Representation Learning on Networks

Jure Leskovec, William L. Hamilton, Rex Ying, Rok Sosic
Stanford University
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

Figure 3: Higher-order cluster in the *C. elegans* neuronal network (28). A: The 4-node “bi-fan” motif, which is over-expressed in the neuronal networks (1). Intuitively, this motif describes a cooperative propagation of information from the nodes on the left to the nodes on the right.

B: The best higher-order cluster in the *C. elegans* frontal neuronal network based on the motif in (A). The cluster contains three ring motor neurons (RMEL/V/R; cyan) with many outgoing connections, serving as the source of information; six inner labial sensory neurons (IL2DL/VR/R/DR/VL; orange) with many incoming connections, serving as the destination of information; and four URA neurons (purple) acting as intermediaries. These RME neurons have been proposed as pioneers for the nerve ring (20), while the IL2 neurons are known regulators of nictation (21), and the higher-order cluster exposes their organization. The cluster also reveals that RIH serves as a critical intermediary of information processing. This neuron has incoming links from all three RME neurons, outgoing connections to five of the six IL2 neurons, and the largest total number of connections of any neuron in the cluster.

C: Illustration of the higher-order cluster in the context of the entire network. Node locations are the true two-dimensional spatial embedding of the neurons. Most information flows from left to right, and we see that RME/V/R/L and RIH serve as sources of information to the neurons on the right.

Representation Learning on Networks, snap.stanford.edu/proj/embeddings-www, WWW 2018
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!

Raw Data ➔ Structured Data ➔ Learning Algorithm ➔ Model

- Feature Engineering
  - Automatically learn the features

- Downstream prediction task
Feature Learning in Graphs

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

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

Feature representation, embedding
Example

Zachary’s Karate Club Network:

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.
This talk

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

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

3) Applications