
# NLP - Graph Embedding

A Brief Introduction into GE Models utilizing Random Walk

Supervisor: M.Sc. Matthias Assenmacher

Noah Hurmer

tbd

**Abstract**

TODO

# Contents

## List of Tables

## List of Figures

## List of Abbreviations

| Term | Abbreviation |
|---|---|
| DL | Deep Learning |
| GE | Graph Embedding |
| RW | Random Walk |

# 1 Introduction

## 1.1 Motivation

## 1.2 Applications

- uses in social networks
- visualization
- Network compression
- Network Partitioning
- Node Classification
- Link Prediction
- fake news detection

perhaps omit some examples as to not overcrowd or overcomplicate this passage.

## 1.3 Graph types

- What is a Graph?
- Nodes and edges
- Nodes are the entities
- Edges are the relations between nodes
- nodes can consist of all sort of Data, images, text, etc even mixed
- nodes can have additional attributes

### 1.3.1 Graph modes

- homogeneous, heterogeneous
- graphs can be directional, weighted, semantik, knowledge based

## 1.4 Embedding types

- node embedding
- edge embedding
- subgraph/Groups
- entire Graphs

# 2 Problem and Dataset

introduce the Problem and the associated Dataset on which the algorithms below should be run on.

# 3 Deep Learning and random Walk

Short Paragraph explaining how one Approach to (node)Graph embedding utilizes Deep Learning methods. List the other approaches not using DL with a brief explanation, and especially mention how the computational advantage of the Random Walk approach can be preferable to MF.

## 3.1 Random Walk

Introduce the random Walk concept as another subsection of Deep Learning tools for Graph embedding. What is the Aim of the Random Walks?

## 3.2 SkipGram

Briefly explain the NLM SkipGram and how its paired with Random Walks. Also introduce the problem of calculating the associated softmax equation.

### 3.2.1 DeepWalk as an example for Hierarchical softmax

### 3.2.2 node2vec as an example for negative sampling

## 3.3 (LSTM)

Optional Section considered as an excursion to another Random Walk method using a different Model and heterogeneous Graphs as Inputs. Perhaps using HSNL as the algorithm example.

# 4 Performance Evaluation

Discuss how to evaluate Graph Embedding methods and algorithms and then evaluate the above used ones.

# 5 References


Alicia Frame, PhD. 2019. "Graph Embeddings." 2019. https://www.youtube.com/watch?v=oQPCxwmBiWo&ab_channel=Neo4j.

Cai, HongYun, Vincent W. Zheng, and Kevin Chen-Chuan Chang. 2018. "A Comprehensive Survey of Graph Embedding: Problems, Techniques, and Applications." *IEEE Transactions on Knowledge and Data Engineering* 30 (9): 1616–37. https://doi.org/10.1109/TKDE.2018.2807452.

Goyal, Palash, and Emilio Ferrara. 2018. "Graph Embedding Techniques, Applications, and Performance: A Survey." *Knowledge-Based Systems* 151: 78–94. https://doi.org/https://doi.org/10.1016/j.knosys.2018.03.022.

Grover, Aditya, and Jure Leskovec. 2016. "Node2vec: Scalable Feature Learning for Networks." *CoRR* abs/1607.00653. http://arxiv.org/abs/1607.00653.

Perozzi, Bryan, Rami Al-Rfou, and Steven Skiena. 2014. "DeepWalk: Online Learning of Social Representations." *CoRR* abs/1403.6652. http://arxiv.org/abs/1403.6652.

Pilehvar, Mohammad Taher, and Jose Camacho-Collados. 2020. "Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning." *Synthesis Lectures on Human Language Technologies* 13 (4): 1–175. https://doi.org/10.2200/S01057ED1V01Y202009HLT047.