

1 Độ đo cơ bản của mạng (Basic Network Metrics)

1.1 Average Degree

Average Degree là một chỉ số cơ bản trong phân tích mạng. Nó được tính bằng trung bình số liên kết của các đỉnh trong mạng. Công thức:

$$\text{Average Degree} = \frac{2E}{N}$$

trong đó:

- E là số cạnh của mạng.
- N là số đỉnh của mạng.

Chỉ số này cho biết mức độ kết nối trung bình của các đỉnh trong mạng.

1.2 Network Diameter

Network Diameter (Đường kính mạng) là khoảng cách dài nhất giữa hai đỉnh bất kỳ trong mạng. Đây là một chỉ số quan trọng để đánh giá khả năng truyền thông tin trong mạng.

Công thức:

$$\text{Network Diameter} = \max_{u,v \in V} \text{dist}(u, v)$$

trong đó:

- V là tập hợp các đỉnh trong mạng.
- $\text{dist}(u, v)$ là khoảng cách ngắn nhất giữa hai đỉnh u và v .

Phạm vi giá trị:

- Giá trị đường kính càng nhỏ thì mạng càng "ngắn", nghĩa là thông tin có thể truyền nhanh hơn giữa các đỉnh.
- Đối với mạng phân tán hoặc mạng xã hội, một đường kính nhỏ thường là đặc trưng của một mạng hiệu quả.

Chỉ số tốt:

- Một chỉ số đường kính tốt phụ thuộc vào mục tiêu của mạng.
- Trong các mạng cần tính hiệu quả truyền thông cao (như mạng xã hội hoặc mạng cảm biến), một đường kính nhỏ là mong muốn.
- Tuy nhiên, trong các mạng lớn (như Internet), đường kính có thể lớn nhưng vẫn đảm bảo chức năng.

1.3 Graph Density

Graph Density (Mật độ đồ thị) đo lường mức độ liên kết giữa các đỉnh trong mạng, tức là tỷ lệ giữa số lượng cạnh hiện có và số lượng cạnh tối đa có thể có trong mạng.

Công thức:

$$\text{Density} = \frac{2E}{N(N-1)}$$

trong đó:

- E là số cạnh của đồ thị.
- N là số đỉnh của đồ thị.

Phạm vi giá trị:

- $0 \leq \text{Density} \leq 1$
- $\text{Density} = 0$: khi đồ thị không có cạnh (đồ thị rỗng).
- $\text{Density} = 1$: khi đồ thị đầy đủ (mỗi đỉnh được kết nối với tất cả các đỉnh khác).

Ý nghĩa:

- Mật độ cao (gần 1) chỉ ra rằng mạng có rất nhiều liên kết giữa các đỉnh, thường là đặc điểm của các mạng nhỏ hoặc nhóm kết nối chặt chẽ.
- Mật độ thấp (gần 0) thường xuất hiện trong các mạng lớn, phân tán, chẳng hạn như mạng xã hội hoặc mạng thông tin.

Chỉ số tốt:

- Tùy thuộc vào ứng dụng, mật độ cao hay thấp đều có ý nghĩa riêng:
 - Mật độ cao phù hợp với các mạng cần tính liên kết mạnh, như mạng nội bộ hoặc nhóm nhỏ.
 - Mật độ thấp phù hợp với các mạng phân tán lớn, như Internet hoặc mạng cảm biến.

1.4 Connected Components

Connected Components (Thành phần liên thông) là các tập hợp con của đồ thị mà trong đó bất kỳ hai đỉnh nào cũng được kết nối với nhau bởi các đường đi, và không có đỉnh nào trong tập hợp được kết nối với đỉnh nằm ngoài tập hợp đó.

Ý nghĩa:

- Thành phần liên thông giúp xác định các phần tách biệt trong mạng.
- Trong một đồ thị vô hướng:
 - Nếu có nhiều thành phần liên thông, mạng bị chia thành nhiều phần tách biệt.
 - Nếu chỉ có một thành phần liên thông, mạng được gọi là liên thông.

Công thức:

- Xác định thông qua thuật toán duyệt đồ thị (DFS hoặc BFS).

Phạm vi giá trị:

- Số lượng thành phần liên thông ≥ 1 .
- Trong một đồ thị liên thông, số lượng thành phần liên thông bằng 1.
- Trong đồ thị không có cạnh, mỗi đỉnh là một thành phần liên thông (số thành phần liên thông bằng số đỉnh).

Chỉ số tốt:

- Một mạng có số lượng thành phần liên thông ít (đặc biệt là chỉ 1) thường hiệu quả hơn trong việc truyền thông tin.
- Trong các ứng dụng cụ thể (ví dụ: phân tích cộng đồng), nhiều thành phần liên thông có thể đại diện cho các nhóm hoặc cụm khác nhau trong mạng.

1.5 Average Path Length

Average Path Length (Độ dài đường đi trung bình) là giá trị trung bình của tất cả các đường đi ngắn nhất giữa mọi cặp đỉnh trong mạng. Đây là một chỉ số quan trọng để đánh giá hiệu quả truyền thông tin trong mạng.

Công thức:

$$\text{Average Path Length} = \frac{\sum_{i \neq j} \text{dist}(i, j)}{N(N-1)}$$

trong đó:

- N là số đỉnh trong mạng.
- $\text{dist}(i, j)$ là khoảng cách ngắn nhất giữa hai đỉnh i và j .

Phạm vi giá trị:

- $0 \leq \text{Average Path Length} < \infty$
- Giá trị thấp cho thấy mạng có khả năng truyền thông tin nhanh giữa các đỉnh.
- Giá trị cao thường gặp ở các mạng lớn hoặc ít liên kết.

Ý nghĩa:

- Độ dài đường đi trung bình càng nhỏ, mạng càng "nhỏ" và hiệu quả trong việc truyền thông tin.
- Trong các mạng xã hội hoặc mạng cảm biến, giá trị nhỏ thể hiện tính kết nối mạnh và khả năng lan truyền thông tin nhanh chóng.

Chỉ số tốt:

- Một chỉ số "tốt" phụ thuộc vào mục đích mạng:
 - Trong các mạng cần tính hiệu quả cao (như mạng Internet hoặc mạng giao thông), độ dài trung bình ngắn là mục tiêu.
 - Trong các mạng phân tán (như mạng phân phối điện), giá trị trung bình lớn hơn có thể chấp nhận được.

1.6 Average Clustering Coefficient

Average Clustering Coefficient (Hệ số cụm trung bình) đo lường mức độ mà các đỉnh trong mạng có xu hướng hình thành các cụm (nhóm liên kết chặt chẽ). Đây là một chỉ số quan trọng để đánh giá cấu trúc của mạng, đặc biệt trong mạng xã hội và mạng phức tạp.

Công thức:

$$\text{Clustering Coefficient} = \frac{\text{Number of closed triplets}}{\text{Number of all triplets (open and closed)}}$$

$$\text{Average Clustering Coefficient} = \frac{1}{N} \sum_{i=1}^N C_i$$

trong đó:

- C_i là hệ số cụm của đỉnh i , được tính bằng:

$$C_i = \frac{2E_i}{k_i(k_i - 1)}$$

với:

- E_i là số cạnh thực tế giữa các láng giềng của đỉnh i .
- k_i là số láng giềng của đỉnh i (bậc của đỉnh i).

- N là tổng số đỉnh trong mạng.

Phạm vi giá trị:

- $0 \leq C_i \leq 1$
- $C_i = 0$: Đỉnh i không có kết nối nào giữa các láng giềng của nó.
- $C_i = 1$: Tất cả các láng giềng của đỉnh i được kết nối đầy đủ với nhau.

Ý nghĩa:

- Hệ số cụm cao cho thấy các cụm hoặc nhóm trong mạng được kết nối chặt chẽ, thường gặp trong các mạng xã hội hoặc mạng cộng đồng.
- Hệ số cụm thấp cho thấy các kết nối ít tập trung hơn.

Chỉ số tốt:

- Mạng có hệ số cụm trung bình cao thường phù hợp với các ứng dụng yêu cầu tính liên kết mạnh giữa các nhóm, như mạng xã hội hoặc mạng giao thông nội bộ.
- Trong các mạng phân tán lớn (như Internet), hệ số cụm trung bình có thể thấp nhưng vẫn đảm bảo chức năng.

2 Độ đo tính trung tâm (Centrality Metrics)

2.1 Degree Centrality (In-degree và Out-degree với đồ thị có hướng)

Degree Centrality (Tính trung tâm theo bậc) đo lường mức độ quan trọng của một đỉnh trong mạng dựa trên số lượng liên kết của nó. Với đồ thị có hướng, Degree Centrality được chia thành hai loại: In-degree (số liên kết đi vào) và Out-degree (số liên kết đi ra).

Công thức:

- **In-degree Centrality:**

$$C_{in}(v) = \deg_{in}(v)$$

- **Out-degree Centrality:**

$$C_{out}(v) = \deg_{out}(v)$$

trong đó:

- $\deg_{in}(v)$: Số liên kết đi vào đỉnh v .
- $\deg_{out}(v)$: Số liên kết đi ra từ đỉnh v .

Phạm vi giá trị:

- Với đồ thị có N đỉnh:

$$0 \leq C_{in}(v), C_{out}(v) \leq N - 1$$

- Giá trị cao của $C_{in}(v)$ cho thấy đỉnh v nhận được nhiều kết nối (nó là "điểm thu hút").

- Giá trị cao của $C_{out}(v)$ cho thấy đỉnh v có ảnh hưởng lớn (nó là "điểm phát tán").

Ý nghĩa:

- **In-degree Centrality:** Đo lường mức độ phổ biến của đỉnh (được nhận biết bởi nhiều đỉnh khác).
- **Out-degree Centrality:** Đo lường mức độ ảnh hưởng của đỉnh (kết nối đến nhiều đỉnh khác).

Chỉ số tốt:

- Một chỉ số $C_{in}(v)$ cao là tốt trong các hệ thống như mạng xã hội (ví dụ: tài khoản có nhiều người theo dõi).
- Một chỉ số $C_{out}(v)$ cao phù hợp trong các hệ thống yêu cầu phát tán thông tin hiệu quả (ví dụ: trạm phát tin tức).

2.2 Betweenness Centrality

Betweenness Centrality (Tính trung tâm trung gian) đo lường mức độ mà một đỉnh nằm trên đường đi ngắn nhất giữa các cặp đỉnh khác trong mạng. Nó cho biết vai trò của đỉnh như một "cầu nối" để truyền thông tin hoặc kết nối các nhóm trong mạng.

Công thức:

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

trong đó:

- s, t : Các đỉnh khác nhau trong mạng ($s \neq t, s \neq v, t \neq v$).
- σ_{st} : Tổng số đường đi ngắn nhất giữa s và t .
- $\sigma_{st}(v)$: Số đường đi ngắn nhất giữa s và t đi qua đỉnh v .

Phạm vi giá trị:

- $0 \leq C_B(v) \leq 1$, khi được chuẩn hóa.
- Giá trị $C_B(v) = 0$: Đỉnh không nằm trên bất kỳ đường đi ngắn nhất nào.

- Giá trị $C_B(v) = 1$: Đỉnh nằm trên tất cả các đường đi ngắn nhất giữa mọi cặp đỉnh.

Ý nghĩa:

- Betweenness Centrality cao cho thấy đỉnh đóng vai trò quan trọng trong việc kết nối các phần của mạng.
- Đỉnh có $C_B(v)$ cao thường là "nút cổ chai" trong mạng, có thể kiểm soát hoặc làm chậm sự lan truyền thông tin.

Chỉ số tốt:

- Trong mạng xã hội: Một đỉnh có Betweenness Centrality cao có thể đại diện cho một cá nhân có khả năng kết nối các nhóm khác nhau.
- Trong mạng giao thông hoặc mạng truyền thông: Đỉnh có $C_B(v)$ cao là các trạm hoặc nút quan trọng cần tối ưu hóa để tránh tắc nghẽn.

Ứng dụng:

- Phân tích mạng xã hội để xác định người có ảnh hưởng lớn.
- Tối ưu hóa hệ thống giao thông hoặc mạng viễn thông.
- Phát hiện các nút quan trọng trong mạng để bảo vệ hoặc kiểm soát.

2.3 Closeness Centrality

Closeness Centrality (Tính trung tâm gần) đo lường mức độ gần gũi của một đỉnh với tất cả các đỉnh khác trong mạng. Chỉ số này được tính dựa trên tổng khoảng cách ngắn nhất từ đỉnh đó đến tất cả các đỉnh khác. Một đỉnh có Closeness Centrality cao thường dễ dàng tiếp cận các đỉnh khác trong mạng.

Công thức:

$$C_C(v) = \frac{N - 1}{\sum_{u \neq v} \text{dist}(v, u)}$$

trong đó:

- N : Tổng số đỉnh trong mạng.

- $\text{dist}(v, u)$: Khoảng cách ngắn nhất từ đỉnh v đến đỉnh u .

Phạm vi giá trị:

- $0 < C_C(v) \leq 1$
- Giá trị cao hơn cho thấy đỉnh v gần hơn với các đỉnh khác, tức là có khả năng truyền thông tin nhanh chóng hơn trong mạng.

Ý nghĩa:

- Closeness Centrality cao cho thấy đỉnh có vị trí chiến lược để tiếp cận hoặc ảnh hưởng đến các phần khác của mạng.
- Closeness Centrality thấp cho thấy đỉnh nằm xa các đỉnh khác, có thể gặp khó khăn trong việc truyền hoặc nhận thông tin.

Chỉ số tốt:

- Trong mạng xã hội: Đỉnh có Closeness Centrality cao có thể đại diện cho một người có khả năng giao tiếp nhanh với tất cả các thành viên khác.
- Trong mạng giao thông: Đỉnh có chỉ số cao đại diện cho vị trí trung tâm hoặc nút giao thông chính yếu.

Ứng dụng:

- Xác định các vị trí quan trọng trong mạng lưới giao thông, truyền thông.
- Phân tích mạng xã hội để tìm các cá nhân có khả năng truyền thông tin nhanh chóng.
- Tối ưu hóa việc truyền tải thông tin hoặc hàng hóa trong mạng.

2.4 Eigenvector Centrality

Eigenvector Centrality (Tính trung tâm véc-tơ riêng) đo lường tầm quan trọng của một đỉnh trong mạng dựa trên mức độ kết nối của đỉnh đó với các đỉnh quan trọng khác. Không chỉ tính số lượng kết nối, Eigenvector Centrality còn xem xét chất lượng của các đỉnh được kết nối.

Công thức:

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

trong đó:

- $C_E(v)$: Eigenvector Centrality của đỉnh v .
- A_{vu} : Phần tử của ma trận kề A , biểu thị sự kết nối giữa đỉnh v và u .
- λ : Giá trị riêng lớn nhất (eigenvalue) của ma trận kề A .
- V : Tập hợp tất cả các đỉnh trong mạng.

Phạm vi giá trị:

- Giá trị của $C_E(v)$ lớn hơn cho thấy đỉnh v có kết nối đến các đỉnh quan trọng khác trong mạng.
- Eigenvector Centrality được chuẩn hóa để nằm trong khoảng $0 \leq C_E(v) \leq 1$.

Ý nghĩa:

- Một đỉnh có Eigenvector Centrality cao không chỉ có nhiều kết nối mà còn được kết nối với các đỉnh quan trọng khác.
- Chỉ số này đặc biệt hữu ích trong các mạng nơi sự ảnh hưởng hoặc tầm quan trọng của một đỉnh phụ thuộc vào kết nối của nó với các đỉnh quan trọng khác (ví dụ: mạng xã hội hoặc mạng truyền thông).

Chỉ số tốt:

- Trong mạng xã hội: Đỉnh có Eigenvector Centrality cao có thể đại diện cho một cá nhân có nhiều kết nối với những người có tầm ảnh hưởng lớn.
- Trong mạng kinh tế: Đỉnh có chỉ số cao có thể đại diện cho một công ty có kết nối mạnh với các công ty quan trọng khác.

Ứng dụng:

- Xác định các nút quan trọng trong mạng xã hội, chẳng hạn như những người có tầm ảnh hưởng lớn.
- Phân tích mạng lưới để tìm các trung tâm truyền thông hoặc nút có sức ảnh hưởng cao.
- Được sử dụng trong thuật toán PageRank để xếp hạng các trang web dựa trên tầm quan trọng của chúng.

2.5 PageRank

PageRank là một thuật toán đo lường tầm quan trọng của các đỉnh trong một đồ thị có hướng, được sử dụng rộng rãi để xếp hạng các trang web trong công cụ tìm kiếm. Chỉ số này dựa trên ý tưởng rằng một đỉnh trở nên quan trọng nếu nó được nhiều đỉnh quan trọng khác trỏ đến.

Công thức:

$$PR(v) = \frac{1-d}{N} + d \sum_{u \in \text{In}(v)} \frac{PR(u)}{\text{OutDegree}(u)}$$

trong đó:

- $PR(v)$: PageRank của đỉnh v .
- N : Số lượng đỉnh trong mạng.
- d : Hệ số giảm dần (damping factor), thường được chọn là 0.85.
- $\text{In}(v)$: Tập hợp các đỉnh trỏ đến v .
- $\text{OutDegree}(u)$: Số lượng liên kết đi ra từ đỉnh u .

Ý nghĩa:

- PageRank cao cho thấy một đỉnh có nhiều liên kết từ các đỉnh khác, đặc biệt là từ các đỉnh có PageRank cao.
- Một đỉnh có PageRank thấp có thể không được trỏ đến nhiều hoặc chỉ được trỏ đến bởi các đỉnh có tầm quan trọng thấp.

Cách tính:

1. Khởi tạo PageRank ban đầu cho tất cả các đỉnh, thường bằng $1/N$.
2. Sử dụng công thức trên để tính giá trị PageRank lặp đi lặp lại cho đến khi hội tụ (sai số giữa các lần tính nhỏ hơn ngưỡng đặt trước).

Phạm vi giá trị:

- $0 \leq PR(v) \leq 1$, với tổng PageRank của tất cả các đỉnh bằng 1 (khi được chuẩn hóa).

Chỉ số tốt:

- Đỉnh có PageRank cao:
 - Là các trang web hoặc nút được liên kết bởi nhiều đỉnh khác, đặc biệt là từ các đỉnh có PageRank cao.
 - Thích hợp để xác định các trang web quan trọng hoặc có sức ảnh hưởng lớn trong mạng Internet.
- Đỉnh có PageRank thấp:
 - Là các đỉnh ít được liên kết đến, thường là những nút ít quan trọng hoặc ngoại vi trong mạng.
- Cân nhắc hệ số giảm dần (d):
 - Hệ số giảm dần d giúp mô phỏng hành vi người dùng khi duyệt web, với giá trị phổ biến là 0.85. Việc điều chỉnh d có thể ảnh hưởng đến cách phân phối điểm PageRank trong mạng.

Ứng dụng:

- **Công cụ tìm kiếm:** PageRank được sử dụng trong Google để xếp hạng các trang web.
- **Phân tích mạng xã hội:** Xác định các nút có tầm ảnh hưởng lớn.
- **Đánh giá tài liệu tham khảo:** Tìm các bài báo hoặc tài liệu được trích dẫn nhiều trong nghiên cứu.
- **Hệ thống khuyến nghị:** Xếp hạng các mục hoặc sản phẩm dựa trên mức độ phổ biến trong mạng lưới.

2.6 HITS (Hub and Authority)

HITS (Hyperlink-Induced Topic Search) là một thuật toán đo lường tầm quan trọng của các đỉnh trong đồ thị có hướng, bằng cách chia tầm quan trọng thành hai loại:

- **Hub:** Các đỉnh liên kết đến nhiều đỉnh quan trọng khác (Authority).
- **Authority:** Các đỉnh được liên kết bởi nhiều đỉnh quan trọng khác (Hub).

Công thức: Thuật toán HITS sử dụng giá trị lặp để tính toán điểm Hub và Authority:

- **Hub Score:**

$$h(v) = \sum_{u \in \text{Out}(v)} a(u)$$

- **Authority Score:**

$$a(v) = \sum_{u \in \text{In}(v)} h(u)$$

trong đó:

- $h(v)$: Điểm Hub của đỉnh v .
- $a(v)$: Điểm Authority của đỉnh v .
- $\text{Out}(v)$: Tập hợp các đỉnh được liên kết từ v (liên kết ra).
- $\text{In}(v)$: Tập hợp các đỉnh liên kết đến v (liên kết vào).

Quy trình tính toán:

1. Bắt đầu với điểm Hub và Authority ban đầu (thường được khởi tạo là 1 cho tất cả các đỉnh).
2. Tính toán điểm Authority cho mỗi đỉnh dựa trên điểm Hub của các đỉnh liên kết đến nó.
3. Tính toán điểm Hub cho mỗi đỉnh dựa trên điểm Authority của các đỉnh mà nó liên kết đến.

4. Chuẩn hóa điểm Hub và Authority để tránh bùng nổ giá trị.
5. Lặp lại quá trình cho đến khi các giá trị hội tụ.

Ý nghĩa:

- **Hub:** Đại diện cho các đỉnh cung cấp thông tin tốt, liên kết đến nhiều nguồn quan trọng.
- **Authority:** Đại diện cho các đỉnh chứa thông tin chất lượng cao, được nhiều nguồn liên kết đến.

Phạm vi giá trị:

- Điểm Hub và Authority được chuẩn hóa để tổng điểm trên toàn mạng là 1. **Chỉ số tốt:**
 - Một đỉnh có điểm Hub cao:
 - * Thích hợp cho các trang web hoặc nút có vai trò kết nối đến nhiều tài nguyên quan trọng (ví dụ: trang web hướng dẫn hoặc thư mục tham khảo).
 - Một đỉnh có điểm Authority cao:
 - * Phù hợp với các trang chứa thông tin có giá trị cao, được nhiều nguồn trỏ đến (ví dụ: bài báo học thuật hoặc trang chủ đề chính yếu).
 - Sự cân bằng giữa điểm Hub và Authority:
 - * Một đỉnh có thể vừa là Hub vừa là Authority nếu nó kết nối đến các đỉnh quan trọng và đồng thời được các đỉnh quan trọng liên kết lại.

Ứng dụng:

- Xếp hạng các trang web trong mạng Internet (đặc biệt trong các hệ thống truy vấn chuyên biệt).
- Phân tích mạng xã hội để tìm các cá nhân hoặc nhóm có ảnh hưởng.
- Tìm các tài liệu quan trọng trong một hệ thống tham khảo học thuật.

2.7 Eccentricity

Eccentricity (Độ lệch tâm) đo lường khoảng cách xa nhất từ một đỉnh cụ thể đến bất kỳ đỉnh nào khác trong mạng. Đây là một chỉ số quan trọng để hiểu sự phân tán của các đỉnh trong mạng.

Công thức:

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

trong đó:

- V : Tập hợp tất cả các đỉnh trong mạng.
- $\text{dist}(v, u)$: Khoảng cách ngắn nhất giữa đỉnh v và u .

Phạm vi giá trị:

- $0 \leq \text{Eccentricity}(v) < \infty$
- Giá trị thấp cho thấy đỉnh v nằm gần các đỉnh khác trong mạng.
- Giá trị cao cho thấy đỉnh v nằm xa hơn so với các đỉnh khác.

Ý nghĩa:

- Độ lệch tâm của một đỉnh phản ánh vị trí tương đối của đỉnh đó trong mạng.
- Trong mạng liên thông, Eccentricity của một đỉnh xác định khoảng cách xa nhất mà thông tin cần truyền từ đỉnh đó.

Chỉ số tốt:

- Đỉnh có Eccentricity thấp là trung tâm hơn và có thể dễ dàng truyền thông tin đến các đỉnh khác.
- Đỉnh có Eccentricity cao thường nằm ở rìa mạng và có thể gặp khó khăn trong việc giao tiếp.

Ứng dụng:

- Xác định các đỉnh trung tâm hoặc rìa trong mạng giao thông, mạng truyền thông.

- Phân tích cấu trúc mạng để tìm các nút quan trọng.
- Tối ưu hóa vị trí các nút trung gian trong các hệ thống phân phối.

Liên hệ với các chỉ số khác:

- Đỉnh có Eccentricity thấp nhất trong mạng là **trung tâm đồ thị** (Graph Center).
- Đỉnh có Eccentricity cao nhất trong mạng là **đỉnh rìa đồ thị** (Graph Periphery).