Explanation of Revisions

This document explains the changes made to our paper ”RECIPE-TKG: From Sparse History to Structured Reasoning for LLM-based Temporal Knowledge Graph Completion” since the previous submission. Our revisions address several core concerns that led to the initial rejection while maintaining the paper's central contributions and claims.

Expanded Evaluation and Baselines We have substantially expanded our experimental comparison by including four additional recent LLM-based methods for temporal knowledge graph completion: PPT (ACL 2023), CoH (ACL 2024), and HFL (IPM 2025). These additions demonstrate that RECIPE-TKG outperforms contemporary approaches across multiple datasets. The expanded comparison provides a more comprehensive evaluation of our method within the rapidly evolving landscape of LLM-based temporal reasoning. We also clarified why certain embedding-based models use different dataset splits and preprocessing steps, making direct comparison with our LLM-based approach potentially misleading.

Quantification of Framework Efficiency We have added concrete metrics to support our claim that RECIPE-TKG is a ”lightweight” framework. The paper now quantifies that we fine-tune only 0.81% of the parameters in the base LLM (54.3M out of 6.74B for LLaMA2-7B). We also demonstrate that our contrastive fine-tuning introduces negligible training overhead compared to standard fine-tuning (actually showing a slight 0.34% decrease in training time), while our test-time filtering adds only 16.6% inference overhead while providing substantial performance improvements. Additionally, we measured the rule mining efficiency across datasets, showing that this process completes in under 20 seconds even for the largest dataset, representing negligible computational overhead.

Cross-Dataset Generalization Analysis To address concerns about the broader applicability of our approach, we conducted additional experiments demonstrating the effectiveness of our similarity-based filtering mechanism across all four benchmark datasets. The results show consistent improvements in Hits@10, with gains ranging from 7.1 to 9.4 percentage points across datasets. We also analyzed the sensitivity of our method to threshold settings, showing robust performance across reasonable parameter variations. These analyses confirm that the benefits of our approach are not dataset-specific but reflect general principles of semantic consistency that transfer across different temporal knowledge domains.

Clarified Contributions and Improved Presentation We refined several sections of the paper to improve clarity and readability. This includes streamlining the conclusion and ablation study sections, improving table layouts, and adding clearer explanations of our methodological choices. We've also enhanced the discussion of how our approach differs from and improves upon prior work, particularly in handling sparse historical contexts where previous methods struggle.

These revisions maintain the original contributions of the paper while providing stronger empirical support and clearer explanations for our claims. The improvements address the primary concerns that led to rejection in the previous review cycle, making the case for RECIPE-TKG's effectiveness as an approach for temporal knowledge graph completion using large language models.