



Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Analysis of bibliographic networks

Vladimir Batagelj

IMFM Ljubljana IAM UP Koper and NRU HSE Moscow

Daria Maltseva

NRU HSE Moscow

XXXIX Sunbelt Social Networks Conference
Montreal, June 18, 2019



Outline

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

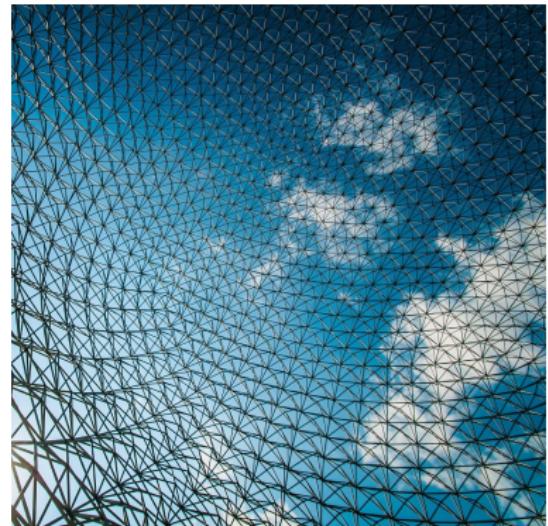
Derived Ns

Temporal Ns

References

Appendix

- 1 Networks
- 2 Statistics
- 3 Citation
- 4 Two-mode Ns
- 5 Multiplication
- 6 Derived Ns
- 7 Temporal Ns
- 8 References
- 9 Appendix



Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si

Daria Maltseva: d_malceva@mail.ru

Current version of slides (June 18, 2019 at 04:21): [slides PDF](#)

<https://github.com/bavla/biblio/blob/master/doc/WS/bibnet19b.pdf>



Networks

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

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

Additional data about nodes or links may be known – their *properties* (attributes). For example: name/label, type, age, value, ...

Network = Graph + Data

A *network* $\mathcal{N} = (\mathcal{V}, \mathcal{L}, \mathcal{P}, \mathcal{W})$ consists of:

- a *graph* $\mathcal{G} = (\mathcal{V}, \mathcal{L})$, where \mathcal{V} is the set of nodes, \mathcal{A} is the set of arcs, \mathcal{E} is the set of edges, and $\mathcal{L} = \mathcal{E} \cup \mathcal{A}$ is the set of links.
 $n = |\mathcal{V}|$, $m = |\mathcal{L}|$
- \mathcal{P} *node value functions* / properties: $p: \mathcal{V} \rightarrow A$
- \mathcal{W} *link value functions* / weights: $w: \mathcal{L} \rightarrow B$



Types of networks

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

In a *two-mode* network $\mathcal{N} = ((\mathcal{V}_1, \mathcal{V}_2), \mathcal{L}, \mathcal{P}, \mathcal{W})$ its set of nodes is split to two subsets. Each link has its end-nodes in both sets.

In a *multi-relational* network $\mathcal{N} = (\mathcal{V}, (\mathcal{L}_i, i \in I), \mathcal{P}, \mathcal{W})$ the set of its links is partitioned into several mutually disjoint subsets – relations. (Subject Verb Object).

In a *temporal* network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, \mathcal{T}, \mathcal{P}, \mathcal{W})$ the time \mathcal{T} is added. To each node and to each link its *activity* set is assigned. Also properties and weights can change through time – temporal quantities.

A *collection* of networks consists of some networks with common subsets of nodes.

Types of networks can be combined – for example: a temporal two-mode multi-relational network.



Analyses

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

The saved records from a data base can still contain some inconsistencies. Some of them are detected as results of the analyses. The simplest way to deal with them is to correct them in the saved data base file and rerun the creation of Pajek's files and analyses.

To improve the quality of the data some tools for detecting (possible) inconsistencies could be developed.

Check (in Pajek) the obtained networks for multiple lines and remove them, if they exist. Remove also the loops from the citation network.



Cleaning networks

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Using on a file *N.net* the commands

Info/Network/General

Net/Transform/Remove/Loops

Net/Transform/Remove lines/Single line

we get the information about the number of loops and multiple links.
Remove loops, and replace multiple links with single links. The obtained network we save (Options - Save coordinates [OFF]) to file *Nclean.net*.



Network boundary problem

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

If the obtained collection of networks contains also the citation network it can be used to address the *network boundary problem*. Otherwise we assume that we included in our bibliography (all/most of) the relevant works.

The first option is to limit the network to the works with complete descriptions – records from the WoS file, $DC > 0$. Since for cited-only works only the first author (no keywords, ...) is known this option is used for most analyses.



Network boundary problem

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

We can get a richer network if we decide to include also some cited-only works that are cited often – at least k times; we delete nodes for which it holds $(0 < \text{indeg}(v) < k) \wedge (\text{outdeg}(v) = 0)$.

```
Net/Partition/Degree/Input  
Partition/Binarize [1-(k-1)]  
Net/Partition/Degree/Output  
Partition/Binarize [0]  
[select partition 1]  
[select partition 2]  
Partitions/Min(V1,V2)  
Operations/Extract from Network/Partition [0]
```

For some (most frequent) of these additional works we can augment the WoS file with their descriptions (without CR data).



PEERE network

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

To the *Web of Science* (WoS) we put the query "peer review*". In May and June 2015 we got (from Web of Science Core Collection) 17053 hits, and additional 2867 hits for the query refereeing.

In March 2016 we updated the data by adding hits for the years 2015 and 2016 and manually prepared short descriptions for the most cited works (fields: AU, PU, TI, PY, PG, KW; but without CR data).

The analysis in 2015 revealed many papers without WoS descriptions having large indegrees in the citation network. We manually searched in WoS for each of them (with indegree larger or equal to 20) and, if found, we added them into the data set. Important earlier papers often did not use the now established terminology and were therefore overlooked by our queries.

The final run of the program WoS2Pajek produced networks with sets of the following sizes: works $|W| = 721547$, authors $|A| = 295849$, journals $|J| = 39988$, and keywords $|K| = 36279$. In both phases 22981 records were collected. There were 887 duplicates (considered only once).

We removed multiple links and loops (resulting from homonyms) from the networks. The cleaned citation network **CiteAll** has $n = 721547$ nodes and $m = 869821$ arcs. [paper](#)



PEERE – most cited works

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

n	freq	first author	title
1	173	Cohen, J	Statistical Power Analysis for the Behavioral Sciences. Routledge, 1988
2	164	Peters, DP	Peer-review practices of psychological journals - the fate of ... Behav Brain Sci
3	151	Egger, M	Bias in meta-analysis detected by a simple, graphical test. Brit Med J, 1997
4	150	Stroup, DF	Meta-analysis of observational studies in epidemiology - A proposal for reporting
5	135	Dersimonian, R	Metaanalysis in clinical-trials. Control Clin Trials, 1986
6	130	Zuckerman, H	Patterns of evaluation in science - institutionalisation, structure and function
7	130	Higgins, JPT	Cochrane Handbook for Systematic Reviews of Interventions. Cochrane, 2011
8	126	Moher, D	Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement
9	125	Higgins, JPT	Measuring inconsistency in meta-analyses. Brit Med J, 2003
10	121	Cicchetti, DV	The reliability of peer-review for manuscript and grant submissions - ... Behav
11	119	Hirsch, JE	An index to quantify an individual's scientific research output. P Natl Acad Sci
12	114	Mahoney, M	Publication prejudices: An experimental study of confirmatory bias ... Cognit
13	114	van Rooyen, S	Effect of open peer review on quality of reviews and on reviewers' recommend
14	114	Easterbrook, PJ	Publication bias in clinical research. Lancet, 1991
15	110	Landis, JR	Measurement Of Observer Agreement For Categorical Data. Biometrics, 1977
16	109	Godlee, F	Effect on the quality of peer review of blinding reviewers and asking them to
17	108	Horrobin, DF	The philosophical basis of peer-review and the suppression of innovation. J Am
18	107	Moher, D	Preferred Reporting Items for Systematic Reviews and Meta-Analyses: PRISMA
19	107	Jadad, AR	Assessing the quality of reports of randomized clinical trials: Is blinding necessary?
20	105	Mcnutt, RA	The effects of blinding on the quality of peer-review - a randomized trial. J Am
21	104	Cole, S	Chance and consensus in peer-review. Science, 1981
22	103	Moher, D	Improving the quality of reports of meta-analyses of randomised controlled tri
23	98	Justice, AC	Does masking author identity improve peer review quality? - A randomized c
24	97	Lock, S	A Difficult Balance: Editorial Peer Review in Medicine. Nuffield Trust, 1985
25	95	van Rooyen, S	Effect of blinding and unmasking on the quality of peer review - A randomiz
26	92	Black, N	What makes a good reviewer and a good review for a general medical journal
27	91	Scherer, RW	Full publication of results initially presented in abstracts - a metaanalysis. J Am
28	90	Higgins, JPT	Quantifying heterogeneity in a meta-analysis. Stat Med, 2002
29	90	Smith, R	Peer review: a flawed process at the heart of science and journals. J Roy Soc
30	87	Goodman, SN	Manuscript quality before and after peer-review and editing at Annals of Inter



Distributions

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

We read in Pajek the citation network `cite.net` for 'centrality' literature. First we remove loops and multiple lines. Then we determine the indegrees and outdegrees. We dispose the normalized degree vectors, transform both partitions into vectors and call R from Pajek submitting all vectors.

```
#####
# R called from Pajek
# The following vectors read:
v3 : From partition 1 (548600)
v4 : From partition 2 (548600)
-----
> inTab <- table(v3)
> indeg <- as.integer(names(inTab))
> inDeg <- indeg[indeg>0]
> inFreq <- as.vector(inTab[indeg>0])
> plot(inDeg,inFreq,log='xy',main="in-degree distribution")
> ouTab <- table(v4)
> outdeg <- as.integer(names(ouTab))
> outDeg <- outdeg[outdeg>0]
> outFreq <- as.vector(ouTab[outdeg>0])
> plot(outDeg,outFreq,log='xy',main="out-degree distribution")
```



Distributions

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

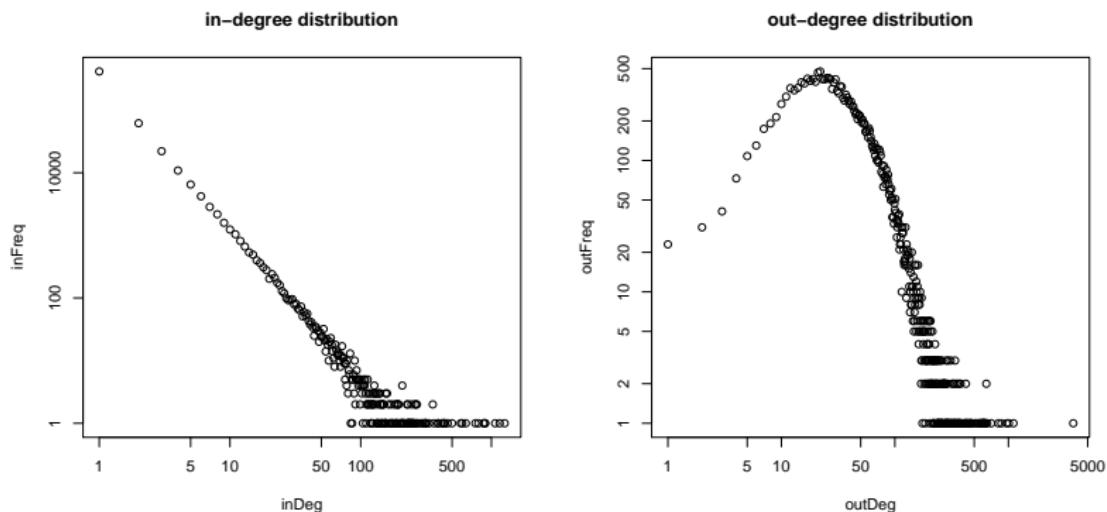
Multiplication

Derived Ns

Temporal Ns

References

Appendix



The in-degree distribution is "scale-free"-like. The parameters can be determined using the package of [Clauset, Shalizi and Newman](#). See also [Stumpf, et al.: Critical Truths About Power Laws](#).



Distributions

SN5 citation network input degrees – scale-free fit

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

SN5 (“social network*” AND SO=(Social networks)) plus most frequently cited works plus around 100 SNA researchers (collected January 2008).

The function `plfit(x)` fits the distribution $p(x) = \frac{\alpha-1}{x_{min}} \left(\frac{x}{x_{min}}\right)^{-\alpha}$. It returns estimates of *alpha*, *x_{min}* and *D* - the Kolmogorov-Smirnov goodness-of-fit statistic. Install library VGAM (zeta function).

```
> source("http://tuvalu.santafe.edu/~aaronc/powerlaws/plfit.r")
> infile <- "https://raw.githubusercontent.com/bavla/biblio/master/dat/indegCite.vec"
> d <- read.table(infile,skip=1,header=FALSE)
> t <- table(d)
> x <- as.numeric(names(t))
> plot(x,t,log="xy",pch=16,main="SN5 cite input degrees / logs")
> len <- length(t)
> z <- t[len:1]; y <- cumsum(z); z <- y[len:1]
> plot(x,z,log="xy",pch=16,cex=0.6,xlab='deg',ylab='num',main="BA cum logs")
> D <- as.vector(d$V1); D <- D[D>0]
> r <- plfit(D)
> names(r)
[1] "xmin"  "alpha"  "D"
> r$alpha
[1] 2.45
> r$xmin
[1] 3
> r$D
[1] 0.006133737
> plot(x,t,log="xy",pch=16,main="SN5 cite input degrees / logs")
> b <- (r$alpha-1)*r$xmin***(r$alpha-1)
> abline(b-r$xmin+1,-r$alpha,col="red",lw=2)
```



Distributions

SN5 citation network input degrees – scale-free fit

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

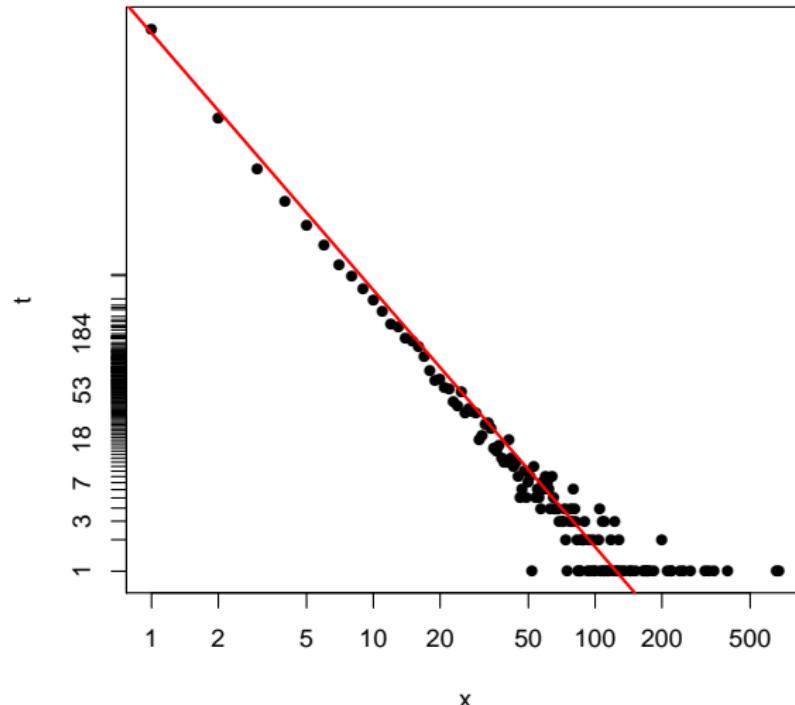
Multiplication

Derived Ns

Temporal Ns

References

Appendix





Distributions

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

From the file `Year.clu` we can get the distribution of *citations by years*.
For the centrality network we get:

```
> setwd("C:/Users/Batagelj/work/Python/WoS/Central")
> years <- read.table(file="Year.clu",header=FALSE,skip=2)$V1
> t <- table(years)
> year <- as.integer(names(t))
> freq <- as.vector(t[1950<=year & year<=2009])
> y <- 1950:2009
> plot(y,freq)
> model <- nls(freq~c*dlnorm(2010-y,a,b),start=list(c=350000,a=2,b=0.7))
> model
Nonlinear regression model
  model: freq ~ c * dlnorm(2010 - y, a, b)
  data: parent.frame()
      c      a      b 
5.427e+05 2.491e+00 6.624e-01 
 residual sum-of-squares: 20474181

Number of iterations to convergence: 7
Achieved convergence tolerance: 3.978e-06
> lines(y,predict(model,list(x=2010-y)),col='red')
```

It can be well approximated by the *lognormal distribution*, but also by the *generalized reciprocal power exponential curve* $c * (x + d)^{\frac{a}{b+x}}$.



Distributions

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

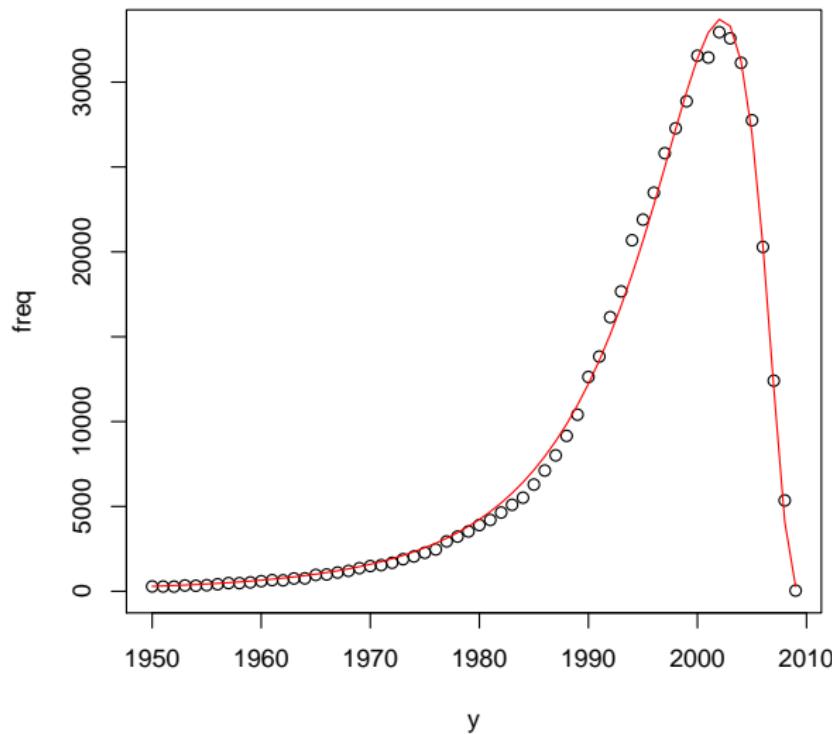
Multiplication

Derived Ns

Temporal Ns

References

Appendix





Zentralblatt MATH networks

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Data from the ZB data base for years 1990–2010.

Network	WA	WJ	WK	WM
Size of the first set	1339201	1339201	1339201	1339201
Size of the second set	557104	3158	143513	12390
Number of arcs	2550437	1331036	15062377	3370820

See the [paper](#).



Distributions – number of keywords

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

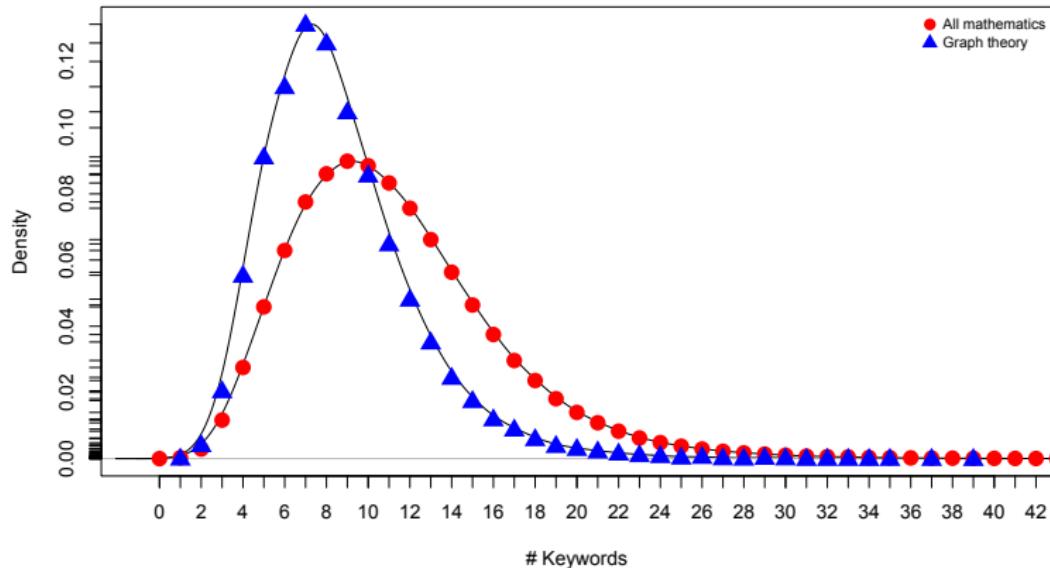
Multiplication

Derived Ns

Temporal Ns

References

Appendix





Distributions – keywords by the number of works using a keyword in their description

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

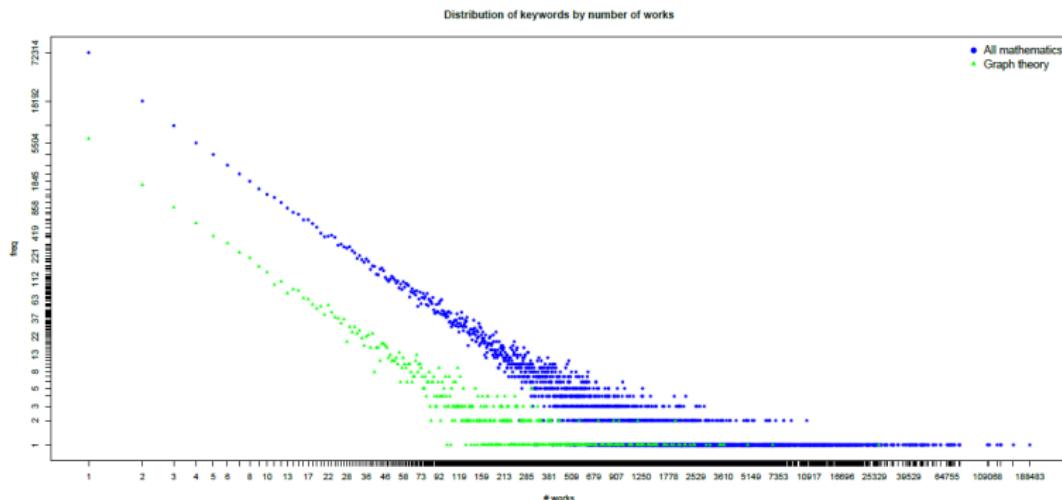
Multiplication

Derived Ns

Temporal Ns

References

Appendix



Citation networks

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

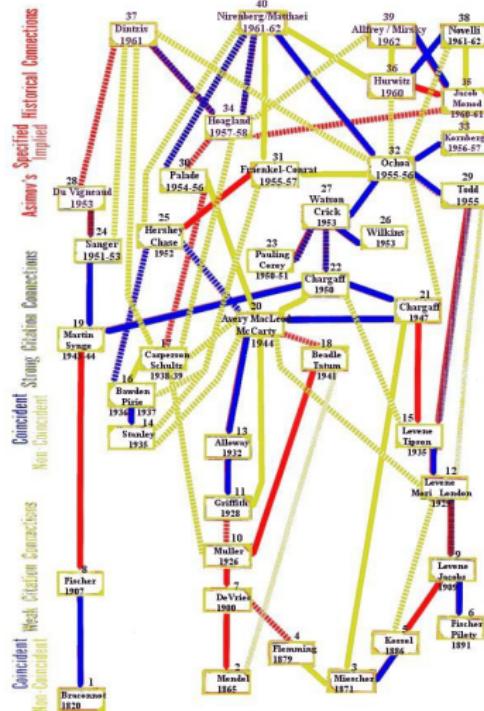
Multiplication

Derived Ns

Temporal Ns

References

Appendix





... Citation networks

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

In a given set of works/nodes W (articles, books, works, etc.) we introduce a *citing relation*/set of arcs $\mathbf{Ci} \subseteq W \times W$

$$u \mathbf{Ci} v \equiv u \text{ cites } v$$

which determines a *citation network* $\mathcal{N} = (W, \mathbf{Ci})$.

A citing relation is usually *irreflexive* (no loops) and (almost) *acyclic*. We shall assume that it has these two properties. Since in real-life citation networks the strong components are small (usually 2 or 3 nodes) we can transform such network into an acyclic network by shrinking strong components and deleting loops.



Preparing the citation network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

For further analysis the citation network has to be acyclic – has no nontrivial strong component. To identify nontrivial strong components and extract them use the commands:

Net/Components/Strong [2]

Operations/Extract from Network/Partition [1-*]

Operations/Transform/Remove Lines/Between Clusters

Save the obtained network to a file *Nstrong.net*.



... Preparing the citation network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

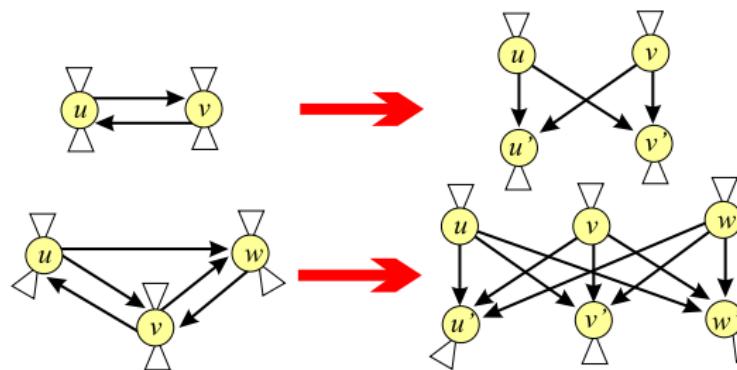
Multiplication

Derived Ns

Temporal Ns

References

Appendix



To transform the network *PRciteR.net* into an acyclic network using the preprint transformation use the Pajek's command

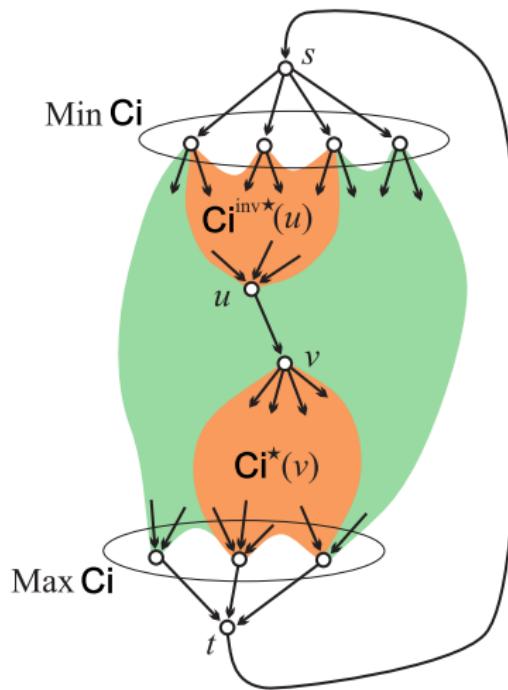
`Network/Acyclic Network/Transform/Preprint Transformation`

Standardized citation network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks
Statistics
Citation
Two-mode Ns
Multiplication
Derived Ns
Temporal Ns
References
Appendix



We assume that the citation relation Ci is acyclic. It is useful to transform a citation network to its *standardized* form by adding a common *source* node $s \notin W$ and a common *sink* node $t \notin W$. The source s is linked by an arc to all minimal elements of Ci ; and all maximal elements of Ci are linked to the sink t . We add also the 'feedback' arc (t, s) .



Search path count method

Hummon and Doreian

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

The *search path count* (SPC) method is based on counters $n(u, v)$ that count the number of different paths from s to t through the arc (u, v) . To compute $n(u, v)$ we introduce two auxiliary quantities: $n^-(v)$ counts the number of different paths from s to v , and $n^+(v)$ counts the number of different paths from v to t . Then

$$n(u, v) = n^-(u) \cdot n^+(v)$$

There exists a very efficient algorithm to compute counters $n^-(v)$ and $n^+(v)$.

The quantities used to compute the arc weights n can be used also to define the corresponding *node weight* t_c

$$t_c(u) = n^-(u) \cdot n^+(u)$$

It is counting the number of paths through the node u .



Properties of SPC weights

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

The values of counters $n(u, v)$ form a flow in the citation network – the *Kirchoff's vertex law* holds: For every node u in a standardized citation network $\text{incoming flow} = \text{outgoing flow}$:

$$\sum_{v: v \text{ Ci } u} n(v, u) = \sum_{v: u \text{ Ci } v} n(u, v) = n^-(u) \cdot n^+(u) = t_c(u)$$

The weight $n(t, s)$ equals to the total flow through network and provides a natural normalization of weights

$$w(u, v) = \frac{n(u, v)}{n(t, s)} \quad \Rightarrow \quad 0 \leq w(u, v) \leq 1$$

and if C is a minimal arc-cut-set $\sum_{(u,v) \in C} w(u, v) = 1$.

In large networks the values of weights can grow very large. This should be considered in the implementation of the algorithms.



Cuts

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

The standard approach to find interesting groups inside a network was based on properties/weights – they can be *measured* or *computed* from network structure (for example Kleinberg's **hubs and authorities**).

The *node-cut* of a network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, p)$, $p : \mathcal{V} \rightarrow \mathbb{R}$, at selected level t is a subnetwork $\mathcal{N}(t) = (\mathcal{V}', \mathcal{L}(\mathcal{V}'), p)$, determined by the set

$$\mathcal{V}' = \{v \in \mathcal{V} : p(v) \geq t\}$$

and $\mathcal{L}(\mathcal{V}')$ is the set of links from \mathcal{L} that have both endnodes in \mathcal{V}' .

The *link-cut* of a network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, w)$, $w : \mathcal{L} \rightarrow \mathbb{R}$, at selected level t is a subnetwork $\mathcal{N}(t) = (\mathcal{V}(\mathcal{L}'), \mathcal{L}', w)$, determined by the set

$$\mathcal{L}' = \{e \in \mathcal{L} : w(e) \geq t\}$$

and $\mathcal{V}(\mathcal{L}')$ is the set of all endnodes of the links from \mathcal{L}' .

```
File/Network/read eatRS.net
Info/Network/Line values ... >= 70
Net/Transform/Remove/Lines with Value/lower than 70
Net/Partitions/Degree/All
Operations/Extract from Network/Partition 1-*
Net/Components/Weak
Draw/Draw-Partition
```



Citation weights

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

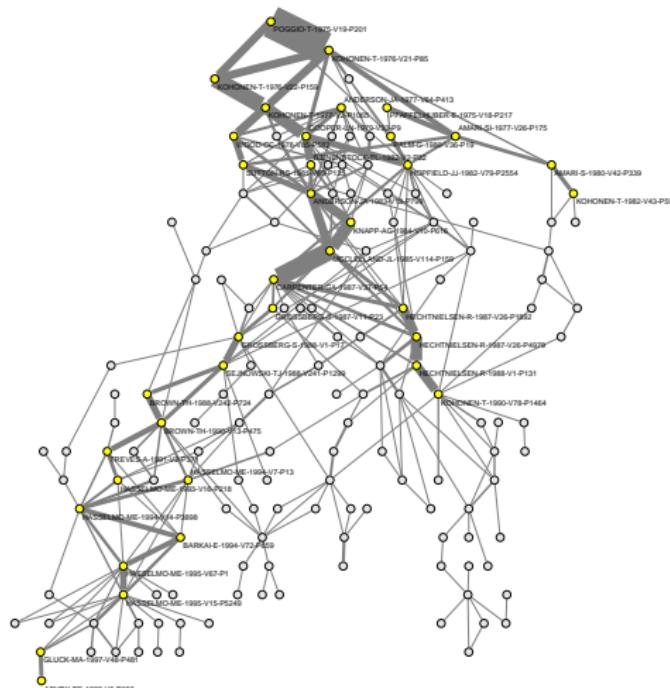
Multiplication

Derived Ns

Temporal Ns

References

Appendix



Main subnetwork (arc cut at level 0.007) of the SOM (selforganizing maps) citation network (4470 nodes, 12731 arcs).
See [paper](#).

Cores and generalized cores

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

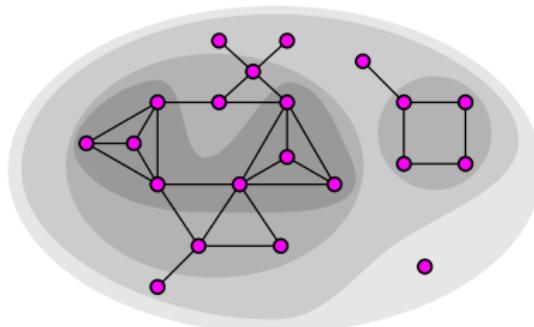
Multiplication

Derived Ns

Temporal Ns

References

Appendix



The notion of core was introduced by Seidman in 1983. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph. A subgraph $\mathcal{H} = (\mathcal{C}, \mathcal{E}|_{\mathcal{C}})$ induced by the set \mathcal{C} is a ***k*-core** or a ***core of order k*** iff $\forall v \in \mathcal{C} : \deg_{\mathcal{H}}(v) \geq k$, and \mathcal{H} is a maximal subgraph with this property. The core of maximum order is also called the ***main*** core.

The ***core number*** of a node v is the highest order of a core that contains this node. The degree $\deg(v)$ can be: in-degree, out-degree, in-degree + out-degree, etc., determining different types of cores.



Properties of cores

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

From the figure, representing 0, 1, 2 and 3 core, we can see the following properties of cores:

- The cores are nested: $i < j \implies \mathcal{H}_j \subseteq \mathcal{H}_i$
- Cores are not necessarily connected subgraphs.

An efficient algorithm for determining the cores hierarchy is based on the following property:

If from a given graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ we recursively delete all nodes, and edges incident with them, of degree less than k , the remaining graph is the k -core.



... Properties of cores

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

The cores, because they can be determined very efficiently, are one among few concepts that provide us with meaningful decompositions of large networks. We expect that different approaches to the analysis of large networks can be built on this basis. For example: we get the following bound on the chromatic number of a given graph \mathcal{G}

$$\chi(\mathcal{G}) \leq 1 + \text{core}(\mathcal{G})$$

Cores can also be used to localize the search for interesting subnetworks in large networks since: if it exists, a k -component is contained in a k -core; and a k -clique is contained in a k -core.
For details see the [paper](#).



Generalized cores

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

The notion of core can be generalized to networks. Let $\mathcal{N} = (\mathcal{V}, \mathcal{E}, w)$ be a network, where $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a graph and $w : \mathcal{E} \rightarrow \mathbb{R}$ is a function assigning values to edges. A *node property function* on \mathbf{N} , or a *p-function* for short, is a function $p(v, U)$, $v \in \mathcal{V}$, $U \subseteq \mathcal{V}$ with real values. Let $N_U(v) = N(v) \cap U$. Besides degrees and (corrected) clustering coefficient, here are some examples of *p-functions*:

$$p_S(v, U) = \sum_{u \in N_U(v)} w(v, u), \text{ where } w : \mathcal{E} \rightarrow \mathbb{R}_0^+$$

$$p_M(v, U) = \max_{u \in N_U(v)} w(v, u), \text{ where } w : \mathcal{E} \rightarrow \mathbb{R}$$

$$p_k(v, U) = \text{number of cycles of length } k \text{ through the node } v \text{ in } (U,$$

The subgraph $\mathcal{H} = (C, \mathcal{E}|_C)$ induced by the set $C \subseteq \mathcal{V}$ is a *p-core at level* $t \in \mathbb{R}$ iff $\forall v \in C : t \leq p(v, C)$ and C is a maximal such set.



Additional p -functions

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

relative density

$$p_\gamma(v, \mathcal{C}) = \frac{\deg(v, \mathcal{C})}{\max_{u \in N(v)} \deg(u)}, \text{ if } \deg(v) > 0; 0, \text{ otherwise}$$

diversity

$$p_\delta(v, \mathcal{C}) = \max_{u \in N^+(v, \mathcal{C})} \deg(u) - \min_{u \in N^+(v, \mathcal{C})} \deg(u)$$

average weight

$$p_a(v, \mathcal{C}) = \frac{1}{|N(v, \mathcal{C})|} \sum_{u \in N(v, \mathcal{C})} w(v, u), \text{ if } N(v, \mathcal{C}) \neq \emptyset; 0, \text{ otherwise}$$



Generalized cores algorithm

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

The function p is *monotone* iff it has the property

$$C_1 \subset C_2 \Rightarrow \forall v \in \mathcal{V} : (p(v, C_1) \leq p(v, C_2))$$

The degrees and the functions p_S , p_M and p_k are monotone. For a monotone function the p -core at level t can be determined, as in the ordinary case, by successively deleting nodes with value of p lower than t ; and the cores on different levels are nested

$$t_1 < t_2 \Rightarrow \mathcal{H}_{t_2} \subseteq \mathcal{H}_{t_1}$$

The p -function is *local* iff $p(v, U) = p(v, N_U(v))$.

The degrees, p_S and p_M are local; but p_k is **not** local for $k \geq 4$. For a local p -function an $O(m \max(\Delta, \log n))$ algorithm for determining the p -core levels exists, assuming that $p(v, N_C(v))$ can be computed in $O(\deg_C(v))$.

For details see the [paper](#).



Cores and generalized cores / Pajek commands

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

File/Network/Read [Geom.net]
Net/Partitions/Core/All
Info/Partition
Operations/Extract from Network/Partition [13-*]
Draw/Draw-Partition
Layout/Energy/Kamada-Kawai
Options/Values of lines/Similarities
Layout/Energy/Kamada-Kawai
Operations/Extract from Network/Partition [21]
Draw
Layout/Energy/Kamada-Kawai
Options/Values of lines/Forget
Layout/Energy/Kamada-Kawai
[select Geom.net]
Net/Vector/PCore/Sum/All
Info/Vector
Vector/Make Partition/by Intervals/Selected Thresholds [45]
Info/Partition
Operations/Extract from Network/Partition [2]
Draw
Options/Values of lines/Similarities
Layout/Energy/Fruchterman-Reingold



Cores of orders 10–21 in Computational Geometry

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

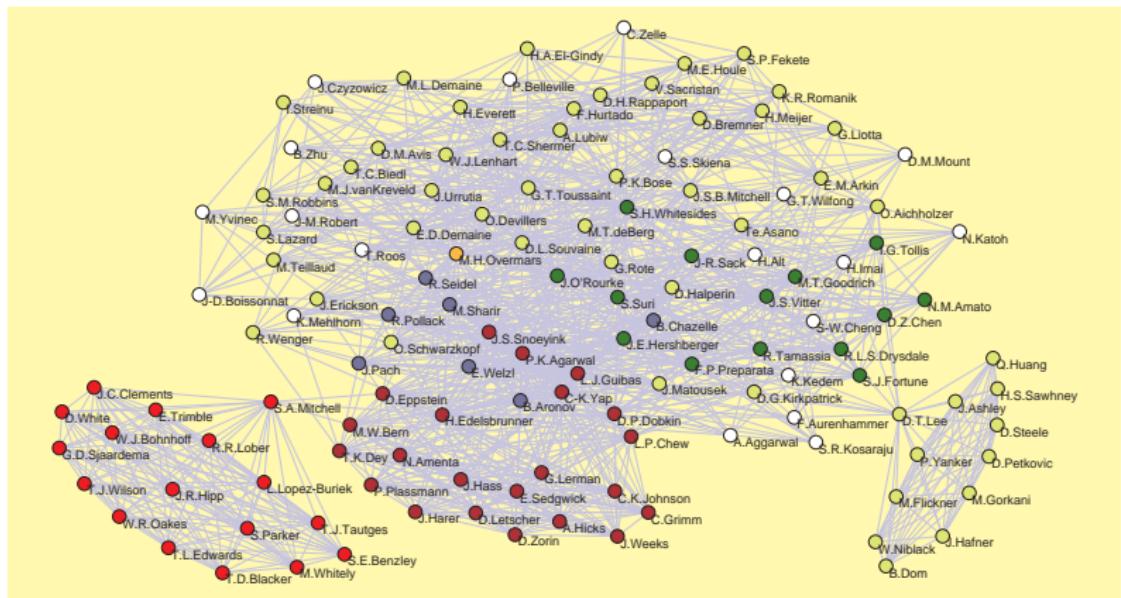
Multiplication

Derived Ns

Temporal Ns

References

Appendix





p_S -core at level 46 in Computational Geometry network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

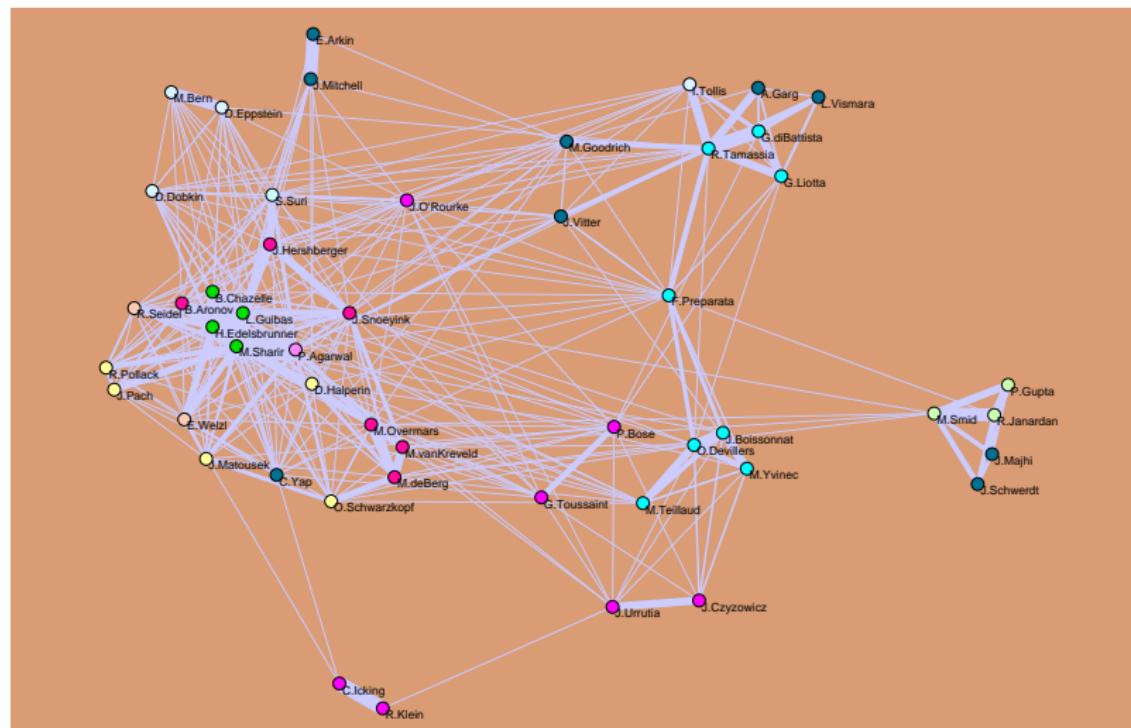
Multiplication

Derived Ns

Temporal Ns

References

Appendix





Islands

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

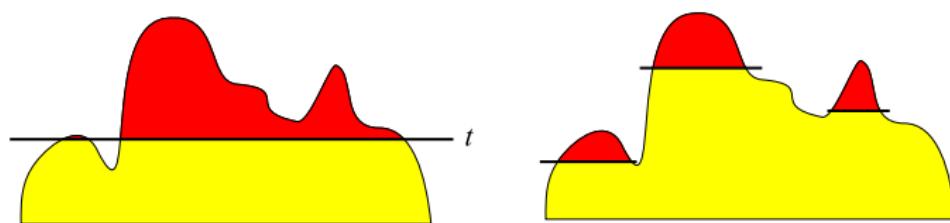
Derived Ns

Temporal Ns

References

Appendix

If we represent a given or computed value of nodes / links as a height of nodes / links and we immerse the network into a water up to selected level we get *islands*. Varying the level we get different islands.



We developed very efficient algorithms to determine the islands hierarchy and to list all the islands of selected sizes.
See [details](#).



... Islands

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Islands are very general and efficient approach to determine the 'important' subnetworks in a given network.

We have to express the goals of our analysis with a related property of the nodes or weight of the links. Using this property we determine the islands of an appropriate size (in the interval k to K).

In large networks we can get many islands which we have to inspect individually and interpret their content.

An important property of the islands is that they identify locally important subnetworks on different levels. Therefore they detect also emerging groups.

The set of nodes $\mathcal{C} \subseteq \mathcal{V}$ is a *local node peak*, if it is a regular node island and all of its nodes have the same value. Node island with a single local node peak is called a *simple node island*. In similar way we define simple link island.



... Islands

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

A set of nodes $C \subseteq \mathcal{V}$ is a *regular node island* in a network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, p)$, $p : \mathcal{V} \rightarrow \mathbb{R}$ iff it induces a connected subgraph and the nodes from the island are 'higher' than the neighboring nodes

$$\max_{u \in N(C)} p(u) < \min_{v \in C} p(v)$$

A set of nodes $C \subseteq \mathcal{V}$ is a *regular link island* in a network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, w)$, $w : \mathcal{L} \rightarrow \mathbb{R}$ iff it induces a connected subgraph and the links inside the island are 'stronger related' among them than with the neighboring nodes – in \mathcal{N} there exists a spanning tree \mathcal{T} over C such that

$$\max_{(u,v) \in \mathcal{L}, u \notin C, v \in C} w(u,v) < \min_{(u,v) \in \mathcal{T}} w(u,v)$$

Network/Create Partition/Islands/Line Weights
Operations/Network+Vector/Islands/Vertex Property



Some properties of islands

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

- The sets of nodes of connected components of node/link-cut at selected level t are regular node/link islands.
- The set $\mathcal{H}_p(\mathcal{N})$ of all regular node islands of a network \mathcal{N} is a complete hierarchy:
 - two islands are disjoint or one of them is a subset of the other
 - each node belongs to at least one island
- The set $\mathcal{H}_w(\mathcal{N})$ of all nondegenerated regular link islands of a network \mathcal{N} is a hierarchy (not necessarily complete):
 - two islands are disjoint or one of them is a subset of the other
 - Node/link islands are invariant for the strictly increasing transformations of the property p / weight w .
 - Two linked nodes cannot/may belong to two disjoint regular node/link islands.

A *simple island* is an island with only one peak.





US patents

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

US patents network ([Nber, US Patents](#)) has 3774768 nodes and 16522438 arcs (1 loop). Without the loop it is acyclic. The weight of an arc is the proportion of paths through the arc from some initial node to some terminal node. We determined all (2,90)-link islands. The corresponding subnetwork has 470137 nodes, 307472 arcs and different k : $C_2 = 187610$, $C_5 = 8859$, $C_{30} = 101$, $C_{50} = 30$, ... islands. [Rolex](#)

[1]	0	139793	29670	9288	3966	1827	997	578	362	250
[11]	190	125	104	71	47	37	36	33	21	23
[21]	17	16	8	7	13	10	10	5	5	5
[31]	12	3	7	3	3	3	2	6	6	2
[41]	1	3	4	1	5	2	1	1	1	1
[51]	2	3	3	2	0	0	0	0	0	1
[61]	0	0	0	0	1	0	0	2	0	0
[71]	0	0	1	1	0	0	0	1	0	0
[81]	2	0	0	0	0	1	2	0	0	7

The **Main path** starts in a link with the largest SPC weight and expands in both directions following the adjacent link with the largest SPC weight.

The **CPM path** is determined using the Critical Path Method from Operations Research (the sum of SPC weights along a path is maximal).



Distribution of island size

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

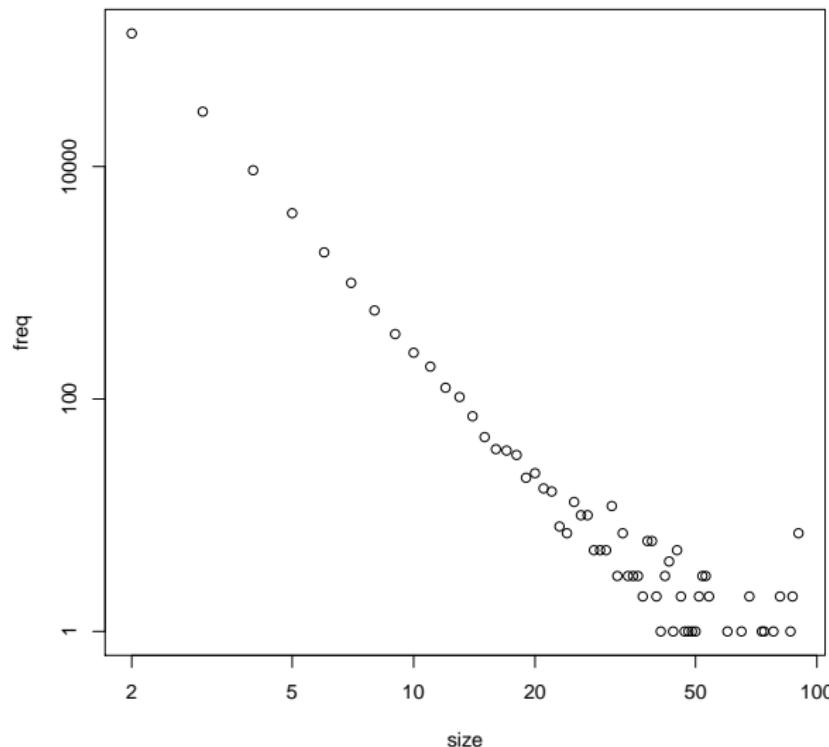
Multiplication

Derived Ns

Temporal Ns

References

Appendix





Main path and main island in US Patents

Nber, US Patents; $n = 3774768$, $m = 16522438$

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

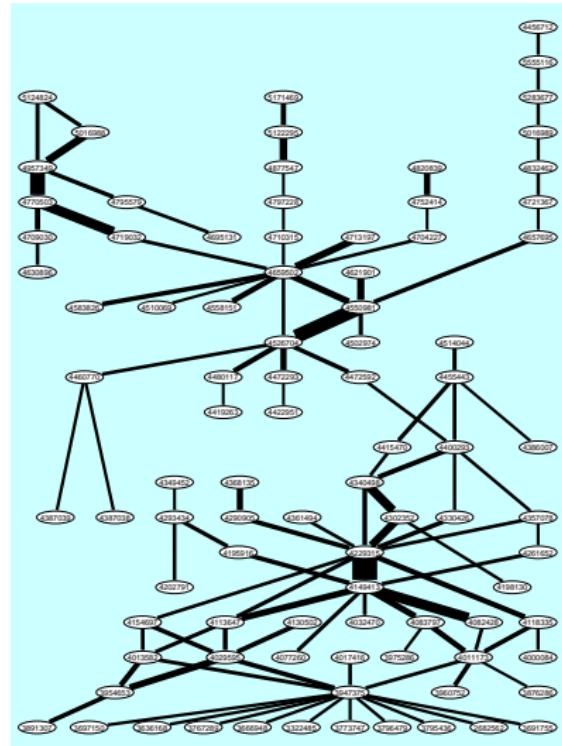
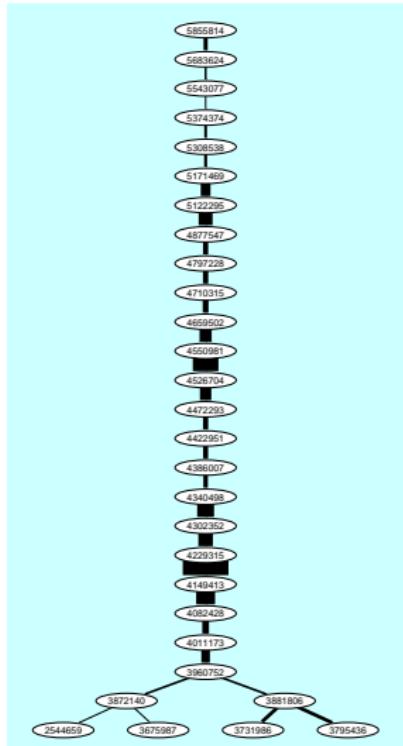
Multiplication

Derived Ns

Temporal Ns

References

Appendix





Main island – Liquid crystal display

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Table 1: Patents on the liquid-crystal display

patent	date	author(s) and title
2049629	Mar 13, 1951	Wender, et al. Preparation of polyesters containing short and the like and the formation and use thereof
2082562	Jul 29, 1954	Wender, et al. Reduction of aromatic carbonyl groups by the addition of elements or groups containing organic nematic compound
3323485	May 30, 1967	Josephson. Preparation of polyacrylate ester nematic compounds
3636168	Jun 18, 1972	Goto, et al. Liquid crystal compositions and devices having an undistorted image on a disturbed background
3691755	Sep 19, 1972	Goto, et al. Liquid crystal compositions and devices
3697150	Oct 19, 1972	Wysoki. Electro-optic systems in which an electrophoretic material is suspended in a liquid crystal and thinnest a liquid crystal to reduce the turn-off time
3731966	May 8, 1973	Fenggas. Display devices utilizing liquid crystal light
3767289	Oct 23, 1973	Arivazhagan. Class of stable ester-thiophene compounds, some displaying nematic or mesomorphic at or near room temperature and others, a range of 0°C to 100°C
3773747	Nov 20, 1973	Steinmesser. Substituted nancy benzene compound which exhibits the Kerr effect at isotropic temperatures
3795436	Mar 3, 1974	Hofrichter, et al. Electro-optical liquid-crystal device
3797479	Mar 12, 1974	Hofrichter, et al. Electro-optical liquid-crystal which exhibits the Kerr effect at isotropic temperatures
3872140	Mar 18, 1975	Klaudermon, et al. Liquid crystalline compositions and methods of use
3876286	Aug 8, 1975	Dentzschek, et al. Use of nematic liquid crystalline substances in electro-optical devices
3890230	Mar 28, 1976	Tanakadate, et al. Phase control of the voltages applied to opposite electrodes for a cholesteric to nematic phase transition
3947375	Mar 30, 1976	Goto, et al. Liquid crystal materials and devices
3959633	May 4, 1976	Yamamoto. Liquid crystal compositions and high dielectric constant materials for electro-optic devices incorporating same
3960752	Jun 1, 1976	Klaudermon, et al. Liquid crystal compositions
3973286	Aug 17, 1976	Goto, et al. Liquid crystal compositions and method of synthesis
4000894	Dec 28, 1976	Held, et al. Liquid crystal mixtures for electro-optical devices
4011173	Mar 8, 1977	Steinmesser. Modified nematic mixtures with increased viscosity
4013582	Mar 22, 1977	Gavrilev. Liquid crystal compounds and electro-optic devices incorporating them
4071746	Apr 12, 1977	Method for preparing some 4-alkyl-4'-biphenylcarboxylates, method for preparing some liquid crystal compositions using same
4029095	Jun 14, 1977	Novel liquid crystal compounds and electro-optic devices incorporating them
4032470	Jun 28, 1977	Goto, et al. Liquid crystal compositions
4077260	Mar 7, 1978	Goto, et al. Optically active cyano-biphenyl compounds and liquid crystal materials containing them
4082428	Apr 4, 1978	Goto, et al. Liquid crystal composition and method

Table 2: Patents on the liquid-crystal display

patent	date	author(s) and title
4113647	Apr 11, 1979	Goto, et al. Liquid crystal compositions
4113647	Sep 12, 1979	Goto, et al. Liquid crystalline materials
4118320	Oct 3, 1979	Krause, et al. Liquid crystalline materials of reduced viscosity
4120520	Oct 17, 1979	Goto, et al. Optically active liquid crystal mixture and liquid crystal devices containing them
4149412	May 15, 1979	Goto, et al. Liquid crystal derivative benzoylphenyl derivatives
4150012	Aug 1, 1979	Custod, et al. Liquid crystal materials
4189130	Aug 15, 1980	Goto, et al. Liquid crystal mixtures
4202794	May 13, 1980	Osawa, et al. Nematic liquid crystalline materials
4211362	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213652	Aug 14, 1980	Goto, et al. Liquid crystal compounds and materials and devices containing them
4260052	Sep 22, 1980	Goto, et al. Liquid crystal compositions
4293434	Oct 6, 1980	Dentzschek, et al. Liquid crystal compounds
4302325	Nov 24, 1980	Eldenbach, et al. Polyphenylketone compounds, the preparation of liquid crystal mixtures of liquid crystal dioxetanes
4304242	Jan 10, 1981	Eldenbach, et al. Cyclohexylbiphenyl, their preparation and use as liquid crystal components and liquid crystal mixtures
4304498	Jul 20, 1982	Sugimoto. Halogenated ether derivative
4349452	Sep 14, 1982	Ousman, et al. Cyclohexylbicyclohexanes
4352075	Nov 2, 1982	Goto, et al. Liquid crystal compositions containing an aldehyde ring and exhibiting a low dielectric anisotropy and liquid crystal mixtures and devices incorporating such compositions
4361494	Nov 20, 1982	Osawa. Anisotropic compounds with a positive or positive dielectric anisotropy and devices incorporating such compositions
4368135	Jan 11, 1983	Krause, et al. Liquid crystalline naphthalene derivatives
4369667	May 31, 1983	Krause, et al. Liquid crystalline naphthalene derivatives
4370208	Jun 7, 1983	Eldenbach, et al. 4-(Trans-4'-alkylcyclohexyl) benzoic acid
4387020	Jun 7, 1983	Sugimoto, et al. Trans-4-(trans-4'-alkylcyclohexyl)-cyclohexane carboxylic acid and -cyclohexylbenzoic acid
4400253	Aug 23, 1983	Eldenbach, et al. Liquid crystal fluorine-containing derivatives
4413470	Nov 15, 1983	Goto, et al. Liquid crystal mixtures and electro-optical display elements based thereon
4413925	Dec 6, 1983	Prufrock, et al. Liquid crystalline cyclohexylcarboxylic acid
4422953	Dec 27, 1983	Sugimoto, et al. Liquid crystal benzene derivatives
4435441	Jan 19, 1984	Takatori, et al. Nematic liquid crystal composition
4435442	Jan 19, 1984	Takatori, et al. Nematic liquid crystal composition
4460770	Jul 17, 1984	Petrzilka, et al. Liquid crystal mixture
4472205	Sep 18, 1984	Goto, et al. Liquid crystal substances of two range and liquid crystal compositions containing the same
4472206	Sep 18, 1984	Takatori, et al. Nematic liquid crystal compositions
4481250	Oct 23, 1984	Goto, et al. Liquid crystal compositions
4502974	Mar 5, 1985	Sugimoto, et al. High temperature liquid-crystaline ether compounds
4530069	Aug 9, 1985	Goto, et al. Cyclohexane derivatives

Table 3: Patents on the liquid-crystal display

patent	date	author(s) and title
6113944	Apr 30, 1987	Jul 2, 1985
6120794	Sep 1, 1987	Petrzilka, et al. Melting liquid crystal esters
6120795	Sep 1, 1987	Petrzilka, et al. Liquid crystal esters and mixtures
6155152	Dec 16, 1987	Tokutomi, et al. Novel liquid crystalline compounds
6158326	Dec 22, 1987	Petrzilka, et al. Phenylbenzenes
6160327	Dec 22, 1987	Petrzilka, et al. Nematic liquid crystal mixture
6169096	Dec 23, 1987	Petrzilka, et al. Benzene-based liquid crystal mixture
6170305	Dec 23, 1987	Saito, et al. Substituted pyridines
6170306	Dec 23, 1987	Elshabani, et al. Electro-optic pyridines
6095131	Sep 22, 1987	Balwali, et al. Dibenzotetralin ethanes and their use in liquid crystal displays and liquid crystal esters
6170427	Sep 22, 1987	Krasen, et al. Liquid crystal compounds
6170930	Nov 24, 1987	Petrzilka, et al. Novel liquid crystal mixture
6170931	Nov 24, 1987	Petrzilka, et al. Novel liquid crystal composition and liquid crystal mixtures thereof
6213137	Dec 15, 1987	Eldenbach, et al. Nitroso-containing heterocyclic compounds
6219032	Dec 15, 1987	Yoshimura, et al. Cyclohexane derivatives
6213138	Dec 15, 1987	Eldenbach, et al. Cyclohexylbenzene derivatives
6213139	Dec 15, 1987	Yoshimura, et al. Liquid crystal device
6213141	Dec 15, 1987	Eldenbach, et al. Cyclohexylbenzene derivatives
6213144	Dec 15, 1987	Yoshimura, et al. Cyclohexylbenzene derivatives
6213145	Dec 15, 1987	Bachelder, et al. Liquid crystalline compounds
6213146	Dec 15, 1987	Bachelder, et al. Liquid crystalline compounds
6213147	Dec 15, 1987	Goto, et al. Cyclohexylbenzene derivatives and their use in liquid crystal display devices
6213148	Dec 15, 1987	Goto, et al. Cyclohexane derivative and liquid crystal
6208303	Aug 11, 1989	Krasen, et al. Nitroso-containing heterocyclic esters
6208304	Aug 11, 1989	Krasen, et al. Nitroso-containing heterocyclic esters
6277547	Oct 21, 1990	Weber, et al. Liquid crystal display element
6375749	Sep 15, 1990	Chen, et al. Active matrix screen for the color display of liquid crystal devices, control system and process for producing said screen
6309088	May 21, 1991	Imazu, et al. Liquid crystal display device with a liquid crystal layer
6309089	May 21, 1991	Otsuka, et al. Liquid crystal elements with improved contrast and liquid crystal display
5122295	Jan 16, 1992	Weber, et al. Matrix liquid crystal display
5124924	Jan 23, 1992	Kozaki, et al. Liquid crystal display device comprising a liquid crystal layer and a polarizing plate having principal refractive index in the thickness direction
5283877	Feb 1, 1994	Takatori, et al. Liquid crystal material and liquid crystal display device
5300539	May 3, 1994	Takatori, et al. Liquid crystal display element with green regions between terminal electrodes
5327446	Dec 20, 1994	Weber, et al. Separated liquid-crystal display
5343977	Aug 6, 1995	Hoyer, et al. Nematic liquid-crystal composition
5351018	Aug 22, 1995	Hoyer, et al. Nematic liquid-crystal composition having adjacent electrode terminals not equal in length
5380300	Apr 4, 1995	Shigematsu, et al. Liquid crystal composition
5352514	Jan 5, 1995	Mitsui, et al. Liquid crystal composition and liquid crystal display elements

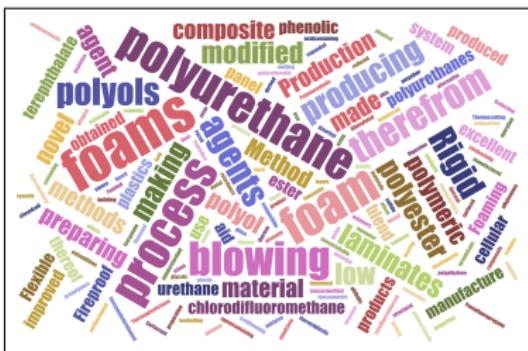


Word clouds for LCD island and foam island

Bibliographic networks

V. Batagelj,
D. Maltseva

- Networks
- Statistics
- Citation**
- Two-mode Ns
- Multiplication
- Derived Ns
- Temporal Ns
- References
- Appendix





Main SPC island in SN5

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

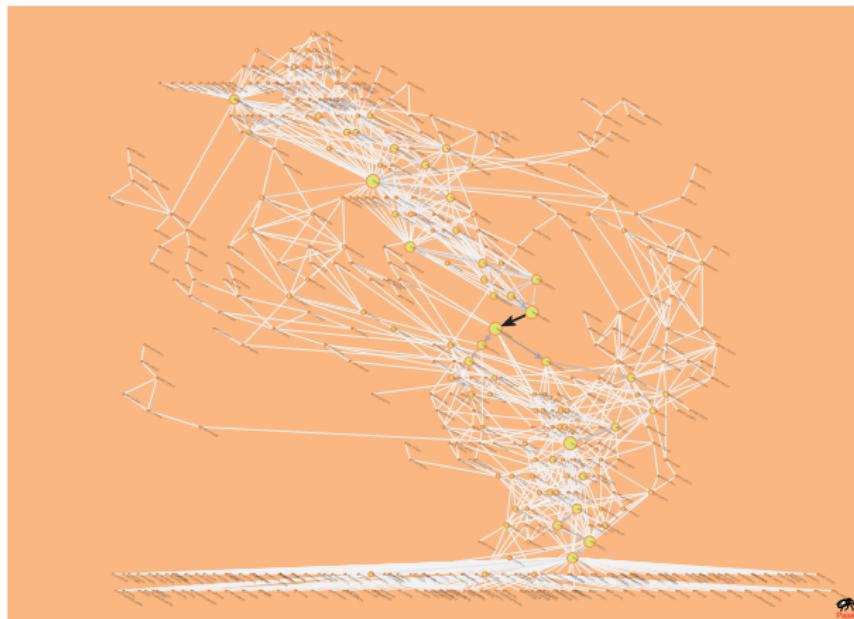
Derived Ns

Temporal Ns

References

Appendix

Network SN5 (2008): for "social network*" + most frequent references + around 100 social networkers;
 $|W| = 193376, |C| = 7950, |A| = 75930, |J| = 14651, |K| = 29267$. Citation networks are acyclic.
Acyclic networks can be displayed by levels – run macro Layers1.mcr from the map Geneo.





PEERE – Main path

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

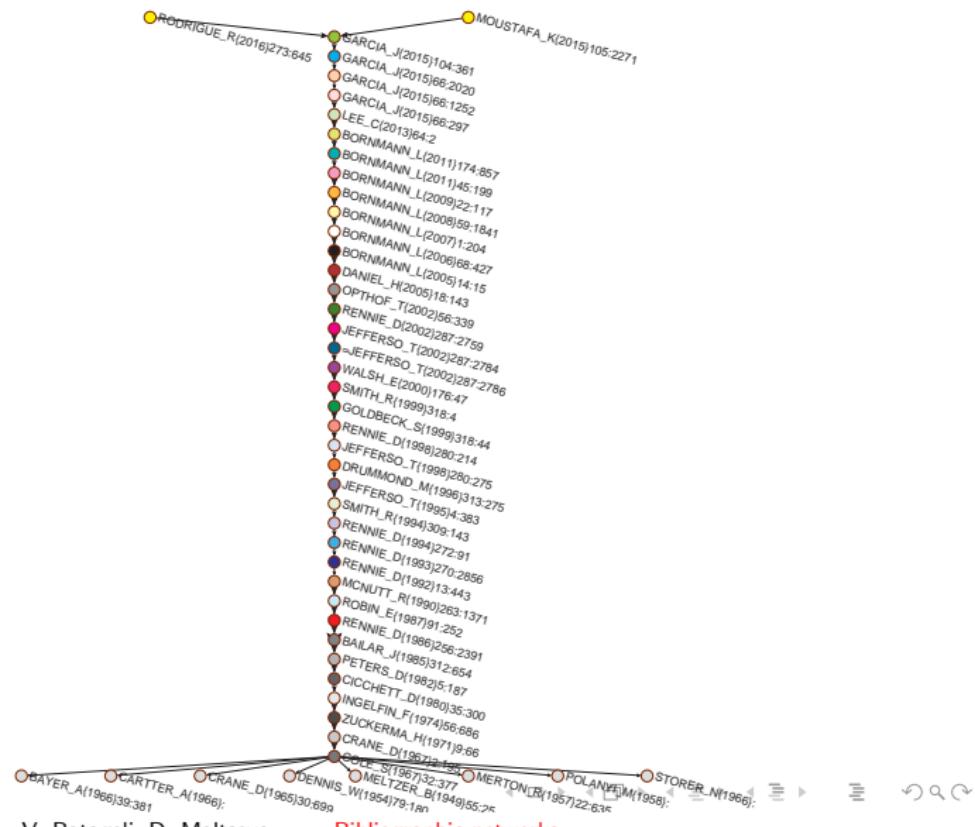
Multiplication

Derived Ns

Temporal Ns

References

Appendix





List of publications on main path

Bibliographic networks

	year	first author	title	journal
V. Batagelj, D. Maltseva	1949	Meltzer BN	The productivity of social scientists	AmJSociol
Networks	1954	Dennis W	Bibliographies of eminent scientists	ScientificM
Statistics	1957	Merton RK	Priorities in scientific discovery - a chapter in the sociology of sci...	AmSocSciRev
Citation	1958	Polanyi M	Personal Knowledge: Towards a Post-Critical Philosophy	UPChicago
Two-mode Ns	1965	Crane D	Scientists at major and minor universities	AmSocioRev
Multiplication	1966	Bayer AE	Some correlates of citation measure of productivity in science	SociolEduc
Derived Ns	1966	Storer NW	The Social System of Science	HRW
Temporal Ns	1966	Cartter A	An Assessment of Quality in Graduate Education	ACE
References	1967	Crane D	Gatekeepers of science - some factors affecting selection...	AmSociol
Appendix	1967	Cole S	Scientific output and recognition - study in operation of reward...	AmSociolRev
	1971	Zuckerman H	Patterns of evaluation in science - institutionalisation, struct...	Minerva
	1974	Ingelfinger FJ	Peer review in biomedical publication	AmJMed
	1980	Cicchetti DV	Reliability of reviews for the american-psychologist	AmPsychol
	1982	Peters DP	Peer-review practices of psychological journals - the fate...	BehavBrainS
	1985	Bailar JC	Journal peer-review - the need for a research agenda	NewEnglJMe
	1986	Rennie D	Guarding the guardians - a conference on editorial peer-review	Jama
	1987	Robin ED	Peer-review in medical journals	Chest
	1990	Mcnutt RA	The effects of blinding on the quality of peer-review	Jama
	1992	Rennie D	Suspended judgment - editorial peer-review - let us put it on trial	ControlClinT
	1993	Rennie D	More peering into editorial peer-review	Jama
	1994	Rennie D	The 2nd international-congress on peer-review in biomedical...	Jama
	1994	Smith R	Promoting research into peer-review	BritMedJ
	1995	Jefferson T	Are guidelines for peer-reviewing economic evaluations necessary	HealthEcon
	1996	Drummond M	Guidelines for authors and peer reviewers of economic submis...	BritMedJ
	1998	Jefferson T	Evaluating the BMJ guidelines for economic submissions	Jama
	1998	Rennie D	Peer review in Prague	Jama



List of works on main path cont.

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

	year	first author	title	journal
	1999	Smith R	Opening up BMJ peer review - A beginning that should lead to...	BritMedJ
	1999	Goldbeck-W. S	Evidence on peer review - scientific quality control or smokescreen?	BritMedJ
	2000	Walsh E	Open peer review: a randomised controlled trial	BritJPsych
	2002	Jefferson T	Effects of editorial peer review - A systematic review	Jama
Networks	2002	Rennie D	Fourth International Congress on Peer Review in Biomedical Pub...	Jama
	2002	Ophof T	The significance of the peer review process against ... bias	Cardiovasc
Statistics	2002	Jefferson T	Measuring the quality of editorial peer review	Jama
	2005	Bornmann L	Committee peer review at an international research foundation	ResEvaluat
Citation	2005	Daniel HD	Publications as a measure of scientific advancement and of...	LearnPubl
	2006	Bornmann L	Selecting scientific excellence through committee peer review	Scientometr
Two-mode Ns	2007	Bornmann L	Convergent validation of peer review decisions using the h index	JInformetr
Multiplication	2008	Bornmann L	Selecting manuscripts for a high-impact journal through peer review	JAmSocInfl
	2009	Bornmann L	The luck of the referee draw: the effect of exchanging reviews	LearnPubl
Derived Ns	2011	Bornmann L	Scientific Peer Review	AnnuRevIn
	2011	Bornmann L	A multilevel modelling approach to investigating the predictive...	JRStatSoc
Temporal Ns	2013	Lee CJ	Bias in peer review	JAmSocInfl
References	2015	Garcia JA	The Principal-Agent Problem in Peer Review	JAssocInflS
	2015	Garcia JA	Adverse selection of reviewers	JAssocInflS
Appendix	2015	Garcia JA	Bias and effort in peer review	JAssocInflS
	2015	Garcia JA	The author-editor game	Scientometr
	2015	Moustafa K	Don't infer anything from unavailable data	Scientometr
	2016	Rodriguez-S. R	Evolutionary games between authors and their editors	ApplMathO



The main path publications

Phases

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

- before 1982: social science journals;
- from 1982 to 2002: biomedical journals;
- after 2002: specialized journals on science studies.



The main path publications till 1982

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Journals: social science journals (sociological, psychological, educational,...) and three books.

The most **influential authors:** Meltzer (1949), Dennis (1954), Merton (1957), Polany (1958), Crane (1965, 1967), Bayer and Folger (1966), Storer (1966), Cartter (1966), Cole and Cole (1967), Zuckerman and Merton (1971), Ingelfinger (1974), Cicchetti (1980) and Peters and Ceci (1982).

Topics: scientific productivity, bibliographies, knowledge, citation measures as measures of scientific accomplishment, scientific output and recognition, evaluation in science, referee system, journal evaluation, peer-evaluation system, review process, peer review practices.



The main path publications from 1983 to 2002

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Journals: biomedical journals, mainly JAMA. From 1986 the International Congress on Peer Review and Biomedical Publication is organized every four years.

The most **influential authors:** Rennie (1986, 1992, 1993, 1994, 1998, 2002), Smith (1994, 1999), Jefferson (1995, 1998, 2002), and their collaborators.

Topics: the effects of blinding on review quality, research into peer review, guidelines for peer reviewing, monitoring the peer review performance, open peer review, bias in peer review system, measuring the quality of editorial peer review. Development of meta-analysis and systematic reviews studies of peer-review.



The main path publications from 2003 on

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Journals: specialized journals on science studies: Scientometrics, Research Evaluation, Journal of Informetrics, JASIST.

The most **influential authors:** Bornmann and Daniel (2005, 2006, 2007, 2008, 2009, 2011) and Garcia, Rodriguez-Sanchez and Fdez-Valdivia (4 papers in 2015, 2016). Others are Lee et al. (2013) and Moustafa (2015).

Topics: Bornmann and Daniel studied the validity of committee peer review process for awarding long-term fellowship to post-graduate researchers, the use of h-index and pre-screening of applications at Boehringer Ingelheim Fonds. They also analysed citations of accepted and rejected papers at a prime chemistry journal, the effect of exchanging reviews, the peer review process in this journal, the validity of its editorial decisions. The other papers studied bias in peer review, selection of reviewers, and the author-editor game.



PEERE – Main paths for 100 largest weights

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

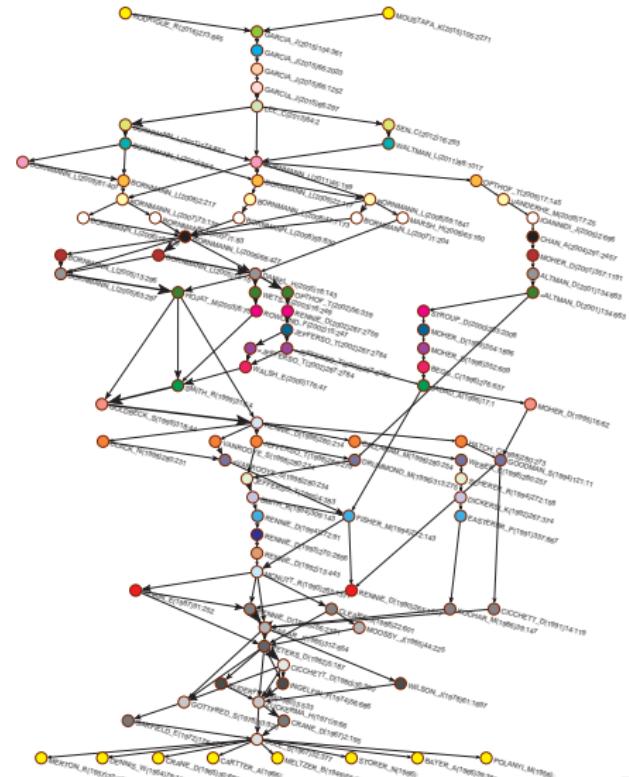
Multiplication

Derived Ns

Temporal Ns

References

Appendix





PEERE – SPC link islands [20 200]

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

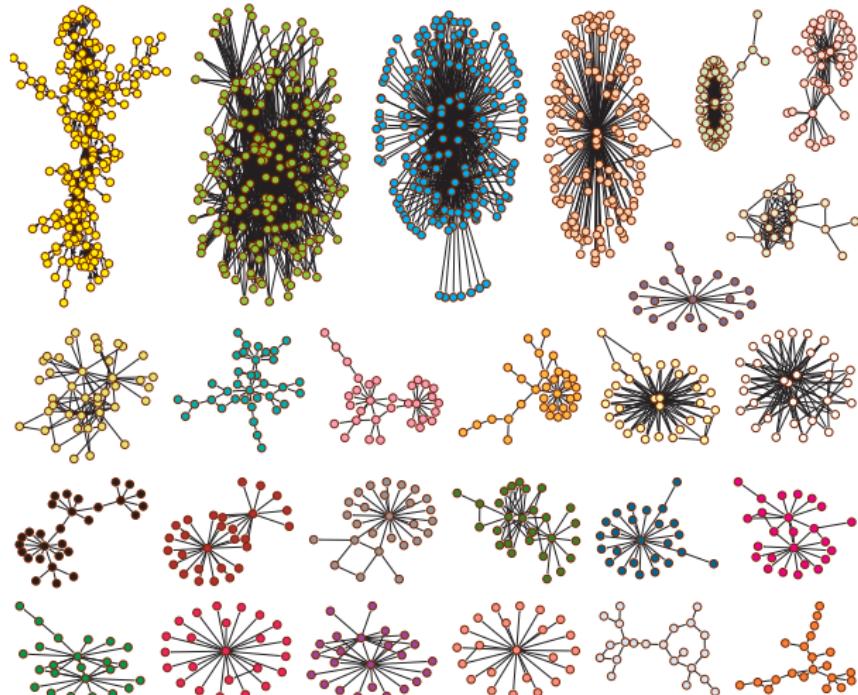
Multiplication

Derived Ns

Temporal Ns

References

Appendix





PEERE – SPC – Link island1

$$w_{max} = 0.297$$

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

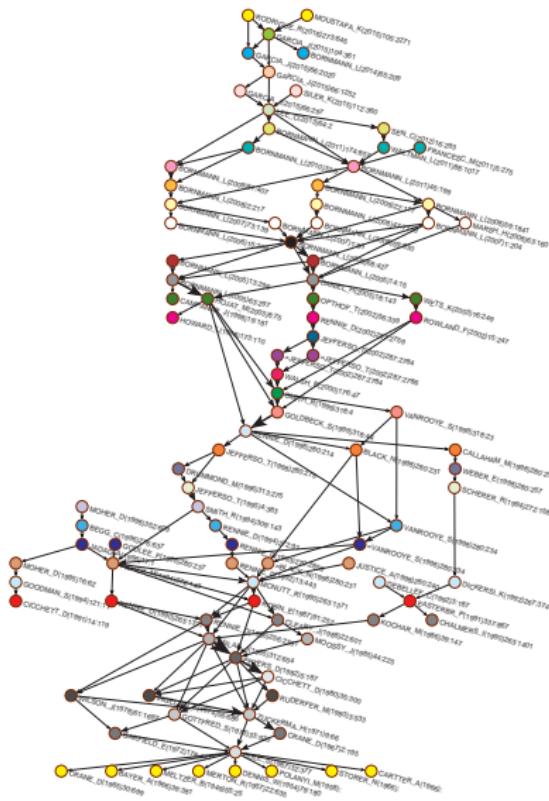
Multiplication

Derived Ns

Temporal Ns

References

Appendix



This island is very similar to the main paths for 100 largest weights and includes main path.



Islands positioning

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

A recent extension of the main paths approach

Network/Acyclic network/Create (sub)network/Main paths/
Local search/Through vertices in Cluster

Network/Acyclic network/Create (sub)network/Main paths/
Global search/Through vertices in Cluster

enables a researcher to determine main paths through a given set of nodes (works) in a citation network. This can be used to position a given set of nodes in a citation network – it can, either attach to the principal main paths, or form a separate stream.

In the literature on network clustering we considered works on valued networks, { ZIBERNA_A(2007)29:105, NORDLUND_C(2007)29:59, ZIBERNA_A (2008)32:57, ZIBERNA_A(2009)6:99, ZIBERNA_A(2013):, ZIBERNA_A (2013)10:99, ZIBERNA_A(2014)39:46, ZIBERNA_A(2016)12:137, NORDLUND_C(2016)44:160 }, as shown in figure in the following slide.



Islands

positioning – valued networks in clustering

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

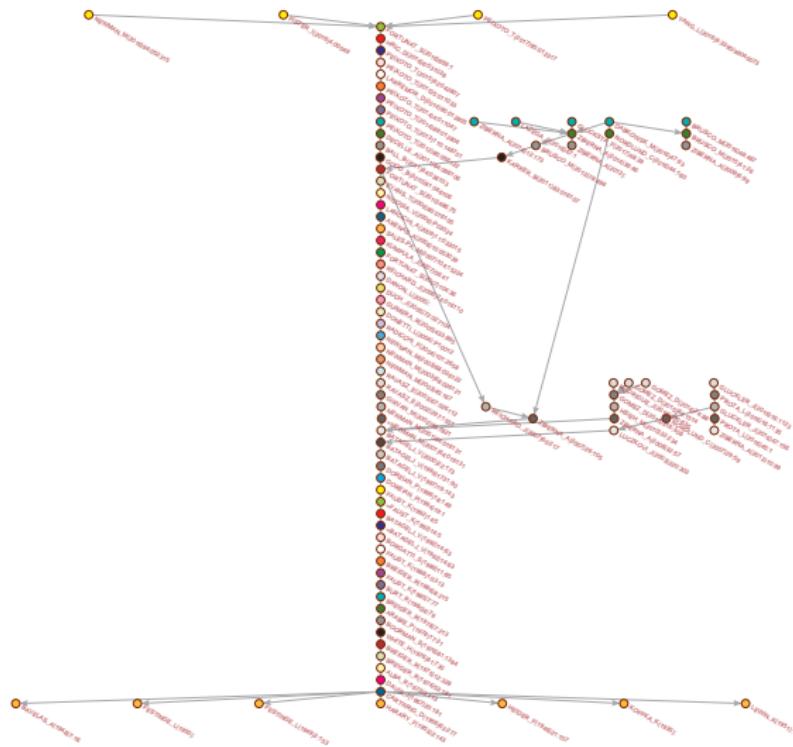
Multiplication

Derived Ns

Temporal Ns

References

Appendix





Two-mode networks

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

In a *two-mode* network $\mathcal{N} = ((\mathcal{U}, \mathcal{V}), \mathcal{L}, \mathcal{P}, \mathcal{W})$ the set of nodes consists of two disjoint sets of nodes \mathcal{U} and \mathcal{V} , and all the links from \mathcal{L} have one end-node in \mathcal{U} and the other in \mathcal{V} . Often also a *weight* $w : \mathcal{L} \rightarrow \mathbb{R} \in \mathcal{W}$ is given; if not, we assume $w(u, v) = 1$ for all $(u, v) \in \mathcal{L}$.

A two-mode network can also be described by a rectangular matrix $\mathbf{A} = [a_{uv}]_{\mathcal{U} \times \mathcal{V}}$.

$$a_{uv} = \begin{cases} w_{uv} & (u, v) \in \mathcal{L} \\ 0 & \text{otherwise} \end{cases}$$

Examples: (persons, societies, years of membership),
(buyers/consumers, goods, quantity),
(parliamentarians, problems, positive vote),
(persons, journals, reading),
(papers, keywords, is described by), etc.



Deep South

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix



Classical example of two-mode network are the Southern women (Davis 1941).

Davis.paj. Freeman's overview.

NAME OF PARTICIPANTS OR GROUP I	CODE NUMBERS AND DATES OF SOCIAL EVENTS REPORTED IN <i>Old City Herald</i>													
	(1) 6/27	(2) 3/2	(3) 4/12	(4) 9/16	(5) 3/25	(6) 5/19	(7) 3/25	(8) 9/16	(9) 4/6	(10) 6/10	(11) 3/23	(12) 4/7	(13) 11/21	(14) 8/3
1. Mrs. Evelyn Jefferson.....	X	X	X	X	X	X	X	X
2. Miss Laura Mandeville.....	X	X	X	X	X	X	X	X
3. Miss Theresa Anderson.....		X	X	X	X	X	X	X	X
4. Miss Brenda Rogers.....			X	X	X	X	X	X	X
5. Miss Charlotte McDowell.....	X			X	X	X	X	X	X
6. Miss Frances Anderson.....			X	X	X	X	X	X	X
7. Miss Eleanor Nye.....				X	X	X	X	X	X
8. Miss Pearl Oglethorpe.....					X	X	X	X	X
9. Miss Ruth DeSand.....						X	X	X	X
10. Miss Verne Sanderson.....							X	X	X	X	X
11. Miss Myra Liddell.....								X	X	X	X	X
12. Miss Katherine Rogers.....									X	X	X	X	X
13. Mrs. Sylvia Avondale.....										X	X	X	X	X
14. Mrs. Norm Fayette.....										X	X	X	X	X
15. Mrs. Helen Lloyd.....										X	X	X	X	X
16. Mrs. Dorothy Murchison.....											X	X
17. Mrs. Olivia Carleton.....											X	X
18. Mrs. Flora Price.....											X	X



Approaches to two-mode network analysis

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

The usual approach to analyze a two-mode network is to transform it to a one-mode network and use standard methods on it.

For direct analysis of two-mode networks we can use the *eigen-vector approach* – a two-mode variant of Kleinberg's hubs and authorities. The weight vector (\mathbf{x}, \mathbf{y}) on $\mathcal{U} \cup \mathcal{V}$ is determined by relations $\mathbf{y} = \mathbf{Ax}$ and $\mathbf{x} = \mathbf{A}^T \mathbf{y}$.

Network/2-Mode Network/Important Vertices

There are also special methods for *clustering* and *blockmodeling* in two-mode networks.

In this lecture we will present two additional direct methods: *two-mode cores* and *4-rings*.



Internet Movie Database <http://www.imdb.com/>

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

12th Annual Graph Drawing Contest, 2005. The IMDb network is two-mode and has $1324748 = 428440 + 896308$ nodes and 3792390 arcs.



Two-mode cores

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

The subset of nodes $C \subseteq \mathcal{V}$ is a **(p, q) -core** in a two-mode network $\mathcal{N} = (\mathcal{V}_1, \mathcal{V}_2; \mathcal{L})$, $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$ iff

- a. in the induced subnetwork $\mathcal{K} = (C_1, C_2; \mathcal{L}(C))$,
 $C_1 = C \cap \mathcal{V}_1$, $C_2 = C \cap \mathcal{V}_2$ it holds $\forall v \in C_1 : \deg_{\mathcal{K}}(v) \geq p$
and $\forall v \in C_2 : \deg_{\mathcal{K}}(v) \geq q$;
- b. C is the maximal subset of \mathcal{V} satisfying condition a.

Properties of two-mode cores:

- $C(0, 0) = \mathcal{V}$
- $\mathcal{K}(p, q)$ is not always connected
- $(p_1 \leq p_2) \wedge (q_1 \leq q_2) \Rightarrow C(p_1, q_1) \subseteq C(p_2, q_2)$
- $\mathcal{C} = \{C(p, q) : p, q \in \mathbb{N}\}$. If all nonempty elements of \mathcal{C} are different it is a lattice.



Algorithm for two-mode cores

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

To determine a (p, q) -core the procedure similar to the ordinary core procedure can be used:

repeat

remove from the first set all nodes of degree less than p ,

and from the second set all nodes of degree less than q

until no node was deleted

It can be implemented to run in $O(m)$ time.

Interesting (p, q) -cores? Table of cores' characteristics

$n_1 = |C_1(p, q)|$, $n_2 = |C_2(p, q)|$ and k – number of components in $\mathcal{K}(p, q)$:

- $n_1 + n_2 \leq$ selected threshold
- 'border line' in the (p, q) -table.



Table ($p, q : n_1, n_2$) for Internet Movie Database

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Network/2-Mode Network/Core/2-Mode Border

1	1590:	1590	1	16	39:	2173	678	44	14:	29	83
2	516:	788	3	17	35:	2791	995	46	13:	29	94
3	212:	1705	18	18	32:	2684	1080	49	12:	26	95
4	151:	4330	154	19	30:	2395	1063	52	11:	16	79
5	131:	4282	209	20	28:	2216	1087	56	10:	34	162
6	115:	3635	223	21	26:	1988	1087	62	9:	31	177
7	101:	3224	244	22	24:	1854	1153	66	8:	29	198
8	88:	2860	263	24	23:	34	39	72	7:	22	203
9	77:	3467	393	27	22:	31	38	96	6:	7	114
10	69:	3150	428	29	20:	35	52	119	5:	6	137
11	63:	2442	382	32	19:	34	57	141	4:	8	258
12	56:	2479	454	35	18:	33	61	186	3:	3	186
13	50:	3330	716	36	17:	33	65	247	2:	2	247
14	46:	2460	596	39	16:	29	70	1334	1:	1	1334
15	42:	2663	739	42	15:	28	76				



(247,2)-core and (27,22)-core

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

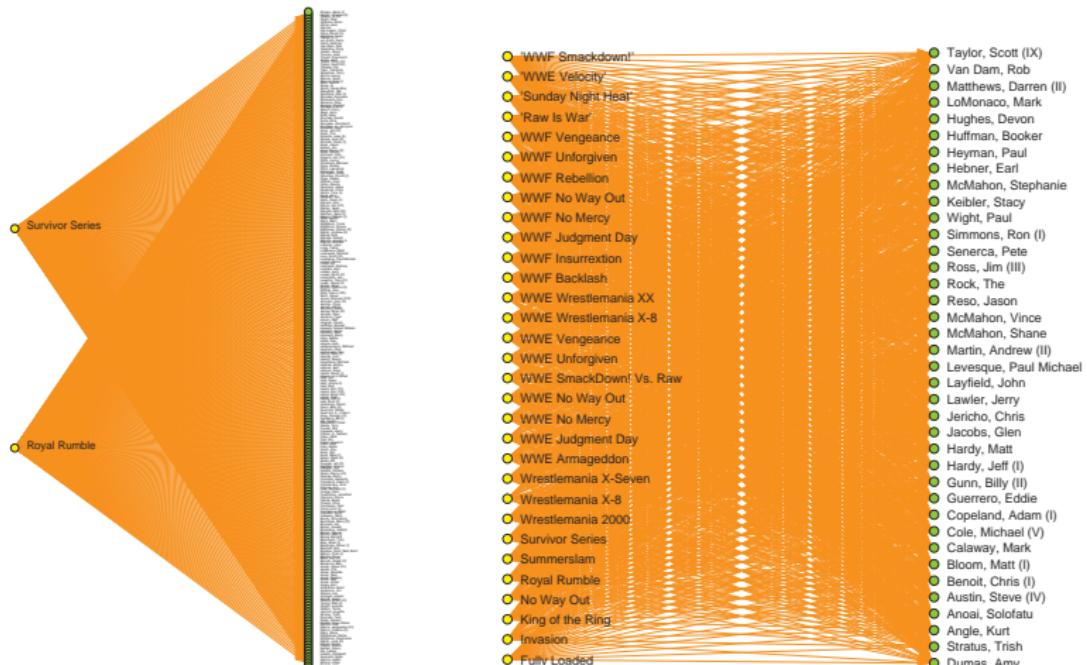
Multiplication

Derived Ns

Temporal Ns

References

Appendix





(2,516)-Hard core

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

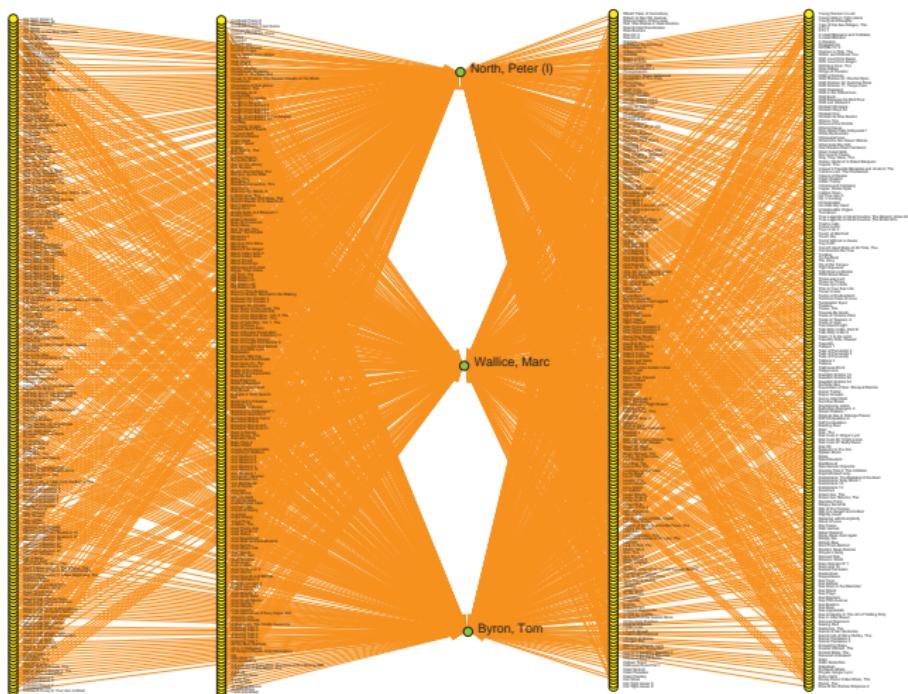
Multiplication

Derived Ns

Temporal Ns

References

Appendix



4-rings and analysis of two-mode networks

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

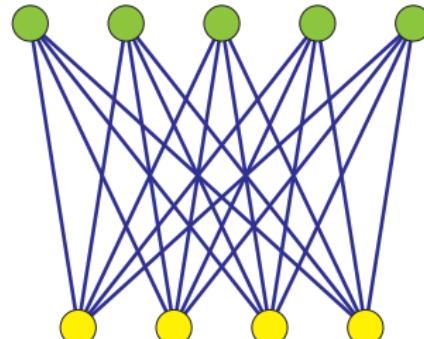
References

Appendix

A *k-ring* is a simple closed chain of length k . Using k -rings we can define a weight of edges as

$$w_k(e) = \# \text{ of different } k\text{-rings containing the edge } e \in \mathcal{E}$$

In two-mode network there are no 3-rings. The densest substructures are complete bipartite subgraphs $K_{p,q}$. They contain many 4-rings.



There are

$$\binom{p}{2} \binom{q}{2} = \frac{1}{4} p(p-1)q(q-1)$$

4-rings in $K_{p,q}$; and each of its edges e has weight

$$w_4(e) = (p-1)(q-1)$$

Network/Create New Network/with Ring Counts.../4-Rings/Undirected

Directed 4-rings

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

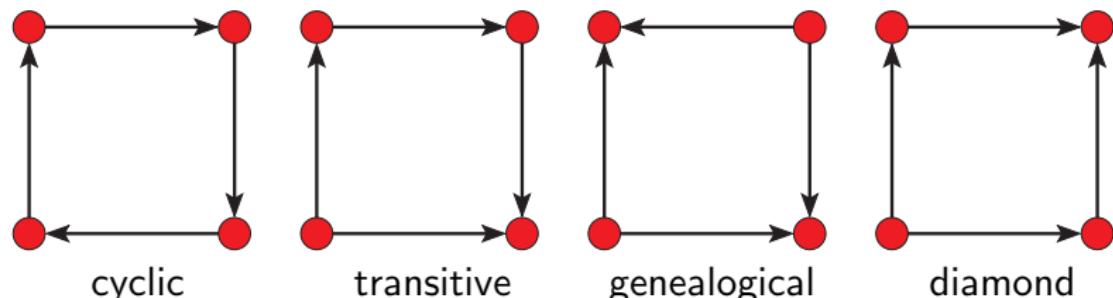
Derived Ns

Temporal Ns

References

Appendix

There are 4 types of directed 4-rings:



In the case of transitive rings Pajek provides a special weight counting on how many transitive rings the arc is a *shortcut*.

Network/Create New Network/with Ring Counts/4-Rings/Directed



Simple link islands in IMDb for w_4

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

We obtained 12465 simple link islands on 56086 nodes. Here is their size distribution.

	Size	Freq									
	2	5512		20	19		38	4		59	2
	3	1978		21	18		39	3		61	1
	4	1639		22	15		40	2		64	1
	5	968		23	9		42	2		67	1
	6	666		24	13		43	3		70	1
	7	394		25	12		45	3		73	1
	8	257		26	6		46	4		76	1
	9	209		27	6		47	5		82	1
	10	148		28	5		48	1		86	1
	11	118		29	6		49	2		106	1
	12	87		30	3		50	2		122	1
	13	55		31	6		51	1		135	1
	14	62		32	5		52	2		144	1
	15	46		33	3		53	1		163	1
	16	39		34	1		54	2		269	1
	17	27		35	5		55	1		301	1
	18	28		36	4		57	1		332	2
	19	29		37	7		58	1		673	1



Example: Islands for w_4 Charlie Brown and Adult

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

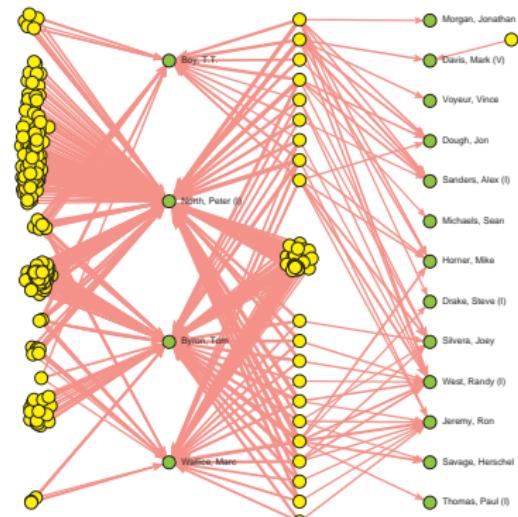
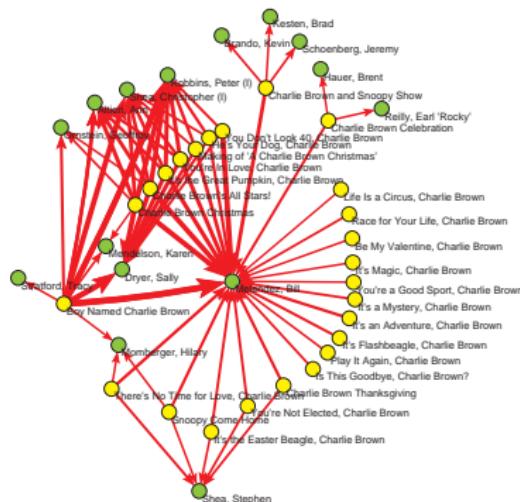
Multiplication

Derived Ns

Temporal Ns

References

Appendix





Example: Islands for w_4

Mark Twain and Abid

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

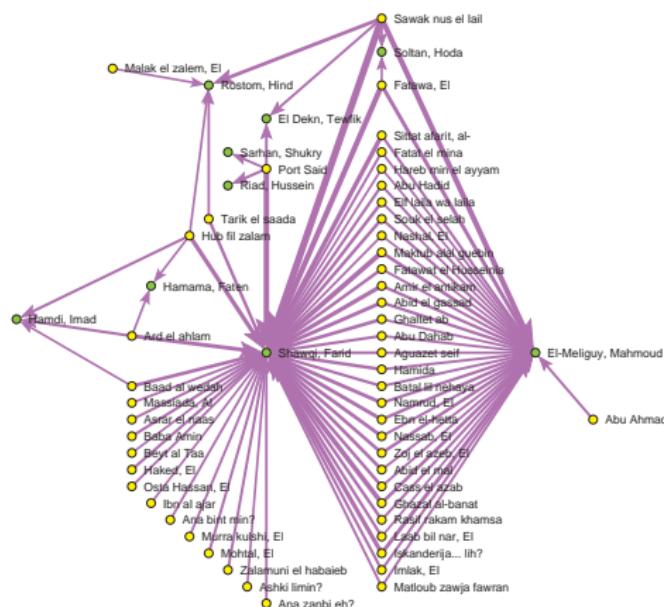
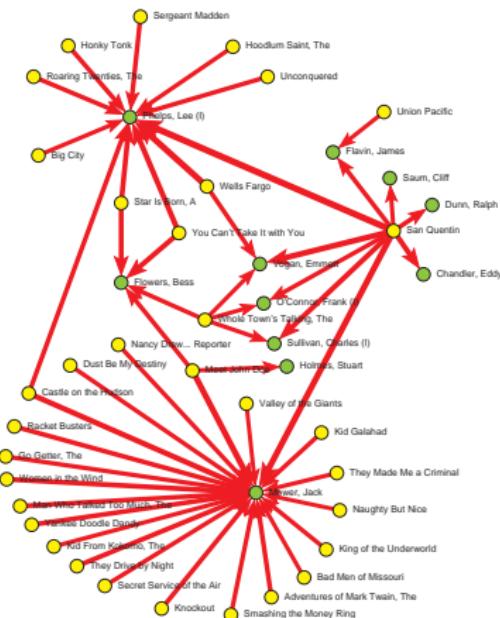
Multiplication

Derived Ns

Temporal Ns

References

Appendix



Example: Island for w_4

Polizeiruf 110 and Starkes Team

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

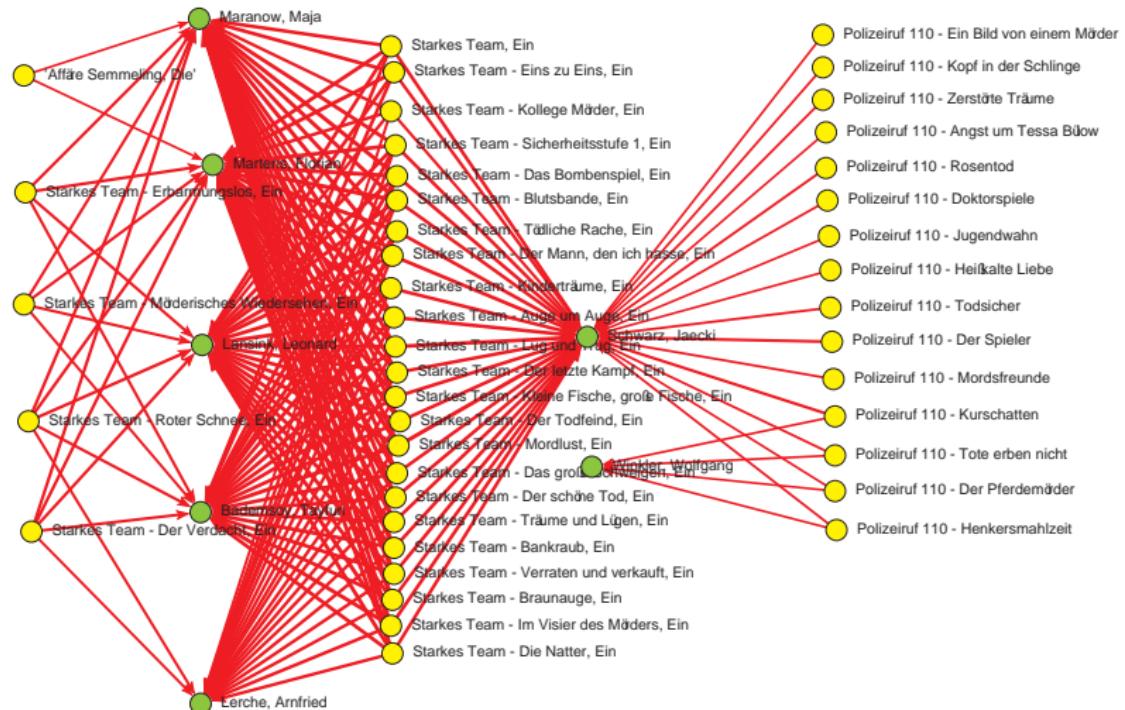
Multiplication

Derived Ns

Temporal Ns

References

Appendix





Multiplication of networks

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

To a simple (no parallel arcs) two-mode **network** $\mathcal{N} = (\mathcal{I}, \mathcal{J}, \mathcal{A}, w)$; where \mathcal{I} and \mathcal{J} are sets of **nodes**, \mathcal{A} is a set of **arcs** linking \mathcal{I} and \mathcal{J} , and $w : \mathcal{A} \rightarrow \mathbb{R}$ (or some other semiring) is a **weight**; we can assign a **network matrix** $\mathbf{W} = [w_{i,j}]$ with elements: $w_{i,j} = w(i,j)$ for $(i,j) \in \mathcal{A}$ and $w_{i,j} = 0$ otherwise.

Given a pair of compatible networks $\mathcal{N}_A = (\mathcal{I}, \mathcal{K}, \mathcal{A}_A, w_A)$ and $\mathcal{N}_B = (\mathcal{K}, \mathcal{J}, \mathcal{A}_B, w_B)$ with corresponding matrices $\mathbf{A}_{\mathcal{I} \times \mathcal{K}}$ and $\mathbf{B}_{\mathcal{K} \times \mathcal{J}}$ we call a **product of networks** \mathcal{N}_A and \mathcal{N}_B a network $\mathcal{N}_C = (\mathcal{I}, \mathcal{J}, \mathcal{A}_C, w_C)$, where $\mathcal{A}_C = \{(i,j) : i \in \mathcal{I}, j \in \mathcal{J}, c_{i,j} \neq 0\}$ and $w_C(i,j) = c_{i,j}$ for $(i,j) \in \mathcal{A}_C$. The product matrix $\mathbf{C} = [c_{i,j}]_{\mathcal{I} \times \mathcal{J}} = \mathbf{A} * \mathbf{B}$ is defined in the standard way

$$c_{i,j} = \sum_{k \in \mathcal{K}} a_{i,k} \cdot b_{k,j}$$

In the case when $\mathcal{I} = \mathcal{K} = \mathcal{J}$ we are dealing with ordinary one-mode networks (with square matrices).



Multiplication of networks

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

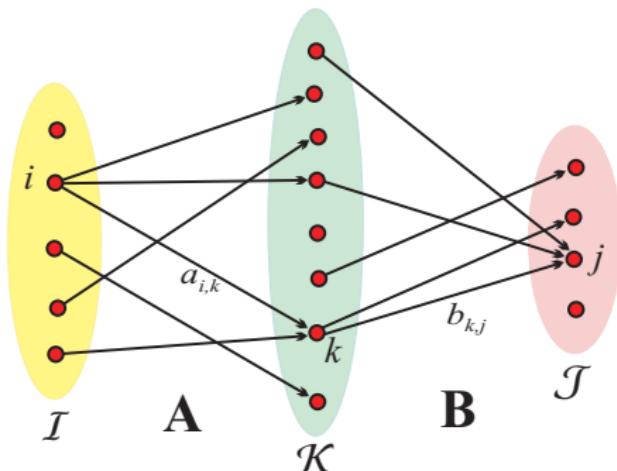
Multiplication

Derived Ns

Temporal Ns

References

Appendix



$$c_{i,j} = \sum_{k \in N_A(i) \cap N_B^-(j)} a_{i,k} \cdot b_{k,j}$$

If all weights in networks \mathcal{N}_A and \mathcal{N}_B are equal to 1 the value of $c_{i,j}$ counts the number of ways we can go from $i \in \mathcal{I}$ to $j \in \mathcal{J}$ passing through \mathcal{K} .



Multiplication of networks

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

The standard matrix multiplication has the complexity $O(|\mathcal{I}| \cdot |\mathcal{K}| \cdot |\mathcal{J}|)$ – it is too slow to be used for large networks. For sparse large networks we can multiply much faster considering only nonzero elements.
In general the multiplication of large sparse networks is a 'dangerous' operation since the result can 'explode' – it is not sparse.

If at least one of the sparse networks \mathcal{N}_A and \mathcal{N}_B has small maximal degree on \mathcal{K} then also the resulting product network \mathcal{N}_C is sparse.

If for the sparse networks \mathcal{N}_A and \mathcal{N}_B there are in \mathcal{K} only few nodes with large degree and no one among them with large degree in both networks then also the resulting product network \mathcal{N}_C is sparse.



Multiplication of networks – details

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

From the network multiplication algorithm we see that each intermediate node $k \in \mathcal{K}$ adds to a product network a complete two-mode subgraph $K_{N_A^-(k), N_B(k)}$ (or, in the case $\mathcal{I} = \mathcal{J}$, a complete subgraph $K_{N(k)}$). If both degrees $\deg_A(k) = |N_A^-(k)|$ and $\deg_B(k) = |N_B(k)|$ are large then already the computation of this complete subgraph has a quadratic (time and space) complexity – the result 'explodes'.

For more details see the [paper](#).



Two-mode network analysis by conversion to one-mode network – projections

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Often we transform a two-mode network $\mathcal{N} = (\mathcal{U}, \mathcal{V}, \mathcal{E}, w)$ into an ordinary (one-mode) network $\mathcal{N}_1 = (\mathcal{U}, \mathcal{E}_1, w_1)$ or/and $\mathcal{N}_2 = (\mathcal{V}, \mathcal{E}_2, w_2)$, where \mathcal{E}_1 and w_1 are determined by the matrix $\mathbf{W}^{(1)} = \mathbf{WW}^T$, $w_{uv}^{(1)} = \sum_{z \in \mathcal{V}} w_{uz} \cdot w_{zv}^T$. Evidently $w_{uv}^{(1)} = w_{vu}^{(1)}$. There is an edge $(u : v) \in \mathcal{E}_1$ in \mathcal{N}_1 iff $N(u) \cap N(v) \neq \emptyset$. Its weight is $w_1(u, v) = w_{uv}^{(1)}$.

The network \mathcal{N}_2 is determined in a similar way by the matrix $\mathbf{W}^{(2)} = \mathbf{W}^T \mathbf{W}$.

The networks \mathcal{N}_1 and \mathcal{N}_2 are analyzed using standard methods.

Network/2-Mode Network/2-Mode to 1-Mode/Rows



Authorship networks

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Let **WA** be the works \times authors two mode authorship network;
 $wa_{pi} \in \{0, 1\}$ is describing the authorship of author i of work p .

$$\forall p \in W : \sum_{i \in A} wa_{pi} = \text{outdeg}_{WA}(p) = \# \text{ authors of work } p$$

Let **N** be its normalized version

$$\forall p \in W : \sum_{i \in A} n_{pi} \in \{0, 1\}$$

obtained from **WA** by $n_{pi} = wa_{pi} / \max(1, \text{outdeg}_{WA}(p))$, or by some other rule determining the author's contribution – the *fractional* approach.



Some transformations of networks

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Binarization $b(\mathcal{N})$ is a network obtained from the \mathcal{N} in which all weights are set to 1.

Transposition \mathcal{N}^T or $t(\mathcal{N})$ is a network obtained from \mathcal{N} in which to all arcs their direction is reversed. $\mathbf{AW} = \mathbf{WA}^T$, $\mathbf{KW} = \mathbf{WK}^T$, ...

(Out) normalization $n(\mathcal{N})$ is a network obtained from \mathcal{N} in which the weight of each arc a is divided by the sum of weights of all arcs having the same initial node as the arc a . For binary networks

$$n(\mathbf{A}) = \text{diag}\left(\frac{1}{\max(1, \text{outdeg}_{WA}(i))}\right)_{i \in \mathcal{I}} * \mathbf{A}$$

$$\mathbf{N} = n(\mathbf{WA}), \mathbf{WA} = b(\mathbf{N})$$



First co-authorship network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

$$\mathbf{Co} = \mathbf{AW} * \mathbf{WA}$$

$$co_{ij} = \sum_{p \in W} wa_{pi} wa_{pj} = \sum_{p \in N^-(i) \cap N^-(j)} 1$$

co_{ij} = the number of works that authors i and j wrote together

co_{ii} = the total number of works that author i wrote

It holds: $co_{ij} = co_{ji}$.

Using the weights co_{ij} we can determine the Salton's cosine similarity or Ochiai coefficient between authors i and j as

$$\cos(i, j) = \frac{co_{ij}}{\sqrt{co_{ii} co_{jj}}}, \quad \text{for } co_{ij} > 0$$



Cores of orders 20–47 in $\text{Co}(\text{SN5})$

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

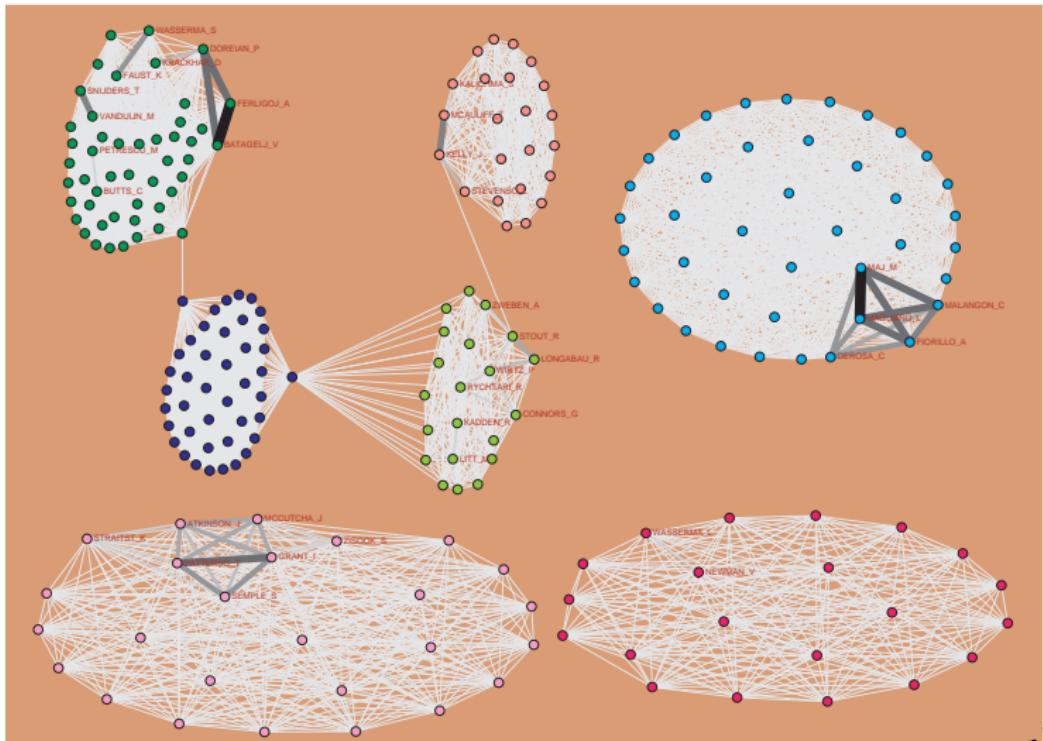
Derived Ns

Temporal Ns

References

Appendix

Network SN5 (2008): for "social network*" + most frequent references + around 100 social networkers;
 $|W| = 193376, |C| = 7950, |A| = 75930, |J| = 14651, |K| = 29267$





Papers by number of authors

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Problem: The **Co** network is composed of complete graphs on the set of work's authors. Works with many authors produce large complete subgraphs and are over-represented, thus blurring the collaboration structure.

outdeg	frequency	outdeg	frequency	paper
1	2637	12	8	
2	2143	13	4	
3	1333	14	3	
4	713	15	2	
5	396	21	1	Pierce et al. (2007)
6	206	22	1	Allen et al. (1998)
7	114	23	1	Kelly et al. (1997)
8	65	26	1	Semple et al. (1993)
9	43	41	1	Magliano et al. (2006)
10	24	42	1	Doll et al. (1992)
11	10	48	1	Snijders et al. (2007)



Snijders et al. (2007)

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Snijders et al.(2007): Snijders, T.A.B., Robinson, T., Atkinson, A.C., Riani, M., Gormley, I.C., Murphy, T.B., Sweeting, T., Leslie, D.S., Longford, N.T., Kent, J.T., Lawrence, T., Airoldi, E.M., Besag, J., Blei, D., Fienberg, S.E., Breiger, R., Butts, C.T., Doreian, P., Batagelj, V., Ferligoj, A., Draper, D., van Duijn, M.A.J., Faust, K., Petrescu-Prahova, M., Forster, J.J., Gelman, A., Goodreau, S. M., Greenwood, P.E., Gruenberg, K., Francis, B., Hennig, C., Hoff, P.D., Hunter, D.R., Husmeier, D., Glasbey, C., Krackhardt, D., Kuha, J., Skrondal, A., Lawson, A., Liao, T. F., Mendes, B., Reinert, G., Richardson, S., Lewin, A., Titterington, D.M., Wasserman, S., Werhli, A.V. and Ghazal, P.. *Discussion on the paper by Handcock, Raftery and Tantrum.* Journal of the Royal Statistical Society: Series A - Statistics in Society, 170 (2007), pp. 322-354.



p_S -core at level 20 of **Co(SN5)**

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

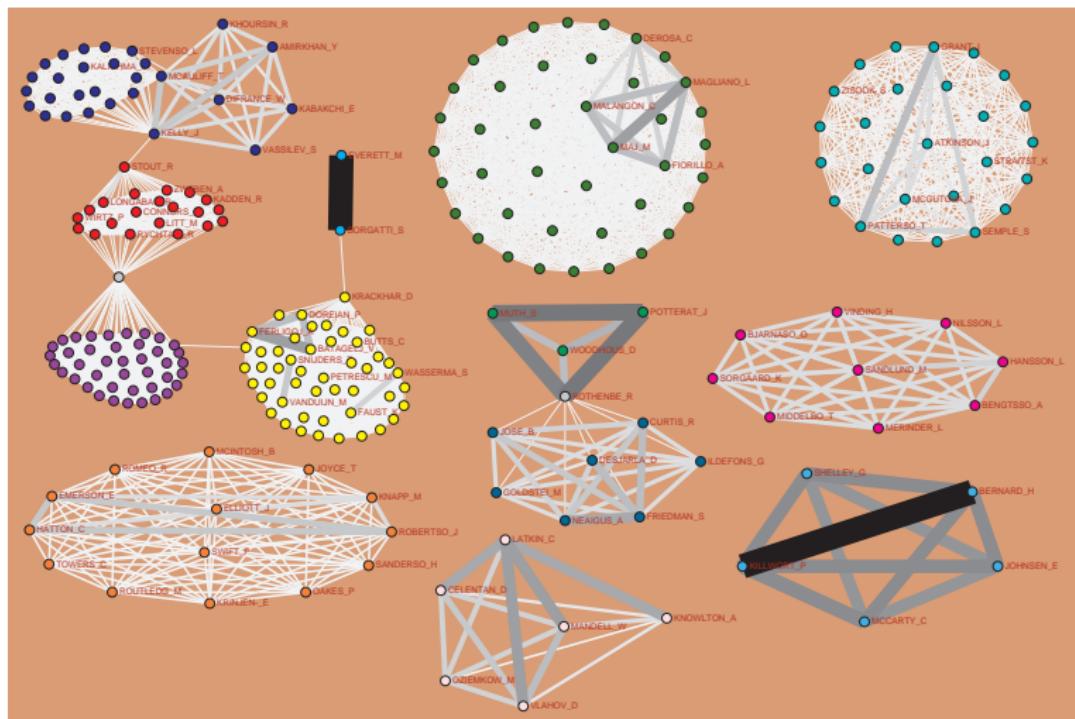
Multiplication

Derived Ns

Temporal Ns

References

Appendix





Second co-authorship network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

$$\mathbf{Cn} = \mathbf{AW} * \mathbf{N}$$

$$cn_{ij} = \sum_{p \in W} wa_{pi} n_{pj} = \sum_{p \in N^-(i) \cap N^-(j)} n_{pj}$$

cn_{ij} = contribution of author j to works, that (s)he wrote together with the author i .

It holds $\sum_{j \in A} \sum_{p \in A} wa_{pi} n_{pj} = \text{outdeg}_{WA}(p)$ and $\sum_{j \in A} cn_{ij} = \text{indeg}_{WA}(i)$

$cn_{ii} = \sum_{p \in N(i)} n_{pi}$ is the contribution of author i to his/her works.

Self-sufficiency: $S_i = \frac{cn_{ii}}{\text{indeg}_{WA}(i)}$

Collaborativeness: $K_i = 1 - S_i$

$$\sum_{i \in A} \sum_{j \in A} cn_{ij} = \sum_{i \in A} \text{indeg}_{WA}(i) = m_{WA}$$

To compute the table we prepared a macro in Pajek.



The "best" authors in Social Networks

Bibliographic networks

	i	author	cn_{ii}	total	K_i	i	author	cn_{ii}	total	K_i
V. Batagelj, D. Maltseva	1	Burt,R	43.83	53	0.173	26	Latkin,C	10.14	37	0.726
	2	Newman,M	36.77	60	0.387	27	Morris,M	9.98	20	0.501
Networks	3	Doreian,P	34.44	47	0.267	28	Rothenberg,R	9.82	28	0.649
Statistics	4	Bonacich,P	30.17	41	0.264	29	Kadushin,C	9.75	11	0.114
Citation	5	Marsden,P	29.42	37	0.205	30	Faust,K	9.72	18	0.460
Two-mode Ns	6	Wellman,B	26.87	41	0.345	31	Batagelj,V	9.69	20	0.516
Multiplication	7	Leydesdorff,L	24.37	35	0.304	32	Mizruchi,M	9.67	15	0.356
Derived Ns	8	White,H	23.50	33	0.288	33	[Anon]	9.00	9	0.000
Temporal Ns	9	Friedkin,N	20.00	23	0.130	34	Johnson,J	8.89	21	0.577
References	10	Borgatti,S	19.20	41	0.532	35	Fararo,T	8.83	16	0.448
Appendix	11	Everett,M	16.92	31	0.454	36	Lazega,E	8.50	12	0.292
	12	Litwin,H	16.00	21	0.238	37	Knoke,D	8.33	11	0.242
	13	Freeman,L	15.53	20	0.223	38	Ferligoj,A	8.19	19	0.569
	14	Barabasi,A	14.99	35	0.572	39	Brewer,D	8.03	11	0.270
	15	Snijders,T	14.99	30	0.500	40	Klov Dahl,A	7.96	17	0.532
	16	Valente,T	14.80	34	0.565	41	Hammer,M	7.92	10	0.208
	17	Breiger,R	14.44	20	0.278	42	White,D	7.83	15	0.478
	18	Skvoretz,J	14.43	27	0.466	43	Holme,P	7.42	14	0.470
	19	Krackhardt,D	13.65	25	0.454	44	Boyd,J	7.37	13	0.433
	20	Carley,K	12.93	28	0.538	45	Kilduff,M	7.25	16	0.547
	21	Pattison,P	12.10	27	0.552	46	Small,H	7.00	7	0.000
	22	Wasserman,S	11.72	26	0.549	47	Iacobucci,D	7.00	12	0.417
	23	Berkman,L	11.21	30	0.626	48	Pappi,F	6.83	10	0.317
	24	Moody,J	10.83	15	0.278	49	Chen,C	6.78	12	0.435
	25	Scott,J	10.47	15	0.302	50	Seidman,S	6.75	9	0.250



Third co-authorship network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

$$\mathbf{Ct} = \mathbf{N}^T * \mathbf{N}$$

ct_{ij} = the total contribution of ‘collaboration’ of authors i and j to works.

It holds $ct_{ij} = ct_{ji}$ and

$$\sum_{i \in A} \sum_{j \in A} n_{pi} n_{pj} = 1$$

The total contribution of a complete subgraph corresponding to the authors of a work p is 1.

$\sum_{j \in A} ct_{ij} = \sum_{p \in W} n_{pi}$ = the total contribution of author i to works from W .

$$\sum_{i \in A} \sum_{j \in A} ct_{ij} = |W|$$



Components in $\text{Ct}(\text{SN5})$ cut at level 0.5

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

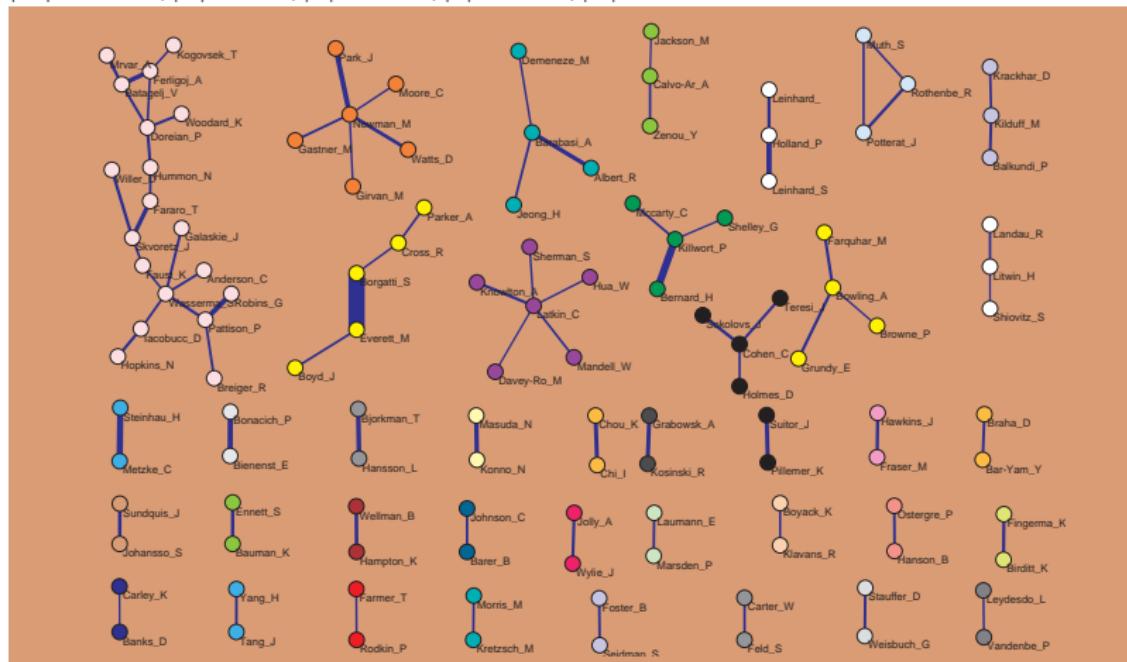
Derived Ns

Temporal Ns

References

Appendix

Network SN5 (2008): for "social network*" + most frequent references + around 100 social networkers;
 $|W| = 193376, |C| = 7950, |A| = 75930, |J| = 14651, |K| = 29267$





p_S -core at level 0.75 in $\text{Ct}(\text{SN}_5)$

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

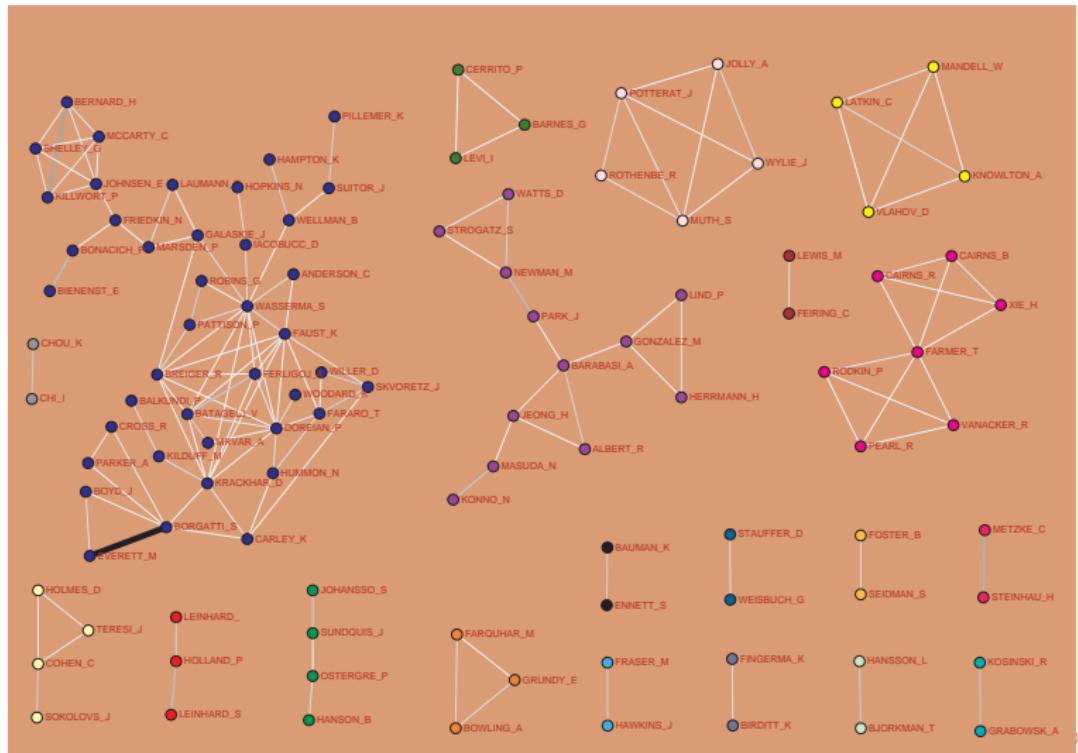
Multiplication

Derived Ns

Temporal Ns

References

Appendix





Some link islands [5,20] in $\mathbf{Ct}(\mathbf{SN5})$

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

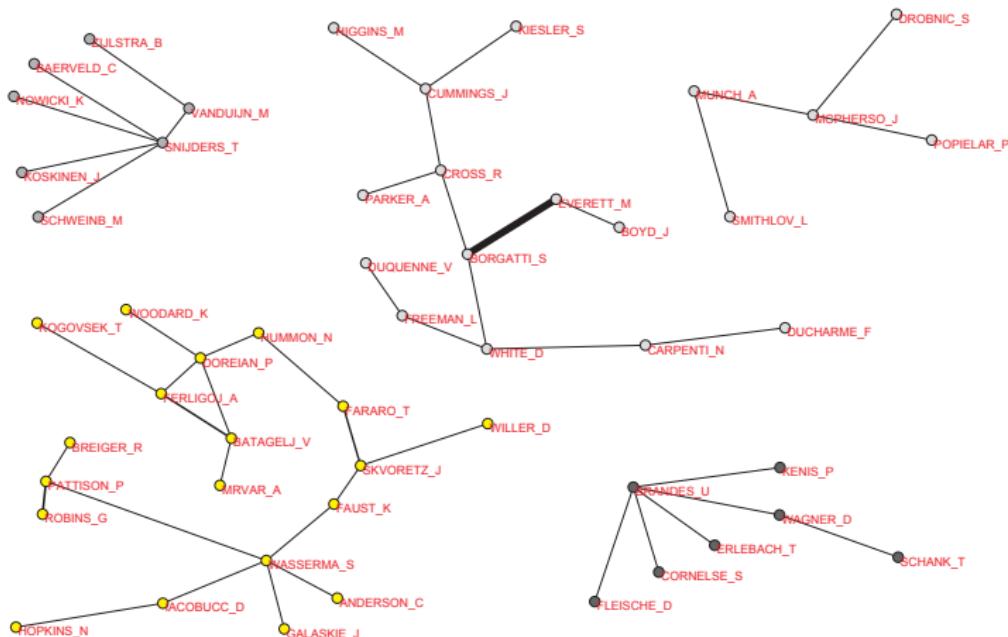
Multiplication

Derived Ns

Temporal Ns

References

Appendix





Fourth co-authorship network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

$$\mathbf{Ct}' = \mathbf{N}^T * \mathbf{N}', \text{ where } n'_{pi} = wa_{pi} / \max(1, \text{outdeg}_{WA}(p) - 1)$$

ct'_{ij} = the total contribution of 'strict collaboration' of authors i and j to works.

In Pajek we can use macros to save sequences of commands to produce different co-authorship networks.

The final result is returned as an undirected simple network with weights (for $i \neq j$)

$$ct'_{ij} = \sum_p \frac{2 \cdot wa_{pi} \cdot wa_{pj}}{\max(1, \text{outdeg}_{WA}(p)) \cdot \max(1, \text{outdeg}_{WA}(p) - 1)}$$



Authors' citations network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

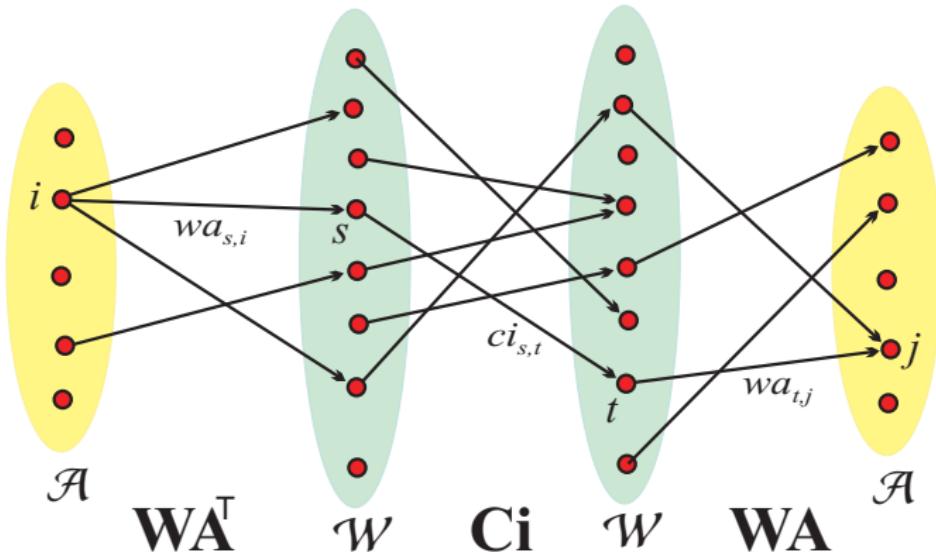
Multiplication

Derived Ns

Temporal Ns

References

Appendix



Ca = AW * Ci * WA is a network of citations between authors.
The weight $w(i,j)$ counts the number of times a work authored by i is citing a work authored by j .



Islands in SN5 authors citation network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

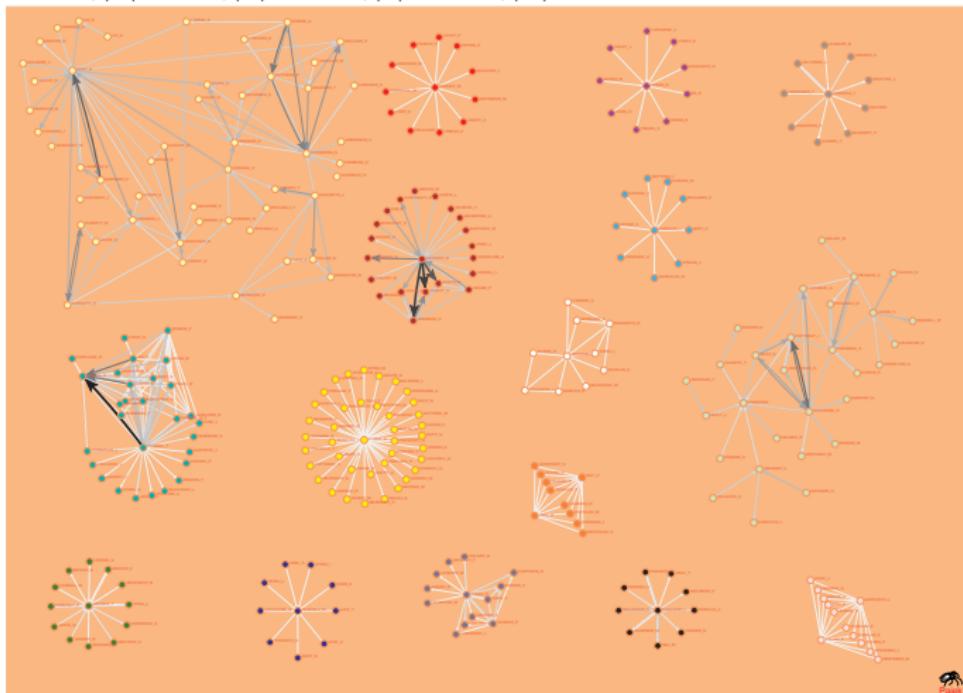
Derived Ns

Temporal Ns

References

Appendix

Network SN5 (2008): for "social network*" + most frequent references + around 100 social networkers;
 $|W| = 193376, |C| = 7950, |A| = 75930, |J| = 14651, |K| = 29267$





Bibliographic Coupling

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

In WoS2Pajek the citation relation means $p\mathbf{Ci}q \equiv$ work p cites work q . Therefore the **bibliographic coupling** network **biCo** can be determined as

$$\mathbf{biCo} = \mathbf{Ci} * \mathbf{Ci}^T$$

$bico_{pq} = \# \text{ of works cited by both works } p \text{ and } q$. $bico_{pq} = bico_{qp}$.

Again we have problems with works with many citations, especially with review papers. To neutralize their impact we can introduce a normalized measure such as

$$\mathbf{biCon} = \frac{1}{2}(n(\mathbf{Ci}) * \mathbf{Ci}^T + \mathbf{Ci} * n(\mathbf{Ci})^T)$$

It is easy to verify that $bicon_{pq} \in [0, 1]$ and $bicon_{pq} = bicon_{qp}$ (symmetry). It also holds: $bicon_{pq} = 1$ iff the works p and q are referencing the same works.



Co-Citation and others

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

The *co-citation* network **coCi** can be determined as

$$\mathbf{coCi} = \mathbf{Ci}^T * \mathbf{Ci}$$

$coci_{pq} = \#$ of works citing both works p and q .

$coci_{pq} = cocici_{qp}$.

The weight $w(a, p)$ in the *author citation* network

$$\mathbf{ACi} = \mathbf{AW} * \mathbf{Ci}$$

counts the number of times author a cited work p .

The *author co-citation* network can be obtained as

$$\mathbf{ACo} = b(\mathbf{ACi}) * t(b(\mathbf{ACi}))$$

Authors using keywords **AK** = **AW** * **WK**.



Temporal quantities

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

We introduce a notion of a *temporal quantity*

$$a(t) = \begin{cases} a'(t) & t \in T_a \\ \text{\#} & t \in \mathcal{T} \setminus T_a \end{cases}$$

where T_a is the *activity time set* of a and $a'(t)$ is the value of a in an instant $t \in T_a$, and \# denotes the value *undefined*.

We assume that the values of temporal quantities belong to a set A which is a *semiring* $(A, +, \cdot, 0, 1)$ for binary operations $+ : A \times A \rightarrow A$ and $\cdot : A \times A \rightarrow A$.

Let $A_{\text{\#}}(\mathcal{T})$ denote the set of all temporal quantities over $A_{\text{\#}}$ in time \mathcal{T} . To extend the operations to networks and their matrices we first define the *sum* (parallel links) $a + b$ as

$$(a + b)(t) = a(t) + b(t) \quad \text{and} \quad T_{a+b} = T_a \cup T_b.$$

The *product* (sequential links) $a \cdot b$ is defined as

$$(a \cdot b)(t) = a(t) \cdot b(t) \quad \text{and} \quad T_{a \cdot b} = T_a \cap T_b.$$



Sum and product of temporal quantities

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

```
a = [(1, 5, 2), (6, 8, 1), (11, 12, 3), (14, 16, 2),  
      (17, 18, 5), (19, 20, 1)]  
b = [(2, 3, 4), (4, 7, 3), (9, 10, 2), (13, 15, 5), (16, 21, 1)]
```

The following are the sum $s = a + b$ and the product $p = a \cdot b$ of temporal quantities a and b over combinatorial semiring.

```
s = [(1, 2, 2), (2, 3, 6), (3, 4, 2), (4, 5, 5), (5, 6, 3),  
      (6, 7, 4), (7, 8, 1), (9, 10, 2), (11, 12, 3),  
      (13, 14, 5), (14, 15, 7), (15, 16, 2), (16, 17, 1),  
      (17, 18, 6), (18, 19, 1), (19, 20, 2), (20, 21, 1)]  
p = [(2, 3, 8), (4, 5, 6), (6, 7, 3), (14, 15, 10),  
      (17, 18, 5), (19, 20, 1)]
```

They are visually displayed at the bottom half of figures on the following slides.



Addition of temporal quantities.

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

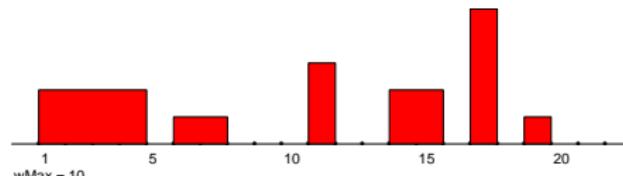
Derived Ns

Temporal Ns

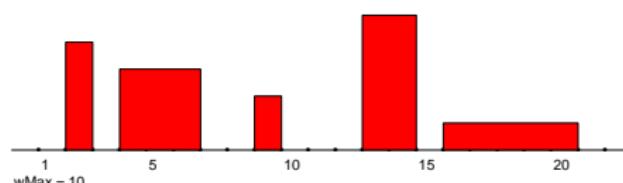
References

Appendix

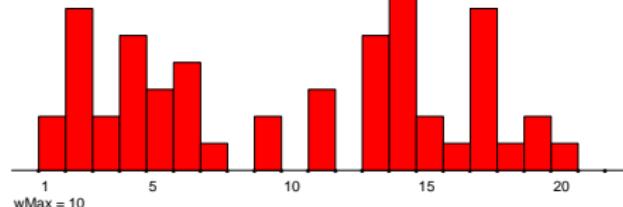
$a :$



$b :$



$a + b :$





Multiplication of temporal quantities.

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

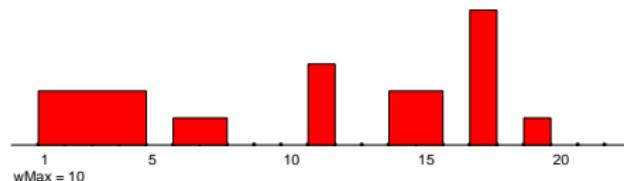
Derived Ns

Temporal Ns

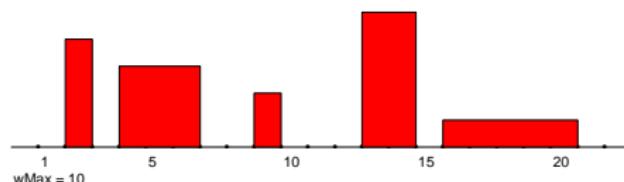
References

Appendix

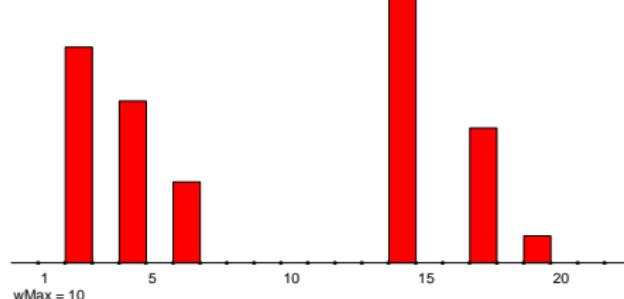
$a :$



$b :$



$a \cdot b :$





Temporal affiliation networks

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Let the binary **affiliation** matrix $\mathbf{A} = [a_{ep}]$ describe a two-mode network on the set of events E and the set of participants P :

$$a_{ep} = \begin{cases} 1 & p \text{ participated in the event } e \\ 0 & \text{otherwise} \end{cases}$$

The function $d : E \rightarrow \mathcal{T}$ assigns to each event e the date $d(e)$ when it happened. $\mathcal{T} = [\text{first}, \text{last}] \subset \mathbb{N}$. Using these data we can construct two temporal affiliation matrices:

- **instantaneous $\mathbf{Ai} = [ai_{ep}]$** , where

$$ai_{ep} = \begin{cases} [(d(e), d(e) + 1, 1)] & a_{ep} = 1 \\ [] & \text{otherwise} \end{cases}$$

- **cumulative $\mathbf{Ac} = [ac_{ep}]$** , where

$$ac_{ep} = \begin{cases} [(d(e), last + 1, 1)] & a_{ep} = 1 \\ [] & \text{otherwise} \end{cases}$$



Multiplication of temporal affiliation networks

Instantaneous

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Instantaneous **A** on $P \times A$ and **B** on $P \times B$. **C** = $\mathbf{A}^T \cdot \mathbf{B}$ on $A \times B$.

$$c_{ij}(t) = \sum_{p \in P} a_{pi}(t)^T \cdot b_{pj}(t)$$

$a_{pi} = [(d_{pi}, d_{pi} + 1, v_{pi})]$ and $b_{pj} = [(d_{pj}, d_{pj} + 1, v_{pj})]$
for $t = d$ we get

$$c_{ij} = [(d, d + 1, \sum_{p \in P: d_{pi}=d_{pj}=d} v_{pi} \cdot v_{pj})]_{d \in \mathcal{T}}$$

for $v_{pi} = v_{pj} = 1$ we finally get

$$v_{ij}(d) = |\{p \in P : d_{pi} = d_{pj} = d\}|$$

For binary temporal two-mode networks **A** and **B** the value $v_{ij}(d)$ of the product $\mathbf{A}^T \cdot \mathbf{B}$ is equal to the number of different members of P with which both i and j have contact in the instant d .



Multiplication of temporal affiliation networks

Cumulative

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Cumulative **A** on $P \times A$ and **B** on $P \times B$. $\mathbf{C} = \mathbf{A}^T \cdot \mathbf{B}$ on $A \times B$.

$$c_{ij}(t) = \sum_{p \in P} a_{pi}(t)^T \cdot b_{pj}(t)$$

$a_{pi} = [(d_{pi}, \text{last} + 1, v_{pi})]$ and $b_{pj} = [(d_{pj}, \text{last} + 1, v_{pj})]$
for $t = d$ we get

$$c_{ij} = [(d, d + 1, \sum_{p \in P: (d_{pi} \leq d) \wedge (d_{pj} \leq d)} v_{pi} \cdot v_{pj})]_{d \in \mathcal{T}}$$

for $v_{pi} = v_{pj} = 1$ we finally get

$$v_{ij}(d) = |\{p \in P : (d_{pi} \leq d) \wedge (d_{pj} \leq d)\}|$$

For binary temporal two-mode networks **A** and **B** the value $v_{ij}(d)$ of the product $\mathbf{A}^T \cdot \mathbf{B}$ is equal to the number of different members of P with which both i and j have contact in all instants up to including the instant d .



Temporal co-authorship networks

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

Using the multiplication of temporal matrices over the combinatorial semiring we get the corresponding instantaneous and cumulative co-occurrence matrices

$$\mathbf{Ci} = \mathbf{Ai}^T \cdot \mathbf{Ai} \quad \text{and} \quad \mathbf{Cc} = \mathbf{Ac}^T \cdot \mathbf{Ac}$$

A typical example of such a matrix is the papers authorship matrix **WA** where E is the set of papers W , P is the set of authors A and d is the publication year.

The triple (s, f, v) in a temporal quantity ci_{pq} tells that in the time interval $[s, f)$ there were v events in which both p and q took part.

The triple (s, f, v) in a temporal quantity cc_{pq} tells that in the time interval $[s, f)$ there were in total v accumulated events in which both p and q took part.

The diagonal matrix entries ci_{pp} and cc_{pp} contain the temporal quantities counting the number of events in the time intervals in which the participant p took part.



Temporal co-authorship network for SN5

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

BibTime

SN5 (2008)

	W	A	K	J
raw	193376	75930	29267	14651
DC=1	7950	12458		

In Pajek we extract a subnetwork **WAc** and a corresponding partition **SN5yearC**. Using a program twoMode2netJSON we transform them into temporal network in the netJSON format.

Bibliographic networks are usually sparse. The network **WAcInst** has 19488 arcs. The co-authorship network

ColInst = **WAcInst**^T * **WAcInst** has 64980 edges; the corresponding matrix in the package **TQ** has $12458^2 = 155201764$ entries. Using a package **Graph** the co-authorship network is computed in a second and half – a big speed-up.



multiply.py

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

```
gdir = 'c:/users/batagelj/work/python/graph/graph'
wdir = 'c:/users/batagelj/work/python/graph/JSON/SN5'
cdir = 'c:/users/batagelj/work/python/graph/chart'
import sys, os, datetime, json
sys.path = [gdir]+sys.path; os.chdir(wdir)
import TQ
from GraphNew import Graph
# file = 'C:/Users/batagelj/work/Python/graph/JSON/WAtest.json'
file = 'C:/Users/batagelj/work/Python/graph/JSON/SN5/WAcInst.json'
# file = 'C:/Users/batagelj/work/Python/graph/JSON/SN5/WAcCum.json'
# file = 'C:/Users/batagelj/work/Python/graph/JSON/Gisela/papInst.json'
t1 = datetime.datetime.now()
print("started: ",t1.ctime(),"\n")
G = Graph.loadNetJSON(file)
t2 = datetime.datetime.now()
print("\nloaded: ",t2.ctime(),"\n time used: ", t2-t1)
# T = G.transpose()
# Co = Graph.TQmultiply(T,G,True)
# CR = G.TQtwo2oneRows()
CC = G.TQtwo2oneCols()
t3 = datetime.datetime.now()
print("\ncomputed: ",t3.ctime(),"\n time used: ", t3-t2)
ia = { v[3]['lab']: k for k,v in CC._nodes.items() }
# CC._links[(ia['BORGATTI_S'],ia['EVERETT_M'])][4]['tq']
# CC._links[(ia['IDI/B'],ia['HCL/B'])][4]['tq']
```



Temporal co-authorship network for SN5

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

```
===== RESTART: C:\Users\batagelj\work\Python\graph\graph\multiply.py =====
started: Sun Nov 20 00:26:51 2016
loaded: Sun Nov 20 00:26:51 2016
time used: 0:00:00.425024
computed: Sun Nov 20 00:26:52 2016
time used: 0:00:01.165066
>>> BB = CC._links[(ia['BORGATTI_S'],ia['BORGATTI_S'])][4]['tq']
>>> BE = CC._links[(ia['BORGATTI_S'],ia['EVERETT_M'])][4]['tq']
>>> BB
[((1988, 1990, 2), (1990, 1991, 4), (1991, 1992, 2), (1992, 1993, 4),
(1993, 1994, 2), (1994, 1995, 3), (1996, 1997, 1), (1997, 1998, 2),
(1998, 1999, 1), (1999, 2000, 3), (2001, 2002, 2), (2002, 2003, 1),
(2003, 2004, 4), (2005, 2006, 3), (2006, 2007, 2), (2007, 2008, 3)])
>>> BE
[((1988, 1989, 1), (1989, 1990, 2), (1990, 1991, 4), (1991, 1992, 1),
(1992, 1995, 2), (1996, 1998, 1), (1999, 2000, 3), (2003, 2004, 1),
(2005, 2007, 1)])
>>> TQmax = 8; Tmin = 1970; Tmax = 2009; w = 600; h = 120
>>> tit = 'BORGATTI_S'
>>> Graph.TQshow(BB,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='orange')
>>> tit = 'BORGATTI_S - EVERETT_M'
>>> Graph.TQshow(BE,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='orange')
>>> NN = CC._links[(ia['NEWMAN_M'],ia['NEWMAN_M'])][4]['tq']
>>> NN
[((1999, 2000, 2), (2000, 2001, 4), (2001, 2002, 7), (2002, 2003, 8),
(2003, 2004, 7), (2004, 2005, 11), (2005, 2006, 7), (2006, 2007, 11),
(2007, 2008, 3)])
>>> tit = 'NEWMAN_M'; TQmax = 12; h = 150
>>> Graph.TQshow(NN,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='orange')
```



Visualization

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

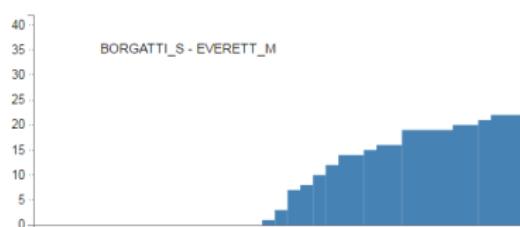
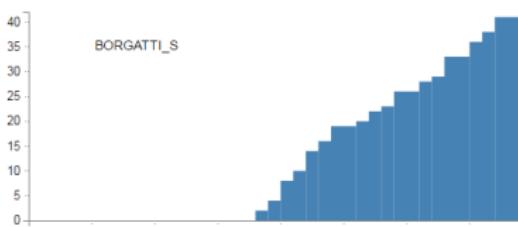
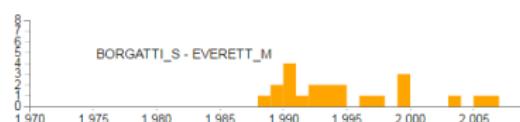
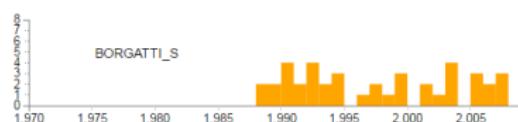
Multiplication

Derived Ns

Temporal Ns

References

Appendix





Understanding large networks

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

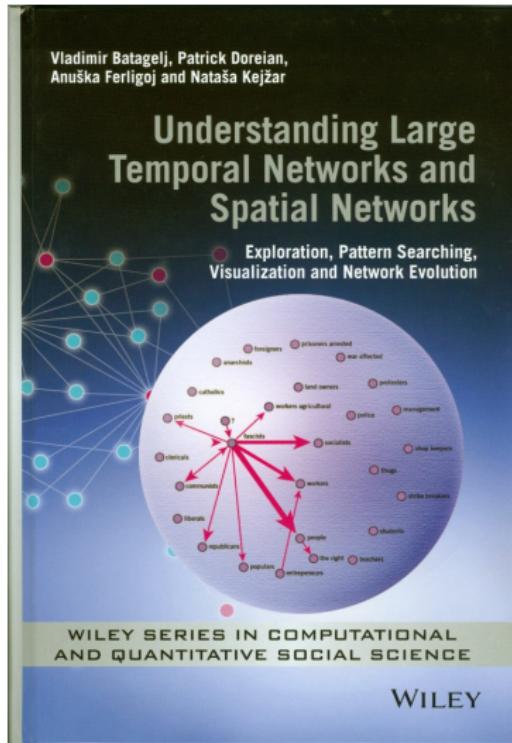
Multiplication

Derived Ns

Temporal Ns

References

Appendix



This lecture is closely related to chapters 2 and 3 in the book:

Vladimir Batagelj, Patrick Doreian, Anuška Ferligoj and Nataša Kejžar: Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution. Wiley Series in Computational and Quantitative Social Science. **Wiley**, October 2014.



References |

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

- Ahmed, A., Batagelj, V., Fu, X., Hong, S.-H., Merrick, D., Mrvar, A.: Visualisation and analysis of the Internet movie database. Asia-Pacific Symposium on Visualisation 2007 (IEEE Cat. No. 07EX1615), 2007, p 17-24.
- Barabasi, A.L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the social network of scientific collaborations. *Physica* **311** (2002) 590–614
- Batagelj, V.: Wos2pajek – networks from web of science (2007).
<http://vladowiki.fmf.uni-lj.si/doku.php?id=pajek:wos2pajek>
- Batagelj, V., Cerinšek, M.: On bibliographic networks. *Scientometrics* **96** (2013) 3, 845-864.
- Batagelj, V., Praprotnik, S.: An algebraic approach to temporal network analysis based on temporal quantities. *Social Network Analysis and Mining*, **6**(2016)1, 1-22.
- Batagelj, V., Zaveršnik, M.: Fast algorithms for determining (generalized) core groups in social networks. *Advances in Data Analysis and Classification*, 2011. Volume 5, Number 2, 129-145.



References II

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

- Cerinšek, M., Batagelj, V.: Network analysis of Zentralblatt MATH data. *Scientometrics*, 102(2015)1, 977-1001.
- Cerinšek, M., Batagelj, V.: Generalized two-mode cores. *Social Networks* 42 (2015), 80–87.
- De Nooy, W., Mrvar, A., Batagelj, V.: Exploratory Social Network Analysis with Pajek; Revised and Expanded Second Edition. *Structural Analysis in the Social Sciences*, Cambridge University Press, September 2011.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. *American Association for Artificial Intelligence Magazine* (1996), 37–54
- J. G. Fletcher, "A more general algorithm for computing closed semiring costs between vertices of a directed graph," *CACM* (1980), pp. 350-351.
- Kejžar, N., Korenjak Černe, S., Batagelj, V.: Network Analysis of Works on Clustering and Classification from Web of Science. *Classification as a Tool for Research*. Hermann Locarek-Junge, Claus Weihs eds. *Proceedings of IFCS 2009. Studies in Classification, Data Analysis, and Knowledge Organization*, 525-536, Springer, Berlin, 2010.



References III

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

- Newman, M.E.: The structure of scientific collaboration communities. *Proceedings of the National Academy of Science (PNAS)* **98** (2001) 404–409
- Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. *Physical Review E* **69** (2004)
- Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435** (2005) 814
- Palla, G., Barabasi, A.L., Vicsek, T.: Quantifying social group evolution. *Nature* **446** (2007) 664-667
- Perianes-Rodriguez, A., Waltman, L., Van Eck, N.J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics*, 10(4), 1178-1195.
- Wasserman, S., Faust, K.: Social network analysis: methods and applications. Cambridge Univ. Press, Cambridge, 1997.
- Zaveršnik, M., Batagelj, V.: Islands. In: XXIV International Sunbelt Social Network Conference, Portorož, Slovenia (2004)



References IV

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix



Pajek's wiki. <http://pajek.imfm.si>



Vladimir Batagelj, Andrej Mrvar: [Pajek manual](#).



Wouter De Nooy, Andrej Mrvar, Vladimir Batagelj: Exploratory Social Network Analysis with Pajek; Revised and Expanded Second Edition. Structural Analysis in the Social Sciences, Cambridge University Press, September 2011.



Names of works in WoS2Pajek

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

The usual *ISI name* of a work (field CR)

LEFKOVITCH LP, 1985, THEOR APPL GENET, V70, P585

has the following structure

AU+' , '+PY+' , '+SO[:20]+' , V'+VL+' , P'+BP

All its elements are in upper case.

In WoS the same work can have different ISI names. To improve the precision the program WoS2Pajek supports also *short names* (similar to the names used in HISTCITE output). They have the format:

LastNm[:8]+' _'+FirstNm[0]+' ('+PY+') '+VL+': '+BP

For example: LEFKOVIT_L(1985)70:585

From the last names with prefixes VAN, DE, ... the space is deleted.

Unusual names start with character * or \$.



... Names of works

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

In the CR field other forms of ISI names and several errors and inconsistencies can be found:

NEWMAN MEJ, 2004, PHYS REV E 2, V69, ARTN 066133
PALLA G, 2005, NATURE, V435, P814, DOI 10.1038/nature03607
PAPIN JA, 2004, TRENDS BIOCHEM SCI, V29, P641, DOI
10.1016/j.tibs.2004.10.001
DOLCINI MM, 2005, J ADOLESCENT HEALTH, V36, UNSP 267.E6-15
EVANS JD, 2001, GENOME BIOL, V2, UNSP RESEARCH0001
NEWMAN MEJ, 2001, IN PRESS COMPLEX NETUNSP 215239
GRANOVET MS, 1973, AM J SOCIOLOG, V78, P1360
GRANOVETTER M, 1983, SOCIOLOGICAL THEORY, V1, P203
BORGATTI SP, 2002, UGINET WINDOWS SOFTW
BORGATTI S, 1999, UCINET V USERS GUIDE
CANTAZARO M, 2005, PHYS REV E, V71, UNSP 027103
CANTAZARO M, 2005, PHYS REV E, V71, UNSP 056104
CATANZARO M, 2005, PHYS REV E 2, V71, ARTN 056104
BRICKER PD, 1968, OCT M PSYCH SOC ST L : BRICKER

We decided to treat in short names the ARTN and UNSP values as BP values.
We also remove the DOI parts. There are also irregular names in AU field:

AU BENSON, , C
KULHAVY, , W
AU SCHONEMÄ. PH

The user can correct the typing errors and nonuniformities on the WoS file.



Program WoS2Pajek

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

For converting WoS file into networks in Pajek's format a program **WoS2Pajek** was developed (in Python). It produces the following files:

- citation network: works \times works, **Ci**;
- authorship (two-mode) network: works \times authors, **WA**; for works without complete description only the first author is known;
- keywords (two-mode) network: works \times keywords, **WK**; only for works with complete description;
- journals (two-mode) network: works \times journals, **WJ**; field J9;
- partition of works by the publication year, **year**;
- partition of works, **DC** – complete description (1) / ISI name only (0);
- vector number of pages, **np**; PG or EP – BP + 1.



Program WoS2Pajek

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

The keywords are obtained from the fields TI (title), ID, DE and AB (abstract). From the text the **stopwords** are removed and a list of words is produced. The words are lemmatized using **MontyLingua** package.

In future versions additional networks can be derived: works × discipline, works × countries, ...

In version 0.7 a GUI support (based on Tkinter) for specifying the program parameters was implemented.

Similar package of programs was produced also by **Loet Leydesdorff**.



Program WoS2Pajek

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

The current version of WoS2Pajek requires 7 parameters to be given by the user:

- MontyLingua directory: path to the directory in which the MontyLingua package is installed (put it also in the PATH env-variable);
- project directory: where the output files are saved;
- WoS file;
- maxnum – estimate of the number of all vertices (number of records + number of cited Works) – $30 * \text{number of records}$;
- step – prints info about each $k * \text{step}$ record as a trace; $\text{step} = 0$ – no trace.
- use ISI name / short name;
- make a clean WoS file without duplicates;
- boolean list [DE, ID, TI, AB] specifying which fields are sources of keywords.



Collecting the data from Web of Science

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

When preparing for exploration of selected topic we first search the WoS for records matching selected keys. WoS allows storing the hits on a file, but only 500 at once.

For the list of hits we can in the **Citation report**, using the option **View citing articles**, get also the list of all works citing the works from the first list. We also save these records and join all the files into a single file to which we apply WoS2Pajek.

Using Pajek we identify the important only cited works – out-degre = 0 and large in-degree. We search for them in WoS and try to add their descriptions.

In this way we produce the final source data set for our analysis.



Types on DC file

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Appendix

When we combine partial files with saved records from WoS into a single file required by the program WoS2Pajek we can include into this file some additional lines:

Comments have the form

**** comment**

Besides this we can specify different types of input records using the lines of the form

***T n**

where n is a type number (1, 2, ...). Since the same record can appear in different parts of the file its class is determined as the set of all corresponding types transformed in integer. For example:
 $\{3, 1\} \rightarrow 2^2 + 2^0 = 5$.