Graph Embedding Techniques,

Applications, and Performance: A Survey

Graph Embedding Techniques,

Applications, and Performance: A Survey

Graph Embedding Techniques,

Applications, and Performance: A Survey

Graph Embedding Techniques,

Applications, and Performance: A Survey

**Graph Embedding Techniques,Applications, and Performance: A Survey**

GRAPH analysis has been attracting increasing attention in the recent years due the ubiquity of networks in the real world. Graphs used to denote information in various areas eg. biology (Protein-Protein interaction networks), social sciences (friendship networks) and linguistics (word co-occurrence networks).

Modeling the interactions between entities as graphs has enabled researchers to understand the various network systems in a systematic manner.

**Graph analytic tasks:**

**(a) node classification:** aims at determining the label of nodes (a.k.a. vertices) based on other labeled nodes and the topology of the network

**(b) link prediction:** refers to the task of predicting missing links or links that are likely to occur in the future.

**(c) clustering:** is used to find subsets of similar nodes and group them together.

**(d) visualization:** helps in providing insights into the structure of the network.

**What are the challenges?**

1. Choice of property: Graph has various properties, difficult to choose and apply or perform.
2. Scalability: Large graphs need the embedding methods to be scalable enough for processing.
3. Dimensionality of the embedding: Get more dimensions will increase the precision but take more space and time complexity.

**What are the contributions?**

1. Propose taxonomy for graph embedding and distinguish their differences.
2. Analyze various graph embedding models in detail, find their performances on different tasks.
3. Present GEM for further researches.

**Approaches from Preliminaries:**

1. The higher the edge weight, the more similar the two nodes are expected to be.
2. Edge weight called First-order proximity, used to measure similarity.
3. Second-order proximity describe the proximity of pair’s neighborhood structure.
4. Embedding process maps each node to a low dimensional feature vector and preserve the connection strength between vertices.

**Graph Embedding Evolution**

Change from part of dimensionality reduction technique to obtaining scalable technique which leverage the sparsity of real world network. Time complexity change from O(|V|^2) to O(|E|).

**Categories of Graph Embedding Method**

1. Factorization based method (Logically Linear Embedding, Laplacian Eigenmaps, Graph Factorization, GraRep, HOPE) [**not capable of learning arbitrary functions**]
2. Random Walk based method (DeepWalk, node2vec) [**can model a wide range if functions]**
3. Deep learning based method (SDNE) []

“Given enough parameters, they can learn the mix of community and structural equivalence, to embed the nodes such that the reconstruction error is minimized. We can interpret the weights of the autoencoder as a representation of the structure of the graph.”

**What are the applications of embedding?**

1. Network compression (Store network more efficient, run algorithm faster, partition graph and reduce the number of edges)

**Graph Reconstruction**

-Reconstruct the graph with their proximity, rank pair of nodes, calculate the ratio of real links in top k predictions by using the mean, sd, MAP

-Cause dimension increase, MAP value increase.

1. Visualization (Different embedding method s cause different network structures then change the interpretation of node visualization as well)
2. Clustering (network partition, k-mean is the most popular way)
3. Link predication (task of predicting either missing interactions or links in the future, based on similarity, MLE, probability)

-predicting unobserved links in the graph

-when dimensions increase, not all performance improved.

1. Node classification (Predicting the missing labels in the network, example as Logistic Regression and Naïve Bayes)