

Lab05

Quốc Tiến 2186400239

1. Số liệu Mạng Cơ bản

Các số liệu này cung cấp một cái nhìn tổng quan về các thuộc tính cấu trúc của toàn bộ mạng.

- **Bậc Trung bình (Average Degree):**

- **Định nghĩa:** Bậc trung bình của một mạng là số kết nối (cạnh) trung bình mà mỗi nút có trong mạng. Nó cung cấp một thước đo tổng quát về khả năng kết nối của mạng.
- **Công thức:** Đối với đồ thị vô hướng, bậc trung bình ($\langle k \rangle$) được tính bằng:

$$k = \frac{(2|E|)}{|V|}$$

Trong đó $|E|$ là tổng số cạnh và $|V|$ là tổng số nút trong mạng. Đối với đồ thị có hướng, công thức được sửa đổi thành tổng các bậc chia cho số nút.

- **Phạm vi:** Bậc trung bình là một số không âm và có thể thay đổi tùy thuộc vào mạng. Trong một đồ thị đơn giản, giới hạn dưới là 0 (không có cạnh, dẫn đến không có bậc), trong khi giới hạn trên sẽ thay đổi tùy thuộc vào cấu trúc liên kết của đồ thị. Thông thường, điều này sẽ không bao giờ đạt đến $|V| - 1$ vì khi đó mỗi nút sẽ được kết nối với nhau.
- **Diễn giải:**
 - **Bậc trung bình cao:** Cho thấy một mạng dày đặc và được kết nối tốt. Thông tin hoặc ảnh hưởng có thể lan truyền nhanh chóng và rộng rãi.
 - **Bậc trung bình thấp:** Cho thấy một mạng thưa thớt với các kết nối hạn chế, có thể có tốc độ khuếch tán thông tin chậm hơn.
 - **Giá trị “tốt”:** Phụ thuộc vào ngữ cảnh của mạng đang được nghiên cứu. Điều quan trọng là phải xác định xem cấu trúc đồ thị có phù hợp với ngữ cảnh hay không, không có nghĩa là mong đợi một đồ thị mạng xã hội có bậc trung bình thấp.

- **Đường kính Mạng (Network Diameter):**

- **Định nghĩa:** Đường kính mạng là khoảng cách lớn nhất giữa hai nút bất kỳ trong mạng. Khoảng cách ở đây có nghĩa là đường đi ngắn nhất (số bước nhảy) giữa các nút. Nó đại diện cho đường đi ngắn nhất dài nhất trong đồ thị. Nó chỉ ra phạm vi tổng thể của mạng và độ dài đường đi tiềm năng mà một thông điệp phải đi qua để đi từ một điểm đến bất kỳ điểm nào khác trong đồ thị.
- **Công thức:** Đường kính của đồ thị là: $d = \max\{d(u,v) \mid u,v \in V\}$

Trong đó $d(u,v)$ là đường đi ngắn nhất giữa u và v . Việc tính toán đường kính đòi hỏi phải tính toán các đường đi ngắn nhất giữa tất cả các cặp nút.

- **Phạm vi:** Phạm vi bắt đầu từ 1 đối với đồ thị đầy đủ đến tối đa là $|V| - 1$ đối với đồ thị đường đi (tất cả các nút trên một đường thẳng).
- **Diễn giải:**

- **Đường kính nhỏ:** Cho thấy giao tiếp hiệu quả và đường đi ngắn hơn giữa các nút.
- **Đường kính lớn:** Cho thấy mạng có thể mất nhiều bước nhảy để thông tin di chuyển giữa các phần xa của đồ thị.
- **Giá trị “tốt”:** Phụ thuộc vào loại mạng đang được nghiên cứu. Một mạng thế giới nhỏ thường có đường kính nhỏ.

- **Mật độ Đồ thị (Graph Density):**

- **Định nghĩa:** Mật độ đồ thị là tỷ lệ giữa số cạnh thực tế trong mạng so với số cạnh tối đa có thể có trong mạng. Nó đo lường mức độ gần của mạng với một đồ thị đầy đủ. Nó cho thấy sự kết nối tổng thể của mạng.
- **Công thức:** Đối với đồ thị vô hướng, mật độ (ρ) được tính bằng:

$$\rho = \frac{(2|E|)}{(|V|(|V| - 1))}$$

Đối với đồ thị có hướng, nó được tính bằng:

$$\rho = \frac{|E|}{(|V|(|V| - 1))}$$

Trong đó $|E|$ là số cạnh và $|V|$ là số nút.

- **Phạm vi:** Từ 0 (không có cạnh) đến 1 (một đồ thị đầy đủ).
- **Diễn giải:**
 - **Mật độ cao (gần 1):** Cho thấy một mạng mà hầu hết các nút được kết nối với nhiều nút khác.
 - **Mật độ thấp (gần 0):** Cho thấy một mạng thưa thớt với ít cạnh so với số lượng nút.
 - **Giá trị “tốt”:** Một lần nữa, phụ thuộc vào ngữ cảnh, một mạng xã hội sẽ thưa thớt hơn và có mật độ thấp, trong khi một mạng tương tác protein có thể có mật độ cao hơn.

- **Thành phần Liên thông (Connected Components):**

- **Định nghĩa:** Một thành phần liên thông là một mạng con mà trong đó mọi nút có thể đạt đến mọi nút khác bằng cách đi theo các cạnh. Nếu toàn bộ mạng là một thành phần liên thông, thì mọi nút đều có thể đạt đến mọi nút khác.
- **Công thức:** Không có công thức cụ thể; xác định bằng cách duyệt mạng theo tất cả các cạnh và phát hiện các đồ thị con rời rạc.
- **Phạm vi:** Từ 1 (toàn bộ mạng được kết nối) đến số nút tối đa ($|V|$, trong đó mỗi nút bị ngắt kết nối).
- **Diễn giải:**

- **Một thành phần lớn:** Cho thấy mạng được kết nối và thống nhất phần lớn.
- **Nhiều thành phần:** Cho thấy mạng bị phân mảnh, với các nhóm nút bị cô lập với nhau.
- **Giá trị “tốt”:** Phụ thuộc vào cấu trúc mạng mong muốn. Trong một số trường hợp, bạn muốn có một thành phần khổng lồ, trong những trường hợp khác, bạn muốn xác định các cụm riêng biệt và chỉ tương tác với nhau.

- **Độ dài Đường đi Trung bình (Average Path Length):**

- **Định nghĩa:** Độ dài đường đi trung bình là số bước nhảy trung bình giữa tất cả các cặp nút trong mạng. Điều này cung cấp thông tin chi tiết về mức độ hiệu quả của thông tin có thể được truyền từ bất kỳ nút nào đến bất kỳ nút nào khác.
- **Công thức:** Độ dài đường đi trung bình (L) được tính bằng:

$$l_G = \frac{1}{n \cdot (n - 1)} \cdot \sum_{i \neq j} d(v_i, v_j),$$

Trong đó $d(u, v)$ là độ dài đường đi ngắn nhất giữa nút u và v , và N là tổng số nút. Độ dài đường đi trung bình cũng có thể được tính là độ dài đường đi trung bình của tất cả các nút được kết nối chia cho số cặp nút được kết nối.

- **Phạm vi:** Từ 1 (tất cả các nút được kết nối trực tiếp) đến một giá trị tối đa về mặt lý thuyết, nhưng thường thấp hơn đường kính mạng và thường thấp hơn nhiều so với $|V| - 1$.
- **Diễn giải:**
 - **Độ dài đường đi trung bình nhỏ:** Cho thấy giao tiếp hiệu quả và hầu hết các nút đều gần nhau.
 - **Độ dài đường đi trung bình lớn:** Cho thấy mạng có thể không dễ điều hướng hoặc truyền bá thông tin qua.
 - **Giá trị “tốt”:** Độ dài đường đi trung bình nhỏ hơn thường được coi là tốt hơn, đặc biệt là trong các mạng truyền thông. Các mạng thế giới nhỏ thường có độ dài đường đi trung bình nhỏ.

- **Hệ số Kết cụm Trung bình (Average Clustering Coefficient):**

- **Định nghĩa:** Hệ số kết cụm đo lường mức độ mà các nút trong mạng có xu hướng kết cụm với nhau. Đối với một nút đã cho, nó đo lường tỷ lệ kết nối giữa các nút lân cận của nút so với số lượng kết nối tối đa có thể có giữa các nút lân cận của nó.
- **Công thức:**
 - Đối với một nút đơn lẻ (i): $C_i = (2 \cdot E_i) / (k_i \cdot (k_i - 1))$, trong đó E_i là số cạnh giữa các nút lân cận của nút và k_i là bậc của nút.

- Hệ số kết cụm trung bình (C) là trung bình cộng của các hệ số kết cụm của tất cả các nút trong mạng:

$$C = \sum(C_i) / |V|$$

Trong đó $|V|$ là số nút trong mạng.

- **Phạm vi:** Từ 0 (không kết cụm) đến 1 (tất cả các nút lân cận của mỗi nút được kết nối với nhau).
- **Diễn giải:**
 - **Hệ số kết cụm cao:** Cho thấy mạng có cấu trúc cộng đồng mạnh mẽ, trong đó các nút có xu hướng tạo thành các nhóm được kết nối chặt chẽ.
 - **Hệ số kết cụm thấp:** Cho thấy các nút trong mạng không được kết cụm nhiều.
 - **Giá trị “tốt”:** Hệ số kết cụm cao có thể chỉ ra sự hiện diện của các cộng đồng hoặc cấu trúc con được kết nối chặt chẽ, điều này thường quan trọng trong việc hiểu các mạng trong thế giới thực.

2. Số liệu Tính Trung tâm

Các thước đo tính trung tâm giúp xác định các nút có ảnh hưởng hoặc quan trọng nhất trong mạng, dựa trên vị trí và kết nối của chúng.

- **Tính Trung tâm Bậc (Degree Centrality):**

- **Định nghĩa:** Tính trung tâm bậc đo số kết nối mà một nút có. Trong một đồ thị có hướng, chúng ta phân biệt bậc vào (cạnh đến) và bậc ra (cạnh đi). Nó cho thấy những nút nào được kết nối nhiều hơn những nút khác, với bậc cao hơn cho thấy ảnh hưởng hoặc khả năng kết nối lớn hơn.
- **Công thức:** Đối với đồ thị vô hướng, tính trung tâm bậc là bậc của nút:

$$\text{degree}(i) = k_i$$

Trong đồ thị có hướng, có tính trung tâm bậc vào (số cạnh đi vào) và tính trung tâm bậc ra (số cạnh đi ra)

$$\text{in-degree}(i) = k_i(\text{in})$$

$$\text{out-degree}(i) = k_i(\text{out})$$

- **Phạm vi:** Từ 0 đến $|V| - 1$.
- **Diễn giải:**
 - **Tính trung tâm bậc cao:** Cho thấy một nút có nhiều kết nối và có khả năng đóng vai trò nổi bật trong mạng.
 - **Tính trung tâm bậc thấp:** Cho thấy một nút có ít kết nối và có thể có ít ảnh hưởng hơn.
 - **Giá trị “tốt”:** Tính trung tâm bậc cao hơn thường được đánh đồng với tầm quan trọng lớn hơn trong một mạng; tuy nhiên, các số liệu khác cũng nên được xem xét để đánh giá đầy đủ vai trò của một nút.

- **Tính Trung tâm Trung gian (Betweenness Centrality):**

- **Định nghĩa:** Tính trung tâm trung gian đo tần suất một nút nằm trên các đường đi ngắn nhất giữa các cặp nút khác trong mạng. Nó xác định các nút bắc cầu các phần khác nhau của mạng.
- **Công thức:** Tính trung tâm trung gian (CB) của một nút i được tính bằng:

$$CB(i) = \sum (\sigma(u, v | i) / \sigma(u, v))$$

trong đó $\sigma(u, v)$ là số đường đi ngắn nhất giữa các nút u và v , và $\sigma(u, v | i)$ là số đường đi ngắn nhất giữa u và v đi qua nút i . Tổng được lấy trên tất cả các cặp nút u và v ngoại trừ $u=i$ và $v=i$.

- **Phạm vi:** Phạm vi phụ thuộc vào cấu trúc mạng, nhưng thường nằm trong khoảng từ 0 đến giá trị tối đa liên quan đến số lượng nút và vị trí của chúng.
- **Diễn giải:**
 - **Tính trung tâm trung gian cao:** Cho thấy rằng một nút là cầu nối giữa các phần khác nhau của mạng. Nút này là một điểm kiểm soát quan trọng đối với dòng thông tin.
 - **Tính trung tâm trung gian thấp:** Cho thấy nút nằm trên ít đường đi ngắn nhất và có khả năng ít ảnh hưởng hơn trong việc kiểm soát dòng thông tin qua mạng.
 - **Giá trị “tốt”:** Các nút có tính trung tâm trung gian cao hơn thường là các điểm quan trọng trong mạng.

- **Tính Trung tâm Gần (Closeness Centrality):**

- **Định nghĩa:** Tính trung tâm gần đo khoảng cách trung bình của một nút đến tất cả các nút khác trong mạng. Nó xác định các nút ‘gần’ với tất cả các nút khác trong mạng. Số liệu này có thể tiết lộ mức độ dễ dàng mà một nút có thể lan tỏa ảnh hưởng đến phần còn lại của mạng.
- **Công thức:** Tính trung tâm gần (CC) của một nút i được tính bằng:

$$CC(i) = |V| / \sum (d(i, j))$$

trong đó $d(i, j)$ là khoảng cách ngắn nhất giữa các nút i và j , và tổng được lấy trên tất cả các nút j trong mạng ngoại trừ $i=j$.

- **Phạm vi:** Từ 0 đến 1. Trong đó 1 có nghĩa là nút nằm trong 1 bước nhảy của mỗi nút khác và các số thấp hơn có nghĩa là nó ở xa hơn so với mức trung bình.
- **Diễn giải:**
 - **Tính trung tâm gần cao:** Cho thấy rằng một nút là trung tâm và có thể dễ dàng tiếp cận các nút khác.
 - **Tính trung tâm gần thấp:** Cho thấy nút ở xa hơn các nút khác và có khả năng bị cô lập hơn.
 - **Giá trị “tốt”:** Các nút có tính trung tâm gần cao hơn có thể có khả năng lan truyền thông tin hoặc ảnh hưởng hiệu quả hơn trong mạng.

- **Tính Trung tâm Vector Riêng (Eigenvector Centrality):**

- **Định nghĩa:** Tính trung tâm vector riêng đo lường ảnh hưởng của một nút bằng cách xem xét ảnh hưởng của các nút lân cận của nó. Ý tưởng là các nút được kết nối với các nút được kết nối nhiều khác có tính trung tâm vector riêng cao hơn.
- **Công thức:** Tính trung tâm vector riêng (x) được xác định bằng phương trình vector riêng:

$$Ax = \lambda x$$

trong đó A là ma trận kề của đồ thị và λ là một giá trị riêng. Tính trung tâm vector riêng của nút i (x_i) là thành phần thứ i của vector riêng liên kết với giá trị riêng lớn nhất.

- **Phạm vi:** Các giá trị không âm, với các giá trị cao hơn đại diện cho các nút có ảnh hưởng hơn.
- **Diễn giải:**
 - **Tính trung tâm vector riêng cao:** Cho thấy một nút được kết nối với nhiều nút được kết nối nhiều khác.
 - **Tính trung tâm vector riêng thấp:** Cho thấy một nút không được kết nối với các nút có ảnh hưởng.
 - **Giá trị “tốt”:** Nút có tính trung tâm vector riêng cao nhất là quan trọng nhất.

- **PageRank:**

- **Định nghĩa:** PageRank là một thuật toán ban đầu được phát triển để tối ưu hóa công cụ tìm kiếm và đo lường tầm quan trọng của một nút bằng cách xem xét có bao nhiêu nút khác trỏ đến nó. Nó đánh trọng số các liên kết đến dựa trên tầm quan trọng của các nguồn của chúng.
- **Công thức:** PageRank được xác định bằng công thức đệ quy:

$$PR(i) = (1-d) + d * \sum (PR(j) / L(j))$$

trong đó $PR(i)$ là PageRank của nút i , d là hệ số giảm chấn (thường được đặt thành 0,85), $L(j)$ là số liên kết đi từ nút j và tổng được lấy trên tất cả các nút j có liên kết đến nút i .

- **Phạm vi:** Các giá trị không âm, trong đó điểm số cao hơn biểu thị các nút có ảnh hưởng hơn.
- **Diễn giải:**
 - **PageRank cao:** Cho thấy nút được nhiều trang quan trọng trỏ đến.
 - **PageRank thấp:** Cho thấy rằng nút có ít kết nối đến hoặc các kết nối đó là từ các trang không quan trọng.
 - **Giá trị “tốt”:** Nút có PageRank cao nhất là nút quan trọng nhất.

- **HITS (Hub and Authority):**

- **Định nghĩa:** HITS (Hyperlink-Induced Topic Search) tính hai điểm cho mỗi nút: điểm trung tâm (hub) và điểm thẩm quyền (authority). Các trung tâm là

các nút trở đến nhiều thẩm quyền và các thẩm quyền được nhiều trung tâm trở đến. Điều này cung cấp các thước đo khác nhau cho sự phù hợp và vai trò của một nút trong mạng.

- **Công thức:** Điểm trung tâm (h) và điểm thẩm quyền (a) được tính bằng phương pháp lặp:

$h(i) = \sum(a(j))$, trong đó tổng là trên tất cả các nút j mà nút i trở đến.

$a(i) = \sum(h(j))$, trong đó tổng là trên tất cả các nút j trở đến nút i .

Các giá trị trung tâm và thẩm quyền ban đầu thường được giả định là đồng nhất.

- **Phạm vi:** Các giá trị không âm, với các giá trị cao hơn cho các trung tâm và thẩm quyền có ảnh hưởng lớn hơn.
- **Diễn giải:**
 - **Điểm trung tâm cao:** Cho thấy một nút liên kết đến nhiều thẩm quyền quan trọng.
 - **Điểm thẩm quyền cao:** Cho thấy một nút được nhiều trung tâm liên kết đến.
 - **Giá trị “tốt”:** Nút có điểm trung tâm cao nhất là một trung tâm tốt, nút có điểm thẩm quyền cao nhất là một thẩm quyền tốt.
- **Độ lệch tâm (Eccentricity):**
 - **Định nghĩa:** Độ lệch tâm của một nút là đường đi ngắn nhất tối đa từ nút đó đến tất cả các nút khác.
 - **Công thức:**

$$ecc(i) = \max\{d(i,j) \mid j \in V\}$$

Trong đó $d(i,j)$ là đường đi ngắn nhất giữa nút i và nút j .

- **Phạm vi:** Một giá trị không âm, từ 0 trong đồ thị một nút đến đường kính mạng.
- **Diễn giải:**
 - **Độ lệch tâm cao:** Cho thấy một nút ở xa một số nút trong mạng.
 - **Độ lệch tâm thấp:** Cho thấy một nút có vị trí trung tâm và “gần” với tất cả các nút khác, không chỉ ở mức trung bình như trong Tính Trung tâm Gần.
 - **Giá trị “tốt”:** Các nút có độ lệch tâm thấp thường là các nút trung tâm.

