Scientometrics, Vol. 4. No. 4 (1982) 283–295

TRANSITION FROM THE MACROLEVEL TO THE MICROLEVEL OF INFORMATION AT RANK DISTRIBUTION INVESTIGATIONS OF THE REPORT LITERATURE OF AN INTERNATIONAL INFORMATION SYSTEM

M. BONITZ, P. SCHMIDT

Central Institute for Nuclear Research, Academy of Sciences of the GDR, 8051 Dresden/GDR, P.O.B. 19 (GDR)

(Received April 27, 1981, in revised form July 28, 1981)

The coincidence method proposed earlier by one of the authors is applied to rank distribution studies of the report literature of the International Nuclear Information System, INIS, and a two-level concept is used to discuss the results of the present and previously reported investigations. Transitions between macrolevel and microlevel of information for constant forms of communication as well as transitions between different forms of communication at constant levels are compared. Escape of the information avalanche for the highly specialized single scientist, greater efficiency of the report literature compared with journal literature in the field of nuclear research, non-compatibility of macrolevel and microlevel of information could be confirmed quantitatively. It is an open question how to transform microlevel distributions into macrolevel distributions.

Coincidence method and two-level concept

The coincidence method was proposed and described in Ref.[1]. It is used to create the experimental data for our investigations and works during a routine SDI-process on the data base of the International Nuclear Information System, INIS. Coincidences occur between a set of user profiles, on the one hand, and a set of journals containing relevant articles or a set of institutions issuing relevant reports, on the other. Coincidence events are stored is matrices (working memories) for "article coincidences" and "report coincidences", respectively. After filling these files, journals and report issuing institutions are ranked according to decreasing coincidence numbers.

Two-level concept means, that ranking can be performed either for each single profile (ranking at the microlevel) or for the overall profile which is the logic sum of the single profiles (ranking at the macrolevel).

Scanning the experimental data in the described way and considering the two different levels permits to extract a multiplicity of information which cannot be obtained by conventional methods. This multiplicity of possible uses of the coincidence data results from the fact that different vehicles of communication – articles from journals, and reports – one the one hand, and different levels of communication – macrolevel and microlevel – on the other, can be taken into consideration and can be investigated in their combination, see Table 1.

The present paper closes a cycle of investigations. It was preceded by papers.[1–4]

The first paper[1] was devoted to the rank distribution of journals at the macrolevel. Here, of course, considerations came into play that had already been brought forward by *Bradford*;[5] the results, too, were similar, and their practical importance for the completion of stocks of journals is well known. From this point of view, the paper[1] represents just one out of many similar investigations conducted in widely varying fields and communicated increasingly in the literature of our field, even when these results were obtained by applying a novel and more sophisticated method; Z_{macr} in Table 1.

The investigation reported in paper[2] was already directed at another form of communication, i.e., reports, but it still was conducted at the macrolevel of information, that means, the results provided statements for an overall profile and the practical conclusions concerned the information center of an institution as a whole. But the fact that at the transition from the communication form 'journal article' to the communication form 'report' the other parameters remained unchanged (profiles, length of time for the experiment) made possible a statement which could not have been obtained by conventional ways and means: the communication form 'report' possesses a higher valency than the communication form 'journal article'; R_{macr} and → Ref.[2] in Table 1. In this way, the coincidence method making feasible the assessment comparison between different forms of communication yields already more informations than investigations would have yielded conducted on Z_{macr} and R_{macr} under noncomparable external conditions.

As far as we know, the step from the macrolevel to the microlevel of information has been taken for the first time in paper[3] Now, the rank distribution for single profiles, Z_{micr}, and derived from this, a 'mean' rank distribution Z_{micr} for an 'average individual scientist' were the center of attention, rather than the rank distribution for the overall profile. When at first considerations played a certain role whether and how *Bradford's* ways of thinking and description could be continued into the microlevel,[3] an increasing understanding developed for the fact that there exist fundamental qualitative differences between the macrolevel and the microlevel, and that events at the microlevel cannot adequately be described by concepts and approaches commonly and successfully applied at the macrolevel.

Table 1
Ways of utilizing the coincidence method

Form of scientific communication	Level of scientific communication: macrolevel	Level of scientific communication: microlevel
journal article		→ Refs[3,4]
journal article	Z_{macr} Ref.[1]	Z_{micr}, $\bar{Z}_{micr}$ Refs[3,4]
↓	Ref.[2] ↓	↓ [*] ↓ [*]
report	R_{macr} Ref.[2]	R_{micr}, $\bar{R}_{micr}$ [*]
report		→ [*]

The various investigational types result from the utilization of one and the same working memory by combining the two levels of scientific communication and (in our case) the two forms of scientific communication.

Z_{macr} cumulative frequency curve for journal articles at the macrolevel, all profiles; R_{macr} cumulative frequency curve for reports at the macrolevel, all profiles; Z_{micr} cumulative frequency curves for journal articles at the microlevel, single profiles; $\bar{Z}_{micr}$ cumulative frequency curve for journal articles at the microlevel, 'mean' single profile; R_{micr} cumulative frequency curves for reports at the microlevel, single profiles; $\bar{R}_{micr}$ cumulative frequency curve for reports at the microlevel, 'mean' single profile. Arrows signify what kinds of transition – from one level into the other at the same communication form, and from one communication form to the other at the same level, respectively – may be effected and discussed.

*The present paper.

The individual scientist's microworld of information is intangible and cannot be comprehended in the language of the macroworld of the information centers; Z_{micr}, $\bar{Z}_{micr}$ and → Refs[3,4] in Table 1.

Our paper [*] describes the step from the macrolevel into the microlevel for the communication form 'report'. With that, the last square in Table 1 is filled in. Again, identical experimental conditions result in gaining information that could not be obtained by conventional rank distribution investigations: in addition to the cumulative frequency curve $\overline{R}_{micr}$ for a 'mean' profile (for an 'average' scientist), it is not only the differences between R_{macr} at the macrolevel and $\overline{R}_{micr}$ at the microlevel that become accessible to investigation, but also the transitions between the different forms of communication within the microlevel itself where both the transitions $\overline{Z}_{micr} \rightarrow \overline{R}_{micr}$ and $Z_{micr} \rightarrow R_{micr}$ for every single profile can be examined; R_{micr}, $\overline{R}_{micr}$ and $\rightarrow$ [*] in Table 1.

Rank distributions of report institutions at the macrolevel. Experimental results

Table 2 illustrates the fact that the outcome of report investigations is the ranking of institutions that publish reports. The practical significance of such investigations lies in the identification of the 'key (core) institutions' (analogous to key (core) journals), the exchange of reports with which is recommendable because they are responsible for the greatest part of the total report potential (in relation to the overall profile). The 30 institutions ranking foremost in Table 1 (out of a total of 330) publish two thirds of the reports relevant to our overall profile. It is quite obvious that these institutions are of interest not only for the exchange of reports.

The working memory corresponding with Table 2 is presented in different ways in Figures 1 and 2. In Fig. 1 the presence of coincidences is indicated by a point at the corresponding memory location, whereas in Fig. 2 the number of stored coincidences is indicated, too.

The representation by points has been provided additionally in order to permit 'at first sight' an impression of the 'density distribution' of the points across the working memory.

The rank distributions of the report institutions have been established for each profile, each profile having a rank distribution of its own (not indicated here) which does not in general coincidence with that of the overall profile. The statistically most significant and useful eight distributions have been selected, the corresponding cumulative frequency curves R_{micr} being shown in Fig. 3. It is well known that this presentation is to be preferred for practical purposes in spite of the fact that it is somewhat less illustrative than rank distributions are, because it is possible to read with ease that rank at which, for example, one third, two thirds etc. of the informations are achieved. Besides, one glance will be enough to see just how strong the deviation from *Bradford*'s law of distribution is which,

Table 2
Rank distribution of the report institutions. From reference 2.

Rank z^*	Designation of report	Number N_{z^*}	Publishing Institution
1	CONF	457	USAEC Technical Information Center, Oak Ridge Tenn.
2	INIS-mf	430	International Atomic Energy Agency, Vienna
3	AED-Conf	309	Zentralstelle für Atomkernenergie-Dokumentation, Leopoldshafen
4	KFK	193	Kernforschungszentrum Karlsruhe
5	ORNL-TM	155	Oak Ridge National Lab., Tenn.
6	ORNL	116	Oak Ridge National Lab., Tenn.
7	DOCKET	98	USAEC Technical Information Center, Oak Ridge, Tenn.
8	CEA-CONF	95	Commissariat à l'Énergie Atomique
9	EUR	94	Commission of the European Communities, Brussels
10	LA	88	Los Alamos Scientific Lab., N. Mex.
11	JUEL	87	Kernforschungsanlage Juelich G.m.b.H.
12	BARC	82	Bhabha Atomic Research Centre, Bombay
13	GA-A	78	Gulf General Atomic, Inc., San Diego, Calif. (General Atomic Co., San Diego, Calif.)
14	IAEA	73	International Atomic Energy Agency, Vienna
15	COO	72	USAEC Chicago Operations Office, Ill.
16	ANL	56	Argonne National Lab., Ill.
17	ICP	54	Allied Chemical Corp., Idaho Falls, Idaho. (Idaho Chemical Corp., Idaho Falls)
18	JAERI-M	54	Japan Atomic Energy Research Inst., Tokyo
19	UCRL	53	California Univ., Berkeley, California Univ., Livermore
20	ZfK	45	Zentralinstitut für Kernforschung, Rossendorf bei Dresden
21	BNWL	43	Battelle-Northwest, Richland, Wash. (Battelle Pacific Northwest Labs., Richland, Wash.)
22	HEDL-SA	40	Hanford Engineering Development Lab., Richland, Wash., (WADCO Corp., Richland, Wash.)
23	HEDL-TME	37	Hanford Engineering Development Lab., Richland, Wash.
24	BNWL-SA	36	Battelle-Northwest, Richland, Wash.
25	WASH	35	USAEC, Washington, D.C.
26	ERDA	31	Energy Research and Development Administration, Washington, D.C.
27	CEA-R	30	Commissariat à l'Énergie Atomique
28	AECL	30	Atomic Energy of Canada Ltd., Chalk River, Ontario
29	MLM	28	Mound Lab., Miamisburg, Ohio
30	ARH	27	Atlantic Richland Hanford Co., Richland, Wash.

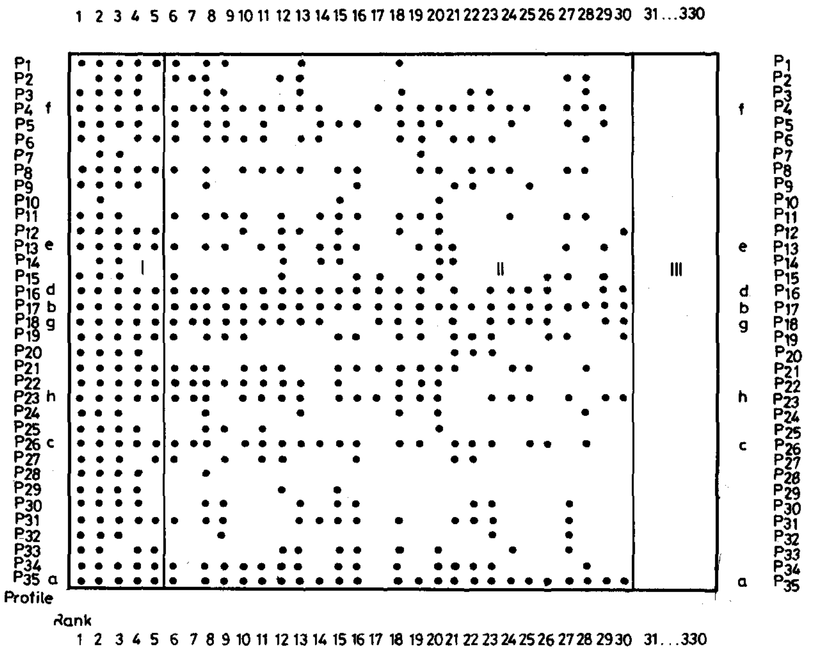

Fig. 1.Working memory for the report investigations. $P_1 \ldots P_{35}$ profiles. a, b, c, d, e, f, g, h profiles selected for the investigation. Points: presence of a coincidence between report institution and profile. Empty spaces: there is no coincidence between report institution and profile. Ranking order: corresponding to the table showing the report institutions, see Table 2. I, II, III zones with the same number of reports N/3, corresponding ranks 1–5, 6–20, 31–330

to be sure, should be a straight line in the semi-logarithmic presentation chosen by us.

The mean curve $\overline{R}_{micr}$ formed from the eight single curves of Fig. 3 which in a sense is representative for an 'average' individual scientist is contrasted with the cumulative frequency curve R_{macr} for the overall profile in Fig. 4. This comparison forms a first basis for answering the question: what happens at the transition from the macrolevel to the microlevel?

The transition from the communication form 'journal article' to the communication form 'report' at the macrolevel has already been discussed in paper.[2] The present paper now offers the chance to investigate the transition at the microlevel, too, that is to say at the level of the individual scientist. For the mean curves

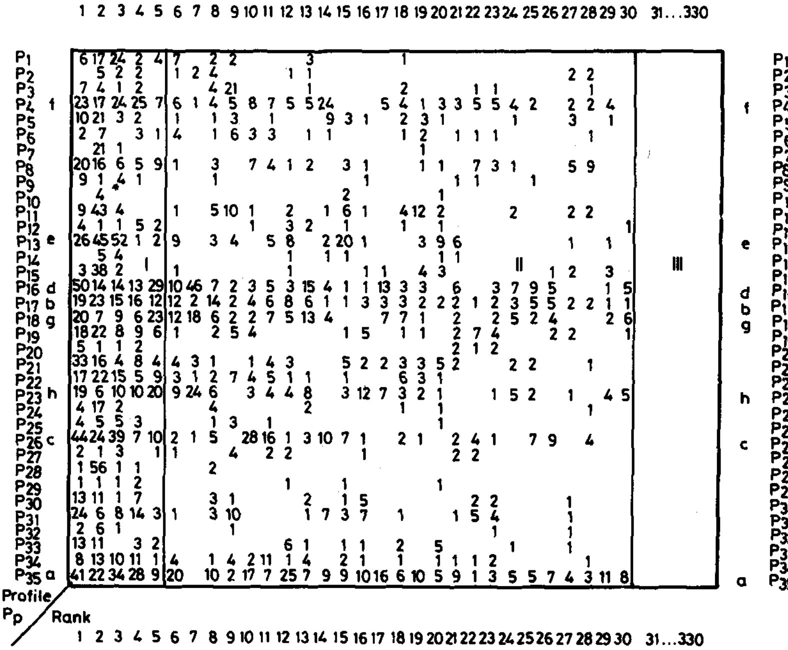

Fig. 2. Working memory for the report investigations. Structure as in Fig. 1, points have been replaced by the number of coincidences. Any report relevant to more than one profile is assessed according to its frequency.

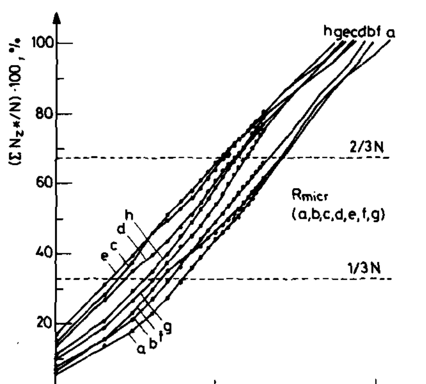

Fig. 3. Cumulative frequency curves R_{micr} at the microlevel for profiles a, b, c, d, e, f, g, h. ΣN_{z*} number of reports up to rank z*. N total number of relevant reports. To the various ranks there correspond generally different report institutions for different curves, that is, each of the selected profiles is provided with a different rank distribution.

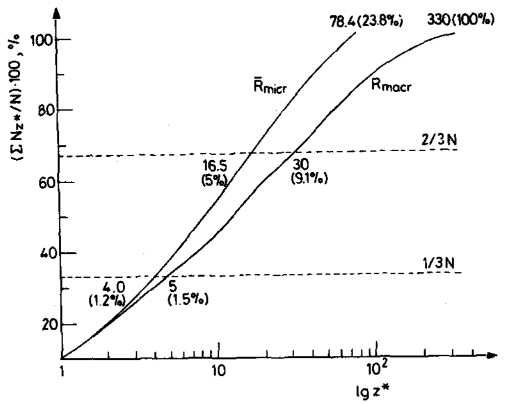

Fig. 4. Cumulative frequency curve R_{micr}. R_{micr} is the mean curve for a single profile at the microlevel, derived from the curves a, b, c, d, e, f, g, h in Fig. 2. R_{macr} cumulative frequency curve for reports at the macrolevel, all profiles (overall profile). Percentages in parentheses refer to the total number of 330 report institutions. R_{macr} from Ref.[2]

Z_{micr} and R_{micr} the basis is provided by Fig. 6, for selected *single profiles with* the corresponding curves Z_{micr} and R_{micr} by Fig. 5. Here the main question is whether phenomena of the macrolevel will make their appearance at the microlevel in the same way. This is of fundamental importance for deciding whether or not the differentiation between the macrolevel and the microlevel of information which is suggested by the coincidence method, will prove fruitful. To this purpose, all four cumulative frequency curves obtained by the investigations up to now will be presented once more in Fig. 6: Z_{macr}, $\bar{Z}_{micr}$, R_{macr} and R_{micr}.

Discussion of results

From the intense point density near the foremost ranks there follows (Fig. 1) that the cumulative frequency curves for the single profiles (Fig. 3) and the curve R_{micr} derived from them for a mean profile (or, rather, for the profile of an individual scientist generated by averaging) must strongly be shortened as compared with the corresponding curve R_{macr} at the macrolevel. For this conclusion we depend on experiences made and presented in papers.[3,4]

A rough measure for the 'density distribution' is provided by the mean allocation of coincidences with single profiles to the ranks in the three *Bradford* zones I,

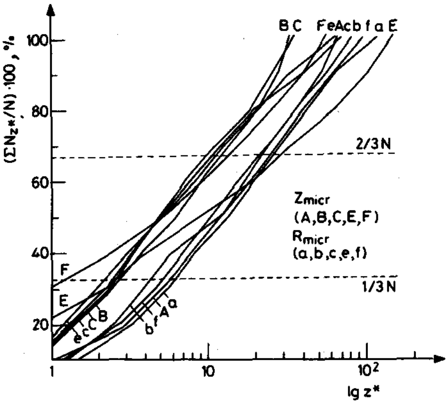

Fig. 5. Cumulative frequency curves at the microlevel. Z_{micr} cumulative frequency curves for journal articles for the identical profiles as used for the report investigations: curves A, B, C, E, F from Refs[3,4]. R_{micr} cumulative frequency curves for reports, single curves a, b, c, e, f (present paper). Profiles P_{35}: (A, a), P_{17}: (B, b), P_{26}: (C, c), P_{13}: (E, e), P_4: (F, f) correspond to curves Z_{micr} and R_{micr}

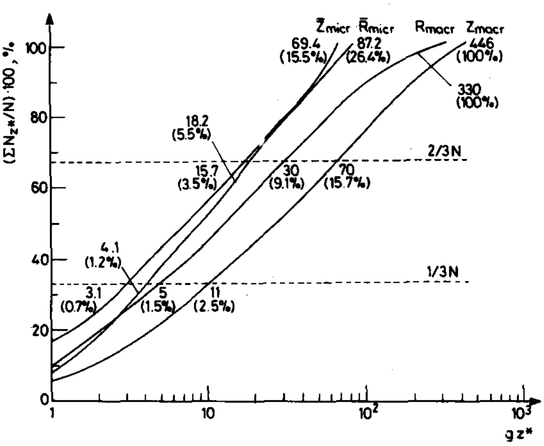

Fig. 6. Cumulative frequency curves. Confrontation of all results corresponding to Table 1: Z_{macr}, R_{macr}, $\overline{Z}_{micr}$, $\overline{R}_{micr}$. The percentages in parentheses refer to the respective total numbers of units of the corresponding communication form (333 report institutions and 446 titles of journals, respectively). Ranks z* cannot be made to show actual titles of journals or names of institutions in this presentation.

Table 3
Mean allocation of profiles to ranks

	I	II	III
Journal articles Ref.[1]	12 (ranks 1–11)	5 (ranks 12–70)	1.5 (ranks 71–446)
Reports Ref.[2]	29 (ranks 1–5)	15 (ranks 6–30)	3 (ranks 31–330)

Numbers indicate within the respective zones the mean number of profiles per rank for which coincidences are present. In each of the three zones I, II, III there are N/3 coincidences present. With journal articles, for instance, there appear in zone II coincidences for 5 profiles per rank on an average; with reports for 15 profiles per rank.

II and III. In Table 3 these profile allocations of the ranks for journal articles and for reports are confronted with each other. While, for example, in the case of journal articles the 11 'key (core) journals' (zone I) being responsible for one third of all relevant articles, 'serve', on an average, 12 profiles (out of 35), yet in the case of reports the 'key institutions' publish one third of all relevant reports for 29 profiles on an average. This means an enormous increase in effectivity achieved by the transition from the communication form 'journal article' to the communication form 'report'. We would like to stress that this statement which should lead in practice to a far better utilization of the communication form 'report', has only been made possible and could only be formulated by the inclusion of informations from the microlevel.

The higher valency of the communication form 'report' can be proved quantitatively in still another way. In Table 4 the respective total numbers of relevant journal articles and reports for 5 profiles are shown in relation to each other. It is clearly to be seen that on an average, a profile 'needs' 1.8 times as many articles as reports. Some profiles are considerably below this value, some others are above. The value $\Sigma N_{art} / \Sigma N_{rep} = 1.3$ should be representative, resulting from the macrocurves for the overall profile (see reference 2). This value lies clearly below the values 2 and 2.5, respectively, presenting the relation of articles to reports in the INIS overall file (Ref.[6] and/or Ref.[7]).

It goes without saying that 'higher valency of the communication form report' does not mean that perhaps in reports more important things are being published than in journal articles. It is a typical statement of the macrolevel which needs not hold true for every single case at the microlevel.

Table 4
N-values for 5 selected profiles

N_{art}	N_{rep}	N_{art}/N_{rep}
A: 570	a: 559	1.0
B: 327	b: 270	1.2
C: 396	c: 295	1.3
E: 1531	e: 311	4.9
F: 250	f: 317	0.8
$\Sigma N_{art} = 3075$	$\Sigma N_{rep} = 1752$	$\Sigma N_{art}/\Sigma N_{rep} = 1.8$

N_{art} total number of relevant journal articles, curves A, B, C, E, F; N_{rep} total of relevant reports, curves a, b, c, e, f.

Two important results may be derived for certain by means of information from the microlevel: first, the shortening (displacement to the left) of the curve for the individual scientist (R_{micr}) relative to the macrocurves for journal articles (Z_{macr}) and reports (R_{macr}); second, the higher valency of the communication form 'report' compared with the communication form 'journal article' within the system of scientific communication.

The shortening of the cumulative frequency curve for the individual scientist, R_{micr}, is effected to such a measure that the total number of report institutions which he has to watch remains small enough for him to handle. By this, proof is provided quantitatively that the individual scientist will still even today be capable of realizing the fundamental principle of all scientific work – to be fully informed about all the scientific information relevant to his tasks and objectives in scientific and technical practice. This conclusion seems to us to be of particular importance since quite frequently the validity of this statement is doubted in the face of the information avalanche. In simple language, there will be no information avalanche for the individual scientist at the microlevel of information provided his tasks and objectives are sufficiently and clearly differentiated and defined. We already arrived at analogous conclusions when investigating the communication form 'journal article' in the papers.[3,4]

The higher valency of the communication form 'report' was derived from the higher profile allocation density and from the higher percentage of utilization, as well as earlier[2] at the macrolevel from the displacement to the left of R_{macr} as opposed to Z_{macr}. It would surely be wrong, however, to suppose a priori that for providing this proof use could be made of the displacement to the left of the curve R_{micr} relative to Z_{micr}, or even for every single profile, of R_{micr} relative to Z_{micr}. Figures 5 and 6 will be useful in illustrating this point.

While in Fig. 6 the mean curves $\bar{Z}_{micr}$ and R_{micr} practically coincide so that there is no displacement to the left, the presentation in Fig. 5 of 5 pairs of curves Z_{micr}, R_{micr} (A, a; B, b; C, c; E, e; F, f) does not show any regularities at all with respect to the relative position of the curves: neither do the curves for reports always lie to the left of the curves for journal articles, nor vice versa, but, rather, very different combinations are realized. There is no clear picture as compared to the one at the macrolevel (cf. Fig. 6: R_{macr} quite significantly displaced to the left with respect to Z_{macr}!). This, then, is an illustrative proof of the fact that relations and conditions at the microlevel are not at all a mere 'reduction in size', a 'copy' or, 'image' of relations and conditions at the macrolevel, but that at the transition from one level into the other there take place processes that are not yet accessible to our understanding.

There is, however, no doubt that the two levels are components of one and the same system of scientific communication – we actually use one and the same data base generated by means of the coincidence methodology for all our investigations. It is equally certain that the meachanisms connecting both levels with each other can only be found when the processes at the microlevel have been investigated and understood in detail. The special demand will have to be formulated that all attempts at describing rank distributions at the macrolevel will have to be made in such a way that they will also be helpful in explaining the rank distributions at the microlevel.*

Concluding remarks

The coincidence method and the two-level concept have been tested in a further example of application – scientific reports in an international information system. Confirmation was supplied for the high valency of the communication form 'report' and for the fact that it will be possible for the individual scientist

*Analogy: In 1937, Soviet physicist P. L. *Kapitsa* discovered the phenomenon of the suprafluidity of liquid helium. This could not be explained by classical concepts. *Kapitsa* suggested the microstructure of matter to be responsible for this phenomenon. Later on, Soviet physicist *L. D. Landau* supplied the microscopic description of this phenomenon within the framework of the quantum theory.

even under the conditions created by the 'information avalanche' still to realize the fundamental principle of scientific work.

It remains an open question by what mechanism the transition of microdistributions into the macrodistribution is effected.

The application potential of the coincidence method and the usability of the two-level concept have not yet been exhausted by far. Their utilization for the investigation of the system of scientific communication, for the assessment of different forms of communication, for scientific information work, for science-of-science investigations, etc. has only just begun. What is necessary now is to define more closely the preconditions and the limitations of the coincidence method, but also its automation.

Concerning the two-level concept, more precise statements on the relation of information center to the individual scientist, on the relation of information science (informatics) to the science of science, and also on the relation of informetrics to scientometrics will eventually become feasible.

We expect that in the near future the coincidence method will provide researchers in the field of informatics and the science of science with an effective and easy-to-handle tool, and they will be supplied by the two-level concept with an efficient frame of thought.

References

1. M. BONITZ, Eine Methode zur Optimierung des Zeitschriftenbestandes in großen Forschungsbibliotheken, in: *Schriftenreihe Dokumentation/Information des Instituts für Informationswissenschaft, Erfindungswesen und Recht*, Heft 31, Ilmenau, 1975, p. 25–40.
2. M. BONITZ, P. SCHMIDT, The Growing Importance of R&D Reports as Shown by Nuclear Science Reports, *International Forum on Information and Documentation*, 3 (1978) No. 1, 8–12; *Journal of Radioanalytical Chemistry*, 57 (1980) No. 1, 7–16.
3. M. BONITZ, Evidence for the Invalidity of the Bradford Law for the Single Scientist, *Scientometrics*, 2 (1980) No. 3, 203–214.
4. M. BONITZ, P. SCHMIDT, Übergang von der Makroinformationsebene zur Mikroinformationsebene bei Rangverteilungsuntersuchungen an einem internationalen Informationssystem, *Informatik*, 27 (1980) No. 5, 18–22.
5. S. C. BRADFORD, *Documentation*, Crosby Lockwood, London, 1948.
6. *INIS Circular Letter*, No. 74, IAEA, Vienna, 1978.
7. *INIS Circular Letter*, No. 89, IAEA, Vienna, 1980.