Scalable Manifold-Regularized Attributed Network Embedding via Maximum Mean Discrepancy

Jun Wu, Jingrui He

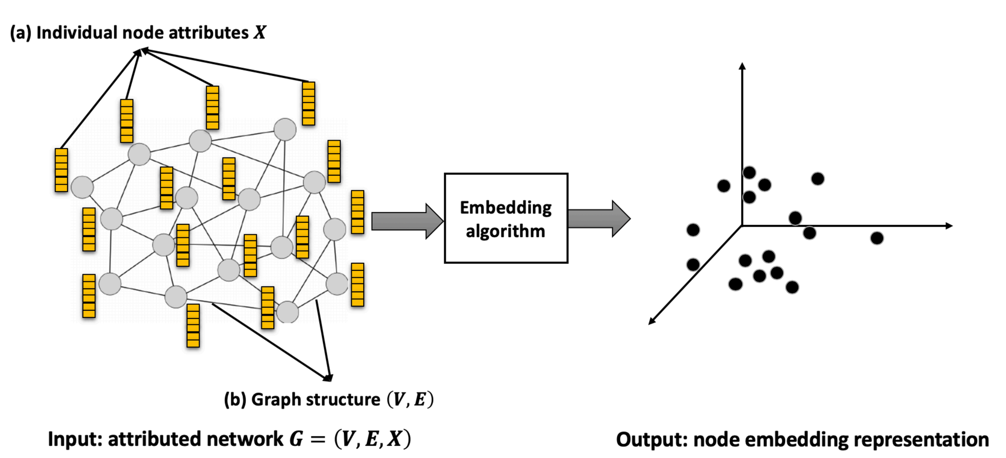
University of Illinois at Urbana-Champaign

junwu3@illinois.edu, jingrui@illinois.edu

Target Problem

Unsupervised Attributed Network Embedding:

- ***Input:*** An attributed network $G=(V,E,X)$ with adjacent matrix $A\in\mathbb{R}^{|V|\times|V|}$ and node attributes $X\in\mathbb{R}^{|V|\times D}$
- ***Output:*** A low-dimensional vector representation $u_i\in\mathbb{R}^d$ for every node $v_i\in V$ where node representation consistently encodes both the graph structure and individual node attributes
- ***Application:*** node classification, node clustering, link prediction, etc.



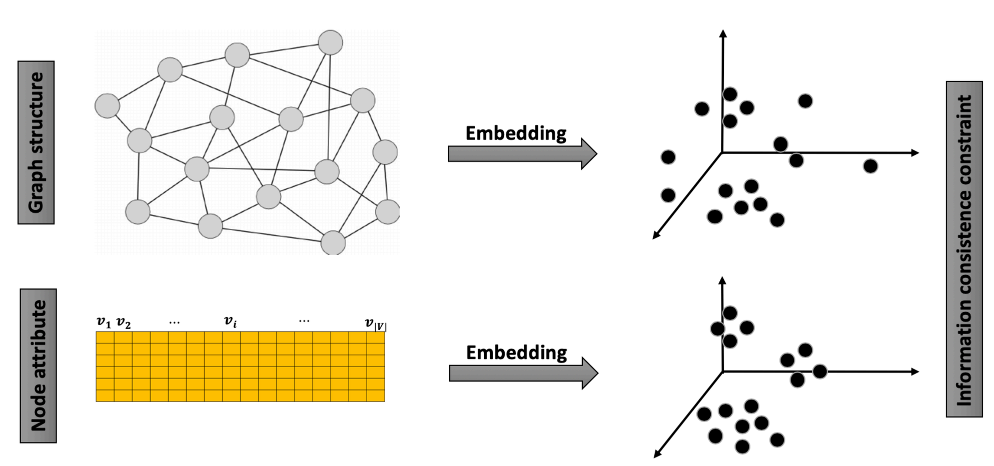
Challenges:

- ***Information heterogeneity:*** graph topological structure and individual node attributes
- ***Long-range spatial dependency:*** each node is associated with others within the connected component of the graph
- ***Scalability:*** real networks are induced by millions of nodes and edges or high-dimensional node attributes

Motivation

Assumption:

- The learned node representation from heterogeneous graph information should share similar data distribution



- The objective function is formulated as follow.

$$\mathcal{L}(U)=\boxed{L_{reg}(A,U)}+\lambda\,\boxed{MMD^2(U,X)}$$

Graph Laplacian regularization *(encode the graph structure information)*

Information discrepancy *between the structure- and attribute-based representation distribution in a Reproducing Kernel Hilbert Space (RKHS)*

Method

Manifold-Regularized Network Embedding (MARINE)

- The objective function for the proposed model is:

$$\mathcal{L}(U)=\underbrace{\boxed{L_{reg}(A,U)}}_{\text{Graph Laplacian}}+\lambda\overbrace{\boxed{\sum_{i=1}^{|V|}\|f(u_i)-x_i\|_2^2}}^{\text{Information consistence}}+\eta\underbrace{\boxed{\mathcal{R}(f)}}_{\text{Parameter regularization}}$$

- ***Graph Laplacian term:*** it encodes the global graph structure by assuming that two linked node have the similar embedding representation

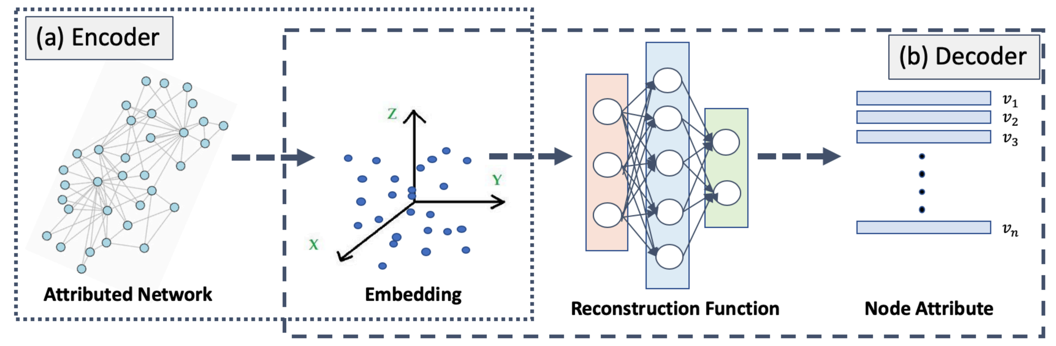
$$L_{reg}(A,U)=\frac{1}{2}\sum_{i,j}a_{ij}\left\|\frac{u_i}{\sqrt{D_{ii}}}-\frac{u_j}{\sqrt{D_{jj}}}\right\|_2^2,\qquad D_{ii}=\sum_j a_{ij}$$

- ***Information consistence:*** the node embedding representation shared similar data distribution with individual node attributes. Specifically, there ***exists a mapping function*** $f(\cdot)$ to reconstruct the node attribute from the learned representation

$$\begin{aligned}MMD^2(U,X)&=\left\|\frac{1}{|V|}\sum_{i=1}^{|V|}\phi(u_i)-\frac{1}{|V|}\sum_{i=1}^{|V|}\phi(x_i)\right\|_{\mathcal{H}}^2\\&\le\frac{1}{|V|}\sum_{i=1}^{|V|}\|\phi(u_i)-\phi(x_i)\|_{\mathcal{H}}^2\\&\le\frac{1}{|V|}\sum_{i=1}^{|V|}\|f(u_i)-x_i\|_{\mathcal{H}}^2\end{aligned}$$

Alternative Interpretations

- Encode-decoder architecture:



- Matrix factorization: when $f(u)=Wu$, it can be seen as a matrix factorization problem on node attributes X associated with the manifold regularization loss that captures the graph structure information.
- Feature diffusion (propagation): when $f(u)=u$, it has the closed-form solution as follow.

$$U=\underbrace{\boxed{\frac{\lambda}{1+\lambda}\left(I_{|V|\times|V|}-\frac{1}{1+\lambda}S\right)^{-1}}}_{\text{Manifold ranking scores measuring the node similarity on the graph}}X,\qquad S=D^{-1/2}AD^{-1/2}$$

Manifold ranking scores measuring the node similarity on the graph

Results

❖ **Data Sets**

	# nodes	# edges	# classes	# attributes
Cora	2, 708	5, 429	7	1, 433
Citeseer	3, 327	4, 732	6	3, 703
Wiki	2, 405	17, 981	19	4, 973
Pubmed	19, 717	44, 338	3	500

Table 1: Statistics of four publicly available networks

❖ **Performance Evaluation**

❑ ***Node Classification***

	Cora	Citeseer	Pubmed	Wiki
LE [1]	0.816	0.566	0.794	0.679
DeepWalk [9]	0.799	0.573	0.795	0.658
GraRep [3]	0.787	0.572	0.802	0.659
TADW [12]	0.777	0.717	0.859	0.795
AANE [8]	0.842	0.724	0.854	0.778
Graph2Gauss [2]	0.815	0.712	0.804	0.577
MARINE	**0.858**	**0.758**	**0.872**	**0.820**

Table 2: Node classification on real networks (the accuracy is reported)

❑ ***Node Clustering***

	Citeseer			Pubmed		
	NMI	AC	ARI	NMI	AC	ARI
LE [1]	0.168	0.323	0.042	0.208	0.509	0.087
DeepWalk [9]	0.180	0.408	0.131	0.293	0.656	0.284
GraRep [3]	0.076	0.292	0.040	0.090	0.477	0.069
TADW [12]	0.402	0.662	0.402	0.243	0.623	0.219
AANE [8]	0.280	0.498	0.219	0.292	0.631	0.269
Graph2Gauss [2]	0.372	0.624	0.381	0.268	0.678	0.278
MARINE	**0.427**	**0.685**	**0.432**	**0.357**	**0.696**	**0.334**

Table 3: Node clustering on real networks (K-means clustering is applied for node clustering using the learned node representation, and three popular measures are employed to evaluate the clustering performance including Normalized Mutual Information (NMI), clustering Accuracy (AC) and Adjusted Rand Index (ARI).)

❑ ***Link Prediction***

	Cora	Citeseer	Pubmed	Wiki
TADW [12]	0.871	0.816	0.907	0.509
AANE [8]	0.736	0.816	0.452	0.725
Graph2Gauss [2]	0.842	0.847	0.802	0.752
MARINE	**0.982**	**0.988**	**0.976**	**0.871**

Table 4: Link prediction on real networks (the AUC score is reported)

Higher is better for all the experimental results

Results

❖ **Efficiency Analysis**

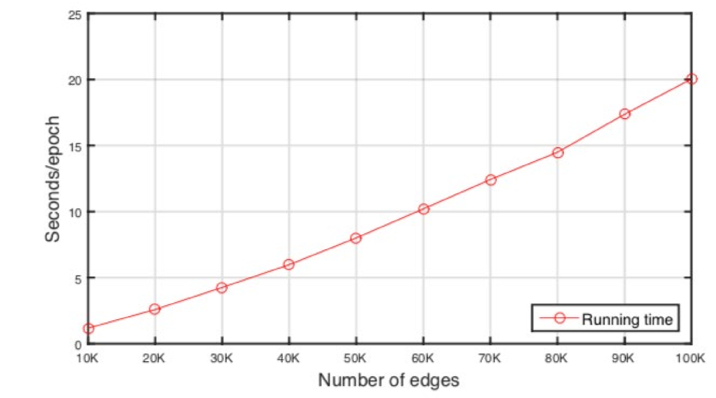


Figure 1: Scalability analysis on a synthetic network (It scales linearly with respect to the number of edges in the network)

❖ **Sensitivity Analysis**

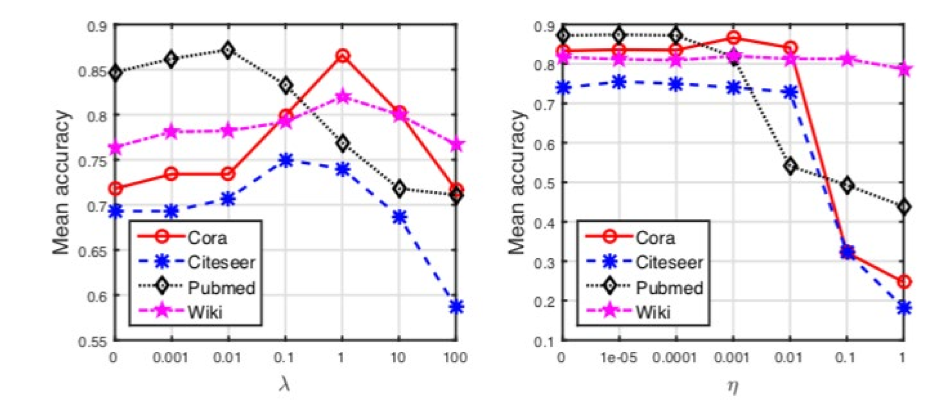


Figure 2: Parameter sensitivity analysis on real networks

Conclusions

- ➢ Association of attribute network embedding with a RKHS to minimize the information discrepancy in the attributed network
- ➢ A novel unsupervised attributed network embedding algorithm (MARINE) for encoding the heterogeneous graph information
- ➢ Extensive results demonstrating the proposed MARINE method on several graph mining tasks

Acknowledgments

This work is supported by the United States Air Force and DARPA under contract number FA8750-17-C-0153, National Science Foundation under Grant No. IIS-1552654 and Grant No. IIS-1813464, the U.S. Department of Homeland Security under Grant Award Number 17STQAC00001-02-00, and an IBM Faculty Award. The views and conclusions are those of the authors and should not be interpreted as representing the official policies of the funding agencies or the government.

ILLINOIS