Sushovan Majhi

Visiting Assistant Professor, George Washington University
Samson Hall 302, George Washington University, Washington, D.C., USA
s.majhi[AT]gwu.edu | Homepage | GitHub | Google Scholar | LinkedIn

Research Interests

Applied Topology, **Topological Data Analysis (TDA)**, **Computational Geometry**, **Pattern and Shape Matching**, and **Statistical Finance**.

My research primarily revolves around the interface of *computer science*, *mathematics*, and the mathematical foundations of *data science*. More specifically, I am motivated to develop provable inference techniques for data science that are inspired by topology and geometry. I also keep a keen interest in applying TDA to fascinating, real-world problems arising in fields, like biology, medicine, genetics, finance, and dynamical systems.

Teaching Interests

Mathematics: Algebraic Topology, Computational Topology, Differential Geometry, Manifold Theory, Analysis, Linear Algebra, Calculus, Applied Mathematics
Data Science: Probability Theory, Mathematical Statistics, Regression Analysis, Topological Data Analysis, Machine Learning, Data Mining, Algorithm Design for Data Science
Computer Science: Algorithms and Data Structures, Computational Geometry, Complexity Theory

Work Experience

- **Visiting Assistant Professor** **August 2023–*current***
  Data Science Program, George Washington University, Washington D.C., USA
  Role: Teaching data science and computer science courses to graduate students in the Data Science Program.

- **Postdoctoral Research Fellow** **January 2021–July 2023**
  School of Information, University of California, Berkeley, USA
  Role: The responsibilities included conducting research broadly in data science, forging new research collaborations, organizing research webinar series.

- **Data Science Instructor** **January 2021–July 2023**
  School of Information, University of California, Berkeley, USA
  Role: According to Fortune[1], MIDS is the **No.2-ranked** Master of Information and Data Science program in the US. Alongside instructing *Statistics for Data Science* for the program, I developed course materials, devised and maintained technology to facilitate teaching and learning.

- **Lecturer (MIDS Program)** **August 2020–December 2020**
  School of Information, University of California, Berkeley, USA
  Role: The position offered me (off-campus) work experience during my doctoral studies, and has served as a Curricular Practical Training (CPT).

Education

- **Doctor of Philosophy** in **Mathematics** **August 2014–December 2020**
  Tulane University, New Orleans, LA, USA.
  Advisor: **Prof. Carola Wenk**
  Courses: computational geometry, computational topology, topological data analysis, differential geometry, differentiable manifolds, algorithms, data structures, computational complexity, applied mathematics, scientific computing.

[1] https://fortune.com/education/information-technology/masters/rankings/best-online-masters-in-data-science

- **Master of Science** in **Mathematics** **August 2009–May 2012**
  Tata Institute of Fundamental Research, Bangalore, India
  Courses: ordinary and partial differential equations, probability theory, complex analysis, functional analysis, numerical linear algebra, measure theory, mechanics.

- **Bachelor of Science** in **Mathematics** (Hons.) **July 2006–May 2009**
  Ramakrishna Mission Vidyamandira, **Calcutta University**, West Bengal, India
  Courses: calculus, real analysis, linear algebra, numerical analysis, game theory, statistics, physics.

Graduate Student Advisees

- Khush Shah, current data science masters student, George Washington University, **August 2023–*current***
  Project Title: Matching geometric graphs

- Shiny Chakraborty, current project assistant, Indian Institute of Technology, India **August 2023–*current***
  Project Title: Discrete Morse theory in Road-network reconstruction

- Anish Rai, current PhD student **August 2022–*current***
  National Institute of Technology, Sikkim, India
  Project Title: Prediction of stock market crashes using topological data analysis

Preprints

2. Henry Adams, Florian Frick, **SM**, Nichola McBride*. Hausdorff VS Gromov–Hausdorff Distances, 22 pages, Sep 2023. Available at: arXiv:2309.16648[math.MG]

1. **SM**. Demystifying Latschev's theorem: Manifold Reconstruction from noisy data, 20 pages, June 2023. Available at: arXiv:2305.17288[math.AT]
   • Submitted to *Discrete and Computational Geometry*

Journal Papers and Preprints

5. **SM** and Carola Wenk. Distance Measures for geometric graphs. *Computational Geometry: Theory and Applications*, 2023. DOI: 10.1016/j.comgeo.2023.102056. Also available at: arXiv:2209.12869[cs.CG]

4. **SM**. Vietoris–Rips complexes of metric spaces near a metric graph. *Journal of Applied and Computational Topology*, May 2023. DOI: 10.1007/s41468-023-00122-z
   Also available at: arXiv:2204.14234[math.AT]

3. **SM**, Jeffrey Vitter, and Carola Wenk. Approximating Gromov-Hausdorff distance in Euclidean space. *Computational Geometry: Theory and Applications*, 116:102034, 2024.
   DOI: 10.1016/j.comgeo.2023.102034. Also available at: arXiv:1912.13008[math.MG]

2. Anish Rai, Ajit Mahata, Md Nurujjaman, **SM**, and Kanish Debnath. A sentiment-based modeling and analysis of stock price during the COVID-19: U- and Swoosh-shaped recovery. *Physica A: Statistical Mechanics and its Applications*, 592:126810, 2022. DOI: 10.1016/j.physa.2021.126810

1. Brittany Terese Fasy, Rafal Komendarczyk, **SM**, and Carola Wenk. On the reconstruction of geodesic subspaces of $\mathbb{R}^N$. *International Journal of Computational Geometry & Applications*, 32(1):91–117, 2022. DOI: 10.1142/S0218195922500066. Also available at: arXiv:1810.10144[math.AT]

* Undergraduate student at the time of research or submission

Peer-reviewed Conference Papers

2. Erin Chambers, Brittany Fasy, Benjamin Holmgren*, **SM**, and Carola Wenk. Metric and path-connectedness properties of the Fr̆echet distance for paths and graphs. In Proceedings of the *34th Canadian Conference on Computational Geometry (CCCG)*, 2023. Available at: arXiv:2308.00900[cs.CG]

1. **SM**. Graph mover's distance: An efficiently computable distance measure for geometric graphs. In Proceedings of the *34th Canadian Conference on Computational Geometry (CCCG)*, 2023. Available at: arXiv:2306.02133[cs.CG]

Workshop Contributions

4. E. Chambers, B. Fasy, B. Holmgren* , **SM**, and C. Wenk. Path-Connectivity of Fréchet Spaces of Graphs. Computational Geometry: Young Researchers Forum, 2022

3. **SM** and Carola Wenk. Distance Measures for Geometric Graphs, At *Fall Workshop on Computational Geometry (FWCG)*, 2022

2. Brittany Terese Fasy, **SM** and Carola Wenk. Threshold-based graph reconstruction using discrete Morse theory. In *Fall Workshop on Computational Geometry (FWCG)*, 2018. Abstract available at: Link

1. Brittany Terese Fasy, Rafal Komendaczyk, **SM**, and Carola Wenk. Topological reconstruction of metric graphs in $\mathbb{R}^N$. At *Fall Workshop on Computational Geometry (FWCG)*, 2017. Abstract available at: Link

Graduate Teaching

- **Algorithm Design (DATS 6001)** **Spring 2024**
  Data Science Program, George Washington University, USA
  Topics: basics of computational complexity, data structures (array, stack, queue, tree, etc), algorithms (search, sort, etc), and programming paradigms like dynamic and greedy algorithms.

- **Topological Data Analysis (CS)** **Fall 2023**
  National Institute of Technology, Sikkim, India
  Topics: topological spaces, metric spaces and their examples, simplicial complexes, homology in $\mathbb{Z}/2$ coeff, persistent homology, bottleneck and Wasserstein distance, and non-linear time-series analysis using TDA.

- **Computer Science Foundations (DATS 6450)** **Fall 2023**
  Data Science Program, George Washington University, USA
  Topics: computer design, programming in Python, object-oriented programming

- **Introduction to Data Mining (DATS 6103)** **Fall 2023**
  Data Science Program, George Washington University, USA
  Topics: data wrangling, linear and logistic regression, classification, clustering, data visualization, support vector machines, machine learning algorithms.

- **Statistics for Data Science (MIDS 203)** **August 2020–July 2023**
  School of Information, University of California, Berkeley, USA
  Topics: probability theory, sampling distributions, estimators and convergence theorems, confidence intervals, hypothesis testing, and regression.

- **Linear Algebra, Complex Analysis** **January 2013–June 2013**
  Christ University and Scimetric Pvt Ltd, Bangalore, India

- **Analysis, Linear Algebra, Complex Analysis** **November 2011–July 2012**
  GATE-IIT Coaching Institute, JP Nagar, Bangalore, India
  Graduate level, for competitive national exams, e.g., National Eligibility Test

* Undergraduate student at the time of research or submission

- **Analysis and Linear Algebra** **February 2012–July 2012**
  MES College, Department of Mathematics, Malleswaram, Bangalore, India

Undergraduate Teaching

- **Undergraduate Statistics for Business Students (MATH 1140)** **Summer 2019**
  Tulane University, USA
  Topics: sampling methods, descriptive statistics, probability theory, random variables, limit theorems, confidence intervals, hypothesis testing, and linear regression.

Other Teaching

- **Teaching Assistant** **Fall 2014–Spring 2017**
  Tulane University, USA

Research Experience

- **Hausdorff vs Gromov–Hausdorff Distances on Manifolds** **April 2023–*current***
  Collaborators: Henry Adams, University of Florida; Florian Frick, Carnegie Mellon University
  In shape reconstruction from noisy samples, both the Hausdorff and Gromov–Hausdorff distances are commonly used to facilitate noise models for the sample. But the relationship between the two distance measures is not well-understood. In this project, we further our understanding of their relationship, through the lenses of Čech and Vietoris–Rips complexes.

- **Vietoris–Rips Complexes near a Euclidean polytope** **June 2023–*current***
  Collaborator: Rafal Komendarczyk, Tulane University
  Finite reconstruction of Euclidean shapes that are not manifolds is a challenging computational problem. The project is motivated by the problem of topological reconstruction of Euclidean shapes that the stratified manifolds, examples include graphs, polytopes, etc.

- **Topological and Statistical Methods in Predicting the Crash and Recovery of Stock Markets** **March 2021–*current***
  Collaborator: Md. Nurujjaman, NIT Sikkim, India
  In the aftermath of stock market crash due to COVID-19, not all sectors recovered in the same way. We proposed novel models to capture the different types of recovery profiles for Indian stocks. We also employed the **Empirical Mode Decomposition** (EMD) for a statistical significance analysis of our model.

  We currently look into the possibility of predicting a future crash in a financial sector—using tools from **Topological Data Analysis** (TDA).

- **Topological and Geometric Signature-Based Shape Comparison** **October 2021–*current***
  Collaborators: Erin Chambers, Liz Munch, and Carola Wenk
  We consider geometric and topological signatures to concisely represent large datasets to facilitate their easy description and efficient comparison. To this end, we look for signatures in the class of algebraic, geometric, and graphical signatures. We have proposed new similarity measures for geometric graphs, and studied their computational aspects.

- **Computation of Gromov-Hausdorff Distance in Euclidean Space** **April 2019–December 2020**
  Collaborators: Helmut Alt, Jeffrey Vitter, and Carola Wenk
  We investigate the computational aspects of the Gromov–Hausdorff distance between sets equipped with the Euclidean metric. We used the Hausdorff distance under isometry to develop an approximation algorithm for the Gromov–Hausdorff distance on the real line with a tight approximation factor of $\left(1+\frac{1}{4}\right)$.

- **Topological Reconstruction of Geodesic Spaces** **December 2016–May 2019**
  Collaborators: Brittany Fasy and Rafal Komendarczyk
  Role: Research Assistant
  PI: Carola Wenk (NSF CCF-161846)

We investigate the reconstruction of geodesic subspaces of Euclidean spaces using the Vietoris-Rips and Čech complexes from a dense sample around it. We propose two new sampling parameters: **distortion** of embedding and **convexity radius** of the underlying geodesic space. We guarantee a successful computation of the Betti numbers. For the special case of planar graphs, we also develop an algorithm for its geometric reconstruction.

- **Dynamics and Prognosis of Chronic Myelogenous Leukemia (CML)** **August 2012–November 2013**
  National Center for Biological Sciences, TIFR, Bangalore, India
  Role: Junior Research Fellow
  PI: Seema Nanda
  In this joint effort to develop better prognostic tools for doctors, computational scientists teamed up with medical officers and biologists to understand the dynamics of CML by modeling the disease by systems of **differential equations**. In our parameter fitting, we made use of the big existing data collected from a large pool of CML patients. We also performed (statistical) **sensitivity analysis** to better understand the parameter spaces for our model.

Academic Services

- I have been a reviewer for Journal of Foundations of Data Science.
- I have been a reviewer for Journal of Combinatorial Optimization.
- I have been a reviewer for conferences, like International Symposium on Computational Geometry, International Symposium on Spatial and Temporal Databases, European Workshop on Computational Geometry, ACM International Conference on Advances in Geographic Information Systems, WADS Algorithms and Data Structures Symposium.
- I organized SIAM Graduate Student Chapters at Tulane University.
- I have been organizing data science webinar series at the University of California, Berkeley.

Computational Skills

Java, C, R, Python, Ruby, JavaScript, SQL, Bash.

Software Projects

- **Simplicial Complexes in JS** GitHub
  JavaScript implementation of some of the widely used computations on simplicial complexes. The library also implements the Smith Normal Form in order to compute the homology groups of an abstract complex.
- **Shape Reconstruction Visualization** WebApp | GitHub
  To complement my PhD research, I implemented my topological reconstruction algorithm for planar metric graphs in this library. The library is written in JavaScript and made available to users as a web-app.
  Skills: JavaScript, HTML, CSS.

Entrepreneurial Experience

- **Scimetric Edulabs Private Limited** **December 2012–April 2017**
  Bangalore, India
  Role: **co-founder** and **director**
  In this start-up venture, our objective was to motivate and train students in higher education. We won franchise to work with several private colleges in India. We coached science students for standardized entrance tests for PhD and academic jobs. The company employed 6 trainers.

Awards, Scholarships, Recognition

- **Travel Grant**, University of California, Berkeley
- **UGC-CSIR NET Research Fellowship**, India, June 2012
- **TIFR Junior Research Fellowship** for pursuing Integrated PhD studies at TIFR-CAM, Bangalore, India. August 2009.
- **Secured grade "A" in SCIENCE TALENT SEARCH EXAMINATION**
  conducted by JATIYA VIJNAN PARISAD and INDIAN SCIENCE CONGRESS ASSOCIATION

Invited Talks and Accepted Abstracts

- **Title:** Demystifying Latschev's Theorem — **August 23, 2023**
  Applied Algebraic Topology Research Network (AATRN)
  Available on: YouTube
- **Title:** Graph Mover's Distance — **August 2–4, 2023**
  The 34th Canadian Conference on Computational Geometry, Montreal, Canada
  Available on: Google Drive
- **Title:** Similarity Measures for Geometric Graphs — **October 14–15, 2022**
  The 30th Fall Workshop on Computational Geometry
  North Carolina State University, USA
- **Title:** Topological Methods in the Reconstruction and Comparison of Shapes — **February, 2022**
  Mathematics Department, ICFAI University, Tripura, India
- **Title:** A Taste of Topological Data Analysis (TDA): Reconstruction of Shapes — **September, 2021**
  Department of Mathematics, Hunter College, NY, USA
- **Fall Workshop on Computational Geometry** — **October, 2018**
  Queens College, New York, USA
- **Fall Workshop on Computational Geometry** — **November, 2017**
  SUNY (Stony Brook), New York, USA

Hobby

In my spare time, I write tutorials on *random* topics in order to make mathematics and statistics a little more interactive; they can be found here: https://www.smajhi.com/tutorials. I also enjoy playing the piano and classical guitar.