

Lab05.01

Hung Nguyen

December 2024

1 Độ đo cơ bản của mạng

Average Degree

- **Ý nghĩa:** Độ trung bình mô tả số cạnh trung bình trên mỗi nút trong đồ thị
- **Công thức:** Average Degree = $\frac{m}{n}$
- **Phạm vi:** $[0, 1]$
- **Chỉ số tốt:** Giá trị cao hơn cho thấy sự dày đặc trong kết nối giữa các đỉnh.

Network Diameter

- **Ý nghĩa:** Đường kính của mạng là độ dài đường đi dài nhất giữa hai đỉnh trong mạng.
- **Công thức:** Là giá trị lớn nhất của khoảng cách ngắn nhất giữa tất cả các cặp đỉnh.
- **Phạm vi:** Từ 1 đến số đỉnh trong đồ thị.
- **Chỉ số tốt:** Giá trị nhỏ cho thấy mạng có khả năng kết nối nhanh chóng giữa các đỉnh.

Graph Density

- **Ý nghĩa:** Mật độ đồ thị mô tả mức độ liên kết của đồ thị, là tỷ lệ số cạnh thực tế so với số cạnh tối đa có thể.
- **Công thức:** Density = $\frac{2k}{n(n-1)}$
- **Phạm vi:** $[0, 1]$
- **Chỉ số tốt:** Giá trị cao hơn chỉ ra rằng đồ thị có nhiều cạnh hơn, cho thấy mức độ kết nối cao.

Connected Components

- **Ý nghĩa:** là độ dài trung bình của đường đi ngắn nhất giữa tất cả các cặp đỉnh.
- **Công thức:** Đếm số nhóm đỉnh mà mỗi nhóm có thể truy cập lẫn nhau nhưng không truy cập được với các đỉnh ngoài nhóm.
- **Phạm vi:** $[1, n]$
- **Chỉ số tốt:** Giá trị thấp hơn cho thấy mạng liên thông tốt hơn.

Average Path Length

- **Ý nghĩa:** Hệ số phân cụm trung bình cho biết mức độ mà các đỉnh trong mạng có xu hướng tạo thành các nhóm kín.
- **Công thức:** Average Path Length = $\frac{\sum_{u \neq v} d(u, v)}{\binom{n}{2}}$
- **Phạm vi:** $[1, n]$
- **Chỉ số tốt:** Giá trị nhỏ hơn cho thấy độ hiệu quả cao trong việc truyền thông tin giữa các đỉnh.

Average Clustering Coefficient

- **Ý nghĩa:** Số thành phần liên thông trong đồ thị, nơi không còn đường đi nào giữa các thành phần này.
- **Công thức:** Average Clustering Coefficient = $\frac{1}{n} \sum_{v \in V} C(v)$
- **Phạm vi:** từ 1 đến đường kính của mạng.
- **Chỉ số tốt:** Giá trị cao hơn cho thấy sự liên kết chặt chẽ giữa các đỉnh, hình thành các cộng đồng.

2 Độ đo tính trung tâm

Degree Centrality

- **Ý nghĩa:** Số lượng các mối quan hệ trực tiếp của một tác nhân với các thành viên khác trong mạng xã hội.
- **Công thức:** $C_D(v) = \frac{\deg(v)}{n-1}$
trong đó $\deg(v)$: tổng số liên kết trực tiếp đến đỉnh v (bậc của đỉnh)
- **Phạm vi:** $[0, 1]$
- **Chỉ số tốt:** Giá trị càng gần 1 tính trung tâm càng lớn

Betweenness Centrality

- **Ý nghĩa:** Đo lường mức độ một đỉnh nằm trên đường đi giữa các cặp đỉnh khác, cho thấy khả năng kiểm soát thông tin trong mạng.
- **Công thức:** $C_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$
trong đó
 $\sigma_{st}(v)$: số đường đi ngắn nhất từ s đến t
 σ_{st} : số đường đi ngắn nhất từ s đến t qua v
Chuẩn hóa: $C'_B(v) = \frac{C_B(v)}{(n-1)(n-2)/2}$ (vô hướng)
 $C'_B(v) = \frac{C_B(v)}{(n-1)(n-2)}$ (có hướng)
- **Phạm vi:** $[0, 1]$
- **Chỉ số tốt:** Số đo này càng lớn thì actor càng quan trọng trong việc kiểm soát thông tin và giao dịch trong mạng.

Closeness Centrality

- **Ý nghĩa:** Đo lường mức độ gần gũi của một đỉnh với tất cả các đỉnh khác trong mạng
- **Công thức:** $C_C(v) = \frac{1}{\sum_{t \in V \setminus \{v\}} d_G(v, t)}$
trong đó $d_G(v, t)$ là chiều dài đường đi ngắn nhất từ đỉnh v tới đỉnh t
Chuẩn hóa: $CC(v) = (n-1)C_c(v)$
- **Phạm vi:** $[0, 1]$
- **Chỉ số tốt:** Khoảng cách càng nhỏ, khả năng truyền tin càng lớn.

Eigenvector Centrality

- **Ý nghĩa:** Một biến thể của tính trung tâm bậc, nó không chỉ tính số liên kết trực tiếp mà còn đánh giá cả ảnh hưởng của các liên kết gián tiếp qua các đỉnh khác.
- **Công thức:** $x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t$
- **Phạm vi:** $[0, 1]$
- **Chỉ số tốt:** Giá trị cao hơn cho thấy đỉnh không chỉ kết nối tốt mà còn được kết nối với các đỉnh có ảnh hưởng cao.

PageRank

- **Ý nghĩa:** Thuật toán nổi tiếng của Google để xếp hạng các trang web dựa trên số lượng và chất lượng các liên kết đến trang đó.
- **Công thức:** $PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}$
- **Phạm vi:** $[0, 1]$
- **Chỉ số tốt:** Giá trị cao hơn cho thấy trang hoặc đỉnh có ảnh hưởng và độ tin cậy cao trong mạng.

HITS (Hub and Authority)

- **Ý nghĩa:** Là một thuật toán xếp hạng phụ thuộc truy vấn tìm kiếm cho việc tìm kiếm thông tin trên web.
- **Công thức:** $auth(p) = \sum_{q \in P_{to}} hub(q)$
 $hub(p) = \sum_{q \in P_{from}} auth(q)$
- **Phạm vi:** phụ thuộc vào q
- **Chỉ số tốt:** Hub scores cao cho thấy đỉnh là một trung tâm tốt cho thông tin, trong khi authority scores cao cho thấy đỉnh là một nguồn thông tin có uy tín.

Eccentricity

- **Ý nghĩa:** Độ lệch tâm của một đỉnh đo lường khoảng cách xa nhất từ đỉnh đó đến bất kỳ đỉnh nào khác trong đồ thị.
- **Công thức:**

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

trong đó $d(v, u)$ là khoảng cách ngắn nhất từ đỉnh v đến đỉnh u trong đồ thị, và V là tập hợp tất cả các đỉnh trong đồ thị.

- **Phạm vi:** Từ 1 đến đường kính của đồ thị.
- **Chỉ số tốt:** Giá trị thấp hơn cho thấy đỉnh có khả năng kết nối nhanh hơn với các đỉnh khác trong mạng.