Network Embedding

Introductory Talk by

Akash Anil

Research Scholar
OSINT Lab
Dept. of Computer Science & Engineering
Indian Institute of Technology Guwahati

September 11, 2019

Table of contents

Network Embedding

Suppose $G(V, E)$ represents a network then Network Embedding refers to generating low dimensional network features corresponding to Node, Edge, Substructure, and the Whole-Graph [1].

Network Embedding

Suppose $G(V, E)$ represents a network then Network Embedding refers to generating low dimensional network features corresponding to Node, Edge, Substructure, and the Whole-Graph [1].

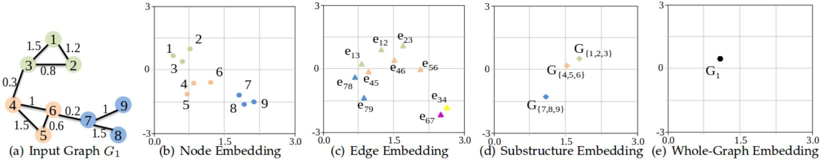


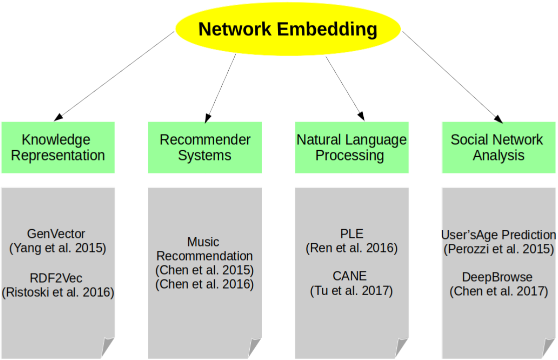
Figure: Different taxonomies of Network Embedding, Picture Source: Cai et. al. "https://arxiv.org/pdf/1709.07604.pdf"

Applications

1. Automatic feature vector generation helps in solving traditional problems on graph e.g. node classification, relation prediction, clustering, etc.

Applications

1. Automatic feature vector generation helps in solving traditional problems on graph e.g. node classification, relation prediction, clustering, etc.
2. Recent Uses



Why Neural Network based Network Embedding ??

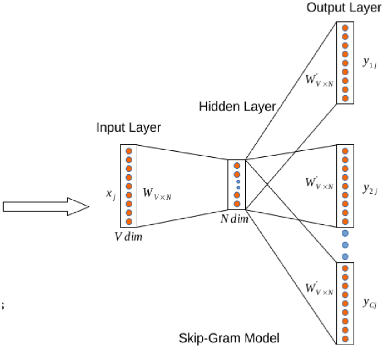
- Traditional approaches based on matrix factorization (e.g. SVD) are not scalable to networks with large number of nodes.
- Recent advances in unsupervised word embedding using single layer neural network (e.g. Word2Vec [3]).

Word Embedding (Word2Vec)

Barak Obama was the president of United States of America

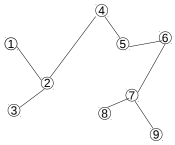
Window = 1

Input (I)	Context / Output (C)
Barak	Obama
Obama	Barak
Obama	president
president	Obama
president	United
United	president
United	States
States	United
Sates	America
America	States

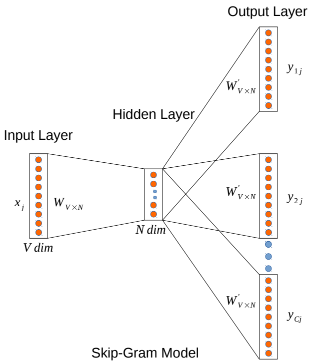


Max P(C|E(I))

A generalized framework used for Node Embedding

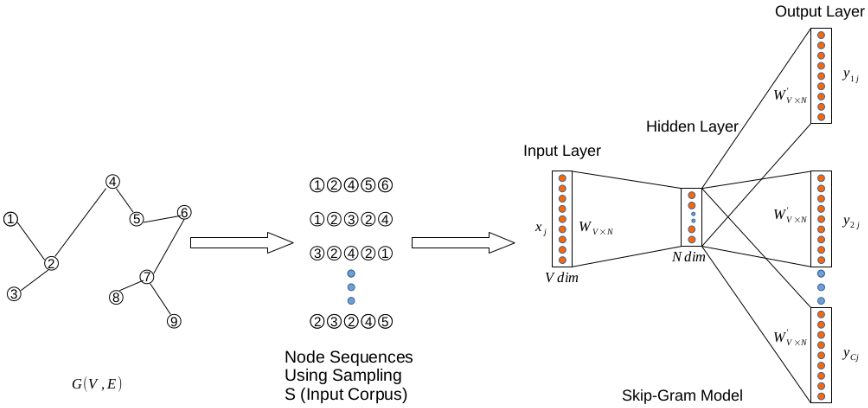


$G(V, E)$



Skip-Gram Model

A generalized framework used for Node Embedding



A General Framework for Unsupervised Node Embedding using Skip-Gram

DeepWalk [4]

- DeepWalk is the first network embedding model exploiting neural networks.
- Scalable to the large real-world networks.

Designing DeepWalk Model

1. Generate node sequences (input corpus) using truncated random walk.
2. Iterate random walks from same source node 80 times for convergence.
3. Supply the node sequences as input to skip-gram model.
4. Maximize the probability of neighborhoods for the given node.

Multi-label Classification for Blog-Catalog Data

	% Labeled Nodes	10%	20%	30%	40%	50%	60%	70%	80%	90%
Micro-F1(%)	DeepWalk	**36.00**	**38.20**	**39.60**	**40.30**	**41.00**	**41.30**	41.50	41.50	42.00
	SpectralClustering	31.06	34.95	37.27	38.93	39.97	40.99	**41.66**	**42.42**	**42.62**
	EdgeCluster	27.94	30.76	31.85	32.99	34.12	35.00	34.63	35.99	36.29
	Modularity	27.35	30.74	31.77	32.97	34.09	36.13	36.08	37.23	38.18
	wvRN	19.51	24.34	25.62	28.82	30.37	31.81	32.19	33.33	34.28
	Majority	16.51	16.66	16.61	16.70	16.91	16.99	16.92	16.49	17.26
Macro-F1(%)	DeepWalk	**21.30**	**23.80**	25.30	26.30	27.30	27.60	27.90	28.20	28.90
	SpectralClustering	19.14	23.57	**25.97**	**27.46**	**28.31**	**29.46**	**30.13**	**31.38**	**31.78**
	EdgeCluster	16.16	19.16	20.48	22.00	23.00	23.64	23.82	24.61	24.92
	Modularity	17.36	20.00	20.80	21.85	22.65	23.41	23.89	24.20	24.97
	wvRN	6.25	10.13	11.64	14.24	15.86	17.18	17.98	18.86	19.57
	Majority	2.52	2.55	2.52	2.58	2.58	2.63	2.61	2.48	2.62

Multi-label Classification for Flickr Data

	% Labeled Nodes	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
Micro-F1(%)	DeepWalk	**32.4**	**34.6**	**35.9**	**36.7**	**37.2**	**37.7**	**38.1**	**38.3**	**38.5**	**38.7**
	SpectralClustering	27.43	30.11	31.63	32.69	33.31	33.95	34.46	34.81	35.14	35.41
	EdgeCluster	25.75	28.53	29.14	30.31	30.85	31.53	31.75	31.76	32.19	32.84
	Modularity	22.75	25.29	27.3	27.6	28.05	29.33	29.43	28.89	29.17	29.2
	wvRN	17.7	14.43	15.72	20.97	19.83	19.42	19.22	21.25	22.51	22.73
	Majority	16.34	16.31	16.34	16.46	16.65	16.44	16.38	16.62	16.67	16.71
Macro-F1(%)	DeepWalk	**14.0**	17.3	**19.6**	**21.1**	**22.1**	**22.9**	**23.6**	**24.1**	**24.6**	**25.0**
	SpectralClustering	13.84	**17.49**	19.44	20.75	21.60	22.36	23.01	23.36	23.82	24.05
	EdgeCluster	10.52	14.10	15.91	16.72	18.01	18.54	19.54	20.18	20.78	20.85
	Modularity	10.21	13.37	15.24	15.11	16.14	16.64	17.02	17.1	17.14	17.12
	wvRN	1.53	2.46	2.91	3.47	4.95	5.56	5.82	6.59	8.00	7.26
	Majority	0.45	0.44	0.45	0.46	0.47	0.44	0.45	0.47	0.47	0.47

Limitations of DeepWalk

- Relying on rigid notion of network neighborhood or local characteristics.
- Fails to captures proximity of different semantics.

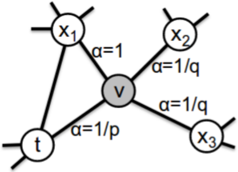
Node2Vec [2]

- Uses 2nd Order Random walk to generate corpus.
- Presents a semi-supervised model which balances the trade-offs of capturing local and global network characteristics.
- Scalable model applicable to any type of graph e.g., (un)directed, (un)weighted, etc.

Designing Node2Vec (I)

Suppose a random walker just traversed edge $(t, v) \in E$ and now resting at node v. To estimate the transition probability to visit next node x originating from v, Node2Vec sets the transition probability w_{vx} to $\pi_{vx} = \alpha_{pq}(t, v).w_{vx}$, where

$$\alpha_{pq}(t, v) = \begin{cases} \frac{1}{p} & \textit{if} \quad d_{tx} = 0 \\ 1 & \textit{if} \quad d_{tx} = 1 \\ \frac{1}{q} & \textit{if} \quad d_{tx} = 2 \end{cases}$$



here d_{tx} is the shortest distance between nodes t to x.

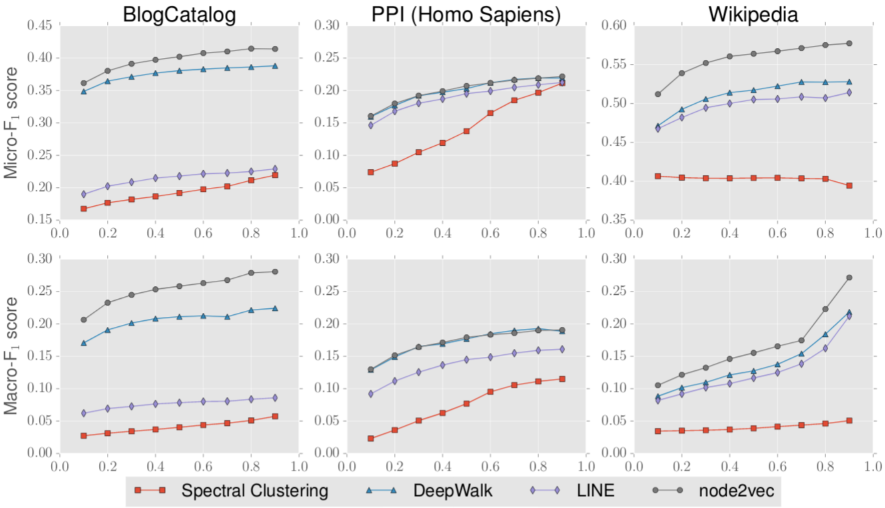
Designing Node2Vec (II)

- p is treated as Return Parameter.
- q is treated as in-out parameter.
- High q gives BFS like behaviour and low represents DFS.
- BFS is helpful in capturing local proximities between nodes.
- DFS is helpful in capturing global proximities between nodes.

Designing Node2Vec (II)

- p is treated as Return Parameter.
- q is treated as in-out parameter.
- High q gives BFS like behaviour and low represents DFS.
- BFS is helpful in capturing local proximities between nodes.
- DFS is helpful in capturing global proximities between nodes.
- Node sequences are generated using truncated random walks of length 80.
- From each node random walk iterates 10 times.
- Using 10% of the dataset sample, Node2Vec sets the hyper-parameters p and q.
- Supply the node sequences as input to skip-gram model and maximize the neighborhood probability.

Efficiency of Node2Vec over Multi-label classification

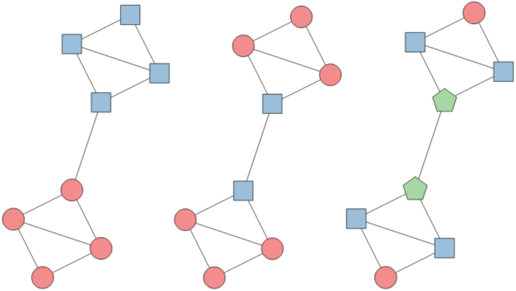


Limitations of DeepWalk and Node2Vec

- fail to capture different types of similarity naturally observed in real-world networks.

Limitations of DeepWalk and Node2Vec

- fail to capture different types of similarity naturally observed in real-world networks.



(a) Community structure (b) Roles (c) Structural equivalence

Versatile Graph Embeddings (VERSE) [5]

- Proposes a model capable of capturing different types of similarity distributions.
- Uses state-of-the-art similarity measures to instantiate the model.

Designing VERSE

- Select a similarity measure, such as Personalized PageRank, SimRank, etc. and generate the similarity distribution matrix Sim_G.
- Initialize the Embedding space Sim_E with random weights.
- Minimize the KL-divergence between distributions Sim_G and Sim_E: $\sum_{v \in V} KL(Sim_G(v,.)||Sim_E(v,.))$

Efficiency of VERSE over Multi-class classification for Co-cit data

	labelled nodes, %				
method	1%	3%	5%	7%	9%
fVERSE	27.52	29.83	31.01	31.68	32.24
VERSE	27.32	29.42	30.67	31.32	31.83
DeepWalk	26.81	29.27	30.37	31.04	31.43
GraRep	27.68	29.21	30.24	30.23	30.79
LINE	23.68	26.90	27.89	28.49	28.80
HOPE	22.81	26.63	27.59	28.19	28.58
hsVERSE	27.46	29.45	30.67	31.38	31.92
Node2vec	27.45	29.66	30.82	31.54	32.04

References I

Hongyun Cai, Vincent W Zheng, and Kevin Chang.
A comprehensive survey of graph embedding: problems, techniques and applications.
TKDE, 2018.

Aditya Grover and Jure Leskovec.
Node2vec: Scalable feature learning for networks.
In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 855–864, New York, NY, USA, 2016. ACM.

References II

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean.
Distributed representations of words and phrases and their compositionality.
In *Advances in neural information processing systems*, pages 3111–3119, 2013.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena.
Deepwalk: Online learning of social representations.
In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710. ACM, 2014.

References III

Anton Tsitsulin, Davide Mottin, Panagiotis Karras, and Emmanuel Müller.
Verse: Versatile graph embeddings from similarity measures.
In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, pages 539–548, Republic and Canton of Geneva, Switzerland, 2018. International World Wide Web Conferences Steering Committee.


Thank you