Academic Mobility from a Big Data Perspective

Laura Pollacci (laura.pollacci@di.unipi.it)
University of Pisa
Letizia Milli
University of Pisa
Tuba Bircan
Vrije Universiteit Brussel
Giulio Rossetti
National Research Council

Research Article

Keywords: scientific mobility, network analysis, scientific networks

Posted Date: April 4th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1510153/v1

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]ISTI, CNR - National Research Council, Via Giuseppe Moruzzi, 1, Pisa, PI, 56124, Italy.

*Corresponding author(s). E-mail(s): laura.pollacci@di.unipi.it; Contributing authors: letizia.milli@di.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 [7, 23, 24]). 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 [1] and a lacking a world migration survey [51]. 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 [44]. 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 [13, 26, 53] and Microsoft Academic Graph [50] to examine the scientific ethnic and mobility networks and [2, 47]. Besides few exceptions (see [29, 40]), 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.

[1] Scopus, URL: https://www.scopus.com/

[2] Web of Science, URL: https://www.webofknowledge.com

[3] Microsoft Academic Knowledge Graph, URL: https://ma-graph.org/. The MAKG dataset is licensed under the Open Data Commons Attribution License (ODC-By).

2 Academic Mobility, Academic Networks and Knowledge Transfer

Analysing how, why and where highly-skilled individuals, in particular researchers, move has attracted accelerating interest in recent decades due to the socio-political evolution, globalisation and the knowledge-based economic approaches around the globe. In the context of internationalisation of academia, 'migration' and 'mobility' have been used interchangeably [41], 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 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 [46], but also a system for global knowledge transfer [10]. Having said that, international academic movements, flows and networks are recognised as beneficial transnational and transferable identity capital that are antitheses to intellectual parochialism [31]. 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 the international knowledge transfer.

Mobile academics are conceptualised through the interplay of multiple movements where knowledge is used as power and mobility as resource [16,

[4] The UN Migration Agency (IOM) defines a migrant as any person who is moving or has moved across an international border or within a State away from his/her habitual place of residence, regardless of (1) the person's legal status; (2) whether the movement is voluntary or involuntary; (3) what the causes for the movement are; or (4) what the length of the stay is. https://www.un.org/en/global-issues/migration

39]. 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 captures the official registration of residential movements. Several researchers have deployed varied scientific data sources and investigated scientific mobility from diverse perspectives including linking career evolution and international mobility [49], measuring knowledge transfer [3], analysing the convergence or discrepancy of countries in academic mobility and collaboration [11]. Moreover, scientific data have been exploited to study scale-free networks [6], temporal sequence analysis [5, 35], network statistical properties [33], measure international scientific collaboration [52, 32], and scientist mobility [27, 29, 12].

Together with multiple perspectives and approaches, state of the art is also rich in scientific data sources used to measure academic mobility and networks. In this sense, the fundamental discriminant is the type of data used: official registers (e.g., census data) versus unconventional data (e.g., social media). Focusing on data, resources are very heterogeneous in terms of distribution, access, necessary skills, content and size. Much of the scientific data-related literature exploits Scopus data [27, 29, 19, 45] or some other uses Web of Science data [11, 40]. Scopus provides access to more than 60 mln [22] scientific, technical and medical (STM) journal articles and their references. It is sold as an annual subscription and based on the number of researchers in the "full-time equivalent"[5] (FTE) [8]. For instance, Moed et al. [27] analyse mobility between institutions in Germany, Italy and the Netherlands. Leveraging bibiliometric data from Scopus, the authors profile academics, e.g., distinguishing "young researchers", and analyse the accuracy of links between academics and institutions. On the other hand, Web of Science is one of the most frequently used indexed database [34] which is currently maintained by Clarivate Analytics. Regarding scientific mobility, Robinson-Garcia et al. [40] analysed individual publication records based on publications covered in the Web of Science for the 2008–2015 period to distinguish between academic migrants (authors who disengaged from their country of origin) and academic travellers (authors who gain additional affiliations but maintain affiliation with their country of origin).

Another scientific data source is the Microsoft Academic Graph (MAG)[6] [43]. In [20] 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 [15] and rank conferences into ratings [14]. Finally, Panagopoulos et al. [36] 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 state of the art. Some works are focused on specific research areas such as biopharmaceuticals [9], molecular life sciences [21], and computer vision [18]. Moreover, studies can be limited in space, as in [21, 12, 26, 28] and focusing on specific regions such as China, Germany, Italy, the Netherlands, United Kingdom, United States of America and Mexico, rather than a holistic or global approach, as in [40].

This study will be the first attempt that will go beyond the existing scholarly work by working with an extensive dataset with wide time span (1800-2020) and high geographical coverage (180 countries) as well as by developing a new index (YDCI) to measure the international annual knowledge exchange on country level.

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 analyses to achieve these goals are elaborated below.

[5]The full-time equivalent represents the effort made or planned to carry out an activity or a project in terms of full-time resources.

[6]The MAG is available with a subscription and is licensed under the Open Data Commons Attribution License. Url: https://bit.ly/3sFB1F2.

3.1 MAKG

Our study is based on bibliometric data from Microsoft Academic Knowledge Graph[7] (MAKG) [17]. The dataset composes of several scientific collaboration-related data, split into 18 sub-sets. From these, we focus on:

- Authors: information about researchers, such as name and affiliation (253,641,783 entities).
- Affiliations: information on scientific institutions, e.g., research centers, academies, hospitals, etc., including name and Wikipedia url (25,431 entities).
- FieldOfStudy: information on the fields of study associated with the papers (229,716 entities).
- Papers: information on publications, including the year of publication and authors (209,792,741 entities). Publications belong to five categories: *JournalArticle* (82,886,342), *PatentDocument* (51,526,166), ConferencePaper (4,533,280), *BookChapter* (2,679,696), and Book (2,086,307).

The MAKG covers 180 worldwide countries and includes publications spanning from 1800 to 2020. We restrict our analysis those papers published from 1980 to 2019. Moreover, we focused only on "active" authors (according to [27]), 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

Starting from collaborations, we define a measure to describe the tendency of researchers to collaborate with colleagues working in the same country or not. This can help describe researchers and their dynamics and can be related to researchers' movements to understand its impact. The Yearly Degree of Collaborations' Internationality is calculated as the binary entropy of each researcher's colleagues' probability distribution of working countries annually. The result is multiplied by -1 if most countries are different from the researcher's country of affiliation.

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

where $\#c_{dif}$ is the number of countries other than that of the researcher, while $\#c_{same}$ is the number of countries equal to that of the researcher.

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 of countries different from their own. Conversely, a YDCI closer to 1 represents the tendency to collaborate with geographically heterogeneous 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.

We use the YDCI to cluster and describe researchers based on their collaborations types (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 $[1980, 2019]$. The value in cell (m, n) is, therefore, 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[8] 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.

[7] Version 2019-12-26, DOI 10.5281/zenodo.3930398.

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

As a second goal, we measure the worldwide knowledge transfer focusing on researchers' movements over affiliations from a geographical and temporal point of view. Given a country C, the *incoming mobility score* ($In(C)$) defines the countries' degree of mobility based on yearly incoming researchers. Similarly, the *outgoing mobility score* ($Out(C)$) defines the countries' degree of mobility based on yearly outgoing researchers. Further, the *mobility balance* estimates whether a country has more incoming or outgoing traffic of authors. 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 "report" of movements (outgoing or incoming accordingly with 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).

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

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:

$$F_X(\varkappa) = P(X \leq \varkappa) \tag{3}$$

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.

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

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, a large dataset of scientific publications and related entities (Section 3.1). As preprocessing steps, a semi-supervised Natural Language Processing (NLP) pipeline (which leverages Wptools[9] and Pycountry[10] python libraries) allows to geolocate affiliations with respect to countries. Further, the annual ego networks of authors are computed as the undirected graph of their scientific collaborators. The set of the preprocessing methods allows us to compute the Yearly Degree of Collaborations' Internationality (Section 3) by extracting the lists of the countries of affiliations of researchers' colleagues (co-authors).

By applying the K-Means[11] to the YDCIs, three well-separated clusters emerge. Figure 1

[9]Wptools: https://pypi.org/project/wptools/.
[10]Pycountry: https://pypi.org/project/pycountry/.
[11]Best GridSearch performance average silhouette 0.54.

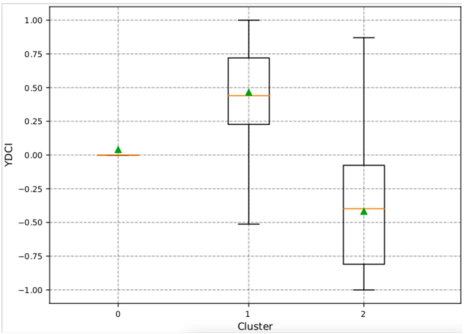

Fig. 1: Clustering of authors according to the YDCI.

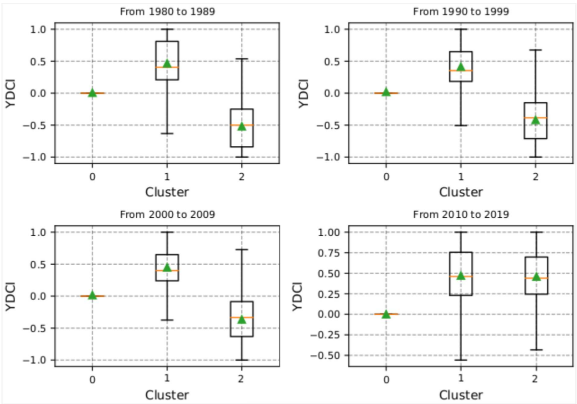

Fig. 2: Clustering of authors according to the YDCI over decades.

shows in the x-axe the clusters and in the y-axe the YDCI value for each cluster. *Cluster0* includes the 89.8% of the dataset (8, 008, 741 authors) and is composed of authors who tend to work alone or establish collaborations only with a few researchers from the same country. *Cluster1* represents the 3.11% of the dataset (277, 324 authors). This cluster identifies authors who tend to collaborate with geographically heterogeneous groups composed of researchers of countries different from theirs. Finally, *Cluster2* is composed of the 7.09% of the dataset (632, 679 authors). It is the opposite of *Cluster1*, and groups together those researchers that prevalently collaborate with geographically homogeneous groups composed of researchers from their same countries.

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 (Figure 2). The three-clustered structure still emerges in each decade, and the overall behaviour of the groups remains the same. However, data show that domestic contributions are increasing in the fourth decade, suggesting a contraction in prevalently inter-country collaboration patterns. Given our first research question focuses on knowledge transfer by analysing collaborations we focus on *Cluster1* and *Cluster2*. 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 (a-b) shows the maps for *Cluster1* and *Cluster2* respectively[12]. By considering both maps, it can be seen that trends in collaborations are geographically homogeneously distributed. Indeed, regardless of the cluster, most countries with the most extreme tendencies are the same, e.g., Mauritania, Guyana, and Libya. Focusing on *Cluster1* - which includes authors who tend to collaborate with colleagues located abroad - it emerges, at least a global scale, that American countries (both North and South), European and Australian ones obtain YDCI values of $[-0.50;\ 0]$, with non-extreme trends. On the contrary, the Asian and African continents show higher collaborations with foreign countries, with values close to -1. Moving to *Cluster2*, which includes authors who collaborate within their own country, similar to *Cluster1*, the whole of America, Europe, and Australia have YDCI values in the range $[0.25;\ 0.50]$. At the same time, Asian and African countries show higher YDCI values, between $[0.50;\ 1]$.

In addition, we analyse the decade-wise YDCI segmentation aggregated at the country level. Considering *Cluster1* in Figure 4 (top) we note a slow opening towards foreign collaborations over the decades. This trend, initially slightly stronger

[12]In both maps, the countries in white do not have an assigned score due to missing data.

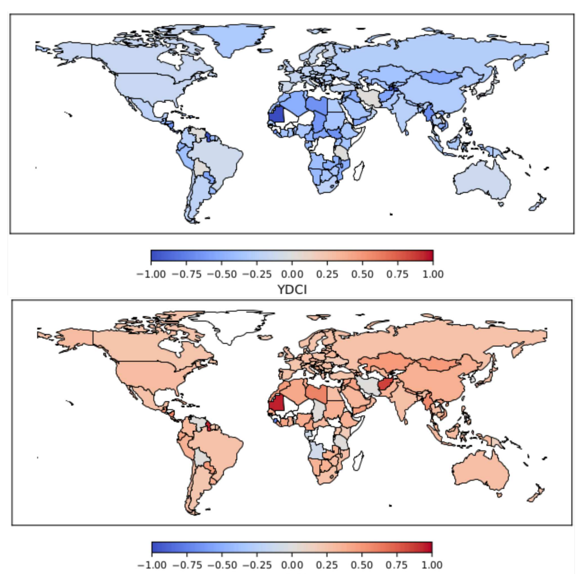

Fig. 3: (a) Average YDCI for *Cluster1* and (b) *Cluster2* countries.

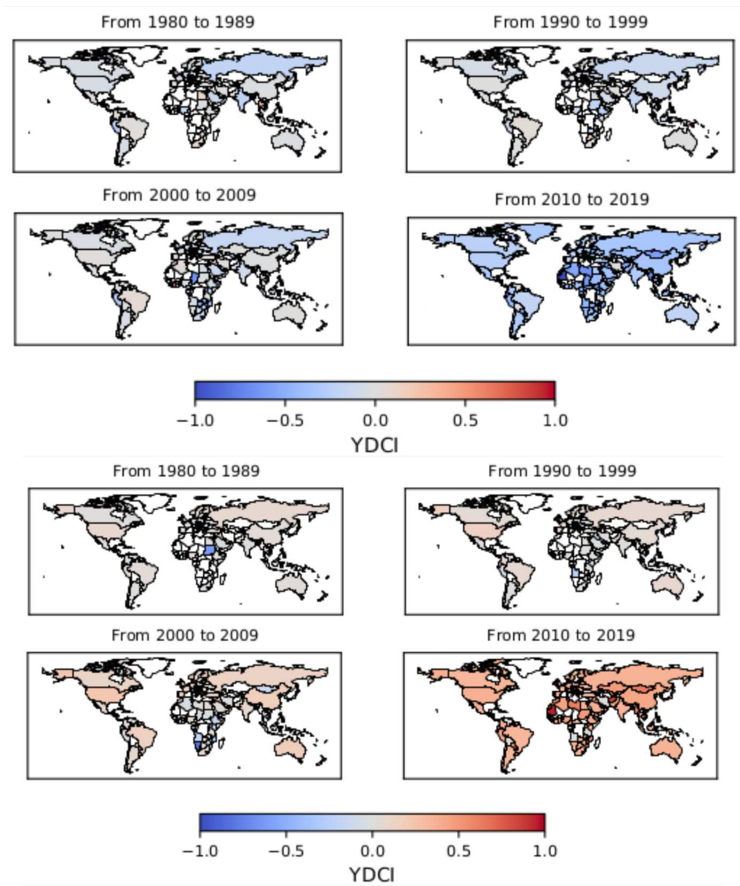

Fig. 4: (top) Average YDCI for *Cluster1* and (bottom) for *Cluster2* countries over decades.

on the Asian continent, becomes more drastic in the window from 2010 to 2019 worldwide. Even looking at *Cluster2* in Figure 4 (bottom), we note that the initial trend is mild on average, excluding Sudan and Iceland, where the YDCI is negative. From 1990 to 2009, the trend towards intra-national collaboration tends to grow, except in Africa where some countries show greater foreign collaboration, Namibia in the first place and follow Mozambique, Angola, Ethiopia, Kenya and Gabon. In the last decade, the trend to collaborate with colleagues of the same country spread worldwide.

To comprehensively observe the evolution in collaboration trends, we aggregate authors of the two clusters and calculate the averages of the YDCI scores again by country and decades. As shown in Figure 5, at a high level, we observe that American (both North and South) and Chinese authors tend to collaborate intra-nationally. On the contrary, some countries, including Russia, Canada, Argentina and Saudi Arabia, are moving from an initial - more or less intense - trend towards international to intra-national collaboration. The African continent has the most heterogeneous landscape, where the southern states, excluding South Africa, seem more open to foreign collaborations than the northern states. Based on these analyses, we can state, in a general way, that the temporal evolution of the YDCI shows progressive generalised work segregation and intra-national collaboration. Focusing on European countries (Figure 6), data shows heterogeneous YDCI distribution during the first decade (1980-1989). For instance, Spanish, Portuguese and Belgian researchers tend to collaborate intra-nationally, while Slovak, Lithuanian and Greek ones internationally. During the second decade (1990-1999), values capturing both domestic and international collaborations tend to

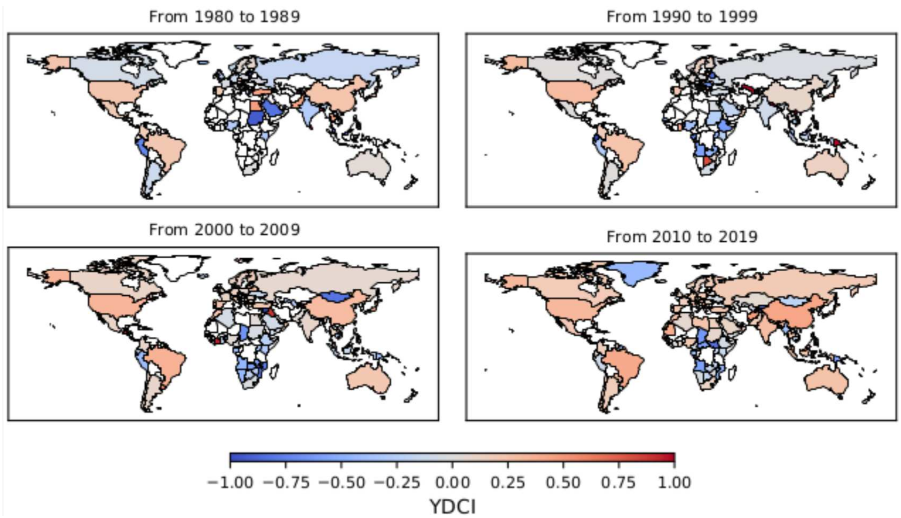

Fig. 5: YDCI over decades.

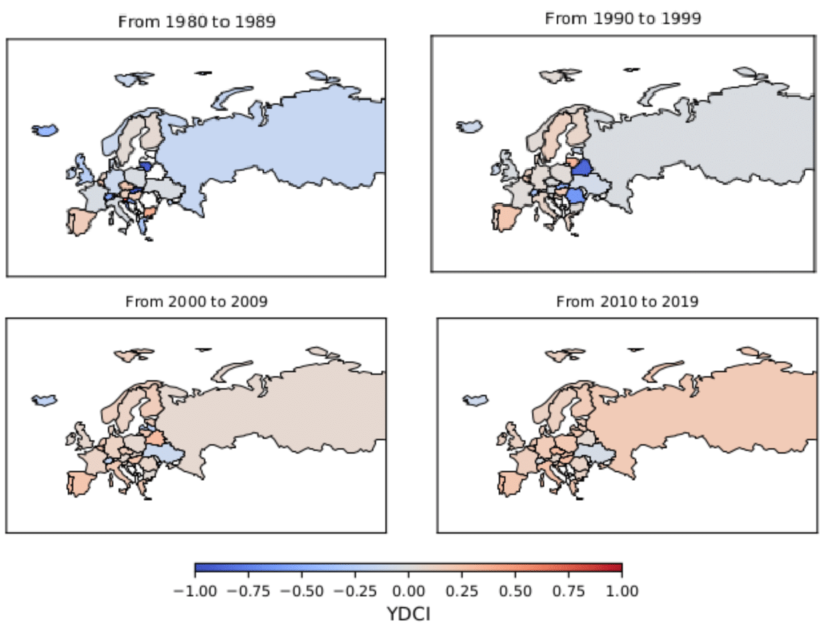

Fig. 6: YDCI for European countries over decades.

diminish. Furthermore, we observe a trend reversal in Lithuania. Finally, in the third and fourth decades (2000-2009 and 2010-2019, respectively), we observe an evident and consolidated generalised trend of collaboration between scholars from the same country.

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 mobility over time. China, like Russia, on the other hand, shows medium-high incoming mobility during the first decade, which then tends

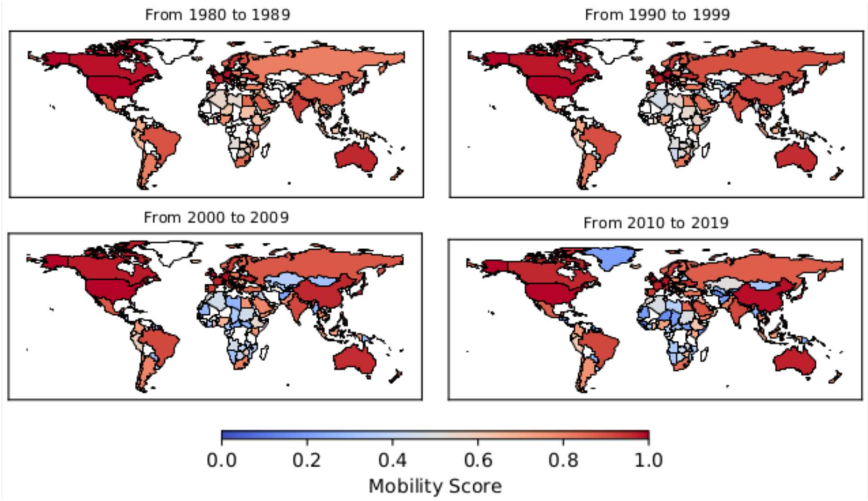

Fig. 7: Incoming mobility over decades.

to increase over time. However, in the Asian continent, there are countries with low and medium-low incoming mobility, i.e., Mongolia, Afghanistan, 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, while the United States shows medium-high mobility over time, China and Russia show increasing outgoing mobility over decades. Within the Asian continent, countries with low and medium-low incoming mobility are generally the same as those with low and medium-low outgoing mobility. Africa, for which we have 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. Every countries in our dataset generally show values in $[-0.10; \ 0]$, which means that, although with little difference, 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, where more authors tend to leave countries than enter.

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 and spatial dimensions and at different resolutions.

Secondly, we focus on the movements of researchers over affiliations worldwide over time. We define two mobility scores (In and Out) describing countries based on incoming and outgoing researchers. As a next step, we use these to compute the mobility balance, which estimates the difference between incoming and outgoing flows providing a comprehensive worldwide perspective.

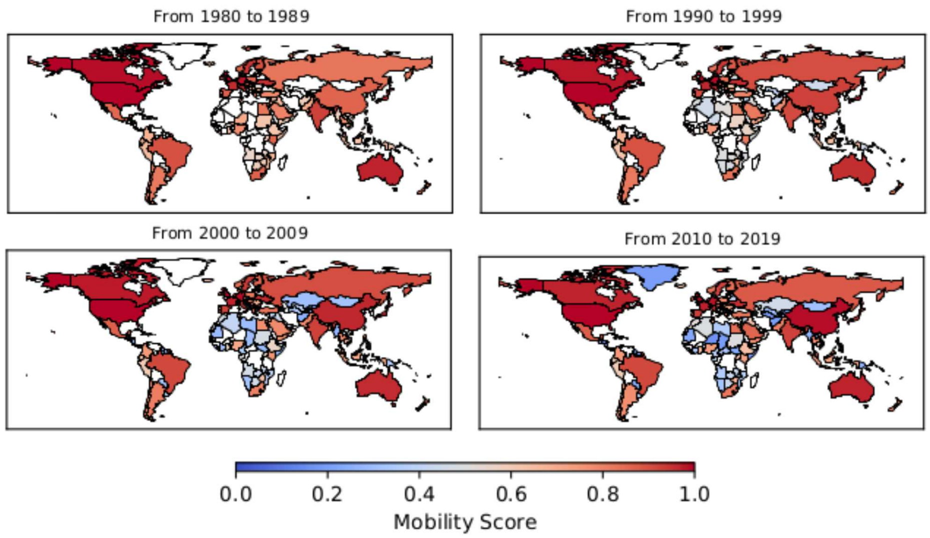

Fig. 8: Outgoing mobility over decades.

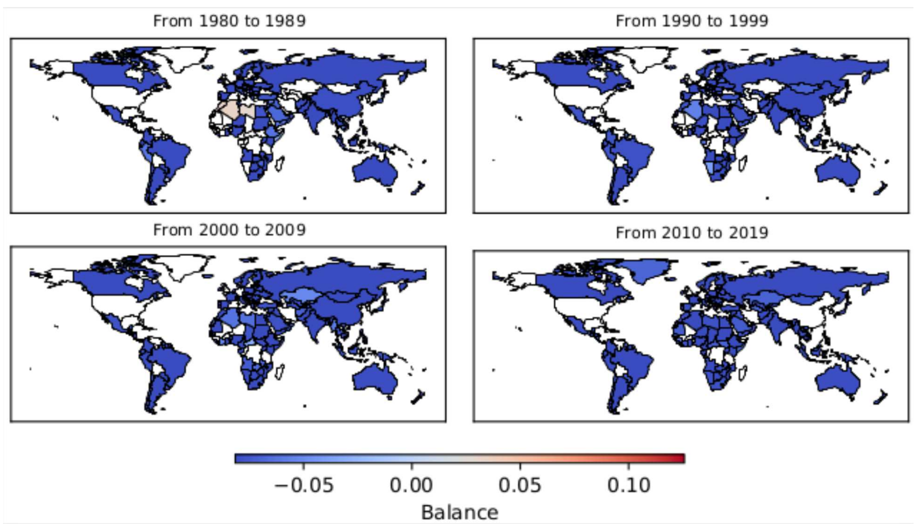

Fig. 9: Mobility balance.

Our findings indicate an ever-increasing trend towards intra-national collaboration and knowledge transfer. 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, impact of contextual factors could be examined to develop better understanding of the mobility patterns and changes in time. For instance, countries' YDCI trends and researchers movements 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.

Acknowledgments. This work is supported by the European Union – Horizon 2020 Program under the scheme "INFRAIA-01-2018-2019 – Integrating Activities for Advanced Communities", Grant Agreement n.871042, "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics" (http://www.sobigdata.eu) and by the Horizon2020 European project "HumMingBird – Enhanced migration measures from a multidimensional perspective" (Grant Agreement n. 870661).

Funding. This work is supported by the European Union – Horizon 2020 Program under the scheme "INFRAIA-01-2018-2019 – Integrating Activities for Advanced Communities", Grant Agreement n.871042, "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics" (http://www.sobigdata.eu) and by the Horizon2020 European project "HumMingBird – Enhanced migration measures from a multidimensional perspective" (Grant Agreement n. 870661).

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

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

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

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

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

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

[19] 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.

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

[21] 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.

[22] 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.

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

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

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

[26] 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.

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

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

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

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

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

[36] 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.

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

[38] 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.

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

[40] 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.

[41] 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.

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

[43] 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.

[44] 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.

[45] 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.

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

[47] 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.

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

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

[50] 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.

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

[52] 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.

[53] 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.