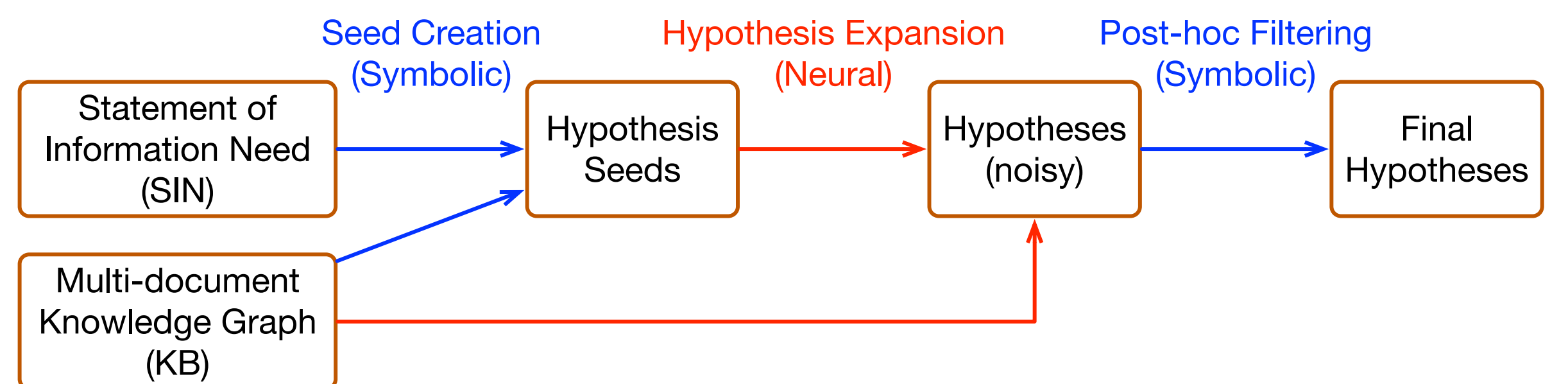


The UTexas system for TAC 2019 SM-KBP Task 3: Hypothesis detection with graph convolutional networks

Pengxiang Cheng, Alex Tomkovich, Eric Holgate, Su Wang, Katrin Erk
{pxcheng, alexander.tomkovich, holgate, shrekwang, katrin.erk}@utexas.edu

Overview

- **SM-KBP Task 3:** generation of hypotheses on emerging topics from a multi-document knowledge graph (the output of Task 2, which usually contains conflicting claims).
- **The UTexas system:** a hybrid neural-symbolic system.



Hypothesis Seed Creation



- Entry point matching is hard on TA2 KBs (some entry points cannot be matched in any of 15 TA2 KBs).
- We re-rank all hypothesis seeds to encourage diversity of entry point fillers.

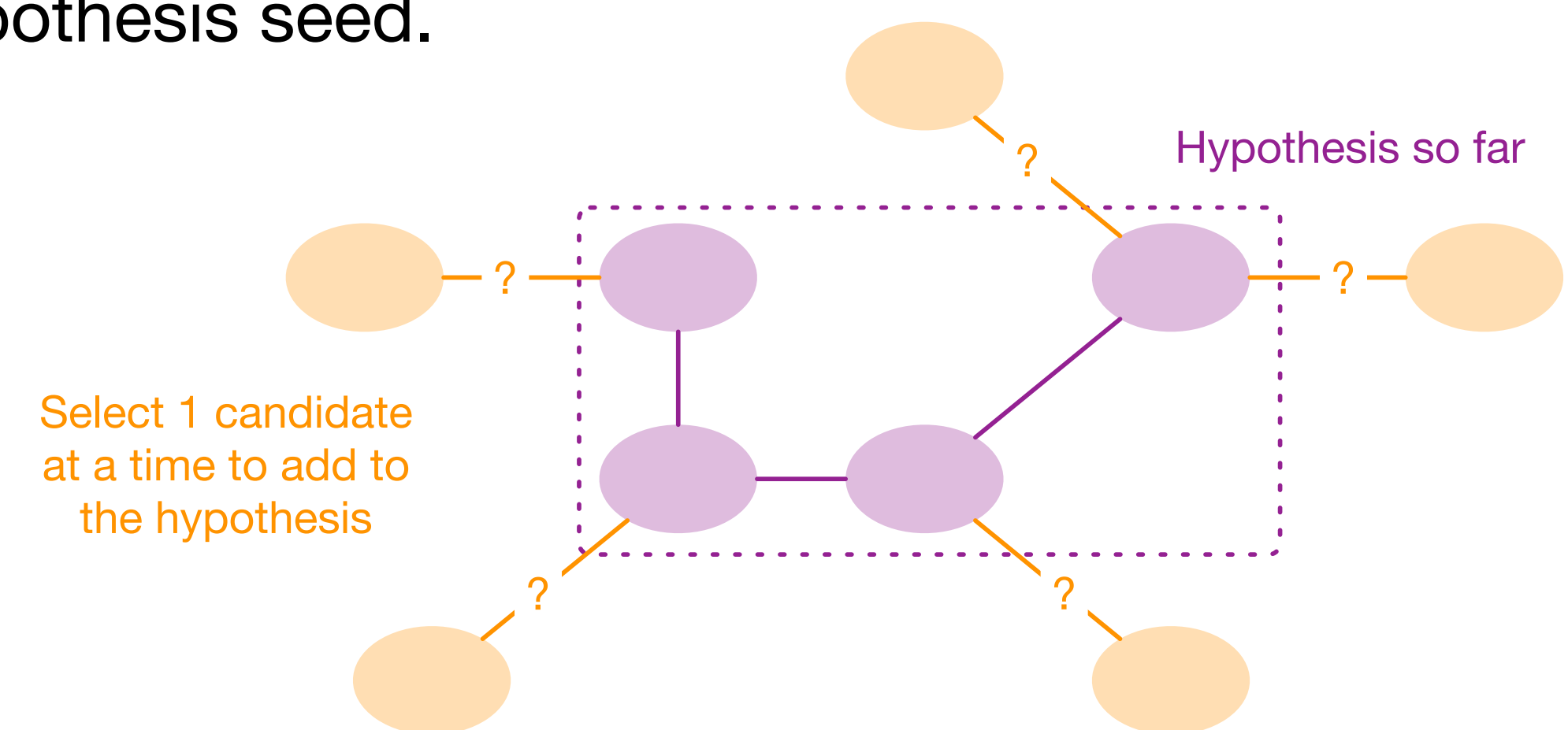
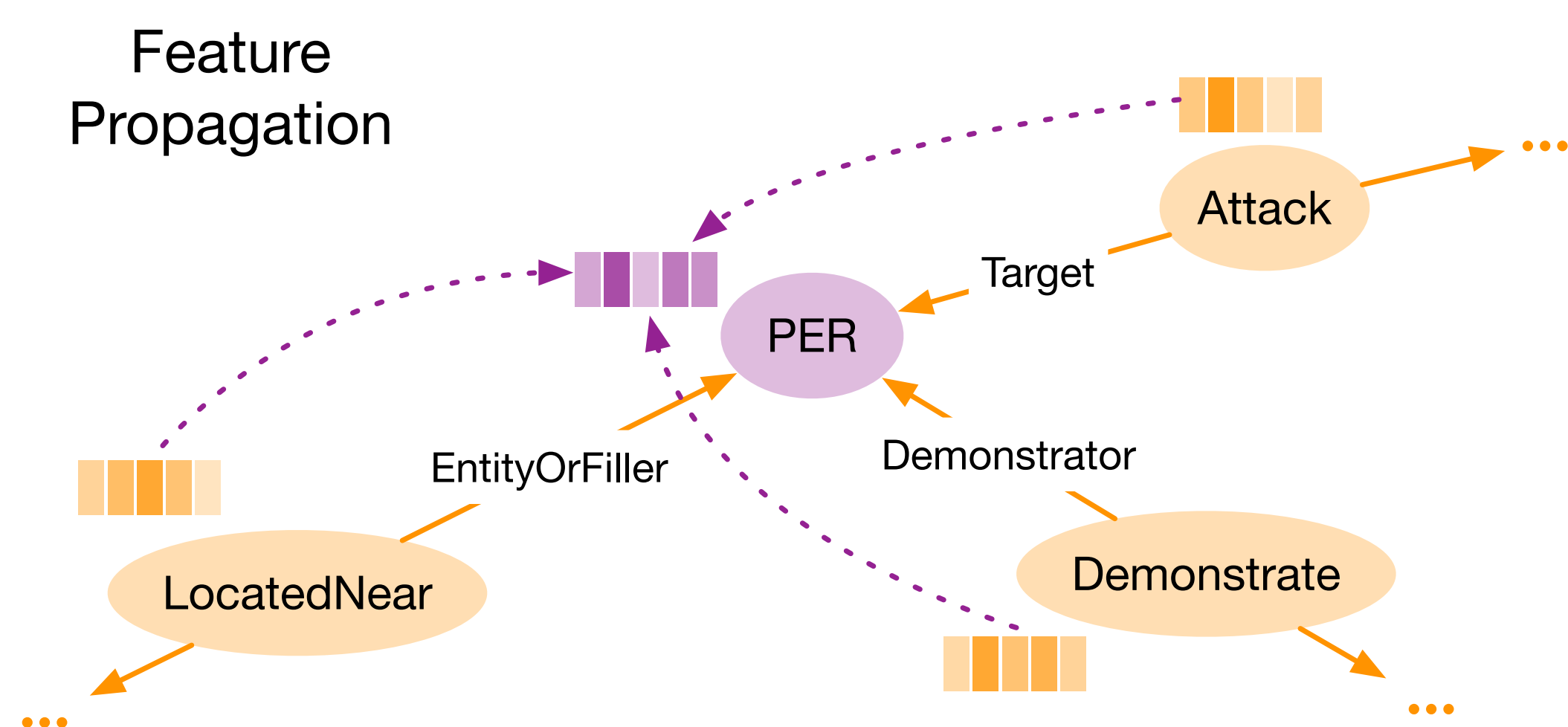
Post-hoc Filtering

- General: a relation node must have two arguments.
- Domain-specific: an attacker cannot attack himself.

% of statements removed	E101	E102	E103
LDC KB	5%	50%	1%
TA2 KBs	24%	55%	39%

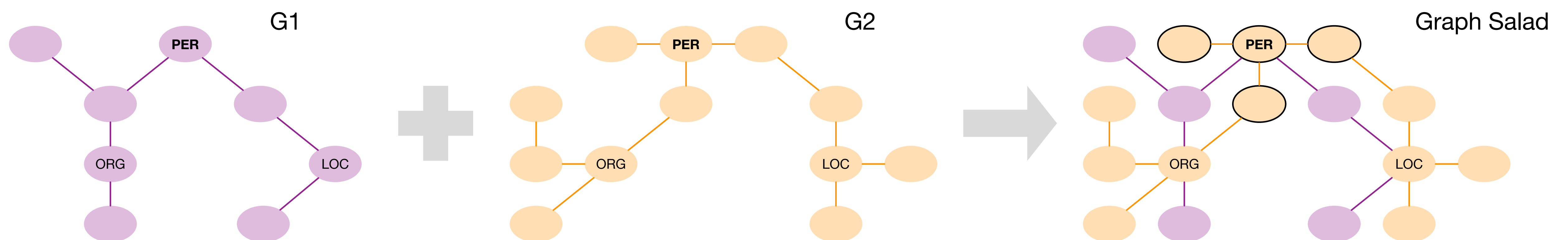
Hypothesis Expansion

- We first encode the knowledge graph with a 2-layer graph convolutional network (GCN).
- Then we iteratively add statements to the hypothesis, one at a time, via attention computation, starting from the hypothesis seed.



Training Data

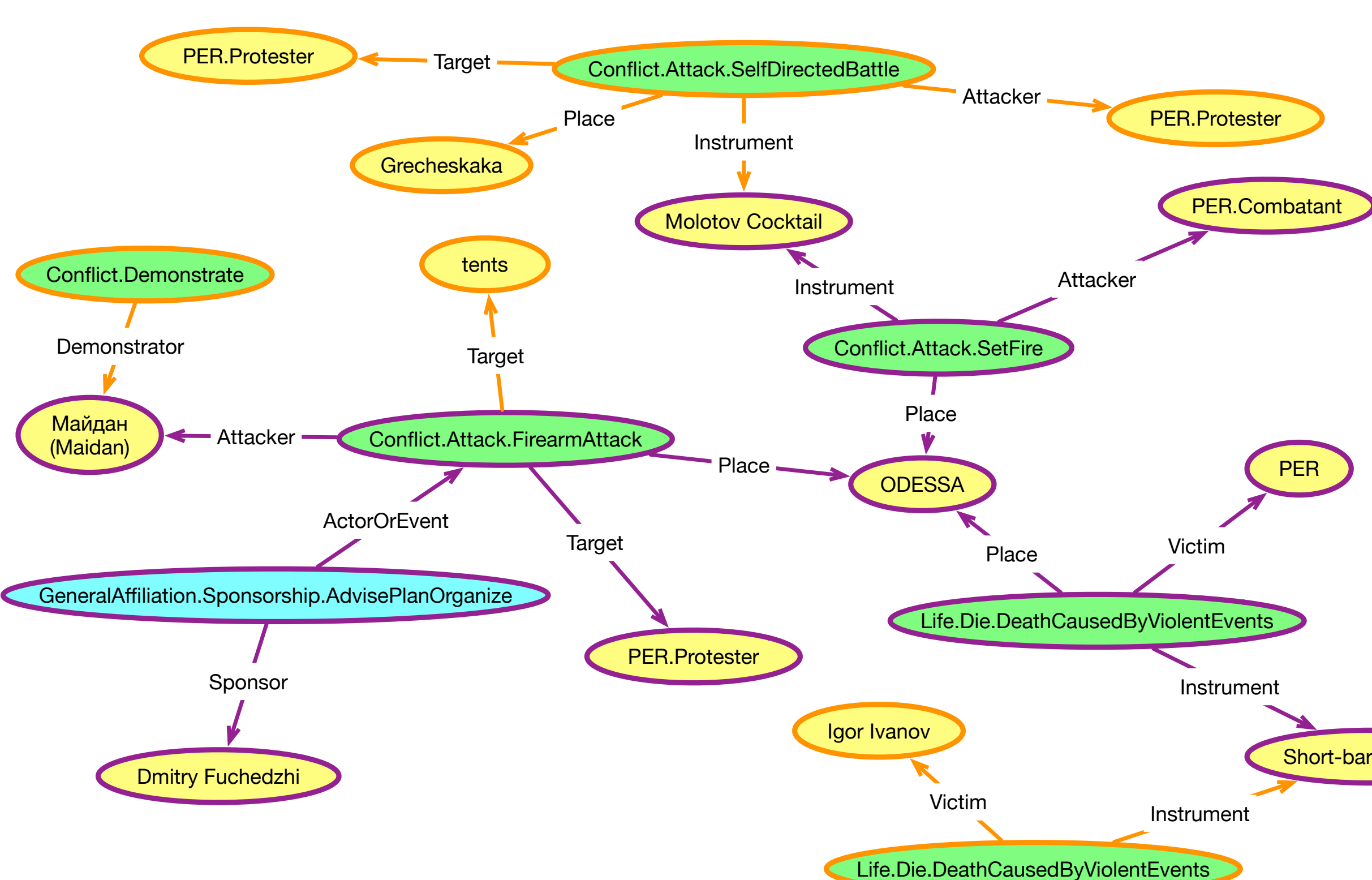
- There is no in-domain training data available, so we create synthetic "graph salads" for training.



- We collect 72k Wikipedia articles on wars and conflicts, and 1,893 unsequestered documents from AIDA M09, and create 100k "Wiki-Wiki" mixtures and 20k "Wiki-AIDA" mixtures.

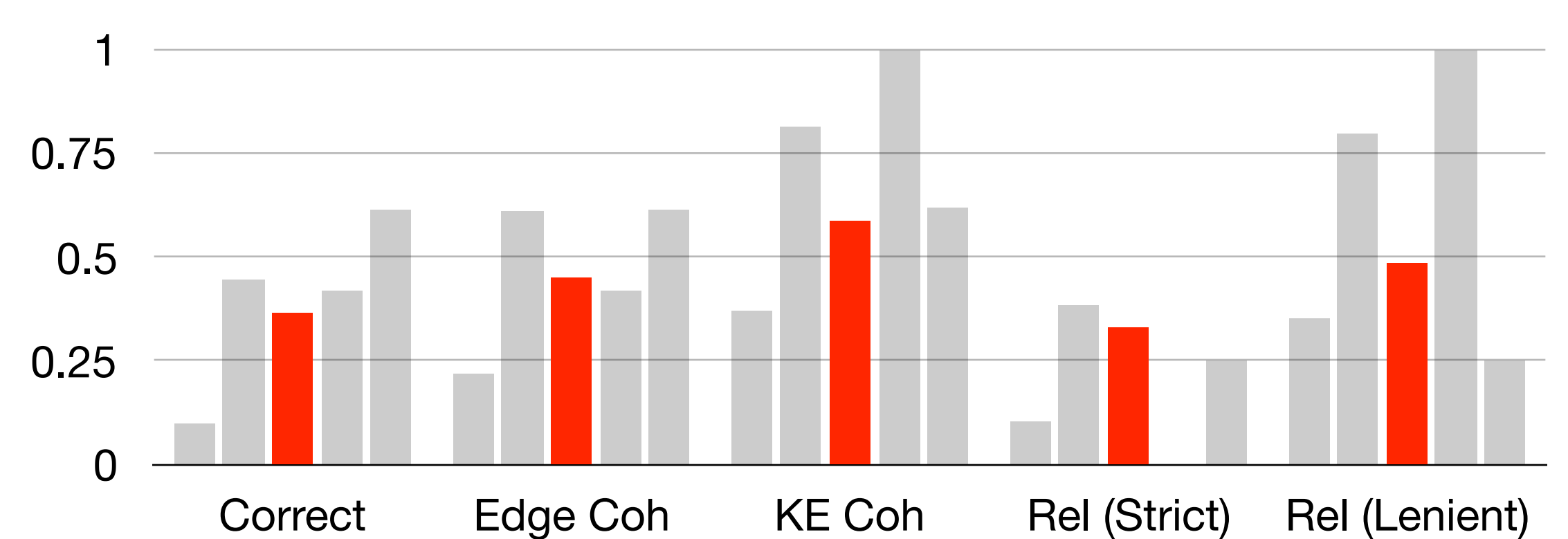
Example

An example hypothesis on E102 generated from LDC KB. (Purple nodes/edges from seed, orange from expansion.)



Results

Performance on TA2 KBs (among 5 TA3 teams)



Performance on LDC KB (among 5 TA3 teams)

