

Bài lab 05 - 01

2186400330 – Lê Nguyễn Anh Nhật

Bài thực hành số 1: Tìm hiểu các chỉ số thông kê cơ bản sử dụng Gephi

Hãy cho biết ý nghĩa của các chỉ số bên dưới, công thức, phạm vi, như thế nào là chỉ số tốt?

1. Độ đo cơ bản của mạng (Basic Network Metrics):

- Average Degree
- Network Diameter
- Graph Density
- Connected Components
- Average Path Length
- Average Clustering Coefficient

2. Độ đo tính trung tâm (Centrality Metrics):

- Degree Centrality (In-degree và Out-degree với đồ thị có hướng)
- Betweenness Centrality
- Closeness Centrality
- Eigenvector Centrality
- PageRank
- HITS (Hub and Authority)
- Eccentricity

Hình 1 – Đề bài tập lab05.01

Câu 1:

1.1. Average Degree

Average Degree (Độ trung bình):

- **Ý nghĩa:** Trung bình số lượng kết nối (liên kết) của mỗi nút trong mạng.
- **Công thức:** $\text{Average Degree} = \frac{2E}{N}$
- Trong đó E là số lượng cạnh và N là số lượng nút.
- **Phạm vi:** [0, 1] (đối với đồ thị không có hướng) hoặc [0, 1] (đối với đồ thị có hướng).
- **Chỉ số tốt:** Tùy thuộc vào mục đích phân tích. Giá trị cao cho thấy mạng có nhiều liên kết, giá trị thấp thể hiện mạng rời rạc. Giá trị càng cao càng tốt và tốt nhất là 1.

1.2. Network Diameter

- **Ý nghĩa:** Khoảng cách dài nhất giữa hai nút trong đồ thị (trong đường đi ngắn nhất).
 - **Công thức:** $\text{Diameter} = \max(d(i, j))$, với $d(i, j)$ là khoảng cách ngắn nhất giữa hai nút i, j .
- **Phạm vi:** Từ 1 đến vô cùng (nếu mạng không kết nối).
- **Chỉ số tốt:** Diameter nhỏ cho thấy đồ thị kết nối tốt.

1.3. Graph Density

- **Ý nghĩa:** Tỷ lệ giữa số cạnh hiện tại và số cạnh tối đa mà mạng có thể có.
- **Công thức:**

$$\text{Density} = \frac{2E}{N(N-1)} \quad (\text{Đồ thị vô hướng})$$

Hoặc:

$$\text{Density} = \frac{E}{N(N-1)} \quad (\text{Đồ thị có hướng})$$

□ **Phạm vi:** $[0, 1]$.

□ **Chỉ số tốt:** Mật độ cao (gần 1) cho thấy mạng dày đặc, mật độ thấp (gần 0) cho thấy mạng thưa.

1.4. Connected Components

- **Ý nghĩa:** Số lượng cụm (subgraph) trong mạng mà mọi nút đều kết nối trực tiếp hoặc gián tiếp.
- **Công thức:** Tính bằng DFS hoặc BFS.
- **Phạm vi:** $[1, N]$.
- **Chỉ số tốt:** Số lượng nhỏ cho thấy mạng có tính liên kết tốt.

1.5. Average Path Length

- **Ý nghĩa:** Trung bình khoảng cách ngắn nhất giữa tất cả các cặp nút.
- **Công thức:**

$$\text{Average Path Length} = \frac{\sum_{i \neq j} d(i, j)}{N(N - 1)}$$

- **Phạm vi:** $[1, \infty)$.
- **Chỉ số tốt:** Giá trị nhỏ cho thấy thông tin trong mạng lan truyền nhanh chóng.

1.6. Average Clustering Coefficient

- **Ý nghĩa:** Đo lường mức độ mà các nút có xu hướng tạo thành các cụm.
- **Công thức:**

$$C = \frac{1}{N} \sum_{i=1}^N \frac{2e_i}{k_i(k_i - 1)}$$

- **Phạm vi:** $[0,1]$.
- **Chỉ số tốt:** Giá trị cao (gần 1) cho thấy mạng có cấu trúc phân cụm rõ ràng.

2. Độ đo trung tâm (Centrality Metrics)

1. Degree Centrality (Trung tâm bậc):

- **Ý nghĩa:** Số kết nối trực tiếp của một nút (In-degree và Out-degree đối với đồ thị có hướng).

$$C_D(v) = \frac{\deg(v)}{N - 1}$$

- **Công thức:**
- **Phạm vi:** $[0,1]$.
- **Chỉ số tốt:** Giá trị cao cho thấy nút có tầm quan trọng cao trong mạng.

2. Betweenness Centrality (Trung tâm trung gian):

- **Ý nghĩa:** Đo lường mức độ mà một nút nằm trên đường đi ngắn nhất giữa các cặp nút khác.

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

- **Công thức:**

- **Phạm vi:** $[0,1]$.
- **Chỉ số tốt:** Giá trị cao cho thấy nút là cầu nối quan trọng.

3. Closeness Centrality (Trung tâm gần):

- **Ý nghĩa:** Đo lường mức độ gần của một nút với tất cả các nút khác.
- **Công thức:**

$$C_C(v) = \frac{1}{\sum_u d(v, u)}$$

-
- **Phạm vi:** $[0,1]$.
- **Chỉ số tốt:** Giá trị cao cho thấy nút có khả năng tiếp cận các nút khác nhanh chóng.

4. Eigenvector Centrality (Trung tâm vector riêng):

- **Ý nghĩa:** Đo tầm quan trọng của một nút dựa trên tầm quan trọng của các nút kết nối với nó.
- **Công thức:**

$$x_i = \frac{1}{\lambda} \sum_j A_{ij} x_j$$

- Với A_{ij} là ma trận kề, λ là giá trị riêng lớn nhất.
- **Phạm vi:** $[0,1]$.
- **Chỉ số tốt:** Giá trị cao cho thấy nút liên kết với các nút quan trọng.

5. PageRank:

- **Ý nghĩa:** Đo mức độ quan trọng của một nút dựa trên số lượng và chất lượng của các liên kết đến nút đó.

- **Công thức:**

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

- Với d là hệ số giảm (thường là 0.85).
- **Phạm vi:** $[0,1]$.
- **Chỉ số tốt:** Giá trị cao cho thấy nút được tham chiếu nhiều.

6. HITS (Hub and Authority):

- **Ý nghĩa:** Đánh giá mức độ của các nút như Hub (trở đến nhiều Authority) và Authority (được trở đến bởi nhiều Hub).
- **Công thức:**
 - Hub score: $h(v) = \sum_{u \in \text{out}(v)} a(u)$
 - Authority score: $a(v) = \sum_{u \in \text{in}(v)} h(u)$
- **Phạm vi:** $[0,1]$.
- **Chỉ số tốt:** Giá trị cao cho thấy nút có vai trò quan trọng.

7. Eccentricity (Độ lệch tâm):

- **Ý nghĩa:** Khoảng cách lớn nhất từ một nút đến bất kỳ nút nào khác trong mạng.
- **Công thức:**

$$E(v) = \max_u d(v, u)$$

- **Phạm vi:** $[1, \infty)$.
- **Chỉ số tốt:** Giá trị nhỏ cho thấy nút gần gũi với các nút khác.

