

11th Scientific Meeting of the SIS Group
"Statistics for the Evaluation and Quality in Services"

BOOK OF **SHORT PAPERS**

Editors

Andrea Bucci

Alfredo Cartone

Adelia Evangelista

Andrea Marletta



**STATISTICAL METHODS
FOR EVALUATION AND QUALITY:
TECHNIQUES, TECHNOLOGIES AND TRENDS (T³)**

**IES 2023 - Statistical Methods for Evaluation and Quality:
Techniques, Technologies and Trends (T³)**

BOOK OF SHORT PAPERS

Editors: Andrea Bucci, Alfredo Cartone, Adelia Evangelista and Andrea Marletta

Book of Short papers
11th International Conference **IES 2023**
Statistical Methods for Evaluation and Quality: Techniques, Technologies and
Trends (T³)

University 'G. d'Annunzio' of Chieti-Pescara



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

Maurizio Carpita	University of Brescia
Pietro Amenta	University of Sannio
Matilde Bini	European University of Rome
Antonello D'Ambra	University of Campania 'L. Vanvitelli'
Luigi D'Ambra	University of Naples 'Federico II'
Paolo Mariani	University of Milan 'Bicocca'
Marica Manisera	University of Brescia
Monica Palma	University of Salento
Pasquale Sarnacchiaro	University of Naples 'Federico II'

Program Committee of the conference IES 2023

Chair: Paolo Mariani	University of Milan 'Bicocca'
Matilde Bini	European University of Rome
Mario Bolzan	University of Padova
Giuseppe Bove	University of Rome 3
Eugenio Brentari	University of Brescia
Ida Camminatiello	University of Campania 'L. Vanvitelli'
Maurizio Carpita	University of Brescia
Carlo Cavicchia	Erasmus University Rotterdam
Roy Cerqueti	University of Rome 'La Sapienza'
Enrico Ciavolino	University of Salento
Corrado Crocetta	University of Bari 'Aldo Moro'
Antonello D'Ambra	University of Campania 'L. Vanvitelli'
Pierpaolo D'Urso	University of Rome 'La Sapienza'
Tonio Di Battista	University 'G. d'Annunzio' of Chieti-Pescara
Simone di Zio	University 'G. d'Annunzio' of Chieti-Pescara
Mauro Ferrante	University of Palermo
Francesca Fortuna	University of Rome 3
Michele Gallo	University of Naples 'Orientale'
Claudiu Herteliu	Bucharest University of Economic Studies
Salvatore Ingrassia	University of Catania
Ron Kenett	Samuel Neaman Institute & KPA Group
Michele La Rocca	University of Salerno
Rosaria Lombardo	University of Campania 'L. Vanvitelli'
Antonio Lucadamo	University of Sannio
Marica Manisera	University of Brescia
Andrea Marletta	University of Milan 'Bicocca'
Lucio Masserini	University of Pisa
Fabrizio Maturo	University Mercatorum
Matteo Mazziotta	ISTAT
Rodolfo Metulini	University of Bergamo
Stefania Mignani	University of Bologna
Enrico Ripamonti	University of Brescia

Silvia Salini	University of Milano
Pasquale Sarnacchiaro	University of Naples 'Federico II'
Violetta Simonacci	University of Naples 'Federico II'
Maurizio Vichi	University of Rome 'La Sapienza'
Mariangela Zenga	University of Milan 'Bicocca'

Organizing Committee

Chair: Tonio Di Battista	University 'G. d'Annunzio' of Chieti-Pescara
Andrea Bucci	University 'G. d'Annunzio' of Chieti-Pescara
Alfredo Cartone	University 'G. d'Annunzio' of Chieti-Pescara
Marco Di Marzio	University 'G. d'Annunzio' of Chieti-Pescara
Marta Di Nicola	University 'G. d'Annunzio' of Chieti-Pescara
Simone Di Zio	University 'G. d'Annunzio' of Chieti-Pescara
Stefania Fensore	University 'G. d'Annunzio' of Chieti-Pescara
Lara Fontanella	University 'G. d'Annunzio' of Chieti-Pescara
Stefano Antonio Gattone	University 'G. d'Annunzio' of Chieti-Pescara
Luigi Ippoliti	University 'G. d'Annunzio' of Chieti-Pescara
Eugenia Nissi	University 'G. d'Annunzio' of Chieti-Pescara
Paolo Postiglione	University 'G. d'Annunzio' of Chieti-Pescara
Annalina Sarra	University 'G. d'Annunzio' of Chieti-Pescara
Pasquale Valentini	University 'G. d'Annunzio' of Chieti-Pescara
Adelia Evangelista (Secretariat)	University 'G. d'Annunzio' of Chieti-Pescara

Editors

Andrea Bucci University of Macerata
Alfredo Cartone University 'G. d'Annunzio' of Chieti-Pescara
Adelia Evangelista University 'G. d'Annunzio' of Chieti-Pescara
Andrea Marletta University of Milan 'Bicocca'



Il Viandante - Copyright © 2023 Edizioni BACME S.r.l.s.
Sede legale: Via Silvino Olivieri, 111 - 66100 Chieti (CH)

ISBN 979-12-803-3369-8

DOI 10.60984/978-88-94593-36-5-IES2023

<https://doi.org/10.60984/978-88-94593-36-5-IES2023>

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.

Sponsored by



Preface

Statistical thinking, design and analysis play a crucial role in social life and are useful to society at large. Besides, promoting advanced methodological research is useful to facilitate the dissemination of ideas related to various fields of interest. For this purpose, experts in statistics, data analysis, data mining, statistical methods for decision making, machine learning and related methods come together to understand and analyse phenomena through data.

In line with this objective, the Statistics Group for the Evaluation and Quality of Services (SVQS; www.svqs.it) of the Italian Statistical Society (SIS) has been organizing the Innovation and Society (IeS) conference biennially since 2009, focusing on new developments and ideas in statistics applied to the evaluation and quality of public and private services, attracting national and international statisticians and data scientists. The meeting contributes to spot light on the main statistical approaches and methodologies for the evaluation of public services currently in use in different contexts, as well as to facilitate discussion on the impact of innovative statistical evaluation systems for these services, involving various economic and social policy actors.

The conference “Statistical Methods for Evaluation and Quality: Techniques, Technologies and Trends (T³)” recorded valuable contributions that are reported in this volume. The papers underscore how the growing availability of data has tasked social and economic actors, organizations, and researchers with the management and analysis of large volumes of unstructured and heterogeneous data. In recent years, many tools for both qualitative and quantitative models have been developed to better describe and understand complex systems and their underlying behaviors, and the papers reported in this volume bear witness to this.

Techniques, technologies and trends: the study of data complexity presents the potential to provide analyses with increased frequency and timeliness, accuracy and objectivity, and to define sustainable models. Traditional quantitative methods for capturing socioeconomic data have often shown limitations in their ability to examine underlying systems, and with the three ‘T’ just mentioned, the outlines of future developments are starting to emerge.

The volume reports 127 contributions in the following areas:

- Advanced statistical methods for pattern recognition
- Advances in statistical learning from high-dimensional data
- Data analysis for web sources
- Distance and depth-based statistical learning methods for robust data analysis

- Economics and environment
- Education and labour
- Inequalities in the labour market
- Innovations and challenges in official statistics
- Labour market: trends, perspectives and new challenges
- Methodological and applicative contributions for evaluating sustainable development
- Methodological developments and applications for the assessment of student competencies
- Networks data analysis: new perspectives and applications
- New advanced statistical methods for data science
- Recent advances in statistical learning and data analysis
- Statistical analysis and modeling of environmental pollution data
- Statistical methods and complexity for evaluation in finance
- Statistical methods and composite indicators for healthcare
- Statistical methods and models for land monitoring with spatio-temporal data
- Statistical methods for environmental monitoring and sustainability
- Statistical methods for the analysis of university student choices and academic performance
- Statistical methods for the assessment of transport services and sustainable emissions
- Statistical methods for education and educational services
- Statistics in sports
- Tourism and territory.

The Conference event attracted many contributions as well as numerous Authors, not just from Italy but also from abroad. Over the three-day meeting, the Community has the opportunity to witness some of the state-of-the arts, new trajectories, and methodological challenges in 24 solicited sessions, 7 sessions of free contributes, two round tables - organized by Maurizio Vichi and Matilde Bini respectively - and three keynotes sessions with Ron S. Kennet of Samuel Neaman Institute of Israel, Luigi D'Ambra of Federico II University of Naples, and the former Minister Enrico Giovannini from University of Tor Vergata.

Organizers

Paolo Mariani

Chair of the Program Committee

Tonio Di Battista

Chair of Local Organizing Committee

Maurizio Carpita

Coordinator of the SVQS Group

Contents

Plenary Session	1
Befitting Cross Validation with Three Case Studies	2
<i>Kenett R.S.</i>	
Solicited Session SS1 - <i>Statistical analysis and modeling of environmental pollution data</i>	3
Assessing environmental quality by clustering a structural equation model based index: An application to European cities air pollution	4
<i>Bottazzi Schenone M., Grimaccia E. and Vichi M.</i>	
Evaluating the nonlinear association between PM ₁₀ and emergency department visits	10
<i>Bucci A., Sanmarchi F., Santi L., Giostra F., Tubertini E., Rosa S., Nante N. and Golinelli D.</i>	
Estimating spatially varying Gaussian Graphical Models to unveil relationships among pollutants in the Red River Delta in Vietnam	16
<i>Pronello N., Cucco A., Ignaccolo R. and Ippoliti L.</i>	
Solicited Session SS2 - <i>Statistics in sports</i>	22
Clustering Athlete Performances in Track and Field Sports	23
<i>Argiento R., Colombi A., Modotti L. and Montagna S.</i>	
A Cross-Country Analysis of Engagement in Physical Activity and Sport Practice Learnt from Eurobarometer Survey Data	29
<i>Simone R.</i>	
Strong eras in male professional tennis	35
<i>Breznik K., Candila V., Milekhina A. and Restaino M.</i>	
NonParametric Combination method for data analytics in basketball matches	41
<i>Barzizza E., Biasetton N., Ceccato R., Disegna M. and Vezzosi G.</i>	
Solicited Session SS3 - <i>Statistical methods for the analysis of university student choices and academic performance</i>	47
The influence of labor market conditions on students' career disruption: first insights from Italy	48
<i>Usala C., Sulis I. and Porcu M.</i>	
Socio-economic aspects that may affect South-North students' mobility in Italy	53
<i>Genova V.G. and Boscaïno G.</i>	
An analysis of student's performance in bachelor's degree	59

<i>La Rocca M., Niglio M. and Restaino M.</i>	
An exploratory strategy for analyzing students' mobility data	65
<i>Primerano I. and Giordano G.</i>	
Solicited Session SS4 - <i>Statistics in football</i>	71
Community Detection in Sport Market Networks: The Case of Italian Professional Football	72
<i>Randinelli R. and Ievoli R.</i>	
An Original Application to Football of PLS-SEM for the xG Model	78
<i>Cefis M. and Carpita M.</i>	
Performance Assessment of Football Players and Combined Permutation Tests with application to Home-Field Advantage	84
<i>Bonnini S.</i>	
A First Proposal of the Triad Census for Weighted Networks: an Application to Football	90
<i>Randinelli R. and Palazzo L.</i>	
Solicited Session SS5 - <i>Advances in statistical learning from high-dimensional data</i>	96
PCA approaches for vector functional time series	97
<i>Aguilera A.M., Alonso F.J. and Acal C.</i>	
Conformal Prediction for Functional Kriging Models	102
<i>Diana A., Romano E. and Adzic J.</i>	
Measuring Public-Private Connectedness in Financial Markets	108
<i>Sánchez García J. and Cruz Rambaud S.</i>	
An original approach to anomalies in intertemporal choices through Functional Data Analysis: Theory and application for the study of Hikikomori syndrome	112
<i>Martino R., Ventre V., Cruz Rambaud S. and Maturo F.</i>	
Solicited Session SS6 - <i>Labour market: trends, perspectives and new challenges</i>	118
Enriching Job Vacancy Official Information with Online Job Advertisements: Chances and Limits	119
<i>Lucarelli A. and Righi A.</i>	
Innovation in Management: towards the Open Manager	126
<i>Bruttini P., Gallo M., Mariani P. and Menini T.</i>	
Solicited Session SS7 - <i>Data analysis for web sources</i>	132
Enhancing SMEs default prediction with web-scraped data	133
<i>Crosato L., Domenech J. and Liberati C.</i>	
Web data as enabler for informed decisions in Labour Market	137
<i>Maggioni G.</i>	
The metaverse & luxury fashion brands: strategic communication exercise	141
<i>Forciniti A. and Zavarrone E.</i>	
Increasing the Geographical Granularity of Economic Indicators with Google Trends	147

Domenech J. and Marletta A.

Solicited Session SS8 - <i>Methodological and applicative contributions for evaluating sustainable development</i>	153
Evaluating sustainable development in EU countries through synthetic indicators	154
<i>Alaimo L.S. and Cucci M.</i>	
Naples and tourism sustainability: A survey of citizens' perceptions	160
<i>Aria M., Pagliara F., D'Aniello L. and Della Corte V.</i>	
Modelling inequalities for sustainable development in Italy countries	166
<i>Musella M., Borrata G., Camminatiello I. and Lombardo R.</i>	
Food Security and Sustainability: A Science Mapping Analysis	172
<i>Piscitelli A.</i>	
Solicited Session SS9 - <i>Inequalities in the labour market</i>	178
Skill similarities across Italian regions: an analysis based on the online job advertisements	179
<i>Kahlawi A., Buzzigoli L., Grassini L., Martelli C. and Giambona F.</i>	
Italian Labour Market reform and gender inequalities	185
<i>Marini C. and Nicolardi V.</i>	
Intergenerational transmission of disadvantages in the Italian labour market: evidence from AD-SILC data	191
<i>Busetta A., Fabrizi E., Ragozini G. and Sulis I.</i>	
Solicited Session SS10 - <i>Statistical methods and complexity for evaluation in finance</i>	197
Financial networks resilience and shocks propagation	198
<i>Cerqueti R., Cinelli M., Ferraro G. and Iovanella A.</i>	
How the choice of one parameter impacts the numerical stability of the efficient frontier	204
<i>Fassino C. and Uberti P.</i>	
Dynamic shrinkage for minimum variance combination of forecasts	210
<i>Mattera R.</i>	
Exploring the perception of the gender issue of Italian female entrepreneurs	217
<i>Castellano R., Riccioni J. and Rinaldi A.</i>	
Solicited Session SS11 - <i>Networks data analysis: new perspectives and applications</i>	223
Describing Italian mobility trajectories in higher education	224
<i>Genova V.G., Giordano G., Ragozini G. and Vitale M.P.</i>	
Collaboration networks: methodological issues and updated empirical evidence on Italian statisticians	230
<i>De Stefano D., Fabbrucci Barbagli A.G., Santelli F. and Zaccarin S.</i>	
Mapping Ashtma Complexity with Graph Theory: an Integrative Approach	236
<i>Cucco A., Simpson A., Murray C., Fontanella S. and Custovic A.</i>	
Investigating the patterns of Italian internal mobility: a network analysis at provincial level	242

Sarra A., D'Ingiullo D., Evangelista A., Nissi E., Quaglione D. and Di Battista T.

Solicited Session SS12 - <i>Innovations and challenges in official statistics</i>	248
Formal and informal networks of care for the elderly: regional profiles compared	249
<i>Sicuro L., Tucci D. and Coniglio R.</i>	
Gender Gap: a multidimensional approach	255
<i>Acampora C., Fusco D., Liguori M.A. and Pagliuca M.M.</i>	
Using Whatsapp in Official Statistics: a New tool for managing the Agriculture Census	259
<i>Fabi C.</i>	
Solicited Session SS13 - <i>Statistical methods and composite indicators for healthcare</i>	265
Longitudinal composite indicators to measure the quality of health services	266
<i>Crocetta C., Antonucci L., Cataldo R. and Mazza R.</i>	
Past and Future of Doctor-Patient Communication	272
<i>Tedesco N., Zavarrone E. and Forciniti A.</i>	
Network Analysis approach to customer satisfaction and service quality detection: an application to health-care services	277
<i>Crocetta C., Grassia M.G., Marino M., Mazza R., Simonacci V. and Stavolo A.</i>	
A project evaluation study on multiset Likert scale data	283
<i>Simonacci V., Marino M., Grassia M.G. and Gallo M.</i>	
Solicited Session SS14 - <i>Distance and depth-based statistical learning methods for robust data analysis</i>	289
Robust distance-based predictive models	290
<i>Boj E., Grané A. and Parron D.</i>	
Data depth for mixed-type data through multidimensional scaling. An application to biological age imputation	294
<i>Cascos I., Grané A. and Qian J.</i>	
A compared protocol to improve clustering procedures	298
<i>Grané A., Riani M. and Salini S.</i>	
Robust diagnostics for Linear Mixed Models with the Forward Search . . .	304
<i>Corbellini A., Grossi L. and Laurini F.</i>	
Solicited Session SS15 - <i>Advanced statistical methods for pattern recognition</i>	310
Unsupervised classification of NPLs recovery curves	311
<i>Carleo A. and Rocci R.</i>	
Living alone in Italian municipalities	317
<i>Vellucci P., Benassi F., Naccarato A. and Gallo G.</i>	
Supervised learning from high-dimensional data through dynamic updating of functional classification rules	323

<i>Maturo F., Fortuna F. and Di Battista T.</i>	
Assessing the effectiveness of coordination among public authorities in cohesion expenditure	329
<i>Coco G., Monturano G. and Resce G.</i>	
Solicited Session SS16 - <i>Recent advances in statistical learning and data analysis</i>	335
A Predictive Functional Principal Component Analysis of Well-Being Data	336
<i>Marcis L., Pagliarella M.C. and Salvatore R.</i>	
Detecting the partition in the extended hierarchy of a dendrogram: an application on biomedical data	343
<i>PolICASTRO V., Palazzo L. and VISTOCCO D.</i>	
Concordance measure for rankings	350
<i>Bissiri P.G. and Nai Ruscone, M.</i>	
Quadratic discriminant scoring for selecting clustering solutions	355
<i>Coraggio L. and Coretto P.</i>	
Solicited Session SS17 - <i>Statistical methods for education and educational services</i>	361
Association between INVALSI scores and students' mobility in Italy: a preliminary analysis	362
<i>Bacci S., Bertaccini B., Lombardi G. and Tocchioni V.</i>	
Modelling Responses and Response Times: an application to Mathematics INVALSI data	368
<i>Bungaro L., Desiderati R. and Mignani S.</i>	
Latent potential outcomes: An analysis of the effects of programs aimed at improving students' non-cognitive skills	374
<i>Pennoni F., Bartolucci F. and Vittadini G.</i>	
Cognitive Skills and Non Cognitive Skills to Analyze School and Students Performances	380
<i>Vittadini G.</i>	
Solicited Session SS18 - <i>Statistical methods for the assessment of transport services and sustainable emissions</i>	386
Sustainability assessment of urban transport by an LCA comparison on different technologies vehicles	387
<i>Della Ragione L. and Meccariello G.</i>	
Passenger comfort prediction via time-series classification	393
<i>Vanacore A., Pellegrino M.S. and Ciardiello A.</i>	
A statistical model to analyse driving behavior: a case study	399
<i>Rodia G., Sarnacchiaro P. and Acciarino V.</i>	
Aggregating judgments in non negotiable group decisions in transport system	403
<i>Amenta P. and Lucadamo A.</i>	
Solicited Session SS19 - <i>New advanced statistical methods for data science</i>	408

A unified framework for two-dimensional clustering on preference-approvals: an analysis of Eurobarometer data	409
<i>Albano A., Sciandra M. and Plaia A.</i>	
Pandemic Data Quality Modelling: A Bayesian Approach	415
<i>Ferrari L., Manzi G., Micheletti A., Nicolussi F. and Salini S.</i>	
Explainable AI for Peer-to-Peer Credit Risk Management	421
<i>Babaei G., Pagnottoni P. and Do T. T.</i>	
Tackling misclassification in surveys about undeclared work via the EM algorithm	427
<i>Arezzo M.F., Guagnano G. and Vitale D.</i>	
Solicited Session SS20 - <i>Tourism, territory and data analysis</i>	432
Tourism, sustainability, and territorial impact: an input-output analysis . .	433
<i>Garau G. and El Meligi A.K.</i>	
The Impact of Big Data in Tourism	439
<i>Ciuffreda R., Choedon C. and Simonetti B.</i>	
Neural network-based prediction of domestic tourists' length of stay in Italy	443
<i>Antolini F. and Cesarini S.</i>	
The management of cultural heritage in contexts of undertourism: a model for assessing the economic sustainability of public-private partnerships	450
<i>Calabrò F.</i>	
Solicited Session SS21 - <i>Statistical methods and models for land monitoring with spatio-temporal data</i>	457
Geo-referenced data and complex networks for measuring road accident risk	458
<i>Cantaluppi G., Clemente C., Della Corte F. and Zappa D.</i>	
A comparison of geospatial models for car crash risk	464
<i>Cantaluppi G., Giardino G. and Zappa D.</i>	
Geostatistical modelling of livestock-related PM _{2.5} pollution and scenario analysis for policymakers - Work in progress	470
<i>Fassò A., Rodeschini J., Fusta Moro A. and Finazzi F.</i>	
Functional clustering methods for space-time big data from mobile phone networks	476
<i>Perazzini S., Metulini R. and Carpita M.</i>	
Solicited Session SS22 - <i>Methodological developments and applica- tions for the assessment of student competencies</i>	483
Modeling the main drivers of mathematical literacy of school-leaving stu- dents. Some evidence from the Invalsi tests	484
<i>Davino C., Palumbo F., Romano R. and Vistocco D.</i>	
Educational Data Mining: clustering students' performance over time . . .	490
<i>Taraborrelli G. and Farnè M.</i>	
The nexus of cultural capital with participation in early childhood education	496
<i>Ripamonti E.</i>	
High- and Low-Performing students and future career: a gender and social issue	502
<i>Falzetti P. and Ricci R.</i>	

Solicited Session SS23 - <i>Statistical methods for environmental monitoring and sustainability</i>	508
Clustering spatial data through optimal transports	509
<i>Balzanella A. and Verde R.</i>	
New interpretative insights for environmental air quality by means of FDA	514
<i>Terzi S., Naccarato A. and Fortuna F.</i>	
A Bayesian State-Space Model to Mitigate Unmeasured Confounding . . .	520
<i>Zaccardi C., Valentini P. and Ippoliti L.</i>	
Mining social media data for damage assessment in environmental disasters	526
<i>del Gobbo E., Cafarelli B., Ippoliti L. and Fontanella L.</i>	
Solicited Session SS24 - <i>Satisfaction and behavior in tourism</i>	531
The evaluation of the hotel stay through a new development of correspondence analysis coping with ordinal variables	532
<i>D'Ambra A. and Amenta P.</i>	
Assessing the role of knowledge and authenticity in the formation of attendee loyalty at cultural festivals	536
<i>Rivetti F., Lucadamo A. and Rossi C.</i>	
Residents' Opinions and Perceptions of Tourism Development in the Historic City of Matera	542
<i>Sarnacchiaro P., Di Gennaro R. and Di Taranto E.</i>	
Exploring tourism at religious sites: The case of Assisi	547
<i>Rivetti F., Dini M. and Splendiani S.</i>	
Session of free contributes SFC1 - <i>Education and labour</i>	552
Local concordance among the items of questionnaires on student's opinion (OPIS)	553
<i>Terzi S. and Petrarca F.</i>	
High School Proficiency of Future University Students: An Analysis based on INVALSI Data	560
<i>Santelli F., Di Credico G. and Di Caterina C.</i>	
Employment vulnerability of immigrants in the labour market – Does origin matter?	566
<i>Bittaye M.</i>	
The effect of pricing policies on students' use of university canteens	572
<i>Masserini L., Bini M. and Lorenzoni V.</i>	
Gig workers' identikit	576
<i>Zavarrone E. and Forciniti A.</i>	
Session of free contributes SFC2 - <i>Tourism and territory</i>	582
Multi-source approach for statistics in tourism sector	583
<i>Bianchino A., Fusco D., Giordano P., Liguori M.A. and Summa D.</i>	
Statistical analysis of tourism sustainability in Campania: post Covid-19 review	589
<i>Giacalone M., Basile V. and Bellucci M.</i>	
Investigating recent changes in dietary behavior	596
<i>D'Uggento A., d'Ovidio F.D., Toma E. and Onorati M.G.</i>	

Depopulation in the Abruzzo municipalities	603
<i>Carulli A.L., Di Spalatro D. and Valentini A.</i>	
The Productions System of Inland Areas	607
<i>Madia Carucci A.M. and Regano A.</i>	
Session of free contributes SFC3 - <i>Social issues</i>	613
Modelling the gender gap in youth mortality with an Age-Period-Cohort analysis	614
<i>Lanfiuti Baldi G. and Nigri A.</i>	
Random forest for classifying odor emission sources	620
<i>Distefano V., Palma M., De Iaco S. and Mazuruse G.</i>	
An Experimental Annotation Task Investigating Annotator Agreement Within a Misogynistic Dictionary and Corpus	626
<i>Tontodimamma A., Ignazzi E., Anzani S., Fontanella L. and Di Zio S.</i>	
Statistical analysis of COVID19 impact on Italian mortality	632
<i>Franchetti G. and Politano M.</i>	
Measuring multidimensional deprivation using objective and subjective data: an application of the Voronoi ranking method	638
<i>Ciommi M., Mariani F., Polinesi G. and Recchioni M.C.</i>	
Session of free contributes SFC4 - <i>Trends</i>	644
The role of big data analytics in circular supply chains: A bibliometric analysis	645
<i>Khan F. and Rapposelli A.</i>	
Estimation of the ranking of incentive policies for the adoption of 4.0 technologies	652
<i>Bonnini S. and Borghesi M.</i>	
Risk Management and Future Scenarios. A proposal based on a mixed-method approach	658
<i>Di Zio S., Bolzan M., Marozzi M. and Scioni M.</i>	
Explainable artificial intelligence (XAI) through artificial intelligence from a human in the loop (HITL) perspective: an interview with ChatGPT	664
<i>Santarcangelo V., Lamacchia A., Vitullo S., Di Lecce M. and Giacalone M.</i>	
Relevance in official statistics: information needs, satisfaction with data quality, some results and future perspectives	670
<i>Loporcaro M.F.</i>	
Session of free contributes SFC5 - <i>Economic issues</i>	676
Evolutionary trends of start-ups in Italy: a case study	677
<i>Duttilo P., Caruso G., Iannone B. and Gattone S.A.</i>	
Permanent establishments and efficiency analysis with global enterprises	682
<i>Frenda A. and Sepe E.</i>	
Techniques and constructs in some recent market and organizational research	685
<i>Sciascia I.A.</i>	
The value of buildings in the Italian general government balance sheet	689
<i>Santoro P. and Regano A.</i>	

Linear and nonlinear factors affecting default risk in the peer-to-peer lending market	695
<i>Giordano F., Milito S. and Parrella M.L.</i>	
Session of free contributes SFC6 - <i>Methodological issues</i>	701
Macroeconomic Time Series Classification by Nonparametric Trend Estimation	702
<i>Feo G., Giordano F., Niglio M. and Parrella M.L.</i>	
A Normalization Method for Space-time Analysis of Evaluation and Quality Indicators	709
<i>Mazziotta M. and Pareto A.</i>	
Unveiling Latent Structures: exploring Multidimensional IRT Models using Dirichlet Process Mixtures	713
<i>Valentini P., Fontanella S. and Fontanella L.</i>	
On a technique to detect accounting data manipulation	718
<i>Passamonti C.</i>	
Session of free contributes SFC7 - <i>Economics and environment</i>	724
Determinants of Water Conservation Behaviour and Spatial Heterogeneity in their Coefficients	725
<i>Mammadli R. and Gigliarano C.</i>	
Modeling the economic burden of grass pollen allergoid immunotherapy . .	731
<i>Bilancia M. and Di Bona D.</i>	
Swine fever in Liguria: who does pay for economic losses? A causal analysis	735
<i>Baggetta C., Cavalletti B. and Corsi M.</i>	
Measuring Multidimensional Poverty of the Italian Regions in the era of COVID-19	741
<i>Chelli F.M.C., Ciommi M., Gigliarano C. and Polinesi G.</i>	
Drivers of inflation: relationships changing over time	747
<i>Sokolenko O., Palumbo A., Fortuna F., Naccarato A. and Marie J.</i>	
Environmental accounting and sustainable cities: an explorative bibliometric-based literature analysis	753
<i>Kaur A., Javed A., D'Andreamatteo A. and Rapposelli A.</i>	

An exploratory strategy for analyzing students' mobility data

Una strategia esplorativa per analizzare dati di mobilità studentesca

Ilaria Primerano and Giuseppe Giordano

Abstract This contribution deals with Italian students' decision to churn in the transition from the first to the second year of a bachelor's degree program. Based on these choices, three different churn scenarios are identified according to the decision to move towards a different university, to change degree programs, or both. Exploratory data analysis and Social Network methods are used to trace and visualize the links among universities defined considering these students' flows. The analysis is conducted considering students enrolled in universities located in the Campania region (Italy). The aim is to define an index of students' mobility flows considering their retention ability and students churn.

Abstract *Questo contributo analizza il tema della mobilità studentesca considerando la decisione degli studenti di trasferirsi nel passaggio dal primo al secondo anno di un corso di laurea triennale (churn risk). Da queste decisioni emergono diversi scenari di churn che dipendono dalla scelta di iscriversi in un'altra università, cambiare corso di laurea o entrambi. La Social Network Analysis permette di tracciare e visualizzare i legami tra le università definiti in base a questo specifico tipo di mobilità. L'analisi è stata condotta considerando gli studenti iscritti alle università situate nella regione Campania (Italia). L'obiettivo è definire un indice dei flussi degli studenti basato sulle decisioni di trasferimento e la capacità di ritenzione delle università.*

Key words: Churn risk, Correspondence Analysis, Directed Graphs, Origin Destination Matrix, Weighted Network

Ilaria Primerano

Dept. of Political and Social Studies, University of Salerno, e-mail: iprimerano@unisa.it

Giuseppe Giordano

Dept. of Political and Social Studies, University of Salerno, e-mail: ggiordano@unisa.it

1 Introduction

Tracing the main patterns of students mobility in higher education is one of the key issues in territorial policy. In fact, both the entire university and territorial governance systems are interested in understanding the factors behind students' decisions and outcomes with respect to university and degree program choice, and also inherent in post-enrollment choices such as inter-universities and/or degree program re-locations, and dropping out.

There are many factors that influence students' individual choices referred to the transition either from high school to bachelor's degree (first level mobility) or from bachelor's degree to master's degree (second level mobility). Sometime, these choices could lead to drop or churn out from the university system also in year-to-year transitions.

In literature, several studies have analyzed different aspects related to student mobility. Some scholars have dealt with outlining the typical patterns of student mobility in Italy, identifying the main trajectory from North to South [1, 2]. Others have focused on the definition of university attractiveness indices based on students flows among universities and territories [3, 4], also trying to highlight the push and pull factors behind the Italian students choices [5, 6]. Despite Universities implemented several measures to improve students' retention, some studies have identified important risk factors in university dropout [7, 8] and churn decisions [9, 10].

Moving from this framework, in this contribution we focus on students' churn decisions in the transition from the first to the second year of the Bachelor's degree program, defined on the bases of their choice to change disciplinary field and/or to move from the university of first enrolment. Looking at this churn decisions, three different settings could be observed: i) change Bachelor's degree program (i.e., moving from a Bachelor's degree program of one university to another Bachelor's degree program of the same university); ii) change university (i.e., moving from a Bachelor's degree program of one university to the same Bachelor's degree program in another university); and iii) change Bachelor's degree program and university (i.e., moving from a Bachelor's degree program of one university to another Bachelor's degree program in another university). Moreover, all these kinds of churn could be observed in Bachelor's, Master's and single-cycle degree programs, for all year-to-year transitions.

This contribution aims to define a quantitative measure that characterizes the flows between pairs of universities arising from students' choices. The use of Factorial Methods and Social Network Analysis allow to visualize the connections among universities according to students' flows. Based on a previous study [11], the proposed index weights a network characterizing student churn activity accounting for both the size of the university considered as a proxy of retention, and flows direction based on churning choices.

2 The churn propensity index

In this contribution we introduce the definition of a churn propensity index based on students' mobility flows among universities. We rely upon administrative data on the Italian University students collected from MOBYSU.IT database¹.

The analysis is conducted considering only students enrolled in a Bachelor's degree program. We consider only freshmen in the cohort 2018-2019 enrolled for the first time in a university who decide to churn in the transition from the first to the second year of their Bachelor's degree program. At this stage, all the disciplinary fields of the Bachelor's program are jointly considered.

Data are organized into a contingency table that considers on the rows the universities of enrollment (origin) and on the columns the universities (destination) chosen by students who changed university after the first year. In the scope of Social Network Analysis [12], this matrix can be seen as an origin-destination adjacency matrix. Its entries are interpreted as flows of students among universities. Consequently, when reading flows by rows, we refer to churning out students; while reading flows by columns, we refer to incoming students.

Let us notice that the sum of all the off-diagonal elements accounts for the total amount of students who decide to churn; while the sum of the elements on the main diagonal accounts for the total amount of stayers. Specifically, for each university, the corresponding element on the main diagonal is a proxy of its retention capability, the off-diagonal elements summed-up by each row gives the number of churning students as a degree of university disenrollment risk; while the off-diagonal elements summed-up by each column gives the number of incoming students as a degree of university attractiveness. In the language of social network analysis, these two quantities correspond to the nodes' out-degree and in-degree.

From this adjacency matrix, we obtain a one-mode, directed and weighted network representing the flows of churning students, where the nodes are the universities, and the edges weight depend on the occurrences of students that decide to churn. Thus, the origin nodes are the universities selected by students as first choice at the time of enrollment (i.e., in 2018), while the destination nodes are the universities choice after the churn decision (i.e, in 2019).

Thus, we propose a global churn propensity index as the ratio between the total number of churning students and the total number of stayers. This index can be specialized for each university as the ratio between the sum of the off-diagonal elements in a row and the corresponding diagonal element. Furthermore, we can specify this index also for each pair of universities as the ratio between each cell and the corresponding diagonal element in a row.

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

2.1 Some evidences from the Campania Region

In this contribution, we focus on students enrolled in a Bachelor's degree program at the universities located in the Campania region, in Italy. The seven universities in Campania, according to their main geographical province of reference, are: University of Naples Federico II, University of Naples "L'Orientale", Suor Orsola Benincasa University of Naples, and Parthenope University in province of Naples, University of Campania Luigi Vanvitelli in province of Caserta, University of Sannio in province of Benevento; and University of Salerno in province of Salerno. Moreover, we do not consider students enrolled in telematic universities, single-cycle and health professions degree programs.

The adjacency matrix holding the seven universities in the region counts the flows of 18,894 students enrolled in the academic year 2018-2019, of whom the 96% are stayers in the next year (a.y. 2019-2020). Only 4% of the students in this cohort churned after the first year, moving from the university of origin in 2018 toward a different university in 2019, regardless the disciplinary field.

We compute the churn propensity indices introduced in section 2 for all the university in the Campania region, for each single university, and for all possible pairs of universities. In the period under analysis, the global churn propensity index is equal to 4.18%, and it confirms a fairly low global propensity to churn. However, some differences emerge among the seven universities. Looking at the churn propensity indices computed for each university, the lower value is observed for the University of Salerno (1.8%), while the University of Sannio (12.5%) shows the higher one. Instead, by considering the indices computed for each pair of universities, relevant values are observed for the University of Sannio towards the universities of Salerno (4.3%), Naples Federico II (4.1%), and Vanvitelli (4.1%). Moreover, the highest churn values are observed from all universities in Campania towards the University of Naples Federico II; and from this latter towards Vanvitelli (2.2%) and Parthenope (1.3%) universities.

In the graph in Figure 1 (left side), both the churn flows and the role of each university are showed as a network; the size of the node-university is proportional to its churn propensity index, while the width of the edges depends on the occurrences of students flows linking pairs of universities

Furthermore, on the right side of Figure 1 there is the Biplot of the Correspondence Analysis computed for the matrix holding the churn propensity indices for all pairs of universities. Here, the proximity among origin universities (blue) shows similar churning out profiles and are displayed as barycenters of the universities of destination (red).

In conclusion, the graph representation reveals the central role of the University of Sannio in terms of churning students towards other universities (disenrollment risk); and the central position of the University of Naples Federico II who receives students from all the other universities (attractiveness). The important role of the University of Naples Federico II appears evident also in the biplot of the Correspondence Analysis, for both its role of origin and destination of the flows. In fact, in the first quadrant, the central position of the University of Naples Federico II (blue) can

be seen among the universities chosen by its churning students (Parthenope, Vanvitelli, Salerno and Orientale universities). While, in the third quadrant, we note the proximity between all the other universities in Campania whose students move towards the University of Naples Federico II (red) to continue their academic studies.

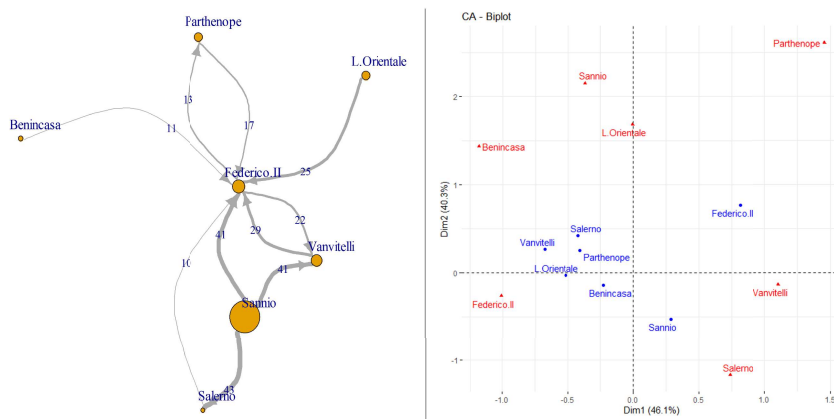


Fig. 1 Graphical representation of the churn-retention propensity index among the seven universities in the Campania region. Left side: network of churning students flows. Right side: Biplot of Correspondence Analysis of the churn propensity index.

3 Concluding remarks

This contribution proposes an exploratory strategy of analysis based on a numerical measure, a visual tool and a graph reading to discover university roles and students attitude to churn out from university of enrollment in the transition from the first to the second year of the Bachelor's degree program. Specifically, we propose a quantitative measure of university churn propensity based on students individual choices, defined at different levels of analysis, which allows us to identify the most relevant paths among universities according to students outgoing flows. The proposed approach considers a relative measure of churn propensity that take into account the retention capability of each university. It is assumed that churning flows are more politically relevant in those situation where the retention capability is low with respect to the whole amount of enrolled students. We presented a case study on the propensity to churn of students enrolled in universities in the Campania region in order to show the applicability of the proposed indices. Our analysis highlights some differences between universities located in the same territory, made even more evident by the proposed graphical representations. On the one hand, the network highlights the central position of the universities that receive the largest number of students who decided to churn, while the presence of many students leaving the origin universities is rendered by the size of the node-university. On the other hand,

Ilaria Primerano and Giuseppe Giordano

the biplot of the Correspondence Analysis, allows us to identify the universities with the most similar profiles, as well as showing the central position of the origin universities compared to the destination ones.

The churn propensity index can be further specified by considering, for example, students flows occurring within the same university, or among universities located into different territorial aggregations, and/or the flows within and between different degree programs, grouped according to the ISCED-F classification. It can also be taken into account in analyzing the attractiveness of the universities by considering the enrollments of churning students. Therefore, the attractiveness of the degree programs offered and the areas in which they are geographically located could also be analyzed. In general, it is best suited in all situation where the retention capability are relatively different with respect to the size (number of students enrolled) of the universities.

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., Albano, A.: Dalla triennale alla magistrale: continua la fuga dei cervelli dal Mezzogiorno d'Italia. *Neodemos*, ISSN: 2421–3209 (2019)
2. 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)
3. Columbu, S., Porcu, M., Primerano, I., Sulis, I., Vitale, M. P.: Correction to: Analysing the determinants of Italian university student mobility pathways. *Genus*, 78(1) (2022)
4. Columbu, S., Porcu, M., Primerano, I., Sulis, I., Vitale, M.P.: Geography of Italian student mobility: A network analysis approach. *Socio-Economic Planning Sciences*, 73, 100918 (2021)
5. Columbu, S., Porcu, M., Sulis, I.: University choice and the attractiveness of the study area: insights on the differences amongst degree programmes in Italy based on generalised mixed-effect models. *Socio-Economic Planning Sciences*, 74, 100926 (2021)
6. Santelli, F., Ragozini, G., Vitale, M.P.: Assessing the effects of local contexts on the mobility choices of university students in Campania region in Italy. *Genus* 78(5) (2022)
7. Contini, D., Salza, G.: Too few university graduates. Inclusiveness and effectiveness of the Italian higher education system. *Socio-Economic Planning Sciences*, 71, 100803 (2020)
8. Sarra, A., Fontanella, L., Di Zio, S. Identifying Students at Risk of Academic Failure Within the Educational Data Mining Framework. *Social Indicators Research* 146, 41-60 (2019)
9. La Rocca, M., Niglio, M., Restaino, M. Predicting university students' churn risk, In: Lombardo, R., Camminatello, I., Simonacci, V. (eds.) *Book of short Papers IES 2022 Innovation & Society 5.0: statistical and economic methodologies for quality assessment*, PKE srl (2022)
10. Primerano I., Santelli F., Usala C., Ragozini, G.: Exploiting students' inter-degree relocations to assess Italian universities' attractiveness. In *Book of Abstracts 9th International Conference on Risk Analysis*, ISBN: 978-972-674-919-6, pp. 60–62 (2022)
11. Giordano, G., Primerano, I., Vitale, P. A Network-Based Indicator of Travelers Performativity on Instagram. *Social Indicators Research*, 156(2), 631-649 (2020)
12. Wasserman, S., Faust, K. *Social network analysis: Methods and applications*, Cambridge University Press (1994)