

Gender bias in journals of gender studies

**Hiltrun Kretschmer, Ramesh Kundra,
Donald deB. Beaver & Theo Kretschmer**

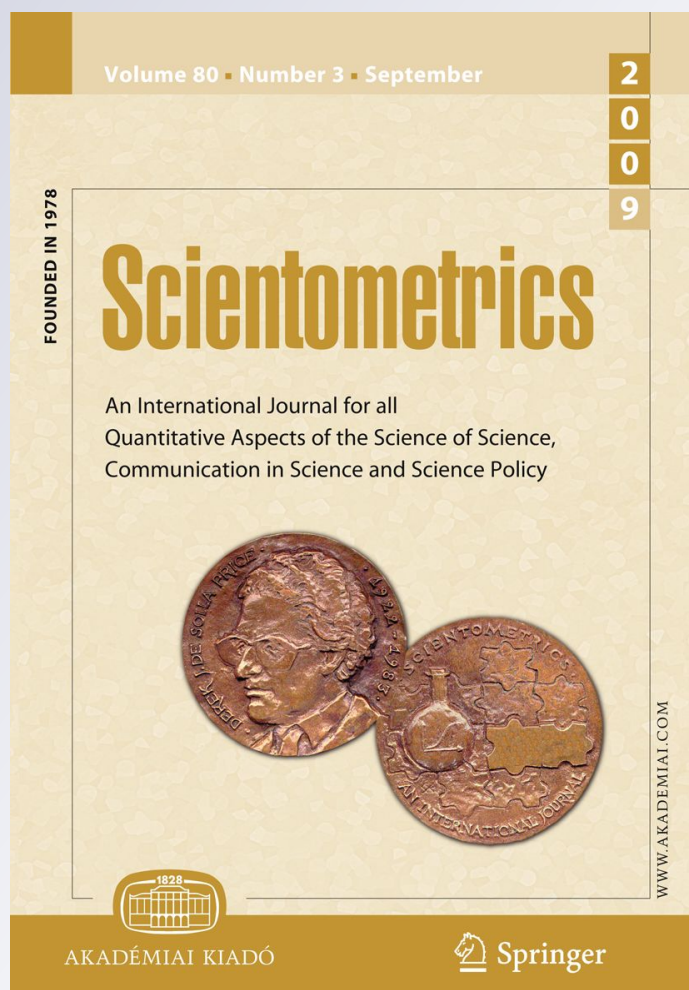
Scientometrics

An International Journal for all
Quantitative Aspects of the Science of
Science, Communication in Science and
Science Policy

ISSN 0138-9130

Scientometrics

DOI 10.1007/s11192-012-0661-5



Your article is protected by copyright and all rights are held exclusively by Akadémiai Kiadó, Budapest, Hungary. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Gender bias in journals of gender studies

Hildrun Kretschmer · Ramesh Kundra · Donald deB. Beaver ·
Theo Kretschmer

Received: 23 January 2012
© Akadémiai Kiadó, Budapest, Hungary 2012

Abstract The causes of gender bias favoring men in scientific and scholarly systems are complex and related to overall gender relationships in most of the countries of the world. An as yet unanswered question is whether in research publication gender bias is equally distributed over scientific disciplines and fields or if that bias reflects a closer relation to the subject matter. We expected less gender bias with respect to subject matter, and so analysed 14 journals of gender studies using several methods and indicators. The results confirm our expectation: the very high position of women in co-operation is striking; female scientists are relatively overrepresented as first authors in articles. Collaboration behaviour in gender studies differs from that of authors in PNAS. The pattern of gender studies reflects associations between authors of different productivity, or “masters” and “apprentices” but the PNAS pattern reflects associations between authors of roughly the same productivity, or “peers”. It would be interesting to extend the analysis of these three-dimensional collaboration patterns further, to see whether a similar characterization holds, what it might imply about the patterns of authorship in different areas, what those patterns might imply about the role of collaboration, and whether there are differences between females and males in collaboration patterns.

Keywords Gender bias · Co-operation · Social networks · Co-authorship · Collaboration patterns

H. Kretschmer (✉)
Faculty of Business Administration/Business Computing, University of Applied Sciences,
Bahnhofstrasse, 15745 Wildau, Germany
e-mail: kretschmer.h@onlinehome.de

H. Kretschmer · T. Kretschmer
COLLNET Center, Borgsdorfer Str. 5, 16540 Hohen Neuendorf, Germany

R. Kundra
Gurgaon, India
e-mail: r_kundra@yahoo.com

D. deB. Beaver
Williams College, 117 Bronfman Science Center, 18 Hoxsey St., Williamstown, MA 01267, USA
e-mail: dbeaver@williams.edu

Mathematical Subject Classification (2000) 62 · 68 · 91 · 94

JEL Classification C0 · C02 · C3 · C31 · C46

Introduction

The scientific and scholarly systems reflect a strong gender bias favoring men which makes it more difficult for women researchers to fully develop their potential and careers. The causes of that gender bias are complex and related to overall gender relationships in most of the countries of the world. An as yet unanswered question is whether in research publication gender bias is equally distributed over scientific disciplines and fields or if that bias reflects a closer relation to the subject matter.

Because we expected that gender bias might vary with respect to subject matter, we analysed 14 journals of gender studies using several methods and indicators. Our results are reported in three sections:

- Bibliometric indicators of gender co-operation
- Author order in the by-line and concentration measures
- Three-dimensional collaboration patterns of the journals “PNAS”, “Psychology of Women Quarterly” and of the mixed bibliography of 14 journals of gender studies

The special methods and indicators are explained in each section.

Data

Bibliometric analysis of the indicators in the 14 journals of gender studies (cf. Table 1) is based on a data sample of 8,649 papers published during the years 1976–2011 written by 12,691 authors in all; 10,867 of them are females and 1,823 males.

Table 1 Titles of the 14 journals

Journal	Number of papers	Number of authors
Affilia	620	1,058
Feminism and psychology	666	999
Gender and society	704	1,034
Gender technology and development	230	319
Men and masculinities	298	433
Psychology of women quarterly	1,111	2,473
Sexualities	427	536
Signs	1,396	1,676
Social politics	283	364
Women's studies international forum	1,631	2,056
European journal of women studies	366	454
Feminist theory	228	258
Indian journal of gender studies	291	373
Feminist economics	398	658
SUM	8,649	12,691

Bibliometric indicators of gender co-operation

Naldi and Parenti (2002) introduced three new bibliometric indicators and have used these indicators in order to process publications produced by co-operation among authors of different gender:

- Participation* counts the number of publications with at least one author of a given gender
- Contribution* measures the involvement of each gender in the production of a publication assuming that each author contributed the same amount
- Number of authors* total count of the authors of a given gender in each publication

Table 2 exemplifies the calculation of the three indicators in the case of a publication produced by four authors (Table 2 according to Naldi et al.).

Table 3 shows the results from three studies:

- Naldi et al. (2004)—Science and Technological Performance by Gender
- COLLNET (COLLNET-Collaboration Network in Science and in Technology: www.collnet.de)
- Journals of Gender Studies (Kretschmer, Kundra, Beaver and Kretschmer)

The COLLNET results in comparison with the results by Naldi et al. have already been published by Kretschmer and Aguillo (2004).

The bibliometric study of Naldi and Parenti (2002) is based on a data sample of 10,000 items published during the year 1995 in scientific journals of international relevance and written by 35,000 authors from six European countries. Women’s Participation amounted to only 45.8% of all items as opposed to the much greater male Participation of 94.7%. Women’s Contribution amounted to about 1/5 (19.5%), approximately the same as the Number of female authors, 22.2% of all authors. Although there are differences in these results related to disciplines and countries, in general the low position of women in co-operation is striking.

The bibliometric study of 64 COLLNET members from 20 countries examined lifetime productivity until June 2003. This study is based on a data sample of 223 multi-authored

Table 2 Calculation of female participation, contribution and total count

Gender				Female participation	Female contribution	Female total count
F	M	M	M	1	1/4	1
F	F	M	M	1	2/4	2
F	F	F	M	1	3/4	3
F	F	F	F	1	4/4	4

Table 3 Bibliometric indicators of gender co-operation

	Participation of women in %	Participation of men in %	Contribution of women, in%	Number of female authors in %
Naldi et al.	45.8	94.7	19.5	22.2
COLLNET bibliometric	65.3	76	45	47.9
14 Journals of gender studies	91.6	17.3	87.5	85.6

publications between at least two COLLNET members. Women's Participation amounted to 65.3% of all items and men's Participation 76%. Although the difference between the participation of women and the participation of men is statistically significant (χ^2 test, $p < 0.01$) it is clearly less than in the Naldi and Parenti study. Women's Contribution, 45%, almost equalled the Number of female authors, 47.9%. In sum, female and male COLLNET members are rather equally distributed in co-operation.

The results of the analysis of the co-operation among COLLNET members differ strongly from those of gender studies in the natural sciences (Naldi et al.) which show a very low participation rate of women in collaboration activities. Female COLLNET members' collaboration patterns are nearly equally as distributed as male members'. But further, the new results found in Journals of Gender studies even surpass the COLLNET results, insofar as female collaboration is concerned.

Bibliometric analysis of 14 journals of gender studies shows that Women's Participation amounted to 91.6% of all papers as opposed to the much lesser male Participation of only 17.3%. As before, women's Contribution, 87.5%, and the Number of female authors, 85.6%, were approximately equal. Although there are some minor differences in these results related to each journal individually, the very high position of women in co-operation is striking and most probably related to the subject matter of the journals.

Author order in the by-line and concentration measures

The order of the authors in the by-line is taken into consideration with help of concentration measures. The concentration of females (COF) in position x ($x = 1, 2, 3, 4, \dots$) in the by-line is defined here as the ratio between the percentage of females in position x and the percentage of females in total (in the present study of gender studies' journals: Number of female authors as a percentage of all authors = 85.6%).

If there is equal concentration by gender, the expected percentage of females in position x of the by-line should be equal to the percentage of females in total: $COF = 1$. If, for example, COF is higher than 1 in position 1 (first author), then females are relatively overrepresented as first authors, and vice versa if lower.

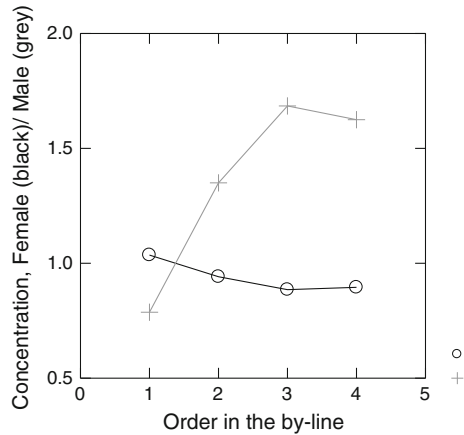
By analogy, the concentration of males (COM) in position x in the by-line is defined here as the ratio between the percentage of males in position x and the percentage of males in total (in the present study of gender studies' journals: Number of male authors as a percentage of all authors = 14.4%).

The results are presented in Fig. 1. In all of the journals the female scientists are relatively overrepresented on the first place in the by-line. This result confirms the above mentioned results obtained by bibliometric indicators of gender co-operation. However, on the contrary, from the second place on men are relatively overrepresented. Their slight downturn and the women's slight upturn for positions 4 or higher result from the very few collaborative papers with 7 or more authors, all of which are women.

Three-dimensional collaboration patterns of the journals PNAS, "Psychology of Women Quarterly" and of the mixed bibliographies of 14 journals

For about a decade social network analysis (SNA) has been used successfully in the information sciences, as well as in studies of collaboration in science. A variety of

Fig. 1 Concentration measure values of females (black: open circle) and males (grey: plus) according to author order in the by-line (1 first place, 2 second, 3 third, 4 fourth and higher order). Female scientists are relatively overrepresented in the first place in the by-line. However, from the second place on the men are relatively overrepresented



applications of SNA is available (Wasserman and Faust 1994; Otte and Rousseau 2002) both for studies in large and in small networks.

Special structures can be found in many networks, for example power-laws and others. In the present paper we present *three-dimensional special network structures that also occur in many networks*.

“Many investigations of scientific collaboration are based on statistical analyses of *large networks* constructed from bibliographic repositories. These investigations often rely on a *wealth of bibliographic data*, but very little or *no other information about the individuals in the network*, and thus, fail to illustrate the broader social and academic landscape in which collaboration takes place” (Pepe et al. 2009, p. 1).

In other words, in investigations of large networks information about “Who is collaborating with whom” is mostly missing.

However, the model of well-ordered three-dimensional distributions of co-author pairs’ frequencies in networks (Kretschmer and Kretschmer, 2007, 2009) says that, depending on the personal characteristics of collaborators (for example, author productivity or others), *a special fundamental principle of social group formation is a determining factor in shaping preferences in co-authorship between individual scientists*. This principle is based on similarities/dissimilarities and the corresponding consideration of this and other complementarities are a crucial determinant of the mathematical model.

Consequently, fundamental principles of social group formation produce well-ordered structures (called “Social Gestalts”) with different shapes depending on changing personalities and situations. This model has already been applied to 52 large co-authorship networks (Kretschmer and Kretschmer 2009). For 96% of them the squared multiple R is larger than 0.98 and for 77% of the 52 networks even larger than 0.99.

Method of counting co-author Pairs, based on SNA

For the purposes of analysis, a social network can be considered as consisting of two sets, a set of n nodes (individuals) and a set of m edges (undirected relations) between pairs of the nodes. The degree of a node F_x with x ($x = 1, 2, \dots, n$) is equal to the number of nodes (or

edges) that are attached to the node F_x . In co-authorship networks between two authors (nodes) F_x and F_y , there exists an edge if both have published at least one publication together.

An author's productivity is measured by his number of publications. The number of publications i per author F_x or j per possible co-author F_y , respectively, are determined by using the 'normal count procedure'. Each time the name of an author appears, it is counted.

The n authors F_x are grouped according to their productivities i or j , respectively. The co-author pairs of authors F_{xi} , (who have the number of publications i) in co-authorship with authors F_{yj} (who have the number of publications j), are counted. The resulting sum of co-author pairs N_{ij} is equal to the sum of degrees of the authors F_{xi} to the co-authors F_{yj} . Therefore, the matrix of N_{ij} is symmetrical (cf. Table 4).

In other words: N_{ij} is equal to the sum of co-author pairs of authors who have the number of publications i in co-authorship with authors who have the number of publications j . N is equal to the total sum of degrees of all n nodes (all authors F_x) in a network, equal to the total sum of pairs.

Logarithmic binning procedure

Distributions of this kind of co-author pairs' frequencies (N_{ij}) have already been published (Kretschmer and Kretschmer 2007; Kundra et al. 2008; Hanning et al. 2008). However, these distributions were restricted to $i_{max} = 31$.

Usually the stochastic noise increases with higher productivity because of the decreasing number of authors. We intend to overcome this problem in this paper with help of the *logarithmic binning procedure*. Newman has already proposed in 2005 using the logarithmic binning procedure for the log-log scale plot of power functions. To get a good fit of a straight line (log-log scale plot of power functions, for example Lotka's distribution), we need to bin the data i into exponentially wider bins. Each bin is a fixed multiple wider than the one before it. For example, choosing the multiplier of 2 we receive the intervals 1-2, 2-4, 4-8, 8-16, etc... For each bin we have ordered the corresponding first value of i (or j) to this bin. Thus, the sequence of bins i' or j' is:

i' ($i' = 1, 2, 4, 8, 16, 32, 64, 128, 256...$). The same holds for the bins j' . The sizes or widths of the bins ($\Delta i'$) are: 1, 2, 4, 8, 16 etc... The same holds for ($\Delta j'$).

However, because of the bivariate presentation the width of a bin ($cell_{i',j'}$) in the matrix is the product of $\Delta i'$ and $\Delta j'$ ($\Delta i' \cdot \Delta j'$). The sum of co-author pairs in a bin ($cell_{i',j'}$)

Table 4 Artificial table of co-author pairs N_{ij}

i/j	1	2	3	N_i
1	30	20	10	60
2	20	25	5	50
3	10	5	2	17
N_j	60	50	17	$N = 127$

Note $N_i = \sum_j N_{ij}$ is the sum of co-authors of all authors with i publications per author

$N_j = \sum_i N_{ij}$ is the sum of co-authors of all authors with j publications per author

$N =$ Total sum of degrees of all nodes in a network, equal to the total sum of pairs including F_x each, with x ($x = 1, 2, \dots, n$)

is called N_{ij}^S , cf. Tables 5, 6 and 7. The total sum of N_{ij}^S is equal to the total number of co-author pairs N of a co-authorship network:

$$N = \sum_{ij} N_{ij}^S.$$

Method of visualizing the original data

For visualizing the original data we use the sum of co-author pairs in a bin ($cell_{i'j'}$), i.e. N_{ij}^S directly in dependence on i' (bin) and j' (bin), (cf. Tables 5, 6 and 7). Because $\log 0$ is not given, we are using the value “0” for presentation of N_{ij}^S in the tables (cf. Tables 5, 6 and 7) but not for regression analysis.

Method of visualization the three-dimensional collaboration patterns

As the next step in the logarithmic binning procedure: N_{ij}^S of a cell ($cell_{i'j'}$) has to be divided by the width of the bin: $(\Delta i' \cdot \Delta j')$. In other words, the new value in a bin is simply the arithmetic average of all the points in the bin. This new value is called the average co-author pairs' frequency N_{ij}' .

Using the log–log–log presentation after the logarithmic binning procedure, the sequence of $\log i'$ (rows) is as follows: $\log i'$ ($\log i' = 0, 0.301, 0.602, 0.903, 1.204, 1.505, 1.806, \dots$); the same holds for $\log j'$ (columns).

In three-dimensional presentations $\log i'$ is placed on the X-axis, $\log j'$ on the Y-axis and $\log N_{ij}'$ on the Z-axis, cf. Figure 2 in the paragraph headed *Results*.

The mathematical function for describing the three-dimensional distribution of co-author pairs' frequencies (N_{ij} or N'_{ij} after logarithmic binning) is a *special case* derived

Table 5 Matrix of N_{ij}^S (PNAS) with $N = 634,014$

i' (bin)/ j' (bin)	1	2	4	8	16	32	64	Sum
1	13,2068	77,429	41,720	18,484	6,847	1,954	196	278,698
2	77,429	54,516	30,168	14,087	5,564	1,708	192	183,664
4	41,720	30,168	17,390	8,203	3,184	1,021	105	101,791
8	18,484	14,087	8,203	3,718	1,371	428	52	46,343
16	6,847	5,564	3,184	1,371	528	153	16	17,663
32	1,954	1,708	1,021	428	153	24	3	5,291
64	196	192	105	52	16	3	0	564
Sum	27,8698	18,3664	10,1791	46,343	17,663	5,291	564	$N = 634,014$

Table 6 Matrix of N_{ij}^S (Psychology of Women Quarterly) with $N = 4,324$

i' (bin)/ j' (bin)	1	2	4	8	Sum
1	2,228	499	186	143	3,056
2	499	154	53	52	758
4	186	53	34	24	297
8	143	52	24	12	231
Sum	3056	758	297	231	$N = 4,324$

Table 7 Matrix of N_{ij}^S (mixture of 14 journals in women's and gender studies) with $N = 11,996$

i' (bin)/ j' (bin)	1	2	4	8	16	Sum
1	6,332	1,415	424	205	44	8,420
2	1,415	574	204	69	31	2,293
4	424	204	160	50	8	846
8	205	69	50	22	3	349
16	44	31	8	3	2	88
Sum	8,420	2,293	846	349	88	$N = 11,996$

from Kretschmer's *mathematical model for the intensity function of interpersonal attraction* (cf. "Appendix").

We use this mathematical model of social Gestalts for describing co-author pairs' frequencies (Kretschmer and Kretschmer 2007, 2009) in form of the log-log-log presentation after logarithmic binning:

$$\log N'_{ij} = c + \alpha \cdot \log(|X - Y| + 1) + \beta \cdot \log(4 - |X - Y|) + \gamma \cdot \log(X + Y + 1) + \delta \cdot \log(7 - X - Y)$$

with $X = \log i'$ and $Y = \log j'$ and with $c = \text{constant}$.

For visualizing the theoretical patterns (Social Gestalts) we use the Function Plot of SYSTAT and the Scatterplot for the empirical patterns.

After regression analysis we obtain four parameters α , β , γ , and δ plus a constant c which are entered into the Function Plot (Z is the dependent variable and X and Y are the independent):

$$Z = c + \alpha \cdot \log(|X - Y| + 1) + \beta \cdot \log(4 - |X - Y|) + \gamma \cdot \log(X + Y + 1) + \delta \cdot \log(7 - X - Y)$$

Scale Range: The maximum and minimum values to appear on the axis are specified. Any data values outside these limits will not appear on the display. The minimum for the X -axis is specified as 0 ($(\log i')_{\min} = 0$) and the maximum is equal to $(\log i')_{\max}$ of the empirical data (For example, in Table 5: $(\log i')_{\max} = \log 64$). The same holds for the Y -axis ($\log j'$). The minimum and maximum values for the Z -axis are selected according to the minimum and maximum values of the whole Gestalt produced by the function. In case there are empirical values greater or less than these two theoretical values, the minimum or maximum of the Z -axis has to be extended accordingly. The Surface and Line Style dialog box is used to customize the appearance of lines or surfaces. The used XY Cut Lines are in two directions. The number of cuts in the grid has to be specified by the number of bins i' (or j' , respectively) minus 1 in the data set. For example, a special data set has 7 bins as in Table 5 (PNAS); the number of cuts in the grid is specified by $7 - 1 = 6$. The resulting number of lines of the theoretical pattern (Gestalt) is equal to the double of the number of bins i' ($2 \times 7 = 14$, cf. Fig. 2, first row in the paragraph *Results*). The number of points where two of the lines intersect, is equal to the square of the number of bins i' ($7^2 = 49$). The Scale Range of the empirical pattern has to be about equal to the theoretical Gestalt.

After the overlay of the empirical distribution and the theoretical pattern into a single frame the goodness-of-fit is highest in the case where the empirical values (dots) are

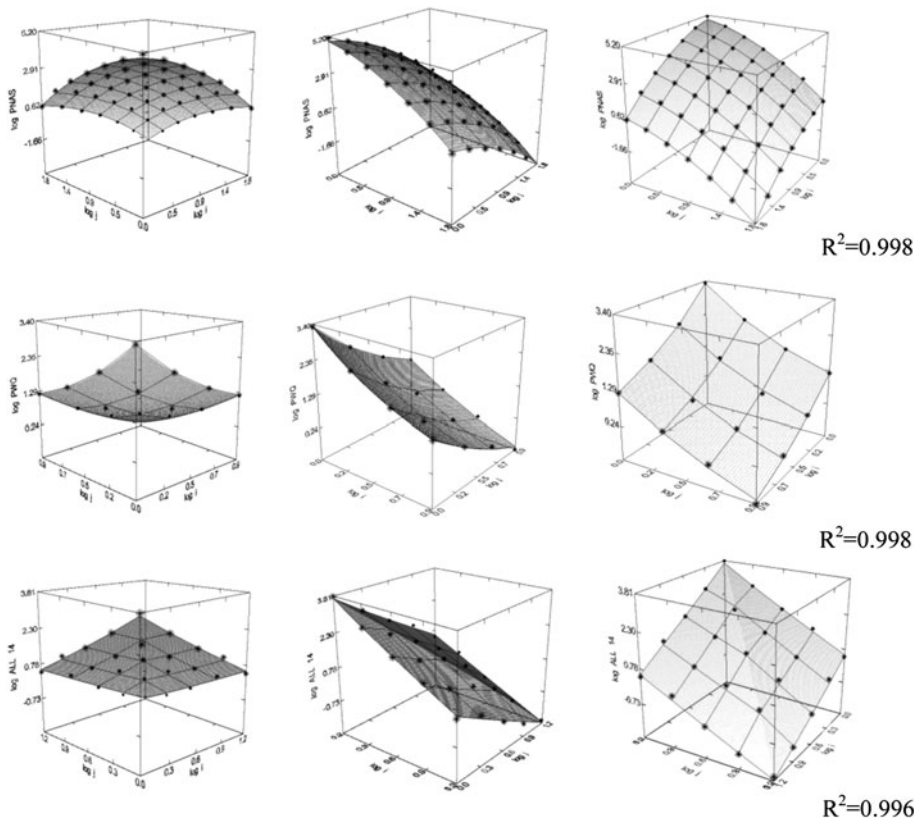


Fig. 2 Well-ordered distributions of co-author pairs' frequencies (Z-axis: $\log N_{ij}$) determined by the productivities of the co-authors (the logarithmic binning procedure is used: X-axis: $\log i'$ and Y-axis: $\log j'$). The leftmost patterns in each row are rotated *clockwise* as viewed from the top 90 degrees twice in succession, resulting in three patterns per row. First row: Distribution based on data obtained from the journal PNAS (1980–1998) Total number of co-author pairs (N) = 634,014; Authors: 80,058; Articles: 32,486, Regression analysis: $R^2 = 0.998$. Second row: Journal “Psychology of Women Quarterly” (1976–2011). $N = 4,342$; Authors: 2,569; Articles: 1,146; $R^2 = 0.998$, Third row: Mixture of 14 journals in women’s and gender studies (1976–2011); $N = 11,996$; Authors: 16,493; Articles: 5,990; $R^2 = 0.996$

directly placed on the points where two of the theoretical lines intersect. As the distance between the intersection points and the dots increases, the goodness-of-fit decreases.

Results

The results of the application of the model to three networks are shown in Fig. 2.

Three different well-ordered shapes of distributions of co-author pairs' frequencies (Z-axis: $\log N_{ij}$) reflecting the relative productivities of the co-authors (X-axis: $\log i'$ and Y-axis: $\log j'$ with i' or j' -number of publications of co-authors) are presented:

- The journal PNAS (1980–1998)
- Journal “Psychology of Women Quarterly” (1976–2011)

- Mixed bibliographies of 14 journals in women's and gender studies (1976–2011) are the source for this current study

All of these three-dimensional distributions are well-ordered with changing shapes. The shapes depend on the accentuation of either similarities or, vice versa, dissimilarities, cf. Fig. 2. Whereas the convex distribution in the first row obtained from the data of the journal *PNAS (1980–1998)* shows the accentuation of similarities of the co-authors regarding their productivities, the concave distribution in the second row obtained from the data of the journal "*Psychology of Women Quarterly*" (1976–2011) shows on the other hand the accentuation of dissimilarities of the co-authors. The shape of the distribution in the third row (source: Mixed Bibliographies of 14 Journals) falls in the middle between those of the *PNAS* and the journal "*Psychology of Women Quarterly*".

Explanation: In convex distributions (as in *PNAS*) the co-author pairs' frequencies between authors with the same number of publications are higher than those with different numbers of publications. Thus, accentuation of similarities is expressed by convex distributions.

On the contrary in concave distributions (like in the *Psychology of Women Quarterly*) the co-author pairs' frequencies between authors with the same number of publications are lower than those with different numbers of publications. Consequently, accentuation of dissimilarities is expressed by concave distributions.

An example for the theoretical predictions for the places of the empirical values in the theoretical patterns is also shown in Fig. 2. The lines of the theoretical patterns are obtained from the mathematical model. The points where two of the lines intersect are the theoretical predictions for the empirical values at those coordinates. The goodness-of-fit is highest in the case where the empirical values correspond exactly to the points obtained from the theoretical Gestalt (under these conditions we obtain after regression analysis: $R^2 = 1$). In Fig. 2 the empirical values are presented in form of dots.

Conclusion

In accordance with our expectations that potential gender bias may be related to the subject matter of journals, we have analysed fourteen journals of gender studies using several methods and indicators. We have obtained the following results:

1. The very large percentage of women in co-operation is striking and most probably related to the subject matter of the journals.
2. Female scientists are relatively overrepresented as first authors. This result confirms the above mentioned results obtained by bibliometric indicators of gender co-operation. In contrast to the women, however, from the second place on in the by-line men are relatively overrepresented.
3. The three-dimensional collaboration patterns are well-ordered, however, the shape is different from the well-ordered shape of *PNAS*. Collaboration behaviour in gender studies is different from that in the natural sciences. The accentuation of similarities in productivity of co-authors is shown in *PNAS* but the accentuation of dissimilarities can be observed in gender studies, especially in the Journal "*Psychology of Women Quarterly*"

The results confirm our expectation that the strength of gender bias is related to the subject matter of journals, and that it is less expressed in the journals of gender studies.

It would be interesting to extend the analysis of three-dimensional collaboration patterns further, to see whether such a characterization continues to hold, what it might imply about the patterns of authorship in different fields, what those patterns might imply about the role of collaboration, and whether there are differences between females and males in collaboration patterns.

Acknowledgment Part of this work by one of the authors (Kretschmer H) was supported by the 7th Framework Program by the European Commission, SIS-2010-1.3.3.1. Project full title: “Academic Careers Understood through Measurement and Norms “, Project acronym: ACUMEN.

Appendix

Theory and mathematical model for the intensity function of interpersonal attraction

The mathematical function for describing the three-dimensional distribution of co-author pairs' frequencies (N_{ij}) is a special case derived from Kretschmer's *mathematical model for the intensity function of interpersonal attraction* (*Who is attracting whom? “Intensity” means the extent of this attraction*).

In the wake of a tangible change of paradigm in science a number of holistic theories have emerged which are based on the idea of holographic interacting entities in the world, with several of them also implying a *field concept*.

For example:

- magnetic fields in physics
- morphogenetic fields of living organisms in evolutionary biology
- psychological fields in psychology or sociology (Gestalts)
- etc.

The field concept says a force, which emanates from a field generates a balanced evenness among all the individual components taken in their totality. However, the field fails to determine completely the behaviour of individual components in terms of the predictability of these individual components.

This is called *conciseness* (or ‘*Prägnanz*’) *tendency* in Gestalt psychology, i.e. there is a *'tendency towards a good Gestalt'* of the totality. The stable final state is, if possible, built up in a simple, well-ordered, harmonic and uniform manner in line with definite rules. Several authors take the view that these fields can be mathematically described.

Interpersonal attraction is a major area of study in social psychology.

Whereas in physics, attraction may refer to gravity or to the electromagnetic force, *interpersonal attraction can be thought of force acting between two people tending to draw them together.*

When measuring interpersonal attraction, one must refer to the *qualities of the attracted* as well as *the qualities of the attractor*. *That means one must refer to their personal characteristics.* For example, in terms of the degree of the node F_x and the degree of the node F_y (Newman 2002) or in terms of productivity: $X = \log i$ of co-author F_x and $Y = \log j$ of co-author F_y (Kretschmer and Kretschmer 2007, 2009).

The notion of “birds of a feather flock together” points out that *similarity is a crucial determinant of interpersonal attraction.*

But: Do birds of a feather flock together or do opposites attract?

This leads to a *model of complementarities: Complementarities are a crucial determinant* of the Intensity Function of Interpersonal Attraction.

Derivation of the Intensity Function of Interpersonal Attraction:

We assume the *intensity structure of mutual attraction* Z_{XY} can be described by a function of a special power functions' combination (X is the value of a special personality characteristic (quality) of an attracted and Y is the value of the same personality characteristic (quality) of the attractor and in case of mutual attraction also vice versa).

The *crucial determinant of interpersonal attraction (similarity or dissimilarity)* suggests considering the *distance* A between the qualities of persons ($A = |X - Y|$) as the independent variable of a power function:

$$Z^* = c_1 \cdot (A + 1)^\alpha$$

with $c_1 = \text{constant}$; the 1 is added because $\log A$ is not possible in case $A = 0$. We see that *as A increases, dissimilarity increases*.

A power function with only one parameter (unequal to zero) is either only monotonically decreasing or only monotonically increasing; when referred to both proverbs we obtain: *either "birds of a feather flock together" or "the opposites attract"*, cf. Fig 3.

In order to fulfil the inherent requirement that both proverbs with their extensions can be included in the representation, the second step of approximation follows.

Information in brief: There is a *complementary variation of similarity and dissimilarity*. As dissimilarity increases between persons, similarity decreases, and vice versa. Similarity is greatest at the minimum of A and least at the maximum and vice versa, dissimilarity is greatest at the maximum and least at the minimum.

A is a variable with the two opposite poles A_{\min} and A_{\max} . The sum of A_{\min} and A_{\max} is a constant. Thus,

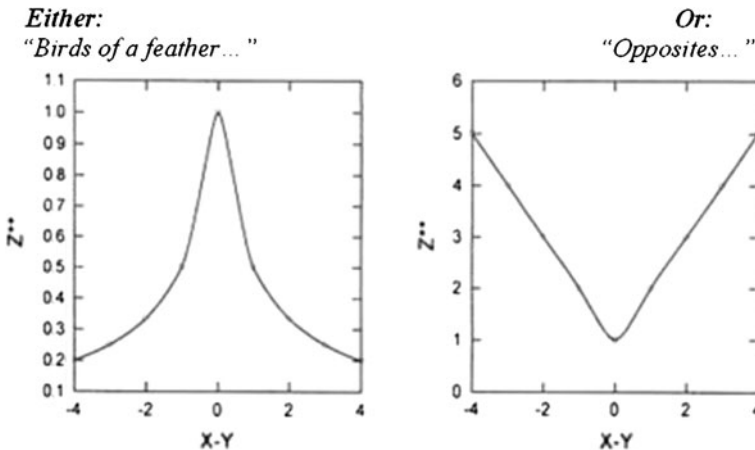


Fig. 3 Power functions with different values of parameter α (non-log presentation). In both patterns $X - Y$ is the abscissa with $X - Y = 0$ (similarity is highest) in the *middle* and Z^* is the ordinate. On the *left* pattern, the parameter α is negative: "Birds of a feather flock together", i.e. decrease of interpersonal relations with increasing dissimilarity. On the *right* pattern, the parameter α is positive: "Opposites attract", i.e. increase of interpersonal relations with increasing dissimilarity (this figure is a copy of a figure in Kretschmer and Kretschmer 2007)

$$A_{\text{complement}} = A_{\text{min}} + A_{\text{max}} - A$$

That means, the variable $A_{\text{complement}}$ increases by the same amount as the variable A decreases and vice versa.

Example: $A_{\text{min}} = 0, A_{\text{max}} = 3$:

A	$A_{\text{complement}}$
0	3
1	2
2	1
3	0

- the model of *complementarities* leads to the conclusion to use additionally the “complement of the distance A ” ($A_{\text{complement}}$) as the independent variable of a second power function:

$$Z^{**} = c_2 \cdot (A_{\text{complement}} + 1)^\beta$$

$$Z_A = \text{constant}_A \cdot (A + 1)^\alpha \cdot (A_{\text{complement}} + 1)^\beta.$$

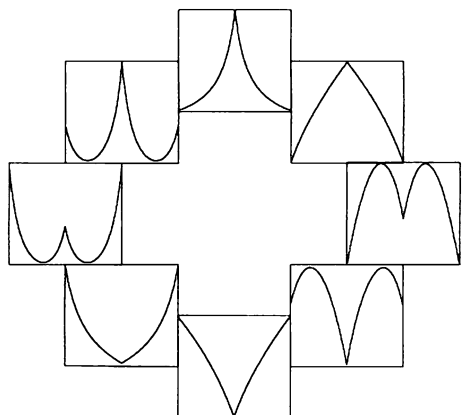
The relationships of the two parameters α and β to each other determine the expressions of the complementarities (similarities, dissimilarities) in each of the eight shapes, cf. Fig. 4. In correspondence with changing relationships of the two parameters α and β to each other a systematic variation is possible from “Birds of a feather flock together” to “Opposites attract” and vice versa.

While in the upmost pattern “Birds of the feather flock together” is more likely to be in the foreground, the bottom pattern reveals that “Opposites attract” is more likely to be salient.

Starting pattern by pattern counter clockwise from the upmost pattern towards the bottom pattern, “Birds of the feather flock together” diminishes as “Opposites attract” emerges. Vice versa, starting pattern by pattern counter clockwise from the bottom pattern towards the upper pattern, “Opposites attract” diminishes as “Birds of the feather flock together” emerges.

For the *purpose of completion*,

Fig. 4 Patterns with varying combinations of the two parameters α and β (non-log presentation). In all of the eight patterns $X - Y$ is the abscissa with $X - Y = 0$ in the middle and Z_A is the ordinate



- Let the addition ($B = X + Y$) as the *opposite* of subtraction ($A = |X - Y|$), be the independent variable of the third power function

$$Z^{***} = c3 \cdot (B + 1)^7$$

- and the complement ($B_{\text{complement}}$) be the independent variable of the fourth power function

$$Z^{****} = c4 \cdot (B_{\text{complement}} + 1)^\delta$$

In analogy to A and $A_{\text{complement}}$:

$$B_{\text{complement}} = B_{\text{min}} + B_{\text{max}} - B$$

$$Z_B = \text{constant}_B \cdot (B + 1)^\gamma \cdot (B_{\text{complement}} + 1)^\delta$$

Because the function Z_A can vary independently from the function Z_B we assume the intensity of mutual attraction Z_{XY} is proportional to the product of the two functions Z_A and Z_B :

$$Z_{XY} \sim Z_A \cdot Z_B$$

Therefore, the Intensity Function of Interpersonal Attraction (Social Gestalt) can be formalized as follows (Prototypes of Social Gestalts, cf. Fig. 5):

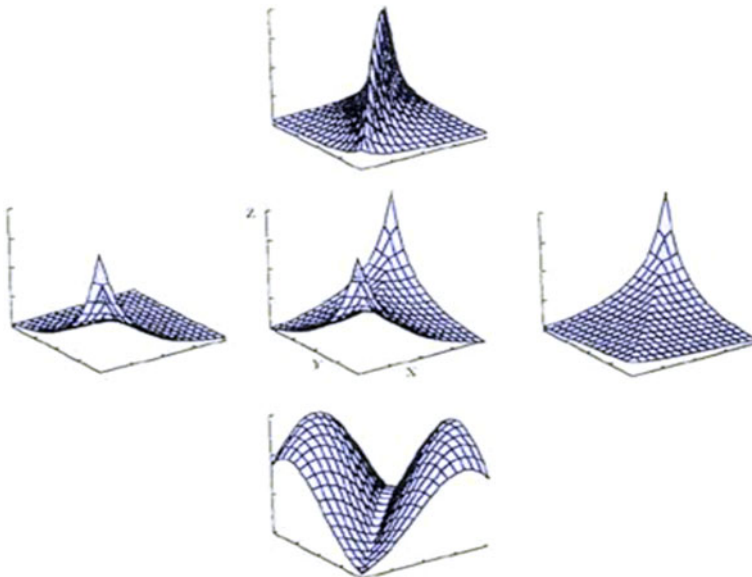


Fig. 5 Prototypes of social Gestalts (non-logarithmic presentation). Several empirical patterns matching the five Prototypes were already taken out and presented in Kretschmer (2002). The distribution of co-author pairs' frequencies N_{ij} is one of the examples. The non-logarithmic presentation is similar to the left prototype. However, in this paper we are showing the corresponding log-log-log presentation only ($\log N_{ij}$ with $\log i$ and $\log j$)

$$Z_{XY} = \text{constant} \cdot (A + 1)^\alpha \cdot (A_{\text{complement}} + 1)^\beta \cdot (B + 1)^\gamma \cdot (B_{\text{complement}} + 1)^\delta$$

with $A = |X - Y|$ and $B = X + Y$

$$A_{\text{complement}} = A_{\text{min}} + A_{\text{max}} - A$$

$$B_{\text{complement}} = B_{\text{min}} + B_{\text{max}} - B$$

$$A_{\text{min}} = (|X - Y|)_{\text{min}}$$

$$A_{\text{max}} = (|X - Y|)_{\text{max}}$$

$$B_{\text{min}} = (|X + Y|)_{\text{min}}$$

$$B_{\text{max}} = (|X + Y|)_{\text{max}}$$

Measurement of the variables X , Y and Z_{XY} including $X_{\text{min}} = Y_{\text{min}}$ and $X_{\text{max}} = Y_{\text{max}}$ depends on the subject being studied.

Examples (types) of social interactions (Z_{XY}) are collaboration, friendships, marriages, etc., while examples (types) of characteristics or of qualities of these individual persons (X or Y) are age, labor productivity, education, professional status, degree of a node in a network, etc.

Whereas Z_A and Z_B are each alone produce two-dimensional patterns, the bivariate function Z_{XY} shows three-dimensional patterns (non-logarithm presentation).

We show one example of how to measure the variables X and Y in relation to the function of the distribution of co-author pairs' frequencies $Z_{XY} = N_{ij}$. The physicist and historian of science de Solla Price (1963) conjectured that the logarithm of the number of publications has greater importance than the number of publications per se.

Thus, using the logarithm of the number of publications ($\log i$ or $\log j$, respectively) as an indicator of the personal characteristic 'productivity', we define:

$$X = \log i \quad Y = \log j$$

$$A = |\log i - \log j| \quad B = \log i + \log j$$

Consequently:

$$A_{\text{min}} = |X - Y|_{\text{min}} = 0 \text{ with } \log i = \log j$$

$$A_{\text{max}} = |X - Y|_{\text{max}} = |(\log i)_{\text{max}} - \log 1| = |\log 1 - (\log j)_{\text{max}}| = (\log i)_{\text{max}} = (\log j)_{\text{max}}$$

$$B_{\text{min}} = (X + Y)_{\text{min}} = \log 1 + \log 1 = 0$$

$$B_{\text{max}} = (X + Y)_{\text{max}} = (\log i)_{\text{max}} + (\log j)_{\text{max}} = 2(\log i)_{\text{max}} = 2(\log j)_{\text{max}}$$

Let us assume a specific value for the maximum possible number of publications i (or j , respectively) of an author as a standard for such studies, which does not vary depending upon the given sample. We assume that the maximum possible number of publications of an author is equal to 1000, i.e.

$$A_{\text{max}} = \log 1000 = 3 \quad B_{\text{max}} = 2A_{\text{max}} = 6$$

Thus, it follows that:

$$A_{\text{COMPLEMENT}} = 3 - |\log i - \log j|, \text{ with } A_{\text{COMPLEMENT}} + 1 = 4 - |\log i - \log j|$$

$$B_{\text{COMPLEMENT}} = 6 - (\log i + \log j), \text{ with } B_{\text{COMPLEMENT}} + 1 = 7 - (\log i + \log j)$$

$$= 7 - \log i - \log j$$

Thus, the theoretical mathematical function for describing the social Gestalts of the distribution of co-author pairs' frequencies results in the previously mentioned logarithmic version ($\log N_{ij}$):

$$\log N_{ij} = c + \alpha \cdot \log(|X - Y| + 1) + \beta \cdot \log(4 - |X - Y|) + \gamma \cdot \log(X + Y + 1) + \delta \cdot \log(7 - X - Y)$$

with $X = \log i$ and $Y = \log j$ and with $c = \text{constant}$.

References

- Hanning, G., Kretschmer, H., & Liu, Z. (2008). Distribution of co-author pairs 'frequencies of the Journal of Information Technology. *COLLNET Journal of Scientometrics and Information Management*, 2(1), 73–81.
- Kretschmer, H. (2002). Similarities and dissimilarities in co-authorship networks; gestalt theory as explanation for well-ordered collaboration structures and production of scientific literature. *Library Trends*, 50(3), 474–497.
- Kretschmer, H., & Aguillo, I. F. (2004). Visibility of collaboration on the Web. *Scientometrics*, 61(3), 405–426.
- Kretschmer, H., & Kretschmer, T. (2007). Lotka's distribution and distribution of co-author Pairs' frequencies. *Journal of Informetrics*, 1, 308–337.
- Kretschmer, H., & Kretschmer, T. (2009). Invited keynote speech. Who is collaborating with whom? Explanation of a fundamental principle. In: H. Hou, B. Wang, S. Liu, Z. Hu, X. Zhang, M. Li (Eds.), *Proceedings of the 5th International Conference on Webometrics, Informetrics and Scientometrics and 10th COLLNET Meeting*, 13–16 September 2009, Dalian, China (CD-ROM for all participants and for libraries).
- Kundra, R., Beaver, D., Kretschmer, H., & Kretschmer, T. (2008). Co-author pairs' frequencies distribution in journals of gender studies. *COLLNET Journal of Scientometrics and Information Management*, 2(1), 63–71.
- Naldi, F., Parenti, I.V. (2002). Scientific and technological performance by gender: a feasibility study on patent and bibliometric indicators. Vol. II: methodological report. European Commission Research, EUR 20309.
- Naldi, F., Luzi, D., Valente, A., & Parenti, I. V. (2004). Scientific and technological performance by gender. In H. F. Moed, et al. (Eds.), *Handbook of quantitative science and technology research* (pp. 299–314). The Netherlands: Kluwer Academic Publishers.
- Newman, M. E. J. (2002). Assortative mixing in networks. *Physical Review Letters*, 89, 208701.
- Newman, M. E. J. (2005). Power laws, pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323–351.
- Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, 28, 443–455.
- Pepe, A., & Marko, A. R. (2009). Collaboration in sensor network research: an in-depth longitudinal analysis of assortative mixing patterns. *Scientometrics*, 84(3), 687–701. This article is published with open access at Springerlink.com.
- Price, D. de Solla (1963). *Little science, big science*. New York: Columbia University Press.
- Wasserman, S., & Faust, K. (1994). *Social network analysis. Methods and applications* (p. 1994). Cambridge: Cambridge University Press.