Università degli Studi di Palermo

Dottorato in Scienze Economiche e Statistiche

Dipartimento di Scienze Economiche, Aziendali e Statistiche

SECS-S/05 - Statistica Sociale

Student mobility: A systemic approach to preferential patterns in higher education

IL DOTTORE

Vincenzo Giuseppe Genova

IL COORDINATORE

Andrea Consiglio

IL TUTOR

Massimo Attanasio

IL CO-TUTOR

Michele Tumminello

CICLO XXXIV

ANNO CONSEGUIMENTO TITOLO 2022

"You can be anything you want to be.

Just turn yourself into anything you think that you could ever be.

Be free with your tempo, be free, be free!

Surrender your ego, be free, be free to yourself!"

— Innuendo, Queen

To my family

Contents

List of Tables

List of Figures

Introduction

Human migration is an important issue and one which has almost always been present in history. From the point of view of the country of origin migration is seen as a loss of human capital in favour of the receiving nation. For receiving countries, meanwhile, inflows promote the formation of ethnic minorities that can, over the long run, change the social, cultural, economic, and political aspects of the country [22]. This migration process is often a community action–fostered by social, economic, and political change. It affects both sending and receiving areas. Furthermore, sometimes migration is facilitated by increasing access to education and information. Thus, migration is complex and is best studied across different disciplines with a variety of approaches and theories [16, 22]. These theories can be grouped into two broad families: those that are "functionalist" and those that are "historical-structural". Functionalist theories analyse society as a system composed of interdependent parts that are in equilibrium. Historical-cultural theories emphasise how social, economic, cultural, and political structures condition individuals in ways that are generally not in balance [16, 22].

Neoclassical theories of migration, broadly following the functionalist approach [22, 75], are based on the idea that human migration is mostly influenced by labour market opportunities. Migration slows down as soon as the labour market balances supply and demand, in both the place of origin and the place of destination. Thus, neoclassical theories describe migration in terms of geographical disparities in labour supply and demand [53, 66]. Furthermore, Lee [64] argued that the decision to leave is influenced by two types of factors. Push factors–everything that pushes people to leave their place of origin, and pull factors–everything that attracts migrants towards a particular destination.

Recently, neoclassical theories have been criticised for unrealistic assumptions. For instance, there are the notions that people are rational actors that decide to move if costs are less than benefits; migrants have perfect knowledge of the opportunities at destination; and that markets are perfect and accessible to everyone [22]. Actually, in the recent literature migration choices depend on many factors such as age, gender, social contacts, preferences, perceptions of the outside world, and relationships with kin, friends, and community members. Indeed, sociologists have stressed that migrant behaviour is strongly influenced by family connections and community dynamics [90].

From the historical-structuralist point of view, migration is seen as a strongly patterned process instead of a random mobility process. Indeed, scholars of historical-structural theories have argued that migrants' choices are constrained by structural forces such as social stratification, market access, inequalities, background *etcetera*. Historical-structural theory, then, describes migrants as people forced to move as an effect of the global political-economic system [22]. In other words, historical-structural theory assumes that migration boosts the profits of the receiver country, draining capital and skills from the areas of origin. Such a mechanism increases the unbalance between the country of origin and destination [23, 27, 100].

As well as neoclassical theories, historical-structural theories have also been strongly criticised for portraying migrants as victims of global capitalism. Both neoclassical theories and historical-structural theories are too restrictive for understanding the complexity of migration. In an attempt to overcome the limitations of neoclassical and historical-structural theories, the *new economics of labour migration* has emerged. Exponents of the new economics of labour migration such as Stark [102, 61] have argued that migration processes are not at an individual level, but usually at the family or household level. Hence, with respect to previous theories based on income maximisation, this approach is based on income risk minimization: here families or households share the risks of migration [104]. The household approach adopted by the new economics of labour migration is useful for explaining migration in both countries affected by economic and social disparities and in wealthy countries. Indeed, when social securities are scarce and we are in a context of high income risks, mutual help and risk sharing within families proves important [103, 63, 61, 74].

Whatever the motivation that drives individuals to move (income, access to new markets, *etcetera*), these theories do not allow us to understand entirely the reasons why transnational movements tend to perpetuate in space and time even when migratory choice has finished producing its effects [76]. Thus, considering migration as the result of a decision taken within a social network of relatives and friends helps us to explain this kind of persistence, that is, chain migration [46, 9]. Chain migration is based on the *migration network theory* that explains how migrants create and maintain social ties facilitating further migration over time. Mabogunje in [69] argued the importance of a feedback mechanism between migrants at destination and fellow villagers at the point of origin. Indeed, positive information transmitted to the place of origin encourages further migration and establishes structured mobility patterns between the country of origin and destination [69]. As Massey *et al.* put it in [76]: *migrant networks can be defined as sets of interpersonal ties that connect migrants, former migrants, and non-migrants in origin and destination areas through bonds of kinship, friendship, and shared community origin.* Furthermore, these migrant networks reduce the economic, social, and psychological costs related to migration, and they increase the probability of moving for the next generation, of whatever social origin. *Once the number of network connections in an origin area reaches a critical level, migration becomes self-perpetuating because migration itself creates the social structure to sustain it* [74]. With respect to "functionalist" and "historical-structural" theories based on financial and human capital, social capital now joins them as a third resource that affects the aspiration and ability to migrate.

This thesis focuses on a particular type of migration: the mobility of university students, which has its specific features, sometimes different from the ones usually studied in demography. Indeed, student mobility is usually characterized by a "temporary" perspective and we can distinguish two types of student mobility: credit mobility [18, 10, 60, 51, 17, 92], and degree mobility [113, 50, 72, 18]. Credit mobility refers mainly to Erasmus students, degree mobility refers to those students that decide to study outside their home region. In this work, we deal with the domestic degree mobility discussed in literature by many authors such as Barrioluengo *et al.*, Dolinska *et al.*, and Van Bouwel *et al.*. They have illustrated how domestic mobility is inhomogeneous within countries. They note that only some universities inside a given country are much in demand from students (*e.g.* Cyprus, Hungary, Lithuania, and Poland) [8, 34]. Van Bouwel *et al.* in [15] focuses on two different perspectives: the "consumption perspective" which is not strictly related to universities prestige or educational quality, but, rather, to the urban services in which universities are located; and the "investment perspective" which is related to a university's prestige. In the UK, there is a specific strain of the literature on domestic student mobility (*e.g.* [35]), which focuses on the role played by "prestigious" universities, such as Oxford and Cambridge, belonging to the "Russell group" [55]. These universities are attractive due to both their prestige and the urban services of the cities they are based in.

In Italy, in the last twenty years, there have been significant changes in terms of student flows and mobility. Student enrolment has decreased significantly, especially since the economic crisis of 2008, with consistent recovery in the last five to six years. On the other hand, student mobility from the South to the Center and North of the country has increased, 30% of students living in the South decided to enrol in universities in the Center-North in 2017 [4]. Boscaino *et al.* in [14] argue that studying in Center-North universities gives better job-market opportunities to students. Furthermore, Santelli *et al.* in [99] showed how southern regions are affected by an increasing rate of students–especially from Sicily–moving to other regions arguing that mobility to the North is driven by job-market opportunities at destination. D'Agostino *et al.* [29] and Impicciatore *et al.* [57] note how this mobility is also affected by contextual factors such as the social class and family background of students.

All these mechanisms involve public information and contextual factors that cannot be used to explain specific South-to-North mobility patterns. Since our main goal is that of studying the presence of preferential mobility patterns, we invoke the paradigm of chain migration in demography. Specifically, we look at "student chain migration" within the broader context of chain migration in demography. This is, by no means, straightforward since with respect to classical chain migration–where movers share information with primary social contacts about labour market opportunities and/or quality of life at destination–"student chain migration" refers also to the enrollment process at university. Here primary social contacts enrolled in a university provide broad information on the university and the place of destination, help newcomers during the enrollment process and provide concrete help on arrival [89].

In the literature, studies on chain migration in students' mobility are very few and mostly based on qualitative methods [89, 88, 19]. In the case of Pérez and McDonough [88] the analysis of chain migration in students' mobility is conducted through semi-structured interviews collected in a non-representative sample of the students' population. They showed in [88] that the presence of members of Latinos' community in the USA played a role in the Latinos college choice. Similarly, Brooks and Waters showed in [19], through the same technique, the importance of social/family networks in the choice of study in the UK for 83 international students.

The work of Person and Rosenbaum [89] combines the use of semi-structured interviews with the use of a survey conducted on students from 14 colleges. They conclude that a friends/kinship network at the destination university increases the probability of enrollment in that specific university. An interesting aspect that clearly emerges from the work of Person and Rosenbaum [89] is that information about destination plays a fundamental role in the choice of college for all students but Latino students. These are characterised by a lack of information compared to non-Latino students, highlighting the way that a social network is the main tool for trying to fill in this lack of information.

The objective of this dissertation is to analyse student mobility from the South to the North of Italy and to investigate the presence of chain migration effects in student migration over the last 15 years. Our research questions can be reduced to: is there a chain migration effect in student mobility? If so, what are the most significant mobility patterns? Are there similar mobility patterns in different areas of origin? And are mobility patterns stable over time?

To the best of our knowledge, in the literature there are no quantitative techniques available for identifying patterns of student mobility on student chain migration. To answer to these research questions, we apply two methods of data analysis from the edge of Statistics and Network Theory. We consider three "big" case studies (Apulia, Sardinia and Sicily), for identifying patterns of student mobility that can be partially rooted in chain migration. These methods reveal statistically significant extra-flows with respect to a null hypothesis that fully takes into account the intrinsic heterogeneity of the entities involved in the analysis (universities, regions, areas of origin, *etc.*). Chapter 1 and Chapter 3 focus on the preferential patterns of mobility originating in Sicily. Chapter 2, meanwhile, provides a comparative study of three different regions of origin, namely, Apulia, Sardinia, and Sicily.

The structure of the thesis is as follows. Chapter 1 includes the introduction to the dataset and the network analysis of preferential patterns of mobility originating in Sicily from the high-school to enrolment in bachelor degrees; in Chapter 2 we illustrate some descriptive statistics and the statistical method for measuring the chain migration effect for Apulia, Sardinia, and Sicily; Chapter 3 illustrates the network analysis of all preferential mobility patterns from high-school to a master's degree–for students from Sicily. Finally, we offer some conclusions.

Chapter 1

Student mobility in higher education: the Sicilian outflow network and chain migration

Abstract

The most important student mobility (SM) flow in Italy is from the Southern to the Central-Northern regions, a phenomenon that has been magnified by an increasing number of outgoing students from Sicily over the last decade. In this chapter, we rely upon micro-data of university enrollment and students' personal records for three cohorts of freshmen, in order to investigate preferential patterns of SM from Sicily toward universities in other regions. Our main goal is to reveal the existence of chain migrations, where students from a particular geographical area move towards a particular destination to follow other students that have previously moved. The chapter provides aspects that are innovative under the view of the data, of the application, and of the statistical method. The data from each cohort is represented as a tripartite network with three sets of nodes, namely, clusters of Sicilian municipalities, students, and universities. The tripartite network is projected in a bipartite weighted network of clusters and universities, which is, then, filtered, in order to obtain a statistically validated bipartite network (SBVN). The SBVNs of the three cohorts may suggest the existence and evolution of chain migration patterns over time, which are also gender specific.

1.1 Introduction

The Italian public universities are subsidized within a competitive framework that economically awards excellence, efficiency, and the capacity of universities to attract students from Italian regions other than its own. However, repeated cuts to public spending over time greatly disadvantaged the Southern universities, also considering the well-known Italian infrastructural and economic North-South divide. Therefore investigating the mechanisms that might explain the presence of a net flow of students from the South to Northern universities appears very important, and “chain migration” might be one of such mechanisms. Over the last decade, student mobility (SM) flows from Southern regions [3, 110] to Northern universities have been growing at an increasing rate, especially from Sicily, in spite of the presence of four universities in the island. Such unidirectional flows imply an increasing loss of human capital for the Southern regions, as most of the students does not come back in the origin region at the end of their studies [36]. In the literature, the Italian SM has been studied by following two approaches of analysis, depending on the available data. The first approach uses macro-data to describe flows of students moving from an area to another one and to detect the determinants of mobility. This approach uses information at an aggregate level of macro-area, such as province, region, or macro-region [48, 20]. The second approach is based on micro-data, which allows to include individual student characteristics [29, 68]. Recently, availability of micro-data at national level allowed to perform longitudinal analyses [39]. The previous contributions highlight the reasons for SM, in particular from South and Islands, are based on: *i)* individual student characteristics, such as high school type and final mark, *ii)* the attractiveness of the universities, and *iii)* the regions where these are located in. At the level of university, some determinants of the attractiveness can be represented by the quality of both research and teaching [26], a different training offer, and the availability of facilities and scholarships [31]. Instead, at the level of regions, the Central-Northern ones offer higher quality of life and better job opportunities after graduation [37]. For these reasons, Central-Northern universities located in towns such as Turin, Milan, Bologna and Rome are often the preferred destination for Southern students who decide to migrate. However, while it is clear which the destination universities are, we believe these determinants are only a part of the reasons that boost the student to migrate. Indeed, students may also choose an out-of-region university by following a chain migration. According to the concept of chain migration [70], a student can find initial accommodation and other facilities by means of primary social relationships with previous movers, or migrants in general. In the broadest sense, such a phenomenon has been extensively studied and theorized in economic and sociological contexts [54]. Literature contributions on chain migration in higher education mainly focus on international student mobility. Pèrez in [88] investigate on chain migrations of Latinos students in Los Angeles, Camila in[32] analyzes the migration of Peruvian students enrolled in Brazil, while Serdar [109] deals with foreign students in Turkey. Other examples are [19, 86].

The chapter provides new insights from the perspective of both statistical method and application. The statistical method developed in this chapter is a generalization of the exact test developed by Tumminello *et al.*[108] to check for the hypothesis of random co-occurrences in a tripartite network[1]. The application is innovative because we apply a statistically validated network analysis to administrative micro-data of migration flows to investigate the possible existence of a chain migration in SM. The network analysis through the proposed method is to be intended as an exploratory tool and it does not allow to evaluate the simultaneous effects of exogenous factors which could explain the SM. Indeed, our main objective is to test, beyond any other reason of university attractiveness [26], the possible existence of a chain migration [19], in which students move to follow other students coming from the same area of origin. Our hypothesis is that flows of Sicilian moving students can significantly differ in the path "area of origin - University of destination", with respect to a null hypothesis of random flows. To define the area of origin, we consider an aggregation level of the Sicilian municipalities, based on some homogeneity criteria, thinner than the 9 Sicilian provinces. The areas we used are 38 clusters of the 390 Sicilian municipalities (Figure 1.1), aggregated by geographical proximity, and economic and commercial criteria according to D'Agostino *et al.* [30]. In particular, these clusters were defined starting from 38 main municipalities, detected as gravitational areas, because their larger flows of commuting for work and study, with respect to other municipalities geographically close. Moreover, to enforce the homogeneity of the clusters, the authors hypothesize that the higher the economic and commercial levels, the larger the use of the Italian language instead of the local dialects. The definition criteria of these clusters are consistent with the existence of a social network of communication among people within each cluster. Through the above definition of areas of origin it may be possible to reduce the effect of exogenous factors such as the ones specified at economic and commercial

[1] The proposed method is equivalent to test that *integer* weights in a projected bipartite network occur by chance.

levels. Of course the method does not allow to consider the individual characteristics of the students.

We rely upon micro-data of university first enrollment students and students' personal records for three cohorts of freshmen, the 2008/09, the 2011/12 and the 2014/15.

The data from each cohort is represented as a tripartite network with three sets of nodes, namely, clusters of Sicilian municipalities, students, and universities. The tripartite network is projected in a bipartite weighted network of clusters and universities, which is, then, filtered, in order to obtain a statistically validated bipartite network (SBVN), which represents a generalization of the method introduced by Tumminello *et al.*[108], to deal with tripartite systems. Specifically, a directed edge from a Sicilian cluster to a university is set if the flow is statistically significant with respect to a null hypothesis of random flow of students, which takes into account the heterogeneity of both universities (total inflow) and clusters of municipalities (total outflow).

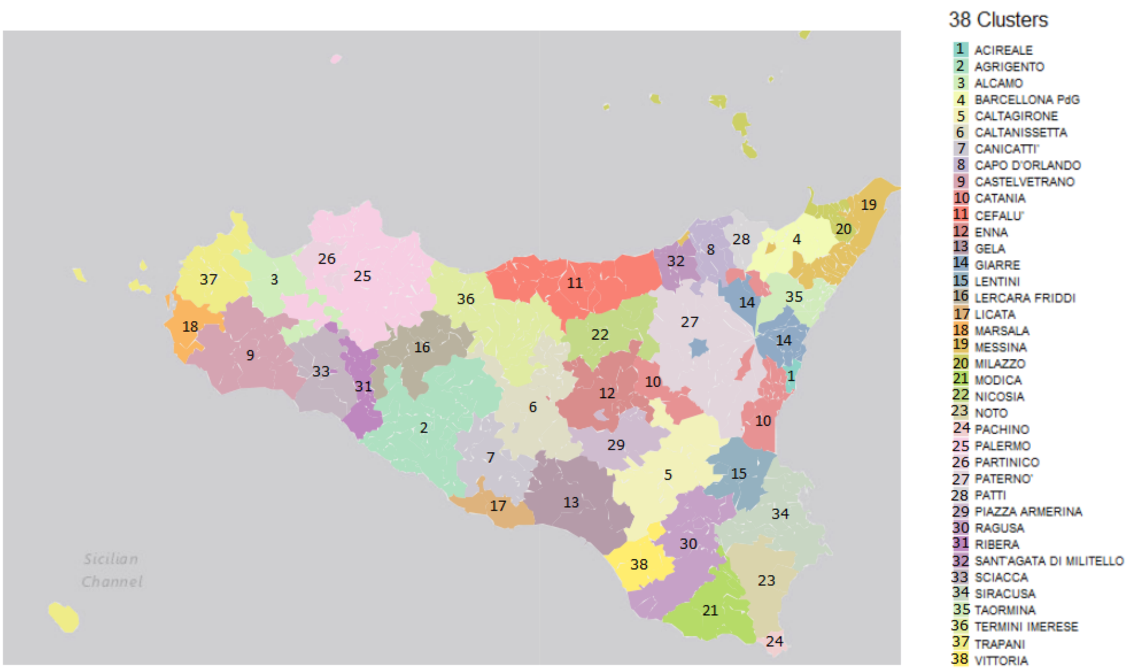


Figure 1.1: The 38 clusters of Sicilian municipalities according to D'Agostino *et al.* [30].

The structure of the chapter is the follow. Section 1.2 includes a detailed description of the micro-data used throughout the chapter; section 1.3 describes the network representation of the system and the statistical method used to analyse and prune the projected bipartite network of universities and clusters; section 1.4 includes empirical results from the network analysis of the system; finally, in section 1.5, we draw our conclusions.

1.2 Data description

Since 2000, the Ministry of Education and Research (MIUR) has yearly collected administrative data–Anagrafe Nazionale Studenti (ANS)–at the student-level from all higher education institutions. The empirical analysis reported in this thesis relies upon data extracted from the MOBYSU.IT database [79]. This database includes data from the ANS and is specifically built to longitudinally monitor the academic career of all the students enrolled in any Italian university since 2008. We had access to the MOBYSU.IT database thanks to a research protocol between MIUR and the Universities of Palermo, Cagliari, Florence, Naples Federico II, Sassari, Siena and Turin. The main objective of the protocol is to study the mobility of Italian students within the country, starting from micro-data included in the ANS that allow to reconstruct the entire career of all Italian students. The database also includes socio-demographic information of all the students, together with their educational achievements from the high school to the Master degree. MOBYSU.IT database consists of about 300 variables per record, where each record is a student, and the total number of records ranges between 270.000 and 295.000, depending on the cohort. According to the objective of the study reported in this chapter, information about students coming from Sicilian municipalities has been extracted from MOBYSU.IT for the time period 2008-2014[2]. The extracted subset includes more than 26,000 records per cohort. In this work, we consider the cohorts of enrolled students as coming from a metapopulation, that is, a (numerable infinite) set of subjects, belonging to certain sub-populations, which are independent subsets of the common metapopulation. The subpopulations are renewed every year at the enrollment, according to a process that makes them independent of the previous and the following subpopulations, each one with structure and characteristics common to the metapopulation (representativeness), but, at the same time, carrying peculiar characteristics. Therefore, we assume that they are metasamples, which can be studied from a statistical point of view, by using inferential methods.

Table 1.1 shows that the percentage of females is grater than the percentage of males at both national and local level.

[2] Checked and organized data for the cohort 2017 became available in 2020. At this time, *Electronic Journal of Applied Statistical Analysis* already published the paper associated with Chapter 1 of the present thesis, and we were working on the analyses reported in Chapters 2 and 3 of the thesis. Therefore, we included that cohort in the latter studies, but it is missing in the first one.

Table 1.1: Percentage of freshmen by Gender and Year

	Italy			Sicily		
	Years			Years		
Gender	2008	2011	2014	2008	2011	2014
F	57%	56%	55%	57%	57%	55%
M	43%	43%	44%	42%	42%	44%

Table 1.2 shows the distribution of students with respect to the high-school they come from. It turns out that there is no apparent difference between Sicilian students and students coming from other regions, in spite of the cohort, although, as expected, we notice relevant differences with respect to the gender of students.

Table 1.2: Distribution of Sicily freshmen and out of Sicily freshmen students by Highschool and Year

		Sicily			All other regions		
		Year			Year		
Gender	Highschool	2008	2011	2014	2008	2011	2014
F	Other	25%	23%	23%	26%	27%	27%
	Classical	24%	28%	28%	18%	18%	18%
	Professional	5%	3%	4%	5%	5%	5%
	Scientific	31%	35%	36%	32%	34%	34%
	Technical	15%	11%	11%	19%	26%	16%
M	Other	5%	5%	5%	9%	9%	9%
	Classical	14%	14%	14%	9%	9%	9%
	Professional	4%	5%	5%	5%	5%	5%
	Scientific	42%	50%	50%	43%	48%	48%
	Technical	35%	26%	27%	35%	29%	29%

Before analyzing the flows from a cluster (source node) to a university (target node), we provide a descriptive statistics of outgoing students (outflow). Specifically, Table 1.3 shows an increasing number of outgoing students from Sicily over time, a trend that is the opposite of the one observed for the total number of newly enrolled students, *i.e.*, freshmen, which, jointly, amplify the relative magnitude of the phenomenon.

Table 1.3: Total and outgoing sicilian-freshmen by level degree courses over time, in parenthesis the percentage of outgoing Sicilian students by degree courses and cohort

	Year					
	2008		2011		2014	
Degree	Total	Outgoing	Total	Outgoing	Total	Outgoing
Bachelor	21327	2997 (14.1%)	17210	4115 (23.9%)	16719	4682 (28%)
Master (5/6 years)	5264	675 (12.8%)	5791	1058 (18.3%)	5261	1087 (20.7%)
Total	**26591**	**3672 (13.8%)**	**23001**	**5173 (22.5%)**	**21980**	**5769 (26.2%)**

Furthermore, our analysis aims at distinguishing among 38 Sicilian territorial areas, which are internally homogeneous. This choice is a trade-off between clustering data using 9 provinces (too large and heterogeneous), and just referring to the 300 Sicilian municipalities (too small to analyze migratory effects). In fact, being the selected areas aggregated on the ground of socio-economics characteristics it is possible to evaluate such chain migration effects in spatio-temporal terms by applying a network analysis. According to D'Agostino *et al.* [30], we have classified the variable *residence city* into the 38 areas described by D'Agostino and Ruffino creating a new variable that represents the homogenous residence area of the *i-th* student. In this framework the variables of interest have used in this analysis are the cluster of residence, and university of enrollment.

As it can be seen in Figure 1.2, a strong increase of the outgoing rate from Sicily is observed from the first cohort, 2008, (panel *a*) to the last one ,2014, (panel *c*). Indeed, the number of clusters with an outgoing rate of more than 20% increased a lot through the time window 2008–2014.

Figures 1.3, 1.4, and 1.5 are the heatmaps of Sicilian students flows from the 38 Sicilian clusters (by row) towards the off Sicily universities (by column). The ordering of universities and the one of areas of origin is provided by the Complete Linkage Cluster Analysis (CLCA) [2] as applied to the Euclidean distances between universities and between areas of origin, respectively. CLCA has been selected among other hierarchical clustering procedures, since it is the one that provides the best separation among groups [2]. Colours in the figures are used to describe the fraction of outgoing students and to distinguish between geographical macro-regions. A straightforward comparison of the figures shows, especially for the last cohort, that such flows are more concentrated (colours orange and yellow) in Central-Northern universities (colours green and grey on the top of the plot). The clustering of 38 areas of origins, which is based on student flows, suggests the presence of homogeneous aggregations based on geographical proximity of the areas. For example, looking at the 2014 cohort, the cluster formed by Siracusa, Modica, Ragusa, Noto, Caltagirone, and Pachino can be easily interpreted along this line of thinking, since these territories are all located in the Southern-Eastern Sicily.

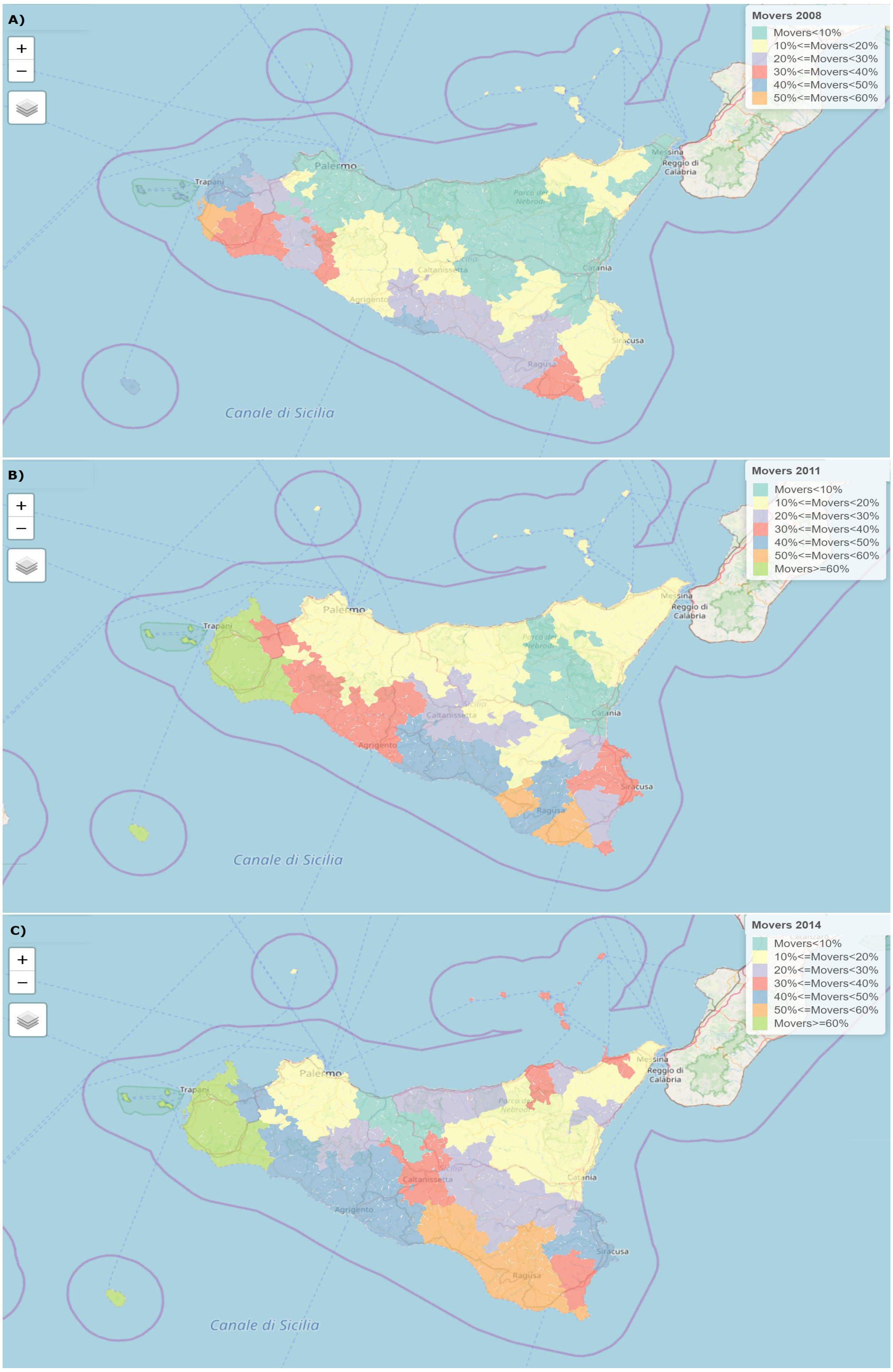


Figure 1.2: Panels *a*), *b*), and *c*) report the outgoing rates arranged in 7 levels by cohort per cluster for years 2008, 2011, 2014.

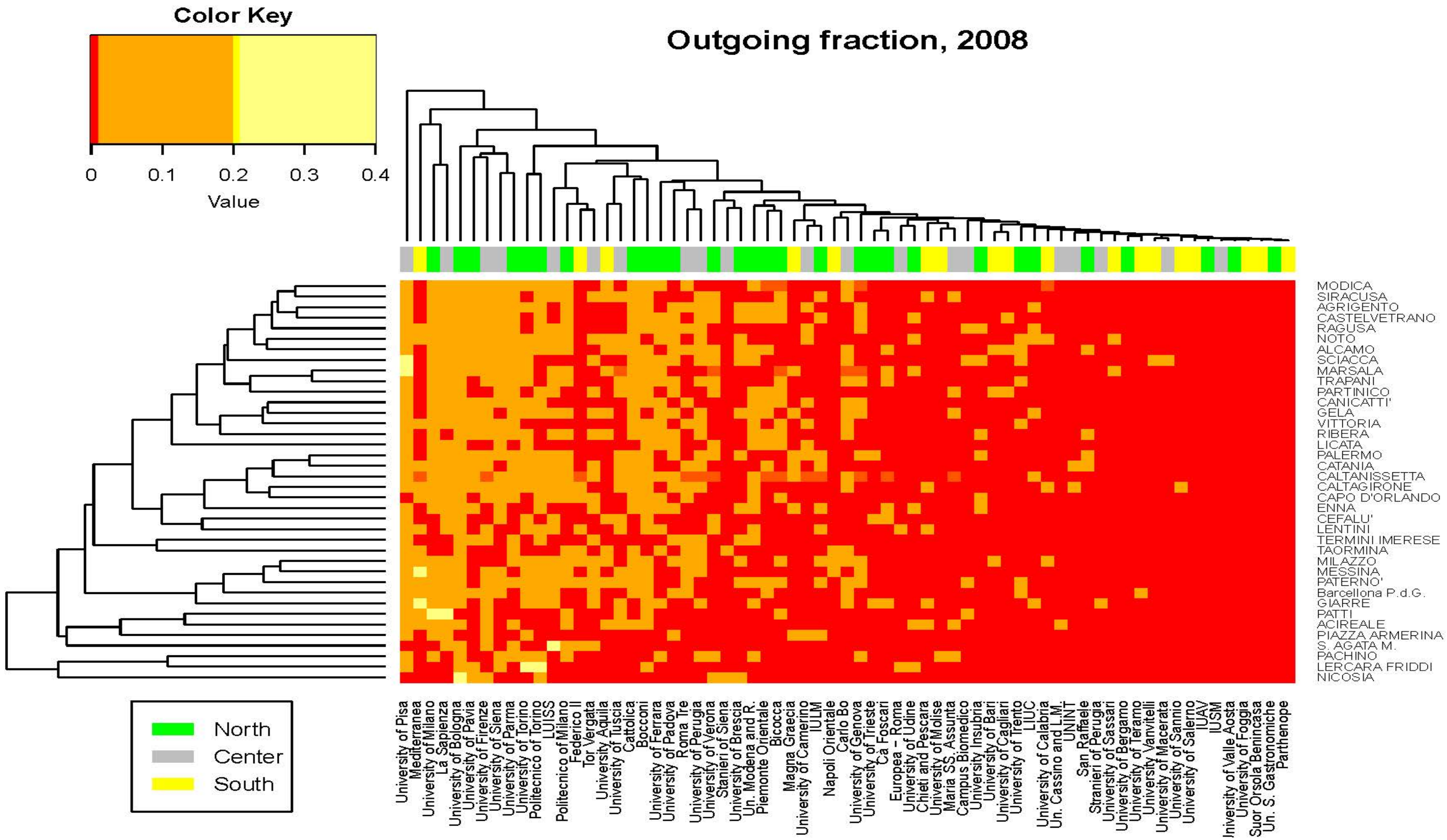


Figure 1.3: Heatmap representation of Sicilian students flows. Rows correspond to the 38 Sicilian territorial clusters of origin, while columns correspond to off Sicily universities of destination. The order of elements by row and by column is provided by complete linkage cluster analysis. Data correspond to year 2008. Heatmap colours, ranging between red and yellow, describe the fraction of outgoing students, while the other colours distinguish between geographical macro-regions.

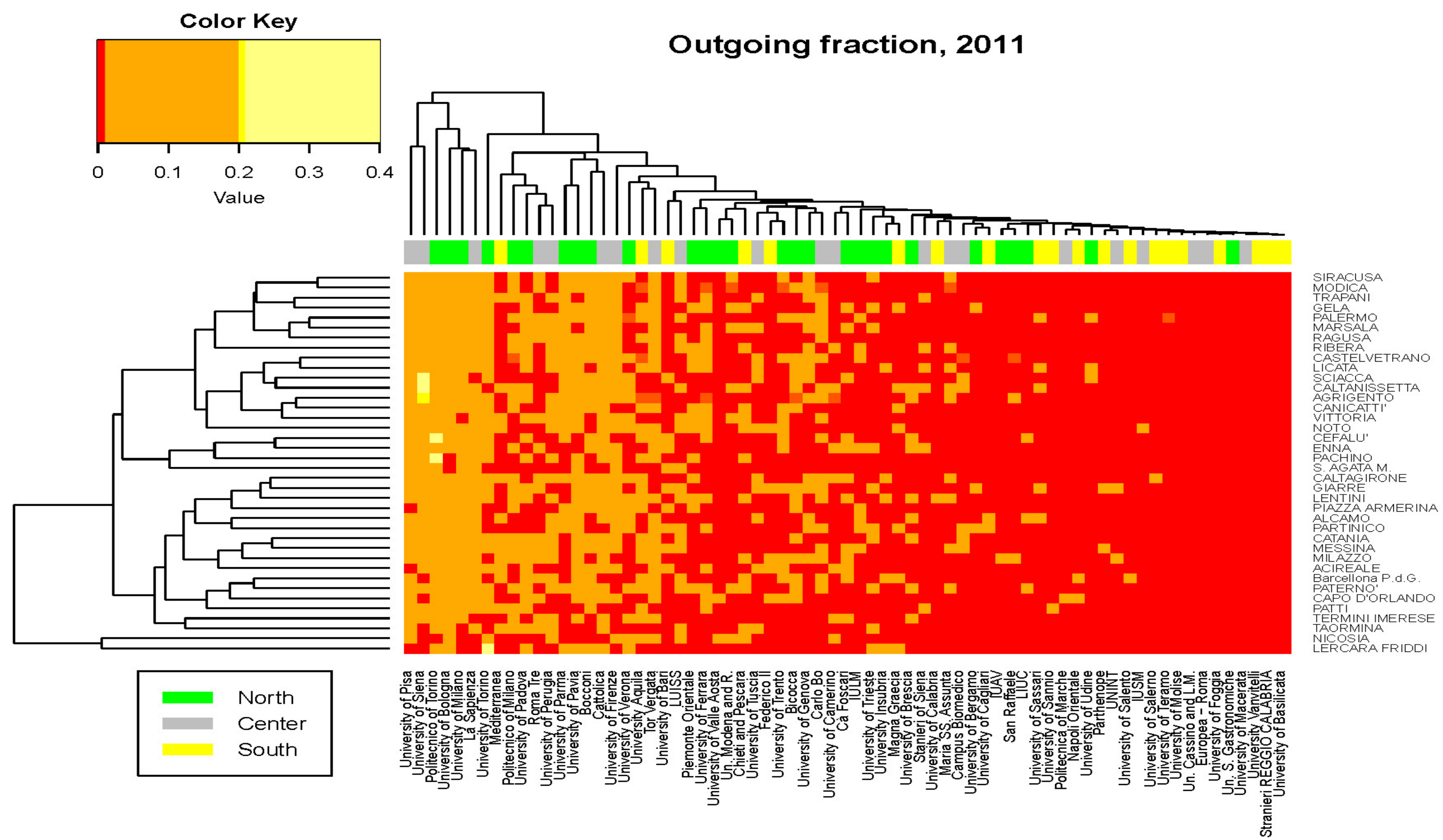


Figure 1.4: Heatmap representation of Sicilian students flows. Rows correspond to the 38 Sicilian territorial clusters of origin, while columns correspond to off Sicily universities of destination. The order of elements by row and by column is provided by complete linkage cluster analysis. Data correspond to year 2011. Heatmap colours, ranging between red and yellow, describe the fraction of outgoing students, while the other colours distinguish between geographical macro-regions.

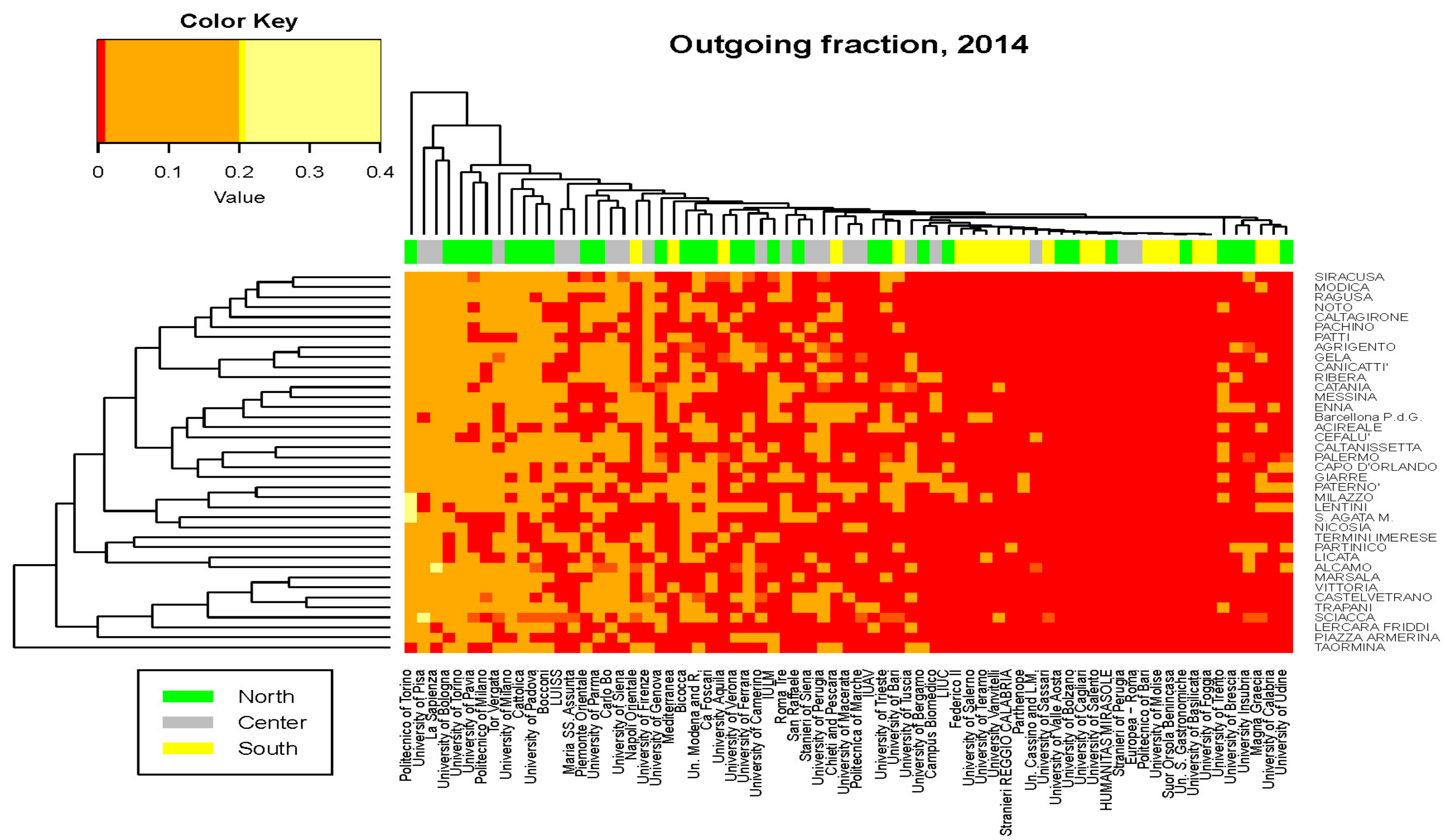


Figure 1.5: Heatmap representation of Sicilian students flows. Rows correspond to the 38 Sicilian territorial clusters of origin, while columns correspond to off Sicily universities of destination. The order of elements by row and by column is provided by complete linkage cluster analysis. Data correspond to year 2014. Heatmap colours, ranging between red and yellow, describe the fraction of outgoing students, while the other colours distinguish between geographical macro-regions.

1.3 Methods

Students' mobility system displays several typical features of complex systems, such as, interconnectedness of elements (which is at the basis of the chain migration hypothesis), heterogeneity of elements, that is, heterogeneity of the clusters of municipalities according to number of outgoing students, of the universities, according to the number of enrolled students, heteroscedasticity, modular structure of elements (*e.g.*, of universities), non stationarity, and tipping points (reflecting, for instance, an economic downturn) [7]. Such features make the system very difficult to analyze by using traditional means based on system's decomposition and reductionism [101, 33], and taking a weak-holistic approach [101], such as the one provided by network theory [82], seems to be more appropriate. The most natural network representation of the mobility system is obtained by considering a tripartite network associated with each cohort, where three sets of nodes, namely, clusters of Sicilian municipalities, students, and universities, are distinguished. In this network, a link between any two elements of the same set cannot occur. Specifically, in the tripartite mobility network, a link is set between a student and a university if the student is enrolled in that university, and between a student and a cluster of municipalities if the student comes from a municipality that belongs to that cluster.[3]

The objective of the present chapter is to understand whether or not student mobility might be driven, in part, by a chain migration. Along this line, we are interested in eliciting from data the presence of an *excess* of flow from one cluster of municipalities to a university, with respect to a null hypothesis of random flow that takes into account the heterogeneity of both clusters and universities. Such an analysis can be done by considering a straightforward generalization of the Statistically Validated Network method to tripartite networks [108]. Specifically, we adapt this technique to construct a Statistically Validated Bipartite Network (SBVN) of clusters and universities, as detailed in the next section.

[3] It is worth noticing that the structure of the network is such that no link can be set between a university and a cluster of municipalities (see Fig. 1.6). This fact implies that the whole information contained in the tripartite network can actually be mapped in a bipartite network with weighted links, that is, a network with only two sets of nodes, clusters and universities, and weighted links connecting elements of the two sets, the weight being equal to the number of students coming from a cluster and enrolled in a university. However, such a simpler representation, would make it harder to describe the configurational model underlying the null hypothesis of random flows described later in the chapter. Therefore, we prefer to use the tripartite representation of the system.

1.3.1 Statistically validated bipartite networks

Starting from the tripartite network of clusters, students, and universities, a bipartite weighted network is obtained by projecting the set of students onto the other sets– see Figure 1.6. Specifically, a link between a cluster of municipalities and a university is set if at least one student is linked to both nodes in the original tripartite network. The weight of such link in the bipartite weighted network corresponds to the number of students linked to both nodes in the tripartite network [82, 108].

Figure 1.6 shows the original tripartite network (Phase 1), where intermediate nodes between the 38 Sicilian clusters and the 80 universities are the $n_q,\ q = 1, 2, 3$, outgoing students[4] in a given cohort under analysis, as well as the corresponding weighted bipartite network between clusters and universities, which is obtained as a projection of the original tripartite network. As discussed above, our objective is to determine the presence in the data of an excess flow between a municipality and a university, with respect to a null hypothesis of random flow in a system with a double heterogeneity, which could be interpreted as a mark of the existence of a chain migration in the mobility system. Such an excess of flow can be revealed by performing the following analysis.

1.3.2 Statistically validated networks: construction

To validate links in the (projected) weighted bipartite network of clusters and universities, the method proposed by Tumminello *et al.* [108] has been generalized to deal with tripartite systems. In its most general setting, this approach requires the validation of each link of the projected network, against the null hypothesis of random co-occurrence of shared neighbor nodes (students in this case). The null hypothesis of random connections between nodes with assigned degree has been proposed in the seminal paper by Xulvi-Brunet *et al.* [116] in terms of a configurational model, which is also able to reproduce assortative mixing. However, in our case, we consider a null hypothesis that does not assume the presence of such an effect. Therefore, the configurational model, where links (of a real network) are iteratively and randomly selected in pairs and swapped is assumed. Accordingly, the co-occurrence of first neighbors in either set can be analytically described through the hypergeometric distribution. In the original formulation of the method, subsets of projecting nodes (students in this case) were considered, in order to deal with the heterogeneity of node degree in the projecting set. However, here, such an heterogeneity

[4] $n_1 = 3672$ in 2008, $n_2 = 5173$ in 2011, $n_3 = 5769$ in 2014. See Table 1.3

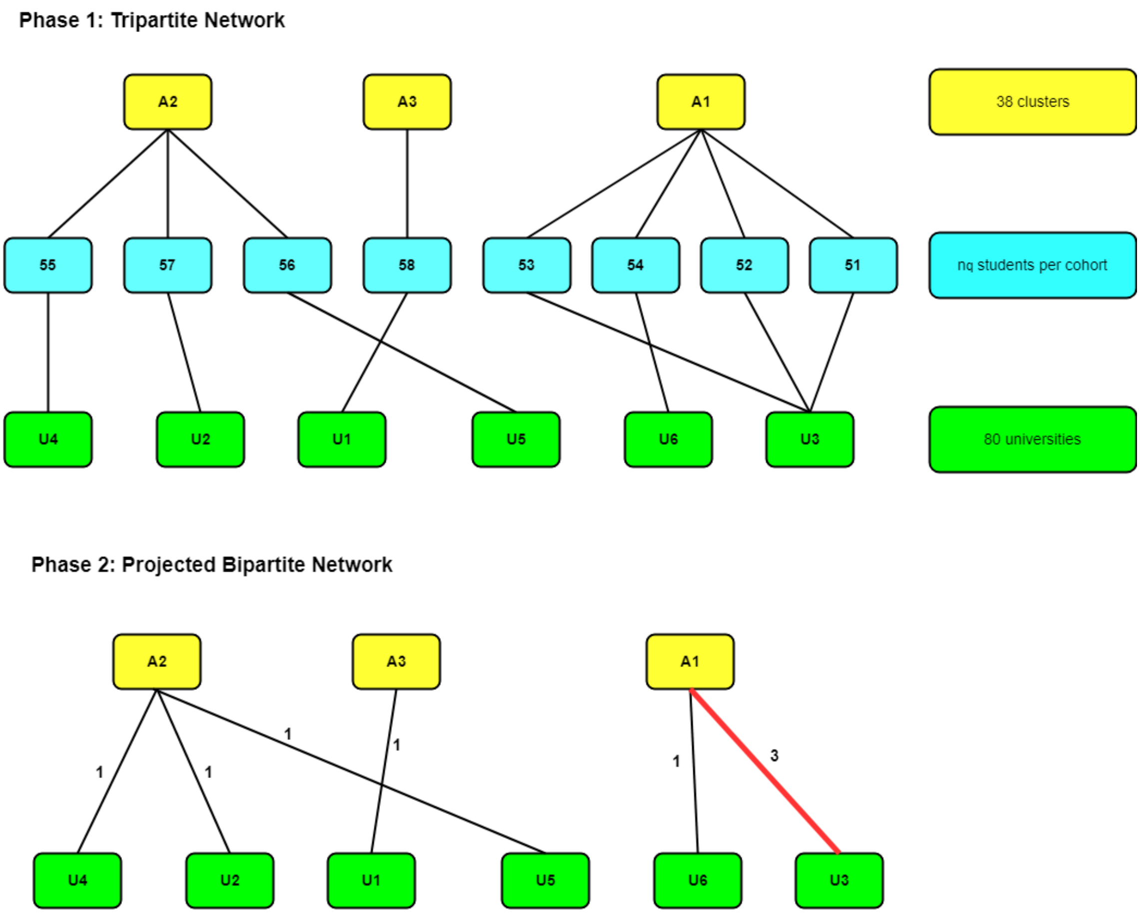


Figure 1.6: The applied projection of the tripartite students' mobility network to the bipartite one, red line shows a statistically validated link and the thickness the intensity of the flow

is missing, since students only have degree equal to two in the original tripartite network, where one link connects a student to a cluster of municipalities and one link connects her to a university. Therefore, to test the statistical significance of an edge weight in the bipartite weighted network for a given cohort against a null hypothesis (H_0) of random flow, we consider the following distribution of the weight X of a link in the (randomly rewired) projected network that describes hypothesis H_0:

$$H(X|N_A, N_B, N) = \frac{\binom{N_A}{X}\binom{N-N_A}{N_B-X}}{\binom{N}{N_B}}, \tag{1.1}$$

where:

- N is the total number of students in the tripartite network,
- N_A is the number of students coming from municipality A,

- N_B is the number of students enrolled in university B,
- X is the number of co-occurrences of A and B in the (random) tripartite network associated with H_0, *i.e.*, it is the variable that describes the number of students *flowing* from cluster A to university B.

Indeed, probability (1.1) can be obtained by a straightforward calculation of the number of "favorable events" divided by the total number of possible events. Specifically, the result can be obtained by assuming that university B randomly selects N_B students (which allows to keep information about the heterogeneity of universities in the null hypothesis) from the overall N students in the population, N_A of which coming from cluster A (which allows to keep information about the heterogeneity of clusters of municipalities in the null hypothesis) and $N - N_A$ from the other clusters. Therefore, if X represents the number of students picked by university B that come from cluster A, then the number of favorable cases is given by the product of the number of ways in which N_A students can be combined in groups of size X, that is, $\binom{N_A}{X}$, times the number of ways in which the other $N_B - X$ students can be drawn from the set of students coming from the other clusters, that is, $\binom{N-N_A}{N_B-X}$. Finally, the probability (1.1) is obtained by noticing that all of the possible ways in which N_B students can be drawn from the overall set of students, with size N, is $\binom{N}{N_B}$.

The hypergeometric distribution described in (1.1) is then used to associate a *p-value* with each pair cluster-municipality, say j and k, respectively, which are connected in the (projected) weighted bipartite network and have a weight (number of outgoing students) $n_{j,k}$:

$$p_{j,k}(n_{j,k}) = \sum_{i=n_{j,k}}^{min(N_A,N_B)} \frac{\binom{N_A}{i}\binom{N-N_A}{N_B-i}}{\binom{N}{N_B}}. \tag{1.2}$$

Since the structure of the bipartite weighted network allows one to represent it as a two-way contingency table, one may argue about the advantage of using Eq. (1.2) to associate a *p-value* with each link, instead of the one that could be provided by the standardized residuals of a χ^2 distribution. Actually, the proposed test is exact and, therefore, on the one hand it allows one to better deal with "extreme values" (right tail of the distribution), and, on the other hand, it works even with small samples. The *p-value* calculation reported in equation (1.2) should be repeated for each link in the weighted bipartite network. Therefore, multiple-test corrections on the threshold of statistical significance of a *p-value* should be considered. The most conservative correction with respect to the family-wise error rate (FWER) is the Bonferroni correction [78], which also applies to the case of dependent tests, whereas, a less restrictive correction is the False Discovery Rate (FDR) [11]. To say if a connection between two nodes is statistically significant the threshold on the *p-value*, according to Bonferroni, is

$$S_{Bon} = \frac{\alpha}{T}, \tag{1.3}$$

while, according to the False Discovering Rate, it is

$$S_{FDR} = \frac{\alpha k}{T}, \tag{1.4}$$

where α is the univariate significance level (*e.g.*, 0.01), T is the number of tests, *i.e.*, the number of links in the weighted bipartite network, and k is the rank of the largest tested *p-value* such that $p_{k_{max}} < k_{max}\alpha/T$. Accordingly, being the Bonferroni correction more conservative than the FDR one, if a link is validated according to Bonferroni, it is also validated according to FDR, whereas, of course, the vice versa does not hold true.

1.4 Results

Figures 1.7, 1.8, and 1.9 report the bipartite validated networks for Sicilian outgoing students by cohort. There are two types of nodes: the first type corresponds to universities (red circles), with node size proportional to the number of enrolled students from Sicily, while the second one corresponds to clusters of municipalities (cyan rectangles), with size proportional to the number of outgoing students. The solid line is used to represent links that belong to the *Bonferroni* network (and therefore to the FDR too), while the dashed line is used to represent links that only belong to the FDR network. Moreover, the thickness of a link is proportional to the number of outgoing students from a cluster to a given university. The network representation is obtained through the force directed layout method with weight proportional to the binary Pearson's correlation coefficient between a university and a cluster, which implies that the longer the link, the lower the correlation between two nodes is.

According to both Bonferroni and FDR networks, the most attractive universities are the University of Bologna, Milan, Turin, Florence, Pisa, La Sapienza, and Cattolica. Moreover, we observe that the number of (statistically significant) links increases overtime. To better frame this phenomenon from a quantitative a point of view, we include a table with the number of links and nodes for each cohort. Table 1.4 shows that the number of both nodes and links involved in the statistically validated networks increases between 2008 and 2014. Both the apparent structure of the networks and the increasing number of links over time can be interpreted within the framework of complex systems, by looking at chain migration as reflecting the relevance of private information in the decision process preliminary to the selection of a university by students, and the increasing relevance of such an information with respect to public information. Indeed, in an efficient purely competitive system, where all of the agents share and (rationally) process the same (public) information [38], the networks reported in Fig.s 1.7 through 1.9 should be empty: preferential patterns of mobility shouldn't be observed, since the null hypothesis exactly takes into account the heterogeneity of clusters of municipalities and universities. Such a consideration implies that private information flow acts as a positive feedback mechanism in the system: students from cluster A enroll in university B at time t, then they share their (*positive*) experience, not only concerning the university, but also the city, *etc.*, with their friends at home (information flow from B to A), which, in turn, reinforces mobility from A to B at time $t + 1$. Such an information flow can also act in reverse (negative feedback) and break a preferential connection: students from cluster A enroll in university B at time t (preferential pattern from A to B at time t), then they share their (*negative*) experience with their friends at home, discouraging them from enrolling in university B at time $t+1$, which, as a result, tends to cancel out the preferential pattern of mobility from A to B. The fact that the number of validated links in the mobility network increases over time should therefore be interpreted as reflecting the increasing importance of private information to determine the structure of the mobility system, at the expense of public information.

A question that remains yet to be answered concerns the imbalance between the relative influence of the two feedback mechanisms discussed above. If only a positive feedback mechanism was effective to orient the decision of students to enroll in a certain university, then the statistically validated network associated with the cohort of 2014 should properly contain the statistically validated network associated with the cohort of 2008. Therefore, an indicator to look at is the proportion of links from the 2008 statistically validated networks that also belong to the 2014 corresponding networks. Looking at the Bonferroni network, this proportion is 41%, while it is 44% for the FDR network. This result indicates that *old* mobility patterns are destroyed, which suggests that the aforementioned negative feedback mechanism is actually influencing the evolution of the mobility system. On the other hand, given the overall increase of mobility patterns, we can conclude that the role played by the positive feedback mechanism is even bigger, likely driven also by exogenous factors, such as the 2011 sovereign crisis, and the subsequent austerity policy introduced and enforced in 2012 and 2013, which affected differently the South and the North of the country, especially with respect to the labour market, which should have favored the formation of new mobility patterns from the South to the North of the Country.

Finally, to make the discussion about the evolution of mobility system more comprehensive, and to address the question of whether gender-specific patterns occur or not, we look at the similarity between the networks that are separately obtained for male and female students over time. As discussed already, an increasing similarity between networks for subsequent cohorts implies that migration patterns are settling over time, whereas a decreasing trend might suggest a progressively reducing impact of chain migration on students' mobility, with respect to the negative feedback mechanism discussed in the previous paragraphs and other influential factors, such as marketing campaigns, and, broadly speaking, the varying attractiveness of "specific" universities. Here we focus on the difference between gender specific patterns of mobility over time. Similarity between networks can be easily evaluated by associating with each one of the potential links ($N_u \times N_c$, where N_u is the number of universities and N_c the number of Sicilian clusters) a binary variable that takes value 1 if the link is set in the network and 0 otherwise.

Table 1.4: Number of Links and Nodes per cohort by female, male, and overall.

	Links			Nodes		
Cohorts	Female	Male	Overall	Female	Male	Overall
2008	62	55	95	46	45	52
2011	80	69	110	47	48	54
2014	72	81	123	48	52	62

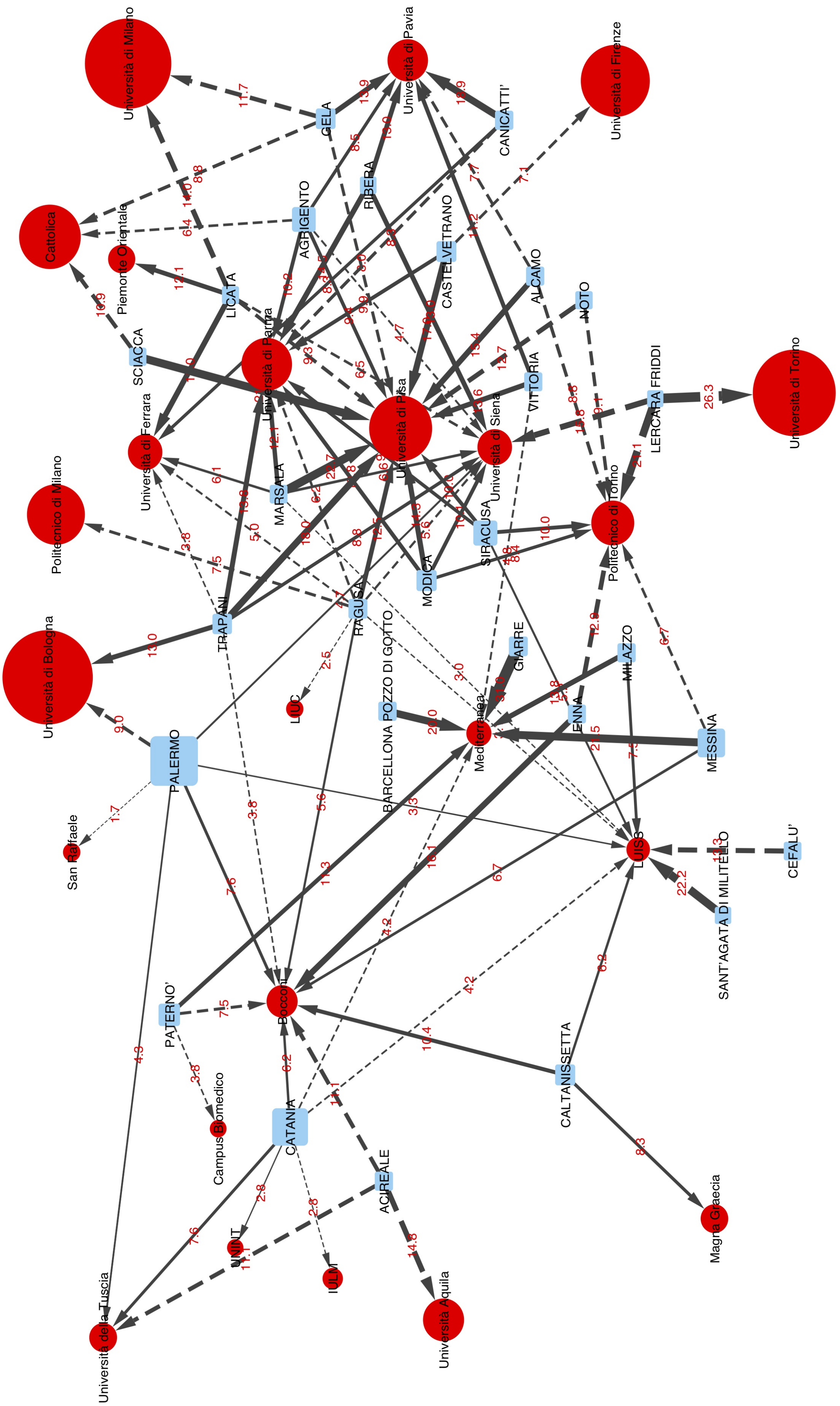


Figure 1.7: Bipartite validated network for Sicilian outgoing students, year 2008

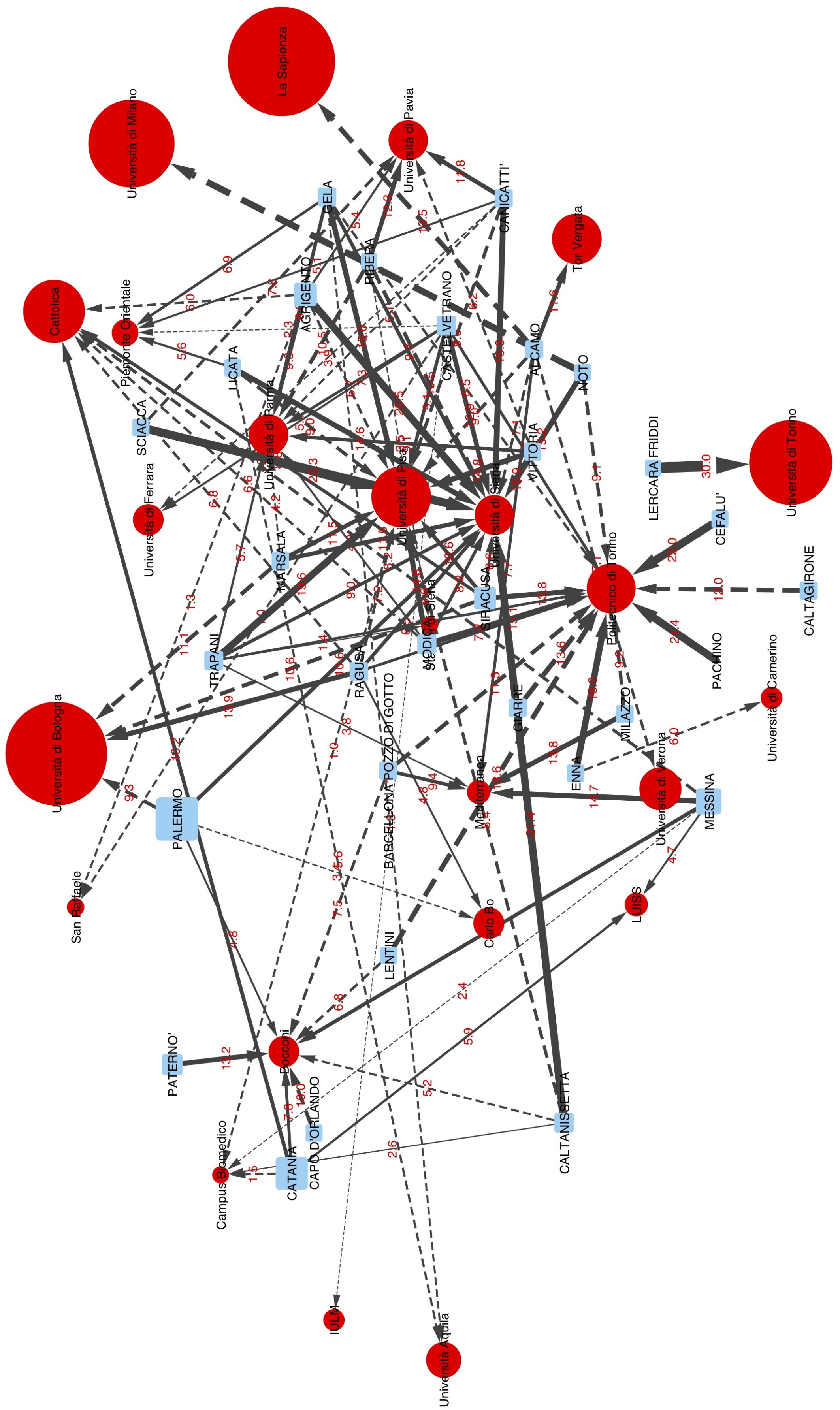


Figure 1.8: Bipartite validated network for Sicilian outgoing students, year 2011

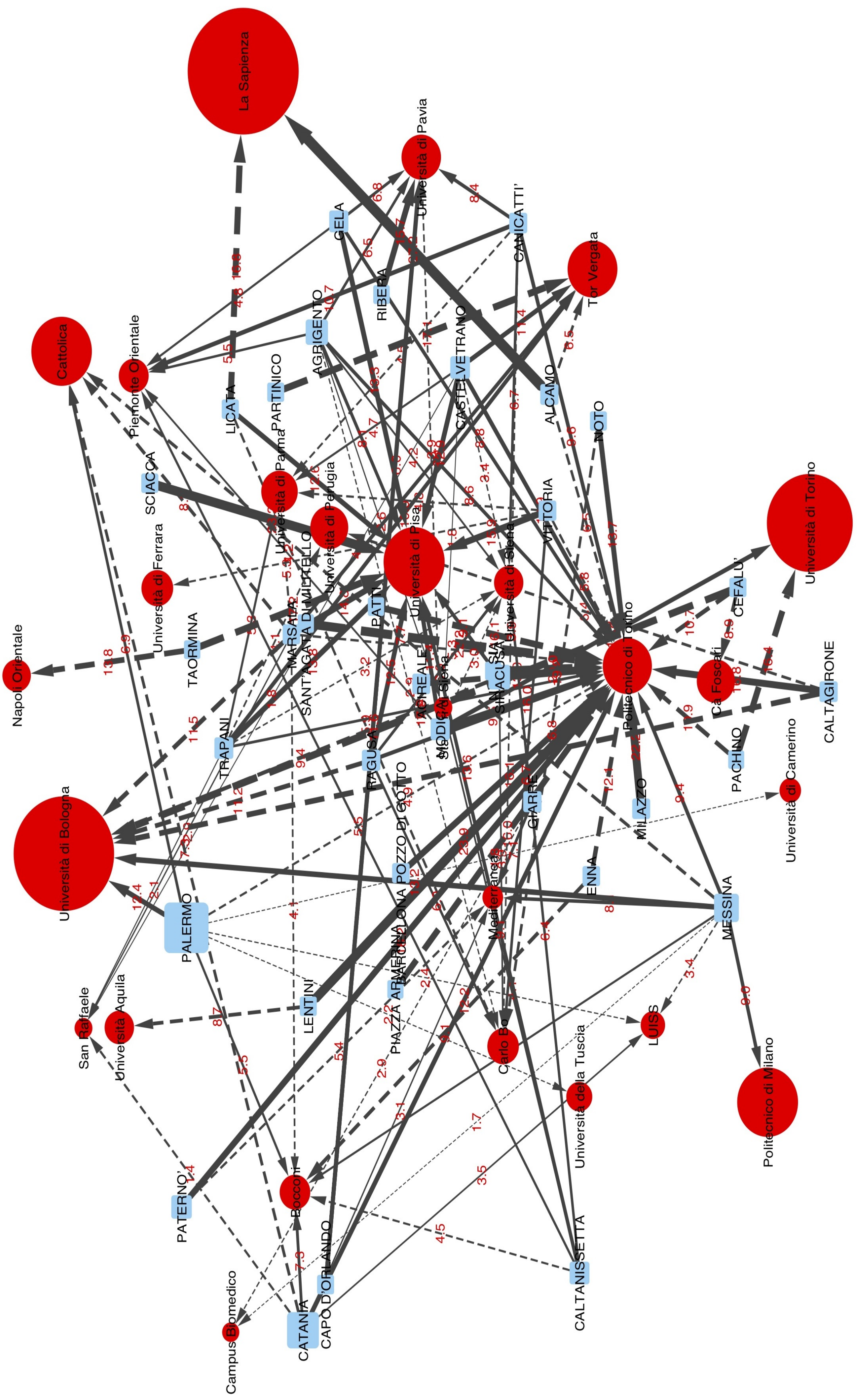


Figure 1.9: Bipartite validated network for Sicilian outgoing students, year 2014

Here, we use the Matthews correlation coefficient (MCC) [117] to evaluate the similarity between two networks, since it properly takes into account the heterogeneity of both clusters and universities, and it is symmetric with respect to the exchange of the networks under comparison. The Matthews correlation coefficients for male and female Bonferroni (FDR) networks is on average 0.47 (0.46), and stable over time, which indicates that the relative frequency of gender specific patterns is rather stable throughout the investigated time window.

In summary, these results support the hypothesis that the impact of austerity policies, introduced in 2012 and 2013 in Italy, favoured a change of migration patterns in 2014, by determining the creation of new patterns from Sicily to universities in the North (increasing trend of the number of preferential patterns), but it did not affect the imbalance of gender specific preferences.

Within the framework of oriented networks, the degree of a university indicates the "popularity" of the university in terms of number of Sicilian clusters displaying a preferential pattern pointing to it. Accordingly, we decided to calculate the degree ranking of universities as reported in Table 1.5. The order of universities in the Table is based on the ranking of the median degree rank (reported in parenthesis in the Table) across all of the networks reported in the Table. Universities that showed a degree equal to zero in at least one network are not displayed in the Table.

Looking at Table 1.5 is clear that University of Pisa, on average, is the big favourite, but decomposing its degree by gender, it's clear that University of Siena is the most chosen university for female students overtime. In light of these results, we focused our case-study network analysis on University of Siena.

Table 1.5: Degree ranking of universities by cohort in the FDR networks, for female, male, and overall. In parenthesis the conditional ranking (see text for details on the selection of listed universities).

		2008			2011			2014		
Ranking	Universities	Female	Male	All	Female	Male	All	Female	Male	All
1	University of Pisa	9 (2)	9 (1)	13 (1)	13 (2)	7 (3)	14 (3)	9 (2)	13 (2)	13 (2)
2	University of Siena	10 (1)	2 (8)	9 (3)	17 (1)	15 (1)	17 (1)	3 (11)	8 (3)	7 (4)
3	Polytechnic University of Turin	1 (15)	3 (7)	7 (7)	7 (4)	11 (2)	15 (2)	10 (1)	23 (1)	28 (1)
4	Bocconi	3 (8)	5 (3)	9 (5)	6 (5)	4 (5)	8 (5)	5 (3)	4 (5)	6 (7)
5	University of Pavia	4 (7)	4 (5)	6 (8)	4 (8)	4 (6)	5 (8)	5 (4)	5 (4)	7 (5)
6	University of Parma	9 (3)	6 (2)	9 (4)	8 (3)	6 (4)	9 (4)	2 (12)	1 (16)	5 (8)
7	Mediterranea	6 (5)	5 (4)	7 (6)	3 (9)	3 (9)	5 (7)	4 (7)	3 (9)	4 (11)
8	University of Bologna	1 (10)	1 (14)	2 (12)	2 (10)	1 (12)	4 (9)	4 (5)	1 (13)	7 (3)
9	LUISS	7 (4)	2 (12)	9 (2)	1 (18)	2 (10)	2 (15)	1 (20)	3 (10)	3 (14)

Figures from 1.10 to 1.12 show the statistically validated subnetworks that include all of the Sicilian clusters pointing to University of Siena. Regardless the gender, in 2008 the Sicilian cluster involved in the flow to Siena are 9, where the clusters of Lercara Friddi, Ribera, and Modica have the greater flow. In 2011, clusters involved in the choice of Siena University increase rapidly–from 9 clusters observed in 2008, they become 17 in 2011 (for all students). Such an effect, however, slowed down in 2014, where the clusters involved for all the students are 7. Looking at the network links according to gender, the number of clusters from where female students have moved to the university of Siena increase from 10 in 2008 to 17 in 2011, and it has been the most favorite destination for female in 2011, but this growth slowed down in 2014 for both gender. As it can be seen in Table 1.5, Polytechnic University of Turin moves over from the 7^{th} position in 2008 to the 1^{st} position in 2014.

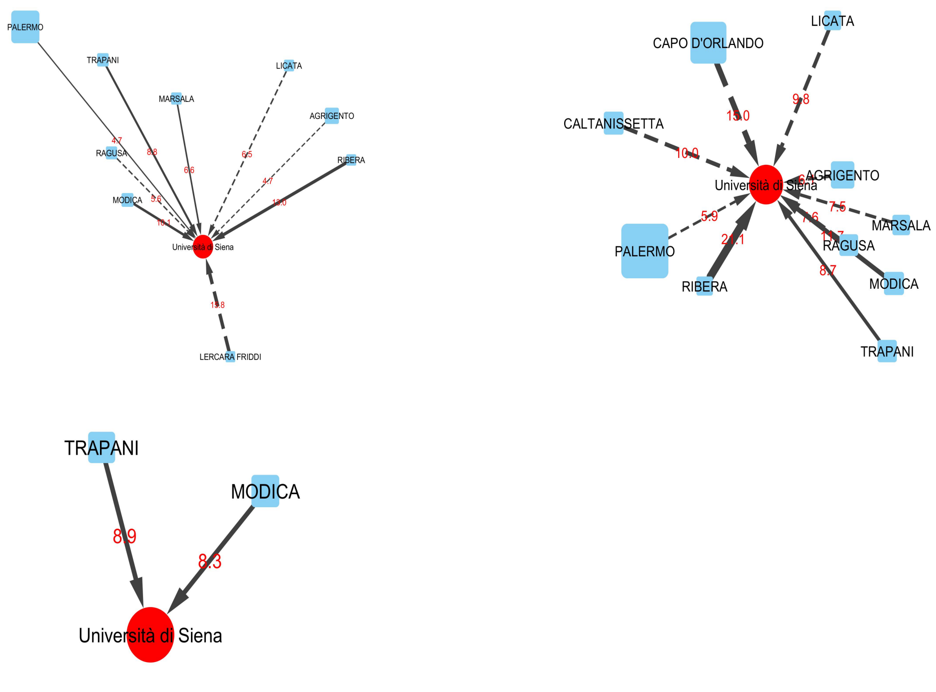


Figure 1.10: Bipartite validated network for University of Siena, focus on Sicilian outgoing students (left top), female students (top right), and male students (bottom left), year 2008

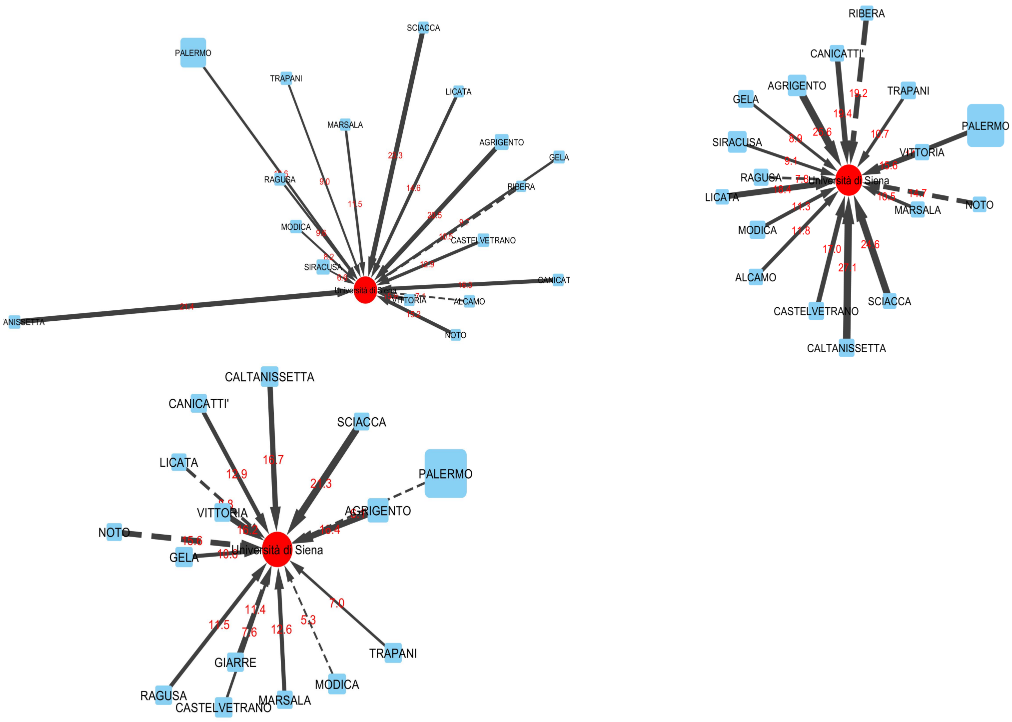


Figure 1.11: Bipartite validated network for University of Siena, focus on Sicilian outgoing students (left top), female students (top right), and male students (bottom left), year 2011

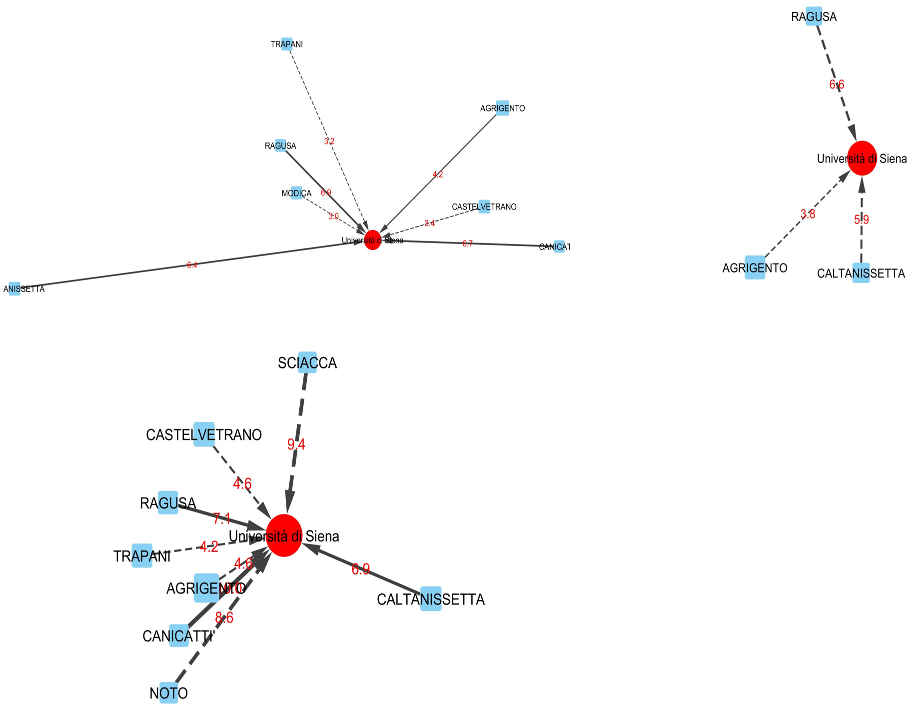


Figure 1.12: Bipartite validated network for University of Siena, focus on Sicilian outgoing students (left top), female students (top right), and male students (bottom left), year 2014

Finally, it's worth to note that contiguous clusters might display similar (statistically significant) patterns of mobility, likely due to the presence of exogenous factors influencing specific mobility routes. Indeed, the considered null hypothesis, which is modeled through the hypergeometric distribution, allows one to indirectly take into account the presence of factors influencing the nodes–*e.g.*, the labour market of the cluster of origin, and the attractiveness of specific universities–through the parameters N_A (the number of students coming from municipality A) and N_B (the number of students enrolled in university B). However, it is possible that some cluster-university link is favored (or disfavored) by factors affecting the connection itself, instead of the connected nodes, factors such as the transport system. For instance, let's consider the case of the contiguous clusters of Marsala and Trapani. People from these clusters live close to the same airport, namely, Birgi airport. Looking at the connections of these clusters in the Statistically Validated Bipartite Network (SVBN) obtained by using the FDR correction, it turns out Trapani is linked to $n_T = 6$ universities, Marsala to $n_M = 5$, and $n_{TM} = 4$ destinations are in common, in 2008. Considering that, overall, there are 78 universities in the system, one can apply the original method developed by Tumminello *et al.* [108] to the (bipartite) SVBN and calculate the probability to observe a value of co-occurrence at least as extreme as the observed one ($n_{TM} = 4$) is 0.00005, according to the hypergeometric distribution. Similar results are obtained, by looking at the cohorts of 2011 and 2014. Specifically $n_T = 5$, $n_M = 4$ and $n_{TM} = 3$, in 2011, with an associated p-value of 0.0005, whereas, $n_T = 7$, $n_M = 6$ and $n_{TM} = 3$, in 2014, determining a p-value of 0.007[5]. This results suggest the presence of common factors influencing the mobility of students from the two clusters. However, we believe that such an analysis deserves more attention and it will be considered for future research.

1.5 Conclusions

The descriptive statistics of mobility data reveals an increase of the movers from Sicily towards the central and northern Italian universities over time. In fact, it has been observed an increase of both the rate of outgoing students from Sicily and the number of clusters with an outgoing rate of at least 30%, which, initially (2008's cohort), were mainly located in the western clusters of Sicily, whereas, at the end (2014's cohort), they were uniformly distributed over the whole Center and South of Sicily, from the western clusters

[5] The total number of considered universities in the system is 80 in 2014.

to the eastern ones.

The network analysis of preferential patterns of mobility has shown that, among universities of central Italy, Pisa and Siena are (on average) the most attractive ones for students coming from western and southern Sicily, and that Polytechnic University of Turin is the favorite northern university for the last cohort of students (2014).

In addition, the network analysis provides some elements to support the existence of migratory chains, over time, by means of the increase of the number of:

- new links, which become significant in the networks of the most recent cohorts and were not in the previous ones, and their thickness which also increased in the networks of the last cohort;
- the new nodes (and their size) that appeared in the networks of the last cohort.

Specifically, it has been observed an increase of the number of clusters with movers and the number of off-Sicily universities chosen by the movers (preferential patterns). At the same time, the number of both movers from some clusters and students towards a given off-Sicily universities (the thickness of their link–outgoing flows), as well as the number of movers from Sicily who enrolled in off-Sicily universities increased.

By investigating the similarity between the networks according to the sex of movers (through the MCC), we observe a stability of similarity over time. This result implies that, even if mobility patterns vary over time, such a variation occurs by keeping almost constant the similarity between gender specific patterns. The case of Siena also suggests that gender specific patterns may both vary over time in a non-monotonous way (pick in 2011).The proportion of preferential links that are preserved over time slightly exceeds 40%. This result indicates that some *old* patterns of mobility are destroyed, while new ones appear, especially in the networks of the last cohort as a possible drawback on the financial crisis, which affected differently the South and the North of the country, especially with respect to the labour market, and favoured the formation of new mobility patterns from the South to the North of the Country. Finally, the observed 40% stable links, however, implies that the role played by chain migration in explaining mobility patterns of students is rather important.

Finally, it is worth to include an assessment of the strengths and limitations of the present study, in relation to the statistical methods used to highlight chain-migration effects on students' mobility. The null hypothesis considered in the chapter involves a probability distribution, the hypergeometric distribution, which is conditioned to both the total number of students moving from each cluster and the total number of students enrolling in each university. Therefore, implicitly, such a null hypothesis incorporates information about the specific attractiveness of universities, as well as cluster-specific factors that might influence the decision of students to enroll in a university out of the region. So, the considered null hypothesis appears to be appropriate to reveal chain-migration effects, since it allows one to untangle chain migration from factors mainly related to node-specific characteristics. Furthermore, the considered null hypothesis properly takes into account the heterogeneity of clusters, in terms of number of moving students, as well as the one of universities, in terms of number of enrolling students. However, factors associated with a specific cluster-university link, such as an airport that directly connects a cluster to a specific destination, are not incorporated in the null hypothesis, neither directly or indirectly. In the last paragraph of the discussion section, we have provided an example of the impact of such link-specific factors, by showing how they may determine a significant similarity between the patterns of mobility from two contiguous clusters. The example of Trapani and Marsala clusters suggests that, based on the present analysis, we cannot claim that chain-migration is the only factor determining the revealed preferential patterns of mobility, since other link-specific factors (*e.g.*, the transport network) might also come into play. A possibility to deal with such an issue would be to use the transport network to group together clusters with a similar connectivity, and perform the analysis of preferential patterns separately for each group of clusters. Another possibility might be to use the Wallenius non-central hypergeometric distribution, by assigning a specific weight to groups of students, sorted according to the groups of clusters of origin and groups of destinations with similar connectivity in the transport network [91]. Such an analysis left for future research.

Chapter 2

The good old ideas: the concept of chain migration to explain Italian student mobility

Abstract

The neoclassical macroeconomic approach to migration postulates that the driving force behind migration is the labour market. On the other and, exponents of the new economics of labour migration argue that the decision to move is not made at an individual level. Rather, it is a family decision, the outcome of a bargaining process aimed at minimising family's risks. Considering migration as the result of a decision taken within a social network of relatives and friends helps us to explain the so called chain migration. In this chapter, we pay attention to a particular migratory chain: the migratory chain of university students. This chapter introduces a statistical technique to "classify" migratory chains starting from enrollment data of students who obtained a high school diploma in Sicily, Sardinia or Apulia, and enrolled at the college in some centre-north region in the period from 2008 to 2017. Results show that many of the origin-destination patterns are enhanced by chain migration effects. Furthermore, a "medium-large city to large city" and a "medium-small city to medium-small city" patterns are observed for students from Sardinia and Sicily. Indeed, students from Palermo, Catania, Messina, and Cagliari move to the two main cities of the country–Milan and Rome–while students from small cities prefer cities of a size similar to their home town.

2.1 Introduction

The neoclassical macroeconomic approach to migration postulates that the driving force behind migration is the labour market. Suppose there are different conditions, in terms of real wages (or even simply in the expectations of a future emergence of a wage gap) between two regions. In that case, the labour force of the disadvantaged region would move towards the the region with a labour market that guarantees higher remuneration, thus determining a progressive readjustment in both the level of employment and the wages in the two territories (the increase/reduction of labour supply would determine a reduction/increase of salary in the originally advantaged/disadvantaged place) until an equilibrium is reached in which migration flows cancel each other out [107, 53]. In this view, migration is only a temporary phenomenon that should stop due to the economic development of disadvantaged areas.

From a microeconomic point of view, the decision to migrate is the result of a rational process in which the costs related to the decision to emigrate are compared with the benefits that derive from it. If the latter exceeds the former, the individual decides to move towards the destination that allows him to maximise her expected utility throughout life (see, for example, the famous work by Borjas [13])[1].

According to the scheme proposed by Lee [64], the decision to leave is therefore influenced by two main forces: the so-called push factors consisting of everything that pushes people to leave their place of origin (poor employment opportunities and insufficient income prospects), and the so-called pull factors (*i.e.* everything that attracts migrants towards a particular destination).

The emphasis placed by individual-level analysis on the expected income associated with migration is the main object of criticism by the exponents of the new economics of labour migration [103, 63, 61]. According to them, the existence of a wage gap between two areas is neither a necessary nor a sufficient condition for the occurrence of a migratory movement, and the decision to move is not made at an individual level. Still, it is the result of a real process of rational family choice aimed at minimising the risks for the family itself. Through the emigration of one or more of its members, a family can have si-

[1] An excellent review of the main migration theories can be found in Massey *et al.* [76] and more recently in Hager-Zanker [52]

multaneous access to other labour markets and social security systems, thus implementing a diversification strategy.

Whatever the motivation that drives individuals to move (income, access to new markets), the above theories do not allow us to understand the reasons why transnational movements tend to perpetuate in space and time when migratory choice has finished producing its effects [76].

In any case, considering migration as the result of a decision taken within a social network of relatives and friends helps us to explain such persistence in terms of cumulative causation migration or, more commonly, chain migration.

In demography, the term chain migration refers to the process through which migrants from one geographic area move to areas where there are migrants from their geographic area who moved previously [81, 70]. This process causes more people–from the same location–to move to a particular destination. This migration occurs as a result of the positive and encouraging information about the destination that early migrants give to residents in their area of origin. This encourages men and women from the same area to move in the hope of finding better economic opportunities and a better quality of life, turning to first migrants that can help them (hospitality, information on housing and work, economic and emotional support, *etc.*). It is also plausible that first emigrants encourage family members and fellows to move and settle near them to partially recreate a familiar environment in the new place of emigration. Thus, the existence of a social network in the place of destination reduces the costs of migrating and, at the same time, opens up to the possibility of a better integration.

The existence of such processes has been highlighted by several studies conducted in different destination countries and considering various ethnic groups. To cite some examples, MacDonald and MacDonald (1964) [70] analyse Italian migration to the USA, MacDonald and MacDonald (1970) [71] focus on Italians in Australia, Palloni *et al.* [85] consider Mexican migration to the USA, Liu *et al.* [67] focus on Filipinos in the North American country while Böcker [12] studies the chain migration of the Turkish community in the Netherlands or Eurenius [41] who emphasises the role of chains in Swedish emigration to the United States in the late nineteenth century. Banejeree [6] highlights the role of social network in the internal movement between countryside and city in India. Finally, Reyneri (1998) [94] investigates the case of immigrants employed in the informal sector in Italy, pointing out the fundamental role of chains in determining the attractiveness of the

Peninsula. See also the volume edited by Reyneri himself in 1979 [93] on the topic of the effects of the migratory chain in the country of origin and in the country of destination.

Obviously, the prerequisite for a migratory chain to be realised is the existence of social capital within a community, that is, the set of relational resources that are formed within it and that help its social development. Within this general concept, the migrant's network is in fact defined by Massey as: "sets of interpersonal ties that link migrants, former migrants and non-migrants in origin and destination areas by ties of kinship, friendship and shared community origin" ([73], p.7).

In this chapter, we pay attention to a particular migratory chain: the migratory chain of university students. It certainly has quite different characteristics from those of the migratory chains usually studied in demography, for several reasons. The first is certainly characterised by a reduced temporal space and by a "temporary" perspective, most of the time linked to the period of study, in fact in the case of students we prefer to use the term mobility, precisely because of its temporary peculiarity. Person and Rosenbaum [89] speak of chain enrollment to identify the phenomenon and highlight its difference from the more traditional concept of chain migration. Bearing in mind, therefore, the diversity of the two phenomena, in this chapter we will use the term student migration chains. Studies on this topic are very few in the international literature and mostly based on qualitative methods [89, 88, 19, 47].

In the case of Pérez and McDonough [88] the analysis is conducted through semi-structured interviews collected in a non-representative sample of the student population in which they show the influence of the presence of other members of their community in the choice of college by Latinos in the USA. Similarly, Brooks and Waters' paper [19] shows, through the same technique, the importance of one's social/family networks in determining the choice of studying in the UK by 83 international students. The work of Person and Rosenbaum [89] combines the use of semi-structured interviews with the use of a survey conducted on students from 14 colleges, concluding that a network of friends/relationships in the destination university increases the probability of enrollment in that specific location by a freshman especially if of South American origin. An interesting aspect that clearly emerges from the work of Person and Rosenbaum [89] is that the information possessed about the destination plays a fundamental role in determining the choice of college for all students, but that students from the Latino community are characterised by a lack of information to their disadvantage compared to non-Latinos and that the use of their social network is the main tool to try to fill it.

Genova *et al.* [47] using data on the entire student population of three freshman cohorts from Sicily show through a network analysis that 40% of the link movements identified between origin and destination areas can be explained by the existence of student migration chains.

There are, however, no techniques in the literature for quantifying this phenomenon. To the best of our knowledge, there is no measurement of this phenomenon that is immediately usable to identify possible routes in which the student migration chain could play an important role. This chapter aims to propose a statistical technique to "classify" migratory chains.

The research, which comes closest to the present one, is a study limited to Sardinia [97] whose primary objective, however, was to understand the main destinations of Sardinian students without offering a key to interpret the phenomenon. The current contribution therefore seeks both to integrate the analysis by considering other regions of southern Italy (Sicily and Apulia) and to read the routes observed in the light of chain migration theory. It was decided to focus exclusively on the movements from the southern regions towards the centre-north since previous studies have shown that mobility goes almost entirely in that direction (see *e.g.* Attanasio and Enea [3] and Ruiu *et al.* [96]).

The chapter will be articulated as follows: in the next section, the data used will be presented and more details on the objectives of the analysis will be provided; then, the proposed method to measure the student migration chain will be described; finally, the obtained results will be discussed.

2.2 Data and objectives

2.2.1 Objectives

The objective of this study is to investigate the presence of migratory chains in the mobility of students coming from Sicily, Sardinia and Apulia who enroll for the bachelor at universities in the North and Center of Italy, in the decade 2008-2017. More precisely, the analysis aims to measure three effects: the "attraction effect" exerted by some destination areas of the country that play a polarising role; the "expulsion effect" from other areas of the country that experience an increasing mobility of their students; and, finally, a "residual effect" that can be partially attributed to family or friendship ties. In general, migration theories use the paradigm of the "differential" between country of origin and country of destination (see, for example, Borjas 1989 [13]) as a synthesis to interpret migration processes. This differential is certainly a multidimensional concept. It is well known, in fact, that the decision to migrate depends on the difference between the expected income over the life cycle in the country of residence and the one in the country of (possible) destination. However, also aspects of non-monetary nature play a significant role in forming the decision to emigrate, such as the different political, religious and social conditions at destination [77, 65]. Part of this component, as previously mentioned, could be attributed to the presence of friends or relatives already present in the country of destination as source of information, which is generally referred to as migratory chain. In (the rather simplified) Figure 2.1, M_i and A_j are the students who leave the *i-th* area of origin ($AreaOri$) and "land" at the *j-th* destination ($Dest$), respectively, without a reason that can be traced back to the migratory chain, while C_{ij} students are the ones who leave the origin i for the destination j as result of migratory chain effects.

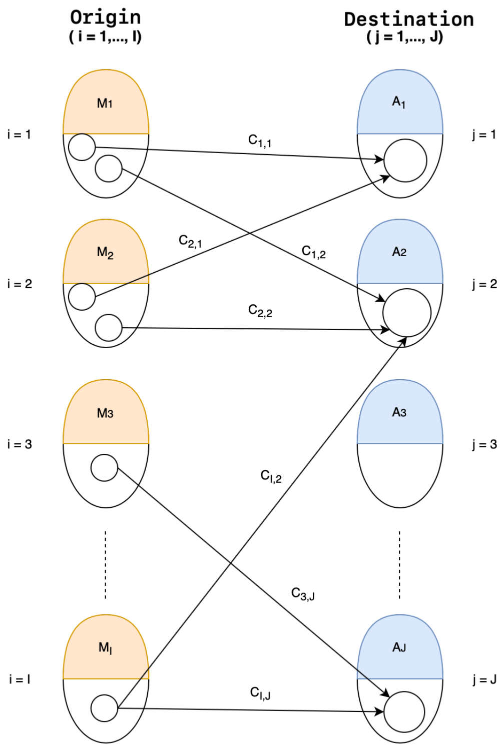


Figure 2.1: Simplified scheme of students flows and chain migration flows C_{ij}

Our work aims to identify and eventually measure this component through a longitudinal analysis, based on the comparison of C_{ij} migration flows in two time intervals, 2008-2012 and 2013-2017. Specifically, the research questions we pose are: does chain migration influence student mobility? If so, what are the most significant patterns? Is there a link between student mobility and labour migration in the past? Are there similarities between the mobility patterns starting from different areas of origin?

2.2.2 Data

The data used in this chapter refers to the MOBYSU.IT database [79] mentioned in section 1.2. In accordance with the objectives of this work, the analysis focused on students who obtained a high school diploma from a secondary school in Sicily, Sardinia and Apulia, and eventually enrolled in a university located in Piedmont, Lombardy, Emilia-Romagna, Veneto, Tuscany or Lazio from 2008 to 2017. Students who enrolled in an online degree course or in a degree course in the health area were excluded from the analysis. The former were excluded because they do not actually move, while the latter were excluded because their mobility is strongly conditioned by the rules of admission to those degree courses. Given the small number of movers for each cohort[2], we decided to aggregate the cohorts into two five-year periods, the first one from 2008 to 2012 and the second one from 2013 to 2017.

As reported in the introduction, this analysis was conducted on 3 Italian regions: Sicily, Sardinia and Apulia, which are characterised by a high rate of students who decide to study outside the region. The first two regions have in common their insularity: this implies that out-of-region mobility (with the exclusion of the Strait of Messina area) is non-commuting–both daily and weekly–while Apulian out-of-region mobility also includes weekly commuting to Campania and Lazio. Moreover, in the case of Apulia, some mobility routes may be favored by the existence of historical transportation routes, such as the Adriatic railway.

2.3 Methods

The method proposed to pursue the objectives of this work is developed in the following steps: *i)* determining the areas of origin (*AreeOri*) as areas in which communities of students can directly communicate and eventually trigger a mechanism of chain migration; *ii)* constructing multidimensional origin-destination matrices (by disciplinary field and gender); *iii)* building origin-destination matrices based on the assumption of independence;

[2] Less than one hundred for some origin-destination patterns.

iv) performing the complete linkage cluster analysis (CLINK) on the residuals origin-destination matrices in points ii) and iii). Each one of these steps will be described in the following sections.

2.3.1 Determination of areas of origin

The construction of areas of origin (*AreaOri*) assumes that student communities in one *AreaOri* directly communicate within the *AreaOri* and do not directly communicate with other *AreaOri*s. The *AreaOri*s are formed by several municipalities gravitating around a hub municipality, which is definitely home to at least one secondary school. Moreover, we assume that the part of student mobility reflecting chain migration only depends on the area of origin, neglecting the mobility deriving from family and/or friendship ties in the destination region of historical type. Finally, the *AreaOri*s were constructed on the basis of students who enrolled in 2008/09 and were kept constant until 2017/18, to make the comparison possible.

In light of these assumptions, the *AreaOri* were formed independently for each considered region (Sicily, Sardinia and Apulia) as follows:

1. We construct an origin-destination matrix $M(i,j)$, in which each row contains the municipalities of residence and each column contains the municipalities of secondary school (Table 2.1).

2. From $M(i,j)$ we choose J destination municipalities home to at least one high school with at least 200 students who enrolled in a university afterwards (see total columns). These municipalities are the hubs used as starting point for the determination of the areas of origin (*AreeOri*). (*e.g.* Table 2.1: Alcamo, Palermo, Trapani and Sciacca as they have a number of students greater than 200).

3. If the municipality of residence i is not a hub, we identify j^*, with $i \neq j^*$, as the municipality with the highest number of graduated people from i.

   a) j^* is hub $\Rightarrow$ the *i-th* municipality is attributed to j^* (*e.g.* Table 2.1: since the maximum value of the cells $(Carini, j)$ is 50, the hub of Carini is Palermo)

   b) j^* is not a hub $\Rightarrow$ the students of the *i-th* municipality are assigned to the *j-th* hub municipality that is closest in terms of physical distance (*e.g.*, with reference to Table 2.1, the students of the cells $(Capaci, j)$ (j=1,...,n) are assigned to the hub of Palermo, which is the geographically closest hub)

4. If the municipality of residence i is a hub, students are assigned to hub j, with $i = j$ (*e.g.*, in Table 2.1, the students in cells $(Palermo, j)$ (j=1,...,n) are assigned to the Palermo hub.

5. The procedure up to step 4 assigned students residing in k non-hub municipalities (C_1, C_2,..., C_K) to hub municipalities. The source area of the generic hub j is:

$$AreaOri(hub_j) = hub_j \cup \left\{ \bigcup_{k=1}^{K} C_k \right\} \tag{2.1}$$

For example, it can be deduced from Table 2.1 that: the Palermo *AreaOri* is given by Palermo, Carini, Capaci and Isola delle Femmine; instead, Buseto Palizzolo is assigned to the nearest hub, *i.e.*, Trapani.

Table 2.1: Example of an origin-destination matrix $M(i, j)$ for the construction of origin areas ($AreeOri$)

Origin	Destination							
	Alcamo	Carini	...	Palermo	...	Trapani	...	Sciacca
Carini	0	30	...	50	...	2	...	10
Palermo	2	10	...	400	...	0	...	0
Capaci	3	50	...	20	...	0	...	5
Isola delle Femmine	2	30	...	40	...	2	...	0
Balestrate	70	3	...	10	...	50	...	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Buseto Palizzolo	10	50	...	30	...	30	...	5
Total	220	100	...	1000	...	300	...	320

Tables 2.2 through 2.7 show the obtained origin-destination matrices for Sicily, Sardinia and Apulia, respectively. The tables consider the regions of destination as the "destination" point. Column "OTHER" includes all the regions not explicitly listed in the table, though the heterogeneity within this group can be high–from a few units of the flows connecting the three regions of origin (Sicily, Sardinia and Apulia) to a few hundreds for other regions such as Umbria and Friuli. Figures 2.2 and 2.3, 2.12 and 2.13, 2.22 and 2.23 show the percentage of movers from each *AreaOri*, for each one of the three considered regions.

Finally, Figures 2.4 through 2.11, 2.14 through 2.21, and 2.24 through 2.31 graphically synthesise the information reported in both the origin-destination matrices and the Tree Maps. In particular, the area of each monochromatic rectangle is proportional to the number of students moving from each of the *AreaOri*, while the areas drawn inside each of these rectangles are proportional to the number of students moving to the region identified by the associated label.

Sicily

For Sicily it is possible to notice that the starting areas with the greatest outflow are the same in the two periods examined (Trapani, Castelvetrano, Ragusa, Siracusa, Agrigento) and that the phenomenon is accentuating, probably thanks to the economic recovery, reaching peaks of more than 50% of students moving out from the areas in the period 2013-17. On the other hand, the main Sicilian cities, namely, Palermo, Catania and Messina, each one site of a university, record the lowest relative share of outgoing students. As far as destinations are concerned, the Tree Maps suggest two distinct patterns between the large cities and the Sicilian provinces. Students from Catania, Palermo and Messina tend more to head towards the universities of big cities, such as, Milan, Rome, and Turin[3], while students from the other Sicilian *AreaOri*s appear to prefer Tuscany and Emilia-Romagna.

[3] especially male students who are strongly attracted by the Polytechnic of Turin

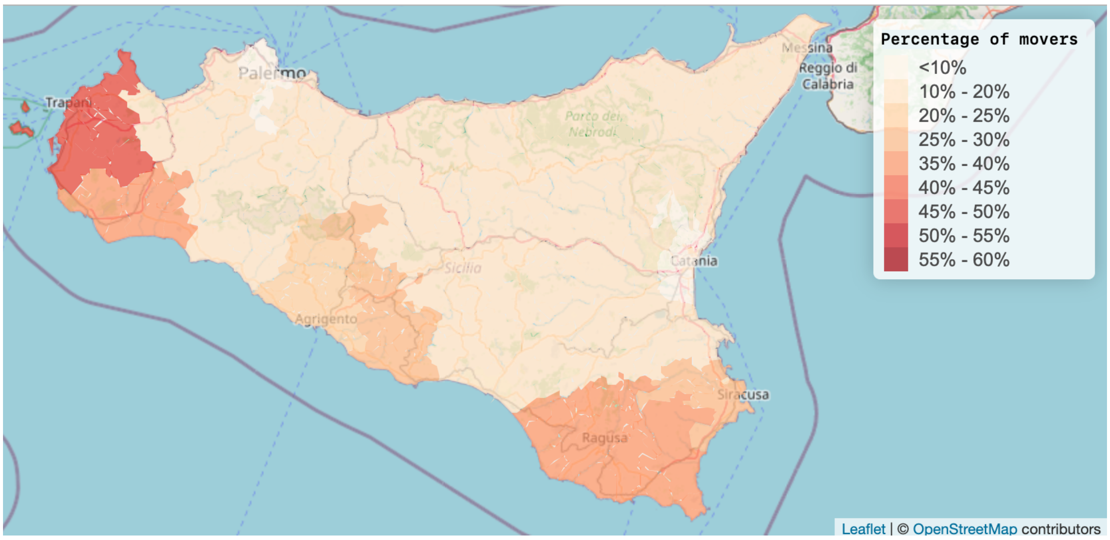


Figure 2.2: Sicily, percentage of student movers by area of origin, 2008-2012.

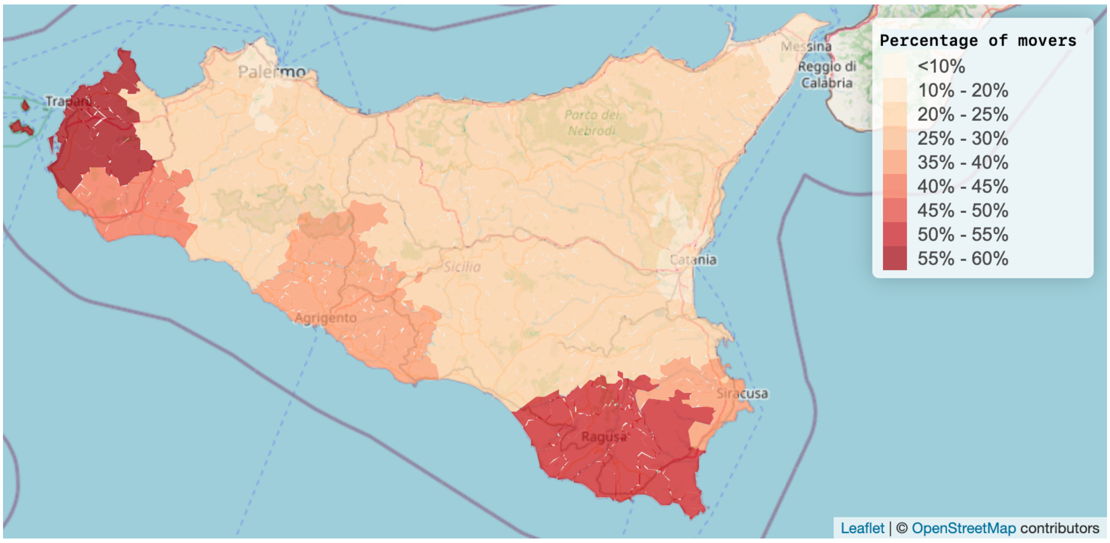


Figure 2.3: Sicily, percentage of student movers by area of origin, 2013-2017.

Table 2.2: Distribution of students enrolled in a bachelor degree residing in Sicily by cohort, area of origin and region of destination (row percentage values).

Period	*AreeOri*	Destinations								Total
		EMILIA-ROMAGNA %	LAZIO %	LOMBARDY %	PIEDMONT %	TUSCANY %	VENETO %	OTHER %	SICILY %	
2008-2012	AGR	4.5	1.6	5.3	2	5.5	0.8	2.4	77.9	**4338**
	CAN	6.5	2.9	6.4	2.9	5.8	0.8	3.6	71.2	**3696**
	CAS	11.1	4.1	8.3	3.7	11.4	1.9	6	53.4	**2782**
	CAT	0.6	1.6	2.2	0.5	0.6	0.3	1.2	92.9	**12517**
	MES	1.2	2.1	3.2	1.1	1.1	0.7	3.1	87.6	**8264**
	PAL	0.9	1.2	2.3	0.4	0.8	0.3	1	93	**15544**
	RAG	6.9	4.7	7.9	6.3	9.8	1	4.4	59	**5731**
	SIR	4.5	4.3	5.3	3.4	4.3	0.8	2.8	74.6	**4743**
	TRA	12.4	5.3	6.2	3.7	15.4	2	6.6	48.3	**4313**
	VIT	8.2	3.7	6.1	2.2	9.9	1.4	5.5	63.1	**1831**
	Other *AreaOri*	1.8	1.8	3	1.6	2.2	0.5	2.2	86.9	**41228**
Total 2008-2012		**2.9**	**2.3**	**3.7**	**1.8**	**3.4**	**0.6**	**2.5**	**82.7**	**104987**
2013-2017	AGR	5.5	2.3	7.4	6.6	7.6	1.3	4.6	64.7	**3666**
	CAN	6.3	2.6	6.8	7.6	8.3	1.2	5.3	62	**3463**
	CAS	9.1	4	8.9	7.7	13.4	1.7	7.9	47.3	**2648**
	CAT	1.4	1.6	3.2	1.5	0.8	0.6	1.7	89.1	**11577**
	MES	2.2	2.6	4.3	2.7	1.3	0.9	3.6	82.4	**7326**
	PAL	1.8	2	2.8	1.2	0.9	0.6	1.2	89.5	**14590**
	RAG	8.3	4.4	8.8	11.2	12.1	1.3	5.8	48.1	**5144**
	SIR	7.2	3.5	7.1	9.4	5.9	1.2	5.3	60.5	**3956**
	TRA	16.1	5.4	7.5	9.2	12.1	2.8	9.6	37.3	**3918**
	VIT	10.7	3.3	7.2	10.2	11.1	1.5	6	49.9	**1657**
	Other *AreaOri*	3	1.9	3.9	3.5	3.2	1	2.8	80.5	**37964**
Total 2013-2017		**4.1**	**2.4**	**4.7**	**4.3**	**4.2**	**1**	**3.4**	**76**	**95909**

Table 2.3: Distribution of Sicilian students by period, area of origin and universities of destination. The percentages for destination universities are calculated with respect to the total number of students enrolled outside the region and the percentage of students enrolled in Sicily is with respect to the total number of Sicilian students enrolled in an Italian university.

Period	*AreeOri*	Universities outside the region												SICILY	Total
		UniPi %	PoliTo %	UniBo %	UniPr %	UniTo %	Cattolica %	UniSi %	La Sapienza %	Bocconi %	PoliMi %	Other %	Total Outside Region	%	
2008-2012	**AGR**	10.8	4.9	7.7	11	2.7	6.5	10.3	3.2	2.6	3.6	36.7	**957**	77.9	**4338**
	CAN	11.3	5.6	8.8	4.9	3.2	4	5.8	2.7	2.9	2.1	48.5	**1063**	71.2	**3696**
	CAS	15	5.4	9	11.1	1.8	3.1	4.9	3.5	2.9	1.9	41.6	**1297**	53.4	**2782**
	CAT	4.4	4.9	7.3	0.5	2.4	5.9	1.4	5.6	9.7	5.5	52.4	**885**	92.9	**12517**
	MES	4.3	6.8	7.5	0.9	1.8	5.5	1	5.8	5.8	7.5	53.2	**1021**	87.6	**8264**
	PAL	5.4	3.5	9.9	1.5	2.4	6.7	3.9	5.3	8.9	4.9	47.7	**1094**	93	**15544**
	RAG	13	10.3	8.7	4.8	4.7	4	6.9	4.7	3.1	4.7	35.1	**2351**	59	**5731**
	SIR	10.7	10.8	7	8	2.5	5.6	4.6	5.1	2.8	4.5	38.4	**1207**	74.6	**4743**
	TRA	19.9	5	11.3	9.7	2	2.2	6.3	3.9	2.5	1.7	35.4	**2228**	48.3	**4313**
	VIT	15.2	4.3	8.1	8.4	1.6	5.2	7.8	6.1	0.7	1.9	40.5	**676**	63.1	**1831**
	Other *AreaOri*	8.6	8.3	7.2	3.8	3.2	4.6	4.5	5.9	5.1	3.1	45.8	**5412**	86.9	**41228**
Total 2008-2012		**11**	**7.1**	**8.4**	**5.6**	**2.8**	**4.5**	**5.2**	**4.9**	**4.3**	**3.5**	**42.7**	**18191**	**82.7**	**104987**
2013-2017	**AGR**	10.8	8.9	6.4	6.3	4.8	3.9	3.9	1.8	2.9	2.8	47.6	**1294**	64.7	**3666**
	CAN	12.2	9	5.8	4.4	6.7	2.7	4.2	1.8	1.1	1.4	50.8	**1316**	62	**3463**
	CAS	16.5	9.1	7.2	6	3.2	2.9	3.3	1.6	1.9	2.4	45.9	**1395**	47.3	**2648**
	CAT	2.5	10.2	9.1	1	3.6	6.4	1.3	2.9	7.8	5.6	49.8	**1257**	89.1	**11577**
	MES	3.9	10.9	9.2	0.6	4	5.1	0.6	4.3	4.3	6.4	50.8	**1287**	82.4	**7326**
	PAL	4.7	7.3	13.7	1.1	3.2	6.6	1.7	2.1	7	4.4	48	**1538**	89.5	**14590**
	RAG	14.4	12	8.8	3.2	8.3	2.5	4.5	2.8	1.8	2.6	38.9	**2672**	48.1	**5144**
	SIR	8.3	17.3	9.7	3.5	6.1	4.2	3.1	3.4	2.6	2.4	39.4	**1563**	60.5	**3956**
	TRA	13	8.5	11.3	9.2	4.8	2.5	2.2	2.6	2.1	1.8	42	**2457**	37.3	**3918**
	VIT	15.5	11.8	9.2	4.7	7.6	4	3.1	2.5	0.6	2.9	38.1	**830**	49.9	**1657**
	Other *AreaOri*	8.5	11.7	8.3	3.3	5	3.5	3.2	3.1	2.9	3.5	47	**7386**	80.5	**37964**
Total 2013-2017		**9.9**	**10.9**	**8.9**	**4**	**5.2**	**3.7**	**3**	**2.8**	**3**	**3.2**	**45.3**	**22995**	**76**	**95909**

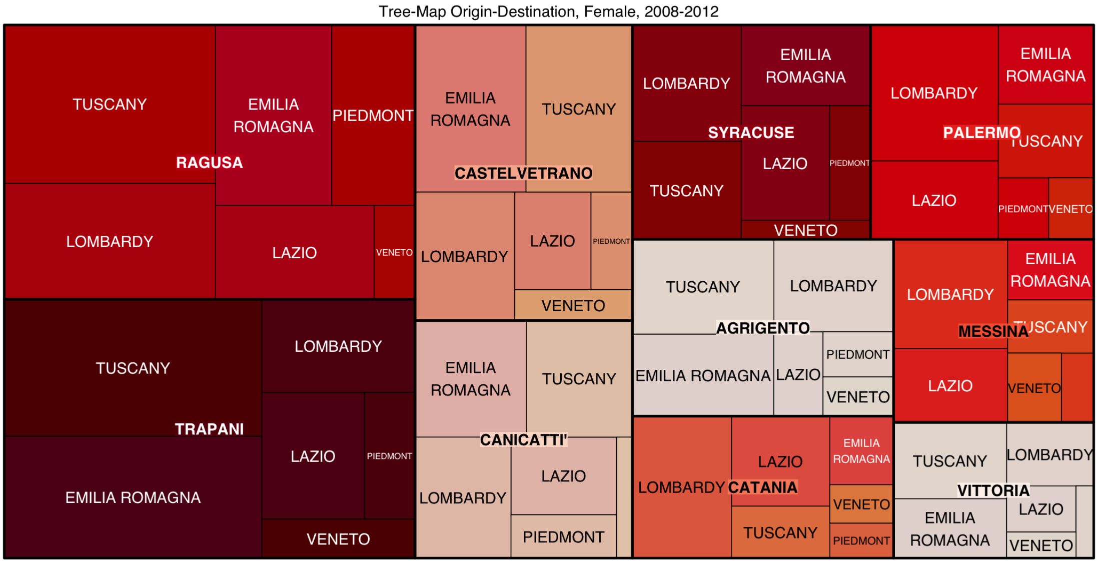


Figure 2.4: Tree-map of the origin-destination flows. The monochromatic rectangles are the selected source areas, while the inner rectangles are the destination regions. The size of the monochromatic rectangles is proportional to the population leaving the source area and the inner rectangles are proportional to the inflows into the destination region, period 2008-2012.

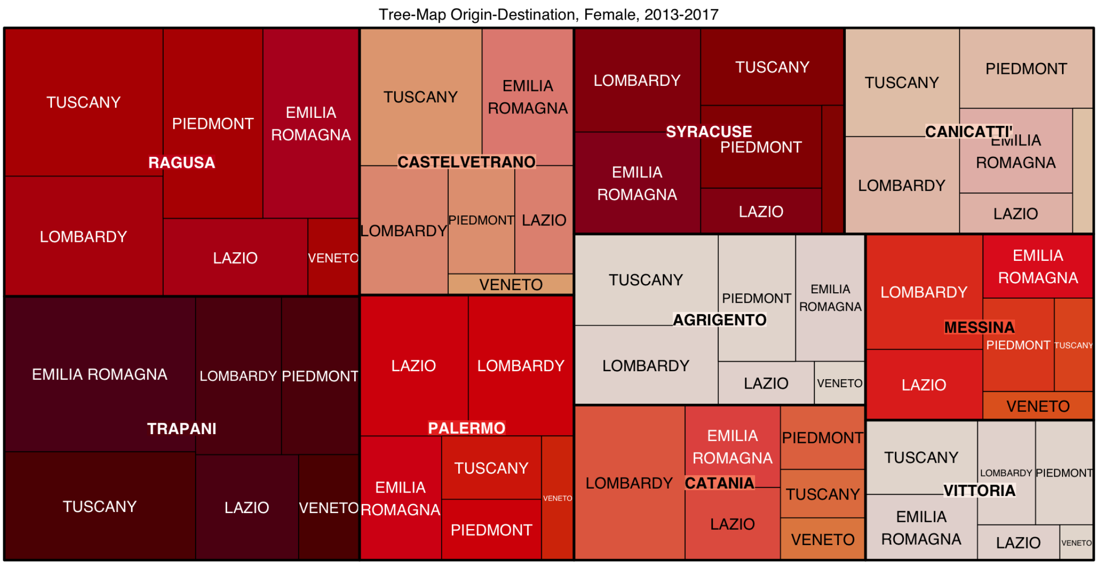


Figure 2.5: Tree-map of the origin-destination flows. The monochromatic rectangles are the selected source areas, while the inner rectangles are the destination regions. The size of the monochromatic rectangles is proportional to the population leaving the origin area and the inner rectangles are proportional to the inflows into the destination region, period2013-2017.

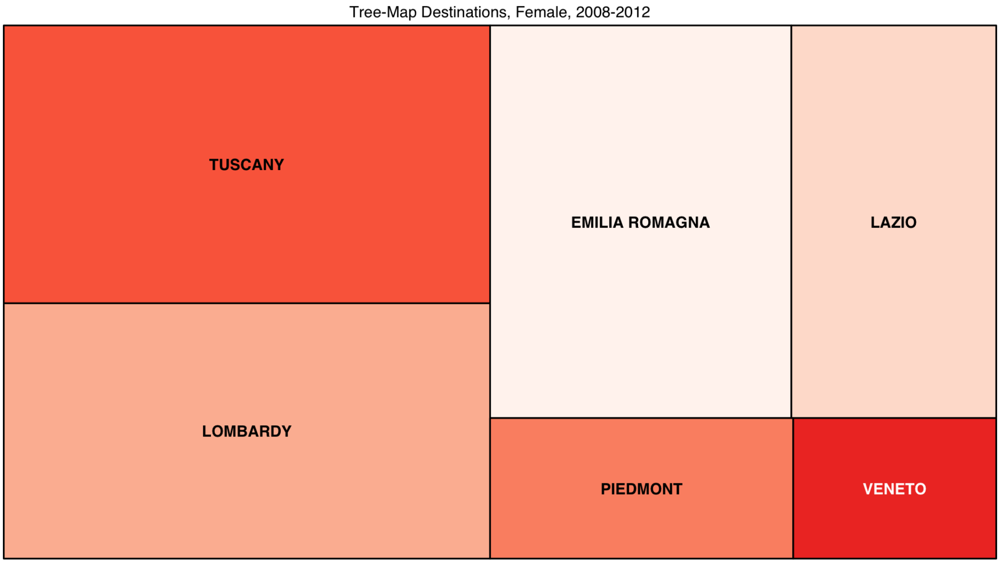


Figure 2.6: Tree-map of destination regions. The size of the rectangles is proportional to the inflows into the destination region, period 2008-2012.

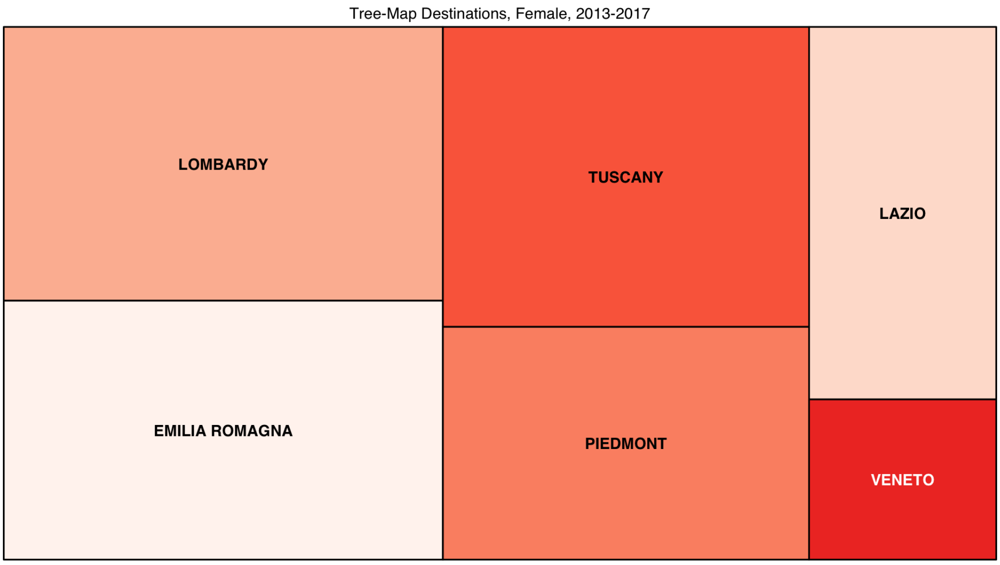


Figure 2.7: Tree-map of destination regions. The size of the rectangles is proportional to the inflows into the destination region, period 2013-2017.

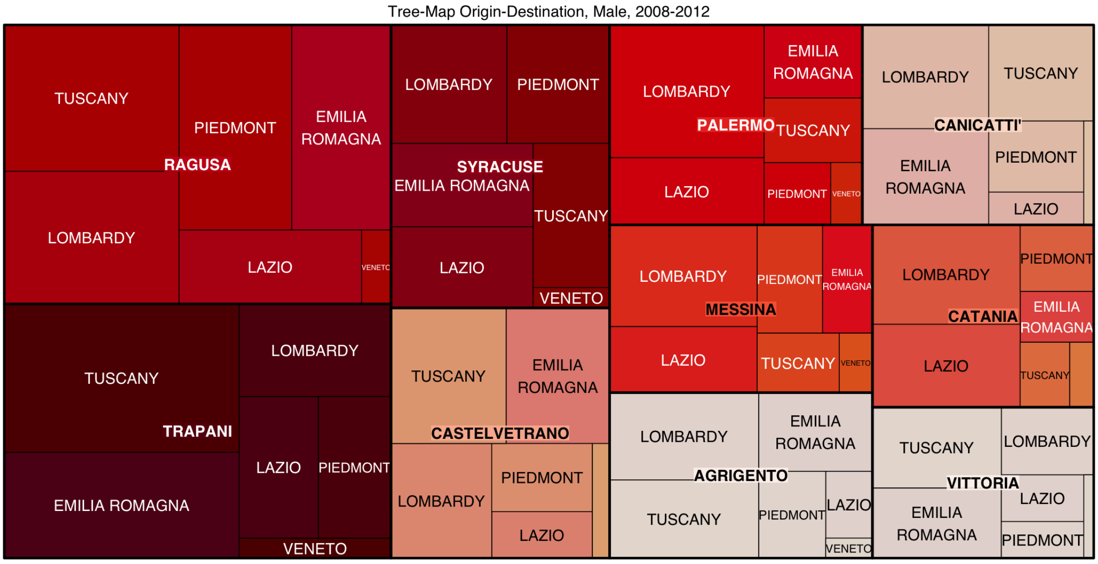


Figure 2.8: Tree-map of the origin-destination flows. The monochromatic rectangles are the selected source areas, while the inner rectangles are the destination regions. The size of the monochromatic rectangles is proportional to the population leaving the source area and the inner rectangles are proportional to the inflows into the destination region, period 2008-2012.

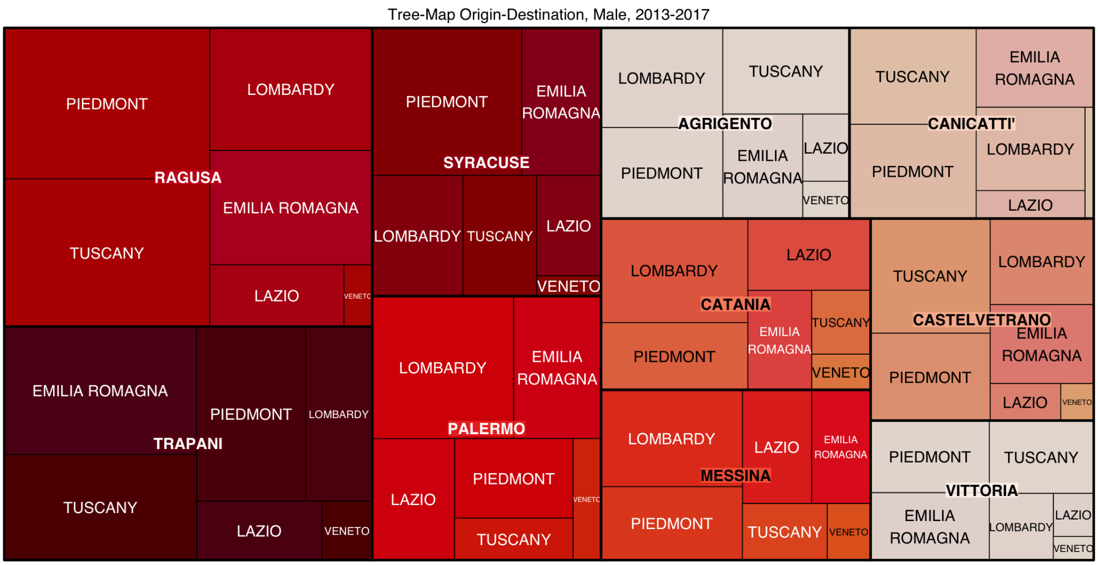


Figure 2.9: Tree-map of the origin-destination flows. The monochromatic rectangles are the selected source areas, while the inner rectangles are the destination regions. The size of the monochromatic rectangles is proportional to the population leaving the origin area and the inner rectangles are proportional to the inflows into the destination region, period2013-2017.

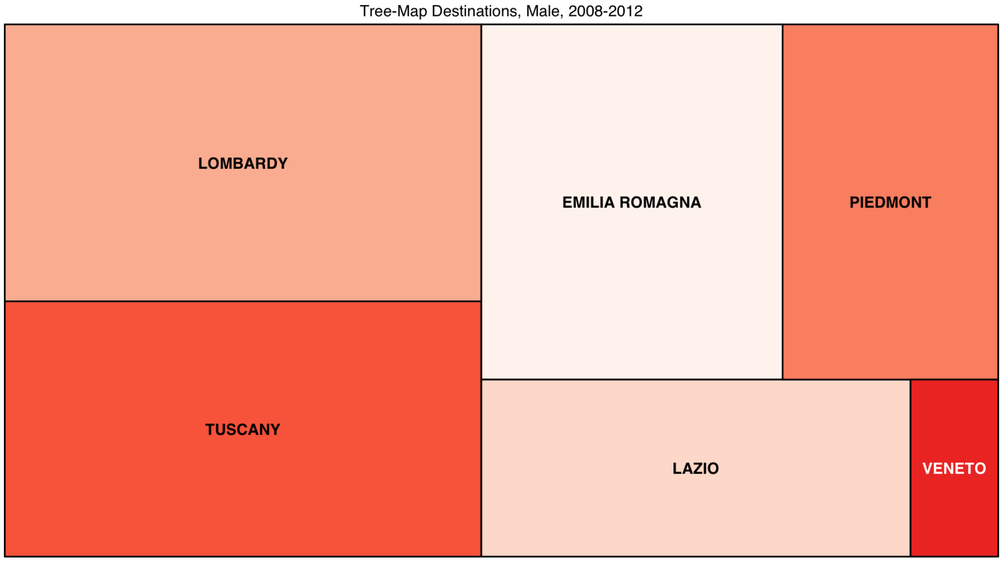


Figure 2.10: Tree-map of destination regions. The size of the rectangles is proportional to the inflows into the destination region, period 2008-2012.

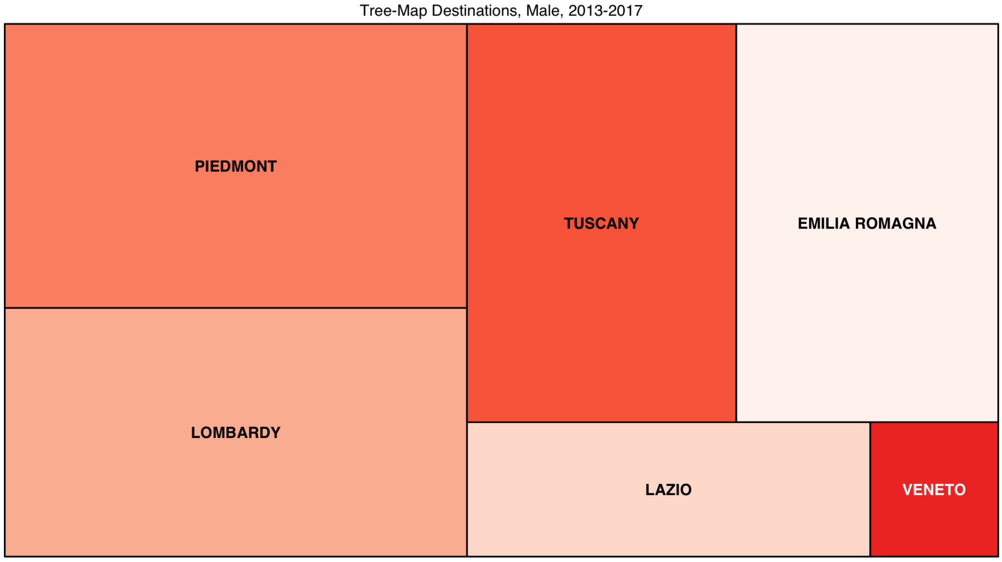


Figure 2.11: Tree-map of destination regions. The size of the rectangles is proportional to the inflows into the destination region, period 2013-2017.

Sardinia

Similarly to Sicily, the areas of Sardinia with the greatest outflow are the same both periods (Olbia-Tempio and Nuoro), although, in this case, the phenomenon seems to attenuate/accentuate for the Gallura/Nuorese territory over time. The percentages of movers are in general much lower than the corresponding values for Sicily. However, also in the Sardinian case, they are considerable: approximately 40% of the graduates of Olbia-Tempio leave the island. The territory Cagliari is the area characterised by the lowest quota of outgoing students in the first period, while the quota increases in the second period, exceeding 10% and reaching the levels of the territories of Sassari and Oristano. As far as destinations are concerned, Piedmont exerts a strong attraction on male students from Sassari, Oristano and Nuoro[4], since they presumably prefer the Polytechnic of Turin to the university of Cagliari for the studies in engineering. However, this destination also becomes relatively important in Cagliari in the second period, while it loses appeal in the Olbia-Tempio territory. Sassari is the territory characterised by the greatest stability of patterns in the two time periods, especially for male students (Piedmont is always on top of the preferences, followed by Lombardy and Tuscany), while it is interesting to note the rise of Piedmont in the attrattivity exerted on female students from Sassari, which becomes, together with Emilia-Romagna, the most popular destination in the second five-year period (the latter region does not seem to attract particularly male students). Cagliari is characterised by quite similar patterns between males and females, who show a strong preference for Lombardy, Lazio and Tuscany in the first period, whereas they see Lazio losing its attractiveness in the second period. Even with the relative loss of importance of Rome in the last five years, Cagliari also seems to follow at some extent the "city-to-city" pattern found for Palermo, Catania and Messina. Even if understanding the reasons for this type of mobility routes is beyond the scope of this chapter, one could hypothesise that moving to a large city has a higher psychological cost (especially if one does not have a social network in the place of arrival) for a student coming from a relatively small context with respect to a student coming from a large city. On the other hand, the decision to move to study is still a provisional choice that presumably lasts as long as the studies. Although the labour market of a big city like Milan or Rome may be more attractive to a young person from the province, this attractiveness mainly influences definitive (or

[4] For Nuoro, this destination loses appeal in favour of Tuscany in the second period.

semi-definitive) migration for work, and it is relatively less important in the choice of the university. Going outside one's own region is still considered an upgrade over staying at home, but, once out, one probably looks for the most similar context to the one in which one used to live[5].

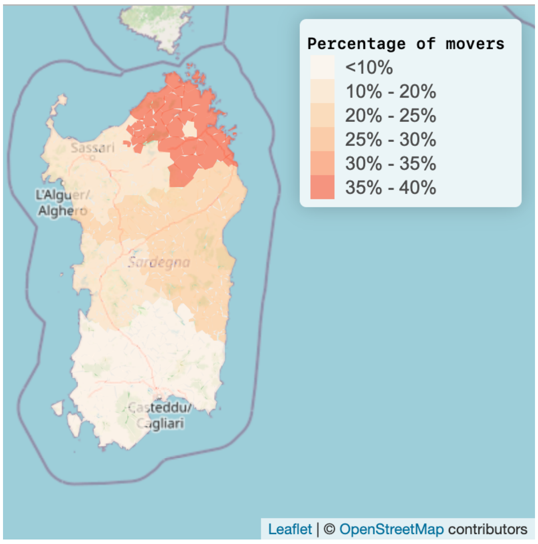


Figure 2.12: Sardinia, percentage of student movers by area of origin, 2008-2012.

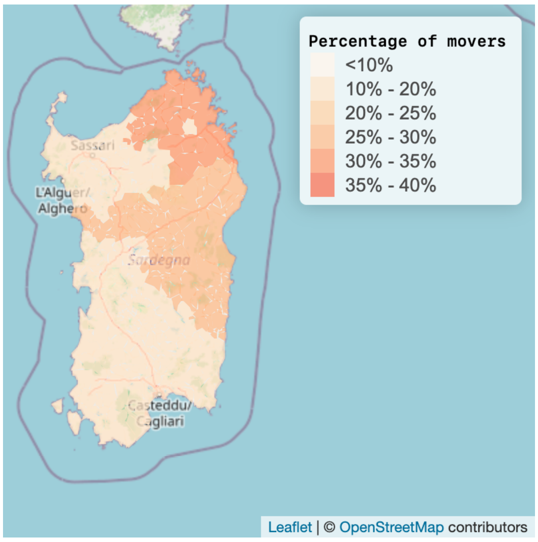


Figure 2.13: Sardinia, percentage of student movers by area of origin, 2013-2017.

[5] Semi-structured interviews with students moving from a small town in the South to a small university town in the Centre-North could be a tool to better understand if mechanisms similar to those described motivate their choices.

Table 2.4: Distribution of students enrolled in a three-year degree programme resident in Sardinia by cohort, area of origin and region of destination (row percentage values).

Period	*AreeOri*	Destinations								Total
		EMILIA-ROMAGNA %	LAZIO %	LOMBARDY %	PIEDMONT %	TUSCANY %	VENETO %	OTHER %	SARDINIA %	
2008-2012	**CAG**	1.2	1.4	1.9	0.9	1.2	0.5	1	91.9	**14562**
	NUO	3.8	2.9	3	4.1	4.6	1	3	77.7	**5117**
	OLB	4	7	8	5.3	6.7	1.7	2.8	64.5	**2349**
	ORI	3	1.9	1.8	3	3	1.3	1.8	84.1	**3012**
	SAS	1.8	1.7	2.9	4.1	2.7	0.9	1.1	84.7	**6541**
Total 2008-2012		**2.1**	**2.1**	**2.7**	**2.6**	**2.7**	**0.8**	**1.5**	**85.3**	**31581**
2013-2017	**CAG**	1.8	1.2	2.3	1.9	1.6	0.7	1.2	89.4	**13709**
	NUO	4.2	2.6	3.6	5.5	5.9	1.7	4	72.4	**4862**
	OLB	5	3.6	7	5.3	6.4	1.7	5.3	65.8	**2323**
	ORI	3.6	1.4	2.3	4	2.9	2	2.7	81	**2554**
	SAS	2.4	1.3	2.5	4.7	2.9	1.2	1.7	83.4	**6217**
Total 2013-2017		**2.7**	**1.6**	**2.9**	**3.5**	**3**	**1.1**	**2.2**	**82.8**	**29665**

Table 2.5: Distribution of Sardinian students by period, area of origin and universities of destination. The percentages for destination universities are calculated with respect to the total number of students enrolled outside the region and the percentage of students enrolled in Sardinia is compared to the total number of Sardinian students enrolled in an Italian university.

		Universities outside the region											SARDINIA	
Period	***AreeOri***	**PoliTo %**	**UniPi %**	**UniBo %**	**UniTo %**	**La Sapienza %**	**UniFi %**	**Bocconi %**	**UniPd %**	**UniMi %**	**PoliMi %**	**Other %**	**Total Outside Region**	**%**	**Total**
2008-2012	**CAG**	6.2	8.1	12.5	4.4	7.5	4.8	6.7	1.9	3.8	4	40	**1180**	91.9	**14562**
	NUO	12.3	13.1	11.8	5.9	5.3	4.7	2.5	2.1	2	1.2	39	**1141**	77.7	**5117**
	OLB	10.3	11.2	7.2	4.3	8.3	6.1	2.9	3.7	4.2	4.6	37.2	**833**	64.5	**2349**
	ORI	15.2	9	15	3.3	4.2	5.4	0.6	4.6	1.7	1.7	39.2	**479**	84.1	**3012**
	SAS	23.2	14.3	9.2	3.9	4.3	2.9	4.6	3.9	1.9	3.5	28.3	**1000**	84.7	**6541**
Total 2008-2012		**13**	**11.3**	**10.9**	**4.5**	**6.1**	**4.7**	**3.9**	**3**	**2.8**	**3.1**	**36.7**	**4633**	**85.3**	**31581**
2013-2017	**CAG**	9.9	9.8	13.9	7.6	2.8	3.2	5.2	2.4	3.8	3.6	37.6	**1457**	89.4	**13709**
	NUO	9.7	13.2	10	10.1	5	5.4	1.7	3.7	2.2	1.7	37.3	**1344**	72.4	**4862**
	OLB	7.7	12.6	10.4	7.5	3.3	4.3	1.9	3.5	4.3	3.5	41	**795**	65.8	**2323**
	ORI	14	9.9	14	7.2	1.6	3.7	1.4	3.5	3.7	1.9	39.1	**486**	81	**2554**
	SAS	21.6	13	10.9	6.4	2.6	3.1	2.7	4.9	2	2.5	30.2	**1029**	83.4	**6217**
Total 2013-2017		**12.2**	**11.8**	**11.7**	**8**	**3.3**	**4**	**2.9**	**3.5**	**3.1**	**2.7**	**36.7**	**5111**	**82.8**	**29665**

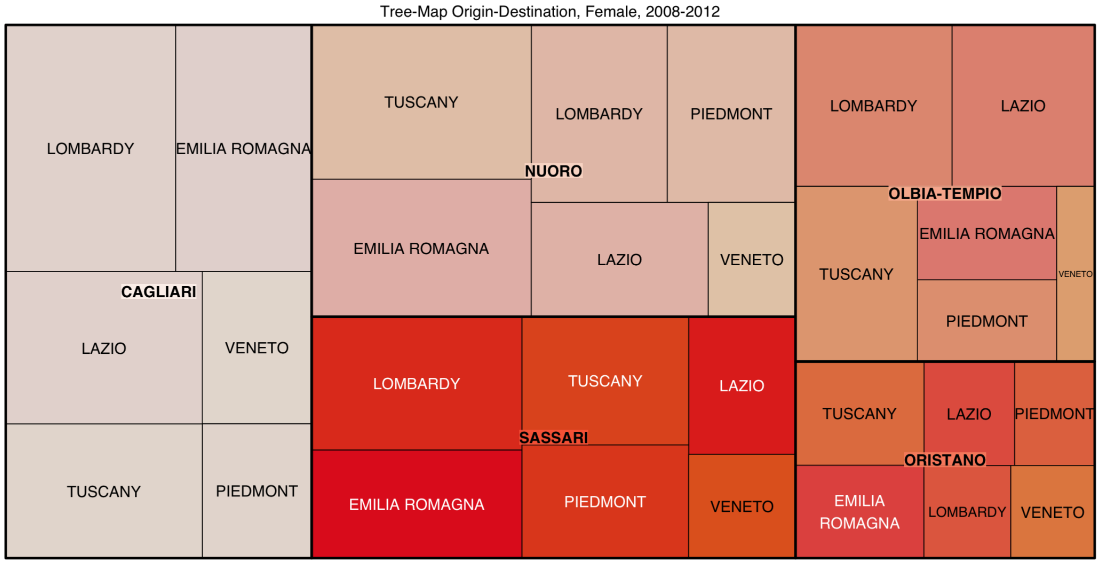


Figure 2.14: Tree-map of the origin-destination flows. The monochromatic rectangles are the selected source areas, while the inner rectangles are the destination regions. The size of the monochromatic rectangles is proportional to the population leaving the source area and the inner rectangles are proportional to the inflows into the destination region, period 2008-2012.

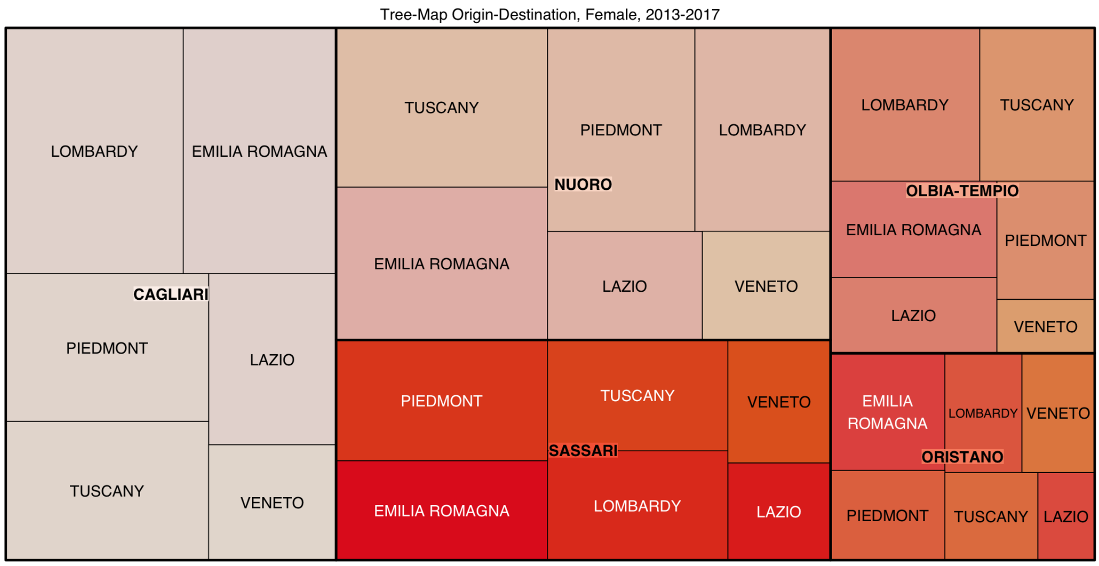


Figure 2.15: Tree-map of the origin-destination flows. The monochromatic rectangles are the selected source areas, while the inner rectangles are the destination regions. The size of the monochromatic rectangles is proportional to the population leaving the origin area and the inner rectangles are proportional to the inflows into the destination region, period 2013-2017.

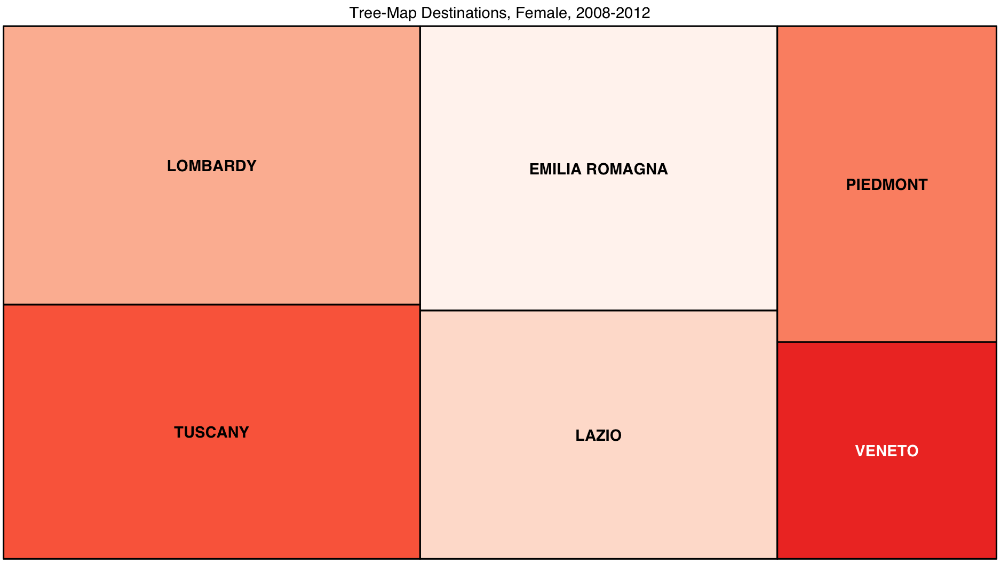


Figure 2.16: Tree-map of destination regions. The size of the rectangles is proportional to the inflows into the destination region, period 2008-2012.

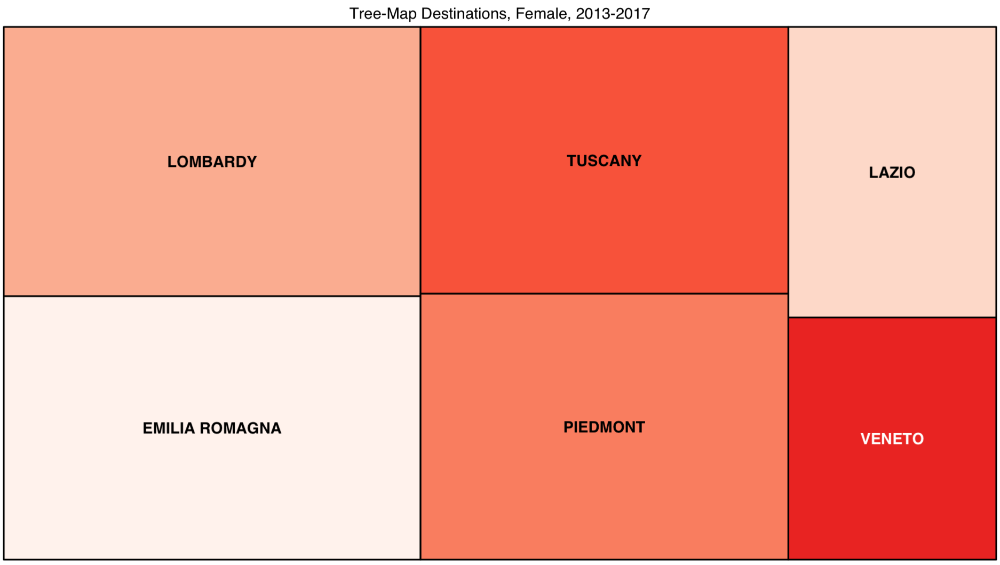


Figure 2.17: Tree-map of destination regions. The size of the rectangles is proportional to the inflows into the destination region, period 2013-2017.

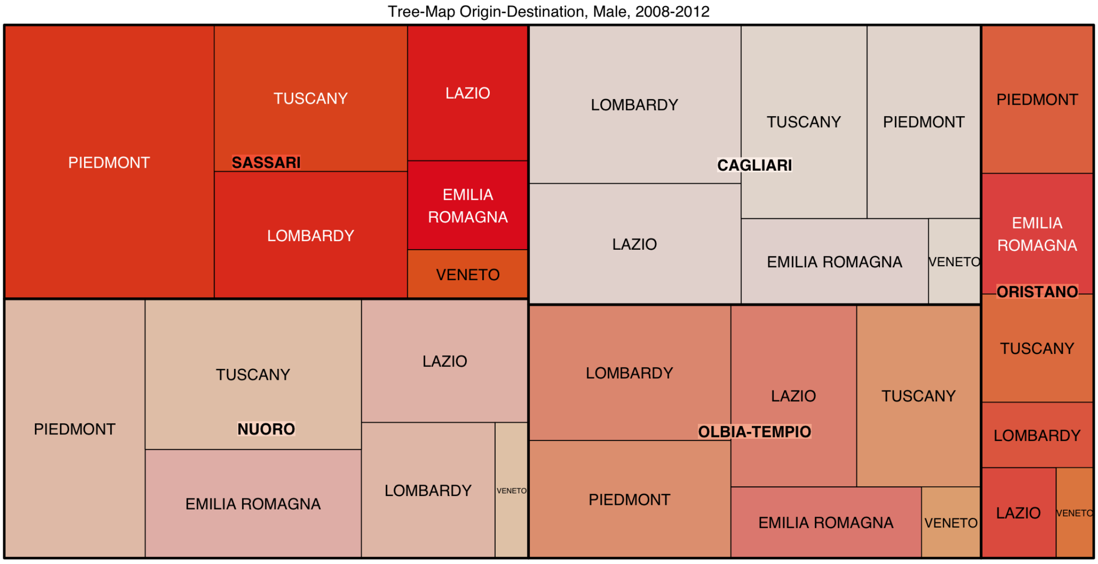


Figure 2.18: Tree-map of the origin-destination flows. The monochromatic rectangles are the selected source areas, while the inner rectangles are the destination regions. The size of the monochromatic rectangles is proportional to the population leaving the source area and the inner rectangles are proportional to the inflows into the destination region, period 2008-2012.

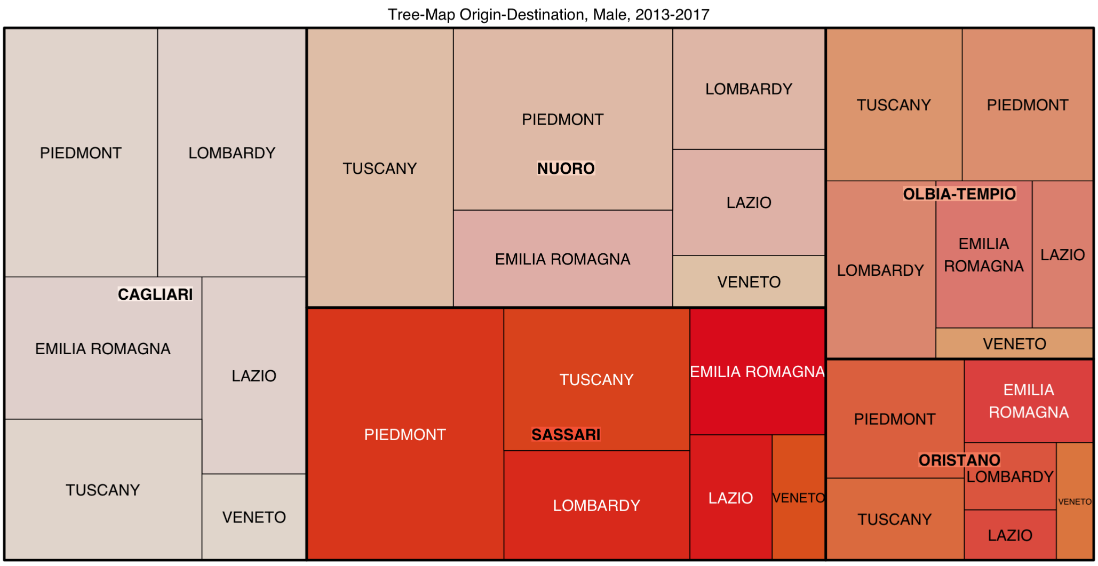


Figure 2.19: Tree-map of the origin-destination flows. The monochromatic rectangles are the selected source areas, while the inner rectangles are the destination regions. The size of the monochromatic rectangles is proportional to the population leaving the origin area and the inner rectangles are proportional to the inflows into the destination region, period 2013-2017.

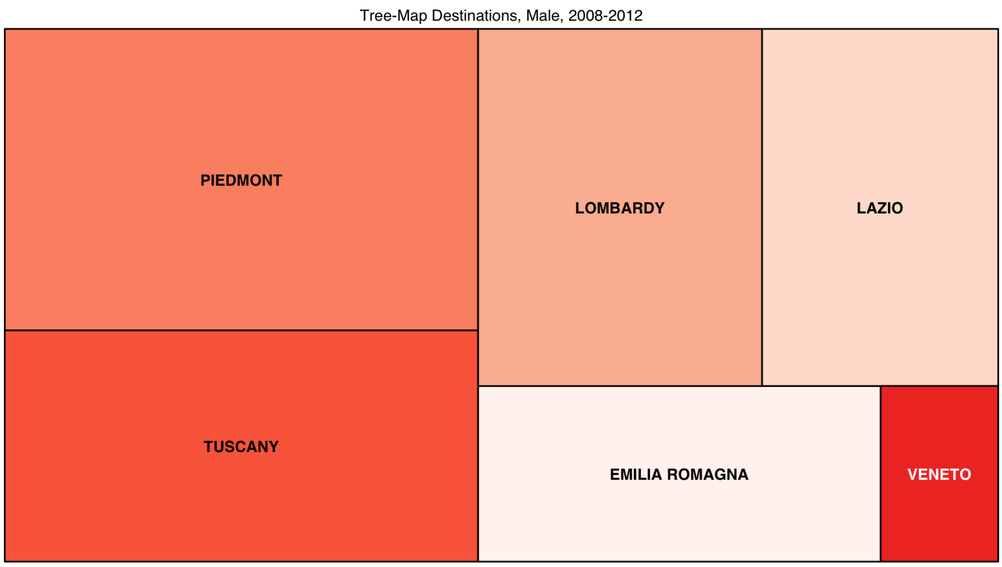


Figure 2.20: Tree-map of destination regions. The size of the rectangles is proportional to the inflows into the destination region, period 2008-2012.

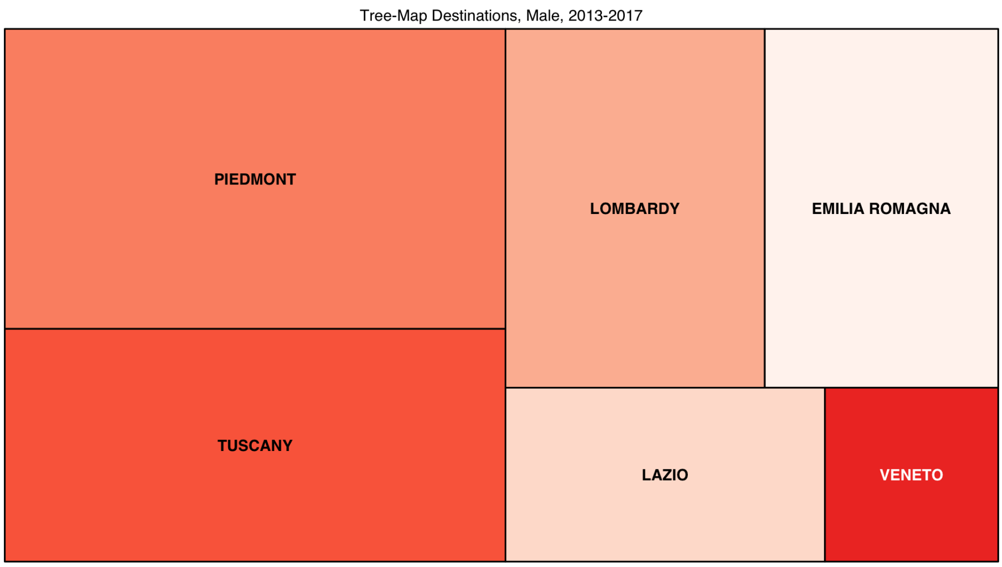


Figure 2.21: Tree-map of destination regions. The size of the rectangles is proportional to the inflows into the destination region, period 2013-2017.

Apulia

Finally, we focus the attention on Apulia. It worth noticing that some trends are similar to the ones observed in Sicily and Sardinia. Specifically, the percentages of movers in the AreeOri of Foggia, Maglie, Casarano and Tricase exceeds 50% and, therefore, it's comparable to the percentages observed in Sicily in the second period. Furthermore, Bari (the capital) seems to be the area least affected by the student mobility phenomenon. However, unlike Sicily, relatively large cities, such as Taranto, Foggia and Lecce (the latter two are also the site of a University, while the first one hosts a branch of the University of Bari) register quite high percentages of outgoing students. Piedmont does not seem to exert the strong attractiveness recorded in Sardinia and Sicily among male students, even if it is the most popular destination in the second five-year period for some areas (Tricase, Martinafranca, Maglie, Casarano). This leads one to hypothesize that the mobility towards Turin could become stronger in the next few years, also in Apulia. Almost everywhere throughout the region, Emilia-Romagna exerts a very strong attraction on both male and female students. This result differentiates the Apulian territory from Sicily and Sardinia, and perhaps reflects the presence of a historical link determined by the so called "Adriatic line". Another destination that seems to attract many Apulian students is Lombardy, which exerts the strongest attraction on students from the territory of Lecce in both periods, on Foggia in the second five-year period, and on Bari, even if, as mentioned above, the overall proportion of movers from the capital is quite limited.

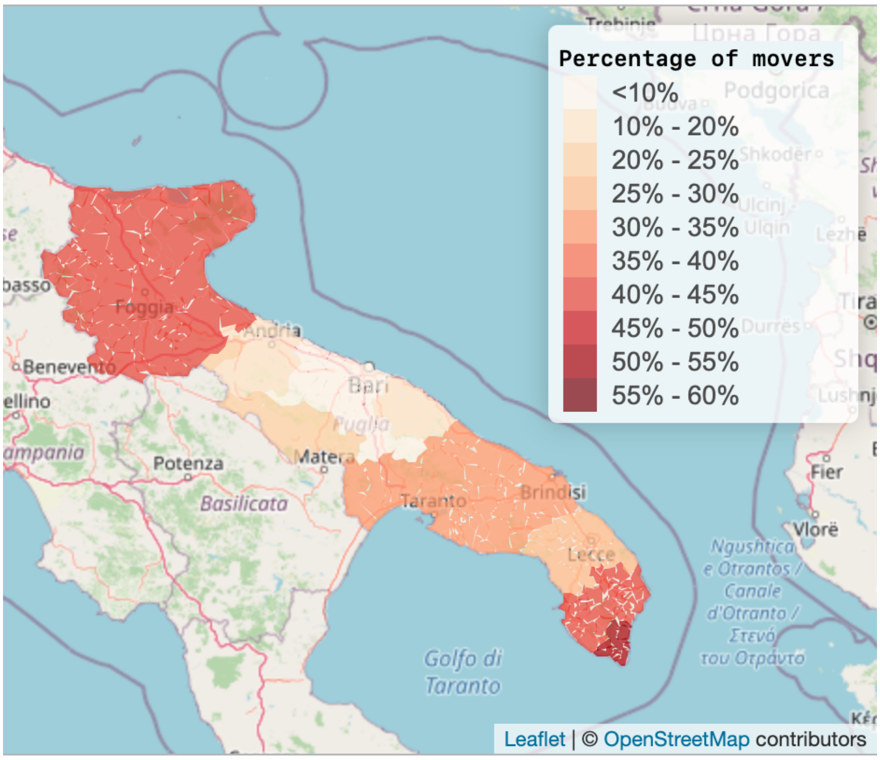


Figure 2.22: Apulia, percentage of student movers by area of origin, 2008-2012.

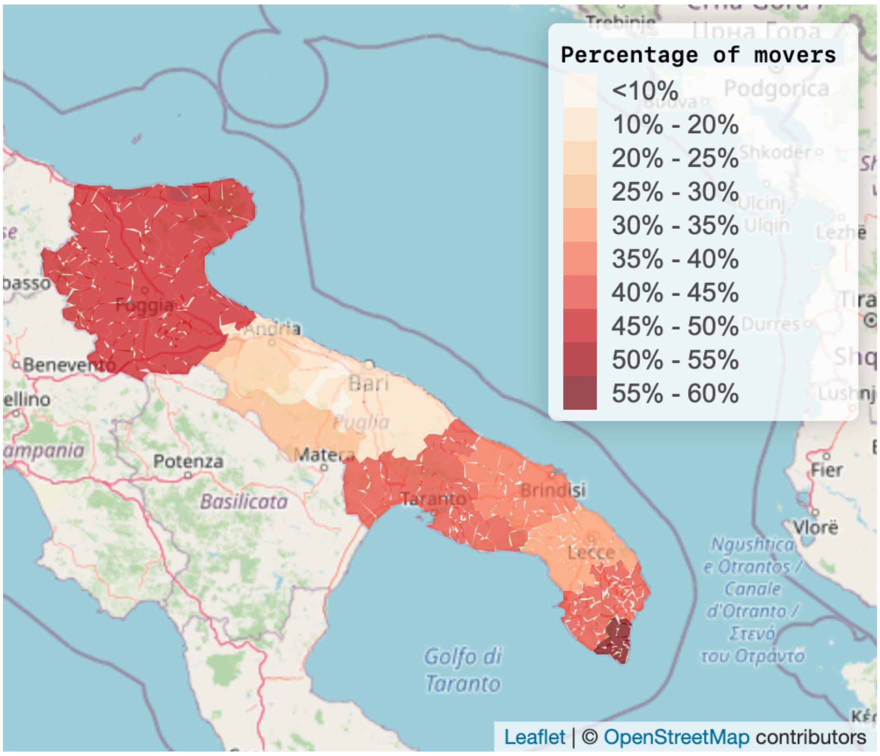


Figure 2.23: Apulia, percentage of student movers by area of origin, 2013-2017.

Table 2.6: Distribution of students enrolled in a bachelor degree residing in Apulia by cohort, area of origin and region of destination (row percentage values).

		Destinations									
Period	*AreeOri*	**CAMPANIA %**	**EMILIA-ROMAGNA %**	**LAZIO %**	**LOMBARDY %**	**PIEDMONT %**	**TUSCANY %**	**VENETO %**	**OTHER %**	**APULIA %**	**Total**
2008-2012	**ALT**	0.3	2.2	3.6	4.1	1.6	1.7	0.4	9.8	76.3	**4687**
	AND	0.4	2.6	2.4	3.2	1.3	1.5	0.5	6	82	**4383**
	BARI	0.2	0.8	1.9	2.3	0.6	0.7	0.3	2.2	91	**13931**
	BAR	0.2	1.7	1.6	2.2	1.2	1.2	0.4	3.8	87.7	**2592**
	BIT	0.3	0.9	1.1	1.1	0.5	0.4	0.5	1.9	93.4	**1707**
	BRI	0.3	7.8	4.6	6.4	4.7	2.5	0.6	6.7	66.4	**7882**
	CAS	0.3	10.5	5.7	7.7	4.8	3.3	0.9	7.3	59.5	**3510**
	CON	0.4	1.9	2.3	2.6	1	1.3	0.5	3.1	86.9	**5800**
	FOG	1.4	7.5	5.3	4.1	1.7	2	0.4	21.6	56.1	**15457**
	LEC	0.3	5.5	5.6	6.3	3.5	2.1	0.8	5.9	70	**10172**
	MAG	0.2	10	6.3	7.9	5	3.6	0.9	7.3	58.9	**3796**
	MAR	0.4	5.5	4.5	5.4	4	4.5	0.8	7.1	67.7	**3843**
	MOL	0.4	0.8	1.1	2.1	0.5	0.8	0.4	2.8	91	**4236**
	TAR	0.5	6.2	5.3	4.5	2.3	4.7	0.7	8.6	67.3	**11421**
	TRI	0.2	16.2	8.8	7.5	5.5	3.5	1.3	10.6	46.4	**2160**
Total 2008-2012		**0.5**	**5.1**	**4.1**	**4.4**	**2.3**	**2.2**	**0.6**	**8.3**	**72.5**	**95577**
2013-2017	**ALT**	0.7	2.9	2.5	5.6	2.4	2.7	0.8	10.5	71.8	**4655**
	AND	1.2	3.3	2	4.3	3.4	1.7	1	6.6	76.5	**4350**
	BARI	0.5	1.9	1.7	3	1.3	0.7	0.6	3.2	87.1	**12721**
	BAR	0.5	1.7	1.6	3.3	1.4	1.4	0.4	4.9	84.8	**2597**
	BIT	0.4	0.6	0.9	1.1	0.8	0.6	0.7	2.1	92.7	**1577**
	BRI	0.6	8.7	3.7	6.5	7.7	2.8	0.7	9	60.2	**6709**
	CAS	0.3	10.9	2.9	7.3	8.8	3.4	1	8.8	56.5	**3293**
	CON	0.4	2.8	1.9	2.7	2.2	1.4	0.6	4.3	83.7	**5111**
	FOG	2.5	8.4	4.2	5	2.9	2.1	0.9	21.1	52.9	**14586**
	LEC	0.6	5.5	3.5	7.3	4.9	2	0.9	6.5	68.6	**9397**
	MAG	0.3	9.3	4.4	9.3	7.4	3.5	0.9	8.9	55.9	**3359**
	MAR	0.7	8	3.5	5.9	7.8	3.8	1.4	10.3	58.6	**3465**
	MOL	0.6	1.5	1.2	2.8	1.3	0.7	0.7	4.1	87	**4163**
	TAR	0.9	7	5	5.9	5.2	5	1	12.3	57.7	**9719**
	TRI	0.5	15.5	4.2	9.7	9.4	4.3	0.9	12	43.6	**1856**
Total 2013-2017		**0.9**	**5.8**	**3.1**	**5.2**	**4**	**2.4**	**0.8**	**9.6**	**68.2**	**87558**

Table 2.7: Distribution of Apulian students by period, area of origin and destination universities. The percentages for destination universities are calculated with respect to the total number of students enrolled outside the region and the percentage of students enrolled in Apulia is compared to the total number of Apulian students enrolled in an Italian university.

Period	*AreeOri*	Universities outside the region												APULIA	Total
		UniCh %	UniBo %	PoliTo %	UniPr %	La Sapienza %	Bocconi %	UniPi %	PoliMi %	Cattolica %	Carlo Bo %	Other %	Total Outside Region	%	
2008-2012	**ALT**	12.3	4.8	4.6	2.3	6	4.3	2.2	3	4.9	2.2	53.5	**1113**	76.3	**4687**
	AND	16.8	8.8	6.2	3.9	2.9	4.8	1.8	3.7	4.4	3.4	43.2	**787**	82	**4383**
	BARI	10.9	6.2	5.4	1.4	5.5	9	3.7	4.6	5.4	2.2	45.9	**1251**	91	**13931**
	BAR	15	9.4	9.4	2.2	2.2	5.9	5.3	4.4	3.8	2.5	40	**320**	87.7	**2592**
	BIT	8	8	7.1	0.9	5.3	3.5	1.8	4.4	2.7	3.5	54.9	**113**	93.4	**1707**
	BRI	6.6	9.4	10.7	8.5	5.7	3.7	4.3	2.9	4.8	3.5	39.8	**2647**	66.4	**7882**
	CAS	6.8	10.3	9.1	9.6	8.4	2.8	3.9	3	4.6	3.1	38.4	**1421**	59.5	**3510**
	CON	8.8	9.9	5.4	3	5.8	4.7	4.6	6.1	2.9	3.8	44.9	**758**	86.9	**5800**
	FOG	24.8	8.7	2.9	3.9	4.7	2.3	1.8	2	1.9	2.3	44.5	**6789**	56.1	**15457**
	LEC	7.6	8.8	9.6	5.5	7.1	5.3	2.9	4.4	3.9	2.5	42.4	**3047**	70	**10172**
	MAG	6.7	11.1	9.9	9	7.1	3	4.2	3.7	3.6	3.5	38.3	**1560**	58.9	**3796**
	MAR	11.5	6.8	9.9	6.9	6	4.6	5	3.4	3.3	1.6	41	**1240**	67.7	**3843**
	MOL	13.1	6.3	4.7	0.5	3.4	7.6	2.9	2.9	5	0.5	53.3	**383**	91	**4236**
	TAR	10.1	7.1	5.4	7.1	5.7	3.3	8.6	3	2.2	2.1	45.3	**3736**	67.3	**11421**
	TRI	7.2	8.5	7.3	10.1	8	2	4	2.4	2.1	3.5	45	**1157**	46.4	**2160**
Total 2008-2012		**13.2**	**8.4**	**6.6**	**5.8**	**5.8**	**3.8**	**3.9**	**3.1**	**3.2**	**2.6**	**43.6**	**26322**	**72.5**	**95577**
2013-2017	**ALT**	8.8	6	5.8	1.4	2.9	4.3	4	3.7	3.8	2.6	56.6	**1312**	71.8	**4655**
	AND	11.4	8.2	11.9	1.6	2.9	3.9	3.1	4.3	4.6	3	45	**1023**	76.5	**4350**
	BARI	7	10.7	6.9	1	3.5	7.4	2.1	4.6	4.9	2.4	49.6	**1646**	87.1	**12721**
	BAR	8.6	6.1	7.8	1.8	2.3	7.8	2.5	3	3.3	0.8	56.1	**396**	84.8	**2597**
	BIT	13	6.1	7	0	0.9	6.1	3.5	3.5	2.6	1.7	55.7	**115**	92.7	**1577**
	BRI	7.1	10.6	14.5	5.1	3.4	2.8	3.6	3.1	3.5	4.6	41.7	**2667**	60.2	**6709**
	CAS	6.7	9.4	13.1	7.1	2.8	2.5	2.7	3.1	4.5	5.1	42.8	**1433**	56.5	**3293**
	CON	6.2	11.5	10.1	1.6	2.6	4.6	4	4.1	3.2	5.5	46.6	**834**	83.7	**5111**
	FOG	18.4	9.1	4.4	3.1	3.4	1.9	1.7	2.6	1.8	2.4	51.2	**6873**	52.9	**14586**
	LEC	7.8	9.1	12	3.1	2.7	5.1	2.3	6.1	3.8	3.3	44.6	**2948**	68.6	**9397**
	MAG	5.9	10.1	10.4	4.7	4.1	3.8	3.8	5.8	3.8	6.4	41.1	**1481**	55.9	**3359**
	MAR	8.3	10.5	12.4	5.1	3.3	4	4.2	1.8	2.7	2.7	45.1	**1435**	58.6	**3465**
	MOL	12.2	7.8	8.1	1.1	1.9	6.5	2	4.3	3	2.4	50.7	**540**	87	**4163**
	TAR	9.7	7.2	8.4	4.2	3.3	2	5.9	2.9	2.4	3.4	50.5	**4113**	57.7	**9719**
	TRI	6.5	10.7	10.2	4.9	3.2	2	3	3.7	2.7	5.4	47.7	**1047**	43.6	**1856**
Total 2013-2017		**10.6**	**9.1**	**8.9**	**3.5**	**3.2**	**3.4**	**3.2**	**3.6**	**3.1**	**3.4**	**47.9**	**27863**	**68.2**	**87558**

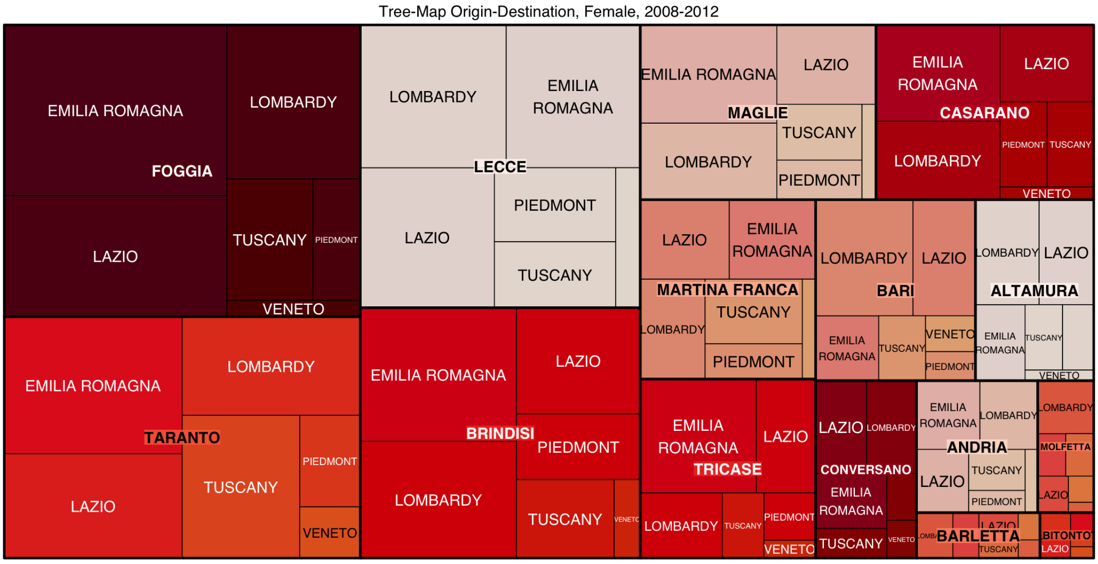


Figure 2.24: Tree-map of the origin-destination flows. The monochromatic rectangles are the selected source areas, while the inner rectangles are the destination regions. The size of the monochromatic rectangles is proportional to the population leaving the source area and the inner rectangles are proportional to the inflows into the destination region, period 2008-2012.

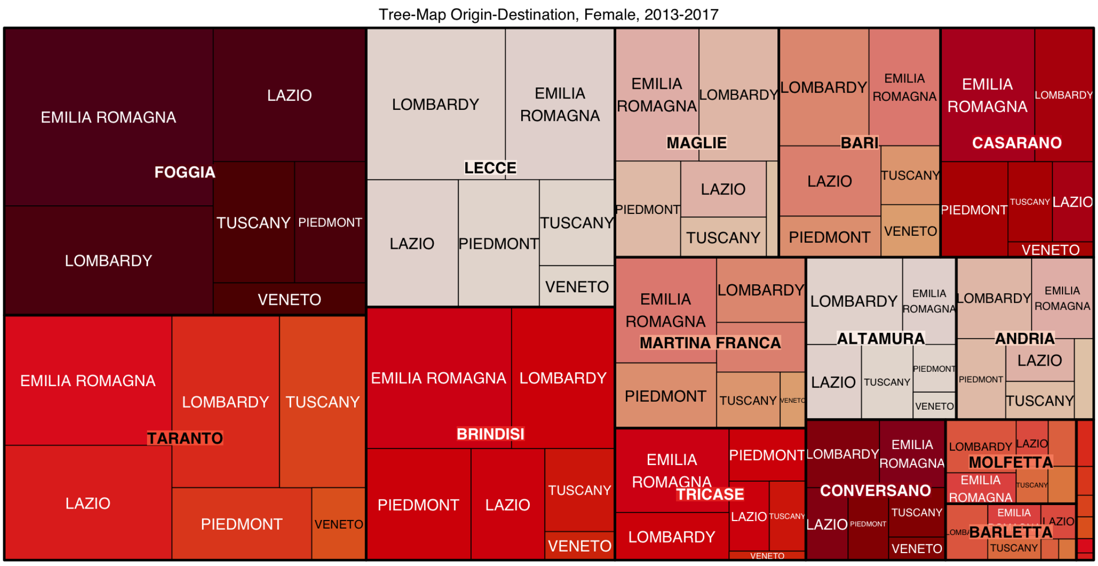


Figure 2.25: Tree-map of origin-destination flows. The monochromatic rectangles are the selected source areas, while the inner rectangles are the destination regions. The size of the monochromatic rectangles is proportional to the population leaving the origin area and the inner rectangles are proportional to the inflows into the destination region, period 2013-2017.

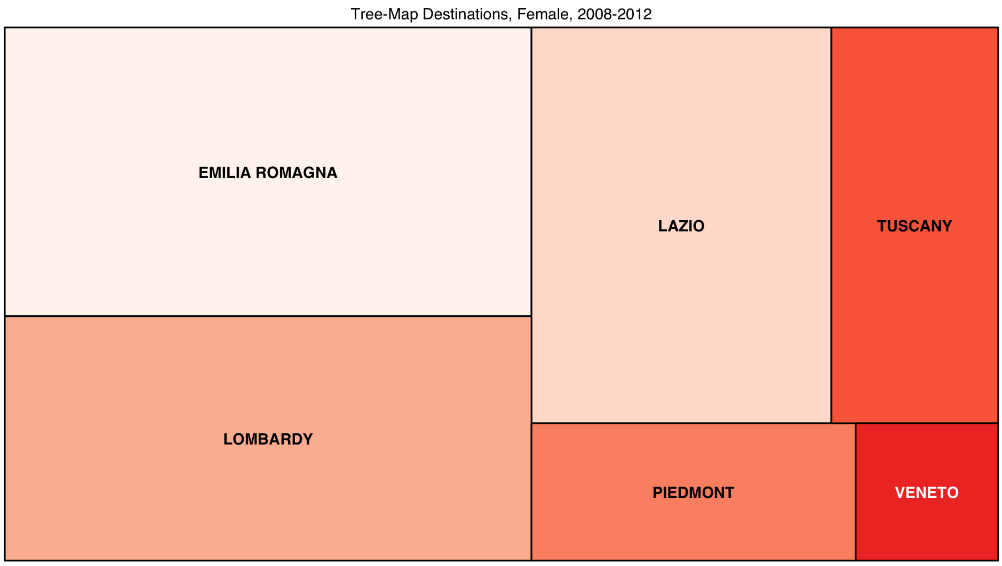


Figure 2.26: Tree-map of destination regions. The size of the rectangles is proportional to the inflows into the destination region, period 2008-2012.

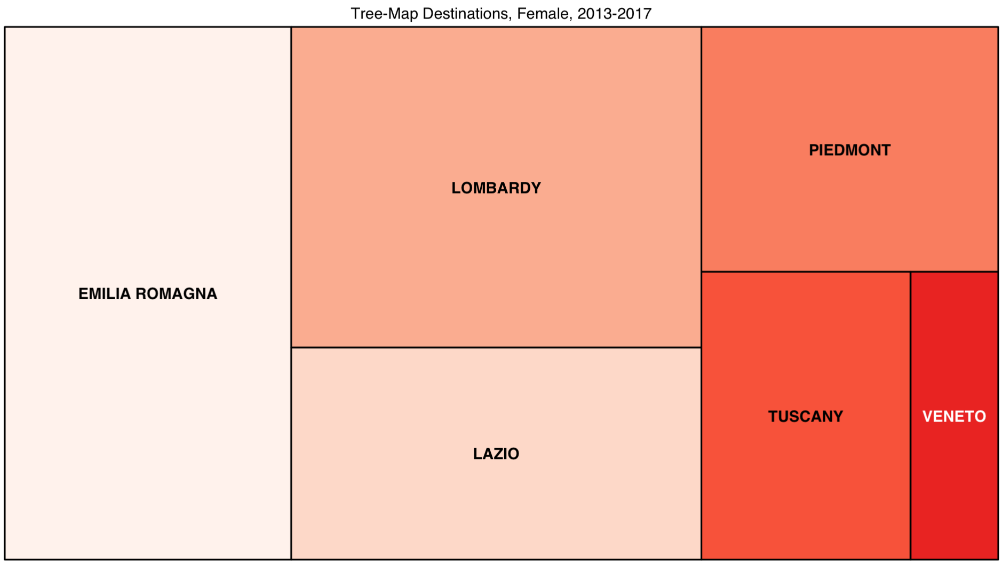


Figure 2.27: Tree-map of destination regions. The size of the rectangles is proportional to the inflows into the destination region, period 2013-2017.

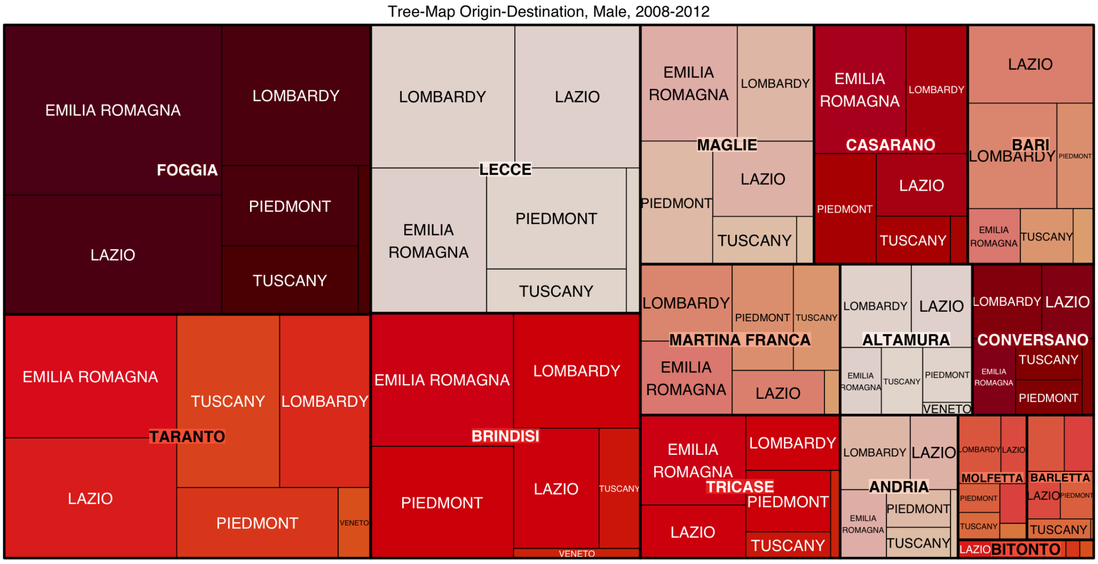


Figure 2.28: Tree-map of the origin-destination flows. The monochromatic rectangles are the selected source areas, while the inner rectangles are the destination regions. The size of the monochromatic rectangles is proportional to the population leaving the source area and the inner rectangles are proportional to the inflows into the destination region, period 2008-2012.

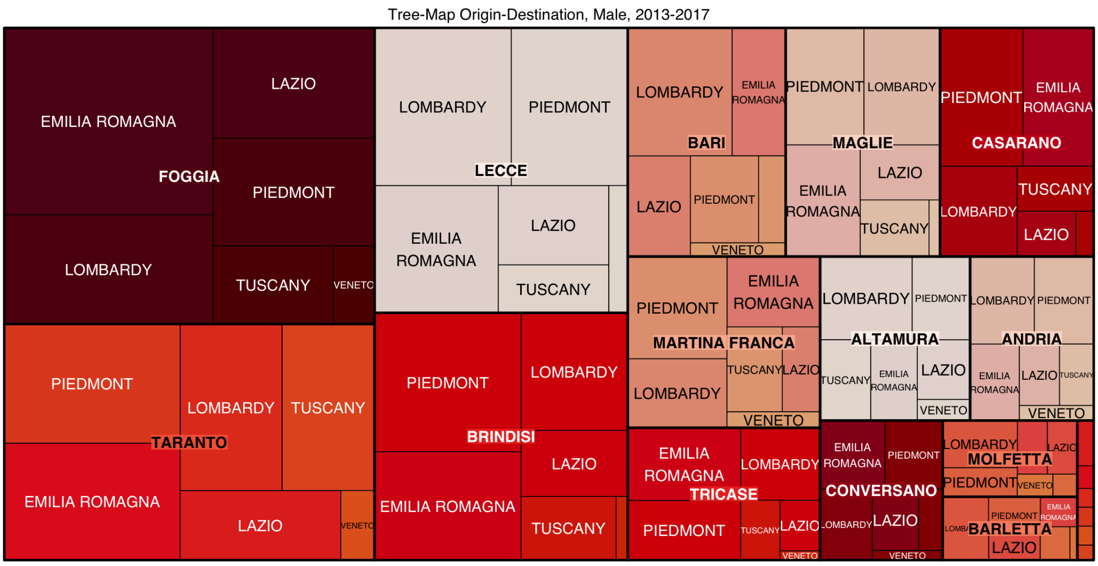


Figure 2.29: Tree-map of the origin-destination flows. The monochromatic rectangles are the selected source areas, while the inner rectangles are the destination regions. The size of the monochromatic rectangles is proportional to the population leaving the origin area and the inner rectangles are proportional to the inflows into the destination region, period 2013-2017.

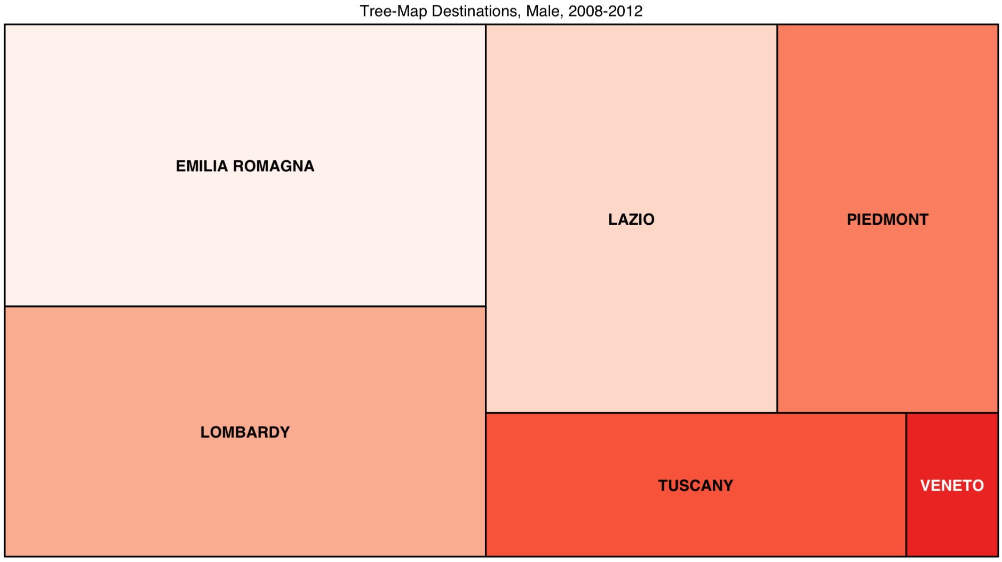


Figure 2.30: Tree-map of destination regions. The size of the rectangles is proportional to the inflows into the destination region, period 2008-2012.

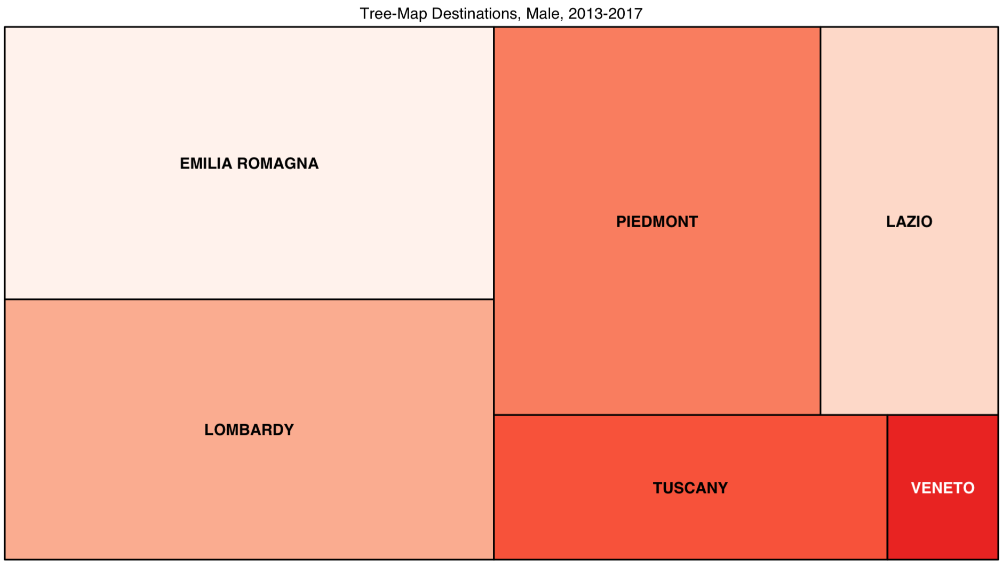


Figure 2.31: Tree-map of destination regions. The size of the rectangles is proportional to the inflows into the destination region, period 2013-2017.

2.3.2 Estimation of residuals and clustering process

Log-linear model of independence and residuals

As mentioned above, the objective of this study is to analyse the presence of chain migration effects on student mobility. To this end, a log-linear model of independence has been constructed that allows us to isolate "the attraction effect" and "the expulsion effect".

$$\log \mu_{ij} = \lambda + \lambda_i^{AreeOri} + \lambda_j^{RegDest}, \tag{2.2}$$

where

- μ_{ij} is the expected value of the number of students that move from the *i-th* Source *AreaOri* to the *j-th* Destination Region (*RegDest*);
- $\lambda_i^{AreeOri}$ and $\lambda_j^{RegDest}$ are the row and column effects, respectively.

In eq. (2.2), $\lambda_i^{AreeOri}$ and $\lambda_j^{RegDest}$ represent the *i-th* row effect and the *j-th* column effect, respectively. In our model, they carry information about the "repulsiveness" of the *i-th AreeOri* and the "attractiveness" of the *j-th* destination region, respectively. Significant deviations from the expected values provided in eq. (2.2) may be due to chain migration effects.

Therefore, standardised residuals are used to measure how much the observed flows deviate from those estimated according to the model:

$$\epsilon_{ij} = \frac{n_{ij} - \mu_{ij}}{\sqrt{\mu_{ij}(1 - p_{i+})(1 - p_{+j})}} \tag{2.3}$$

Where:

- n_{ij} are the flows observed from *AreaOri* i towards destination region j;
- μ_{ij} are the flows from i to j estimated from equation (2.2);
- p_{i+} and p_{+j} are the marginal proportions of the *i-th* row and *j-th* column, respectively.

All residues $|\epsilon_{ij}| > 2$ correspond to pairs (i, j) with an attraction level ($\epsilon_{ij} > 2$) and a level of repulsion ($\epsilon_{ij} < -2$) higher/lower than expected under the independence assumption. The matrix of residuals ϵ_{ij} is calculated for each time period: $t_1 = 2008-2012$ and $t_2 = 2013 - 2017$. If both $\epsilon_{i^{'}j^{'}t_1} > 2$ and $\epsilon_{i^{'}j^{'}t_2} > 2$, then we hypothesize a migratory chain effect.

Clustering process and graphical representation

To better understand the presence of mobility patterns that depend on time, sex of the students, and field of study, we apply a hierarchical clustering algorithm to the modalities of the pairs $(AreeOri, RegDest)$ to two different sets of covariates: *i)* the five-year periods and the disciplinary field (Scientific/non Scientific), and *ii)* the five-year periods and the gender. Indeed, our interest is not strictly constrained to the composition of resulting clusters, rather it focuses on the ordering of variables implicitly provided by the hierarchical clustering that allows to highlight pattern specificities in the data.

The choice of covariates to investigate allows one to reveal the persistence or not in time of attraction/repulsion of the origin-destination pairs by isolating the pairs $(AreeOri, RegDest)$ with respect to the five-year period. This choice also allows one to investigate if such "persistence" eventually depends on gender and disciplinary field. The interest in gender and disciplinary area derives from the fact that these two variables are associated with university mobility, mostly due to the greater propensity to mobility of male students interested in the STEM disciplinary areas [4]. Therefore, to analyse the presence of a possible migratory chain taking into account gender and disciplinary area as determinants of student mobility, we decided to estimate the model in (2.2), and its residuals in (2.3), by adding separately the gender component and the disciplinary area component. In this way, we obtained the following models:

$$\log \mu_{ijk} = \lambda + \lambda_i^{AreeOri} + \lambda_j^{RegDest} + \lambda_k^{Gender} \tag{2.4}$$

$$\log \mu_{ijk} = \lambda + \lambda_i^{AreeOri} + \lambda_j^{RegDest} + \lambda_k^{STEM} \tag{2.5}$$

Also in this case, both the residuals of the model (2.4) and those of the model (2.5) have been calculated for each one of the considered five-year periods ($t_1 = 2008 - 2012$; $t_2 = 2013-2017$). These models provide an estimate of the expected value of student flows under the hypothesis of independence against which measuring the presence of an extra-flow of students in the data corresponding to each mobility pattern–origin $i^{'}$, destination

j'–controlling for the third variable, k'. Then, we consider the residuals from these models (eq.s 2.4 and 2.5), and apply a hierarchical clustering algorithm to investigate the similarity among preferential mobility patterns.

The hierarchical clustering procedure, as applied to the Euclidean distance matrix of residuals, allows one to elicit from data the similarity structure of origin-destination patterns. Indeed, hierarchical clustering constructs a dendrogram of origin-destination units, by iteratively grouping them together at reduced levels of similarity, until they merge into only one cluster–the root. In other words, increasing levels of root depth correspond to larger levels of aggregation. In particular, the clustering algorithm used is the Complete Linkage Cluster Analysis that, compared to methods such as the Single Linkage, or the Average Linkage (weighted and unweighted) guarantees a better separation of the groups [2] by avoiding the so-called chaining phenomenon, which is typical of clustering algorithms based on the nearest-neighbour distance. Complete Linkage effectively reduces chaining phenomenon with respect to the other clustering methods since the iterative aggregation is based on the farthest-neighbour distance [42], thus creating more compact and homogeneous clusters [2, 42].

2.4 Results

The results of the cluster analysis for Sicily (see Figures 2.32 and 2.33) confirm that student chain migration plays an important role in explaining the patterns that emerge from the reading of the Tree-Map. In particular, if we consider the humanities (Figure 2.32), the residuals are particularly high in both five-year periods–thus highlighting a number of students permanently higher than what would be expected on the sole basis of pull and push effects–in the following paths: Palermo-Lazio, Catania-Lazio, Messina-Lazio, Palermo-Lombardy, Messina-Lombardy, Catania-Lombardy, Trapani-Tuscany, Trapani-Emilia-Romagna,Vittoria-Emilia-Romagna, Agrigento-Tuscany, Vittoria-Tuscany, Ragusa-Tuscany, Castelvetrano-Emilia-Romagna. All these paths point to Northern regions, specifically to Tuscany, Emilia-Romagna, and Lombardy, which underscores the preference of Sicilian students in the humanities for specific universities, *i.e.*, Siena (Tuscany), Bologna (Emilia-Romagna), and Cattolica (Milan, Lombardy).

Even for patterns that display rather low values of flow, such as Messina-Veneto and Catania-Veneto, the present analysis allows to discover excess-flows that support the role of chain migration effects. The patterns that include Piedmont and in which the number of students is stably higher than expected refer to the non-humanistic sectors. This result further supports the hypothesis that the "Politecnico di Torino" represents an important basin of attraction especially for students coming from the Sicilian smaller cities (in fact, no strong chain effects are apparent for Palermo and Catania). Therefore, the popularity of this site among Sicilian students coming from small cities not only derives from the undeniable quality of the educational offer, but also from chain migration effects. On the contrary, for the non-humanistic sectors, Piedmont scores number of students lower than expected under the hypothesis of proportional flows. This reading is further supported by the cluster analysis carried out by gender over time (Figure 2.33). Likely due to persistent problems of gender stereotypes [24], scientific subjects continue to be the prerogative of men and, therefore, the patterns in which the destination is Piedmont are mainly composed of male students.

As far as Sardinia is concerned, the strong link in both periods between Sassari and Piedmont for scientific subjects is very evident (Figure 2.34). The link between Sassari and Piedmont for scientific subjects is milder but still significant, while, as shown by the Tree-Map, a process of student chain migration only becomes apparent in the last period for students coming from Cagliari. In the field of scientific subjects also Tuscany seems to be linked by a chain migration effect to the territories of Sassari, Nuoro, and Olbia-Tempio. It is also worth mentioning that the network effect seems to negatively contribute to the mobility pattern of students interested in the humanities from the territory of Sassari towards Tuscany. The link between Veneto and Oristano for the humanities is not surprising given the historical relations between these two territories nourished by the presence of numerous families of Venetian origin who arrived in this area of the island during the fascist land reclamation (see Ruiu *et al.* 2020 [97]). This evidence therefore goes to support the goodness of the proposed method to capture chain effects.

Among the non-scientific disciplines, it is also reported that chain effects seem to fuel Cagliari's links with Lazio, Lombardy and Emilia-Romagna. The latter region is also strongly linked to Oristano and Nuoro, while students from Olbia-Tempio seem to have deeper links with Lombardy. When gender is also considered for Sardinia (Figure 2.35), it is confirmed that patterns concerning scientific subjects are driven by flows of male students.

Finally, apparent differences emerge between Apulia (Figures 2.36 and 2.37) and the other two investigated regions: preferential patterns revealed for scientific subjects are pretty similar to the ones revealed for humanistic subjects–see Figure 2.36. Our interpretation of this difference is that geographical and logistic reasons mostly influence the pattern formation. Sardinia and Sicily are islands and, therefore, do not have "neighboring regions" that can easily be reached by car, bus, or train. Furthermore, all the destinations in the Centre-North of Italy, perhaps excluding Milan and Rome (for which many low-cost flights are available) are similarly expensive to reach from both Sicily and Sardinia. For students from Apulia it is instead relatively easy to reach also Emilia-Romagna, which, indeed, seems to be the region that attracts the most from both areas of knowledge. The ease of moving by lowering both the monetary and psychological cost of it implies that students are more likely to go outside the region to study, independently of the subject area and the gender. Along this line of thinking, gender differences may be lowered by the fact that preferential patterns of mobility are more dependent on the transportation network than on chain migration. However, looking in more detail at the identified patterns, it is reasonable that chain migration also plays a role in shaping mobility patterns. This conclusion can be reached by noticing that some destinations, which are equally appealing in terms of quality of education, labour market, and transportation network from Apulia, such as Veneto and Lombardy, show very different mobility patterns of students. Specifically, many preferential patterns from Apulia point to Lombardy, whereas just a few patterns point to Veneto.

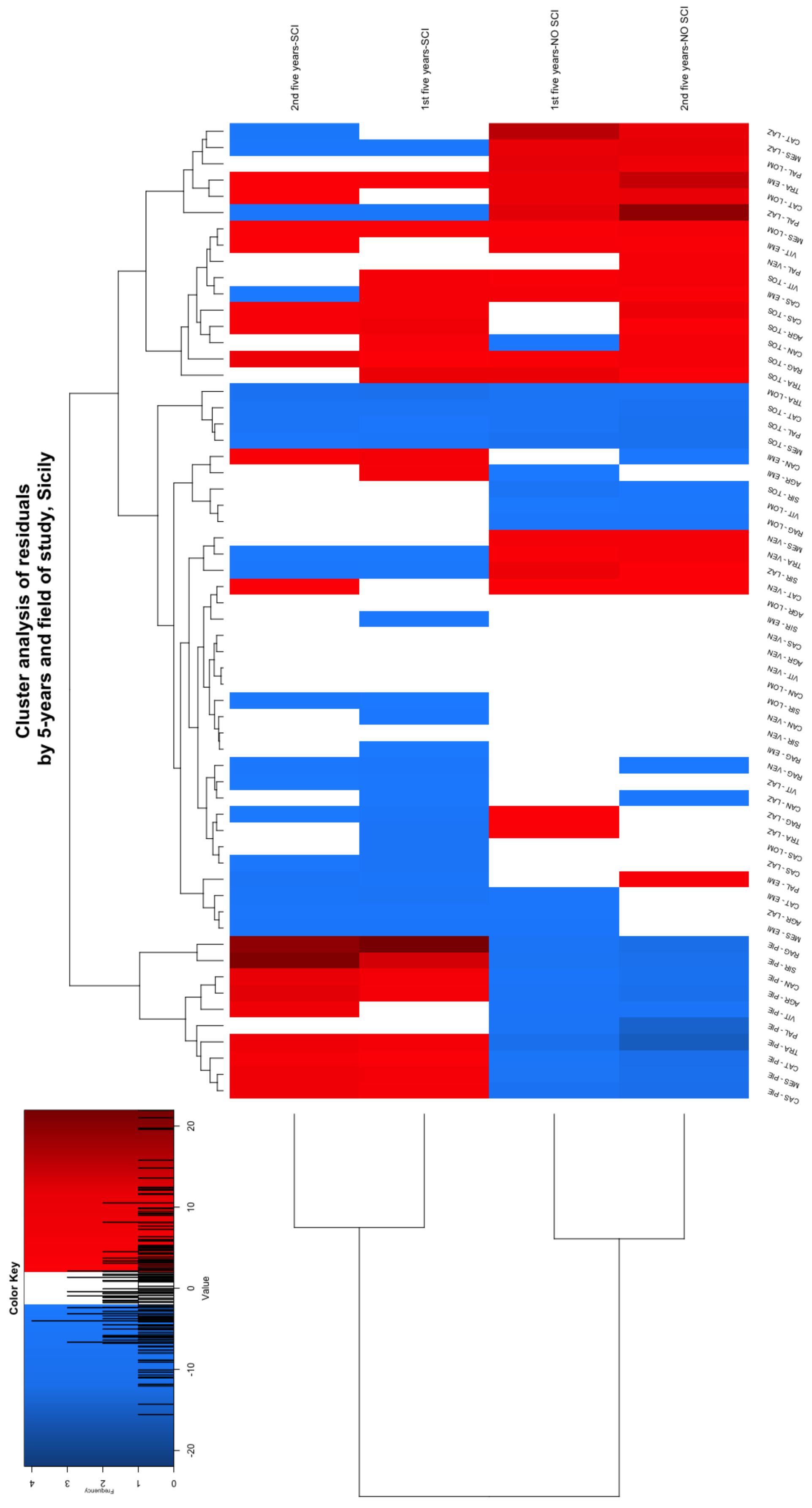


Figure 2.32: Heatmap of residuals clustered by five-years, subject area, and origin-destination. Residuals less than −2 correspond to blue rectangles, between −2 and 2 to white rectangles, and greater than 2 to those in red.

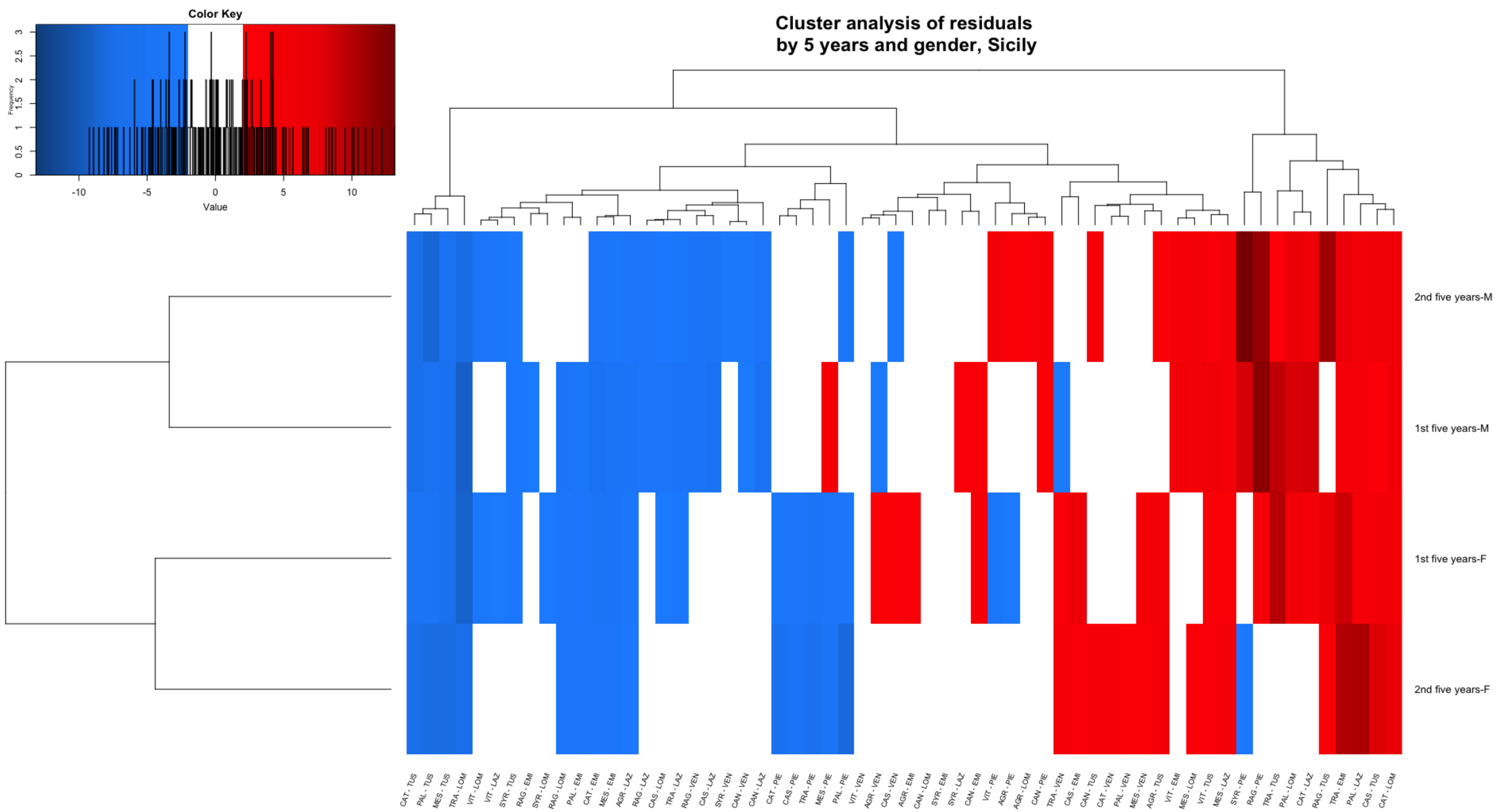


Figure 2.33: Heatmap of residuals clustered by five-years, gender, and origin-destination. Residuals less than -2 correspond to blue rectangles, between -2 and 2 to white rectangles, and greater than 2 to those in red.

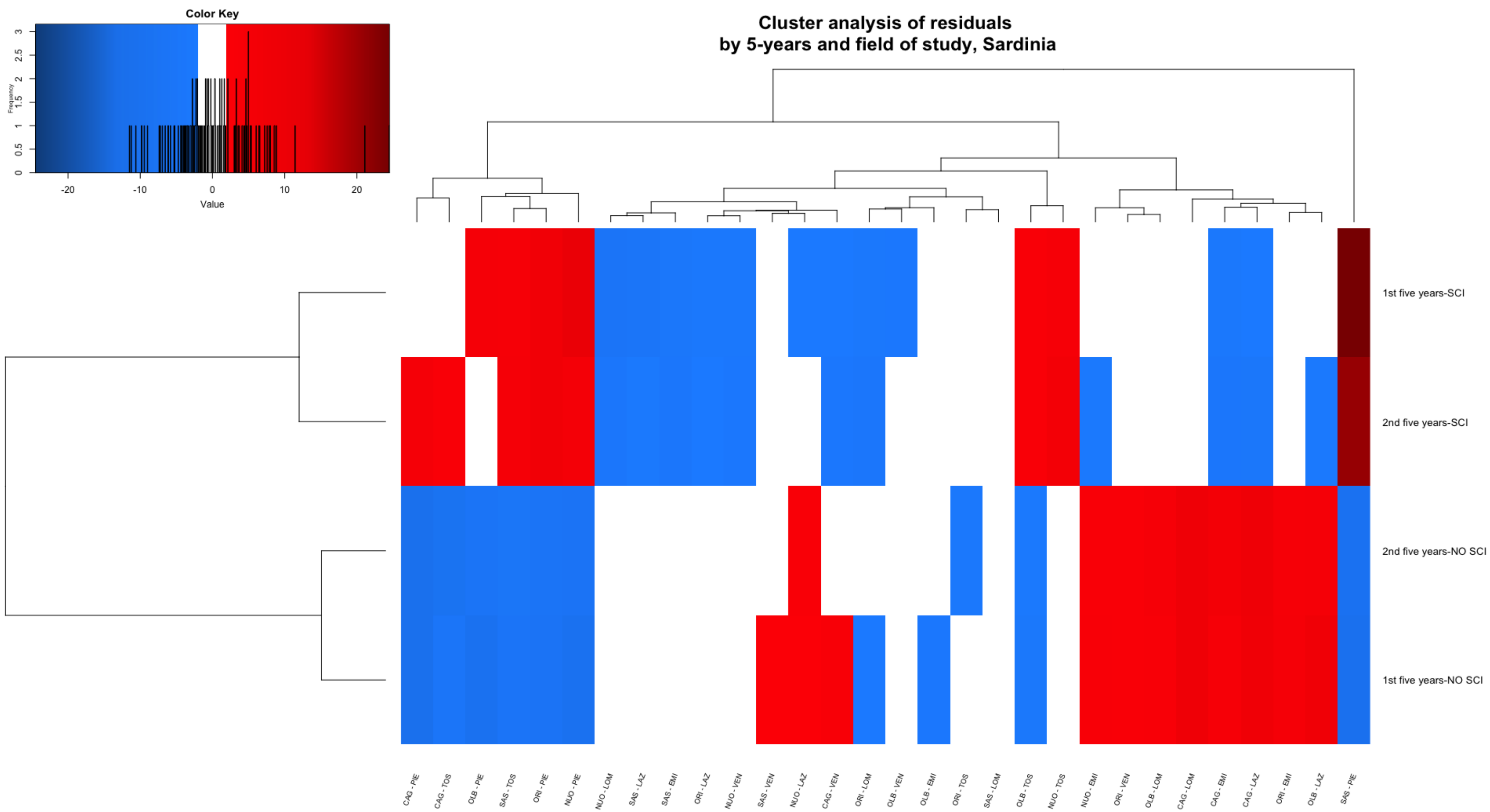


Figure 2.34: Heatmap of residuals clustered by five-years, subject area, and origin-destination. Residuals less than -2 correspond to blue rectangles, between -2 and 2 to white rectangles, and greater than 2 to those in red.

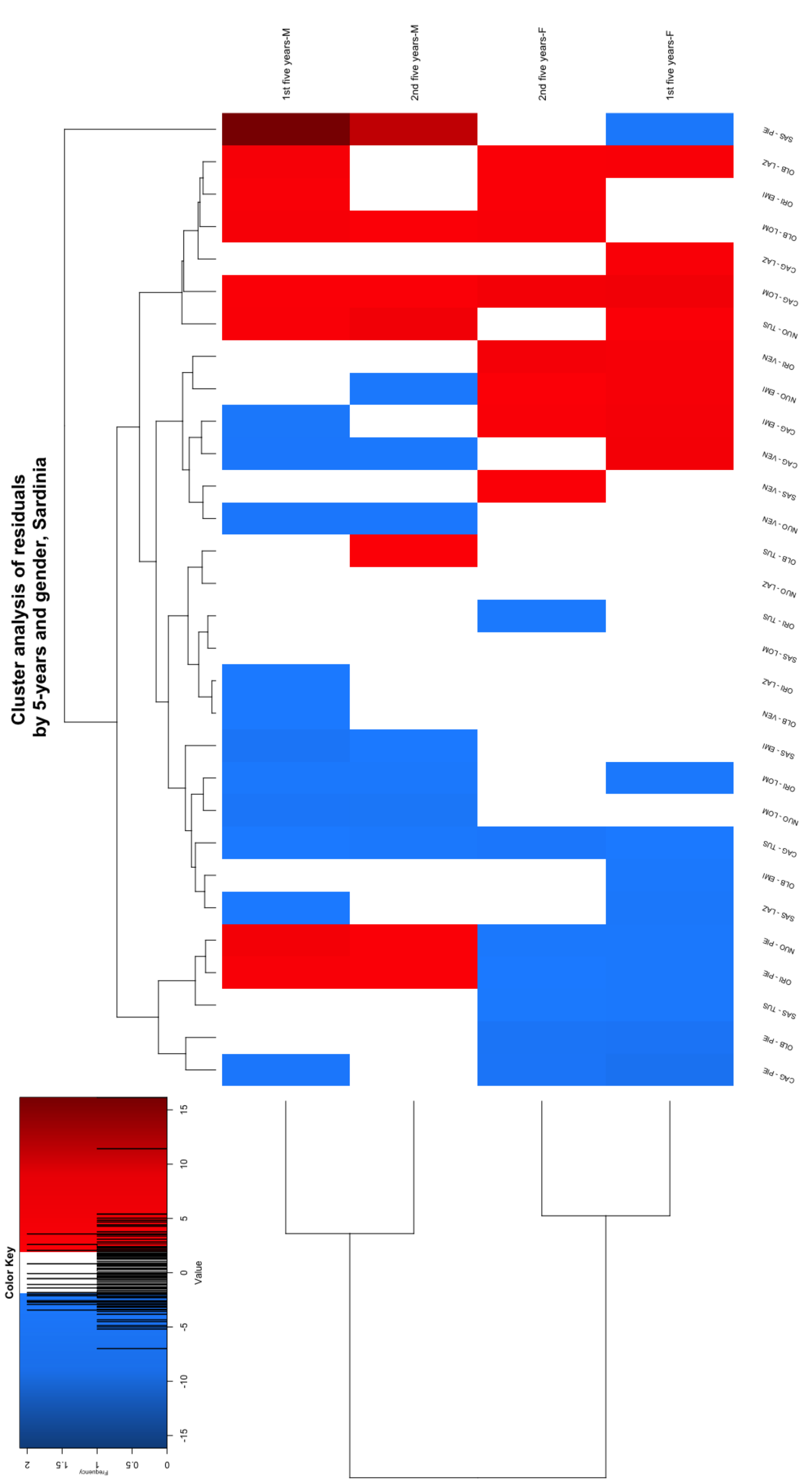


Figure 2.35: Heatmap of residuals clustered by five-years, gender, and origin-destination. Residuals less than −2 correspond to blue rectangles, between −2 and 2 to white rectangles, and greater than 2 to those in red.

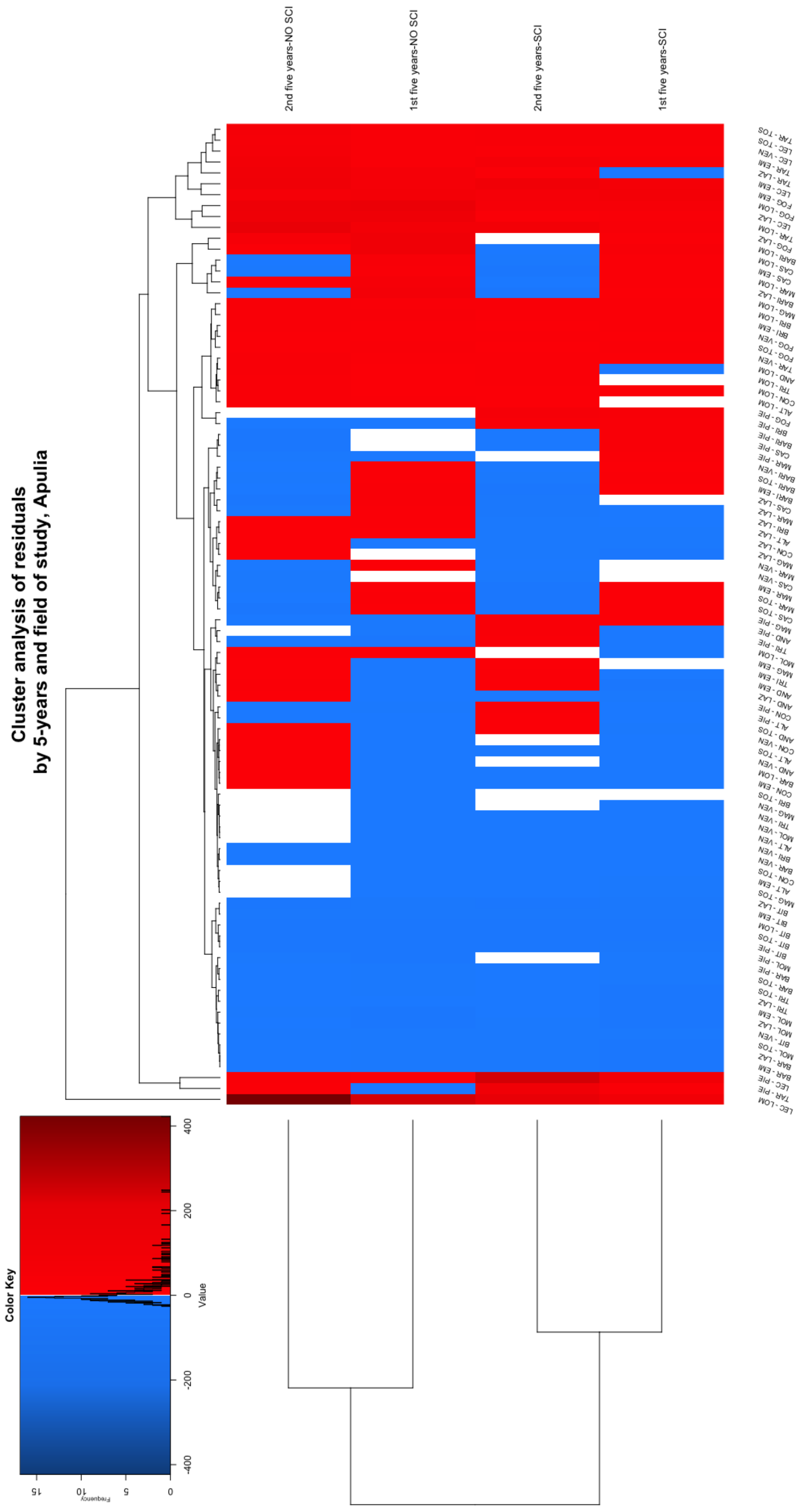


Figure 2.36: Heatmap of residuals clustered by five-years, subject area, and origin-destination. Residuals less than −2 correspond to blue rectangles, between −2 and 2 to white rectangles, and greater than 2 to those in red.

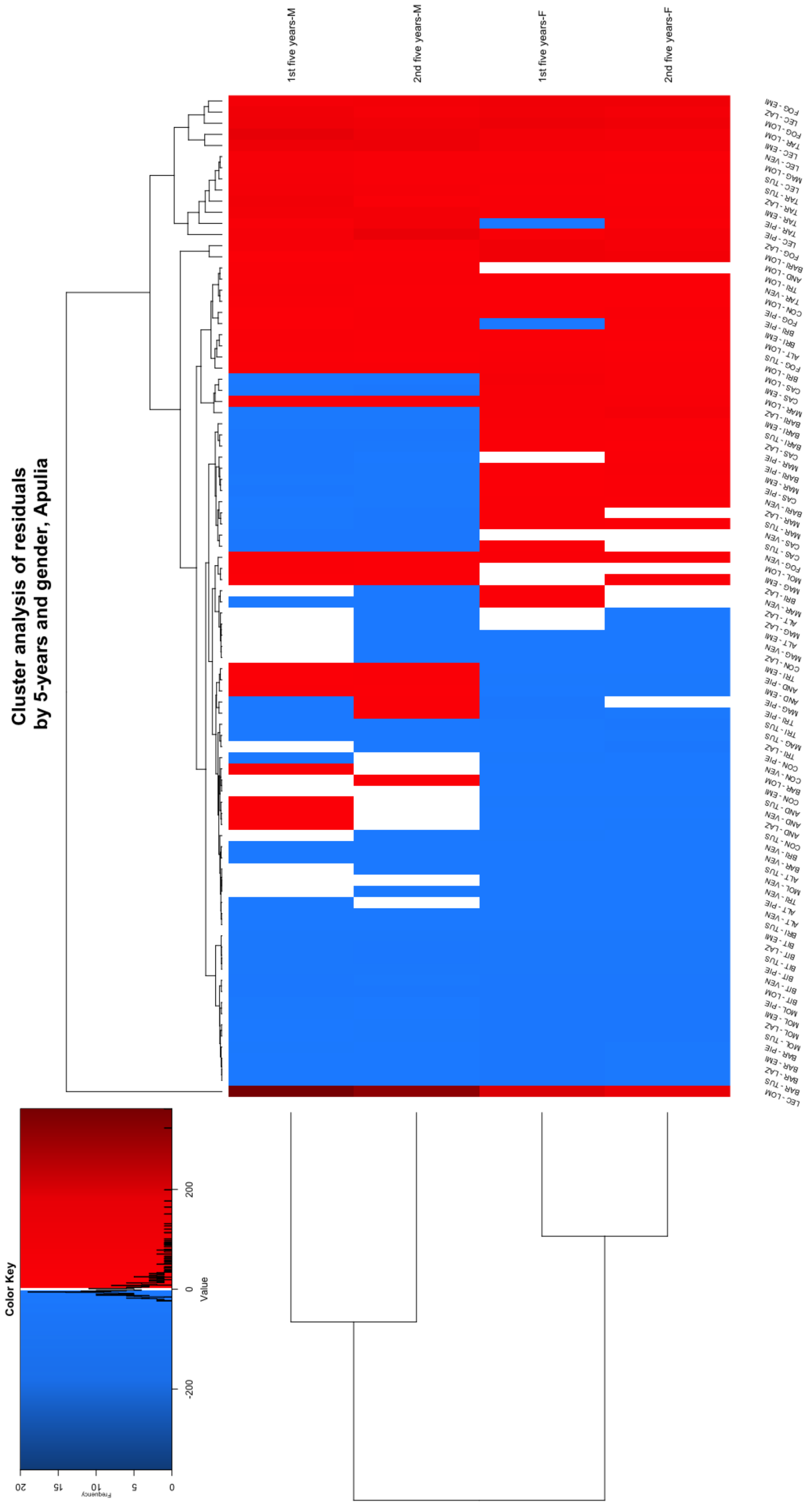


Figure 2.37: Heatmap of residuals clustered by five-years, gender, and origin-destination. Residuals less than −2 correspond to blue rectangles, between −2 and 2 to white rectangles, and greater than 2 to those in red.

2.5 Conclusions

Understanding the reasons behind the choice of a university location is increasingly crucial in a country like Italy where the financing of the universities is parameterised, for a consistent part, on their student pool and, therefore, the very survival of universities is jeopardised in those areas of the country that see a constant haemorrhage of freshmen. Indeed, the Italian high education system fails to attract students from abroad and, on the other hand, is characterised by a deep demographic drowsiness that will lead to a reduction in the size of student cohorts in the coming years (unless policy changes will determine a strong reduction in school dropouts). These negative effects are partially compensated in the North of the country by the net inflow of students coming from Southern regions, which, instead, deepens the crisis of universities in the South. The bottom line is that the net out-flow of students from the Southern regions contributes significantly to increase the so-called North-South divide in Italy.

In this chapter, a statistical method has been proposed and applied to three "big" case studies (Apulia, Sardinia and Sicily), in order to identify the patterns of student mobility that can be at least partially explained by chain migration effects. The hypotheses used in the work are essentially three: *i)* the areas of origin of the flows can be identified by analysing the mobility flows during secondary school and identifying communities that gravitate around the same hub provider of school services; *ii)* the forces of attraction and repulsion towards/from a destination/origin can be quantified by using the column/row marginals in an origin-destination matrix; *iii)* the difference between the realised flows and the expected flows generated by the factors of attraction/repulsion is due to the effect of the chain migration. The first hypothesis is reasonable since communities gravitating around the same centre for the provision of education services are likely to largely share similar cultural traits and economic wealth. The second and third hypotheses are well-established within the analysis of qualitative data of student mobility, although, to the best of our knowledge, the present work is the first to use this technique to perform a comprehensive quantitative analysis of micro-data of student mobility.

Obviously, the reasonableness of the assumptions does not imply that the work is free from limitations. The first is certainly having focused the attention only on bachelor students. The reason for this choice has in fact more practical than scientific motivations.

The numbers of bachelor's degree students, especially in Sardinia, are small and this implies difficulties in identifying areas of origin that are not too extensive in terms of territory, as well as greater variability that could make it difficult to identify patterns. A solution to this issue is provided in the next chapter, where multi-step preferential patterns are revealed through a different statistical method.

The chain migration effect, which is intended as a residual compared to the effects of push and pull factors, should be further supported by an analysis of students' motivations to move carried out at an individual level through a survey and/or in depth interviews. With these limitations in mind, some interesting results emerge. First, the hypothesis that many of the origin-destination patterns of mobility are powered by chain migration effects is supported by our data. Both Sicily and Sardinia show patterns of mobility that are different depending on the field of study: science VS humanities. Instead, Apulia tends to replicate the same patterns of mobility for both fields of study. This difference between Apulia and the Islands is likely driven by factors related to the transportation network that connects Apulia to the Centre-North of the country, which is rather different than the one available in the Islands. The ease of moving from Apulia to a destination in the Centre-North lowers the overall moving barrier, but, at the same time, contributes to shape the preferential mobility patterns, partially hiding chain migration effects.

Another noteworthy aspect is the applicability of a "medium-large city to large city" and "medium-small city to medium-small city" model to student mobility data, especially for Sardinia and Sicily. In fact, students from Palermo, Catania and Messina and those from Cagliari tend to move to the largest cities in the Centre-North of the country, namely, Milan and Rome, whereas students from small cities tend to prefer universities located in small-medium cities.

Chapter 3

A network analysis of multi-step student mobility patterns: from high school to master's degree

Abstract

Human migration involves the movement of people from one place to another. An example of undirected migration is Italian student mobility where students move from the South to the Center-North. This kind of mobility has become of general interest, and this work explores student mobility from Sicily towards universities outside the island. The data used in this chapter regards six cohorts of students, from 2008/09 to 2013/14. In particular, our goal is to study the 3-step migration path: the area of origin (Sicilian provinces), the regional university for the bachelor's degree, and the regional university for the master's. Our analysis is conducted by building a multipartite network with four sets of nodes: students; Sicilian provinces; bachelor region of studies; and the master region of studies. By projecting the students' set onto the others, we obtain a tripartite network where the number of students represents the link weight. Results show that the big Sicilian cities–Palermo, Catania, and Messina–have different preferential paths compared to small Sicilian cities. Furthermore, the results reveal preferential paths of 3-step mobility that only, in part, reflect a south-north orientation in the transition from the region of study for the bachelor degree to that for the master's.

3.1 Introduction

Human and cultural capital, including education and qualification, can be considered a crucial resource for development and innovation. Educational mobility plays a role in the development of a country [1]. In the words of the European Higher Education Council *"learning mobility is widely considered to contribute to enhancing the employability of young people through the acquisition of key skills and competences, including especially language competences and intercultural understanding, but also social and civil skills, entrepreneurship, problem-solving skills and creativity in general"*. We can distinguish two types of student mobility: credit mobility [18, 10, 60, 51, 17, 92], and degree mobility. The former refers mainly to Erasmus students, the latter refers to those students that decide to study outside their home region. In this work, we deal with the domestic degree mobility discussed in literature by many authors such as Barrioluengo *et al.* [8], Dolinska *et al.* [34], and Van Bouwel *et al.* [15]. In particular, those authors showed how domestic mobility is inhomogeneous within countries. They note that only some universities inside a given country are much in demand from students (*e.g.* Cyprus, Hungary, Lithuania, and Poland) [8, 34]. Van Bouwel *et al.* in [15] focuses on two different perspectives: the "consumption perspective" which is not strictly related to universities prestige or educational quality, but, rather, to the urban services; and the "investment perspective" which is related to a university's prestige. In the UK there is an important literature on student mobility: there are upper class "prestige" universities, such as Oxford and Cambridge, and some other London universities. These universities are attractive both due to their prestige and the urban services in the cities where they are based.

Thinking now of Italy, in the last twenty years, Italian universities have been through significant changes in terms of student flows and mobility, as well as university governance. Student enrolment has decreased significantly, especially after the economic crisis of 2008, with consistent recovery in the last five to six years. On the other hand, student mobility from the South to the Center and North of the country has increased, 30% of students living in the South decided to enrol in universities in the Center-North in 2017 [4]. Boscaino *et al.* in [14] argue that this kind of mobility can be affected by better job-market opportunities in the Center-North, so it looks as if university quality is less important. Furthermore, Santelli *et al.* in [99] showed how southern regions are affected by increasing rates of outward bound students–especially in Sicily–arguing that mobility to the North is driven by job-market opportunities at destination. D'Agostino *et al.* [29] and Impicciatore *et al.*

[57] note how this mobility is also affected by contextual factors such as students' social class and family background. These findings suggest the Italian mobility is in line with that of students found in the international literature [40, 83, 5].

Italian student mobility has been mainly investigated with aggregate administrative data released by the Ministry of Education (Miur), for different territorial levels. Some studies have focused on the structural determinants of student mobility and examined aggregate migration flows of students across Italian provinces, with macroeconomic models [43, 98]. On the other hand, individual determinants of internal mobility have been studied mainly by exploiting survey data on secondary school graduates and university graduates, focusing on the role played by individual factors such as family background and schooling career [84, 26, 58]. Recently, the Ministry of Education released micro-data students' database careers (Database MOBYSU [79]). This new individual dataset has fostered several Italian students' mobility analyses, concerning general and specific issues [3, 48]. Most of the concern has been devoted to mobility at undergraduate level. This represents the "usual" choice for a student who attends a university faraway from his/her region, while less attention was devoted to mobility between the undergraduate and the master's level [39, 28]. Indeed, the aim of this chapter is to look at this "second term" mobility, which has not been studied thoroughly because it is more difficult to analyse: single universities do not record the master's of their undergraduates when these move to another university.

This chapter is connected to another article [47] where the analysis was focused on the *HStoBA*: high school province – Bachelor's university region trajectory, while now our focus is on the whole path *HStoBA* and *BAtoMA*: high school province – Bachelor's university region – Master's university region. As in the previous paper, we apply a network methodology and limit our study to Sicilian first-years at university, even if it would be possible to extend the study to other geographical areas. The novelty of this analysis is twofold: first, it provides an overall network analysis with a triadic perspective over the entire university experience; and second, it offers the generalization of the test for link flow overexpression to a 3-mode motif (patterns).

3.2 Data and aims

3.2.1 Premise

This section introduces some definitions and quantities that will be used throughout the chapter. Specifically, we define:

- *MoversBA*: Sicilian first-years enrolled in a non-Sicilian university;
- *StayersBA*: Sicilian first-years enrolled in a Sicilian university;
- *MoversMA*: Sicilian BAs enrolled in a non-Sicilian MA course, who received a BA degree from a Sicilian university;
- *StayersMA*: Sicilian BAs enrolled in a Sicilian MA course, who received a BA degree from a Sicilian university;
- *MMoversMA*: Sicilian students who got both their BA and MA degrees outside Sicily;
- *BackMA*: Sicilian students who got a BA degree outside Sicily but the MA degree in Sicily;
- *Others*: Sicilian students who do not belong to the previous groups.

Another definition concerns the transition *HStoBA*, where the "origin" is the province (there are nine provinces in Sicily): that is where the high school of the student is located, while the "destination" is the region of the BA university (we selected just the most popular destinations *e.g.* Lombardy, Lazio, Emilia-Romagna, Piedmont *etc.*). Similarly, the transition *BAtoMA* indicates the regions where the BA university and the MA university are located.

3.2.2 The data

Using the MOBYSU database described in section **??**, we restrict our analysis to students who attended a Sicilian high school, who enrolled in a BA course in an Italian university in the period from 2008 to 2013 and who enrolled in an MA course within five years of first enrollment. We excluded online universities. The first cohort, in 2008, is the year of the current university structure, whereas the last cohort, 2013, is the year that allows us to follow up on students' university progress for five years (Figure 3.1). Because of the low number of students in each cohort and assuming negligible variations over time, we aggregated the cohorts into two 3-year periods, namely, one from 2008 to 2010 (1^{st} 3-year), and the other one from 2011 to 2013 (2^{nd} 3-year). This was a trade-off between having enough observations and the dynamism of the series to be analysed. In our study, we considered only the variables necessary to perform the network analysis.

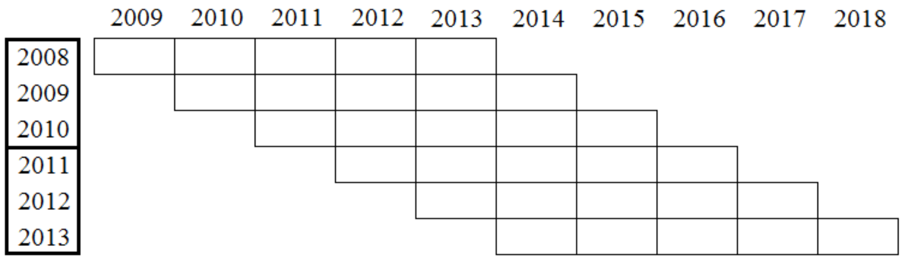


Figure 3.1: Data structure divided into two 3-year groups. The observation time is five years since the first enrolment for each cohort.

3.2.3 Aims

Our main goal has been to analyse student mobility from Sicily to the rest of Italy through the transition *BAtoMA*, by taking into account the previous path, that is, *HStoBA*. Our intention is to be able to answer questions such as: How many students migrate and how many stay? Does mobility change across the Sicilian provinces? Is there a common geographical pattern in the transition *HStoBA* and *BAtoMA*? Is there a special attraction between some provinces and regions? Is that pattern stable over time? As already noted, those questions are related to previous work [47] concerning the HStoBA transition, but with a larger focus, since we are now looking at the whole higher education path.

3.3 Descriptive statistics

To analyse the percentage of movers in the transition *HStoBA* and in the transition *BAtoMA*, the quantities in equations (1) and (2) were used:

$$R(HStoBA(i)) = \frac{MoversBA(i)}{MoversBA(i) + StayersBA(i)} \times 100; \tag{3.1}$$

$$R(BAtoMA(i)) = \frac{MoversMA(i)}{MoversMA(i) + StayersMA(i)} \times 100. \tag{3.2}$$

Where i is the 3-year period grouped as:

- 1^{st} 3-year period (2008/09, 2009/10 and 2010/11), for $i = 1$;

- 2^{nd} 3-year period (2011/12, 2012/13 and 2013/14), for $i = 2$.

Figure 3.2 shows $R(HStoBA)$ mobility percentages for the two 3-year periods.

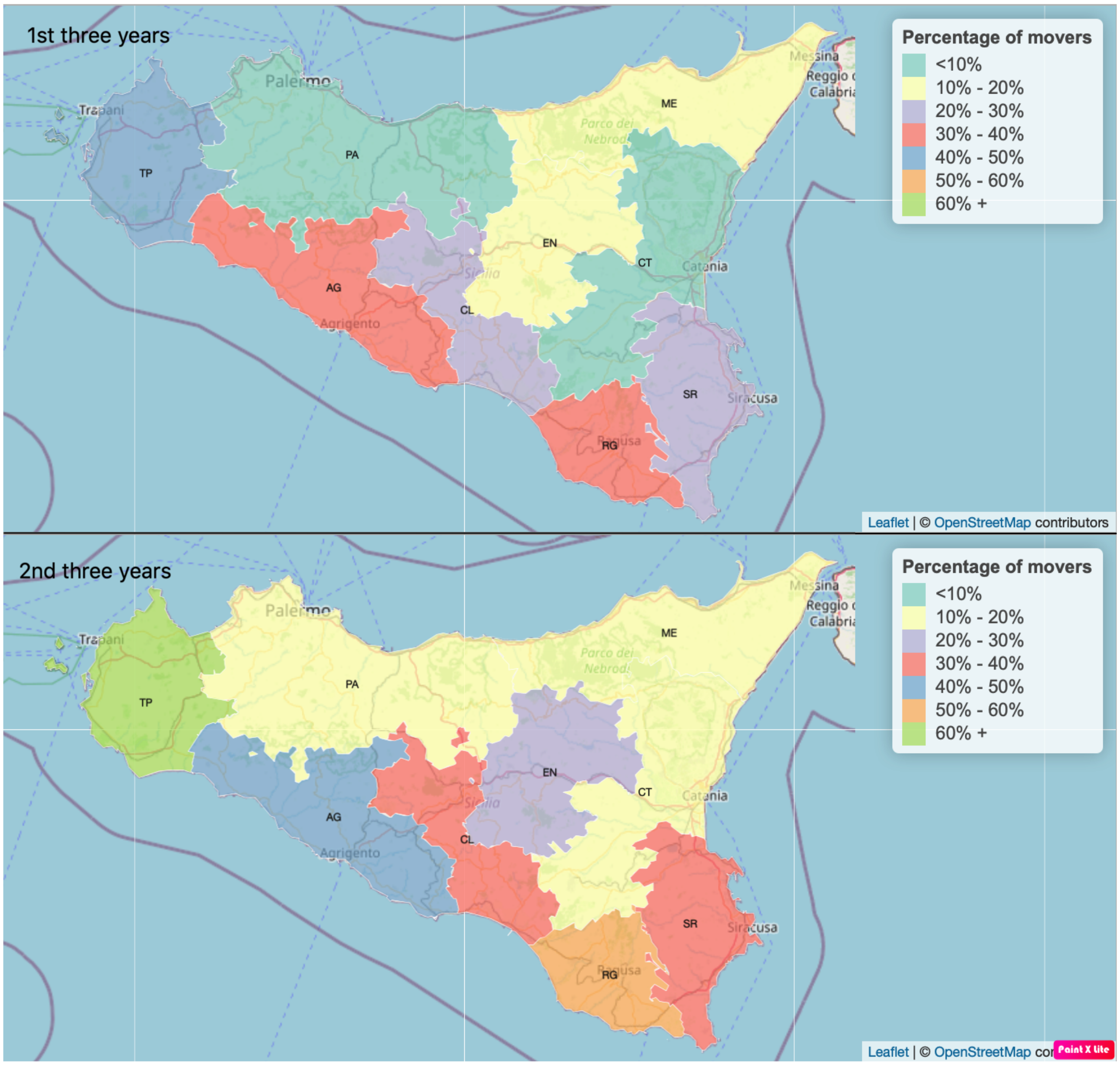


Figure 3.2: Percentage of outgoing students at BA, 1^{st} three year period (top panel) and 2^{nd} three year period (bottom panel).

They were 18.0% in all Sicily in 2008-2010 and 25.1% in 2011-2013. As expected, lower rates occur in provinces where there are universities and the number of movers increased over time in all the provinces, save Caltanisetta and Enna, which saw a slight decrease. Figure 3.3 shows that $R(BAtoMA)$ mobility percentages are much higher than the $R(HStoBA)$ ones: for all Sicily, there were 33.4% in 2008-2010 and 35.8% in 2011-2013. The difference between the two periods is smaller than 5% in all the provinces, save Messina, where there is an 11% increase in the mobility rate. Actually, these rates show a stable and dramatic detachment from Sicilian universities, as the number of incoming students from other regions in both degree levels is negligible.

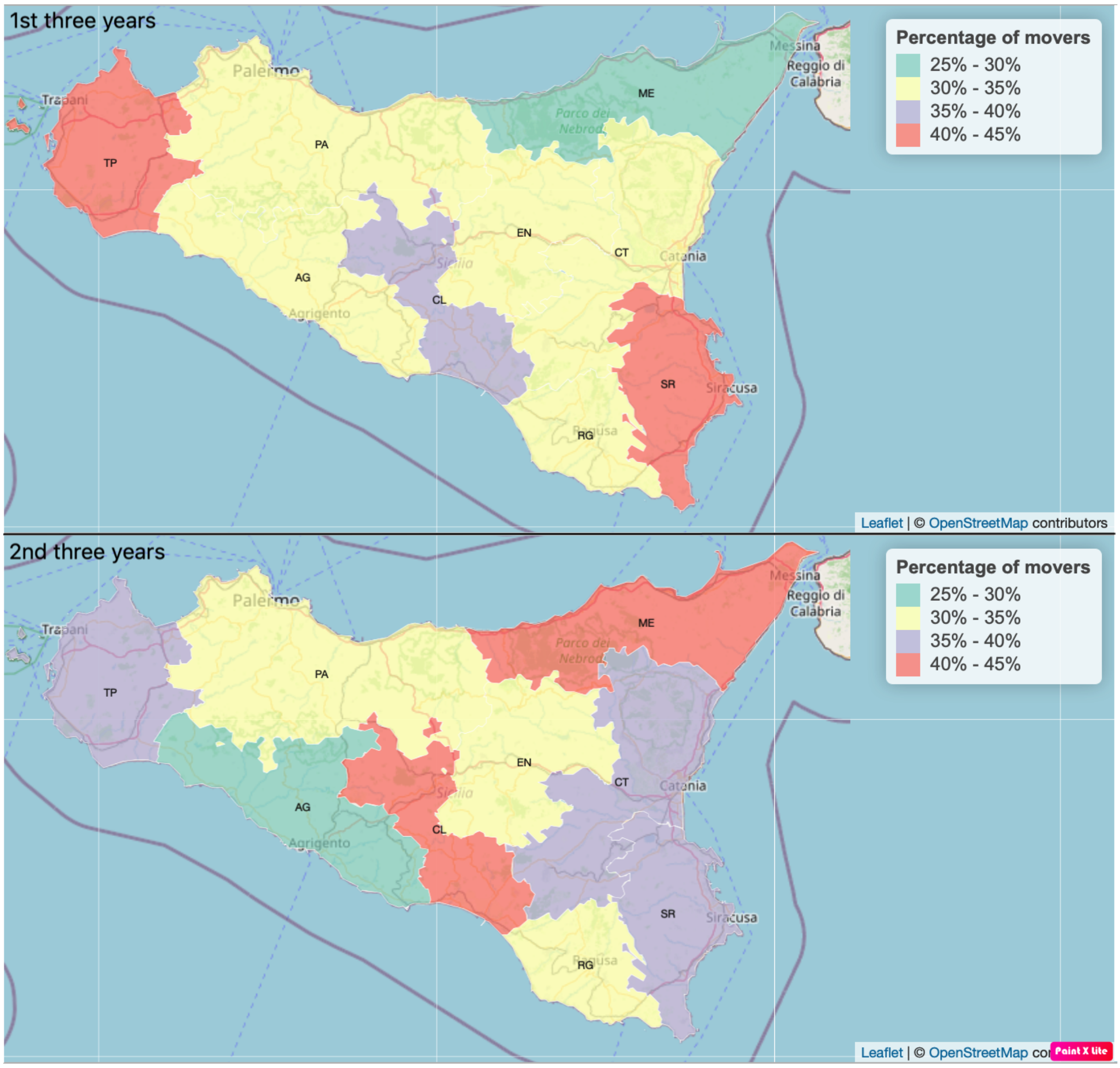


Figure 3.3: Percentage of outgoing students at the MA, 1^{st} three year period (top panel) and 2^{nd} three year period (bottom panel).

To analyse all the trajectories *HStoBAtoMA* of Sicilian students, according to their status, as previously defined, we constructed Table 3.1, Table 3.2 and Figures 3.4, 3.5, and 3.6.

In Table 3.1 *MoversBA* increased from 18.0% to 25.8% from 2008-10 to 2011-13, while *MoversMA* increased from 19.9% to 22.4% and *MMoversMA* increased from 51.2% to 52.4%. It is worth noting that *MoversBA* increases while *MMoversMA* does not change. The *BackMA* group is negligible in both periods and the *Others* group does not change.

Table 3.2 reports MA students according to their status: there are about 25% more females than males but males are much more likely to move outside Sicily. In both groups the number of movers increased in the second period. Few students return to Sicily to attend a MA course in the groups.

Table 3.1: Absolute values and *percentages* of Sicilian students in the trajectories *HStoBA* and *BAtoMA*, according to the their status (see subsection 3.2.1).

BA enrolled	2008–2010	StayersBA	46164 (*82*%)	BA	25322 (*54.8*%)	StayersMA	10999 (*43.4*%)
						MoversMA	5031 (*19.9*%)
						Others	9292 (*36.7*%)
				Others	20841 (*45.2*%)		
		MoversBA	10141 (*18*%)	BA	7069 (*69.7*%)	BackMA	357 (*5.1*%)
						MMoversMA	3620 (*51.2*%)
						Others	3092 (*43.7*%)
				Others	3072 (*30.3*%)		
	2011–2013	StayersBA	36874 (*74.2*%)	BA	20015 (*54.3*%)	StayersMA	8817 (*44.1*%)
						MoversMA	4492 (*22.4*%)
						Others	6706 (*33.6*%)
				Others	16859 (*45.7*%)		
		MoversBA	12823 (*25.8*%)	BA	8724 (*68*%)	BackMA	393 (*4.5*%)
						MMoversMA	4573 (*52.4*%)
						Others	3758 (*43.1*%)
				Others	4099 (*32*%)		

Table 3.2: *BAtoMA* mobility status by gender and cohorts.

3-year period	Gender	StayerMA	MoverMA	MMoverMA	BackMA	Total
2008-10	Female	6985	2820	1897	209	11911
	(%)	*(58.6)*	*(23.7)*	*(15.9)*	*(1.8)*	*(100.0)*
	Male	4014	2211	1723	148	8096
	(%)	*(49.6)*	*(27.3)*	*(21.3)*	*(1.8)*	*(100.0)*
2011-13	Female	5530	2533	2367	255	10685
	(%)	*(51.8)*	*(23.7)*	*(22.2)*	*(2.4)*	*(100.0)*
	Male	3287	1959	2206	138	7590
	(%)	*(43.3)*	*(25.8)*	*(29.1)*	*(1.8)*	*(100.0)*

Figures 3.4, 3.5, and 3.6 include a new variable *Presence of university* (PresUniv), which splits the Sicilian provinces into two groups. The first group has Palermo, Catania, Messina and Enna. The other has the remaining provinces. All the blank rectangles, as well as the bands, are proportional to their total. For the sake of brevity, we report just for the *HStoBA* for the 2011-2013 period, because the differences are negligible between the two periods.

The Figure 3.4 concerns *HStoBA*. Each figure is double, one for females (there are 17,223) and the other for males (there are 11,576). As already noted, females are less likely to leave Sicily. In fact the female-male mobility rates are, respectively, 27.5% and 34.1%. Moreover, males and females are differently distributed with respect to their field of study. Males prefer scientific areas, while females prefer all other areas.

Figures 3.5 and 3.6 show the *BAtoMA* transition, analysed for both periods because there are some differences. Males and females are differently distributed with respect to their field of study, following the usual preferences.

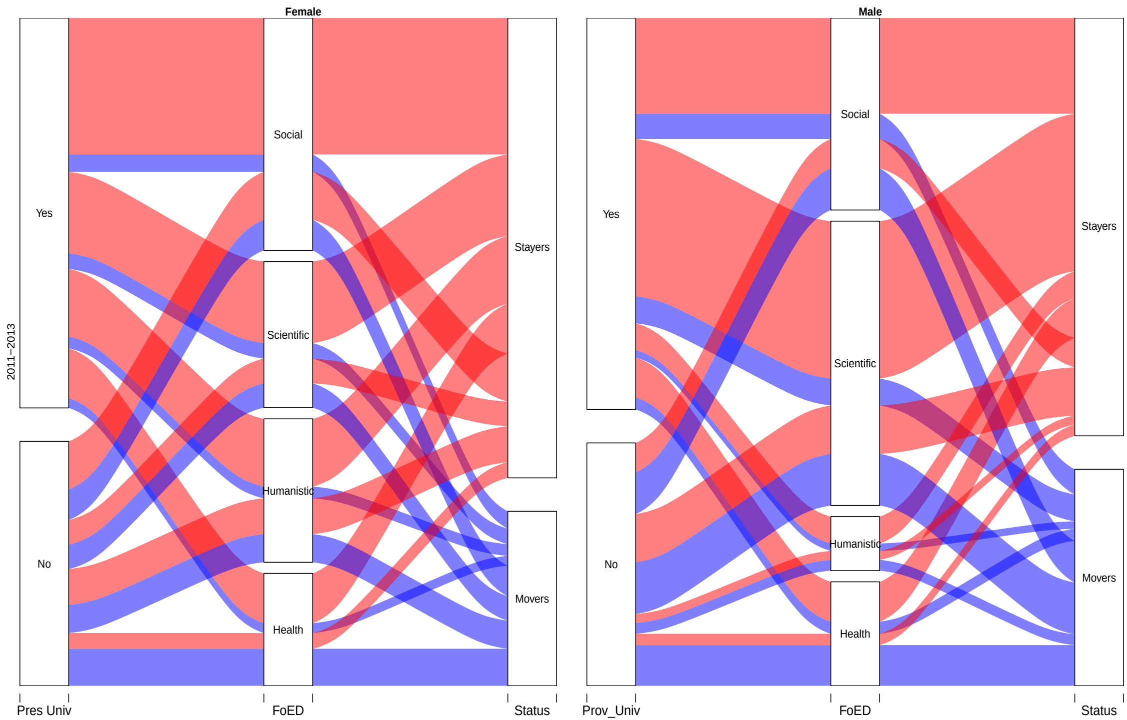


Figure 3.4: *HStoBA* mobility status by gender, presence of a university, and field of study, for the cohorts 2011-2013.

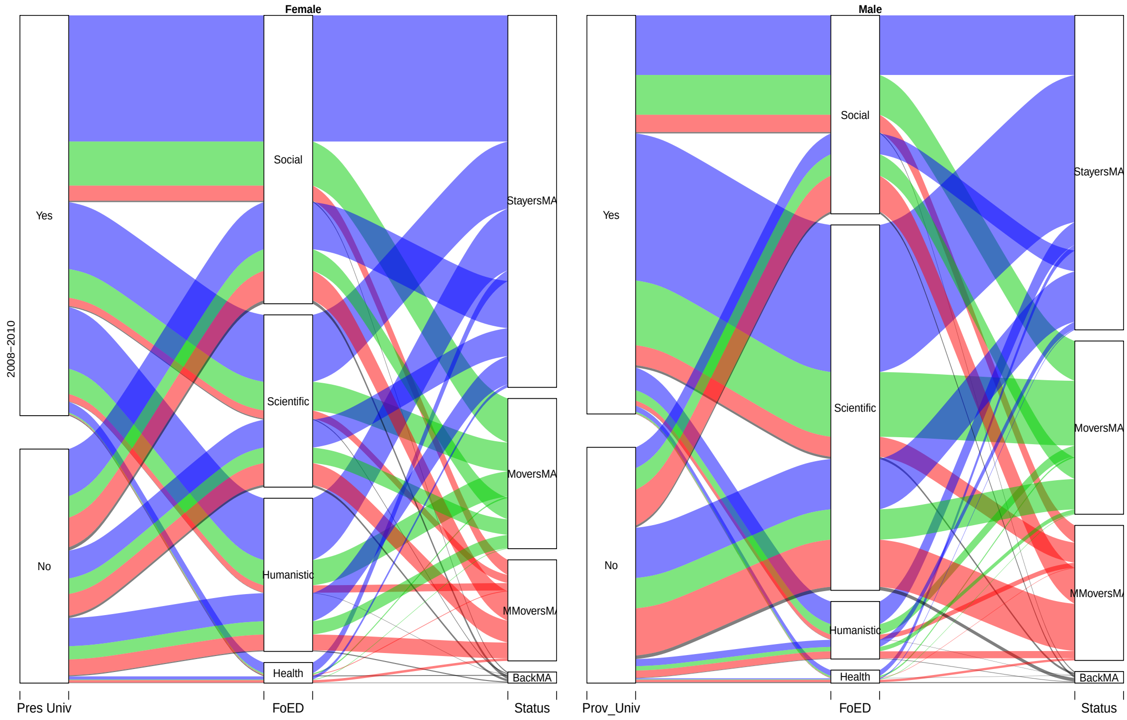


Figure 3.5: *BAtoMA* mobility status by gender, presence of a university, and field of study, for the cohorts 2008-2010.

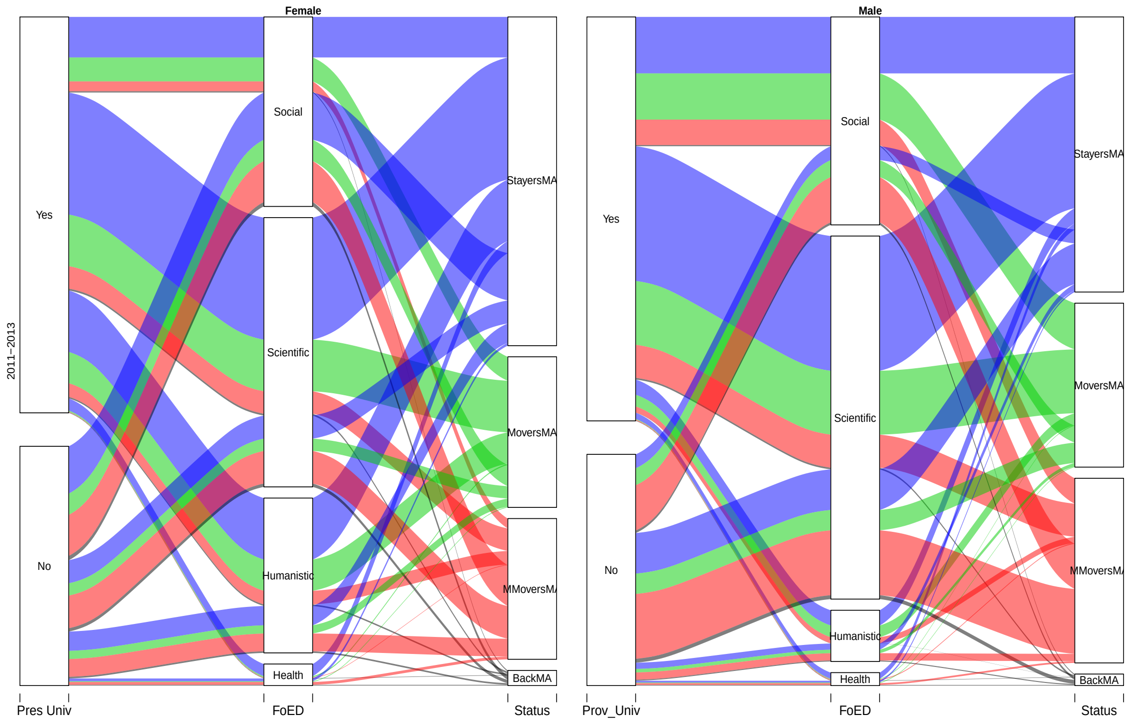


Figure 3.6: *BAtoMA* mobility status by gender, presence of a university, and field of study, for the cohorts 2011-2013.

3.4 Methods

3.4.1 Network structure

Fararo *et al.* [44] generalised Wilson's bipartite network representation [115] to a *restricted tripartite representation*. This representation is based on the following two axioms:

1. *there are three types of nodes*;

2. *ties exist only between nodes of different types.*

In our work we propose a generalisation of this *restricted tripartite representation* by introducing a quadripartite representation with four sets of nodes (*"groups"*), namely, *Students* (S_1), *Provinces of origin* (S_2), *Bachelor region* (S_3), and *Master region* (S_4). Besides a straightforward generalisation of the axioms proposed by Fararo *et al.* [44],

1. *there are four types of nodes*;

2. *ties exist only between nodes of different types*;

we add the following axiom:

3. all ties involve one node of type S_1.

Figure 3.7 reports a schematic representation of the proposed network structure. It is worth noting that the "students" set of nodes is homogeneous by degree–which is equal to three. As explained in subsection 3.2.2, the present analysis considers only students coming from a Sicilian province, who enrolled in both a bachelor course (BA) and a master course (MA) in Italy. This kind of homogeneity guarantees the suitability of the test proposed in subsection 3.4.2 on the projected tripartite network reported in Figure 3.8.

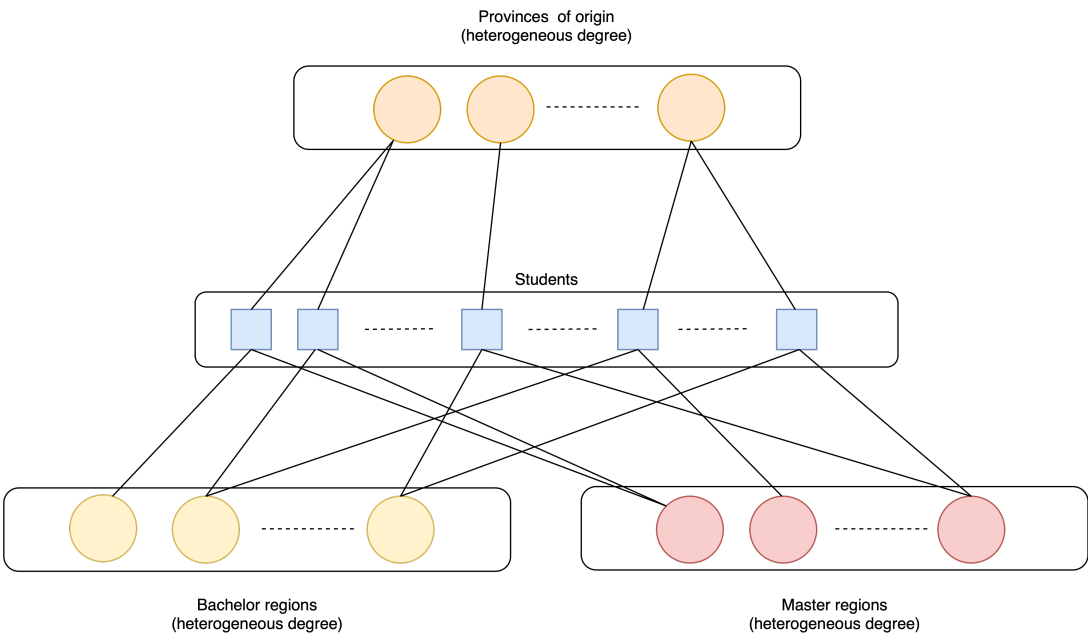


Figure 3.7: An example of the Sicilian student mobility network structure before the projection.

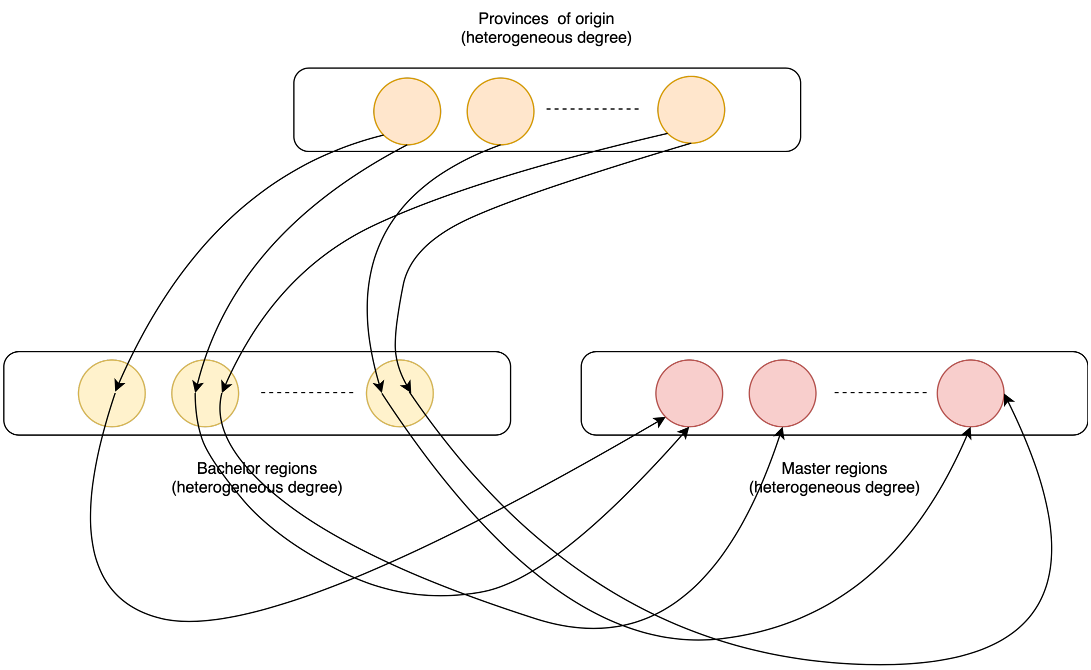


Figure 3.8: An example of the Sicilian student mobility network structure after the projection.

The weight of each tie in the projected network (Figure 3.8) is equal to the number of students that move from a province of origin to an MA region of studies passing through a BA region of studies. Thus, each flow represents a unique path from the province of origin to the MA region of studies and not a juxtaposition of the two transitions from provinces of origin to the BA region and from the BA region to the MA region. Indeed, unlike the work by Columbu *et al.* [28] and with our previous work [47], where ties between each couple of nodes were tested, here we test the flow of students for each three-step pattern against the null hypothesis presented in the 3.4.2 subsection. The main reason to use this joint approach, instead of separately analysing *HStoBA* and *BAtoMA* patterns, is that our approach can identify significant joint flows *HStoBAtoMA* even in cases where either or both separate *HStoBA* and *BAtoMA* flows are not statistically significant. This possibility is important for investigating possible chain-migration patterns of mobility since the two transitions involve different decision processes. Indeed, a preferential pattern *HStoBA* may reflect mostly family decisions, whereas a preferential pattern *BAtoMA* reflects more individual and/or peer influenced decisions. Therefore, a simple juxtaposition of the networks describing the two transitions would not provide insights about chain-migration effects from HS through BA to MA because the processes underlying the two transitions are different. Chain migration effects in the first transition (*HStoBA*) were already studied in ref. [47]. Several studies on human mobility, *e.g.*, ref.s [17, 92, 99], emphasise the role of regions and/or countries as good "receivers" and "senders" through network metrics and indexes such as hub and authority indexes. Here, instead, we focus on mobility patterns and analyse the deviations of the observed flows from a null hypothesis of random flows. In other words, our analysis deals with patterns, not with nodes, and our aim is to reveal deviations from random flows, which account for the heterogeneity of nodal inflows and outflows, instead of highlighting the backbone of the mobility network. This approach may help to untangle the role played by the attractiveness (and unattractiveness) of nodes and the chain-migration effects in shaping student mobility patterns. Our null hypothesis incorporates the actual inflows and outflows of the origin provinces and the destination regions. Therefore, statistically significant deviations from the proposed null hypothesis may highlight preferential patterns of mobility reflecting chain-migration effects.

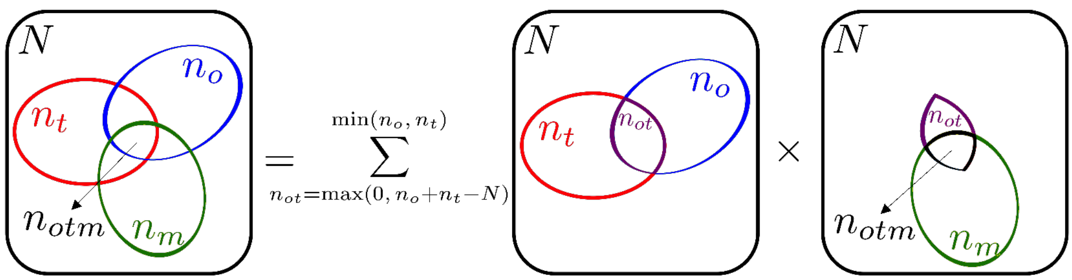
Table 3.3 reports some descriptive statistics of the projected tripartite network, before the filtering procedure illustrated in section 3.4.2 is applied.

Table 3.3: Synoptic table of the unfiltered tripartite network. We report the following statistics for both the 3-year periods under analysis: the number of nodes per set of nodes, the number of triplets, the total number of students moving from origin o ($\sum n_o$), the total number of students enrolled at a university of region t in an undergraduate program ($\sum n_t$), and the total number of students enrolled at a university of region m for their graduate studies ($\sum n_m$).

		Period	
		1^{st} 3-year	2^{nd} 3-year
N. of nodes by sets	Provinces	9	9
	BA regions	15	17
	MA regions	15	16
Triplets		300	430
$\sum n_o$		600	899
$\sum n_t$		586	893
$\sum n_m$		578	883

3.4.2 Hypothesis testing

We aim to reveal the three-step preferential patterns of students' mobility from their Sicilian province of origin through the region chosen for their undergraduate studies (BA region), to the region of their graduate studies (MA region). To accomplish this goal, we first define the concept of random mobility as a flow of students across province of origin, BA region, and MA region, which occurs uniformly under the constraints determined by the empirical marginal values of the inflow and outflow of each one of these nodes. The operationalization of this concept provides the null hypothesis H_0 against which actual flows are tested through the probability scheme reported in Figure 3.9 and in Eq. (3.3):



$$P(n_{otm}|n_o,n_t,n_m,N) = \sum_{n_{ot}=\max(0,\,n_o+n_t-N)}^{\min(n_o,\,n_t)} \frac{\binom{n_o}{n_{ot}}\binom{N-n_o}{n_t-n_{ot}}}{\binom{N}{n_t}} \times \frac{\binom{n_{ot}}{n_{otm}}\binom{N-n_{ot}}{n_m-n_{otm}}}{\binom{N}{n_m}}$$

Figure 3.9: Construction scheme of the probability mass function reported in Eq.(3.3)

$$P(n_{otm}|n_o, n_t, n_m, N) =$$

$$= \sum_{n_{ot}=\max(0,n_0+n_t-N)}^{\min(n_o,n_t)} P(n_{otm}|n_{ot}, n_m, N)P(n_{ot}|n_o, n_t, N) =$$

$$= \sum_{n_{ot}=\max(0,n_0+n_t-N)}^{\min(n_o,n_t)} H(n_{otm}|n_{ot}, n_m, N)H(n_{ot}|n_o, n_t, N) =$$

$$= \sum_{n_{ot}=\max(0,n_0+n_t-N)}^{\min(n_o,n_t)} \frac{\binom{n_{ot}}{n_{otm}}\binom{N-n_{ot}}{n_m-n_{otm}}}{\binom{N}{n_m}} \frac{\binom{n_o}{n_{ot}}\binom{N-n_o}{n_t-n_{ot}}}{\binom{N}{n_t}}, \quad (3.3)$$

where:

- n_{otm} = number of students moving from origin o through university region t for an undergraduate degree to university region m for graduate studies (flow *HStoBAtoMA*);
- n_{ot} = number of students moving from origin o to university region t for an undergraduate degree (flow *HStoBA*);
- n_{tm} = number of students moving from university region t for an undergraduate degree to university region m for graduate studies (flow *BAtoMA*);
- n_o = number of students moving from origin o;
- n_t = number of students enrolled at a university of region t for an undergraduate degree;
- n_m = number of students enrolled at a university of region m for their graduate studies;
- N = number of students moving from Sicily to a university in another region.

The probability mass function reported in Eq.(3.3), which represents the operationalization of the null hypothesis H_0, can be easily obtained under the assumption of random flows, by noting that $P(n_{ot}|n_o, n_t, N)$ is provided by a hypergeometric distribution with parameters n_o, n_t, and N, and the same goes for $P(n_{otm}|n_{ot}, n_m, N)$ given n_{ot}, which is given by a hypergeometric distribution [95] with parameters n_{ot}, n_m, and N. Furthermore, Eq.(3.3) allows one to associate a *p-value* with any observed value of variable n_{otm}, say q, as follows:

$$p - value(q|n_o, n_t, n_m, N) =$$

$$= \sum_{n_{otm}=q}^{\min(n_o,n_t,n_m)} \sum_{n_{ot}=n_{otm}}^{\min(n_o,n_t)} \frac{\binom{n_{ot}}{n_{otm}}\binom{N-n_{ot}}{n_m-n_{otm}}}{\binom{N}{n_m}} \frac{\binom{n_o}{n_{ot}}\binom{N-n_o}{n_t-n_{ot}}}{\binom{N}{n_t}}. \qquad (3.4)$$

It is worth noting that, given the presence of the sum over n_{ot}, a low *p-value*, according to Eq.(3.4), does not merely reflect an excess of the values of n_{ot}, or n_{tm}. In other words, it is possible to obtain a low *p-value* associated with observation q, even if the observed values of n_{ot} and n_{tm} are such that their individual right tail *p-values*, according to the standard model of Statistically Validated Networks (SVN) [108], are large enough that the hypothesis of (marginal) random flow cannot be rejected. Therefore, the model presented in Eq.(3.3) represents a genuine generalization of the SVN method to detect over-represented three-node motifs [56], in spite of the representation of (marginal) two-mode motifs, that is, links, in the system [47]. Furthermore, our method is able to highlight significant patterns through vertices with low weights since the test exactly takes into account the heterogeneity of all three nodes involved in each pattern. This possibility represents a clear difference in the proposed method with respect to classical network filtering methods, which are generally based on the application of a threshold on the link weight. Indeed, unlike a backbone extraction with unconditional threshold–where the network structure strongly depends on the chosen threshold [80]–our method preserves structural and multi-scale features of the network controlling for the number of students at origin (n_o), the number of students in the BA region (n_t), the number of students in the MA region (n_m), and all the students involved in the system (N). In the study of migration flows the *bistochastic filter* was proposed as a backbone extraction of the network, but, as reported in [45], this method may alter the network structure due to the edge weight matrix manipulation. Similarly to the *bistochastic filter*, Yongwan *et al.* in [25] proposed the application of an eigenvector spatial filtering procedure in migration flows as a backbone extraction, but again this kind of procedure requires non-trivial data manipulation. In our view, the method proposed in Eq. (3.3) is an unsupervised and data-driven way to evaluate statistically significant links between nodes. Furthermore, it is important to stress that our method does not aim to reveal the backbone structure of the mobility network. Rather it aims to reveal preferential patterns of mobility, *i.e.*, patterns that display a significant deviation in the observed flow from the null hypothesis of random flow.

3.4.3 Network construction

The projected network under investigation is a network with three sets of nodes, namely, the provinces of origin (set S_o), the BA regions (set S_t), and the MA regions (set S_m), as detailed in subsection 3.4.1. Furthermore, a directed link can only occur between a node of S_o and a node of S_t, and between a node of S_t and a node of S_m (see Figure 3.8). The weight of each link is equal to the number of students flowing between the two nodes. If we indicate the number of nodes in set S_i with $\#node_i$, $(i = o, t, m)$, then the number of performed tests of statistical significance, according to Eq.(3.4), is $T = \#node_o \cdot \#node_t \cdot \#node_m$. Therefore, we set a univariate statistical threshold at 5% and correct *p-values* through the *Bonferroni* [78] correction for multiple hypothesis testing. This kind of correction guarantees that the family-wise error rate is controlled, in spite of any dependence among the tests. This property of the *Bonferroni* correction is particularly important in the present analysis, since two tested triplets can differ for only one node, which clearly introduces a dependence between the corresponding tests. Finally, for the sake of pattern readability, we pictorially represent the *Bonferroni* Statistically Validated Network by joining together the nodes of S_o and S_t in meta-nodes defined by a pair $(v_{o,i}, v_{t,j})$. That pair represents the first part of a statistically significant three-node motif involving node $v_{o,i}$ from S_o and node $v_{t,j}$ from S_t.

3.5 Results

As stated in subsection 3.2.3, the central goal of this work is to analyse students' mobility in the whole path from the provinces of origin, through the BA region, to the MA region. This analysis was carried out on six cohorts of students, from 2008 to 2013, that enrolled to the second level degree within five years of the first level enrolment. The work by Genova *et al.* [47] analysed the path from area of origin to bachelor universities, while in this chapter, we have analysed three-node subnetworks, in order to understand whether there are preferential patterns of mobility that involve both the transitions *HStoBA* and *BAtoMA*.

Figures 3.10 and 3.11 show the tripartite validated networks of outward bound Sicilian students with this new validation link procedure. For the sake of readability, instead of highlighting statistically significant three-node motifs *HStoBAtoMA*, we have merged the first two nodes for each patterns, that is, *HStoBA*. The solid lines represent links that are significant at level α=0.05 after applying the *Bonferroni* correction [78] for multiple hypothesis testing. Moreover, the thickness of any link is proportional to the number of students flowing from the merged nodes (blue nodes) to the MA regions (red node) and, the red circle size is proportional to the indegree.

Looking at the results, Figures 3.10 and 3.11 show that the most attractive regions, in the period under study, are Lombardy and Emilia-Romagna. The number of statistically validated links increases over time, in particular from 40 in the first period to 74 in the second. This increase is mainly due to the provinces of Agrigento, Ragusa, and Trapani (see Table 3.4). On the other hand, big cities, in particular those with universities, show a smaller number of connections. These results can be interpreted in terms of student behaviour. Students coming from small cities seem have more varied patterns than students from large cities.

Table 3.4: Number of links by province and period.

Provinces	1st-3years	2nd-3years	Change (%)
AG	3	11	267%
CL	2	6	200%
CT	2	5	150%
EN	0	1	–
ME	5	7	40%
PA	6	6	0%
RG	10	16	60%
SR	4	9	125%
TP	8	13	63%

It is important to stress the distribution of the "straightforward" triplets, that is, when the BA region is equal to the MA region, compared to the "non-straightforward" triplets. The first ones are common in Lombardy for both time periods, while the "non-straightforward" triplets are more common in Emilia-Romagna and Piedmont in both periods.

As noted before, the thickness of the links in Figure 3.10 and 3.11 is proportional to the number of students that move from the blue source nodes to the red target nodes. In the first 3-year period it is worth stressing that the biggest flows are the Trapani–Emilia-Romagna–Emilia-Romagna paths[1] and the Trapani–Emilia-Romagna–Lombardy paths. Moreover, the "straightforward" triplets with the biggest link thickness point to Lombardy. The "non-straightforward" triplets point to Emilia-Romagna. In the second period, such an effect still persists for these regions plus Piedmont. This result implies the main student mobility flows point to Lombardy, Emilia-Romagna, and Piedmont. Besides the usual south-to-north direction of mobility roots, Fig. 3.12 also shows roots with a different orientation. Specifically, the 3-step paths, which are stable across both cohorts, include: Trapani–Tuscany–Lazio; Trapani–Tuscany–Lombardy; Messina–Lombardy–Lazio; Ragusa–Emilia-Romagna–Piedmont; and Ragusa–Umbria–Tuscany. This result indicates that 3-step mobility patterns only partially reflect a south-to-north orientation in the transition from the BA to the MA region of study.

Finally, to evaluate which preferential paths persist over time, Figure 3.12 shows the intersection between the networks represented in Figures 3.10 and 3.11.

[1] These patterns cannot be trivially explained by the presence of students that enrol for the MA courses in the same universities where they earned the BA degree, since such students were excluded from the analysis.

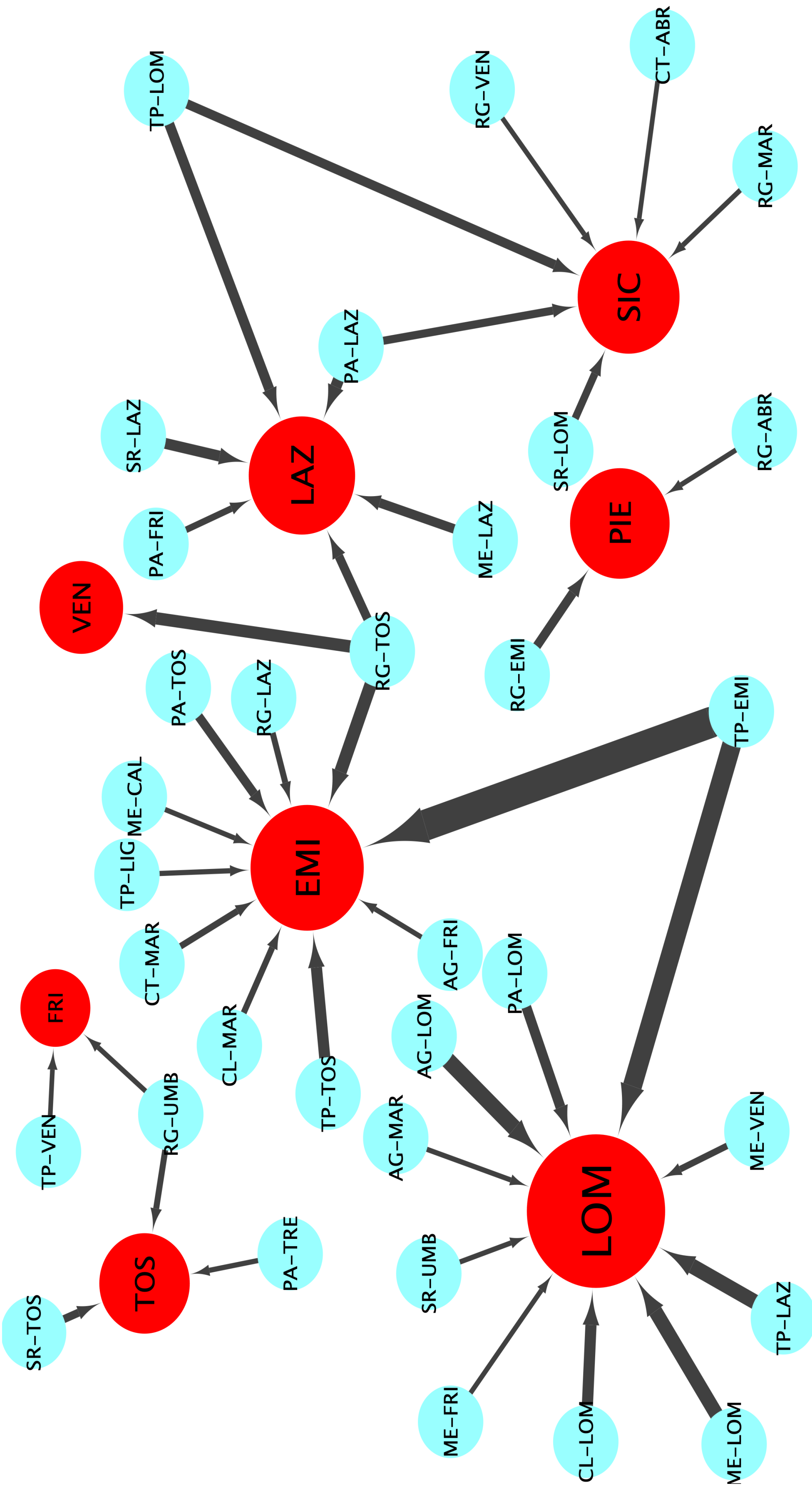


Figure 3.10: Statistically validated network from origin to MA region, first three year period. The red nodes represent the MA region, the blue ones the couple: province–BA region. The black solid lines are the statistically significant links, and their thickness is proportional to the students' flows.

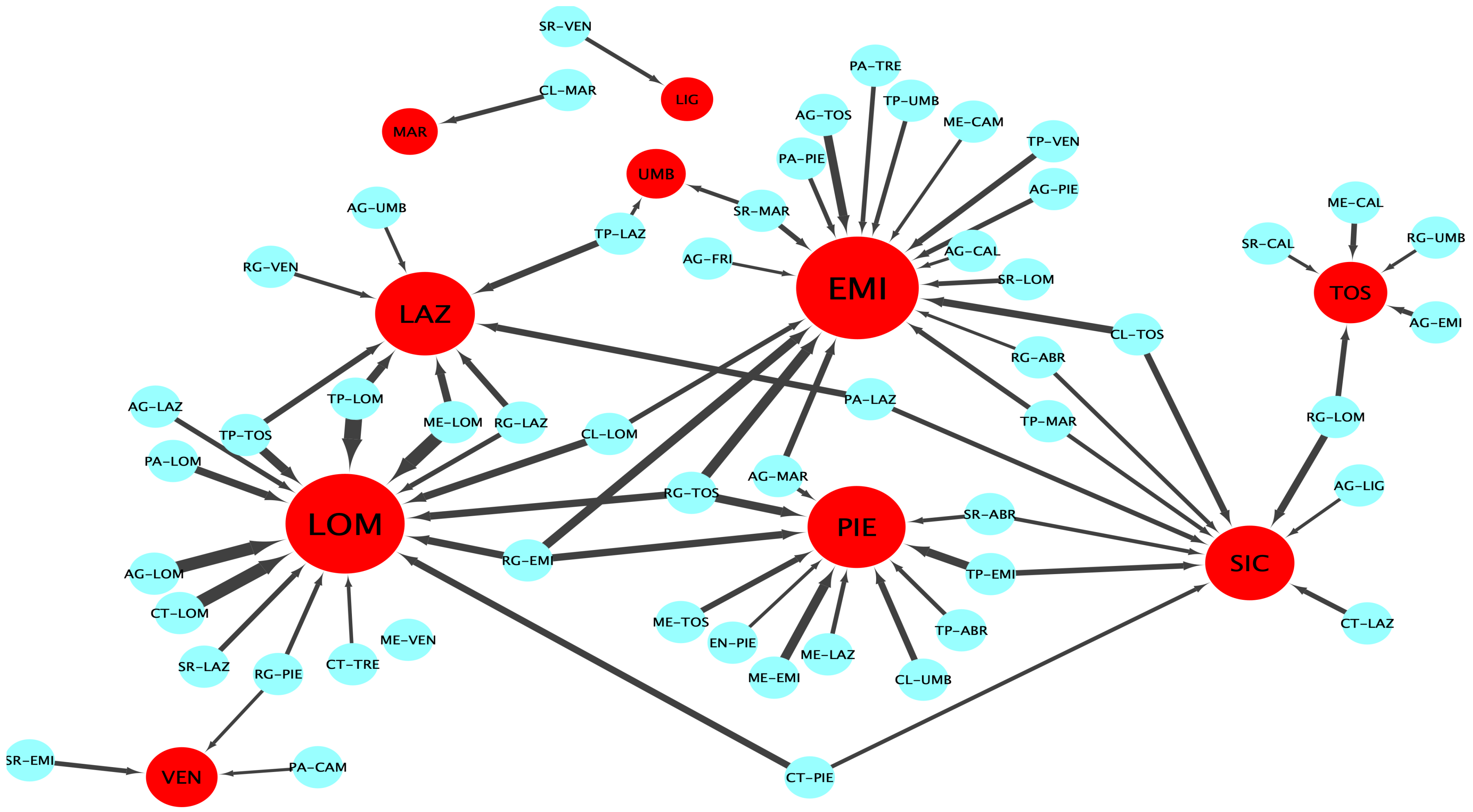


Figure 3.11: Statistically validated network from origin to MA region, second three year period. The red nodes represent the MA region, the blue ones the couple: province–BA region. The black solid lines are the statistically significant links, and their thickness is proportional to the students' flows.

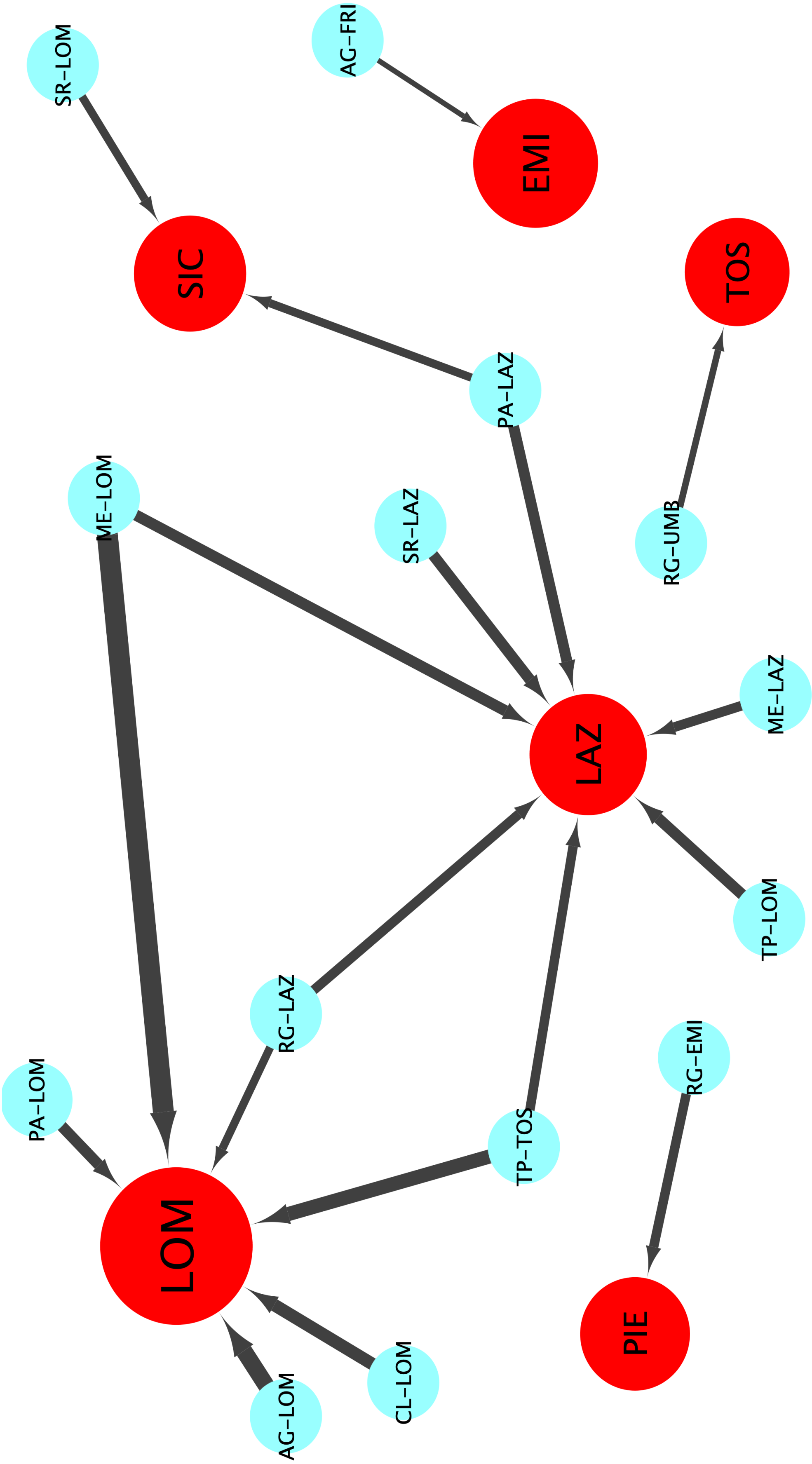


Figure 3.12: The intersection of the two statistically validated network in figures 3.10 and 3.11.

A comparative analysis of the networks associated with the two periods shows that 11 triplets persist through time, and that 15% of the first-period links are still present in the second period. In particular, the MA regions that persist through time are Lombardy, Lazio, Emilia-Romagna, Piedmont, and Tuscany.

3.6 Conclusions

We provided a comprehensive analysis of the three-node patterns of mobility of Sicilian students enrolled in the 2008-2013 period. These patterns were studied through a specifically devised statistical test of over-representation in a tripartite network. The test and the associated statistically validated networks represent a generalization of the SVN method [108] to tripartite networks. The empirical analysis of student mobility patterns indicates that students coming from small Sicilian cities show more varied patterns of mobility than students coming from large Sicilian cities. The revealed patterns, which are discussed in Section 3.5, turn out to be stable over the two analysed 3-year time periods. This supports the idea that such patterns not only depend on economic factors but also on socio/cultural aspects. Indeed, the second period considered in this study covered the sovereign debt crisis, and, in particular, 2013 was a black year for Italian household expenditure on goods and services. The prominent role of socio/cultural factors may also explain the presence of patterns that do not reflect the usual south-to-north orientation. Furthermore, the descriptive statistics showed that females are generally less inclined to move than males, and that the propensity to move also depends on the field of study. Indeed, males prefer scientific areas, whereas females prefer social and humanities studies. We plan, next, to investigate mobility patterns, by also taking into account both gender and field of study. This analysis, though, would require a larger sample size.

Conclusions

In this thesis, we have provided an empirical analysis of student mobility based on microdata of enrolment for several cohorts of Italian students. The main objective of the study was to reveal preferential patterns of mobility linked to chain migration effects. Indeed, we invoked the "paradigm" of chain migration to investigate South-to-North mobility patterns, unlike previous work on Italian student mobility that mostly considered mechanisms based on public information and contextual factors [4, 14, 99, 29, 57]. This is not straightforward, since student chain migration is different from classical chain migration–where movers share information with primary social contacts about labour market opportunities and/or quality of life at destination. Specifically, student chain migration also refers to the enrollment process at the university of destination–where primary social contacts already enrolled there *i)* provide broad information on the university and destination, *ii)* help newcomers during the enrollment process, and *iii)* provide concrete help on arrival [89].

As highlighted in the introduction to this thesis, there are just a few studies that consider chain migration and student mobility and they essentially use a qualitative approach [89, 88, 19]. For instance, Pérez and McDonough [88] study chain migration effects in student mobility through semi-structured interviews on a sample of students that are not representative of the population. Similarly, Brooks and Waters [19] use the same technique for bringing out the importance of social/family networks in the choice of studying in the UK. Finally, Person and Rosenbaum [89] combine semi-structured interviews with a survey conducted on students from 14 colleges. They conclude that a student with a friend or kinship network in a given university has a higher probability of enrolling in that same university.

To the best of our knowledge, no quantitative methods have been proposed in the literature for empirically identifying the patterns of student mobility favoured by chain migration. Thus, we proposed two quantitative methods of data analysis, which seem to reveal preferential patterns of student mobility, and we applied them to identify specific

South-to-North mobility patterns among Italian students.

Overall Contribution

The objective of this dissertation was to analyse student mobility from the South to the North of Italy and to untangle chain migration effects from push and pull factors. We analysed the transition from the high-school to bachelor enrolment in Chapter 1 and 2, whereas we analysed the whole preferential patterns of mobility from the high-school to master's in Chapter 3. In particular, we provided answers to the following research questions: is there a chain migration effect in student mobility? If so, what are the most significant mobility patterns? Are there similarities between mobility patterns from different areas of origin? And are mobility patterns stable over time?

To answer these research questions we proposed two methods of data analysis on the edge of Statistics and Network Theory. We applied these methods to three "big" case studies (Apulia, Sardinia and Sicily). We did so to identify patterns of student mobility that can be partially rooted to chain migration. Indeed, these methods reveal statistically significant extra-flows with respect to a null hypothesis that fully takes into account the intrinsic heterogeneity of the entities involved in the analysis (universities, regions, areas of origin, *etc.*). In particular, in Chapter 3, we developed a method for revealing statistically significant 3-mode motifs in a four-partite network. The method was applied to elicit from data (statistically significant) 3-step preferential patterns of mobility from the province of origin to the region of studies for the masters, through the region of bachelor studies. The main reason for using this joint approach is that it can identify significant joint flows even in cases where either or both separate flows are not statistically significant. In other words, a 3-step preferential pattern may occur, even if the two 2-step patterns that go to make it up are not significant.

Results in all cases indicate that student mobility is not merely a random process from the South to the North of Italy. Instead, student mobility is a strongly patterned process from the South to the North, where student choices are motivated by factors such as interpersonal relationships, private information, and strong and weak ties at destination [49, 62] *etc.*. These compete to shape migration patterns. Indeed, as suggested by Pearson and Rosenbaum in [89], students that received private information from trusted family members and friends (strong ties) are less motivated to consider enrolling in a university on the basis of public information conveyed by official channels, or from acquaintances (weak ties).

The network analysis of preferential patterns of mobility performed in Chapter 1 shows that, among the universities of central Italy, Pisa and Siena are (on average) the most attractive–in terms of preferential patterns of students coming from western and southern Sicily, whereas the Polytechnic University of Turin is the favourite northern university for the 2014 cohort of Sicilian students. In addition, the network analysis provides some elements for supporting the existence of migratory chains. This includes the presence of new links–which become significant in the networks of the most recent cohorts and that were not in the previous ones–and their thickness which also increased in the networks of the last cohort. Specifically, we observed an increase in the number of clusters in the network with movers and the number of off-Sicily universities chosen by the movers (preferential patterns). By investigating the similarity between the gender specific networks, we observe a stability of similarity over time. This result implies that, even if mobility patterns vary over time, such a variation occurs by keeping the similarity between gender specific patterns fairly constant. The case of Siena also suggests that gender specific patterns may both vary over time in a non-monotonous way (pick in 2011). About 40% of preferential links revealed in 2014 were already present in 2008. This implies that the role played by chain migration in explaining the mobility patterns of students is rather important.

Furthermore, results in Chapter 2 confirm the hypothesis that many of the origin-destination mobility patterns seem to be powered by chain effects. Indeed, the cluster analysis for Sicily confirms the idea that student migration chains can play an important role in explaining the patterns that emerge from the Tree-Maps. In particular, for the humanities, residuals are particularly high in both five-year periods. This highlights a consistently higher number of students than what we would expected on the basis of pull and push factors. In Chapter 2, we dealt with the "unusual" connections in detail between the clusters of origins and destinations, particularly respect to gender and field of study.

Theoretical implications

Migration network theory suggests that strong ties can facilitate migration flows in that they provide help to new migrants [62, 76]. In particular, strong ties are identified as links in a network where *"members know one another, interact on a routine basis and are privy to the same information regarding the social environment, including job opportunities"* [114]. On the other hand, scholars on network theory argue that weak ties, too, can facilitate access to information related to job market opportunities. Brut [21] argues that *"dense networks tend to convey redundant information, while weaker ties can be sources of new knowledge and resource"*. Thus strong and weak ties play a role in migration. Strong ties provide concrete help at the beginning of a migration process, weak ties reduce the redundancy in information coming in from a dense and close community of relatives and friends.

Keeping in mind how chain migration is influenced by weak and strong ties, it is worth highlighting how appropriate a holistic approach, such as network analysis, is for investigating the chain migration process. Indeed, the migration process is intrinsically a complex phenomenon, one that is influenced by several endogenous and exogenous factors– *e.g.*, country development; access to information and education; and communication and transport development. Thus instead of modeling the process through linear models– that disregard the non linear dependence on endogenous and exogenous factors–holistic approaches oriented to path discovery can help us to reveal preferential mobility patterns, without the need to explicitly model chain migration, or weak and strong tie dependence.

Furthermore, looking at the wider mobility pattern from high-school to masters, results show that there are roots, besides the usual south-to-north mobility roots, with different directionality (*i.e.* Trapani–Tuscany–Lazio and Messina–Lombardy–Lazio). This result indicates that 3-step mobility patterns only partially reflect south-to-north orientation in the transition from the bachelor to the master region of study. Many factors can influence this change of direction. For instance, a possible explanation may be provided by considering the theory of weak and strong ties [62, 114, 76, 21]. It is possible that strong ties (families and close friends) shape the transition from high-school to a bachelor's: a student decides to move from the area of origin to another region for their bachelor, guided by relatives and friends that can concretely help her/him at destination. During the bachelor degree, the same student can expand her/his network to include many new acquaintances. These can provide new information about job market opportunities or other information that motivate the student to change his/her direction in the transition from the bachelor to the master level.

Furthermore, our analysis indicates that big cities have a smaller number of preferential connections with respect to smaller cities. Students coming from small cities display more varied patterns than students from large cities. This behaviour is probably related to the "medium-large city to large city" and "medium-small city to medium-small city" preference presented in Chapter 2, where students coming from small (big) cities tend to move to small (big) cities in the transition from high-school to the bachelor's. This model, though, is not predictive anymore for the transition from bachelor to master studies. Indeed, results from Chapter 3 highlight a transition from minor cities in the bachelor to main cities in the master's (*e.g.* Agrigento–Marche–Lombardy). This may be due to an increased tendency among students to look for job opportunities, and to the growing importance of weak ties in the decision. Furthermore, the stability of three-step preferential patterns over time supports the idea that such patterns not only depend on economic factors. They mostly depend on socio/cultural aspects, especially in the transition from high school to the bachelor's. Indeed, preferential patterns remain highly stable even in the second period considered in this study. This covers the sovereign debt crisis, and 2013 that was the *black year* of Italian household expenditure on goods and services [59].

Finally, the persistency of preferential patterns of student mobility from high school to the masters' degree suggests that chain migration in student mobility is a social phenomenon that covers the whole degree path and not only the first transition from high-school to the bachelor's. Along this line of thinking, student mobility is not only affected by public information on university quality [14, 99] (*e.g.* CENSIS or Shanghai ranking). There is also private information provided by a network of relatives, friends and weak ties at destination. These help students in the decision for enrollment.

Limitations and Future Research

In this dissertation, we introduced statistical methods to empirically detect patterns of student mobility that cannot be considered as random flows in the student-mobility network. Revealed preferential patterns suggest that chain-migration effects, among others, influence the university enrollment decisions of Italian students. In this section, before considering future developments, it is important to stress the limitations and the assumptions behind this work.

As for the limitations, we are aware that:

1) to test the statistical significance of mobility preferential patterns, we implicitly assume that cohorts of enrolled students are samples of a meta-population. The meta-population is made up of all the students who enrolled in an Italian university at some point in the investigated time window. Furthermore, we assume that each subpopulation displays a structure and characteristics common to the meta-population (representativeness). Specifically, assuming that the cohorts of students are samples from the same meta-population allows us to perform an inferential analysis of data and to reveal the preferential mobility patterns;

2) we have no information about the Italian students enrolled in foreign universities. This lack of information does not strongly affect the present analysis since only a small fraction of students from the South choose to enroll at a university abroad;

3) our data does not provide information about the motives that induce students to enroll at a university outside the region of residence. Motives may relate to an anticipation of job mobility, and the appeal of *i)* specific study programs, *ii)* student welfare policies, and the *iii)* university's prestige.

4) Another important limitation concerns the neglected sources of heteroscedasticity. The methodological approaches considered throughout the thesis involve probabilistic models conditioned to both the total number of students moving out from each

cluster and the total number of students enrolling at each university. These properties make the considered null hypotheses appropriate for revealing chain migration effects, since they allow one to untangle chain migration from factors mainly related to node-specific characteristics. However, factors associated with some eventual origin-destination links, such as the presence of devoted connections between the two (*e.g.*, airline flights) are not incorporated, either directly or indirectly, in the null hypotheses.

Keeping in mind these limitations, the rationale of the thesis relies upon specific assumptions that together form the "paradigm" beyond our work:

i) the forces of attraction towards the area of destination and any repulsion from the area of origin can be quantified using the column and row marginals of an origin-destination matrix;

ii) the analysed database provides comprehensive information on student-mobility patterns.

Under these limitations and assumptions, we cannot claim that chain migration is the only factor determining the preferential patterns of mobility that we have found. After all, other specific factors (*e.g.*, the transport network and social class) may also come into play.

As for the future developments, we distinguish between methodological improvement and data enrichment:

1) It would be possible to include in the null hypothesis data concerning special connections between some origin-destination nodes. This could be done by relaxing the hypergeometric distribution hypothesis and considering the Wallenius non-central hypergeometric distribution [91] instead. Though extremely challenging from the computational point of view, this methodology can help to reduce type II errors in the analysis;

2) There would also be the possibility of conducting an *ad-hoc* survey that analyses the motivation behind Italian student mobility. Specifically, to achieve an in-depth understanding of movers' trajectories, two different questionnaires could be created: one specifically devised for (former) students who decided not to go to the college,

and another one for university students–both movers and stayers. This analysis would be particularly helpful for stepping up research into the motives that induce students to leave Southern regions and to enroll at universities in the North, and for investigating the impact of the university context and the role of migration in forming the decision to move. Furthermore, we can jointly consider the results of the present study and the ones of the survey so as to propose specific policy changes for improving the Italian higher-education system, and for reducing the South-North divide.

Appendixes

Appendix A - Tables of abbreviations

Table A1: Abbreviations of target regions.

Destination regions	Abbreviations
Abruzzo	ABR
Basilicata	BAS
Calabria	CAL
Campania	CAM
Emilia-Romagna	EMI
Friuli Venezia Giulia	FRI
Lazio	LAZ
Liguria	LIG
Lombardy	LOM
Marche	MAR
Molise	MOL
Piedmont	PIE
Apulia	PUG
Sardinia	SAR
Sicily	SIC
Tuscani	TOS
Trentino Alto Adige	TRE
Umbria	UMB
Valle D'Aosta	VAL
Veneto	VEN

Table A2: Abbreviations of Sicilian areas of origin.

Sicilian areas	Abbreviations
Agrigento	AGR
Canicattì	CAN
Castelvetrano	CAS
Catania	CAT
Messina	MES
Palermo	PAL
Ragusa	RAG
Syracuse	SIR
Trapani	TRA
Vittoria	VIT

Table A3: Abbreviations of Sardinian areas of origin.

Sardinian areas	Abbreviations
Cagliari	CAG
Nuoro	NUO
Olbia-Tempio	OLB
Oristano	ORI
Sassari	SAS

Table A4: Abbreviations of Apulian areas of origin.

Apulian areas	Abbreviations
Altamura	ALT
Andria	AND
Bari	BARI
Barletta	BAR
Bitonto	BIT
Brindisi	BRI
Casarano	CAS
Conversano	CON
Foggia	FOG
Lecce	LEC
Maglie	MAG
Martina Franca	MAR
Molfetta	MOL
Taranto	TAR
Tricase	TRI

Table A5: Abbreviations of Sicilian provinces

Sicilian provinces	Abbreviations
Agrigento	AG
Caltanissetta	CL
Catania	CT
Enna	EN
Messina	ME
Palermo	PA
Ragusa	RG
Syracuse	SR
Trapani	TP

Appendix B - Software and libraries

Statistical data analyses were performed in `R4.1.1`. Data preparation and descriptive statistics were performed using the `tidyverse` [112] library. As for the cartograms, we used the `sf` [87] and `tmap` [105] libraries. The treemaps and heatmaps were performed using `treemap` [106] and `gplots` [111] libraries, respectively. The statistically validated networks proposed in sections 1.3.2 and 3.4.2 were implemented in `R4.1.1`. The visualization of statistically validated networks reported in section 1.3.2 and 3.4.2 were obtained by using `Cytoscape 3.8.2`.

References

[1] Abramovsky, L., Harrison, R., Simpson, H.: University research and the location of business r&d. The Economic Journal **117**(519), C114–C141 (2007)

[2] Anderberg, M.R.: Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks, vol. 19. Academic press (2014)

[3] Attanasio, M., Enea, M.: La mobilità degli studenti universitari nell'ultimo decennio in italia. In: G. De Santis, E. Pirani, M. Porcu (eds.) Rapporto sulla popolazione. L'istruzione in Italia, pp. 43–58. Bologna, Il Mulino (2019)

[4] Attanasio, M., Priulla, A.: Chi rimane e chi se ne va? Un'analisi statistica della mobilità universitaria dal Mezzogiorno d'Italia. In: M. Attanasio, O. Giambalvo, G. Ragozini, M. Porcu (eds.) Verso Nord. Le nuove e vecchie rotte delle migrazioni universitarie. Franco Angeli, Cham (2020)

[5] Ballarino, G., Panichella, N.: Social origins, geographical mobility and occupational attainment in contemporary Italy. Genus **77**(1), 1–24 (2021)

[6] Banerjee, B.: Social networks in the migration process: empirical evidence on chain migration in india. The Journal of Developing Areas **17**(2), 185–196 (1983)

[7] Bar-Yam, Y.: Dynamics of complex systems. Reading, Massachussetts, Addison-Wesley (1997)

[8] Barrioluengo, M.S., Flisi, S.: Student Mobility in Tertiary Education: institutional factors and regional attractiveness. JRC Working Papers JRC108895, Joint Research Centre (Seville site) (2017)

[9] Bartram, D., Poros, M., Monforte, P.: Key concepts in migration. Sage (2014)

[10] Beech, S.E.: Adapting to change in the higher education system: International student mobility as a migration industry. Journal of Ethnic and Migration Studies **44**(4), 610–625 (2018)

[11] Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) **57**(1), 289–300 (1995)

[12] Böcker, A.G.: Chain migration over legally closed borders: Settled immigrants as bridgeheads and gatekeepers. The Netherlands' Journal of Social Sciences **30**(2), 87–106 (1994)

[13] Borjas, G.J.: Economic theory and international migration. International migration review **23**(3), 457–485 (1989)

[14] Boscaino, G., Sottile, G., Adelfio, G.: Migration and students' performance: detecting geographical differences following a curves clustering approach. Journal of Applied Statistics pp. 1–15 (2020). DOI 10.1080/02664763.2020.1845624

[15] Bouwel, L.V., Veugelers, R.: The determinants of student mobility in Europe: the quality dimension. European Journal of Higher Education **3**(2), 172–190 (2013). DOI 10.1080/21568235.2013.772345

[16] Brettell, C.B., Hollifield, J.F.: Migration theory: Talking across disciplines. Routledge (2014)

[17] Breznik, K., Skrbinjek, V.: Erasmus student mobility flows. European Journal of Education **55**(1), 105–117 (2020)

[18] Brooks, R., Waters, J.: International higher education and the mobility of uk students. Journal of Research in International Education **8**(2), 191–209 (2009)

[19] Brooks, R., Waters, J.: Social networks and educational mobility: the experiences of UK students. Globalisation, Societies and Education **8**(1), 143–157 (2010). DOI 10.1080/14767720903574132

[20] Bruno, G., Genovese, A.: A spatial interaction model for the representation of the mobility of university students on the italian territory. Networks and Spatial Economics **12**(1), 41–57 (2012). DOI 10.1007/s11067-010-9142-7

[21] Burt, R.S.: Structural Holes: The Social Structure of Competition. Harvard University Press (1992)

[22] Castles, S., de Haas, H., Miller, M.J.: The age of migration: International population movements in the modern world. Palgrave Macmillan Basingstoke (2014)

[23] Castles, S., Kosack, G.: Immigrant workers and class structure in Western Europe. Oxford University Press (1973)

[24] Cervia, S., Biancheri, R.: Women in science: The persistence of traditional gender roles. a case study on work–life interface. European Educational Research Journal **16**(2-3), 215–229 (2017)

[25] Chun, Y., Griffith, D.A.: Modeling network autocorrelation in space–time migration flow data: An eigenvector spatial filtering approach. Annals of the Association of American Geographers **101**(3), 523–536 (2011). DOI 10.1080/00045608.2011.561070

[26] Ciriaci, D.: Does University Quality Influence the Interregional Mobility of Students and Graduates? The Case of Italy. Regional Studies **48**(10), 1592–1608 (2014). DOI 10.1080/00343404.2013.821569

[27] Cohen, R.: The new helots: Migrants in the international division of labour. Gower Publishing Company, Limited (1988)

[28] Columbu, S., Porcu, M., Primerano, I., Sulis, I., Vitale, M.P.: Geography of Italian student mobility: A network analysis approach. Socio-Economic Planning Sciences **73**, 100918 (2021). DOI https://doi.org/10.1016/j.seps.2020.100918

[29] D'Agostino, A., Ghellini, G., Longobardi, S.: Out-migration of university enrolment: the mobility behaviour of italian students. International Journal of Manpower **40**(1), 56–72 (2019). DOI 10.1108/IJM-07-2017-0169

[30] D'Agostino, M., Ruffino, G.: I rilevamenti sociovariazionali – Linee Progettuali. Atlante Linguistico della Sicilia. Palermo, Centro studi filologici e linguistici siciliani. Dipartimento di Scienze Filologiche e Linguistiche, University of Palermo (2005)

[31] Dal Bianco, A., Spairani, A., Ricciari, V.: La mobilità degli studenti in Italia: un'analisi empirica. Rivista di Economia e Statistica del Territorio **1**(1), 123–143 (2010)

[32] Daniel, C.: Building a south-south connection through higher education: the case of peruvian university students in brazil. Cahiers de la Recherche sur l'éducation et les savoirs **13**, 119–137 (2014)

[33] Dekker, S.: Drift into failure. From hunting broken to understanding complex systems. Farnham, Ashgate (2011)

[34] Dolinska, A., Jonczy, R., Rokita-Poskart, D.: Post-secondary-school migration of young people to large regional centres as a factor of depopulation and disharmonious regional development in Poland. European Research Studies Journal **23**(3), 260–279 (2020)

[35] Donnelly, M., Gamsu, S.: Home and away : social, ethnic and spatial inequalities in student mobility. Project report, DU, London (2018). URL `http://dro.dur.ac.uk/27367/`

[36] Dotti, N., Fratesi, U., Lenzi, C., Percoco, M.: Local labour markets and the interregional mobility of italian university students. Spatial Economic Analysis **8**(4), 443–468 (2013). DOI 10.1080/17421772.2013.833342

[37] Dotti, N., Fratesi, U., Lenzi, C., Percoco, M.: Local labour market conditions and the spatial mobility of science and technology university students: evidence from italy. Review of Regional Research **34**(2), 119–137 (2014). DOI 10.1007/s10037-014-0088-y

[38] Easley, D., Kleinberg, J.: Networks, Crowds, and Markets. Cambridge, Cambridge University Press (2010)

[39] Enea, M.: From south to north? mobility of southern italian students at the transition from the first to the second level university degree. In: C. Perna, M. Pratesi, A. Ruiz-Gazen (eds.) Studies in Theoretical and Applied Statistics, pp. 239–249. Springer International Publishing, Cham (2018)

[40] Etzo, I.: The determinants of the recent interregional migration flows in Italy: A panel data analysis. Journal of Regional Science **51**(5), 948–966 (2011)

[41] Eurenius, A.M.: A family affair: Evidence of chain migration during the mass emigration from the county of halland in sweden to the united states in the 1890s. Population studies **74**(1), 103–118 (2020)

[42] Everitt, B., Landau, S., Leese, M., Stahl, D.: Cluster Analysis. Wiley Series in Probability and Statistics. Wiley (2011)

[43] Faggian, A., McCann, P., Sheppard, S.: Some evidence that women are more mobile than men: gender differences in uk graduate migration behavior. Journal of Regional Sciences **47**(3), 517–539 (2007)

[44] Fararo, T.J., Doreian, P.: Tripartite structural analysis: Generalizing the breiger-wilson formalism. Social Networks **6**(2), 141–175 (1984). DOI https://doi.org/10.1016/0378-8733(84)90015-7

[45] Foti, N.J., Hughes, J.M., Rockmore, D.N.: Nonparametric sparsification of complex multiscale networks. PLOS ONE **6**(2), 1–10 (2011). DOI 10.1371/journal.pone.0016431

[46] Fussell, E.: The cumulative causation of international migration in latin america. The Annals of the American Academy of Political and Social Science **630**(1), 162–177 (2010)

[47] Genova, V.G., Tumminello, M., Enea, M., Aiello, F., Attanasio, M.: Student mobility in higher education: Sicilian outflow network and chain migrations. Electronic Journal of Applied Statistical Analysis **12**(4), 774–35 (2019). DOI 10.1285/i20705948v12n4p774

[48] Giambona, F., Porcu, M., Sulis, I.: Students mobility: assessing the determinants of attractiveness across competing territorial areas. Social Indicators Research **133**(3), 1105–1132 (2017)

[49] Giulietti, C., Wahba, J., Zenou, Y.: Strong versus weak ties in migration. European Economic Review **104**, 111–137 (2018)

[50] Grabher, A., Wejwar, P., Unger, M., Terzieva, B.: Student mobility in the EHEA: Underrepresentation in student credit mobility and imbalanced degree mobility; Study commissioned by the Austrian Ministry of Science (BMWF) (2014)

[51] Gümüş, S., Gök, E., Esen, M.: A review of research on international student mobility: Science mapping the existing knowledge base. Journal of Studies in International Education **24**(5), 495–517 (2020)

[52] Hagen-Zanker, J.: Why do people migrate? A review of the theoretical literature. A Review of the Theoretical Literature (January 2008). Maastrcht Graduate School of Governance Working Paper No (2008)

[53] Harris, J.R., Todaro, M.P.: Migration, unemployment and development: a two-sector analysis. The American economic review pp. 126–142 (1970)

[54] Haug, S.: Migration networks and migration decision-making. Journal of Ethnic and Migration Studies **34**(4), 585–605 (2008)

[55] Hemsley-Brown, J.: Getting into a russell group university: High scores and private schooling. British Educational Research Journal **41**(3), 398–422 (2015)

[56] Holland, P.W., Leinhardt, S.: The statistical analysis of local structure in social networks (1974)

[57] Impicciatore, R., Tosi, F.: Student mobility in Italy: The increasing role of family background during the expansion of higher education supply. Research in Social Stratification and Mobility **62**, 100409 (2019). DOI https://doi.org/10.1016/j.rssm.2019.100409

[58] Impicciatore, R., Tuorto, D.: Mobilità interna e istruzione universitaria: risorse familiari, individuali e opportunità di ascesa sociale nell'occupazione. SOCIOLOGIA DEL LAVORO **121**(1), 51–78 (2011). DOI 10.3280/SL2011-121004

[59] Istituto Nazionale di Statistica: Rapporto annuale 2014. La situazione del Paese. Tech. rep., ISTAT, Via Cesare Balbo, 16 - Roma (2014). URL `https://www.istat.it/it/files//2014/05/Rapporto-annuale-2014.pdf`

[60] Javed, B., Zainab, B., Zakai, S.N., Malik, S.: Perceptions of international student mobility: A qualitative case study. Journal of Education and Educational Development **6**(2), 269–287 (2019)

[61] Körner, H., Stark, O.: The migration of labor. International Migration Review **26**, 1462 (1991)

[62] Kuschminder, K.: Strong ties, weak ties: Exploring the role of networks in domestic worker migration from ethiopia to the middle east. Asian and Pacific migration Journal **25**(4), 401–421 (2016)

[63] Lauby, J., Stark, O.: Individual migration as a family strategy: Young women in the philippines. Population studies **42**(3), 473–486 (1988)

[64] Lee, E.S.: A theory of migration. Demography **3**(1), 47–57 (1966)

[65] Lewer, J.J., Van den Berg, H.: Religion and international trade: does the sharing of a religious culture facilitate the formation of trade networks? American Journal of Economics and Sociology **66**(4), 765–794 (2007)

[66] Lewis, J.R.: International labour migration and uneven regional development in labour exporting countries. Tijdschrift voor economische en sociale geografie **77**(1), 27–41 (1986)

[67] Liu, J.M., Ong, P.M., Rosenstein, C.: Dual chain migration: Post-1965 Filipino immigration to the United States. International Migration Review **25**(3), 487–513 (1991)

[68] Lupi, C., Ordine, P.: Family income and students' mobility. Giornale degli Economisti e Annali di Economia **68 (Anno 122)**(1), 1–23 (2009)

[69] Mabogunje, A.L.: Systems approach to a theory of rural-urban migration. Geographical analysis **2**(1), 1–18 (1970)

[70] MacDonald, J.S., MacDonald, L.D.: Chain migration ethnic neighborhood formation and social networks. The Milbank Memorial Fund Quarterly **42**(1), 82–97 (1964)

[71] MacDonald, J.S., MacDonald, L.D.: Italian migration to australia: manifest functions of bureaucracy versus latent functions of informal networks. Journal of Social History pp. 249–275 (1970)

[72] Marshall, D.: The degree mobility spectrum: The tiering of canadian degrees. Mount Royal College (2005)

[73] Massey, D.S.: Social structure, household strategies, and the cumulative causation of migration. Population Index pp. 3–26 (1990)

[74] Massey, D.S., Alarcón, R., Durand, J., González, H.: Return to aztlan. University of California Press (1990)

[75] Massey, D.S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., Taylor, J.E.: Theories of international migration: A review and appraisal. Population and Development Review **19**(3), 431–466 (1993)

[76] Massey, D.S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., Taylor, J.E.: Theories of international migration: A review and appraisal. Population and development review pp. 431–466 (1993)

[77] Mayda, A.M.: International migration: a panel data analysis of economic and non-economic determinants. Available at SSRN 725441 (2005)

[78] Miller, R.G.: Simultaneous Statistical Inference. New York, Springer-Verlag (1981)

[79] MOBYSU.IT: Database MOBYSU.IT, *Mobilità degli studi universitari italiani*, Protocollo di ricerca MIUR - Università degli Studi di Cagliari, Palermo, Siena, Torino, Sassari, Firenze e Napoli Federico II, Fonte dei dati ANS-MIUR/CINECA (2016)

[80] Neal, Z.: The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors. Social Networks **39**, 84–97 (2014). DOI https://doi.org/10.1016/j.socnet.2014.06.001

[81] Nelson, P.: Migration, real income and information 1. Journal of Regional Science **1**(2), 43–74 (1959)

[82] Newman, M.: Networks. An Introduction. Oxford, Oxford University Press (2011)

[83] Nifo, A., Vecchione, G.: Do Institutions Play a Role in Skilled Migration? The Case of Italy. Regional Studies **48**(10), 1628–1649 (2014). DOI 10.1080/00343404.2013.835799

[84] Ordine, P., Rose, G.: Students' mobility and regional disparities in quality and returns to education in italy. Giornale degli Economisti e Annali di Economia **66 (Anno 120)**(2), 149–175 (2007)

[85] Palloni, A., Massey, D.S., Ceballos, M., Espinosa, K., Spittel, M.: Social capital and international migration: A test using information on family networks. American Journal of Sociology **106**(5), 1262–1298 (2001)

[86] Parr, N., Lucas, D., Mok, M.: Branch migration and the international dispersal of families. International Journal of Population Geography **6**(3), 213–227 (2000)

[87] Pebesma, E.: Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal **10**(1), 439–446 (2018). DOI 10.32614/RJ-2018-009. URL https://doi.org/10.32614/RJ-2018-009

[88] Pérez, P., Mcdonough, P.: Understanding latina and latino college choice: A social capital and chain migration analysis. Journal of Hispanic Higher Education **7**(3), 249–265 (2008). DOI 10.1177/1538192708317620

[89] Person, A.E., Rosenbaum, J.E.: Chain enrollment" and college "enclaves": Benefits and drawbacks for latino college students. New directions for community colleges: Community colleges and Latino educational opportunity **133**, 51–60 (2006)

[90] Portes, A., Böröcz, J.: Contemporary immigration: Theoretical perspectives on its determinants and modes of incorporation. International Migration Review **23**(3), 606–630 (1989)

[91] Puccio, E., Vassallo, P., Piilo, J., Tumminello, M.: Covariance and correlation estimators in bipartite complex systems with a double heterogeneity. Journal of Statistical Mechanics: Theory and Experiment **5**(30 May 2019), e53404 (2019)

[92] Restaino, M., Vitale, M.P., Primerano, I.: Analysing international student mobility flows in higher education: A comparative study on European Countries. Social Indicators Research **149**(3), 947–965 (2020)

[93] Reyneri, E.: La catena migratoria: il ruolo dell'emigrazione nel mercato del lavoro di arrivo e di esodo, vol. 105. Il Mulino (1979)

[94] Reyneri, E.: Immigrazione ed economia sommersa. Stato e mercato **18**(2), 287–318 (1998)

[95] Rice, J.A.: Mathematical Statistics and Data Analysis, 3^{rd} edn. Duxbury Press (2007)

[96] Ruiu, G., Fadda, N., Ezza, A., Esposito, M.: An investigation of mobility of italian ph. doctors. Electronic Journal of Applied Statistical Analysis **12**(4) (2019)

[97] Ruiu, G., Genova, V.G., Attanasio, M., Breschi, M.: L'isola che se ne va. In: M. Delogu, M. Meleddu (eds.) Riflessioni trasversali sopra il capitale umano, pp. 105–112 (2020)

[98] Sà, C., Florax, R., Rietveld, P.: Determinants of the regional demand for higher education in the netherlands: a gravity model approach. Regional Studies **38**(4), 375–392 (2004)

[99] Santelli, F., Scolorato, C., Ragozini, G.: On the determinants of student mobility in an interregional perspective: a focus on Campania region. Statistica Applicata-Italian Journal of Applied Statistics **31**(1), 119–142 (2019)

[100] Sassen, S.: The mobility of labor and capital: A study in international investment and labor flow (1988). DOI 10.1017/CBO9780511598296

[101] Simon, H.A.: The Sciences of the Artificial. Cambridge, MIT Press (1996)

[102] Stark, O.: Economic-demographic interactions in agricultural development: The case of rural-to-urban migration, vol. 6. Food & Agriculture Org. (1978)

[103] Stark, O., Bloom, D.E.: The new economics of labor migration. The American Economic Review **75**(2), 173–178 (1985)

[104] Stark, O., Levhari, D.: On migration and risk in ldcs. Economic development and cultural change **31**(1), 191–196 (1982)

[105] Tennekes, M.: tmap: Thematic maps in R. Journal of Statistical Software **84**(6), 1–39 (2018). DOI 10.18637/jss.v084.i06

[106] Tennekes, M.: treemap: Treemap Visualization (2021). URL `https://CRAN.R-project.org/package=treemap`. R package version 2.4-3

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

[108] Tumminello, M., Miccichè, S., Lillo, F., Piilo, J., Mantegna, R.: Statistically validated networks in bipartite complex systems. PLoS ONE **6**(3), e17994 (2011)

[109] Ünal, S.: The new actors of international migration: a comparative analysis of foreign students' experiences in a medium-sized city in turkey. People's Movements in the 21st Century: Risks, Challenges and Benefits p. 231 (2017)

[110] Viesti, G.: Università in declino. Un'indagine sugli atenei italiani da Nord a Sud. Roma, Donzelli (2016)

[111] Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B.: gplots: Various R Programming Tools for Plotting Data (2020). URL `https://CRAN.R-project.org/package=gplots`. R package version 3.1.1

[112] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H.: Welcome to the tidyverse. Journal of Open Source Software **4**(43), 1686 (2019). DOI 10.21105/joss.01686

[113] Wiers-Jenssen, J.: Degree mobility from the nordic countries: Background and employability. Journal of Studies in International Education **17**(4), 471–491 (2013)

[114] Wilson, T.: Weak ties, strong ties: Network principles in mexican migration. Human Organization **57**(4), 394–403 (1998)

[115] Wilson, T.P.: Relational networks: An extension of sociometric concepts. Social Networks **4**(2), 105–116 (1982). DOI https://doi.org/10.1016/0378-8733(82)90028-4

[116] Xulvi-Brunet, R., Sokolov, I.: Reshuffling scale-free networks: From random to assortative. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics **70**(6 2), 066102/1–066102/6 (2004). DOI 10.1103/PhysRevE.70.066102

[117] Yingbo, L., Jiujun, C., Xiao, W., Fuzhen, C.: Research on the matthews correlation coefficients metrics of personalized recommendation algorithm evaluation. International Journal of Hybrid Information Technology **8**(1), 163–172 (2005)

CRediT Author Statement

- **Chapter 1** - ***Vincenzo Giuseppe Genova:*** conceptualization, software, formal analysis, investigation, data curation, writing - original draft, visualization. ***Michele Tumminello:*** methodology, formal analysis, investigation, writing - review and editing, supervision. ***Marco Enea:*** formal analysis, writing - review and editing. ***Fabio Aiello:*** investigation, writing - review and editing. ***Massimo Attanasio:*** supervision, project administration, funding acquisition.

- **Chapter 2** - ***Vincenzo Giuseppe Genova:*** methodology, software, formal analysis, investigation, data curation, writing - original draft, visualization. ***Gabriele Ruiu:*** conceptualization, writing - original draft, writing - review and editing. ***Massimo Attanasio:*** conceptualization, formal analysis, investigation, supervision, project administration, funding acquisition. ***Marco Breschi:*** conceptualization, formal analysis, investigation, supervision.

- **Chapter 3** - ***Vincenzo Giuseppe Genova:*** conceptualization, methodology, software, investigation, data curation, writing - original draft, visualization. ***Michele Tumminello:*** conceptualization, methodology, formal analysis, supervision. ***Fabio Aiello:*** conceptualization, formal analysis, writing - review and editing. ***Massimo Attanasio:*** supervision, project administration, funding acquisition.

Outputs of the PhD research

During the PhD program I co-authored three papers, each one corresponding to a chapter of the thesis. Specifically:

i) V.G. Genova, M. Tumminello, M. Enea, F. Aiello, M. Attanasio (2019). Student mobility in higher education: Sicilian outflow network and chain migrations. Electronic Journal of Applied Statistical Analysis, 12 (4), pg. 774-35. **Published**.

ii) V.G. Genova, G. Ruiu, M. Attanasio, M. Breschi (2021). The good old ideas: an attempt to adapt the concept of chain migration to explain Italian student mobility. **To be published**.

iii) V.G. Genova, M. Tumminello, F. Aiello, M. Attanasio (2021). A network analysis of student mobility patterns from high school to master's. Stat Methods Appl, DOI: https://doi.org/10.1007/s10260-021-00592-4 **Published**.

Moreover, I co-authored other publications:

- M.P. Vitale, V.G. Genova, G. Giordano, G. Ragozini (2021). Community detection in tripartite networks of university student mobility flows, Book of abstracts and short papers CLADAG 2021.

- G. Boscaino, V.G. Genova (2021). Exploring drivers for Italian university students' mobility: first evidence from AlmaLaurea data, Book of Short Papers SIS 2021.

- B. Margo, V.G. Genova et al. (2021). Predicting in-hospital mortality from Coronavirus Disease 2019: A simple validated app for clinical use, PLOS ONE 16(1).

- P. Miano, M. Bellomare, V.G. Genova (2021). Personality Correlates of Gaslighting Behaviours in Young Adults, Journal of Sexual Aggression.

- G. Ruiu, V.G. Genova, M. Attanasio, M. Breschi (2020). L'isola che se ne va, In Riflessioni trasversale sopra il capitale umano (a cura di Delogu M., Meleddu M.).

- M. Attanasio, M. Enea, V.G. Genova (2020). Analisi della predittività dei test d'accesso per Economia. In Il test di ingresso: valenza predittiva e rilevanza nei percorsi di orientamento universitario. Le attività del CISIA al servizio degli Atenei, Consorzio interuniversitario Sistemi integrati per l'accesso.
- V.M.R. Muggeo, A. Consiglio, G. Sottile, V.G. Genova, G. Bertolazzi, M. Porcu (2020). Analisi e monitoraggio della diffusione del Covid19 in Italia: il gruppo CoViSTAT19, STATISTICA & SOCIETÀ, vol. anno IX, p. 1-4.
- G. Boscaino, V.G. Genova (2020). Pull factors for university students' mobility: a gravity model approach, Book of Short Papers—SIS 2020.
- M. Sciandra, G. Boscaino, V.G. Genova (2019). Parceling in Multilevel Structural Equation Models for the measure of a latent construct, Statistical Methods for Service Quality Evaluation, Book of short papers of IES 2019.
- A. M. Parroco, V.G. Genova, L. Mancuso, F. Giannone (2019). Assessing Mental Health Therapeutic Communities Functioning. Book of Short Papers ASA2019 - Scientific Conference on Statistics for Health and Well-being with editors Carpita M. and Fabbris L.

During the three years of the Ph.D. program, I participated in the following conferences where I presented some of the content of this thesis:

- 5^{th} European Conference on Social Networks EUSN 2021 – 6-10/09/2021, Virtual (Naples). **Contributed Talk**.
- Models for people in motion in the Mediterranean basin, Istituto Nazionale di Alta Matematica "F. Severi", Rome (online), 23/04/2021. **Invited speaker**.
- Seventh International Workshop on Social Network Analysis Multilayer, Multilevel and Multimode Networks – 30-31/10/2019, Vietri sul Mare (Italy). **Contributed Talk**.

Funding

This thesis has been supported from Italian Ministerial grant PRIN 2017 "From high school to job placement: micro-data life course analysis of university student mobility and its impact on the Italian North-South divide.", n. 2017HBTK5P - CUP B78D19000180001.