Research Statement: Sushovan Majhi

My research primarily revolves around the interface of ***mathematics*** and ***computer science***. More specifically, I am interested in problems arising in ***topological data analysis (TDA)***, ***computational and applied algebraic topology***, and ***computational geometry***. I am also very enthusiastic about applying TDA to other fields of science and developing computational ***libraries and software***.

Motivation. I drew on the affinity for applications of mathematics since a very early stage of my academic career. As a master's student at a research institute for ***applicable mathematics***, I was introduced to both the foundational and applicable aspects of mathematics. For hands-on research experience, I made an effort to get involved in ***cancer research*** there. Working closely with medical officers, pathologists, and mathematicians, I was intrigued by the endeavor to solve challenging real-life problems as a team – divided by their field of expertise but united by their common goal. Such an exposure to team-work has encouraged me to appreciate ***collaborative research*** and to acquire the skill of explaining abstract mathematical concepts to non-experts. From bringing out new medicine and equipment for curing ailments to building tools to enhance the convenience of our daily life, I found mathematics in action serving a multitude of purposes. My aspiration to a career as a mathematician was further enlivened when I embarked on a mathematics PhD at Tulane University.

Training. During my course work, I grew a keen interest in ***algebraic topology*** and ***differential and computational geometry***. I was also exposed to the fields of topological data analysis and applied algebraic topology. My research experience debouched into a broader realm when I got involved in different projects for my dissertation under the supervision of ***Carola Wenk*** and ***Brittany Fasy***. The opportunity gradually matured my research experience and perspective of mathematics profoundly. The challenge in applying mathematics to solve a real-world problem is to bring up the right mathematical tool to approach the problem. Unfortunately, the problem at hand does not necessarily inform us about the underlying mathematics of it. The objective becomes two-fold: to develop the ***right formulation*** of the problem and to devise a ***mathematical tool*** that fits the formulation. In such a pursuit, new mathematical theories, like persistent homology, are sometimes bestowed upon the regime of pure mathematics as well. My research interests encompass both the applied side of topological data analysis as well as developing such mathematical theories triggered by applications.

To further my theoretical training beyond the confines of mathematics, I took a good many number of computer science classes: algorithms, computational geometry, and computational complexity. My sheer curiosity about geometric algorithms and geometric data structures eventually shaped my research interest in computational geometry.

Current Research

Geodesic Space Reconstruction via Čech and Rips Complex. Application of algebraic topological methods in reconstruction of graph-like or filamentary structures claims the lion's share of my current research. These projects are primarily motivated by the problem of reconstructing road networks from noisy GPS locations. Road-networks can be modeled as metric graphs embedded in the two-dimensional Euclidean space.

The challenges in developing geometrically close and computationally efficient reconstruction algorithms are ensued from the noise present in the data.

We realize that the available topological methods for shape reconstruction only work for compact Euclidean subsets with positive reach or, at least, positive weak feature size (see [12, 3]). However, our shapes of interest fail to evince such properties. In [5], we introduce the notion of *geodesic feature size* and develop a persistence-based algorithm for computing the 1-Betti number of the underlying of planar metric graph.

In order to further our understanding of this new feature size, we extend our formulation to be able to reconstruct a larger class of Euclidean shapes: geodesic spaces. In [6], we introduce two new sampling parameters: convexity radius and distortion of embedding into the Euclidean ambient. When the point-cloud is sampled around a small Hausdorff proximity of the underlying geodesic space, the Euclidean Čech and Vietoris-Rips complexes of the point-cloud both are used to correctly compute the homotopy (and homology) groups of the underlying shape. When considered the persistent topological features at small scales, we guarantee a correct topological reconstruction. For a geometric reconstruction, in the special case of metric graphs, we further compute a Vietoris-Rips complex of the sample with respect to a different metric, which, unlike the Euclidean metric, embodies the underlying geodesic metric of the shape. Then, the shadow (as defined in [2]) of this complex is shown to be homotopy equivalent to the ground truth. Given the probability of correct reconstruction, we also estimate the smallest sample size.

In order to demonstrate and visualize our theoretical development, I coded a Java Script library, which is hosted on Github (repository link). The library runs on a Web-app (link).

Graph Reconstruction via Discrete Morse Theory. The regime of the reconstruction of road-networks changes considerably in the presence of non-Hausdorff type outliers in the sample. If outliers are far away from the shape of interest, they contribute to an uncontrollably large Hausdorff distance. To this end, I collaborated with Carola Wenk and Brittany Fasy in order to apply a density-based approach to the problem. In [4], the authors successfully used discrete Morse theory (see [8]), coupled with persistence, in the reconstruction of graphs. In their threshold-based approach, a homotopy-type reconstruction is guaranteed, however it failed to produce a geometrically-close reconstruction. In [7], we propose a double threshold-based noise model to overcome such a caveat. Our work on this project is still in progress. We are now particularly working on avoiding the costly persistence computation in the algorithm.

Approximating Gromov-Hausdorff Distance. Along with reconstruction, I am also interested in comparison of shapes. The Gromov-Hausdorff (d_{GH}) distance has been shown (in [11, 10]) to provide a reasonable framework for comparing shapes. Although the authors of [1] show the NP-hardness of approximation of the GH distance between metric trees, computing the distance between Euclidean subsets is still elusive. To address such questions in the Euclidean space, I started a collaborative project with Carola Wenk and Jeffrey Vitter. In our effort to approximate the GH distance for subsets of $\mathbb{R}^d$, we look into its relationship with $d_{H,iso}$, the minimum Hausdorff distance under the class of Euclidean isometries. For $d = 1$, we show that

$$d_{H,iso} \leq \left(1 + \frac{1}{4}\right) d_{GH}.$$

We also show that the bound is tight. This gives rise to an $O(n \log n)$-time algorithm for approximating d_{GH} with an approximation factor of $\left(1 + \frac{1}{4}\right)$; see [9]. Like metric trees, we believe that computing the GH is also NP-hard for Euclidean subsets. The nature of the project is ongoing. Currently, we investigate such hardness questions, along with questions of approximating d_{GH} for $d > 1$.

Prospective Research Projects.

With the success of TDA in analyzing data and solving real-world problems, comes the attention from various other fields. The growing complexity of the problems now demands group effort and expertise beyond a single field. As a student researcher, I have been persistently encouraged to value collaborative research. Sharing ideas with peers, attending conferences and seminars, paying research visits to the experts in the field have eventually become an integral part of my research. In the effort to organize my teams and boost active engagement with collaborators, I found online platforms, like Slack, Github, Overleaf, etc, very useful.

TDA Software. Although, my current research projects mostly concern the foundational aspects of TDA, I am also enthusiastic about applications of TDA and more product-oriented research projects. As a coding hobbyist, I develop tools, libraries, and software to supplement my research. Along with following the standard coding styles and conventions, I also value some of the very important aspects of software engineering, like design patters and code reusability. As the field of TDA is growing apace, a large number of TDA software and libraries are being made available by researchers around the globe. Most of the software are ad-hoc and designed and written by academician, not by software engineers. And more often than not, they lack the industry standards. I look forward to grant proposals from programs, like CDS&E at the National Science Foundation (NSF), to fund, organize, and partake in the development of platforms to house TDA libraries with the special objectives of uniformity, adaptability, reusability, and long-term support.

Grants. For a professional researcher, writing successful grant proposals is a principle skill to have. As a research assistant at Tulane, I was funded from an NSF grant. The grant has fostered my early research career. By providing funding for full-time research, the grant helped me to fully engage in research, reclaiming time from other responsibilities as a teaching assistant. I went to a good many number of workshops and conferences to broaden my academic horizons, expose myself to related research opportunities, and connect with people. Without the travel funding from the grant, such opportunities might not have been otherwise explored. In order to contrive and continue productive collaborations, I will strive to procure grants from programs, like DMS, CISE (Algorithmic Foundations), CAREER, etc, at NSF.

With the theoretical knowledge and training in research as a graduate student, I now seek opportunities to emerge as an independent researcher. Being an energetic researcher with a great passion for teaching and graduating students, I consider myself a good fit for the position. Given the opportunity to thrive as an academician, I ensure to make all the efforts to further my field of research by creating new research collaborations, fetching scholarships and grants, and exposing graduate and undergraduate students to extensive research activities.

[1] Pankaj K. Agarwal, Kyle Fox, Abhinandan Nath, Anastasios Sidiropoulos, and Yusu Wang. Computing the gromov-hausdorff distance for metric trees. *ACM Trans. Algorithms*, 14(2):24:1–24:20, April 2018. URL: http://doi.acm.org/10.1145/3185466, doi:10.1145/3185466.

[2] Erin W. Chambers, Vin de Silva, Jeff Erickson, and Robert Ghrist. Vietoris–Rips complexes of planar point sets. *Discrete & Computational Geometry*, 44(1):75–90, Jul 2010. doi:10.1007/s00454-009-9209-8.

[3] Frédéric Chazal and André Lieutier. Stability and computation of topological invariants of solids in $\mathbb{R}^n$. *Discrete & Computational Geometry*, 37(4):601–617, 2007.

[4] Tamal K. Dey, Jiayuan Wang, and Yusu Wang. Graph reconstruction by discrete Morse theory. In *34th International Symposium on Computational Geometry*, pages 31:1–31:15, 2018.

[5] Brittany Terese Fasy, Rafal Komendaczyk, Sushovan Majhi, and Carola Wenk. Topological reconstruction of metric graphs in $\mathbb{R}^n$. In *Fall Workshop on Computational Geometry*, New York, NY, October 2017. URL: https://arxiv.org/abs/1912.03134.

[6] Brittany Terese Fasy, Rafal Komendarczyk, Sushovan Majhi, and Carola Wenk. On the Reconstruction of Geodesic Subspaces of $\mathbb{R}^n$. *arXiv:1810.10144 [math.AT]*, 2018. URL: https://arxiv.org/abs/1810.10144, arXiv:1810.10144.

[7] Brittany Terese Fasy, Sushovan Majhi, and Carola Wenk. Threshold-based graph reconstruction using discrete Morse theory. In *Fall Workshop on Computational Geometry*, New York, NY, November 2018. URL: https://arxiv.org/abs/1911.12865.

[8] Robin Forman. A user's guide to discrete Morse theory. *Séminaire Lotharingien de Combinatiore*, (48):Art. B48c, 2002.

[9] Sushovan Majhi, Jeffrey Vitter, and Carola Wenk. Approximating Gromov-Hausdorff Distance in Euclidean Space. *arXiv:1912.13008 [math.MG]*, 2019. URL: https://arxiv.org/abs/1912.13008, arXiv:1912.13008.

[10] Facundo Memoli. On the use of Gromov-Hausdorff Distances for Shape Comparison. In M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007. doi:10.2312/SPBG/SPBG07/081-090.

[11] Facundo Mémoli and Guillermo Sapiro. A Theoretical and Computational Framework for Isometry Invariant Recognition of Point Cloud Data. *Foundations of Computational Mathematics*, 5(3):313–347, July 2005. URL: http://link.springer.com/10.1007/s10208-004-0145-y, doi:10.1007/s10208-004-0145-y.

[12] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete And Computational Geometry*, 39. 1-3:419–441, 2008.