Similarities and Dissimilarities in Coauthorship Networks: Gestalt Theory as Explanation for Well-ordered Collaboration Structures and Production of Scientific Literature

HILDRUN KRETSCHMER

ABSTRACT

BASED ON GESTALT THEORY, the author assumes the existence of a field-force equilibrium to explain how, according to the conciseness principle, mathematically precise gestalts could exist in coauthorship networks. A simple mathematical function is developed for the description of these gestalts which can encompass complementary tendencies (as in the principle of Yin and Yang) in their dynamic interplay and, thus, can reflect the change in gestalts. For example, "Birds of a feather flock together" and "Opposites attract" are explained as complementary tendencies.

The data are obtained by SCI. In analyzing the coauthorship networks, coauthorship relations Z between scientists (third dimension) are recorded from the point of view of every scientist with productivity X (first dimension) to all the other scientists with productivity Y (second dimension).

According to the conciseness principle, three-dimensional well-ordered gestalts from different science disciplines are presented. The results of the study have confirmed Metzger's conjectures that the conciseness principle also has validity for social systems, and is valid even with the same conciseness as in the psychology of perception.

It is possible that the presented mathematical function has assumed a more general character and, in consequence, is also more likely applicable to the description of citation networks or the spreading of information.

INTRODUCTION

In every science discipline, basic research and applied research are complementary tendencies interacting dynamically with each other. Progress in scientometrics and informetrics is possible only in this manner. Evalua-

Hildrun Kretschmer, Honorary Professor at Henan Normal University, China, Borgsdorfer Str. 5, D-16540 Hohen Neuendorf, Germany
LIBRARY TRENDS, Vol. 50, No. 3, Winter 2002, pp. 474–497

tion of research institutions by science indicators can be successful only with thorough basic knowledge. For example, citations and coauthorships are reflections of general social relations in networks of people.

The present study is basic research oriented. It will start with general theoretical considerations, followed by applications to coauthorship networks in science.

In the wake of a tangible change of paradigm in science, by the end of the twentieth century a number of holistic theories have emerged (e.g., Bohm, 1980; Stapp, 1993; Prigogine & Stengers, 1984; Sheldrake, 1988; Laszlo, 1997; just to mention a few) that operate on the idea of holographic interacting entities in the world, with several of them also implying a field concept. According to Pribram (1997, p. 12), field concepts are being used when remote-field effects have to be explained.

In psychology, the specialty "'Gestalt'-psychology" originated at the end of the nineteenth century, with due consideration of psychological processes and with holistic organizational patterns playing a role that comprised humanity and the environment. These holistic entities are often designated as psychological fields. Their tendency towards a stable state of order is called conciseness tendency, a "tendency towards a good gestalt." The stable final state is, if possible, built up in a simple, well-ordered, harmonic, and uniform manner in line with definite rules.

Metzger's definition of "gestalt" reads as follows (Metzger, 1954, quoted in line with Metzger, 1986):

> The form of an object is called 'gestalt' if it is not attributable to the rigidity of material and not based on fixing each individual point as such, but rather on an equilibrium of forces (tensions, etc.). In addition, the form of a process or its course is also called 'gestalt' if it is not fixed by impenetrable conduits, or confined to one degree of freedom, but if it had emerged from the free play of field forces (in case of a diverse number of freedom degrees) . . . Thus, we generally call such objects as gestalts which, as correctly noted by PIAGET, owe their gestalt to be balancing interactions of forces.
>
> In this context the opposite notion to gestalt would be the mosaic. While in a mosaic the individual parts are arranged within an externally defined array, with the parts 'not knowing of each other' to a certain degree, the parts and points of a gestalt are to be found in a more or less close dynamic state of communication and interaction: every one interacts with every other and, if something like an ordered array is brought about, every part and point carried and keeps every other and is instantly carried and kept by the totality of the others. (pp. 130 f.)

The conciseness principle was discovered while studying the phenomena of perception: No doubt, perception is an active process; that is, the objects perceived represent a more regular entity than the physical objects existing in the environment. Metzger presumed that this conciseness principle could be generalized and applied to other fields of the psyche, and to socio-psychology as well. Here Metzger had in mind the succinct forms of group structure.

In 1967 Metzger wrote (quoted according to Metzger, 1986):

> If the conciseness principle is validly applicable over the entire psychic sector, its efficacy could also be extended to those wholes, the natural parts of which are human beings: to social group formations, especially to the spontaneously configuring natural small-size groups. I cannot present here any accepted theses, but only preliminary presumptions. (p. 142)

Metzger (1986) mentioned:

> An order of behaviour that without any constraints builds itself up due to internal vectors should qualify as an excellent, a succinct, like the order in the field of perception. (p. 203)

As for the structure of social groups, Metzger (1982, quoted by Metzger 1986, p. 196) suggested that already in the prehistoric times of higher vertebrates—birds and mammals— two succinctly distinguishable conciseness forms of group structures had apparently existed that are also identifiable in humans: Step structure and ring structure. The step structure reveals individual members arranged in an hierarchical sequence (pecking order of the chicken run), whereas the "ring"-members, with their common concern in the center, are distributed "over equal heights."

In his deliberations about the formation of a group, Metzger (1986, p. 222) also touched upon the proverb "Birds of a feather flock together," and gave it a grain of truth. At the same time, however, he suggested that similarity could only be viewed as one factor among many, irrespective of whether it may turn out to be an indispensable or just sufficient requirement for group formation.

This point of view was adopted and, in this study, extended to additional knowledge from the literature on the characteristics of structures in social systems. The results of studies, as contained below, indicate that Metzger's definition of gestalt, which implies the balancing interaction of forces (tensions, field forces, etc.), can be fully applied to social systems, even while retaining the validity of the conciseness principle in a still more precise form than it would have been thought possible by Metzger himself. Hence, there are structures existing in social systems that are strictly mathematically describable.

Without assuming the existence of a field-force equilibrium, it would be difficult to explain how such mathematically precise gestalts—which are thought to have been established by the free cooperation (self-organization) of scientists around the world—could exist. (Cf. all three-dimensional figures of gestalts in the coauthorship networks of this study.)

General Characteristics of Structures in Interpersonal Relations in Social Networks

When discussing the structural characteristics of interpersonal relations in social networks, the author references one of Wolf's works (1996), rather than the many studies conducted and contained in the literature. As a result, one can identify a definite structure underlying a great 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" (see below).

The author illustrates these three aspects on the basis of an empirical example (Wolf, 1996, p. 35). Independently of whether or not socio-demographic 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 be expected from persons with accidental associations. Relations may qualify as friendships, marriages, professional contacts, or other types of relationship.

Wolf, in one of his empirical examples, studied similarity underlying relations of friendship due to common education. It was unequivocal that those persons preferred to become friends with individuals who had achieved the same level of education. These data can also be used to observe the edge effect. The edge effect designates the more pronounced similarity of friendly couples observable at the edges of status features (referring both to persons at the lowest and the highest levels of education). Using Wolf's data file, it is possible to identify four-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 less clearly perceptible with persons having medium-level school degrees. As a result, at the same level of education a U-curve of data arose.

Two hypotheses are primarily suggested that should explain the edge effect. On the one hand, it is maintained that the persons of the lowest and the highest group would be visibly exposed due to their social position and, thus, developed a stronger sense of affiliation than people having a medium-level social status. In addition, those people at a medium status display a stronger orientation towards career so that they are reluctant to have frequent contacts with people of the same level. On the other hand, it is suggested that the choices of people who are either at the very bottom or at the top are blocked in one direction.

Quite similar results were obtained in other studies, for example, the distribution of persons within age groups. The persons belonging to the youngest and those to the oldest groups display a much stronger inclination to remain among their groups than is the case for the medium-age groups.

The well-known proverb "Birds of a feather flock together" can be conveniently integrated into this theory, together with the empirical results published. Far less evidence is found, however, for the opposite saying, "Opposites attract"—although several efforts have been put into proving its correctness; for example, Winch, Ktsanes, & Ktsanes (1954), who considered the complementarity of personality features the decisive factor for partnership relations.

The descriptions available in literature on the crucial specifics of social structures refer to important and special aspects of individual phenomena. In Wolf's empirical example, and in those of many other authors, it became obvious that only one of the two proverbs was used ("Birds of a feather flock together"), leaving out the other, with its opposite meaning. In addition, such examples used the U-curve only in one of its positions (edge effect), and not vice versa. But apart from the U-curve, the assumptions were linear, for instance, Wolf's assumption (1996) on the "Decrease of contacts with declining similarity" (an extension of the proverb "Birds of a feather . . . "), or monotonously falling (e.g., Marsden's "Unidimensional Social Distance Model" (1981)).

Marsden (1981) operated on the premise that: ". . . the likelihood of sociable intercourse between persons in groups is an inverse function of the distance between those groups along a single dimension" (p. 21), with distance being distance in similarity. By contrast, this study will suggest that the opposing proverbs should only be perceived as the conspicuously visible state of a holistic process caused by the 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—that is, with edge effect on the one side, and the reverse case, on the other side.

Development of a Mathematical Function to Describe Gestalts in Social Networks

Matrix of Interpersonal Relations

If one started from the assumption that all individual manifestations of social structure, as invariably mentioned in the literature, have come to interact within a system of the equilibrium of forces, an hypothesis might be established on the emerging forms of the adequate three-dimensional gestalts. These forms should be as simple, ordered, harmonic, and uniform as possible, according to the conciseness tendency, and should be structured in line with definite rules.

Their uniformity could be expressed by the visible retention of the balancing interaction of the different and also opposing individual phenomena in social structures, as known from the literature, and could become visible in only one function. Thus, the diversity of patterns or gestalts is then expressed by the variation of the parameters of this function, with the diversity being dependent upon the conditions causing these patterns (e.g., the environment). These many gestalts can be classified into types in line with their similarity.

Both the opposing proverbs and the U-curves in their contrasting situations give rise to reflect on the notion of complementarity. Capra (1996) wrote that the term "complementarity" (e.g., particle/wave), introduced by Niels Bohr, has become a firm integral part of the conceptual framework within which physicists attentively weigh the problems of nature, and that Bohr had repeatedly indicated that this idea could also be beneficial outside of physics. In conformity with the above, Capra also suggested that the modern notion of complementarity 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, as complementary tendencies interacting dynamically with each other, so that the entire system is kept flexible and open to change. Capra (1996) said:

> It is important and difficult to understand for the people in the western world that these oppositions do not belong to different categories but are opposing poles of only one whole. There is no separate Yin and no separate Yang. All natural phenomena are manifestations of a continuous interplay between both poles, all transitions proceed in a direct and uninterrupted sequence. The natural order manifests itself in a dynamic equilibrium between Yin and Yang. (p. 32; translated from German by the author)

It is consistent with all above considerations to seek a simple mathematical function (the conciseness principle) for the description of gestalt that can encompass the complementary tendencies (Yin and Yang) in their dynamic interplay and, accordingly, also the change of gestalt. The basic requirement for establishing this function is, however, the classification of persons according to a variable of personality characteristics, for example, age or education.

Following the interpersonal relations between these persons (variable Z)—for example, friendship or coauthorship—will be recorded from the point of view of every individual person with value X of the variable of personality characteristics to all the other authors with value Y of this variable. If the relations are recorded from the point of view of every individual person (with X) to all the other persons (with Y), then a symmetrical matrix of Z_{XY} is obtained. For example, there are three friends classified according to education (elementary school: X,Y = 1; junior high school: X,Y = 2; grammar school: X,Y = 3; university: X,Y = 4):

- person A with X (or Y resp.) = 1
- person B with X (or Y resp.) = 4
- person C with X (or Y resp.) = 3

From the viewpoint of A with X = 1, there is one relation recorded to B with Y = 4, that is, Z_{14} and one relation to C with Y = 3, that is, Z_{13}.

From the viewpoint of B with X = 4, there is one relation recorded to A with Y = 1, that is, Z_{41} and one relation to C with Y = 3, that is, Z_{43}.

From the viewpoint of C with X = 3, there is one relation recorded to A with Y = 1, that is, Z_{31} and one relation to B with Y = 4, that is, Z_{34}.

See Table 1 for a symmetrical matrix of friendship relations Z_{XY}.

Table 1. Symmetrical Matrix of Friendship Relations Z_{XY} Between Three Friends Classified According to Education (X or Y Respectively).

X/Y	*1*	*2*	*3*	4
1			1	1
2				
3	1			1
4	1		1	

In general, according to this principle, matrices of interpersonal relations between persons classified according to a variable of personality characteristics can be obtained.

The mathematical function Z = f(X,Y) to describe three-dimensional gestalts in such social networks should depend on the above named three pivotal aspects of the structure of social networks.

Three coordinated steps of approximation to the description of gestalt will be discussed. Both the first and the second steps are only related to similarity or dissimilarity, but the third one concerns the three aspects of structures in interpersonal relations in social networks in total.

Similarity and Dissimilarity

Dissimilarity or contrary similarity between two groups of persons can be measured by the difference between X and Y:

$$X - Y$$

The difference is chosen because of the above mentioned symmetry in its absolute form:

$$|X - Y|$$

There is both a minimum of the difference,

$$|X - Y|_{min}$$

and a maximum of the difference

$$|X - Y|_{max}$$

The similarity is highest at the minimum and lowest at the maximum and vice versa, that is, the dissimilarity is highest at the maximum and lowest at the minimum. Moreover, there is a complementary variation of similarity and dissimilarity: With increasing dissimilarity, the similarity is decreasing and vice versa.

Under the condition dissimilarity A is defined as difference,

$$A = |X - Y| = \text{Dissimilarity}$$

similarity has to be defined as complement $A_{COMPLEMENT}$.
Therefore, with increasing distance D_A of the dissimilarity A from the minimum,

$$D_A = A - |X - Y|_{min} = |X - Y| - |X - Y|_{min}$$

similarity has to decrease according to the same distance from the maximum:

$$A_{COMPLEMENT} = |X - Y|_{max} - D_A = |X - Y|_{max} + |X - Y|_{min} - |X - Y| = \text{Similarity}$$

Accordingly, if the dissimilarity is moving to the maximum, the similarity is moving to the minimum and vice versa. Both the first and the second steps of approximation are two-dimensional representations of patterns only.

First step of approximation. The initial ideas on the mathematical function Z = f(X,Y) were developed in pursuit of quantitative science research. It has for decades been shown that the overwhelming majority of distributions of bibliometric data can be represented as a power function—that is, as a Zipf-distribution instead of a Gaussian distribution as used in psychology and the natural sciences.

For reasons of simplicity, a power function was chosen as the starting point for considerations:

> As a first step of approximation we can say the interpersonal relations are at least dependent on a power function of the dissimilarity between persons.

Since, in case of "equals," the value 0 cannot be raised to a negative power, 1 is added to the term |X – Y|, resulting in the power function:

$$Z^{**} = \text{constant} \cdot (|X - Y| + 1)^{\alpha}$$

$$Z^{**} = \text{constant} \cdot (A + 1)^{\alpha}$$

If the parameter α should be positive, then the idea of the proverb "Opposites attract" would be fulfilled in connection with the assumption of "Increase of interpersonal relations with increasing dissimilarity" (cf. example in Table 2 and Figure 1, right).

The proverb "Birds of a feather flock together," and extended version with the assumption of Wolf's "Decrease of personal relations with declining similarity" or Marsden's Unidimensional Social Distance Model, would all be complied with by the power function in which the parameter α is negative (cf. example in Table 3 and Figure 1, left).

Table 2. Example with α = +1 and Constant = 1.

$X - Y$	$\|X - Y\|$	$\|X - Y\| + 1$	$(\|X - Y\| + 1)^1$	$1 \cdot (\|X - Y\| + 1)^1$
−4	4	5	5	5
−3	3	4	4	4
−2	2	3	3	3
−1	1	2	2	2
0	0	1	1	1
1	1	2	2	2
2	2	3	3	3
3	3	4	4	4
4	4	5	5	5

Table 3. Example with α = −1 and Constant = 1.

$X - Y$	$\|X - Y\|$	$\|X - Y\| + 1$	$(\|X - Y\| + 1)^{-1}$	$1 \cdot (\|X - Y\| + 1)^{-1}$
−4	4	5	0.2	0.2
−3	3	4	0.25	0.25
−2	2	3	0.3	0.33
−1	1	2	0.5	0.5
0	0	1	1	1
1	1	2	0.5	0.5
2	2	3	0.3	0.33
3	3	4	0.25	0.25
4	4	5	0.2	0.2

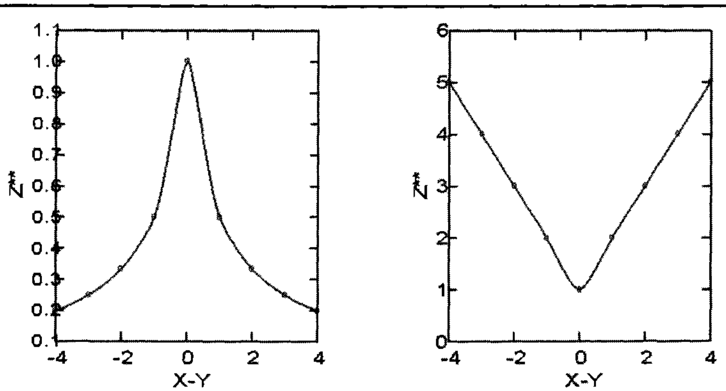

Figure 1. Power Function of the Dissimilarity Z** = constant · (A + 1)α. On the left, the parameter α is negative: "Birds of a feather flock together" and "Decrease of interpersonal relations with increasing dissimilarity." On the right, the parameter α is positive: "Opposites attract" and "Increase of interpersonal relations with increasing dissimilarity."

A power function with only one parameter (unequal to zero) is either only a monotonically declining or a monotonically rising function, when referring to both proverbs: Either Yin or Yang. According to Chinese philosophy, Yin and Yang are the opposite poles of a single whole. There is neither an isolated, exclusive Yin, nor an isolated, exclusive Yang. All transitions occur with a direct and uninterrupted sequence. The natural order is secured by the dynamic equilibrium between Yin and Yang.

In order to fulfill the inherent requirement that both proverbs and their extensions be included in the representation, the second step of approximation will follow.

Second step of approximation. As mentioned above, with increasing dissimilarity, similarity is decreasing and vice versa. Dissimilarity A and similarity $A_{COMPLEMENT}$ are two opposed varying factors and have to be inserted into the equation with one parameter each. It depends upon the parameters to what extent Yin has retracted itself in favour of Yang or vice versa.

As a second step of approximation, we can say that the interpersonal relations are at least dependent on both a power function of the dissimilarity between persons and another power function of the complement:

$$Z^* = \text{constant} \cdot (A + 1)^{\alpha} \cdot (A_{COMPLEMENT} + 1)^{\beta}$$

In an attempt to convey, in theory, a graphic idea of this function, a systematic parameter variation was made and the results are shown in Figure 2. It is a two-dimensional portrayal of patterns. In every box the difference $X - Y$ is always the abscissa, as in Figure 1, and Z^* is the ordinate axis. In the middle of the abscissa is $X - Y = 0$. The relationships of the two parameters to each other determine the expressions of Yin and Yang in each of the patterns. While in the upper pattern with $\alpha = -1$ and $\beta = 0$ Yin is more likely to be in the foreground ("Birds of a feather flock together"), the pattern below with $\alpha = .5$ and $\beta = 0$ reveals that Yang is more likely to be accentuated ("Opposites attract").

Starting from the upper pattern in the direction of the pattern below, from pattern to pattern Yin has retracted itself in favour of Yang, for example, the right pattern with $\alpha = .75$ and $\beta = 1$.

As mentioned above, the mathematical function $Z = f(X,Y)$ to describe gestalts in social networks should depend on three pivotal aspects of the structure of social networks. Two of the three aspects are already included.

However, if you still want to incorporate the third pivotal aspect called "edge effect"—that is, both forms of the U-curve rather than only the two proverbs—it is necessary to extend the formula according to the same principle (simplicity, conciseness, Yin/Yang) to the sum of X and Y, that is, the formula that so far included only the difference between X and Y.

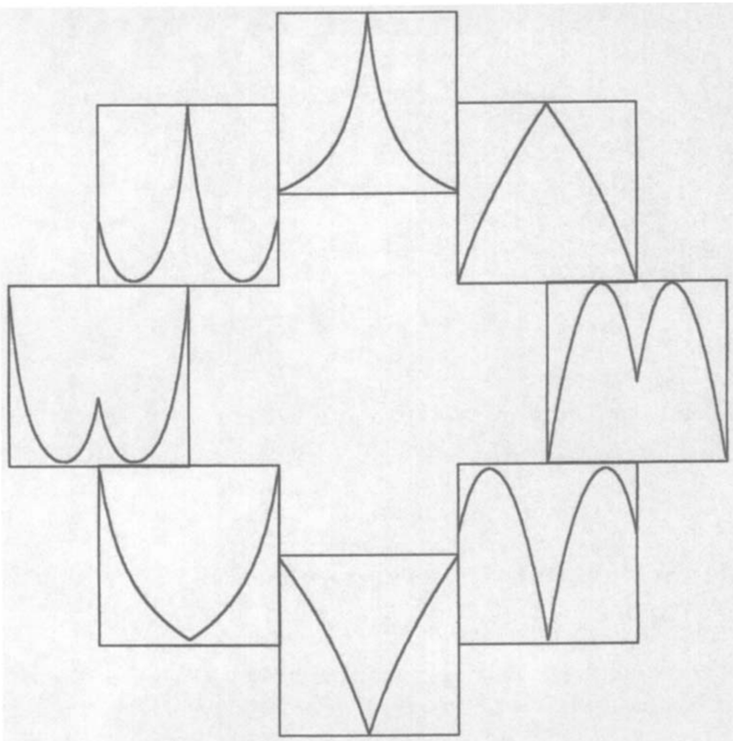

Figure 2. Product of Both the Power Function of the Dissimilarity and the Power Function of the Complement $Z^* = \text{constant} \cdot (A + 1)\alpha \cdot (A_{COMPLEMENT} + 1)\beta$. Systematic variation of both parameters. In every box the difference X – Y is always the abscissa and Z* the ordinate axis. In the middle of the abscissa is X – Y = 0. The relationships of the two parameters to each other determine the expressions of Yin and Yang in each of the patterns.

Edge Effect

Third step of approximation. Interpersonal relations Z at the main diagonal (X = Y) are more striking at the edges than in the middle, although the differences between X and Y did not vary:

$$A = |X - Y| = 0 = \text{constant}$$

The values of Z_{XY}, Z_{11} or Z_{55} at the edges are higher than the values of Z_{22}, Z_{33}, or Z_{44} in the middle.

Whereas the differences between X and Y are constant, the sums are varying.

Therefore, when we put A = |X – Y| and the opposite B = X + Y, the following formula is obtained under the condition that $B_{COMPLEMENT}$ will be calculated according to the same principle as $A_{COMPLEMENT}$:

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

As a third step of approximation, we can say the gestalt of interpersonal relations can be described by the product of the four power functions: First of dissimilarity; second, of its complement; third, of the sum of the values of personality characteristics; and fourth, of its complement. Accordingly, we have obtained a function with four parameters and one constant. This function can encompass the complementary tendencies (Yin and Yang) in their dynamic interplay on two dimensions each (A and B).

What do the three-dimensional gestalts look like that are described in this form? In Figure 3 five prototypes of gestalts are shown. Proceeding in an example from $X_{min} = 1$ and $X_{max} = 5$, or from $Y_{min} = 1$ and $Y_{max} = 5$ respectively, you can obtain the patterns by way of variation of parameters, as shown in Figure 3. Such gestalts can also be generated with other values for minimum and maximum values of X and Y.

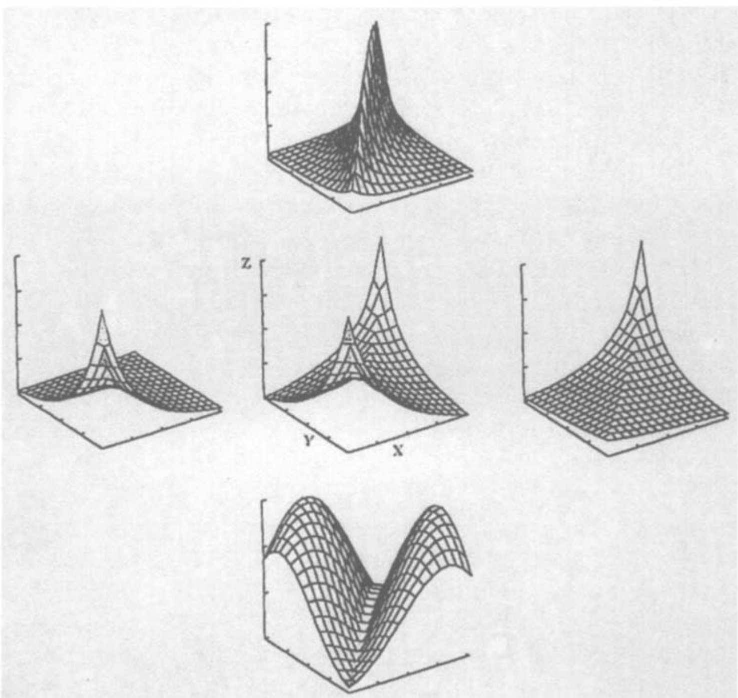

Figure 3. Prototypes of Gestalts in Social Systems.

In the center of the gestalt, the proverb "Birds of a feather flock together," along with the extended version "Decrease of interpersonal relations with declining similarity," becomes conspicuously visible, as does the U-curve with the edge effect. In the lower gestalt, this tendency is less apparent. "Opposites attract," with the extended version "Increase of interpersonal relations with declining similarity," on the other hand, has become more strongly perceptible.

The reversed U-curve is especially conspicuous in the upper gestalt. In the left and the right ones the U-curve has converted into a one-sided tilt. A large number of patterns could be drawn with the same simple function.

At first, in reviewing the hypothesis of social relations in science, the starting point is the social relations, especially in coauthorship networks. In the process of study, one might re-examine whether such or similar gestalts are also applicable outside this field.

The gestalts obtained from coauthorship networks might be more or less similar to those in Figure 3, but they will be somewhat more robust. However, this is not attributable to defective deviations, but rather to the discrete values on the X- and Y-axis in contrast to the first ones. The same applies to the gestalts estimated according to regression analyses.

Social Structures in Collaboration in Science and Hypothesis

This paper shall test the hypothesis whether or not the above mentioned central specifics underlying the structure of personal relations are reflected in scientific communities.

In 1979, Donald deB. Beaver—disciple, assistant, and later coauthor of de Solla Price—developed jointly with R. Rosen a comprehensive and empirically tested theory on scientific collaboration, which in the second half of this century, had become perceptible in extended coauthorships. In their fundamental analysis, Beaver and Rosen referred also to several additional growth-exceeding aspects of structure-forming processes in scientists' communities. The political and economic elites, outside of the science institutions, had become the conduits of financial support in science, but they did not decide on the distribution of those funds among the scientists' communities. That was a task that was to be fulfilled by the scientific elite itself within the science system. This process had intensified the degree of *stratification* in science. It was significant to realize *who collaborated with whom* and how did this collaboration become recognizable within the entire scientists' community by way of headlines of publications.

In the natural sciences and medicine, the second half of the twentieth century has been marked by *teamwork and coauthorship,* with about 60–70 percent of published scientific papers being coauthored. The development towards collaboration and cooperation has become such a prevailing trend that it is highly imperative to study it so as to gain fundamental knowledge on the intensification of research, which will be indispensable given the probable deceleration of science growth in the future.

What do these structures look like? Beaver and Rosen gave some indications: There is an intensified stratification in science observable due to professionalization. The initially prevailing "Master-master collaboration" (that is, "Birds of a feather flock together") has shifted towards a "Master-apprentice collaboration" system (that is, "Opposites attract").

These studies were related to the nineteenth century. However, a continuation for the second half of the twentieth century appears possible. Enquiries would have to be conducted on the status of every scientist who had made a contribution to the coauthorship network analyzed. Something like that should be done in the future. Today, however, larger networks with several thousands of authors would require a considerable labor expenditure.

By contrast, a relatively small amount of labour is needed for the analysis of an appreciably large amount of bibliometrical data. Therefore, this study has tried to find out whether a scientist's productivity is a characteristic that generates behavioral patterns, as is done by his/her status (master, apprentice, etc.).

Productivity, which is determined quantitatively by the number of publications per author, is coupled with several criteria—such as ability, professional recognition, endurance, social rank, communication, associative relationship with a team, and many others—that might have an influence on who collaborates with whom.

Of course, the number of a scientist's publications is not identical with his status. However, this number nonetheless provides a correlative indication. Even if the correlations for individuals might be moderate or low, they are prone to increase with the number of persons studied. In addition, an elitist scientist has, on average, far more publications than a normal scientist. As a result, a bibliometrical analysis assumes a somewhat different function than the studies by Beaver and Rosen. Yet a comparison between them is possible. Apart from such a comparison, this study is generally concerned with pattern formation; that is, pattern formation is not dependent upon these differences between the studies of the nineteenth century and the bibliometric analysis.

Hypothesis

Three-dimensional gestalts obtained from international coauthorship networks of different science disciplines can be shown to be strictly mathematically describable as a mathematical function.

Data

Twenty bibliographies of the international literature of physics and medicine were analysed. These bibliographies were obtained by SCI and compiled by ISSRU Budapest. The following bibliographies were analyzed:

- Ten bibliographies of the international literature of physics (each: five years time period), including theoretical physics, experimental physics, and cross-disciplinary physics. All the data for the ten physics bibliographies were compiled into a single count:
  Years: 1980–1989
  Total number of articles: 21,730
  On an average: 2.5 coauthors per article, maximum about 10
  Total number of coauthorships: 366,000
- Ten bibliographies of the international literature of medicine (each: five years time period), including biochemistry, pharmacology, clinical, and biochemical analyses. All the data for the ten medicine bibliographies were compiled into a single count:
  Years: 1979–1990
  Total number of articles: 40,596
  On an average: 3 coauthors per article, maximum 10
  Total number of coauthorships: 361,000

Methods for the Analysis of Coauthorship Networks

Three different kinds of matrices are, independently of each other, the basis for the search for gestalts that arise between scientists due to coauthorships. It has to be shown that all the three matrices are relevant for gestalts.

The matrix of relative frequencies of coauthorship relations F_{XY} between scientists with X and Y publications per scientist is one of the three matrices.

The second one will be the matrix of observed frequencies of coauthorship relations C_{XY} and the third one will be the matrix of a special interaction index H_{XY}, which is used in sociology for studies of this kind (cf. Wolf, 1996).

Each of F_{XY}, C_{XY}, and H_{XY} will be used, acting as variable "Z" (relations between coauthors) of the mathematical function.

Matrix of Observed Coauthorships: Matrix C_{ij}

The relative frequency of coauthorships of scientists F_{ij} with i and j publications per author is expressed by the relationship between observed coauthorships C_{ij} with the statistically expected ones W_{ij}.

First let us find out the matrix of C_{ij}. Given is a bibliography (partly represented, names of authors A, B, C . . .)

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

etc.

The number of publications per author i 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, i.e., i = 3: Once in the first article, and each once in the third and fourth article).

It should be noted here that the term "article" is used in relation to a work or paper which was jointly written by one or several authors, compare 1., 2., 3., . . . etc. articles in the bibliography. By contrast, the term "publication" refers to persons.

If the relations in the by-line of an article are recorded from the point of view of every individual author to all the other authors, then a symmetrical matrix is obtained. As an example, in the fourth article there is, from the viewpoint of author D with i = 2, one relation recorded to author A with j = 3 and to F with j = 1. Furthermore, in the same article, from the viewpoint of A with i = 3, there is a relation recorded to D with j = 2 and to F with j = 1. From the viewpoint of F with i = 1, there is a relation recorded to D with j = 2 and to A with j = 3.

The same procedure has to be continued with all of the articles. Generally, from the hypothetically assumed complete bibliography—it is only partly represented in the upper example—it is the matrix of the observed coauthorship relations of each author to all the other ones: Matrix C_{ij}.

Matrix of Expectation Values: Matrix W_{ij}

From the hypothetically assumed complete bibliography, Table 4 was established. A_i are the number of authors with i publications per author. For example, there are $A_2 = 64$ authors with two publications per author. A_i is distributed according to Lotka's law (1926).

The product of i and A_i is the number of publications of all authors (or group of authors respectively) with i publications per author. For example, the number of publications of the group of authors with two publications per author is the following one:

$$i \cdot A_i = 2 \cdot 64 = 128$$

Table 4. Distribution of Authors and Publications of the Hypothetical Bibliography ($\Sigma_j j \cdot A_j = 540$).

Number of Publications per Author	Number of Authors	Number of Publications of All Authors (A_i)	Relative Frequency of Publications of All Authors (A_i)
i	A_i	$i \cdot A_i$	$f_i = i \cdot A_i/\Sigma_j j \cdot A_j$
1	167	167	0.30926
2	64	128	0.23704
3	39	117	0.21667
4	27	108	0.20000
..	..	..	..
20	1	20	0.03704

The relative frequency of publications of one group of authors f_i with i publications per author is the ratio of the number of publications of this group divided by the total sum of publications of all groups of authors. For example, the relative frequency of publications of the group of authors with two publications per author is the following one:

$$f_i = \frac{i \cdot A_i}{\Sigma_j j \cdot A_j} = \frac{128}{540} = 0.237$$

Probability p_{ij}, on the assumption of independence, is that a publication of the authors' group with i publications per author coincides with a publication of the authors' group with j publication per author, equals $f_i \cdot f_j$:

$$p_{ij} = f_i \cdot f_j$$

A matrix of p_{ij} will be established according to the usual rules and following a matrix of expectation values W_{ij}:

$$W_{ij} = f_i \cdot f_j \cdot T$$

with T = Total sum of coauthorships ($\Sigma_i \Sigma_j C_{ij}$).

Classification of Data

There are very large bibliographies, for example, with more than fifty publications per author. In order to avoid statistical fluctuations, the data are classified according to the logarithm of the number of papers. There is a conjecture by Price (1963) that the logarithm of the number of publications is of a higher degree of importance than the number of publications per se.

Both the line variable i and the column variable j can be separately classified according to the logarithm, which results in the conversion of the large matrix, initially available in a raw form, into a smaller one by summing up lines and columns (class X = 1 contains those authors with one publication per author; class X = 2, authors with two to three publications; X = 3, authors with four to seven publications; X = 4, authors with eight to fifteen publications; and X = 5, authors with sixteen and more publications—by analogy the same applies to Y).

This limitation to five classes was established in order to compare behavioural patterns of different science disciplines. However, in most bibliographies there are only very few authors to be found with more than thirty-two publications. But, since the patterns become more stable with an increasing number of individuals, classes with individual authors would distort the picture.

Both for matrix W_{ij} and for matrix C_{ij} it is possible to determine the appropriate sum of data for every cell, that is, resultant C_{XY} and W_{XY}.

The data are classified now into 5 · 5 = 25 classes—15 of them are independent of each other because of symmetry.

The relative frequency of coauthorships F_{XY} is

$$F_{XY} = \frac{C_{XY}}{W_{XY}}$$

Matrix of Special Interaction Indices

In some sociological studies of interpersonal relations in social networks of men (Wolf, 1996), a special interaction 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, which would otherwise be expected in case of statistical independence from characteristics. In order to calculate this index, we have to convert the matrix of observed frequencies C_{XY} into a new matrix using geometric mean. The special interaction index H_{XY} is defined as:

$$H_{XY} = C_{XY} \cdot \frac{G}{G_X \cdot G_Y}$$

where G = geometric mean of all matrix data
G_X = geometric mean of the data in row X
G_Y = geometric mean of the data in column Y

Regression Analysis

In order to calculate the correlation coefficients and error probabilities, the logarithm was taken of the mathematical function above so that it was possible to carry out the classic linear regression analysis. Fifteen out of twenty-five data of the matrix were only evaluated due to symmetry. With four parameters and one constant, a degree of freedom of df = 10 was obtained.

Mixture of Bibliographies

When several bibliographies are mixed with each other, per class, both the observed and the statistically expected values of these bibliographies are added. After it, the usual procedures are carried out.

Three-Dimensional Gestalts in Coauthorship Networks in International Science

In a former study (Kretschmer, 1996), the author showed that after regression analyses, the correlation between empirical and theoretical values is increasing with the rising number of mixed individual patterns and finally tend to one. The same holds good for the bibliographies with increasing scope of data; that is, the "tendency towards a good gestalt" will enhance.

Therefore, from a "mixture" of ten bibliographies of international physics (source: SCI), the following matrix of relative frequencies is obtained, acting as an example for other matrices, compare Table 5. The corresponding gestalt is shown in Figure 4, bottom, left.

Table 5. Relative Frequencies of Coauthorships F_{XY} in International Bibliographies of Physics.

X/Y	1	2	3	4	5
1	1.866	0.864	0.647	0.536	0.417
2	0.864	1.391	0.968	0.795	0.655
3	0.647	0.968	1.322	1.09	0.959
4	0.536	0.795	1.09	1.408	1.237
5	0.417	0.655	0.959	1.237	1.856

Note: Hints on X, Y and F_{XY} cf. section "Methods"

Regarding the gestalts of the relative frequencies of international physics and international medicine (a "mixture" of ten bibliographies each) separately, regression analyses were carried out with a view towards estimating the parameters and the constant for every behavioural pattern. Based on these estimates, the twenty-five values of Z_{XY} of a matrix were again calculated. Figure 4 contains, at the right side, the behavioural patterns esti-

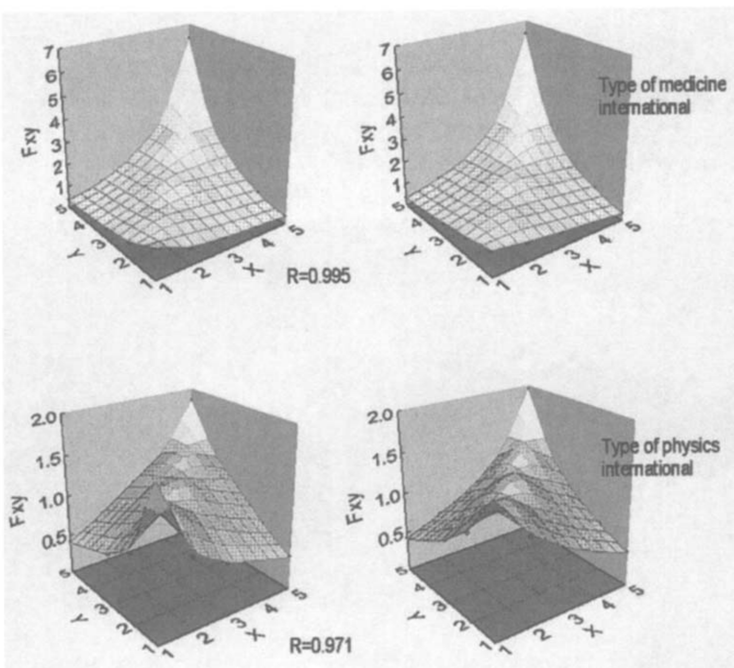

Figure 4. Observations and Estimations of Two Gestalts of Relative Frequencies F_{XY}. Above: International medicine. Bottom: International physics. This figure contains, at the right side, the estimations and, at the left side, the actually observed gestalts.

mated according to the regression analysis and, at the left side, the actually observed behavioural patterns.

From a "mixture" of these twenty bibliographies of international physics and medicine, with altogether some 62,000 articles, some 2.75 coauthors on an average per article, some 730,000 coauthorships, the following three gestalts were obtained (cf. Figure 5):

- Observed frequency C_{XY} with $R = 0.997$, $P < 10^{-10}$
- Special interaction index H_{XY} with $R = 0.994$, $P = 2 \cdot 10^{-9}$
- Relative frequency F_{XY} with $R = 0.978$, $P = 9 \cdot 10^{-7}$

The gestalts of observed frequencies and special interaction indices could, if possible, provide the greatest concurrence with the curvilinear function but, presumably, twenty bibliographies do not appear to be conclusive enough to substantiate the assumption. Irrespective of it, this approach was used to verify still another three international "mixtures."

The total sum of coauthorships in these five gestalts amounts to some 880,000 taken from altogether forty-one bibliographies.

Since these five gestalts are all very similar to each other, even when compared with the gestalt of observed frequencies in Figure 5, the differences were clearly shown in a representation that was selected to find the logarithm to the base 10 in the C_{XY}-axis (cf. Figure 6).

Figure 6 shows the five gestalts of observed frequencies, with two of them contained as "mixture" in the previous figure (source of the first two: SCI, the other three are derived from other sources, e.g., MEDLINE, PSYCINFO, etc.).

Gestalts 1 and 2 are gestalts of physics: $R_1 = 0.993$, $P_1 < 3 \cdot 10^{-9}$ and $R_2 = 0.988$, $P_2 < 5 \cdot 10^{-8}$. Gestalts 3 and 4 are gestalts of medicine: $R_3 = 0.996$, $P_3 < 10^{-10}$ and $R_4 = 0.998$, $P < 10^{-10}$. Gestalt 5 is taken from social sciences: $R_5 = 0.990$, $P < 2 \cdot 10^{-8}$.

At the same time the gestalts in the left part of the figure were turned around 180° and by 90° in the right part.

Both gestalts of physics are similar to each other; the same holds true of both medicine gestalts. However, the socio-scientific one looks somewhat different.

Now the question is whether this is coincidental or whether there are differences in terms of "types." Only additional and comprehensive analyses can give an answer to these questions.

Most of the gestalts display very sharp peaks in the Z dimension. What do those peaks mean? These peaks can be found at the edges of the main diagonal, that is, Z_{11} or Z_{55} can become striking. As mentioned above, one of the three pivotal aspects of general characteristics of structures in interpersonal relations in social networks can be explained as the emergence of the edge effect. On the one hand, it is maintained that the persons of the lowest and the highest group would be visibly exposed due to their social

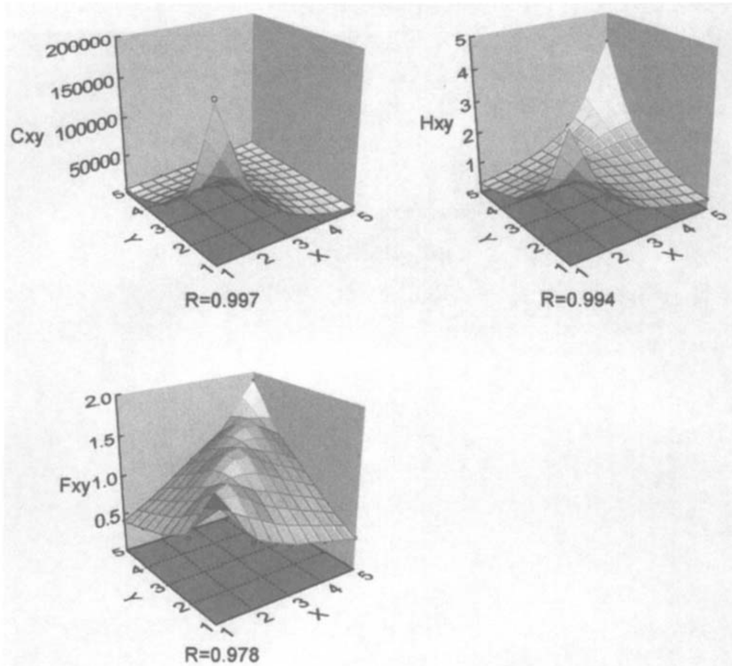

Figure 5. Three Gestalts of the Type from International Medicine and Physics. Gestalts of observed frequencies (C_{XY}), the special interaction index (H_{XY}), and the relative frequencies (F_{XY}).

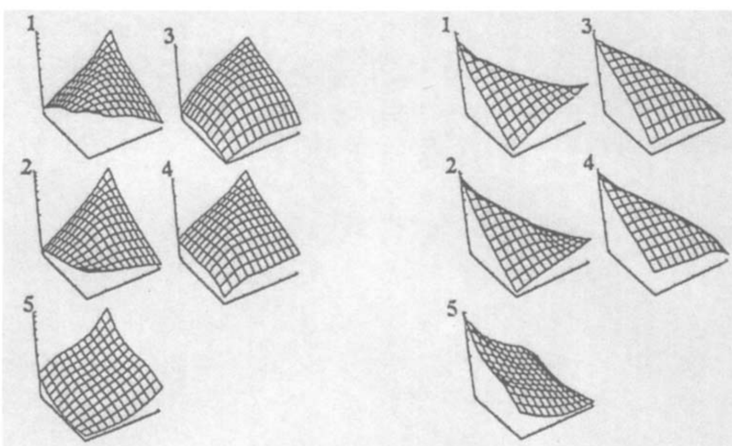

Figure 6. Gestalts of the Logarithms of Observed Frequencies (log C_{XY}) from Five Types in International Science. X and Y are like usual and log C_{XY} are attached to the third dimension. Two gestalts are from physics (1 and 2), two from medicine (3 and 4), one type from social sciences (5). In the left part, the gestalts are turned around by 180°, and, in the right part, they are turned around by 90°.

position and thus developed a stronger sense of affiliation (Z_{11} and Z_{55}) than people having a medium-level social status. In addition, those people at a medium status display a stronger orientation towards career so that they are reluctant to have frequent contacts with people of the same level (Z_{22}, Z_{33}, Z_{44}). On the other hand, it is also suggested that the choices of people who are either at the very bottom (Z_{11}) or at the top (Z_{55}) are blocked in one direction.

Concluding Remarks

The results of the studies have confirmed Metzger's conjectures that the conciseness principle also has validity for social systems and is valid even with the same conciseness as in the psychology of perception.

The three-dimensional gestalts in coauthorship networks, which are widely spread over the entire world, are obviously real objects that owe their shape to the balancing interaction of forces, namely to the dynamic equilibria interacting between Yin and Yang in the sense of ancient Chinese philosophy.

Now let us revert to the theoretical considerations where we found that, for describing a change of gestalts, a mathematical function was derived from ancient Chinese thought. Several characteristic gestalts that were explained in this study were taken out and presented in Figure 7.

The upper and the lower gestalts are new ones. The upper one represents a pattern of a Dutch institute of physics; international coauthorships are included. The lower one represents a pattern of an institutional coauthorship network from which all coauthorships based on authors who are not employed in this institution were removed. In fact, it is the "Kaiser-Wilhelm-Institut für Kohleforschung" (Institute for Coal Research) of the 1920s and 30s. It is really a very small network with only 350 coauthorships. That's why the gestalt is not evenly proportioned like the others. Undoubtedly, it would be worthwhile conducting and continuing such studies also at other institutions in order to find out whether the other gestalts resemble those obtained from the coal-research institute. The same one is valid for the Dutch institute of physics.

Under the condition that the conciseness principle underlies not only the coauthorship network as presented here, but beyond it, a great number of social processes in scientific communities—such as citation networks or the spreading of information—could be developed in the direction of application in information science. For example, it could perhaps be employed in designing search algorithms in databases.

At present, there are some further theoretical developments, including the conciseness principle and the Yin/Yang teaching. Lotka's law (1926) states that scientists will be counted who have i publications included in the bibliography. Couples of scientists will be counted under the condition of both the first scientist's count who has i publications, and the second scientist's count who has j publications included in the bibliography. The fol-

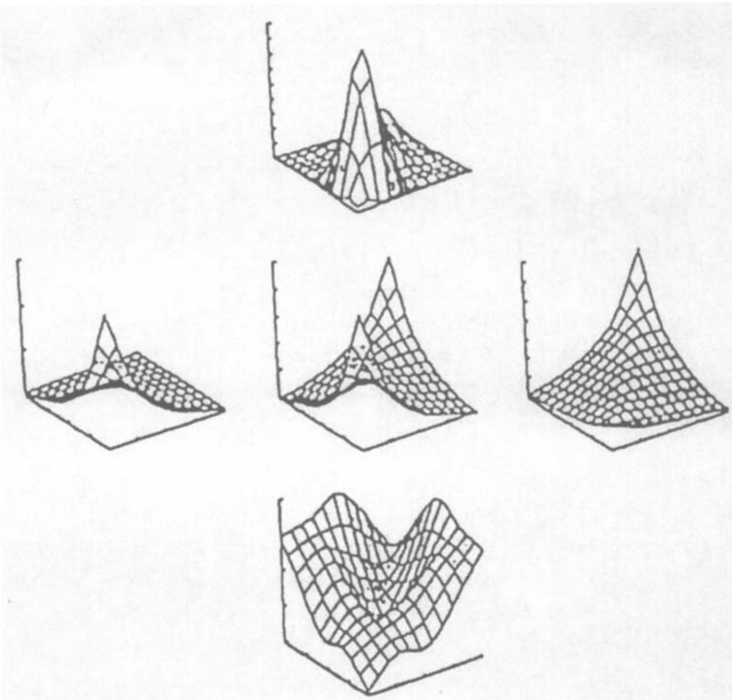

Figure 7. Gestalts Put Together from this Study. Above: TUD-Delft University of Technology 1990–94, relative frequencies F_{XY}. Center, left: Total mix of international gestalts, observed frequencies C_{XY}. Center, right: Mix of international medicine (SCI), relative frequencies F_{XY}. Bottom: Kaiser-Wilhelm Institut für Kohleforschung, relative frequencies F_{XY}.

lowing question arises: Is there any regularity for the distribution of coauthor couples in journals? Is there a continuation of Lotka's law on the third dimension?

In conclusion, the author suggests considering whether or not the conciseness principle, in all its succinctness, is only verifiable in scientists' communities, or whether it can also be largely extended to other social systems (see Figure 8)?

Acknowledgments

The author extends her appreciation to the Technology Foundation Utrecht (STW), the Netherlands, which had provided financial support. Work on this article was carried out partly at the Center for Science and Technology Studies (CWTS), Leiden University, the Netherlands.

The author thanks Wolfgang Glänzel (Budapest) and R. Wagner-Döbler (München) for making available the bibliographies in this paper.

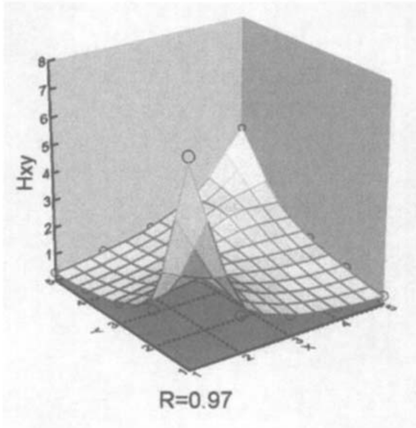

Figure 8. Gestalt Taken from the Special Interaction Index (H_{XY}). Source from which the gestalt in this study was developed: Marsden (1981, Table 1, p. 4). The values from the table were symmetrized (2,450 data).

References

Beaver, D. deB. & Rosen, R. (1979). Studies in scientific collaboration. Part III. Professionalization and the natural history of modern scientific co-authorship. *Scientometrics, 3*(1), 231–245.

Bohm, D. (1980). *Wholeness and the implicate order.* London: Routledge and Kegan Paul.

Capra, F. (1996). *Wendezeit. Bausteine für ein neues Weltbild.* München: Deutscher Taschenbuch Verlag, GmbH & Co.

Kretschmer, H. (1996). Die Gemeinschaft der Wissenschaftler—Ein nichtlineares dynamisches System im Verlaufe von 5 000 Jahren. In H. Meyer (Hrsg.), *25 Jahre Wissenschaftsforschung* (Heft 10, S. 34–55). Berlin: Schriftenreihe des Wissenschaftssoziologie Und—statistik e.V.

Laszlo, E. (1997). *Kosmische Kreativität.* Frankfurt am Main und Leipzig: Insel Verlag.

Lotka, A. J. (1926). The frequency distribution of scientific production. *Journal of the Washington Academy of Science, 16*(12), 317–323.

Marsden, P. V. (1981). Models and methods for characterizing the structural parameters of groups. *Social Networks, 3*(1), 1–27.

Metzger, W. (1986). *Gestalt-Psychologie. Ausgewählte Werke aus den Jahren 1950 bis 1982 herausgegeben und eingeleitet von Michael Stadler und Heinrich Crabus.* Frankfurt am Main: Verlag Waldemar Kramer.

Pribram, K. (1997). Vorwort. In E. Lazslo (Hrsg.), *Kosmische Kreativität* (S. 11–16). Frankfurt am Main und Leipzig: Insel Verlag.

Price, D. J. de Solla. (1963). *Little science, big science.* New York: Columbia University Press.

Prigogine, I. & Stengers, I. (1984). *Order out of chaos: Man's new dialogue with nature.* New York: Bantam Books.

Sheldrake, R. A. (1988). *The presence of the past.* New York: Times Books.

Stapp, H. P. (1993). *Matter, mind, and quantum mechanics.* New York: Springer-Verlag.

Winch, R. F.; Ktsanes, T.; & Ktsanes V. (1954). The theory of complementary needs in mate-selection: An analytical and descriptive study. *American Sociological Review, 19*(3), 241–249.

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