Abstract

Approximate sub-graph matching is important in many graph data mining fields. At present, current solutions can be difficult to implement, have an expensive pre-processing phase, or only work for given types of graph. In this paper a novel generic approach is presented which addresses these issues. An approximate sub-graph matcher (A-SGM) calculates the distance between the topological characteristics (footprint) of the sub-graphs to be matched, applying a weighting to the different sub-graph characteristics and those of neighbor nodes. The weights are calibrated for each dataset with a simulated annealing process using sample sets of graph nodes to reduce computational cost, and an exact isomorphism matcher as a fitness function which takes into account how well the match maintains the neighboring node degree distributions. Benchmarking is performed on several state of the art methods and real and synthetic graph datasets to evaluate the precision, recall and computational cost. The results show that the A-SGM is competitive with state of the art methods in terms of precision, recall and execution time.
References


Index Terms
Computer Science Information Sciences

Keywords
Graph Matching, topology, graph characteristics, weight calibration, simulated annealing, graph queries.