Social Network Analysis

Lecture 1: Introduction to Graph & Network Analysis

Dr. Hung-Nghiep Tran
nghiepth@uit.edu.vn

University of Information Technology, VNU-HCM, Vietnam
2025

Mục tiêu buổi học

- Giới thiệu sơ lược Graphs & Networks
- Giới thiệu sơ lược Analysis & Applications

- Làm quen với phương pháp học trong môn này
- Thiết lập thói quen đọc trước & thảo luận

Đọc và diễn giải

Tài liệu đọc

- Reading:
  - Introduction to Graphs and Networks
  - Starter Topic: Structure of Graphs
  - Choice of Network Representations

Câu hỏi kiểm tra

Quiz

1. A graph in our course is best described as:
A. A set of tables and SQL joins
B. A set of entities (nodes) and relationships (edges)
C. A time series of stock prices
D. A probability distribution

2. Which is a good example of a bipartite network?
A. Friendships among students in a class
B. Webpages and hyperlinks among them
C. Users and movies with rating edges
D. Airports connected by flight routes

Quiz

3. Directed graphs are needed when:
A. Relationships have direction (e.g., “follows”, “links to”)
B. Only weighted edges exist
C. We never have self-loops
D. We must visualize with Gephi

4. A weighted edge is useful when:
A. Edges are always binary
B. We encode strength/frequency of interactions
C. We need to remove parallel edges
D. The graph is bipartite

5. Which task is PRIMARILY about finding influential nodes?
A. Community detection
B. Link prediction
C. Centrality analysis
D. Diffusion modeling

Quiz

1. A graph in our course is best described as:
A. A set of tables and SQL joins
B. A set of entities (nodes) and relationships (edges)
C. A time series of stock prices
D. A probability distribution
Answer: B

2. Which is a good example of a bipartite network?
A. Friendships among students in a class
B. Webpages and hyperlinks among them
C. Users and movies with rating edges
D. Airports connected by flight routes
Answer: C

Quiz

3. Directed graphs are needed when:
A. Relationships have direction (e.g., “follows”, “links to”)
B. Only weighted edges exist
C. We never have self-loops
D. We must visualize with Gephi
Answer: A

4. A weighted edge is useful when:
A. Edges are always binary
B. We encode strength/frequency of interactions
C. We need to remove parallel edges
D. The graph is bipartite
Answer: B

5. Which task is PRIMARILY about finding influential nodes?
A. Community detection
B. Link prediction
C. Centrality analysis
D. Diffusion modeling
Answer: C

Hands-on lab

Setup

- Install python
  - with Anaconda
- Install IDE:
  - (Microsoft) vscode
  - Jupyter Notebook
- Install networkx
  - with pip

Mini hands-on

```
import networkx as nx
G = nx.read_edgelist('toy_edges.txt')  # create a small toy file with 8–12 edges
print(G.number_of_nodes(), G.number_of_edges())
print(nx.average_clustering(G))
print(sorted(nx.degree(G), key=lambda x: x[1], reverse=True)[:5])
```

Data in *toy_edges.txt*:

```
Alice Bob
Bob Charlie
Alice Dan
```

Chuẩn bị cho tuần tới

Chuẩn bị trước tuần sau

- Reading:
  - (Re-read) Choice of Network Representations
  - SNAP library introduction
  - Maths review

- Câu hỏi định hướng (trả lời ngắn):
  - Compare edge list, adjacency list, adjacency matrix: memory & time trade-offs for common tasks
  - Compare NetworkX, SNAP.py, and Gephi. What's it great at? What's inconvenient?

- Muddiest point: "Điểm nào mù mờ nhất sau khi đọc? (< 50 từ)"

Thank you for listening

Q & A