Ranking of Translator result graphs: ARAX's approach

Stephen Ramsey and David Koslicki[1]

Team Expander Agent

October 26, 2021

[1] `github:RTXteam/RTX/code/ARAX/ARAXQuery/ARAX_ranker.py`

ARAX result-graph ranking method

A result g is a weighted multi-digraph $g = (\mathcal{V}, \mathcal{P}, \mathcal{E}, \mathcal{W})$, where $\mathcal{V}$ is the vertex-set, $\mathcal{P}$ is a predicate set, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V} \times \mathcal{P}$ is the set of directed edges, and $\mathcal{W} : \mathcal{E} \to [0, 1]$ contains edge weights. Let $\boldsymbol{A}_g$ be the induced weight matrix with components $(\boldsymbol{A}_g)_{v,v'} = \sum_{p \text{ sth. } (v,v',p)\in\mathcal{E}}(\mathcal{W}(v, v', p))$. Let G be the set of result-graphs. For each g, ARAX computes three $\mathbb{R}$-scalar scores:

1. $S_1(g) = ||\boldsymbol{A}_g||_F$, the Frobenius norm
2. $S_2(g) = \max\left(\text{max-flow}(\boldsymbol{A}_g)\right)$
3. $S_3(g) = \left\langle((\boldsymbol{A}_g)^{L(g)})_{i,j}\right\rangle_{(i,j)\in P_g(L(g)}/L(g)!$

where $\text{max-flow}(\boldsymbol{A}_g)$ denotes the maximum-flow matrix computed for the weighted digraph via the Push-relabel algorithm; and $L(g)$ denotes the maximum unweighted geodesic path length of g; and $P_g(l)$ denotes the set of node pairs with geodesic path length l. The scores are combined by $\sum_{k\in\{1,2,3\}} \text{rank}_{S_k(G)}(S_k(g))/3$.