


复旦大学数学科学学院
数学综合报告会

报告题目：DAG-Math: Graph-Guided Mathematical Reasoning in LLMs

报告人：刘方辉 (上海交通大学)

时间：2025-12-22 星期一 14:00-15:00

地点：光华楼东主楼 2001

报告摘要：

Large Language Models (LLMs) demonstrate strong performance on mathematical problems when prompted with Chain-of-Thought (CoT), yet it remains unclear whether this success stems from search, rote procedures, or rule-consistent reasoning. In this talk, I will discuss about modeling CoT as a certain rule-based stochastic process over directed acyclic graphs (DAGs), where nodes represent intermediate derivation states and edges encode rule applications. Within this framework, we introduce logical closeness, a metric that quantifies how well a model's CoT trajectory (i.e., the LLM's final output) adheres to the DAG structure, providing evaluation beyond classical PASS@k metrics. Building on this, we introduce the DAG-MATH CoT format and construct a benchmark that guides LLMs to generate CoT trajectories in this format, thereby enabling the evaluation of their reasoning ability under our framework. Across standard mathematical reasoning datasets, our analysis uncovers statistically significant differences in reasoning fidelity among representative LLM families-even when PASS@k is comparable-highlighting gaps between final-answer accuracy and rule-consistent derivation. Our framework provides a balance between free-form CoT and formal proofs systems, offering actionable diagnostics for LLMs reasoning evaluation. Talk is based on https://arxiv.org/abs/2510.19842, joint work with Yuanhe Zhang (Warwick), Ilja Kuzborskij (Google DeepMind), Jason D. Lee (UC Berkeley) Chenlei Leng (PolyU HK).