The role of immigrants, emigrants and locals in the historical formation of European knowledge agglomerations

Philipp Koch, Viktor Stojkoski & César A. Hidalgo

    View supplementary material ↗

    Published online: 27 Nov 2023.

    Submit your article to this journal ↗

    View related articles ↗

    View Crossmark data ↗

Routledge
Taylor & Francis Group

RSA Regional Studies Association

Check for updates

The role of immigrants, emigrants and locals in the historical formation of European knowledge agglomerations

Philipp Koch[a,b] ⓘ, Viktor Stojkoski[a,c] ⓘ and César A. Hidalgo[a,d,e] ⓘ

ABSTRACT
Did migrants make Paris a mecca for the arts and Vienna a beacon of classical music? Or was their rise a pure consequence of local actors? We use data on more than 22,000 historical individuals born between the years 1000 and 2000 to estimate the contribution of famous immigrants, emigrants and locals to the knowledge specialisations of European regions. We find that the probability that a region develops or keeps specialisation in an activity (based on the birth of famous physicists, painters, etc.) grows with both the presence of immigrants with knowledge about that activity and immigrants with knowledge in related activities. In contrast, we do not find robust evidence that the presence of locals with related knowledge explains entries and/or exits. We address some endogeneity concerns using fixed-effects models considering any location–period–activity-specific factors (e.g., the presence of a new university attracting scientists).

KEYWORDS
migration; knowledge spillovers; relatedness; economic history; economic complexity

1. INTRODUCTION

Migrants help carry knowledge across space (Cipolla, 1972; Kerr et al., 2017; Lissoni, 2018; Trippl & Maier, 2011; Williams, 2006), shaping the geography of cultural and economic activities (Elekes et al., 2019; Miguelez & Morrison, 2022; Morrison, 2023; Neffke et al., 2018; Putterman & Weil, 2010). But most studies documenting the role of migrants in the diffusion of knowledge use recent data on patents (Bahar et al., 2020; Bernstein et al., 2022; Bosetti et al., 2015; Breschi et al., 2017; Fassio et al., 2019; Hunt & Gauthier-Loiselle, 2010; Miguelez & Morrison, 2022; Miguelez & Noumedem Temgoua, 2020; Miguélez & Moreno, 2013, 2015), research (Bosetti et al., 2015; Trippl, 2013) or product exports (Bahar & Rapoport, 2018), or analyse historical spillovers within activities (Borowiecki, 2012; Borowiecki & Graddy, 2021; Collins, 1974; Diodato et al., 2022; Ganguli, 2015; Hornung, 2014; Mitchell, 2019; Moser et al., 2014; Scoville, 1952a, 1952b; Waldinger, 2010, 2012), leaving questions about the role of migrants in the

historical formation of knowledge agglomerations relatively unexplored.

To explore the role of migrants in the historical formation of knowledge agglomerations we use biographical data on more than 22,000 famous individuals – artists, physicists, explorers, philosophers, etc. – living in Europe between the years 1000 and 2000. We use these data to investigate how immigrants, emigrants and locals explain the probability that famous individuals specialised in an activity – that was not yet present in a region – are born during the next century. That is, we study how the knowledge of migrants and locals contributes to explain, for example, Paris becoming the birthplace of painters and Vienna of composers.

We can explore these questions by creating measures of knowledge spillovers within and between locations and activities. Consider spillovers across locations within the same activity. The knowledge that migrants carry across borders may impact a location's ability to give birth to famous figures in the activity that the migrants specialise in. That is, immigrant mathematicians may increase the

probability that a city or region begets famous mathematicians. Similarly, emigrating mathematicians may decrease that probability. To capture such spillovers, we identify whether a region experiences a larger than expected inflow or outflow of famous individuals specialised in an activity.

Now consider spillovers across both locations and activities. Migrants and locals specialised in an activity (e.g., mathematics) can impact a region's ability to give birth to famous figures in a related activity (e.g., physics). To capture such spillovers, we use measures of relatedness (Balland et al., 2022; Boschma, 2017; Hidalgo, 2021; Hidalgo et al., 2007, 2018), which exploit information on the co-location of activities to estimate how 'cognitively close' a location is to an activity.

During the past decades, measures of relatedness have been validated as robust predictors of the probability that countries, regions and cities enter or exit an activity, such as product exports (Hidalgo et al., 2007; Pinheiro et al., 2021; Poncet & de Waldemar, 2015), technologies (Balland et al., 2019; Balland & Boschma, 2021; Boschma et al., 2015; Juhász et al., 2021; Petralia et al., 2017; Rigby, 2015; Uhlbach et al., 2022), industries (Deegan et al., 2021; Essletzbichler, 2015; Neffke et al., 2011, 2017) and research areas (Boschma et al., 2014; Chinazzi et al., 2019; Guevara et al., 2016). Recent contributions to this literature have focused on unpacking relatedness by considering multiple channels (Bahar et al., 2019; Boschma & Capone, 2015; Cortinovis et al., 2017; Diodato et al., 2018; Farinha et al., 2019; Jara-Figueroa et al., 2018; Jun et al., 2020; Zhu et al., 2017). For instance, does industry- or occupation-specific knowledge contribute to the growth and survival of firms? (Jara-Figueroa et al., 2018). Or do value chains or knowledge agglomerations explain the co-location of firms? (Diodato et al., 2018). To the best of our knowledge no study has yet unpacked relatedness in the context of historical migration. Here we use a dataset spanning 1000 years of history in Europe to explore how the knowledge of immigrants, emigrants and locals explains the probability that a famous cultural figure specialised in an activity is born in a specific region. This contributes to both understanding the role of migrants in the geography of knowledge and unpacking relatedness metrics in the context of migration.

Our findings show that migrants play a crucial role in knowledge agglomerations. Specifically, we find that the probability that a European region enters a new activity grows on average by between 1.7 and 4.6 percentage points if that region received an excess number of immigrants specialised in that activity during the last century. Moreover, we find this correlation is enhanced by immigrants specialised in related activities. Similarly, we find the probability that a European region loses one of its existing specialisations decreases on average by 5.0–10.2 percentage points if that region received an excess number of immigrants specialised in that activity. This correlation is also enhanced by immigrants specialised in related activities. In contrast, we do not find a statistically significant and robust role of the related knowledge of locals (people born in that region) in entries or exits.

To tackle some important endogeneity concerns (migration is often a motivated choice), we employ a highly restrictive fixed effects structure controlling for all possible unobserved factors that are specific to a broad occupational category in a region during a century. These are factors that might affect both, migration patterns and the birth of famous individuals, such as a new university attracting scientists and leading to the birth of more famous scientists in the future, or a prosperous city attracting and begetting more artists. In addition, we control for unobserved factors that are specific to a more granular occupational category in a century which might affect both migration and births, for example, the emergence of a new technology (e.g., photography) begetting a new occupational category (photographers). This captures, for instance, that musicians and singers are likely to have different migration and birth patterns across time than other artists such as painters or actors. Lastly, we tackle some concerns of reverse causality by focusing on excess migration and estimating the expected number of migrants in a location. Although we control for multiple possible observed and unobserved factors to limit endogeneity concerns, we want to stress that we are not able to make strictly causal claims.

Together, these findings advance our understanding of the role of immigrants, emigrants and locals in the historical formation of knowledge agglomerations. They contribute to both the literature on the role of migrants in knowledge diffusion (Bahar et al., 2020; Bahar & Rapoport, 2018; Borowiecki, 2012; Borowiecki & Graddy, 2021; Bosetti et al., 2015; Breschi et al., 2017; Cipolla, 1972; Diodato et al., 2022; Elekes et al., 2019; Fassio et al., 2019; Ganguli, 2015; Hornung, 2014; Hunt & Gauthier-Loiselle, 2010; Kerr et al., 2017; Lissoni, 2018; Miguelez & Morrison, 2022; Miguelez & Noumedem Temgoua, 2020; Miguélez & Moreno, 2013, 2015; Mitchell, 2019; Moser et al., 2014; Neffke et al., 2018; Putterman & Weil, 2010; Trippl, 2013; Trippl & Maier, 2011; Williams, 2006) and that on relatedness (Bahar et al., 2019; Balland et al., 2019, 2022; Balland & Boschma, 2021; Boschma, 2017; Boschma et al., 2014, 2015; Boschma & Capone, 2015; Chinazzi et al., 2019; Cortinovis et al., 2017; Deegan et al., 2021; Diodato et al., 2018; Essletzbichler, 2015; Farinha et al., 2019; Guevara et al., 2016; Hidalgo, 2021; Hidalgo et al., 2007, 2018; Jara-Figueroa et al., 2018; Juhász et al., 2021; Jun et al., 2020; Neffke et al., 2011, 2017; Petralia et al., 2017; Pinheiro et al., 2021; Poncet & de Waldemar, 2015; Rigby, 2015; Uhlbach et al., 2022; Zhu et al., 2017). Moreover, by developing measures of the related knowledge of migrants, we combine both migration and relatedness in a framework that can be used to study how knowledge spillovers across space and across activities combine in more recent settings. Lastly, this study provides a long-term perspective on the evolution of regional specialisations in Europe, a perspective which is underrepresented in the field of economic geography (Henning, 2019).

2. DATA AND METHODS

2.1. Data

We use the 2020 version of Pantheon (Yu et al., 2016), a publicly available dataset including information on famous individuals with a Wikipedia page in more than 15 different language editions. We focus on the 22,847 famous individuals born or died in Europe between the years 1000 and 2000. We choose Pantheon because it assigns individuals using a controlled taxonomy of 101 occupations, such as painter, writer, composer, physicist, chemist, mathematician, etc. Pantheon provides a good sectoral disaggregation compared with other datasets which either have few sectors (Schich et al., 2014) or use uncontrolled taxonomies with duplicate entries, for example, film director and movie director (Laouenan et al., 2022). This granularity is needed to construct measures of specialisation and relatedness. The full taxonomy and descriptive statistics are provided in Section 1.1 in the supplemental data online.

We use geographical coordinates to assign the place of birth and death of each biography to European administrative regions (NUTS-2 or regions of similar size for countries outside the EU, for example, Russian *oblasts*; see Section 1.2 in the supplemental data online). Figures 1 (a) and (b) show the places of birth and death of all individuals in our dataset within the applied administrative borders. Due to a lack of data on the full trajectory of individuals, we follow the literature investigating migration patterns of famous individuals (De La Croix & Licandro, 2015; Laouenan et al., 2022; Schich et al., 2014; Serafinelli & Tabellini, 2022) and use places of birth and death as rough proxies for migration. Manual inspection of a random sample of 200 biographies revealed places of death to be a valid proxy of an important living place for around 90% of biographies and corresponded to a place of major impact for 75% of biographies (see Section 1.3 in the supplemental data online).

Finally, we assign each individual to a century t based solely on his or her year of birth. That is, a famous person who is born in the 18th century in Brussels and died in Paris (in the 18th or 19th centuries) is considered a local in Brussels and an immigrant in Paris in the 18th century. We choose this approach since we do not have information on the time of migration. We take this into account in the regression models by lagging the independent variables (see also Section 1.1 in the supplemental data online).

2.2. Descriptive statistics: migration and spatial concentration patterns

We find that most of the migration of famous Europeans over the past 1000 years took place within countries and towards large cities (e.g., from smaller cities in France to Paris). Figures 1(c) and (d) visualise the migration network. Migration is common among famous individuals. In fact, going back to the 11th century, the share of migrants in our dataset never drops below 65%. In the 19th century, almost 80% of famous individuals in our dataset died in a different region than the one in which they were born (Figure 1e).

These migration patterns are not random but follow a process of preferential attachment, clustering individuals in major cities (Borowiecki, 2013; Borowiecki & Dahl, 2021; O'Hagan & Borowiecki, 2010; O'Hagan & Hellmanzik, 2008; Schich et al., 2014; Serafinelli & Tabellini, 2022) and leading to a higher spatial concentration for places of death than birth. For instance, 416 famous individuals were born in Paris in the 19th century, but 934 died there (see Section 2.1 in the supplemental data online).

We use information entropy H to quantify the spatial concentration of births and deaths across regions. Information entropy (base 2) estimates the number of yes/no questions that we would need to answer – on average – to find the place of birth or death of an individual (see Section 2.1 in the supplemental data online). If deaths are more concentrated than births, we will need less questions to guess a place of death than one of birth. We can use entropy H to estimate the effective number of places of birth or death as $E = 2^H$, which is the number of regions effectively experiencing the birth or death of a famous individual.

Figure 1(f) shows the effective number of places of birth and death E for each century. Before the 15th century, the spatial concentration of famous births and deaths was similar. But starting in the 15th century, places of death have become more spatially concentrated and places of birth more widespread. In fact, by the 19th century famous individuals were effectively born in more than 200 (out of 405) regions across Europe, while they effectively died in only 100 regions (Figure 1f).

2.3. Methods
2.3.1. Relatedness of immigrants, emigrants and locals

To explore how the knowledge of immigrants, emigrants and locals shapes the geography of knowledge, we estimate the probability that a region gives birth to a famous individual specialised in an activity as a function of estimates of knowledge spillovers within and between regions and activities.

To capture the knowledge spillovers of migration within the same activity, we calculate the ratio between the observed number of famous immigrants ($N_{ik,t}^{immi}$) or emigrants ($N_{ik,t}^{emi}$) with a certain activity and their expected number (respectively $\hat{N}_{ik,t}^{immi}$ and $\hat{N}_{ik,t}^{emi}$), where i denotes the region, k the occupation and t the century.

Taking the ratio between the observed and expected number of migrants allows us to create measures of excess immigration or excess emigration, and thus, to control for the natural attractiveness of a location and the characteristics of an activity. This is important to address reverse causality concerns, since the effects of migrants could be simply a reflection of local factors making a place attractive for migrants with a certain specialisation.
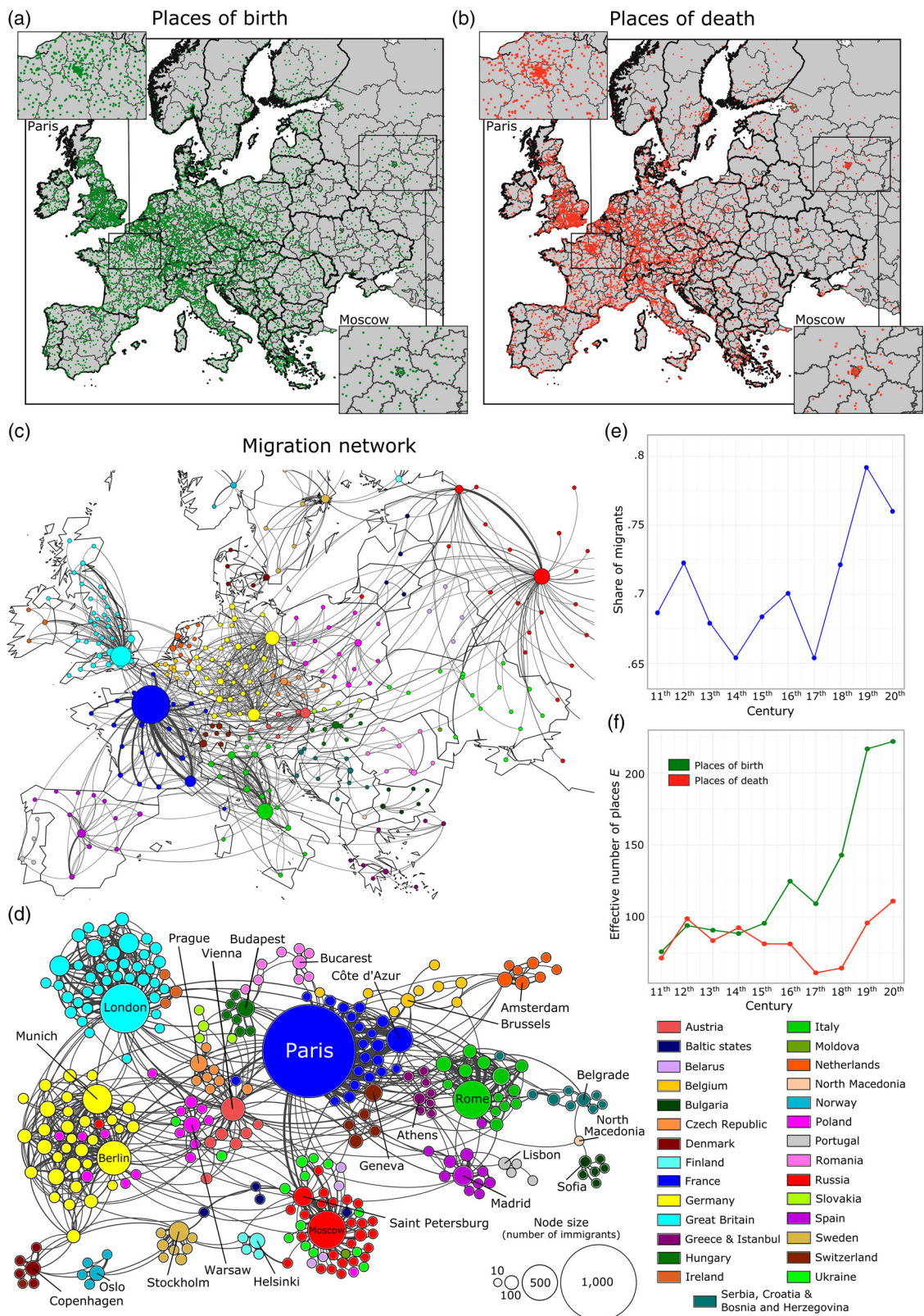
Figure 1. Places of birth, places of death and migration patterns of famous individuals in Europe over the past 1000 years. Maps of (a) places of birth and (b) places of death included in the analysis (NUTS-2 regions for the EU, comparable regions for other countries, for example, *oblasts* in Russia; see Section 1.2 in the supplemental data online). Migration network of famous individuals within Europe over the past 1000 years, using (c) geography or (d) a force-directed algorithm for visualisation. The latter reveals that famous individuals tend to move within countries towards large regions. (e) Share of migrants in the dataset per century. (f) Effective number of places of birth and death *E* derived from Shannon entropy (see Section 2.1 in the supplemental data online). Starting in the 15th century, the places of death of famous individuals are more spatially concentrated than their places of birth.

It is worth mentioning that migration decisions can be influenced by multiple local factors. Creatives, for instance, are more likely to move towards places that are already populated by other creatives (Borowiecki & Graddy, 2021) or potential patrons (Haskell, 1996/2000; Oevermann et al., 2007). Geographical and cultural distance (Caragliu et al., 2013; Lewer & Van den Berg, 2008), such as a common language or the presence of fellow countrymen, can also play a role (Rephann & Vencatasawmy, 2000). Lastly, migration can also be exogenously forced due to conflict (Borowiecki, 2012) or climate (Abel et al., 2019). By focusing on excess migrants instead of total migrants, in a restrictive fixed-effects model we help mitigate the risks of reverse causality. Mathematically, this involves taking the ratio between the observed and expected number of immigrants or emigrants:

$$R_{ik,t}^{immi} = \frac{N_{ik,t}^{immi}}{\hat{N}_{ik,t}^{immi}}$$

$$R_{ik,t}^{emi} = \frac{N_{ik,t}^{emi}}{\hat{N}_{ik,t}^{emi}}, \tag{1}$$

where the values are for individuals in region i and activity k born in century t.

Here we use two models for the expected number of migrants ($\hat{N}_{ik,t}$). The first considers the number of individuals in a location and the number of individuals specialised in an activity. That is a 'bins and balls' model for the expected number of immigrants or emigrants, making equation (1) the revealed comparative advantage (Balassa, 1965) or location quotient, a common measure of specialisation:

$$\hat{N}_{ik,t} = \frac{\sum_k N_{ik,t} \sum_i N_{ik,t}}{\sum_{i,k} N_{ik,t}}. \tag{2}$$

The second model expands on this by taking into account the attractiveness of a location in a specific activity (Neffke et al., 2011). We model $\hat{N}_{ik,t}$ using a negative binomial regression where we control for the observed number in the previous century ($N_{ik,t-1}$), the previous specialisation of the location in the activity based on famous individuals born there ($S_{ik,t-1}^{births}$),

$$S_{ik,t-1}^{births} = \frac{N_{ik,t-1}^{births}}{\left(\frac{\sum_k N_{ik,t-1}^{births} \sum_i N_{ik,t-1}^{births}}{\sum_{i,k} N_{ik,t-1}^{births}} \right)}. \tag{3}$$

where $N_{ik,t}^{births}$ denotes the number of famous individuals born in location i specialised in activity k in century t, and fixed effects for each location–time (θ_{it}) and activity–time (ϑ_{kt}) to account for unobserved factors. That is, we estimate:

$$\hat{N}_{ik,t} = f(\alpha_0 + \alpha_1 N_{ik,t-1} + \alpha_2 S_{ik,t-1}^{births} + \theta_{it} + \vartheta_{kt}), \tag{4}$$

where f denotes the negative binomial probability density (see Section 3.4.1 in the supplemental data online for results).

If the observed number of immigrants or emigrants in an activity exceeds the expected number, we say that region received excess immigrants, or produced excess emigrants, on that activity and time period.

Next, we create two specialisation matrices for famous immigrants ($M_{ik,t}^{immi}$; died 'here' but born elsewhere) and emigrants ($M_{ik,t}^{emi}$; born 'here' but died elsewhere):

$$\begin{aligned} M_{ik,t}^{immi} &= \begin{cases} 1 & \text{if } R_{ik,t}^{immi} \geq 1 \\ 0 & \text{otherwise} \end{cases} \\ M_{ik,t}^{emi} &= \begin{cases} 1 & \text{if } R_{ik,t}^{emi} \geq 1 \\ 0 & \text{otherwise}. \end{cases} \end{aligned} \tag{5}$$

Figures 2(a) and (b) show these two matrices using data for individuals born in the 19th century. The matrices are characterised by a nested structure that we recover by sorting locations by diversity (respectively $\sum_k M_{ik,t}^{immi}$ and $\sum_k M_{ik,t}^{emi}$), and activities by ubiquity (respectively $\sum_i M_{ik,t}^{immi}$ and $\sum_i M_{ik,t}^{emi}$). This structure is typical for matrices summarising the geography of activities (Bustos et al., 2012; Hausmann & Hidalgo, 2011) (see Section 2.2 in the supplemental data online), but also for networks describing species interactions in ecology (Almeida-Neto et al., 2008; Bascompte et al., 2003; Bastolla et al., 2009).

To capture spillovers across activities we use measures of relatedness (Balland et al., 2022; Boschma, 2017; Hidalgo, 2021; Hidalgo et al., 2007, 2018). Relatedness exploits information on the co-location of activities to estimate their affinity with a location. We create three separate measures of relatedness for immigrants, emigrants and locals.

These measures build on the specialisation matrices described in equation (5). This time, however, we need to create specialisation matrices for locals, which we define as famous individuals who were born in a region, no matter if they died there or elsewhere. We use this definition because of the large share of migrants among famous individuals (Figure 1d), which would reduce our number of observations drastically if we defined locals as individuals who were born and died in the same place. Controlling for the related knowledge of emigrants, however, relatedness based on all births is a valid proxy for the related knowledge of individuals who were born and died in the same region (see Section 2.3 in the supplemental data online).

That is, as before, we calculate the ratio between observed and expected births of famous individuals:

$$R_{ik,t}^{births} = \frac{N_{ik,t}^{births}}{\hat{N}_{ik,t}^{births}}. \tag{6}$$

Again, we can apply both the naïve model described in equation (2) or estimate the expected number of births given local factors (see equation 3 in Section 3.4.1 in the supplemental data online) before creating binary
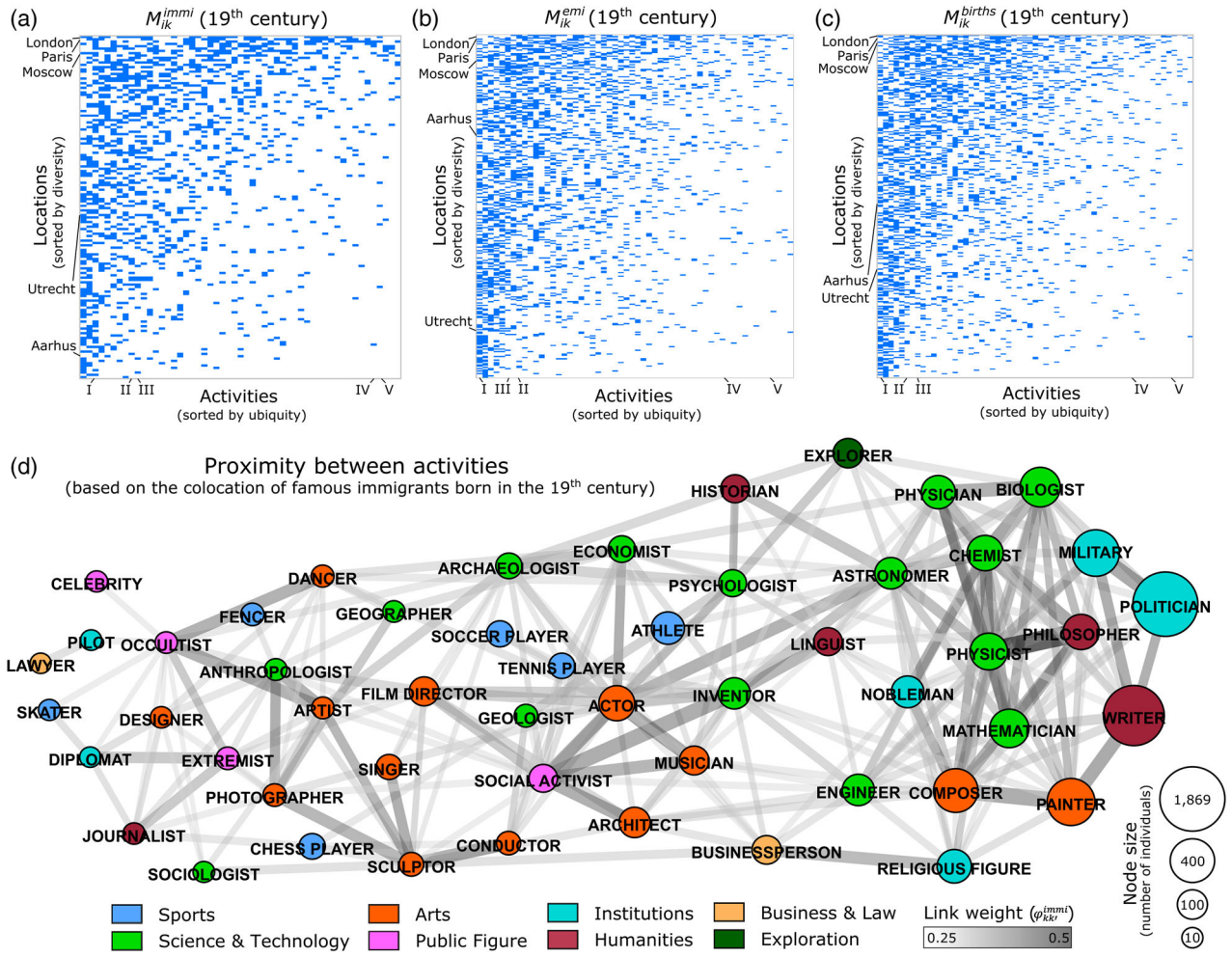
Figure 2. Nested specialisation matrices and the similarity between activities in the 19th century. Specialisation matrices based on (a) immigrants, (b) emigrants and (c) locals in the 19th century (see equations 5 and 7). Examples of locations and activities (I, writer; II, mathematician; III, physicist; IV, journalist; V, pilot) are highlighted. (d) Proximities between activities based on the co-location of famous immigrants born in the 19th century using the naïve model described in equation (2) to determine the expected number of immigrants. Node size is proportional to the number of famous individuals specialised in the respective activity and born in the 19th century.

specialisation matrices for locals:

$$M_{ik,t}^{births} = \begin{cases} 1 & if\ R_{ik,t}^{births} \geq 1 \\ 0 & otherwise \end{cases} \quad (7)$$

This matrix also exhibits a nested structure (Figure 2c).

Next, we define the proximity or similarity between two activities as the minimum of the conditional probability that a location is specialised in both of them (Hidalgo et al., 2007):

$$\varphi_{kk',t}^{immi} = \frac{\sum_i M_{ik,t}^{immi} M_{ik',t}^{immi}}{max(\sum_i M_{ik,t}^{immi},\ \sum_i M_{ik',t}^{immi})},$$

$$\varphi_{kk',t}^{emi} = \frac{\sum_i M_{ik,t}^{emi} M_{ik',t}^{emi}}{max(\sum_i M_{ik,t}^{emi},\ \sum_i M_{ik',t}^{emi})}, \quad (8)$$

$$\varphi_{kk',t}^{births} = \frac{\sum_i M_{ik,t}^{births} M_{ik',t}^{births}}{max(\sum_i M_{ik,t}^{births},\ \sum_i M_{ik',t}^{births})},$$

and use these proximities to calculate the relatedness between locations and activities as:

$$\omega_{ik,t}^{immi} = \frac{\sum_{k'} M_{ik',t}^{immi} \varphi_{kk',t}^{immi}}{\sum_{k'} \varphi_{kk',t}^{immi}},$$

$$\omega_{ik,t}^{emi} = \frac{\sum_{k'} M_{ik',t}^{emi} \varphi_{kk',t}^{emi}}{\sum_{k'} \varphi_{kk',t}^{emi}}, \quad (9)$$

$$\omega_{ik,t}^{births} = \frac{\sum_{k'} M_{ik',t}^{births} \varphi_{kk',t}^{births}}{\sum_{k'} \varphi_{kk',t}^{births}}.$$

These measures quantify how far, for example, immigrants to Paris are from being specialised in archaeology, emigrants from Madrid are from being specialised in singing or locals in Berlin are from being specialised in philosophy.

We note that the relatedness densities calculate with the naïve and binomial model are highly correlated ($R^2 >$ 0.9). So, going forward, we present results using the

naïve model and provide additional results using the negative binomial model in Section 3.4.1 in the supplemental data online.

Since the multiple factors contributing to the co-location of activities can be different when looking at immigration, emigration and births, we create separate measures of proximity ($\varphi_{kk',t}^{immi}$, $\varphi_{kk',t}^{emi}$, $\varphi_{kk',t}^{births}$, equation 8). But as a robustness check, we also consider a joint measure of proximity ($\varphi_{kk',t}^{joint}$) using co-location at birth and death (see Section 2.5 in the supplemental data online). Nevertheless, we find the separate measures of proximity provide valuable nuance (see Figure A6 online). Consider explorers and military personnel. Explorers and military personnel share many required capabilities such as navigating, planning, commanding, etc., that may be explained by local factors such as military academies for education, distance to the sea, recency of a war or naval technology. Also, exploration teams often involve soldiers and military personnel, which could then become famous as explorers. Hence, explorers and military personnel are likely to share a geographical origin. Yet, since exploration and military campaigns tend to involve different locations, these two activities are less likely to collocate at death. Now consider composers and noblemen. For these two activities, the proximity based on immigration patterns is higher than the proximity based on births. It makes sense that these activities are to some extent related when looking at places of birth: noblemen are known to be patrons for the arts. Hence, noblemen born in a location will likely create institutions that promote the cultivation of the talent of composers born in this location. But it is also plausible that these activities are even more related when looking at immigration patterns. Given that we observe a disproportional migration flow of noblemen towards a certain location, we can view this location as highly related to composers, since the institutional factors attracting noblemen likely play a role in attracting and cultivating the talent of composers as well.

These examples highlight why we believe that generating separate measures of proximity for immigrants, emigrants and births provides a nuanced perspective that helps unpack relatedness (see Section 2.5 in the supplemental data online for more details).

We illustrate the structure of these proximity networks for immigrants born in the 19th century ($\varphi_{kk',t}^{immi}$) (Figure 2d). A high proximity between two activities indicates similarity or complementarity among them. Like measures of propensity, measures of proximity capture the combined presence of multiple factors that may be contributing to the co-location of two activities. For example, we find a high proximity between biologists and physicians, mathematicians and physicists, and musicians and actors (Figure 2d). While the latter may be considered an example of co-location due to high complementarity (musicians and actors may perform together), associations between mathematicians and physicists, or biologists and physicians, may indicate similarity in knowledge or skills.

2.3.2. Entries and exits

We use our measures of relatedness to study the entry and exit of activities in European regions. We do this by estimating logistic models explaining the probability that a region starts to give birth to a disproportionately large number of famous individual specialised in an activity (entries) or stops doing so (exits). That is, a region enters the activity 'philosophy' if more philosophers are born there in a certain century than expected, while this has not been the case in the prior century. Similarly, we explain the probability that a region loses an existing specialisation. A region exits the activity 'physics' if fewer physicists are born there than expected, while this has not been the case in the prior century. The variables $Entry_{ik,t}$ and $Exit_{ik,t}$ emerge directly from the specialisation matrix defined in equation (7).

Specifically, we define:

$$Entry_{ik,t} = \begin{cases} 1 & if\ M_{ik,t-1}^{births} = 0\ and\ M_{ik,t}^{births} = 1, \\ 0 & otherwise \end{cases},$$

$$Exit_{ik,t} = \begin{cases} 1 & if\ M_{ik,t-1}^{births} = 1\ and\ M_{ik,t}^{births} = 0 \\ 0 & otherwise \end{cases}. \tag{10}$$

That is, a region i enters (exits) an occupation k in century t if the observed births of famous individuals with that occupation during the considered century is larger (lower) than expected, while this was not the case in the prior century.

Defining entries and exits looking at places of birth is a rather conservative approach. For a region to enter an activity, it needs to become a place where the required knowledge to cultivate a certain talent can be absorbed through formal or informal institutions and social ties. Indeed, early exposure to local knowledge in an individual's life is highly relevant in shaping his or her career, both for inventors nowadays (Bell et al., 2019) and artists centuries ago (Galenson, 2009, p. 278). A different approach of describing the geography of knowledge would be, for instance, to focus on all individuals living at a certain place. But this would require having data on all places of living.

We explain entries and exits using measures of the presence of immigrants and emigrants in that activity ($M_{ik,t}^{immi}$, $M_{ik,t}^{emi}$) and of the related activities that we can attribute to immigrants, emigrants and locals ($\omega_{ik,t}^{immi}$, $\omega_{ik,t}^{emi}$, $\omega_{ik,t}^{births}$). For instance, a significantly positive correlation between $M_{ik,t}^{immi}$ and entries would point towards migrants bringing into the region the knowledge needed to carry out activity k. That is, a high influx of mathematicians would increase the probability that the region begets its own famous mathematicians. This would be consistent with research showing that migrants help carry the knowledge needed to enter an activity (Bahar et al., 2020; Bahar & Rapoport, 2018; Borowiecki, 2012; Borowiecki & Graddy, 2021; Bosetti et al., 2015; Breschi et al., 2017; Diodato et al., 2022; Fassio et al., 2019; Ganguli, 2015; Hornung, 2014; Hunt & Gauthier-

Loiselle, 2010; Miguelez & Morrison, 2022; Miguelez & Noumedem Temgoua, 2020; Miguélez & Moreno, 2013, 2015; Mitchell, 2019; Moser et al., 2014; Trippl, 2013).

Similarly, a significant correlation between $\omega_{ik,t}^{immi}$ and entries would support the idea that the related knowledge brought by migrants also impacts the probability that a region develops a new activity. That is, the knowledge of famous immigrants specialised in mathematics diffuses to related fields, such as physics or chemistry, and increases the probability that a region begets its own physicists or chemists.

Lastly, a significant correlation between $\omega_{ik,t}^{births}$ and entries, after controlling for $\omega_{ik,t}^{emi}$, would indicate that the related knowledge of locals contributes to entering a new activity. That is, a region with many locals already specialised in mathematics has a higher probability of branching into physics or chemistry.

The entry of a region into a new activity could be the result of multiple factors other than migration. For instance, the creation of a new university could attract scientists, and the expansion of a port could create conditions attractive to merchants. We address such endogeneity concerns by using highly restrictive fixed effects models accounting for unobserved factors that could affect both migration and the probability that a region enters an activity. Specifically, we control for these unobserved factors by using fixed effects specific to a broad occupational category, region, and century (γ_{mit}, i.e., a three-way interaction). Index m denotes one of eight broad occupational categories such as 'arts', 'science and technology', 'humanities' or 'sports' (see Table S1, column 1, in the supplemental data online).

In addition, we control for unobserved factors affecting both migration and future births that are specific to a more granular occupational category and time (δ_{lt}). Index l denotes one of 26 occupation categories, which distinguish, for instance, between social sciences, natural sciences and engineering within the broad category 'science and technology' or music, design, and film and theatre within the broad category 'arts' (see Table S1, column 2, the supplemental data online). The latter fixed effects capture, for instance, that the invention of motion picture technology at the end of the 19th century likely affected migration and birth patterns among film directors and actors differently than among other artists, such as painters or sculptors.

We also control for several other observed factors that might correlate with the probability of entry or exit and that are not captured in the fixed effects. This includes an activity's ubiquity (i.e., the number of locations specialised in it) and how close a region already is to having or losing a specialisation ($R_{ik,t-1}^{births}$; see equation 6). Lastly, we account for knowledge diffusion across space due to other reasons than migration by creating measures of the spatial proximity to other regions with specialisations in that specific activity or in related activities (see Section 2.4 in the supplemental data online). We provide descriptive statistics and discuss the explanatory variables in more detail in Section 3.1 online.

In sum, we define $Y_{ik,t} = \{Entry_{ik,t}, \ Exit_{ik,t}\}$ and estimate:

$$\begin{aligned}
P(Y_{ik,t}) = g(&\beta_1 M_{ik,t-1}^{immi} + \beta_2 M_{ik,t-1}^{emi} \\
&+ \ \beta_3 \omega_{ik,t-1}^{immi} + \beta_4 \omega_{ik,t-1}^{emi} + \beta_5 \omega_{ik,t-1}^{births} \quad (11) \\
&+ \ \boldsymbol{\alpha}' \boldsymbol{X}_{ik,t-1} + \gamma_{mit} + \delta_{lt} + \varepsilon_{ik,t}),
\end{aligned}$$

where g denotes the logistic probability density, $\boldsymbol{X}_{ik,t-1}$ denotes a vector of observed control variables, and γ_{mit}, δ_{lt} the fixed effects.

We calculate average marginal effects based on this logistic regression by computing the marginal effect for each data point and taking the average.

3. RESULTS

Table 1 and Figure 3 show the relationship between the activities of immigrants, emigrants and locals and the number of observed entries and exits. For entries (Table 1, columns 1–5), the probability correlates positively with an excess inflow of migrants specialised in an activity during the previous century ($M_{ik,t-1}^{immi} = 1$). Specifically, an excess of immigrants increases the probability of entry on average by 4.6 percentage points (Figure 3a). Figure 3(b) plots the probability of entry as a function of $M_{ik,t-1}^{immi}$.

We also find that the probability of entry grows with the related knowledge of immigrants. A standard deviation increase of $\omega_{ik,t-1}^{immi}$ increases the probability of entry on average by 5.8 percentage points (Figure 3a). Figure 3(c) visualises the results by plotting the average probability of entry as a function of the relatedness density of immigrants ($\omega_{ik,t-1}^{immi}$). In accordance with the literature (Alshamsi et al., 2018; Hidalgo et al., 2018), the average probability of entry grows super-linearly, from 1.1% if no related knowledge of famous immigrants is present in a region ($\omega_{ik,t-1}^{immi} = 0$) to 16.4% if all related activities are present ($\omega_{ik,t-1}^{immi} = 100$). Moreover, we find a positive correlation ($p < 0.1$) between $\omega_{ik,t-1}^{births}$ and entries, but unlike the estimate of the related knowledge of immigrants, this correlation is not robust (see Section 3.4 in the supplemental data online).

When we look at exits (Table 1, columns 6–10), we find similar relationships but with the opposite sign. An excess inflow of famous individuals specialised in an activity during the previous century ($M_{ik,t-1}^{immi} = 1$) reduces the probability of exit significantly by 10.2 percentage points on average (Figure 3d). Also, the related knowledge of immigrants ($\omega_{ik,t-1}^{immi}$) helps prevent losing specialisation in an activity. Figure 3(e) and (f) visualise these results by plotting the probability of exit as a function of $M_{ik,t-1}^{immi}$ and $\omega_{ik,t-1}^{immi}$, respectively.

These results are robust to estimating the expected number of immigrants, emigrants and locals in equations (1) and (6) using the negative binomial regression model described in equation (3) (see Section 3.4.1 in the supplemental data online). By accounting for local factors, we are able to obtain a more accurate estimate of the

Table 1. Main results of logistic regression models explaining entries and exits of activities.

	Dependent variable: $Entry_{ik,t}$					Dependent variable: $Exit_{ik,t}$				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
$M_{ik,t-1}^{immi}$	0.334***	0.303***	0.336***	0.331***	0.300***	−0.603***	−0.584***	−0.591***	−0.587***	−0.571***
	(0.080)	(0.075)	(0.086)	(0.080)	(0.076)	(0.127)	(0.134)	(0.120)	(0.126)	(0.126)
$M_{ik,t-1}^{emi}$	0.115	0.045	0.106	0.121	0.018	0.310	0.330	0.233	0.306	0.291
	(0.261)	(0.278)	(0.261)	(0.255)	(0.270)	(0.240)	(0.232)	(0.216)	(0.222)	(0.203)
$\omega_{ik,t-1}^{immi}$		0.027***			0.028***		−0.067***			−0.064***
		(0.006)			(0.007)		(0.016)			(0.011)
$\omega_{ik,t-1}^{emi}$			−0.006		−0.024			−0.048		−0.025
			(0.012)		(0.019)			(0.038)		(0.063)
$\omega_{ik,t-1}^{births}$				0.011	0.027*				−0.059***	−0.034
				(0.008)	(0.015)				(0.018)	(0.041)
Further controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fixed effects										
Broad category–region–century	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Category–century	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	3944	3944	3944	3944	3944	1051	1051	1051	1051	1051
Pseudo-R^2	0.213	0.214	0.213	0.213	0.215	0.224	0.230	0.226	0.226	0.232
BIC	9537.0	9539.4	9545.0	9544.5	9553.1	3619.6	3618.0	3623.4	3623.3	3628.8

Note: The fixed effects in these models are highly restrictive, amounting to more than 700 parameters in columns (1) to (5) and more than 350 parameters in columns (6) to (10). All regions included in the regression model exhibit a minimum number of births and migrants such that measures of specialisation and relatedness are defined (see Section 2.2 in the supplemental data online). Standard errors are clustered by region and period. For the full regression tables with all control variables, see Sections 3.2 and 3.3 online. BIC, Bayesian information criterion. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.
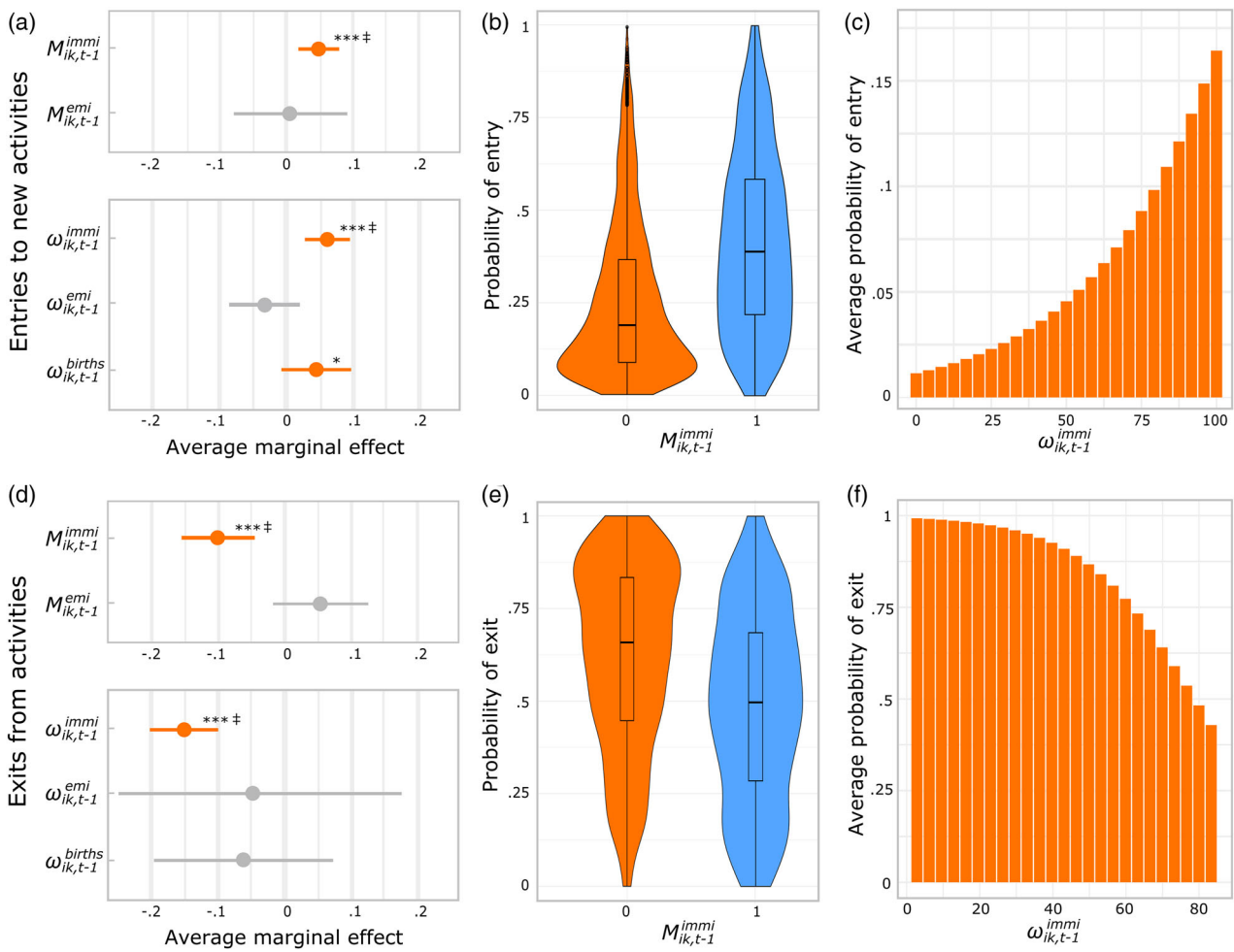
Figure 3. Visualisation of the main results. (a) Average marginal effects on the probability of entry to new activities based on the logistic regression model in Table 1, column 5. $M_{ik,t-1}^{immi} = 1$ increases the probability of entry on average by 4.6 percentage points, while a standard deviation increase of $\omega_{ik,t-1}^{immi}$ correlates with an increase in the average probability of entry of 5.8 percentage points. Probability of entry to a new activity as a function of (b) $M_{ik,t-1}^{immi}$ and (c) the immigrants' related knowledge, $\omega_{ik,t-1}^{immi}$. (d) Average marginal effects on the probability of exit from activities based on the logistic regression model in Table 1, column 10. $M_{ik,t-1}^{immi} = 1$ reduces the probability of exit on average by 10.2 percentage points, while a standard deviation increase of $\omega_{ik,t-1}^{immi}$ correlates with a reduction in the average probability of exit of 15.1 percentage points. Probability of exit from an existing area of specialisation as a function of (e) $M_{ik,t-1}^{immi}$ and (f) the immigrants' related knowledge, $\omega_{ik,t-1}^{immi}$.
Note: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Average marginal effects are computed by taking the average of the marginal effects across observations; lines indicate the 95% confidence interval; ‡ denotes robustness of the results (see Section 3.4 in the supplemental data online).

expected number of immigrants, emigrants and locals and, thus, of a disproportionate migration flow. This mitigates some of the endogeneity concerns.

The highly restrictive fixed effects specification, however, reduces the number of observations in the regression model. To assure the robustness of our results, we estimate the logistic regression models with several less restrictive specifications. This also allows us to include observed variables previously captured by the fixed effects, such as urban population (Bairoch et al., 1988; Buringh, 2021) or a location's diversity of activities. We find that the knowledge of immigrants remains a significant predictor for both entries and exits (see Section 3.4.2 in the supplemental data online). We acknowledge that, over such long periods, travel times are not constant but decrease with improvements in infrastructure and/or technology.

Hence, we allow for century-specific effects of spatial proximity, $\rho_{ik,t-1}^{M}$ and $\rho_{ik,t-1}^{\omega}$, leaving our results unchanged (see Section 3.4.3 online). Also, our sample of famous individuals is not balanced over time. Our findings, however, are robust to excluding the twentieth century from the analysis as well as looking at the 20th century alone (see Section 3.4.4 online). Moreover, our results do not change if we redefine entries and exits as the first or last birth of a famous individual with a specific occupation in a location instead of developing or losing specialisation in an activity (see Section 3.4.5 online). In addition, we explore the explanatory power of interaction terms between various relatedness densities on entries, following the literature on migrants as agents of structural change (Elekes et al., 2019; Miguelez & Morrison, 2022; Neffke et al., 2018). We find a significant, but quantitatively

negligible negative interaction term between ω_{ik}^{immi} and ω_{ik}^{births}, indicating that the related knowledge of immigrants and locals are weak substitutes (see Section 3.4.6 online). Also, it may be that our findings of knowledge spillovers are different for different activities. We explore potential heterogenous effects by estimating our regression model separately for aggregate occupational categories (see Section 3.4.7 online). We find, for instance, a stronger correlation of the presence of immigrants specialised in the same activity ($M_{ik,t-1}^{immi}$) on entries in sciences and public institutions, and an increased correlation of related knowledge of immigrants (ω_{ik}^{immi}) in humanities and sports. Another source of heterogeneity can be city size when size plays a relevant role in generating knowledge spillovers (see Section 3.4.8 online). We find that most entries take place in large cities, and thus, the effects of migration and relatedness are mainly urban. Whereas for exits, we see that spillovers across activities are more important in larger cities. But the probability of exiting an activity in small cities grows massively with the emigration of individuals specialised in the same activity, pointing towards a pronounced role of talent loss in shaping regional specialisations of small agglomerations. This finding relates to the recent literature on left-behind places (Rodríguez-Pose, 2018; Rodríguez-Pose et al., 2023).

Lastly, although the ratio of observed above expected numbers (equations 1 and 6) fulfils the purpose of controlling for size and reverse causality, these models are opaque, not telling us whether our results are driven by changes in the observed or expected number (or both). Hence, we run our main regression model including all terms of the ratio as a robustness check (see Section 3.4.9 in the supplemental data online). We find that the observed number of immigrants with a specific occupation ($N_{ik,t-1}^{immi}$) correlates positively with future entries and negatively with future exits, confirming our main results with composite indices. One additional immigrant to a region with a specific occupation correlates with an average increase in the probability of entry by 1.68 percentage points and a reduction in the probability of exit by 5.04 percentage points (see Section 3.4.9 online).

4. DISCUSSION

Labour mobility and migration are core tenets of the United States and the European Union, because policymakers intuit that migrants carry knowledge across space and activities (Cipolla, 1972; Kerr et al., 2017; Lissoni, 2018; Trippl & Maier, 2011; Williams, 2006). Yet, despite multiple studies documenting the role of migrants in the diffusion of knowledge (Bahar et al., 2020; Bahar & Rapoport, 2018; Borowiecki, 2012; Borowiecki & Graddy, 2021; Bosetti et al., 2015; Breschi et al., 2017; Diodato et al., 2022; Fassio et al., 2019; Ganguli, 2015; Hornung, 2014; Hunt & Gauthier-Loiselle, 2010; Miguelez & Morrison, 2022; Miguelez & Noumedem Temgoua, 2020; Miguélez & Moreno, 2013, 2015; Mitchell, 2019; Moser et al., 2014; Scoville, 1952a, 1952b; Trippl, 2013; Waldinger, 2010, 2012), there is

little historical quantitative evidence of the role of migrants in the historical evolution of knowledge agglomerations.

Here, we used biographical data on more than 22,000 famous individuals – sculptors, composers, politicians, chemists, etc. – living in Europe between the years 1000 and 2000 to explore how the knowledge of immigrants, emigrants and locals explains the probability that a region enters or exits an activity.

Our findings show that migrants play a crucial role in the historical geography of knowledge. Specifically, we find that the probability that a European region enters a new activity grows with the presence of immigrants with knowledge on that activity. Also, using measures of relatedness (Balland et al., 2022; Boschma, 2017; Hidalgo, 2021; Hidalgo et al., 2007, 2018), we find that this correlation is enhanced by spillovers across related activities. Put differently, the probability that a region begets famous mathematicians grows with an excess immigration of mathematicians and with immigrants from related fields, such as physics or chemistry. Similarly, we find that the probability that a European region loses one of its existing areas of specialisation decreases with the presence of immigrants specialised in that activity and in related activities. However, we do not find that locals with related knowledge play the same statistically significant and robust role in entries or exits.

These findings advance our understanding of the evolution of European agglomerations over the past millennium and of the role of migrants and locals therein. Specifically, we find robust evidence that European agglomerations did not only evolve path-dependently (Nunn, 2021), but also that they benefited from spillovers generated by the migration of famous individuals. This supports the literature on the role of migrants in the diffusion of knowledge (Bahar et al., 2020; Bahar & Rapoport, 2018; Borowiecki, 2012; Borowiecki & Graddy, 2021; Bosetti et al., 2015; Breschi et al., 2017; Cipolla, 1972; Collins, 1974; Diodato et al., 2022; Elekes et al., 2019; Fassio et al., 2019; Ganguli, 2015; Hornung, 2014; Hunt & Gauthier-Loiselle, 2010; Kerr et al., 2017; Lissoni, 2018; Miguelez & Morrison, 2022; Miguelez & Noumedem Temgoua, 2020; Miguélez & Moreno, 2013, 2015; Mitchell, 2019; Moser et al., 2014; Neffke et al., 2018; Trippl, 2013; Trippl & Maier, 2011; Williams, 2006) and contributes to the literature on relatedness (Balland et al., 2022; Boschma, 2017; Hidalgo, 2021; Hidalgo et al., 2007, 2018) explaining changes in specialisation patterns (Balland et al., 2019; Balland & Boschma, 2021; Boschma et al., 2014, 2015; Chinazzi et al., 2019; Deegan et al., 2021; Essletzbichler, 2015; Guevara et al., 2016; Juhász et al., 2021; Neffke et al., 2011, 2017; Petralia et al., 2017; Pinheiro et al., 2021; Poncet & de Waldemar, 2015; Rigby, 2015; Uhlbach et al., 2022).

Migrants are known agents of structural change enabling the development of unrelated activities (Elekes et al., 2019; Miguelez & Morrison, 2022; Neffke et al., 2018). Our findings differ slightly from that by emphasising migration as a channel of related diversification and path-dependent development, adding to the literature unpacking

the principle of relatedness (Bahar et al., 2019; Boschma & Capone, 2015; Cortinovis et al., 2017; Diodato et al., 2018; Farinha et al., 2019; Jara-Figueroa et al., 2018; Jun et al., 2020; Zhu et al., 2017). Recently, this intersection between evolutionary economic geography, regional diversification and migration has been identified as a promising field of research (Morrison, 2023). We contribute methodologically to this literature by disentangling relatedness measures for immigrants, emigrants and locals. These novel measures make it possible to explore how knowledge spillovers across space and across activities combine (see Section 2.5 in the supplemental data online). Lastly, this study provides a long-term perspective on the evolution of regional specialisations in Europe, a perspective which has been underrepresented in economic geography (Henning, 2019).

Unfortunately, we do not observe the mechanisms explaining the entry or exit of regions in activities. There are, however, several potential mechanisms responsible for these results, which can be subsumed as horizontal and vertical socialisation (Mokyr, 2017, p. 37). For instance, immigrating physicists could teach at a university, leading to a local flourishing of the field of physics and increasing the probability that a famous physicist emerges in the future. Also, immigrating physicists may bring new ideas and approaches with them, which can stimulate creative thinking and cross-pollination of ideas among local scientists in related fields such as chemistry or mathematics. This could lead to the development of new methods as well as new ways of thinking about problems, which could in turn contribute to an increased probability of giving birth to famous chemists or mathematicians in the future. The mechanisms may be different in other activities such as the arts or humanities. The presence of immigrating musicians may create a critical mass of artists, making it profitable to build cultural infrastructure due to economies of scale (Borowiecki & Graddy, 2021), from which artists in related activities such as singers, composers or dancers benefit as well. Shedding light on these different mechanisms is a promising avenue for future research.

Our study has also other limitations. First, we observe only a small and highly mobile subset of the overall population. That is, 22,000 of the most famous individuals living in Europe over the past 1000 years. A more comprehensive dataset would allow for a more accurate and granular estimation of a location's related knowledge and the geography of activities. Indeed, we suspect that the limited sample is a likely reason for why we do not observe a statistically significant and robust relevance for locals in shaping the historical geography of knowledge. That being said, the related knowledge of locals plays a significant role in several specifications, for instance if estimating the expected number of famous individuals to define specialisations (see Section 3.4.1 in the supplemental data online) or for large cities (see Section 3.4.8 online). Continuing to investigate the role of locals in the historical geography of knowledge can be an interesting avenue for future research.

Second, we do not observe the full migration trajectory of individuals, but only their place of birth and place of death. Although this approach follows the literature (De

La Croix & Licandro, 2015; Laouenan et al., 2022; Schich et al., 2014; Serafinelli & Tabellini, 2022) and provides a good proxy of migration (see Section 1.3 in the supplemental data online), more detailed data on where famous individuals lived and when could provide a better analytical basis to explore the evolution of agglomerations (Lucchini et al., 2019; Menini et al., 2017). Indeed, based on a small number of famous individuals living between 1450 and 1750, it is estimated that they moved on average 3.72 times during their lifetime (Mokyr, 2005). Third, we focus only on Europe. So, it may be that the principles behind the historical geography of knowledge uncovered here are different for other parts of the world. Lastly, migration is influenced by multiple factors such as geography and culture (Caragliu et al., 2013; Lewer & Van den Berg, 2008), agglomeration (Borowiecki & Graddy, 2021), patrons (Haskell, 1996/2000; Oevermann et al., 2007) or conflict (Abel et al., 2019; Borowiecki, 2012), evoking reverse causality and endogeneity concerns in our study. We tackled these concerns by using highly restrictive fixed-effects and estimating the expected number of immigrants, emigrants and locals to define specialisations. Despite these efforts, we want to stress that we are not able to make strictly causal claims, a task that can be challenging using historical observational data.

Yet, despite these limitations, our study provides evidence of migration playing a central role in the evolution of European knowledge agglomerations. Also, while being a historical study, our study concerns a topic that is highly relevant in today's economic policy. The effects of migration on local economies have been debated intensively, both in academia (Borjas, 1994; Card, 2001; Ottaviano & Peri, 2012; Putterman & Weil, 2010) and in policy circles (OECD, 2021; World Bank, 2019). Our findings add to this debate by showing that the immigration of high-skilled individuals correlates with entering and exiting specialisations of regions. Yet, our results can neither be interpreted causally nor tell us whether these findings remain for migration that is incentivised by policy instruments, since we observe migration involving multiple forces, from forced displacement due to war, to organic forms of migration.

ACKNOWLEDGEMENTS

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

FUNDING

ORCID

Philipp Koch http://orcid.org/0000-0002-3248-9686
Viktor Stojkoski http://orcid.org/0000-0002-7594-4431
César A. Hidalgo http://orcid.org/0000-0002-6977-9492

REFERENCES

Abel, G. J., Brottrager, M., Crespo Cuaresma, J., & Muttarak, R. (2019). Climate, conflict and forced migration. *Global Environmental Change*, *54*, 239–249. https://doi.org/10.1016/j.gloenvcha.2018.12.003

Almeida-Neto, M., Guimarães, P., Guimarães, P. R., Loyola, R. D., & Ulrich, W. (2008). A consistent metric for nestedness analysis in ecological systems: Reconciling concept and measurement. *Oikos*, *117*(8), 1227–1239. https://doi.org/10.1111/j.0030-1299.2008.16644.x

Alshamsi, A., Pinheiro, F. L., & Hidalgo, C. A. (2018). Optimal diversification strategies in the networks of related products and of related research areas. *Nature Communications*, *9*(1), 1328. https://doi.org/10.1038/s41467-018-03740-9

Bahar, D., Choudhury, P., & Rapoport, H. (2020). Migrant inventors and the technological advantage of nations. *Research Policy*, *49*(9), 103947. https://doi.org/10.1016/j.respol.2020.103947

Bahar, D., & Rapoport, H. (2018). Migration, knowledge diffusion and the comparative advantage of nations. *The Economic Journal*, *128*(612), F273–F305. https://doi.org/10.1111/ecoj.12450

Bahar, D., Rosenow, S., Stein, E., & Wagner, R. (2019). Export take-offs and acceleration: Unpacking cross-sector linkages in the evolution of comparative advantage. *World Development*, *117*, 48–60. https://doi.org/10.1016/j.worlddev.2018.12.016

Bairoch, P., Batou, J., & Chèvre, P. (1988). *La population des villes européennes de 800 à 1850*. Librairie Droz.

Balassa, B. (1965). Trade liberalisation and revealed comparative advantage. *The Manchester School*, *33*(2), 99–123. https://doi.org/10.1111/j.1467-9957.1965.tb00050.x

Balland, P.-A., & Boschma, R. (2021). Complementary interregional linkages and smart specialisation: An empirical study on European regions. *Regional Studies*, *55*(6), 1059–1070. https://doi.org/10.1080/00343404.2020.1861240

Balland, P.-A., Boschma, R., Crespo, J., & Rigby, D. L. (2019). Smart specialization policy in the European union: Relatedness, knowledge complexity and regional diversification. *Regional Studies*, *53*(9), 1252–1268. https://doi.org/10.1080/00343404.2018.1437900

Balland, P.-A., Broekel, T., Diodato, D., Giuliani, E., Hausmann, R., O'Clery, N., & Rigby, D. (2022). The new paradigm of economic complexity. *Research Policy*, *51*(3), 104450. https://doi.org/10.1016/j.respol.2021.104450

Bascompte, J., Jordano, P., Melián, C. J., & Olesen, J. M. (2003). The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, *100*(16), 9383–9387. https://doi.org/10.1073/pnas.1633576100

Bastolla, U., Fortuna, M. A., Pascual-García, A., Ferrera, A., Luque, B., & Bascompte, J. (2009). The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, *458*(7241), 1018–1020. https://doi.org/10.1038/nature07950

Bell, A., Chetty, R., Jaravel, X., Petkova, N., & Van Reenen, J. (2019). Who becomes an inventor in America? The importance of exposure to innovation. *The Quarterly Journal of Economics*, *134*(2), 647–713. https://doi.org/10.1093/qje/qjy028

Bernstein, S., Diamond, R., Jiranaphawiboon, A., McQuade, T., & Pousada, B. (2022). *The contribution of high-skilled immigrants to innovation in the United States* (NBER Working Paper Series w30797). https://doi.org/10.3386/w30797

Borjas, G. J. (1994). The economics of immigration. *Journal of Economic Literature*, *32*(4), 1667–1717.

Borowiecki, K. J. (2012). Are composers different? Historical evidence on conflict-induced migration (1816–1997). *European Review of Economic History*, *16*(3), 270–291. https://doi.org/10.1093/ereh/hes001

Borowiecki, K. J. (2013). Geographic clustering and productivity: An instrumental variable approach for classical composers. *Journal of Urban Economics*, *73*(1), 94–110. https://doi.org/10.1016/j.jue.2012.07.004

Borowiecki, K. J., & Dahl, C. M. (2021). What makes an artist? The evolution and clustering of creative activity in the US since 1850. *Regional Science and Urban Economics*, *86*, 103614. https://doi.org/10.1016/j.regsciurbeco.2020.103614

Borowiecki, K. J., & Graddy, K. (2021). Immigrant artists: Enrichment or displacement? *Journal of Economic Behavior & Organization*, *191*, 785–797. https://doi.org/10.1016/j.jebo.2021.09.007

Boschma, R. (2017). Relatedness as driver of regional diversification: A research agenda. *Regional Studies*, *51*(3), 351–364. https://doi.org/10.1080/00343404.2016.1254767

Boschma, R., Balland, P.-A., & Kogler, D. F. (2015). Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. *Industrial and Corporate Change*, *24*(1), 223–250. https://doi.org/10.1093/icc/dtu012

Boschma, R., & Capone, G. (2015). Institutions and diversification: Related versus unrelated diversification in a varieties of capitalism framework. *Research Policy*, *44*(10), 1902–1914. https://doi.org/10.1016/j.respol.2015.06.013

Boschma, R., Heimeriks, G., & Balland, P.-A. (2014). Scientific knowledge dynamics and relatedness in biotech cities. *Research Policy*, *43*(1), 107–114. https://doi.org/10.1016/j.respol.2013.07.009

Bosetti, V., Cattaneo, C., & Verdolini, E. (2015). Migration of skilled workers and innovation: A European perspective. *Journal of International Economics*, *96*(2), 311–322. https://doi.org/10.1016/j.jinteco.2015.04.002

Breschi, S., Lissoni, F., & Miguelez, E. (2017). Foreign-origin inventors in the USA: Testing for diaspora and brain gain effects. *Journal of Economic Geography*, *17*(5), 1009–1038. https://doi.org/10.1093/jeg/lbw044

Buringh, E. (2021). The population of European cities from 700 to 2000: Social and economic history. *Research Data Journal for the Humanities and Social Sciences*, *6*(1), 1–18. https://doi.org/10.1163/24523666-06010003

Bustos, S., Gomez, C., Hausmann, R., & Hidalgo, C. A. (2012). The dynamics of nestedness predicts the evolution of industrial ecosystems. *PLoS One*, *7*(11), e49393. https://doi.org/10.1371/journal.pone.0049393

Caragliu, A., Del Bo, C., de Groot, H. L. F., & Linders, G.-J. M. (2013). Cultural determinants of migration. *The Annals of Regional Science*, *51*(1), 7–32. https://doi.org/10.1007/s00168-012-0531-3

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, *19*(1), 22–64. https://doi.org/10.1086/209979

Chinazzi, M., Gonçalves, B., Zhang, Q., & Vespignani, A. (2019). Mapping the physics research space: A machine learning approach. *EPJ Data Science*, *8*(1), 33. https://doi.org/10.1140/epjds/s13688-019-0210-z

Cipolla, C. M. (1972). The diffusion of innovations in early modern Europe. *Comparative Studies in Society and History*, *14*(1), 46–52. https://doi.org/10.1017/S0010417500006496

Collins, H. M. (1974). The TEA Set: Tacit knowledge and scientific networks. *Science Studies*, *4*(2), 165–185. https://doi.org/10.1177/030631277400400203

Cortinovis, N., Xiao, J., Boschma, R., & van Oort, F. G. (2017). Quality of government and social capital as drivers of regional diversification in Europe. *Journal of Economic Geography*, *17*(6), 1179–1208. https://doi.org/10.1093/jeg/lbx001

De La Croix, D., & Licandro, O. (2015). The longevity of famous people from Hammurabi to Einstein. *Journal of Economic Growth*, *20*(3), 263–303. https://doi.org/10.1007/s10887-015-9117-0

Deegan, J., Broekel, T., & Fitjar, R. D. (2021). Searching through the haystack: The relatedness and complexity of priorities in smart specialization strategies. *Economic Geography*, *97*(5), 497–520. https://doi.org/10.1080/00130095.2021.1967739

Diodato, D., Morrison, A., & Petralia, S. (2022). Migration and invention in the age of mass migration. *Journal of Economic Geography*, *22*(2), 477–498. https://doi.org/10.1093/jeg/lbab032

Diodato, D., Neffke, F., & O'Clery, N. (2018). Why do industries coagglomerate? How Marshallian externalities differ by industry and have evolved over time. *Journal of Urban Economics*, *106*, 1–26. https://doi.org/10.1016/j.jue.2018.05.002

Elekes, Z., Boschma, R., & Lengyel, B. (2019). Foreign-owned firms as agents of structural change in regions. *Regional Studies*, *53*(11), 1603–1613. https://doi.org/10.1080/00343404.2019.1596254

Essletzbichler, J. (2015). Relatedness, industrial branching and technological cohesion in US metropolitan areas. *Regional Studies*, *49*(5), 752–766. https://doi.org/10.1080/00343404.2013.806793

Farinha, T., Balland, P.-A., Morrison, A., & Boschma, R. (2019). What drives the geography of jobs in the US? Unpacking relatedness. *Industry and Innovation*, *26*(9), 988–1022. https://doi.org/10.1080/13662716.2019.1591940

Fassio, C., Montobbio, F., & Venturini, A. (2019). Skilled migration and innovation in European industries. *Research Policy*, *48*(3), 706–718. https://doi.org/10.1016/j.respol.2018.11.002

Galenson, D. W. (2009). *Conceptual revolutions in twentieth-century art*. Cambridge University Press.

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

Guevara, M. R., Hartmann, D., Aristarán, M., Mendoza, M., & Hidalgo, C. A. (2016). The research space: Using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics*, *109*(3), 1695–1709. https://doi.org/10.1007/s11192-016-2125-9

Haskell, F. (1996/2000). *Maler und Auftraggeber: Kunst und Gesellschaft im italienischen Barock* (trans. A. Sahm). DuMont.

Hausmann, R., & Hidalgo, C. A. (2011). The network structure of economic output. *Journal of Economic Growth*, *16*(4), 309–342. https://doi.org/10.1007/s10887-011-9071-4

Henning, M. (2019). Time should tell (more): evolutionary economic geography and the challenge of history. *Regional Studies*, *53*(4), 602–613. https://doi.org/10.1080/00343404.2018.1515481

Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics*, *3*(2), 92–113. https://doi.org/10.1038/s42254-020-00275-1

Hidalgo, C. A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D. F., Morrison, A., Neffke, F., Rigby, D., Stern, S., Zheng, S., & Zhu, S. (2018). The principle of relatedness. In A. J. Morales, C. Gershenson, D. Braha, A. A. Minai, & Y. Bar-Yam (Eds.), *Unifying themes in complex systems IX* (pp. 451–457). Springer International Publishing. https://doi.org/10.1007/978-3-319-96661-8_46.

Hidalgo, C. A., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science (New York, N.Y.)*, *317*(5837), 482–487. https://doi.org/10.1126/science.1144581

Hornung, E. (2014). Immigration and the diffusion of technology: The Huguenot Diaspora in Prussia. *American Economic Review*, *104*(1), 84–122. https://doi.org/10.1257/aer.104.1.84

Hunt, J., & Gauthier-Loiselle, M. (2010). How much does immigration boost innovation? *American Economic Journal: Macroeconomics*, *2*(2), 31–56. https://doi.org/10.1257/mac.2.2.31

Jara-Figueroa, C., Jun, B., Glaeser, E. L., & Hidalgo, C. A. (2018). The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proceedings of the National Academy of Sciences*, *115*(50), 12646–12653. https://doi.org/10.1073/pnas.1800475115

Juhász, S., Broekel, T., & Boschma, R. (2021). Explaining the dynamics of relatedness: The role of co-location and complexity. *Papers in Regional Science*, *100*(1), 3–21. https://doi.org/10.1111/pirs.12567

Jun, B., Alshamsi, A., Gao, J., & Hidalgo, C. A. (2020). Bilateral relatedness: Knowledge diffusion and the evolution of bilateral trade. *Journal of Evolutionary Economics*, *30*(2), 247–277. https://doi.org/10.1007/s00191-019-00638-7

Kerr, S. P., Kerr, W., Özden, Ç, & Parsons, C. (2017). High-skilled migration and agglomeration. *Annual Review of Economics*, *9*(1), 201–234. https://doi.org/10.1146/annurev-economics-063016-103705

Laouenan, M., Bhargava, P., Eyméoud, J.-B., Gergaud, O., Plique, G., & Wasmer, E. (2022). A cross-verified database of notable people, 3500 BC–2018 AD. *Scientific Data*, *9*(1), 290. https://doi.org/10.1038/s41597-022-01369-4

Lewer, J. J., & Van den Berg, H. (2008). A gravity model of immigration. *Economics Letters*, *99*(1), 164–167. https://doi.org/10.1016/j.econlet.2007.06.019

Lissoni, F. (2018). International migration and innovation diffusion: An eclectic survey. *Regional Studies*, *52*(5), 702–714. https://doi.org/10.1080/00343404.2017.1346370

Lucchini, L., Tonelli, S., & Lepri, B. (2019). Following the footsteps of giants: Modeling the mobility of historically notable individuals using Wikipedia. *EPJ Data Science*, *8*(1), 36. https://doi.org/10.1140/epjds/s13688-019-0215-7

Menini, S., Sprugnoli, R., Moretti, G., Bignotti, E., Tonelli, S., & Lepri, B. (2017). RAMBLE ON: Tracing movements of popular historical figures. In Proceedings of the software demonstrations of the 15th conference of the European Chapter of the Association for Computational Linguistics (pp. 77–80). https://aclanthology.org/E17-3020

Miguelez, E., & Morrison, A. (2022). Migrant inventors as agents of technological change. *The Journal of Technology Transfer*. https://doi.org/10.1007/s10961-022-09927-z

Miguelez, E., & Noumedem Temgoua, C. (2020). Inventor migration and knowledge flows: A two-way communication channel? *Research Policy*, *49*(9), 103914. https://doi.org/10.1016/j.respol.2019.103914

Miguélez, E., & Moreno, R. (2013). Research networks and inventors' mobility as drivers of innovation: Evidence from Europe.

Regional Studies, *47*(10), 1668–1685. https://doi.org/10.1080/00343404.2011.618803

Miguélez, E., & Moreno, R. (2015). Knowledge flows and the absorptive capacity of regions. *Research Policy*, *44*(4), 833–848. https://doi.org/10.1016/j.respol.2015.01.016

Mitchell, S. (2019). London calling? Agglomeration economies in literature since 1700. *Journal of Urban Economics*, *112*, 16–32. https://doi.org/10.1016/j.jue.2019.05.002

Mokyr, J. (2005). Mobility, creativity, and technological development: David Hume, Immanuel Kant and the Economic development of Europe. In Session on 'Creativity and the economy', German Association of Philosophy. https://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=CF581541DADAFCD43FED5624F5F3BAE3?doi=10.1.1.524.4828&rep=rep1&type=pdf

Mokyr, J. (2017). *A culture of growth: The origins of the modern economy*. Princeton University Press. https://doi.org/10.1515/9781400882915.

Morrison, A. (2023). Towards an evolutionary economic geography research agenda to study migration and innovation. *Cambridge Journal of Regions, Economy and Society*, rsad013. https://doi.org/10.1093/cjres/rsad013

Moser, P., Voena, A., & Waldinger, F. (2014). German Jewish Émigrés and US invention. *American Economic Review*, *104*(10), 3222–3255. https://doi.org/10.1257/aer.104.10.3222

Neffke, F., Hartog, M., Boschma, R., & Henning, M. (2018). Agents of structural change: The role of firms and entrepreneurs in regional diversification. *Economic Geography*, *94*(1), 23–48. https://doi.org/10.1080/00130095.2017.1391691

Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography*, *87*(3), 237–265. https://doi.org/10.1111/j.1944-8287.2011.01121.x

Neffke, F., Otto, A., & Weyh, A. (2017). Inter-industry labor flows. *Journal of Economic Behavior & Organization*, *142*, 275–292. https://doi.org/10.1016/j.jebo.2017.07.003

Nunn, N. (2021). History as evolution. In A. Bisin & G. Federico (Eds.), *The handbook of historical economics* (pp. 41–91). Elsevier. https://doi.org/10.1016/B978-0-12-815874-6.00010-1

OECD. (2021). *International migration outlook 2021*. OECD. https://doi.org/10.1787/29f23ae9d-en.

Oevermann, U., Süssmann, J., & Tauber, C. (Eds.). (2007). *Die Kunst der Mächtigen und die Macht der Kunst: Untersuchungen zu Mäzenatentum und Kulturpatronage*. Akademie Verlag.

O'Hagan, J., & Borowiecki, K. J. (2010). Birth location, migration, and clustering of important composers: Historical patterns. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *43*(2), 81–90. https://doi.org/10.1080/01615441003729945

O'Hagan, J., & Hellmanzik, C. (2008). Clustering and migration of important visual artists: Broad historical evidence. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *41*(3), 121–136. https://doi.org/10.3200/HMTS.41.3.121-136

Ottaviano, G. I. P., & Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, *10*(1), 152–197. https://doi.org/10.1111/j.1542-4774.2011.01052.x

Petralia, S., Balland, P.-A., & Morrison, A. (2017). Climbing the ladder of technological development. *Research Policy*, *46*(5), 956–969. https://doi.org/10.1016/j.respol.2017.03.012

Pinheiro, F. L., Hartmann, D., Boschma, R., & Hidalgo, C. A. (2021). The time and frequency of unrelated diversification. *Research Policy*, 104323. https://doi.org/10.1016/j.respol.2021.104323

Poncet, S., & de Waldemar, F. S. (2015). Product relatedness and firm exports in China. *The World Bank Economic Review*, *29*(3), 579–605. https://doi.org/10.1093/wber/lht037

Putterman, L., & Weil, D. N. (2010). Post-1500 population flows and The long-run determinants of economic growth and inequality. *The Quarterly Journal of Economics*, *125*(4), 1627–1682. https://doi.org/10.1162/qjec.2010.125.4.1627

Rephann, T. J., & Vencatasawmy, C. P. (2000). Determinants of the spatial mobility of immigrants: Evidence from Sweden. *The Review of Regional Studies*, *30*(2), 189–213. https://doi.org/10.52324/001c.8624

Rigby, D. L. (2015). Technological relatedness and knowledge space: Entry and exit of US cities from patent classes. *Regional Studies*, *49*(11), 1922–1937. https://doi.org/10.1080/00343404.2013.854878

Rodríguez-Pose, A. (2018). The revenge of the places that don't matter (and what to do about it). *Cambridge Journal of Regions, Economy and Society*, *11*(1), 189–209. https://doi.org/10.1093/cjres/rsx024

Rodríguez-Pose, A., Terrero-Dávila, J., & Lee, N. (2023). Left-behind versus unequal places: Interpersonal inequality, economic decline and the rise of populism in the USA and Europe. *Journal of Economic Geography*, lbad005. https://doi.org/10.1093/jeg/lbad005

Schich, M., Song, C., Ahn, Y.-Y., Mirsky, A., Martino, M., Barabási, A.-L., & Helbing, D. (2014). A network framework of cultural history. *Science*, *345*(6196), 558–562. https://doi.org/10.1126/science.1240064

Scoville, W. C. (1952a). The Huguenots and the diffusion of technology. I. *Journal of Political Economy*, *60*(4), 294–311. https://doi.org/10.1086/257238

Scoville, W. C. (1952b). The Huguenots and the diffusion of technology. II. *Journal of Political Economy*, *60*(5), 392–411. https://doi.org/10.1086/257274

Serafinelli, M., & Tabellini, G. (2022). Creativity over time and space: A historical analysis of European cities. *Journal of Economic Growth*, *27*(1), 1–43. https://doi.org/10.1007/s10887-021-09199-6

Trippl, M. (2013). Scientific mobility and knowledge transfer at the interregional and intraregional level. *Regional Studies*, *47*(10), 1653–1667. https://doi.org/10.1080/00343404.2010.549119

Trippl, M., & Maier, G. (2011). Knowledge spillover agents and regional development. In P. Nijkamp, & I. Siedschlag (Eds.), *Innovation, growth and competitiveness* (pp. 91–111). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-14965-8_5

Uhlbach, W.-H., Balland, P.-A., & Scherngell, T. (2022). Public R&D funding and new regional specialisations: The contingent role of technological relatedness. *Industry and Innovation*, *29*(4), 511–532. https://doi.org/10.1080/13662716.2022.2043147

Waldinger, F. (2010). Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany. *Journal of Political Economy*, *118*(4), 787–831. https://doi.org/10.1086/655976

Waldinger, F. (2012). Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany. *The Review of Economic Studies*, *79*(2), 838–861. https://doi.org/10.1093/restud/rdr029

Williams, A. M. (2006). Lost in translation? International migration, learning and knowledge. *Progress in Human Geography*, *30*(5), 588–607. https://doi.org/10.1177/0309132506070169

World Bank. (2019). *Leveraging economic migration for development: A briefing for the world bank board*.

Yu, A. Z., Ronen, S., Hu, K., Lu, T., & Hidalgo, C. A. (2016). Pantheon 1.0, a manually verified dataset of globally famous biographies. *Scientific Data*, *3*(1), 150075. https://doi.org/10.1038/sdata.2015.75

Zhu, S., He, C., & Zhou, Y. (2017). How to jump further and catch up? Path-breaking in an uneven industry space. *Journal of Economic Geography*, *17*(3), 521–545. https://doi.org/10.1093/jeg/lbw047