Representation Learning on Networks

Slides based on WWW 2018 Tutorial on Representation Learning on Networks by Jure Leskovec, William L. Hamilton, Rex Ying, Rok Sosic (Stanford University)
snap.stanford.edu/proj/embeddings-www
Why networks?

Networks are a general language for describing and modeling complex systems.
Network!
Many Data are Networks

Social networks

Economic networks

Biomedical networks

Information networks: Web & citations

Internet

Networks of neurons
Why Networks? Why Now?

- Universal language for describing complex data
  - Networks from science, nature, and technology are more similar than one would expect
- Shared vocabulary between fields
  - Computer Science, Social science, Physics, Economics, Statistics, Biology
- Data availability (+computational challenges)
  - Web/mobile, bio, health, and medical
- Impact!
  - Social networking, Social media, Drug design
Machine Learning with Networks

Classical ML tasks in networks:

- Node classification
  - Predict a type of a given node
- Link prediction
  - Predict whether two nodes are linked
- Community detection
  - Identify densely linked clusters of nodes
- Network similarity
  - How similar are two (sub)networks
Example: Node Classification
Example: Node Classification

Classifying the function of proteins in the interactome!

Example: Link Prediction
Example: Link Prediction

Content recommendation is link prediction!
Machine Learning Lifecycle

- (Supervised) Machine Learning Lifecycle: This feature, that feature. Every single time!
Feature Learning in Graphs

Goal: Efficient task-independent feature learning for machine learning in networks!

\[ f: u \rightarrow \mathbb{R}^d \]

node \( u \) → vec

Feature representation, embedding
Example

Zachary’s Karate Club Network:

Input

Output

Image from: Perozzi et al. 2014. DeepWalk: Online Learning of Social Representations. KDD.
Why Is It Hard?

- Modern deep learning toolbox is designed for simple sequences or grids
  - CNNs for fixed-size images/grids...
  - RNNs or word2vec for text/sequences...
Why Is It Hard?

- But networks are far more complex!
  - Complex topographical structure (i.e., no spatial locality like grids)
  - No fixed node ordering or reference point (i.e., the isomorphism problem)
  - Often dynamic and have multimodal features.
Today’s Class

1) Node embeddings
   ▪ Map nodes to low-dimensional embeddings

2) Graph neural networks
   ▪ Deep learning architectures for graph-structured data

3) Reasoning over KGs