SURBHI KUMAR

Dallas, TX | sxk230046@utdallas.edu | linkedin | Personal Website

PROFESSIONAL SUMMARY

Ph.D. Candidate in Mathematics specializing in graph representation learning and topological methods for biomedical applications, including brain network modeling and ligand-based virtual screening. Experienced in building scalable GPU-accelerated machine learning pipelines on high-performance computing clusters, with emphasis on robust evaluation and reproducibility.

EDUCATION

The University of Texas at Dallas — Expected - May 2028
Ph.D. in Mathematics | GPA: 3.75/4

The University of Texas at Dallas — Expected - Dec 2026
M.S. in Mathematics (Data Science Specialization) | GPA: 3.75/4

National Institute of Technology, Rourkela — May 2017
Master of Mathematics | GPA: 7.87/10

Guru Nanak Dev University — June 2015
Bachelor of Science (Majors: Mathematics, Computer Science, Physics) | GPA: 7.9/10

TECHNICAL SKILLS

- **Programming Languages:** Python, C, C++, Java, SQL, R, MATLAB, Mathematica
- **Libraries & Frameworks:** PyTorch, PyTorch Geometric, DGL, TensorFlow, scikit-learn, NetworkX, RDKit, XGBoost, NumPy, SciPy, Gudhi, Ripser
- **Tools & Platforms:** CUDA, Slurm (HPC), Linux, Neo4j, Git, Databricks, Jupyter Notebook, LaTeX, MS Office

PROJECTS

- **TopU-LBVS: A Realistic Multi-Target Benchmark for Ligand-Based Virtual Screening (In Progress):** Constructed the unbiased TopU95 multi-target dataset from CHEMBL35 and established a rigorous benchmarking protocol. Evaluated naïve models, classical machine learning baselines, deep learning graph models, and modern foundation models under a unified framework for fair and reproducible comparison. This debiasing strategy reduces artificial class separation and provides a realistic evaluation setting across diverse protein targets.
- **Same-Graph Cross-Task Transfer in GNNs (Under Review, ICML 2026):** Proposed a leakage-free protocol with fixed node/edge splits and controlled negative sampling for transfer between node classification and link prediction. Showed transfer is directional and predictable from graph properties (e.g., homophily) and introduced CoTask Score to measure shared encoder utility across tasks.
- **Low-Shot Graph Learning with Topological and Spectral Embeddings (Accepted, LOG 2025):** Introduced a prototype-based low-shot graph learning model using Betti vectors and spectral DoS embeddings, outperforming GNN and Graph Transformer models under scarce supervision.

EXPERIENCE

Teaching Assistant — Aug 2023 – Present
The University of Texas at Dallas

- Led problem-solving sessions and assisted with grading for undergraduate calculus courses, supporting large enrollment classes.

Subject Matter Expert - Advanced Mathematics — Jun 2017 - Jul 2023
Ed-Tech Industry (Remote)

- Developed advanced mathematics content for leading Ed-Tech platforms including Chegg and Meritnation.

RESEARCH PUBLICATION

- **Journal Publication:** Kumar, S., Saha Ray, S. Numerical treatment for Burgers–Fisher and generalized Burgers–Fisher equations. Math Sci 15, 21–28 (2021). https://doi.org/10.1007/s40096-020-00356-3