

Leonardo Pellegrina

Postdoctoral Researcher

June, 2022

Department of Information Engineering
University of Padova
Via Gradenigo 6b, Padova, Italy, IT 35131
leonardo.pellegrina@unipd.it
<http://www.dei.unipd.it/~pellegri>

EDUCATION

Ph.D. in Information Engineering, University of Padova (XXXIII cycle) 24th March 2021
Thesis: *Rigorous and Efficient Algorithms for Significant and Approximate Pattern Mining*.
http://www.dei.unipd.it/~pellegri/thesis/leonardo_pellegrina_tesi.pdf
Advisor: Prof. Fabio Vandin.

M.S. in Computer Engineering, University of Padova, *summa cum laude* 12th September 2017
Thesis: *Efficient Incremental Mining of Significant Patterns*.
<https://www.dei.unipd.it/~pellegri/thesis/MS-thesis-LP.pdf>
Advisor: Prof. Fabio Vandin.

B.S. in Information Engineering, University of Padova 24th November 2015
Thesis: *Network Performance Evaluation of Fat-Tree and B-Cube Topologies*.
Advisor: Prof. Leonardo Badia.

ACADEMIC APPOINTMENTS

Postdoctoral Researcher (Junior type B grant), Department of Information Engineering,
University of Padova 1st December 2020 - 30th November 2022

Visiting Research Fellow, Department of Computer Science, Brown University, RI (USA),
advised by Prof. Eli Upfal. 1st January 2019 - 30th June 2019

TEACHING ACTIVITIES

Lecturer (with responsibility of the course), *Dati e Algoritmi 1 (Data structures and Algorithms 1)*, B.S. in Computer Engineering, Department of Information Engineering, University of Padova
A.A. 2021/2022

Teaching Assistant (Tutor Junior), *Machine Learning*, Department of Information Engineering,
University of Padova 2017 - 2021

Teaching Assistant (Tutor Junior), *Foundations of Computer Science*, Department of
Information Engineering, University of Padova 2019 - 2020

Teaching Assistant (Didattica Integrativa), *Big Data Computing*, Department of Information
Engineering, University of Padova 2017 - 2019

Teaching Assistant (Tutor Junior), *Tutorato Formativo*, Department of Information Engineering,
University of Padova 2017 - 2019

PUBLICATIONS

Preprints (under review)

- [1] *Sharper convergence bounds of Monte Carlo Rademacher Averages through Self-Bounding functions*, **Leonardo Pellegrina**, arXiv (2020). <https://arxiv.org/abs/2010.12103>
- [2] *SILVAN: Estimating Betweenness Centralities with Progressive Sampling and Non-uniform Rademacher Bounds*, **Leonardo Pellegrina** and Fabio Vandin, arXiv (2021). <https://arxiv.org/abs/2106.03462>

During my master studies and before starting my Ph.D., I engaged with research through reproducibility studies of algorithms for Information Retrieval, and in the design of a novel system for the use of tethers in microgravity conditions, a multidisciplinary project in collaboration with the Department of Industrial Engineering of the University of Padua. My main responsibilities were in the design and implementation of the software system for the controlled deployment of a tethered probe, and in the analysis of the data that we collected from various sensors during the microgravity experiments at the ZARM facility in Bremen (Germany).

- [1] *An innovative space tether deployer with retrieval capability: Design and test of STAR experiment*, Gilberto Grassi, Alessia Gloder, **Leonardo Pellegrina**, Mattia Pezzato, Alvise Rossi, Francesco Branz, Matteo Duzzi, Riccardo Mantellato, Lorenzo Olivieri, Francesco Sansone, Enrico C. Lorenzini, Alessandro Francesconi, Proceedings of the 68th International Astronautical Congress (**IAC 2017**). <http://hdl.handle.net/11577/3243555>
- [2] *Design and test in microgravity of a space tether length and length rate measurement device*, Gilberto Grassi, Mattia Pezzato, Alessia Gloder, Riccardo Mantellato, Alessandro Francesconi, Enrico Lorenzini, Alvise Rossi and **Leonardo Pellegrina**, IEEE International Workshop on Metrology for AeroSpace (**MetroAeroSpace 2017**). <https://doi.org/10.1109/MetroAeroSpace.2017.7999602>
- [3] *Towards open-source shared implementations of keyword-based access systems to relational data*, Alex Badan, Luca Benvegnú, Matteo Basetton, Giovanni Bonato, Alessandro Brighente, Alberto Cenzato, Piergiorgio Ceron, Giovanni Cogato, Stefano Marchesin, Alberto Minetto, **Leonardo Pellegrina**, Alberto Purpura, Riccardo Simionato, Nicolò Soletti, Matteo Tessarotto, Andrea Tonon, Federico Vendramin, Nicola Ferro, Workshop Proceedings of the joint 20th International Conferences on Extending Database Technology and Database Theory (**EDBT/ICDT 2017**). <http://hdl.handle.net/11577/3254670>
- [4] *Keyword-based access to relational data: To reproduce, or to not reproduce?*, Alex Badan, Luca Benvegnu, Matteo Basetton, Giovanni Bonato, Alessandro Brighente, Stefano Marchesin, Alberto Minetto, **Leonardo Pellegrina**, Alberto Purpura, Riccardo Simionato, Matteo Tessarotto, Andrea Tonon, Nicola Ferro, Proceedings of the Symposium on Advanced Database Systems (**SEBD 2017**). <http://hdl.handle.net/11577/3254675>

Note that papers in computer science conference proceedings are fully refereed publications (often with very low acceptance rates). In all the following papers, authors are sorted by contribution; as first author (or equal first author, denoted by *), I was the principal contributor in the design and analysis of the proposed algorithmic methods, in their implementation and experimental evaluations.

- [5] *Efficient Mining of the Most Significant Patterns with Permutation Testing*, **Leonardo Pellegrina** and Fabio Vandin, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (**KDD 2018**). (18.4% acceptance rate) <https://doi.org/10.1145/3219819.3219997>
- [6] *Fast Approximation of Frequent k-mers and Applications to Metagenomics*, **Leonardo Pellegrina**, Cinzia Pizzi and Fabio Vandin, Proceedings of the 23th Annual International Conference on Research in Computational Molecular Biology (**RECOMB 2019**). https://link.springer.com/chapter/10.1007/978-3-030-17083-7_13
- [7] *SPuManTE: Significant Pattern Mining with Unconditional Testing*, **Leonardo Pellegrina**, Matteo Riondato and Fabio Vandin, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (**KDD 2019**). (14.2% acceptance rate) <https://dl.acm.org/doi/abs/10.1145/3292500.3330978>
- [8] *MCRapper: Monte-Carlo Rademacher Averages for Poset Families and Approximate Pattern Mining*, **Leonardo Pellegrina**, Cyrus Cousins, Fabio Vandin and Matteo Riondato, Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (**KDD 2020**). (16.8% acceptance rate) <https://doi.org/10.1145/3394486.3403267>

- [9] *SPRISS: Approximating Frequent k-mers by Sampling Reads, and Applications*, Diego Santoro*, **Leonardo Pellegrina***, and Fabio Vandin, 25th Annual International Conference on Research in Computational Molecular Biology (**RECOMB 2021**).
<https://arxiv.org/abs/2101.07117>

Journals

- [1] *Fast Approximation of Frequent k-mers and Applications to Metagenomics*, **Leonardo Pellegrina**, Cinzia Pizzi and Fabio Vandin, **Journal of Computational Biology Special Issue of Best Papers from the RECOMB 2019 conference**, 27, 4, 534-549 (2020).
<https://doi.org/10.1089/cmb.2019.0314>
- [2] *Efficient Mining of the Most Significant Patterns with Permutation Testing*, **Leonardo Pellegrina** and Fabio Vandin, **Data Mining & Knowledge Discovery** 34, 1201-1234 (2020). <https://doi.org/10.1007/s10618-020-00687-8>
- [3] *SPRISS: Approximating Frequent k-mers by Sampling Reads, and Applications*, Diego Santoro, **Leonardo Pellegrina**, Matteo Comin, and Fabio Vandin, *Bioinformatics* (2022), <https://doi.org/10.1093/bioinformatics/btac180>
- [4] *MCRapper: Monte-Carlo Rademacher Averages for Poset Families and Approximate Pattern Mining*, **Leonardo Pellegrina**, Cyrus Cousins, Fabio Vandin and Matteo Riondato, *ACM Transactions on Knowledge Discovery and Data Mining* (2022)
<https://doi.org/10.1145/3532187>
- [5] *Discovering Significant Evolutionary Trajectories in Cancer Phylogenies*, **Leonardo Pellegrina** and Fabio Vandin, to appear in *Bioinformatics* (2022).
<https://github.com/VandinLab/MASTRO/blob/main/mastro.pdf>

Thesis

- [1] *Efficient Incremental Mining of Significant Patterns*, **Leonardo Pellegrina**, M.S. Thesis (2017). <https://www.dei.unipd.it/~pellegrini/thesis/MS-thesis-LP.pdf>
- [2] *Rigorous and Efficient Algorithms for Significant and Approximate Pattern Mining*, **Leonardo Pellegrina**, Ph.D. Thesis (2021).
http://www.dei.unipd.it/~pellegrini/thesis/leonardo_pellegrina_tesi.pdf

Patents

Patent no. IT201700039763A1 for the invention *Automatic equipment for the launch and controlled recovery of a tethered mass in orbit*, c/o Italian Office for Patents and Trademarks (2017). Submitted to the World Intellectual Property Organization (WO2018189633A1).
<https://patents.google.com/patent/WO2018189633A1>

PRESENTATIONS AT INTERNATIONAL CONFERENCES

Efficient Mining of the Most Significant Patterns with Permutation Testing, 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2018) (short talk), London (UK). 21th August 2018

Fast Approximation of Frequent k-mers and Applications to Metagenomics, 23th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2019), Washington, DC (USA). 8th May 2019

SPuManTE: Significant Pattern Mining with Unconditional Testing, 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2019) (short talk), Anchorage, AK (USA). 6th August 2019

MCRapper: Monte-Carlo Rademacher Averages for Poset Families and Approximate Pattern Mining, 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2020), San Diego, CA (USA), virtual event. 26th August 2020

Discovering Significant Evolutionary Trajectories in Cancer Phylogenies, 26th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2022) (poster), La Jolla, San Diego, USA, 2022. 23th May 2022

Discovering Significant Evolutionary Trajectories in Cancer Phylogenies, 30th Conference on Intelligent Systems for Molecular Biology (ISMB 2022), Madison, Wisconsin, USA, 2022. 10-14th July 2022

Discovering Significant Evolutionary Trajectories in Cancer Phylogenies, 21th European Conference on Computational Biology (ECCB 2022), Sitges, Barcelona, ES, 2022. 12-21th September 2022

INVITED TALKS

Rigorous Mining of the Most Significant Patterns, Workshop on Foundations of Learning from Data (FouLarD), Bertinoro, Italy. 10th September 2018

Efficient Mining of the Most Significant Patterns with Permutation Testing, Kick off meeting of the PRIN project AHeAd, Luiss University, Rome, Italy. 23th September 2019

SPuManTE: Significant Pattern Mining with Unconditional Testing, Workshop on Computation and Statistics in Data science (CaStleD), Bertinoro, Italy. 4th October 2019

Fast Approximation of Frequent k-mers and Applications, invited lecture for the *Machine Learning* course, Department of Information Engineering, University of Padova, Italy. 25th November 2019

SPRISS: Approximating Frequent k-mers by Sampling Reads, and Applications, 9th RECOMB Satellite on Computational Methods in Genetics (RECOMB-Genetics). 2th September 2021

FELLOWSHIPS AND AWARDS

Hans von Muldau Team Award for the Best Team Project at the International Astronautical Congress (IAC) 2017, Adelaide, Australia. September 2017

Cariparo Foundation Ph.D. Fellowship, Padova, Italy. 2017 - 2021

Luciano Iglesias Foundation Fellowship, Padova, Italy. May 2018

Ing. Aldo Gini Foundation Fellowship, Padova, Italy. November 2018

ISCB RECOMB 2019 Travel Fellowship. February 2019

Honorable mention for 2021 SIGKDD Dissertation Award.
<https://kdd.org/awards/view/2021-sigkdd-dissertation-award-winners> August 2021

Honorable mention for Best Program Committee member at ACM The Web Conference 2022. <https://www2022.thewebconf.org/awards/#mention-2> April 2022

ISCB ECCB 2022 Travel Fellowship. June 2022

SCIENTIFIC ACTIVITIES

Tutorials

- [1] *Multiple Hypothesis Testing and Statistically-sound Pattern Mining*, **Leonardo Pellegrina**, Matteo Riondato, Fabio Vandin, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (**KDD 2019**).
<https://doi.org/10.1145/3292500.3332286>
- [2] *Multiple Hypothesis Testing and Statistically-sound Pattern Mining*, **Leonardo Pellegrina**, Matteo Riondato, Fabio Vandin, Proceedings of the 2021 SIAM International Conference on Data Mining (**SDM 2021**).
https://meetings.siam.org/session/dsp_programsess.cfm?SESSIONCODE=71922

Program Committees

RECOMB Satellite Workshop on Massively Parallel Sequencing (RECOMB-Seq) 2021.
ACM International Conference on Web Search and Data Mining (WSDM) 2022.
SIAM International Conference on Data Mining (SDM) 2022.
ACM The Web Conference (TheWebConf) 2022.
ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2022 (Research and Applied Data Science tracks).

Conference Reviewing

Workshop on Algorithms in Bioinformatics (WABI) 2018.
International Conference on Research in Computational Molecular Biology (RECOMB) 2019, 2020.
ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD) 2019, 2020, 2021.
The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD) 2019, 2020, 2021.
ACM International Conference on Information and Knowledge Management (CIKM) 2019.
IEEE International Conference on Data Mining (ICDM) 2019, 2020.
ACM International WSDM Conference on Web Search and Data Mining (WSDM) 2020.
AAAI Conference on Artificial Intelligence 2020.
International Conference on Intelligent Systems for Molecular Biology (ISMB) 2020.
ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB) 2021.
ACM The Web Conference (TheWebConf) 2021.

Journal Reviewing

Journal of Graph Algorithms and Applications (JGAA) 2020.
Journal of Machine Learning Research (JMLR) 2021.
IEEE Journal of Biomedical and Health Informatics (JBHI) 2022.

Professional Memberships

Member of the International Society for Computational Biology (ISCB).
Member of the Association for Computing Machinery (ACM).

SOFTWARE PACKAGES

- TopKWY, an efficient algorithm for mining the most significant patterns with permutation testing. <https://github.com/VandinLab/TopKWY>
- SAKEIMA, a fast algorithm to compute accurate approximations of frequent k -mers from large sequencing data sets based on random sampling of k -mers. <https://github.com/VandinLab/SAKEIMA>
- SPuManTE, an efficient algorithm for mining significant pattern with unconditional testing. <https://github.com/VandinLab/SPuManTE>
- MCRapper, a practical algorithm to compute rigorous approximations of pattern languages with a poset structure using Monte-Carlo Empirical Rademacher Averages. <https://github.com/VandinLab/MCRapper>
- SPRISS, a flexible and fast algorithm to compute sound approximations of frequent k -mers by sampling reads from large sequencing data sets. <https://github.com/VandinLab/SPRISS>
- SILVAN, a fast sampling-based algorithm based on non-uniform Monte-Carlo Empirical Rademacher Averages to compute rigorous approximations of the betweenness centrality of nodes in large graphs. <https://github.com/VandinLab/SILVAN>
- MASTRO, our algorithm to discover statistically significant conserved evolutionary trajectories of alterations from phylogenetic tumor trees. <https://github.com/VandinLab/MASTRO>