Comparison of Rules in Bibliographic and Web Networks

Hildrun Kretschmer Theo Kretschmer

Abstract

The emergence of the Internet and the Web have led to changes in the process of scholarly publishing and communication. In particular, the way that scientists and scholars search for and find information in patterns of national and international collaboration has changed. The development of information and library sciences together with science studies will, among other things, be fashioned by the development of quantitative studies conducted in this field (The way from librametry to webometrics).

The EU has recently financed a new project (WISER) to further investigate the potential of creating new indicators of the Web for use in science and technology policy making. The study of collaboration in e-science is one focus of this project, including questions that examine the extent to which collaboration structures visible in the Web follow similar rules to collaboration networks measured by traditional bibliometric data.

There are three specific key objectives:

- *To develop methods or webometric indicators for measurement of Web visibility of collaboration networks*
- *To collect general rules in bibliographic collaboration structures that are suitable for testing in Web networks*
- *To compare bibliographic collaboration networks and networks visible in the Web*

There are few general rules in bibliographic co-authorship networks available from the literature. A new approach is proposed by the author: Distribution of co-author couples in journals: "Continuation" of Lotka´s law on the 3rd dimension.

Keywords : Webometrics, Librametry

0. Introduction

Newly emerging information and communication technologies (ICT) are becoming increasingly important in libraries as well as in science and in scholarly research. The emergence of the Internet and the Web have led to changes in the process of scholarly publishing and communication. In particular, the way that scientists and scholars search for and find information in patterns of international collaboration has changed.

The development of information and library sciences together with science studies will, among other things, be fashioned by the development of quantitative studies conducted in this field.

1. The Way from Librametry to Webometrics

Some of the ideas regarding the way from librametry to webometrics are already mentioned in Kretschmer & Thelwall 2003.

Around fifty years ago, at the Aslib Conference in Leamington Spa in 1948, the term 'Librametry' was established by the famous Indian librarian S. R. Ranganathan: "...it is necessary for librarians to develop librametry in the lines of biometry, econometry and psychometry since *many of the matters connected with library work and services involve large numbers...*" (Aslib Proceedings 1949, p. lO2). His suggestions were welcomed at the conference, notably by Bernal.

In preparation for this great event, Ranganathan in the course of the preceding 20 years, i.e. since 1925, had already successfully practised the application of 'Elements of Statistical Calculus' on library problems. This success inspired him to publicly introduce the term 'Librametry' to the above-mentioned conference (Gopinath, 1992). For many years Ranganathan had worked in the library of the University of Madras.

Within the sphere of libraries, the term 'librametry' was applied to a wide range of problems, practiced by Ranganathan and other Indian scientists. The following uses account only for some of them (Ranganathan, 1996, 1995 reprinted).

- Librametry in the day-to-day work of a library
- Librametry in organising national or state library systems
- Operations research in library work
- Librametry and book selection
- Librametry and classification
- Library administration
- Library services

In the first few years following Ranganathan's achievements, there were hardly any responses to his work in the western world. As a result, more than 20 years after the introduction of the term 'Librametry' - sometimes also called 'Librametrics' - A. Prichard, the information scientist, coined the term, in 1969 , 'Bibliometrics' as a redevelopment of the term 'Statistical Bibliography' for the quantitative analysis of bibliographies (Sen, 1995). This term was very rapidly accepted internationally, with 'Bibliometrics' eventually also being accepted by Indian librarians.

However, at a later date, the well-known English information scientist B. C. Brookes pointed out regretfully (l990, p. 41): "Had I known of Ranganathan's term in time, I would have adopted librametrics for information studies and bibliometrics for information science. But it was too late. Librarians liked bibliometrics too". Brookes (1990) attached importance to a distinction to be made between Information Science and Information Studies, with the former term apparently being more theory-oriented and the latter term more application-oriented, i.e. going along the lines of the use of techniques with a view to optimising Library Administrations and Library Services.

The technical redevelopment of methods of communication via the Internet presents a challenge for information scientists to cultivate novel quantitative methods and techniques in order to measure rates of information exchange in this new medium.

Probably the biggest changes in the methods of information exchange in the West over the past ten years have all been driven by the potential of the Internet. Digital libraries (Fox & Urs, 2002) are one visible incursion into the domain of librarians, and today much more general information is electronic and available over the Internet. In the West, librarians seem to be keeping their traditional roles, perhaps in a reduced form, but moving into new areas, helping users to search for information from a much greater variety of sources, many online. Information science research has also changed, with much research taking place into how the new technologies are being used, particularly email (Herring, 2002) and the web (Cronin, 2001; Kling & McKim, 2000). In addition to user studies, there have been attempts to extract new kinds of information from the Web by, for example, examining the relationship between areas of the

Web by counting the number of hyperlinks between them (Ingwersen, 1998). These kinds of studies called "webometrics" have grown out of bibliometric analyses of the citations in published journal articles (Vohora *et al*., 2001; Borgman, & Furner, 2002).

2. The European Union-funded project WISER

A new European Union-funded project into this area is a part of the emerging field of webometrics.

An increasing part of online scientific communication and research is not (or only partially) visible in traditional S&T indicators. The EU has recently financed a new project run by a consortium from England, The Netherlands and Spain to further investigate the potential of creating new indicators of the Web for use in science and technology policy making. The WISER project started in November 2002, and runs for three years. Its objective is to explore the possibilities and problems inherent in developing a new generation of Web based S&T indicators (www.webindicators.org).

The study of collaboration in e-science is one focus of this project, including questions that examine the extent to which collaboration structures visible in the Web follow similar rules to collaboration networks measured by traditional bibliometric data.

There are three specific key objectives:

- To develop methods or webometric indicators for measurement of Web visibility of collaboration networks
- To collect general rules in bibliographic collaboration networks that are suitable for testing in Web networks
- To compare bibliographic collaboration networks and networks visible on the Web

2.1 Indicators for webometric measurement of Web visibility of collaboration

In former studies, Kretschmer and Aguillo (forthcoming) have defined new webometric indicators to measure the visibility of collaboration in the Web. A method of Vaughan and Shaw (forthcoming) to search article quotations in the Web (*Web citations*) was used successfully, albeit in a slightly modified form, to measure the visibility of the collaboration in the Web with the following definitions of new indicators:

- The Web visibility of a co-authored publication won by bibliographic data (WVP) is measured as a frequency of the different Websites on which this bibliographic publication is mentioned after entering the full title of the co-authored publication into Google or Alltheweb.
- The Web visibility of a pair of collaborators (WVC) is equal to the sum of Web visibility, WVP, of all of their co-authored publications.

Results of a pilot study in brief (Kretschmer & Aguillo, forthcoming) show that a high percentage (78%) of all bibliographic multi- authored publications of the actors of a network became visible through the use of search engines in the Web.

2.2 Collection of general rules in bibliographic collaboration networks suitable for testing in Web networks

There are few general rules in bibliographic co-authorship networks available from the literature.

Therefore two new approaches are proposed by the author:

- Author productivity and geodesic distance (Kretschmer, forthcoming)
- Distribution of co-author couples in journals: "Continuation" of Lotka´s law on the 3rd dimension

The first approach has been accepted for publication in *Scientometrics* and the second is presented here.

2.3 Comparison of bibliographic collaboration networks and networks visible on the Web

The first results of a pilot study (Kretschmer & Aguillo, forthcoming) have shown that the newly created webometric indicators of Web visibility could be usefully applied as methods of investigating similarities and dissimilarities in bibliographic networks and in the Web. Both networks (co-authorship network won by bibliometric data and the corresponding co-authorship network won by webometric data) are very similar, with only very slight deviations. However, the pilot study was limited to 64 actors in the network.

Therefore the pilot study should be continued with larger samples of scientists to allow for the comparison of general rules in bibliographic as well as in Web networks.

General Rules in Bibliographic Networks Suitable for Testing in the Web: Distribution of Co-Author Couples in Journals:

"Continuation" of Lotka´s Law on the 3rd Dimension

Over time, collaboration is increasing in science and in technology. Single authored bibliographies are changing to multi-authored bibliographies with an increasing number of co-authors per paper.

In single authored bibliographies only single scientists distributions can be found. However, in multi-authored bibliographies single scientists distributions, couples distributions, triples distributions, etc., can be presented. Several methods are possible for counting single scientists, couples, triples etc. These are:

- Normal count procedure, in which each co-author (single scientists P counting like in Lotka´s law) or each couple (couples P,Q counting), etc., receives full credit
- Weighted (or fractional) counting, in which each co-author of m co-authors per paper receives 1/mth credit etc.

The famous Lotka's law, discovered in 1926, remains a source of fascination for scientists from all fields.

In Lotka´s law, based on single scientists, P, counting is valid for:

- single authored bibliographies, and,
- under the condition of normal count procedure, also for multi-authored bibliographies

Whereas regarding Lotka´s law single scientists P distribution (both in single authored and in multi-authored bibliographies) is of interest, in the future couples P,Q distribution, triples P,Q,R distribution, etc. should be considered using normal counting procedure.

Starting with couples distribution, the following question arises in the present paper: Is there any pattern for the distribution of co-author couples in journals?

3. Methods

Different versions of counting couples P,Q are possible. Two of them will be presented below:

Given is a bibliography (partly represented, names of authors A, B, ...)

1.	A	4.	D, A, F	7.	H,G
2.	B	5.	C	8.	H,G,A
3.	D,E	6.	G, H	etc.	etc.

The number of publications, i, per author P is determined by resorting to the ´normal count procedure´. Each time the name of an author appears, it is counted (e.g. A three times: once in the first article, and once each in the 4th and 8th article).

It should be noted here the term ´article´ is used with reference to a work or paper, which was jointly written by one or several authors, (cf. 1., 2., 3., ...) etc. articles in the bibliography. By contrast, the term ´publication´ refers to persons.

First version : Couples P,Q are counted analog to single authors P in Lotka´s law. The number of publications n_c per couple P,Q is determined (D,E once, D,A once, D,F once, A,F once, G,H once, H,A once, G,A once, H,G twice).

Second version : Couples P,Q are counted under the condition of both the first authors P count (i) and the second authors Q count (j), i.e. the authors are ordered according to i or j respectively (cf. Table 1). In the present paper the symmetrical form of matrix 1 will be studied, see Table 2. In the symmetrical matrix, one can determine for each author P the number of his collaborators N_P as well as the total sum of his relationships through co-authorships C_P.

Table 1: Matrix 1

i / j		1				2	3		
		B	C	E	F	D	A	G	H
	B								
	C								
1	E								
	F								
2	D			1	1		1		
	A				1				
3	G						1		1
	H						1	2	

Table 2: Symmetrical form of matrix 1

i / j		1				2	3				
		B	C	E	F	D	A	G	H	Np	Cp
	B										
	C										
1	E					1				1	1
	F					1	1			2	2
2	D			1	1		1			3	3
	A				1	1		1	1	4	4
3	G						1		3	2	4
	H						1	3		2	4
									SUM	14	18

The number of collaborators N_P is the number of couples assigned to the author P in the corresponding row of the matrix. The number of relationships through co-authorships can differ between the couples.

For example, the author P=H has two collaborators (N_H = 2), namely Q_1 = A and Q_2 = G. While the author H has one relationship with the collaborator A through co-authorships, there are three relationships with the collaborator G.

The sum of the relationships through co-authorships for the author P = H is, thus, four (C_H=4).

Two different matrices can be derived from matrix 2:

- Table 3 is the representation of the number of couples N_{ij} with authors who have i publications per author, with authors who have j publications per author included in the bibliography. A_i is the number of authors with i publications per author. $N_i = S_j N_{ij}$ is the number of collaborators of all authors with i publications per author. $R_i = k = N_i/A_i$ is the average number of collaborators per author. A former investigation on the topic of collaboration and Lotka proposed to use the average number of collaborators per author k to predict the productivity strata (Jian Qin 1995). An experiment was introduced in Lotka's law with the new variable (k).

- Table 4 is the representation of the number of co-authorship relations C_{ij} between authors with i publications per author, and authors with j publications per author. The matrix C_{ij} was the foundation of an earlier study (Kretschmer 1999). Due to the possible fluctuation at that time, however, a classification of the data i and j was done corresponding to the logarithm resulting in a matrix C_{XY}. The structure of this matrix C_{XY}, as also of the corresponding relative values, has been described. The results of this study have been taken as a pre-condition for the present work.

Tabel 3 : Matrix of N_{ij}

i / j	1	2	3	N_i	A_i	R_i
1	0	2	1	3	4	0.75
2	2	0	1	3	1	3
3	1	1	6	8	3	2.67
			SUM	14	8	

Table 4: Matrix of C_{ij}

i / j	1	2	3	C_i
1	0	2	1	3
2	2	0	1	3
3	1	1	10	12
			SUM	18

4. Hypotheses

The hypotheses were derived from the following considerations:

- How would the distribution of the number of couples look like, if derived alone from the marginal sum N_i, that is from the number of collaborators of all authors with i publications per author?
- Measurement of the influence of characteristic features of social structures on the distribution of N_{ij} regardless of the marginal sum N_i
- Derivation of a formula for describing the distribution of the number of couples N_{ij} including both the distribution of the marginal sums and the characteristic features of social structures

Possible distribution of the number of couples N_{ij} , derived from the marginal sums N_i

Jian Qin (1995) showed, in her example, that the number of collaborators N_i is distributed in the same way as the total number of publications of all authors with i publications per author (T_i):

$$T_i = i \cdot A_i \qquad (1)$$

This means, that the marginal sums N_i should be distributed according to an inverse power function in line with Lotka's law, however, with a different parameter:

$$N_i = \text{constant}/i^a \qquad (2)$$

Due to the symmetry of the matrix, both the distributions of the marginal sums of the matrix are equal (row = column).

Assuming the condition that the productivity of the authors has no social influence whatsoever, whichever author collaborates with which other author, the distribution of the number of couples could be determined within the matrix solely on the basis of the marginal sums:

$N'_{ij} = N_i \cdot N_j / S_i N_i$ (3)

Under the condition the formula (2) is valid there could be a relationship here with Lotka's law:

$N'_{ij} = \text{constant}/(i \cdot j)^a$ (4)

Measurement of the influence of social structures on the distribution of the number of co-authorship couples N_{ij} regardless of the marginal sums N_i

To determine, whether the distribution of data within a matrix shows additional characteristic features, which have arisen independent of the distribution of the marginal sums, the homophylic index H_{ij} can be used.

In some sociological studies of interpersonal relations in social networks of men (Wolf 1996), this special homophylic index is used. This index provides information on the factor by which the observed frequency in a cell of a matrix deviates from the occupancy of this cell that would otherwise be expected in case of statistical independence from its characteristics. In order to calculate this index, we have to covert the matrix of observed frequencies N_{ij} into a new matrix using geometric mean. The special homophylic index H_{ij} is defined as:

$$H_{ij} = N_{ij} \cdot \frac{G}{G_i \cdot G_j} \quad (5)$$

where G - geometric mean of all matrix data

G_i - geometric mean of the data in row i

G_j - geometric mean of the data in column j

Under the condition the distribution of data within the matrix is determined by the distribution of the marginal sums the resulting homophylic index H'_{ij} is equal to 1 in each cell of the matrix.

Thus, H'_{ij} is valid starting from the distribution of N'_{ij} produced according to formula (3).

On the other hand, if the distribution of the actually observed N_{ij} is determined beyond the marginal sums through the structures, which are in general valid in social structures, these structures must be provable in the distribution of H_{ij}. The distribution of H_{ij} must be different from that of H'_{ij}.

5. How do these social structures appear ?

When discussing the structural characteristics of interpersonal relations in social networks, reference will be made to one of Wolf's works (1996) rather than to the many studies conducted and contained in the literature. As a result, a definite structure can be identified that underlies a large number of social processes of a distributive character, such as the spreading of diseases, the propagation of information, the change of views or the distribution of innovations. A generalization of this structure reveals three pivotal aspects:

1. Over-coincidental similarity among persons in contact with each other ("Birds of a feather flock together");
2. Decrease of interpersonal relations with declining similarity
3. Emergence of the 'edge effect' (cf. below);

The author illustrates these three aspects on the basis of an empirical example investigated by using the homophylic index. (Wolf, 1996, p. 35). Independently of whether or not sociodemographic features, socio-structural characteristics or general approaches are taken into account, it has repeatedly been shown that persons with social contacts reveal greater characteristic similarities than could have expected from persons with accidental associations. Relations may qualify as friendships, marriages, professional contacts, collaboration or other types of relationship.

Wolf, in one of his empirical examples, studied the similarity underlying relations of friendship due to common education. It was quite clear that the people studied chose to become friends with others who had achieved the same level of education. These data of the same level of education can also be used to observe the edge effect. This term designates the more pronounced similarity of friendly couples that is observable at the edges of status features (referring also to persons at the lowest and also highest levels of education). Using Wolf's data file it is possible to identify 4-times-higher relations between high-school leavers and university graduates than it would be expected at a fortuitous choice of friends. The tendency to choose status-homogeneous friends is not as clearly perceptible with persons having medium-level school degrees. For those with the same level of education a U-curve of data arose.

In the above-mentioned earlier paper (Kretschmer 1999) it was shown that these general basic structures of social networks are also valid in co-authorship networks. This was shown by analyzing the relationships through co-authorships (Matrix of C_{ij}). The present paper will further examine whether these basic structures also appear in the analysis of couples (Matrix of N_{ij}).

Derivation of a formula for describing the distribution of the number of couples N_{ij}

The first pivotal aspect of the interpersonal relations in social networks, i.e. the over-coincidental similarity among persons in contact with each other, the well known proverb "Birds of a feather flock together" could be conveniently integrated into the theory, together with the empirical results published (Kretschmer 1999).

In the literature of sociology far less evidence is found, however, for the opposite saying of "Opposites attract", although several efforts have also been put into proving its correctness; just think of Winch et al., 1954.

By contrast, the author (Kretschmer 1999) has made an attempt to suggest that both opposing proverbs should only be perceived as the conspicuously visible state of a holistic process caused by conditions to which the system under study was subjected at the time of investigation. In addition, the same applies to both opposing views of U-curves, i.e. with edge effect on the one side, and the reverse case, on the other side.

In this context it can be shown that four factors that can vary independently of one another, influence the authors; who makes a co-authorship with whom:

- Dissimilarity (A=| ln i – ln j |) regarding their productivity
- Complementary to that similarity (A_{max} – A)
  (Both the first two factors concern “Birds of a feather flock together” or “Opposites attract”).
- Another independent, influencing factor is the sum B= lni + lnj
- Complementary to this (B_{max} – B).

(For example, in the case of A= 0= constant on one hand the authors could be having low productivity with B_{min}, or on the other hand highly productive with B_{max} or also authors with any average productivity)

Both the last two factors concern the edge effect, visible through the U-curve or the reverse case.

The four influencing factors, that are independent of one another, have been included in the formula for the description of co-authorship structures. Similar to the formula of the earlier paper (Kretschmer 1999), it is assumed here for describing the distribution of H_{ij}:

$$H_{ij} = const \cdot (A + 1)^t \cdot (A_{max}-A+1)^w \cdot (B+1)^x \cdot (B_{max}-B+1)^y \quad (6)$$

with A = | ln i – ln j | and B = ln i + ln j and with the parameters t, w, x, and y.

It has been mentioned, in the earlier paper, that the values of the individual parameters depend on the interplay of the influencing factors and their complements. For example, the lengthening of the environmental conditions can have an effect on this interplay, and hence can affect a change of the parameter. In this context, different types of collaboration patterns have been presented.

How far the influence of the similarity of the authors regarding their productivity reduces for the benefit of the effect of dissimilarity and vice versa, depends on the ratio of the parameters t and w. These changes take place continuously. The same is also true of the parameters x and y with respect to the edge effect.

The logarithm of the maximum possible number of publications of an author corresponds to A_{max}. Thereby, $B_{max}= 2 \cdot A_{max}$. In the present study the maximum number of publications of an author was equal to 256.

However, it is also possible to lay down a specific value in future as standard for such studies, which does not vary depending upon the given sample.

It is assumed that the formula given above for the distribution of H_{ij} is also suitable for describing the number of couples N_{ij}, because N_{ij}

- is influenced on one hand by the social structures, and,
- independent of that, on the other hand, also by the distribution of the marginal sums N_i

Since the distribution of N'_{ij} can approximately be described by the product of the third and the fourth factor of the formula, it is assumed that while changing the value of the parameter, the distribution of the number of co-author couples can be described by the following formula:

$$N_{ij} = const \cdot (A + 1)^a \cdot (A_{max}-A+1)^b \cdot (B+1)^g \cdot (B_{max}-B+1)^d \quad (7)$$

with A = | ln i – ln j | and B = ln i + ln j and with the parameters a, b, g, and d

6. Data and Results

The articles from 1980-1998 of the journal *Science* were studied with 47,117 authors and N_{ij}=418,458 co-author couples.

Figure 1 shows the distribution of the number of authors A_i with i publications per author contrasted with the distribution of the number of co-authorship couples N_{ij} for clarity´s sake and optimum visualization restricted to authors with at most ten articles i(i=1,2…10).

The upper row reflects the Lotka's law. In this row, the distribution of A_i is given on the left, and on the right the corresponding double logarithmic presentation (R=0.997, F=1325.6 with n=10).

In the lower row, the distribution of the number of co-authorship couples N_{ij} is given on the left, and the corresponding triple logarithmic presentation on the right. The number of couples N_{ij} or log N_{ij} respectively are plotted at the Z-axis. Under the condition of logarithmic presentation after regression analysis the following correlation could be found:

R=0.998, F=3317.5 with n=55 values (because of symmetry, half matrix only plus main diagonal) for log N_{ij}.

R=0.993, F=615.6 with n=10 are valid regarding the distribution of log N_i according to an inverse power function, see formula (2)

Under the condition the data were cut off after i=15 about 99.9% of the total number of authors are included:

- R=0.996, F=1586.9, n=15 regarding the distribution of log A_i
- R=0.993, F=899.5, n=15 regarding the distribution of log N_i
- R=0.995, n=120 F=2881.7 regarding the distribution of log N_{ij}

Under the condition of using all data the following correlations could be found:

- R=0.978, F=619.7 with n=30 regarding the distribution of log A_i in comparison with Lotka´s law
- R=0.934, F=194.5 with n=30 regarding the distribution of log N_i
- R=0.980, F=1668.0 with n=280 regarding the distribution of log N_{ij} (Here the five authors with the highest number of publications are excluded because of matrix size).

Figure 2 shows the homophylic index, i(i=1,2…10). The distribution of H'_{ij} resembles a plane surface because the homophylic index H'_{ij} is equal to 1 in each cell of the matrix, see the four cases. The figures on the right are rotated by 90°.

In the upper row, the distribution of H'_{ij} is compared with the distribution of the theoretical values H_{Tij}. These theoretical values are determined through regression analysis of the empirical values H_{ij} regarding the formula (6) in the hypothesis (R=0.840, F=30.2 with n=55 for log H_{ij}).

In the lower row, the distribution of H'_{ij} is compared with the distribution of the empirical values H_{ij}.

Figure 2 clearly shows, that the general social structures – independent of the marginal sums N_i – have exercised an influence on the distribution of N_{ij}.

Fig. 3 shows the comparison of the empirical distribution of log N_{ij} with two different distributions of theoretical values. The figures on the right are rotated by 90°.

In the first row, the distribution of theoretical values is shown under the condition

$$\log N''_{ij} = constant + a' \cdot (\log i + \log j) \qquad (8)$$

The second row shows the empirical distribution of log N_{ij}.

In the 3rd row the theoretical values are shown under the condition:

$$\log N_{ij} = const + a \cdot \log (A + 1) + b \cdot \log(A_{max} - A + 1) + g \cdot \log(B+1) + d \cdot \log(B_{max} - B + 1) \qquad (9)$$

R=0.988 and F=2178.3 with n=55 after regression analysis of the empirical distribution of log N_{ij} with the theoretical pattern in the first row.

R=0.998 and F=3317.5 with n=55 after regression analysis of the empirical distribution of log N_{ij} with the theoretical pattern in the third row.

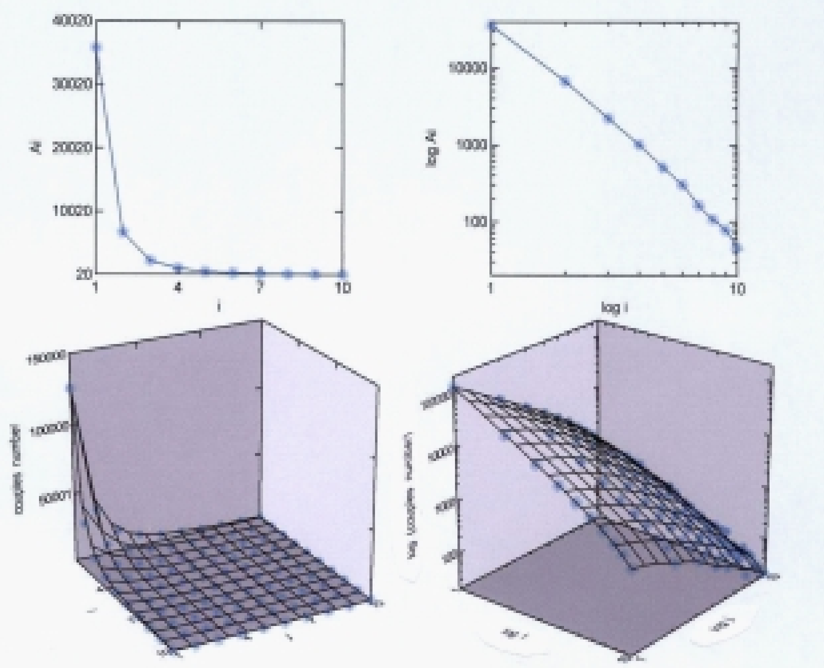

Fig. 1: Distribution of the number of authors A_i with i publications per author contrasted with the distribution of the number of co-authorship couples N_{ij} with i(i=1,2…10).
The upper row reflects the Lotka's law. In this row, the distribution of A_i is given on the left, and on the right the corresponding double logarithmic presentation.
In the lower row, the distribution of the number of co-authorship couples N_{ij} is given on the left, and the corresponding triple logarithmic presentation on the right.

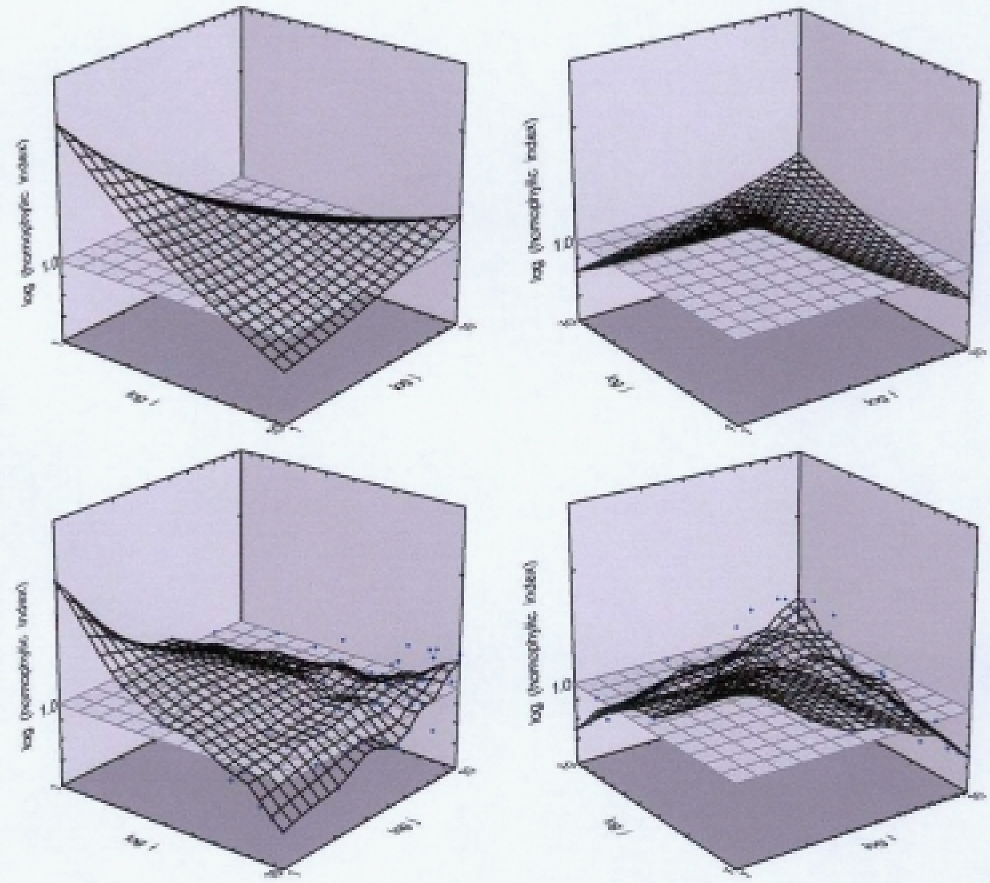

Fig. 2: Homophylic index.
The distribution of H'_{ij} resembles a plane surface, see the four cases. The figures on the right are rotated by 90^o.
In the upper row, the distribution of H'_{ij} is compared with the distribution of the theoretical values H_{Tij} (theoretical values determined through regression analysis). In the lower row, the distribution of H'_{ij} is compared with the distribution of the empirical values H_{ij}.
This figure clearly shows, that the general social structures – independent of the marginal sums N_i – have exercised an influence on the distribution of N_{ij}.

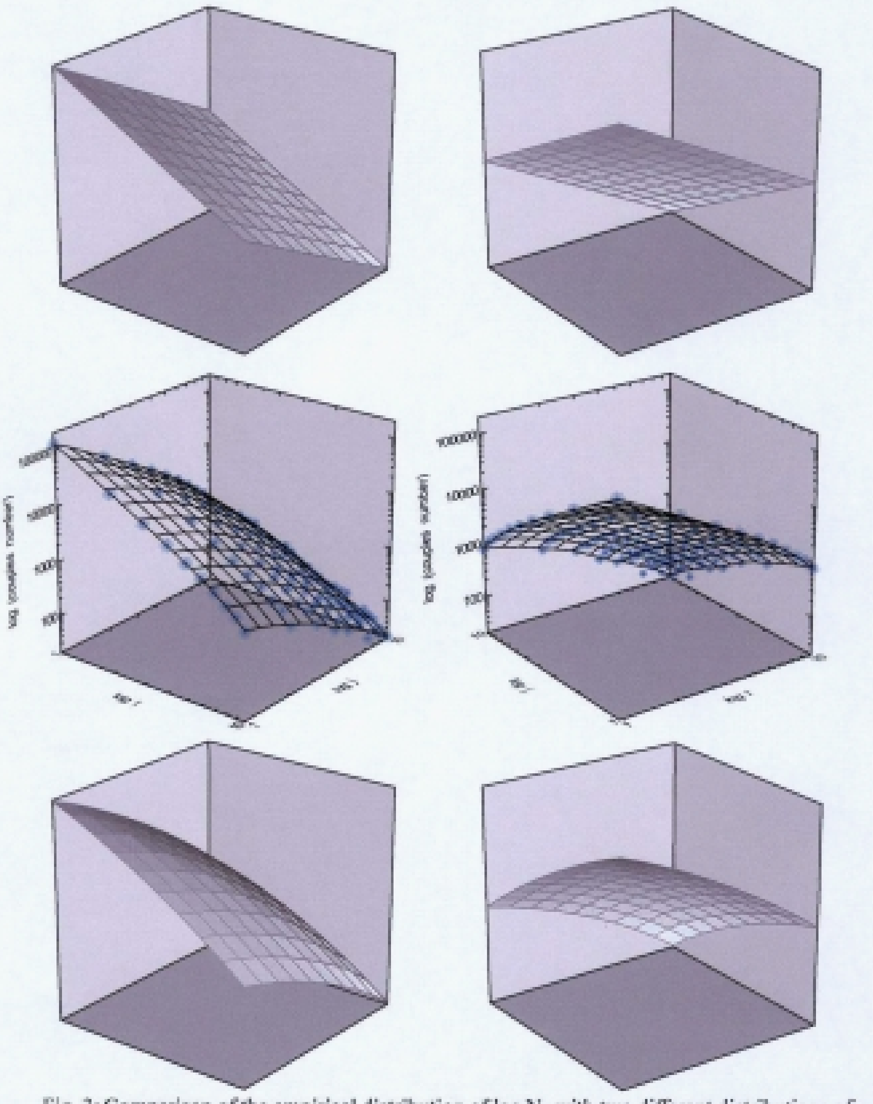

Fig. 3: Comparison of the empirical distribution of log N_{ij} with two different distributions of theoretical values. The figures on the right are rotated by 90°

In the first row, the distribution of theoretical values is shown under the condition:

log N''_{ij} =constant+a•(log i +log j). The 2^{nd} row shows the empirical distribution of log N_{ij}.

In the 3rd row the theoretical values are shown under the condition:

$\log N_{Tij} = const + \alpha \bullet \log(A + 1) + \beta \bullet (A_{max} - A + 1) + \gamma \bullet \log(B+1) + \delta \bullet \log(B_{max} - B + 1)$

7. Concluding Remarks

There are few general rules in bibliographic co-authorship networks available from the literature. A new approach is proposed by the author: Distribution of co-author couples in journals: “Continuation” of Lotka´s law on the 3rd dimension,

The following rules were found in the bibliographic co-authorship network of the journal *Science:*

- The number of collaborators of all authors with i publications per author is distributed according to an inverse power function in line with Lotka´s law.
- Beyond this, patterns were found based on general rules in social networks related to “Who is in contact with whom” and was published in Kretschmer, 1999.
- The distribution of co-author couples in journals on the 3rd dimension is independently influenced by:
- Lotka´s distribution and
- the general characteristic features in social systems

As mentioned above, the study of collaboration in e-science is one focus of the EU- project, WISER, including the questions dealing with the extent to which collaboration structures visible in the Web follow similar rules like collaboration networks measured by traditional bibliometric data.

The results of the size limited pilot study regarding the similarity of bibliographic co-authorship and Web networks are encouraging for testing the extent to which the rules in the distribution of co-author couples in journals with a large amount of data are also reflected in the Web network.

8. Acknowledgments

This work was supported by a grant from the Common Basis for Science, Technology and Innovation Indicators part of the Improving Human Research Potential specific programme of the Fifth Framework for Research and Technological Development of the European Commission. It is part of the WISER project (Web indicators for scientific, technological and innovation research) (Contract HPV2-CT-2002-00015) (www.webindicators.org).

The authors want to thank Holger Heitsch and Jan Johannes for programming. Editorial discussions with Colin Reddy are acknowledged.

9. References

1. Borgman, C & Furner, J. (2002). Scholarly communication and bibliometrics. In: Cronin, B. (ed.), Annual Review of Information Science and Technology 36, Medford, NJ: Information Today Inc., pp. 3-72.
2. Brookes, B.C. (1990). Biblio-, sciento-, infor-metrics??? What are we talking about? In: Egghe, L. & Rousseau, R. (Ed.). Informetrics (89/90). Amsterdam: Elsevier Science Publisher. 31-44
3. Cronin, B. (2001). Bibliometrics and Beyond: Some thoughts on Web-based citation analysis. Journal of Information Science, 27(1), 1-7.
4. Fox, E. A. & Urs, S. R. (2002), Digital libraries. In: Cronin, B. (ed.), Annual Review of Information Science and Technology 36, Medford, NJ: Information Today Inc., pp. 503-589.

5. Gopinath, M.A. (1992). Shiyali Ramamrita Ranganathan: A Profile in Relation to Librametry. In: Ravichandra Rao, I.K. (Ed.). Informetrics 91. Bangalore: Sarada Ranganathan Endowment Library Science, 9-16

6. Herring, S. C. (2002). Computer-Mediated Communication on the Internet. In: Cronin, B. (ed.), Annual Review of Information Science and Technology 36, Medford, NJ: Information Today Inc., pp. 109-168.

7. Ingwersen, P. (1998). The calculation of Web Impact Factors. Journal of Documentation, 54(2), 236-243.

8. Kling, R. & McKim, G. (2000). Not Just a Matter of Time: Field Differences in the Shaping of Electronic Media in Supporting Scientific Communication. Journal of the American Society for Information Science, 51(14), 1306-1320.

9. Kretschmer, H. (1999). A New Model of Scientific Collaboration. Part I: Types of Two Dimensional and Three-Dimensional Collaboration Patterns. Scientometrics. Vol.46.No.3 (1999) 501-518

10. Kretschmer, H. and R. Rousseau (2001). Author Inflation Leads to a Breakdown of Lotka's Law. Journal of the American Society for Information Science and Technology, 52(8), 610-614

11. Kretschmer, H. and M. Thelwall (2003): The development of information professionals: The European perspective: The way from librametry to webometrics. MALA: Madras. 13-25

12. Kretschmer, H & I. Aguillo (forthcoming). New indicators of gender visibility in Web networks.

13. Kretschmer, H & I. Aguillo (forthcoming). Visibility of collaboration networks on the Web Lotka, A.J. (1926). The frequency distribution of scientific production. Journal of the Washington Academy of Science, 16, 317-323.

14. Ranganathan, S.R. (1995, Reprint from 1969). Librametry and its Scope. The International Journal of Scientometrics and Informetrics, 1(1), 15-21.

15. Rousseau, R. (1992). Breakdown of the robustness property of Lotka's law: the case of adjusted counts for multi-authorship attribution. Journal of the American Society for Information Science, 43, 645-647

16. Sen, S.K.(1995). Introduction. The Int. Journal of Scientometrics and Informetrics, 1(1), 3-5.

17. Qin, J.(1995). Collaboration and publication productivity: An experiment with a new variable in Lotka´s law. In: Proceedings of the 5th International Conference on Scientometrics and Informetrics, June 7-10, 1995, River Forest (USA). M.E.D. Koenig and A. Bookstein (Eds.). Medford: Learned Information. 445-454

18. Vaughan, L. & Shaw Deborah (forthcoming), Bibliographic and Web citations: What is the difference?, (To be published in JASIST)

19. Vohora, S. B., Shah, Z. A. & Vohora, D. (2001). Low impact factors for Indian journals. Do factors other than quality influence? Current Science 81(8), 867.

20. Wasserman, S. & Faust, K. (1994). Social network analysis. Methods and applications. Cambridge: Cambridge University Press 1994

21. Winch, R. F., T. Ktsanes und V. Ktsanes (1954). The Theory of Complementary Needs in Mate-Selection: An Analytical and Descriptive Study. American Sociological Review, 19, 241-249

22. Wolf, Ch. (1996). Gleich und gleich gesellt sich. Individuelle und strukturelle Einflüsse auf die Entstehung von Freundschaften. Hamburg: Verlag Dr. Kovac.

About Authors

Mrs. Hildrun Kretschmer is Research Associate at Networked Research and Digital Information (NERDI), The Royal Netherlands Academy of Arts and Sciences (NIWI-KNAW), PO Box 95110, 1090 HC Amsterdam, The Netherlands. She is also the coordinator of COLLNET, Honorary Professor at Henan Normal University, China, Private Lecturer (PD) at Free University Berlin, Germany. She is professionally associated with varoius Universities/Institues such as NISTADS in New Delhi, India, University of New South Wales in Sydney, Australia, Dalian University of Technology in China, University of Leiden in the Netherlands, Bibliometrics, KAI eV in Berlin, Academy of Sciences of the GDR, Berlin, and Steel Factory, Hennigsdorf etc. She is an international expert in the areas of Bibliometrics, Informetrics, Scientometrics, Citation Studies and contributed huge number of reserarch papers and edited international conference proceedings and journals.
E-mail: Hildrun.Kretschmer@niwi.knaw.nl, kretschmer.h@t-online.de.

Mr. Theo Kretschmer is associated with COLLNET, Borgsdorfer Str. 5, D-16540 Hohen Neuendorf, Germany.
E-mail: kretschmer.h@onlinehome.de