

ANALYZING THE AUTHOR-READER RELATIONSHIP THROUGH THE PERSPECTIVE OF VNEXPRESS

Phân tích mối quan hệ tác giả - độc giả qua góc nhìn của Vnexpress

Nguyễn Văn Đạt - 2186400229
Lê Nguyên Anh Nhật - 2186400330
Ngô Thị Hồng Uyên - 2186400213

Đại học Công Nghệ Tp. HCM
GVHD: ThS. Lê Nhật Tùng



HUTECH
Đại học Công nghệ Tp.HCM

Nội Dung

- 1 Giới thiệu
- 2 Chuẩn bị dữ liệu
- 3 Tính toán số đo
- 4 Phân cụm đồ thị
- 5 Kết luận

Giới thiệu

Lý do chọn đề tài

Bối cảnh nghiên cứu:

- Nền tảng truyền thông trực tuyến đang trở thành không gian cho thảo luận và xây dựng cộng đồng.
- Góc nhìn (VnExpress) là ví dụ tiêu biểu cho sự tương tác giữa tác giả và độc giả.

Mục tiêu nghiên cứu:

- Hiểu cấu trúc và động lực của mạng lưới VnExpress.
- Phân tích các vai trò (tác giả, bình luận viên, người phản hồi).
- Nghiên cứu tính toán các số đo và cách cộng đồng hình thành.

Chuẩn bị dữ liệu

Thu thập và tiền xử lý dữ liệu

Thu thập dữ liệu:

- Dữ liệu được thu thập từ báo điện tử Vnexpress.
- Dữ liệu ban đầu gồm 3265 bài báo: Author ID, Discription, Category, Commenter, Replier.

Author ID	Category	Description	Comments
998	Kinh doanh & Quản trị	Chuyên gia kinh tế	['Nickname': 'RO là GOAT', 'Reply Nickname': [], 'Nickname': 'MU dẫn', 'Reply Nickname': [], 'Nickname': 'MU dẫn', 'Reply Nickname': []]
1955	Giáo dục & tri thức	Nhân viên xã hội và chuyên gia phát triển cộng đồng	['Nickname': 'nth fl', 'Reply Nickname': ['Angry Joker', 'Tu anh', 'Account solely for anti toxic anti-fans'], 'Nickname': 'Giao ha', 'Reply Nickname': ['Tu anh', 'v', 'nth fl']]

Table: Mô tả bảng dữ liệu

Thu thập và tiền xử lý dữ liệu

Tiền xử lý dữ liệu ban đầu: Loại bỏ các dòng dữ liệu bị thiếu, chuyển thành đồ thị có hướng, loại bỏ các **isolated nodes** (nút cô lập):

- **Node:** Author ID (tác giả), Nickname (người bình luận), Reply Nickname (người phản hồi).
 - **Edge:** Author ID \rightarrow Nickname ID (tác giả \rightarrow người bình luận) và Nickname \rightarrow Reply Nickname (người bình luận \rightarrow người phản hồi).
- \Rightarrow Kết quả: Đồ thị gồm 60.191 nodes và 130.776 edges.

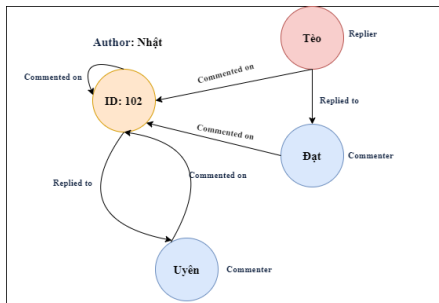


Figure: Đồ thị mẫu

Tích toán số đo

Average Degree và Graph Density

Average Degree:

- Giá trị: **4.59**
- Ý nghĩa:
 - Trung bình, mỗi nút trong mạng (bài viết hoặc bình luận) kết nối với khoảng 4.59 nút khác.
 - Thể hiện mức độ kết nối thấp, phản ánh tương tác yếu trong mạng lưới.

Graph Density:

- Giá trị: **0.000055**
- Ý nghĩa:
 - Mật độ rất thấp, thể hiện tỉ lệ kết nối thực tế so với tổng số kết nối có thể có là cực kỳ nhỏ.
 - Cấu trúc mạng thưa thớt, các nút không kết nối mạnh mẽ với nhau.

Degree Centrality

- **In-Degree Centrality:** Thể hiện khả năng thu hút bình luận và tương tác của các tác giả.
- **Out-Degree Centrality:** Đánh giá mức độ chủ động tương tác và trả lời của các bình luận viên.

Top 10 In-Degree Centrality

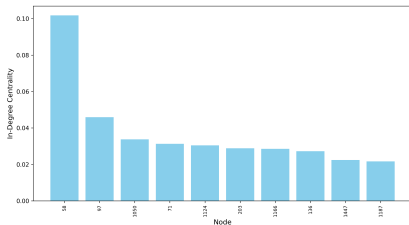


Figure: In-Degree Centrality Bar Chart

Top 10 Out-Degree Centrality

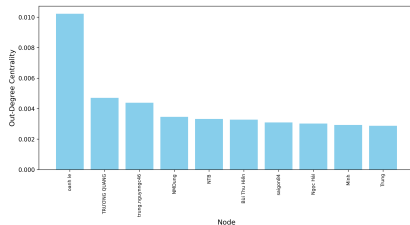


Figure: Out-Degree Centrality Bar Chart

Betweenness Centrality

- Thể hiện vai trò của từng thành viên trong việc kết nối cộng đồng.
- Những người có giá trị cao đóng vai trò cầu nối giúp thông tin lan truyền hiệu quả hơn trong mạng lưới.

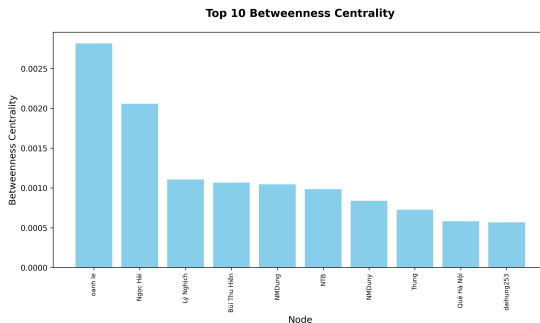


Figure: Betweenness Centrality Bar Chart

Closeness Centrality

- Đánh giá mức độ khả năng kết nối nhanh chóng của một thành viên đến tất cả các thành viên khác trong mạng.
- Giá trị cao cho thấy thành viên có vị trí chiến lược, giúp thông tin được truyền tải hiệu quả.

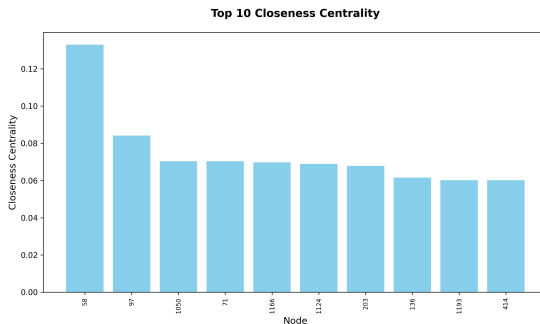


Figure: Closeness Centrality Bar Chart

Phân cụm đồ thị

Phân cụm đồ thị

Table: Kết quả phân cụm đồ thị

Model	Number of Communities	Modularity
Infomap	2223	0.453
Label Propagation	212	0.211
Louvain	43	0.519
Leiden	42	0.541

- **Infomap:** Phát hiện nhiều cụm nhất nhưng các cụm bị phân mảnh, kết nối yếu.
- **Label Propagation:** Hiệu quả thấp, số lượng cụm vừa phải nhưng liên kết nội bộ yếu.
- **Louvain và Leiden:** Tạo ra các cụm rõ ràng và liên kết chặt chẽ hơn.

Phân tích cụm

❶ Cụm lớn nhất - Cluster 0:

- Số lượng: Author (94), Commenter (5.466), Replier (2.503).
- Vai trò: Cộng đồng cân bằng về vai trò (Author, Replier, Commenter).
- Chủ đề chính: Kinh doanh & Quản lý, Giáo dục & Tri thức, Văn hóa & Lối sống.

❷ Cụm lớn thứ 2 - Cluster 2:

- Số lượng: Author (7), Commenter (3.632.), Replier (236).
- Tương tác: Tập trung vào bình luận, ít phản hồi.

❸ Các cụm khác:

- Cluster 5 và Cluster 6: Tập trung vào vai trò người phản hồi, nhấn mạnh các trao đổi qua lại.
- Cluster 41: Chỉ có 2 nút (1 tác giả, 1 bình luận viên), phản ánh mức độ tương tác tối thiểu.
- Cluster 19 và Cluster 6: Giáo dục & Tri thức, Văn hóa & Lối sống, Sức khỏe & Đời sống.

Kết luận

Kết luận

Về mặt lý thuyết:

- Đề xuất mô hình phân tích mối quan hệ tác giả - độc giả qua mạng lưới xã hội.
- Khám phá cấu trúc cộng đồng và vai trò của các thành phần trong mạng lưới.

Về mặt thực tiễn:

- Hỗ trợ nhận diện cụm cộng đồng và tối ưu hóa nội dung trên nền tảng.
- Nâng cao chiến lược tương tác và phát triển môi trường thảo luận trực tuyến.

Cảm ơn mọi người đã lắng
nghe!