

PROJECT 1A:

PHÂN TÍCH CỤM MẠNG LƯỚI QUAN HỆ DIỄN VIÊN ĐIỆN ẢNH VIỆT NAM

1st Bùi Minh Huy, 2nd Trần Lê Văn, 3rd Nguyễn Thị Thanh Tâm and Le Nhat Tung
HUTECH University, Vietnam

huybm.ds@gmail.com, vtran0712004@gmail.com, ttam99852@gmail.com, and lenhattung@hutech.edu.vn

Tóm tắt nội dung

Ngành điện ảnh được nhắc tới là một hệ sinh thái phức tạp, trong đó mối liên hệ của các diễn viên giữ vai trò trung tâm. Trong khi quốc tế đã có nhiều nghiên cứu mối quan hệ giữa các diễn viên của các nền điện ảnh lớn, nhưng tại Việt Nam thì vẫn còn thiếu những nghiên cứu chuyên sâu về đề tài này. Chính vì vậy, chúng tôi đã tiến hành thực hiện nghiên cứu này nhằm tạo nên mạng hợp tác của các diễn viên Việt nam dựa trên cơ sở dữ liệu phim chiếu rạp từ năm 2020, trong đó mỗi một diễn viên được minh họa bằng một nút và sự hợp tác trong cùng một bộ phim được minh họa bởi cạnh có trọng số. Bên cạnh đó, áp dụng các chỉ số thống kê mạng, phân tích các chỉ số trung tâm, phát hiện cộng đồng bằng các thuật toán Louvain, Leiden, Infomap và Label Propagation. Kết quả của nghiên cứu cho thấy mạng có tính chất "thế giới nhỏ" với các cụm hợp tác rõ ràng, làm nổi bật một vài cá thể đóng vai trò cầu nối. Những phát hiện này không chỉ mở rộng nghiên cứu học thuật về mạng xã hội mà còn cung cấp hàm ý thực tiễn về phát triển điện ảnh Việt Nam.

Từ khóa

Social Network Analysis, Community Detection, Actor Collaboration, Vietnamese Cinema, Complex Networks

I. GIỚI THIỆU

Trong bối cảnh hội nhập và phát triển, ngành công nghiệp điện ảnh Việt Nam đang từng bước khẳng định vị thế, không chỉ là một phương tiện giải trí mà còn là một lĩnh vực kinh tế - văn hóa năng động. Các tác phẩm điện ảnh đã và đang đóng một vai trò quan trọng trong việc hình thành và ảnh hưởng đến các khía cạnh của cuộc sống, không chỉ cuốn hút người xem bởi nội dung đa dạng và khả năng kể chuyện sinh động mà còn mang đến những trải nghiệm thẩm mỹ phong phú, những phút giây giải trí và còn là công cụ phản ánh, lưu giữ và lan tỏa bản sắc văn hóa của quốc gia. Trong 5 năm gần đây, điện ảnh đã có sự phát triển mạnh mẽ cả về số lượng lẫn chất lượng của các tác phẩm, qua đó càng khẳng định vị trí quan trọng trong đời sống tinh thần của khán giả cũng như trong chiến lược phát triển công nghiệp văn hóa đất nước. Trong quá trình sản xuất, mỗi một bộ phim điện ảnh được tạo nên từ sự kết hợp của nhiều cá nhân và tập thể với nhiều vai trò khác nhau. Trong đó, các diễn viên tham gia diễn xuất đóng vai trò trung tâm, bởi chính họ là người trực tiếp thể hiện câu chuyện và kết nối tác phẩm tới khán giả. Trong hệ sinh thái điện ảnh, mối quan hệ hợp tác giữa các diễn viên tưởng chừng là riêng lẻ nhưng thực chất khi kết hợp lại sẽ tạo nên một mạng xã hội phức tạp với cấu trúc và động lực riêng. Mạng lưới này hoạt động liên tục, trong các lựa chọn hợp tác không chỉ phản ánh sự hợp tác về mặt nghệ thuật hay chiến lược của nhà sản xuất, mà còn dẫn định hình nên các dòng chảy sáng tạo và xu hướng của toàn ngành. Chính vì vậy, việc nghiên cứu các mối quan hệ hợp tác giữa các diễn viên có thể mang lại những hiểu biết về cách ngành công nghiệp điện ảnh hoạt động cũng như phản ánh sự gắn kết nghề nghiệp, những xu hướng cộng đồng và phân hóa trong hệ sinh thái điện ảnh.

Để giải mã các cấu trúc mạng hợp tác của các diễn viên, thì phương pháp phân tích mạng xã hội (Social Network Analysis – SNA) là một công cụ nghiên cứu quan trọng trong khoa học dữ liệu, mạnh mẽ, cho phép khám phá cấu trúc và động lực hình thành các mối quan hệ trong nhiều hệ thống phức tạp. Ứng dụng SNA vào lĩnh vực điện ảnh đã được nhiều nghiên cứu quốc tế triển khai, đặc biệt tại Hollywood và Bollywood nhằm phân tích cấu trúc cộng đồng, đánh giá tầm ảnh hưởng của các diễn viên và làm rõ đặc điểm của tổ chức nói riêng của toàn ngành nói chung. Các nghiên cứu tiên phong này đã chỉ ra rằng, việc phân tích mạng lưới hợp tác diễn viên không chỉ là một chủ đề học thuật thú vị mà còn mang lại giá trị thực tiễn to lớn trong việc xác định các cá nhân có tầm ảnh hưởng, phát hiện các cụm cộng đồng sáng tạo và làm rõ đặc tính của toàn ngành.

Để kế thừa và phát triển hướng tiếp cận này với bối cảnh điện ảnh Việt Nam, nghiên cứu này đã được ra đời và được thực hiện với mục tiêu, xây dựng mạng hợp tác dựa trên dòng xuất hiện của diễn viên trong các bộ phim chiếu rạp từ năm 2020 đến nay và áp dụng phương pháp phân tích mạng xã hội để xây dựng, phân tích mạng lưới quan hệ hợp tác giữa các diễn viên điện ảnh. Từ đó, nghiên cứu sẽ cung cấp cái nhìn tổng thể dựa trên dữ liệu; không chỉ về cấu trúc hợp tác trong ngành điện ảnh Việt Nam mà còn hướng đến những ứng dụng thực tiễn, từ việc hỗ trợ quá trình tuyển chọn diễn viên, đề xuất chiến lược hợp tác hiệu quả đến việc đóng góp cho việc hoạch định chính sách phát triển văn hóa-nghệ thuật một cách khoa học và bền vững.

A. Động lực và thách thức

Ngành điện ảnh Việt Nam đang phát triển mạnh mẽ từng ngày. Tuy nhiên các nghiên cứu theo hướng mạng lưới hợp tác giữa các diễn viên vẫn còn hạn chế, chủ yếu tập trung vào doanh thu, thể loại hoặc chiến lược truyền thông hoặc dừng lại ở những thống kê mô tả đơn lẻ và phải đối diện với nhiều thách thức. Ta phải đối mặt với việc thu thập dữ liệu vì thông tin về các bộ phim và danh sách diễn viên mặc dù có thể thu thập được nhưng thường phân tán, thiếu chuẩn hóa và xảy ra tình trạng trùng lặp hoặc sai lệch thông tin. Ranh giới cộng đồng trong mạng diễn viên không chỉ được quyết định bởi yếu tố kỹ thuật mà còn chịu ảnh hưởng từ các bối cảnh thực tiễn như thể loại phim, hãng sản xuất, sự quen biết trong nghề, do đó việc diễn giải kết quả phải có sự phân tích định lượng và với góc nhìn bối cảnh. Cuối cùng, sự biến động nhanh chóng của ngành điện ảnh cũng đặt ra thách thức khi cấu trúc mạng có thể thay đổi đáng kể chỉ trong vài năm, khiến cho các phân tích tĩnh cần được đặt trong mối liên hệ với sự tiến hóa theo thời gian.

Cho đến nay, còn ít công trình xây dựng và phân tích một mạng hợp tác diễn viên điện ảnh dựa trên dữ liệu phim chiếu rạp một cách có hệ thống. Sự thiếu vắng này khiến cho chúng ta chưa có bằng chứng khoa học rõ ràng về mức độ gắn kết, sự phân tầng cộng đồng hay vai trò trung tâm của các diễn viên trong ngành điện ảnh Việt Nam. Những thách thức này nếu không được giải quyết thì sẽ làm hạn chế giá trị của nghiên cứu. Vì vậy, việc thiết kế một phương pháp tiếp cận có hệ thống, bao gồm các bước chuẩn hóa dữ liệu, xây dựng mạng lưới hợp tác hợp lý, áp dụng các thuật toán phát hiện cộng đồng phù hợp, và cuối cùng là diễn giải kết quả trong bối cảnh điện ảnh Việt Nam, chính là động lực thúc đẩy nghiên cứu này.

B. Đóng góp nghiên cứu

Nghiên cứu này mang lại một số đóng góp quan trọng cho lĩnh vực phân tích mạng xã hội và ứng dụng trong điện ảnh Việt Nam. Trước hết chúng tôi xây dựng một bộ dữ liệu thực nghiệm từ các bộ phim chiếu rạp kèm với đó là danh sách các diễn viên tham gia đóng tương ứng với bộ phim đó, các bộ này đều là phim Việt Nam được sản xuất kể từ năm 2020 cho đến nay. Bộ dữ liệu này đã được xử lý và chuẩn hóa nhằm đảm bảo tính toàn vẹn và có thể tái sử dụng trong các nghiên cứu tiếp theo. Dựa trên cơ sở dữ liệu này, nghiên cứu tiếp tục tiến hành phân tích định lượng mạng lưới hợp tác bằng cách tính toán các chỉ số đặc trưng như phân bố bậc (degree distribution), độ trung gian (betweenness centrality), độ gần (closeness centrality) và hệ số cụm (clustering coefficient), qua đó mô tả đặc điểm cấu trúc của mạng lưới diễn viên. Bên cạnh đó, nghiên cứu áp dụng và so sánh các thuật toán phát hiện cộng đồng phổ biến, bao gồm Louvain, Leiden, Infomap và Label Propagation, đồng thời đánh giá chất lượng cộng đồng bằng những thước đo như modularity và conductance. Điểm nổi bật của nghiên cứu này là việc diễn giải các cộng đồng thu được trong bối cảnh thực tế của Việt Nam, từ thể loại phim đến sự hình thành của các ekip quen thuộc, từ đó kết quả không chỉ có giá trị định lượng mà còn phản ánh đúng thực tiễn. Cuối cùng, những kết quả này mang lại các hàm ý thực tiễn quan trọng, có thể hỗ trợ gợi ý casting, xây dựng chiến lược hợp tác sản xuất, cũng như cung cấp bằng chứng khoa học cho việc hoạch định chính sách phát triển văn hóa và công nghiệp điện ảnh một cách bền vững.

C. Tổ chức bài báo

Phần còn lại của bài báo sẽ được tổ chức như sau: Mục II trình bày các công trình nghiên cứu liên quan, bao gồm những nghiên cứu quốc tế về mạng hợp tác diễn viên và các phương pháp phát hiện cộng đồng trong phân tích mạng xã hội. Mục III mô tả khung phương pháp nghiên cứu, trong đó trình bày cách tiếp cận để xây dựng mạng hợp tác diễn viên cũng như các thuật toán phát hiện cộng đồng được áp dụng. Mục IV cung cấp chi tiết về quá trình xây dựng và xử lý dữ liệu, từ khâu thu thập đến chuẩn hóa, nhằm đảm bảo tính toàn vẹn và độ tin cậy của bộ dữ liệu. Mục V mô tả thiết kế thí nghiệm và các tiêu chí đánh giá được sử dụng để phân tích chất lượng cộng đồng. Mục VI trình bày kết quả phân tích mạng lưới, bao gồm thống kê đặc trưng cấu trúc, kết quả phát hiện cộng đồng và diễn giải các phát hiện trong bối cảnh điện ảnh Việt Nam. Cuối cùng, Mục VII đưa ra phần kết luận, đồng thời thảo luận những hạn chế còn tồn tại và đề xuất các hướng nghiên cứu trong tương lai.

II. CÁC NGHIÊN CỨU LIÊN QUAN

Phần này cung cấp một đánh giá toàn diện về các phương pháp hiện có, được tổ chức bởi các nguyên tắc cơ bản và phương pháp tính toán của chúng.

A. Khái quát về phân tích mạng xã hội (SNA)

Phân tích mạng xã hội (Social Network Analysis – SNA) từ lâu đã trở thành một trong những phương pháp nền tảng để nghiên cứu, khám phá cấu trúc và động lực hình thành các mối quan hệ trong những hệ thống phức tạp trong nhiều lĩnh vực khác nhau như sinh học, truyền thông xã hội và khoa học. Với khả năng mô hình hóa các thực thể như nút (nodes) và mối quan hệ giữa chúng là những cạnh (edges), SNA cung cấp cho các nhà nghiên cứu những công cụ mạnh mẽ, cho phép các nhà nghiên cứu nhận diện các đặc trưng quan trọng của mạng lưới chẳng hạn như vai trò trung tâm, mức độ gắn kết và sự hình thành cộng đồng. Các nghiên cứu về mạng phức tạp đã cho thấy rằng nhiều hệ thống xã hội và tự nhiên đều sở hữu những đặc điểm chung [1], [2], bao gồm tính chất “thế giới nhỏ” (small-world property), sự xuất hiện của các nút trung tâm có ảnh hưởng lớn (hubs) và phân bố bậc theo luật lũy thừa, đã được phát hiện trong nhiều loại mạng. Những phát hiện này

khẳng định tính phổ quát của SNA, đồng thời nhấn mạnh tiềm năng ứng dụng của nó trong việc nghiên cứu các lĩnh vực khác nhau và tiềm năng ứng dụng rộng rãi của SNA trong việc phân tích các hệ thống đa dạng.

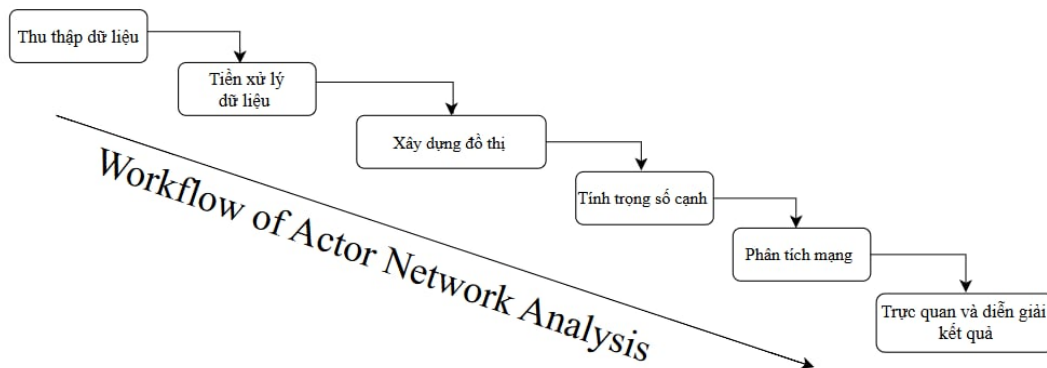
B. Nghiên cứu mạng hợp tác diễn viên trong điện ảnh quốc tế

Trong bối cảnh ngành điện ảnh, mạng hợp tác giữa các diễn viên đã trở thành một vấn đề nghiên cứu điển hình [3] và thu hút nhiều sự quan tâm của cộng đồng khoa học quốc tế. Các công trình tại Hollywood đã chỉ ra rằng mạng hợp tác diễn viên không chỉ phản ánh các cơ chế cộng tác trong sản xuất phim mà còn thể hiện những đặc tính chung của mạng và sự phân tầng theo mức độ ảnh hưởng. Các nghiên cứu cho thấy, thông qua việc phân tích các độ đo trung tâm (centrality measures) có thể xác định được những cá nhân đóng vai trò là trụ cột, có khả năng lan tỏa sức ảnh hưởng và kết nối tới các nhóm khác. Tiếp nối hướng nghiên cứu trên, nhiều công trình đã mở rộng sang các nền điện ảnh lớn khác như Bollywood [4]. Đáng chú ý ở đây là các phân tích không chỉ dừng lại ở việc mô tả cấu trúc, mà còn đi sâu vào việc lý giải các yếu tố văn hóa-xã hội, các yếu tố chi phối sự hình thành mạng lưới, tạo ra các cụm hợp tác bền chặt. Những kết quả đã khẳng định rằng mạng lưới hợp tác của các diễn viên trong giới nghệ thuật không chỉ là một cấu trúc kỹ thuật mà còn là sản phẩm của bối cảnh lịch sử, văn hóa và cơ chế vận hành đặc thù của mỗi nền điện ảnh.

C. Phương pháp phát hiện cộng đồng

Một trong những nhiệm vụ trọng tâm và mang lại nhiều giá trị nhất trong Phân tích Mạng xã hội là phát hiện cộng đồng (community detection). Trong mạng lưới hợp tác diễn viên, cộng đồng có thể được hiểu là các nhóm nghệ sĩ có tần suất hợp tác với nhau dày đặc hơn một cách đáng kể so với các thành viên bên ngoài nhóm. Việc xác định các cộng đồng này có ý nghĩa quan trọng, vì nó không chỉ giúp nhận diện các "trường phái" hay "ê-kíp" sáng tạo nòng cốt mà còn làm lộ rõ cấu trúc phân mảnh hay gắn kết của toàn ngành. Nhiều thuật toán phát hiện cộng đồng đã được phát triển và chứng minh hiệu quả, trong đó phổ biến là các thuật toán như Louvain (tối ưu modularity), Infomap (dựa trên quá trình random walk) [5], [6]. Mỗi thuật toán cung cấp một góc nhìn khác nhau về cấu trúc cộng đồng, và việc áp dụng chúng cho phép các nhà nghiên cứu có được một bức tranh đa chiều và sâu sắc hơn về các nhóm hợp tác trong mạng lưới. Điều này cho thấy phương pháp phân tích cộng đồng không chỉ giúp hiểu rõ cấu trúc nội tại của mạng lưới hợp tác, mà còn mang lại các giá trị thực tiễn trong quản trị và phát triển điện ảnh.

Các nghiên cứu quốc tế đã chứng minh được giá trị của phân tích mạng xã hội đối với sự hợp tác trong điện ảnh, sự thiếu vắng các công trình tương tự tại Việt Nam tạo ra một khoảng trống cả về học thuật lẫn thực tiễn. Nghiên cứu này nhằm lấp đầy những khoảng trống đó bằng cách xây dựng và phân tích mạng hợp tác của diễn viên điện ảnh Việt Nam dựa trên dữ liệu phim chiếu rạp từ năm 2020 cho đến nay. Công trình không chỉ nằm ở việc áp dụng các phương pháp phát hiện cộng đồng, mà còn ở cách diễn giải kết quả trong bối cảnh đặc thù của ngành điện ảnh Việt Nam, nơi các yếu tố văn hóa, lịch sử và tổ chức sản xuất có vai trò quyết định trong việc hình thành mạng lưới. Bằng việc kết hợp phân tích định lượng và diễn giải bối cảnh, nghiên cứu định vị mình như một bước khởi đầu có hệ thống cho việc nghiên cứu cộng đồng trong điện ảnh Việt Nam, đồng thời đóng góp những hàm ý thiết thực cho chiến lược phát triển của ngành công nghiệp văn hóa – nghệ thuật quốc gia.



Hình 1. Phương pháp nghiên cứu và phân tích mạng cộng đồng diễn viên

III. PHƯƠNG PHÁP NGHIÊN CỨU

Bài nghiên cứu này được triển khai theo một khung phương pháp có hệ thống nhằm xây dựng và phân tích mạng hợp tác của các diễn viên điện ảnh Việt Nam.

A. Xây dựng mạng hợp tác diễn viên

Một đồ thị G sẽ được kí hiệu: $G = (V, E)$ với V là tập hợp chứa các đỉnh, và E là tập hợp chứa các cạnh, mỗi cạnh có dạng một cặp giá trị $\{u, v\}$ (có thể được viết thành uv) [7]. Tập hợp đỉnh V của đồ thị G được kí hiệu $V(G)$, tập hợp cạnh được kí hiệu $E(G)$. Dữ liệu được thu thập từ danh sách phim chiếu rạp Việt Nam từ năm 2020 cho đến nay, trong đó mỗi một bộ phim có các thuộc tính như tên phim, đạo diễn, diễn viên, thể loại, ngày khởi chiếu. Tiến hành tiền xử lý để loại bỏ các trùng lặp và chuẩn hóa tên diễn viên thì mỗi diễn viên được mô hình hóa thành một nút $v \in V$, mỗi quan hệ hợp tác giữa hai diễn viên được biểu diễn bằng một cạnh $e = (u, v) \in E$. Để phản ánh mức độ hợp tác của mỗi diễn viên, mỗi cạnh được gán một trọng số dựa trên số lần hai diễn viên đo xuất hiện cùng nhau trong các tác phẩm điện ảnh. cách tiếp cận này đảm bảo rằng mạng hợp tác phản ánh trung thực cường độ hợp tác thực tế giữa các diễn viên.

B. Đặc trưng cấu trúc mạng

Tiến hành xây dựng đồ thị hoàn chỉnh, chúng tôi tiến hành phân tích các đặc trưng cơ bản nhằm mô tả chi tiết cấu trúc của mạng. Các chỉ số được sử dụng bao gồm:

- **Bậc trung bình:**

$$\bar{k} = \frac{2m}{n} \quad (1)$$

trong đó m là số nút và n là số cạnh.

- **Mật độ mạng:**

$$D = \frac{2m}{n(n-1)} \quad (2)$$

- **Hệ số gom cụm trung bình:**

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i \quad (3)$$

- **Độ dài đường đi trung bình:**

$$L = \frac{1}{n(n-1)} \sum_{i \neq j} d(i, j) \quad (4)$$

- **Đường kính mạng:** được xác định là khoảng cách ngắn nhất dài nhất giữa hai nút bất kỳ.

C. Phát hiện cộng đồng

Một trọng tâm của nghiên cứu là khám phá cấu trúc cộng đồng, có nghĩa là các nhóm diễn viên có xu hướng hợp tác thường xuyên hơn với nhau so với các thành viên bên ngoài nhóm. Chúng tôi sử dụng thuật toán Louvain [5] để phát hiện cộng đồng thông qua việc tối ưu hóa chỉ số modularity, được định nghĩa như sau:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \quad (5)$$

trong đó A_{ij} là ma trận kề của mạng, k_i và k_j là bậc của các nút i và j , còn $\delta(c_i, c_j)$ nhận giá trị 1 nếu hai nút thuộc cùng một cộng đồng, và 0 nếu khác cộng đồng.

Để phát hiện cộng đồng, chúng tôi sử dụng thuật toán **Louvain**, một trong những thuật toán hiệu quả và phổ biến nhất để tối ưu hóa chỉ số modularity trong các mạng lớn. Phương pháp này không chỉ cho phép nhận diện các ê-kíp sáng tạo nòng cốt mà còn làm rõ mức độ phân mảnh hay gắn kết của toàn bộ hệ thống hợp tác điện ảnh. Bên cạnh Louvain, nghiên cứu này còn triển khai và so sánh thêm ba thuật toán khác nhằm tăng tính toàn diện: **Leiden** [8] là một cải tiến trực tiếp từ Louvain, giúp khắc phục hạn chế về tính ổn định và cho ra các cộng đồng cân bằng hơn, **Infomap** [9] dựa trên cơ chế random walk nhằm nén thông tin dòng chảy trong mạng, từ đó phát hiện được các cộng đồng có cấu trúc chi tiết và phân mảnh hơn, **Label Propagation** [10] là một phương pháp dựa trên lan truyền nhãn giữa các nút theo quy tắc đa số, nổi bật nhờ tốc độ xử lý rất nhanh, tuy nhiên kết quả thường kém ổn định hơn. Sử dụng bốn thuật toán (Louvain, Leiden, Infomap và Label Propagation) cho phép đối chiếu và đánh giá chéo, từ đó lựa chọn được kết quả cộng đồng có chất lượng tối ưu, thay vì phụ thuộc vào một phương pháp duy nhất.

D. Các chỉ số trung tâm

Bên cạnh việc phân tích cộng đồng, chúng tôi tập trung xác định các diễn viên giữ vai trò quan trọng trong mạng thông qua nhiều thước đo trung tâm.

- **Degree Centrality (Số đo bậc trung tâm): Degree Centrality:**

$$C_D(v) = \frac{\deg(v)}{n-1} \quad (6)$$

- **Weighted Degree Centrality (Độ trung tâm có trọng số):**

$$C_{WD}(v) = \sum_{u \in N(v)} w(u, v) \quad (7)$$

- **Betweenness Centrality (Độ trung gian):**

$$C_B(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (8)$$

trong đó σ_{st} là tổng số đường đi ngắn nhất giữa s và t , còn $\sigma_{st}(v)$ là số đường đi trong đó nút v xuất hiện.

- **Closeness Centrality (Độ gần gũi):**

$$C_C(v) = \frac{1}{\sum_{t \in V \setminus v} d_C(v, t)} \quad (9)$$

trong đó n là số lượng nút và $d(u, v)$ là độ dài đường đi ngắn nhất từ u đến v .

- **Eigenvector Centrality (Độ ảnh hưởng):**

$$C_E(v) = \frac{1}{\lambda} \sum_{u \in N(v)} A_{vu} C_E(u) \quad (10)$$

- **PageRank:**

$$PR(v) = \frac{1-d}{n} + d \sum_{u \in N(v)} \frac{PR(u)}{\deg(u)} \quad (11)$$

trong đó d là hệ số suy giảm.

Việc kết hợp nhiều chỉ số trung tâm cho phép phân biệt rõ những diễn viên “phổ biến” với nhiều kết nối trực tiếp (degree), những diễn viên giữ vai trò cầu nối quan trọng (betweenness), và những diễn viên có ảnh hưởng sâu rộng trong mạng (eigenvector, PageRank).

E. Trực quan hóa mạng

Chúng tôi sử dụng công cụ Gephi [11] với bộ cục ForceAtlas2 [12] để hỗ trợ cho quá trình trực quan hóa mạng. Mạng được biểu diễn ở dạng tổng quan với kích thước nút cố định và màu sắc đồng nhất để làm rõ cấu trúc liên kết tự nhiên. Sau đó, chúng tôi điều chỉnh kích thước nút theo *Degree Centrality* và tô màu theo cộng đồng được phát hiện bởi các thuật toán đã nêu (Louvain [6], Leiden [8], Infomap [9], Label Propagation [10]). Cách tiếp cận này cho phép quan sát trực quan sự phân cụm và nhận diện rõ các cá nhân trung tâm trong toàn mạng, đồng thời so sánh trực quan sự khác biệt giữa các phương pháp phát hiện cộng đồng.

IV. XÂY DỰNG VÀ CHUẨN BỊ DỮ LIỆU

A. Nguồn dữ liệu và cách thu thập

Nghiên cứu sử dụng hai nguồn dữ liệu chính. Thứ nhất là bộ dữ liệu phim chiếu rạp Việt Nam giai đoạn 2020–nay, bao gồm 176 tác phẩm với thông tin chi tiết về tên phim, đạo diễn, danh sách diễn viên, thể loại và ngày công chiếu. Thứ hai là bộ dữ liệu diễn viên Việt Nam với 618 diễn viên, bổ sung thêm các thông tin về tiểu sử, quốc tịch, quá trình hoạt động nghệ thuật và hình ảnh minh họa. Hai nguồn dữ liệu này là nền tảng để xây dựng mạng cộng tác, trong đó các nút là diễn viên và các cạnh phản ánh quan hệ hợp tác qua từng bộ phim.

Bảng 1 minh họa cấu trúc dữ liệu phim. Mỗi hàng đại diện cho một bộ phim, trong đó trường “Diễn viên” chứa nhiều tên cần được tách, chuẩn hóa và loại bỏ trùng lặp trong giai đoạn tiền xử lý.

Bảng I
VÍ DỤ VỀ CẤU TRÚC DỮ LIỆU PHIM CHIẾU RẠP VIỆT NAM TỪ NĂM 2020

Tên phim	Đạo diễn	Diễn viên	Thể loại	Khởi chiếu
Đôi mắt âm dương	Nhật Trung	Thu Trang, Quốc Trường, NSND Ngọc Giàu, Trung Dân...	Tâm lý, Kinh dị	25/01/2020
30 chưa phải Tết	Nguyễn Quang Huy	NSND Hồng Vân, Trường Giang, Mạc Văn Khoa, Đức Phúc...	Hài	25/01/2020
Gái già lắm chiêu 3	Bảo Nhân, Nam Cito	Ninh Dương Lan Ngọc, NSND Lê Khanh, Jun Vũ...	Hài, Lãng mạn	25/01/2020

B. Thống kê và đặc trưng dữ liệu

Bảng II trình bày các đặc trưng thống kê chính.

Bảng II
THỐNG KÊ ĐẶC TRƯNG CỦA MẠNG CỘNG TÁC

Dữ liệu	Số nút	Số cạnh	Bậc trung bình	Hệ số gom cụm	Đường kính	Số cộng đồng
Mạng phim	563	3,405	12.096	0.7907	7	7

Kết quả cho thấy mạng phim có quy mô lớn hơn nhưng mức độ gắn kết cộng đồng thấp hơn, trong khi mạng diễn viên nhỏ hơn nhưng chặt chẽ hơn, với hệ số gom cụm cao.

C. Quy trình tiền xử lý dữ liệu

Để đảm bảo chất lượng, dữ liệu được xử lý theo các bước:

- **Làm sạch:** Loại bỏ cạnh trùng lặp, cạnh tự nối và các nút cô lập.
- **Lọc thành phần:** Giữ lại thành phần liên thông lớn nhất để duy trì tính kết nối toàn cục.
- **Lọc theo bậc:** Loại bỏ nút có bậc nhỏ hơn 2 để giảm nhiễu.
- **Sắp xếp thời gian:** Các cạnh được sắp xếp theo mốc công chiếu để đảm bảo tính thời gian.
- **Kiểm định chất lượng:** So sánh phân bố bậc, số thành phần và các chỉ số mạng sau xử lý.

D. Đảm bảo chất lượng dữ liệu

Để xác thực dữ liệu, áp dụng các biện pháp:

- **Kiểm định thống kê:** So sánh phân bố bậc, đường kính, hệ số gom cụm.
- **Nhất quán thời gian:** Duy trì trật tự cạnh theo mốc chiếu.
- **Xác thực cộng đồng:** Đánh giá modularity và conductance.
- **Đầy đủ dữ liệu:** Đảm bảo không mất nút hoặc cạnh.

Bộ dữ liệu cuối cùng phản ánh trung thực cấu trúc mạng cộng tác điện ảnh Việt Nam, tạo nền tảng vững chắc cho các phân tích tiếp theo.

V. THỰC NGHIỆM VÀ ĐÁNH GIÁ

Trong chương này, chúng tôi trình bày quá trình thực nghiệm cùng với việc phân tích và đánh giá các kết quả thu được từ mạng lưới cộng tác giữa các diễn viên.

A. Dữ liệu thực nghiệm

Dữ liệu thực nghiệm được xây dựng từ hai nguồn chính: bộ phim chiếu rạp Việt Nam giai đoạn 2020–nay với 176 tác phẩm và bộ dữ liệu 618 diễn viên từ các diễn viên đóng trong các phim chiếu rạp Việt Nam có thông tin tiểu sử, quốc tịch và quá trình hoạt động nghệ thuật. Sau khi tiền xử lý, mạng cộng tác được hình thành với 563 nút và 3,405 cạnh, trong đó trọng số cạnh thể hiện số lần cùng tham gia nhiều phim. Mạng có bậc trung bình 12.096, hệ số gom cụm 0.7907 và đường kính 7, phản ánh quy mô lớn và mức độ gắn kết cộng đồng cao. Bộ dữ liệu này được coi là đại diện cho bức tranh hợp tác điện ảnh Việt Nam, đảm bảo tính đầy đủ, khách quan và tạo nền tảng vững chắc cho các phân tích trung tâm cũng như cấu trúc cộng đồng.

B. Môi trường thực nghiệm

Các thực nghiệm được tiến hành trên máy tính cá nhân cấu hình chip Apple M3, RAM 16GB, hệ điều hành macOS. Môi trường lập trình sử dụng Python 3.9 với các thư viện pandas, numpy, networkx, matplotlib. Ngoài ra, công cụ Gephi phiên bản 0.10.1 cùng các plugin hỗ trợ cũng được chúng tôi sử dụng cho việc trực quan hóa và phân tích mạng.

C. Tiền xử lý dữ liệu

Chúng tôi tiến hành tiền xử lý dữ liệu qua một chuỗi bước nhằm chuẩn hoá và chuẩn bị cho việc xây dựng mạng cộng tác. Đầu tiên, dữ liệu gốc được làm sạch bằng cách loại bỏ ký tự thừa, chuẩn hoá khoảng trắng, bỏ dấu tiếng Việt và chuyển toàn bộ về một định dạng thống nhất, từ đó giảm thiểu sai lệch khi so khớp tên. Tiếp theo, danh sách diễn viên trong từng phim được tách riêng dựa trên dấu phẩy hoặc chấm phẩy, đồng thời xử lý các trường hợp đặc biệt như nhóm diễn viên trong ngoặc hoặc bí danh. Sau khi tách, chúng tôi loại bỏ trùng lặp trong cùng một phim để đảm bảo một diễn viên chỉ xuất hiện một lần cho mỗi tác phẩm.

Trên cơ sở đó, dữ liệu được chuẩn hoá và gán mã định danh duy nhất cho từng phim và từng diễn viên. Từ đây, chúng tôi xây dựng hai bảng riêng biệt: một bảng liệt kê thông tin diễn viên (danh sách các nút trong mạng) và một bảng thể hiện quan hệ cộng tác giữa các cặp diễn viên (danh sách các cạnh trong mạng). Ở bảng quan hệ, nếu hai diễn viên cùng tham gia nhiều phim thì trọng số cạnh sẽ được tăng tương ứng, phản ánh số lần hợp tác. Cuối cùng, dữ liệu được kiểm tra toàn vẹn, loại bỏ cạnh tự nối và các giá trị không hợp lệ, đảm bảo mạng được xây dựng có cấu trúc rõ ràng và chính xác, sẵn sàng cho các bước phân tích tiếp theo.

D. Các chỉ số thống kê đặc trưng

Bảng III
CÁC THỐNG KÊ ĐẶC TRƯNG CỦA MẠNG

Chỉ số	Giá trị
Số nút	563
Số cạnh	3405
Bậc trung bình	12.096
Mật độ mạng	0.0215
Thành phần liên thông	7 (lớn nhất: 545 nút)
Hệ số gom cụm trung bình	0.7907
Đường kính (giant)	7
Độ dài đường đi trung bình	3.1644

Bảng V-D trình bày các thống kê cơ bản phản ánh đặc trưng cấu trúc của mạng nghiên cứu. Mạng bao gồm tổng cộng 563 nút và 3,405 cạnh, cho thấy đây là một hệ thống có quy mô trung bình nhưng sở hữu số lượng liên kết tương đối dày đặc. Với bậc trung bình đạt 12.096, mỗi nút trong mạng thường kết nối trực tiếp với hơn 12 nút khác. Chỉ số này phản ánh mức độ tương tác cao, gợi ý rằng thông tin hoặc ảnh hưởng có khả năng lan truyền nhanh chóng trong toàn hệ thống. Tuy nhiên, mật độ mạng chỉ đạt 0.0215, cho thấy so với số liên kết tối đa có thể hình thành, thì cấu trúc mạng vẫn ở trạng thái thưa. Điều này là đặc điểm phổ biến của các mạng xã hội và mạng phức tạp thực tế, nơi không phải mọi nút đều kết nối với nhau, mà liên kết thường tập trung vào một số cụm hoặc nhóm đặc thù.

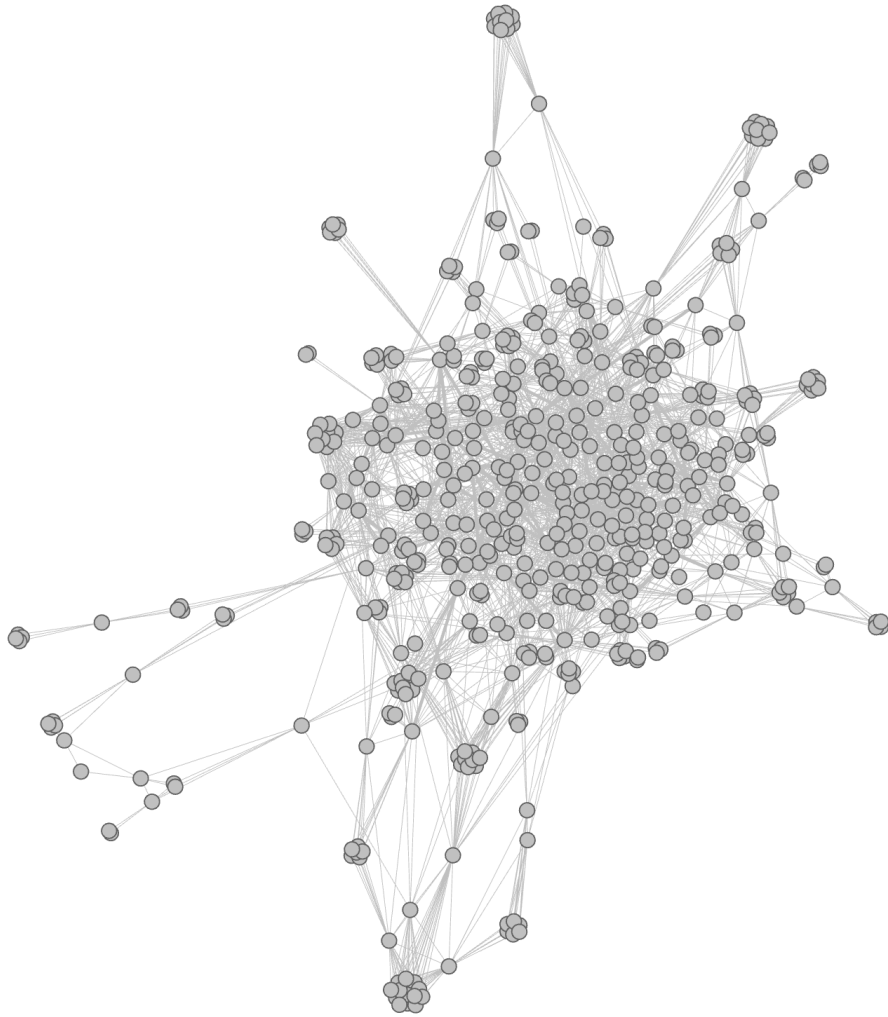
Xét về tính liên thông, mạng được chia thành 7 thành phần riêng biệt, trong đó thành phần lớn nhất chứa 545 nút, chiếm phần lớn tổng thể mạng. Sự tồn tại của thành phần khổng lồ này cho thấy đa số các nút đều được kết nối với nhau trong một cấu trúc lớn, trong khi một số nhỏ nút bị cô lập hoặc hình thành các cụm tách biệt. Đặc điểm này có ý nghĩa quan trọng, bởi các hiện tượng như sự lan truyền thông tin, sự hình thành ảnh hưởng hoặc các tiến trình động lực học khác chủ yếu diễn ra trong thành phần lớn nhất của mạng. Hệ số gom cụm trung bình đạt 0.7907, một giá trị khá cao so với nhiều loại mạng ngẫu nhiên có cùng kích thước và mật độ. Chỉ số này cho thấy các nút trong mạng có xu hướng hình thành những cụm liên kết chặt chẽ, phản ánh đặc điểm “bạn của bạn cũng có khả năng là bạn của nhau”. Đây chính là đặc tính tiêu biểu của mạng xã hội và mạng cộng đồng, trong đó các kết nối thường mang tính tập trung và có cấu trúc nhóm rõ rệt.

Xét về khoảng cách, đường kính của thành phần khổng lồ là 7, nghĩa là khoảng cách xa nhất giữa hai nút bất kỳ trong thành phần này không vượt quá 7 bước. Bên cạnh đó, độ dài đường đi trung bình chỉ đạt 3.1644, cho thấy các nút trong mạng thường kết nối với nhau thông qua rất ít trung gian. Đây là minh chứng rõ rệt cho hiệu ứng “thế giới nhỏ” (small-world), nơi mà mặc dù quy mô mạng lớn, khoảng cách trung bình giữa hai cá thể bất kỳ vẫn rất ngắn. Đặc điểm này có ý nghĩa quan trọng trong nghiên cứu các quá trình lan truyền, bởi nó cho thấy thông tin, ảnh hưởng hoặc dịch bệnh có thể lan tỏa nhanh chóng trong hệ thống. Các chỉ số thống kê cho thấy mạng nghiên cứu vừa mang đặc điểm điển hình của mạng xã hội (mức độ gom cụm cao, cấu trúc thế giới nhỏ), vừa thể hiện tính phân tầng nhất định (sự tồn tại của nhiều thành phần liên thông nhỏ bên ngoài giant component). Những đặc trưng này không chỉ phản ánh bản chất tổ chức của mạng mà còn tạo nền tảng quan trọng cho các phân tích tiếp theo về chỉ số trung tâm, cấu trúc cộng đồng và trực quan hóa mạng.

E. Trực quan hóa và biểu diễn mạng

Để giúp hình dung rõ hơn về cấu trúc mạng cộng tác giữa các diễn viên, chúng tôi tiến hành trực quan hóa đồ thị bằng công cụ Gephi với bố cục ForceAtlas2. Trước hết, mạng được biểu diễn ở dạng tổng quan ban đầu với kích thước và màu sắc nút mặc định, nhằm phản ánh cấu trúc liên kết tự nhiên. Sau đó, để làm nổi bật vai trò của từng diễn viên và sự phân cụm

cộng đồng, chúng tôi điều chỉnh kích thước nút theo chỉ số Degree Centrality và tô màu theo cộng đồng được phát hiện bởi thuật toán Louvain.



Hình 2. Trực quan mạng cộng tác diễn viên với bố cục ForceAtlas2, nút hiển thị mặc định.

Quan sát từ Hình 2 có thể thấy mạng mang đặc trưng *core-periphery*: phần lõi tập trung nhiều cạnh dày đặc phản ánh nhóm diễn viên thường xuyên hợp tác, trong khi các nút ở rìa đại diện cho những cá nhân ít kết nối hoặc chỉ tham gia một vài dự án nhỏ. Trong trực quan này, cạnh được gán trọng số theo số lần hai diễn viên cùng xuất hiện trong một bộ phim; do đó, các cạnh dày hơn biểu diễn tần suất hợp tác cao hơn, trong khi cạnh mảnh hơn phản ánh sự kết nối ít thường xuyên. Điều này phù hợp với các chỉ số thống kê toàn mạng: mật độ chung thấp nhưng hệ số gom cụm cao cho thấy các nhóm nhỏ liên kết chặt chẽ trong khi toàn mạng vẫn khá thưa.

F. Các chỉ số trung tâm

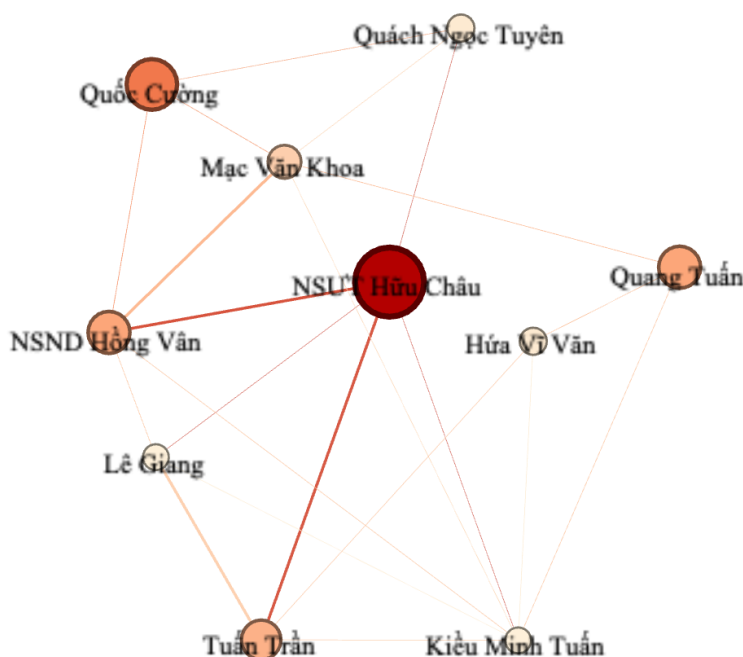
Ở phần này, chúng tôi sẽ tập trung vào việc phân tích các chỉ số trung tâm nhằm xác định những diễn viên giữ vai trò quan trọng trong mạng lưới cộng tác. Mỗi chỉ số phản ánh một khía cạnh khác nhau của mức độ ảnh hưởng hay vị trí chiến lược của một nút trong mạng. Cụ thể, chúng tôi xem xét sáu chỉ số chính: Degree Centrality, Weighted Degree, Betweenness Centrality, Closeness Centrality, Eigenvector Centrality và PageRank. Với mỗi chỉ số, chúng tôi trình bày bảng xếp hạng Top 10 diễn viên cùng đồng thời trực quan minh họa và phân tích vai trò đặc trưng của họ trong mạng.

1) *Degree Centrality*: Degree Centrality đo lường số lượng kết nối trực tiếp của một nút trong mạng, phản ánh mức độ “phổ biến” hay khả năng hợp tác rộng rãi của một diễn viên. Nói cách khác, những diễn viên có Degree cao thường tham gia vào nhiều bộ phim và kết nối với nhiều cộng sự khác.

Hình 3 trình bày danh sách Top 10 diễn viên có Degree Centrality cao nhất. Kết quả cho thấy NSUT Hữu Châu đứng đầu với 78 kết nối, theo sau là Quốc Cường (63), Quang Tuấn (56) và NSND Hồng Vân (56). Các vị trí còn lại đều có trên 40 kết nối, cho thấy đây là những diễn viên có mạng lưới hợp tác rộng lớn.

Id	Label	Degree ▾
A000274	NSUT Hữu Châu	78
A000381	Quốc Cường	63
A000255	NSND Hồng Vân	56
A000373	Quang Tuấn	56
A000493	Tuấn Trần	54
A000249	Mạc Văn Khoa	49
A000378	Quách Ngọc Tuyên	44
A000143	Hứa Vĩ Văn	44
A000177	Kiều Minh Tuấn	43
A000197	Lê Giang	43

Hình 3. Top 10 diễn viên theo Degree Centrality



Hình 4. Trực quan Top 10 diễn viên theo Degree Centrality (kích thước và màu sắc nút phản ánh số lượng kết nối)

Để trực quan hơn, Hình 4 minh họa mạng con của 10 diễn viên có Degree Centrality cao nhất cùng các kết nối trực tiếp của họ. Trong hình, kích thước nút tỷ lệ thuận với số lượng kết nối, còn màu sắc chuyển từ nhạt đến đậm phản ánh giá trị Degree tăng dần. Nhờ đó, có thể dễ dàng nhận thấy NSUT Hữu Châu và NSND Hồng Vân giữ vị trí nổi bật nhất, đóng vai trò hạt nhân kết nối nhiều nhóm cộng đồng. Bên cạnh đó, các diễn viên như Mạc Văn Khoa hay Lê Giang tuy Degree thấp hơn nhưng vẫn góp phần quan trọng vào việc mở rộng mạng lưới hợp tác.

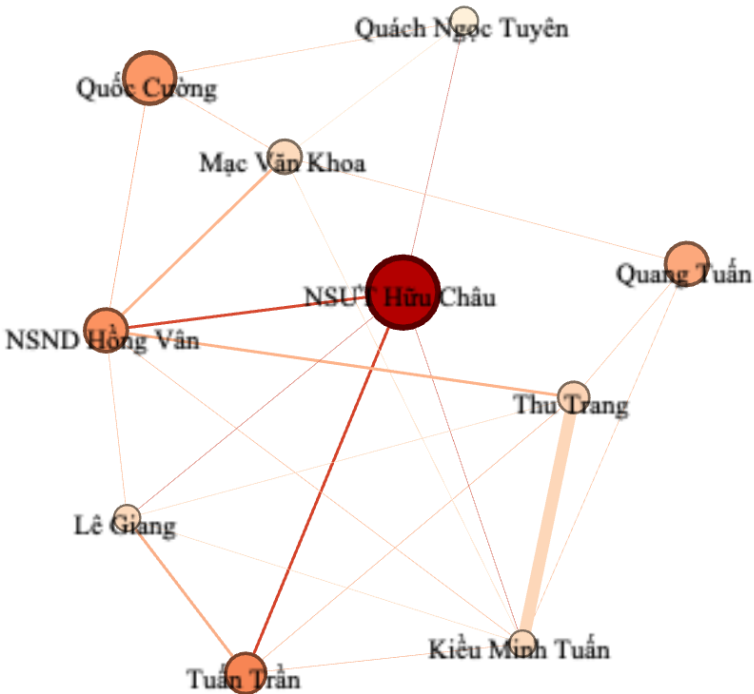
Như vậy, Degree Centrality đã giúp xác định những diễn viên có số lượng kết nối trực tiếp lớn nhất trong mạng, phản ánh mức độ phổ biến và khả năng hợp tác rộng rãi. Tuy nhiên, chỉ số này mới dừng lại ở việc đo lường số lượng, chưa tính đến

tần suất hay mức độ bền chặt của các mối quan hệ hợp tác. Để khắc phục hạn chế này, phần tiếp theo sẽ phân tích Weighted Degree Centrality, nhằm làm rõ những diễn viên không chỉ có nhiều kết nối mà còn duy trì các mối quan hệ cộng tác lặp lại và chặt chẽ hơn.

2) *Weighted Degree Centrality*: Khác với Degree Centrality chỉ phản ánh số lượng kết nối trực tiếp, **Weighted Degree Centrality** còn tính đến trọng số của cạnh, tức là số lần hợp tác giữa các diễn viên. Do đó, chỉ số này không chỉ cho biết ai có nhiều mối quan hệ, mà còn làm nổi bật những diễn viên thường xuyên hợp tác lặp lại, thể hiện sự gắn kết bền chặt trong mạng.

Id	Label	Weighted Degree ▾
A000274	NSUT Hữu Châu	83.0
A000493	Tuấn Trần	66.0
A000255	NSND Hồng Vân	64.0
A000381	Quốc Cường	63.0
A000373	Quang Tuấn	60.0
A000425	Thu Trang	52.0
A000249	Mạc Văn Khoa	52.0
A000177	Kiều Minh Tuấn	51.0
A000197	Lê Giang	51.0
A000378	Quách Ngọc Tuyên	47.0

Hình 5. Top 10 diễn viên có Weighted Degree cao nhất



Hình 6. Mạng con gồm 10 diễn viên có Weighted Degree cao nhất

Kết quả cho thấy NSUT Hữu Châu tiếp tục giữ vị trí trung tâm với Weighted Degree cao nhất (83), thể hiện ông không chỉ hợp tác rộng rãi mà còn lặp lại hợp tác nhiều lần với các đồng nghiệp. Các diễn viên như Tuấn Trần (66), NSND Hồng Vân (64) và Quốc Cường (63) cũng nổi bật nhờ duy trì nhiều mối quan hệ hợp tác bền chặt. Điều này cho thấy Weighted Degree làm rõ chiều sâu của mạng lưới, nhấn mạnh vai trò của những diễn viên thường xuyên cộng tác trong nhiều dự án khác nhau,

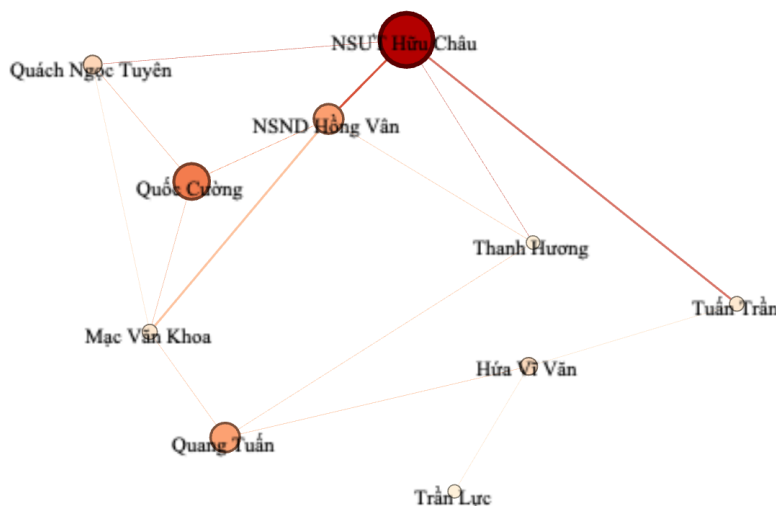
chứ không chỉ là số lượng kết nối. Các cạnh dày trong hình trực quan thể hiện tần suất hợp tác cao, minh chứng cho các cặp diễn viên có mối quan hệ cộng tác lâu dài.

Tuy nhiên, cả Degree và Weighted Degree đều chủ yếu phản ánh số lượng và tần suất kết nối trực tiếp, mà chưa thể hiện rõ vai trò của một diễn viên trong việc kết nối các nhóm khác nhau trong mạng. Vì vậy, phần tiếp theo sẽ phân tích Betweenness Centrality, một chỉ số nhấn mạnh những diễn viên giữ vị trí cầu nối giữa các cộng đồng, giúp cho sự lan tỏa hợp tác trong toàn bộ mạng lưới.

3) *Betweenness Centrality*: Betweenness Centrality đo lường mức độ một diễn viên đóng vai trò cầu nối trong mạng, tức là số lần nút xuất hiện trên các đường đi ngắn nhất giữa các cặp nút khác. Diễn viên có giá trị betweenness cao thường là những người giữ vai trò trung gian, giúp kết nối các cộng đồng khác nhau trong mạng.

Id	Label	Betweenness Centrality ∨	top10bet
A000274	NSUT Hữu Châu	16053.505244	0.101836
A000381	Quốc Cường	11311.741356	0.071756
A000255	NSND Hồng Vân	9689.500236	0.061466
A000373	Quang Tuấn	9600.04853	0.060898
A000378	Quách Ngọc Tuyên	6706.935989	0.042546
A000143	Hứa Vi Văn	6432.052757	0.040802
A000249	Mạc Văn Khoa	5778.444777	0.036656
A000493	Tuấn Trần	5679.260224	0.036027
A000412	Thanh Hương	5380.485134	0.034131
A000474	Trần Lực	5322.479461	0.033763

Hình 7. Bảng Top 10 diễn viên theo chỉ số Betweenness Centrality (bao gồm giá trị chuẩn hóa).



Hình 8. Trực quan mạng con của Top 10 diễn viên theo chỉ số Betweenness Centrality, kích thước và màu sắc tỷ lệ với giá trị betweenness.

Kết quả cho thấy NSUT Hữu Châu có chỉ số betweenness cao nhất, thể hiện vai trò cầu nối quan trọng giữa nhiều nhóm diễn viên khác nhau. Quốc Cường, NSND Hồng Vân và Quang Tuấn cũng giữ vị trí nổi bật với giá trị betweenness cao, cho thấy khả năng kết nối và trung gian giữa các cụm cộng đồng. Điều này bổ sung cho Degree Centrality và Weighted Degree, khi Betweenness không chỉ phản ánh số lượng hay tần suất hợp tác, mà còn nhấn mạnh vai trò chiến lược trong việc duy trì sự gắn kết toàn mạng.

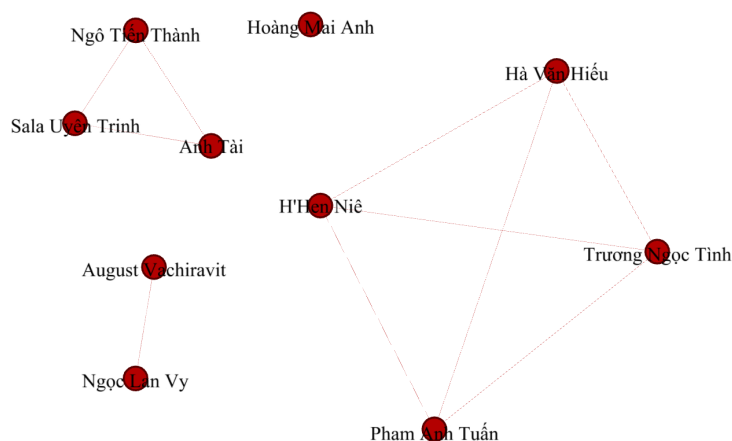
Tuy vậy, Betweenness Centrality chủ yếu nhấn mạnh vai trò trung gian, mà chưa phản ánh khả năng một diễn viên tiếp cận nhanh chóng tới toàn bộ các diễn viên khác trong mạng. Do đó, trong phần tiếp theo chúng tôi tập trung vào chỉ số Closeness Centrality, nhằm đo lường mức độ trung tâm toàn cục và khả năng lan tỏa trong mạng cộng tác.

4) *Closeness Centrality*: Chỉ số Closeness Centrality đo lường mức độ “gần gũi” trung bình của một diễn viên so với tất cả những người khác trong mạng. Giá trị closeness càng cao, diễn viên đó càng có khả năng nhanh chóng tiếp cận các nút

khác thông qua số bước ngắn nhất. Trong dữ liệu này, có đến 19 diễn viên đạt cùng giá trị cực đại (1.0) do đặc trưng mạng có nhiều thành phần nhỏ và chặt chẽ. Để đảm bảo tính thống nhất, chúng tôi chỉ liệt kê 10 diễn viên tiêu biểu trong Bảng 9.

Id	Label	Closeness... ▾
A000007	Anh Tài	1.0
A000324	Ngô Tiến Thành	1.0
A000394	Sala Uyên Trinh	1.0
A000015	August Vachiravit	1.0
A000328	Ngọc Lan Vy	1.0
A000077	H'Hen Niê	1.0
A000119	Hà Văn Hiếu	1.0
A000360	Phạm Anh Tuấn	1.0
A000461	Trương Ngọc Tình	1.0
A000091	Hoàng Mai Anh	1.0

Hình 9. Top 10 diễn viên theo chỉ số Closeness Centrality.



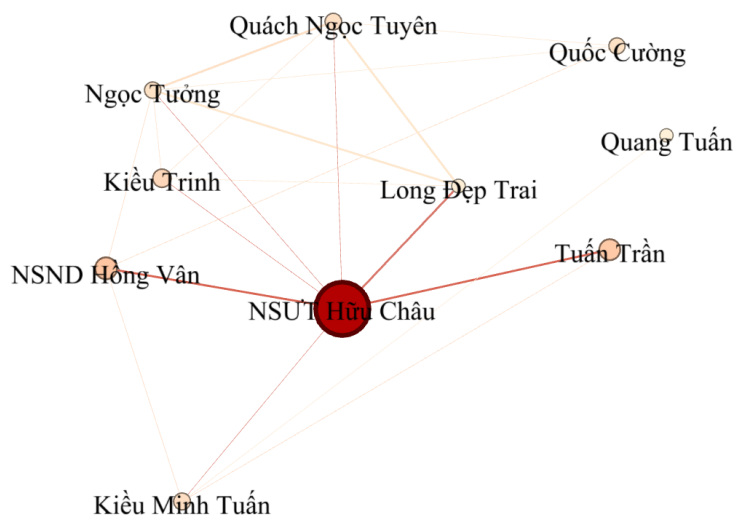
Hình 10. Trực quan mạng con của Top 10 diễn viên theo chỉ số Closeness Centrality, kích thước và màu sắc tỷ lệ với giá trị closeness.

Kết quả cho thấy nhiều diễn viên ít tên tuổi hoặc chỉ tham gia trong các thành phần nhỏ (ví dụ: Anh Tài, Ngô Tiến Thành, Sala Uyên Trinh, H'Hen Niê, Johnny Trí Nguyễn, NSND Tự Long) vẫn đạt giá trị closeness bằng 1.0. Điều này phản ánh tính chất kỹ thuật của chỉ số closeness trong mạng phân mảnh: các nút trong tiểu thành phần nhỏ thường có khoảng cách trung bình tối thiểu tới các nút khác cùng thành phần, dẫn đến điểm số tối đa. Tuy vậy, closeness chưa đủ để khẳng định vai trò ảnh hưởng rộng, mà chỉ nhấn mạnh khả năng tiếp cận nội bộ trong cụm. Do đó, chỉ số này cần được bổ sung bằng các chỉ số khác như Eigenvector Centrality hay PageRank để đánh giá ảnh hưởng dựa trên kết nối với các nút quan trọng trong toàn mạng.

5) *Eigenvector Centrality*: Chỉ số Eigenvector Centrality không chỉ xét số lượng kết nối trực tiếp, mà còn đánh giá mức độ quan trọng của một diễn viên dựa trên việc họ kết nối với những diễn viên khác cũng quan trọng. Nói cách khác, đây là thước đo “sức ảnh hưởng lan tỏa” trong mạng.

Id	Label	Eigenvector Centrality ▾
A000274	NSUT Hữu Châu	1.0
A000255	NSND Hồng Vân	0.633247
A000493	Tuấn Trần	0.62257
A000179	Kiều Trinh	0.590355
A000177	Kiều Minh Tuấn	0.574042
A000378	Quách Ngọc Tuyên	0.571126
A000333	Ngọc Tường	0.567876
A000381	Quốc Cường	0.56579
A000186	Long Đẹp Trai	0.536807
A000373	Quang Tuấn	0.53286

Hình 11. Top 10 diễn viên theo chỉ số Eigenvector Centrality.



Hình 12. Trục quan mạng con của Top 10 diễn viên theo chỉ số Eigenvector Centrality, kích thước và màu sắc tỷ lệ với giá trị eigenvector.

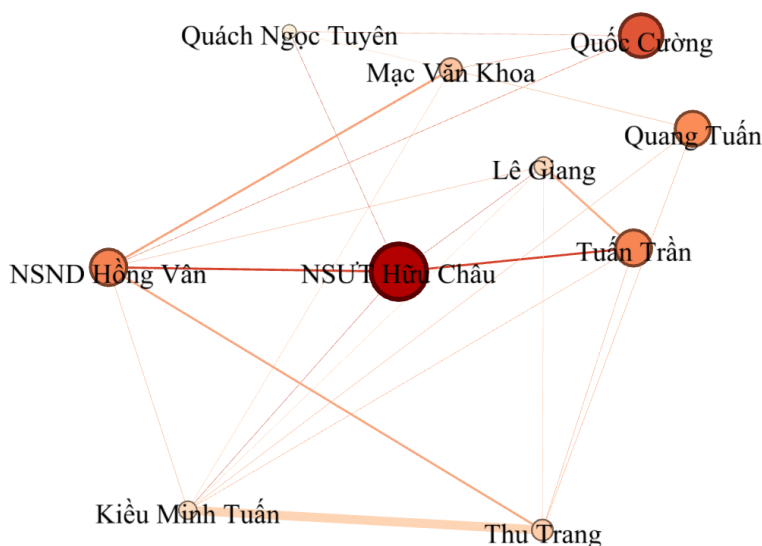
Kết quả cho thấy NSUT Hữu Châu đạt giá trị eigenvector cao nhất (1.0), tiếp theo là NSND Hồng Vân, Tuấn Trần, Kiều Trinh và Kiều Minh Tuấn. Điều này cho thấy các diễn viên này không chỉ có nhiều kết nối, mà còn liên kết trực tiếp với những diễn viên trung tâm khác, làm tăng tầm ảnh hưởng của họ trong mạng. Các diễn viên như Quách Ngọc Tuyên, Quốc Cường và Quang Tuấn cũng giữ vị trí đáng chú ý, minh họa cho sự quan trọng lan tỏa vượt ra ngoài số lượng kết nối đơn thuần.

So với Degree Centrality hay Betweenness Centrality, Eigenvector Centrality làm rõ hơn khía cạnh ảnh hưởng gián tiếp: một diễn viên có thể ít hợp tác hơn, nhưng nếu họ kết nối với các “ngôi sao” lớn thì họ vẫn trở thành nút quan trọng. Đây là cơ sở để tiếp tục phân tích chỉ số PageRank, một biến thể cải tiến dựa trên ý tưởng tương tự.

6) *PageRank*: Cuối cùng, chúng tôi xem xét chỉ số PageRank, một biến thể dựa trên ý tưởng của Eigenvector Centrality nhưng có bổ sung cơ chế “truyền ảnh hưởng” lặp lại. PageRank đánh giá một diễn viên không chỉ qua số lượng hay tầm quan trọng của các kết nối trực tiếp, mà còn cân nhắc khả năng “lan truyền sự chú ý” trong toàn bộ mạng, giống như cách thuật toán tìm kiếm của Google xếp hạng trang web.

Id	Label	PageRank ▼
A000274	NSUT Hữu Châu	0.009227
A000381	Quốc Cường	0.008087
A000255	NSND Hồng Vân	0.007492
A000493	Tuấn Trần	0.007462
A000373	Quang Tuấn	0.007377
A000249	Mạc Văn Khoa	0.006288
A000425	Thu Trang	0.006139
A000197	Lê Giang	0.005923
A000177	Kiều Minh Tuấn	0.005857
A000378	Quách Ngọc Tuyên	0.005482

Hình 13. Top 10 diễn viên theo chỉ số PageRank.



Hình 14. Trục quan mạng con của Top 10 diễn viên theo chỉ số PageRank, kích thước và màu sắc tỷ lệ với giá trị PageRank.

Kết quả cho thấy NSUT Hữu Châu tiếp tục đứng đầu, theo sau là Quốc Cường, NSND Hồng Vân, Tuấn Trần và Quang Tuấn. Điều này khẳng định vai trò then chốt của các diễn viên này không chỉ nhờ số lượng hay vị trí cầu nối, mà còn bởi sức ảnh hưởng lan tỏa mạnh mẽ qua mạng lưới hợp tác. Đáng chú ý, sự xuất hiện của Thu Trang và Lê Giang trong Top 10 PageRank cho thấy họ tuy không quá vượt trội ở các chỉ số khác, nhưng lại đạt ảnh hưởng ổn định trong mạng nhờ liên kết trực tiếp với các nút trung tâm.

Như vậy, sáu chỉ số trung tâm đã được phân tích toàn diện, từ số lượng kết nối, cường độ hợp tác, vai trò cầu nối, khả năng tiếp cận, đến sức ảnh hưởng lan tỏa. Phần tiếp theo sẽ tập trung vào phân tích cấu trúc cộng đồng bằng thuật toán Louvain và chỉ số Modularity, nhằm khám phá sự hình thành và đặc điểm của các nhóm hợp tác trong mạng.

7) *Tổng kết các chỉ số trung tâm:* Qua sáu chỉ số trung tâm đã phân tích (Degree, Weighted Degree, Betweenness, Closeness, Eigenvector và PageRank), có thể thấy rằng mỗi chỉ số làm nổi bật một khía cạnh khác nhau trong vai trò của các diễn viên. Một số diễn viên như NSUT Hữu Châu, NSND Hồng Vân và Quốc Cường xuất hiện lặp lại trong nhiều bảng xếp hạng, khẳng định vai trò hạt nhân cả về độ phủ kết nối, tần suất hợp tác, cũng như ảnh hưởng gián tiếp trong mạng. Bên cạnh đó, các chỉ số cũng cho thấy sự phân hóa chức năng: có diễn viên nổi bật về mức độ hợp tác rộng rãi, có diễn viên giữ vai trò cầu nối giữa các cụm, trong khi một số khác duy trì ảnh hưởng ổn định nhờ liên kết với các nút trung tâm.

Tuy nhiên, các chỉ số trung tâm chủ yếu tập trung vào vai trò của từng cá nhân. Để hiểu rõ hơn về cách các diễn viên liên kết thành các nhóm và cấu trúc cộng đồng, cần mở rộng phân tích sang cấp độ cao hơn. Do đó, phần tiếp theo sẽ tập trung vào việc phát hiện cộng đồng bằng thuật toán Louvain và đánh giá chất lượng phân cụm thông qua chỉ số modularity.

G. Phân tích cấu trúc cộng đồng

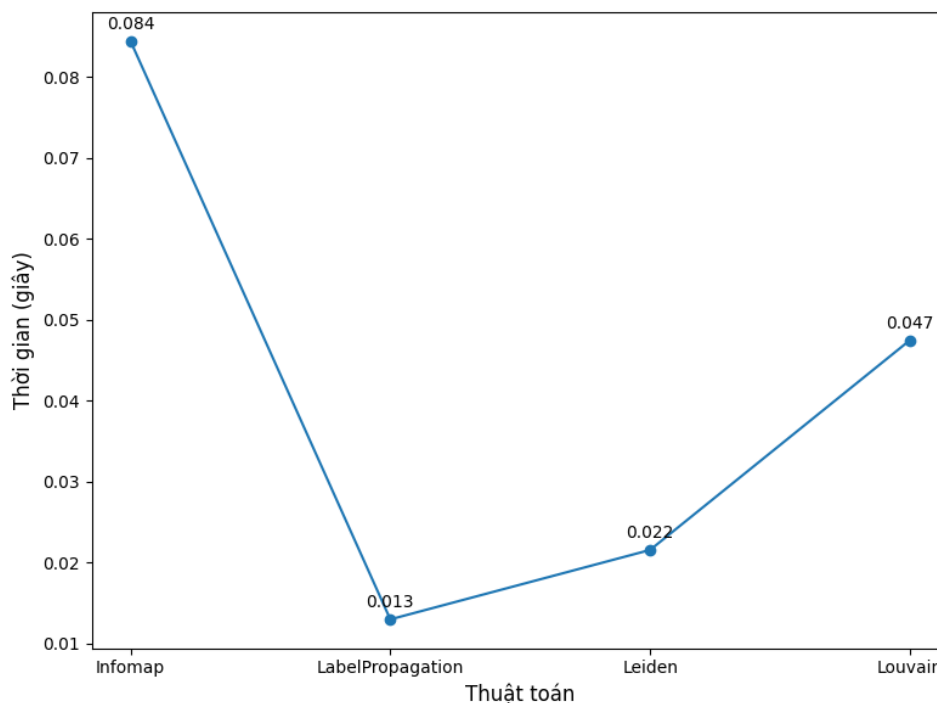
Sau khi phân tích các chỉ số trung tâm, chúng tôi tiếp tục áp dụng các thuật toán phát hiện cộng đồng nhằm khám phá cấu trúc tiềm ẩn trong mạng. Các thuật toán được so sánh bao gồm Louvain, Leiden, Infomap và Label Propagation. Do các thuật toán này đều có tính ngẫu nhiên trong quá trình tối ưu nên chúng tôi tiến hành chạy lặp lại 30 lần với các seed khác nhau, sau đó lựa chọn kết quả có giá trị modularity cao nhất để đảm bảo tính khách quan và ổn định của so sánh. Kết quả tổng hợp được trình bày trong Bảng IV.

Bảng IV
SO SÁNH CÁC THUẬT TOÁN PHÁT HIỆN CỘNG ĐỒNG (BEST OVER 30 RUNS)

Thuật toán	Modularity	Cộng đồng	Kích thước TB	Coverage_w	Conductance_avg
Louvain	0.6135	28	20.32	0.7080	0.1919
Leiden	0.6088	26	21.88	0.6954	0.1964
Infomap	0.5718	57	9.98	0.6137	0.3452
Label Propagation	0.4342	38	14.97	0.8291	0.2770

Kết quả trong Bảng IV cho thấy Louvain đạt modularity cao nhất (0.6135), phản ánh chất lượng phân cụm tốt nhất. Leiden có modularity gần tương đương (0.6088) nhưng tạo ra số cộng đồng ít hơn, với kích thước trung bình lớn hơn và cân bằng hơn, cho thấy tính ổn định cao. Infomap phát hiện nhiều cộng đồng hơn (57) nhưng modularity thấp (0.5718), dẫn đến phân mảnh mạng. Trong khi đó, Label Propagation có coverage cao (0.8291) nhưng modularity thấp (0.4342) và độ tách biệt cộng đồng chưa tốt, phản ánh tính ổn định kém. Nhìn chung, Louvain và Leiden là hai thuật toán phù hợp nhất để phân tích cấu trúc cộng đồng trong mạng.

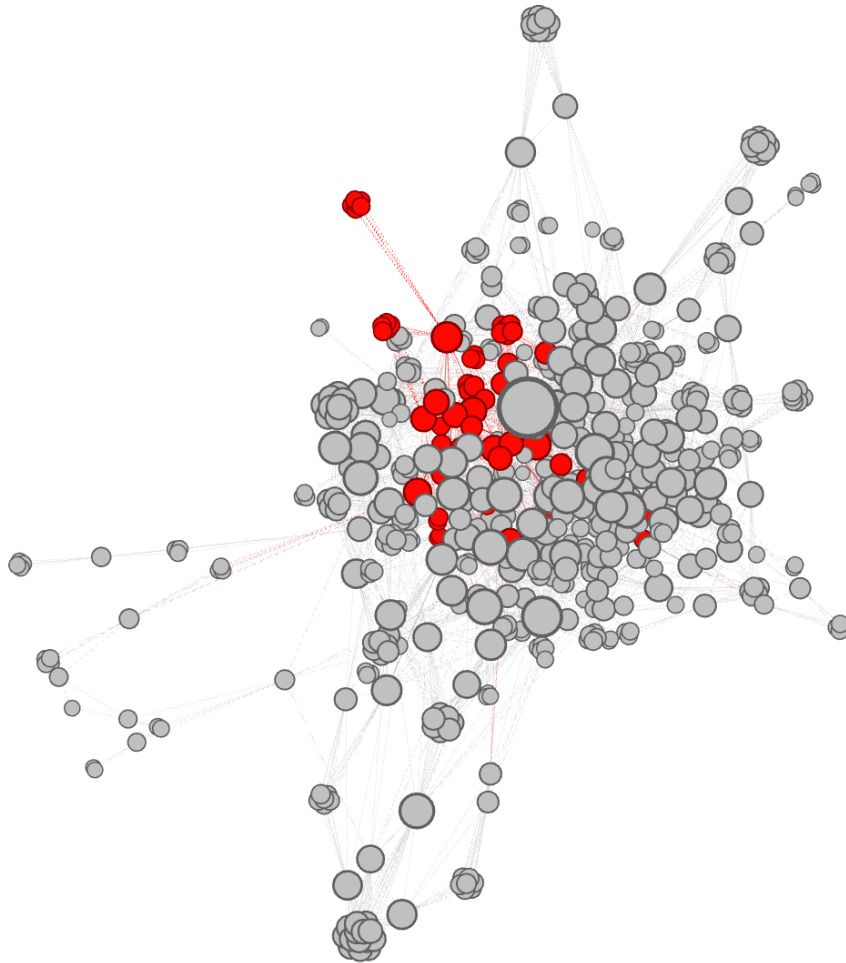
Tiếp theo, chúng tôi so sánh về mặt thời gian chạy của các thuật toán, kết quả được minh họa trong Hình 15.



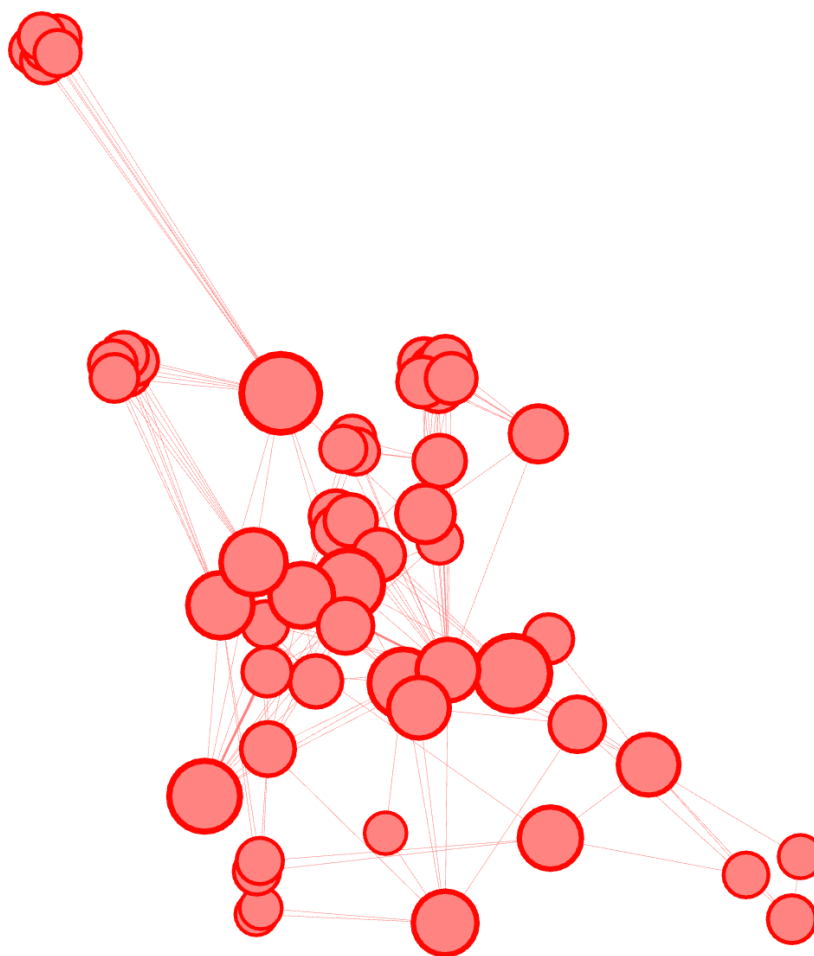
Hình 15. So sánh thời gian chạy của các thuật toán phát hiện cộng đồng

Hình 15 cho thấy Label Propagation có thời gian chạy ngắn nhất (0.013 giây), tiếp đến là Leiden (0.022 giây). Louvain mất nhiều thời gian hơn (0.047 giây) nhưng vẫn ở mức chấp nhận được so với lợi ích modularity cao. Ngược lại, Infomap chậm nhất (0.084 giây) trong khi chất lượng phân cụm không vượt trội, do đó ít phù hợp hơn. Điều này phản ánh rõ sự khác biệt về độ phức tạp tính toán và trade-off giữa chất lượng và tốc độ của từng thuật toán.

1) *Phân tích cộng đồng lớn nhất*: Dựa trên so sánh các thuật toán phát hiện cộng đồng, Louvain được lựa chọn để phân tích chi tiết hơn vì đạt modularity cao nhất và cho kết quả cộng đồng rõ rệt. Theo kết quả từ Louvain, mạng được phân tách thành 28 cộng đồng, trong đó cộng đồng lớn nhất có 75 diễn viên (chiếm khoảng 13,2% toàn mạng). Cộng đồng này được trực quan hóa ở hai mức: (i) được highlight trên toàn mạng để thấy vị trí và vai trò trung tâm (Hình 16); và (ii) được phóng to riêng để quan sát cấu trúc nội bộ (Hình 17).



Hình 16. Trực quan cộng đồng lớn nhất trên toàn bộ mạng.



Hình 17. Trực quan cấu trúc cộng đồng lớn nhất.

Để làm rõ vai trò từng diễn viên, bảng V liệt kê Top 10 diễn viên trong cộng đồng theo nhiều chỉ số trung tâm. Trong đó, NSND Hồng Vân nổi bật nhất với Degree = 56, Weighted Degree = 64, Betweenness rất cao (9689.5), và Eigenvector = 0.493, PageRank = 0.151. Điều này cho thấy Hồng Vân vừa là trung tâm liên kết nội bộ, vừa đóng vai trò cầu nối quan trọng trong mạng hợp tác. Các diễn viên khác như Trần Phong, Hoàng Mèo, Công Ninh hay Trần Nghĩa cũng nổi bật ở nhiều chỉ số, phản ánh vai trò nòng cốt trong duy trì sự gắn kết và lan tỏa trong cộng đồng.

Bảng V
TOP 10 DIỄN VIÊN CÓ CHỈ SỐ QUAN TRỌNG TRONG CỘNG ĐỒNG LỚN NHẤT

Id	Diễn viên	Degree	WDegree	Betweenness	Closeness	Eigenvector	PageRank
A000255	NSND Hồng Vân	56	64.0	9689.50	0.440	0.493	0.1506
A000477	Trần Phong	28	31.0	5031.06	0.356	0.398	0.1539
A000093	Hoàng Mèo	26	27.0	1217.60	0.400	0.441	0.4328
A000056	Công Ninh	23	24.0	5006.57	0.356	0.392	0.1898
A000475	Trần Nghĩa	22	23.0	1461.69	0.381	0.419	0.2710
A000055	Công Dương	21	21.0	1986.68	0.388	0.422	0.2257
A000262	NSND Thanh Nam	19	19.0	1440.52	0.356	0.390	0.1520
A000472	Trần Kim Hải	19	19.0	1437.14	0.353	0.388	0.1557
A000156	Jun Vũ	18	18.0	1942.50	0.371	0.404	0.1737
A000265	NSND Việt Anh	18	20.0	376.92	0.351	0.382	0.1589

Kết quả phân tích cho thấy mạng cộng tác diễn viên có cấu trúc cộng đồng rõ rệt, với cộng đồng lớn nhất giữ vai trò trung tâm, gắn kết chặt chẽ bên trong và liên kết mạnh ra ngoài thông qua các diễn viên hạt nhân. Những phát hiện này không chỉ

khẳng định đặc trưng phân cụm và tính chất thể giới nhỏ của mạng xã hội, mà còn gợi mở nhiều hướng khai thác tiếp theo về ảnh hưởng, lan tỏa và mối quan hệ giữa các nhóm. Trong phần tiếp theo chúng tôi sẽ trình bày thảo luận và kết luận chung, tổng hợp các kết quả đạt được và đề xuất hướng nghiên cứu mở rộng.

VI. THẢO LUẬN

Phần này tổng hợp kết quả nghiên cứu, làm rõ ý nghĩa của việc phân tích mạng lưới hợp tác giữa các diễn viên trong điện ảnh Việt Nam, gợi ý những hướng mở rộng cho các nghiên cứu tiếp theo

A. Các phát hiện chính

Phân tích mạng lưới của 563 diễn viên với hơn 3,400 mối quan hệ hợp tác cho thấy nhiều điểm nổi bật về cấu trúc và vai trò của từng người trong mạng

- **Nhóm trung tâm và nhóm xung quanh:** Các chỉ số như Degree, Weighted Degree, Betweenness, Eigenvector và PageRank cho thấy một nhóm nhỏ diễn viên đóng vai trò trung tâm, vừa có nhiều mối quan hệ hợp tác, vừa kết nối được giữa các nhóm nhỏ khác nhau [13] Trong đó, NSƯT Hữu Châu, NSND Hồng Vân, Quốc Cường và Quang Tuấn là những người nổi bật nhất. Không chỉ tham gia nhiều bộ phim mà còn giữ vị trí kết nối quan trọng, giúp mạng lưới trở nên gắn kết hơn
- **Tính cộng đồng rõ rệt:** Phân tích bằng các thuật toán Louvain và Leiden cho thấy mạng có cấu trúc cộng đồng khá rõ, phản ánh xu hướng các ê-kíp diễn viên quen thuộc thường xuyên hợp tác lặp lại trong nhiều dự án [6] Kết quả modularity cao (0.6135 đối với Louvain) cho thấy sự gắn bó giữa các nhóm, đồng thời vẫn có những cầu nối liên kết giữa các cộng đồng lớn
- **Hiệu ứng “thế giới nhỏ”:** Độ dài đường đi trung bình chỉ 3.16 và đường kính mạng là 7, nghĩa là hầu hết các diễn viên đều có thể kết nối với nhau chỉ qua vài bước trung gian [14] cho thấy trong môi trường điện ảnh Việt Nam, thông tin, xu hướng hay cơ hội hợp tác có thể lan tỏa rất nhanh giữa các cá nhân
- **Sự khác biệt về vai trò cá nhân:** Một số diễn viên có số lượng hợp tác không nhiều nhưng lại có vị trí đặc biệt trong việc kết nối giữa các nhóm khác nhau, thể hiện rõ ở các chỉ số như Betweenness hoặc PageRank cao, họ là “cầu nối” trong mạng. Đây là điểm khác biệt giữa mức độ “nổi tiếng” (số lượng hợp tác) và “tầm ảnh hưởng” (khả năng kết nối)

B. Ứng dụng thực tế

Kết quả nghiên cứu có thể mang lại nhiều ứng dụng hữu ích trong thực tiễn, đặc biệt là trong lĩnh vực sản xuất và quảng bá phim

- **Chiến lược quảng bá phim:** Các diễn viên trung tâm trong mạng thường có sức lan tỏa lớn, có thể được xem như “điểm tựa ảnh hưởng” giúp truyền thông phim đến nhiều nhóm khán giả khác nhau [15] Việc xác định các cá nhân này giúp nhà sản xuất tối ưu chiến lược quảng bá, tận dụng hiệu ứng lan truyền trong mạng lưới nghệ sĩ.
- **Tổ chức và sản xuất phim:** Việc phát hiện các nhóm cộng tác ổn định giúp các nhà sản xuất hiểu rõ hơn về các ê-kíp quen thuộc, từ đó lựa chọn diễn viên có sự phối hợp tốt, giảm rủi ro khi tuyển chọn diễn viên mới.
- **Phát triển hệ thống gợi ý:** Các chỉ số mạng có thể được dùng để xây dựng công cụ gợi ý diễn viên cho các dự án phim mới, dựa trên khả năng hợp tác hiệu quả hoặc những cặp diễn viên tiềm năng chưa từng làm việc chung.
- **Nghiên cứu truyền thông và hình ảnh cá nhân:** Phân tích mạng giúp hiểu rõ con đường lan tỏa thông tin, xác định ai là người có tầm ảnh hưởng lớn trong giới nghệ sĩ, từ đó hỗ trợ xây dựng thương hiệu cá nhân và chiến lược truyền thông hiệu quả.

C. Ý nghĩa của việc phân tích cộng đồng

Phân tích cộng đồng bằng Louvain cho thấy mạng điện ảnh Việt Nam vừa tập trung quanh một số gương mặt lớn, vừa có nhiều nhóm nhỏ hơn phản ánh các ê-kíp sản xuất quen thuộc Cộng đồng lớn nhất gồm 75 diễn viên (chiếm 13.2% toàn mạng), trong đó NSND Hồng Vân giữ vai trò trung tâm, vừa kết nối nội bộ chặt chẽ vừa liên kết ra ngoài. Trong môi trường điện ảnh, các mối quan hệ nghề nghiệp thường được hình thành dựa trên sự tin tưởng và quen biết lâu dài, tạo thành các cụm làm việc ổn định. Chính các “cầu nối” giữa các nhóm này lại giúp mang đến sự mới mẻ, giúp các xu hướng hoặc phong cách diễn xuất lan tỏa ra toàn mạng.

Bản đồ mạng cộng tác được xem như “bản đồ nghề nghiệp” của ngành điện ảnh Việt Nam, cho thấy cách các diễn viên kết nối, hợp tác và ảnh hưởng lẫn nhau trong thực tế

D. Điểm mạnh và hạn chế của nghiên cứu

Điểm mạnh:

- Dữ liệu thu thập thực tế, bao gồm hơn 500 diễn viên trong nhiều dự án phim khác nhau, khá đầy đủ mạng hợp tác hiện nay.
- Kết hợp giữa tính toán các chỉ số mạng và hình ảnh trực quan, giúp người đọc dễ hình dung và kiểm chứng.
- Sử dụng nhiều chỉ số trung tâm khác nhau để đánh giá vai trò của từng diễn viên, từ mức độ hợp tác, ảnh hưởng cho tới vị trí cầu nối.

Hạn chế:

- Dữ liệu hiện tại chỉ bao gồm phim điện ảnh, chưa có truyền hình hay phim trực tuyến.
- Phân tích chỉ phản ánh một thời điểm, chưa theo dõi sự thay đổi theo thời gian.
- Chưa xem xét yếu tố bên ngoài như doanh thu, giải thưởng hay tương tác truyền thông để đánh giá mức độ ảnh hưởng thực tế của từng diễn viên.
- Bộ dữ liệu vẫn còn trùng lặp ở một số tên nghệ sĩ hoặc nghệ danh khác nhau, quá trình xử lý phải thực hiện thủ công.

E. Hướng nghiên cứu tiếp theo

- **Mạng theo thời gian:** Phân tích sự thay đổi của mạng lưới hợp tác qua các giai đoạn, ví dụ trước và sau COVID-19 hoặc trong các thời kỳ phát triển của điện ảnh Việt Nam.
- **Mạng đa tầng:** Kết hợp dữ liệu từ phim chiếu rạp, truyền hình, âm nhạc và mạng xã hội để xây dựng bức tranh toàn diện hơn về ngành điện ảnh.
- **Ứng dụng học máy:** Dùng các đặc trưng mạng để huấn luyện mô hình dự đoán cặp diễn viên có khả năng hợp tác trong tương lai hoặc thành công khi cùng tham gia dự án.
- **Liên kết với dữ liệu thực tế:** Kết hợp thêm dữ liệu doanh thu, giải thưởng, và phản ứng khán giả để hiểu rõ hơn mối liên hệ giữa cấu trúc hợp tác và thành công thương mại.
- **So sánh liên ngành:** Áp dụng phương pháp tương tự cho sân khấu, âm nhạc hay truyền hình để so sánh mức độ hợp tác và lan tỏa trong các lĩnh vực khác.
- **Khắc phục chất lượng dữ liệu:** Hoàn thiện quy trình thu thập và xử lý dữ liệu theo hướng tự động, nhằm loại bỏ tình trạng trùng lặp và sai khác trong thông tin diễn viên, giúp bộ dữ liệu đầy đủ và chính xác hơn cho các nghiên cứu tiếp theo.

VII. KẾT LUẬN

Trong nghiên cứu này, chúng tôi xây dựng và phân tích mạng cộng tác giữa các diễn viên phim chiếu rạp Việt Nam giai đoạn 2020–nay, dựa trên dữ liệu được thu thập và xử lý cẩn thận, tạo nên một mạng gồm 563 nút và 3.405 cạnh. Kết quả cho thấy mạng mang đặc trưng của mạng xã hội phức tạp: mật độ tổng thể còn thưa nhưng hệ số gom cụm cao, đường đi trung bình ngắn, thể hiện rõ hiệu ứng “thế giới nhỏ” khi các diễn viên dễ dàng kết nối qua rất ít trung gian.

Việc áp dụng các chỉ số trung tâm (Degree, Weighted Degree, Betweenness, Closeness, Eigenvector, PageRank) giúp nhận diện những nghệ sĩ giữ vai trò quan trọng, vừa trong hợp tác trực tiếp, vừa trong việc kết nối giữa các cộng đồng. Đặc biệt, các nghệ sĩ gạo cội như NSƯT Hữu Châu, NSND Hồng Vân, Quốc Cường, Quang Tuấn hay Tuấn Trần được xác định là những nút trung tâm nổi bật, góp phần duy trì sự gắn kết và định hình cấu trúc hợp tác toàn mạng.

Những kết quả này mang lại giá trị thực tiễn: về lý thuyết, nghiên cứu chứng minh tính ứng dụng của khoa học mạng trong lĩnh vực văn hóa nghệ thuật; về thực tiễn, chúng có thể hỗ trợ nhà làm phim trong công tác casting, dự đoán khả năng hợp tác, thiết kế chiến lược quảng bá và định hướng phát triển thị trường điện ảnh. Bản đồ cộng tác cũng phản ánh quá trình vận động của ngành phim Việt Nam trong bối cảnh hội nhập, nơi sự kết nối giữa các thế hệ diễn viên tạo nên sản phẩm đa dạng và hấp dẫn.

Trong tương lai, nghiên cứu có thể mở rộng theo hướng phân tích mạng động để làm rõ sự biến đổi qua thời gian, kết hợp thêm dữ liệu về đạo diễn, biên kịch, nhà sản xuất để xây dựng mạng hợp tác đa chiều, hoặc áp dụng học máy nhằm dự đoán xu hướng hợp tác và gợi ý cộng tác mới. Qua đó, không chỉ giúp hiểu rõ hơn cơ chế hình thành và phát triển mạng cộng tác trong điện ảnh Việt Nam, mà còn cung cấp công cụ hỗ trợ cho việc hoạch định chiến lược trong ngành công nghiệp văn hóa và sáng tạo.

TÀI LIỆU

- [1] M. Newman, “The structure and function of networks,” *Computer Physics Communications*, vol. 147, no. 1, pp. 40–45, 2002, proceedings of the Europhysics Conference on Computational Physics Computational Modeling and Simulation of Complex Systems.
- [2] A.-L. Barabási, *Network Science*. Cambridge University Press, 2016.
- [3] J. Park and M. E. J. Newman, “A network-based analysis of the u.s. movie collaboration network,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 02, p. P02001, 2005.
- [4] S. Jalan, C. Sarkar, A. Madhusudanan, and S. Dwivedi, “Uncovering randomness and success in society,” *PloS one*, vol. 9, p. e88249, 06 2014.
- [5] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.
- [6] V. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” vol. 2008, 04 2008.
- [7] J. A. Bondy, U. S. R. Murty *et al.*, *Graph theory with applications*. Macmillan London, 1976, vol. 290.
- [8] V. Traag, L. Waltman, and N. J. van Eck, “From louvain to leiden: guaranteeing well-connected communities,” *Scientific Reports*, vol. 9, p. 5233, 03 2019.
- [9] M. Rosvall and C. T. Bergstrom, “Maps of random walks on complex networks reveal community structure,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1118–1123, 2008.
- [10] N. Raghavan, R. Albert, and S. Kumara, “Near linear time algorithm to detect community structures in large-scale networks,” *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 76, p. 036106, 10 2007.
- [11] M. Bastian, S. Heymann, and M. Jacomy, “Gephi: An open source software for exploring and manipulating networks,” *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 3, no. 1, pp. 361–362, Mar. 2009.
- [12] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, “Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software,” *PloS one*, vol. 9, p. e98679, 06 2014.
- [13] L. C. Freeman, “A set of measures of centrality based on betweenness,” *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.
- [14] M. E. J. Newman, *Networks: An Introduction*. Oxford University Press, 2010.
- [15] L. Page, S. Brin, R. Motwani, and T. Winograd, “The pagerank citation ranking: Bringing order to the web,” in *Technical Report, Stanford InfoLab*, 1999.