



Photo: Vladimir Batagelj: *Dragon*

Integrating Quantitative Narrative Analysis, Network Analysis and R

Vladimir Batagelj
Andrej Mrvar

University of Ljubljana

Roberto Franzosi

Emory University

ECPR Summer School, July 19 – August 4, 2007
Faculty of Social Sciences, University of Ljubljana

Introduction

Combination of

- Course 11: Quantitative Narrative Analysis (R. Franzosi)
- Course 13: Network Analysis (V. Batagelj, A. Mrvar, N. Kejžar)
- Crash Course 4: R (A. Blejec)

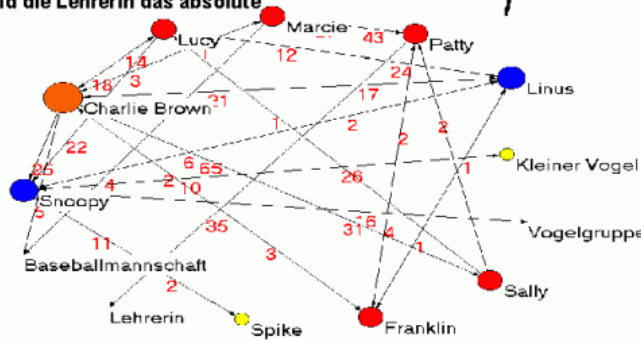
These slides

Data and programs

Networks

Aber damit geben sich Soziologen noch lange nicht zufrieden. Sie wollen zum Beispiel noch wissen, wer ist hier wichtig in diesem Netzwerk?

Dabei gibt es in der Art des Wichtig-Seins Unterschiede. Beispielsweise kann man fragen, wer ist am besten eingebunden, wer hat die meisten Beziehungen? Hier wurde gemessen, wer die meisten Beziehungen zu anderen aufgenommen hat, und da ist Charlie Spitzenreiter, während die Vogelgruppe, die Baseballmannschaft und die Lehrerin das absolute Schlußlicht bilden



Alexandra Schuler/ Marion Laging-Glaser:

Analyse von Snoopy Comics

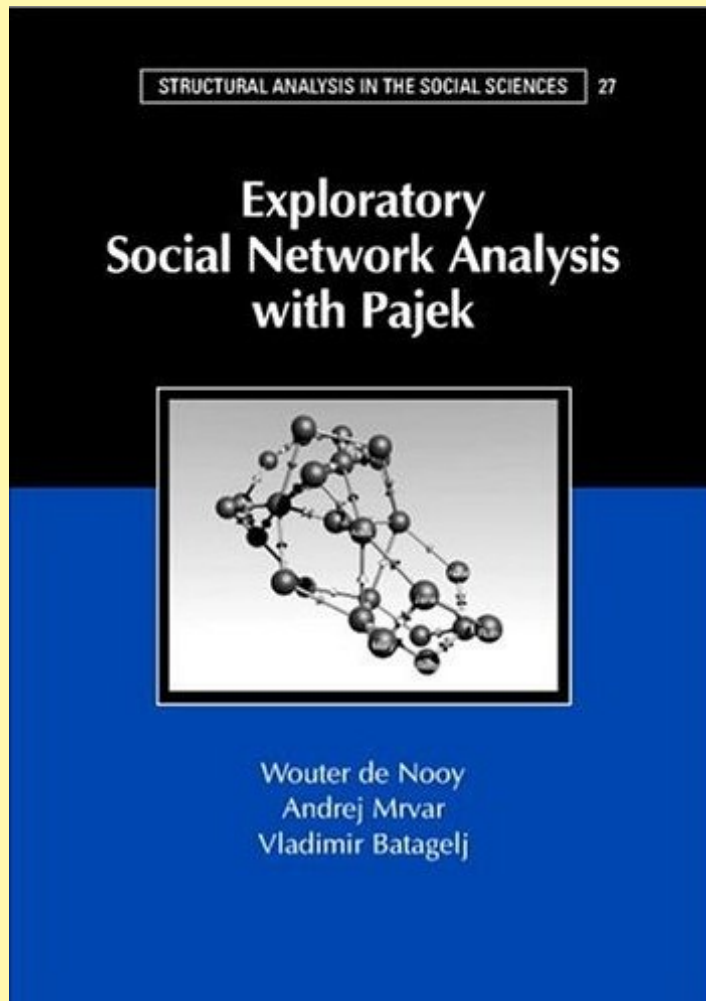
A *network* is based on two sets – set of *vertices* (nodes), that represent the selected *units*, and set of *lines* (links), that represent *ties* between units. They determine a *graph*. A line can be *directed* – an *arc*, or *undirected* – an *edge*.

Additional data about vertices or lines can be known – their *properties* (attributes). For example: name/label, type, value, ...

Network = Graph + Data

The data can be measured or computed.

ESNA Pajek



For network analysis of the data we shall use the program **Pajek** since it supports the temporal multirelational networks that are needed for description of the data.

An introduction to social network analysis with **Pajek** is available in the book **ESNA** (de Nooy, Mrvar, Batagelj 2005).

Pajek – program for analysis and visualization of large networks is freely available, for noncommercial use, at its web site.

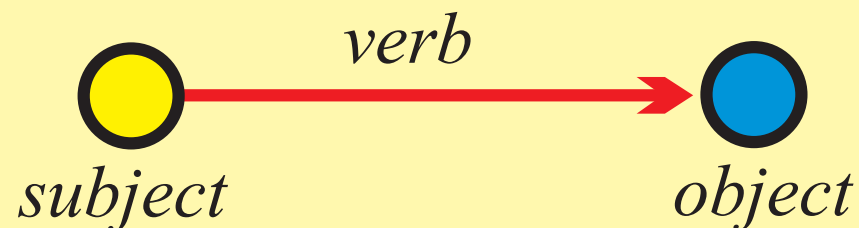
<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>

... approaches to CaTA

In thematic TA the units are coded as rectangular matrix *Text units* \times *Concepts* which can be considered as a two-mode network.

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

In semantic TA the units (often clauses) are encoded according to the S-V-O (*Subject-Verb-Object*) model or its improvements.



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

This coding can be directly considered as network with *Subjects* \cup *Objects* as vertices and lines labeled with *Verbs*.

See also **RDF** triples in **semantic web**.

PC-ACE to Pajek

```

Date;Subject;Action;Object
19.jul.19;?;fanno retate;individui
19.jul.19;?;fanno retate;socialisti
19.jul.19;?;fanno retate;anarchici
8.sep.19;?;fanno eccidi;?
8.feb.20;?;fa sassaiuola;lavoratori
1.jan.20;?;ferisce;soldati
1.jan.20;?;uccide;cittadini
13.jan.20;?;ferisce;carabinieri
20.jan.20;?;ferisce;carabinieri
5.mar.20;?;ferisce;lavoratori
21.feb.20;?;ferisce;militari
21.feb.20;?;ferisce;operai
.:.
22.dec.20;fascisti;picchiano;socialisti
12.jan.21;abitanti;sparano;individui
19.avg.19;agenti;caricano;lavoratori
20.avg.19;agenti;percuotono;donne
9.maj.19;agenti;si lanciano;socialisti
9.maj.19;agenti;malmenano;socialisti
9.maj.19;agenti;percuotono;socialisti
.:.
9.nov.20;ufficiali;picchiano;individui
9.nov.20;ufficiali;bastonano;individui
9.nov.20;ufficiali;sparano;folla
4.dec.19;ufficiali;sparano;operai
11.apr.20;viaggiatori;maltrattano;personale

```

PC-ACE to Pajek

```

% Recoded by PCACEmonths, Tue Jul 31 11:00:30 2007
*vertices 310
1 "?" [1-*]
2 "abitanti" [1-*]
3 "abitanti-passanti" [1-*]
4 "adulti" [1-*]
5 "agenti" [1-*]
6 "agenti in borghese" [1-*]
...
307 "vedove di guerra" [1-*]
308 "viaggiatori" [1-*]
309 "viandanti" [1-*]
310 "villeggianti" [1-*]
*arcs :128 "aggredisce"
*arcs :14 "aggrediscono"
*arcs :57 "ammazzano"
...
*arcs :79 "uccidere"
*arcs :8 "uccidono"
*arcs

1: 1 113 1 [7]
1: 1 293 1 [7]
1: 1 16 1 [7]
2: 1 1 1 [9]
3: 1 226 1 [14]
4: 1 295 1 [13]
5: 1 48 1 [13]
4: 1 42 1 [13]
4: 1 42 1 [13]
4: 1 226 1 [15]
4: 1 239 1 [14]
4: 1 249 1 [14]
...
15: 85 293 1 [24]
11: 2 113 1 [25]
16: 5 226 1 [8]
17: 5 74 1 [8]
18: 5 293 1 [5]
19: 5 293 1 [5]
17: 5 293 1 [5]
...
15: 303 113 1 [23]
12: 303 113 1 [23]
11: 303 90 1 [23]
11: 303 249 1 [12]
24: 308 257 1 [16]

```

PC-ACE to Pajek in R

PCACEdays and PCACEmonths transform

date;actor1;relation;actor2

into *temporal multirelational* network

relNum : act1Num act2Num 1 [time]

```
# PCACEdays
# recoding of PC-ACE files into Pajek's multirelational temporal files
# granularity is 1 day
# -----
# Vladimir Batagelj, 30. July 2007
# based on KEDSdays
# Vladimir Batagelj, 27. November 2004
# -----
# Usage:
#   PCACEdays(PCACE_file,Pajek_file)
# Examples:
#   PCACEdays('VIOLENCE.csv','violenceD.net')
# -----
# http://www.pc-ace.com/
# http://vlado.fmf.uni-lj.si/pub/networks/pajek/
# -----
```


... PC-ACE to Pajek in R

```

PCACEdays <- function(fdat, fnet) {
  get.codes <- function(line) {
    nlin <- nlin + 1;
    z <- unlist(strsplit(line, ";"))
    if (length(z)==4) {
      d <- unlist(strsplit(z[1], "\\."))
      if (length(d)==3) {
        t <- as.numeric(d[1])+100*as.numeric(monthN[d[2]])+
          10000*as.numeric(d[3])
        if (t<t0) t0 <- t
        u <- z[2]; v <- z[4]; r <- z[3]
        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)) {
          nrel <- nrel + 1; assign(r,nrel,env=rel) }
      } else {
        cat("*** error in line ",nlin," : ",line,"\n")
      }
    } else {
      cat("*** error in line ",nlin," : ",line,"\n")
    }
  }
}

```

... PC-ACE to Pajek in R

```

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

Days <- function(d, m, y){
  p <- (0 == y %% 4) & (0 != y %% 100) | (0 == y %% 400)
  z <- y - 1; n <- z %% 4 - z %% 100 + z %% 400
  n <- n + 365*y + 30*(m - 1) + d + e[m]
  if (p & (m > 2)) {n <- n + 1}
  n
}

```

... PC-ACE to Pajek in R

```

cat('PCACEdays: PC-ACE -> Pajek\n')
ts <- strsplit(as.character(Sys.time()), " ")[[1]][2]
e <- c(1,2,0,1,1,2,2,3,4,4,5,5)
monthN <- seq(12)
names(monthN) <- list("jan", "feb", "mar", "apr", "maj", "jun", "jul", "avg",
  "sep", "okt", "nov", "dec")
act <- new.env(TRUE, parent = baseenv())
rel <- new.env(TRUE, parent = baseenv())
dat <- file(fdat, "r"); net <- file(fnet, "w")
# lst <- file('keds.dbg', "w"); dni <- 0
nver <- 0; nrel <- 0; t0 <- 9999999; nlin <- 0
line <- readLines(dat, n=1)
lines <- readLines(dat); close(dat)
sapply(lines, get.codes)
a <- sort(ls(envir=act)); n <- length(a)
cat(paste('% Recoded by PCACEdays, ', date()), "\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 :", get(b[i], env=rel, inherits=FALSE), ' ',
  b[i], '"\n', sep='', file=net) }
t0 <- Days(t0 %% 100, (t0 %% 10000) %% 100, 1900 + t0 %% 10000) - 1
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')
}

#PCACEdays('VIOLENCE.csv', 'violenceD.net')

```

Notes

For the analysis the granularity by months was selected. In **Pajek** the time points (months) are numbered with numbers from 1 to 48 according to the following table

	1919	1920	1921	1922
January	1	13	25	37
February	2	14	26	38
March	3	15	27	39
April	4	16	28	40
May	5	17	29	41
June	6	18	30	42
July	7	19	31	43
August	8	20	32	44
September	9	21	33	45
October	10	22	34	46
November	11	23	35	47
December	12	24	36	48

Network analysis

```
File/Network/Read [ViolenceM.net]
```

```
Info/Network/General
```

```
Info/Network/Multiple Relations
```

The network has 310 vertices (actors) and 149 relations (kinds of violence) – see next slide for the most frequent.

For this analysis we decided to treat the network as a single relation network:

```
Net/Transform/Multiple Relations/Change Relation Number;Label  
[1-*] [1] [Violence] [Yes]
```

This network has 4802 arcs (99 loops).

Network analysis / Kinds of violence

rel	freq	label			
7	1001	feriscono	30	17	attaccano
11	725	sparano	24	16	maltrattano
12	606	bastonano	90	15	fanno spedizioni
14	319	aggrediscono	53	14	danno la caccia
15	218	picchiano	48	11	fanno fuoco
25	215	fanno violenze	55	10	si gettano
8	145	uccidono	5	10	uccide
23	115	fanno violenza	120	10	picchia
17	108	percuotono	21	10	fanno carica
13	94	terrorizzano	26	9	calpestando
41	86	commettono violenze	51	9	fanno collutazione
40	81	inseguono	92	9	legnano
6	76	lanciano	33	9	assassinano
4	60	ferisce	52	9	sbandano
62	57	obbligano	83	8	costringere
32	49	assalgono	28	8	combattono
44	47	colpiscono	79	7	uccidere
78	42	si scontrano	20	6	molestano
19	38	malmenano	46	6	seviziano
63	38	fanno spedizione	104	6	fa violenza
66	35	costringono	71	6	lancia
16	33	caricano	86	6	pugnalano
18	30	si lanciano	43	5	si precipitano
27	30	fanno rappresaglie	45	5	si azzuffano
29	27	assaltano	56	5	si avventano
10	26	spara	99	5	fa violenze
9	22	fanno conflitto	1	5	fanno retate
91	18	schiaffeggiano	111	5	fanno violazioni
93	18	fanno rappresaglia	70	5	carica
22	18	si scagliano			

Network analysis / Classes of actors

Also among the 310 actors there are several similar/equivalent names. Roberto prepared a partition `similar.clu` that reduces their number to 29 classes of actors – see next slide, left column. The file `similarOrg2.clu` contains a 'working' version of the partition file – see next slide, right column.

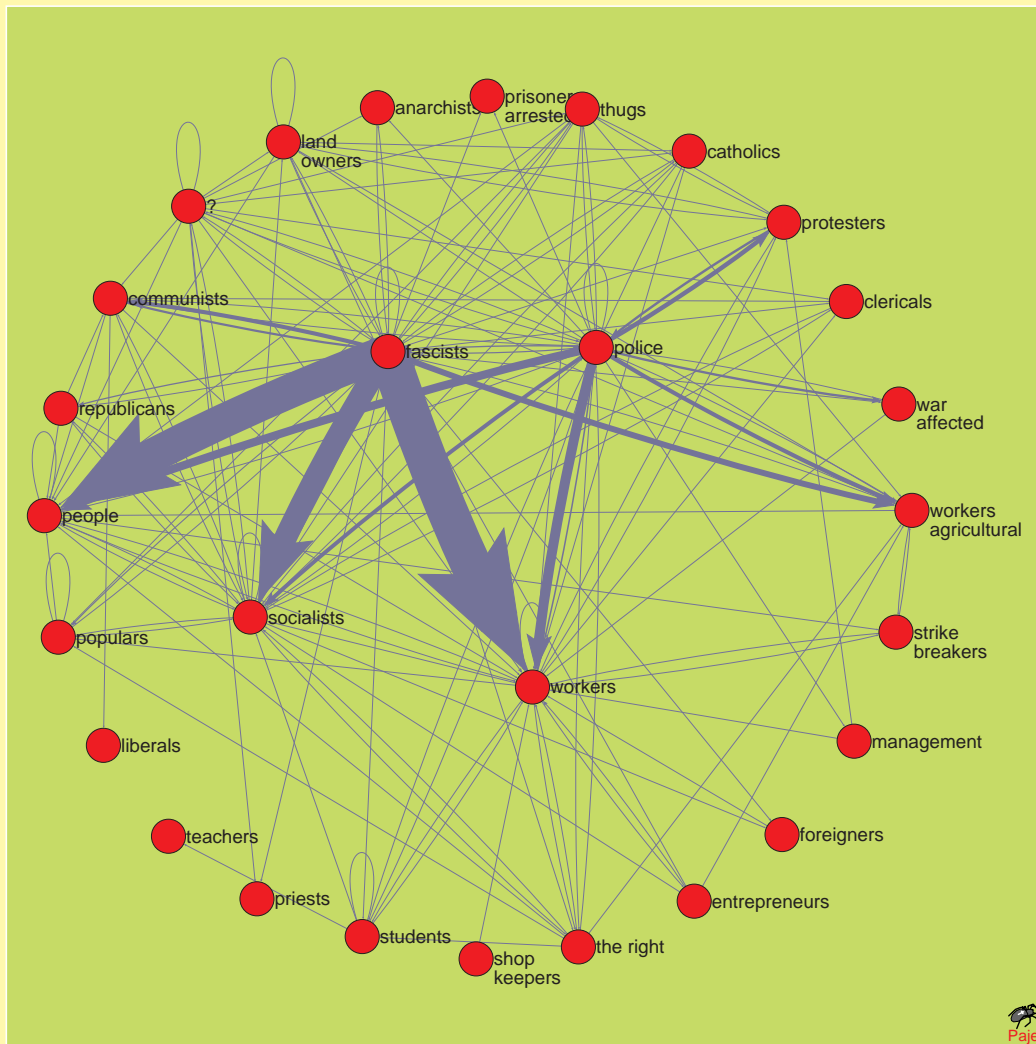
We shrink similar/equivalent units determined by a partition `similar.clu`

```
File / Partition / Read [similar.clu]
Operations / Extract from Network / Partition [1-*]
Operations / Shrink Network / Partition [1][0]
[Edit partition] change class names to English (val is the code!)
File / Partition / Save [classes.clu]
Draw / Draw
Draw: Layout / Energy / Fruchterman Reingold / 2D [factor = 8]
Draw: Export / 2D / EPS/PS [picAll.eps]
File / Network / Save [violenceAll.net]
```

...Network analysis / Classes of actors

0	undefined	*vertices	310
1	?	1	1 "?" [1-*]
2	people	2	2 "abitanti" [1-*]
3	police	2	3 "abitanti-passanti" [1-*]
4	land owners	2	4 "adulti" [1-*]
5	anarchists	3	5 "agenti" [1-*]
6	fascists	3	6 "agenti in borghese" [1-*]
7	communists	3	7 "agenti investigativi" [1-*]
8	workers (agricultural)	3	8 "agenti-investigativi" [1-*]
9	socialists	3	9 "agenti-scorta" [1-*]
10	clericals	0	10 "aggressori" [1-*]
11	war affected	4	11 "agrari" [1-*]
12	protesters	4	12 "agricoltori" [1-*]
13	strike breakers	2	13 "amici" [1-*]
14	management	2	14 "ammalati" [1-*]
15	workers	0	15 "amministratori" [1-*]
16	the right	5	16 "anarchici" [1-*]
17	entrepreneurs	5	17 "anarcoidi" [1-*]
18	populars	2	18 "anziani" [1-*]
19	students	6	19 "arditi" [1-*]
20	catholics	7	20 "arditi del popolo" [1-*]
21	shop keepers	26	21 "arrestati" [1-*]
22	teachers	3	22 "artiglieri" [1-*]
23	priests	
24	republicans	2	306 "vecchie" [1-*]
25	thugs	11	307 "vedove di guerra" [1-*]
26	prisoners/arrested	2	308 "viaggiatori" [1-*]
27	foreigners	2	309 "viandanti" [1-*]
28	liberals	2	310 "villeggianti" [1-*]

...Network analysis / Classes of actors



Picture of the reduced network displays an overall structure of violence among classes in all four years.

The picture also provides coordinates of vertices in time slices.

Network analysis

Work in progress !!!

Make time slices Generate in time; for each

Draw the sequence in **Pajek**: **Pajek project file** with sequence of time slices – read into **Pajek**, draw the first network and inspect the sequence using `next` and `previous` buttons.

SVGanim: you need a web browser with **SVG viewer** installed. See **animation** (transition 4s, pause 5s).

Determine degrees for each slice

Export vectors to R

Draw time series

R / Police (blue) and Fascists (black) activity

