Accelerated Attributed Network Embedding

Xiao Huang, † Jundong Li, ‡ and Xia (Ben) Hu †

†Computer Science & Engineering, Texas A&M University, College Station, TX, USA
‡Computer Science & Engineering, Arizona State University, Tempe, AZ, USA

Emails: {xhuang,xiahu}@tamu.edu, jundongl@asu.edu
What is Network Embedding

Network

$n$-dimensional

Embedding Representation

\[
H = \begin{bmatrix}
0.54 & 0.27 & n_1 \\
0.22 & 0.91 & n_2 \\
0.55 & 0.28 & n_3 \\
0.98 & 0.11 & n_4 \\
0.32 & 0.87 & n_5 \\
0.26 & 0.11 & n_6 \\
\end{bmatrix}
\]

d \ll n

Application

- Classification
- Clustering
- Link Prediction
- Visualization
- ...

- Learn a low-dimensional vector representation for each node, such that all the geometrical structure information is preserved.

- Similar nodes have similar representations, and the informative latent space benefits real-world applications.
What is Attributed Network

- In real-world information systems, nodes are not just vertices.

- Both node-to-node dependencies & node attribute information are available.
Why Attributes Benefit Embedding

- Node attributes are rich and informative.
- Homophily & social influence: network and node attributes influence each other and are inherently correlated.

- High correlation of user posts and following relationships.
- Strong association between paper topics and citations.
Major Challenges

- Hard to jointly assessing node proximity from heterogeneous information.
  - Node attribute information such as text is distinct from network topological structure.

- Number of nodes and dimension of attributes could be large.
  - Classical algorithms such as eigen-decomposition and gradient descent cannot be applied.
  - It might be expensive to store or manipulate the high-dimensional matrices such as node attribute similarity.
Define Attributed Network Embedding

Given $W$ and $A$, we aim to represent each node as a $d$-dimensional row $h_i$, such that $H$ can preserve node proximity both in network and node attributes.

Nodes with similar topology or attributes would have similar representations.
Major Contributions

- Propose a scalable framework AANE to jointly learn node proximity from network and node attributes.

- Present a distributed optimization algorithm to accelerate by decomposing the task into low complexity sub-problems.

- Strategies for filling the gap:
  
  I. Assimilate the two information in the similarity space to tackle heterogeneity, but without calculating the network similarity matrix.

  II. Avoid high-dimensional matrix manipulation.

  III. Make sub-problems independent to each other to allow parallel computation.
Based on the decomposition of attribute similarity and penalty of embedding difference between connected nodes.

\[
\min_H \ J = \|S - HH^\top\|_F^2 + \lambda \sum_{(i,j) \in E} w_{ij} \|h_i - h_j\|_2^2
\]

- $\ell_2$ norm alleviates the impacts from outliers and missing data.
- Fused lasso clusters the network without similarity matrix.
- $\lambda$ adjusts the size of clustering group.
Framework AANE: Strategy II

- Make a copy of $\mathbf{H}$ and reformulate into a linearly constrained problem.

$$
\min_{\mathbf{H}} \sum_{i=1}^{n} \| \mathbf{s}_i - \mathbf{h}_i \mathbf{Z}^T \|_2^2 + \lambda \sum_{(i,j) \in \mathcal{E}} w_{i,j} \| \mathbf{h}_i - \mathbf{z}_j \|_2,
$$

subject to $\mathbf{h}_i = \mathbf{z}_i, i = 1, \ldots, n$.

- Given fixed $\mathbf{H}$, all the row $\mathbf{z}_i$ could be calculated independently.
- Each sub-problem only needs row $\mathbf{s}_i$, not the entire $\mathbf{S}$.
- Time complexity of updating $\mathbf{h}_i$ is $\mathcal{O}(d^3 + dn + d|N(i)|)$, with space complexity $\mathcal{O}(n)$.
- Alternating Direction Method of Multipliers (ADMM) converges to a modest accuracy in a few iterations.
Framework AANE: Strategy III

Worker 1:

Problem 1 \( S_1 \) = \( ? \times Z^T \)

Problem 2 \( S_2 \) = \( ? \times Z^T \)

... Problem 5 \( S_5 \) = \( ? \times Z^T \)

... Problem 6 \( S_6 \) = \( ? \times Z^T \)

Updating

Problem 7 \( S_1^T \) \( H \) = \( ? \times \)

Problem 8 \( S_2^T \) \( H \) = \( ? \times \)

... Problem 11 \( S_5^T \) \( H \)

... Problem 12 \( S_6^T \) \( H \)
Experimental Setup

- Classification on three real-world network:
  - BlogCatalog
  - Flickr
  - Yelp

- Three types of baselines:
  - Scalable network embedding, DeepWalk & LINE.
  - Node attribute modeling based on PCA.
  - Attributed network representation learning, multispec & LCMF.
AANE achieves higher performance than the state-of-the-art embedding algorithms with different training percentage and latent dimension d.
Efficiency Evaluation

- AANE takes much less running time than the attributed network representation learning methods even with single-thread.

![Diagram showing running time comparison between LCMF, MultiSpec, and AANE for Flickr and Yelp datasets.](image)
Conclusions

- The proposed accelerated attributed network embedding (AANE) framework is scalable, efficient, and effective.

- Future work:
  - Embedding of large-scale and dynamic attributed networks.
  - Semi-supervised attributed network embedding.
Acknowledgement

- DATA Lab and collaborators

[Logo: Texas A&M University]

Data Analytics at Texas A&M (DATA Lab)

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

- Everyone attending the talk