On the Migration of Researchers across Scientific Domains

Soumajit Pramanik[1], **Surya Teja Gora**[1], **Ravi Sundaram**[2], **Niloy Ganguly**[1] and **Bivas Mitra**[1]
[1]Department of Computer Science & Engineering, IIT Kharagpur, India
[2]College of Computer and Information Science, Northeastern University, USA
soumajit.pramanik@cse.iitkgp.ernet.in, surya.cse.iitkgp@gmail.com, koods@ccs.neu.edu,
niloy@cse.iitkgp.ernet.in, bivas@cse.iitkgp.ernet.in

Abstract

Shift of research interest is an inherent part of a scientific career. Accordingly, researchers tend to migrate from one field of research to another. In this paper, we systematically study the publication records of more than $200,000$ researchers working in Computer Science domain and propose a simple algorithm to identify the migrating researchers. Just like human migration, this kind of research field migration is driven by various latent factors. Inspired by the classical theories of human migration, here we present a *theoretical framework* which models the decision-making processes of the individual migrating researchers and helps us to derive those latent factors. We further investigate the impact of these key factors in regulating a researcher's decision to migrate to a specific research field and observe the effect of such migration on her career. We note that in general publication quantity & quality, collaborator profile, fields' popularity contribute to a researcher's decision of field-migration. Importantly, effects of migration are not only limited to just one individual's career but also extend to the prospect of the research fields associated with it. Despite few initial capacity issues, field migration in general contribute in flourishing the research field people migrate into, in long term.

Introduction

Human migration is traditionally defined as the movement of people from one place to another with the intention of settling temporarily or permanently in the new location. A variety of theories have been developed for explaining the dynamics of human migration (Adams Jr and Page 2005; Sjaastad 1962; Todaro 1969). These theories can be broadly classified into two categories - (a) *micro-level* theories which focus on analyzing individual migration decisions, and (b) *macro-level* theories which concentrate on aggregated migration trends and explain these trends with macro-level explanations (Massey et al. 1993; Hagen-Zanker 2008).

Analogous to human migration (Massey and Zenteno 1999), scientific researchers migrate from one field of research to another, primarily driven by the desire to excel in their careers, which demands a continuous flow of high quality publications. It is observed that a large fraction of productive researchers do not restrict themselves into one or two research fields over their career life-span. Rather they migrate from one field of research to another as and when necessary. Given its broad impact on individual careers, it is an interesting research problem to investigate the shift in the research interests of individual researchers. Besides serving an individual researchers interest, more profoundly, migration phenomenon essentially enriches the entire scientific knowledge base by enhancing the flow of knowledge across multiple fields. Indeed it may be immensely useful to uncover the underlying dynamics behind the field migration of scientists, as it affects the ways in which people plan career, agencies fund research, researchers organize and discover knowledge, and governments recognize & reward excellence.

In this paper, we leverage on a large empirical dataset of scientific publications in computer science and attempt to study the local and the global dynamics regulating researchers' migration across scientific fields. We pose the following research questions - (i) Which are the motivating factors leading a researcher to migrate from one field to another? (ii) How does a migrating researcher select her field of research at different points of time in her career? (iii) What are the short-term and long-term impacts of field migration on her career as well as on the research fields she joins (or leaves)? Notably, classical *macro* and *micro* theories of human migration already provide succinct answers to similar questions raised in the context of human migration (Massey et al. 1993). Naturally, the question comes whether such theories can be effectively adopted in the context of scientific migration for addressing the aforementioned research questions. In this direction, our paper extends the *macro* and *micro* theories of human migration to the context of scientific migration in the following manner. We define (a) a *macro* analysis of scientific migration as dealing with the aggregated view of scientific migration, uncovering the broad reasons of large scale migrations of researchers and their impacts on affected fields of research, and (b) a *micro* analysis of scientific migration as investigating the motivation and decision-making procedures of individual migrating researcher as well as understanding the impact of migration on individual's scientific career (Section 'Macro and micro dynamics of migration').

We initially investigate certain macro-level features guiding migration, such as field choices and the phenomenon

of exodus from a field. In order to explain the researchers' choice of fields during migration, we group the similar fields into 'research domains' and examine the temporal flow of migrating researchers across them. Additionally, we discover the massive influx and outflux of researchers to (from) certain research fields within a short time span, which we designate as *mass migration*. Furthermore, we reveal the impact of such migrating behavior on the overall evolution of the corresponding research fields (Section 'Macro Dynamics').

We adopt a classical microeconomic model of human migration in our context (Massey et al. 1993; Borjas 1990), whereby a set of potentially motivating factors such as publication quantity & quality, collaborators, field's popularity are identified for investigation. Subsequently, we thoroughly analyze the contribution of each of these motivating factors to a researcher's decision of field migration, along with their statistical significance[1]. Moreover, we reveal the impact of such field migration on the researchers' publication & citation profiles. In general, migration is observed to positively influence the career of the researchers, however, collaborating with the prominent researchers of the joining field is essential to substantially improve her quality of publication in the new field, after migration. Finally, leveraging on the aforesaid features, we develop a prediction model to estimate the propensity of a researcher to migrate to a specific research field in near future; this model further confirms the adaptability of the microeconomic theories in the context of scientific migration (Section 'Micro Dynamics').

Related Works

To the best of our knowledge, understanding of how scientists choose and shift their research focus over time remains pretty scant. During the 80's, there was a surge in the research on field mobility and field migration (Vlachý 1981; Van Houten et al. 1983; Hargens 1986; Mulkay 1974).In those endeavours, field mobility has been discussed as the driving force for the exploration of new territories in the landscape of science (Urban 1982). Most of the experiments relied on the personal interviews and surveys to trace academic careers (primarily in physics domain); evidently those approaches were mostly restricted to small case studies and could not be generalized & scaled (Van Houten et al. 1983).

However, with the availability of large scale data sources such as Scopus, PubMed, Google Scholar, Microsoft Academic etc, the domain of 'Science of Science' (*SciSci*) has observed an enormous growth in recent times (Fortunato et al. 2018; Qazvinian and Radev 2009; Pramanik, Yerra, and Mitra 2015). This literature encompasses all the studies leading to a quantitative understanding of the genesis of scientific discovery, creativity, and practice and developing tools and policies aimed at accelerating scientific progress. In the domain of 'Science of Science', the term 'scientific mobility' is typically perceived as movement of researchers across countries or universities (Deville et al. 2014; Franzoni, Scellato, and Stephan 2014). Hence, most of these mobility studies have focused on quantifying the brain drain and intellectual gain of a country/region (Van Noorden 2012; Arrieta, Pammolli, and Petersen 2017). For instance, in (Franzoni, Scellato, and Stephan 2014) Franzoni et al. showed that migrant scientists exhibit higher productivity compared to domestic scientists, irrespective of their prior experience of international mobility.

In the gamut of *SciSci*, attempts have been made in bits and pieces on investigating research field selection and career diversification of researchers. One school focused on developing models to mimic the notion of field selection process of researchers (Chakraborty et al. 2015; Jia, Wang, and Szymanski 2017; Braun 2012). For instance, Jia et. al. (Jia, Wang, and Szymanski 2017) aimed to model the research interest evolution of scientific researchers using a simple random walk. Moreover, macro level studies have been performed on the authors performing research in multiple fields simultaneously in their career. For instance, Abramo et al. (Abramo, DAngelo, and Di Costa 2018) showed that a scientist's outputs resulting from research diversification are more often the result of collaborations with multidisciplinary teams.

In a nutshell, we highlight the following major limitations of the state of the art literature in the context of scientific migration. (a) First of all, literature failed to study the temporal events of field migration of the researchers. Precisely, they overlooked the conditions (motivating factors) under which an author decides to migrate and the immediate effects of such migration on her own career and collectively, on the research field. (b) Secondly, none of these works aimed to borrow the concepts from classical human migration theories and explained the field migration of scientific researchers. Our paper takes an important step towards this direction.

Dataset and Migrator Identification

First, we introduce the empirical dataset of scientific publications, which we use to study the dynamics of research migrations. Next, we define the behavior of migrating authors and propose a simple algorithm for identifying them. Finally, we exhibit few salient characteristics of the migrating authors.

Data description

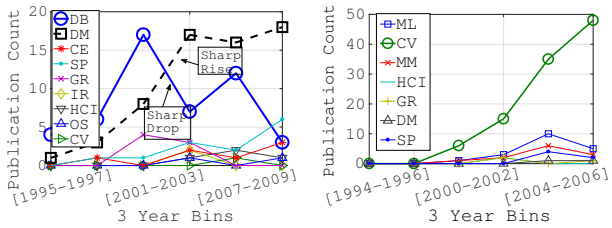
In this paper, we consider *'MAS'*, a large corpus of publications (published during 1960-2010) in Computer Science, from the Microsoft Academic Search portal[2]. This dataset contains around 4 million $(3,787,483)$ distinct papers contributed by nearly 3 million $(2,951,394)$ researchers and distributed over a set of distinct research fields of computer science (see Table. 1). Moreover, each paper comes along with various bibliographic information - the title of the paper along with the abstract and keywords, the list of author(s), the year of publication, the publication venue, references, citation contexts and the related field(s) of the paper. Apart from its massive size, another unique advantage provided by this dataset is that a unique identity is associated with each author, paper and publication venue. We conduct

[1]Notably, our empirical conclusions mostly rely on the correlations observed across different factors.

[2]http://academic.research.microsoft.com

Domain	Fields	Acronyms	% of Papers
C_1	Bioinformatics and Computational Biology	BCB	4.14
C_2	Computer Vision	CV	4.65
	Machine Learning and Pattern Recognition	ML	3.67
	Graphics	GR	3.65
C_3	Natural Language Processing	NLP	0.42
C_4	Hardware and Architecture	HA	9.23
	Real Time and Embedded Systems	RTE	7.39
C_5	Data Mining	DM	3.96
	Information Retrieval	IR	0.03
	Databases	DB	7.05
	Computer Education	CE	4.42
	World Wide Web	WWW	0.85
C_6	Human Computer Interaction	HCI	2.37
C_7	Software Engineering	SE	5.55
	Programming Languages	PL	1.19
C_8	Operating Systems	OS	15.60
	Distributed & Parallel Computing	DPC	0.40
	Security and Privacy	SP	3.59
C_9	Artificial Intelligence	AI	6.52
	Simulation	SM	1.58
C_{10}	Multimedia	MM	3.57
	Networks and Communications	NC	8.51
C_{11}	Scientific Computing	SC	1.65

Table 1: Field-wise paper count distribution and corresponding acronyms. Similar fields are further grouped into 'domains'. The grouping procedure is mentioned in detail in Section 'Macro Dynamics'.



(a) Migrator example. (b) Non-migrator example.

Figure 1: Example of a migrator and a non-migrator. Each color represents publication counts in a field.

a simple and straightforward preprocessing of the dataset to make it suitable for migration analysis. First, we eliminate all the researchers who published in only one research field throughout her career (since no migration is possible for single field authors). Additionally, we also drop the authors who remained active in research for less than 10 years[3]. Finally, we are left with $299,544$ authors, which we refer as 'Filtered' dataset in the rest of this paper. The experiments conducted in this paper are mostly concentrated on this 'Filtered' dataset.

Defining migration

We define a 'migrating author' as a researcher who migrates from one research field to another at some point of her career. Precisely, for a migrator, we observe a significant drop in publication in one field, and a subsequent surge in another field. For instance, Fig. 1(a) demonstrates the publication profile of a migrating author; in year 2000-2001 her publication drops in the 'Databases (DB)' field, side by side a rise can be observed in 'Data Mining (DM)'. Hence, this author can be characterized as a 'migrator'. On the contrary, Fig. 1(b) shows the profile of a 'non-migrator' who consistently publishes in the fields she has chosen since the beginning of her career (although occasionally publishing in allied fields).

[3]We remove people with shorter career spans from our list as we believe in order to discover stable patterns, we need to consider people who have embraced research as their long term career.

Detection of migrating authors

In this section, we propose an algorithm for automatically detecting the migrators. This algorithm revolves around the index I_{MIG}, which identifies if an author migrates from one field to another in her research career. First we divide the author's career into non-overlapping $w = 3$ year windows[4] where the collection of windows $W = \{w_1, w_2, ..., w_n\}$ define the entire career of the author. I_{MIG} evaluates positive for an author if there exists at least one pair of consecutive time windows $[w_i, w_{i+1}]$ where (a) the top publishing field (say f_a) in the previous time window (w_i) drops and becomes a non-top publishing field in the subsequent window (w_{i+1}) and (b) a non-top publishing field (f_b) in the previous time window (w_i) rises to become the new top publishing field in the next window (w_{i+1}). The steeper the slopes of this fall and rise, higher the index indicating clearly observable migration behavior. We define an index $I_{MIG}(w_i, w_{i+1})$ [5] as the product of these two slopes to identify migration between the window pair w_i and w_{i+1}. We precisely detect a field migration between time window $[w_i, w_{i+1}]$ if $I_{MIG}(w_i, w_{i+1}) > 0$; this indicates the migration from the field f_a to f_b. We extend this expression to identify field migration of an author in her career as

$$I_{MIG} = \max_{\forall i \in \{1,2,...,|W|-1\}} \left(I_{MIG}(w_i, w_{i+1}) \right) \quad (3)$$

[4]We experimented with different window sizes. Three year time window provides us a perfect trade-off between the data sparsity and information loss.

[5]

$$I_{MIG}(w_i, w_{i+1}) = \max_{\forall f_a, f_b \in F, f_a \neq f_b} \Bigg(\underbrace{\left(\frac{Pub_{f_a}^{w_i} - Pub_{f_a}^{w_{i+1}}}{Pub_{f_a}^{w_i}} \times \mathbf{1}_{Top(w_i)}(f_a) \times (1 - \mathbf{1}_{Top(w_{i+1})}(f_a)) \right)}_{F_{Drop}} \times \underbrace{\left(\frac{Pub_{f_b}^{w_{i+1}} - Pub_{f_b}^{w_i}}{Pub_{f_b}^{w_{i+1}}} \times \mathbf{1}_{Top(w_{i+1})}(f_b) \times (1 - \mathbf{1}_{Top(w_i)}(f_b)) \right)}_{F_{Rise}} \Bigg) \quad (1)$$

where F is the set of all fields, $Top(w_i)$ is the top-publishing fields of an author in window w_i and

$$\mathbf{1}_{Top(w_i)}(f) = \begin{cases} 1 & if\ f \in Top(w_i) \\ 0 & otherwise \end{cases} \quad (2)$$

is the indicator function. $Pub_{f_a}^{w_i}$ is the number of articles the author has published in field f_a during window w_i. Evidently, the factor F_{Drop} indicates the fall of the top publishing field f_a in window w_{i+1} and the factor F_{Rise} indicates the rise of the field f_b in the window w_{i+1}.

Finally, we consider the author profile (say Bob) as input. $I_{MIG}(Bob) > 0$ identifies author Bob as a migrator.

Validation

We apply the proposed migrator detection algorithm on all the $299,544$ authors from the 'Filtered' dataset. Following the algorithm, we obtain $66,008$ migrators and $233,536$ non migrators. Notably, all non-migrators have I_{MIG} value 0.0 due to the presence of indicator function in the I_{MIG} definition. The mean I_{MIG} value of the migrators is found to be 0.79 (with standard deviation 0.25), distant from that of the non-migrators (0.0) and with $5.5 * 10^{-9}$ p-value ('two-sample t' test) indicating perfect separation between them.

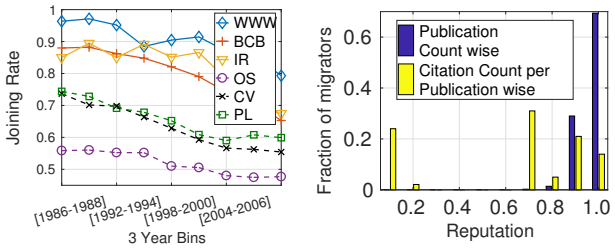
In order to further substantiate the correctness of migrator - non-migrator segregation, we randomly choose 300 authors (68 migrators) from the $299,544$ 'Filtered' authors and manually annotate them based on their publication profiles. Precisely, we employ three annotators (two males and a female - all university graduates and aged between 25-30) in this experiment and annotate each author following majority voting. The annotators have been asked to examine the career and publication profile of each author carefully and label them accordingly. The annotation results in exactly 68 authors as migrators (same as algorithm result) and rest as non migrators (with high inter-annotator agreement; $\kappa = 0.83$). This indicates statistically significant separation results in a near-perfect segregation.

First glimpse

Preliminary analysis on the 'Filtered' dataset reveals several interesting observations regarding migration behavior - two of which are reported below.

Ranking fields according to migration rate: We rank the research fields based on the (i) joining rate and (ii) departing rate of the migrating authors in a specific time window w. In Fig. 2(a), we illustrate the rank of fields based on their joining rates, over a time period. Interestingly, the fields ranked top are relatively new and trendy topics such as 'World Wide Web', 'Information Retrieval', 'Bioinformatics and Computational Biology' etc. On the other side, fields ranked bottom are relatively classical fields such as 'Operating Systems', 'Programming Languages', 'Computer Vision' etc. Similarly, we rank the fields based on their departing rates and observe that the trendy fields like 'Information Retrieval', 'Distributed & Parallel Computing', 'World Wide Web' again appear on top whereas the same three traditional fields fall at the bottom (not shown here). Evidently, the upcoming and trendy fields demonstrate highly dynamic & migrating behavior, whereas classical fields exhibit saturation.

Reputation of migrating authors: In this experiment, we aim to identify the reputation of the migrating authors in their respective fields before migration. We measure the reputation T_u^w of a migrating author u as the percentile of the author with respect to the total population (N^w) in the migrating window w, hence $T_u^w = 1 - \frac{R_u^w}{N^w}$ where R_u^w is the rank of the author in the population. We characterize the rank R_u^w of the author u from two different perspectives (a) publication count (b) citation count per publication (Vanclay and Bornmann 2012; Raj and Zainab 2012; Duffy et al. 2008). In Fig. 2(b), we show the distribution of reputation of the migrating (departing) authors from 'Networks and Communications' field (similar patterns have been observed for other fields too). We observe that almost all the migrating (departing) authors are mostly the top (20%) reputed researchers of that field. Interestingly, albeit in general most of the migrating authors exhibit high reputation in terms of publication count, however few of them show poor citation count per publication (bottom 20%), depicting their low impact in that field.



(a) Top and bottom ranked fields based on joining rate during different time windows.

(b) Reputations of migrating researchers leaving 'Networks and Communications' field based on publication count and citation count per publication.

Figure 2: Preliminary analysis of the 'Filtered' dataset.

Macro and micro dynamics of migration

In the context of human migration, two different dynamics, called (a) *micro* and (b) *macro* have been studied in the literature (Richmond 1988). In this section, we present an overview of these dynamics and subsequently adapt them in the context of field migration of the researchers.

Micro dynamics: In human migration literature, the *micro* level analysis studies the socio-psychological factors discriminating migrants from non-migrants, as well as developing models explaining the motivation behind migration, decision-making and satisfaction of the migrants (Sjaastad 1962; Todaro 1969).For instance, (Massey et al. 1993; Borjas 1990) proposed a microeconomic model for estimating the net (say, monetary) benefit of migration to a different place, which may help the human migrator to take the decision of migration in the light of cost-benefit tradeoff. This microeconomic model can be analytically represented as (Massey et al. 1993; Borjas 1990),

$$ER(0) = \int_0^n [P_1(t)P_2(t)Y_d(t) - P_3(t)Y_o(t)]e^{-rt}dt - C(0) \quad (4)$$

where interpretation of individual parameters have been specified in Table. 2 (and r is a discount factor). We borrow the clues from this model and adapt it in the context of scientific migration of the researchers. In the Table. 2, we highlight the plausible factors playing role behind the field migration of individual researcher. In the following, the 'Micro Dynamics' section extensively investigates the suitability of this model to explain the micro dynamics of the migrating researcher.

Parameter	Human Migration	Scientific Migration
$ER(0)$	Expected net monetary return	Expected career benefit
$P_1(t)$	Probability of avoiding deportation from the destination	Probability of getting accepted in the destination field
$P_3(t), P_2(t)$	Probability of employment at the source and destination	Popularity of the source and destination fields
$Y_o(t), Y_d(t)$	Earnings if employed in the source and destination	Performance of the researcher in the source and destination fields
$C(0)$	Total costs of movement (including psychological costs)	Distance between the source and destination fields
t	Time	Phase of career

Table 2: Conceptual mapping the parameters of human migration into the context of scientific migration.

Macro dynamics: The macro level analysis of human migration focuses on the migration streams, identifying those conditions under which large-scale movements occur and describing the demographic, economic and social characteristics of the migrants in aggregate terms. For instance, Adams et al. (Adams Jr and Page 2005) advocate that the international migration and corresponding remittances have a strong impact on reducing poverty in the developing world (source countries). Side by side, volume of human migration critically impacts the infrastructure & economy of the destination places (developed countries) however, brings the diversity in the respective society. In the similar vein, in the following section, we shed some light on the macro dynamics of the scientific migration. Precisely, we concentrate on the (i) effects of migration on the departing & the joining fields, (ii) demonstrate the role of individual research fields behind migration and (iii) show the volume of migrating researchers across different fields. The detail follows.

Macro Dynamics

Analysis of macro dynamics investigates the following - (a) role of research fields on migration (b) volume of the migrating researchers including large-scale movement (mass migration) across fields (c) short-term and long-term effects of migration on the departing & joining fields.

(A) Migration dynamics across research fields

The similarity & overlap between multiple research fields may instigate researchers to migrate from one field to another close field[6]. In order to identify the similar and dissimilar fields, we first introduce the notion of proximity between different research fields and subsequently detect *domains*, which are the group of fields close to each other.

Field proximity: discovery of domains We estimate the proximity between two fields f_i and f_j as the total number of mutual citations between them (this reflects the distance between two fields $C(0)$ as shown in Table. 2); we denote two fields as similar if they have high number of mutual citations.

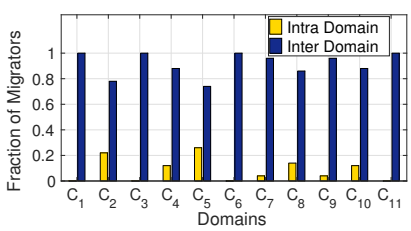
Consider $F = \{f_1, f_2, ..., f_N\}$ be the set of all research fields in our dataset. We construct a field graph $G = \{V, E\}$ where $V = \{v_i : \forall f_i \in F\}$ is the set of vertices and $E = \{(v_i, v_j, d_{ij}) : v_i \in V, v_j \in V \ \& \ d_{ij} \in [0, 1]\}$ is the set of weighted directed edges. Each vertex v_i corresponds to one research field f_i and each directed edge (v_i, v_j, d_{ij}) denotes the aggregated citation links from the articles in field v_j to articles in v_i. The edge weight d_{ij} denotes the fraction of citations coming from the field f_j to the field f_i. We apply directed Louvain (Dugué and Perez 2015) community detection algorithm on G to obtain a cover $\mathcal{C} = \{C_1, C_2, ..., C_m\}$ of the set of fields V. Each of these C_i denotes a community of research fields highly citing each other; we refer each community C_i as a *Domain*. Precisely, we discover the eleven research domains $(C_1, C_2, \ldots C_{11})$ shown in Table. 1.

Concurring with our intuition, the fields within each detected domain are found to be deeply related with each other. For instance, all three fields in domain C_8 are closely connected with computer systems; similarly, in case of C_{10}, knowledge in 'Networks and Communications (NC)' is essential for developing online multimedia systems.

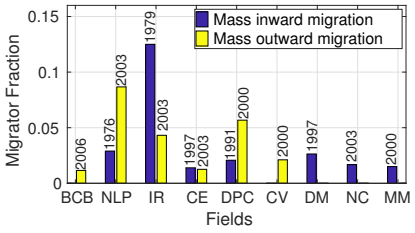
Domain Migration Based upon the domain classification, we dissect the migration dynamics of the researchers. General intuition expects the migration of researchers across the fields within a single domain; this expectation stems from the expertise overlap between different fields of a domain. However, in Fig. 3(a), the empirical observation appears counter intuitive. We observe that a major fraction of authors migrate across different domains. Hence, we assert that if a researcher decides to migrate, she prefers to come out from her comfort zone, develop new expertise and migrate to a different research domain. However, we observe that domains C_2={Computer Vision, Machine Learning and Pattern Recognition, Graphics} and C_5={Data Mining, Information Retrieval, Databases, Computer Education, World Wide Web} are the only two typical domains where a significant fraction (just above 20%) of authors migrate within a single domain (intra-domain migration).

In order to further understand the nature of this cross-domain migration, we investigate whether there is a preference of authors from a certain home domain (say) C_{10} to migrate to a target domain C_8. For that we build up a bipartite graph considering the home and target domains and then run a community detection algorithm (Barber 2007) to discover the penchant. We identify certain communities whereby scientists working in a group of domains have a tendency to move to another target group of domains. For example, we identify **(a)** Domains C_4, C_{10} in the home partition, group with C_3, C_8 domains in the target partition, **(b)** Similarly domains C_2, C_9, C_{10} group with C_1, C_2 domains and **(c)** Domains C_1, C_7, C_8 group with C_4, C_7, C_9, C_{11} domains. We observe that there is a back and forth movement from applied fields (eg. MM) to systems field (eg. OS, architecture). We believe this is also driven by the requirement

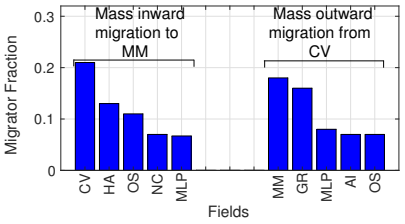
[6]In natural migration, distance is often considered as a cost of movement while determining the propensity of migration from origin to destination (Massey et al. 1993).



(a) Fraction of migrators migrating within same (intra) and to different (inter) domains.



(b) Fraction of authors migrated in and out of the fields during *mass migration*. The year labeled in each bar indicates the starting year of the time-window exhibiting *mass inward/outward migration* in/to the corresponding field.



(c) The distribution of contributions from different fields in the *mass inward migration* of 'Multimedia (MM)' field and the distribution of contributions towards different fields in the *mass outward migration* of 'Computer Vision (CV)' field during (2000, 2002).

Figure 3: (a) Migration across domains, (b), (c) Mass migration

of the field - for example, development of efficient multimedia algorithms need more understanding of the computer operating systems etc. Hence there is a tendency to become *MM-specialized OS researchers*.

(B) Mass migration

We further delve deep the dynamics of field migration. We define *mass migration*[7] phenomenon as a massive influx (or outflux) of authors to (or from) a research field within a relatively short period of time. In order to identify the fields exhibiting mass migration, first we compute $in^g_{w_i}$ (and $out^g_{w_i}$), the fraction of researchers migrating to (from) a research field g in each time window w_i. We designate a field g exhibiting mass *inward* migration in time window(s) w_i if the joining population to field g in window(s) w_i is substantially higher than the rest of the windows. Precisely, we detect mass *inward* migration by identifying the window(s) w_i exhibiting $in^g_{w_i} > \mu_{in^g} + 3\sigma_{in^g}$ (following the principle of outlier detection (Miller 1991; Grubbs 1969)), where μ_{in^g} & σ_{in^g} are respectively the mean and standard deviation of the fraction of joining migrators to field g across all the time windows. Similarly, we detect a field g exhibiting mass *outward* migration in time window(s) w_i if we identify window(s) w_i showing $out^g_{w_i} > \mu_{out^g} + 3\sigma_{out^g}$, where μ_{out^g} & σ_{out^g} are respectively the mean and standard deviation of the fraction of departing migrators from field g across all time windows.

Our experiments identify the mass (inward and outward) migrations in fields shown in Fig. 3(b); mass migration is a rare event as one field experience it only once in the entire time period. In Fig. 3(b), we show the fraction of authors joining (and departing) during the mass migration; notably, there can be a surge (or drop) of $8\% - 10\%$ of researchers in the mass migration time window. Interestingly, Fig. 3(b) clinically demarcates the time-periods when a particular field becomes highly trendy (attracting researchers) or stagnant (repelling researchers) - for instance, the 'Distributed & Parallel Computing (DPC)' field started to gain popularity around 1991 causing mass inward migration whereas it observes mass outward migration in the year 2000-2001.

[7]In many of the scenarios, mass immigration across countries are observed in human race (Massey and Zenteno 1999).

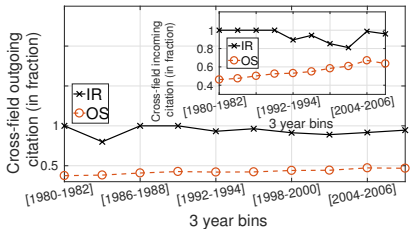
One step further, we attempt to examine whether a mass outward migration from one field f may result in mass inward migration to another field g in same time window. Albeit rare, we observe such a scenario for the pair of fields 'Computer Vision (CV)' and 'Multimedia (MM)' experiencing mass outward and inward migrations respectively in year $[2000 - 2002]$. Fig. 3(c) clearly depicts that majority of authors migrating out from field CV during $[2000 - 2002]$ joined MM contributing to *mass inward migration*.

(C) Effects of migration on fields

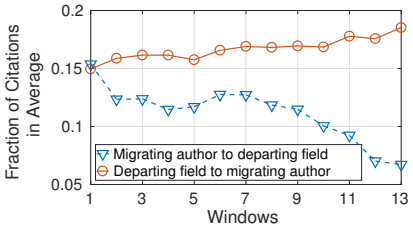
Migration of the researchers can directly effect the two fields closely associated with this migration event - (a) the departing field from which she leaves (b) the joining field to which she migrates. Hence, in the following, we examine how the migration of researchers affect the corresponding departing and joining fields in short and long time span. In order to perform few anecdotal experiments, we handpick Information Retrieval (IR) and Operating Systems (OS) as the top and bottom fields respectively in terms of joining (& departing) rate of the migrators (see Fig. 2(a))

Effects on departing field **(a) Gain in impact:** Impact of a research field is measured as the fraction of incoming citations it receives *only* from the other fields, denoted as *cross field incoming citations*. Interestingly, in the inset of Fig. 4(a), we show that the top fields based on departing rate (say IR) always enjoy a high impact. Delving deep, we observe that as more researchers start departing a field and join new fields, they keep on citing articles published in their previous fields. This *increases* the cross field incoming citations as well as improves the overall impact of the departing fields. However, this is a *short term gain* as the citations made by migrated researchers, towards the departing fields drop over time (see Fig. 4(b)).

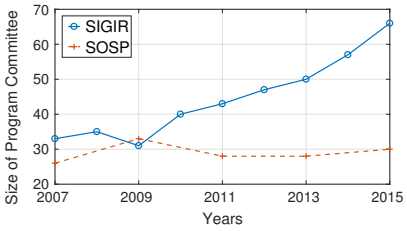
Effects on joining field **(a) Loss - building up pressure on the field's ecosystem:** Continuous influx of researchers in certain research fields build up pressure in the ecosystem of that field. For instance, continuous migration into a research field in turn significantly increases the volume of submitted articles, which immediately creates a huge constraint on the peer-review based publication infrastructure. As an



(a) Correlation of joining and leaving rates with the fraction of cross-field outgoing and incoming citations.



(b) Avg. number of citations towards departing field from the migrating authors and vice versa after migration.



(c) Size of Program Committee of 'SIGIR' and 'SOSP' during 2007-2015.

Figure 4: Effects of migration- (a) Gain of fields with high joining and leaving rates (b) Gain of departing field (c) Pressure on joining field's infrastructure as indicated by the Program Committee sizes of top conferences.

anecdote, we handpicked 'SIGIR' and 'SOSP', two top-tier conferences from the fields of 'Information Retrieval' and 'Operating Systems' respectively and examined the size of their respective Program Committees. Fig. 4(c) illustrates that over the years, the size of 'SIGIR Program Committee' has been grown almost at a double rate than the 'SOSP Program Committee'[8] to cope up with the volume of submitted articles. Naturally, in such a scenario, there is always a possibility of compromising with the quality of the accepted papers. Substantiating this claim, we observe that only 11% of the researches are able to publish articles multiple times in the top five venues of 'Information Retrieval'. On the contrary, researchers working in 'Operating Systems' exhibit better consistency; 25% of them are able to publish articles multiple times in top five Operating Systems venues.

(b) Gain - joining researchers improve diversity of a field: 'Diversity' of a research field is measured as the fraction of *outgoing citations* (citing papers of other fields) of the articles published in that field. High 'diversity' essentially indicates that the field is getting enriched by incorporating ideas and techniques from various allied domains. Fig. 4(a) depicts the fact that the research fields with higher joining rate displays high 'diversity'. Analogous to the preceding argument, we observe that as new authors join a field, they keep on citing articles published in their previous fields, hence, increasing the *cross field outgoing citations* (aka 'diversity') of the newly joined field.

(c) Gain - joining researchers improve impact of a field: Interestingly, as a migrating author starts working in a new field, her past collaborators and followers (top citers), who still continue to work in the departing field, grow interest and start citing her articles published in the new field. This indirectly increases the 'impact' of the joining field, albeit at a slower rate. In Fig. 4(b), we observe this phenomenon to get intense over time, indicating a *long term* gain of the joining field.

Summarizing, analysis of macro dynamics reveals that (a) Researchers mostly perform cross-domain migration. However, there are certain domains like Data Mining, Information Retrieval, Databases where steady in-domain migration is also observed. (b) At certain point of times, mass scale movement of researchers indeed occur from one field to another. (c) Amidst migration, the departing field gets benefited by the incoming citations from the migrators for a short duration. (d) In long term, the joining field receives increasing number of incoming citations from the migrators' past collaborators and followers still working in the departing field; however, too much migration might disrupt the ecosystem of the field.

[8]SIGIR (organized yearly) has slightly different organization than SOSP (organized biyearly). For SIGIR, we use the Senior PC Committee (SPC) as its PC contains any and all reviewers (avg. size 291); For SOSP, we have Program Committee (PC) and a separate set of external reviewers, hence used the PC for SOSP.

Micro Dynamics

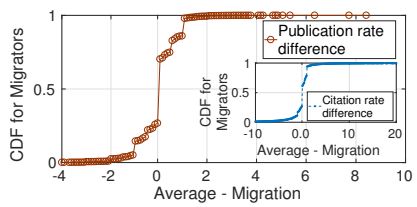
In this section, we perform a *micro* level analysis of the scientific migration following the microeconomic model introduced in Eq. 4 and the respective adaptation showed in Table. 2. We leverage on those identified factors and delve deep to investigate their role in motivating the researcher for the field migration. We classify those motivating factors into two categories (a) researcher's individual research profile (b) researcher's current working fields. Subsequently, we demonstrate the effects of the field migration on the career of the researcher. Finally, leveraging on the features discovered from the Eq. 4, we develop a prediction model to estimate the propensity of a researcher to migrate to a specific research field; this shows the adaptability of the microeconomic model in the context of scientific migration.

(A) Motivating Factors behind Migration

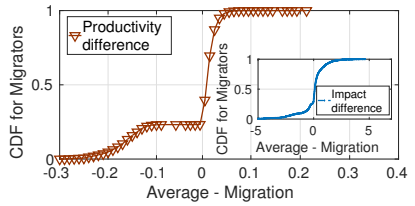
Motivation behind the field migration stems from the (a) researcher's own research profile & (b) profile of her current research fields.

(a) Researcher's individual research profile We consider that an author u migrates from the field f in time window w_i to another field g in window w_{i+1}. In the following, we explore the signatures that we receive in window w_i regarding her migration.

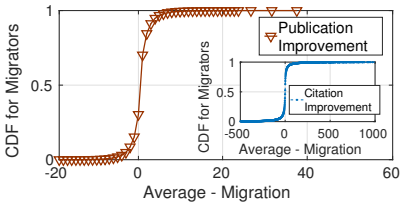
(i) Publication profile: The publication profile & quality (Vanclay and Bornmann 2012; Duffy et al. 2008) of researcher u in field f depict her performance in f, which plays an important role in her migration decision (see $Y_0(t)$ in Table. 2). One of the probable reasons behind migration can be the *decline in publication rate* in the current field. In the following, first we define $r_f^{w_i}$ as the publication rate



(a) Cumulative distribution of drop in publication rate compared to average rate $(b_f^{w_i} - r_f^{w_i})$ for the migrators. Inset shows their cumulative distribution of degradation in quality compared to average $(\hat{b}_f^{w_i} - \hat{r}_f^{w_i})$.



(b) Cumulative distribution of drop in leaving field productivity compared to average productivity $(bProd_f^{w_i} - Prod_f^{w_i})$ for the migrators. Inset shows their cumulative distribution of leaving field impact compared to average impact $(bImp_f^{w_i} - Imp_f^{w_i})$.



(c) Cumulative publication quantity $(\left|Pub_g^{w_{i+1}}\right| - \left|Pub_f^{w_i}\right|)$ and publication quality difference $(Q_g^{w_{i+1}} - Q_f^{w_i})$ distributions before and after migration.

Figure 5: (a) Reasons of migration- Individual publication profile, Inset: Individual quality degradation; (b) Reasons of migration- Field productivity, Inset: Field impact (c) Effects of migration- Individual publication quantity, Inset: Individual publication quality.

of the author in the time window w_i and $b_f^{w_i}$ as her average publication rate [9] prior to time window w_i.

Let $P_f^{y_i}$ be the set of publications of u in the departing field f in year y_i. Therefore, u's rate of publication in departing field f during window w_i can be defined as

$$r_f^{w_i} = \frac{\sum_{j=1}^{|w_i|-1} \frac{\left|P_f^{y_{j+1}^i}\right| - \left|P_f^{y_j^i}\right|}{\left|P_f^{y_j^i}\right|}}{|w_i| - 1} \quad (5)$$

where y_1^i, y_2^i and y_3^i are the three consecutive years in window w_i. We further define her average baseline publication rate in field f during w_1 to w_{i-1} as

$$b_f^{w_i} = \frac{\sum_{j=1}^{i-1} r_f^{w_j}}{i-1} \quad (6)$$

In Fig. 5(a), we plot the cumulative distribution of differences between the two rates (i.e. $(b_f^{w_i} - r_f^{w_i})$) for all the migrating authors. We observe that for 72.7% of cases, the migrating authors' publication rate $(r_f^{w_i})$ in field f in window w_i is substantially lower than her average publication rate $b_f^{w_i}$ in f. This observation gives an indication that a substantial decline in publication in a field f may trigger migration.

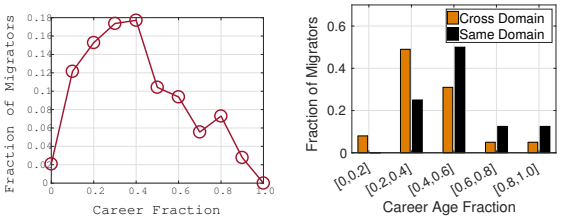
For testing the statistical significance of the aforementioned result, we propose a 'Null hypothesis' where we compute the rate difference $((b_f^{w_i} - r_f^{w_i}))$ of the non-migrators. In order to carry out a fair comparison, for each migrating author, we select a random set of 10 non-migrating authors of same 'reputation' (in terms of publication count, career age) as the migrating author, with respect to her top publishing field during migration [10]. Subsequently, we compute the average rate difference of those 10 non-migrators for comparing with the corresponding migrators. We obtain the mean rate difference of the migrators as 0.02, distant from that of the non-migrators (-0.01) and the low p-value $(3.68 * 10^{-9})$ indicates the perfect separation between them (from standard 'two-sample t' test).

Similarly, recent growth in the publication rate of u in the joining field g can also motivate her to migrate into g (see $Y_d(t)$ in Table. 2). We observe that for 84.6% of migrators the publication rate in joining field g undergoes a hike (with p-value $2.89 * 10^{-4}$) in window w_i compared to the baseline rate $(r_g^{w_i} - b_g^{w_i}) > 0$.

(ii) Quality degradation: Degradation in the quality of publications in the departing field f may be an important reason behind the possible migration of u (see $Y_0(t)$ in Table. 2). We consider the incoming citations, that an article receives, as a proxy to measure its quality (Vanclay and Bornmann 2012; Duffy et al. 2008). Extending this line of argument, we define the quality $Q_f^{y_i}$ of u's publications in field f in year y_i as the per paper citation count of her articles (of field f) published in year y_i. Subsequently, similar to publication profile (as shown in Eq. 5) we define u's rate of quality improvement in field f during window w_i as $\hat{r}_f^{w_i}$ and the average improvement rate before window w_i as $\hat{b}_f^{w_i}$ (similar to Eq. 6). In the inset of Fig. 5(a), we observe that for 72% of migrating authors the quality improvement rate of their publications in field f in window w_i becomes inferior than their average quality improvement rate in field f till w_i (i.e. $(\hat{b}_f^{w_i} - \hat{r}_f^{w_i}) > 0$). Essentially, the quality degradation (diminishing number of citations) of articles in field f works as a precursor of migration. Similar to publication profile, here we propose a 'Null hypothesis' considering the average $(\hat{b}_f^{w_i} - \hat{r}_f^{w_i})$ value of the equivalent non-migrators. The low p-value $(1.93 * 10^{-6})$ confirms the statistical significance of the aforesaid result. Additionally, for 85.9% of migrators the rate of obtaining citations in the joining field g is observed to rise during window w_i compared to average rate of past windows $((\hat{r}_g^{w_i} - \hat{b}_g^{w_i}) > 0)$ (with p-value $1.59 * 10^{-8}$) (see $Y_d(t)$ in Table. 2). Hence, this hike in the rate of obtaining citations in field g can also potentially attract the researchers to migrate into it.

[9] We deal with window-wise rates in stead of cumulative publication counts because considering window-wise rates normalizes the disparity of efforts provided to the same field across different windows.

[10] In the rest of the paper, whenever we refer to 'non-migrators', it would mean this set of same 'reputation' non-migrators.



(a) Distribution of career phase during migration of authors.

(b) Distribution of authors' career phases during cross domain and same domain migrations.

Figure 6: Career Phase at which researchers prefer to migrate within same domain, cross domain and overall.

(iii) Fields of close collaborators: We investigate the fields of close collaborators vis-a-vis the field (g) to which a migrating author u joins. We define the close collaborators of u as the top ten co-authors of u in the time window w_i, just before she migrates. We denote the set of close collaborators having the respective joining field g in one of their top 5 publishing fields as $CO_u^{w_i}$. Higher size of $CO_u^{w_i}$ reflects her acceptability in field g (see $P_1(t)$ in Table. 2). We observe that for 75.7% of migrating authors, $|CO_u^{w_i}| > 0$, which reveals that authors are inclined to migrate to fields in which her collaborators are actively working.

(iv) Career phase: We investigate the role of the career phase (Hu, Chen, and Liu 2014) of an author u on her field migration (see t in Table. 2). Considering that the career length of the author u spreads over N_u time windows $\{w_1, w_2, \dots w_{N_u}\}$and we observe her *first* field migration event during time $[w_i, w_{i+1}]$, we can identify her migrating career phase $p_u^{w_i}$ as i/N_u. In Fig. 6(a), we plot the distribution of $p_u^{w_i}$ for all the migrating authors. It is clearly evident that an author on average prefers to migrate in the *middle phase* of her career. Intuitively, this happens as the researchers may experience saturation in their currently working fields and explore the new domains for possible migration. Additionally, we observe (see Fig. 6(b)) that most of the researchers prefer cross domain migration at relatively earlier phase of their career. This is possibly due to the fact that cross domain migration generally demands more effort & time (also risk) to cope up with a completely new domain of research, which researchers prefer to undertake at the early phase. On the contrary, researchers performing migration late in their career prefer to migrate within the same domain.

(b) Researcher's working fields We investigate the collective role that the departing field f and the joining field g play to motivate author u in window w_i to migrate from field f to field g in window w_{i+1}. We examine two different aspects (i) Popularity of the field, and (ii) Impact of the field.

(i) Field popularity: The overall stagnancy of a field can also motivate the researchers to move out of it (see $P_3(t)$ in Table. 2). We denote $Pop_f^{w_i}$ as the total volume of papers published in field f (Vanclay and Bornmann 2012; Duffy et al. 2008) in time window w_i and $Auth_f^{w_i}$ as the number of authors involved in publishing them. We define the productivity ($Prod_f^{w_i}$) of a field in time window w_i as the ratio of this two terms (publication per author) i.e. $Prod_f^{w_i} = \frac{Pop_f^{w_i}}{Auth_f^{w_i}}$. The diminishing productivity of a field f may instigate the researchers to stop publishing in field f and initiate migration. We further define the average baseline productivity of the field f during w_1 to w_{i-1} as $bProd_f^{w_i}$ where,

$$bProd_f^{w_i} = \frac{\sum_{j=1}^{i-1} Prod_f^{w_j}}{i-1} \tag{7}$$

Indeed, in Fig. 5(b), we observe that for 75.6% of migrators the overall productivity of the currently working field f in window w_i experiences a fall. We build a 'Null hypothesis' considering the $(bProd_f^{w_i} - Prod_f^{w_i})$ values of all the non-migrating authors (averaged over all the windows) and observe that in case of migrating authors, influence of diminishing field popularity is statistically significant (p-value $1.17 * 10^{-8}$) than non-migrators.

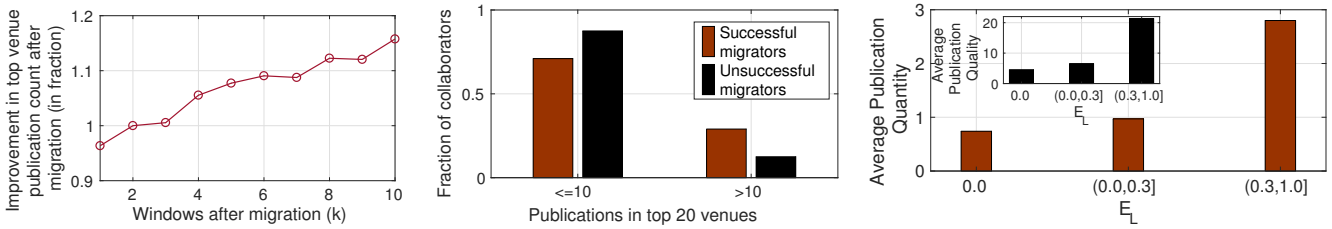
Similarly, recent hike in popularity of the joining field g may also motivate the researchers to migrate into it (see $P_2(t)$ in Table. 2). We observe that for 77.4% of migrators the target field g undergoes a hike in popularity in window w_i compared to window w_{i-1} (i.e. $(Prod_g^{w_i} - Prod_g^{w_{i-1}}) > 0$) (with p-value $1.11 * 10^{-6}$).

(ii) Field impact: The impact of a field f in an window w_i can be measured as the volume of paper-wise incoming citations (Vanclay and Bornmann 2012; Duffy et al. 2008) it receives in that time window w_i from other fields (reflects field popularity; see $P_3(t)$ in Table. 2). The lack of innovation of new concepts in a field f essentially reduces its impact and hence, may provoke the researchers working in f to migrate to other fields. We thereby define the overall impact of field f during time window w_i as $Imp_f^{w_i} = \frac{QF_f^{w_i}}{Pop_f^{w_i}}$ (where $QF_f^{w_i}$ is the total incoming citations towards field f during w_i). The average impact of field f before time window w_i is denoted as $bImp_f^{w_i}$. In the inset of Fig. 5(b), we observe that for 69.4% of migrators the overall impact of departing field f in window w_i becomes lower than the average impact of f till w_i (i.e. $(bImp_f^{w_i} - Imp_f^{w_i}) > 0$) (with p-value $2.46 * 10^{-5}$ showing significant distinctions from non-migrators). This overall quality degradation of articles in field f may work as a motivating factor for migration of author u.

Additionally, for 78.8% of migrators the impact of the joining field g is observed to rise during window w_i compared to window w_{i-1} (i.e. $(Imp_g^{w_i} - Imp_g^{w_{i-1}}) > 0$)(with p-value $3.19{*}10^{-5}$) (see $P_2(t)$ in Table. 2). Hence, this surge of impact in field g can potentially attract the researchers to migrate to g.

(B) Effect of migration on researcher's career

We consider that the author u migrates from the field f in time window w_i to another field g in window w_{i+1}. We demonstrate the implication of this migration on the career of author u in terms of her publication quantity and quality in the succeeding time window w_{i+1}.



(a) Improvement in top venue publication count after migration.

(b) Effects of collaborators on publication quality improvement.

(c) Correlation between the estimated intensity of migration (using Random Forest based regression) and effect of migration.

Figure 7: Effects of migration- (a) Improvement in top venue publication; (b) Impact of collaborators on quality of venues; (c) Correlation between the estimated intensity and actual effect of migration.

(i) Publication quantity: We define $Pub_f^{w_i}$ as the set of publications of u in field f during time window w_i. In the inset of Fig. 5(c), we show the cumulative distribution of u's improvement in publication count $(\left|Pub_g^{w_{i+1}}\right| - \left|Pub_f^{w_i}\right|)$ after migration. We observe that this difference yields positive values for 70% of migrators, indicating that the decision of migration proves to be beneficial in terms of publication quantity.

We reveal the benefit of the migrators vis-a-vis non-migrators in terms of publication quantity in their respective career. For each author, we compute the difference between the publication counts in the middle and the end of her career. It is observed that on average migrating authors enjoy a larger hike in publication count (16.5) at the end of their career compared to non-migrating authors (5.9).

(ii) Publication quality: We measure the quality of a publication from two aspects (a) incoming citations (b) prestige of publication venue. (a) First, we rely on the incoming citation count to determine the quality of a publication. Suppose, $Q_x^{w_i}$ measures the number of in-citations obtained by author u's publications in field x during w_i time window and f and g are the departing and joining field respectively. Plotting the cumulative distribution of $(Q_g^{w_{i+1}} - Q_f^{w_i})$ in Fig. 5(c), we observe that migrating to field g proves beneficial in terms publication quality for a significant fraction (61%) of migrating authors. Similar to publication count, in case of publication quality also we find that on average migrating authors obtain a larger boost (1061.2) in their overall citation counts compared to midway with respect to non-migrating authors (257.5).

(b) Quality of a publication is also indicated by the prestige of publication venue (conferences & journals). We rank the venues of every field by the total number of in-citations received by the articles published at those venues. Suppose, we denote the set of top k venues of the field f as Top_f^k and $PubTop_{f_k}^{w_i}$ as the set of publications of author u in the top k venues of field f during window w_i. First we compute the number of publications $\left|PubTop_{f_{20}}^{w_i}\right|$ of author u in the top 20 venues of the departing field f just before migration (in window w_i). Subsequently, we compute the top venue publications of u in the joining field g for each subsequent windows $(w_{i+1}, w_{i+2}, w_{i+3}, \ldots)$ after the migration event (as $\left|PubTop_{g_{20}}^{w_{i+1}}\right|, \left|PubTop_{g_{20}}^{w_{i+2}}\right|, \left|PubTop_{g_{20}}^{w_{i+3}}\right|, \ldots$ respectively). Fig. 7(a) which plots the ratio of $\left|PubTop_{g_{20}}^{w_{i+k}}\right| / \left|PubTop_{f_{20}}^{w_i}\right| \;\forall\; k = 1 \; to \; 10$, depicts that immediately after migration, the overall quantity of top venue publications degrade a little. However, after spending a short amount of time in the joining field, migrators cope up and exhibit a sharp improvement, even compared to the departing field.

(iii) Role of the collaborators: Furthermore, we aim to uncover the effect of collaborators on publishing in top quality venues after field migration. We first designate migrators as 'successful' and 'unsuccessful' based on whether they are able to improve their top-venue (considering top 20 venues) publication counts after the migration (i.e. $\left|PubTop_{g_{20}}^{w_{i+1}}\right| - \left|PubTop_{f_{20}}^{w_i}\right| > 0$). Subsequently, for each migrating author, we identify the most cited researcher collaborating with her just after migration. Fig. 7(b) shows that the collaborators of the 'successful' migrators have published relatively more papers in top venues (10.5 in average) compared to the collaborators of the 'unsuccessful' migrators (6 in average). Evidently, the choice of collaborators just after field migration indeed has a deep impact on the plausibility of migrating authors publishing in top venues of the newly joined field.

Migration prediction model

This is comforting for us to observe that each motivating factor derived from the microeconomic model of human migration is individually shown to be significantly correlating with the migration behavior of the scientific researchers. In the following, we propose a proof of concept *migration prediction model* to further substantiate the importance of the field migration factors introduced in this section. We develop the classification model to predict (*yes* or *no*) the future migration of a researcher in the next time window w_{i+1} to a specific research field g.

Extracting features & labeling authors Input of the prediction model is the profile of a migrating author u, till the prediction window w_i and a joining field g. Subsequently, for each migrating and non-migrating window of author u, we extract all the features introduced in this section & sum-

Type	Feature	Rank (+ve/-ve)
Publication Profile (f)	$(r_f^{w_i} - b_f^{w_i})$	1 (-)
Publication Profile (g)	$(r_g^{w_i} - b_g^{w_i})$	2 (+)
Quality Degradation (f)	$(\hat{r}_f^{w_i} - \hat{b}_f^{w_i})$	4 (-)
Quality Degradation (g)	$(\hat{r}_g^{w_i} - \hat{b}_g^{w_i})$	3 (+)
Fields of close Collaborators	$\lvert CO_u^{w_i} \rvert$	10 (+)
Career Phase	$p_u^{w_i}$	9 (-)
Field Popularity (f)	$(Prod_f^{w_i} - bProd_f^{w_i})$	8 (-)
Field Popularity (g)	$(Prod_g^{w_i} - Prod_g^{w_{i-1}})$	7 (+)
Field Impact (f)	$(Imp_f^{w_i} - bImp_f^{w_i})$	5 (-)
Field Impact (g)	$(Imp_g^{w_i} - Imp_g^{w_{i-1}})$	6 (+)
Domain	0 (same) or 1 (different)	11 (+)

Table 3: Features used in the prediction and their corresponding ranks as per the feature-weights returned by LR where 'Domain' indicates whether the leaving (f) and joining (g) fields belong to the same (0) or different (1) domains and '+ve/-ve' denotes the sign of the assigned weight.

marized in Table. 3. As the joining and leaving fields do not exist for any non-migrating window of author u, we consider her top publishing field of window w_i as the departing field f and her second top field in window w_i as the joining field g. For each migrating author u, we label the migrating window as 1 and remaining (non-migrating) windows as 0. We label the profile of all 66008 migrating authors in the 'Filtered' dataset. During model training, we under-sample the non-migrating) windows to balance the positive and negative samples and use 10-fold cross validation.

Evaluation First, we perform a *multi-variate* analysis of the extracted features in order to confirm their ability to discriminate the migrating and non-migrating windows. We choose standard Kruskal-Wallis test (Kruskal and Wallis 1952) and n-way ANOVA test (Morrison 2005) for this purpose and both of them indicate p-values much less than 0.05 ($1.09 * 10^{-5}$ and $6.41 * 10^{-5}$ respectively) proving the discriminating ability of our extracted features.

Next, we implement the prediction model using five standard Machine Learning (ML) algorithms - Support Vector Machine (SVM), Decision Tree (DT) (J48), Logistic Regression (LR), Random Forest (RF) & Multi-Layer Perceptron (MLP). In Table. 4, we show the elegance of the proposed models with high prediction performance (Random Forest achieves highest accuracy 87% and AUC 0.95.), confirming the utility of the features illustrated in Table. 3 as the motivating factors of migration. In addition, we also rank the features according to the weights (absolute) returned by the Logistic Regression based model. Evidently, 'Publication profile' and 'Quality degradation' related features ranked top in the list. The features corresponding to the departing field obtain a negative weight (indicating negative correlation with migration) whereas the features corresponding to the joining field obtain a positive weight (indicating positive correlation with migration), concurring with the Eq. 4.

Finally, we examine the career benefit of the researcher, obtained from the prediction model, with the effect of migration on the researcher's career. Instead of classification, we compute the career benefit E_L from the regression model; this is equivalent to the expected net return $ER(0)$ of Eq. 4. In Figure. 7(c), we show that both publication quality and quantity of the researcher after migration correlate with the estimated career benefit E_L. In a nutshell, our prediction model demonstrates that the features identified by the microeconomic model for human migration can be successfully adapted in the context of scientific migration.

Models	Acc.	Pr.	Re.	F_1	AUC
SVM	0.824	0.829	0.825	0.827	0.825
LR	0.821	0.825	0.822	0.823	0.897
DT (J48)	0.853	0.854	0.853	0.853	0.856
RF	**0.875**	**0.878**	**0.876**	**0.875**	**0.947**
MLP	0.849	0.850	0.849	0.849	0.929

Table 4: Classification Accuracy (Acc.), Precision (Pr.), Recall (Re.), F_1 score (F_1) and area under the ROC curve (AUC) values for the model using all five ML techniques.

Summarizing, in this section, we borrow the individual-centric and field-centric factors from standard human migration theory, to show their role in motivating authors to migrate. We observe that (a) the rate at which a migrating author publishes (and the corresponding publication qualities), generally diminishes during the time window before migration, (b) poor condition of the departing field and rise in the popularity of joining field (based on rate of publications as well as field impact) might also motivate an author to migrate. (c) While analyzing the effects of migration from micro dynamics perspective, we discover that individual publication quantity and quality improve in general after migration; however, publishing consistently in top-tier venues at the joining field after migration might take some time. (d) Finally, with the help of machine-learning based prediction model, we show that the microeconomic model for human migration can be successfully adapted in the context of scientific migration.

Conclusion

To the best of our knowledge, this is the first work which suitably adopts the process of scientific migration from classical human migration. We noticed that similar to human migration, there are two facets of the dynamics of scientific migration - the *individual side (micro perspective)* whereby highly 'reputed' (high publication/citation count) individuals in order to be part of a vibrant community (economy) may decide to 'relocate' (shift research interest); and the *collective side (macro perspective)* whereby sudden hike in the popularity of a research field drives a group of individuals to move into it. The decision of migration may come at a juncture of the migrator's personal crisis (drop in publication rate) or at crisis of the field (country) she is presently working. Of course, the transition becomes smoother and successful if one has an eminent collaborator in the new field (sponsor in the country one is immigrating) or she is young. However, the search for better scientific activity (living) by an individual/group behavior has a profound *impact on the field*. While major influx may strain the present infrastructure of a field but it enriches the field in the longer run by

bringing in diversity of knowledge; it enhances the depth and quality of the field by accumulating relevant knowledge from various 'older' fields and ensures new creative collaborations among researchers. Notably, with a 'proof of concept' classification model, we established the sanity of our analysis as well as obtained a relative ranking of the (impacts of) various factors affecting the migration dynamics. To conclude, we believe contextualizing the classical human migration theories unveils a new way of looking into the migration dynamics of scientific researchers and should motivate a trail of new research works in this area in future.

References

Abramo, G.; DAngelo, C. A.; and Di Costa, F. 2018. The effect of multidisciplinary collaborations on research diversification. *Scientometrics* 116(1):423–433.

Adams Jr, R. H., and Page, J. 2005. Do international migration and remittances reduce poverty in developing countries? *World development* 33(10):1645–1669.

Arrieta, O. A. D.; Pammolli, F.; and Petersen, A. M. 2017. Quantifying the negative impact of brain drain on the integration of european science. *Science Advances* 3(4):e1602232.

Barber, M. J. 2007. Modularity and community detection in bipartite networks. *Physical Review E* 76(6):066102.

Borjas, G. J. 1990. *Friends or strangers: The impact of immigrants on the US economy*. Basic Books.

Braun, D. 2012. Why do scientists migrate? a diffusion model. *Minerva* 50(4):471–491.

Chakraborty, T.; Tammana, V.; Ganguly, N.; and Mukherjee, A. 2015. Understanding and modeling diverse scientific careers of researchers. *Journal of Informetrics* 9(1):69 – 78.

Deville, P.; Wang, D.; Sinatra, R.; Song, C.; Blondel, V. D.; and Barabási, A.-L. 2014. Career on the move: Geography, stratification, and scientific impact. *Scientific reports* 4:4770.

Duffy, R. D.; Martin, H. M.; Bryan, N. A.; and Raque-Bogdan, T. L. 2008. Measuring individual research productivity: A review and development of the integrated research productivity index. *Journal of counseling psychology* 55(4):518.

Dugué, N., and Perez, A. 2015. *Directed Louvain: maximizing modularity in directed networks*. Ph.D. Dissertation, Université d'Orléans.

Fortunato, S.; Bergstrom, C. T.; Börner, K.; Evans, J. A.; Helbing, D.; Milojević, S.; Petersen, A. M.; Radicchi, F.; Sinatra, R.; Uzzi, B.; Vespignani, A.; Waltman, L.; Wang, D.; and Barabási, A.-L. 2018. Science of science. *Science* 359(6379).

Franzoni, C.; Scellato, G.; and Stephan, P. 2014. The mover's advantage: The superior performance of migrant scientists. *Economics Letters* 122(1):89–93.

Grubbs, F. E. 1969. Procedures for detecting outlying observations in samples. *Technometrics* 11(1):1–21.

Hagen-Zanker, J. 2008. Why do people migrate? a review of the theoretical literature. *SSRN Electronic Journal*.

Hargens, L. 1986. Migration patterns of us ph. d. s among disciplines and specialties. *Scientometrics* 9(3-4):145–164.

Hu, Z.; Chen, C.; and Liu, Z. 2014. How are collaboration and productivity correlated at various career stages of scientists? *Scientometrics* 101(2):1553–1564.

Jia, T.; Wang, D.; and Szymanski, B. K. 2017. Quantifying patterns of research-interest evolution. *Nature Human Behaviour* 1(4):0078.

Kruskal, W. H., and Wallis, W. A. 1952. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association* 47(260):583–621.

Massey, D. S., and Zenteno, R. M. 1999. The dynamics of mass migration. *Proceedings of the National Academy of Sciences* 96(9):5328–5335.

Massey, D. S.; Arango, J.; Hugo, G.; Kouaouci, A.; Pellegrino, A.; and Taylor, J. E. 1993. Theories of international migration: A review and appraisal. *Population and development review* 431–466.

Miller, J. 1991. Short report: Reaction time analysis with outlier exclusion: Bias varies with sample size. *The quarterly journal of experimental psychology* 43(4):907–912.

Morrison, D. F. 2005. Multivariate analysis of variance. *Encyclopedia of biostatistics* 5.

Mulkay, M. 1974. Conceptual displacement and migration in science: A prefatory paper. *Science Studies* 4(3):205–234.

Pramanik, S.; Yerra, P. H.; and Mitra, B. 2015. Whom-to-interact: does conference networking boost your citation count? In *IKDD CoDS*, 39–48. ACM.

Qazvinian, V., and Radev, D. R. 2009. The evolution of scientific paper title networks. In *ICWSM*.

Raj, R.-G., and Zainab, A. N. 2012. Relative measure index: a metric to measure the quality of journals. *Scientometrics* 93(2):305–317.

Richmond, A. H. 1988. Sociological theories of international migration: the case of refugees. *Current Sociology* 36(2):7–25.

Sjaastad, L. A. 1962. The costs and returns of human migration. *Journal of political Economy* 70(5, Part 2):80–93.

Todaro, M. P. 1969. A model of labor migration and urban unemployment in less developed countries. *The American economic review* 59(1):138–148.

Urban, D. 1982. Mobility and the growth of science. *Social Studies of Science* 12(3):409–433.

Van Houten, J.; Van Vuren, H.; Le Pairs, C.; and Dijkhuis, G. 1983. Migration of physicists to other academic disciplines: situation in the netherlands. *Scientometrics* 5(4):257–267.

Van Noorden, R. 2012. Global mobility: Science on the move. *Nature News* 490(7420):326.

Vanclay, J. K., and Bornmann, L. 2012. Metrics to evaluate research performance in academic institutions: a critique of era 2010 as applied in forestry and the indirect h2 index as a possible alternative. *Scientometrics* 91(3):751–771.

Vlachý, J. 1981. Mobility in physics. *Czechoslovak Journal of Physics B* 31(6):669–674.