Analysis of bibliographic networks

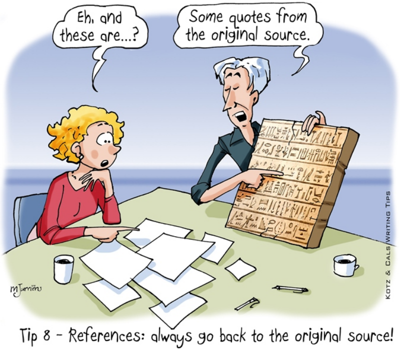
Vladimir Batagelj

IMFM Ljubljana and IAM UP Koper

PhD program on Statistics

University of Ljubljana, March 17, 2022

Outline

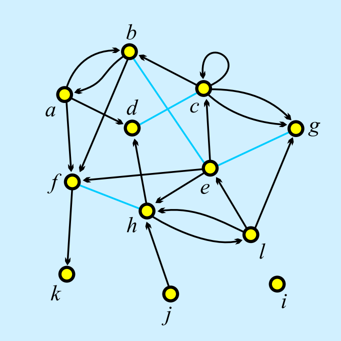


tip 8

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

Current version of slides (March 17, 2022 at 02 : 36): slides PDF
http://vladowiki.fmf.uni-lj.si/doku.php?id=notes:vis:nds

Networks



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

Network

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: $p : \mathcal{V} \to A$
- $\mathcal{W}$ *link value functions* / weights: $w : \mathcal{L} \to B$

Types of networks

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 mutally 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.

In a *linked* or *multimodal* network $\mathcal{N} = ((\mathcal{V}_1, \ldots, \mathcal{V}_j), (\mathcal{L}_1, \ldots, \mathcal{L}_k), \mathcal{P}, \mathcal{W})$ the set of nodes $\mathcal{V}$ is partitioned into subsets (*modes*) $\mathcal{V}_i$, $\mathcal{L}_s \subseteq \mathcal{V}_p \times \mathcal{V}_q$, and properties and weights are usually partial functions.

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.

Networks from bibliographies

For data from the Web of Science (Knowledge) we can obtain the corresponding networks using the program `WoS2Pajek`:

- citation network **Ci**: works $\times$ works;
- authorship network **WA**: works $\times$ authors, for works without complete description only the first author is known;
- keywords network **WK**: works $\times$ keywords, only for works with complete description;
- journals network **WJ**: works $\times$ journals;
- partition *year* of works by the publication year;
- partition *CD* of works – complete description (1) / ISI name only (0);

Similar programs exist also for other bibliographic sources/formats: Scopus, BibTEX, Zentralblatt Math, Google Scholar, DBLP, IMDB, etc.

Record from Web of Science

```
PT J
AU Dipple, H
   Evans, B
TI The Leicestershire Huntington's disease support group: a social network
   analysis
SO HEALTH & SOCIAL CARE IN THE COMMUNITY
LA English
DT Article
C1 Rehabil Serv, Troon Way Business Ctr, Leicester LE4 9HA, Leics, England.
RP Dipple, H, Rehabil Serv, Troon Way Business Ctr, Sandringham
   Suite,Humberstone Lane, Leicester LE4 9HA, Leics, England.
CR BORGATTI SP, 1992, UCINET 4 VERSION 1 0
   FOLSTEIN S, 1989, HUNTINGTONS DIS DISO
   SCOTT J, 1991, SOCIAL NETWORK ANAL
NR 3
TC 3
PU BLACKWELL SCIENCE LTD
PI OXFORD
PA P O BOX 88, OSNEY MEAD, OXFORD OX2 ONE, OXON, ENGLAND
SN 0966-0410
J9 HEALTH SOC CARE COMMUNITY
JI Health Soc. Care Community
PD JUL
PY 1998
VL 6
IS 4
BP 286
EP 289
PG 4
SC Public, Environmental & Occupational Health; Social Work
GA 105UP
UT ISI:000075092200008
ER
```

WoS2Pajek

Cleaning networks

The saved records from a bibliographic database 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 database file and rerun the creation of the network files and analyses.

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

Unit identification problem

Entity Resolution

The main two problems with units (works, authors, journals, keywords, etc.) are *equivalence* (different words representing the same unit) and *ambiguity* (same word representing different terms).

Synonymy:

Otfried Cheong (formerly Otfried Schwarzkopf): German computational geometer working in South Korea at KAIST.

Michel Marie Deza (formerly Mikhail Efimovich Tylkin): a Soviet and French mathematician, specializing in combinatorics, discrete geometry, and graph theory.

Borštnik, N. S. Mankoč; Mankoč Borštnik, N.; Mankoč-Borštnik, Norma; Mankoč Borštnik, Norma Susana; Mankoc-Borstnik, N.S.; and Mankoč Borštnik, N.S. belong to the same author.

NUCLEIC ACIDS RES, NUCL ACIDS RES, NUCLEIC ACIDS RES S, NUCLEIC ACIDS RES S2, NUCL ACID RES, NUCL ACIDS RES S2, NUCL ACIDS S SER, NUCL ACIDS RES S, NUCL AC RES, NUCLEIC ACIDS RES S1, Nucleic Acids Res, NUCL ACIDS RES S1.

Unit identification problem

Homonymy (ambiguity):

Lorenzo Bartolini from Letters to Juliet.

Smith, John W. - publications of the author(s) with this name spanned from 1868 to 2007.

Chinese, $\approx$ 100 surnames – "three Zhang, four Li" (there are some hundreds of different mathematicians with the name Zhang, Li in the MathSciNet Database).
Kim, Lee, and Park account for nearly one-half of all Koreans.
Nguyen is a surname held by 40% of the Vietnamese population, etc.

Solutions for **names**: ResearcherID, ORCID, ISNI, AMS;
for **works**: International Standard Serial Number ISSN, Digital Object Identifier DOI, International Standard Book Number ISBN;
for **words**: dictionaries, stemming, lemmatization; for **countries**: ISO 3166; for **languages**: ISO 639

Network boundary problem

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.

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 $(0 < \text{indeg}(v) < k) \wedge (\text{outdeg}(v) = 0)$.

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

Example networks

SN5: WoS, January 2008, ("social network*" AND SO=(Social networks)) + most frequently cited works + around 100 SNA researchers. $|W| = 193376$, $|C| = 7950$, $|A| = 75930$, $|J| = 14651$, $|K| = 29267$.

ZBMath: Data from the Zentralblatt MATH data base for years 1990–2010. $|W| = 1339201$, $|A| = 557104$, $|J| = 3158$, $|K| = 143513$, $|M| = 12390$. paper.

PEERE: WoS, March 2016, using the queries `"peer review*"` and `refereeing` + manually prepared descriptions of frequently cited only works. `WoS2Pajek` produced networks with sets of the following sizes: works $|W| = 721547$, authors $|A| = 295849$, journals $|J| = 39988$, keywords $|K| = 36279$ and $m = 869821$ arcs. 22981 records were collected. paper

SNA17: WoS, January 2018, $|W| = 1297133/70792$, $|A| = 395971/93011$, $|K| = 32409/32409$, $|J| = 69146/8943$. paper

Statistics

Global/overall properties.

Extreme (minimal/maximal values) units.

Distributions of values.

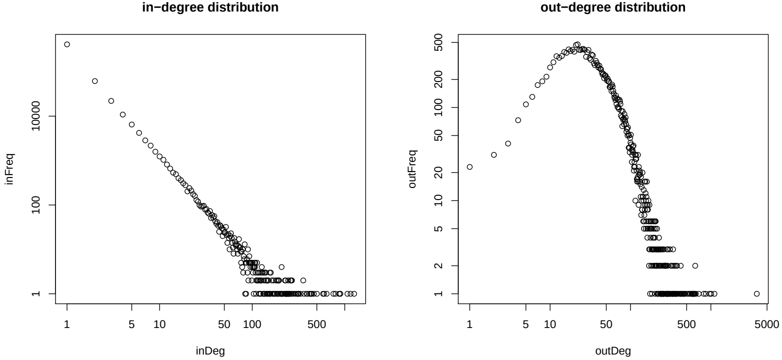
degree of node v, $\deg(v)$ = number of links with v as an endnode;
indegree of node v, $\text{indeg}(v)$ = number of links with v as a terminal node (endnode is both initial and terminal);
outdegree of node v, $\text{outdeg}(v)$ = number of links with v as an initial node.
source node $v \Leftrightarrow \text{indeg}(v) = 0$
sink node $v \Leftrightarrow \text{outdeg}(v) = 0$

PEERE – most cited works: indegree in **Ci**

n	freq	first author	title
1	173	Cohen, J	Statistical Power Analysis for the Behavioral Sciences. Routledge,1988
2	164	Peters, DP	Peer-review practices of psychological journals - the fate of . . . Behav Brain Sci, 1982
3	151	Egger, M	Bias in meta-analysis detected by a simple, graphical test. Brit Med J, 1997
4	150	Stroup, DF	Meta-analysis of observational studies in epidemiology - A proposal for reporting. JAMA, 2000
5	135	Dersimonian, R	Metaanalysis in clinical-trials. Control Clin Trials, 1986
6	130	Zuckerma.H	Patterns of evaluation in science - institutionalisation, structure . . . Minerva, 1971
7	130	Higgins, JPT	Cochrane Handbook for Systematic Reviews of Interventions. Cochrane, 2011
8	126	Moher, D	Preferred Reporting Items for Systematic Reviews and Meta-Analyses. Plos Med, 2009
9	125	Higgins, JPT	Measuring inconsistency in meta-analyses. Brit Med J, 2003
10	121	Cicchetti, DV	The reliability of peer-review for manuscript and grant submissions - . . . Behav Brain Sci, 1991
11	119	Hirsch, JE	An index to quantify an individual's scientific research output. P Natl Acad Sci Usa, 2005
12	114	Mahoney, M	Publication prejudices: An experimental study of confirmatory bias . . . Cognitive T& R, 1977
13	114	van Rooyen, S	Effect of open peer review on quality of reviews and on reviewers' . . . Brit Med J, 1999
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 sign . . . JAMA, 1998
17	108	Horrobin, DF	The philosophical basis of peer-review and the suppression of innovation. JAMA, 1990
18	107	Moher, D	Preferred Reporting Items for Systematic Reviews and Meta-Analyses. Ann Intern Med, 2009
19	107	Jadad, AR	Assessing the quality of reports of randomized clinical trials. Control Clin Trials, 1996
20	105	Mcnutt, RA	The effects of blinding on the quality of peer-review - a randomized trial. JAMA, 1990
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 trials. Lancet, 1999
23	98	Justice, AC	Does masking author identity improve peer review quality? JAMA, 1998
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 randomized trial. JAMA, 1998
26	92	Black, N	What makes a good reviewer and a good review for a general medical journal? JAMA, 1998
27	91	Scherer, RW	Full publication of results initially presented in abstracts - a metaanalysis. JAMA, 1994
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 Med, 2006
30	87	Goodman, SN	Manuscript quality before and after peer-review and editing at . . . . Ann Intern Med, 1994
31	87	Chubin, D	Peerless Science: Peer Review and U.S. Science Policy. SUNY Press, 1990

Distributions

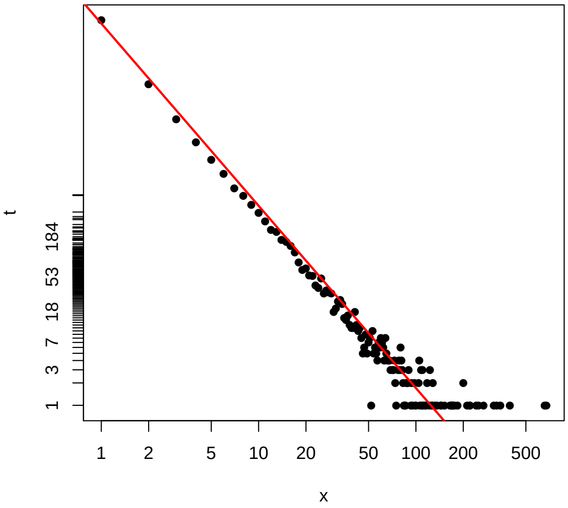
indegree and outdegree in citation network



The indegree 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.

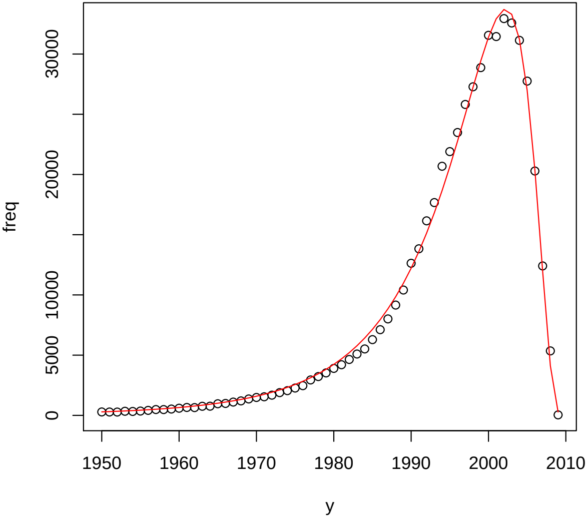
Distributions

SN5 citation network input degrees – scale-free fit



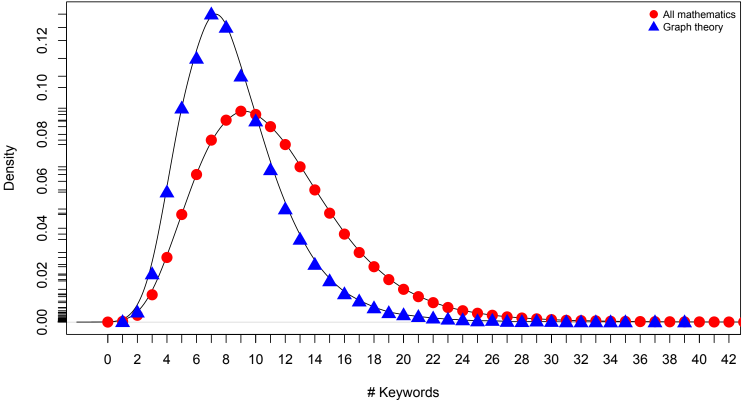
Distributions

year partition lognormal distribution fit



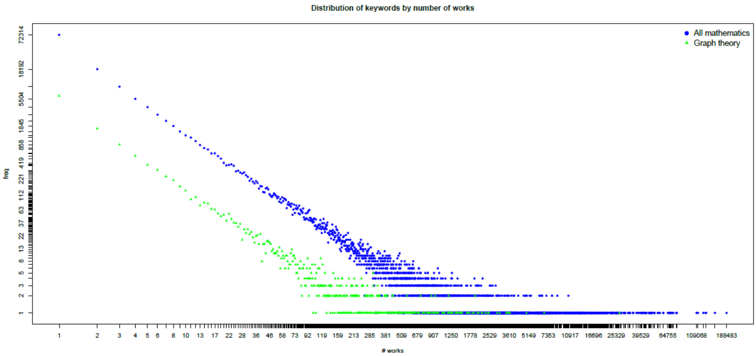
Distributions

number of keywords in **ZBMath**



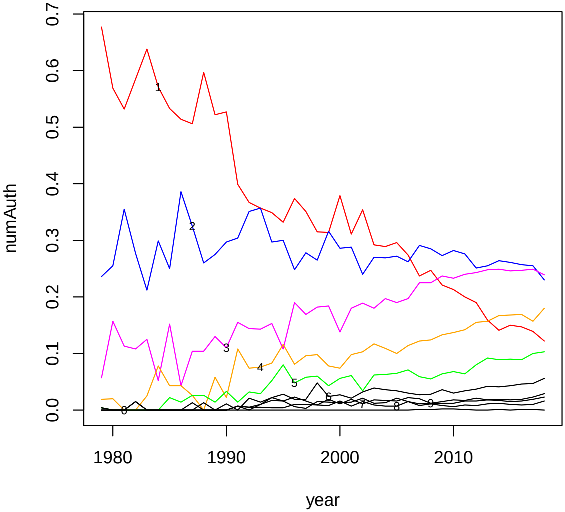
Distributions

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



Temporal distribution

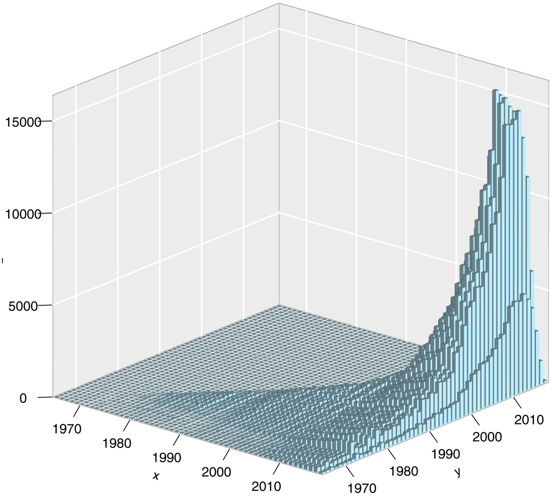
Number of authors SNA 2018



SocNet

Citations between years

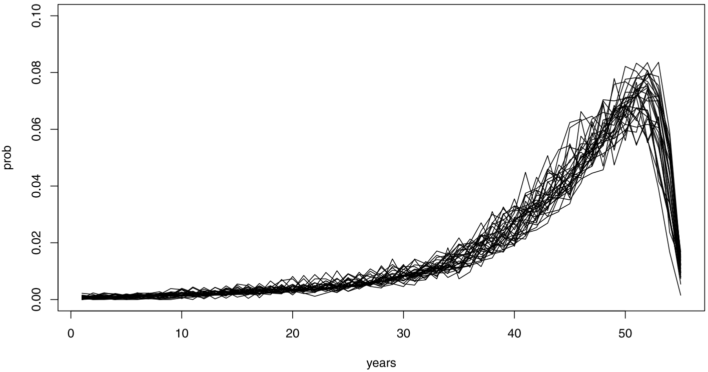
SNA 2018



SocNet

Citations per year

SNA 2018; normalized curves



SocNet

Citation networks

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

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

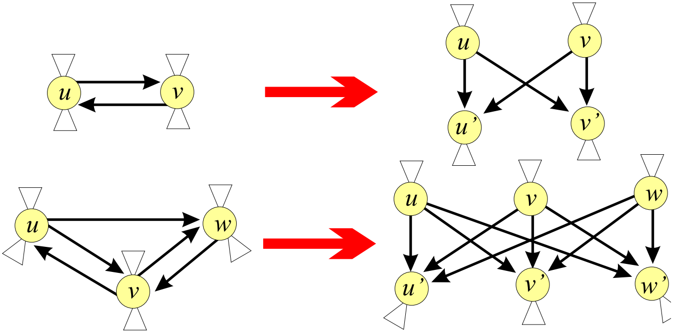
which determines a *citation network* $\mathcal{N} = (W, \mathbf{Ci})$.

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.

A better way is the preprint transformation.

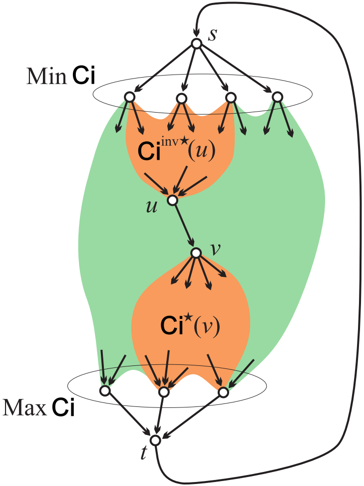
Citation network

Preprint transformation



Transforming a citation network into an acyclic network using the preprint transformation.

Standardized citation network



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

Search path count method

Hummon and Doreian

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).
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 *incoming flow* = *outgoing flow*:

$$\sum_{v:v\,\mathbf{Ci}\,u} n(v,u) = \sum_{v:u\,\mathbf{Ci}\,v} n(u,v)$$

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

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

and if C is a minimal arc-cut-set $\sum_{(u,v)\in C} w(u,v) = 1$.
The value $w(u, v)$ is equal to the probability that a random s-t path passes through the arc (u, v).

Cuts

The standard approach to find interesting groups inside a network is 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 $\mathcal{N} = (\mathcal{V}, \mathcal{L}, p)$, $p : \mathcal{V} \to \mathbb{R}$, 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 $\mathcal{V}'$.

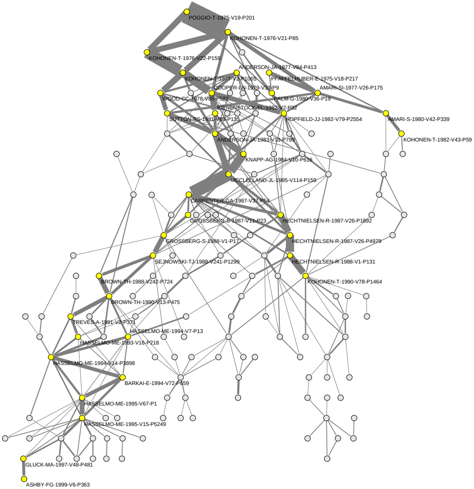
The *link cut* of a network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, w)$, $w : \mathcal{L} \to \mathbb{R}$, at selected level t is a subnetwork $\mathcal{N}(t) = (\mathcal{V}(\mathcal{L}'), \mathcal{L}', w)$, determined by the set

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

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

Citation weights

Link cut for SPC in SOM



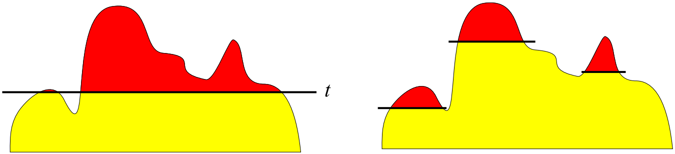
Main subnetwork (arc cut at level 0.007) of the SOM (selforganizing maps) citation network (4470 nodes, 12731 arcs).

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

See paper.

Islands

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

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

A set of nodes $C \subseteq \mathcal{V}$ is a *regular node island* in a network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, p)$, $p : \mathcal{V} \to \mathbb{R}$ 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 $\mathcal{N} = (\mathcal{V}, \mathcal{L}, w)$, $w : \mathcal{L} \to \mathbb{R}$ 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

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

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

US patents

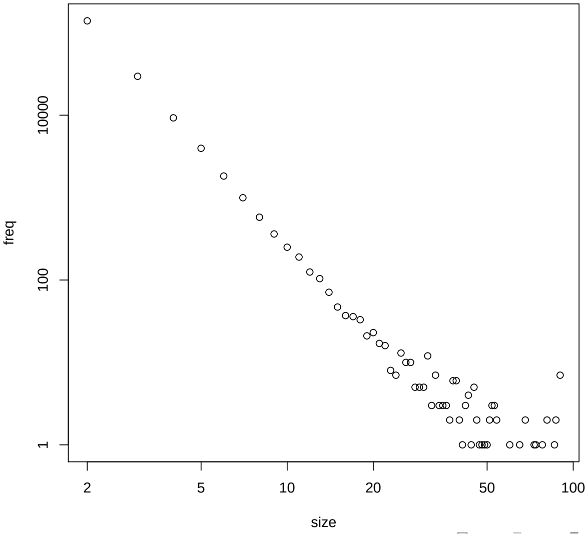
US patents network (Nber, US Patents) has 3774768 nodes and16522438 arcs (1 loop). Without the loop it is acyclic. The weight of an arc is the proportion of paths through the arc from some initial node to some terminal node. We determined al (2,90)-link islands. The corresponding subnetwork has 470137 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.

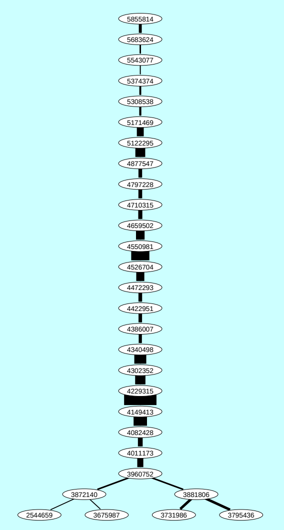
The *CPM path* is determined using the Critical Path Method from Operations Research (the sum of SPC weights along a path is maximal).

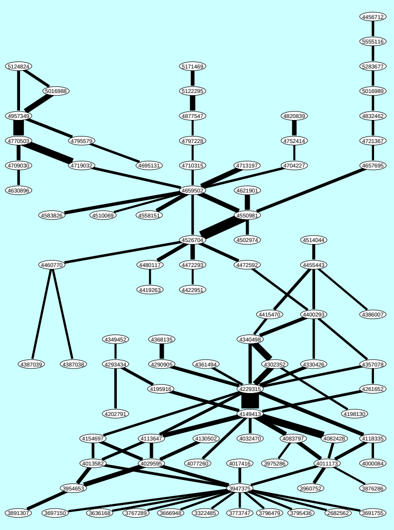
Distribution of island size



Main path and main island in US Patents

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





Main island – Liquid crystal display

Table 1: Patents on the liquid-crystal display

patent	date	author(s) and title
2544659	Mar 13, 1951	Dreyer. Dichroic light-polarizing sheet and the like and the formation and use thereof
2682562	Jun 29, 1954	Wender, et al. Reduction of aromatic carbinols
3322485	May 30, 1967	Williams. Electro-optical elements utilazing an organic nematic compound
3636168	Jan 18, 1972	Josephson. Preparation of polynuclear aromatic compounds
3666948	May 30, 1972	Mechlowitz, et al. Liquid crystal termal imaging system having an undisturbed image on a disturbed background
3675987	Jul 11, 1972	Rafuse. Liquid crystal compositions and devices
3691755	Sep 19, 1972	Girard. Clock with digital display
3697150	Oct 10, 1972	Wysochi. Electro-optic systems in which an electrophoretic-like or dipolar material is dispersed throughout a liquid crystal to reduce the turn-off time
3731986	May 8, 1973	Fergason. Display devices utilizing liquid crystal light modulation
3767289	Oct 23, 1973	Aviram, et al. Class of stable trans-stilbene compounds, some displaying nematic mesophases at or near room temperature and others in a range up to 100°C
3773747	Nov 20, 1973	Steinstrasser. Substituted azoxy benzene compounds
3795436	Mar 5, 1974	Boller, et al. Nematogenic material which exhibit the Kerr effect at isotropic temperatures
3796479	Mar 12, 1974	Helfrich, et al. Electro-optical light-modulation cell utilizing a nematogenic material which exhibits the Kerr effect at isotropic temperatures
3872140	Mar 18, 1975	Klanderman, et al. Liquid crystalline compositions and method
3876286	Apr 8, 1975	Deutscher, et al. Use of nematic liquid crystalline substances
3881806	May 6, 1975	Suzuki. Electro-optical display device
3891307	Jun 24, 1975	Tsukamoto, et al. Phase control of the voltages applied to opposite electrodes for a cholesteric to nematic phase transition display
3947375	Mar 30, 1976	Gray, et al. Liquid crystal materials and devices
3954653	May 4, 1976	Yamazaki. Liquid crystal composition having high dielectric anisotropy and display device incorporating same
3960752	Jun 1, 1976	Klanderman, et al. Liquid crystal compositions
3975286	Aug 17, 1976	Oh. Low voltage actuated field effect liquid crystals compositions and method of synthesis
4000084	Dec 28, 1976	Hsieh, et al. Liquid crystal mixtures for electro-optical display devices
4011173	Mar 8, 1977	Steinstrasser. Modified nematic mixtures with positive dielectric anisotropy
4013582	Mar 22, 1977	Gavrilovic. Liquid crystal compounds and electro-optic devices incorporating them
4017416	Apr 12, 1977	Inukai, et al. P-cyanophenyl 4-alkyl-4'-biphenylcarboxylate, method for preparing same and liquid crystal compositions using same
4029595	Jun 14, 1977	Ross, et al. Novel liquid crystal compounds and electro-optic devices incorporating them
4032470	Jun 28, 1977	Bloom, et al. Electro-optic device
4077260	Mar 7, 1978	Gray, et al. Optically active cyano-biphenyl compounds and liquid crystal materials containing them
4082428	Apr 4, 1978	Hsu. Liquid crystal composition and method

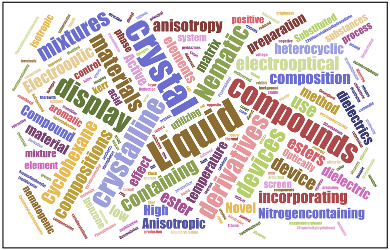
Table 2: Patents on the liquid-crystal display

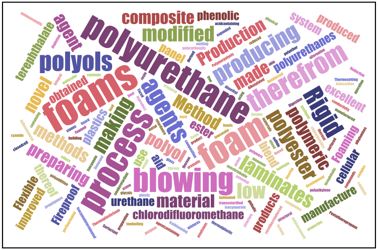
patent	date	author(s) and title
4083797	Apr 11, 1978	Oh. Nematic liquid crystal compositions
4113647	Sep 12, 1978	Coates, et al. Liquid crystalline materials
4118335	Oct 3, 1978	Krause, et al. Liquid crystalline materials of reduced viscosity
4130502	Dec 19, 1978	Eidenschink, et al. Liquid crystalline cyclohexane derivatives
4149413	Apr 17, 1979	Gray, et al. Optically active liquid crystal mixtures and liquid crystal devices containing them
4154697	May 15, 1979	Eidenschink, et al. Liquid crystalline hexahydroterphenyl derivatives
4195916	Apr 1, 1980	Coates, et al. Liquid crystal compounds
4198130	Apr 15, 1980	Boller, et al. Liquid crystal mixtures
4202791	May 13, 1980	Sato, et al. Nematic liquid crystalline materials
4229315	Oct 21, 1980	Krause, et al. Liquid crystalline cyclohexane derivatives
4261652	Apr 14, 1981	Gray, et al. Liquid crystal compounds and materials and devices containing them
4290905	Sep 22, 1981	Kanbe. Ester compound
4293434	Oct 6, 1981	Deutscher, et al. Liquid crystal compounds
4302352	Nov 24, 1981	Eidenschink, et al. Fluorophenylcyclohexanes, the preparation thereof and their use as components of liquid crystal dielectrics
4330426	May 18, 1982	Eidenschink, et al. Cyclohexylbiphenyls, their preparation and use in dielectrics and electrooptical display elements
4340498	Jul 20, 1982	Sugimori. Halogenated ester derivatives
4349452	Sep 14, 1982	Osman, et al. Cyclohexylcyclohexanoates
4357078	Nov 2, 1982	Carr, et al. Liquid crystal compounds containing an alicyclic ring and exhibiting a low dielectric anisotropy and liquid crystal materials and devices incorporating such compounds
4361494	Nov 30, 1982	Osman, et al. Anisotropic cyclohexyl cyclohexylmethyl ethers
4368135	Jan 11, 1983	Osman. Anisotropic compounds with negative or positive DC-anisotropy and low optical anisotropy
4386007	May 31, 1983	Krause, et al. Liquid crystalline naphthalene derivatives
4387038	Jun 7, 1983	Fukui, et al. 4-(Trans-4'-alkylcyclohexyl) benzoic acid 4'''-cyano-4''-biphenylyl esters
4387039	Jun 7, 1983	Sugimori, et al. Trans-4-(trans-4'-alkylcyclohexyl)-cyclohexane carboxylic acid 4'''-cyanobiphenyl ester
4400293	Aug 23, 1983	Romer, et al. Liquid crystalline cyclohexylphenyl derivatives
4415470	Nov 15, 1983	Eidenschink, et al. Liquid crystalline fluorine-containing cyclohexylbiphenyls and dielectrics and electro-optical display elements based thereon
4419263	Dec 6, 1983	Praefcke, et al. Liquid crystalline cyclohexylcarbonitrile derivatives
4422951	Dec 27, 1983	Sugimori, et al. Liquid crystal benzene derivatives
4455443	Jun 19, 1984	Takatsu, et al. Nematic halogen Compound
4456712	Jun 26, 1984	Christie, et al. Bismaleimide triazine composition
4460770	Jul 17, 1984	Petrzilka, et al. Liquid crystal mixture
4472293	Sep 18, 1984	Sugimori, et al. High temperature liquid crystal substances of four rings and liquid crystal compositions containing the same
4472592	Sep 18, 1984	Takatsu, et al. Nematic liquid crystalline compounds
4480117	Oct 30, 1984	Takatsu, et al. Nematic liquid crystalline compounds
4502974	Mar 5, 1985	Sugimori, et al. High temperature liquid-crystalline ester compounds
4510069	Apr 9, 1985	Eidenschink, et al. Cyclohexane derivatives

Table 3: Patents on the liquid-crystal display

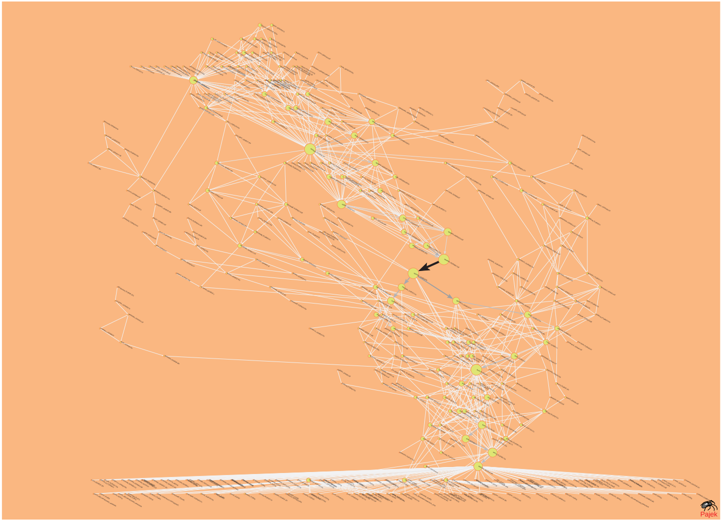
patent	date	author(s) and title
4514044	Apr 30, 1985	Gunjima, et al. 1-(Trans-4-alkylcyclohexyl)-2-(trans-4'-(p-sub stituted phenyl) cyclohexyl)ethane and liquid crystal mixture
4526704	Jul 2, 1985	Petrzilka, et al. Multiring liquid crystal esters
4550981	Nov 5, 1985	Petrzilka, et al. Liquid crystalline esters and mixtures
4558151	Dec 10, 1985	Takatsu, et al. Nematic liquid crystalline compounds
4583826	Apr 22, 1986	Petrzilka, et al. Phenylethanes
4621901	Nov 11, 1986	Petrzilka, et al. Novel liquid crystal mixtures
4630896	Dec 23, 1986	Petrzilka, et al. Benzonitriles
4657695	Apr 14, 1987	Saito, et al. Substituted pyridazines
4659502	Apr 21, 1987	Fearon, et al. Ethane derivatives
4695131	Sep 22, 1987	Balkwill, et al. Disubstituted ethanes and their use in liquid crystal materials and devices
4704227	Nov 3, 1987	Krause, et al. Liquid crystal compounds
4709030	Nov 24, 1987	Petrzilka, et al. Novel liquid crystal mixtures
4710315	Dec 1, 1987	Schad, et al. Anisotropic compounds and liquid crystal mixtures therewith
4713197	Dec 15, 1987	Eidenschink, et al. Nitrogen-containing heterocyclic compounds
4719032	Jan 12, 1988	Wachtler, et al. Cyclohexane derivatives
4721367	Jan 26, 1988	Yoshinaga, et al. Liquid crystal device
4752414	Jun 21, 1988	Eidenschink, et al. Nitrogen-containing heterocyclic compounds
4770503	Sep 13, 1988	Buchecker, et al. Liquid crystalline compounds
4795579	Jan 3, 1989	Vauchier, et al. 2,2'-difluoro-4-alkoxy-4'-hydroxydiphenyls and their derivatives, their production process and their use in liquid crystal display devices
4797228	Jan 10, 1989	Goto, et al. Cyclohexane derivative and liquid crystal composition containing same
4820839	Apr 11, 1989	Krause, et al. Nitrogen-containing heterocyclic esters
4832462	May 23, 1989	Clark, et al. Liquid crystal devices
4877547	Oct 31, 1989	Weber, et al. Liquid crystal display element
4957349	Sep 18, 1990	Clerc, et al. Active matrix screen for the color display of television pictures, control system and process for producing said screen
5016988	May 21, 1991	Iimura. Liquid crystal display device with a birefringent compensator
5016989	May 21, 1991	Okada. Liquid crystal element with improved contrast and brightness
5122295	Jun 16, 1992	Weber, et al. Matrix liquid crystal display
5124824	Jun 23, 1992	Kozaki, et al. Liquid crystal display device comprising a retardation compensation layer having a maximum principal refractive index in the thickness direction
5171469	Dec 15, 1992	Hittich, et al. Liquid-crystal matrix display
5283677	Feb 1, 1994	Sagawa, et al. Liquid crystal display with ground regions between terminal groups
5308538	May 3, 1994	Weber, et al. Supertwist liquid-crystal display
5374374	Dec 20, 1994	Weber, et al. Supertwist liquid-crystal display
5543077	Aug 6, 1996	Rieger, et al. Nematic liquid-crystal composition
5555116	Sep 10, 1996	Ishikawa, et al. Liquid crystal display having adjacent electrode terminals set equal in length
5683624	Nov 4, 1997	Sekiguchi, et al. Liquid crystal composition
5855814	Jan 5, 1999	Matsui, et al. Liquid crystal compositions and liquid crystal display elements

Word clouds for LCD island and foam island

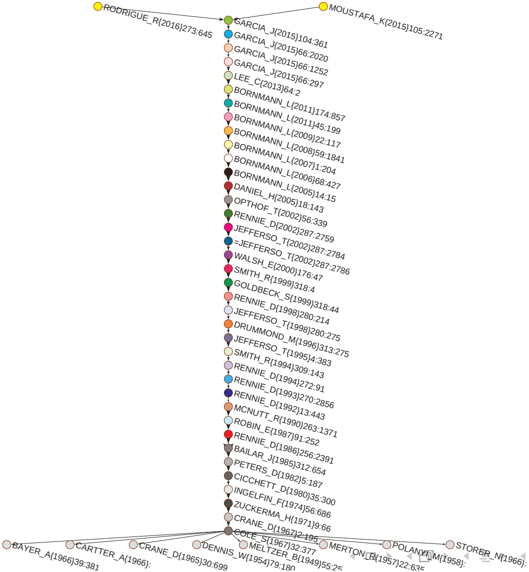




Main SPC island in SN5



PEERE – Main path



PEERE – List of publications on main path

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...	AmSociolRev
1958	Polanyi M	Personal Knowledge: Towards a Post-Critical Philosophy	UPChicago
1965	Crane D	Scientists at major and minor universities	AmSociolRev
1966	Bayer AE	Some correlates of citation measure of productivity in science	SociolEduc
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...	AmSociolRev
1971	Zuckerman H	Patterns of evaluation in science - institutionalisation, struct...	Minerva
1974	Ingelfinger FJ	Peer review in biomedical publication	AmJMed
1980	Cicchetti DV	Reliability of reviews for the american-psychologist	AmPsychol
1982	Peters DP	Peer-review practices of psychological journals - the fate...	BehavBrainSci
1985	Bailar JC	Journal peer-review - the need for a research agenda	NewEnglJMed
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	ControlClinTrials
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

PEERE – . . . List of works on main path

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	BritJPsychiat
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	Opthof T	The significance of the peer review process against . . . bias	CardiovascRes
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	Scientometrics
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	JAmSocInfSciTe
2009	Bornmann L	The luck of the referee draw: the effect of exchanging reviews	LearnPubl
2011	Bornmann L	Scientific Peer Review	AnnuRevInform
2011	Bornmann L	A multilevel modelling approach to investigating the predictive...	JRStatSocASta
2013	Lee CJ	Bias in peer review	JAmSocInfSciTe
2015	Garcia JA	The Principal-Agent Problem in Peer Review	JAssocInfSciTec
2015	Garcia JA	Adverse selection of reviewers	JAssocInfSciTec
2015	Garcia JA	Bias and effort in peer review	JAssocInfSciTec
2015	Garcia JA	The author-editor game	Scientometrics
2015	Moustafa K	Don't infer anything from unavailable data	Scientometrics
2016	Rodriguez-S. R	Evolutionary games between authors and their editors	ApplMathComp

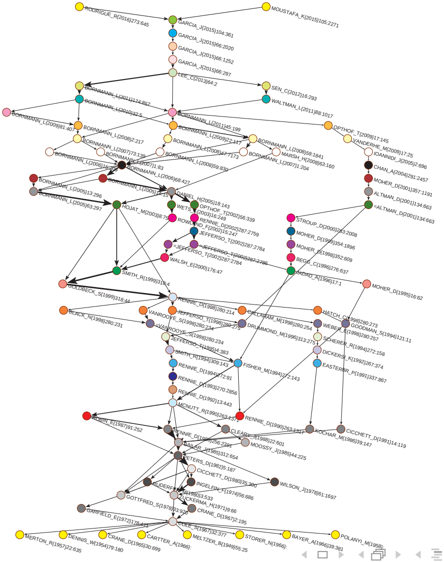
The main path publications

Phases

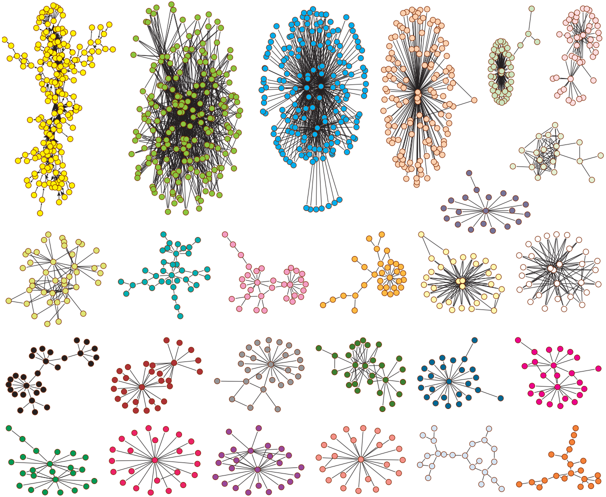
- before 1982: social science journals;
- from 1982 to 2002: biomedical journals;
- after 2002: specialized journals on science studies.

Journals, Main authors, Topics.

PEERE – Main paths for 100 largest weights

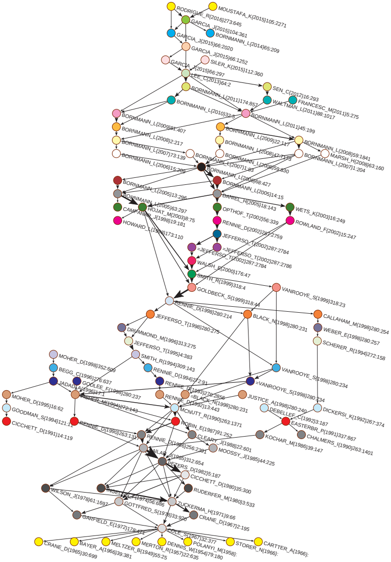


PEERE – SPC link islands [20 200]



PEERE – SPC – Link island1

$w_{max} = 0.297$



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

Islands

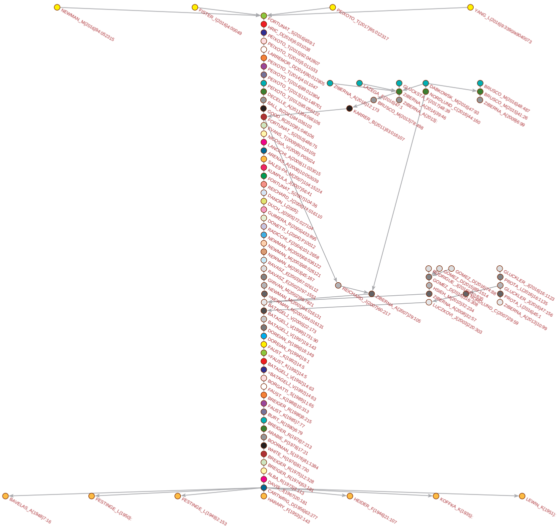
positioning

A recent extension of the main paths approach 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 in **NetClus** – valued networks in clustering



Two-mode networks

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* $w : \mathcal{L} \to \mathbb{R} \in \mathcal{W}$ 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), (parlamentarians, problems, positive vote), (persons, journals, reading), (papers, keywords, is described by), etc.

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

Multiplication of networks

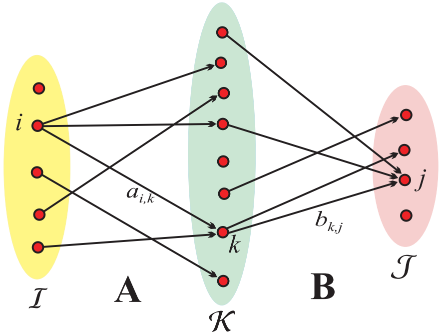
To a simple (no parallel arcs) two-mode *network* $\mathcal{N} = (\mathcal{I}, \mathcal{J}, \mathcal{A}, w)$; where $\mathcal{I}$ and $\mathcal{J}$ are sets of *nodes*, $\mathcal{A}$ is a set of *arcs* linking $\mathcal{I}$ and $\mathcal{J}$, and $w : \mathcal{A} \to \mathbb{R}$ (or some other semiring) 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



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

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

Multiplication of networks

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 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.

For more details see the paper.

Two-mode network analysis

by conversion to one-mode network – projections

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 $\mathbf{W}^{(1)} = \mathbf{W}\mathbf{W}^T$, $w^{(1)}_{uv} = \sum_{z\in\mathcal{V}} w_{uz} \cdot w^T_{zv}$. Evidently $w^{(1)}_{uv} = w^{(1)}_{vu}$. 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 $w_1(u, v) = w^{(1)}_{uv}$.

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.

Authorship networks

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

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

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

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

First co-authorship network

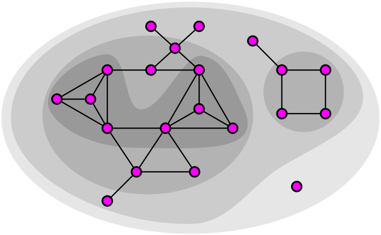
$\mathbf{Co} = \mathbf{AW} * \mathbf{WA}$

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

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

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

Cores and generalized cores



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

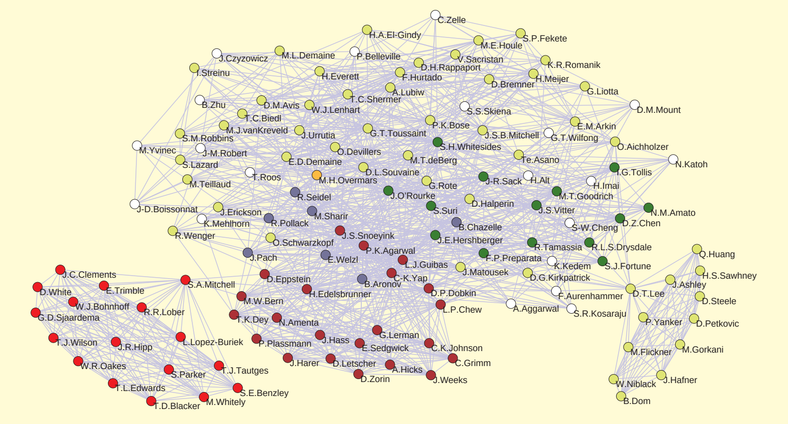
The *core number* of a node v is the highest order of a core that contains this node. The degree $\deg(v)$ can be: in-degree, out-degree, in-degree + out-degree, etc., determining different types of cores.

p_S-cores

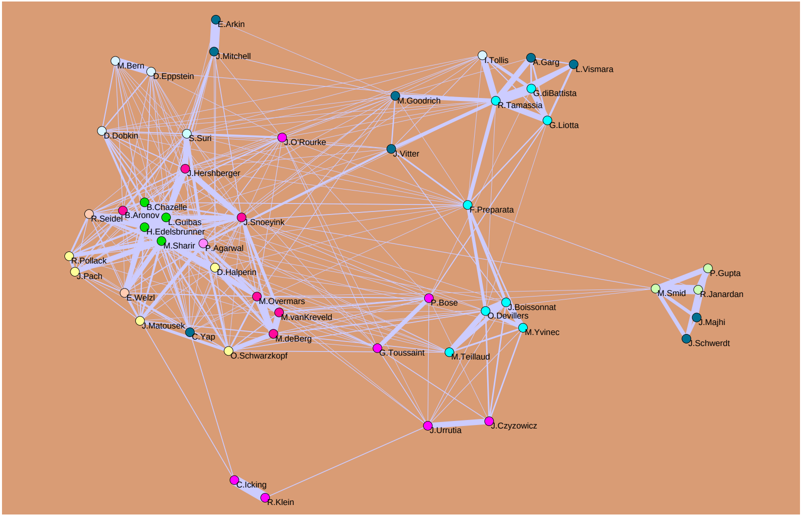
$$p_S(v, U) = \sum_{u \in N_U(v)} w(v, u), \text{ where } w : \mathcal{E} \to \mathbb{R}_0^+$$

The cores can be determined very efficiently: paper 1, paper 2.

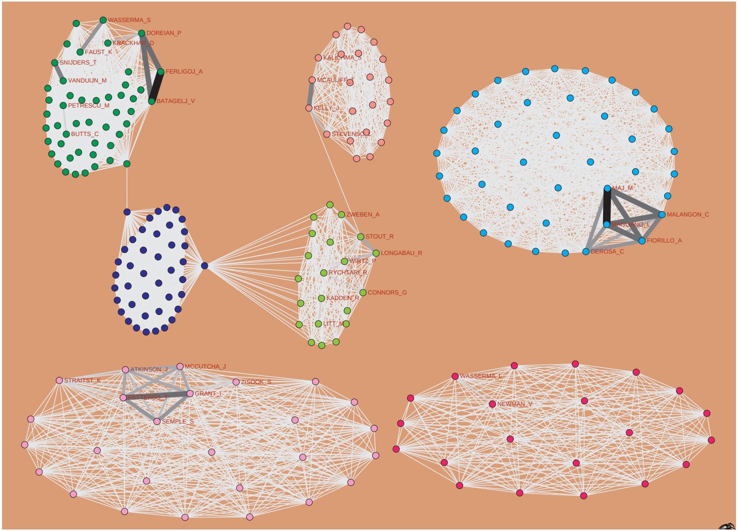
Cores of orders 10–21 in Computational Geometry



p_S-core at level 46 in Computational Geometry network



Cores of orders 20–47 in **Co**(SN5)

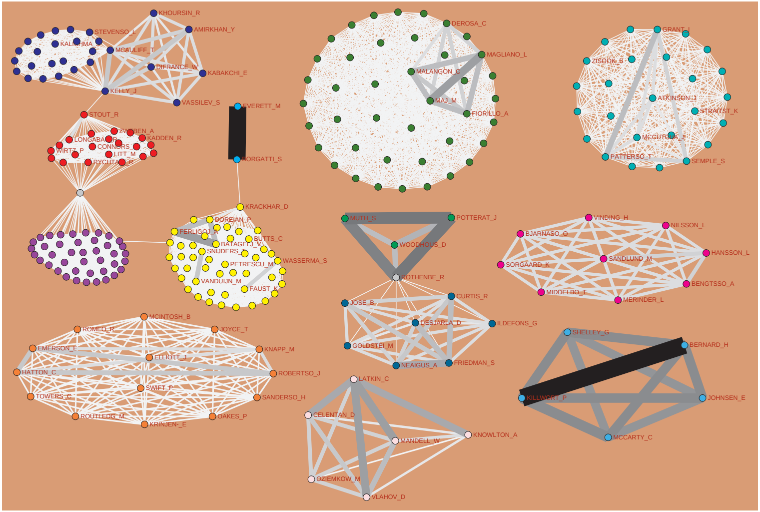


Papers by number of authors

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 bluring the collaboration structure.

outdeg_{WA}	frequency	outdeg_{WA}	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)

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



Second co-authorship network

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

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

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

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

Collaborativness: $K_i = 1 - S_i$

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

The "best" authors in Social Networks

i	author	cn_{ii}	total	K_i	i	author	cn_{ii}	total	K_i
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
3	Doreian,P	34.44	47	0.267	28	Rothenberg,R	9.82	28	0.649
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
6	Wellman,B	26.87	41	0.345	31	Batagelj,V	9.69	20	0.516
7	Leydesdorf,L	24.37	35	0.304	32	Mizruchi,M	9.67	15	0.356
8	White,H	23.50	33	0.288	33	[Anon]	9.00	9	0.000
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
11	Everett,M	16.92	31	0.454	36	Lazega,E	8.50	12	0.292
12	Litwin,H	16.00	21	0.238	37	Knoke,D	8.33	11	0.242
13	Freeman,L	15.53	20	0.223	38	Ferligoj,A	8.19	19	0.569
14	Barabasi,A	14.99	35	0.572	39	Brewer,D	8.03	11	0.270
15	Snijders,T	14.99	30	0.500	40	Klovdahl,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

For vectors $x = [x_1, x_2, \ldots, x_n]$ and $y = [y_1, y_2, \ldots, y_m]$ 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

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

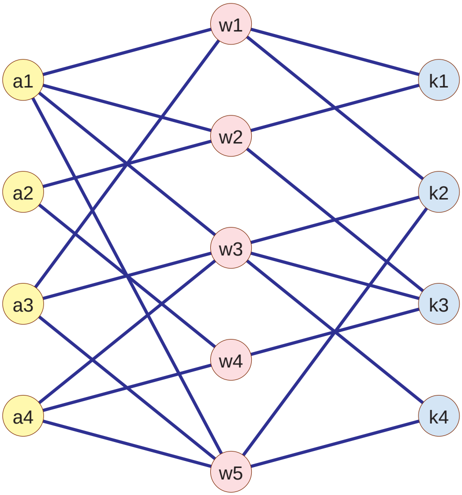
Therefore, if $S_x = S_y = 1$ then also $S = 1$.

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

$$\mathbf{AK} = \mathbf{AW} \cdot \mathbf{WK} = \sum_w \mathbf{H}_w \quad \text{where} \quad \mathbf{H}_w = \mathbf{WA}[w, \cdot] \circ \mathbf{WK}[w, \cdot]$$

For binary (weights) networks we have $\mathbf{H}_w = K_{N_{WA}(w), N_{WK}(w)}$.

$WA^T \cdot WK$



Example

$$\mathbf{H} = \mathbf{AW} \cdot \mathbf{WK}, \quad \mathbf{WA} = \begin{array}{cc} & \begin{matrix} a_1 & a_2 & a_3 & a_4 \end{matrix} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{matrix} & \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \end{array}, \quad \mathbf{WK} = \begin{array}{cc} & \begin{matrix} k_1 & k_2 & k_3 & k_4 \end{matrix} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{matrix} & \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \end{array}$$

$$\begin{array}{cc} \mathbf{H} & \begin{matrix} k_1 & k_2 & k_3 & k_4 \end{matrix} \\ \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} & \begin{bmatrix} 2 & 3 & 2 & 2 \\ 1 & 0 & 2 & 0 \\ 1 & 3 & 1 & 2 \\ 0 & 2 & 2 & 2 \end{bmatrix} \end{array} = \begin{array}{cc} \mathbf{H}_1 & \begin{matrix} k_1 & k_2 & k_3 & k_4 \end{matrix} \\ \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} & \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array} + \begin{array}{cc} \mathbf{H}_2 & \begin{matrix} k_1 & k_2 & k_3 & k_4 \end{matrix} \\ \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} & \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array} +$$

$$\begin{array}{cc} \mathbf{H}_3 & \begin{matrix} k_1 & k_2 & k_3 & k_4 \end{matrix} \\ \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \end{array} + \begin{array}{cc} \mathbf{H}_4 & \begin{matrix} k_1 & k_2 & k_3 & k_4 \end{matrix} \\ \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} & \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array} + \begin{array}{cc} \mathbf{H}_5 & \begin{matrix} k_1 & k_2 & k_3 & k_4 \end{matrix} \\ \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \end{array}$$

Fractional approach

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

and define the *normalized* matrices

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

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

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

$$\mathrm{diag}(\frac{1}{S^{WA}}) = \begin{array}{c c} & \begin{array}{ccccc} w_1 & w_2 & w_3 & w_4 & w_5 \end{array} \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{array} & \left[\begin{array}{ccccc} 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/3 \end{array}\right] \end{array}$$

$$\mathrm{diag}(\frac{1}{S^{WK}}) = \begin{array}{c c} & \begin{array}{ccccc} w_1 & w_2 & w_3 & w_4 & w_5 \end{array} \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{array} & \left[\begin{array}{ccccc} 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \end{array}\right] \end{array}$$

$$\mathbf{WAn} = \begin{array}{c c} & \begin{array}{cccc} a_1 & a_2 & a_3 & a_4 \end{array} \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{array} & \left[\begin{array}{cccc} 1/2 & 0 & 1/2 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/2 & 0 & 1/2 \\ 1/3 & 0 & 1/3 & 1/3 \end{array}\right] \end{array}, \quad \mathbf{WKn} = \begin{array}{c c} & \begin{array}{cccc} k_1 & k_2 & k_3 & k_4 \end{array} \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{array} & \left[\begin{array}{cccc} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{array}\right] \end{array}$$

...example

normalized matrices

$$\mathbf{F}_1 = \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array} \begin{array}{c} \begin{array}{cccc} k_1 & k_2 & k_3 & k_4 \end{array} \\ \begin{bmatrix} 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \qquad \mathbf{F}_5 = \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array} \begin{array}{c} \begin{array}{cccc} k_1 & k_2 & k_3 & k_4 \end{array} \\ \begin{bmatrix} 0 & 1/6 & 0 & 1/6 \\ 0 & 0 & 0 & 0 \\ 0 & 1/6 & 0 & 1/6 \\ 0 & 1/6 & 0 & 1/6 \end{bmatrix} \end{array}$$

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

Third co-authorship network

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

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

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

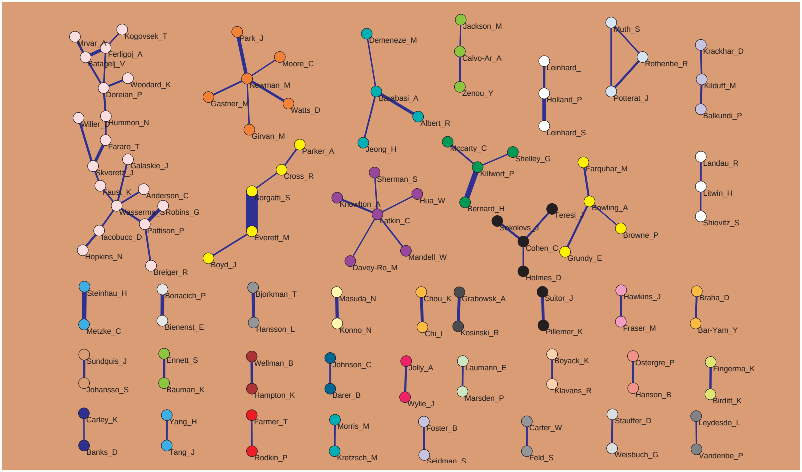
$$\sum_{i \in A} \sum_{j \in A} n_{pi} n_{pj} = 1$$

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

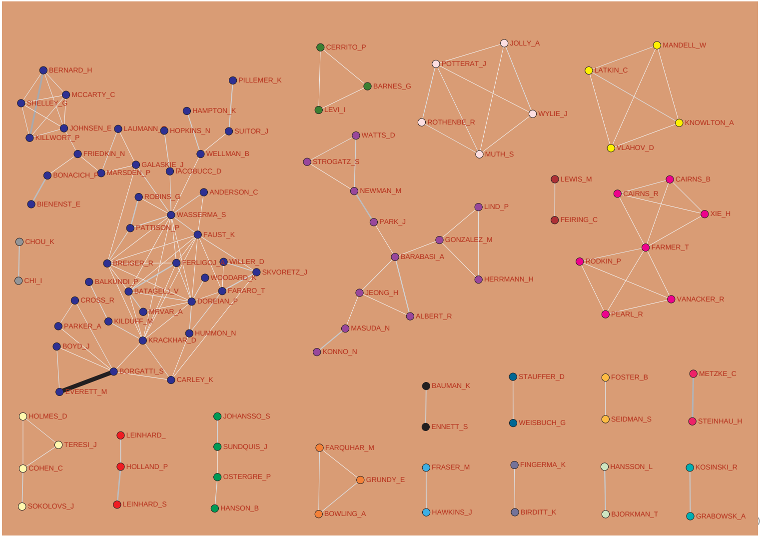
$\sum_{j \in A} ct_{ij} = \sum_{p \in W} n_{pi}$ = the total contribution of author i to works from W.

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

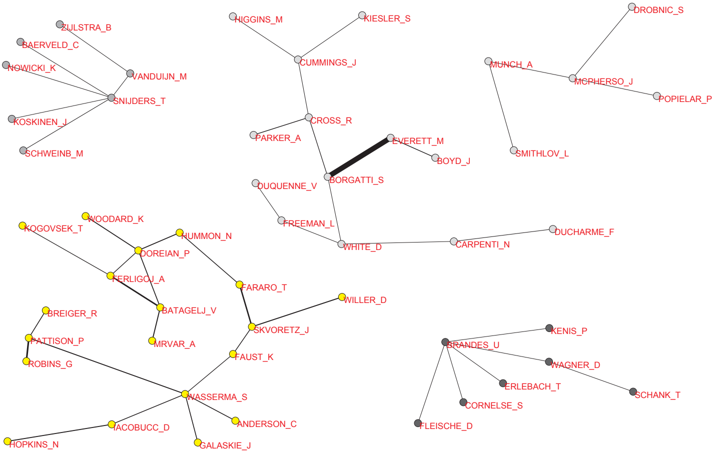
Components in **Ct**(SN5) cut at level 0.5



p_S-core at level 0.75 in **Ct**(SN5)



Some link islands [5,20] in **Ct**(SN5)



Fourth co-authorship network

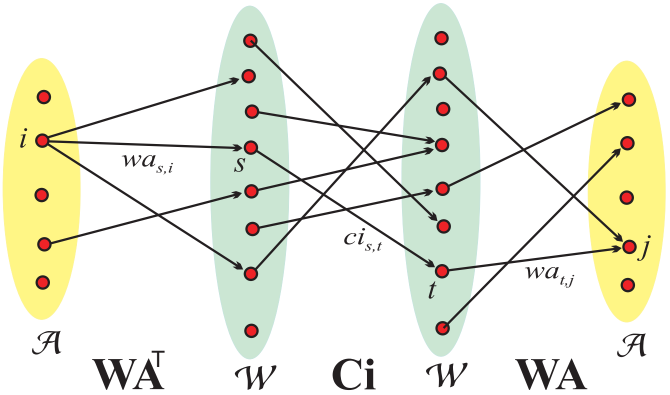
$\mathbf{Ct}' = \mathbf{N}^T * \mathbf{N}'$, where $n'_{pi} = wa_{pi} / \max(1, \text{outdeg}_{WA}(p) - 1)$ is the Newman's normalization.

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

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

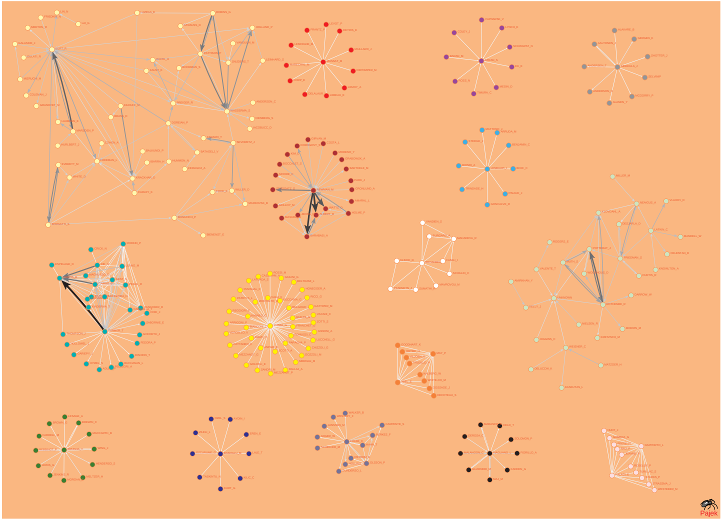
$$ct'_{ij} = \sum_p \frac{2 \cdot wa_{pi} \cdot wa_{pj}}{\max(1, \text{outdeg}_{WA}(p)) \cdot \max(1, \text{outdeg}_{WA}(p) - 1)}$$

Authors' citations network



$\mathbf{Ca} = \mathbf{AW} * \mathbf{Ci} * \mathbf{WA}$ 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



Linking through a network

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:

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

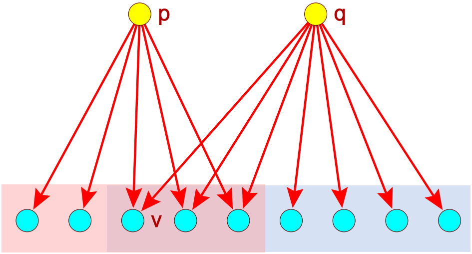
It is easy to verify that:

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

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

Bibliographic Coupling

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

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

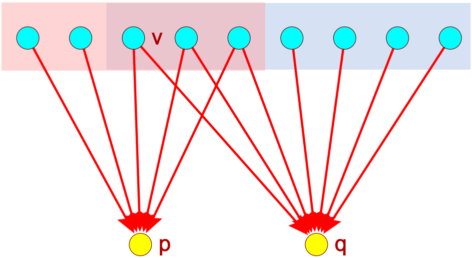
$$\mathbf{biCo} = \mathbf{Ci} * \mathbf{Ci}^T$$

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

Co-citation is a concept with strong parallels with bibliographic coupling (Small and Marshakova 1973). 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

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

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

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

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

Normalization of Bibliographic Coupling

The fractional approach can not bi directly applied to bibliographic coupling – to get the outer product decomposition work we would need to normalize **Ci** 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

$$\mathbf{biC} = \mathbf{Cin} \cdot \mathbf{Ci}^T$$

we have

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

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

Normalization of Bibliographic Coupling

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

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

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

or, may be more interesting

$$\mathbf{biCog}_{pq} = \sqrt{\mathbf{biC}_{pq} \cdot \mathbf{biC}_{qp}} = \frac{|\mathbf{Ci}(p) \cap \mathbf{Ci}(q)|}{\sqrt{|\mathbf{Ci}(p)| \cdot |\mathbf{Ci}(q)|}}$$ Geometric mean Salton cosinus

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

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

All these measures are similarities.

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

Normalization of Bibliographic Coupling

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 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

We introduce a notion of a *temporal quantity*

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

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 $\maltese$ 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 $+ : A \times A \to A$ and $\cdot : A \times A \to A$.
Let $A_\maltese(\mathcal{T})$ denote the set of all temporal quantities over $A_\maltese$ 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


```
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)]
```

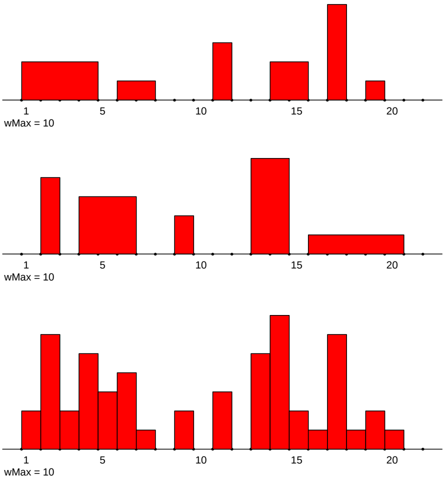
They are visually displayed at the bottom half of figures on the following slides.

Addition of temporal quantities.

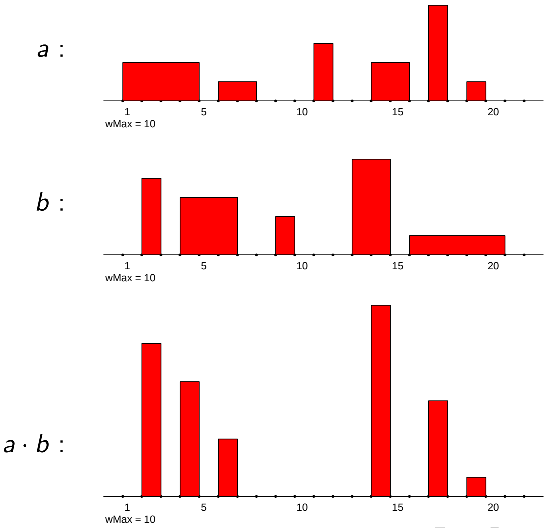
$a:$

$b:$

$a+b:$



Multiplication of temporal quantities.



Temporal affiliation networks

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

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

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

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

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

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

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

Cumulative properties

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

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

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* $cum(a)$ is defined as

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

where $s_{k+1} = last$ and $V_i = \sum_{j=1}^{i} v_j$.
A temporal network is cumulative for a weight w iff all its values are cumulative.

Temporal properties

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

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

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

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

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

Group productivity

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

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

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

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

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

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

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

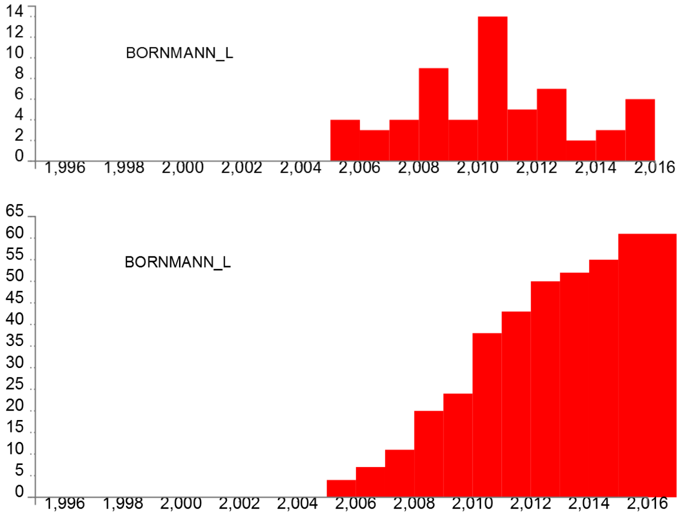
see the top of the figure. The corresponding cumulative productivity is

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

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

Temporal properties in networks on peer review

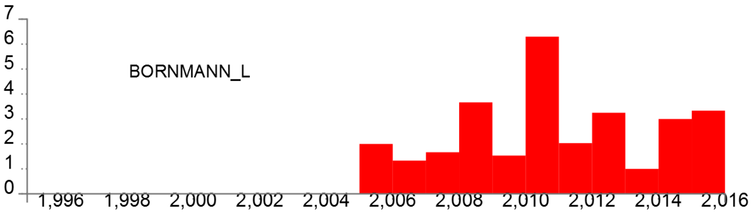
Productivity and cumulative productivity of Lutz Bornmann



The fractional productivity of Lutz Bornmann is

$$
\begin{aligned}
fpr \quad = \quad & [(2005, 2006, 2.0), (2006, 2007, 1.333), (2007, 2008, 1.667), \\
& (2008, 2009, 3.667), (2009, 2010, 1.533), (2010, 2011, 6.3), \\
& (2011, 2012, 2.033), (2012, 2013, 3.25), (2013, 2014, 1.0), \\
& (2014, 2015, 3.0), (2015, 2016, 3.333)]
\end{aligned}
$$

see the figure



Citations to a paper through years

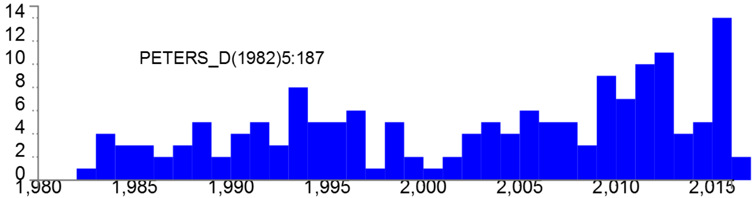
In the citation network **Cite** for the peer review bibliography the most cited, 164, paper is Peters, D. P., Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. Behavioral and Brain Sciences, 5(2), 187-255. The temporal quantity $citP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

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

See the figure at the next slide.

Peer review

Citations to Peters



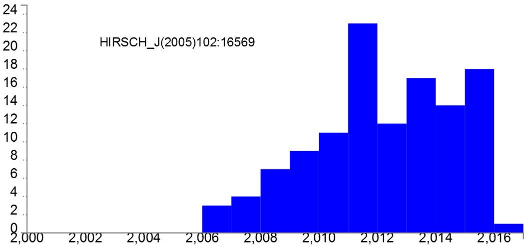
Similarly we could look at the number of works by year $wy = \sum_{j \in J} iS(\mathbf{WJi}, j)$, the popularity of a keyword k: $pop(k) = iS(\mathbf{WKi}, k)$, etc.

Peer review

Citations to Hirsch

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
$citH = iS(\mathbf{Cii}, \text{HIRSCH\_J(2005)102:16569})$

$$\begin{aligned} citH \quad = \quad & [(2006, 2007, 3), (2007, 2008, 4), (2008, 2009, 7), (2009, 2010, 9), \\ & (2010, 2011, 11), (2011, 2012, 23), (2012, 2013, 12), (2013, 2014, 17), \\ & (2014, 2015, 14), (2015, 2016, 18), (2016, 2017, 1)] \end{aligned}$$



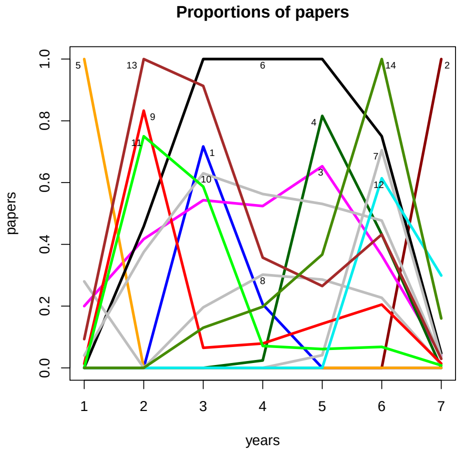
Main journals publishing on peer review

To identify the main journals publishing on peer review 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 the 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

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



i	journal
1	BEHAV BRAIN SCI
2	BMJ OPEN
3	BRIT MED J
4	CUTIS
5	J ASSOC OFF AGR CHEM
6	JAMA-J AM MED ASSOC
7	J SEX MED
8	LANCET
9	MED J AUSTRALIA
10	NATURE
11	NEW ENGL J MED
12	PLOS ONE
13	SCIENCE
14	SCIENTOMETRICS

Multiplication of temporal networks

Let **A** on $A \times P$ and **B** on $P \times B$ be (matrices of) co-occurence 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 ⌘) 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

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

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

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

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

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

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.

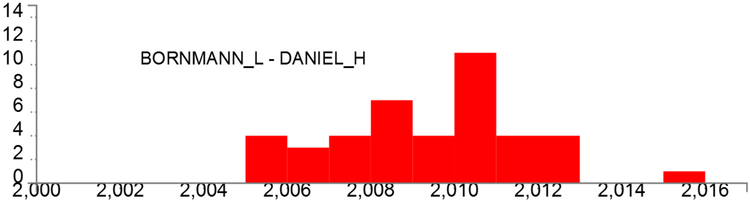
Example: Temporal coauthorship network

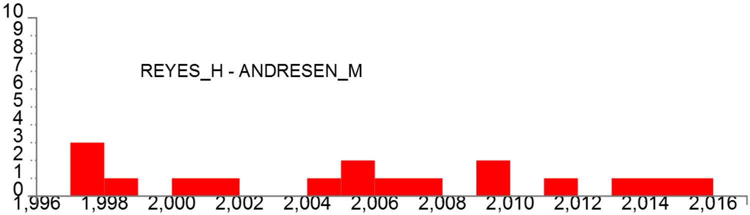
For the peer review data we get the largest values
$co[\text{BORNMANN\_L}, \text{DANIEL\_H}] = 42$,
$co[\text{MOHER\_D}, \text{ALTMAN\_D}] = 24$,
$co[\text{REYES\_H}, \text{ANDRESEN\_M}] = 17$.
The corresponding temporal quantities
$bd = tq(\text{BORNMANN\_L}, \text{DANIEL\_H})$ and
$ra = tq(\text{REYES\_H}, \text{ ANDRESEN\_M})$ are

$$\begin{aligned}
bd &= [(2005, 2006, 4), (2006, 2007, 3), (2007, 2008, 4), (2008, 2009, 7),\\
&\quad (2009, 2010, 4), (2010, 2011, 11), (2011, 2013, 4), (2015, 2016, 1)]\\
ra &= [(1997, 1998, 3), (1998, 1999, 1), (2000, 2002, 1), (2004, 2005, 1),\\
&\quad (2005, 2006, 2), (2006, 2008, 1), (2009, 2010, 2), (2011, 2012, 1),\\
&\quad (2013, 2016, 1)]
\end{aligned}$$

Both temporal quantities are presented on the next slide.

Example: Temporal coauthorship network





Example: Temporal citations between journals

The derived network describing citations between journals is obtained as

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

Note that the third network in the product is cumulative.
The weight of the element jcj_{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

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

$$
\begin{aligned}
jm \quad = \quad & [(1973, 1976, 1), (1988, 1989, 1), (1989, 1990, 2), (1990, 1991, 16), (1991, 1992, 1),\\
& (1992, 1993, 11), (1993, 1994, 4), (1994, 1995, 44), (1995, 1996, 9), (1996, 1997, 2),\\
& (1997, 1998, 3), (1998, 1999, 68), (1999, 2000, 14), (2000, 2001, 10), (2001, 2002, 7),\\
& (2002, 2003, 60), (2003, 2004, 11), (2004, 2005, 4), (2005, 2006, 1), (2006, 2007, 16),\\
& (2007, 2008, 8), (2008, 2009, 2), (2009, 2010, 4), (2012, 2013, 4), (2013, 2014, 3),\\
& (2014, 2015, 7), (2015, 2016, 5)]\\
sm \quad = \quad & [(1991, 1992, 1), (1995, 1996, 2), (1998, 1999, 2), (2001, 2002, 1), (2003, 2004, 1),\\
& (2005, 2006, 6), (2006, 2007, 10), (2007, 2009, 4), (2009, 2010, 7), (2010, 2011, 16),\\
& (2011, 2012, 14), (2012, 2013, 9), (2013, 2014, 34), (2014, 2015, 19), (2015, 2016, 17),\\
& (2016, 2017, 1)]
\end{aligned}
$$

and are presented on the next slide.

Example: Journals Selfcitations

Bibliographic networks

V. Batagelj

Networks

Statistics

Citation

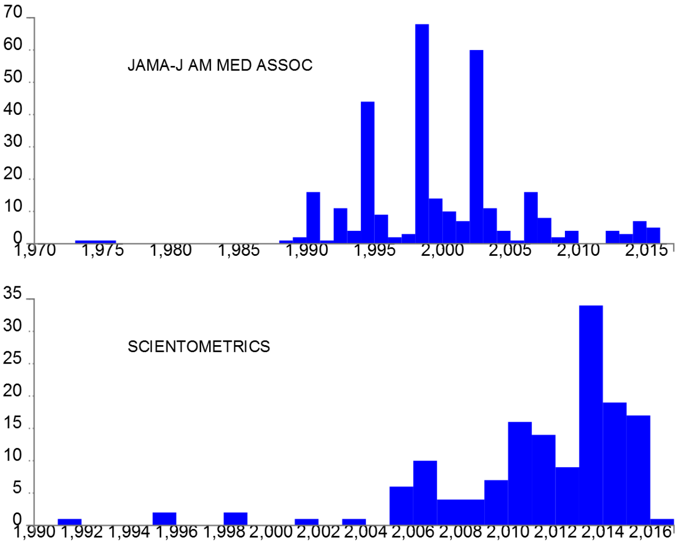
Two-mode Ns

Multiplication

Derived Ns

Temporal Ns

References

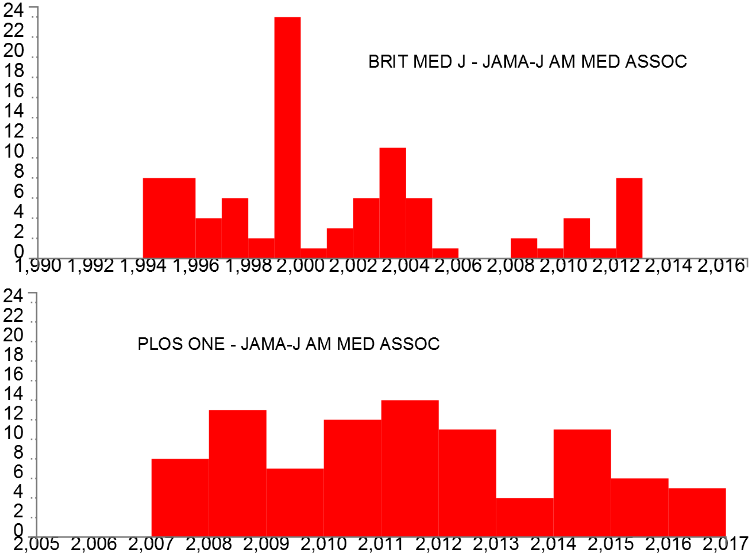


Example: Temporal citations between journals

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

$$
\begin{aligned}
bj &= [(1994, 1996, 8), (1996, 1997, 4), (1997, 1998, 6), (1998, 1999, 2), (1999, 2000, 24),\\
&\quad (2000, 2001, 1), (2001, 2002, 3), (2002, 2003, 6), (2003, 2004, 11), (2004, 2005, 6),\\
&\quad (2005, 2006, 1), (2008, 2009, 2), (2009, 2010, 1), (2010, 2011, 4), (2011, 2012, 1),\\
&\quad (2012, 2013, 8)]\\
pj &= [(2007, 2008, 8), (2008, 2009, 13), (2009, 2010, 7), (2010, 2011, 12), (2011, 2012, 14),\\
&\quad (2012, 2013, 11), (2013, 2014, 4), (2014, 2015, 11), (2015, 2016, 6), (2016, 2017, 5)]
\end{aligned}
$$

Example: Temporal citations between journals



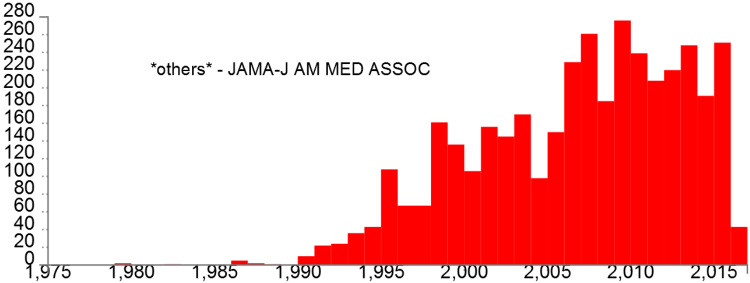
Example: Temporal citations between journals

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 $jci = iS(\mathbf{JCJ}, \text{JAMA-J AM MED ASSOC})$:

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

See the figure on the next slide.

Example: Temporal citations between journals



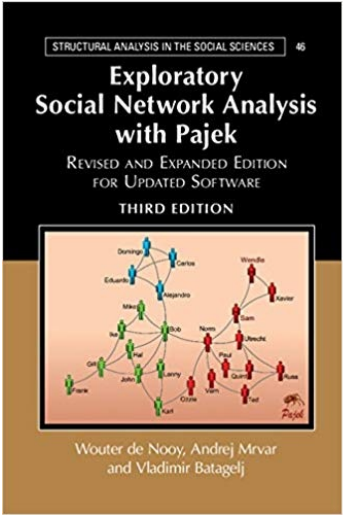
Citations between authors

Similarly we get the temporal network describing citations between authors

$$\mathbf{ACA} = \mathbf{WAi}^T \cdot \mathbf{CiI} \cdot \mathbf{WAc}$$

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



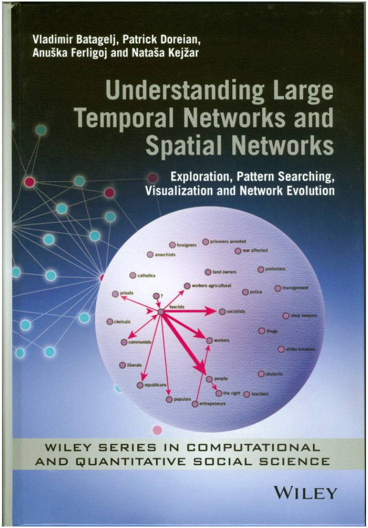
An introduction to social network analysis with Pajek is available in the book ESNA 3 (de Nooy, Mrvar, Batagelj, CUP 2005, 2011, 2018).

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



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 I

- 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 scientic 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.

Bibliographic networks

V. Batagelj

References II

- Batagelj, V.: On fractional approach to analysis of linked networks. Scientometrics (2020). DOI: 10.1007/s11192-020-03383-y
- Batagelj, V., Maltseva, D.: Temporal bibliographic networks. Journal of Informetrics, Volume 14, Issue 1, February 2020, 101006. DOI: 10.1016/j.joi.2020.101006
- Batagelj, V., Ferligoj, A., Squazzoni, F.: The emergence of a field: a network analysis of research on peer review. Scientometrics (October 2017) Volume 113, Issue 1, pp 503–532.
- 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; Third Edition. Structural Analysis in the Social Sciences, Cambridge University Press, September 2018.

References III

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.

Maltseva, D., Batagelj, V. iMetrics: the development of the discipline with many names. Scientometrics 125 (2020), pages 313–359.

Maltseva, D., Batagelj, V. Journals publishing social network analysis. Scientometrics 126 (2021), pages 3593–3620.

Maltseva, D., Batagelj, V.: Towards a systematic description of the field using keywords analysis: main topics in social networks. Scientometrics, 2020. DOI: 10.1007/s11192-020-03365-0

References IV

Maltseva, D., Batagelj, V.: Social network analysis as a field of invasions: bibliographic approach to study SNA development. Scientometrics, 121(2019)2, 1085-1128. DOI 10.1007/s11192-019-03193-x

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.

References V

- 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)
- Pajek's wiki. http://pajek.imfm.si
- Vladimir Batagelj, Andrej Mrvar: Pajek manual.