Efficient Computation of Relationship Centrality in Large Entity-Relationship Graphs

Stephan Seufert    Srikanta Bedathur    Johannes Hoffart
Andrey Gubichev    Klaus Berberich

Relationship Centrality

Given a large Entity-Relationship graph (such as Wikipedia) and two sets of entities \(S, T\), identify entities that play an important role in the relationship between \(S\) and \(T\).

Ingredients:
- Underlying graph structure: Connectivity of the entities
- Semantic relatedness between individual entities to measure connection strength

Query sets can be specified manually or, when combined with a knowledge base such as YAGO, in the form of SPARQL queries.

Semantic Relatedness between Entities

Link-Based
Count overlapping links
MW [Milne & Witten, WIKIAI 2008]

Keyphrase-Based
Intersect weighted keyphrases
KORE [Hoffart et al., CIKM 2012]

Measure

Centrality of vertex \(v\) with respect to the query sets \(S\) and \(T\):

\[
c_R(v) = \frac{1}{\sum_{s \in S \cap T} \rho(s, v, t)}
\]

Using a connection penalty of the form

\[
\rho(s, v, t) = (1 + d(s, v))(1 + d(v, t))
\]

- Distance \(d(s, v)\) captures the dissimilarity between the entities and depends on the semantic relatedness measure, for example

\[
d(s, v) = 1 - \text{KORE}(s, v)
\]

- Computation: \(|S|+|T|\) rounds of Dijkstra’s algorithm
- Speedup: Stop expansion after specified distance bound \(\Delta\)

Example Results

- Query 1: Events between European politicians and politicians from the United States
- Query 2: Movies between US action movie stars and Asian action movie stars
- Query 3: Events between countries from Middle East/Central Asia and Western countries

Efficiency

- Execution times over the Wikipedia graph (~37M edges)
- Different choice of distance bound \(\Delta\)