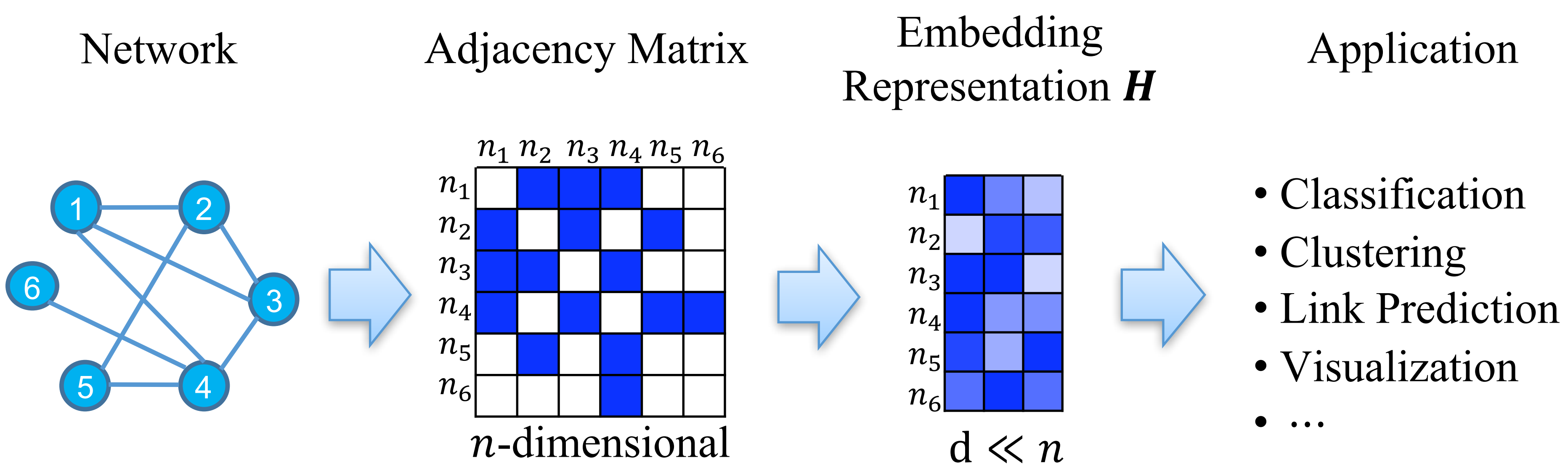
# Label Informed Attributed Network Embedding

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**

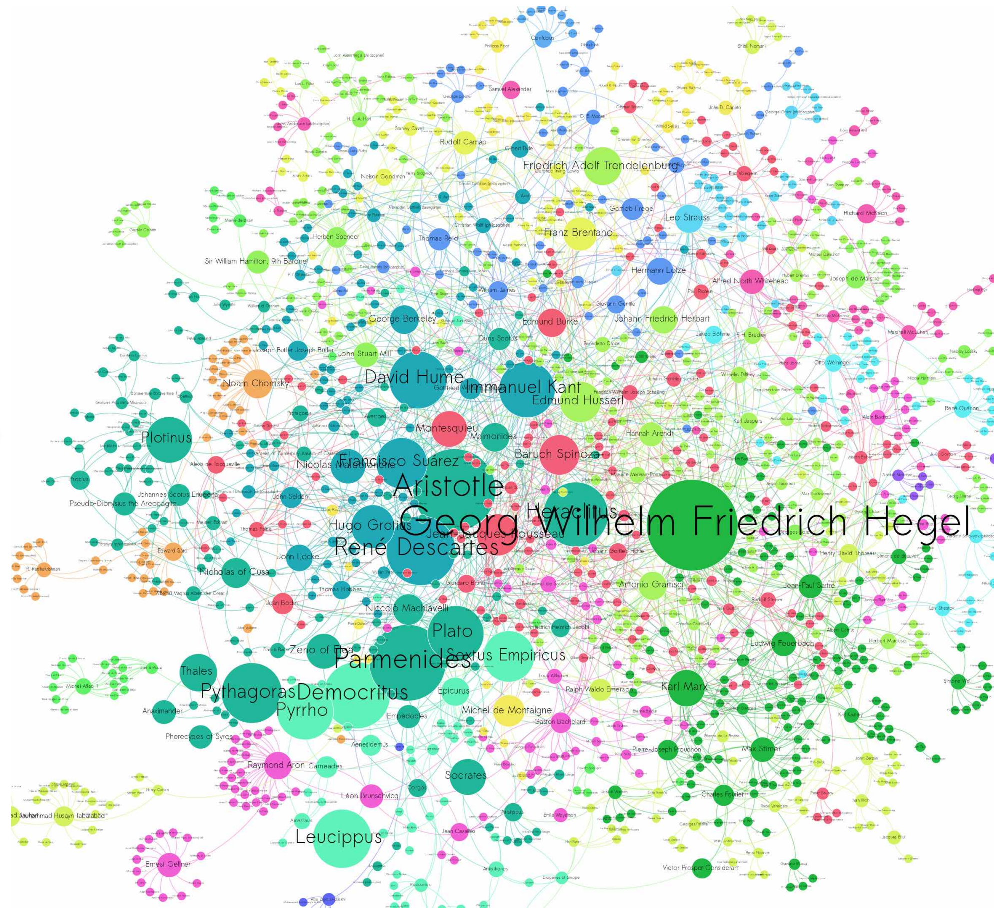
- Preserve the geometrical structure by mapping each node into a continuous low-dimensional vector space
- Pave the way for numerous applications




Source: http://www.perozzi.net/projects/deepwalk/

# What is **Attributed Network**

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

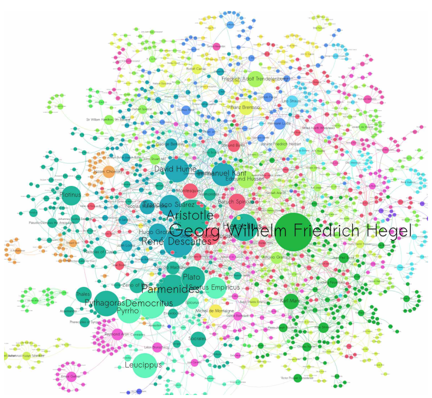



Source: http://www.designandanalytics.com/visualizing-the-history-of-philosophy-as-a-social-network-the-problem-with-hegel

# Why **Label Informed**

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

Colors ?

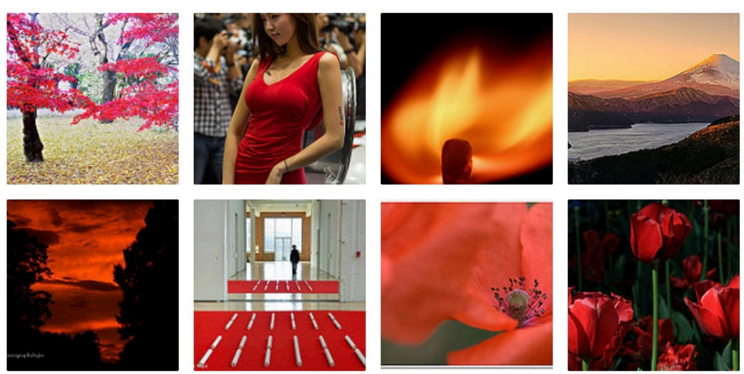



**Green is Beautiful**

88.1k 787.7k 29

**Red Rule**

56.4k 637.8k 152

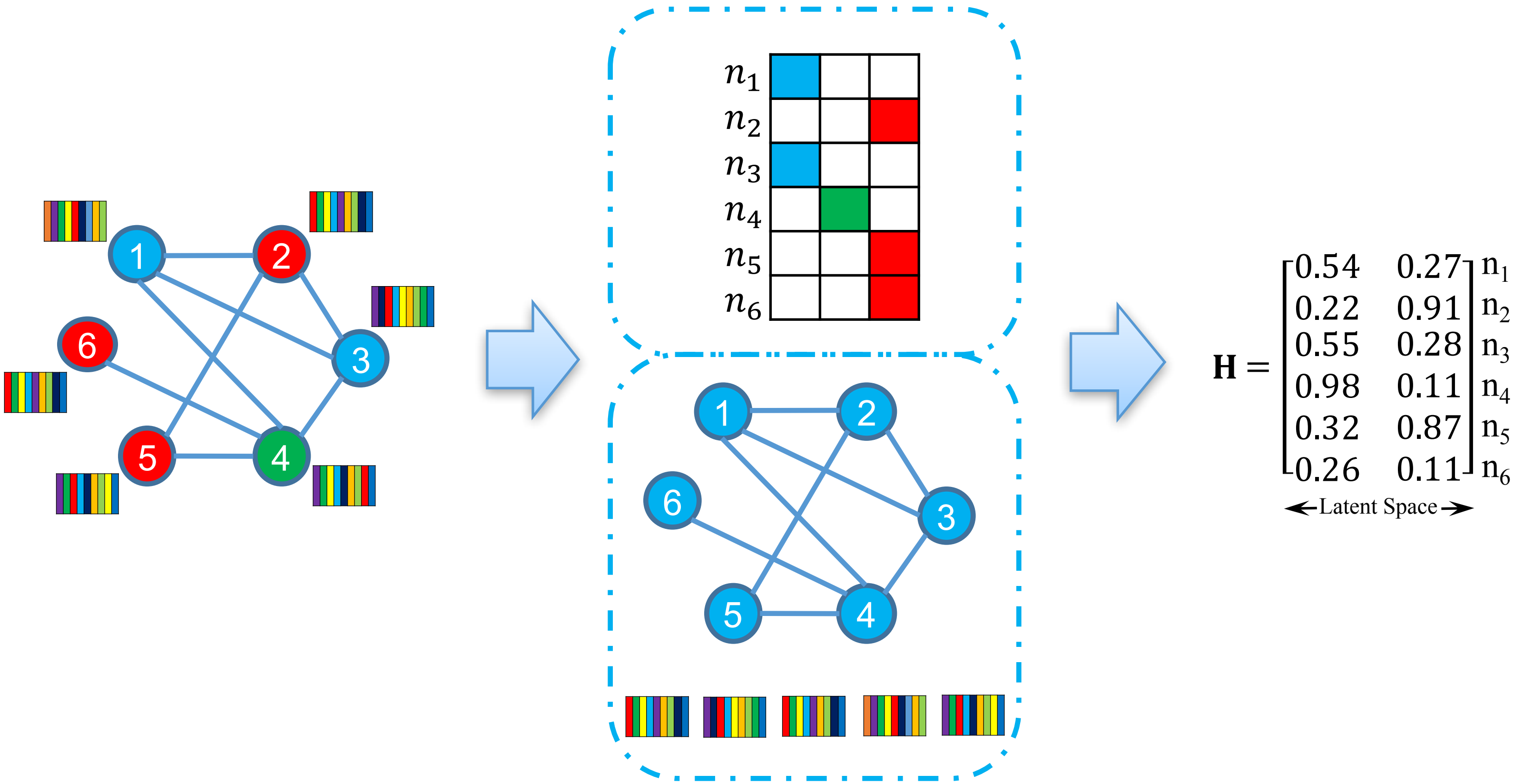


Same labels:

- 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



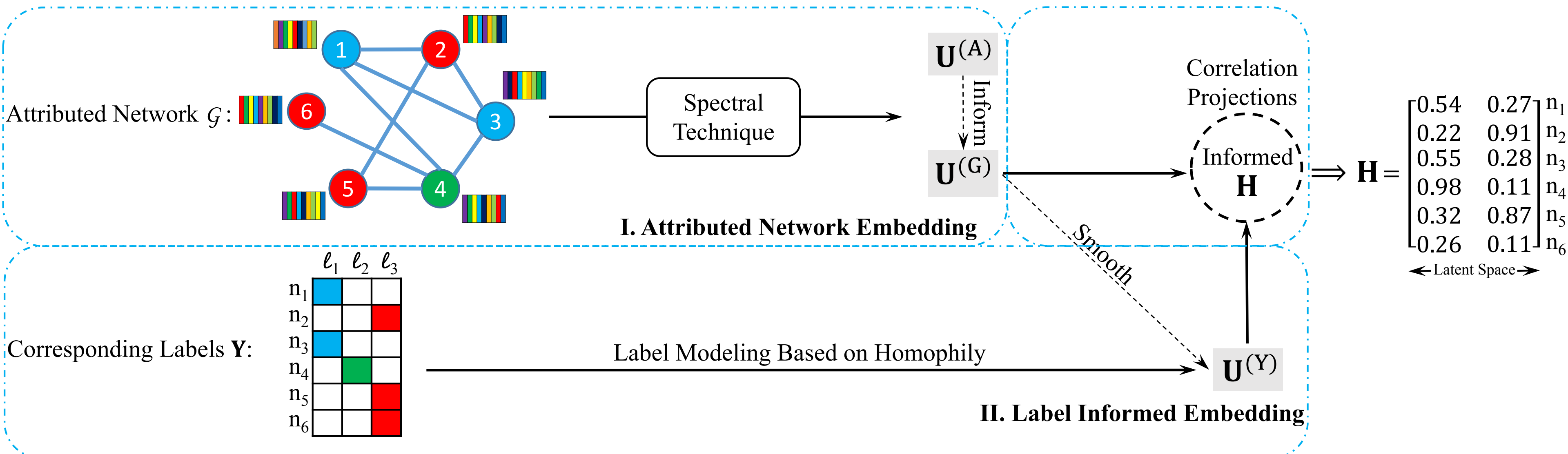
# Opportunities & Challenges

- 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**

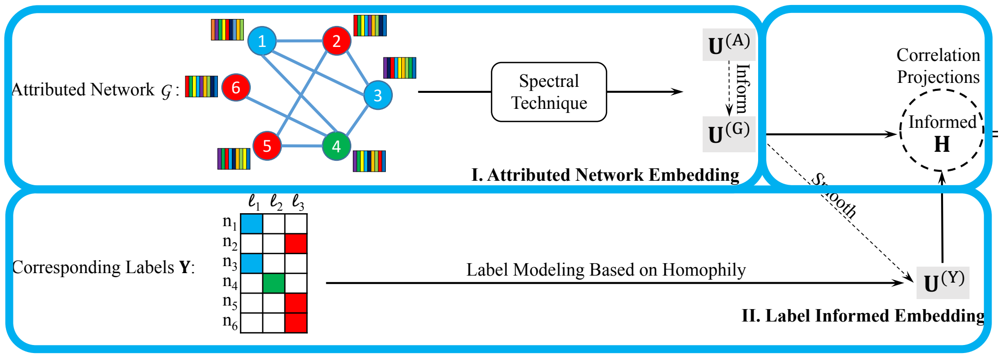
- Propose a framework LANE that embeds nodes with similar network structure, attribute proximity, or same label into similar vector representations



# Framework **LANE**

I

II




III

I. Collectively model network proximity and node attribute info via spectral technique

$$\underset{\mathbf{U}^{(G)},\mathbf{U}^{(A)}}{\text{maximize}} \quad \mathcal{J}_G + \alpha_1(\mathcal{J}_A + \rho_1) = \operatorname{Tr}\left(\mathbf{U}^{(G)\top}\mathcal{L}^{(G)}\mathbf{U}^{(G)} + \alpha_1\mathbf{U}^{(A)\top}\mathcal{L}^{(A)}\mathbf{U}^{(A)} + \alpha_1\mathbf{U}^{(A)\top}\mathbf{U}^{(G)}\mathbf{U}^{(G)\top}\mathbf{U}^{(A)}\right)$$

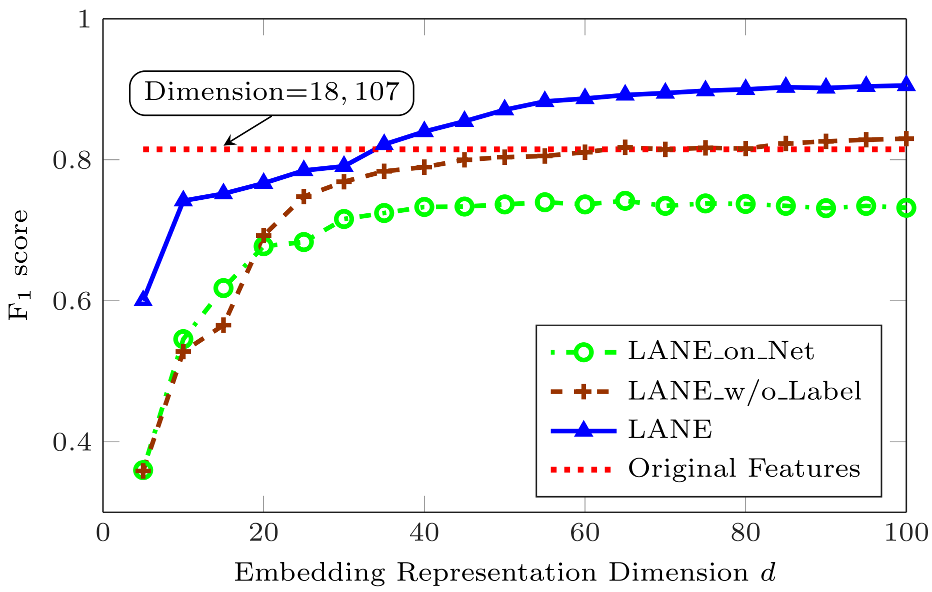
II. Consider nodes with the same label as a clique, and employ the learned network proximity to smooth the label info

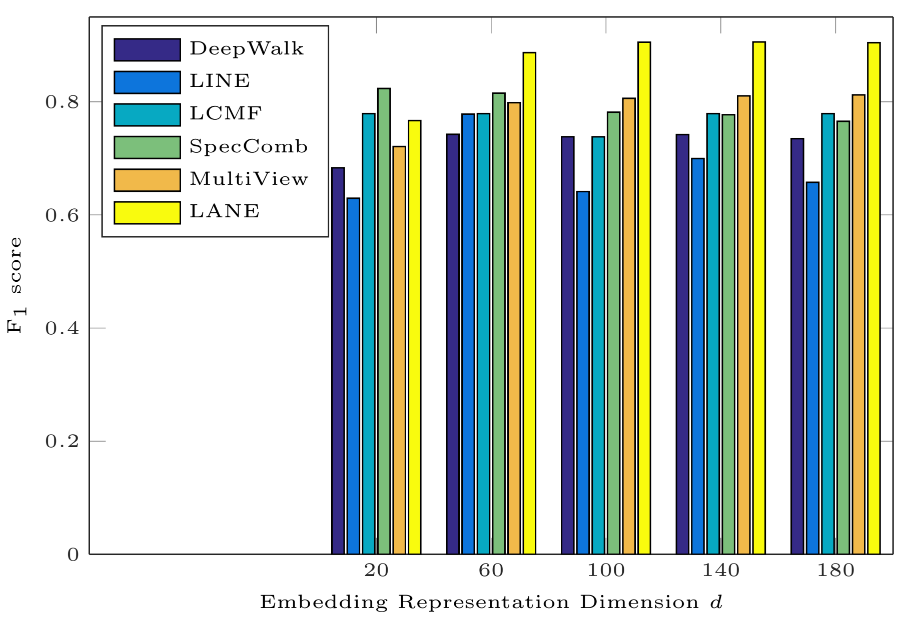
$$\underset{\mathbf{U}^{(G)},\mathbf{U}^{(Y)}}{\text{maximize}} \quad \mathcal{J}_Y = \mathbf{U}^{(Y)\top}(\mathcal{L}^{(YY)} + \mathbf{U}^{(G)}\mathbf{U}^{(G)\top})\mathbf{U}^{(Y)}$$

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



# Acknowledgement

➢ DATA Lab and collaborators

COMPUTER SCIENCE & ENGINEERING
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