Network aspects of structural, dynamical and evaluative scientometrics

ANDRÁS SCHUBERT

MTA–KSZI, Budapest

Contents

- Fields of scientometrics
  - Structural scientometrics
  - Dynamical scientometrics
  - Evaluative scientometrics
- Structural maps of scientific links
- Dynamics of "scientopographic" maps
- Evaluation through network characteristics
- Network aspects of Hirsch's h-index
- Conclusions

"...*the isolated man does not develop any intellectual power. It is necessary for him to be immersed in an environment of other men, whose techniques he absorbs during the first twenty years of his life. He may then perhaps do a little research of his own and make a very few discoveries which are passed on to other men. From this point of view [research] must be regarded as carried out by the human community as a whole, rather than by individuals."*
(Alan Turing)

- The quantitative study and assessment of scientific activity (i.e., scientometrics) must, therefore, focus its attention much more on the connections among individual actors than on the individual actors themselves.

Fields of scientometrics

- Structural scientometrics

  Structural mapping of scientific communities, sets of documents, concepts, etc.

Fields of scientometrics

- Structural scientometrics

  Structural mapping of scientific communities, sets of documents, concepts, etc.

- Dynamical scientometrics

  Studying the space-time behavior of scientific information

Fields of scientometrics

- Structural scientometrics

  Structural mapping of scientific communities, sets of documents, concepts, etc.

- Dynamical scientometrics

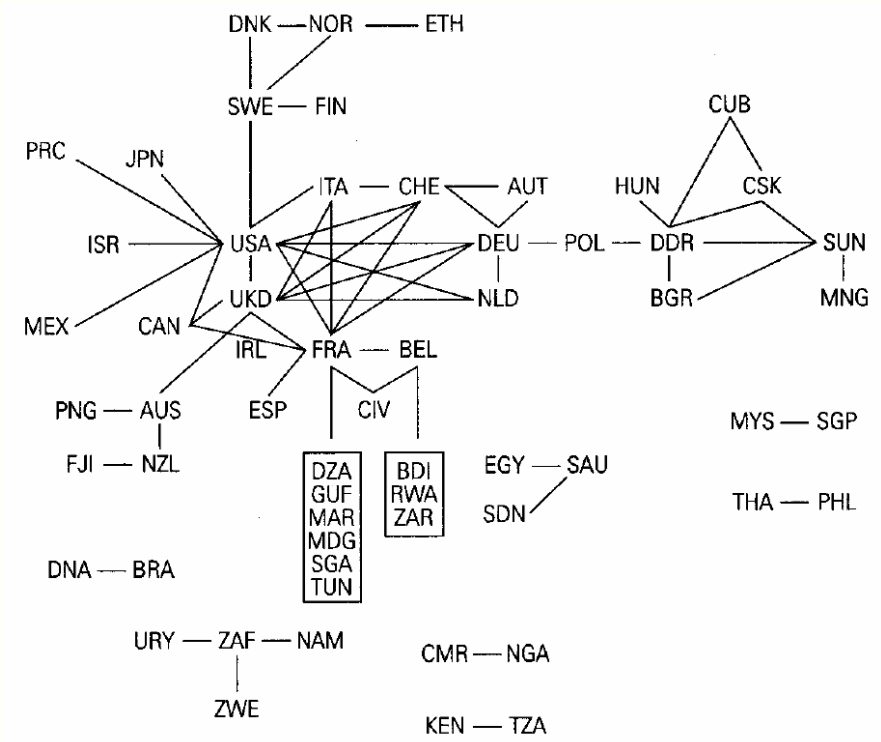
  Studying the space-time behavior of scientific information

- Evaluative scientometrics

  Assessing the performance of the actors of scientific research

Structural maps of scientific links: "Scientopography"

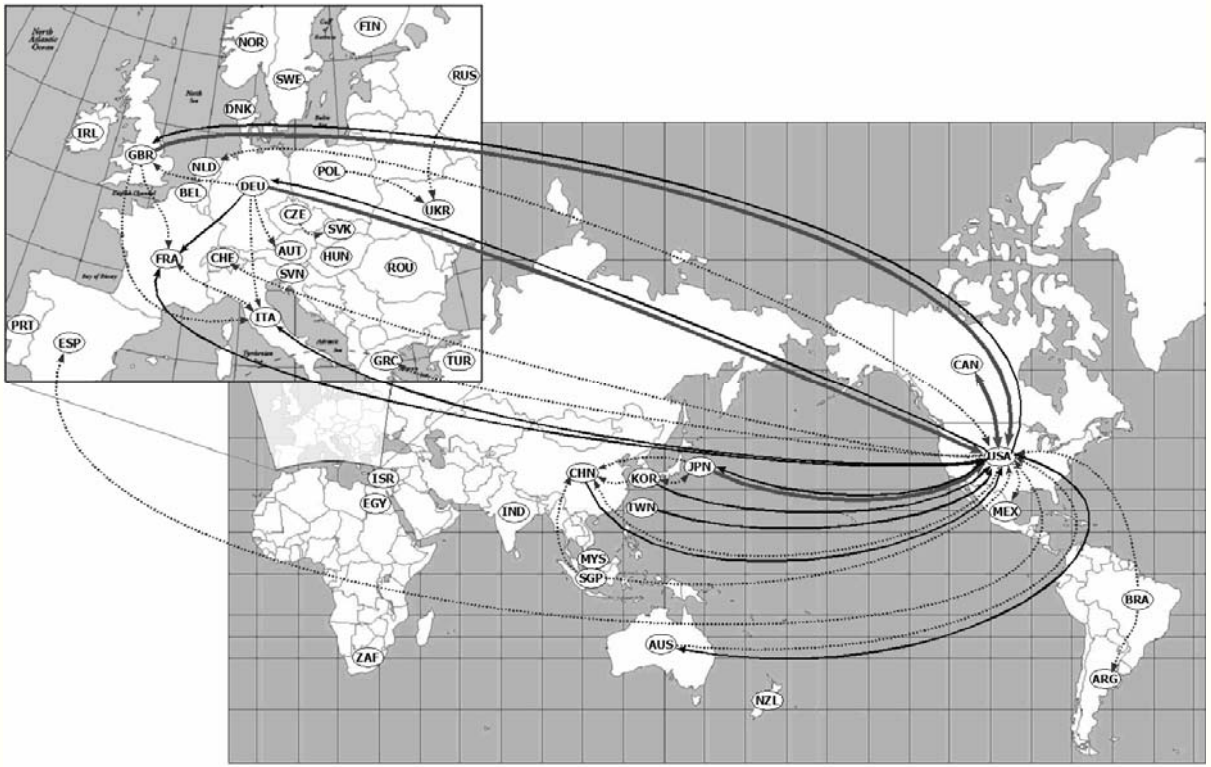
Co-authorship links, 1985



A. Schubert, T. Braun: Scientopography: World maps and charts based on scientometric indicators. A bird's eye view on the metropolis and beyond', In: Waast R. (ed.), *20th Century Sciences: Beyond the Metropolis.* Vol. 6. Paris, ORSTOM, pp. 65–72, 1996.

Structural maps of scientific links: "Scientopography"

Attendance in scientific conferences



W. Glänzel, B. Schlemmer, A. Schubert, B. Thijs: Proceedings literature as additional data source for bibliometric analysis, *Scientometrics*, 68 (2006) 457–473

Structural maps of scientific links: Ternary plots of science

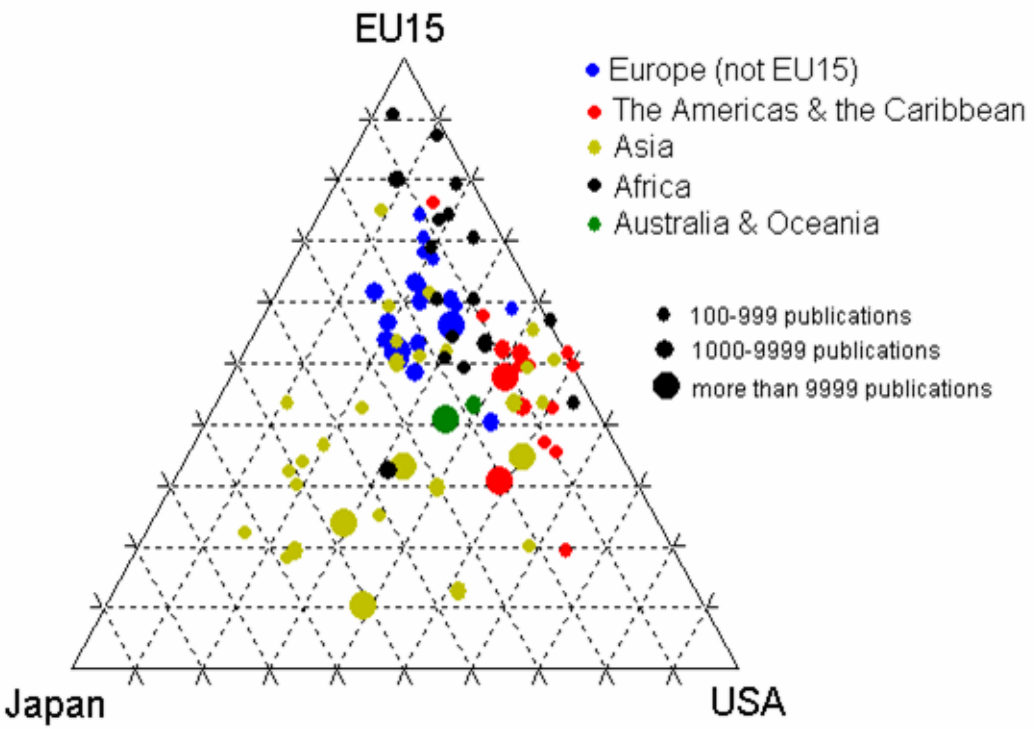
- The concept of the Triad (EU–USA–Japan) conceived by Kenichi Ohmae in the eighties, became widely used in international analyses both in economics and in S&T, and became a cliché by the turn of the millennium.

- The position of the "rest-of-the-world" countries relative to the Triad (either as co-authoring partners or as source of citations) can be impressively presented in ternary plots.

K. Ohmae: Triad Power: The *Coming Shape of Global Competition*. Free Press, New York, 1985.
A. Schubert, W. Glänzel: Ternary plots of science in a tripolar world, *ISSI Newletters*, No. 15 (September, 2008) pp. 51–52.

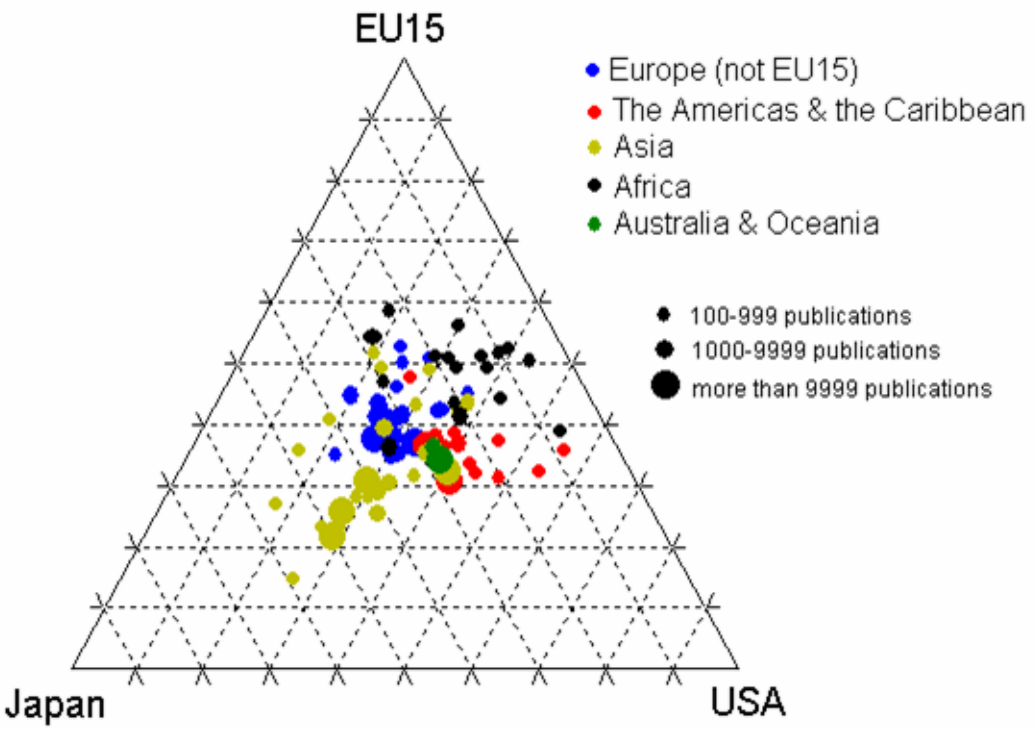
Structural maps of scientific links: Ternary plots of science

Ternary plot of co-authorship



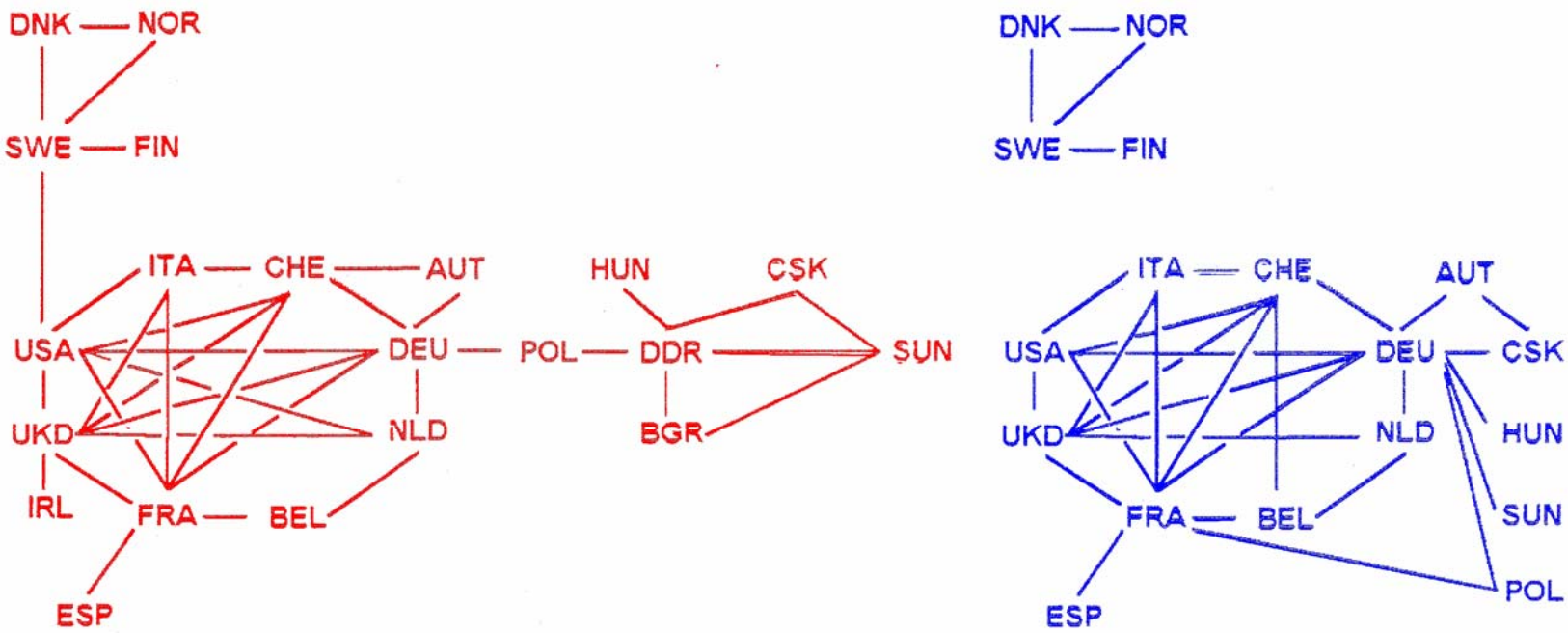
Structural maps of scientific links: Ternary plots of science

Ternary plot of cross-citations



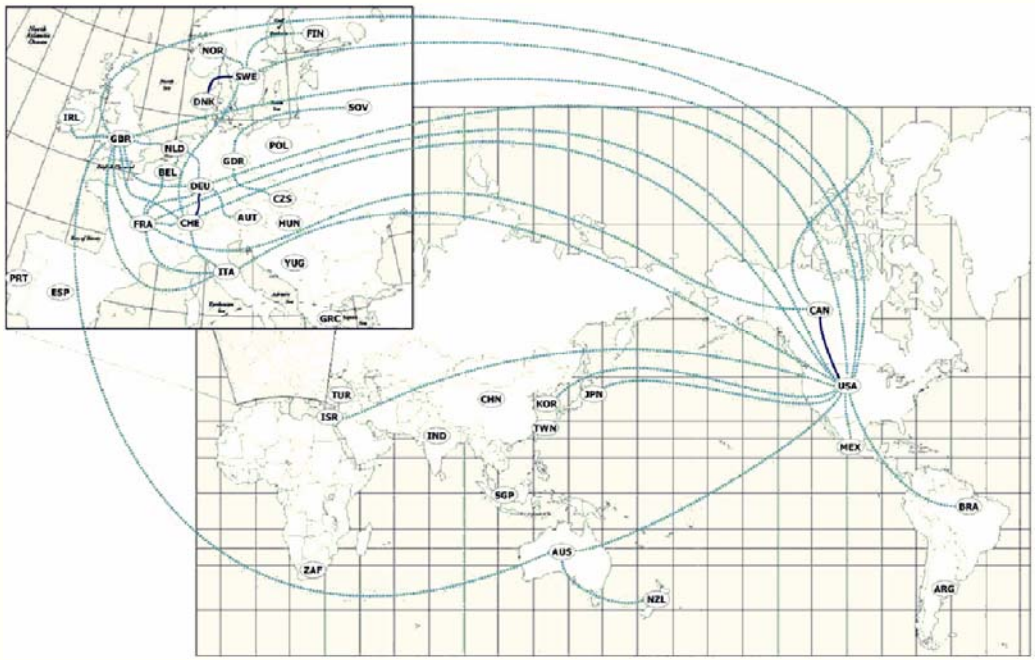
Dynamics of "scientopographic" maps

Co-authorship links, 1985 –> 1995



Dynamics of "scientopographic" maps

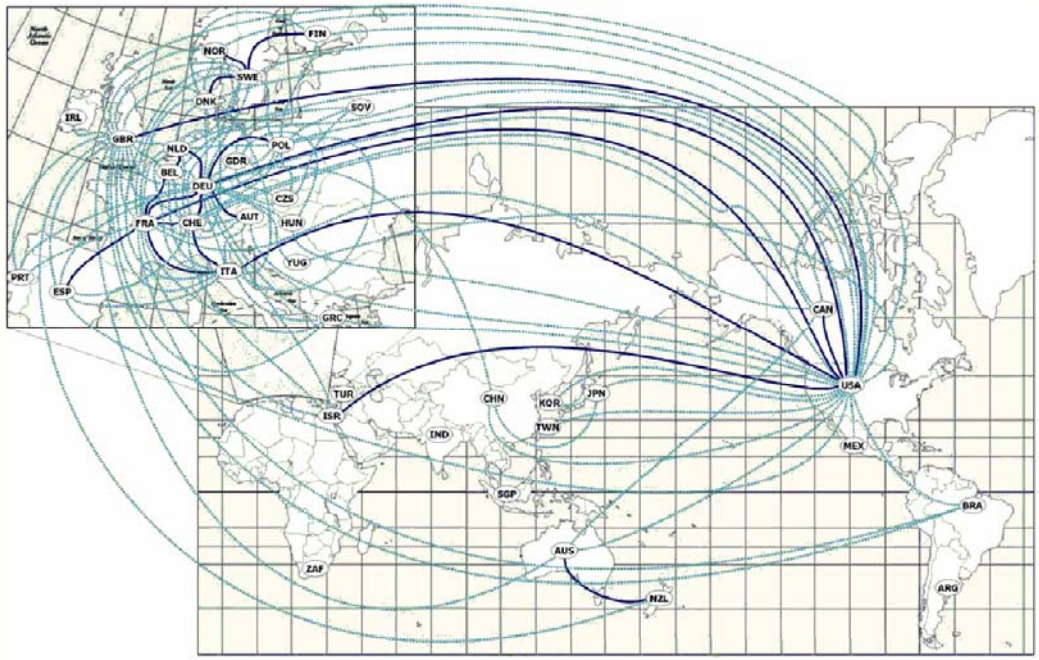
Co-authorship links, 1980



W. Glänzel, A. Schubert: Analyzing scientific networks through co-authorship, Ch 11 in *Handbook of Quantitative science and Technology Research,* 2004

Dynamics of "scientopographic" maps

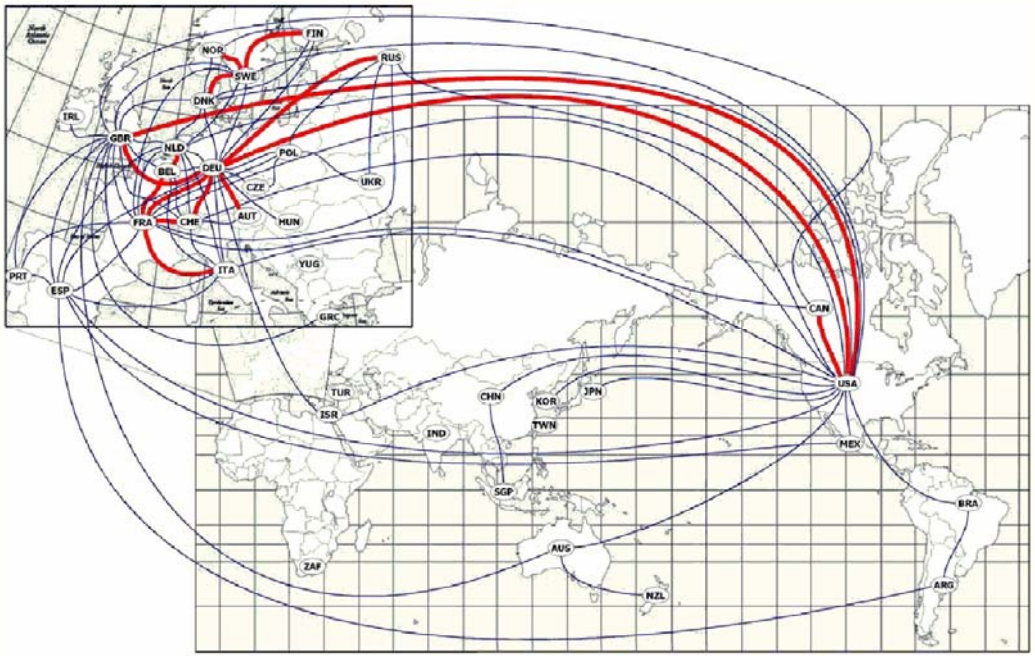
Co-authorship links, 1990



W. Glänzel, A. Schubert: Analyzing scientific networks through co-authorship, Ch 11 in *Handbook of Quantitative science and Technology Research,* 2004

Dynamics of "scientopographic" maps

Co-authorship links, 1990



W. Glänzel, A. Schubert: Analyzing scientific networks through co-authorship, Ch 11 in *Handbook of Quantitative science and Technology Research,* 2004

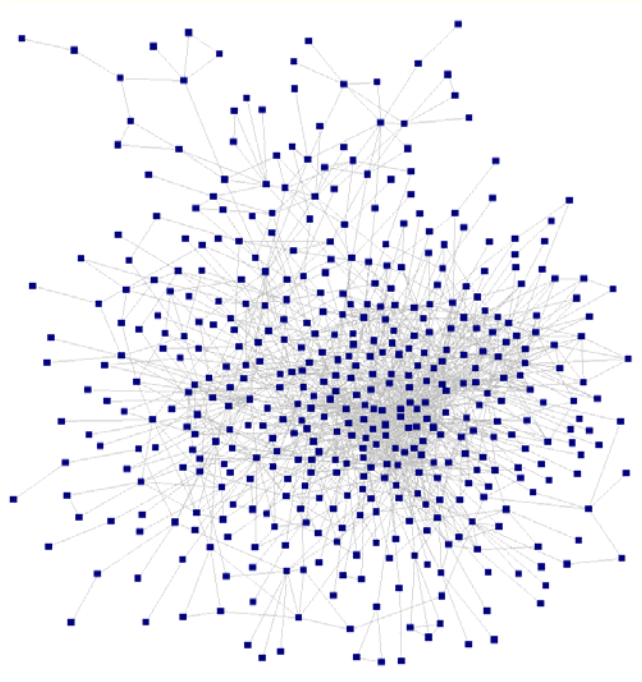
Evaluation through network characteristics: The Erdős-number

The Erdős-number is the archetype of structural–evaluative scientometric indicators

Definition:
The Erdős-number of Paul Erdős is 0.
The Erdős-number of an author is *n*,
if the minimum value of her/his
co-authors' Erdős-numbers is *n*–1.

Evaluation through network characteristics: The Erdős-number

The co-authorship network of
Paul Erdős (1416 papers, 509 co-authors)



V. Krebs, http://www.orgnet.com/Erdos.html

Evaluation through network characteristics: The Erdős-number

- Small Erdős number as a figure of merit

Nobel-prize winners and several other prominent members of the scientific community generally have an Erdős-number between 2 and 4. (Einstein: 2, Neumann: 3, Pauling: 4, Eccles: 3, Gates: 4)
((Schubert: 3))

- Measuring "Erdősness"

What indicators may characterize the position of an author in the co-authorship network?

Evaluation through network characteristics: Centrality and centralization

- Centrality measures of network nodes
  - Degree centrality
  - Closeness centrality
  - Betweenness centrality
  - Eigenvalue centrality
- The centralization of the entire network can be measured by the inequality of the node centralities

Evaluation through network characteristics: Centrality and centralization

The centrality of Paul Erdős within his co-authorship network is doubtlessly towering, but the centralization of the network itself is not particularly high;
the structure of the network suggests the existence of some kind of functional optimum.

Evaluation through network characteristics: "Outperform your neighbors"

- International comparisons are generally made on regional or global level, and usually result in regional or global rankings.
- The subjective perception of a country's position may not be perfectly correlated with its rank in a list. An individual's perceived satisfaction with her/his living standard is largely depends on how it compares with that of her/his next-door neighbors.
- This reasoning lead us to the attempt to construct indices which measure certain national indicators to the average of that of its bordering countries.

A. Schubert, M. Schubert: Outperform your neighbors, *Scientometrics*, 80(2) (2009) 555–560

Evaluation through network characteristics: "Outperform your neighbors"

Neighborhood-level indicators

Rank	Country	E(cit/pub)	Neighbors
1	Israel	2.887	EGY, JOR, LBN, SYR
2	Hong Kong	2.770	CHN
3	Japan	1.886	RUS, KOR, TWN
4	United States	1.748	CAN, CUB, MEX, RUS
5	Finland	1.586	NOR, RUS, SWE, EST
6	Switzerland	1.403	AUT, DEU, FRA, ITA, LIE
7	Thailand	1.398	KHM, LAO, MMR, MYS
8	Sweden	1.308	FIN, NOR, DNK, DEU, POL, EST, LVA, LTU
9	Singapore	1.271	IDN, MYS
10	England	1.156	SCO, WAL, IRL, NIR, FRA, NLD, BEL, NOR
11	Hungary	1.142	AUT, HRV, ROM, SRB, SVK, SVN, UKR
12	Australia	1.142	NZL, PNG, IDN
13	Chile	1.131	ARG, BOL, PER
14	Netherlands	1.128	BEL, DEU, ENG
15	Argentina	1.127	BOL, BRA, CHL, PRY, URY
16	South Africa	1.124	BWA, LSO, MOZ, NAM, SWZ, ZWE
17	Denmark	1.119	DEU, SWE, NOR, ENG
18	Bulgaria	1.083	GRC, MKD, ROM, SRB, TUR
19	Saudi Arabia	1.067	ARE, IRQ, JOR, KWT, OMN, QAT, YEM, EGY, IRN, SDN
20	Scotland	1.050	ENG, NOR, IRL
21	India	1.042	BGD, BTN, CHN, MMR, NPL, PAK, LKA
22	Venezuela	1.000	BRA, COL, GUY

Network aspects of Hirsch's h-index

- Definition:

An author's h-index is *h* if not more than *h*
of her/his papers received not less than h citations.

- Generalization for other sets of documents

(journals, institutes, countries, arbitrary set
of retrieved records in Web of Science)

Network aspects of Hirsch's h-index: Measuring centrality

- Generalization to network nodes:

  A node's h-index is *h* if not more than *h* of its neighbors have a degree not less than *h*.

- Generalization to entire networks:

  A network's h-index is *h* if not more than *h* of its nodes have a degree not less than *h*.

- The h-index of nodes is a possible (non-normalized) measure of centrality, the h-index of networks is that of centralization.

A. Korn, A. Schubert, A. Telcs: Lobby index in networks, *Physica A*, 388 (2009) 2221–2226
A. Schubert, A. Korn, A. Telcs: Hirsch-type indices for characterizing networks, *Scientometrics*, 78(2) (2009) 375–382

Network aspects of Hirsch's h-index: Indirect citation influence of highly cited publications

- Indirect citation influence: the citation does not only document the attention focussed on a publication but also generates it.

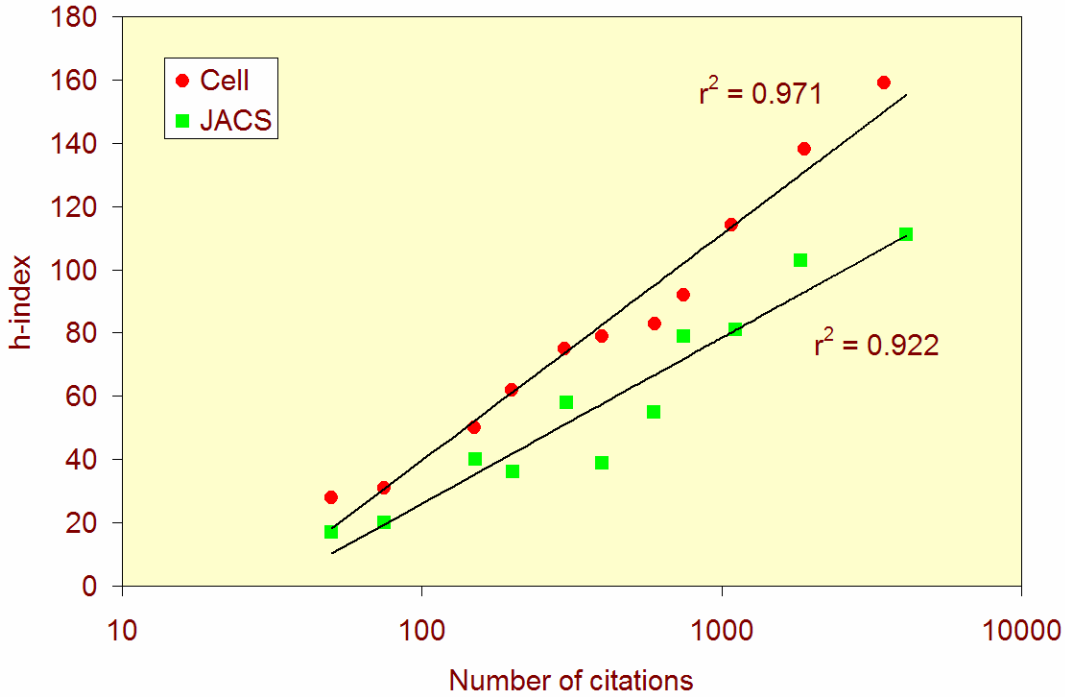
- The h-index of a publication can be defined as the h-index of the set of publications citing it.

A. Schubert: Using the h-index for assessing single publications, *Scientometrics*, 78 (3) (2009) 559–565

Network aspects of Hirsch's h-index:

Indirect citation influence of highly cited publications

Example



The h-index of a *Cell* paper is higher than the h-index of a *JACS* paper cited at the same rate. It is because papers citing *Cell* are usually higher cited than papers citing *JACS*.

Network aspects of Hirsch's h-index: The "influential weight" of author communities in journals

The weight of a community is determined by its size (number of members) and its coherence (connectivity among members).

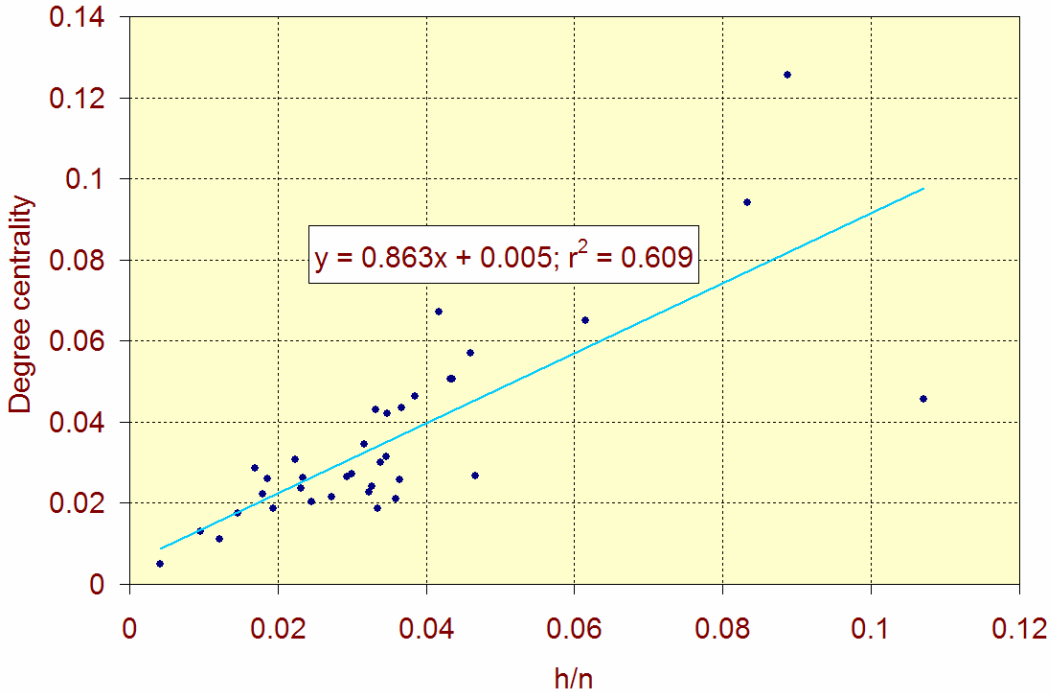
- Let us consider two authors of a journal connected if they have at least one joint paper. The h-index of the so defined co-author network is a possible measure of its "influential weight".
- Empirical study: the author community of 36 journals in the field of Dentistry & Oral Medicine was studied in papers published in 1999.

A. Schubert, A. Korn, A. Telcs: Hirsch-type indices for characterizing networks, *Scientometrics*, 78(2) (2009) 375–382

Network aspects of Hirsch's h-index:

The "influential weight" of author communities in journals

Empirical study

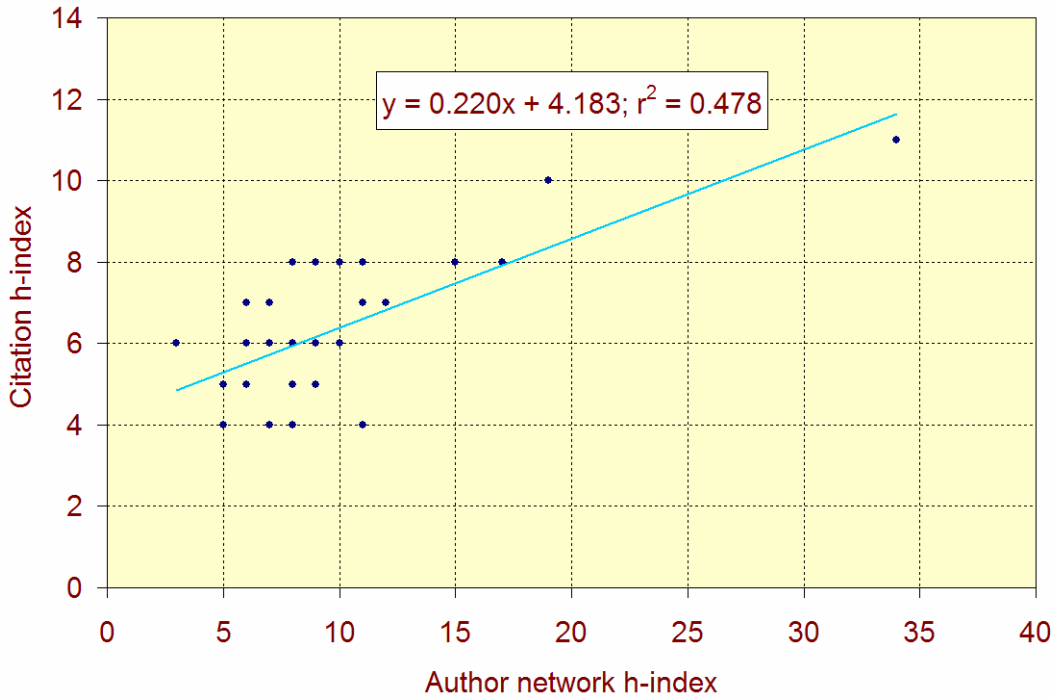


The "h-fraction" (the h-index divided by the number of authors) strongly correlates with one of the simplest measures of centrality: the degree centrality.

Network aspects of Hirsch's h-index:

The "influential weight" of author communities in journals

Empirical study



What is even more surprising: the citation h-index of journals appears to correlate with the degree h-index, as well. Thus the "influential weight" seems to influence the citation rate of the papers (partly, of course, through such 'condemnable' practices like self-citation of cross-citation).

Network aspects of Hirsch's h-index: h-similarity of journals

Let us define the h-core of a ranked cited journal list (like the ones found in the Citing Journal Package of the Journal Citation Reports) as the largest top h-element list with each item having at least h citations. Then we can define the h-similarity of two journals as the number of joint titles in their h-core divided by the total number of different titles in the union of the two h-cores:

$$h_{A,B} = |H_A \cap H_B| / | H_A \cup H_B | ,$$

where $h_{A,B}$ is the h-similarity between journals A and B, the sets H_A and H_B are the h-cores of A and B, respectively, ∩ denotes the intersection, U denotes the union and | | denotes the cardinality of sets.
The h-similarity is a Jaccard-type similarity measure with a range [0,1] that has value 0 if and only if the two h-cores have no common elements and has value 1 if and only if the two h-cores contain identical elements in whatever order.

A. Schubert: A reference-based Hirschian similarity measure for journals, *Scientometrics*, DOI 10.1007/s11192-009-0072-4 (to be published in 2009)

Network aspects of Hirsch's h-index: h-similarity of journals

The concept of h-similarity can advantageously used to construct standardized impact factors. Namely,

$$\xi_A = x_A/(\Sigma_B h_{A,B} x_B/\Sigma_B h_{A,B}) \ ,$$

where ξ_A is the standardized impact factor of journal A, x_A and x_B are the JCR impact factors of journals A and B, $h_{A,B}$ is the h-similarity between journals A and B, and Σ_B denotes summation over all journals B other than A.

Network aspects of Hirsch's h-index: h-similarity of journals

The top list of non-review journals by the standardized impact factor includes several celebrated high-prestige titles, but also some journals of less overall visibility which, nevertheless, are preeminent in their reference environment.

Rank	Title	Std IF	JCR IF
1	NEW ENGL J MED	6.448	51.296
2	MIS QUART	3.436	4.731
3	INT J NONLINEAR SCI	3.421	4.386
4	INT J COMPUT VISION	3.280	6.085
5	LANCET	3.254	25.800
6	JAMA-J AM MED ASSOC	3.136	23.175
7	INT J PLASTICITY	3.104	4.113
8	ANN MATH	2.834	2.426
9	PROD OPER MANAG	2.783	2.516
10	IEEE T AUTOMAT CONTR	2.665	2.772
11	NAT IMMUNOL	2.624	27.596
12	SCIENCE	2.612	30.028
13	ACM T GRAPHIC	2.588	4.081
14	IEEE T PATTERN ANAL	2.537	4.306
15	ARCH GEN PSYCHIAT	2.526	13.936

Conclusions

Network aspects are of ever growing importance in the theory and practice of scientometrics not only at the structural but also at the dynamical and evaluative levels of investigations.

There is a bidirectional exchange in the toolkit of network studies and scientometrics; for example, the h-index, originally conceived and developed as a tool of evaluative scientometrics, seems to open new directions in the study of general networks, as well.

Thanks are due to

- my co-authors
  for their essential contribution

Thanks are due to

- my co-authors
  for their essential contribution
- **the organizers of this conference**
  for the honoring invitation

Thanks are due to

- my co-authors
  for their essential contribution
- the organizers of this conference
  for the honoring invitation
- **the audience**
  for their interest and patience