Jointly published by Elsevier Science Ltd, Oxford
and Akadémiai Kiadó, Budapest

Scientometrics,
Vol. 44, No. 2 (1999) 235–265

STRUCTURE AND DYNAMICS OF SCIENTIFIC NETWORKS. PART II: THE NEW ZIPF'S LAW, THE CLUSTERS OF CO-CITATIONS AND THE MODEL OF THE DESCRIPTOR PRESENCE

R. RUIZ-BAÑOS,[1] R. BAILÓN-MORENO, [2] E. JIMÉNEZ-CONTRERAS,[1] J.-P. COURTIAL[3]

[1]*Departamento de Biblioteconomía y Documentación, Facultad de Biblioteconomía y Documentación,*
Universidad de Granada, Campus de Cartuja, 18071 Granada (Spain)
[2]*Departamento de Ingeniería Química, Facultad de Ciencias, Universidad de Granada,*
Campus de Fuentenueva, 18071 Granada (Spain)
[3]*Laboratoire de Psychologie-Education-Cognition Développement (LabEcd),*
Université de Nantes, B.P. 1025, 44036 Nantes Cedex (France)

(Received October 27, 1998)

Here, the quantitative theory of translation is shown to be of great utility in describing scientific networks. In fact, we deduce a new Zipf's Law for the descriptors of a set of documents, based on the concepts of centres of interest and of irreversible parallel translations. This new law can be generalized to other phenomena, such as the distribution of the sizes of co-citation clusters. Finally, we have established the model, for descriptor presence in a network, which closely fits the values recorded.

Introduction

Scientometrics owes its existence primarily to the classic bibliometric studies based on indicators of scientific activity. The principal laws which give this bibliometrics meaning are on the one hand those referring to the size of science (Price's growth law and Brookes' obsolescence law), and on the other those referring to overall distributions of the actors: the laws of Lotka, Bradford and Zipf (these three laws being materially equivalent). Afterwards, with the advent of the first- and second-generation relational indicators (co-citation and co-word analyses), scientometrics finally makes its appearance as we know it today. In scientometrics, science is conceived as a network of actors who can be represented in the form of maps, either using relationships through bibliographical references or by using word association, normally descriptors of scientific articles or patents.[1–4] The theory of translation, proposed by *Latour*, provides a qualitative study of the dynamics of science as a continual interrelating of interests held by the actors.[5,6]

In Part I of the present work, we attempted to lay the mathematical and quantitative foundations of translation. That is, we presented a tool that we consider capable of accounting for both the classic laws of bibliometrics as well as the structure and dynamics of scientific networks expressed in the form of co-citations and co-words. We began with the precept that for a group of actors to undergo a translation process, that group must be capable of overcoming what we term the translation "barrier," a process governed by the Maxwell-Boltzmann distribution. We offer a conception of translation as the derivative of the quality function with respect to the spatial co-ordinates of the translation, where quality is comprised of those properties which differentiate one actor from another, and space being the setting in which the quality under consideration develops. We demonstrated with these premises the well-known principle of "success breeds success." Translations were divided into elemental and complex. The former are irreversible and in equilibrium, while the latter are combinations either in series or parallel to the former. With these complexes, it should be possible to demonstrate any bibliometric phenomenon.

Objective

In Part II of the present work, we seek to validate our model experimentally by performing the following:

1. Zipf's Law will be demonstrated for the descriptors of a set of scientific articles that cover an extensive scientific area.

2. The size data will be retaken for the co-citation clusters offered by *Van Raan*[7,8] covering science as a whole, and these will be analysed according to our criteria. Both *Van Raan* and *Small*[9] considered it feasible, from the distribution of these clusters, to view science as a fractal structure.

3. Finally, the concepts of irreversible translation and translation in equilibrium will be applied to the structure of a scientific network shown by the co-word analysis.

With these applications, we seek to generalize and validate our model for all sciencimetric spheres, from the standpoint both of classic bibliometrics and of relational co-citations and co-words.

Methods

The scientific field chosen was archaeology from 1980 to 1993, using Francis as the database, since this is the most complete in this field. The articles were indexed by expert professionals, and therefore the descriptors present in the documents are of high relevance and present no reason for their goodness to be distrusted. Nevertheless, despite the uniformity of criteria evident in the indexing, occasional disparities appear in gender, number, writing or simply typographical errors. Therefore, for the sake of precision, we proceeded to purify the descriptors, attempting not to detract from the original indexing. In addition, these were translated from French into Spanish.

The articles were divided into 12 periods, as shown in Table 1. The period listed as 1980 contains the articles from before 1980 as well as through 1983. The periods from 1983 to 1992 each include works published in the year considered plus the preceding and succeeding year. The period 1993 includes only 1992 and 1993.

Table 1
Distribution of records by periods

Period	Years	Records
1980	<1980, 1980, 1981, 1982, 1983	2.759
1983	1982, 1983, 1984	4.213
1984	1983, 1984, 1985	5.357
1985	1984, 1985, 1986	6.376
1986	1985, 1986, 1987	7.036
1987	1986, 1987, 1988	8.109
1988	1987, 1988, 1989	9.322
1989	1988, 1989, 1990	8.706
1990	1989, 1990, 1991	7.724
1991	1990, 1991, 1992	6.122
1992	1991, 1992, 1993	4.980
1993	1992, 1993	2.886

The archaeology network was reconstructed for each period using the co-words analysis (Leximappe program), using a minimum size cluster or theme of 4 descriptors and a maximum of 10. The occurrence and co-occurrence which was required of the descriptors presented different values for each period according to the number of the articles present, with an attempt to hold the vocabulary to roughly 700 words. A more detailed explanation of this procedure can be found in *Ruiz-Baños.*[10]

Results

In the Appendix, a sample of the first words of the all used by Leximappe in the making of the maps have been related, that is, all those words which in each period present an occurrence equal to or greater than that required. These are the words which enter to form part of the network. In the list, the occurrence rank (*Ro*) is specified as well as the frequency of appearance or relative occurrence expressed on a per-thousand basis with respect to the total occurrences (*Oc*) the number of periods in which the word appears as principal (S_P) thematic (S_T) or non-thematic (S_E) as well as the overall number of appearances (S_G). A principal word is one found within a cluster or theme in a central position; thematic means that the word forms part of the cluster, directly connected to a principal word; and an extra-thematic word is one used by Leximappe but does not appear in any theme.

Discussion

From the results, we can follow two approaches:

1. We can analyse the distribution of the overall vocabulary in a classical way in terms of occurrence-rank of the words, according to the Zipf-type distribution, and verify whether or not this vocabulary behaves in the same way as does the vocabulary of a sufficiently large natural language either written or spoken.

2. Given that the descriptors which form the network are not uniformly organized, but rather seek characteristic positions within the network, wich we call principal, thematic and extra-thematic, it would be useful to determine the probability which a descriptor or key word has of situating itself in any of these positions according to its occurrence rank. Finally, with this, we attempt to characterize the structure of the network in probabilistic terms.

Firstly, we shall study the vocabulary, following the first approach, bibliometric in character, to follow later the relational scientometric aspect of the second approach.

Inapplicability of Zipf's law for scientific networks

When we arrange the words of a text in descending order of their frequency of appearance or occurrence, *O*c, and denote the rank as *R*o, usually the distribution fits any of the following equations:

$$Oc = k_z Ro^{-1} \tag{1}$$

$$Oc = k_{\mathrm{b}} Ro^{B} \tag{2}$$

$$Oc = k_{\mathrm{brk}} (Ro + a)^{-1} \tag{3}$$

$$Oc = k_{\mathrm{m}} (Ro + m)^{B} \tag{4}$$

and which are considered by *Condon-Zipf*, *Booth* and *Federowicz*, and *Mandelbrot*, respectively.[11–15] All these equations require a double-logarithmic graphic representation. Therefore, the occurrence values (*Oc*) expressed in terms of per thousand with respect to the total occurrences (second column of the Appendix) have been represented in this way against the occurrence rank (*Ro*) in Figure 1.

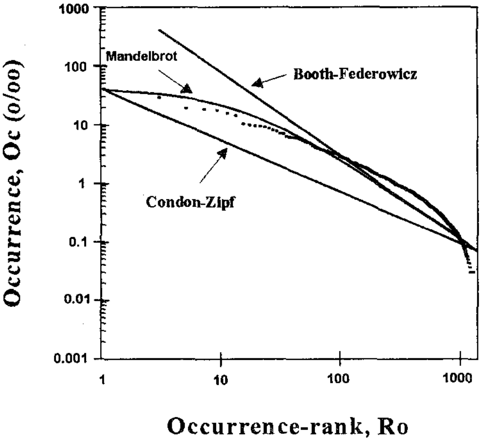

Fig. 1. Occurrence against occurrence rank. Application of reverse power laws

Together with these values, lines have been drawn corresponding to the four above equations. However, none of these adequately fit the observed values. Specifically, the line for *Booth* and *Federowicz* is expressed as:

$$Oc = 3468/Ro^{1.609} \tag{5}$$

with a very bad R^2 value (0.859). The Mandelbrot equation is expressed as:

$$Oc = 3468/(Ro + 14.7)^{1.609} \tag{6}$$

This latter equation adjusts somewhat better to extremely low occurrence-rank values (below 150), and badly to all others. The Brookes equation was not even depicted, since, on being less flexible than that of Mandelbrot, it would not give a satisfactory result.

Thus, we concluded that the distribution of the terms which make up the network do not fit any of the equations proposed to date for texts: Condon-Zipf, Booth and Federowicz, Brookes, and Mandelbrot. Consequently, it was necessary to search for an original expression which would satisfy the fit.

Deduction of a new Zipf's law. Centres of interest

Let there be a scientific network formed by a set of scientific articles, *N*. Let *W* be the vocabulary of all the key words used to index these articles. Let us arrange the key words in descending order of occurrence, noting as the occurrence rank (*Ro*) the position which each descriptor has in this list. The space being considered is one-dimensional, *Ro* being the geometric co-ordinate comprising this space.

Let us consider the quality "to be indexed by rank *Ro*" of the key words, which we shall quantify either as the occurrence of this key word (*Oc*) or, the equivalent, as the number of scientific articles which are indexed by the key word. The occurrence (*Oc*) of a key word is the result of the sum of three types of interests.

a) Interest as a fundamental centre (*F*).

b) Interest as a structural centre(*S*).

c) Interest as complementary centre (*C*).

That is:

$$Oc = F + S + C \tag{7}$$

The interest as a fundamental centre strongly diminishes as the rank (*Ro*) increases, according to a pseudo-irreversible translation:

$$\mathrm{F} \xrightarrow{k_1} \text{Not fundamental interest}$$

Analogously, the interests as structural centre and as a complementary centre diminish as the rank of the following form increases.

$$\mathrm{S} \xrightarrow{k_2} \text{Not structural interest}$$

$$\mathrm{C} \xrightarrow{k_3} \text{Not complementary interest}$$

An extremely low-rank key word, that is, with a very high occurrence, presents the three interests, the fundamental, the structural and the complementary. For somewhat higher occurrence ranks, the fundamental interest disappears and the structural and complementary remain. In extremely high ranks, with low frequencies, only complementary interest remains. This distribution of interests is because $k_1>k_2>k_3$.

The differential model is:

$$\mathrm{d}F/\mathrm{d}Ro = -k_1F \tag{8}$$

$$\mathrm{d}S/\mathrm{d}Ro = -k_2S \tag{9}$$

$$\mathrm{d}C/\mathrm{d}Ro = -k_3C \tag{10}$$

When we integrate the Eq. (8) from the rank Ro=1 to Ro=Ro, we get:

$$\int_{F_1}^{F} \frac{\mathrm{d}F}{F} = -k_1 \int_{1}^{Ro} \mathrm{d}Ro \tag{11}$$

$$F = F_1\, \mathrm{e}^{-k_1(Ro-1)} \tag{12}$$

$$F = F_1\, \mathrm{e}^{k_1}\, \mathrm{e}^{-k_1 Ro} \tag{13}$$

After the following:

$$F_0 = F_1\, \mathrm{e}^{-k_1} \tag{14}$$

we get:

$$F = F_0\, \mathrm{e}^{-k_1 Ro} \tag{15}$$

and analogously, from Eqs (9) and (10):

$$S = S_0\, \mathrm{e}^{-k_2 Ro} \tag{16}$$

$$C = C_0\, \mathrm{e}^{-k_3 Ro} \tag{17}$$

where F_0, S_0 and C_0 are the values of F, S and C in the imaginary rank Ro=0.

In agreement with Eq. (7), the occurrence of a key word with Ro is:

$$Oc = F_0\, \mathrm{e}^{-k_1 Ro} + S_0\, \mathrm{e}^{-k_2 Ro} + C_0\, \mathrm{e}^{-k_3 Ro} \tag{18}$$

This equation represents the mathematical expression of the new Zipf's Law for scientific networks. It contains 3 negative exponential terms representative of each of the different interests: fundamental, structural and complementary.

Validation of the new Zipf's law

In the deduction of the new Zipf's Law, Eq. (18), we assumed that $k_1>k_2>k_3$. This means that when the rank is high, the first and second negative exponential terms are practically nil. Under these circumstances, the occurrence is expressed by the following equation:

$$Oc = C_0 \, e^{-k_3 Ro} \tag{19}$$

for which the complementary interest is dominant. When we take logarithms, the preceding equation is linearized:

$$\ln Oc = \ln C_0 - k_3 Ro \tag{20}$$

In fact, when we represent $\ln(Oc)$ against Ro, a straight line should result. In Fig. 2, the observed occurrence values from the Appendix are represented using a semi-logarithmic diagram. We confirmed that for ranks greater than 200, the points aligned perfectly, indicating the only predominance of the exponential term of complementary interest. Beyond the rank 1100, the values fall somewhat, due to a cut-off effect. This distribution was obtained from the 12 partial distributions corresponding to each of the periods in which the interval under study was subdivided. The cut-off was set as a rank of roughly 700. In periods with few articles, this corresponds to very low occurrences, but in periods with many articles, it corresponds to somewhat higher occurrences. In these latter cases, key words of lesser occurrence are no longer considered and their values are not added to the overall distribution. Again, we refer to *Ruiz-Baños*[10] for a more extensive explanation of these procedures.

By linear regression of the values with ranks from 200 to 1100, and using Eq. (20), we get an extraordinarily good fit. The results are as follows:

$$k_3 = 3.610\ 10^{-3}$$
$$C_0 = 1.770$$
$$R^2 = 0.9991$$

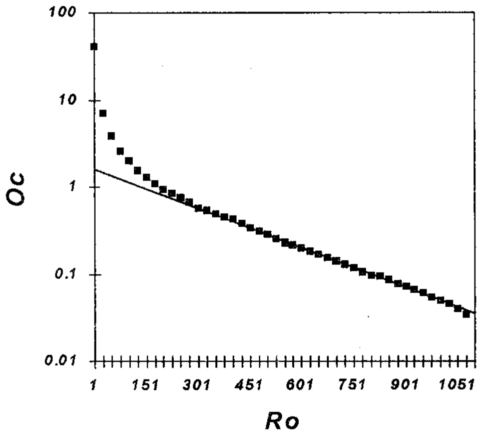

Fig. 2. Occurrence against occurrence rank. Semi-logarithmic diagram

When we consider ranks somewhat lower than 200, the negative exponential term representative of the structure interest gains importance. Under these conditions, the occurrence is given by the expression:

$$Oc = S_0 \, e^{-k_2 Ro} + C_0 \, e^{-k_3 Ro} \tag{21}$$

where Oc represents the values observed, and the second term the values calculated using Eq. (19), Oc_{calc}, the parameters of which we already know. Next:

$$Oc = S_0 \, e^{-k_2 Ro} + Oc_{\mathrm{calc}} \tag{22}$$

$$Oc - Oc_{\mathrm{calc}} = S_0 \, e^{-k_2 Ro} = \mathit{differences1} \tag{23}$$

$$\ln(\mathit{differences1}) = \ln S_0 - k_2 Ro \tag{24}$$

Figure 3 shows the logarithm of *differences1* against *Ro*, and we confirm that we have a straight line between the ranks 20 and 200, which we consider to be the zone of the predominant influence on the structural interest without interference from the fundamental one (this latter appears between approximately ranks 1 and 20). By linear regression, we obtain the following results for the parameters of structural interest:

$k_2 = 2.268\ 10^{-2}$
$S_0 = 7.812$
$R^2 = 0.990$

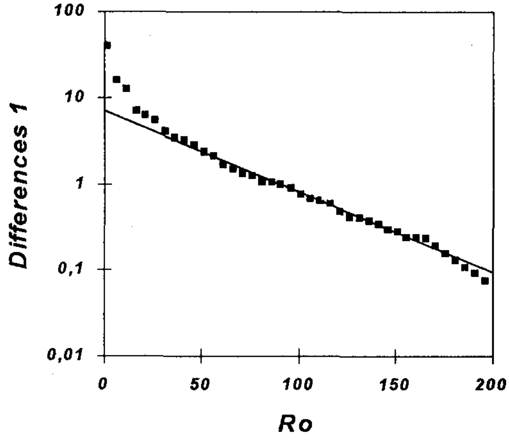

Fig. 3. Differences 1 against occurrence rank

Similarly:

$$\ln(\mathit{differences2}) = \ln F_0 - k_1 Ro \qquad (25)$$

where "*differences2*" now represent the difference between the observed and calculated values in Eq. (21). In Fig. 4, the diagram represents the logarithm of *differences2* against *Ro*. As the cloud of points in this case is more scattered than in previous cases, the linear regression was weighted, using as the weight of the dependent variable its value squared. In this way, we adjusted preferentially for the high occurrences. The results are as follows:

$k_1 = 0.1321$
$F_0 = 31.487$
$R^2 = 0.93$

With all the values obtained from the parameters, the occurrence can be calculated for any rank by:

$$Oc = 31.487\ e^{-0.1321Ro} + 7.812\ e^{-0.02268Ro} + 1.770\ e^{-0.00361Ro} \qquad (26)$$

Figure 5 presents the extraordinary agreement between the values observed and the line drawn in Eq. (26). In addition, as further confirmation, Fig. 6 shows the values of the occurrence logarithm calculated against the observed logarithm. The alignment over the diagonal is perfect, substantiating the validity of Eq. (18), and of the model which we have proposed for the distribution of the descriptors in the network, based on one complex translation comprised of 3 parallel irreversible translations representative of the fundamental, structural and complementary interests.

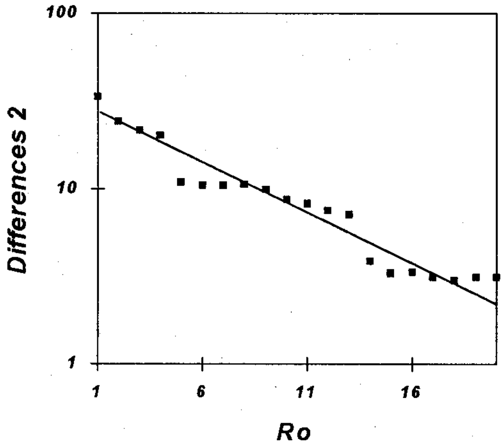

Fig. 4. Differences 2 against occurrence rank

Expansion of the new law of co-citation clusters

The new Zipf's Law which we have formulated satisfactorily reproduces the observed values of descriptor occurrence in scientific articles. It can also be generalized and may be applicable to other types of distributions, taking into account that, finally, these key words represent centres of interest. In fact, if a co-citation cluster has a large size, that is, it is representative of a great number of articles, then we can be sure that this represents a zone of the network which is of substantial scientific interest. Similarly, if a cluster represents a small number of works, its interest is weak. However, regardless of the size of the aggregate, this will always represent a centre of interest. Therefore, if we begin with this premise, the model proposed for the occurrence of descriptors may be valid for the size of the co-citation clusters.

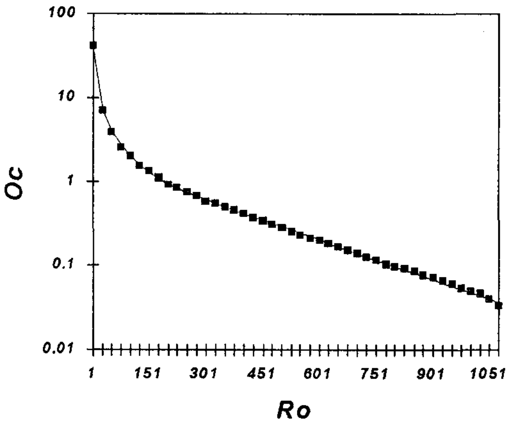

Fig. 5. Fit of descriptor occurrence values by the new Zipf's Law

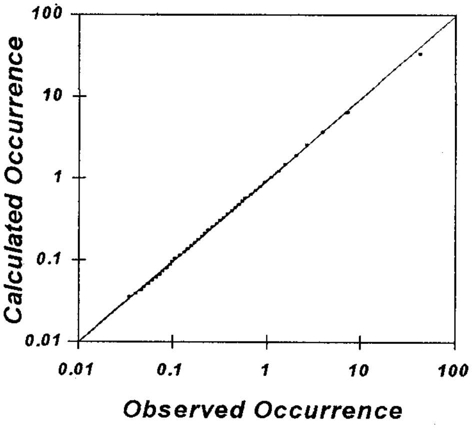

Fig. 6. Validation of the new Zipf's Law: calculated occurrences against observed occurrences

If we use the values of co-citation clusters C2 and C3,[7,8] we find excellent agreement with our model. A reverse power regression is possible up to the rank of roughly 400 (only 30% of the total distribution) at C2 and up to about the rank 100 (some 56% of the total distribution) at C3. On the other hand, with the new Zipf's Law which we have formulated, it is possible to adjust with fine precision 100% of the C2 and C3 distributions.

Figure 7 shows that the C2 distribution has been divided into 3 types of interests: from rank 1 to approximately 20, the large clusters represent centres of discipline-type interest, from 20 to 180, the clusters are centres of interest representative of fields, and finally the small clusters are centres of interest of the sub-field type. Translations are parallel and represented according to the following scheme:

$$\mathrm{Di} \longrightarrow \text{Not discipline}$$

$$\mathrm{Fi} \longrightarrow \text{Not field}$$

$$\mathrm{Sf} \longrightarrow \text{Not sub-field}$$

According to our new Zipf's Law, the equation representative for the C2 distribution would be the type:

$$g(r) = Di_0\mathrm{e}^{-k_1 r} + Fi_0\mathrm{e}^{-k_2 r} + Sf_0\mathrm{e}^{-k_3 r} \tag{27}$$

After the same fit technique as in the previous section, Eq. (27) becomes:

$$g(r) = 10^4\,\mathrm{e}^{-0.1367r} + 2.1\ 10^3\,\mathrm{e}^{-0.0173r} + 834\,\mathrm{e}^{-0.00243r} \tag{28}$$

with the value of R^2=0.997 for the fit of the calculated against the observed *g(r)* logarithm. The solid line in Fig. 7 represents the values calculated with Eq. (28), confirming the excellent agreement with the observed values over the entire distribution. The broken line represents the cluster sizes using a reverse power fit. We have verified that from the rank 300 or 400 the divergence becomes progressively more pronounced, and thus it is unacceptable to consider the structure of science to be fractal, as proposed by *Van Raan*[7,8] and *Small.*[9]

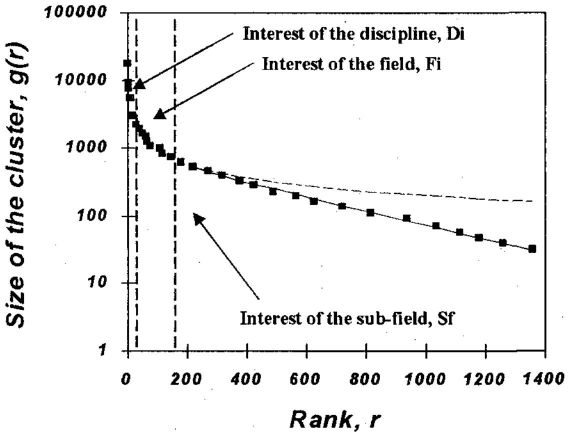

Fig. 7. Size of co-citation clusters C2 as a function of rank; interests of discipline, field and subfield

Now, on considering the C3 distribution, we find that the parallel translation model responds to only two types of interest, that of the disciplines and that of the fields (Fig. 8). The interest of the subfields disappears due to the effect of re-aggregating on going from C2 to C3. The model is then:

$$\mathrm{Di} \longrightarrow \text{Not discipline}$$

$$\mathrm{Fi} \longrightarrow \text{Not field}$$

The size of the cluster in C3 is calculated now by the equation:

$$g(r) = 3.59\ 10^4\ \mathrm{e}^{-0.1904r} + 4.4\ 10^3\ \mathrm{e}^{-0.0223r} \qquad (29)$$

Figure 8 shows the fit by means of Eq. (29) (the solid line), and the fit by means of the reverse power equation (the broken line). Again, we find that the reverse power law diverges notably from the observed values in the medium and high ranks, thus rejecting the hypothesis of the overall fractality of science. In the zone of discipline-type interest, we find that the exponential law proposed can calculate cluster sizes somewhat larger than those observed, while the reverse power law correctly determines some and underestimates others. This mixed situation appears to indicate that the large clusters do present a certain fractal behaviour. This phenomenon could be explained considering that the high k_1 value responds to a very strong translational capacity in the centres of discipline-type interest. Any actor approaching these centres would be immediately translated according to the discipline-type interests, without any possibility of the actor appreciably modifying the discipline. Let us consider the hypothetical publication of an article on the chemistry of the non-ionic tensioactives based on ethylene oxide. This article alone could scarcely change the foundations of the discipline chemistry, and clusters on the periphery of the discipline. The process would in this case be controlled by the diffusion of the actors or their drawing near to the interest centres (the tensioactives, the chemistry); we would be in a situation, quite common in natural processes, known as DLA (diffusion-limited aggregation).[16] On the other hand, the hypothetical article would indeed be capable of translating the interests of the tensioactive scientific subfield and of course those of non-ionic tensioactives based on ethylene oxide.

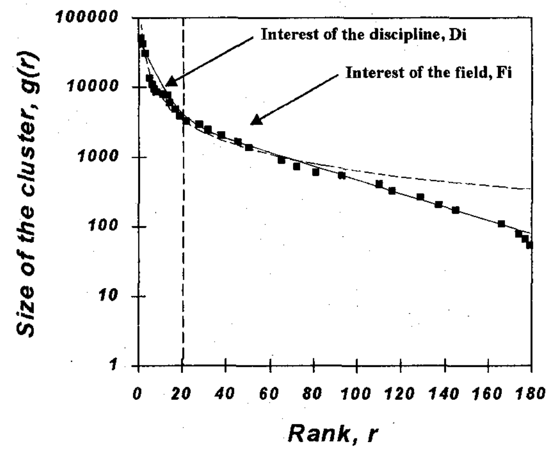

Fig. 8. Size of clusters of co-citations C3 as a function of rank; interests of discipline and field

Presence of the descriptors in a network

Let there be a scientific network generated by an extensive group of scientific articles and defined by the association of the descriptors of these articles according to the co-words analysis. Let us divide the network into a set of sub-networks or topics with a star-like structure. These star structures contain a central descriptor which we shall call the principal descriptor, and a group of descriptors, all directly joined to the principal one, called a thematic descriptor. Any descriptor, associated or not to others, which does not join directly to the principal one is called an extra-thematic descriptor.

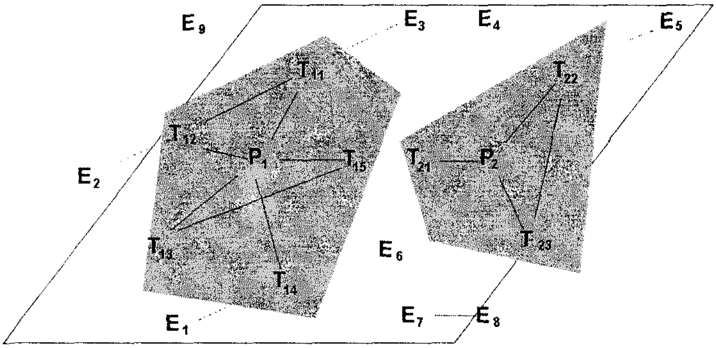

Fig. 9. Example of a network with two themes, defined by principal words P1 and P2

Figure 9 shows a small network formed by two themes and defined by the principal words P_1 and P_2. Around P_1 the descriptors T_{11} to T_{15} join and around the work P_2, the descriptors T_{21} to T_{23}. There are also 9 extra-thematic descriptors. Let us consider that the network evolves and is modified over n equal periods of time (for example, *n* years). With the passage of time, the descriptors change, both by disappearance and by the appearance of new ones. In addition, their position within the network can change, becoming sometimes principal, sometimes thematic and sometimes extra-thematic. On this set of premises, we shall define the following magnitudes:

a) Presence, S. Fraction of times that a given descriptor appears in a network over the *n* period of time.

$$S = N_a/n \tag{30}$$

where N_a is the number of periods in which the descriptor appears.

The presence can be of three types: principal (S_P), thematic (S_T) and extra-thematic (S_E), and in these cases N_a refers to appearances in the principal, thematic or extra-thematic, respectively. In addition, we shall consider a fourth type of presence, called overall (S_G), which is the sum of all the foregoing types of presence of a given descriptor. Figure 10 shows the presence of the descriptors in the archaeology network. We see that, despite fluctuations, regularity prevails. Thus, for example, the overall presence (S_G) drawn with a thick solid line is very high in the low occurrence rank, descending gently at first, and later diminishing more abruptly to stabilize in the high ranks. However, according to the figure, the presence (S) as such, being subject to unpredictable fluctuations, gives a blurry view and thus effaces the background of the positioning phenomenon of the descriptors in the network.

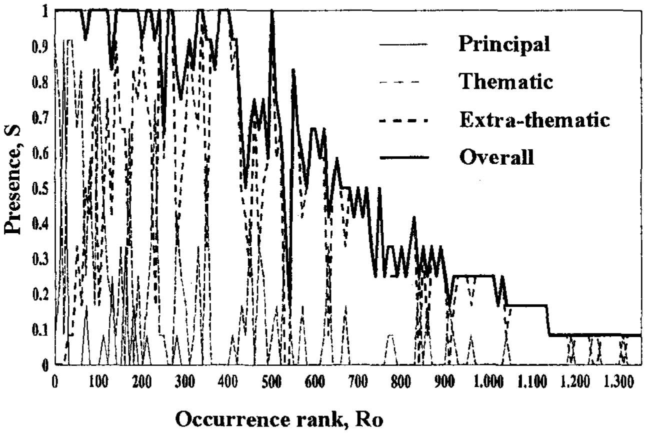

Fig. 10. Presence, S, of the descriptors in the archaeological network as a function of the occurrence rank

b) Probable presence. This is the presence that a descriptor is expected to have. It can also be overall (G), principal (P), thematic (T) or extra-thematic (E). This is determined experimentally as the mean value of the observed presence values, given that we are dealing with probability. *Egghe* and *Rousseau*[17] present a softening algorithm by weighted means, previously described by Winston, which consists of an iterative process of relaxation when each point is softened by adding to it the weighted mean of its left and right points. We have modified this technique, introducing more than one point to the left and more than one point to the right. Using four points to each side considered, all with identical weight and equal to 0.01, and after 33 steps in the iteration, we arrive at the probable presence values of Fig. 11. The overall presence (G) represented by the thick solid line descends to approximately the rank of 400, at first lineally and afterwards non-lineally. The probable thematic presence (T) passes through a maximum approximately in the 50-70 rank and afterwards progressively descends.

The extra-thematic presence (E) follows a similar profile, but in this case the maximum is situated towards the higher ranks, around 400. On the other hand, the principal presence (P) descends uniformly until becoming nullified at the rank of 200. We confirmed that the fundamental and structural interests of our new Zipf's Law reach precisely to this rank.

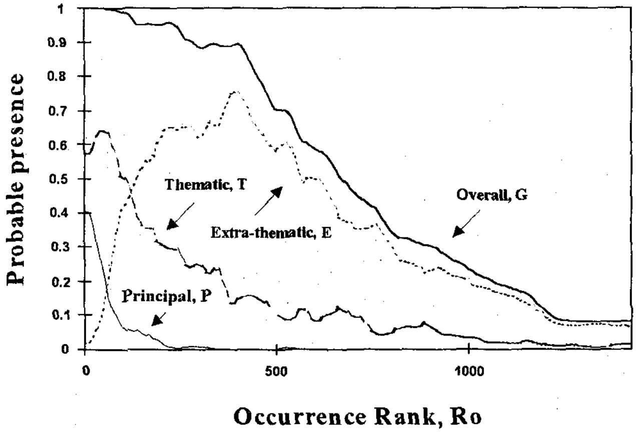

Fig. 11. Probable presence P, T, E and G of the descriptors in the archaeological network

c) Relative probable presence: This is a fraction of descriptor presence with respect to the overall probable presence. Logically, it can be principal (p), thematic (t), or extra-thematic (e).

$$p = P/G \tag{31}$$

$$t = T/G \tag{32}$$

$$e = E/G \tag{33}$$

Figure 12 shows the probable relative presence, except for the overall one, which naturally is constant with a value of unity. The analysis of the figure immediately leads us to the following assumptions:

a) The principal relative presence (p) decreases until nullifying itself. This induces us to conclude that we are dealing with an irreversible translation.

b) After a transition period, the thematic relative presence (t) stabilizes at ranks of over 400. The extra-thematic relative presence (e) similarly stabilizes at these ranks. Undoubtedly, this is a translation in equilibrium.

c) In short, it appears that we are confronted with a complex translation formed by the association in series of an irreversible translation and a translation in equilibrium. In the following sections, we shall mathematically develop this hypothesis.

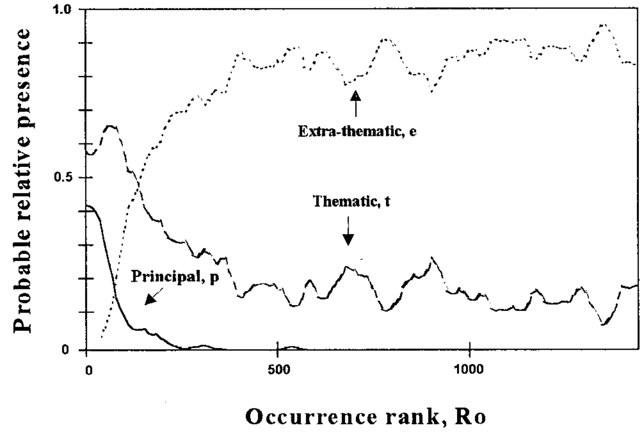

Fig. 12. Probable relative presence p, t and e of the descriptors in the archaeological network

Differential model of descriptor presence

The overall relative presence (g), which is equal to unity, is the sum of the relative probable presences of the principal (p), thematic (t) and extra-thematic (e) descriptors:

$$p + t + e = 1 \tag{34}$$

If we derive the foregoing expression with respect to the occurrence rank, the sum of the derivatives is null:

$$\mathrm{d}p/\mathrm{d}Ro + \mathrm{d}t/\mathrm{d}Ro + \mathrm{d}e/\mathrm{d}Ro = 0 \tag{35}$$

We accept that our model is such that as we move towards the higher occurrence ranks, the tendency is to lower the relative probability that a descriptor should be principal in favour of its being thematic. Finally, an equilibrium is established between the probability of being thematic and extra-thematic. Schematically, the model takes the following form:

$$\mathrm{P} \xrightarrow{k_1} \mathrm{T} \underset{k_3}{\overset{k_2}{\rightleftarrows}} \mathrm{E}$$

where k_1, k_2 and k_3 are translation constants of the elemental translations.

From here, we deduce that the relative presence of the principal descriptors decreases proportionally to p; that is:

$$\mathrm{d}p/\mathrm{d}Ro = -k_1 p \qquad (36)$$

The presence of the thematic descriptors decreases proportionally to t, and increase proportionally to p and e:

$$\mathrm{d}t/\mathrm{d}Ro = k_1 p - k_2 t + k_3 e \qquad (37)$$

Finally, the relative presence of the extra-thematic descriptors increases proportionally to t and decreases proportionally to e:

$$\mathrm{d}e/\mathrm{d}Ro = k_2 t - k_3 e \qquad (38)$$

We can corroborate that the sum of these three equations equals zero, as predicted in Eq. (35). To validate this model, it would be necessary to integrate it in order to compare the observed values against those calculated with the integrated equations.

Integral model of descriptor presence

The system formed by Eqs (36), (37) and (38) can be integrated analytically without difficulty. Here, in the interest of brevity, we will not dwell step-by-step detail in examining the integration process. Rather, we shall simply show the integrated equations of the relative presences and comment briefly on their significance.

a) principal probable relative presence:

$$p = p_1 \, \mathrm{e}^{-k_1(Ro-1)} \tag{39}$$

where p_1 is the relative presence of the word with an occurrence range equal to 1, that is, the relative presence of the most abundant word of the network. According to eq. (39), the principal relative presence declines according to a negative exponential function.

b) Thematic probable relative presence:

$$t = C_1 + C_2 \, \mathrm{e}^{-(k_2+k_3)Ro} + C_3 \, \mathrm{e}^{-k_1 Ro} \tag{40}$$

The thematic probable relative presence is a combination of two negative exponentials, one dependent on the in equilibrium translation constants, and the other on the irreversible translation constant. The constant C_3 presents the following complex expression:

$$C_3 = (k_3-k_1)/(k_1-k_2-k_3)p_1 \, \mathrm{e}^{k_1} \tag{41}$$

If *Ro* tends to infinity (it would suffice to be adequately large), the exponentials of Eq. (40) are nullified and then:

$$t_\infty = C_1 \tag{42}$$

The constant C_1 is the limit value reached by the relative probable presence of the thematic descriptors when *Ro* is infinitely high. In practice, it is the value reached in the equilibrium zone.

If the occurrence rank is equal to unity, replacing in the Eq. (40) and solving it, we find that:

$$C_2 = (t_1-C_1-C_3 \, \mathrm{e}^{-k_1}) \, \mathrm{e}^{k_2+k_3} \tag{43}$$

where t_1 is the thematic probable presence for the rank one.

c) Extra-thematic probable relative presence: As the sum of the relative probable presences is equal to unity, as indicated by Eq. (34), the calculation of *e* is immediate:

$$e = 1 - p - t \tag{44}$$

When the rank is high and *p* is eliminated, the extra-thematic relative presence reaches equilibrium and its constant value is equal to 1–*t*.

d) Overall probable presence: Once the equations are found for the relative probable presences *p*, *t* and *e*, we can determine the probable presences *P*, *T* and *E* with the help of Eqs (31), (32) and (33) if we know the integrated equation of *G*.

In the model that we are developing, we accept that two clearly differentiated zones exist: the pre-equilibrium zone or pre-critical, and the equilibrium zone or post-critical. The transition between the two is smooth with respect to principal, thematic and extra-thematic probable presence, but not with respect to overall presence (G). In fact, we perceive a nucleus of descriptors of high frequency with a behaviour markedly different from the rest of the descriptors and which is sharply separated with regard to what we call the critical point.

In the pre-equilibrium zone, the descriptors present such translation capacity that the decrease in overall presence is constant and independent of G. In fact, if G_0 is the maximum value attainable by the quality of overall presence, and G_i is the value capable of overcoming the translation barrier, the variation of G with respect to Ro will be proportional to G_i:

$$\mathrm{d}G/\mathrm{d}Ro = -aG_i \tag{45}$$

We admit that below the rank of the critical point (Ro_c), practically all the actors have sufficient capacity to surmount the translation barrier, and therefore in this case $G_i=G_0$ and the previous equation takes on the following aspect:

$$\mathrm{d}G/\mathrm{d}Ro = -aG_0 \tag{46}$$

and as a and G_o are constant:

$$\mathrm{d}G/\mathrm{d}Ro = -q \tag{47}$$

The integration of this equation leads us to a straight line with ordinate at the origin equal to 1 and with slope q:

$$G = 1 - qRo \text{ for } Ro \leq Ro_c \tag{48}$$

where Ro_c is the rank of the critical point.

On the other hand, in the equilibrium zone, the decrease in the overall presence will govern as an irreversible translation; afterwards:

$$\mathrm{d}G/\mathrm{d}Ro = -k_G G \tag{49}$$

which, after integration becomes, as expected, a negative exponential function:

$$G = G_0\, \mathrm{e}^{-k_G Ro} \text{ for } Ro \geq Ro_c \tag{50}$$

where K_G is the translation constant of the process, and G_0 is a hypothetical overall presence in a rank equal to zero. For the equations of the pre-equilibrium and equilibrium to be cut off at the critical point, the constant G_0 must have the value:

$$G_0 = G_c \, e^{(1-G_c)/q} \tag{51}$$

where G_c is the presence at the critical point.

It is evident that the equations which we have deduced for the presence of the descriptors are valid, and therefore our model is correct, only when they fit well with the observed values. Table 2 shows the values calculated from the equation constants. In Fig. 13, the observed values for probable relative presence are compared with those calculated with Eqs. (39), (40) and (44). We verified that, although the form of the curves is complex, particularly that of the thematic relative presence (which passes through the maximum), the agreement is excellent. Figure 14 provides comparisons of the calculated values for P, T, E and G solving the Eqs. (31), (32), (33), (48) and (50), with the observed values. The fit is very good, both in the pre-equilibrium and the equilibrium zones. All these graphic representations corroborate the goodness of our model for the presence of descriptors in the network, and strongly supports our quantitative theory of translation.

Table 2
Values of the parameters in the model for descriptor presence in the archaeological network

Parameter	Symbol	Value
Translation constant of P to T	k_1	0.0214
Translation constant of T to E	k_2	0.0094
Translation constant of E to T	k_3	0.0018
Principal probable relative presence at rank 1	p_1	0.541
Thematic probable relative presence at equilibrium	$C_1=t_e$	0.160
Pre-exponential constant	C_2	1.475
Pre-exponential constant	C_3	-1.066
Slope of probable overall presence at zone of pre-equilibrium	q	-3.61×10^{-4}
Translation constant of probable overall presence at zone of equilibrium	K_G	-2.398×10^{-3}
Probable overall presence at hypothetical zero rank	G_0	2.277

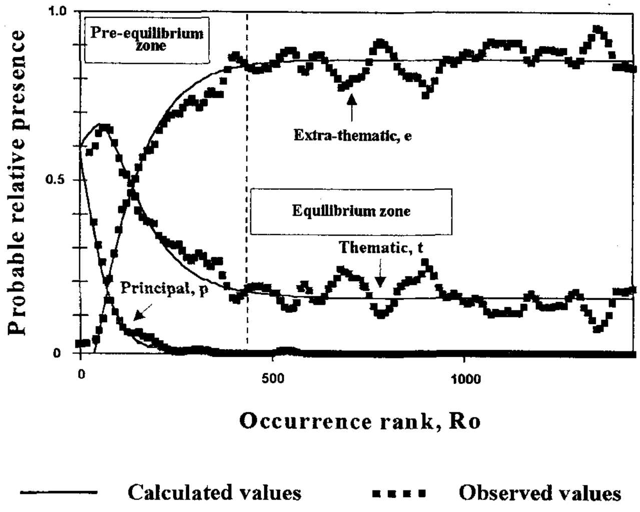

Fig. 13. Comparison of the probable relative presence p, t and e, observed and calculated

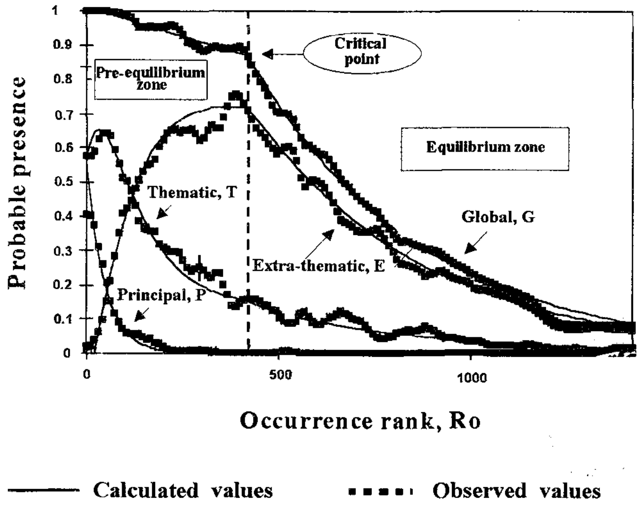

Fig. 14. Comparison of the probable presences P, T, E and G, observed and calculated

Conclusions

We have used our quantitative translation model in the three fundamental spheres of scientometrics: classic bibliometrics, co-citation analysis and co-word analysis. In all cases, we have been able to account satisfactorily for the phenomena observed, using the concepts of irreversible translation, equilibrium translation and centres of interest.

In addition, we have confirmed that the equations proposed to represent the frequency of appearance of words in a text, such as that of Condon-Zipf, Booth-Federowicz, Brookes and Mandelbrot, are inadequate to represent the occurrence of descriptors from a group of scientific articles. This demonstrates that the phenomenon of constructing a speech, written or spoken, differs in nature from the constructing of scientific facts. In fact, Zipf's classic reverse power distribution laws reveal a fractal construction, as proposed by Mandelbrot, whereas, as we demonstrate here, the descriptors present a different behaviour, explicable by a complex translation consisting of three parallel irreversible translations. Depending on the rank, the descriptors present one, two or three of these translations simultaneously, depending on the representation at the complementary, structural or fundamental interest centres within the network. Consequently, the mathematical expression of the new Zipf's Law deduced bears three addends, each formed by a negative exponential function.

This idea of simultaneous action of three types of interests can be generalized to the structure of scientific networks. Taking the size values of the co-citation clusters C2 and C3, published by *Van Raan*, and analysing them in the same way, we arrive at the conclusion that scientific networks are constructed around interest centres of the sub-field, field and discipline type, as *Van Raan* also proposed, but according to a process of translation and not as a basically fractal construction. Fractality clearly appears only in extremely large clusters of the discipline type in the C3 distribution, and this can be explained as a consequence of a highly intense translation of the disciplines on works and authors approaching them. Under these circumstances, though the phenomenon of translation exists, the construction of the facts is controlled by a diffusion-limited aggregation process (DLA) of fractal nature.

Finally, we have established the concept of presence of a descriptor in a network to achieve more than a simple count of key words in a Zipf-type distribution or a count of the number of articles in a co-citation aggregate. Persistence over a given time period of a descriptor in a network and its relative position within the clusters or themes, indicated by a co-word analysis, inform us much more clearly about the structure of the network and of the dynamics of the construction of scientific facts. We have verified that if we arrange the descriptors in descending order of their frequency within a time period, the probability that they will form part of the network decreases lineally until reaching a rank, called the critical point, in which it will decrease according to a negative exponential function. This critical point divides the network into two environments. In the first, the pre-critical or pre-equilibrium, formed fundamentally by key words of principal and thematic positions, the capacity of translation is so strong that all of these are capable of overcoming, without any appreciable difficulty, the translation barrier. In this case, the gradient of the overall presence is constant and does not govern by the principle "success breeds success." From the critical point onwards, within the post-critical or equilibrium zone, not all the descriptors have sufficient capacity to jump the translation barrier, and their probability of appearing in the network decreases exponentially. When we analyse the relative probable presences, we corroborate that the passage from a principal position to a thematic one is made by an irreversible translation. By contrast, when the rank is sufficiently high, an equilibrium translation is established between the thematic and extra-thematic positions.

In short, this work suggests that general bibliometric laws can be viewed as a consequence not of independent natural laws of human knowledge, but of interaction processes in which both the natural laws and the cultural ones interact. Our quantitative translation model is not only a purely mathematical model, but is an innovative way to conceive of scientometric laws. It is the first time that scientometric laws have been successfully explained by an interactive process that combines the natural and the cultural, that is, nature and society.

References

1. SMALL, H., Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24 (1973) No. July-August, 265-269.
2. CALLON, M., LAW, J., RIP, A., Eds. *Mapping the dynamics of science and technology: Sociology of science in the real world.* London: Mc Millan, 1986.
3. CALLON, M., COURTIAL, J.P., LAVILLE, F., Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry. *Scientometrics*, 22 (1991), No. 1, 155-205.
4. COURTIAL, J. P., *Introduction à la scientométrie. De la bibliométrie à la veille technologique.* Paris: Anthropos, 1990.
5. LATOUR, B., WOOLGAR, S., *Laboratory life.* London: Sage, 1979.
6. LATOUR, B., *Science in action.* Cambridge: Harvard University Press, 1987.
7. VAN RAAN, A. F. J., Fractal geometry of information space as represented by co-citation clustering, *Scientometrics*, 20 (1990), No. 3, 439-449.
8. VAN RAAN, A. F. J., Fractal dimension of co-citations, *Nature*, 347 (1990), 18 October, 626.
9. SMALL, H., Macro-level changes in the structure of co-citations clusters: 1983-1989, *Scientometrics*, 26 (1993), No. 1, 5-20.

10. RUIZ-BAÑOS, R., *Ciencimetría de Redes. Análisis de la investigación internacional sobre Arqueología mediante el Método de las Palabras Asociadas (1980-1993)*, doctoral thesis, Granada, 1997.
11. CONDON, E. U., Statistics of vocabulary. *Science*, 68 (1928), No. 300, p. 1733.
12. ZIPF, G. K., *Human behaviour and the principle of least effort.* Cambridg: Addison-Wesley Press, Inc., 1949.
13. FEDEROWICZ, J. E., A zipfian model on automatic bibliographic system: an application to MEDLINE. *Journal of the American Society for Information Science*, 33 (1982), No. 4, p. 223-232.
14. BROOKES, B. C., Ranking techniques and the empirical log law. *Information Processing and Management*, 20 (1984), p. 16-37.
15. MANDELBROT, B., *The Fractal Geometry of Nature*, W.H. San Francisco: Freeman and Co., 1982.
16. HARRISON, A., *Fractal in Chemistry*, Oxford University Press, 1995, p. 23-31.
17. EGGHE, L., ROUSSEAU, R., *Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science*, Amsterdam, etc.: Elsevier, 1990.

Appendix
Occurrence of words in the archaeological network

Ro	Oc	S_P	S_T	S_E	S_G	KEYWORD
1	41.241	1	11	0	12	EGIPTO
2	31.597	7	5	0	12	ARQUITECTURA
3	28.795	3	9	0	12	EXCAVACION
4	27.119	2	10	0	12	CERAMICA
5	17.909	7	5	0	12	EPIGRAFIA
6	17.289	2	10	0	12	CHINA
7	17.172	9	3	0	12	CRONOLOGIA: EGIPTO
8	17.123	7	5	0	12	ICONOGRAFIA
9	16.285	6	6	0	12	PALESTINA
10	14.929	4	8	0	12	SEPULTURA
11	14.274	3	9	0	12	ISLAM
12	13.442	4	7	1	12	INSCRIPCION
13	12.890	6	6	0	12	HISTORIA
14	9.530	8	4	0	12	CRONOLOGIA: CHINA
15	8.869	12	0	0	12	METODO
16	8.786	6	6	0	12	BRONCE (OBJETO)
17	8.446	1	11	0	12	ESCULTURA
18	8.197	11	1	0	12	ASIA MENOR
19	8.188	3	9	0	12	PINTURA
20	8.065	2	10	0	12	TEMPLO
21	7.957	11	1	0	12	INDIA
22	7.705	0	10	2	12	CRONOLOGIA
23	7.639	6	5	1	12	TIPOLOGIA
24	7.436	7	5	0	12	SIRIA
25	7.216	2	10	0	12	COLECCION
26	7.044	2	10	0	12	ARQUEOLOGIA
27	6.652	7	4	1	12	CRONOLOGIA: ISLAM
28	6.555	4	8	0	12	CRONOLOGIA: PALESTIN
29	6.303	4	8	0	12	MONEDA
30	6.261	5	7	0	12	MESOPOTAMIA
31	5.674	0	11	1	12	MOBILIARIO FUNERARIOv
32	5.631	0	8	4	12	PIEDRA (OBJETO)
33	5.620	2	10	0	12	ROMA
34	5.551	7	5	0	12	SOCIEDAD
35	4.925	0	12	0	12	DATACION
36	4.922	0	11	1	12	HABITAT
37	4.845	0	11	1	12	ECONOMIA
38	4.805	9	3	0	12	JAPON
39	4.788	6	5	1	12	DIOS Y HEROE: EGIPTO
40	4.776	4	8	0	12	PROXIMO ORIENTE
41	4.756	0	11	1	12	RELIEVE
42	4.693	0	11	1	12	ESTELA
43	4.596	6	4	2	12	CATALOGO
44	4.539	2	10	0	12	MUSEO
45	4.399	3	7	2	12	IRAN
46	4.327	2	9	1	12	ESTATUARIA
47	4.256	2	10	0	12	GLIPTICO
48	4.147	11	1	0	12	MAYA
49	4.093	4	7	1	12	COMERCIO
50	3.838	7	3	2	12	FENICIA

Ro	Oc	S_P	S_T	S_E	S_G	KEYWORD
51	3.830	0	8	4	12	TECNOLOGIA
52	3.815	1	11	0	12	PAPIRO
53	3.764	2	10	0	12	TABLETAS
54	3.755	2	6	4	12	RELIGION
55	3.718	0	10	2	12	SIMBOLICO
56	3.526	7	4	1	12	JORDANIA
57	3.495	0	11	1	12	DECORACION
58	3.226	0	7	5	12	LITERATURA
59	3.220	0	9	3	12	NECROPOLIS
60	3.152	4	3	5	12	BIBLIOGRAFIA
61	3.089	0	10	2	12	PROSPECCION
62	3.032	0	11	1	12	CRONOLOGIA: ASIA MEN
63	2.954	0	12	0	12	CRONOLOGIA: INDIA
64	2.929	0	12	0	12	MEJICO
65	2.886	1	7	4	12	ARTE
66	2.880	0	9	3	12	CASA
67	2.843	0	7	5	12	MANUSCRITO
68	2.794	0	11	1	12	TRADUCCION
69	2.734	6	1	5	12	HIDRAULICO
70	2.711	0	10	2	12	BUDISMO
71	2.700	2	3	6	11	SITIO ARQUEOLOGICO
72	2.691	0	8	4	12	URBANISMO
73	2.657	1	8	3	12	CHIPRE
74	2.625	0	11	1	12	ESPAÑA
75	2.594	2	8	2	12	MOSAICO
76	2.574	3	7	2	12	CRONOLOGIA: MESOPOTA
77	2.548	0	6	6	12	BIZANCIO
78	2.457	0	12	0	12	CRONOLOGIA: JAPON
79	2.437	0	5	7	12	CONGRESO
80	2.408	4	4	4	12	ARABIA
81	2.397	0	7	5	12	FORTIFICACION
82	2.391	1	8	3	12	CRONOLOGIA: SIRIA
83	2.371	0	11	1	12	IGLESIA
84	2.368	0	2	10	12	CIUDAD
85	2.351	2	3	7	12	GEOGRAFIA HISTORICA
86	2.328	3	2	7	12	TEXTIL
87	2.325	0	4	8	12	ORIGEN
88	2.297	0	4	8	12	SELLO
89	2.282	6	4	2	12	CARTAGO
90	2.277	1	8	3	12	PALACIO
91	2.274	0	2	10	12	EXPOSICION
92	2.259	1	4	6	11	RELACION CULTURAL
93	2.254	0	5	7	12	ESTATUILLA
94	2.248	0	5	7	12	COLOQUIO
95	2.239	0	7	4	11	ESTILISTICA
96	2.128	0	8	4	12	AGRICULTURA
97	2.111	1	4	7	12	CIVILIZACION
98	2.105	0	4	8	12	BIOGRAFIA
99	1.996	2	2	8	12	PERU
100	1.988	0	6	6	12	PORCELANA
101	1.988	0	10	2	12	ADMINISTRACION
102	1.956	0	9	3	12	ARMA
103	1.953	0	11	1	12	POLITICA
104	1.905	0	6	6	12	ONOMASTICA
105	1.885	0	11	1	12	RESTAURACION
106	1.882	0	8	4	12	IMPERIO NUEVO

Ro	Oc	S_P	S_T	S_E	S_G	KEYWORD
107	1.873	1	6	5	12	ARQUEOLOGIA SUBACUAT
108	1.839	2	4	6	12	BARCO
109	1.833	0	5	7	12	RITO
110	1.828	0	4	8	12	FORTALEZA
111	1.822	1	6	5	12	ASIRIA
112	1.810	0	12	0	12	PTOLOMEOS
113	1.810	1	4	7	12	CULTO
114	1.802	4	1	7	12	SUDAN
115	1.796	0	12	0	12	DIOS Y HEROE: INDIA
116	1.765	1	6	5	12	ADQUISICION
117	1.753	0	7	5	12	UTIL LITICO
118	1.682	2	5	5	12	BABILONIA
119	1.647	0	9	3	12	MEZQUITA
120	1.645	0	3	9	12	JOYA
121	1.616	0	3	9	12	MADERA (OBJETO)
122	1.559	0	12	0	12	CONSERVACION
123	1.553	1	4	7	12	ASIA CENTRAL
124	1.547	0	4	8	12	CULTURA
125	1.539	0	5	7	12	SARCOFAGO
126	1.533	0	1	10	11	ARQUITECTURA (ELEMEN
127	1.522	0	4	6	10	INFLUENCIA
128	1.516	0	6	2	8	BUQUE
129	1.510	0	0	12	12	BARRO COCIDO
130	1.504	0	5	7	12	SANTUARIO
131	1.502	3	2	5	10	AFRICA DEL NORTE
132	1.499	5	0	2	7	PERSIA
133	1.493	1	4	7	12	ESCRITURA
134	1.481	4	0	8	12	ARQUEOZOOLOGIA
135	1.467	3	1	8	12	METALURGIA
136	1.456	0	6	6	12	TEBAS
137	1.450	0	5	7	12	METAL (OBJETO)
138	1.444	0	7	5	12	NUMISMATICA
139	1.439	0	9	1	10	TRANSLITERACION
140	1.427	0	3	9	12	METROLOGIA
141	1.404	0	1	11	12	TOPOGRAFIA
142	1.393	0	7	5	12	OSTRACA
143	1.356	0	5	1	6	AÑO 1989
144	1.353	0	8	4	12	SIGLO +20
145	1.347	0	4	8	12	ESTRATIGRAFIA
146	1.338	0	4	8	12	ASTRONOMIA
147	1.330	0	7	5	12	ANALISIS QUIMICO
148	1.316	0	8	4	12	VASO
149	1.316	0	1	11	12	LEGISLACION
150	1.310	0	7	5	12	JERUSALEN, PALESTINA
151	1.307	0	4	8	12	VIAJERO
152	1.296	0	1	11	12	BARRO (OBJETO)
153	1.276	0	0	12	12	LEXICOGRAFIA
154	1.273	1	0	8	9	VIETNAM
155	1.264	0	1	11	12	CALIGRAFIA
156	1.250	1	2	9	12	ORO (OBJETO)
157	1.247	0	3	8	11	EVOLUCION
158	1.238	0	6	5	11	PLATA (OBJETO)
159	1.233	4	2	6	12	PALEOANTROPOLOGIA
160	1.230	0	10	2	12	INFORMATICA
161	1.227	4	0	8	12	SIGLO +19
162	1.224	2	5	3	10	TUNEZ (PAIS)

Ro	Oc	S_P	S_T	S_E	S_G	KEYWORD
163	1.218	0	11	1	12	CRONOLOGIA: MAYAS
164	1.213	0	1	11	12	PIRAMIDE
165	1.213	0	8	4	12	COPTO
166	1.207	0	12	0	12	RADIOCARBONO
167	1.195	0	3	8	11	ESTATUA
168	1.195	0	9	3	12	CUNEIFORME
169	1.175	0	4	8	12	VIDRIO (OBJETO)
170	1.150	0	6	6	12	CRONOLOGIA: IRAN
171	1.144	0	8	4	12	CRONOLOGIA: JORDANIA
172	1.138	0	2	10	12	TOPONIMIA
173	1.130	0	6	6	12	GRECIA
174	1.130	1	2	2	5	AÑO 1990
175	1.113	1	10	1	12	MING
176	1.093	0	0	12	12	ALIMENTACION
177	1.087	1	2	9	12	MOMIA
178	1.087	0	5	6	11	GUERRA
179	1.055	0	0	12	12	MONASTERIO
180	1.050	0	2	10	12	HELENISTICO
181	1.050	2	0	10	12	AMERICA DEL NORTE
182	1.035	0	1	11	12	AMULETO
183	1.032	0	3	9	12	PINTURA MURAL
184	1.027	0	4	8	12	ARCHIVO
185	1.021	0	5	1	6	CRONOLOGIA: PERSA
186	1.012	0	2	10	12	MAGIA
187	1.012	0	11	1	12	ANALISIS
188	0.992	0	0	11	11	ESCARABEO
189	0.990	0	1	11	12	HORNO
190	0.984	0	7	5	12	IMPERIO ANTIGUO
191	0.981	0	3	9	12	OBJETO
192	0.978	3	3	5	11	ITALIA
193	0.972	0	9	3	12	BRITISH MUSEUM, LOND
194	0.967	0	0	11	11	FLORA
195	0.961	0	1	11	12	SAQQARAH
196	0.947	0	1	11	12	VIDA COTIDIANA
197	0.938	0	0	12	12	PROSOPOGRAFIA
198	0.938	0	2	5	7	AÑO 1988
199	0.935	0	10	2	12	OTOMANO
200	0.932	0	7	5	12	SELLO CILINDRICO