Naive Retrieval

Problem ▭ Motivation ▭ Mention of Methodology ▭

Train Retriever

Web Corpora

Citation Graph

Query | S |
Title: BERT-of-Theseus: Compressing BERT by Progressive Module Replacing
Abstract: ...model compression approach to effectively compress BERT... Our approach first divides the original BERT into several modules and builds their compact substitutes. Then, we randomly replace the original modules... deeper level of interaction between the original and compact models... ...does not introduce any additional loss function.

Seed Paper

Ranked Retrieval

1 **Title:** A Language Model based Evaluator for Sentence Compression
Abstract: ... We here in present a language-model based evaluator for deletion-based sentence compression, and viewed this task as a series of deletion-and-evaluation operations using the ...

2 **Title:** DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter
Abstract: ...a method to pre-train a smaller general purpose language representation model ...introduce a triple loss combining language modeling, distillation and cosine-distance losses..

...

Corpus

S

Non-Methodology Relevant Paper ▭

Methodology Inspiration Retrieval

Extract **Problem** & **Motivation**

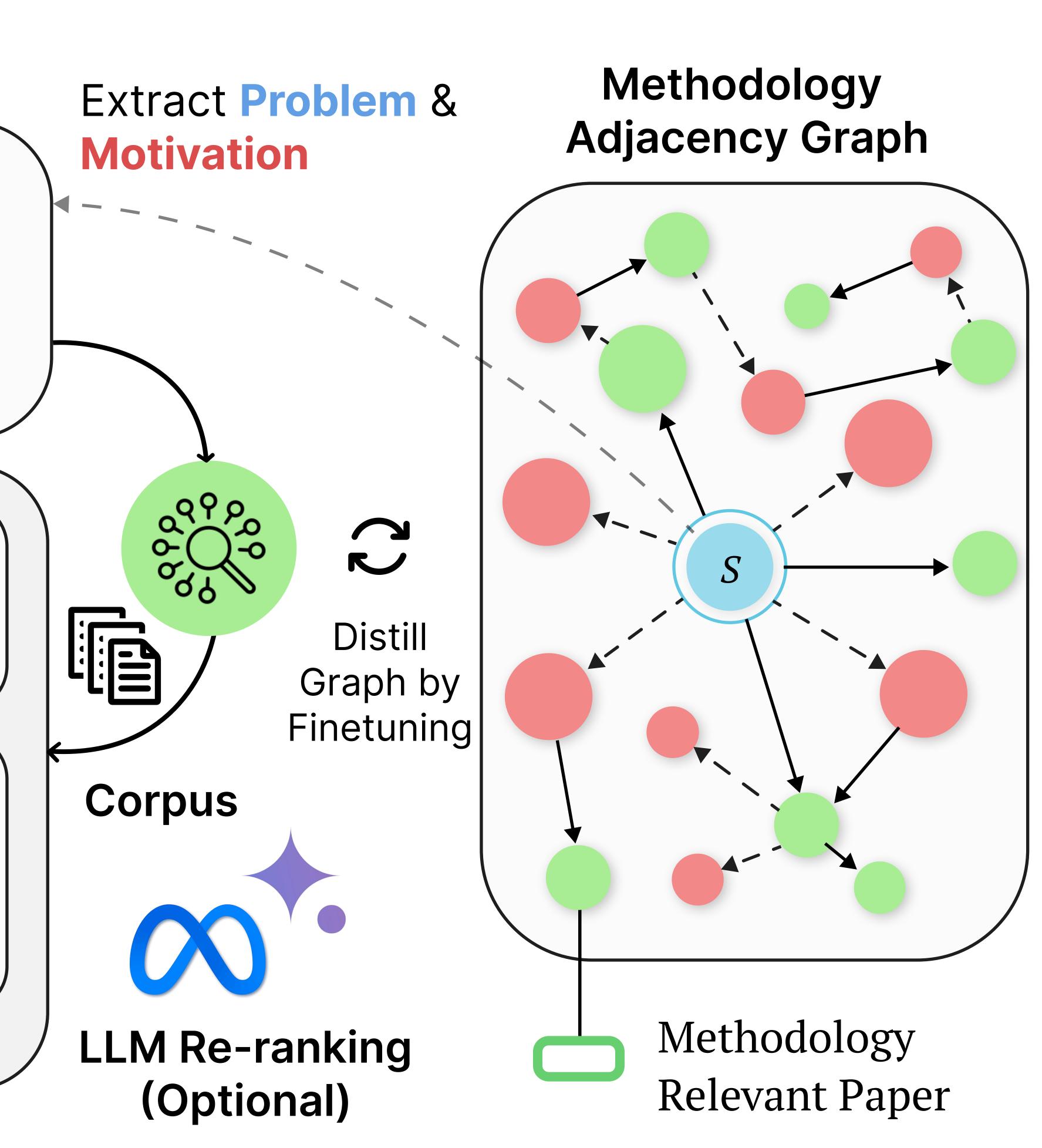**Methodology Adjacency Graph**

Query | S |
Proposal: The research aims to effectively compress BERT, a large language model, while maintaining its performance. The motivation is to improve the efficiency of BERT by reducing its size without sacrificing performance, improve existing knowledge distillation approaches by introducing deeper level of interaction between the original and compact models.

Ranked Retrieval

1 **Title:** DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter
Abstract: ...a method to pre-train a smaller general purpose language representation model ...introduce a triple loss combining language modeling, distillation and cosine-distance losses..

2 **Title:** Patient Knowledge Distillation for BERT Model Compression
Abstract: ...alleviate this resource hunger in large-scale ...high demand for computing resources in training such models hinders... ...patient distillation schemes enable the exploitation of rich information in the teacher's hidden layers...

...

Corpus

Distill Graph by Finetuning

S

LLM Re-ranking (Optional)

Methodology Relevant Paper ▭