Mathias Czaika, Heidrun Bohnet and Akira Soto-Nishimura

Spatial and categorical dependence of European migration flows

Deliverable 5.2

QuantMig has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870299.

History of changes

Version	Date	Changes
1.0	15 March 2021	Issued for Consortium Review
1.1	30 March 2021	First version submitted as official deliverable to the EC

Suggested citation

Czaika, M, H Bohnet and A Soto-Nishimura (2021) Spatial and categorical dependence of European migration flows. QuantMig Project Deliverable 5.2.

Dissemination level

PU Public

Acknowledgments

This document reflects the authors' view and the Research Executive Agency of the European Commission are not responsible for any use that may be made of the information it contains. We are very grateful for useful comments and advice from Marta Bivand Erdal, Jakub Bijak, and Constantinos Melachrinos.

Cover photo: iStockphoto.com/Guenter Guni

Table of Contents

1. Introduction

Europe has become a prime destination for migrants from all over the world. In 2019 alone, about 2.5 million migrants, of whom about 230,000 received asylum, entered one of the 28 EU member states (Eurostat 2020). Migrants generally move for very different reasons and motivations, but the regulatory and legal framework for entering the EU foresees basically four main pathways or legal categories, namely for labour, family reunion, education, and asylum. Yet, how have migration flows within these legal categories evolved over recent periods? And, to what extent are flows between these legal pathways mutually interdependent, either in terms of a *categorical dependence* – flows of different legal pathways between two countries are linked – or, in terms of *spatial dependence*, i.e. flows of a certain legal pathway across similar or nearby countries of origin or destination are linked?

Although many studies addressing the role of migration drivers (including migration policies) of flows of certain legal categories of migration have already been put forward, empirical evidence remains almost absent when it comes to the interlinkages between the different entry types of European immigration. While it has been pointed out that 'fixed' migration flow categories (Crawley and Skleparis 2018, p. 52), such as of asylum, refugee or labour migrant or the "forced-voluntary dichotomy's stickiness" (Erdal and Oeppen 2018), do not fully consider the complexity of different migration drivers (Bakewell 2008) as well as "their shifting significance for individuals over time and space" (Crawley and Skleparis 2018, p. 48, Zetter 1991), we are not aware of a study that empirically assessed on a larger scale how, and to what extent, migration flows within and between different legal categories are actually interconnected.

Moreover, previous studies have failed to highlight and analyse simultaneously the spatial and categorical interconnectedness of multiple migration categories. Even though categorical and spatial 'substitution effects' (de Haas 2011) of migration flows are presumed to exist, for instance in terms of a categorical deflection of asylum seekers into irregularity (Czaika and Hobolth 2016), whether and to what extent different entry categories are indeed spatially and categorically interlinked has not received thorough empirical attention on a broader scale.

This paper fills this gap by simultaneously analysing migration flow data on asylum-seeking, labour and family migration, as well as student mobility to explore the extent to which different legal pathways for migration are spatially and categorically interconnected and mutually reinforcing, or not. We further explore the extent to which different types of migration flows towards Europe are indeed spatially and categorically clustered and how these clusters have changed over time. We consider both space and migration flow categories together, thus, focusing on cross-cutting clusters.

Over the past three decades, major political and economic developments, such as the EU enlargement in 2004, the financial and economic crisis in 2007-8, the rise in refugee and other migration inflows around 2015, as well as the Brexit referendum in 2016 and its implementation had significant ramifications on the composition of migration flows into Europe. In addition, various economic, societal or political developments and events in Europe, but also in countries of origin, have led to shifts in global incentive and opportunity structures for migrating towards Europe.

We argue here that these contextual shifts affected not only the overall number of migrants, but more importantly, the composition of EU immigration. A changing social, political or economic environment within the EU and its member states as well as outside the EU may have led to categorical – for example, humanitarian migrants entering rather on a study visa, or labour migrants entering through the family migration route – and spatial shifts – for example, asylum seekers

targeting country A rather than country B as EU destination – in migration flows of third country nationals into the EU over the last three decades. Consequently, we hypothesize that migration flows are not only geographically and dynamically interconnected but also that migrants can and, in fact, do switch between different legal entry categories leading to so-called 'categorical dependence' in the flows of migrants who enter the EU on a work, family, study, or humanitarian visa.

To test this, we first explore bilateral migrant stocks (since 1990) and flow data (since 2008) from 171 countries of origin into the 28 European Union member states, i.e. including the United Kingdom. Second, for the period since 2008 (earlier data of this type is not available), we use data on first residence permits issued for reasons of asylum, labour, family, and education to analyse the extent to which migration flow categories are geographically and spatially interconnected. Our key hypothesis is based on the assumption that bilateral migration flows, that is migration of people from country *i* to country *j*, are not independent from migration flows from or to other countries, but spatially dependent. We further hypothesize that migration flows of a certain legal category, that is migration for reasons of labour, family, education or asylum, are not independent from flows within other legal categories from country *i* to country *j*, thus testing categorical dependence of bilateral migration flows.

The implication of spatial and categorical dependence is that migration clusters are formed which are characterised by distinct compositions of the migrant population. Based on a cluster evolution analysis (Ramon-Gonen & Gelbard 2017), we identify similar migration clusters and patterns and detect dynamic changes in cluster size and characteristics. This is mainly due to the fact that drivers and uncertainties affecting the decision to migrate vary for different types of flows. For example, while family migration is a relatively stable type of migration as largely dependent on the size of diasporas, differences in quality of life between origin and destination, the recruitment of third country nationals as workers (or as students), depends mostly on economic cycles (and study opportunities). In comparison, conflict-induced humanitarian migration is the most uncertain and hence most volatile type of migration flow (Bijak and Czaika 2020).

This paper demonstrates that spatio-categorical changes in one category of migration does affect the evolution of other migration flows and their direction and composition at both the origin and the European destination side. These spatial, categorical, and intertemporal interdependencies have implications not only for our understanding of the complexity of international migration processes, but also on the scope and limitations of migration policy in influencing migration flows in certain categorical ways and spatial directions. This calls for more comprehensive approaches for migration policy making to be effective in influencing migration outcomes.

Our findings show that changes in one entry category can have significant implications for the use of other legal entry categories and that, as a consequence, migration flows are indeed spatially clustered. Since these spatial and categorical dependence effects can be significant in size, governments must be aware of these interdependencies of different migration flow categories when designing policies with the aim of influencing migration flows in a certain categorical or spatial direction. If ignored, unintended policy effects or policy failure are the result.

The remainder of this paper is organized as follows. Section 2 explores global patterns of European immigration based on a comprehensive assessment of bilateral migrant stocks data since 1990. Section 3 employs a dynamic cluster analysis identifying and exploring clusters of countries of origin that share similar yet changing compositions of their emigrant population. Section 4 identifies spatial and categorical dependence of four different migration flow categories (or, legal pathways) into Europe, namely, asylum, family, education, and labour. We conclude with some policy implications of our findings.

2. Europe since 1990 - a migration hub in transition

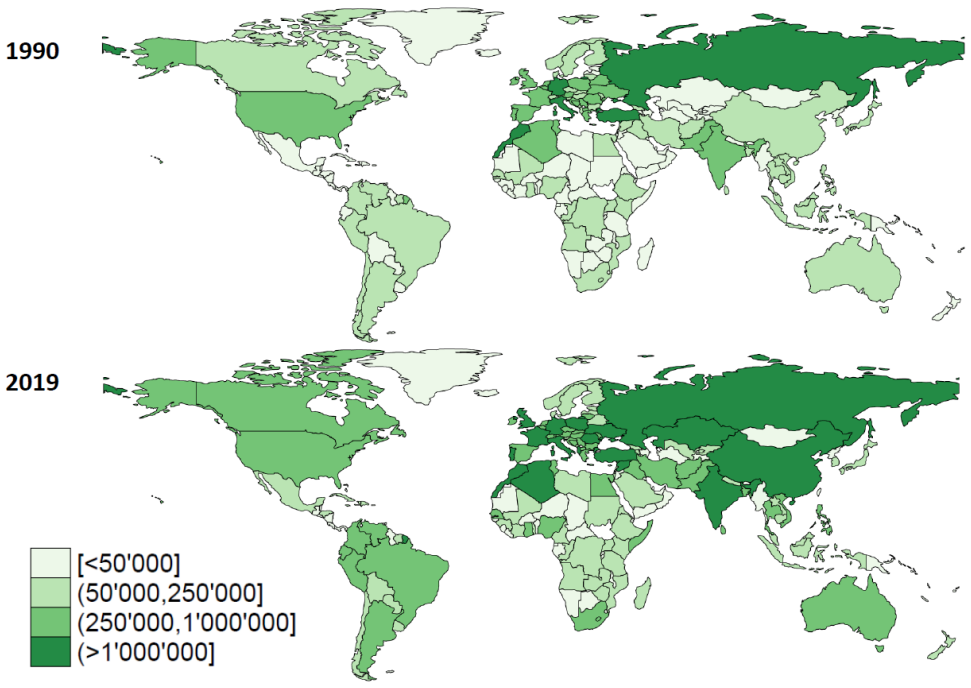
Using stock data from 1990-2019 data of third-country nationals in the EU by UNDESA (2020), we show in the following how migration patterns have changed since 1990. This will serve as a base for identifying critical changes and trends of past migration flows into the EU to be able, in a next step, to determine specific cluster of migration flows.

As pointed out by Czaika and de Haas (2014) and IOM (2020), on the global scale, international migration has both accelerated and diversified over the past decades. We know that international migrants travel over increasingly long distances (World Bank 2018), and immigration has made host societies much more diverse in terms of origins and the composition of their immigrant population (Vertovec 2007). Even though the average migration rate, approximated by the proportion of the world population that is residing outside their country of birth or citizenship, has been surprisingly stable over recent decades, the absolute increase in the stock of international migrants by about 120 million between 1990 and 2019 to more than 275 million could suggest numerous shifts and transitions at more regional levels.

Amid this global migration transition, Europe has become a gravity centre. Over centuries a continent of emigration, more recently Europe has turned into a, or rather *the* major global destination for migrants from all corners of the globe. In fact, Europe turned from a net emigration to a net immigration continent already more than half a century ago, but the complex and accelerating migration transitions Europe has seen since the early 1990s have been at a different level and consolidated Europe's uncontested role as a global migration hub.

Between 1990 and 2020, driven by an annual inflow of about 2-3 million non-EU citizens, Europe's non-EU migrant population has increased from about 16.5 million to 38 million third-country nationals. In relation to its total resident population of 513 million in 2019, about 7.4 per cent of the EU residents are born outside the EU (UNDESA 2020). Since 1990, this amounts to about 21 million new foreign-born residents, or an increase of more than 130 per cent.

Figure 1: Stocks of third country nationals (intra-EU nationals) in the EU by non-EU (EU) country of origin, 1990 vs. 2019

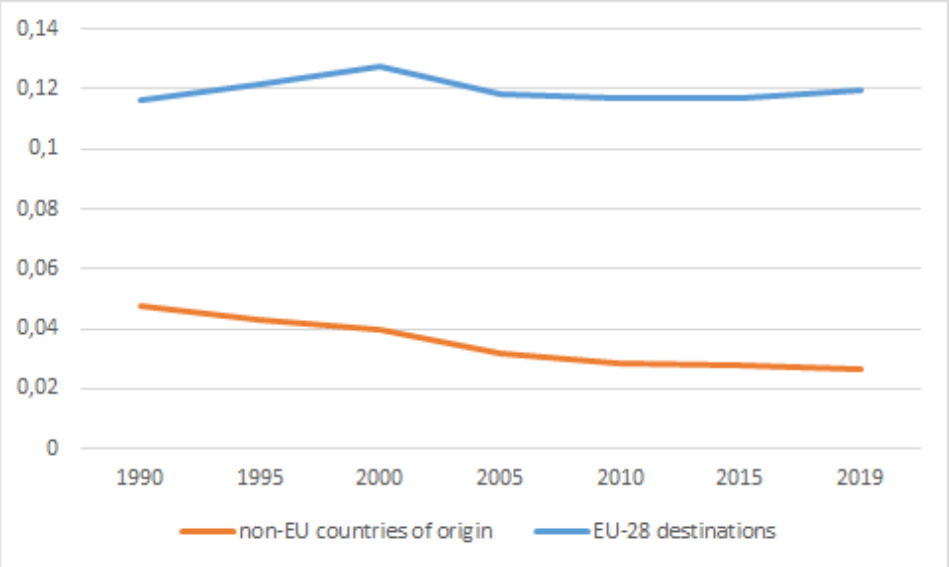


Source: Own elaboration based on data provided by UNDESA (2020).

Figure 1 maps the major countries of origin in terms of the absolute number of migrants in the EU. In 2019, a quarter (about 10 million) of third country nationals in the EU are either born or citizens of one of the top five countries of origin (Morocco, Turkey, Russia, Algeria, and India), while the other 75 per cent (or 28 million migrants) are originating from 195 countries of the world.

The concentration of the EU migrant population in terms of the composition of the stock of third country nationals can be measured by the Herfindahl index H, defined as the sum of squared proportions, which indicates the level of concentration by which the EU immigrant population $\sum_i \sum_{j=1}^{28} M_{ij}$ originates from a diverse set of non-EU countries i: $H^{origin} = \sum_i \left(\frac{\sum_{j=1}^{28} M_{ij}}{\sum_i \sum_{j=1}^{28} M_{ij}} \right)^2$. Since 1990, the diversity of the non-EU immigrant population, measured by this Herfindahl index of concentration, is decreasing from 0.05 to slightly over 0.02 in 2019 (Figure 2). This implies that Europe as a continent has not only turned into global migration hub, but the non-European immigrant population has also become for heterogeneous in terms of a more even spread over a growing number of countries of origin. At the same time, immigration of more than 20 million migrants over the past three decades has not led to a 'de-concentration' of the non-EU migrant population *within* the EU. Rather the contrary, as Figure 2 also shows, the Herfindahl index of concentration of the non-EU immigrant population $H^{EU} = \sum_{j=1}^{28} \left(\frac{\sum_i M_{ij}}{\sum_i \sum_{j=1}^{28} M_{ij}} \right)^2$ across the 28 EU member states has slightly increased from 0.116 to 0.119 between 1990 and 2019. This suggests that even though the EU immigrant population comes from an increasingly diverse array of origin countries, its spread within the EU has not increased of the past three decades. In other words, a continuously small number of EU destinations attract the bulk of EU immigrants from an increasingly diverse array of origin countries, which largely mirrors long-term trend in global migration patterns (Czaika de Haas 2013).

Figure 2: Concentration (Herfindahl index) of EU migrant population (third country nationals) across non-EU countries of origin and EU destination, 1990-2019



Source: Own elaboration, based on bilateral stocks data from UNDESA (2020).

The stagnating (high) level of immigrant concentration within the European Union can partly be explained by the fact that a significant number of non-EU nationals are not only coming from a few major countries of origin, but they are also linked to some specific European destination countries.

The significance of relatively few bilateral (origin-destination) migration corridors in comparison to the total is evidenced by the fact that the 38 million third country nationals are spread very unevenly across the 5600 bilateral corridors that exist between about 200 countries of origin and 28 EU destination countries. In fact, the 18 largest bilateral migrant corridors listed in Table 1, that is 0.3 per cent of all corridors connecting a non-EU origin country with an EU destination country comprise about one third (~12.5 million) of all third-country nationals in the EU.

Table 1: Top 18 bilateral corridors (>250,000 migrants) in 2019 between non-EU migrant origin and EU destination country, percentage change since 1990

Corridor	in 2019	in 1990	Change in %
Argentina ⟶ Spain	259,946	42,923	506%
Turkey ⟶ France	327,508	241,148	36%
Colombia ⟶ Spain	367,816	12,548	2831%
Bosnia-Herzegovina ⟶ Croatia	373,838	312,821	20%
Ecuador ⟶ Spain	415,310	3,734	11022%
Albania ⟶ Greece	426,449	63,963	567%
Tunisia ⟶ France	427,897	276,216	55%
Morocco ⟶ Italy	450,557	169,285	166%
Albania ⟶ Italy	475,196	44,935	958%
Syria ⟶ Germany	589,628	15,330	3746%
Pakistan ⟶ United Kingdom	605,016	228,321	165%
Morocco ⟶ Spain	711,792	133,341	434%
India ⟶ United Kingdom	917,686	399,526	130%
Kazakhstan ⟶ Germany	940,296	10,199	9119%
Russia ⟶ Germany	999,162	77,318	1192%
Morocco ⟶ France	1,020,162	713,987	43%
Turkey ⟶ Germany	1,531,333	1,586,121	-3%
Algeria ⟶ France	1,575,528	788,914	100%

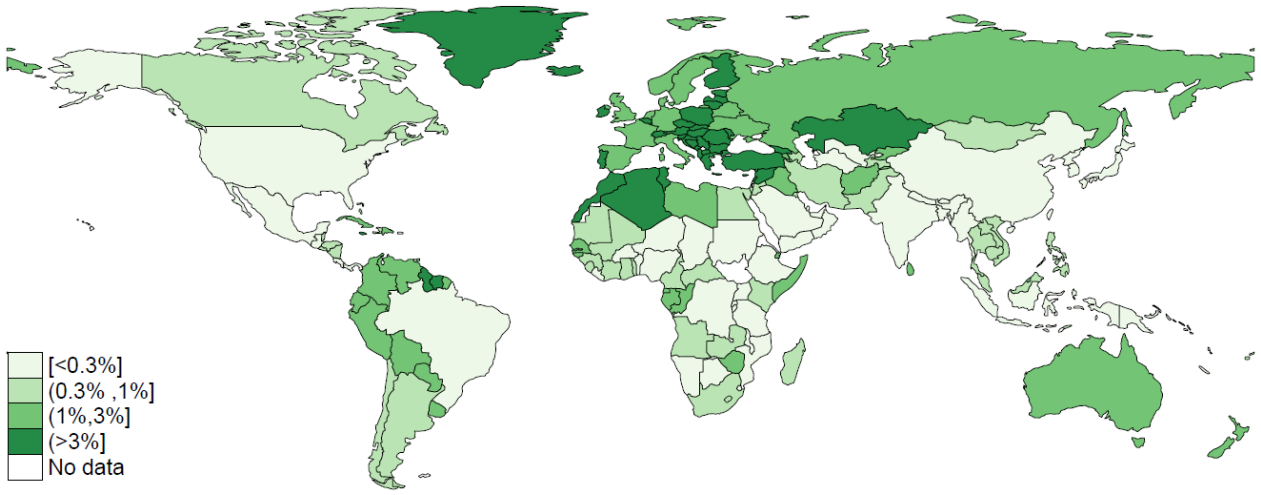
Source: Own elaboration based on data from UNDESA (2020).

Putting this into a temporal context shows that the concentration of immigration to Europe within a small number of bilateral corridors has disproportionally increased over the past three decades. While the overall population of third country nationals residing in the EU has increased by 130 per cent since 1990, the migrant population within these top 18 corridors has increased by 142 per cent, and if we ignore the longstanding and stagnating Turkish-German corridor, it has even increased by more than a factor of three (by 208 per cent).

Yet not only the high concentration within a few bilateral (non-EU / EU) corridors is a major feature of European immigration, but also that for some non-EU countries of origin the EU as a whole is hosting a significant and growing percentage of their population (Figure 3). In 2019, for instance, 42 non-EU countries had more than 3 per cent of their home-born population residing in one of the 28 EU member states, and another 37 countries had between 1 and 3 per cent of their population living

in the EU.

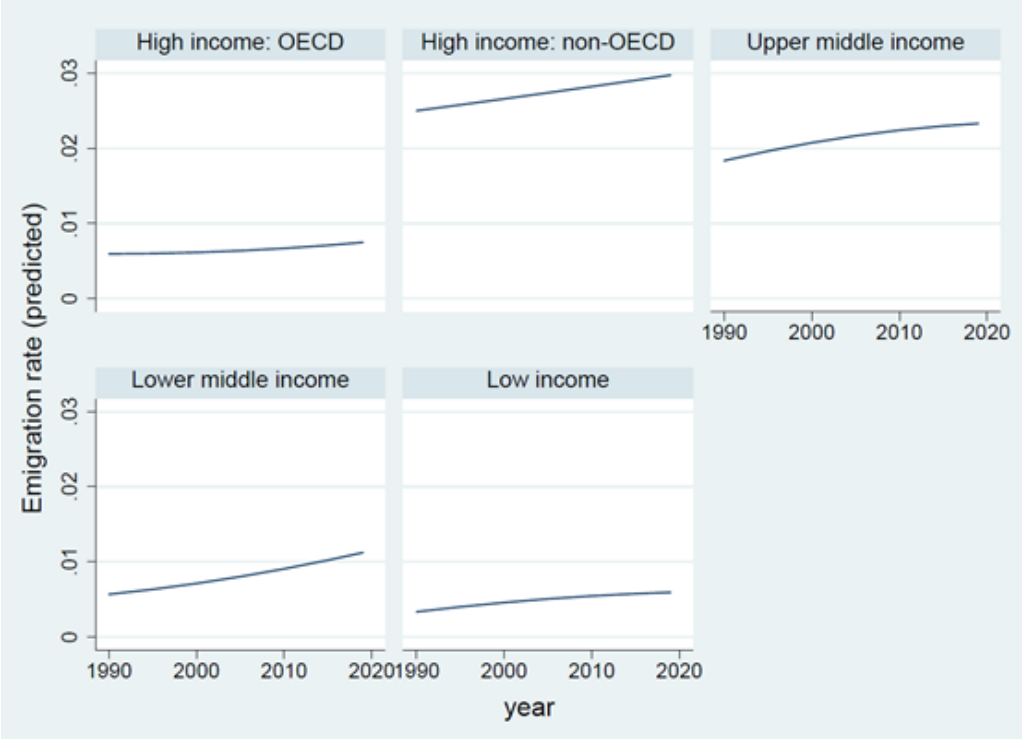
Figure 3: Emigration propensity (emigration rate) of non-EU migrant population towards the EU, in % of home country population, 2019



Source: Own elaboration based on data provided by UNDESA (2020).

Once emigration rates towards the EU are clustered by origin countries' level of economic development, we find the inverted U-shaped association between development and emigration confirmed (Clements 2014). While the average emigration rate of non-EU OECD countries and of the poorest countries of the world is well below one per cent, emigration to the EU from middle-income countries, and most significantly, from high-income non-OECD countries, is three to four times higher (Figure 4).

Figure 4: Emigration propensity (migration rate) to EU in non-EU origin countries, in % of home country population, clustered by level of economic development, 1990-2019



Source: Own elaboration based on data provided by UNDESA (2020) and World Bank (2020).

In sum, we can identify three major trends of EU migration since 1990:

1. Non-EU migration into the EU has increased significantly (more than 130 per cent) over the past three decades. Europe has become a prime destination for migrants worldwide and migrant stocks have diversified.
2. While over the years the non-EU migrant population has been spreading out more evenly across a larger number of countries of origin, the growing non-EU immigrant population remains concentrated within a few EU member states.
3. An agglomeration of relatively few bilateral migration corridors exists. One third of all third-country nationals in the EU come from just 18 corridors connecting a non-EU origin country with an EU destination country. Among them, for example, Argentina-Spain and Syria-Germany.

In the following, we analyse how the non-EU immigrant population has been growing through the inflow of third country nationals and how these flows have been clustered spatially and categorically by their legal pathways.

3. The evolution of spatio-categorical clusters in European migration since 2008

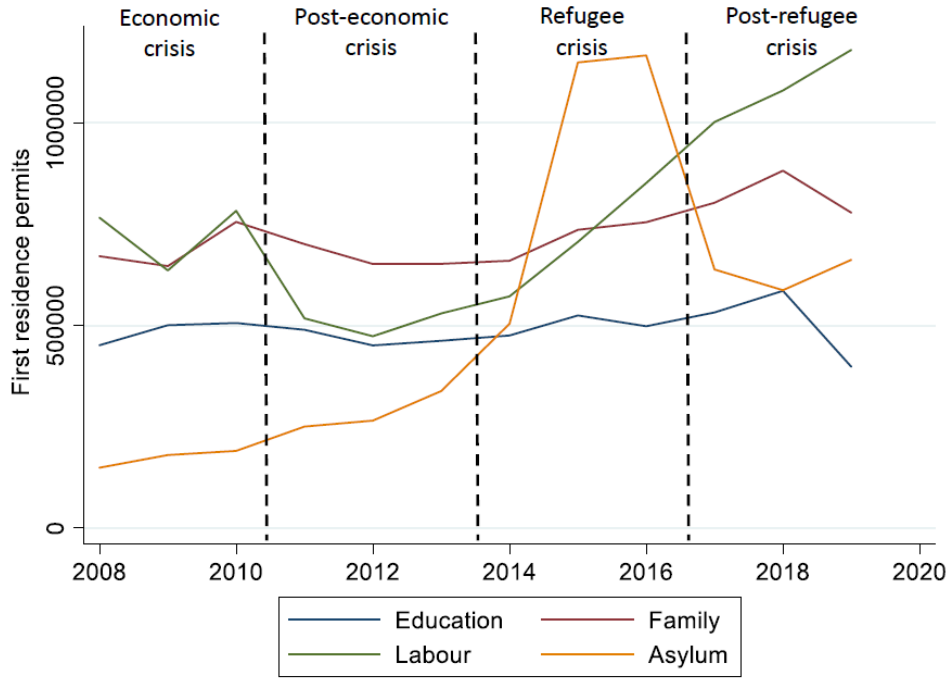
3.1 Data and Background

The previous analysis has shown some key features of dynamically changing geographical patterns over the last three decades in the stock of third country nationals in the European Union. Migrants generally move to Europe for very different reasons and motivations, but the regulatory and legal framework for entering the EU foresees basically four main legal pathways: labour, family reunification, education, and asylum. According to these politically motivated and legally constructed categories, Eurostat provides aggregate data for each EU member state on the annual number of first residence permits issued to third-country nationals by the type of permit since 2008.

Between 2008 and 2019, the EU-28 member states have issued about 30 million first residence permits across these four legal categories. While numbers have been relatively stable over this time period for those who have entered the EU through the family or education-based entry route (about 500-700 thousand first residence permit recipients per year), both asylum and labour migration have been fluctuating significantly (Figure 5).

Pivotal changes in the numbers of immigrants can be linked to major structural changes over this time period, which is why we have split this 12-year time frame into four periods. First, the economic crisis (2008-2010) which has hit most European economies and many non-European countries of origin most severely and led to deep but in most cases rather short recessions, followed by a period of economic recovery in the early 2010s. This post-economic crisis period (2011-2013) was followed in Europe by the so-called refugee crisis, emerging and culminating in 2014 to 2016. After an unprecedented number of asylum applications submitted in many European destination countries, asylum numbers came significantly down in 2017 heralding a post-refugee crisis period (2017-2019).

Figure 5: Total number of first residence permits issued by legal category (reason) in all EU-28 countries to third country nationals, 2008-2019



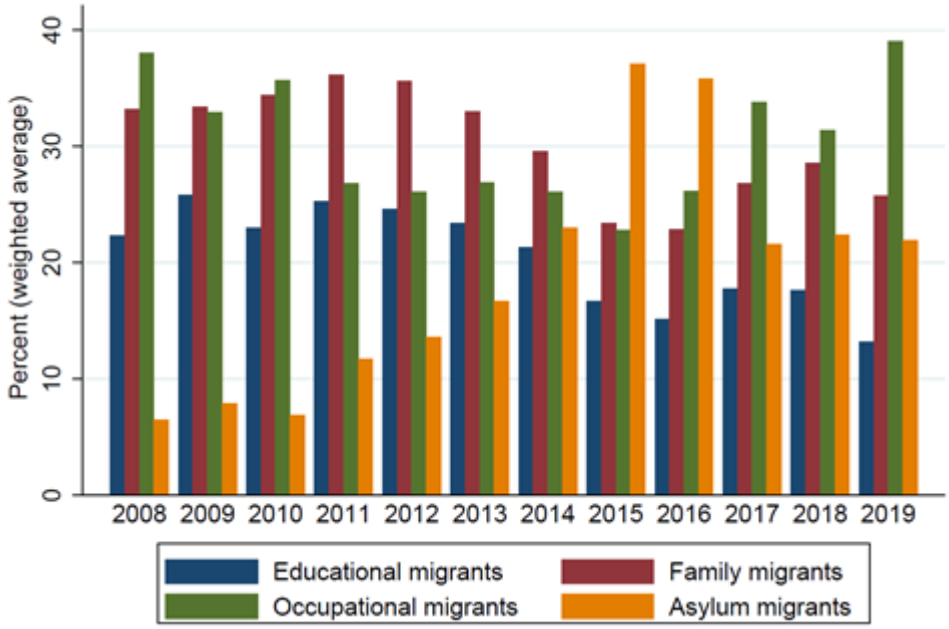
Source: Own elaboration based on data provided by Eurostat (2020).

While during and after the economic crisis immigration for labour has been significantly declining and only reached pre-crisis levels around the year 2015, asylum migration has been very low at the end of the first decade but increased rapidly between 2013 and 2016 before returning to levels before the so-called refugee crisis. These changes seem not surprising if considering the economic shocks and the developments of 2015 regarding asylum.

While citizens of almost all countries in the world access the European Union through all four legal pathways, most countries of origin have very distinct patterns when it comes to the main routes. Figure 6 displays the population-weighted average percentages across all non-EU countries of origin admitted for each of the four legal categories. While the average percentage (of all emigrants to the EU) that has left a country of origin for seeking asylum in the EU did not exceed 10 to 15 per cent before the refugee crisis, this average percentage increased to more than 30 per cent of all legal migration from a country of origin to the EU during the years 2014-2016.

On the other hand, the population-weighted average percentage of emigrants who are moving into the EU on a family visa is relatively stable at around 30 per cent, or on a study visa at around 20 per cent, even though this percentage has slightly decreased for the more recent past. The percentage of emigrants entering the EU on a work visa has, as one would expect, been declining since the outbreak of the global financial and economic crisis in 2007/08 but reached a long-term low only in 2015 when only about one fifth of an origin country's EU migrants entered the EU on a work visa (Figure 6).

Figure 6: Composition of migration flows into EU of third country nationals by different legal entry pathways, 2008-19



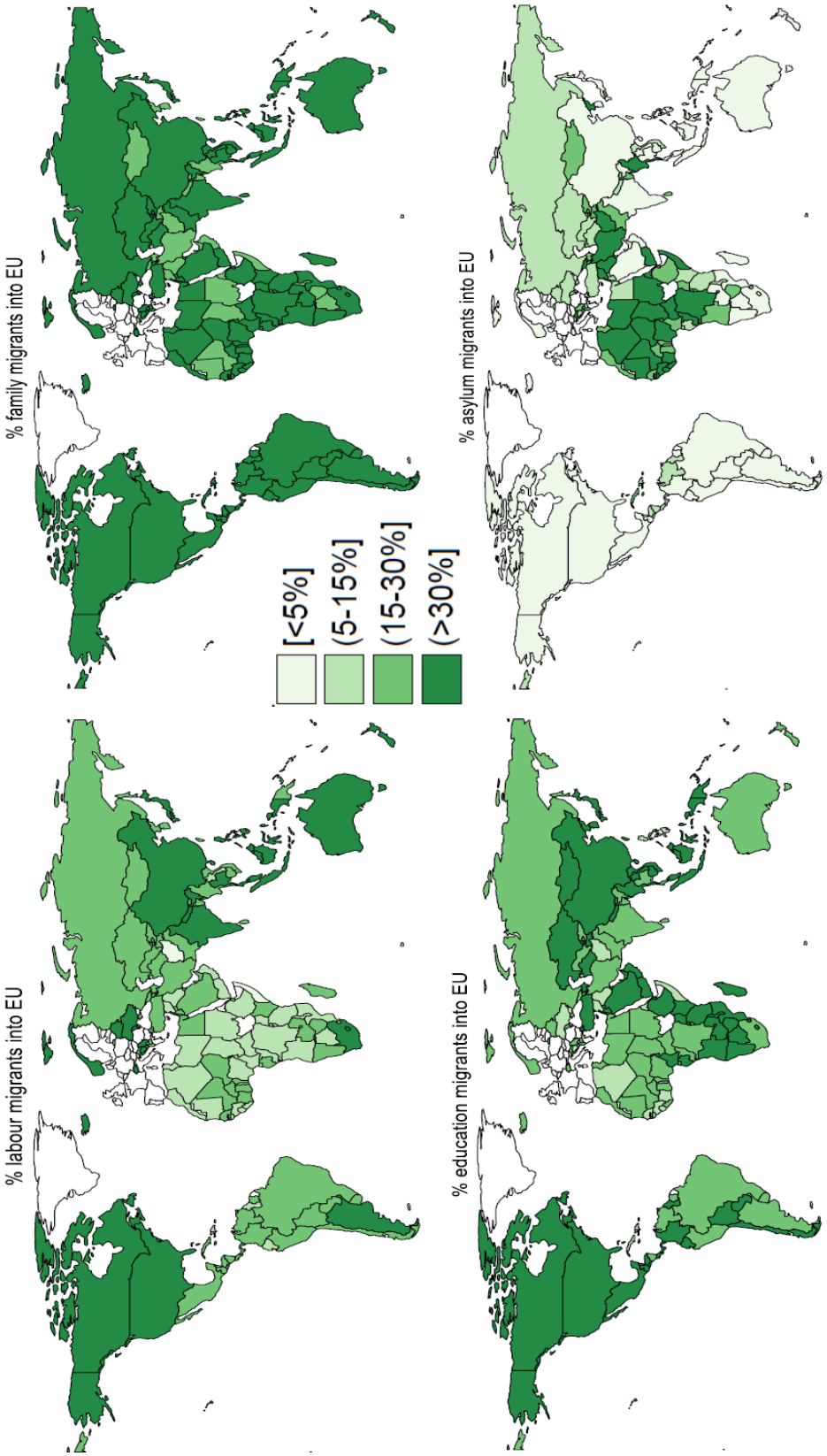
Source: Own elaboration based on data provided by Eurostat (2020).

The composition of bilateral migration flows into the EU by the type of the first residence permit varies significantly across countries of origin (Figure 7). In 85 non-EU countries, more than 30 per cent of all EU migrants have been migrating on a study visa. For countries like Oman, Qatar or Taiwan, more than half of all EU migrants move on a study visa (Oman: 61 per cent, or 16,000 students; Qatar: 58 per cent, 14,000 students; Taiwan: 55 per cent, 70,000 students). For more than 130 countries, emigration to the EU for family reasons has been the dominating mode of entry. For instance, some island states (Dominican Republic: 118,000; Cape Verde: 38,000, Cuba: 81,000) having more than two thirds of their EU migrants entering the EU on a family visa.

On the other side, only 34 non-EU emigration countries had their EU migrants entering on a work visa with the Philippines (48 per cent), India (43 per cent) and the Ukraine (42 per cent) as top labour-sending countries. For conflict-ridden countries of Somalia, Eritrea or Afghanistan, asylum is the dominating mode of entering the EU. More than three quarters of all migrants from these countries who have entered the EU between 2008 and 2019 have used the humanitarian route and received a protection status as legitimation of a first residence permit.

While most non-EU countries of origin have a dominating migration flow category, the composition of all four entry categories can vary significantly across countries, as illustrated in Table 2 for the example of the top 10 EU migrant origins between 2008 and 2019. While, for the Ukraine, family and labour migration have been equally balanced including about 80 per cent of all migrants in the EU, for Morocco family migration is the primary mode of entry with education and asylum being secondary forms of entering the EU.

Figure 7: Composition of EU migrant-origin flows by type of first residence permit (2008-2019)



Source: Own elaboration based on first residence data as provided by Eurostat (2020).

Table 2: Top 10 countries of origin of third country nationals in EU, by type of first residence permit, 2008-2019

Country	First residence permits issued to third country nationals				
	Total	Education	Family	Labour	Asylum
Ukraine	4'020'551	13%	37%	42%	8%
China	1'915'747	32%	30%	34%	4%
India	1'710'948	22%	31%	43%	4%
Syria	1'458'620	10%	28%	8%	58%
Morocco	1'388'049	16%	52%	13%	20%
USA	1'260'942	33%	33%	34%	0%
Russia	806'993	18%	44%	26%	15%
Pakistan	802'343	19%	36%	18%	29%
Brazil	723'530	27%	43%	30%	1%
Afghanistan	720'997	9%	22%	4%	73%

Source: Own elaboration based on first residence data as provided by Eurostat (2020). Corresponding to the shading in Figure 7, dark green (medium green; light green) indicates percentages of EU emigrants using the respective legal pathway being > 30% (between 15 and 30per cent; between 5 and 15 per cent).

3.2 A dynamic cluster analysis

The following analysis disentangles the spatial and categorical clustering process of EU immigrant a little further. Cluster analysis is a segmentation or taxonomy analysis that has been used in multiple research fields including genetics (Fürtges et al. 2017), media use (Shensa et al. 2018), and welfare support (Thomann & Rapp 2018), to name just a few. Within the field of migration there have been several applications. Škuflić et al. (2018), for example, clustered EU countries based on their economic performance and migration stocks and flows identifying three clusters for the period 2006 to 2015. For instance, member states with a relatively high average income, a relatively low unemployment rate, a relatively large net migration level and a relatively high stock of foreign citizens were stable members of the same clusters. Manafi et al. (2017) did a similar cluster analysis using a mix of economic, social, and migration indicators to cluster EU and EFTA countries. They also observed three clusters and noted that the cluster composition was relatively persistent over time.

Even though cluster analyses are often static, clusters can also be compared over time by using two alternative methods. One way is to compare shared objects, or countries in our analysis. Two clusters that contain many shared objects can be defined as similar (Wagner & Wagner 2007). The other method is to compare cluster characteristics. If two clusters share the same characteristics they are defined as similar. Typical tests for similarity between clusters rely on distance calculations and statistical tests, such as the t-test, Mann-Whitney, or Chi-square.
We compare cluster characteristics using the moving average of cluster centroid technique developed by Ramon-Gonen and Gelbard (2017). In order to elaborate on intertemporal changes in compositional migration features we employ a dynamic cluster analysis on both non-EU countries. We hereby identify similarities between clusters of countries and trace changes in cluster types over time. After aggregating the first residence permit data along four three-year time periods (2008-2010, 2011-2013, 2014-2016, and 2017-2019), corresponding to the major shifts in migration-relevant

developments in the EU and beyond (in particular the economic and refugee crisis), we split the data into four separate datasets, one for each time period, and calculate the proportion of each legal category per country and per time period. We then standardized the proportions to obtain z-scores and employ the k-means clustering algorithm.

In the analysis of origin country clusters, the number of clusters identified was always four for all periods, while for the analysis of EU destination country clusters the recommended number of clusters varied between four to eight over the four three-year periods. Reasons why there are more clusters and more changes in the number of destination clusters over the four time period could be that the EU destination countries are relatively more heterogeneous in the composition of migrant categories but also because the EU destination countries became slightly less similar in their migrant composition.

Once the number of clusters was determined for each time period, we identified dynamic changes in clusters that is whether clusters in one time period were similar in characteristics in the following time period. Each cluster in time period two (T2, 2011-2013), represented by the centroid vector and the distance in terms of SSE of each cluster, was compared to each cluster in time period one (T1, 2008-2010). If the distance was larger than a threshold value, the cluster was considered a new cluster. If, however, the distance was smaller than the threshold value, the cluster in T2 was matched to the cluster in T1 with the minimum distance. Once matched, the centroid vector was adjusted to the mean value of the matched clusters, thereby calculating a moving average of the centroid. Thus, the clusters in T3 (2014-2016) would be compared to any clusters from the previous time period (T2) that were not matched, as well as the moving average of clusters that were matched.

We defined cluster names based on the average values from the cluster centers. For origin country clusters it was clear that the four clusters corresponded to one of the four dominant migration categories. Names for the destination clusters are more nuanced as at times multiple clusters were dominated by one migrant category but to different degrees.

3.2.1 EU migrant origin clusters

Clustering of the non-EU countries of origin identified only four clusters throughout all time periods, which implies that the cluster characteristics did not change enough to be greater than the threshold value. On the EU destination side, however, clustering for T3 (2014-16) brought up a cluster that was different enough from previous clusters so that it was deemed a new destination cluster.[1]

Figure 8: Dominant EU migrant-origin clusters at different time periods between 2008 and 2019

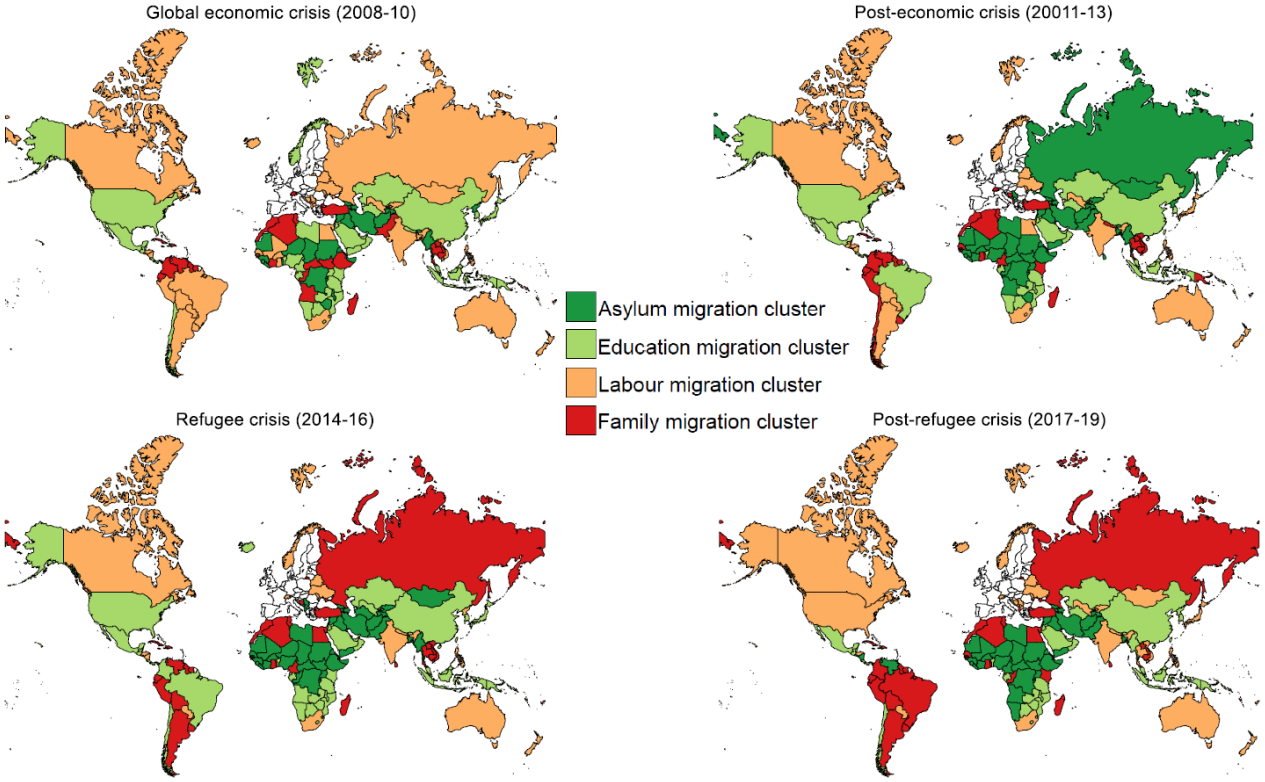


Figure 8 shows the four origin country clusters and their respective persistence and changes over time. While for the first period 2008-10 the labour migration cluster included large parts of Latin America and Eastern Europe, the asylum cluster was dominant in large parts of Central and Western Africa, the Middle East and Central Asia.

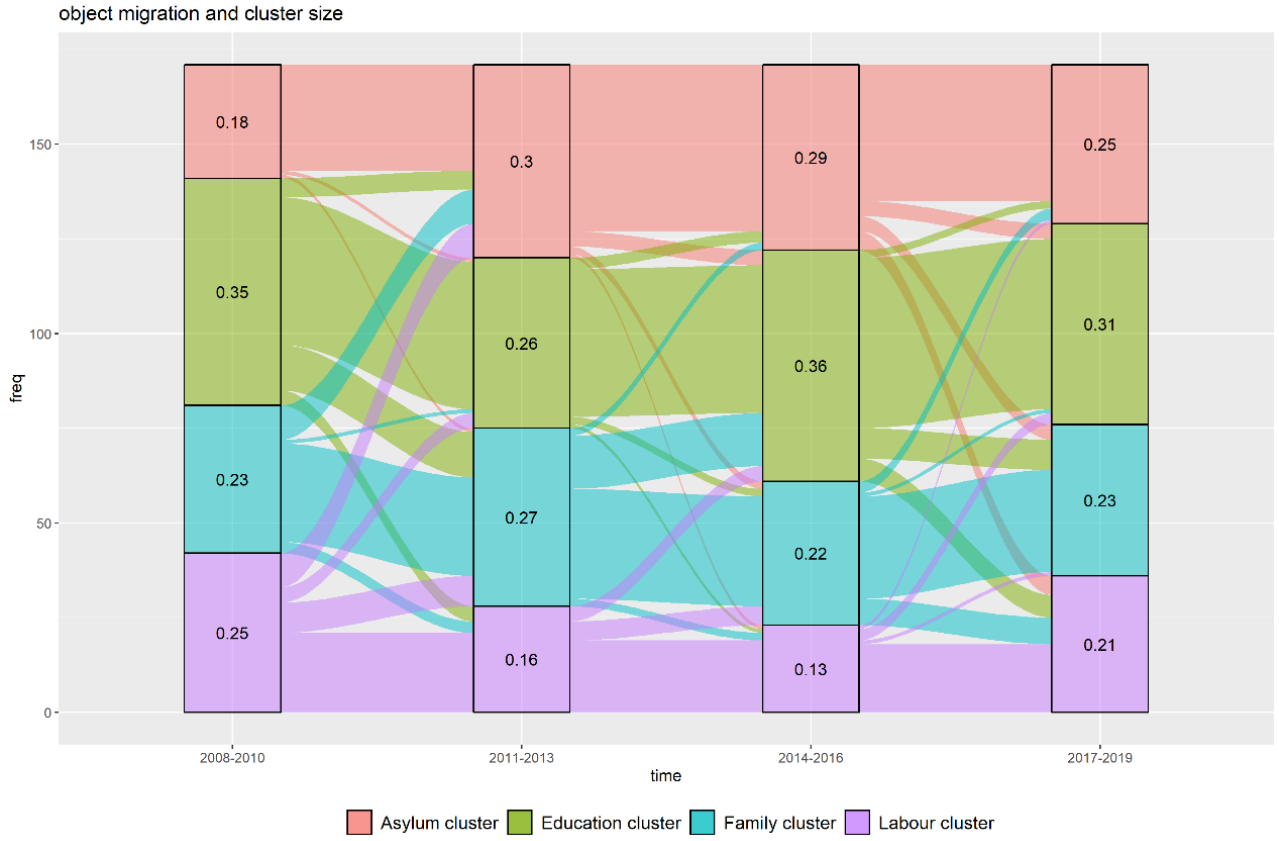
Interestingly, Latin America has transitioned significantly from labour migration (2008-10), to EU migration dominated by labour and education (2011-13), then to education-family migration (2014-16), and in the period 2017 to 2019, to largely family migration dominated region. On the other hand, countries on the African continent were rather diverse in terms the legal pathways into the EU in the height of the global and European economic crisis of 2008-2010 but turned then increasingly in a geographical cluster dominated in large parts by asylum migration - except for the Southern African countries, which have largely remained source countries dominated by education and labour (in the case of South Africa) migration.

Figure 9 displays the proportional sizes of these four clusters, that is, the percentage of countries of the total 171 non-EU countries in a cluster, and the 'movement of countries' from one cluster to

[1] This analysis was also done on the yearly data and results are available upon request. The names of the clusters were determined by look at the centre values for the four variables (asylum, family, labour, and education) produced by the k-means algorithm.

another cluster over time. While for about 60 per cent of countries labour or education was still the dominant mode of entry during the period of the economic crisis in Europe, this has changed significantly in the subsequent period of 2011-13 when only for 42 per cent of the 171 countries of origin these two migration categories have been dominant. This decline in the two cluster sizes corresponded primarily to the increase in the asylum and family migration clusters. Over the course of these four time periods, we observe some modest cluster movements with an average cluster changing rate of 23 per cent. Cluster persistence is therefore the most common feature. Persistence is relatively high in the education cluster representing around 25 per cent of all movements. In 23 per cent of all movements a country transitioned from the education cluster to a different cluster. Thirty countries were in the education cluster for all four time periods (2008-2019). When a country did change cluster membership from the education cluster it was most often to the family cluster (4 per cent of all possible movements).

Figure 9: EU migrant-origin cluster size, persistence and movements

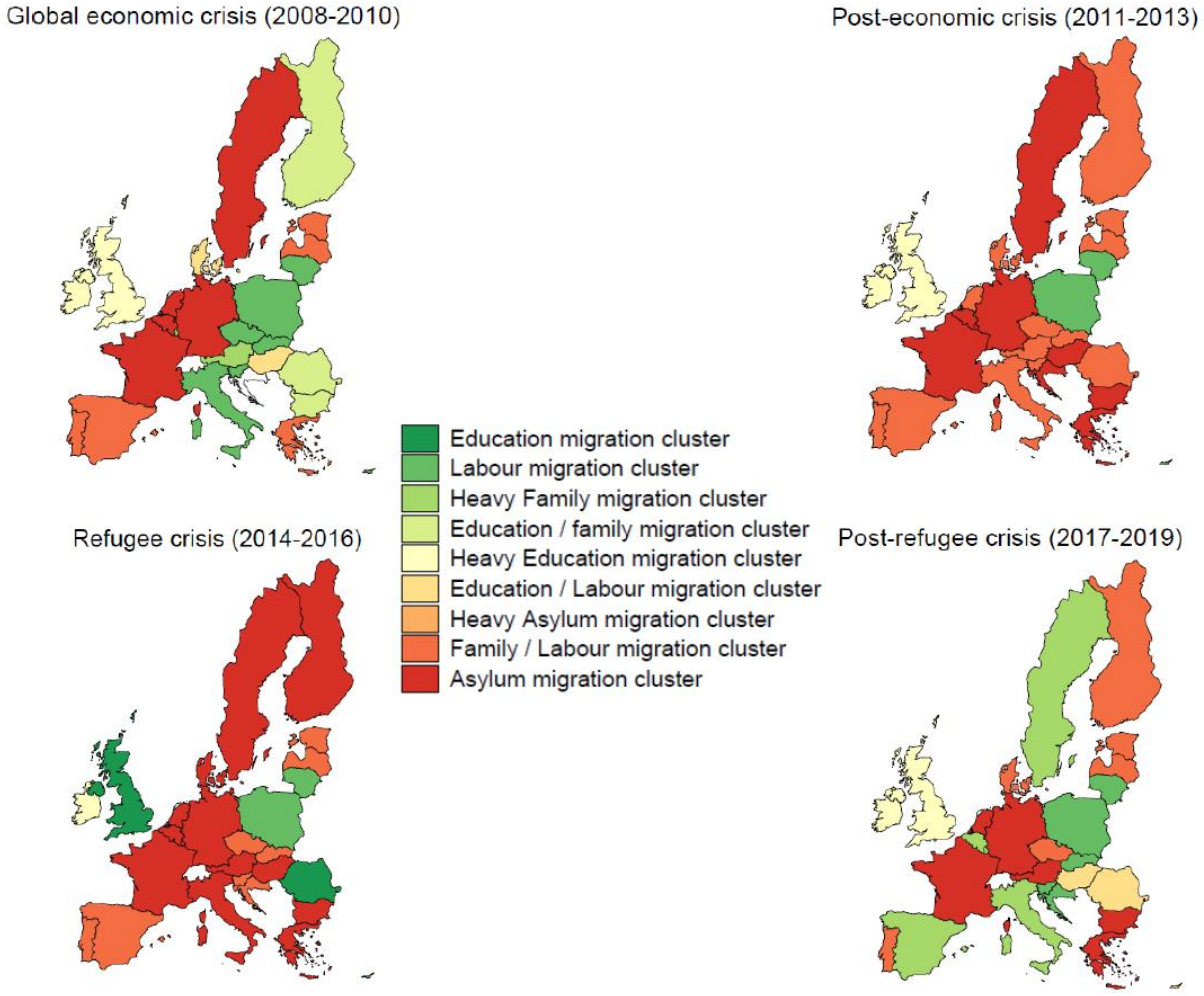


The asylum cluster was the most persistent cluster with 23 countries being in the asylum cluster throughout all four time periods. Only 17 per cent of the time a country left the asylum cluster. When a country did leave the asylum cluster it was most often to the education cluster. The labour migration cluster was the least persistent cluster. 38 per cent of the time countries left this cluster and most often moved to the family cluster. Overall, this analysis of the composition of the EU immigration flows from 171 non-EU countries of origin reveals significant persistence over time in terms of the dominant modes of entry.

3.2.2 EU migration destination clusters

The cluster analysis on the destination side identifies in total nine clusters with some distinct spatial-categorical patterns and changes in the composition of third country nationals across the 28 EU member states over the four periods between 2008 and 2019 (Figure 10).

Figure 10: Dominant EU migrant-destination clusters at time periods between 2008 and 2019

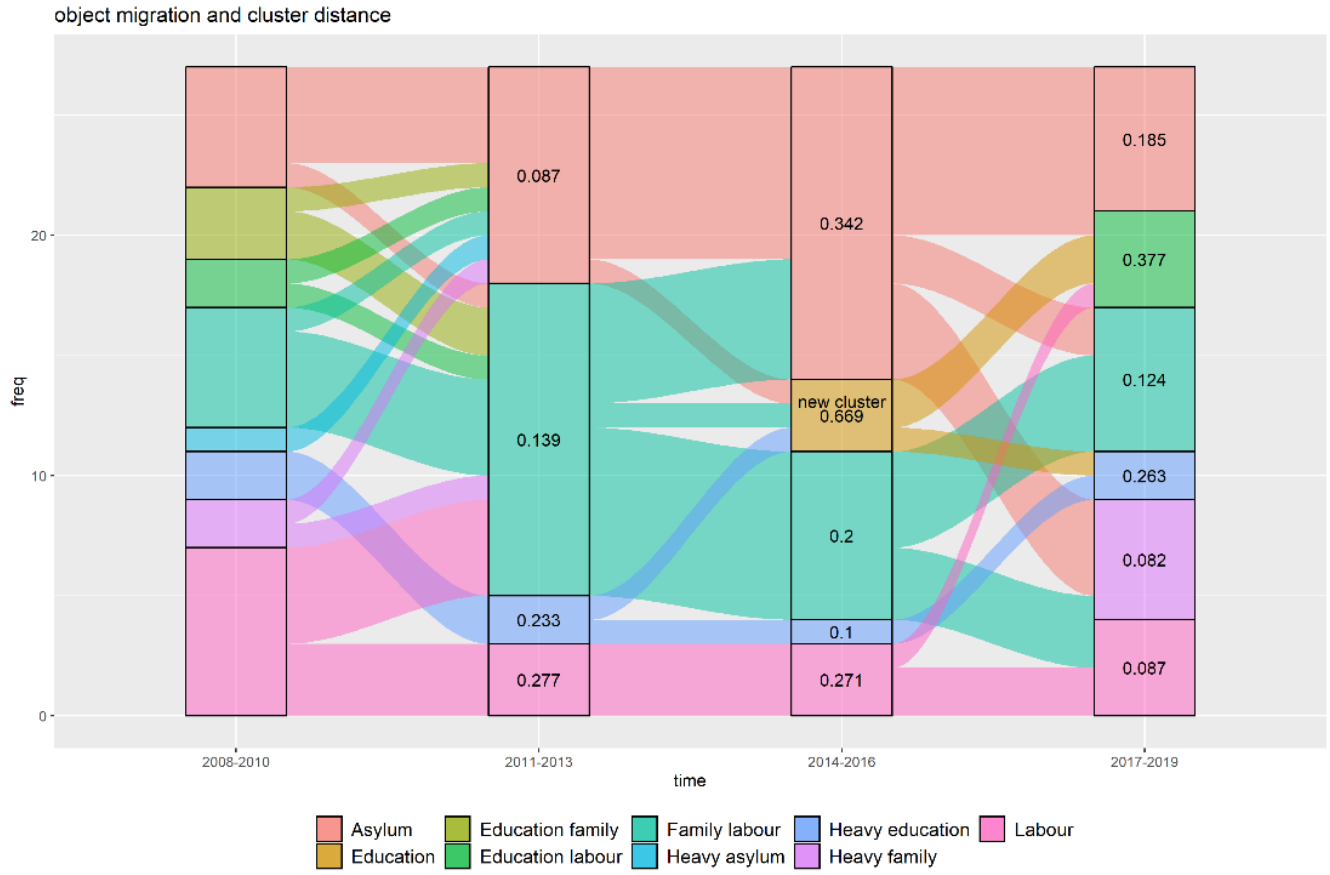


There are some countries that never change clusters such as Germany and France (persistent members of the asylum cluster), Ireland (persistent member of the education migration cluster), and the group of Estonia, Portugal, and Latvia (persistent members of the family-labour migration cluster). Other EU countries have seen compositional changes and transitions over time. In fact, 22 of the 28 EU destination countries changed clusters at least once over the course of these 12 years.[2] For instance, while inflows into Spain have been dominated by both family and labour migrants in the first three periods, in the latest period Spain became member of a cluster just dominated by family migration. Italy is a special case in this respect: It has changed cluster membership in each time period; first in the labour migration cluster, then moving to the family-occupation cluster, then in the height of the refugee crisis being part of the asylum migration cluster, and in the latest period, it moved to the family migration cluster.

[2] See Table A-3 in the appendix for the cluster movement of all destination countries.

Figure 11 displays the proportional sizes of the clusters, that is, the percentage of all 28 EU countries being member in a particular cluster, and the 'movement of countries' from one cluster to another over time. Notably in T1 the labour cluster was the largest at 26 per cent of the countries but it shrinks to 11 per cent in T2. From T1 to T2 the family occupation cluster grew the largest (19 per cent to 48 per cent) but declines back to 26 per cent in T3. A number of clusters that appear in T1 only appear in T1. The heavy family cluster disappears in T2 and T3 but re-emerges in T4. Unsurprisingly the asylum cluster is largest at T3 the period covering the height of the 'refugee crisis'.

Figure 11: EU migrant-destination cluster size, persistence and movements



There was a total of (28*3) 84 possible chances for cluster membership to change. Most often (53 per cent of the time) countries did not change cluster membership. Of all the possible movements, including staying in the same cluster, the most common feature was staying in the asylum cluster with around 21 per cent of all possible 'movements' (Figure 9 and 10). Thus, the most common state transition was a stationary. Some 64 per cent of the time EU destination countries in the asylum cluster remained in the asylum cluster. The most common state transition, excluding stationary, was from the family-labour migration cluster to the asylum cluster. This movement represented 7 per cent of all possible movements. However, 57 per cent of the time countries remained in the family occupation cluster.

4. Spatial and categorical dependence of bilateral migration flows into the Europe

In this section, we go one step further and analyse how the different migration flow categories (family, asylum, labour and education) are spatially and categorically interconnected. The previous explorative cluster analysis identified significant changes in the composition of origin and destination country clusters between 2008 and 2019, whereby some countries 'moved' for one or more periods to different clusters that were characterised by a different migration flow composition.

Changes in the categorical composition of bilateral migration flows may indicate that migrants switch between different legal pathways, for instance because of political or economic developments, or due to changes in admission policies in EU destination countries that may affect the relative attractiveness of a particular entry category. Whether and to what extent different legal pathways are indeed 'substitutes' and indeed interconnected both spatially and categorically is the focus of the following analysis.

4.1 Conceptualising spatial and categorical dependence

Despite an abundance of research on the drivers of migration, most studies focus solely on one entry category without considering the interdependence of migration flows, that is, the spatial and categorical substitutability of alternative modes and destinations of entry.[3] While previous research underlines the fact that multiple drivers may be interconnected (cf. Van Hear et al. 2018) and migration flows across different categories might interact (Erdal and Oeppen 2018, Crawley and Skleparis 2018), a rigorous investigation of the hypothesis that migration flows are categorically and spatially interdependent has so far not been put forward.

While migration research has pointed to evasive behaviour of migrants as the often unintended and unexpected consequences of migration policy interventions, categorical and spatial interdependencies have not been thoroughly analysed on a broader empirical basis. De Haas (2011, p. 27) differentiates between substitution effects that may occur when immigration restrictions are put into place: spatial substitution results as "the diversion of migration to countries with less restrictive regulations for similar categories of migrants," while categorical substitution reflects the reorientation of migrants towards legal or unauthorized entry channels in a situation where one entry channel has become more restrictive. The lack of legal immigration pathways may compel, for instance, low-skilled workers to apply for family reunion or asylum, rather than for work permit as actually intended (Massey 2004; Castles 2004; van Liempt and Doomernik 2006). For instance, after the suspension of guest worker programmes in the early 1970s, migration from the Maghreb or Turkey to Western Europe continued largely because de facto labour migrants had to switch to family visa as an alternative option, which was in most cases the only option for legal entry (de Haas 2014; Natter 2014). Consequently, some migration policies, including those targeting highly skilled workers (Czaika and Parsons 2017), might to some extent be ineffective because of categorical substitution effects (Czaika and de Haas 2013).

The lack of any viable options for legal entry has generally been argued to be a key reason for a

[3] For a synthesis of knowledge on the drivers of migration, see e.g. Czaika and Reinprecht (2020).

categorical deflection towards unauthorized channels (Massey 2004, Massey and Pren 2012, Castles 2004, de Haas 2008). In the European context, for instance, the introduction of visa requirements by Spain and Italy in 1991 shifted former temporary labour migrants from the Maghreb, or later also migrants from sub-Saharan countries, towards irregular modes of entry. Czaika and Hobolth (2016) have tested this 'deflection into irregularity' for the EU and demonstrate that more restrictive asylum and visa policies may indeed divert migrants into irregularity.

Similarly, spatial deflection of migration towards different destinations because of policy interventions has been conceptualised but rarely studied empirically on a broader basis. Spatial substitution occurs when restrictive admission or border policies divert migrants to countries with more liberal regulations or encourage migrants to follow alternative geographical itineraries (de Haas el al. 2019). For instance, increasing immigration restrictions by France, Belgium, and the Netherlands over the 1970s and 1980s contributed to a diversification of destinations for Moroccan emigrants, particularly to Spain and Italy and, mainly for the higher-skilled, to the United States and Canada (Berriane et al. 2015; de Haas 2014; Natter 2014).

Flahaux and Vezzoli (2018) have found that Caribbean countries whose borders with the former colonizing state were closed experienced a higher diversification of migration destinations than countries that retained a free mobility regime with the former metropole (de Haas el al. 2019). Vezzoli (2015) showed that migration restrictions introduced before independence diverted migration from Britain to North America. Czaika and Neumayer (2017) find that the introduction of a visa restriction by a destination country for citizens from a particular origin country significantly deters tourism flows but that some of the deterred flows in tourists are redirected to other (visa-free) destinations.

Also, the timing and sequencing of policy interventions can significantly affect the spatial substitution dynamic of migration. Even though the actual size of a spatial deflection effect seems relatively small (Czaika and Neumayer 2017), destination substitutability tends to be stronger when destination societies are similar in terms of culture, language, and opportunities, in which case migrants are more likely to opt for alternative destinations rather than giving up on their migration plans.

Barthel and Neumayer (2015) are the first who provide systematic evidence for spatial dependence of bilateral asylum flows for a larger sample of country dyads. They find that spatial dependence in asylum flows exists not only between geographically proximate countries of origin, but also between destinations. This indicates that that network effects are not only crucial within country dyads (between a country of origin and destination), but also between country dyads.

These network effects or spatial dependence among geographically proximate countries can be explained by factors including a common history, shared colonial past, common language, as well as political and economic ties (Barthel and Neumayer 2015). Brekke et al. (2017) have further pointed out that bilateral migration flows do not solely depend on a single dyad (pair of countries), but possibly also "on the opportunities in other destinations" (p. 3).[4] Nevertheless, Brekke et al. (2017) and Barthel and Neumayer (2015) have left out other types of migration flows, such as labour, family and education, in their analysis and did not examine the categorical dependence between different flows. The following analysis intends to fill this gap.

[4] In an earlier paper, Willekens and Baydar (1983) already have underlined the importance of "a distribution component" in forecasting migration whereby all possible destinations are taken into account and spatial interactions, such as distance, come into play.

4.2 Assessing spatial and categorical dependence

Even though Czaika and de Haas (2017) and de Haas et al. (2019) have already underlined the need for assessing the simultaneous effect of various migration flows and pointed out the unintended consequences of migration policies whereby deterrence and deflection of migration flows can arise, these have not yet been systematically studied on a broad empirical scale and simultaneously for multiple types of migration flows.

While spatial and categorical deflection effects might be driven by (changes in) policies, we may assume that these can also be the result of various other structural (political or economic) developments, such as the EU enlargement, the financial crisis or the so-called refugee crisis, that unevenly affect certain types of migrants and the relative attractiveness of respective legal pathways for third country nationals. Using theoretical arguments from the debate about policy effectiveness, the role of migrant networks and transnational ties in migration processes, as well as our findings from the cluster analysis above, we argue and hypothesize that bilateral migration flows are spatially and categorically interdependent.

4.2.1 Spatial dependence

Migrant networks are commonly agreed to "play a crucial role for facilitating continued migration" (Böcker 1995, Hatton 2004, de Haas 2011). These networks not only include the same nationality but may also include ethnic ties or religious groups and communities that go beyond one specific origin country (Massey et al. 1993, Beine et al. 2010, Bauböck and Faist 2010). For example, not just "Ghanaian community but Western African community" could facilitate migration from proximate origin countries (Barthel and Neumayer, 2015, p. 7).

Moreover, it has been shown that migration flow dependence is particularly likely across origin countries that are economically similar (Beverelli and Orefice 2019). We assume that this is especially true for those countries that are spatially close. Thus, we hypothesize that the flow of non-EU migrants from a specific country of origin to an EU country is influenced by other geographically proximate migrant flows of the same origin country. Or more specifically:

> (*H1*) An increase in the flow of one type of migration (labour, family, education, asylum) to a specific European destination country *j* from other countries of origin *spatially proximate* to country of origin *i* is associated with an increase in the flow from country of origin *i* to the same EU destination *j*.
>
> (*H2*) An increase in the flow of one type of migration (labour, family, education, asylum) to a specific European destination country *j* from other countries of origin of the *same cluster type* as country of origin *i* is associated with an increase in the flow from country of origin *i* to the same EU destination *j*.

We conjecture that both hypotheses may basically hold for all four types of migration flows: asylum, education, labour and family.

Migration research has highlighted that migration policies may be ineffective or cause unintended side-effects (de Haas 2011, Czaika and de Haas 2017). This phenomenon is explained by the power of migrants' individual or collective agency by "defying immigration rules and adopting new migration strategies and pathways" (de Haas, 2011, p. 25). Spatial substitution effects of migration flows may arise once restrictive policies are implemented in a destination country so that migration flows may be deflected to neighbouring countries that are less restrictive (Collyer 2005, Holzer 2000, Barthel and Neumayer 2015, p. 11-12).

Besides policy changes, also economic conditions can play a role in the deflection of migrant flows

(Bertoli et al. 2016, Beverelli and Orefice 2019). Substitutability of destinations within certain limits is based on the idea that, if necessary, migrants may choose alternative destinations within a set of possible destinations ranked in subjective hierarchy of most favoured destinations. The substitutability is expected to be strongest where countries of destination are culturally, linguistically and economically relatively similar (Barthel and Neumayer 2015). We assume that this is mainly the case where destination countries are geographically proximate or part of the same type of migration cluster. Consequently, we hypothesize that the flow of non-EU migrants to a specific a destination country influences geographically proximate migrant flows to geographically proximate destination countries. More specifically, this translates into the following hypotheses:

> (*H3*) An increase in the flow of migrants of a certain form (labour, family, education, asylum) from country of origin *i* to other *spatially proximate* EU destination countries is associated with a decrease the flow of migrants of the same type from country of origin *i* to EU destination *j*.
>
> (*H4*) An increase in the flow of migrants of a certain form (labour, family, education, asylum) from country of origin *i* to other EU destination countries of the *same cluster type* is associated with a decrease the flow of migrants of the same type from country of origin *i* to EU destination *j*.

We suspect that this holds for all migration flow categories.

4.2.2 Categorical dependence

Since categorical dependence has rarely been subject to research so far, its theoretical basis is almost non-existent. While categorical deflection of regular migrants into unauthorised migrants has been analysed (Czaika and Hobolth 2016, Massey and Pren 2012), to our knowledge similar analyses are not available for testing categorical substitution effects between legal migration channels. But anecdotal evidence as well as our cluster analysis above provides an indication that migration flows within different legal pathways might be categorically interconnected, that is, flows of migrants within one legal pathway are associated with flows in another legal pathway. For instance, the cluster analysis has shown that family migration is a significant entry category in almost all origin and destination country clusters, and for many countries, it is even the dominant legal pathway (Figure 8 and 10). This is not surprising knowing that family migration is legally bound to all other entry categories. We therefore hypothesize that categorical dependence between family migration and the other categories is relevant:

> *(H5)* The higher the inflow of migrants under labour, education or asylum, the higher the subsequent number of family migrants.

However, we expect that for labour, categorical dependence with family migration to be the highest since regulations on family reunification after entry of a (sponsoring) labour migrant is overall more liberal (or, less restrictive) than in the case of the other legal categories education or asylum (Czaika and de Haas 2013). Moreover, migration research suggests that restrictions on labour entry can defer potential migrants to opt for family, asylum, or educational pathways instead (Massey 2004; Castles 2004; van Liempt and Doomernik 2006; Czaika and de Haas 2013). Even high-skilled workers often use the family visa as a mode of entry as work visas are often quota-based and only available in limited numbers (e.g. H1B visa in the US). Consequently, we hypothesize that:

> *(H6)* Family migration is positively interlinked with all other entry categories, but strongest with labour migration.

Furthermore, while labour and study pathways are often legally interlinked, the cluster analysis has shown (e.g. for the case of many Latin American) countries that education and labour migration

clusters have been attached over time (see Figure 8 and 9). Therefore, we hypothesize:

> *(H7)* The flow of migrants under entry category labour is mainly interlinked with educational and family migration flows. The higher the inflow of educational and family flows, the higher the subsequent inflow of labour migrants, and vice versa.

The cluster analysis has further shown that for many origin countries, but also some EU destination countries, migration flow compositions have turned from asylum-dominated to family–dominated (see Figure 10 and 11). For instance, this has been the case for Albania as a migrant-origin country or Austria as an EU destination. We assume that this could be an indication for a categorical dependence between asylum migration and subsequent family reunion. Yet, focusing on the effect of asylum on other categories, we only, if at all, expect small effects, also because the asylum cluster showed high persistency over time (Figure 11). Only in the case of well-educated persons who are seeking protection from conflict and persecution, the educational pathway, if available at all, may be an alternative legal pathway instead of applying for asylum. Thus, to that end, we hypothesize that:

> *(H8)* Asylum migration is positively interlinked with family migration but negatively with educational migration.

4.3 Data and Methodology

To test the stated hypotheses, we use the first permit data and asylum applications from Eurostat (2020)[5] that we also used for the cluster analysis in 3.2. The observation period is still 2008 to 2019 and the unit of analysis is dyad-periods of 28 EU destination countries and 171 countries of origin with some data gaps creating an unbalanced dataset. The dependent variables are the total numbers of the different migration flow categories, that is, labour, family, education and asylum. All flow data is log-transformed[6] to reduce the skewness in the distributions.

To test spatial dependence between different migration flows, we first created spatial lag variables for each entry category. As we assume that migration flows are geographically linked and clustered, we use the inverted distance as a weight measure. The population-weighted bilateral distance (defined as geographical distance between most populated cities) comes from the CEPII GeoDist dataset (Mayer and Zignago 2011). Because we anticipate spatial dependence among both source and target countries, we calculate both specific source and target contagion variables. We follow here the method from Neumayer and Plümper (2010), Plümper and Neumayer (2010) and Barthel and Neumayer (2015) who have used inverted distance as one of their spatial weight variables. Formally, the estimated model is:

$$M_{ijt}^{a} = \beta_0 + \beta_1^{s\neq a} M_{ijt-1}^{s\neq a} + \omega_{ikt-1} \sum_{k\neq i} M_{kjt-1}^{a} + \omega_{jmt-1} \sum_{m\neq j} M_{imt-1}^{a} + \gamma_{ij}^{a} + \mu_{it}^{a} + \rho_{jt}^{a} + \varepsilon_{ijt}^{a}, \qquad (1)$$

where M_{ijt}^{a} is the natural log of migration flows of type $a \in \{\text{labour}, \text{family}, \text{education}, \text{asylum}\}$ from country of origin i to EU destination country j in time period t; $M_{ijt-1}^{s\neq a}$ is the natural log of the migration flow of type $s \neq a$ from country of origin i to EU destination country j in time period $t-1$; $\omega_{ikt-1} \sum_{k\neq i} M_{kjt-1}^{a}$ models the lagged spatial dependence among countries of origin, and $\omega_{jmt-1} \sum_{m\neq j} M_{imt-1}^{a}$ the lagged spatial dependence between EU destinations regarding migration

[5] See for more info on the Eurostat first permit data: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Residence_permits_-_statistics_on_first_permits_issued_during_the_year&oldid=507019.

[6] Zeros have been added +1 to before log-transformation.

flow of type a; and ε_{ijt}^{a} is the idiosyncratic error term in the model of migration flow of type a. To control for unobserved spatial heterogeneity and spatial clustering, we include origin-time μ_{it}^{a} and destination-time fixed effects ρ_{jt}^{a}, as well as ixj dyad-specific fixed effects γ_{ij}^{a} to capture any form of unobserved heterogeneity. This provides the added advantage of not having to include specific drivers of migration which are often hard to measure and incomplete for a large dataset like ours.

Spatial autoregressive models introduce some endogeneity due to the presence of spatial lags in the estimation models. Tackling this by a spatial maximum likelihood or spatial instrumental variable model (Lee and Yu 2010) would be extremely difficult given the very large number of fixed effects. However, as also noted by Barthel and Neumayer (2015), if the true degree of spatial dependence is relatively small, then Monte Carlo analysis suggests that the bias of an ordinary least squares (OLS) model is small (Franzese and Hays 2007).

Model (1) represents four migration equations where we may assume that error terms ε_{ijt}^{a} across the equations are correlated. In this case may face possible bias in our estimates and inefficiency due to dependence in the error terms which is why we fit a four-equation seemingly unrelated regression (SUR) model including in each equation the set of spatial and categorical dependence variables and the set of fixed effects. For robustness reasons we have also run in addition the four OLS regressions separately including origin-time, destination, and dyad-specific fixed effects (Table A-1 in the appendix).

4.4 Results

Estimation results of model (1), as displayed in Table 3, show for all four types of migration flows, labour, family, education and asylum, supportive evidence for the existence of both spatial and categorical dependence. The estimates of the spatial lag variables indicate that migration flows of a certain category are simultaneously influenced by bilateral migration flows towards EU destinations originating in geographically proximate countries of origin. Hypothesis 1 is thus confirmed. This effect of geographical clustering of countries of origin on bilateral migration flows towards Europe is strongest in the area of asylum migration, for which we estimate that a 10 per cent increase in asylum flows from proximate countries increasing asylum flows from a country of origin by about 3.7 per cent. This estimate is largely in line with the respective estimate of Barthel and Neumayer (2015) who estimate an elasticity of 0.443 (i.e., a 10 per cent increase in asylum flows in proximate countries of origin increasing flows *i* to *j* by about 4.4 per cent). Spatial source dependence estimates for other forms of migration are all highly statistically significant at 1 per cent levels, but smaller in effect size compared to spatial source dependence in asylum flows.

For spatial target dependence in flows of certain forms of migration between EU destinations, we find negative and statistically significant effects of the respective spatial lag terms. This is in line with our hypothesis 3. The effect size, however, is smaller for target than for source dependence, which implies that proximity between countries of destination affects the direction of flows by less than distance between countries of origin. While spatial dependence between EU destinations seems strongest for educational migration, the effect size of spatial dependence between countries of origin is largest in the case of asylum. But possibly more important than the actual effect size is the direction of the effect. While bilateral flows from countries of origin towards a specific EU destination are complements, flows from a specific country of origin 'splitting' between EU destinations are rather negatively associated.

Table 1: Spatial and categorical dependence of European immigrant flow categories, 2008-2019, SUR regression

VARIABLES	(1) **Labour**	(2) **Family**	(3) **Education**	(4) **Asylum**
Spatial dependence				
Labour migration (other origins)	0.166***			
	(0.0166)			
Labour migration (other destinations)	-0.0604**			
	(0.0276)			
Family migration (other origins)		0.238***		
		(0.0165)		
Family migration (other destinations)		-0.124***		
		(0.0261)		
Education migration (other origins)			0.321***	
			(0.0180)	
Education migration (other destinations)			-0.261***	
			(0.0311)	
Asylum migration (other origins)				0.370***
				(0.0213)
Asylum migration (other destinations)				-0.196***
				(0.0471)
Categorical dependence				
Labour migration		0.433***	0.368***	0.0270**
		(0.00718)	(0.00888)	(0.0124)
Family migration	0.427***		0.346***	0.546***
	(0.00743)		(0.00910)	(0.0119)
Education migration	0.325***	0.264***		0.0178
	(0.00690)	(0.00682)		(0.0109)
Asylum migration	0.0150***	0.213***	-0.0196***	
	(0.00549)	(0.00510)	(0.00621)	
Constant	-0.676***	-0.649***	-1.348***	-1.656***
	(0.169)	(0.173)	(0.188)	(0.260)
Observations	15,176	15,176	15,176	15,176
R-squared	0.884	0.908	0.858	0.760

Standard errors in parentheses: *** p<0.01, ** p<0.05. All dependent and independent variables are logged to interpret estimates as elasticities. Independent variables are lagged by time (3yrs) period. Specific target and source spatial lag variables with inverted distance as weighting matrix. All models include origin-time and destination-time fixed effects.

The strongest interdependence exists for education and asylum migration flows where a 10 per cent increase in proximate European destination countries reduces flows to country *j* by about 2.6 per cent in the case of education flows, and of about 2 per cent in asylum flows. The 'weakest' spatial interdependence between flows across EU destinations we identify in the area of labour migration.

These results point to some non-negligible substitutability within certain migrant categories across destination countries but also some complementarity between countries of origin. That is, the larger a migration flow in one legal category the larger the flows of migrants of the same type from geographically close origin countries moving towards the same EU destination country. But within the EU, a larger inflow of migrants of one type results in lower inflows of the same type into other proximate EU destination countries.

Results do not only indicate to spatial dependence between alternative forms of migration flows of third country nationals into Europe, but also substitutability between legal pathways by which migrants enter an EU member state. We identify the strongest categorical dependence effects between education and labour migration flows as well as between labour and family migration flows as hypothesized in H6 and H7. For instance, an increase in the number of labour migrants by 10 per cent in one period increases inflows of family migrants by 4.3 per cent in the following three-year period. Family migration, as also expected in H6, is not only dependent on prior labour migration flows, but also responds to educational inflows (elasticity of 0.264) and to the inflow of asylum migrants (elasticity of 0.213). Thus, family migration is categorically dependent on all other entry categories.

Educational migration is positively associated with labour and family migration, but negatively with asylum migration, which could be an indication for some limited substitutability between these two legal pathways. However, this substitution effect seem to exist only in one direction - from asylum towards the educational pathway -, the reverse is not significant. At the same time, asylum migration is positively associated with labour and family migration indicating to the possibility that prior labour or family migrants may facilitate subsequent asylum migration towards the same destination.

These results have shown to be robust to alternative model specifications and estimation techniques. For instance, rather than running simultaneously estimating a multi-equation SUR model which assumes dependency of the respective error terms, we have also run separate regressions on the four types of migration flows including the maximum amount of fixed effects (Table A-1). We also have run model (1) by using alternative spatial lag variables that have been calculated using the information from the cluster analysis of countries using countries' shared cluster membership as the weighting matrix for creating the respective spatial lag variables. Results of this alternative specification of the spatial lag variables are reported in Table A-2 in the annex and largely confirm the estimates discussed above using (inverse) distance as weighting matrix.

5. Summary and conclusion

This study has shown that migration flows within different legal pathways into Europe are spatially and categorically clustered and interconnected. Using longitudinal data from Eurostat on first residence permits for reasons of asylum, family reunification, labour or education, we have, in a first step, identified distinct migration clusters of countries sharing similar compositions of bilateral migration flows. Our empirical assessment reveals that these migration clusters are far from static but can change significantly over time, particularly after the economic crisis, and during, as well as after the refugee crisis. While some countries of origin of EU immigrants follow persistent cluster trajectories, such as those for asylum, other countries are regularly changing cluster membership in regard to the legal pathway predominantly used by their EU migrants. Dominant legal pathways may shift, for example, between labour and education or labour to family migration, reflecting rather flexible and fluid migration trajectories. Similarly on the migrant-receiving side of EU member states in terms of the composition of third country nationals by their legal status: while immigration in some countries like Germany, France, or Poland has been dominated by one distinct type of immigrants, other countries like Italy, Denmark or Austria have seen some nuanced shifts in the composition of their non-EU immigrant population since 2008.

In the second step, we explored the extent to which European immigration flows of third country nationals are both categorically and spatially dependent. Our findings suggest that bilateral flows of labour, family, education and asylum migrants are not only categorically clustered, but also multilaterally dependent on proximate flows on the side of the non-EU countries of origin as well

as on the side of EU destinations. Therefore, due to their spatial dependence, geographically proximate and clustered migration flows are to be taken into account when wanting to manage or to predict migration flows into Europe. For instance, any unilateral migration policy intervention affects not only the targeted migration flow but also flows from and towards other countries. Consequently, migration policy interventions in other countries are of relevance in any attempt to influence migration flows in a certain direction.

Moreover, we find robust evidence that bilateral migration flows are not only spatially dependent across country dyads, but also within country dyads different legal pathways categorically dependent. The findings show that especially family migration is a cross-cutting legal category that is structurally connected to flows of the other legal pathways. Yet also education and labour migration flows are categorically interconnected, as well as asylum and family.

Consequently, migration pathways intersect and are neither spatially nor categorically independent. An increase in one migration flow category is associated with an increase in other categories. For instance, an increase in asylum or labour migration is very likely to be followed by an increase in family migration, even though to a different extent.

These results highlight the importance and acknowledgement of spatial and categorical interdependencies in our general understanding of migration processes. Our own review of literature has demonstrated a lack of theoretical elaborations of spatial and categorical dependence. If at all, migration systems theory is providing some explanations for the interconnectedness of international migration flows (in addition to other forms of cross-border flows of goods, services, and commodities). But how exactly various forms of migration are functionally interlinked is largely undertheorized and requires more consideration.

Our results are also of relevance for policy makers who aim to influence migration flows in a certain direction - often by means of unilateral action. This study suggests that policymakers must be aware of the side effects their policy interventions may have on other than directly targeted migration flows. Any form of 'pathway-specific migration management', that is, national approaches for controlling certain types of migration independently from other legal migration pathways within the same country or to and from other proximate countries, are largely bound to fail. Sometimes even worse, they may suffer from various unexpected and unintended externalities. This underlines the need for more coordination within national regulatory systems, but calls as well for a more comprehensive migration flow management across national approaches. Consequently, there is a need for greater coordination within the EU but also beyond the EU both with non-EU destinations and countries of origin in order to embrace spatial and categorical dependence of migration flows as an empirical fact that cannot be ignored if migration policy objectives are to be achieved.

6. Annex

Table A-1: Spatial and categorical dependence of European immigrant flow categories, 2008-2019, FE panel regression

DV	(1) **Labour**	(2) **Family**	(3) **Education**	(4) **Asylum**
Spatial dependence				
Labour migration (other origins)	0.198***			
	(0.026)			
Labour migration (other destinations)	-0.751***			
	(0.027)			
Family migration (other origins)		0.111***		
		(0.028)		
Family migration (other destinations)		-0.471***		
		(0.037)		
Education migration (other origins)			0.163***	
			(0.026)	
Education migration (other destinations)			-0.640***	
			(0.033)	
Asylum migration (other origins)				0.348***
				(0.023)
Asylum migration (other destinations)				-0.535***
				(0.029)
Categorical dependence				
Labour migration		0.095***	0.065***	-0.005
		(0.005)	(0.005)	(0.007)
Family migration	0.055***		0.032***	0.028***
	(0.006)		(0.006)	(0.008)
Education migration	0.098***	0.066***		0.032***
	(0.006)	(0.005)		(0.007)
Asylum migration	0.001	0.030***	0.019***	
	(0.004)	(0.004)	(0.004)	
Constant	3.102***	3.167***	3.280***	1.609***
	(0.157)	(0.195)	(0.164)	(0.117)
Observations	40,809	40,962	40,879	42,559
R-squared	0.445	0.447	0.417	0.287
Number of dyads	4,676	4,676	4,676	4,676

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1. All dependent and independent variables are logged to interpret estimates as elasticities. Independent variables are lagged by one year. Specific target and source spatial lag variables with inverted distance as weighting matrix. All models include origin-time, destination-time and dyad-specific fixed effects.

Table A-2: EU destination country cluster trajectories

Country code (iso3)	Period t1: 2008-2010	Period t2: 2011-2013	Period t3: 2014-2016	Period t4: 2017-2019
AUT	Heavy family cluster	family occupation cluster	asylum cluster	asylum cluster
BEL	asylum cluster	asylum cluster	asylum cluster	Heavy family cluster
BGR	education family cluster	asylum cluster	asylum cluster	asylum cluster
CYP	occupation cluster	occupation cluster	occupation cluster	education occupation cluster
CZE	occupation cluster	family occupation cluster	family occupation cluster	family occupation cluster
DEU	asylum cluster	asylum cluster	asylum cluster	asylum cluster
DNK	education occupation cluster	family occupation cluster	asylum cluster	family occupation cluster
ESP	family occupation cluster	family occupation cluster	family occupation cluster	Heavy family cluster
EST	family occupation cluster	family occupation cluster	family occupation cluster	family occupation cluster
FIN	education family cluster	family occupation cluster	asylum cluster	family occupation cluster
FRA	asylum cluster	asylum cluster	asylum cluster	asylum cluster
GBR	Heavy education cluster	Heavy education cluster	education cluster	Heavy education cluster
GRC	family occupation cluster	asylum cluster	asylum cluster	asylum cluster
HRV	NA	asylum cluster	family occupation cluster	occupation cluster
HUN	education occupation cluster	asylum cluster	asylum cluster	education occupation cluster
IRL	Heavy education cluster	Heavy education cluster	Heavy education cluster	Heavy education cluster
ITA	occupation cluster	family occupation cluster	asylum cluster	Heavy family cluster
LTU	occupation cluster	occupation cluster	occupation cluster	occupation cluster
LUX	Heavy family cluster	asylum cluster	asylum cluster	Heavy family cluster
LVA	family occupation cluster	family occupation cluster	family occupation cluster	family occupation cluster
MLT	Heavy asylum cluster	asylum cluster	education cluster	education occupation cluster
NLD	asylum cluster	family occupation cluster	asylum cluster	asylum cluster
POL	occupation cluster	occupation cluster	occupation cluster	occupation cluster
PRT	family occupation cluster	family occupation	family occupation	family occupation cluster

		cluster	cluster	
ROU	education family cluster	family occupation cluster	education cluster	education occupation cluster
SVK	occupation cluster	family occupation cluster	family occupation cluster	occupation cluster
SVN	occupation cluster	family occupation cluster	family occupation cluster	occupation cluster
SWE	asylum cluster	asylum cluster	asylum cluster	Heavy family cluster

Figure A-1 Destination country cluster distance

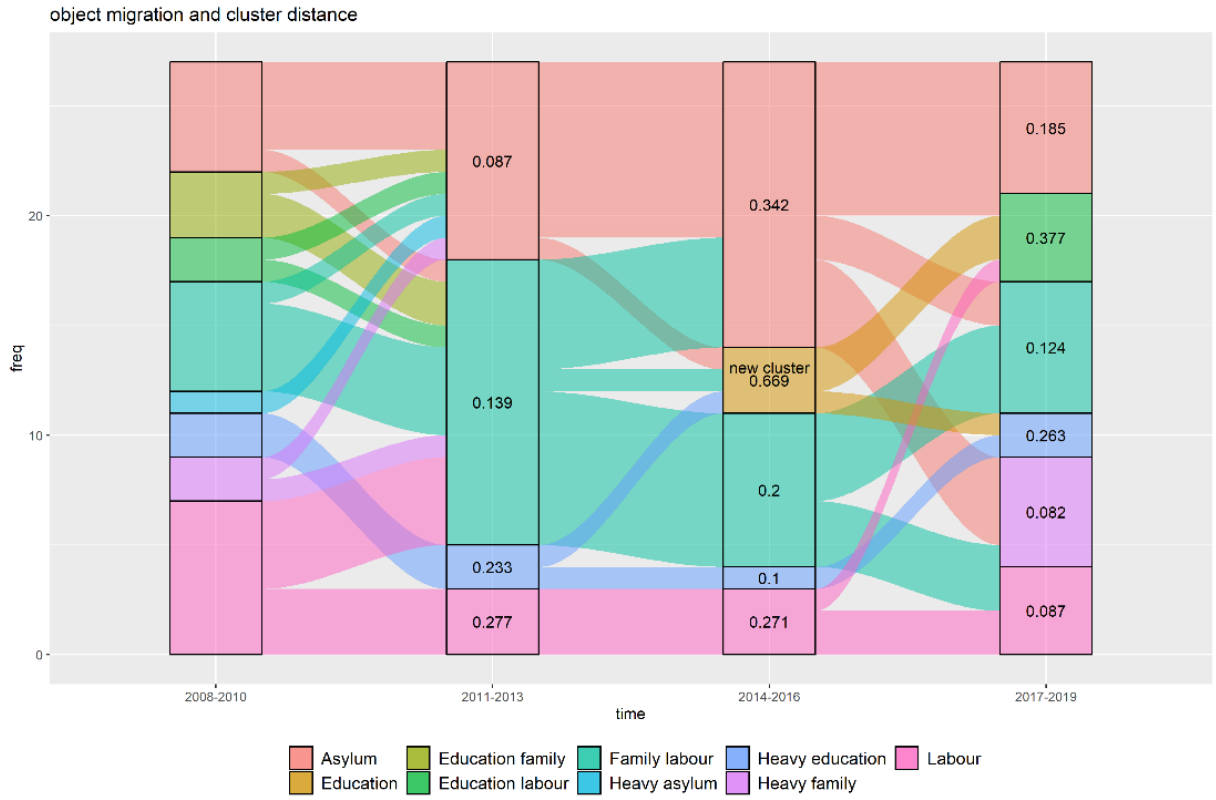
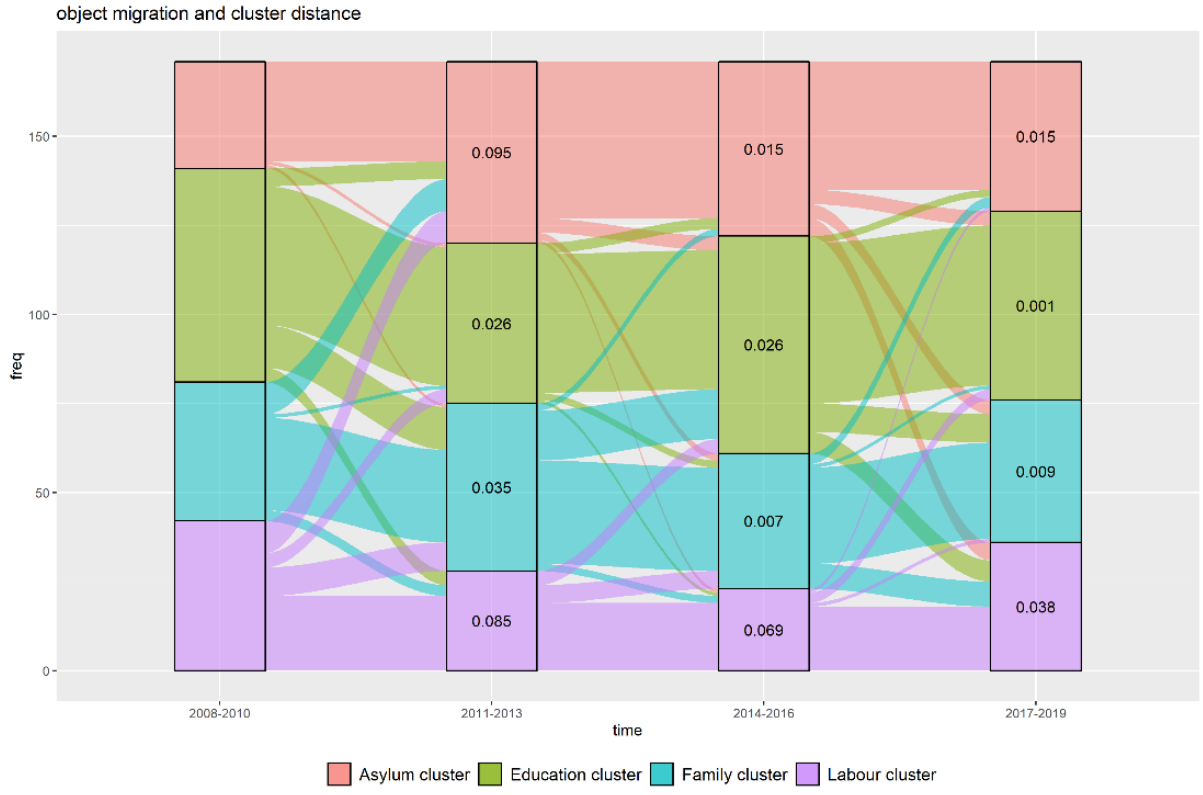


Figure A-2 Origin country cluster distance



7. References

Bakewell, O. (2008) Research beyond the categories: The importance of policy irrelevant research into forced migration. *Journal of Refugee Studies*, *21*(4), 432-453.

Bijak, J. and Czaika, M. (2020) Assessing Uncertain Migration Futures –A Typology of the Unknown. QuantMig Project Deliverable D1.1. Southampton/ Krems: University of Southampton and Danube University Krems.

Barthel, F. and Neumayer, E. (2015) Spatial Dependence in Asylum Migration, *Journal of Ethnic and Migration Studies*, 41(7), 1131-1151.

Bauböck, R. and Faist, T. (eds.) (2010). Diaspora and Transnationalism: Concepts, Theories and Methods, IMISCOE: Amsterdam University Press.

Beine, M., Docquier, F., and Özden, C. (2011) Diaspora Effects in International Migration: Key Questions and Methodological Issues, *World Bank Policy Research Working Paper* 5721, Washington, DC: World Bank.

Berriane, M., De Haas, and Natter, K (2015) Introduction: revisiting Moroccan migrations, *The Journal of North African Studies*, 20(4), 503-521

Beverelli, C., and Orefice, G. (2019) Migration deflection: The role of Preferential Trade Agreements. *Regional science and urban economics*, *79*, 103469.

Brekke, J.-P., Røedy, M., and Schøn, P. (2017) Reduction or deflection? The effect of asylum policy on interconnected asylum flows, *Migration Studies*, 1-32.

Böcker, A.G.M. (1995) Migration Networks: Turkish Migration to Western Europe. In: R. van Erf & L. Heering (Eds.), Causes of International Migration. Proceedings of a Workshop, 151-171.

Castles, S. (2004) Why migration policies fail. *Ethnic and Racial Studies* 27(2), 05-27.

Collyer, M. (2005) When do social networks fail to explain migration? Accounting for the movement of Algerian asylum-seekers to the UK, *Journal of Ethnic and Migration Studies*, 31(4), 699-718.

Crawley, H., and Skleparis, D. (2018) Refugees, migrants, neither, both: categorical fetishism and the politics of bounding in Europe's 'migration crisis'. *Journal of Ethnic and Migration Studies*, *44*(1), 48-64.

Czaika, M. and De Haas, H. (2013) The Effectiveness of Immigration Policies, *Population and Development Review*, 39: 487-508.

Czaika, M., and De Haas, H. (2014) The globalization of migration: Has the world become more migratory?, *International Migration Review*, 48(2), 283-323.

Czaika, M. and Hobolth, M. (2016) Do restrictive asylum and visa policies increase irregular migration into Europe?, *European Union Politics*, 17(3), 345-365.

Czaika, M. and de Haas, H. (2017) The effect of visas on migration processes, *International Migration Review*, 51(4), 893-926.

Czaika, M. and E. Neumayer (2017) Visa restrictions and Economic Globalisation, *Applied Geography*, 84, 75 – 8.

Czaika, M., and Parsons, C. R. (2017). The gravity of high-skilled migration policies. *Demography*, *54*(2), 603-630.

De Haas, H. (2011) The determinants of international migration: Conceptualising policy, origin and destination effects, *DEMIG project paper 2*, IMI/Oxford Working Paper, 32.

De Haas, H., Czaika, M., Flahaux, M. L., Mahendra, E., Natter, K., Vezzoli, S., and Villares-Varela, M. (2019) International migration: Trends, determinants, and policy effects. *Population and Development Review*, *45*(4), 885-922.

Erdal, M. B., and Oeppen, C. (2018) Forced to leave? The discursive and analytical significance of describing migration as forced and voluntary. *Journal of Ethnic and Migration Studies*, *44*(6), 981-998.

Eurostat (2020) Residence permits - statistics on first permits issued during the year, Eurostat, Brussels, https://ec.europa.eu/eurostat/statistics-explained/, accessed January 2021

Flahaux, M. L., and Vezzoli, S. (2018) Examining the role of border closure and post-colonial ties in Caribbean migration. *Migration Studies*, *6*(2), 165-186.

Franzese, R. and Hays, J. (2007) Spatial-econometric models of cross-sectional interdependence inpolitical-science panel and time-series-cross-section data, *Political Analysis* 15(2): 140–164.

Fürtges, L., Conradt, D., Schätzle, M. A., Singh, S. K., Kraševec, N., Rižner, T. L., …, Husain, S. M. (2017) Phylogenetic Studies, Gene Cluster Analysis, and Enzymatic Reaction Support Anthrahydroquinone Reduction as the Physiological Function of Fungal 17β-Hydroxysteroid Dehydrogenase. *ChemBioChem*, 18(1), 77-80.

Hatton, T. (2004) Seeking Asylum in Europe, *Economic Policy*, 19(38), 5-62.

Holzer, T., Schneider, G. and Widmer, T. (2000) The Impact of Legislative Deterrence Measures on the Number of Asylum Applications in Switzerland (1986-1995), *International Migration Review*, 34(4), 1182-1216.

IOM (2020) World migration report 2020, Geneva: International Organization for Migration.

Lee, L. F., and Yu, J. (2010). Some recent developments in spatial panel data models. *Regional Science and Urban Economics*, *40*(5), 255-271.

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.

Massey, D. S. (2004) Social and Economic Aspects of Immigration, *Annals of the New York Academy of Sciences* 1038(1): 206–212.

Mayer, T., and Zignago, S. (2011) Notes on CEPII's distances measures: The GeoDist database.

Natter, K. (2014) The Formation of Morocco's Policy Towards Irregular Migration (2000–2007): Political Rationale and Policy Processes. *International Migration*, 52, 15-28.

Plümper, Thomas, and Neumayer, Eric (2010). Model specification in the analysis of spatial dependence. *European Journal of Political Research*, *49*(3), 418-442.

Ramon-Gonen, R., and Gelbard, R. (2017). Cluster evolution analysis: Identification and detection of similar clusters and migration patterns. *Expert Systems with Applications*, *83*, 363-378.

Shensa, A., Sidani, J. E., Dew, M. A., Escobar-Viera, C. G., and Primack, B. A. (2018) Social media use and depression and anxiety symptoms: A cluster analysis. *American Journal of Health Behavior*, 42(2), 116-128.

Škuflić, L., Krpan, M., and Žmuk, B. (2018). Migration and the economic crisis in the European

Union member states: cluster analysis. *Entrepreneurship and Sustainability Issues*, 6(2), 979-1001.

Thomann, E., and Rapp, C. (2018). Who deserves solidarity? Unequal treatment of immigrants in Swiss welfare policy delivery. *Policy Studies Journal*, 46(3), 531-552.

Massey, D. and Pren, K. (2012) Unintended consequences of US immigration policy: Explaining the post-1965 surge from Latin America. *Population and Development Review* 38(1): 1–29.

Neumayer, E., and Plümper, T. (2010) Spatial effects in dyadic data. *International Organization*, 145-166.

UNDESA (2020) International Migrant Stock database, United Nations Department of Social and Economics Affairs, New York, accessed in January 2021

Van Hear, N. Bakewell, O. and Long, K. (2018) Push-pull plus: Reconsidering the drivers of migration. *Journal of Ethnic and Migration Studies*, 44(6), 927-944.

Van Liempt, I., & Doomernik, J. (2006). Migrant's agency in the smuggling process: theperspectives of smuggled migrants in the Netherlands. *International Migration*, 44(4), 165-190.

Vertovec, S. (2007) Super-diversity and its implications. *Ethnic and racial studies*, 30(6), 1024-1054.

Vezzoli, S. (2015) Borders, independence and post-colonial ties: The role of the state in Caribbean migration. Maastricht: Boekenplan

Wagner, S., and Wagner, D. (2007). Comparing clusterings: an overview (pp. 1-19). Karlsruhe: Universität Karlsruhe, Fakultät für Informatik.

World Bank (2018) Moving for prosperity: Global migration and labor markets. The World Bank.

Zetter, R. (1991) Labelling refugees: Forming and transforming a bureaucratic identity. *Journal of refugee studies*, *4*(1), 39-62.

■