W1. **"Figure 1 could be potentially improved for clarity and readability. Specifically, it would be nice to make the figure self-contained and self-explanatory by adding more illustrations and/or more descriptions in the figure caption."**

Figure 1

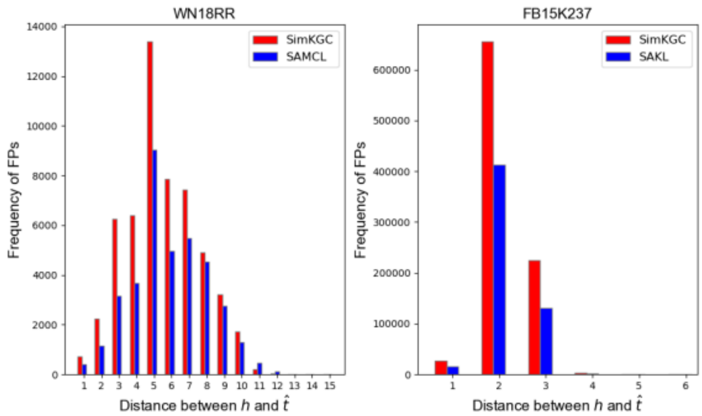

Figure 1: Number of false positives (FPs) based on the distance (i.e., length of shortest path) between a positive head and a false positive tail. The gap in this number between SimKGC and SAMCL is large, especially for small distances (e.g., 2-5 hops in WN18RR, 2-3 hops in FB15k-237).

Figure 2

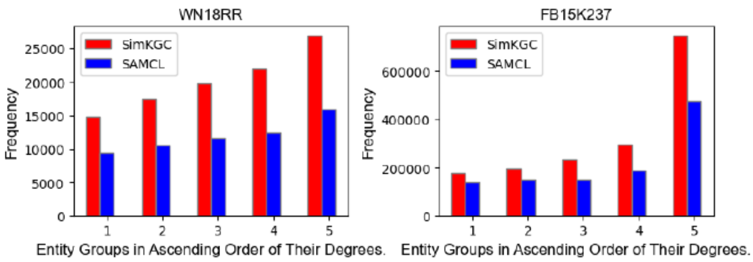

Figure 2: Frequency of the false positive triples proportional to their heads' degrees.

Figure 3.

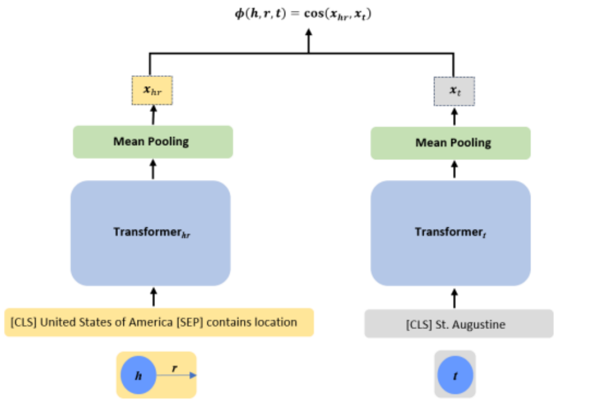

Figure 3: Bi-encoder architecture of SAMCL. Encoder *Transformer*$_{hr}$ takes the concatenation of a head's and a relation's name and description as input to produce embedding ($\mathbf{x}_{hr}$). Encoder *Transformer*$_t$ takes a tail's name and description to obtain embedding ($\mathbf{x}_t$). The score of a triple is calculated by the cosine similarity between ($\mathbf{x}_{hr}$) and ($\mathbf{x}_t$).

Figure 4.

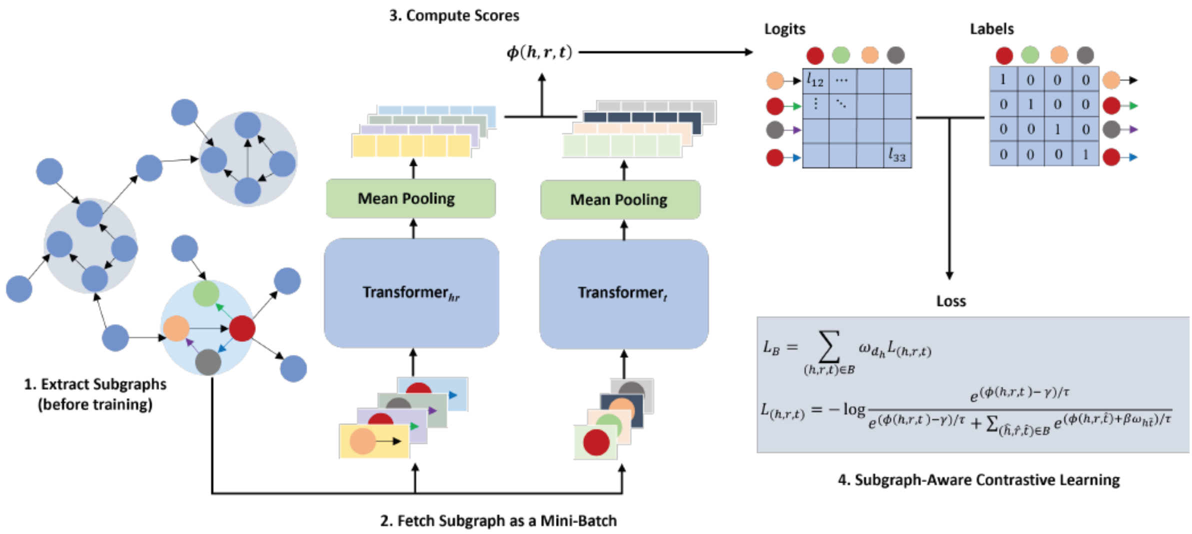

Figure 4: Overview of the proposed framework for subgraph sampling and training, which consists of: (i) extracting subgraphs from KG (which is performed in advance before training); (ii) fetching a mini-batch of triples in the subgraph with the least frequently visited triple; (iii) calculating a similarity between every (head, relation) pair and every tail in the mini-batch based on the embeddings produced by two encoders; (iv) contrastive learning over the mini-batch via infoNCE loss incorporated with two structure-aware factors.