Cosmin HOLEAB, PhD
E-mail: cosmin.holeab@gmail.com
UNESCO Chair on Science and Innovation Policies, SNSPA
Professor Mihai PĂUNICĂ, PhD
E-mail: mihai.paunica@cig.ase.ro
The Bucharest University of Economic Studies
Professor Adrian CURAJ, PhD
E-mail: adrian.curaj@gmail.com
POLITEHNICA University of Bucharest and UNESCO Chair on Science and Innovation Policies, SNSPA

A COMPLEX METHOD OF SEMANTIC BIBLIOMETRICS FOR REVEALING CONCEPTUAL PROFILES AND TRENDS IN SCIENTIFIC LITERATURE. THE CASE OF FUTURE-ORIENTED TECHNOLOGY ANALYSIS (FTA) SCIENCE

Abstract. *The aim of this paper is to present a complex bibliometric method based on blending semantic and network analysis that enables the combined operation of complex parametric and non-parametric models, such as structural and loose semantic algorithms together with mathematical and statistical algorithms for dynamic visualization of data.*

Subsequently, the results of the analysis aim at substantiating the current profile and trends of the academic discipline of Future-oriented Technology Analysis (FTA) – based on the special issues` publications of five scientific journals published after four FTA international conferences (2004 – 2011). As such, the paper will contribute to enabling further scientific dialogue on FTA and moreover enhancing the big picture of FTA research for a better understanding of current approaches and future prospects.

We elaborate on the analytical relevance of 'classic' bibliometrics (word counting) and semantic analysis focusing on methodological operationalization as we endeavor to expand the current investigative focus and broaden the dialogue on future FTA research and innovative scientometrics.

Keywords*: Semantic analysis, Network analysis, Big Data visualization, Bibliometrics, Scientific literature, Future-oriented Technology Analysis (FTA), Conceptual structures, Scientific trends*

JEL Classification: O33

1. Introduction

On FTA and research trends:

The term "Future-oriented Technology Analysis"(FTA) was first used in 2004 in the context of a *foresight* seminar organized by the Institute for Prospective Technological Studies (IPTS); at that moment, it was defining a so-called 'umbrella' covering a number of different methods of technology analysis in the fields of technology foresight, technology forecasting and technology assessment [1]. FTA was later described as a scientific discipline aiming to explain a broad range of future-looking activities involving foresight, forecasting, futures, and technology assessment among the others [2].

Over time, FTA started to be treated as a kind of *future management* concept and FTA research began to develop in two parallel trends (when looking at the scientific focus and applicability):

(1) *technological* and
(2) *decision-making*[3].

The *technological* approach is concerned with methods and tools for analyzing, assessing and predicting the development of technologies [4] as well as managing their future [5]; the second approach addresses FTA as a tool for policy-making [6][7][8].

At the same time, in terms of structural views, three major trends have been recently referred to by foresight scholars:

(3) firstly, FTA (and foresight) research focus has shifted from intra-organizational planning and forecasting in science and technology to open and inter-organizational 'strategizing' with inclusion of external stakeholders opinions on alternative futures;
(4) secondly, FTA research is taking a more systemic approach (both in practice and in understanding of innovation);
(5) also, FTA research has (lately) started to focus more on the regional or sectorial dimensions of innovation/ policy [9].

2. Method and data

2.1. Research methodology and tools

In order to perform a semantic analysis of FTA scientific publications, we used the open-source semantic software Tropes [10].

The theory that the software is based on is integrating two distinct work models: *propositional discourse analysis* [11]and *predicative propositional analysis* [12].

This analytic approach derives from the need to identify a *cognitive unit* for primary information processing and a *syntactic unit* to allow 'clipping' the discourse. The minimal unit that meets both requirements is the *sentence*. Moving forward to content analysis, it can be observed that the sentence is exposing microworlds more or less articulated among each other, more or less completed.

The refrom *actors*(*actants* and *acted on*) appear, highlighted by acts (predicates) as being embodied by the argumentative strategies and the constraints that are constitutional to the linguistic system. The number of *references* (microworlds) that evolve around a topic depends on a number of central objects, referred as 'nodal references', which are the structural elements of the given semantic universe. Finally, a logical model for construction of discourse is needed in order to mark out the *cause-consequence*; this model facilitates the identification of the node that is generating the references, which plays an essential role in shaping and analyzing the discourse [11][13].

The main outputs of Tropes – relevant for the bibliometric analysis presented in this paper – consist in matrixes of references (central topics containing default or/and user-defined keywords) with directional cause-consequence relations (depending on the relative positions of each reference in the sentence).

These outputs can be further processed with network analysis tools, as we show in this paper. For that matter, we used another open-source software (with the aim to facilitate further replications of our method) for network analysis – Gephi.

The basic principle of the two software blending is that the network analysis software interprets the matrixes of *references* and their *relations* (or *semantic ontology*, Tropes output) as *nodes* and *edges*. The substantial benefit is that network analysis tools enable employment of both Big Data visualization algorithms suitable for semantic analysis and complex algorithms for further parametrization of data. An example is the multi-level *modularity class* aggregation for decomposition of networks and identification of communities [14] – that is, in semantic terms – for identifying the modular subnetworks of references that are the discursive episodes.

When talking about data clustering, there are a few basic concepts which need to be discussed, such as distance metric, similarity matrix and clustering algorithms. Conventional clustering methods mainly consist of two parts: the construction of a similarity matrix between documents and the construction of clusters using a clustering algorithm.

2.2. Literature analysis:

The semantic analysis was performed on a corpus of 87 scientific articles published in special issues of five international foresight journals after the FTA conferences in 2004, 2006, 2008 and 2011 (Table 1). Working with Scopus (the largest database of academic journal articles) we retrieved both the full-text and the abstracts of the articles (enabling also an in-depth comparison of conceptual frameworks revealed by abstracts and full-texts).

Table 1. Methodological data: Number of articles selected from foresight journals

JOURNAL NAME	YEAR	NO. OF PAPERS
Foresight	2012	6
Foresight	2013	5
Futures	2011	7
Futures	2014	6
Science and Public Policy (SPP)	2010	7
Science and Public Policy	2012	12
Technological Forecasting and Social Change (TFSC)	2005	8
Technological Forecasting and Social Change	2008	7
Technological Forecasting and Social Change	2009	13
Technological Forecasting and Social Change	2013	9
Technology Analysis and Strategic Management (TASM)	2012	7
	Total	87

The manifold process of the semantic analysis blended with network analysis consisted of:

(1) (for analysis of full-texts) a preparatory phase in which we have formatted and cleaned the textual information by removing the redundant abstract, footnotes, bibliography, and acknowledgement of each paper, that could result in inaccurate semantic ontologies and statistics;

(2) the processing of the Scopus indexed keywords of the selected scientific articles, as the input for building the semantic dictionary needed for performing the analysis;

(3) building and refining a semantic dictionary – the analysis scenario (relevant to the analytical purpose of our paper and also reflecting the actual semantic content of the 87 articles) with a four-levels tree structure containing 137 *semantic references* i.e. the main topics / concepts, with 1.253 corresponding *keywords* (performed with Tropes software);

(4) building the visual representations (Figure 3, 4, 5A, and 5B) of the semantic networks of the textual corpus by employing force-directed graph layout algorithms on the semantic ontologies per years and per journal exported from Tropes, i.e. ForceAtlas2 with LinLog[1] and low-scaling

[1]a – r = 1 in LinLog, meaning that visual densities in the graph denote structural densities, that is when the attraction force of the nodes depends less on distance, and the repulsion force depends more**Invalid source specified.**.

'Dissuade Hubs'[2] modes distribution showing modularity classes[3] (performed with Gephi software).

3. Results and discussions

3.1. On the proposed method

The common practice in bibliometric (and related semantic) analyses is taking into consideration rather general information on or from scientific papers, such as keywords and/ or abstract texts; recent bibliometric research on FTA academic trends carried by Andersen & Alkærsig is based on "analysis of keywords and abstracts"[9, p. 6].

A first methodological remark on that practice is that a semantic analysis – different from word counters – lacks in relevance when not looking at the full-texts of selected publications. That is because semantic analysis allowsin-depth mining of a sampled textual corpus and can reveal conceptual frameworks (the most representative structures and substructures of discourse), thematic structures of discourse, causal relations between concepts, and other nuances intrinsic to the linguistic system.

As an example of data sampling of keywords and abstracts vs. full-texts, Figure 1 shows the differences between abstracts (a) and full-texts (b) in volume (occurrence and frequency) of keywords describing the reference "foresight". It can be seen that the selected 87 papers are referring to "foresight" 205 times in their abstracts vs. 4230 times in the full-text while full-texts introduce at least 2 times more concepts/ keywords related to foresight. That is to say we are only discussing here a first layer of general statistics (occurrence and frequency).

[2]The "Dissuade Hubs" mode affects the shape of the graph by dividing the attraction force of each node by its degree plus one for nodes it points to, meaning it grants authorities (nodes with a high indegree) a more central position than hubs (nodes with a high outdegree)**Invalid source specified.**.

[3]The modularity classes (represented with colors in Figure 3, 4, 5A, and 5B) describe semantic subnetworks consisting in nodes/ references with strong and complex (semantic) relations, while between nodes in different modules there are sparse connections.

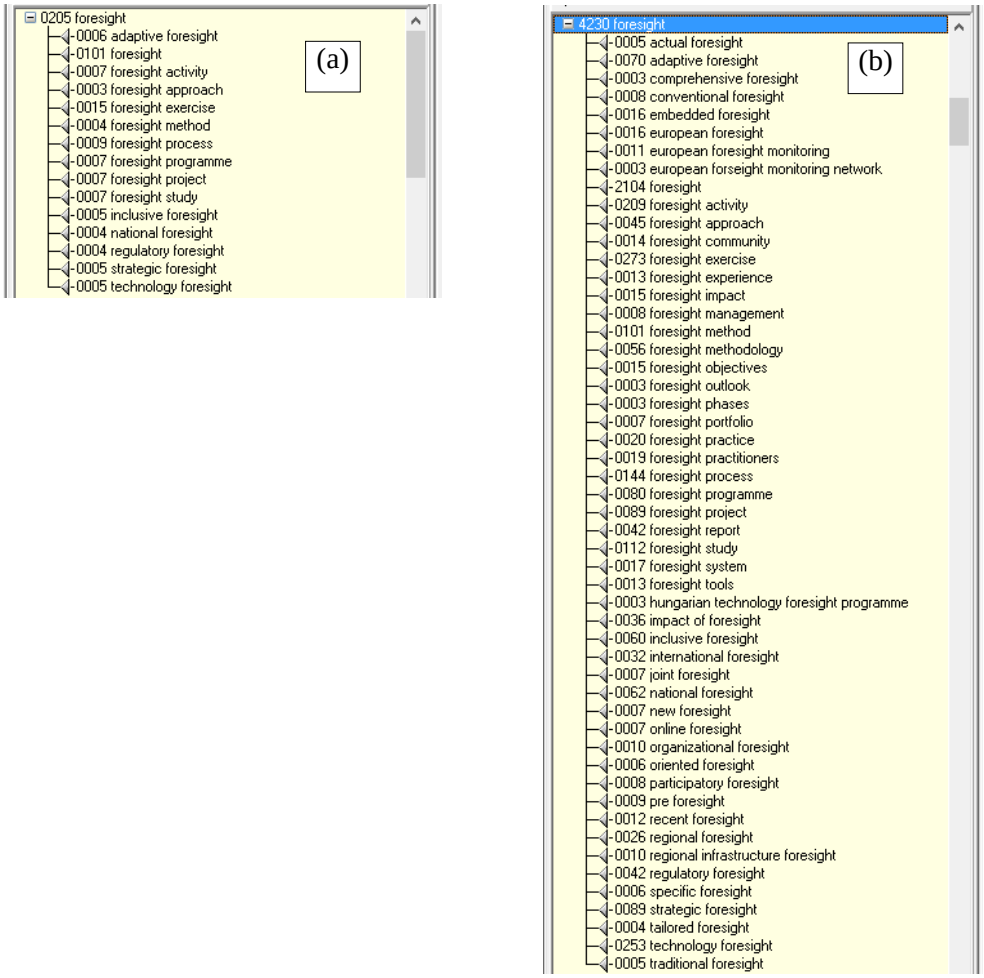


Figure 1. Example of the tree-structure of the reference"foresight" and its keywords, (a) in abstracts, (b) in full-texts (of the same semantic dictionary – in Tropes)

Going one step further with the semantic analysis there is even more conclusive that textual statistics (word counting) alone are significantly less relevant for describing research trends reflected by scientific publications than semantic analysis. We are talking here about a fundamental function of semantic analysis, i.e. showing the relevant relations between references. For that matter, Tropes is addressing the cause-effect by filtering the relevant co-occurrences and displaying only non-aleatory relations – considering that a single co-occurrence of

two references in the same sentence is likely an aleatory fact and, in any terms, has no statistical relevance.

In the example considering the abstracts of the papers published in 2010, although the reference counter (word cloud) (Figure 2) shows a somewhat complex conceptual 'landscape' with rather high occurrences of words (for instance "foresight", with 13 occurrences), the semantic analysis of those abstracts show a different conceptual framework, also with a rather low statistical relevance for contextual assumptions(Figure 3); moreover, the reference "foresight", with the highest occurrence in the word cloud, has been excluded in the semantic analysis because it has no statistically relevant connection with any of the other references (those displayed in the word cloud).

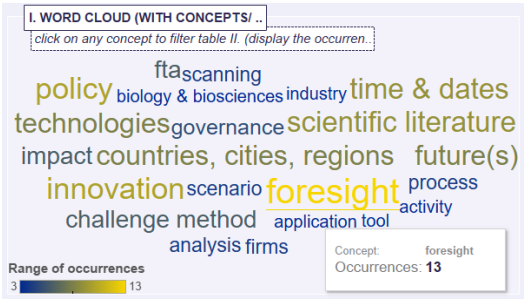


Figure 2. The word cloud of references (from the working dictionary) in the abstracts of the 7 papers published in 2010 in SPP

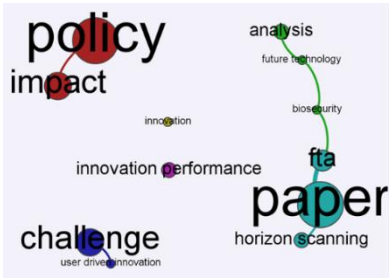
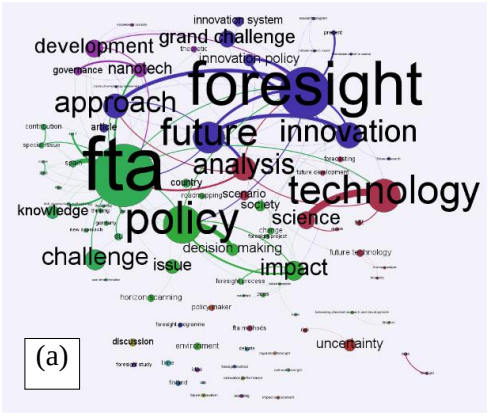


Figure 3.The semantic network of the abstracts of the 7 papers published in 2010 in SPP[4,5]

[4]The nodes in the graph (built with Gephi) are both the references and their keywords defined in Tropes, contained in the semantic dictionary; their size represent the use

Finally, going back to the discussion on analysis of abstracts vs. full-texts, the full-fledged semantic analysis reveals the complexity of the conceptual networks of the 87 scientific articles (Figure 4).



(a)

(b)

frequency of the term, and the lines represent the in / out relations (and their frequency) of each two terms based on their relative positions in each sentence.

[5]The two nodes that have no conections are present in the network because they represent self-references (co-occurrances in the same sentence within the analysed texts).

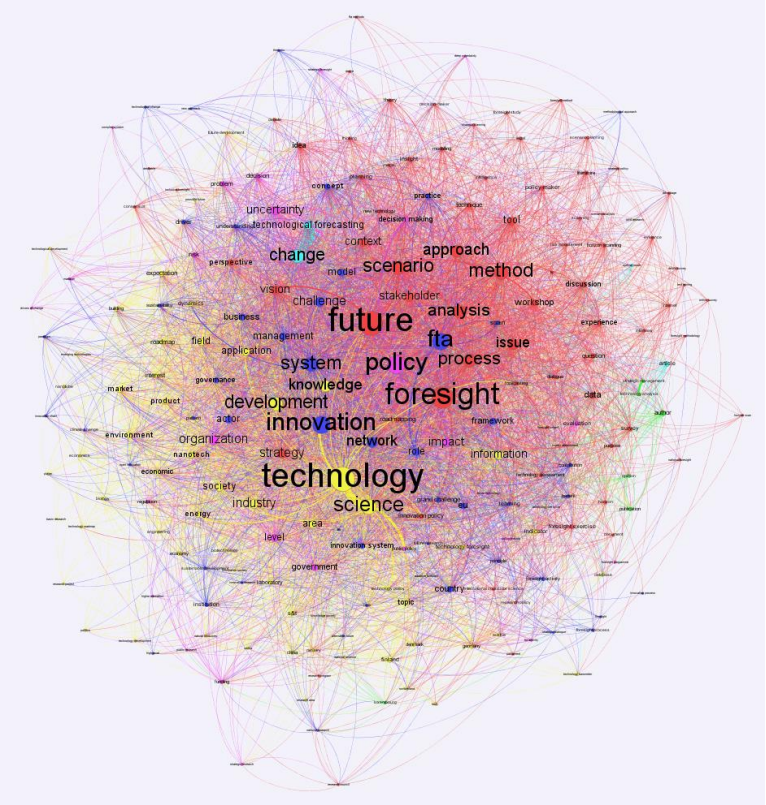


Figure 4.The semantic network of the 87 (a) abstracts and (b) full-texts

3.2. On FTA research trends

While the differences in Figure 4 (a) and (b) are self-evident, what the analyzed abstracts are mainly showing is:

(1) a complex web of relations from and to “FTA”,
(2) a focus on “FTA”, on “policy” and on “technology” (aside the general focus on “foresight”),
(3) an existing connection between “FTA” and “policy” – although not a strong one,
(4) between “foresight” and “innovation” but also “innovation policy”,
(5) that “FTA” is mainly connected to “knowledge” (that is scientific knowledge),
(6) and that “technology” is strongly related to “science”.

To what is concerning the first two trends described in the *Introduction* of our paper (*technological* and *decision-making* focus), the semantic analysis of the 87 abstracts (Figure 4 (a)) is showing a relatively distinctive approach of policy and of technology, although the rather small thematic subnetwork of "technology" and the focus on "foresight" (and its whole subnetwork represented in violet color) are hindering a clear conclusion.

Nevertheless, the semantic network of the full-texts (Figure 4 (b)) reveals – visible by colored modularity classes:

(1) a distinctive focus on "FTA" and "technology" at "systemic" level (blue color) – consistent with the second trend identified by Andersen & Alkærsig[6]: FTA research is taking a more systemic approach [9],
(2) a focus on "policy" and "impact" at "organizational level" but also at "governmental level"(pink color),
(3) and a clear and complex focus on "technology" in relation with "science" "knowledge", "development", "industry" and "society" (with yellow color),
(4) while "foresight" – as consistently addressed by *futures* literature over the last 10 to 20 years – relates to "future" and all the "methodological" aspects of various "approaches", to "scenario", "vision" and "strategy" (red color).

When looking into detail at the temporal evolution of the semantic networks over the period 2005-2014 (Figure 5A and Figure 5B), one can see that the scientific interest in FTA – as it is reflected by the frequency of using the term[7]–has been constantly increasing between 2005 and 2012.The frequencies have increased at a steady pace from 2005 to 2010 (see also Figure 6), when in 2011 they have skyrocketed – in terms of relative occurrences in the total corpus of references of articles published in that year; in fact, the absolute frequency have almost tripled in the following year, 2012 (when "FTA" recorded the highest frequency of all the references of the textual corpus), but the size of the "FTA" node in the right 2012 in Figure 5B is smaller than the one in the left 2011 because the substantially higher number of published articles led to a general increase in references` frequencies. It is also obvious that the focus on FTA has lately 'faded' to some extent (in 2013 and 2014). All in all, this frequency dynamic secures an upward trend in scientific interest for FTA (see Figure 6).

Aside the relative frequencies of references and the trends they disclose, the semantic networks also outline structural trends, such as the increasingly consistent relation of FTA to *policy* (starting with 2010 – Figure 5A and 5B), a

[6]Marked as number (4) in the *Introduction* section of this paper.
[7]The reference have been marked out with a yellow circle in Figure 5A and Figure 5B. The circle size indicate the yearly relative frequency of the reference (in relation to the maximum frequency recorded in that respective year).

systemic focus of FTA research in 2012 and 2014 (with violet color in Figure 5B), and the recent interest for *organizations* and *strategic management* (in 2014) – to a certain extent consistent with the trends identified by Andersen & Alkærsig [9].

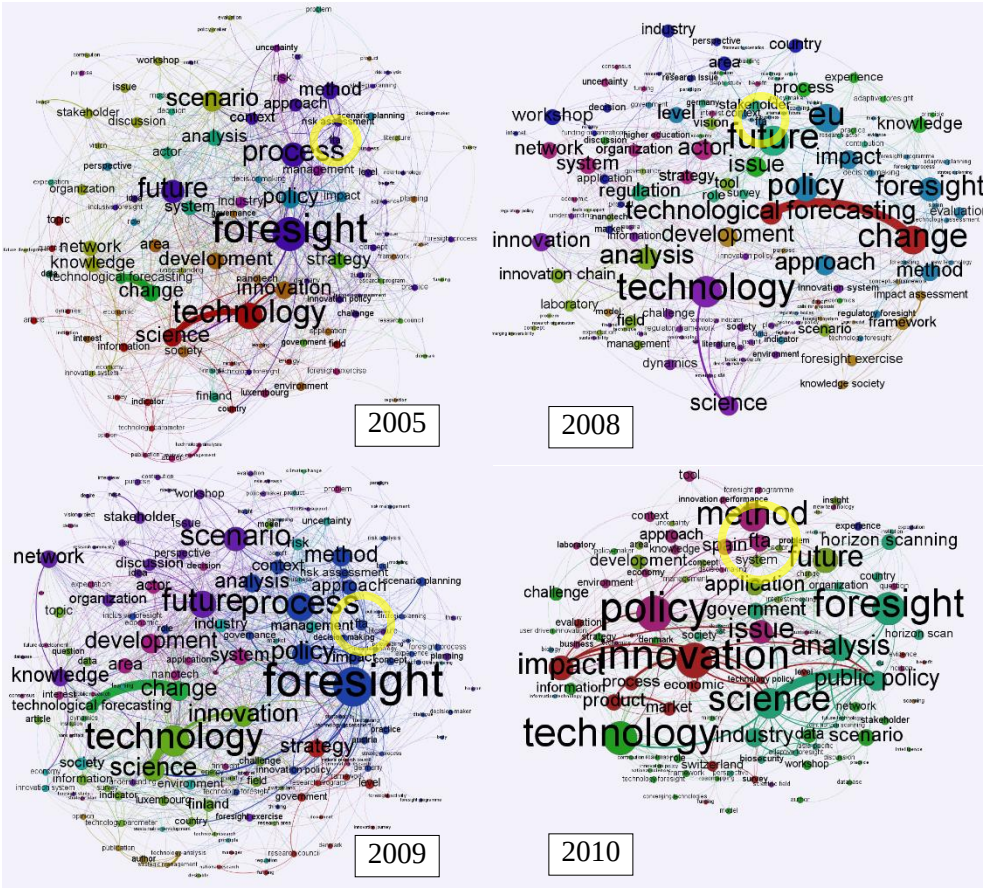


Figure 5A.The semantic networks of the 87 scientific articles distributed over time by publication date: 2005-2010

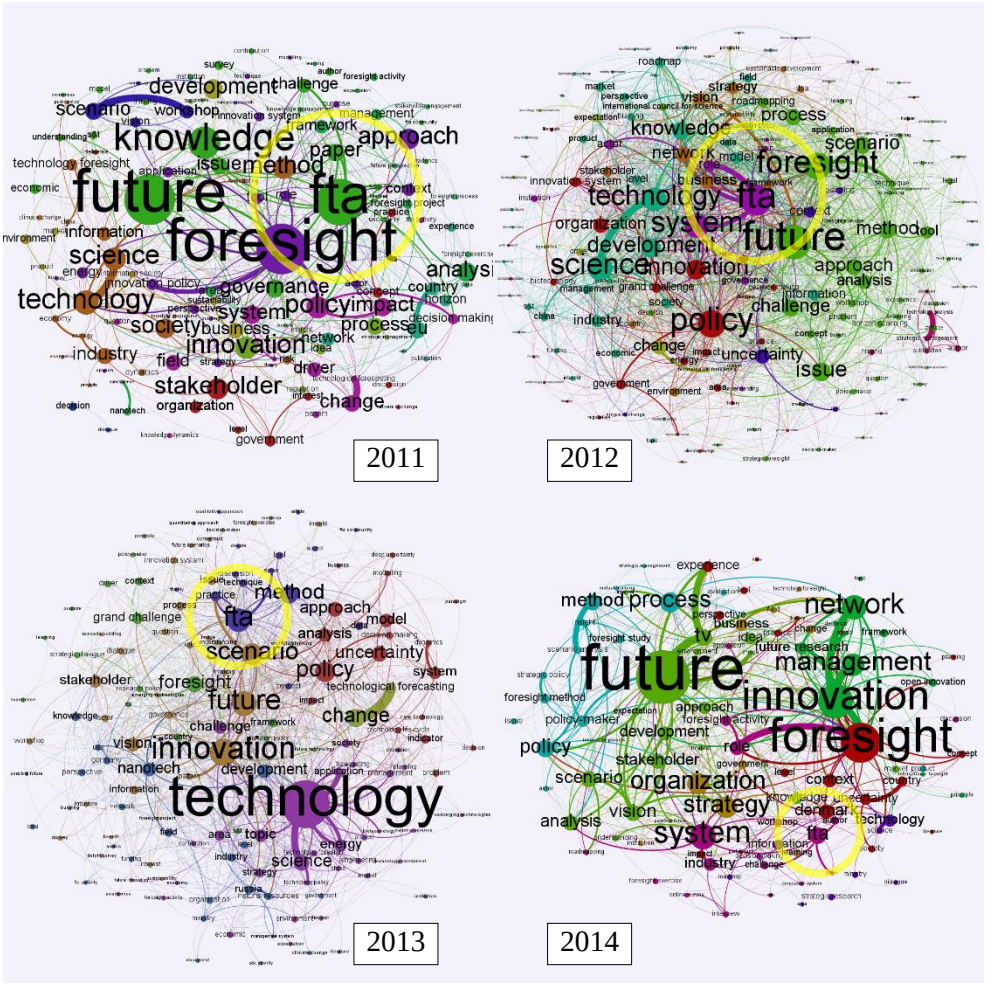


Figure 5B.The semantic networks of the 87 scientific articles distributed over time by publication date: 2011-2014

Indeed, as shown also in Figure 6[8], there is a clear trend of including *stakeholders* – reflected by the whole corpus of articles[9]; there is an obvious trend in FTA research for addressing *policy* and *governance*; there is also a clear increase in the interest for *innovation*[10]. When looking at the trendline associated with "technologies", one might say that this stable focus of FTA research in the

[8]Market with the dashed red lines.
[9]In line with the first trend identified by Andersen & Alkærsig, marked as number (4) in the *Introduction* section of this paper [9].
[10]Finally, substantiating the third trend marked as number (5) in the *Introduction*.

context of increasing interest for *innovation* could lead eventually to a shift in FTA research towards the *systemic dimension of innovation* (obviously, not entirely away from *technology*).

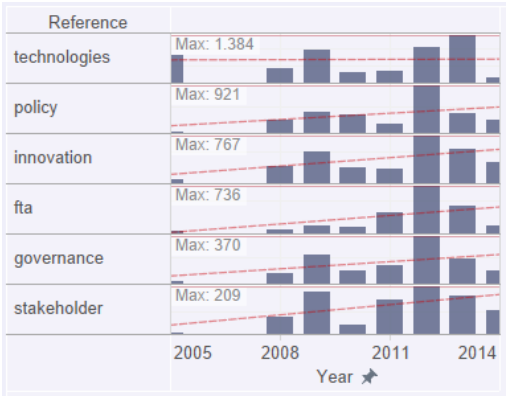


Figure 6.Evolution of frequencies over time – selected references

4. Conclusions

The proposed bibliometric method of blended semantic and network analysis is a valuable addition to the methodological repertoire of scientometrics. It draws on 'classic' textual statistics and employs parametric algorithms of semantic analysis, adding the power of Big Data visualizations. Moreover, given the complexity and 'accuracy' of our methodology, we argue that it (together with this paper) has the potential to further improve and focus the scientometrics research.

A notable advantage of our research consists in the (interoperational) use of open-source tools that will enable future replications of our methodology.

There are also certain limitations of our method that have to be considered. The majority of them fall under the issue of scale – i.e. the semantic networks cannot be deconstructed back to sentences as Big Data does not account for case details –,and can be overcome with cautious deductions.

The semantic analysis of the scientific literature presented in this paper not only substantiates the trends in FTA research that have been recently identified by experts in futures studies, but also discloses various nuances and possible future trends, offering at the same time a bird's eye view on the conceptual construction of the scientific research in the field.

Finally, we believe that this paper expands the current investigative focus and broadens the dialogue on the rapidly growing research areas of FTA and scientometrics.

REFERENCES

[1] **C. Cagnin, M. Keenan, R. Johnston, F. Scapolo and R. Barré(2008),** ***Future-oriented Technology Analysis. Strategic Intelligence for an Innovative Economy****; Springer-Verlag;*

[2] **O. Saritas, S. Burmaoglu and A. Tabak (2014),** ***The Evolution of the Use of Foresight Methods: A Bibliometric Analysis of Global Research Output for Cutting-edge FTA Approaches,*** in: *5th International Conference on Future-Oriented Technology Analysis (FTA) – Engage today to shape tomorrow,* Brussels;

[3] **K. Halicka (2016),** ***Innovative Classification of Methods of the Future-oriented Technology Analysis****; Technologycal and Economic Development of Economy,* vol. 22, no. 4, pp. 574-597;

[4] **L. Huang, Y. Guo, J. Youties and A. Porter (2012),** ***Early Commercialization Pattern Profiling: Nano-enhanced Biosensors****,* in *Technology Management for Emerging Technologies (PICMET 2012),* Vancouver, BC;

[5] **A. Cheng, C. Chen and C. Chen (2008),** ***A Fuzzy Multiple Criteria Comparison of Technology Forecasting Methods for Predicting the New Materials Development****; Technological Forecasting and Social Change,* vol. 75, no. 1, pp. 131-141;

[6] **C. Cagnin, D. Loveridge and O. Saritas (2011),** ***FTA and Equity: New Approaches to Governance****; Futures,* vol. 43, no. 3, pp. 279-291;

[7] **L. Georghiou and J. Harper (2013),** ***Rising to the Challenges - Reflections on Future-oriented Technology Analysis****; Technological Forecasting & Social Change,* vol. 80, no. 3, pp. 386-397;

[8] **S. Marinho and C. Cagnin (2014),** ***The Roles of FTA in Improving Performance Measurement Systems to Enable Alignment between Business Strategy and Operations: Insights from three Practical Cases;*** *Futures,* vol. 59, pp. 50-61;

[9] **P. Andersen and L. Alkærsig (2014),** ***Profile and Trends in FTA Approaches: A Bibliometric Analysis of Special Issues of International Journals from four FTA,*** in: *5th International Conference on Future-Oriented Technology Analysis (FTA) - Engage Today to Shape Tomorrow,* Brussels;

[10] **ACETIC; CYBERLEX, "Semantic knowledge - Tropes," ACETIC; CYBERLEX, 2002.** [Online]. Available: https://www.semantic-knowledge.com/tropes.htm. [Accessed 1 February 2016].

[11] **R. Ghiglione, C. Kekenbosch and A. Landré(1995),** ***L'analyse cognitivo-discursive,*** Grenoble: Presses Universitaires de Grenoble, 1 pp. 1-2;

[12] **W. Kintsch and T. Van Dijk (1978),** ***Toward a Model of Text Comprehension and Production****; Psychological Review,* vol. 85, no. (5), pp. 363-394;

[13] **D. Caragea and A. Curaj(2013),** ***Analiza automată a discursului***; Bucharest: *Academia Româna Publishing;*

[14] **V. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre (2008),** ***Fast Unfolding of Communities in Large Networks*****;** *Journal of Statistical Mechanics: Theory and Experiment*, vol. 10.