

A textual analysis and scientometric mapping of the dynamic of knowledge in and around the IFSA community.

Marc Barbier INRA, Paris-Grignon, France

Véronique Garandel - Batifol INRA, Toulouse, France

Marianne Bompert INRA, Toulouse, France

Abstract

This communication proposes to depict the dynamic of the conceptual area of IFSA, and to bring a support to a discussion about the future of the IFSA community. It presents the results of a textual and scientometric analysis of the descriptors and concepts of the IFSA community. A general view of the knowledge of the community is delivered on the basis of the data mining of the corpus built up with the communications of the 6 previous IFSA Symposia. The empirical study details the key concepts and domains of investigation and their dynamics throughout the previous Symposia. This work has been realized in relation with a group of experts of the IFSA scientific domain.

In order to contextualize the scientific positioning of the IFSA community, this study also delivers a set of scientometric mappings in order to analyze the corpus that have been extracted from international databases (CAB Abstract, Web of Science) based on the identification of the main IFSA conceptual keywords with IFSA experts. In conclusion the dynamic of the concepts and the positioning of the IFSA community is discussed.

Keywords:

Farming System; Knowledge Dynamic; Information Retrieval; Data Mining; Textual Analysis; Co-word Analysis;

1. Introduction

The objective of this communication is to present to the members of the IFSA community and to the participants to this symposium, the result of our exploration of the dynamic of knowledge in and around this community. The purpose is thus to deliver an analysis of the conceptual domain of IFSA in order to empower the future positioning of this community. The methodology of this analysis has been defined, built-up and adjusted thanks to regular exchanges with some members of the Board of IFSA. The results that are delivered in this communication represent only a small part of all the work done, which will be delivered to the board of IFSA in a kind of a report of activities and results. We shall thank especially B.Dedieu, B.Hubert and M.Cerf for their help in connecting us to this community.

The main results exposed here have been produced thanks to the construction of a corpus of all retrievable communications presented to previous IFSA symposium and of two corpus of scientific bibliographic notices extracted from CAB and SCI of the Web of Knowledge. Those corpuses have then been mined and analysed with two software that are dedicated to textual and scientometric analysis of corpus¹.

Co-word analysis is a small branch of network analysis which is largely grounded in Actor-Network-Theory (Callon et al., 1983) and in the implementation of specific algorithm of mapping scientific knowledge. Born in relation to the evaluation and policy of science (Callon et al., 1986; Law et al., 1988), cword analysis is a critical prolongation of the early approaches of co-citation (Small, 1973) and it relays largely on techniques of full text indexation. The relevance of cword analysis for mapping large scientific domain has received critiques in relation to the significance of the relationships of word and its context of enunciation (see lately Leydesdorff & Hellstein, 2006). Thus, it is to be noticed that other types of characterization exists and that we only propose one possible way of characterizing knowledge dynamics.

At present, the evolution of the analysis of scientific networks is largely attached to the question of characterizing collaborative and cognitive dynamics of knowledge production (Powell et al., 2005) and to the emergence of multi or trans-disciplinary emerging fields of research (Lucio-Arias, Leydesdorff, 2007) or paradigmatic field of research (Chavalarias, Cointet, 2008). Tracing and mapping knowledge in scientific database or in other electronic sources represents a huge field of problems for many disciplines dealing with information. It is also the case for co-word analysis (Mogoutov, Kahane, 2007). More locally, in relation to specific area of research, mapping

¹ One, called Beluga, is free downloadable software recently designed (<http://genome.jouy.inra.fr/~turenne/beluga.html>), and the other, called RéseauLu is well know within the community of Scientometrics and powered by Aguidel Technology (<http://www.aguidel.com/en/>).

heterogeneous networks appears to help the understanding of social dynamic of research activities (Cambrosio, Keating, Mogoutov, 2004; Cambrosio et al., 2006; Bourret et al., 2006). Though our communication remains descriptive, it participates to the effort of social studies of regime of knowledge production in agricultural science and of the significance of sustainability (Barbier, Mogoutov et al., 2007).

The use of those tools in the context of an interactive work with members representing a scientific community is a significant way of realizing a kind of participatory sociology of scientific knowledge, trying notably to avoid an evaluative perspective and more to co-design a situation of using tools in a comprehensive way and in relation to a purpose of maieutic intervention. This attitude toward network using co-word analysis mapping in interaction with a scientific community shares many ideas of shifting the use of tool from a scientific context to a science policy context (Noyons, 2001).

2. Methodological framework

Using methods and technologies of mining textual corpus and analysis the co-occurrence of terms allows then to build a general view of the knowledge of a community. Based on an empirical study of some descriptors of the social and cognitive dimension of IFSA community, we propose such general view of the concepts and the kind knowledge that characterise this particular community and its scientific context.

The notions of “knowledge” and “concept” are – in this context- reduced to what is written in communications and scientific articles. Therefore, we do not pretend that this strong limitation is the one-best-way of tackling with knowledge dynamics; on the contrary, we expect to create the possibility of framing interaction with other ways of realizing and discussing knowledge in a scientific community through formal or informal exchanges. Nevertheless we assume that the IFSA community might need this kind of confrontation to formal and reductionist account of scientific production in order to establish its own way of exploring the future.

The possibility of this kind of science mapping and the method of the present study have been presented and been discussed with colleagues of the IFSA (and in a couple of meeting with Benoit Dedieu; B.Hubert; Marianne Cerf particularly). Thus, the orientation of this study is the result of collaborative work which aimed at mobilizing some representatives of the community in a bottom-up process to establish both the objectives of the analyses and an intermediary set of momentum of discussion to make decisions and exchange on word significance. The following results have been also presented in a final oral restitution.

The general methodological framework was defined as followed. A first phase has consisted in the exploration of the content of the database of all communications proposed at IFSA previous symposia

and of the general scientific background defined by the keywords [farming system] in WOS DataBase (CAB and SCI). A second phase followed the decision to realise a content analysis of the IFSA DB in order to characterize the dynamic of the concepts of the community on one side, and to realise a supervised retrieval of bibliographic notices in CAB and SCI both at the world level and the European level. The definition of the query to retrieve bibliographic notices has been particularly a long and difficult process since it relayed on three approaches that has been pooled: the result of a systematic selection of terms according to a measure of the specificity of keywords; the result of a terminological extraction realised on the title, keywords, abstract of the IFSA communication; the views of IFSA experts on those collections of terms.

The general purpose was thus to create a pool of data and partial analysis in order to ground the reflection of the IFSA scientific board and more largely IFSA members on this two task: construction, data mining and information extraction of the IFSA Database (section 3) and information retrieval plus coward analysis on the contextual database (section 4). The development and results of these two tasks are going to be presented separately in the next sections, though they have been realized simultaneously and interactively.

3. Constitution and analysis of the IFSA corpus

3.1. Creating data and quality of the corpus

The very first work was to build up a database of the communications submitted to the IFSA during the last six symposiums. To run that task we recovered the texts of all communications and worked with the constitution of tagged standardized files, in order to allow the non-ambiguous recognition of the fields and the contents of fields. It is a rather ungrateful and meticulous work which leads to the constitution of a database under Excel containing 325 communications covering the 1996-2006 period of time and described with 15 fields. Each communication has been processed to lead to standard fields. An important work was carried out on the affiliations of the authors in order to be able to work on the affiliation of the members of the community. This preliminary work aimed at allowing thereafter a lexical indexing with Beluga and some frequential measure of fields.

Working on the contents of communications shows a strong heterogeneity even if each symposium is observed, this is obviously related to the multi-field character of the community. The bibliographical references are present in all the communications but rules of presentation are very diverse. Thus we should have re-built a common structure of references but we did not want to spend too much time on this issue since the task was huge and the benefit of it would be low because of a small number of communications and a large diversity of references. This work of ordering data allowed realizing

analysis by simple counting of occurrences within the fields. Those results are presented below in graph 3 and make it possible to grasp in a synthetic way how the IFSA structure its own textual memory.

One notes a very low level of standardization of the inner structure of communications as simply as the presence of a summary and key-words. But that varies through time with few key words and the part of communication that are labelled “Discussion”. It seems that the first 2 years, a notorious effort of standardization was done with a climax in 2002 for the symposium of Florence, then it seems that the homogeneity of communications got less structured in terms of text composition (20% of the communications without summaries appear a high figure). This, of course, does not say anything at all about the scientific quality of the contents.

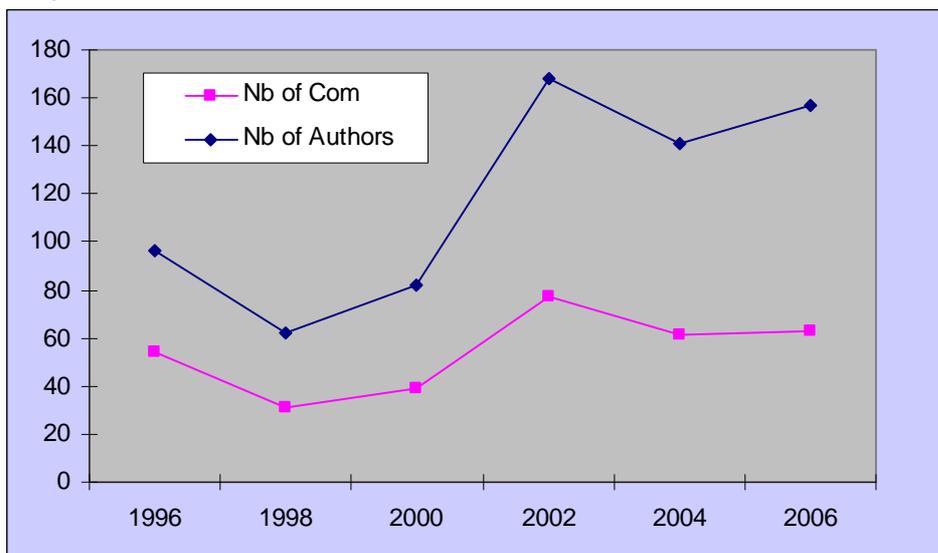
3.2. Results

The following graphic shows the number of documents and authors each year and the ratio for each symposium (see Table 1). The evolution shown in graphic 1 accounts for the importance of the 2002 symposium of IFSA in Florence. A more Les auteurs publiants représentent 5% du nombre d’auteurs total concernés par l’étude et assurent 36% du total des communications présentés lors des symposiums IFSA de 1996 à 2006.

Table 1.

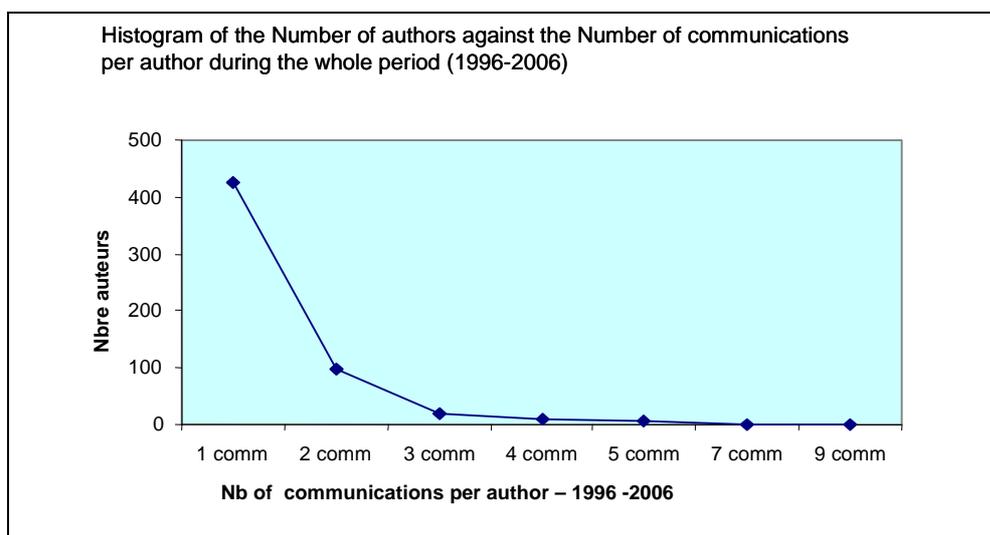
Grenade : Technical and Social Systems Approaches for Sustainable Rural Development	1996
Hohenhein : Rural and Farming Systems Analyses Environmental Perspectives	1998
Volos : European Farming and Rural Systems Research and Extension into the next Millenium. Environnemental, agricultural, and socio economic issues.	2000
Florence : Farming and Rural Systems Research and Extension Local identities and globalisation	2002
Vila Réal : European farming and society in search of a new social contract - learning to manage change	2004
Wageningen : Changing european farming systems for a better future	2006

Graphic 1.



A second result shows the distribution of the number of communication per author during the whole period (graphic 2). It establishes the fact that a large number of participants have been contributing only once or twice during the period. This is certainly the effect of participants that are in doctoral period and thus might not stay after it for various reasons. But an analysis of the trend of contribution through time shows that it is also due to the fact that there is a strong turn over of contributors.

Graphic 2.



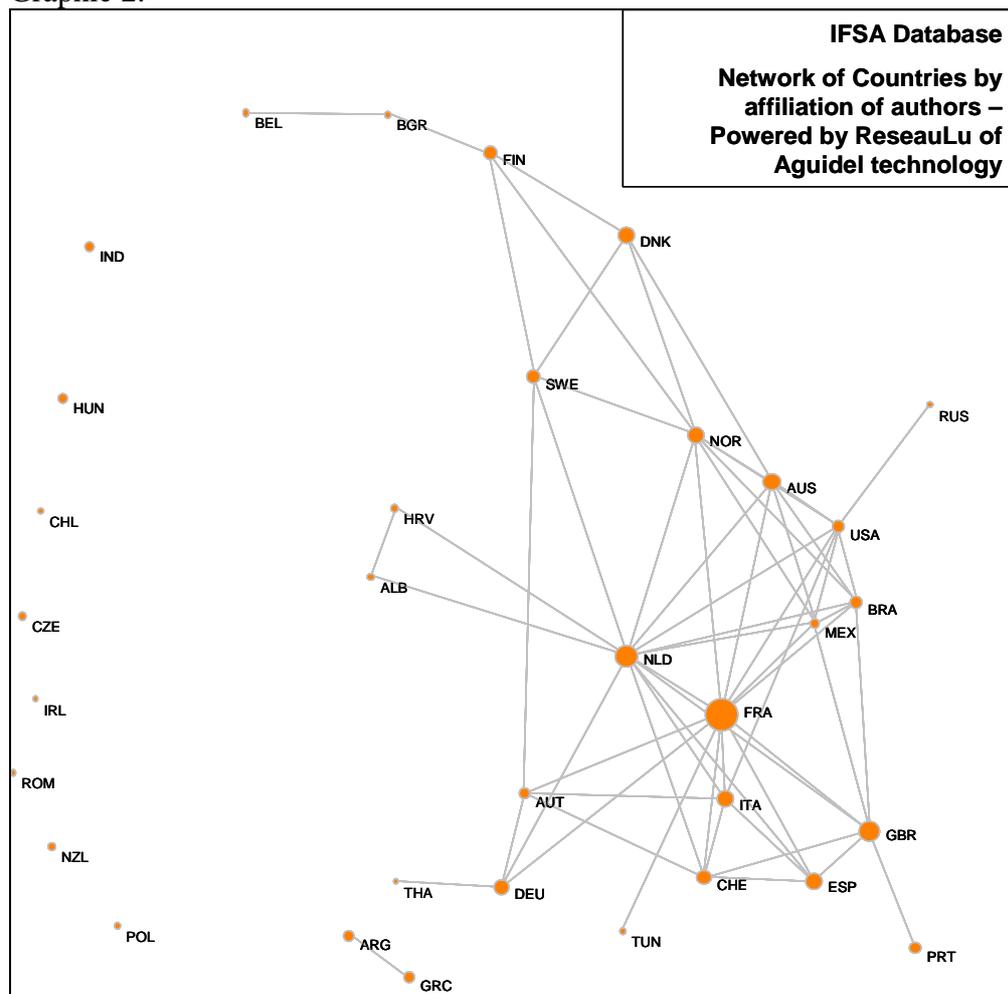
We looked also at the affiliation of authors as person working in an institutions placed somewhere in the world (there is no question of nationality there). We established the evolution of the contribution countries through time and selected the more represented (at least 4 times) in table 2. In order to complete this general view we propose a mapping of the relation between countries that are built on the co-authorship in communication (Graphic 2)

Table 2.

	1996	1998	2000	2002	2004	2006	Total
1 France	18	1	8	11	10	14	62
2 Allemagne	5	9	4	9	6	3	36
3 Pays-Bas	3	3	2	6	7	12	33
4 Royaume-Uni	8	6	3	4	7	2	30
5 Australie		2		4	8	6	20
6 Italie	7		1	8	2	2	20
7 Suède	2	1	4	5	3	4	19
8 Danemark	2	1	1	3	5	6	18
9 Espagne	7	1		3	4	1	16
10 Portugal	3	1		2	1	5	12
11 Grèce	1	2	5	2		1	11
12 Brésil			1	2	2	4	9
13 Finlande			1	3	3	2	9
14 Norvège					4	5	9
15 Argentine			5	3			8
16 Autriche				3	1	3	7
17 États-Unis				3	2	2	7
18 Mexique		1				3	4

NB : the contribution of a country when many co-authors of this country contribute is counted only once.

Graphic 2.



Finally in order to tackle with the lexical composition abstract of the communications, though the corpus is heterogeneous through time (see Graphic 3), we realised very lately an Upstream Hierarchical Classification of terms of abstract with Alceste Software (Reinert, 1990). The quality of the analysis is very good according to the standard in use for this software since more than 80% of the Elementary Context Unit (e.c.u.) of the corpus is classified and since the repartition is realized in 4 classes, see table 3 (1st class 405 e.c.u. and 6793 analysable terms; 2nd class 580 e.c.u. and 9733 analysable terms; 3rd class 652. e.c.u. and 10758 analysable terms; 4th class 184. e.c.u. and 3149 analysable terms).

Table 3.

Class 1. describes a certain type of vocabulary dealing with problems of rural and agricultural areas with a clear emphasis on the way family farming is performed in difficult or marginal milieux, it is also clearly the class that deals with cattle and grazing.

Class 2. describes a certain type of vocabulary dealing very clearly with the issue of learning and participation in rural development, extension activities, knowledge transfert situation and research intervention; it also refers a lot to reflexivity and interdisciplinary.

Class 3. describes a certain type of vocabulary dealing very clearly with socio-economical and anthropological approaches of belief, knowledge, rationales, markets in relation to global context of knowledge society or to public action (CAP, standards, globalization).

Class 4. describes a certain type of vocabulary dealing with agri-environmental terminology largely centred on fertilization issues and agronomic question of cropping system in a context of sustainable development (water quality, biodiversity, landscape, erosion).

3.3. Conclusion

It is necessary to remind that the results we have produced are very much dependant on the quality of the data of the IFSA. One could recommend to the organizers of IFSA to provide the community with a robust information system that would also help to realise the kind of analysis we have delivered here. We show that something happened in 2002 since we acknowledge a kind of revival of the community perhaps due to the place of the conference. The community is very much European with major leading countries (France, Germany, the Netherlands and the UK). It seems that the turn over of participants is quite high and the more contributive authors are very few. Though we could not perform an analysis of the citations of the communications because of a high heterogeneity of format, we noticed that there is not much memory of seminal works like important books or articles that intend to define the community. Finally it is important to notice that our lexical classification does not match very precisely the distribution of thematic workshop proposed during the symposium.

4. Building and mapping the scientific context of IFSA community

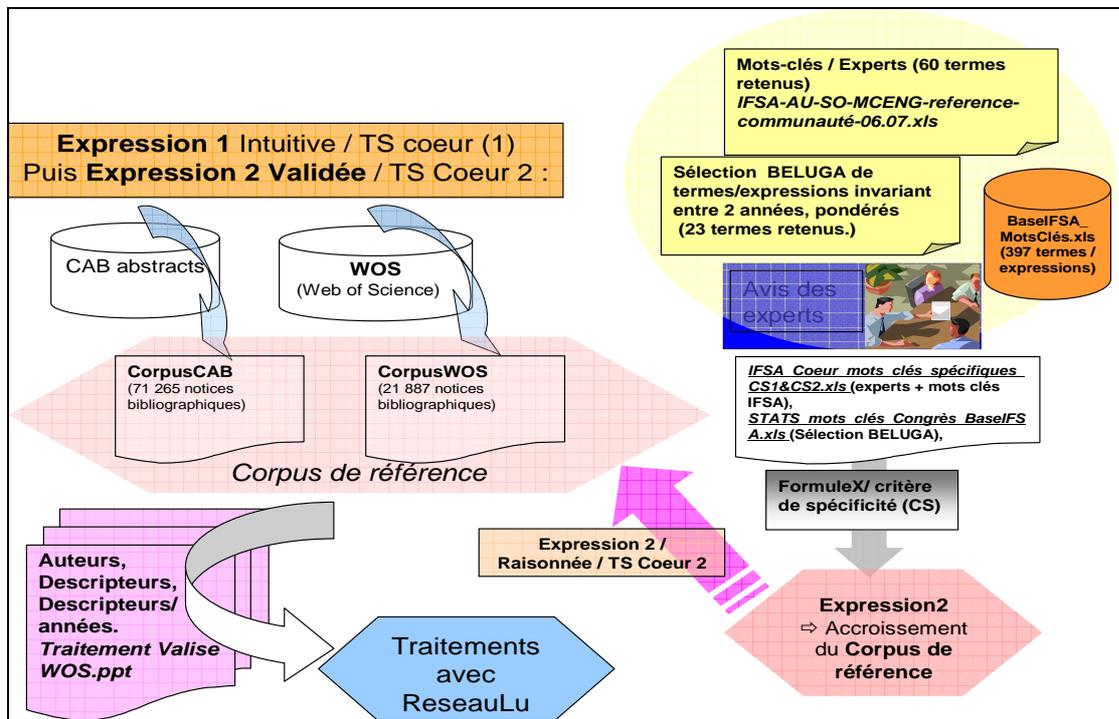
4.1. Methodology

The purpose of the second part of the study was to build and to map corpus of references of the IFSA community. This supposed that a specific query would be designed and used in the Web Of Science Databases in order to retrieve corpus of bibliographic notices both at the European and at the world level. In order to realize this task we have chosen an incremental method of query definition (see Scheme 1 for a visual representation).

We have firstly defined a simple query based on terms defining the community in terms of systemic approaches dealing with agriculture, agroforestry and farming. Than we have refined this query including the collaboration and critical views of IFSA experts on one hand (58 new terms were proposed and their specificity tested), and also a list of terms and expression that were taken from the terminological extraction of the IFSA database with Beluga (23 terms tested). It is also to be noticed that the specificity of each term comparing its occurrence in the corpus and in the WOS in general has been performed and discussed with IFSA experts. This methodology is largely due to a work realized

also in a similar study to define an appropriate query to represent the field of sustainable development, if any (see Barbier et al., 2008).

Scheme 1.



The result of this work holds in the table 4, but it is a time consuming and very systematic work which means a complete different practice than any common uses of the WOS by researchers. It is also as such a result to be tested or used by others.

Table 4.

Initial Query1: TS=("agroforestry systems" OR "agroforestry system") OR TS=("agrosilvicultural systems" OR "agrosilvicultural system") OR TS=("farming systems research" OR TS="farming system research") OR TS=("grazing systems" OR "grazing system") OR TS=("farming systems" OR "farming system") OR TS=("cropping systems" OR "cropping system") OR TS=("agricultural systems" OR "agricultural system") AND PY=(1996-2005 OR 2006)
Final Query =(("farming systems" OR "farming system") OR TS=("cropping systems" OR "cropping system") OR TS=("agricultural systems" OR "agricultural system") OR TS=("agricultural knowledge") OR TS=("farmers participation") OR TS=("natural resource management") OR TS=("nature conservation") OR TS=(small scale farm* OR smallholder farm* OR family farm*) OR TS=("livestock systems" OR "livestock system") OR TS=("organic agriculture") OR TS=("livestock farming system" OR "livestock farming systems") OR TS=("rural system" OR "rural systems") OR TS=("agrarian system" OR "agrarian systems") OR TS=("local food" OR "local foods") OR TS=("pluriactivity" OR "pluriactivities") OR TS=("social learning" OR "social learnings") OR TS=("Farm management")) OR TS=(("livelihood" or "livelihoods" or "system approach" or "systems approach" or "household" or "households" or "R&D" or "research development" or "extensions systems" or "extension system") AND TS=(agricult* farm or farming or rural))
CAB --- 71 265 bibliographic notices WOS --- 21 887 bibliographic notices

4.1. Mapping of corpus

The ReseauLu approach is based on the construction of association matrix and on the mapping of the co-occurrence of indexed terms that are issued by a systematic parsing of database using Access at the back. The ReseauLu mapping algorithm is object oriented and optimizes the positioning of objects in a two dimensional space focusing on the existence of strong links between objects. The positioning of objects and links in maps is not at odds and results of an optimization process using a step by step positioning that depends on explicit rules which mobilize concepts of network analysis (structural equivalence, centrality and betweenness of objects). Thus, the mapping has to be read according to the idea of global scale that shows the poles that structure the space of object in major opposition of distribution of links; then on give poles, the composition of the pole is to be approached; and finally at the local scale specific description might be establish on the links between two objects. It is to be noticed that the size of the object is proportional to its number of occurrences and that the links are without any measure of intensity, they only indicate the existence of a link (of course a heavy object is having many links). The results are presented in the following selection of significant Mappings. We have chosen to deliver mapping of co-authorships and mapping of key-words co-occurrence in order to present a social and cognitive description of the corpus of reference and we did it at the world level and the European level. It was also much needed to deliver a representation of the dynamic of networks of key-words through time, and ReseauLu powers a very useful possibility of establishing heterogeneous networks of Year and Key words. All these mappings request oral comments in order to explain what kind of structure is at work in each case. We only proposed here a set of map that are commented in the next section..

4.2. Results

The difficulties encountered but overcome to create database IFSA indicate that the IFSA community must be more vigilant in the future with the compulsory structuring of communications in order to support a follow up of the scientific production, and not only the kind of work we have proposed here. This supposes perhaps a permanent information system used by the IFSA symposium organizers. Secondly it appears that conceptual positioning of the IFSA community, described through its list of authors is not fitting to a world wide or even a European mapping of the scientific knowledge context (see Mapping 3, 4 and 5). The IFSA community is not much visible as such in the network of co-authorship. Moreover prominent authors of the IFSA context are not really taking part to the IFSA symposium. It is to be noticed that the CAB and SCI databases show differences, but this differences do not change the fact that there is not a social visibility of the IFSA community in its own scientific context.

The evolution of keyword positioning through time (Mapping 7) shows clearly a polarization on what is sometimes called an ecologization of knowledge production. Issues of plant protection, nature conservation, biodiversity appears to be at the heart of the more recent production of knowledge. The agri-environmental wave seems to have passed away in 2001. At the world wide level (see Mapping 1 and 2), the evolution of the scientific context of IFSA is showing a shift from classical agronomic system approach of crops and farming in the mid 90's to the uses dealing with sustainability in the end of the 90's. Then it appears a second shift in the early years of 2000 in relation to issues like erosion, soil fertility basically in Africa. The next significant shift in the mid 2000 concerns grazing, soil quality and natural resource management and deforestation: climate change is becoming central. It will be interesting to study whether the 2008 symposium reflects also this evolution.

The mapping of the European context of IFSA (Mapping 6) reveals a structure of 5 poles: 1. nature conservation and grassland; 2. sustainability and farming systems; 3. crop protection; 4. cropping system, fertilizing and rotations; 5. farm management. It is to be notices that [organic farming] is a powerful connector between pole 2 and 3 and [land use] a connector between pole 1 , 2 and 3 while [simulation models] connects pole 4 and 5. This structure is not completely fitting to the classification proposed in Table 3.

Conclusion

In order to open the discussion in a provocative way, one might take the opportunity of this general view of IFSA repertoire and IFSA scientific context to question the positioning and constitution of this scientific community. Firstly, it seems that the central position of this community is to produce scientific knowledge which is very attached to areas, territories and human or/and natural milieu. The contribution of this type of knowledge to the IFSA scientific context does not seems to be very visible. This kind of scientific contribution might be addressed by the claim that it does not says much about how it takes part to large issues and public problems such as global change, poverty, migration, food security (see Mapping 1) and to the way it could also find a significant position within the buzz concept of sustainability (see Mapping 8 taken from Barbier et al., 2008).

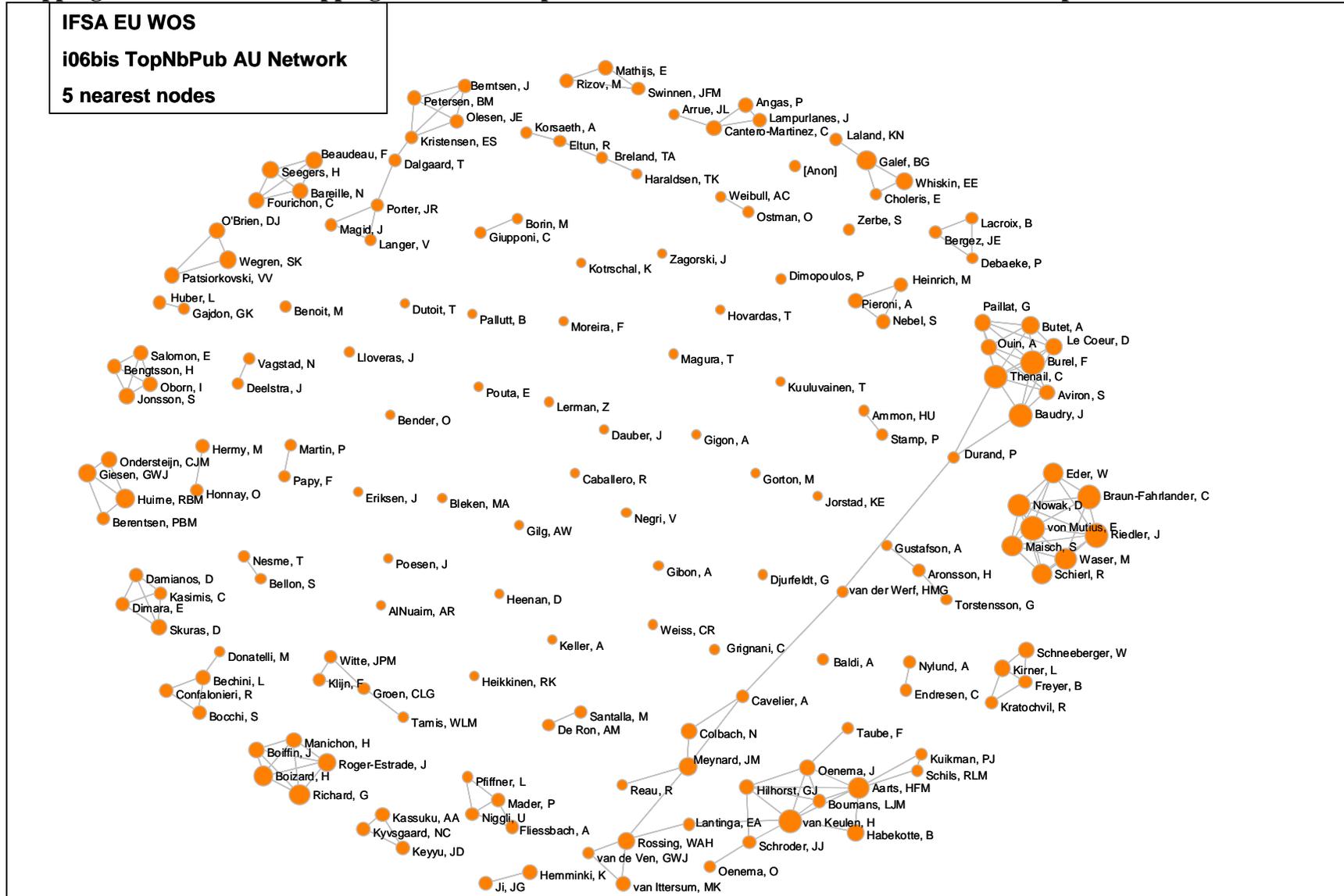
Secondly, our study and interactions with the colleagues of IFSA indicates that the nature of the community is also a matter of debates. Is IFSA a scientific community? Certainly yes, since most of its members belong to the scientific community or are working in close connection. Nevertheless, one might also say that it is not a scientific community academically driven. This might certainly explain the feeling of a distance between the type of knowledge production in IFSA and the kind of certified knowledge that is to be found in the database that are supposed to mirror the conceptual orientation of IFSA community. It also means that scientometric tools and lexical approaches certainly encounter limits there.

At last, it would be certainly misleading to understand this conclusion as a critique of a supposed “insufficiently developed academic nature” of IFSA. The type of issues addressed by scientists of both IFSA community and more largely by the farming system community indicates that the problems we are facing with climate change, food security, energy and sustainability are not going to be solved solely by a more modern or efficient scientific production regime of any kind. It means then that some hybrid type of organizations are needed, organization that could be as much dedicated to product scientific knowledge than to enlighten the political use of knowledge and knowing when science is realized in societies. Is IFSA one of those? At least, it is not said in the label.

REFERENCES

- Barbier M., François-Moll F., Vichnevskaja T. et Mogoutov A. 2008. Figures de la "soutenabilité" dans des bases de données scientifiques du "Web of Science". Exploration et discussion de traitements scientométriques, *Communication au Séminaire Changement climatique, expertise et fabrications du futur : Modèles et scénarios, politique et gouvernance*, Centre Koyré, Paris, 8 Avril 2008.
- Bourret P., Mogoutov A., Julian-Reynier C., and Cambrosio A., (2006). A New Clinical Collective for French Cancer Genetics A Heterogeneous Mapping Analysis, *Science, Technology, & Human Values*, 31 (4): 431-464
- Callon, M., J. Law, A. Rip (1986), *Mapping the Dynamics of Science and Technology*. London: The MacMillan Press Ltd.
- Callon, M., J. P. Courtial, W. A. Turner, S. Bauin (1983), From translations to problematic networks: An introduction to co-word analysis, *Social Science Information*, 22: 191-235.
- Cambrosio A., Keating P., Mercier S., Lewison G., and Mogoutov A., (2006). Mapping the emergence and development of translational cancer research, *European journal of cancer*, 24: 3140-3148
- Cambrosio A., Keating P., Mogoutov A. (2004). Mapping collaborative work and innovation in biomedicine: a computer assisted analysis of antibody reagent workshops, *Social Studies of Science*, 34 (3): 325-364.
- Chavalarias D, Cointet JP. (2008). Bottom-up scientific field detection for dynamical and hierarchical science mapping, methodology and case study, *Scientometrics*. 75(1): 37-50.
- Leydesdorff L, Hellsten I., (2006). Measuring the meaning of words in contexts: An automated analysis of controversies about 'Monarch butterflies,' 'Frankenfoods,' and 'stem cells', *Scientometrics*, 67 (2): 231-258.
- Lucio-Arias D, Leydesdorff L, (2007). Knowledge emergence in scientific communication: from "fullerenes" to "nanotubes", *Scientometrics*, 70 (3): 603-632
- Noyons E., (2001). Bibliometric mapping of science in a policy context, *Scientometrics*, 50(1): 83-98.
- Powell et al., 2005 W.W. Powell, D.R. White, K.W. Koput and J. Owen-Smith, Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences, *American Journal of Sociology*, 110 (2005), pp. 901–975.
- Reinert, M. (1990). ALCESTE, une méthodologie d'analyse des données textuelles et une application: Aurélia de Gérard de Nerval, *Bulletin de Méthodologie Sociologique*, 26 : 24-54.
- Small, H. (1973), Co-citation in scientific literature: A new measure of the relationship between publications, *Journal of the American Society for Information Science*, 24: 265-269.
- Turenne N., Barbier M., (2004). BELUGA : un outil pour l'analyse dynamique des connaissances de la littérature scientifique d'un domaine. Première application au cas des maladies à prions, In *Proceedings of Extraction et Gestion de Connaissances*, Hébrail G. and Lebart L. eds., Clermont-Ferrand, 2004, France

Mapping 4. Scientometric mapping on SCI – European Context of IFSA – network of Co-authorships



Mapping 5. Scientometric mapping on CAB – European Context of IFSA – network of Co-authorships

