MASTER

MESTRADO EM ECONOMIA E GESTÃO DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO

MASTER´S FINAL WORK

DISSERTATION

THE EVOLUTION OF INTERNATIONAL BUSINESS RESEARCH: A CONTENT ANALYSIS OF EIBA'S CONFERENCE PAPERS (1999-2011)

JOÃO PEDRO MAGALHÃES DA ROCHA

NOVEMBER - 2020

MASTER
MESTRADO EM ECONOMIA E GESTÃO DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO

MASTER´S FINAL WORK
DISSERTATION

THE EVOLUTION OF INTERNATIONAL BUSINESS RESEARCH: A CONTENT ANALYSIS OF EIBA'S CONFERENCE PAPERS (1999-2011)

JOÃO PEDRO MAGALHÃES DA ROCHA

SUPERVISION:
VÍTOR DUARTE CORADO SIMÕES

NOVEMBER - 2020

AIB – Academy of International Business.

EIASM – European Institute for Advanced Studies in Management.

EIBA – European International Business Academy.

IB – International Business.

IBR – International Business Review.

JIBS – Journal of International Business Studies.

LDA – Latent Dirichlet Allocation.

NLP – Natural Language Processing.

ABSTRACT

This study seeks to analyse the evolution of the European International Business Academy's Annual Conferences between the years 1999 and 2011. A collection of the 2221 documents presented across the defined period was processed with the use of a computer-aided tool – the Latent Dirichlet Allocation, powered by artificial intelligence systems – to facilitate the content analysis. The study utilized the R software environment as the platform to apply the method, with the support of a group of compatible libraries to support in the pre-processing, topic modelling and result plotting stages of this study. The method was able to identify 30 underlying research topics across all the documents, and a label was manually assigned to each topic according to its overall theme. Results show an overall growth in number of papers presented, as well as new authors throughout the years, indicating an increase in degree of openness in the Association's Annual Conferences. Specific research topics have also shown to be more discussed across the documents than others, and research on the topics of 'Dynamic capabilities, resource-based view and firm internationalization' and 'Institutional approaches to International Business research and theory' showed to be trending in recent years of the Conferences. With little need for human intervention, this study was able to successfully apply an automated method to identify the Association's research themes and their evolution throughout the years, reducing researcher's bias and allowing for the efficient analysis of large volumes of text content.

KEYWORDS: Latent Dirichlet Allocation; International Business; EIBA; Content Analysis; Artificial Intelligence; Innovation.

JEL CODES: C10; F20; M16; O19; O30.

RESUMO

Este estudo busca analisar a evolução das Conferências Anuais da European International Business Academy entre os anos 1999 e 2011. Um conjunto de 2221 documentos apresentados durante o período foi processado com o uso de uma ferramenta computadorizada – Latent Dirichlet Allocation, potencializada por sistemas de inteligência artificial – para facilitar a análise de conteúdo. O estudo utilizou o sistema de software R como a plataforma de aplicação do método, com o apoio de bibliotecas compatíveis para auxiliar as etapas de pré-processamento, de modelagem de tópicos, e de plotagem de resultados deste estudo. O método foi capaz de identificar 30 tópicos de investigação subjacentes ao grupo de documentos, e um rótulo foi manualmente atribuído a cada tópico de acordo com seu tema geral. Resultados mostram um crescimento generalizado no número de artigos apresentados, bem como novos autores ao longo dos anos, indicando um aumento no grau de abertura das Conferências Anuais da Associação. Determinados tópicos de investigação também se mostraram mais discutidos nos documentos em relação a outros tópicos, e investigação nos tópicos de 'Capacidades dinâmicas, visão baseada em recursos e internacionalização de firma' e 'Abordagens institucionais à investigação e teoria em Negócios Internacionais' se mostraram tópicos de tendência em anos recentes das Conferências. Com pouca necessidade de intervenção humana, este estudo pôde aplicar com sucesso um método automatizado para identificar os temas de investigação da Associação e sua evolução ao longo dos anos, reduzindo o viés do investigador e permitindo uma análise eficiente de grandes volumes de conteúdo textual.

PALAVRAS-CHAVE: Latent Dirichlet Allocation; Negócios Internacionais; EIBA; Análise de Conteúdo; Inteligência Artificial; Inovação.

CÓDIGOS JEL: C10; F20; M16; O19; O30.

TABLE OF CONTENTS

TABLE OF FIGURES

1. Introduction

Though the field of International Business (IB) is young when compared to other research areas, with its first rise in interest dating back to the end of the Second World War (Engwall, Pahlberg & Persson, 2018), much has been done in terms of research in the field. Throughout the years, IB research has evolved from a field clustered around five categories (Elahee, 2007) to a much more complex subject, tackling on different issues related to the internationalization of business activities (Buckley & Lessard, 2005).

To assist in the communication, socialization and exchange of information among the community with an interest on IB, scholarly associations were created, and two can be identified as being the most relevant in the field: the Academy of International Business (AIB) and the European International Business Academy (EIBA). As the associations grew, they felt the need to provide their members with tools to interact, to publish their research and to keep them abreast of the main developments in the field; among other initiatives, this led AIB to the create the *Journal of International Business Studies* (Fayerweather, 1986), and EIBA to adopt the *International Business Review* as its official journal (EIBA, 2020a).

As EIBA has been operating for over 40 years, the Association launched a project led by John Cantwell, Philippe Gugler and Vítor Corado Simões with the objective of documenting the Association's history and providing more information about the evolution of IB research as a whole. An important vector in the project seeks to analyse EIBA's Annual Conference proceedings, since these can provide a longitudinal view about the evolution of IB research. As the Academy's archives include a large volume of text documents spanning for those 40 years, analysing such content may prove to be difficult and costly.

This study has, therefore, the objective of verifying how IB research presented at EIBA's Annual Conferences evolved throughout the years. For that, a computer-aided method is applied to identify underlying themes in EIBA's Annual Conference proceedings. A group of 2221 documents presented at the Association's Conferences between 1999 and 2011 (except for 2001) are analysed. The use of computer-aided tools in content analysis, especially in the IB field, has not been widely explored, an important exception being Piepenbrink and Gaur (2017). This study provides a successful

application of a computer-aided method to analyse a large volume of textual data by providing an encompassing view and analysis of EIBA's Conference proceedings in a relatively long time window.

The research question proposed in this study is contextualized and responded in five chapters. First, a contextualization of IB as a research field is made, explaining how its scholarly associations were created, and how their evolution led to the research question. Second, the topic modelling method is described, explaining how the textual data was processed and analysed. Third, the results generated from the model are presented. Fourth, the results are then interpreted and discussed. Finally, a synthesis of the findings, as well as the study's limitations and recommendations for future research are presented as a conclusion.

By utilizing a computer-aided approach to analyse the evolution of EIBA's Conference research, this study presents an innovative contribution towards the advancement of Science, Technology and Innovation. Such method applied a machine-learning algorithm to study the evolution of a scientific field, specifically allowing for a better comprehension of how EIBA's research has contributed to the field of IB by utilizing a recent technological development. As it makes use of a relatively modern method, this study provides an innovative approach to content analysis, and seeks to further disseminate the method as a valid tool to assist other studies.

2. FORMULATING THE RESEARCH QUESTION

To formulate the research question that will guide this study, it is necessary to first provide a historical contextualization of the IB field. The first section of this chapter will, therefore, address the evolution of IB Literature, while the second will present how the main scholarly associations of IB were founded and developed.

As IB progressively became an independent research field, it helped pave the way to the creation of scholarly associations responsible for assisting both IB scholars and professionals by acting as a gateway for the exchange of information and contacts, as well as promoting IB-related education, research and connected activities. Such associations played an important role in the development of IB as a research field.

Through the contextualization of the development of IB research as well as the creation and evolution of its main scholarly associations, it is possible to understand the circumstances that led to the formulation of this study's research question.

2.1. Evolution of International Business literature

IB as a formal research field is a relatively modern one. Its history began with the evolution of a multidisciplinary shift of attention towards Multinational Enterprise's (MNE) activities. The rise in IB studies can be dated back to the end of the Second World War, when an increasing interest in international business activities could be observed (Engwall et al, 2018). While interest in IB faced a rapid growth during its first years, it was still considered an international aspect of management, and not an independent field (Dunning, 2002).

In its early years, academics such as Lee C. Nehrt, J. Frederick Truitt and Richard W. Wright were able to categorize IB research developments until the year 1968 into five different areas: "International Business Strategy and Structure; Functional Aspects of International Business; International Business and National Environments; Cultural Factors; and Other" (Elahee, 2007: 148). This categorization facilitated the development of further studies and activities related to IB, resulting in a continuous evolution process of the field.

Throughout the years, IB research has changed and evolved by tackling issues specific to a period in time related to business activities and internationalization (Buckley &

Lessard, 2005). As time went on, divergences in the classification of IB studies can be observed, allowing for the identification of research trends that emerge from the main issues tackled by scholars and professionals in a given timeframe (Engwall et al, 2018). These different issues led to new IB approaches, resulting in a consistent cycle of improvement and expansion in the field.

While IB can be characterized as having an everchanging and evolving nature, research in the field has converged around a set of main authors throughout its history. Morrison and Inkpen (1991) identified the beginning of the consolidation of main authors and journals in the field. The most-cited works in IB were mainly published in the same journals throughout the years (Griffith, Cavusgil & Xu, 2008), and this feature can still be observed in recent years. This shows the importance of the mainstream journals, especially JIBS, in the evolution of the field (García-Lillo, Claver-Cortés, Marco-Lajara & Úbeda-García, 2019).

Historically, the main journals of IB have been anchored to partnered institutions, such as universities or professional associations (Morrison & Inkpen, 1991). These associations played a crucial role in the development of IB as a formal research field by providing a centralized place to promote and organize research and activities related to IB (Engwall et al, 2018) and to foster socialization and the development of joint research among their members.

2.2. International Business Research Associations

Throughout the development of IB research, some scholarly associations have been created. The main purpose of these associations is to promote IB research and related activities, while serving as a gateway to facilitate the communication, socialization, and exchange of information among researchers and professionals interested in the field. Although there are other associations, two are considered the most significant in IB research: AIB and EIBA.

The AIB was founded in 1959 in the United States of America, after a group of nineteen people with different backgrounds met the previous year seeking to create an association where they could address issues they had previously experienced (Fayerweather, 1986). Three main points were raised during the meeting: first, the members wanted to expand and strengthen the network of scholars to better exchange

information about IB; second, they intended to establish a centralized medium to organize and promote research activities; and, finally, they wanted to further professionalize and spread knowledge about IB (Fayerweather, 1986). In December 1959, the Association for Education in International Business was formally created, and, in 1972, changed its name to AIB. When the Association was first created, its efforts were put on IB education, and not research, but throughout the years that focus would change into a more research-oriented approach (Fayerweather, 1986).

As the years went by, the AIB grew to become the largest academic IB association in the world, having over 3000 members spread throughout the world as of 2020 (AIB, 2020). To provide "a good medium for publication of papers and a source of information" (Fayerweather, 1986: 23), the AIB launched the *Journal of International Business Studies* (*JIBS*), a decision that was quickly supported by the growth in the field of IB research, but also contributed to further develop the field. The journal was initially small and had a relatively high rejection rate (Fayerweather, 1986). In recent years, JIBS has been increasingly more open to the entry of first-time contributors and faces a rise in international research collaborations (Cantwell, Piepenbrink, Shukla & Vo, 2016). As of 2020, JIBS is ranked first among other IB journals (Google Scholar, 2020).

Although the AIB already had achieved an international level of activities, the European region also started developing an increasing interest in IB research. In the years 1973 and 1974, the European Institute for Advanced Studies in Management (EIASM), in cooperation with the Centre d'Enseignement Supérieur des Affaires, hosted two workshops with the objective of discussing the creation of a communications network in Europe for IB scholars (Simões, Cantwell & Gugler, forthcoming). These workshops, along with the support provided by the EIASM and the European Foundation for Management Development, let to discussions about creating a scholarly association to further develop the IB network in Europe (Simões, Cantwell & Gugler, forthcoming).

In 1974, fifteen years after AIB's foundation, the European International Business Association was founded in Brussels, and was officially renamed to European International Business Academy in 1995. It sought to become "a professional society for academics and practitioners with an interest in the growing field of International Business" (EIBA, 2020a, para. 1). The EIBA has the mission of becoming the "core

communication network for disseminating information and promoting international exchange in the field of International Business" (EIBA, 2020a, para. 2).

Unlike the AIB, EIBA did not launch its own official journal, but instead adopted, in 2000, an already existing journal: *International Business Review* (IBR). Through IBR, the Association "provides a forum for academics and professionals to share the latest developments and advances in knowledge and practice of international business" (IBR, 2020, para. 1). As it is more recent than JIBS, IBR can still be considered a smaller journal, but it is currently ranked second among all IB journals (Google Scholar, 2020), and serves as one of the main journals for publishing IB research developments.

2.3. Research Question

In 2014, EIBA hit a maturity milestone by operating for 40 years. Within its already long history, it is possible to find contributions from scholars and professionals from multiple generations and disciplines. With the goal of providing an overall perspective of EIBA's evolution by better documenting and describing its history, the Association is developing the EIBA History project. The project, led by John Cantwell, Philippe Gugler and Vítor Corado Simões, is also intended to draw from the collective work of the Association's members. Through the project, EIBA seeks to provide more information about IB history, allowing both scholars and professionals to better understand the historical path of the Academy's developments, contributing to the IB academic community.

Every year, EIBA hosts its Annual Conference that "features paper presentations, workshops, sessions and panels on new and on-going research and practice" (EIBA, 2020b, para. 1). Therefore, a large share of EIBA's community research developments is presented at its Annual Conferences. With thousands of individual works presented throughout the years, the proceedings archives are of great importance to the study of EIBA's history.

Much of the Association's history is archived in its Annual Conference history, as many developments in the IB field at both European and global levels stemmed from such conferences. This leads to this study's main research question: **How has the IB research presented at EIBA's Annual Conferences evolved throughout the years?**

Through EIBA's history project, much work has already been done to map and document the Association's developments throughout the years. Among other tasks inside the project, Vítor Corado Simões and his colleagues have already begun documenting EIBA's Conference Papers. Throughout the years 1982 and 1997, the conference proceedings are only available, for most years, on a paper format. Only from 1998 onwards were the papers published digitally. This study focuses, therefore, on the Conference Papers made available at EIBA's proceedings database, limiting the studied documents to those presented in the 1999-2011 period, excluding the year 2001, as it was not made available. As such, the number of papers presented amounts to 2221.

To answer the proposed question, this study makes use of topic modelling, a computer-aided method, to process the large volume of the Association's Conference papers. The preparation and application of the method for this study are explained in the following chapter.

3. METHOD

To properly study the evolution of EIBA's Annual Conferences' research themes, it is necessary to analyse the set of 2221 conference papers. Although there are many ways to tackle this issue, there is still a general lack of tools to assist in analysing data stored in textual formats (Piepenbrink & Gaur, 2017). There have, however, been relevant developments in the field of text processing research. Through Natural Language Processing (NLP), researchers were able to develop computer-aided tools that can extract and subsequently interpret meaning contained in a sequence of words (Chowdhary, 2020).

As proposed by Grimmer and Stewart (2013), there are several methods to process textual data through computers, differing mainly on the type of data to be analysed and what parameters define the algorithms. The text data analysed in this study consists of a large number of research papers with no explicitly known research themes, thus meaning that their themes must be identified through this study. As such, it is necessary to apply a language processing method that identifies previously unknown themes in text.

A popular area of NLP is topic modelling, a group of computer-aided tools capable of processing large volumes of textual documents and, more importantly, identifying its previously unknown themes or categories (Piepenbrink & Gaur, 2017). Through the use of topic modelling to process documents containing EIBA's conference papers, it is possible to identify the main research themes contained within the text articles, while at the same time avoiding the risks inherent to researcher's bias when categorizing textual data. This allows for the study of EIBA conference papers evolution through the analysis of its key themes throughout the years, thus fulfilling this research's main goal.

3.1. Topic Modelling and Latent Dirichlet Allocation

For this study, it is necessary to define the basic terms that compose the topic modelling vocabulary. A word is an item from a vocabulary, a document is a sequence of words, and a corpus is a collection of documents (Blei, Ng & Jordan, 2003).

As proposed by Dunning (1989), IB demands an increasingly diverse mix of factors and capabilities, making it desirable for IB research to approach issues from a multidisciplinary perspective. As such, EIBA's research developments have drawn from the contributions from different disciplines.

The corpus analysed in this study contains a large group of documents with different and initially unknown themes. In addition, the documents may contain a mix of themes to the interdisciplinary nature of IB research. This poses issues regarding the topic modelling method. First, the chosen method must be able to identify unknown themes in methods. Second, due to the documents' interdisciplinary nature, the method must also consider the possibility that the documents' contents may be grouped into more than one category.

In response to the issues, Grimmer and Stewart (2013) present a representation of different methodologies to analyse textual data. The most appropriate method capable of handling both unknown topics and the possibility of multiple topics being present in one document is the Latent Dirichlet Allocation (LDA). LDA is an NLP method in which "documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words" (Blei et al, 2003: 996). Through the use of artificial intelligence systems, the method can identify previously unknown topics in documents, as well as attributing topic probability scores for each document processed.

Applications of LDA in other areas have provided encouraging results when identifying and analysing large groups of textual data. In the field of Political Science, Grimmer and Stewart (2013) show that LDA is able to assist in classifying documents when properly validated. A different approach to LDA was able to identify Marketing-relevant information from user-generated information posted online (Tirunillai & Tellis, 2014; Guo, Barnes & Jia, 2017). Another application of the method has even shown the possibility of identifying themes in a digital vehicle for political activism, allowing for the study of the social behaviour of specific communities (Hagen, 2018). As such, LDA has shown to be a useful tool in past research in different fields, assisting researchers in identifying and further studying the evolution of themes in text documents.

Making use of the LDA approach on this study brings many benefits, as described by Piepenbrink and Gaur (2017). First, it is not necessary to have prior knowledge about the corpus' topics, allowing for the study of multiple papers without previous interaction. The method also allows the researcher to process a large group of texts at once, considerably speeding up the analysis process. It can also consider context while analysing words by utilizing the concepts of polysemy and synonymy, bringing more depth to the analysis.

Lastly, the method is able to process new documents in an already-fitted model. These benefits assist the researcher by "quickly and efficiently deriving the thematic structure of large amounts of text documents" (Maier, Waldherr, Miltner, Wiedemann, Niekler, Keinert, Pfetsch, Heyer, Reber, Häussler, Schmid-Petri & Adam, 2018: 93).

It is important, however, to bear in mind that LDA as a quantitative approach to analyse text data suffers from some limitations. LDA and other distributional models may identify degrees of semantic relations between words, but are not able to specify the kind of relation, entailing a possible loss of contextual information which might have been able to assist in the study (Lenci, 2008). LDA also assumes that topics are necessarily independent of each other, not considering the possibility of one topic being part of another larger topic, for example (Maier et al, 2018). Though LDA has this limitation, its use also brings multiple positive outcomes. First, the method is very efficient, as it does not require many resources and is able to handle large volumes of data in a fast manner. Second, other methods may not be validated as LDA is, which may limit certain works. Finally, a large group of libraries supporting the development of LDA-based applications is readily available for different platforms (Piepenbrink & Gaur, 2017).

Although no single quantitative model of language is fully correct, some of them can be quite useful, as they may significantly enhance a researcher's ability to analyse language (Grimmer & Stewart, 2013). It is important, however, to maintain a constant cycle of validation in order to make the best use of the method (Grimmer & Stewart, 2013; Maier et al, 2018).

To conduct a proper topic modelling via LDA, Piepenbrink and Gaur (2017) propose a 5-step routine composed of: corpus collection and definition; pre-processing of the corpus; topic modelling; topic model validation; and topic model analysis. For this research, a similar routine will be followed.

3.2. Corpus Collection and Definition

The corpus for this study considered all Competitive and Workshop papers presented to EIBA's annual conferences between the years 1999 and 2011, excluding the year 2001, since its papers are unavailable. Conference papers until 1998 were published in a paper format and would require efforts that would not fit this study's time constraints, while papers presented after 2011 were not yet made available in the Academy's database. In

total, 2221 papers were considered for this study. The respective files were available in three distinct formats: HTM, PDF and TXT. To read and convert the documents into a pure-text corpus, the R software environment version 4.0.3 (R Core Team, 2020) was used.

To import HTM files into the software, the htm2txt library (Park, 2017) was utilized, as it is able to read HTM files, a common web document format, and convert them into raw text data. To read and convert the PDF files into raw text, the pdftools library (Ooms, 2020) was selected, as it behaves in a similar way as the htm2txt library, being able to quickly extract textual data from PDF files. As for TXT files, these are already in a plain text format and therefore did not need to be converted. As such, the readtext library (Benoit, Obeng, Watanabe, Matsuo, Nulty & Müller, 2020) was used, as it provides tools that allow for reading and handling either plain or formatted text files.

Once all files were successfully read and imported into R, a corpus was created via the tm library (Feinerer & Hornik, 2019). This library is a powerful tool and has many different functions related to the practice of text mining, or the extraction of information from textual data. It allows for further work on the text data, as it formats all documents in the same standard and stores it in a single corpus object.

3.3. Definition of document unit

Once the corpus has been defined and properly created, the next step regards the definition of the document unit. This step defines what units will be considered for the textual analysis. For the purpose of this study, the entirety of the documents was considered, with the exception of irrelevant data identified and removed during the pre-processing stage of the analysis. By applying a pre-processing exercise on the corpus, the text data was cleaned up and structured for a proper interpretation by the model.

Though a paper's abstract may contain enough relevant information to properly identify a corpus' underlying themes, not all abstracts provide sufficient information. Such information must be, therefore, found elsewhere on the paper. On the one hand, this approach ensures that all the necessary information for the proper topic modelling will be present. On the other, much more data will need to be pre-processed and, if necessary, removed, as not all information is necessarily useful for this study, thus generating extra workload.

3.4. Pre-processing of the Corpus

The pre-processing stage of text data analysis consists of "applying methods for cleaning up and structuring the input text for further analysis" (Feinerer, Hornik & Meyer, 2008: 23). This step makes sure that all text data is properly cleaned up, removing unnecessary or unwanted content from the corpus.

For this research, the pre-processing stage consisted of multiple steps, as follows: (1) all non-alphabetic characters were removed via a standard regex function in R, leaving only words with alphabetic characters (A to Z); (2) all words were converted to lower-case to make sure the software would correctly analyse them; (3) all punctuation and special characters were removed; (4) all stopwords, or words with no meaning to the analysis, were removed according to the stopword dictionary provided in the tm library; (5) all words were stemmed, reducing them to their root form in order to consolidate related words into one single unit, through the use of Porter's stemming algorithm (Porter, 2006) provided in the tm library; and (6) only words with more than two characters were then considered.

The pre-processing helps remove unwanted data from the analysis process, as LDA is sensitive to differences in its corpus' structure, bringing significant results to the topic modelling, both in the definition of the optimal number of topics, as well as in the modelling itself (Denny & Spirling, 2017). A good pre-processing may require a considerable amount of resources, and it may take more effort than manually analysing the documents, depending on the corpus complexity. However, a properly defined pre-processing stage may be reused in other works of the same nature with little effort, significantly speeding up other analyses.

3.5. Topic Modelling

As LDA requires some human-input parameters before it is run, the first stage of the topic modelling is the definition of the optimal number of topics. This can be done in a few different ways, like human experimentation and validation, or mathematical estimation via the use of statistical models.

For the estimation of the ideal number of topics, this study makes use of the ldatuning library (Nikita & Chaney, 2020). This library is intended to facilitate the estimation of the best number of topics for the model through a visual representation of four different

estimation metrics. A maximum of 40 possible topics was defined as a parameter in the library in order to first explore the estimation's results, generating the plot shown in Figure 1. To guarantee future reproducibility of the results, a pre-defined seed was set as a parameter for the estimation, as results tend to vary when different seeds are set.
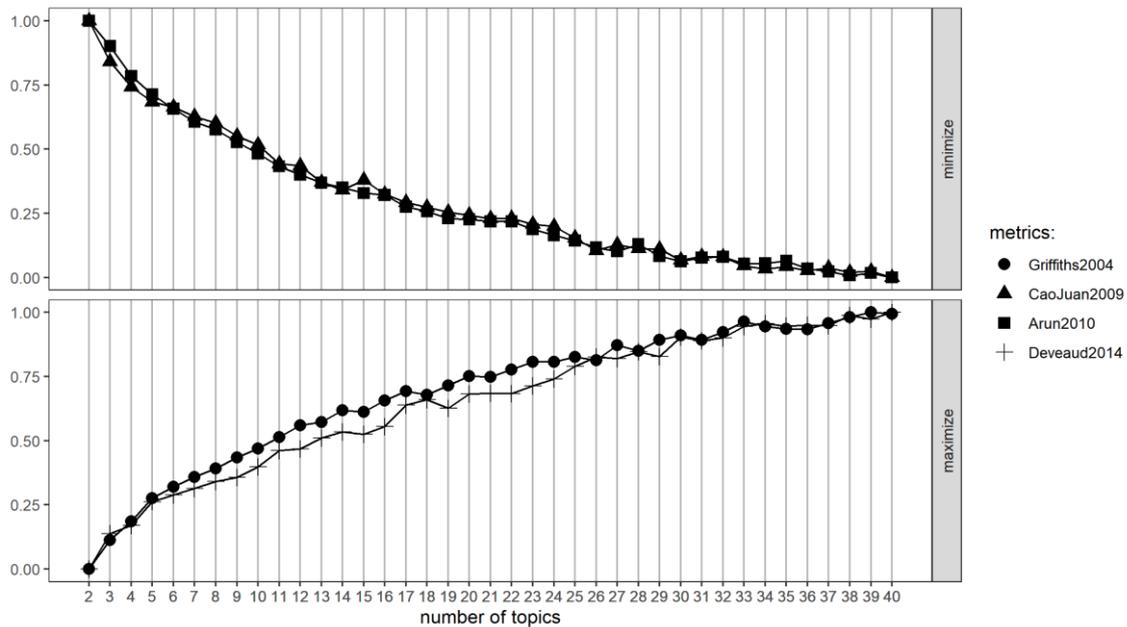
Figure 1 – Estimated Number of Topics. Source: LDA analysis by the author.

The plot presented in Figure 1 contains four different methodologies grouped in two metrics: one consists of maximizing the values to obtain the ideal result – CaoJuan2009 (Cao, Xia, Li, Zhang & Tang, 2009), and Arun2010 (Arun, Suresh, Veni Madhavan & Narasimha Murthy, 2010); the other consists of minimizing the values – Griffiths2004 (Griffiths & Steyvers, 2004) and Deveaud2014 (Deveaud, SanJuan & Bellot, 2014). An examination of the values in Figure 1 shows the number 30 as a possible candidate for optimal topic numbers, as there is little gain when raising the number, and lower values still show great volatility when changing values. For this estimation, a number of 40 topics is shown to be too high, as not much information is gained when raising the number of topics over 30, according to the four metrics. When the number of topics is too high, the model may no longer be semantically useful (Roberts, Stewart, Tingley, Lucas, Leder-Luis, Gadarian, Albertson & Rand, 2014), a concept explained in the following subchapter. Further tests were run by grouping the model in different numbers of topics, and the number 30 showed the most robust results.

Following the estimation, an LDA analysis is run, fitting a topic model with 30 different topics and a predefined seed to ensure the results can be reproduced should any change be made. For the analysis, the topicmodels library (Grün & Hornik, 2011) is used. This library contains a set of statistical methods for generating different topic models, including an LDA topic model.

Once the model is properly fitted, the results go through one final processing stage in order to create better visualizations for validation and discussion. Two important probability metrics are considered at this stage: word-topic probabilities, or *beta*, which are the probabilities of given terms being generated by a specific topic; and document-topic probabilities, or *gamma*, which denotes the probabilities of all terms in a given document being generated from a specific topic.

To better organize and visualize the model's results, a set of tools is used to assist in the process. The tidytext (Silge & Robinson, 2016) library is able to convert text data into a structured format, facilitating text mining applications. The dplyr (Wickham, François, Henry & Müller, 2020) library contains a function set aimed at manipulating data in an organized and user-friendly way. The ggplot2 (Wickham, 2016) library is then utilized to draw visual plots of the treated results to facilitate the data analysis. A matrix containing the *beta* probabilities of all terms in the 30 topics was generated, and a plot containing the top 10 terms by highest *beta* is found in Appendix 1.

3.6. Topic Model Analysis

To analyse and validate the topic model, Roberts et al (2014) propose a two-step validation approach to achieve a semantically interpretable topic. First, a topic must have a high *cohesion*, which indicates that words with higher probability should occur across multiple documents in the corpus. For this study, cohesion will be measured in two intervals: high and medium, indicating how well a topic's top words make sense in an article about that topic, as proposed by Piepenbrink and Gaur (2017). If a topic's top words tend to reoccur within documents, a high cohesion value is attributed.

Second, a topic must have *exclusivity*, which means a topic's top words should not reoccur as top words in other topics. In less exclusive topics, their top words commonly reappear as top words in other topics, while more exclusive topics' top words tend to not appear as top words in other topics of the model. For the model generated in this study, it

is possible to identify reoccurring top words across different topics, which, according to Roberts et al (2014), tends to lead to less semantically useful topics. It is still possible, however, to extract semantic meaning from the model, as the overlapping terms are very few and commonly used words in IB research, such as 'firm' or 'country'.

To analyse the model's cohesion, an assessment was made by evaluating the 20 documents with the highest gamma for each topic and checking whether their respective top words could be found in the selected documents. Only 6 out of 30 topics presented a medium cohesion score, as some of the top words did not reoccur frequently in those topics' documents. This assessment also allowed to manually assign a label to each topic by comparing its top words with the selected documents' contents and interpreting how that information could be grouped up in a broader topic label. Another form of evaluation done was the assessment of documents for each topic. The top 3 documents' abstracts were analysed and compared how well they fit the topic. Overall, the documents' contents fit the topics at hand, indicating a generally high assessment of cohesion[1].

According to Roberts et al (2014)'s definition, this model is considered semantically useful and can, therefore, provide meaningful information about EIBA's conference proceedings' underlying themes. The results identified in the model are described in the next chapter.

---

[1] One anomaly was identified during this assessment, however, as a paper presented in the 2009 Conference contained text in mostly Spanish. The model assigned the paper a high probability of belonging to a specific topic due to very specific English words contained within the document. The anomaly did not steer the results differently, as an assessment of other documents in the topic still indicate the existence of cohesion.

4. RESULTS

In this chapter, the results obtained in this study are presented. The data generated by the topic model is reflected in two main concepts: One refers to the word-topic probability – *beta* –, or the probability of an assigned word being generated in a given topic over another; the other is the document-topic probability –*gamma* –, or the probability of a given document belonging to one topic over the others. A summarized version of this study's results is also shown in Appendix 2, containing each topic's label, top words, documents and gamma values, as well as an assessment of cohesion. The 30 topics generated by the model alongside their respective labels are shown in Table I:

TABLE I

LIST OF TOPICS DERIVED FROM THE ANALYSIS

Topic	Label
1	Inter-organisational relationships, alliances and networks
2	Family firms and business in Asian countries (with special reference to the global drinks industry)
3	International taxation, finance and public policy issues
4	International Human Resource management
5	International strategy and R&D activities abroad
6	International marketing (branding and country of origin perceptions)
7	Internationalisation of firms from non-traditional locations
8	International aspects of innovation and technology
9	International Joint-Ventures
10	Foreign Subsidiary roles, autonomy and management issues
11	International marketing strategy and export performance
12	International finance issues, acquisitions and performance
13	Evolutionary perspectives on Organisational learning and knowledge sharing
14	Born Globals and International Entrepreneurship
15	Cross-border Mergers and Acquisitions
16	Strategic scope and multinationals' social responsiveness
17	Dynamic capabilities, resource-based view and firm internationalization
18	FDI, Trade and Catching-up
19	FDI: Spatial distribution, relocation and knowledge spillovers
20	MNEs' R&D activities: agglomeration, location patterns and spillover effects
21	Qualitative approaches to IB research and theorizing
22	FDI and regional integration in transition economies (with specific references to the banking industry)
23	Offshoring, service activities, and international survey methods
24	Development of scales in IB research
25	Qualitative IB research, business relationships and networks
26	Institutional approaches to IB research and theory
27	Cross-cultural issues
28	Entry mode choices
29	Company valuation, information disclosure and internet-based approaches
30	International corporate governance issues

Source: LDA analysis by the author.

By ordering the terms based on their beta probabilities, the top words for each topic can be identified. A similar approach may be applied to identify each topic's top documents. For that, the maximum *gamma* value for each document is identified, attributing a most probable topic to each document. As LDA models do not allow for topic correlation, documents may not reoccur in other topics when utilizing this approach. The analysis of the model's *beta* and *gamma* values allows for a better assessment of the results.

It is possible to assign papers to a specific topic by maximizing their *gamma* probabilities, indicating that a given paper has the highest probability of belonging to a specific topic. The number of papers assigned to each topic utilizing this approach is shown in Figure 2. Topics 10 (Foreign Subsidiary roles, autonomy and management issues), 20 (MNEs' R&D activities: agglomeration, location patterns and spillover effects) and 14 (Born Globals and International Entrepreneurship) contain the highest number of papers presented in EIBA's Conferences. The topics with the lowest number of papers presented are topics 2 (Family firms and business in Asian countries), 23 (Offshoring, service activities, and international survey methods) and 5 (International strategy and R&D activities abroad).
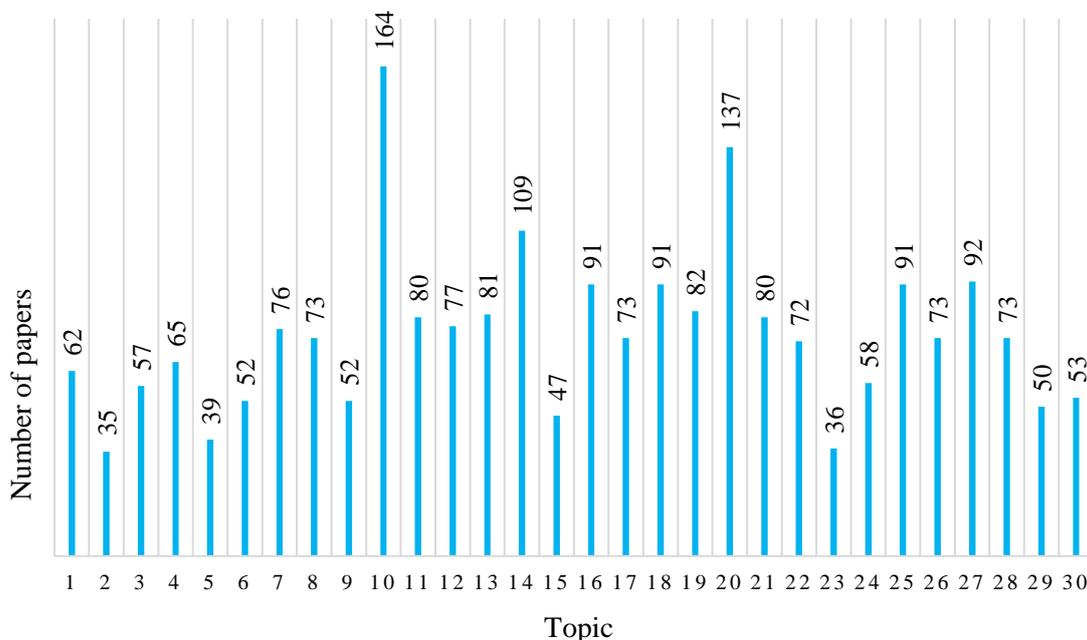
Figure 2 – Number of Papers per Topic by Highest *Gamma* Values. Source: LDA analysis by the author.

Another result derived from the analysis of each document's probabilities is the *popularity* of each topic, which provides information on how much it is addressed throughout the entire corpus. The *popularity* of a given topic is analyzed by applying an arithmetic mean over the *gamma* of all documents for that topic (Piepenbrink & Gaur, 2017). For this study, a uniform distribution would define a popularity of 3.33% for each topic, obtained by dividing 100% by 30, the number of topics in the model. Topics with a popularity above 3.33% tend to be more represented across the textual content of all documents, while topics below the value tend to be less represented. In this model, the most popular topic is Topic 12 – International finance issues, acquisitions and performance – with a popularity share of 5.84%, while the least discussed topic is Topic 5 – International strategy and R&D activities abroad –, with a share of 1.39%.

The findings also allow for the identification of topic trends over time. For this approach, the papers were grouped into three main time windows: years 1999 to 2002, excluding 2001; 2003 to 2006; and 2007 to 2011. The time windows were defined in accordance with the periodization defined for the EIBA history text. As displayed in Figure 3, the number of papers presented in the Annual Conferences has increased over the years.
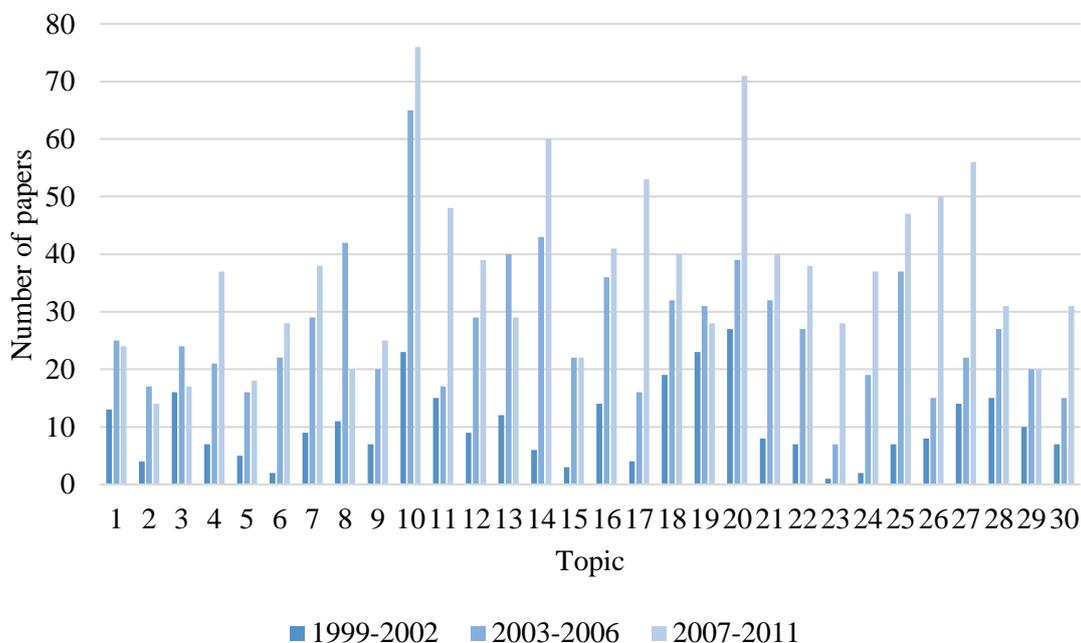
Figure 3 – Number of papers per topic in each time window. Source: LDA analysis by the author.

An overall increase in the distribution of the papers presented over time is observed, as shown by a growth in the number of papers presented for most topics throughout the years. Topics 3 (International taxation, finance and public policy issues), 8 (International aspects of innovation and technology), 13 (Evolutionary perspectives on Organizational learning and knowledge sharing) and 19 (FDI: Spatial distribution, relocation and knowledge spillovers), on the other hand, showed a decrease in number of papers during the 2007-2011 time window in relation to 2003-2006.

The proportion of papers presented in each time window is shown in Figure 4. A large number of topics displayed a significant growth in interest in the 2007-2011 time window, as indicated by the higher proportion of papers presented during that period, representing approximately 50% of all presented papers. On the other hand, the weight of presented papers was much lower between years 1999 and 2002, being responsible for approximately 14% of the total of papers, indicating a much lower publication rate in earlier years of the Association.

A notable increase in the proportion of papers presented on Topics 23 (Offshoring, service activities, and international survey methods), 17 (Dynamic capabilities, resource-based view and firm internationalization) and 26 (Institutional approaches to IB research and theory) in 2007-2011 was observed. When compared to the previous time windows, the three topics display a very rapid growth spurt in more recent. Although research on the three topics was already developed in previous years, the sudden emergence in number of papers may indicate the existence of research trends in specific time windows. A further, more detailed analysis of the existence of trending subjects is presented in the following chapter, along with the discussion of the results presented in this chapter.
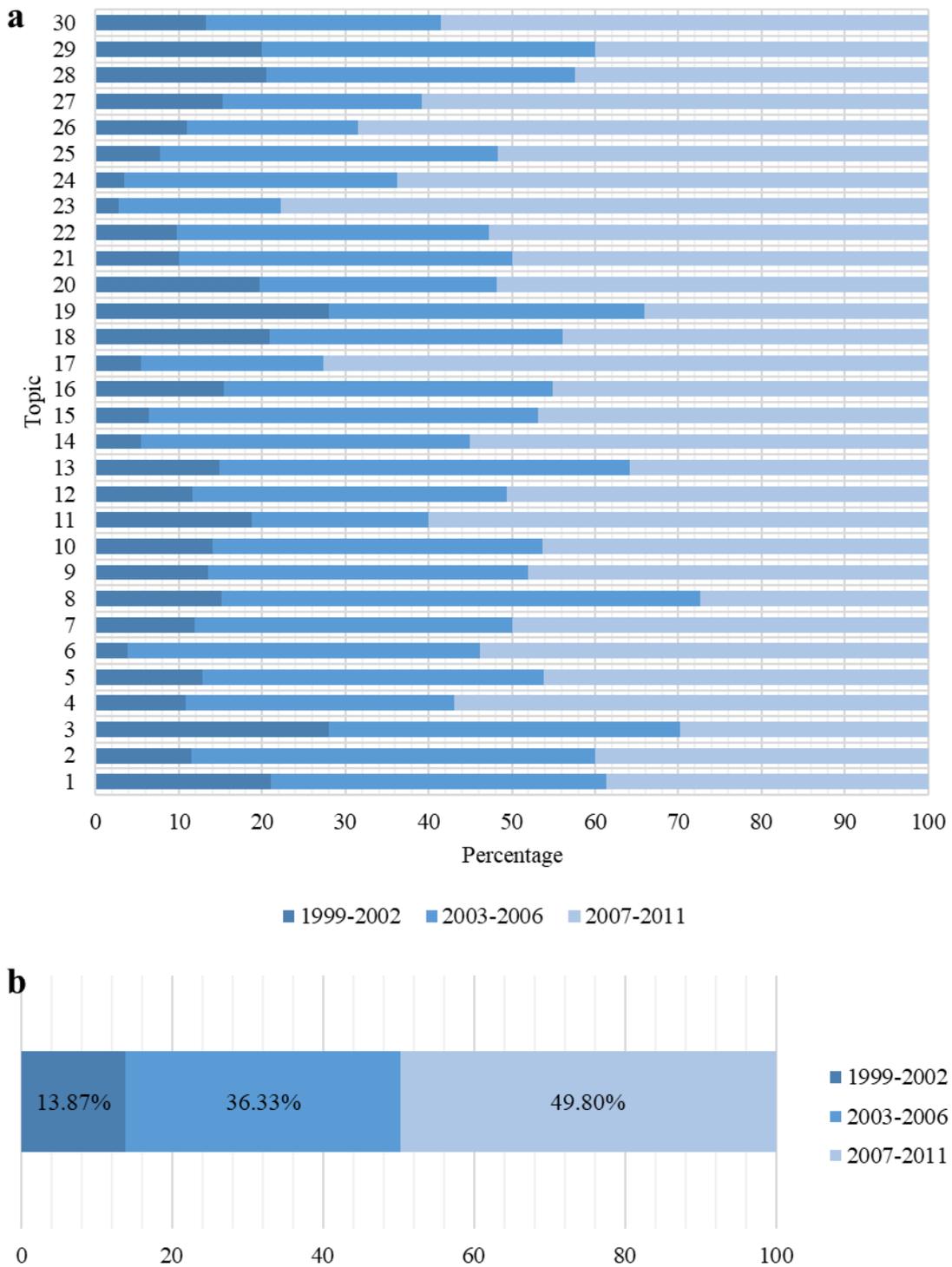
Figure 4 – **a** Percentage of papers per topic over time periods. **b** Percentage of all papers over time periods. Source: LDA analysis by the author.

5. Discussion

By analysing the results generated by the model, it is possible to identify different trends in EIBA's Conference Paper themes over the years. The proposition that IB should be an interdisciplinary field (Dunning, 1989) is corroborated by the results obtained from analysing each topic's theme and its top words. Multiple disciplines, such as Economics, Management, Finance, Marketing or Psychology can be identified in the topics, indicating a high degree of interdisciplinarity.

It is possible to observe a rapid increase in the number of papers presented in EIBA's Annual Conferences throughout the years. According to Cantwell et al (2016), the JIBS, AIB's official journal, tends to be increasingly open to the entry of new researchers interested in IB. Although it is another Academy's journal, the increasing openness in JIBS can also be observed in EIBA's research patterns. The documents analysed in this study show that the number of papers presented in the Association's Annual Conferences faced an overall increase each year. The number of new authors also grew throughout the years, indicating a higher degree of openness to new researchers in the Association's Conferences, reflecting the behaviour identified in IBR (Rialp, Merigó, Cancino & Urbano, 2019). This is important not only for the growth of EIBA as a scholarly association of IB, but also for the further dissemination and consolidation of the field of IB research.

Although the number of papers presented in EIBA's Conferences increased as time went on, it is also important to analyse the proportion of each topic in specific time frames to better comprehend the evolution of the Association's research. By studying how EIBA's Conference research evolved over the years, it is possible to identify possible research trends in given time periods. Figure 5 shows a visualization of the 10 topics with highest proportion of papers in each time window. When looking at the different time windows, it is visible that the 10 highest topics have changed throughout the years, and time-specific research trends are identified.

Topic 1: Inter-organisational relationships, alliances and networks
Topic 3: International taxation, finance and public policy issues
Topic 8: International aspects of innovation and technology
Topic 10: Foreign Subsidiary roles, autonomy and management issues
Topic 11: International marketing strategy and export performance
Topic 13: Evolutionary perspectives on Organisational learning and knowledge sharing
Topic 14: Born Globals and International Entrepreneurship
Topic 16: Strategic scope and multinationals' social responsiveness
Topic 17: Dynamic capabilities, resource-based view and firm internationalization
Topic 18: FDI, Trade and Catching-up
Topic 19: FDI: Spatial distribution, relocation and knowledge spillovers
Topic 20: MNEs' R&D activities: agglomeration, location patterns and spillover effects
Topic 21: Qualitative approaches to IB research and theorizing
Topic 25: Qualitative IB research, business relationships and networks
Topic 26: Institutional approaches to IB research and theory
Topic 27: Cross-cultural issues
Topic 28: Entry mode choices

Figure 5 – Weight of top ten topics per time period. Source: LDA analysis by the
author.

Although the number of papers presented in the Association's Annual Conferences
experienced an increase throughout the years, it is possible to identify a decline in interest
in research on specific topics. Topic 18 (FDI, Trade and Catching-up) has seen its
proportion among the top 10 topics drop throughout the years, as it was responsible for

10.61% of all papers among the top 10 topics in 1999-2002, but dropped to 7.38% during 2007-2011. Even though the number of Conference papers about the topic rose throughout the years, this decrease in proportion indicates that there has been a higher interest in developing research on other topics.

Historically, specific research topics have played a very important role in further accelerating the rhythm of publications in IBR (Rialp et al, 2019). Although EIBA's Conference papers are not all necessarily published in IBR, key subjects still helped develop the Association's research. In earlier time windows, Topics 17 (Dynamic capabilities, resource-based view and firm internationalization) and 26 (Institutional approaches to IB research and theory) were not among the topics with highest number of papers. However, during 2007-2011, both topics saw a rapid growth in interest, and quickly appeared among the top 10 topics in that period. Paired with their overall increase in proportion in recent periods, as shown in Figure 4a, topics 17 and 26 emerge as trending research themes in the 2007-2011 window.

On the other hand, topics 1 (Inter-organisational relationships, alliances and networks), 3 (International taxation, finance and public policy issues), 19 (FDI: Spatial distribution, relocation and knowledge spillovers) and 28 (Entry mode choices) displayed a steep decline in proportion of papers presented over the years. Topics 1, 3 and 19 were no longer part of the top 10 topics after 1999-2002, as confirmed by the drop in absolute number of papers on those topics in 2007-2011 observed in Figure 3. Topic 28, however, did not face such a decline, which indicates that, although its number of papers has increased, other topics grew at a faster rate and thus showed displayed a higher proportion of papers.

By analysing the *popularity* of the topics across the corpus, it is possible to verify how the documents' contents tend to discuss each topic. A topic's *popularity* is obtained by calculating the average *gamma* probabilities for each topic. For this discussion, the 10 most popular topics are analysed, represented from highest to lowest popularity. All 10 topics observed have an average gamma value above 3.33%, indicating that they are, on average, more discussed across the corpus than other topics with lower *gamma*. Results of this analysis are shown in Figure 6. The results show that the 10 topics presented received the highest attention across the entire corpus.
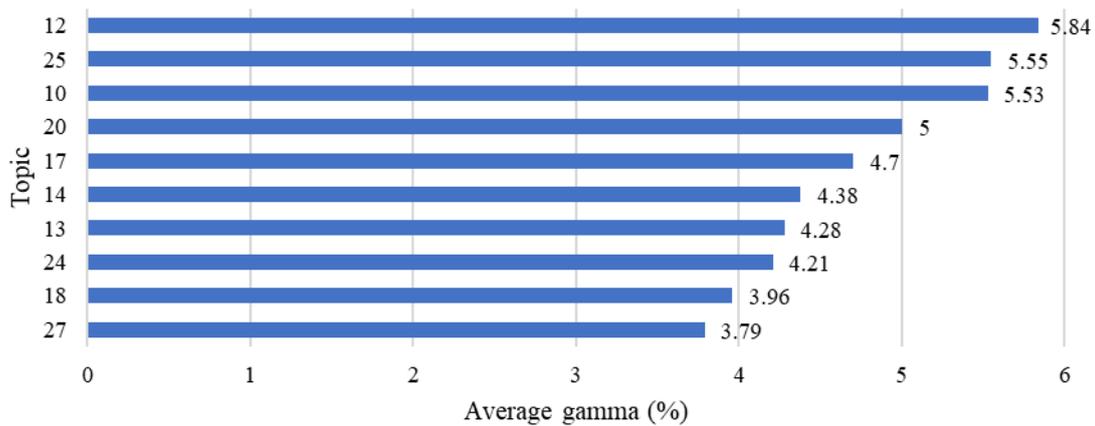
Topic 12: International finance issues, acquisitions and performance
Topic 25: Qualitative IB research, business relationships and networks
Topic 10: Foreign Subsidiary roles, autonomy and management issues
Topic 20: MNEs' R&D activities: agglomeration, location patterns and spillover effects
Topic 17: Dynamic capabilities, resource-based view and firm internationalization
Topic 14: Born Globals and International Entrepreneurship
Topic 13: Evolutionary perspectives on Organisational learning and knowledge sharing
Topic 24: Development of scales in IB research
Topic 18: FDI, Trade and Catching-up
Topic 27: Cross-cultural issues

Figure 6 – Top 10 most popular topics by average *gamma*. Source: LDA analysis by the author.

Among the most popular topics, it is also possible to identify topic 17, one of the trending themes in 2007-2011. This indicates that even though the topic did not contain an overall high number of papers, as shown in Figure 2, it was still highly popular across the entire corpus, solidifying the topic as very important in the Association's research history.

As years went on, the number of EIBA's Conference Papers presented experienced a constant overall growth. This points to an increasing degree of openness of the Association, as new authors develop research for the Conference every year. This openness is important, as it allows for the entry of new researchers in the field of IB research, and, therefore, further consolidating EIBA as one of the world's highest regarded scholarly association of International Business.

The method utilized in this study made possible for the identification of the research subjects discussed across all documents and allowed for the analysis of how EIBA's conference research has evolved throughout the years.

6. CONCLUSION

This concluding chapter is structured in four parts: First, a synthesis of the findings is presented; secondly, the contributions of this study to IB research and practice are discussed; the third part points the main limitations of the research; finally, suggestions for further research are presented.

6.1. Synthesis of Results

This dissertation was focused on performing an artificial intelligence approach to analyse a large number of textual documents. It made use of the R software (R Core Team, 2020), and benefited from its academic community-created libraries as the environment to read, pre-process and process the documents. Since further information is continuously generated through research, it is important to develop and utilize such tools to assist in performing content analysis on the large volumes of data constantly made available.

Through the use of a topic modelling approach, it was possible to identify and analyse 30 underlying topics in a corpus of 2221 documents in varying formats (PDF, HTM and TXT), from the years of 1999 until 2011, not including papers from the year 2001, for the reasons provided in section 3.2 above.

Little human intervention was necessary during the process, allowing for an efficient processing of the textual data (Maier et al, 2018). The 30 topics identified are part of multiple disciplines that permeate the field of International Business. The model was able to identify the top words belonging to each topic, as well as the probability of any given document containing the contents of a topic.

Results were found to be both exclusive, indicating low reoccurrence of topic words from one topic in another, and cohesive, indicating the common reoccurrence of top words from one topic within all of its documents; they allowed for the extraction of semantically useful information from the model (Roberts et al, 2014).

As documents from multiple years were analysed in the model, it was possible to better understand how the European International Business Academy's Conference papers evolved throughout the period under analysis. The paper distribution over time shows an overall increase in papers in all topics as years went on, indicating an increasing

degree of openness in the Association. Analysis of the model also allowed for the identification of trending research subjects throughout the years.

The number of Conference papers on the subjects of topics Papers on topics 17 (Dynamic capabilities, resource-based view and firm internationalization) and 26 (Institutional approaches to IB research and theory) went through a rapid growth in the 2007-2011 period, pointing towards the emergence of the two topics as trending during the period.

6.2. Contributions of this Study

This study contributes to the IB field in multiple facets. First, it provides a successful application of LDA as a computer-aided method for the analysis of large volumes of textual data, allowing for an efficient identification of underlying themes with minimum human intervention needed. As the method is still not widely used, this study aims to help further disseminate topic modelling as a valid content analysis tool.

Second, this study has provided important information for the documentation of EIBA's history, as the analysis of the Association's Conference proceedings allows for a deeper understanding of how its research has evolved throughout the years. With a sample of 2221 documents, this study was able to identify 30 underlying topics present across all Conference Papers, and has proved to be a useful tool for the documentation of the Association's history.

Third, this research was able to identify the existence of important conferences that directed how IB research was developed, leading to the emergence of trends in the field. Identifying and studying such driving events is important, as relevant information may be extracted from this exercise, such as the relevance of geographical location in the number of publications in a Conference.

Lastly, this study provides an innovative approach to the content analysis of large groups of documents. Other approaches, such as bibliometric analyses, can extract relevant information from such documents, but suffer from certain limitations that LDA does not, such as requiring longer periods of time to conduct the research, or requiring constant human intervention, requiring more resources and potentializing the impact of researcher's bias. In addition, this study was conducted on a group of conference

proceedings instead of papers published in a journal, and provides a broader view of developments in conference research in IB.

6.3. Limitations

Although topic modelling may provide meaningful information for content analysis, there are specific limitations to the use of the method. First, the Latent Dirichlet Allocation used in this study implies that topics are independent of each other (Maier et al, 2018), which does not allow for the analysis of certain topics as being part of a larger theme. Second, the pre-processing stage of the study, necessary to format and clean up data for the model, can require a high amount of human intervention, which may, in some cases, cost too many resources when documents of different formats are introduced, something common in a temporal analysis in which no file standard has been adopted early on. Finally, this study could only conduct the analysis on Conference Papers presented between the years 1999 and 2011, excluding 2001 due to an unavailability of data. For a better understanding of the evolution of EIBA's research history, it is important to conduct a similar approach on documents from earlier and later years as well.

6.4. Suggestions for Future Studies

For future studies, it is recommended that researchers conduct a topic modelling approach on documents from earlier years of the Association, if made available, as well as more recent years, as it can provide a broader understanding of EIBA's Conference proceedings evolution. Topic modelling methods may also be applied to specific fragments of EIBA's conference proceedings, such as abstracts, introductions or conclusions, to better understand how the method is able to identify underlying themes in different sections of documents, as well as provide an even faster approach of analysing the evolution of research themes in the Academy's conferences.

It may be relevant to conduct a topic modelling analysis to compare the results from this study with other research published in journals, namely JIBS and IBR, to provide a broader view of how IB research has evolved in both conferences and journal publications. Findings from this study may also be compared to those of other IB literature adopting LDA as their content analysis method, in order to better figure out how different researches approach and validate the method and how their results behave under different contexts.

Similar methods may also be compared to improve our understanding about research themes. The use of methods that allow for topic correlation, for example Correlated Topic Models (Blei & Lafferty, 2005), may provide results that allow for the study of the evolution of research on both macro and microscopic perspectives. Efforts to develop a standardized and efficient way of conducting the pre-processing step of the text analysis may provide significant support in further spreading the use of computer-aided approaches in content analysis. It is important, however, to ensure a constant process of validation and, when necessary, improvement when utilizing quantitative models to analyse language in order to correctly extract useful information from textual data (Grimmer & Stewart, 2013; Maier et al, 2018).

The field of the Economy and Management of Science, Technology and Innovation contains a large history of research publications on many different subjects. As such, a considerable volume of textual documents can be found. Through the application of LDA, similar works may be developed in the field to better understand the evolution of Science, Technology and Innovation and assist in related activities, such as the prospection of new technologies, or the analysis of a group of patents, for example.

By utilizing a computer-aided approach to analyse the evolution of EIBA's Conference research, this study presents a contribution towards the advancement of Science, Technology and Innovation. Such method allowed for a better comprehension of how EIBA's research has approached the field of IB by utilizing a recent technological development, providing an innovative approach to the analysis of research documents. As it makes use of a relatively modern method, this study provides an innovative approach to content analysis, and seeks to further disseminate the method as a valid tool to assist future studies.

REFERENCES

Academy of International Business. (2020). AIB Statistics. Retrieved October 25, 2020, from https://aib.msu.edu/statistics.asp.

Arun, R., Suresh, V., Veni Madhavan, C. E., & Narasimha Murthy, M. N. (2010). On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In: *Advances in Knowledge Discovery and Data Mining* (pp. 391–402). Springer Berlin Heidelberg.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3: 993–1022.

Blei, D. M., & Lafferty, J. D. (2005). Correlated Topic Models. *NIPS '06: Advances in Neural Information Processing Systems*, 18.

Benoit, K., Obeng, A., Watanabe, K., Matsuo, A., Nulty, P., & Müller, S. (2020). readtext: Import and Handling for Plain and Formatted Text Files. R package version 0.80. Retrieved October 14, 2020, from https://cran.r-project.org/package=readtext.

Buckley, P. J., & Lessard, D. R. (2005). Regaining the edge for international business research. *Journal of International Business Studies*, 36(6), 595–599.

Cantwell, J., Piepenbrink, A., Shukla, P., & Vo, A. (2016). The changing landscape of JIBS authorship. *Journal of International Business Studies*, 47(7), 749–777.

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7–9), 1775–1781.

Chowdhary K. R. (2020) Natural Language Processing. In: *Fundamentals of Artificial Intelligence*. Springer, New Delhi.

Denny, M. J., & Spirling, A. (2017). Text pre-processing for unsupervised learning: Why it matters, when it misleads, and what to do about it. New York University. *Political Analysis*, 26(2), 168–189.

Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document Numérique*, 17(1), 61–84.

Dunning, J. H. (1989). The Study of International Business: A Plea for a More Interdisciplinary Approach. *Journal of International Business Studies*, 20(3), 411–436.

Dunning, J. H. (2002). Perspectives on International Business Research: A Professional Autobiography Fifty Years Researching and Teaching International Business. *Journal of International Business Studies*, 33(4), 817–835.

Elahee, M. (2007). Lee C. Nehrt (1926-): a pioneering international business scholar. *European Business Review*, 19(2), 142–159.

Engwall, L., Pahlberg, C., & Persson, O. (2018). The development of IB as a scientific field. *International Business Review*, 27(5), 1080–1088.

European International Business Academy. (2020a). Mission. Retrieved October 24, 2020, from https://www.eiba.org/r/mission.

European International Business Academy. (2020b). Future Conferences. Retrieved October 23, 2020, from https://www.eiba.org/r/future-conferences.

Fayerweather, J. (1986). A History of the Academy of International Business from Infancy to Maturity: The First 25 Years. Retrieved October 21, 2020, from https://www.aib.world/wp-content/uploads/2019/04/John-Fayerweather-First-25-Years-AIB-History-parts-1-and-2.pdf.

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, 25(5), 1–54.

Feinerer, I., & Hornik, K. (2019). Tm: Text Mining Package. R package version 0.7-7. Retrieved October 14, 2020, from https://cran.r-project.org/package=tm.

García-Lillo, F., Claver-Cortés, E., Marco-Lajara, B., & Úbeda-García, M. (2019). Identifying the 'knowledge base' or 'intellectual structure' of research on international business, 2000–2015: A citation/co-citation analysis of JIBS. *International Business Review*, 28(4), 713–726.

Google Scholar. (2020). Top 20 publications matching international business. Retrieved October 25, 2020, from

https://scholar.google.com/citations?hl=en&view_op=search_venues &vq=international+business.

Griffith, D. A., Cavusgil, S. T., & Xu, S. (2008). Emerging themes in international business research. *Journal of International Business Studies*, 39(7), 1220–1235.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Supplement 1), 5228–5235.

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3), 267–297.

Grün, B., & Hornik, K. (2011). topicmodels: An R Package for Fitting Topic Models. Journal of Statistical Software, 40(13), 1–30.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.

Hagen, L. (2018). Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models? *Information Processing & Managemen*t, 54(6), 1292–1307.

International Business Review. (2020). International Business Review. Retrieved October 24, 2020, from https://www.elsevier.com/journals/personal/international-business-review/0969-5931.

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, 20(1), 1–31.

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.

Morrison, A. J., & Inkpen, A. C. (1991). An Analysis of Significant Contributions to the International Business Literature. *Journal of International Business Studies*, 22(1), 143–153.

Nikita, M., & Chaney, N. (2020). ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters. R package version 1.0.2. Retrieved October 14, 2020, from https://cran.r-project.org/package=ldatuning.

Ooms, J. (2020). pdftools: Text Extraction, Rendering and Converting of PDF Documents. R package version 2.3.1. Retrieved October 14, 2020, from https://cran.r-project.org/package=pdftools.

Park, S. (2017). htm2txt: Convert Html into Text. R package version 2.1.1. Retrieved October 14, 2020, from https://cran.r-project.org/package=htm2txt.

Piepenbrink, A., & Gaur, A. S. (2017). Topic models as a novel approach to identify themes in content analysis: The example of organizational research methods. Paper presented at 77th Annual Meeting of the Academy of Management, AOM 2017, Atlanta, United States.

Porter, M. F. (2006). An algorithm for suffix stripping. *Program*, 40(3), 211–218.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved October 24, 2020, from http://www.R-project.org/.

Rialp, A., Merigó, J. M., Cancino, C. A., & Urbano, D. (2019). Twenty-five years (1992–2016) of the International Business Review: A bibliometric overview. *International Business Review*, 28(6), 101587.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, 58(4), 1064–1082.

Silge, J., & Robinson, D. (2016). tidytext: Text Mining and Analysis Using Tidy Data Principles in R. Journal of Statistical Software, 1(3).

Simões, V. C., Cantwell, J., & Gugler, P. (forthcoming). *The History of the European International Business Academy (EIBA)*, Chapter 1, first draft. Mimeo.

Tirunillai, S., & Tellis, G. J. (2014). Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation. *Journal of Marketing Research*, 51(4), 463–479.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4.

Wickham, H., François, R., Henry, L., & Müller, K. (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.2. Retrieved October 24, 2020, from https://cran.r-project.org/package=dplyr.

APPENDICES

Appendix 1: 30 Research Topics and Respective Top Ten Terms by *Beta* values. Source:
LDA analysis by the author.
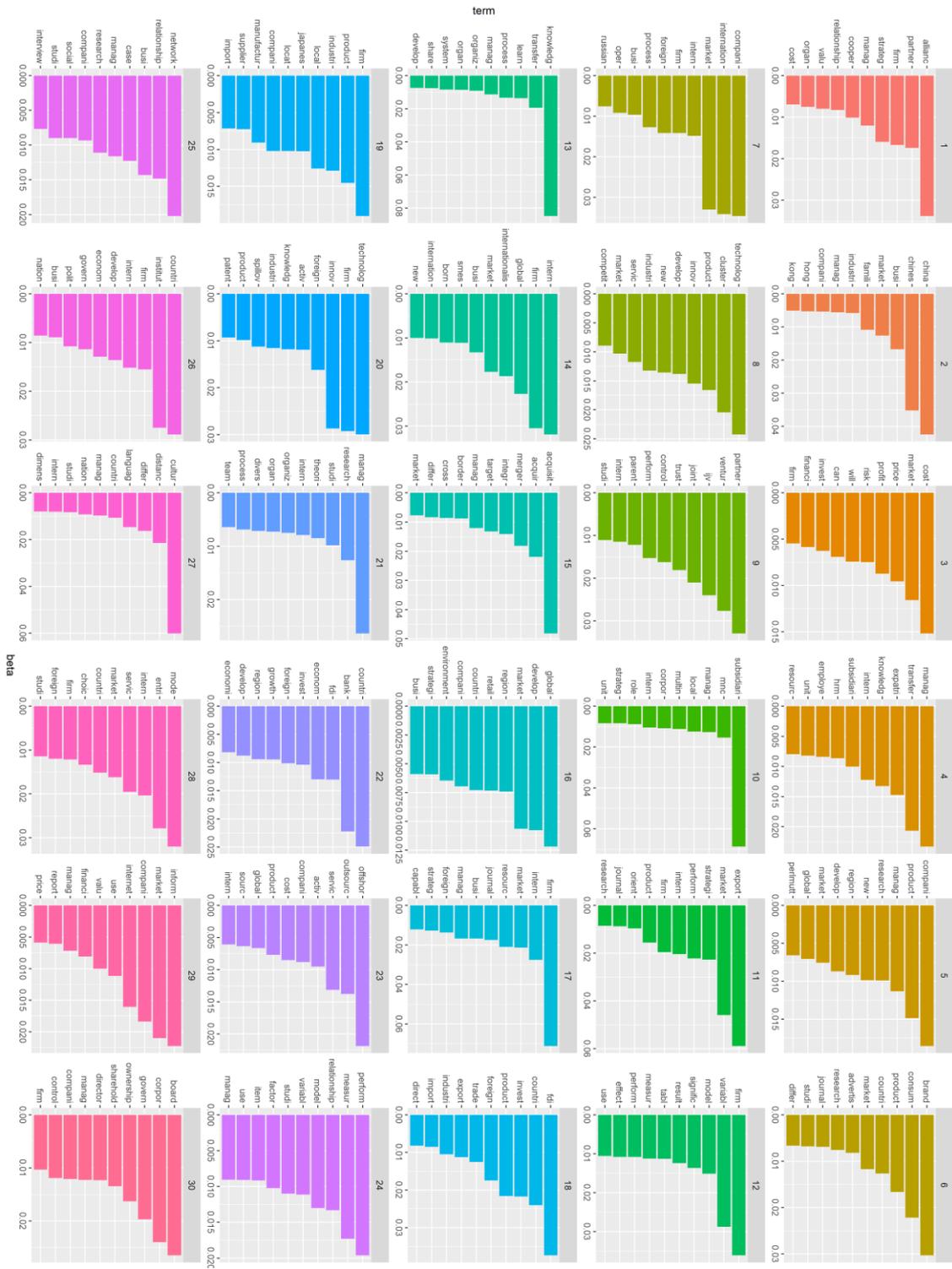
Appendix 2: Summarized Table of Results Obtained with 30-Topic Model. Source: LDA analysis by the author.

Topic Number	Label	Top 10 words	Assessment of cohesion	Top 3 articles and gamma (%)	Assessment of top 3 articles	Average gamma (%)
1	Inter-organisational relationships, alliances and networks	alliance, partner, firm, strategi, manag, cooper, relationship, valu, organ, cost	High	Marschan & Lehtoonen (2004) – 75 / Ouden & Bell (2003) – 72 / Kudic & Banaszak (2009) – 70	High / High / High	2.81
2	Family firms and business in Asian countries (with special reference to the global drinks industry)	china, chines, busi, market, famili, industri, manag, compani, hong, kong	High	Forte&Brandão (2006) – 89 / Lopes (2007) – 69 / Ibeh et al (2002) – 63	Low / Low / Medium	1.61
3	International taxation, finance and public policy issues	cost, market, price, profit, risk, will, can, invest, financi, firm	High	Marques & Mano (2009) – 87 / Forte & Brandão (2005) – 80 / Marques (2009) – 76	High / High / High	3.15
4	International Human Resource management	manag, transfer, expatri, knowledg, intern, subsidiari, firm, employe, unit, resourc	High	Bjorkman et al (2009) – 73 / Suutari et al (2009) – 70 / Suutari & Makela (2006) – 67	High / High / High	2.62
5	International strategy and R&D activities abroad	compani, product, manag, research, new, region, develop, market, global, perform	Medium	Schmid & Grosche (2009) – 70 / Schmid & Machulik (2005) – 64 / Schlegelmilch (2003) – 64	High / High / High	1.39
6	International marketing (branding and country of origin perceptions)	brand, consum, product, countri, market, advertis, research, journal, studi, differ	High	Lai & Hung (2010) – 91 / Pregsamatz & Ofstad (2005) – 91 / Gabrielsen et al (2005) – 90	High / High / High	2.04
7	Internationalisation of firms from non-traditional locations	compani, internation, market, intern, firm, foreign, process, busi, oper, russian	Medium	Vissak (2008) – 83 / Vissak (2009) – 82 / Kovacs et al (2011) – 77	High / High / High	3.65
8	International aspects of innovation and technology	technolog, cluster, product, innov, develop, new, industri, servic, market, competi	High	Bijaoui (2003) – 96 / Edman (2004) – 89 / Wegberg (2000) – 84	High / High / High	3.28
9	International Joint-Ventures	partner, ventur, jjv, joint, trust, control, perform, patent, intern, studi	High	Nguyen (2008a) – 92 / Nguyen (2008b) – 92 / Ali & Larimo (2006) – 85	High / High / High	2.08
10	Foreign Subsidiary roles, autonomy and management issues	subsidiari, mnc, manag, local, multin, corpor, intern, role, strategi, unit	High	Manolopoulos (2006) – 92 / Schmid et al (2000) – 84 / Morschett et al (2010) – 78	High / High / High	5.53
11	International marketing strategy and export performance	export, market, strategi, perform, intern, firm, product, orient, journal, research	High	Wheeler & Taggart (1999) – 96 / Sousa & Alserhan (2002) – 92 / Stoian et al (2010) – 87	High / High / High	3.72
12	International finance issues, acquisitions and performance	firm, variabl, model, signific, result, tabl, measur, perform, effect, use	Medium	Gioia (2003) – 74 / Eckert et al (2008) – 69 / Eckert et al (2010) – 63	High / High / Medium	5.84
13	Evolutionary perspectives on Organisational learning and knowledge sharing	knowledg, transfer, learn, process, manag, organiz, organ, system, share, develop	High	Borghoff (2007) – 88 / Nieminen (2004) – 84 / Hakanson (2004) – 84	High / High / High	4.28
14	Born Globals and International Entrepreneurship	intern, firm, global, internationalis, market, busi, smes, born, internation, new	High	Evers (2007) – 88 / Schlange & Lehmann (2004) – 79 / Madsen et al (2008) – 77	High / High / High	4.38
15	Cross-border Mergers and Acquisitions	acquisit, acquir, merger, integr, target, manag, border, cross, differ, market	High	Chevalier & Redor (2011) – 71 / Chevalier & Redor (2005) – 70 / Hamberg et al (2009) – 66	High / High / High	1.54

No.	Topic	Keywords	Level	References	Rating	Score
16	Strategic scope and multinationals' social responsiveness	global, develop, market, region, retail, countri, compani, environment, strategi, busi	Medium	Shanker (2006) – 92; Rugman (2003) – 82; Sammartino (2007) – 73	High; High; High	3.34
17	Dynamic capabilities, resource-based view and firm internationalization	firm, intern, market, resource, journal, busi, manag, foreign, strategi, capabl	Medium	Cellard & Prange (2007) – 72; Thomas et al (2005) – 66; Clarke et al (2010) – 66	High; High; High	4.70
18	FDI, Trade and Catching-up	fdi, countri, invest, product, foreign, trade, export, industri, import, direct	High	Haar (2007) – 92; Damijan & Rojec (2004) – 89; Gorynia et al (2007) – 84	High; High; High	3.96
19	FDI: Spatial distribution, relocation and knowledge spillovers	firm, product, industri, local, japanes, local, compani, manufactur, supplier, import	Medium	Puig & Marques (2009) – 74; Sannasse & Pearce (2004) – 73; da Rocha et al (2006) – 68	Medium; Medium; Medium	3.29
20	MNEs' R&D activities: agglomeration, location patterns and spillover effects	technolog, firm, innov, foreign, activ, knowledg, industri, spillov, product, patent	High	Cantwell & Piscitello (2002) – 96; Cantwell & Piscitello (2004) – 96; Cantwell & Piscitello (2000) – 96	High; High; High	5.00
21	Qualitative approaches to IB research and theorizing	manag, research, studi, theori, intern, organiz, organ, divers, process, team	High	Welch et al (2009) – 83; Madureira (2007) – 80; Vincze (2006) – 78	High; High; High	3.66
22	FDI and regional integration in transition economies (with specific references to the banking industry)	countri, bank, fdi, econom, invest, foreign, growth, region, develop, economi	High	Alexe & Tatomir (2011) – 88; Paas & Tafenau (2004) – 85; Bacic et al (2004) – 81	High; High; High	3.28
23	Offshoring: service activities, and international survey methods	offshor, outsourc, servic, activ, compani, cost, product, global, sourc, intern	High	Falk & Peng (2011) – 68; Chidlow & Ghauri (2011) – 63; Pajak (2011) – 63	High; High; High	1.57
24	Development of scales in IB research	perform, measur, relationship, model, variabl, studi, factor, item, use, manag	High	Lu (2006) – 76; Carneiro et al (2006) – 76; Gumus (2008) – 74	Medium; High; High	4.21
25	Qualitative IB research, business relationships and networks	network, relationship, busi, case, manag, research, compani, social, studi, interview	High	Michailova et al (2010) – 82; Welch et al (1999) – 77; Pereira & Barandas (2009) – 73	High; High; High	5.55
26	Institutional approaches to IB research and theory	countri, institut, firm, intern, develop, econom, govern, polit, busi, nation	High	Voinea & van Kranenburg (2011) – 73; Cuervo-Cazurra & Dau (2009) – 71; Hemisz et al (2005) – 68	High; High; High	3.27
27	Cross-cultural issues	cultur, distanc, differ, languag, countri, manag, nation, studi, intern, dimens	High	Quer et al (2005) – 75; León-Darder et al (2010) – 75; Teodosio & Robalo (2011) – 79	High; High; High	3.79
28	Entry mode choices	mode, entri, intern, servic, market, countri, choic, firm, foreign, studi	High	Fisher et al (2002) – 83; Lu (2006) – 80; Pla-Barber & León-Darder (2002) – 73	High; High; High	2.86
29	Company valuation, information disclosure and internet-based approaches	inform, market, compani, internet, use, valu, financi, manag, report, price	High	Gupta et al (1999) – 81; Gupta et al (2000) – 81; Holland (2006) – 79	Medium; High; High	1.56
30	International corporate governance issues	board, corpor, govern, ownership, sharehold, director, manag, compani, control, firm	High	Rueda & Correa (2009) – 99; Schutter (2007) – 99; Tacheva & Simon (2005) – 84	Low; Low; High	2.02