

Varying Shapes of Co-author Pairs' Distributions^{*}

Hildrun Kretschmer^{1,2}

Theo Kretschmer²

10 September 2008

Abstract

There are different versions of counting coauthor pairs. We refer to two of them in the present paper including two corresponding hypotheses. Whereas the first version leads to a power function distribution of the co-author pairs the other version shows a bivariate distribution of co-author pairs' frequencies hence producing three-dimensional graphs. The regularities for the second kind of distributions are very different from simple power law distributions. These new regularities may be described by a model for the intensity function of interpersonal attraction (Social Gestalt). Complementarities are a crucial determinant of this model. In conjunction with these complementarities, various shapes of the distribution of observed co-author pairs' frequencies emerge. As the difference between the two shapes increases, the correlation between these two shapes decreases. After mixing these two shapes a new shape of social Gestalt emerges which can be successfully described with the help of the mathematical function for a social Gestalt. The change of any shape to any other shape of the social Gestalt can be shown by simulation (or morphing). For comparison of two or more shapes the overlay of these shapes into a single frame is possible: either overlay of standardized shapes or of the originals.

The present paper is focused on a theoretical approach verified by empirical studies explained in two other papers published in the Proceedings (Guo et.al.2008 and Kundra et.al, 2008). These three papers are also published in the June 2008 issue of the *COLLNET Journal of Scientometrics and Information Management, Vol.2, No 1, 2008, 45-81*(in detail: cf. References).

1. Introduction and Hypotheses

Power laws are successful in describing complex networks of interactions. Scale-free network models describe many natural and social phenomena, for example networks of interacting components of a living cell or social networks (Guimerà, R., et al., 2003).

We refer in our paper to social networks in scientific communities and we present power law distributions if one counts the number of publications of co-author pairs and furthermore other regularities *different* from *simple* power law distributions. Those regularities are discussed in conjunction with a *model for the intensity function of interpersonal attraction (Social Gestalt)*.

Lotka's law (1926) is based on *single scientists P counting*. With increasing collaboration in science and in technology the study of the *frequencies of pairs P,Q* or triples of co-authors P,Q,R ... etc, is highly relevant.

However, *there are different versions of counting pairs*. We refer to two of them in the present paper including the corresponding two hypotheses.

^{1st} Counting the number of publications of co-author pairs: The pairs are *counted as units (P,Q)* in analogy to single authors P in Lotka's Law, where k is the number of joint publications of the pair P,Q (For example: Smith & Miller).

We assume there is a regularity for the distribution of coauthor pairs' frequencies B_k with k publications per co-author pair (P,Q) in form of a power law distribution:

$$B_k=f(k)=B_k=c/k^b,$$

^{*} This paper is published in: *COLLNET Journal of Scientometrics and Information Management Vol.2, No 1(2008),45-61*

¹WISELAB, Dalian University of Technology, Dalian, 116023, China

²COLLNET Center, Borgsdorfer Str. 5, D-16540 Hohen Neuendorf, Germany, kretschmer dot h at onlinehome dot de

with $c=\text{constant}$ (1)

Morris, S. A. and M. L. Goldstein (2007) have already shown this kind of regularity in one of their empirical studies. In this connection Egghe (to appear) has presented a theoretical model for the size-frequency function of co-author pairs.

2nd The pairs P,Q are counted *under the condition of both the first authors P count (i) and the second authors Q count (j)*. A bivariate distribution of co-author pairs' frequencies N_{ij} is studied hence producing three-dimensional graphs.

In a former paper (Kretschmer and Kretschmer 2007) we have verified there are regularities for the distribution of $N_{ij}=f(i,j)$, different from a simple power law distribution. These regularities can be described by a model of social Gestalts:

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

with $c=\text{constant}$ (2)

Complementarities are a crucial determinant of this model. We expect, in conformity with these complementarities there are varying shapes of the distributions of co-author pairs' frequencies in correspondence with varying shapes of the social Gestalt. This phenomenon is different from simple power law functions.

The present paper is focused on a theoretical approach verified by empirical studies explained in two papers published in the Proceedings and in the June 2008 issue of the *COLLNET Journal of Scientometrics and Information Management. Vol.2, No1*, together with the present paper:

- Guo, Kretschmer & Liu, 2008
- Kundra, deB. Beaver, Kretschmer & Kretschmer, 2008

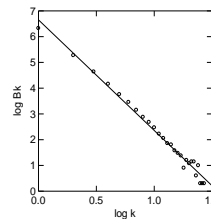
2. Power Law Distributions of Co-author Pairs' Frequencies B_k

We have studied the regularities for distributions if one counts the number of publications of the

co-author pairs both in the *Journal of Biochemistry* and in the data set of the combination of four high impact SCI journals (*Science, Nature, PNAS and Phys Rev B Condensed Matter*) from 1980-1998. We could find in both studies power law distributions, Fig. 1.

$B_k=c/k^b$, with $c=\text{constant}$

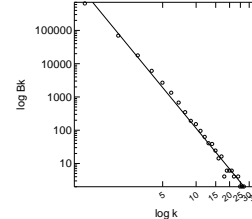
Mix of Four High SCI Impact Journals



$n=15$
 $R=0.996, R^2=.991$
 $F\text{-Ratio}=2777$

$B_k=6.13-4.07 \log k$

Journal of Biochemistry



$n=26$
 $R=0.987, R^2=.975$
 $F\text{-Ratio}=508.65$

$B_k=5.39-4.539 \log k$

Fig. 1: Power Law Distributions of Co-author Pairs' Frequencies B_k

Although power law functions are successful in describing complex networks of interactions, these functions are only descriptive having no explanatory features as expected in the function (1) of co-author-pairs' frequencies.

However, considering the co-author pairs' frequencies N_{ij} means, a *new quality* is involved because we have to consider the question: who is in contact with whom. In contrast to simple power functions a rather new function (2) of co-author-pairs' frequencies N_{ij} (Kretschmer 1999, 2002) having both descriptive and explanatory features will be discussed in this paper.

The rather new version of counting pairs leads to bivariate distributions hence producing three-dimensional graphs. In comparison with the power law distributions of co-author pairs in Fig. 1 this paper is focused on graphs, very different from Fig. 1.

3. Bivariate Distributions of Co-author Pairs' Frequencies N_{ij} : A Short Introduction

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 (For example, in terms of age or of productivity, i.e. of *i* and *j*) to achieve predictive accuracy. It is suggested that to determine attraction, *personality and situation (or environment)* must be taken into account. The notion of "birds of a feather flock together" points out that *similarity is a crucial determinant* of interpersonal attraction. Do birds of a feather flock together, or the opposites attract? This leads to a model of *complementarities* (We refer in this connection to the varying shapes of the social Gestalt to be explained in section 4).

The modern notion of complementarities introduced by *Niels Bohr* had existed already in a clear-cut manner in old Chinese thought, in the Yin/Yang teaching. Yin and Yang have to be seen as polar forces of only one whole, as *complementary tendencies interacting dynamically with each other*, so that the *entire system is kept flexible and open to change* (For example, expressed in varying shapes of the social Gestalt).

Based on this background Kretschmer (1999, 2002) has already created a model for social Gestalts valid for social networks in general. This model is also applied for description of the distribution of co-author pairs' frequencies N_{ij} (Kretschmer & Kretschmer 2007).

The fundamental "formula" of Gestalt theory might be expressed in this way: there is a *whole*, the behaviour of which is not determined by that of their individual elements, but where the *part-processes* are themselves determined by the *intrinsic nature of the whole* (Gestalt Theory by Max Wertheimer, 1924).

With this in mind, the *distribution of observed co-author pairs' frequencies can be considered to be a special reflection of a social Gestalt* (Further details, cf. section 4.4).

The model of social Gestalts mathematically describes, and textual explains (cf. section 4.1), *well-ordered fields of mutual attraction* between a large numbers of individual persons. In conformity with complementarities or Yin/Yang theory these fields change their

shapes depending on *changing personalities and situations*.

Although the field of mutual attraction (or social Gestalt) fails to determine completely individual pairs of persons in terms of the predictability of these individual pairs, the force that emanates from this field generates a statistically balanced evenness among all the individual pairs in their totality (*Tendency towards a good social Gestalt*). This statistically balanced evenness is enhanced as the number of co-author pairs' frequencies rises. We refer to *Bernoulli's law of large numbers*.

Obtaining large sample sizes is possible in studies of large bibliographies (journals or research areas). On the other hand, the phenomenon of Gestalts could be considered as originally related to morphogenetic fields or morphogenetic Gestalts: Overlapping or mixing of co-author pairs frequencies' distributions in analogy to overlapping of faces (morphing).

The result of the superposition mixed or average faces was more attractive than the individual ones and these mixed faces became more attractive, the greater the number of overlappings. The 'more attractive' faces were considered more well-ordered and harmonic and generally more proportional.

By analogy, we have shown the higher the number of mixed distributions of co-author pairs' frequencies, the higher becomes the tendency towards a good social Gestalt overall (Kretschmer & Kretschmer 2007: Paragraphs 3.2.3 and 7.3, Fig. 2). In another example, after overlapping or mixing eleven distributions of co-author pairs in women's and gender studies the data reveal the same tendency to a good Gestalt (cf. in this issue: Kundra, deB. Beaver, Kretschmer and Kretschmer 2008, Fig. 3)

In summary:

The model of social Gestalts can become successfully applied to the study of a large number of individual persons. However, in the case of *small sample sizes* the tendency towards a good or well-ordered social Gestalt can be shown after overlapping or mixing of several co-author pairs' frequencies distributions.

In conjunction with the model of complementarities, there are different shapes of a social Gestalt depending on changing personality and

situation or environment (For example, one of the empirical proofs is given in Kretschmer, Liang, Kundra 2001). A copy of the corresponding figure can be found in Fig. 3.

In conclusion, *as the difference between two shapes increases, the correlation between these two shapes decreases!* At first glance, the question arises whether mixing of such shapes is suitable under these conditions.

However despite first doubts, after mixing these two shapes *a new shape of social gestalt* emerges and can be successfully described with help of the mathematical function of social Gestalt (cf. section 4.3).

The change of any shape to any other shape of the social Gestalt can be shown by simulation (or morphing) (cf. in this issue: Kundra et.al. 2008, Fig. 6).

In conclusion, mixing of several distributions of co-author pairs' frequencies has two functions:

- Generating a statistically balanced evenness or a tendency towards a good social Gestalt respectively
- Emergence of distributions reflecting a new emerging shape of a social Gestalt

For comparison of two or more shapes the overlay of these shapes into a single frame is possible: either overlay of standardized shapes or of the originals. The overlay of two original shapes, i.e. of the empirical patterns of the *Journal of Information Technology* with the *Journal of Biological Chemistry* is shown in this issue: Guo, Kretschmer & Liu, 2008, Fig. 5.

4. Complementarities and Varying Shapes

4.1 Model for the Intensity Function of Interpersonal Attraction (Social Gestalt)

The function of N_{ij} in (2) is a special case of the social Gestalts in social networks described by the general function Z_{XY} in (11). The explanation for the special case is given in section 4.4. Thus, starting with the general model of social Gestalts we will focus later on the distribution of co-author pairs' frequencies.

Whereas the empirical shapes of the Lotka's power law distributions are rather similar with each other the empirical shapes of the three dimensional distributions show *manifold shapes*.

A few of the already published shapes are shown in Fig. 2 (First row, left picture: Kretschmer 1999, Fig. 2; right picture: Kretschmer, Liang, Kundra, 2001, Fig. 1. Second row, left picture: Kretschmer & Kretschmer 2007, Fig. 1; right picture: Guo et al. 2008 in this issue, Fig. 3). The pictures in the first row are obtained by the distributions of ratios between observed to statistical expectation values of co-author pairs' frequencies. (For more details, cf. section 4.4) The pictures in the second row show the results of triple logarithmic presentations ($\log N_{ij}$).

Because of the phenomenon mentioned above we need a new function for possible description of all of these manifold shapes!

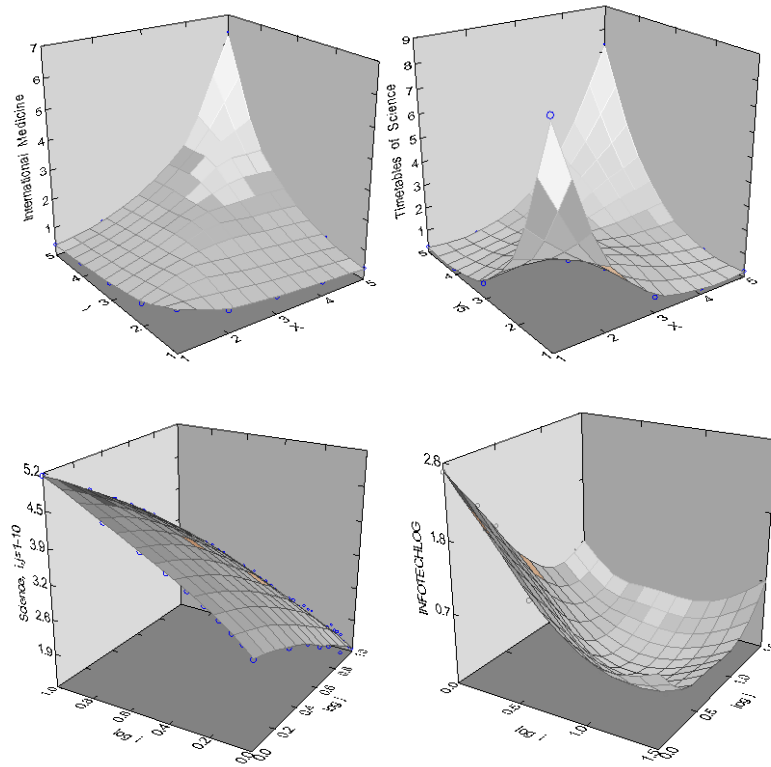


Fig. 2: Several Empirical Shapes of Co-authorship Patterns

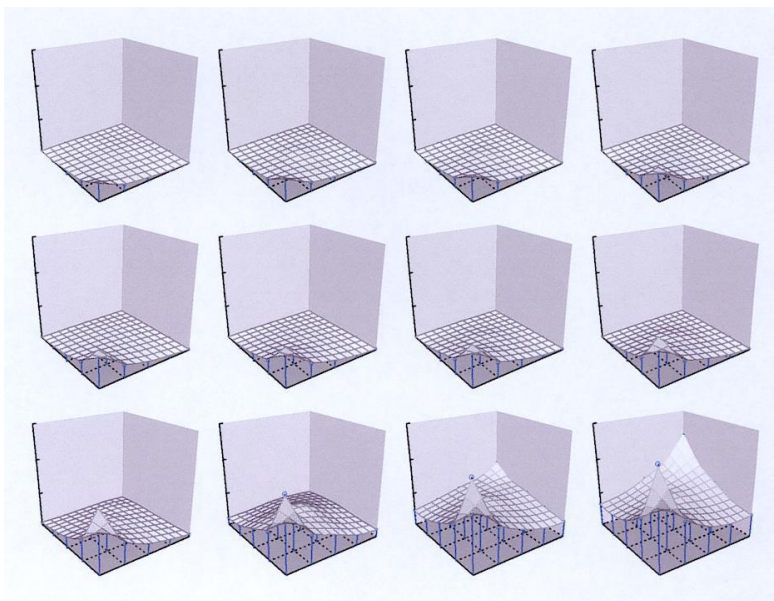


Fig.3: Change of the Shapes dependent on the changing environment and men (Indian medicine during 30 years).

Let us come back to the power distributions. 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 of an attracted and Y is the value of the same personality characteristic of the attractor and in case of mutual attraction also vice versa).

Because of the textual explaining character of the model, the power functions, we are considering for use in the model, are not any arbitrary selected power functions but they are especially created on basis of both well-known determinants of interpersonal attraction and complementarities mentioned above:

- the crucial determinant of interpersonal attraction (similarity) suggests to consider the distances A between the qualities of persons ($A=|X-Y|$) as independent variable of a power function:

$$Z^* = c_1 \cdot (A+1)^\alpha \quad (3)$$
with $c_1 = \text{constant}$; the 1 is added because log A is not possible in case $A=0$

- the model of *complementarities* leads to the conclusion to use additionally the complements of these distances ($A_{\text{complement}}$) as independent variable of the second power function:

$$Z^{**} = c_2 \cdot (A_{\text{complement}}+1)^\beta \quad (4)$$

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,

$$A + A_{\text{complement}} = A_{\text{min}} + A_{\text{max}} \quad (5)$$

$$A_{\text{complement}} = A_{\text{min}} + A_{\text{max}} - A \quad (6)$$

That means, $A_{\text{complement}}$ is increasing according to the same amount as A is decreasing and vice versa.

The relationships of the two parameters α and β to each other determine the expressions of the complementarities (similarities, dissimilarities) in each of the shape. Demonstrations of *complementary tendencies* are given in section 4.2 with help of an example.

- for the *purpose of completion*, the addition ($B=X+Y$) as *opposite* of subtraction ($A=|X-Y|$), is supposed as independent variable of the third power

function

$$Z^{***} = c_3 \cdot (B+1)^\gamma \quad (7)$$

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

$$Z^{****} = c_4 \cdot (B_{\text{complement}}+1)^\delta \quad (8)$$

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

$$B + B_{\text{complement}} = B_{\text{min}} + B_{\text{max}} \quad (9)$$

$$B_{\text{complement}} = B_{\text{min}} + B_{\text{max}} - B \quad (10)$$

The sums B and ($B_{\text{complement}}$) are covering another well-known general characteristic of structures in interpersonal relations (For more details, cf. Kretschmer & Kretschmer 2007).

We assume the *intensity of mutual attraction* Z_{XY} is proportional to the product of the four above mentioned different power functions:

$$Z_{XY} = \text{constant} \cdot (A+1)^\alpha \cdot (A_{\text{complement}}+1)^\beta \cdot (B+1)^\gamma \cdot (B_{\text{complement}}+1)^\delta \quad (11)$$

Further, we assume the *frequency of social interactions* Z'_{XY} between pairs of persons (or frequency of pairs) is a reflection of the intensity of their mutual attraction Z_{XY} . Thus - according to Bernoulli's law of large numbers - with increasing sample size, Z'_{XY} is approximating to Z_{XY} . (Bernoulli's law of large numbers - (statistics) law stating that a large number of items taken at random from a population will (on the average) have the population statistics)

The measurement of the variables X, Y and Z'_{XY} including $X_{\text{min}}=Y_{\text{min}}=m_1$ and

$$X_{\text{max}}=Y_{\text{max}}=m_2 \quad (12)$$

$$X_{\text{max}}=Y_{\text{max}}=m_2 \quad (13)$$

is depending on the studied object.

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, etc.

The measurement of these variables for the study of co-author pairs' distributions are shown in section 4.4.

H. Kretschmer & F. Havemann (Eds.): Proceedings of WIS 2008, Berlin

Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting

Humboldt-Universität zu Berlin, Institute for Library and Information Science (IBI)

This is an Open Access document licensed under the Creative Commons License BY

<http://creativecommons.org/licenses/by/2.0/>

4.2 Example for Demonstration Complementary Tendencies in Connection with Varying Shapes

The Function Plot of SYSTAT produces two- and three-dimensional function plots. Three variable equations have the form $z = f(x,y)$ and are plotted here in a rectangular coordinate system. We have typed the equation (11) with $A=|X-Y|$ and $B=X+Y$. Further we have defined: $m_1=1$ and $m_2=5$, $c=1$, resulting in:
 $Z = (|X-Y|+1)^\alpha \cdot (5-|X-Y|)^\beta \cdot (X+Y+1)^\gamma \cdot (13-x-y)^\delta$

For demonstration of complementary tendencies with varying shapes we have entered different values for the parameters but c_1 , m_1 and m_2 are fixed. As mentioned above the relationships of the parameters to each other determine the expressions of the complementarities.

In Fig. 5 we show the four simplest shapes of the social Gestalt using 1 for one of the parameters and 0 for the three others. The left picture each is turned around two times resulting in the other two pictures. Fig. 4 is used for explanation of the pictures in Fig. 5. We have selected the left picture of the first row in Fig. 5 for use in Fig. 4.

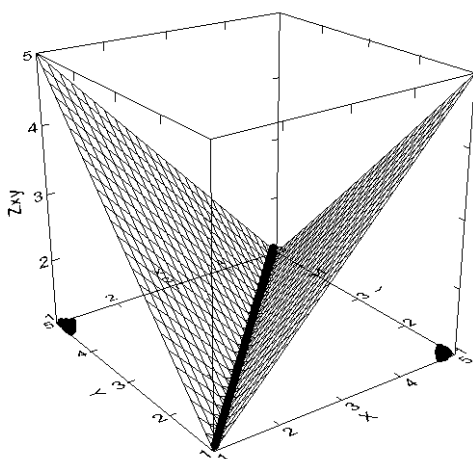


Fig. 4: Simple shape of the social Gestalt. Thick line: main diagonal
 X is equal to Y in the main diagonal (Fat line), $X=Y$, i.e. $A=0$. The maximum difference between X and Y can be found at the left and the right corners (Special marked), i.e. $A_{max}=5-1=4$.

Fig. 4 shows the intensity of mutual attraction. Z_{XY} is lowest for equal persons (main diagonal) and it is increasing up to the most unequal pairs. In other words "Opposites attract". For simplicity, let us say "Yin".

The second row in Fig. 5 shows the complementary picture: "Birds of a feather flock together". For simplicity, let us say "Yang".

In analogy, related to the sum of X and Y , i.e. $B=X+Y$, the two complementary pictures are drawn in the third and fourth rows of Fig. 5. The two complementary pictures are presenting the opposite Yin/Yang pair in relation to the first mentioned Yin/Yang pair.

For demonstration how the two polar forces (first Yin/Yang pair) can interact as complementary tendencies dynamically with each other, so that the entire system is kept flexible and open to change the shapes, let us have a view at Fig. 6, first row. The expression of Yin can be found in the left picture ($\alpha=1$, $\beta=0$) and the expression of Yang in the right ($\alpha=0$ and $\beta=1$).

The three pictures between are produced by systematical variation of the two parameters α and β (Second picture: $\alpha=0.75$ and $\beta=0.25$, third: $\alpha=0.5$ and $\beta=0.5$, fourth: $\alpha=0.25$ and $\beta=0.75$). The relationships of the two parameters to each other determine the expressions of Yin and Yang in each of the pictures. Starting from the left side of the first row in the direction of the right picture, from picture to picture Yin has retracted itself in favor of Yang.

The first row shows only 5 pictures as typical examples. However, according to Chinese philosophy, Yin and Yang are the opposite poles of a single whole. There is not an isolated exclusive Yin, or only a Yang. All transitions occur with a direct and uninterrupted sequence. The natural order is secured by the dynamic equilibrium between Yin and Yang. That means, theoretically an infinite number of pictures should have been produced in row one for presentation of all of the possible pictures.

The same principle as in the first row of Fig. 6 is also working in the second row but with the other Yin/Yang pair. The two complementary pictures on the left and the right sides of the second row are taken from the third and fourth rows (middle pictures) of Fig. 5.

The third row in Fig. 6 shows the same principle as in the other two rows. However, all four parameters (α , β , γ and δ) are involved.

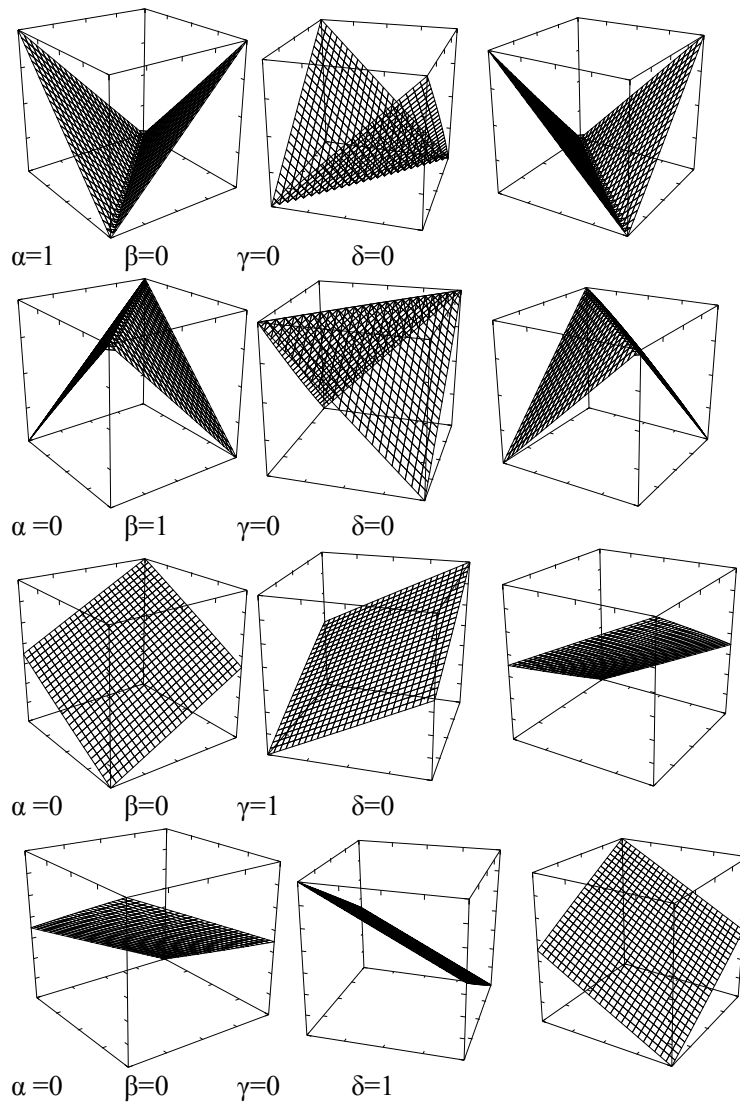


Fig. 5: The Four Simplest Shapes of the Social Gestalt

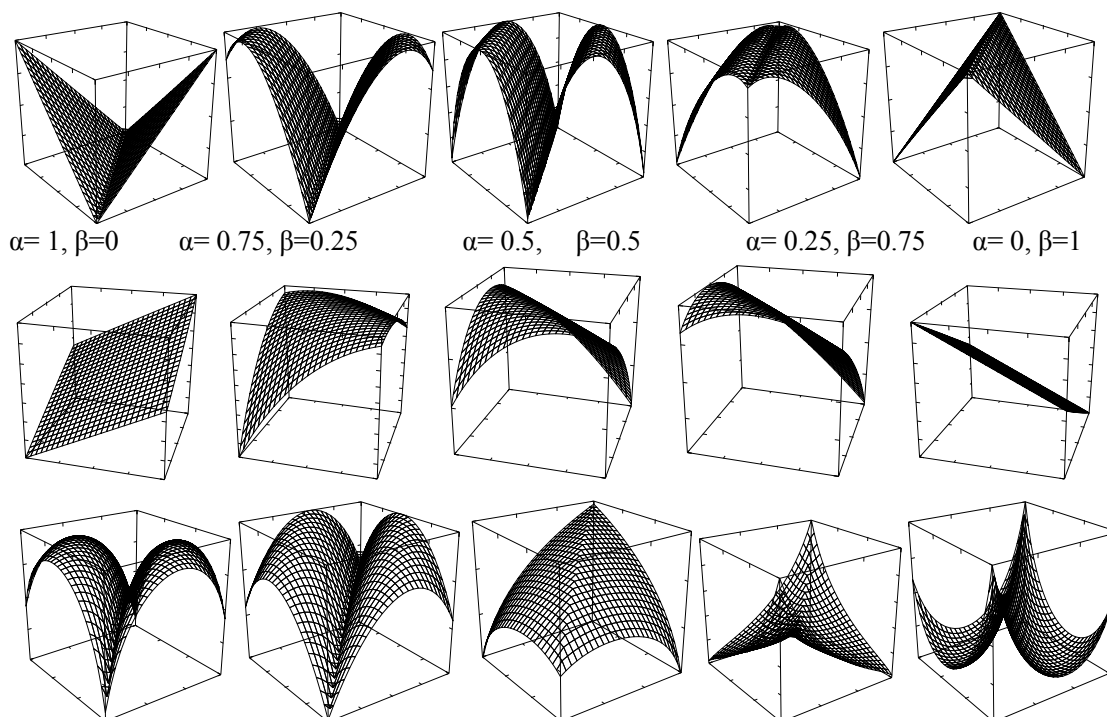


Fig. 6: Presentation how polar forces can interact as complementary tendencies dynamically with each other, leading to the emergence of new shapes (First row: left – opposites attract, right-birds of a feather flock together. The relationships of the two parameters α and β to each other determine the expressions of Yin and Yang in each of the pictures. Starting from the left side in the direction to the right picture, from picture to picture Yin has retracted itself in favour of Yang. Second row: analogously, changing relationships of γ and δ . Third row: analogously, changing relationships of all of the four parameters)

In summary, social Gestalts can change the shapes resulting in a diversity of different shapes. These many shapes, found in empirical studies, are classified into five Prototypes; cf. Fig. 7, upper picture (Kretschmer 2002, p. 316).

However, what is happening in case we have a fixed proportion of the four parameters to each other but the values of these parameters are decreasing from up to down?

Given is a simple example:

$$|c| = \alpha = \beta = \gamma = \delta = \text{Constant}$$

Let us say, this is the **Basic Shape (BS)** of social Gestalt.

In case the constant is decreasing continuously from c to $c' > 0$, resulting, the shape is not changing!

The opposite simple example:

$$-c = -\alpha = -\beta = -\gamma = -\delta = -\text{Constant}$$

Let us say, this is the **Mirror Image of the Basic Shape**

In case the constant is increasing continuously from $-c$ to $-c' < 0$, again, the shape is not changing!

But there is not any social Gestalt with $c=0$!

Whereas the **Basic Shape (BS)** shows similarities with the lower Prototype in Fig. 7 (This was found in studies about collaboration in institutionalized communities: Opposites attract) the **Mirror Image of the Basic Shape** has some similarities with the middle Prototype found in collaboration structures of invisible colleges (Birds of a feather flock together), cf. Fig. 7. This main principle is already discussed in Kretschmer 1994, however without corresponding pictures.

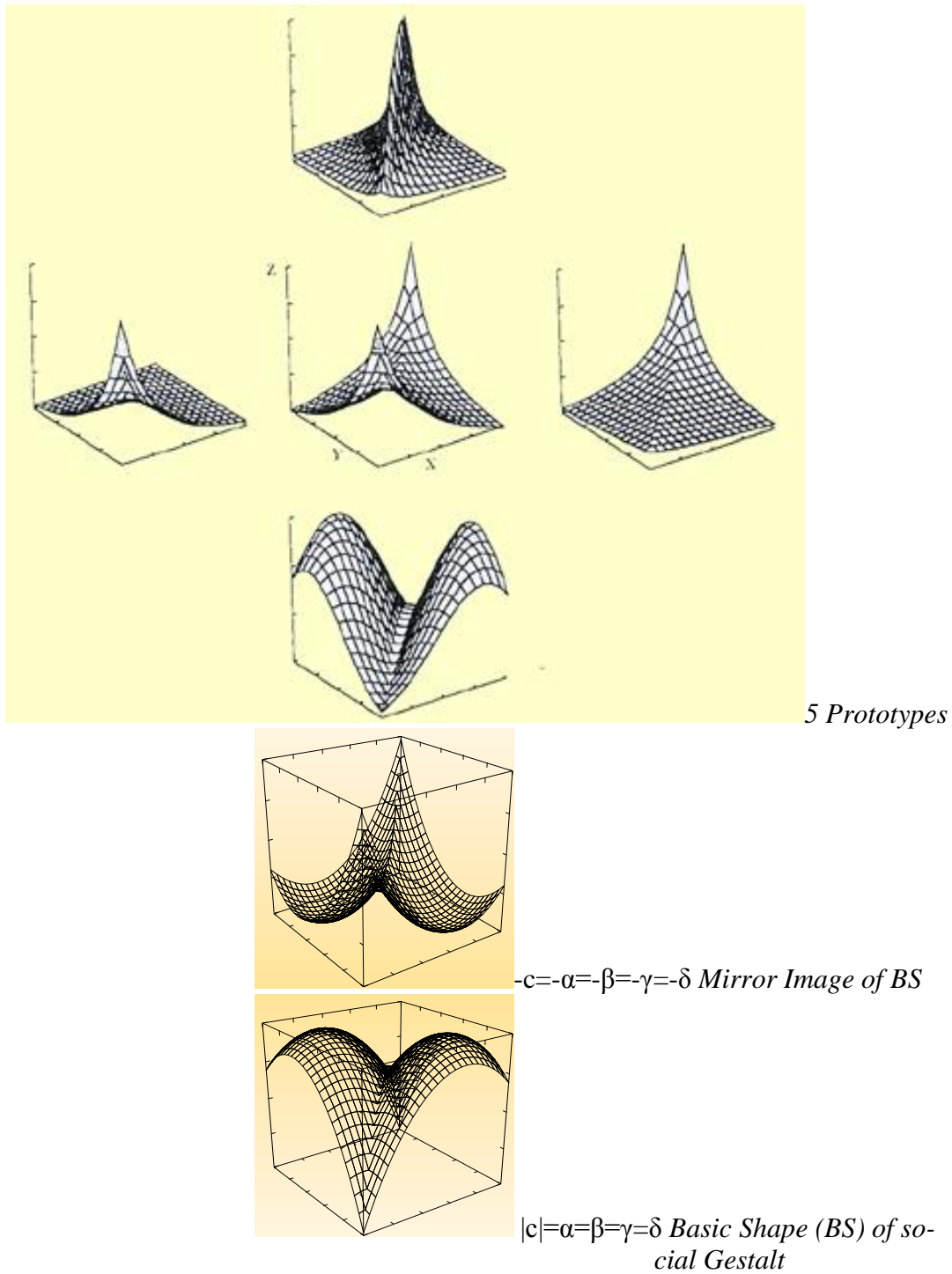


Fig. 7: Five Prototypes of empirical studies in comparison with the Basic Shape (BS) and its Mirror Image

4.3 Overlapping or Mixing of Shapes and Standardizing Procedure

As mentioned above, in conjunction with the model of complementarities, there are different shapes of a social Gestalt and with increasing difference between two shapes the correlation between these two shapes is decreasing. However, after mixing two shapes a new shape of social Gestalt is emerging describable successfully with help of the mathematical function of the social Gestalt, cf. Fig. 8.

The empirically obtained distributions can become mixed according to the same procedures as the mixed shapes because according to our assumption the frequency distribution of social interactions is a reflection of a social Gestalt.

For comparison of the shapes as well as for mixing, sometimes standardizing of the Z_{XY} values (Z_{XYS}) is recommendable. For comparison of two or more shapes the overlay of these shapes into a single frame is possible: either overlay of standardized shapes or of the originals.

The following two paragraphs in this section show

- the procedure of mixing shapes
- the mathematical proof that after mixing a new shape of social Gestalt is emerging and that this new shape is successfully describable with help of the mathematical function of the social Gestalt.
- the procedure of standardizing

Procedure of mixing shapes and mathematical proof

Shapes are mixed with fixed m_1 (12) and fixed m_2 (13) as in the studies of co-author pairs' frequencies (11).

Mixing of shapes is possible on the one hand using the original data directly (morphing) or by mixing the logarithmic data either using standardized or non-standardized data.

Mix ($\log Z_{XYM2}$) of 2 shapes by logarithmic data with fixed m_1 and m_2 :

$$\log Z_{XYM2} = (\log Z_{XYS1} + \log Z_{XYS2})/2 \quad (14)$$

$$\log Z_{XYM2} = (\text{constant}_1 + \text{constant}_2)/2 + ((\alpha_1 + \alpha_2)/2) \cdot (A+1) + ((\beta_1 + \beta_2)/2) \cdot (A_{\text{complement}}+1) + ((\gamma_1 + \gamma_2)/2) \cdot (B+1) + ((\delta_1 + \delta_2)/2) \cdot (B_{\text{complement}}+1) \quad (15)$$

The mix of two shapes can become extended to the mix of N shapes:

$$\log Z_{XYMN} = \text{average constant} + (\text{average } \alpha) \cdot (A+1) + (\text{average } \beta) \cdot (A_{\text{complement}}+1) + (\text{average } \gamma) \cdot (B+1) + (\text{average } \delta) \cdot (B_{\text{complement}}+1) \quad (16)$$

The function (16) is the mathematical proof, after mixing two (or N) shapes a new shape of social Gestalt is emerging – independently on the increasing difference between the two (or N) mixed shapes- describable successfully with help of the mathematical function of social Gestalt.

This procedure of standardizing or mixing is also possible for $\log N_{ij}$.

There is an exception: A shape cannot be mixed with its mirror image because all of the averages are equal to 0!

Procedure of Standardizing ($\log Z_{XYS}$ or Z_{XYS}):

$$\log Z_{XYS} = (\log Z_{XY} - \log Z_{XY\min}) / (\log Z_{XY\max} - \log Z_{XY\min}) \quad (17)$$

$$\log Z_{XYS} = a + b \cdot \log Z_{XY} \quad (18)$$

$$\text{with } a = (-\log Z_{XY\min}) / (\log Z_{XY\max} - \log Z_{XY\min}) \quad (19)$$

$$\text{with } b = 1 / (\log Z_{XY\max} - \log Z_{XY\min}) \quad (20)$$

$\log Z_{XY\min}$ and $\log Z_{XY\max}$ should be taken from the full theoretical shape (Using the function plot for obtaining the full theoretical shape). The four parameters and the constant can be obtained by regression analysis for use in the model of social Gestalt, i.e. using function (11) or (2) respectively, m_1 and m_2 are taken from the empirical data.

$$Z_{XYS} = 10^{(\log Z_{XYS})} \quad (21)$$

The shapes are not changing the form after standardization

With help of an example we intend to show:

- how to calculate two shapes of the social Gestalt
- how to standardize shapes for comparison
- how to mix two or N shapes
- the overlay of these two shapes into a single frame for comparison

Example:

In empirical investigations the distributions of the frequencies of social interactions (reflection of shapes of the social Gestalt) are the results of the empirical studies and the exponents (parameters) as well as the constant can be obtained by regression analysis.

Given is a personality characteristic with a

huge number of values of the variables X and Y with $m_1 = 1$ and $m_2 = 5$. Following $A_{\min} = 0$ and $A_{\max} = 4$, $B_{\min} = 2$ and $B_{\max} = 10$

For the purpose of simplicity we have selected 3 values only for X and Y each in

Table 1. However, the Figures are not limited to these 3 values each.

All of the variations of the X,Y pairs are used for calculations (In this example: $3^2 = 9$).

Table 1 shows the calculation of two shapes with given exponents.

Table 1: Results of calculation of two shapes with given exponents, results of standardization and mixing by logarithmic data

X	Y	A+1= X-Y +1	A _{complement} = 4- X-Y +1	B+1= X+Y+1	B _{complement} = 12- (X+Y)+1	Z _{XY} of Shape 1 $\alpha = -0.1$ $\beta = 0.1$ $\gamma = -1$ $\delta = -1$	Z _{XY} of Shape 2 $\alpha = -0.01$ $\beta = 1$ $\gamma = 0.1$ $\delta = 5$	Z _{XYS} Shape 1	Z _{XYS} Shape 2	Mix
1	1	1	5	3	11	0.0356	898764	10	10	10
3	1	3	3	5	9	0.022	210379	2.2	5.9	3.6
5	1	5	1	7	7	0.017	20749	1	2.6	1.6
1	3	3	3	5	9	0.022	210379	2.2	5.9	3.6
3	3	1	5	7	7	0.024	102087	2.8	4.5	4.5
5	3	3	3	9	5	0.022	11808	2.2	2.1	2.1
1	5	5	1	7	7	0.017	20749	1	2.6	1.6
3	5	3	3	9	5	0.022	11808	2.2	2.1	2.1
5	5	1	5	11	3	0.036	1544	10	1	3.1

In Fig. 8 two different standardized shapes of the social Gestalt are shown, i.e. shape 1 and shape 2 from Table 1 (Shape 1 is similar to the right picture in the first row of Fig. 2). After regression analysis the correlation between these

two shapes is very low: $R = 0.062$ with F-ratio equal to 0.089 and Error probability $P = 0.77$.

The correlation coefficients, the F-ratios and the Error Probabilities are equal using the original data ($\log Z_{XY}$) or the standardized data (\log

H. Kretschmer & F. Havemann (Eds.): Proceedings of WIS 2008, Berlin

Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting
Humboldt-Universität zu Berlin, Institute for Library and Information Science (IBI)

This is an Open Access document licensed under the Creative Commons License BY

<http://creativecommons.org/licenses/by/2.0/>

Z_{XY}). However, because of standardization both the constants and the coefficients are different, cf. (18).

After mixing these two different shapes a new shape of social Gestalt is emerging describable successfully with help of the mathematical function (11), i.e. the correlation between the mixed data and social Gestalt are equal to one ($R=1$).

On one hand the Mix is similar to shape 1 but on other hand there are also similarities with shape

2. Further, mixing the Mix with shape 1 or with shape 2 is possible, etc., etc.

In total: The change of the shape 1 towards the shape 2 or vice versa can be shown by simulation or by animation. The change of the shape from the Gestalt of the journals of gender studies towards the shape of four mixed SCI journals is shown in this issue in Kundra et. al. 2008

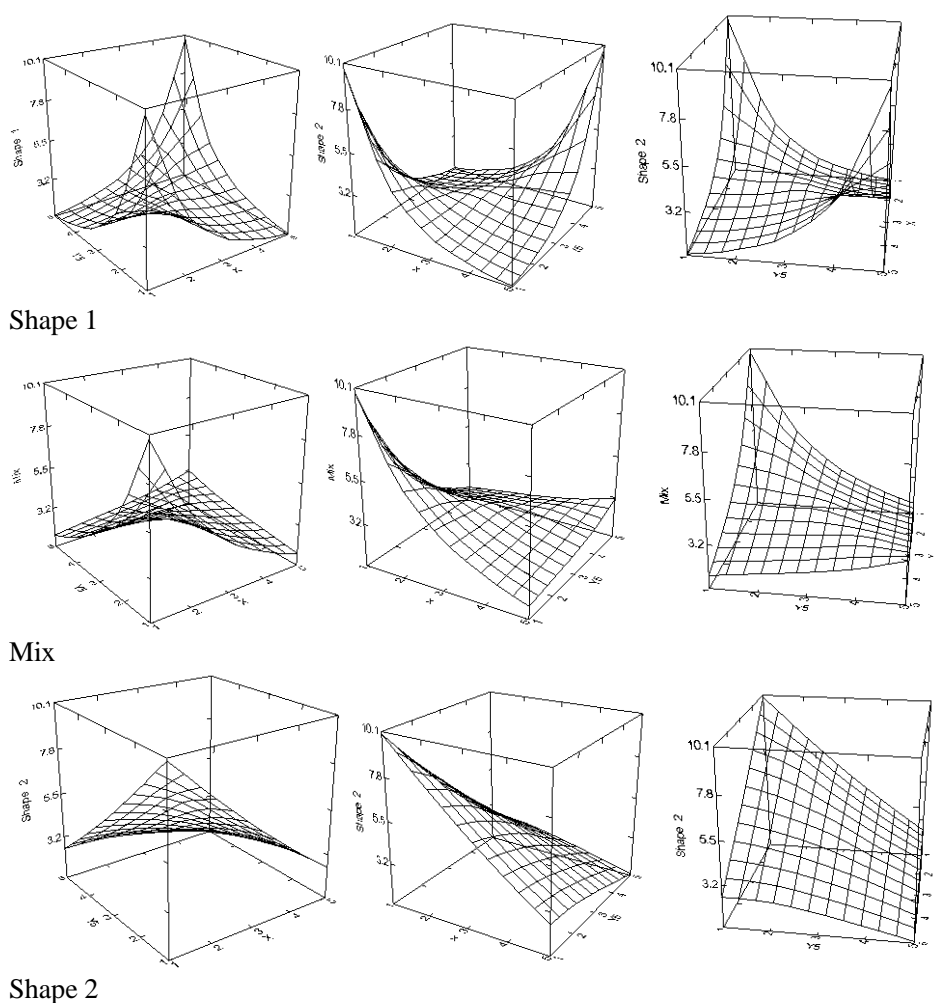


Fig.8: Two shapes of social Gestalt and the Mix of Shape 1 and 2. The pictures in the left column are turned around two times resulting in the pictures of the second and third columns.

For comparison of the shapes 1 and 2 the overlay of these two shapes into a single frame is shown in Fig. 9. The overlay of shapes is also possible with non-standardized shapes or distributions (cf. an empirical example, Fig. 5 in Guo et.al., 2008 in this issue).

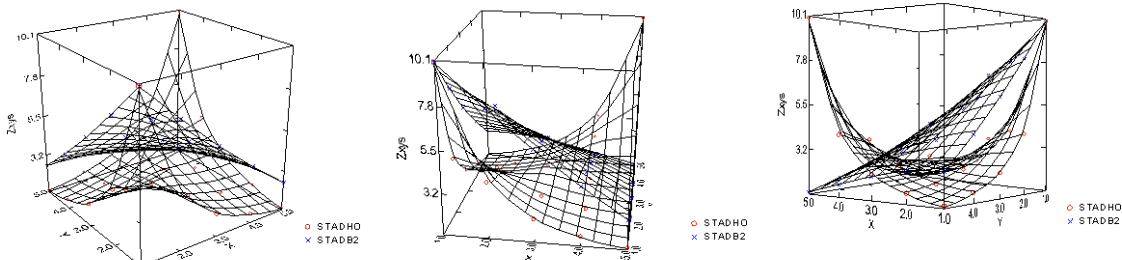


Fig. 9: Overlay of Shape 1 and Shape 2. The left picture is turned around two times resulting in the two other pictures

4.4 Distributions and Shapes of Co-author Pairs' Frequencies $N_{ij}=f(i,j)$

As mentioned in section 3, the distribution of observed co-author pairs' frequencies can be considered to be a special reflection of a social Gestalt, i.e. a reflection in conjunction with Lotka's power distribution. That means, the empirical distribution is based on both the interpersonal attraction and the decreasing frequency of authors A_i (or A_j respectively) with increasing i (or j) publications per author.

For determination the shape of a social Gestalt, independent on Lotka's power distribution, a statistical index (cf. Kretschmer, 2002 and 2007) can be used. That index provides information on the factors, by which the observed frequencies deviate from the expected values in case of statistical independence from the social Gestalt. The pictures in the upper row of Fig. 2 show distributions of these factors (Shapes of the social Gestalt). The pictures in the second row are the triple logarithmic presentations of observed co-author pairs' frequencies.

One of the examples, how to measure the variables X and Y can be shown in relation to the function of co-author pairs' frequencies $N_{ij}=Z'_{XY}$. There is a conjecture by de Solla Price (1963), physicist and science historian, that the logarithm of the number of publications is of a higher degree of importance than the number of publications per se.

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

$$X = \log i \tag{22}$$

$$Y = \log j \tag{23}$$

$$A = |\log i - \log j| \tag{24}$$

$$B = \log i + \log j \tag{25}$$

$$A_{\min} = |X - Y|_{\min} = 0 \text{ with } \log i = \log j \tag{26}$$

$$A_{\max} = |X - Y|_{\max} = |(\log i)_{\max} - \log 1| = |\log 1 - (\log j)_{\max}| = (\log i)_{\max} = (\log j)_{\max} \tag{27}$$

$$B_{\min} = (X + Y)_{\min} = \log 1 + \log 1 = 0 \tag{28}$$

$$B_{\max} = (X + Y)_{\max} = (\log i)_{\max} + (\log j)_{\max} = 2(\log i)_{\max} = 2(\log j)_{\max} \tag{29}$$

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

$$A_{\max} = \log 1000 = 3 \tag{30}$$

$$B_{\max} = 2 A_{\max} = 6 \tag{31}$$

$$A_{\text{COMPLEMENT}} = 3 - |\log i - \log j| \tag{32}$$

$$B_{\text{COMPLEMENT}} = 6 - (\log i + \log j) \tag{33}$$

The theoretical mathematical function for describing the social Gestalts of the distribution of co-author pairs' frequencies is resulting in (2).

Examples:

Because of high skewed distribution of co-author pairs' frequencies $N_{ij}=f(i,j)$ usually we are using the triple logarithmic presentation of the distribution or shape.

We show four empirical distributions and

shapes of a social Gestalt for comparison. The first rows of each of the four figures show the full shapes but the second rows the partial only. There is a cut of full shapes in the second row done by the maximum empirical data of X,Y.

First we want to show an example of the high impact SCI journal, *Journal of Biological Chemistry* (Kretschmer & Kretschmer 2007). The data are taken from 1980-1998 (SCI). After regression analysis the correlation between the theoretical Gestalt and $n=961$ empirical values of $\log N_{ij}$ is equal to $R=0.99$, with $F\text{-ratio}=1,660.3$. The error probability is equal to $P=0.000000000$ (very high statistical significance). The four, from the regression analysis resulting parameters with constant, are used in the model of social Gestalts (mathematical function (2)):

$$\alpha= 1.4224; \beta=5.326; \gamma=1.198; \delta=20.54 \text{ and } \text{const}=-15.692$$

Fig.10 shows the non-standardized triple loga-

rithmic presentation of the social Gestalt in form of a lattice ($Z=\log N_{ij}$, $X=\log i$, $Y=\log j$). The corresponding empirical data are entered in blue colored dots. The left pictures of the Gestalt in Fig.10 are turned around two times (3 pictures of the same shape in each row in total). The full non-standardized shape of the Gestalt of the *Journal of Biochemistry* is presented in the first row and the partial presentation in the second.

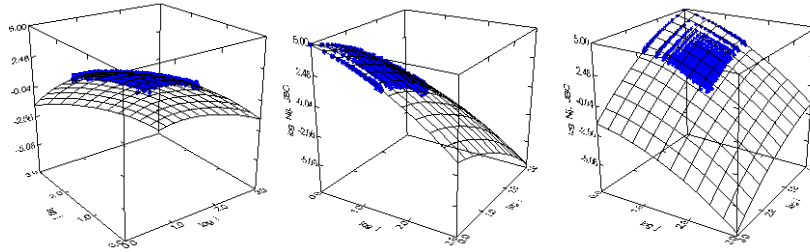
The next example (Fig. 11) is the study of the data set of the combination of four high impact SCI journals (*Science, Nature, PNAS and Phys Rev B Condensed Matter*) The data are taken from 1980-1998 (SCI), (Kretschmer & Kretschmer 2007): $n=961$; $R=0.99$;

$F\text{-ratio}=1,868.3$; $P=0.000000000$.

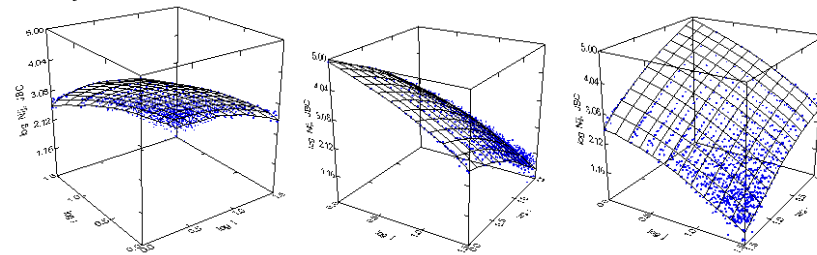
Parameters and constant:

$$\alpha= 1.621; \beta=7.021; \gamma=-1.046; \delta=17.412 \text{ and } \text{const}=-13.245$$

The figures 12 and 13 refer to the *Journal of Information Technology* and to the mix of 11 journals in women's and gender studies.

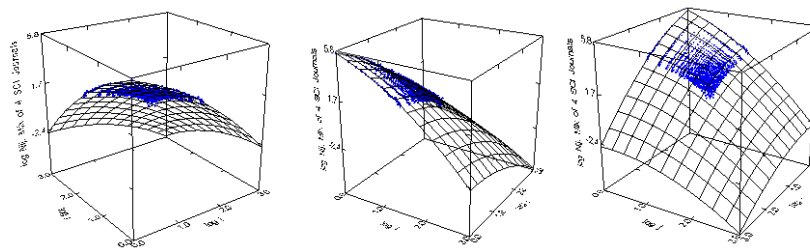


$\log N_{ij}$ (JBC), with $Z_{\min}=-7.6$, $Z_{\max}=5$, $X_{\min}=Y_{\min}=0$, $Y_{\min}=Y_{\max}=3$

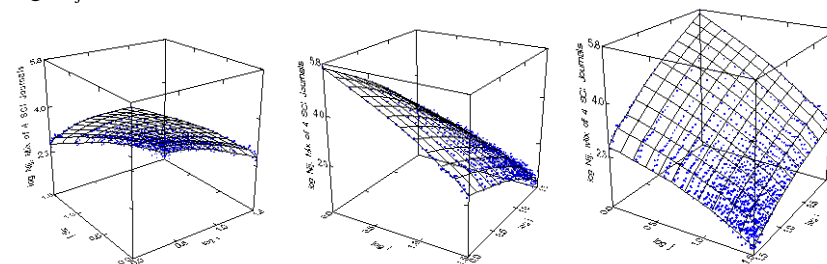


$\log N_{ij}$ (JBC), with $Z_{\mincut}=0.2$, $Z_{\max}=5$, $X_{\min}=Y_{\min}=0$, $X_{\maxcut}=Y_{\maxcut}=1.5$

Fig.10: Presentation of the full non-standardized shape of the Gestalt of the Journal of Biochemistry (first row) and the partial presentation (Second row)

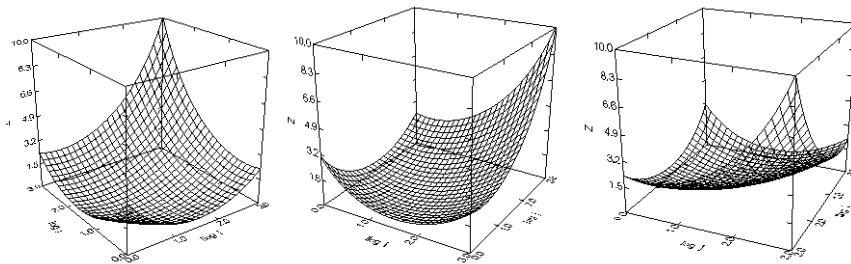


$\log N_{ij}$ (4Mix), with $Z_{\min}=-6.5$, $Z_{\max}=5.8$, $X_{\min}=Y_{\min}=0$, $Y_{\min}=Y_{\max}=3$

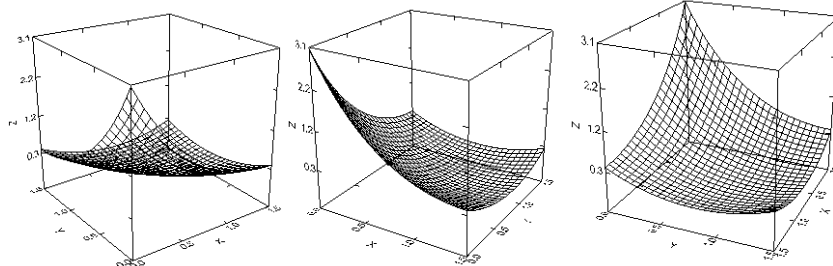


$\log N_{ij}$ (4Mix), with $Z_{\min}=.5$, $Z_{\max}=5.8$, $X_{\min}=Y_{\min}=0$, $Y_{\min}=Y_{\max}=1.5$

Fig. 11: Presentation of the full non-standardized shape of the Gestalt of the Mix of four High Impact SCI Journals (first row) and the partial presentation (Second row)

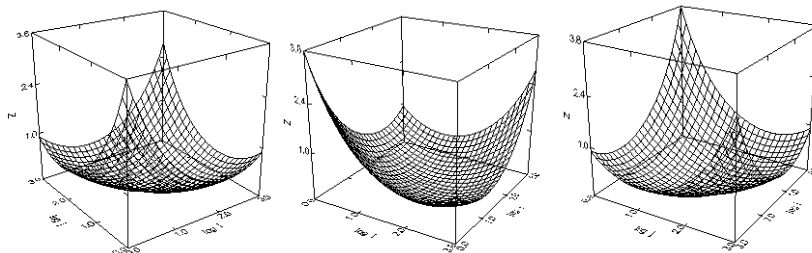


log N_{ij} (JIT), with $Z_{\min}=-0.2$, $Z_{\max}=10$, $X_{\min}=Y_{\min}=0$, $Y_{\min}=Y_{\max}=3$

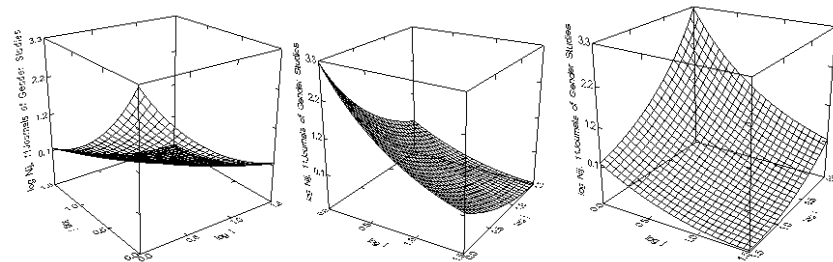


log N_{ij} (JIT), partial presentation $Z_{\min}=-0.2$, $Z_{\maxcut}=3.1$, $X_{\min}=Y_{\min}=0$, $X_{\maxcut}=Y_{\maxcut}=1.5$

Fig.12: Presentation of the full non-standardized shape of the Gestalt of the Journal of Information Science (first row) and the partial presentation (second row)



log N_{ij} (Gender), with $Z_{\min}=-.35$, $Z_{\max}=3.8$, $X_{\min}=Y_{\min}=0$, $Y_{\min}=Y_{\max}=3$



log N_{ij} (Gender), partial presentation $Z_{\mincut}=-.35$, $Z_{\maxcut}=3.8$, $X_{\min}=Y_{\min}=0$, $X_{\maxcut}=Y_{\maxcut}=1.5$

Fig.13: Presentation of the full standardized shape of the Gestalt of the mixed 11 Gender journals (first row) and the partial presentation (second row)

5. Discussion and conclusion

The present paper is focused on a theoretical approach verified by empirical studies explained in two papers published in these Proceedings and in the June 2008 issue of the *COLLNET Journal of Scientometrics and Information Management*. Vo.2, No.1 together with the present paper:

- Guo, Kretschmer & Liu, 2008
- Kundra, deB. Beaver, Kretschmer & Kretschmer, 2008

We have referred to two different versions of counting co-author pairs in the present paper. Whereas the first version leads to a power function distribution of the co-author pairs the other version shows a bivariate distribution of co-author pairs' frequencies hence producing three-dimensional graphs.

The distribution of observed co-author pairs' frequencies can be considered to be the reflection of a social Gestalt (Well-ordered fields of mutual attraction between a large numbers of individual persons) in conjunction with Lotka's power distribution. Based on complementary tendencies a variety of shapes of the distributions of co-author pairs' frequencies emerges.

We have shown the shapes of four distributions (Figures 10-13). The shapes of the high impact factor SCI journals are similar to each other (Figures 10 and 11). However, these two shapes are different from the other two shapes.

In contrast to the first two examples the other two (Figures 12 and 13) are not selected from high impact SCI journals. Additionally, the data sets are very small compared to the first two examples. However, the methods of collection and presentation of the data are the same.

Looking at the *partial* presentations of the four examples: The two high impact SCI journal distributions show similar partial shapes and the two partial shapes of the other two journals are also similar to each other.

Whereas the *full* shapes of the first two examples (Figures 10 and 11) are similar this is not the case for the full shapes of the other two examples (Figures 12 and 13).

Further, the shapes of the first two examples

show self-similarities (full and partial shape) but not the other two shapes.

Maybe all of these differences are caused by the size of the journals, the impact factor or the research area (science or social sciences). Further investigations are necessary to clarify these questions.

As mentioned in the paper, for comparison of two or more shapes the overlay of these shapes into a single frame is possible. We have selected 5 shapes produced in continuation of the shapes in the figures 2, 5 and 7 (cf. 14):

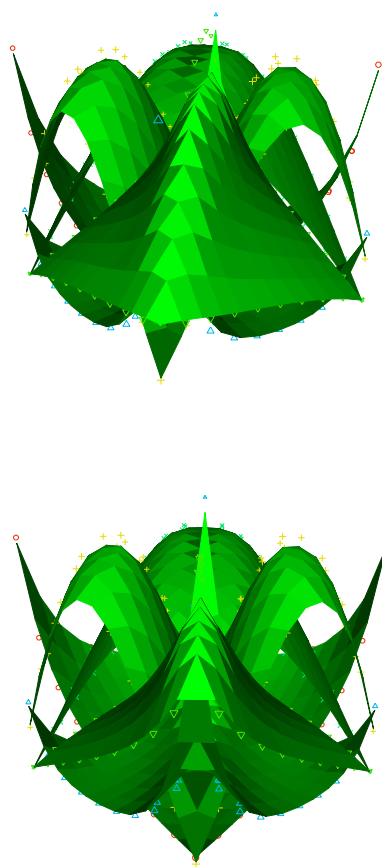


Fig.14: Overlay of Miscellaneous Shapes

We assume the visible three-dimensional behavioural patterns of the scientific communities are emerging through self-organization. In this connection one of the central problems of science will be touched:

How do the forms arise in the nature?

There are various theories in this direction but up today without any empirical proof.

According to new theories the evolution of organisms isn't determined by the genetic code and selection alone but also by self-organization. We have seen the forms or surfaces of the social Gestalt and the change of the Indian shapes. However, there is not any genetic code but the shapes emerge through self-organization.

6. Acknowledgements

The authors wish to acknowledge contributions by Prof. Donald deB. Beaver.

7. References

- Eghe, L. (to appear in 2008). A model for the size-frequency function of co-author pairs
- Guimerà, R., L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas (2003): Self-similar community structure in a network of human interactions. *Phys. Rev. E* 68, 065103
- Guo H., H. Kretschmer & Z. Liu (2008): Distribution of Co-Author Pairs' Frequencies of the Journal of Information Technology. *COLLNET Journal of Scientometrics and Information Management*. Vol. 2, No 1, 73-81
- Kretschmer, H. (1994). Coauthorship networks of invisible colleges and institutionalized communities. *Scientometrics* Vol. 30 No.1 (1994) 363-369
- 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, 501-518
- Kretschmer, H., L. Liang, R. Kundra (2001): Chinese-Indian-German collaboration results that provided the impetus for the foundation of COLLNET. *Scientometrics*. Vol. 52. No. 3, 445-456
- 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*. Vol. 50 No.3, 474-497
- Kretschmer, H. & T. Kretschmer (2007): Lotka's Distribution and Distribution of Co-Author Pairs' Frequencies. *Journal of Informetrics*.1, 308-337
- Kretschmer, H. & T. Kretschmer (2008): Varying Shapes of co-author pairs' distributions. *COLLNET Journal of Scientometrics and Information Management*. Vol. 2, No 1,45-61
- Kundra, R., D. deB. Beaver; H. Kretschmer and T. Kretschmer (2008): Co-author pairs' frequencies distribution in journals of gender studies. *COLLNET Journal of Scientometrics and Information Management*. Vol. 2, No 1,63-71
- Lotka, A.J. (1926). The frequency distribution of scientific production. *Journal of the Washington Academy of Science*, 16, 317-323.
- Morris, S. A. and M. L. Goldstein (2007): Manifestation of research teams in journal literature: a growth model of papers, authors, collaboration, coauthorship, weak ties, and Lotka's law. *Journal of the American Society for Information Science and Technology*.58(12).1764-1782
- Wertheimer, M. (1924). Über Gestalttheorie [an address before the Kant Society, Berlin, 7th December, 1924], Erlangen, 1925. Gestalt Theory. In the translation by Willis D. Ellis published in his "Source Book of Gestalt Psychology," New York: Harcourt, Brace and Co, 1938. Reprinted by the Gestalt Journal Press, New York 1997