

Lab 05 - Exercise 01

Contents

1.	Các độ đo cơ bản của mạng (Basic Network Metric)	2
1.1.	Average Degree	2
1.2.	Network Diameter	2
1.3.	Graph Density	3
1.4.	Connected Components	3
1.5.	Average Path Length	4
1.6.	Average Clustering Coefficient	4
2.	Độ đo trung tâm (Centrality Metrics)	5
2.1.	Degree Centrality	5
2.2.	Betweenness Centrality	5
2.3.	Closeness Centrality	5
2.4.	Eigenvector Centrality	6
2.5.	PageRank	6
2.6.	HITS (Hub and Authority)	7
2.7.	Eccentricity	7

1. Các độ đo cơ bản của mạng (Basic Network Metric)

1.1. Average Degree

- **Định nghĩa:** Average Degree (Độ trung bình) của một đồ thị là trung bình số lượng cạnh kết nối với mỗi đỉnh trong đồ thị. Đây là một chỉ số cơ bản để mô tả độ kết nối tổng quan của một mạng hoặc đồ thị.

- **Công thức:**

- o Vô hướng: $G = (V, E)$

$$A = 2 \frac{|E|}{|V|}$$

Với

- $|V|$: Số lượng đỉnh trong đồ thị.
- $|E|$: Số lượng cạnh trong đồ thị.
- Hệ số 2 trong công thức xuất hiện vì mỗi cạnh được tính cho cả hai đỉnh mà nó kết nối.
 - o Có hướng

$$A = \frac{|E|}{|V|}$$

- **Phạm vi:**

- o Đồ thị không hướng: $0 \leq \text{Average Degree} \leq |V| - 1$.
 - o Đồ thị có hướng: Phạm vi tương tự nhưng không nhân đôi số cạnh vì cạnh có hướng chỉ tính một lần.

- **Đánh giá:** Không có một giá trị cố định nào là "tốt" cho Average Degree, vì điều này phụ thuộc vào ngữ cảnh của bài toán và loại mạng. Tuy nhiên:

- o **Chỉ số cao:** chỉ ra mạng có tính kết nối cao. Các đỉnh có xu hướng có nhiều kết nối, thường gặp trong các mạng xã hội lớn hoặc đồ thị hoàn chỉnh. Tuy nhiên, nếu quá cao, có thể báo hiệu rằng mạng đang quá phức tạp và khó phân tích.
 - o **Chỉ số thấp:** chỉ ra mạng thưa thớt. Điều này phổ biến trong các mạng chuyên biệt hoặc mạng có cấu trúc rõ ràng như cây hoặc mạng không đồng nhất. Mạng thưa thớt dễ phân tích hơn nhưng có thể kém bền vững nếu xảy ra lỗi (ví dụ: mất đỉnh hoặc cạnh).

1.2. Network Diameter

- **Định nghĩa:** Network Diameter (Đường kính mạng) là độ dài lớn nhất của tất cả các đường ngắn nhất (shortest paths) giữa bất kỳ hai đỉnh nào trong một đồ thị. Nói cách khác, nó là khoảng cách xa nhất mà một thông điệp hoặc tài nguyên cần phải đi qua trong mạng.
- **Công thức:**

$$D = \max_{u,v \in V} \{d(u,v)\}$$

- Phạm vi: $1 \leq D \leq n - 1$
- Đánh giá:
 - **Đường kính nhỏ:** thể hiện rằng mạng có tính kết nối cao, dẫn đến truyền tải thông tin nhanh chóng. Ví dụ: Mạng xã hội hoặc mạng internet có đường kính nhỏ để cải thiện hiệu quả giao tiếp.
 - **Đường kính lớn:** Có thể chỉ ra rằng mạng bị phân mảnh hoặc có cấu trúc chuỗi dài (linear structure). Điều này thường không mong muốn trong các mạng yêu cầu tính hiệu quả hoặc độ bền cao.

1.3. Grap Desity

- **Định nghĩa:** Graph Density (Mật độ đồ thị) là một chỉ số đo lường mức độ kết nối giữa các đỉnh trong đồ thị. Nó được tính bằng tỉ lệ giữa số cạnh thực tế so với số cạnh tối đa có thể có trong đồ thị.
- **Công thức:**
 - Với đồ thị vô hướng

$$A = \frac{2E}{N(N-1)}$$

Trong đó:

E: Số cạnh trong đồ thị.

N: Số đỉnh trong đồ thị.

Với đồ thị có hướng thì lược bỏ việc *2

- **Phạm vi:** $0 \leq D \leq 1$
- **Đánh giá:** chỉ số tốt hay xấu phụ thuộc vào ngữ cảnh.
 - Với D cao: Thể hiện đồ thị có tính kết nối chặt chẽ, phù hợp với mạng lưới xã hội hoặc các hệ thống mạnh mẽ.
 - Còn D thấp: Biểu thị mạng thưa, thường gặp trong mạng quy mô lớn hoặc các hệ thống phân tán.

1.4. Connected Components

- **Định nghĩa:** Connected components là các tập hợp con của đồ thị, trong đó các đỉnh được kết nối trực tiếp hoặc gián tiếp với nhau và không kết nối với các đỉnh khác ngoài tập hợp đó.
- **Công thức:** Không có công thức cụ thể, nhưng các thành phần liên thông có thể được xác định bằng các thuật toán như DFS hoặc BFS.
- **Phạm vi:** Nếu là 1 được xem là đồ thị liên thông, còn lớn hơn 1 là đồ thị nhiều thành phần liên thông
- **Đánh giá:**
 - **Số thành phần nhỏ (1 hoặc vài cụm):** Thể hiện mạng liên thông tốt, phù hợp với các hệ thống yêu cầu giao tiếp giữa các thành phần.
 - **Số thành phần lớn:** Biểu thị mạng bị phân mảnh, không lý tưởng cho các hệ thống liên kết mạnh (như mạng xã hội)

1.5. Average Path Length

- **Định nghĩa:** Average path length là độ dài trung bình của tất cả các đường đi ngắn nhất giữa mọi cặp đỉnh trong đồ thị.
- **Công thức:**

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} d(v_i, v_j)$$

- **Phạm vi:** $L \geq 1$

Trong đó:

- $L=1$: Đồ thị đầy đủ, mọi đỉnh kết nối trực tiếp.
 - L càng lớn: Đồ thị càng ít liên kết.
- **Đánh giá:**
 - Chỉ số nhỏ: Mạng lưới hiệu quả, các đỉnh dễ tiếp cận nhau, tốt trong mạng xã hội hoặc hệ thống giao thông.
 - Chỉ số lớn: Mạng kém kết nối, không phù hợp cho các ứng dụng yêu cầu truyền thông nhanh.

1.6. Average Clustering Coefficient

- **Định nghĩa:** Average clustering coefficient đo mức độ mà các đỉnh trong đồ thị có xu hướng hình thành tam giác (3 đỉnh kết nối với nhau).
- **Công thức:**

$$C = \frac{1}{N} \sum_{i=1}^N c_i$$

- **Phạm vi:** $0 \leq C \leq 1$

Trong đó:

- Với $C = 0$ thì không có tam giác
 - Với $C = 1$ thì tất cả đỉnh đều kết nối với nhau
- **Đánh giá:**
 - **Chỉ số cao:** Thể hiện mạng có tính cục bộ cao, các đỉnh có xu hướng tạo cụm, phù hợp trong mạng xã hội hoặc sinh học.
 - **Chỉ số thấp:** Mạng có cấu trúc phân tán, thường gặp trong mạng ngẫu nhiên hoặc các hệ thống ít tương tác.

2. Độ đo trung tâm (Centrality Metrics)

2.1. Degree Centrality

- **Định nghĩa:** Độ đo này xác định tầm quan trọng của một đỉnh dựa trên số lượng cạnh kết nối trực tiếp với nó. Với đồ thị có hướng, nó được chia thành:
 - In-degree centrality: Số cạnh đi vào đỉnh.
 - Out-degree centrality: Số cạnh đi ra từ đỉnh.
- **Công thức:**

- **In-degree Centrality:**

$$C_{in}(v) = \frac{\text{Số cạnh vào đỉnh } v}{N - 1}$$

- **Out-degree Centrality:**

$$C_{out}(v) = \frac{\text{Số cạnh ra từ đỉnh } v}{N - 1}$$

- **Phạm vi:** $0 \leq C \leq 1$
- **Đánh giá:**
 - Giá trị cao: Đỉnh có nhiều kết nối, thường quan trọng trong mạng lưới.
 - Giá trị thấp: Đỉnh ít được kết nối, có vai trò ít quan trọng hơn.

2.2. Betweenness Centrality

- **Định nghĩa:** Đo tầm quan trọng của một đỉnh dựa trên số lượng đường đi ngắn nhất giữa các cặp đỉnh khác đi qua nó.
- **Công thức:**

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma(s, v, t)}{\sigma(s, t)}$$

- **Phạm vi:** $0 \leq C \leq 1$
- **Đánh giá:**
 - Giá trị cao: Đỉnh là trung gian quan trọng trong mạng lưới.
 - Giá trị thấp: Đỉnh ít đóng vai trò trung gian.

2.3. Closeness Centrality

- **Định nghĩa:** Đo khả năng tiếp cận của một đỉnh tới tất cả các đỉnh khác trong đồ thị thông qua các đường đi ngắn nhất.

- Công thức

$$C_C(v) = \frac{N - 1}{\sum_{t \neq v} d(v, t)}$$

- Phạm vi : $0 \leq C \leq 1$
- Đánh giá:
 - o Giá trị cao: Đỉnh có khả năng tiếp cận nhanh với các đỉnh khác.
 - o Giá trị thấp: Đỉnh bị cô lập hoặc khó tiếp cận.

2.4. Eigenvector Centrality

- Định nghĩa: Đo tầm quan trọng của một đỉnh dựa trên mức độ kết nối với các đỉnh quan trọng khác.

- Công thức

$$C_E(v) = \frac{1}{\lambda} \sum_{u \in \text{Neighbours}(v)} A_{vu} C_E(u)$$

- Phạm vi: không giới hạn, phụ thuộc vào cấu trúc của đồ thị
- Đánh giá:
 - o Giá trị cao: Đỉnh có kết nối với nhiều đỉnh quan trọng.
 - o Giá trị thấp: Đỉnh ít quan trọng trong mạng.

2.5. PageRank

- **Định nghĩa:** Một thuật toán xác định tầm quan trọng của các đỉnh dựa trên số lượng và chất lượng các kết nối đến đỉnh đó.
- Công thức:

$$PR(v) = (1 - d) + d \sum_{u \in \text{InNeighbours}(v)} \frac{PR(u)}{\text{OutDegree}(u)}$$

- Phạm vi: $0 \leq C \leq 1$
- Đánh giá:
 - o Giá trị cao: Đỉnh quan trọng, nhận được nhiều kết nối từ các đỉnh khác.

- Giá trị thấp: Đỉnh kém quan trọng.

2.6. HITS (Hub and Authority)

- **Định nghĩa:** HITS đo lường tầm quan trọng của đỉnh qua hai khái niệm:
 - Hub: Đỉnh trở tới nhiều Authority.
 - Authority: Đỉnh được trở tới bởi nhiều Hub.

- Công thức

- **Authority:**

$$A(v) = \sum_{u \in \text{InNeighbours}(v)} H(u)$$

- **Hub:**

$$H(v) = \sum_{u \in \text{OutNeighbours}(v)} A(u)$$

- **Phạm vi:** không có giới hạn cụ thể
- **Đánh giá:**
 - Giá trị cao (Hub): Đỉnh trở tới nhiều đỉnh quan trọng.
 - Giá trị cao (Authority): Đỉnh nhận được nhiều liên kết từ các Hub.

2.7. Eccentricity

- **Định nghĩa:** Eccentricity đo khoảng cách lớn nhất từ một đỉnh tới bất kỳ đỉnh nào khác trong đồ thị.
- **Công thức**

$$E(v) = \max_{u \in V} d(v, u)$$

- **Phạm vi:** $1 \leq E \leq \infty$
- **Đánh giá:**
 - Giá trị thấp: Đỉnh có vị trí trung tâm hơn trong đồ thị.
 - Giá trị cao: Đỉnh ở vị trí xa trong đồ thị.