SIS | 2022

51st Scientific Meeting
of the Italian Statistical Society

Caserta, 22-24 June

Università
degli Studi
della Campania
Luigi Vanvitelli

SIS
Società
Italiana di
Statistica

www.unicampania.it

Book of the Short Papers

Editors: Antonio Balzanella, Matilde Bini, Carlo Cavicchia, Rosanna Verde

Città di Caserta

800 1222-2022 ANNI

Università degli Studi di Padova

Dipartimento di Scienze Statistiche

Stata

sas

Università degli Studi del Sannio
Benevento

Pearson

Matilde Bini (Chair of the Program Committee) - *Università Europea di Roma*
Rosanna Verde (Chair of the Local Organizing Committee) - *Università della Campania "Luigi Vanvitelli"*

PROGRAM COMMITTEE

Matilde Bini (Chair), Giovanna Boccuzzo, Antonio Canale, Maurizio Carpita, Carlo Cavicchia, Claudio Conversano, Fabio Crescenzi, Domenico De Stefano, Lara Fontanella, Ornella Giambalvo, Gabriella Grassia - Università degli Studi di Napoli Federico II, Tiziana Laureti, Caterina Liberati, Lucio Masserini, Cira Perna, Pier Francesco Perri, Elena Pirani, Gennaro Punzo, Emanuele Raffinetti, Matteo Ruggiero, Salvatore Strozza, Rosanna Verde, Donatella Vicari.

LOCAL ORGANIZING COMMITTEE

Rosanna Verde (Chair), Antonio Balzanella, Ida Camminatiello, Lelio Campanile, Stefania Capecchi, Andrea Diana, Michele Gallo, Giuseppe Giordano, Ferdinando Grillo, Mauro Iacono, Antonio Irpino, Rosaria Lombardo, Michele Mastroianni, Fabrizio Maturo, Fiammetta Marulli, Paolo Mazzocchi, Marco Menale, Giuseppe Pandolfi, Antonella Rocca, Elvira Romano, Biagio Simonetti.

ORGANIZERS OF SPECIALIZED, SOLICITED, AND GUEST SESSIONS

Arianna Agosto, Raffaele Argiento, Massimo Aria, Rossella Berni, Rosalia Castellano, Marta Catalano, Paola Cerchiello, Francesco Maria Chelli, Enrico Ciavolino, Pier Luigi Conti, Lisa Crosato, Marusca De Castris, Giovanni De Luca, Enrico Di Bella, Daniele Durante, Maria Rosaria Ferrante, Francesca Fortuna, Giuseppe Gabrielli, Stefania Galimberti, Francesca Giambona, Francesca Greselin, Elena Grimaccia, Raffaele Guetto, Rosalba Ignaccolo, Giovanna Jona Lasinio, Eugenio Lippiello, Rosaria Lombardo, Marica Manisera, Daniela Marella, Michelangelo Misuraca, Alessia Naccarato, Alessio Pollice, Giancarlo Ragozini, Giuseppe Luca Romagnoli, Alessandra Righi, Cecilia Tomassini, Arjuna Tuzzi, Simone Vantini, Agnese Vitali, Giorgia Zaccaria.

ADDITIONAL COLLABORATORS TO THE REVIEWING ACTIVITIES

Ilaria Lucrezia Amerise, Ilaria Benedetti, Andrea Bucci, Annalisa Busetta, Francesca Condino, Anthony Cossari, Paolo Carmelo Cozzucoli, Simone Di Zio, Paolo Giudici, Antonio Irpino, Fabrizio Maturo, Elvira Romano, Annalina Sarra, Alessandro Spelta, Manuela Stranges, Pasquale Valentini, Giorgia Zaccaria.

Copyright © 2022

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891932310

Contents

Preface XXVII

1 Plenary Sessions 1

Causal inference in air pollution epidemiology 2
Francesca Dominici

Clustering of Attribute Data and Network 11
Anuška Ferligoj

Bayesian approaches for capturing the heterogeneity of neuroimaging experiments 17
Francesco Denti, Laura D'Angelo and Michele Guindani

2 Specialized Sessions 30

Advances in Bayesian nonparametric methodology 31

Repulsive mixture models for high-dimensional data 32
Lorenzo Ghilotti, Mario Beraha and Alessandra Guglielmi

Bayesian nonparametric mixtures of directed acyclic graph models 37
Federico Castelletti and Guido Consonni

Bayesian Clustering of Brain Regions via Extended Stochastic Block Models 45
Sirio Legramanti, Tommaso Rigon and Daniele Durante

Data Science skills for next generation statisticians 52

Cluster based oversampling for imbalanced learning 53
Gioia Di Credico and Nicola Torelli

Estimating the effect of remote teaching for university students through generalised linear mixed models 65
Silvia Bacci, Bruno Bertaccini, Simone Del Sarto, Leonardo Grilli and Carla Rampichini

Perceived stress across EU countries: does working from home impact? 71
Stefania Capecchi, Francesca Di Iorio and Nunzia Nappo

4 Contributed Sessions 844

Applications in Machine Learning 845

A neural network approach to survival analysis with time-dependent covariates for modelling time to cardiovascular diseases in HIV patients 846
Federica Corso, Agostino Lurani Cernuschi, Laura Galli, Chiara Masci, Camilla Muccini, Anna Maria Paganoni and Francesca Ieva

Analyzing the Correlation Structure of Financial Markets Using a Quantile Graphical Model 852
Beatrice Foroni, Luca Merlo and Lea Petrella

Neural Network for statistical process control of a multiple stream binomial process with an application to HVAC systems in passenger rail vehicles 858
Gianluca Sposito, Antonio Lepore, Biagio Palumbo and Giuseppe Giannini

Sparse signal extraction via variational SVM 864
Cristian Castiglione and Mauro Bernardi

Bayesian modelling and inference 1 870

Bayesian Inference for the Multinomial Probit Model under Gaussian Prior Distribution 871
Augusto Fasano, Giovanni Rebaudo and Niccolo Anceschi

Mapping Indicators on the Unit Interval: the tipsae Shiny App 877
Silvia De Nicolò and Aldo Gardini

A Bayesian spatio–temporal model of PM10 pollutant in the Po Valley 883
Matteo Gianella, Alessandra Guglielmi and Giovanni Lonati

Construction if a proper prior for a Bayesian envelope model 889
Andrea Mascaretti

Hilbert principal component regression for bimodal bounded responses 895
Enea G. Bongiorno, Agnese M. Di Brisco, Aldo Goia, and Sonia Migliorati

Methods of causal inference 901

Bayesian causal mediation analysis through linear mixed-effect models 902
Chiara Di Maria, Antonino Abbruzzo and Gianfranco Lovison

Bootstrap-aggregated adjustment set selection 908
Lorenzo Giammei

Exploiting partial knowledge to evaluate the average causal effect via an ABC perspective 914
Giulia Cereda, Fabio Corradi and Cecilia Viscardi

Intertemporal propensity score matching for casual inference: an application to covid-19 lockdowns and air pollution in Northern Italy 920
Daniele Bondonio and Paolo Chirico

Methods for Spatio-temporal data 926

Local Spatio-Temporal Log-Gaussian Cox Processes for seismic data analysis 927
Nicoletta D'Angelo, Giada Adelfio, and Jorge Mateu

Spatial explorative analysis of thyroid cancer in Sicilian volcanic areas 933
Francesca Bitonti and Angelo Mazza

Using geo-spatial topic modelling to understand the public view of Italian Twitter users: a climate change application 939
Yuri Calleo and Francesco Pilla

Comparing local structures of spatio-temporal point processes on linear networks 945
Nicoletta D'Angelo, Giada Adelfio, and Jorge Mateu

DISTATIS-based spatio-temporal clustering approach: an application to business cycles' time series 951
Raffaele Mattera and Germana Scepi

Developments in composite indicators 957

Bayesian Networks for monitoring the gender gap 958
Flaminia Musella, Lorenzo Giammei, Silvana Romio, Fulvia Mecatti and Paola Vicard

An Alternative Aggregation Function for the UNDP Human Development Index 964
Manuela Scioni and Paola Annoni

An ultrametric model for building a composite indicator system to study climate change in European countries 970
Giorgia Zaccaria and Pasquale Sarnacchiaro

Functional Weighted Malmquist Productive Index: a proposal for a dynamic composite indicator 975
Annalina Sarra, Eugenia Nissi and Tonio Di Battista

CFA & PLS-PM for UX-AI Product infused 981
Emma Zavarrone and Rosanna Cataldo

Fertility, adulthood, and economic uncertainty 987

Uncertainty and fertility intentions: a comparison between the Great Recession and the Covid-19 crisis 988
Chiara Ludovica Comolli

Interpreting the relationship between life course trajectories and explanatory factors. An example on the transition to adulthood 996
Danilo Bolano, Matthias Studer and Reto Buergin

The relationship between economic news and fertility: the case of Germany 1002
Maria Francesca Morabito, Raffaele Guetto, Matthias Vollbracht and Daniele Vignoli

Leaving home among Millennials in Italy: does economic uncertainty matter? 1008
Silvia Meggiolaro and Fausta Ongaro

Adverse pregnancy outcomes in The United Kingdom following unexpected job loss 1014
Alessandro Di Nallo and Selin Koksal

Bayesian modelling and inference 2 1020

A Bayesian beta linear model to analyze fuzzy rating responses 1021
Antonio Calcagnì, Massimiliano Pastore, Gianmarco Altoe and Livio Finos

A Mixture Model for Multi-Source Cyber-Vulnerability Assessment 1028
Mario Angelelli, Serena Arima and Christian Catalano

Hierarchical Bayesian models for analysing fish biomass data 1034
Rita Fici, Antonino Abbruzzo, Luigi Augugliaro and Giacomo Milisenda

Insights into the derivative-based method for nonlinear mediation models 1040
Claudio Rubino and Chiara Di Maria

An exploration of Approximate Bayesian Computation (ABC) and dissimilarities 1046
Laura Bondi, Marco Bonetti and Raffaella Piccarreta

Advances in Categorical and Preference data 1052

On the predictability of a class of ordinal data models 1053
Rosaria Simone and Domenico Piccolo

Multivariate analysis of binary ordinal data using graphical models 1059
Camilla Caroni, Fabio Alberto Comazzi, Andrea Deretti and Federico Castelletti

Multinomial Thompson Sampling for adaptive experiments with rating scales 1065
Nina Deliu

Ranking extraction in nested partially ordered data systems 1071
Marco Fattore, Barbara Cavalletti, Matteo Corsi and Alessandro Avellone

Towards the definition of distance measures in the preference-approval structures 1077
Alessandro Albano, Mariangela Sciandra and Antonella Plaia

Covid-19 Assessment and Evaluation 1 1083

Covid-19 impact assessment and inequality decomposition methods 1084
Federico Attili and Michele Costa

Multiversal methods for model selection: COVID-19 vaccine coverage and relative risk reduction 1090
Venera Tomaselli and Giulio Giacomo Cantone

Efficiency and feasibility of two stage sampling designs for estimating SARS-CoV-2 epidemic 1096
Pietro Demetrio Falorsi, Vincenzo Nardelli and Giuseppe Arbia

Evaluating the impacts of Covid-19 on the overall Italian death process via Functional Data Analysis 1102
Riccardo Scimone, Alessandra Menafoglio, Laura M. Sangalli and Piercesare Secchi

Developing countries, migration and migrants 1107

Domestic violence in Africa: a glance through the DHS survey 1108
Micaela Arcaio, Daria Mendola and Anna Maria Parroco

Inequalities in undernutrition among Roma and non-Roma children in Western Balkans: an analysis of the determinants 1114
Annalisa Busetta, Valeria Cetorelli and Chiara Puglisi

The manual, communicative and quantitative abilities of native and foreign workers according to their level of education in Italy 1120
Camilla Pangallo, Oliviero Casacchia and Corrado Polli

HIV Prevalence in some African Territories: Socio-Economic Drivers 1126
Micaela Arcaio, Daria Mendola and Anna Maria Parroco

A longitudinal cross country comparison of migrant integration policies via Mixture of Matrix-Normals 1132
Leonardo Salvatore Alaimo, Francesco Amato and Emiliano Seri

Education and job placement 1138

Measuring happiness at work with categorical Principal Component Analysis 1139
Ulpiana Kocollari, Maddalena Cavicchioli and Fabio Demaria

Early and accurate: a Machine Learning approach to predict students' final outcome with registry data 1146
Lidia Rossi, Marta Cannistrà and Tommaso Agasisti

Students' experience with distance learning during Covid 19 pandemic in Southern Italy 1153
Angela Maria D'Uggento and Nunziata Ribecco

Time series methods and Applications 1159

Trend and cycle decomposition in nonlinear time series 1160
Maddalena Cavicchioli

Asymptotic properties of the SETAR parameters: a new approach 1166
Marcella Niglio and Guy Mélard

Food prices forecast using post-sampled crowdsourced data with Reg-ARMA model: the case of Nigeria 1172
Ilaria Lucrezia Amerise, Gloria Solano Hermosilla, Vincenzo Nardelli and Giuseppe Arbia

Universal change point testing for dependent data 1178
Federica Spoto, Alessia Caponera and Pierpaolo Brutti

Change point detection in fruit bioimpedance using a three-way panel model 1184
F. Marta L. Di Lascio and Selene Perazzini

Bayesian modelling and inference 3 1190

A dynamic power prior approach to non-inferiority trials for normal means with unknown variance 1191
Francesco Mariani, Fulvio De Santis and Stefania Gubbiotti

Bayesian Change-Point Detection for a Brownian Motion with a Total Miss Criterion 1197
Bruno Buonaguidi

On the comparison of alternative Bayesian measures of posterior discrepancy 1203
Fulvio De Santis and Stefania Gubbiotti

A Bayesian Test for the comparison of two independent populations 1209
Mara Manca, Silvia Columbu and Monica Musio

A contribution to the L. J. Savage problem 1215
Francesco Bertolino, Silvia Columbu and Mara Manca

Methods for Complex Data 1221

Optimization of delayed rejection adaptive metropolis 1222
Daniele Raffo and Antonietta Mira

Dealing with multicollinearity and outliers in multinomial logit model: a simulation study 1228
Ida Camminatiello and Antonio Lucadamo

A tool to validate the assumptions on ratios of nearest neighbors' distances: the Consecutive Ratio Paths 1233
Francesco Denti and Antonietta Mira

Dimensionality reduction and visualization for interval-valued data via midpoints-ranges principal component analysis 1239
Viviana Schisa, Alfonso Iodice D'Enza and Francesco Palumbo

Data-driven design-based mapping of forest resources 1245
Sara Franceschi, Rosa Maria Di Biase, Lorenzo Fattorini, Marzia Marcheselli and Caterina Pisani

Environmental data and Climate change 1252

Ensemble model output statistics for temperature forecasts in Veneto 1253
Gaetan Carlo, Giummole Federica, Mameli Valentina and Siad Si Mokrane

State of the urban Environment in Italy. A comparative analysis of selected composite indicators 1259
Giuseppe Lecardane

A Functional Data Analysis approach for Climate Model Selection: the case study of Campania Region 1266
Veronica Villani, Elvira Romano and Paola Mercogliano

Evolution of scientific literature on climate change: a bibliometric analysis 1273
Gianpaolo Zammarchi, Giulia Contu, Maurizio Romano

Energy and material demand of the Italian Regions 1279
Flora Fullone, Giulia Iorio, Assunta Lisa Carulli

Health and survivorship 1285

Increasing Inequalities in Mortality by Socioeconomic Position in Italy 1286
Chiara Ardito, Nicolás Zengarini, Roberto Leombruni, Angelo d'Errico and Giuseppe Costa

The role of health conditions in the relationship between socio-economic status and well-being: the counterfactual approach in mediation models 1296
Sara Manzella and Margherita Silan

Excess economic burden of multimorbidity: a population-based study in Italy 1302
Chiara Seghieri, Niccolò Borri, Gaia Bertarelli and Sabina Nuti

Depression-free life expectancy among 50 and older Americans by gender, race/ethnicity and education: the effect of marital disruption 1308
Alessandro Feraldi and Cristina Giudici

Disability-free grandparenthood in Italy. Trends and gender differences 1314
Margherita Moretti, Elisa Cisotto and Alessandra De Rose

Advances in regression models 1320

Semiparametric M-quantile regression for modelling georeferenced housing price data 1321
Riccardo Borgoni, Antonella Carcagnì, Alessandra Michelangeli, Nicola Salvati and Francesco Schirripa Spagnolo

Resampling-based inference for high-dimensional regression 1327
Anna Vesel, Jelle J. Goeman, Angela Andreella and Livio Finos

Quantile regression coefficient modeling for counts to evaluate the productivity of university students 1333
Viviana Carcaiso and Leonardo Grilli

Adaptive smoothing spline using non-convex penalties 1339
Daniele Cuntrera and Vito M.R. Muggeo

Conditional tests for generalized linear models 1345
Riccardo De Santis, Jelle J. Goeman, Anna Vesely and Livio Finos

Methods and applications in economics and finance 1351

Mixed models for anomaly detection in anti-money laundering aggregate reports 1352
Stefano Iezzi and Marianna Siino

On the drivers of Greenwashing risk: evidence from Eurostoxx600 1358
Yana Kostiuk, Costanza Bosone and Paola Cerchiello

Modelling Financial Returns with Finite Mixtures of GED 1364
Pierdomenico Duttilo and Stefano Antonio Gattone

Risk Parity strategy for portfolio construction: a kurtosis-based approach 1370
Maria Debora Braga, Consuelo Rubina Nava and Maria Grazia Zoia

Fully reconciled probabilistic GDP forecasts from Income and Expenditure sides 1376
Tommaso Di Fonzo and Daniele Girolimetto

Latent Class models 1382

Latent thresholds model in classification tasks 1383
Giuseppe Mignemi, Andrea Spoto and Antonio Calcagnì

Adaptive filters for time-varying correlation parameters 1389
Michele Lambardi di San Miniato, Ruggero Bellio, Luca Grassetti and Paolo Vidoni

Bayesian structural learning for Latent Class Model with an application to Record Linkage 1395
Davide Di Cecco

Multilevel Latent Class modelling to advise students in self-learning platforms: an application in the context of learning Statistics 1401
Roberto Fabbricatore, Zsuzsa Bakk, Roberto Di Mari, Mark de Rooij and Francesco Palumbo

Latent Markov models with associated mixed responses 1407
Alfonso Russo and Alessio Farcomeni

Methods for health studies 1413

Beyond the fragility index 1414
Piero Quatto and Enrico Ripamonti

Evaluation of the diagnostic-therapeutic paths for schizophrenic patients through state sequences analysis 1419
Laura Savaré, Giovanni Corrao and Francesca Ieva

Optimal timing of bone-marrow transplant in myelodysplastic syndromes through multi-state modeling and microsimulation 1425
Caterina Gregorio, Marta Spreafico and Francesca Ieva

A fully Bayesian approach for sample size determination of Poisson clinical trials 1431
Susanna Gentile and Valeria Sambucini

Compartmental models in epidemiology: Application on Smoking Habits in Tuscany 1437
Alessio Lachi, Cecilia Viscardi, Maria Chiara Malevolti, Giulia Carreras and Michela Baccini

Covid-19 Assessment and Evaluation 2 1443

We are in the same storm but not in the same boat: Impact of COVID-19 on UK households 1444
Demetrio Panarello and Giorgio Tassinari

A network approach to investigate learning experiences and social support in higher education 1450
Ilaria Primerano, Maria Carmela Catone, Giuseppe Giordano, Maria Prosperina Vitale

Physical and cultural activity, internet use and anxiety of Italian university students during the pandemic 1456
Giovanni Busetta, Maria Gabriella Campolo and Demetrio Panarello

The digital divide in Italy before and during the pandemic phase 1462
Laura Zannella

Covid-19 and financial professional advice 1468
Marianna Brunetti and Rocco Ciciretti

Bayesian modelling and inference 4 1472

Bayesian functional mixed effects model for sports data 1473
Patric Dolmeta, Raffaele Argiento and Silvia Montagna

Bayesian Optimization with Machine Learning for Big Data Applications in the Cloud 1479
Bruno Guindani, Danilo Ardagna and Alessandra Guglielmi

Confidence distributions and fusion inference for intractable likelihoods 1485
Elena Bortolato and Laura Ventura

Wasserstein distance and applications to Bayesian nonparametrics 1491
Marta Catalano, Hugo Lavenant, Antonio Lijoi and Igor Prunster

Network Analysis and community detection 1497

Community detection in networks: a heuristic version of Girvan Newman algorithm 1498
Ilaria Bombelli and Lorenzo Di Rocco

Geographically weighted regression for spatial network data: an application to traffic volumes estimation 1504
Andrea Gilardi, Riccardo Borgoni and Jorge Mateu

Asymmetric Spectral Clustering: a comparison between symmetrizations 1510
Cinzia Di Nuzzo and Donatella Vicari

Community detection of seismic point processes 1516
Valeria Policastro, Nicoletta D'Angelo and Giada Adelfio

An Explorative analysis of Different Distance Metrics to Compare Unweighted Undirected Networks 1522
Anna Simonetto, Matteo Ventura and Gianni Gilioli

Gender, attitudes and family ties 1528

Parents of a disabled child in Italy: less healthy but more civically engaged 1529
Nicoletta Balbo and Danilo Bolano

Searching the nexus between women's empowerment and female genital cutting (FGC) 1535
Patrizia Farina, Liva Ortensi, Thomas Pettinato and Enrico Ripamonti

Social stratification, gender, and attitudes towards voluntary childlessness in Europe: A double machine learning approach 1539
Danilo Bolano and Francesco C. Billari

Integrating structuralism and diffusionism to explain the new Italian emigration 1545
Francesca Bitonti

On the effects of rooted family ties in business networks: The South of Italy in the 19th century 1551
Roberto Rondinelli, Giancarlo Ragozini and Maria Carmela Schisani

Methods and Applications in Clustering 1557

A semi-supervised clustering method to extract information from the electronic Word Of Mouth 1558
Giulia Contu, Luca Frigau, Maurizio Romano and Marco Ortu

Spectral approach for clustering three-way data 1564
Cinzia Di Nuzzo and Salvatore Ingrassia

Double clustering with a matrix-variate regression model: finding groups of athletes and disciplines in decathlon's data 1570
Mattia Stival, Mauro Bernardi, Manuela Cattelan and Petros Dellaportas

Classification of the population dynamics 1576
Federico Bacchi and Laura Neri

Locating γ-Ray Sources on the Celestial Sphere via Modal Clustering 1582
Anna Montin, Alessandra R. Brazzale and Giovanna Menardi

Sampling and Official Statistics 1588

Fisher's Noncentral Hypergeometric Distribution for Population Size Estimation 1589
Veronica Ballerini and Brunero Liseo

Small area models for skew and kurtotic distributions 1595
Maria Rosaria Ferrante and Lorenzo Mori

The use of remotely sensed data in sampling designs for forest monitoring 1601
Chiara Bocci, Gherardo Chirici, Giovanni D'Amico, Saverio Francini and Emilia Rocco

Analyzing different causes of one–inflation in capture recapture models for criminal populations 1607
Davide Di Cecco, Andrea Tancredi and Tiziana Tuoto

Administrative database and official statistics: an IT and statistical procedure 1613
Caterina Marini and Vittorio Nicolardi

Spatial modeling and Analyses 1619

Spatial statistics analysis using microdata: an application at agricultural sector 1620
Daniela Fusco, Maria Antonietta Liguori, Valerio Moretti and Francesco Giovanni Truglia

Bayesian spatial modeling of extreme precipitation 1627
Federica Stolf

A proposal to adjust local Moran's I for measuring residential segregation 1632
Antonio De Falco and Antonio Irpino

Accurate directional inference for gaussian graphical models 1637
Claudia Di Caterina, Nancy Reid and Nicola Sartori

Advances in Classification 1643

Measures of interrater agreement based on the standard deviation 1644
Giuseppe Bove

A Comparison of accuracy measures for Classification tasks 1650
Amalia Vanacore and Maria Sole Pellegrino

Iterative Threshold-based Naive Bayes Classifier: an efficient Tb-NB improvement 1656
Maurizio Romano, Gianpaolo Zammarchi and Giulia Contu

Reprogramming FairGANs with Variational Auto-Encoders: A New Transfer Learning Model 1662
Beatrice Nobile, Gabriele Santin, Bruno Lepri and Pierpaolo Brutti

Robust statistics 1669

Combinatorial Analysis of Factorial Designs with Ordered Factors 1670
Roberto Fontana and Fabio Rapallo

Robustifying the Rasch model with the forward search 1676
Anna Comotti and Francesca Greselin

A novel estimation procedure for robust CP model fitting 1682
Valentin Todorov, Violetta Simonacci, Michele Gallo and Nikolay Trendafilov

A robust approach for functional ANOVA with application to additive manufacturing 1688
Fabio Centofanti, Bianca Maria Colosimo, Marco Luigi Grasso, Alessandra Menafoglio, Biagio Palumbo and Simone Vantini

Modeling unconditional M-quantiles in a regression framework 1692
Luca Merlo, Lea Petrella and Nicola Salvati

Model-based clustering 1696

Bayesian mixtures of semi-Markov models 1697
Rosario Barone and Andrea Tancredi

Specification of informative priors for capture-recapture finite mixture models 1703
Pierfrancesco Alaimo Di Loro, Gianmarco Caruso, Marco Mingione, Giovanna Jona Lasinio and Luca Tardella

Clustering multivariate categorical data: a graphical model-based approach 1709
Francesco Rettore, Michele Russo, Luca Zerman and Federico Castelletti

The Gaussian mixture model-based clustering for the comparative analysis of the Healthcare Digitalization Index in the Italian local health authorities 1715
Margaret Antonicelli, Michele Rubino and Filomena Maggino

Student performance evaluation 1721

Rasch model versus Rasch Mixture model: strengthens and limits in identifying factors affecting students’ performance in mathematics 1722
Clelia Cascella

Does taking additional Maths classes improve university performance? 1728
Martina Vittorietti, Andrea Priulla and Massimo Attanasio

University dropout and churn in italy: an analysis over time 1734
Barbara Barbieri, Mariano Porcu, Luisa Salaris, Isabella Sulis, Nicola Tedesco and Cristian Usala

The ANOGI for detecting the impact of education and employment on income inequality 1740
Elena Fabrizi, Alessio Guandalini and Alessandra Spagnoli

What causes juvenile crime? a case-control study 1747
Elena Dalla Chiara and Federico Perali

Methods and Applications in Survival analysis 1753

Recursive partitioning for survival data 1754
Ambra Macis

Detecting survival patterns in a digital learning platform 1760
Marta Cannistrà, Mara Soncin and Federico Frattini

An extension of proper Bayesian bootstrap ensemble tree models to survival analysis 1766
Elena Ballante

Modelling time to university dropout by means of time-dependent frailty COX PH models 1771
Mirko Giovio, Paola Mussida and Chiara Masci

Family history in survival and disease development 1777
Maria Veronica Vinattieri and Marco Bonetti

Text mining 1783

Topics & metaverse: an explorative analysis 1784
Emma Zavarrone, Alessia Forciniti, Emanuele Parisi, Maria Gabriella Grassia

Applying Topic Models to bibliographic search: some results in basketball domain 1791
Manlio Migliorati and Eugenio Brentari

Exploiting Text Mining and Network Analysis for future scenarios development: an application on remote working 1797
Yuri Calleo, Simone Di Zio and Vanessa Russo

Emotion recognition in Italian political language to predict positionings and crises government 1803
Alessia Forciniti and Emma Zavarrone

What does your self-description reveal about you? 1809
Riccardo Ricciardi

Variable selection and complete matrix approaches 1815

A Statistical Approach for the Completion of Input-Output Tables 1816
Rodolfo Metulini, Giorgio Gnecco, Francesco Biancalani and Massimo Riccaboni

On multivariate records over sequences of random vectors with Marshall-Olkin dependence of components 1822
A. Khorrami Chokami and Simone A. Padoan

The joint censored gaussian graphical lasso model 1829
Gianluca Sottile, Luigi Augugliaro and Veronica Vinciotti

Variable selection with unbiased estimation: the cdf penalty 1835
Daniele Cuntrera, Vito M.R. Muggeo and Luigi Augugliaro

Automatic variable selection for MIDAS regressions: an application 1841
Consuelo Rubina Nava, Luigi Riso and Maria Grazia Zoia

Distribution Theory and Estimation 1847

A general framework for unit distributions 1848
Francesca Condino, Filippo Domma and Bozidar V. Popovic

Prediction intervals based on multiplicative model combinations 1854
Valentina Mameli and Paolo Vidoni

Some advances on pairwise likelihood estimation in ordinal data latent variable models 1860
Giuseppe Alfonzetti and Ruggero Bellio

Functional Data Analysis 1866

A new functional clustering method: the Functional Clustering and Dimension Reduction model 1867
Adelia Evangelista and Stefano Antonio Gattone

Nonparametric functional prediction bands: theory with an application to bike sharing mobility demand in the city of Milan 1873
Jacopo Diquigiovanni, Matteo Fontana and Simone Vantini

An R package for the statistical process monitoring of functional data 1878
Christian Capezza, Fabio Centofanti, Antonio Lepore, Alessandra Menafoglio, Biagio Palumbo and Simone Vantini

Trend filtering for functional regression 1884
Federico Ferraccioli, Alessandro Casa and Marco Stefanucci

Conformal prediction for spatio-functional regression models 1890
Diana, Romano, Irpino

Tourism and sport studies 1895

Assessing satisfaction of tourists visiting Italian museums: evidence from the eWOM 1896
Daria Mendola and Valentina Oddo

COVID-19 pandemic and tourism demand: a comparison between Spain and Italy 1902
Caterina Sciortino, Ludovica Venturella and Stefano De Cantis

A compositional analysis of tourism in Europe 1908
Francesco Porro

Improving administrative data quality on tourism using Big Data 1914
Antonella Bianchino, Armando d'Aniello and Daniela Fusco

Geographical variations of socio-demographic issues 1920

Elderly HCE and health care need: comparing spatially unexplained levels 1921
Irene Torrini, Laura Rizzi and Luca Grassetti

Measuring sustainable development at the regional level. The case of Italy 1927
Marianna Bartiromo and Enrico Ivaldi

Socio-economic deprivation and COVID-19 infection: a Bayesian spatial modelling approach 1933
Antonino Abbruzzo, Andrea Mattaliano, Alessandro Arrigo, Salvatore Scondotto and Mauro Ferrante

Applications in Economics 1939

The measurement of economic security through relative indicators 1940
Alessandro Gallo, Silvia Pacei and Maria Rosaria Ferrante

A regional analysis of the efficiency by energy's producers in Italy 1946
Gianna Greca, Giuseppe Cinquegrana and Giovanni Fosco

On investigating social and financial aspects of Cardano 1953
Stefano Vacca, Marco Ortu, Gianpaolo Zammarchi and Giuseppe Destefanis

Combined permutation test on the effect of age of micro enterprises on the propensity to Circular Economy 1959
Stefano Bonnini and Michela Borghesi

Comparison of Two Different Approaches to Measure Economic Access to Food and Insecurity: an Application to Mexican data 1965
Stefano Marchetti, Luca Secondi and Adrian Vargas-Lopez

Image analysis and visual methods 1971

Bias correction of the maximum likelihood estimator for Emax model at the interim analysis 1972
Caterina May and Chiara Tommasi

Visual and automated methods in digital microscopy to evaluate fungal colonisation on plant roots 1977
Ivan Sciascia, Andrea Crosino and Andrea Genre

From satellite images to road pavement type: an object-oriented classification approach 1983
Arianna Burzacchi, Matteo Landrò and Simone Vantini

Valid inference for group analysis of functionally aligned fMRI images 1987
Angela Andreella, Riccardo De Santis and Livio Finos

Topological persistence for astronomical image segmentation 1993
Riccardo Ceccaroni, Pierpaolo Brutti, Marco Castellano, Adriano Fontana and Emiliano Merlin

Statistical assessment and empirical estimation 1999

Confidence regions for optimal sensitivity and specificity of a diagnostic test 2000
Gianfranco Adimari, Duc-Khanh To and Monica Chiogna

On the sensitiveness to the memory parameter in the network of tennis 2006
Alberto Arcagni, Vincenzo Candila and Rosanna Grassi

Two-part model with measurement error 2011
Maria Felice Arezzo, Serena Arima, and Giuseppina Guagnano

Statistical assessment of practical significance 2017
Andrea Ongaro, Sonia Migliorati, and Enrico Ripamont

Autoregressive and mixed effects models 2023

Asymptotic Properties of the Nonlinear Least Squares Estimator in HE-HAR Models 2024
Emilija Dzuverovic and Edoardo Otranto

A note on testing for threshold non-linearity in presence of heteroskedasticity in time series 2030
Simone Giannerini and Greta Goracci

The conditional autoregressive Whart-G model 2036
Massimiliano Caporin and Marco Girardi

Semi-parametric generalized linear mixed effects models for binary response for the analysis of heart failure hospitalizations 2042
Alessandra Ragni, Chiara Masci, Francesca Ieva and Anna Maria Paganoni

Issues in Data science 2048

etree: Classification and Regression With Structured and Mixed-Type Data in R 2049
Riccardo Giubilei, Tullia Padellini and Pierpaolo Brutti

Deep Learning framework for ungrouping coarsely aggregated vital rates 2055
Andrea Nigri

Inside the metaverse: analysis of the state of the art and development of a new usage approach based on quality and ethics 2061
Vito Santarcangelo, Emilio Massa, Saverio Gianluca Crisafulli, Antonio Ruoto, Angelo Lamacchia, Alessandro D'Alcantara, Alessandro Verderame and Massimiliano Giacalone

Preface

This book includes the contributions presented at the 51st Scientific Meeting of the Italian Statistical Society (SIS) held in Caserta at the Università della Campania "Luigi Vanvitelli", from the 22nd to 24th of June, 2022.

The conference has registered more than 300 presentations, including 4 keynotes in plenary invited sessions and 9 presentations in 3 guest sessions, 48 presentations collected in 16 specialized sessions and 68 presentations in 17 solicited sessions, all dealing with specific themes in methodological and/or applied statistics and demography. Furthermore, more than 200 contributions, with one or more authors, have been spontaneously submitted to the Program Committee and arranged in 43 contributed sessions.

The high number of contributions and the large participation at the conference show that researchers have met the challenge of pursuing working even in the face of the pandemic period from which we are only now emerging. The research activity in our field therefore has never stopped, and the desire to participate in scientific events, as a place for exchange and discussion on new developments in our field, remains a living characteristic of our scientific community.

With the publication of this book, we wish to offer to all members of the Italian Statistical Society, all international academics, researchers, Ph.D. students, and all interested practitioners, a good snapshot of the on-going research in the statistical and demographic fields. We deeply thank all contributors for having submitted their works to the conference and all the researchers for their remarkable job in acting as referees accurately and timely. We also would like to thank the International Biometric Society (IBS) – Italian region, the European Network for Business and Industrial Statistics (ENBIS) and the Italian Society of Statistical Physics (SIFS) we had the pleasure of hosting. A special thanks is addressed to the Scientific and Organizational Committees for their great efforts devoted to all the organizational aspects, to the Università della Campania "Luigi Vanvitelli" and to the Department of Mathematics and Physics who made this event possible, as well as to the Municipality of the Town of Caserta who has patronized the event and to all the funders for their supports.

Finally, we wish to express our gratitude to the publisher Pearson Italia for all the support received.

Asymmetric Spectral Clustering: a comparison between symmetrizations

Spectral clustering asimmetrico: un confronto tra simmetrizzazioni

Cinzia Di Nuzzo and Donatella Vicari

Abstract In this work, spectral clustering of asymmetric data is addressed. In particular, two different methods to perform clustering have been compared: the application of spectral clustering on directed graphs represented by an asymmetric matrix; and the application of the classical spectral clustering algorithm once transformed the directed graph into an undirected one. To this end, some symmetrizations are described to convert the directed graph to an undirected one.

Abstract *In questo lavoro viene trattato lo spectral clustering di dati asimmetrici. In particolare, sono state confrontate due diverse metodologie per effettuare il clustering: l'applicazione dello spectral clustering su grafi diretti rappresentati da una matrice asimmetrica; e l'applicazione del classico algoritmo di spectral clustering dopo la trasformazione del grafo diretto in uno non diretto. A tal fine, sono stati descritte alcune simmetrizzazioni per convertire il grafo diretto in uno non diretto.*

Key words: spectral clustering, directed graph, symmetrizations

1 Introduction

Clustering on directed graphs has several useful applications, just think of network and social network analysis. So far, few spectral methods have been proposed for clustering this type of data, see [8], [3] and [2]. Many clustering methods on directed graphs rely on the symmetrization of the weighted matrix of the data by taking into account the out-degrees between the vertices of the graph. Once such a symmetrization has been computed, an undirected graph can be associated with the resulting symmetric matrix, and the standard clustering methods on undirected graphs can be applied. Here, we propose to apply the spectral clustering method proposed by [5]. In particular, three different ways to symmetrize an asymmetric matrix are presented: the first method is based on the symmetrization of the asymmetric weighted matrix associated with the directed graph; the second one is based on the symmetrization of the weighted matrix considering the out-degrees of the graph, while the last symmetrization takes into account both the in-degrees and out-degrees of

---

Cinzia Di Nuzzo
Dipartimento di Scienze Statistiche - Sapienza Università di Roma e-mail: `cinzia.dinuzzo@uniroma1.it`

Donatella Vicari
Dipartimento di Scienze Statistiche - Sapienza Università di Roma e-mail: `donatella.vicari@uniroma1.it`

the directed graph. The spectral clustering, proposed in [5] and applied to these three symmetrizations, has been compared with the asymmetric spectral clustering proposed by [8]. The rest of the paper is structured as follows: in Section 2, some definitions on directed graphs are given. In Section 3, the symmetrizations of the asymmetric matrix associated with the directed graph, and the undirected spectral clustering method are summarized. In Section 4 the asymmetric spectral clustering proposed by [8] is summarized. Section 5 provides a numerical example where a comparison is carried out.

2 Background theory

Let $\mathscr{G}=(V,E)$ be a directed graph, where $V=\{v_1,\ldots,v_n\}$ is the set of n vertices or nodes, $E\subseteq V\times V$ is the set of the edges or arcs. In a directed graph, an edge is an ordered pair (v_i,v_j), where $v_i,v_j\in V$ for $i,j\in\{1,\ldots,n\}$. A tuple of vertices $(v_1,v_2,\ldots,v_p)$ with $(v_h,v_{h+1})\in E$ for $1\leq h\leq p-1$, is a *path*. If for each pair of vertices v_i and v_j there is a path where $v_1=v_i$ and $v_p=v_j$, then the directed graph is *strongly connected*. When $\mathscr{G}$ is a strongly connected graph, there is an integer $l\geq 1$ and a unique partition $V=V_0\cup V_1\cup\cdots\cup V_{l-1}$ such that for all $0\leq r\leq l-1$ each edge $(v_i,v_j)\in E$ with $v_i\in V_r$ and $v_j\in V_{r+1}$, and where $V_l=V_0$. The graph $\mathscr{G}$ is *weighted* when there is a weight function $w:E\rightarrow\mathbb{R}^+$ which associates a positive value $w((v_i,v_j))$ to each edge $(v_i,v_j)\in E$. Let $\mathscr{G}=(V,E,W)$ be a weighted directed graph, where W is the *asymmetric weighted matrix* associated to $\mathscr{G}$. The out-degree and the in-degree of a vertex v_i are defined by $d_i^+=\sum_{j=1}^n w_{ij}$ and $d_i^-=\sum_{j=1}^n w_{ji}$, respectively. Let $D_o=$diag(d^+) be the diagonal matrix of the vertex-out degrees. Given a weighted directed graph $\mathscr{G}$, there is a natural random walk on the graph with *transition probability matrix of the Markov chain* equal to matrix $P=D_o^{-1}W$ with entries $p_{ij}=w_{ij}/d_i^+$. When $\mathscr{G}$ is assumed to be a strongly connected graph, the Markov chain P is *irreducible* and has a unique stationary distribution π, i.e. a unique probability distribution π (with n non-negative entries summing to one) that satisfies the following balance equations

$$P'\pi=\pi, \tag{1}$$

where $\pi=(\pi_1,\ldots,\pi_n)'$. Let us denote Π the diagonal matrix having π as main diagonal.

3 Symmetrizations of the asymmetric matrix W and spectral clustering algorithm

Several attempts have been made to generalize the spectral graph theory of undirected graphs to directed graphs and they have been mainly addressed towards the symmetrization of the weight matrix associated to the graph. The graph symmetrizations proposed in literature are summarized below (see [6] for further details).

First symmetrization: $W+W'$. The simplest symmetrization is obtained as

$$U=W+W'. \tag{2}$$

Note that this symmetrization does not take into account the in- or out- links between the vertices of the graph.

Second symmetrization: Random walk. The random walk symmetrization proposed by [1] is defined as

$$U_{rw} = \frac{\Pi P + P'\Pi}{2}. \tag{3}$$

This symmetrization is based on the stationary distribution Π of the graph and does not take into account the in-degrees in each vertex.

Third symmetrization: Degree-discounted. In order to symmetrize the graph, both in- and out- degrees are taken into account in [6].The idea behind this is to look for a similarity measure that takes into account the in- and out-degrees between the vertices of the directed graph. In fact, nodes with high degrees will share many common edges (in- or out-) with other nodes simply by virtue of their higher degrees. The idea is to weigh the edges so that vertices with high (in- or out-) degree are weighted less in the symmetrization process. The reason is very simple: vertices with high (in- or out-) degrees are actually less informative in describing the links between the vertices of the graph than nodes with lower degrees. The symmetrization proposed in [6] is given by the following expression

$$U_d = D_o^{-1/2} W D_i^{-1/2} W' D_o^{-1/2} + D_i^{-1/2} W' D_o^{-1/2} W D_i^{-1/2} \tag{4}$$

where D_i =diag(d^-) is the diagonal matrix of the in-degrees.

Once a directed graph has been transformed into an undirected graph, the spectral clustering algorithm proposed in [5] can be can be run as summarized in Algorithm 1.

Algorithm 1 Spectral Clustering algorithm (see [5])

Let $\mathscr{G}_u = (V_u, E_u, W_u)$ be a undirected weighted graph, where $V_u = \{v_i^u, \ldots, v_n^u\}$ is the set of the vertices, E_u is the set of the edges, and $W_u = \left(w_{ij}^u\right)$ is the symmetric weighted matrix associated to $\mathscr{G}_u$.

Input: Symmetric weighted matrix W_u; number of clusters K.

1. Compute the degree matrix D_u =diag(d_i^u) associated to the undirected graph $\mathscr{G}_u$, where d_i^u is the degree of the vertex v_i^u associated to the undirected graph $\mathscr{G}_u$ and it is computed as $d_i^u = \sum_{i \neq j} w_{ij}^u$
2. Calculate the normalized symmetric Laplacian matrix $L_{\text{sym}} = D_u^{-1/2} W_u D_u^{-1/2}$.
3. Consider matrix Y having the eigenvectors associated to the K largest eigenvalues of L_{sym} as columns
4. Re-normalize the rows of Y to have unit length yielding $X \in \mathbb{R}^{n \times K}$.
5. Perform k-means algorithm on matrix X.

---

4 Asymmetric Spectral Clustering

Let us suppose to divide the graph into K subgraphs. In order to find the best partition of the graph $\mathscr{G}$, in the standard undirected spectral clustering method a normalized cut is introduced and the clustering is obtained thanks to a minimization of this cut (see [7] and [4]). An extension to directed graphs has been introduced by [8], where the directed $Ncut$ of a cluster C is defined as

$$Ncut_{dir}(C) = \frac{\sum_{i \in C, j \in \bar{C}} \pi_i p_{ij}}{\sum_{i \in C} \pi_i} + \frac{\sum_{i \in C, j \in \bar{C}} \pi_j p_{ji}}{\sum_{j \in \bar{C}} \pi_j}. \tag{5}$$

In [8], the $Ncut_{dir}$ is minimized by the eigenvectors of the following matrix

$$\Theta = \frac{\Pi^{1/2}P\Pi^{-1/2} + \Pi^{-1/2}P'\Pi^{1/2}}{2}. \tag{6}$$

The Asymmetric Spectral Clustering is summarized in Algorithm 2.

Algorithm 2 Asymmetric Spectral Clustering algorithm (see [8])

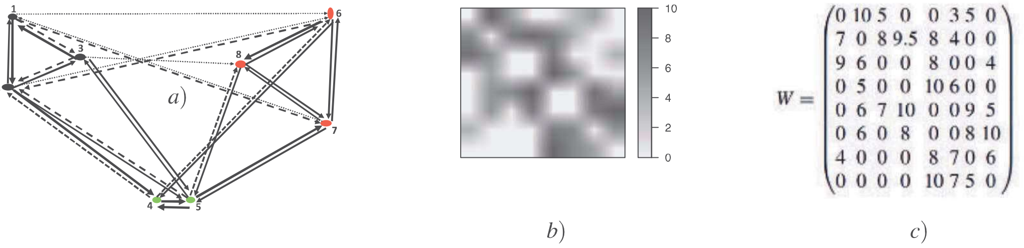
```
Input: Asymmetric weighted matrix $W$ associated to directed graph $\mathscr{G}$; number of clusters $K$.

1. Compute the transition matrix $P$ and the diagonal matrix $\Pi$ from its unique stationary distribution $P$.
2. Compute the (symmetric) Laplacian matrix $\Theta$ (6).
3. Consider the eigenvectors of $\Theta$ and introduce matrix $Z$ with columns equal to the eigenvectors $\gamma_1, \dots, \gamma_K$ associated with the $K$ largest eigenvalues of $\Theta$.
4. Perform $k$-means algorithm on matrix $Z$.
```

5 Numerical experiments

In order to compare the methods described in Sections 3 and 4, we present an example with artificial data.

The artificial data are described in Figure 1, the graph $\mathscr{G} = (V, E, W)$ is composed by three groups: $C_1 = \{v_1, v_2, v_3\}$, $C_2 = \{v_4, v_5\}$ and $C_3 = \{v_6, v_7, v_8\}$. Specifically, in Figure 1-a) the edges of the graph $\mathscr{G}$ are displayed with different thickness according to their weights (i.e., solid , dashed and dotted lines represent weights in descending order, respectively).



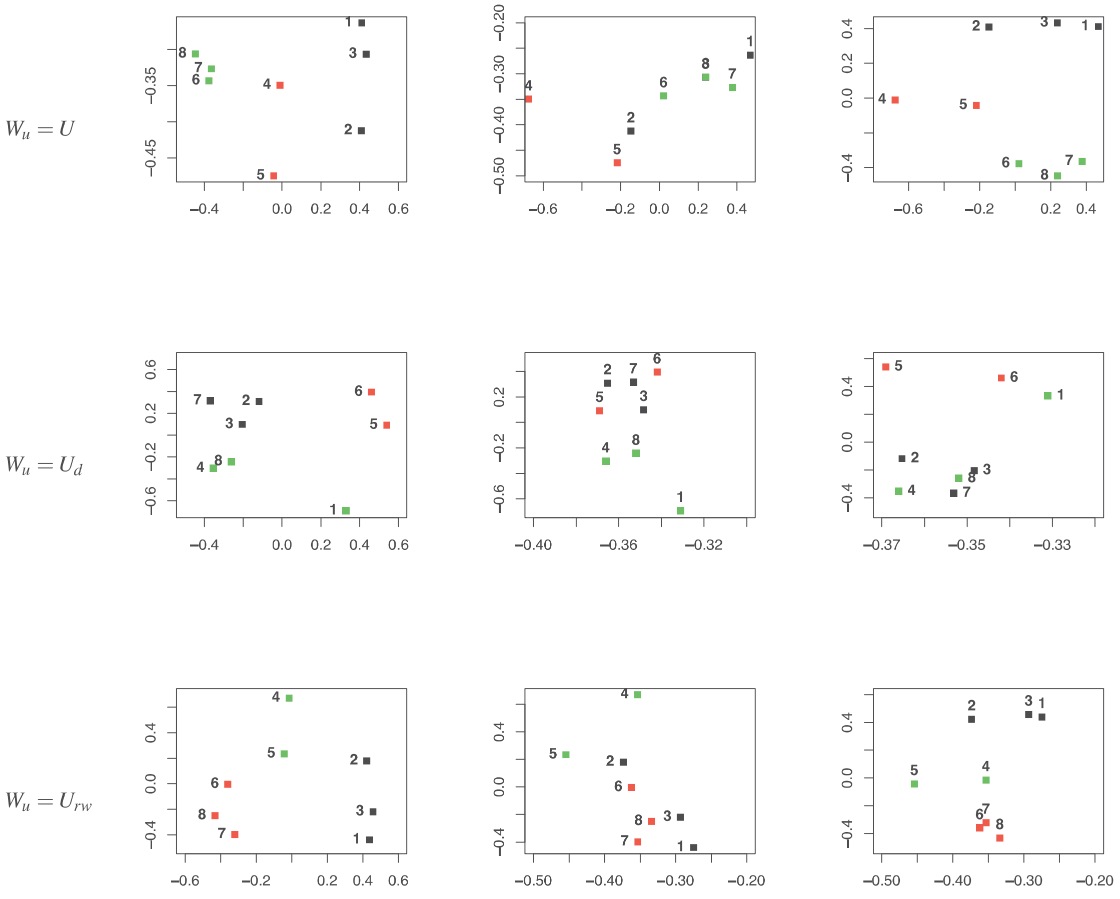
$$W = \begin{pmatrix} 0 & 10 & 5 & 0 & 0 & 3 & 5 & 0 \\ 7 & 0 & 8 & 9.5 & 8 & 4 & 0 & 0 \\ 9 & 6 & 0 & 0 & 8 & 0 & 0 & 4 \\ 0 & 5 & 0 & 0 & 10 & 6 & 0 & 0 \\ 0 & 6 & 7 & 10 & 0 & 0 & 9 & 5 \\ 0 & 6 & 0 & 8 & 0 & 0 & 8 & 10 \\ 4 & 0 & 0 & 0 & 8 & 7 & 0 & 6 \\ 0 & 0 & 0 & 0 & 10 & 7 & 5 & 0 \end{pmatrix}$$

Fig. 1: *Artificial data*. a) Representation of the graph $\mathscr{G}$. b) Heat map of the asymmetric matrix W. c) Asymmetric weighted matrix W associated to graph $\mathscr{G}$.

In Figure 2, the geometric features of the spectral clustering algorithm (Algorithm 1) applied on the three symmetrizations described in Section 3 have been represented. The correct clustering results are attained from the U and U_{rw} symmetrizations; conversely, the spectral clustering applied on the U_d symmetrization does not provide the correct classification, because the accuracy is equal to 0.5 and the confusion matrix is

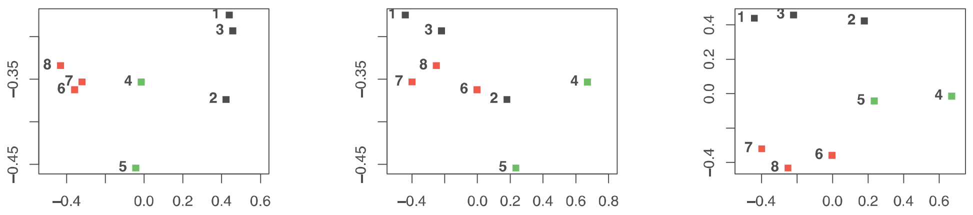
$$\begin{bmatrix}2&0&1\\1&1&1\\0&1&1\end{bmatrix}.$$

The asymmetric spectral clustering described in Algorithm 2 has been also carried out, the geometric features of the algorithm are shown in Figure 3, and the correct clustering is obtained.



a) *b*) *c*)

Fig. 2: *Spectral clustering after symmetrizations*. Embedded data associated to L_{sym} according to: *a*) the first two eigenvectors; *c*) the first and third eigenvector; *d*) the last two eigenvectors (different colors denote different clusters).



a) *b*) *c*)

Fig. 3: *Asymmetric spectral clustering (Algorithm 2)*. Embedded data associated to Θ according to: *a*) the first two eigenvectors; *c*) the first and third eigenvector; *d*) the last two eigenvectors (different colors denote different clusters).

To conclude, the asymmetric spectral clustering described in Algorithm 2 provides the correct classification, while the spectral clustering applied after the symmetrizations provides the correct results only when the U_{rw} and U symmetrizations are used. Finally, the Laplacian embedding structure provided by the U_d symmetrization, shown in Figure 2, is different from the others, because U_d also takes into account and incorporates the information from the in-degrees of the nodes of the directed graph $\mathscr{G}$. Therefore, this lays the groundwork for further studies on the information extracted from the different symmetrizations by carrying out studies on both artificial and real data.

References

1. Gleich, D.: Hierarchical Directed Spectral Graph Partitioning. MS&E 337 - Information Networks. (2006)
2. Lejay, A.: Asymmetric Spectral clustering. Technical Report. Inria Nancy - Grand Est. url: https://hal.inria.fr/hal-02372570. (2019)
3. Meila, M. and Pentney, W. : Clustering by weighted cuts in directed graphs. doi: 10.1137/1.9781611972771.13. (2007)
4. Meila, M., Shi, J. : A random walks view of spectral segmentation. In: 8th International Workshop on Artificial Intelligence and Statistics (AISTATS). (2001)
5. Ng, A., Jordan, M., and Weiss, Y.: On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, **14**. (2002).
6. Satuluri, V., Parthasarathy, S.: Symmetrizations for Clustering Directed Graphs. In: EDBT/ICDT'11: Proceedings of the 14th International Conference on Extending Database Technology, pp 343–354. (2011)
7. Shi, J.,Malik, J.: Normalized cuts and image segmentation. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 888–905. (2000)
8. Zhou, D., Huang, J., and Schölkopf, B.: Learning from labeled and unlabeled data on a directed graph. In: ICML '05, pp.1036-1043. (2005)