Final Report: Enhancing Graph Recognition with Multimodal AI

Dr. Mina Samizadeh
Department of Computer Science
Lincoln University

Faculty Research and Development
Summer 2025

Project Overview

This project investigated how large multimodal foundation models can be adapted to perform specialized graph-theoretic reasoning from visual inputs. While modern vision–language models can visually perceive graph structures, they often lack the domain-specific knowledge required to accurately identify complex graph properties. The goal of this work was to evaluate whether targeted fine-tuning could significantly improve model performance on such tasks.

Specifically, this project focused on training a multimodal model to recognize graph properties directly from images, including Hamiltonicity, tree structure, planarity, and claw-freeness. The work builds on the original funded proposal, which aimed to advance AI-driven mathematical reasoning through multimodal learning.

Graph Properties Studied

The project addressed several fundamental graph-theoretic properties:

- **Hamiltonicity**: Whether a graph contains a path (or cycle) that visits each vertex exactly once. This is a classic NP-complete problem in graph theory.
- **Trees**: Connected graphs with no cycles, characterized by unique paths between all vertex pairs.
- **Planarity**: Whether a graph can be drawn in the plane without edge crossings.
- **Claw-Free Graphs**: Graphs that do not contain a $K_{1,3}$ (claw) as an induced subgraph.

These problems are inherently challenging, as exact solutions typically require explicit mathematical computation and structural analysis over graph adjacency matrices. An emerging approach converts graph adjacency structures into visual representations and performs reasoning directly over images; however, existing multimodal foundation models generally lack the capability to reliably detect specific graph properties. To the best of my knowledge, this work represents the first efforts to explicitly teach a foundation model to perform graph property detection through targeted multimodal fine-tuning.

Methodology

The project followed a structured experimental workflow:

Infrastructure and Setup

Google Cloud Platform services were configured using a dedicated service account. Authentication, cloud storage, and Vertex AI services were verified, and cloud-based computation was enabled to support model fine-tuning.

Dataset Preparation

A dataset of graph images paired with accurate textual descriptions of their properties was curated. The data were converted into the JSONL format required for multimodal fine-tuning and uploaded to Google Cloud Storage.

Figure 1 shows an example graph image from the dataset along with its corresponding JSON caption used for training.

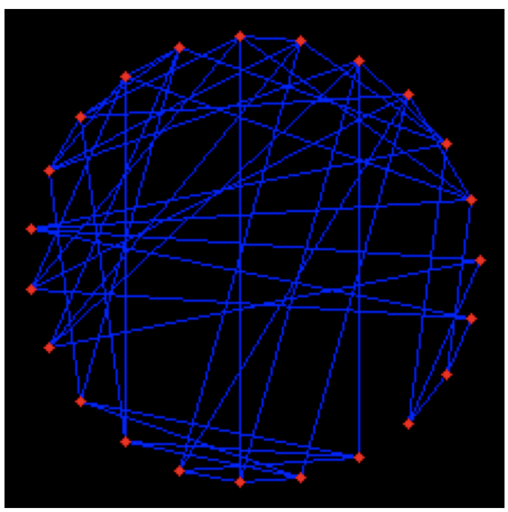

Figure 1: Example graph image used in the training dataset. Nodes are represented by red markers and edges by blue lines.

The corresponding caption and label information for this image were encoded in JSON format as follows:


```
{
    "description": "The image shows a graph with 20 nodes arranged in a circle.
```

```
    Each node is represented by a red diamond. The nodes are connected
    by blue lines, forming a complex network of edges. The graph appears
    to be highly connected, but it is difficult to determine the degree
    of each node by visual inspection.",
    "is_hamiltonian": true
}
```

Baseline Evaluation

The base multimodal model (Gemini 2.0 Flash) was evaluated using structured prompts that required JSON-formatted outputs. Initial testing revealed a significant limitation: while Hamiltonian graphs were correctly identified, non-Hamiltonian graphs were consistently misclassified as Hamiltonian. This confirmed that the base model lacked specialized graph-theoretic reasoning capabilities.

Model Fine-Tuning

The base multimodal model (Gemini 2.0 Flash) was fine-tuned using Google Cloud Vertex AI on the curated dataset of graph images and property annotations. The fine-tuning process employed supervised multimodal learning, where each training instance paired a graph image with a structured textual description and corresponding property labels.

Training hyperparameters were configured within Vertex AI, including a learning rate multiplier of 1.0 and an adapter size of 1, and the tuning job was executed using cloud-based computational resources. Throughout training, model progress and system logs were monitored via the Vertex AI platform to ensure stable convergence and successful job completion.

The resulting fine-tuned model was specifically adapted to analyze graph images and to infer structural properties such as Hamiltonicity, planarity, tree structure, and claw-freeness, demonstrating improved domain-specific reasoning compared to the base model.

Results and Findings

After fine-tuning, the model was evaluated on the same test set used in the baseline experiments. The fine-tuned model correctly identified both Hamiltonian and non-Hamiltonian graphs, achieving a substantial improvement in accuracy.

Key outcomes include:

- **Improved Accuracy**: Classification accuracy improved from 50% with the base model to 100% on the evaluation set for the Hamiltonian Cycle problem.
- **Domain Adaptation**: Fine-tuning enabled the model to capture graph-theoretic regularities from the multimodal data distribution, learning associations between visual graph structures and their corresponding textual property descriptions.

- **Feasibility for NP-Complete Problems**: The results demonstrate the potential of fine-tuned foundation models for AI-assisted reasoning on computationally difficult graph problems.

Technologies Utilized

The project leveraged the following technologies:

- Google Cloud Platform for authentication, storage, and computation
- Vertex AI for model fine-tuning and evaluation
- Python for data processing and experimentation
- Pandas and Matplotlib for analysis and visualization

Conclusion

This project demonstrates that general-purpose multimodal foundation models can be effectively adapted to perform specialized mathematical reasoning tasks through targeted fine-tuning. By supplying graph images with accurate property annotations, a model that initially failed to reason about Hamiltonicity was transformed into one capable of reliable graph property detection.

The findings support the broader goal of developing AI systems that can reason about mathematically constrained visual data and provide a foundation for future work in AI-driven graph analysis and visualization.

Dissemination and Future Work

The results of this project are being prepared for submission to the *Workshop on Emerging Directions in Data for Multimodal Foundation Models* at CVPR 2026. Future work will focus on expanding the dataset, extending the approach to additional NP-complete graph problems, and developing interactive tools to visualize and explain model reasoning.