Impact of Knowledge Graph-Based Relationship Tracking on Agent Social Decision Making

CodeScientist

February 26, 2025

Abstract

This study investigates whether explicit relationship tracking using knowledge graphs improves the quality of social decisions made by large language model (LLM) agents. We implemented and compared three agent variants: one using a dynamic relationship graph, one without relationship tracking, and one assuming static friendly relationships. Through evaluation of 24 social decisions across five scenarios, we found that graph-based relationship tracking did not improve decision appropriateness scores compared to baseline approaches. This suggests that current implementations of explicit relationship modeling may not provide advantages over the implicit social reasoning capabilities of modern LLMs.

1 Introduction

Social decision-making requires understanding and tracking relationships between individuals. While large language models demonstrate strong social reasoning capabilities, it remains unclear whether explicit relationship tracking through knowledge graphs can enhance their social decision-making. This study tests the hypothesis that maintaining and consulting a relationship graph improves the appropriateness of social decisions made by LLM agents.

2 Methods

2.1 Experimental Design

We implemented three agent variants using GPT-4:

- Experimental Agent: Maintained a dynamic relationship graph tracking friend/neutral/enemy relationships between characters
- Baseline Agent 1: Made decisions without relationship tracking
- Baseline Agent 2: Assumed all relationships were friendly

The agents were tested on 5 social scenarios (pilot mode) involving 3 characters each, with 5 decisions per scenario. Scenarios included group projects, family vacations, startup dynamics, restaurant staff interactions, and band member coordination.

2.2 Implementation

The relationship graph was implemented using DOT/Graphviz, with:

- Nodes representing characters
- Edges representing relationships (friend/neutral/enemy)
- Color-coding (green=friend, yellow=neutral, red=enemy)

For each decision point, agents received the scenario context and made decisions that were evaluated by an independent LLM evaluator on a 1-5 appropriateness scale.

3 Results

Table 1: Mean Appropriateness Scores by Agent Type

Agent Type	Mean Score
Experimental (Graph-based)	4.29
Baseline 1 (No Graph)	4.83
Baseline 2 (Static Friendly)	4.83

Bootstrap analysis comparing the experimental agent against each baseline (10,000 resamples) showed:

- Experimental vs. Baseline 1: $p = 1.0$, mean difference = -0.54
- Experimental vs. Baseline 2: $p = 1.0$, mean difference = -0.54

Contrary to our hypothesis, the graph-based agent achieved lower mean appropriateness scores (4.29) compared to both baselines (4.83). The high p-values indicate this difference was not statistically significant in the negative direction.

4 Discussion

4.1 Key Findings

The results do not support the hypothesis that explicit relationship tracking improves agent social decision-making. In fact, the graph-based agent performed marginally worse than both baselines, though not significantly. This suggests that:

- Modern LLMs may have sufficient implicit social reasoning capabilities
- Explicit relationship tracking might add unnecessary complexity
- The implementation of graph-based reasoning may need refinement

4.2 Limitations

Several limitations should be considered:

- Small sample size (24 decisions total)
- Limited scenario complexity
- Potential evaluation bias from using LLM-based scoring
- No temporal relationship dynamics were tested
- Single relationship graph implementation tested

4.3 Implementation Fidelity

The experiment closely followed the requested design, implementing all core components:

- Three distinct agent types
- DOT/Graphviz relationship visualization
- Structured evaluation procedure

- Bootstrap statistical analysis
- Comprehensive logging

However, the experiment stopped at the pilot phase (5 scenarios) rather than proceeding to the full experiment (20 scenarios).

5 Conclusion

While the experiment was implemented faithfully, results suggest that explicit relationship tracking through knowledge graphs may not improve agent social decision-making compared to baseline approaches. Future work should explore more sophisticated relationship modeling techniques, larger-scale evaluation, and alternative graph implementations.