

Francesco Durazzi

francesco.durazzi2@unibo.it

Born in Fossombrone (PU), Italy, 15/6/1994 (29 years).

Address: viale Zagabria 7, Bologna, BO, 40127.

Stats (Scholar). *h*-index: 4, publications: 6, citations: 301

WORK

Fixed-term Researcher (type A), Department of Physics and Astronomy, University of Bologna

11/2023 – Now

Postdoctoral researcher, Department of Physics and Astronomy, University of Bologna

02/2023 – 11/2023

Visiting PhD student, Digital Epidemiology Lab, EPFL (CH)

09/2021 – 12/2021

Dynamics of social media behavior before and after SARS-CoV-2 infection[1]. Keywords: NLP, Deep Learning, Twitter data, time series, statistical testing.

PhD student, Department of Physics and Astronomy, University of Bologna

11/2019 – 01/2023

Thesis: “Data surveillance for infectious diseases: models, complex networks and machine learning.” (embargo until 2026)[2]

Topics and projects:

- Complex networks for biology and computational social science. Network centrality measures, community detection, evolution of the network communities over time, network diffusion. Applications: protein-protein-interaction networks, mobility networks between cities, online social networks (Twitter data).
 - Epidemiological modelling at multiple scales. Development of deterministic and stochastic compartmental models (SIR, SEAIR). Modelling the diffusion of bacteria in a hospital ward, implementing control measures and fitting model parameters on epidemiological curves. Modelling the spread of an infectious disease across multiple cities using mobility data as in- and out-fluxes.
 - Deep Natural Language Processing for text and protein sequences. Deep Learning Language models to represent text data as numerical vectors (BiLSTM neural networks, Transformers (BERT)). Short text classification with Deep Learning classifiers (BERT model with zero-shot learning, Named Entity Recognition, regular expressions). Predicting functional properties of proteins starting from the neural network embeddings with machine learning.
 - Survival analysis on cancer data. Harmonizing hospital dataset to perform survival analysis with Cox model. Penalized Cox model to deal with highly dimensional datasets (genomics, radiomics). Predicting risk of severe cases with logistic regression.
-
-

EDUCATION

Master's degree in Physics, curriculum of Applied Physics for Complex Systems and Health, University of Bologna

Final mark: 110/110 cum Laude, 14/12/2018

Thesis (paper): "Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota"[3]

Topics and exams: complex networks, pattern recognition, statistical mechanics. Projects: data analysis, dynamic processes simulation for graphs, network analysis, image recognition with convolutional neural networks, multivariate regression with machine learning and applications to econometrics.

Bachelor's degree in Physics, University of Bologna

Final mark: 110/110 cum Laude, 28/10/2016

High School diploma, Liceo Classico G.L. Storoni, Pesaro

Final mark: 100/100 cum Laude, 07/2013

SUMMER SCHOOLS

DeepLearn Summer School

25-29 July 2022, Las Palmas de Gran Canaria, Spain

Topics: representation learning, deep learning transformers and synthetic data generation.

Behavioral Digital Trace Data in Response to the COVID-19 Pandemic

26-31 July 2021, Online

Topics: computational social science, behavioral digital traces, social media data, epidemiological modelling, mobility data.

Summer school on Parallel Computing

13-24 May 2019, CINECA, Bologna, Italy

CONFERENCES

VEO symposium 2023 (EU project: Versatile Emerging infectious disease observatory)

12-14 June 2023, Rotterdam, Netherlands

Invited speech at YoungVEO: protein sequence embeddings through Deep Learning for the predictability of functional properties from protein sequences of viruses.

OBVIOUS symposium 2022 (Observatory on Vaccine Hesitancy in Italy – Online UniBo Surveys)

26 October 2022, Bertinoro, Italy

Results on vaccine hesitancy analysis on Twitter data (joint presentation with prof. D. Remondini).

VEO symposium 2022 (EU project)

May 2022, The Hague, Netherlands

Mentioned in the best posters: “Language-based Deep Learning for protein sequences vectorization and phenotype prediction”.

COLLABORATIONS

VEO (EU project: Versatile Emerging infectious disease observatory)

- With prof. M. Salathé, Digital Epidemiology Lab, EPFL: analysis of COVID-19 related social media posts from Twitter (more than 350M). Retweets network and NLP analysis to quantify the attention shifts in the first months of the pandemic[4]. Dynamics of social media behavior before and after SARS-CoV-2 infection[1]: Deep Learning- and regex-based NLP to detect a cohort of 11K Twitter users reporting their SARS-CoV-2 infection on Twitter.
- With Dr. S. Lycett (University of Edinburgh), prof. R. de Vries (EMC), prof. B. Haagmans (EMC): embedding of SARS-CoV-2 protein sequences with Deep Learning language models. Comparison of Deep Learning embeddings and phylogenetic distances. Predicting antigenic properties for SARS-CoV-2 variants starting from protein sequences (immune escape).
- With prof. M. Koopmans: analysis of COVID-19 related citation network on PubMed (more than 150K papers/nodes). Community detection and word frequency analysis to characterize the communities.
- With prof. R. Fouchier, EMC: predicting antigenic phenotype properties for influenza viruses starting from protein sequences (Deep Learning Language Models, BiLSTM neural networks, Transformers, chemical properties). Clustering, dimensionality reduction, prediction on artificially generated protein sequences, feature importance. Abstract submitted at Epidemics 9, 9th International Conference on Infectious Disease Dynamics, 28th November-1st December 2023: “From geno to pheno: reconstructing Influenza antigenic maps through Natural Language approaches to HA protein sequences”.
- With prof. F. Pasquali, University of Bologna: integrated analysis of viral load of SARS-CoV-2 in Bologna wastewater, hospitalizations, road traffic and phylogenetic profiling of variants. Paper in preparation for submission: “Multi-factorial analysis of SARS-CoV-2 detection in wastewater integrated with clinical and mobility data in Bologna metropolitan area”.
- Participation at bi-weekly meetings, presentation of results regularly since 2020.

OBVIOUS (Observatory on Vaccine Hesitancy in Italy – Online UniBo Surveys), with prof. M. P. Fantini, Dr. D. Gori, Department of Biomedical and Neuromotor Sciences, University of Bologna

- Social media surveillance of Italian Twitter posts related to vaccines. Collection and storage of more than 18M tweets. Time trend analysis[5], network analysis to quantify the polarization of the debate.
- Organization of manual annotation of tweets as ProVax, Neutral and AntiVax with around 30 students from Master’s degree in Physics and Specialty School of Medicine (Public Health). Analysis of manually annotated tweets and users to quantify vaccine hesitancy over time[6].
- Classification of tweets as ProVax/AntiVax using both Deep Learning embeddings for text and retweets network features at user level.

EPI-Net (EU project: The Epidemiology Network), with prof. E. Tacconelli, Division of Infectious Diseases, Department of Diagnostics and Public Health, University of Verona

- Development of a compartmental model to describe the diffusion of an antimicrobial resistant (AMR) bacterium in a hospital setting. Implementation of AMR- and infection-control interventions in the model, according to literature and hospital data. Mechanistic interpretations of model parameters and association with clinical data. Validation on longitudinal point prevalence time series of *Carbapenem resistant Klebsiella Pneumonia*. Sensitivity analysis to quantify the effect of intervention and policies.
- Publications: “Modelling antimicrobial resistance transmission to guide personalized antimicrobial stewardship interventions and infection control policies in healthcare setting: a pilot study.”[7]

Prof. A. Bazzani, Department of Astronomy and Physics, University of Bologna

- Epidemiological model with N meta-populations to describe the spread of SARS-CoV-2 in Veneto (Italy) using aggregated human mobility data provided by phone companies (collaboration with prof. E. Tacconelli, University of Verona). Machine learning to predict infection peaks and size from network features and geographic data.

Systemic (EU project: circular solutions for biowaste)

- Exploratory analysis of agricultural-related social media posts from Twitter and identification of main social media authorities. Characterization of main peaks in the time series in association with external events.
- Automated search of Scopus scientific papers database about “irrigation management”, “water availability”, “crop yield” and “crop evapotranspiration”. Deep Learning embedding of abstracts, clustering, dimensionality reduction, trending topics.

Health Big Data (IRCCS network, funded by Italian Ministry of Health)

- With prof. M. Carman: NLP tools to extract information from hospital data in text format. Document retrieval, question answering, anonymization.
- Attendance to weekly meetings and online challenges.

E-MUSE (EU project: Complex microbial Ecosystems MUltiScale modElling: mechanistic and data driven approaches integration)

- Diffusion-based detection of new AMR genes in *E. Coli* starting from a list of known AMR genes. Dataset: *E. Coli* protein-protein-interaction network and ResFinder database of AMR genes.

Genomed4all (EU project: Genomics For Next Generation Healthcare)

- Survival analysis on multiple myeloma dataset to develop a clinical support decision system (CSDS). Integration of clinical exams (PET, blood tests), copy number variation data and radiomics features.
- Preparation of a deliverable on CSDS and open source software.

Microbiology Lab, Pievesestina, with prof V. Sambri, Department of Medical and Surgical Sciences, University of Bologna

- Characterization of a dataset of 5K SARS-CoV-2 genomes from Emilia Romagna, to extract geographic related patterns in the spread of the pandemic. Deep Learning language model embeddings of proteins, variants identification.
- (About to start): Project of Relevant National Interest (PRIN) on machine learning to predict sepsis in hospital infections of *Klebsiella Pneumoniae*. (Link unibo magazine: <https://magazine.unibo.it/archivio/2023/06/19/un-sistema-di-intelligenza-artificiale-per-prevedere-il-rischio-di-sepsi>)

Local health unit, Bologna (AUSL)

- Member of the team providing weekly updates and prediction about COVID-19 hospitalizations to the local health unit of Bologna. Epidemiological modelling, compartmental models.
 - Analysis of risk factors for severe cases of COVID-19 (comorbidities, anagraphic data). Under review at PLOS ONE: “Risk Factors for Admission into COVID-19 General Wards, Sub-Intensive and Intensive Care Units among SARS-CoV-2 Positive Subjects in the Municipality of Bologna, Italy”.
-
-

TEACHING AND DISSEMINATION

Scientific degrees plan (Piano Lauree Scientifiche)

2022; 2023; 20 hours yearly

Lectures for high school students about text and DNA analysis, Markov chain generative models, simple neural networks and Python basics.

Models and analysis of biomedical and epidemiological Big Data

A.A. 2021/2022, 8 hours (1CFU)

Optional course for medicine students in Ravenna, University of Bologna, integrated course with prof. M. P. Fantini, Dr. D. Gori. Topics: epidemiological modelling with compartmental models, social media as a source of epidemiological surveillance, opportunities in artificial intelligence for health.

Complex Networks and Pattern Recognition

2020-2023, ~4 hours yearly

Lectures for Master’s degree students in Physics for the course of Complex Networks (prof. D. Remondini) and Pattern Recognition (prof. G. Castellani). Lectures about current research on: online social network analysis, recursive neural networks, Deep Learning Language models for text and proteins.

European Researchers’ Night @ Bologna

29th September 2023

Assistant at Virtual reality lab realized for Euclid satellite project, with Augmented Alma.

E-MUSE training event

28th November – 2nd December 2022, 3 hours

Lecture for international PhD students and postdocs of the E-MUSE EU project, about applied complex network analysis in R. Dataset: protein-protein-interaction network of *E. Coli*.

Tutor at Camplus, Bologna

2019-2021, ~100 hours

Private lessons to University students about mathematical analysis and physics.

Mentions in Unibo Magazine (scientific dissemination)

2023: “Un sistema di intelligenza artificiale per prevedere il rischio di sepsi” (eng: An artificial intelligence system to predict sepsis risk): <https://magazine.unibo.it/archivio/2023/06/19/un-sistema-di-intelligenza-artificiale-per-prevedere-il-rischio-di-sepsi>

2021: “L’influenza della comunità scientifica durante la pandemia nel dibattito online” (eng: The influence of the scientific community during the pandemic in the online debate):

<https://magazine.unibo.it/archivio/2021/11/04/2019influenza-della-comunita-scientifica-durante-la-pandemia-nel-dibattito-online>

2021: "Unibo in finale alla Pandemic Response Challenge" (eng: Unibo in final at Pandemic Response Challenge): <https://magazine.unibo.it/archivio/2021/02/11/unibo-in-finale-alla-pandemic-response-challenge>

Co-supervisor of final dissertation for Bachelor's and Master's students in Physics

- A.A.2021-2022: "Analysis of COVID-related symptoms from a database of Italian Tweets", by K. M. Damiani. Bachelor's degree in Physics.
 - A.A.2022-2023: "Integration of Natural Language Processing and network observables for a classification task: vaccine perception in Twitter", by D. Pignedoli. Master's degree in Applied Physics.
-
-

PEER REVIEWS

- Reviewer for Springer's Social Network Analysis and Mining, ISSN: 1869-5469.
 - Reviewer for PLOS ONE, ISSN: 1932-6203.
 - Member of the Editorial Board of Frontiers in Public Health since May 2023, section "Digital Public Health".
-
-

CERTIFICATES & AWARDS

Physics Teaching licence for High School (A20 class)

2022

Machine Learning course by Andrew Ng, Coursera

2018

IELTS, British Council, English language certificate

2016, Mark: 7.0

Prize for High School students by Fondazione Occhialini

2014, Rank: 2nd, Topic: relativistic mechanics, evaluation by Prof. N. Semprini, Department of Astronomy and Physics, University of Bologna

SKILLS

- 7 years' experience in Python (coding language). Libraries and topics: Pandas, Seaborn, Scikit-learn for machine learning, Lifelines for survival analysis, Deep Learning neural networks (Tensorflow, Torch, Huggingface's Transformers). Networkx and Igraph for complex networks. Lmfit for model fitting.
- Familiarity with R and Matlab (coding languages). Survival analysis and Cox model, Neural Networks for image recognition, differential analysis of microbiomes with DeSeq2.
- 4 years of experience in Scientific writing: peer-reviewed publications, thesis, EU projects deliverables and presentations.

REFERENCES

- [1] F. Durazzi, F. Pichard, D. Remondini, e M. Salathé, «Dynamics of social media behavior before and after SARS-CoV-2 infection», *Front. Public Health*, vol. 10, 2023, Consultato: 2 luglio 2023. [Online]. Disponibile su: <https://www.frontiersin.org/articles/10.3389/fpubh.2022.1069931>
- [2] F. Durazzi, «Data surveillance for infectious diseases: models, complex networks and machine learning». 31 gennaio 2023.
- [3] F. Durazzi, C. Sala, G. Castellani, G. Manfreda, D. Remondini, e A. De Cesare, «Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota», *Sci. Rep.*, vol. 11, fasc. 1, 2021, doi: 10.1038/s41598-021-82726-y.
- [4] F. Durazzi, M. Müller, M. Salathé, e D. Remondini, «Clusters of science and health related Twitter users become more isolated during the COVID-19 pandemic», *Sci. Rep. 2021 111*, vol. 11, fasc. 1, pp. 1–11, ott. 2021, doi: 10.1038/s41598-021-99301-0.
- [5] D. Gori, C. Reno, D. Remondini, F. Durazzi, e M. P. Fantini, «Are we ready for the arrival of the new covid-19 vaccinations? Great promises and unknown challenges still to come», *Vaccines*, vol. 9, fasc. 2, 2021, doi: 10.3390/vaccines9020173.
- [6] D. Gori *et al.*, «Mis-tweeting communication: a Vaccine Hesitancy analysis among twitter users in Italy», *Acta Bio-Medica Atenei Parm.*, vol. 92, fasc. S6, 2021, doi: 10.23750/abm.v92iS6.12251.
- [7] F. Durazzi *et al.*, «Modelling antimicrobial resistance transmission to guide personalized antimicrobial stewardship interventions and infection control policies in healthcare setting: a pilot study», *Sci. Rep.*, vol. 13, fasc. 1, Art. fasc. 1, set. 2023, doi: 10.1038/s41598-023-42511-5.