A STATISTICAL ANALYSIS ON THE TRANSITION FROM
UNIVERSITY TO LABOUR MARKET

Adriana Monte, Gabriella Schoier

1. Introduction

As it is well known there are many reasons for going to university, among them higher education is seen as a basic step to obtain a good job. Many higher education programs are explicitly vocational, including medicine, actuarial sciences, engineering, accountancy and law. Other courses are less directly vocational; nevertheless, they help students to develop analytical and presentational skills which are highly valued in our modern economy. In fact in an increasingly knowledge-driven and global economy, universities need to ensure that they produce graduates who are employable, innovative, and in possession of skills which are relevant to national and international needs (Lombardo et al. 2011).

The aim of this study is to analyze some problems regarding the transition from university to labor market, in particular exploring the ties between degree and economic sector of the first post-degree job (Salas-Velasco 2007, Pozzoli 2009). In order to perform the analysis we use two large administrative databases.

The data are used to understand the ties between different types of degrees and different economic sector of employment. After the integration of the two databases and the examination of data quality, we explore the existing ties between type of degree and economic sector.

The individuation of the subjects and the specification of the ties allow to define a network. Different classical social network theories can be used; among these the small-world literature has shown that there is a high degree of local clustering in the networks, this suggests that an approach for studying the structure of large networks would involve first the identification of local clusters and then the analysis of the relations within and between clusters.

Our aim here is to explore such phenomena, in terms of the presence of clusters and of the ties between employers and graduates, trying to understand the origins and motivations of such elements.
The application concerns the local labor market of the Friuli Venezia Giulia Region (F.V.G.), a region situated in the north-east of Italy, and the graduates of the University of Trieste.

2. The two databases

In this study we used administrative data provided by two different sources, the University of Trieste as regards the graduates and the Labour Market Regional Observatory of F.V.G. as regards the employees.

The former administrative database, regarding the graduates, is formed by 40216 records, from January 2000 to July 2010. Each record contains some general information related to the graduate (fiscal code, sex, birth date, place of residence, high school degree, etc.) and several information related to the degree (date, vote, etc.). There is also a variable identifying the profile of the degree (standard, “honoris causa”, transfer degree and other).

In this study we considered only the standard profile in so doing the graduates, identified by fiscal codes, are 36131, each of them may have more degrees (in this database there are up to five degrees for a person).

In order to obtain information related to the first post-degree job (Chies et al. 2013) we used a database containing the data of the activations and terminations given by the Labour Market Regional Observatory of F.V.G that refers to all the employments (but not to self-employments). This latter database refers to all the activations and terminations of employment relationships (see e.g. Chelli et al. 2012) referred to firms located in the F. V. G. region from 2000 to 2012. The interest in this database is due to the existence of the date of activation, the date of termination, the type of contract and the economic sector of the firm for each employment relationship. This database is formed by 1565424 records; each record refers to an activation.

The first step was a reorganization of the graduates database, so for each fiscal code (person) there is only one record. The second step was the merging with the employment database using the fiscal code. The resulting database contains information related to the jobs and to the degrees of the person.

We selected only the jobs referred to graduates, so the dimension of the database decreased up to 62014 records. These are activations in firms located in F. V. G. region before or post degree.

For each person we only consider the highest degree (keeping out the doctorate) categorized into “short degree” (three years courses) and “long degree” (four, five or six years courses). The distance between the degree attainment date and the activation date is measured in days. We focused the attention on the first activation
post degree (15437 records) and the sectors of the firms in order to study the link between type of degree (scientific or humanistic) and sector of the firm. The variables choice has been determined by the object of the research *i.e.* the relation between the degree and the sector of firm in the first post graduate job. The variables concerning the graduate are sex, age, type of degree (scientific or humanistic), vote, waiting time (time between the degree achievement and the first activation) and work before degree (dichotomic variable). The variables concerning firms are schedule, firm location, position assigned to the graduate and economic activity sector. We recoded the sectors exploiting the Eurostat (Eurostat, 2014) aggregations of the manufacturing industry according to technological intensity (high-technology, medium high-technology, medium low-technology and low-technology) and of the services according to knowledge intensity (knowledge-intensive and less knowledge-intensive). We judged this classification more suitable than Ateco (Italian version of Nace) to study the transition to work of graduates. We also used the variable duration of the work relationship (the period between the date of activation and the date of termination).

### 3. Social Network Analysis and Two Step Cluster Analysis

As we said before the object of this research is to find relation(s) between the degree and the sector of the firm in the first post graduate job. After the pre-processing phases, the data consist of a finite set of units (the graduate) on which three relations (duration of the work relationship, waiting time, characteristics of work relationship) are measured, this forms a network.

A large network needs to individuate local clusters before that the internal structures of the clusters or relations among the clusters themselves may be analysed. Different methods for the individuation of adherent subgroups have been proposed in the field of social network analysis. They lead to: subgroups based on the complete reciprocity of the ties among the elements of the subgroup, subgroups based on the diameter and on the reaching of the units of the subgroups, subgroups based on the degrees of the nodes of the network, subgroups based on the comparison between the intern and extern ties of the members of a subgroup, subgroups based on the permutation of matrices and classical methods of data analysis as clustering algorithms (see Wasserman et al. (1994)).

Various clustering algorithms have been developed to group data into clusters however they work effectively either on numerical or categorical data but most of them perform poorly on mixed categorical and numerical data. Two step cluster analysis allows to avoid this problem (Bacher et al. 2004, Ming-Y et al. 2010). In
particular SPSS two step clustering developed by Chiu et al. 2001 for the analysis of large data sets consists of two steps:

**Step 1:** Pre-clustering of cases. A sequential approach is used to pre-cluster the cases. The aim is to compute a new data matrix with fewer cases for the next step; in order to reach this aim, the computed pre-clusters and their characteristics (cluster features) are used as new cases. The pre-clusters are defined as dense regions in the analyzed attribute space. The results may depend on the input order of cases therefore it is recommended to use random order.

**Step 2:** Clustering of cases. A model based hierarchical technique is applied. Similar to agglomerative hierarchical techniques, the pre-clusters are merged stepwise until all clusters are in one cluster. In contrast to agglomerative hierarchical techniques, an underlying statistical model is used. The model assumes that the continuous variables are within clusters independent normal distributed and the categorical variables are within clusters independent multinomial distributed. Two distance measures are available: euclidean distance and a log-likelihood distance in case of mixed types of attributes.

### 4. The application

We describe the results of the application to the graduates’ activations in 2009. The graduates and the firms are the units we consider while the relations are: waiting time, duration and characteristics of the relationship. In Table 1 we show the values (mean or percentage) of the variables for the population: the graduates activated in 2009 are mainly females, with “long degree”, with no work before the degree. The firms activity is mostly in services. As regards the ties most graduates have a full time schedule with a waiting time of 1.94 years.

**Table 1 – Variables used in the application. Mean values and percentage**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Graduates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>28.04</td>
<td></td>
</tr>
<tr>
<td>Final grades</td>
<td>103.65</td>
<td></td>
</tr>
<tr>
<td>Sex (females)</td>
<td></td>
<td>64.3%</td>
</tr>
<tr>
<td>Scientific degree</td>
<td></td>
<td>53.5%</td>
</tr>
<tr>
<td>“Long degree”</td>
<td></td>
<td>59.5%</td>
</tr>
<tr>
<td>No work before degree</td>
<td></td>
<td>58.1%</td>
</tr>
<tr>
<td><strong>Firms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-technology manufacturing industries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium high-technology manufacturing industries</td>
<td></td>
<td>2.1%</td>
</tr>
</tbody>
</table>
Ties
Waiting time (years)
- Length <6 months
- Length >3 years
Full time schedule

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Waiting time (years)</th>
<th>Duration of relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1 – The Clusters.

Cluster ONE
- More graduates
- Higher average waiting time (2.7 vs. 1.9 years)
- Length (3 years: 34% vs. 42.4% and
- Medium positions and full time schedule (96.9% vs. 66.1%)
- No work before degree (64.1% vs. 58.1%)

Cluster TWO
- More graduates
- Higher average waiting time (3 vs. 1.9 years)
- Length (3 years: 54.1% vs. 42.4% and
- Medium positions and full time schedule (96%) vs. 66.1%)
- Location in Trento province (57.8% vs. 47.8%)

Figure 1 – The Clusters.
5. Conclusions

The use of the two administrative databases presents two relevant aspects: we can analyse all the graduates and their first job, not only a sample, and can have precise measures of some variables (for example waiting time and duration of relationship are measured in days). The merging of databases is the starting point to
perform several analysis. In particular we focus our interest in economic sector activity and in ties between graduates and firms.

Two step cluster analysis allows to deal with categorical and quantitative variables simultaneously and consequently we can point out both quantitative and qualitative ties between network units. A result of the study is that these ties (waiting time, duration, timetable, position) are able to identify different categories of graduates and different categories of firms. Our future approach will be the study of the network on the bases of these ties.

References


A Statistical Analysis on the Transition from University to Labour Market

The aim of this study is to analyze some problems regarding the transition from University to Labour Market, in particular by exploring the ties between degree and economic sector of the first post-degree job. In order to perform the analysis we use two large administrative databases: the former concerning the graduates of University of Trieste, the latter the activations in Friuli Venezia Giulia region.

Our aim here is to explore such phenomena, in terms of the presence of clusters in terms of the ties between employers and graduates. The attention is focused on methods to highlight clusters and the ties.

After the integration of the two databases and the examination of data quality, we try to explore the existing ties between type of degree and economic sector.

Adriana MONTE, University of Trieste, adriana.monte@deams.units.it
Gabriella SCHOIER, University of Trieste, gabriella.schoier@deams.units.it