

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

1 Độ Đo Cơ Bản của Mạng (Basic Network Metrics)

1.1 Average Degree

Ý nghĩa: Trung bình số lượng cạnh kết nối với mỗi nút.

Công thức:

$$\text{Average Degree} = \frac{2 \times |E|}{|V|}$$

với $|E|$ là số cạnh và $|V|$ là số nút trong đồ thị.

Phạm vi: Từ 0 (không có cạnh) đến $|V| - 1$ (nếu là đồ thị đầy đủ).

Chỉ số tốt: Giá trị cao biểu thị mạng kết nối dày đặc.

1.2 Network Diameter

Ý nghĩa: Khoảng cách lớn nhất giữa hai nút xa nhất.

Công thức:

$$\text{Diameter} = \max_{u,v \in V} \text{Shortest Path Length}(u, v)$$

Phạm vi: Từ 1 (đồ thị đầy đủ) đến $|V| - 1$ (đồ thị hình đường thẳng).

Chỉ số tốt: Giá trị nhỏ biểu thị mạng hiệu quả trong truyền tải thông tin.

1.3 Graph Density

Ý nghĩa: Tỷ lệ giữa số cạnh thực tế và số cạnh tối đa có thể tồn tại.

Công thức:

$$\text{Density} = \frac{2 \times |E|}{|V| \times (|V| - 1)}$$

Phạm vi: Từ 0 (không có cạnh) đến 1 (đồ thị đầy đủ).

Chỉ số tốt: Giá trị cao cho thấy mạng dày đặc kết nối.

1.4 Connected Components

Ý nghĩa: Số thành phần liên thông của đồ thị.

Công thức: Đếm số nhóm các nút mà mọi nút trong nhóm đều được kết nối.

Phạm vi: Từ 1 (đồ thị liên thông hoàn toàn) đến $|V|$ (không có cạnh).

Chỉ số tốt: Ít thành phần liên thông biểu thị mạng mạnh hơn.

1.5 Average Path Length

Ý nghĩa: Khoảng cách trung bình giữa tất cả các cặp nút trong mạng.

Công thức:

$$APL = \frac{\sum_{u,v \in V} \text{Shortest Path Length}(u, v)}{|V| \times (|V| - 1)}$$

Phạm vi: Từ 1 (đồ thị đầy đủ) đến một giá trị lớn hơn trong mạng thưa thớt.

Chỉ số tốt: Giá trị thấp biểu thị mạng hiệu quả.

1.6 Average Clustering Coefficient

Ý nghĩa: Trung bình mức độ mà các nút lân cận của một nút cũng được kết nối với nhau.

Công thức:

$$\text{Clustering Coefficient} = \frac{1}{|V|} \sum_{v \in V} \frac{\text{Number of edges between neighbors of } v}{\text{Maximum possible edges between neighbors of } v}$$

Phạm vi: Từ 0 (không có kết nối giữa các hàng xóm) đến 1 (tất cả hàng xóm kết nối).

Chỉ số tốt: Giá trị cao biểu thị mạng có tính cụm mạnh.

2 Độ Đo Tính Trung Tâm (Centrality Metrics)

2.1 Degree Centrality

Ý nghĩa: Đo lường mức độ kết nối của một nút, dựa trên số cạnh trực tiếp kết nối với nó.

Công thức:

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

với $\deg(v)$ là số cạnh kết nối với nút v và $|V|$ là số nút trong đồ thị.

Phạm vi: Từ 0 (nút cô lập) đến 1 (kết nối với tất cả các nút khác).

Chỉ số tốt: Giá trị cao cho thấy nút có tầm ảnh hưởng mạnh.

2.2 Betweenness Centrality

Ý nghĩa: Đo lường tầm quan trọng của một nút dựa trên số đường đi ngắn nhất đi qua nó.

Công thức:

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

với σ_{st} là số đường đi ngắn nhất giữa s và t , và $\sigma_{st}(v)$ là số đường đi đó đi qua v .

Phạm vi: Từ 0 (nút không nằm trên đường đi ngắn nào) đến một giá trị lớn hơn (khi nút là trung gian quan trọng).

Chỉ số tốt: Giá trị cao biểu thị nút là cầu nối quan trọng.

2.3 Closeness Centrality

Ý nghĩa: Đo lường mức độ gần gũi của một nút với tất cả các nút khác trong mạng.

Công thức:

$$C_C(v) = \frac{|V| - 1}{\sum_{u \in V} d(u, v)}$$

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

Phạm vi: Giá trị cao hơn biểu thị nút có vị trí chiến lược, gần với các nút khác.

2.4 Eigenvector Centrality

Ý nghĩa: Đo lường mức độ ảnh hưởng của một nút dựa trên sự kết nối với các nút quan trọng khác.

Công thức: Dựa trên giá trị riêng của ma trận kề:

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

với A_{vu} là phần tử trong ma trận kề và λ là giá trị riêng lớn nhất.

Phạm vi: Giá trị cao biểu thị nút có tầm ảnh hưởng lớn trong mạng.

2.5 PageRank

Ý nghĩa: Được sử dụng trong xếp hạng các trang web, dựa trên độ phổ biến và quan trọng của các trang kết nối.

Công thức:

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

với d là hệ số suy giảm (thường là 0.85).

Phạm vi: Giá trị cao cho thấy nút quan trọng trong mạng.

2.6 HITS (Hub and Authority)

Ý nghĩa: Đánh giá nút dựa trên hai khía cạnh: nút là trung tâm (Hub) hay có độ tin cậy (Authority).

Công thức:

$$\text{Authority}(v) = \sum_{u \in \text{in-neighbors}(v)} \text{Hub}(u)$$

$$\text{Hub}(v) = \sum_{u \in \text{out-neighbors}(v)} \text{Authority}(u)$$

Phạm vi: Giá trị cao cho thấy nút là trung tâm (Hub) hoặc có độ tin cậy cao (Authority).

2.7 Eccentricity

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

Công thức:

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

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

Phạm vi: Từ 1 (đồ thị đầy đủ) đến ∞ (đồ thị không liên thông).

Chỉ số tốt: Giá trị thấp biểu thị nút trung tâm hơn.