Graz University of Technology

Institute of Geometry

ICC Colloquium - Computational Discrete Mathematics

10.10.2025, 10:45

Hörsaal BE01, Steyrergasse 30

When Topology Meets XAI

Bei Wang Phillips

(University of Utah)

Large language models learn from massive amounts of text, yet the internal structure of their "thinking" remains largely hidden. In this talk, we explore how tools from topological data analysis—specifically mapper graphs—can help uncover patterns within these models' high-dimensional embedding spaces. Mapper graphs reveal how groups of similar representations connect and evolve, much like a map charting neighborhoods and roads in an unfamiliar city. To make these complex structures interpretable, we introduce the Explainable Mapper, an interactive workspace equipped with intelligent mapper agents. These agents can summarize clusters, compare regions, and probe how small perturbations affect the structure, shedding light on the kinds of linguistic information the model has learned and how ideas flow between different parts of its hidden space. This approach brings us closer to explaining how large language models organize knowledge—transforming opaque black boxes into navigable landscapes.

Michael Kerber