Attributed Network Embedding

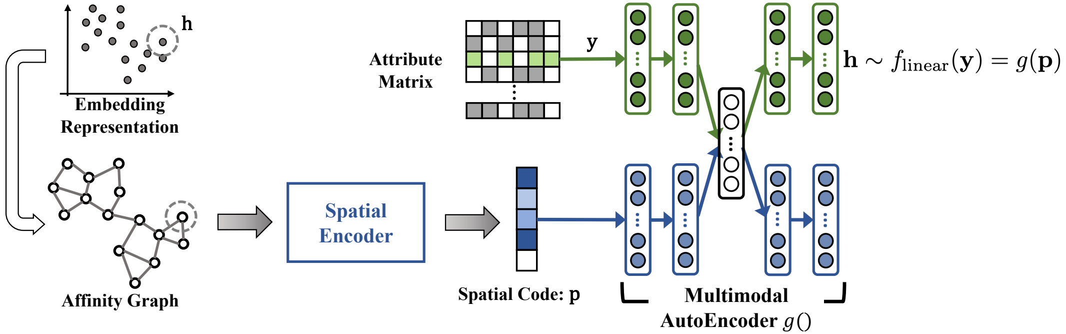
- **Motivations & challenges**

- **Mining attributed networks with shallow embedding**
  - Coupled spectral embedding
  - Coupled matrix & tri-factorization
  - Random walk based embedding

- **Mining attributed networks with deep embedding**
  - Objective function based deep embedding
  - Graph neural networks

- **Human-centric network analysis**
  - Interpretable node representation learning
  - Attributed network analysis with humans in the loop

Interpretable node representation learning



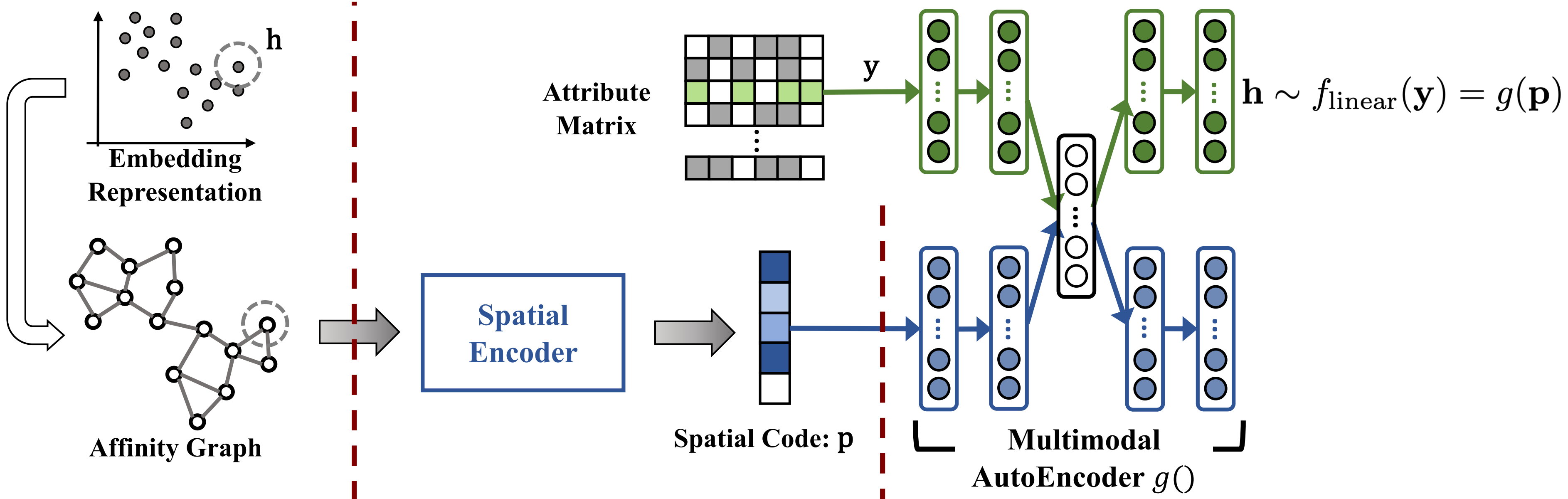
[Liu et al. WSDM, 2019]

- Opacity of embedding space
  - How representation vectors distribute in the embedding space?
  - What information is encoded in different embedding space regions?
  - Existing methods for explaining classifiers are not directly applicable

- Comprehensible node attributes are available

- Goal: Mining explainable structures and identifying characteristic factors from the mass of representation vectors

Spatial encoding and multimodal analytics

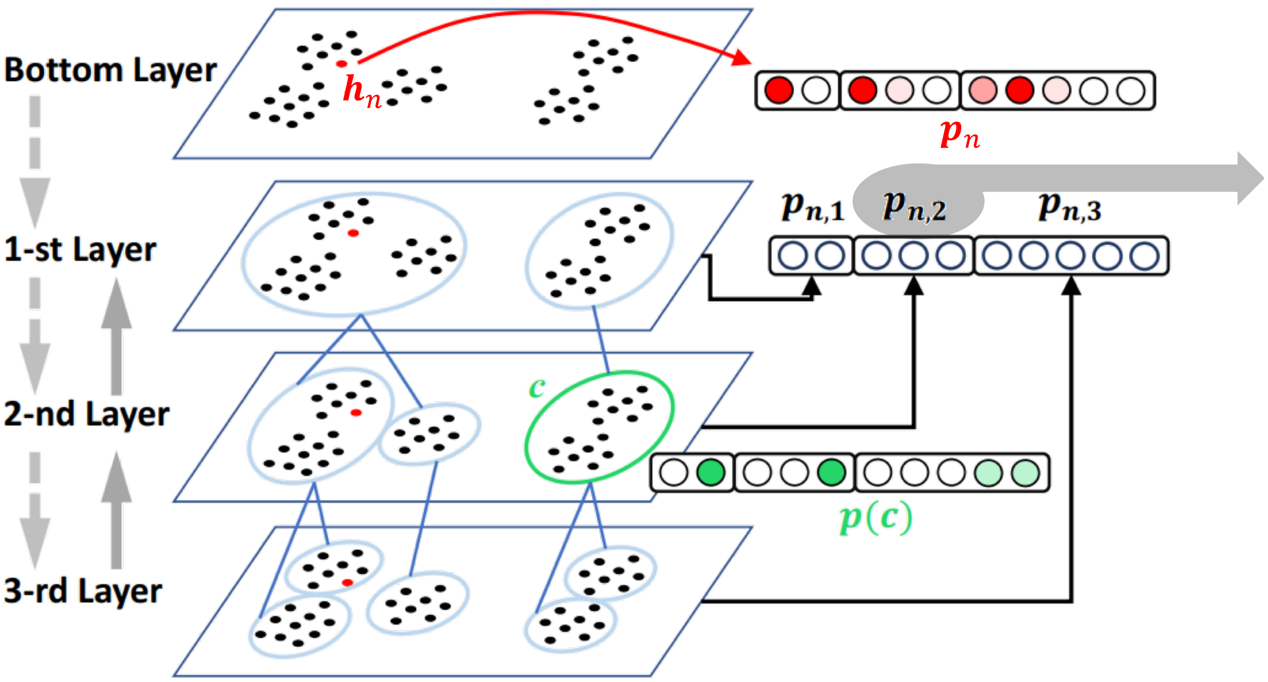


Affinity Graph Construction

Encoding Spatial Position

Interpretation Generation

Spatial encoding



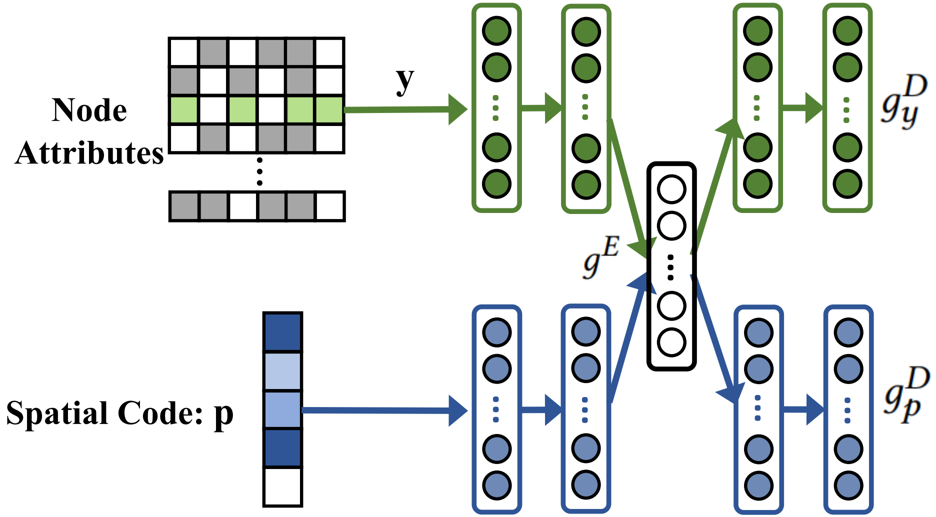
$$\mathbf{p}_{n,l} = \mathbf{e}_n \mathbf{W}_L \mathbf{W}_{L-1} \ldots \mathbf{W}_l$$

$\mathbf{e}_n \in \mathbb{R}^N$ is a one-hot vector, where $\mathbf{e}_n(i) = 1$ for $i = n$

The spatial code for node n is $\mathbf{p}_n = [\hat{\mathbf{p}}_{n,1}, \hat{\mathbf{p}}_{n,2}, \ldots, \hat{\mathbf{p}}_{n,L-1}, \hat{\mathbf{p}}_{n,L}]$

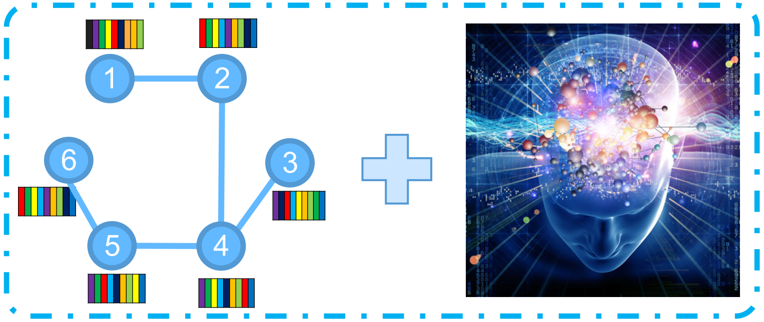
Multimodal autoencoder

- $\boldsymbol{y}$ are comprehensible node attributes
- Variational autoencoder is used to reconstruct $\boldsymbol{y}$ and $\boldsymbol{p}$
- After training the autoencoder, the interpretation for embedding representation $\boldsymbol{h}$ is,
  - $\mathbf{h} \sim f_{\text{linear}}(\mathbf{y}) = g(\mathbf{p}) = g_y^D \circ g^E(\mathbf{p}, \mathbf{0})$
  - The input to the node attribute side is set to be absent
  - The output from node attribute decoder is used as the interpretation



Attributed network analysis with humans in the loop

Initial Attributed Network



Embedding Representation

$$\mathbf{H} = \begin{bmatrix} 0.54 & 0.27 \\ 0.22 & 0.91 \\ 0.55 & 0.28 \\ 0.98 & 0.11 \\ 0.32 & 0.87 \\ 0.26 & 0.11 \end{bmatrix}$$

Tasks

- Classification
- Clustering
- Link Prediction
- Visualization
- …

[Huang et al. WSDM, 2018]

- Attributed network embedding (ANE) serves as infrastructures of various real-world applications
- We aim to learn cognition from experts and incorporate it into ANE to advance downstream analysis algorithms

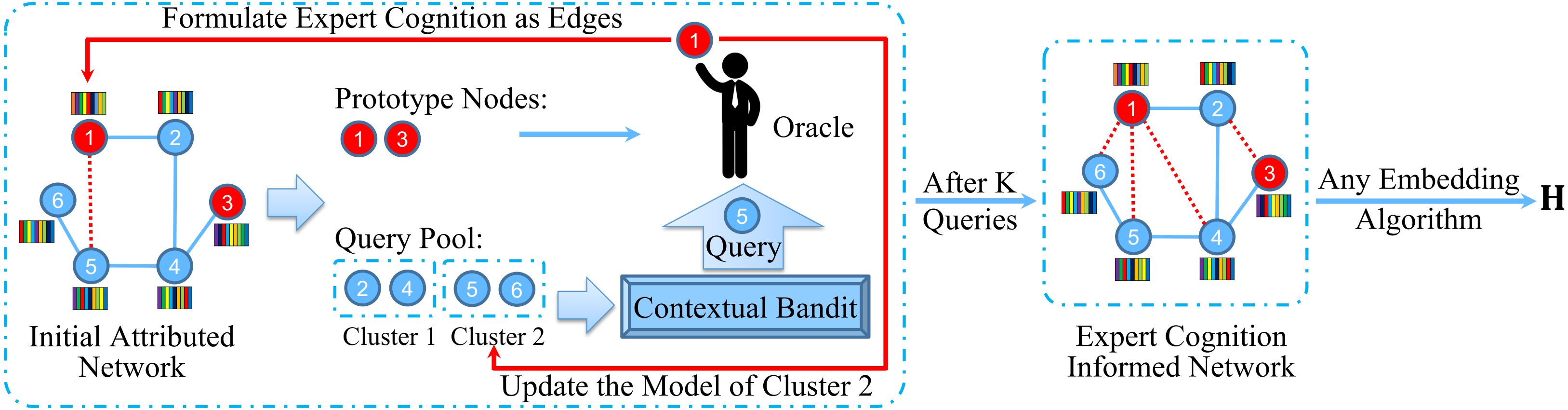
Expert cognition benefits data analysis

- Definition: Meaningful and Intelligence-related info that experts know beyond the data

- Understanding of domain knowledge
- Awareness of conventions
- Perception of latent relations

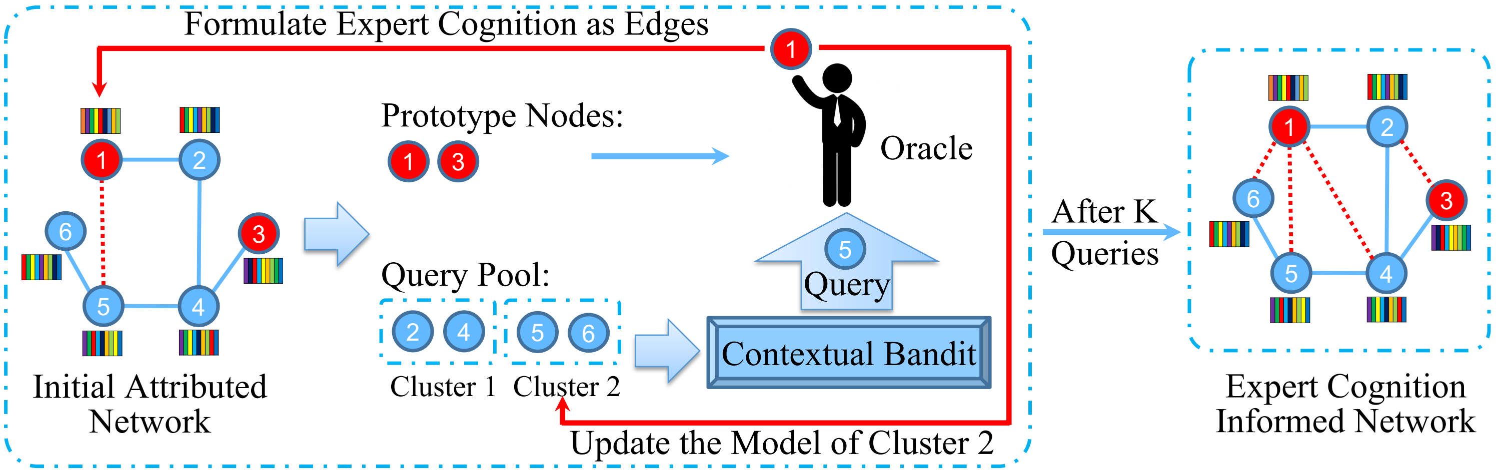
- Example: Human understand the sentiment in product reviews. This cognition could be applied to enhance the recommendations

Network embedding with expert cognition - NEEC



- Convert the abstract and meaningful cognition of domain experts into concrete answers
- Incorporate answers into ANE towards a more informative $\mathbf{H}$
- Employ a general and concise form of queries to learn expert cognition from the oracle while greatly saving his/her effort

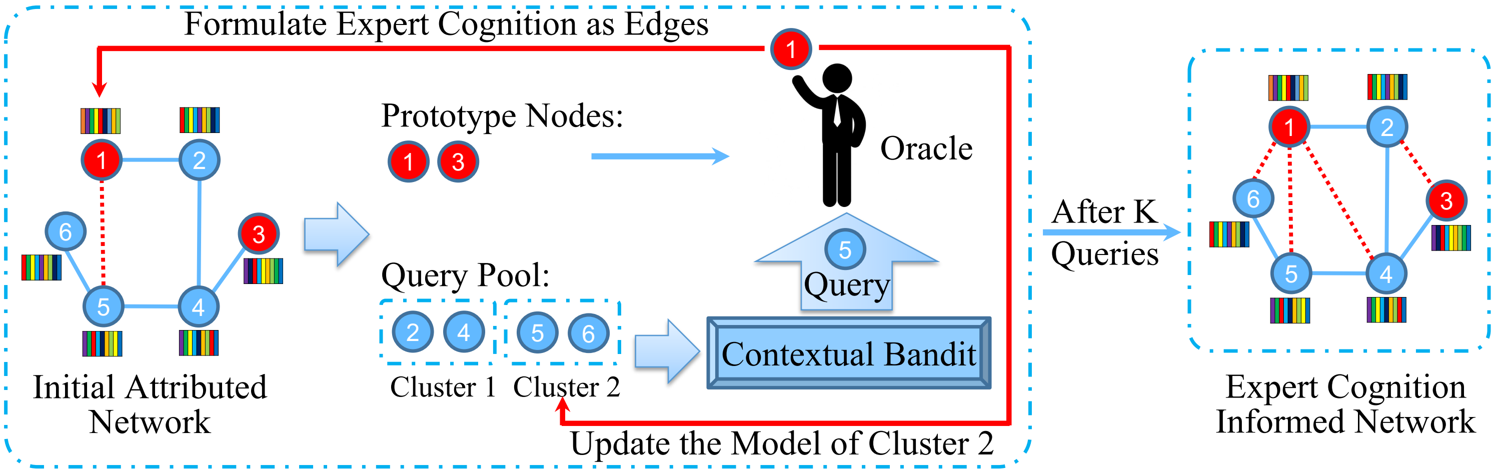
Strategies of framework NEEC



- Two steps to find the top K meaningful queries
  - Find few representative and distinct nodes (in red) as prototypes
  - Iteratively select K nodes from the remaining nodes (in blue) with the largest amount of expected learned expert cognition

- Oracle needs to indicate a node from the prototypes (e.g., $j = 1$) that is the most similar to the queried node $i = 5$

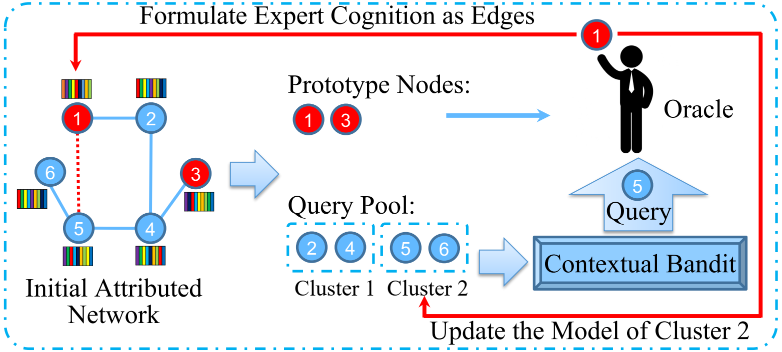
Strategies of framework NEEC



- Answers will be added into the network structure in the form of weighted edges, named as cognition edges (red dotted lines)
- With these cognition edges, different ANE methods can be directly applied to the expert cognition informed network to learn $\mathbf{H}$

Human-centric network analysis

- Focuses:
  Interpretable embedding, & utilizing network embedding to incorporate human knowledge
- Methods:
  Interpretable node representation learning
  Attributed network analysis with humans in the loop
- Techniques:
  Linking embedding with interpretable node attributes, converting knowledge into links, etc.



Summary of attributed network embedding

- ANE learns low-dimensional vectors to represent all nodes, bridging the gap between real-world systems & ML algorithms
- Challenges: Heterogeneity, large-scale, & Data Characteristics Vary Significantly
- Compare with other research topics
  - Multiview learning: Learn a unified representation of instances from multiple feature matrices observed from different aspects
  - Multimodal learning: Embed multiple sources with distinct modalities such as networks, images, and audio
  - Attributed network embedding: Preserve proximity information in networks and (one or multiple types of) node attributes

Summary of Attributed Network Embedding

- Shallow attributed network embedding:
  - Coupled spectral embedding
  - Coupled matrix & tri-factorization
  - Random walk based embedding
- Deep attributed network embedding:
  - Objective function based deep embedding
  - Graph neural networks
- Comprehensible node attributes help humans interact with systems.
  - Interpretable node representation learning
  - Attributed network analysis with humans in the loop

Acknowledgments

- DATA Lab and collaborators

COMPUTER SCIENCE & ENGINEERING
TEXAS A&M UNIVERSITY

Arizona State University

DATA LAB

- Funding agencies
  - National Science Foundation
  - Defense Advanced Research Projects Agency

- Everyone attending the talk

References

- Hongchang Gao and Heng Huang. 2018. Deep Attributed Network Embedding. In IJCAI. 3364–3370.
- Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In NIPS. 1024–1034.
- Xiao Huang, Jundong Li, and Xia Hu. 2017. Accelerated Attributed Network Embedding. In SDM. 633–641.
- Xiao Huang, Jundong Li, and Xia Hu. 2017. Label Informed Attributed Network Embedding. In WSDM. 731–739.
- Xiao Huang, Jundong Li, Na Zou, and Xia Hu. 2018. A General Embedding Framework for Heterogeneous Information Learning in Large-Scale Networks. TKDD 12 (2018).
- Xiao Huang, Qingquan Song, Jundong Li, and Xia Hu. 2018. Exploring Expert Cognition for Attributed Network Embedding. In WSDM. 270–278.
- Xiao Huang, Qingquan Song, Fan Yang, and Xia Hu. 2019. Large-Scale Heterogeneous Feature Embedding. In AAAI.
- Xiao Huang, Qingquan Song, Yuening Li, and Xia Hu. 2019. Graph Recurrent Networks with Attributed Random Walks. In KDD.
- Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In ICLR.
- David Hallac, Jure Leskovec, and Stephen Boyd. 2015. Network Lasso: Clustering and Optimization in Large Graphs. In KDD. 387–396.
- Abhishek Kumar, Piyush Rai, and Hal Daume. 2011. Co-regularized Multi-view Spectral Clustering. In NIPS. 1413–1421.
- Jundong Li, Harsh Dani, Xia Hu, Jiliang Tang, Yi Chang, and Huan Liu. 2017. Attributed Network Embedding for Learning in a Dynamic Environment. In CIKM. 387–396.
- Jiongqian Liang, Peter Jacobs, Jiankai Sun, and Srinivasan Parthasarathy. 2018. Semi-Supervised Embedding In Attributed Networks With Outliers. In SDM. 153–161.
- Ninghao Liu, Xiao Huang, and Xia Hu. 2017. Accelerated Local Anomaly Detection via Resolving Attributed Networks. In IJCAI. 2337–2343.
- Shirui Pan, Jia Wu, Xingquan Zhu, Chengqi Zhang, and Yang Wang. 2016. Tri- Party Deep Network Representation. In IJCAI. 1895–1901.
- Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: Online Learning of Social Representations. In KDD. 701–710.
- Guo-Jun Qi, Charu Aggarwal, Qi Tian, Heng Ji, and Thomas S. Huang. 2012. Exploring Context and Content Links in Social Media: A Latent Space Method. TPAMI 34, 5 (2012), 850–862.
- Ulrike von Luxburg. 2007. A Tutorial on Spectral Clustering. Statistics and Computing 17, 4 (2007), 395–416.

- Hongchang Gao and Heng Huang. 2018. Deep Attributed Network Embedding. In IJCAI. 3364–3370.

- Lizi Liao, Xiangnan He, Hanwang Zhang, and Tat-Seng Chua. 2018. Attributed Social Network Embedding. TKDE 30, 12 (2018), 2257–2270.
- Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y. Chang. 2015. Network Representation Learning with Rich Text Information. In IJCAI. 2111– 2117.
- Shenghuo Zhu, Kai Yu, Yun Chi, and Yihong Gong. 2007. Combining Content and Link for Classification Using Matrix Factorization. In SIGIR. 487–494.
- Zhen Zhang, Hongxia Yang, Jiajun Bu, Sheng Zhou, Pinggang Yu, Jianwei Zhang, Martin Ester, and Can Wang. 2018. ANRL: Attributed Network Representation Learning via Deep Neural Networks. In IJCAI. 3155–3161.