Label Informed 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**

- Preserve the geometrical structure by mapping each node into a continuous low-dimensional vector space
- Pave the way for numerous applications

Network | Adjacency Matrix | Embedding Representation $H$ | Application
---|---|---|---

$n$-
dimensional

$\begin{array}{cccccc}
  n_1 & n_2 & n_3 & n_4 & n_5 & n_6 \\
  n_1 & 1 & 0 & 0 & 1 & 0 \\
  n_2 & 0 & 1 & 1 & 0 & 0 \\
  n_3 & 0 & 1 & 1 & 1 & 0 \\
  n_4 & 0 & 1 & 1 & 0 & 1 \\
  n_5 & 0 & 1 & 1 & 0 & 1 \\
  n_6 & 0 & 0 & 0 & 0 & 0 \\
\end{array}$

$\begin{array}{cccc}
  n_1 & n_2 & n_3 & n_4 \\
  n_1 & 1 & 0 & 0 \\
  n_2 & 0 & 1 & 0 \\
  n_3 & 0 & 0 & 1 \\
  n_4 & 0 & 0 & 0 \\
\end{array}$

$d \ll n$

- Classification
- Clustering
- Link Prediction
- Visualization
- …

Source: http://www.perozzi.net/projects/deepwalk/
What is Attributed Network

- Powerful in modeling real-world information systems
- Network topological structure & node attribute information

Why Label Informed

- Abundant label info observed: group, community, category
- Labels and attributed network affect and depend on each other

Colors?

Same labels:
- Similar photos
- Interact with each other
Problem Statement

- Label Informed Attributed Network Embedding (LANE): leverage both labels and node proximity in attributed network to learn a more efficient latent representation

\[
\begin{bmatrix}
0.54 & 0.27 \\
0.22 & 0.91 \\
0.55 & 0.28 \\
0.98 & 0.11 \\
0.32 & 0.87 \\
0.26 & 0.11
\end{bmatrix}
\]

\( H = \)

Latent Space
Opportunities & Challenges

- **Labels are informative:**
  - They are strongly influenced by and inherently correlated to the attributed network
  - Jointly exploiting them with node proximity in attributed network benefits various data mining applications

- **Noise & Heterogeneity:**
  - Data could be sparse, incomplete and noisy
  - Label info is distinct from topological structure and node unique features
Major Contributions

- Propose a framework LANE that embeds nodes with similar network structure, attribute proximity, or same label into similar vector representations.
I. Collectively model network proximity and node attribute info via spectral technique

\[
\max_{U^{(G)}, U^{(A)}} J_G + \alpha_1 (J_A + \rho_1) = \text{Tr} \left( U^{(G)\top} L^{(G)} U^{(G)} + \alpha_1 U^{(A)\top} L^{(A)} U^{(A)} + \alpha_1 U^{(A)\top} U^{(G)} U^{(G)\top} U^{(A)} \right)
\]

II. Consider nodes with the same label as a clique, and employ the learned network proximity to smooth the label info

\[
\max_{U^{(G)}, U^{(Y)}} J_Y = U^{(Y)\top} (L^{(YY)} + U^{(G)} U^{(G)\top}) U^{(Y)}
\]

III. Uniformly and jointly model proximities of heterogeneous info
Experimental Results

- LANE and its variation outperform Original Features
- LANE achieves significantly better performance than the state-of-the-art embedding algorithms

![Graph showing F1 score vs Embedding Representation Dimension d]

Dimension=18, 107

LANE on Net
LANE w/o Label
Embedding Representation Dimension d

Experimental Results

Ø LANE and its variation outperform Original Features
Ø LANE achieves significantly better performance than the state-of-the-art embedding algorithms
Acknowledgement

- DATA Lab and collaborators

DATA Analytics at Texas A&M (DATA Lab)

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

- Everyone attending the talk