

# Lab05.01

December 28, 2024

## 1 Độ đo cơ bản của mạng

### 1.1 Average Degree

Average Degree, hay Độ trung bình bậc của mạng, đo lường trung bình số lượng cạnh kết nối đến mỗi nút trong mạng. Chỉ số này cung cấp cái nhìn tổng quát về mức độ kết nối của các nút trong mạng, từ đó đánh giá sự gắn kết và mức độ liên thông của mạng.

#### Ý nghĩa

- **Mạng xã hội:** Average Degree cho thấy mức độ trung bình mà một cá nhân (nút) kết nối với các cá nhân khác. Giá trị cao có thể biểu thị một cộng đồng tương tác mạnh mẽ.
- **Mạng thông tin:** Trong các mạng phân phối thông tin, Average Degree cao thường cho thấy khả năng lan truyền thông tin hiệu quả hơn.
- **Mạng giao thông:** Đánh giá mức độ kết nối trung bình của các nút giao thông, từ đó xác định hiệu suất của hệ thống giao thông.

#### Công thức

$$AverageDegree = \frac{2E}{N} \quad (1)$$

Trong đó  $E$  là tổng số cạnh (edge) trong mạng.

#### Phạm vi giá trị

**Giá trị Average Degree nằm trong khoảng từ 0 đến  $N-1$**

- 0: Khi mạng không có cạnh nào (mạng rỗng), tất cả các nút đều có bậc bằng 0.

### 1.2 Network Diameter

**Định nghĩa:** Đường kính của mạng (graph diameter) là khoảng cách (shortest path) lớn nhất giữa hai nút bất kỳ trong mạng

**Công thức**

$$Diameter = \max d(u, v) \quad (2)$$

Trong đó  $d(u, v)$  là độ dài đường đi ngắn nhất giữa 2 nút  $u$  và  $v$

**Phạm vi:**

- Nhỏ nhất là 1 (mạng đầy đủ - complete graph, mọi nút đều kết nối trực tiếp).
- Lớn nhất là  $N - 1$  (một mạng chuỗi đường thẳng hoặc mạng cây gần nhánh dài, v.v.).

**Chỉ số tốt:**

- Đường kính nhỏ gợi ý mạng có tính kết nối cao, dễ lan truyền thông tin, ít trung gian.
- Đường kính lớn cho thấy một số cặp nút nằm cách xa nhau, có thể khiến chi phí truyền thông tin lớn.

**1.3 Graph Density**

**Định nghĩa:** Mật độ của đồ thị mô tả tỉ lệ số cạnh hiện có trên số cạnh tối đa (nếu mạng là đồ thị đầy đủ).

**Công thức:**

- Đồ thị vô hướng  

$$Density = \frac{2E}{N(N-1)}$$
- Đồ thị có hướng (không tính cạnh bắc ngược lại chính nó):  

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

**Phạm vi:** Từ 0 (không có cạnh nào) đến 1 (mạng đầy đủ).

**Chỉ số tốt:**

- Density cao cho thấy mạng dày đặc.
- Density thấp cho thấy mạng thưa, có thể dẫn tới phân mảnh hoặc khó lan truyền thông tin, nhưng cũng có thể tiết kiệm chi phí liên kết.

**1.4 Connected Components**

**Định nghĩa:** Một thành phần liên thông (connected component) là một tập các nút trong đó mọi nút đều có thể đi tới (và nhận được) từ các nút khác. Với đồ thị vô hướng, nếu mạng phân tách thành nhiều cụm, mỗi cụm là một connected component.

**Phạm vi:**

- Tối thiểu 1 (tức là đồ thị liên thông hoàn toàn).
- Tối đa  $N$  (mỗi nút tách biệt, không có cạnh nào).

**Chỉ số tốt:**

- Với mục tiêu kết nối, số lượng thành phần liên thông càng ít càng tốt (lý tưởng là 1).
- Nếu mục tiêu phân mảnh (ví dụ phân vùng để bảo mật), thì số lượng thành phần nhiều có thể phù hợp.

## 1.5 Average Path Length

**Định nghĩa:** Là độ dài đường đi ngắn nhất trung bình (trung bình khoảng cách) giữa tất cả các cặp nút trong mạng. Thường dùng để đánh giá “độ nhỏ của thế giới” (small-world effect).

**Công thức:**

$$AveragePathLength = \frac{1}{N^2} \sum_{u \neq v} d(u, v) \quad (3)$$

Trong đó  $d(u, v)$  là khoảng cách ngắn nhất giữa hai nút  $u, v$

**Phạm vi:**

- Tối thiểu 1 (mạng đầy đủ).
- Tối đa có thể gần  $N - 1$  (mạng dạng đường dài).

**Chỉ số tốt:**

- Giá trị nhỏ ám chỉ mạng có tính “nhỏ” (small-world), dễ kết nối giữa các nút.
- Giá trị lớn cho thấy chi phí đi lại/lan truyền cao.

## 1.6 Average Clustering Coefficient

**Định nghĩa:** Độ đo trung tâm dựa trên bậc (số liên kết trực tiếp). Nút có bậc cao được coi là “trung tâm” vì có nhiều kết nối tới các nút khác.

**Công thức:**

$$C_{degree}(v) = \frac{deg(v)}{N - 1} \quad (4)$$

**Phạm vi:** Từ 0 đến 1

**Chỉ số tốt:**

- Cao cho thấy nút có nhiều “tiếp xúc” trực tiếp, là trung tâm về mặt kết nối.
- Thấp có nghĩa nút ít được kết nối trực tiếp.

## 2 Độ đo tính trung tâm (Centrality Metrics)

### 2.1 Betweenness Centrality

**Định nghĩa:** Đo lường mức độ một nút nằm trên nhiều đường đi ngắn nhất giữa các cặp nút khác. Nút có Betweenness cao là “cầu nối” quan trọng.

**Công thức:**

$$C_{betweenness}(v) = \sum_{s \neq v \neq t} \frac{\sigma_s(v)}{\sigma_s} \quad (5)$$

**Phạm vi:** Giá trị tuyệt đối phụ thuộc số lượng đường đi; tuy nhiên có thể chuẩn hoá về  $[0,1]$ .

**Chỉ số tốt:**

- Cao: nút đó có vai trò môi giới (broker). Mất nút này có thể làm gián đoạn mạnh mẽ mạng.
- Thấp: ít ảnh hưởng lên các đường đi ngắn nhất.

### 2.2 Closeness Centrality

**Định nghĩa:** Đo lường mức độ một nút gần với tất cả các nút khác thông qua khoảng cách đường đi ngắn nhất. Nút có Closeness cao là nút có khoảng cách trung bình đến các nút khác nhỏ.

**Công thức:**

$$C_{closeness}(v) = \frac{N - 1}{\sum_{u \neq v} d(v, u)} \quad (6)$$

$d(v, u)$  là khoảng cách ngắn nhất từ  $v$  đến  $u$

**Phạm vi:** Từ 0 đến 1 (sau khi chuẩn hoá).

**Chỉ số tốt:**

- Cao: nút đó “gần” đa số nút khác (lan truyền thông tin nhanh).
- Thấp: nút “xa”, khó tiếp cận các nút khác.

## 2.3 Eigenvector Centrality

**Định nghĩa:** Đo lường tầm quan trọng của một nút dựa trên tầm quan trọng của các nút mà nó liên kết. Một nút có liên kết đến những nút “quan trọng” thì càng được đánh giá cao.

**Công thức:**

$$C_{\text{eigenvector}}(v_i) = \frac{1}{\lambda} \sum_j A_{ij} C_{\text{eigenvector}}(v_j) \quad (7)$$

**Phạm vi:** Sau khi chuẩn hoá, thường trong khoảng  $[0,1]$ .

**Chỉ số tốt:**

- Cao: nút được kết nối (trực tiếp hoặc gián tiếp) với nhiều nút quan trọng khác.
- Thấp: ít hoặc không liên kết tới những nút có tầm quan trọng cao.

## 2.4 PageRank

**Định nghĩa:** Đây là biến thể của Eigenvector Centrality, ban đầu được Google đề xuất để xếp hạng trang web. PageRank xem xét cả số liên kết đến (inbound links) và mức quan trọng của nguồn liên kết (tương tự ý tưởng “bỏ phiếu”).

**Công thức**

$$PR(v) = \frac{1 - \alpha}{N} + \alpha \sum_{u \rightarrow v} \frac{PR(u)}{\text{OutDegree}(u)} \quad (8)$$

**Phạm vi:** Sau khi tính, PageRank được chuẩn hoá để tổng các giá trị PageRank bằng 1, hoặc chuẩn hoá thành  $[0,1]$  tùy mục đích.

**Chỉ số tốt:**

- Cao: nút có nhiều liên kết từ các nút “có chất lượng” (PageRank cao).
- Thấp: ít được liên kết hoặc chỉ được liên kết từ những nút có giá trị thấp.

## 2.5 HITS (Hyperlink-Induced Topic Search)

**Định nghĩa:** HITS phân tách khái niệm “trung tâm” thành hai loại: Hub và Authority.

- **Authority:** Trang (nút) có nội dung uy tín về chủ đề (nhiều hub tốt trở đến).
- **Hub:** Trang (nút) trở đến nhiều authority tốt.

**Công thức**

$$A(v) = \sum_{u \rightarrow v} H(u) \quad (9)$$

$$H(v) = \sum_{v \rightarrow w} A(w) \quad (10)$$

**Phạm vi:** Thường được chuẩn hoá về  $[0,1]$ .

**Chỉ số tốt:**

- Hub cao: nút trở đến nhiều authority chất lượng.
- Authority cao: nút được nhiều hub chất lượng trở đến.

## 2.6 Eccentricity

**Định nghĩa:** Với mỗi nút, eccentricity là khoảng cách lớn nhất từ nút đó đến bất kỳ nút nào khác trong mạng (tức là “bán kính cục bộ” của nút đó).

**Công thức:**

$$e(v) = \max_d(v, u) \quad (11)$$

**Phạm vi**

- Giá trị eccentricity nhỏ có nghĩa nút đó có đường kính cục bộ nhỏ, dễ kết nối mọi người.
- Giá trị lớn có nghĩa nút nằm xa trong cấu trúc mạng.