

**Algorithms
for
Intelligent Data Analytics**

Theses Proposals

May 2024

AIDA lab

Algorithms for Intelligent Data Analytics



<https://aidalabdei.github.io/>

Focus: algorithms for extracting useful information from large datasets

Topics for theses:

- Modern computing architectures
- Pattern mining and learning
- Networked, temporal, and evolving data
- Unsupervised learning (clustering, diversity)

Modern Computing Architectures

Theses topics:

- Design algorithms that fully exploit modern architectures
 - Memory hierarchy and parallel architectures
 - Hardware accelerators (e.g., tensor cores)
 - Processing-in-memory architectures
- Lower bounds methods for analyzing running times in modern architectures

Contacts:

- Francesco Silvestri (francesco.silvestri@unipd.it)

Pattern Mining and Learning

Theses topics:

- **Significant pattern mining**: efficient algorithms to find statistically significant patterns (sampling, statistical learning)
- **Algorithms with predictions**: how to use machine learning to improve algorithms
- **Graph neural networks**: generalizability, application to combinatorial problems (e.g., subgraph counting)
- **Applications to biology and medicine**: significant alterations in cancer, patterns relevant for clinical predictions

Contacts:

- Leonardo Pellegrina (leonardo.pellegrina@unipd.it)
- Andrea Pietracaprina (andrea.pietracaprina@unipd.it)
- Geppino Pucci (geppino.pucci@unipd.it)
- Fabio Vandin (fabio.vandin@unipd.it)

Networked, Temporal, and Evolving Data

Theses topics:

- **Algorithms for (temporal) networks:** sampling for subgraph counting, centralities, significant patterns
- **Algorithms for trajectories:** trajectory similarity, differential privacy for trajectories
- **Algorithms for mobility data:** analysis of mobility data for understanding human mobility
- **Algorithms for data streams:** data analytics for insertion-only, sliding window, and fully-dynamic settings
- **Algorithms for time series:** motif discovery, anomaly detection, indexing

Contacts:

- Matteo Ceccarelo (matteo.ceccarelo@unipd.it)
- Leonardo Pellegrina (leonardo.pellegrina@unipd.it)
- Andrea Pietracaprina (andrea.pietracaprina@unipd.it)
- Geppino Pucci (geppino.pucci@unipd.it)
- Francesco Silvestri (francesco.silvestri@unipd.it)
- Fabio Vandin (fabio.vandin@unipd.it)

Unsupervised Learning

Theses topics:

- **Clustering of very large noisy datasets:** clustering with outliers/uncertain data and outlier detection, in distributed and streaming settings
- **Clustering under novel objective functions:** silhouette-based clustering, clustering under additional constraints (e.g., matroid, fairness, knapsack)
- **Similarity search:** randomized algorithms for searching similarities in high dimensional data (e.g., Locality-Sensitive Hashing)
- **Dimensionality reduction:** algorithms for embedding high-dimensional in small embeddings (e.g., Johnson-Lindstrauss)

Contacts:

- Matteo Ceccarelo (matteo.ceccarelo@unipd.it)
- Andrea Pietracaprina (andrea.pietracaprina@unipd.it)
- Geppino Pucci (geppino.pucci@unipd.it)
- Francesco Silvestri (francesco.silvestri@unipd.it)
- Fabio Vandin (fabio.vandin@unipd.it)

Common Themes

Rigorous algorithms and techniques

Scalability on big data

Networked data (graphs, networks)

Privacy and fairness

Efficient implementation