

Core Topics Discovery in Sustainable Supply Chain Literature: An Automatic Approach

Carlos Montenegro¹, Edison Loza-Aguirre¹, Marco Segura¹, Eddie Yanez¹

¹Escuela Politécnica Nacional, Quito-Ecuador

carlos.montenegro@epn.edu.ec, edison.loza@epn.edu.ec, marco.segura@epn.edu.ec,
eddie.yanez@epn.edu.ec

Abstract. The study of Sustainable Supply Chain (SSC) has evolved and expanded over the last two decades. This study uses text mining and machine learning methods for automatically identify and classify the topics that permeate a collection of documents. The topics of SSC research were identified, using the Latent Dirichlet Allocation model, from 684 articles published between 2001 and 2017 in 13 top journals. Then, we explored trends by examining changes in the classification of topics in different periods and by identifying the hot and cold topics of SSC research. The relationships of these topics with the journals were also determined. Finally, applying the Competitive Neural Network learning model, the topics were classified according to the Elkington's Triple Bottom Line precepts. The findings of this study are expected to provide clues for researchers and policymakers in the field of SSC.

1. Introduction

Reviews concerning the literature on Sustainable Supply Chain (SSC) has focused on identifying intellectual pillars on the subject including influential works, journals, and authors [1] [2] [3] [4] [5]; or with the orientation for identifying core topics [6] [7] [8] [9].

The first publications that combined themes of sustainable development (SD) with logistics and supply chain management appeared in the early 2000s. The first notion to appear was Green Supply Chain Management (GSCM) which results from the combination of environmental management and supply chain management [10] [11]. Later, the researchers introduced the Logistics Social Responsibility (LSR) perspective that involved the social aspects of the SD, which have been set aside by GSCM [1] [12]. A holistic notion of integrating social and environmental aspects was born later. Seurin [13] proposes a definition of SSC Management based on the principles of Elkington's triple bottom line model (TBL) [14]. His notion stated the need for a balance of all three SD dimensions (economic, environmental and social). However, even though the balance between the three dimensions of SD has been recognized as the most desirable scenario, researchers have identified an unbalance on the consideration given to the three dimensions of SD [15] [11]. So, in the identification of the topics included in the SSC domain, the current SD dimensions constitute an important analysis factor of the technical literature content [16] [17].

Frequently, the analysis of a corpus of documents employs the conventional approach of manual assignments in which each article is classified into a single category of predetermined topics based on subjective judgement. See, for example, the studies of Monzer et al. [18] and Signori et al. [19]. In this paper, we adopted an alternative approach for the analysis to discover core topics in SSC Literature. This study reviews 684 Articles of top 13 journals, from 2001 to 2017, using Latent Dirichlet



Allocation (LDA) and Competitive Neural Network models for configuring an automatic approach for topics discovery.

2. Related work

The use of discrete manual assignments in the literature review cannot adequately capture potential topics of large scholarly data. Firstly, manual allocation based on reading abstracts or author keywords always entails the risk of classification errors. Secondly, the predetermined categories are by no means exhaustive; relatively new and emerging topics are likely to be ignored. Convergence topics are also difficult to handle. Thirdly, an article usually contains two or more topics.

Some studies describe the use of alternative methods to analyse the content of the corpus of documents, focused on research based on computational methods, which helps to reduce workload and processing errors, avoiding particular appraisals [20]. These studies show that the use of machine learning and text mining techniques has the potential to decrease the workload involved in the screening phase of or systematic literature mappings and reviews [21] [22] [23].

Topic models techniques automatically can discover the core topics that pervade a large and unstructured collection of documents [24]. These algorithms do not require any prior labelling of the documents; the topics emerge from the analysis of the original texts. Additionally, posterior labelling of topics can be done using an unsupervised learning technique. Given these advantages, recent years have seen an increased impetus to use the automatic approach in a variety of academic domains such as historical analysis [25], statistics [26], hydropower [27], among others.

3. Methods

A survey in several journal rankings is conducted to identify top-tier journals. It is validated the journals that are classed on the quartiles Q1 and Q2 on the SCImago Journal Rank (SJR). According to the contents of the SSC domain, a set of queries to collect the articles of the corpus is used.

Then, the Latent Dirichlet Allocation (LDA), a probabilistic unsupervised topic model is used to identify the topics from Bag of Words (BoW) codification, where the sequence of words has a probability which is not affected by the order in which they appear [24]. According to Griffiths and Steyvers [28], it is straightforward to analyse the dynamics of topics, as a means to obtain a vision of the dynamics of science. Here we use an analysis based on a post hoc examination of the estimates of probabilities produced by LDA, being able to identify the “hot” and “cold” topics. These topics are providing quantitative measures of the prevalence of particular types of research that may be useful for historical purposes and the determination of objectives of scientific interest. Next, Topics and SSC journals relationships are identified.

Competitive Neural Network [29], an unsupervised learning model, is used to classify the topics. The model is based on the concept of the winning neuron, which is as the one whose weight vector is the closest to the current input vector according to the distance (Euclidean in this case). During the learning phase, the weights of the winning neurons are modified to extract the average characteristic of the input patterns. Finally, the classified topics allow identifying the implicit relationships of SD dimensions in the corpus.

4. Results

The use of sections to divide the text of the paper is optional and left as a decision for the author. Where the author wishes to divide the paper into sections the formatting shown in table 2 should be used.

4.1. Corpus Collection

It is conducted a review in several journal rankings to determine the top-tier journals, retaining the journals that were ranked at one the two higher level on at least one classification. Since the research was confined to articles published before 2018, we consulted the standing rankings at that time.

For testing the pertinence and validity of the journal selection approach, we reviewed the quartile indicator of each journal on the SCImago (and SCOPUS) Journal Rank (SJR). It is validated the journals that were classed on the quartiles Q1 and Q2. Accordingly, six Production and Operations journals and seven Supply Chain and Logistics journals were retained. Then, it is collected the articles of these journals for the review. Systematically are applied the following filters of Table 1. The process carried out obtained 684 articles that form part of the corpus

Table 1. Filters to collect the Articles

(Sustainable AND supply chain) OR (Sustainable AND logistics) OR
 (Green AND supply chain) OR (Green AND logistics) OR (Sustainability
 AND supply chain) OR (Sustainability AND logistics) OR (Social AND
 sustainable AND supply chain) OR (Social AND sustainable AND
 logistics) OR (Social AND sustainability AND supply chain) OR (Social
 AND sustainability AND logistics) OR (Social AND responsibility AND
 supply chain) OR (Social AND responsibility AND logistics)

Table 2 presents the journals consulted; alike, Fig. 1 shows the articles per Year

Table 2. Journals consulted to collect the Corpus

ID	Journals	Articles
IJLM	Inter. J. of Logistics Management	25
IJLRA	Inter. J. of Logistics: Research and Applications	43
IJOPM	Inter. J. of Operations and Production Management	32
IJDLM	Inter. J. of Physical Distribution & Logistics Management	58
IJPE	Inter. J. of Production Economics	201
IJPR	Inter. J. of Production Research	99
JBL	Journal of Business Logistics	20
JOM	Journal of Operations Management	13
JSCM	Journal of Supply Chain Management	35
POM	Production and Operations Management	24
PPC	Production Planning and Control	59
SCMAIJ	Supply Chain Management	23
TRPE	Transportation Research Part E	52

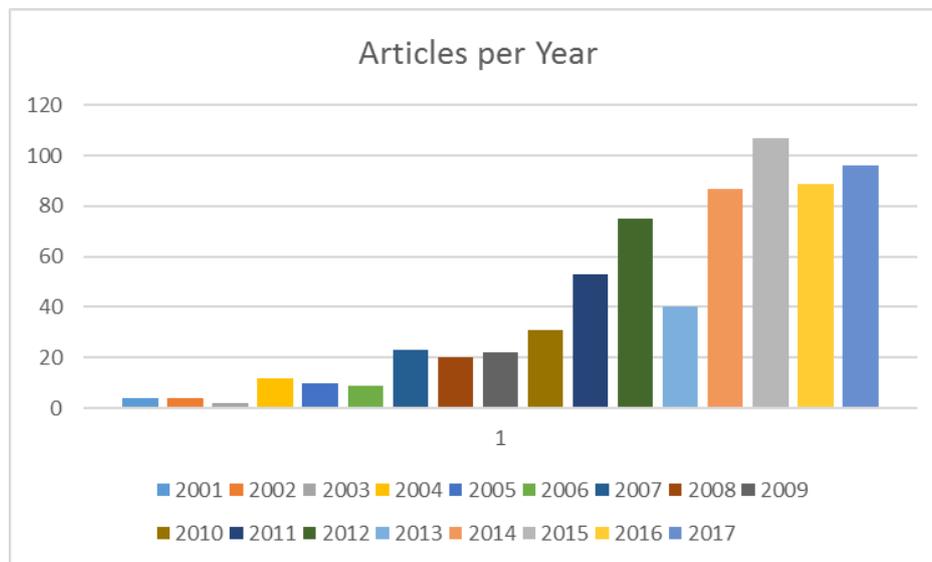


Figure 1. Corpus articles per year

4.2. LDA Codification and Topics

Once the articles are collected, some text pre-processing is required before conducting the LDA inference. Mainly, all standard and user stop words are eliminated. Additionally, on the content of articles must be removed the title, headers, footers, references, special characters, punctuation, whitespaces, and numbers

Gibbs sampling [30], a form of Markov Chain Monte Carlo, it is used to conduct the LDA implementation, which provides a relatively efficient method for extracting a set of topics from a corpus. Since there does not exist an optimal solution to estimate the number of topics, it is followed a best practice [31] which combine maximization [32] [33] and minimization [34] [35] criteria. By following this approach, the estimated value was 48, even though the first twenty cover the 80% of the corpus content, and the first ten, 60%.

Table 3. Topics in Corpus and Articles (Excerpt)

Topics in Corpus		Topics in Articles		First ten frequent Terms of Topics	Interpretation
Rank	Prob. %	Rank	Freq.		
1	19.9	1	163	Research, management, study, data, level, company, process, literature, analysis, operations	Research about companies, management, process, and operations
2	9.8	2	78	Model, costs, demand, product, transportation, objective, period, problem, network, capacity	Economic considerations of supply chain models
3	5.5	5	32	Environmental, green, performance, management, practices, manufacturing, supplier, supply, product, chain	Production, environment and supply chain
4	5.1	4	42	Performance, model, results, items, research, study, survey, factor, variables, sample	Research about models performance
				Criteria, decision, supplier,	Research about

5	4.0	3	44	performance, selection, factors, evaluation, fuzzy, green, environmental	environmental models performance
6	3.7	9	6	Sustainability, sustainable, social, environmental, economic, sscm, practices, development, theory, organizations	Sustainable environmental and economic theories and practices in organizations
7	3.6	10	2	Supply, chain, chains, product, industry, management, global, design	Design of global supply chains
8	3.6	8	17	Reverse, product, products, logistics, recycling, recovery, returns, collection, materials, value	Recycling
9	3.0	7	19	Transport, logistics, transportation, emissions, freight, road, vehicle, service, environmental, delivery	Environmental logistics, transport and delivery
10	3.0	6	24	Manufacturer, retailer, csr, price, channel, chain, supply, contract, demand, product	Manufacturing Supply chain and distribution channels

The extracted and labelled ten topics of SSC research are shown in Table 3 with ten frequent words and their interpretation. For the calculation, this research used R with recommended values of LDA parameters [36]. Topics are numbered in descending order of probabilities in the corpus. The number of articles in which each of the topics has the highest proportion is also provided. For example, Topic 3 has a rank of 3 in the corpus, according to the probability of appearance (5.5%); but, it has a rank of 5 for the appearance as the most critical topic in 32 articles of the corpus

Examining the dynamic changes in favourite topics over time can also provide fruitful implications for SSC researchers. The 17 years under study is divided into four periods. Next, it is presented the analysis based on dynamic changes in favourite topics over time (Table 4). Periods are arbitrary and cover since the date of the first publication in 2001.

Table 4. Dynamic changes in favorite topics over time

Topic Rank	Period 1 2001-2005	Period 2 2006-2010	Period 3 2011-2015	Period 4 2016-2017
1	0.25061	0.26351	0.18225	0.18499
2	0.03745	0.06103	0.11178	0.10178
3	0.09699	0.06725	0.05950	0.03090
4	0.10195	0.04093	0.05263	0.04478
5	0.00350	0.03460	0.04671	0.03737
6	0.01746	0.04910	0.03741	0.03735
7	0.03098	0.04505	0.03720	0.02995
8	0.08489	0.05036	0.03425	0.02201
9	0.04299	0.03630	0.02765	0.02574
10	0.00604	0.02198	0.03136	0.03601

It is possible to observe graphically what topics were prevalent in each period and how topic rankings have changed (Fig. 2). Eight of the ten topics are less important than 10%. The research about economic considerations of supply chain models is the hottest topic of the SSC field (Topic 2).

The combination of production, environment and supply chain (Topic 3), as well as the research about product recycling (Topic 8) and generic model performance (Topic 4), are cold topics that diminish their importance in the period of analysis. Topic 1, related to research, tends to recover. Other topics are less critical and do not have a clear tendency of development.

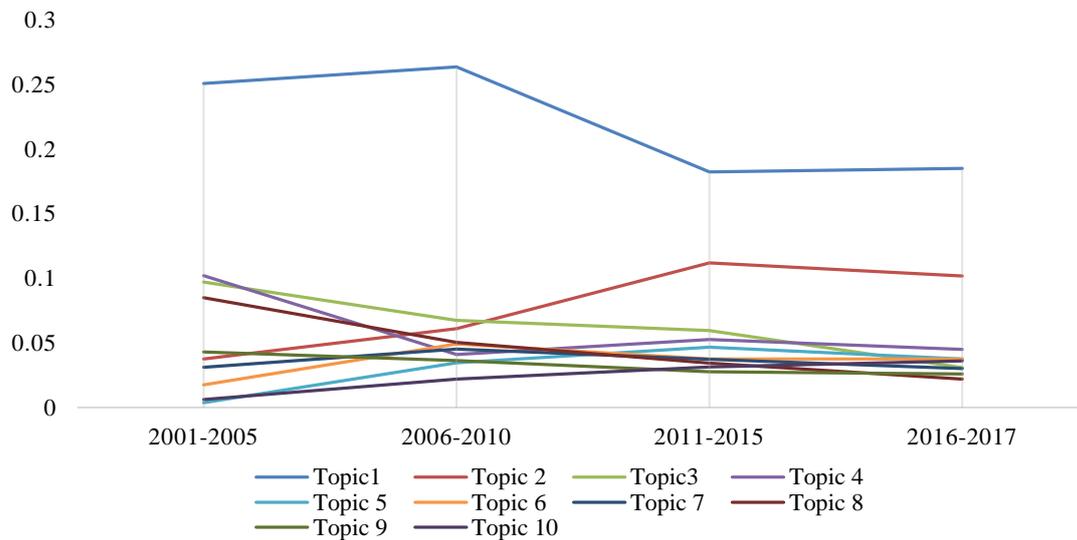


Figure 2. Hot and cold topics of the corpus

4.3. Topics and SSC journals relationships

Table 5 presents five top topics for each of the journals. Crossing this information with Table 3, it is evident that all journals publish researches about companies, management, process, and operations (Topic 1), and economic considerations of supply chain models (Topic 2). But, each journal includes distinctive topics; for example, IJPE publishes researches about environmental models performance (Topic 5), while IJPR includes recycling models (Topic 8). The group of IJPE, IJPR, and PPC include most content into the ten top topics of the corpus

Table 5. Topics and Journals Relationships

Journal	Articles	Five Top Topics
IJLM	25	1, 9, 4, 15, 18
IJLRA	43	1, 9, 32, 5, 2
IJOPM	32	1, 3, 4, 6, 18
IJDLM	58	1, 6, 9, 3, 11
IJPE	201	2, 1, 10, 5, 3
IJPR	99	1, 2, 5, 8, 3
JBL	20	1, 31, 4, 2, 39
JOM	13	1, 6, 18, 21, 4
JSCM	35	1, 6, 4, 12, 26
POM	24	1, 23, 10, 24, 43
PPC	59	1, 2, 5, 3, 8
SCMAIJ	23	1, 18, 3, 7, 20
TRPE	52	2, 1, 14, 3, 4

4.4. Topic Labeling and SD dimensions of the SSC Corpus

The topics are classified using the Competitive model. The logical number of categories is seven, as the number of SD dimensions and intersections. For data processing, the *ntool* of MATLAB was used with 1000 epochs and a learning rate of 0.1. Table 6 shows the aggregate results.

The topics are classified using the Competitive model. The logical number of categories is seven, as the number of SD dimensions and intersections. For data processing, the *ntool* of MATLAB was used with 1000 epochs and a learning rate of 0.1. Table 6 shows the aggregate results.

Table 6. Aggregate dimensions of the Topics

SD Dimension	Aggregate probability
Only Environmental	0.0517
Only Economic	0.2567
Only Social	0.0534
Environmental + Economic	0.4049
Environmental + Social	0.0985
Economic + Social	0.0481
Economic + Environmental + Social	0.0465

Fig. 3 shows the mapping using the TBL model to represent SD dimensions; it presents the theoretical model and the model using the real data of Table 6. Fig. 3b is obtained using a routine based on the Monte Carlo process. As is evident, the social dimension has minor relative development in the period of analyzing.

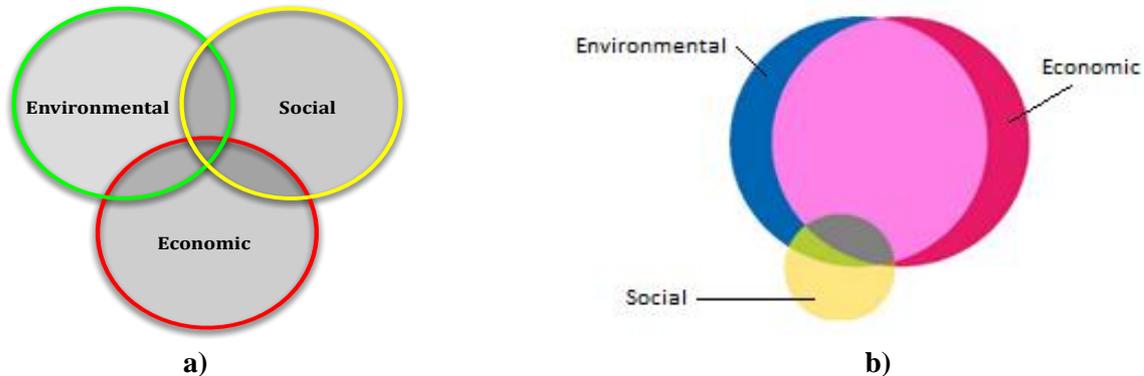


Figure 3.

a) TBL theoretical model

b) TBL using real data

5. Discussion and Conclusion

For continued progress in SSC research, the state of the art of research must be understood. It is hoped that the findings of this study will provide fruitful implications for researchers, journal editors and policymakers.

Researchers can judge whether their current research topics are hot or cold, and select appropriate journals to submit their papers based on the main topics and relationships with the journals. Newcomers to the field, both in the academic world and in practice, can obtain an overview of the SSC's research. The list of hot topics can be a useful reference for the allocation of grants and for designing R & D programs to promote SSC research.

The results of the study, in the period of analysis, evidence the following:

- a. SSC research has one current hot topic that contents an economic approach. There are some cold topics related to environmental issues that are diminished their importance. Research about companies, management, process, and operations tend to stay current.
- b. The number of publications per year tends to stabilize. IJPE and IJPR are the essential journals regard to the number of published articles
- c. Current TBL model suggests that most of the academic contributions are oriented to topics combining both environmental and economic aspects over social issues.
- d. Apart from the results concerning core topics on SSC research, this paper is another example of the validity of data mining and machine learning models as an alternative for analysing a corpus.

These results have an academic approach. A pertinent question is if the business has the same approach about the orientation of SSC development. This theme is an option for future work.

Finally, the main originality of this contribution lies in the analyse the corpus of technical information about the SSC theme, using a synthesis of data mining and learning models. The automatized approach would allow, on further research, the analysis on corpora in other fields. The procedure can be used not only in academia but also by governmental agencies or private organizations.

References

- [1] Carter. C. R. and Jennings, M. M. "Social responsibility and supply chain relationships," *Transport Res. E-Log.*, vol. 38, no. 1, pp. 37-52, 2002.
- [2] Teuteberg F. and Wittstruck, D. "Systematic Review of Sustainable Supply Chain Management Research," in *MKWI 2010 – Betriebliches Umwelt- und Nachhaltigkeitsmanagement*, 2010.
- [3] Amarasuriya, D. "A Systematic Review of Literature on Theories Available on Sustainable Supply Chain Management," *International Journal of Engineering Research And Management*, vol. 5, no. 9, pp. 1-5, 2018.
- [4] Saeed, M., Waseek I, and Kersten, W. "Literature Review of Drivers of Sustainable Supply Chain Management," in *Proceedings of the Hamburg International Conference of Logistics (HICL)*, 2017.
- [5] Massaroni, E., Cozzolino A., and Wankowicz, E. "Sustainability in supply chain management," *Sinergie*, vol. 33, no. 98, pp. 331-355, 2015.
- [6] Vlachos, I., Huatuco, L., ShakirUllah, G., and Roa-Atkinson, A. "A Systematic Literature Review on Sustainability and Disruptions in Supply Chains," in X. Liu (ed.), *Environmental Sustainability in Asian Logistics and Supply Chains*, 2019, pp. 85-96.
- [7] Tebaldi, L., Bigliardi B., and E. Bottani. "Sustainable Supply Chain and Innovation: A Review of the Recent Literature," *Sustainability*, vol. 10, no. 3946, pp. 1-29, 2018.
- [8] Ntabe, E., LeBel, L., Munson, A., and L. Santa-Eulalia, "A systematic literature review of the supply chain Operations reference (SCOR) model application with special attention to environmental issues," *International Journal of Production Economics*, 2015.
- [9] Sodhi M. and Tang, C. "Corporate social sustainability in supply chains: a thematic analysis of the literature," *International Journal of Production Research*, pp. 1-20, 2017.
- [10] Srivastava, S. K. "Green supply-chain management: A state-of-the-art literature review," *Int. J. Manag. Rev.*, vol. 9, no. 1, pp. 53-80, 2007.
- [11] Carter C. R. and Rogers, D. S. "A framework of sustainable supply chain management: moving toward new theory," *Int. J. Phys. Distr. Log.* , vol. 38, no. 5, pp. 360-387, 2008.
- [12] Murphy P. and Poist, R. " Socially responsible logistics: An exploratory study.," *Transport. J.*,

vol. 41, no. 4, pp. 22-35, 2002.

- [13] Seuring, S. M. M. "Core issues in sustainable supply chain management - a Delphi study," *Bus. Strategy. Environ.*, vol. 17, no. 8, pp. 455-466, 2008.
- [14] Elkington, J. *Cannibals with forks: the triple bottom line of 21st-century business*, Gabriola Island, BC.: New Society Publishers, 1998.
- [15] Pagell M. and S. A., "Why Research in Sustainable Supply Chain Management Should Have no Future," *J. Supply Chain Manag.*, vol. 50, no. 1, pp. 44-51, 2014.
- [16] Vimal K. and Vinodh, S. "Development of checklist for evaluating sustainability characteristics of manufacturing processes.," *Int. J. Proc. Manage. Bench.*, vol. 3, no. 2, pp. 213-232, 2013.
- [17] Sloan, T. "Measuring the Sustainability of Global Supply Chains: Current Practices and Future Directions.," *J. Glob. Bus. Manage.*, vol. 6, no. 1, pp. 1-16, 2010.
- [18] Monzer, D., Rebs, T., Khalid R., and Brandenburg, M. "Sustainable Supply Chain Management at the Base of Pyramid: A Literature Review," in *M. Brandenburg et al. (eds.), Social and Environmental Dimensions of Organizations and Supply Chains, Greening of Industry Networks Studies 5*, Springer International Publishing, 2018, pp. 235-257.
- [19] Signori, P., Flint, D., and Golicic, S. "Toward sustainable supply chain orientation (SSCO): mapping managerial perspectives," *International Journal of Physical Distribution & Logistics Management*, vol. 45, no. 6, pp. 536-564, 2015.
- [20] Hurtado, J., Agarwal A., and Zhu, X. "Topic discovery and future trend forecasting for texts," *Journal of Big Data*, vol. 3, no. 7, pp. 2-21, 2016.
- [21] Gulo C. and Rubio, T. "Text Mining Scientific Articles using the R Language," in *Proceedings of the 10th Doctoral Symposium in Informatics Engineering - DSIE'15*, 2015.
- [22] Thomas, J., McNaughton J., and Ananiadoub, S. "Applications of text mining within systematic reviews," *Research Synthesis Methods*, vol. 2, pp. 1-14, 2011.
- [23] Shemilt, I., Simon, A., Hollands, G., Marteau, T., Ogilvie, D., O'Mara, A., Kelly M., and Thomas, J. "Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening workload in extremely large scoping reviews," *Research Synthesis Methods*, vol. 5, pp. 31-49, 2013.
- [24] Blei, D. "Probabilistic Topic Models," *Communications of the ACM*, vol. 55, no. 4, pp. 77-84, 2012.
- [25] Wang, W., Mayfield, E., Naidu S., and Dittmar, J. "Historical Analysis of Legal Opinions with a Sparse Mixed-Effects Latent Variable Model," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Jeju, Republic of Korea, 2012.
- [26] De Battisti, F., Ferrara A., and Salini, S. "A decade of research in statistics: A topic model approach," *Scientometrics*, vol. 10, no. 32, p. 413-433, 2015.
- [27] Jiang, H., Qiang M., and Lin, P. "A topic modeling based bibliometric exploration of hydropower research," *Renewable and Sustainable Energy Reviews*, vol. 57, p. 226-237, 2016.
- [28] Griffiths T. and Steyvers, M. "Finding Scientific Topics," 2004.
- [29] Barreto, G., Mota, J., Souza, L., Frota, R., Aguayo, L., Yamamoto J., and Macedo, P. "Competitive Neural Networks for Fault Detection and Diagnosis in 3G Cellular Systems," in *J.N. de Souza et al. (Eds.): ICT 2004*, Berlin, 2004.
- [30] Liu, L., Tang, L., Dong, W., Yao S., and Zhou, W. "An overview of topic modeling and its current applications in bioinformatics," *SpringerPlus*, vol. 5, no. 1608, pp. 1-22, 2016.
- [31] Murzintcev, N. "Select number of topics for LDA Model," 24 October 2016. [Online]. Available:

- <https://cran.r-project.org/web/packages/ldatuning/vignettes/topics.html>. [Accessed January 2019].
- [32] Griffiths, T., Steyvers M., and Tenenbaum, J. "Topics in Semantic Representation," *Psychological Review*, vol. 114, no. 2, pp. 211-244, 2007.
- [33] Deveaud, R., Sanjuan E., and Bellot, P. "Accurate and Effective Latent concept for Ad Hoc Information Retrieval," *Revue des Sciences et Technologies de l'Information - Serie Document Numerique*, pp. 61-84, 2014.
- [34] Arun, R., Suresh, V., Veni C., and Murthy, M. "On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations," in *Advances in Knowledge Discovery and Data Mining*, M. Zaki and J. Xu, Eds., Berlin Heidelberg, Springer, 2010, p. 391–402.
- [35] Cao, J., Xia, T., Li, J., Zhang Y., and Tang, S. "A density-based method for adaptive LDA model selection," in *Neurocomputing — 16th European Symposium on Artificial Neural Networks*, 2009.
- [36] "Parameter estimation for text analysis," [Online]. Available: <http://www.arbylon.net/publications/text-est.pdf>.