A BIBLIOMETRIC APPROACH TO SUPPORT REDEFINING MANAGEMENT OF TECHNOLOGY FOR THE POST-DIGITAL WORLD

RUDOLPH OOSTHUIZEN
CSIR and University of Pretoria, Department of Engineering and Technology Management, South Africa
rudolph.oosthuizen@up.ac.za

ABSTRACT

Management of Technology (MoT) has evolved since its inception in the 1980s and definitions from the 1990s. However, the field's definition may not be keeping up with the ever-increasing changes in our world. This paper implements bibliometrics, through natural language processing and topic modelling, of published literature on MoT to trace the evolution of research focus areas. The processed literature consists of an extensive sample from a keyword search of publication databases. Analysing the topic priorities over time indicates how research in the field evolved. Comparing these focus areas to the different definitions provide inputs for improving the definition of MoT. The topics extracted in this paper over the history of MoT offers a base from where to initiate such an investigation.

Key words: Management of Technology, Definition, Topic Modelling, Machine-learning.

INTRODUCTION

Management of Technology (MoT) was defined in the 1980s but needed to evolve along with changing environment. Although some MoT definitions have stood the test of time, the ever-increasing rate of change necessitates an in-depth investigation into definitions for the future (Botha, 2020). Over the past decades, the study of MoT and its related fields has continued to evolve and expand, establishing itself as an academic discipline (Lee & Kang, 2018). During the 1970s, governmental and practitioner authors took the lead to develop the need for a capability to manage the development and implementation of technology.

In the 1980s, the focus was on economics, trade policy, basic technology and education to achieve international industrial competitiveness. However, this approach overlooked the ability to manage technology. As a result, Herink et al. (1987) defined MoT as linking "engineering, science, and management disciplines to plan, develop, and implement technological capabilities to shape and accomplish an organisation's strategic and operational objectives". This definition focuses on technology strategy (management) to make an enterprise more competitive. During this period, business schools took the lead in maturing the field (Gerybadze, 2020).

The 1990s saw the emergence of MoT educational programmes to promote the field as MoT was considered a competitive advantage. Research and development (R&D) was a priority during this period. However, research was biased towards technology-push as large multinational corporations dominated innovation (Gerybadze, 2020). Porter et al. (1991) also made a distinction between industrial management and technology management, shifting the focus of MoT to integrate technology with softer issues of people.

However, since the early 2000s, innovation and R&D became global and diversified (Gerybadze, 2020). Hence, t expanded understanding of MoT includes "all knowledge, products, processes, tools, methods and systems employed in the creation of goods and in providing services" (Khalil, 2000). The definition includes the notion of wealth creation as a concept broader than only money to have the factors of knowledge, intellectual capital, resources, the natural environment, and raising of the standard of living. (Chanaron & Grange, 2006) also updated the vision of MoT as involving the mobilisation of resources to create and implement knowledge and know-how on markets. Innovation is seen as the source of growth due to the relationships between science, technology and society for sustainable development on a macro level. Therefore, MoT adds value to products and services through the application of knowledge and know-how. This highlights the intertwined and blurred relationship between MoT and the management of innovation. As innovation is related to change, MoT may also be considered to manage change (Botha, 2020).

Management of Technology (MoT) will have to evolve to remain relevant in a post-digital world (Botha, 2020). Tracing and mapping the evolution of perceptions from researchers and practitioners about what MoT entails over this time should help understand the requirements of a "new" definition that will address the challenges of the post-digital world. Research aims to create knowledge by understanding, explaining and predicting phenomena observed within a field. Knowledge development in a scientific discipline involves learning through empirical observation, formulation of theories, and experimenting to test theories (Adam & Fitzgerald, 2000; Caillaud et al., 2016). It is challenging to keep track of dynamic research where new sciences continually evolve to split or merge and gain or lose importance (Lamba & Madhusudhan, 2019).

This paper aims to examine the historical trends in MoT research to identify the trends in focus areas. Firstly, bibliometrics is discussed to motivate it as a valid method to trace the development in a research field. Natural Language Processing (NLP) of bibliometric data will then be performed to extract data on the evolution of MoT research. Unsupervised machine learning-based algorithms will then be applied to extract research topics from the abstracts and titles from papers published about MoT. Processing this data supports generating temporal trends for research on specific topics. The research in this paper builds on previous topic modelling related research within the field of MoT.

MANAGEMENT OF TECHNOLOGY

The definitions listed above may be inadequate when considering the future and a post-digital world where humanities and social sciences are coupled to new planning, development and implementation of technology principles. In future, this may happen at the point of use through consumer innovation and personal fabrication for technology adoption. As a result, the technology manufacturing enterprise may need to focus on marketing raw materials and algorithms for point of use manufacturing and customisation instead of finished products. Therefore, the definitions for MoT should be aligned demand-driven nature of the post-digital world and account for the time factor of fast change (Botha, 2020). The evolution of MoT objectives since 1980 can be summarised as follows (Botha, 2020; Chanaron & Grange, 2006):

i. 1982 – 1992: Modernise production
ii. 1992 – 2002: Deploy and integrate information systems as well as strategic planning of technological products
iii. 2002 – 2012: Innovate and optimise the life cycle

It is not easy to provide a single concise definition for MoT due to the fluid and dynamic processes in the field, which experiences constant incremental and evolutionary change (Badawy, 2009).

Therefore, Badawy (2009) updated the definition for MoT as the architecture or configuration of management systems, policies, and procedures governing its strategic and operational functioning to achieve its goals and objectives.

Future technology requires the inherent flexibility to sense changes, reshape and re-assemble to adapt to instantaneous need shifts. This place new requirements on MoT to design, manufacture and use products or services with digital technology. With the rapid progress of technological evolution and its role in adding value, the definition of MoT should be revisited, considering the disruptive forces of the post-digital world (Botha, 2020).

METHOD

This paper analyses the bibliometric data from research papers published on MoT to extract semantic topics. The prominence of these topics is tracked since 1980 and related to the relevant definitions of MoT. Academic publications provide a platform for researchers to codify their outputs. Peer reviews of these papers offer validation for the captured knowledge. Many scientific fields already apply this form of quantitative analysis of published data to map research and knowledge evolution (Keathley et al., 2015). Bibliometric analysis may focus on publication performance (citations) or science mapping (conceptual structure) to describe the research progress in a scientific field. This paper applies unsupervised machine-learning-based topic modelling to capture the published research trends in MoT since its inception (Lamba & Madhusudhan, 2019).

Manual and automated topic modelling methods can extract core topics from published research. The traditional method for topic modelling is manually assigning papers to a predetermined topic list derived from subject-matter experts. However, as the scientific publication repository size increases, manually reading and sorting each into topic fields becomes too time-consuming and challenging, with a high risk of bias. This may also miss the hidden, latent, or emerging topics in a large text corpus. Furthermore, due to the multi-disciplinary nature of MoT, a single paper may contain multiple topics. Therefore, automated machine-learning methods should be a viable alternative to improve the topic modelling of large text corpora (Lee & Kang, 2018).

Automated topic modelling implements NLP to interpret human language and convert text into numerical values for machine-learning algorithms to process. NLP algorithms also categorise and cluster text for extracting information to classify and summarise documents. Topic modelling is an unsupervised text classification method with quantitative statistical algorithms to extract semantic information from text. The process does not require any prior understanding of the corpus documents to remove the thematic concepts (Agrawal et al., 2018).

Topic modelling works on the assumption that a document contains latent topics intended by the author. The topic modelling algorithm extracts these words' co-occurrence and hidden relationships in the documents to define each topic as a probability distribution of selected terms (Jiang et al., 2016; Tong & Zhang, 2016). Latent Dirichlet Allocation (LDA) is a standard topic modelling algorithm that provides a generative statistical model to explain the similarity of the data from groups of words. The algorithm learns the statistical distribution of topics in each document in an extensive corpus based on the significant thematic clusters to forms document-topic and topic-word pairs (Hecking & Leydesdorff, 2019; Maier et al., 2018).

The primary parametric input to the LDA process is the required number of topics for extraction. The algorithm evaluates each publication in the text corpus for association with a set of topics using a probability value. Interpreting, naming, and describing each topic still requires expert and domain knowledge. However, topic modelling results may not be deterministic due to built-in stochastic processes with random seed values and influenced by input parameters, resulting in different sets of topics with each run of the algorithm. The selection of balanced algorithm parameters will stabilise the outputs and limit the inaccuracies (Agrawal et al., 2018; Eker et al., 2019; Tong & Zhang, 2016). Therefore, a valid output requires a cleaned and prepared text corpus, a suitable set of parameters followed by thorough inspection and naming of the topics (Maier et al., 2018).

Despite some challenges, topic modelling still provides a more comprehensive and faster result than other manual methods for performing a literature review, especially when analysing large amounts of literary publications. Furthermore, since the topic modelling algorithms record all the data from processing steps, useful data is available for validation and post-processing (Asmussen & Møller, 2019). The process to capture data, extract topics and perform post-processing is shown in Figure 1.

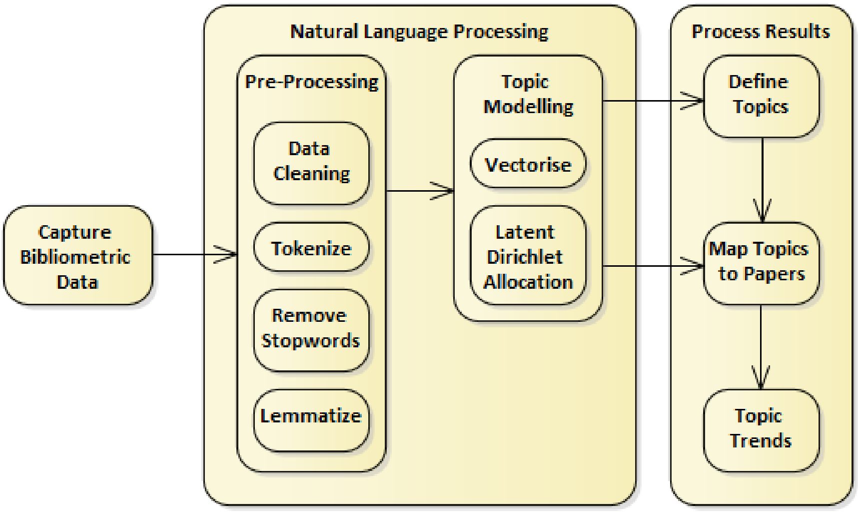

Figure 1: Topic Modelling and Analysis of Technology Management Bibliometric Data

The first step is to capture bibliometric data on papers published on MoT. For this paper, the search terms ("Management of Technology" OR "Technology Management") were used to download 6470 papers from the Scopus and 3631 from the Web of Science (WoS) repositories. The combined data set consists of the publication date, authors, title, citation, publication name, publication type and repository source. As not all the papers were research papers required for the processing, the papers without authors, titles and abstracts were manually removed. Removed papers also included editorials, corregidums, errata, correspondence and obituaries. Since a notable overlap between the papers from Scopus and WoS occurred, duplicates also required removal. The final combined list that remained for topic modelling processing consisted of 6602 papers from 1969 to 2021.

NLP was performed on the captured bibliometric text using the SpaCy library in Python. The raw captured bibliometric text requires pre-processing to prepare it for analysis. The titles and abstracts were combined to increase the size and richness of the text sample for processing. The titles contain some of the most representative words on the focus topic of the article, while the abstract summarises the problem description and method along with some of the research results (Agrawal et al., 2018;

Lee & Kang, 2018). The following pre-processing steps were performed to transform the text into a format suitable for analysis (Eker et al., 2019):

i. Data Cleaning. The text contains unwanted words (e.g. published, copyright, Elsevier, etc.), punctuations, numbers, capitalised letters, special characters, and redundant spaces that hamper the NLP algorithms unless removed. Some key abbreviations and acronyms, such as "R&D" (Research and Development), were replaced with the whole words to improve processing accuracy and consistency.

ii. Tokenize. Tokenisation extracts the linguistic building blocks for sentences as words using the spaces in between.

iii. Remove Stopwords. The stopwords tend to be common words, such as "as", "and", "the", "if", "a", etc., not adding meaning to the text. In addition, stopwords have a high frequency that adds noise when processing the text. Due to the academic nature of research articles, some common words need to be removed that do not provide information on the specific research topics in the field of MoT, such as "paper", "study", "research", "describe", "article", "introduce", "example", "literature" etc.

iv. Lemmatisation. Lemmatization also normalises and reduce text dimensionality by combining derivatives of words, such as plurals and past tenses. The lemma for each word is determined through morphological analysis and a vocabulary with part-of-speech information.

The Scikit-learn library in Python was implemented to vectorises the text data and perform the topic modelling LDA algorithm. The CountVectoriser function transforms the remaining text (terms) into a document term matrix using a minimum and maximum document frequency setting (max_df and min_df) of the words per document frequency. This setting enables the algorithm to exclude uncommon or too common terms throughout the text corpus. The total number of documents, the number of words per document, and the distribution of individual words over the documents may affect the optimum values selected for max_df and min_df parameters. The LDA function processed the document term matrix to extract the predefined number of topics.

The next topic modelling step is to define and name each of the extracted topics. The LDA algorithm can only cluster documents by their topics without identifying a name for each topic. Manual analysis, with domain knowledge, is applied to interpret the LDA results and associated bibliometric data to assign a descriptive topic to each cluster of terms. However, this is still a subjective process that may introduce bias (Lee & Kang, 2018). The algorithms also assigned the extracted topics to each paper by best-fit probability.

The LDA model and the output topics still require evaluation. Topic model evaluation is still immature despite accelerated progress over the past few years. Validation of topics is challenging in the absence of external ground truth. The probabilistic nature of the LDA algorithm may produce non-deterministic results due to random seed values to initiate training of the model. Selecting a good set of parameters with prepared data will produce relatively minor differences in the topics (Maier et al., 2018). The primary consideration should be the utility or usefulness, i.e. interpretability, replicability, external validity, and internal coherence of the model to understand the research field under consideration (Hagen, 2018; Isoaho et al., 2021). Experts in the field may provide semantic validation in terms of interpretability and plausibility of the topics. This external validation ensures the relevance of the topics without being misleading, uninformative, or wrong.

RESULTS

The topic modelling in this paper approached the data set from different perspectives. Firstly, the complete data set of 6602 articles were processed first to extract 50 topics. The output is an extracted set of fine-grain topics. However, the focus will only be on the 20 most popular topics to identify the main focus areas in the field. The fine grain of the topics aims to enable unanimous identification of the extracted topics and consistency between algorithm code executions. Next, topic prominence values were calculated by summing the probability of each topic being assigned per paper over the whole data set. Only the first ten topics per paper were added for simplicity as it accounts for an average of 96% of the total allocation percentage, as seen in Figure 2.

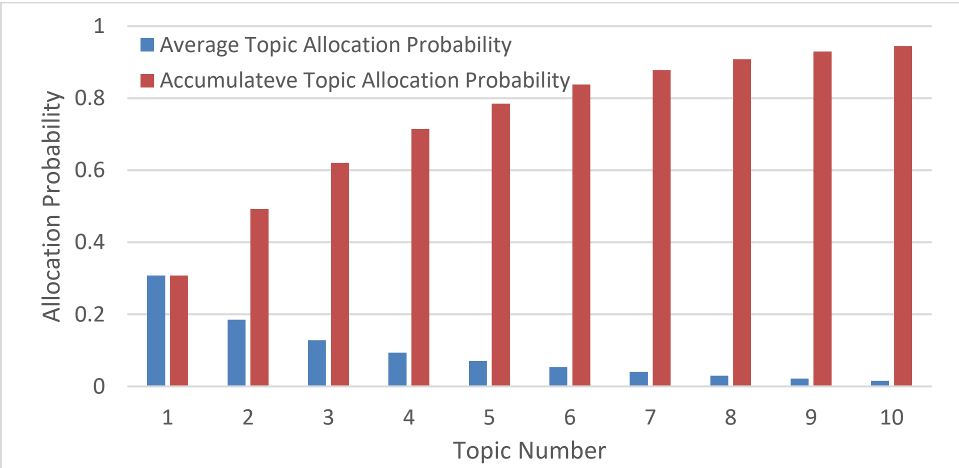

Figure 2: Topic Allocation Probability per Paper

It comes as no surprise that the most popular topic is "Management of Technology" with a score of 496, which is nearly double the next topic. However, as this topic adds no value for analysing the focus areas within the field, it is ignored in the detailed topic analysis. The topics are defined in Table 1 using their terms and a short description. The relative importance of the topics is shown in Figure 3.

The papers were also split and grouped into different periods. Papers from the 1980s and 1990s were combined to provide a data set (1029) large enough for effective topic modelling. Papers from the 2000s accounted for 2260 papers. The 2010s consisted of an extensive paper set, which necessitated a split in 2015. The period from 2010 to 2015 contained 1600 papers, while the final period will be from 2016 to 2021, processed 1600 papers. Splitting the data into the different periods helps identify the prominent topics that may become insignificant when looking at the complete set.

The topic modelling process was applied to each of these data sets to determine the priority of topics per period as the field of MoT evolved. As each period have a reduced collection of papers, only 40 topics were extracted from this reduced set of papers to achieve a fine grain. The 15 most popular topics with their level of relative popularity are provided in Figures 4, 5, 6, and 7.

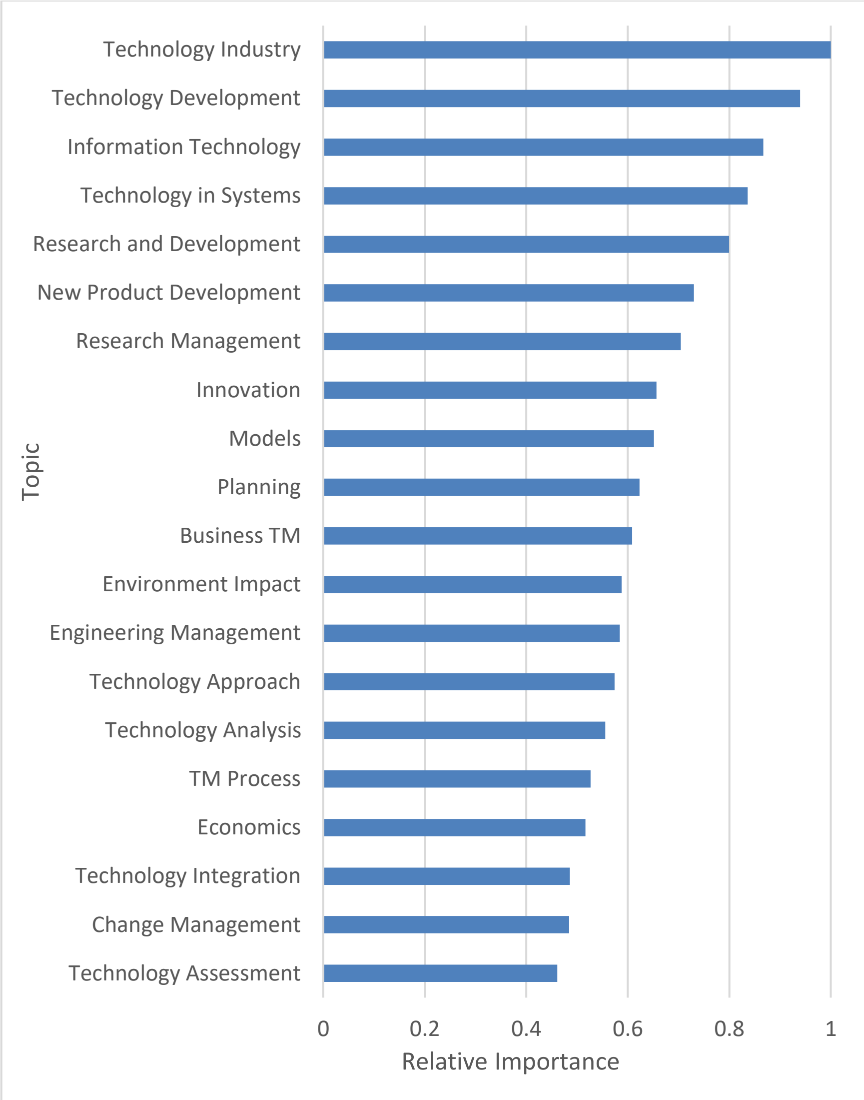

Figure 3: Relative Topic Allocation Prominence of the Full Data Set

Table 1 provides a summary of the most prominent topics over the four periods. Again, these topics compare well with the topics derived from the whole data set, although the relative priorities may differ a little. Finally, figure 8 shows how the popularity of selected topics from the processed complete data set changes over time. Again a reasonable comparison is noticeable between the different data sets.

Comparing the outputs from the data processed in this paper to the literature about the evolution of MoT, clear relationships can be identified. The 1980s and 1990s were seen as the phase when MoT was established as a discipline. During this period, the prominent topics were the principles of technology management, business and technology strategies, R&D, and educational programmes to achieve international industrial competitiveness.

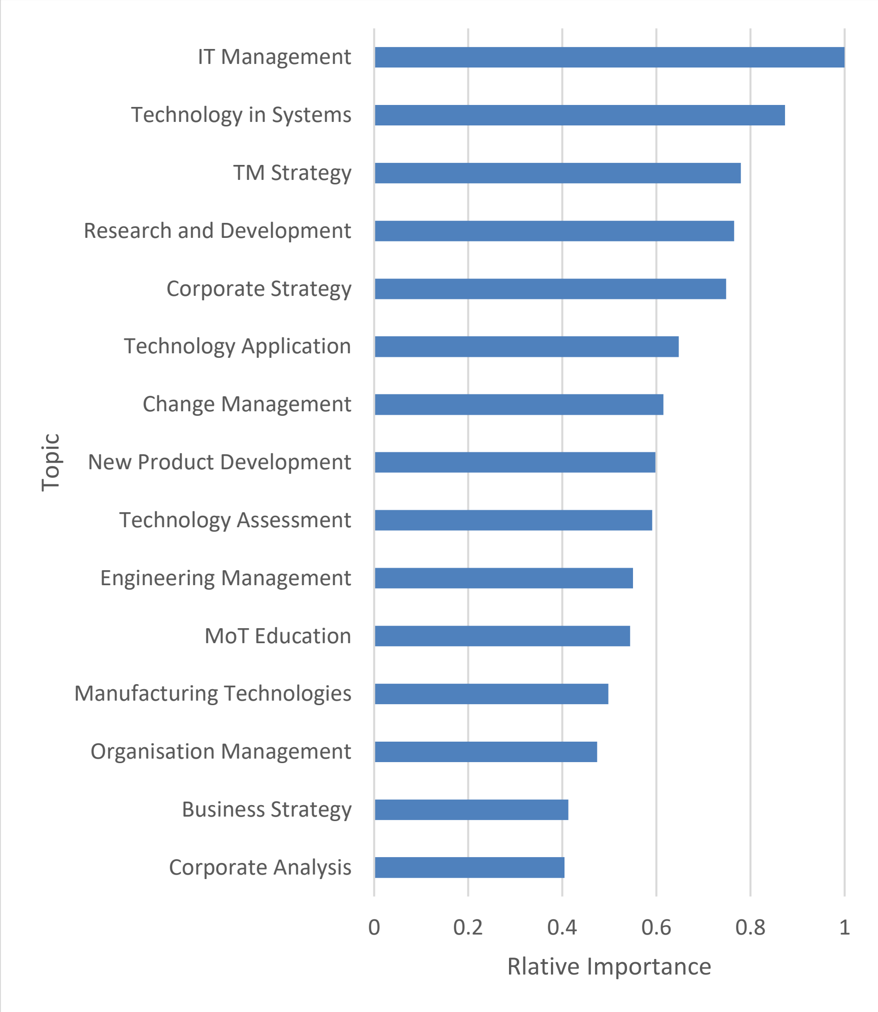

Figure 3: Relative Topic Allocation Prominence of the 1980s and 1990s

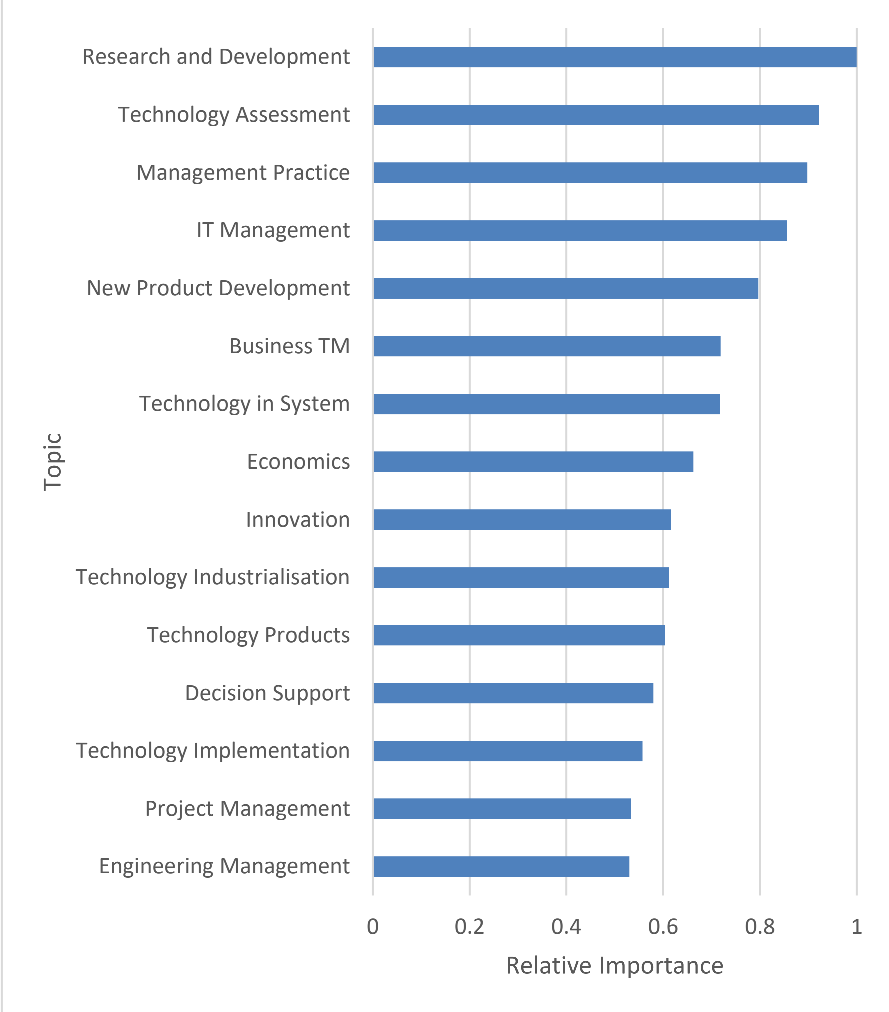

Figure 4: Relative Topic Allocation Prominence of 2000 to 2009

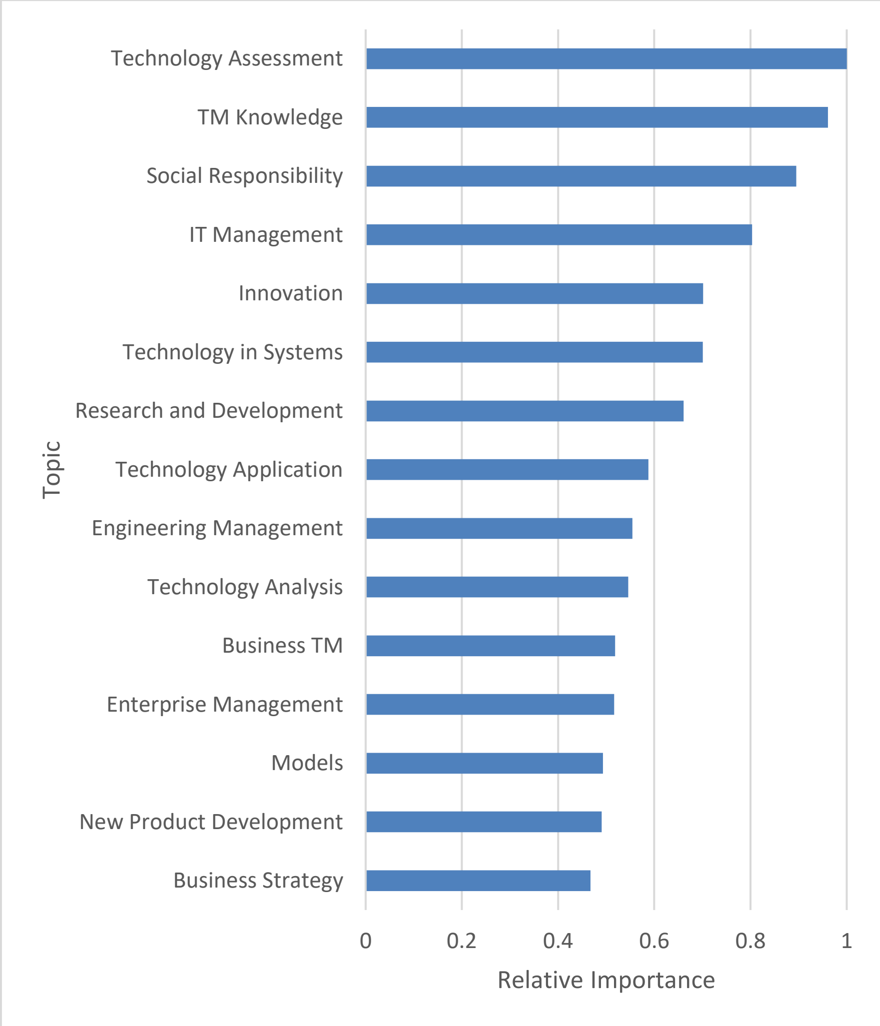

Figure 5: Relative Topic Allocation Prominence of 2010 to 2015

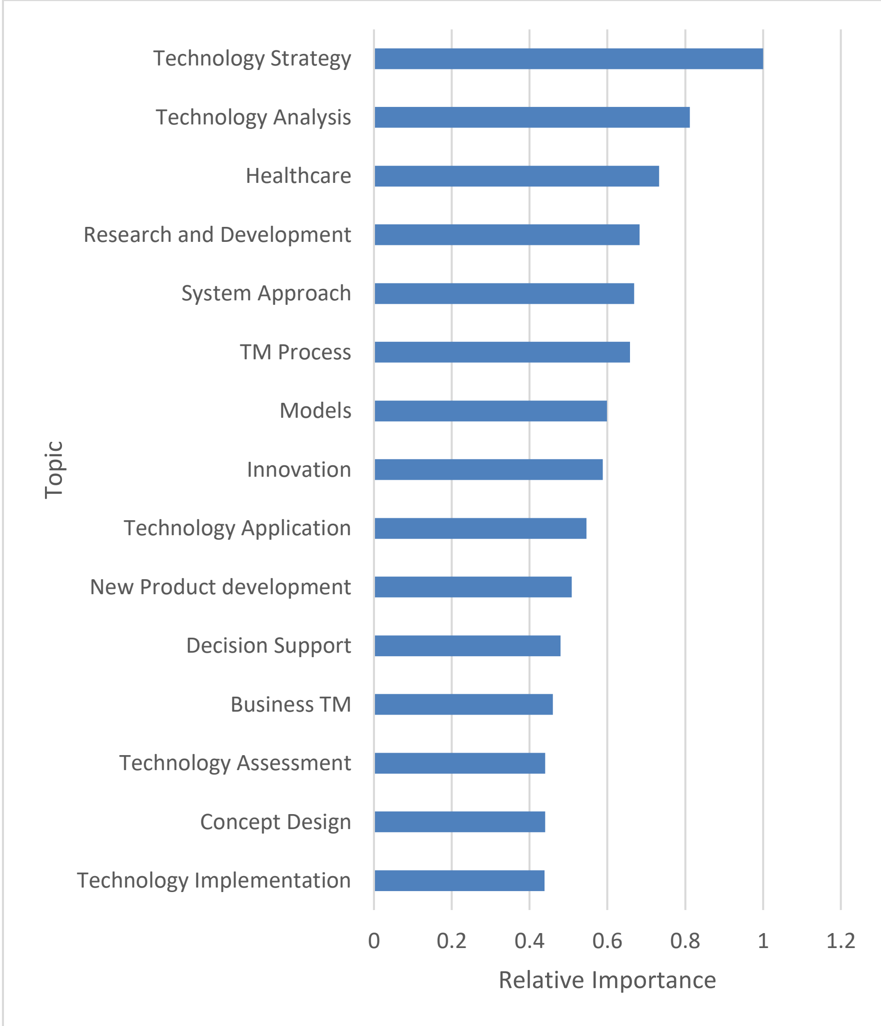

Figure 6: Relative Topic Allocation Prominence of 2015 to 2021

Table 1:Summary of Topics

No	Topic Name	Whole Set	1980-1990	2000s	2010s	2016s
1	Research and Development	0.80	0.77	1.00	0.66	0.68
2	Technology in Systems	0.84	0.87	0.72	0.70	0.67
3	IT Management	0.87	1.00	0.86	0.80	
4	Technology Assessment	0.46	0.59	0.92	1.00	0.44
5	New Product Development	0.73	0.60	0.80	0.49	0.51
6	Innovation	0.66		0.62	0.70	0.59
7	Business TM	0.61		0.72	0.52	0.46
8	Engineering Management	0.58	0.55	0.53	0.55	
9	Technology Analysis	0.56			0.55	0.81
10	Technology Application		0.65		0.59	0.55
11	TM Strategy		0.78			1.00
12	Models	0.65			0.49	0.60
13	TM Process	0.53				0.66
14	Economics	0.52		0.66		
15	Change Management	0.48	0.61			

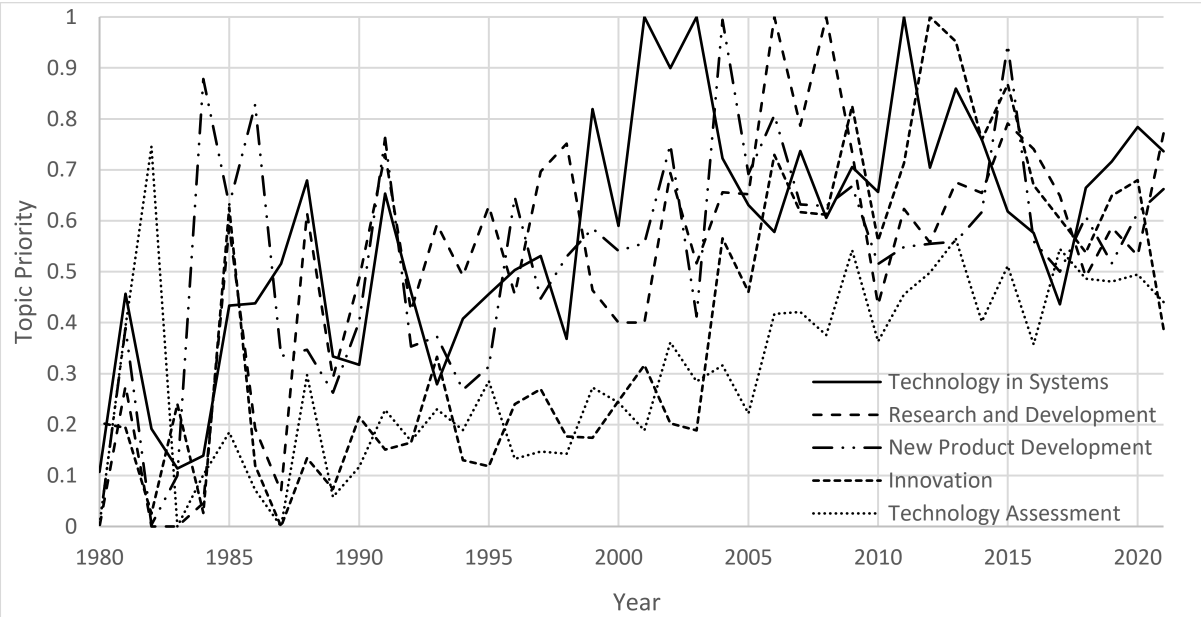

Figure 7: Topic Trends over the Whole Data Set

CONCLUSION

Comparing the outputs from the data processed in this paper to the literature about the evolution of MoT, clear relationships can be identified. The 1980s and 1990s were seen as the phase when MoT was established as a discipline. During this period, the prominent topics were the principles of technology management, business and technology strategies, R&D, and educational programmes to achieve international industrial competitiveness.

According to the literature, R&D became a clear focus in the 2000s (Gerybadze, 2020), as seen in Figure 5. This was supported by innovation for globalisation and diversification; other topics are also prominent in Figure 5. The theme from the literature on wealth creation is reflected in the topic of economics. During the early part of the 2010s, to focus shifted to technology assessment, knowledge management, and the social (including environmental) aspects of technology implementation. During the latter half of the last decade, healthcare appeared as a significant theme in MoT. This is supported by a resurgence of technology strategy and analysis.

Another output of this paper demonstrates the utility of topic modelling to analyse a research field to define the main themes. These themes are helpful to enable defining or mapping the domain. The data extracted in this paper now provides a platform for improving the definition of MoT that encapsulates recent research and can be projected into the near future.

REFERENCES

Adam, F., & Fitzgerald, B. (2000). The status of the information systems field : historical perspective and practical orientation. *Information Research*, *5*(4).

Agrawal, A., Fu, W., & Menzies, T. (2018). What is wrong with topic modeling? And how to fix it using search-based software engineering. *Information and Software Technology*, *98*, 74–88. https://doi.org/10.1016/j.infsof.2018.02.005

Asmussen, C. B., & Møller, C. (2019). Smart literature review: a practical topic modelling approach to exploratory literature review. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0255-7

Badawy, A. M. (2009). Technology management simply defined: A tweet plus two characters. *Journal of Engineering and Technology Management - JET-M*, *26*(4), 219–224. https://doi.org/10.1016/j.jengtecman.2009.11.001

Botha, A. (2020). Management of Technology in a Post-Digital World. *29th Annual Conference of the International Association for Management of Technology (IAMOT 2020)*, 889–900.

Caillaud, E., Rose, B., Goepp, V., Caillaud, E., Rose, B., Goepp, V., Caillaud, E., Rose, B., Goepp, V., & Strasbourg-icube, I. (2016). ScienceDirect systems engineering : for systems engineering : engineering : some systems engineering : *IFAC-PapersOnLine*, *49*(12), 1567–1572. https://doi.org/10.1016/j.ifacol.2016.07.803

Chanaron, J. J., & Grange, T. (2006). Towards a Re-definition of technology management. *ICMIT 2006 Proceedings - 2006 IEEE International Conference on Management of Innovation and Technology*, *2*, 955–959. https://doi.org/10.1109/ICMIT.2006.262363

Eker, S., Rovenskaya, E., Langan, S., & Obersteiner, M. (2019). Model validation: A bibliometric analysis of the literature. *Environmental Modelling and Software*, *117*, 43–54. https://doi.org/10.1016/j.envsoft.2019.03.009

Gerybadze, A. (2020). Technology and innovation management in a global perspective. *Managing Innovation in a Global and Digital World*, 207–225. https://doi.org/10.1007/978-3-658-27241-8_14

Hagen, L. (2018). Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models? *Information Processing and Management*, *54*(6), 1292–1307. https://doi.org/10.1016/j.ipm.2018.05.006

Hecking, T., & Leydesdorff, L. (2019). Can topic models be used in research evaluations? Reproducibility, validity, and reliability when compared with semantic maps. *Research Evaluation*, *28*(3), 263–272. https://doi.org/10.1093/reseval/rvz015

Herink, R. H., Adler, P. ., Anderson, R. M. ., Bean, A. S. ., Chen, K. ., Condit, P. M. ., Fowler, R. E., &

Hanna, A. M. (1987). *Management of Technology The Hidden Competitive Advantage*.

Isoaho, K., Gritsenko, D., & Mäkelä, E. (2021). Topic Modeling and Text Analysis for Qualitative Policy Research. *Policy Studies Journal*, *49*(1), 300–324. https://doi.org/10.1111/psj.12343

Jiang, H., Qiang, M., & Lin, P. (2016). A topic modeling based bibliometric exploration of hydropower research. In *Renewable and Sustainable Energy Reviews* (Vol. 57, pp. 226–237). Elsevier Ltd. https://doi.org/10.1016/j.rser.2015.12.194

Keathley, H., Bean, A., Chen, T., Vila, K., Ye, K., Tech, V., & Gonzalez-Aleu, F. (2015). BIBLIOMETRIC ANALYSIS OF AUTHOR COLLABORATION IN ENGINEERING MANAGEMENT RESEARCH. *Proceedings of the American Society for Engineering Management*.

Khalil, T. (2000). *Management of Technology: The Key to Competitiveness and Wealth Creation*. McGraw-Hill Higher Education.

Lamba, M., & Madhusudhan, M. (2019). Mapping of topics in DESIDOC Journal of Library and Information Technology, India: a study. *Scientometrics*, *120*(2), 477–505. https://doi.org/10.1007/s11192-019-03137-5

Lee, H., & Kang, P. (2018). Identifying core topics in technology and innovation management studies: a topic model approach. *Journal of Technology Transfer*, *43*(5), 1291–1317. https://doi.org/10.1007/s10961-017-9561-4

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, *12*(2–3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Porter, A. L. ., Roper, A. T. ., Mason, F. A. ., & Banks, J. (1991). *Forecasting and Management of Technology* (D. F. Kocaoglu (ed.)). Wiley Series in Engineering and Technology Management.

Tong, Z., & Zhang, H. (2016). *A Text Mining Research Based on LDA Topic Modelling*. 201–210. https://doi.org/10.5121/csit.2016.60616