Accurate graph inference from images and documents: Datasets and challenges

Rahul Sakar

Many of the mathematical literature in topology involves commutative diagrams, a class of graphs with labeled nodes, edges, and multiple edge types. The literature on this topic consists of published books and research papers that need to be digitized to train AI models capable of mathematical reasoning. A major bottleneck to their digitization is the lack of automated tools that can extract graph structures accurately from such media. To address this gap and with the long-term goal of building AI systems capable of accurately reasoning about topological questions in mathematics, we present our ongoing work on creating a visual graph understanding dataset benchmark for open-source vision large language models (vLLMs). The benchmark assesses fine-grained graph understanding skills, such as node and edge detection, common neighbor identification, and clique recognition using skill-specific prompts and verification metrics (e.g., graph edit distance for node and edge detection) of open-source vLLMs. The dataset covers diverse graph domains and styles, including commutative diagrams, maps, molecular graphs, and knowledge graphs. We evaluate the state of the art open source vLLMs on these specific tasks, along with a few baseline methods. Developing better algorithms for this task extends to other domains involving graph extraction in computer science and other allied disciplines.