



# Exploring Sixty-Two Years of Research on Immigrants' Integration Using Structural Topic Modeling-Based Bibliometric Analysis

Robert Djogbenou<sup>1</sup> · Vissého Adjiwanou<sup>2</sup> · Solène Lardoux<sup>1</sup>

Accepted: 24 February 2024

© The Author(s), under exclusive licence to Springer Nature B.V. 2024

## Abstract

The research on immigrant integration is a dynamic interdisciplinary domain with rich and diverse literature. Considering the numerous studies in this domain, it is valuable to provide a comprehensive mapping to understand the research landscape resource and facilitate collaborations. This article combines Structural Topic modeling with bibliometric analysis to identify key research topics on immigrant integration. Applying these methods to 70890 abstracts published between 1960 and 2022, we identified 30 key research topics. We also tracked their prevalence over time, correlations and distributions within institutions and countries, funding effects, and countries' collaborations. The results indicate that the most discussed topics are integration theory, economic integration, education, residential segregation, integration policy, language, religious diversity, cultural participation, identity & belonging, racism & discrimination, political participation, health & welfare, research methods, demographic issues, gender & violence. Moreover, some of these topics were highly prominent in earlier periods and nearly non-existent in later years, while others emerged only recently. Specific topics maintained consistent significance over time. Analyses of correlations and trends reveal clusters of topics and diverse distributions across countries and research institutions. The implications of these results can benefit researchers, helping them better understand the current state of research and design future research projects. The finding could also help stakeholders in migration and integration governance and funding agencies to guide policies regarding the integration of immigrants.

**Keywords** Immigration · Immigrants' integration · Ethnic diversity · Topic modeling · Bibliometric analysis

---

✉ Robert Djogbenou  
yao.robert.djogbenou@umontreal.ca

<sup>1</sup> Department of Demography, Université de Montréal, Montréal, Québec H3C 3J7, Canada

<sup>2</sup> Department of Sociology, Université du Québec à Montréal, Montréal, Québec H2L 2C4, Canada

## Introduction

Over the past five decades, the world has experienced an increase in migration flows, resulting in more culturally, religiously, and ethnically diverse populations in many Western countries (Alba & Foner, 2015; Vertovec, 2007). The increasing numbers and diverse origins of immigrants in many countries have led to political and public concerns about their successful integration. Although it is the subject of several debates, integration remains a central focus of many studies and discussions regarding the settlement of newcomers in the host society. It is a long-term intergenerational and multidimensional process that refers to the “settlement process, interaction with the host society, and social changes that follow immigration” (Penninx & Garcés-Mascreñas, 2016, p. 11). When immigrants arrive in a host society, they must secure a place. Finding a home in a social and cultural sense will enable them to establish cooperation and interaction with other individuals and groups, learn to know and use the host society’s institutions, and be recognized and accepted in their cultural specificity (Penninx & Garcés-Mascreñas, 2016). As a result, integration is a complex process encompassing many aspects such as religion, language, culture, education, employment, housing, family, health, legal, juridical recognition, gender equality, identity, and many others (Ager & Strang, 2008).

To date, the field of integration research has grown considerably, and many articles have been published. Along with the growth of these studies, various efforts have been made to assess the progress. This body of work includes systematic literature reviews (Bekteshi & Kang, 2020; Brzozowski, 2019; Kaufmann, 2021; Laurentsyeva & Venturini, 2017; Salehi, 2010; Smith et al., 2019) and bibliometric analyses (Atçeken & Dik, 2023; Gao & Wang, 2022; Picanço Cruz & de Queiroz Falcão, 2016; Shuangyun & Hongxia, 2020; Sweileh et al., 2018). These studies aim to synthesize, critique, and reflect on integration. While these studies conduct literature searches and subsequently summarize findings, most have evaluated the content of research articles using systematic content analysis, which has four limitations. Firstly, the growing volume of publications in the integration field, coupled with the rise of big data, has limited the efficiency of manual efforts to conduct content analysis. Secondly, studies employing manual coding predefine categories of topics, which may change over time. Moreover, manual coding is often susceptible to subjective biases that can impact results. Thirdly, manually coding review articles is a demanding task, usually leading to limited analysis of a relatively small number of articles. Finally, existing studies focus on a specific aspect of integration or a short period, thus missing an overall view of the integration of immigrants. Overall, previous research has lacked in one or more of the following aspects: (i) a large-scale approach, (ii) a content-related approach, and (iii) an inductive approach.

This article addresses these limitations and aims to examine the evolution of the topic in studies on immigrants’ integration. It seeks to address the following research questions:

- Q1: What research topics have been covered in immigrant integration studies?  
Q2: How have these topics varied over time? How have they correlated?  
Q3: How have these topics been distributed among countries and influential research institutions? How have the countries' collaborations, and the article funding affected them?

To address these questions, we combine the Structural Topic Model (STM) and bibliometric analysis to examine 70,890 abstracts of articles on immigrant integration. We downloaded these articles from 117 journals in Web of Science, Scopus, and Dimensions, as well as three book series from 1960 to 2022. With the digitization of scholarly databases, bibliometric data has become increasingly employed to map scientific studies (Alburez-Gutierrez et al., 2021; Kashyap et al., 2023). Advances in computer technology have made it possible to use a large-scale approach, automate content analysis, and use the inductive method, enabling researchers to make comparisons over long periods. We employ Structural Topic Modeling to identify the topics that emerged from our analysis, their evolution, and their interactions. Additionally, we explore the distribution of topics across bibliometric indicators such as countries, research institutions, scientific collaborations, and funding effects.

Combining bibliometric analysis with the STM method in this study offers several advantages. Identifying substantial topics, their interactions, evolutions, and emerging areas, particularly within a longitudinal context, effectively captures the essence of a research domain. It helps track current and future developments and shapes research priorities (Chen et al., 2020, 2021; Chen & Xie, 2020). It is, therefore, essential to know how research topics gradually emerge to understand science's role in society. This way, results obtained using STM methods and bibliometrics could considerably broaden the scope of knowledge on integration of immigrants compared with the results of traditional manual content coding analyses.

## Literature Review

Immigrant integration has been described as an interdisciplinary field as it has historically been studied across several broader academic disciplines, including sociology, demography, economics, geography, anthropology, and psychology (Tran, 2015). Furthermore, the field has been described as an open discipline, including researchers from diverse academic backgrounds and contributions from media actors, practitioners, and researchers specializing in the field. Moreover, studies on integration employ a variety of theories and statistical methods, data from diverse sources conducted in many countries, albeit dominated by North American and European countries (Pisarevskaya et al., 2019). Previous studies have pointed out that integration is a policy goal for initiatives aimed at facilitating the settlement of immigrants and refugees (Rodríguez-García, 2015; Ager & Strang, 2008; Chen & Wang, 2015). It is a much complex, contested, politicized, and contextual concept, and there is a lack of clarity in its measurement (Rodríguez-García, 2015; Chen & Wang, 2015; Phillimore & Goodson, 2008; Goodman & Kirkwood, 2019). This demonstrates that the exact nature of integration is

vague and can be used to support a range of immigrant related aspects and more inclusive host country policies (Goodman & Kirkwood, 2019).

Based on previous research, three approaches offer a panoramic view of immigrant integration. The first approach, known as close reading, relies on peer-reviewed articles. For instance, Salehi (2010) used 38 peer-reviewed articles and 20 pieces of gray literature to critically overview existing research on the health of young immigrants in Canada. Similarly, Kaufmann (2021) systematically reviewed 319 sources to map topics on youth integration in Canadian literature. Using the PRISMA method (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), Bekteshi and Kang (2020) synthesized results from 30 studies published between 2000 and 2015 that investigated the influence of several sociodemographic and cultural contexts on acculturative stress among Latino immigrants in the United States. Brzozowski (2019) presents a critical analysis of research on immigrants' economic integration based on a sample of 44 articles. Although very useful, such a method poses problems for analyzing large amounts of textual data. One of the limitations of this method is its susceptibility to the knowledge view and subjective judgment of peers, often leading to incomplete representation of important literature and emerging topics. Similarly, this approach is limited in its ability to cover various research topics within the expansive literature on the integration of immigrants and its changing landscape over time. This underscores the necessity for large-scale analyses capable of accommodating the vast numbers of journals, articles, and diverse topics within the field.

The second approach employs a bibliometric analysis of large corpora. Bibliometric analysis, defined as “the quantitative study of bibliometric data”, has become a vital, efficient, and reliable tool for assessing scientific publications (Chen & Xie, 2020). With the rapid advances of computers and the availability of analytical tools, its use has grown considerably in recent years. It objectively assesses researchers' quality and productivity in a specific field, retrospectively explores research history, and identifies frontiers in various disciplines or research areas (Chen et al., 2020). In the field of migration and integration, several studies have employed large-scale bibliometric data to explore academic mobility to and from a specific country (Miranda-González et al., 2020; Subbotin & Aref, 2021; Zhao et al., 2021) and to study specific integration dimensions in literature (Atçeken & Dik, 2023; Gao & Wang, 2022; Picanço Cruz & de Queiroz Falcão, 2016; Shuangyun & Hongxia, 2020; Sweileh et al., 2018). For instance, Shuangyun and Hongxia (2020) used the bibliometric method on 1,557 articles to analyze the research institutions, key authors, keywords, and citations to map knowledge about ethnic identity and immigrant acculturation. Picanço Cruz and de Queiroz Falcão (2016) conducted a study on the economic integration of immigrants, analyzing 676 articles published on the Web of Science. Their bibliometric analysis involved quantifying yearly publication numbers, qualitative analysis of the 40 most-cited articles, and examining the research network of 20 international authors who had contributed the most to the topic. Similarly, Sweileh et al. (2018) used a large-scale bibliometric analysis of 21,457 documents (research articles, review articles, letters, notes, editorials, conference papers, and news) to assess the mapping of the literature on international migrant health. Their analysis focuses on over 70 health keywords in titles and abstracts, author networks,

affiliations, research areas, and most active institutions. Recently, Atçeken and Dik (2023) analyzed 1836 journal articles on academic literature regarding Syrian migration between 2011 and 2021 using co-citation analysis of keywords. The authors grouped keywords into six clusters: temporary protection, governance strategy, labor market, Turkey, Syrian refugee students, and connected immigrants.

While these studies offer valuable insights into migration and integration dynamics and trends, certain aspects have been missed. Firstly, prior studies include in their analyses reviews, editorials, and book reviews, yield fewer original results than research articles. However, it is recommended to use research articles exclusively as they provide more original results, while including other publication types might negatively impact findings (Chen & Xie, 2020). Secondly, these studies often analyze citation patterns and frequently use keywords, emphasizing influence rather than research content. Moreover, analyses based on keywords, titles, and journal categorization are intrinsically flawed, lacking depth and substantive content analysis. These studies prioritize a broader analysis but come at the expense of comprehensive content analysis, offering limited insights into the diverse array of topics within the research field.

A third inductive approach, combined with quantitative text analysis, offers new possibilities beyond these limitations. Developing new automated techniques for large-scale text analysis promises significant advancements in the study of immigrant integration (Drouhot et al., 2023). These techniques typically involve using a human reference to classify a small portion of the data and then employing machine learning to identify similar patterns in the remaining data, thereby combining automated scalability with the depth of human interpretation (Drouhot et al., 2023). Quantitative text analysis enables researchers to uncover the corpus structure before imposing preconceptions on the analysis. This approach is precious due to the expansive number and diversity of available publications and the scattered nature of the domain of immigrant integration. Using an inductive approach offers the possibility of moving from pure subjective judgment to the combination of subjective judgment and objective measurement (Daenekindt & Huisman, 2020). Quantitative text analysis, a process of deriving trends from text, allows for a comprehensive semantic content analysis of large document corpora to identify relevant studies in scientific journal articles quickly (Malaterre et al., 2019; O'Mara-Eves et al., 2015; Wang et al., 2016). This method differs from and transcends conventional approaches such as citation analysis, systematic reviews, and meta-analysis, which are designed to address a specific research question using a restricted number of publications (Peters et al., 2015). Such analysis can help map an “*academic landscape*” (Kajikawa et al., 2007, p.223; Pisarevskaya et al., 2019).

In the migration domain, little research has employed an inductive approach with quantitative text analysis, with a few exceptions, such as the study by Pisarevskaya et al., (2019). Pisarevskaya et al. (2019) used the Latent Dirichlet Allocation method (Blei et al., 2003) to examine 22,140 article abstracts from 1986 to 2017, identifying the topics covered in migration studies. The authors found that the increase in migration studies did not necessarily result in a more diverse range of topics within the field. Instead, the focus shifted from demographic, statistical, and governance issues to a growing emphasis on mobility, migration-related diversity, gender, and health. Our analysis follows a similar goal in the field of integration of immigrant

studies. However, some aspects are still not considered in the study of Pisarevskaya et al. (2019). Specifically, most articles published after 2017 were not included in their analysis. Given that migration and integration research is increasing in recent years, it is essential to consider the latest studies for a complete and up-to-date analysis. Furthermore, we extract latent topics from abstract content using the Structural Topic Model (STM), a newer extension of the Latent Dirichlet Allocation method. STM considers independent variables during the exploration and estimation of latent topics (for more details, see the methodology section). As a result, we incorporate variables derived from bibliometric data, such as year of publication, collaboration type, funding status of the article, countries, and institutions. This approach allows us to explore how the addressed topics vary based on these bibliometric indicators.

Although the STM method has been little used to map the scientific literature on migration and immigrant integration, this type of method is becoming increasingly popular in studies of media coverage of migration (Erhard et al., 2022), parliamentary discourses on migration (Hajdinjak et al., 2020), and host community attitudes towards immigrants and refugees (Kelling & Monroe, 2022). For instance, Hajdinjak et al. (2020) employed STM to assess the framing of migration policies in the U.S. House of Representatives and the Canadian House of Commons from 1994 to 2016. Their findings indicated that U.S. Democrats and Canadian Liberals emphasized well-being and humanitarian aid, while conservative groups focused on security and the legal aspects of migration. On the other hand, Erhard et al. (2022) used STM to examine discourse in German media coverage on migration from 2001 to 2016. They identified topics including education, economics, immigration law, refugee crises, multiculturalism, religion, notably Islam, domestic violence, and football.

This extensive literature review shows that relatively few studies have applied STM to track the evolution of immigrant integration over a long period. This article contributes to existing studies, combining the STM method with bibliometric analysis to explore the prominent trends in immigrant integration research over time. This type of combination is increasingly used in other fields (Chen et al., 2020, 2021; Chen & Xie, 2020, for AI-assisted human brain research; Sharma et al., 2021, for information management research). An empirical investigation into the knowledge structure and topic composition changes within studies on immigrant integration is critical for understanding field evolution, identifying potential research avenues, and addressing complex research questions which were previously impossible to address.

## Data and Methods

### Data Collection

We compile a dataset of abstracts from articles published in 117 social science journals in Web of Science, Scopus, and Dimensions. In addition to considering research articles only, we also collect data from three book series (International Migration, Integration and Social Cohesion in Europe; Migration, Diasporas; and Citizenship, International Perspectives in Migration). Including these book series aligns with existing studies (Pisarevskaya et al., 2019), as they exclusively publish

articles on migration and immigrant integration, likely to influence topics in this domain. Our concentration on abstracts aligns with previous research (see, for example, Schwemmer & Wieczorek, 2019; Sweileh et al., 2018) and is informed by four criteria: abstracts (i) are more readily accessible for automated scraping from Web of Science, Scopus and Dimensions, and book series compared to full texts; (ii) are freely available than full texts, (iii) provide a concise summary of the article, (iv) are relatively consistent in format or style across various journals. We identify the most relevant literature sources and focus on high-impact factor journals for which publication data are available. We then extract abstracts using three keyword subgroups. The first subgroup is related to integration and includes the following topics: participation, insertion, assimilation, adaptation, acculturation, sense of belonging, discrimination, and ethnic identity. The second subgroup is related to the dimensions of integration. The dimensions selected are social, economic, cultural, civic, family, political, citizenship, linguistic, demographic, health, and housing. The third subgroup concerns immigrant populations, including minority groups, visible minorities, ethnocultural minorities, and cultural minorities. These subgroups correspond to three distinct research fields and are linked by the programming term 'AND.' The equivalents of all subgroups are entered in their respective search fields separated by the topical expression 'OR.'

The data collection was done based on three criteria. Firstly, articles had to be written in English and published between 1960 and 2022. We limit the search between 1960 and 2022 because very few articles were published before 1960, and abstracts were unavailable in 1960 in the Web of Science, Scopus, Dimensions databases, and book series. Secondly, we limited our analysis to journal articles because they typically present more original research findings and offer explicit details about authors and their affiliations. Thirdly, we ensured that terms in each article's title, abstract, or keywords matched at least one of the keywords used in our final queries. The literature search in the three databases (Web of Science, Scopus, and Dimensions) generated 119,190 results. In addition, we manually collected 941 abstracts of available chapters in the three-book series mentioned above. We excluded articles that did not focus on aspects of immigrants' integration or published in fields outside of social sciences. Duplicates and articles without abstracts were excluded from the final dataset, which includes 70,890 articles from 117 journals and the three book series (Supplementary Table S1). The bibliometric information for each article was verified. Authors' names, research institutions, and countries were extracted from author affiliation information. We used standard metrics in bibliometric studies, such as the number of articles and citations, to evaluate the performance of influential countries and institutions. The number of articles and citations measures the productivity and influence of a researcher's scientific output (Chen et al., 2020; Chen et al., 2021). For each extracted country, we grouped articles into two mutually exclusive categories: (1) articles with a single institution or at least two institutions from the same country (local/national collaboration) and (2) articles with at least two institutions from at least two different countries (international collaboration). Funding was measured using information available in the acknowledgment section of each article. Articles are grouped into two groups: articles with and without funding.



## Data Pre-Processing

We apply a series of standard text pre-processing techniques. We converted all abstracts to a word format where words were normalized to lowercase. Punctuation, numbers, white space, and URLs were removed. Words included on the article page, such as copyright information, were removed because they provide no semantic value about an article (see Schwemmer & Wieczorek, 2019). In addition, general words, such as ‘introduction,’ ‘background,’ ‘objective,’ ‘method,’ ‘result,’ ‘conclusion,’ ‘contribution,’ ‘keywords,’ as well as stop words (e.g., the, for, pronouns, adjectives, adverbs, verbs, etc.) have been removed from the abstracts. Next, we use lemmatization based on orthographic normalization, a linguistic technique that helps combine words with similar semantic meanings (Malaterre et al., 2019; Schwemmer & Wieczorek, 2019). For example, applying lemmatization to the words ‘good,’ ‘better,’ and ‘best’ will be reduced to ‘good’. Since structural topic modeling is based on the co-occurrence of words in a corpus, keeping only words that reach a certain frequency threshold is essential. For the robustness of the results, we removed words appearing in more than 50% of all abstracts. We also normalize spelling differences in the U.K. and the U.S. to have only one word in the corpus for terms such as ‘labor’ and ‘labour,’ or ‘analyse’ and ‘analyze.’

## Methods

### Structural Topic Modeling

This study employs a form of quantitative text analysis known as Structural Topic Modeling (STM), developed by Roberts et al. (2014). This method enables researchers to uncover topics within documents and understand how metadata covariates affect the contribution of each topic in the documents. Beyond the advantage of accounting for covariates, STM explicitly models inherent correlations or co-occurrences between topics and estimates multiple topics simultaneously present in a text, distinguishing it from traditional Latent Dirichlet Allocation methods (Rodriguez & Storer, 2019). This attribute is desirable because researchers can use structural topic modeling to observe which topics are highly correlated with each other at the document level (Lindstedt, 2019). STM is beneficial in this study because we are interested in understanding how the prevalence of topics discussed in abstracts has changed over the past sixty years and their co-occurrence, variations across countries, research institutions, collaborations, funding effects, and citation-based impact. Model diagnostics were examined using the searchK function in the STM package within the R software. We use the spectral initialization method as it ensures that globally optimal parameters are obtained and, compared to other initialization methods, is faster and produces better results (Roberts et al., 2016). While there are methods for estimating the likelihood that the number of topics selected is correct, there is



no scientific consensus on how best to choose the number of topics (Grimmer & Stewart, 2013; Rodriguez & Storer, 2019). However, the authors of the STM package recommend examining diagnostic figures and looking at the trade-off between semantic coherence<sup>1</sup> and exclusivity<sup>2</sup> (Supplementary Figures S1 and S2). We generate a set of candidate models with different numbers of topics, ranging from 5 to 100. This allowed for an examination of the ideal of granularity in the data, which was between 25 and 35 topics. By estimating additional models within the range of 25-35, we ultimately select the solution with 30 topics as the optimal number. The model with 30 topics was chosen because it offers the best balance between semantic coherence and exclusivity, accounting for topic diversity in the domain, offering a high capacity for analytical interpretation, and representing topics about our research questions more effectively than other models. Conversely, models with more than 30 topics exhibit low semantic coherence but greater exclusivity, indicating that these topics were peripheral to the domain and may not be as interpretable. Furthermore, the model with 30 topics allows for a nuanced overview of the topic sub-components of integration of immigrants (e.g., culture, language, education, employment, housing, legal recognition, security, discrimination, health, and many others) or other categories such as integration policies and research category.

### Topic Labeling

Following the application of STM, certain words are linked with specific topics. The process of topic labeling is crucial because unsupervised machine learning does not automatically produce explicit definitions, labels, or intuitive meanings for topics. The researcher must determine the representation of topics based on the frequency of various words appearing within the topics. The human coder grouping step aids in achieving this goal. To ensure objectivity and consistency of topics, the three authors coded each topic and clustered the topics separately to assign the final label. Carefully reading the documents strongly associated with each topic further helps us interpret and validate the model.

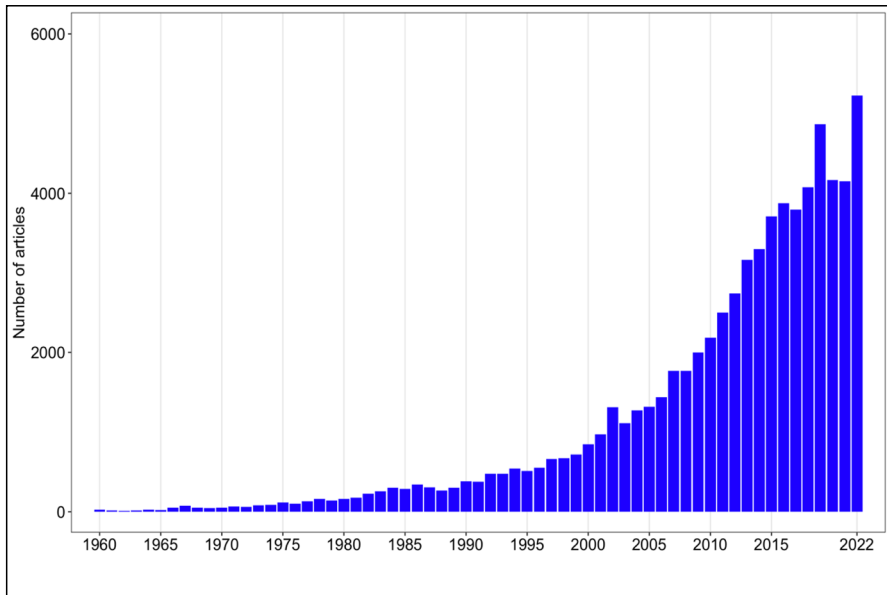
### Topic Network Analysis

The STM algorithm determines the proportions of all observed topics in each abstract. Therefore, each abstract may contain several topics of substantial importance. Based on the matrix generated by the STM model with proportions of topics per abstract, we calculate the Pearson correlation coefficient between topics at the level of all abstracts. Correlation figures were generated using the Fruchterman-Reingold algorithm available in the visNetwork package in R.

---

<sup>1</sup> Defined as the probability for a set of words in topics to coexist in the same document.

<sup>2</sup> Defined as the probability for a word to appear primarily in the top rankings of a single topic.



**Fig. 1** Number of publications and abstracts in the database

## Results

### Evolution of Articles on Integration Over Time

Figure 1 shows the dataset's total number of abstracts per year. After a stable phase until the 1990s, there was an exponential explosion of publications related to integration in the following years. The exponential growth in articles demonstrated increased authors' interest and enthusiasm toward immigrant integration research.

### Geographical Distribution of the Number of Publications

Figure 2 illustrates the geographic distribution of the number of articles on the integration of immigrants. A total of 167 countries contributed to the publication of these articles. The United States has demonstrated considerable research publications, with the highest number of articles (28,014), followed by the United Kingdom (9,881) and Canada (5,267). These findings suggest that scholars based in North America and Europe drive much of the research on integrating immigrants.

### Proportion of Topics

Figure 3 presents the expected proportions of topics in the corpus along with each topic's seven most important words. Supplementary Table S2 presents detailed results

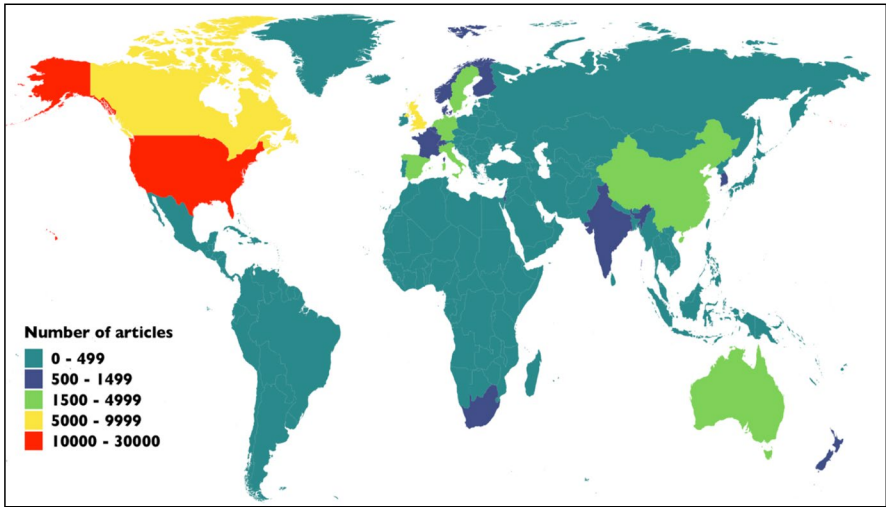


Fig. 2 Geographical distribution of the number of publications

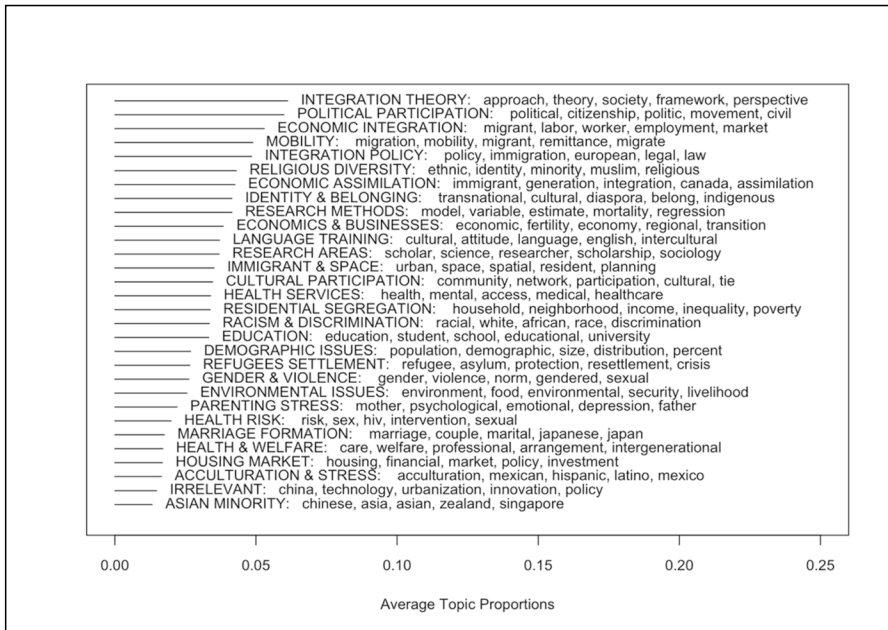


Fig. 3 Expected average proportions of the 30 topics

for the 30 topics and topic categories. The figure indicates the existence of various topics in studies on the integration of immigrants. Of the 30 topics, 23 are specific to immigrant integration studies. These topics account for 76.6% of the entire dataset and include integration theory, economic integration, economics & businesses,

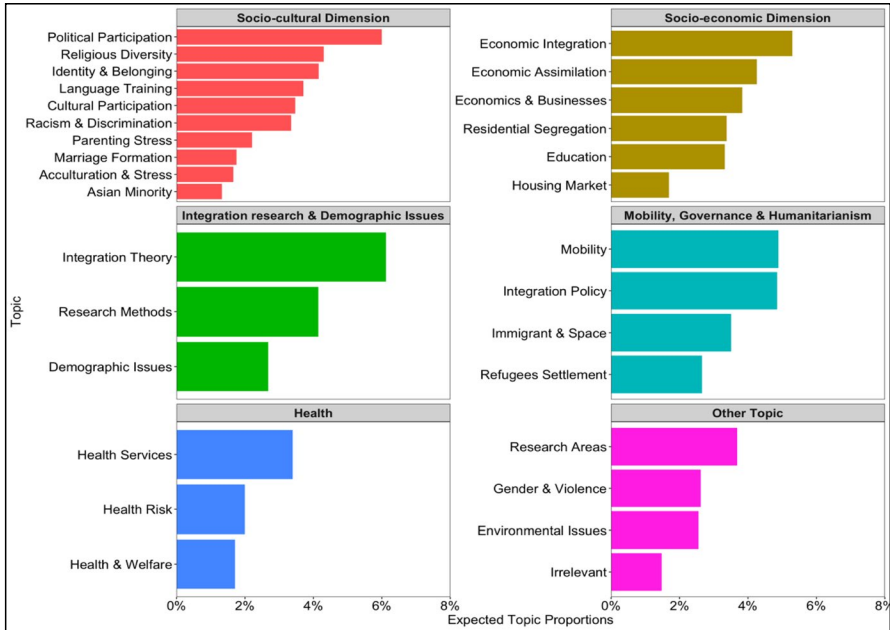


Fig. 4 Clustering of the 30 topics

economic assimilation, education, the housing market, residential segregation, mobility, refugees settlement, immigrant & space, integration policy, language training, acculturation & stress, religious diversity, cultural participation, identity & belonging, racism & discrimination, marriage formation, asian minority, parenting stress, political participation, health risk, health services, health & welfare. Three other topics concern theoretical frameworks, methods, and demographic issues, comprising 10% of the total data. These topics are research methods, integration theory, and demographic issues. The remaining topics are too generic in migration and immigrants’ integration studies. These topics represent 13.3% of the total data, including gender & violence, irrelevant, environmental issues, and research areas.

The six main topics with the highest proportions in the dataset are: integration theory (6%), political participation (6%), economic integration (5%), mobility (5%), integration policies (5%), and religious diversity (5%).

### Categorization of Topics in Immigrants’ Integration Studies

The results of the STM model suggest that the 30 obtained topics can be grouped into six clusters: (1) the socioeconomic dimension of integration, (2) the socio-cultural dimension of integration, (3) mobility, governance, and humanitarian aspects, (4) health, (5) research and demographic issues, and (6) other categories. Figure 4 presents the six categories of topics. We provide here the most representative document for each topic.

## Socio-Economic Dimension

The six topics grouped in this category discuss employment or the economy (economic integration, economy & businesses, economic assimilation), housing (housing market, residential segregation), and education (education). For instance, the five most essential words for the topic we call «economic integration» are: «migrant, labor, worker, employment, market, job, and skilled». The article most strongly associated with this topic examines how the recession impacts the labor market performance of immigrants from Eastern Europe living in the U.K. (Khattab & Fox, 2016). The topic «economic assimilation» includes words such as «immigrant; generation; integration; Canada; assimilation». One article that aligns closely with the topic explores variations in the income assimilation of immigrants from the former Soviet Union into Germany and Israel between 1994 and 2005 (Haberfeld et al., 2011). Another article strongly linked to this topic studies the earnings of immigrant cohorts and Canadian workers between 1980 and 2000 (Frenette & Morissette, 2005). Another topic in this category, «education», includes «education; student; school; educational; university; and academic». The article most connected to this topic examines the educational and occupational outcomes of adult children of immigrants in the United States (Feliciano & Rumbaut, 2005). The topic «residential assimilation» includes «household, neighborhood, income, inequality, poverty, residential, segregation». The article most associated with this topic explores the impact of the ethnoracial composition of neighborhoods on the mobility of first and second-generation immigrants in France (McAvay, 2018).

## Identity/Culture and Religion

As expected, this category encompasses a broader range of topics related to ethnic and religious diversity, identity and a sense of belonging, language education, marriage, parenthood, acculturation, racial discrimination, cultural participation, political engagement, and asian minorities. For instance, the topic we call «language training» includes «cultural, language, English». The article most closely associated with this topic investigates English as a language for success in primary schools in England (Demie, 2018). The «cultural participation» topic encompasses words such as «community; network; participation; cultural». The article strongly associated with this topic examines the role of cultural similarity between the immigrants' culture of origin and host cultures in second-generation immigrants' community engagement in Hong Kong and the United States (Li et al., 2019). Similarly, other topics address cultural aspects, such as «acculturation stress», with words such as «acculturation, mexican, hispanic, latino». The article linked with this topic evaluates how adaptation to American culture by U.S.-born Hispanics and foreign-born Hispanics living in the U.S. might influence physiological stress-related factors and health (Cedillo et al., 2021). The topic «political participation» incorporates words such as «political, citizenship, politics». The article associated with this topic explores immigrant involvement trajectories within Irish political parties (Szlovak, 2017). Another article linked to this topic discusses immigrant residence and voting rights (Lenard, 2015). The topic labeled «religious diversity» includes words such

as «ethnic, identity, minority, muslim, and religious». The most associated article investigates the role of religious commitment and host national identity in shaping interreligious sentiments among Sunni Muslim and Alevi minority groups in Europe (Martinovic & Verkuyten, 2016). On the other hand, the topic «identity and belonging» involves words such as «transnational, cultural, belong, identity». The articles linked to this topic discuss the super-diversity (Meissner & Vertovec, 2015; Vertovec, 2007). The topic named «racism and discrimination» include «racial, African, race, discrimination». The most associated article examines perceptions of African Americans, Afro-Caribbeans, and non-Hispanic Whites regarding their proximity to other racial and ethnic groups in the United States (Thornton et al., 2012). Another article on this topic delves into the relationship between socially attributed race and experiences of discrimination among Latinos or Mexicans in the United States (Vargas et al., 2016).

### **Governance and Policy**

This cluster includes topics related to integration policies, security laws, immigrant protection, refugee rights, and mobility. For example, the topic named «integration policy» encompasses words such as «policy; immigration; integration; legal; law; Europe». The article associated with this topic explores views on implementing migrants' integration policy in a context characterized by decentralization in the United Kingdom (Galandini et al., 2019). Other articles associated with this topic examine how and why different relationships between national and local governments affect the governance of immigrants' integration (Garcés-Mascareñas & Penninx, 2016; Scholten & Penninx, 2016). The second topic in this group, «Refugees Settlement», contains words such as «refugee, asylum, protection, resettlement, and crisis». The most closely related article on this topic documents the settlement experiences of Syrian refugees in the province of Alberta in Canada during their first year of resettlement (Agrawal, 2019). Moreover, the article contrasts the sponsorship programs- government and private- to determine which program more effectively supports refugees' resettlement and integration in Canada.

### **Health**

In this category, three topics concern aspects of immigrants' health: «health risk», «health services», and «health & welfare». This new set of topics highlights the relevance of studying immigrant health and well-being across diverse health-related issues and services, including smoking and alcoholism, clinical services, and mental health. For instance, the topic named «health services» includes words such as «health, mental, access, medical, and healthcare». The article associated with this topic investigates disparities in preventive health services among Somali immigrants and refugees in the United States (Morrison et al., 2012). The topic «Health & Welfare» includes words such as «care, welfare, professional, arrangement, and intergenerational». The most associated article examines the relief experiences among family and paid immigrant caregivers (del Carmen Salazar et al., 2016).

## Scientific Research and Statistics

In this category, topics related to theoretical, conceptual, and methodological perspectives on researchers' contributions to the analysis of immigrants' integration questions and issues emerge. For instance, the «Integration Theory» topic encompasses words such as «approach, theory, society, framework, perspective, diversity, theoretical». The most associated article with this topic explores the complex relationship between structure and agency and how it has been incorporated into migration theory (Bakewell, 2008). The topic «Research Methods» includes «model, variable, estimate, mortality, regression and method». «Demographic Issues» includes «population, demographic, size, distribution, percent, statistic, composition, projection».

### How Have the Immigrant Integration Topics Changed Between 1960 and 2022?

Figure 5 presents the proportions of 20 topics related to immigrants' integration during the publication year (for the rest of the topics, refer to Supplementary Figure S3). Some topics were significant in the early periods of the study and almost non-existent a few years later. Others have only recently emerged. Still, others have maintained a relatively stable significance over time. For example, the topics on «integration theory», «political participation», «the immigrant and space» and «integration policies» show a downward temporal trend, indicating that academic research focused on these topics was more likely to be published earlier and less so in recent years. More striking patterns concern the prevalence of topics related to the «economy and business», «residential segregation», and «language training» between 1960 and 1980. These trends then experienced a sharp decline from the 2000s until 2022.

On the other hand, topics linked to «mobility», «religious diversity», «cultural participation», and «education» showed a continuous presence. This reflects the maturity of these topics in the field of immigrants' integration. Figure 5 also reveals another critical pattern involving the rise of topics related to «racism and discrimination», «acculturative stress», «ethnic identity and sense of belonging», «economic assimilation», and «economic integration» throughout the study period. This suggests that more recent articles have been published on these topics. It is worth noting the emergence of topics related to «access to healthcare services» and «health and well-being» over the past two decades, although these remain smaller in proportion. This indicates a growing awareness of studies on health and well-being in the realm of immigrants' integration. These findings align with previous studies that demonstrated immigrants often encounter difficulties accessing appropriate healthcare services and face social and linguistic barriers that may affect their physical and mental health and the quality of healthcare services they receive (Sweileh et al., 2018).

### Evolution of Topic Clusters

Figure 6 presents the evolution of the six selected topic categories over time. The width of each stream is proportional to the percentage. Overall, all topic categories,



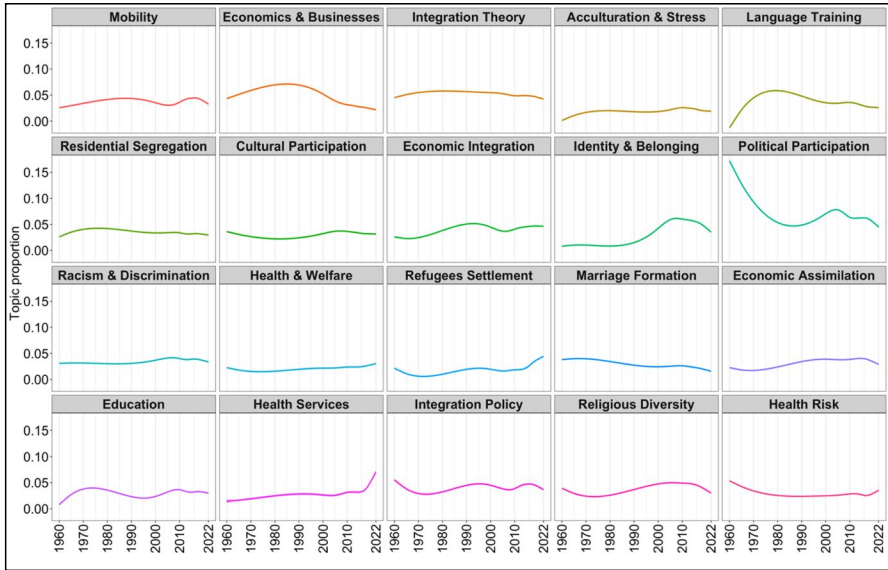
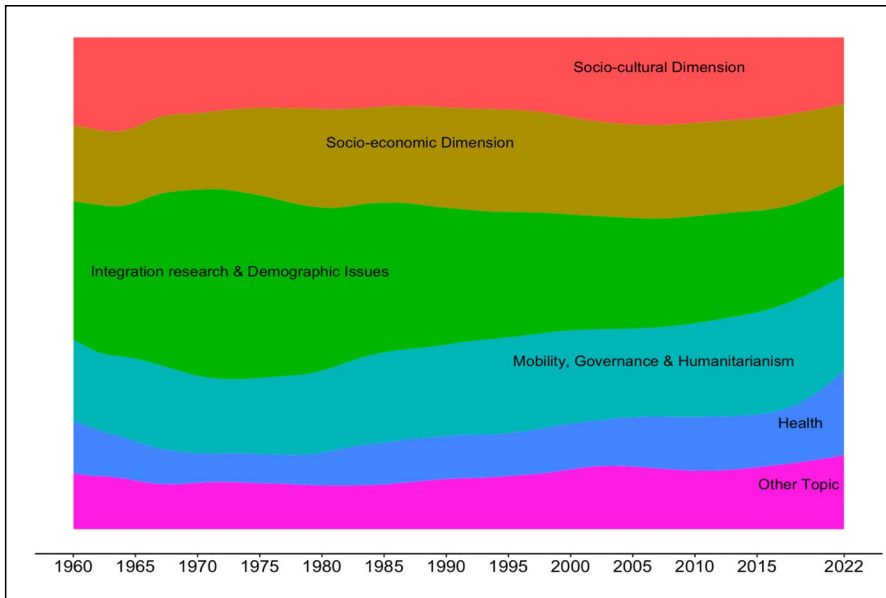


Fig. 5 Evolution of topics between 1960–2022

notably «Mobility, Governance, and Humanitarian Aspects», «Health», and «Research and Demographic Issues», experienced rapid growth, with the most significant growth observed around the turn of the 2000s. This indicates that the field of immigrants’ integration has shifted from a focus on economic and theoretical issues, demographic and methodological aspects to an increasing emphasis on governance, mobility, health, and immigrant well-being over the past decades.

**Correlation of Topics**

Figure 7 shows the correlation network of the topics. In the figure, each topic is represented by a circle, with size proportional to its prevalence in the entire corpus. Short distances between nodes and connecting lines indicate strong correlations and co-occurrence of topics within the same document. Colors are used to guide the reader through four distinct emergent topic groups. The first group (yellow color) comprises topics such as «Immigrant & Space», «Asian Minority», «Housing Market», «Economics & Businesses», and «Residential Segregation». These links suggest closely related research areas with a propensity for mutual influence. The second group (green color) encompasses «Integration theory, Political participation, Identity & belonging, religious diversity, Racism and discrimination, and cultural participation». In this group, the central position of «Integration theory» suggests its influence. Correlations within this group reflect the researcher’s interest in racial discrimination, ethnic and religious diversity, social cohesion, and immigrant sociocultural integration. This is expected due to the ethnic scope of migration and



**Fig. 6** Evolution of topic categories between 1960–2022

integration research (Pisarevskaya et al., 2019). Societies are increasingly marked by cultural diversity on an unprecedented scale, sometimes referred to as «super-diversity» (Vertovec, 2007). In this context, ethnic minorities often experience discrimination, impacting their participation and integration in the host society (Bilodeau, 2017). The strong connection between topics on «religious diversity» and «racial discrimination» may reflect the increasing attention to structural and interpersonal racism not only in the United States, perhaps reflecting the #blacklivesmatter movement, but also in Europe, where the idea of white Europeaness has figured in many public discourses (Pisarevskaya et al., 2019). The third cluster (red color) group topics such as «Economic Integration», «Economic Assimilation», «Education», «Language Training», «Mobility», and «Integration policy». This suggests sustained interest in immigrants' socioeconomic positions and discussions on their professional and linguistic skills. The fourth group (blue color) encompasses topics such as «Health Risk», «Health Services», «Health & Welfare», «Acculturation & Stress» and «Parenting Stress». Connectivity among these topics indicates that immigrants' integration is a complex process influenced by various migration-related events. Immigrants may experience integration as a period of shock and stress, managing factors such as loss of social status, language, housing, employment, schooling for their children, and access to healthcare in a new cultural environment (Penninx & Garcés-Mascareñas, 2016). Notably, «Health access» and «Health risk» are more distant, indicating an emerging or expanding field of immigrant health integration.

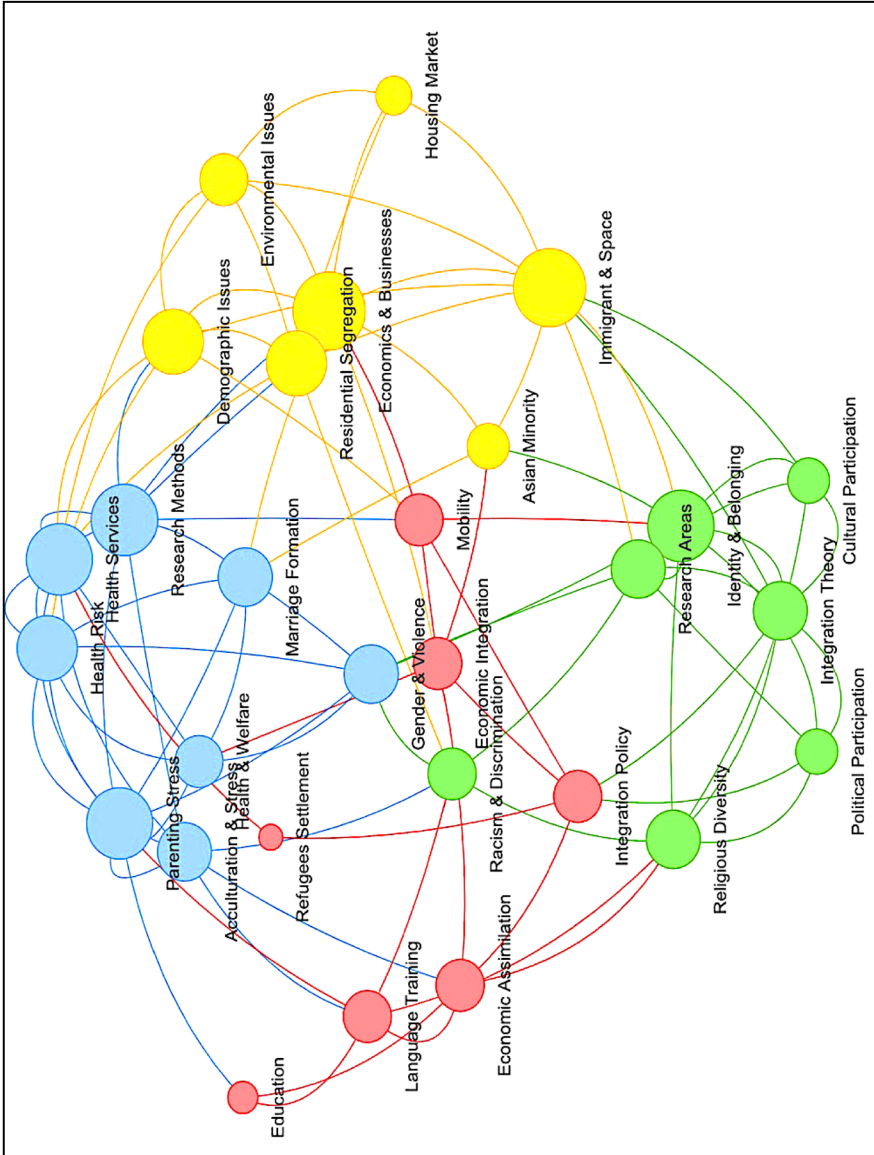


Fig. 7 Correlation network of topics

### Distribution of Topics by Countries

Figure 8 displays the distribution of topics among the countries. We observe a relatively balanced interest in each topic in countries such as the United States, the United Kingdom, Canada, Australia, and others. The United States, the United Kingdom, Australia, and Canada were mainly focused on topics related to «integration theories» and «economic integration». Research conducted in the United States also delved into topics of «racial discrimination», «residential segregation», «political participation», and «research methodologies». The United Kingdom exhibited a particular interest in topics concerning «immigrant mobility», «immigrants and space», and «identity, and sense of belonging». Canada and Australia showed significant enthusiasm for questions regarding «political participation», «cultural participation», «healthcare access», and «identity, sense of belonging». China's interest centered around «residential segregation», «immigrants and space», and «education». France was more productive in addressing «demographic issues», «mobility», «political participation», and «integration policies».

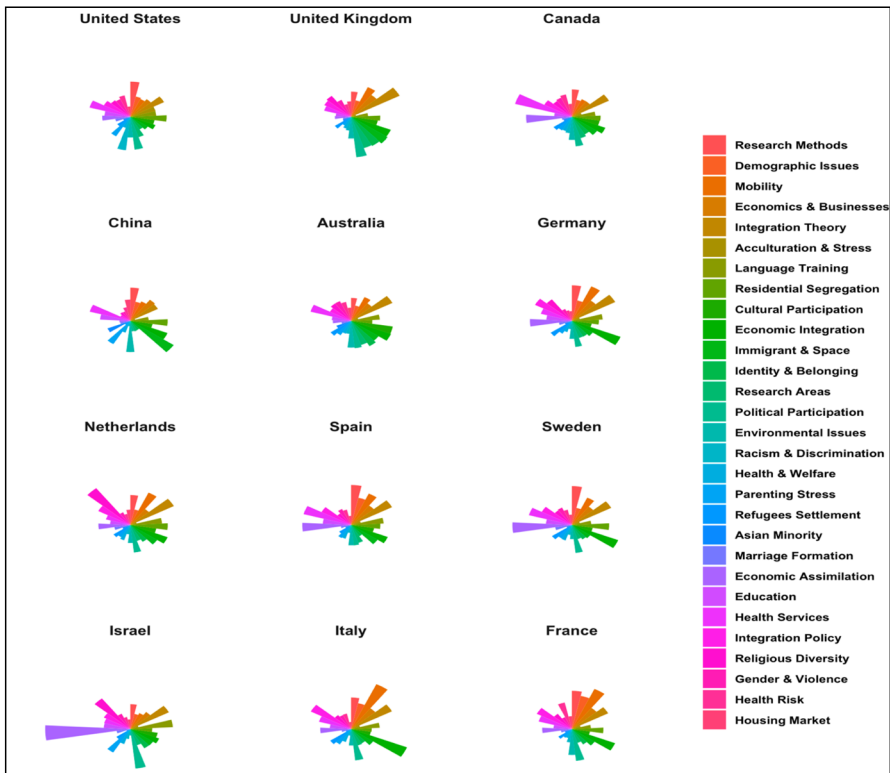


Fig. 8 Distribution of topic proportions by countries

### Distribution of Topics by Research Institutions

Figure 9 presents the distribution of topics by research institutions. Most of the research institutions have shown interest in topics related to «integration theories» and «research methods». In addition to these topics, each institute has interests in other areas. For example, American universities such as Princeton University, Harvard University, University of Michigan, and the University of California, Berkeley are interested in topics related to «Residential Segregation», «Political Participation», «Racism & Discrimination», «Health Services», «Economics & Businesses», «Demographic Issues», «Mobility», «Economic Integration», «Integration Policy». Regarding U.K. universities, the University of Oxford is more productive in topics such as «Economics & Businesses», «Economic Integration», «Political Participation», and «Integration Policy». In Canada, the University of British Columbia and the University of Toronto are also more productive in the fields of «Identity and belonging», «Political Participation», «Health Services», «Residential Segregation», and «Cultural Participation».

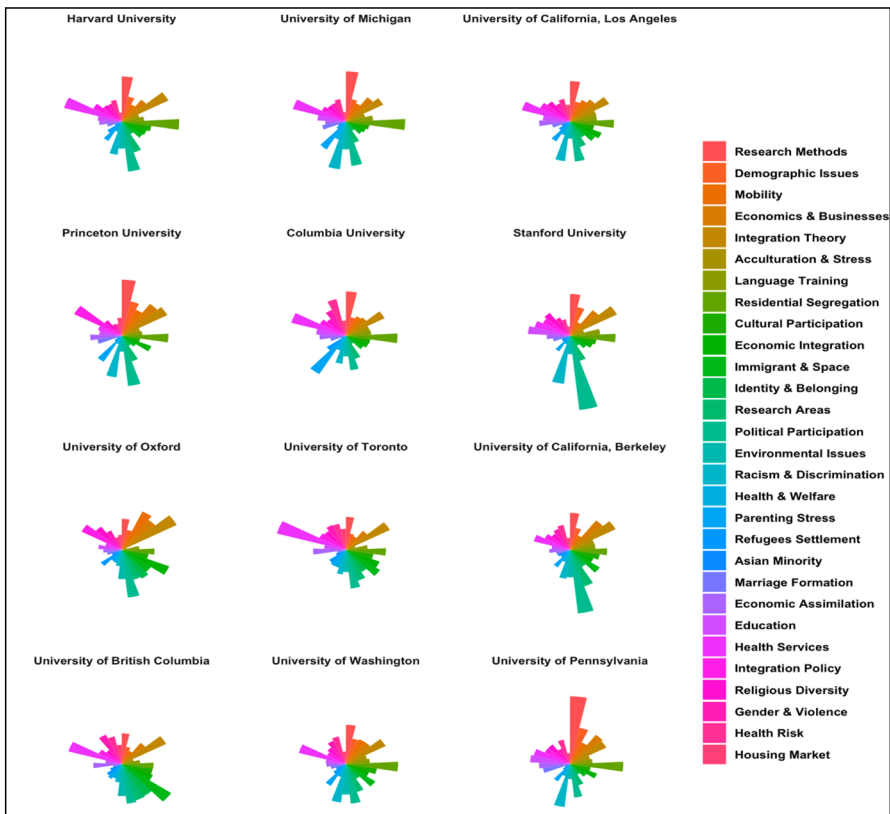


Fig. 9 Distribution of topic proportions by research institutions

### Topic Association by Funding Effect of Articles

Figure 10 shows the contrast in topic prevalence between funded and non-funded articles. Funded articles discuss topics related to «Education, Racism and discrimination, Health and Welfare, Parenting Stress, Refugees Settlement, Asian Minority, Research Areas, Political Participation, Economic Integration, Language Training, Research Methods, Demographic Issues, Mobility, Economic and business, Gender & Violence, Health Risk, Housing Market» much more compared to unfunded articles. On the other hand, topics such as «Integration Theory, Acculturation and stress, Cultural Participation, Identity and belonging, Marriage Formation, Economic Assimilation, Health Services, and Integration Policy» appear more frequently in unfunded research. Topics such as «Residential Segregation, Immigrant & Space, Religious Diversity, Environmental Issues» do not show significant differences between funded and unfunded articles.

### Topic Differences in Country Collaboration

Figure 11 presents the differences in topic prevalence between articles with international and national collaboration. Topics such as «Demographic Issues,

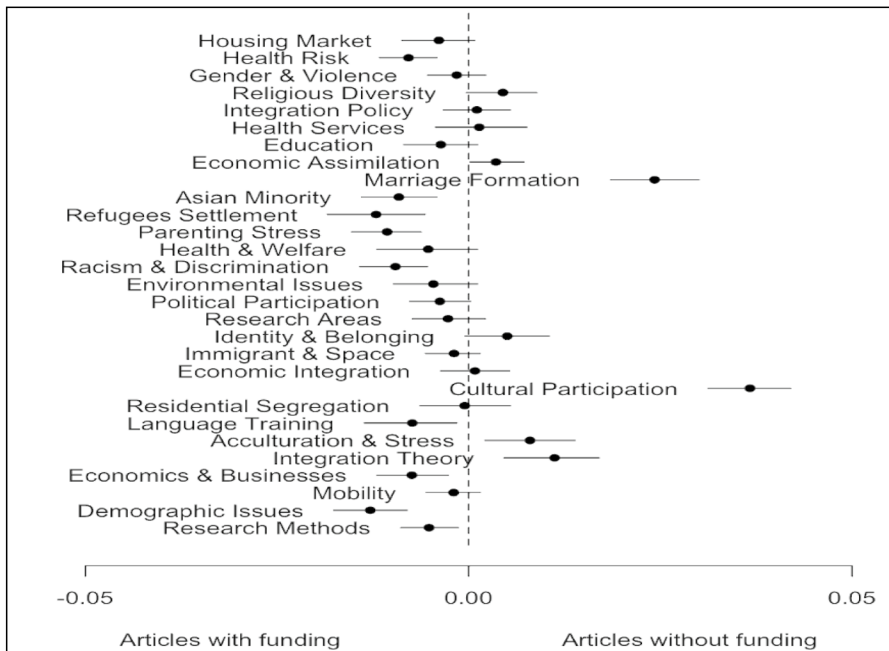


Fig. 10 Prevalence of topics by funding effect

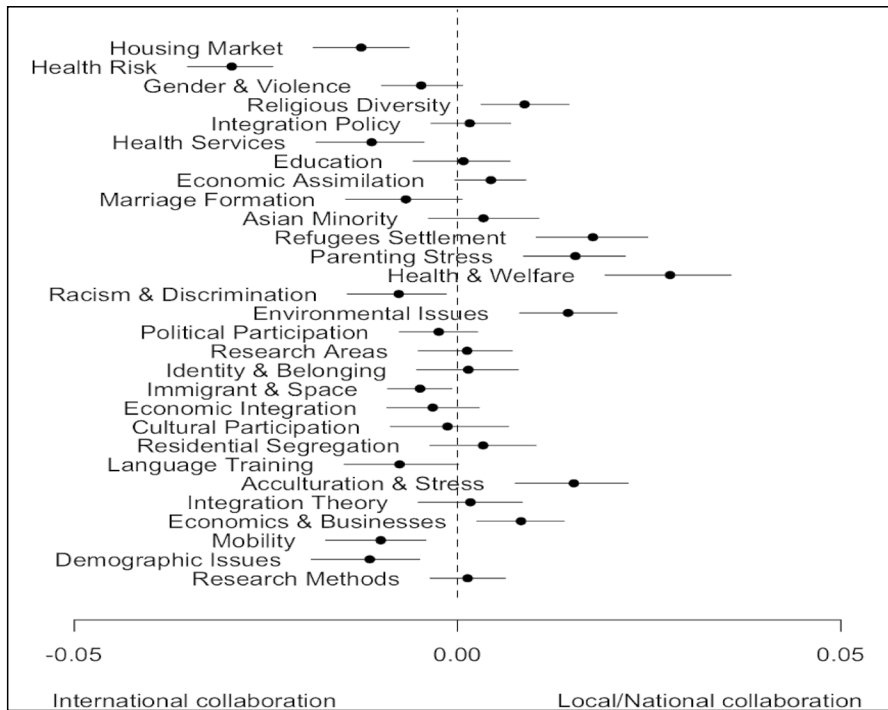


Fig. 11 Prevalence of topics by country collaboration

Mobility, Language Training, Economic Integration, Immigrant and space, Racism and discrimination, Marriage Formation, Education, Cultural Participation, Health Services, Gender and violence, Health Risk, and Housing Market» are more frequently addressed in articles with international collaboration. Conversely, topics such as «Economics and businesses, Integration Theory, Acculturation and stress, Residential Segregation, Identity and belonging, Environmental Issues, Health and Welfare, Parenting Stress, Refugees Settlement, Economic Assimilation, Integration Policy, and Religious Diversity» are more commonly addressed in articles with local or national collaboration.

## Discussion and Conclusion

This article presents a comprehensive mapping of the research field on immigrant integration. We combine the Structural Topic Modeling approach and bibliometric analysis on a corpus of 70,890 articles published in 117 journals and three book series covering 1960–2022. We identify 30 key topics that can be used better to understand the literature in this evolving interdisciplinary domain. We also document the temporal prevalence of these topics and their correlations. Furthermore, we visualize the distribution of topics across different countries and research institutions over time and the importance of funding effects and scientific collaborations



in shaping topic development. By employing a large-scale, inductive, and content-related approach, our analysis provides a robust foundation for understanding various topics and their variations across research units. Based on the findings, several conclusions can be drawn.

Firstly, our analysis reveals an exponential growth trajectory in terms of publication volume within the field of integration studies over the last decades. Our findings align with the conclusions of other studies in the context of migration (Pisarevskaya et al., 2019). This can be explained by the fact that the integration of immigrants is studied across various broader disciplines. This growth can also be attributed to the influence of several researchers from immigration countries and research institutions, notably in North America, Europe, and Australia, which have collectively shaped the research landscape on immigrants' integration. Pisarevskaya et al. (2019) have pointed out, English-language studies on migration-related diversity have been dominated by researchers based in English-speaking and Northern European countries.

Secondly, this study highlights several promising topics covering socioeconomic, cultural, and political dimensions of integration, health, immigration and integration governance, demographic issues, and theoretical and methodological concerns. Furthermore, the attention given to topics in this field has varied over time. While some topics such as «integration theory», «political participation», and «integration policies» showed a downward trend, topics related to «racism and discrimination», «acculturative stress», «ethnic identity, sense of belonging», and «economic integration» showed an upward trend. Additionally, topics such as «mobility», «residential segregation», «language training», «religious diversity», «cultural participation», and «education» demonstrated temporal stability in terms of prevalence. Finally, the emergence of new topics in recent years, notably topics such as «access to services and healthcare», «health and well-being», «gender and violence», and «environmental issues», is noteworthy. The variations of topic over the years imply that research on the integration of immigrants undergoes constant flux. In addition to the diversity of topics, their classification into groups based on their importance or priority assigned by researchers has also changed over time. The field has shifted from focusing on socioeconomic dimensions and demographic issues to an increasing emphasis on governance, mobility, humanitarian aspects, health issues, and sociocultural aspects. These findings align with other studies (Pisarevskaya et al., 2019).

Thirdly, our findings suggest that topics tend to cluster together but also differently. For instance, topics such as «Political participation», «Identity & belonging», «Religious diversity», «Racism and discrimination», and «Cultural participation» are investigated together, implying that this could be a potential research domain on the integration of immigrants. However, our correlation analysis among the topics identifies gaps in the literature, meaning that some topics are seldom discussed simultaneously in existing research. For example, we can critically examine the topic correlation figure to identify missing links in addition to those that are present. Our study reveals minimal correlation or a lack of connections among topics related to «Gender and violence», «Marriage formation», «Parenthood» and «Demographic issues» prompting further studies for enhanced theorization on women, marriage, family, immigrants' integration, and the significance of gender inequality. The absence of specific topic links could hinder

theoretical development in immigrants' integration. This suggests that actions must be implemented to facilitate greater collaboration among researchers from different disciplines. We also recommend that journals and scientific and academic responsible for promoting researchers value interdisciplinary articles.

Fourthly, large countries and institutions had a more significant influence on several specific topics. For example, the distributions of topics concerning all aspects of immigrants' integration are almost similar between countries such as the USA, the UK, Canada, China, Australia, and many other countries of immigration. Furthermore, topic trends among countries have remained relatively consistent over time. This implies that the issue of immigrants' integration is discussed equitably across immigration countries over time.

The article presents a methodological contribution by providing a guide on how automated text analysis, particularly Structural Topic Modeling, combined with bibliometric analysis, can be used to study complex and contested research questions in social sciences. These methods allow researchers to analyze large amounts of textual data efficiently and effectively. We can employ topic modeling to identify unexplored research directions in the field. By repeating the analysis in 20 years, we could assess whether stronger connections emerge between various topics and whether the area gives more attention, for example, to these questions. Additionally, we could observe whether issues such as economics, diversity, cultural and religious identity, sense of belonging, and cultural participation – which seem to receive more attention in the field – emerge as distinct topics over time.

Despite these contributions, our study has a few limitations that are important to note. The first limitation concerns the definition of the search terms we use to generate a results list. While the search terms were relatively broad, it is conceivable that adding other terms could have augmented the number of pertinent articles. The second limitation is that our electronic search was restricted to only three bibliometrics databases and three series of books. Unfortunately, the electronic catalogs of Scopus, Web of Science, and Dimensions do not list all articles and abstracts published by all journals. Their policy is to collect articles and abstracts whenever available. Consequently, numerous original articles were excluded from our dataset because they are not indexed in these repositories. Although we maintain that our data selection criteria provide an acceptable representation of the primary integration research journals, we recognize that our selection is somewhat subjective. Future research could combine data from JSTOR with the databases used in this study to test the validity and robustness of our findings. Thirdly, the corpus analyzed here is not exhaustive of all research legitimately belonging to the field of immigrants' integration. By limiting the analysis to articles published in English, we exclude important articles published in other languages. Future research could integrate articles in English and other languages to evaluate the robustness of our findings. The fourth limitation relates to the modeling of topics, as it is not as thorough as a close reading of the texts (Grimmer & Stewart, 2013). This approach cannot reveal «why specific themes are likely to be found together, nor can it explain why specific themes have changed over time» (Mostafa, 2023).

Nevertheless, topic modeling offers valuable opportunities for social science researchers, as we often use content analysis with data that appear impossible to analyze due to their considerable volume. Indeed, employing the bibliometric approach

based on STM methods provides a valuable methodological strategy to identify the linguistic contexts surrounding social institutions (Daenekindt & Huisman, 2020) and to assess developmental trends in an academic field (Chen et al., 2021). Further extensions of this work should incorporate result validation through human coding of a portion of the selected dataset, comparison with existing qualitative studies, or combining topic modeling techniques with qualitative methods.

In conclusion, as the results indicate, the bibliometric approach based on STM methods applied to this corpus generates commonly discussed topics and other emerging topics in immigrant integration studies. Addressing the gaps identified by our analysis provides valuable starting points for future research. The implications of this analysis can benefit researchers, helping them better understand the current state of research and design future research projects. The findings could also be helpful for stakeholders in immigration and integration governance and funding agencies to guide policies regarding immigrants' integration.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s12134-024-01139-8>.

**Data Availability** The data used in this study are available on request from the corresponding author.

## Declarations

**Conflict of interest** No potential conflict of interest was reported by the authors.

## References

- Ager, A., & Strang, A. (2008). Understanding integration: A conceptual framework. *Journal of Refugee Studies*, 21(2), 166–191. <https://doi.org/10.1093/jrs/fen016>
- Agrawal, S. K. (2019). Canadian refugee sponsorship programs: experience of Syrian refugees in Alberta, Canada. *Journal of International Migration and Integration*, 20(4), 941–962. <https://doi.org/10.1007/s12134-018-0640-7>
- Alba, R., & Foner, N. (2015). *Strangers no more: immigration and the challenges of integration in North America and Western Europe*. Princeton Univ. Press.
- Alburez-Gutierrez, D., Mason, C., & Zagheni, E. (2021). The sandwich generation revisited: global demographic drivers of care time demands. *Population and Development Review*, 47(4), 997–1023.
- Atçeken, K., & Dik, E. (2023). The 11 Years bibliometric analysis of Syrian migration. *Journal of Identity and Migration Studies*, 17(1), 67–85.
- Bakewell, O. (2008). Research beyond the categories: The importance of policy irrelevant research into forced migration. *Journal of Refugee Studies*, 21(4), 432–453. <https://doi.org/10.1093/jrs/fen042>
- Bekteshi, V., & Kang, S. (2020). Contextualizing acculturative stress among latino immigrants in the United States: A systematic review. *Ethnicity & Health*, 25(6), 897–914. <https://doi.org/10.1080/13557858.2018.1469733>
- Bilodeau, A. (2017). Mobilisation or demobilisation? Perceived discrimination and political engagement among visible minorities in Quebec. *Political Science*, 69(2), 122–138. <https://doi.org/10.1080/00323187.2017.1332955>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993–1022.
- Brzozowski, J. (2019). Entrepreneurship and economic integration of immigrants: A critical review of literature. *International Journal of Entrepreneurship and Innovation Management*, 23(6), 584. <https://doi.org/10.1504/IJEIM.2019.102803>

- Cedillo, Y. E., Bertrand, B., Baker, E., Cherrington, A. L., Beasley, T. M., & Fernández, J. R. (2021). Assimilation, Acculturation, and allostatic load in U.S.- and foreign-born hispanics. *Journal of Immigrant and Minority Health*, 23(1), 35–44.
- Chen, Y., & Wang, J. (2015). Social integration of new-generation migrants in Shanghai China. *Habitat International*, 49, 419–425. <https://doi.org/10.1016/j.habitatint.2015.06.014>
- Chen, X., & Xie, H. (2020). A structural topic modeling-based Bibliometric Study of Sentiment Analysis Literature. *Cognitive Computation*, 12(6), 1097–1129. <https://doi.org/10.1007/s12559-020-09745-1>
- Chen, X., Chen, J., Cheng, G., & Gong, T. (2020). Topics and trends in artificial intelligence-assisted human brain research. *PLoS One*, 15(4), e0231192. <https://doi.org/10.1371/journal.pone.0231192>
- Chen, X., Zou, D., Xie, H., & Cheng, G. (2021). A topic-based bibliometric review of computers in human behavior: contributors, collaborations, and research topics. *Sustainability*, 13(9), 4859. <https://doi.org/10.3390/su13094859>
- Daenekindt, S., & Huisman, J. (2020). Mapping the scattered field of research on higher education. A correlated topic model of 17,000 articles, 1991. *Higher Education*, 80(3), 571–587. <https://doi.org/10.1007/s10734-020-00500-x>
- del Carmen Salazar, M., Martínez, L. M., & Ortega, D. (2016). Sowing the Semillas of critical multicultural citizenship for Latina/o undocumented youth: Spaces in school and out of school. *International Journal of Multicultural Education*, 18(1), 88–106. <https://doi.org/10.18251/ijme.v18i1.1082>
- Demie, F. (2018). English as an additional language and attainment in primary schools in England. *Journal of Multilingual and Multicultural Development*, 39(3), 210–223. <https://doi.org/10.1080/01434632.2017.1348508>
- Drouhot, L. G., Deutschmann, E., Zuccotti, C. V., & Zagheni, E. (2023). Computational approaches to migration and integration research: Promises and challenges. *Journal of Ethnic and Migration Studies*, 49(2), 389–407.
- Erhard, L., Heiberger, R. H., & Windzio, M. (2022). Diverse effects of mass media on concerns about immigration: new evidence from Germany, 2001. *European Sociological Review*, 38(4), 629–647. <https://doi.org/10.1093/esr/jcab063>
- Feliciano, C., & Rumbaut, R. G. (2005). Gendered paths: Educational and occupational expectations and outcomes among adult children of immigrants. *Ethnic and Racial Studies*, 28(6), 1087–1118. <https://doi.org/10.1080/01419870500224406>
- Frenette, M., & Morissette, R. (2005). Will they ever converge? Earnings of immigrant and Canadian-Born workers over the last two decades. *The International Migration Review*, 39(1), 228–257.
- Galardini, S., Mulvey, G., & Lessard-Phillips, L. (2019). Stuck between mainstreaming and localism: views on the practice of migrant integration in a devolved policy framework. *Journal of International Migration and Integration*, 20(3), 685–702. <https://doi.org/10.1007/s12134-018-0626-5>
- Gao, H., & Wang, S. (2022). The intellectual structure of research on rural-to-urban migrants: a bibliometric analysis. *International Journal of Environmental Research and Public Health*, 19(15), 9729. <https://doi.org/10.3390/ijerph19159729>
- Garcés-Mascareñas, B., & Penninx, R. (2016). Introduction: Integration as a three-way process approach? In B. Garcés-Mascareñas & R. Penninx (Eds.), *Integration Processes and Policies in Europe: Contexts, Levels and Actors* (pp. 1–9). Springer International Publishing. [https://doi.org/10.1007/978-3-319-21674-4\\_1](https://doi.org/10.1007/978-3-319-21674-4_1)
- Goodman, S., & Kirkwood, S. (2019). Political and media discourses about integrating refugees in the UK. *European Journal of Social Psychology*, 49(7), 1456–1470. <https://doi.org/10.1002/ejsp.2595>
- 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. <https://doi.org/10.1093/pan/mps028>
- Haberfeld, Y., Cohen, Y., Kalter, F., & Kogan, I. (2011). Differences in earnings assimilation of immigrants from the Former Soviet Union to Germany and Israel during 1994–2005: The interplay between context of reception, observed, and unobserved immigrants' attributes. *International Journal of Comparative Sociology*, 52(1), 6–2. <https://doi.org/10.1177/0020715210377157>
- Hajdinjak, S., Morris, M. H., & Amos, T. (2020). Migration policy framing in political discourse: evidence from Canada and the USA. In E. Deutschmann, J. Lorenz, L. G. Nardin, D. Natalini, & A. F. X. Wilhelm (Eds.), *Computational Conflict Research* (pp. 83–111). Springer International Publishing. [https://doi.org/10.1007/978-3-030-29333-8\\_5](https://doi.org/10.1007/978-3-030-29333-8_5)
- Kajikawa, Y., Ohno, J., Takeda, Y., Matsushima, K., & Komiyama, H. (2007). Creating an academic landscape of sustainability science: An citation network analysis. *Sustainability Science*, 2(2), 221. <https://doi.org/10.1007/S11625-007-0027-8>
- Kashyap, R., Rinderknecht, R. G., Akbaritabar, A., Albrez-Gutierrez, D., Gil-Clavel, S., Grow, A., Kim, J., Leasure, D. R., Lohmann, S., Negraia, D. V., Perrotta, D., Rampazzo, F., Tsai, C. J., Verhagen, M. D.,

- Zagheni, E., & Zhao, X. (2023). Digital and computational demography. *Research Handbook on Digital Sociology* (pp. 48–86). Edward Elgar Publishing. <https://doi.org/10.4337/9781789906769>
- Kaufmann, L. (2021). Integration in Canada: A systematic review of the youth experience [Review]. *International Journal of Intercultural Relations*, 84, 52–64. <https://doi.org/10.1016/j.ijintrel.2021.06.010>
- Kelling, C., & Monroe, B. L. (2022). Analyzing community reaction to refugees through text analysis of social media data. *Journal of Ethnic and Migration Studies*. <https://doi.org/10.1080/1369183X.2022.2100551>
- Khattab, N., & Fox, J. (2016). East-European immigrants responding to the recession in Britain: Is there a trade-off between unemployment and over-qualification? *Journal of Ethnic and Migration Studies*, 42(11), 1774–1789. <https://doi.org/10.1080/1369183X.2016.1166040>
- Laurentsyeva, N., & Venturini, A. (2017). The social integration of immigrants and the role of policy a literature review. *Intereconomics - Review of European Economic Policy*, 2017(5), 285–292. <https://doi.org/10.1007/s10272-017-0691-6>
- Lenard, P. T. (2015). Residence and the right to vote. *Journal of International Migration and Integration*, 16(1), 119–132. <https://doi.org/10.1007/s12134-014-0328-6>
- Li, M., Lin, H. C., Li, L. M. W., & Frieze, I. H. (2019). Cross-cultural study of community engagement in second-generation immigrants. *Journal of Cross-Cultural Psychology*, 50(6), 763–788.
- Lindstedt, N. C. (2019). Structural topic modeling for social scientists: A brief case study with social movement studies literature, 2005–2017. *Social Currents*, 6(4), 307–318. <https://doi.org/10.1177/2329496519846505>
- Malaterre, C., Chartier, J. F., & Pulizzotto, D. (2019). What is this thing called philosophy of Science? A computational topic-modeling perspective, 1934–2015. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, 9(2), 215–249. <https://doi.org/10.1086/704372>
- Martinovic, B., & Verkuyten, M. (2016). Inter-religious feelings of Sunni and Alevi Muslim minorities: The role of religious commitment and host national identification. *International Journal of Intercultural Relations*, 52, 1–1.
- McAvay, H. (2018). How durable are Ethnoracial Segregation and spatial disadvantage? Intergenerational contextual mobility in France. *Demography*, 55(4), 1507–1545. <https://doi.org/10.1007/s13524-018-0689-0>
- Meissner, F., & Vertovec, S. (2015). Comparing super-diversity. *Ethnic and Racial Studies*, 38(4), 541–555. <https://doi.org/10.1080/01419870.2015.980295>
- Miranda-González, A., Aref, S., Theile, T., & Zagheni, E. (2020). Scholarly migration within Mexico: Analyzing internal migration among researchers using Scopus longitudinal bibliometric data. *EPI Data Science*, 9(1, 1), 1–26.
- Morrison, T. B., Wieland, M. L., Cha, S. S., Rahman, A. S., & Chaudhry, R. (2012). Disparities in preventive health services among Somali immigrants and refugees. *Journal of Immigrant and Minority Health*, 14(6), 968–974. <https://doi.org/10.1007/s10903-012-9632-4>
- Mostafa, M. M. (2023). A one-hundred-year structural topic modeling analysis of the knowledge structure of international management research. *Quality & Quantity*, 57(4), 3905–3935.
- O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: A systematic review of current approaches. *Systematic Reviews*, 4(1), 5. <https://doi.org/10.1186/2046-4053-4-5>
- Penninx, R., & Garcés-Masareñas, B. (2016). The concept of integration as an analytical tool and as a policy concept. In B. Garcés-Masareñas & R. Penninx (Eds.), *Integration Processes and Policies in Europe: Contexts, Levels and Actors* (pp. 11–29). Springer International Publishing. [https://doi.org/10.1007/978-3-319-21674-4\\_2](https://doi.org/10.1007/978-3-319-21674-4_2)
- Peters, M. D. J., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., & Soares, C. B. (2015). Guidance for conducting systematic scoping reviews. *JBI Evidence Implementation*, 13(3), 141–146. <https://doi.org/10.1097/XEB.0000000000000050>
- Phillimore, J., & Goodson, L. (2008). Making a place in the global city: The relevance of indicators of integration. *Journal of Refugee Studies*, 21(3), 305–325. <https://doi.org/10.1093/jrs/fen025>
- Picanço Cruz, E., & de Queiroz Falcão, R. P. (2016). A bibliometric review of immigrant and ethnic entrepreneurship: Revisão bibliométrica no tema Empreendedorismo Imigrante E Étnico. *Intermetx: Revista Eletrônica De Negócios Internacionais Da ESPM*, 11(3), 78–94. <https://doi.org/10.18568/1980-4865.11378-94>
- Pisarevskaya, A., Løving, N., Scholten, P., & Jansen, J. (2019). Mapping migration studies: an empirical analysis of the coming of age of a research field. *Migration Studies*, mnz031, 1–27. <https://doi.org/10.1093/migration/mnz031>
- Roberts, M., 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. <https://doi.org/10.1111/ajps.12103>

- Roberts, M., Stewart, B. M., & et Tingley, D. (2016). Navigating the local modes of big data: the case of topic models. In R. M. Alvarez (dir.), *Computational Social Science* (p. 51–97). Cambridge University Press. <https://doi.org/10.1017/CBO9781316257340.004>.
- Rodriguez, M. Y., & Storer, H. (2019). A computational social science perspective on qualitative data exploration: using topic models for the descriptive analysis of social media data\*. *Journal of Technology in Human Services*, 1–32. <https://doi.org/10.1080/15228835.2019.1616350>
- Rodríguez-García, D. (2015). Inter-marriage and integration revisited: international experiences and cross-disciplinary approaches. *The Annals of the American Academy of Political and Social Science*, 662(1), 8–36. <https://doi.org/10.1177/0002716215601397>
- Salehi, R. (2010). Intersection of health, immigration, and youth: a systematic literature review. *Journal of Immigrant and Minority Health*, 12(5), 788–797. <https://doi.org/10.1007/s10903-009-9247-6>
- Scholten, P., & Penninx, R. (2016). The Multilevel Governance of Migration and Integration. In B. Garcés-Masareñas & R. Penninx (Eds.), *Integration Processes and Policies in Europe: Contexts, Levels and Actors* (pp. 91–108). Springer International Publishing. [https://doi.org/10.1007/978-3-319-21674-4\\_6](https://doi.org/10.1007/978-3-319-21674-4_6).
- Schwemmer, C., & Wieczorek, O. (2019). The methodological divide of sociology: evidence from two decades of journal publications. *Sociology-the Journal of The British Sociological Association*, 0038038519853146. <https://doi.org/10.1177/0038038519853146>
- Sharma, A., Rana, N. P., & Nunkoo, R. (2021). Fifty years of information management research: A conceptual structure analysis using structural topic modeling. *International Journal of Information Management*, 58, 102316. <https://doi.org/10.1016/j.ijinfomgt.2021.102316>
- Shuangyun, Y., & Hongxia, L. (2020). Knowledge mapping of ethnic identity and acculturation based on the method of bibliometric. *Journal of Ethnic and Cultural Studies*, 7(3), 17–33. <https://doi.org/10.29333/ejecs/370>
- Smith, R., Spaaij, R., & McDonald, B. (2019). Migrant integration and cultural capital in the context of sport and physical activity: a systematic review. *Journal of International Migration and Integration*, 20(3), 851–868. <https://doi.org/10.1007/s12134-018-0634-5>
- Subbotin, A., & Aref, S. (2021). Brain drain and brain gain in Russia: analyzing international migration of researchers by discipline using Scopus bibliometric data 1996. *Scientometrics*, 126(9), 7875–7900.
- Sweileh, W. M., Wickramage, K., Pottie, K., Hui, C., Roberts, B., Sawalha, A. F., & Zyoud, S. H. (2018). Bibliometric analysis of global migration health research in peer-reviewed literature (2000–2016). *BMC Public Health*, 18(1), 777. <https://doi.org/10.1186/s12889-018-5689-x>
- Szlovak, P. (2017). Exploring the reasons for migrant under-representation in Irish electoral politics: The role of Irish political parties. *Ethnic and Racial Studies*, 40(5), 851–870. <https://doi.org/10.1080/01419870.2016.1259489>
- Thornton, M. C., Taylor, R. J., & Chatters, L. M. (2012). African American, black caribbean, and non-hispanic white feelings of closeness toward other racial and ethnic groups. *Journal of Black Studies*, 43(7), 749–772.
- Tran, V. (2015). Immigrant assimilation. In M. Jeff (Ed.), *Obo in sociology*. Oxford University Press.
- Vargas, E. D., Winston, N. C., Garcia, J. A., & Sanchez, G. R. (2016). Latina/o or Mexicana/o? The relationship between socially assigned race and experiences with discrimination. *Sociology of Race and Ethnicity (Thousand Oaks Calif)*, 2(4), 498–515. <https://doi.org/10.1177/2332649215623789>
- Vertovec, S. (2007). Super-diversity and its implications. *Ethnic and Racial Studies*, 30(6), 1024–1054. <https://doi.org/10.1080/01419870701599465>
- Wang, S. H., Ding, Y., Zhao, W., Huang, Y. H., Perkins, R., Zou, W., & Chen, J. J. (2016). Text mining for identifying topics in the literature about adolescent substance use and depression. *BMC Public Health*, 16(1), 279. <https://doi.org/10.1186/s12889-016-2932-1>
- Zhao, X., Aref, S., Zagheni, E., & Stecklov, G. (2021, April 27). International migration in academia and citation performance: an analysis of German-affiliated researchers by gender and discipline using Scopus Publications 1996–2020. <https://doi.org/10.48550/arXiv.2104.12380>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.