); rel <- new.env(TRUE,NULL)
dat <- file(fdat,"r"); net <- file(fnet,"w")
lst <- file('WEIS.lst',"w"); dni <- 0
nver <- 0; nlin <- 0; t0 <- 9999999
lines <- readLines(dat); close(dat)
sapply(lines,get.codes)
a <- sort(ls(envir=act)); n <- length(a)
cat(paste('% Recoded by WEISmonths,',date()),"\n",file=net)
cat("% from http://www.ku.edu/~keds/data.html\n",file=net)
cat("*vertices",n,"\n",file=net)
for(i in 1:n){ assign(a[i],i,env=act);
  cat(i,' "',a[i],'" [1-*]\n',sep='',file=net) }
b <- sort(ls(envir=rel)); m <- length(b)
for(i in 1:m){ assign(a[i],i,env=act);
cat("*arcs :",as.numeric(b[i]),' "',
get(b[i],env=rel,inherits=FALSE),'"\n',sep='',file=net) }
t0 <- 12*(1900 + t0 %/% 10000)
slice <- 0
cat("*arcs\n",file=net); nlin <- 0
sapply(lines,recode)
cat(' ',nlin,'lines processed\n'); close(net)
te <- strsplit(as.character(Sys.time())," ")[[1]][2]
cat('  start:',ts,'  finish:',te,'\n')
}

WEISmonths('Balkan.dat','BalkanMonthsR.net')
```

Opomba: Slovarjem (dictionary) se v R-ju reče *environment*.

Omrežja sosednjih točk

Recimo, da imamo na množici enot $\mathcal{U}$ dano mero različnosti $d(u, v)$. Glede na d lahko vpeljemo dve vrsti omrežij:

k-najbližjih sosedov: $\mathcal{N}(k) = (\mathcal{U}, \mathcal{A}, d)$

$$(u, v) \in \mathcal{A} \Leftrightarrow v \text{ je med } k \text{ najbližjimi sosedi točke } u$$

Za utež povezave $a(u, v) \in \mathcal{A}$ postavimo $w(a) = d(u, v)$.

Omrežje r-okolic: $\mathcal{N}(r) = (\mathcal{U}, \mathcal{E}, d)$

$$(u : v) \in \mathcal{E} \Leftrightarrow d(u, v) \leq r$$

Za utež povezave $e(u : v) \in \mathcal{E}$ postavimo $w(e) = d(u, v)$.

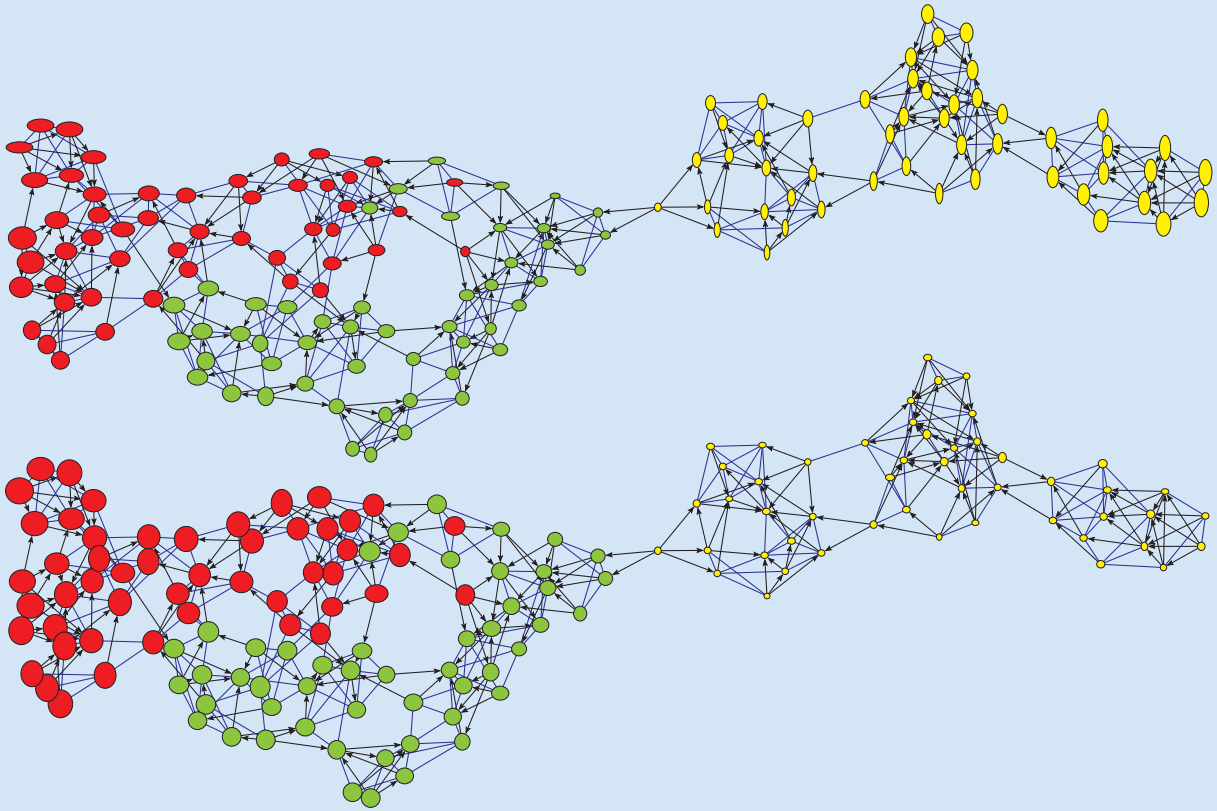
Ta omrežja so povezava z običajno analizo podatkov. Še vedno odprto vprašanje: učinkoviti postopki za določanje teh omrežij.

Multivariatne pajčevine – določitev omrežij sosedov v R-ju.

Najbližjih k sosedov v R-ju

```
k.neighbor2Net <-
# stores network of first  k  neighbors for
# dissimilarity matrix  d  to file  fnet  in Pajek format.
function(fnet,d,k){
  net <- file(fnet,"w")
  n <- nrow(d); rn <- rownames(d)
  cat("*vertices",n,"\n",file=net)
  for (i in 1:n) cat(i," \"",rn[i],"\"\n",sep="",file=net)
  cat("*arcs\n",file=net)
  for (i in 1:n) for (j in order(d[i,])[1:k+1]) {
    cat(i,j,d[i,j],"\n",file=net)
  }
  close(net)
}
stand <-
# standardizes vector  x .
function(x){
  s <- sd(x)
  if (s > 0) (x - mean(x))/s  else  x - x
}
data(iris)
ir <- cbind(stand(iris[,1]),stand(iris[,2]),stand(iris[,3]),
  stand(iris[,4]))
k.neighbor2Net("iris5.net",as.matrix(dist(ir)),5)
```


Fisherjeve perunike z dodatnimi podatki



`Draw/Draw-Partition-2Vectors`

Velikost točk je sorazmerna normaliziranima (Sepal.Length, Sepal.Width) oziroma (Petal.Length, Petal.Width). Barva točk je določena z razbitjem iz podatkov. *Podatki*.

r-sosedi v R-ju

```
r.neighbor2Net <-
# stores network of  r-neighbors (d(v,u) <= r) for
# dissimilarity matrix  d  to file  fnet  in Pajek format.
function(fnet,d,r){
  net <- file(fnet,"w")
  n <- nrow(d); rn <- rownames(d)
  cat("*vertices",n,"\n",file=net)
  for (i in 1:n) cat(i," \"",rn[i],"\"\n",sep="",file=net)
  cat("*edges\n",file=net)
  for (i in 1:n){
    s <- order(d[i,]); j <- 1
    while (d[i,s[j]] <= r) {
      k <- s[j]; if (i < k) cat(i,k,d[i,k],"\n",file=net)
      j <- j+1
    }
  }
  close(net)
}
```

Transformacije

Besedni graf – točke so besede; besedi sta povezani, če lahko eno dobimo iz druge s spremembo ene črke. DIC28, *Članek*.

Omrežja iz besedil – besedi sta povezani, če se v besedilu pojavita dovolj blizu skupaj. Utež povezave je število takih ponovitev. Primer *CRA*.

Grafi iger – točke so stanja v igri, povezave pa dovoljeni prehodi med njimi.

Internetska omrežja

KartOO network

Internet Mapping Project.

Sosednost na spletu (Najdi.si), Grobelnik.

E-mail, blogi, strežniški dnevniki in druge storitve.

Povezave med stranmi na spletu: KartOO, TouchGraph.

Cybergeography, *CAIDA*.

Orodja za pridobivanje podatkov s spleta: *MedlineR*, *SocSciBot*.

Rabutanje

Za pridobivanje izbranih podatkov iz (večih) spletnih strani lahko napišemo posebne programe *web wrappers*. Ti iz posamezne strani izluščijo iskane podatke in jih shranijo – pogosto v obliki XML.

Primeri v R-ju: Naslovi patentov, Knjige z Amazon.

Ker je pisanje teh programov za običajnega uporabnika prezapleteno, je bilo razvitih več *orodij*, ki jih ustvarijo iz uporabnikovih opisov/zahtev (članek / seznam / LAPIS).

Med prostimi orodji sta zanimiva še XWRAP (opis / stran) in TSIMMIS (opis / stran).

Med tržnimi orodji trenutno prevladuje lixto.

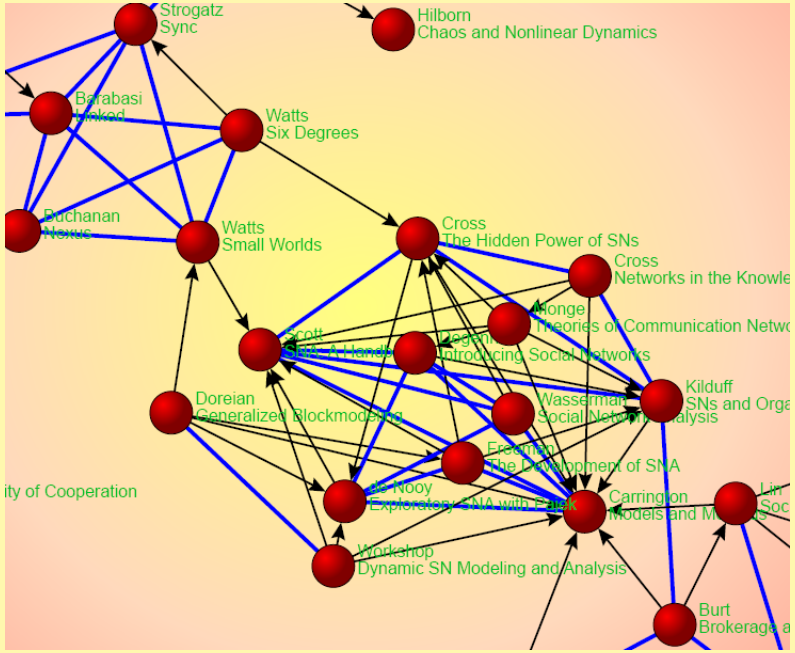
Še nekaj naslovov 1, 2, 3.

Omrežje z Amazona

```
amazon <- function(fvtx,flnk,ftit,maxver){
# ustvari omrezje knjig s spletisca Amazon
#  amazon('v.txt','a.txt','t.txt',10)
# Vladimir Batagelj, 20-21. nov. 2004 / 10. nov. 2006
  opis <- function(line){
    i <- regexpr('\">',line); l <- i[1]+attr(i,"match.length")[1]
    j <- regexpr('</a>',line); r <- j[1]-1; substr(line,l,r)
  }
  vid <- new.env(hash=TRUE,parent=emptyenv())
  vtx <- file(fvtx,"w");  cat('*vertices\n', file=vtx)
  tit <- file(ftit,"w");  cat('*vertices\n', file=tit)
  lnk <- file(flnk,"w");  cat('*arcs\n',file=lnk)
  url1 <- 'http://www.amazon.com/exec/obidos/tg/detail/-/'
  url2 <- '?v=glance';
  book <- '0521840856'
  auth <- "Patrick Doreian"
  titl <- "Generalized Blockmodeling"
  narc <- 0;  nver <- 1
  page <- paste(url1,book,url2,sep='')
  cat(nver, ' "', book, '" URL "',page,'"\n', sep='', file=vtx)
  cat(nver, ' "', auth, ':\\n',titl, '"\n', sep='', file=tit)
  assign(book,nver,env=vid)
  cat('new vertex ',nver,' - ',book,'\n')
  books <- c(book)
```


```
  while (length(books)>0){
    bk <- books[1]; books <- books[-1]
    vini <- get(bk,env=vid); cat(vini,'\n')
    page <- paste(url1,bk,url2,sep='')
    stran <- readLines(con<-url(page)); close(con)
    i <- grep("Customers who bought",stran,ignore.case=TRUE)[1]
    if (is.na(i)) break
    j <- grep("Explore Similar Items",stran,ignore.case=TRUE)[1]
    izrez <- stran[i:j]; izrez <- izrez[-which(izrez=="")]
    izrez <- izrez[-which(izrez=="   ")]
    ik <- regexpr("/dp/",izrez); ii <- ik+attr(ik,"match.length")
    for (k in 1:length(ii)) {
      j <- ii[k];
      if (j > 0) {
        bk <- substr(izrez[k],j,j+9); cat('test',k,bk,'\n')
        if (exists(bk,env=vid,inherits=FALSE)){
          vter <- get(bk,env=vid,inherits=FALSE)
        } else {
          nver <- nver + 1; vter <- nver; line <- izrez[k]
          assign(bk,nver,env=vid)
          if (nver <= maxver) {books <- append(books,bk)}
          cat(nver,' "',bk,'" URL "',url1,bk,url2,'"\n',sep='',file=vtx)
          cat('new vertex ',nver,' - ',bk,'\n');
          t <- opis(line); line <- izrez[k+1]
          if (substr(line,1,2)=='by') {a <- substr(line,4,100)}
            else { a <- 'UNKNOWN' }
          cat(nver, ' "', a, ':\\n', t, '"\n', sep='', file=tit)
        }
        narc <- narc + 1; cat(vini,vter,'\n', file=lnk)
      }
    }
    flush.console()
  }
  close(lnk); close(vtx); cat('Amazon - END\n')
}
```


Omrežje z Amazona – knjige o analizi omrežij



Knjige o analizi omrežij z Amazona, 10. november 2006; začetna točka P. Doreian &: Generalized Blockmodeling.

Slika SVG. Datoteke/ZIP.

Program v R-ju je le zasnova. Možne izpopolnitve:¡br¿ seznam začetnih točk;¡br¿ nadaljevanje po prekinitvi zveze; . . .

Slučajna omrežja

Omrežja lahko tudi sami ustvarimo z nekim slučajnim postopkom. Ozadja teh postopkov bomo spoznali kasneje.

Vgrajeni so v Pajka (`Net / Random network`), lahko pa jih tudi sami zapišemo z razmeroma kratkimi postopki v R-ju.

Na voljo je tudi program `GeneoRnd` za ustvarjanje slučajnih rodovnikov.

Slučajni neusmerjeni Erdős-Rényijev graf

```
dice <- function(n=6){return(1+trunc(n*runif(1,0,1)))}

ErdosRenyiNet <-
# generates a random undirected graph of Erdos-Renyi type
# with n vertices and m edges, and stores it on the file
# fnet  in Pajek's format.
# Example:
#   ErdosRenyiNet('testER.net',100,175)
# ------------------------------------------------
# by Vladimir Batagelj, R version: Ljubljana, 20. Dec 2004
# based on ALG.2 from: V. Batagelj, U. Brandes:
#   Efficient generation of large random networks
function(fnet,n,m){
  net <- file(fnet,"w"); cat("*vertices",n,"\n",file=net)
  cat('% random Erdos-Renyi undirected graph G(n,m) / m = ',
    m,'\n',file=net)
#  for (i in 1:n) cat(i," \"v",i,"\"\n",sep="",file=net)
  cat("*edges\n",file=net); L <- new.env(TRUE,NULL)
  for (i in 1:m){
    repeat { u <- dice(n); v <- dice(n)
      if (u!=v) {
        edge <- if (u<v) paste(u,v) else paste(v,u)
        if (!exists(edge,env=L,inherits=FALSE)) break }
    }
    assign(edge,0,env=L); cat(edge,'\n',file=net)
  }
  close(net)
}
```

Dodatni viri

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

2. V. Hlebec, T. Kogovšek: *Merjenje socialnih omrežij*. Scripta, Študentska založba, Ljubljana, 2006.