

Bài thực hành số 1: Tìm hiểu các chỉ số thông kê cơ bản sử dụng Gephi

Hãy cho biết ý nghĩa của các chỉ số bên dưới, công thức, phạm vi, như thế nào là chỉ số tốt?

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

- Average Degree
- Network Diameter
- Graph Density
- Connected Components
- Average Path Length
- Average Clustering Coefficient

2. Độ đo tính trung tâm (Centrality Metrics):

- Degree Centrality (In-degree và Out-degree với đồ thị có hướng)
- Betweenness Centrality
- Closeness Centrality
- Eigenvector Centrality
- PageRank
- HITS (Hub and Authority)
- Eccentricity

Yêu cầu nộp bài

- 1. Báo cáo (định dạng PDF) (tên file [LAB05/Lab05.01.pdf](#))**

Bài thực hành số 2: Phân tích cộng đồng trong mạng xã hội sử dụng Gephi

Mục tiêu

- Hiểu và thực hành các phương pháp phát hiện cộng đồng trong mạng xã hội
- Sử dụng thành thạo công cụ Gephi để phân tích và trực quan hóa mạng lưới
- So sánh kết quả từ các thuật toán phân cụm khác nhau

Dữ liệu

Sử dụng tập dữ liệu tự chọn - một mạng xã hội tự chọn

Yêu cầu thực hiện

Phần 1: Phân tích cấu trúc mạng Tính toán và so sánh các độ đo tính trung tâm (Centrality Measures) sau:

- Degree Centrality
- Betweenness Centrality
- Closeness Centrality Hãy xác định 3 nút có độ trung tâm cao nhất theo mỗi độ đo và giải thích ý nghĩa của chúng trong mạng lưới.

Phần 2: Phát hiện cộng đồng Thực hiện phân cụm mạng lưới sử dụng 3 thuật toán sau:

1. Thuật toán Louvain
2. Thuật toán Girvan-newman (gephi.org/plugins/#/plugin/girvan-newman-clustering hoặc gephi.org/plugins/#/plugin/newman-girvan-plugin)
3. Thuật toán LPA (gephi.org/plugins/#/plugin/label-propagation-clustering)

Với mỗi thuật toán, hãy:

- Ghi lại số lượng cộng đồng được phát hiện
- Tính toán độ đo Modularity của kết quả phân cụm
- Lưu ảnh kết quả phân cụm với các node được tô màu theo cộng đồng

Phần 3: Trực quan hóa Tạo một bản trực quan hóa đẹp và có ý nghĩa cho mạng lưới bằng cách:

1. Sử dụng thuật toán layout ForceAtlas2 với các tham số phù hợp

2. Điều chỉnh kích thước node theo độ đo trung tâm đã tính
3. Tô màu node theo kết quả phân cụm từ thuật toán cho kết quả tốt nhất
4. Thêm nhãn cho các node quan trọng (có độ trung tâm cao)

Phần 4: Báo cáo và đánh giá Viết báo cáo ngắn bao gồm:

1. So sánh kết quả của 3 thuật toán phân cụm, nêu ưu và nhược điểm của mỗi phương pháp
2. Giải thích ý nghĩa của các cộng đồng được phát hiện trong ngữ cảnh của mạng xã hội
3. Đề xuất phương pháp phân cụm phù hợp nhất cho loại dữ liệu này và lý do

Yêu cầu nộp bài

2. File Gephi project (Lab05/**Lab05.02.gephi**)
3. Báo cáo phân tích (bao gồm các ảnh trực quan hóa) (định dạng PDF) (tên file **LAB05/Lab05.02.pdf**)
4. Bảng số liệu thống kê (có thể đính kèm trong báo cáo hoặc file Excel riêng)