MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH

Konrad-Zuse-Strasse 1 · D-18057 Rostock · Germany · Tel +49 (0) 3 81 20 81 - 0 · Fax +49 (0) 3 81 20 81 - 202 · www.demogr.mpg.de

MPIDR Working Paper WP 2025-028 I September 2025
https://doi.org/10.4054/MPIDR-WP-2025-028

It's who you know — unless you're famous: Professional networks and prestige in scholarly mobility

Alexandra Rottenkolber
Ola Ali
Gergely Mónus
Jiaxuan Li
Jisu Kim
Daniela Perrotta
Aliakbar Akbaritabar I akbaritabar@demogr.mpg.de

This working paper has been approved for release by: Emilio Zagheni (sekzagheni@demogr.mpg.de),
Head of the Laboratories of Migration and Mobility and Population Dynamics and Sustainable Well-Being.

It's who you know — unless you're famous: Professional networks and prestige in scholarly mobility

Alexandra Rottenkolber[1,6,+], Ola Ali[2,6,+], Gergely Mónus[3,6], Jiaxuan Li[4,6], Jisu Kim[5,6], Daniela Perrotta[6], and Aliakbar Akbaritabar[6*]

[1]Institute for Analytical Sociology, Linköping University, Linköping, Sweden.
[2]Complexity Science Hub, Vienna, Austria.
[3]Corvinus University Budapest, Budapest, Hungary.
[4]University of Oxford, Oxford, England.
[5]Utrecht University, Utrecht, the Netherlands.
[6]Max Planck Institute for Demographic Research, Rostock, Germany.
[+]Contributed equally.
[*]Corresponding author, Email: akbaritabar@demogr.mpg.de

Abstract

Mobility of researchers is a key driver of knowledge diffusion, innovation, and international collaboration. While prior research highlights the role of networks in shaping migration flows, the extent to which personal and institutional ties influence the direction of scientific mobility remains unclear. This study leverages large-scale digital trace data from Scopus, capturing complete mobility trajectories, co-authorship networks, and collaboration histories of 172,000 authors. Using multinomial logistic regressions and discrete choice modelling, we systematically assess the effects of first- and second-order co-authorship ties and institutional linkages on scholars' mobility outcomes, focusing on their first career move. Our findings demonstrate that not only first-, but also second-order co-authorship ties — connections to a scholar's collaborators' collaborators — are a strong predictor for the direction of a move. Scholars with extensive individual professional networks, as well as those migrating abroad, are more likely to move along individual ties. In contrast, those from prestigious institutions, as well as those moving nationally, tend to follow institutional routes more often. Discrete choice models further confirm that both individual and institutional ties increase the probability of moving to specific research institutions, with individual connections being more influential than institutional ones. This research provides empirical evidence for the role that individual and institutional connections play in shaping high-skilled labour mobility. Furthermore, it has important implications for migration theory and policy, emphasising the need to support national and international collaborative networks, both individual and institutional, to foster scientific exchange.

Keywords: Scientific mobility; Science of Science; Co-authorship Networks; Bibliometric data

Introduction

Scholarly migration is a significant and growing form of mobility that not only influences individual careers (Biazoli et al., 2025; Sugimoto et al., 2017; Venturini et al., 2024), but also shapes the development of entire research fields, institutional collaborations, and innovation systems (Chou et al., 2019; Mahroum, 2000; Trippl, 2013). Scholars are key actors in the global flow of knowledge, and their movements across borders can accelerate the spread of ideas (Aman, 2020; Ganguli, 2015), foster innovation (Bosetti et al., 2015), and strengthen international partnerships. Understanding how co-authorship and institutional connections drive these patterns is critical for informing policies aimed at enhancing scientific cooperation and knowledge transfer within and between countries.

Unlike many other migrant groups, scholars are extensively documented through publicly available data on their publications, collaborations, and career trajectories, which are increasingly used for demographic research at subnational and international levels (Akbaritabar, Theile, & Zagheni, 2024; Akbaritabar et al., 2025). Traditional migration data, such as census records, visa applications, and employment statistics or surveys, often lack the granularity needed to track scholars' movements across institutions and countries. In addition, these data sources do not provide evidence on scientific collaboration networks that might influence mobility decisions. In contrast, digital trace data from bibliometric sources provides real-time, longitudinal insights into individual career trajectories, co-authorship networks, and institutional affiliations (Kashyap et al., 2023; Sîrbu et al., 2021), offering a level of detail and transparency typically not available for other migrant groups. By analysing this rich data, we can better understand how academic networks and institutional partnerships shape scholarly migration, thereby contributing to broader theories of high-skilled labour mobility and the knowledge economy.

These same networks also offer a rare opportunity to empirically examine a longstanding insight from migration theory: that networks shape the opportunity structures individuals face when making mobility decisions (Massey, 1990). Both kin ties, such as family and friendships, and weaker connections to broader communities help individuals navigate the migration process. These networks serve as channels for information and sources of social and economic support (Blumenstock et al., 2023). Despite their importance, these network effects remain difficult to quantify (Lin, 1999). Sociological and anthropological studies of migration provide strong ethnographic evidence that social and institutional networks shape migration flows, often using diaspora communities to demonstrate how household-, community-, and macroeconomic-level contextual factors contribute to stable migration pathways between specific origins and destinations (Massey & Espinosa, 1997; McPherson et al., 2001). With regard to scholars, both co-authorship connections and institutional relationships (e.g., between universities or research organisations) play critical roles in facilitating movement (Lu & Zhang, 2015). The availability of these networks, both individual and institutional ones, can shape career opportunities, access to research resources, and long-term professional success (Albarrán et al., 2017; Haines et al., 2011; Herz & Altissimo, 2021; Jonkers & Tijssen, 2008; Liu et al., 2023; Netz et al., 2020; Schaer, 2022a). This makes scholarly migration not only a relevant and increasingly significant phenomenon but also a unique case for analysing migration through the lens of network effects, and leads us to our first two research questions

RQ1: What role do individual-level connections play in shaping scholarly mobility? Specifically, do

scholars preferentially move to institutions where they have prior co-authors or are connected to via their co-authors?

RQ2: Do institutional partnerships and connections influence the likelihood of scholars moving between specific institutions? Specifically, do scholars preferentially move to institutions that are connected to their home institution?

Under the assumption that both individual and institutional ties may affect mobility, the strength of these ties is likely to matter, too. Recent research has emphasised not only direct ties but also higher-order connections as influential in career advancement and information diffusion (Christakis & Fowler, 2013; Granovetter, 1973; Petersen, 2015). However, it is unclear whether individual or institutional ties more strongly shape mobility decisions, and whether these effects are additive, non-linear, or whether simply having 'at least one' connection is sufficient. This leads to our third question:

RQ3: Do the effects of individual- and institutional-level connections on mobility outcomes scale with the strength of those connections?

Finally, the literature on academic stratification highlights the role of institutional prestige in structuring career trajectories (Burris, 2004; Holding et al., 2024; Kwiek, 2021; Nevin, 2019). Prestige can influence who gets hired where and may condition how scholars are able to leverage their networks. Because high-status departments already attract a large applicant pool, they can screen candidates based on publicly visible merit signals, thereby reducing their dependence on personal referrals (Podolny, 1993). For lower-ranked universities, the situation often looks different, and hiring through trusted co-author links can mitigate informational uncertainty (Fernández & Weinberg, 1997). Following these lines of thought, we hypothesise that prestige moderates the use of networks. Likewise, scholars with larger personal networks may be better positioned to access desirable opportunities, while those with smaller networks may face structural constraints (Akbaritabar, Castro Torres, & Larivière, 2024; Akbaritabar et al., 2025). Together, these factors may structure scientific mobility. Thus, we ask:

RQ4: Do institutional prestige and the size of a scholar's professional network moderate the relationship of a scholar's embedding on the direction of academic mobility?

To answer these research questions, we use a 2020 extract of Scopus data, which includes complete historical records of authors' publications and affiliations up to that year. Mobility events are identified by analysing changes in author affiliations at both subnational and international levels (Akbaritabar et al., 2025). We focus on the first career moves of scholars and the direction of mobility, which are key transitions that often shape academic trajectories in the long term (Schaer, 2022b; Woolley et al., 2008). We deliberately set aside the equally important question of whether a scholar becomes mobile, which may be governed by different mechanisms.

To assess the role of individual- and institutional-level connections on mobility outcomes, we construct co-authorship networks at both individual and institutional levels. At the individual level, we measure first- and second-order ties between scholars — that is, direct collaborators as well as the collaborators of collaborators — and track their affiliations. At the institutional level, we assess collaborative links between institutions. Figure 1 illustrates this analytical strategy. Our

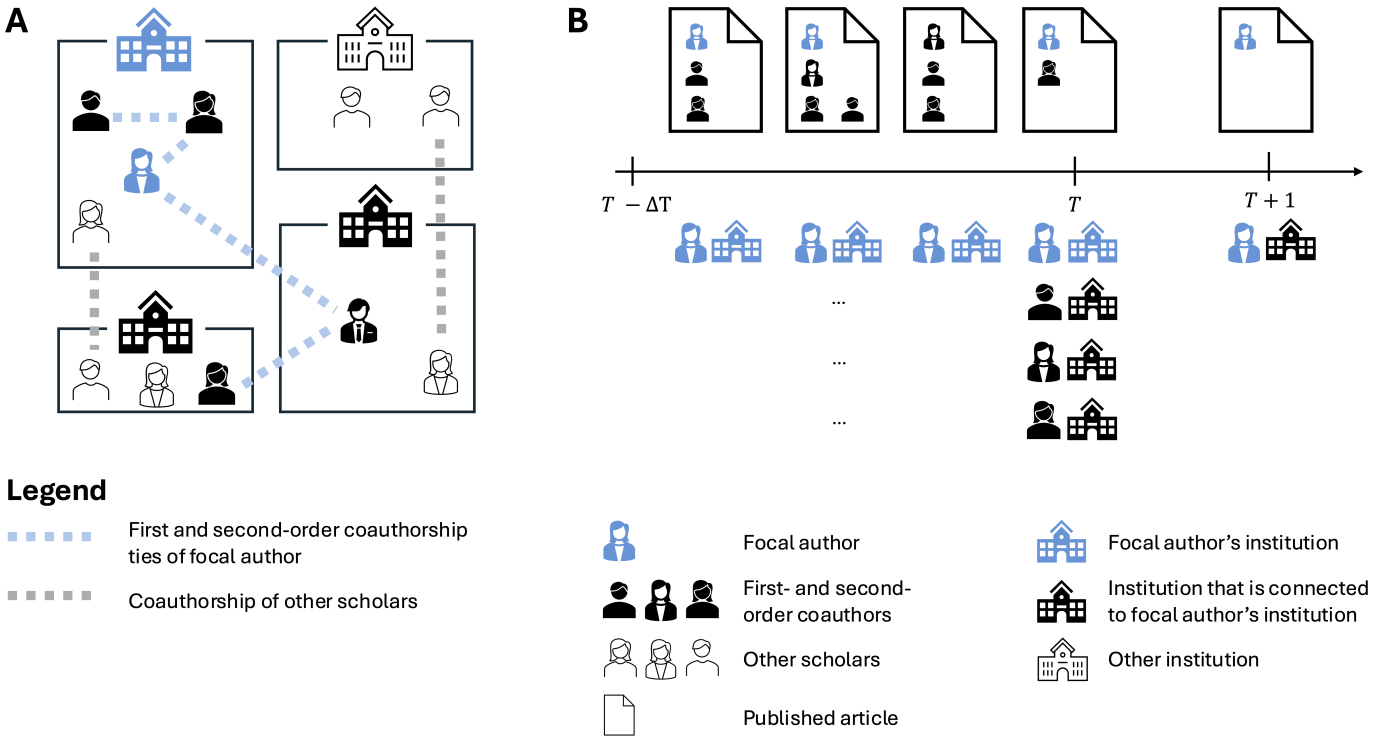

Fig. 1: **Illustration of the analytical strategy and retrospective measurement of scientific collaboration and mobility. A. Mapping individual-level and institutional-level collaboration ties.** Visualisation of the focal author's individual connections to different institutions prior to the mobility event (blue dotted lines). We trace the focal author's direct co-authors as well as the co-authors of co-authors, and record the institutions they are affiliated with. The institutional-level connections are independent of the focal author and represented by the blue dotted lines in combination with the grey ones. The snapshot at time T, i.e., the year before the mobility event, is used to map past connections and calculate the respective connection strengths. **B. Detecting migration events through affiliation addresses in publication records.** Timeline showing the stream of papers published by both the focal author (in blue) and their first- and second-order co-author (in black) prior to the observed mobility event of the focal author at time $T+1$. All authors' affiliations are documented for each publication, and a mobility event is recorded when the focal author's affiliation changes consistently between different years.

network variables capture co-authorship-documented professional ties, a conservative proxy for scholars' broader social and mentoring networks (Barber & Blumstein, 2022; European Comission et al., 2017; Katz & Martin, 1997). Because many influential relationships, e.g. advisors, informal mentors, conference contacts, do not always culminate in joint publications, our estimates should be interpreted as lower-bound effects of network embeddedness (see Data & Methods section for details). The connection strengths are measured as the proportion of a scholar's network affiliated with a given institution (individual ties) and as the relative intensity of collaboration between institutions (institutional ties). Institutional prestige is captured using the Leiden Ranking (Van Eck, 2024). To model mobility outcomes, we employ multinomial logistic regressions to categorise mobility events based on network ties and use discrete choice models to evaluate how the connection strength influences destination selection. The models account for both individual characteristics, such as gender, career age, and academic discipline, as well as institutional characteristics, including prestige ranking, size, and geographic location. This approach allows for a systematic comparison of how different forms of embeddedness structure academic mobility.

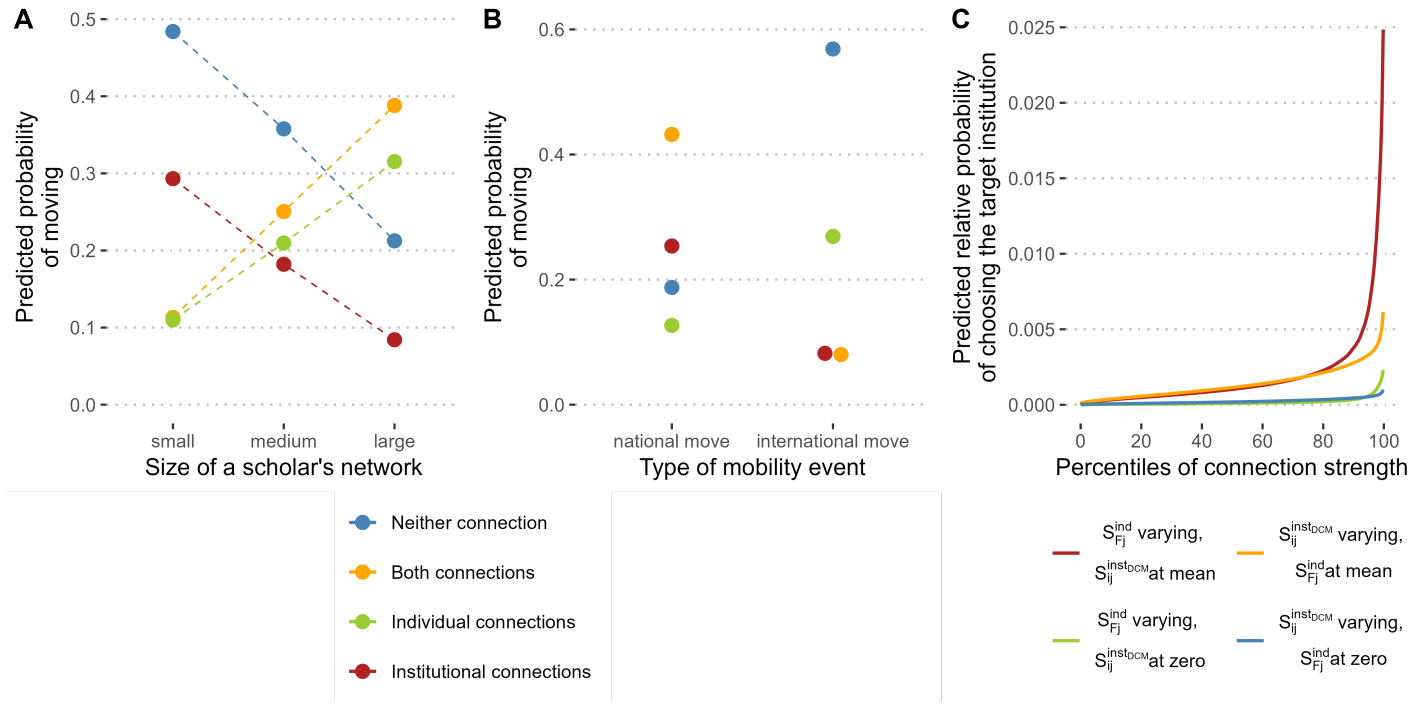

Fig. 2: **Effect of the size of a scholar's network (A), the type of mobility event (B), and the strength of a scholar's connection to a potential target institution (C) on the likelihood of a specific mobility outcome. A.** Predicted probabilities of moving along one of the four pathways in the event of scholarly mobility plotted against the size of a scholar's personal network (measured by first- and second-order co-authorship ties), as estimated by the multinomial logistic regression. While the likelihood of moving along individual ties or a combination of individual and institutional connections increases with a scholar's network size, the likelihood of moving along no ties or institutional connections decreases. **B.** Similar analysis as in panel A, but here the predicted probabilities are plotted against the type of mobility, i.e., national or international. Whereas most scholars who move nationally do so along institutional ties or a combination of institutional and individual ties, it is most common for scholars who move internationally not to follow any connection at all, or if they do, they move along individual-level connections. **C.** Marginal relative predicted probabilities of choosing a certain alternative institution plotted against the percentile rank of the individual and institutional strength measure, all else equal, as estimated by a discrete choice model (conditional multinomial logistic regression). The probabilities should only be interpreted in relative terms, not in absolute ones, as the absolute values are conditional on the number of options to choose from (i.e. the size of the choice set). The percentile ranks were only calculated for connection strength values that were different from zero, i.e., where a connection was present at all. The use of percentile ranks allows for the comparison of the individual and institutional strength measures. If a potential target institution is connected to an author via both individual and institutional ties, the probability of being chosen by the scholar is much higher than if only one of the two connection types exists. Furthermore, the probability of a potential target institution being chosen increases disproportionately at high percentile ranks, especially if both types of connections are present. Note that the plotted marginal predicted relative probabilities as shown in panel C should only be interpreted in relative terms. An interpretation in absolute terms is not meaningful here.

Results

All results are conditional on mobility: they describe how a researcher's individual and institutional embedding is associated with the choice of destination, not with the prior decision to move at all.

Network size matters. The size of a scholar's individual professional network plays a crucial role in shaping the direction of mobility. Our multinomial models consistently indicate that the number of first and second-order connections – measured through coauthors and coauthors of coauthors – is among the most influential factors relating to scholars' movement patterns. This finding is further supported by an additional analysis using a Decision Tree to predict mobility outcomes (see Supplementary Material SI-2). As shown in Fig. 2A, scholars with large personal

networks are significantly more likely to follow individual connections or a combination of individual and institutional ties (see 'individual connections' and 'both connections' in the Figure). In contrast, those with smaller networks mostly follow pathways unrelated to pre-existing connections (i.e., 'neither connection'). Scholars with medium-sized networks show a relatively higher likelihood of moving along institutional pathways compared to highly connected scholars, and are more likely to rely on individual ties and a combination of both types of ties compared to those with smaller networks.

Analysing discrete choice models allows us to assess the relative importance of individual- and institutional-level ties in scholars' destination choices, conditional on having at least one of the two possible connections. Figure 2C visualises how the probability of selecting a target institution varies with the strength of a connection. We observe a nonlinear relationship: When connected to an institution through only one type of tie (individual or institutional), a noticeable increase in the probability of choosing that institution occurs only for connection strengths in the highest percentiles (80th percentile and above). In contrast, when both types of ties are present (i.e., a combination of individual and institutional ties, see red and orange lines), the likelihood of choosing an institution increases almost linearly with connection strength up to the 75th strength percentile, and then rises more rapidly with especially strong ties. The strongest effect is observed when scholars have strong individual ties and an average institutional embedding (red line).

Individual and institutional connections are differently relevant for national and international mobility. Mobility direction patterns differ significantly between national and international moves (see Fig. 2B). For national moves, the largest proportion of scholars follows a combination of individual and institutional ties, or, less frequently, institutional ties only, reinforcing the importance of overlapping personal and institutional ties. In contrast, international moves frequently fall into the 'neither' category, with the highest predicted probability among all types (close to 0.6), indicating that many international moves occur outside established career networks. When ties do influence international moves, they are most likely individual rather than institutional, thus providing crucial support for navigating cross-border transitions. Institutional ties, however, appear to play a limited role in shaping international mobility and are mainly at play in national mobility.

Prestige conditions the role of connections. The prestige of both source and target institutions substantially shapes how connections influence migration decisions. Scholars affiliated with higher-ranked source institutions are more likely to follow institutional connections and less likely to rely on individual ties (see Fig 3A.1), suggesting prestige homophily in hiring practices – a finding consistent with previous work (Clauset et al., 2015). In addition, we did not find any differential effects in how connection strength (individual or institutional) operates based on the prestige of the source institution (see Fig 3B.1 & 3C.1). The prestige of the target institution, however, exerts a more complex effect: The probability of moving along individual ties increases with institutional rank, but peaks at medium-high ranks before declining at the highest levels. In contrast, the probability of moving along institutional pathways, or via a combination of both individual and institutional ties, rises steadily as the target institution's prestige increases (see Fig 3A.2). These findings suggest that moving *from* a high-prestige institution is commonly facilitated by institutional pathways, whereas moving *to* a high-prestige institution relies more on individual

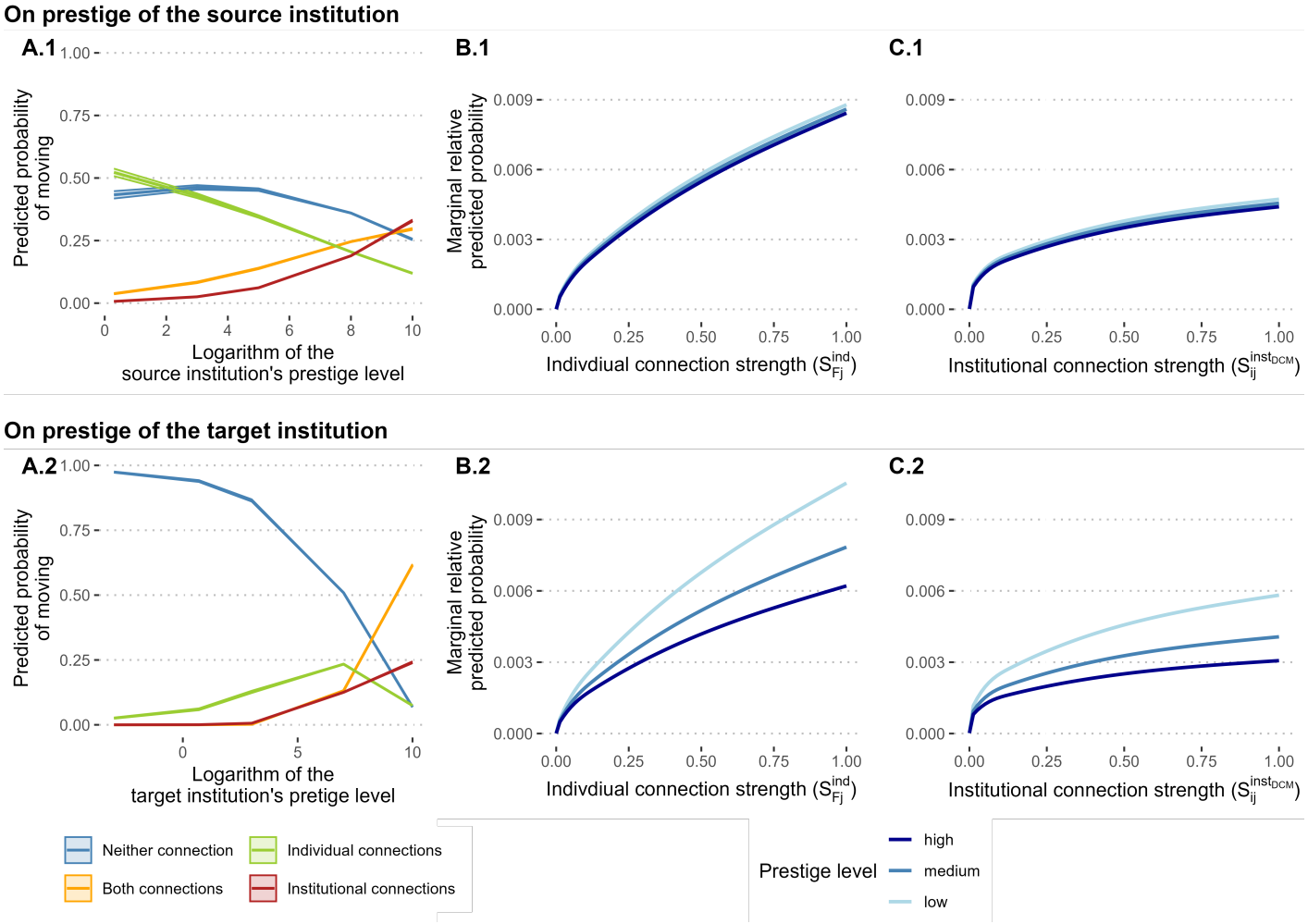

Fig. 3: **Effect of the prestige of an institution on the predicted probabilities to move from there or be chosen as the destination.** Results of the multinomial logistic regression are visualised in **A.1** and **A.2** for the prestige level of the source and target institutions, respectively. If the prestige of the source institution is low, scholars mostly move along individual ties or do not follow any connection. If it is high, however, scholars become more likely to move either along institutional connections or a combination of institutional and individual ties, and it becomes less likely for them to go somewhere where they have no connection to yet. With regard to the target institution, if the prestige is low, scholars mostly move there without any previous connections. Target institutions of a high prestige ranking, on the contrary, are more frequently reached via institutional connections or a combination of institutional and individual ties compared to having no ties or having individual ties only. Results of the discrete choice model (conditional multinomial logistic regression) are shown in Figure B.1 – C.2 with **B.1** and **B.2** illustrating how the prestige of the source and target institution interact with the individual connection strength and **C.1** and **C.2** with the institutional connection strength, given that there is at least one connection present. While we did not find any heterogeneity in the effect of both the individual and institutional connection strength with regard to the prestige of the source institution (see top row panels on the figure), we find some heterogeneity with respect to the target institution's prestige (bottom row panels). Institutions of lower prestige rank benefit from having ties to potential recruits more than institutions of higher prestige rank. This pattern is consistent across all levels of connection strengths for both individual and institutional connections, but gets more pronounced the stronger the connections are (see panel B.2 and C.2).

connections.

When examining how the effect of the connection strength varies by the prestige level of the target institutions, we find that lower-ranked institutions rely more heavily on both individual and institutional ties when recruiting scholars (see Fig. 3B.2 & 3C.2). This effect becomes more pronounced as connection strength increases. In contrast, high-ranked institutions exhibit less reliance

on pre-existing ties, suggesting that other selection mechanisms, such as merit-based evaluation, play a stronger role in these contexts.

This result aligns with previously published arguments and findings: Institutional prestige can be viewed as a positional resource (location in a recognised status hierarchy) that shapes both supply and demand in academic labour markets. Highly prestigious institutions routinely attract a larger pool of high-quality applicants and can screen them on universally visible merit signals, such as citations, grants, and awards, which ultimately reduces dependence on personal referrals (Clauset et al., 2015; Podolny, 1993). In contrast, mid- and lower-ranked institutions face higher information uncertainty and therefore have to rely more heavily on trusted channels, such as referrals from colleagues, to identify suitable hires (Fernández & Weinberg, 1997). Furthermore, mobility along institutional pathways can be interpreted as being consistent with status-closure dynamics and prestige homophily, whereby high-status institutions preferably exchange talent among each other (Burris, 2004; Rivera, 2012). Our findings also echo a cumulative advantage logic: An early placement at a prestigious institution amplifies future opportunities independently of existing ties (Allison & Long, 1990; Clauset et al., 2015).

Other moderating factors. While the literature has discussed other personal and institutional characteristics which might also shape the direction of mobility decisions, their effects, in our analysis, tend to be less pronounced. Career age negatively correlates with network-directed migration — older scholars are more likely to fall into the 'neither' category. This could mean that more senior scholars have accumulated enough research prestige through their prior work and publications that they could select destinations without prior ties to move to. Productivity, measured by past publications, has a positive impact on moving along individual or a combination of both types of ties, indicating that more productive scholars are better positioned to leverage their networks. Gender differences are minor but present: men have a slightly lower probability of moving along individual ties, or a combination of both individual and institutional connections (see Supplementary Material SI-4). Discrete choice models confirm that the direction of migration decisions is strongly shaped by connection structures, particularly in combination with institutional prestige. Notably, we did not find any gender effect in the extent to which men and women use their resources (see Supplementary Material SI-5). However, we did not investigate whether men and women differ structurally in their resources, i.e., which institutions they are placed in/they are connected to. This question is still open for further investigation.

Overall, our findings underscore the intertwined roles of personal connections, institutional prestige, and migration context in shaping academic mobility.

Discussion and Limitations

High-skilled labour migration is often framed as a rational response to institutional incentives or career advancement opportunities. Yet, the decision of where to move is deeply embedded in social structures. In this study, we use digital trace data from 172,000 scholars and their co-authorship networks to examine how individual and institutional connections shape academic mobility. Leveraging multinomial logistic and discrete choice models, we infer how professional connections jointly structure researchers' opportunity spaces in their early academic careers. The methodological ap-

proach adopted here, i.e., the use of large-scale bibliometric data to track migration and map relational structures, enables us to quantify these dynamics in ways that traditional migration data sources, such as surveys, cannot.

Rather than treating ties in isolation or combining different sets of ties without proper differentiation, we demonstrate that individual and institutional professional networks play distinct yet interconnected roles in shaping scholarly migration. We show that individual professional networks, and especially higher-order connections, are more strongly associated with where scholars move than institutional connections, although both matter, especially when they operate together. For example, scholars with large individual networks are more likely to follow individual connections, while those with smaller networks often move outside predefined career pathways. This suggests that professional individual connections not only provide opportunities but also enhance agency in mobility decisions. Additionally, international moves often occur outside established networks, with individual ties playing a crucial role in facilitating cross-border transitions. This highlights the challenges of international mobility and the importance of personal connections in overcoming them, which is consistent with longstanding insights from migration literature (Massey, 1990). In contrast, national moves are more likely to follow institutional connections or combined ties, potentially amplifying within-country stratification.

We further find that institutional prestige moderates how a scholar's embedding facilitates mobility. Researchers affiliated with high-ranking institutions are more likely to follow institutional pathways or a combination of institutional and individual-level professional ties, reflecting prestige homophily (Clauset et al., 2015). Conversely, scholars affiliated with lower-ranked institutions tend to depend more on individual professional networks when making mobility decisions. Our results also show that prestigious institutions are able to attract researchers with less reliance on pre-existing individual ties, while institutions with lower prestige benefit disproportionately more from strong individual-level connections during recruitment. This asymmetry underscores how institutional reputation interacts with social networks in structuring academic mobility.

While this study advances our understanding of how networks shape the migration of researchers, several limitations must be acknowledged. First, scholarly migration represents a specific form of high-skilled mobility, and the role of networks may differ in other migration contexts. Second, while Scopus provides extensive data (Baas et al., 2020), its temporal and spatial coverage is skewed towards Western countries and English-language publications. Third, co-authorship networks may not capture all personal connections (Katz & Martin, 1997; Laudel, 2002), potentially underestimating the influence of individual ties. Fourth, while we examine the role of prestige, we do not explore how connections are used to move *up or down* the prestige hierarchy. Fifth, while we investigate gender differences in the use of ties, we do not address structural disparities in access to resources. Finally, our models treat mobility as a scholar-driven decision, whereas academic hiring is a two-sided matching process, involving both applicant and institution. Future work should integrate employer demand, for instance, by combining bibliometric mobility traces with job postings or department growth metrics.

Despite these limitations, we argue that our study advances our understanding of the social foundations of scientific mobility by showing how different types of network ties interact to shape scholars' opportunity spaces. Our findings reaffirm that mobility decisions are not made in isolation

but are fundamentally shaped by the relationships scholars form within the academic system (Lin, 1999; Massey, 1990) as well as the reputation of the institutions they interact with. We propose several directions for future research, namely: i) integrating a first-stage 'move vs. stay' decision to distinguish between anchoring effects and steering effects of networks, ii) exploring structural disparities in access to resources across genders, iii) modelling the two-sided nature of academic hiring more explicitly, and iv) examining the role of digital platforms in shaping mobility opportunities. By deepening our understanding of these dynamics, we can better support global scientific exchange and foster a more equitable and efficient allocation of talent in academia.

Data and Methods

Data

For this study, we utilise a 2021 excerpt of Scopus, covering all records until the end of 2020 (Baas et al., 2020), a widely recognised data source for scholarly migration studies (Akbaritabar, Theile, & Zagheni, 2024; Akbaritabar et al., 2025; Zhao et al., 2023). This data is provided by the German Competence Network for Bibliometrics (Schmidt et al., 2025) via the Max Planck Digital Library. Our dataset includes around 172,000 authors (see Supplementary Material SI-1), for which we identify mobility events by tracking consistent changes in authors' affiliations over time, following the approach suggested by Akbaritabar, Theile, and Zagheni (2024). To achieve this, we geocode affiliation addresses to subnational regions, enabling us to distinguish between internal (within a country) and international (between countries) mobility events (Akbaritabar et al., 2025). By taking these steps, we are able to construct complete individual-level mobility histories for each author. In addition to the mobility histories, we create authorship records, including authors' first- and second-order co-authorship networks, publication histories, the country of an author's affiliation, as well as their discipline, which is inferred based on the discipline classification of the publication using Scopus' All Science Journal Classifications (Baas et al., 2020). An author's gender is inferred using their first name, provided by Zhao et al. (2023). As we are interested in studying how researchers' network embedding influences the direction of their move, we focus on mobile authors only. Furthermore, we focus on first mobility events only (the mobility step after an author's first publication, which is usually their first position after obtaining a PhD), as this stage is often especially critical for career trajectories and long-term outcomes (Schaer, 2022b; Woolley et al., 2008).

To measure individual connections, we use one of the most widely applied proxies for it: co-authorship ties. However, we acknowledge that this proxy only partially captures our interest in individual connections, as not all professional collaborative ties are reflected in the form of shared publications (Katz & Martin, 1997; Laudel, 2002). Hence, we expect our findings to be more on the conservative side. For each migrant scholar — those who moved at least once — we calculate the number of co-authors they had before their move, as well as the number of second-order ties, which represent the co-authors of their co-authors. This distinction is crucial, as it allows us to differentiate between the influence of direct connections and 'friend of a friend' effects. For institutional pathways, we aggregate all co-authorships and migration events between research

institutions. Using these distinct measures for both individual and institutional networks, we categorize the direction of movements based on whether the scholar's move followed individual or institutional ties, resulting in four possible directions, namely: along individual ties only, along institutional ties only, along a combination of individual and institutional ties (referred to as 'both'), and not along any ties (referred to as 'neither'). We then tested these categories using different models. Figure 1 shows a schematic visualisation of these individual and institutional ties before the mobility event and our retrospective approach in measuring collaboration and mobility.

We use institutional rankings as a proxy for prestige, considering it as a potential factor influencing migration decisions (Holding et al., 2024; Nevin, 2019). Specifically, we use the Leiden Ranking's[1] metric for $P(\text{top } 50\%)$ publications, which captures the share of publications ranked among the top 50% most frequently cited within the same field and publication year. This metric encompasses approximately 1,400 universities from 2006 to 2018 (Van Eck, 2024).

As the predictors in our observational data vary endogenously only, and we cannot rule out all potential unobserved confounders, such as individual talent or quality of individuals' research output, we interpret the estimated associations descriptively rather than causally.

Strength measures

In our study, we aim to determine whether an individual (a scholar) moves along an individual direction — that is, influenced by their co-authorship connections — or along an institutional direction, shaped by broader affiliations between institutions. We condition on the event of a move, as we are interested in how a scholar's embedding influences the direction of a move, and not whether it triggers a move in the first place.

Co-authorship-documented ties are a conservative proxy for scholars' broader professional networks — just the visible tip of a much larger iceberg. Survey evidence and previous literature shows that many mobility-relevant relationships — e.g. mentor sponsorship, previous colleagues and acquaintances, informal conference contacts — never culminate in joint papers, while some multi-author papers mask only peripheral interaction (Barber & Blumstein, 2022; European Comission et al., 2017; Katz & Martin, 1997). We therefore treat co-authorship as necessary but not sufficient evidence of professional connection, and interpret our coefficients as lower-bound estimates of network embeddedness. In the following, we describe how we operationalise these two dimensions to quantify the strength of these connections.

Constructing the individual-level tie strength measure

To quantify the individual-level embedding of an author before the migration event occurs, we propose the individual-based strength measure, $S^{\text{ind}}_{Fj}(T)$, specified in (Eq. 1). It describes how strongly an individual F is connected to an institution j via their first- and second-order co-authorship network in the years before a mobility event is observed. It is defined as follows:

$$S^{\text{ind}}_{Fj}(T) = \frac{C^{\text{1st \& 2nd}}_{Fj}(T)}{\sum_j C^{\text{1st \& 2nd}}_{Fj}(T)} \tag{1}$$

[1] https://www.leidenranking.com/

where $C_{Fj}^{\text{1st \& 2nd}}$ represents the number of first- and second-order co-authors the individual F has at institution j at time T, and $\sum_j C_{Fj}^{\text{1st \& 2nd}}$ is the total number of first- and second-order co-authors the individual F has at the same time point T, regardless of their institution.

We also examined the observed effects when only first-order connections were considered. The respective formula can be found in the Supplementary Material (Eq. SI-1).

Constructing the institutional-level connection measure

To quantify the institutional-level connection strengths, we propose a time-dependent measure based on past collaborations. The connection strength is evaluated over the five years preceding the year of the migration event. Two institutions are considered linked if they co-authored an article in a given year. The collaboration-based strength measure, $S_{ij}^{\text{inst}}(T)$, between institution i and institution j at a given year T takes the following form (Eq. 2):

$$S_{ij}^{\text{inst}}(T) = \sum_{t=T-5}^{T} \frac{A_{ij}(t)}{\sum_{j,j\neq i} A_{ij}(t)} \tag{2}$$

where $A_{ij}(t)$ is the total number of articles co-authored by scholars at the two institutions i and j at time t and $\sum_j A_{ij}(t)$ is the total number of articles co-authored by scholars at institution i, excluding those co-authored with institution j.

Institutional collaboration networks are denser than individual networks. Hence, for the Decision Trees and the multinomial logistic regressions, we collapsed our measure into a binary one. We consider an individual to have moved along an institutional tie only if the connection strength is within the top 10% of connected institutions to their previous institution at that time.

For the discrete choice models (i.e., the conditional multinomial logistic regressions), our measure needed to be continuous. We used a slightly altered version of $S_{ij}^{\text{inst}}(T)$ (as stated in Eq. 2) to enhance computational efficiency and address constraints related to computing time. Instead of using a rolling time window, we calculated the institutional connection strengths sensitive to different time periods ΔT, i.e., before 2000, 2000–2005, 2005–2010, 2010–2015, and later than 2015, as well as fields (Eq. 3).

$$S_{ij}^{\text{inst}_{DCM}}(\Delta T) = \frac{A_{ij}^{field}(\Delta T)}{\sum_{j,j\neq i} A_{ij}^{field}(\Delta T)} \tag{3}$$

We also tested an alternative specification for the institutional strength measure, considering past mobility between institutions instead of past collaborations. The corresponding formula can be found in the Supplementary Material (Eq. SI-2).

Modelling strategies

For the multinomial logistic regression models, our dependent variable is categorised into four outcomes: whether migration follows individual-level ties, institutional-level connections, both types of connections, or neither. Here, the question is whether a scholar falls into one of these four categories. Using discrete choice models, we further analyse how the strength of individual and institutional connections affects the probability of moving to specific institutions. For this part,

the focus shifts to how the strength of a connection — whether individual, institutional, or both — changes the probability of a scholar moving to a certain target institution.

Multinomial logistic regression

To analyse how individual and contextual variables influence the likelihood of a researcher following a specific network direction, we use Multinomial Logistic Regression (MLR) models. In MLR models, we predict outcomes that fall into multiple categories by selecting one category as the baseline and expressing the log-odds of the remaining categories relative to the baseline. The probabilities of each category are denoted by π_j for $j = 1, \ldots, c$, where c is the total number of categories. Each category (except the baseline) has a separate logit equation, allowing us to model the effects of explanatory variables on the odds of being in one category versus the baseline. The choice of the baseline category is arbitrary, and the same maximum likelihood parameter estimates are obtained for a pair of categories regardless of which baseline category is used.

For a model with c categories and p explanatory variables, the logit for category j, relative to the baseline category $c, j \neq c$, can be expressed as (Eq. 4):

$$\log\left(\frac{\pi_j}{\pi_c}\right) = \alpha_j + \beta_{j1}x_1 + \beta_{j2}x_2 + \cdots + \beta_{jp}x_p, \quad j = 1, \ldots, c-1. \tag{4}$$

In this form, α_j represents the intercept for category j, and β_{jk} denotes the effect of the k^{th} explanatory variable x_k on the log-odds of being in category j rather than the baseline. Each category thus has distinct parameter values, allowing flexible modelling of varied influences across categories.

For the MLR model, the set of control variables includes the following personal characteristics: (1) gender, (2) career age at the time of migration (measured as the difference between the year of migration and the year of a scholar's first publication in our database), (3) academic field, (4) productivity (measured by the total number of papers published before migration), and size of the network (measured by the size of the first- and second-order co-author network). Productivity and network size are stratified into tertiles and quantiles within each academic field, respectively, to account for variations in publishing behaviour. For the size of the network, those in the first quantile are classified as having small networks, those between the second and third quantiles as having medium-sized networks, and those in the top quantile as having large networks. We also include contextual variables to account for the environment in which migration decisions are made (Vaccario et al., 2021). These variables include (5) whether a move was national or international, (6) the individual's academic cohort (determined by the year of their first observed publication), (7) the timing of the move (the time period in which the move happened), (8) the region of the source and target institutions, and, importantly, (9) the prestige rankings of both source and target institutions.

Discrete Choice Models

While the results from the MLR analysis can provide insights into the likelihood of a scholar with certain characteristics (gender, career age, etc.) exhibiting a particular type of move (along

individual ties, institutional ties, along ties of both kind, or without ties), the question of how much the actual connection strength matters for an individual's decision-making process, and whether individual or institutional ties more strongly shape the opportunity space for scholars, remains unresolved.

To address these questions, we can formulate the analysis within the framework of discrete choice analysis (Ben-Akiva & Lerman, 1985; E. Bruch & Swait, 2019): Here, the decision-making process of every author is modelled at the individual level, taking the characteristics of the alternatives as well as the individual into account. This allows us to model the connection strengths as independent variables that influence the choice process of an author. In this modelling framework, authors are presented with different institutions they can choose from – hypothetical ones and the actually chosen ones. These institutions differ in their attributes, in our case, the individual and institutional connection strengths, as well as the prestige ranking. The probability that an individual scholar i chooses to move to an institution j is then given by Eq. 5:

$$Pr(y_{i,j} = 1) = \frac{\exp\left(\boldsymbol{\beta}^\top \mathbf{X}_{i,j} - \log(q_{i,j})\right)}{\sum_{(j')\in\mathcal{C}_i} \exp\left(\boldsymbol{\beta}^\top \mathbf{X}_{ij'} - \log\left(q_{i,j'}\right)\right)}. \tag{5}$$

where $\mathbf{X}_{i,j}$ is a matrix of predictors that describe an individual i's attributes as well as alternative-level attributes for the institution j as seen by the individual i.

In the discrete choice analysis, we included the following variables: (1) a binary indicator for whether an individual has at least one individual-level tie to a potential target institution, (2) a binary indicator for whether an alternative is among the top 10% of institutions in terms of collaborations based on Eq. 2, (3) a continuous indicator, that given (1), indicates the strength of the connection as specified in Eq. 1, (4) a continuous indicator, that given (2), indicates the strength of the connection as specified in Eq. 3, (5) the prestige ranking of the target institution, (6) the prestige ranking of the source institution, (7) whether a move is international or national, and (8) in which region of the world an alternative is located. We furthermore included interaction terms between the different strength measures and the ranking of the source and target institution to check for heterogeneity in the effect of existing ties, i.e., between (1) and (5), (1) and (6), (2) and (5), (2) and (6), (3) and (5), (3) and (6), (4) and (5), (4) and (6). Furthermore, j indexes the alternatives present in individual i's choice set.

Because vacancy data are unavailable, we make an assumption about which alternatives a scholar realistically could have had available as options. For our specific case, we used a stratified sampling approach to generate choice sets of size 50. The four mutually exclusive sets of institutions from which we sampled correspond to the categories used to classify mobility events in the decision tree and multinomial logistic regression framework, namely: institutions that an individual was connected to via (1) both institutional and individual level ties (sampled in at 20%), (2) institutional level ties only (35%), (3) individual-level ties only (35%), or (4) neither tie (10%, or as much was needed to fill the remaining open spots). This stratified sampling approach implies that choice sets vary across individual decision-makers. As the alternatives present in the choice set were not chosen randomly, we included an adjustment parameter in the model to account for their uneven likelihoods of inclusion: q_{ij}, where q_{ij} represents the probability of sampling alternative j into the choice set of respondent i (E. E. Bruch & Mare, 2012; Jarvis, 2019).

Sampling a choice set using prior knowledge can improve the precision of the estimates, as the choice sets will exhibit more of the alternatives that an individual is *a priori* more likely to choose, rather than noise from alternatives which are unlikely to be selected. As most institutions in the total set of institutions are not connected to an individual of interest, i.e., having a connection is a rare event, weighted sampling ensured more realistic choice sets (Ben-Akiva & Lerman, 1985). The marginal probability plots were produced by constructing a new prediction dataset using a sample of the individuals who were presented with the full choice set (all possible institutions). Note that the plotted marginal predicted relative probabilities should not be interpreted in absolute terms: They are meaningful only when interpreted relative to each other.

Competing Interests

The authors declare to have no competing interests.

Funding

This research was funded by the Max Planck Institute for Demographic Research (MPIDR) as part of the Population and Social Data Science Summer Incubator Program of 2024. This study has received access to the bibliometric data through the project 'Kompetenznetzwerk Bibliometrie' (Schmidt et al., 2025) and we acknowledge their funder, the German Bundesministerium für Bildung und Forschung (grant number 16WIK2101A).

Author Contributions

- Conceptualization: A.R., O.A., G.M., J.L., J.K., D.P., A.A.
- Data curation: AA
- Formal Analysis: A.R., O.A., G.M., J.L.
- Methodology: A.R., O.A., G.M., J.L., A.A., J.K., D.P
- Supervision: A.A., J.K., D.P.
- Visualization: A.R., O.A., G.M., J.L.
- Writing - original draft: A.R., O.A., G.M., J.L.
- Writing - review & editing: A.R., O.A., G.M., J.L., J.K., D.P., A.A.

Data availability

The raw and processed data are not available due to Scopus license terms. Interested researchers can reach out to Elsevier/Scopus for individual agreements on accessing this data. Data and scripts to replicate the figures in the manuscript will be made available in a public repository upon acceptance.

Acknowledgements

We thank Logan Walls, University of Michigan, and Benjamin Jarvis, Institute for Analytical Sociology, Linköping University, for extensive discussions on the specification and set-up of the discrete choice modelling. We furthermore thank Tom Theile, Max Planck Institute for Demographic Research, for support in downloading the raw bibliometric data. Additionally, we thank the participants of the Max Planck Institute for Demographic Research Population and Social Data Science Summer Incubator Program of 2024 for their comments on an earlier version of the results, as well as participants of the 'Bridging approaches in the sciences studying science' workshop held at the Cognition, Values and Behaviour Lab, Ludwig Maximilian University of Munich, participants of the Population Association of America's Annual Meeting (PAA 2025), the Annual Conference of the International Network of Analytical Sociology (INAS 2025), and participants of the International Conference on Computational Social Science (IC2S2 2025) for helpful comments and feedback.

References

Akbaritabar, A., Castro Torres, A. F., & Larivière, V. (2024). A global perspective on social stratification in science. *Humanities and Social Sciences Communications*, *11*(1), 914. https://doi.org/10.1057/s41599-024-03402-w

Akbaritabar, A., Dańko, M. J., Zhao, X., & Zagheni, E. (2025). Global subnational estimates of migration of scientists reveal large disparities in internal and international flows. *Proceedings of the National Academy of Sciences*, *122*(15), e2424521122. https://doi.org/10.1073/pnas.2424521122

Akbaritabar, A., Theile, T., & Zagheni, E. (2024). Bilateral flows and rates of international migration of scholars for 210 countries for the period 1998-2020. *Scientific Data*, *11*(1), 816.

Albarrán, P., Carrasco, R., & Ruiz-Castillo, J. (2017). Are migrants more productive than stayers? Some evidence from a set of highly productive academic economists. *Economic Inquiry*, *55*(3), 1308–1323.

Allison, P. D., & Long, J. S. (1990). Departmental Effects on Scientific Productivity. *American Sociological Review*, *55*(4), 469–478. https://doi.org/10.2307/2095801

Aman, V. (2020). Transfer of knowledge through international scientific mobility: Introduction of a network-based bibliometric approach to study different knowledge types. *Quantitative Science Studies*, *1*(2), 565–581.

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, *1*(1), 377–386.

Barber, V., & Blumstein, M. (2022). *On Finding and Choosing a Postdoc*. https://mitcommlab.mit.edu/cee/2022/06/27/on-finding-and-choosing-a-postdoc/?utm_source

Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press.

Biazoli, L., de Paula Fonseca, B., Çetinkaya-Rundel, M., Araújo, E., & de Oliveira, I. R. C. (2025). International mobility boosts scientific careers: A synthetic control analysis of Brazilian researchers. *Scientometrics*, *130*(5), 3029–3051. https://doi.org/10.1007/s11192-025-05327-w

Blumenstock, J. E., Chi, G., & Tan, X. (2023). Migration and the value of social networks. *The Review of Economic Studies*, rdad113. https://doi.org/10.1093/restud/rdad113

Bosetti, V., Cattaneo, C., & Verdolini, E. (2015). Migration of skilled workers and innovation: A European Perspective. *Journal of International Economics*, *96*(2), 311–322.

Bruch, E., & Swait, J. (2019). Choice Set Formation in Residential Mobility and Its Implications for Segregation Dynamics. *Demography*, *56*(5), 1665–1692. https://doi.org/10.1007/s13524-019-00810-5

Bruch, E. E., & Mare, R. D. (2012). Methodological Issues in the Analysis of Residential Preferences, Residential Mobility, and Neighborhood Change. *Sociological Methodology*, *42*(1), 103–154. https://doi.org/10.1177/0081175012444105

Burris, V. (2004). The Academic Caste System: Prestige Hierarchies in PhD Exchange Networks. *American Sociological Review*, *69*(2), 239–264. https://doi.org/10.1177/000312240406900205

Chou, M.-H., Wang, J., & Li, A. (2019). Collaboration patterns of mobile academics: The impact of international mobility. *Science and Public Policy*, *46*(3). https://doi.org/10.1093/scipol/scy073

Christakis, N. A., & Fowler, J. H. (2013). Social contagion theory: Examining dynamic social networks and human behavior. *Statistics in Medicine*, *32*(4), 556–577. https://doi.org/10.1002/sim.5408

Clauset, A., Arbesman, S., & Larremore, D. B. (2015). Systematic inequality and hierarchy in faculty hiring networks. *Science Advances*, *1*(1), e1400005. https://doi.org/10.1126/sciadv.1400005

European Comission, IDEA Consult, WIFO, & Technopolis. (2017). *MORE3 study: Support data collection and analysis concerning mobility patterns and career paths of researchers*. European Commission - Directorate-General for Research and Innovation European Research Area. LU. Retrieved July 11, 2025, from https://data.europa.eu/doi/10.2777/710643

Fernández, R. M., & Weinberg, N. (1997). Sifting and sorting: Personal contacts and hiring in a retail bank. *American Sociological Review*, *62*(6), 883. https://doi.org/10.2307/2657345

Ganguli, I. (2015). Immigration and Ideas: What Did Russian Scientists "Bring" to the United States? *Journal of Labor Economics*, *33*(S1), S257–S288. https://doi.org/10.1086/679741

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, *78*(6), 1360–1380. http://www.jstor.org/stable/2776392

Haines, V. A., Godley, J., & Hawe, P. (2011). Understanding interdisciplinary collaborations as social networks. *American Journal of Community Psychology*, *47*(1–2), 1–11. https://doi.org/10.1007/s10464-010-9374-1

Herz, A., & Altissimo, A. (2021). Understanding the structures of transnational youth im/mobility: A qualitative network analysis. *Global Networks*, *21*(3), 500–512. https://doi.org/10.1111/glob.12315

Holding, B. C., Acciai, C., Schneider, J. W., & Nielsen, M. W. (2024). Quantifying the mover's advantage: Transatlantic migration, employment prestige, and scientific performance. *Higher Education*, *87*(6), 1749–1767.

Jarvis, B. F. (2019). Estimating Multinomial Logit Models with Samples of Alternatives. *Sociological Methodology*, *49*(1), 341–348. https://doi.org/10.1177/0081175018793460

Jonkers, K., & Tijssen, R. (2008). Chinese researchers returning home: Impacts of international mobility on research collaboration and scientific productivity. *Scientometrics*, *77*(2), 309–333. https://doi.org/10.1007/s11192-007-1971-x

Kashyap, R., Rinderknecht, R. G., Akbaritabar, A., Alburez-Gutierrez, D., Gil-Clavel, S., Grow, A., Kim, J., Leasure, D. R., Lohmann, S., Negraia, D. V., Perrotta, D., Rampazzo, F., Tsai, C.-J., Verhagen, M. D., Zagheni, E., & Zhao, X. (2023, March 31). Chapter 3: Digital and computational demography. Retrieved February 18, 2025, from https://www.elgaronline.com/edcollchap/book/9781789906769/book-part-9781789906769-10.xml

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18. https://doi.org/10.1016/S0048-7333(96)00917-1

Kwiek, M. (2021). The prestige economy of higher education journals: A quantitative approach. *Higher Education*, *81*(3), 493–519. https://doi.org/10.1007/s10734-020-00553-y

Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation*, *11*(1), 3–15. https://doi.org/10.3152/147154402781776961

Lin, N. (1999). Building a network theory of social capital. *Connections*, *22*(1), 28–51.

Liu, Y., Huang, C., Zhang, F., & Ma, H. (2023). Locational preferences of high-level overseas talent returning to China: Evidence from microdata analysis. *Population, Space and Place*, *29*(5), e2650. https://doi.org/10.1002/psp.2650

Lu, X., & Zhang, W. (2015). The Reversed Brain Drain: A Mixed-method Study of the Reversed Migration of Chinese Overseas Scientists. *Science, Technology and Society*, *20*(3), 279–299. https://doi.org/10.1177/0971721815597127

Mahroum, S. (2000). Scientific mobility: An agent of scientific expansion and institutional empowerment. *Science Communication*, *21*(4), 367–378.

Massey, D. S. (1990). Social Structure, Household Strategies, and the Cumulative Causation of Migration. *Population Index*, *56*(1), 3–26. https://doi.org/10.2307/3644186

Massey, D. S., & Espinosa, K. E. (1997). What's Driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis. *American Journal of Sociology*, *102*(4), 939–999. https://doi.org/10.1086/231037

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*, 415–444. http://www.jstor.org/stable/2678628

Netz, N., Hampel, S., & Aman, V. (2020). What effects does international mobility have on scientists' careers? A systematic review. *Research Evaluation*, *29*(3), 327–351.

Nevin, A. D. (2019). Academic hiring networks and institutional prestige: A case study of canadian sociology. *Canadian Review of Sociology/Revue canadienne de sociologie*, *56*(3), 389–420. https://doi.org/10.1111/cars.12252

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Petersen, A. M. (2015). Quantifying the impact of weak, strong, and super ties in scientific careers. *Proceedings of the National Academy of Sciences*, *112*(34), E4671–E4680. https://doi.org/10.1073/pnas.1501444112

Podolny, J. M. (1993). A Status-Based Model of Market Competition. *American Journal of Sociology*, *98*(4), 829–872. https://doi.org/10.1086/230091

Rivera, L. A. (2012). Hiring as Cultural Matching: The Case of Elite Professional Service Firms. *American Sociological Review*, *77*(6), 999–1022. https://doi.org/10.1177/0003122412463213

Schaer, M. (2022a). Early-Career Academics' Transnational Moves: The Gendered Role of Vertical Social Ties in Obtaining Academic Positions Abroad. In E. Keskiner, M. Eve, & L. Ryan (Eds.), *Revisiting Migrant Networks: Migrants and their Descendants in Labour Markets* (pp. 205–226). Springer International Publishing. https://doi.org/10.1007/978-3-030-94972-3_10

Schaer, M. (2022b). From mobility attractiveness to mobility fatigue: The impact of repeated transnational mobility on the lives and aspirations of early-career academics. *Population, Space and Place*, *28*(5), e2536. https://doi.org/10.1002/psp.2536

Schmidt, M., Rimmert, C., Stephen, D., Lenke, C., Donner, P., Gärtner, S., Taubert, N., Bausenwein, T., & Stahlschmidt, S. (2025). The Data Infrastructure of the German Kompetenznetzwerk Bibliometrie: An Enabling Intermediary between Raw Data and Analysis. *Quantitative Science Studies*, 1–36. https://doi.org/10.1162/qss.a.20

Sîrbu, A., Andrienko, G., Andrienko, N., Boldrini, C., Conti, M., Giannotti, F., Guidotti, R., Bertoli, S., Kim, J., Muntean, C. I., Pappalardo, L., Passarella, A., Pedreschi, D., Pollacci, L., Pratesi, F., & Sharma, R. (2021). Human migration: The big data perspective. *International Journal of Data Science and Analytics*, *11*(4), 341–360. https://doi.org/10.1007/s41060-020-00213-5

Sugimoto, C. R., Robinson-Garcia, N., Murray, D. S., Yegros-Yegros, A., Costas, R., & Larivière, V. (2017). Scientists have most impact when they're free to move. *Nature*, *550*(7674), 29–31. https://doi.org/10.1038/550029a

Trippl, M. (2013). Scientific mobility and knowledge transfer at the interregional and intraregional level. *Regional studies*, *47*(10), 1653–1667.

Vaccario, G., Verginer, L., & Schweitzer, F. (2021). Reproducing scientists' mobility: A data-driven model. *Scientific Reports*, *11*(1), 10733. https://doi.org/10.1038/s41598-021-90281-9

Van Eck, N. J. (2024). CWTS leiden ranking open edition 2023 - results. *Zenodo*. https://doi.org/10.5281/zenodo.10587952

Venturini, S., Sikdar, S., Rinaldi, F., Tudisco, F., & Fortunato, S. (2024). Collaboration and topic switches in science. *Scientific Reports*, *14*(1), 1258.

Woolley, R., Turpin, T., Marceau, J., & Hill, S. (2008). Mobility Matters: Research Training and Network Building in Science. *Comparative Technology Transfer and Society*, *6*(3), 159–184. https://doi.org/10.1353/ctt.0.0014

Zhao, X., Akbaritabar, A., Kashyap, R., & Zagheni, E. (2023). A gender perspective on the global migration of scholars. *Proceedings of the National Academy of Sciences*, *120*(10). https://doi.org/10.1073/pnas.2214664120

Supplementary Material

SI-1 Descriptive view of the data

Table SI-1 summarizes key characteristics of the study sample comprising 171,746 mobile scholars identified using Scopus-based bibliometric data. On average, individuals had published approximately 10 papers (SD = 20.51) and collaborated with over 15 first-order coauthors and more than 700 first- and second-order coauthors before their first move. The mean academic age at migration was 5.9 years (SD = 3.86). Institutional prestige, measured via the Leiden Ranking[2], was broadly distributed, with average target and source institution rankings around 3,800. The sample is skewed towards male scholars (66.7%) and is largely concentrated in Physical Sciences (44.6%), followed by Life Sciences (24.4%) and Health Sciences (19.7%), which is similar to the overall composition of scholars across fields of science in Scopus data (Baas et al., 2020). Most mobility events were national (55.3%), although international moves were also common (44.7%). The sample also spans a wide range of academic cohorts from 1996 to 2020, allowing to capture temporal variation in mobility patterns. Strength measures indicate that approximately 47.2% of scholars moved along at least one individual-level tie, while 46.0% moved along an institutional tie. Notably, only 19.1% had strong first-order-only individual connections, and 22.7% moved along alternative institutional pathways.

SI-2 Decision Tree Results

We constructed a decision tree to analyze the direction of scholarly mobility, classifying each move into one of four categories: along individual ties, along institutional ties, along both, or along neither. The model incorporates individual-level features (e.g., gender, career age, and connectedness) as well as contextual characteristics (e.g., move type and institutional rankings) to predict the direction of a scholar's movement. Our primary aim with this model was predictive—to identify the most important features that influence mobility direction.

We trained the model using the `DecisionTreeClassifier` from Python's `sklearn.tree` package (Pedregosa et al., 2011). The decision tree offers a transparent and interpretable framework, revealing both the hierarchy and relative importance of input variables. Unlike more complex models, its structure allows for a direct understanding of how specific features influence the classification outcome, i.e., the type of the move.

The dataset was randomly split into a 70% training set and a 30% test set. The classifier was configured with `criterion= entropy`, `max_depth=10`, and `min_samples_leaf=10`, and class weights were balanced to mitigate class imbalance (Table SI-2). The model achieved a weighted F1-score between 0.66 and 0.74 across categories (Table SI-4), and an overall accuracy between 0.72 and 0.77 (Table SI-3), indicating a solid and balanced performance across the four mobility pathways.

Among all input features, the size of a scholar's second-order network—measured by the number of coauthors and co-coauthors—emerged as the most influential predictor, contributing approxi-

[2] https://www.leidenranking.com/

Variable	Overall (N=171,746)
Cumulative Number of Publications	9.97 (20.51)
Number of First Order Coauthors	15.36 (44.05)
Number of First and Second Order Coauthors	725.90 (2417.14)
Career Age (Years)	5.90 (3.86)
Target Institution Ranking	3,815.43 (3442.37)
Source Institution Ranking	3,908.10 (3307.17)
Individual connection strength, continuous	0.03 (0.08)
Institutional connection strength, continuous	0.01 (0.03)
Gender	
Female	57,185 (33.3%)
Male	114,561 (66.7%)
Academic Cohort	
1996-2000	24,514 (14.3%)
2001-2005	43,065 (25.1%)
2006-2010	59,687 (34.8%)
2011-2015	41,897 (24.4%)
2016-2020	2,583 (1.5%)
ASJC Classification (Top)	
Health Sciences	33,844 (19.7%)
Life Sciences	41,836 (24.4%)
Physical Sciences	76,665 (44.6%)
Social Sciences	19,109 (11.1%)
Unknown	292 (0.2%)
Move Type	
International move	76,814 (44.7%)
National move	94,932 (55.3%)
Strength Measures	
$S_{Fj}^{\text{ind}_{alt}}(T)$ Binary = 1	32,861 (19.1%)
$S_{Fj}^{\text{ind}}(T)$ Binary = 1	81,088 (47.2%)
$S_{ij}^{\text{inst}}(T)$ Binary = 1	79,080 (46.0%)
$S_{ij}^{\text{inst}_{alt}}(T)$ Binary = 1	38,985 (22.7%)

Table SI-1: Descriptive Statistics of the Study Sample

mately 40% to the model's decision-making process (Figure SI-1). This finding underscores the critical role of extended personal networks in shaping mobility patterns, suggesting that indirect connections (i.e., "friends-of-friends") may be particularly instrumental in facilitating moves.

The next most important features were the type of move (national vs. international) and the ranking of the target institution. The importance of move type supports the idea that international mobility often follows different mechanisms than domestic moves—potentially relying more on individual-level connections and less on institutional pipelines. Likewise, the influence of institutional rankings suggests that prestige considerations weigh heavily in scholars' decision-making processes, potentially guiding them toward highly ranked destinations when such opportunities are

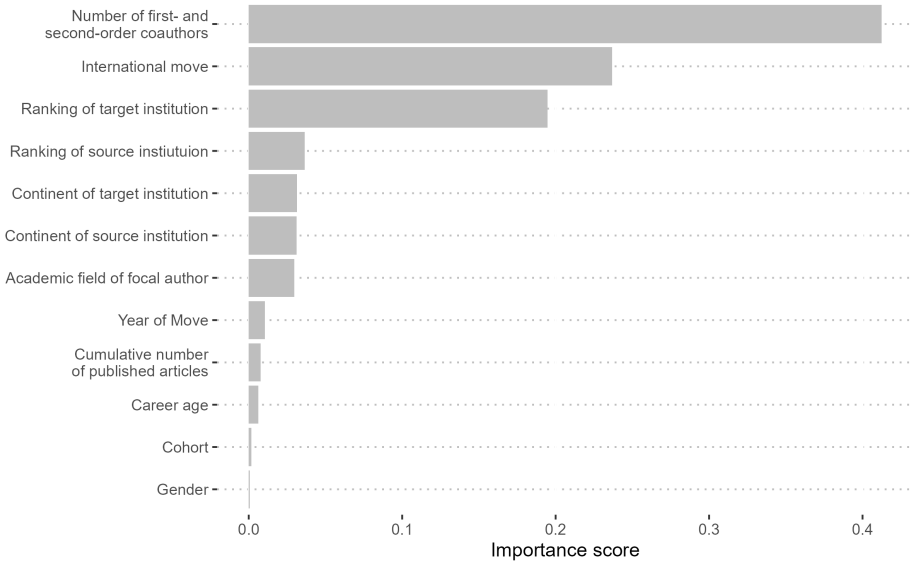

Fig. SI-1: **Decision tree.** The higher the importance score the more predictive a variable is for the type of mobility event observed (whether a move followed individual ties only, institutional ties only, both, or none of them).

available.

Parameter	Value
class_weight	balanced
criterion	entropy
max_depth	10
max_features	None
min_samples_leaf	10
min_samples_split	2

Table SI-2: Decision Tree Parameters

Direction	Accuracy
Following Individual Direction	0.72
Following Institutional Direction	0.74
Not Following Either Direction	0.74
Following Both Direction	0.77

Table SI-3: Accuracy of the decision tree in prediction of each direction

SI-3 Alternative operationalization of the strength measures

Acknowledging that past collaborations only serve as a limited proxy for individual and institutional connections (Katz & Martin, 1997; Laudel, 2002), we tested two additional measures to quantify individual and institutional connection strengths.

Direction	Precision	Recall	F1-score	Support
Following Individual Direction	0.70	0.75	0.72	24458
Following Institutional Direction	0.70	0.77	0.74	23760
Not Following Either Direction	0.64	0.67	0.65	18060
Following Both Direction	0.58	0.76	0.66	14723
Micro avg	0.66	0.74	0.70	81001
Macro avg	0.65	0.74	0.69	81001
Weighted avg	0.66	0.74	0.70	81001
Samples avg	0.65	0.69	0.63	81001

Table SI-4: Classification Report of the decision tree

SI-3.1 Alternative individual-level tie strength measure

For the individual level metric, we also evaluated our analysis when taking only first-order ties into account. The corresponding measure $S_{Fj}^{\mathrm{ind}_{alt}}(T)$, specified in Eq. SI-1, quantifies how strongly an individual F is connected to an institution j via their first-order co-authorship network in the year before a mobility event. It is defined as follows:

$$S_{Fj}^{\mathrm{ind}_{alt}}(T) = \frac{C_{Fj}^{\mathrm{1st}}(T)}{\sum_j C_{Fj}^{\mathrm{1st}}(T)} \tag{SI-1}$$

where C_{Fj}^{1st} represents the number of first-order co-authors that individual F has at institution j at time T and $\sum_j C_{Fj}^{\mathrm{1st}}$ is the total number of first-order co-authors that individual F has at the same time point T, the year before the mobility event, regardless of their institution.

SI-3.2 Alternative institutional-level tie strength measure

The alternative measure we tested for the strength of the institutional connection is based on the past mobility of scholars between institutions. Specifically, two institutions are considered linked if individuals migrated from one to the other in a given year. The mobility-based strength measure, $S_{ij}^{\mathrm{inst}_{alt}}(T)$, between institution i and institution j at a given year T takes the following form (Eq.SI-2):

$$S_{ij}^{\mathrm{inst}_{alt}}(T) = \sum_{t=T-5}^{T} \frac{M_{ij}(t)}{\sum_{j \neq i} M_{ij}(t)} \tag{SI-2}$$

where $Mij(t)$ is the total number of individuals migrating between the two institutions and $\sum_j M_{ij}(t)$ is the total number of individuals who migrated from or to institution i, excluded those who migrated from or to institution j.

SI-3.3 Comparison of measures

The two measures for the institutional level, where one is based on past collaborations and one on past mobility events, could be considered different proxies for the influence an institution has on the individual migration direction. For the main analysis, we took $S_{ij}^{\mathrm{Inst}_{alt}}$ as it covers more percentage of the individuals (Table SI-5).

	Individual-level connection strength		Institutional-level connection strength	
Strength measure	$S_{Fj}^{\text{ind}_{alt}}$	S_{Fj}^{ind}	S_{ij}^{inst}	$S_{ij}^{\text{inst}_{alt}}$
% of individuals	22.0%	45.0%	35.0%	18.7%

Table SI-5: Percentages of individuals who moved along a certain connection.

SI-4 Multinominal Logistic Regression Models

As a robustness measure, we fit our network direction-based multinomial logistic regression models with all possible measurement combinations. There are four models all together, combining the two main measure and two alternative measure. Our results show that while effect sizes vary between the different models, the significances and directions are consistent across them.

Table SI-6: Regression table - Multinominal logistic regressions - Individual direction measured with first-order coauthorship connections ($S_{Fj}^{\text{ind}_{alt}}$), institutional direction measured with aggregated past coauthorships between institutions (S_{Fj}^{inst})

	Dependent variable:		
	both	individual_direction	institutional_direction
	(1)	(2)	(3)
Gender (male == TRUE)	−0.027	−0.049**	−0.010
	(0.020)	(0.021)	(0.014)
National move (TRUE)	2.357***	−0.146***	2.403***
	(0.023)	(0.022)	(0.016)
Career Age	−0.008	−0.031***	−0.004
	(0.005)	(0.005)	(0.004)
Network size (Medium)	0.793***	0.679***	0.051***
	(0.028)	(0.027)	(0.016)
Network size (High)	1.607***	1.349***	0.070***
	(0.030)	(0.030)	(0.020)
Productivity (Medium)	0.109***	0.150***	−0.022
	(0.026)	(0.027)	(0.016)
Productivity (High)	0.508***	0.524***	−0.034*
	(0.029)	(0.030)	(0.021)
log(Ranking target institution)	0.972***	0.090***	0.962***
	(0.011)	(0.010)	(0.008)
log(Ranking source institution)	0.255***	−0.046***	0.427***
	(0.011)	(0.010)	(0.008)
Cohort fixed effects	Included	Included	Included
Period fixed effects	Included	Included	Included
Field fixed effects	Included	Included	Included
Geographic controls	Included	Included	Included
Constant	−10.089***	−1.696***	−9.938***
	(0.155)	(0.151)	(0.114)
Akaike Inf. Crit.	322,959.6	322,959.6	322,959.6

Notes: *p<0.1; **p<0.05; ***p<0.01. Cohort fixed effects include indicators for cohorts 2001-2005, 2006-2010, 2011-2015, and 2016-2020. Period fixed effects include indicators for move periods 2011-2015 and 2016-2020. Field fixed effects include indicators for Life Sciences, Physical Sciences, Social Sciences, and other fields. Geographic controls include indicators for current and previous continents (Asia, Australia, Europe, Europe/Asia, North America, and South America).

Table SI-7: Regression table - Multinominal logistic regressions - Individual direction measured with first-order coauthorship connections ($S_{Fj}^{\text{ind}_{alt}}$), institutional direction measured with aggregated past mobility of scholars between institutions ($S_{Fj}^{\text{inst}_{alt}}$)

	Dependent variable:		
	both	individual_direction	institutional_direction
	(1)	(2)	(3)
Gender (male == TRUE)	0.014	-0.035^{**}	0.032^{**}
	(0.025)	(0.016)	(0.016)
National move (TRUE)	2.742^{***}	-0.065^{***}	2.687^{***}
	(0.036)	(0.017)	(0.022)
Career age	-0.007	-0.025^{***}	-0.014^{***}
	(0.007)	(0.004)	(0.004)
Network size (Medium)	0.735^{***}	0.726^{***}	-0.019
	(0.036)	(0.023)	(0.019)
Network size (High)	1.528^{***}	1.457^{***}	-0.025
	(0.039)	(0.024)	(0.024)
Productivity (Medium)	0.150^{***}	0.129^{***}	-0.020
	(0.034)	(0.022)	(0.019)
Productivity (High)	0.412^{***}	0.541^{***}	-0.129^{***}
	(0.038)	(0.024)	(0.025)
log(Ranking target institution)	0.935^{***}	0.088^{***}	0.907^{***}
	(0.014)	(0.008)	(0.009)
log(Ranking source institution)	0.376^{***}	-0.083^{***}	0.573^{***}
	(0.014)	(0.008)	(0.009)
Cohort fixed effects	Included	Included	Included
Period fixed effects	Included	Included	Included
Field fixed effects	Included	Included	Included
Geographic controls	Included	Included	Included
Constant	-12.813^{***}	-1.408^{***}	-12.449^{***}
	(0.206)	(0.122)	(0.138)
Akaike Inf. Crit.	285,709.6	285,709.6	285,709.6

Notes: *p<0.1; **p<0.05; ***p<0.01. Cohort fixed effects include indicators for cohorts 2001-2005, 2006-2010, 2011-2015, and 2016-2020. Period fixed effects include indicators for move periods 2011-2015 and 2016-2020. Field fixed effects include indicators for Life Sciences, Physical Sciences, Social Sciences, and other fields. Geographic controls include indicators for current and previous continents (Asia, Australia, Europe, Europe/Asia, North America, and South America).

Table SI-8: Regression table - Multinominal logistic regressions - Individual direction measured with first- and second-order coauthorship connections (S_{Fj}^{ind}), institutional direction measured with aggregated past coauthorships between institutions (S_{Fj}^{inst})

	Dependent variable:		
	both	individual_direction	institutional_direction
	(1)	(2)	(3)
Gender (male == TRUE)	−0.050***	−0.047***	0.010
	(0.016)	(0.016)	(0.017)
National move (TRUE)	2.795***	0.357***	2.238***
	(0.019)	(0.017)	(0.019)
Career age	−0.020***	−0.022***	−0.009**
	(0.004)	(0.004)	(0.005)
Network size (Medium)	1.098***	0.948***	−0.174***
	(0.020)	(0.020)	(0.019)
Network size (High)	2.057***	1.877***	−0.424***
	(0.023)	(0.022)	(0.025)
Productivity (Medium)	0.182***	0.261***	−0.024
	(0.019)	(0.019)	(0.019)
Productivity (High)	0.527***	0.622***	−0.059**
	(0.022)	(0.021)	(0.024)
log(Ranking target institution)	1.182***	0.285***	0.885***
	(0.009)	(0.008)	(0.009)
log(Ranking source institution)	0.267***	−0.098***	0.452***
	(0.009)	(0.008)	(0.009)
Cohort fixed effects	Included	Included	Included
Period fixed effects	Included	Included	Included
Field fixed effects	Included	Included	Included
Geographic controls	Included	Included	Included
Constant	−10.968***	−2.317***	−9.748***
	(0.135)	(0.125)	(0.133)
Akaike Inf. Crit.	356,441.9	356,441.9	356,441.9

Notes: *p<0.1; **p<0.05; ***p<0.01. Cohort fixed effects include indicators for cohorts 2001-2005, 2006-2010, 2011-2015, and 2016-2020. Period fixed effects include indicators for move periods 2011-2015 and 2016-2020. Field fixed effects include indicators for Life Sciences, Physical Sciences, Social Sciences, and other fields. Geographic controls include indicators for current and previous continents (Asia, Australia, Europe, Europe/Asia, North America, and South America).

Table SI-9: Regression table - Multinominal logistic regressions - Individual direction measured with first- and second-order coauthorship connections ($S_{Fj}^{\text{ind}_{alt}}$), institutional direction measured with aggregated past mobility of scholars between institutions ($S_{Fj}^{\text{inst}_{alt}}$)

	Dependent variable:		
	both	individual_direction	institutional_direction
	(1)	(2)	(3)
Gender (male == TRUE)	−0.015	−0.056***	0.046**
	(0.019)	(0.014)	(0.022)
National move (TRUE)	3.254***	0.552***	2.490***
	(0.026)	(0.014)	(0.029)
Career age	−0.020***	−0.022***	−0.022***
	(0.005)	(0.004)	(0.006)
Network size (Medium)	1.115***	1.027***	−0.236***
	(0.024)	(0.016)	(0.024)
Network size (High)	2.062***	2.042***	−0.510***
	(0.027)	(0.019)	(0.034)
Productivity (Medium)	0.197***	0.247***	−0.024
	(0.022)	(0.016)	(0.024)
Productivity (High)	0.432***	0.625***	−0.063**
	(0.025)	(0.018)	(0.032)
log(Ranking target institution)	1.196***	0.342***	0.833***
	(0.011)	(0.007)	(0.013)
log(Ranking source institution)	0.420***	−0.110***	0.566***
	(0.011)	(0.007)	(0.013)
Cohort fixed effects	Included	Included	Included
Period fixed effects	Included	Included	Included
Field fixed effects	Included	Included	Included
Geographic controls	Included	Included	Included
Constant	−13.816***	−2.416***	−11.514***
	(0.163)	(0.107)	(0.171)
Akaike Inf. Crit.	320,691.3	320,691.3	320,691.3

Notes: *p<0.1; **p<0.05; ***p<0.01. Cohort fixed effects include indicators for cohorts 2001-2005, 2006-2010, 2011-2015, and 2016-2020. Period fixed effects include indicators for move periods 2011-2015 and 2016-2020. Field fixed effects include indicators for Life Sciences, Physical Sciences, Social Sciences, and other fields. Geographic controls include indicators for current and previous continents (Asia, Australia, Europe, Europe/Asia, North America, and South America).

SI-5 Discrete Choice Models

Regression table for the discrete choice model. We estimated three different models: (1) on the full data, (2) on male scholars only, and (3) on female scholars only. The latter two were performed to test for gender differences, but none were observed.

Table SI-10: Regression table

	Dependent variable:		
	choice	choice	choice
	(1) all	(2) men only	(3) women only
S_{Fj}^{ind} binary (TRUE)	-6.994^{***}	-6.191^{***}	-8.609^{***}
	(0.398)	(0.487)	(0.695)
$S_{ij}^{\text{inst}_{DCM}}$ binary (TRUE)	-1.128^{***}	-1.223^{***}	-0.911^{**}
	(0.247)	(0.298)	(0.440)
S_{Fj}^{ind} continuous	0.728^{***}	0.669^{***}	0.846^{***}
	(0.027)	(0.033)	(0.047)
$S_{ij}^{\text{inst}_{DCM}}$ continuous	0.480^{***}	0.477^{***}	0.488^{***}
	(0.018)	(0.022)	(0.033)
International move (TRUE)	0.569^{***}	0.581^{***}	0.524^{***}
	(0.053)	(0.064)	(0.097)
log(Ranking target institution)	-0.385^{***}	-0.421^{***}	-0.327^{***}
	(0.052)	(0.063)	(0.091)
Region (Asia)	-0.105^{***}	-0.095^{***}	-0.143^{***}
	(0.025)	(0.031)	(0.046)
Region (Australia)	0.170^{***}	0.109^{***}	0.288^{***}
	(0.028)	(0.034)	(0.049)
Region (Europe)	-0.245^{***}	-0.303^{***}	-0.129^{***}
	(0.025)	(0.031)	(0.045)
Region (North America)	0.208^{***}	0.160^{***}	0.304^{***}
	(0.025)	(0.031)	(0.046)
Region (South America)	-0.003	-0.106^{***}	0.172^{***}
	(0.030)	(0.037)	(0.052)
S_{Fj}^{ind} binary (TRUE):log(Ranking target institution)	0.101^{***}	0.112^{**}	0.086
	(0.038)	(0.046)	(0.066)
$S_{ij}^{\text{inst}_{DCM}}$ binary (TRUE):log(Ranking target institution)	-0.114^{***}	-0.093^{***}	-0.151^{***}
	(0.020)	(0.024)	(0.036)
S_{Fj}^{ind} continuous:log(Ranking target institution)	-0.027^{***}	-0.028^{***}	-0.024^{***}
	(0.003)	(0.003)	(0.005)
$S_{ij}^{\text{inst}_{DCM}}$ continuous:log(Ranking target institution)	-0.025^{***}	-0.025^{***}	-0.026^{***}
	(0.002)	(0.002)	(0.003)
S_{Fj}^{ind} binary (TRUE):log(Ranking source institution)	-0.034	-0.145^{***}	0.186^{**}
	(0.042)	(0.052)	(0.073)
$S_{ij}^{\text{inst}_{DCM}}$ binary (TRUE):log(Ranking source institution)	0.058^{**}	0.062^{**}	0.042
	(0.024)	(0.029)	(0.043)
S_{Fj}^{ind} continuous:log(Ranking source institution)	0.011^{***}	0.020^{***}	-0.006
	(0.003)	(0.004)	(0.005)
$S_{ij}^{\text{inst}_{DCM}}$ continuous:log(Ranking source institution)	0.006^{***}	0.005^{**}	0.008^{***}
	(0.002)	(0.002)	(0.003)
Observations	8,582,100	5,724,300	2,857,800
Log Likelihood	$-482{,}282.900$	$-323{,}538.200$	$-158{,}550.700$
LR Test (df = 19)	$615{,}618.600^{***}$	$401{,}570.700^{***}$	$214{,}435.800^{***}$

Note: *p<0.1; **p<0.05; ***p<0.01