

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

Introduction to bibliographic network analysis

Vladimir Batagelj

UP Koper and IMFM Ljubljana

Faculty of Economics in Osijek
Osijek, October 2, 2023

Outline

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

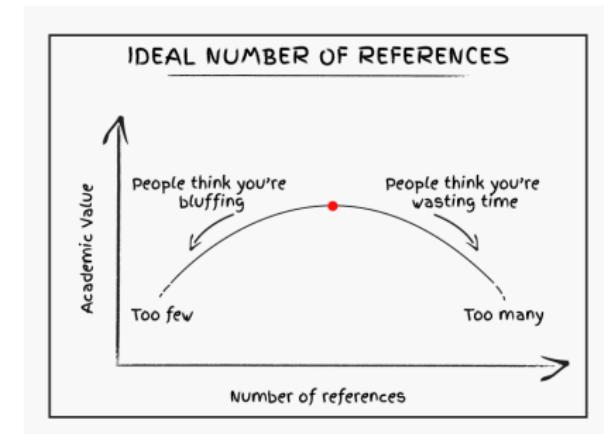
Normalizations

Groups

Bibliographic coupling
and co-citation

Resources

- 1 Introduction
- 2 Statistics
- 3 Citation
- 4 Binary projections
- 5 Fractional approach
- 6 Normalizations
- 7 Groups
- 8 Bibliographic coupling and co-citation
- 9 Resources



J-Gate

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

Current version of slides (October 2, 2023 at 01:33): [slides PDF](#)
<https://github.com/bavla/biblio/>

Motivation

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

Bibliographic databases such as WoS (Web of Science), Scopus, Google Scholar, etc. contain a lot of interesting information.

Usually, they are used for searching for works, evaluation of researchers or groups or institutions or journals, etc.

We can get more – higher-order bibliographic services.

- Student preparing a thesis. Which works to read?
- Journal editor. Select a reviewer for a given paper.
- Wrote a paper. To which journal to submit it?
- Preparing a project. Whom to invite to collaborate?
- etc.

Bibliographic network analysis provides methods to support such services.

Networks

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

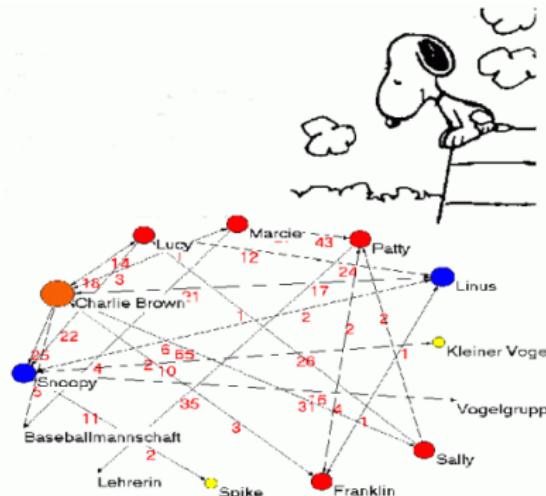
Citation

Binary projections
Fractional approach

Normalizations
Groups

Bibliographic coupling and co-citation

Resources



Alexandra Schuler/ Marion Laging-Glaser:
Analyse von Snoopy Comics

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

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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

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

Graph

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

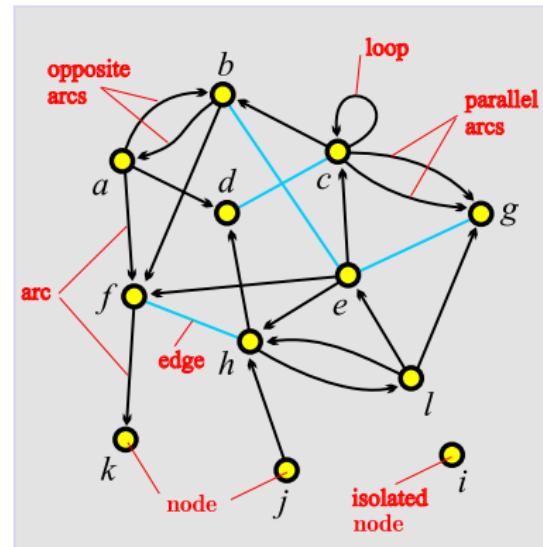
Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources



unit, actor – node, vertex
tie, line – link, edge, arc

arc = directed link, (a, d)
 a is the *initial* node,
 d is the *terminal* node.

edge = undirected link,
 $(c: d)$
 c and d are *end* nodes.

Types of networks

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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

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

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

In a *linked* or *multimodal* network $\mathcal{N} = ((\mathcal{V}_1, \dots, \mathcal{V}_j), (\mathcal{L}_1, \dots, \mathcal{L}_k), \mathcal{P}, \mathcal{W})$ 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.

Record from Web of Science

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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

Networks from bibliographies

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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

- citation network **Ci**: works × works;
- authorship network **WA**: works × authors, for works without complete description only the first author is known;
- keywords network **WK**: works × keywords, only for works with complete description;
- journals network **WJ**: works × 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.

Example networks

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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](#)

SNA18: WoS, January 2018, $|W| = 1297133/70792$,
 $|A| = 395971/93011$, $|K| = 32409/32409$, $|J| = 69146/8943$. [paper](#)

Statistics

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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.

weighted degree of node v , $w\deg(v)$ = sum of the weights of links with v as an endnode;

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	Zuckerman, 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

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

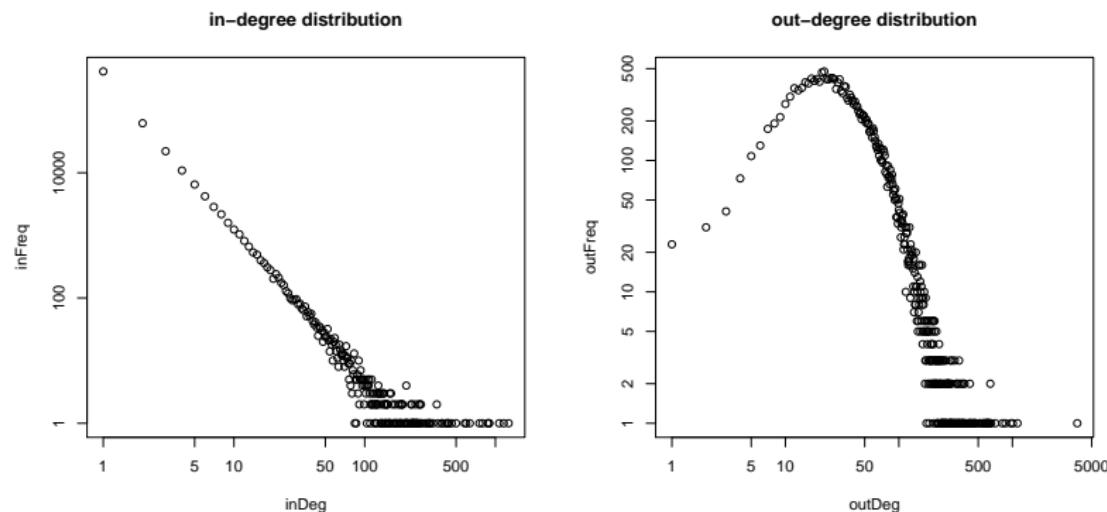
Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources



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

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

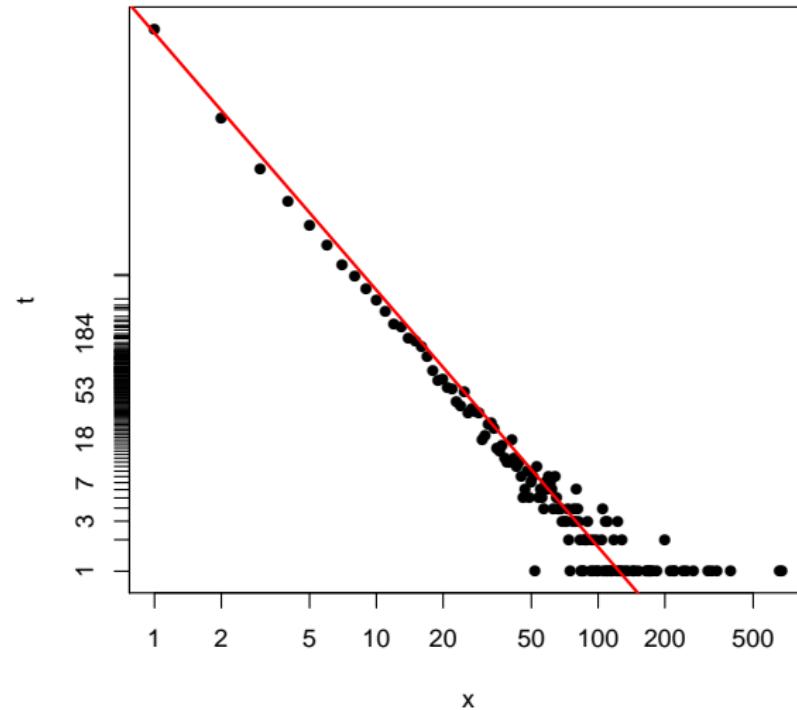
Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources



Distributions

year partition lognormal distribution fit

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

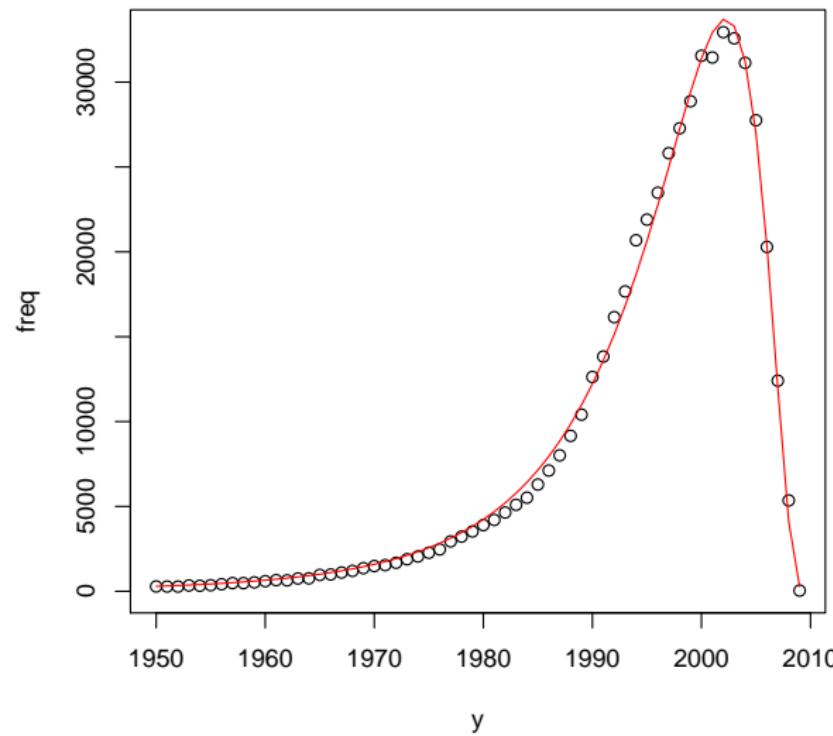
Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources



Distributions

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

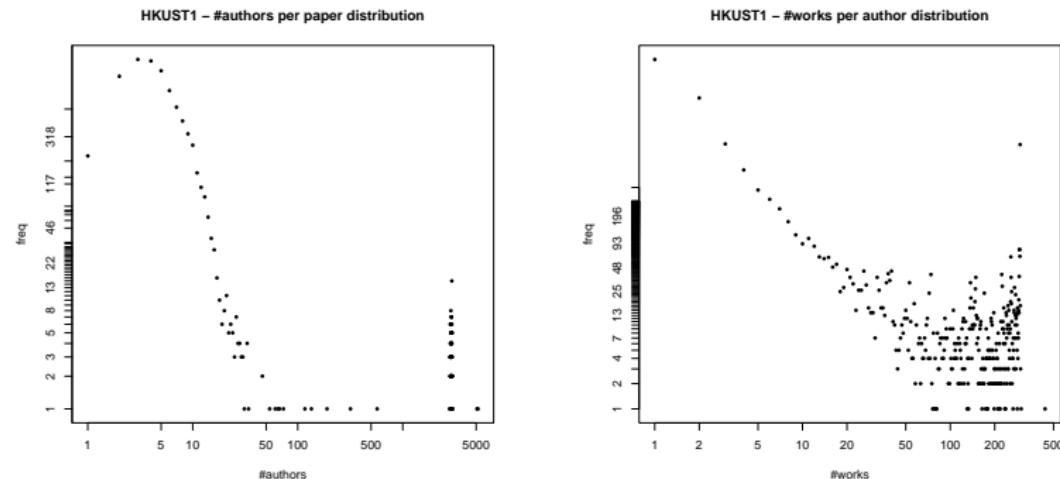
Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources



The Hong Kong University of Science and Technology (HKUST) 2017–2019

5215 co-authors: The ATLAS collaboration, & The CMS collaboration (2019). Combinations of single-top-quark production cross-section measurements and $|f_{LV} V_{tb}|$ determinations at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS experiments. Journal of High Energy Physics, 2019(5), Article 88.
[https://doi.org/10.1007/JHEP05\(2019\)088](https://doi.org/10.1007/JHEP05(2019)088) / Springer

Distributions

number of keywords in **ZBMath**

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

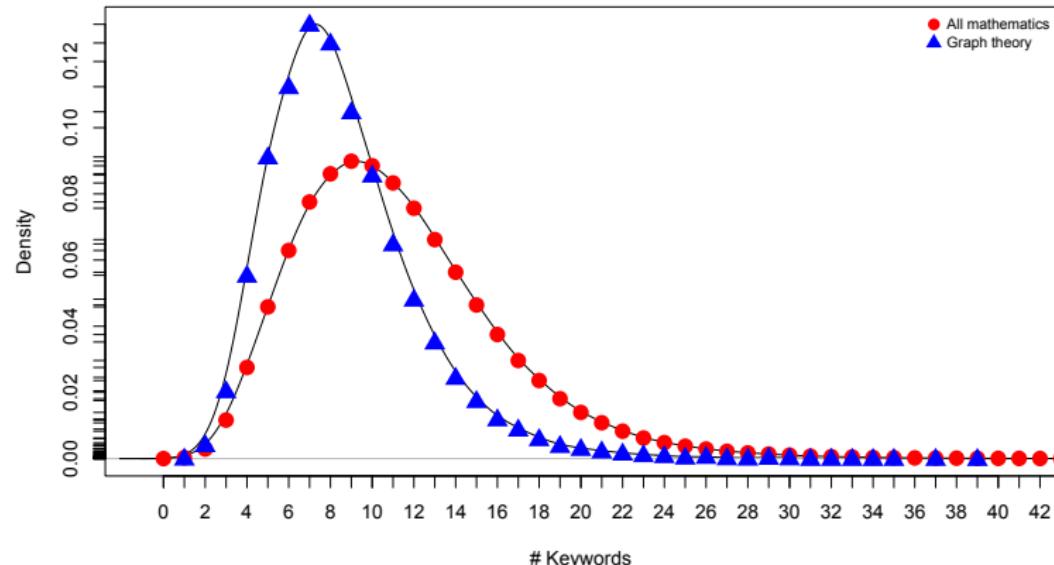
Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources



Distributions

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

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

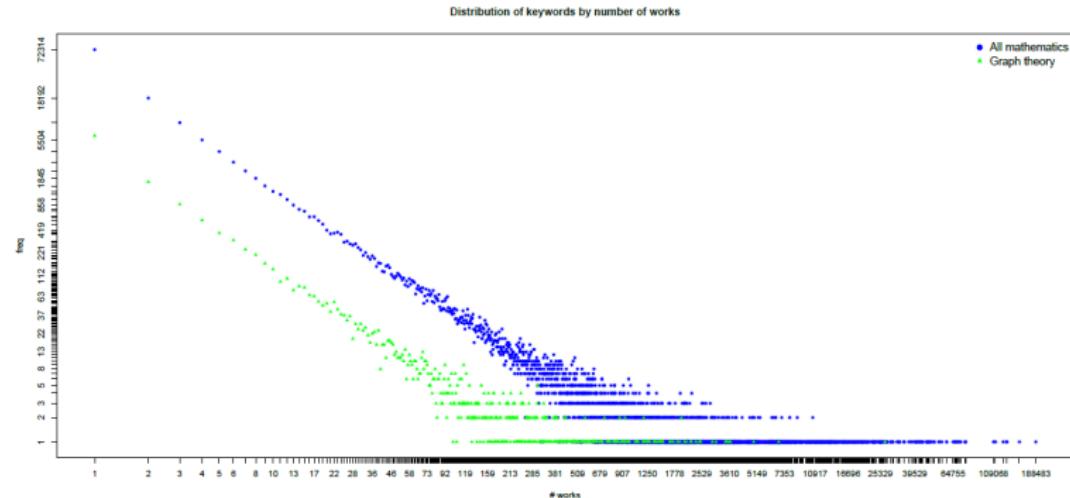
Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources



Temporal distribution

Number of authors SNA 2018

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

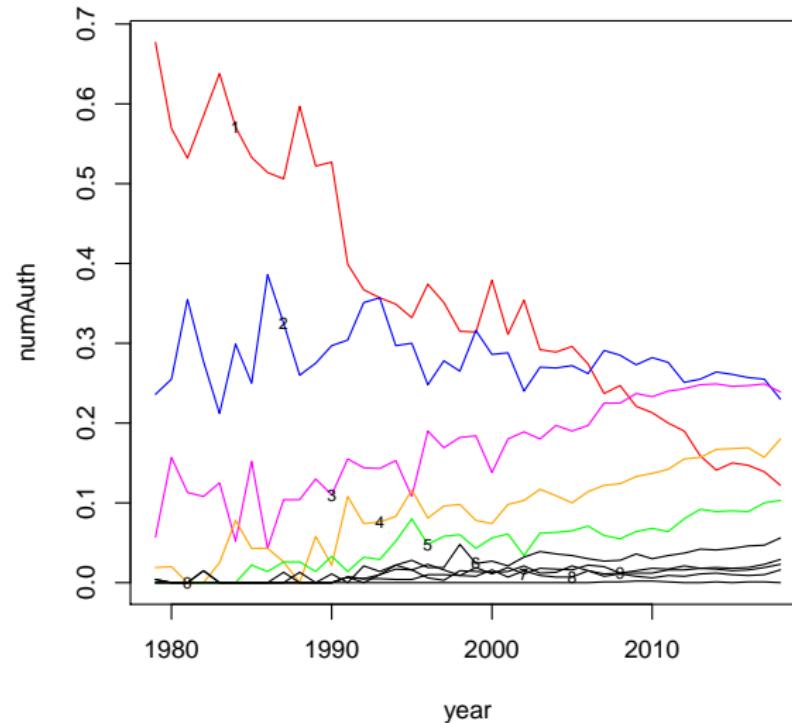
Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

SocNet



Citations between years

SNA 2018

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

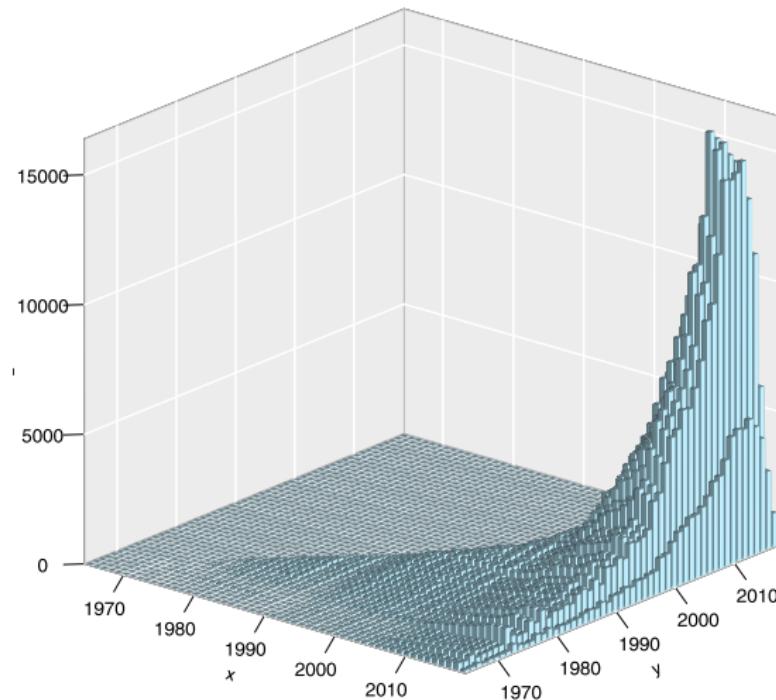
Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

SocNet



Citations per year

SNA 2018; normalized curves

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

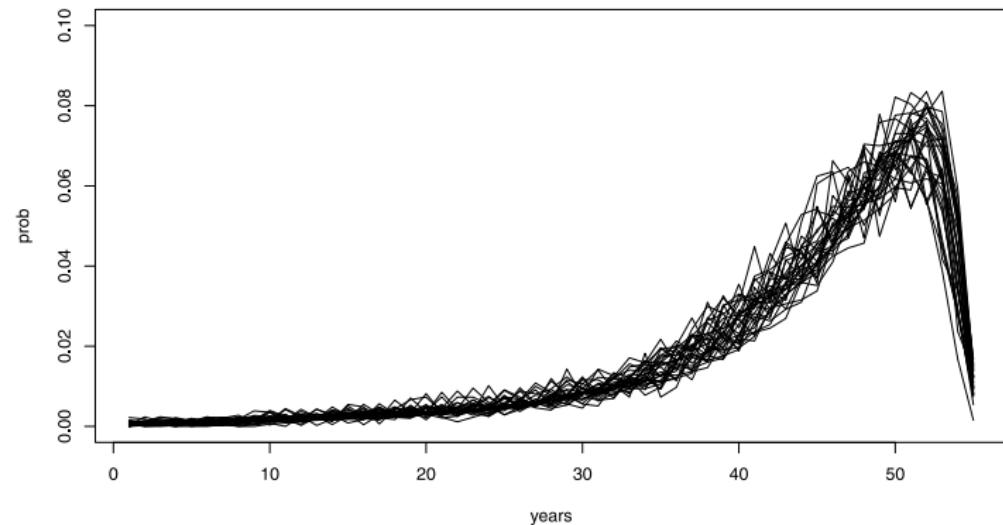
Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources



SocNet

Citation networks

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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

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

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 a network into an acyclic network by shrinking strong components and deleting loops.

A better way is the preprint transformation.

Citation network

Preprint transformation

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

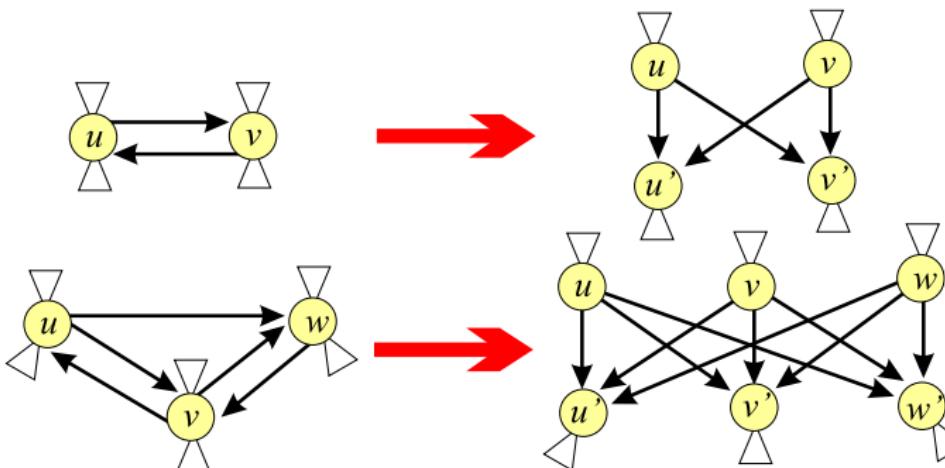
Citation

Binary projections
Fractional approach

Normalizations
Groups

Bibliographic coupling and co-citation

Resources



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

Standardized citation network

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

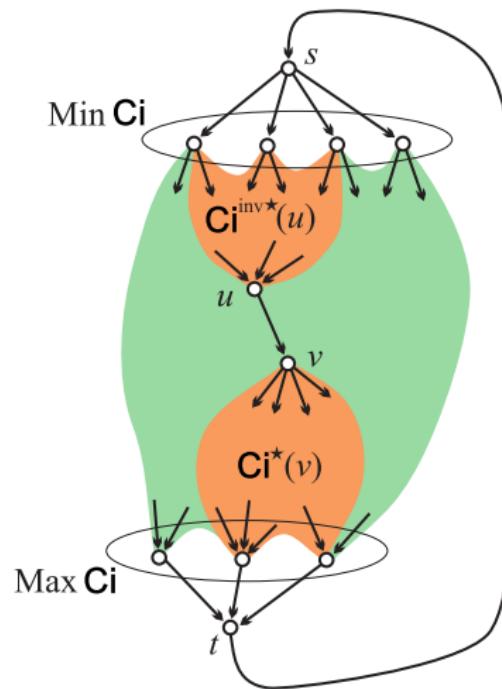
Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources



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

Search path count method

Hummon and Doreian

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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*:

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

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

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

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

The value $w(u, v)$ is equal to the probability that a random $s-t$ path passes through the arc (u, v) .

Cuts

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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

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

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

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

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

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

Citation weights

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

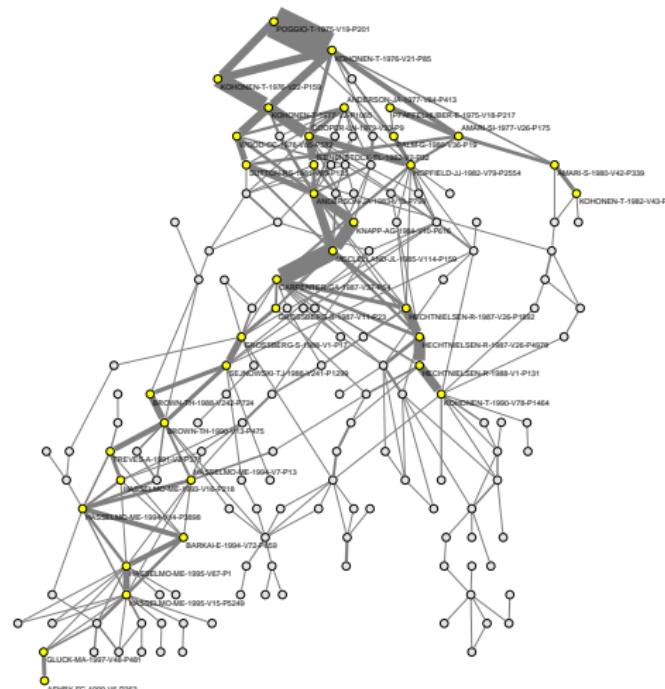
Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources



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

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

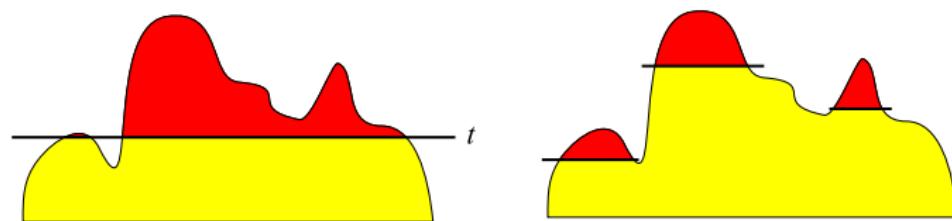
Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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



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

... Islands

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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

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

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

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

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

... Islands

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

A set of nodes $C \subseteq \mathcal{V}$ is a *regular node island* in a network

$\mathcal{N} = (\mathcal{V}, \mathcal{L}, p)$, $p : \mathcal{V} \rightarrow \mathbb{R}$ iff it induces a connected subgraph and the nodes from the island are 'higher' than the neighboring nodes

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

A set of nodes $C \subseteq \mathcal{V}$ is a *regular link island* in a network

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

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

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

US patents

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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

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

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

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

Distribution of island size

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

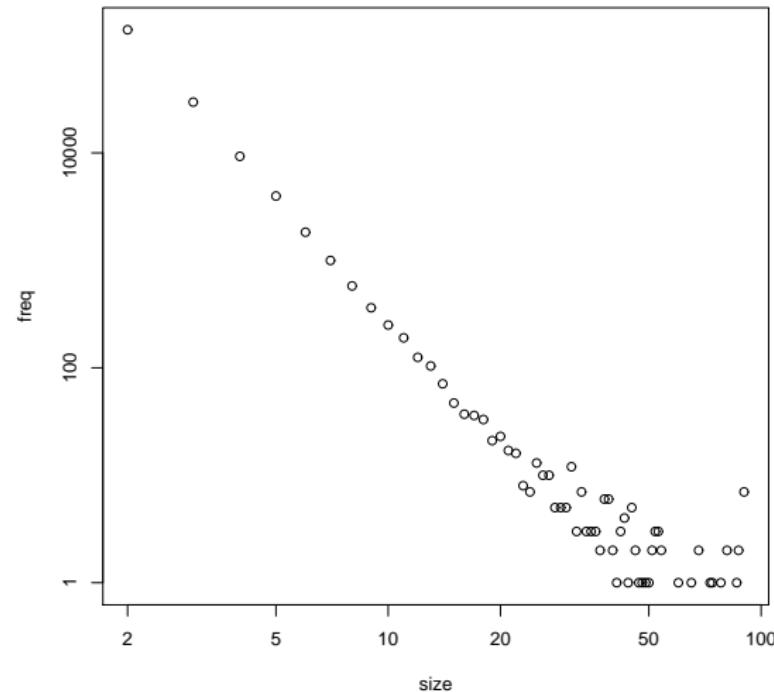
Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources



Main path and main island in US Patents

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

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

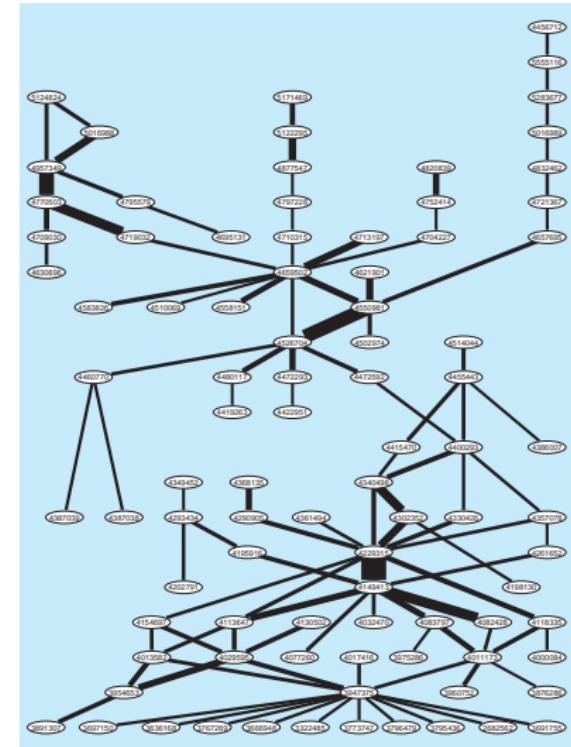
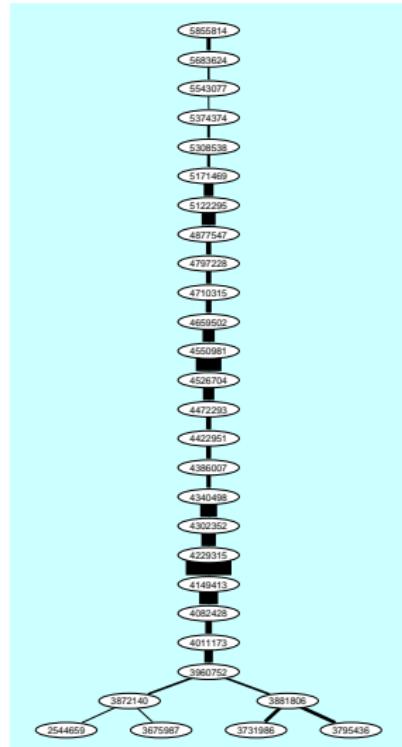
Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources



Main island – Liquid crystal display

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources

Table 1: Patents on the liquid-crystal display

patent	date	author(s) and title
2049629	Mar 13, 1951	Wender, et al. Preparation of polyesters containing short and the like and the formation and use thereof
2082562	Jul 29, 1954	Wender, et al. Reduction of aromatic carbonyl groups by the addition of aromatic elements utilizing an organic nematic compound
3322485	May 30, 1967	Josephson. Preparation of polyacrylate ester compounds having an undeterred image on a disturbed background
3636168	Jun 18, 1972	Goto, et al. Liquid crystal compositions and devices
3691755	Sep 19, 1972	Tokutomi, et al. A range of nematic liquid crystal compositions
3697150	Oct 19, 1972	Wysoki. Electro-optic systems in which an electrophoretic particle moves in response to an electric field through a liquid crystal to reduce the turn-off time
3731966	May 8, 1973	Fengues. Display devices utilizing liquid crystal light
3767289	Oct 23, 1973	Arivazhagan, et al. Class of stable trans-styrene compounds, some displaying nematic or mesomeric at or near room temperature and others at a range of 0°C
3773477	Nov 20, 1973	Steinmesser. Substituted nancy benzene compounds which exhibit the Kerr effect at isotropic temperatures
3795436	Mar 3, 1974	Hofrichter, et al. Electro-optical liquid-crystalline cell
3797479	Mar 12, 1974	Hofrichter, et al. Electro-optical liquid-crystalline cell which exhibits the Kerr effect at isotropic temperatures
3872140	Mar 18, 1975	Klaesnerman, et al. Liquid crystalline compositions and methods of use
3876286	Aug 8, 1975	Dentzschek, et al. Use of nematic liquid crystalline substances in electro-optical devices
3890200	Mar 28, 1976	Yamada, et al. Liquid crystal materials and devices
3891307	Jun 24, 1976	Tanakozumi, et al. Phase control of the voltages applied to opposite electrodes for a cholesteric to nematic phase transition
3947375	Mar 30, 1976	Goto, et al. Liquid crystal materials and devices
3958053	May 4, 1976	Yamada, et al. Liquid crystal materials and devices having high dielectric constants
3960752	Jun 1, 1976	Klaesnerman, et al. Liquid crystal compositions
3973286	Aug 17, 1976	Goto, et al. Liquid crystal compositions and method of synthesis
4000894	Dec 28, 1976	Held, et al. Liquid crystal mixtures for electro-optical applications
4011173	Mar 8, 1977	Steinmesser. Modified nematic mixtures with increased viscosity
4013582	Mar 22, 1977	Gavrilev. Liquid crystal compounds and electro-optic devices incorporating them
4071746	Apr 12, 1977	Method for preparing some liquid crystal compositions using same
4029095	Jun 14, 1977	Novel liquid crystal compounds and electro-optic devices incorporating them
4032470	Jun 28, 1977	Goto, et al. Novel liquid crystal compounds and electro-optic devices incorporating them
4077260	Mar 7, 1978	Goto, et al. Optically active cyano-biphenyl compounds and liquid crystal materials containing them
4082428	Aug 4, 1978	Goto, et al. Liquid crystal composition and method

Table 2: Patents on the liquid-crystal display

patent	date	author(s) and title
4113647	Apr 11, 1979	Goto, et al. Liquid crystal compositions
4126704	Sep 12, 1979	Custodis, et al. Liquid crystalline materials
4148320	Oct 3, 1979	Krause, et al. Liquid crystalline materials of reduced viscosity
4150252	Oct 17, 1979	Goto, et al. Optically active liquid crystal mixture and liquid crystal devices containing them
4149412	Oct 17, 1979	Goto, et al. Liquid crystal derivative benzoyldiphenyl boronates
4149407	May 15, 1979	Custodis, et al. Liquid crystal materials
4189130	Aug 1, 1980	Goto, et al. Liquid crystal mixtures
4202794	May 13, 1980	Osawa, et al. Nematic liquid crystalline materials
4221362	Aug 1, 1980	Goto, et al. Liquid crystal compositions
4261652	Apr 14, 1981	Goto, et al. Liquid crystal compounds
4269055	Sep 22, 1981	Goto, et al. Liquid crystal compounds and materials and devices containing them
4293434	Oct 6, 1981	Dentzschek, et al. Liquid crystal compounds
4302325	Nov 24, 1981	Eidenschink, et al. Polyphenylketone compounds, the preparation of their esters and ester amides of liquid crystal dioxetanes
4302326	Nov 24, 1981	Eidenschink, et al. Cyclohexylbiphenyl, their preparation and use as liquid crystal components
4304049	Jul 20, 1982	Sugimori, et al. Halogenated ether derivatives
4349452	Sep 14, 1982	Osema, et al. Cyclohexylbicyclohexanes
4352075	Nov 2, 1982	Goto, et al. Liquid crystal compositions containing an aldehyde ring and exhibiting a low dielectric anisotropy and liquid crystal materials and devices incorporating such compositions
4361494	Nov 20, 1982	Osawa, et al. Anisotropic cyclohexane derivatives and ethers
4368135	Jan 11, 1983	Osawa. Anisotropic compounds with a positive Kerr constant
4368607	May 31, 1983	Krause, et al. Liquid crystalline naphthalene derivatives
4370208	Jun 7, 1983	Al, et al. 4-(Trans-4'-alkylcyclohexyl) benzoic acid
4387020	Jun 7, 1983	Sugimori, et al. Trans-4-(trans-4'-alkylcyclohexyl)-cyclohexane carboxylic acid and its salt
4400253	Aug 23, 1983	Eidenschink, et al. Liquid fluorescent heterocyclic compounds and method of synthesis
4413570	Nov 15, 1983	Goto, et al. Liquid fluorescent and electro-optical display elements based thereon
4413925	Dec 6, 1983	Prufrock, et al. Liquid crystalline cyclohexylcarboxylic acid
4422953	Dec 27, 1983	Sugimori, et al. Liquid crystal benzene derivatives
4435441	Jan 19, 1984	Takatori, et al. Nematic liquid crystal composition
4435442	Jan 19, 1984	Takatori, et al. Nematic liquid crystal composition
4460770	Jul 17, 1984	Petrzilka, et al. Liquid crystal mixture
4472250	Sep 18, 1984	Goto, et al. Liquid crystal substances of four rings and liquid crystal compositions containing the same
4472252	Sep 18, 1984	Takatori, et al. Nematic liquid crystal compounds
4481250	Oct 2, 1984	Goto, et al. Liquid crystal compositions
4502974	Mar 5, 1985	Sugimori, et al. High temperature liquid-crystaline ester compounds
4530069	Aug 9, 1985	Al, et al. Cyclohexane derivatives

Table 3: Patents on the liquid-crystal display

patent	date	author(s) and title
6113944	Apr 30, 1987	Baldwell, et al. Dibenzotetralin ethanes and their use in liquid crystal displays and liquid crystal esthane
6114947	Nov 2, 1987	Krause, et al. Liquid crystal compounds
6179030	Dec 1, 1987	Petrzilka, et al. Novel liquid crystal mixture
6180030	Dec 1, 1987	Petrzilka, et al. Novel liquid crystal composition and liquid crystal mixtures therewith
6213357	May 25, 1988	Eidenschink, et al. Heterocyclic derivatives
6219032	Jun 12, 1988	Yoshimura, et al. Liquid crystal device
6219114	Jun 12, 1988	Eidenschink, et al. Cyclohexane derivatives
6219115	Jun 12, 1988	Bachelder, et al. Liquid crystalline compounds
6219116	Jun 12, 1988	Bachelder, et al. Liquid crystalline compounds
6219379	Jun 13, 1988	Bachelder, et al. Liquid crystalline compounds
6220227	Jun 18, 1989	Goto, et al. Cyclohexane derivative and liquid crystal composition
6230269	Aug 11, 1989	Krause, et al. Nitro-Containing heterocyclic ethers
6230270	Aug 11, 1989	Krause, et al. Nitro-Containing heterocyclic ethers
6277547	Oct 21, 1989	Weber, et al. Liquid crystal display element
6357349	Sep 18, 1990	Chen, et al. Active matrix screen for the color display of liquid crystal devices, control system and process for producing said screen
6369888	May 21, 1991	Ueda, et al. Liquid crystal display device with a liquid crystal layer
6370093	May 21, 1991	Otsuka, et al. Liquid crystal elements with improved contrast and liquid crystal display devices
6372295	Jun 16, 1992	Weber, et al. Matrix liquid crystal display
6324924	Jun 23, 1992	Kozai, et al. Liquid crystal display device comprising a polymer film having a different refractive index in the thickness direction
6371409	Dec 15, 1992	Takatori, et al. Liquid crystal terminal part
6371410	Feb 1, 1993	Takatori, et al. Liquid crystal terminal part with ground regions between terminals
6380539	May 3, 1993	Weber, et al. Separated liquid-crystal display
6374374	Dec 20, 1993	Hoyer, et al. Nematic liquid-crystal composition
6343707	Aug 6, 1995	Hoyer, et al. Separating liquid-crystal display element
6350108	Aug 14, 1995	Shigematsu, et al. Liquid crystal composition
6380374	Jan 4, 1995	Mitsui, et al. Liquid crystal composition and liquid crystal display elements
6353414	May 5, 1995	Mitsui, et al. Liquid crystal composition and liquid crystal display elements

Word clouds for LCD island and foam island

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

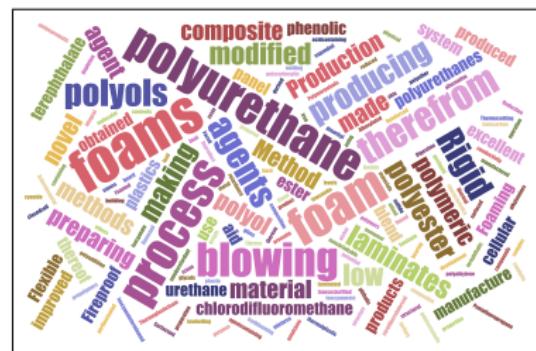
Fractional approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources



Main SPC island in SN5

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

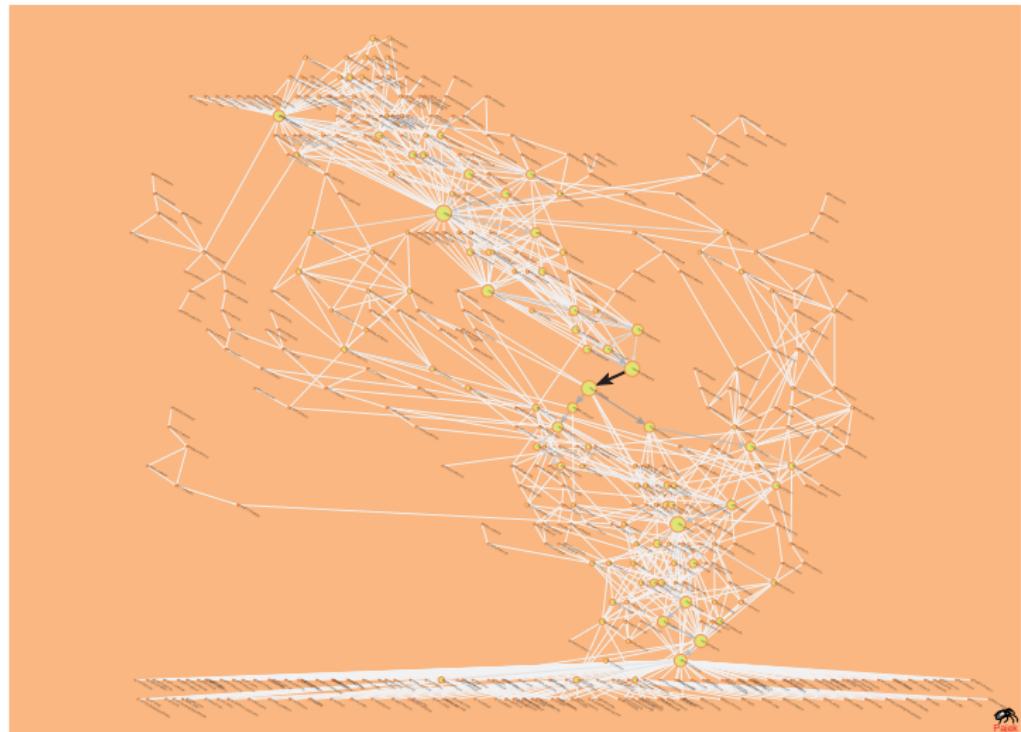
Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources



PEERE – Main path

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

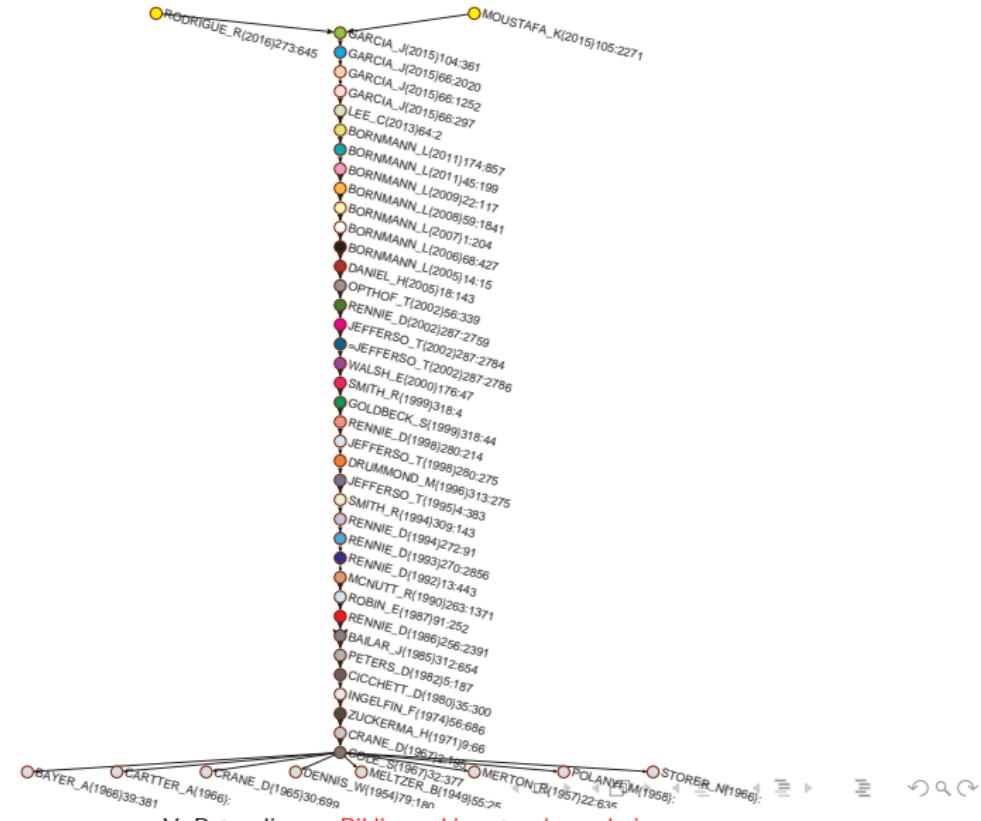
Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources



V. Batagelj

Bibliographic networks analysis

PEERE – List of publications on main path

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources

	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
Introduction	1966	Bayer AE	Some correlates of citation measure of productivity in science	SociolEduc
Statistics	1966	Storer NW	The Social System of Science	HRW
Citation	1966	Carter A	An Assessment of Quality in Graduate Education	ACE
Binary projections	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
Fractional approach	1982	Peters DP	Peer-review practices of psychological journals - the fate...	BehavBrainSci
Normalizations	1985	Bailar JC	Journal peer-review - the need for a research agenda	NewEnglJMed
Groups	1986	Rennie D	Guarding the guardians - a conference on editorial peer-review	Jama
Bibliographic coupling and co-citation	1987	Robin ED	Peer-review in medical journals	Chest
Resources	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

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources

	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	BritJPsiychiat
Introduction	2002	Jefferson T	Effects of editorial peer review - A systematic review	Jama
Statistics	2002	Rennie D	Fourth International Congress on Peer Review in Biomedical Pub...	Jama
Citation	2002	Ophof T	The significance of the peer review process against ... bias	CardiovascRes
Binary projections	2002	Jefferson T	Measuring the quality of editorial peer review	Jama
Fractional approach	2005	Bornmann L	Committee peer review at an international research foundation	ResEvaluat
Normalizations	2005	Daniel HD	Publications as a measure of scientific advancement and of...	LearnPubl
Groups	2006	Bornmann L	Selecting scientific excellence through committee peer review	Scientometrics
Bibliographic coupling and co-citation	2007	Bornmann L	Convergent validation of peer review decisions using the h index	JInformatr
Resources	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...	JRStatSocAStat
	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

PEERE – Main paths for 100 largest weights

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

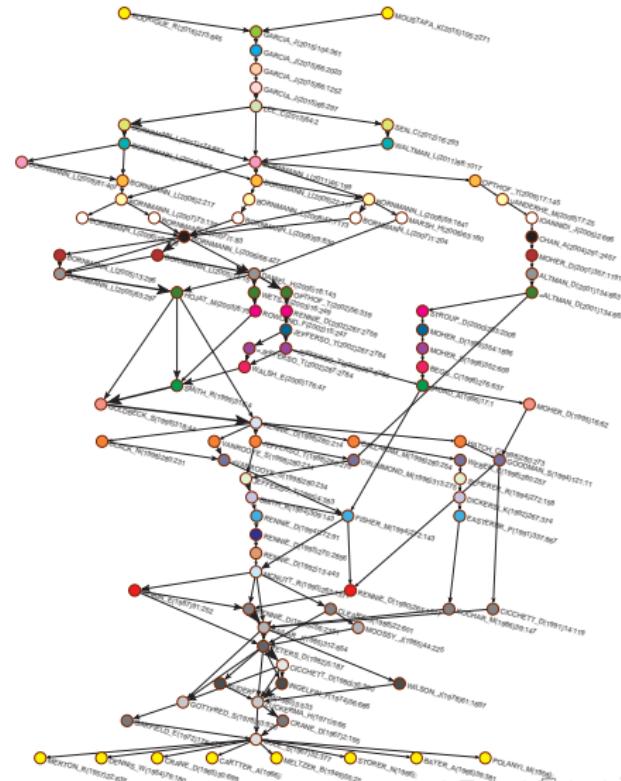
Fractional
approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources



PEERE – SPC – Link island1

$$W_{max} = 0.297$$

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

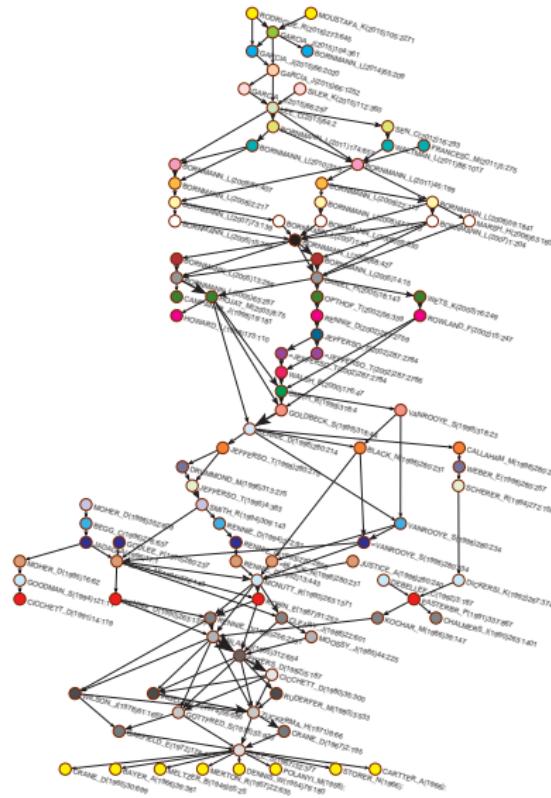
Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources



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

Matrix representation of a network

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

A simple (no parallel links) two-mode network

$\mathcal{N}_A = ((I, K), L_A, a)$ can be represented by a matrix

$\mathbf{A} = [A[i, k]]_{I \times K}$ where $A[i, k] = a(i, k)$ iff $(i, k) \in L_A$ and \square otherwise.

We extend the codomain \mathbb{R} of the weight $a : L_A \rightarrow \mathbb{R}$ with a *structural zero* \square such that

$$\square + x = x \quad \text{and} \quad \square \cdot x = \square$$

In the case $a : L_A \rightarrow \mathbb{R}^+$ we can set $\square = 0$.

Multiplication of networks

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

The product $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$ of two compatible matrices $\mathbf{A}_{I \times K}$ and $\mathbf{B}_{K \times J}$ is defined in the standard way

$$C[i, j] = \sum_{k \in K} A[i, k] \cdot B[k, j]$$

(semirings !!!)

The product of two compatible networks $\mathcal{N}_A = ((I, K), L_A, a)$ and $\mathcal{N}_B = ((K, J), L_B, b)$ is the network $\mathcal{N}_C = ((I, J), L_C, c)$ where $L_C = \{(i, j) : c[i, j] \neq \square\}$ and the weight c is determined by the matrix \mathbf{C} , $c(i, j) = C[i, j]$.

Multiplication of networks

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

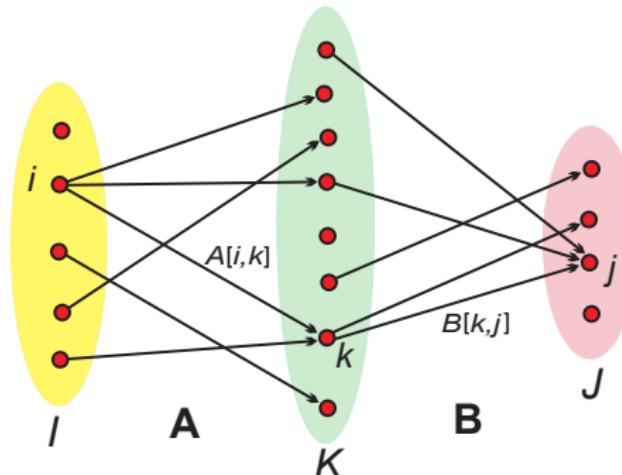
Citation

Binary projections
Fractional approach

Normalizations
Groups

Bibliographic coupling and co-citation

Resources



If all weights in networks \mathcal{N}_A and \mathcal{N}_B are equal to 1, the value of $C[i, j]$ of $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$ counts the number of ways we can go from the node $i \in I$ to the node $j \in J$ passing through K , $C[i, j] = |N_A(i) \cap N_B^-(j)|$.

$$C[i, j] = \sum_{k \in N_A(i) \cap N_B^-(j)} A[i, k] \cdot B[k, j]$$

Multiplication of networks

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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.

```
for k in K do
    for (i,j) in N_A^-(k) × N_B(k) do
        if ∃ci,j then ci,j := ci,j + ai,k · bk,j
        else new ci,j := ai,k · bk,j
```

In general, the multiplication of large sparse networks is a 'dangerous' operation since the result can 'explode' – it is not sparse.

Multiplication of networks

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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

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

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

Outer product decomposition

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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

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

then we can express the product **C** of two compatible matrices **A** and **B** as the *outer product decomposition*

$$\mathbf{C} = \mathbf{A} \cdot \mathbf{B} = \sum_k \mathbf{H}_k \quad \text{where} \quad \mathbf{H}_k = \mathbf{A}[:, k] \circ \mathbf{B}[k, :],$$

A[:, k] is the k -th column of matrix **A**, and **B**[k , :] is the k -th row of matrix **B**.

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

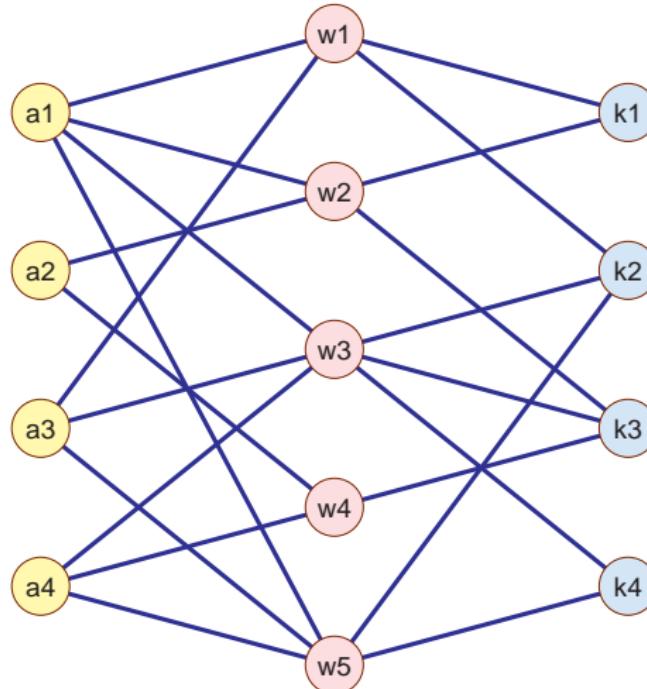
Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources



Example

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

$$\mathbf{H} = \begin{bmatrix} k_1 & k_2 & k_3 & k_4 \end{bmatrix} = \mathbf{H}_1 + \mathbf{H}_2 + \mathbf{H}_3 + \mathbf{H}_4 + \mathbf{H}_5$$

$$\begin{matrix} a_1 & 2 & 3 & 2 & 2 \\ a_2 & 1 & 0 & 2 & 0 \\ a_3 & 1 & 3 & 1 & 2 \\ a_4 & 0 & 2 & 2 & 2 \end{matrix} = \begin{matrix} a_1 & 1 & 1 & 0 & 0 \\ a_2 & 0 & 0 & 0 & 0 \\ a_3 & 1 & 1 & 0 & 0 \\ a_4 & 0 & 0 & 0 & 0 \end{matrix} + \begin{matrix} a_1 & 1 & 0 & 1 & 0 \\ a_2 & 1 & 0 & 1 & 0 \\ a_3 & 0 & 0 & 0 & 0 \\ a_4 & 0 & 0 & 0 & 0 \end{matrix} + \begin{matrix} a_1 & 0 & 1 & 0 & 1 \\ a_2 & 0 & 0 & 0 & 0 \\ a_3 & 0 & 1 & 0 & 1 \\ a_4 & 0 & 1 & 0 & 1 \end{matrix}$$

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

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

Derived networks

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

WA – works × authors – authorship network

WK – works × keywords

Ci – works × works – citation network

JD – journal × discipline

AK = $\mathbf{WA}^T * \mathbf{WK}$ – authors × keywords

Co = $\mathbf{WA}^T * \mathbf{WA}$ – coauthorship

ACi = $\mathbf{WA}^T * \mathbf{Ci} * \mathbf{WA}$ – citations between authors

DCi = $\mathbf{JD}^T * \mathbf{WJ}^T * \mathbf{Ci} * \mathbf{WJ} * \mathbf{JD}$ – citations between disciplines

Interpretation!!!

Projections

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

A standard approach to the analysis of a two-mode network \mathcal{N} is to transform it into the corresponding one-mode networks determined by:

row projection to U : $\text{row}(\mathbf{A}) = \mathbf{A} \cdot \mathbf{A}^T$, or
column projection to V : $\text{col}(\mathbf{A}) = \mathbf{A}^T \cdot \mathbf{A}$

and analyze the obtained weighted network.

$$\text{col}(\mathbf{A}) = \mathbf{A}^T \cdot \mathbf{A} = \mathbf{A}^T \cdot (\mathbf{A}^T)^T = \text{row}(\mathbf{A}^T), \quad \text{row}(\mathbf{A}) = \text{col}(\mathbf{A}^T)$$

Binary projections

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

In the following, we will consider a two-mode network $\mathcal{N} = ((W, A), L, wa)$ described with the corresponding binary matrix **WA**. To support our intuition, we can interpret it as an authorship network linking a work $w \in W$ to its authors from A .

Its column projection $\mathbf{Co} = col(\mathbf{WA}) = \mathbf{WA}^T \cdot \mathbf{WA}$ has entries

$Co[a, b] = |N(a) \cap N(b)| = Co[b, a] = \# \text{ of works that authors } a \text{ and } b \text{ coauthored}$

Co is a *co-appearance* / coauthorship matrix.

$Co[a, a] = |N(a)| = \text{indeg}_{WA}(a) = \# \text{ of works (co)authored by the author } a$.

... nodes

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

A real-life network \mathcal{N} can contain nodes $w \in W$ of degree 0 (works with no author) and 1 (single author works). Works with no author do not contribute to the matrix **Co**. Single author works contribute only to the author's diagonal entry in the matrix **Co**.

We denote $W_{[d]} = \{w \in W : \deg(w) \geq d\}$ and
 $\mathcal{N}_{[d]} = ((W_{[d]}, A), L(W_{[d]}), wa/W_{[d]})$.

Using the entries of the matrix **Co** we can express the *Salton's cosine similarity* between nodes a_i and a_j as

$$S(i, j) = \cos[a_i, a_j] = \frac{\text{Co}[a_i, a_j]}{\sqrt{\text{Co}[a_i, a_i] \cdot \text{Co}[a_j, a_j]}}$$

Properties of Salton's cosine similarity

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

The Salton's cosine has the following properties

- ① $S(i, j) \in [-1, 1]$
- ② $S(i, j) = S(j, i)$
- ③ $S(i, i) = 1$
- ④ $wa_{ki} \in \mathbb{R}_0^+ \Rightarrow S(i, j) \in [0, 1]$
- ⑤ $S(\alpha i, \beta j) = S(i, j), \quad \alpha, \beta > 0$
- ⑥ $S(\alpha i, i) = 1, \quad \alpha > 0$

Total weight

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

The ***total weight*** of links in the network $\mathcal{N} = (V, L, w)$, $w: L \rightarrow \mathbb{R}$

$$T(\mathcal{N}) = \sum_{(u,v) \in L} w(u,v)$$

On the basis of outer product decomposition, we have

$$T(\mathbf{C}) = T\left(\sum_k \mathbf{H}_k\right) = \sum_k T(\mathbf{H}_k)$$

$$T(\mathbf{H}_k) = \left(\sum_i A[i, k]\right) \cdot \left(\sum_j B[k, j]\right) = \text{windeg}_A(k) \cdot \text{woutdeg}_B(k)$$

Therefore for **Co** we have $\mathbf{H}_w = \mathbf{K}_{N(w)}$ and

$$T(\mathbf{Co}) = \sum_{w \in W} \text{outdeg}_{WA}(w)^2$$

p_S -core at level 46 in Computational Geometry co-authorship network

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

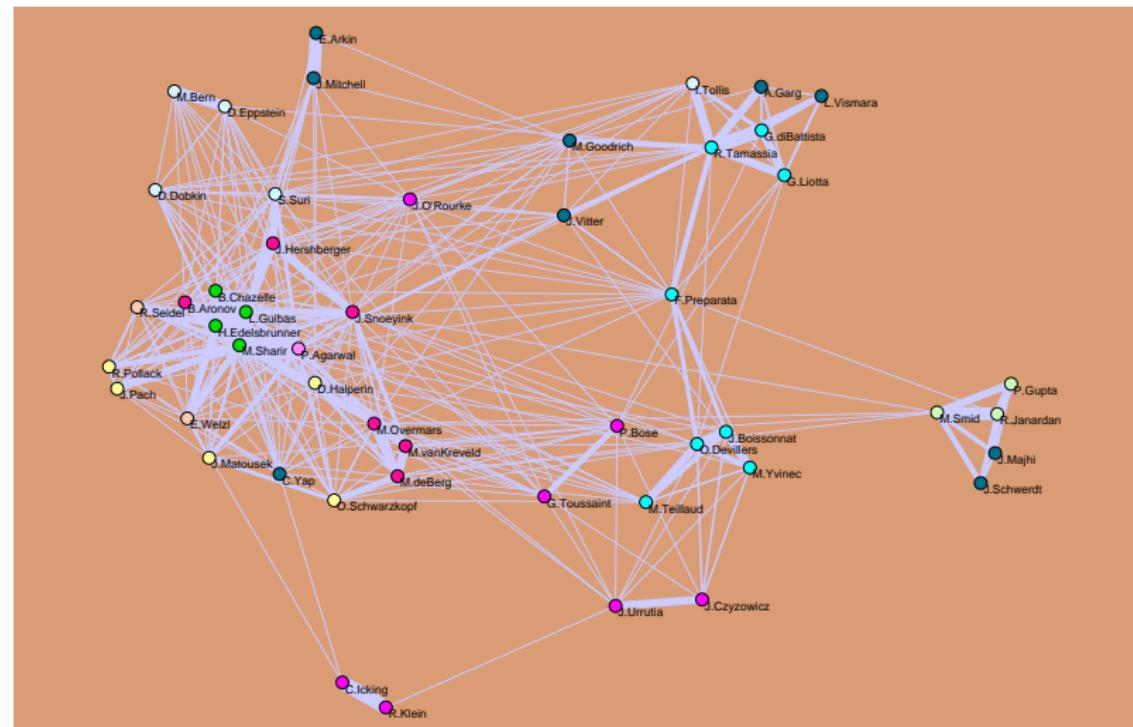
Binary projections
Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources



Structure of projection

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources

In words

- a projection network is a sum of complete subgraphs;
- the contribution of a node w from the other set W to the total is $\text{outdeg}(w)^2$.

This means that the nodes with a large degree in W are over-represented in the projection.

To make a contribution of each active node $k \in W_{[1]}$ equal to 1 we normalize matrices **A** and **B** such that for the normalized matrices $n(\mathbf{A})$ and $n(\mathbf{B})$ we have

$$\text{windeg}_{n(A)}(k) = 1 \quad \text{and} \quad \text{woutdeg}_{n(B)}(k) = 1$$

– *fractional approach*.

... fractional approach

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

We define the *normalized* matrices

$$n(\mathbf{WA}) = \text{diag}\left(\frac{1}{\deg_{WA}}\right)\mathbf{WA}, \quad n(\mathbf{WK}) = \text{diag}\left(\frac{1}{\deg_{WK}}\right)\mathbf{WK}$$

Then the *normalized product* matrix is

$$\mathbf{AKn} = n(\mathbf{WA})^T \cdot n(\mathbf{WK})$$

Denoting $\mathbf{F}_w = \frac{1}{\deg_{WA}(w) \deg_{WK}(w)} \mathbf{H}_w$ the outer product decomposition gets form

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

Example

normalized matrices

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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

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

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

... example

normalized matrices

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

$$\mathbf{F}_1 = \begin{matrix} & k_1 & k_2 & k_3 & k_4 \\ \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} & \begin{bmatrix} 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

$$\mathbf{F}_5 = \begin{matrix} & k_1 & k_2 & k_3 & k_4 \\ \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} & \begin{bmatrix} 0 & 1/6 & 0 & 1/6 \\ 0 & 0 & 0 & 0 \\ 0 & 1/6 & 0 & 1/6 \\ 0 & 1/6 & 0 & 1/6 \end{bmatrix} \end{matrix}$$

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

Fractional approach

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

The normalized 2-mode network $n(\mathbf{WA})$ has weights

$$n(\mathbf{WA})[w, a] = \frac{\mathbf{WA}[w, a]}{\max(1, \text{outdeg}_{\mathbf{WA}}(w))}$$

$$\sum_{a \in A} n(\mathbf{WA})[w, a] = 1, \quad w \in W_{[1]}$$

Normalized projection $\mathbf{Cn} = n(\mathbf{WA})^T \cdot n(\mathbf{WA})$

Network: opposite arcs replaced by an edge with double weight.

$$T(\mathbf{Cn}) = |W_{[1]}|$$

Each work (with at least one author) has a value 1 that is distributed over the links of the projection.

... fractional approach

Collaborativeness

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

Note: instead of the matrix $n(\mathbf{WA})$ we could use a matrix of authors' contributions $\mathbf{N} = [N[w, a]]$ such that

$\sum_{a \in A} N[w, a] = 1$. Unfortunately, it is very difficult to get such data.

In the network $n(\mathbf{WA})$,

- $\text{indeg}_{n(WA)}(a) = \# \text{ of works (co-)authored by the author } a$
- $\text{windeg}_{n(WA)}(a) = \text{fractional contribution of the author } a \text{ to his/her works}$

$$\text{Self-sufficiency: } S(a) = \frac{\text{windeg}_{n(WA)}(a)}{\text{indeg}_{WA}(a)}$$

$$\text{Collaborativeness: } K(a) = 1 - S(a)$$

... fractional approach

SNA18 collaborativness

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources

#	Author	Total # works	Total contribution	Collaborativeness, %	#	Author	Total # works	Total contribution	Collaborativeness, %
1	LATKIN_C	130	32.99	75	31	EVERETT_M	44	22.58	49
2	VALENTE_T	97	34.96	64	32	MORRIS_M	43	17.22	60
3	DUNBAR_R	91	40.02	56	33	CONTRACT_N	43	14.15	67
4	NEWMAN_M	81	50.02	38	34	WHITE_H	42	27.28	35
5	CHRISTAK_N	74	22.89	69	35	SKVORETZ_J	42	20.07	52
6	DOREIAN_P	72	46.19	36	36	PENTLAND_A	41	14.12	66
7	CARLEY_K	72	28.11	61	37	MOODY_J	40	17.7	56
8	BURT_R	71	55.73	22	38	SMITH_A	40	14.2	65
9	BORGATTI_S	71	29.72	58	39	MARSDEN_P	39	30.17	23
10	SNIJDERS_T	67	29.63	56	40	BERKMAN_L	39	14.3	63
11	BARABASI_A	67	27.61	59	41	SMITH_M	39	13.29	66
12	FOWLER_J	65	20.14	69	42	KRACKHAR_D	38	18.24	52
13	KAZENKO_P	64	21.97	66	43	JACKSON_M	38	17.78	53
14	ROBINS_G	64	19.67	69	44	THELWALL_M	37	18.41	50
15	WELLMAN_B	63	36.43	42	45	FRIEDKIN_N	36	28.17	22
16	FALOUTSO_C	60	17.86	70	46	SINGH_A	36	14.5	60
17	RAHMAN_M	59	19.18	67	47	WASSERMA_S	35	15.64	55
18	PATTISON_P	58	18.94	67	48	BRANDES_U	35	14.39	59
19	JOHNSON_J	54	21.19	61	49	GONZALEZ_A	35	14.13	60
20	MARTINEZ_M	53	21.9	59	50	KLEINBERG_J	34	15.05	56
21	GONZALEZ_M	52	17.76	66	51	FARINE_D	34	14.04	59
22	RODRIGUE_J	52	15.9	69	52	BATAGELJ_V	33	14.64	56
23	SCHNEIDER_J	52	13.89	73	53	BREIGER_R	31	19.73	36
24	LEYDESDO_L	51	33.28	35	54	WILLIAMS_A	31	14.5	53
25	LITWIN_H	50	32.42	35	55	SCOTT_J	28	17.54	37
26	RICE_E	48	13.09	73	56	MASUDA_N	28	14.26	49
27	BONACICH_P	46	34	26	57	FREEMAN_L	27	20.03	26
28	RODRIGUE_M	46	13.21	71	58	WATTS_D	27	13.67	49
29	NGUYEN_H	46	12.76	72	59	LAZEGA_E	26	14.17	46
30	CROFT_D	46	11.6	75	60	FAUST_K	25	13.5	46

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

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

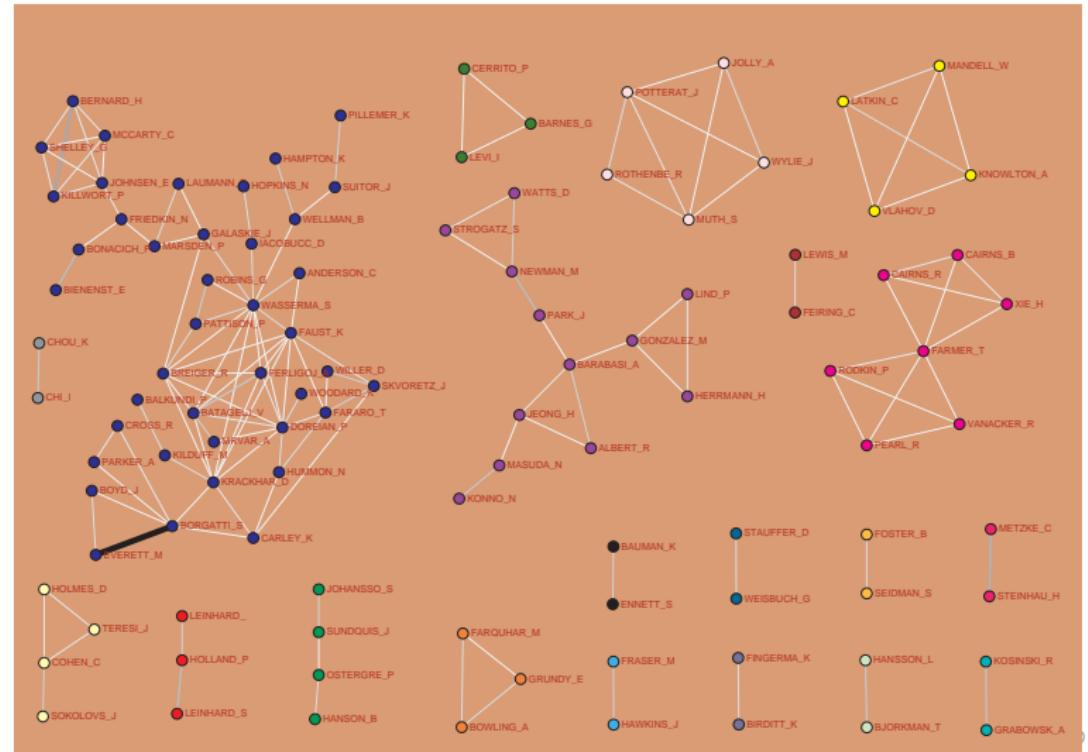
Citation

Binary projections
Fractional approach

Normalizations
Groups

Bibliographic coupling and co-citation

Resources



Newman's normalization – strict coauthorship

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

Mark Newman proposed an alternative normalization that considers only coauthorship between different authors – single-author works and self-coauthorship are excluded.

The *Newman's normalized* 2-mode network $n'(\mathbf{WA})$ has weights

$$n'(\mathbf{WA})[w, a] = \begin{cases} \frac{\mathbf{WA}[w, a]}{\deg(w) - 1} & w \in W_{[2]} \\ 0 & \text{otherwise} \end{cases}$$

Strict projection $\mathbf{Ct} = D_0(n(\mathbf{WA})^T \cdot n'(\mathbf{WA}))$.

$D_0(\mathbf{M})$ sets the diagonal of the square matrix \mathbf{M} to 0.

Network: opposite arcs replaced by an edge with double weight.

$$T(\mathbf{Ct}) = |W_{[2]}|$$

Example: SNA18[18] /selected strict islands 10-30

$$|W| = 70792, |A| = 93011$$

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

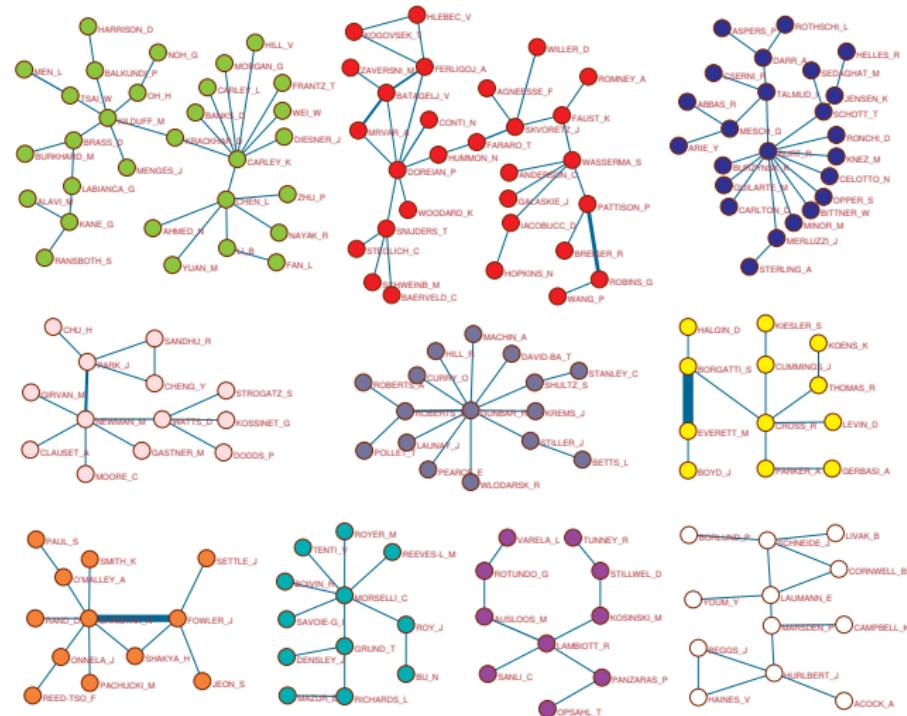
Fractional
approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources



Normalizations

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

In networks obtained from large two-mode networks, there are often huge differences in weights. Therefore it is not interesting to compare the nodes according to the raw data – the nodes with large weights will prevail. First, we have to *normalize* the network to make the weights comparable.

There exist several ways how to do this. Assume that the network is described with a square matrix $\mathbf{C}_{V \times V}$.

Balassa or activity normalization:

$$A[u, v] = \frac{C[u, v] \cdot T(C)}{\text{woutdeg}(u) \cdot \text{windeg}(v)}$$

If $A[u, v] > 1$ the measured weight is larger than expected.

$$c_B[u, v] = \log_2 A[u, v] \quad \text{for } A[u, v] > 0$$

... Normalizations

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

$$c_{\min}[u, v] = \frac{C[u, v]}{\min(C[u, u], C[v, v])} \quad c_{\max}[u, v] = \frac{C[u, v]}{\max(C[u, u], C[v, v])}$$

$$c_{\min\text{Dir}}[u, v] = \begin{cases} \frac{C[u, v]}{C[u, u]} & C[u, u] \leq C[v, v] \\ 0 & \text{otherwise} \end{cases}$$

$$c_{\max\text{Dir}}[u, v] = \begin{cases} \frac{C[u, v]}{C[v, v]} & C[u, u] \leq C[v, v] \\ 0 & \text{otherwise} \end{cases}$$

After a selected normalization the important parts of the network are obtained by link-cuts or islands approaches.

minDir of Slovenian journals 2000

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

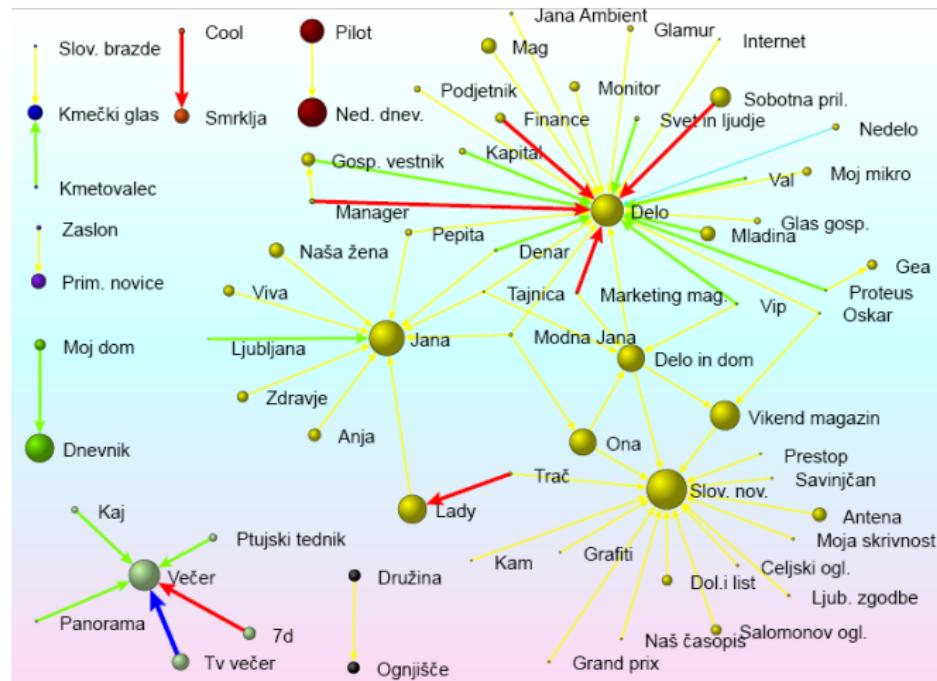
Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources



Over 100000 people were asked in the years 1999 and 2000 about the journals they read. They mentioned 124 different journals. (source Cati)

Keywords and journals for selected groups of authors

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

To uncover the research topic and disciplinary identities of a selected group of authors we constructed networks of authors and keywords

AKn = $n(\mathbf{WA})^T * n(\mathbf{WK})$ and authors and journals

AJn = $n(\mathbf{WA})^T * n(\mathbf{WJ})$. Both networks are normalized. In the network **AKn**, the weight $AKn[a, k]$ is equal to the fractional use (in publishing) of author a of keyword k . In the network **AJn**, the weight $AJn[a, j]$ is equal to the fractional use of author a of journal j . For a given keyword k and journal j , it can be extended to a group of authors $C \subseteq A$

$$\mathbf{AKn}[C, k] = \sum_{a \in C} AKn[a, k] \quad \text{and} \quad \mathbf{AJn}[C, j] = \sum_{a \in C} AJn[a, j]$$

SNA18[18] / keywords for 3 selected islands

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

Wasserman et al.		Dunbar et al.		Newman et al.	
Rank	Value	Keyword	Value	Keyword	Value
1	30.02	<i>network</i>	3.80	<i>social</i>	11.27
2	20.11	<i>social</i>	3.60	<i>network</i>	5.05
3	8.62	<i>model</i>	2.97	<i>size</i>	4.13
4	7.36	<i>analysis</i>	1.41	<i>human</i>	3.80
5	6.01	<i>graph</i>	1.31	evolution	3.54
6	5.50	<i>structure</i>	1.17	<i>brain</i>	2.68
7	3.19	<i>datum</i>	1.13	<i>group</i>	2.51
8	3.03	<i>structural</i>	0.90	primate	2.43
9	3.00	correction	0.84	support	2.31
10	2.96	<i>exchange</i>	0.79	<i>perspective</i>	2.24
11	2.80	equivalence	0.75	constraint	1.96
12	2.68	<i>random</i>	0.69	<i>neocortex</i>	1.82
13	2.54	<i>theory</i>	0.69	<i>relationship</i>	1.46
14	2.53	<i>power</i>	0.61	communication	1.39
15	2.51	<i>markov</i>	0.60	hypothesis	1.28
16	2.41	<i>evolution</i>	0.57	<i>evolutionary</i>	1.16
17	2.28	<i>group</i>	0.55	<i>behavior</i>	1.10
18	2.25	<i>statistical</i>	0.52	<i>cooperation</i>	1.03
19	2.19	<i>method</i>	0.51	cognitive	0.93
20	2.18	dynamics	0.49	<i>difference</i>	0.92
21	1.88	<i>generalized</i>	0.47	<i>kinship</i>	0.90
22	1.82	<i>journal</i>	0.44	<i>altruism</i>	0.74
23	1.80	regression	0.43	<i>bond</i>	0.70
24	1.78	<i>exponential</i>	0.40	<i>organization</i>	0.69
25	1.78	blockmodel	0.40	emotional	0.67
26	1.76	<i>logit</i>	0.38	<i>female</i>	0.67
27	1.73	<i>balance</i>	0.38	<i>closeness</i>	0.66
28	1.73	<i>p</i>	0.36	<i>analysis</i>	0.65
29	1.68	<i>measure</i>	0.36	<i>individual</i>	0.64
30	1.66	<i>algorithm</i>	0.35	baboon	0.64
31	1.66	cluster	0.35	internet	0.62
32	1.64	<i>approach</i>	0.35	dynamics	0.60
33	1.62	<i>actor</i>	0.34	<i>structure</i>	0.60
34	1.59	logistic	0.34	<i>personal</i>	0.59
35	1.55	<i>relation</i>	0.33	<i>volume</i>	0.56
36	1.54	<i>introduction</i>	0.33	<i>world</i>	0.56
37	1.54	<i>bias</i>	0.32	community	0.56
38	1.51	<i>dynamic</i>	0.32	friendship	0.55
39	1.45	blockmodeling	0.31	<i>theory</i>	0.53
40	1.44	<i>friendship</i>	0.30	<i>society</i>	0.52

SNA18[18] /journals for 3 selected islands

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources

Wasserman et al.			Dunbar et al.			Newman et al.		
Rank	Value	Journal	Value	Journal	Value	Journal	Value	Journal
1	122.60	Soc Networks	4.67	Hum Nature-Int Bios	35.02	Phys Rev E		
2	25.18	J Math Sociol	4.17	Evol Hum Behav	8.62	P Natl Acad Sci		
3	10.75	Psychometrika	3.33	Philos T R Soc B	5.25	Nature		
4	8.08	J Math Psychol	2.95	Soc Networks	4.75	Phys Rev Lett		
5	7.33	Sociol Method Res	2.75	Anim Behav	3.00	Soc Networks		
6	7.10	J Classif	2.57	P Roy Soc B-Biol Sci	2.92	Science		
7	7.08	Qual Quant	2.33	Biol Lett-Uk	2.83	Lect Notes Comput Sc		
8	5.04	Sociol Methodol	2.25	J Theor Biol	2.75	Eur Phys J B		
9	4.00	J Am Stat Assoc	2.20	Roy Soc Open Sci	2.50	Am J Sociol		
10	4.00	Contemp Sociol	2.00	Pers Indiv Differ	2.00	J Stat Mech-Theory		
11	3.90	Am Sociol Rev	2.00	Trends Cogn Sci	1.97	Plos One		
12	3.59	Lect Notes Comput Sc	2.00	Behaviour	1.67	Siam Rev		
13	3.58	Am J Sociol	1.70	Sci Rep-Uk	1.33	Reliab Eng Syst Sa		
14	3.53	Scientometrics	1.67	Comput Hum Behav	1.33	Secur Commun Netw		
15	3.40	Soc Forces	1.62	Plos One	1.25	Nat Commun		
16	3.38	Sociol Method	1.58	Hum Nature	1.12	J Comput Sci-Neth		
17	3.17	Brit J Math Stat Psy	1.57	Adapt Hum Behav Phys	1.08	Scientometrics		
18	3.00	J Am Soc Inform Sci	1.53	P Natl Acad Sci	1.03	J Supercomput		
19	2.98	Annu Rev Sociol	1.50	Brit J Psychol	1.00	Inform Process Ma		
20	2.75	Soc Psychol Quart	1.25	Soc Cogn Affect Neur	1.00	Comm Com Inf Sc		
21	2.50	Inform Process Manag	1.00	Int J Dev Disabil	1.00	Symmetry-Basel		
22	2.19	Plos One	1.00	Group Dyn-Theor Res	1.00	Electron Libr		
23	2.17	Netw Sci	1.00	Curr Dir Psychol Sci	1.00	Contemp Phys		
24	2.00	Poetics	1.00	Brit J Dev Psychol	1.00	Comput Phys Commun		
25	2.00	Sociol Rev	1.00	Soc Dev	1.00	J Consum Res		
26	2.00	Soc Sci Res	1.00	J Community Appl Soc	1.00	J Stat Phys		
27	2.00	Comput Stat Data An	1.00	Folia Primatol	1.00	Harvard Bus Rev		
28	1.92	Methods Ser	1.00	Hist Hum Sci	1.00	Annu Rev Sociol		
29	1.92	Organ Sci	1.00	Underst Complex Syst	1.00	Jpn J Polit Sci		
30	1.75	Sociol Perspect	1.00	Orig Hum Behav	1.00	Phys Lett A		

Citations among authors from the network clustering literature [13]

We consider the bibliometric data on the network clustering literature from the book [13]. We analyze the normalized network of citations among authors

$$\mathbf{nAcite} = n(\mathbf{WA})^T \cdot n(\mathbf{Cite}) \cdot n(\mathbf{WA})$$

The weight $\mathbf{nAcite}[u, v]$ is equal to the fractional share of works co-authored by u that are citing a work co-authored by v .

We removed loops (self-citations). We identified clusters such that the corresponding induced subnetworks are connected and contain a single center.

The **nAcite** weights are similarities, $s \in [\infty, 0]$. To convert them to dissimilarities d we selected $d = 1 - s/s_{\max} \in [0, 1]$ with $s_{\max} = 2.52$. On the obtained network, we applied, in Pajek, the hierarchical clustering with relational constraints procedure with the Maximum/Leader strategy and determined the partition of units into clusters of size at most 50. There are 257 such clusters.

Advances in Network Clustering and Blockmodeling

[13] / Wasserman

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

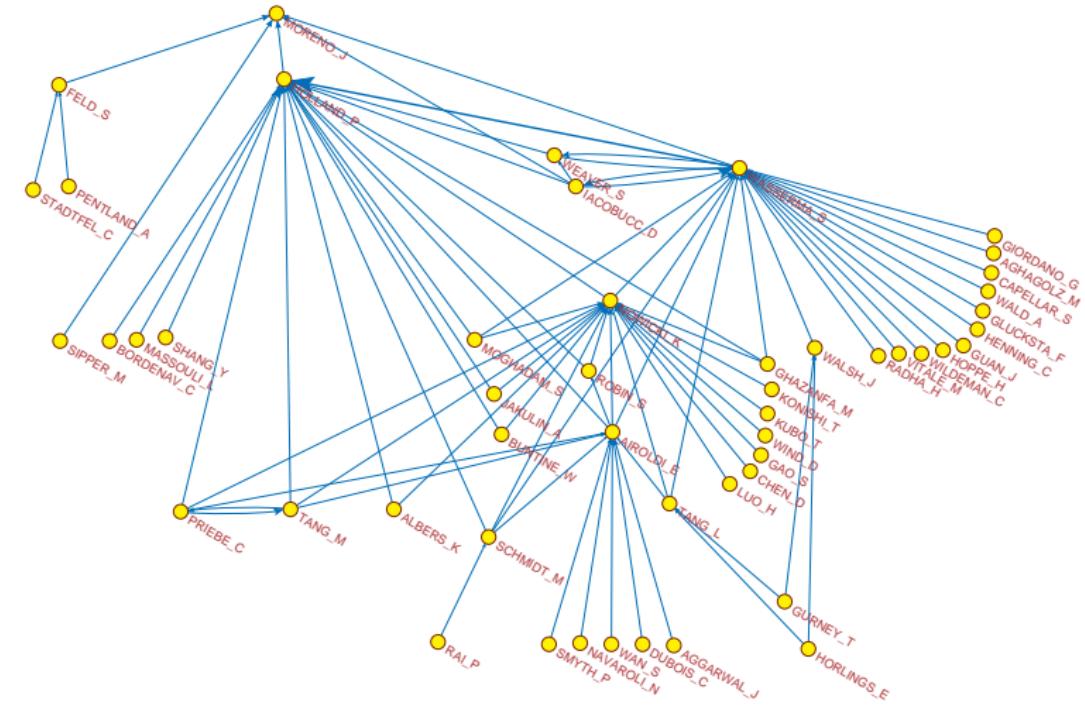
Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources



Advances in Network Clustering and Blockmodeling

[13] / Ward

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

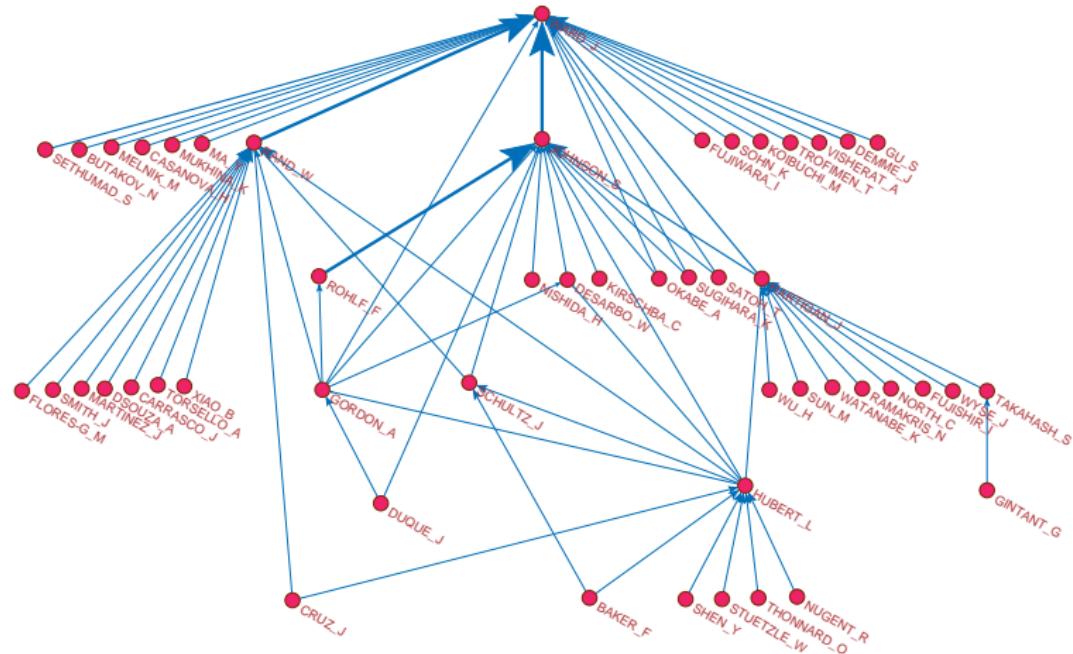
Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources



Advances in Network Clustering and Blockmodeling

[13] / Harary

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

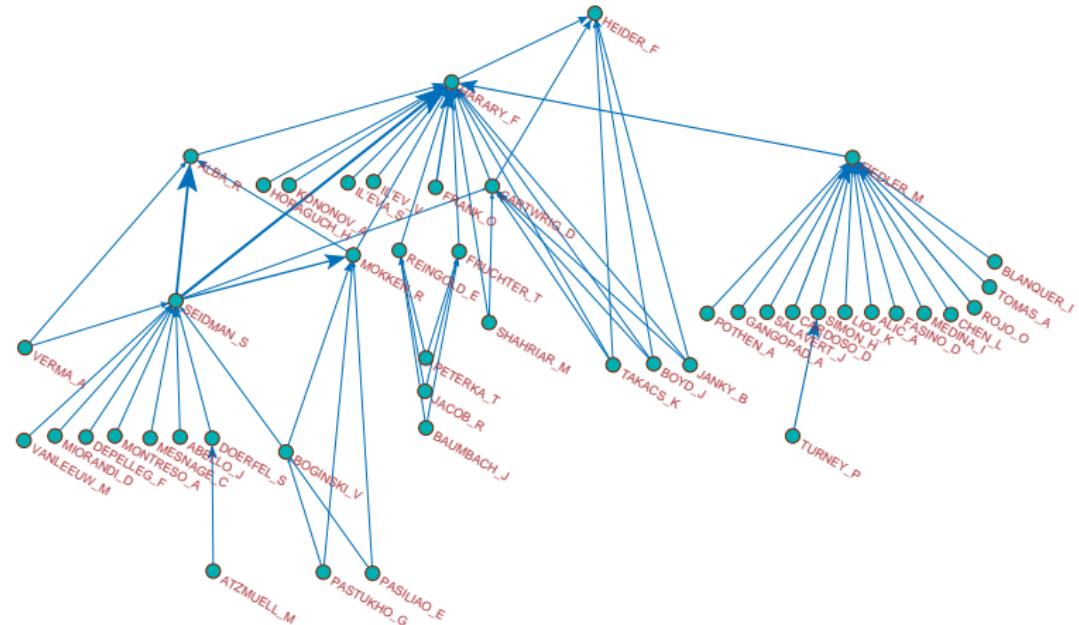
Binary projections

Fractional approach

Groups

Bibliographic
coupling and
co-citation

Resources



Advances in Network Clustering and Blockmodeling

[13] / Batagelj + Ferligoj

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

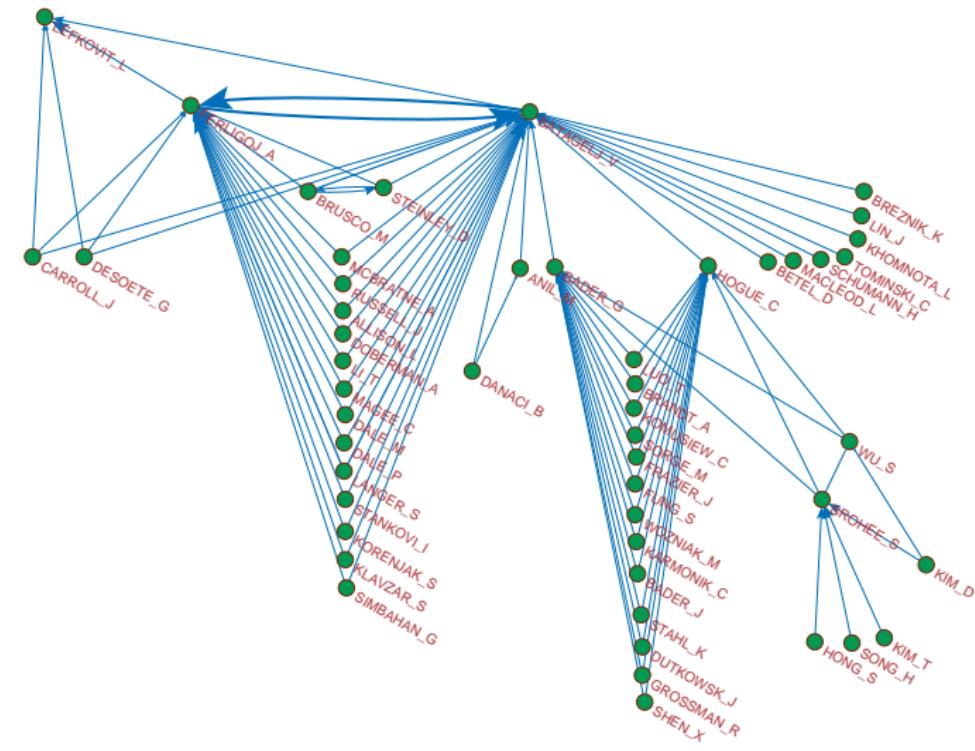
Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources



Linking through a network

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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 $W_{[1]} = W$, 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
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

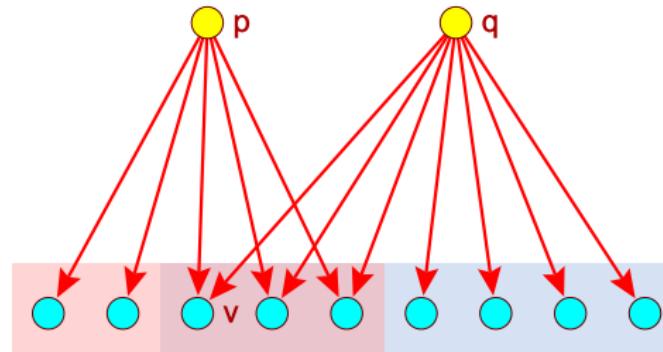
Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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



Bibliographic Coupling

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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

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

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

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

Co-citation

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

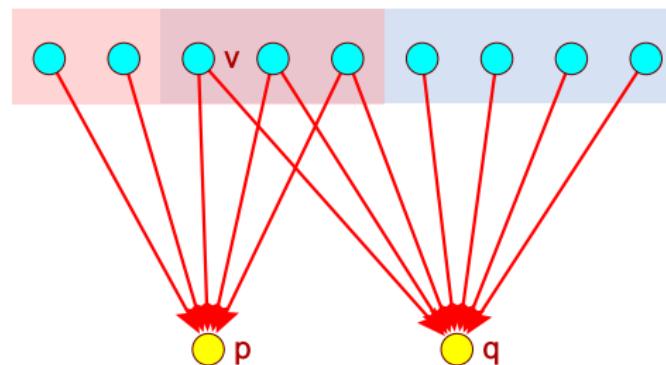
Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources

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

Normalization of Bibliographic Coupling

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources

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

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

we have

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

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

Normalization of Bibliographic Coupling

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

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

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

We have different options to construct normalized symmetric measures such as

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

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

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

Normalization of Bibliographic Coupling

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 $\mathbf{biCoX}_{pq} \in [0, 1]$ and: $\mathbf{biCoX}_{pq} = 1$ iff the works p and q are referencing the same works, $\mathbf{Ci}(p) = \mathbf{Ci}(q)$.

Normalization of Bibliographic Coupling

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources

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

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

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

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

where \oplus denotes the symmetric difference of sets.

Bibliographic Coupling, Co-citation and Linking through a network

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources

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.

Books

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

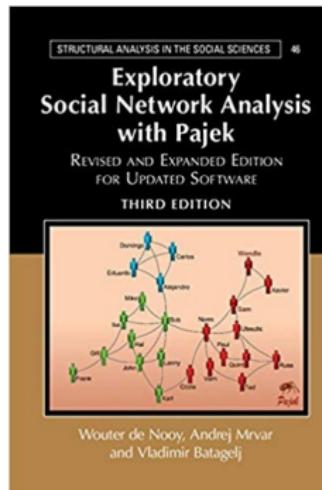
Fractional
approach

Normalizations

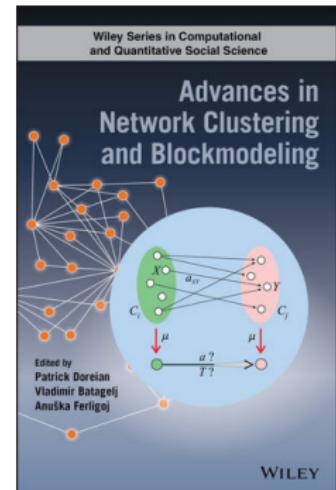
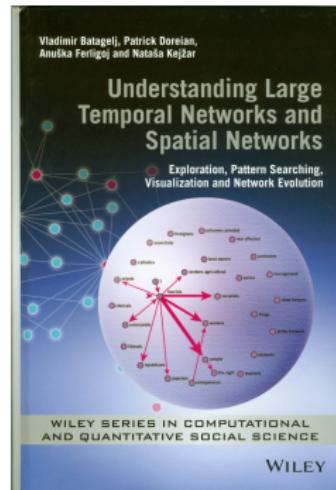
Groups

Bibliographic
coupling and
co-citation

Resources



Pajek



References |

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources

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

References II

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources

-  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

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources

-  Doreian, P., Batagelj, V., Ferligoj, A.: *Advances in Network Clustering and Blockmodeling*. Wiley, 2020.
-  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
-  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

References IV

Bibliographic networks analysis

V. Batagelj

Introduction

Statistics

Citation

Binary projections

Fractional approach

Normalizations

Groups

Bibliographic coupling and co-citation

Resources

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

References V

Bibliographic
networks
analysis

V. Batagelj

Introduction

Statistics

Citation

Binary
projections

Fractional
approach

Normalizations

Groups

Bibliographic
coupling and
co-citation

Resources



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



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