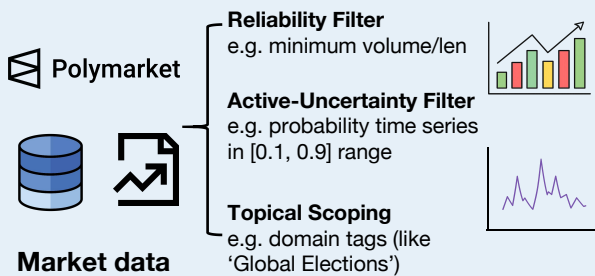
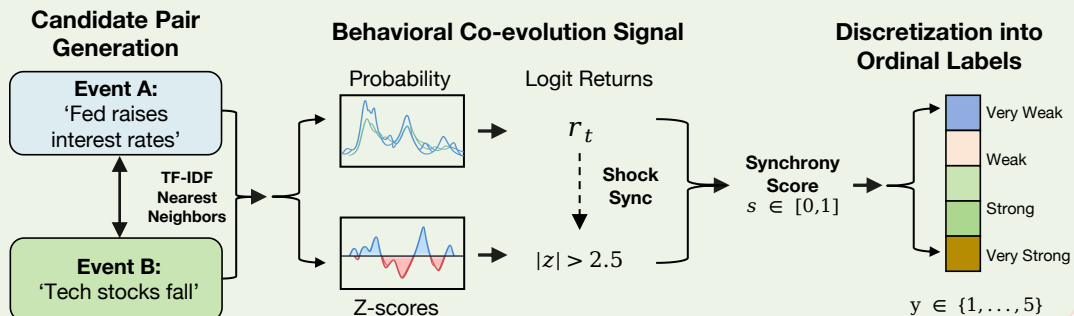


Benchmarking Foundation

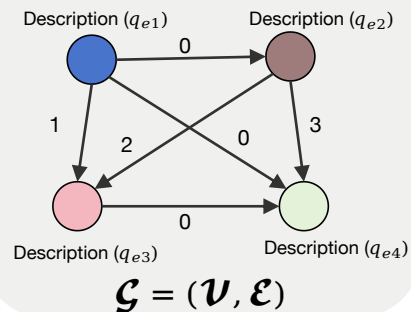
1. Event Node Collection



2. Edge Construction Pipeline

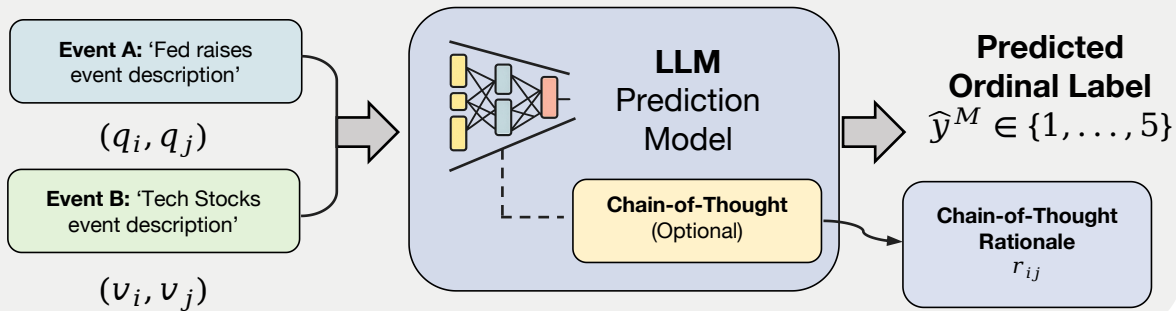


3. Opinion Graph

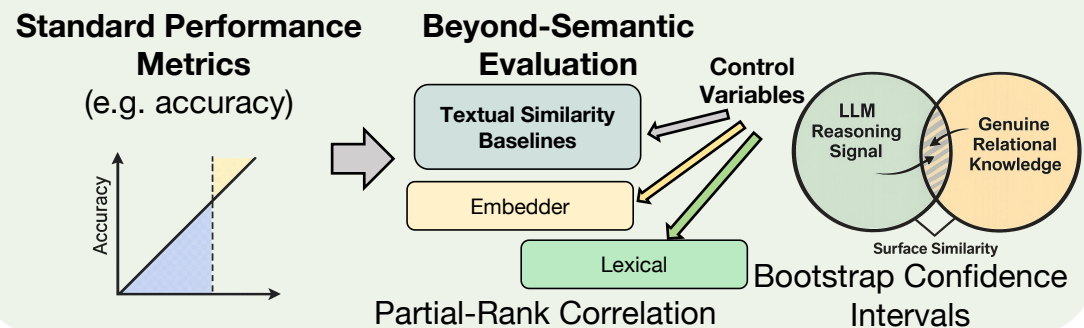


Evaluation Model Reasoning

Edge-Label Prediction Task

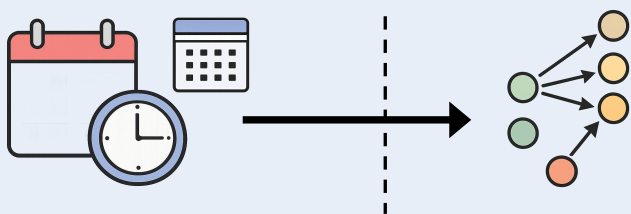


Beyond-Semantic Analysis



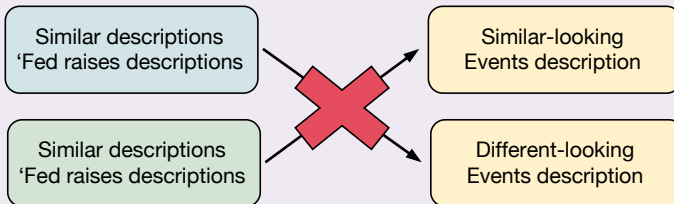
Core Contributions and Insight

Refreshes with Post-Cutoff Data



Mitigate training-set contamination by refreshing with post-cutoff data

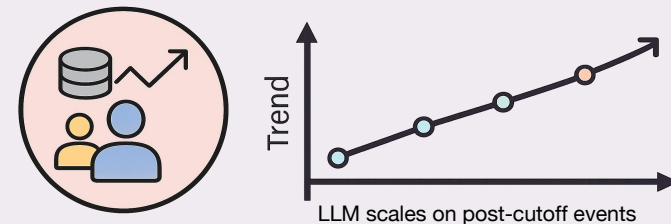
Behavioral Label, Not Text



High textual similarity but low co-evolution

Low similarity high co-evolution

Evidence of Event Co-evolution Knowledge



LLMs encode event-relational signal beyond text, with a clear within-family scaling pattern