Bài thực hành số 1

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

* Ý nghĩa: Average degree, hay bậc trung bình, là một chỉ số quan trọng trong lý thuyết đồ thị, thể hiện mức độ kết nối trung bình của các đỉnh trong một đồ thị. Nó cho thấy một cái nhìn tổng quan về độ dày đặc của các cạnh, tức là trung bình một đỉnh sẽ kết nối với bao nhiêu đỉnh khác. Đồ thị có average degree cao thường có nhiều cạnh hơn so với số đỉnh, cho thấy các đỉnh có xu hướng kết nối với nhiều đỉnh khác và ngược lại.
* Công thức:

(avg\_deg) = (2 \* ) /

* là số lượng cạnh của đồ thị
* là số lượng đỉnh của đồ thị
* Phạm vi: Average degree là một số thực không âm và có thể có giá trị là số nguyên hoặc thập phân. Với đồ thị vô hướng, average degree nằm trong khoảng từ 0 (đồ thị rỗng) đến |V| - 1 (đồ thị đầy đủ). Trong khi đó, với đồ thị có hướng, average degree có thể vượt quá giới hạn |V| - 1 vì mỗi cạnh có hướng riêng biệt.
  1. Network Diameter
* Ý nghĩa: Network diameter, hay đường kính mạng, cho biết số bước đi tối đa cần thiết để đi từ một đỉnh bất kỳ đến một đỉnh bất kỳ khác trong đồ thị. Một đường kính mạng nhỏ thường cho thấy mạng kết nối tốt hơn và có hiệu quả cao hơn.
* Công thức:

diam(G) = max{d(u, v) | u, v V}

* d(u, v): Khoảng cách ngắn nhất giữa đỉnh *u* và đỉnh *v.*
* diam(G): Đường kính của đồ thị *G*
* V là tập hợp các đỉnh của đồ thị *G*.
* max là hàm tìm giá trị lớn nhất.
* Phạm vi: Network diameter, hay đường kính mạng, là một số nguyên không âm, thể hiện khoảng cách lớn nhất giữa bất kỳ cặp đỉnh nào trong một đồ thị. Giá trị này có ý nghĩa quan trọng trong việc đánh giá tính kết nối và hiệu quả của mạng. Phạm vi của network diameter phụ thuộc vào cấu trúc và tính liên thông của đồ thị. Phạm vi từ 1 (đồ thị đầy đủ) đến |V| - 1 (đồ thị đường thẳng) trong đồ thị liên thông, và có thể là vô cùng (∞) trong đồ thị không liên thông. Giá trị thực tế của đường kính mạng sẽ phụ thuộc vào cấu trúc cụ thể của từng đồ thị.
  1. Graph Density
* Ý nghĩa: Graph density, hay mật độ đồ thị, là một chỉ số đo lường mức độ “đặc” của một đồ thị. Nó cho biết tỷ lệ giữa số lượng cạnh thực tế trong đồ thị so với số lượng cạnh tối đa mà đồ thị đó có thể chứa. Một đồ thị có mật độ cao có nghĩa là các đỉnh kết nối với nhau nhiều, trong khi một đồ thị có mật độ thấp thì các đỉnh ít kết nối hơn. Graph density giúp chúng ta so sánh mức độ kết nối giữa các đồ thị khác nhau.
* Công thức:
* Với đồ thị vô hướng:

Density = (2 \* ) / ( \* (

Công thức này xuất phát từ việc số cạnh tối đa trong đồ thị vô hướng là |V| \* (|V| - 1) / 2, và chúng ta nhân 2 ở tử số để loại bỏ phép chia 2 ở mẫu số.

* Với đồ thị có hướng:

Density = |E| / (|V| \* (|V| - 1))

Trong trường hợp này, số cạnh tối đa có thể có là |V| \* (|V| - 1), vì mỗi cạnh có một hướng xác định.

* Phạm vi: Graph density là một số thực nằm trong khoảng từ 0 đến 1. Giá trị 0 tương ứng với một đồ thị rỗng (không có cạnh nào), và giá trị 1 tương ứng với một đồ thị đầy đủ (mọi đỉnh đều kết nối với mọi đỉnh khác). Giá trị density càng gần 1, đồ thị càng “đặc”, và ngược lại, càng gần 0, đồ thị càng “thưa”.
  1. Connected Components
* Ý nghĩa: Trong lý thuyết đồ thị, một connected component (thành phần liên thông) của một đồ thị vô hướng là một tập hợp các đỉnh mà trong đó có một đường đi giữa bất kỳ hai đỉnh nào trong tập đó. Nói cách khác, các đỉnh trong cùng một thành phần liên thông có thể “đến được” nhau thông qua các cạnh của đồ thị. Thành phần liên thông giúp chúng ta hiểu rõ cấu trúc kết nối của đồ thị, phân tách nó thành các “khối” độc lập.
* Phạm vi: Trong đồ thị có hướng, khái niệm “thành phần liên thông” có hai biến thể: weakly connected components (thành phần liên thông yếu), nơi các đỉnh liên thông nếu bỏ qua hướng cạnh, và strongly connected components (thành phần liên thông mạnh), nơi có một đường đi có hướng giữa mọi cặp đỉnh
  1. Average Path Length
* Ý nghĩa: Average Path Length (APL), hay độ dài đường đi trung bình, là một chỉ số quan trọng trong lý thuyết đồ thị, đo lường độ dài trung bình của đường đi ngắn nhất giữa tất cả các cặp đỉnh trong một đồ thị. Nó cho biết trung bình cần bao nhiêu bước để đi từ một đỉnh bất kỳ đến một đỉnh bất kỳ khác. APL càng nhỏ, các đỉnh càng “gần gũi” nhau, cho thấy tính kết nối tốt của mạng.
* Công thức:

APL = (1 / (n \* (n - 1))) \* ∑ ∑ d(u, v)

* n là số lượng đỉnh
* d(u, v) là khoảng cách ngắn nhất giữa đỉnh u và v
* Phạm vi: APL là một số thực không âm. Trong một đồ thị liên thông, APL có giá trị tối thiểu là 1 (khi đồ thị là đồ thị đầy đủ) và có thể xấp xỉ n/3 trong trường hợp đồ thị là đường thẳng. Với đồ thị không liên thông, APL không được xác định rõ vì không có đường đi giữa một số cặp đỉnh; trong trường hợp này, ta có thể tính APL của từng thành phần liên thông riêng lẻ.
  1. Average Clustering Coefficent
* Ý nghĩa: Average Clustering Coefficient (ACC), hay hệ số cụm trung bình, là một chỉ số quan trọng trong lý thuyết đồ thị, đặc biệt khi phân tích mạng xã hội và các mạng phức tạp khác. Nó đo lường mức độ mà các đỉnh trong đồ thị có xu hướng hình thành các “cụm” hay “tam giác”. ACC cho biết, trung bình, các láng giềng của một đỉnh có xu hướng kết nối với nhau như thế nào, hay nói cách khác, nó đo lường mật độ của các đỉnh trong mạng.
* Công thức:

ACC = (1 / |V|) \* ∑ C(v)

* |V| là số lượng đỉnh
* C(v) là Clustering Coefficient của đỉnh v
* Phạm vi: ACC là một số thực nằm trong khoảng từ 0 đến 1. Một đồ thị với ACC bằng 0 cho thấy các láng giềng của các đỉnh hầu như không kết nối với nhau, còn một đồ thị có ACC bằng 1 cho thấy tất cả các láng giềng của mỗi đỉnh đều kết nối với nhau (tạo thành các clique). ACC càng gần 1, đồ thị càng có tính “cụm” mạnh, cho thấy các đỉnh có xu hướng hình thành các nhóm chặt chẽ.

1. Độ đo tính trung tâm (Centrality Metrics)
   1. Degree Centrality (In-degree và Out-degree với đồ thị có hướng)

* Ý nghĩa: Degree Centrality là một thước đo đơn giản nhưng mạnh mẽ, dùng để xác định tầm quan trọng của một đỉnh trong mạng dựa trên số lượng kết nối trực tiếp mà nó có. Trong đồ thị vô hướng, Degree Centrality của một đỉnh đơn giản là số lượng cạnh mà đỉnh đó kết nối với các đỉnh khác. Nó phản ánh mức độ kết nối của một đỉnh trong mạng. Một đỉnh có nhiều kết nối hơn sẽ có Degree Centrality cao hơn, cho thấy nó có vai trò trung tâm hơn trong mạng.
* Công thức:

=

* *n*: là số đỉnh của đồ thị
* *deg*(*v*): tổng số liên kết trực tiếp đến đỉnh v (bậc của đỉnh)
* Phạm vi: Degree Centrality (và In-degree, Out-degree Centrality) là một số nguyên không âm. Giá trị tối thiểu là 0 (đỉnh cô lập), và giá trị tối đa là |V| - 1 (trong đó |V| là số lượng đỉnh) nếu không có self-loop. Các giá trị có thể được chuẩn hóa bằng cách chia cho |V| - 1 để có giá trị trong khoảng [0,1].
  1. Betweenness Centrality
* Ý nghĩa: Betweenness Centrality (BC) là một thước đo đánh giá tầm quan trọng của một đỉnh trong mạng dựa trên số lần nó xuất hiện trên các đường đi ngắn nhất giữa các cặp đỉnh khác trong mạng. Nói cách khác, nó đo lường mức độ mà một đỉnh nằm “giữa” các đỉnh khác, đóng vai trò là cầu nối hoặc trung gian trong luồng thông tin, tài nguyên, hoặc ảnh hưởng trong mạng.
* Công thức:

BC(v) = ∑ (σ(s, t|v) / σ(s, t))

* v là đỉnh mà chúng ta đang tính betweenness centrality.
* s và t là hai đỉnh khác nhau trong mạng, không bao gồm v
* σ(s, t): Tổng số đường đi ngắn nhất giữa đỉnh s và đỉnh t.
* σ(s, t|v): Số đường đi ngắn nhất giữa đỉnh s và đỉnh t mà đi qua đỉnh v.
* ∑ : Tổng của tất cả các cặp đỉnh s, t khác nhau.
* Phạm vi: Betweenness Centrality là một số thực không âm. Giá trị tối thiểu là 0 (đỉnh không nằm trên bất kỳ đường đi ngắn nhất nào giữa các cặp đỉnh khác). Giá trị tối đa phụ thuộc vào cấu trúc đồ thị, nhưng thường xảy ra ở các đỉnh nằm ở trung tâm của các mạng có cấu trúc “cây” hoặc “sao”. Giá trị BC có thể được chuẩn hóa bằng cách chia cho số cặp đỉnh có thể có để có giá trị trong khoảng [0, 1], giúp so sánh giữa các đồ thị khác nhau.
  1. Closeness Centrality
* Ý nghĩa: Closeness Centrality là một số thực không âm. Giá trị tối thiểu là 0 (đỉnh không thể kết nối đến tất cả các đỉnh khác hoặc đồ thị không liên thông). Giá trị tối đa thường xảy ra ở các đỉnh trung tâm có thể tiếp cận tất cả các đỉnh khác một cách nhanh chóng. CC không có ý nghĩa trong đồ thị không liên thông.
* Công thức:

CC(v) = (|V| - 1) / ∑ d(v, u)

trong đó d(v, u) là khoảng cách ngắn nhất giữa đỉnh v và đỉnh u, và tổng ∑ được tính trên tất cả các đỉnh u khác v. Chúng ta cũng có thể chuẩn hóa công thức thành CC(v) = (|V| - 1) / ∑ d(v, u) để có giá trị dễ so sánh hơn.

* Phạm vi: Closeness Centrality là một số thực không âm. Giá trị tối thiểu là 0 (đỉnh không thể kết nối đến tất cả các đỉnh khác hoặc đồ thị không liên thông). Giá trị tối đa thường xảy ra ở các đỉnh trung tâm có thể tiếp cận tất cả các đỉnh khác một cách nhanh chóng. CC không có ý nghĩa trong đồ thị không liên thông.
  1. Eigenvector Centrality
* Ý nghĩa: Eigenvector Centrality (EC) là một thước đo đánh giá tầm quan trọng của một đỉnh trong mạng dựa trên ý tưởng rằng một đỉnh quan trọng là đỉnh được kết nối với các đỉnh quan trọng khác. Thay vì chỉ đếm số lượng kết nối trực tiếp (như Degree Centrality), EC xem xét chất lượng của các kết nối, tức là tầm quan trọng của các đỉnh láng giềng. Một đỉnh có EC cao không chỉ có nhiều kết nối, mà còn có nhiều kết nối với các đỉnh có tầm ảnh hưởng lớn khác.
* Công thức:
* Phạm vi: Eigenvector centrality là một số thực không âm. Giá trị tối thiểu là 0, thường xảy ra ở các đỉnh không có kết nối với các đỉnh quan trọng khác. Giá trị tối đa phụ thuộc vào cấu trúc đồ thị và không có giới hạn cụ thể. EC chỉ có ý nghĩa trong các đồ thị liên thông hoặc trong các thành phần liên thông riêng lẻ.
  1. PageRank
* Ý nghĩa:
* Công thức:

PR(i) = (1-d) + d \* ∑ (PR(j) / L(j))

* PR(i): PageRank của trang web *i*.
* d: Dampening factor (thường là 0.85), biểu thị xác suất người dùng tiếp tục truy cập trang web thông qua một liên kết, thay vì truy cập ngẫu nhiên.
* j: Các trang web mà liên kết đến trang web *i*.
* PR(j): PageRank của trang web *j*.
* L(j): Số lượng liên kết ra khỏi trang web *j*.
* ∑: Tổng của tất cả các trang web *j* liên kết đến *i.*
* Phạm vi: Giá trị PageRank là một số thực không âm, giá trị tối thiểu có thể là 0, và giá trị tối đa phụ thuộc vào cấu trúc của mạng liên kết. PageRank có thể bị ảnh hưởng bởi các thủ thuật SEO, chẳng hạn như “trang trại liên kết”, và không phải là yếu tố duy nhất quyết định thứ hạng trong kết quả tìm kiếm.
  1. HITS (Hub and Authority)
* Ý nghĩa: Thuật toán Hubs and Authorities, thường được gọi là HITS, là một phương pháp phân tích mạng liên kết để xác định hai loại trang web quan trọng: “hubs” (trung tâm) và “authorities” (thẩm quyền). Hubs là các trang web đóng vai trò như danh mục hoặc bộ sưu tập liên kết đến nhiều trang web authority về một chủ đề cụ thể. Authorities là các trang web chứa thông tin chất lượng và được nhiều hub trỏ đến. HITS giúp chúng ta đánh giá không chỉ chất lượng nội dung mà còn vai trò của trang web trong mạng lưới liên kết.
* Công thức:

authority(i) = ∑ hub(j)

* j là các trang web liên kết đến trang web i.

hub(i) = ∑ authority(k)

* k là các trang web mà trang web i liên kết đến
* Phạm vi: Các giá trị hub và authority là các số thực không âm, thường được chuẩn hóa để có tổng bình phương bằng 1 hoặc một giá trị khác. Giá trị tối thiểu là 0, cho các trang không có liên kết hoặc không được trỏ đến bởi các trang khác. Phạm vi giá trị phụ thuộc vào cấu trúc của mạng liên kết.
  1. Eccentricity
* Ý nghĩa: Eccentricity (độ lệch tâm) của một đỉnh trong đồ thị là khoảng cách lớn nhất giữa đỉnh đó đến bất kỳ đỉnh nào khác trong đồ thị. Nói cách khác, nó đo lường khoảng cách xa nhất mà một đỉnh phải đi để đến được một đỉnh nào đó trong đồ thị. Eccentricity thể hiện mức độ “ngoại vi” hoặc “trung tâm” của một đỉnh trong mạng.
* Công thức:

eccentricity(v) = max {d(v, u)}

trong đó d(v, u) là khoảng cách ngắn nhất giữa đỉnh v và đỉnh u, và max là giá trị lớn nhất của khoảng cách

* Phạm vi: Eccentricity là một số nguyên không âm. Giá trị tối thiểu là 0, thường chỉ xảy ra trong đồ thị có một đỉnh. Không có giới hạn tối đa cụ thể cho eccentricity, giá trị này phụ thuộc vào kích thước và cấu trúc của đồ thị. Eccentricity có thể không được xác định trong đồ thị không liên thông, vì có thể có những đỉnh không thể đến được từ đỉnh đang xét.