# IES 2022 Innovation & Society 5.0: Statistical and Economic Methodologies for Quality Assessment

# BOOK OF SHORT PAPERS

Editors: Rosaria Lombardo, Ida Camminatiello and Violetta Simonacci

Book of Short papers
10th International Conference **IES 2022**
Innovation and Society 5.0: Statistical and Economic Methodologies for Quality Assessment

Department of Economics, University of Campania "L. Vanvitelli",
January 27th - 28th 2022

**Scientific Committee of the group of the Italian Statistical Society on Statistics for the Evaluation and Quality of Services – SVQS**

Pietro Amenta -University of Sannio
Matilde Bini -European University of Roma
Luigi D'Ambra -University of Naples "Federico II"
Maurizio Carpita -University of Brescia
Paolo Mariani -University of Milan "Bicocca"
Marica Manisera -University of Brescia
Monica Palma -University of Salento
Pasquale Sarnacchiaro -University of Rome Unitelma Sapienza

**Program Committee of the conference IES 2022**

Chair: Rosaria Lombardo -University of Campania "L. Vanvitelli"
Fabio Bacchini -ISTAT
Laura Baraldi -University of Campania "L. Vanvitelli"
Eric Beh -University of Newcastle, Australia
Wicher Bergsma -The London School of Economic and Political Science, UK
Enrico Bonetti -University of Campania "L. Vanvitelli"
Eugenio Brentari -University of Brescia
Clelia Buccico -University of Campania "L. Vanvitelli"
Rosalia Castellano -University of Naples "Parthenope"
Ida Camminatiello -University of Campania "L. Vanvitelli"
Carlo Cavicchia -University of Rotterdam, The Netherlands
Enrico Ciavolino -University of Salento
Corrado Crocetta -University of Foggia
Claudio Conversano -University of Cagliari
Antonello D'Ambra -University of Campania "L. Vanvitelli"
Antonio D'Ambrosio -University of Naples "Federico II"
Alfonso Iodice D'Enza -University of Naples "Federico II"
Tonio Di Battista -University of Chieti "G. D'Annunzio"
Michele Gallo -University of Naples "L'Orientale"
Francesco Gangi -University of Campania "L. Vanvitelli"
Michele La Rocca -University of Salerno
Amedeo Lepore -University of Campania "L. Vanvitelli"
Riccardo Macchioni -University of Campania "L. Vanvitelli"
Filomena Maggino -University of Rome "La Sapienza"
Angelos Markos -Demokritus University of Thrace, Greece
Lucio Masserini -University of Pisa
Stefania Mignani -University of Bologna
Nicola Moscariello -University of Campania "L. Vanvitelli"
Francesco Palumbo -University of Naples "Federico II"
Alessandra Petrucci -University of Florence
Alessio Pollice -University of Bari

Donato Posa -University of Salento
Luca Secondi -University of Tuscia
Amalia Vanacore -University of Naples "Federico II"
Michel van de Velden -University of Rotterdam, The Netherlands
Rosanna Verde -University of Campania "L. Vanvitelli"
Donatella Vicari -University of Rome "La Sapienza"
Grazia Vicario -Polytechnic University of Turin
Maurizio Vichi -University of Rome "La Sapienza"

**Organizing Committee**

Ida Camminatiello -University of Campania "L. Vanvitelli"
Antonello D'Ambra -University of Campania "L. Vanvitelli"
Rosaria Lombardo -University of Campania "L. Vanvitelli"
Elvira Romano -University of Campania "L. Vanvitelli"
Luca Rossi -University Cusano
Violetta Simonacci -University of Naples "L'Orientale"

# Editors

Rosaria Lombardo - University of Campania “L. Vanvitelli”, Italy
Ida Camminatiello - University of Campania “L. Vanvitelli”, Italy
Violetta Simonacci - University of Naples “L’Orientale”, Italy

PKE - Professional Knowledge Empowerment s.r.l.
Sede legale: Villa Marelli - Viale Thomas Alva Edison, 45 - 20099 Sesto San Giovanni (MI)
Sede operativa: Villa Marelli - Viale Thomas Alva Edison, 45 - 20099 Sesto San Giovanni (MI)
Ufficio Di Rappresentanza: Via Giacomo Peroni, 400 - 00131 Roma (RM)
CF / P.I. 03167830920 — www.pke.it; e-mail info@pke.it — Privacy

February 2022 PKE s.r.l.
ISBN 978-88-94593-35-8 on print
ISBN 978-88-94593-36-5 online
All rights reserved.
This work is protected by copyright law.
All rights, in particular those relating to translation, citation, reproduction in any form, to the use of illustrations, tables and the software material accompanying the radio or television broadcast, the analogue or digital recording, to publication and dissemination through the internet are reserved, even in the case of partial use. The reproduction of this work, even if partial or in digital copy, is admitted only and exclusively within the limits of the law and is subject with the authorization of the publisher. Violation of the rules involves the penalties provided for by the law.
PKE Publisher

# Preface

This Book of Short Papers includes all peer-reviewed long-abstracts submitted to the IES2022 conference, titled "Innovation & Society 5.0: Statistical and Economic Methodologies for Quality Assessment", held at the University of Campania "L. Vanvitelli" on January 27-28, 2022. IES2022 is the 10th meeting of the biennial international conference proposed by the permanent group Statistics for the Evaluation and Quality in Services (SVQS) of the Italian Statistical Society (SIS). The SVQS group, born in 2004, focuses on national research programs and applied research activities, on statistical methods and methodologies for the evaluation of the quality of services in public and private fields. For further information, please visit https://www.svqs.it/. IES2022 has been sponsored by the Italian Statistical Society (SIS), the European Network for Business and Industrial Statistics (ENBIS), and the International Association for Statistical Computing (IASC). In addition, also the two SIS groups Statistics and Data Science (SDS) and Enhancement of Public Statistics (VSP) actively supported the conference. IES2022 aims at stimulating a scientific debate on the challenges of Society 5.0 with respect to quality assessment. The conference provides an important moment of reflection for the development of new ideas and methodologies by promoting the rethinking of the open issues in service evaluation within the new paradigm of an interconnected cyber-social system. Service quality assessment represents the starting point for the development of effective policies for private and public institutions, which is crucial for the development of society. Big data, heterogeneous multi-layered structure and designs, cutting-edge analytical tools, and advanced data harvesting techniques have become fundamental for research; nonetheless, they require a continuous effort in terms of proper treatment, interpretation, and supervision to ensure the centrality of human and social problems. In this perspective, IES 2022 main goals are:

- to promote and coordinate the statistical and economic methodologies for the evaluation of a human-centered society emphasizing how statistical thinking, design, and analysis may be of use to a Society 5.0;
- to foster advanced methodological research supporting the assessment of the quality of social services;

- to be a platform where the experts of Statistics, Data Mining, Data Science, Machine Learning, and related disciplines meet for analyzing Big Data.

The high turn-out of the conference, with a total of 107 presentations organized in 22 solicited sessions and 11 contributed sessions, two plenary talks, and the participation of over 300 authors, made evident a very alive interest in evaluation topics. Previous IES editions include:

- IES2009 was held at the University of Brescia (June 24-26, 2009) with selected papers published in special issues of Electronic Journal of Applied Statistical Analysis (EJASA) and Statistica & Applicazioni;
- IES2011 was held at the University of Florence (May 30 – June 1, 2011) with selected papers published in a special issue of the Journal of Applied Quantitative Methods;
- IES2013 held at the University of Milan "Bicocca" (December 9 – 13, 2013) with selected papers published in the Procedia Economics & Finance (Elsevier Publisher);
- IES2015 was held at the University of Bari "Aldo Moro" (June 8 – 9, 2015) with selected papers published in a special issue of Quality & Quantity;
- IES2017 held at the University of Naples "Federico II" (September 6 – 7, 2017) with selected papers published in special issues of Social Indicator Research, Quality & Quantity, and EJASA;
- IES2019 was held at the European University of Rome (July 4 – 5, 2019) with selected papers published in special issues of Socio-Economic Planning Science and EJASA.

All IES2022 contributions are based on the development of innovative statistical methodologies or interesting applications. The topics covered in the numerous presentations range over the following fields: Sustainability, Health, Wellness, Sport, Tourism, Education, Training and Research, Bank and FinTech, Transportation, Environment, Enterprise, Cultural changes and values, Industry and Finance, E-commerce, Digital Marketing, Labour Market, Public Administration, Advertising, Political preferences, Justice System. Several short papers deal with the shock of the COVID-19 pandemic and its impact in different areas such as poverty and sustainability, education and distance learning, student satisfaction, environment, health services, and social interactions. From a methodological standpoint, many of the short papers deal with challenging structures such as high-dimensional data, complex survey designs, constrained variability, sparsity, multicollinearity, and multidimensional longitudinal series. A wide range of statistical tools and models have been employed, including functional data analysis, various types of regression models (high-dimensional, logit, quantile, OLR, LASSO, etc.), machine learning algorithms for classification, methods for multi-way data and contingency tables,

generalized discriminant analysis, multidimensional Item Response Theory, PLS-SEM, advanced visualization techniques, compositional data analysis, Bayesian methods and so on. Extended versions of selected IES2022 papers will be included in a special issue of the Computational Statistics Journal titled "High-dimensional Data Analysis and Visualisation to Assess Service Quality" and of Annals of Operations Research Journal, titled "Statistical Methods and Data-Driven for Decision Making in Public Sector".

Rosaria Lombardo, Ida Camminatiello and Violetta Simonacci
*Editors*

Luigi D'Ambra
*Onorary Coordinator of SVQS Group*
Maurizio Carpita
*Coordinator of SVQS Group*

# Contents

**Solicited Session SS1** – *Compositional Data Analysis*
**Organizer and Chair: Gianna Monti** **1**
Rieser C. and Filzmoser P.. *Compositional Data and graph theory* . . . . 2
Egozcue J. J., Pawlowsky-Glahn V. and Buccianti A. *Compositional deviations from linear and non-linear equilibria* . . . . . 8
Monti G. S. and Filzmoser P. *The knockoff filter for FDR control in robust ZeroSum regression in microbiome analysis* . . . . . 14

**Solicited Session SS2 -** *Monitoring progress towards SDGs: statistical approaches and methods for measuring poverty, inequalities and food insecurity*
**Organizer and Chair: Luca Secondi** **18**
Marchetti S., Giusti C., Pratesi M. and Biggeri L. *Poverty indicators adjusted using local price indexes* . . . . . 19
Tonutti G., Bertarelli G., Giusti C. and Pratesi M. *Assessing the targeting of the anti-poverty measure "Reddito di Cittadinanza" using Small Area Estimation methods* . . . . . 25
Vargas-Lòpez A. and Secondi L. *Household consumption and food insecurity in Mexico: COVID19 and sustainable development* . . . . . 31

**Session of free contributes SCL1** – *Big Data, Proximity data, Multi way data*
**Chair: Donatella Vicari** **37**
Metulini R. and Carpita M. *Forecasting Traffic Flows with Complex Seasonality using Mobile Phone Data* . . . . . 38
Simonacci V., Menini T. and Gallo M. *CP decomposition of 4th-order tensors of compositions* . . . . . 44
Bove G. *A strategy of analysis of symmetry and skew-symmetry in asymmetric relationships* . . . . . 50

**Session of free contributes SCL2** –*Industry and Society*

**Chair: Germana Scepi** **56**

Crosato L., Domenech J. and Liberati C. *Toward an early detection of SME's default with websites' indicators* . . . . . 57

Angelone R. *Italians' Culture and Values after two years of pandemic* . . 63

Aria M., Cuccurullo C., D'Aniello L. and Spano M. *Thematic evolution of Academic Medical Centers' research: a focus on Italian public owned AOUs in metropolitan areas* . . . . . 67

**Session of free contributes SCL3** – *Society and Tourism*
**Chair: Emma Zavarrone** **73**

Rondinelli R., Palazzo L. and Ievoli R. *Local clustering coefficient to measure intra-regional tourism in Italy* . . . . . 74

Zavarrone E., M. and Forciniti A. *Local clustering coefficient to measure intra-regional tourism in Italy* . . . . . 80

Firza N. *Sustainable tourism: The case of Albania* . . . . . 86

Romano M., Zammarchi G. and Conversano C. *Threshold-based Naïve Bayes Classifier Customer Satisfaction evaluation* . . . . . 90

**Session of free contributes SCL4** –*Society and Innovation*
**Chair: Michelangelo Misuraca** **95**

Seri E., Alaimo L. S., di Bella E., Cataldo R. and Piscitelli A. *Migrant Integration Policy Index (MIPEX): an analysis of countries via Gaussian mixture modelbased clustering* . . . . . 96

Rossi L. and Daddi S. *The measure of the BES: a proposal for the aggregation of the indicator education and training* . . . . . 102

Marino M., Mazza R., Misuraca M. and Stavolo A. *Monitoring consumer sentiment using control charts* . . . . . 108

**Session of solicited contributes SS3** – *Statistical methods for the assessment of student careers in higher education*
**Organizer and Chair: Maria Prosperina Vitale** **114**

La Rocca M., Niglio M. and Restaino M. *Predicting university students' churn risk* . . . . . 115

Primerano I., Santelli F. and Usala C. *Discovering archetypal universities in Italian higher education mobility flows* . . . . . 121

Cascella C. and Ragozzini G. *Measuring quality of students' careers in Higher Education: a systematic literature review* . . . . . 127

Porcu M., Sulis I. and Usala C. *Estimating the peers effect on students' university choices* . . . . . 134

**Session of solicited contributes SS4** – *Conformity assessment and quality predictions- itENBIS*
**Organizer and Chair: Amalia Vanacore** **140**
Vanacore A., Pellegrino M. S. and Ciardiello A. *Testing the predictive performance of multi-class classifiers* . . . . . 141
Borgoni R., Gilardi A. and Zappa D. *Optimal Subgrids from Spatial Monitoring Networks* . . . . . 148
Pennecchi F. and Kuselman I. *Extension of the JCGM 106:2012 - Conformity assessment of multicomponent items and finite statistical samples* . . . . . 153

**Session of solicited contributes SS5** – *Big Data and Large-dimensional Data*
**Organizer and Chair: Stefania Mignani** **160**
Farné M. and Vouldis A. *ROBOUT: a conditional outlier detection methodology for large-dimensional data* . . . . . 161
Camillo F. *Behaviours, emotions and opinions in modern citizen or customer relationship systems: a correct integration of small and big data for hyper-targeting, personal advertising and look-alike* . . . . . 168
Camminatiello I. and Lucadamo A. *A model for assessing sea environmental quality* . . . . . 172

**Session of solicited contributes SS6** – *Health Quality*
**Organizer and Chair: Paolo Mariani** **178**
Bartolini B., Bertoldi S., Benedan L., Galeone C., Mariani P., Sofia F. and Zenga M. *The uneasiness index in a patient-designed quality of life questionnaire* . . . . . 179
Benedan L., Hachem M. E., Galeone C., Mariani P., Pilo C. and Tadini G. *Assessing the Quality of Life of patients with Epidermolysis Bullosa (EB): Development of a patient-centered questionnaire* . . . . . 184
Marletta A. and Morandi M. *Survival analysis in a business context: how to control the abandons of my subscribers* . . . . . 190

**Session of free contributes SCL5** – *Assessing Performance*
**Chair: Cristina Davino** **194**
di Trapani G. *Political performance measuring and tracking through a system based on the Political Performance Indicator (Iep): Naples 2021 case* . . . . . 195
Montanari G. E. and Doretti M. *A class of case-mix adjusted probability-based indices for performance evaluation* . . . . . 202

Cavicchia C., Sarnacchiaro P., Vichi M. and Zaccaria G. *An ultrametric model to build a Composite Indicators system* . . . . . 208

**Session of free contributes SCL6** – *Statistical Learning*
**Chair: Massimo Aria** **212**

Migliorati M. and Brentari E. *Feature definition for NBA result prediction through Deep Learning* . . . . . 213

Levantesi S., Lizzi M. and Nigri A. *An application of contrast trees for mortality models diagnostic and boosting* . . . . . 219

Aria M., Gnasso A. and D'Aniello L. *Twenty Years of Random Forest: preliminary results of systematic literature review* . . . . . 225

Park H., Hong J., Shin Y. and Park J-S. *A study on the GEV activation function for classification of class imbalance data* . . . . . 231

**Session of free contributes SCL7** – *Society and Disparity*
**Chair: Pasquale Sarnacchiaro** **236**

Alaimo L. S., D'Urso P. and Nigri A. *The gender gap in lifespan disparity as a social indicator of international countries: A fuzzy clustering analysis approach* . . . . . 237

Iannario M. and Tarantola C. *Modelling scale effects via a Bayesian approach: an application to decision making in public sector* . . . . . 243

Gangi F. , Daniele L. M. and Coscia M. *Board Gender Diversity and Social engagement: evidence from the banking industry* . . . . . 249

**Session of free contributes SCL8** – *Education*
**Chair: Matilde Bini** **256**

Davino C. and Lamberti G. *Assessing heterogeneity in students' performance. The case of the Massive Open Online Courses* . . . . . 257

Primerano I., Catone M. C., Giordano G. and Vitale M. P. *Assessing undergraduate students' perceptions of distance learning during the COVID-19 pandemic* . . . . . 263

Crisci A., Lucadamo A. and Amenta P. *PhD satisfaction analysis in Italian University via Classification tree, Bagging and Random Forest* 269

Cervellera S., Cusatelli C. and Giacalone M. *Comparative Analysis of Student Learning: Technical, Methodological and Result Assessing of PISA-OECD and INVALSI-Italian Systems* . . . . . 275

**Session of solicited contributes SS7** – *SEM with PLS: Theory and Applications*
**Organizer and Chair: Enrico Ciavolino** **281**

Wang S., Cheah J. and Roldan J. L. *PLS-SEM basics and its potential applications: A quick journey* . . . . . 282
Cefis M. and Carpita M. *A PLS-SEM confirmatory composite analysis for football goalkeeper's performance validation* . . . . . 288
Pasca P., Misuraca M., Meloni A. and Ciavolino E. *Text-mining and PLS-SEM combination to measure food satisfaction with Google Review: When the gut (re)counts!* . . . . . 294

**Session of solicited contributes SS8** – *Applications of non standard statistical tools to real-life*
**Organizer and Chair: Antonio D'Ambrosio** **300**
Iorio C. and Pandolfo G. *A robust strategy for building a financial portfolio* 301
Ruscone M. N. and De Luca G. *Conditional copula a financial application* 307
Ortu M., Frigau L. and Contu G. *Explaining Student Satisfaction Assessments: A Natural Language Processing Approach* . . . . . 313

**Session of solicited contributes SS9** – *Innovation and Value Co-creation in Society*
**Organizers and Chairs: Alessandra De Chiara & Anna D'Auria** **319**
Mauro S. *The smart working towards a Society 5.0* . . . . . 320
Del Vacchio E., Carignani F., Laddaga C. and Bifulco F. *Innovative interaction in Society 5.0: insights from the cultural sector* . . . . . 326
D'Auria A. and De Chiara A. *Society 5.0: a bibliometric analysis* . . . . 333

**Session of solicited contributes SS10** – *Statistical learning for mainstream press, health and fiscal data*
**Organizer and Chair: Claudio Conversano** **339**
Baldassarre A. and Carullo D. *The regression trunk model for partitioning Italian municipalities based on their fiscal capacities and its determinants* . . . . . 340
Pandolfo G. and Iorio C. *An analysis of Italian healthcare mobility through a depth-based clustering procedure* . . . . . 346
Dossou B. F. P. and Wilhelm A. F. X. *Automatic Fake News Detection to Ensure Quality of News Articles* . . . . . 352

**Session of solicited contributes SS11** – *Multi-way Methods for Evaluation Service*
**Organizer and Chair: Michele Gallo** **358**
Simonacci V., Marino M., Grassia M. G. and Gallo M. *Multiple factor analysis with external information on PISA survey data* . . . . . 359

Bocci L. and D. Vicari D. *A three-way analysis of well-being in Italy over time* . . . . . 365
Cerqueti R., Mattera R. and Scepi G. *Multiway approach for clustering time series with time varying parameters* . . . . . 371

**Session of solicited contributes SS12** – *Assessment of Management Quality*
**Organizer and Chair: Clelia Fiondella** **377**
Belfiore A., Cuccurullo C. and Aria M. *Using partial triadic analysis for depicting the temporal evolution of Italian private healthcare organizations* . . . . . 378
Bollani L., Celegato A., Barbero F. and Fontemaggi F. *Management of the human factor into the company. An experience from the aeronautic sector.* . . . . . 385

**Session of solicited contributes SS13** – *New technologies for students learning assessment and evaluation*
**Organizer and Chair: Alfonso Iodice D'Enza** **391**
Themelis E. and Markos A.*A non parametric cognitive diagnostic method in classroom assessment conditions* . . . . . 392
Iannario M., Iodice D'Enza A. and Romano R. *Hybrid unfolding models to Likert-scale data to assess distance learning perception in higher education* . . . . . 398
Pacella D., Fabbricatore R., Galluccio C. and Palumbo F. *Classification of Statistics learners using multi-dimensional latent class IRT model and archetypal analysis: the ALEAS app* . . . . . 404

**Session of solicited contributes SS14** – *Labor Market and Enterprises*
**Organizer and Chair: Lucio Masserini** **408**
Maggioni G., Mariani P., Marletta A. and Zenga M. *Searching for new trends and dynamics in Labour Market: a statistical approach for the recruiting process* . . . . . 409
Bruttini P., Mariani P., Marletta A., Masserini L. and Zenga M. *A new definition of the professional figure Open Manager* . . . . . 413

**Session of solicited contributes SS15** – *Statistical Approaches to Environmental Sustainability*
**Organizer and Chair: Alfonso Piscitelli** **417**
Aspinall R. *Measuring sustainability as an emergent property of whole system dynamics* . . . . . 418

Alaimo L. S. and Finocchiaro G. *Tourism sustainability in the Italian regions: a fuzzy approach* . . . . . 424
D'Uggento A. M. *How young people perceive environmental issues, react to ecological concerns and commit themselves to sustainable behaviours* 430

**Session of solicited contributes SS16** – *Statistical Methods for Environmental, Natural Resources and Health Assessment*
**Organizer and Chair: Alessio Pollice** **437**
Ferretti A., Ippoliti L. and Valentini P. *Spatial-ARFIMA models for the statistical analysis of environmental lattice processes* . . . . . 438
Arima S., Pasculli G. and Polettini S. *A Bayesian non parametric approach for bias correction for underreported data* . . . . . 444
Lasinio G. J., Mastrantonio G., Pollice A., Ventura D., Mancini G. and Ardizzone G. *Assessment of the impact of anthropic pressures on the Giglio island meadow of Posidonia oceanica* . . . . . 450

**Session of solicited contributes SS17** – *Functional Data Analysis Methodologies for Quality Assessment*
**Organizer and Chair: Elvira Romano** **456**
Fortuna F., Maturo F. and Di Battista T. *Improving the quality of questionnaires via the combined use of functional outlier detection and Item Response Theory* . . . . . 457
Naccarato A., Fortuna F. and Terzi S. *Assessing government effectiveness over time: a functional data analysis approach* . . . . . 462
Balzanella A. and Verde R. *Mining Distributed Acoustic Sensing data for vehicle traffic monitoring* . . . . . 467

**Session of solicited contributes SS18** – *Evaluation and assessment of cognitive and learning processes*
**Organizer and Chair: Francesco Palumbo** **473**
Ponticorvo M. and Argiuolo T. *Neuropsychological Assessment supported by Technology: the E-BTT case* . . . . . 474
Milano N. and Gigliotta O. *Mining Introducing OpenAi-ES in Interactive Data. Clustering with R-EVOK* . . . . . 480
Davino C., Gherghi M., Palumbo F. and Vistocco D. *Modeling heterogeneity in student's satisfaction during the Covid-19 pandemia* . . . 484

**Session of solicited contributes SS19** – *Substainability and Environment*
**Organizer and Chair: Ida Camminatiello** **490**

Lombardo R. and Beh E. J. *Partitioning the Cressie-Read divergence statistic for three-way contingency tables: a study on environmental sustainability data* . . . . . 491
Prahadchai T., Hong J., Busababodhin P. and Park J-S.. *Analysis of maximum precipitation in Thailand using non-stationary extreme value models* . . . . . 498
Tregua M. and Scaglione M. *Assessing citizens' participation to urban transformation: a review of quantitative methods* . . . . . 504

**Session of solicited contributes SS20** – *Statistical methods for health and environmental impact assessment*
**Organizer and Chair: Fabrizio Maturo** **512**
Gattone S. A. and Di Battista T. *Density estimation via Functional Data Analysis* . . . . . 513
Evangelista A., Acal C., Aguilera A. M., Sarra A., Di Battista T. and Palermi S. *A new multivariate functional ANOVA approach for assessing air quality data amid COVID-19 pandemic* . . . . . 517
Diana A., Romano E. and Irpino A. *Conformal Prediction for Geographically Weighted Functional Regression models: an application for environmental impact assessment.* . . . . . 523

**Session of solicited contributes SS21** – *Local sustainability assessment: challenges in building quality indicators*
**Organizers and Chairs: Francesca Fortuna & Alessia Naccarato** **529**
Grimaccia E. *Urban Sustainability Assessment: A Proposal for an Index Based on SDGs' Indicators* . . . . . 530
Liberati P. and Resce G. *Between and Within Country Inequality in Regional Well Being* . . . . . 536
Di Battista T., Nissi E. and Sarra A. *Equitable and sustainable well-being over time: a functional approach* . . . . . 542

**Session of free contributes SCL9** – *Health and Covid-19*
**Chair: Maria Sole Pellegrino** **548**
Cao N., Calcagní A. and Finos L. *Twitter about COVID-19: An application of Structural Topic Models to a sample of Italian tweets* . . . . 549
Parretti C., Tartaglia R., Sbrana G., Mandó M. and Pacchi S. *Assessing the quality of a health service through the risk profile number (RPN)* 555
Di Lorenzo G., Franchetti G. and Politano M. *The insurance premium structure for a covid-19 insurance policy* . . . . . 562

**Session of solicited contributes SS22** – *Statistics, culture and tourism*
**Organizer and Chair: Marica Manisera** **568**
Cristiani P. *How data can influence the promotion and the consumption of cultural experience* . . . . . 569
Carpita M., Manisera M. and Zuccolotto P. *Mobile phone data to monitor the impact of social and cultural events of Brescia* . . . . . 575
Capecchi S., Quaranta G. and Salvia R. *Analysing opinions on sustainable tourism in the Vallo di Diano area, Campania, Italy* . . . . . 582

**Session of free contributes SCL10**– *Time Series data, Panel data and Circular Economy*
**Chair: Anna Crisci** **588**
Scaccabarozzi D., Toninelli D., Zurlo D., Bacchini F. and Iannaccone R. *Testing the Participation Gap Inclusion within the Wage Phillips Curve* . . . . . 589
Fusco D., Liguori M. A. and Moretti V. *Multisource approach for trends evaluation. An application at the agricultural sector* . . . . . 596
Bonnini S. and Borghesi M. *A permutation test on the relationship between Circular Economy and firm size* . . . . . 601
Mele S., Izzo F. and Tomnyuk V. *Circular economy and business models: a literature review* . . . . . 607

**Session of free contributes SCL11** –*Modelling Extreme Values, High dimensional, time series data*
**Chair: Violetta Simonacci** **611**
Shin Y., Busababodhin P. and Park J-S. *Modeling extreme values using the r-largest four parameter distribution* . . . . . 612
Sabri M., Maturo F., Verde R., Riffi J. and Yahyaouy A. E. *Classification of ECG signals based on functional data analysis and machine learning techniques* . . . . . 618
Park J-S., Shin Y., Shin Y. and Hong J. *Determining shape parameters in a climate multi-model ensemble* . . . . . 624

# Discovering archetypal universities in higher education mobility flows in Italy

## *Analisi delle università archetipiche nei flussi di mobilità degli studenti universitari in Italia*

Ilaria Primerano, Francesco Santelli and Cristian Usala

**Abstract** The aim of this contribution is to identify the archetypal universities in the Italian students' mobility network in terms of their attitude in attracting students. We define a set of networks according to the disciplinary groups by relying upon administrative data regarding students' mobility between bachelor's and master's degrees. For each disciplinary group, a network has been defined by considering the universities as nodes and the flows of students moving between nodes as links. Then, in each network, the set of archetypal universities is based on several network centrality indexes. Finally, these archetypes are used as benchmarks to identify the main determinants of universities' performances.

**Abstract** *L'obiettivo dell'analisi è quello di identificare le università archetipiche all'interno della rete di mobilità degli studenti universitari sulla base della loro capacità di attrarre o meno studenti. Sulla base dei dati amministrativi riguardanti le scelte di mobilità degli studenti tra la triennale e la magistrale, le singole reti sono state definite considerando, per ogni gruppo disciplinare, le università come nodi, e il flusso di studenti tra i nodi come legami. Per ogni gruppo disciplinare, è stato identificato il set di università archetipiche a partire dagli indici di centralità delle reti. Infine, questi archetipi sono stati utilizzati come riferimento per analizzare le determinanti delle performance delle università.*

**Key words:** University mobility, Archetypal Analysis, Network centrality indexes, Multivariate data

Ilaria Primerano
Department of Political and Social Studies, University of Salerno, Italy e-mail: iprimerano@unisa.it

Francesco Santelli
Department of Political Sciences, University of Naples Federico II, Italy e-mail: francesco.santelli@unina.it

Cristian Usala
Department of Political and Social Sciences, University of Cagliari, Italy e-mail: cristian.usala@unica.it

## 1 Introduction

In the last decades, the Italian student migration propensity has increased, highlighting the main flows from South to Northern Italy [1] [2]. The study of mobility flows can help measure the attractiveness of Italian universities. Different research approaches have been developed to study the determinants of students' mobility by means of different statistical methods, form longitudinal analysis [3] to Network Analysis considering geographical macro-area aggregations [4] [5], and different fields of study [6]. Most of the works refer to the first-level mobility, i.e. from high school diploma to bachelor degree, and only few recent contributions have dealt with the students' migration in the transition from Bachelor to Master degree programmes (i.e. second-level mobility), by focusing on Southern Italian students [7]. In studying second level mobility flows, a multilevel multinomial logit model has been adopted to assess the effects of university centrality role in the network (i.e. in terms of Hub and Authorities) on students' choices [6], while a multiplex network approach has been used to highlight the presence of groups of universities that play a fundamental role in each layer through the detection of the core universities [8].

Moving from this framework, this contribution investigates the Italian second-level mobility network, by considering the flows traced by students who change university for their master's degree by accounting also for differences existing among disciplinary fields. Specifically, a student who decides to change university when enrolling at the master is here considered as a student in mobility. Students' flows are analyzed to assess the similarity among Italian universities by identifying subgroups of universities within each disciplinary field that share common behavior in terms of their attitude in attracting students. In this context, students' flows are read into the scope of network analysis and, starting from a set of network centrality measures, the Archetypal analysis is applied to identify groups of similar universities within each disciplinary field.

The contribution is structured as follows: Section 2 describes the methodological approach; Section 3 presents an overview on the dataset used and the main results.

## 2 Network definition, centrality measures and archetypes

Based on Social Network Analysis [9], we start visualizing and analyzing Italian students' flows to get descriptive insights into the different network structures defined for each disciplinary field. We consider as nodes of source the universities where students achieved their bachelor's degree, and as nodes of destinations the universities where students enrolled for their master's degree. The flows of students moving in the second-level mobility network define the links connecting the Italian universities.

Specifically, starting from the ISCED-F 2013 classification, we define a set of one-mode, weighted, and directed networks [9]. Formally, each of these networks can be described as a graph $\mathcal{G}_i(V,L,W)$, where $V$ is the set of units, $L$ is the set of

directed links, and $W$ is the set of weights. Let $\mathbf{A_i}$ be the corresponding adjacency matrices, with elements $a_{ij}$ holding the presence of one or more links between node of origin $v_i$ and node of destination $v_j$, otherwise $a_{ij} = 0$.

To gain insights on the characteristics of these networks, for each disciplinary field we have computed several network centrality indexes to measure universities' attractiveness and to determine whether the observed universities are top receivers of incoming students or, on the opposite, top senders of outgoing students. To classify the universities, we have defined a matrix holding by rows the Italian Universities and by columns the set of centrality indexes computed. This matrix is then used to perform an Archetypal analysis [10].

Archetypal analysis is a method of unsupervised learning that aims to represent each object in a dataset as a mixture of *individuals of pure type*, known as archetypes.Formally, archetypes are defined as $m$ points $\left\{a_j\right\}$, $j = 1,...,m$ contained in archetype matrix $\mathbf{A}$ that are in the euclidean space that satisfy $x_i' = \alpha_i'\mathbf{A}$. The computation of the archetypes is a non-linear least squares problem, which is solved using an alternating minimizing algorithm. The key idea is that, given a $\mathbf{X}_{(n\times p)}$ data matrix, with $n$ individuals and $p$ variables, each archetype is computed as a linear combination of the original data based on the following constraints: $\beta_{ji} \geq 0$; $\sum_j \beta_{ji} = 1$; $\alpha_{ij} \geq 0$; $\sum_{i=1}^{m} \alpha_{ij} = 1$. The optimal $\alpha_{ij}$ are found by minimizing the following:

$$RSS = \sum_i \left\| x_i - \sum_{j=1}^{m} \alpha_{ij} A \right\|^2 \tag{1}$$

where $\alpha_{ij}$ are the coefficients of the archetypes while $\beta_{ji}$ are the coefficients of the data set. Archetypes are useful in unsupervised learning also due to their location properties. Given a Convex Hull ($CH$) of original data points, if $k = 1$, only one archetype is identified, and the sample mean is the solution to minimize $RSS$; if $1 < k < n$, all the $a_j$ vectors of archetypes lie on the boundary of $CH$ to minimize $RSS$; if $k = n$ the number of archetypes is equal to $n$ the $RSS = 0$ [10]. Using both sets of coefficients $\alpha_{ij}$ and $\beta_{ji}$, the $RSS$ in (1) can be written using matrices notation, and thus matrix $\Gamma$ including all the $\alpha$'s, matrix $B$ of the coefficients $\beta$'s and the starting matrix $X$ of the data, obtaining in Frobenius norm:

$$\min_{\boldsymbol{\Gamma}_k, \mathbf{B}_k} RSS_k = \min_{\boldsymbol{\Gamma}_k, \mathbf{B}_k} \left\| \mathbf{X} - \boldsymbol{\Gamma}_k \mathbf{B}_k^T \mathbf{X} \right\|_F \tag{2}$$

In the space spanned by archetypes by exploiting the property of the compositional space defined by the Aitchinson distance [11], a k-means cluster analysis is performed to find consistent group of universities with respect to the archetypes.

## 3 Data description and main findings

We have collected data on students' mobility choices from the micro-data database MOBYSU.IT [12] that includes information on students university careers.[1] We consider the population of Italian students enrolled in a bachelor program in an Italian university between a.y. 2011-12 and a.y. 2016-17 that have enrolled in a master's degree program between the 2014 and 2019. We retain in our data only students that have graduated in this time frame (639,505 students, the 56.6% of the population), and that have enrolled in a master's degree program. Therefore, our data includes 400,049 students grouped in 92 universities (of which 10 are e-learning institutions). Moreover, to define the networks, we classify the students according to their disciplinary group. In particular, the set of available degree programs are classified into 10 disciplinary groups according to the ISCED-F 2013 classification [13]. Data are described in Table 1.

**Table 1** Descriptive statistics on master's students mobility choices. Students are classified as *in mobility* if they have changed their university in the transition from bachelor's to master's degrees

| | Master students | | In mobility students | |
|---|---|---|---|---|
| ISCED - F 2013 | Total N | In mobility N (%) | Same field (%) | Different field (%) |
| All fields | 400,049 | 110,442 (27.61) | 85.30 | 14.70 |
| **By disciplinary field in origin university:** | | | | |
| N1 - Agriculture, forestry, fisheries and veterinary | 12,385 | 3,132 (25.29) | 82.25 | 17.75 |
| N2 - Arts and humanities | 71,807 | 22,649 (31.54) | 82.59 | 17.41 |
| N3 - Business, administration and law | 62,539 | 16,896 (27.02) | 85.36 | 14.64 |
| N4 - Education | 13,508 | 3,569 (26.42) | 92.23 | 7.77 |
| N5 - Engineering, manufacturing and construction | 86,518 | 14,347 (16.58) | 98.62 | 1.38 |
| N6 - Health and welfare | 12,390 | 4,719 (38.09) | 76.52 | 23.48 |
| N7 - Information and Communication Technologies (ICTs) | 4,809 | 938 (19.51) | 85.15 | 14.85 |
| N8 - Natural sciences, mathematics and statistics | 46,023 | 12,743 (27.69) | 95.68 | 4.32 |
| N9 - Services | 14,149 | 4,720 (33.36) | 36.71 | 63.29 |
| N10 - Social sciences, journalism and information | 75,921 | 26,729 (35.21) | 76.12 | 23.88 |

In the following, to show the results of our analytical procedure, we focused on the Education Field. The results of archetypal analysis performed on the network measures show that the best solution found by minimizing the RSS is obtained with 6 archetypes. In this space, the clustering algorithm identifies 12 groups of universities. The archetypes description with respect to original indexes (figure 1) and the clustering analysis (figure 2) show that the $6^{th}$ archetype has a peculiar trait with very high values for both attractiveness and exporting indexes (Roma Tre University is very close to this archetype), while the $5^{th}$ archetype represents an extreme type of university characterized by few flows (some Telematic universities are very

[1] Data drawn from the Italian 'Anagrafe Nazionale della Formazione Superiore' has been processed according to the research project 'From high school to the job market: analysis of the university careers and the university North-South mobility' carried out by the University of Palermo (head of the research program), the Italian 'Ministero Università e Ricerca', and INVALSI.

close to this archetype). Archetype $4^{th}$ is characterized by a greater attitude to export rather than import students (some Southern Universities), while the $3^{th}$ archetype, which is characterized by a very high value in both in–strenght and in–closeness network measures, fully describes University Pegaso and it is very close to Milano Bicocca, Bologna and Milano Cattolica.
The archetypal analysis on the network indexes highlights the peculiar behavior of the on–line universities in the mobility flows, and simultaneously a strong geographical component clearly emerges in most of the identified clusters.

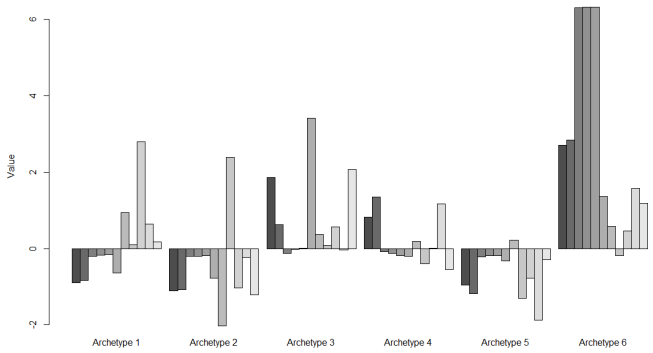


**Fig. 1** Barplots for each archetype with respect to original variables (as archetypal coefficients). Columns are, from left to right: In–strength, Out–strength, Hub, Authority, Eigen–centrality, Page–rank, In–closeness, Out–closeness, Vertex–betweenness, N. of Recursive Deps, N. of reverse Recursive Deps

**Acknowledgements** This contribution 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.

## References

1. Attanasio, M., Enea, M. & Priulla, A.: Quali atenei scelgono i diplomati del Mezzogiorno d’Italia?. Neodemos, ISSN: 2421-3209, (2019)
2. D’Agostino, A., Ghellini, G., & Longobardi, S.: Exploring determinants and trend of STEM students internal mobility. Some evidence from Italy. Electron. J. App. Stat. Anal. **12(4)**, 826–845, (2019)
3. Attanasio, M., Enea, M., & Albano, A.: Dalla triennale alla magistrale: continua la ‘fuga dei cervelli’ dal Mezzogiorno d’Italia. Neodemos, ISSN: 2421-3209, (2019)
4. Columbu, S. & Primerano, I.: A multilevel analysis of university attractiveness in the network flows from bachelor to master’s degree. In A. Pollice, N. Salvati, and F. Schirippa Spagnolo, editors, Book of short Papers SIS 2020, pages 480–485, (2020)

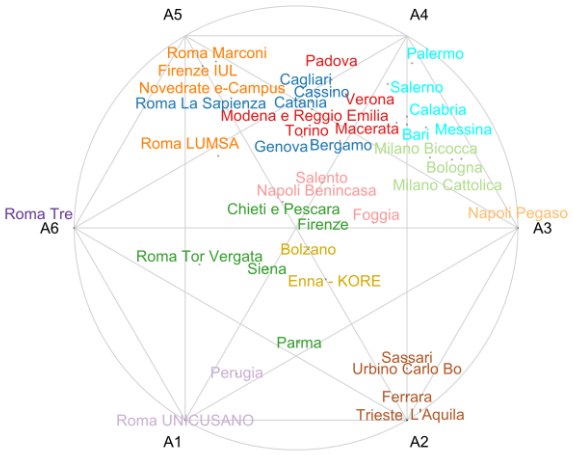


**Fig. 2** Simplex plot of the space spanned by 6 archetypes. Universities are row–points. Clustering is performed to obtain 12 groups

5. Genova, V. G., Tumminello, M., Enea, M., Aiello, F., & Attanasio, M.: Student mobility in higher education: Sicilian outflow network and chain migrations. Electron. J. App. Stat. Anal. **12(4)**, 774–800, (2019)
6. Columbu, S., Porcu, M., Primerano, I., Sulis, I., & Vitale, M.P.: Geography of Italian student mobility: A network analysis approach. Socio-Econ. Plan. Sci. **73**, 100918 (2021)
7. 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, and A. Ruiz-Gazen, editors, Studies in Theoretical and Applied Statistics, 239–249, (2018)
8. Primerano, I., Santelli, F., Usala, C.: A multiplex approach to study Italian Students' Mobility. In C. Perna, N. Salvati, and Schirripa Spagnolo F. (a cura di), Book of Short Papers SIS 2021, 473-478, (2021)
9. Wasserman, S., & Faust, K.: Social network analysis: Methods and applications, Cambridge University Press, (1994)
10. Cutler, A. & Breiman, L.: Archetypal analysis. *Technometrics* **36**, 338-347 (1994)
11. Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J. & Pawlowsky-Glahn, V.: Logratio analysis and compositional distance. *Mathematical Geology* **32**, 271-275, (2000)
12. Database MOBYSU.IT [Mobilità degli Studi Universitari in Italia], research protocol MUR - Universities of Cagliari, Palermo, Siena, Torino, Sassari, Firenza, Cattolica and Napoli Federico II, Scientific Coordinator Massimo Attanasio (UNIPA), Data Source ANS-MUR/CINECA.
13. UNESCO Institute for Statistics. ISCED Fields of Education and International Standard Classification of Education 2011, Montrèal, (2014)