# Social Network Analysis:

## Bibliographic Network Analysis of the Field and its Evolution

Daria Maltseva, Vladimir Batagelj

IMFM Ljubljana, IAM UP Koper and NRU HSE Moscow

**Sredin Seminar 1282 / 1283**
November 21 / 28, 2018

# Outline

# Inroduction





Social Network Analysis (SNA) has moved from a fragmented direction represented by the works of individual scientific groups unrelated to each other, to a discipline whose representatives by 1990 have formed an "invisible college" and achieved the status of what Kuhn had labeled a "normal science" [Freeman, 2004; Hummon and Carley, 1993].

Starting from that time, the field has grown significantly, which can be seen by the number of scientific publications [Otte and Rousseau, 2002] in different scientific fields, including Natural Sciences, which lead to the so called "physicists' invasion" into SNA [Batagelj et al., 2014] and resulted with the development of Network Science discipline.

This calls into a question whether the field remains unified and which scientific groups (by disciplines, thematic agenda, etc.) it is currently formed of. Thus, the aim of the current study is to trace the evolution of the field of Social Network Analysis using bibliographic approach.

# Previous studies



1. Hummon N.P., Doreian P., Freeman L.C. (1990). Analyzing the Structure of the Centrality-Productivity Literature Created Between 1948 and 1979 / Science Communication. 11, 4, 459 – 480.
2. Freeman, L. (2004). The development of social network analysis. A Study in the Sociology of Science, 1.
3. Hummon, N. P., Carley, K. (1993). Social networks as normal science. Social networks, 15(1), 71-106.
4. Otte, E., Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. Journal of information Science, 28(6), 441-453.
5. Borgatti, S. P., Foster, P. C. (2003). The network paradigm in organizational research: A review and typology. Journal of management, 29(6), 991-1013.
6. Leydesdorff L., Schank T., Scharnhorst A., De Nooy W. (2008). Animating the development of Social Networks over time using a dynamic extension of multidimensional scaling / El Profesional de Informacion, 17(6).
7. Lazer, D., Mergel, I., and Friedman, A. (2009). Co-citation of prominent social network articles in sociology journals: The evolving canon. Connections, 29(1):43:64.
8. Brandes, U., Pich, C. (2011). Explorative visualization of citation patterns in social network research. Journal of social structure, 12(8), 1-19.
9. Varga, A. V., Nemeslaki, A. (2012) Do organizational network studies constitute a cohesive communicative field? Mapping the citation context of organizational network research. Journal of Sociology and Social Anthropology, 5(64), XV, 349-364.
10. Batagelj, V., Doreian P., V., Ferligoj, A., Kejžar N. (2014). Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution.
11. Groenewegen, P., Hellsten, I., Leydesdorff, L. (2015) Social Networks as a looking glass on the social networks community. International Sunbelt XXXV Conference. Hilton Metropole, Brighton, UK, June 23 – 28, 2015. Abstracts, 118.

# WoS

To the Web of Science (WoS), 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,
Social Network Analysis and Mining,
Journal of Complex Networks
```

Other network-related journals *are not presented* in WoS:

```
Computational Social Networks, Applied Network Science,
Online Social Networks and Media, Connections,
Journal of Social Structure
```

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.

# WoS record

```
PT J
AU GRANOVET.MS
TI STRENGTH OF WEAK TIES
SO AMERICAN JOURNAL OF SOCIOLOGY
LA English
DT Article
C1 JOHNS HOPKINS UNIV, BALTIMORE, MD 21218 USA.
CR BARNES JA, 1969, SOCIAL NETWORKS URBA
   BECKER MH, 1970, AM SOCIOL REV, V35, P267
   BERSCHEID E, 1969, INTERPERSONAL ATTRAC
   BOISSEVAIN J, 1968, MAN, V3, P542
   BOTT E, 1957, FAMILY SOCIAL NETWOR
NR 61
TC 2156
PU UNIV CHICAGO PRESS
PI CHICAGO
PA 5720 S WOODLAWN AVE, CHICAGO, IL 60637
SN 0002-9602
J9 AMER J SOCIOL
JI Am. J. Sociol.
PY 1973
VL 78
IS 6
BP 1360
EP 1380
PG 21
SC Sociology
GA P7726
UT ISI:A1973P772600003
ER
SK IP
```

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

As the data set SNA5 till 2007 was already completed, we made the additional search only for works 2008-* in July 2018.

# Types of networks and partitions

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. the 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*.

# ISI names

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

`GRANOVETTER M, 1985, AM J SOCIOL, V91, P481`

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: `GRANOVET_M(1985)91:481`
From the last names with prefixes `VAN`, `DE`, ...the space is deleted.
Unusual names start with character $*$ or \$.

# Equivalent works, authors and journals

```
BOYD_D(2007)13 | BOYD_D(2008)13:210
GRANOVET_M(1973)78:1360 | GRANOVET_M(1973)78:6
COLEMAN_J(1988)94:95 | COLEMAN_J(1988)94:S95

GRANOVET_M
GRANOVET_

63656 1312696 10849 SONEANMI | SOCIAL NETWORK ANAL
63657 1330776 3 SONEANMI | SOCIAL NETWORKS ANAL
63658 1311789 645 SONEANMI | SOC NETW ANAL MIN
63659 1313366 7 SONEANMI | SOCIAL NETW ANAL MIN
63660 1315722 7 SONEANMI | SOC NETW ANAL MINING
...
55366 1351847 54714 1 PSPOSC  | PS POLITICAL SCIENCE
55768 1320199 23066 5 POSC | POLITICAL SCI
55769 1320573 23440 3 POSC | POLIT SCI
56082 1297982 849 224 PSSCPO | PS-POLIT SCI POLIT
56083 1298064 931 110 PSSCPO | PS-POLITICAL SCI POL
```

# Equivalent works, authors and journals



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.

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 $p = worksEQ$, we shrink using Pajek the Citation network **Cite, WA, WJ**, and **WK**. We had to shrink also partitions *year*, *DC* and the vector *NP*.

The total cleaning of **journals** would require too much time. We decided to clean only the high frequency elements (cited 200 times and more) and consider bugs in low frequency elements as a noise. We got the list *journalK*100*.csv* with 3714 titles for inspection.

We also produced a list of frequent journal names (of length at most 5), have chosen all the cases that could be regarded as abbreviations (*CACM, JACM, JASA*), and manually searched for them.

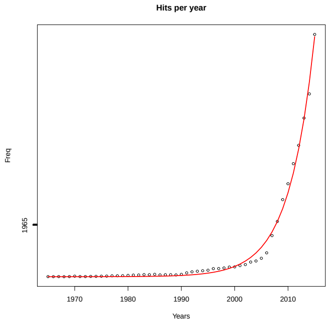
# Sizes of Basic and Reduced networks

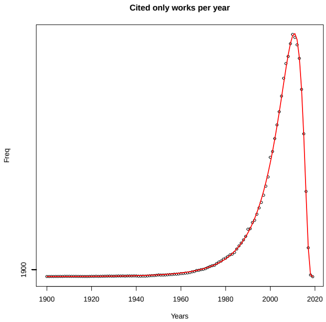
| | # nodes (sum) | # nodes 1 | # nodes 2 | # arcs |
|---|---|---|---|---|
| CiteN | 1,297,133 | | | 2,753,633 |
| **CiteR** | **70,792** | | | 398,199 |
| WAn | 1,693,104 | 1,297,133 | 395,971 | 1,442,240 |
| **WAr** | 163,803 | **70,792** | **93,011** | 215,901 |
| WKn | 1,329,542 | 1,297,133 | 32,409 | 1,167,670 |
| **WKr** | 103,201 | **70,792** | **32,409** | 1,167,666 |
| WJn | 1,366,279 | 1,297,133 | 69,146 | 720,044 |
| **WJr** | 79,735 | **70,792** | **8,943** | 61,741 |

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

# CiteN network

## Distribution of works by years




The distribution of hits fits to the *exponential model*.

$c \cdot a^{year-1965}$, where $a = 1.2338$, and $c = 0.2526$.

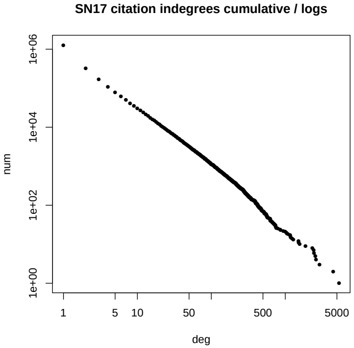
The obtained values shows that the amount of works almost dounles in each 3 years ($\log(2)/\log(1.2338) = 3.299148$).

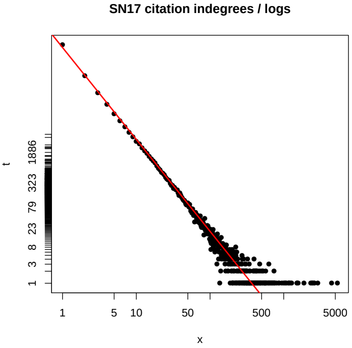
The distributions of cited only works fits the *log normal* distribution:

$c \cdot \text{dlnorm}(2018 - year, a, b)$, where $a = 1.501$, $b = 0.9587$, and $c = 7.11010^4$.

# CiteN network

## Indegree distribution




$fx = cx^{-\alpha}$

Fitted $\alpha = 2.3007$, $c = 749338$.

# Cite net

## The most cited works - indegree

| 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 | MCPHERSO_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 |

Labels ending with : represent books, **bold** for social science, * for physics

# CiteN net

## The most citing work - outdegree

| 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.

# WAr net

## Authors with the largest number of papers - indegree

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.

| Rank | Orig rank | Author | Value | Rank | Orig rank | Author | Value |
|---|---|---|---|---|---|---|---|
| 1 | 45 | LATKIN_C | 130 | 26 | 211 | SCHNEIDE_J | 52 |
| 2 | 72 | VALENTE_T | 97 | 27 | 212 | LEYDESDO_L | 51 |
| 3 | 84 | DUNBAR_R | 91 | 28 | 217 | LITWIN_H | 50 |
| 4 | 102 | NEWMAN_M | 81 | 29 | 228 | RICE_E | 48 |
| 5 | 121 | CHRISTAK_N | 74 | 30 | 232 | KAWACHI_I | 47 |
| 6 | 126 | DOREIAN_P | 72 | 31 | 233 | BONACICH_P | 46 |
| 7 | 127 | CARLEY_K | 72 | 32 | 234 | PARK_Y | 46 |
| 8 | 129 | BURT_R | 71 | 33 | 237 | RODRIGUE_M | 46 |
| 9 | 130 | BORGATTI_S | 71 | 34 | 238 | NGUYEN_H | 46 |
| 10 | 139 | SNIJDERS_T | 67 | 35 | 239 | CROFT_D | 46 |
| 11 | 140 | BARABASI_A | 67 | 36 | 249 | EVERETT_M | 44 |
| 12 | 146 | FOWLER_J | 65 | 37 | 252 | FERNANDE_M | 44 |
| 13 | 149 | KAZIENKO_P | 64 | 38 | 255 | CONTI_M | 44 |
| 14 | 150 | ROBINS_G | 64 | 39 | 256 | MORRIS_M | 43 |
| 15 | 152 | WELLMAN_B | 63 | 40 | 259 | CONTRACT_N | 43 |
| 16 | 163 | FALOUTSO_C | 60 | 41 | 266 | WHITE_H | 42 |
| 17 | 167 | RAHMAN_M | 59 | 42 | 267 | SKVORETZ_J | 42 |
| 18 | 172 | PATTISON_P | 58 | 43 | 275 | PENTLAND_A | 41 |
| 19 | 176 | TUCKER_J | 58 | 44 | 276 | WILLIAMS_M | 41 |
| 20 | 181 | HOSSAIN_L | 56 | 45 | 280 | MOODY_J | 40 |
| 21 | 187 | JOHNSON_J | 54 | 46 | 289 | FRIEDMAN_S | 40 |
| 22 | 194 | NGUYEN_T | 54 | 47 | 290 | MARSDEN_P | 39 |
| 23 | 196 | MARTINEZ_M | 53 | 48 | 292 | BERKMAN_L | 39 |
| 24 | 207 | GONZALEZ_M | 52 | 49 | 301 | KRACKHAR_D | 38 |
| 25 | 209 | RODRIGUE_J | 52 | 50 | 306 | MORENO_M | 38 |

Authors with Chinese names are omitted. Orig rank – original rank of an author in indegree distribution.

# WAr net

## Number of authors in works - outdegree

| Outdeg | Value | Freq% | Outdeg | Value | Freq% |
|---|---|---|---|---|---|
| 0 | 44 | 0,062 | 19 | 6 | 0,008 |
| 1 | 13157 | 18,585 | 20 | 2 | 0,003 |
| 2 | 18635 | 26,324 | 21 | 4 | 0,006 |
| 3 | 16661 | 23,535 | 22 | 3 | 0,004 |
| 4 | 10617 | 14,997 | 23 | 4 | 0,006 |
| 5 | 5759 | 8,135 | 24 | 2 | 0,003 |
| 6 | 2802 | 3,958 | 25 | 1 | 0,001 |
| 7 | 1322 | 1,867 | 26 | 2 | 0,003 |
| 8 | 686 | 0,969 | 27 | 5 | 0,007 |
| 9 | 384 | 0,542 | 28 | 2 | 0,003 |
| 10 | 247 | 0,349 | 29 | 1 | 0,001 |
| 11 | 155 | 0,219 | 31 | 3 | 0,004 |
| 12 | 90 | 0,127 | 36 | 1 | 0,001 |
| 13 | 70 | 0,099 | 41 | 1 | 0,001 |
| 14 | 54 | 0,076 | 42 | 1 | 0,001 |
| 15 | 32 | 0,045 | 43 | 1 | 0,001 |
| 16 | 12 | 0,017 | 48 | 1 | 0,001 |
| 17 | 14 | 0,020 | 53 | 1 | 0,001 |
| 18 | 9 | 0,013 | 126 | 1 | 0,001 |
| Sum | | | | 70792 | 100,000 |

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 |

# WAr net

## Works with the largest number of authors - outdegree

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 Kleigrewe, 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 Engene, 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, Wenyuan 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.

# WJn and WJr nets

## The most used journals - indegree

| WJn - Journals used in all publications | | | WJr - Journals used by hits | | |
|---|---|---|---|---|---|
| Rank | Value | Id | Rank | Value | Id |
| 1 | 7757 | LECT NOTES COMPUT SC | 1 | 2009 | LECT NOTES COMPUT SC |
| 2 | 3866 | SOC SCI MED | 2 | 1134 | ***SOC NETWORKS*** |
| 3 | 3414 | **J PERS SOC PSYCHOL** | 3 | 806 | COMPUT HUM BEHAV |
| 4 | 2741 | P NATL ACAD SCI USA | 4 | 667 | PLOS ONE |
| 5 | 2734 | COMPUT HUM BEHAV | 5 | 531 | LECT NOTES ARTIF INT |
| 6 | 2631 | SCIENCE | 6 | 470 | PHYSICA A |
| 7 | 2609 | AM J PUBLIC HEALTH | 7 | 399 | COMM COM INF SC |
| 8 | 2208 | NATURE | 8 | 375 | SOC SCI MED |
| 9 | 2111 | **AM SOCIOL REV** | 9 | 319 | **PROCD SOC BEHV** |
| 10 | 1945 | PHYSICA A | 10 | 314 | PHYS REV E |
| 11 | 1825 | ANIM BEHAV | 11 | 283 | PROCEDIA COMPUTER SCIENCE |
| 12 | 1812 | **AM J SOCIOL** | 12 | 273 | **SOC NETW ANAL MIN** |
| 13 | 1780 | JAMA-J AM MED ASSOC | 13 | 238 | ADV INTELL SYST |
| 14 | 1763 | LANCET | 14 | 231 | SCIENTOMETRICS |
| 15 | 1759 | SCIENTOMETRICS | 15 | 225 | CYBERPSYCHOL BEHAV |
| 16 | 1703 | **ACAD MANAGE J** | 16 | 216 | EDULEARN PROC |
| 17 | 1668 | LECT NOTES ARTIF INT | 17 | 215 | GERONTOLOGIST |
| 18 | 1642 | ***SOC NETWORKS*** | 18 | 198 | INTED PROC |
| 19 | 1573 | **J APPL PSYCHOL** | 19 | 194 | SCI REP-UK |
| 20 | 1517 | AM ECON REV | 20 | 188 | J MED INTERNET RES |
| 21 | 1450 | **J MARRIAGE FAM** | 21 | 186 | P NATL ACAD SCI USA |
| 22 | 1441 | EXPERT SYST APPL | 22 | 180 | EXPERT SYST APPL |
| 23 | 1403 | BRIT MED J | 23 | 176 | INFORM SCI |
| 24 | 1399 | CHILD DEV | 24 | 170 | BMC PUBLIC HEALTH |
| 25 | 1379 | **RES POLICY** | 25 | 167 | **NEW MEDIA SOC** |
| 26 | 1372 | COMMUN ACM | 26 | 160 | IEEE T KNOWL DATA EN |
| 27 | 1365 | NEW ENGL J MED | 27 | 153 | IEEE ACCESS |
| 28 | 1311 | PHYS REV E | 28 | 145 | AIDS BEHAV |
| 29 | 1287 | **SOC FORCES** | 29 | 140 | INFORM COMMUN SOC |
| 30 | 1279 | GERONTOLOGIST | 30 | 139 | STUD COMPUT INTELL |

# WKn net

## The most used keywords - indegree

Keywords were extracted also from titles. Phrases were split to component words. Stop words were removed. Words were lemmatized.

| 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 |

# Cite net

## Boundary problem

The network CiteN has 1,297,133 nodes and 2,753,767 arcs.

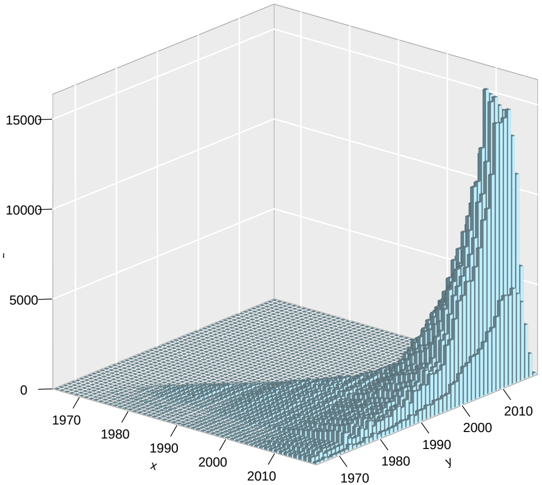
$(DC = 1) = 70{,}792 \; (DC = 0) = 1{,}226{,}341$

| indeg | 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 ($DC = 0$) or nodes cited only once (indegree=1). We decided (*boundary problem*) to include in our networks nodes with $DC > 0$ or indeg $> 2$. They determine a subnetwork **CiteB** with 222,086 nodes and 1,521,434 arcs.

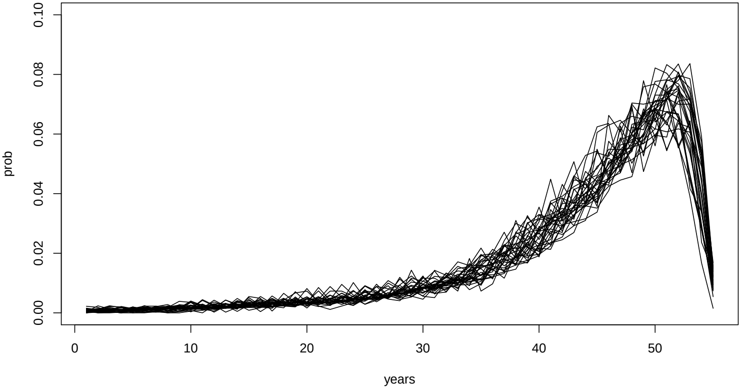
# CiteB net

## Citations among years



# CiteB net

## Citations among years – normalized



# CiteB net

## Components

The citation network CiteB has 41 nontrivial strong components. 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 network arcs, and determined

- CPM path / Main path = 59 nodes
- Key-routes = 127 nodes
- SPC link islands [weights] of sizes [10, 200] = 5 islands of 138, 65, 13, 12, and 11 nodes
- SPC node islands [weights] of szes [10, 200] = 1 island of 200 nodes

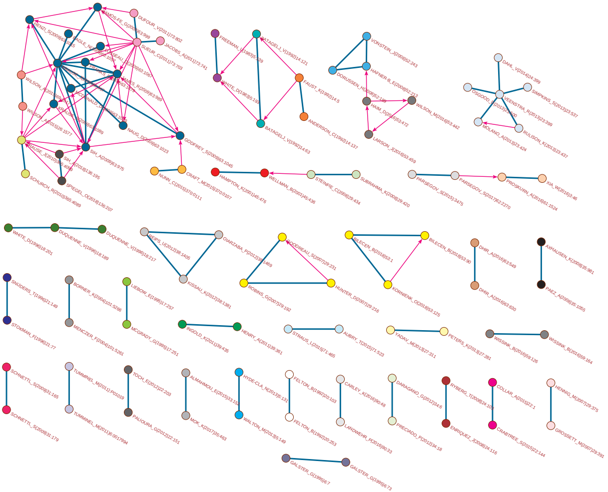
We computed the Probabilistic flow, and determined 200 nodes with the largest values.

# Strong components

## from SPC network

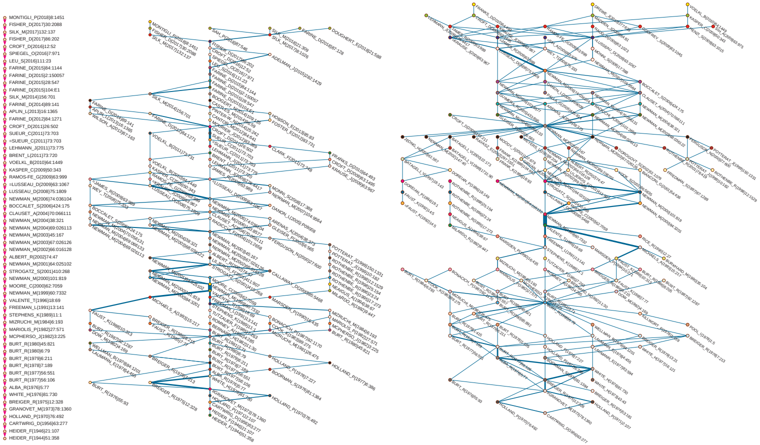






# Main path, Key Routes, and Island 4

## from SPC network



# Main path

## from SPC network

1. 1944–1996: 20 works from the field of Social Science in journals *Social networks, Administrative Science Quarterly, Annual Review of Sociology, American Sociological Review, Social Forces, Sociological Methods & Research, Journal of Mathematical Psychology, Psychological Review, The Journal of Psychology*
2. 1999-2007: 14 works from Physics in journals *Physical Review E, Journal of Statistical Physics, Reviews of Modern Physics, European Physical Journal B, Physics Reports, Nature*, and *SIAM Review*
3. 2008-2018: 25 works from Animal social networks in journals *Animal Behaviour, American Journal of Primatology, Primates, Journal of Evolutionary Biology, Journal of Animal Ecology, Journal of Evolutionary Biology, Trends in Ecology & Evolution*

# Key-routes

## from SPC network



**Third period (2008–2018)**: 49 works of behavioural ecologists.

1. Krause, James (2009): *general works* on animal social network analysis
2. Ramos-Fernandez, Kasper, Voell, Lehmann, Brent, Sueur (2009–2011): *social networks of Nonhuman Primates* (monkeys, baboons).
3. Croft (2011): practical guide on *hypothesis testing* in animal social networks
4. 2014: research on *mixed-species groups* (Farine), *killer whales* (Foster), *sharks* (Mourier), *dolphins* (Cantor), published in 2012, and *birds* (Silk) and *starlings* (Boogert)
5. 2013-2014: *methodological issues* – Hobson (*An analytical framework for quantifying and testing patterns of temporal dynamics in social networks*), Castels (*Social networks created with different techniques are not comparable*), Pinter-Wollman (*The dynamics of animal social networks: analytical, conceptual, and theoretical advances*)

# Key-routes

## from SPC network

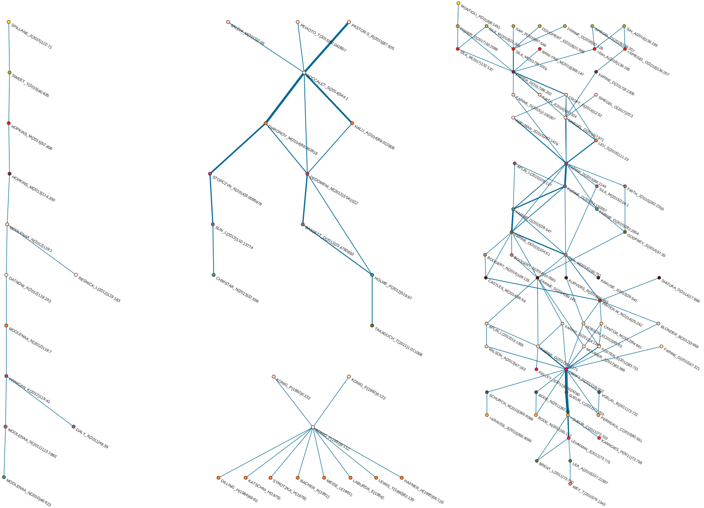
1. Farine (2015): *methodological issues on constructing, conducting and interpreting animal social network analysis*, study of the *wild birds territory acquisition*.
2. 2013-2014: *social personality and phenotypic types* (Wilson, Alpin, Farine).
3. 2016–2018: studies on *desease transmission* (Adelman, Sah, Silk, Dougherty), *animal paths tracking* (Leu, Spiegel); works on *theoretical issues* (*Current directions in animal social networks* by Croft, *Social traits, social networks and evolutionary biology* by Fisher); *implementation of different models of network analysis to animal behaviour research*: exponential random graph models and statistical network models (Silk), the potential of stochastic actor-oriented models (Fisher), dynamic vs. static social network analysis (Farine).

# Islands 1-3, 5

## from SPC network



1 - Education, 2 - Physics, 3 - Neuropsychiatry, 5 - Animal SNA



# Most important works

## from Probabilistic Flow network



The values were multiplied by 1,000,000.

| Rank | Value | Id | Rank | Value | Id |
|---|---|---|---|---|---|
| 1 | 4691 | WASSERMA_S(1994): | 31 | 545 | BLONDEL_V(2008):P10008 |
| 2 | 2941 | WATTS_D(1998)393:440 | 32 | 527 | KATZ_L(1953)18:39 |
| 3 | 2676 | GRANOVET_M(1973)78:1360 | 33 | 526 | NEWMAN_M(2010): |
| 4 | 2445 | BOYD_D(2007)13:210 | 34 | 520 | STROGATZ_S(2001)410:268 |
| 5 | 2241 | BARABASI_A(1999)286:509 | 35 | 517 | PALLA_G(2005)435:814 |
| 6 | 1926 | FREEMAN_L(1979)1:215 | 36 | 499 | CLAUSET_A(2004)70:066111 |
| 7 | 1396 | GIRVAN_M(2002)99:7821 | 37 | 497 | ERDOS_P(1960)5:17 |
| 8 | 1299 | NEWMAN_M(2003)45:167 | 38 | 488 | ROGERS_E(2003): |
| 9 | 1227 | MCPHERSO_M(2001)27:415 | 39 | 485 | NEWMAN_M(2006)103:8577 |
| 10 | 1158 | ALBERT_R(2002)74:47 | 40 | 481 | COLEMAN_J(1990): |
| 11 | 1105 | SCOTT_J(2000): | 41 | 478 | BRIN_S(1998)30:107 |
| 12 | 1098 | BURT_R(1992): | 42 | 477 | AMARAL_L(2000)97:11149 |
| 13 | 1045 | MILGRAM_S(1967)1:61 | 43 | 475 | ERDOS_P(1959)6:290 |
| 14 | 1013 | NEWMAN_M(2004)69:026113 | 44 | 465 | WATTS_D(1999): |
| 15 | 928 | KAPLAN_A(2010)53:59 | 45 | 462 | LAVE_J(1991): |
| 16 | 878 | FREEMAN_L(1977)40:35 | 46 | 460 | KLEINBER_J(1999)46:604 |
| 17 | 852 | PUTNAM_R(2000): | 47 | 449 | SCOTT_J(1991): |
| 18 | 847 | COLEMAN_J(1988)94:95 | 48 | 446 | BOLLOBAS_B(1985): |
| 19 | 835 | * BLEI_D(2003)3:993 | 49 | 442 | PAGE_L(1999): |
| 20 | 742 | GRANOVET_M(1985)91:481 | 50 | 440 | NEWMAN_M(2001)64:025102 |
| 21 | 731 | CHRISTAK_N(2007)357:370 | 51 | 436 | NEWMAN_M(2004)69:066133 |
| 22 | 727 | EVERETT_M(2002): | 52 | 431 | REDNER_S(1998)4:131 |
| 23 | 726 | NEWMAN_M(2001)98:404 | 53 | 429 | CHRISTAK_N(2008)358:2249 |
| 24 | 719 | * ALBERT_R(1999)401:130 | 54 | 424 | ADOMAVIC_G(2005)17:734 |
| 25 | 701 | * O'REILLY_T(2005): | 55 | 424 | KEMP_D(2003):137 |
| 26 | 669 | BORGATTI_S(2002): | 56 | 423 | DOMINGOS_P(2001):57 |
| 27 | 667 | FORTUNAT_S(2010)486:75 | 57 | 423 | MITCHELL_J(1969): |
| 28 | 633 | HANNEMAN_R(2005): | 58 | 415 | ALBERT_R(2000)406:378 |
| 29 | 569 | STEINFIE_C(2007)12:1143 | 59 | 415 | GLASER_B(1967): |
| 30 | 549 | ZACHARY_W(1977)33:452 | 60 | 410 | ROGERS_E(1995): |

# Cite net

## Overlapping of important parts

SNA. Bibliographic Network Analysis

D. Maltseva, V. Batagelj



| i | name | title | jour | comp |
|---|---|---|---|---|
| 1 | Granovet M | Strength of weak ties | amer j sociol | 1, 2, 4, 5, 6 |
| 2 | Newman M | The structure and function of complex networks | siam rev | 1, 2, 4, 5, 6 |
| 3 | Albert R | Statistical mechanics of complex networks | rev mod phys | 1, 2, 4, 5, 6 |
| 4 | Boccaletti S | Complex networks: structure and dynamics | phys rept | 1, 2, 4, 5, 6 |
| 5 | White H | Soc. str. from mult. nets. Blockmodels | amer j sociol | 1, 2, 4, 5, 6 |
| 6 | Newman M | Clustering and pref.l attach. in growing nets | phys rev e | 1, 2, 4, 5, 6 |
| 7 | Newman M | Finding and evaluating comm. struct. in nets | phys rev e | 1, 2, 4, 5, 6 |
| 8 | Newman M | Mixing patterns in networks | phys rev e | 1, 2, 4, 5, 6 |
| 9 | Strogatz S | Exploring complex networks | nature | 1, 2, 4, 5, 6 |
| 10 | Newman M | Detecting community structure in nets | eur phys j b | 1, 2, 4, 5, 6 |
| 11 | Newman M | Spread of epidemic disease on nets | phys rev e | 1, 2, 4, 5, 6 |
| 12 | Newman M | Finding community str. in nets using eigenvectors | phys rev e | 1, 2, 4, 5, 6 |
| 13 | Cartwright D | Structural balance - a generaliz. of heider theory | psychol rev | 1, 2, 4, 5, 6 |
| 14 | Clauset A | Finding community struct. in very large nets | phys rev e | 1, 2, 4, 5, 6 |
| 15 | Newman M | Models of the small world | j statist phys | 1, 2, 4, 5 |
| 16 | Newman M | Scaling and percolation in small-world net model | phys rev e | 1, 2, 4, 5 |
| 17 | Valente T | Social net thresholds in the diff. of innov. | soc networks | 1, 2, 4, 5 |
| 18 | Burt R | Cohesion versus structural equivalences as a basis for net subgroups | soc meth res | 1, 2, 4, 5 |
| 19 | Stephenson K | Rethinking centrality - methods and examples | soc networks | 1, 2, 4, 5 |
| 20 | Breiger R | Algorithm for clustering relational data | j math psychl | 1, 2, 4, 5 |
| 21 | Freeman L | Centrality in valued graphs - a measure of betweenness based on net flow | soc networks | 1, 2, 4, 5 |
| 22 | Burt R | Models of network structure | annu rev soc | 1, 2, 4, 5 |
| 23 | Holland P | Method for detecting structure in sociom. data | amer j sociol | 1, 2, 4, 5 |
| 24 | Alba R | Intersection of social circles | socl meth res | 1, 2, 4, 5 |
| 25 | Moore C | Exact solution of site and bond percolation on small-world net | phys rev e | 1, 2, 4, 5 |
| 26 | Mcpherson J | Hypernetwork sampling - duality and differentiation among voluntary organizations | soc networks | 1, 2, 4, 5 |
| 27 | Mariolis P | Centrality in corporate interlock networks | adm sci quart | 1, 2, 4, 5 |
| 28 | Burt R | Positions in multiple network systems 1. General conception of stratification and prestige | soc forces | 1, 2, 4, 5 |
| 29 | Burt R | Positions in multiple network systems 2. Stratification and prestige among elite | soc forces | 1, 2, 4, 5 |
| 30 | Mizruchi M | Interlock groups, cliques, or interest-groups | soc networks | 1, 2, 4, 5 |

1- Key Routes, 2- Main Path (CPM), 3- Island5, 4 - Island 4, Node Island, 5 - Prob Flow Island

# Authors Collaboration

## Network Co

$$\mathbf{Co} = \mathbf{WAr}^T * \mathbf{WAr}$$

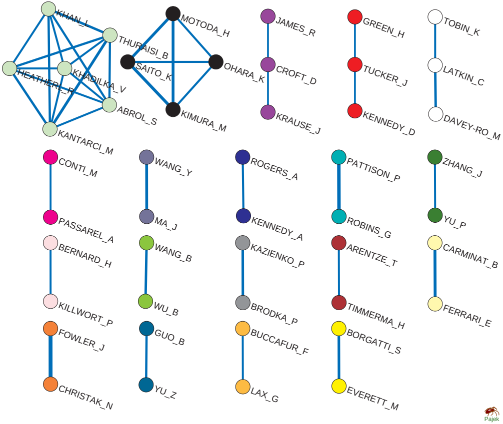
The weight of the edges between the nodes $i$ and $j$ is equal to total number of works author i and j wrote together.

The loops in *Co* are equal to the total number of works that each author have (which is also equal to the indegree values of the *WA* network).

The proposed approach has some *limitations*, such as the overrating of the contribution of works with many authors.

# Authors Collaboration

## **Co** net (20 and more works written together)



# Authors Collaboration

## Network **Cn**

The contribution of authors to their own works and works written with co-authors is considered. The normalization create network $n(\mathbf{WA})$ where the weight of each arc is divided by the sum of weights of all arcs having the same initial node as this arc (outdegree or outsum of a node).

$$n(\mathbf{WAr})[w,a] = \frac{\mathbf{WAr}[w,a]}{\max(1, \mathrm{outdeg}[w])}$$

then

$$\mathbf{Cn} = \mathbf{WAr}^T * n(\mathbf{WAr})$$

The weight of the edges between the nodes (authors) is equal to the contribution of author i to works, that he or she wrote together with author j (can be not symmetric).
The total contribution for an author is equal to the number of works that he or she co-authored (indegree of **WA**).
The diagonal (loops) of the matrix is equal to the total contribution of author to his or her own works.

Batagelj and Cerinšek (2013):

- **self-sufficiency index Si** as a proportion of author's contribution to all her/his works and his/her total number of works (which is equal to the value of indegree of **WA** network):

$$S_i = \frac{cn_{ii}}{\mathrm{indeg}_{\mathbf{WA}}(i)}$$

- **collaborativness index Ki**, which is complementary to it (is equal to 1 minus self-sufficiency):

$$K_i = 1 - S_i$$

# Authors Collaboration

## Collaborativeness index from **Cn** net (from **WAr**)

| # | Author | Tot Contr | Tot #works | Collab | # | Author | Tot Contr | Tot #works | Collab |
|---|---|---|---|---|---|---|---|---|---|
| 1 | BURT_R | 55,73 | 71 | 0,22 | 31 | **PATTISON_P** | 18,94 | 58 | 0,67 |
| 2 | NEWMAN_M | 50,02 | *81 | 0,38 | 32 | **THELWALL_M** | 18,41 | 37 | 0,50 |
| 3 | DOREIAN_P | 46,19 | 72 | 0,36 | 33 | **KRACKHAR_D** | 18,24 | 38 | 0,52 |
| 4 | **PARK_H** | 41,94 | *113 | 0,63 | 34 | **FALOUTSO_C** | 17,86 | 60 | 0,70 |
| 5 | **DUNBAR_R** | 40,02 | *91 | 0,56 | 35 | **JACKSON_M** | 17,78 | 38 | 0,53 |
| 6 | WELLMAN_B | 36,43 | 63 | 0,42 | 36 | **GONZALEZ_M** | 17,76 | 52 | 0,66 |
| 7 | **VALENTE_T** | 34,96 | *97 | 0,64 | 37 | **MOODY_J** | 17,7 | 40 | 0,56 |
| 8 | **PARK_S** | 34,59 | *109 | 0,68 | 38 | SCOTT_J | 17,54 | 28 | 0,37 |
| 9 | BONACICH_P | 34 | 46 | 0,26 | 39 | **MORRIS_M** | 17,22 | 43 | 0,60 |
| 10 | LEYDESDO_L | 33,28 | 51 | 0,35 | 40 | **RODRIGUE_J** | 15,9 | 52 | 0,69 |
| 11 | **LATKIN_C** | 32,99 | *130 | 0,75 | 41 | **WASSERMA_S** | 15,64 | 35 | 0,55 |
| 12 | LITWIN_H | 32,42 | 50 | 0,35 | 42 | **KLEINBER_J** | 15,05 | 34 | 0,56 |
| 13 | MARSDEN_P | 30,17 | 39 | 0,23 | 43 | **BATAGELJ_V** | 14,64 | 33 | 0,56 |
| 14 | **BORGATTI_S** | 29,72 | 71 | 0,58 | 44 | **WILLIAMS_A** | 14,5 | 31 | 0,53 |
| 15 | **SNIJDERS_T** | 29,63 | 67 | 0,56 | 45 | **SINGH_A** | 14,5 | 36 | 0,60 |
| 16 | FRIEDKIN_N | 28,17 | 36 | 0,22 | 46 | **BRANDES_U** | 14,39 | 35 | 0,59 |
| 17 | **CARLEY_K** | 28,11 | 72 | 0,61 | 47 | **BERKMAN_L** | 14,3 | 39 | 0,63 |
| 18 | **BARABASI_A** | 27,61 | 67 | 0,59 | 48 | MASUDA_N | 14,26 | 28 | 0,49 |
| 19 | WHITE_H | 27,28 | 42 | 0,35 | 49 | **SMITH_A** | 14,2 | 40 | 0,65 |
| 20 | **CHRISTAK_N** | 22,89 | 74 | 0,69 | 50 | LAZEGA_E | 14,17 | 26 | 0,46 |
| 21 | EVERETT_M | 22,58 | 44 | 0,49 | 51 | **CONTRACT_N** | 14,15 | 43 | 0,67 |
| 22 | **KAZIENKO_P** | 21,97 | 64 | 0,66 | 52 | **GONZALEZ_A** | 14,13 | 35 | 0,60 |
| 23 | **MARTINEZ_M** | 21,9 | 53 | 0,59 | 53 | **PENTLAND_A** | 14,12 | 41 | 0,66 |
| 24 | **JOHNSON_J** | 21,19 | 54 | 0,61 | 54 | **FARINE_D** | 14,04 | 34 | 0,59 |
| 25 | **FOWLER_J** | 20,14 | 65 | 0,69 | 55 | **SCHNEIDE_J** | 13,89 | 52 | 0,73 |
| 26 | **SKVORETZ_J** | 20,07 | 42 | 0,52 | 56 | WATTS_D | 13,67 | 27 | 0,49 |
| 27 | FREEMAN_L | 20,03 | 27 | 0,26 | 57 | FAUST_K | 13,5 | 25 | 0,46 |
| 28 | BREIGER_R | 19,73 | 31 | 0,36 | 58 | **SMITH_M** | 13,29 | 39 | 0,66 |
| 29 | **ROBINS_G** | 19,67 | 64 | 0,69 | 59 | **RODRIGUE_M** | 13,21 | 46 | 0,71 |
| 30 | **RAHMAN_M** | 19,18 | 59 | 0,67 | 60 | **RICE_E** | 13,09 | 48 | 0,73 |

# Authors Collaboration

## Network **Ct'**

Newman's normalization: "strict collaboration" – the weight is a proportion of time spent for the collaboration with each co-author. The weight of each arc is divided by the sum of weights of all arcs having the same initial node as this arc (outdegree of a node) subtracting the initial author.

$$n'(\mathbf{WAr})[w,a] = \frac{\mathbf{WAr}[w,a]}{\max(1, \mathrm{outdeg}(w) - 1)}$$

then

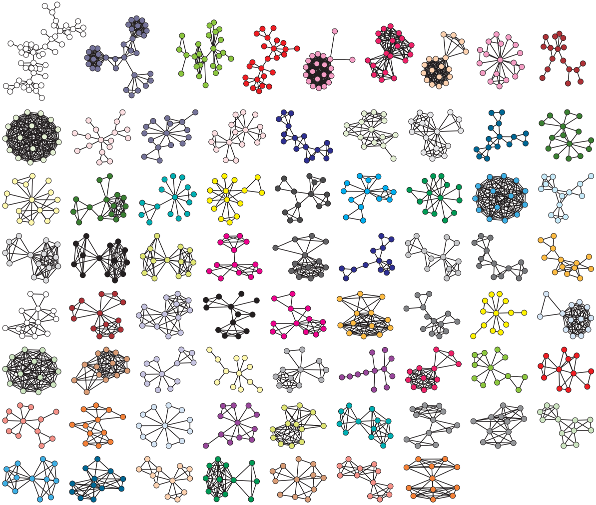
$$\mathbf{Ct}' = n(\mathbf{WAr})^T * n'(\mathbf{WAr})$$

The final $Ct'$ is undirected without loops. The contribution of a complete subgraph corresponding to each work is 1.
The weights of the edges between the nodes (authors) are equal to the total contribution of "strict collaboration" of authors i and j to works they wrote together.
The total contribution for an author is counted by line weights – it is equal to the sum of the weights of all the works he or she co-authored.

# Authors Collaboration

## Main simple islands from **Ct'** net (from **WAr**)



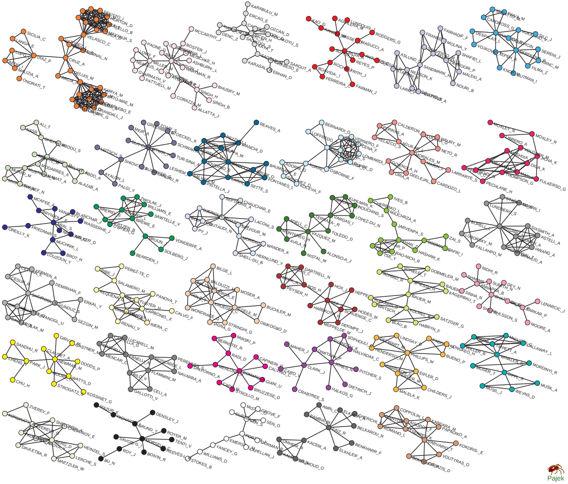


# Authors Collaboration

## Selected islands from **Ct'** net

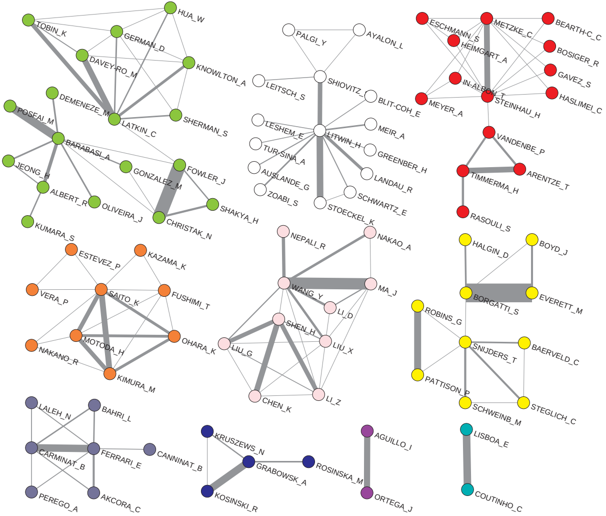






# Authors Collaboration

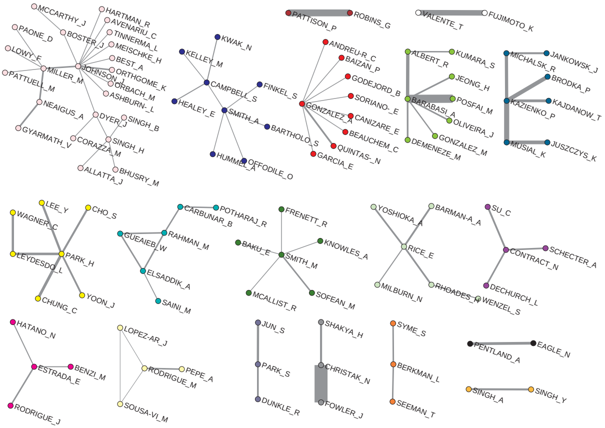
## Authors with largest link weights from **Ct'** net



# Authors Collaboration

## Selected authors from **Ct'** net





$$\mathbf{AK} = n(\mathbf{WAr})^T * n(\mathbf{WKr})$$

The weight of the edges between the nodes $a$ and $k$ is equal to the fractional contribution for a given keyword $k$ to the works of author $a$ or a group of authors $C$.

$$\mathbf{AK}[\mathbf{C}, \mathbf{k}] = \sum_{a \in C} \mathbf{AK}[a, k]$$

# Key words in coauthorship islands

## **AK** net

| | BORGATTI_S | | BARABASI_A | | CHRISTAKIS_K | |
|---|---|---|---|---|---|---|
| Rank | Value | Id | Value | Id | Value | Id |
| 1 | 4.9303 | network | 7.0709 | network | 3.1788 | network |
| 2 | 2.5918 | social | 2.0782 | social | 2.9358 | social |
| 3 | 2.0858 | graph | 1.7068 | dynamics | 1.0204 | spread |
| 4 | 1.4210 | centrality | 1.6670 | complex | 1.0192 | behavior |
| 5 | 1.4202 | analysis | 1.6362 | scale | 0.7261 | health |
| 6 | 1.3399 | role | 1.5946 | web | 0.5512 | large |
| 7 | 1.2780 | regular | 1.5516 | community | 0.5169 | model |
| 8 | 1.2424 | equivalence | 1.4709 | world | 0.4778 | smoking |
| 9 | 1.0530 | semigroup | 1.3622 | internet | 0.4522 | human |
| 10 | 1.0000 | correction | 1.1906 | model | 0.4479 | cooperation |
| 11 | 0.9891 | structure | 1.1858 | free | 0.4313 | obesity |
| 12 | 0.7755 | clique | 1.0210 | evolve | 0.4125 | influence |
| 13 | 0.7576 | homomorphism | 1.0087 | science | 0.3973 | life |
| 14 | 0.7241 | relation | 0.9808 | random | 0.3728 | dynamics |
| 15 | 0.6346 | power | 0.9476 | wide | 0.3715 | evolution |
| 16 | 0.6301 | betweenness | 0.8178 | human | 0.3463 | analysis |
| 17 | 0.6287 | exchange | 0.8076 | theory | 0.3286 | cosponsorship |
| 18 | 0.6232 | algorithm | 0.7561 | small | 0.3044 | norm |
| 19 | 0.6167 | similarity | 0.7536 | graph | 0.3036 | trial |
| 20 | 0.5595 | ebloc | 0.6603 | phenomenon | 0.2985 | study |
| Total: | 63.0810 | | 76.6373 | | 46.8865 | |

# Key words in coauthorship islands

## from **AK**

| | PATTISON_P | | SNIJDERS_T | | VALENTE_T | |
|---|---|---|---|---|---|---|
| Rank | Value | Id | Value | Id | Value | Id |
| 1 | 2.2196 | network | 2.6375 | network | 2.5536 | network |
| 2 | 2.0729 | social | 2.0902 | social | 1.9553 | social |
| 3 | 1.7567 | model | 1.6702 | model | 1.0000 | untitled |
| 4 | 1.3084 | graph | 1.0692 | graph | 0.9419 | health |
| 5 | 0.8939 | random | 0.8857 | dynamics | 0.8737 | diffusion |
| 6 | 0.8583 | markov | 0.7390 | markov | 0.7802 | behavior |
| 7 | 0.8531 | logit | 0.6903 | random | 0.7402 | innovation |
| 8 | 0.8220 | logistic | 0.6734 | friendship | 0.6974 | model |
| 9 | 0.8220 | regression | 0.6228 | datum | 0.6521 | use |
| 10 | 0.8012 | exponential | 0.5932 | statistical | 0.6349 | peer |
| 11 | 0.7055 | analysis | 0.5780 | behavior | 0.6216 | adolescent |
| 12 | 0.6752 | p | 0.5547 | analysis | 0.5717 | influence |
| 13 | 0.5530 | statistical | 0.5423 | peer | 0.5610 | smoking |
| 14 | 0.5038 | structure | 0.5383 | inference | 0.5371 | analysis |
| 15 | 0.3561 | semigroup | 0.5346 | influence | 0.5247 | prevention |
| 16 | 0.3522 | asterisk | 0.4623 | stochastic | 0.4987 | cigarette |
| 17 | 0.3368 | process | 0.4612 | actor | 0.4979 | opinion |
| 18 | 0.3333 | multirelational | 0.4480 | selection | 0.4860 | leader |
| 19 | 0.3249 | family | 0.4372 | longitudinal | 0.4545 | risk |
| 20 | 0.3031 | dynamics | 0.3785 | orient | 0.4491 | intervention |
| Total: | 38.6110 | | 46.6732 | | 44.8812 | |

## CiteA and CiteAn nets



$$\mathbf{CiteA} = (\mathbf{WAr})^T * \mathbf{CiteR} * \mathbf{WAr}$$

The value of the element **CiteA[u,v]** is equal to the **number of citations** from works coauthored by $u$ to works coauthored by $v$.

$$\mathbf{CiteAn} = (\mathbf{WAr})^T * n(\mathbf{CiteR}) * \mathbf{WAr}$$

where

$$n(\mathbf{CiteR})[u,v] = \frac{\mathbf{CiteR}[u,v]}{\max(1, \mathrm{outdeg}(u))}$$

The value of element **CiteAn[u,v]** is equal to the sum of **fractional contribution** of citations from works coauthored by $u$ to works coauthored by $v$.

# Citation among authors

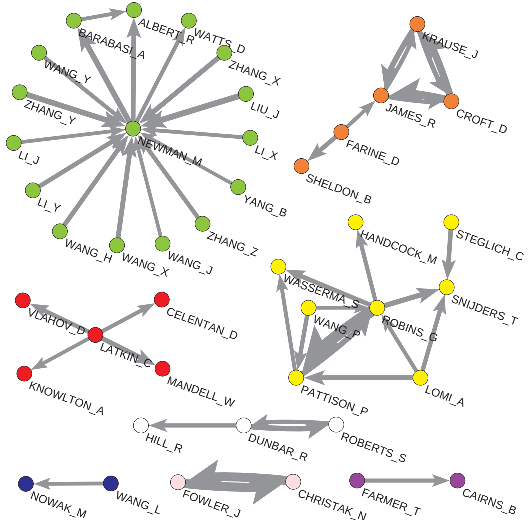
## self-citation from **CiteA** net

| Rank | Id | Value | Rank | Id | Value |
|---|---|---|---|---|---|
| 1 | DUNBAR_R | 589 | 11 | BARABASI_A | 201 |
| 2 | LATKIN_C | 387 | 12 | FARINE_D | 191 |
| 3 | CHRISTAK_N | 292 | 13 | SNIJDERS_T | 188 |
| 4 | VALENTE_T | 280 | 14 | WELLMAN_B | 153 |
| 5 | BURT_R | 268 | 15 | DOREIAN_P | 148 |
| 6 | NEWMAN_M | 248 | 16 | BORGATTI_S | 146 |
| 7 | ROBINS_G | 232 | 17 | ZENOU_Y | 146 |
| 8 | PATTISON_P | 224 | 18 | RICE_E | 143 |
| 9 | FOWLER_J | 221 | 19 | JAMES_R | 142 |
| 10 | CROFT_D | 204 | 20 | KRAUSE_J | 141 |

# Citation among authors

## Authors with largest line weights from **CiteA** net





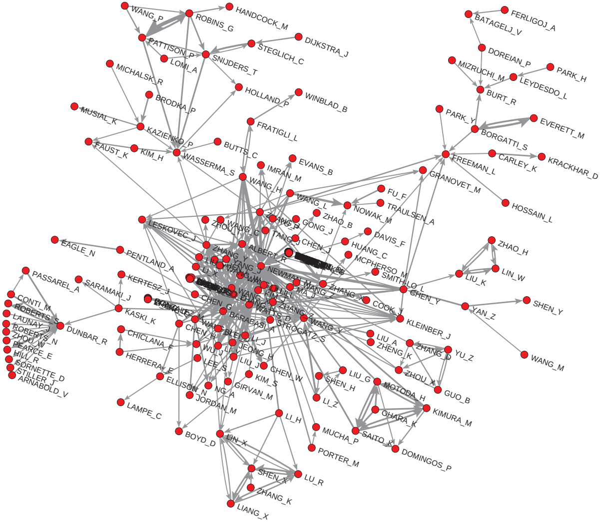
# Citation among authors

## self-citation from **CiteAn** net

| # | Author | Value | % | # | Author | Value | % |
|---|---|---|---|---|---|---|---|
| 1 | TANG_J | 9.786 | 0.025 | 26 | FALOUTSO_C | 8.169 | 0.014 |
| 2 | FARINE_D | 9.778 | 0.074 | 27 | HANSON_B | 8.091 | 0.041 |
| 3 | PENTLAND_A | 9.746 | 0.016 | 28 | SUEUR_C | 8.025 | 0.092 |
| 4 | MARATHE_M | 9.528 | 0.034 | 29 | FRANK_K | 7.995 | 0.062 |
| 5 | ZENOU_Y | 9.466 | 0.077 | 30 | LI_X | 7.966 | 0.020 |
| 6 | EVERETT_M | 9.342 | 0.012 | 31 | MORENO_M | 7.724 | 0.034 |
| 7 | KRAUSE_J | 9.035 | 0.022 | 32 | THELWALL_M | 7.698 | 0.033 |
| 8 | CHEN_H | 8.975 | 0.032 | 33 | SHEN_X | 7.679 | 0.039 |
| 9 | BERKMAN_L | 8.949 | 0.007 | 34 | KENNEDY_A | 7.619 | 0.070 |
| 10 | POTTERAT_J | 8.899 | 0.027 | 35 | GARLAND_S | 7.606 | 0.087 |
| 11 | MORRIS_M | 8.861 | 0.025 | 36 | ZHANG_D | 7.586 | 0.041 |
| 12 | KAZIENKO_P | 8.802 | 0.067 | 37 | NOWAK_M | 7.554 | 0.021 |
| 13 | SHEN_H | 8.799 | 0.049 | 38 | MAGLIANO_L | 7.542 | 0.051 |
| 14 | LIU_J | 8.763 | 0.034 | 39 | BONACICH_P | 7.540 | 0.020 |
| 15 | XU_Q | 8.667 | 0.085 | 40 | LU_R | 7.458 | 0.045 |
| 16 | TUCKER_J | 8.496 | 0.061 | 41 | WANG_J | 7.414 | 0.023 |
| 17 | SKVORETZ_J | 8.481 | 0.056 | 42 | WANG_L | 7.345 | 0.038 |
| 18 | THAI_M | 8.453 | 0.068 | 43 | SAITO_K | 7.335 | 0.055 |
| 19 | BATAGELJ_V | 8.421 | 0.032 | 44 | CHEN_W | 7.245 | 0.012 |
| 20 | MUTH_S | 8.382 | 0.026 | 45 | FERRARI_E | 7.209 | 0.038 |
| 21 | MARTINEZ_M | 8.313 | **0.141** | 46 | COHEN_S | 7.202 | 0.010 |
| 22 | LITWIN_H | 8.297 | 0.052 | 47 | RYAN_L | 7.193 | 0.080 |
| 23 | STANTON_N | 8.250 | **0.216** | 48 | MEYBODI_M | 7.175 | **0.314** |
| 24 | TUREL_O | 8.227 | **0.161** | 49 | KIMURA_M | 7.139 | 0.053 |
| 25 | ABDELZAH_T | 8.176 | 0.085 | 50 | KIM_H | 7.121 | 0.037 |

# Citation among authors

## Main island from **CiteAn** net



# Citation among journals

## **CiteJ** and **CiteJn** nets

Networks take into account citations from papers published in journal *i* to papers published in journal *j*, which appeared in the works included into the **WJr** network. In the obtained network, the value of the element **CiteJ[u,v]** is equal to the **number of citations** from journal *i* to journal *j*.

$$\mathbf{CiteJ} = (\mathbf{WJr})^T * \mathbf{CiteR} * \mathbf{WJr}$$

Line weights in **CiteJn** take into account *fractional* contribution of citations from papers published in journal *i* to papers published in journal *j*.

$$\mathbf{CiteJn} = (\mathbf{WJr})^T * n(\mathbf{CiteR}) * \mathbf{WJr}$$

Computation was done with networks **WJr** and **CiteR**.

# Citation among journals

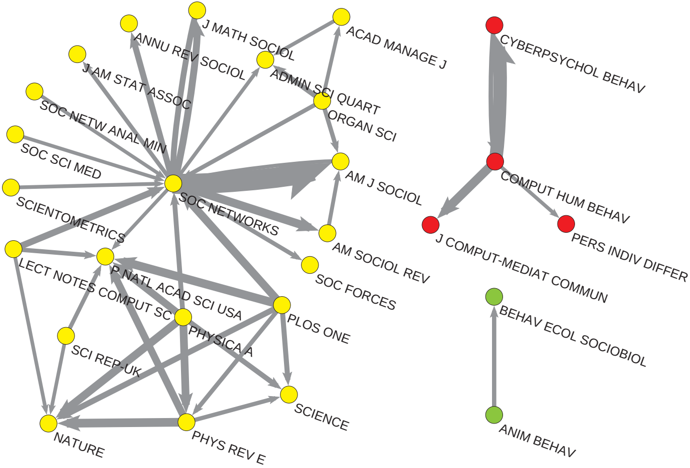
## self-citation from **CiteJ** net

| Rank | Value | Id |
|---|---|---|
| 1 | 4443 | SOC NETWORKS |
| 2 | 2058 | COMPUT HUM BEHAV |
| 3 | 569 | PHYSICA A |
| 4 | 429 | PHYS REV E |
| 5 | 382 | LECT NOTES COMPUT SC |
| 6 | 339 | CYBERPSYCHOL BEHAV |
| 7 | 328 | SOC SCI MED |
| 8 | 315 | AM J SOCIOL |
| 9 | 303 | PLOS ONE |
| 10 | 258 | ANIM BEHAV |
| 11 | 246 | SCIENTOMETRICS |
| 12 | 232 | J MED INTERNET RES |
| 13 | 226 | P NATL ACAD SCI USA |
| 14 | 209 | ORGAN SCI |
| 15 | 194 | BEHAV ECOL SOCIOBIOL |

# Citation among journals

## journals with the largest line weights from **CiteJ** net





# Citation among journals

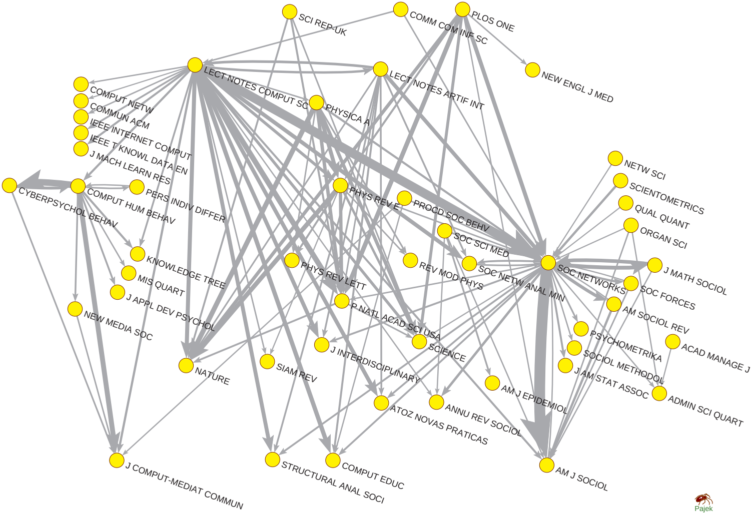
## self-citation from **CiteJn** net

| # | Value | % | Journal | # | Value | % | Journal |
|---|---|---|---|---|---|---|---|
| 1 | **355.65** | **0.34** | SOC NETWORKS | 16 | 18.35 | 0.17 | ANIM BEHAV |
| 2 | **168.39** | 0.22 | COMPUT HUM BEHAV | 17 | 17.03 | 0.12 | AIDS BEHAV |
| 3 | **122.57** | 0.09 | LECT NOTES COMPUT SC | 18 | 16.03 | 0.19 | AM J COMMUN PSYCHO |
| 4 | 57.75 | 0.13 | PHYSICA A | 19 | 14.87 | 0.10 | INFORM SCI |
| 5 | 43.00 | 0.14 | SOC SCI MED | 20 | 14.14 | 0.14 | KNOWL-BASED SYST |
| 6 | 42.18 | **0.24** | J MED INTERNET RES | 21 | 12.64 | 0.19 | PROF INFORM |
| 7 | 41.49 | 0.21 | CYBERPSYCHOL BEHAV | 22 | 12.35 | **0.23** | COMUNICAR |
| 8 | 33.16 | 0.05 | PLOS ONE | 23 | 12.00 | 0.18 | BEHAV ECOL SOCIOBI |
| 9 | 32.93 | 0.11 | PHYS REV E | 24 | 11.87 | **0.25** | AM J EPIDEMIOL |
| 10 | 30.22 | 0.13 | SCIENTOMETRICS | 25 | 11.01 | 0.11 | DECIS SUPPORT SYST |
| 11 | 24.16 | 0.14 | P NATL ACAD SCI USA | 26 | 10.58 | 0.14 | J ETHN MIGR STUD |
| 12 | 23.15 | **0.26** | AM J SOCIOL | 27 | 10.43 | 0.13 | COMPUT EDUC |
| 13 | 20.04 | 0.05 | LECT NOTES ARTIF INT | 28 | 10.31 | 0.18 | SEX TRANSM DIS |
| 14 | 19.31 | 0.12 | EXPERT SYST APPL | 29 | 10.19 | **0.28** | NATURE |
| 15 | 18.77 | 0.14 | NEW MEDIA SOC | 30 | 9.85 | 0.09 | ORGAN SCI |

# Citation among journals

## Main island from **CiteJn** net





# Citation among journals

## Pairs of journals with largest line weights from **CiteJn** net

| # | value | from journal | to journal |
|---|---|---|---|
| 1 | 8,1 | IEEE GLOB COMM CONF | IEEE INFOCOM SER |
| 2 | 6,26 | HIST COMUN SOC | COMUNICAR |
| 3 | 4,63 | J YOUTH ADOLESCENCE | J RES ADOLESCENCE |
| 4 | 4,44 | INT J GERIATR PSYCH | J PSYCHIAT RES |
| 5 | 4,38 | INT MIGR | INT MIGR REV |
| 6 | 4,31 | J BUS ETHICS | ACAD MANAGE REV |
| 7 | 3,99 | DEMOGR RES | DEMOGRAPHY |
| 8 | 3 | J INTELL FUZZY SYST | J APPL MATHE COMPUT |
| 9 | 3 | J INT DEV | TROP MED INT HEALTH |
| 10 | 3 | PERVASIVE MOB COMPUT | INT CONF PERVAS COMP |
| 11 | 2,78 | J CONSTR ENG M | J CONSTR ENG M ASCE |
| 12 | 2,68 | PHYS EDUC RES CONF | PHYS REV SPEC TOP-PH |
| 13 | 2,59 | ENERGY RES SOC SCI | ENERG POLICY |
| 14 | 2,5 | INT P ECON DEV RES | TECHNOVATION |
| 15 | 2,37 | COMPUT ASSIST LANG L | LANG LEARN TECHNOL |
| 16 | 2,33 | INFORM SOC-ESTUD | PERSPECT CIENC INF |
| 17 | 2,33 | WORLD DEV | ECON J |
| 18 | 2,31 | J PEACE RES | J CONFLICT RESOLUT |
| 19 | 2,22 | HEALTH RES POLICY SY | HEALTH POLICY PLANN |
| 20 | 2,1 | SEX HEALTH | INT J STD AIDS |
| 21 | 2 | REV LAT COMUN SOC | PALABRA CLAVE |
| 22 | 2 | J RETAIL CONSUM SERV | AUSTRALAS MARK J |
| 23 | 2 | ETHN DIS | HEART LUNG |
| 24 | 2 | IEEE INT SYMP INFO | IEEE T INFORM THEORY |
| 25 | 2 | REV BRAS ENFERM | REV LAT-AM ENFERM |

# Bibliographic coupling

## Jaccard **biCo**, **JCoj** and **ACoj** nets

$$\mathbf{biCo} = \mathbf{CiteR} * (\mathbf{CiteR})^T$$

Normalization create networks $n(\mathbf{WJr})$ and $n(\mathbf{WAr})$ where the weight of each arc is divided by the sum of weights of all arcs having the same initial node (journal or author) as this arc (outdegree of a node).

Weights in the obtained networks take into account *fractional* similarity of journals $i$ and $j$, or authors $u$ and $v$.

$$\mathbf{JCoj} = n(\mathbf{WJr})^T * \mathbf{biCoj} * n(\mathbf{WJr})$$

$$\mathbf{ACoj} = n(\mathbf{WAr})^T * \mathbf{biCoj} * n(\mathbf{WAr})$$

The values of links from biCoj are redistributed in JCoj and ACoj.

$$\sum_{e \in E(\mathbf{JCoj})} \mathbf{JCoj}[e] = \sum_{e \in E(\mathbf{biCoj})} \mathbf{biCoj}[e]$$

The total sum of link weights is preserved.

# Bibliographic coupling

## Jaccard **biCo**, **JCoj** and **ACoj** nets

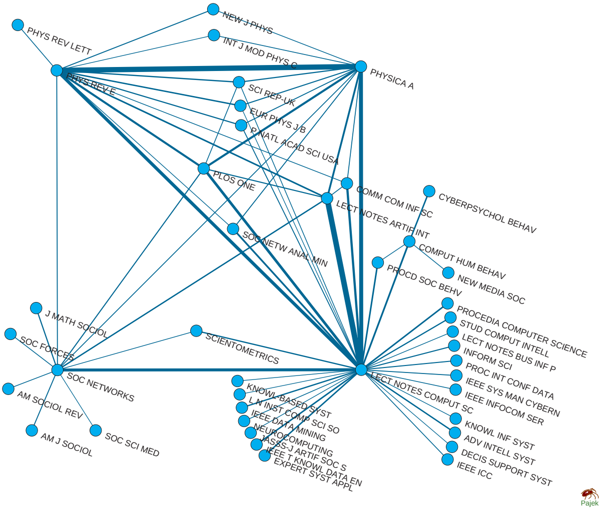
The produced Jaccard network **biCo** contained a large number of links – 62.079.457, – and the computation of networks **JCoj** and **ACoj** would be quite time consuming.

The desision was made to make a link cut in the Jaccard **biCo** at the level 0.085, and then use the obtained network for the multiplication with **WJsr** and **WAsr** networks, respectively.

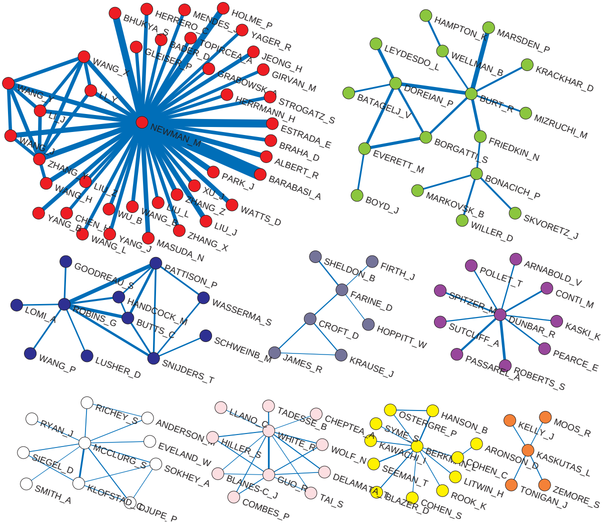
The obtained **biCo** network contained 70.792 works and 13.208.451 links. After multiplication, we got **JCoj** network with 8.943 nodes and 4.966.617 arcs, and **ACoj** network with 93.011 nodes and 127.220.243 arcs. In both obtained networks, the loops were deleted, the bidirected arcs were converted to edges (with summation of values) before the further analysis.

# Bibliographic coupling

## Co-citation among journals **JCoj**



## Co-citation among authors **ACoj**



# Keywords co-occurence

## KK net

$$\mathbf{KKn} = n(\mathbf{WK})^T * n(\mathbf{WK})$$

where

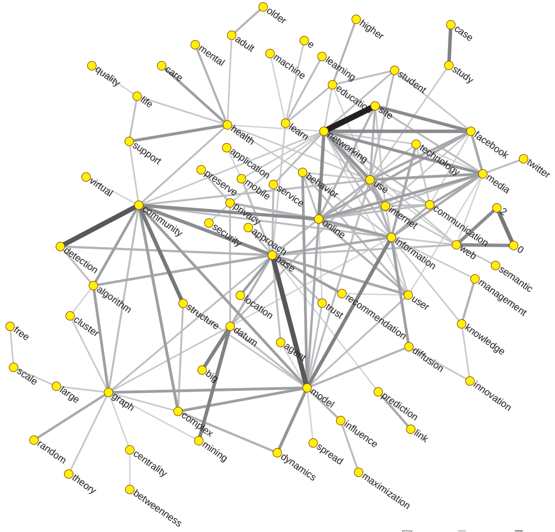
$$n(\mathbf{WK})[w,k] = \frac{\mathbf{WK}[w,k]}{\max(1, outdeg(w))}$$

Fractional approach:

The fractional contribution of each complete subgraph in **KKn** is 1 – all works are contributing equally (the works with large number of keywords are not overrepresented).
The weight of the edges between the nodes (keywords) is equal to the *fractional* co-occurence of keywords $i$ and $j$ in the same works.

# Keywords co-occurence

## **KK** net Main island



# Temporal networks



A *temporal network* $\mathcal{N}_T = (\mathcal{V}, \mathcal{L}, \mathcal{T}, \mathcal{P}, \mathcal{W})$ is obtained by attaching the *time*, $\mathcal{T}$, to an ordinary network where $\mathcal{T}$ is a set of *time points*, $t \in \mathcal{T}$.

In a temporal network, nodes $v \in \mathcal{V}$ and links $l \in \mathcal{L}$ are not necessarily present or active in all time points. Let $T(v)$, $T \in \mathcal{P}$, be the *activity set* of time points for node $v$ and $T(l)$, $T \in \mathcal{W}$, the activity set of time points for link $l$.

Besides the presence/absence of nodes and links also their properties can change through time.

# Temporal quantities

We introduce a notion of a *temporal quantity*

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

where $T_a$ is the *activity time set* of $a$ and $a'(t)$ is the value of $a$ in an instant $t \in T_a$, and ⌘ denotes the value *undefined*.

We assume that the values of temporal quantities belong to a set $A$ which is a *semiring* $(A, +, \cdot, 0, 1)$ for binary operations $+$ and $\cdot$ .

We can extend both operations to the set $A_{\text{⌘}} = A \cup \{\text{⌘}\}$ by requiring that for all $a \in A_{\text{⌘}}$ it holds

$$a + \text{⌘} = \text{⌘} + a = a \quad \text{and} \quad a \cdot \text{⌘} = \text{⌘} \cdot a = \text{⌘}.$$

The structure $(A_{\text{⌘}}, +, \cdot, \text{⌘}, 1)$ is also a semiring.

# Operations with temporal quantities

Let $A_{⌘}(\mathcal{T})$ denote the set of all temporal quantities over $A_{⌘}$ in time $\mathcal{T}$. To extend the operations to networks and their matrices we first define the *sum* (parallel links) $a + b$ as

$$(a + b)(t) = a(t) + b(t) \quad \text{and} \quad T_{a+b} = T_a \cup T_b.$$

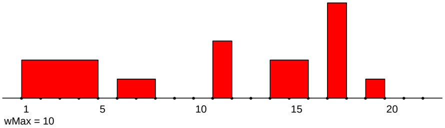
The *product* (sequential links) $a \cdot b$ is defined as

$$(a \cdot b)(t) = a(t) \cdot b(t) \quad \text{and} \quad T_{a \cdot b} = T_a \cap T_b.$$

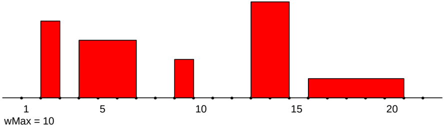
Let us define the temporal quantities $\mathbf{0}$ and $\mathbf{1}$ with requirements $\mathbf{0}(t) = ⌘$ and $\mathbf{1}(t) = 1$ for all $t \in \mathcal{T}$. Again, the structure $(A_{⌘}(\mathcal{T}), +, \cdot, \mathbf{0}, \mathbf{1})$ is a semiring.

# Addition of temporal quantities

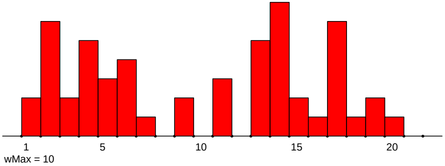
$a$ :



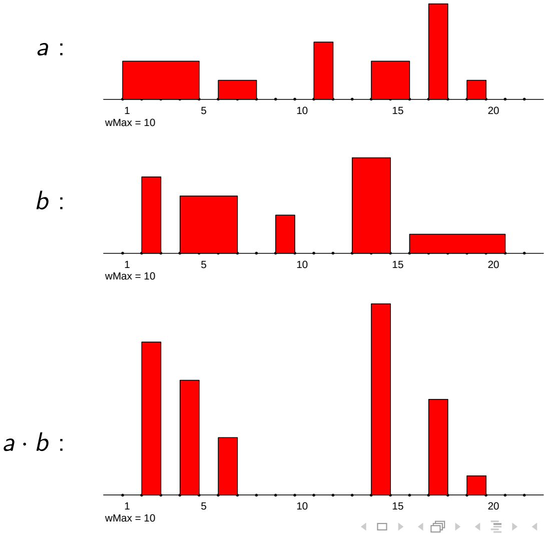

$b$ :




$a + b$ :



# Multiplication of temporal quantities



# Temporal co-occurrence networks



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

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

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

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

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

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

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

# Temporal networks

```
>>> net = wdir+"/WAins.json"
>>> WAi = N.loadNetJSON(net)
>>> I = WAi.Index()
>>> I["ERDOS_P(1959)6:290"]
776
>>> WAi._nodes[776]
[{}, {}, {71246: [1091], 89670: [214562]},
{'mode': 1, 'lab': 'ERDOS_P(1959)6:290', 'act': [[1959, 2019, 1]]}]
>>> WAi._nodes[71246]
[{}, {776: [1091], 11539: [25213], 11540: [25214], 15565: [33820],
42898: [129167]}, {},
{'mode': 2, 'lab': 'ERDOS_P', 'act': [[1894, 2019, 1]]}]
>>> WAi._nodes[89670]
[{}, {776: [214562], 15565: [215596]}, {},
{'mode': 2, 'lab': 'RENYI_A', 'act': [[1894, 2019, 1]]}]
>>> WAi._links[1091]
[776, 71246, True, None, {'tq': [[1959, 1960, 1]]}]
>>> WAi._links[214562]
[776, 89670, True, None, {'tq': [[1959, 1960, 1]]}]
```

# Multiplication of co-occurence networks

## Instantaneous

Instantaneous $\mathbf{A}$ on $P \times A$ and $\mathbf{B}$ on $P \times B$. $\mathbf{C} = \mathbf{A}^T.\mathbf{B}$ on $A \times B$.

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

$a_{pi} = [(d_{pi}, d_{pi} + 1, v_{pi})]$ and $b_{pj} = [(d_{pj}, d_{pj} + 1, v_{pj})]$
for $t = d$ we get

$$c_{ij} = [(d, d+1, \sum_{p \in P : d_{pi} = d_{pj} = d} v_{pi} . v_{pj})]_{d \in \mathcal{T}}$$

for $v_{pi} = v_{pj} = 1$ we finally get

$$v_{ij}(d) = |\{p \in P : d_{pi} = d_{pj} = d\}|$$

For binary temporal two-mode networks $\mathbf{A}$ and $\mathbf{B}$ the value $v_{ij}(d)$ of the product $\mathbf{A}^T.\mathbf{B}$ is equal to the number of different members of $P$ with which both $i$ and $j$ have contact in the instant $d$.

# Multiplication of co-occurence networks

## Cumulative

Cumulative $\mathbf{A}$ on $P \times A$ and $\mathbf{B}$ on $P \times B$. $\mathbf{C} = \mathbf{A}^T.\mathbf{B}$ on $A \times B$.

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

$a_{pi} = [(d_{pi}, last + 1, v_{pi})]$ and $b_{pj} = [(d_{pj}, last + 1, v_{pj})]$
for $t = d$ we get

$$c_{ij} = [(d, d+1, \sum_{p \in P:(d_{pi} \le d) \wedge (d_{pj} \le d)} v_{pi}.v_{pj})]_{d \in \mathcal{T}}$$

for $v_{pi} = v_{pj} = 1$ we finally get

$$v_{ij}(d) = |\{p \in P : (d_{pi} \le d) \wedge (d_{pj} \le d)\}|$$

For binary temporal two-mode networks $\mathbf{A}$ and $\mathbf{B}$ the value $v_{ij}(d)$ of the product $\mathbf{A}^T.\mathbf{B}$ is equal to the number of different members of $P$ with which both $i$ and $j$ have contact in all instants up to including the instant $d$.

# Temporal networks

Using the **Temporal quantities approach** [Batagelj and Praprotnik, 2016] and Python packages **Nets.py** and **TQ.py**, out of the reduced networks **WAr, WJr, WKr**, and **CiteR** we constructed the set of temporal networks of two types – *instantaneous* (with values given per each year) and *cumulative* (with cumulative values over the years) in `.json` format. These are the networks **WAins, WAcum, WJins, WJcum, WKrins, WKcum**, and **CiteIns, CiteCum**.

These temporal networks were used for construction of other **derived** temporal networks.

# Temporal networks

Example of node with temporal quantities - *instantaneous*:

['NEWMAN_M', [(1999, 2000, 2), (2000, 2001, 4), (2001, 2002, 9), (2002, 2003, 10), (2003, 2004, 9), (2004, 2005, 15), (2005, 2006, 7), (2006, 2007, 13), (2007, 2008, 4), (2008, 2009, 1), (2009, 2010, 2), (2010, 2011, 3), (2016, 2017, 2)]]

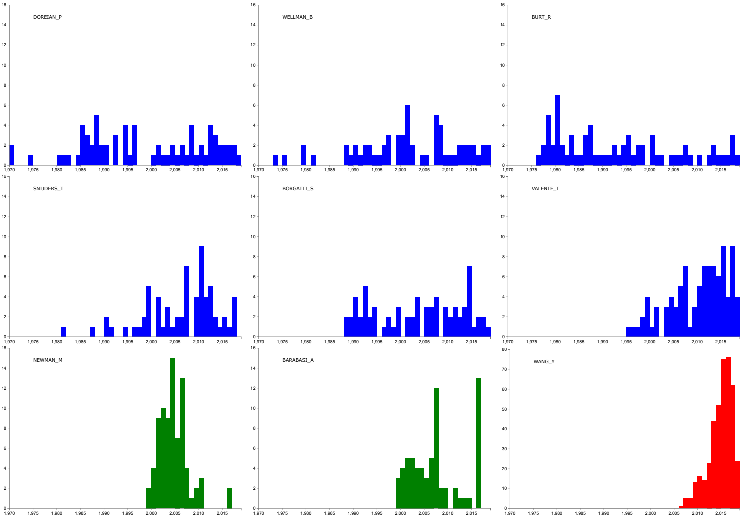
Example of node with temporal quantities - *cumulative*:

['NEWMAN_M', [(1999, 2000, 2), (2000, 2001, 6), (2001, 2002, 15), (2002, 2003, 25), (2003, 2004, 34), (2004, 2005, 49), (2005, 2006, 56), (2006, 2007, 69), (2007, 2008, 73), (2008, 2009, 74), (2009, 2010, 76), (2010, 2016, 79), (2016, 2019, 81)]]

# Number of publications for selected authors

## Indegree of WAins network





## Works of author *a* in period *time*



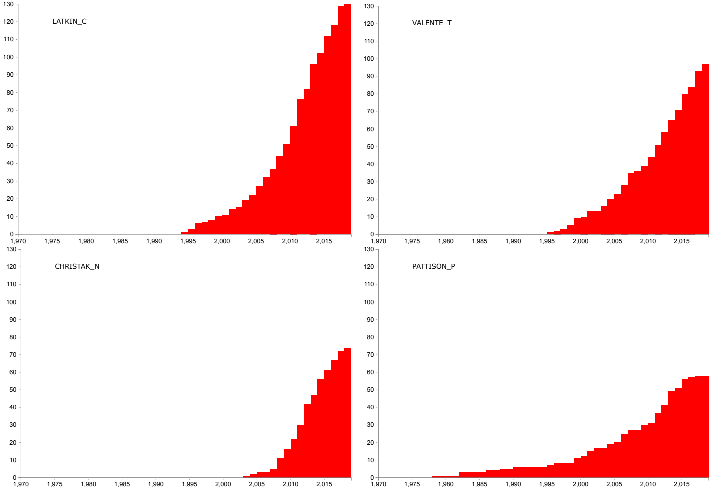
```
>>> def works(a,time):
   return [ WAi._nodes[WAi._links[p][0]][3]['lab']
          for p in WAi.inStar(a)
          if WAi._links[p][4]['tq'][0][0] in time ]

>>> works(I["BARABASI_A"],[2016])
['BARABASI_A(2016):1', 'DEVILLE_P(2016)113:7047', 'BARABASI_A(2016):VII',
'BARABASI_A(2016):21', 'BARABASI_A(2016):43', 'BARABASI_A(2016):73',
'BARABASI_A(2016):113', 'BARABASI_A(2016):165', 'BARABASI_A(2016):203',
'BARABASI_A(2016):233', 'BARABASI_A(2016):271', 'BARABASI_A(2016):321',
'BARABASI_A(2016):379']

>>> works(I["BARABASI_A"],range(2010,2019))
['BARABASI_A(2016):1', 'ONNELA_J(2011)6:0016939', 'BAGROW_J(2011)6:0017680',
'DEVILLE_P(2016)113:7047', 'WANG_P(2013)12:383', 'GAO_L(2014)4:srep03997',
'PALCHYKO_V(2012)2:srep00370', 'BARABASI_A(2016):VII', 'BARABASI_A(2016):21',
'BARABASI_A(2016):43', 'BARABASI_A(2016):73', 'BARABASI_A(2016):113',
'BARABASI_A(2016):165', 'BARABASI_A(2016):203', 'BARABASI_A(2016):233',
'BARABASI_A(2016):271', 'BARABASI_A(2016):321', 'BARABASI_A(2016):379']
```

Barabasi, Albert-Laszlo; Posfai, Marton (2016) Network Science. Cambridge University Press.

# Number of publications for selected authors

## Indegree from WAcum network





# Temporal co-authorship networks



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

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

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

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

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

The diagonal matrix entries $ci_{pp}$ and $cc_{pp}$ contain the temporal quantities counting the number of events in the time intervals in which the participant $p$ took part.

# Temporal co-authorship networks

The matrices

$$\mathbf{CoIns} = \mathbf{WAins}^T \cdot \mathbf{WAins} \quad \text{and} \quad \mathbf{CoCum} = \mathbf{WAcum}^T \cdot \mathbf{WAcum}$$

describe the instantaneous co-autorship temporal network and the cumulative co-autorship temporal network, counting the number of works two authors wrote together.

# Collaboration networks CCn and CCtt

## Constructed from WAins and WAcum networks

$$n(\mathbf{WAcum})[w,a] = \frac{\mathbf{WAcum}[w,a]}{\max(1, \mathrm{outdeg}(w))}$$

then

$$\mathbf{CCn} = (\mathbf{WAcum})^T * n(\mathbf{WAcum}),$$

The weight of the edges between the nodes (authors) is equal to the contribution of author $i$ to works, that he or she wrote together with author $j$.

$$n'(\mathbf{WAins})[w,a] = \frac{\mathbf{WAins}[w,a]}{\max(1, \mathrm{outdeg}(w) - 1)}$$

then

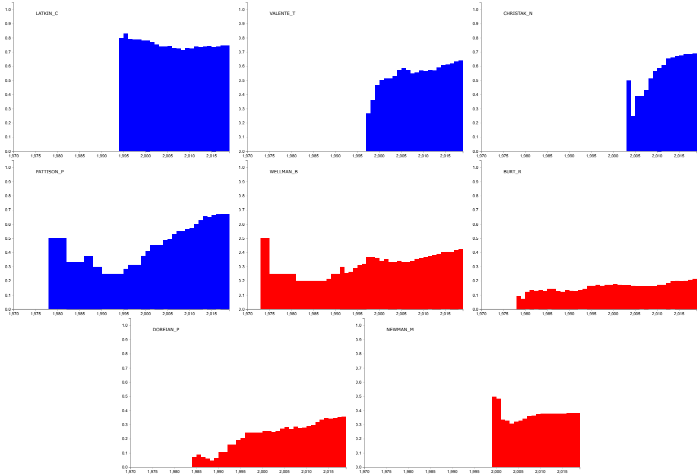
$$\mathbf{CCtt} = n(\mathbf{WAins})^T * n'(\mathbf{WAins}),$$

The weights of the edges between the nodes (authors) are equal to the total contribution of “strict collaboration” of authors $i$ and $j$ to works they wrote together.

# Temporal collaborativenness for selected authors

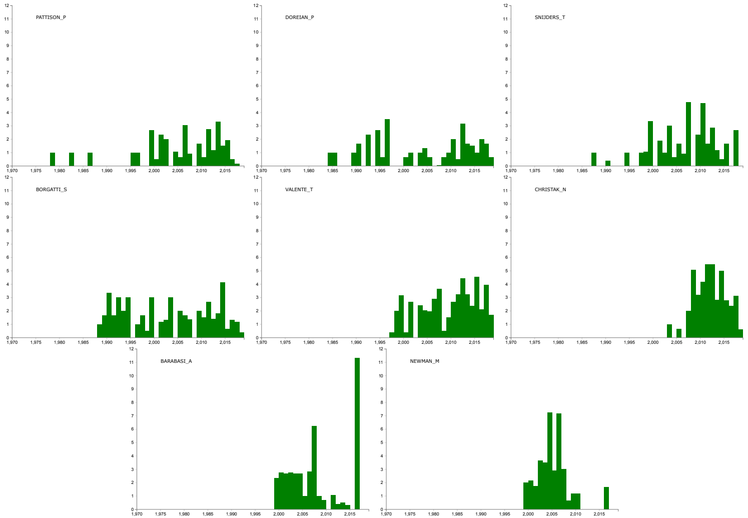
## from Collaboration network CCn





# Strict temporal collaboration for selected authors

## from Collaboration network CCtt



# Temporal citation networks





A citation matrix **Ci** describes the citation relation $p$ cites $q$. Note that $\quad p \text{ cites } q \Rightarrow d(p) \geq d(q)$.
Then we can construct its instantaneous version **Cii**:

$$cii_{pq} = [(d(p), d(p)+1, 1)] \quad \text{iff} \quad ci_{pq} = 1$$

and its cumulative version **Cic**:

$$cic_{pq} = [(d(p), last+1, 1)] \quad \text{iff} \quad ci_{pq} = 1$$

Temporal versions of:
Bibliographic coupling $\mathbf{biCo} = \mathbf{Ci} \cdot \mathbf{Ci}^T$.
Co-citation $\mathbf{coCi} = \mathbf{Ci}^T \cdot \mathbf{Ci}$.
Citations between authors $\mathbf{Ca} = \mathbf{WA}^T \cdot \mathbf{Ci} \cdot \mathbf{WA}$.

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

# Citations among authors

## ACA and ACAn nets from WAins, WAcum and CiteIns nets

$$\mathbf{ACA} = (\mathbf{WAins})^T * \mathbf{CiteIns} * \mathbf{WAcum}$$

The value of weight of the element $ACA[u, v]$ is equal to the **number of citations** per year from works coauthored by $u$ to works coauthored by $v$.

$$\mathbf{ACAn} = (\mathbf{WAins})^T * n(\mathbf{CiteIns}) * \mathbf{WAcum}$$

where

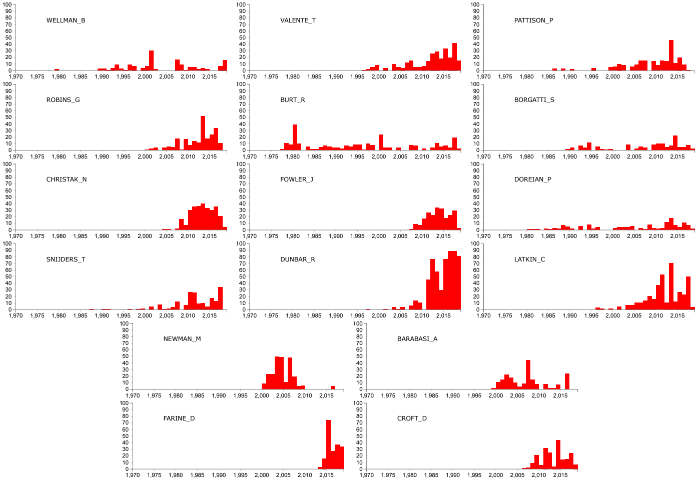
$$n(\mathbf{CiteIns})[u, v] = \frac{\mathbf{CiteIns}[u, v]}{\max(1, \mathrm{outdeg}(u))}$$

The value of element ACAn[u,v] is equal to the number of **fractional contribution** of citations per year from works coauthored by $u$ to works coauthored by $v$.

# Self-citations for selected authors

## Loops from ACA network



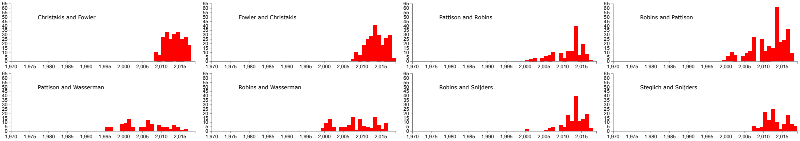


# Citations between authors with largest link weights

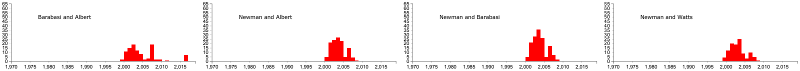
## from ACA network with extracted loops



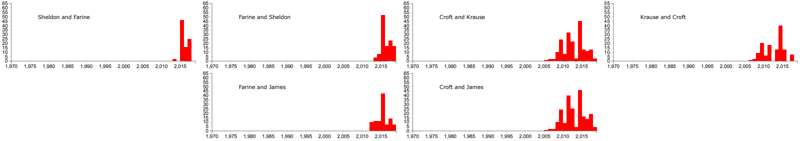
### Social Sciences




### Physics




### Animal social networks

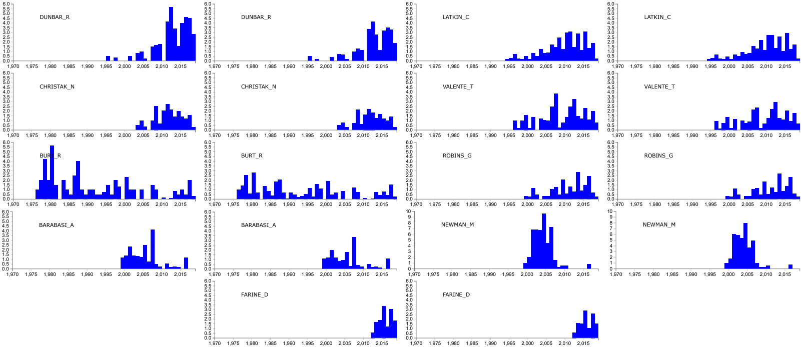


# Outgoing citations for selected authors

## from ACAn network



OutSum of ACAn: how much authors are cited by a selected author

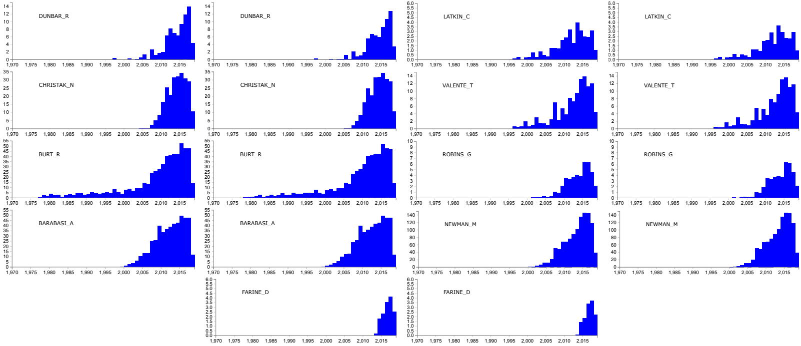


# Incoming citations for selected authors

## from ACAn network

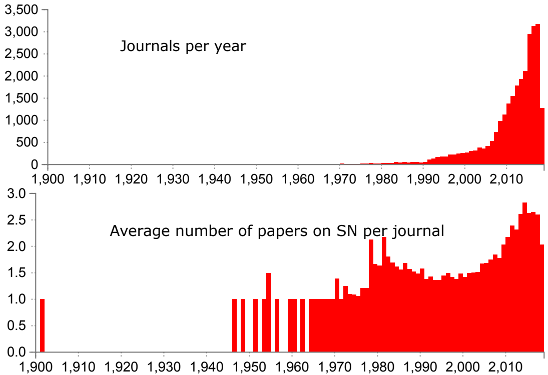


InSum of ACAn: how much authors cite selected authors (note different scales)



# Number of journals and papers

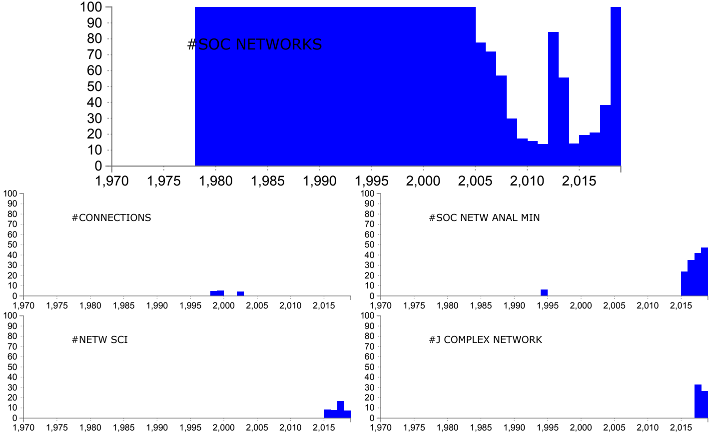
## from WJins network




We counted a proportion of the *Number of works* published in each journal in each year and the *Maximum number of papers* published per year (in any journal). This proportions aims to normalize the input of the selected journal in each year in range from 0 to 100% and shows their importance trough time.

# Importance of SNA journals

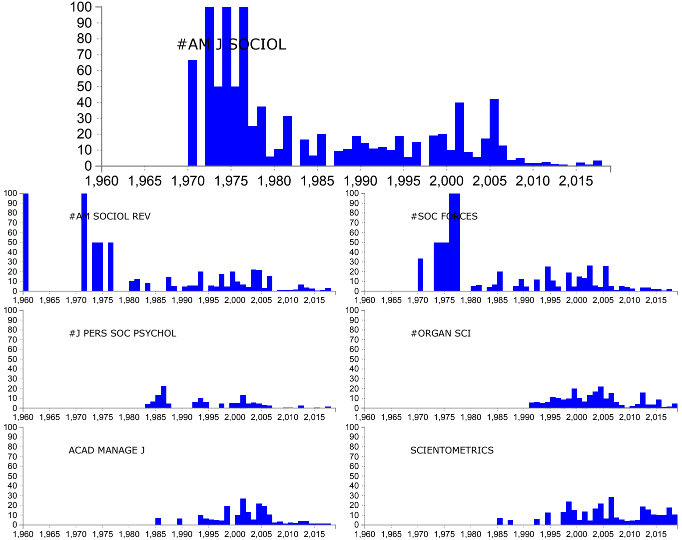
## from WJins norm network



# Importance of journals in Social Sciences

## from WJins norm network

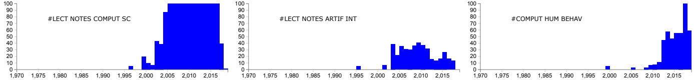




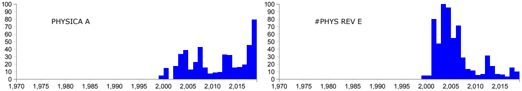
# Importance of other journals

## from WJins norm network

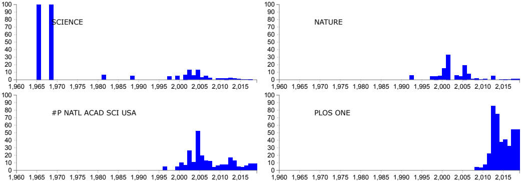
### Computer Science




### Physics



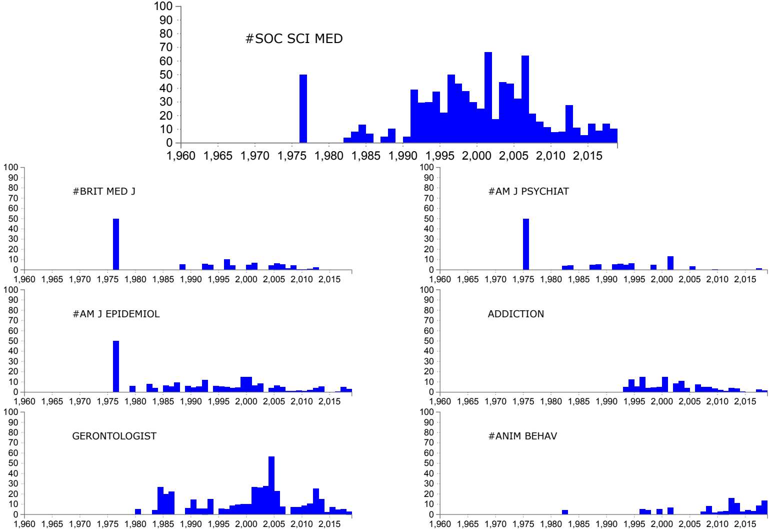

### General scientific journals



# Importance of other journals

## from WJins norm network





# Citations among journals

## JCJ and JCJn from WJins, WJcum, CiteIns nets

$$\mathbf{JCJ} = (\mathbf{WJins})^T * \mathbf{CiteIns} * \mathbf{WJcum}$$

The value of weight of the element $JCJ[i,j]$ is equal to the **number of citations** per year from journal $i$ to journal $j$.

$$\mathbf{JCJn} = (\mathbf{WJins})^T * n(\mathbf{CiteIns}) * \mathbf{WJcum}$$

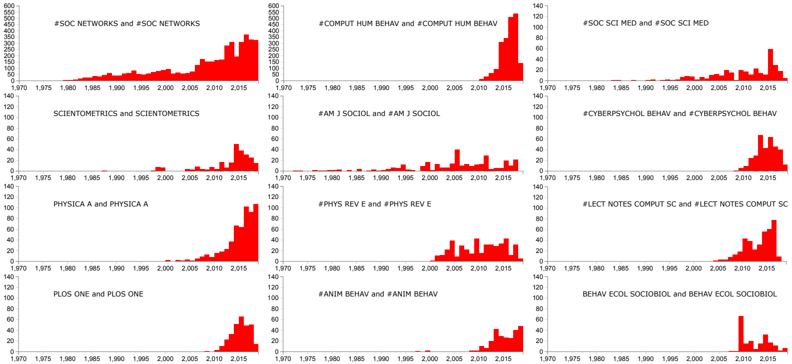
where

$$n(\mathbf{CiteIns})[u,v] = \frac{\mathbf{CiteIns}[u,v]}{\max(1, \mathrm{outdeg}(u))}$$

The value of element $JCJn[i,j]$ is equal to the number of **fractional contribution** of citations per year from journal $i$ to journal $j$.

# Self-citations of journals

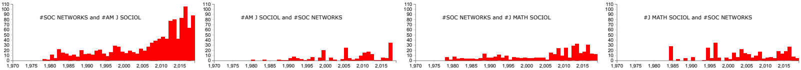
## Loops from JCJ network



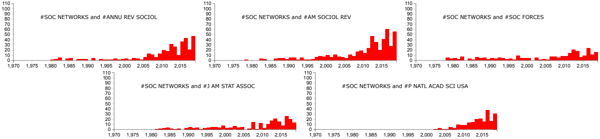
# InSum and OutSum of JCJ network without loops



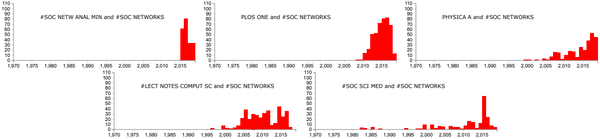
## OutSum and InSum:




## OutSum:



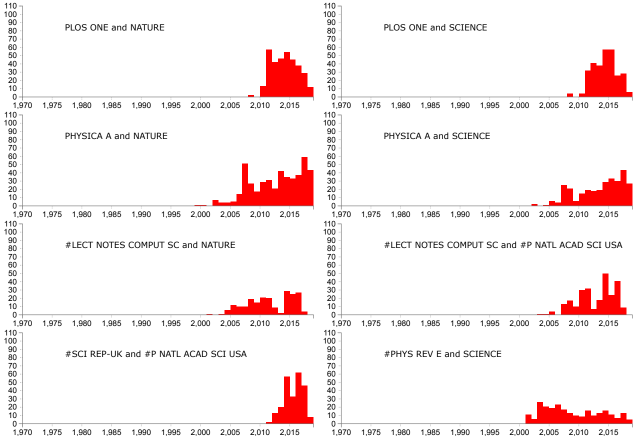

## InSum:



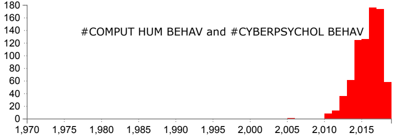
# Citation of general scientific journals

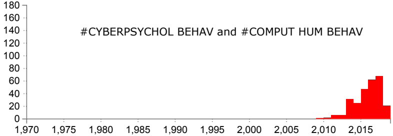
## from JCJ network without loops

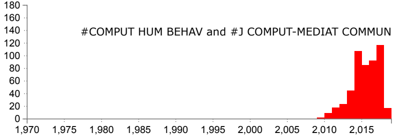


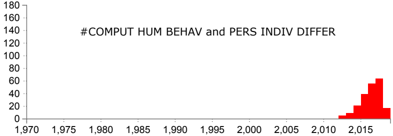


### Computer Science:

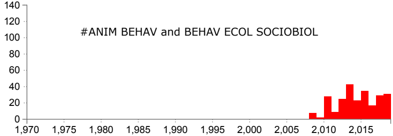


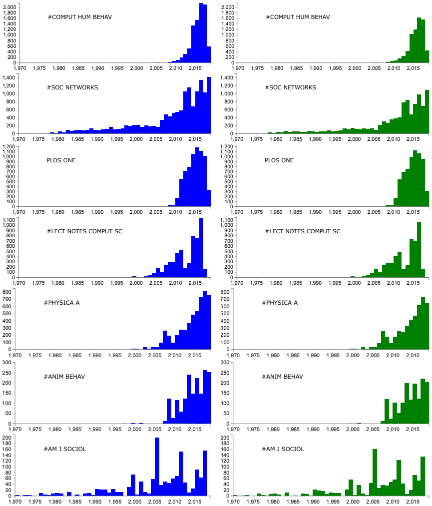




### Animal social networks:



# Outgoing citations with/without loops

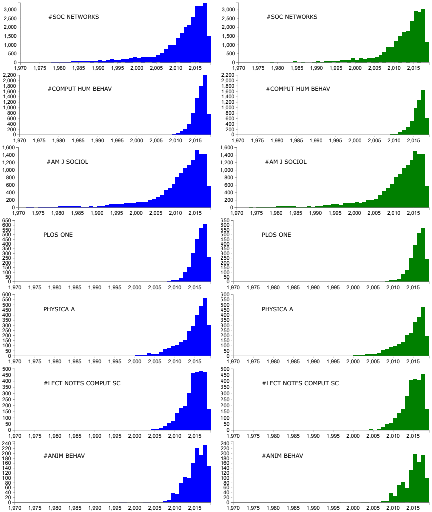
## from JCJn network



# Incoming citations with/without loops

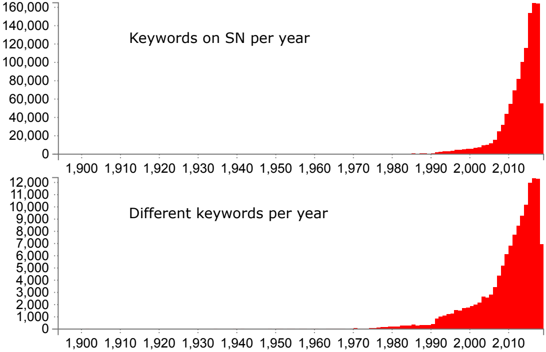
## from JCJn network





# Number of keywords

## from WKins network

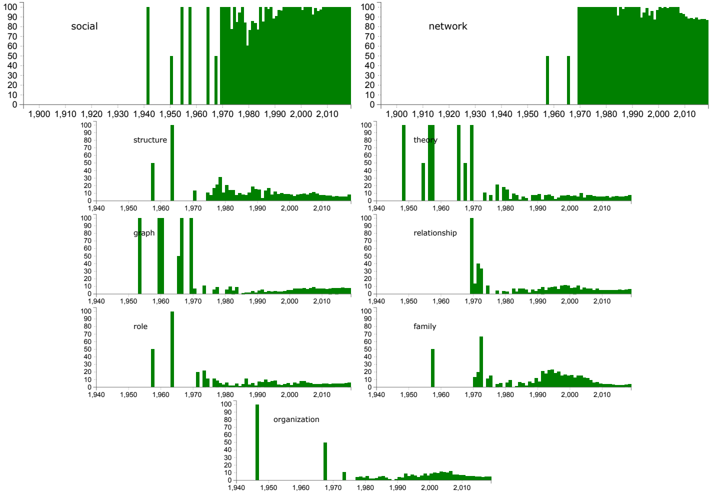



Basing on **WKins** network, we counted a proportion of the *Number of each keyword* per year and the *Maximum number of keywords* for each year. This proportions aims to normalize the importance of selected keywords through time in range from 0 to 100%.

# Importance of selected keywords

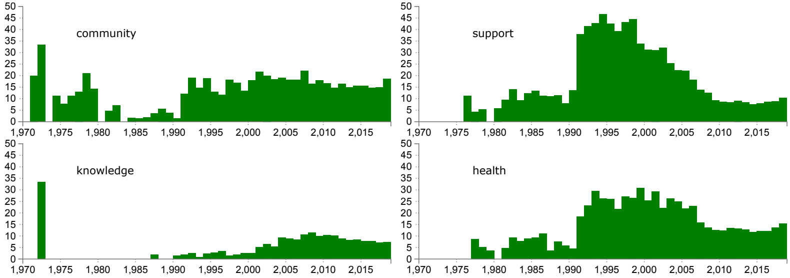
## from WKins network [100 scale]





# Importance of selected keywords
## from WKins network [50 scale]

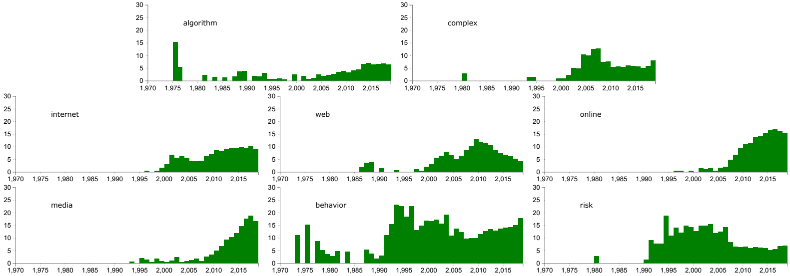




# Importance of selected keywords

## from WKins network [30 scale]

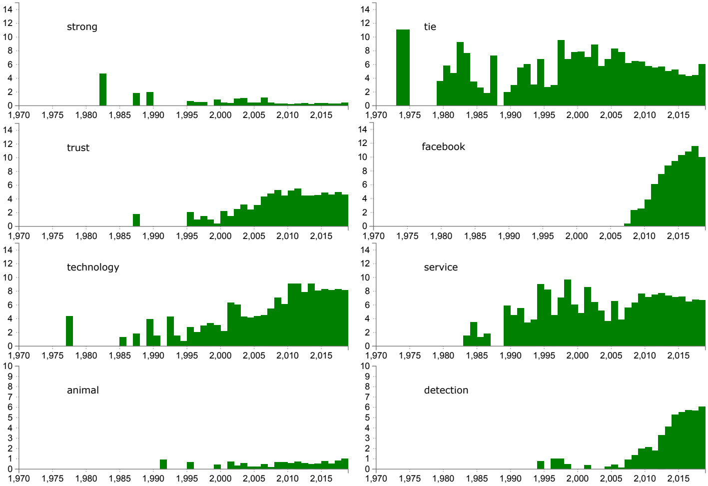




# Importance of selected keywords

## from WKins network [14 and 10 scale]





# Bibliography

1. Batagelj, V. (2007) WoS2Pajek. Networks from Web of Science. Version 0.3. Manual. URL: http://vlado.fmf.uni-lj.si/pub/networks/pajek/WoS2Pajek/WoS2Pajek.pdf
2. Batagelj V., Cerinšek M.(2013). On bibliographic networks. Scientometrics. 96 (3), 845-864
3. Batagelj, V., Doreian P., V., Ferligoj, A., Kejzar N. Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution, 2014.
4. Batagelj, V., Ferligoj, A. Squazzoni, F. The emergence of a field: a network analysis of research on peer review. Scientometrics (2017) 113: 503. https://doi.org/10.1007/s11192-017-2522-8