Academic Mobility from a Big Data Perspective

Laura Pollacci[1,3*], Letizia Milli[1,3†], Tuba Bircan[2†] and Giulio Rossetti[3†]

[1*]Department of Computer Science, University of Pisa, Largo Bruno Pontecorvo, 3, Pisa, PI, 56100, Italy.
[2]Department of Sociology, Vrije Universiteit of Brussel, Bd de la Plaine 2, Ixelles, 1050, Brussel, Belgium.
[3]Institute of Information Science and Technologies "A. Faedo" (ISTI), National Research Council (CNR), Via Giuseppe Moruzzi, 1, Pisa, PI, 56124, Italy.

*Corresponding author(s). E-mail(s): laura.pollacci@unipi.it;
Contributing authors: letizia.milli@unipi.it; tuba.bircan@vub.be; giulio.rossetti@isti.cnr.it;
[†]These authors contributed equally to this work.

Abstract

Understanding the careers and movements of highly skilled people plays an ever-increasing role in today's global knowledge-based economy. Researchers and academics are sources of innovation and development for governments and institutions. Our study uses scientific-related data to track careers evolution and Researchers' movements over time. To this end, we define the Yearly Degree of Collaborations Index, which measures the annual tendency of researchers to collaborate intra-nationally, and two scores to measure the mobility in and out of countries, as well as their balance.

Keywords: scientific mobility, network analysis, scientific networks

1 Introduction

Knowledge has become a valuable resource for exchange and international mobility plays a key role in scientific production, education, and policy-making and research careers of highly qualified personnel. Given the importance of highly skilled personnel, career analyses and pattern mobility models are increasingly attracting the attention of both institutions and researchers. As an intersection of two significant discourses (1) the internalisation in the global academia and (2) researchers as highly-skilled migrants, there exists a notable gap in the contemporary knowledge environment of migration and mobility of researchers, who are also named as "academics",

and "scientists". Despite the increasing global trends of highly-skilled migration and emergent interest in migration/mobility studies, migrant researchers have captured a limited interest (for exceptions see [8, 27, 28]). One of the challenges with demographic modelling highly skilled migration and movements is the significant gaps in international statistics considering definitions, and specific socio-economic indicators for migrants such as education levels [2] and a lacking a world migration survey [56]. To extend the knowledge gained by inferring the mobility and migratory patterns of researchers through traditional data such as register statistics, alternative data sources open the way for new perspectives.

The availability of massive data describing both publications and researchers' careers, together with its multifaceted nature, have made scientific mobility a fertile research ground for multiple fields of study [49]. Researchers have benefited from the advantages of alternative sources such as bibliometric repositories such as Scopus[1] and Web of Science[2] to study the academic collaboration networks and to develop scientific mobility indicators [16, 30, 58] and Microsoft Academic Graph [55] to examine the scientific ethnic and mobility networks and [3, 52]. Besides few exceptions (see [33, 44]), the international worldwide mobility patterns have not been fully explored. With our work, we aim to provide a global vision of scientific knowledge exchange and researchers' mobility at different temporal resolutions based on data from the Microsoft Academic Knowledge Graph (MAKG)[3].

The contribution of this paper is twofold:

(1) We investigate the collaborative environment of academia and scientific exchange by focusing our analyses on scientific collaborations observed through the proxy of article co-authorship. We will accordingly develop a measure, *Yearly Degree of Collaboration Index (YDCI)*, which captures the tendency for a scientist to collaborate with colleagues working in the same country or abroad on annual basis. This index enables identifying different (homogeneous) groups of scientists, which we describe based on spatial and temporal dimensions.

(2) We focus on the evolution of highly specialised academic mobility flows and propose a *mobility score* to describe academic outbound and inbound migrants on the country level. Based on this mobility score, the *mobility balance* index, which allows estimating the difference between inflows and outflows, will be derived.

The article is structured as follows: Section 2 draws the conceptual framework of our study by contextualising academic mobility and knowledge transfer in the existing literature and by discussing how our approach differs from previous efforts. In Section 3, we describe the data and our methodological approach including the data collection and pre-processing phase and the preliminary and necessary steps for our analyses, such as the calculation of the YDCI, mobility score and mobility balance. The core of our work is set out in Section 4, where we provide the description of our analytical approach and discuss of the observed outcomes. Finally, Section 5 summarises our conclusions and interpretations, with some suggestions for future developments.

2 Academic Mobility, Academic Networks and Knowledge Transfer

Analysing how, why and where highly-skilled individuals such as researchers move has attracted accelerating interest in recent decades due to the socio-political evolution, globalisation, and knowledge-based economic approaches around the globe. In the context of internationalisation of academia, "migration" and "mobility" have been used interchangeably [46], however, mobility of academic go beyond the commonly accepted migrant[4] approach which encompasses long-term change of residence by a cross-border (physical) mobility. Nevertheless, despite the fact that the gist of the interest in the highly-skilled migration is being mostly economic, internationalisation and mobility of researchers can be recognised as not only a physical mobility [51] but also a system for global knowledge transfer [12]. Having said that, international academic movements, flows, and networks are recognised as beneficial transnational and transferable identity capital that are antitheses to intellectual parochialism [35]. In short, internationalisation in academia covers not only the cross-border (both short-term and long-term) mobility of the researchers but also the cross-country collaborations which facilitate international knowledge transfer.

Mobile academics are conceptualised through the interplay of multiple movements where knowledge is used as power and mobility as a resource

---

[19, 43]. Since academic mobility and freedom of movement of knowledge are a global multidimensional phenomenon, studying academic mobility within the migration framework requires more complex data than the population registers that capture the official registration of residential movements.

Although in the debate on the internationalisation of higher education, academics' and researchers' mobility has been less investigated than student mobility [51], the literature includes different approaches based on traditional (e.g., official registers and census data) or *innovative* (e.g., big and social media data) sources. Regarding traditional data sources, UNESCO, OECD, and the European Union, through EUROSTAT, collect educational-related statistics. However, these data often do not include information about citizenship and mobility of academics and information cannot be fully comparable [51, 45]. Due to this, several surveys rise, such as the GlobSci Survey [59], the Changing Academic Profession (CAP) [60], MORE2 [61], albeit with notable differences in sample size and geographic coverage. Conversely, by exploiting innovative and big data, the research has focused on linking career evolution and international mobility [54], measuring knowledge transfer [4], analysing the convergence or discrepancy of countries in academic mobility and collaboration [13]. Moreover, scientific data have been exploited to study scale-free networks [7], temporal sequence analysis [6, 39], network statistical properties [37], measure international scientific collaboration [57, 36], and scientist mobility [31, 33, 14].

As pointed out in [51], a quantitative analysis of the mobility of academics has to cope with the wide range of terms used and definitions adopted. Due to the tendency to classify academic personnel into categories (e.g., scientists, qualified personnel, highly skilled workers, R&D personnel, and researchers), integrating and comparing data from different resources is often complex. Moreover, available data resources are very heterogeneous in terms of distribution, access, necessary skills, content, and size. Most of the literature exploits Scopus data [9, 31, 33, 23, 50] while some others use Web of Science data [13, 44], one of the most frequently used indexed database [38]. Moed et al. [31] analysed mobility between institutions in Germany, Italy, and the Netherlands. Leveraging bibliometric data from Scopus, the authors profile academics, e.g., distinguishing "young researchers", and analysed the accuracy of links between academics and institutions. Moreover, in [9], the authors emplyoed Scopus data analyse academic mobility by observing researchers of various fields and countries also considering career stages and gender. Also, Robinson-Garcia et al. [44] analysed individual publication records based on publications covered in the Web of Science, from 2008 to 2015, to distinguish between "academic migrants", i.e., authors who disengaged from their country of origin, and "academic travellers", i.e., authors who gain additional affiliations but maintain affiliation with their country of origin. Other well-known scientific data sources are the Microsoft Academic Knowledge Graph (MAKG) [1, 20] and its parent dataset, the Microsoft Academic Graph (MAG) [48], a heterogeneous graph about scientific publication records and actors involved in these, e.g., authors, institutions, and journals. In [24], an in-depth analysis is proposed to highlight the characteristics of the MAG and compare it with other publicly available research publication datasets. Effendy et al. examine trends in computing using citation counts [18] and rank conferences into ratings [17]. Finally, Panagopoulos et al. [40] focus on evaluating the impact of authors based on both collaborative networks and citations by research areas. Scientific disciplines and geographical coverage are other distinctive characteristics of academic mobility in the state-of-the-art. Some works focus on specific research areas such as bio-pharmaceuticals [11], molecular life sciences [25], and computer vision [21]. Moreover, studies can be limited in space, as in [25, 14, 30, 32] and focusing on specific regions rather than a holistic or global approach, as in [44].

This study presents a new approach to study knowledge transfer analysing scientific collaborations and the international mobility of researchers. Even in [9] is presented a study based on joint investigation of scholarly publications and movements of researchers over time. And, as in our study, authors' affiliation in publication is employed to track changes of affiliations over time. However, in [9], only publications indexed in Scopus and authors with a Scopus author ID are considered. On the contrary, here we jointly

employ Microsoft Academic Graph and Knowledge Graph. Although the Microsoft Academic website and underlying APIs were retired on Dec. 31, 2021[5], we believe that this dataset offers us a peerless source of information in terms of quantity (e.g., number of publications and authors) and geographical-temporal coverage (i.e., from 1800 to 2020 and about 180 countries). Further in [9], "mobility" is defined as "having a co-affiliation or multiple affiliations". In this study are proposed two new measures. The first, is a country level measure referring to scientific collaborations and exchange on a yearly basis, the *Yearly Degree of Collaborations' Internationality index* (YDCI). The second measure is a country level *mobility score* which allows to estimate annual inflows and outflows and balance for researchers mobility. The YDCI was also used to identify homogeneous groups of researchers who were analyzed with respect to geographical and temporal dimensions at different spatial resolutions (country, european, and worldwide level). As regards the mobility score, we define two versions, *In* and *Out*, based on the flows of researchers entering and leaving the countries, respectively. Finally, on the basis of these, we calculate the *mobility balance* to estimate the difference between the two flows providing a comprehensive worldwide overview.

3 Data and Methodology

The aim of this study is internationalisation and knowledge transfer through, firstly collaboration, and, secondly mobility of researchers. The data source and the method to achieve these goals are elaborated below.

3.1 Dataset

Our study is based on bibliometric data from Microsoft Academic Knowledge Graph[6] (MAKG). The dataset composes of several scientific collaboration-related data, split into 18 subsets. From these, we focus on:

- *Authors*: information about researchers, such as name and affiliation (about 254 million entities).

- Affiliations: information on scientific institutions, e.g., research centers, academies, hospitals, etc., including name and Wikipedia url (about 25 thousand entities).
- *FieldOfStudy*: information on the fields of study associated with the papers (about 230 thousand entities).
- *Papers*: information on publications, including the year of publication and authors (210 million entities). Publications belong to five categories: *JournalArticle* (about 82 million), *PatentDocument* (about 51 million), ConferencePaper (about 4 million), *BookChapter* (about 2 million), and Book (about 2 million).

The MAKG covers 180 countries worldwide and includes publications from 1800 to 2020. We restrict our analysis to those papers published from 1980 to 2019. Moreover, we focused only on "active" authors (according to [31]), filtering out those without publications yearly. Doing so, we obtained a dataset composed of 9 million authors - having at least a specified affiliation during their research activity - and all their papers.

3.2 Methodological Approach

To observe the researcher exchange between countries and internationalisation, it is necessary to geolocate the institutions to which the authors refer. Since this information is unavailable from the MAKG data, it was necessary to resort to the latest version of MAG[7], in which each publication is modeled as a triple $< paper, author, institution >$. The obtained dataset is pre-processed following a semi-supervised Natural Language Processing (NLP) pipeline (which leverages Wptools[8] and Pycountry[9] python libraries) allows to geolocate affiliations with respect to countries. Then, the authors' annual *ego networks* are computed as their scientific collaborators' undirected graph. In brief, ego-centric networks, also called *ego* networks, consist of a central node, the "ego", the nodes to whom ego is directly connected to, which are called "alters", and the tiles among the alters, if any. Thus, in this study, we build a network

---

where each author in turn acts as the ego and the annual co-authors are the alters. Here no ties among the alters are taken into account. The pre-processed data allows us to compute the Yearly Degree of Collaborations' Internationality, a measure to describe the tendency of researchers to collaborate with colleagues working in the same country or not as follows:

1. The ego network of each researcher is extracted for each year in the dataset.
2. From each ego network is extracted the list of countries of affiliation of the co-authors.
3. The co-author country lists are converted into distributions by chance. For instance, given the co-author country list *[Italy, Italy, Germany]* where each country has a distribution of $1/3$, we obtain as probability distribution list $[0.67, 0.33])$.
4. For each ego, thus the researcher acting as the center of the ego networks is calculated the binary entropy of probability distributions, and the result is multiplied by -1 in case the majority of countries of co-authors within the ego network is different from the country of affiliation of the ego node.

Thus, the Yearly Degree of Collaborations' Internationality is calculated as the binary entropy of the probability distribution obtained starting from the list of countries of affiliation of the nodes present within each ego network, ego excluded. In other words, the YDCI is the binary entropy of each researcher's colleagues' probability distribution of working countries annually.

By defining k as a set of co-authors of an ego, $\#c_{dif}$ the number of countries different from that of the network ego, and as $\#c_{same}$ the number of countries equal to that of the network ego we have, the YDCI is expressed as:

$$\text{YDCI} = \begin{cases} (-\sum_k P_k \log_2 P_k) * (-1) \\ \text{if } \#c_{dif} > \#c_{same} \\ \\ -\sum_k P_k \log_2 P_k \\ \text{otherwise} \end{cases} \quad (1)$$

where P_k is the probability distribution of co-authors' countries in a given year, which is derived by extracting from egos the lists of affiliations countries of each co-author. YDCI ranges from -1 to 1, where a YDCI closer to -1 represents the researcher's tendency to collaborate with geographically heterogeneous groups composed of researchers from countries different from their own. Conversely, a YDCI closer to 1 represents the tendency to collaborate with geographically homogeneous groups composed of researchers from their own country.

Thus the YDCI measures the researcher's annual tendency to collaborate with colleagues working in their own country and establish intra- and international scientific collaborations. Furthermore, by aggregating the authors following different criteria, the YDCI allows studying trends at different geographic (e.g., national, continental, and world level) and temporal (e.g., globally and for decades) scales. The YDCI is employed to cluster and describe researchers based on their collaboration types, i.e., inter- vs. intra- national, with respect to temporal and geographical dimensions. To this end, authors are represented as vectors by using their YDCI values in time. We identify with $\mathbf{X}_{m,n}$ the matrix, where the mth row corresponds to an author, and the nth columns represent a year in the range $[1980, 2019]$. Therefore, the value in cell (m, n) is the YDCI of author m at time n. In case of missing values, we complete the trends considering the average of the values of the column, e.g., the global average YDCI of the given year. We use GridSearch[10] to find the best k and optimise the silhouette to apply the K-Means clustering algorithm. Further, clusters of researchers based on YDCI are computed independently over four decades to observe their stability temporally.

As a second investigation, we measure the worldwide knowledge transfer focusing on researchers' movements over affiliations from a geographical and temporal point of view. Given a country C and a year Y, the *incoming mobility score* $(In(C))$ defines the countries' degree of mobility based on yearly incoming researchers, where a *yearly incoming researcher* is a researcher who published in a year previous than Y in country $C_X \neq C$ and in year Y in country C. Similarly, the *outgoing mobility score* $(Out(C))$ defines the countries' degree of mobility based on

---

[10]We perform GridSearch on the 80% of the dataset, with $k \in [2; 10]$.

yearly outgoing researchers where a *yearly outgoing researcher* is a researcher who published in a year previous than Y in country C, and in year Y in country $C_X \neq C$. Further, the *mobility balance* estimates whether a country has more incoming or outgoing traffic of authors. To calculate these scores, first, we build two matrices representing the incoming $(\mathbf{X}_{In})$ and outgoing $(\mathbf{X}_{Out})$ researchers for each country annually. Each matrix has as many rows as countries and has many columns as the years in the time window (1980-2019). By construction, each matrix column represents an annual worldwide count of movements (outgoing or incoming accordingly to the matrix). To prevent what in terms of probability distributions are called "outliers", i.e., a few high values vs. a high number of low values, these are converted in the $[0, 1]$ range by using the *quantile transformation* (Formula 2).

$$G^{-1}(F(X)) \qquad (2)$$

where, F is the cumulative distribution function of features, i.e., values of X, and G^{-1} is the quantile function of the desired distribution in output, i.e., G.

Given a distribution probability, i.e., values in a generic column x of $\mathbf{X}_{In}$ and $\mathbf{X}_{Out}$ matrices, its cumulative function represents the probability that a random variable X takes a value less than or equal to $\varkappa$. This can be expressed as Formula 3:

$$F_X(\varkappa) = P(X \leq \varkappa) \qquad (3)$$

The quantile function returns a threshold $\varkappa$ below which a random extraction from the probability distribution, i.e., cumulative distribution (Formula 3), will fall most of times, as expressed in Formula 4.

$$G(\rho) = \inf\{\varkappa \in \mathbb{R} : \rho \leq F(\varkappa)\} \qquad (4)$$

Formula 4 uses the following principles: a) if X is a random variable with cumulative distribution F, then $F(X)$ is uniformly distributed in $[0, 1]$, and b) if U is a random variable uniformly distributed in $[0, 1]$, then $G^{-1}(U)$ has G as distribution.

The *incoming* and *outgoing mobility scores* are calculated by applying Formula 4 to probability distributions of countries. Then, given a country C, the mobility balance is computed as the difference between the incoming and the outgoing

mobility scores. The defined mobility scores are studied based on different temporal resolutions to observe changes in trends over time.

4 Analysis

The method proposed in Section 3 has been applied to bibliometric data from Microsoft Academic Knowledge Graph (Section 3.1).

As shown in Figure 1, three well-separated clusters emerge by applying the K-Means[11] to the YDCIs.

- *Cluster0* includes the 89.8% of the dataset $(8,008,741$ authors) and is composed of authors

---

[11]Best GridSearch performance average silhouette 0.54.
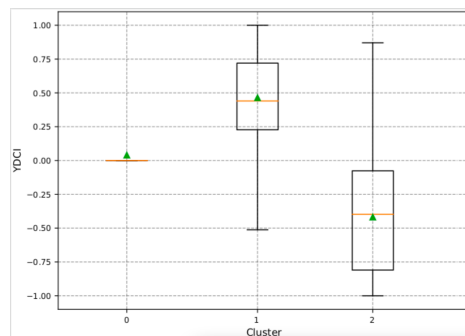

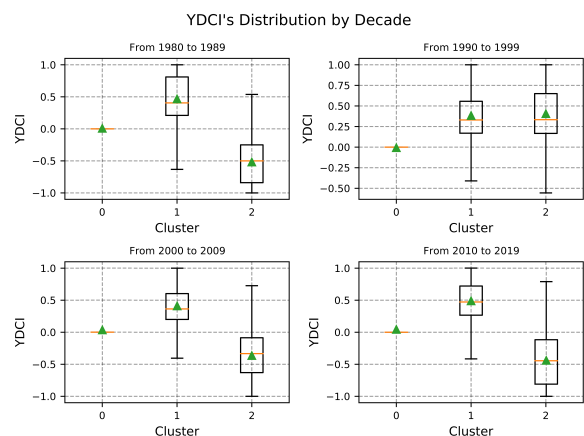**Fig. 1**: Clustering of authors according to the YDCI.

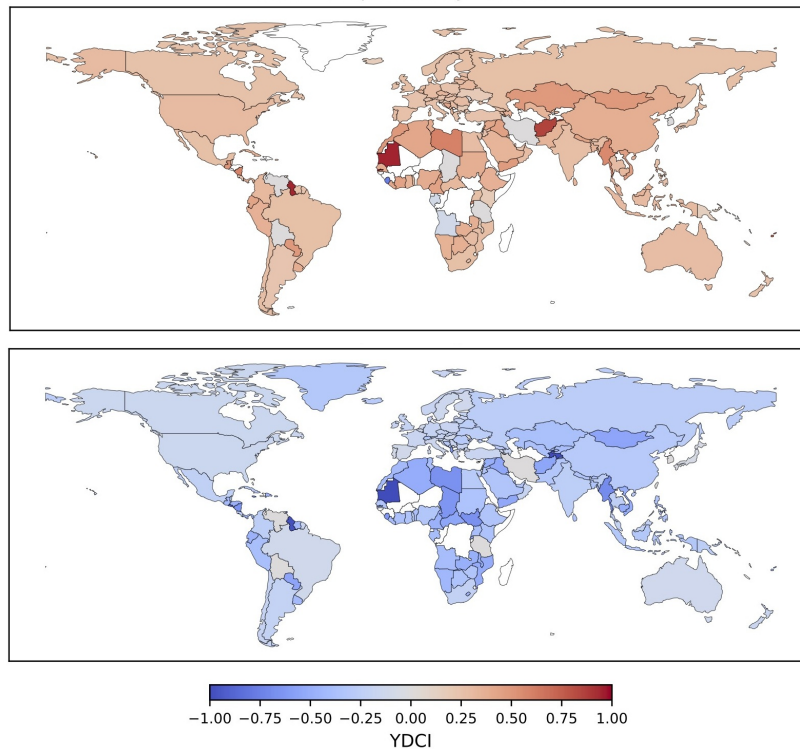**Fig. 2**: Clustering of authors according to the YDCI over decades.

Fig. 3: (top) Average YDCI for *Cluster1* and (bottom) *Cluster2* countries.

who tend to work alone or establish collaborations only with a few researchers from the same country.

- *Cluster1* is composed of the 7.09% of the dataset (632, 679 authors). This cluster groups together those researchers that prevalently collaborate with geographically homogeneous groups composed of researchers from the same countries.
- *Cluster2* represents the 3.11% of the dataset (277, 324 authors). It is the opposite of *Cluster1* and identifies authors who tend to collaborate with geographically heterogeneous groups composed of researchers from countries different from theirs.

To observe the stability of the identified clusters over time, we replicate the clustering over four decades, i.e., from 1980-1989, 1990-1999, 2000-2009, and from 2010 to 2019, as shown in Figure 2. The three-clustered structure still emerges in each decade, and the overall behaviour of the groups remains unchanged. However, data suggests that

local contributions tend to increase in the second decade (1990-1999), as testified by the change values of *Cluster2*.

Although *Cluster0* is the cluster that includes the majority of researchers, the latter shows a tendency to work alone. In contrast, although less populated, *Cluster1* and *Cluster2* include researchers who generally collaborate with other colleagues. Given that our research focuses on knowledge transfer, we deeper investigate the dynamics of collaborations of researchers in *Cluster1* and *Cluster2*.

In order to observe YDCI distribution globally, for the two selected clusters, we calculate the average of the annual scores of the authors of each country. Figure 3 shows the obtained maps for *Cluster1* (top) and *Cluster2* (bottom), respectively[12]. By considering both maps, it can be seen that trends in collaborations are generically geographically homogeneously distributed. Focusing on *Cluster1* (top)- which includes authors who

---

[12]In both maps, the countries in white do not have an assigned score due to missing data.
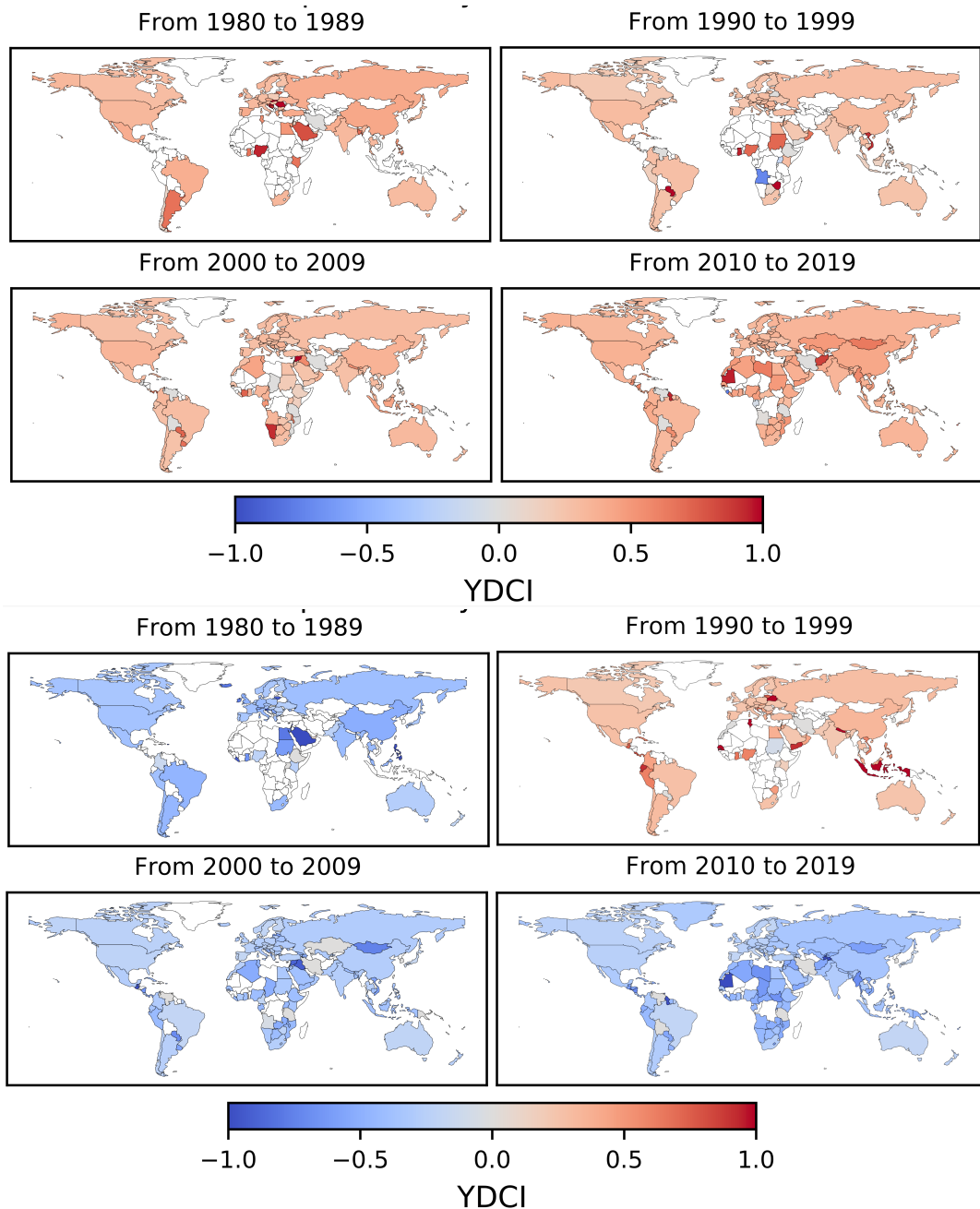
Fig. 4: (top) Average YDCI for *Cluster1* and (bottom) for *Cluster2* countries over decades.

tend to collaborate within their own country- it emerges, at least on a global scale, that American countries (both North and South), European and Australian ones obtain YDCI values of [0.25; 0.50], with non-extreme trends. On the contrary, the Asian and African continents show more heterogeneous values of collaboration

intra-nation. Moving to *Cluster2* (bottom),- which includes authors who tend to collaborate with colleagues located abroad -American countries (especially in the north), Europe, and Australia have YDCI values in the range [−0.25; −0.50],
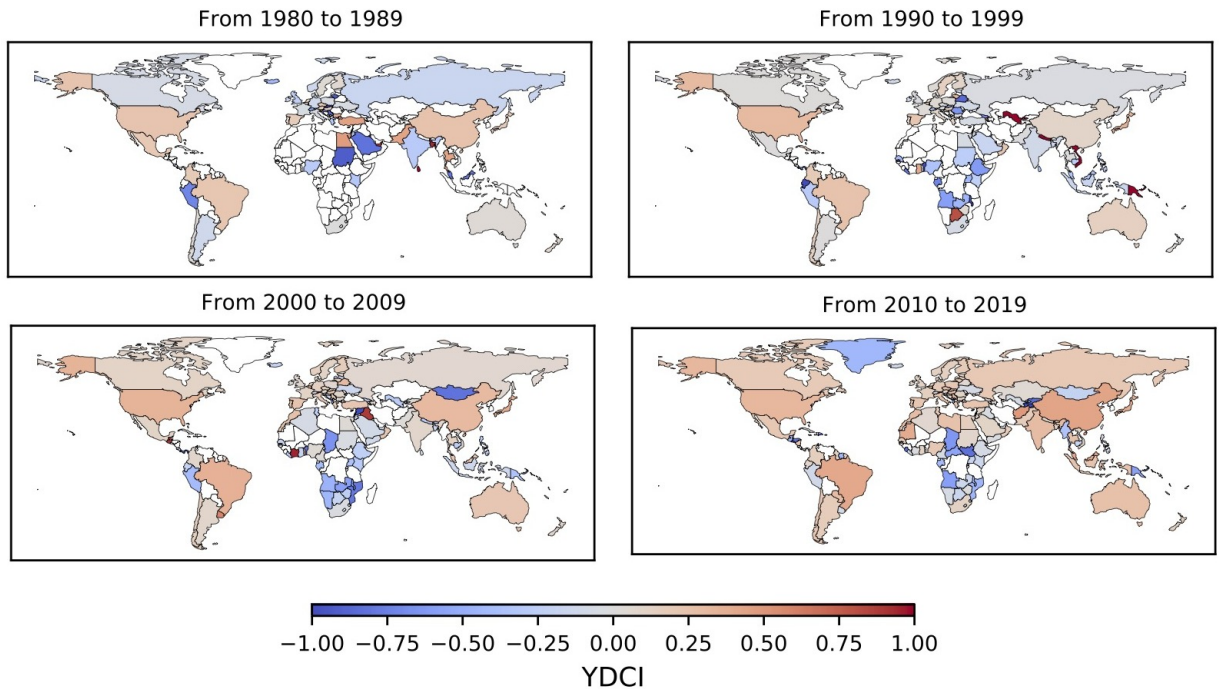
Fig. 5: Evolution of YDCI values over decades.

which mean some degree of international collaborations, as *Cluster1*. At the same time, in particular African countries, and some from Asia, show the lowest YDCI values, between $[-0.50; -1.00]$. As noticed, most countries with the most extreme tendencies are the same, e.g., Mauritania, Guyana, and Libya. Nevertheless, this behaviour could be due in part to the fact that these are most likely countries with a relatively small share of academic publications and staff in the MAKG data.

In addition, we analyse the decade-wise YDCI segmentation aggregated at the country level. Considering *Cluster1* on top in Figure 4, we note a constant and diffuse closure towards foreign collaborations over the decades. This trend becomes more evident globally in the fourth decade (2010-2019), especially in Asia. When looking at *Cluster2* in Figure 4 (bottom), we note a diffuse tendency to collaborate with colleagues in countries other than their own. In the following decade, from 1990 to 1999, a trend reversal is observed, with a large diffusion of collaborations between researchers in the same country. However, from 2000 onwards, the YDCI values settled again on

negative values around -0.5 showing growing collaborations at an international level. Our hypothesis is that, on the one hand, globalization, ease of travel and growing agreements between institutions have contributed to a greater circulation of researchers around the world. On the other hand, it is possible that part of the international collaborations are finalised or at least lead to a transfer - more or less long - to a new institution. Thus, our suggestion is that co-authoring one or more articles could then act as an initial point of contact leading to a temporal or permanent period of direct collaboration – and thus a move.

To comprehensively observe the evolution and general trend in collaborations, we aggregate authors of the two observed clusters and calculate the averages of the YDCI scores again by country and decade. As shown by Figure 5, we are witnessing a general trend towards intra-national collaboration over time. Observing the individual *choropleth maps* in more detail, it is noted-above all in the first decade (1980-1989)- two well-defined groups of countries seem to coexist. On the one hand, the United States of America, Mexico, Brazil and China show a greater tendency towards intra-national collaborations. On the other hand,
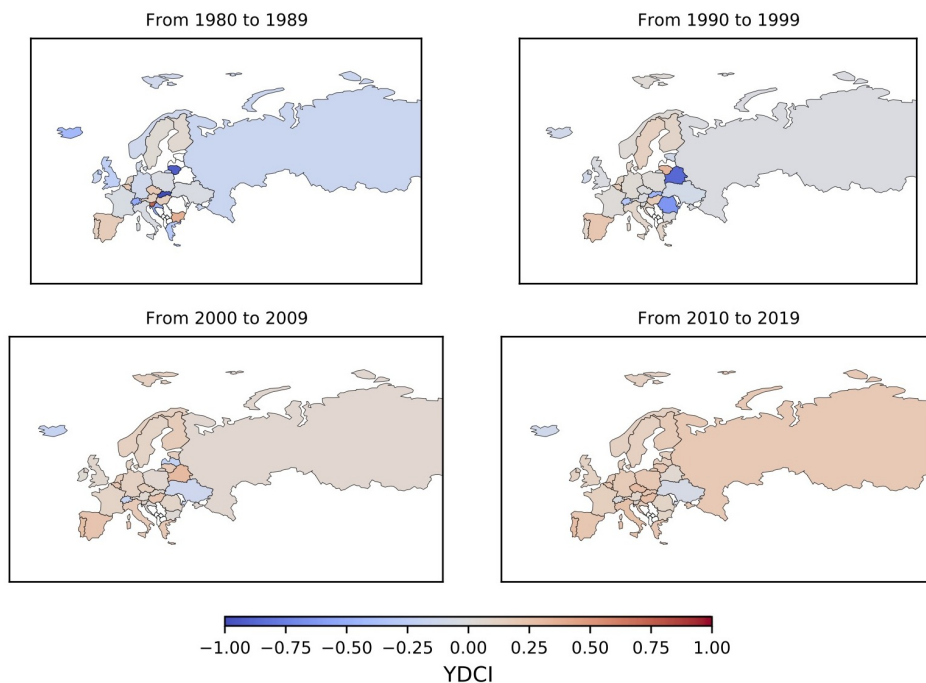
Fig. 6: YDCI for European countries over decades.

Canada, Peru, Argentina, Russia, India and Australia show a -albeit slight- certain tendency to collaborate with abroad, with YDCI values in the range [-0.25,-0.50]. Already from the second decade, and even more accentuated in the third, the negative YDCI values tend to reverse their trend. Finally, in the fourth decade, most countries show positive YDCI values and a greater tendency towards intra-national collaboration, albeit with some small exceptions. Again, these exceptions refer to countries that may be underrepresented in MAG and MAKG. Focusing on European countries shown in Figure 6, we observe the same general trend observed worldwide (Figure 5). Europe shows more heterogeneous YDCI distribution during the first and second decades (1980-1989 and 1990-1999), with a prevalence of countries characterized by collaborations at the international level, with YDCI values around −0.25. However, there are some exceptions, e.g., Spain, Portugal, Sweden, and Finland, which show values slightly positive. As already observed worldwide, starting from the second decade onwards, there is an evident and progressive growing trend towards intra-national collaboration throughout Europe,

although with YDCI values around 0.5. Internationalization noted at the European level mirrors what is observed at the global level. A further explanation for this behavior could derive from the diffusion of research centers and institutions. In fact, we believe that part of the shifts observed in the first decades may be due to the need for researchers to physically reach institutions due to a) lack of a team/institution/technologies related to their area of study in their own country; b) difficulties in remote communication and collaboration.

Moving to study worldwide knowledge transfer based on researchers' movements over affiliations, we calculate countries' incoming mobility score, outgoing mobility score and mobility balance (Section 3). The map in Figure 7 shows the distribution of the incoming mobility score by decades. We observe that the United States maintains constant high incoming mobility over time, acting as a particularly attractive country. China, like Russia, on the other hand, shows medium-high incoming mobility during the first decade, which then tends to increase over time. However, in the Asian continent, there are countries with low and medium-low incoming mobility, i.e., Mongolia, Afghanistan,
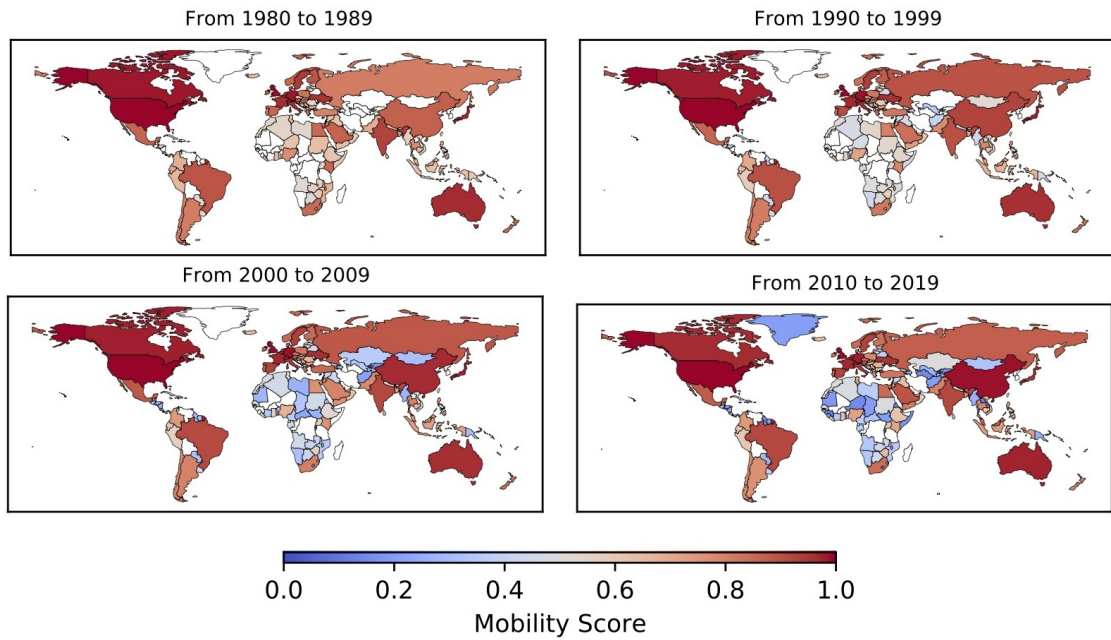
From 1980 to 1989   From 1990 to 1999

From 2000 to 2009   From 2010 to 2019

Fig. 7: Incoming mobility over decades.

From 1980 to 1989   From 1990 to 1999

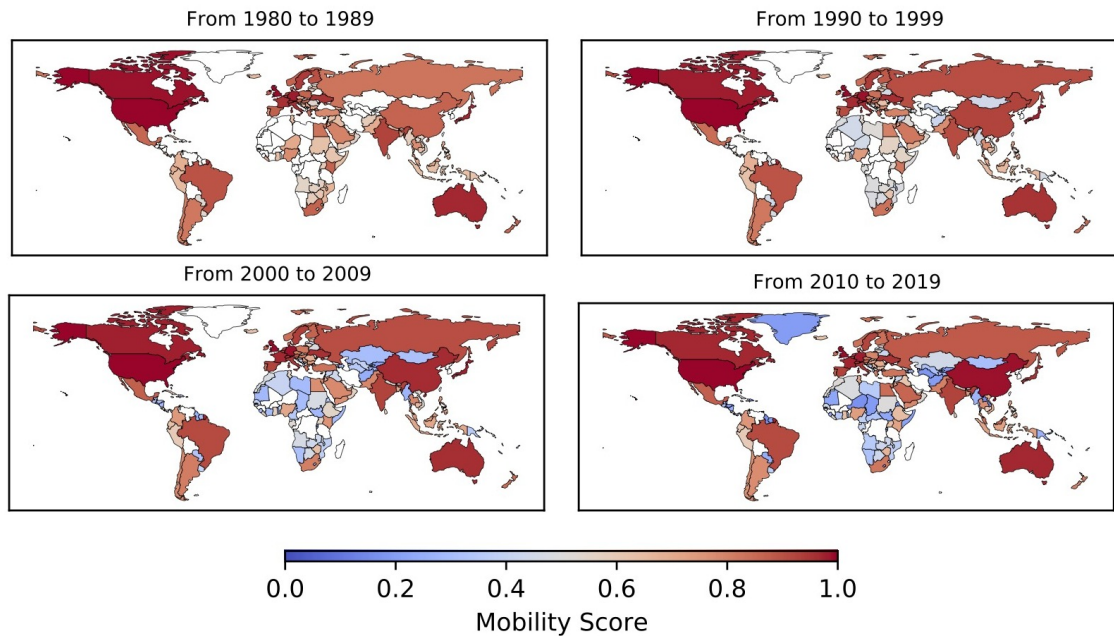From 2000 to 2009   From 2010 to 2019

Fig. 8: Outgoing mobility over decades.

Burma and Turkmenistan. Although we did not have full coverage of information, Africa shows very low inflows over time, except a few countries, Egypt and South Africa.

The distribution of the outgoing mobility score for each decade is shown in Figure 8. As before,
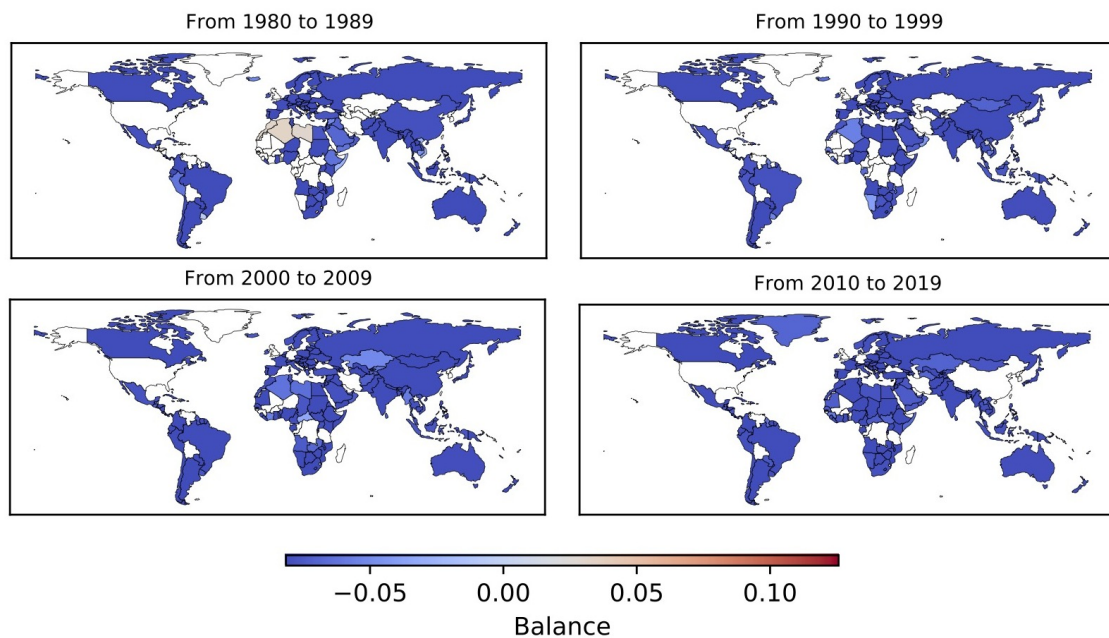
From 1980 to 1989 · From 1990 to 1999 · From 2000 to 2009 · From 2010 to 2019

Fig. 9: Mobility balance.

while the United States shows a stable medium-high outgoing mobility over time, China and Russia show a lower outgoing mobility which increases over time. Within the Asian continent, countries with low and medium-low outgoing mobility are generally the same as those with low and medium-low incoming mobility. Africa, for which we have very spurious data in the first decade, initially shows slight outward mobility, depending on the country. The scenario becomes increasingly heterogeneous from the second decade, with countries showing medium-low mobility, i.e., Mauritania, Niger and Chad, and others with medium-high mobility, i.e., South Africa, Egypt. Finally, Figure 9 shows the map relating to the mobility balance over decades. Almost all countries in our dataset show values in $[-0.10; 0]$, which means that outbound mobility tends to prevail over inbound mobility. Going into the details of the decades, we note that in the first, only a few countries of Africa and Central America, i.e., Algeria, Libya, Morocco and Honduras, have incoming mobility slightly higher than outgoing. Over the decades, this trend reverses and aligns itself with the world trend, with outgoing flows of researchers greater than incoming ones.

5 Discussion and Conclusions

This paper presents a new approach to study knowledge transfer through collaborations and the international mobility of researchers by developing two new measures: (1) a country level *Yearly Degree of Collaborations' Internationality index* to understand the collaborative environment of academia and scientific exchange on a yearly basis, (2) a *mobility score* to estimate annual inflows and outflows differentials for academic mobility on the country level.

Accordingly, we first define the YDCI index which measures the degree of inter-nationality of researchers' collaborations around the globe annually. The YDCI allows us particularly to identify three separate groups of researchers using K-Means. The clusters found are deeply studied and described with respect to geographical, temporal and spatial dimensions and at different resolutions.

Secondly, we focus on the movements of researchers over affiliations worldwide over time by defining two mobility scores (In and Out) to describe countries based on incoming and outgoing researchers. As a next step, we use these to compute the mobility balance, which estimates the

13

difference between incoming and outgoing flows providing a comprehensive worldwide perspective.

Our findings indicate an ever-increasing trend of collaborating with geographically homogeneous groups composed of researchers from their own countries, especially Europe and North and South America. However, international collaboration seems to prevail in the African continent, Central America, and some countries in Asia. On the other hand, researchers move more often and in a homogeneous way concerning both continents and individual countries.

A possible interpretation of results could be that the networks of researchers are steady at a certain degree that their mobility patterns are consistent for reaching particular research groups or institutions with which to collaborate in certain geographies. With this study, we illustrated two new measures to investigate academic mobility and knowledge exchange. Given the temporal dimension in these measures, as future work, the impact of contextual factors could be examined to develop a better understanding of the mobility patterns and changes in time. For instance, countries' YDCI trends and movements of researchers can be compared with socio-cultural events, e.g., Ĉhernobyl' disaster (1986), the fall of the Berlin Wall (1989), the Dissolution of the Soviet Union (1991), and the collapse of the Twin Towers (2001), or Ukranian war (2022) to study the influence of global poignant events on academic mobility. Moreover, information from authors' collaborative networks can help identify and describe professional and geographical patterns in researchers' careers.

Conflict of interest. On behalf of all authors, the corresponding author states that there is no conflict of interest.

Author contributions. All authors contributed equally to this work.

Research involving Human Participants and/or Animals. This research does not involve Human Participants and / or Animals.

Informed consent. This research does not require informed consent.

References

[1] Färber, M. The Microsoft Academic Knowledge Graph: A Linked Data Source with 8 Billion Triples of Scholarly Data. *Proceedings Of The 18th International Semantic Web Conference.* pp. 113-129 (2019), https://doi.org/10.1007/978-3-030-30796-7%5C_8

[2] Ahmad Wali Ahmad-Yar and Tuba Bircan. Anatomy of a misfit: International migration statistics. *Sustainability*, 13(7):4032, 2021.

[3] Bedoor K AlShebli, Talal Rahwan, and Wei Lee Woon. The preeminence of ethnic diversity in scientific collaboration. *Nature communications*, 9(1):1–10, 2018.

[4] Valeria Aman. A new bibliometric approach to measure knowledge transfer of internationally mobile scientists. *Scientometrics*, 117(1):227–247, 2018.

[5] S. Appelt, B. van Beuzekom, F. Galindo-Rueda, and R. de Pinho. Which factors influence the international mobility of research scientists? *Global mobility of research scientists* 177–213, 2015.

[6] A BARABAaSI, H Jeong, Z Néda, et al. Evolution of the social network of scientific collaborations. *Physica A: Sta-tistical Mechanics and its Applications*, 311(3):590–614, 2002.

[7] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: statistical mechanics and its applications*, 281(1-4):69–77, 2000.

[8] B Bönisch-Brednich. Rituals of encounter: campus life, liminality and being the familiar stranger. *Crossing boundaries and weaving intercultural work, life, and scholarship in globalizing universities*, pages 118–130, 2016.

[9] Momeni, F., Karimi, F., Mayr, P., Peters, I. & Dietze, S. The many facets of academic

mobility and its impact on scholars' career. *Journal Of Informetrics*. **16**, 101280 (2022)

[10] Judy F Burnham. Scopus database: a review. *Biomedical digital libraries*, 3(1):1–8, 2006.

[11] Clara Calero, Thed van Leeuwen, and Robert Tijssen. Research cooperation within the bio-pharmaceutical industry: Network analyses of co-publications within and between firms. *Scientometrics*, 71(1):87–99, 2007.

[12] Carolina Cañibano. Scientific mobility and economic assumptions: From the allocation of scientists to the socioeconomics of network transformation. *Science as Culture*, 26(4):505–519, 2017.

[13] Zaida Chinchilla-Rodríguez, Lili Miao, Dakota Murray, Nicolás Robinson-García, Rodrigo Costas, and Cassidy R Sugimoto. A global comparison of scientific mobility and collaboration according to national scientific capacities. *Frontiers in research metrics and analytics*, 3:17, 2018.

[14] Sonia Conchi and Carolin Michels. Scientific mobility: An analysis of germany, austria, france and great britain. Technical report, Fraunhofer ISI Discussion Papers-Innovation Systems and Policy Analysis, 2014.

[15] Arnab, S., Zhihong, S., Yang, S., Hao, M., Darrin, E., Paul, H. & Kuansan, W. An overview of microsoft academic service (MAS) and applications. *Proceedings Of The 24th International Conference On World Wide Web (WWW 2015 Companion)*. pp. 243-246 (2015)

[16] Marco Conti. Dynamics of scientific collaboration networks due to academic migrations. In *Social Informatics: 12th International Conference, SocInfo 2020, Pisa, Italy, October 6-9, 2020, Proceedings*, volume 12467, page 283. Springer Nature, 2020.

[17] Suhendry Effendy and Roland HC Yap. Investigations on rating computer sciences conferences: An experiment with the microsoft academic graph dataset. In *Proceedings of the 25th international conference companion on world wide web*, pages 425–430, 2016.

[18] Suhendry Effendy and Roland HC Yap. Analysing trends in computer science research: A preliminary study using the microsoft academic graph. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 1245–1250, 2017.

[19] Johannah Fahey and Jane Kenway. International academic mobility: Problematic and possible paradigms. *Discourse: Studies in the cultural politics of education*, 31(5):563–575, 2010.

[20] Michael Färber. The microsoft academic knowledge graph: A linked data source with 8 billion triples of scholarly data. In *International Semantic Web Conference*, pages 113–129. Springer, 2019.

[21] Takao Furukawa, Nobuyuki Shirakawa, and Kumi Okuwada. Quantitative analysis of collaborative and mobility networks. *Scientometrics*, 87(3):451–466, 2011.

[22] Siekierski, P., Lima, M. & Borini, F. International mobility of academics: Brain drain and brain gain. *European Management Review*. **15**, 329-339 (2018)

[23] Saeed-Ul Hassan, Anna Visvizi, and Hajra Waheed. The 'who'and the 'what'in international migration research: data-driven analysis of scopus-indexed scientific literature. *Behaviour & Information Technology*, 38(9):924–939, 2019.

[24] Drahomira Herrmannova and Petr Knoth. An analysis of the microsoft academic graph. *D-lib Magazine*, 22(9/10), 2016.

[25] Koen Jonkers and Robert Tijssen. Chinese researchers returning home: Impacts of international mobility on research collaboration and scientific productivity. *Scientometrics*, 77(2):309–333, 2008.

[26] Samiya Khan, Xiufeng Liu, Kashish A Shakil, and Mansaf Alam. A survey on scholarly data: From big data perspective. *Information Processing & Management*, 53(4):923–944, 2017.

[27] Terri Kim. Academic mobility, transnational identity capital, and stratification under conditions of academic capitalism. *Higher Education*, 73(6):981–997, 2017.

[28] Sin Yee Koh and I Lin Sin. Academic and teacher expatriates: Mobilities, positionalities, and subjectivities. *Geography Compass*, 14(5):e12487, 2020.

[29] Jean-Baptiste Meyer. Network approach versus brain drain: lessons from the diaspora. *International migration*, 39(5):91–110, 2001.

[30] Andrea Miranda-González, Samin Aref, Tom Theile, and Emilio Zagheni. Scholarly migration within mexico: analyzing internal migration among researchers using scopus longitudinal bibliometric data. *EPJ Data Science*, 9(1):34, 2020.

[31] Henk F. Moed, M'hamed Aisati, and Andrew Plume. Studying scientific migration in Scopus. *Scientometrics*, 94(3):929–942, jul 2013.

[32] Henk F Moed, M'hamed Aisati, and Andrew Plume. Studying scientific migration in scopus. *Scientometrics*, 94(3):929–942, 2013.

[33] Henk F Moed and Gali Halevi. A bibliometric approach to tracking international scientific migration. *Scientometrics*, 101(3):1987–2001, 2014.

[34] Izabela Moise, Edward Gaere, Ruben Merz, Stefan Koch, and Evangelos Pournaras. Tracking Language Mobility in the Twitter Landscape. In *IEEE International Conference on Data Mining Workshops, ICDMW*, volume 0, pages 663–670. IEEE Computer Society, jul 2016.

[35] Louise Morley, Nafsika Alexiadou, Stela Garaz, José González-Monteagudo, and Marius Taba. Internationalisation and migrant academics: the hidden narratives of mobility. *Higher Education*, 76(3):537–554, 2018.

[36] Francis Narin, Kimberly Stevens, and Edith S Whitlow. Scientific co-operation in europe and the citation of multinationally authored papers. *Scientometrics*, 21(3):313–323, 1991.

[37] M. E.J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2):404–409, jan 2001.

[38] Michael Norris and Charles Oppenheim. Comparing alternatives to the web of science for coverage of the social sciences' literature. *Journal of Informetrics*, 1(2):161–169, 2007.

[39] Gergely Palla, Albert-László Barabási, and Tamás Vicsek. Quantifying social group evolution. *Nature*, 446(7136):664–667, 2007.

[40] George Panagopoulos, Christos Xypolopoulos, Konstantinos Skianis, Christos Giatsidis, Jie Tang, and Michalis Vazirgiannis. Scientometrics for success and influence in the microsoft academic graph. In *International Conference on Complex Networks and Their Applications*, pages 1007–1017. Springer, 2019.

[41] Bartosz Paszcza. *Comparison of Microsoft academic (graph) with web of science, scopus and google scholar.* PhD thesis, University of Southampton, 2016.

[42] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani. Activity driven modeling of time varying networks. *Scientific Reports*, 2(1):1–7, jun 2012.

[43] Susan L Robertson. Critical response to special section: International academic mobility. *Discourse: Studies in the cultural politics of education*, 31(5):641–647, 2010.

[44] Nicolás Robinson-Garcia, Cassidy R Sugimoto, Dakota Murray, Alfredo Yegros-Yegros, Vincent Larivière, and Rodrigo Costas. The many faces of mobility: Using bibliometric data to measure the movement of scientists. *Journal of Informetrics*, 13(1):50–63, 2019.

[45] Schiller, D. & Cordes, A. Measuring researcher mobility: A comparison of different datasets and methods with an empirical application of micro-data for the United States and Germay. *OECD Blue Sky Forum.* (2016)

[46] M Rostan and E Hohle. The international mobility of faculty, the internationalisation of the academy: Changes, realities and prospects. *The changing academy–The changing academic profession in international comparative perspective*, 10, 2014.

[47] Roberta Sinatra, Dashun Wang, Pierre Deville, Chaoming Song, and Albert László Barabási. Quantifying the evolution of individual scientific impact. *Science*, 354(6312), nov 2016.

[48] Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-June Hsu, and Kuansan Wang. An overview of microsoft academic service (mas) and applications. In *Proceedings of the 24th international conference on world wide web*, pages 243–246, 2015.

[49] Alina Sîrbu, Gennady Andrienko, Natalia Andrienko, Chiara Boldrini, Marco Conti, Fosca Giannotti, Riccardo Guidotti, Simone Bertoli, Jisu Kim, Cristina Ioana Muntean, Luca Pappalardo, Andrea Passarella, Dino Pedreschi, Laura Pollacci, Francesca Pratesi, and Rajesh Sharma. Human migration: the big data perspective. *International Journal of Data Science and Analytics*, pages 1–20,

mar 2020.

[50] Alexander Subbotin and Samin Aref. Brain drain and brain gain in russia: Analyzing international migration of researchers by discipline using scopus bibliometric data 1996-2020. *arXiv preprint arXiv:2008.03129*, 2020.

[51] Ulrich Teichler. Academic mobility and migration: What we know and what we do not know. *European Review*, 23(S1):S6–S37, 2015.

[52] Giacomo Vaccario, Luca Verginer, and Frank Schweitzer. The mobility network of scientists: analyzing temporal correlations in scientific careers. *Applied Network Science*, 5(1):1–14, 2020.

[53] Richard Van Noorden. Global mobility: Science on the move. *Nature*, 490(7420):326–329, oct 2012.

[54] Caroline S Wagner and Koen Jonkers. Open countries have strong science. *Nature News*, 550(7674):32, 2017.

[55] Kuansan Wang, Zhihong Shen, Chiyuan Huang, Chieh-Han Wu, Yuxiao Dong, and Anshul Kanakia. Microsoft academic graph: When experts are not enough. *Quantitative Science Studies*, 1(1):396–413, 2020.

[56] Frans Willekens, Douglas Massey, James Raymer, and Cris Beauchemin. International migration under the microscope. *Science*, 352(6288):897–899, 2016.

[57] Fuyuki Yoshikane and Kyo Kageura. Comparative analysis of coauthorship networks of different domains: The growth and change of networks. *Scientometrics*, 60(3):435–446, 2004.

[58] Zhenyue Zhao, Yi Bu, Lele Kang, Chao Min, Yiyang Bian, Li Tang, and Jiang Li. An investigation of the relationship between scientists' mobility to/from china and their research performance. *Journal of Informetrics*, 14(2):101037, 2020.

[59] Franzoni, C., Scellato, G. & Stephan, P. Foreign born scientists: Mobility patterns for sixteen countries. (National Bureau of Economic Research,2012)

[60] Teichler, U., Arimoto, A. & Cummings, W. The changing academic profession. *Dordrecht Etc: Springer*. (2013)

[61] Consult, I. & Others Support for continued data collection and analysis concerning mobility patterns and career paths of researchers. *Final Report MORE*. **2** (2013)