Designing healthy and sustainable food systems: how is research contributing?

Agénor Lahatte, Elisabeth de Turckheim, Lucile Chalumeau

Designing healthy and sustainable food systems: how is research contributing?

Agénor Lahatte[1], Élisabeth de Turckheim[1,2] and Lucile Chalumeau[1]

[1] agenor.lahatte, elisabeth.de-turckheim, lucile.chalumeau)@hceres.fr
OST, High Council for Evaluation of Research and Higher Education (Hcéres), 2 rue Albert Einstein, 75013, Paris (France)

[2] Délégation à l'évaluation, INRA, 147 rue de l'Université, 75338 Paris Cedex 07 (France)

Abstract

A method for delineating the research area addressing a complex global problem is developed in a context where no core set of documents is available and where the objectives of the research are described as the production of broad knowledge related to societal outcomes. The method, entirely based on textual analysis has two steps: (a) multi-terms selected from a policy document are used to retrieve documents from a publication database, (b) a LDA topic model fitted to the retrieved corpus allows to identify irrelevant topics and to remove the corresponding documents. Once the corpus is cleaned, topics are clustered on the base of their semantic proximities and of their co-occurrence in documents for structuring the domain into main research themes. The method is applied to a selected facet of the food security challenge and provides a representative sample of research works in the domain. Structuring the corpus into six research themes allows to compare the thematic profiles and the research practices of twelve countries most involved in the domain.
Keywords: corpus delineation, societal challenge, topic model, food security, sustainable food system

Introduction

Assessing the contribution of research addressing complex global challenges has become increasingly important in science policy as research framework programme or strategies tend to align on societal demands. This raises the issue of identifying the research areas that may contribute to the challenge either to define scientific programmes or to assess the achievement of such research policies. In these situations, the first problem is to delineate a relevant corpus of scientific documents.

This issue of delineating a research domain is regularly addressed in science policy or scientometrics. This is for instance a key point for emerging scientific domains (Milanez, Noyons & de Faria, 2016; Raimbault, Cointet & Joly, 2016). In such cases, usual methods select specific scientific multi-terms and, after some trial and error, use these multi-terms as queries to retrieve documents from a database. The resulting corpus is then either cleaned or extended to reach a satisfactory balance between recall and precision of the final corpus.

However, when the objectives of a programme are oriented towards social needs, they are not directly characterized by scientific keywords as they point to expected societal outcomes of research and should leave open a diversity of scientific options (Wallace & Rafols, 2015). In such cases, queries have to be more general to retrieve documents in any possibly contributing field. But queries based on general language terms, partly because of words polysemy, may harvest documents that are out of the scope of the programme. Therefore an intensive cleaning step may be necessary.

Topic models are adapted for screening large datasets. They only rely on textual information and have been presented as an efficient method to extract relevant documents from very large corpora (Klavans and Boyack, 2014) or to find the thematic structure of sets of non academic documents as web pages or press articles (Di Maggio, Nag and Blei, 2013). We show here

how topic models can be used to clean a corpus that may contain important subsets of irrelevant documents.

As a case study, we choose to delineate a corpus of publications addressing the challenge of food security and we focus on a facet of the challenge entitled *Designing safe, sustainable and innovative food systems* from a policy document of the French Ministry of Higher Education and Research (MESRI, 2014). We extract from this document the main questions addressed to researchers and delineate a related corpus. This corpus is then analysed through six main research themes. We compare the scientific production of various countries to the domain and characterize the degree of interdisciplinarity of themes and of countries contributions.

Corpus delineation

The delineation based on a policy document consists in two steps: first a selection of key phrases that are used as queries to retrieve documents from the chosen database and, second, a cleaning step of the retrieved corpus.

Choosing queries

The policy document is a workshop report of the Expert group of the Ministry (MESRI, 2014) that focuses on a set of selected research questions which are: How to improve the efficiency of the food supply chain? What are the determinants of consumers' behaviour? How and why diets are changing? How diets are influencing people nutritional status and health? What are the environmental and social impacts of food systems? Which policy and institutional actions could enhance food security and nutrition?

We reorganize these research questions as 8 items. An overall issue and 7 elements of food systems according to the conceptual framework of food systems of the FAO Committee on World Food Security (HLPE, 2017, page 26)

1. Sustainable food systems ensuring food security
2. Food chain efficiency with a focus on food process optimisation, food circulation, localized agri-food, food losses and waste reduction
3. Food environments impacting consumers' behaviour: food availability, food information...
4. Consumers' behaviour
5. Diets and diet transitions
6. Environmental and social impacts of food systems
7. Nutrition and health outcomes of diet patterns
8. Policy and institutional actions to move towards sustainable food systems.

According to these research questions, we design 32 queries (Table 1).

We use the OST home version of the WoS database restricted to citable documents (article, letter, review), with at least one address and a WoS category. We query for documents published between 2012 and 2017 where the selected key phrases appear in the text that merges title, abstract and authors' keywords of each article.

Table 1. Selected research questions and associated queries

Questions to scientific research Item #	Query#	Queries
1 Sustainable and healthy food systems, that ensure food and nutrition security	1	sustainable AND food system%
	2	sustainable food
	3	food system% AND (safe OR safety)
	4	food system% AND (secure OR security)
	5	food security OR nutrition security
2 Food chain efficiency: Optimisation of food processing and production, optimization of food circuits, reduction and recycling of food losses and waste, localized agri-food production	6	(efficien% OR optim%) AND (food process% OR food production)
	7	eco\-design AND food
	8	food preservation AND (technolog% OR process%)
	9	preservation AND (food process% OR food technol%)
	10	optim% AND (food circuit% OR food supply)
	11	short AND (food circuit% OR food supply)
	12	food waste%
	13	food loss%
	14	local% (agrifood OR agri\-food OR agrofood OR agro\-food)
3 Food environments that impact consumer behaviour: Food availability and supply, food information, food quality and safety	15	food availab% and (choice% OR consumer%)
	16	food supply AND (choice% OR consumer%)
	17	food (information OR advertising OR marketing OR promotion) AND (choice% OR consumer%)
	18	food quality AND (choice% OR consumer%)
	19	(food safety OR safe food) AND (choice% OR consumer%)
	20	(healthy food OR healthy eating) AND (choice% OR consumer%)
4 Consumer behaviour	21	consumer% behavio% AND food
	22	food choice% AND consumer%
5 Diets and diet transitions	23	(food OR diet OR dietary) pattern%
	24	(food OR diet OR dietary OR nutrition) transition%
6 Impacts of food systems	25	social impact% AND (diet OR dietary OR food system%)
	26	environment% impact% AND (diet OR dietary OR food system%)
	27	impact% on environment AND (diet OR dietary OR food system%)
7 Nutrition and health outcomes	28	health benefit% AND (diet OR dietary OR food pattern%)
	29	healthy (diet OR dietary OR food pattern%)
	30	health risk% AND (diet OR dietary OR food pattern%)
8 Policy and governance	31	(food OR nutrition) polic%
	32	(food OR nutrition) governance

Cleaning the initial corpus by identifying irrelevant topics

The resulting corpus C1 contains 24,890 documents. To screen this corpus and control that there are not too many documents out of the scope of the study, we fit a LDA topic model (Blei, Nag & Jordan, 2003; Blei, 2011) to identify possible irrelevant documents. We choose the grain to examine the corpus as approximately 1,000 documents so that a model with 25 topics is used to scrutinize corpus C1. Among the 25 fitted topics, 3 of them are considered out of the scope of the study. A topic on ecology, with papers on marine ecology, evolutionary developmental ecology (evo-devo) and other specialities in ecology was collected with the queries on diet patterns or on food availability, because these terms are common for human diets and for other animal species. A second topic on plant genetics and plant science is considered marginal for this study, despite the fact that improving crop yield with selection

and genetic engineering is a way to improve food production. We consider that the articles collected here do not correctly represent the whole research on crop genetics and plant physiology. We therefore decide not to include this research area in our study. The third topic is focused on the impact of particular food components on the metabolism, through clinical and experimental research. Again we decide not to include this level of scientific investigation in the study.

As the same queries retrieve documents in both relevant and irrelevant topics, it is not possible to remove initial queries. It also appears difficult to refine queries, for instance so that they only apply to human diets. Therefore, in order to remove the documents focused on unwanted topics, we use a similar method as Milanez, Noyons & de Faria (2016) and select terms that are specific to these irrelevant topics to remove off-domain documents. Such terms are easily found with the LDAvis software (Sievert & Shirley, 2014) setting the parameter λ of the *relevance indicator* very low, thus selecting terms with a high probability in the topic and a very low probability in the rest of the corpus. These specific terms are used as counter-queries: documents of corpus C1 that use them are removed to get corpus C2. We also remove articles in journals of WoS categories of the specific scientific approaches that we do not want to include in our study (Table 2). This provides a second corpus C2 of 20,500 documents where the irrelevant topics do not appear any more.

Table 2. Counter-queries and selected WoS categories to remove documents focused on irrelevant topics

Irrelevant topics	Counter-queries	Removed WoS categories
Ecology, marine and freshwater biology, evo-devo	prey%, predator%, trophic, foraging, benthic, coral, reef%, juvenile, zooplancton, phytoplancton, pelagic, invertebrate, neolithic, archeological, webs, nest%, catches	ECOLOGY MARINE & FRESHWATER BIOLOGY ZOOLOGY ENTOMOLOGY ORNITHOLOGY
Plant genetics, plant science	genome%, landraces, accession%, loci, allele%, allelic, qtl%, arabidopsis, transcriptom%, snps, nucleotide, chromosome%, microsatellite%, heterosis, heterozygo%, inbred, barcoding	PLANT SCIENCES GENETICS & HEREDITY
Metabolism and endocrinology	insulin, rats, mice, hdl (high density lipoprotein), ldl (low dentity lipoprotein), lipoprotein, nafld (non alcoholic fat liver disease), hfd (high fat diet), crp, (C reactive protein), adipose, adinopectin, leptin, tnf (tumor necrosis factor), homa (homeostasis model assessment), steatosis, interleukin, lps (lipopolysaccharide), pcos (polycystic ovary syndrome), wistar (a laboratory rat), ppar (peroxisome proliferator-activated receptor), cortisol, macrophage%	ENDOCRINOLOGY & METABOLISM

Semantic analysis of the corpus

Fitting a LDA model with 20 topics on this C2 corpus shows that the chosen research questions in Table 1 are present in this corpus and that one to three topics correspond to each question. Table 3 displays a title of these topics based on the most relevant terms (Sivert & Shirley, 2014) and on the titles of the 30 articles most focused on each selected topic. We

organize (and renumber) them into 8 groups according to the initial research questions. Some questions of Table 1 are merged in the same topic - as the issues of food policies and systems governance that are merged into the first group. In Table 3, the fist column shows the relationship between the topics and the initial questions.

Table 3. The 20 topics of corpus C2

Item #	Topic #	Topic title
1,8	1	Food systems: sustainability, policies, governance
	2	Food security: markets and trade, agriculture efficiency, land management
	3	Food insecurity: social factors
2	4	Food processing technologies
	5	Food waste treatment and biogas production
	6	Detection and control of food microbial contamination
3,4	7	Factors impacting consumers' behaviour
	8	Nutritional education
	9	Factors influencing young people's nutrition
5	10	Diet patterns
6	11	Environmental impact of the food supply chain
7	12	Health risks associated with diet pattern, health benefits of (mediterranean) diet
	13	Health risks: children and adolescents lifestyle and obesity
	14	Health risks associated with diet pattern: diabetes, depression, cancer...
	15	Food contaminants and dietary exposure to heavy metals, pesticides residues...
7	16	Healthy food components: nutraceuticals, probiotics, prebiotics....
	17	Healthy food components: antioxidants (unsaturated fatty acids, phenolic components...)
	18	Agriculture and climate change
	19	Farm animal feeding
	20	Methods and models for estimation of crop production and monitoring

The first group is composed of articles on food and nutrition security and policies (Topic 1) and on economic and trade aspects of food systems (Topic 2 and Topic 3). Three topics are focused on food technology: food processing (Topic 4), treatment of food waste and manure for biogas production (Topic 5), detection and control of food microbial pathogens, food maturation and preservation with food microbiota (Topic 6). Topics 7, 8 and 9 are devoted to consumers' behaviour and its determinant factors as food quality perception, food marketing and nutritional education. Topic 10 is a general topic on diets and diet changes. Health risks of diets are present in Topics 12, 13 and 14 with Topic 13 focused on young people diet and lifestyle. Food contaminations by heavy metals and pesticides are explored in Topic 15. One topic is about environmental impacts of food systems (Topic 11). Finally, Topics 16 and 17 are focused on beneficial food components as antioxidants and nutraceuticals (food with health benefits beyond their nutritional value). As shown on Figure 1, topics in the same topic group have similar vocabularies.

Three last topics appear in this corpus though there were not explicitly requested: Topic 18 that deals with the interactions between agriculture and climate change is linked to food (in)security and therefore close to Topics 1 and 2. It also has a vocabulary similar to the topic on environmental impacts of agri-food systems (Topic 11). However, the retrieved publications do not represent properly the whole research on climate change and agriculture interactions but only those articles that explicitly cite the impact on food security. This domain related with the food security issue should either be more intensively included or treated separately. Topic 20 is an ancillary topic, with methodologies for data collection and analysis and tools for image treatment, surveys etc. Finally, Topic 19 on farm animal feeding (mainly zootechnics) is out of the scope of the study. This topic was not detected in the first topic model and we later remove the documents mainly focused on this topic (326 documents).

The 17 first topics properly cover the expected research themes, except the social impacts of food systems that do not appear as a specific topic. This research question is not visible at this grain, probably merged in Topic 1.

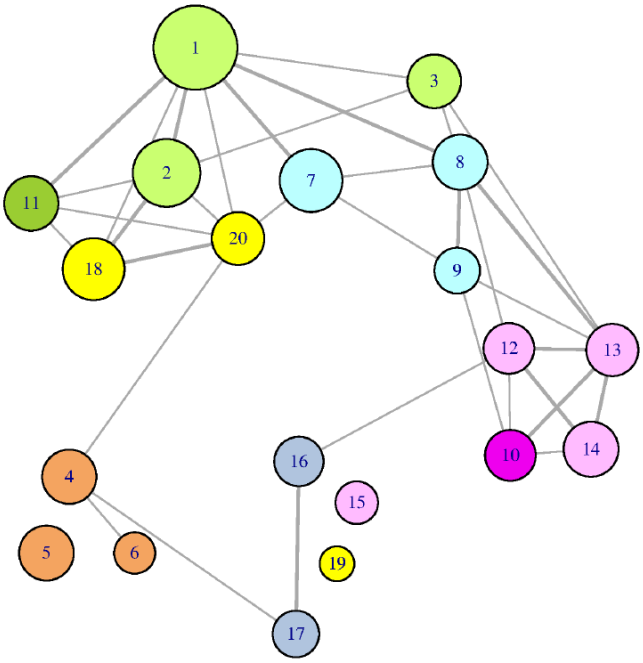

Figure 1. Map of the 20 topics of corpus C2. Topic positions are related to topic vocabulary: two topics are close in the figure if they use similar vocabularies. Bubble surface is proportional to topic weight. Edges show two levels of co-occurrence of topics in documents with a threshold chosen so as to select about 20% of the total number of possible edges (Jensen-Shannon divergence less than 0.5 or less than 0.55) .

Topic co-occurrences

In a topic model, documents combine topics. Various types of topic sharing may occur. At one end, a topic could appear mainly in specific documents where the topic weight is very high

and no other topic is important. At the other end, an ancillary topic appears in many documents with a moderate weight. This is often the case with topics gathering general scientific terms as jargon or about methodology for data collection and analysis as statistics, data modelling, image processing, survey methodology etc. that are often associated with core topics. Between these two extremes types, some topics could play both roles and groups of related topics could compose small cliques. Various measure of topic co-occurrence are possible. In a preceding paper (Cassi, Lahatte & Rafols, 2017), we used the number of documents where the weight of each of two topics is above a threshold (say 25%) to measure topic co-occurrence. In this work, we use a dissimilarity measure between the two vectors of the document:topic matrix (in this case, the Jensen-Shannon divergence).

Figure 1 shows that Topic 1 is the most connected topic with 7 edges (or degree 11 if a weight 2 is given to the 13 strongest edges of the network). Not surprisingly, the methodology Topic 20 is sharing numerous documents with other topics (6 edges) as well as Topic 8 on nutritional education (6 edges). It makes sense to say that nutritional education is a factor driving consumers' behaviour (Topics 7 and 9) and is related with social aspects of food systems (Topic 3) and with food policies (Topic 1). Topic 13 on young people diet (6 edges) shares documents with Topics 8 and 9 (on young people exposure to harmful influences) and with papers on health risks of inappropriate diets. At the other end, topics on food components (Topics 15, 16, 17) are not sharing many documents with the rest of the corpus. The three topics on food technology are also mostly specialised, with only two edges with other topics.

This connected network suggests to reconfigure the initial research questions of the policy document into five groups of topics or five research themes structuring the domain. In addition, considering that the nutritional and health status of young people is an important societal challenge and that topics 8, 9 and 13 are strongly linked, we also select a sixth theme, partially overlapping two other themes (Table 4). We leave aside four marginal topics thus focusing on a third corpus C3.

Table 4. The six reconfigured research themes

	Item #	Research theme	Topics	Document #	Theme %
Th1	1,6,8	Food security, food systems, social factors, policies	1, 2, 3, 11	6257	0.31
Th2	3,4	Consumers' behaviour and factors impacting it	7, 8, 9	3159	0.15
Th3	5,7	Diet patterns and health risks	10, 12, 13, 14	3629	0.18
Th4	7	Healthy foods	16, 17	1563	0.08
Th5	2	Food chain efficiency	4, 5, 6	2874	0.14
		Marginal topics	15, 18, 19, 20	3018	0.15
Th6		Young people nutritional and health status	8, 9, 13	2572	0.13

Analysis of country publications

Country publications

Among the countries with the largest contribution to the corpus C3, USA, Great Britain and China have more than 1,500 publications in the domain during the period 2012-2017 (Figure 2, whole counting). Five other European countries (Italy, Spain, Germany, The Netherlands and France) have more than 600 publications. Outside Europe, Australia, Canada, Brazil and India are also among the 12 most most productive countries. We focus on these 12 countries to compare their thematic profiles.

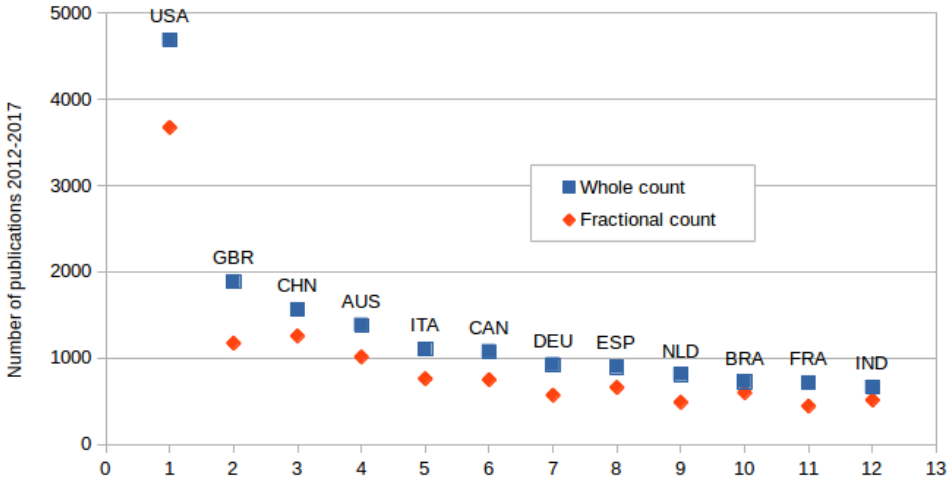

Figure 2. Number of publications of the 12 countries with the largest contributions to corpus C3

Country thematic profiles

The thematic profile of a country is obtained by attributing each document to its heaviest topic. This is a more convenient rule than assigning fractions of documents as the underlying model assumes. We also assign a document to a country as soon as it has an address in the country (whole counting).

Thematic profiles of countries show major differences (Figure 3). Canada, Great Britain, Germany and The Netherlands have a higher involvement than the world average on food systems sustainability and food policies while Spain is much less covering this theme than the world average. China and India are more involved in improving food chain efficiency but are less specialised on diet health risks. In an opposite way USA, Australia, and The Netherlands have a higher relative commitment on food risks, balanced with a low one on food chain efficiency. India, Italy and Spain seem to publish more research about healthy food, possibly in relation with regional diets as the Mediterranean or vegetarian diet. Two countries, USA and Australia, are more concerned by young people diet and lifestyle than the rest of the world and this may be related with the high prevalence of obesity in USA, Canada and Australia (OECD, 2017). These first findings seem consistent with the general context of food patterns and food system issues in the different countries.

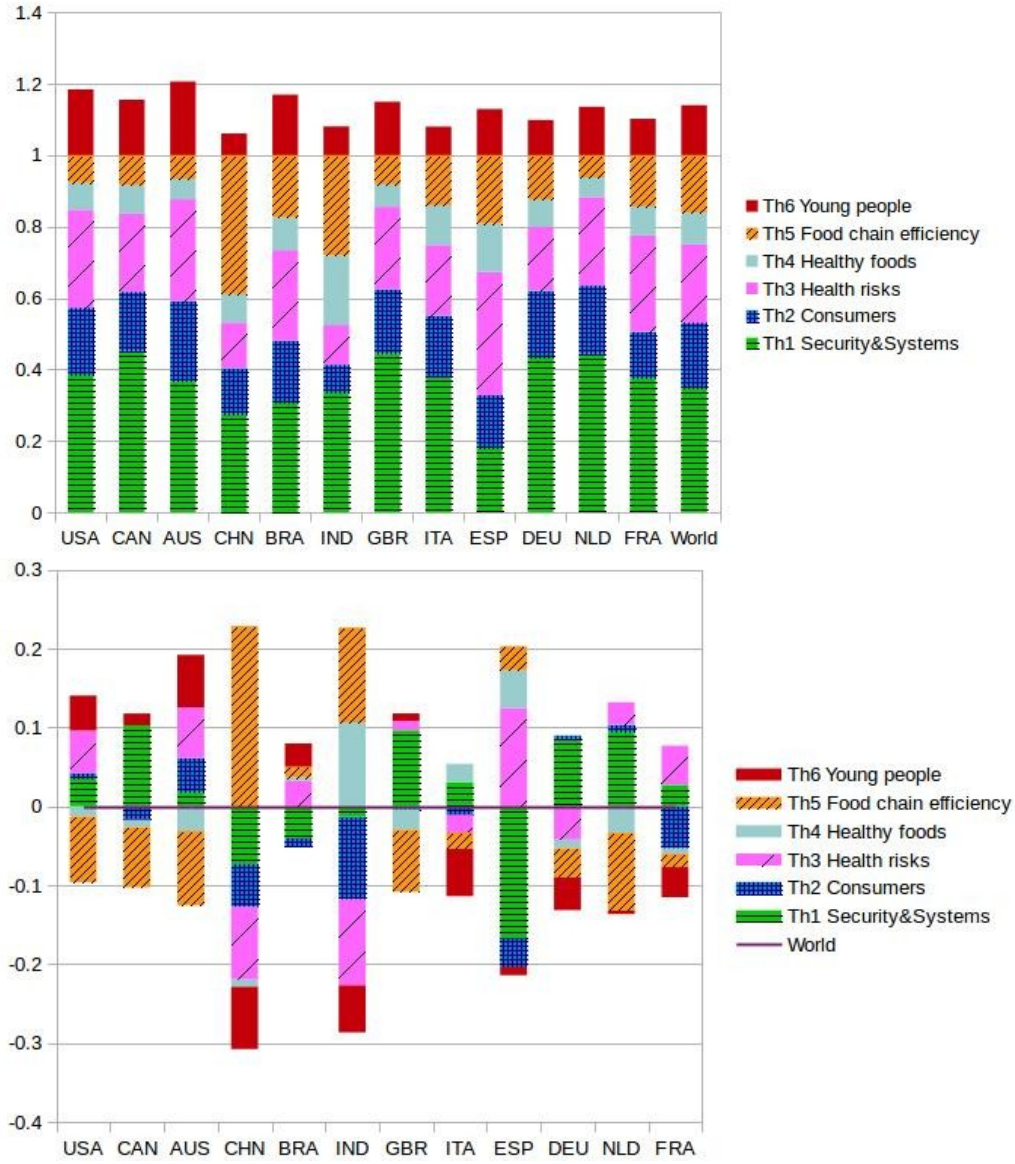

Figure 3. Thematic profile of 12 countries: distribution of publications of corpus C3 by country (whole counts) in each thematic group (top figure), difference of proportions with the world (bottom figure)

How interdisciplinary are themes and country contributions?

An expected feature of the research oriented by a complex global problem is its ability to cross disciplinary approaches (Ledford, 2015; Molas-Gallart, Rafols & Tang, 2014). It is not only the disciplinary diversity of the articles that is relevant but also the ability to combine different approaches, theories, concepts or methods in a same research work. The repartition of the corpus in WoS categories confirms the diversity of scientific fields involved but it does not inform about interdisciplinary that is achieved at the article level. To measure it, we use the integration score proposed by Porter, Cohen & Roessner (2007) based on the Rao-Stirling diversity index of the categories of references and we compare this indicator between the 6 themes and the 12 countries.

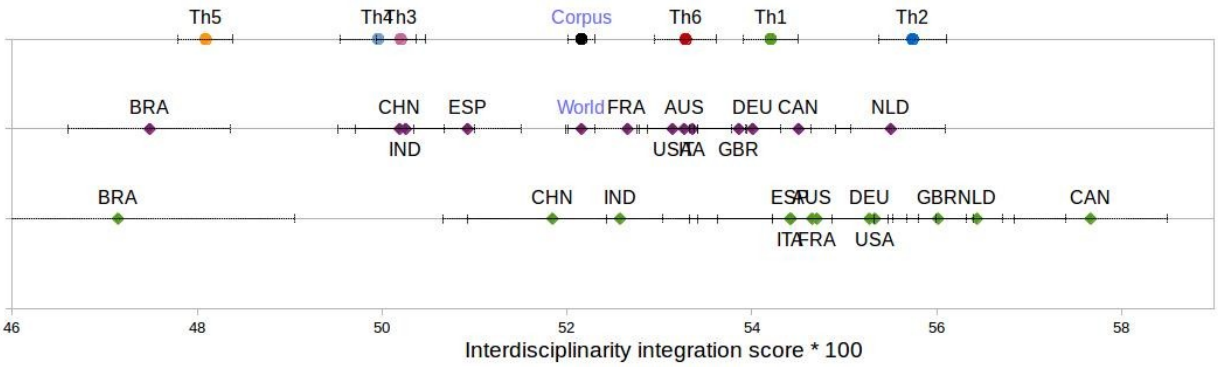

Figure 4. Interdisciplinary integration scores of themes (top line), countries (middle line), of and of countries for Theme 1 (bottom line), with probability 0.9 confidence intervals

The most interdisciplinary theme is Theme 2 on consumers behaviour. Theme 1 on food systems and policies and Theme 6 on young people diet are also more interdisciplinary than the three other themes. More data on the position of country indexes by theme shows that Brazil has very low interdisciplinary indexes in themes 1, 2, 4 and that India has the lowest index in Theme 3, just before Brazil (Figure not displayed). As China, India and Spain are most specialised in the less interdisciplinary themes, this also contributes to their low overall interdisciplinarity index. At the other side, The Netherlands have the highest index in five themes. For Theme 1, the striking position of Canada is to be mentioned, significantly above every other country but not statistically different from The Netherlands (Figure 4). The three most interdisciplinary countries for this theme (Canada, The Netherlands, Great Britain) are also the three countries most involved in the theme. This suggests a correlation between the specialisation in this theme and the disciplinary integration of these studies.

Discussion

The first results about the country profiles on the six final research themes suggest that the delineation method is relevant to select an appropriate corpus of publications representing the research on food security and food system sustainability. However, some methodological issues remain insufficiently resolved. The query selection is a trial and error process with examining samples of retrieved documents to further improve the queries. The part of arbitrary decision of this step is reasonably corrected by the cleaning process.

The cleaning process itself, based on a topic model, is dependent on various parameters. The number of topics is the first issue and this number is related with the desired degree of precision of the process. With a lazy choice (an average of 1,000 documents per topic), we had to remove a smaller off-domain topic at the second step. The second issue is that the fitting algorithm has a part of randomness and with another algorithm or just with another seed, topics may be different. As main topics are generally found in each fit with the same core terms, small topics may be arranged differently. However, the cleaning step is finally defined by the selected counter-queries and rejected categories. Though they have been

suggested by topics, they are to be validated *per se* and this is not too much time consuming for experts.

These points are to be taken into consideration but, in our experience, they are manageable and topic modelling appears to be an interesting method to assist and correct a delineation task in a context of imprecise queries.

Two other issues have not been considered in this work and should be solved before the method is used in an assessment process. First, the cleaning process makes it possible to control the precision of the delineation process, but there is no control of the recall so far. Second, the choice of the database has an impact on the domain delineation (Rafols, Ciarli & Chavarro, 2015). In the present case, it is expected that other relevant publications are to be found in more comprehensive databases, particularly regarding developing countries, as CABI or in databases with a better coverage of open access publications.

The second part of this paper intended to reveal a structure of the corpus and to compare various actors. The question of the efficiency of topic models for such a task must be raised. The complexity of topic models is balanced by its flexibility and the ability to unveil two different relationships between topics and documents. However, the variability of the output of a fitting algorithm is problematic as it has an impact on document assignment to themes and therefore on quantitative indicators of actors. Thus, quantifying this (in)stability is desirable (Hecking & Leydesdorff, 2018). As a partial answer to this issue, we fitted 10 times a 20-topic model with different seeds and calculated average country profiles to possibly challenge the results. For these 10 runs, the results for 4 themes were confirmed but, for the two themes Healthy food (Th4) and Young people diet (Th6), there was too much variability to confirm our findings. Therefore more is to be done to stabilize the themes. A solution could be to use a larger number of topics to recover the same 6 themes with refined frontiers and hopefully less variability. Such improvement is still to be experienced and its success would help to validate the findings about country profiles and interdisciplinarity indicators.

Conclusion

Despite these shortcomings, this work suggests that using topic modelling is a flexible and efficient method to answer the issue of delineating a research domain defined by large policy objectives associated to a complex societal challenge. It allows to reveal unexpected scientific approaches as well as multi- and interdisciplinary research work.

For the second purpose of revealing the structure of the domain and deriving indicators to compare various actors, topic models provide interesting analyses but with a serious cost. They allow to cross information on vocabulary similarity and on topic combination in documents. But more work is necessary to ensure the stability of the indicators derived from topics. Such a control is time consuming and, for the moment, no general method is available to achieve this control. This work therefore suggests that topic modelling is an interesting but costly method to provide a quantitative analysis of a domain.

Acknowledgements

This work is inspired by an on-going project of the OST department of Hcéres for the French Ministry of Higher Education and Research with contributions of experts in the field. Their opinions were useful to consolidate the delineation method. However, in this paper, designed as a methodological contribution, the scope has been freely restricted by the authors. Consequently, the views expressed in this publication, as well as the information included in it, do not reflect the opinion or position of the Hcéres or of the Ministry and in no way commit either of them.

References

Blei, D. M., Nag, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of machine learning research 3, 993-1022

Blei, D., M. (2012). Probabilistic topic models. Communications of the ACM 55 (4) 77-83. doi:10.1145/ 2133806.2133826

Cassi, L., Lahatte, A., Rafols, I., Sautier, P. & de Turckheim, É. (2017) Improving fitness: Mapping research priorities against societal needs on obesity. Journal of Informetrics 11 (4) 1095--1113 https://doi.org/10.1016/j.joi.2017.09.010

Di Maggio, P., Nag, M., & Blei, D. (2013) Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of the U.S. governments arts funding. *Poetics*, *41*, 570-606. doi:10.1016/j.poetic.2013.08.004

Hecking, T. & Leydesdorff, L. (2018) Topic Modelling of Empirical Text Corpora:Validity, Reliability, and Reproducibility in Comparison to Semantic Maps http://export.arxiv.org/pdf/1806.01045

HLPE. 2017. Nutrition and food systems. A report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security, Rome. http://www.fao.org/3/a-i7846e.pdf # 12

Klavans, R., & Boyack, K. W. (2014). Mapping altruism. *Journal of Informetrics*, 8(2), 431-447. doi/10.1016/j.joi.2014.02.002

Ledford, H. (2015). How to solve the world's biggest problems. *Nature* 525, 308-311

MESRI, Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation. (2014). SNR, Bilan de l'atelier 5 : Sécurité alimentaire et défi démographique http://www.enseignementsup-recherche.gouv.fr/cid78802/strategie-nationale-recherche-bilan-des-travaux-des-ateliers.html

Milanez, D.H., Noyons, E. & de Faria, L.I.L. (2016). A delineating procedure to retrieve relevant publication data in research areas: the case of nanocellulose. Scientometrics 107: 627. doi:10.1007/s11192-016-1922-5

Molas-Gallart, J., Rafols, I. & Tang, P. (2014) On the relationship between interdisciplinarity and impact: different modalities of interdisciplinarity lead to different types of impact. *Journal of Science Policy and Research Management* 29 (2), 69-89.https://arxiv.org/pdf/1412.6684.pdf

OECD (2017).Obesity update http://www.oecd.org/els/health-systems/Obesity-Update-2017.pdf

Porter, A.L., Cohen, A.S., Roessner, D.J., & Perreault, M. Measuring researcher interdisciplinarity. Scientometrics. 2007;72(1):117–147.

Rafols, I., Ciarli, T., & Chavarro, D. (2015) *Under-reporting research relevant to local needs in the global south. Database biases in the representation of knowledge on rice.* Available at http://digital.csic.es/bitstream/10261/132530/1/knowledgerice.pdf

Raimbault, B. Cointet,J-P & Joly, P-B (2016) Mapping the Emergence of Synthetic Biology. PLoS ONE 11(9): e0161522. https://doi.org/10.1371/journal.pone.0161522

Sievert, C., & Shirley., K., E. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* 63-70. Available at https://CRAN.R-project.org/package=LDAvis

Wallace, M. L., & Rafols, I. (2015). Research portfolios in science policy: moving from financial returns to societal benefits. *Minerva* 53 (2), 89-115. doi:10.1007/s11024-015-9271-8