

PROGRAM and BOOK of ABSTRACTS

JOCLAD2021

9 - 11 DECEMBER

COVILHÃ, PORTUGAL



XXVIII MEETING OF THE PORTUGUESE ASSOCIATION FOR CLASSIFICATION AND DATA ANALYSIS
XXVIII JORNADAS DE CLASSIFICAÇÃO E ANÁLISE DE DADOS



Program and Book of Abstracts

XXVIII Meeting of the Portuguese Association for Classification and Data Analysis (CLAD)

9–11 December 2021

Covilhã, Portugal

www.joclad.ipt.pt/joclad2021/

Sponsors

Associação Portuguesa de Classificação e Análise de Dados
Universidade da Beira Interior
Instituto Nacional de Estatística
Banco de Portugal
Câmara Municipal da Covilhã
Quinta dos Termos
Queijos Braz
Natura IMB Hotels

Program and Book of Abstracts

XXVIII Meeting of the Portuguese Association for Classification
and Data Analysis (JOCLAD 2021)

Editors: José G. Dias, João Cordeiro, M. Paula Brito, Célia Nunes, Sebastião Pais,
Conceição Rocha, M. Eugénia Ferrão

Publisher: CLAD

ISBN 978-989-98955-8-4

Preface

Welcome to JOCLAD 2021! The JOCLAD 2021 – Meeting of the Portuguese Association for Classification and Data Analysis aims to bring together researchers and professionals in the field of Data Science. This is already the twenty-eighth meeting of the CLAD. After many meetings throughout Portugal - 2017 in Porto, 2018 in Almada, 2019 in Viseu, 2020 in Lisbon; this year, from 9 to 11 December, JOCLAD 2021 takes place at the University of Beira Interior (UBI), which co-organizes the event. UBI is situated in the beautiful and historical city of Covilhã, near Serra da Estrela, the tallest Portuguese mountain ($\approx 2000\text{m}$). The city roots are anchored to the old traditions of wool production, going back at least to the XII century. Covilhã has just been classified as “Creative City of Design” by UNESCO. From the local organizers, receive a warm welcome to our city and university, wishing you a pleasant discovery of the wonders of this region. The JOCLAD 2021 program includes a mini-course, on December 9, on the topic of “Co-clustering”, held by Mohamed Nadif (Université Paris Descartes, Paris), three Keynote Sessions - “Efficient Search for Good Neural Data Processors” (Luís Alexandre, Universidade da Beira Interior, Covilhã), “Clustering with Data Embedding” (Mohamed Nadif, Université Paris Descartes, Paris), and “A Regression Perspective of Binary and Multi-Class Support Vector Machines” (Patrick J.F. Groenen, Erasmus School of Economics, Rotterdam). Their talks present a representative cross-section of research in Data Science. In addition, the program includes the Fernando Nicolau Award, five thematic sessions - CLAD 2021 Scholarship, CLAD-SPE, INE, Banco de Portugal, and CLAD Corporate -, 35 oral communications, and 17 posters. A Thematic Session is aimed at students who have received a CLAD scholarship 2021, whose members of the evaluation committee were José G. Dias (Chair), Irene Oliveira, and Victor Lobo. We also thank the organizers of the other Thematic Sessions: Pedro Campos (INE - Instituto Nacional de Estatística), Luís Teles Dias (Banco de Portugal), Conceição Amado and Ricardo São João (CLAD-SPE), and Pedro Campos (CLAD Corporate).

Additionally, this volume contains all the abstracts of talks and posters presented at regular oral and poster sessions. Each abstract published in this volume has been double-blind reviewed by at least one anonymous member of the scientific committee. We thank all authors who submitted an abstract to our meeting and the reviewers, who supported the editorial work with their quick and constructive responses. These procedures contribute to strengthening the overall quality of the JOCLAD 2021 program. Also, we thank all the chairs of these sessions.

Our deep gratitude goes to the members of the CLAD Board, especially Conceição Rocha and Pedro Campos, who have so voluntarily devoted their time to supporting

JOCLAD 2021. Last but not least, we are delighted to thank the sponsors for their support. Our institutional sponsors deserve a special mention: Instituto Nacional de Estatística (INE) and Banco de Portugal.

Finally, thank you all for your support in helping us keep our annual meeting going! With your high-quality work, CLAD will continue its tradition of excellence in the advancement of Data Science!

Lisboa, December 2021

Chair of the Scientific Program

José G. Dias

Conference Chair

João Cordeiro

President of CLAD

Maria Paula Brito

Organization

President of the CLAD

Maria Paula Brito

Chair of the JOCLAD 2021

João Cordeiro (Universidade da Beira Interior)

Local Organizing Committee

Célia Nunes (Universidade da Beira Interior, CMA-UBI)

João Cordeiro (Universidade da Beira Interior, INESC TEC)

Maria Eugénia Ferrão (Universidade da Beira Interior, CEMAPRE)

Sebastião Pais (Universidade da Beira Interior, NOVA LINCS)

Conceição Rocha (INESC TEC - Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência)

Chair of the Scientific Program Committee

José G. Dias (ISCTE - Instituto Universitário de Lisboa)

Scientific Program Committee

A. Manuela Gonçalves (Universidade do Minho)

Adelaide Figueiredo (Universidade do Porto)

Adelaide Freitas (Universidade de Aveiro)

Ana Lorga da Silva (Universidade Lusófona)

Ana Sousa Ferreira (Universidade de Lisboa)

Anabela Afonso (Universidade de Évora)

Anabela Marques (Instituto Politécnico de Setúbal)

Carla Henriques (Instituto Politécnico de Viseu)

Carlos Ferreira (Universidade de Aveiro)
Carlos Soares (Universidade do Porto)
Catarina Marques (ISCTE - Instituto Universitário de Lisboa)
Conceição Amado (Universidade de Lisboa)
Conceição Rocha (INESC-TEC)
Fátima Salgueiro (ISCTE - Instituto Universitário de Lisboa)
Fernanda Sousa (Universidade do Porto)
Helena Bacelar-Nicolau (Universidade de Lisboa)
Irene Oliveira (Universidade de Trás-os-Montes e Alto Douro)
Isabel Silva Magalhães (Universidade do Porto)
Luís Miguel Grilo (Instituto Politécnico de Tomar)
Manuela Neves (Universidade de Lisboa)
Margarida Cardoso (ISCTE - Instituto Universitário de Lisboa)
Maria Filomena Teodoro (Escola Naval-Marinha Portuguesa)
Paula Vicente (ISCTE - Instituto Universitário de Lisboa)
Paulo Infante (Universidade de Évora)
Pedro Campos (Universidade do Porto)
Pedro Duarte Silva (Universidade Católica Portuguesa)
Rosário Oliveira (Universidade de Lisboa)
Sónia Dias (Instituto Politécnico de Viana do Castelo)
Susana Faria (Universidade do Minho)
Victor Lobo (Universidade Nova de Lisboa)

Contents

Program Overview	xi
Program	xv
Abstracts	1
Mini-Course	3
Co-clustering	5
Keynote Sessions	7
Efficient search for good neural data processors	9
Clustering with data embedding	11
A Regression Perspective of Binary and Multi-Class Support Vector Machines	13
Thematic Session I: CLAD 2021 Scholarship	15
Bias in Citizen Science: an application to the BioDiversity4All project	17
The Shape of Collaboration in Biodiversity Monitoring	19
Modeling the cost associated with body injuries in workplace accidents	21
The use of the EM and the CEM algorithms in variable selection for mixtures of linear models with random effects	23
Short-Term Forecast Models for Meteorological Variables	25
Spatio-temporal variability of distribution and abundance of sardine in Por- tuguese continental coast: environmental effects	27
Thematic Session II: Statistics Portugal	29
Second Child: An Uncertain Transition	31
Covid-19 pandemic: an unprecedented shock to consumer confidence	33
Data analysis during Covid time: the e-invoice case	35
Inter-organizational networks of EuroGroups Register - a Supervised Clustering Algorithm for Network Data	37
Thematic Session III: Banco de Portugal	39
Using Isolation Forest in the quality control of the securities database of Banco de Portugal	41
Machine Learning models applied to ITENF data quality control	43
An application of cluster analysis to the interest rates reported to the Portuguese Central Credit Register	45

Thematic Session IV: CLAD–SPE	47
Prediction Models in Medicine	49
Is age at menopause decreasing? The consequences of not completing the generational cohort	51
Analysis of cutoff point estimation for determining seropositivity in the context of SARS-CoV-2 infections	53
Thematic Session V: CLAD Corporate	55
How to Measure the Outdoor Advertising Audience and the Actual Mobility in Portugal?	57
Contributed Sessions	59
Modelling interval-valued data: a clusterwise regression approach	61
Multivariate Parametric Analysis of Distributional Data	63
Qualitative-quantitative synergy	65
Adapting the sampling design of research surveys to improve the biomass estimation of non-target species - the case study of <i>Raja clavata</i>	67
Estimating echolocation clicks rates in narwhals (<i>Monodon monoceros</i>)	69
A new selection index to perform polyclonal selection in ancient grapevine varieties	71
Waste management on subsurface ships: a case study	73
Sparse Divisive Feature Clustering	75
Clustering domestic water consumption profiles: a Portuguese case study	77
Clustering times series of electricity consumption	79
Clustering of EFCs in European Economies	81
Visually supported curriculum development	83
How to analyze online behavior as a source for political information in the Portuguese 2019 European Parliament election?	85
The role of the pre-university life path in the performance of students who access higher education: the case study of the Master’s Degree in Civil Engineering at FEUP	87
Data Analysis to Advance Immersive Technologies: examples from Augmented Reality-based assembly procedures	89
Poster Sessions	91
Statistical Modeling of User Influx to the HESE’s Emergency Service	93
Quality control techniques for tidal data in near real time	95
Time series synchronization: An application to COVID-19 data	97
A Case Study of Pavement Texture Performance using Linear Mixed Models	99
Comparing Time Series Models for Forecasting Meteorological Data	101
Determinants for the existence of victims in road accidents in the district of Setúbal	103
Likelihood function approximation through the delta method in mixed SDE models	105
The impact of e-government on sustainable development: a logit model	107

A hurdle-gamma regression model for the average number of undeveloped pine nuts per cone	109
How the indicators for EU countries in the period 2010-2019 are approaching targets of Europe 2020 agenda?	111
A state space framework for daily temperature forecasting	113
A sociological portrait of the Portuguese, based on their religiosity and values: A cross-time comparison with Europe	115
A longitudinal analysis of the severity of road accidents in the district of Setúbal between 2016 and 2019	117
Data, Technology and Journalism	119
Field test validation for predicting VO ₂ max in the Portuguese Military Academy	121
Mechanical Behavior of Skin: an ANOVA Approach	123
The Relationship of the Human Capital Index with the Level of Education and the Adult Survival Rate	125
Author Index	127

Program Overview



Thursday, 9 December

-
- 9:30 Registration – Hall of Anfiteatros Verde and Azul, Faculdade de Ciências da Saúde
- 10:30 **Mini-course**
- 12:30 Lunch Time
- 13:30 **Mini-course** (cont.)
- 14:30 **Opening Session**
- 15:00 **Keynote Session I**
- 16:00 Coffee Break
- 16:20 **Fernando Nicolau Award**
- 16:50 **Contributed Session I - Symbolic and Complex data**
- 18:30 Visit to the *Real Fábrica Veiga*
-

Friday, 10 December

-
- 8:30 Registration – Hall of Anfiteatros Verde and Azul, Faculdade de Ciências da Saúde
- 9:00 **Contributed Session II - Data Science applications in Life Sciences**
- 10:20 **Poster Session I**
- 10:40 Coffee Break
- 11:00 **Thematic Session I - Scholarship CLAD 2021**
- 13:00 Lunch Time
- 14:00 **Keynote Session II**
- 15:00 **Thematic Session II - Statistics Portugal**
- 16:20 Coffee Break
- 16:40 **Thematic Session III - Banco de Portugal**
- 18:00 General Assembly of CLAD
- 20:00 Meeting Dinner – Hotel Puralã
-

Saturday, 11 December

-
- 8:30 Registration – Hall of Anfiteatros Verde and Azul, Faculdade de Ciências da Saúde
- 9:00 **Contributed Session III - Clustering**
- 10:20 **Poster Session II**
- 10:40 Coffee Break
- 11:00 **Contributed Session IV - Data Science applications in Social Sciences**
- 12:20 **Thematic Session IV - CLAD - SPE**
- 13:20 Lunch Time
- 14:20 **Thematic Session V - CLAD Corporate**
- 15:00 **Keynote Session III**
- 16:00 **Closing Session**
- 16:15 Coffee Break
-

Program



Thursday, 9 December

9:30 Registration – Hall of Anfiteatros Verde and Azul, Faculdade de Ciências da Saúde

10:30 **Mini-course**
Co-clustering
Mohamed Nadif, p. 5

12:30 **Lunch Time**

13:30 **Mini-course** (cont.)

14:30 **Opening Session**

15:00 **Keynote Session I**
Efficient search for good neural data processors
Luís A. Alexandre, p. 9

Chair: João Cordeiro

16:00 **Coffee Break**

16:20 **Fernando Nicolau Award**

Chair: Helena Bacelar-Nicolau

16:50 **Contributed Session I - Symbolic and Complex data**

Chair: Margarida Cardoso

16:50 **Modelling interval-valued data: a clusterwise regression approach**
Sónia Dias, Paula Brito, and Nikhil Suresh, p. 61

17:10 **Multivariate Parametric Analysis of Distributional Data**
Paula Brito and A. Pedro Duarte Silva, p. 63

17:30 **Qualitative-quantitative synergy**
Ana Lorga da Silva and Artur Parreira, p. 65

18:30 **Visit to the *Real Fábrica Veiga***

Friday, 10 December

8:30	Registration – Hall of Anfiteatros Verde and Azul, Faculdade de Ciências da Saúde
9:00	Contributed Session II - Data Science applications in Life Sciences Chair: A. Manuela Gonçalves
9:00	Adapting the sampling design of research surveys to improve the biomass estimation of non-target species - the case study of <i>Raja clavata</i> Daniela Silva, <u>Raquel Menezes</u> , Ivone Figueiredo, Bárbara Serra-Pereira, and Manuela Azevedo, p. 67
9:20	Estimating echolocation clicks rates in narwhals (<i>Monodon monoceros</i>) <u>Diana Marques</u> , Tiago Marques, Susanna B. Blackwell, Mads Peter Heide-Jørgensen, and Carolina Marques, p. 69
9:40	A new selection index to perform polyclonal selection in ancient grapevine varieties <u>Sónia Surgy</u> , Jorge Cadima, and Elsa Gonçalves, p. 71
10:00	Waste management on subsurface ships: a case study <u>M. Filomena Teodoro</u> , Suzana Lampreia, and Tomás Mendes, p. 73
10:20	Poster Session I Chair: Sebastião Pais
	– Statistical Modeling of User Influx to the HESE’s Emergency Service , <u>Loide Ascenso</u> , Hugo Quintino, Paulo Infante, and Gonçalo Jacinto, p. 93
	– Quality control techniques for tidal data in near real time , <u>Dora Carinhas</u> , Margarida Alves, Paulo Infante, and António Martinho, p. 95
	– Time series synchronization: An application to COVID-19 data , José G. Dias, p. 97
	– A Case Study of Pavement Texture Performance using Linear Mixed Models , Adriana Santos, <u>Susana Faria</u> , and Elisabete Freitas, p. 99
	– Comparing Time Series Models for Forecasting Meteorological Data , A. Manuela Gonçalves, Marco Costa, and Cláudia Costa, p. 101
	– Determinants for the existence of victims in road accidents in the district of Setúbal , <u>Paulo Infante</u> , Gonçalo Jacinto, Anabela Afonso, Rodrigo Cesar, Pedro Nogueira, Marcelo Silva, Vitor Nogueira, José Saias, Paulo Quaresma, Patrícia Gois, and Paulo Rebelo Manuel, p. 103
	– Likelihood function approximation through the delta method in mixed SDE models , Nelson T. Jamba, <u>Patrícia A. Filipe</u> , Gonçalo Jacinto, and Carlos A. Braumann, p. 105
	– The impact of e-government on sustainable development: a logit model , <u>Cristina Lopes</u> and Conceição Castro, p. 107
10:40	Coffee Break

11:00 **Thematic Session I - Scholarship CLAD 2021**

Chair: José G. Dias

11:00 **Bias in Citizen Science: an application to the BioDiversity4All project**

João Alves, Guilherme Correia, Ana Almeida Matos, Ana Subtil, Francisco C. Santos, and M. Rosário Oliveira, p. 17

11:20 **The Shape of Collaboration in Biodiversity Monitoring**

Guilherme Correia, João Alves, Ana Almeida Matos, Ana Subtil, M. Rosário Oliveira, Patrícia Tiago, and Francisco C. Santos, p. 19

11:40 **Modeling the cost associated with body injuries in workplace accidents**

Ana Moreira, Filipe Gonçalves, Luís Maranhão, and Susana Faria, p. 21

12:00 **The use of the EM and the CEM algorithms in variable selection for mixtures of linear models with random effects**

Luísa Novais and Susana Faria, p. 23

12:20 **Short-Term Forecast Models for Meteorological Variables**

F. Catarina Pereira, p. 25

12:40 **Spatio-temporal variability of distribution and abundance of sardine in Portuguese continental coast: environmental effects**

Daniela Silva, Raquel Menezes, Ana Moreno, Ana Teles-Machado, and Susana Garrido, p. 27

13:00 **Lunch Time**

14:00 **Keynote Session II**

Clustering with data embedding

Mohamed Nadif, p. 11

Chair: Paula Brito

15:00 **Thematic Session II - Statistics Portugal**
Challenges in Official Statistics X

Chair: Pedro Campos

15:00 **Second Child: An Uncertain Transition**

Rita Brazão Freitas, Andreia Maciel, and Maria Filomena Mendes, p. 31

15:20 **Covid-19 pandemic: an unprecedented shock to consumer confidence**

Ana Raquel Marques, p. 33

15:40 **Data analysis during Covid time: the e-invoice case**

João Poças and Sofia Rodrigues, p. 35

16:00 **Inter-organizational networks of EuroGroups Register - a Supervised Clustering Algorithm for Network Data**

Bárbara Santos and Pedro Campos, p. 37

16:20 **Coffee Break**

16:40 **Thematic Session III - Banco de Portugal**
New tools for outliers detection in big databases

Chair: Luís Teles Dias

16:40 **Using Isolation Forest in the quality control of the securities database of Banco de Portugal**

André Fernandes and Rafael Figueira, p. 41

17:00 **Machine Learning models applied to ITENF data quality control**

C. Ferreira, F. Fonseca, L. Pinto, J. A. Neves, N. Azevedo, and V. Lopes, p. 43

17:20 **An application of cluster analysis to the interest rates reported to the Portuguese Central Credit Register**

André Costa, Francisco Fonseca, and Susana Maurício, p. 45

18:00 General Assembly of CLAD

20:00 **Meeting Dinner** – Hotel Puralã

Saturday, 11 December

8:30 Registration – Hall of Anfiteatros Verde and Azul, Faculdade de Ciências da Saúde

9:00 **Contributed Session III - Clustering**

Chair: Fernanda Sousa

9:00 **Sparse Divisive Feature Clustering**

Ndèye Niang, Mory Ouattara, and Gilbert Saporta, p. 75

9:20 **Clustering domestic water consumption profiles: a Portuguese case study**

Elisa Araújo, Flora Ferreira, Duarte Silva, Estela Bicho, and Wolfram Erlhagen, p. 77

9:40 **Clustering times series of electricity consumption**

Margarida G. M. S. Cardoso, Ana Martins, and João Lagarto, p. 79

10:00 **Clustering of EFCs in European Economies**

Eliana Costa e Silva, Aldina Correia, and Ana Borges, p. 81

10:20 **Poster Session II**

Chair: Célia Nunes

-
- **A hurdle-gamma regression model for the average number of undeveloped pine nuts per cone**, Anabela Afonso, Dulce G. Pereira, and Ana Cristina Gonçalves, p. 109
 - **How the indicators for EU countries in the period 2010-2019 are approaching targets of Europe 2020 agenda?**, Adelaide Figueiredo and Fernanda Otilia Figueiredo, p. 111
 - **A state space framework for daily temperature forecasting**, A. Manuela Gonçalves, F. Catarina Pereira, and Marco Costa, p. 113
 - **A sociological portrait of the Portuguese, based on their religiosity and values: A cross-time comparison with Europe**, Maria Paula Lousão, Cláudia Vasconcelos Silvestre, and José Luís Casanova, p. 115
 - **A longitudinal analysis of the severity of road accidents in the district of Setúbal between 2016 and 2019**, Paulo Infante, Anabela Afonso, Gonçalo Jacinto, Leonor Rego, Pedro Nogueira, Marcelo Silva, Vitor Nogueira, José Saias, Paulo Quaresma, Patrícia Gois, and Paulo Rebelo Manuel, p. 117
 - **Data, Technology and Journalism**, Cláudia Silvestre and Pedro Frazão, p. 119
 - **Field test validation for predicting VO₂max in the Portuguese Military Academy**, Rui Lucena, Lucas Nogueira, Nuno Almeida, Cristiano Almeida, and Paula Simões, p. 121
 - **Mechanical Behavior of Skin: an ANOVA Approach**, M. Filomena Teodoro and Teresa Oliveira, p. 123
 - **The Relationship of the Human Capital Index with the Level of Education and the Adult Survival Rate**, Alexandra Marques, Ana Pinheiro, Maria Carolina Matos, Cristina Torres, Cristina Lopes, and Isabel Vieira, p. 125

10:40 **Coffee Break**

11:00	Contributed Session IV - Data Science applications in Social Sciences	
		Chair: Adelaide Figueiredo
11:00	Visually supported curriculum development	
	<u>Rogério Duarte</u> , <u>Ângela Lacerda Nobre</u> , <u>Fernando Pimentel</u> , and <u>Marc Jacquinet</u> , p. 83	
11:20	How to analyze online behavior as a source for political information in the Portuguese 2019 European Parliament election?	
	<u>Cláudia Silvestre</u> , <u>Rodrigo Pinheiro</u> , and <u>Filipe Montargil</u> , p. 85	
11:40	The role of the pre-university life path in the performance of students who access higher education: the case study of the Master's Degree in Civil Engineering at FEUP	
	<u>Fernanda Campos de Sousa</u> and <u>Isabel Martins Ribeiro</u> , p. 87	
12:00	Data Analysis to Advance Immersive Technologies: examples from Augmented Reality-based assembly procedures	
	<u>Carlos Ferreira</u> , <u>Bernardo Marques</u> , <u>João Alves</u> , <u>Paulo Dias</u> , and <u>Beatriz Sousa Santos</u> , p. 89	
12:20	Thematic Session IV - CLAD - SPE	
		Chairs: <u>Conceição Amado</u> and <u>Ricardo S. João</u>
12:20	Prediction Models in Medicine	
	<u>Ana Luisa Papoila</u> , p. 49	
12:40	Is age at menopause decreasing? The consequences of not completing the generational cohort	
	<u>Rui Martins</u> , <u>Bruno de Sousa</u> , <u>Thomas Kneib</u> , <u>Nadja Klein</u> , <u>Maike Hohberg</u> , <u>Elisa Duarte</u> , and <u>Vítor Rodrigues</u> , p. 51	
13:00	Analysis of cutoff point estimation for determining seropositivity in the context of SARS-CoV-2 infections	
	<u>Tiago Dias Domingues</u> , <u>Helena Mouriño</u> , and <u>Nuno Sepúlveda</u> , p. 53	
<hr/> <hr/>		
13:20	Lunch Time	
<hr/> <hr/>		
14:20	Thematic Session V - CLAD Corporate	
		Chair: <u>Carlos Ferreira</u>
	How to Measure the Outdoor Advertising Audience and the Actual Mobility in Portugal?	
	<u>Paulo Caldeira</u> and <u>João Pequito</u> , p. 57	
15:00	Keynote Session III	
	A Regression Perspective of Binary and Multi-Class Support Vector Machines	
	<u>Patrick J.F. Groenen</u> , p. 13	
		Chair: <u>A. Pedro Duarte Silva</u>
16:00	Closing Session	
16:15	Coffee Break	

Abstracts



11 December, 11:00 - 11:20

Visually supported curriculum development

Rogério Duarte¹, Ângela Lacerda Nobre², Fernando Pimentel¹, Marc Jacquinet³,

¹ ESTSetúbal, CINEA, Instituto Politécnico de Setúbal

² ESCE, Instituto Politécnico de Setúbal

³ DCSG, CEMRI, Universidade Aberta

Curriculum development processes are of paramount importance for the preparation of future professionals. An impediment to these processes lays in stakeholders' communication difficulties. This study considers the combination of information and data science techniques namely, the use of classification, natural language processing and network graphs, to make curriculum representations self-explanatory, bridging communication gaps and helping stakeholders to articulate their expert knowledge.

Keywords: Network graphs, Natural language processing, Classification, Curriculum analytics, Education

Accreditation bodies require “scientific” curriculum development processes reliant on outcomes based education and on constructive alignment principles. To implement these principles *Learning Outcomes Mappings* use courses as building blocks and the visual representation—typically a chart, table or map—depicting vertical (from year to year) and horizontal (within a year) relations between courses conveys a sense to the study program structure. From the analysis of each course Learning Outcomes (LO) it is possible to infer further connections between courses and gain access to the “mechanics” behind a program plan. However, course LO presume tacit understanding of concepts specific to disciplinary and scientific sub-areas and this renders LO-statements seldom clear and unequivocal.

This study reports the combination of information and data science techniques including classification of LO-statements, natural language processing, the use of network graphs, to communicate—visually and quantitatively—information included in the study program that is typically difficult to access and to articulate.

Figure 1, obtained as described in detail in [1], is the end product of a data visualization method that considers the following steps: (1) classification of course learning objective statements into *broader terms*; (2) use of Natural Language Processing (NLP) to convert *broader terms* into quantitative frequencies of key program concepts; (3) visualization of the links between courses based on common key concepts.

In Figure 1, MATH, PHY, STAT, etc. are acronyms for courses included in the study program. The relative position of the courses as well as the number of connecting lines provides both

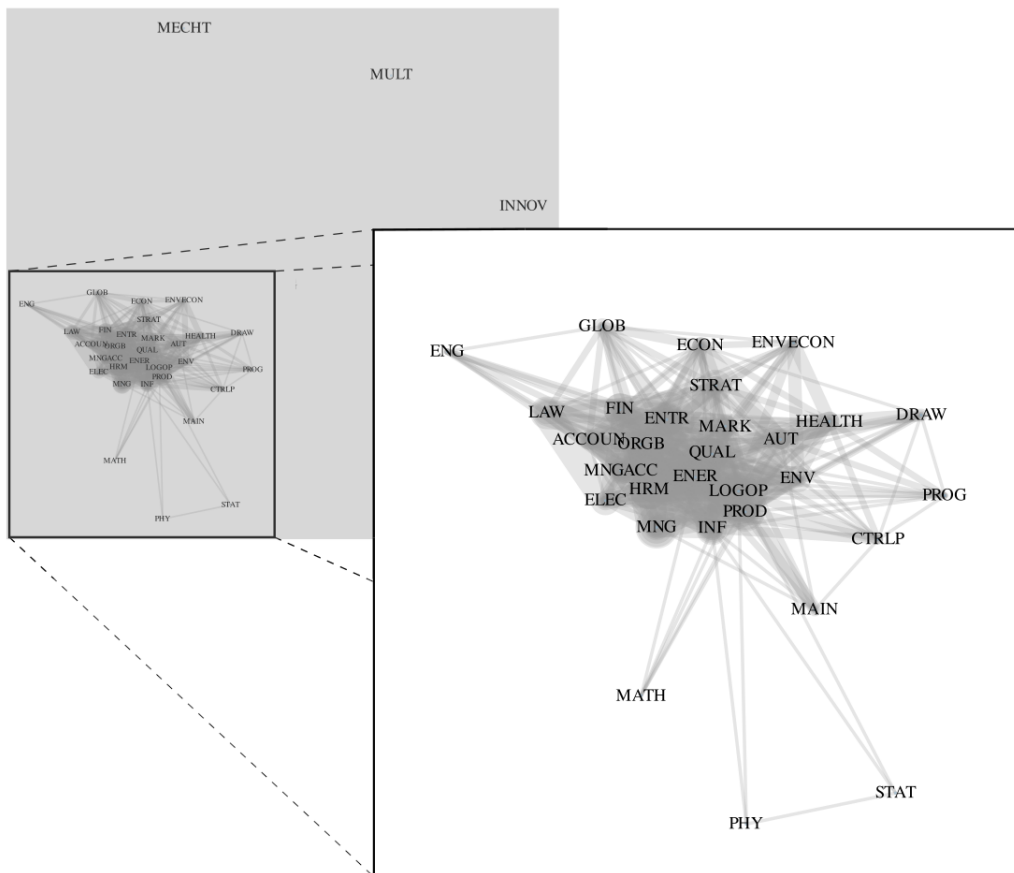


Figure 1: Visualization of links between courses of a bachelor degree in Technology and Industrial Management [1]. The forward (white) plane presents an enlarged detail with 30 of the 33 courses present in the background (gray) plane.

visual and quantitative measure of courses connectivity. Figure 1 highlights the lack of connections to three courses depicted in the background (gray) plane, **MECHT**, **MULT**, **INNOV**. It becomes obvious (inspecting the forward white plane) that fundamental courses such as mathematics, physics, statistics (**MATH**, **PHY**, **STAT**) are detached from the figure’s mass center, where the majority of the courses lay.

References

[1] Rogério Duarte, Ângela Lacerda Nobre, Fernando Pimentel, and Marc Jacquinet. Broader terms curriculum mapping: Using natural language processing and visual-supported communication to create representative program planning experiences. *CoRR*, abs/2102.04811 (<https://arxiv.org/abs/2102.04811>), 2021.

Author Index

- A. Manuela Gonçalves, 101, 113
A. Pedro Duarte Silva, 63
Adelaide Figueiredo, 111
Adriana Santos, 99
Aldina Correia, 81
Alexandra Marques, 125
Ana Almeida Matos, 17, 19
Ana Borges, 81
Ana Cristina Gonçalves, 109
Ana Lorga da Silva, 65
Ana Luisa Papoila, 49
Ana Martins, 79
Ana Moreira, 21
Ana Moreno, 27
Ana Pinheiro, 125
Ana Raquel Marques, 33
Ana Subtil, 17, 19
Ana Teles-Machado, 27
Anabela Afonso, 103, 109, 117
André Costa, 45
André Fernandes, 41
Andreia Maciel, 31
Angela Lacerda Nobre, 83
António Martinho, 95
Artur Parreira, 65
- Bárbara Serra-Pereira, 67
Bárbara Santos, 37
Beatriz Sousa Santos, 89
Bernardo Marques, 89
Bruno de Sousa, 51
- C. Ferreira, 43
Carlos A. Braumann, 105
Carlos Ferreira, 89
Carolina Marques, 69
Cláudia Costa, 101
Cláudia Silvestre, 85, 115, 119
Conceição Castro, 107
- Cristiano Almeida, 121
Cristina Lopes, 107, 125
Cristina Torres, 125
- Daniela Silva, 27, 67
Diana Marques, 69
Dora Carinhas, 95
Duarte Silva, 77
Dulce G. Pereira, 109
- Eliana Costa e Silva, 81
Elisa Araújo, 77
Elisa Duarte, 51
Elisabete Freitas, 99
Elsa Gonçalves, 71
Estela Bicho, 77
- F. Catarina Pereira, 25, 113
F. Fonseca, 43
Fernanda Campos de Sousa, 87
Fernanda Otilia Figueiredo, 111
Fernando Pimentel, 83
Filipe Gonçalves, 21
Filipe Montargil, 85
Flora Ferreira, 77
Francisco C. Santos, 17, 19
Francisco Fonseca, 45
- Gertjan van den Burg, 13
Gilbert Saporta, 75
Gonçalo Jacinto, 93, 103, 105, 117
Guilherme Correia, 17, 19
- Helena Mouriño, 53
Hugo Quintino, 93
- Isabel Martins Ribeiro, 87
Isabel Vieira, 125
Ivone Figueiredo, 67
- J. A. Neves, 43

- João Alves, 17, 19, 89
João Lagarto, 79
João Pequito, 57
João Poças, 35
Jorge Cadima, 71
José G. Dias, 97
José Luís Casanova, 115
José Saias, 103, 117
- L. Pinto, 43
Leonor Rego, 117
Loide Ascenso, 93
Luís A. Alexandre, 9
Luís Maranhão, 21
Luísa Novais, 23
Lucas Nogueira, 121
- M. Filomena Teodoro, 73, 123
M. Rosário Oliveira, 17, 19
Mads Peter Heide-Jørgensen, 69
Maike Hohberg, 51
Manuela Azevedo, 67
Marc Jacquinet, 83
Marcelo Silva, 103, 117
Marco Costa, 101, 113
Margarida Alves, 95
Margarida G. M. S. Cardoso, 79
Maria Carolina Matos, 125
Maria Filomena Mendes, 31
Maria Paula Lousão, 115
Mohamed Nadif, 5, 11
Mory Ouattara, 75
- N. Azevedo, 43
Nadja Klein, 51
Ndèye Niang, 75
Nelson T. Jamba, 105
Nikhil Suresh, 61
Nuno Almeida, 121
Nuno Sepúlveda, 53
- Patrícia A. Filipe, 105
Patrícia Gois, 103, 117
Patrícia Tiago, 19
Patrick J.F. Groenen, 13
Paula Brito, 61, 63
Paula Simões, 121
- Paulo Caldeira, 57
Paulo Dias, 89
Paulo Infante, 93, 95, 103, 117
Paulo Quaresma, 103, 117
Paulo Rebelo Manuel, 103, 117
Pedro Campos, 37
Pedro Frazão, 119
Pedro Nogueira, 103, 117
- Rafael Figueira, 41
Raquel Menezes, 27, 67
Rita Brazão Freitas, 31
Rodrigo Cesar, 103
Rodrigo Pinheiro, 85
Rogério Duarte, 83
Rui Lucena, 121
Rui Martins, 51
- Sónia Dias, 61
Sónia Surgy, 71
Sofia Rodrigues, 35
Susana Faria, 21, 23, 99
Susana Garrido, 27
Susana Maurício, 45
Susanna B. Blackwell, 69
Suzana Lampreia, 73
- Teresa Oliveira, 123
Thomas Kneib, 51
Tiago Dias Domingues, 53
Tiago Marques, 69
Tomás Mendes, 73
- V. Lopes, 43
Vítor Rodrigues, 51
Vitor Nogueira, 103, 117
- Wolfram Erlhagen, 77

SPONSORS

