

DIPARTIMENTO DI  
ECONOMIA E FINANZA

**METODI E ANALISI  
STATISTICHE  
2023**



UNIVERSITÀ  
DEGLI STUDI DI BARI  
ALDO MORO



**DIPARTIMENTO DI  
ECONOMIA E FINANZA**

**METODI E ANALISI  
STATISTICHE  
2023**



Tutti i diritti di traduzione, riproduzione e adattamento, totale o parziale, con qualsiasi mezzo (comprese le copie fotostatiche e i microfilm) sono riservati

Toma E. (a cura di), *Metodi e Analisi Statistiche 2023*, Dipartimento di Economia e Finanza, Università degli studi di Bari Aldo Moro.

© Copyright 2023 by Università degli Studi di Bari Aldo Moro  
[www.uniba.it](http://www.uniba.it)

Prima edizione: dicembre 2023

ISBN: 978-88-6629-063-6

I contributi presentati nel presente volume sono stati tutti oggetto di valutazione interna

Editing finale a cura di E. Toma e F.D. d'Ovidio

# Sommario

Ernesto Toma <b>Presentazione</b> .....	pag. 5
Crescenza Calculli, Alessio Pollice, Angela Carluccio, Porzia Maiorano, Gianfranco D'Onghia <b>Modeling abundance of benthopelagic fauna using in-situ detection data...</b> «	7
Antonella Massari, Viviana D'Addosio, Vittoria Claudia De Nicolò, Samuela L'Abbate <b>Contact tracing dei dati sanitari e tutela della privacy: literature review</b> ..... «	23
Antonella Massari, Delia Vozza, Samuela L'Abbate <b>Blockchain, dati sanitari e tutela della privacy</b> ..... «	45
Laura Antonucci, Antonia Cofano, Corrado Crocetta, Massimo Russo <b>Un modello integrato di valutazione di servizi ICT e sanitari</b> ..... «	59
Maria Grazia Manco, Alessio Pollice, Crescenza Calculli, Nicola Ungaro, Antonietta Porfido <b>Analisi del livello di trofia delle acque marine pugliesi</b> ..... «	71
Giulio Fenicia <b>Le origini della Regia Aeronautica</b> .....	« 87
Domenico Tebala, Monica Carbonara <b>Un indicatore composito per descrivere i cambiamenti climatici in Europa prima e dopo il Covid-19</b> ..... «	113
Marina Annese, Nunziata Ribecco, Crescenza Calculli, Angela Maria D'Uggento <b>Quale futuro? Analisi sulle prospettive future degli studenti di alcune scuole secondarie superiori della Puglia</b> ..... «	123
Alessia Orrico, Angela Maria D'Uggento, Nunziata Ribecco <b>La percezione della donazione tra i giovani: un'indagine nelle scuole pugliesi</b> ..... «	143
Ilaria Pepe, Ernesto Toma <b>A systematic review of standard diagnostic criteria for Cognitive Impairment No Dementia (CIND), risk factors and conversion rates</b> ..... «	159

Diego Battagliese, Alessio Pollice <i>Valutazione dei divari educativi tra le regioni italiane mediante la stima per piccole aree</i> .....	« 175
Sabrina Diomede <i>A note on the choice-based approach in the theory of individual decision-making</i> .....	« 191
Rosa Ceglie, Angela Maria D'Uggento <i>Il fenomeno delle università telematiche in Italia</i> .....	« 199
Margaret Antonicelli, Angela Maria D'Uggento, Raffella Girone, Claudia Marin <i>Investigating the ICT skills of Italian graduates' intellectual migration</i> .....	« 219
Elena Equatore, Angela Maria D'Uggento, Irene Brescia, Mariabeatrice Principi <i>La qualità della vita dei pazienti affetti da infiammazione intestinale</i> .....	« 233
Francesco Grimaldi <i>La dinamica degli assetti proprietari delle società quotate italiane nel decennio 2012-2021. Alcune evidenze empiriche</i> .....	« 249

# Investigating the ICT skills of Italian graduates' intellectual migration

Margaret Antonicelli<sup>1</sup>, Angela Maria D'Uggento<sup>2</sup>,  
Raffaella Girone<sup>3</sup>, Claudia Marin<sup>3\*</sup>

<sup>1</sup>*Department of Statistical Science, La Sapienza, University of Rome,*

<sup>2</sup>*Department of Economics and Finance, University of Bari Aldo Moro,*

<sup>3</sup>*Department of FORPSICOM, University of Bari Aldo Moro*

**Abstract:** The intellectual migration of Italian graduates is a significant and enduring trend, marked by patterns shaping the national academic landscape. The migration primarily flows from southern to northern Italy and then extends to European countries or overseas, particularly prominent in STEM fields. Graduates seek enhanced professional opportunities and career prospects in more prosperous regions. Understanding these migration characteristics is crucial for addressing brain drain factors in Italy and beyond. This study focuses on Italian Master's graduates in 2021-2022, analyzing a dataset of 2,137 CVs from Almalaurea. Various methods, including machine learning and text mining, were used to extract relevant information, including sociographic characteristics (such as gender, place of residence, degree grade, academic discipline, university attended, etc.) and the graduates' educational background. Significant differences were found when the geographical location of the graduates (the university from which they graduated) was taken into account. Therefore, the graduates were divided into three groups for the statistical analysis: Southern, Central and Northern Italy. This cluster approach allowed a comparative study of individual skills (e.g. language and computer skills, Erasmus experience) and future career prospects. These results are crucial for the development of targeted policies and initiatives that effectively counteract brain drain and create an environment that retains and attracts highly skilled workers

**Keywords:** intellectual migration; IT skills; POSET; Hasse diagram; text mining.

---

\* Corresponding author: [claudia.marin@uniba.it](mailto:claudia.marin@uniba.it)

## 1. Introduction

In the context of skill evaluation theories, the skills approach is an effective tool for the management, development and enhancement of human resources, especially within the current organizational contexts. These are increasingly streamlined and de-verticalised, process-oriented rather than individual functions and in which it is necessary to plan and manage horizontal and inter-functional careers of figures in which the professional component is growing.

Companies must be able to attract and adequately value human resources, on which the success of the organization itself depends; the competitive advantages of an organization are in fact increasingly linked to internal expertise and to the skills acquired and developed by these resources.

A first definition of competence takes shape in the 70s by David McClelland who defines it: “a personal characteristic that allows the excellent performance of a specific job in a specific company”. In this definition, competencies are described as knowledge, skills, behaviors and attitudes, linked to the individual, which determine an excellent performance in a specific context such as the job role performed within a given company.

In the 1980s Richard Boyatzis, a collaborator of McClelland, defined competence as: “an intrinsic characteristic of an individual causally linked to effective or superior performance in a task” (Boyatzis, 2008).

Once again, we talk about knowledge, skills, motivations, and aspects of one's self-image related to successful business performance. However, with Boyatzis, skills are considered as a generic personal attribute, which, although always connected to an effective or superior performance, is influenced by the psychological aspect of the person, presenting the lack of an important objective aspect of evaluation. The meaning of the term competence has transformed over time and today an interesting description defines it as: “a set of behaviors expressed by the individual deriving from the possession and application of theoretical knowledge, skills, attitudes and mental orientations”. This definition has the advantage of enhancing the manifest and therefore measurable aspect of competence, i.e. the organizational behaviors acted out by each resource in their daily work, behaviors that become objectifiable as they are measurable. It is precisely this objective meaning that allows the creation of a system of evaluation of skills understood as a set of technical and behavioral dimensions that are decisive for a successful performance.

As far as the skills assessment process is concerned, there are different methodologies that are distinguished by the object on which they focus, but what is fun-

damental, whatever the approach used and the object of investigation, is that the contextualization of skills is contextualized. The space-time context can radically change the value of a skill.

To implement and define this analysis methodology, it is necessary to define the expected competence profiles in terms of set of behavioral skills and set of technical knowledge referring to areas of responsibility and performance indicators contained in a good job description of the position and examining.

Linked to the issue of skills analysis is the issue of job migrations linked to the skills miss match. There is currently no detailed and substantial literature on the subject. Some articles assume that migration must precede research or – less often – that research must precede migration. McCall and McCall (1987) assume such migration (between cities) must precede the job search. Topel (1986) allows workers to move but not search anywhere meaningful sense.

Coulson and Wang (2001) model search in a single metropolitan area and allow agents to do so look in the central business district or the suburbs, but not both at the same time (although they can research – and work – in both markets with a commuting cost); they argue that global search will never happen.

Fahr and Sunde (2006) adapt a research model to allow for endogenous interregional migration, but as in the case the above docs assume that migration should precede research. Spilimbergo and Ubeda (2004) develop a migration model where it focuses on double matching in the labour market and marriage market.

The aim of this work is to define how training and different skill levels, especially in the IT and technological context, can determine and influence internal migratory flows for work in Italy.

In particular, in this work the process of summarizing the information taken from the AlmaLaurea CVs will be shown; the synthesis method used will be the POSET, a non-aggregative approach that allows the identification of profiles that are comparable and not each other.

## **2. Data and methods**

The level of knowledge of IT tools is one of the most important indicators of the preparation achieved by students at the end of their university studies. Among other things, the importance of IT knowledge emerges from the results of the Survey on the Employment Condition of Graduates where it is observed that the probability of being employed, among those who know at least five IT tools, is 29.2%

higher than who knows at most two instruments (AlmaLaurea, 2022). At the end of their university studies, students provide a self-assessment of their level of knowledge of ten IT tools, on a five-mode scale "none", "limited", "fair", "good", "excellent".

The level of knowledge of IT tools varies within the different disciplinary groups. However, it needs to be clarified that skills relating to internet browsing and network communication, word processors, spreadsheets, tools presentation and operating systems tend to be more transversal to the different disciplinary fields, while the remaining skills (programming languages, database, implementation websites, data transmission networks, drawing and assisted design) are strongly affected by the specificities linked to the course of study. In fact, over 60% graduates from the IT and ICT technologies group declares to have "at least good" knowledge of all IT skills mentioned, except for the tools drawing and assisted planning (11.3%). Similarly, are the graduates of the architecture and civil engineering and engineering group industrial and information technology to possess the highest levels of knowledge of drawing and assisted design (respectively for 87.7% and 32.8%) (AlmaLaurea, 2022).

This section of the Almalaurea questionnaire collects information on technical and specialist knowledge in the ICT (Information and Communication Technologies) field. The identification of specific ICT skills is often used by companies in selection processes, and for this reason the section has been structured in such a way as to allow the inclusion of timely information.

The IT skills analyzed are the following:

- a) Office Automation
- b) Application software
- c) Systems and Network Management
- d) Programming
- e) Data Management
- f) Graphics and Multimedia
- g) Safety

Sub-categories have been associated with each category:

- a) *Office Automation*
  - Office suite
  - Spreadsheets
  - Digital communication

- Presentation software
- Web Browsers
- Word processing
- OCR
- Desktop publishing
- b) *Application software*
  - E-learning platforms
  - Project Management
  - Data Visualisation
  - Structural Calculation
  - CMS - Web publishing
  - Statistic analysis
  - ERP
  - Numerical analysis
  - Use of CAD software
  - Use of CAE software
  - Use of CAM software
  - Geographic Information System (GIS)
- c) *Systems and Network Management*
  - Network architectures
  - Network protocols
  - Network components
  - Operating systems
  - Embedded systems
  - Middleware
  - IoT
- d) *Programming*
  - Web programming
  - Build Automation
  - Video game creation systems
  - Development Models
  - Software modeling languages
  - Software Testing Tools
  - Javascript libraries
  - Client/Server Applications
  - Programming languages

- Markup languages
  - Mobile application development
  - Firmware and software for industrial electronics
  - Real-Time Systems
  - Parallel computing
  - Integrated Development Environments (IDEs)
- e) Data Management*
- Database Design
  - Data modeling tools
  - Database Management Systems (DBMS)
  - ETL tools
  - Query languages
  - Big Data
  - Blockchain
- f) Graphics and Multimedia*
- Web graphics
  - Raster graphics editors
  - Vector graphics editor
  - Rendering
  - 3D graphics
  - Creation of animated narratives
  - Video Processing / Editing
  - Audio Processing / Editing
  - Streaming and Video Conferencing
- g) Safety*
- Safety standards
  - Penetration testing
  - Cryptography
  - Anti-SPAM
  - Firewalls
  - Backup and Recovery
  - Virtual Private Network (VPN)
  - Antivirus

The analyzed sample includes 2,137 curricula relating to the 2021-2022 academic year. The sample was defined through random sampling.

## **2.1 Statistical methods**

The measurement of the self-evaluative (or cognitive) dimension has a consolidated tradition in social statistics and is expressed through satisfaction with certain aspects of life. In the self-evaluative process, the interviewee carries out a comparison with a condition that he considers suitable or desirable, considering his own training, comparison with colleagues, aspirations, results achieved by other people who consider themselves closer and more significant (Michalos, 1985). As previously mentioned, in the present case under study, students were asked to evaluate their IT skills on an ordinal scale composed of 5 values. By converting this scale into values, a possible score is obtained for each evaluation area ranging from 1 to 5. For a synthetic representation of these items, the arithmetic mean is often used.

However, the compensatory method is absolutely not correct because it is not a correct method, since we are dealing with ordinal and not continuous data. Therefore, a non-compensatory approach is required, which is suitable for analyzing ordinal data.

In this work, the Partially Ordered Set (POSET) approach was considered suitable (Maggino, 2017; Fattore, 2016). A partial order set is an arrangement such that, for certain pairs of elements, one precedes the other. The word partial is used to indicate that not every pair of elements needs to be comparable; that is, there may be pairs for which neither element precedes the other. Partial orders thus generalize total orders, in which every pair is comparable.

Formally, a partial order is a homogeneous binary relation that is reflexive, transitive and antisymmetric. A partially ordered set (POSET for short) is a set on which a partial order is defined.

The simplification of complex and unobservable concepts is the main reason why this kind of methodology has gained a vivid interest in applied sciences. The chosen method takes into account the characteristics of each item and consequently safeguards the informative role of each dimension.

A POSET  $P = (X; \leq)$  is a set  $X$  equipped with a partial order relation, i.e. with a binary relation satisfying the properties of reflexivity, antisymmetry and transitivity (Fattore, 2016):

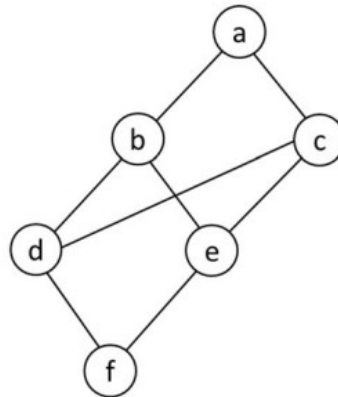
1.  $x \leq x$  for all  $x \in X$  (reflexivity);
2. if  $x \leq y$  and  $y \leq x$  then  $x = y$ ,  $x, y \in X$  (antisymmetry);
3. if  $x \leq y$  and  $y \leq z$ , then  $x \leq z$ ,  $x, y, z \in X$  (transitivity).

If  $x \leq y$  or  $y \leq x$ , then  $x$  and  $y$  are called comparable, otherwise they are called incomparable (written  $x \parallel y$ ). A partial order  $P$  where any two elements are comparable is called a linear order or a complete order or a chain. On the contrary, if any two elements of  $P$  are incomparable, then  $P$  is called an antichain. Given  $x, y \in P$ ,  $y$  is said to cover  $x$  (written  $x < y$ ) if  $x \leq y$  and there is no other element  $z \in P$  such that  $x \leq z \leq y$  (Fattore, 2013).

A finite POSET  $P$  can be easily depicted by means of a Hasse diagram. Hasse diagrams are directed acyclic graphs representing the cover relation generating the partial order; they are drawn according to the following two rules: (1) if  $x \leq y$ , then node  $y$  is placed above node  $x$ ; (2) if  $x < y$ , then an edge is inserted linking node  $y$  to node  $x$ . By transitivity,  $x \leq y$  (or  $y \leq x$ ) in  $P$ , if and only if in the Hasse diagram there is a descending path linking the corresponding nodes; otherwise,  $x$  and  $y$  are incomparable. Examples of Hasse diagrams are reported in Figure 1.

An element  $x \in P$  is called maximal if  $x \leq y$  implies  $x = y$ . An element  $x$  such that  $y \leq x$  for any  $y \in P$  is called (the) maximum (or top, or greatest element) of  $P$ . Dually,  $x$  is minimal if  $y \leq x$  implies  $x = y$ ; if  $x \leq y$  for any  $y \in P$  then it is called (the) minimum (or bottom, or least element) of  $P$ . Any finite POSET has both maximal and minimal elements, but not necessarily maximum and minimum.

**Figure 1.** *Hasse Diagram example with six elements*



### 3. Results

As for the main characteristics of our sample, we summarize the parameters that are of interest to us. From a territorial point of view, 46% of respondents belong to universities in Northern Italy, 32% attend universities in Central Italy, while the

remaining students are enrolled in universities in Southern Italy. It is important to underline that almost 50% of students attending universities in Northern Italy register a domicile in the regions of Central and Southern Italy.

In terms of gender, the sample is evenly distributed with 50.9% men and 49.1% women.

As regards faculties of origin, over 27% come from various engineering fields, with computer science standing out, 21.4% from natural science faculties (mathematics, physics, chemistry, statistics and economics, etc.), just over 15% from law and political science and 13.5% from medicine and dentistry.

If we look at the constructed indicator for self-assessment of IT skills, we can see that young graduates rate their knowledge in the IT field very highly. However, this contradicts the objective text mining analyzes trivially performed on the basis of the completed questionnaire and the analysis of the profiles of a sub-sample of these students (280 units) through the web scraping technique on their LinkedIn profiles.

The analyzed group of students is partially differently ordered in the different dimensions of self-assessment of existing IT skills. Once we have defined the characteristics of the students in the sample in relation to the parameters of interest, it is important to define the incidence matrix. We then proceed to the construction of the Hasse diagram. This approach involves, given the values measured for each company and for each parameter analyzed, identifying all the possible profiles that can be delineated for the two indicators defined in our study (Rubino et al, 2023). In particular, 169 possible profiles were identified in relation to the IT skills possessed. Apart from the difference between the possible profiles that we could have had from a theoretical point of view (given by the linear combination of the elements) and the real profiles identified in our study, the difference detected in the results depends on an important difference both in the parameters analyzed and in their quantification and, as we will see later, in a situation of important incompatibility (Maggino, 2017) between subjects in relation to IT skills.

The measurement of the two indicators analyzed is therefore based on the POSET of all possible profiles (Rubino et al., 2023). Another essential element of the analysis was the identification of the thresholds.

This stage of the analysis inevitably has a subjective component, which, contrary to popular belief, is not necessarily a negative component (Rubino et al., 2023).

As this is a new and innovative approach, there is a lack of sound scientific literature on universal criteria for threshold identification. In view of this, we adopted the approach defined by Arcagni et al. (2019).

When analysing the POSET, we proceed to identify a subset  $\pi_l$  of incomparable profiles (lower threshold); all profiles at  $\pi_l$  or below one of its elements are classified as unsatisfied. At the same time, let's define another subset  $\pi_c$  (upper threshold) such that the profiles in  $\pi_c$  or above one of its elements are considered fully satisfied. As shown in Table 1, we have defined low and high thresholds.

In comparison to the two low thresholds, we also considered the average values (6655655, 6566555 and 7144554) in addition to the profile with the lowest recorded proficiency level (represented by profiles 4126655 and 4114433, respectively).

The approach was also similar for the high thresholds: the highest profile was added, which practically represents people who consider their IT knowledge to be more than optimal. Specifically, the following thresholds were identified: 10 10 10 10 10 10 10, 10 9 10 9 9 10 10, 10 10 9 9 9 10 9 and 10 9 10 10 10 10 9.

The identification of the medium-high and medium-low thresholds derives mainly from the type of studies carried out, even if there has not always been a close relationship between the two parameters.

**Table 1.** Lower and upper thresholds of the IT skills level of perception

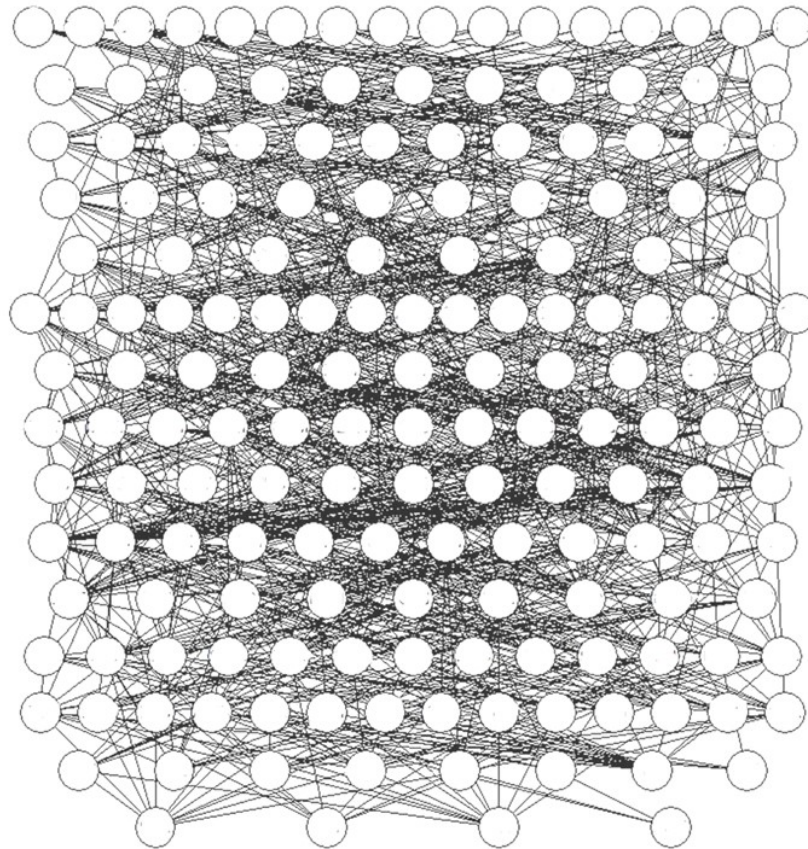
POSET	Lower thresholds	Upper thresholds
IT skills level of perception	4126655 - 4114433 6655655 - 6566555 - 7144554	10 10 10 10 10 10 10 10 9 10 9 9 10 10 10 10 9 9 9 10 9 10 9 10 10 10 10 9

The Hasse diagram of the perception of IT skills is shown in Figure 2. In this diagram, the high and medium-high profiles are shown in the upper part, while the lowest and medium-low profiles are shown in the lower part.

The results obtained with the Hasse diagram show that the situation is extremely articulated. As already mentioned, with this type of statistical approach, the number of possible profiles that can be generated by combining the responses to each of the elements that make up the indicator does not, in most cases, match the actual number of profiles that appear in the empirical analysis. The presence of a high number of profiles identified indicates a significant and not positive heterogeneity among students. This, as already mentioned, largely depends on the different types of studies carried out, together with an unequal self-perception of one's skills in this sector. The portion of students who admitted having a medium-low level of IT skills is decidedly low; these students are represented in the lower part of the Hasse diagram.

The non-objectivity in the recognition of one's skills subsequently affects the mismatch between supply and demand in the world of work or in a continuous and constant search for the ideal job.

**Figure 2.** *The Hasse diagram.*



#### **4. Final remarks**

The analysis of IT capabilities is a topic that has become of great interest over the last decade, both in the field of academic studies and in the corporate world.

A constant and continuous evolution of technologies, an important push towards increasingly digitalized and innovative processes have led students to apply to faculties or specialization and training courses that have these characteristics.

However, the reality is that innovation and digitalization processes, as global

and widely recognized as they are, have a strong territorial connotation in the Italian context and therefore tend to generate significant internal migration flows.

The mismatch between job supply and demand in southern Italy has inevitably led to a continuous exodus of numerous students over the years.

This analysis has not only confirmed what has just been said, but has also highlighted a not very positive aspect: the incorrect assessment of one's own abilities.

The cross-checks have revealed an overestimation of one's own skills, which inevitably leads to dissatisfaction on the part of the companies that continue with the recruitment or that are forced to compensate for the employee's lack of knowledge. In view of what has just been said, we consider it essential to carry out a new study that takes into account the same parameters but decodes the information in an objective way using machine learning techniques.

## References

- Arcagni, A., Barbiano di Belgiojoso, E., Fattore, M., Stefania, M. Rimoldi, L. (2019). Multidimensional analysis of deprivation and fragility patterns of migrants in Lombardy, using partially ordered sets and self-organizing maps. *Social Indicator Research*, 141, 2: 551-579.
- Boyatzis, R.E. and Saatioglu, A. (2008). A 20-year view of trying to develop emotional, social and cognitive intelligence competencies in graduate management education. *Journal of Management Development*, Vol. 27 No. 1, pp. 92-108. <https://doi.org/10.1108/02621710810840785>.
- Consorzio Almalaurea (2022) Indagine sul Profilo dei Laureati.
- Coulson N. E., Laing D., Wang P. (2001), Spatial Mismatch in Search Equilibrium, *Journal of Labor Economics*, vol. 19, No. 4, University of Chicago Press.
- Fahr R., Sunde U. (2009). Did the Hartz reforms speed-up the matching process? A macro-evaluation using empirical matching functions, *German Economic Review* 10(3): 284–316, Wiley.
- Fattore, M. (2013), “La valutazione statistica del benessere: ordinamenti parziali e nuovi paradigmi metodologici”, *Statistica e Società*, Vol. 2 No. 2, pp. 10-13.
- Fattore, M. (2016). Partially ordered sets and the measurement of multidimensional ordinal deprivation. *Social Indicators Research*, 128(2), 835–858.
- Maggino, F. (2017), “Developing indicators and managing the complexity”, in Maggino, F. (Ed.), *Complexity in Society: From Indicators Construction to Their Synthesis*, Springer, Cham, pp. 87-114.
- McCall B. P., McCall J. J. (1987). A Sequential Study of Migration and Job Search, *Journal of Labor Economics*, Vol. 5, No. 4, Part 1, University of Chicago Press.

go Press.

McClelland, D.C. (1975). *Power: The Inner Experience*. New York: Irvington Publishers, Halstead Press.

Michalos, A (1985) Multiple Discrepancies Theory (MDT). *Social Indicators Research* 16(4):347-413. DOI: 10.1007/BF00333288.

Rubino M., Maggino F., Antonicelli M. (2023). Measuring human IT agility and firms' digitalization using POSET: evidence from Italy. *International Journal of Quality & Reliability Management*, Emerald Publishing, <https://doi.org/10.1108/IJQRM-09-2021-0321>.

Spilimbergo A., Ubeda L. (2004). A model of multiple equilibria in geographic labor mobility, *Journal of Development Economics*, Vol. 73, Issue 1, Elsevier.

Topel R. H. (1986). Local Labor Markets, *Journal of Labor Economics*, Vol. 94, No. 3, Part 2, University of Chicago Press.

