Authors as "Artists" or "Heavy Weights" in Scientific Publishing
The Sociological Analysis of Scientific Literature and the BMS

Karl M. van Meter
Authors as "Artists" or "Heavy Weights" in Scientific Publishing
The Sociological Analysis of Scientific Literature and the BMS

Karl M. van Meter

Introduction

1 In 1992, Current Sociology organized and published a thematic issue on sociological AIDS research (Pollac, 1992). For that issue, William Turner and I proposed to analyze and produce a "cognitive map" of sociological AIDS literature based on articles abstracted in Sociological Abstracts. Using a hierarchically ascending method of classification based on co-occurrence of key words (Lexinet and Leximappe), in this case in the abstracts, we published the special annex of that issue of Current Sociology (van Meter and Turner, 1992) which included a "strategic diagram" based on the analysis of the 290 last research articles (all of 1989 and part of 1990) on AIDS in the Sociological Abstracts data base at that time.

2 We followed up that initial publication three year later with a more detailed and extensive analysis of all the AIDS articles abstracted in Sociological Abstracts between 1980 and 1990 (van Meter, Turner and Bizard, 1995). That analysis produced two other strategic diagrams covering the periods 1980-1986 and 1987-1988 which correspond, respectively, with the "birth" of sociological AIDS research literature (low volume and slow initial growth) and with its exponential "growth" (fast growth to "plateau" volume). The initial 1989-1990 period corresponded with "maturity" (high volume but little further growth). Together, these three periods and their corresponding maps or strategic diagrams describe the well-known "S-curve" of evolution in scientific publishing.

3 But that article, with its sections on the "Sociological Abstracts Data Base", "Cognitive Mapping", "Early Period", "Middle Period", and "Final Period", examined the thematic or key-word evolution of sociological AIDS research during that time but did not get into the different roles played by the journals publishing the articles, or into the authors who
wrote the articles. Those topics were treated two years later in another article (van Meter and Turner, 1997).

**The Co-Occurrence o Key Words Method**

In the three research articles mentioned above, we used the Lexinet-Leximappe program which is a hierarchically ascending method of classification based on co-occurrence of key words in the units of text under analysis. In this research, the units were the abstracts published by Sociological Abstracts and covering the scientific sociological AIDS research literature. The more two key words co-occur together in these units (abstracts), the "closer together" they are placed by the classification algorithm, and the closer they appear on the resulting cognitive map of the data set. The PC-based Lexinet-Leximappe program -- and its Apple predecessor, Candide (Teil, 1990) -- was developed in France in the framework of cognitive mapping and the sampling of information flow at the Paris Ecole des Mines (Chartron et al., 1990). The program does not use pre-established dictionaries but generates its own thesaurus of key words in a text, and also an anti-thesaurus of words not retained as key words or replaced by a closely-associated key word.

The program is modular in structure and performs the following operations: segmentation of the text which identifies each word and its context in the corpus; normalization of terms by examining those with similar suffixes; distributional filtering which eliminates the common everyday terms; managing synonyms through the use of a thesaurus and the intervention of a human expert; identification and management of composite terms through a statistical study of their distribution in the corpus; management of the monoterms; and indexing of the retained key words in the documents (in our case, the abstracts). This set of key words constitutes the Lexinet thesaurus or lexicon which is forwarded, along with the indexed documents, to the Leximappe program.

Leximappe calculates the association between all pairs of the 1,500 most frequently occurring lexicon words, using a statistical index for the measurement of co-occurrence of key words (Michelet, 1988). To manage and analyse the general network of all co-occurrences, the network is cut into coherent, manageable sub-networks called "components" which correspond with strongly-associated sub-sets of not more than ten lexicon words. The actual process uses hierarchically ascending classification analysis (a form of "cluster analysis") which -- beginning with the set of all pairs of key words as initial classes -- constructs successively more general classifications, which means with successively fewer classes. When a "component" or class of ten closely-associated key words is obtained, it is withdrawn from the classification, and the procedure continues until all pairs of key words have been associated with a class.

Once these components have been built by Leximappe, they are plotted onto a two-dimensional diagram called the "strategic diagram" mentioned above. Its horizontal axis indicates "centrality" and the vertical axis "density" (Vinc, 1991). Centrality is a proportional measure of how often key words in a class or component are linked (co-occur) with key words outside the class. It's a measure of external tie intensity. Density is a measure of the strength of links within the class or the strength of the mean co-occurrence between members of the class.
Therefore, these two axes divide the strategic diagram into four quadrants: quadrant one (upper right-hand) which is dense and central; quadrant two (upper left-hand) which is dense but not central; quadrant three (lower left-hand) which is neither dense, nor central; and quadrant four (lower right-hand) which is central but not dense. By construction, this means that quadrant one contains components whose key words have strong ties between themselves and strong ties with key words outside the component. Quadrant two contains components whose key words have strong ties between themselves but weak ties with key words outside the component. Quadrant three contains components whose key words have weak ties among themselves and with key words outside the component. Quadrant four contains components whose key words have weak ties among themselves but strong ties with key words outside the component.

**Scientific Innovation and Evolution Over Time**

With the above explanation, it is evident why scientometric research tends to refer to quadrant one as "mainstream science", quadrant two as "ivory tower science", quadrant three as either "up-and-coming science" or "chaos", and quadrant four as "bandwagon science" (Vinck, 1991). In scientometry, there are several competing explanations of scientific innovation and the associated circulation of components on the strategic diagram. One explanation maintains that innovation and the circulation of components follows a counter-clockwise motion from bandwagon science to mainstream science, then to ivory-tower science and chaos, from which a renovated form reappears as bandwagon science. Another explanation maintains that mainstream, ivory-tower and bandwagon science all emerge from up-and-coming science. Still another explanation, which certain of our results support, maintains that bandwagon science generates the components that move progressively toward mainstream science, ivory-tower science or chaos. Only further research will be able to clarify this question.

In the specific case of our sociological AIDS research, using the three periods 1980-1986, 1987-1988 and 1989-1999, we were able to follow over time the evolution of the components and their associated key words, thus describing a "thematic" evolution of this field. We were even able to follow the scientific trajectory of individual authors and the evolution of the network of co-authorship in our data base. It was also possible to compare thematic evolution and author trajectories with publishing practices of the journals figuring in our data base, thus explaining the terms of "Representation and Confrontation of Three Types of Longitudinal Network Data" in the title of our 1997 article.

In the 1997 article, we mentioned that "the evolution of research themes, authors and journals in this domain are quite distinct from each other. Indeed, they each form a separate set of longitudinal network data generated from the same data base. The research themes seem to divide up into solid, less flexible themes which change little with time, and those that are more flexible, even 'adaptive' and change over time. The set of authors also seems to have this aspect of division between those that move in a clearly designated direction and those that change over time" (van Meter and Turner, 1997: 32). It is particularly this latter aspect concerning authors that interests us here.

The distribution of research articles is the following:
The year 1990 was truncated in the version of the Sociological Abstracts CD-ROM which we used in 1991 when this research began. For the period 1980-1986, the strategic diagram identified three clearly distinct clusters with no ties between them: Homosexuality, AIDS and Sociology. For the 1987-1988 period, the strategic diagram identified eight clusters of which six were interconnected and important: Homosexuality, U.S.A., Prevention, Drug Addiction, College Students and Children. For the final 1989-1990 period, there were five clusters with Hispanic Americans and Behavior Modification in the ivory-tower second quadrant, Public Health and U.S.A. in the bandwagon fourth quadrant, and Sexual Behavior in the mainstream first quadrant where Sexual Behavior is a continuation and an expansion of the preceding period's Homosexuality cluster.

**Authors: Heavy Weights vs. Artists & Hedgehogs vs. Foxes**

In most scientometric research, an enormous majority of authors appear only once in a data set, either alone or as a co-author. That is true in our case where only two authors (S. R. Friedman and R. Stall) wrote two or more articles in each of the three periods and only 16 authors wrote one or more articles in each of the three periods. In addition to the two above authors, there are five others (J. W. Curran, D. M. Fox, G. S. Getzel, J. M. Mann, D. G. Ostrow) who published two or more articles in two periods. To these seven authors, one must add 63 other authors to obtain the total of 70 authors who published one or more articles in two or three periods. There are also 39 authors who have published two or more articles but only in a single period (van Meter and Turner, 1997: 40).

Even if it is difficult (in terms of statistical significance) to speak of ties or association between authors, we found an important association between authors and thematic key words. We adopted as a minimal criteria that an author is "associated" with a thematic cluster in any of the three periods when at least two key works of an author's article are members of the thematic cluster. Using this definition, we were able to represent author-theme association directly on the key-word strategic diagram for each of the three periods (see the three strategic diagrams in the annex).

Given this evolution of authors in relation to the thematic key words, which are also in a state of dynamic evolution, it seems that there are two types of authors whose association with the thematic clusters are important: those, such as Ostrow, Ross and Stall, associated with a stable theme which evolves with the authors; those, such as Watters, Des Jarlais, Bayer and Friedman, who appear often associated with key-word clusters but whose thematic content changes relatively quickly. Therefore, in 1997, we made the conjecture that if there is serious competition in a scientific field for resources (both financial and
institutional), such as was the case for sociological AIDS research (Grmek, 1989), then there are only two types of scientists who survive and make an important contribution to the field: those who evolve with a theme which is stable over time; those who move from association with one major theme to another.

Almost by definition, the thematic cluster in the first case will be near or in the mainstream first quadrant, while the clusters in the second case will be in either the ivory-tower second quadrant or the bandwagon fourth quadrant. This may also mean that, in the first case, the authors (and/or themes) manage to obtain the necessary resources to continue on their apparently linear evolution. In the second case, either resources are made available for research on new themes and these resources are obtained by well-known scientists or these scientists find new themes and encourage the allocation of resources to these new themes. At any rate, and despite a small number of pertinent cases, we have not found an author who moves from one form of scientific publishing activity to the other (van Meter and Turner, 19977: 41).

Following the initial publication of these results, we adopted the term of "heavy weights" to describe authors who were associated with stable, major themes over time, and the term of "artists" to describe authors who were associated with important but varying themes over time. Indeed, we have used this terminology and developed these results further in different conferences ("Authors as 'Artists' or 'Heavy Weights' in Scientific Publishing", International Federation of Science Editors (IFSE) Twelfth Conference, 10-14 October 2004, Merida, Yucatan, Mexico; "From Lenin to Hedgehog Heavyweights and Foxy Artists: Social Network Analysis' Political Past and Scientific Future", EINIRAS Conference, 14-16 September, Zurich, Switzerland).

In the meantime, colleagues in scientometrics noted the similarity between our general description of "heavy weights" and "artists" and the often-used metaphor of "foxes and hedgehogs" which seems to go back to the seventh-century-BC Greek poet, Archilochus of Paros, who reported stated: "Foxes know many tricks, hedgehogs only one, but it's a big thing." The metaphor has since evolved into the generalization that "foxes" are "generalists" and "hedgehogs" are "specialists" and produced unsubstantiated claims such as: "Most of the great scientific discoveries are made by hedgehogs, most of the little discoveries by foxes, but science needs both for healthy growth." The claims even generalizes to: "there are two types of people in the world, foxes and hedgehogs". The metaphor is so widely used that a recent Google search for "foxes and hedgehogs" produced -- in the first ten pages -- 815 references, of which almost none could be considered scientific. In short, the formalization above of scientific authors as "heavy weights" or "artists" seems to have struck a very popular vein which, until now, has only functioned as a metaphor, but a widely used one indeed.

Methods for Analyzing the BMS & Publishing "Elites"

It's interesting to note that a formal method of sociological analysis -- co-word analysis of scientific publications -- can generate results that confront and impinge on popular conceptions of a social phenomenon; in this case, scientific innovation. At the Bulletin de Méthodologie Sociologique (BMS), we have been aware of this and decided to use a further development of co-word analysis -- Calliope, an advanced Windows version of Lexinet-Leximappe by Mathilde de Saint Léger -- and a factorial correspondence analysis --
These two methods were used to analyse the entire set of abstracts of articles published in the BMS between 1993 and 2003. The results of these two methods, and their comparison, clearly showed the close association of the BMS with the Research Committee 33 (RC33), "Logic and Methodology", of the International Sociological Association, and the extensive coverage of both language-based data or "qualitative" methods and formalized data or "qualitative" methods. This key-word analysis was largely thematically oriented and, by sheer lack of volume, could not delve into the question of "heavy weight" or "artist" authors: for approximately 200 authors, only 14 wrote two articles, and only 8 wrote more than two articles.

The authors of several articles are often referred to as the publishing "elite" of a scientific domain and sociology of science has a tendency to concentrate its attention on them. But publishing is only one of several important functions of leading scientists as Rafaël Stofer has shown in his research on the French scientific "elite" in cancerology (Stofer, 2001; Stofer, 2002). As the title of his 2002 article indicates, the use of both structural (social network) methods and scientometric methods were used. They showed that there are "elites" in publishing scientific articles, there are "elites" in obtaining financing for research projects, there are "elites" in recruiting the best doctoral candidates, there are "elites" in obtaining and running the newest and most powerful analytical machinery, and these "elites" are each different and negotiate between each other to obtain the resources necessary for their scientific production, of which publishing in only one aspect, although admittedly a major one.

But even in publishing, there are different "elites". The founding editor of Sociological Abstracts, Leo P. Chall, once used his privileged position to do a questionnaire survey of 123 major English-speaking sociologists, including many editors of sociological journals. Chall asked them to rank 90 leading journals abstracted by Sociological Abstracts according to the importance of their scientific contribution as seen by the respondent (Chall, 1988). The analysis of these data showed the structuring of this domain and an important difference between annuals and other journals. Also, the non-American English-speaking "elite" sub-population behaved in a manner different from the American "elite" sub-population. The former placed more internationally-oriented journals, such as International Sociology (published in association with the ISA and UNESCO) near the top of the list while American respondents systematically placed the American Journal of Sociology and the American Sociological Review at the top.

Chall’s survey method explicitly targeted an "elite" and found out that it was not unified. More recently, a far more "democratic" survey method was applied by Philippe Jeannin (2003). His method consisted of crossing several authoritative data bases of periodicals (from libraries and official research documentation services), and then constructing a list of reviews which was widely submitted to the scientific community for comment. The "scientific community" consisted of all those teaching sociology at the university level or doing sociological research. A journal was then considered scientific when so judged by the members of the scientific community which then classed journals in one of three categories: A, central non-specialized journals; B, central but specialized journals; C, non-central journals. Each member of the scientific community could vote on each journal and positively or negatively for each category A, B and C. The tabulation of the votes was then posted on the Web for general public viewing at http://www.iut-tarbes.fr/enquete and
This scientific-community-based method of analysis and evaluation has met with a certain success and was of particular interest to participants at the IFSE'12 conference. Indeed, several participants asked the author if the system could be implemented in their non-English-speaking country to provide a system of evaluation other than the American-dominated citation indexing system. Of course, the answer is "yes"; the system can be implemented in any country which has an organized scientific community in the field to be analyzed, even if the scientific literature read by that community is largely from other countries.

**Analyzing World & "National" Sociologies**

Jeannin’s research is one of the first attempts we have encountered in trying to analyze a "national" sociology, as well as other "national" disciplines. It is widely accepted that sociology as a science is very closely tied to the culture and language in which the research is done. Indeed, there even exists an "Association Internationale de Sociologues de Langue Française". But does this mean one can find in the structure and content of world sociology specific differences associated with "national" sociologies? We addressed this question at the IFSE’12 conference and participants were not surprised by our scientometric results concerning sociology which show that there doesn’t seem to be an overall structure to world sociology with well-defined "national" sociologies (van Meter, 2004: 243): "One clear conclusion is that contemporary sociology has little overall structure ... Only at a detailed level does structure and thematic coherence begin to appear." Using the Lexinet-Leximappe program, we found thematic components associated with "U.S.A.", "Canada", "Federal Republic of Germany", "Eastern Europe", and "Italy", but they were all low-density, unstable clusters (van Meter, 2004: 242).

The problems encountered with this "top down" approach to analyzing "national" sociologies has encouraged us to pursue a "bottom up" approach which resulted in a special thematic issue of the *BMS* (n. 85, January 2005). There, not one, but four different methods were applied to the same data base consisting of all abstracts of the more than 1,000 presentations given at the first congress of the newly-constituted Association Française de Sociologie (AFS) in February 2004. The four methods were Calliope, Trideux, Alceste and Prospéro. At an overall level, the key word "Femme" (woman) dominates in terms of frequency but has little effect on the structuring of the field which appears much more closely associated with the specialized sections ("Réseaux thématiques") which constitute the AFS.

The results of that special thematic issue of the *BMS* have been developed further and are being published as a book (Brossaud *et al.*, 2006) which will be available for the forthcoming second congress of the AFS on 5-8 September 2006 in Bordeaux. In the chapter by Mathilde de Saint Léger and me (de Saint Léger and van Meter, 2006), we use the co-occurrence of key words method, and in particular the program Calliope, to describe in detail the very visible role of the key word "Femme" and the structuring role of the "réseaux thématiques".
The two of us are continuing work with Calliope on the "bottom up" approach to "national" sociologies by attempting to analyze the 418 abstracts of presentations made at the 2004 congress of the Deutsche Gesellschaft für Soziologie (DGS) from which we expect results by the end of this year. At the same time, we are attempting to do the same type of analysis with the abstracts of the 2004 congress of the American Sociological Association. Once we have obtained results for all three with the use of Calliope, we hope to be able to do a concrete comparative analysis of the structure and content of French, German and American sociology.

A very similar "bottom up" project of international comparative research based on the analysis of documents has also been published by the BMS. However, the subject matter was not sociological research literature but general public "political" literature. Valdis Krebs collected data from Web-based book retailers, such as Amazon, which, for a given book, list other books that were also purchased by customers along with the first given book. Using this "buddy book" purchase data as social network tie information, Krebs used his own InFlow network analysis program to show that the American buddy book network consists of two clearly divided groups: a larger and more diffuse left-of-center readership, and a smaller and more closely tied right-of-center readership (Krebs, 2003). In the former, two well-known books -- Michael Moore's *Stupid White Men* and Norm Chomsky's *9/11* -- concentrated the most ties (see the figure in the annex). Between the two clearly-divided groups, there is only one book (at this level of analysis) with ties to both groups: *What Went Wrong*.

This method offers the fascinating possibility of describing and eventually being able to measure a society's partisanship or political "divisiveness". That, in turns, would permit us to follow these developments over time and analyze political developments in terms of their influence on "divisiveness" and vice versa. Moreover, with Amazon branches in France, Germany and elsewhere, we hope to be able to do a comparative contemporaneous analysis of political "divisiveness" in the United States (and perhaps Canada), France and Germany.

**Future Research**

The large variety of methods presented here is only a current portrait of some of the better-known methods available for analyzing scientific and even general public literature. New methods and new applications are being developed all the time, and a constant effort is necessary to try to remain receptive to these new developments. We hope that both future IFSE conferences and the *BMS* will contribute to this scientific effort.
BIBLIOGRAPHY


Karl M. van Meter (2004), "Conclusions and Future Research: A Structure for World Sociology", in Karl M. van Meter (editor), Interrelation between Type of Analysis and Type of Interpretation, Peter Lang, Bern, pp. 227-246.


APPENDIXES

Three strategic diagrams of sociological AIDS research with authors and key words (three pages).

Diagramme stratégique du SIDA 1980-1986
Diagramme stratégique du SIDA 1987-1988
Diagramme stratégique du SIDA 1989-1990

Kreb's diagram of buddy books

ABSTRACTS

There are only two types of major authors who survive in such a new and competitive field of scientific development as early sociological AIDS research. The "artists" survive by publishing in successive, more-or-less neighboring specialties, probably associated with newly available financial or institutional resources. They also appear to change co-authors comparatively often. On the other hand, "heavy weights" dominate a specific specialty closely associated with mainstream scientific developments which are therefore assured of financial and institutional backing. They appear to change co-authors relatively less frequently. These results, based on co-word analysis can be compared with the results of divers other methods of sociological analysis of scientific literature which tend to show that there are two types of world sociology, that there
are multiple, not unique, scientific elites, and that there are different national sociologies which can be compared for content and structure.

Les auteurs comme "artistes" ou "poids lourds" dans la littérature scientifique, l'analyse sociologique de littérature scientifique et le BMS: Il n'y a que deux types importants d'auteurs qui survivent dans un domaine nouveau et compétitif comme les premières recherches sociologiques sur le SIDA. Les "artistes" survivent en publiant successivement dans des sous-spécialités plus ou moins proches, et probablement en association avec de nouvelles ressources financières ou institutionnelles. Ils semblent changer relativement souvent de co-auteurs. En revanche, les "poids lourds" dominent une spécialité étroitement associée à la "grande science" et donc assurée de financement et de soutien institutionnels. Ces auteurs semblent changer relativement moins de co-auteurs. Ces résultats, basés sur une analyse de mots associés, peuvent être comparés avec ceux de diverses autres méthodes d'analyse sociologique de littérature scientifique qui tendent à montrer qu'il existe deux types de sociologies mondiales, qu'il y a de multiples élites scientifiques, et que les différentes sociologies nationales peuvent être comparées par rapport à leur contenu et leur structure.

INDEX

Mots-clés: Analyse par mots associés, Carrières scientifiques, Comparaison de sociologies nationales, Littérature scientifique, Publications scientifiques

Keywords: Co-Word Analysis, Comparing National Sociologies, Scientific Careers, Scientific Literature, Scientific Publishing

AUTHOR

KARL M. VAN METER

CMH-CNRS-EHESS-ENS, Paris, karl.vanmeter@ens.fr