MARU: Representation Learning for Heterogeneous Networks with Meta-context Aware Random Walks

Jyun-Yu Jiang, Zeyu Li, Chelsea J.-T. Ju and Wei Wang

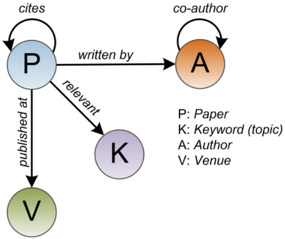
Scalable Analytics Institute (ScAi)
University of California, Los Angeles (UCLA)

CIKM 2020

Information networks are ubiquitous in our lives.

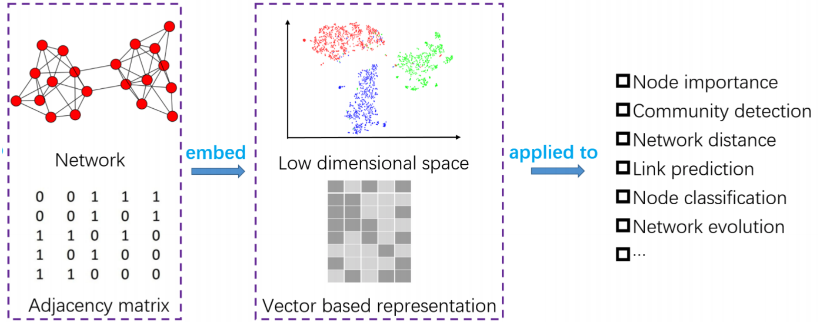
Social Networks

Medical Networks



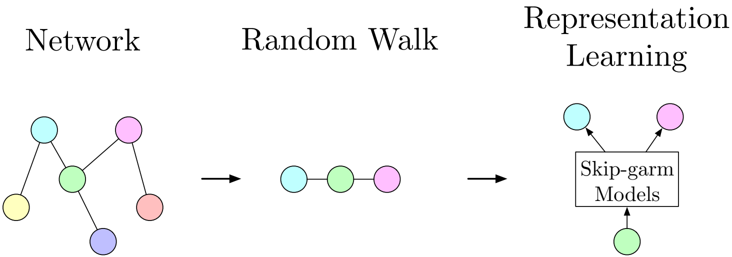
Bibliographic Networks

Network Embedding for Machine Learning Approaches



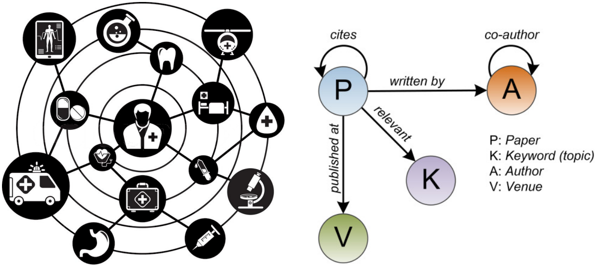
Low-dimensional feature vectors can be directly applied on applications.

State-of-the-art: Learning from Random Walks



Many recent researches focus on generating "good" random walks.

Heterogeneous networks are more intuitive and general.



Medical Networks Bibliographic Networks

Nodes can belong types with different semantic meanings.

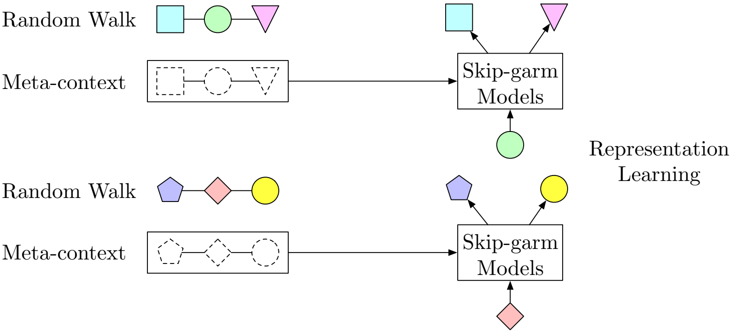
Existing methods are limited for heterogeneous networks.

- Nodes can have different frequencies and distributions.
    - Popular nodes can be over-optimized.
- Node types are ignored.
    - Random walks with varied types can have different semantics.
- Relying on prior knowledge and given strategies.
    - Predefined meta-paths, i.e., short type sequences.
    - Ad-hoc random walk strategies.
    - Random walks are manipulated and limited.

A robust method for learning heterogeneous network representations is needed!

Motivation: Dynamically Learning with Meta-context

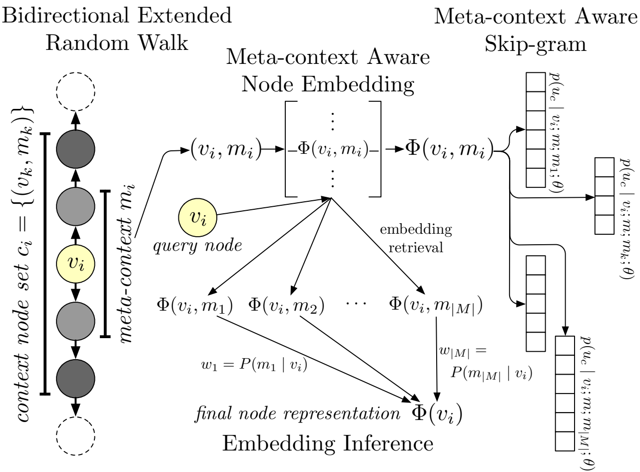
Different contexts of types should be modeled separately.



The number of feasible meta-contexts are usually limited!

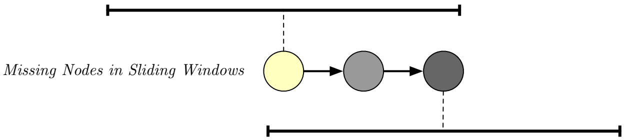
MARU: The Proposed Framework

Representation Learning with Meta-context Aware Random Walks (MARU)

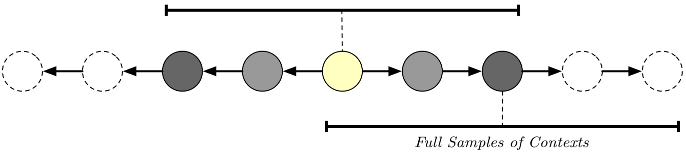


Bidirectional Extended Random Walk

Classical Random Walks



Bidirectional Extended Random Walks



Tail nodes can also be theoretically better optimized.

Theoretical Proof for Bidirectional Extended Random Walk

Corollary

Assume head nodes are never transitioning to tail nodes in random walks, and the probability of transitioning between tail nodes is p. Given a tail starting node u and the walking length $2n+1$, the expected number of tail nodes in a bidirectional random walk is greater than the expected number in a one-directional random walk.

Proof.

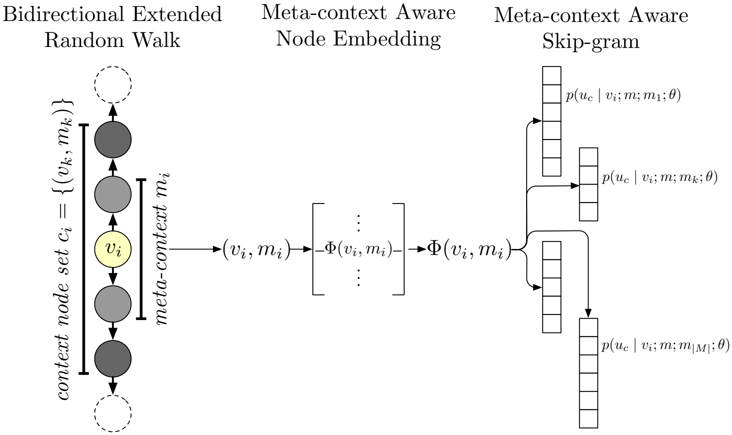
For the one-directional random walk, the expected number of visited tail nodes is $E_o = 1 + \sum_{i=0}^{2n} i \cdot p^i \cdot (1-p)$. For the bidirectional random walk, the expected number of visited tail nodes is $E_b = 1 + 2 \cdot \sum_{i=0}^{n} i \cdot p^i \cdot (1-p)$. Therefore, we have

$$\lim_{n\to\infty} E_b - E_o = \lim_{n\to\infty} (1-p) \cdot \left(\sum_{i=1}^{n} i \cdot p^i - (i+n) \cdot p^{i+n} \right) > 0$$

□

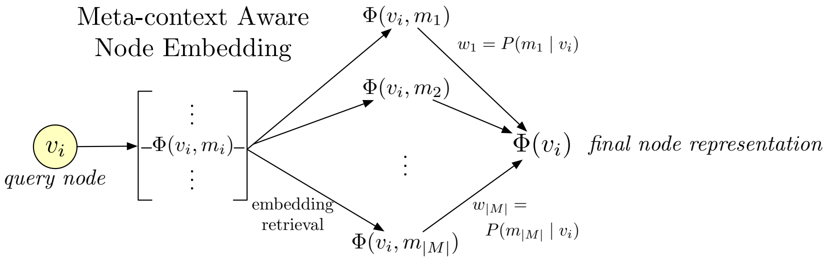
Meta-context Aware Node Embedding and Learning

A node v can have distinct representations $\Phi(v, m)$ for different meta-contexts m.



Embedding Inference

Each node still needs an ultimate representation for applications.



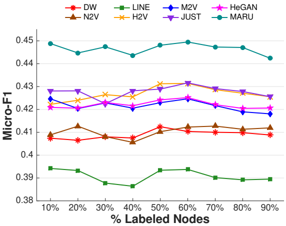
Embedding Inference

Experimental Tasks and Datasets

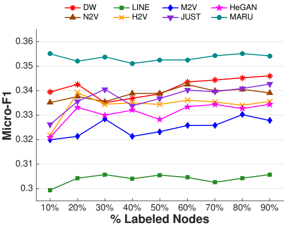
- Three Evaluation Tasks
  - Multi-label Node Classification (Supervised Learning)
  - Node Clustering (Unsupervised Learning)
  - Link Prediction (Supervised Learning)

Dataset	Node Types and Number of Nodes			
DBIS (264,323 edges)	Author (A) 60,694	Paper (P) 72,902	Venue (V) 464	
MovieLens (1,097,495 edges)	Movie (M) 10,197	Actor (A) 95,321	Director (D) 4,060	User (U) 2,113
Yelp (411,263 edges)	User (U) 16,239	Business (B) 14,284	Category (C) 511	Location (L) 47

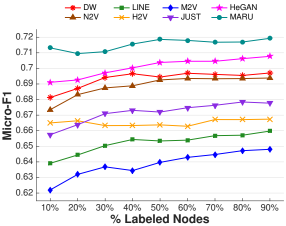
Multi-label Node Classification



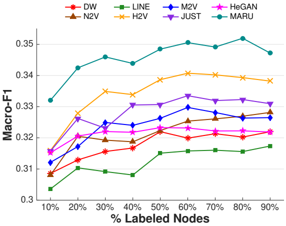
(a) DBIS (Micro-F1)



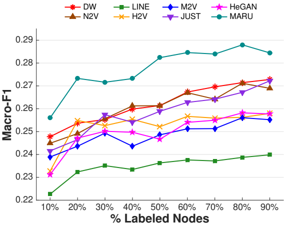
(b) MovieLens (Micro-F1)



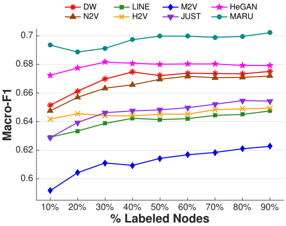
(c) Yelp (Micro-F1)



(d) DBIS (Macro-F1)

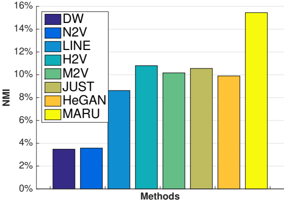


(e) MovieLens (Macro-F1)

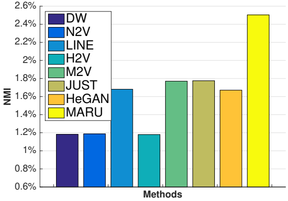


(f) Yelp (Macro-F1)

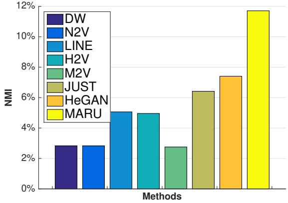
Node Clustering



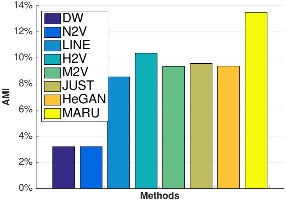
(g) DBIS (NMI)



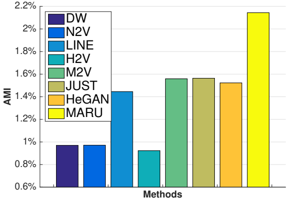
(h) MovieLens (NMI)



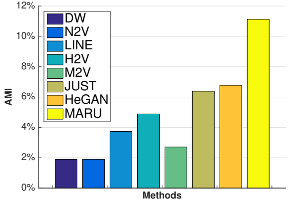
(i) Yelp (NMI)



(j) DBIS (AMI)



(k) MovieLens (AMI)



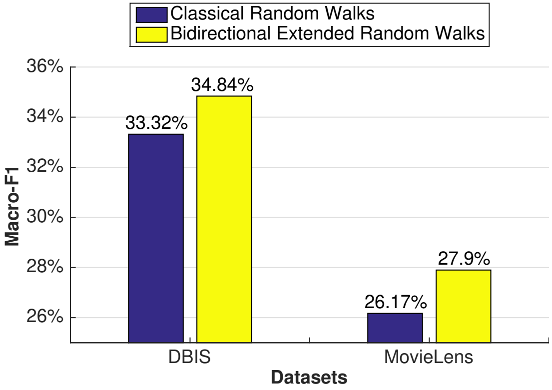
(l) Yelp (AMI)

Link Prediction

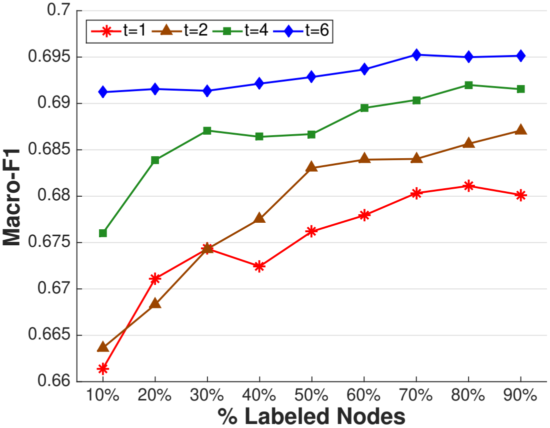
AUC with different operators for feature engineering.

Method	Operator	DBIS	MOVIE	YELP
DeepWalk	Hadamard	0.6367	0.9110	0.7330
	Weighted-L2	0.6094	0.7904	0.6872
node2vec	Hadamard	0.6362	0.9060	0.6622
	Weighted-L2	0.6292	0.7968	0.6848
LINE	Hadamard	0.5001	0.8631	0.5689
	Weighted-L2	0.5751	0.7611	0.6229
HIN2Vec	Hadamard	0.8028	0.9651	0.8117
	Weighted-L2	0.7240	0.7885	0.7137
metapath2vec	Hadamard	0.6778	0.9151	0.7372
	Weighted-L2	0.7363	0.6996	0.8240
JUST	Hadamard	0.6463	0.9119	0.7453
	Weighted-L2	0.6260	0.7845	0.6009
HeGAN	Hadamard	0.9597	0.9207	0.6361
	Weighted-L2	0.6714	0.7970	0.7289
MARU	Hadamard	**0.9979**	**0.9963**	0.7241
	Weighted-L2	0.7468	0.7979	**0.8315**

Effectiveness of Bidirectional Extended Random Walk



Meta-context Space Size vs. Performance



Conclusions

- Proposed MARU to learn representations for heterogeneous networks.
- Bidirectional extended random walks improves tail node embeddings.
- Meta-context aware skip-gram model dynamically learns the representations without any prior knowledge or manipulation.
- Extensive experiments demonstrate the significant improvements of MARU against state-of-the-art baselines in three practical evaluation tasks with three real-world datasets.
- Two analyses also show the effectiveness of our proposed techniques.

Thank you!
Ask me questions on QA sessions and `jyunyu AT cs.ucla.edu`
Personal website: `https://jyunyu.csie.org/`