

# Phần 1

## Degree Report

Id	Label	Interval	Degree ▾
136	136		100
60	60		96
132	132		75

Nút 136: Quan trọng nhất, là trung tâm của toàn bộ mạng, có ảnh hưởng đến nhiều nút khác. Degree Centrality của nút 136 là  $100/(198-1) = 0.5076$ . Nút 136 có giá trị Degree Centrality cao nhất (~50.76%), nghĩa là nó kết nối với khoảng 51% số nút còn lại trong mạng.

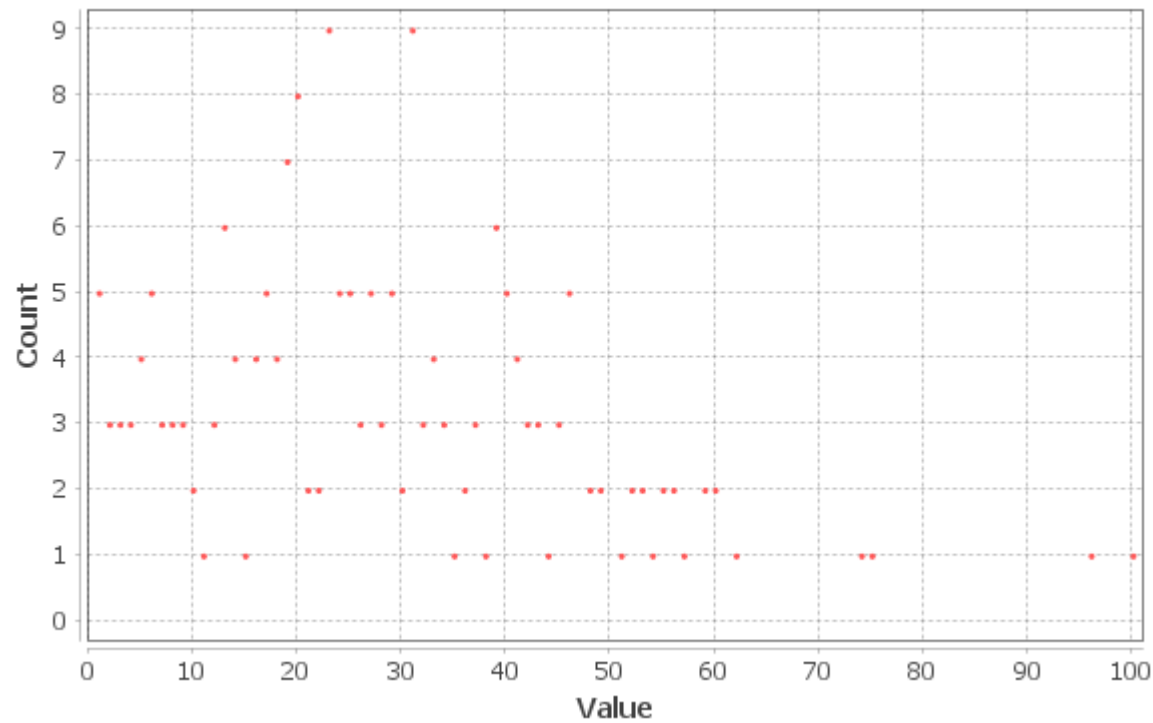
Nút 60: Đầu mối kết nối lớn thứ hai, với khả năng lan tỏa thông tin rộng nhưng thấp hơn nút 136. Degree Centrality của nút 60 là  $96/(198-1) = 0.4873$

Nút 132: Có vai trò kết nối trong phạm vi cục bộ, ảnh hưởng chủ yếu ở các nhóm nhỏ. Degree Centrality của nút 132 là  $75/(198-1) = 0.3807$

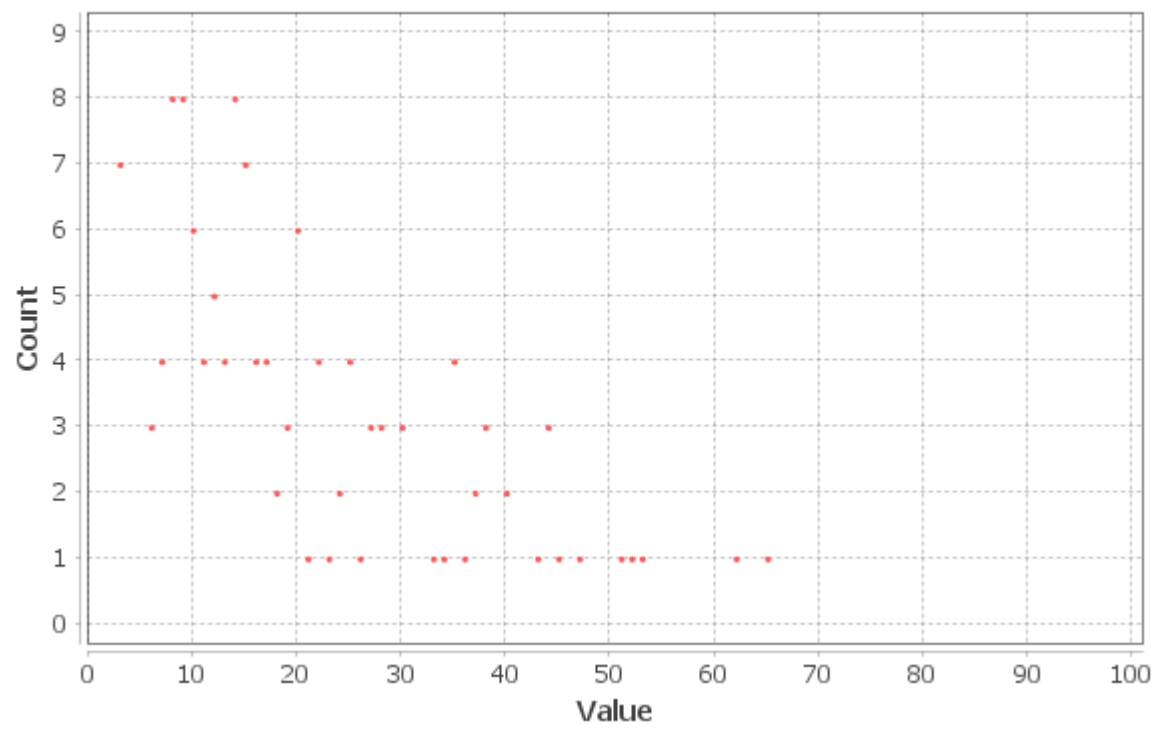
Nút 60 và 132 có giá trị thấp hơn, nhưng vẫn quan trọng với mức kết nối lần lượt là khoảng 49% và 38%.

Average Degree: 13.848 (Mỗi nút trong mạng trung bình có khoảng 13.848 kết nối trực tiếp với các nút khác.)

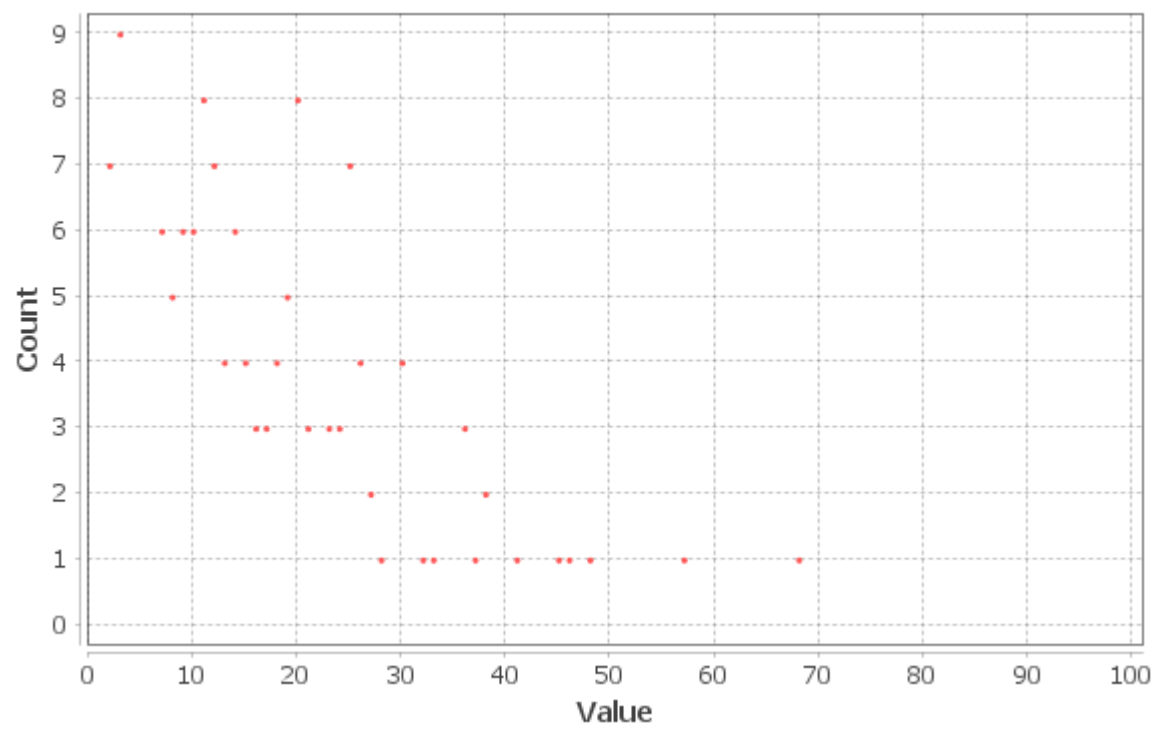
### Degree Distribution



### In-Degree Distribution



### Out-Degree Distribution



# Graph Distance Report

Id	Label	Closeness Centrality ▾
136	136	0.648026
60	60	0.58982
168	168	0.581121

Nút 136 có Closeness Centrality cao nhất (0.648026), nghĩa là nó là trung tâm về khả năng kết nối nhanh chóng đến toàn bộ mạng, đóng vai trò quan trọng trong việc truyền tải thông tin.

Nút 60 và nút 168 là các trung tâm phụ, vẫn có khả năng kết nối hiệu quả nhưng không mạnh bằng nút 136.

Id	Label	Betweenness Centrality ▾
136	136	2916.290104
153	153	1318.446696
60	60	1110.846351

ID: 136 (Betweenness Centrality = 2916.290104): Đây là nút có Betweenness Centrality cao nhất, tức là nó là "cầu nối" quan trọng nhất trong mạng, giúp liên kết các phần khác nhau. Nếu nút này bị loại bỏ, mạng có thể bị phân mảnh hoặc khả năng kết nối giữa các phần khác bị suy giảm nghiêm trọng.

ID: 153 (Betweenness Centrality = 1318.446696): Nút này là "cầu nối phụ", đóng vai trò kết nối một phần mạng với phần khác. Tuy không quan trọng bằng nút 136, nó vẫn đảm bảo một phần khả năng liên kết của mạng.

ID: 60 (Betweenness Centrality = 1110.846351): Đây là nút có Betweenness Centrality cao thứ ba, vẫn đóng vai trò kết nối nhưng mức độ ảnh hưởng thấp hơn so với nút 136 và 153. Là một nút trung gian phụ, có vai trò duy trì kết nối giữa một số thành phần trong mạng.

## Parameters:

Network Interpretation: undirected

## Results:

Diameter: 6

Trong mạng này, để đi từ hai nút xa nhất, cần di chuyển qua tối đa 6 bước.

Radius: 4

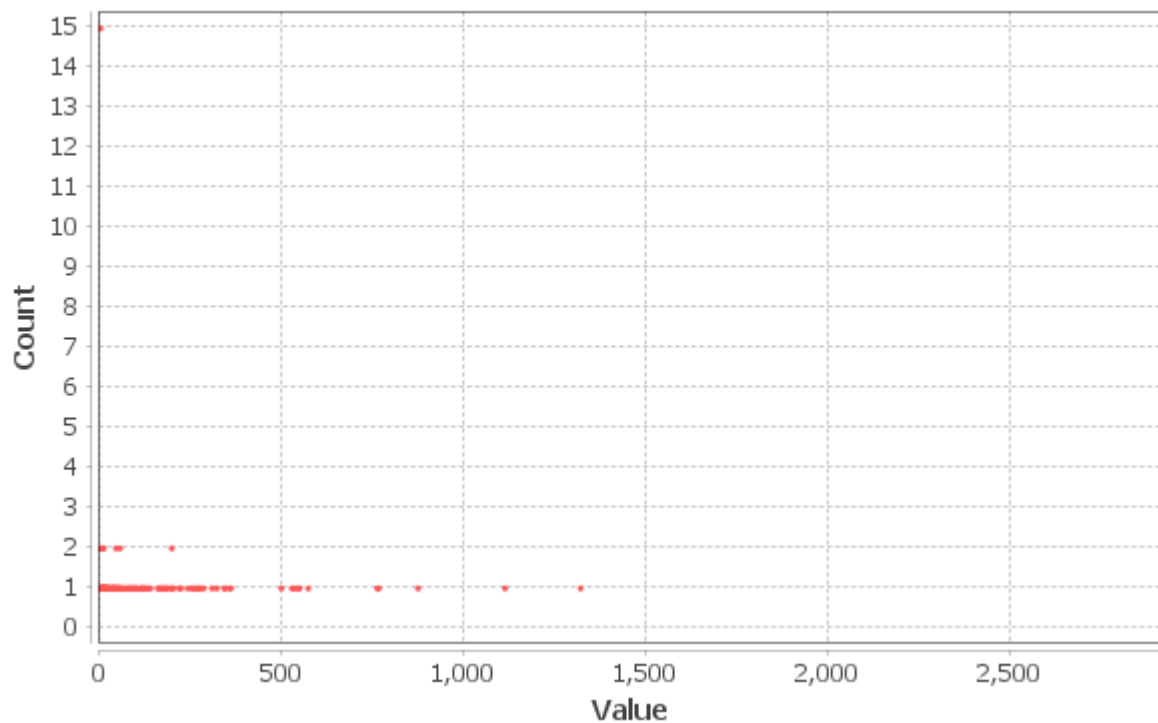
Nút trung tâm (hoặc trung tâm tiềm năng) của mạng này có khoảng cách tối đa là 4 bước đến bất kỳ nút nào khác.

Average Path length: 2.2350407629595446

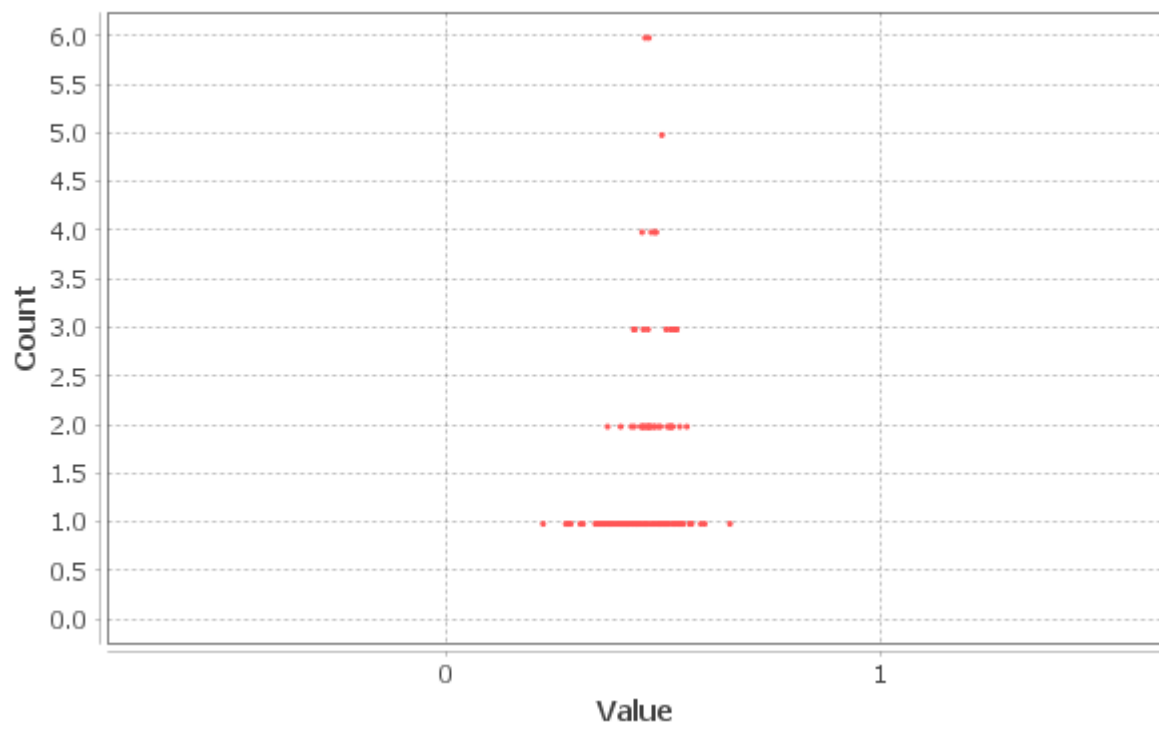
Đây là độ dài trung bình của tất cả các đường ngắn nhất giữa các cặp nút trong mạng.

Với giá trị trung bình khoảng 2.24, mạng này cho thấy rằng, trung bình chỉ cần hơn 2 bước là có thể di chuyển từ một nút bất kỳ đến một nút khác.

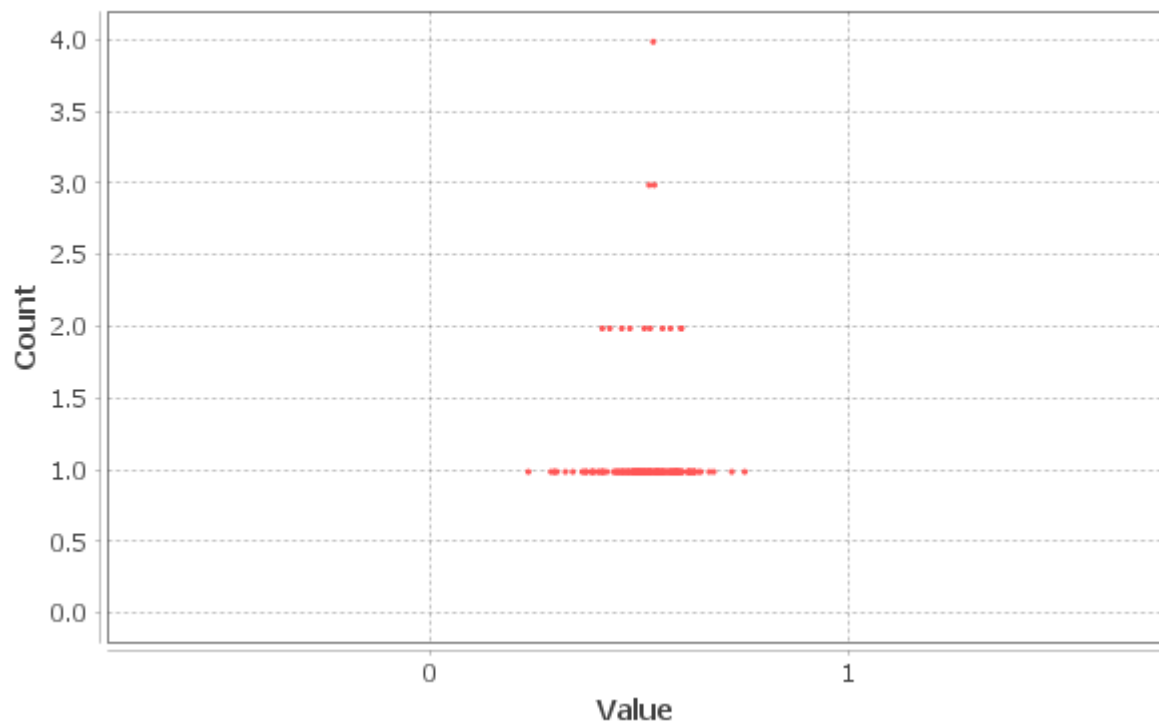
### Betweenness Centrality Distribution

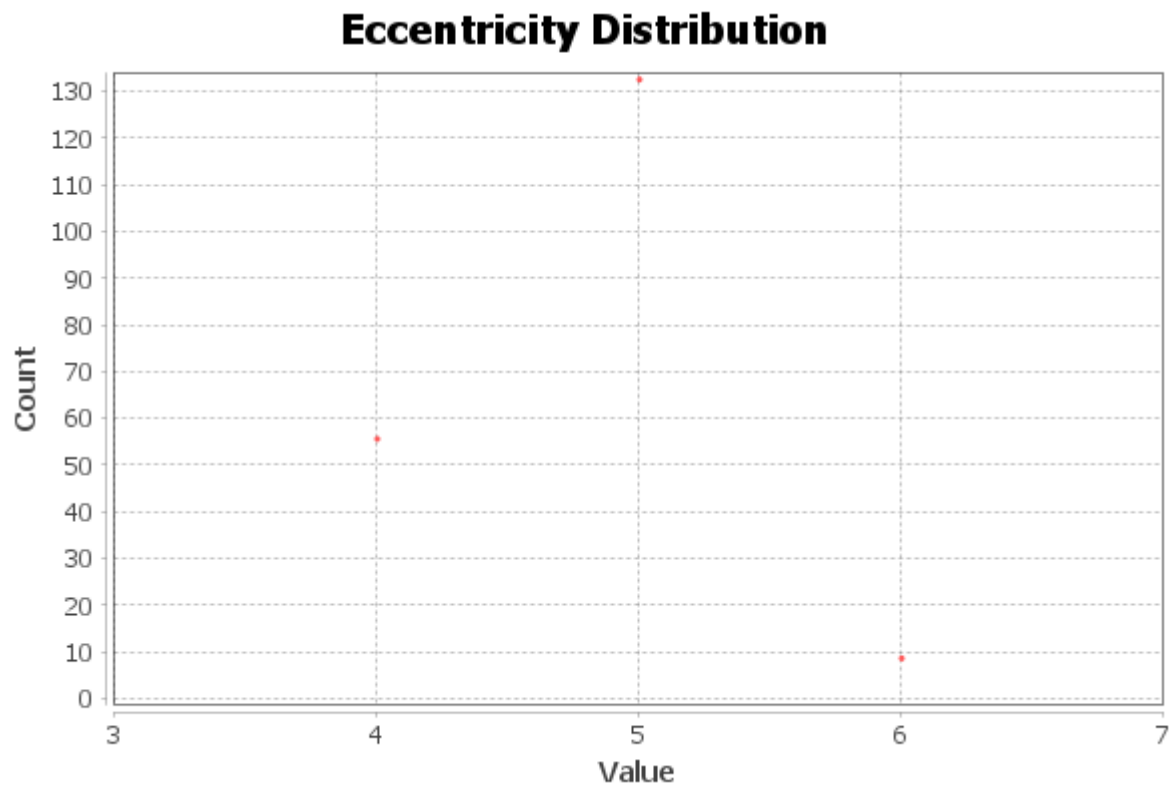


### Closeness Centrality Distribution



### Harmonic Closeness Centrality Distribution





## Phần 2

### Thuật toán Louvain

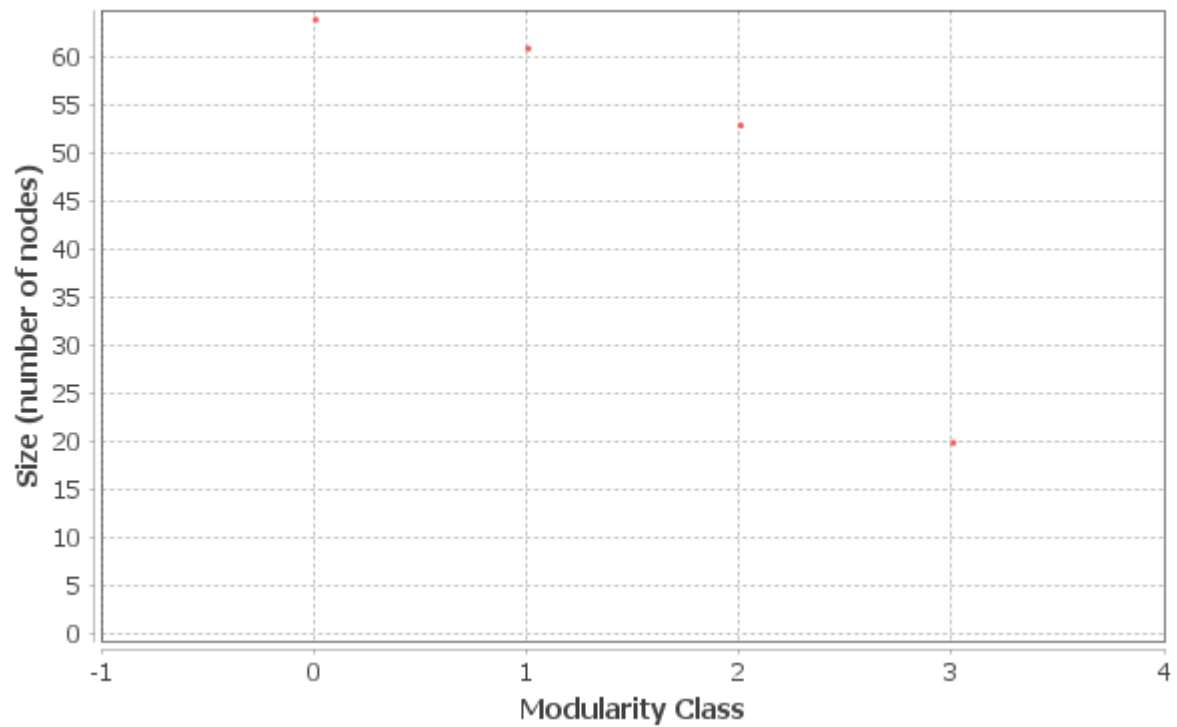
#### Results:

Modularity: 0.445

Modularity with resolution: 0.445

Number of Communities: 4

## Size Distribution



## Thuật toán Girvan-newman

### Processed Graph Data

Nodes: 198

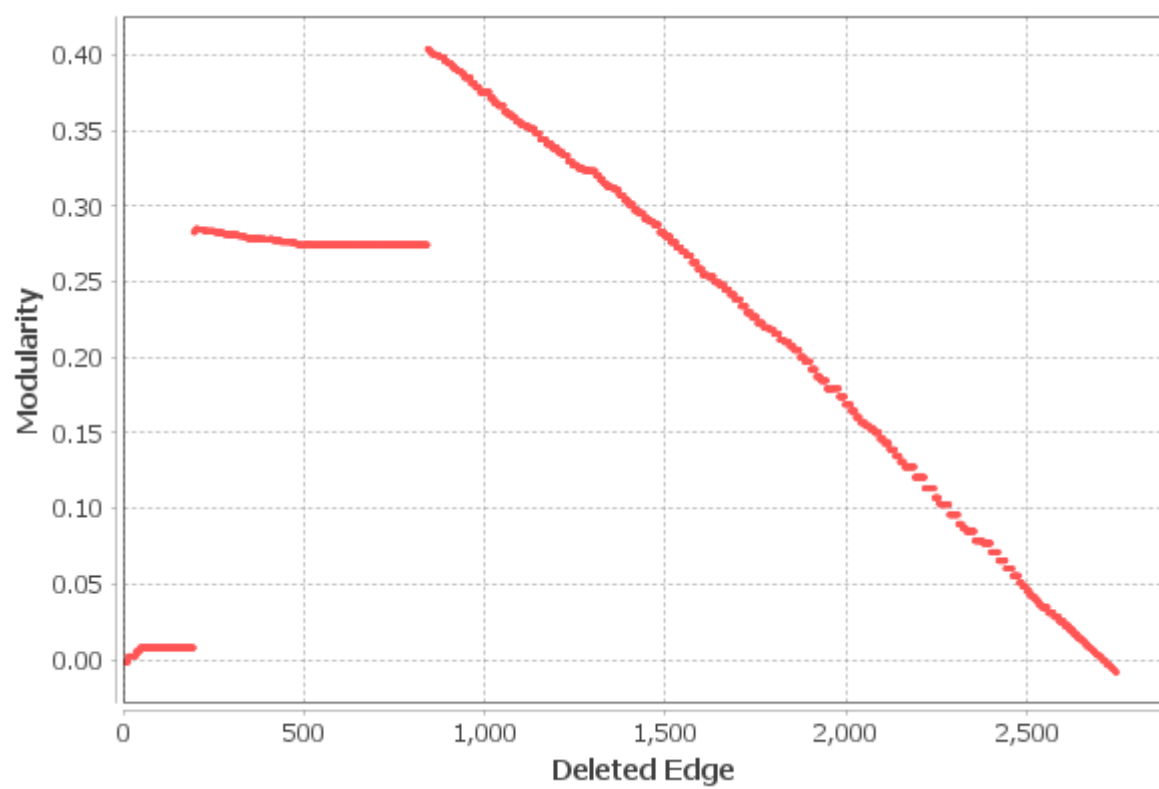
Edges 2742

Processing time: 8.076 sec.

### Communities

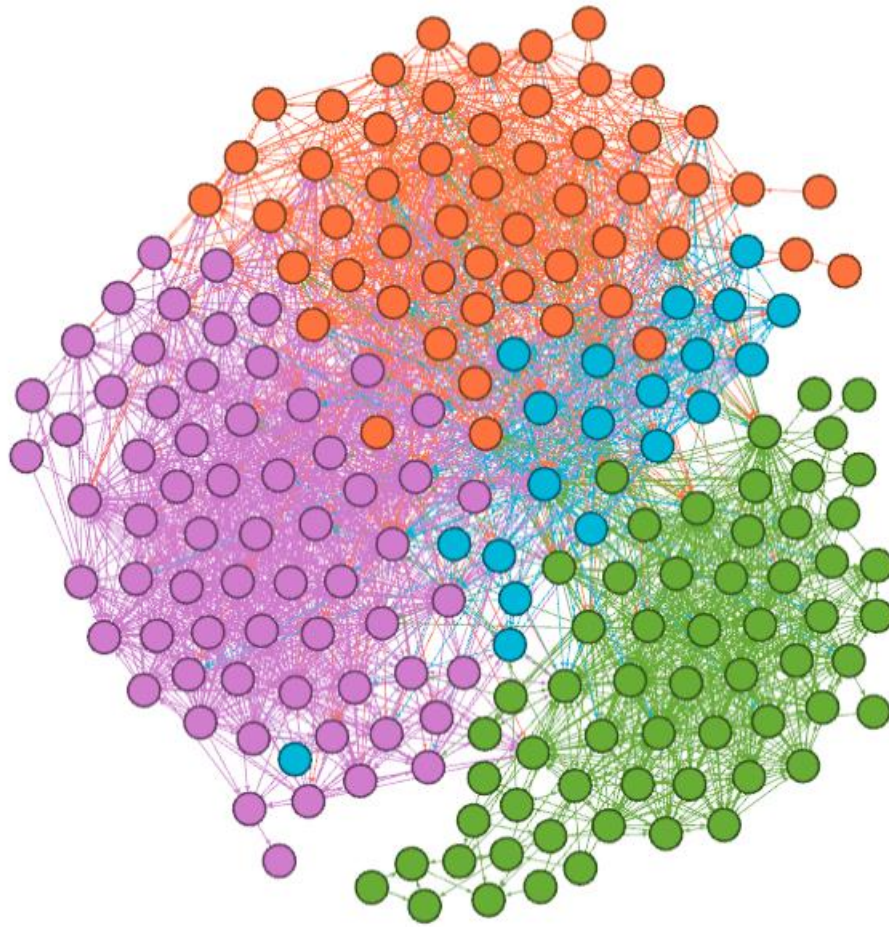
Number of communities: 39

Maximum found modularity: 0.40509886



## Phần 3





## Phần 4

1.

Thuật toán Louvain:

Modularity: 0.445

Số lượng cộng đồng: 4

Ưu điểm:

- Hiệu suất cao: Louvain là một thuật toán phân cụm rất nhanh và hiệu quả, đặc biệt khi áp dụng cho các mạng lớn. Thuật toán này có thể xử lý mạng lưới có số lượng nút và cạnh lớn mà không tốn quá nhiều thời gian.
- Tự động điều chỉnh phân cụm: Louvain tối ưu hóa modularity và tự động phân cụm các nút vào cộng đồng sao cho tăng giá trị modularity tối đa.
- Khả năng phân tích cộng đồng rõ ràng: Vì Louvain chia mạng thành số lượng cộng đồng ít (trong trường hợp của bạn là 4 cộng đồng), nên kết quả phân cụm dễ dàng phân tích và mô tả.

Nhược điểm:

- Số lượng cộng đồng ít: Mặc dù giá trị modularity của Louvain khá cao (0.445), nhưng số lượng cộng đồng rất ít (chỉ 4 cộng đồng). Điều này có thể không phản ánh đầy đủ sự phân bố của các mối quan hệ trong mạng lưới.
- Chưa tối ưu với các mạng phức tạp: Đối với các mạng lưới rất phức tạp hoặc có nhiều sự kết nối chéo, Louvain có thể không phân cụm chính xác tất cả các mối quan hệ.

Thuật toán Girvan-Newman:

Modularity: 0.405

Số lượng cộng đồng: 39

Ưu điểm:

- Phát hiện cộng đồng chi tiết: Thuật toán Girvan-Newman tạo ra số lượng cộng đồng nhiều hơn (39 cộng đồng trong trường hợp này), giúp phân tích mạng ở cấp độ chi tiết hơn. Điều này rất hữu ích khi bạn muốn có cái nhìn sâu sắc hơn về các nhóm trong mạng.
- Dễ dàng hiểu và giải thích: Vì thuật toán này sử dụng phương pháp loại bỏ các cạnh có trọng số cao (edge betweenness), nên bạn có thể dễ dàng giải thích lý do tại sao các nút được phân thành các cộng đồng nhất định.

Nhược điểm:

- Hiệu suất chậm: Thuật toán Girvan-Newman có thể rất tốn thời gian và tài nguyên, đặc biệt với mạng lớn như của bạn (198 nút và 2742 cạnh). Với mạng lưới lớn hơn, thuật toán này có thể trở nên không thực tế do tính toán phức tạp.
- Phân cụm không chính xác cho mạng lớn: Mặc dù số lượng cộng đồng lớn (39 cộng đồng), nhưng modularity (0.405) thấp hơn Louvain, cho thấy các cộng đồng không quá chặt chẽ hoặc liên kết.

3.

Đề xuất thuật toán Louvain là phương pháp phân cụm phù hợp nhất cho bộ dữ liệu này. Dưới đây là lý do chi tiết cho sự lựa chọn này:

Lý do chọn thuật toán Louvain:

- Hiệu suất cao và nhanh chóng: Bộ dữ liệu có 198 nút và 2742 cạnh, thuộc loại mạng có kích thước vừa phải. Thuật toán Louvain rất hiệu quả trong việc phân cụm mạng lớn với tốc độ xử lý nhanh và không yêu cầu quá nhiều tài nguyên tính toán.
- So với thuật toán Girvan-Newman, Louvain có thể xử lý mạng lớn nhanh chóng mà không làm giảm hiệu suất quá nhiều, trong khi Girvan-Newman tốn nhiều thời gian hơn, đặc biệt đối với mạng có số lượng nút và cạnh lớn.
- Phân cụm tự động và tối ưu hóa modularity: Thuật toán Louvain tối ưu hóa độ đo modularity, đảm bảo rằng các cộng đồng được phát hiện có sự phân chia mạnh mẽ, tức là các nút trong cùng một cộng đồng có kết nối mạnh với nhau và ít kết nối với các cộng đồng khác.
- Kết quả modularity của Louvain (0.445) có giá trị khá tốt, cho thấy phân cụm là hợp lý và có sự phân bố rõ ràng giữa các cộng đồng.
- Số lượng cộng đồng ít nhưng hợp lý:
- Louvain phát hiện ra 4 cộng đồng trong bộ dữ liệu của bạn. Mặc dù số lượng cộng đồng ít hơn so với Girvan-Newman, nhưng điều này có thể phù hợp nếu chỉ cần một

cái nhìn tổng quan về các nhóm trong mạng mà không cần phân tích chi tiết từng cộng đồng nhỏ.

- Việc có ít cộng đồng giúp dễ dàng phân tích và trực quan hóa kết quả, tránh việc quá tải thông tin khi có quá nhiều cộng đồng (như trong kết quả của Girvan-Newman với 39 cộng đồng).
- Khả năng mở rộng với mạng lớn hơn trong tương lai: