

Phân tích cộng đồng trong mạng xã hội sử dụng Gephi

Le Thai Duong

Ngày 30 tháng 12 năm 2024

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

Id	Label	Degree ▾
Tyrion	Tyrion	36
Sansa	Sansa	26
Jon	Jon	26

Hình 1: Ba nodes có Degree Centrality ($\text{Degree} / \text{tổng số nodes} - 1$) cao nhất

Id	Label	Betweenness Centrality ▾
Jon	Jon	1279.753353
Robert	Robert	1165.602517
Tyrion	Tyrion	1101.384972

Hình 2: Ba nodes có Betweenness Centrality cao nhất

Id	Label	Closeness Centrality ▼
Tyrion	Tyrion	0.512077
Sansa	Sansa	0.509615
Robert	Robert	0.5

Hình 3: Ba nodes có Closeness Centrality cao nhất

Dựa trên dữ liệu, bảng dưới đây hiển thị các node có giá trị cao nhất theo từng tiêu chí:

Centrality Type	Node	Value
Degree Centrality	Tyrion	33.64% (36/107)
Betweenness Centrality	Jon	1279.753353
Closeness Centrality	Tyrion	0.512077

Bảng 1: Các nodes có giá trị độ đo trung tâm cao nhất

Ý nghĩa của các độ đo

Degree Centrality: Tyrion: 33.64%

Node Tyrion là một "trung tâm" trong mạng lưới, có nhiều mối quan hệ trực tiếp với các nhân vật khác. Đóng vai trò quan trọng trong việc giao tiếp và tương tác với nhiều node trong mạng.

Betweenness Centrality: Jon: 1279.753353

Node Jon là một "người trung gian" quan trọng trong mạng lưới. Đóng vai trò như cầu nối, giúp thông tin hoặc ảnh hưởng lan truyền giữa các nhóm nhân vật khác nhau.

Closeness Centrality: Tyrion: 0.512077

Điều này cho thấy Node Tyrion có vị trí thuận lợi trong mạng lưới, cho phép anh ta tiếp cận với các node khác một cách nhanh chóng. Có thể ảnh hưởng đến toàn bộ mạng chỉ trong một số bước di chuyển thông tin.

Kết luận

Tyrion là nhân vật có nhiều kết nối trực tiếp (Degree) và có vị trí chiến lược trong mạng (Closeness), làm cho anh ta trở thành một trong những nhân vật có ảnh hưởng lớn nhất.

Jon, mặc dù không có nhiều kết nối trực tiếp, nhưng lại đóng vai trò trung gian quan trọng, giúp gắn kết các phần khác nhau của mạng.

2 Phát hiện cộng đồng

2.1 Thuật toán Louvain

Results:

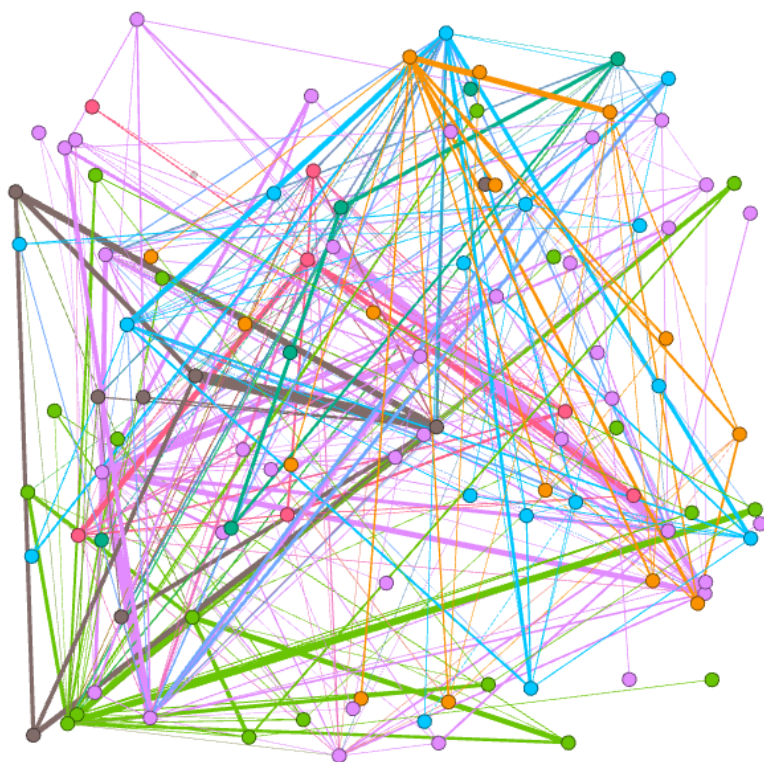
Modularity: 0.598

Modularity with resolution: 0.598

Number of Communities: 7

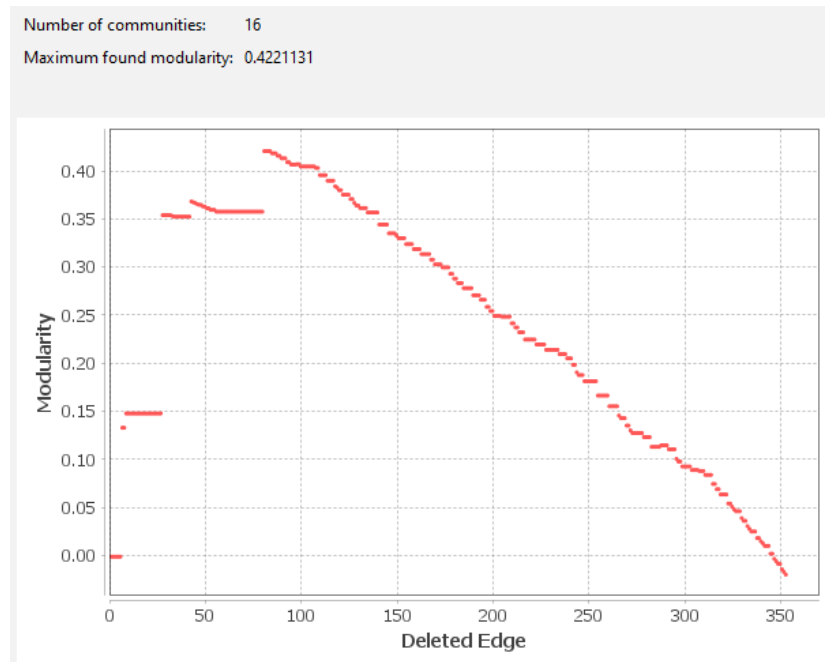
Hình 4: kết quả của thuật toán Louvain.

- Số lượng cộng đồng: 7
- Modularity: 0.598



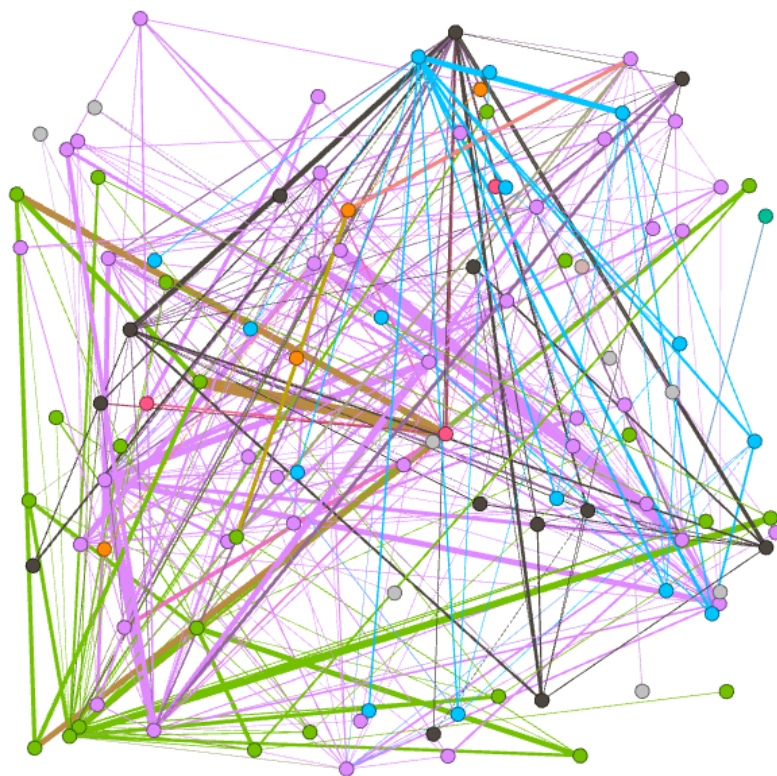
Hình 5: Phân cụm mạng lưới sử dụng Louvain.

2.2 Thuật toán Girvan-Newman



Hình 6: kết quả của thuật toán Girvan-Newman.

- Số lượng cộng đồng: 16
- Modularity: 0.422



Hình 7: Phân cụm mạng lưới sử dụng Girvan-Newman.

2.3 Thuật toán LPA

Results:

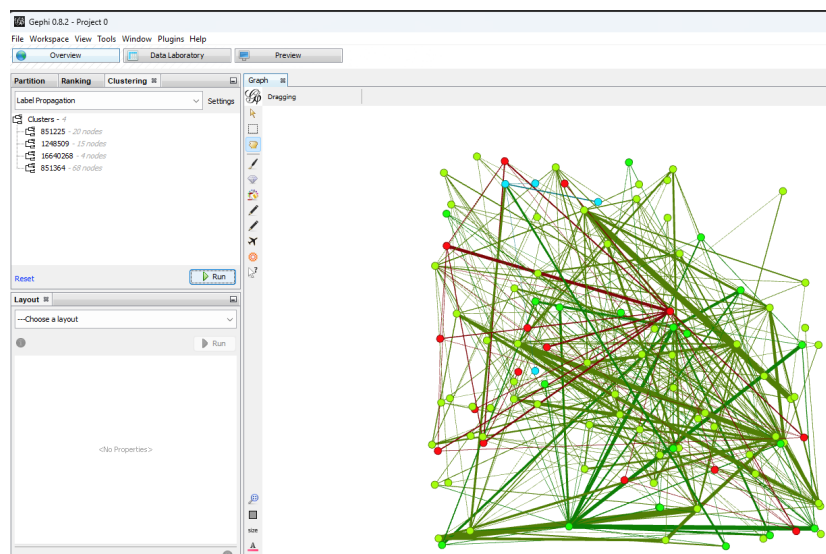
Modularity: 0.467

Modularity with resolution: 0.873

Number of Communities: 4

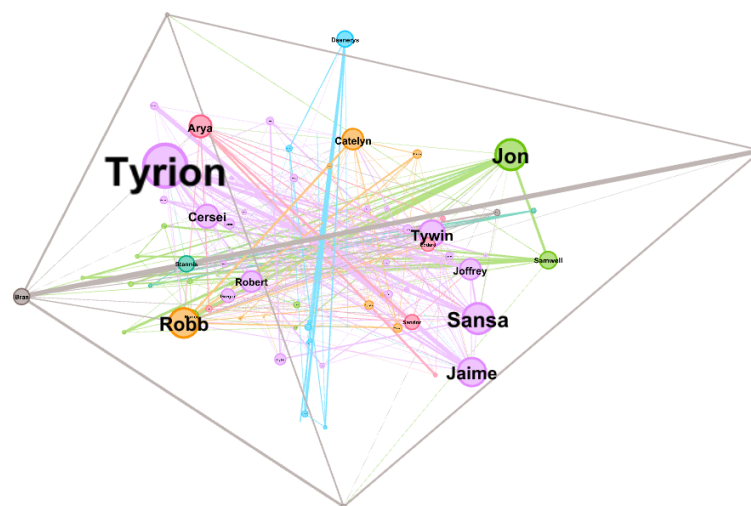
Hình 8: kết quả của thuật toán LPA trên gephi0.8.2beta.

- Số lượng cộng đồng: 4
- Modularity: 0.467



Hình 9: Phân cụm mạng lưới sử dụng LPA trên gephi0.8.2beta.

3 Trục quan hóa (chọn thuật toán phân cụm Louvain thì có Modularity cao nhất)



Hình 10: Trục quan bằng ForceAtlas2

4 Nhận xét

4.1 So sánh kết quả của hai thuật toán phân cụm

- Thuật toán Louvain:

- Ưu điểm:

- * Tối ưu modularity hiệu quả, đặc biệt với mạng lớn.
- * Kết quả phân cụm ổn định, dễ giải thích.
- * Thời gian tính toán nhanh, phù hợp với mạng lưới có nhiều node và cạnh.

- Nhược điểm:

- * Độ chính xác có thể giảm khi gặp mạng lưới nhỏ hoặc không đồng nhất.
- * Phụ thuộc nhiều vào cấu trúc mô-đun của mạng.

- **Thuật toán Girvan-Newman:**

- **Ưu điểm:**

- * Hiệu quả với mạng nhỏ, cho phép phát hiện các cụm nhỏ hoặc tiềm năng.
 - * Tập trung vào việc tìm các cầu nối quan trọng giữa các cụm (betweenness).

- **Nhược điểm:**

- * Tốn nhiều thời gian và tài nguyên tính toán khi mạng lớn.
 - * Modularity thấp hơn, thể hiện sự phân cụm kém rõ ràng hơn so với Louvain.

- **Thuật toán Label Propagation (LPA):**

- **Ưu điểm:**

- * Rất nhanh và hiệu quả, đặc biệt đối với mạng lớn do không yêu cầu tối ưu toàn cục.
 - * Không cần thông tin trước như số cụm hoặc trọng số cạnh.
 - * Dễ triển khai, phù hợp để phân cụm sơ bộ trong mạng phức tạp.

- **Nhược điểm:**

- * Kết quả không ổn định do tính ngẫu nhiên của thuật toán.
 - * Không phù hợp để phát hiện các cụm nhỏ hoặc cấu trúc cụm rõ ràng.

4.2 Ý nghĩa của các cộng đồng được phát hiện

- Trong ngữ cảnh mạng xã hội, các cộng đồng đại diện cho các nhóm người có sự tương tác chặt chẽ với nhau.
- Kết quả phân cụm cho thấy các node trong cùng một cộng đồng có nhiều kết nối nội bộ hơn, thể hiện các nhóm bạn bè, gia đình, hoặc đồng nghiệp.
- Ví dụ:

- Các cộng đồng nhỏ được phát hiện bởi Girvan-Newman có thể đại diện cho các nhóm tương tác cụ thể (như nhóm bạn thân).
- Các cộng đồng lớn được phát hiện bởi Louvain phản ánh mối quan hệ rộng rãi hơn, như các tổ chức hoặc cộng đồng chung sở thích.

4.3 Đề xuất phương pháp phân cụm phù hợp

- Phương pháp được đề xuất: Louvain.
- Lý do:
 - Louvain có modularity cao hơn (0.598 so với 0.422), cho thấy khả năng phân cụm hiệu quả hơn.
 - Thích hợp với mạng lớn, điển hình cho dữ liệu mạng xã hội hiện nay.
 - Thời gian tính toán nhanh hơn, phù hợp để xử lý lượng dữ liệu lớn.