WoS networks on clustering and blockmodeling

Citations and derived networks

Vladimir Batagelj

IMFM Ljubljana and IAM UP Koper

1273. Sredin seminar

Ljubljana, 22. March 2017

Outline

Names of Participants of Group I	Code Numbers and Dates of Social Events Reported in *Old City Herald*													
	(1) 6/27	(2) 3/2	(3) 4/12	(4) 9/26	(5) 2/25	(6) 5/19	(7) 3/15	(8) 9/16	(9) 4/8	(10) 6/10	(11) 2/23	(12) 4/7	(13) 11/21	(14) 8/3
1. Mrs. Evelyn Jefferson	×	×	×	×	×	×		×	×					
2. Miss Laura Mandeville	×	×	×		×	×	×	×						
3. Miss Theresa Anderson		×	×	×	×	×	×	×	×					
4. Miss Brenda Rogers	×		×	×	×	×	×	×						
5. Miss Charlotte McDowd			×	×	×		×							
6. Miss Frances Anderson			×		×	×		×						
7. Miss Eleanor Nye					×	×	×	×						
8. Miss Pearl Oglethorpe						×		×	×					
9. Miss Ruth DeSand					×		×	×	×					
10. Miss Verne Sanderson							×	×	×			×		
11. Miss Myra Liddell								×	×	×		×		
12. Miss Katherine Rogers								×	×	×		×	×	×
13. Mrs. Sylvia Avondale							×	×	×	×		×	×	×
14. Mrs. Nora Fayette						×	×		×	×	×	×	×	×
15. Mrs. Helen Lloyd							×	×		×	×	×		
16. Mrs. Dorothy Murchison								×	×					
17. Mrs. Olivia Carleton									×		×			
18. Mrs. Flora Price									×		×			

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

Version (22. marec 2017, 16 : 51): sreda1273.pdf

WoS

To the Web of Science (WoS)Web of Science, Clarivate Analytics's multidisciplinary databases of bibliographic information, we put the query

```
"block model*" or "network cluster*" or
"graph cluster*" or "community detect*" or
"blockmodel*" or "block-model*" or
"structural equival*" or "regular equival*"
```

We limited the search to the Web of Science Core Collection because for other data bases from WoS the CR-fields (containing citation information) can not be exported. The first search was done on May 16, 2015.

WoS

We call a *terminal* node a node without a description in the collected data set – it appears only in the WoS CR field as a reference. We additionally collected on WoS and Google data for terminal nodes with large indegree in the citation network – highly cited works without description in the collected data set. If a description of a node was not available in WoS we manually constructed a corresponding description without CR data.

On January 6, 2017 we made an update for the years 2014-2017, and another update for the years 2015-2017 on February 22, 2017.

WoS record

```
PT J
AU JOHNSTON, RD
   BARTON, GW
AF JOHNSTON, RD
   BARTON, GW
TI STRUCTURAL EQUIVALENCE AND MODEL-REDUCTION
SO INTERNATIONAL JOURNAL OF CONTROL
LA English
DT Article
RP JOHNSTON, RD (reprint author), UNIV SYDNEY,DEPT CHEM ENGN,SYDNEY,NSW 2006,AUSTRALIA.
CR JOHNSTON RD, 1984, INT J CONTROL, V40, P257, DOI 10.1080/00207178408933271
   JOHNSTON RD, 1984, UNPUB COMPUT CHEM EN
   MORARI M, 1980, AICHE J, V26, P232, DOI 10.1002/aic.690260206
   Morari M., 1977, THESIS U MINNESOTA
NR 4
TC 6
Z9 6
U1 0
U2 0
PU TAYLOR & FRANCIS LTD
PI LONDON
PA ONE GUNDPOWDER SQUARE, LONDON, ENGLAND EC4A 3DE
SN 0020-7179
J9 INT J CONTROL
JI Int. J. Control
PY 1985
VL 41
IS 6
BP 1477
EP 1491
DI 10.1080/0020718508961210
PG 15
WC Automation & Control Systems
SC Automation & Control Systems
GA AQJ42
UT WOS:A1985AQJ4200007
ER
```

Networks

We applied the new WoS2Pajek 1.5 to the collected data.

	2015/05/16	2017/01/06	2017/02/23
number of works	75249	112114	117082
number of authors	44787	60419	62143
number of journals	8993	12271	12652
number of keywords	10095	12715	10269
number of records	2944	5472	6953
number of duplicates	1	62	1255

The following networks were constructed: the authorship network *WA* on works $\times$ authors (from the field AU), the journalship network *WJ* on works $\times$ journals (from the field CR or J9), the keywordship network *WK* on works $\times$ keywords (from the field ID or DE or TI), and the citation network *Cite* on works (from the field CR).

Networks

We obtained also the following partitions: partition *year* of works by publication year, the *DC* partition distinguishing between works with complete description (DC=1) and the cited only works (DC=0); and the vector of number of pages *NP*.

The sizes of the sets are as follows: works $|W| = 117082$, works with complete description $|C| = 5698$, authors $|A| = 62143$, journals $|J| = 12652$, keywords $|K| = 10269$.

An important property of all these networks is that they share as the first node set the same set – the set of works (papers, reports, books, etc.) W.

Networks

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

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

has the following structure

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

All its elements are in upper case.
In WoS the same work can have different ISI names. To improve the precission the program `WoS2Pajek` supports also *short names* (similar to the names used in HISTCITE output). They have the format:

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

For example: `LEFKOVIT_L(1985)70:585`
From the last names with prefixes `VAN`, `DE`, ... the space is deleted.
Unusual names start with character * or $.

Networks
equivalent works reduction

There are two possibilities how to correct the data:

- to make corrections in the local copy of original data (WoS file);
- to make the equivalence partition of nodes and shrink the set of works accordingly in all obtained networks.

We used the second option. For the works with large counts (≥ 30) we prepared lists of possible equivalents and manually determined equivalence classes. With a simple program in Python we produced a Pajek's partition file `worksEQ.clu` used in Pajek for shrinking the set of works. notes

Using the partition $p = \mathit{worksEQ}$, $p : V \to C$, we shrink using Pajek the citation network *cite* to *citeR*. As a byproduct we get also a partition $q : V_C \to V$, such that $q(v) = u \Rightarrow p(u) = v$.

We have to shrink also partitions *year*, *DC* and the vector *NP*.

Networks
equivalent works reduction

Let us describe how this can be done in Pajek. In general, given a mapping $s : V \to B$, we ask for a mapping $r : V_C \to B$ such that if $q(v) = u$ then $s(u) = r(v)$. Therefore, $r(v) = s(u) = s(q(v)) = q * s(v)$ or equivalently $r = q * s$. In Pajek, given a mapping $q : V_C \to V$, the mapping r can be determined as follows:

```
select partition q as First partition
select partition s as Second partition
Partitions/Functional Composition First*Second
```

or

```
select partition q as First partition
select vector s as First vector
Operations/Vector+Partition/
   Functional Composition Partition*Vector
```

For the partition $q = worksEQq$ we computed networks CiteR, WAr, WKr, WJr and partitions YearR and DCr and the vector NPr.

The most cited works

indegree in CiteR

i	freq	id	i	freq	id
1	1096	GIRVAN_M(2002)99:7821	31	211	HOLLAND_P(1983)5:109
2	969	FORTUNAT_S(2010)486:75	32	206	WHITE_H(1976)81:730
3	712	CLAUSET_A(2004)70:066111	33	199	AHN_Y(2010)466:761
4	638	BLONDEL_V(2008):P10008	34	168	KERNIGHA_B(1970)49:291
5	621	NEWMAN_M(2004)69:026113	35	163	AIROLDI_E(2008)9:1981
6	578	NEWMAN_M(2006)103:8577	36	161	NEWMAN_M(2010):
7	553	ZACHARY_W(1977)33:452	37	157	SCHAEFFE_S(2007)1:27
8	544	PALLA_G(2005)435:814	38	155	GOOD_B(2010)81:046106
9	489	FORTUNAT_S(2007)104:36	39	150	KARRER_B(2011)83:016107
10	416	WATTS_D(1998)393:440	40	150	LANCICHI_A(2009)80:016118
11	412	DANON_L(2005):	41	145	BURRIDGE_R(1967)57:341
12	380	NEWMAN_M(2004)38:321	42	145	LANCICHI_A(2011)6:0018961
13	369	LANCICHI_A(2008)78:046110	43	139	GREGORY_S(2010)12:103018
14	351	WASSERMA_S(1994):	44	139	LESKOVEC_J(2010):
15	329	NEWMAN_M(2006)74:036104	45	138	BOCCALET_S(2006)424:175
16	326	ROSVALL_M(2008)105:1118	46	137	GUIMERA_R(2004)70:025101
17	319	RAGHAVAN_U(2007)76:036106	47	129	NEWMAN_M(2004)70:056131
18	307	LANCICHI_A(2009)11:033015	48	127	BRANDES_U(2008)20:172
19	306	RADICCHI_F(2004)101:2658	49	126	BREIGER_R(1975)12:328
20	304	BARABASI_A(1999)286:509	50	126	NOWICKI_K(2001)96:1077
21	292	NEWMAN_M(2003)45:167	51	125	ROSVALL_M(2007)104:7327
22	292	LANCICHI_A(2009)80:056117	52	124	VONLUXBU_U(2007)17:395
23	286	NEWMAN_M(2004)69:1	53	122	NEWMAN_M(2001)64:026118
24	259	GUIMERA_R(2005)433:895	54	119	REICHARD_J(2004)93:218701
25	251	ALBERT_R(2002)74:47	55	118	ARENAS_A(2008)10:053039
26	244	DUCH_J(2005)72:027104	56	118	ERDOS_P(1959)6:290
27	236	LUSSEAU_D(2003)54:396	57	116	FREEMAN_L(1979)1:215
28	216	SHI_J(2000)22:888	58	116	FREEMAN_L(1977)40:35
29	216	LORRAIN_F(1971)1:49	59	113	NEWMAN_M(2001)98:404
30	215	REICHARD_J(2006)74:016110	60	112	SHEN_H(2009)388:1706

The most citing works

outdegree in CiteR

i	refs	id	i	refs	id
1	1095	PRUESSNE_G(2012):1	6	417	NEWMAN_M(2003)45:167
2	863	BOCCALET_S(2006)424:175	7	398	FORTUNAT_S(2010)486:75
3	839	FOUSS_F(2016):1	8	327	HOLME_P(2015)88:e2015-60657-4
4	476	ARABIE_P(1992)43:169	9	321	SIBLEY_C(2012)12:505
5	456	TURCOTTE_D(1999)62:1377	10	310	FRANK_K(1998)23:171

- Gunnar Pruessner: Self-organised criticality: theory, models and characterisation. Thesis. Cambridge University Press, 2012.
- S Boccaletti, V Latora, Y Moreno, M Chavez, DU Hwang: Complex networks: Structure and dynamics. Physics reports 424(2006)4, 175-308.
- F Fouss, M Saerens, M Shimbo: Algorithms and models for network data and link analysis. Cambridge University Press, 2016.
- P Arabie, and L J Hubert: Combinatorial Data Analysis. Annual Review of Psychology, 43(1992): 169-203.

Authors – basic statistics

```
WAc   n = 19071 = 5695+13376  m = 21562  AveDegree = 2.26123433
WKc   n = 15964 = 5695+10269  m = 88953  AveDegree = 11.14419945
WJc   n =  7451 = 5695+1756   m =  5815  AveDegree = 1.56086431
CiteC n =  5695     m = 38400            AveDegree = 13.48551361
```

On the next slide a list of authors with the largest number of papers on our topic is presented.

The large number of Chinese authors in the list is probably a "three Zhang, four Li" effect. It is out of our resources to drill into this. We can only make a warning.

Authors with the largest number of papers

indegree in WAc

i	freq	author	i	freq	author
1	66	ZHANG_X	21	31	CHEN_H
2	57	WANG_Y	22	29	YANG_J
3	56	LIU_J	23	28	HANCOCK_E
4	51	WANG_X	24	28	WANG_W
5	44	LI_J	25	27	CHEN_L
6	42	WANG_H	26	26	LI_H
7	41	LIU_Y	27	26	WU_J
8	41	LI_Y	28	26	ZHANG_H
9	40	NEWMAN_M	29	26	WANG_L
10	39	WANG_J	30	26	TURCOTTE_D
11	39	DOREIAN_P	31	26	BORGATTI_S
12	38	CHEN_Y	32	26	EVERETT_M
13	36	ZHANG_Y	33	26	WANG_C
14	35	WANG_Z	34	24	LI_X
15	35	ZHANG_Z	35	24	LI_L
16	35	ZHANG_J	36	24	LIU_X
17	34	JIAO_L	37	23	LI_S
18	33	ZHANG_S	38	23	ZHOU_Y
19	32	WANG_S	39	23	CHEN_X
20	31	BATAGELJ_V	40	23	LEE_J

The most used journals

indegree in WJr and WJc

i	freq	id	freq	id
1	1058	P NATL ACAD SCI USA	223	LECT NOTES COMPUT SC
2	1014	NATURE	175	PHYS REV E
3	941	LECT NOTES COMPUT SC	151	PHYSICA A
4	908	SCIENCE	122	SOC NETWORKS
5	667	PHYSICA A	88	PLOS ONE
6	639	PHYS REV E	56	LECT NOTES ARTIF INT
7	616	PHYS REV LETT	56	J GEOPHYS RES-SOL EA
8	549	BIOINFORMATICS	45	P NATL ACAD SCI USA
9	548	NUCLEIC ACIDS RES	40	SCI REP-UK
10	522	SOC NETWORKS	39	J STAT MECH-THEORY E
11	519	J GEOPHYS RES-SOL EA	33	NEUROCOMPUTING
12	428	B SEISMOL SOC AM	30	PHYS REV LETT
13	400	TECTONOPHYSICS	28	COMM COM INF SC
14	398	GEOPHYS J INT	27	APPL MECH MATER
15	348	NEUROIMAGE	27	BMC BIOINFORMATICS
16	342	J GEOPHYS RES	27	EUR PHYS J B
17	342	J BIOL CHEM	27	GEOPHYS J INT
18	336	J MOL BIOL	25	PROCEDIA COMPUT SCI
19	330	PHYS REV B	25	BIOINFORMATICS
20	321	IEEE T PATTERN ANAL	24	INFORM SCIENCES
21	285	AM J SOCIOL	23	IEEE DATA MINING
22	274	PATTERN RECOGN	23	KNOWL-BASED SYST
23	272	AM SOCIOL REV	23	J MATH SOCIOL
24	260	GEOPHYS RES LETT	21	SOC NETW ANAL MIN
25	249	GEOLOGY	21	ADV INTELL SYST
26	239	SCIENTOMETRICS	20	MATH PROBL ENG
27	229	LECT NOTES ARTIF INT	20	EXPERT SYST APPL
28	224	EARTH PLANET SC LETT	19	EPL-EUROPHYS LETT
29	220	BIOCHEMISTRY-US	19	INT J MOD PHYS B
30	214	APPL ENVIRON MICROB	19	TECTONOPHYSICS
31	212	J CHEM PHYS	19	ANN STAT
32	207	J NEUROSCI	19	NATURE
33	207	J AM STAT ASSOC	18	IEEE T KNOWL DATA EN
34	205	J GEOPHYS RES-SOLID	18	PATTERN RECOGN LETT
35	201	J AM CHEM SOC	18	AM J SOCIOL

The most used keywords

indegree in WKc

i	freq	id	i	freq	id
1	3204	network	36	227	clustering
2	2064	community	37	220	theory
3	1533	detection	38	213	large
4	1499	model	39	209	self
5	1177	graph	40	205	matrix
6	1135	cluster	41	204	dynamic
7	1104	algorithm	42	204	identification
8	1082	complex	43	197	modeling
9	1080	social	44	197	pattern
10	932	structure	45	195	detect
11	900	analysis	46	194	local
12	880	base	47	190	world
13	727	block	48	186	similarity
14	494	use	49	184	multi
15	430	datum	50	181	evolution
16	407	modularity	51	176	mining
17	398	method	52	166	functional
18	373	dynamics	53	165	behavior
19	357	structural	54	164	simulation
20	317	approach	55	163	state
21	300	blockmodel	56	163	gene
22	294	information	57	160	genetic
23	293	optimization	58	159	centrality
24	293	random	59	157	flow
25	291	earthquake	60	156	classification
26	281	protein	61	155	partition
27	276	stochastic	62	155	hierarchical
28	270	overlap	63	150	application
29	268	fault	64	148	slip
30	265	equivalence	65	146	small
31	241	prediction	66	146	design
32	240	organization	67	146	link
33	237	interaction	68	145	web
34	236	scale	69	144	organize
35	229	time	70	143	spectral

Boundary problem in citation network

The network CiteR has 116906 nodes and 195784 arcs.

indegree	freq
0	4070
1	93246
2	10694
3	3352
4	1610

Most of nodes are terminal ($DCr = 0$) nodes cited only once (indegree=1). We decided (boundary problem) to include in our networks nodes with $DCr > 0$ or indeg > 2 (partition boundary). They determine a subnetwork CiteB with 13540 nodes and 82238 arcs.

Components

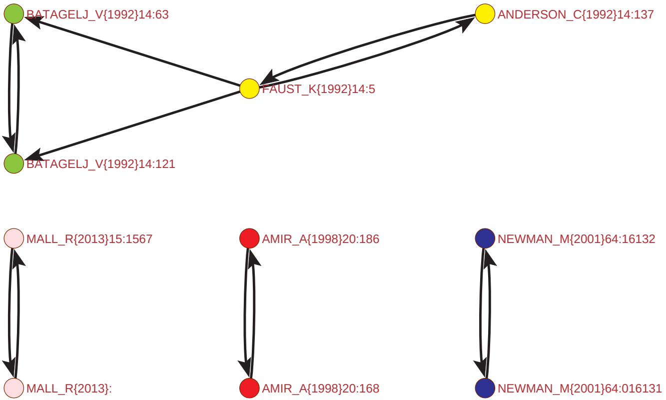
The network CiteB has 690 connected components with sizes 12702, 21, 20, 19, 17, 10, 9, ... We limit our analysis to the largest component CiteMain. It has 5 strong components all of size 2 (see the next slide).

To get an acyclic network we applied the *preprint transformation* to CiteMain. The resulting network CiteMacy has 12712 nodes and 81972 arcs. We computed the SPC weights on its arcs. The total flow is equal to $1.625\ 10^{20}$.

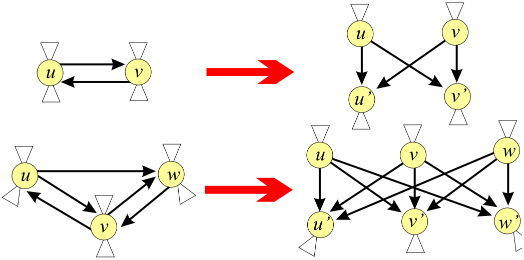
First we determine the CPM path in this network, followed by the key-routes approach. Afterward we determine SPC link islands of sizes [20, 200]. There are 10 islands. From the picture we see that only island 10, 7, 9 and 2 have "interesting" structure.

In Slides 25–29 some details about works from CPM path, main paths and island 10 are presented to be used in their interpretation.

Strong components



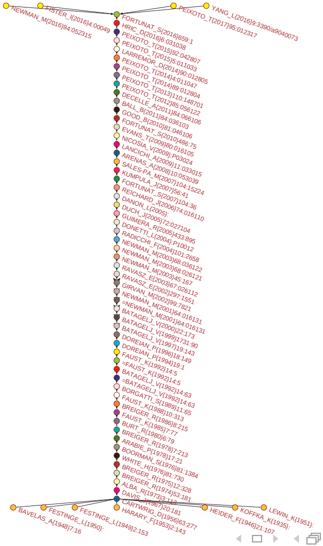
Preprint transformation



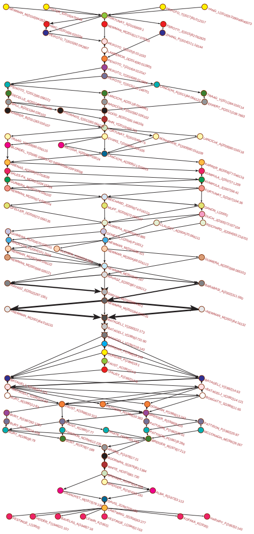
To transform the network into an acyclic network using the preprint transformation we applied the `Pajek`'s command

`Network/Acyclic Network/Transform/Preprint Transformation`

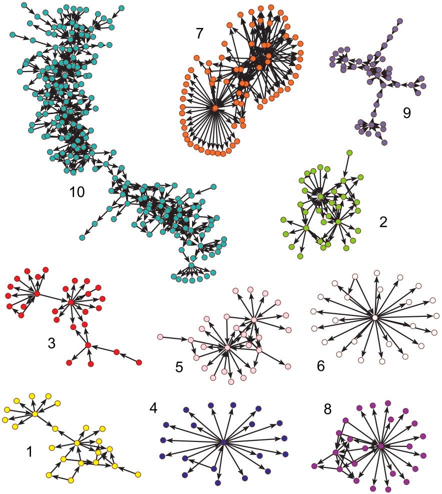
CPM path


NEWMAN_M{2016}94:052315
FISTER_I{2016}4:00049
PEIXOTO_T{2017}95:012317
YANG_L{2016}9:3390/a9040073
FORTUNAT_S{2016}659:1
HRIC_D{2016}6:031038
PEIXOTO_T{2015}92:042807
PEIXOTO_T{2015}5:011033
LARREMOR_D{2014}90:012805
PEIXOTO_T{2014}4:011047
PEIXOTO_T{2014}89:012804
PEIXOTO_T{2013}110:148701
PEIXOTO_T{2012}85:056122
DECELLE_A{2011}84:066106
BALL_B{2011}84:036103
GOOD_B{2010}81:046106
FORTUNAT_S{2010}486:75
EVANS_T{2009}80:016105
NICOSIA_V{2009}:P03024
LANCICHI_A{2009}11:033015
ARENAS_A{2008}10:053039
SALES-PA_M{2007}104:15224
KUMPULA_J{2007}56:41
FORTUNAT_S{2007}104:36
REICHARD_J{2006}74:016110
DANON_L{2005}:
DUCH_J{2005}72:027104
GUIMERA_R{2005}433:895
DONETTI_L{2004}:P10012
RADICCHI_F{2004}101:2658
NEWMAN_M{2003}68:036122
NEWMAN_M{2003}68:026121
NEWMAN_M{2003}45:167
RAVASZ_E{2003}67:026112
RAVASZ_E{2002}297:1551
GIRVAN_M{2002}99:7821
NEWMAN_M{2001}64:016131
=NEWMAN_M{2001}64:016131
BATAGELJ_V{2000}22:173
BATAGELJ_V{1999}1731:90
BATAGELJ_V{1997}19:143
DOREIAN_P{1996}18:149
DOREIAN_P{1994}19:1
FAUST_K{1992}14:5
=FAUST_K{1992}14:5
BATAGELJ_V{1992}14:63
=BATAGELJ_V{1992}14:63
BORGATTI_S{1989}11:65
FAUST_K{1988}10:313
BREIGER_R{1986}8:215
FAUST_K{1985}7:77
BURT_R{1980}6:79
BREIGER_R{1978}7:213
ARABIE_P{1978}17:21
BOORMAN_S{1976}81:1384
WHITE_H{1976}81:730
BREIGER_R{1975}12:328
BREIGER_R{1974}53:181
ALBA_R{1973}3:113
DAVIS_J{1967}20:181
CARTWRIG_D{1956}63:277
HARARY_F{1953}2:143
BAVELAS_A{1948}7:16
FESTINGE_L{1950}:
FESTINGE_L{1949}2:153
HEIDER_F{1946}21:107
KOFFKA_K{1935}:
LEWIN_K{1951}:

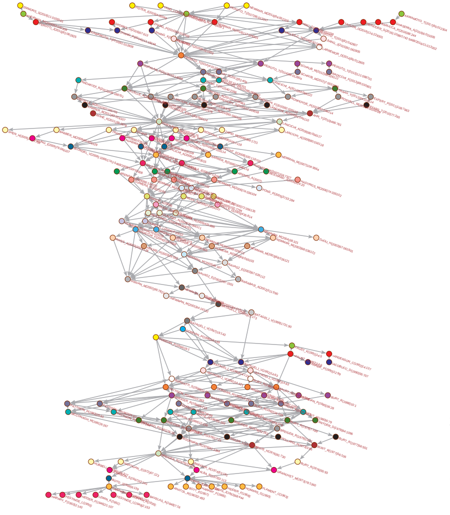
Key-route paths



Link islands [20 200]



Island 10



The island 10 on 200 nodes is unreadable. We reduced its size to 150 nodes. The maximal weight is 0.5785. We get essentially the CPM path put in the context.

List of works on CPM path(1), main paths (2) and island (3) part 1

label	code	first author	title	journal
KOFFKA_K(1935):	12	Koffka, K	Principles of Gestalt Psychology	book
HEIDER_F(1946)21:107	12	Heider, F	Attitudes and cognitive organization	J PSYCHOL
BAVELAS_A(1948)7:16	12	Bavelas, A	A mathematical model for group structure	HUMAN ORG
FESTINGE_L(1949)2:153	12	Festinger, L	The analysis of sociograms using matrix algebra	HUMAN REL
FESTINGE_L(1950):	12	Festinger, L	Informal social communication	PSYCHO REV
LEWIN_K(1951):	12	Lewin, K	Field theory in social science	book
HARARY_F(1953)2:143	12	Harary, F	On the notion of balance of a signed graph	MICH MATH J
CARTWRIG_D(1956)63:277	123	Cartwright, D	Structural balance - a generalization of heider theory	PSYCHOL REV
HUBBELL_C(1965)28:377	3	Hubbell, CH	An input-output approach to clique identification	SOCIOMETRY
DAVIS_J(1967)20:181	123	Davis, JA	Clustering and structural balance in graphs	HUM RELAT
BOYD_J(1969)6:139	3	Boyd, JP	Algebra of group kinship	J MATH PSYCHOL
HARTIGAN_J(1972)67:123	3	Hartigan, JA	Direct clustering of a data matrix	J AM STAT ASSOC
ALBA_R(1973)3:113	123	Alba, RD	Graph-theoretic definition of a sociometric clique	J MATH SOCIOL
GRANOVET_M(1973)78:1360	23	Granovet.MS	The strength of weak ties	AM J SOCIOL
BREIGER_R(1974)53:181	123	Breiger, RL	Duality of persons and groups	SOC FORCES
BREIGER_R(1975)12:328	123	Breiger, RL	Algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional-scaling	J MATH PSYCHOL
WHITE_H(1976)81:730	123	White, HC	Social-structure from multiple networks .1. Blockmodels of roles and positions	AM J SOCIOL
BOORMAN_S(1976)81:1384	123	Boorman, SA	Social-structure from multiple networks .2. Role structures	AM J SOCIOL
BURT_R(1976)55:93	3	Burt, RS	Positions in networks	SOC FORCES
BURT_R(1977)56:106	3	Burt, RS	Positions in multiple network systems .1. General conception of stratification and prestige in a system of actors cast as a social topology	SOC FORCES
BURT_R(1977)56:551	3	Burt, RS	Positions in multiple network systems .2. Stratification and prestige among elite decision-makers in community of Altneustadt	SOC FORCES
ARABIE_P(1978)17:21	123	Arabie, P	Constructing blockmodels - how and why	J MATH PSYCHOL
SAILER_L(1978)1:73	3	Sailer, LD	Structural equivalence - meaning and definition, computation and application	SOC NETWORKS
BURT_R(1978)7:189	23	Burt, RS	Cohesion versus structural equivalence as a basis for network subgroups	SOCIOL METHOD RES
BREIGER_R(1978)7:213	123	Breiger, RL	Joint role structure of 2 communities elites	SOCIOL METHOD RES
SNYDER_D(1979)84:1096	3	Snyder, D	Structural position in the world system and economic-growth, 1955-1970 - multiple-network analysis of transnational interactions	AM J SOCIOL
BREIGER_R(1979)13:21	3	Breiger, RL	Toward an operational theory of community elite structures	QUAL QUANT
BREIGER_R(1979)42:262	3	Breiger, RL	Personae and social roles - network structure of personality-types in small-groups	SOC PSYCHOL
BURT_R(1980)6:79	123	Burt, RS	Models of network structure	ANNU REV SOCIOL
MCCONAGH_M(1981)9:267	23	Mcconaghy, MJ	The common role structure - improved block-modeling methods applied to 2 communities elites	SOCIOL METHOD RES
PATTISON_P(1981)9:286	23	Pattison, PE	A reply to Mcconaghy - equating the joint reduction with block-model common role structures	SOCIOL METHOD RES
BURT_R(1982)16:109	23	Burt, RS	Testing a structural model of perception - conformity and deviance with respect to journal norms in elite sociological methodology	QUAL QUANT

List of works on CPM path(1), main paths (2) and island (3) part 2

label	code	first author	title	journal
PATTISON_P(1982)25:51	23	Pattison, PE	A factorization procedure for finite-algebras	J MATH PSYCHOL
PATTISON_P(1982)25:87	23	Pattison, PE	The analysis of semigroups of multirelational systems	J MATH PSYCHOL
MANDEL_M(1983)48:376	3	Mandel, MJ	Local roles and social networks	AM SOCIOL REV
WHITE_D(1983)5:193	23	White, DR	Graph and semigroup homomorphisms on networks of relations	SOC NETWORKS
FRIEDKIN_N(1984)12:235	23	Friedkin, NE	Structural cohesion and equivalence explanations of social homogeneity	SOCIOL METHOD RES
DOREIAN_P(1985)36:28	3	Doreian, P	Structural equivalence in a journal network	J AM SOC INFORM SCI
FAUST_K(1985)7:77	123	Faust, K	Does structure find structure - a critique of Burt use of distance as a measure of structural equivalence	SOC NETWORKS
FIENBERG_S(1985)80:51	3	Fienberg, SE	Statistical-analysis of multiple sociometric relations	J AM STAT ASSOC
BREIGER_R(1986)8:215	123	Breiger, RL	Cumulated social roles - the duality of persons and their algebras	SOC NETWORKS
BURT_R(1987)92:1287	23	Burt, RS	Social contagion and innovation - cohesion versus structural equivalence	AM J SOCIOL
FAUST_K(1988)10:313	123	Faust, K	Comparison of methods for positional analysis - structural and general equivalences	SOC NETWORKS
DOREIAN_P(1988)13:243	23	Doreian, P	Equivalence in a social network	J MATH SOCIOL
PATTISON_P(1988)10:383	23	Pattison, PE	Network models - some comments on papers in this special issue	SOC NETWORKS
WINSHIP_C(1988)10:209	3	Winship, C	Thoughts about roles and relations - an old document revisited	SOC NETWORKS
BORGATTI_S(1989)11:65	123	Borgatti, SP	The class of all regular equivalences - algebraic structure and computation	SOC NETWORKS
IACOBUCC_D(1990)55:707	3	Iacobucci, D	Social networks with 2 sets of actors	PSYCHOMETRIKA
BURT_R(1990)12:83	23	Burt, Rs	Detecting role equivalence	SOC NETWORKS
BATAGELJ_V(1992)14:63	123	Batagelj, V	Direct and indirect methods for structural equivalence	SOC NETWORKS
BATAGELJ_V(1992)14:121	23	Batagelj, V	An optimizational approach to regular equivalence	SOC NETWORKS
ANDERSON_C(1992)14:137	3	Anderson, CJ	Building stochastic blockmodels	SOC NETWORKS
FAUST_K(1992)14:5	123	Faust, K	Blockmodels - interpretation and evaluation	SOC NETWORKS
DOREIAN_P(1994)19:1	123	Doreian, P	Partitioning networks based on generalized concepts of equivalence	J MATH SOCIOL
DOREIAN_P(1996)18:149	123	Doreian, P	A partitioning approach to structural balance	SOC NETWORKS
BATAGELJ_V(1997)19:143	123	Batagelj, V	Notes on blockmodeling	SOC NETWORKS
BATAGELJ_V(1999)1731:90	123	Batagelj, V	Partitioning approach to visualization of large graphs	LECT NOTES COMPUT SC
BATAGELJ_V(2000)22:173	123	Batagelj, V	Some analyses of Erdos collaboration graph	SOC NETWORKS
NEWMAN_M(2001)64:016131	123	Newman, MEJ	Scientific collaboration networks. I. Network construction and fundamental results	PHYS REV E
NEWMAN_M(2001)64:16132	23	Newman, MEJ	Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality	PHYS REV E
GIRVAN_M(2002)99:7821	123	Girvan, M	Community structure in social and biological networks	P NATL ACAD SCI USA
RAVASZ_E(2002)297:1551	123	Ravasz, E	Hierarchical organization of modularity in metabolic networks	SCIENCE
BARABASI_A(2002)311:590	23	Barabasi, AL	Evolution of the social network of scientific collaborations	PHYSICA A
RAVASZ_E(2003)67:026112	123	Ravasz, E	Hierarchical organization in complex networks	PHYS REV E

List of works on CPM path(1), main paths (2) and island (3) part 3

label	code	first author	title	journal
NEWMAN_M(2003)45:167	123	Newman, MEJ	The structure and function of complex networks	SIAM REV
NEWMAN_M(2003)68:026121	123	Newman, MEJ	Properties of highly clustered networks	PHYS REV E
RIVES_A(2003)100:1128	3	Rives, AW	Modular organization of cellular networks	P NATL ACAD SCI USA
GUIMERA_R(2003)68:065103	23	Guimera, R	Self-similar community structure in a network of human interactions	PHYS REV E
HOLME_P(2003)19:532	3	Holme, P	Subnetwork hierarchies of biochemical pathways	BIOINFORMATICS
NEWMAN_M(2003)68:036122	123	Newman, MEJ	Why social networks are different from other types of networks	PHYS REV E
GLEISER_P(2003)6:565	23	Gleiser, PM	Community structure in jazz	ADV COMPLEX SYST
NEWMAN_M(2004)69:026113	23	Newman, MEJ	Finding and evaluating community structure in networks	PHYS REV E
NEWMAN_M(2004)38:321	23	Newman, MEJ	Detecting community structure in networks	EUR PHYS J B
REICHARD_J(2004)93:218701	23	Reichardt, J	Detecting fuzzy community structures in complex networks with a Potts model	PHYS REV LETT
ARENAS_A(2004)38:373	3	Arenas, A	Community analysis in social networks	EUR PHYS J B
CLAUSET_A(2004)70:066111	23	Clauset, A	Finding community structure in very large networks	PHYS REV E
RADICCHI_F(2004)101:2658	123	Radicchi, F	Defining and identifying communities in networks	P NATL ACAD SCI USA
DONETTI_L(2004):P10012	123	Donetti, L	Detecting network communities: a new systematic and efficient algorithm	J STAT MECH
GUIMERA_R(2004)70:025101	23	Guimera, R	Modularity from fluctuations in random graphs and complex networks	PHYS REV E
GUIMERA_R(2005)433:895	123	Guimera, R	Functional cartography of complex metabolic networks	NATURE
DUCH_J(2005)72:027104	123	Duch, J	Community detection in complex networks using extremal optimization	PHYS REV E
DANON_L(2005):	123	Danon, L	COSIN book	–
PALLA_G(2005)435:814	3	Palla, G	Uncovering the overlapping community structure of complex networks in nature and society	NATURE
MUFF_S(2005)72:056107	23	Muff, S	Local modularity measure for network clusterizations	PHYS REV E
GFELLER_D(2005)72:056135	23	Gfeller, D	Finding instabilities in the community structure of complex networks	PHYS REV E
GUIMERA_R(2005)102:7794	3	Guimera, R	The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles	P NATL ACAD SCI USA
NEWMAN_M(2006)103:8577	3	Newman, MEJ	Modularity and community structure in networks	P NATL ACAD SCI USA
REICHARD_J(2006)74:016110	123	Reichardt, J	Statistical mechanics of community detection	PHYS REV E
BOCCALET_S(2006)424:175	3	Boccaletti, S	Complex networks: Structure and dynamics	PHYS REP
NEWMAN_M(2006)74:036104	23	Newman, MEJ	Finding community structure in networks using the eigenvectors of matrices	PHYS REV E
FORTUNAT_S(2007)104:36	123	Fortunato, S	Resolution limit in community detection	P NATL ACAD SCI USA
KUMPULA_J(2007)56:41	123	Kumpula, JM	Limited resolution in complex network community detection with Potts model approach	EUR PHYS J B
KUMPULA_J(2007)7:L209	23	Kumpula, JM	Limited resolution and multiresolution methods in complex network community detection	FLUCT NOISE LETT
GUIMERA_R(2007)3:63	3	Guimera, R	Classes of complex networks defined by role-to-role connectivity profiles	NAT PHYS
ROSVALL_M(2007)104:7327	3	Rosvall, M	An information-theoretic framework for resolving community structure in complex networks	P NATL ACAD SCI USA
GUIMERA_R(2007)76:036102	23	Guimera, R	Module identification in bipartite and directed networks	PHYS REV E

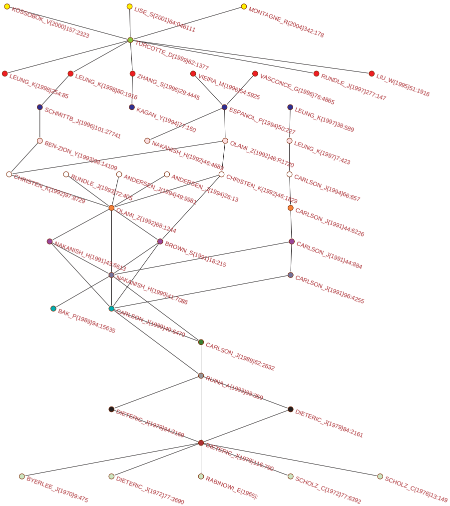
List of works on CPM path(1), main paths (2) and island (3) part 4

label	code	first author	title	journal
SALES-PA_M(2007)104:15224	123	Sales-Pardo, M	Extracting the hierarchical organization of complex systems	P NATL ACAD SCI USA
ARENAS_A(2008)10:053039	123	Arenas, A	Analysis of the structure of complex networks at different resolution levels	NEW J PHYS
CLAUSET_A(2008)453:98	3	Clauset, A	Hierarchical structure and the prediction of missing links in networks	NATURE
KUMPULA_J(2008)78:026109	3	Kumpula, JM	Sequential algorithm for fast clique percolation	PHYS REV E
KARRER_B(2008)77:046119	23	Karrer, B	Robustness of community structure in networks	PHYS REV E
BLONDEL_V(2008):P10008	23	Blondel, VD	Fast unfolding of communities in large networks	J STAT MECH-THEORY E
LEUNG_I(2009)79:066107	3	Leung, IXY	Towards real-time community detection in large networks	PHYS REV E
LANCICHI_A(2009)11:033015	123	Lancichinetti, A	Detecting the overlapping and hierarchical community structure of complex networks	NEW J PHYS
LANCICHI_A(2009)80:016118	23	Lancichinetti, A	Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities	PHYS REV E
RONHOVDE_P(2009)80:016109	23	Ronhovde, P	Multiresolution community detection for megascale networks by information-based replica correlations	PHYS REV E
GOMEZ_S(2009)80:016114	3	Gomez, S	Analysis of community structure in networks of correlated data	PHYS REV E
TRAAG_V(2009)80:036115	23	Traag, VA	Community detection in networks with positive and negative links	PHYS REV E
NICOSIA_V(2009):P03024	123	Nicosia, V	Extending the definition of modularity to directed graphs with overlapping communities	J STAT MECH
EVANS_T(2009)80:016105	123	Evans, TS	Line graphs, link partitions, and overlapping communities	PHYS REV E
LANCICHI_A(2009)80:056117	3	Lancichinetti, A	Community detection algorithms: A comparative analysis	PHYS REV E
BARBER_M(2009)80:026129	3	Barber, MJ	Detecting network communities by propagating labels under constraints	PHYS REV E
FORTUNAT_S(2010)486:75	123	Fortunato, S	Community detection in graphs	PHYS REP
GOOD_B(2010)81:046106	123	Good, BH	Performance of modularity maximization in practical contexts	PHYS REV E
LANCICHI_A(2010)81:046110	3	Lancichinetti, A	Statistical significance of communities in networks	PHYS REV E
RADICCHI_F(2010)82:026102	23	Radicchi, F	Combinatorial approach to modularity	PHYS REV E
LANCICHI_A(2010)5:0011976	3	Lancichinetti, A	Characterizing the Community Structure of Complex Networks	PLOS ONE
AHN_Y(2010)466:761	23	Ahn, YY	Link communities reveal multiscale complexity in networks	NATURE
EVANS_T(2010)77:265	3	Evans, TS	Line graphs of weighted networks for overlapping communities	EUR PHYS J B
MUCHA_P(2010)328:876	3	Mucha, PJ	Community Structure in Time-Dependent, Multiscale, and Multiplex Networks	SCIENCE
GREGORY_S(2010)12:103018	3	Gregory, S	Finding overlapping communities in networks by label propagation	NEW J PHYS
KARRER_B(2011)83:016107	23	Karrer, B	Stochastic blockmodels and community structure in networks	PHYS REV E
EXPERT_P(2011)108:7663	23	Expert, P	Uncovering space-independent communities in spatial networks	P NATL ACAD SCI USA
PSORAKIS_I(2011)83:066114	23	Psorakis, I	Overlapping community detection using Bayesian non-negative matrix factorization	PHYS REV E
TRAAG_V(2011)84:016114	23	Traag, VA	Narrow scope for resolution-limit-free community detection	PHYS REV E
DECELLE_A(2011)107:065701	3	Decelle, A	Inference and Phase Transitions in the Detection of Modules in Sparse Networks	PHYS REV LETT
BALL_B(2011)84:036103	123	Ball, B	Efficient and principled method for detecting communities in networks	PHYS REV E
DECELLE_A(2011)84:066106	123	Decelle, A	Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications	PHYS REV E

List of works on CPM path(1), main paths (2) and island (3) part 5

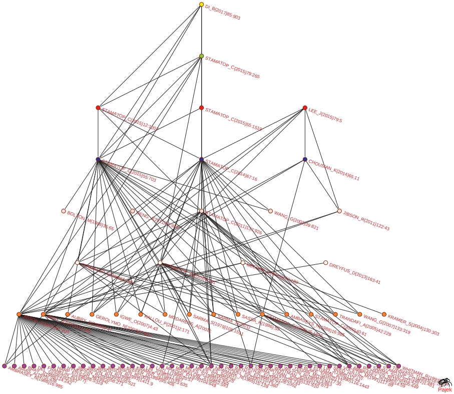
label	code	first author	title	journal
LANCICHI_A(2011)6:0018961	23	Lancichinetti, A	Finding Statistically Significant Communities in Networks	PLOS ONE
LANCICHI_A(2011)84:066122	23	Lancichinetti, A	Limits of modularity maximization in community detection	PHYS REV E
NADAKUDI_R(2012)108:188701	3	Nadakuditi, RR	Graph Spectra and the Detectability of Community Structure in Networks	PHYS REV LETT
PEIXOTO_T(2012)85:056122	123	Peixoto, TP	Entropy of stochastic blockmodel ensembles	PHYS REV E
PEIXOTO_T(2013)110:148701	123	Peixoto, TP	Parsimonious Module Inference in Large Networks	PHYS REV LETT
GOPALAN_P(2013)110:14534	3	Gopalan, PK	Efficient discovery of overlapping communities in massive networks	P NATL ACAD SCI USA
PEIXOTO_T(2014)89:012804	123	Peixoto, TP	Efficient Monte Carlo and greedy heuristic for the inference of stochastic block models	PHYS REV E
PEIXOTO_T(2014)4:011047	123	Peixoto, TP	Hierarchical Block Structures and High-Resolution Model Selection in Large Networks	PHYS REV X
LARREMOR_D(2014)90:012805	123	Larremore, DB	Efficiently inferring community structure in bipartite networks	PHYS REV E
ZHANG_P(2014)111:18144	23	Zhang, P	Scalable detection of statistically significant communities and hierarchies, using message passing for modularity	P NATL ACAD SCI USA
KAWAMOTO_T(2015)91:012809	3	Kawamoto, T	Estimating the resolution limit of the map equation in community detection	PHYS REV E
ZHANG_X(2015)91:032803	3	Zhang, X	Identification of core-periphery structure in networks	PHYS REV E
PEIXOTO_T(2015)5:011033	123	Peixoto, TP	Model Selection and Hypothesis Testing for Large-Scale Network Models with Overlapping Groups	PHYS REV X
JIANG_J(2015)91:062805	3	Jiang, JQ	Stochastic block model and exploratory analysis in signed networks	PHYS REV E
PEIXOTO_T(2015)92:042807	23	Peixoto, TP	Inferring the mesoscale structure of layered, edge-valued, and time-varying networks	PHYS REV E
PEROTTI_J(2015)92:062825	23	Perotti, JI	Hierarchical mutual information for the comparison of hierarchical community structures in complex networks	PHYS REV E
ZHANG_P(2016)93:012303	3	Zhang, P	Community detection in networks with unequal groups	PHYS REV E
VALLES-C_T(2016)6:011036	3	Valles-Catala, T	Multilayer Stochastic Block Models Reveal the Multilayer Structure of Complex Networks	PHYS REV X
NEWMAN_M(2016)117:078301	23	Newman, MEJ	Estimating the Number of Communities in a Network	PHYS REV LETT
HRIC_D(2016)6:031038	123	Hric, D	Network Structure, Metadata, and the Prediction of Missing Nodes and Annotations	PHYS REV X
FORTUNAT_S(2016)659:1	123	Fortunato, S	Community detection in networks: A user guide	PHYS REP
NEWMAN_M(2016)94:052315	123	Newman, MEJ	Equivalence between modularity optimization and maximum likelihood methods for community detection	PHYS REV E
FISTER_I(2016)4:00049	123	Fister, I	Toward the Discovery of Citation Cartels in Citation Networks	FRONT PHYS
YANG_L(2016)9:3390/a9040073	123	Yang, LJ	Community Structure Detection for Directed Networks through Modularity Optimisation	ALGORITHMS
PEIXOTO_T(2017)95:012317	123	Peixoto, TP	Nonparametric Bayesian inference of the microcanonical stochastic block model	PHYS REV E

Island 9



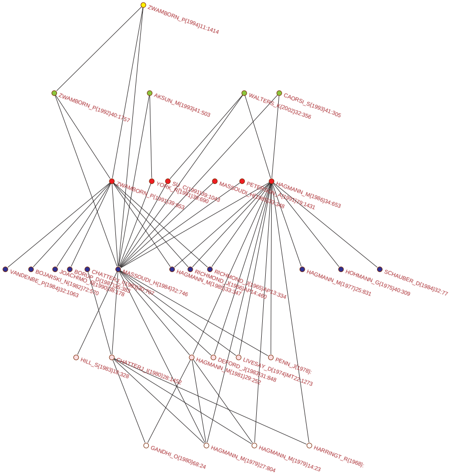
The island 9 has 44 nodes. The maximal weight is 2.416 10^{-14}. Papers in this island deal with eartquake modeling. One among models is a "spring-block model". The main journals are Phys Rev (A, E, lett), Physica A and J GEOPHYS RES.

Island 7



The island 7 has 74 nodes. The maximal weight is 4.9611 10^{-18}. Papers from island 7 deal with landslides (some related to earthquakes). They are using "multi-block modeling of landslides". The main journals are SOIL DYN EARTHQ ENG, ENG GEOL and LANDSLIDES.

Island 2



The island 2 has 33 nodes. The maximal weight is 2.462 10^{-19}. From island.csv we see that papers from this island deal with numerical methods for computation of electromagnetic field. They use block model . . . Most papers are published in the journal IEEE T MICROW THEORY.

Coauthorship networks

Ct is an undirected network obtained from $\mathbf{N}^T * \mathbf{N}$, where

$$\mathbf{N} = \text{diag}\left(\frac{1}{\max(1, \text{outdeg}(p))}\right)\mathbf{WA}$$

by symmetrization.
Authors with the largest P_S-core values in **Ct** are listed in Slide 35. The value of an author is equal to the sum of all his/her fractional contributions to works with authors inside the core. For better readability loops are removed.

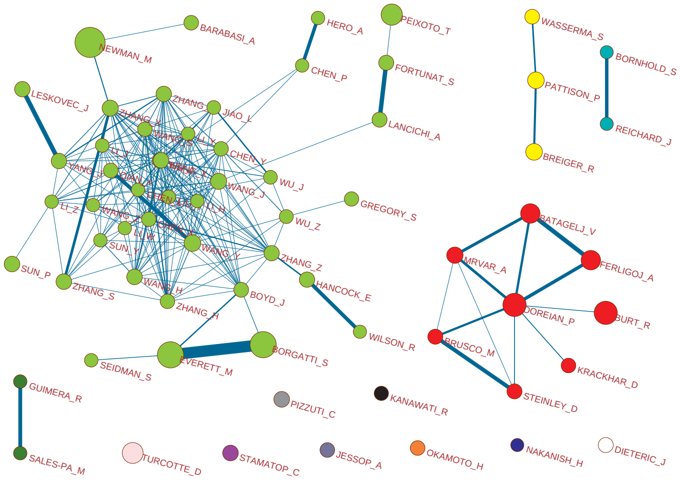
Ct′ is an undirected network without loops obtained from $\mathbf{N}^T * \mathbf{N}'$, where

$$\mathbf{N}' = \text{diag}\left(\frac{1}{\max(1, \text{outdeg}(p) - 1)}\right)\mathbf{WA},$$

by symmetrization and setting the diagonal values to 0.
Authors with the largest P_S-core values in **Ct**′ are listed in Slide 38. The value of an author is equal to the sum of all his/her fractional collaborations with authors inside the core.

In both figures the size of nodes is proportional to its P_S-core value.

P_S-cores at level 4 in **Ct**



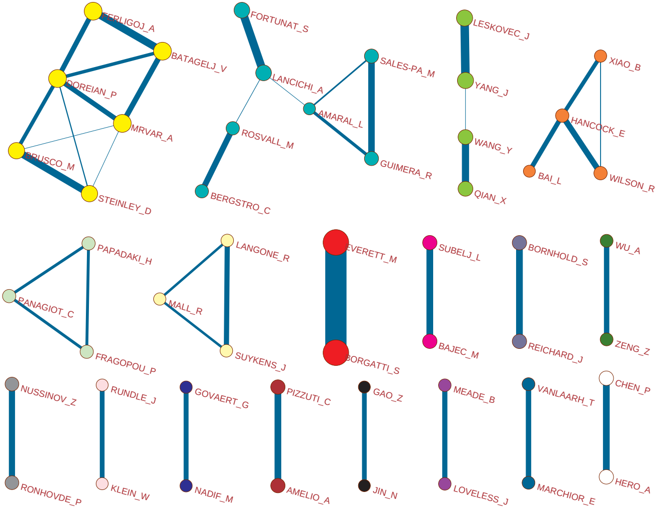
Authors with the largest P_S-core values in **Ct**

i	value	author	i	value	author
1	21.0347	NEWMAN_M	21	5.6307	LIU_J
2	15.9653	BORGATTI_S	22	5.5417	YANG_J
3	15.9653	EVERETT_M	23	5.5417	LESKOVEC_J
4	12.5000	BURT_R	24	5.5417	ZHANG_J
5	12.5000	DOREIAN_P	25	5.4432	HANCOCK_E
6	10.4722	PEIXOTO_T	26	5.4432	ZHANG_Z
7	10.1126	TURCOTTE_D	27	5.2589	STEINLEY_D
8	8.7900	FERLIGOJ_A	28	5.2589	BRUSCO_M
9	8.7900	BATAGELJ_V	29	5.2500	DIETERIC_J
10	6.5115	WANG_Y	30	5.2483	LANCICHI_A
11	6.4097	PATTISON_P	31	5.2483	FORTUNAT_S
12	6.4097	BREIGER_R	32	5.1111	BOYD_J
13	6.2083	MRVAR_A	33	5.0633	WANG_X
14	6.0292	ZHANG_X	34	5.0278	QIAN_X
15	6.0292	WANG_J	35	5.0208	WASSERMA_S
16	5.7500	PIZZUTI_C	36	5.0000	OKAMOTO_H
17	5.7014	STAMATOP_C	37	5.0000	JESSOP_A
18	5.6736	SUN_P	38	4.9881	BARABASI_A
19	5.6669	ZHANG_S	39	4.9775	KRACKHAR_D
20	5.6307	WANG_H	40	4.9112	ZHANG_H

Computing **Ct**$'$ in Pajek

```
select/read WAc network
Network/Create Vector/Centrality/Degree/Output
Network/2-Mode Network/Partition into 2 Modes
Operations/Vector+Partition/Extract Subvector [1] = V1
Vector/Create Constant Vector [5695,1] = V2
select V1 as Second vector
Vectors/Max(First,Second)
Vector/Transform/Invert
Operations/Network+Vector/Vector#Network/Output = N
select V1 as First vector
select V2 as Second vector
Vectors/Subtract (First-Second)
Vectors/Max(First,Second)
Vector/Transform/Invert
select/read WAc network
Operations/Network+Vector/Vector#Network/Output = N'
select network N
Network/2-Mode/Transpose 2-Mode
select N' as Second network
Networks/Multiply Networks [Yes]
Network/Create New Network/Transform/Remove/Loops
Network/Create New Network/Transform/Arcs -> Edges/Biderected Only/Sum
File/Network/Change Label [Ct'(WAc)]
```

P_S-cores at level 3.5 in $\mathbf{Ct}'$



Authors with the largest P_S-core values in **Ct**′

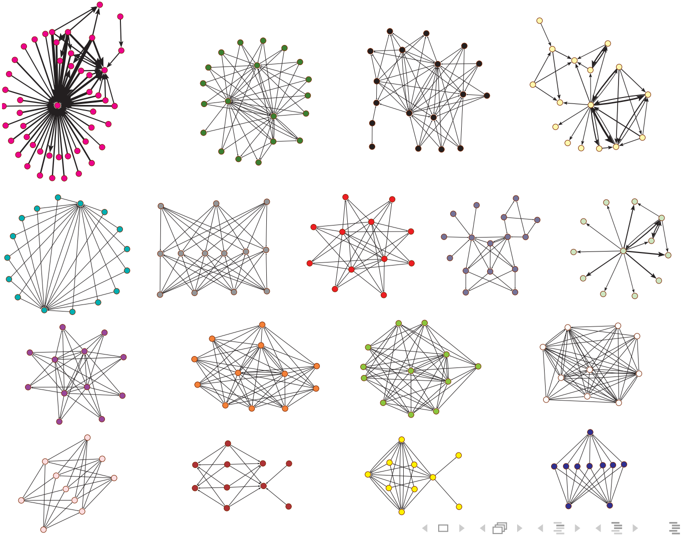
i	value	author	i	value	author
1	15.8333	BORGATTI_S	16	5.0000	AMELIO_A
2	15.8333	EVERETT_M	17	5.0000	BAJEC_M
3	7.6667	FERLIGOJ_A	18	5.0000	SUBELJ_L
4	7.6667	BATAGELJ_V	19	5.0000	CHEN_P
5	7.6667	MRVAR_A	20	5.0000	PIZZUTI_C
6	7.6667	DOREIAN_P	21	5.0000	REICHARD_J
7	6.4333	STEINLEY_D	22	5.0000	BORNHOLD_S
8	6.4333	BRUSCO_M	23	4.8333	SALES-PA_M
9	6.3333	YANG_J	24	4.8333	GUIMERA_R
10	6.3333	LESKOVEC_J	25	4.5833	NUSSINOV_Z
11	6.0000	LANCICHI_A	26	4.5833	RONHOVDE_P
12	6.0000	FORTUNAT_S	27	4.3333	ROSVALL_M
13	5.3333	QIAN_X	28	4.3333	BERGSTRO_C
14	5.3333	WANG_Y	29	4.3333	WILSON_R
15	5.0000	HERO_A	30	4.3333	HANCOCK_E

Citations among authors

The network $\mathbf{Acite} = \mathbf{WA}^T * \mathbf{CiteC} * \mathbf{WA}$ describes the citations among authors.

The value of element $\mathbf{Acite}[u, v]$ is equal to the number of works coauthored by u that are citing a work coauthored by v.

Citations among authors – link islands [10 50]



Citations among authors – link islands 17 and 16