

Proposal of a Method for Review and Content Analysis of Literature: The Case of Industrial Energy Efficiency

Marcos Gonçalves Perroni (doutorado.perroni@gmail.com)

Pontifical Catholic University of Paraná

Sergio Eduardo Gouvea da Costa (s.gouvea@pucpr.br)

Pontifical Catholic University of Parana

Federal University of Technology – Parana

Edson Pinheiro de Lima (e.pinheiro@pucpr.br)

Pontifical Catholic University of Parana

Federal University of Technology – Parana

Wesley Vieira da Silva (wesley.vieira@pucpr.br)

Pontifical Catholic University of Parana

Dilmeire Sant Anna Ramos Vosgerau (dilmeirerv@gmail.com)

Pontifical Catholic University of Parana

Abstract

This paper aims to develop a method integrating techniques: Content Analysis, Bibliometric Analysis, Social Network Analysis and Concept Mapping in a Systematic Literature Review Environment. The application of the method will occur in the area of Industrial Energy Efficiency. This paper hopes to contribute with the field of Research Methodology.

Key-words – Systematic Literature Review, Content Analysis, Industrial Energy Efficiency

Introduction

It is common for a researcher to identify the state of the art in relation a specific research topic. The survey involves mapping the existing scientific knowledge in order to identify previous studies. The theoretical framework is used by the researcher as a map for evaluation of existing intellectual territory. This mapping gives conditions for the researcher to propose original research initiatives that advance knowledge in this specific area (Aquino, et al. 2008; Luft and Shilds, 2003).

With the advance of information technology, the mapping task on the one hand, it has become easier due to the formation of large scientific document databases, but then there is the difficulty of the choice appropriate work to the goal of the research, due the a lot of work in

some areas. Aquino et al. (2008), highlights two common mistakes in this task which are the omission errors (OE) and inclusion errors (IE). An error of omission occurs when a relevant work is left out of the map (theoretical framework). An inclusion error occurs when an irrelevant work is included in the map (theoretical framework).

The solution to the above problems has been the application of a systematic review, which is a structured means to identify all relevant documents for a specific issue or research topic (Armitage and Allen, 2008; Kitchenham, 2004; Lopes and Fracolli, 2008, Plytiuk, et al. 2014).

The focus of this paper is to propose a method for application of a systematic review and content analysis of the literature that seeks to overcome the problems mentioned, integrating techniques such as Content Analysis (Bardin, 2011), Bibliometric Analysis (Du et al. 2013; Pilkiton and Meredith, 2009; Tsay, 2008), Social Network Analysis (Breiger, 1974; Hansen, et al. 2011; Prell, 2012), Concept Maps (Kane and Trochim, 2007; McLinden, 2013; Trochim, 1989) and Text Mining (Zhao, 2012). Will be demonstrated an application in the case of Industrial Energy Efficiency literature.

Literature Review

According to Aquino et al. (2008), the theoretical framework of a study should be evaluated in a similar manner that the empirical part of the scientific work, giving legitimacy to the operation of scientific research. Taylor and Taylor (2009) identify two ways to evaluate a research field: Using bibliometric methods, as the analysis of citations and co-citations, or applying content analysis.

In the energy field several studies have been conducted applying bibliometric method. The bibliometric techniques can be defined as "*statistical method of bibliography counting to evaluate and quantify the growth of literature for a particular subject*" (Tsay, 2008). Tsay (2008) explores the literature on hydrogen energy in the period 1965-2005, identifying a growth rate of 18%, the most productive countries were USA, Japan and China, the main center was Chinese Academy of Sciences. Fernández et al. (2012) investigates the scientific production of renewable energy from 2003-2008. In this study the largest number of publications comes from USA, China and UK, with the main journals, Energy Conversion and Management, Solar Energy Materials and Solar Cells, Renewable Energy, Energy, Earth, Energy and Buildings and Journal of Cleanner Production. Xu and Yaoyang (2013) map the biofuels field, among the most productive countries highlight USA, China, UK, Germany, Sweden and Brazil. Du et al. (2014) analyzed the literature on solar energy identifying USA, China, India, Japan and Germany as the most productive countries and Renewable Energy, Solar Energy, Energy Conversion and Management as the most productive journals, and last, Du et al. (2013) present the first bibliometric analysis of the literature on energy efficiency, identifying USA, China, UK, Canada and Japan as the most productive countries, with the most productive journals Energy Policy, Energy and Buildings, Energy, Energy Conversion and Management.

The content analysis according with Bardin (2011) is a set of techniques of communication analysis in order to obtain by systematic procedures and description objectives, the content of the indicator messages, allowing the inference of knowledge related to the conditions of production and reception of these messages. Some of the techniques suggested by Bardin (2011) were applied by Taylor and Taylor (2009) in the operations management field.

Method Development

The vision of this work, is agreeing with Aquino (2008), that the literature review needs to be addressed in a scientific way. To do this, the method of Figure 1 is proposed. This method is modular (specific steps), with the dorsal spine a structure based on content analysis, integrating techniques such as Bibliometric Analysis, Social Network Analysis, Concept Maps and Statistics, forming a systematic literature review

The first step occurs with the formation of the *corpus*, which is according Bardin (2011) all the documents taken into account to be subjected to analytical procedures. The constitution of the *corpus* implies choices, selections and rules. Initially, the research topic is mapped, thus enabling the construction of an initial theoretical framework. With the initial theoretical framework, the theme is delimited based on the research question (Perroni et al. 2014). Data sources are chosen and further the *corpus* or set of documents is defined.

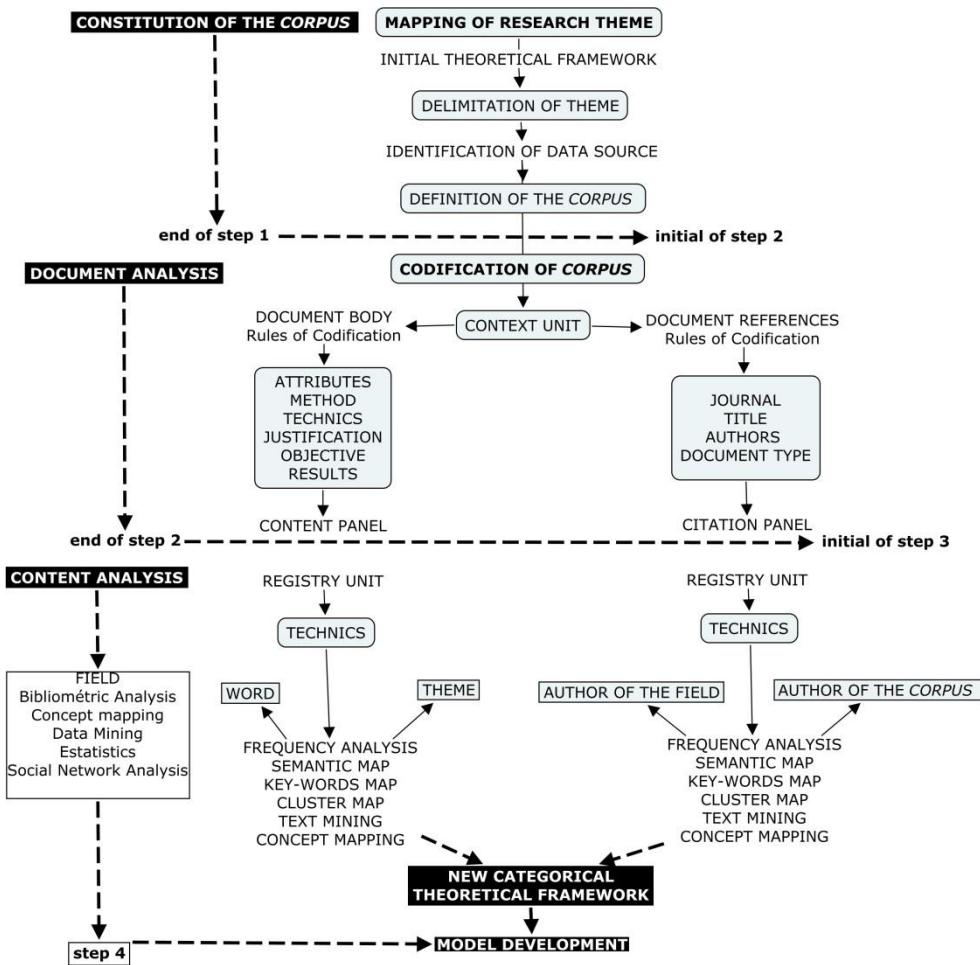


Figure 1 - Method for Review and Content Analysis of Literature

The next step is the codification of the corpus or document analysis. Based on Bardin (2011) the codification corresponds to a transformation of the raw data, allowing a representation of content or expression, susceptible of explanation by the analyst. The context unit serves as

comprehension unit to codification the subsequent registry unit. The context unit corresponds to the message segment, in which the dimensions are superior to the registry units. In this step the documents (*corpus*) are divided in two context units, the body of the document and the references of the document. The natural codification for the document body is to extract the objectives, justification, method, techniques and results, and of course some attributes such as department, institution, journal, city and country in the case of a scientific paper. The codification for the document references is the extraction of the authors, title, journal and document type. As extractions are being made, is being built a content panel for the body of the document and a citation panel for the references of the documents.

The third step is part of content analysis itself. In this step the registry units are established, in the case of content panel, the word and the theme are natural registry units, in the case of citation panel, the first author is another natural registry unit. The registry units are built in such a way that can build an incidence matrix (contingency) and through this matrix the transformation in an adjacency matrix (co-occurrences) allowing the application of various techniques: Social Network Analysis (Breiger, 1974; Hansen, et al, 2011; Prell, 2012), Concept Maps (Kane and Trochim, 2007; McLinden, 2013; Trochim, 1989) and Text Mining (Zhao, 2012). Qualitative data are transformed into quantitative data, applying the same techniques both in the analysis of the content and in the bibliometric analysis. The analysis of co-occurrences, Bardin (2011), seeks to extract of the text, the relationship between the message elements, or more exactly, is dedicated to identify the simultaneous presence of two or more elements in the same context unit. For the construction of the adjacency matrix (co-occurrences), was used a technique developed by Breiger (1974), where A is the incidence matrix and P is the relation author/author, word/word or theme/theme, according to equation 1 below.

$$P_{ij} = \sum_{k=1}^g A_{ik} A_{jk} \quad \text{rewrite } P = A(A^T) \quad (1)$$

Method Application in the Field of Industrial Energy Efficiency

The Industrial Energy Efficiency is understood as a sub-area of energy efficiency. Du et al. (2013) notes an increasing number of publications in the area of energy efficiency. This study has a narrower scope than the study made by Du et al. (2012). The perspective here is the management of energy efficiency in industrial processes. The vision here is to capture the research where energy efficiency is part of the management context in industries, especially manufacturing. Based on Bunse et al. (2011) energy management is an activity of control, monitoring and improving energy efficiency.

Constitution of the Corpus

Initially the research topic has to be delimited, was made an initial search in books, government agencies reports, International Energy Agency and scientific articles. By constructing an initial theoretical framework was found that the field contains a number of significant works (Perroni et al. 2014). It was also found that the Science Direct is the scientific database with the most journals in this area. We performed a search in the database filtered by title, abstract and keywords with the search terms: "*industrial energy efficiency*", "*industrial energy management*", "*energy efficiency manufacturing*", "*energy efficiency and manufacturing*", "*energy management and industry*", "*industrial symbiosis and energy*" and "*energy management and manufacturing*".

The search was performed during the month of May 2014, returning 574 articles, of which 178 downloaded, based on reading the abstract. The main criterion of choice at that stage was whether article addressed the issue industrial energy efficiency in management perspective.

Document Analysis

The document analysis is understood here as a previous step to content analysis, according to Bardin (2011), document analysis is a preliminary stage in the database construction, allowing move from a primary document (gross) to a secondary document (representation of the first). The result of the document analysis in this phase was the construction of two databases calls content panel and citation panel. For the constitution of panels, 104 of the 178 articles were read in full, and the information entered into Excel®. The content panel contains the data as in Figure 1 codification rules. A total of 61.202 words were extracted, with an average of 588 words per article, minimum 288 words, and maximum with 985 words. The citation panel was built based on the rules of Figure 1, but selecting the first author of each reference, the database totaled 4.466 citations where 1.255 are journal authors.

Content Analysis

The content analysis involves a set of techniques in the analysis of communications, including: Categorical, Evaluation, Enunciation, Discourse, Expression and Relations (Bardin, 2011). The results of this study will mainly categorical analysis and display the relationship using techniques described in step 3 of Figure 1. In the citation panel was concentrated in 1255 citations of scientific papers. Figure 2 shows the main countries and journals by the number of articles.

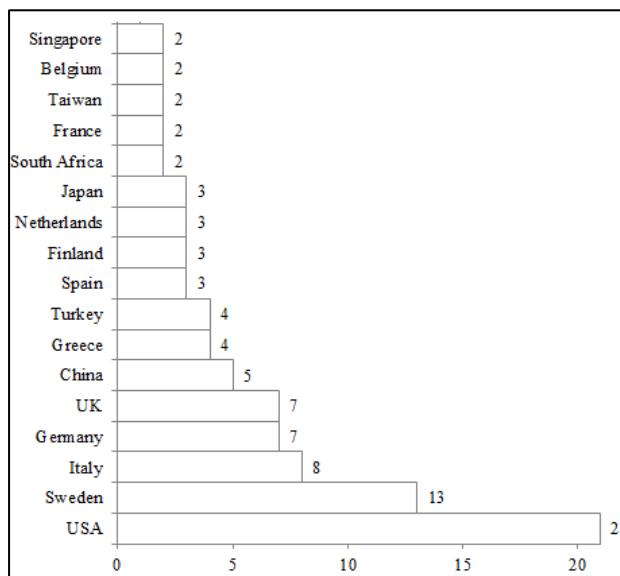


Figure 2a – Articles by Countries (91 de 104)

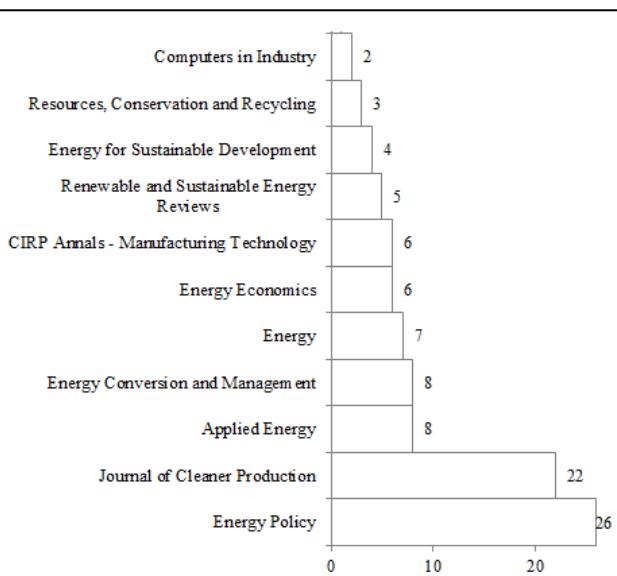


Figure 2b – Articles by Journal (97 de 104)

Figure 3 shows frequency as word cloud built in the package “*wordcloud*” of the R, either the citation panel how the content panel.

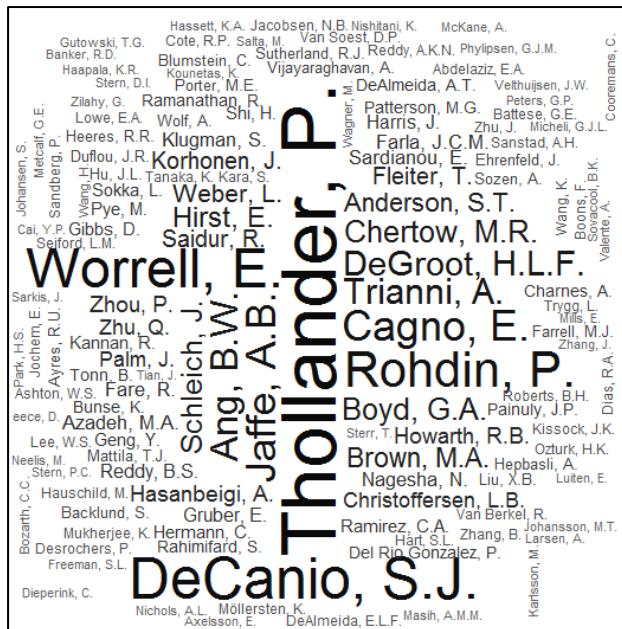


Figure 3a - Frequency of Authors

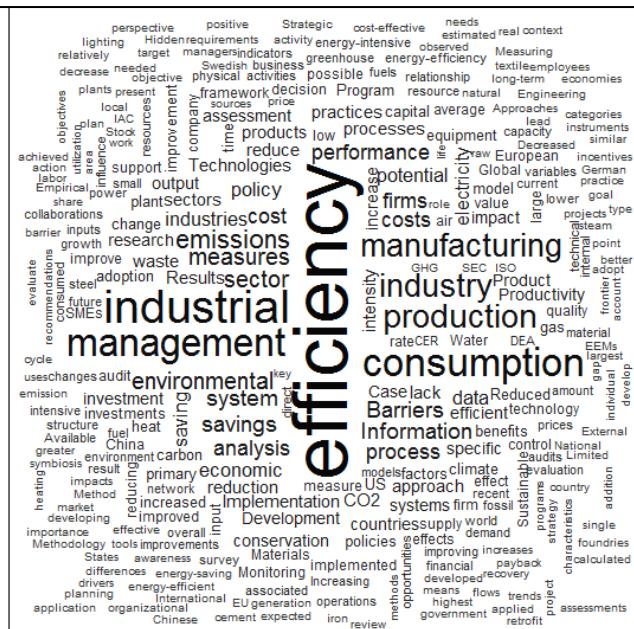


Figure 3b - Frequency of Words

Figure 3a was constructed based on incidence matrix (author-document relation) (Zhao, 2012) of the 150 most cited authors in 104 articles and Figure 3b based on incidence matrix (term-document relation) (Zhao, 2012) of the 300 most cited significant words in content panel of such articles. The word energy was excluded because is an outlier. The frequency analysis shows the most important words and authors for the context, but not shows the relationship between the component parts.

One type of analysis of relations that is becoming popular is the representation through a social network or social network analysis (SNA). Based on Prell (2012) the SNA approach involves theoretical concepts, methods and analytical techniques that aim to discover the social relations between the individual and the group, as well as its structure, as these relationships and structure influence and are influenced by social behavior, attitudes, beliefs and knowledge. McLinden (2013) has argued that the concept map (Kane and Trochim, 2007; Trochim, 1989) has a similar structure that the SNA, representing a conceptual map. These two techniques operate with the construction of a similar incidence matrix. Building a social network here is a map.

Figure 4 shows the social network of the 100 most frequent words and Figure 5 the thematic network. Figure 4 depicts the relative connection, considering that the real connection is virtually 100% of these words. To generate Figure 4 we used the software NodeXL implementing the *Harel-Koren Fast Multiscale* algorithm, the size of vertices is proportional to statistical *outdegree or indegree centrality*, the colors are proportional to statistical *eigenvector centrality*, and differentiation of edges are proportional to the opacity of *weight Edge* (Hansen et al. 2011; Prell, 2012).

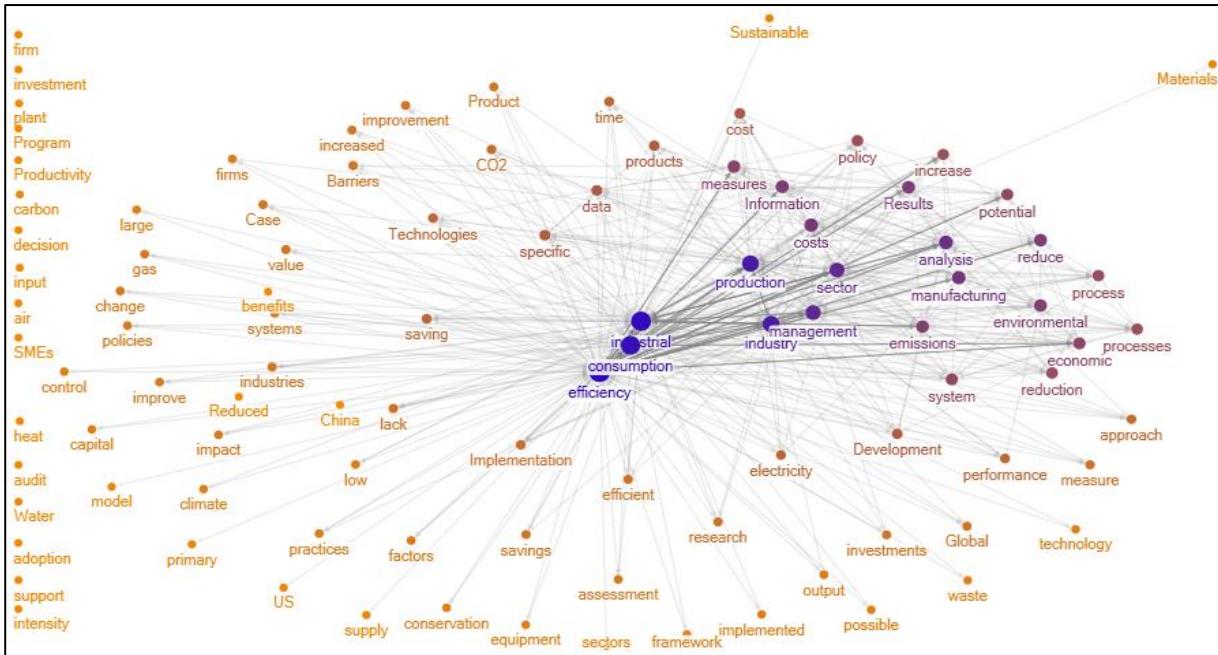


Figure 4 – Social network of the 100 most frequent words

In Figure 4 the word *energy* was excluded because it is an outlier. In the right side of the figure in blue (higher *degree and eigenvector centrality*) can be inferred that there is a concern with reducing energy consumption and pollution (*reduce, reduction*), or increase efficiency (*increase*) in the industrial sector (*industrial, industry, manufacturing*) and the strongest relationship with economic variables (*economic, cost, costs*), environmental (*environmental, emissions*) and policies (*policy*).

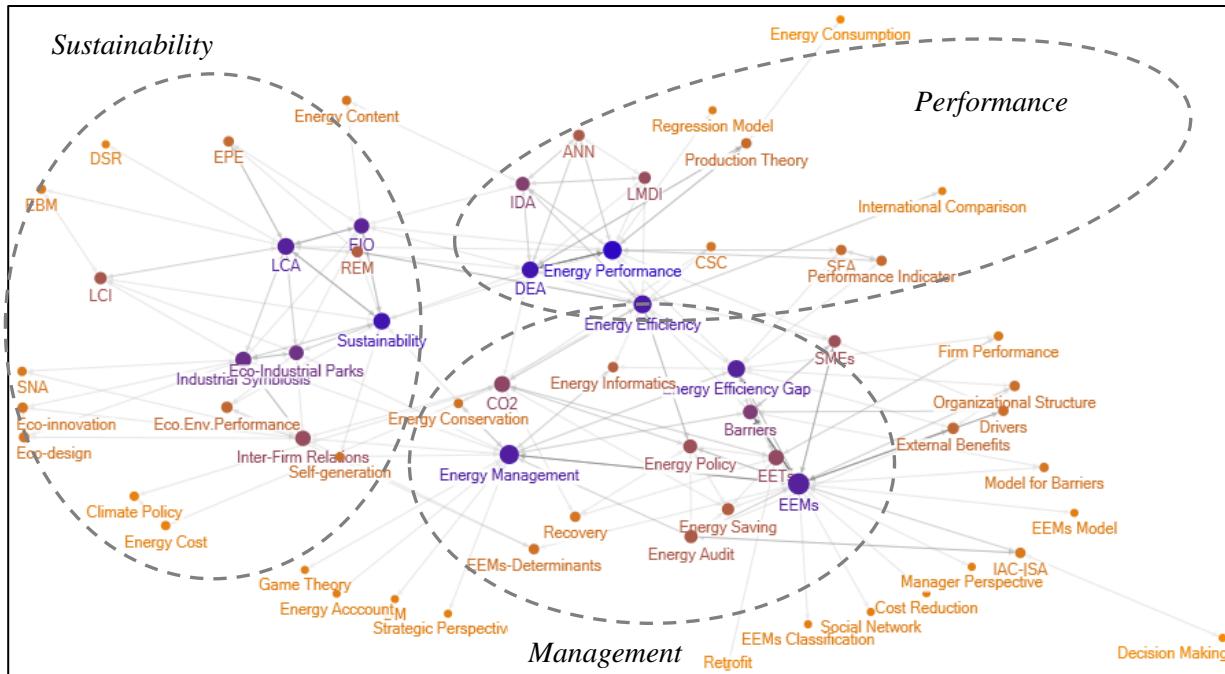


Figure 5 – Thematic network (63 themes in 104 articles)

Figure 5 shows the thematic network based on the extraction of two to five themes in each of the 104 articles treated. We used the same network statistics of Figure 4. Figure 5 can be classified into three major groups: Sustainability in energy, energy management and energy performance modeling. It can be understood that these groups form specific categories or sub-areas of interest. The categories were chosen with these names because have the largest eigenvector centrality statistics (stronger blue color).

Figure 6 shows the social network (co-occurrence) of the 100 authors (journals) most cited of the 104 articles, configured to match the previous network statistics. We can identify several authors, in the formation of the three previous categories. A group of dense authors in the center work with management and energy efficiency, another group in right side works with energy in a broader aspect of sustainability and a group of authors work with several models of quantitative performance in energy.

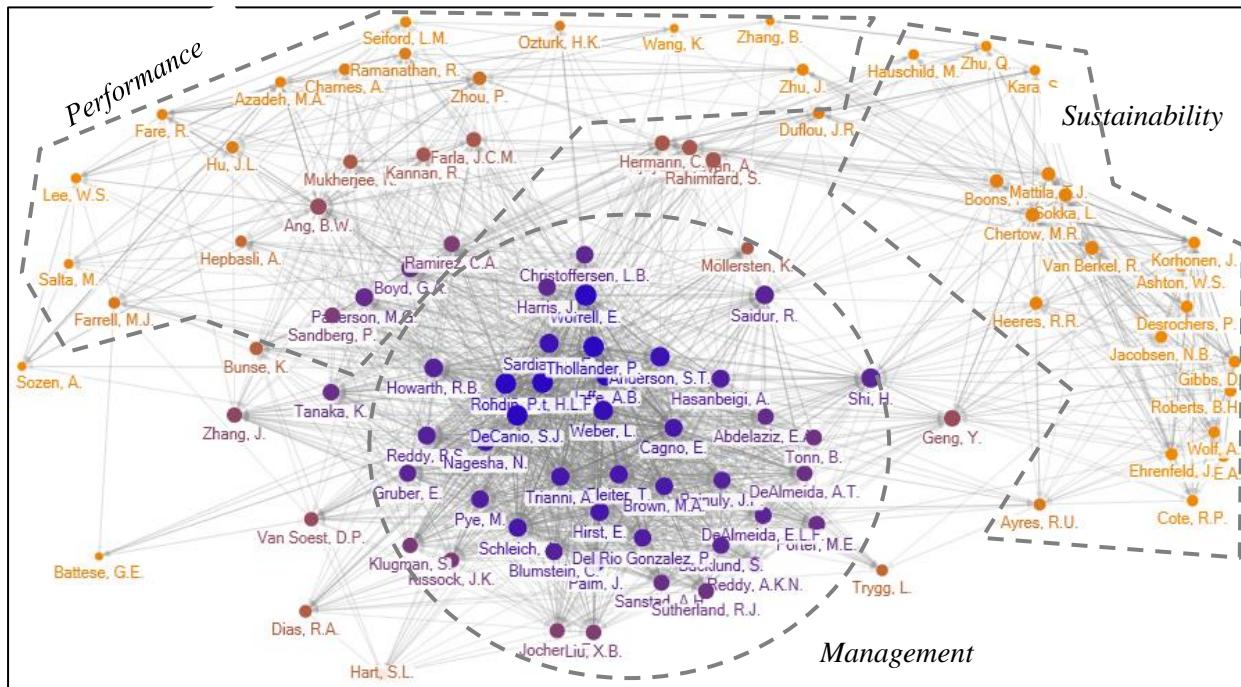


Figure 6 – Network of research field (100 most cited authors)

Figure 6 shows the co-citation of 100 most cited authors of the 1.255 journal authors. Since 1.255 these authors were cited 2.275 times, each author received an average of 1,81 citations. Figure 07 demonstrates how each article of the *corpus* was co-cited by other. Briefly Figure 6 shows the relationship of the authors of the research field and Figure 7 shows the relationship of the authors of the *corpus*. Of the 104 articles, 48 had co-citation, and received 244 citations, with citation average of 5,08 times. This means that the authors co-cited of the *corpus* received almost three times more citations than other journals authors, considering the average value.

The authors co-cited of the *corpus* were divided into four groups as shown in Figure 7, with their respective research topics.

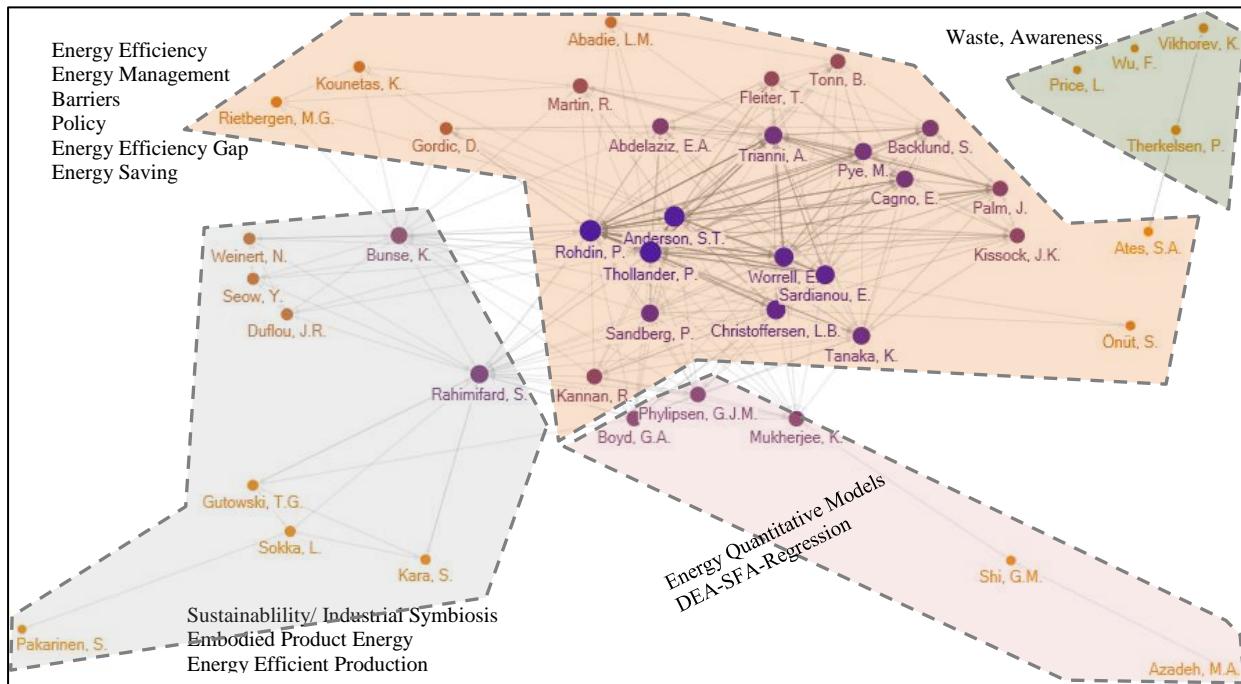


Figure 7 – Network of corpus (all co-cited authors)

Discussion

The method demonstrates an alternative way of presenting a theoretical framework of scientific research. Was implemented in the field of industrial energy efficiency, which is a new area of research. Was possible to access the main topics (categories) discussed, and the relationships between them through Figures 3b, 4 and 5. Based on the frequency and the relationships between the words and themes, some problems of the area can be raised:

- Concerns with existing barriers for energy efficiency measures;
- Cost is a significant barrier;
- Concerns in generate energy performance models;
- Strong relationship between energy consumption and environmental impacts, especially emissions.

By the method was possible to relate the issues and problems discussed with the authors of the group. This can be denoted by Figure 3a, 6 and 7. Although seen by the frequency of citations in figure 3a that the author "Thollander, P." has a number of proportionally higher citations, the co-citation analysis reveals the inexistence of an outlier. This demonstrates the lack of a "guru" in the field. The set of Figures 3 to 7 are the new theoretical framework of this area of research, allowing the proposition of new models with a specific position in the theoretical map.

The main difficulty is in identification of many authors and works for the construction of citation panel. The main limitation is that does not release the complete reading of the *corpus*, so

a very large number of documents (articles) require a longer demand, being necessary a more rigorous selection of the documents.

This work has made the integration by means of different techniques of bibliometric analysis and content analysis within the scope of the systematic literature review. It is recognized that the content analysis as part of qualitative research, goes beyond the techniques employed here. Based on the content panel and citation panel, other analyzes can be undertaken by expanding the present method.

Bibliography

Aquino, A.C.B., Pagliarussi, M.S., Bitti, E.J.S. 2008. *Heuristic Method for Composing a Literature Review*. 19(47):73–88. (In Portuguese)

Armitage, A., Allen, D. K. 2008. *Undertaking a Structured Literature Review or Structuring a Literature Review : Tales from the Field*. The Electronic Journal of Business Research Methods 6(2): 103–114

Bardin, L. 1977. *Content Analysis* – São Paulo. Edições 70, 2011(in portuguese)

Breiger, R.L. 1974. *The duality of persons and groups*, Social Forces, Vol. 53(2): Special Issue, 181-190.

Du, H., Wei, L., Brown, M. A., Wang, Y., Shi, Z. 2013. *A bibliometric analysis of recent energy efficiency literatures: an expanding and shifting focus*. Energy Efficiency 6(1): 177–190.

Du, H., Li, N., Brown, M. A., Peng, Y., Shuai, Y. 2014. *A bibliographic analysis of recent solar energy literatures: The expansion and evolution of a research field*. Renewable Energy 66:696-706.

Fernández, L. M. R., Bote, V. P. G., Anegón, F. M. 2012. *World scientific production on renewable energy, sustainability and the environment*. Energy for Sustainable Development 16(4):500–508.

Hanse, D.L., Shneiderman, B., Smith, M.A. 2011. *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*, Elsivier, Burlington.

Kane, M., Trochim, W.M.K., Kane, M. 2007. *Concept Mapping for Planning and Evaluation*, Sage Publications, California.

Kitchenham, B. 2004. *Procedures for Performing Systematic Reviews*. Keele University Technical Report TR/SE-0401.

Lopes, A. L. M., Fracolli, L. A. 2008. *Systematic Review of Literature and Qualitative Metasynthesis: Considerations about their Application in Nursing Research*. Reflexão 17(4), 771–778 (in Portuguese).

Luft, J., Shields, M.D. 2003. *Mapping managing accounting: graphics and guidelines for theory-consistent empirical research*. Accounting, Organization and Society. 28(2-3):169-249.

McLinden, D. 2013. *Concept maps as network data: Analysis of a concept map using the methods of social network analysis*. Evaluation an Program Planning (36):40-48.

Perroni, M. G., Pinheiro de Lima, E., Gouvea da Costa, S. E. 2014. *Proposal of a Model for Evaluation of Industrial Energy Performance: From Energy Efficiency to Effectiveness*. 7th International Conference on Production Research/American Region, Lima.

Pilkington, A., Meredith, J. 2009. *The evolution of the intellectual structure of operations management—1980–2006: A citation/co-citation analysis*. Journal of Operations Management, 27(3), 185–202

Plytiuk, C. F., Gouvea da Costa, S. E., Pinheiro de Lima, E. 2014. *Lean in Healthcare: A Systematic Literature Review and Social Network Analysis*. POMS 25th Annual Conference, Atlanta, USA.

Prell, C. 2012. *Social Network Analysis: history, theory and methodology*. Sage Publications, California.

Taylor, A., Taylor, M. 2009. *Operations management research: contemporary themes, trends and potential future directions*. International Journal of Operations & Production Management 29(12):1316–1340.

Trochim, W.M.K. 1989. *An Introduction to Concept Mapping for Planning and Evaluation*. Evaluation an Program Planning 12:1-16.

Tsay, M.Y. 2008. *A bibliometric analysis of hydrogen energy literature, 1965-2005*. Scientometrics 75(3):421-438.

Zhao, Y. 2012. *R and Data Mining: Examples and Case Studies*, Elsevier, New York.