Contents

1 Paper Reading Task

1.1 Summary of Paper

1. Maps graph nodes to vector embeddings using 2nd order random walks.

2. By being flexible with the parameters p and q, one can balance homophility (enabled by BFS, with the p parameter) and structural equivalence (enabled by DFS, with the q parameter).

3. node2vec, by virtue of its technique, is relatively fast and very easily scalable even for large networks. Coupled with it's flexibility, it can be used in a number of uses.

1.2 Major Strengths of Paper

- The authors of the paper have used multiple datasets, of varying sizes and complexity, to evaluate node2vec. This eliminates possible bias towards certain properties of graphs (size, diameter, etc.).

- Random walks allow memory-efficient sampling and generation of embeddings.

1.3 Major Weaknesses of Paper

- Largely an extension of Deepwalk with the introduction of parameters p and q. Therefore, not as groundbreaking as the titular correlation with Word2Vec indicates

- Continuing from the previous point, DeepWalk was usually only 0.01-1% better than node2vec, thus making marginal gains in exchange for added complexity.

- The datasets chosen did not include any extreme cases, making it difficult to evaluate the performance of word2vec in cases of general networks where the behaviour can be hard to predict (economic activity).

- There were very few algorithms that were compared, making it seem artifically biased in the favour of node2vec

1.4 Improvements in Paper

- The effect of operators on the accuracy was not uncovered and can be an area of future research. Particularly, is there a class of operators that node2vec can accomodate? Furthermore, what properties do operators satisfy to yield high-accuracy predictions?

- There could have been comparisions with other classical community detection algorithms like InfoMap and Spinglass, allowing for a more fairer comparision on various graph parameters like number of nodes.