



SNA.
Evolution of
the field

V. Batagelj,
D. Maltseva

Introduction

Data

Networks
construction

Statistics on
networks

Citation
network
analysis

Social Network Analysis

The evolution of the field

Vladimir Batagelj, Daria Maltseva

IMFM Ljubljana, IAM UP Koper and NRU HSE Moscow

Applied Statistics

Ribno, 23-26. September 2018



Outline

SNA. Evolution of the field

V. Batagelj,
D. Maltseva

Introduction

Data

Networks
construction

Statistics on
networks

Citation
network
analysis

- 1 Introduction
- 2 Data collection
- 3 Networks construction
- 4 Statistics on networks
- 5 Citation network



Introduction

SNA.
Evolution of
the field

V. Batagelj,
D. Maltseva

Introduction

Data

Networks
construction

Statistics on
networks

Citation
network
analysis

Aim of the study - to trace the development of the field of Social Network Analysis as a discipline.

Previous studies on SNA were done by:

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

"social network*"

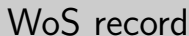
Additionally, all the articles from the following journals were collected.

Social Networks, Network Science,
Computational Social Networks, Applied Network Science,
Social Network Analysis and Mining,
Online Social Networks and Media, Journal of Complex
Networks, Journal of Social Structure, Connections

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 data set was collected in 2007, second - in June, 2018.

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 (using RIS bibliographic format and converting it to WoS).

As the data set of 2007 was already completed, we made this additional search only for works 2008-* in July 2018.



WoS record

V. Batagelj,
D. Maltseva

Data

Networks construction

Statistics on networks

Citation
network
analysis

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

Types of networks and partitions

SNA.
Evolution of
the field

V. Batagelj,
D. Maltseva

Introduction

Data

Networks
construction

Statistics on
networks

Citation
network
analysis

We applied the WoS2Pajek 1.5 to the collected data.

The following networks were constructed:

- ① the authorship network WA on works \times authors (from the field AU),
- ② the journalship network WJ on works \times journals (from the field CR or J9),
- ③ the keywordship network WK on works \times keywords (from the field ID or DE or TI),
- ④ the citation network $Cite$ on works (from the field CR).

We obtained also the following partitions:

- ① partition *year* of works by publication year,
- ② the DC partition distinguishing between works with complete description ($DC=1$) and the cited only works ($DC=0$),
- ③ the vector of number of pages NP .

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

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

has the following structure

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

All its elements are in upper case.

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

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

For example: LEFKOVIT L(1985)70:585

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

Unusual names start with character * or \$.

However, same works can be named by different names:

BOYD_D(2007)13:

BOYD_D(2008)13:210

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 largest counts we prepared lists of possible equivalents and manually determined equivalence classes. With a program in R we produced a Pajek's partition *EQ.clu* file used for shrinking the set of works.

Using the partition $p = \text{worksEQ}$, we shrink using Pajek the Citation network *cite*, *WA*, *WJ*, and *WK*.

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

Works with complete description = 70795

1-mode network *Cite*:

	N of nodes	N of arcs
CiteN	1297133	2753767

2-mode networks *WA*, *WJ*, *WK*:

	N of nodes 1	N of nodes 2	N of nodes (sum)	N of arcs
WA	1297133	395972	1693105	1442242
WJ	1297133	70425	1367558	1301276
WK	1297133	32409	1329542	1167670

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.)

Citation network

Distribution by years

SNA.
Evolution of
the field

V. Batagelj,
D. Maltseva

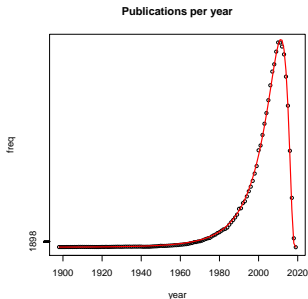
Introduction

Data

Networks
construction

Statistics on
networks

Citation
network
analysis



Citation network

Indegree distribution

SNA.
Evolution of
the field

V. Batagelj,
D. Maltseva

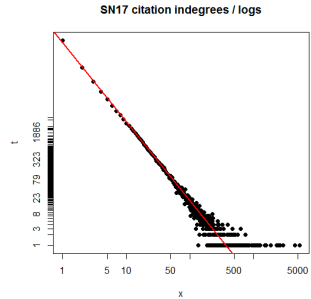
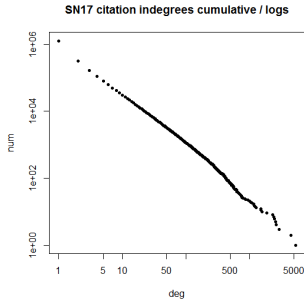
Introduction

Data

Networks
construction

Statistics on
networks

Citation
network
analysis



The most cited works

indegree in Cite

SNA.

Evolution of the field

V. Batagelj,
D. Maltseva

Introduction

Data

Networks construction

Statistics on networks

Citation network analysis

i	freq	id	i	freq	id
1	5348	WASSERMA_S(1994):	31	734	NEWMAN_M(2001)98:404
2	4471	GRANOVET_M(1973)78:1360	32	719	NEWMAN_M(2010):
3	2906	WATTS_D(1998)393:440	33	701	PORTES_A(1998)24:1
4	2614	BARABASI_A(1999)286:509	34	687	BLEI_D(2003)3:993
5	2561	FREEMAN_L(1979)1:215	35	670	BURT_R(2004)110:349
6	2447	BOYD_D(2007)13:210	36	654	HANSEN_M(1999)44:82
7	2429	MCPHERSON_M(2001)27:415	37	639	PALLA_G(2005)435:814
8	2330	BURT_R(1992):	38	634	CLAUSET_A(2004)70:066111
9	1886	COLEMAN_J(1988)94:95	39	629	BONACICH_P(1987)92:1170
10	1572	NEWMAN_M(2003)45:167	40	628	ERDOS_P(1959)6:290
11	1520	GIRVAN_M(2002)99:7821	41	628	UZZI_B(1997)42:35
12	1510	PUTNAM_R(2000):	42	628	ROGERS_E(2003):
13	1285	ALBERT_R(2002)74:47	43	613	PUTNAM_R(1993):
14	1240	GRANOVET_M(1985)91:481	44	593	BERKMAN_L(1979)109:186
15	1192	SCOTT_J(2000):	45	583	ZACHARY_W(1977)33:452
16	1171	EVERETT_M(2002):	46	572	BORGATTI_S(2009)323:892
17	1166	NEWMAN_M(2004)69:026113	47	569	NEWMAN_M(2001)64:025102
18	1093	COLEMAN_J(1990):	48	565	BURT_R(2005):
19	1058	STEINFIE_C(2007)12:1143	49	561	ADLER_P(2002)27:17
20	1034	FORTUNAT_S(2010)486:75	50	559	CHRISTAK_N(2008)358:2249
21	999	BORGATTI_S(2002):	51	555	ROGERS_E(1995):
22	945	CHRISTAK_N(2007)357:370	52	554	MILGRAM_S(1967)1:61
23	867	FREEMAN_L(1977)40:35	53	553	BARON_R(1986)51:1173
24	854	HANNEMAN_R(2005):	54	550	GRANOVET_M(1978)83:1420
25	800	LIN_N(2001):	55	539	FISCHER_C(1982):
26	757	KAPLAN_A(2010)53:59	56	537	BRIN_S(1998)30:107
27	756	BLONDEL_V(2008):P10008	57	524	MARSDEN_P(1990)16:435
28	742	NAHAPIET_J(1998)23:242	58	523	KEMP_D(2003):137
29	740	FORNELL_C(1981)18:39	59	523	KLEINBER_J(1999)46:604
30	740	NEWMAN_M(2006)103:8577	60	517	BOCCALET_S(2006)424:175

The most citing works

outdegree in Cite

SNA.

Evolution of
the field

V. Batagelj,
D. Maltseva

Introduction

Data

Networks
construction

Statistics on
networks

Citation
network
analysis

i	freq	id	i	freq	id
1	1572	CHAPMAN_C(2016):1	11	731	TSATSOU_P(2014):1
2	1406	HRUSCHKA_D(2010)5:1	12	654	GOODALE_E(2017):IX
3	1293	COWARD_F(2015):1	13	649	PEPPER_G(2017)40:S0140525X1700190X
4	1254	FITZGERA_P(2008):1	14	632	STROM_R(2012):1
5	1207	DAVIES_N(2015):V	15	613	SCHACHNE_G(2015)23:49
6	1055	MARSH_C(2009):1	16	597	COSTA_L(2011)60:329
7	942	YUS_F(2011)213:1	17	593	BRANDES_U(2005)3418:1
8	929	BOCCALET_S(2006)424:175	18	586	ROBERTS_J(2014):1
9	799	REEVES_M(2017):1	19	557	GUNTER_B(2016):1
10	768	GROSS_J(2007):1	20	547	CASTELLA_C(2009)81:591

- MUIJS, D., Reynolds, D., CHAPMAN, C. (2015). Educational effectiveness and improvement research and practice: The emergence of the discipline. In The Routledge International Handbook of Educational Effectiveness and Improvement (pp. 33-56). Routledge.
- Hruschka, D. J. (2010). Friendship: Development, ecology, and evolution of a relationship (Vol. 5). Univ of California Press.
- Coward, F., Hosfield, R., Pope, M., Wenban-Smith, F. (Eds.). (2015). Settlement, society and cognition in human evolution. Cambridge University Press.
- Fitzgerald, P., Lambkin, B. (2008). Migration in Irish history 1607-2007. Springer.
- Davies, N.B. Animal Social Networks Foreword. In: Krause, J., James, R., Franks, D. W., Croft, D. P. (Eds.). (2015). Animal social networks. Oxford University Press, USA.
- Marsh, C. J. (2009). Key concepts for understanding curriculum. Routledge.

Authors with the largest number of papers indegree in WA

SNA.

Evolution of the field

V. Batagelj,
D. Maltseva

Introduction

Data

Networks
construction

Statistics on
networks

Citation
network
analysis

Rank	Value	Id	Rank	Value	Id
1	1169	WANG_Y	21	552	KIM_H
2	883	ZHANG_Y	22	550	CHEN_J
3	868	CHEN_Y	23	536	LIU_X
4	847	LI_Y	24	533	WANG_L
5	838	WANG_X	25	509	LI_H
6	819	ZHANG_J	26	490	KIM_Y
7	788	WANG_J	27	485	ZHANG_Z
8	786	LIU_Y	28	474	WANG_Z
9	766	LEE_J	29	471	WANG_S
10	765	LEE_S	30	471	CHEN_X
11	749	LI_J	31	471	NEWMAN_M
12	708	LI_X	32	462	CHEN_L
13	696	CHEN_C	33	461	ZHANG_L
14	690	KIM_J	34	450	YANG_Y
15	620	WANG_H	35	450	ZHANG_H
16	611	ZHANG_X	36	432	WU_J
17	611	LIU_J	37	431	LEE_H
18	570	CHEN_H	38	420	LI_Z
19	557	KIM_S	39	420	WANG_W
20	554	WANG_C	40	417	LI_L

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

Number of authors in works outdegree in WA

SNA.
Evolution of
the field

V. Batagelj,
D. Maltseva

Introduction

Data

Networks
construction

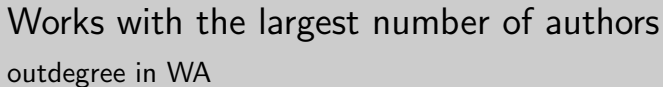
Statistics on
networks

Citation
network
analysis

Cluster	Freq	Freq%	Cluster	Freq	Freq%
1	1239496	95.5566	21	4	0.0003
2	18637	1.4368	22	3	0.0002
3	16661	1.2844	23	4	0.0003
4	10617	0.8185	24	2	0.0002
5	5759	0.4440	25	1	0.0001
6	2802	0.2160	26	2	0.0002
7	1322	0.1019	27	5	0.0004
8	686	0.0529	28	2	0.0002
9	384	0.0296	29	1	0.0001
10	247	0.0190	31	3	0.0002
11	155	0.0119	36	1	0.0001
12	90	0.0069	41	1	0.0001
13	70	0.0054	42	1	0.0001
14	54	0.0042	43	1	0.0001
15	32	0.0025	48	1	0.0001
16	12	0.0009	53	1	0.0001
17	14	0.0011	126	1	0.0001
18	9	0.0007			
19	6	0.0005			
20	2	0.0002			
SUM				1297133	100

Works with the largest number of authors

Rank	Freq	Id
1	126	WANG_M(2016)34:828
2	53	VASHISHT_R(2012)7:0039808
3	48	SNIJDERS_T(2007)170:322
4	43	GUSTAVSS_A(2011)21:718
5	42	DOLL_L(1992)29:1
6	41	MAGLIANO_L(2006)15:219



SNA.
Evolution of
the field

V. Batagelj,
D. Maltseva

Introduction

Data

Networks construction

Statistics on networks

Citation
network
analysis

Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking / Nature Biotechnology volume 34, pages 828–837 (2016)

Mingxun Wang, Jeremy J Carver, Vanessa V Phelan, Laura M Sanchez, Neha Garg, Yao Peng, Don Duy Nguyen, Jeramie Watrous, Clifford A Kapono, Tal Luzzatto-Knaan, Carla Porto, Amina Bouslimani, Alexey V Melnik, Michael J Meehan, Wei-Ting Liu, Max Crüsemann, Paul D Boudreau, Eduardo Esquenazi, Mario Sandoval-Calderón, Roland D Kersten, Laura A Pace, Robert A Quinn, Katherine R Duncan, Cheng-Chih Hsu, Dimitrios J Floros, Ronnie G Gavilan, Karin Kleigrew, Trent Northen, Rachel J Dutton, Delphine Parrot, Erin E Carlson, Bertrand Aigle, Charlotte F Michelsen, Lars Jelsbak, Christian Sohlenkamp, Pavel Pevzner, Anna Edlund, Jeffrey McLean, Jörn Piel, Brian T Murphy, Lena Gerwick, Chih-Chuang Liaw, Yu-Liang Yang, Hans-Ulrich Humpf, Maria Maansson, Robert A Keyzers, Amy C Sims, Andrew R Johnson, Ashley M Sidebottom, Brian E Sedio, Andreas Klitgaard, Charles B Larson, Cristopher A Boya P, Daniel Torres-Mendoza, David J Gonzalez, Denise B Silva, Lucas M Marques, Daniel P Demarque, Egle Pociute, Ellis C O'Neill, Enora Briand, Eric J N Helfrich, Eve A Granatosky, Evgenia Glukhov, Florian Ryffel, Hailey Houson, Hosein Mohimani, Jenan J Kharbush, Yi Zeng, Julia A Vorholt, Kenji L Kurita, Pep Charusanti, Kerry L McPhail, Kristian Fog Nielsen, Lisa Vuong, Maryam Elfeki, Matthew F Traxler, Niclas Emene, Nobuhiro Koyama, Oliver B Vining, Ralph Baric, Ricardo R Silva, Samantha J Mascuch, Sophie Tomasi, Stefan Jenkins, Venkat Macherla, Thomas Hoffman, Vinayak Agarwal, Philip G Williams, Jingqui Dai, Ram Neupane, Joshua Gurr, Andrés M C Rodríguez, Anne Lamsa, Chen Zhang, Kathleen Dorrestein, Brendan M Duggan, Jehad Almaliti, Pierre-Marie Allard, Prasad Phapale, Louis-Felix Nothias, Theodore Alexandrov, Marc Litaudon, Jean-Luc Wolfender, Jennifer E Kyle, Thomas O Metz, Tyler Peryea, Dac-Trung Nguyen, Danielle vanLeer, Paul Shinn, Ajit Jadhav, Rolf Müller, Katrina M Waters, Wenyan Shi, Xueting Liu, Lixin Zhang, Rob Knight, Paul R Jensen, Bernhard Ø Palsson, Kit Pogliano, Roger G Linington, Marcelino Gutiérrez, Norberto P Lopes, William H Gerwick, Bradley S Moore, Pieter C Dorrestein, Nuno Bandeira.

The most used journals

indegree in WJ

SNA.

Evolution of
the field

V. Batagelj,
D. Maltseva

Introduction

Data

Networks
construction

Statistics on
networks

Citation
network
analysis

Rank	Value	Id	Rank	Value	Id
1	7080	LECT NOTES COMPUT SC	31	1258	AM J PSYCHIAT
2	3859	SOC SCI MED	32	1221	J BUS RES
3	3408	J PERS SOC PSYCHOL	33	1217	MANAGE SCI
4	2719	COMPUT HUM BEHAV	34	1185	ACAD MANAGE REV
5	2631	SCIENCE	35	1182	J CONSULT CLIN PSYCH
6	2602	AM J PUBLIC HEALTH	36	1151	ORGAN SCI
7	2599	P NATL ACAD SCI USA	37	1150	ADDICTION
8	2208	NATURE	38	1143	STRATEGIC MANAGE J
9	2058	AM SOCIOL REV	39	1087	J GERONTOL B-PSYCHOL
10	1945	PHYSICA A	40	1075	PEDIATRICS
11	1815	ANIM BEHAV	41	1055	AM J EPIDEMIOLOG
12	1778	JAMA-J AM MED ASSOC	42	1050	COMPUT EDUC
13	1763	LANCET	43	1022	DEV PSYCHOL
14	1759	SCIENTOMETRICS	44	1022	PSYCHOL BULL
15	1734	AM J SOCIOL	45	1007	J ADOLESCENT HEALTH
16	1703	ACAD MANAGE J	46	997	J MARKETING
17	1632	LECT NOTES ARTIF INT	47	996	ARCH GEN PSYCHIAT
18	1573	J APPL PSYCHOL	48	994	AIDS BEHAV
19	1551	SOC NETWORKS	49	972	PERS INDIV DIFFER
20	1509	AM ECON REV	50	949	PERS SOC PSYCHOL B
21	1433	J MARRIAGE FAM	51	947	J BUS ETHICS
22	1400	BRIT MED J	52	939	J MARKETING RES
23	1399	CHILD DEV	53	925	INFORM SCIENCES
24	1373	EXPERT SYST APPL	54	916	HARVARD BUS REV
25	1365	NEW ENGL J MED	55	915	IEEE T KNOWL DATA EN
26	1363	COMMUN ACM	56	901	DRUG ALCOHOL DEPEND
27	1355	RES POLICY	57	900	WORLD DEV
28	1279	GERONTOLOGIST	58	899	AM J PREV MED
29	1275	BRIT J PSYCHIAT	59	895	ADDICT BEHAV
30	1271	SOC FORCES	60	893	J CONSUM RES

The most used keywords indegree in WK

SNA.
**Evolution of
the field**

V. Batagelj,
D. Maltseva

Introduction

Data

Networks
construction

Statistics on
networks

Citation
network
analysis

Rank	Value	Id	Rank	Value	Id
1	51333	social	31	3485	structure
2	46191	network	32	3479	life
3	11751	analysis	33	3444	risk
4	10219	model	34	3358	research
5	8104	community	35	3143	learn
6	8090	use	36	3116	influence
7	7596	base	37	3054	student
8	7439	information	38	3054	impact
9	7061	health	39	3049	perspective
10	7023	behavior	40	3042	complex
11	6745	online	41	3024	theory
12	6087	networking	42	2859	organization
13	5833	media	43	2828	relationship
14	5404	support	44	2802	algorithm
15	5101	communication	45	2776	education
16	5013	study	46	2714	group
17	4759	datum	47	2704	mobile
18	4376	management	48	2698	tie
19	4372	internet	49	2695	adult
20	4164	knowledge	50	2633	approach
21	4126	user	51	2608	care
22	4023	facebook	52	2551	adolescent
23	3984	technology	53	2479	role
24	3907	site	54	2472	state
25	3888	web	55	2467	innovation
26	3855	self	56	2434	pattern
27	3784	graph	57	2385	effect
28	3676	performance	58	2339	people
29	3534	service	59	2333	trust
30	3512	dynamics	60	2332	family

The network Cite has 1297133 nodes and 2753767 arcs.

Cluster	Freq	Freq%	CumFreq	CumFreq%
0	41954	3.2344	41954	3.2344
1	933315	71.9521	975269	75.1865
2	154895	11.9413	1130164	87.1278
3	58141	4.4823	1188305	91.6101
4	29885	2.3039	1218190	93.9140
5	17651	1.3608	1235841	95.2748

Most of nodes are terminal ($DCr = 0$) or 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 222 086 nodes and 1 521 434 arcs.



Components

SNA. Evolution of the field

V. Batagelj,
D. Maltseva

Introduction

Data

Networks
construction

Statistics on
networks

Citation
network
analysis

To get an acyclic network we applied the *preprint transformation* to CiteB. The resulting network CiteT has 222 189 nodes and 1 521 658 arcs. We computed the SPC weights on its arcs. The total flow is equal to (??) $1.625 \cdot 10^{20}$.

First we determine the CPM path in this network, followed by the key-routes approach.

Afterward we determine SPC link islands [Line weights] of sizes [20, 200] and got 5 islands. Also we computed SPC node islands [Vertex weights] of sizes [20, 200] and got 1 island. Node islands were also counted for a network with probabilistic flow (got 1 island).



Cite network

SNA.
**Evolution of
the field**

V. Batagelj,
D. Maltseva

Introduction

Data

Networks
construction

Statistics on
networks

**Citation
network
analysis**

Pictures: Main path Key route Islands Prob flow



Cite network

SNA.

Evolution of the field

V. Batagelj,
D. Maltseva

Introduction

Data

Networks
construction

Statistics on
networks

Citation
network
analysis

N	author	
1	GRANOVET, MS	
2	Newman, MEJ	
3	Albert, R	
4	Boccaletti, S	
5	WHITE, HC	
6	Newman, MEJ	
7	Newman, MEJ	
8	Newman, MEJ	
9	Strogatz, SH	
10	Newman, MEJ	
11	Newman, MEJ	
12	Newman, MEJ	
13	CARTWRIGHT, D	
14	Clauset, A	
15	Newman, MEJ	
16	Newman, MEJ	
17	Valente, TW	
18	BURT, RS	
19	STEPHENSON, K	
20	BREIGER, RL	ALGORITHM FOR CLUSTERING RELATIONAL DATA WITH APPLICATIONS
21	FREEMAN, LC	
22	BURT, RS	
23	HOLLAND, PW	
24	ALBA, RD	
25	Moore, C	
26	MCPHERSON, JM	HYPER
27	MARIOLIS, P	
28	BURT, RS	POSITIONS IN MULTIPLE NETWORK SYSTEMS .1. GENERAL CONCEPTION
29	BURT, RS	POSITIONS IN MULTIPLE NETWORK SYSTEMS .2. STRATEGIC
30	MIZRUCHI, MS	



Cite network

V. Batagelj,
D. Maltseva

Data

Prob flow table

Citation
network
analysis