

SSApproximate Shortest Distance Computing: A Query-Dependent Local Landmark Scheme

Abstract:

Shortest distance query is a fundamental operation in large-scale networks. Many existing methods in the literature take a landmark embedding approach, which selects a set of graph nodes as landmarks and computes the shortest distances from each landmark to all nodes as an embedding. To answer a shortest distance query, the precomputed distances from the landmarks to the two query nodes are used to compute an approximate shortest distance based on the triangle inequality. In this paper, we analyze the factors that affect the accuracy of distance estimation in landmark embedding. In particular, we find that a globally selected, query-independent landmark set may introduce a large relative error, especially for nearby query nodes. To address this issue, we propose a query-dependent local landmark scheme, which identifies a local landmark close to both query nodes and provides more accurate distance estimation than the traditional global landmark approach. We propose efficient local landmark indexing and retrieval techniques, which achieve low offline indexing complexity and online query complexity. Two optimization techniques on graph compression and graph online search are also proposed, with the goal of further reducing index size and improving query accuracy. Furthermore, the challenge of immense graphs whose index may not fit in the memory leads us to store the embedding in relational database, so that a query of the local landmark scheme can be expressed with relational operators. Effective indexing and query optimization mechanisms are designed in this context. Our experimental results on large-scale social networks and road networks demonstrate that the local landmark scheme reduces the shortest distance estimation error significantly when compared with global landmark embedding and the state-of-the-art sketch-based embedding.