David Gentile, PhD

contact@pilesofsand.com · pilesofsand.com · github.com/dcgentile · linkedin.com/in/dcgentile

Mathematics PhD in optimal transport, the geometry of probability distributions, and the numerics that support data-driven experiments, bringing theory to practice. I've published open-source libraries for Python and Julia, built a pipeline for classifying scientific literature using LLMs, and applied transport methods to underwater acoustic and molecular dynamics data sets. My aim is to help build a robust and dependable AI future that produces transparent answers and empowers scientists to pursue the most pressing problems.

EXPERIENCE

Naval Research Enterprise Internship Program Summer 2025
Naval Research Laboratory, Acoustics Division

- Applied optimal transport and segment-based clustering to classify non-uniformly sampled underwater acoustic signals, for the division's work on analog-to-digital conversion. Authored the resulting DTIC technical report.

Senior Data Scientist Intern Summer 2023
Analysis Group, Inc., Data Science Team

- Built a classification pipeline driven by an LLM to cut the time and labor cost of systematic literature review in industrial drug design. Benchmarked it against a colleague's support vector machine over vector-embedded abstracts on the same corpus. Findings fed internal tooling and a conference poster session.

SOFTWARE

- **GraphTransportation.jl** — open-source Julia library for optimal transport and barycentric coding of measures on graphs, from dissertation work supervised by James M. Murphy. Four solver backends (cone programming, Hamiltonian shooting, Chambolle-Pock, Sinkhorn), validated against closed-form solutions and published results. github.com/dcgentile/GraphTransportation.jl
- **graphtransport-py** — Python port of GraphTransportation.jl, built on PyTorch so the transport routines work as differentiable layers. github.com/dcgentile/graphtransport-py
- **MFpy** — scikit-learn-compatible Python library for transport-based change point detection and segment clustering in time series. github.com/dcgentile/mfpy
- **Interactive research demos** — browser demos of Wasserstein barycenters on graphs and in color space, running graphtransport-py. pilesofsand.com/demo

EDUCATION

PhD, Mathematics 2019 – 2026
Tufts University, advised by Prof. James M. Murphy

- Defended April 2026. Dissertation at the intersection of graph signal processing and optimal transport: barycentric coding for families of probability measures on graphs and discrete metric spaces.
- Central problem is a variance minimization over a Riemannian manifold of finitely supported distributions.
- Change point detection via transport distances on molecular dynamics trajectories, for Markov state models.

MS, Mathematics 2019 – 2021
Tufts University, advised by Prof. Bruce M. Boghosian

- Candidacy exam on random asset exchange models: SDEs, Fokker-Planck equations, kinetic theory.

BA, Liberal Arts (Concentration in Mathematics & Computer Science) 2014 – 2018
Sarah Lawrence College

PUBLICATIONS & WRITING

- D. Gentile, J. M. Murphy. "Static and Dynamic Approaches to Computing Barycenters of Probability Measures on Graphs." Under review (arXiv:2603.26940).
- D. Gentile, J. Huang, J. M. Murphy. "Wasserstein-Based Identification of Metastable States in Time Series Data via Change Point Detection and Segment Clustering." Under review (arXiv:2603.01989).
- D. Gentile, J. M. Murphy. "Regularized 2-Wasserstein Barycenters on Discrete Metric Spaces." SampTA 2025, Vienna.
- D. Gentile, L. T. Fialkowski, R. M. Oba. "An Approach to Signal Classification Via Optimal Transport and Machine Learning Using Nonuniformly Sampled Data." NRL/7160/MR–2025/6.

TALKS

- "A Barycentric Coding Model for Measures on Graphs." Invited speaker, FSU Topology & Geometry Seminar, Sept. 2026. pilesofsand.com/talks/seminar

SKILLS

Programming: Python, Julia, MATLAB; PyTorch, scikit-learn, NumPy; Git, GitHub Actions CI
Machine learning: document classification with LLMs, segment-based clustering, change point detection
Mathematics: optimal transport, graph signal processing, Wasserstein and information geometry, Riemannian optimization, non-smooth convex analysis, statistical physics (SDEs, Fokker-Planck)