

Return migration to home country: a systematic literature review with text mining and topic modelling

Cecilia Fortunato^a, Andrea Iacobucci^b, and Elena Ambrosetti^a

^a Sapienza University of Rome; cecilia.fortunato@uniroma1.it, elena.ambrosetti@uniroma1.it

^b Regione Emilia-Romagna; iacobucciandrea@gmail.com

Abstract

A crucial (and less developed) part of migration studies is the exploration of migrant's further mobility and the intention of return to home country at some point in life. Knowing who, why and when returns matters for both the host and the home country. Very few studies have focused on return in a wider perspective, adopting a comparative approach. The present study aims at providing a systematic review of peer-reviewed literature indexed in Scopus database, to understand how return has been dealt with by researchers. The main objectives are: collecting and synthesizing previous studies; comparing approaches, conceptualizations of return, methods and variable of interest. A bibliometric analysis on metadata and content analysis based on text mining and topic modelling techniques has been conducted on a sample of approximately 3,000 publications. With our contribution, we expect to implement a baseline for theoretical development and empirical research, presenting an overview on the evolution of trend topics, with regional and temporal patterns of research focus and identifying knowledge gaps in literature.

Key words: return migration, remigration, home country, systematic literature review, topic modelling

1. Introduction

Migration studies have developed rapidly in recent decades, in terms of size, methods used, interdisciplinarity, heterogeneity of theoretical frameworks and internationalization of research groups (Pisarevskaya et al., 2020; Vargas-Silva, 2012; Bonifazi and Strozza, 2006; King and Skeldon, 2010; Massey et al., 1993).

Migration flows and many aspects of the migration cycle have been widely discussed in the literature, relating to a growing diversity of migratory categories (from refugees and forced migrants to economic or climate migrants) and specific case-studies at national and global level (Abel, 2018; Castels and Miller, 2014).

Researchers have concentrated their efforts in analyzing why people decide to migrate. But a crucial (and poorly developed) part of migration knowledge is the exploration of migrant's further mobility over the life course. A fundamental theoretical and practical question should be: what are future plans and prospects of migrants? What factors influence their decision for permanent stay in host country, for onward migration to a third destination, or for the return to home country, at some point in life? Awareness on the whole migratory cycle is scarce. Yet, knowing who, why and when returns, estimating future migration scenarios and forecasting returns, matters for both the host and the home country. And this aspect could be even more relevant in times of crisis, as shown by the COVID-19 pandemic, when millions of migrants have been returned to their countries of origin, we don't know if permanently, and large numbers of migrants have found themselves "stranded abroad and in need of assistance" (Le Coz and Newland, 2021).

Modern fluid societies, characterized by the globalization of movements, communications, technologies and information, make the ground extremely rough for analysis, because of the volatility of events, the heterogeneity in biographies and experiences and the increasing transnationalism of migrants. Return decision making is a complex process (Carling and Pettersen, 2014; Waldorf, 1995): it is influenced by a variety of

factors related to conditions both in origin and destination countries and motivated by needs at individual, household and social level (Krasniqi and Williams, 2018; Paparusso and Ambrosetti, 2017). Migrants' characteristics and situational conditions affect return probabilities and reintegration prospects of returnees. Furthermore, return to home country is particularly difficult to measure because of the lack of reliable and consistent register data and large-scale survey data, as it often goes spontaneous and unrecorded.

We know from literature and international reports that 1/4 of total migratory events are estimated to be returns; up to 50% of immigrants leave the host country within 5 years (return home or secondary emigration) and return migrant's selection tends to be the reverse of the initial selection process for migration (Azose and Raftery 2019; Abel 2018; OECD, 2017; Åkesson and Baaz, 2015).

However, very few studies have focused on return in a wider perspective, adopting a comprehensive and comparative approach (see Cassarino, 2015; Mohamed and Abdul-Talib, 2020)

2. Main objectives and Research Questions

Knowledge production within the field of migration is accelerating at a tremendous speed, while risking at the same time to remain over-specialized, fragmented and inconsistent (Pisarevskaya, 2020; Snyder, 2019; Denyer and Tranfield, 2009; Kitchenham, 2004; Tranfield, et al., 2003; Webster and Watson, 2002,).

The present study aimed at providing a theoretical basis necessary for the further analysis of factors driving intentions and realizations of migrant's homeland return, and the relationship between initial migration drivers and return intentions. Mapping academic landscape is an important step in addressing new research questions, proposing new conceptual framework, facilitating theory development and identifying knowledge gaps and topics that need further research.

The systematic review of peer-reviewed literature indexed on citation databases is a powerful tool for understanding how return has been dealt with by researchers. What has been studied so far may shed light on what has been relevant in time, assuming that fundamental research questions arise with the development of societies and follow the evolution of phenomena under study (Snyder, 2019).

The main objective of this study was collecting and synthesizing previous studies, comparing approaches, conceptualizations of return, methods and variable of interest, integrating findings and perspectives, testing return theories against international migration theories and comparing the assumptions on which they rest. We aimed at providing a detailed overview of the existing literature through a comprehensive search in publication's repositories, using relevant keywords.

The following questions guided our study:

RQ1: How the studies on return migration developed from 1960 to 2020?

RQ2: How has return been conceptualized in terms of topical focus?

RQ3: What theoretical framework and motivational factors have been taken into account?

RQ4: Is the evolution of "return theories" following the evolution of "migration theories"?

3. Data and Method

Bibliometric analysis and content analysis based on text mining and topic modelling algorithm have been applied to publications, gathered from Sci-Verse Scopus.

The first step was building a valid search query (combinations of key terms and logical operators) that would retrieve (in title, author's keywords or abstract) as many documents as possible with minimum irrelevant results. When possible, we gathered publication and metadata via API, alternatively, we gathered data with search-by-string method (Belter, 2020; Aria and Cuccurullo, 2017-2018). We applied a massive data cleaning procedure to exclude out of scope articles and to clean text with lemmatization and tokenization techniques. For data analysis, we selected specific packages of the statistical software R: dplyr, ldatuning, topicmodels, rscopus, bibliometrix.

The descriptive analysis of the collection and the bibliometric analysis of meta-data aimed at discovering the temporal and spatial distribution of publications, evaluating the impact of a publication in the scientific community, identifying important theoretical contributions and mapping the evolution in international research cooperation networks.

The content analysis of title-keywords-abstract with text mining and topic modelling techniques identified topics or topic clusters that figure centrally in the specific textual landscape. We aimed at

discovering patterns and regularities within the corpus of texts, elaborating classification and categorization of topics, evaluating correlations among topics and following the evolution of trend topics. For text analysis we use the TF-IDF technique (term frequency-inverse document frequency), assessing how relevant a word is to a document in a collection of documents, on the base of how many times the word occurs in the document; and the inverse document frequency of the word across a set of documents. We considered the absolute frequency of a term in the whole corpus of text (frequency) as well as the numbers of documents containing the specific token (count).

For topic modelling, Latent Dirichlet Allocation model has been selected (Farren, 2019). LDA model is a Bayesian probabilistic model of latent topics from the contents of abstracts, based on the assumptions that each document in a corpus discusses multiple topics in differing proportions. A topic can be considered as a collection of words ordered by their probability of occurrence. The topic structure in the corpus of words is hidden and this method seeks to provide an ideal number of topics (k), a matrix with per-topic word probabilities and a matrix with per-document topic proportions. As a result, it was possible to summarize and label the topic in topic clusters and their probability of occurrence. Two metrics have been defined to summarize topic distribution across documents and classify articles: the topic presence, showing the number of document in which the topic is present, equal to 1 if the gamma score of LDA model is above the per topic average of gamma (and 0 otherwise); and the topic dominance, showing the number of document in which the topic is prevailing over other topics, selecting the topic with maximum gamma score for each document.

4. Main results

Our final sample included almost 3.000 documents published in English between 1960 and 2020 by more than 1.400 publishers.

The great majority of the publications focusing on return were articles published in specialized journals (2469) and only the 12.9% of them were published in open access. The first publishers, in terms of number of publications and citation obtained were migration's specialized scientific journals such as *International Migration* and the *Journal of Ethnic and Migration Studies*; the only regional journal among the top publishers was the *Asian and Pacific Migration Journal*. Among the top 15 publishers, two were journals specialized on refugees and forced migration, a fundamental component of the migration phenomenon (*Journal of Refugee Studies* and the *Refugee Survey Quarterly*).

Our results from the descriptive and bibliometric analysis of meta-data showed that first return migration studies can be traced back to the beginning of the 1960s, with an exponential growth of scientific production on the topic starting from the late seventies. However, a proper corpus of studies took shape only in the 1980s, with the politization of international migration movement and the growing importance of the discourse on (economic) development in origin countries, both in academia and in the institutional and public debate (Cassarino, 2015). The great acceleration of the topic's relevance started at the beginning of the new millennium, with researchers looking at the effects of the global economic crises on millions of migrants living in recessing high industrialized countries. Indeed, 60% of total studies have been published between 2010 and 2020.

Looking at the spatial distribution of publications, we highlighted the dominant role of receiving countries involved in the "knowledge production" on return. The 50% of total documents have been published in only 5 countries, namely United States, United Kingdom, Canada, Germany and Australia.

We then analysed the textual landscape presented in publications retrieved, mining the text of title, abstract and keywords. In the overall corpus of text, most frequent tokens present a variegated nature, including all the topical aspects of different migration theories. Among the top frequent tokens, a major role can be attributed to economic, push and pull factors and labour migration theories, followed by systemic and macrolevel theories and transnationalism and social network theories.

We then analysed most frequent tokens divided in three subperiods (1960-1980, 1981-2000 and 2001-2020), showing a shift in scientific focus over the last 40 years, with economic and labor market associated keywords leaving room for more subjective aspects of migration.

The first period considered (1960-1980) has been dominated by the economic paradigm. In this perspective, the return has been considered as the "logical outcome of a calculated strategy" (Cassarino, 2015), underlining the importance of wage differentials, accumulations of savings and, more important, remittances as explanatory factors of the return decision. In the second period considered (1981-2000),

words related to the economic perspective are still relevant, but a growing attention has been paid to the macro-level influence and constraints on return migration, especially related to the increasing number of conflicts and humanitarian crises around the world during the '90s. In that period, experts and researchers from different disciplines implemented a series of new empirical tools ("model") to study migration trends, and most of them started to take into account meso-level factors (group and community) behind migration and return, such as the new role of family in migration strategies and the increasing number of women among international migrants. In the beginning of the new millennium (2001-2020), the phenomenon of return gained another level of interest as well as complexity of analytical frameworks. The increasing globalization of information, facilitated by communication and transportation, the cyclical economic crisis and the higher mobility of high skilled workers, led the researcher to focus on social capital and transnational practices. Many studies still point out that economic factors play a crucial role in homeland return's decision, but there is a slight cultural shift in the economic paradigm as the focus has turned into a matter of "aspirations" for an upward-mobility. As it was predictable, in the last twenty years the discourse on return has been dominated by an increasing attention on the refugee return to post-conflict areas, and it is worth mentioning the growing importance of gender perspective in research landscape.

Topic models applied to our textual data with the LDA algorithm provided the 17 topics that best describe our data. Each topic is defined by a string of words that are semantically connected, according to LDA model. We analysed 10 selected words among the top 30 words defining the topic and assigned a label to each topic. We then analysed relevance of each topic in the whole corpus.

In the overall period, the dominant topic focuses on cultural, ethnic and identity issues, defined by words related to diasporic communities and the difficult compromise between the need of preserving the own cultural identity and the need of getting integrated in the host society. The second dominant topic is the classical economic paradigm anticipated by the top frequent words, concerning skills, labour market opportunities and wage differentials, highlighting the rational benefit-cost perspective upon the possibility of return based on migrants' selection, skills, education, wage and returnees' prospects once back to their origin country (for temporary or permanent return). The third dominant topic is the topic on refugee and post-conflict return. From our results, lower attention has been paid to the topic of circular mobility in conflict areas, a topic mainly defined by a set of words related to specific territory and repeated border crossing between two "national states". Also, the topic of European migration flows, mainly related to internal temporary movement in Europe has attracted a lower interest compared to other topics.

Dividing the whole period in the above mentioned three subperiods, we presented our results starting with topics that have gained importance in researchers' works in the last 60 years. The most relevant topic in 2000-2020 concerns "Cultural, ethnic and political identity", marginal in the first period (dominant in only 3% of documents) but gaining importance in the last twenty years of the last century and dominant in almost the 16% of total document published in the last period considered. The second topic in terms of relevance in the last period (2001- 2020) is related to "Refugee and post-conflict return", experiencing a constant increase in relevance and dominant in the 13% of publications of recent years. As already reported, the topic of "Skills, labour market opportunities and wage differentials" has been relevant in the whole scientific production on return migration, but at the same pace of the growing importance of the neoclassical economic paradigm proposed by Todaro in 1969, it reached the maximum interest in the eighties, being dominant in the 15% of total publications of the period. It can be considered one of the most important topical focus also in the last 20 years, being dominant in the 13% of scientific publications. As anticipated by the top tokens cited in publications in the last period considered, the topic on "Gender, family and vulnerabilities" has gained growing importance in research already in the central period, following an increasing female migration flow, primarily from former Soviet Union Republics, and becoming crucial with the increasing process of family reunifications in "mature" host countries and the aging process and the subsequent higher need of caregivers in high income countries. The topic defined by words expressing a wider "Systemic approach and multiple focus", is dominant in a growing number of publications in the whole period considered, maintaining however a stable proportion of 10% of publications in which the topic is dominant in the three different periods considered.

As for topic that have experienced a vigorous decline in the interest received from the scientific community, we first mention the topic on "South-North migration, Africa and demographic imbalances", which appears to be the most discussed between 1960 and 1980, dominant in almost the 28% of researches published in that period. In recent decades, instead, the topic has been the focal point only in

the 3% of publications. Such decline in relevance could be partially explained by the persistent lack of consistent data in African countries, discouraging researchers always seeking for large-scale data to conduct robust quantitative analysis. Another reason could be the growing complexity of migration phenomena, the consequent questioning of the “South-North” migration model and the growing attention referred to emergent “South-South” mobility patterns. The topic identified as “Rural-urban mobility” and mainly referring to the context of rural Asian countries, has also experienced a sustained and constant decline, being dominant in 15% of publications in the sixties and only in the 7% in the beginning of the millennium. Same destiny has the topic on “Economic-differential and cost-opportunities assessment”, with declining dominance from 10% to 4% of total publications in the whole period. This lower interest in a topic defined by words that are key points of the New Economics of Labour Migration approach (NELM), can be interpreted as an overcome of the classical view of return migration as the logical outcome of a “calculated strategy” in favour of a greater attention toward new paradigms focused on identity and cultural issues, transnational practices and social network effects on return’s decision-making process. An expected result is the decline of interest in the topic related to “EU migration flows and temporary/permanent stay”, dominant in only 1% of recent publications. The first 10 selected words suggested a focus on mobility occurred in the last half of nineteenth century, presenting very different characteristics from nowadays migration patterns in terms of migrants’ selection, education and skills, information and opportunities, migration policies and mobility in Europe, social networks and family settlement. However, understanding what happened to guest workers migrated to European countries in that period could be of great importance to understand present patterns of aspirations to return.

The topic of “Education/qualification opportunities and high skilled mobility”, attracted a small interest in the eighties, probably because of the average higher proportion of unskilled migrants, but regained importance in last 20 years. The same trend is shown for the topic of “Eastern to Northern Europe migration”, probably related to old migration pattern, not highly representative of current situation. And also, the decline in relevance of the topic on “Optimal strategies, return/ onward mobility” can be explained with an overcome of the neoclassical and the NELM paradigms.

Lastly, marginal topics in our dataset, meaning dominant in less than 3% of the publications from the first to the last period are “Circular mobility in conflict areas”, “Southern Asia-Pacific labor migration”, “Socialization, linguistic skills and integration”, (Higher education mobility and opportunities”.

5. Conclusions and further development

Topic modelling provides an affordable way to classify scientific papers and discover hidden characteristics of documents in a corpus of text. However, natural language processing still needs expert’s supervision and interpretation and, in some cases, the full interpretability of topics can be challenging. Also, the way of indexing publications with meta-data and keywords, the query selected for data retrieving, filters applied in data cleaning process and the authors’ classification procedure are, intuitively, of fundamental importance for systematic review results.

We found consistent results with what has been highlighted by the scientific literature in the field and the evolution of international migration theories, keeping in mind that ongoing processes of social transformation will perpetually create the need for theoretical innovation (de Haas 2021).

Our results clearly show the dominance of receiving countries in the knowledge’s production of return migration, stressing the need of sending countries perspectives on return prospects of reintegration and effective returnees.

Moreover, the number of topics identified and the evolution of their relevance in time, suggests an over-fragmentation of the field and the need for a more comprehensive, wider social-scientific perspective, supported by an historical-comparative approach, which is completely absent in our dataset of publications.

The aspiration-capability framework and the conceptualization of return capability within the interaction between structural constraint (macro and meso-level factors) and agency are “great absents” in the scientific field.

With the increasing role of both emigration and immigration states in controlling and managing migration and return, as suggested by literature (de Haas et al. 2019, Hollifield et al. 2014, Waldinger 2015, Casarino 2004) we suggest the relevance of the focus on how changing political contexts influence the possibility of return and reintegration experiences of returnees.

Our results also suggested that lower attention has been paid so far to the development of concept of

return in terms of intrinsic aspirations (subjective: identity, perceptions, networks and ties), instrumental aspirations (objective: costs/opportunities, policies and external factors) and effective capabilities, as mainly stressed by Carling and others (Carling, 2002; Carling and Petersen, 2014; Carling and Schewel, 2018; de Haas, 2014).

We finally reported the need of more attention on emerging migration patterns, such as the possibility of return for climate and environmental refugees, as well as the relevant question on what the age and gender composition of migrants' population can tell us about their probability of return, a topic that can be further investigated with the exploitation of decomposition and microsimulations techniques.

References

References

- [1] Abel G.J.: Estimates of Global Bilateral Migration Flows by Gender between 1960 and 2015. In: *International Migration*, Volume: 52 issue: 3, pp. 809-852 (2018)
- [2] Akesson, L., Baaz, M.E.: *Africa's Return Migrants: The New Developers?* In: Zed Books Ltd. (2015)
- [3] Aria M.: *Bibliometrix: Data Importing and Converting*, <https://www.bibliometrix.org/vignettes/Data-Importing-and-Converting.html> (2018)
- [4] Aria, M., Cuccurullo, C.: *Bibliometrix: An R-tool for comprehensive science mapping analysis*. In: *Journal of Informetrics*, 11(4), pp 959-975, Elsevier (2017)
- [5] Aria M., Cuccurullo C.: *Science Mapping Analysis with bibliometrix R-package*, https://bibliometrix.org/documents/bibliometrix_Report.html (2018)
- [6] Azose, J.J., Raftery, A. E.: Estimation of emigration, return migration, and transit migration between all pairs of countries. In: *Proceedings of the National Academy of Sciences Jan 2019*, 116 (1) 116-122 (2019)
- [7] Belter C.: *Text mining in R* <https://github.com/christopherBelter/textmining> (2020)
- [8] Bonifazi C., Strozza S.: *Conceptual Framework and Data Collection in International Migration*. In: Caselli G., Vallin J., Wunsch G. (eds.), *Demography: Analysis and Synthesis. A Treatise in Population*, Volume IV, Elsevier Inc., USA, 2006, pp. 537-554 (2006)
- [9] Carling, J.: Migration in the age of involuntary immobility: Theoretical reflections and Cape Verdean experiences, *Journal of Ethnic and Migration Studies*, 28:1, 5-42, DOI: 10.1080/13691830120103912 (2002)
- [10] Carling, J., Petersen, S.V.: Return migration intentions in the integration–transnationalism matrix. In: *International Migration*, Vol. 52 No. 6, pp. 13-30 (2014)
- [11] Carling, J., Schewel, K.: Revisiting aspiration and ability in international migration. In: *Journal of Ethnic and Migration Studies*, 44:6, 945-963, DOI: 10.1080/1369183X.2017.1384146 (2018)
- [12] Cassarino, J.P.: Theorising return migration: the conceptual approach to return migrants revisited. In: *International Journal on Multicultural Societies*, Vol. 6 No. 2, pp. 253-279 (2015)
- [13] Castles, S., Miller, M. J.: *The Age of Migration: International Population Movements in the Modern World*, 5th edn. Basingstoke: Palgrave Macmillan (2014)
- [14] Denyer, D., Tranfield, D.: Producing a systematic review. In: Buchanan, D.A. and Bryman, A. (Eds), *The SAGE Handbook of Organizational Research Methods*, pp. 671-689. Entrepreneurship: Theory and Practice, Vol. 20 No. 1, pp. 55-72 (2009)
- [15] Farren T.: *Beginner's Guide to LDA Topic Modelling with R: Identifying topics within unstructured text*, <https://towardsdatascience.com/beginners-guide-to-lda-topic-modelling-with-r-e57a5a8e7a25> (2019)
- [16] King, R. Skeldon, R.: Mind the Gap: Integrating Approaches to Internal and International Migration. In: *Journal of Ethnic and Migration Studies*, 36/10: 1619–46 (2010)
- [17] Kitchenham, B.: *Procedures for performing systematic reviews*, Keele, UK, Keele University, Vol. 33 No. 2004, pp. 1-26 (2004)
- [18] Krasniqi, B.A., Williams, N.: Migration and intention to return: entrepreneurial intentions of the diaspora in post-conflict economies. In: *PostCommunist Economies*, Vol. 31 No. 4, pp. 464-483 (2018)
- [19] Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E.: *Theories of International Migration: A Review and Appraisal*. In: *Population and Development Review*, 19/3: 431–66 (1993)
- [20] Mohamed, M. , Abdul-Talib, A.N.: Push–pull factors influencing international return migration intentions: a systematic literature review. In: *Journal of Enterprising Communities: People and Places in the Global Economy*, Vol. 14 No. 2, pp. 231-246 (2020)
- [21] OECD: *International Migration Outlook 2017*, OECD Publishing, Paris, https://doi.org/10.1787/migr_outlook-2017-en (2017)
- [22] Paparusso, A., Ambrosetti, E.: To stay or to return? Return migration intentions of Moroccans in Italy. In: *International Migration*, Vol. 55 No. 6, pp. 137-155 (2017)

- [23] Pisarevskaya, A., Levy, N., Scholten, P., Jansen, J.: Mapping migration studies: An empirical analysis of the coming of age of a research field. In: *Migration Studies*, Volume 8, Issue 3, pp. 455–481 (2020)
- [24] Snyder, H.: Literature review as a research methodology: An overview and guidelines. In: *Journal of Business Research*, Elsevier, No. 104, pp. 333-339 (2019)
- [25] Tranfield, D., Denyer, D., Smart, P.: Towards a methodology for developing evidence in formed management knowledge by means of systematic review. In: *British Journal of Management*, Vol. 14 No. 3, pp. 207-222 (2003)
- [26] Vargas-Silva, C.: *Handbook of Research Methods in Migration*. Cheltenham: Edward Elgar (2012)
- [27] Waldorf, B.: Determinants of international return migration intentions. In: *The Professional Geographer*, Vol. 47 No. 2, pp. 125-136 (1995)
- [28] Webster, J., Watson, R.T.: Analysing the past to prepare for the future: writing a literature review. In: *MIS Quarterly*, Vol. 26 No. 2, pp. 13-2 (2002)