

# Lab05.02

Hung Nguyen

December 2024

Bộ dữ liệu được sử dụng email-Eu-core từ <https://snap.stanford.edu/>

## 1 Phần 1:

Betweenness Centrality

- Nút 160: Độ trung tâm trung gian là 72626
- Nút 86: Độ trung tâm trung gian là 37695
- Nút 5: Độ trung tâm trung gian là 27174

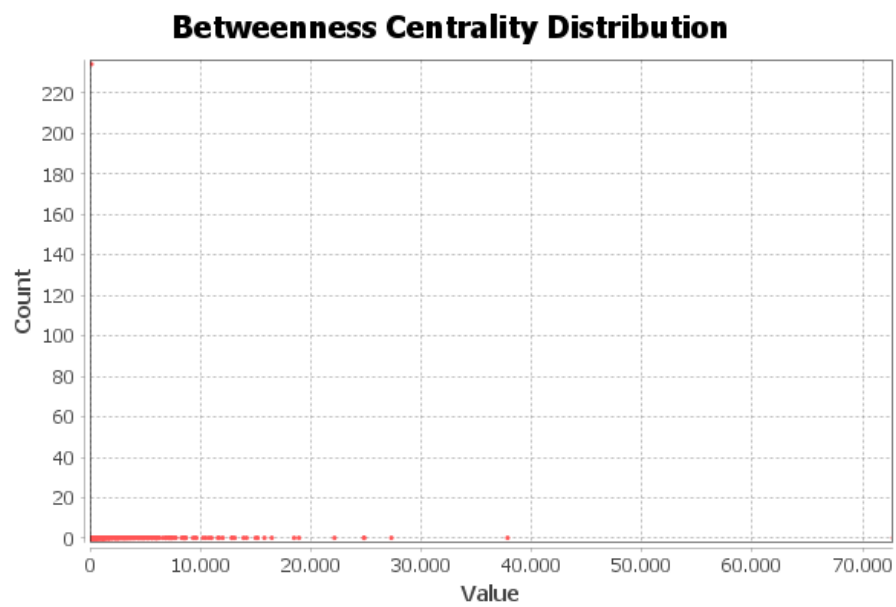
Nút 160, 86, và 5 có khả năng cao là các điểm trung gian quan trọng trong việc truyền thông tin hoặc tài nguyên trong mạng. Chúng có thể là các nút chiến lược để kiểm soát hoặc giám sát thông tin trên toàn bộ mạng, bởi vì các nút này chi phối các đường đi ngắn nhất và có thể ảnh hưởng đến việc truyền thông tin được hiệu quả hơn hoặc ít hiệu quả hơn.

Closeness Centrality

- Nút 846 và 995: Độ trung tâm gần gũi là 1.0
- Nút 160: Độ trung tâm gần gũi là 0.580723
- Nút 82: Độ trung tâm gần gũi là 0.542182

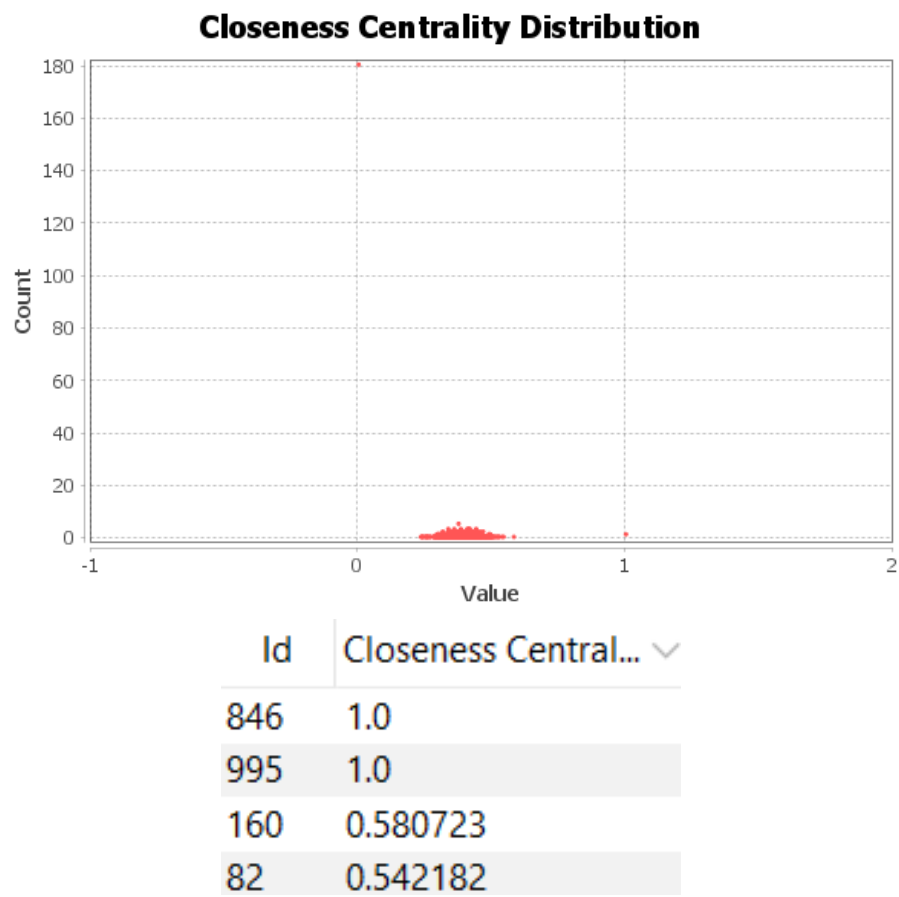
Nút 846 và 995 có giá trị cao nhất, cho thấy nó là nút có thể tiếp cận nhanh nhất đến các nút khác trong mạng, có thể là một trung tâm giao thông quan trọng hoặc một điểm phân phối trọng yếu.

Nút 160 và 82 cũng có giá trị cao, điều này làm cho chúng trở thành các điểm trung gian chủ chốt trong mạng, với khả năng kết nối tốt đến các phần khác của mạng.

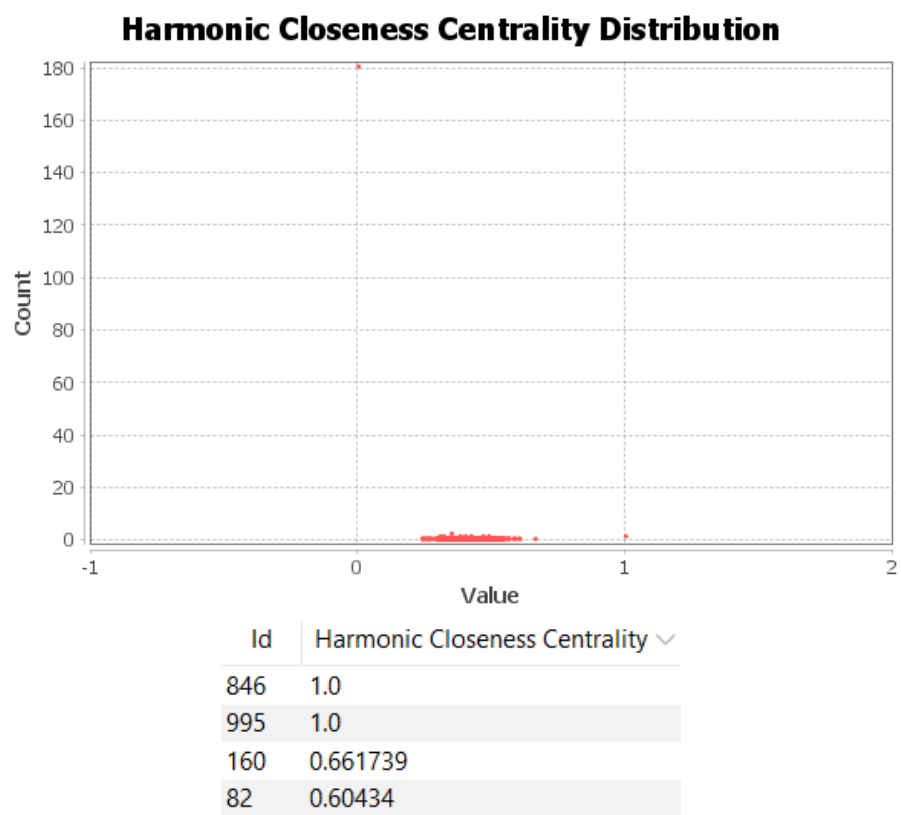


Id	Betweenness Ce... ▾
160	72626.497032
86	37695.391702
5	27174.021691
121	24704.121995

Hình 1: Bảng và kết quả của Betweenness Centrality



Hình 2: Bảng và kết quả của Closeness Centrality



Hình 3: Bảng và kết quả của Harmonic Closeness

#### Harmonic Closeness

- Nút 846 và 995: Độ trung tâm hài hòa là 1.0
- Nút 160: Độ trung tâm hài hòa là 0.661739
- Nút 82: Độ trung tâm hài hòa là 0.60434

Nút 160 và 82 có các giá trị độ trung tâm hài hòa khá cao, cho thấy chúng có mối liên hệ mạnh mẽ và thường xuyên với nhiều nút khác trong mạng, điều này làm tăng khả năng tương tác và truyền thông tin. Nút 846 và 995 có giá trị là 1.0, đây có thể là kết quả từ một đặc điểm cấu trúc mạng đặc biệt hoặc cách tính toán có sự nhầm lẫn, vì thông thường độ trung tâm hài hòa không đạt giá trị tối đa như vậy trừ khi nó nằm ở trung tâm của một mạng nhỏ hoặc rất đơn giản.

## 2 Phần 2.

#### Louvain

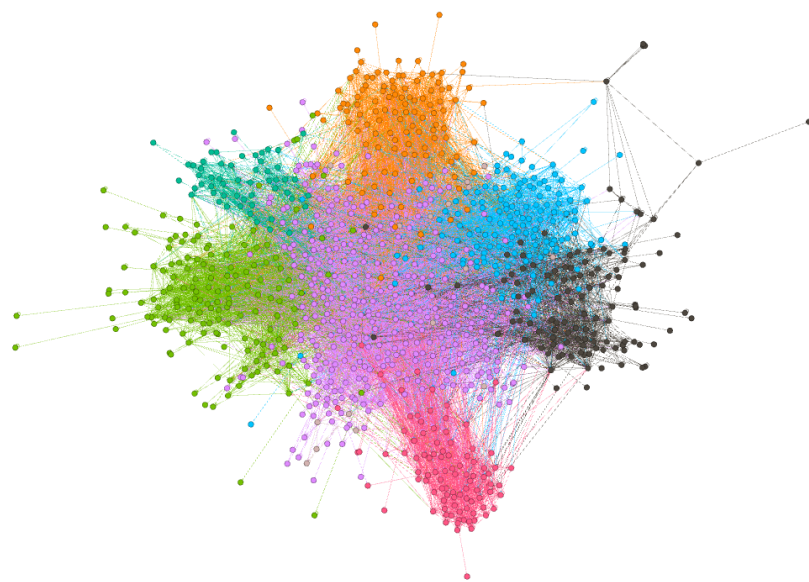
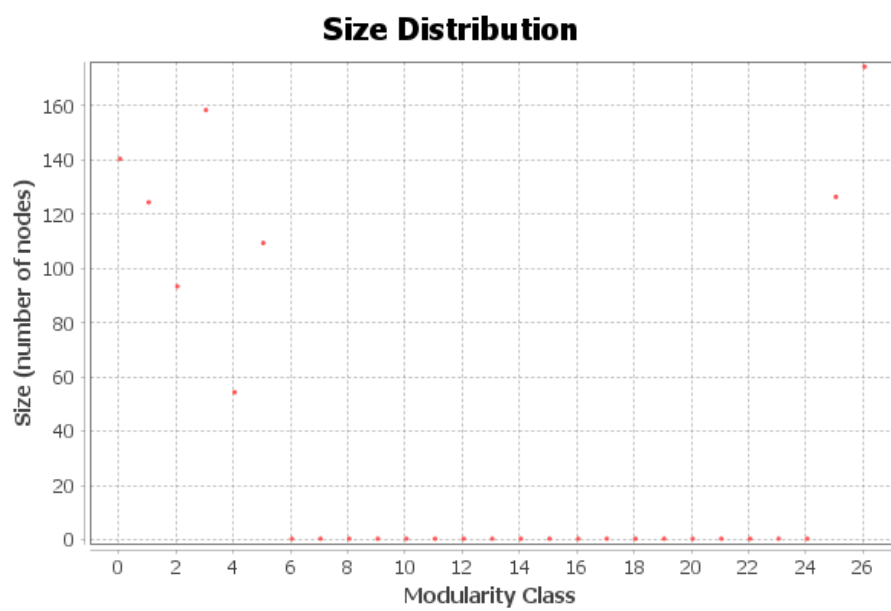
- Modularity: 0,414
- Modularity with resolution: 0,414
- Number of Communities: 27

#### Girvan-newman

- Maximum found modularity: 0.10949308
- Number of Communities: 731

## 3 Phần 4

Đối với Louvain, phương pháp này đã cho kết quả modularity khá cao (0,414), cho thấy có cấu trúc rõ ràng trong mạng nơi các nút được kết nối dày đặc bên trong các cộng đồng nhưng ít hơn giữa chúng. Số lượng cộng đồng được phát hiện là vừa phải, có thể chỉ ra một cái nhìn cân bằng về phân đoạn mạng. Còn khi xét trên Girvan-newman, cách tiếp cận này đã dẫn đến điểm modularity thấp hơn nhiều và số lượng cộng đồng cực kỳ cao lần lượt là 0.10949308 và 731. Điều này cho thấy phương pháp có thể đang phân mạng quá mức, dẫn đến nhiều nhóm nhỏ không hiệu quả trong việc bắt được các cấu trúc cộng đồng lớn hơn một cách hiệu quả.



Hình 4: Bảng kết quả của Louvain



Hình 5: Bảng kết quả của Girvan-newman

### Ưu và nhược điểm;

- Louvain:
  - **Ưu điểm:** Điểm modularity cao, thường liên quan đến các cấu trúc cộng đồng có ý nghĩa. Nó cũng thường nhanh hơn và có khả năng mở rộng tốt với các mạng lớn.
  - **Nhược điểm:** Vấn đề giới hạn độ phân giải, có thể bỏ sót các cộng đồng nhỏ hoặc ép các cộng đồng nhỏ vào các cộng đồng lớn hơn nếu không được xử lý đúng cách.
- Girvan-newman
  - **Ưu điểm:** Dựa trên betweenness của cạnh, có thể lý thuyết phát hiện các cấu trúc cộng đồng dựa trên dòng lưu thông mạng, có thể phát hiện các cấu trúc tinh tế.
  - **Nhược điểm:** Không hiệu quả về mặt tính toán với các mạng lớn và có khả năng phân mảnh cộng đồng quá mức như đã thấy.

**Giải Thích Trong Ngữ Cảnh** Các cộng đồng được phát hiện trong mạng lưới email có thể đại diện cho các nhóm người dùng chia sẻ mối quan hệ chuyên môn hoặc cá nhân mạnh mẽ. Trong một mạng lưới email, một cộng đồng có thể bao gồm những cá nhân làm việc trong cùng một bộ phận, dự án chung hoặc có liên hệ thường xuyên với nhau qua email để trao đổi thông tin liên quan đến công việc. Các mối liên kết email giữa các thành viên trong cùng một cộng đồng thường xuyên hơn và có xu hướng mang tính chất nội bộ, trong khi liên kết giữa các cộng đồng khác nhau có thể hiếm hơn và liên quan đến giao tiếp xuyên bộ phận hoặc xuyên tổ chức.

**Đề Xuất Phương Pháp** Xét đến bản chất của dữ liệu và kết quả: Louvain có vẻ phù hợp hơn do khả năng phát hiện cộng đồng có ý nghĩa một cách hiệu quả và hiệu quả. Điểm modularity cao cho thấy nó có thể nắm bắt cấu trúc của mạng một cách chính xác hơn mà không làm phân mảnh quá mức. Lý do vì Girvan-Newman có xu hướng phân mảnh quá mức và tốn kém tính toán, nó có thể không thực tế cho phân tích quy mô lớn hoặc thời gian thực nơi hiệu quả và rõ ràng của cấu trúc cộng đồng là cần thiết. Kết luận, lựa chọn thuật toán nên phù hợp với các yêu cầu cụ thể về rõ ràng, nguồn lực tính toán và quy mô dữ liệu mạng bạn đang xử lý. Đối với hầu hết các mục đích chung và đặc biệt trong các mạng xã hội, Louvain có lợi thế do sự cân bằng của nó về hiệu quả và hiệu quả trong việc phát hiện các cộng đồng đáng kể.