



Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Analysis of bibliographic networks

Vladimir Batagelj

IMFM Ljubljana, IAM UP Koper and NRU HSE Moscow

Daria Maltseva

NRU HSE Moscow

10th International Summer School “Analysis of Scientific Networks”

Moscow, July, 15–21, 2019



Outline

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

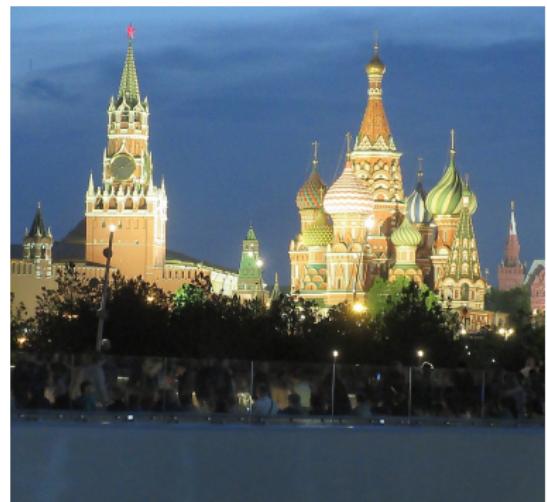
Multiplication

Derived Ns

Temporal Ns

References

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



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

Daria Maltseva: d_malceva@mail.ru

Current version of slides (July 19, 2019 at 00:52): [slides PDF](#)

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



Networks

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

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

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

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 links 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

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

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

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

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

	n	freq	first author	title
Bibliographic networks	1	173	Cohen, J	Statistical Power Analysis for the Behavioral Sciences. Routledge, 1988
V. Batagelj, D. Maltseva	2	164	Peters, DP	Peer-review practices of psychological journals - the fate of ... Behav Brain Sci
Networks	3	151	Egger, M	Bias in meta-analysis detected by a simple, graphical test. Brit Med J, 1997
Statistics	4	150	Stroup, DF	Meta-analysis of observational studies in epidemiology - A proposal for reporting
Citation	5	135	Dersimonian, R	Metaanalysis in clinical-trials. Control Clin Trials, 1986
Two-mode Ns	6	130	Zuckerman, H	Patterns of evaluation in science - institutionalisation, structure and function
Multiplication	7	130	Higgins, JPT	Cochrane Handbook for Systematic Reviews of Interventions. Cochrane, 2011
Derived Ns	8	126	Moher, D	Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA
Temporal Ns	9	125	Higgins, JPT	Measuring inconsistency in meta-analyses. Brit Med J, 2003
References	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

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

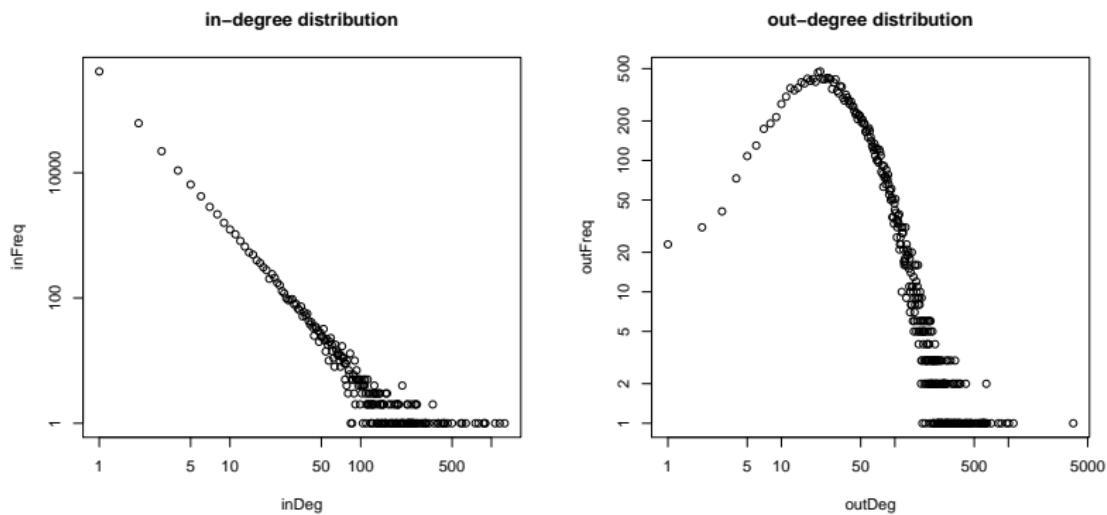
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



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

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

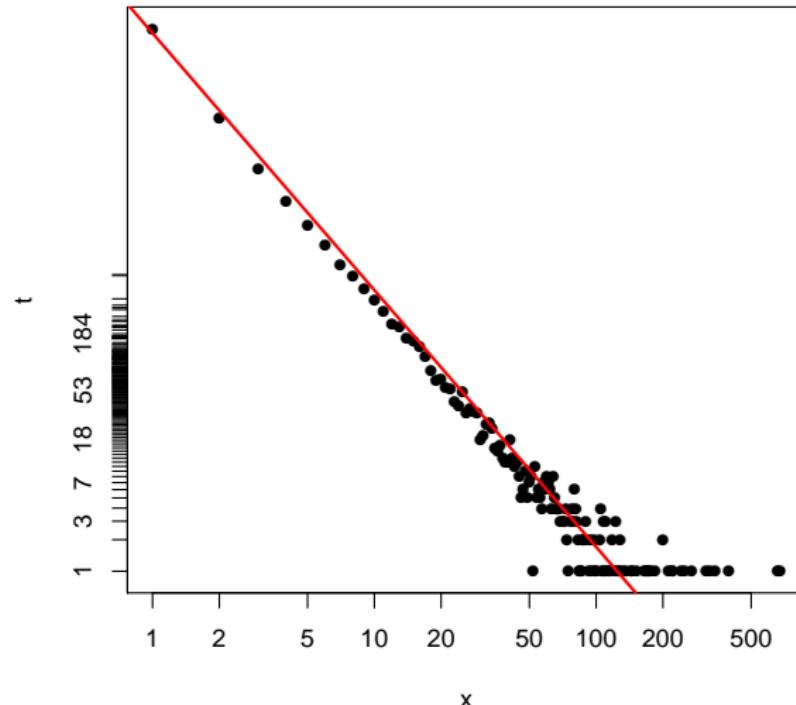
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Distributions

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

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

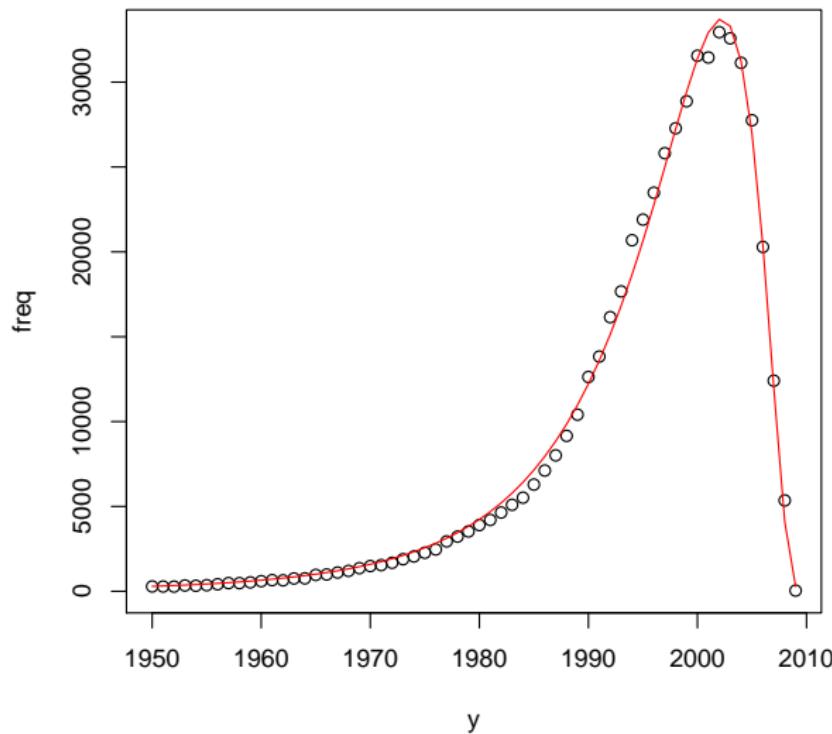
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Zentralblatt MATH networks

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

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

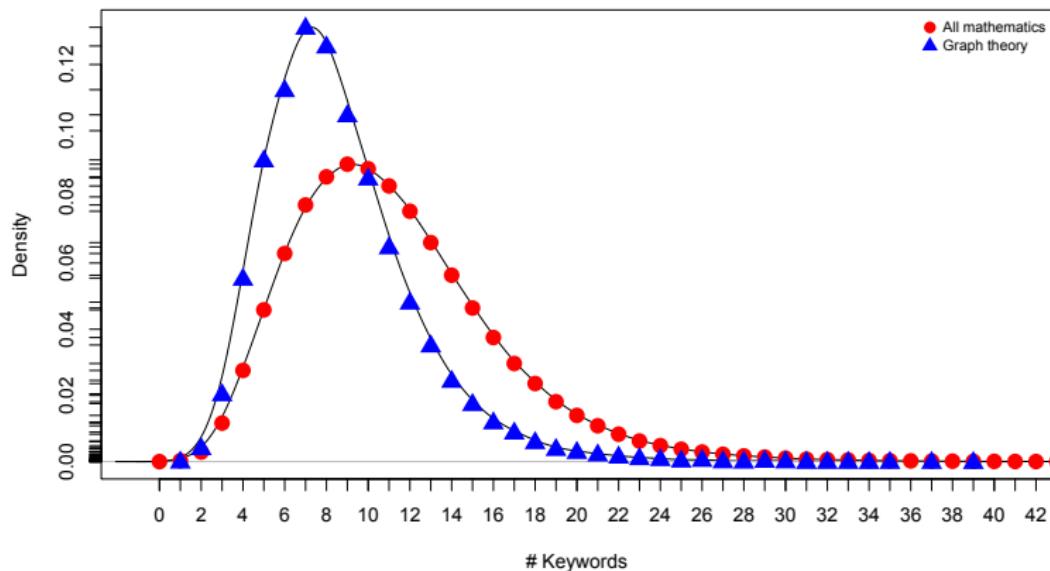
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





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

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

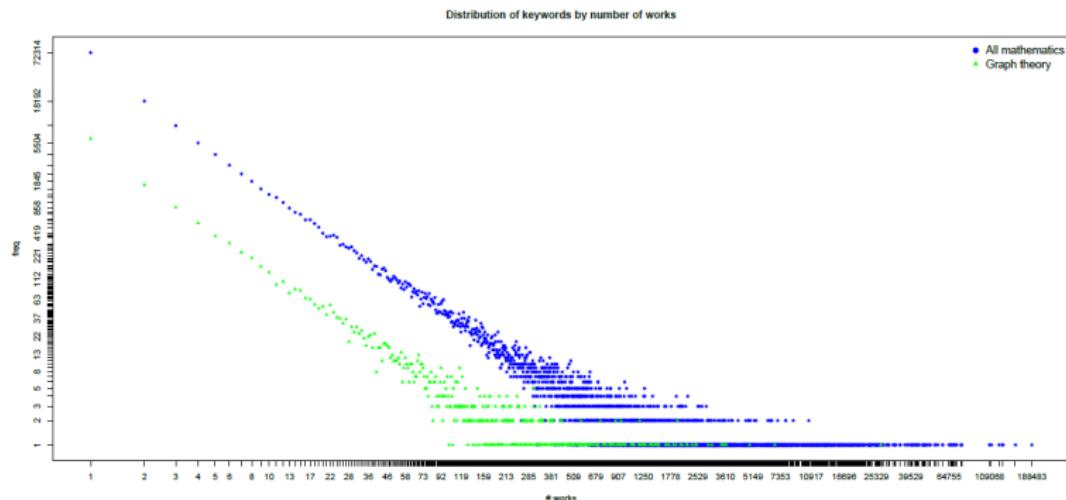
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Number of authors: temporal distribution

SNA 2018

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Combining the number of authors with the publication year we get their temporal distribution, describing how the number of authors is changing through years. See next slide.

The results show that since 1980s, the number of single author papers dropped from 70% to almost 10%. The number of papers authored by a pair of authors is relatively constant – around 25%. The numbers of papers authored by 3 and more authors are increasing (3: from 6% to 25%, 4: from 2% to 17%, 5: from 0% to 10%, etc.).

Besides the general trend to higher collaboration the reason could be also the expansion of SNA to other disciplines (physics, computer science, neuro science, biology, chemistry, etc.) with different writing cultures.

SocNet

Number of authors: temporal distribution

SNA 2018

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

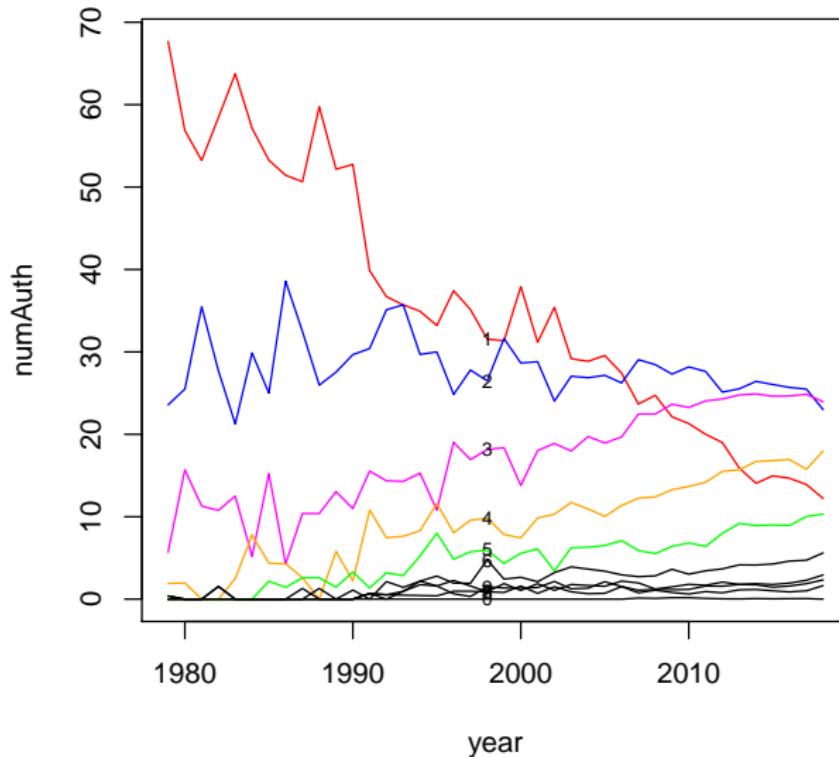
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Citations between years in Bounded network

SNA 2018

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

It is interesting to observe how many citations are made per year. We combined CiteB network with the partition on the year of publication and constructed a network of citations between years, where the values are equal to the number of times that all works published in one year were cited in all works published in another year (the network is directed, only later years can cite previous years).

Figure on the next slide presents the distribution of citations between years in a three-dimensional space. The majority of citations in recent works are made to recent works as well. The years having the largest amount of citations from other years are 2007 (80,129), 2008 (77,595), 2009 (82,294), 2010 (88,840), 2011 (79,843). The largest number of citations are from 2015 and 2016 to 2010 (16,384 and 15,755, respectively) and 2011 (16,026 and 15,944).

Figure on the next slide presents the normalized curves of values of citations per year in the period 1985-2018 (54 years in total). The result shows that the yearly citation patterns do not vary significantly from year to year – there are always noticeably more citations made of recent works, than of works published previously.



Citations between years

SNA 2018

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

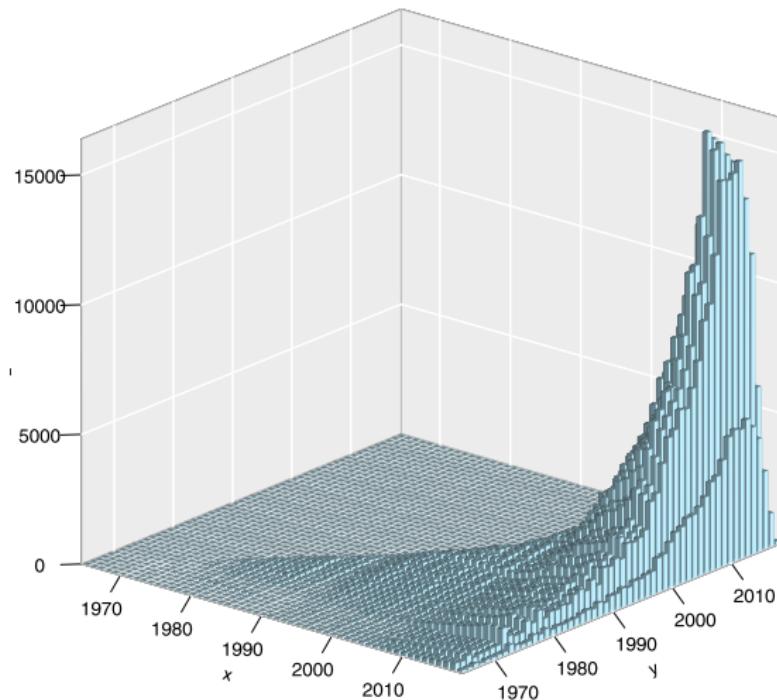
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



SocNet



Citations per year; normalized curves

SNA 2018

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

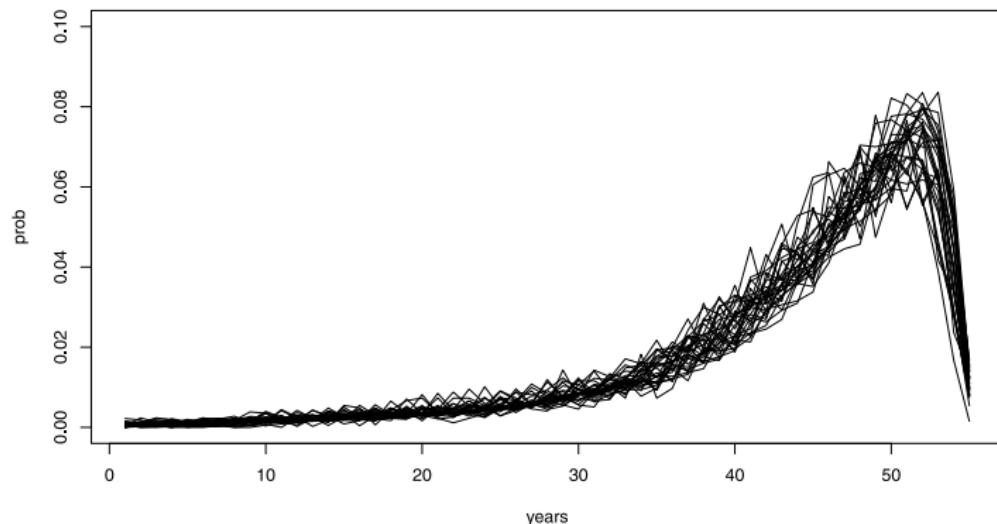
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



SocNet

Citation networks

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

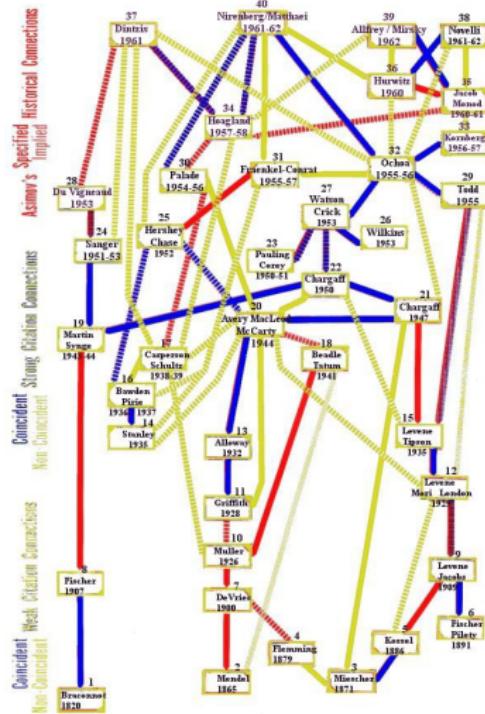
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



The citation network analysis started in 1964 with the paper of **Garfield et al.** In 1989 **Hummon and Dorrian** proposed three indices – weights of arcs that provide us with automatic way to identify the (most) important part of the citation network. We developed an algorithm to efficiently compute SPC weights, provided the citation network is acyclic.



... Citation networks

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

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

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

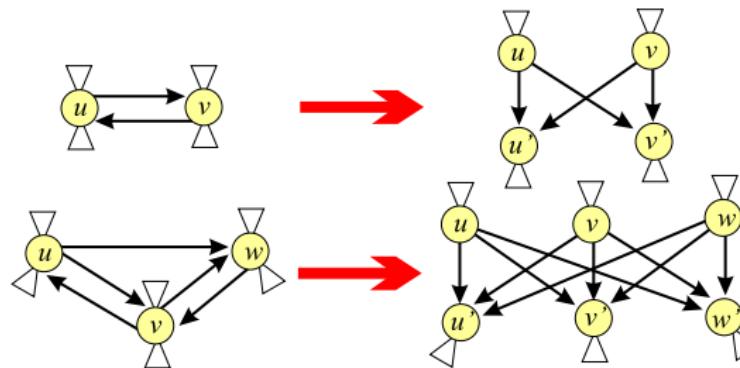
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



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

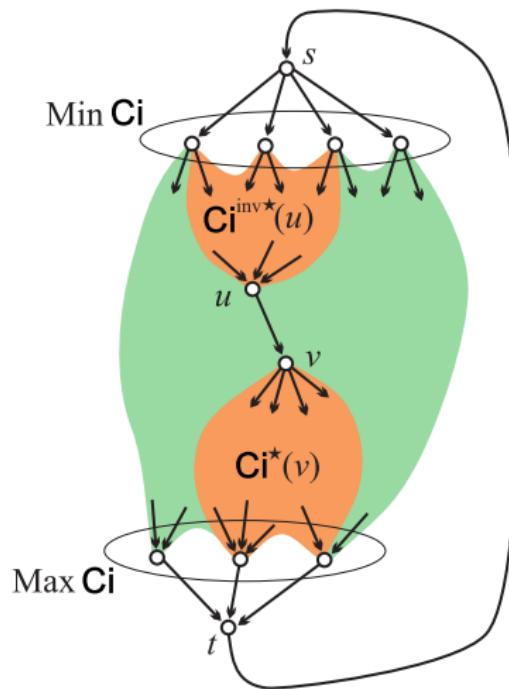
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



We assume that the citation relation \mathbf{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 \mathbf{Ci} ; and all maximal elements of \mathbf{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

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

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

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}' .

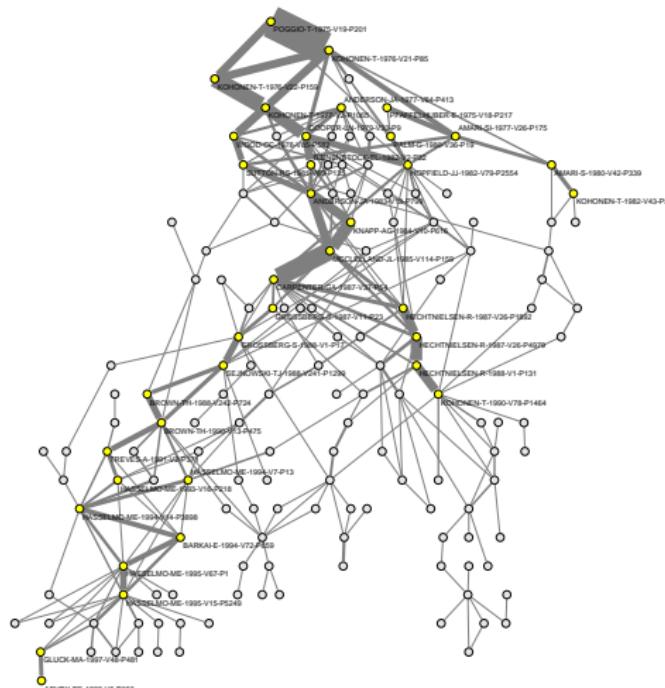
```
File/Network/read eatRS.net
Info/Network/Line values ... >= 70
Net/Transform/Remove/Lines with Value/lower than 70
Net/Partitions/Degree/All
Operations/Extract from Network/Partition 1-*
Net/Components/Weak
Draw/Draw-Partition
```

Citation weights

Bibliographic networks

V. Batagelj,
D. Maltseva

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



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

For visualization of acyclic networks in Pajek use the macro layers.

See [paper](#).

Cores and generalized cores

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

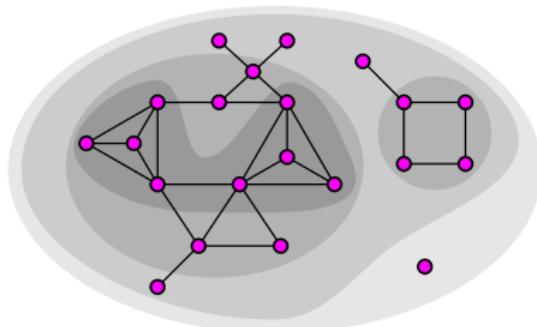
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



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

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

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

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

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

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

```
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

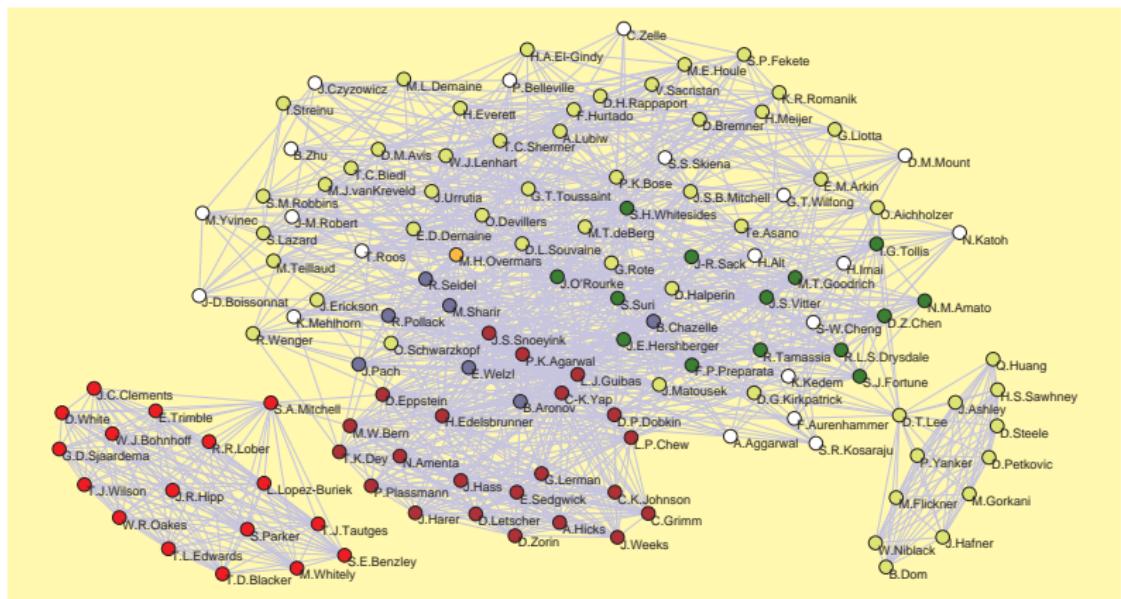
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





p_S -core at level 46 in Computational Geometry network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

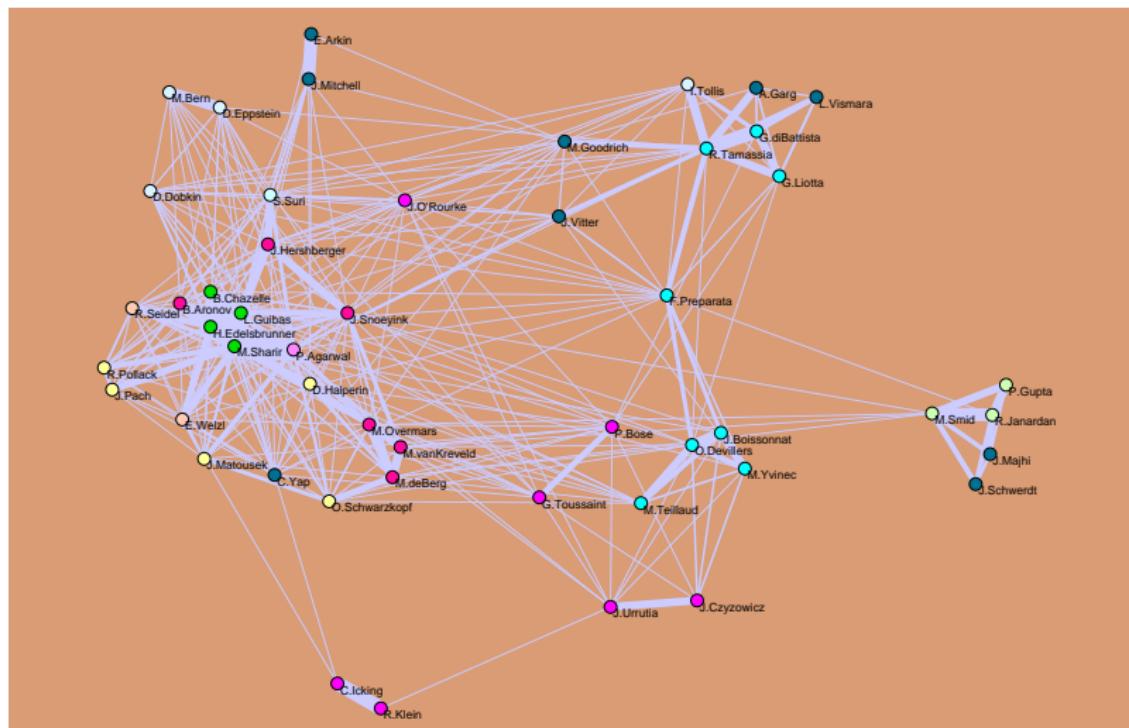
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Islands

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

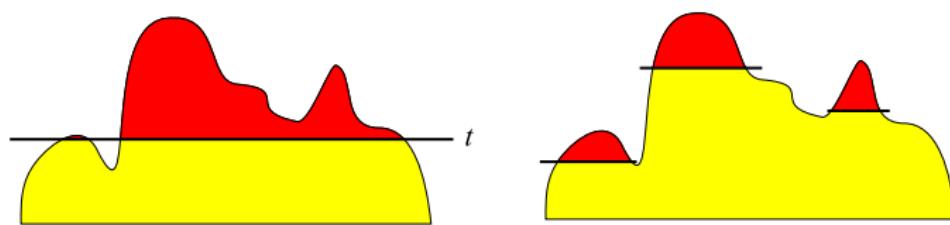
Multiplication

Derived Ns

Temporal Ns

References

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

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

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

- 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

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

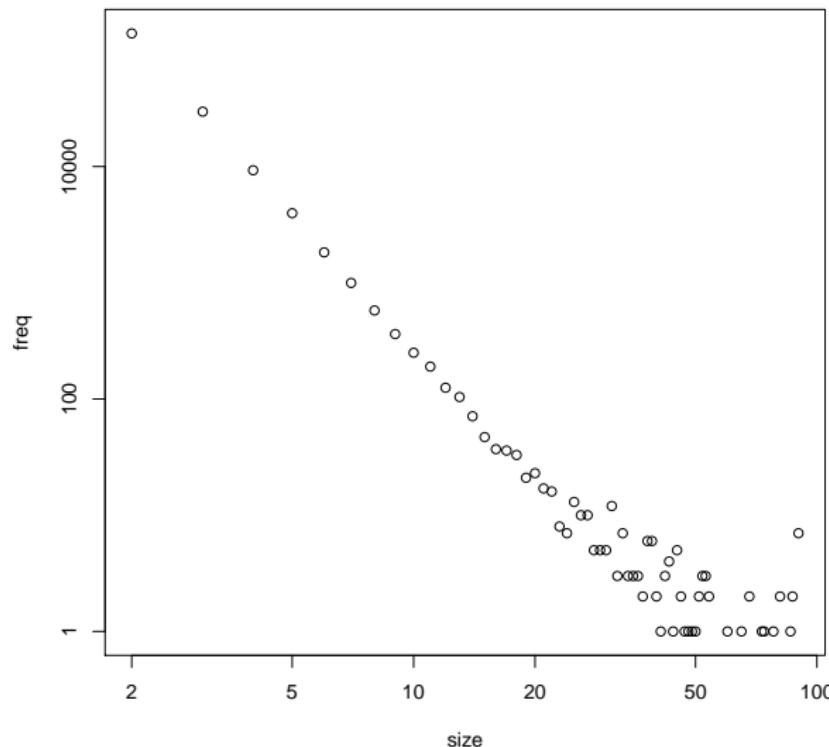
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Main path and main island in US Patents Nber, US Patents; $n = 3774768$, $m = 16522438$

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

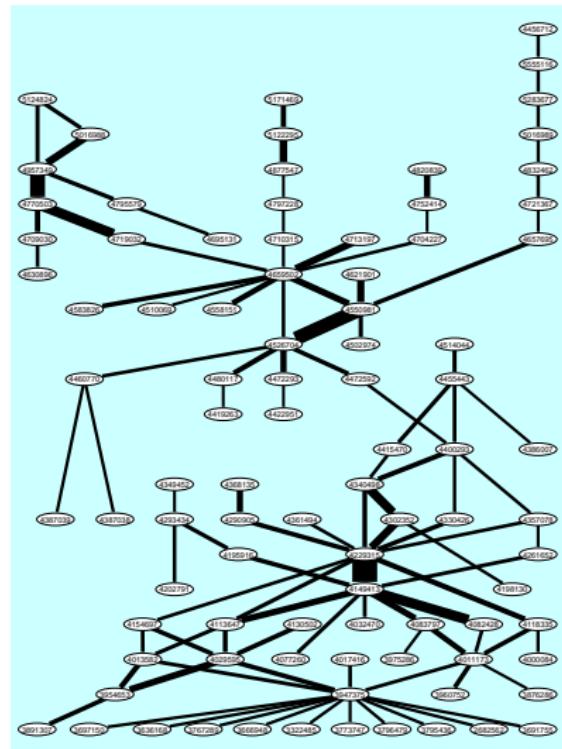
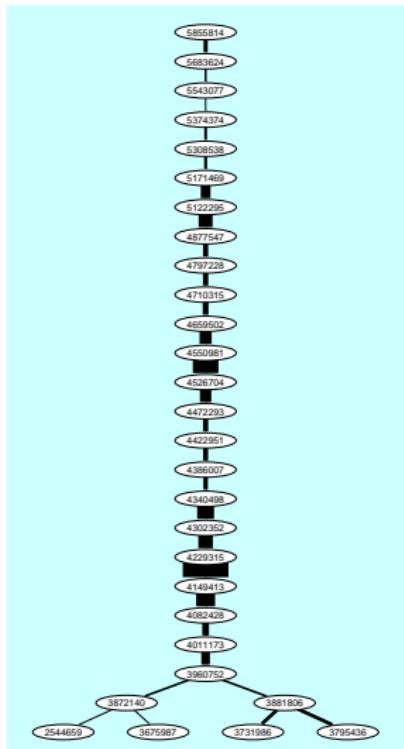
Two-mode Ns

Multiplication

Derived Nc

Temporal Nc

References





Main island – Liquid crystal display

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Table 1: Patents on the liquid-crystal display

patent	date	author(s) and title
2049629	Mar 13, 1951	Method of preparing sheet and the like and the formation and use thereof
2092562	Jul 29, 1954	Wender, et al. Reduction of aromatic carbonyl groups by metallic elements utilizing an organic nematic compound
3323485	May 30, 1967	Strelzow, et al. Electro-optic systems having an undistorted image on a disturbed background
3636168	Jun 18, 1972	Josephson. Preparation of polyacrylate ester compounds containing a chiral substituent having a large optical rotation
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 viscous liquid crystal to reduce the turn-off time
3731966	May 8, 1973	Fenggas. Display devices utilizing liquid crystal light
3767289	Oct 23, 1973	Arivazhagan, et al. Class of stable ester-thiophene compounds, some displaying electro-nematic phases at or near room temperature and others, a range of 0°C to 100°C
3773477	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 materials
3797479	Mar 12, 1974	Hofrichter, et al. Electro-optical liquid-crystal materials which exhibit the Kerr effect at isotropic temperatures
3872140	Mar 18, 1975	Klauder, 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
3880330	Mar 28, 1976	Tanabe, et al. Liquid crystal compositions
3891307	Jun 24, 1976	Tanabe, et al. Phase control of the voltage applied to opposite electrodes for a cholesteric to nematic phase transition
3947375	Mar 30, 1976	Goto, et al. Liquid crystal materials and devices
3958053	May 4, 1976	Yamamoto. Liquid crystal compositions and high dielectric constant materials for electro-optic devices incorporating same
3960752	Jun 1, 1976	Klauder, et al. Liquid crystal compositions
3973286	Aug 17, 1976	Hofrichter, et al. Electro-optical liquid crystal compositions and method of synthesis
4000894	Dec 28, 1976	Hofrichter, et al. Liquid crystal mixtures for electro-optical devices
4011173	Mar 8, 1977	Steinmesser. Modified nematic mixture with reduced viscosity
4013582	Mar 22, 1977	Gavrilev. Liquid crystal compounds and electro-optic devices incorporating them
4017416	Apr 12, 1977	Steinmesser. 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	Yamamoto. Liquid crystal compositions
4077260	Mar 7, 1978	Method for preparing some liquid crystal compositions using same
4082428	Aug 4, 1978	Goto, et al. Optically active cyano-biphenyl compounds and liquid crystal materials containing them. 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. Novel liquid crystal compositions
4120794	Sep 12, 1979	Custodis, et al. Liquid crystalline materials
4118320	Oct 3, 1979	Krause, et al. Liquid crystalline materials of reduced viscosity
4120795	Oct 3, 1979	Goto, et al. Optically active liquid crystal mixtures and liquid crystal devices containing them
4149412	Apr 17, 1979	Goto, et al. Optically active liquid crystal compositions
4149407	May 15, 1979	Goto, et al. Liquid crystal derivatives of bis(4-alkylphenyl)benzenes
4189130	Aug 1, 1980	Custodis, et al. Liquid crystal mixtures
4202794	May 13, 1980	Osawa, et al. Nematic liquid crystalline materials
4213622	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213623	Aug 14, 1980	Goto, et al. Liquid crystal compounds
4213625	Aug 14, 1980	Goto, et al. Liquid crystal materials
4213626	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213627	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213628	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213629	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213630	Aug 14, 1980	Goto, et al. Liquid crystal materials
4213631	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213632	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213633	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213634	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213635	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213636	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213637	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213638	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213639	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213640	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213641	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213642	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213643	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213644	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213645	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213646	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213647	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213648	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213649	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213650	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213651	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213652	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213653	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213654	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213655	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213656	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213657	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213658	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213659	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213660	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213661	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213662	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213663	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213664	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213665	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213666	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213667	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213668	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213669	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213670	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213671	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213672	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213673	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213674	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213675	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213676	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213677	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213678	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213679	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213680	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213681	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213682	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213683	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213684	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213685	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213686	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213687	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213688	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213689	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213690	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213691	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213692	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213693	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213694	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213695	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213696	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213697	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213698	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213699	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213700	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213701	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213702	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213703	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213704	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213705	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213706	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213707	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213708	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213709	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213710	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213711	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213712	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213713	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213714	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213715	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213716	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213717	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213718	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213719	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213720	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213721	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213722	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213723	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213724	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213725	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213726	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213727	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213728	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213729	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213730	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213731	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213732	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213733	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213734	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213735	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213736	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213737	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213738	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213739	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213740	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213741	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213742	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213743	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213744	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213745	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213746	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213747	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213748	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213749	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213750	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213751	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213752	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213753	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213754	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213755	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213756	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213757	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213758	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213759	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213760	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213761	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213762	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213763	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213764	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213765	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213766	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213767	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213768	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213769	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213770	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213771	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213772	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213773	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213774	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213775	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213776	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213777	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213778	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213779	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213780	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213781	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213782	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213783	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213784	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213785	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213786	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213787	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213788	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213789	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213790	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213791	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213792	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213793	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213794	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213795	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213796	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213797	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213798	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213799	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213800	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213801	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213802	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213803	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213804	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213805	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213806	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213807	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213808	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213809	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213810	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213811	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213812	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213813	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213814	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213815	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213816	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213817	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213818	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213819	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213820	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213821	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213822	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213823	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213824	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213825	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213826	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213827	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213828	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213829	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213830	Aug 14, 1980	Goto, et al. Liquid crystal mixtures
4213831	Aug 14, 1980	Goto, et al. Liquid crystal devices containing them
4213832	Aug 14, 1980	Goto, et al. Liquid crystal compositions
4213833	Aug	



Statistics

Citation

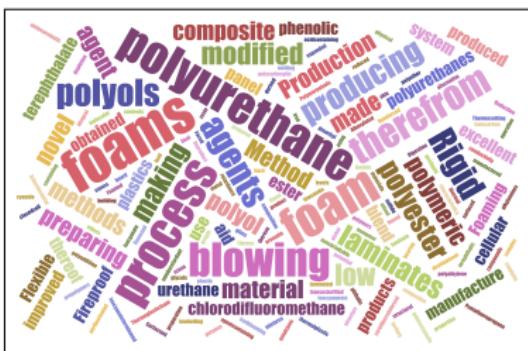
Two-mode Ns

Multiplication

Derived Na

Temporal Na

References





Main SPC island in SN5

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

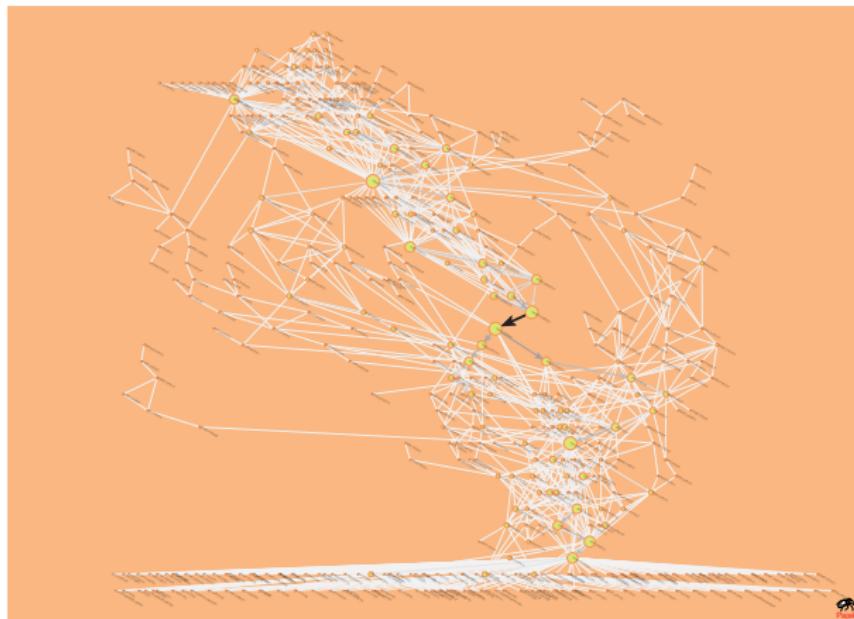
Multiplication

Derived Ns

Temporal Ns

References

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

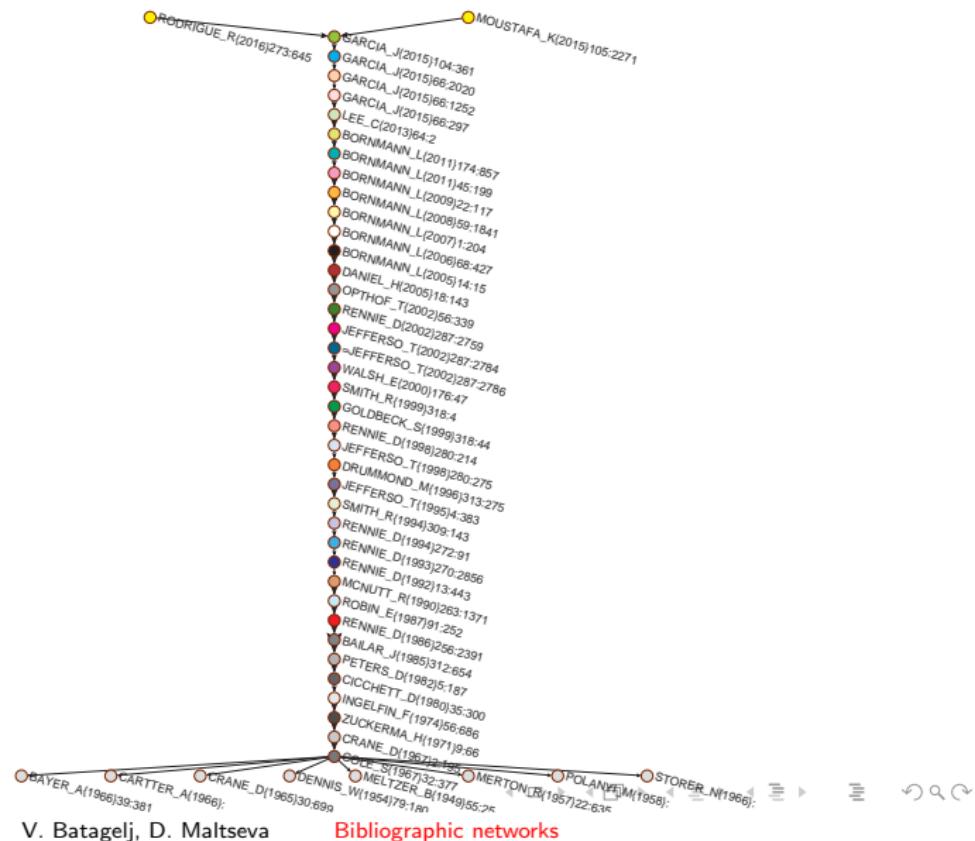
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





List of publications on main path

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

	year	first author	title	journal
	1949	Meltzer BN	The productivity of social scientists	AmJSociol
	1954	Dennis W	Bibliographies of eminent scientists	ScientificM
	1957	Merton RK	Priorities in scientific discovery - a chapter in the sociology of sci...	AmSocSciRev
	1958	Polanyi M	Personal Knowledge: Towards a Post-Critical Philosophy	UPChicago
Networks	1965	Crane D	Scientists at major and minor universities	AmSocSciRev
Statistics	1966	Bayer AE	Some correlates of citation measure of productivity in science	SociolEduc
Citation	1966	Storer NW	The Social System of Science	HRW
	1966	Cartter A	An Assessment of Quality in Graduate Education	ACE
	1967	Crane D	Gatekeepers of science - some factors affecting selection...	AmSociol
	1967	Cole S	Scientific output and recognition - study in operation of reward...	AmSocSciRev
	1971	Zuckerman H	Patterns of evaluation in science - institutionalisation, struct...	Minerva
Two-mode Ns	1974	Ingelfinger FJ	Peer review in biomedical publication	AmJMed
Multiplication	1980	Cicchetti DV	Reliability of reviews for the american-psychologist	AmPsychol
Derived Ns	1982	Peters DP	Peer-review practices of psychological journals - the fate...	BehavBrainS
Temporal Ns	1985	Bailar JC	Journal peer-review - the need for a research agenda	NewEnglJMe
References	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

	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
	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
	2002	Jefferson T	Measuring the quality of editorial peer review	Jama
	2005	Bornmann L	Committee peer review at an international research foundation	ResEvaluat
	2005	Daniel HD	Publications as a measure of scientific advancement and of...	LearnPubl
	2006	Bornmann L	Selecting scientific excellence through committee peer review	Scientometr
	2007	Bornmann L	Convergent validation of peer review decisions using the h index	JInformetr
	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
	2011	Bornmann L	Scientific Peer Review	AnnuRevIn
	2011	Bornmann L	A multilevel modelling approach to investigating the predictive...	JRStatSoc
	2013	Lee CJ	Bias in peer review	JAmSocInfl
	2015	Garcia JA	The Principal-Agent Problem in Peer Review	JAssocInflS
	2015	Garcia JA	Adverse selection of reviewers	JAssocInflS
	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

- 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

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

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

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

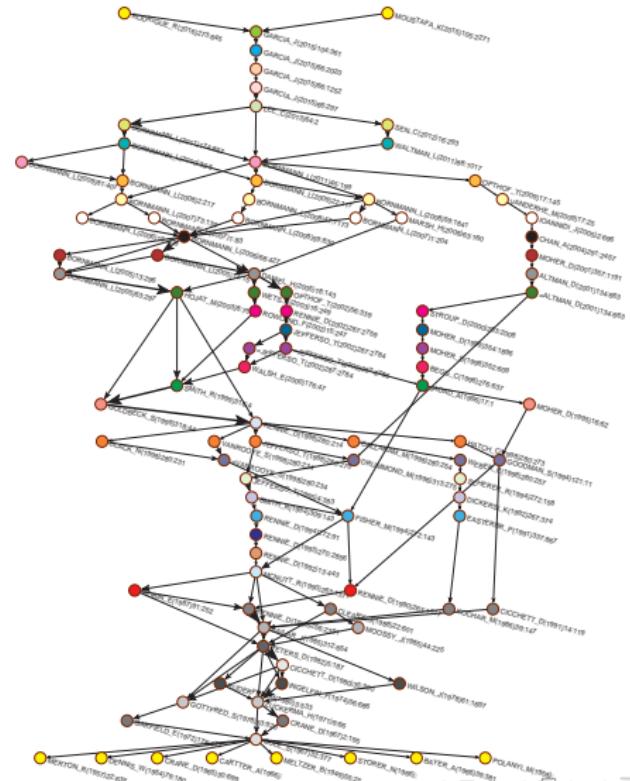
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





PEERE – SPC link islands [20 200]

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

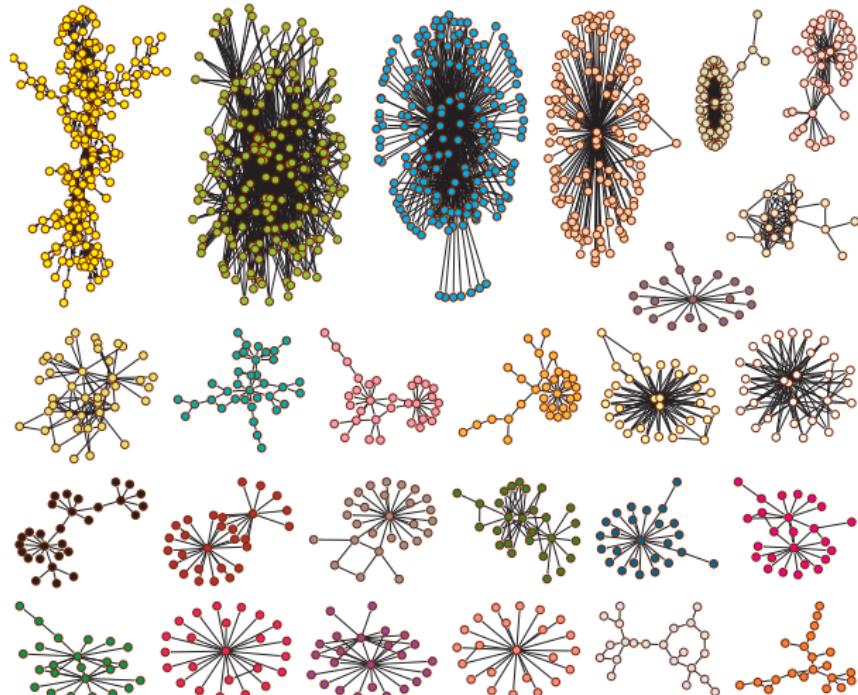
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





PEERE – SPC – Link island1

$$w_{max} = 0.297$$

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

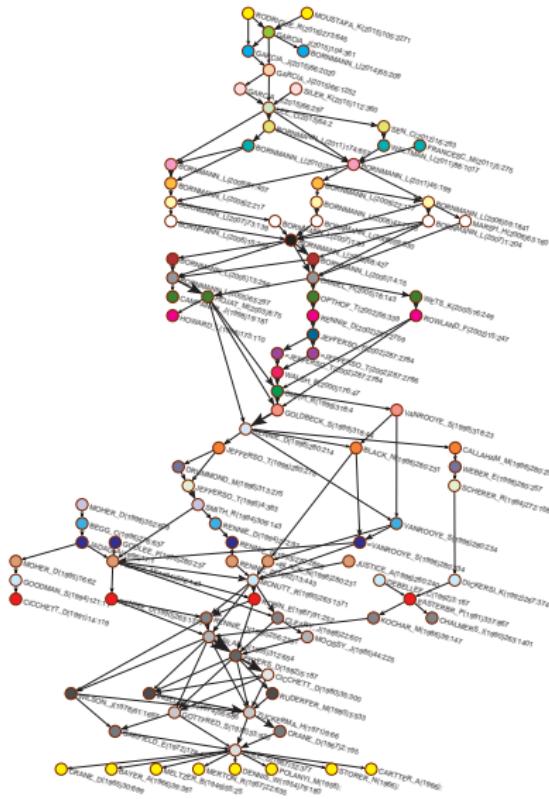
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



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

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

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

The corresponding cluster of nodes is produced manually using the Pajek's command

Network / Info / Vertex Label -> Vertex Number



Islands

positioning – valued networks in clustering

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

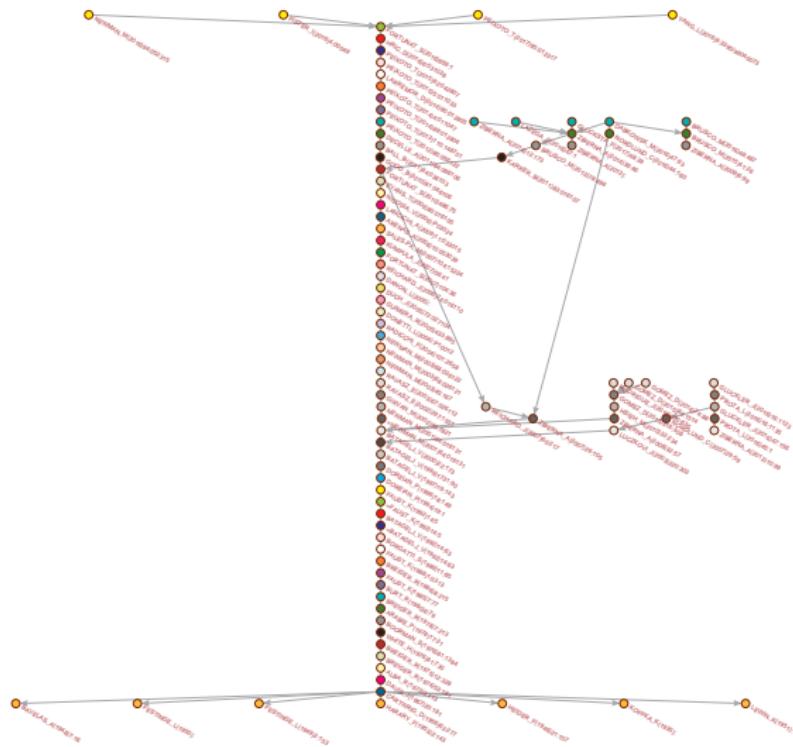
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Two-mode networks

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

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



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	X
4. Miss Brenda Rogers.....
5. Miss Charlotte McDowell.....	X
6. Miss Frances Anderson.....	X	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	X	X	X
12. Miss Katherine Rogers.....	X	X	X	X	X	X	X
13. Mrs. Sylvia Avondale.....	X	X	X	X	X	X	X
14. Mrs. Norm Fayette.....	X	X	X	X	X	X	X	X
15. Mrs. Helen Lloyd.....	X	X	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

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

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

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

- 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 $\forall v \in C_2 : \deg_{\mathcal{K}}(v) \geq q$;
- 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

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

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

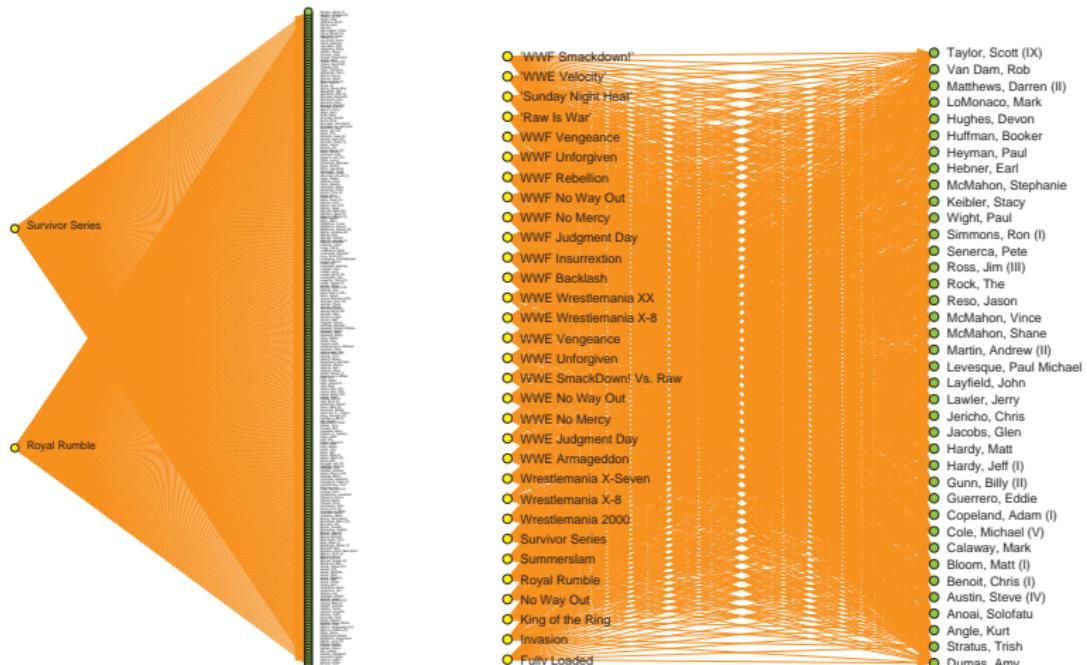
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





(2,516)-Hard core

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

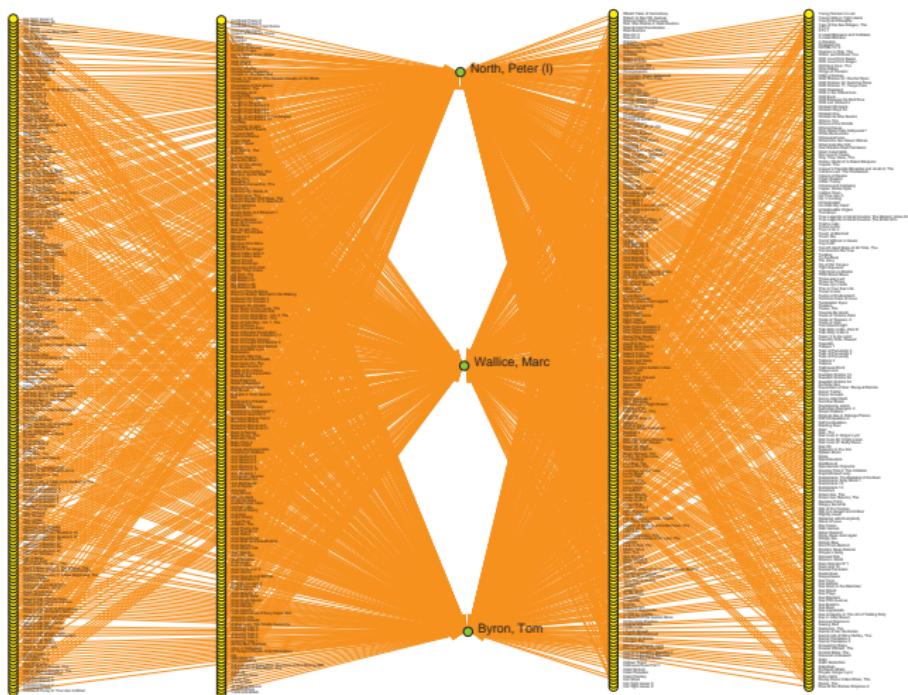
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



4-rings and analysis of two-mode networks

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

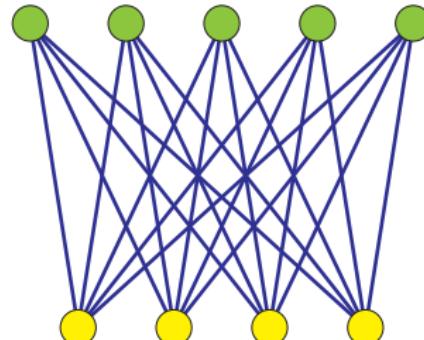
Temporal Ns

References

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

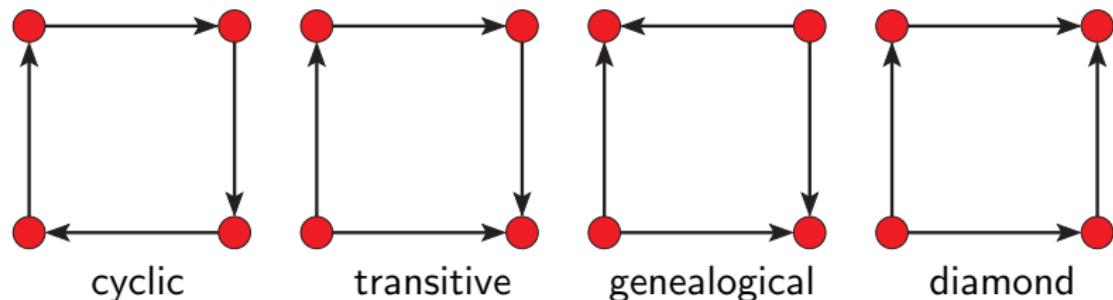
Multiplication

Derived Ns

Temporal Ns

References

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

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

Size	Freq	Size	Freq	Size	Freq	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

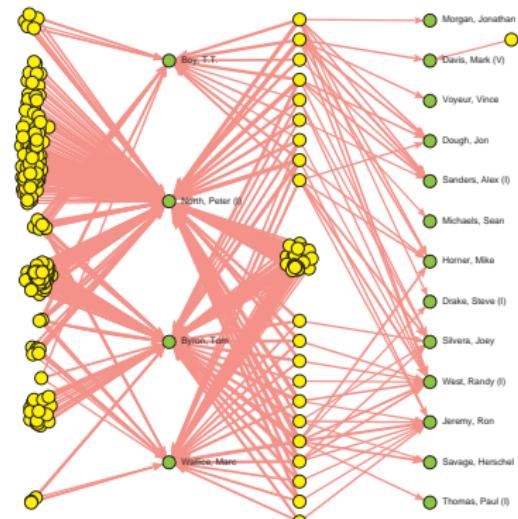
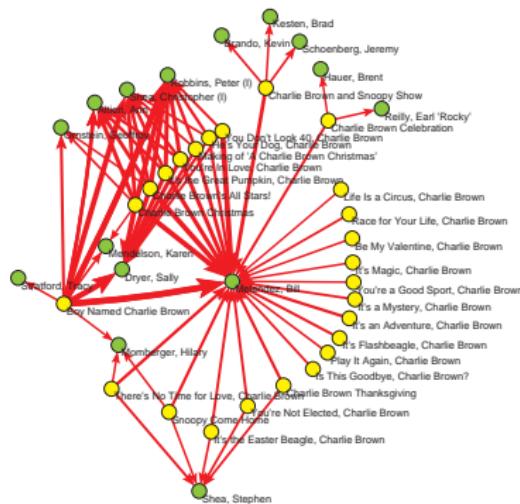
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Example: Islands for w_4

Mark Twain and Abid

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

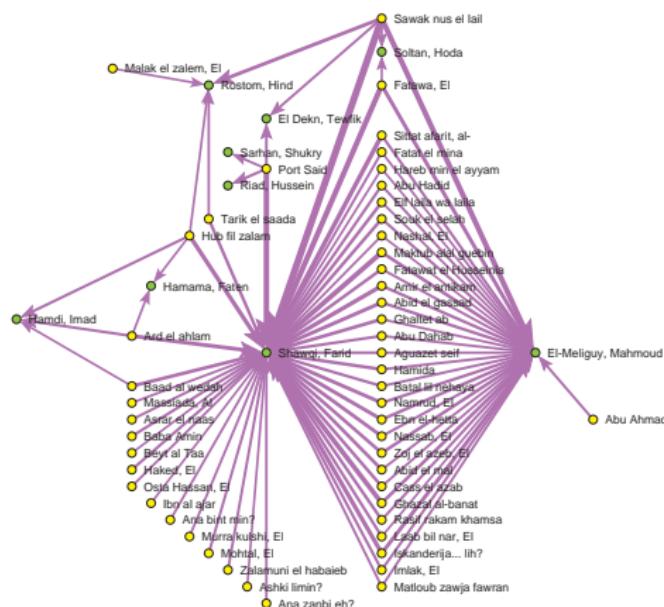
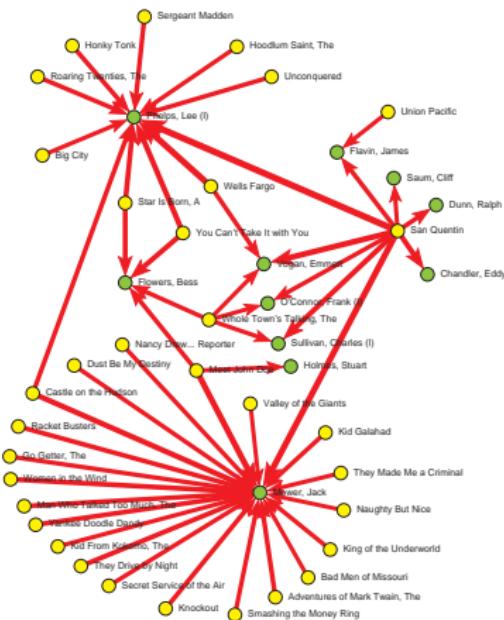
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



Example: Island for w_4

Polizeiruf 110 and Starkes Team

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

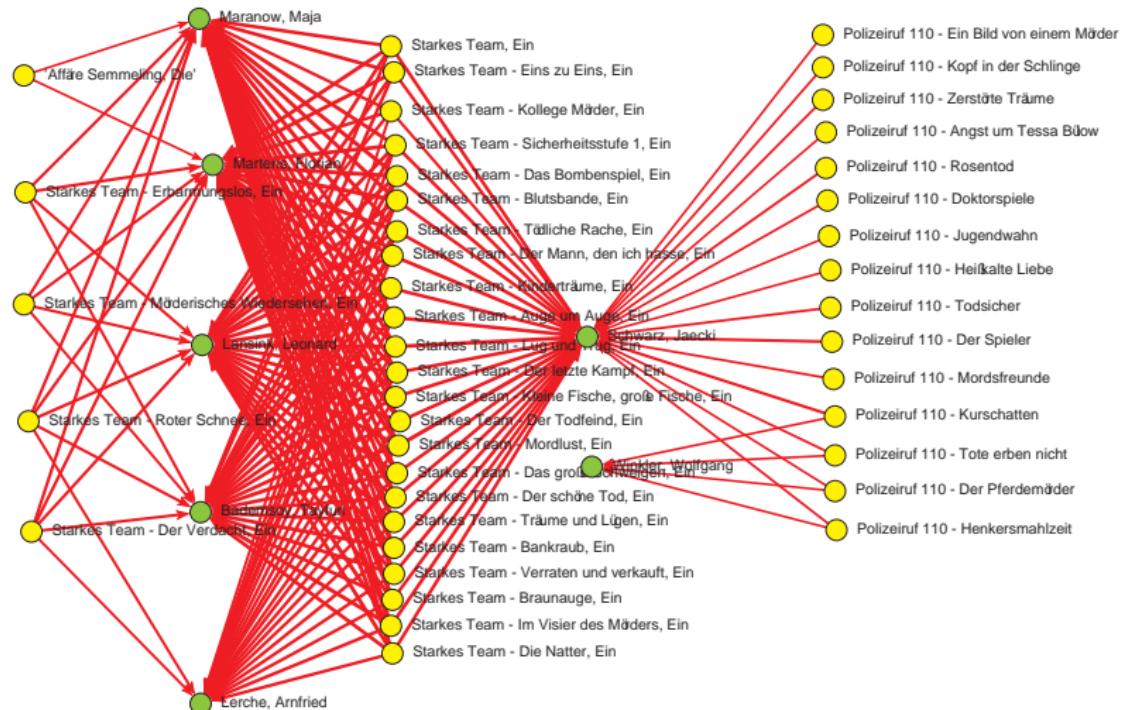
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Multiplication of networks

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

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

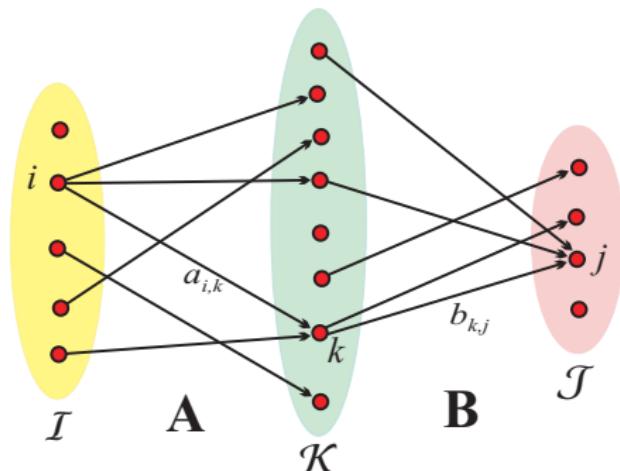
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



$$c_{i,j} = \sum_{k \in N_{\mathcal{A}}(i) \cap N_{\mathcal{B}}^-(j)} a_{i,k} \cdot b_{k,j}$$

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

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

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

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

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

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

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

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

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

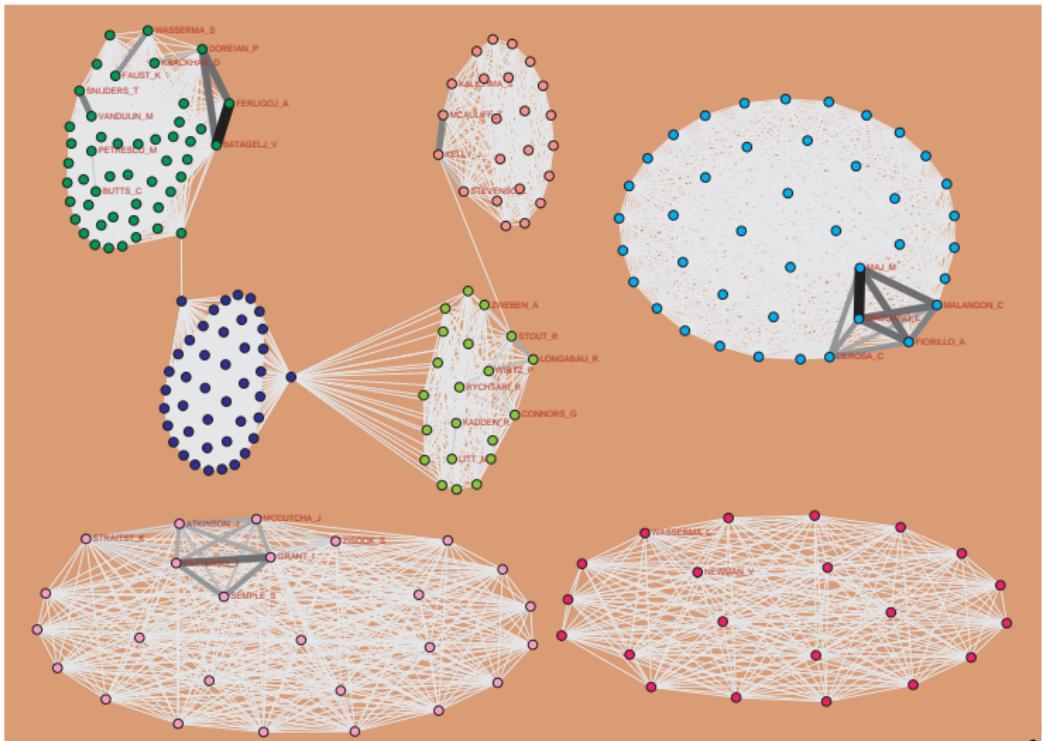
Multiplication

Derived Ns

Temporal Ns

References

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

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

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

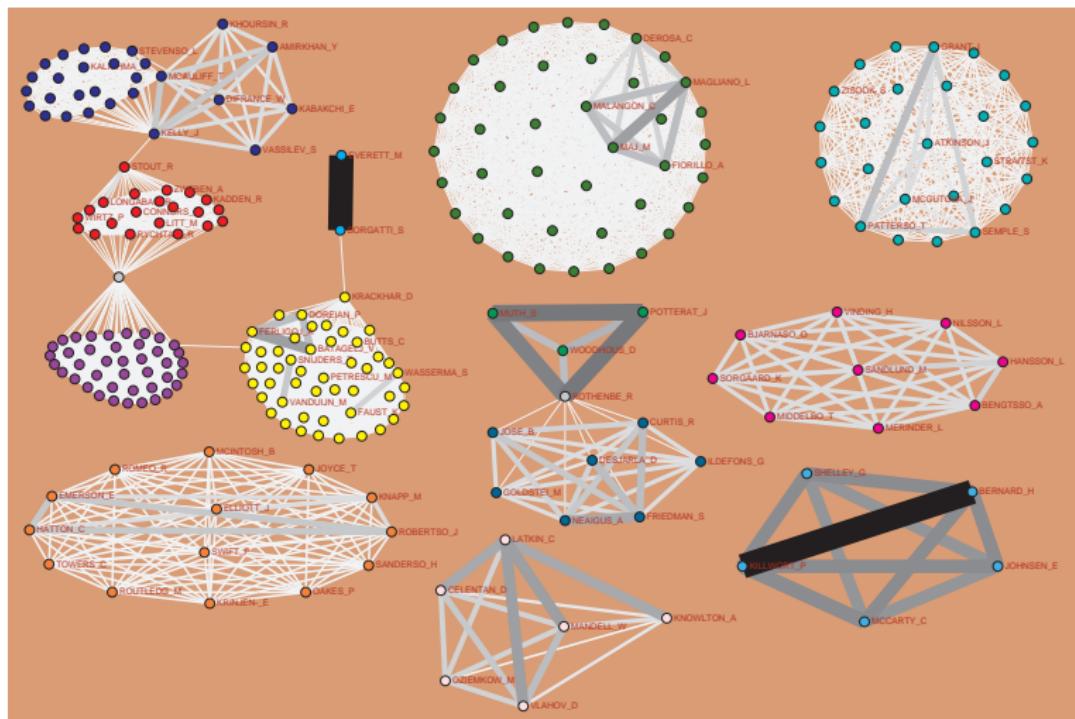
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Second co-authorship network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

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

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

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
	5	Marsden,P	29.42	37	0.205	30	Faust,K	9.72	18	0.460
Citation	6	Wellman,B	26.87	41	0.345	31	Batagelj,V	9.69	20	0.516
	7	Leydesdorff,L	24.37	35	0.304	32	Mizruchi,M	9.67	15	0.356
Two-mode Ns	8	White,H	23.50	33	0.288	33	[Anon]	9.00	9	0.000
Multiplication	9	Friedkin,N	20.00	23	0.130	34	Johnson,J	8.89	21	0.577
	10	Borgatti,S	19.20	41	0.532	35	Fararo,T	8.83	16	0.448
Derived Ns	11	Everett,M	16.92	31	0.454	36	Lazega,E	8.50	12	0.292
Temporal Ns	12	Litwin,H	16.00	21	0.238	37	Knoke,D	8.33	11	0.242
References	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



Outer product decomposition

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

For vectors $x = [x_1, x_2, \dots, x_n]$ and $y = [y_1, y_2, \dots, y_m]$ their *outer product* $x \circ y$ is defined as

$$x \circ y = [x_i \cdot y_j]_{n \times m}$$

Let $S_x = \sum_i x_i$ and $S_y = \sum_j y_j$ then

$$S = \sum_{i,j} (x \circ y)_{ij} = \sum_i \sum_j x_i \cdot y_j = \sum_i x_i \cdot \sum_j y_j = S_x \cdot S_y$$

Therefore, if $S_x = S_y = 1$ then also $S = 1$.

It is easy to verify that the *outer product decomposition* holds

$$\mathbf{H}_w = \mathbf{W}\mathbf{A}[w, \cdot] \circ \mathbf{W}\mathbf{K}[w, \cdot] \quad \text{and} \quad \mathbf{AK} = \mathbf{AW} \cdot \mathbf{WK} = \sum_w \mathbf{H}_w$$

For binary (weights) networks we have

$$\mathbf{H}_w = K_{N_{WA}(w), N_{WK}(w)}$$



$$\mathbf{W} \mathbf{A}^T \cdot \mathbf{W} \mathbf{K}$$

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

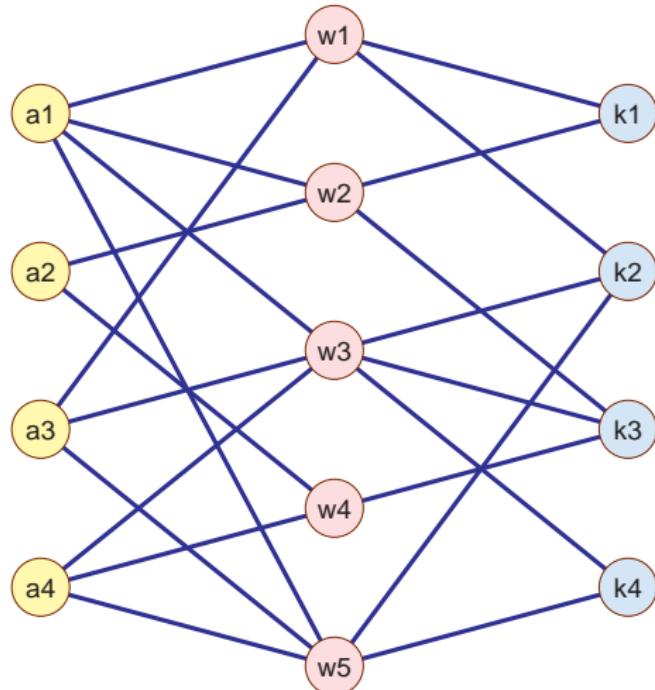
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Example

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

$$\mathbf{H} = \mathbf{AW} \cdot \mathbf{WK}, \quad \mathbf{WA} = \begin{matrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{matrix} \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix}, \quad \mathbf{WK} = \begin{matrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{matrix} \begin{bmatrix} k_1 & k_2 & k_3 & k_4 \end{bmatrix}$$

$$\mathbf{H} = \begin{matrix} k_1 & k_2 & k_3 & k_4 \end{matrix} = \mathbf{H}_1 + \mathbf{H}_2 + \mathbf{H}_3 + \mathbf{H}_4$$

$$= \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} + \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} + \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\mathbf{H}_3 = \begin{matrix} k_1 & k_2 & k_3 & k_4 \end{matrix} = \mathbf{H}_4 + \mathbf{H}_5$$

$$= \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} + \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$



Fractional approach

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

In general each term \mathbf{H}_w in the outer product decomposition has different value $S(\mathbf{H}_w) = \sum_{a,k} (\mathbf{H}_w)_{ak}$ leading to over-representation of works with large values.

To make the contributions of all works equal we can apply the *fractional* approach by normalizing the weights: setting $x' = x/S_x$ we get $S_{x'} = 1$ and therefore $S(\mathbf{H}'_w) = 1$ for all works w .

In the case of two-mode networks **WA** and **WK** we denote

$$S_w^{\mathbf{WA}} = \begin{cases} \sum_a \mathbf{WA}[w, a] & \text{outdeg}_{\mathbf{WA}}(w) > 0 \\ 1 & \text{outdeg}_{\mathbf{WA}}(w) = 0 \end{cases}$$

(and similarly $S_w^{\mathbf{WK}}$)



... fractional approach

Bibliographic
networks

V. Batagelj,
D. Maltseva

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

and define the *normalized* matrices

$$\mathbf{WAn} = \text{diag}\left(\frac{1}{S^{WA}}\right)\mathbf{WA}, \quad \mathbf{WKn} = \text{diag}\left(\frac{1}{S^{WK}}\right)\mathbf{WK}$$

Then the *normalized product* matrix is

$$\mathbf{AKt} = \mathbf{WAn}^T \cdot \mathbf{WKn}$$

Denoting $\mathbf{F}_w = \frac{1}{S_w^{WA} S_w^{WK}} \mathbf{H}_w$ the outer product decomposition gets form

$$\mathbf{AKt} = \sum_w \mathbf{F}_w$$



... fractional approach

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Since

$$T(\mathbf{F}_w) = \begin{cases} 1 & (\text{outdeg}_{\mathbf{WA}}(w) > 0) \wedge (\text{outdeg}_{\mathbf{WK}}(w) > 0) \\ 0 & \text{otherwise} \end{cases}$$

we have further

$$\sum_{a,k} \mathbf{F}[a, k] = \sum_{a,k} \sum_w \mathbf{F}_w[a, k] = \sum_w T(\mathbf{F}_w) = |W^+|$$

where $W^+ = \{w \in W : (\text{outdeg}_{\mathbf{WA}}(w) > 0) \wedge (\text{outdeg}_{\mathbf{WK}}(w) > 0)\}$.

In the network **AKt** the contribution of each work to the bibliography is 1. These contributions are redistributed to arcs from authors to keywords.



Example

normalized matrices

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

$$\text{diag}\left(\frac{1}{S^{WA}}\right) = \begin{bmatrix} w_1 & w_1 & w_2 & w_3 & w_4 & w_5 \\ w_2 & 1/2 & 0 & 0 & 0 & 0 \\ w_3 & 0 & 1/2 & 0 & 0 & 0 \\ w_4 & 0 & 0 & 1/3 & 0 & 0 \\ w_5 & 0 & 0 & 0 & 1/2 & 0 \\ & 0 & 0 & 0 & 0 & 1/3 \end{bmatrix}$$

$$\text{diag}\left(\frac{1}{S^{WK}}\right) = \begin{bmatrix} w_1 & w_1 & w_2 & w_3 & w_4 & w_5 \\ w_2 & 1/2 & 0 & 0 & 0 & 0 \\ w_3 & 0 & 1/2 & 0 & 0 & 0 \\ w_4 & 0 & 0 & 1/3 & 0 & 0 \\ w_5 & 0 & 0 & 0 & 1 & 0 \\ & 0 & 0 & 0 & 0 & 1/2 \end{bmatrix}$$

$$WAn = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ w_1 & 1/2 & 0 & 1/2 \\ w_2 & 1/2 & 1/2 & 0 \\ w_3 & 1/3 & 0 & 1/3 \\ w_4 & 0 & 1/2 & 0 \\ w_5 & 1/3 & 0 & 1/3 \end{bmatrix}, \quad WKn = \begin{bmatrix} k_1 & k_2 & k_3 & k_4 \\ w_1 & 1/2 & 1/2 & 0 \\ w_2 & 1/2 & 0 & 1/2 \\ w_3 & 0 & 1/3 & 1/3 \\ w_4 & 0 & 0 & 1 \\ w_5 & 0 & 1/2 & 0 \end{bmatrix}$$



... example

normalized matrices

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

$$\mathbf{F}_1 = \begin{matrix} & k_1 & k_2 & k_3 & k_4 \\ a_1 & \left[\begin{array}{cccc} 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right] \\ a_2 & \\ a_3 & \\ a_4 & \end{matrix} \quad \mathbf{F}_5 = \begin{matrix} & k_1 & k_2 & k_3 & k_4 \\ a_1 & \left[\begin{array}{cccc} 0 & 1/6 & 0 & 1/6 \\ 0 & 0 & 0 & 0 \\ 0 & 1/6 & 0 & 1/6 \\ 0 & 1/6 & 0 & 1/6 \end{array} \right] \\ a_2 & \\ a_3 & \\ a_4 & \end{matrix}$$

$$\mathbf{AKt} = \mathbf{F} = \begin{matrix} & k_1 & k_2 & k_3 & k_4 \\ a_1 & \left[\begin{array}{cccc} 0.50000 & 0.52778 & 0.36111 & 0.27778 \\ 0.25000 & 0.00000 & 0.75000 & 0.00000 \\ 0.25000 & 0.52778 & 0.11111 & 0.27778 \\ 0.00000 & 0.27778 & 0.61111 & 0.27778 \end{array} \right] \\ a_2 & \\ a_3 & \\ a_4 & \end{matrix}$$



Third co-authorship network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

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

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

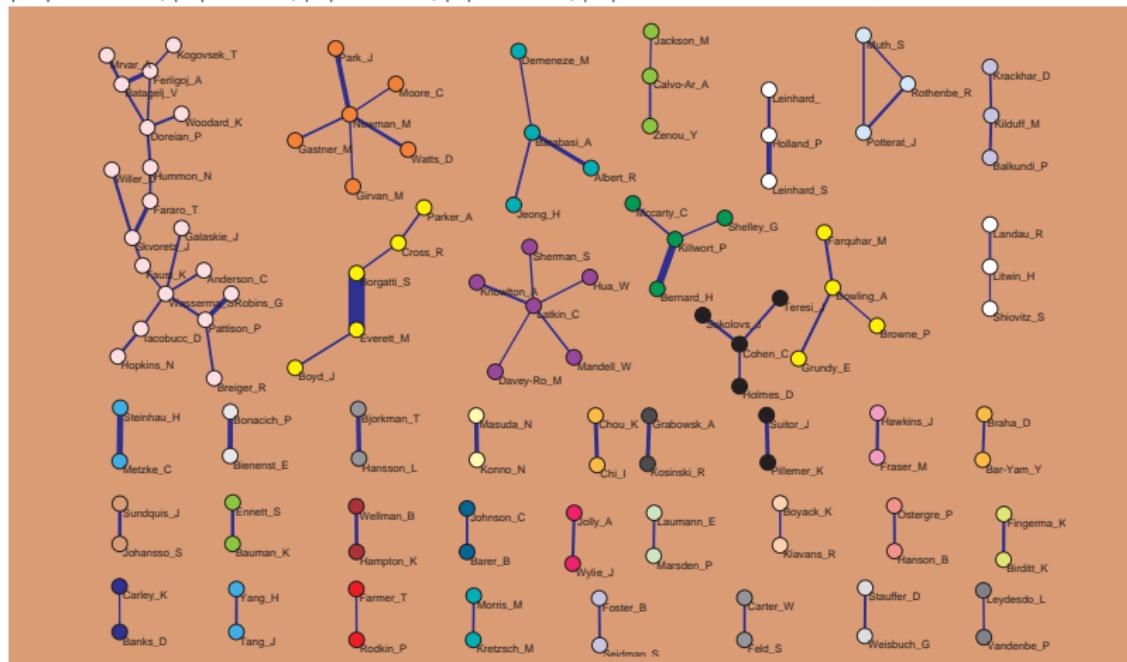
Multiplication

Derived Ns

Temporal Ns

References

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

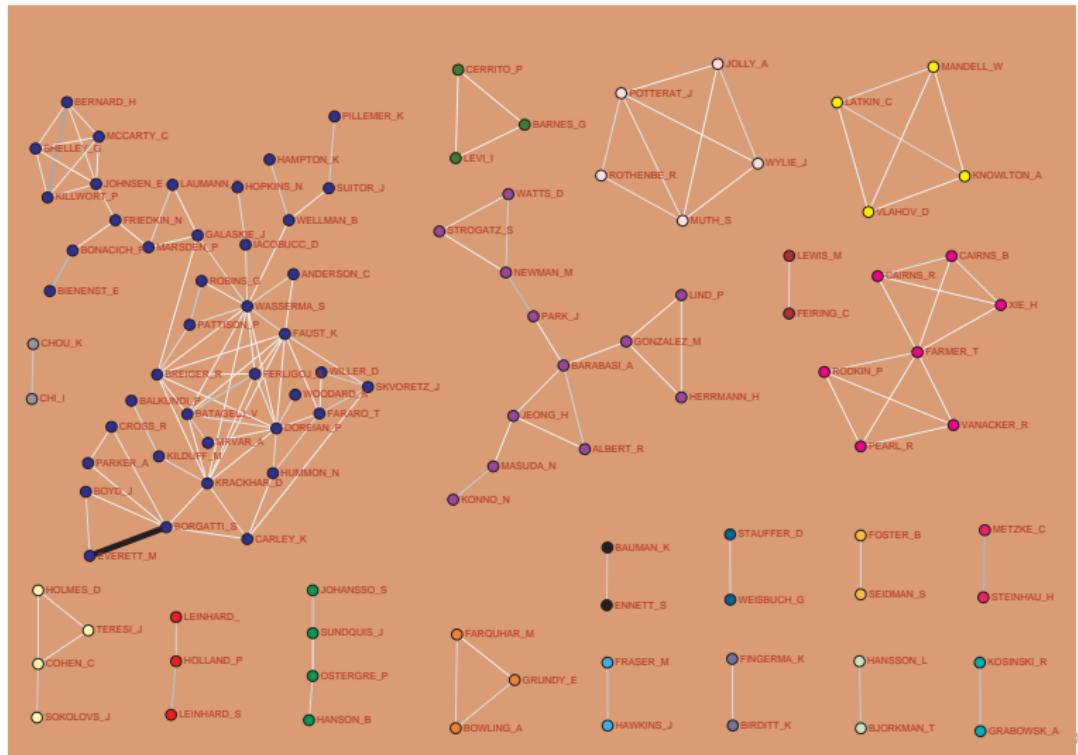
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





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

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

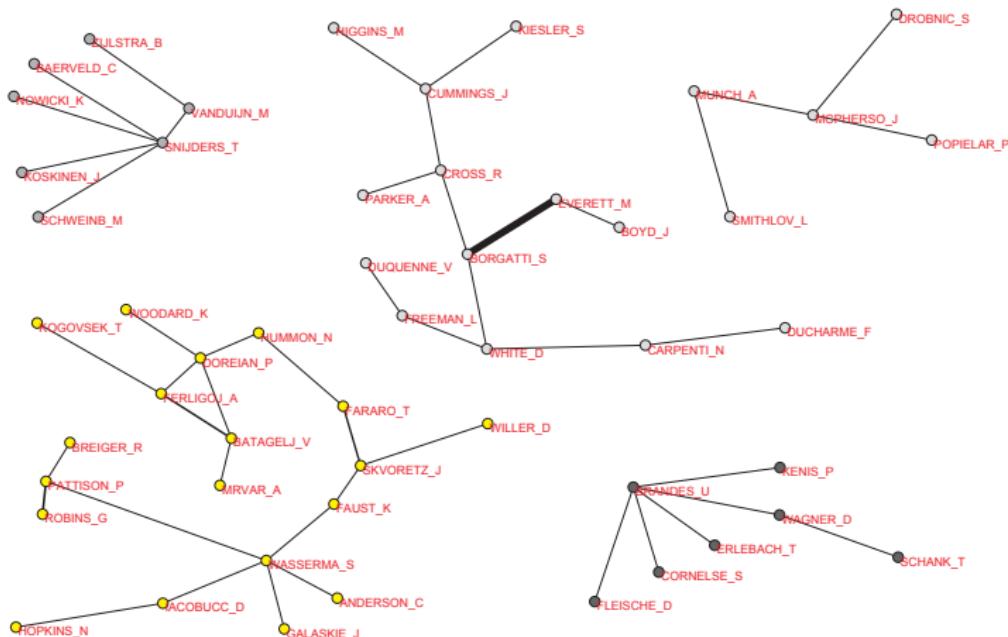
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Fourth co-authorship network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

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

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

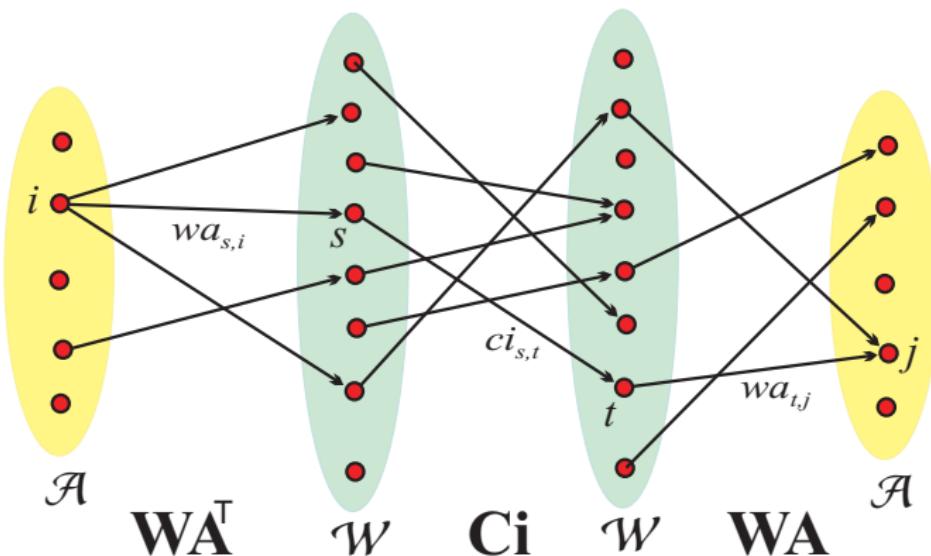
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



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

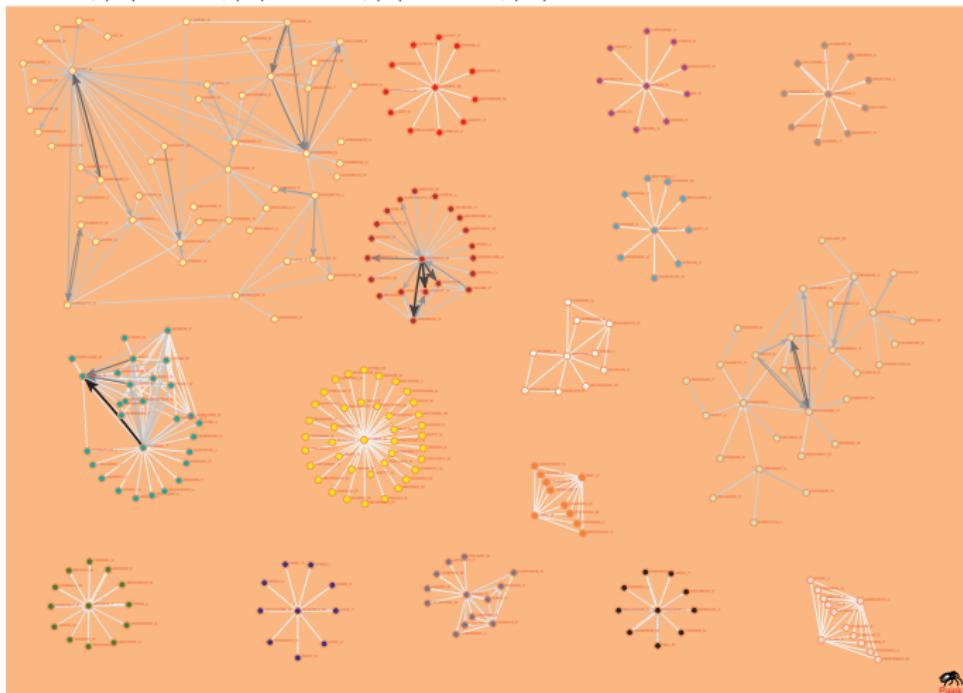
Multiplication

Derived Ns

Temporal Ns

References

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





Linking through a network

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Let a network \mathbf{S} links works to works. The derived network $\mathbf{WA}^T \cdot \mathbf{S} \cdot \mathbf{WA}$ links authors to authors *through* \mathbf{S} . Again, the normalization question has to be addressed. Among different options let us consider the derived networks defined as:

$$\mathbf{C} = \mathbf{WAn}^T \cdot \mathbf{S} \cdot \mathbf{WAn}$$

It is easy to verify that:

- if \mathbf{S} is symmetric, $\mathbf{S}^T = \mathbf{S}$, then also \mathbf{C} is symmetric, $\mathbf{C}^T = \mathbf{C}$;
$$\mathbf{C}^T = (\mathbf{WAn}^T \cdot \mathbf{S} \cdot \mathbf{WAn})^T = \mathbf{WAn}^T \cdot \mathbf{S}^T \cdot (\mathbf{WAn}^T)^T = \mathbf{C}$$
- if $W^+ = \{w \in W : \text{outdeg}_{\mathbf{WA}}(w) > 0\} = W$, the total of weights of \mathbf{S} is redistributed in \mathbf{C} :

$$T(\mathbf{C}) = \sum_{e \in L(\mathbf{C})} c(e) = \sum_{e \in L(\mathbf{S})} s(e) = T(\mathbf{S})$$



... Linking through a network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Since $\sum_{a \in A} wa[p, a] = \text{outdeg}_{WA}(p)$ and

$$wan[p, a] = \begin{cases} \frac{wa[p, a]}{\text{outdeg}_{WA}(p)} & \text{outdeg}_{WA}(p) > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{we get}$$

$$T(\mathbf{C}) = \sum_{e \in L(\mathbf{C})} c(e) = \sum_{a \in A} \sum_{b \in A} c[a, b] =$$

$$\sum_{a \in A} \sum_{b \in A} \sum_{p \in W} \sum_{q \in W} wan[p, a] \cdot s[p, q] \cdot wan[q, b] =$$

$$= \sum_{p \in W^+} \sum_{q \in W^+} \frac{s[p, q]}{\text{outdeg}_{WA}(p)\text{outdeg}_{WA}(q)} \sum_{a \in A} wa[p, a] \sum_{b \in A} wa[q, b] =$$

$$\sum_{p \in W^+} \sum_{q \in W^+} s[p, q]$$

and finally, if $W^+ = W$

$$\sum_{p \in W^+} \sum_{q \in W^+} s[p, q] = \sum_{e \in L(\mathbf{S})} s(e) = T(\mathbf{S})$$



... Linking through a network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

As special cases we get for normalized author's citation networks with
 $W^+ = W$: for $\mathbf{S} = \mathbf{Ci}$

$$\sum_{a \in A} \sum_{b \in A} c[a, b] = \sum_{p \in W} \sum_{q \in W} ci[p, q] = |\mathbf{Ci}|$$

and for $\mathbf{S} = \mathbf{Cin}$

$$\sum_{a \in A} \sum_{b \in A} c[a, b] = \sum_{p \in W} \sum_{q \in W: \text{outdeg}_{\mathbf{Ci}}(q) > 0} \frac{ci[p, q]}{\text{outdeg}_{\mathbf{Ci}}(p)} =$$

$$\sum_{q \in W: \text{outdeg}_{\mathbf{Ci}}(q) > 0} 1 = |W_{\mathbf{Ci}}^+|$$

Bibliographic Coupling

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

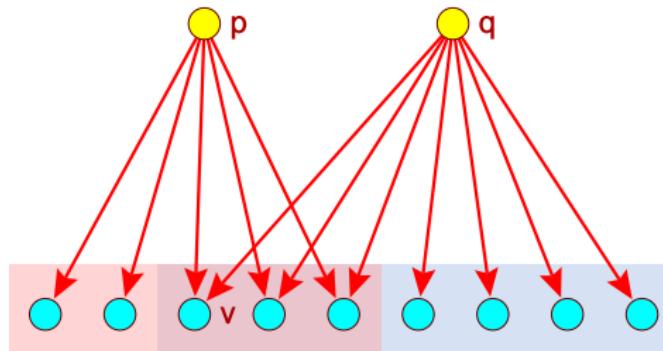
Multiplication

Derived Ns

Temporal Ns

References

Bibliographic coupling occurs when two works each cite a third work in their bibliographies, see figure. The idea was introduced by Kessler (1963) and has been used extensively since then. See figure where two citing works, p and q , are shown. Work p cites five works and q cites seven works. The key idea is that there are three works cited by both p and q . This suggests some content communality for the three works cited by both p and q . Having more works citing pairs of prior works increases the likelihood of them sharing content.





Bibliographic Coupling

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

We assume that the citation relation means
 $p \mathbf{Ci} q \equiv$ work p cites work q . Then the *bibliographic coupling* network **biCo** can be determined as

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

The weight $bico_{pq}$ is equal to the number of works cited by both works p and q ; $bico_{pq} = |\mathbf{Ci}(p) \cap \mathbf{Ci}(q)|$. Bibliographic coupling weights are symmetric: $bico_{pq} = bico_{qp}$:

$$\mathbf{biCo}^T = (\mathbf{Ci} \cdot \mathbf{Ci}^T)^T = \mathbf{Ci} \cdot \mathbf{Ci}^T = \mathbf{biCo}$$

Co-citation

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

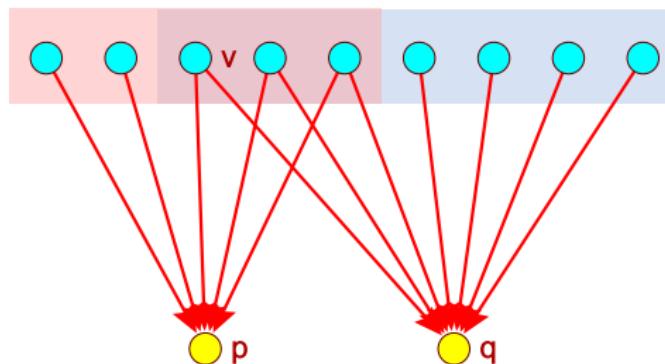
Multiplication

Derived Ns

Temporal Ns

References

Co-citation is a concept with strong parallels with bibliographic coupling (Small and Marshakova 1973), see Figure ??, right. The focus is on the extent to which works are co-cited by later works. The basic intuition is that the more earlier works are cited, the higher the likelihood that they have common content.





Bibliographic Coupling

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

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

$$\mathbf{coCi} = \mathbf{Ci}^T \cdot \mathbf{Ci}.$$

The weight $coci_{pq}$ is equal to the number of works citing both works p and q . The network **coCi** is symmetric $coci_{pq} = coci_{qp}$:

$$\mathbf{coCi}^T = (\mathbf{Ci}^T \cdot \mathbf{Ci})^T = \mathbf{Ci}^T \cdot \mathbf{Ci} = \mathbf{coCi}$$

An important property of co-citation is that $\mathbf{coCi}(\mathbf{Ci}) = \mathbf{biCo}(\mathbf{Ci}^T)$:

$$\mathbf{biCo}(\mathbf{Ci}^T) = \mathbf{Ci}^T \cdot (\mathbf{Ci}^T)^T = \mathbf{Ci}^T \cdot \mathbf{Ci} = \mathbf{coCi}(\mathbf{Ci})$$

Therefore the constructions proposed for bibliographic coupling can be applied also for co-citation.



Normalization

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

What about normalizations? Searching for the most coupled works we have again problems with works with many citations, especially with review papers. To neutralize their impact we can introduce normalized measures. The fractional approach works fine for normalized co-citation

$$\text{CoCit} = \mathbf{Cin}^T \cdot \mathbf{Cin}$$

where $\mathbf{Cin} = \mathbf{D} \cdot \mathbf{Ci}$ and $\mathbf{D} = \text{diag}\left(\frac{1}{\max(1, \text{outdeg}(p))}\right)$. $\mathbf{D}^T = \mathbf{D}$. In the normalized network every work has value 1 and it is equally distributed to all cited works.



Normalization of Bibliographic Coupling

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

The fractional approach can not be directly applied to bibliographic coupling – to get the outer product decomposition work we would need to normalize \mathbf{C}_i by columns – a cited work has value 1 which is distributed equally to the citing works – the most cited works give the least. This is against our intuition. To construct a reasonable measure we can proceed as follows. Let us first look at

$$\mathbf{biC} = \mathbf{C}_{in} \cdot \mathbf{C}_i^T$$

we have

$$\mathbf{biC} = (\mathbf{D} \cdot \mathbf{C}_i) \cdot \mathbf{C}_i^T = \mathbf{D} \cdot \mathbf{biCo}$$

$$\mathbf{biC}^T = (\mathbf{D} \cdot \mathbf{biCo})^T = \mathbf{biCo}^T \cdot \mathbf{D}^T = \mathbf{biCo} \cdot \mathbf{D}$$



Normalization of Bibliographic Coupling

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

For $\text{Ci}(p) \neq \emptyset$ and $\text{Ci}(q) \neq \emptyset$ it holds

$$\mathbf{biC}_{pq} = \frac{|\text{Ci}(p) \cap \text{Ci}(q)|}{|\text{Ci}(p)|} \quad \text{and} \quad \mathbf{biC}_{qp} = \frac{|\text{Ci}(p) \cap \text{Ci}(q)|}{|\text{Ci}(q)|} = \mathbf{biC}_{pq}^T$$

and $\mathbf{biC}_{pq} \in [0, 1]$. \mathbf{biC}_{pq} is the proportion of its references that the work p shares with the work q . The network \mathbf{biC} is not symmetric.

We have different options to construct normalized symmetric measures such as

$$\mathbf{biCoa}_{pq} = \frac{1}{2}(\mathbf{biC}_{pq} + \mathbf{biC}_{qp}) \quad \text{Average}$$

$$\mathbf{biCom}_{pq} = \min(\mathbf{biC}_{pq}, \mathbf{biC}_{qp}) \quad \text{Minimum}$$

$$\mathbf{biCoM}_{pq} = \max(\mathbf{biC}_{pq}, \mathbf{biC}_{qp}) \quad \text{Maximum}$$



Normalization of Bibliographic Coupling

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

or, may be more interesting

$$\text{biCog}_{pq} = \sqrt{\text{biC}_{pq} \cdot \text{biC}_{qp}} = \frac{|\text{Ci}(p) \cap \text{Ci}(q)|}{\sqrt{|\text{Ci}(p)| \cdot |\text{Ci}(q)|}}$$

Geometric mean
Salton cosinus

$$\text{biCoh}_{pq} = 2 \cdot (\text{biC}_{pq}^{-1} + \text{biC}_{qp}^{-1})^{-1} = \frac{2|\text{Ci}(p) \cap \text{Ci}(q)|}{|\text{Ci}(p)| + |\text{Ci}(q)|}$$

Harmonic mean

$$\text{biCoj}_{pq} = (\text{biC}_{pq}^{-1} + \text{biC}_{qp}^{-1} - 1)^{-1} = \frac{|\text{Ci}(p) \cap \text{Ci}(q)|}{|\text{Ci}(p) \cup \text{Ci}(q)|}$$

Jaccard index

All these measures are similarities.

It is easy to verify that $\text{biCoj}_{pq} \in [0, 1]$ and: $\text{biCoj}_{pq} = 1$ iff the works p and q are referencing the same works, $\text{Ci}(p) = \text{Ci}(q)$.



Normalization of Bibliographic Coupling

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

From $m \leq H \leq G \leq A \leq M$ and $J \leq m$,
 $(\frac{|P \cap Q|}{|P \cup Q|} \leq \min(\frac{|P \cap Q|}{|P|}, \frac{|P \cap Q|}{|Q|}))$ we get

$$\mathbf{biCoj}_{pq} \leq \mathbf{biCom}_{pq} \leq \mathbf{biCoh}_{pq} \leq \mathbf{biCog}_{pq} \leq \mathbf{biCoa}_{pq} \leq \mathbf{biCoM}_{pq}$$

The equalities hold iff $\mathbf{Ci}(p) = \mathbf{Ci}(q)$.

To get a dissimilarity we can use transformations $dis = 1 - sim$ or
 $dis = \frac{1}{sim} - 1$ or $dis = -\log sim$. For example

$$\mathbf{biCod}_{pq} = 1 - \mathbf{biCoj}_{pq} = \frac{|\mathbf{Ci}(p) \oplus \mathbf{Ci}(q)|}{|\mathbf{Ci}(p) \cup \mathbf{Ci}(q)|} \quad \text{Jaccard distance}$$

where \oplus denotes the symmetric difference of sets.



Bibliographic Coupling, Co-citation and Linking through a network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Bibliographic coupling and co-citation networks are linking works to works. To get linking between authors, journals or keywords considering citation similarity we can apply the construction from *Linking through a network* to the normalized co-citation or bibliographic coupling network.



Temporal quantities

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

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

```
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

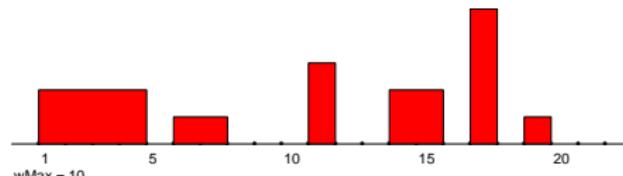
Multiplication

Derived Ns

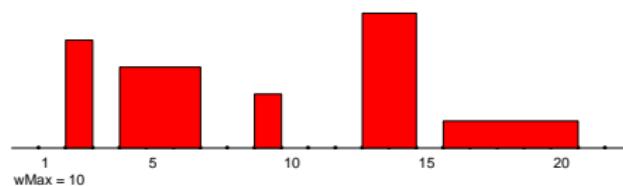
Temporal Ns

References

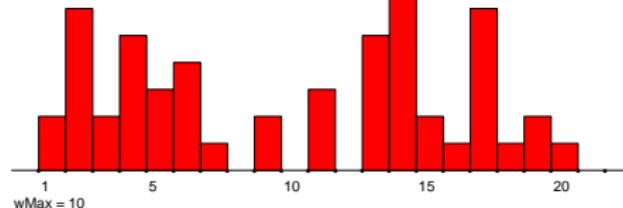
$a :$



$b :$



$a + b :$





Multiplication of temporal quantities.

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

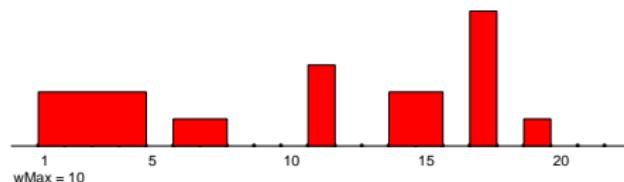
Multiplication

Derived Ns

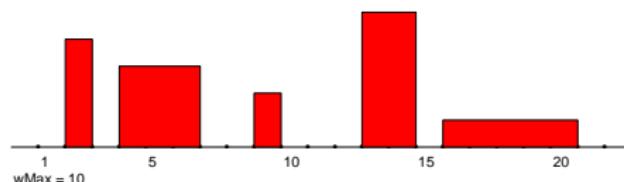
Temporal Ns

References

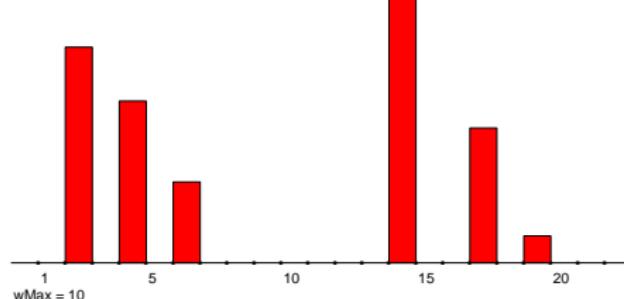
$a :$



$b :$



$a \cdot b :$





Temporal affiliation networks

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

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

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



Cumulative properties

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

In general a temporal quantity a is called *cumulative* iff it has for $t, t' \in \mathcal{T}$ the property

$$t \in T_a \wedge t' > t \Rightarrow t' \in T_a \wedge a(t') \geq a(t)$$

A sum and product (over combinatorial semiring) of cumulative temporal quantities are cumulative temporal quantities.

For a temporal quantity $a = [(s_i, f_i, v_i)]_{i \in 1..k}$ its *cumulative* $\text{cum}(a)$ is defined as

$$\text{cum}(a) = [(s_i, s_{i+1}, V_i)]_{i \in 1..k}$$

where $s_{k+1} = \text{last}$ and $V_i = \sum_{j=1}^i v_j$.

A temporal network is cumulative for a weight w iff all its values are cumulative.



Converting Pajek net and clu files into temporal network in netsJSON

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

The Python code for creating temporal networks from Pajek files for the peer review data is given here.

To set up an environment for computing our examples we have to put in the directory `gdir` Python files (`Nets.py`, `TQ.py`, `search.py`, `coloring.py`, `IndexMinPQ.py`) from the library `Nets`, and in the subdirectory `cdir` the files `TQchart.html`, `d3.v3.min.js` and `barData.js`. The directory `ndir` contains the network data and the directory `wdir` contains the results.



... Converting Pajek net and clu files into temporal network in netsJSON

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

```
gdir = 'c:/path/Nets'
wdir = 'c:/path/Test/peere'
ndir = 'c:/path/WoS/peere2'
cdir = 'c:/path/Nets/chart'
import sys, os, datetime, json
sys.path = [gdir]+sys.path; os.chdir(wdir)
from TQ import *
from Nets import Network as N
net = ndir+"/WAd.net"
clu = ndir+"/Yead.clu"
t1 = datetime.datetime.now(); print("started: ",t1.ctime(),"\n")
WAc = N.twoMode2netsJSON(clu,net,'WAcum.json',instant=False)
t2 = datetime.datetime.now()
print("\nconverted to cumulative TN: ",t2.ctime(),"\netime used: ", t2-t1)
WAi = N.twoMode2netsJSON(clu,net,'WAins.json',instant=True)
t3 = datetime.datetime.now()
print("\nconverted to instantaneous TN: ",t3.ctime(),"\netime used: ", t3-t2)
cit = ndir+"/CiteD.net"
Citei = N.oneMode2netsJSON(clu,cit,'CiteIns.json',instant=True)
t4 = datetime.datetime.now()
print("\nconverted to instantaneous TN: ",t4.ctime(),"\netime used: ", t4-t3)
ia = WAi.Index()
ic = Citei.Index()
```



Temporal properties

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Let \mathbf{N} be a temporal network on $E \times P$. On it we can define some interesting temporal quantities such as *in-sum*:

$$iS(\mathbf{N}, p) = \sum_{e \in E} n_{ep}$$

and *out-sum*:

$$oS(\mathbf{N}, e) = \sum_{p \in P} n_{ep}$$

In a special case where $\mathbf{N} \equiv \mathbf{WAI}$ we get the *productivity of an author* a

$pr(a) = iS(\mathbf{WAI}, a) =$ number of publications of the author a by year

and for $\mathbf{N} \equiv \mathbf{WAc}$ we get the *cumulative productivity of an author* a

$cpr(a) = iS(\mathbf{WAc}, a) =$ cumulative number of publications of a by year.

It holds $cpr(a) = cum(pr(a))$.



Group productivity

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

The productivity of an author can be extended to the *productivity of a group of authors* C

$$pr(C) = \sum_{a \in C} pr(a) = \sum_{a \in C} iS(\mathbf{WAn}, a)$$

There is a problem with the productivity of a group. In the case when two authors from a group co-authored the same paper it is counted twice. To account for a “real” contribution of each author the fractional approach is used. It is based on normalized networks (matrices) – in the case of co-authorship on

$$n(\mathbf{WA}) = \mathbf{WAn} = [wan_{wa}]$$

$$wan_{wa} = \frac{wa_{wa}}{\max(1, \text{outdeg}_{\mathbf{WA}}(w))}.$$

This leads to the *fractional productivity of an author* a

$$fpr(a) = iS(\mathbf{WAni}, a) = \text{fractional contribution of publications of } a \text{ by year}$$



Temporal properties in networks on peer review

Productivity and cumulative productivity of Lutz Bornmann

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

In the analysis of the ordinary authorship network **WA** we get that Lutz Bornmann is the author who wrote the largest number, 61, of works on peer review. To see the dynamics of his publishing we compute his productivity

$$pr = [(2005, 2006, 4), (2006, 2007, 3), (2007, 2008, 4), (2008, 2009, 9), (2009, 2010, 4), (2010, 2011, 14), (2011, 2012, 5), (2012, 2013, 7), (2013, 2014, 2), (2014, 2015, 3), (2015, 2016, 6)]$$

see the top of the figure. The corresponding cumulative productivity is

$$cpr = [(2005, 2006, 4), (2006, 2007, 7), (2007, 2008, 11), (2008, 2009, 20), (2009, 2010, 24), (2010, 2011, 38), (2011, 2012, 43), (2012, 2013, 50), (2013, 2014, 52), (2014, 2015, 55), (2015, 2017, 61)]$$

see the bottom of the figure. Note that $cpr = \text{cum}(pr)$: $7 = 4 + 3$, $11 = 4 + 3 + 4$, ...



Temporal properties in networks on peer review

Productivity and cumulative productivity of Lutz Bornmann

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

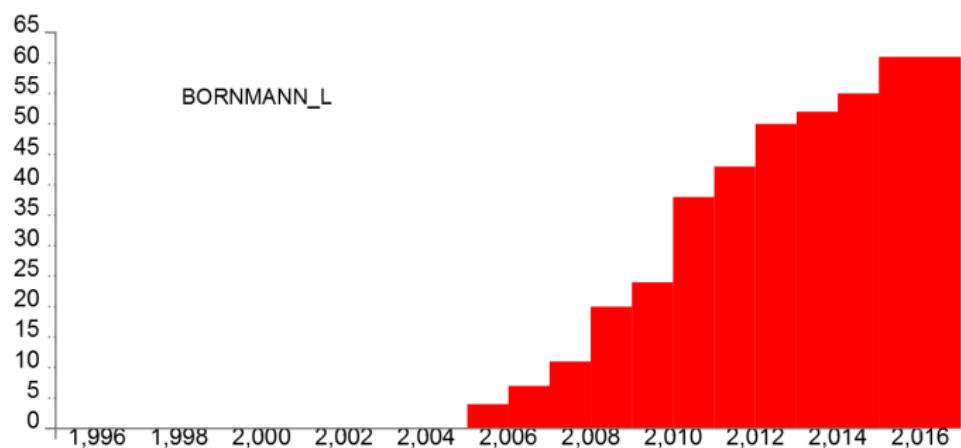
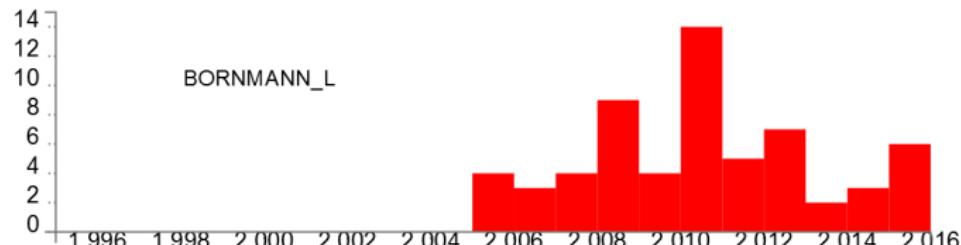
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Temporal properties in networks on peer review

Fractional productivity of Lutz Bornmann

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

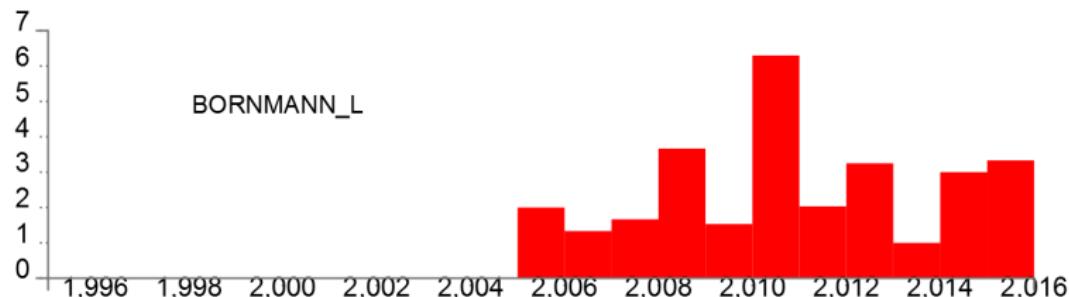
Derived Ns

Temporal Ns

References

$fpr = [(2005, 2006, 2.0), (2006, 2007, 1.333), (2007, 2008, 1.667), (2008, 2009, 3.667), (2009, 2010, 1.533), (2010, 2011, 6.3), (2011, 2012, 2.033), (2012, 2013, 3.25), (2013, 2014, 1.0), (2014, 2015, 3.0), (2015, 2016, 3.333)]$

see the figure. For the Python code see the next slide.





Productivities of authors – Python code

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

```
>>> tit = 'BORNMANN_L'; b = ia[tit]
>>> pr = WAi.TQnetInSum(b)
>>> pr
[(2005, 2006, 4), (2006, 2007, 3), (2007, 2008, 4), (2008, 2009, 9),...]
>>> TQ.TqSummary(pr)
(1900, 2017, 0, 14)
>>> TQmax = 15; Tmin = 1995; Tmax = 2016; w = 600; h = 150
>>> N.TQshow(pr,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='red')
>>> cpr = WAc.TQnetInSum(b)
>>> cpr
[(2005, 2006, 4), (2006, 2007, 7), (2007, 2008, 11), (2008, 2009, 20),...]
>>> TQmax = 65; Tmin = 1995; Tmax = 2016; w = 600; h = 250
>>> N.TQshow(cpr,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='red')
>>> WAni = WAi.TQnormal()
>>> fpr = WAni.TQnetInSum(b)
>>> fpr
[(2006, 2007, 1.3333333333333333), (2007, 2008, 1.6666666666666665),...]
>>> TQmax = 7; Tmin = 1995; Tmax = 2016; w = 600; h = 150
>>> N.TQshow(fpr,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='red')
```



Citations to a paper through years

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

In the citation network **Cite** for the peer review bibliography the most cited, 164, paper is Peters, D. P., Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences*, 5(2), 187-255. The temporal quantity $citP = iS(\mathbf{Cii}, \text{PETERS_D}(1982)5:187)$ describes the number of citations to this paper through years.

$$citP = [(1982, 1983, 1), (1983, 1984, 4), (1984, 1986, 3), (1986, 1987, 2), (1987, 1988, 3), (1988, 1989, 5), (1989, 1990, 2), (1990, 1991, 4), (1991, 1992, 5), (1992, 1993, 3), (1993, 1994, 8), (1994, 1996, 5), (1996, 1997, 6), (1997, 1998, 1), (1998, 1999, 5), (1999, 2000, 2), (2000, 2001, 1), (2001, 2002, 2), (2002, 2003, 4), (2003, 2004, 5), (2004, 2005, 4), (2005, 2006, 6), (2006, 2008, 5), (2008, 2009, 3), (2009, 2010, 9), (2010, 2011, 7), (2011, 2012, 10), (2012, 2013, 11), (2013, 2014, 4), (2014, 2015, 5), (2015, 2016, 14), (2016, 2017, 2)]$$

See the figure at the next slide.



Peer review

Citations to Peters

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

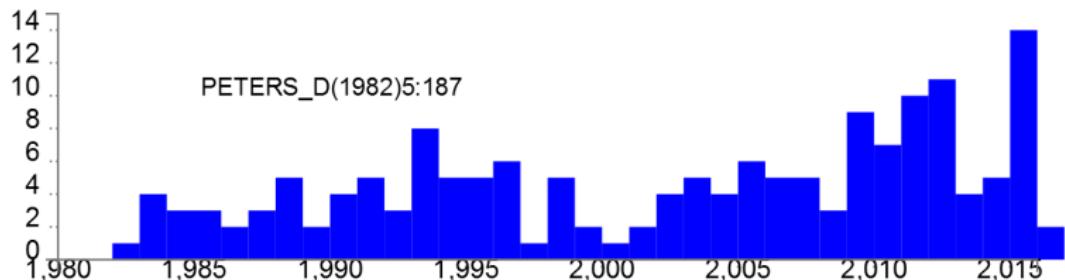
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



Similarly we could look at the number of works by year

$wy = \sum_{j \in J} iS(\mathbf{WJi}, j)$, the popularity of a keyword k :

$pop(k) = iS(\mathbf{WKi}, k)$, etc.



Peer review

Citations to Hirsch

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

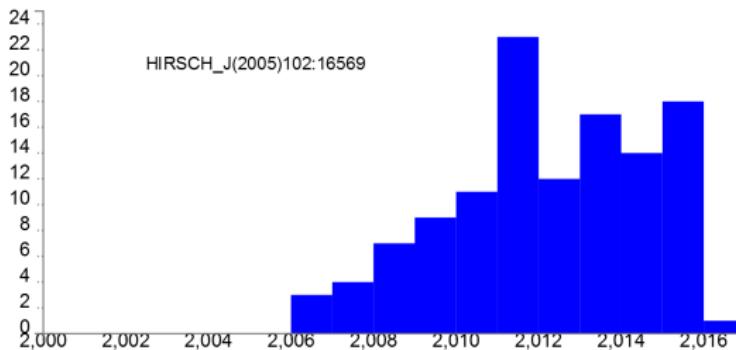
Temporal Ns

References

Another well known paper is Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A. 2005 Nov 15;102(46):16569-72 with 119 citations and
 $citH = iS(\mathbf{Cii}, \text{HIRSCH_J}(2005)102:16569)$

$citH = [(2006, 2007, 3), (2007, 2008, 4), (2008, 2009, 7), (2009, 2010, 9),$
 $(2010, 2011, 11), (2011, 2012, 23), (2012, 2013, 12), (2013, 2014, 17),$
 $(2014, 2015, 14), (2015, 2016, 18), (2016, 2017, 1)]$

For the Python code see the next slide.





Citations between works – Python code

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

```
>>> tit = 'PETERS_D(1982)5:187'; c = ic[tit]
>>> ci = Citei.TQnetInSum(c)
>>> ci
[(1982, 1983, 1), (1983, 1984, 4), (1984, 1986, 3), (1986, 1987, 2), ...]
>>> TQmax = 15; Tmin = 1980; Tmax = 2016; w = 600; h = 150
>>> N.TQshow(ci,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='blue')
>>> tit = 'HIRSCH_J(2005)102:16569'; c = ic[tit]
>>> ci = Citei.TQnetInSum(c)
>>> ci
[(2005, 2006, 0), (2006, 2007, 3), (2007, 2008, 4), (2008, 2009, 7), ...]
>>> TQmax = 25; Tmin = 2000; Tmax = 2017; w = 600; h = 250
>>> N.TQshow(ci,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='blue')
```



Main journals publishing on peer review

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

To identify the main journals publishing on peer review, see Appendix ??, we determined first the temporal inSums Jt in the network **JWi** for all journals. An entry $Jt(j) = iS(\mathbf{JWi}, j)$ contains the temporal quantity counting the number of papers on peer review published in the journal j in each year. Because most of frequencies are small (one digit numbers) we decided to change the time scale (granularity) to time intervals: 1: 1900-1970, 2: 1971-1980, 3: 1981-1990, 4: 1991-2000, 5: 2001-2005, 6: 2006-2010, 7: 2011-2015. The recoded table Jt is labeled Jr . For the table Jr we determined for each time interval three the most frequently used journals – they are listed on the right side of Figure ?? . The corresponding data were exported as `journals.csv` and visualized using R. The picture on the left side presents the trajectories of relative importance (journal's frequency divided with the maximum frequency on the interval) for the selected journals.



... Main journals publishing on peer review

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

The papers on peer review (refereeing) published till 1970 appeared most often in J ASSOC OFF AGR CHEM. Till 2005 the dominant journals were JAMA, SCIENCE, NATURE, BRIT MED J, and LANCET (general medical and science journals). In the period 2006-2010 the leading role was overtaken by a specialized journal SCIENTOMETRICS. In the last period 2011-2015 the primate is shifted to the mega-journals BMJ OPEN and PLOS ONE. Note that the frequencies for SCIENTOMETRICS are 3: 6, 4: 25, 5: 18, 6: 44, 7: 78 and in the period 2011-2015 489 papers on peer review were published in BMJ OPEN.

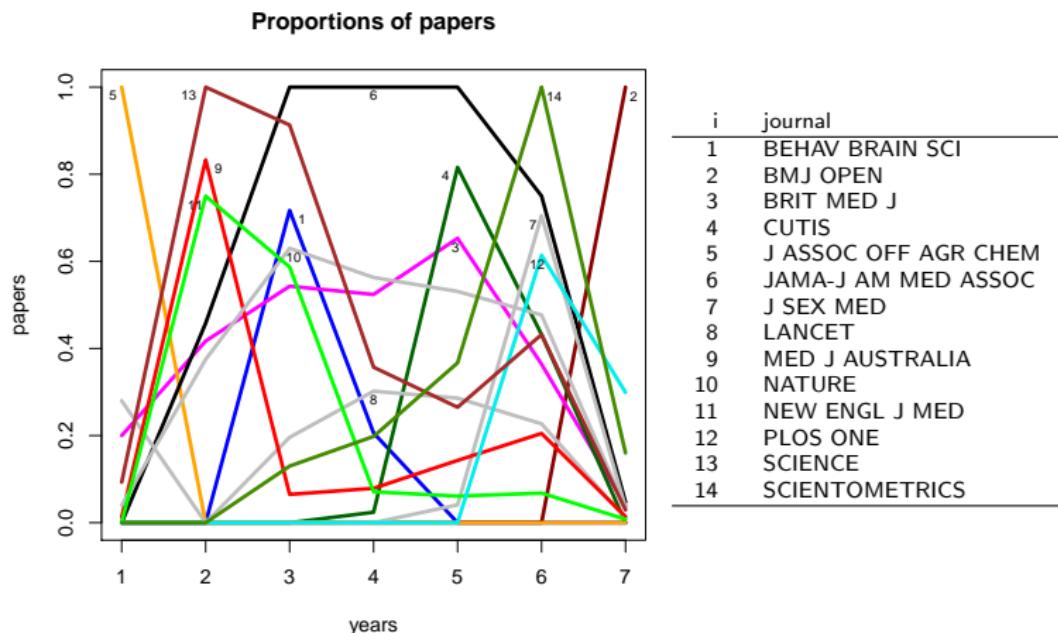


Main journals publishing on peer review

Bibliographic networks

V. Batagelj,
D. Maltseva

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





The most important journals – Python code

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

```
>>> jrn = ndir+"/WJd.net"
>>> WJc = N.twoMode2netJSON(clu,jrn,'WJcum.json',instant=False)
>>> WJi = N.twoMode2netJSON(clu,jrn,'WJins.json',instant=True)
>>> J = list(WJi.nodesMode(2))
>>> Jt = [ (j, WJi._nodes[j][3]['lab'], TQ.cutGT(WJi.TQnetInSum(j),0)) for j in J ]
>>> p = [0,1971,1981,1991,2001,2006,2011,2016,3000]
>>> Jr = [ (j,l,TQ.changeTime(a,p)) for (j,l,a) in Jt ]
>>> I = { Jr[j][1] : j for j in range(len(Jt)) }
>>> JL = [ "BEHAV BRAIN SCI", "BMJ OPEN", "BRIT MED J", "CUTIS",
    "J ASSOC OFF AGR CHEM", "JAMA-J AM MED ASSOC", "J SEX MED",
    "LANCET", "MED J AUSTRALIA", "NATURE", "NEW ENGL J MED",
    "PLOS ONE", "SCIENCE", "SCIENTOMETRICS" ]
>>> IJ = [ I[j] for j in JL ]; Ir = [ Jr[i] for i in IJ]
```

In the library `TQ` we included a new function `changeTime` that recodes a temporal quantity `a` into new time intervals determined by a sequence `p`.



Multiplication of temporal networks

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Let \mathbf{A} on $A \times P$ and \mathbf{B} on $P \times B$ be (matrices of) co-occurrence networks. Then $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$ is a temporal network on $A \times B$. What is its meaning? Consider the value of its item in an instant t

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

For $c_{ij}(t)$ to be defined (different from \emptyset) there should be at least one $p \in P$ such that $a_{pi}(t)$ and $b_{pj}(t)$ are both defined, i.e.
 $t \in T_{a_{pi}} \cap T_{b_{pj}}$. Then there exists g_{pi} such that $(s_{g_{pi}}, f_{g_{pi}}, v_{g_{pi}}) \in a_{pi}$,
 $t \in [s_{g_{pi}}, f_{g_{pi}}]$, and $a_{pi}(t) = v_{g_{pi}}$. Similarly $b_{pj}(t) = v_{h_{pj}}$. Therefore

$$c_{ij}(t) = \sum_{p: t \in T_{a_{pi}} \cap T_{b_{pj}}} v_{g_{pi}} \cdot v_{h_{pj}}$$



... Multiplication of temporal networks

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

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

The product of cumulative networks is cumulative itself. For binary cumulative two-mode networks **A** and **B** the value $c_{ij}(t)$ of the product $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$ is equal to the number of different members of P with which both i and j had contact in instants up to including the instant t .



Temporal co-occurrence networks

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

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

$$\mathbf{C}_i = \mathbf{A}_i^T \cdot \mathbf{A}_i \quad \text{and} \quad \mathbf{C}_c = \mathbf{A}_c^T \cdot \mathbf{A}_c$$

The triple (s, f, v) in a temporal quantity c_{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 (loop) weights c_{pp} and cc_{pp} contain the temporal quantities counting the number of events in the time intervals in which the participant p took part.

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



Example: Temporal coauthorship network

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

The instantaneous coauthorship network **Coi** is obtained as

$$\mathbf{Coi} = \mathbf{WAi}^T \cdot \mathbf{WAi}$$

Bibliographic networks are usually sparse. Often also the product of sparse networks is sparse itself. Considering in computation only non zero elements it can be computed fast. In our example, the network **WA** has 22104 works, 62106 authors and 80021 arcs. The derived network **Coi** has 633977 edges and was computed on a laptop in 12.7 seconds.



Example: Temporal coauthorship network

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

For the peer review data we get the largest values

$$co[BORNMANN_L, DANIEL_H] = 42,$$

$$co[MOHER_D, ALTMAN_D] = 24,$$

$$co[REYES_H, ANDRESEN_M] = 17.$$

The corresponding temporal quantities

$$bd = tq(BORNMANN_L, DANIEL_H) \text{ and}$$

$$ra = tq(REYES_H, ANDRESEN_M) \text{ are}$$

$$\begin{aligned} bd &= [(2005, 2006, 4), (2006, 2007, 3), (2007, 2008, 4), (2008, 2009, 7), \\ &\quad (2009, 2010, 4), (2010, 2011, 11), (2011, 2013, 4), (2015, 2016, 1)] \end{aligned}$$

$$\begin{aligned} ra &= [(1997, 1998, 3), (1998, 1999, 1), (2000, 2002, 1), (2004, 2005, 1), \\ &\quad (2005, 2006, 2), (2006, 2008, 1), (2009, 2010, 2), (2011, 2012, 1), \\ &\quad (2013, 2016, 1)] \end{aligned}$$

Both temporal quantities are presented on the next slide. The Python code is given in the following slide.



Example: Temporal coauthorship network

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

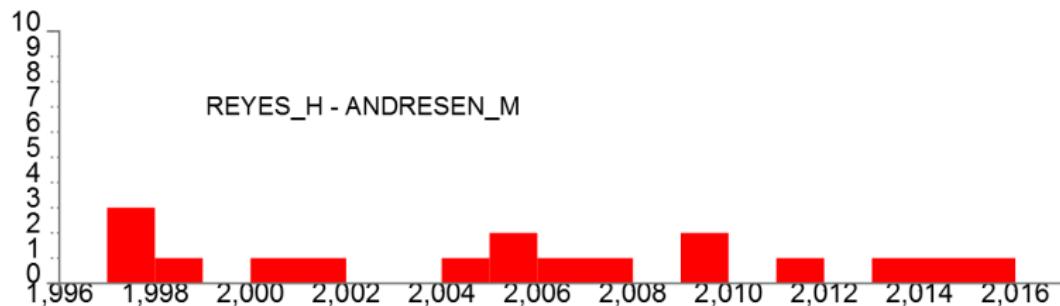
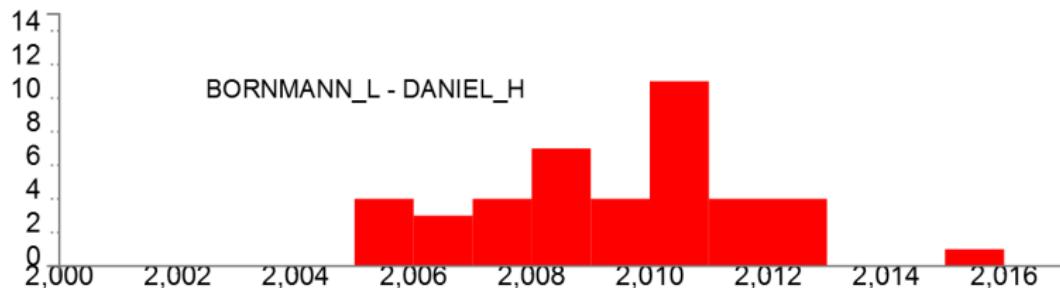
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Example: Temporal coauthorship network

Python code

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

```
>>> Co = WAi.TQtwoZoneCols()
>>> Co.saveNetsJSON('CoIns.json', indent=2)
>>> Co.delLoops()
>>> C = Co.TQtopLinks(thresh=15)
>>> tit = C[0][2]+ '-' +C[0][3]; bd = C[0][5]
>>> TQmax = 15; Tmin = 2000; Tmax = 2017; w = 600; h = 150
>>> N.TQshow(bd, cdir, TQmax, Tmin, Tmax, w, h, tit, fill='red')
>>> tit = C[2][2]+ '-' +C[2][3]; ra = C[2][5]
>>> TQmax = 10; Tmin = 1996; Tmax = 2017; w = 600; h = 150
>>> N.TQshow(ra, cdir, TQmax, Tmin, Tmax, w, h, tit, fill='red')
>>> TQ.total(bd), TQ.total(ra)
(42, 17)
```



Example: Temporal citations between journals

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

The derived network describing citations between journals is obtained as

$$\mathbf{JCJ} = \mathbf{WJi}^T \cdot \mathbf{Cil} \cdot \mathbf{WJc}$$

Note that the third network in the product is cumulative.

The weight of the element jc_{ij} is equal to the number of citations per year from works published in journal i to works published in journal j .

In a special case when $i = j$ we get a temporal quantity describing selfcitations of journal i . In the peer review data the largest number of selfcitations are 320 in JAMA and 148 in Scientometrics.



... Example: Temporal citations between journals

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

The corresponding temporal quantities $jm = jcj[JAMA, JAMA]$ and $sm = jcj[SCIENTOMETRICS, SCIENTOMETRICS]$ are:

$$jm = [(1973, 1976, 1), (1988, 1989, 1), (1989, 1990, 2), (1990, 1991, 16), (1991, 1992, 1), (1992, 1993, 11), (1993, 1994, 4), (1994, 1995, 44), (1995, 1996, 9), (1996, 1997, 2), (1997, 1998, 3), (1998, 1999, 68), (1999, 2000, 14), (2000, 2001, 10), (2001, 2002, 7), (2002, 2003, 60), (2003, 2004, 11), (2004, 2005, 4), (2005, 2006, 1), (2006, 2007, 16), (2007, 2008, 8), (2008, 2009, 2), (2009, 2010, 4), (2012, 2013, 4), (2013, 2014, 3), (2014, 2015, 7), (2015, 2016, 5)]$$

$$sm = [(1991, 1992, 1), (1995, 1996, 2), (1998, 1999, 2), (2001, 2002, 1), (2003, 2004, 1), (2005, 2006, 6), (2006, 2007, 10), (2007, 2009, 4), (2009, 2010, 7), (2010, 2011, 16), (2011, 2012, 14), (2012, 2013, 9), (2013, 2014, 34), (2014, 2015, 19), (2015, 2016, 17), (2016, 2017, 1)]$$

and are presented on the next slide.



Example: Journals Selfcitations

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

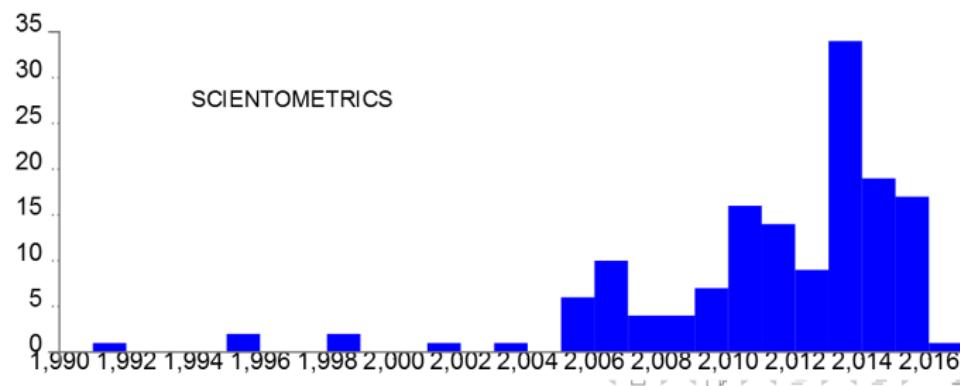
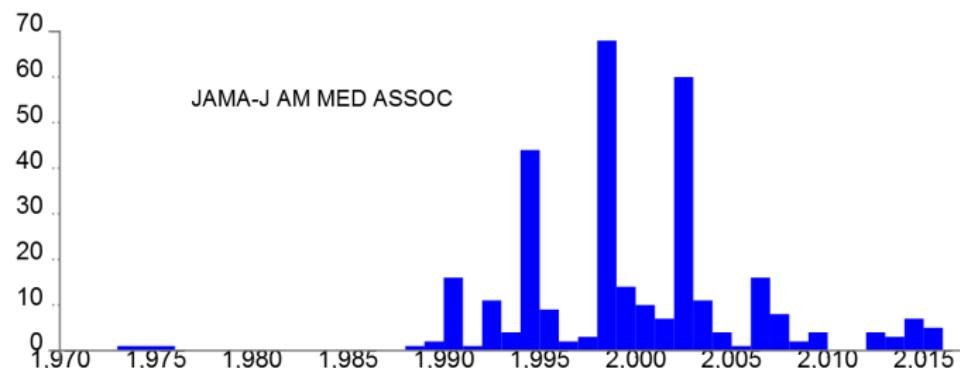
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Example: Temporal citations between journals

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

The largest number of citations are from journals BMJ Open (142) and Scientometrics (108) to the unknown journal ******, followed by $bj = jcj[\text{BRIT MED J, JAMA-J AM MED ASSOC}]$ and $pj = jcj[\text{PLOS ONE, JAMA-J AM MED ASSOC}]$ with totals 96 and 91.

$$bj = [(1994, 1996, 8), (1996, 1997, 4), (1997, 1998, 6), (1998, 1999, 2), (1999, 2000, 24), (2000, 2001, 1), (2001, 2002, 3), (2002, 2003, 6), (2003, 2004, 11), (2004, 2005, 6), (2005, 2006, 1), (2008, 2009, 2), (2009, 2010, 1), (2010, 2011, 4), (2011, 2012, 1), (2012, 2013, 8)]$$

$$pj = [(2007, 2008, 8), (2008, 2009, 13), (2009, 2010, 7), (2010, 2011, 12), (2011, 2012, 14), (2012, 2013, 11), (2013, 2014, 4), (2014, 2015, 11), (2015, 2016, 6), (2016, 2017, 5)]$$



Example: Temporal citations between journals

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

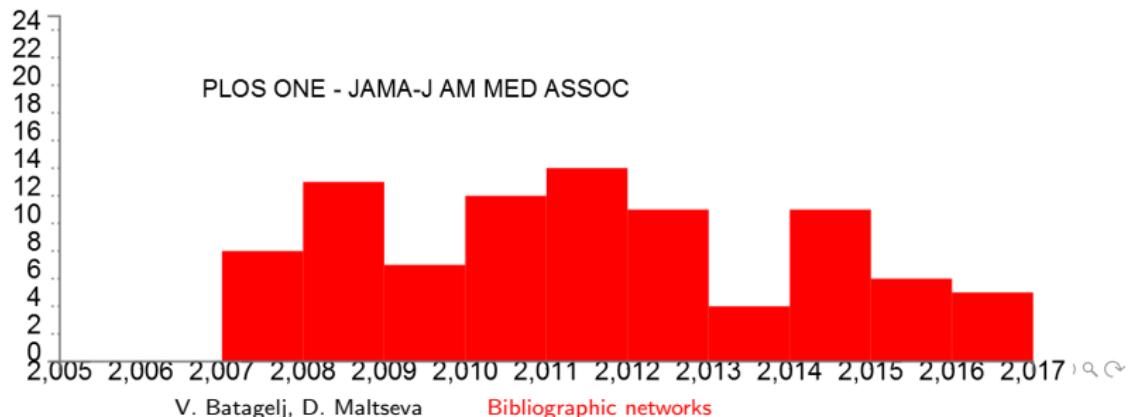
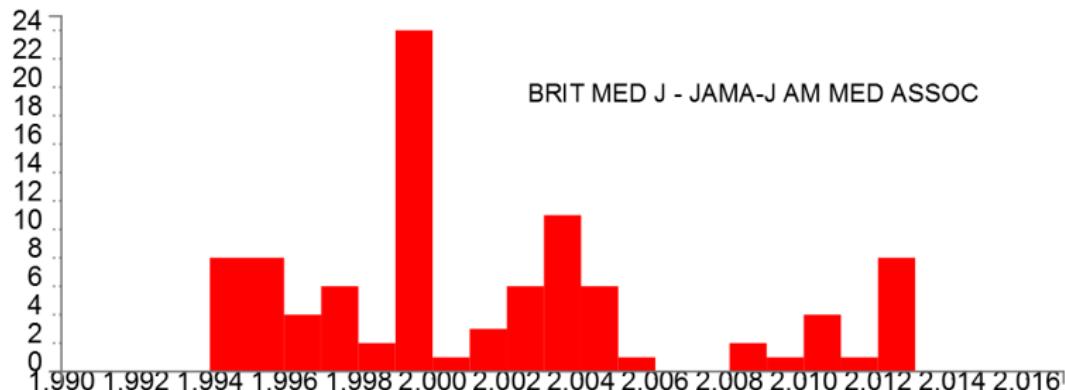
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References





Example: Temporal citations between journals

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

In the peer review data the journal JAMA is the most prominent. To get the temporal quantity describing citations of others to JAMA we compute $jci = iS(\mathbf{JCJ}, \text{JAMA-J AM MED ASSOC})$:

$$jci = [(1979, 1980, 2), (1982, 1983, 1), (1986, 1987, 5), (1987, 1988, 2), (1988, 1989, 1), (1990, 1991, 10), (1991, 1992, 22), (1992, 1993, 24), (1993, 1994, 36), (1994, 1995, 43), (1995, 1996, 108), (1996, 1998, 67), (1998, 1999, 161), (1999, 2000, 136), (2000, 2001, 106), (2001, 2002, 156), (2002, 2003, 145), (2003, 2004, 170), (2004, 2005, 98), (2005, 2006, 150), (2006, 2007, 229), (2007, 2008, 261), (2008, 2009, 185), (2009, 2010, 276), (2010, 2011, 239), (2011, 2012, 208), (2012, 2013, 220), (2013, 2014, 248), (2014, 2015, 191), (2015, 2016, 251), (2016, 2017, 43)]$$

See the figure on the next slide. The Python code is given in the following slide.

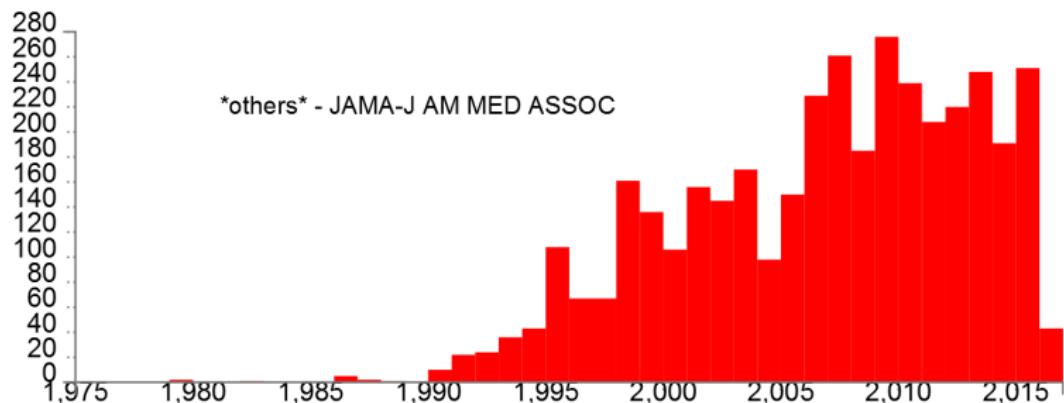


Example: Temporal citations between journals

Bibliographic networks

V. Batagelj,
D. Maltseva

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





Example: Temporal citations between journals

Python code

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

```
>>> JCJ = N.TQmultiply(N.TQmultiply(WJi.transpose(),
    Citei.one2twoMode()),WJc,True)
>>> L = JCJ.TQtopLoops(thresh=100)
>>> tit = L[0][1]; jm = L[0][3]
>>> TQmax = 70; Tmin = 1970; Tmax = 2017; w = 600; h = 200
>>> N.TQshow(jm,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='blue')
>>> tit = L[1][1]; sm = L[1][3]
>>> TQmax = 35; Tmin = 1990; Tmax = 2017; w = 600; h = 200
>>> N.TQshow(sm,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='blue')
>>> JCJ.delLoops()
>>> T = JCJ.TQtopLinks(thresh=70)
>>> tit = T[2][2]+'-'+T[2][3]; bj = T[2][5]
>>> TQmax = 25; Tmin = 1990; Tmax = 2017; w = 600; h = 200
>>> N.TQshow(bj,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='red')
>>> tit = T[3][2]+'-'+T[3][3]; pj = T[3][5]
>>> TQmax = 25; Tmin = 2005; Tmax = 2017; w = 600; h = 200
>>> N.TQshow(pj,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='red')
>>> jci = TQ.cutGE(JCJ.TQnetInSum(T[2][1]),1e-10)
>>> TQ.TqSummary(jci)
(1979, 2017, 1, 276)
>>> TQ.total(jci)
3861
>>> tit = '*others* - '+T[2][3]
>>> TQmax = 280; Tmin = 1975; Tmax = 2017; w = 600; h = 200
>>> N.TQshow(jci,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='red')
```



Citations between authors

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

Similarly we get the temporal network describing citations between authors

$$\mathbf{ACA} = \mathbf{WAI}^T \cdot \mathbf{CIL} \cdot \mathbf{WAC}$$

The weight of the element aca_{ab} is equal to the number of citations per year from works coauthored by author a to works coauthored by author b .



ESNA Pajek

Bibliographic
networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

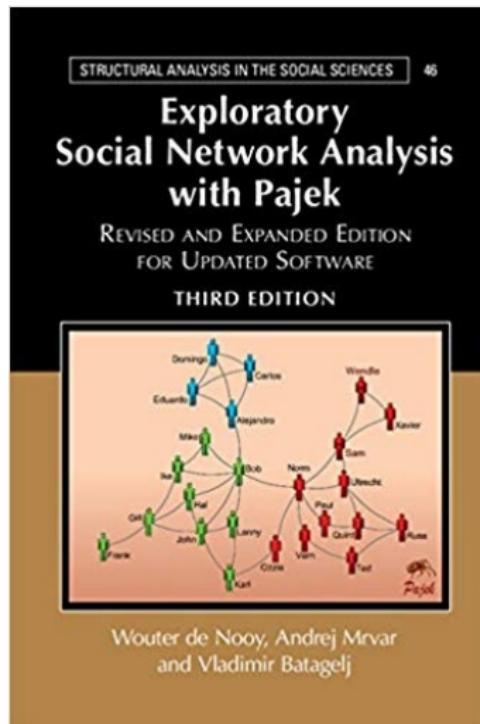
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



V. Batagelj, D. Maltseva

An introduction to social network analysis with Pajek is available in the book **ESNA 3** (de Nooy, Mrvar, Batagelj, CUP 2005, 2011, 2018).

ESNA in Japanese was published by Tokyo Denki University Press in 2010; and in Chinese by Beijing World Publishing in November 2012.

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

<http://mrvar.fdv.uni-lj.si/pajek/>



Understanding large networks

Bibliographic networks

V. Batagelj,
D. Maltseva

Networks

Statistics

Citation

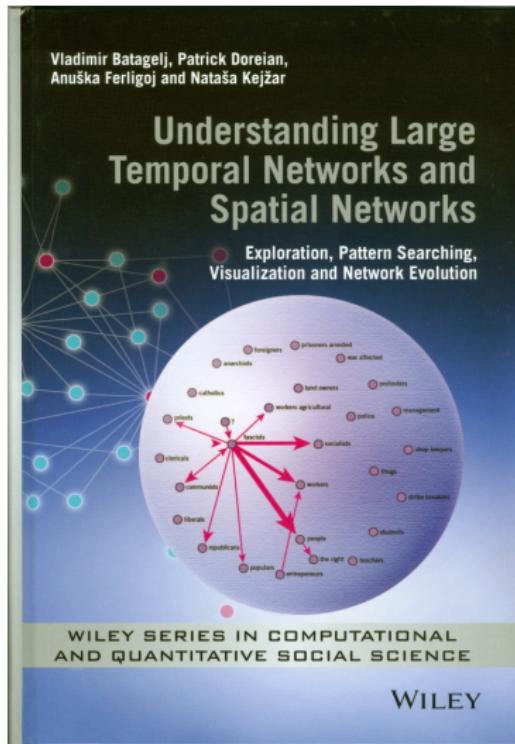
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References



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

- 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

- 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

- 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



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.