

Update-1: TDL Project

Team XYZ
Himanshu Pal
Pranav Gupta
Anantha Muppidi

Overview

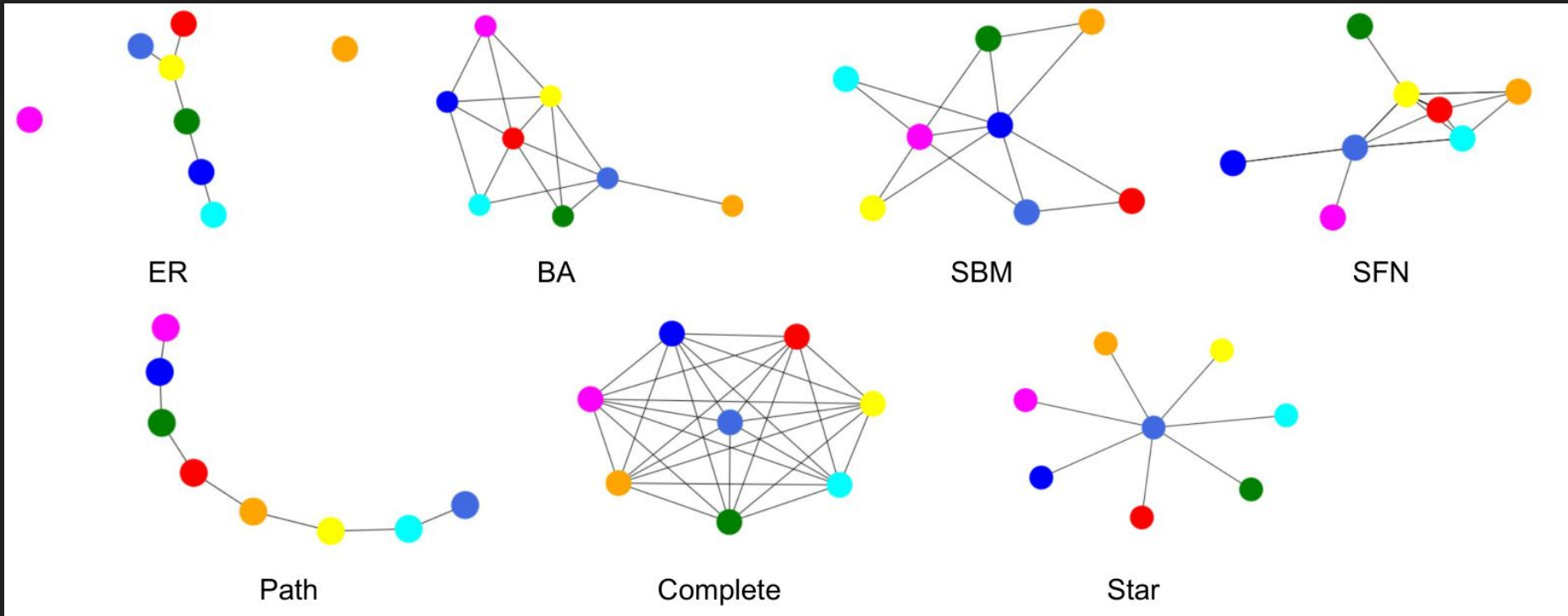
- Explore the untapped potential of large language models (LLMs) in reasoning on graph-structured data.
- Investigate the impact of various graph encoding methods on LLM performance in graph reasoning tasks.
- Explored different prompting methods for improving graph reasoning for LLM methods

Dataset Creation

- The GraphQA dataset is not publicly available.
- We are following the exact details mentioned in the paper to create the dataset.
- We are using “*NetworkX*” modules to create different graphs mentioned in the paper

Dataset Creation

Types of graphs covered in the study:



Dataset Creation






- 500 graphs for each of the following algorithms:
 - ER
 - BA
 - SFN
 - SBM
- Due to less variation in structural representation we are creating 100 graphs for:
 - Path
 - Complete
 - Star
- All graphs have between 5 to 20 nodes.
- For ER graph the probability for edge creation is $[0, 1]$
- For SBM graphs, number of communities has been sampled from 2 to 10

LLM Model

We have finalized the following compute friendly LLMs to reason different graph encodings.

- LLAMAv2-7b
- PaLM 2 XXS
- PaLM 2 XS
- GPT-4

Work Update

- Graph Dataset
 - Graph Encoding Finalization 
 - Prompting Techniques 
 - QA creation 
- LLM model trial
 - LLAMAv2-7b 
 - PaLM 2 XXS
 - PaLM 2 XS
 - GPT-4 

Next Steps

- **Prepare Pipeline:** Set up the pipeline for testing the models with our version of the GraphQA dataset. This involves organizing the data, loading the models, and defining the evaluation metrics.
- **Preprocess Data:** Before testing the models, preprocess the GraphQA dataset. This may include tasks like tokenization, encoding, and any necessary formatting to make the data compatible with the models.
- **Implement Novel Prompting Techniques:** Develop and integrate innovative prompting techniques to gain insights into the global structure of the graphs. This could involve designing new prompting strategies or refining existing ones to better understand the overall graph structure.
- **Evaluate Models:** Finally, run the models on the prepared dataset using the novel prompting techniques. Evaluate their performance and analyze the results to gain insights into how well they understand the global structure of the graphs in your dataset.