



Historical Demography and GIS Databases: New Research Paths

Fornasin, A. Marzona, N. Dan

February 2012

n. 2/2012

Historical Demography and GIS Databases: New Research Paths

A. Fornasin, A. Marzona, N. Dan

February 13, 2012

Abstract

Many demographic sources contain a considerable amount of geographical data and information. A census, for instance, informs us on age, sex and family relationships of each resident as well as his/her place of residence (specifying also the address) and birth place. Marriage records (civil and religious) include information on the marriage's place and information on the residence of the spouses (sometimes of their parents too) before the ceremony too. Our contribution aims at illustrating how to best exploit territorial information in the demographic research. We intend to illustrate the construction criteria and the wide possibilities of analysis offered by a database deriving from more than 80.000 marriage records and including about 500.000 geographical data.

Working Paper DIES n° 2/2012

A first version of this paper was presented at conference «New perspectives in historical demography and the history of the family». Lyon 30 November-1 December 2009. Session 3 - Methodological innovations.

Historical Demography and GIS Databases: New Research Paths

1. Introduction

In 2003 Italian Government financed a project, named FRIULI/in prin, promoted by Udine State Archive Office (Archivio di Stato di Udine). Three were the aims of the project: a) to find historical information on names and surnames; b) to track relationships among people (genealogy trees); c) to carry both aggregate and individual historical-demographical analysis on the collected data.

To support this project, was built a structured database collecting data referred to almost all the municipalities of the province. The database was projected using administrative information preserved inside State Archive, mainly civil marriage certificates of the period 1871-1900 and the lists of military conscripts born in the years 1846-1890.

This paper describes how we use geographical information cited by the sources. We tried to adopt a general model, not dependent on the area covered by our research: they can be adopted without changes to the entire state and even to different territorial areas.

2. The source

To project the structure of the database we used both marriage certificates and the lists of military conscripts, and we made abstractions to guarantee the extension of our archiving system to the registration of personal civil certificates of a different nature, such as birth and decease certificates. In this paper the description of the sources will be limited to the sole marriage certificate: we then describe the steps we followed to build the general model of structure and data processing.

We had to register on the database some 80,000 marriage certificates issued by the Italian Registry Office during the period 1872-1900 and now held in the Udine State Archive.

Marriage certificates provide useful data on spouses, irrespective of where the marriage took place¹. In addition to recording the name and surname of the spouse, their age, often their profession and their previous marital status, they also contain information on the spouses' parents, usually limited to the name of the father and name and surname of the mother.

Marriage certificates are also rich in territorial information. A single record can specify up to nine different geographical indications: the municipality where the marriage took place, the birthplace and residence of each spouse, the residence of their parents.

3. Toponyms and database

All over the project, we adopted some general constraints. One of these is the strong control on the input of territorial information to avoid ambiguities. Our purpose was to collect high quality geographical data to be used in conjunction with GIS to map all the attributes directly or indirectly associated to the persons cited in the database, such as name, surname, age and job.

Geographical information, in the historical certificates, is expressed as toponyms. A toponym, or “name of place”, usually indicates a municipality, but often it represents a single hamlet in a municipality. It could be sometimes very imprecise, i.e. “America”, or just a little more specific: “Egypt” or “France”. Therefore a toponym, on a source, can represent a place, an area, an administrative entity: the territorial level of information is not homogeneous, and can be ambiguous.

According to the conventions adopted in GIS, we made a distinction between places and territories. A place, for example the city of Udine, can be represented on a map as a point. Its position is fixed, almost invariable in time, and can be detected by a set of coordinates. A territory, for example “Municipality of Udine” or “Italy”, is mapped through its boundaries. Therefore a territory can other territories and can contain many places. The data structures we adopted are able to deal with these particularities.

GIS can represent both raster and vectorial maps. The first are static images where geographical coordinates are defined. In the second, maps are defined as a set of interrelated geometric bidimensional shapes; the map is drawn dynamically by composing in an appropriate way its set of elements. The basic shapes are points, lines², polygons; every shape

¹ S. Moretto, *Le statistiche italiane dello stato civile dal 1862 al 1930*, Bollettino di Demografia Storica, 14 (1991): 51-71.

² Lines are not included in our database.

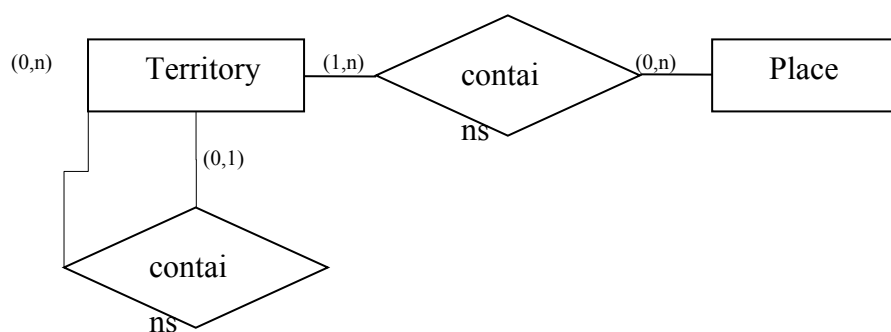
has its proper attributes. The points are dimensionless elements that define a full set of coordinates (i.e. a specific point on a map). They are used to represent exactly the position of a house, a tree, a city, a person. Polygons are composed by a set of lines that delimit a finite portion of a bidimensional space. Their lines set the boundaries of the area. Therefore, an attribute of a polygon is the area covered.

We decided to map cities, villages and hamlets as points of the GIS system, and municipalities, states and continents as polygons.

With this model, we do not need to declare a dimensional scale to describe places, as they are points, dimensionless. In other words, both a spare farm and a big city will be represented as a point on the map. On the other hand, we need to declare in advance the geographic scale to be used to represent territories (in our case, mainly municipalities)³.

For analysis purposes, the conceptual model of territorial data is the one shown in figure 1, with entities (places) that belong to a dynamic hierarchy of territories.

Fig. 1. *Conceptual model of territorial data*



Stated this model, to record and use correctly data we had to solve some practical problems. For instance, the sole transcription of the toponym does not give us the guarantee of its position. If we use this information alone, we can encounter two kinds of error. The first derives from the presence of different places that have the same name (for instance, Sammardenchia, hamlet of Pozzuolo del Friuli and hamlet of Tarcento): in this case we could by mistake give them the same position. The second error consist in the duplication of a single place or territory, when it is referred with different names (for instance, the municipality of Flaibano is referred as Sant’Odorico as well). To identify all the places and

³ A. Fornasin, *I sistemi informativi geografici per la storia del Friuli. Ricerche in corso*, «Annales Series Historia et Sociologia», 15 (2005) 1, 25-38.

the territories cited by the source, each one is given a unique code and input operation are supported by a procedure that helps the operator solve ambiguities.

We adopted two levels of code. The first is an alphanumeric code built in order to identify the coordinates of the places and to help the operator in data input. Each code is composed by a three letter substring to identify the nation, (“IT_” for Italy)⁴, followed by a second substring composed by 9 digits. The first six are the official code of the municipality (ISTAT code) whose the place belongs now⁵. They represent both the province (digits 1-3) and the municipality inside the province (digits 4-6). The following two characters represent the place inside the municipality. To refer the entire municipality, these two characters are “00”. Therefore, in GIS representation, if a code terminates with 00 it represents an area (a polygon), otherwise a point. For instance, the code of Udine as a city is IT_03012901, while the code of Udine as a municipality is IT_03012900. The very last character, set only in few cases, is used to distinguish when needed, two different places of the same municipality that can be named indistinctly together in the source. For instance, Cergneu, hamlet of Nimis, is given the code IT_03006505. When the source specifies Cergneu Inferiore the code used is IT_03006505A, when it specifies Cergneu Superiore the code will be 03006505B.

Unfortunately, often is not possible to distinguish between place and territory (point and polygon from a GIS point of view). Therefore we adopted the following decision criteria: if a toponym refers both a place and a territory, we set the ‘Place’ code if the toponym is the same as the municipality where the certificate was registered (i.e.: Udine as the place of residence of the spouse in a marriage held in Udine will be coded as IT_03012901). If the toponym is cited in a certificate registered in a different municipality, the code used is the “territory” one (i.e.: Udine as the place of residence of the spouse in a marriage held in Cividale will be coded as IT_03012900).

We have not yet defined completely the coding criteria for the foreign places and the territories. The database contains only a small number of places extra Italian boundaries: therefore we adopted the territorial configuration as defined al the date of December 31st 2005. For the once Italian municipalities of Slovenia and Croatia the territorial reference is the one prior the Paris Peace Treaties of February 10th 1947.

This model let us solve easily input ambiguities and let us make some analysis based on data aggregation on the current administrative territorial partition.

Solved the problem of the identification of the toponym, we have to face an other question, related to the time dimension. While the position of a place is almost invariable in time, the

⁴ Soon we will adopt the standard ISTAT Code for foreign nations.

⁵ <http://www.istat.it/strumenti/definizioni/comuni/>.

borders of the territories may change frequently. In a GIS point of view, therefore, points are fixed, polygons are variable in time. As a consequence, at a given date a point can belong to a polygon, but at a different date it can belong to a different polygon. The territorial hierarchy we used to identify the place referred by the toponyms does not consider this variability in time. It is only a method to disambiguate the native information. When we faced this problem, we evaluated two possible solution:

1. to map time information on the database, adding it to the “belong” relationship between a place and a territory
2. to carry data analysis on historical territorial configurations by using a GIS system.

We excluded the first solution, as it would have raise excessively the level of complexity of the data model. We decided instead to built a set of maps, one for each census year⁶; when we need to analyze data based on historical territorial partition, we exploit the properties of analysis and representation of GIS software, that aggregate different elements on the basis of their position in the space.

We have therefore a large amount of data, mostly localized to point in the space, kept on a relational database, and a set of maps in a GIS system we use to analyze the spatial characteristics of the data collected.

4. An application

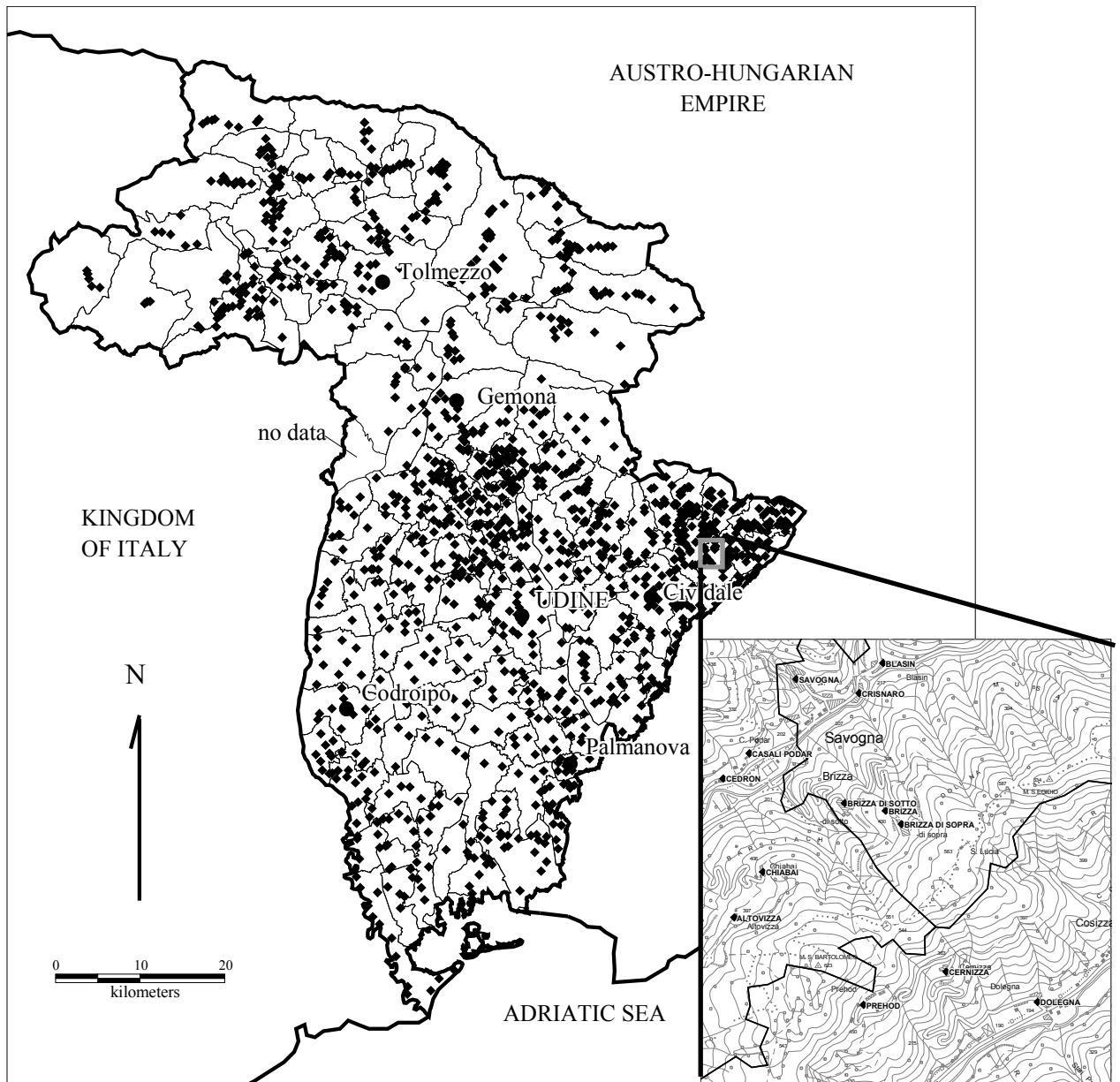
One first application of our geographical data model is the research on exogamy and marital propinquity in Friuli. We described this theme on a recent paper⁷. Here we would evidence one of the ways to take advantage of the spatial information registered in the civil status corpus of documents.

For GIS database, we used georeferenced polygon and point maps of the historical boundaries of Friuli’s municipalities produced by the Department of Statistics of the University of Udine (Figure 2).

⁶ This is a work in progress.

⁷ A. Fornasin A. Marzona, *Exogamy and Marital Propinquity in 19th Century Northeast Italy*, «Historical geography», 37 (2009).

Fig. 2. Polygon and point maps of the Province of Friuli (1881)



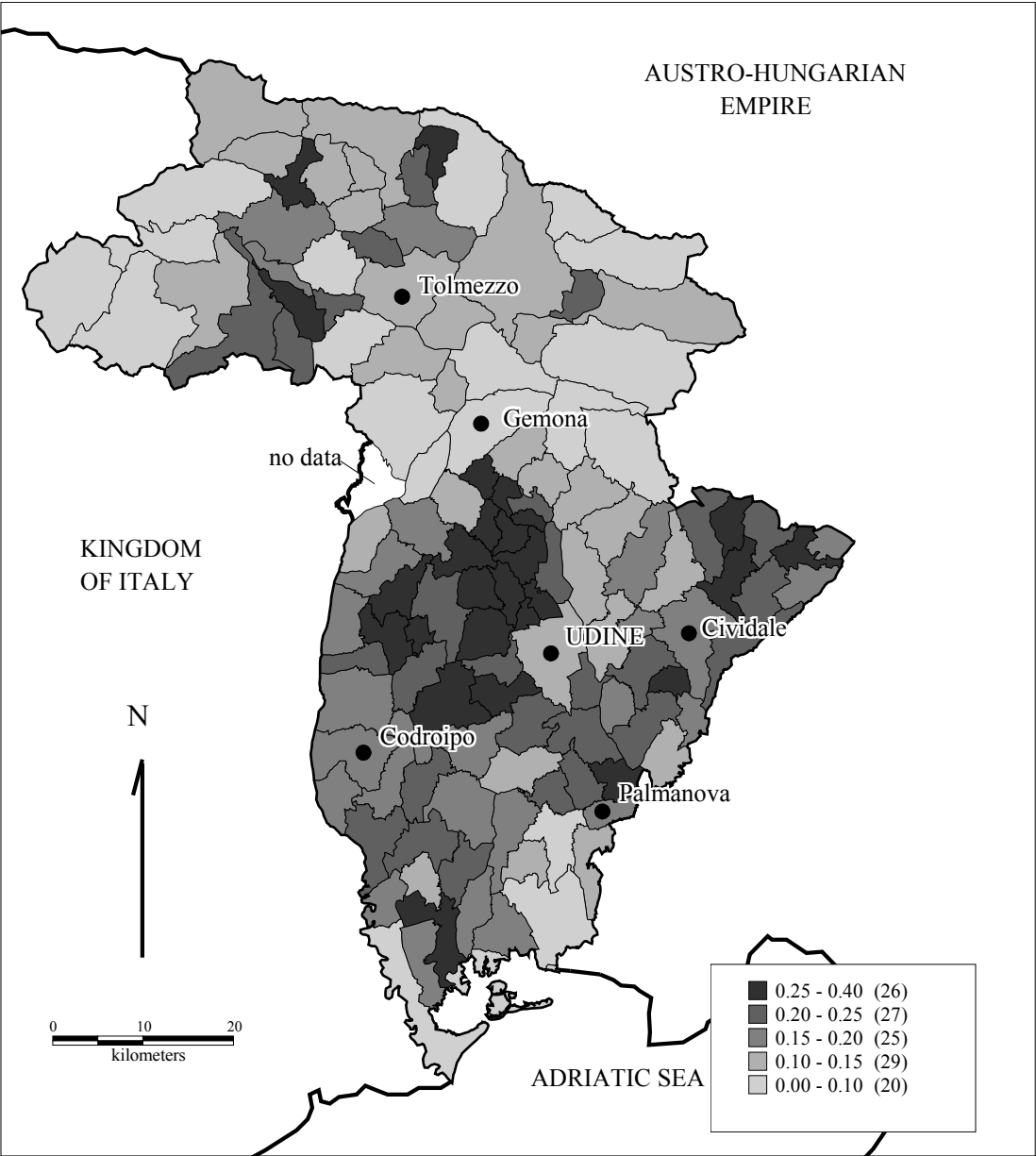
note: the diamond-shaped symbols indicate the localities mentioned in the 1881 census and in the marriage registers.

The polygon map represents the boundaries of 127 municipalities. Since some administrative boundaries changed during the study period, the map we chose as our framework of reference represents the boundaries of December 31st 1881. The point map represents the position of all 656 localities listed in the 1881 census together with the 1,278 localities cited in the marriage certificates. The hamlets are related to the municipalities according to historical information.

Although predominantly based on marriage certificate data, the research also uses other demographic and geographical sources from the National Institute of Statistics (ISTAT). Data provided by the censuses of 1871, 1881 and 1901 on the number of residents in each

municipality were used to calculate the total number of inhabitants in the area under study. The very detailed data provided by the 1881 census, which included population records for each individual hamlet, allowed us to estimate the number of inhabitants in those municipalities whose boundaries had changed over time.

Fig. 3. Exogamy Levels for the Province of Friuli (1872-1900)



A marriage is exogamous when a person who lives in a given place marries another person who lives in a different place. The choice of the word ‘place’ is explicitly generic and is just one example of the many ambiguous terms commonly used in studies on exogamy.

The simplest way to measure territorial exogamy, for instance the exogamy of a parish, is to calculate the ratio between the number of marriages in which one of the spouses is not

native to the parish and the total number of marriages celebrated within that parish. The complementary measure indicates the endogamy level of the parish.

We use this measure to give a first description of how exogamy levels were distributed across the territory: we calculated intra-provincial exogamy as the ratio between the number of marriages involving a resident and a non-resident in a given municipality and the total number of marriages in the same municipality⁸.

The results of our analysis are shown in Figure 3. An analysis of the map can be found in the previously cited paper.

The measurement of exogamy at municipal level poses a number of geographic and demographic problems that derive from both the shape and size of the territory as well as the number of inhabitants. A slightly more accurate way to calculate exogamy would therefore be to take these additional aspects into consideration, and this is something we will do in part – using a variety of different means in accordance with the different types of analysis – further on in this study. In addition to exogamy levels, a second interesting item under examination is marital propinquity (the distance between the places of residence of the spouses).

Now we will give only a brief explanation of the criteria used to construct the map. As we stated previously, we define a marriage as “exogamous” if the distance between the spouses’ places of residence exceeds zero. We further affirmed that the distance is calculated on a municipality basis; we did not however specify how this is measured. There are numerous ways to measure distance and the choice can have a significant bearing on the results.

It could, for example, be measured as the distance between the geographic barycentres of the municipalities (the so called “centroids”); alternatively, we could measure it as the distance between the largest hamlets, following, for optimal precision, road axes. To give this study both geographical and demographic validity, we choose to measure it as the distance between the demographic barycentres of the municipalities. We calculated the longitude and latitude of the demographic barycentre of each municipality using the 1881 census populations together with geographical coordinates obtained from the map⁹. The formula we used to calculate the longitude of the demographic barycentre \bar{x} of a municipality is

⁸ To calculate the denominator we used the data published by MAIC.

⁹ Obviously, the demographic barycentre of the single municipalities changed during the period under examination. The populations of hamlets change over time and thus move the demographic barycentres of the municipality. Census data published by MAIC only allows these movements to be tracked at provincial level. To estimate the variation, we computed the demographic barycentre of the province in accordance with the 1881 boundaries using both demographic barycentres built at individual municipal level and the population rates of municipalities. What we can observe are merely minor provincial movements. The barycentre moves approximately 500 meters north between 1871 to 1881 and slightly more than 200 meters east from 1881 to 1901. In other words, Friuli’s population distribution varied only very marginally during the two intervals; this leads us to assume that the population did not vary at individual municipal level sufficiently to radically alter the entire demographic structure of the province.

$\bar{x} = \frac{\sum_i x_i \cdot p_i}{\sum_i p_i}$, where x_i and p_i are, respectively, the longitude and the number of inhabitants of

hamlet i . A similar equation is adopted for the calculation of the latitude.

We measured the distance between each of the 12,265 spouses at municipality level¹⁰. The values used to construct the distance map were taken by weighting the distance with the mean distance between people living in a given municipality and all the other inhabitants of the province. By applying this method we were able to eliminate the distortion that arise from the geographical location of the municipalities¹¹. The values plotted on the map, therefore, are not the mean distances between the spouses, but mean values that weight the distance between spouses with the distance between inhabitants generally. A value greater than one indicates that the mean distance between spouses is higher than the mean distance between people of that municipality and all the other people of the province; a value less than one indicates the opposite. The results are shown in Figure 4.

From a combined reading of Figures 3 and 4, we can conclude that contrary to what might be reasonably expected, higher exogamy levels do not imply greater distances between the spouses. Many places with a high level of exogamy reveal a tendency for inhabitants to find their spouse within a relatively short distance, whereas in those with lower levels the opposite may be the case. There are some exceptions, the northwest mountain area being one example. The latter, however, has a traditional predisposition for extra-provincial exogamy. This effect could be generated by other spatial distortions not considered in this work; for example, the presence of direct roads and easy communication channels between adjacent municipalities, or the rate between population density and the length of municipal boundaries.

Forthcoming developments

As we anticipated, we can design maps using all the attributes of people cited in the database. Some of the attributes are easy to recognize and to map, for example the average marriage age related to the municipalities of residence. Other attributes can be computed with a greater uncertainty. For instance, currently we are not able to map a reliable territorial

¹⁰ The mean distance of intra-provincial exogamic marriages is 9,5 km.

¹¹ The standardization is necessary because the average distance between persons is larger for municipalities located at the borders of the region than for the ones located in the inner part. We adopted the same method employed in K. Haandrikman, L. van Wissen, C. Harmsen, *Explaining spatial homogamy. Compositional, spatial and regional cultural determinants of regional patterns of spatial homogamy in the Netherlands*, «Applied Spatial Analysis and Policy», 4 (2011), 75-93.

distribution of the professions of individuals, a representation that could tell us information on area specialization and on the social status of people.

A further development of our project is to implement HISCO (Historical International Classification of Occupations) encoding: we would like to evaluate the geographical diffusion of abilities and competences related to different occupations by using a quantitative analysis.

Fig. 4. Marital propinquity index for the Province of Friuli (1872-1900)

