

BÀI THỰC HÀNH SỐ 1

Tìm hiểu các chỉ số thống kê cơ bản sử dụng Gephi

1. Độ đo cơ bản của mạng

Chỉ số	Ý nghĩa	Công thức	Phạm vi	Chỉ số tốt
Average Degree	Trung bình số lượng cạnh của các nút trong mạng	$k = \frac{2E}{N}$ E: số cạnh N: số nút	$[0, N - 1]$,	Mạng có mức kết nối vừa phải, không quá thưa (degree thấp) hoặc quá dày (degree cao).
Network Diameter	Khoảng cách lớn nhất (tính bằng số cạnh) giữa 2 nút xa nhất trong mạng	$D = \max_{ij} d(i,j)$ $d(i,j)$: độ dài đường đi ngắn nhất giữa nút i và j .	$D \geq 1$	Đường kính nhỏ hơn thể hiện khả năng lan truyền thông tin nhanh hơn.
Graph Density	Tỷ lệ số cạnh hiện có so với số cạnh tối đa có thể có trong mạng. Thể hiện mật độ mạng của mạng.	$\frac{2E}{N(N-1)/2}$ E: số cạnh N: số nút	$[0,1]$	Giá trị gần 1 biểu thị mạng kết nối dày đặc, giá trị gần 0 biểu thị mạng thưa thớt.
Connected Components	Số lượng thành phần liên thông trong mạng. Một thành phần liên thông là một nhóm các nút mà giữa chúng luôn có đường đi.	Không có công thức cố định, đếm số lượng thành phần liên thông (connected components).	$[1, N]$	Giá trị nhỏ hơn biểu thị mạng kết nối mạnh, giá trị lớn hơn biểu thị mạng phân mảnh hoặc ít kết nối.
Average Path Length	Độ dài trung bình của đường đi ngắn nhất giữa tất cả các cặp nút trong mạng.	$l = \frac{\sum_{i \neq j} d(i,j)}{N(N-1)}$ $d(i,j)$: độ dài đường đi ngắn nhất giữa nút	$l \geq 1$	Giá trị nhỏ biểu thị mạng kết nối hiệu quả hơn, dễ dàng lan truyền thông tin.

		i và j .		
Average Clustering Coefficient	Trung bình hệ số phân cụm (clustering coefficient) của tất cả các nút. Hệ số này đo lường mức độ liên kết chặt chẽ giữa các nút lân cận.	$C_i = \frac{2e_i}{k_i(k_i - 1)}$ $C = \frac{1}{N} \sum_{i=1}^N C_i$ e_i : số cạnh giữa các lân cận của nút i k_i : degree của nút i .	[0,1]	Giá trị cao biểu thị các nút có xu hướng kết nối thành cụm, giá trị thấp biểu thị mạng ít có cấu trúc cụm.

2. Độ đo tính trung tâm

Chỉ số	Ý nghĩa	Công thức	Phạm vi	Chỉ số tốt
Degree Centrality	Đo số lượng các mối quan hệ trực tiếp của một tác nhân với các thành viên khác trong mạng xã hội.	$C_D = \frac{\deg(v)}{n - 1}$ $\deg(v)$: bậc của đỉnh n : số đỉnh In-degree: $In - degree(v) = \sum_{u \in V} A(u, v)$ Out-degree: $Out - degree(v) = \sum_{u \in V} A(v, u)$	[0,1]	Giá trị cao biểu thị nút có nhiều kết nối vào (in-degree) hoặc ra (out-degree), có vai trò trung tâm trong việc gửi hoặc nhận thông tin.
Betweenness Centrality	Số đo này càng lớn thì actor càng quan trọng trong việc kiểm soát thông tin và giao dịch trong mạng.	$C_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$ $\sigma_{st}(v)$: số đường đi ngắn nhất từ s đến t qua v σ_{st} : số đường đi ngắn nhất từ s đến t .	[0,1]	Giá trị cao biểu thị nút đóng vai trò "cầu nối" trong mạng, quan trọng cho việc truyền thông tin hoặc kết nối các nhóm khác nhau.

Closeness Centrality	Số đo này tương ứng với thời gian cần thiết để thông tin truyền từ một actor tới các actor khác.	$C_c(v) = \frac{1}{\sum_{t \in V} d_G(v, t)}$ $d_G(v, t): \text{ chiều dài đường đi ngắn nhất từ đỉnh } v \text{ tới đỉnh } t.$ <p>Công thức chuẩn hóa: $C_c(v) = (n - 1)C_c(v)$ n: đỉnh đồ thị</p>	[0,1]	Khoảng cách càng nhỏ, khả năng truyền tin càng lớn.
Eigenvector Centrality	Đo mức độ quan trọng của một nút dựa trên tầm quan trọng của các nút mà nó kết nối.	$x = \lambda Ax$ <p>A: ma trận kề λ: giá trị riêng lớn nhất X: vector giá trị riêng</p>	[0,1]	Giá trị cao biểu thị nút có khả năng được truy cập nhiều trong mạng, thường áp dụng trong đánh giá các trang web.
PageRank	Đo tầm quan trọng của một nút trong mạng dựa trên khả năng được "ghé thăm" ngẫu nhiên trong một mạng có hướng.	$PR(v) = (1 - d) + d \sum_{u \in in(v)} \frac{PR(u)}{k_{out}(u)}$ <p>d: hệ số giảm dần</p>	$l \geq 1$	Giá trị nhỏ biểu thị mạng kết nối hiệu quả hơn, dễ dàng lan truyền thông tin.
HITS (Hub and Authority)	Phân tích vai trò của nút trong mạng: Hub: Nút trở tới các nút quan trọng. Authority: Nút được nhiều nút quan trọng trở đến.	<p>Hub:</p> $h(v) = \sum_{u \in out(v)} a(u)$ <p>Authority:</p> $a(v) = \sum_{u \in in(v)} h(u)$	[0,1]	Giá trị Hub cao biểu thị nút cung cấp thông tin tốt; giá trị Authority cao biểu thị nút có thông tin được nhiều nguồn đáng tin cậy trở đến.
Eccentricity	Khoảng cách lớn nhất từ một nút đến bất kỳ nút	$e(v) = \max_{u \in V} d(v, u)$ <p>$d(v, u)$: độ dài đường đi ngắn nhất giữa v và u.</p>	$e \geq 0$	Giá trị thấp biểu thị nút nằm gần trung tâm của mạng, dễ dàng

	nào khác trong mạng.			tiếp cận các nút khác.
--	-------------------------	--	--	---------------------------