

Stage I: Dataset Construction

Real-World Scenario



Tasks

Enumeration

- ◆ Max Degree
- ◆ Triangle
- ◆ Maximum Clique

Exploration

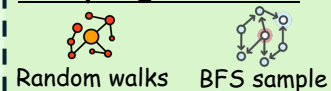
- ◆ Distance-K
- ◆ PathSum
- ◆ Diameter

Intuition

- ◆ MST
- ◆ Maximum K-core
- ◆ DistanceThreshold

Graphs

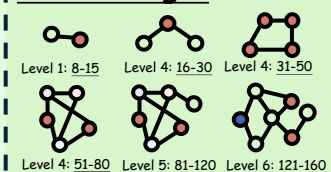
Sampling Methods



Graph Type

Undirected, Weighted, Tree

Node Range:



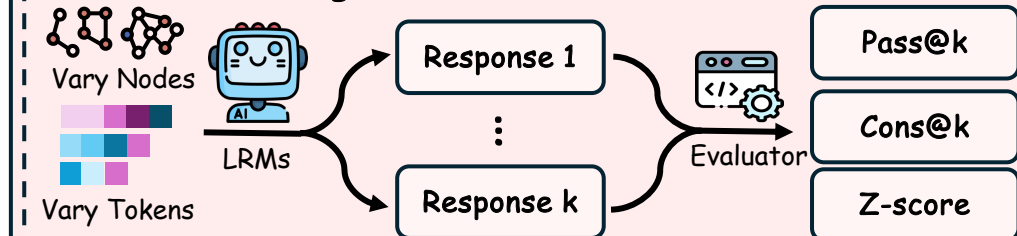
Dataset Construction Pipeline



Stage II: Evaluation

RQ1: Do LRMs excel at long-context reasoning / memory?

Inference in Long contexts



Error Analysis

Algorithm Selection

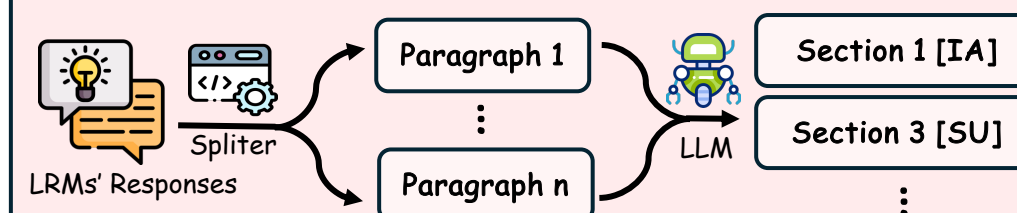
- Incorrect Algorithm
- Suboptimal Algorithm

Algorithm Execution

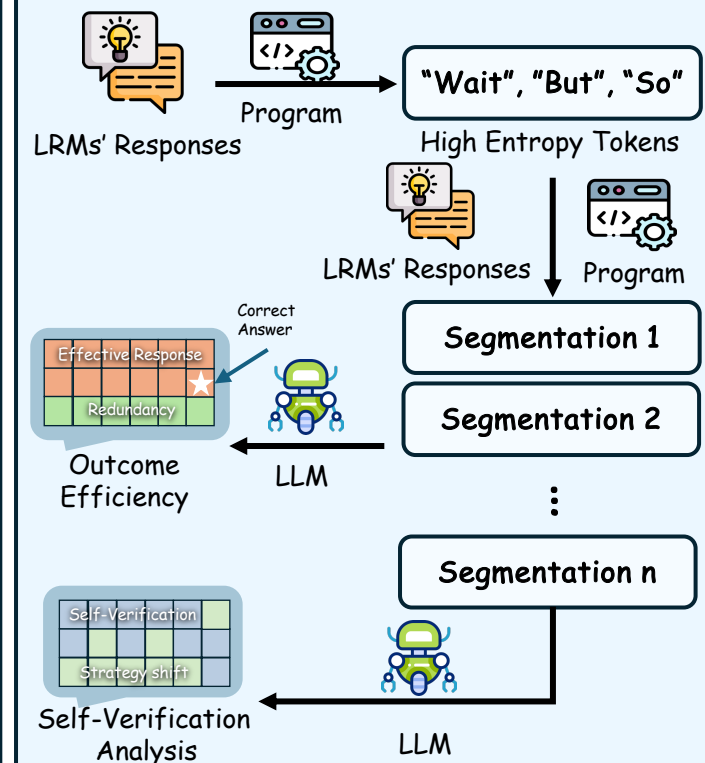
- State Update
- Missing Element
- Condition Misjudgement

Output Quality

- Redundancy
- Information
- Graph Memory



RQ2: What leads to overthinking of LRMs when solving graph algorithm problems?



Reasoning Models



Non-Reasoning Models

