Interrelations and dynamics in thematic networks
How to present bibliometric outcome?

Abstract

Network analyses of term and concept co-occurrences are examined to demonstrate their potential in combining both in one map. Especially in time considerations it is helpful to apply general concepts as fixing points. These concepts can be taken from data bases directly (with possible indexing bias) or be generated ex post by combining certain text terms or descriptors to concepts. Alternative possibilities are discussed. Examples are given from studies on German literature about nutrition and elderly employees.

1: Introduction

Analyses of bibliographic databases often have the objective to make statements about the developments of a field of knowledge. For example, actors, like the most frequent authors or institutions, and their connections are used to interpret scientific networks and their development. The advantage here is that authors are more or less clear and easy to be systematically extracted. Concerning the substantive component, which should illuminate the cognitive structure of an area and the scientific context (Noyons & van Raan 2002), this has to be dealt with differently. If the term-level (e.g. title words) is used, then quite diffuse contents appear in larger text collections and it is difficult to recognize a general pattern. If general concepts are examined, then usually these are too rough classifications which reveal no specific trends as concepts are by nature constituted through abstraction. This problem shall be illuminated more nearly and solutions shown by a network analytical examination of descriptors and classifications of literature from sociological fields: the areas of nutrition sociology and elderly employees.

2: The meaning of concept

According to Satija (2004), following Ingetraut Dahlberg (1974), “A unit of thought [...] is] constituted through abstraction on the basis of properties common to a set of objects; a mental image formed by generalization. It is generally regarded as the product of a ‘system of orderly thinking’”. It means that a concept is not identical with a certain word, but it is a more general idea that lies behind it, including other words and their relations to each other. Concepts can be described by a sum of adequate characteristics but these are not invariant in time and space either; and they must be adapted to their cultural contexts. In content analysis the common procedure to define concepts is the usage of several terms that are variants or partial aspects of the examined concept (see the so-called content analysis dictionaries, e.g. Mohler & Schönbach 1979).

3: Themes of publications concerning nutrition research

The most frequent title words could give an indication of current research trends and special technical terms. As expected in an example of nutrition sociological literature 1985-1996 (retrieved from three different German databases) the language variants of nutrition and eating happen to be the most frequent ones (even after a morphological reduction). On the other hand, the words child, Germany, woman, bulimia already show...
typical application areas. Further titlewords like result, examination say something about the research culture.

To what extent more general concepts in the area of ‘nutrition and society’ in the literature occur, was shaped up by a formal content analysis. For this purpose terms in the original title (of the authors) and the descriptors (of the database manufacturers) as well as their categorization were used (see Tab. 2). The advantage here is that a uniform standard was applied to the literature of the different databases, which have different description rules. Thus it could be coded newly, independent from the original time and indexing. On the other hand, then different possible time-bound meanings contained in the terms could not be taken into account. The idea is that a sufficient number of terms considered in any case meets the sought-after concept; in a more recent study only one consistent database was used and thus the existing classification scheme itself could be used directly (Lange et al. 2008).

3.1: Concept dynamics

A diagram of the publication years can give an indication now, to what extent there are ups and downs of certain topics within a given period. In the retrieved universe of publications the concept PATHOLOGY is quite equally distributed over the years (in any given year there is an emergence of about 3/4 or less than 1/4 of the average rate of publication), whereas the topic HOUSEHOLD is in the lower or upper frequency range within 8 publication years, i.e. it is subject to very strong fluctuations. If the under-represented year 1996 is ignored, for several issues, such as YOUTH, CONSUMPTION, or LIFESTYLE, it is remarkable that in later years they all show somewhat higher rates of publication.

3.2: The associative relationship between concepts

Since concepts never are narrowly defined and static, and get or vary their meaning only by their reference to other concepts, it was examined to what extent they appear together in the same publications. Instead of the frequencies and contingencies the connections are represented here in a simplified correspondence graphic. The main topics are represented in their two main dimensions, i.e. common aspects (horizontal axis: the contrast between HOUSEHOLD and METHOD; vertical axis: partly the difference between METHOD and SENIOR CITIZENS).

If topics in the correspondence map are close to each other, then, as a rule of thumb, they appear in the publications also together and if they are far from each other, they rarely appear together in the publications. With the help of the axes, moreover, we can classify their content. Thus SOCIOLOGY and PREVENTION are relatively well in the middle; so they are not specific with regard to the two main axes. HOUSEHOLD, METHOD, INFANTS and SENIOR CITIZENS on the other hand are really concise topics. PATHOLOGY, PSYCHOLOGY and SEXES as well as LIFESTYLE, WORK, CULTURE and SOCIOLOGY are near at each other, i.e. they frequently appear together, and therefore are not very selective under each other. The axes seem to contain three tendencies:

- negative horizontal axis: Economic and cultural science area
- positive horizontal axis: Psycholo-methodological area
- positive vertical axis: Medico-educational area
A multivariate analysis that locates authors, publishing houses, journals and data bases in this topic space is reported in Bayer et al. (1999) but this will not be our focus here.

The grouping of topics after its correlative or associative proximity (e.g. with cluster or multidimensional scaling) is a common method of topical interpretation. However, this is a quite rough view and projected on the basis of similarities by several characteristics, which not necessarily co-occur in the particular literature. Another critical point is the labelling of the clusters built according to their correlation, which is often identical with the most frequent words occurring in it, whereas only some may be suitable as paraphrase (cf. Callon et al. 1991). We will therefore apply a network analysis which connects the original context of descriptors with their broader concepts (Herfurth et al. 2003).

4: Themes of publications concerning elderly employees

The aim of a literature search on elderly employees over four different German databases was to get as well some clues to research trends. As an empirical basis the descriptor indexing of the literature by the database producers was taken. There was a total of 2564 different types in the databases; therefore only the most common descriptors could be taken. As expected, language variants of older employed person like older labourer and elderly employee were frequent descriptors. The words early retirement, labour market, flexible age limit and, on the other hand, USA already show typical application areas. Development, structure, trends give information on the research culture. Like in the previous discussed study the general subject areas were found by a formal content analysis and should as well represent consciously rough, fuzzy concepts, to express all kinds of descriptors by a more general semantic. The selection of concepts was based on the main issues as well as the most common descriptors. WORK and OLD PEOPLE proved to be the most frequent concepts in accordance with the retrieval query. On the other hand, the subject areas PUBLIC SERVICE and TECHNOLOGY have a rather low frequency. The application areas (like SENIOR CITIZENS or WORK) show a much lower frequency than the fields of science (like CULTURE and SOCIOLOGY). The application area PREVENTION (296 or 43% of the publications) and the science field DEMOGRAPHY (47 or 7% of the publications) are exceptions.

4.1: Connections of the subject areas concerning elderly employees

The relations between these topics can be represented in a concept net under consideration of all relations with each other (see Fig. 1). By this the number of links through the indirect semantic proximities of the concepts is shown, as well as the bivariate strengths of relations (represented here by the Jaccard coefficient). Chains like new Bundeslænder - EAST - FRG are connected with each other only by one single path and cliques like ageing workforce - INTERNATIONAL - USA are directly and indirectly connected to each other.

Absolutely and regarding the (Jaccard-)relations in the covered time period 1985-2000 the main concepts (in MAJOR notation) had to be dominant, at least as each concept was formed by more than one descriptor and as these descriptors were used
differently in the four databases. Two central concept components stand out: One around the larger subject area WORK and a smaller one around the subject area GENDER being in loose connection with each other by demography. WORK is in connection to EARLY RETIREMENT PENSION, RETIREMENT, OLD PEOPLE (resp. elderly employee and old person) and STATE, which are less connected between each other. The subject area STATE is essentially bound to the descriptor labour market. Questions of HEALTH are only indirectly connected by OLD PEOPLE with the thematic complex WORK. Research that is stronger oriented theoretical (SCIENCE) rather refers to old persons generally. GENDER is picked out as topic primarily in connection with the demographic distributions. In the further context also the economic sector and EDUCATION are addressed in this concept cluster. The other concept components (EAST, WORKING HOURS, LAW, BUSINESS, INTERNATIONAL) are obviously no specific questions for WORK or GENDER. Topics which are not in a closer connection with broader concepts are listed in the lower part of the figure.

4.2: Temporal development of topic relations concerning elderly employees
A diagram of the publication years for certain topics could already give an indication to what extent there are peaks and dales of research activities (e.g. by the burst model, see Klineberg 2002). But by a differentiated inspection of the concept frequencies together with their relations to other concepts or to descriptor terms in the various periods can reveal the interactive developments. While relations of words are rarely stable in the course of time, by combination with generic concepts we get the possibility to compare these concepts with their constituent special terms also over different temporal contexts (cf. Leydesdorff 1997 with reference to Hesse 1980). It seems that the methodological and theoretical treatment (SCIENCE) of the topics WORK and OLD PEOPLE increasingly move into the foreground in later years. While part-time work at the WORKING HOURS models is always present, only initially flexibility and reducing of working hours and finally partial retirement are included in the papers. In the entire considered period EDUCATION is only in indirect connection with WORK and OLD PEOPLE: at first with demographic categories like GENDER and age structure but then with continuing education and BUSINESS. BUSINESS appears even only as of 1997 in contact with OLD PEOPLE. LEGAL aspects get more concrete only in the time period as of 1989: first with regard to employment promotion act, then with regard to a statutory pension insurance. EASTern issues (first FRG, then new federal states) are non-specific first and as of 1993 related to specific questions like early retirement or sex ratio. Issues of HEALTH are in the complete observation period in connection with WORK and OLD PEOPLE and are more frequent out only in the last time period together with professional efficiency. SOCIAL SITUATION is in reference to WORK or unemployment apart from 1989-1992 and as of 1989 to old person. Consistently, STATE protection (in relation to SOCIAL SITUATION) is picked out as central theme. 1989-1992 HEALTH aspects, 1993-1996 EAST-West aspects, and 1997-2000 GENDER aspects come into consideration. PRIVACY and EUROPE remain non-specific in all time periods, i.e. they are not connected with special topics of the area elderly employees.

Retirement and early retirement pension are the dominating topics in the years 1985-2000 – at an increasing consideration of “society” (STATE) as a parameter. In contrast,
regarding older workers, much less is published on the social situation, education, health, business issues, and gender roles (and this is usually detached from business work). Considering the temporal evolution, it is striking that the social situation of women, vocational training and legal protections of retirement gain weight at the end of the nineties at least. Not much in appearance come publications on European issues, private households, the new Bundeslaender (or regional approaches), or foreigners. For accompanying measures to increase retiring age these are topics of increasing importance, though. Also international comparative knowledge production supported by scientific methodology and self reflection need to be upgraded.

5: Summary
Bibliometric studies have the problem that thematic questions often start with single terms and might lead then to quite random, chaotic results. A common method is the multivariate reduction (cluster analysis, correspondence analysis), where single concepts are weighted and located in accordance with their overall relationships. Their disadvantage is that the sum of the individual terms often does not reveal the theoretical concept lying behind them and their spatial constellation often does not correspond with their empirical co-occurrence. On the other hand, if only theoretical constructs are examined, which are projected onto the empirical text ex ante or ex post, then rough concepts arise whose empirical meaning gets not very transparent. Therefore in an example concerning elderly employees a thematic network analysis was demonstrated, thus shaping both the context-specific words, as well as broader concepts. By this the concepts get concrete interpretation and vice versa, the special descriptors are arranged around concepts more aptly. Another advantage is the temporal consistency of the mapping of the concepts, while the keywords can vary considerably dependent on time.

References
Mohler P.P., Schönbach K., 1979, Deutsche Diktionäre für computerunterstützte Inhaltsanalyse, ZUMA working report.


---

**Fig. 1:** Topic relations as a semantic net 1985-2000 (example: elderly employees; 3195 documents; minimum frequency of the concepts: 75 = frequency of the 80th concept; minimum frequency of the relations: 21; the line width corresponds to a Jaccard coefficient between 0.2 and 0.7)