Node Embeddings In Practice

This summary was generated by AI.

The thesis provides a comprehensive overview of graph embeddings - techniques converting graph data into low-dimensional vectors while preserving structural and relational information.
Initially, it establishes the mathematical and computational foundations, addressing the non-Euclidean nature of graph data.

The node2vec algorithm, a shallow embedding method, is explored extensively. Node2vec combines random walks on graphs with the Skip-gram neural model, originally developed for language processing tasks.
It defines similarity between nodes based on their co-occurrence in random walks. The method is governed by two parameters: 'return' (p) and 'in-out' (q), balancing breadth-first and depth-first search strategies.

Subsequently, the thesis discusses limitations of shallow embeddings, such as lack of inductivity, computational inefficiency on large graphs, and the inability to leverage node features.

To address these issues, Graph Neural Networks (GNNs) and specifically GraphSAGE, a deep embedding approach, are introduced. GraphSAGE aggregates node information through learned neural network parameters, enabling inductive reasoning—embedding nodes unseen during training.
Different aggregation methods are analyzed, including mean, pooling, and attention-based mechanisms.
Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) are also presented as foundational models.

Empirical evaluation includes experiments on realistic graph datasets and synthetic graphs generated by the Erdős–Rényi model.
The study assesses embedding quality, computational performance, and robustness of the mentioned methods. Results highlight the efficacy of deep embeddings in capturing node and structural attributes, outperforming shallow embeddings in scalability and flexibility.

The thesis concludes that while shallow embeddings like node2vec are useful for fixed-size graphs and simpler tasks, deep embedding methods such as GraphSAGE provide superior performance and broader applicability, particularly in dynamic and large-scale graph scenarios.