node2vec: Scalable Feature Learning for Networks

Authors: Aditya Grover and Jure Leskovec

Chenyu Shi and Shupei Li

Leiden Institute of Advanced Computer Science

November 18, 2022

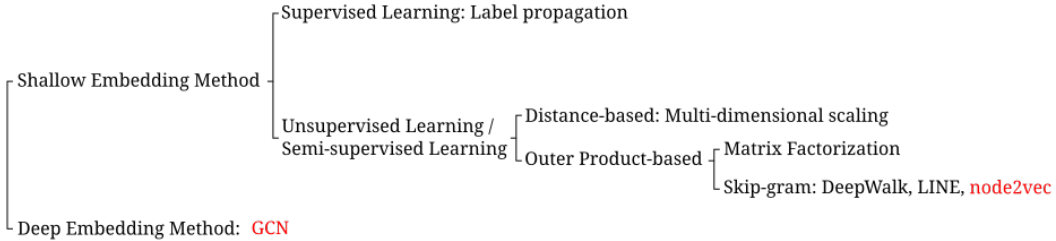
Universiteit Leiden
The Netherlands

Introduction to Graph Embeddings

- Represent graph-structured data.
- Applications:
  Social network analysis, recommender systems, molecular structure modelling, etc.
- Challenge: Limitations of traditional methods.
- Development of techniques specially for graph representations.

Related Work

A taxonomy of graph embedding techniques[1].



[1] Kevin P. Murphy. *Probabilistic Machine Learning: An Introduction*. Adaptive Computation and Machine Learning series. MIT Press, 2022. ISBN: 9780262046824.

Experiment 1: Multi-label Classification

- Task description
  - Labels from a finite set $\mathcal{L}$
  - Training: A fraction of nodes and all their labels.
  - Predict the labels for the remaining nodes.
- Data

Dataset	Nodes	Edges	Labels
BlogCatalog	10,312	333,983	39
Protein-Protein Interactions (PPI)	3,890	76,584	50
Wikipedia	4,777	184,812	40

Experiment 1: Multi-label Classification

- Results

Algorithm	**Dataset**		
	BlogCatalog	PPI	Wikipedia
Spectral Clustering	0.0405	0.0681	0.0395
DeepWalk	0.2110	0.1768	0.1274
LINE	0.0784	0.1447	0.1164
node2vc	**0.2581**	**0.1791**	**0.1552**
node2vec settings (p, q)	0.25, 0.25	4, 1	4, 0.5
Gain of node2vec [%]	**22.3**	**1.3**	**21.8**

Experiment 2: Link Prediction

- Task description
  - A network with a fraction of edges removed.
  - Predict these missing edges.
- Data

Dataset	Nodes	Edges
Facebook	4,039	88,234
Protein-Protein Interactions (PPI)	19,706	390,633
arXiv ASTRO-PH	18,722	198,110

Experiment 2: Link Prediction

- Results

Algorithm	**Dataset**		
	Facebook	PPI	arXiv
Common Neighbors	0.8100	0.7142	0.8153
Jaccard's Coefficient	0.8880	0.7018	0.8067
Adamic-Adar	0.8289	0.7126	0.8315
Pref. Attachment	0.7137	0.6670	0.6996
Spectral Clustering	0.6192	0.4920	0.5740
DeepWalk	**0.9680**	0.7441	0.9340
LINE	0.9490	0.7249	0.8902
mode2vec	**0.9680**	**0.7719**	**0.9366**

Preliminary Results

Future work