A meta path abstracts the paths and provides important insights about paths. This allows a query user to focus on the high-level relationship patterns, rather than on the detailed path instances.

**Importance Definition**

We posit that length, path instances, global and local relationships between the given node pairs are fundamental to evaluate importance of a meta path. Therefore, the importance function of a meta path $P = A_1 \rightarrow A_2 \rightarrow \cdots \rightarrow A_{k+1}$ can be formalized as follows:

$$I(P) = sup_{A_i}(P) \times rare(P, D) \times penalty(P)$$

**Minimum Instances (MI):** $MINI(P) = \min_{\ell \geq |P|} \{ |p| : p \in P_\ell \}$

where $p_i$ the node v in the i-th node inside type(1<e<1) in a path instance p of P. It is easy to see that $MINI(P) \leq |p|_{min}$. 

Support:

$$sup_{A_i}(P) = (\prod_{i=1}^{k+1} min|a_i(A_i)|) \times MINI(P)$$

where $a_i(A_i)$, $l(A_i+1|R_i)$ note the average out-degree and in-degree of $R_i$ from $A_i$ to $A_{i+1}$

Rarity: This is to evaluate how rare P is among similar pairs to (s, t)

$$rare(P, D) = \log \frac{\sum_{i=1}^{k} |p_i|_{min} \times sup_{A_i}(P) \times rare(P, D) \times penalty(P)}{\sum_{i=1}^{k} |p_i|_{min} \times sup_{A_i}(P) \times rare(P, D) \times penalty(P)}$$

Length Penalty: $penalty(P) = |P|^\beta, \beta \in (0, 1)$

Longer meta paths have "diluted" meaning [1], and thus the importance should decrease when P becomes longer.

The upper bound of a meta path’s importance is noted as $I(P)$

$$I(P) = \prod_{i=1}^{k+1} min|a_i(A_i)| \times l(A_{i+1}|R_i) \times sup_{A_i}(P) \times rare(P, D) \times penalty(P)$$

The inequality $(P - R) \leq I(P)$ guarantees the correctness of A* searching.

**Connections with baselines**

**Shortest Meta Path(SMP):**

$$sup_{A_i}(P) = 1, rare(P, D) = 1, penalty(P) = 1$$

**Meta Path Weight Function in WsRel [2] (SLV1):**

$$sup_{A_i}(P) = (\prod_{i=1}^{k+1} min|a_i(A_i)| \times l(A_{i+1}|R_i) \times sup_{A_i}(P) \times rare(P, D) \times penalty(P)$$

**Meta Path Weight Function in HeteRecom [3] (SLV2):**

$$sup_{A_i}(P) = e^{(\alpha |a_i(A_i)| \times l(A_{i+1}|R_i) \times sup_{A_i}(P) \times rare(P, D) \times penalty(P)}}$$

**Case Study on ACM**

We perform the top-k meta path query between Philip Yu and Jiawei Han by different importance function.

**MINI-based**

![MINI-based](image)

**SLV1 and SLV2**

![SLV1 and SLV2](image)

All methods rank P1 the first place, but SLV1 and SLV2 do not consider the direction of the edge and thus they can not differentiate P2 and P3.

**Label-based Connectivity Experiment**

Each paper is labelled by its subject, which is used to generate 100 positive pairs and 100 negative pairs, and we evaluate the performance by the average accuracy of the top-k meta paths, noted by $Accuracy@k$:

$$Accuracy@k = \frac{1}{k} \sum_{i=1}^{k} \frac{P_{ij}}{|P_{ij}| + n_{ij}}$$

$p_{ij}$, the number of connected positive pairs and $n_{ij}$, the number of connected negative pairs by $P_{ij}$, where $P_{ij}$ represents the i-th meta path(out of k) obtained by j-th positive pairs.

![MNIS and Baselines Comparison](image)

![Beta Comparison in MNIS](image)

**References**


[2] Zhu T, Peng Z, Wang S, et al. Measuring the relevance of different meta paths between Philip Yu and Jiawei Han by different importance function.