BCDR: BETWEENNESS CENTRALITY-BASED DISTANCE RESAMPLING FOR GRAPH SHORTEST DISTANCE EMBEDDING

Anonymous authors
Paper under double-blind review

ABSTRACT

Along with unprecedented development in social analysis and molecular biology, Shortest Distance Queries(SDQs) in graph recieve widespread and increasing attention among the public. Approximate algorithm of SDQs with reduced complexity is thus of vital importance to complex graph applications in such contemporary society. Recently, embedding-based distance prediction makes a breakthrough in both efficiency and accuracy, ascribing to significant performance of Graph Representation Learning(GRL). The common idea of embedding-based distance prediction usually leverages truncated random walk and Point Mutual Information(PMI) to pre-embed local structural feature into dense vector on each node and integrates with subsequent predictor for global extraction of nodes' mutual shortest distance. However, they are not without limitation. Random walk as an unstrained node sequence possesses a limited exploration distance, failling to take into account remote node under graph shortest distance metric, while PMI-based maximum likelihood optimization of node embeddings reflects versatile local similarity, which impact preservation of the exact shortest distance relation during the mapping from orignal graph space to pre-embedded vector space.

To solve the above problems, in this paper, we propose a novel graph shortest distance embedding method called Betweenness Centrality-based Distance Resampling(BCDR). First, we prove in statistic perspective that a Betweenness Centrality-based(BC-based) random walk can occupy a wider distance range measured by intrinsic metric in graph domain due to the awareness of path structure. Second, we perform a unique distance resampling from original walk paths before maximum likelihood optimization, and prove that this strategy preserves distance relation with respect to any calibrated node via steering optimization objective to reconstruct a global distance matrix. Our method possesses strong theoretical backdrops and much more satisfactory empirical performance among a broad class of real graph datasets even of large diameters in SDQ problem as well as other graph structure-related applications.