Social Network Analysis

Lecture 3: Measuring Networks & Random Graph Models

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

University of Information Technology, VNU-HCM, Vietnam
2025

Mục tiêu buổi học

- Define, compute, and interpret key structural properties of networks: *degree distribution, path length, clustering, connected components*
- Explain the Erdős-Rényi model and how it serves as a null model
- Contrast properties of real-world networks vs. random graphs
- Understand the small-world property and its mechanisms

Câu hỏi kiểm tra trước

Quiz

1. Which quantity describes the probability that a randomly chosen node has degree k?
A. Path length B. Degree distribution C. Clustering coefficient D. Diameter

2. What does a clustering coefficient measure?
A. Average degree of the graph B. Length of the shortest path
C. Proportion of neighbor pairs that are connected D. Probability of edge existence

3. What is the diameter of a network?
A. Maximum shortest path distance between any two nodes B. Largest connected component
C. Average degree D. Number of triangles in the graph

4. In an Erdős-Rényi random graph edges:
A. Are chosen deterministically B. Appear independently with probability p
C. Are weighted D. Follow power-law distribution

5. Which of these is *not* typically true for real social networks?
A. High clustering B. Heavy-tailed degree distribution C. Short average path length D. Degree distribution is binomial

Quiz

1. Which quantity describes the probability that a randomly chosen node has degree k?
A. Path length **B. Degree distribution** C. Clustering coefficient D. Diameter

2. What does a clustering coefficient measure?
A. Average degree of the graph B. Length of the shortest path
C. Proportion of neighbor pairs that are connected D. Probability of edge existence

3. What is the diameter of a network?
A. Maximum shortest path distance between any two nodes B. Largest connected component
C. Average degree D. Number of triangles in the graph

4. In an Erdős-Rényi random graph edges:
A. Are chosen deterministically **B. Appear independently with probability p**
C. Are weighted D. Follow power-law distribution

5. Which of these is *not* typically true for real social networks?
A. High clustering B. Heavy-tailed degree distribution C. Short average path length **D. Degree distribution is binomial**

Đọc và diễn giải

Tài liệu đọc

- Reading:
  - Measuring Networks & Random Graph Models

Câu hỏi kiểm tra sau

Quiz

1. The degree distribution of $G_{n,p}$ follows:
A. Power law B. Normal distribution C. Uniform distribution D. Binomial (→ Poisson in large nnn)

2. What happens in $G_{n,p}$ when average degree k<1?
A. Giant component emerges B. Graph is fully connected
C. Components remain small (~log n) D. Graph becomes bipartite

3. Which property of real networks is not captured by Erdős-Rényi graphs?
A. Short paths B. High clustering C. Existence of giant component D. Expansion properties

4. Watts-Strogatz small-world model achieves:
A. High clustering + long path length B. Low clustering + short path length
C. High clustering + short path length D. Uniform degree distribution

5. In sampling degree distributions, a key pitfall is:
A. Measuring clustering instead of degree B. Sampling biases distort heavy-tail detection
C. Confusing directed with undirected graphs D. Overcounting triangles

Quiz

1. The degree distribution of $G_{n,p}$ follows:
A. Power law B. Normal distribution C. Uniform distribution **D. Binomial (→ Poisson in large nnn)**

2. What happens in $G_{n,p}$ when average degree k<1?
A. Giant component emerges B. Graph is fully connected
C. Components remain small (~log n) D. Graph becomes bipartite

3. Which property of real networks is not captured by Erdős–Rényi graphs?
A. Short paths **B. High clustering** C. Existence of giant component D. Expansion properties

4. Watts-Strogatz small-world model achieves:
A. High clustering + long path length B. Low clustering + short path length
C. High clustering + short path length D. Uniform degree distribution

5. In sampling degree distributions, a key pitfall is:
A. Measuring clustering instead of degree **B. Sampling biases distort heavy-tail detection**
C. Confusing directed with undirected graphs D. Overcounting triangles

Thảo luận

Chuẩn bị tuần trước

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

Chuẩn bị tuần trước

- Trả lời ngắn câu hỏi định hướng:
  - Small-world property: Graphs with short average path length and high clustering.
    - Indicators: h (average path length), C (clustering coefficient).
  - Degree distribution of social networks: Heavy-tailed (power-law like).
    - Many nodes have small degree, few hubs have very high degree.
  - Why heavy-tailed? Preferential attachment and social mechanisms → hubs form naturally.
  - Log-log plot of degree frequency: Plots logP(k) vs. log k. Reveals straight-line trend for power-law.
  - Linear vs log-log plot: Linear hides tail, log-log reveals heavy-tail clearly.

Chuẩn bị tuần trước

- Trả lời ngắn câu hỏi định hướng:
  - Local vs. global clustering coefficient:
    - Local: proportion of neighbors connected for a node.
    - Global: average over all nodes.
    - High values = strong triadic closure.
  - Short path length + high clustering: Explained by small-world models (Watts-Strogatz) with local connections + random shortcuts.
  - Pitfall in estimating degree distributions: Sampling bias (e.g., BFS sampling oversamples high-degree nodes).

Hands-on lab

Setup

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

Chuẩn bị cho tuần tới

Chuẩn bị trước tuần sau

- Reading:
  - (Stanford CS224W L15) Network Centrality
  - (Stanford CS224W L3) PageRank

- Câu hỏi định hướng (trả lời ngắn):
  - Define degree centrality, betweenness centrality, and eigenvector centrality. How do they differ conceptually?
  - Why might PageRank favor nodes with many incoming links from important neighbors?
  - In what kind of applications is betweenness centrality especially insightful?
  - What's a limitation of degree centrality in capturing influence?
  - Sketch how PageRank can be interpreted as a random walk with teleportation.

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

Thank you for listening

Q & A