Semantic Proximity Search on Graphs with Metagraph-based Learning

Yuan Fang1 Wenqing Lin1 Vincent Zheng2 Min Wu1 Kevin Chang3 Xiao-Li Li1

1 Institute for Infocomm Research, Singapore
2 Advanced Digital Sciences Center, Singapore
3 University of Illinois at Urbana-Champaign, USA

Problem: Semantic Proximity Search on Heterogeneous Graph

On a “typed” object graph that captures users and their attributes on a social network:

Which users are close to related to Bob?

Family? (Alice)
Classmates? (Tom)

Insights: Metagraphs to “Explain” Different Semantic Classes

Training

Definition of Proximity

Proximity of two nodes \( x, y \) on graph

\[
\pi(x, y; w) \triangleq \frac{2}{m_{xy} \cdot w + m_y \cdot w}
\]

\( m_{xy}[i] \): # times \( x, y \) co-occur in instances of metagraph \( i \)
\( m_y[i] \): # times \( y \) occurs in instances of metagraph \( i \)
\( w[i] \): weight for metagraph \( i \)

Basic Learning Model

Pairwise learning to rank

\[
P(q, x, y; w) \triangleq \frac{1}{1 + e^{-\mu(\pi(q, x; w) - \pi(q, y; w))}}
\]

Each example is a triplet: for query \( q \), \( x \) is ranked before \( y \).

Objective function

\[
L(w; \Omega) = \sum_{(q, x, y) \in \Omega} \log P(q, x, y; w)
\]

Matching Metagraphs

- Existing method
  - Backtracking DFS search
  - Node by node until an entire matched instance is found
  - Fail to leverage symmetric components

- Symmetry-based matching
  - Many metagraphs are symmetric
  - Avoid redundant computation

Dual-Stage Training

- Expensive to process/match all metagraphs
- Yet not all metagraphs are useful

Main Results

Datasets:
- College & Coworkers (labelled on LinkedIn)
- Family & Classmate (rule on Facebook)

Baselines:
- MGP: metagraph-based proximity (ours)
- MPP: metapath-based proximity
- MGP-U: all metagraphs have uniform weights
- MGP-B: only use the best metagraph
- SRW: supervised random walk