**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

A blue logo with a planet and a flower

Description automatically generated**---------□&□---------**

**ĐỒ ÁN MÔN HỌC MẠNG XÃ HỘI**

**ĐỀ TÀI NGHIÊN CỨU**

**SOCIAL COMMUNITY DETECTION SCHEME BASED ON SOCIAL-AWARE IN MOBILE SOCIAL NETWORKS**

*Giảng viên hướng dẫn:* **ThS. Thái Bảo Trân**

*Nhóm sinh viên thực hiện:* **Nhóm 13 - IS353.O12.HTCL**

**Trần Trọng Nghĩa** 20521657

**Nguyễn Phương Yến Nhi** 20521715

**Phan Công Thành** 20521923

*Thành phố Hồ Chí Minh, tháng 11 năm 2023*

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

---------------------------------------------

**ĐỒ ÁN MÔN HỌC MẠNG XÃ HỘI**

**ĐỀ TÀI: NGHIÊN CỨU CHỦ ĐỀ**

**SOCIAL COMMUNITY DETECTION SCHEME BASED ON SOCIAL-AWARE IN MOBILE SOCIAL NETWORKS**

*Thành phố Hồ Chí Minh, tháng 11 năm 2023*

**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

# LỜI MỞ ĐẦU

Trong nghiên cứu này, chúng em giới thiệu một phương pháp khám phá cộng đồng xã hội mới dựa trên ý thức về mạng xã hội trong mạng xã hội di động. Phương pháp của chúng em dựa một bài báo nghiên cứu được đăng trên IEEE - Hội Kỹ sư Điện và Điện tử . Chúng em tiếp cận vấn đề này từ góc độ của việc hiểu biết sâu hơn về các mối quan hệ xã hội, quan điểm và hành vi người dùng trong mạng xã hội di động.

Lời đầu tiên chúng em xin chân thành gửi lời cảm ơn đến với giảng viên môn học là cô Thái Bảo Trân. Cảm ơn cô đã giảng dạy, truyền đạt rất nhiều kiến thức bổ ích về mạng xã hội, những kiến thức này vô cùng cần thiết cho chúng em thực hiện đồ án. Cảm ơn cô đã hướng dẫn nhiệt tình, giải đáp thắc mắc cũng như hỗ trợ chúng em rất nhiều trong quá trình thực hiện đồ án.

Với khả năng và thời gian có hạn nên không thể tránh khỏi những thiếu sót, chúng em rất mong được sự quan tâm, giúp đỡ và thông cảm của cô để chúng em hoàn thiện hơn về đồ án của nhóm mình. Một lần nữa, chúng em xin chân thành cảm ơn cô!

Thành phố Hồ Chí Minh, tháng 11 năm 2023

Nhóm sinh viên thực hiện

Nhóm 13

# BẢNG PHÂN CÔNG, ĐÁNH GIÁ THÀNH VIÊN

|  |  |  |  |
| --- | --- | --- | --- |
| MSSV | Họ tên | Phân công | Đánh giá |
| 20521657 | Trần Trọng Nghĩa | Tìm hiểu bài báo khoa học và thuật toán khám phá cộng đồng  Phân tích đề tài nghiên cứu và cách triển khai mô phỏng thuật toán  Tìm tài liệu thực nghiệm cho mô phỏng thuật toán  Thực hiện phần tài liệu về phương pháp nghiên cứu, các quy trình, chỉ số thuật toán và các thuộc tính trong khám phá cộng đồng  Thực hiện tài liệu báo cáo , nội dung thuyết trình | Hoàn thành 100% |
| 20521715 | Nguyễn Phương Yến Nhi | Tìm hiểu bài báo khoa học và thuật toán khám phá cộng đồng  Phân tích đề tài nghiên cứu và cách triển khai mô phỏng thuật toán  Tìm tài liệu thực nghiệm cho mô phỏng thuật toán  Thực hiện phần tài liệu về giới thiệu đề tài, nội dung nghiên cứu, các cơ sở lý thuyết và nghiên cứu liên quan.  Thực hiện tài liệu báo cáo , nội dung thuyết trình | Hoàn thành 100% |
| 20521923 | Phan Công Thành | Tìm hiểu bài báo khoa học và thuật toán khám phá cộng đồng  Phân tích đề tài nghiên cứu và cách triển khai mô phỏng thuật toán.  Tìm nguồn dữ liệu và lập trình mã nguồn mô phỏng thuật toán nghiên cứu.  Thực hiện tài liệu báo cáo, video mô phỏng. | Hoàn thành 100% |

Link drive: [Team13\_IS353.O12.HTCL\_ĐO\_AN\_CK](https://drive.google.com/drive/folders/1Y4Ecw6VW6X62JmJy5Byd7mCEQDM4mEwJ?fbclid=IwAR2SifXYJRYd16DqzARcfXKJT8VVyjUyWtJhbyUVh_fDELY4l1Ier0TBPrE)

# MỤC LỤC

[**LỜI MỞ ĐẦU 4**](#_Toc155019594)

[**BẢNG PHÂN CÔNG, ĐÁNH GIÁ THÀNH VIÊN 5**](#_Toc155019595)

[**MỤC LỤC 6**](#_Toc155019596)

[**DANH MỤC HÌNH ẢNH 9**](#_Toc155019597)

[**DANH MỤC BẢNG 10**](#_Toc155019598)

[**CHƯƠNG 1: GIỚI THIỆU ĐỀ TÀI 11**](#_Toc155019599)

[**1.1 Thông tin tác giả 11**](#_Toc155019600)

[**1.2 Thông tin bài báo 11**](#_Toc155019601)

[**1.3 Tổng quan nội dung 11**](#_Toc155019602)

[**CHƯƠNG 2: ĐẶT VẤN ĐỀ 12**](#_Toc155019603)

[**2.1 Tình hình chung 12**](#_Toc155019604)

[**2.2 Mạng xã hội di động tại Việt Nam 12**](#_Toc155019605)

[**2.3 Nhu cầu nghiên cứu 13**](#_Toc155019606)

[**CHƯƠNG 3: NỘI DUNG NGHIÊN CỨU 14**](#_Toc155019607)

[**3.1 Các thuật ngữ liên quan 14**](#_Toc155019608)

[**3.2 Mục tiêu nghiên cứu 14**](#_Toc155019609)

[**3.3 Phương pháp nghiên cứu: 15**](#_Toc155019610)

[**3.4 Dữ liệu nghiên cứu: 15**](#_Toc155019611)

[**3.5 Đánh giá nghiên cứu: 16**](#_Toc155019612)

[**CHƯƠNG 4: CƠ SỞ LÝ THUYẾT 17**](#_Toc155019613)

[**4.1 Lý thuyết về Nodes và Edges 17**](#_Toc155019614)

[**4.1.1 Thế nào là Nodes và Edges 17**](#_Toc155019615)

[**4.1.2 Edges Direction 17**](#_Toc155019616)

[**4.1.3 Edge Weight 18**](#_Toc155019617)

[**4.2 Lý thuyết độ đo trong mạng xã hội 18**](#_Toc155019618)

[**4.2.1 Degree Centrality 18**](#_Toc155019619)

[**4.2.2 Betweenness Centrality 18**](#_Toc155019620)

[**4.2.3 Closeness Centrality 19**](#_Toc155019621)

[**4.3 Lý thuyết thuận toán phát hiện cộng đồng 19**](#_Toc155019622)

[**4.3.1 Thuật toán Girvan Newman 19**](#_Toc155019623)

[**4.3.2 Thuật toán phân cụm 20**](#_Toc155019624)

[**4.3.3Thuật toán Newman Modularity 21**](#_Toc155019625)

[**CHƯƠNG 5: CÁC NGHIÊN CỨU LIÊN QUAN 23**](#_Toc155019626)

[**5.1 Khám phá cộng đồng trong mạng xã hội 23**](#_Toc155019627)

[**5.2 Khám phá cộng đồng trong mạng xã hội di động 24**](#_Toc155019628)

[**CHƯƠNG 6: KHÁM PHÁ CỘNG ĐỘNG 26**](#_Toc155019629)

[**6.1 Các định nghĩa liên quan 26**](#_Toc155019630)

[**6.2 Đề xuất giao thức khám phá cộng đồng xã hội 29**](#_Toc155019631)

[**1) TƯƠNG ĐỒNG THUỘC TÍNH CỘNG ĐỒNG 29**](#_Toc155019632)

[**2) TƯƠNG ĐỒNG SỞ THÍCH NGƯỜI DÙNG 30**](#_Toc155019633)

[**3) TRỌNG SỐ CẠNH 31**](#_Toc155019634)

[**4) GÓC ĐỘ MÔ ĐUN CỘNG ĐỒNG 32**](#_Toc155019635)

[**5) MÔ HÌNH DỰA TRÊN MỨC ĐỘ MÔ-ĐUN 33**](#_Toc155019636)

[**CHƯƠNG 7: THU THẬP VÀ PHÂN TÍCH DỮ LIỆU 40**](#_Toc155019637)

[**CHƯƠNG 8: THỰC NGHIỆM KHÁM PHÁ CỘNG ĐỒNG 41**](#_Toc155019638)

[**CHƯƠNG 9: KẾT QUẢ THỰC NGHIỆM 42**](#_Toc155019639)

[**CHƯƠNG 10: MÔ PHỎNG NGHIÊN CỨU 46**](#_Toc155019640)

[**10.1 Mô tả dữ liệu 46**](#_Toc155019641)

[**10.1.1 Giới thiệu nguồn dữ liệu 46**](#_Toc155019642)

[**10.1.2 Mô tả dữ liệu 46**](#_Toc155019643)

[**10.1.3 Tiền xử lý dữ liệu 47**](#_Toc155019644)

[**10.2 Độ đo trung tâm và trực quan hóa 47**](#_Toc155019645)

[**10.2.1 Degree Centrality 47**](#_Toc155019646)

[**10.2.2 Betweeness Centrality 50**](#_Toc155019647)

[**10.2.3 Closeness Centrality 52**](#_Toc155019648)

[**10.3 Phát hiện cộng đồng và nhận xét 54**](#_Toc155019649)

[**10.3.1 Girvan Newman 54**](#_Toc155019650)

[**10.3.2 Newman Modularity 55**](#_Toc155019651)

[**10.3.3 Thuật toán của bài báo 56**](#_Toc155019652)

[**CHƯƠNG 10: ĐÁNH GIÁ PHƯƠNG PHÁP 59**](#_Toc155019653)

[**10.1 Ưu điểm của phương pháp 59**](#_Toc155019654)

[**10.2 Nhược điểm của phương pháp 59**](#_Toc155019655)

[**10.3 Kết luận 60**](#_Toc155019656)

[**CHƯƠNG 11: TÀI LIỆU THAM KHẢO 62**](#_Toc155019657)

# DANH MỤC HÌNH ẢNH

[Hình 1 Minh họa nodes và edges 18](https://d.docs.live.net/6233bfe8baf8921b/Máy%20tính/ĐO_AN_CK.docx#_Toc155020102)

[Hình 2 Minh họa thuật toán phân cụm 22](#_Toc155020103)

[Hình 3 Kết quả thực nghiệm 1 35](#_Toc155020104)

[Hình 4 Bảng tập dữ liệu của bài báo 41](#_Toc155020105)

[Hình 5 Kết quả thực nghiệm 2 43](#_Toc155020106)

[Hình 6 Kết quả thực nghiệm 3 44](#_Toc155020107)

[Hình 7 Kết quả thực nghiệm 4 45](https://d.docs.live.net/6233bfe8baf8921b/Máy%20tính/ĐO_AN_CK.docx#_Toc155020108)

[Hình 8 Bảng dữ liệu 47](#_Toc155020109)

[Hình 9 Tiền xử lý dữ liệu 48](#_Toc155020110)

[Hình 10 Kết quả xử lý dữ liệu 48](#_Toc155020111)

[Hình 11 Trực quan độ đo Degree Centrality 49](#_Toc155020112)

[Hình 12 Top 10 node Degree cao nhất 50](#_Toc155020113)

[Hình 13 Tần số node Degree xuất hiện nhiều nhất 50](#_Toc155020114)

[Hình 14 Trực quan độ đo Betweeness Centrality 51](#_Toc155020115)

[Hình 15 Top 10 node Betweeness cao nhất 52](#_Toc155020116)

[Hình 16 Tần số node Betweeness xuất hiện nhiều nhất 52](#_Toc155020117)

[Hình 17 Trực quan độ đo Closeness Centrality 53](#_Toc155020118)

[Hình 18 Top 10 node Closeness cao nhất 54](#_Toc155020119)

[Hình 19 Tần số node Closeness xuất hiện nhiều nhất 54](#_Toc155020120)

[Hình 20 Kết quả cộng đồng Girvan Newman 56](#_Toc155020121)

[Hình 21 Kết quả cộng đồng Newman Modularity 57](#_Toc155020122)

[Hình 22 Kết quả cộng đồng di động 59](#_Toc155020123)

# CHƯƠNG 1: GIỚI THIỆU ĐỀ TÀI

Nhóm thực hiện nghiên cứu đề tài **“Khám phá cộng đồng xã hội dựa trên các yếu tố xã hội trong mạng xã hội di động”**  dựa trên bài báo nghiên cứu khoa học “Social Community Detection Scheme Based on Social-Aware in Mobile Social Networks”

Truy cập bài báo: <https://ieeexplore.ieee.org/document/8915780>

## 1.1 Thông tin tác giả

Bài viết được viết bởi Ke Gu, Dianxing Liu và Keming Wang, những nhà nghiên cứu hàng đầu trong lĩnh vực mạng xã hội di động. Họ đến từ Trường Kỹ thuật Máy tính và Truyền thông, Đại học Khoa học và Công nghệ Changsha, Trung Quốc. Họ là thành viên của IEEE và đã có nhiều công trình nghiên cứu quan trọng về mạng xã hội di động.

## 1.2 Thông tin bài báo

Bài viết có tựa đề “Social Community Detection Scheme Based on Social-Aware in Mobile Social Networks” và được công bố trên tạp chí IEEE Access vào ngày 27 tháng 11 năm 2019. Bài viết là kết quả của một dự án nghiên cứu lâu dài, trong đó các tác giả đã dành rất nhiều thời gian và công sức để phát triển và kiểm tra thuật toán khám phá cộng đồng xã hội mới cho mạng xã hội di động.

## 1.3 Tổng quan nội dung

Bài viết trình bày về việc Khám phá cộng đồng trong mạng xã hội di động dựa trên yếu tố xã hội. Các tác giả đã sử dụng ba yếu tố chính: sự tương đồng thuộc tính xã hội, sự tương đồng quan tâm của nút và di chuyển của nút để xây dựng một thuật toán khám phá cộng đồng mới.

[Kết quả thử nghiệm cho thấy thuật toán có thể khám phá cộng đồng một cách chính xác và hiệu quả](https://edgeservices.bing.com/edgesvc/chat?udsframed=1&form=SHORUN&clientscopes=chat,noheader,udsedgeshop,channelstable,&shellsig=d7310b8b5e4b67387a1320627cb72fb4f7b52c46&setlang=en-US&darkschemeovr=1#sjevt%7CDiscover.Chat.SydneyClickPageCitation%7Cadpclick%7C1%7C4000ee49-be79-47f1-bf62-1db1cd85166c%7C%7B%22sourceAttributions%22%3A%7B%22providerDisplayName%22%3A%221.%20Do%20%C4%91%C3%B3%2C%20...%22%2C%22pageType%22%3A%22pdf%22%2C%22pageIndex%22%3A1%2C%22relatedPageUrl%22%3A%22file%253A%252F%252F%252FD%253A%252FUIT%252FHKVII%252FMXH%252FTV.pdf%22%2C%22lineIndex%22%3A29%2C%22highlightText%22%3A%221.%20Do%20%C4%91%C3%B3%2C%20s%C6%A1%20%C4%91%E1%BB%93%20%C4%91%E1%BB%81%20xu%E1%BA%A5t%20c%E1%BB%A7a%20ch%C3%BAng%20t%C3%B4i%20c%C3%B3%20th%E1%BB%83%20ph%C3%A1t%20hi%E1%BB%87n%20c%E1%BB%99ng%20%C4%91%E1%BB%93ng%20ch%C3%ADnh%20x%C3%A1c%20v%C3%A0%20hi%E1%BB%87u%20qu%E1%BA%A3%20h%C6%A1n%20nh%E1%BA%B1m%20t%C4%83ng%20t%C3%ADnh%20%E1%BB%95n%20%C4%91%E1%BB%8Bnh%20c%E1%BB%A7a%20c%E1%BA%A5u%22%2C%22snippets%22%3A%5B%5D%7D%7D). Cụ thể, thuật toán có thể giảm số lượng và kích thước của các cộng đồng quy mô lớn để tăng tính ổn định của cấu trúc cộng đồng. [Ngoài ra, thuật toán có thể điều chỉnh tỷ lệ ảnh hưởng của các yếu tố xã hội để khám phá các cộng đồng chi tiết hơn](https://edgeservices.bing.com/edgesvc/chat?udsframed=1&form=SHORUN&clientscopes=chat,noheader,udsedgeshop,channelstable,&shellsig=d7310b8b5e4b67387a1320627cb72fb4f7b52c46&setlang=en-US&darkschemeovr=1#sjevt%7CDiscover.Chat.SydneyClickPageCitation%7Cadpclick%7C2%7C4000ee49-be79-47f1-bf62-1db1cd85166c%7C%7B%22sourceAttributions%22%3A%7B%22providerDisplayName%22%3A%22T%C3%93M%20T%E1%BA%AET%20Hi...%22%2C%22pageType%22%3A%22pdf%22%2C%22pageIndex%22%3A1%2C%22relatedPageUrl%22%3A%22file%253A%252F%252F%252FD%253A%252FUIT%252FHKVII%252FMXH%252FTV.pdf%22%2C%22lineIndex%22%3A18%2C%22highlightText%22%3A%22T%C3%93M%20T%E1%BA%AET%20Hi%E1%BB%87n%20t%E1%BA%A1i%20c%C3%B3%20nhi%E1%BB%81u%20y%E1%BA%BFu%20t%E1%BB%91%20c%C3%B3%20th%E1%BB%83%20%E1%BA%A3nh%20h%C6%B0%E1%BB%9Fng%20%C4%91%E1%BA%BFn%20vi%E1%BB%87c%20ph%C3%A1t%20hi%E1%BB%87n%20c%E1%BB%99ng%20%C4%91%E1%BB%93ng%20trong%20c%C3%A1c%20m%E1%BA%A1ng%20x%C3%A3%20h%E1%BB%99i%20di%20%C4%91%E1%BB%99ng%2C%20trong%20%C4%91%C3%B3%20t%C3%ADnh%22%2C%22snippets%22%3A%5B%5D%7D%7D). Thuật toán cũng có thể hoạt động hiệu quả với các mạng có khả năng di chuyển cao của người dùng.

# CHƯƠNG 2: ĐẶT VẤN ĐỀ

## 2.1 Tình hình chung

Với sự phát triển nhanh chóng của công nghệ mạng, cuộc sống của con người đã bước vào một kỷ nguyên mới của mạng xã hội. Trong mọi khía cạnh của đời sống con người, mạng xã hội có thể được nhìn thấy ở khắp mọi nơi. Dưới đây là một số thông tin về tình hình chung của mạng xã hội di động trên toàn cầu vào năm 2023:

* [Số người sử dụng mạng xã hội trên toàn thế giới đã tăng lên 4,80 tỷ người, với 150 triệu người dùng mới đã truy cập trong 12 tháng qua](https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/).
* [Thời gian trung bình hàng ngày mà người dùng dành cho mạng xã hội là 2 giờ 24 phút](https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/).
* [Tỉ lệ thâm nhập mạng xã hội trên toàn cầu đứng ở mức 59,4% vào đầu năm 2023](https://www.statista.com/statistics/282846/regular-social-networking-usage-penetration-worldwide-by-country/).
* [Các nền tảng mạng xã hội phổ biến nhất trên toàn cầu vào tháng 7 năm 2023, được xếp hạng theo số lượng người dùng hàng tháng (tính bằng triệu) bao gồm Facebook, YouTube, WhatsApp, Instagram, WeChat, TikTok, Facebook Messenger, Snapchat, Douyin (TikTok tại Trung Quốc), Telegram, Kuaishou (một nền tảng chia sẻ video phổ biến tại Trung Quốc), QQ (một dịch vụ nhắn tin tức thì phổ biến tại Trung Quốc), Sina Weibo (một nền tảng mạng xã hội phổ biến tại Trung Quốc), Twitter và Pinterest3](https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/).
* [Các nền tảng mạng xã hội có tốc độ phát triển nhanh nhất vào năm 2023 bao gồm BeReal và Twitch](https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/).

## 2.2 Mạng xã hội di động tại Việt Nam

Theo các báo cáo mới nhất, mạng xã hội di động đang phát triển mạnh mẽ tại Việt Nam. Dưới đây là một số thông tin chi tiết:

* Đầu năm 2023, Việt Nam có 77,93 triệu người dùng Internet, chiếm 79,1% tổng dân số. Ngoài ra, số lượng người dùng mạng xã hội cũng đạt con số 70 triệu, tương đương với 71% tổng dân số.
* Tổng số kết nối di động đang hoạt động là 161,6 triệu, tương đương với 164,0% tổng dân số1. Điều này cho thấy Việt Nam có tổng số người dùng Internet và mạng xã hội đáng kể, cùng với số lượng kết nối di động vượt quá tổng dân số.
* Mạng xã hội cũng trở thành nền tảng quan trọng với hơn 70 triệu người tham gia. Điều này đem lại cơ hội và thách thức cho doanh nghiệp trong môi trường số hóa.
* Theo kết quả khảo sát, Youtube với tính năng hỗ trợ xem và chia sẻ video hiện là trang mạng có lượng người dùng lớn thứ hai ở Việt Nam sau Facebook (56,3%); đứng thứ ba là Instagram (24,5%) chuyên xem và chia sẻ ảnh; Zingme (16,8%) hỗ trợ chơi game, nghe nhạc trực tuyến; các mạng Viber, Zalo chiếm tỷ lệ 10%.

## 2.3 Nhu cầu nghiên cứu

Với sự phát triển không ngừng của mạng xã hội di động, việc hiểu rõ về cấu trúc và hoạt động của các cộng đồng trực tuyến trở nên cực kỳ quan trọng. Bên cạnh đó, việc nghiên cứu và phát triển các phương pháp Khám phá cộng đồng xã hội mới trong mạng xã hội di động trở nên vô cùng cần thiết.

Các phương pháp này không chỉ giúp tăng cường sự hiểu biết về cấu trúc và hoạt động của các cộng đồng trực tuyến, mà còn có thể được sử dụng để Khám phá và ngăn chặn các hoạt động không mong muốn hoặc có hại trên mạng xã hội.

# CHƯƠNG 3: NỘI DUNG NGHIÊN CỨU

## 3.1 Các thuật ngữ liên quan

Dưới đây là một số lý thuyết, thuật ngữ liên quan đến bài báo nghiên cứu khoa học:

* **Mạng xã hội** là những cấu trúc phức tạp, trong đó các nút (người dùng) và các cạnh (mối quan hệ) có thể hình thành các nhóm tự nhiên, được gọi là cộng đồng. Khám phá cộng đồng là một vấn đề quan trọng trong nghiên cứu mạng xã hội, vì nó có thể giúp khám phá các đơn vị chức năng, mối quan hệ và đặc điểm ẩn của mạng.
* **Mạng xã hội di động (Mobile social network)**: Là một loại mạng xã hội mà các thành viên có thể kết nối và tương tác với nhau thông qua các thiết bị di động, như điện thoại thông minh, máy tính bảng, v.v. Các mạng xã hội di động có thể hỗ trợ các dịch vụ như chia sẻ tin tức, hình ảnh, video, âm thanh, vị trí, v.v.
* **Khám phá cộng đồng (Community detection)**: Là một quá trình phân loại các nút (thành viên) trong một mạng xã hội thành các nhóm có liên quan đến nhau dựa trên một số tiêu chí, như thuộc tính xã hội, sở thích, hoạt động, v.v. Mục tiêu của Khám phá cộng đồng là khám phá ra cấu trúc và đặc điểm của mạng xã hội.
* **Đặc trưng xã hội (Social-aware):** Là một khái niệm chỉ sự nhận thức và quan tâm đến các yếu tố xã hội trong một mạng xã hội. Các yếu tố xã hội có thể bao gồm: đồng cảm, niềm tin, tương tác, ảnh hưởng, v.v. Đặc trưng xã hội có thể ảnh hưởng đến hành vi và quan hệ của các thành viên trong mạng xã hội.
* **Độ di chuyển của nút (Node mobility)**: Là một thuộc tính chỉ sự thay đổi vị trí của các nút trong một mạng xã hội di động theo thời gian. Độ di chuyển của nút có thể phản ánh sự linh hoạt và đa dạng của các thành viên trong mạng xã hội di động. Độ di chuyển của nút cũng có thể ảnh hưởng đến sự ổn định và hiệu quả của Khám phá cộng đồng.

## 3.2 Mục tiêu nghiên cứu

Mục tiêu chính của bài nghiên cứu là đề xuất một hệ thống Khám phá cộng đồng xã hội mới cho mạng xã hội di động dựa trên yếu tố xã hội, bao gồm sự tương đồng thuộc tính xã hội, sự tương đồng quan tâm của nút và di chuyển của nút.

[Các tác giả muốn phát triển một thuật toán có thể Khám phá cộng đồng một cách chính xác và hiệu quả trong mạng xã hội di động, dựa trên sự tương đồng thuộc tính xã hội, sự tương đồng quan tâm của nút và di chuyển của nút](https://edgeservices.bing.com/edgesvc/chat?udsframed=1&form=SHORUN&clientscopes=chat,noheader,udsedgeshop,channelstable,&shellsig=d7310b8b5e4b67387a1320627cb72fb4f7b52c46&setlang=en-US&darkschemeovr=1#sjevt%7CDiscover.Chat.SydneyClickPageCitation%7Cadpclick%7C0%7C4000ee49-be79-47f1-bf62-1db1cd85166c%7C%7B%22sourceAttributions%22%3A%7B%22providerDisplayName%22%3A%22t%E1%BB%B1%20thu%E1%BB%99c%20t...%22%2C%22pageType%22%3A%22pdf%22%2C%22pageIndex%22%3A2%2C%22relatedPageUrl%22%3A%22file%253A%252F%252F%252FD%253A%252FUIT%252FHKVII%252FMXH%252FTV.pdf%22%2C%22lineIndex%22%3A25%2C%22highlightText%22%3A%22t%E1%BB%B1%20thu%E1%BB%99c%20t%C3%ADnh%20x%C3%A3%20h%E1%BB%99i%2C%20%C4%91%E1%BB%99%20t%C6%B0%C6%A1ng%20t%E1%BB%B1%20s%E1%BB%9F%20th%C3%ADch%20c%E1%BB%A7a%20n%C3%BAt%20v%C3%A0%20t%C3%ADnh%20di%20%C4%91%E1%BB%99ng%20c%E1%BB%A7a%20n%C3%BAt.%22%2C%22snippets%22%3A%5B%5D%7D%7D).

## 3.3 Phương pháp nghiên cứu:

* **Dựa trên các yếu tố xã hội: P**hương pháp này khám phá các yếu tố xã hội, bao gồm sự giống nhau về thuộc tính xã hội, sự giống nhau về sở thích và khả năng di chuyển của các nút.
* **Áp dụng mô hình toán học:** Mô hình này được áp dụng để xác định mức độ tương đồng xã hội giữa các nút trong mạng xã hội di động. Mức độ tương đồng xã hội bao gồm ba yếu tố chính: đồng cảm, niềm tin và tương tác.
* **Sử dụng thuật toán phân cụm:** Thuật toán này được áp dụng để phân nhóm các nút có mức độ tương đồng xã hội cao vào các cộng đồng xã hội.
* **Tiếp cận phương pháp:** Đề xuất đầu tiên của bài báo này là đo lường các thuộc tính xã hội của mỗi nút, sau đó tính toán mức độ tương đồng về thuộc tính xã hội và sở thích giữa các nút. Dựa trên sự tương đồng về mặt xã hội, một phương pháp được xây dựng để tính toán bậc của mô-đun. Cuối cùng, đề xuất này phát hiện các cộng đồng trong mạng xã hội di động dựa trên bậc mô-đun đã được tính toán.
* **Xem xét sự di động:** Trong đề xuất của bài báo này, sự di động được coi là một yếu tố ảnh hưởng để tính toán trọng số của mức độ tương đồng cạnh. Điều này khác biệt so với các kế hoạch liên quan, khi tập trung vào các thuộc tính xã hội của nút và sự di động của nó. Hơn nữa, thông qua việc điều chỉnh yếu tố ảnh hưởng trong khoảng từ 0 đến 1, tỷ lệ của mức độ tương đồng về thuộc tính xã hội và sở thích của nút được kiểm soát.

## 3.4 Dữ liệu nghiên cứu:

Sử dụng hai bộ dữ liệu thực tế là ***Reality Mining*** và ***MIT Social Evolution*** để kiểm tra hiệu quả của phương pháp.  Các bộ dữ liệu này ghi lại các thông tin về vị trí, cuộc gọi, tin nhắn và sự kiện của các thành viên trong mạng xã hội di động.

## 3.5 Đánh giá nghiên cứu:

Bài báo đã dùng hai phương pháp để đánh giá kết quả thuật toán Khám phá cộng đồng xã hội dựa trên đặc trưng xã hội trong mạng xã hội di động.

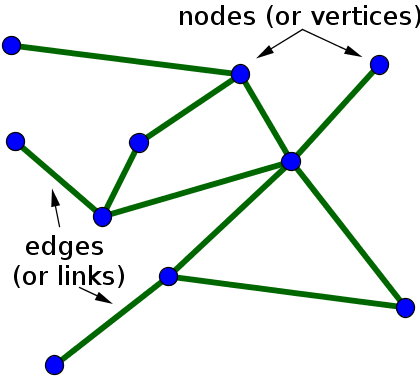
* **So sánh với các phương pháp khác**: Bài báo đã so sánh hiệu quả của phương pháp được đề xuất với các phương pháp khác, như K-means, DBSCAN, Louvain, GN và NM. Các tiêu chí so sánh bao gồm: độ chính xác, độ ổn định và thời gian chạy.
* **Phân tích ảnh hưởng của độ di chuyển của nút**: Bài báo đã phân tích ảnh hưởng của độ di chuyển của nút đến kết quả của phương pháp được đề xuất. Các tiêu chí phân tích bao gồm: số lượng cộng đồng, số lượng thành viên trong cộng đồng lớn nhất và thời gian chạy.

# CHƯƠNG 4: CƠ SỞ LÝ THUYẾT

## Lý thuyết về Nodes và Edges

### Thế nào là Nodes và Edges

Một nút (node) được hiểu là đại điện một thực thể, một actor trong việc tạo ra liên kết mạng.

Cạnh (Edge): các cạnh thể hiện sư hiện diện của một kết nối hoặc mối quan hệ giữa hai nút

Hình 1 Minh họa nodes và edges

### Edges Direction

Có hai loại cạnh: cạnh có hướng và cạnh vô hướng. Trước khi xây dựng lên mạng cần phải giải mã dữ liệu của chúng ta chứa loại cạnh nào khi xây dựng biểu đồ mạng.

+ Các cạnh được định hướng: được áp dụng từ nút này sang nút khác với một nút bắt đầu và một nút kết thúc.

+ Các cạnh không định hướng: các mối quan hệ này được đáp lại bởi cả hai bên mà không có nút bắt đầu và nút kết thúc rõ ràng.

### Edge Weight

Trọng số của một cạnh là số lần cạnh đó xuất hiện giữa hai nút cụ thể.

## Lý thuyết độ đo trong mạng xã hội

**Biện pháp trung tâm (Centrality Measures):** Tính trung tâm là một tập hợp các chỉ số được sử dụng để xác định mức độ quan trọng và ảnh hưởng của một nút cụ thể đối với toàn bộ mạng. Các biện pháp trung tâm được sử dụng trên các nút cụ thể trong mạng và không cung cấp thông tin ở cấp độ mạng. Ví dụ 3 độ do trung tâm dưới đây:

### 4.2.1 Degree Centrality

Độ của một nút là số cạnh mà nút đó có. Số đo này giúp ta đo số lượng của các mối quan hệ trực tiếp của một tác nhân nào đó với các thành viên khác trong mạng xã hội.

CD(v) =

− Trong đó:

+ n: là số đỉnh của đồ thị

+ deg(v): tổng số các liên kết trực tiếp đến đỉnh v (bậc của đỉnh).

### 4.2.2 Betweenness Centrality

Betweenness Centrality: đo lường tầm quan trọng của các kết nối của một nút trong việc cho phép các nút đến các nút khác (trong một bước nhảy). Khoảng giữa của một nút là số đường đi ngắn nhất mà nút được đưa vào chia cho tổng số đường đi ngắn nhất. Điều này sẽ cung cấp tỷ lệ phần trăm đường dẫn ngắn nhất trong mạng mà nút đang ở trong đó. Số đo trung tâm trung gian xác nhận một tác nhân nào đó trong mạng có thể có thể có ít gắn kết với các thành viên khác trong mạng xã hội (số đo bậc trung tâm thấp), cũng không gần gũi lắm với các thành viên khác (số đo trung tâm lân cận thấp), nhưng lại là cầu nối (bridge) hay nhà trung gian cần thiết trong mọi cuộc trao đổi trong mạng.

Ảnh có chứa văn bản, Phông chữ, màu trắng, ảnh chụp màn hình

Mô tả được tạo tự động

Trong đó:

+ σst(u): số đường đi ngắn nhất giữa s và t có chứa u (s≠u≠ t)

+ σst : tổng số đường ngắn nhất giữa s và t (s ≠ u ≠ t)

### 4.2.3 Closeness Centrality

Mức độ gần gũi là độ đo đo lường mức độ kết nối của một nút với mọi nút khác trong mạng. Mức độ gần gũi của một nút là số bước trung bình cần thiết để tiếp cận mọi nút khác trong mạng. Bước nhảy là đường đi của một cạnh từ nút này đến nút khác.

Trong một đồ thị liên thông, độ đo gần trung tâm của một nút là thước đo mức độ trung tâm trong mạng, được tính bằng nghịch đảo của tổng độ dài của các đường đi ngắn nhất giữa nút và tất cả các nút khác trong biểu đồ. Do đó, một nút càng ở trung tâm thì nó càng gần với tất cả các nút khác. Số đo Closeness Centrality được tính bằng công thức sau:

CC(v) =

Trong đó:

+ n: là số đỉnh của đồ thị

+ d(u.v): là khoảng cách ngắn nhất từ đỉnh u đến v của đồ thị

## Lý thuyết thuận toán phát hiện cộng đồng

Cộng đồng được tạo từ các cá nhân sao cho các cá nhân trong cùng một nhóm sẽ tương tác với nhau thường xuyên hơn với các cá nhân nằm ngoài nhóm. Có ba cách tiếp cận để khám phá cộng đồng trong mạng xã hội:

+ Dựa trên cấu trúc mạng xã hội để khám phá cộng đồng.

+ Dựa trên thông tin trao đổi theo các tương tác trên mạng xã hội.

+ Dựa trên vừa cấu trúc và nội dung trao đổi.

### 4.3.1 Thuật toán Girvan Newman

Thuật toán Girvan Newman dựa trên quan niệm cho rằng khi các cộng đồng được gắn kết với nhau thì đường đi giữa cộng đồng này đến cộng đồng khác sẽ đi qua các cạnh nối giữa các cộng đồng với tần suất cao. Mục đích chính của thuật toán là tìm những cạnh nối đó. Thay vì việc xây dựng cộng đồng bằng cách thêm vào các cạnh mạnh mẽ nhất, chúng ta sẽ xây dựng bằng cách loại bỏ dần dần các cạnh nối từ đồ thị ban đầu. Khi đó, các cộng đồng trong mạng sẽ bị ngắt kết nối với nhau, ta có thể xác định được cách phân vùng đồ thị thành các phần nhỏ riêng rẽ. Để làm được việc này, điều quan trọng nhất của thuật toán là việc tính toán như thế nào, sử dụng tính chất nào để phát hiện ra những cạnh nối này, từ đó loại bỏ chúng ra khỏi đồ thị. Thuật toán lần đầu tiên được đề xuất bởi Freeman. Theo Freeman, các cạnh được coi là cạnh có số lượng con đường ngắn nhất giữa các cặp đỉnh khác nhau chạy qua nó. Cạnh nối có ảnh hưởng rất lớn đến dòng chảy của thông tin giữa các nút khác, đặc biệt là trong trường hợp thông tin lưu truyền trong mạng chủ yếu theo con đường ngắn nhất. Thuật toán điển hình nhất trong các thuật toán chia này là thuật toán Girvan-Newman.

Nếu một mạng lưới bao gồm các cộng đồng hoặc nhóm chúng chỉ được liên kết nối yếu bằng một nhóm cạnh, thì tất cả các đường đi ngắn nhất giữa các cộng đồng khác nhau sẽ phải đi dọc theo một trong số ít các cạnh thuộc nhóm cạnh đó. Vì vậy, các cạnh kết nối các cộng đồng sẽ là cạnh có độ trung gian cao. Bằng cách loại bỏ các cạnh, thuật toán Girvan-Newman tách được thành các nhóm riêng biệt. Thuật toán được thực hiện theo các bước sau:

+ Bước 1: Tính độ đo trung gian cho tất cả các cạnh trong mạng.

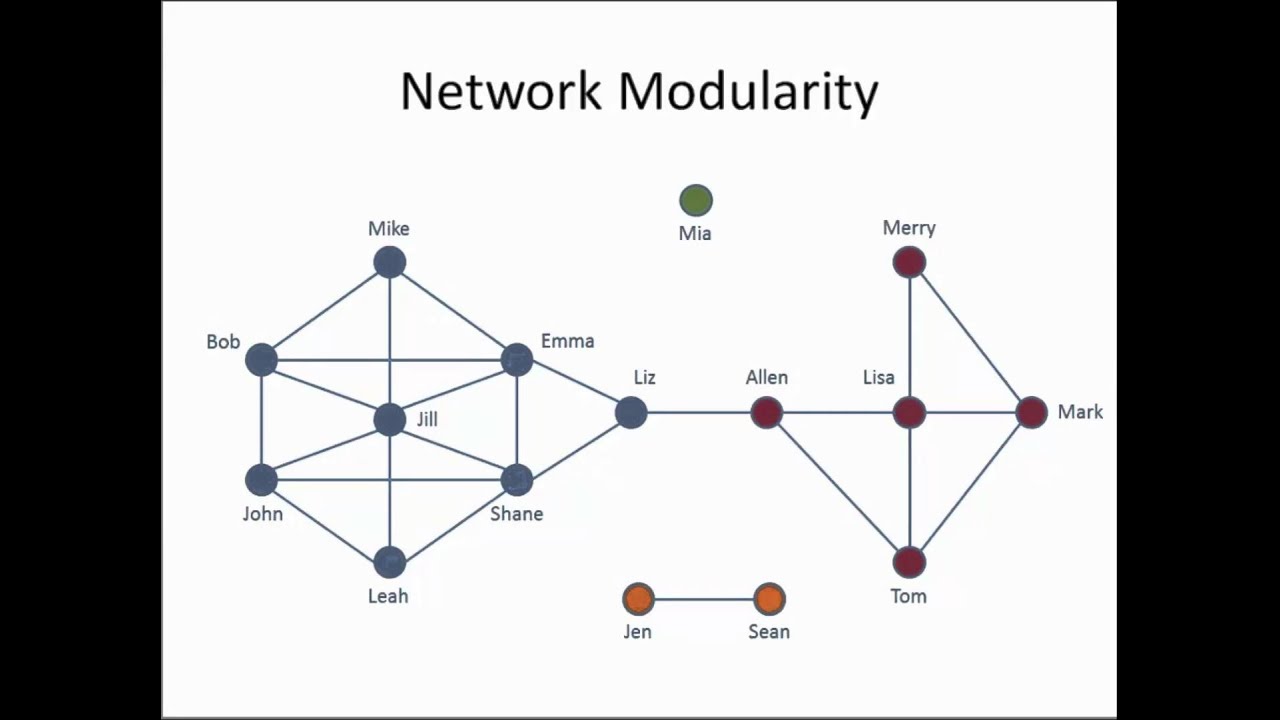
+ Bước 2: Hủy bỏ các cạnh có độ trung gian cao nhất.

+ Bước 3: Tính lại độ trung gian cho tất cả các cạnh bị ảnh hưởng theo các cạnh đã loại bỏ.

+ Bước 4: Lặp lại từ bước 2 cho đến khi không còn các cạnh trung gian

### 4.3.2 Thuật toán phân cụm

Thuật toán phân cụm (clustering algorithm) là một phương pháp máy học được sử dụng để tự động phân loại các điểm dữ liệu thành các nhóm (cụm) dựa trên các đặc trưng chung giữa chúng. Mục tiêu là để các điểm dữ liệu trong cùng một cụm có sự tương đồng lớn nhau và các cụm khác nhau có sự khác biệt đáng kể.



Hình 2 Minh họa thuật toán phân cụm

Thuật toán phân cụm giúp tổ chức dữ liệu thành các nhóm có tính tương đồng, từ đó mang lại sự hiểu biết sâu sắc hơn về cấu trúc và mối quan hệ trong dữ liệu. Mục tiêu chính là tạo ra các cụm có ý nghĩa và giúp hiểu rõ hơn về đặc tính chung của các điểm dữ liệu.

### 4.3.3Thuật toán Newman Modularity

Thuật toán Newman Modularity là một phương pháp phân cụm đồ thị (graph clustering) được thiết kế để tìm ra cấu trúc cộng đồng trong mạng phức tạp. Thuật toán này có nguồn gốc từ nghiên cứu của Mark Newman, một nhà nghiên cứu trong lĩnh vực mạng phức tạp.

Mục tiêu của Newman Modularity là phân chia đồ thị thành các cụm sao cho sự tương tác giữa các nút bên trong mỗi cụm là mạnh hơn so với sự tương tác giữa các nút ở giữa các cụm. Newman Modularity được đo lường bằng một chỉ số được gọi là modularità, có giá trị nằm trong khoảng [-1, 1]. Giá trị càng gần 1 cho thấy mô hình cụm tốt, trong khi giá trị càng gần -1 cho thấy mô hình cụm không tốt.

Thuật toán hoạt động bằng cách phân chia đồ thị thành các cụm và tính modularità của từng cụm. Quá trình này được thực hiện bằng cách so sánh số lượng cạnh nối nút bên trong mỗi cụm với số lượng cạnh dự kiến trong trường hợp ngẫu nhiên. Nếu modularità của cụm là dương, nghĩa là có sự tăng cường tương tác bên trong cụm, thì cụm đó được giữ lại.

Newman Modularity thường được sử dụng trong nhiều lĩnh vực như xã hội học, sinh học mạng, và phân tích dữ liệu mạng để hiểu rõ hơn về cấu trúc cộng đồng trong mạng phức tạp. Thuật toán được thực hiện theo các bước sau:

+ Bước 1: Tính trọng số của từng cạnh trong đồ thị, phản ánh mức độ tương tác giữa các đỉnh.

+ Bước 2: Tính tổng trọng số của toàn bộ đồ thị và tổng bậc của mỗi đỉnh.

+ Bước 3: Sử dụng trọng số và tổng trọng số để xây dựng ma trận Modularity.

+ Bước 4: Áp dụng thuật toán tìm kiếm để phân cụm đỉnh dựa trên ma trận Modularity.

+ Bước 5: Tính giá trị Modularity ban đầu để đánh giá việc phân cụm.

+ Bước 6: Lặp lại: Nếu giá trị Modularity không thay đổi đáng kể, kết thúc thuật toán. Nếu giá trị Modularity thay đổi, tiếp tục với các bước sau: Tính lại giá trị Modularity sau khi loại bỏ một cụm và tính toán lại phân cụm.

# CHƯƠNG 5: CÁC NGHIÊN CỨU LIÊN QUAN

Hiện nay, nhiều học giả tập trung vào nghiên cứu về mạng xã hội, bao gồm khám phá cộng đồng , bảo vệ quyền riêng tư, mạng xã hội và nhiều khía cạnh khác. Khám phá cộng đồng đã trở thành một trung tâm nghiên cứu. Ví dụ, nhiều phương pháp khám phá cộng đồng đã được đề xuất để chia cấu trúc mạng xã hội, bao gồm:

• Thuật toán chia nhỏ dựa trên đồ thị.

• Thuật toán độ mô-đun.

• Thuật toán gom cụm cạnh.

• Thuật toán gom cụm phân cấp.

• Phương pháp phân tán hạt giống.

• Thuật toán duyệt ngẫu nhiên.

• Phương pháp truyền nhãn.

## 5.1 Khám phá cộng đồng trong mạng xã hội

*Girvan* và *Newman* đã đề xuất thuật toán GN (Girvan-Newman) dựa trên việc loại bỏ các cạnh có 'số đường trung gian' lớn nhất. Họ đầu tiên giới thiệu khái niệm 'độ modularity' để đo lường chất lượng của cấu trúc cộng đồng thông qua độ modularity. Tuy nhiên, thuật toán GN chỉ áp dụng cho mạng quy mô nhỏ.

*Huysman* và *Rattigan* sau đó đề xuất các thuật toán riêng của họ để cải thiện hiệu suất của thuật toán GN do sự gia tăng quy mô của mạng xã hội.

*Newman* đề xuất thuật toán FastQ, một thuật toán khám phá cộng đồng nhanh hơn GN bằng cách tối ưu hóa độ modularity thông qua việc gộp các cộng đồng.

*Clauset* đề xuất thuật toán CNM để cải thiện hiệu suất so với FastQ bằng việc sử dụng cấu trúc dữ liệu.

*Danon* đã cải thiện thuật toán FastQ để gộp các cộng đồng lớn.

*Guimera* và *Amaral* sử dụng phương pháp làm nguội mô phỏng để tối ưu hóa độ modularity và thu được cấu trúc cộng đồng tối ưu.

*White* và *Smyth* tối ưu hóa hàm độ modularity bằng phương pháp phổ và đề xuất một phương pháp khám phá cộng đồng dựa trên gom cụm phổ.

*Van Dongen* đề xuất một thuật toán gom cụm nút dựa trên bước đi ngẫu nhiên.

*Rosvall* và *Bergstrom* sử dụng lý thuyết mã để tìm mã ngắn nhất từ đường đi ngẫu nhiên để phân chia toàn bộ mạng.

*Tabrizi* đề xuất phương pháp dựa trên độ modularity và tập hợp dựa trên bộ sưu tập cho Khám phá cộng đồng đa tầng.

Tác giả đã đề xuất một phương pháp riêng cho khám phá cộng đồng xã hội dựa trên độ mô-đun, độ quan tâm và nguyện vọng cá nhân.

*Javed* và đồng nghiệp đã cung cấp một bài đánh giá về các thuật toán khám phá cộng đồng từ truyền thống đến hiện đại và xem xét các thuật toán dựa trên kỹ thuật giảm chiều dữ liệu như ma trận phi tuyệt không âm và phân tích thành phần chính.

## 5.2 Khám phá cộng đồng trong mạng xã hội di động

*Xu* và đồng nghiệp đã đề xuất phương pháp xây dựng cấu trúc cộng đồng dựa trên mức độ kết hợp entropy để phát hiện và dự đoán tình bạn trong mạng xã hội di động.

*Hutair* và đồng nghiệp đề xuất một thuật toán dựa trên vị trí địa lý và sự tương đồng trong sở thích của người dùng để gom cụm nút thành các cộng đồng trong mạng xã hội di động.

*Sarzynska* và đồng nghiệp nghiên cứu về việc sử dụng mô hình trống tích hợp thông tin không gian và đề xuất mô hình trống mới dựa trên mô hình truyền tải dân số để tối ưu hóa độ modularity trong mạng xã hội di động.

Có nhiều phương pháp dựa trên sự thay đổi của topologia và thời gian trong mạng xã hội di động. *Amelio* và *Pizzuti* mở rộng phương pháp gom cụm tiến hóa để phát hiện cộng đồng trong mạng đa lớp thời gian. *Chakrabarit* đề xuất khái niệm gom cụm tiến hóa để nắm bắt quá trình tiến hóa của cộng đồng. *Tantipathananandh* biến đổi vấn đề phát hiện cộng đồng thành vấn đề tô màu đồ thị và đề xuất thuật toán heuristics cho việc kết nối cộng đồng trong các mạng vào các thời điểm khác nhau. Lin đề xuất thuật toán Facetnet để phát hiện cộng đồng và phân tích sự tiến hóa dựa trên cấu trúc cộng đồng lịch sử*. Kim* và *Han* đề xuất thuật toán gom cụm tiến hóa dựa trên kích thước hạt và mật độ. *Palla* đề xuất thuật toán lọc cục bộ để xác định sự kiện chính trong sự tiến hóa của cộng đồng.

Nhiều nghiên cứu cũng tập trung vào ứng dụng trong mạng xã hội di động. Ví dụ, *Guan* và *Wu* chia cộng đồng xã hội dựa trên đặc điểm hoạt động của con người trong mạng lưới cơ hội. *Liu* đề xuất thuật toán định tuyến tiến phân mờ mờ dựa trên tính tương đồng tổng hợp của nút trong mạng xã hội cơ hội. *Zhang* đề xuất phương pháp gom cụm dựa trên phương pháp phối hợp theo cặp để phát hiện cộng đồng tạm thời thông qua quá trình tiếp xúc theo cặp.

Tổng cộng, những nghiên cứu này đem lại một cái nhìn sâu rộng và đa dạng về phương pháp phát hiện và phân tích cộng đồng trong mạng xã hội di động, và cung cấp cơ sở lý thuyết và ứng dụng quan trọng trong lĩnh vực này.

# CHƯƠNG 6: KHÁM PHÁ CỘNG ĐỒNG

## 6.1 Các định nghĩa liên quan

Cho một đồ thị có trọng số và có hướng G(V, E) biểu thị mối quan hệ của người dùng trong mạng xã hội, tập hợp các nút được biểu thị bởi V, tập hợp các cạnh được biểu thị bởi E (được sử dụng để chỉ các thông điệp được chuyển bởi người dùng), số lượng nút được biểu thị bởi n1 = |V(G)| và số lượng cạnh được biểu thị bởi n2 = |E(G)|, sau đó một mạng xã hội với n1 nút có thể được biểu thị bằng ma trận kề Bn1×n1, trong đó Bi,j biểu thị

Bi,j =

và số lượng cạnh được biểu thị là

A number symbols and symbols

Description automatically generated with medium confidence

Chúng ta cũng đặt ra rằng người dùng i trong mạng xã hội có một tập hợp thuộc tính ψi, trong đó ψk ∈ ψi được sử dụng để biểu thị thuộc tính thứ k của người dùng i. Cộng đồng cj có một tập hợp thuộc tính φj, trong đó φk ∈ φj được sử dụng để biểu thị thuộc tính thứ k của cộng đồng cj. Hơn nữa, đối với người dùng i, trọng số của thuộc tính ψk đối với cộng đồng cj được ký hiệu là k,j, và tính di động của người dùng i được ký hiệu là πi.

**Định nghĩa 1:** Bậc của nút (Bậc của Người dùng): nó biểu thị phạm vi ảnh hưởng của người dùng i trong mạng xã hội, tức là số lượng tất cả các cạnh liên quan đến nút (người dùng) i được ký hiệu là bậc của nút, được biểu thị như sau:

A close up of a text

Description automatically generated with medium confidence

trong đó:

- kin là số lượng tất cả các cạnh liên quan đến nút i.

- kout là số lượng tất cả các cạnh liên quan từ nút i.

Vì vậy, kin là bậc vào của nút i và kout là bậc ra của nút i.

**Định nghĩa 2**: Tương đồng trong Cộng đồng Xã hội: nó biểu thị sự tương tự cộng đồng xã hội giữa người dùng i và cộng đồng cj, được đo lường dựa trên các thuộc tính xã hội. Tương tự cộng đồng xã hội giữa người dùng i và cộng đồng cj được định nghĩa như sau:

A black and white math symbols

Description automatically generated

Sự giao nhau i,j của tập hợp thuộc tính xã hội giữa người dùng i và cộng đồng cj được định nghĩa như sau:

A black and white image of a symbol

Description automatically generated with medium confidence

Nó đo lường sự tương đồng giữa các thuộc tính xã hội của người dùng và cộng đồng.

**Định nghĩa 3:** Tương đồng thuộc tính nút: nó biểu thị tương tự thuộc tính nút của người dùng i đối với người dùng z, cũng được đo lường dựa trên các thuộc tính xã hội. Tương tự thuộc tính nút của người dùng i đối với người dùng z được định nghĩa như sau:

A black and white math symbols

Description automatically generated

Sự giao nhau i,j của tập hợp thuộc tính xã hội giữa người dùng i và cộng đồng cj được định nghĩa như sau:

A black and white image of a symbol

Description automatically generated

**Định nghĩa 4:** Sở thích của Nút: nó được sử dụng để đo lường sở thích của một người dùng, trong đó một sở thích của mỗi người dùng được biểu thị dưới dạng < Interest\_type,weight>, trong đó Interest\_type biểu thị sở thích cụ thể và weight biểu thị trọng lượng tương ứng. Do đó, các sở thích của người dùng i được định nghĩa như sau:

A close up of a number

Description automatically generated

trong đó:

- Interesti là tập hợp các sở thích của người dùng i.

- Interest\_typei,k biểu thị loại sở thích thứ k trong vectơ Interesti

- i,k biểu thị trọng lượng của sở thích thứ k trong vectơ Interesti

- K là số lượng sở thích của người dùng i.

**Định nghĩa 5:** Bậc Mô-đun WD: Đây là tỷ lệ giữa mật độ cạnh trong cộng đồng và mật độ cạnh giữa các cộng đồng liên quan, công thức của nó như sau:

A black text on a white background

Description automatically generated

trong đó:

- WD là bậc Mô-đun giữa người dùng i và j.

- là hàm được sử dụng để tính giá trị lớn nhất của số cạnh giữa ci và cj.

- n1 là số lượng cạnh.

- n2 là số lượng nút.

- ci biểu thị cộng đồng mà nút i thuộc về.

- cj biểu thị cộng đồng mà nút j thuộc về.

Từ công thức trên, ta có thể biết rằng khi bậc Mô-đun của mạng là giá trị tối thiểu của 0, sự khởi tạo của mỗi nút là độc lập để tạo thành một cộng đồng đơn; khi bậc Mô-đun của mạng là giá trị tối đa của 1, tất cả các nút được phát hiện vào một cộng đồng. Quá trình phát hiện và phân loại cộng đồng dựa trên nguyên tắc tối ưu hóa độ phân cộng đồng: 1) các thuộc tính có giá trị âm nên được loại trừ; 2) các nút có độ tương tự lớn hơn nên được phân thành cùng một cộng đồng.

## 6.2 Đề xuất giao thức khám phá cộng đồng xã hội

Một giao thức phát hiện cộng đồng xã hội được đề xuất cho các mạng xã hội di động dựa trên các yếu tố xã hội, bao gồm sự tương tự về thuộc tính xã hội và tương tự về sở thích nút, trong đó chúng ta cũng xem xét sự di động của mỗi nút. Giao thức đề xuất của tác giả trước tiên đo lường các thuộc tính xã hội của mỗi nút, sau đó tính toán tương tự về thuộc tính xã hội và tương tự về sở thích nút giữa các nút. Dựa trên sự tương tự về social-ware, tác giả xây dựng một phương pháp để tính bậc mô-đun. Cuối cùng, giao thức đề xuất của tác giả phát hiện các cộng đồng của mạng xã hội di động dựa trên bậc mô-đun tính toán

### 1) TƯƠNG ĐỒNG THUỘC TÍNH CỘNG ĐỒNG

Trong phần này, chúng ta chỉ ra cách đo lường các thuộc tính xã hội của từng nút. Dựa trên một số thuộc tính xã hội chung, chúng ta xây dựng một tập hợp thuộc tính xã hội cho mỗi nút, trong đó các thuộc tính xã hội có thể được sử dụng để đo lường đặc điểm xã hội của mỗi nút. Ví dụ, mỗi người dùng có các thuộc tính xã hội cụ thể, bao gồm tên, nghề nghiệp, tuổi, sở thích, nơi cư trú, ảnh hưởng và nhiều thuộc tính khác. Trong các mạng xã hội di động, tính di động cũng là một đặc điểm xã hội rất quan trọng, có thể ảnh hưởng đến cấu trúc của cộng đồng xã hội. Do đó, đối với một số nút có ảnh hưởng lớn và di chuyển thường xuyên, hành vi di động của họ có thể ảnh hưởng đến sự thay đổi cấu trúc cộng đồng. Ngoài ra, đối với các cộng đồng khác nhau, các thuộc tính khác nhau có hiệu quả khác nhau đối với cấu trúc của cộng đồng.

Theo các định nghĩa liên quan (Định nghĩa 2 và 3), việc tính toán tương tự thuộc tính cộng đồng dựa trên tương tự cộng đồng xã hội và tương tự thuộc tính nút. Việc tính toán tương tự thuộc tính cộng đồng cũng liên quan đến tính di động của nút. Tương tự thuộc tính cộng đồng CNSi,j giữa người dùng i và cộng đồng cj được tính như sau:

A black and white math symbols

Description automatically generated with medium confidence

trong đó:

- là hệ số tác động,

- cj là kích thước của cộng đồng cj,

- k biểu thị người dùng k thuộc về cộng đồng cj.

Trong các mạng xã hội di động, do các thuộc tính của các nút khác nhau, các hệ số tác động đối với việc chia cộng đồng cũng khác nhau. Tác giả xem xét rằng tương tự cộng đồng xã hội giữa người dùng i và cộng đồng cj có ảnh hưởng lớn hơn đến việc phát hiện cộng đồng khi η lớn hơn; ngược lại, nó có ảnh hưởng nhỏ hơn đến việc phát hiện cộng đồng. Khi η = 0, việc phát hiện cộng đồng chỉ dựa trên tương tự thuộc tính nút giữa người dùng i và tất cả các người dùng khác thuộc về cộng đồng cj. Khi η = 1, tương tự thuộc tính cộng đồng chỉ là tương tự cộng đồng xã hội. Khi tương tự thuộc tính cộng đồng giữa người dùng i và cộng đồng cj lớn hơn, khả năng người dùng (nút) bị chia thành cộng đồng cj cao hơn. Do đó, hệ số tác động η được sử dụng để cân bằng giữa thuộc tính cộng đồng và thuộc tính nút.

### 2) TƯƠNG ĐỒNG SỞ THÍCH NGƯỜI DÙNG

Trong phần này, tác giả chỉ ra cách đo lường sở thích của từng người dùng. Bởi vì những người dùng chia sẻ cùng một sở thích thì có khả năng cao hình thành một cộng đồng xã hội, nên sở thích người dùng đóng một vai trò quan trọng trong việc phát hiện cộng đồng. Để tính toán tương tự sở thích của người dùng, tác giả giới thiệu phương pháp gom cụm [58] vào sơ đồ phát hiện cộng đồng của tác giả. Theo Định nghĩa 4 (sở thích của nút), giao của các vector sở thích giữa người dùng i và người dùng j được định nghĩa như sau:

A close up of a word

Description automatically generated

trong đó:

- Interesti biểu thị vector sở thích của người dùng i và Interestj biểu thị vector sở thích của người dùng j.

Dựa trên các vector sở thích, tương đồng sở thích ISM*i,j* của người dùng i đối với người dùng j được tính toán như sau:

A black and white text

Description automatically generated

trong đó:

- k là trọng số của sở thích thứ k trong vector sở thích tương ứng.

### 3) TRỌNG SỐ CẠNH

Để phát hiện cộng đồng, cần phải đo mối quan hệ giữa người dùng i và người dùng k thuộc cộng đồng cj. Vì vậy, dựa trên tương tự thuộc tính cộng đồng giữa người dùng i và cộng đồng cj và tương tự sở thích giữa người dùng i và người dùng k thuộc cj, tương tự cạnh giữa người dùng i và người dùng k được đo lường dựa trên sự kết hợp của tương tự thuộc tính cộng đồng và tương tự sở thích. Trọng số EWi,k của tương tự cạnh được định nghĩa như sau:

A black and white text

Description automatically generated

trong đó:

- CNSi,j là tương đồng thuộc tính cộng đồng giữa người dùng i và cộng đồng cj.

- ISMi,j là tương tự sở thích giữa người dùng i và người dùng k thuộc cj.

- là hệ số tác động, nằm trong khoảng từ 0 đến 1, được sử dụng để cân bằng tác động lên tương tự cạnh. Khi = 1, việc phát hiện cộng đồng chỉ liên quan đến tương tự thuộc tính; khi = 0 việc phát hiện cộng đồng chỉ liên quan đến tương tự sở thích. Tương tự cạnh có thể chỉ ra mức độ thân mật giữa hai nút. Trọng số của tương tự cạnh giữa hai nút càng lớn, khả năng rằng hai nút sẽ được chia thành một cộng đồng càng cao. Ngoài ra, vì sự di động πi được giới thiệu vào CNSi,j và ISMi,j, sự di động được sử dụng như một yếu tố tác động để tính toán trọng số của tương tự cạnh.

### 4) GÓC ĐỘ MÔ ĐUN CỘNG ĐỒNG

Trong phần này, tác giả chỉ cách tính độ module dựa trên tương tự cạnh có trọng số. Độ module cộng đồng là một độ đo rất quan trọng để chia cấu trúc cộng đồng. Nó chỉ ra mức độ thân mật giữa các nút thuộc cùng một cộng đồng. Các nút trong cộng đồng này có khả năng lớn hơn để được chia thành cùng một cộng đồng. Độ module cao hơn của một cộng đồng, cấu trúc cộng đồng đã được phát hiện càng ổn định hơn. Trong đề xuất của tác giả, khi cấu trúc cộng đồng được phát hiện, sự gia tăng của độ module được sử dụng để đo lường sự thay đổi của độ module để đánh giá mức độ thân mật giữa nút được quan sát và cộng đồng được phát hiện. Theo Định nghĩa 5, độ module WD∗ dựa trên tương tự cạnh có trọng số được tính như sau:

A black text on a white background

Description automatically generated

trong đó:

- EWi,j là trọng số của tương tự cạnh giữa nút i và nút j.

- là tổng trọng số của các bậc ra của nút i.

- là tổng trọng số của các bậc vào của nút j.

- SW là tổng trọng số của tương đồng cạnh.

Theo tính toán của WD∗, nếu nút i được chia thành cộng đồng cj, thì sự gia tăng WD∗ của độ module được tính như sau:

A math equations with black text

Description automatically generated with medium confidence

trong đó:

- EWcj là tổng trọng số của tương tự cạnh nội bộ giữa tất cả các nút thuộc cộng đồng cj.

- EWi,cj là tổng trọng số của tương tự cạnh giữa nút i và các nút lân cận khác thuộc cộng đồng cj.

- là tổng trọng số của các bậc vào của cộng đồng cj.

- là tổng trọng số của các bậc ra của cộng đồng cj.

Độ module đo sự mức độ thân mật giữa các nút thuộc cùng một cộng đồng. Các nút thuộc cùng một cộng đồng sẽ có độ module cao hơn, ngụ ý rằng cộng đồng được phát hiện sẽ ổn định hơn.

### 5) MÔ HÌNH DỰA TRÊN MỨC ĐỘ MÔ-ĐUN

Trong phần này, chúng ta trình bày một mô hình chi tiết dựa trên mức độ mô-đun để phát hiện cộng đồng, dựa trên "social-ware", bao gồm tương tự về đặc điểm xã hội và tương tự về sở thích của người dùng. Trong mô hình của tác giả, sự di động của người dùng được xem xét như một yếu tố ảnh hưởng đến việc tính toán trọng số độ tương tự của cạnh. So với các mô hình tương tự, mô hình của tác giả tập trung vào các đặc điểm xã hội và tính di động của người dùng. Hơn nữa, thông qua việc thay đổi yếu tố ảnh hưởng ξ ∈ [0, 1] để kiểm soát tỷ lệ tương tự về đặc điểm xã hội và tương tự về sở thích của người dùng, mô hình của tác giả có thể điều chỉnh tỷ lệ ảnh hưởng lên trọng số của độ tương tự cạnh, từ đó có thể tạo ra cấu trúc cộng đồng ổn định hơn. Mô hình của tác giả được mô tả như sau (xem trong Hình 1):

A group of lines and dots

Description automatically generated

Hình 3 Kết quả thực nghiệm 1

**Bước 1:** Trong bước này, chúng ta nhập đồ thị xã hội kề G(V, E), tổng số n1 của các nút và tập hợp cộng đồng C1. Sau đó, chúng ta tính toán độ tương tự về thuộc tính cộng đồng CNSi,zsvà tương tự về sở thích ISMi,js giữa tất cả các cặp nút có i ∈ 1, 2, ...n1, z ∈1, 2, ...|C1| và j ∈ 1, 2, ...n1. Cuối cùng, chúng ta xuất ra tất cả CNSi,zsvà ISMi,js, trong đó mỗi nút ban đầu được coi như một cộng đồng riêng. Quy trình được mô tả trong Thuật toán 1.

**Thuật toán 1: Tính toán Sự tương đồng của Thuộc tính Cộng đồng và Sở thích Của Node**

**Input:** the adjacent social network G(V, E), the total number n1 of nodes, the community set C1

**Output:** all CNSi,zs and ISMi,js

**Begin**

Initialize each node from G(V, E) to form a community and save the structure to C1;

**for** the node i ∈ V **do**

**for** the community z ∈ C1 **do**

CNSi,z= 0; //initialize CNSi,z

**for** the node j ∈ z **do**

ISMi,j= 0; //initialize ISMi,j

**End for**

**End for**

**End for**

**for** the node i ∈ V **do**

**for** the community z ∈ C1 **do**

Compute CSi,z;

sum = 0;

**for** the node j ∈ z and j belongs to the adjacent node set N(i) of the node i **do**

Compute NSi,j;

Compute ISMi,j;

sum = sum + NSi,j;

**End for**

CNSi,z = η · CSi,z + (1 − η) · ; //|z| is the size of z

**End for**

**End for**

**return** all CNSi,zs and ISMi,js;

**End**

**Bước 2:** Trong bước này, chúng ta nhập G(V, E), n1, C1, tất cả CNSi,zsvà tất cả ISMi,js với i ∈ 1, 2, ...n1, z ∈ 1, 2, ...|C1| và j ∈ 1, 2, ...n1. Chúng ta tính toán trọng số của độ tương tự cạnh giữa tất cả các cặp nút và đo lường sự tăng của mức độ mô-đun. Cuối cùng, chúng ta xuất ra tập hợp C2 các cộng đồng và một cờ. Quy trình được mô tả trong Thuật toán 2. Thuật toán 2 phát hiện cấu trúc cộng đồng dựa trên mức độ mô-đun: 1) Đối với mỗi cộng đồng j (j ∈ G), thuật toán tính toán tất cả các tăng 1WD∗ của mức độ mô-đun, trong đó chúng ta giả định cộng đồng j cố gắng tham gia vào tất cả các cộng đồng kề. 2) Thuật toán tìm cộng đồng kề có giá trị tăng lớn nhất (WD∗ > γ), và sau đó cộng đồng j tham gia vào cộng đồng kề tương ứng, trong đó γ là một tham số được thiết lập trước để kiểm soát kết thúc của quy trình. 3) Miễn là các giá trị của WD∗ đang thay đổi và lớn hơn γ, quá trình hợp nhất các cộng đồng sẽ tiếp tục qua chu kỳ lặp cho đến khi không thể chia cộng đồng thành các cộng đồng cấp cao hơn. 4) Nếu tất cả các giá trị của WD∗ đều nhỏ hơn γ, thuật toán trả về một cờ 'End' và một tập hợp C2 của các cộng đồng, nếu không, thuật toán trả về một cờ 'Null' và tập hợp C2 của các cộng đồng.

**Thuật toán 2: Phân chia cộng đồng dựa trên độ Module**

**Input:** G(V, E), n1, C1, all CNSi,zs and all ISMi,js

**Output:** a set C2 of communities and a flag Flag

**Begin**

C2 ← {{}, ......{}}; //initialize C2

**while** being partitioned into the communities of higher level **do**

Q ← {}; //being emptied

**while** the values of WD∗ are changing and more than γ **do**

**for** the community i ∈ C1 **do**

**for** the community j that belongs to the adjacent community set NC(i) of the community i **do**

Compute the increment WD∗i,j; //the community i tries to join the adjacent

community j

**if** WD∗i,j> γ **then**

Q ←**Save**(1WD∗i,j, j);

**End if**

**End for**

Max\_Index ←**Find\_max**(Q);

//find the community Max\_Index with the maximum of increment

The community i joins the corresponding community Max\_Index;

Update C1;

**End for**

**End while**

**End while**

C2 ← C1;

**if** all the values of WD∗ are less than γ **then**

**return** ‘‘End’’ and C2;

**else**

**return** ‘‘Null’’ and C2;

**End if**

**End**

**Bước 3:** Trong bước này, chúng ta nhập G(V, E), n1, C2 và tất cả ISMi,js, sau đó tính toán lại độ tương tự thuộc tính cộng đồng CNSi,zsgiữa người dùng i và cộng đồng z với i ∈ 1, 2, ...n1 và z ∈ 1, 2, ...|C2|. Sau đó, quay lại Bước 2 với C1 ← C2. Quy trình được mô tả trong Thuật toán 3. Thuật toán 3 tính toán lại độ tương tự thuộc tính cộng đồng CNSi,zs: 1) Nếu Thuật toán 2 trả về một cờ kết thúc, thì Thuật toán 3 sẽ kết thúc, nếu không, Thuật toán 3 sẽ tiếp tục; 2) Dựa trên tập hợp C2 của các cộng đồng, đối với mỗi cộng đồng z (z ∈ C2), thuật toán tính toán độ tương tự thuộc tính cộng đồng CNSi,zsgiữa người dùng i và cộng đồng z theo mô tả ở Mục 3.2.1; 3) Thuật toán 3 gọi Thuật toán 2 với C1 ← C2.

**Thuật toán 3: Khám phá cộng đồng động**

**Input:** G(V, E), n1, C2, all ISMi,js

**Output:** a set C3 of communities

**Begin**

Flag =‘‘Null’’;

**while** Flag <>‘‘End’’ **do**

**for** the node i ∈ V **do**

**for** the community z ∈ C2 **do**

Compute CSi,z;

sum = 0;

**for** the node j ∈ z and j belongs to the adjacent node set N(i) of the node i **do**

Compute NSi,j;

sum = sum + NSi,j;

**End for**

CNSi,z = η · CSi,z + (1 − η) · ;

**End for**

**End for**

(Flag,C2) ←**Call** Algorithm 2(G, n1, C2, all CNSi,zs, all ISMi,js);

**End while**

C3 ← C2;

**return** C3;

**End**

# CHƯƠNG 7: THU THẬP VÀ PHÂN TÍCH DỮ LIỆU

Để đảm bảo tính hiệu quả của hệ thống mà tác giả đã phát triển, tác giả đã tiến hành một loạt các thử nghiệm đầy thú vị. Đầu tiên, để có dữ liệu cơ sở thực tế, tác giả đã sử dụng phần mềm thu thập dữ liệu để tổng hợp thông tin từ Weibo, mạng xã hội lớn tại Trung Quốc. Thông qua quá trình này, tác giả đã thu thập một loạt các tập dữ liệu đa dạng.

Một trong những điểm đáng chú ý là việc tác giả bổ sung tính di động vào dữ liệu của mỗi người dùng. Điều này có ý nghĩa quan trọng vì tính di động có thể ảnh hưởng đáng kể đến hành vi mạng xã hội. Các tác giả đã quyết định tạo ngẫu nhiên giá trị của tính di động cho từng người dùng trong tập dữ liệu để đảm bảo đa dạng và phản ánh đúng hiện thực.

Tác giả không ngừng nỗ lực trong việc thu thập dữ liệu, và từ phần mềm Weibo, tác giả đã có được nhiều tập dữ liệu khác nhau dựa trên các khoảng thời gian khác nhau. Điều này đã giúp tác giả xây dựng bốn tập dữ liệu mạng xã hội độc lập và phân biệt nhau. Bảng 1 hiển thị thông tin chi tiết về các tập dữ liệu này, bao gồm số lượng node và số lượng edge trong mỗi tập dữ liệu tương ứng với các năm 2010, 2012, 2014 và 2015.

A table with numbers and text

Description automatically generated

Hình 4 Bảng tập dữ liệu của bài báo

# CHƯƠNG 8: THỰC NGHIỆM KHÁM PHÁ CỘNG ĐỒNG

Trong các thí nghiệm phát hiện cộng đồng, tác giả đã tiến hành tiền xử lý trước trên các tập dữ liệu kiểm tra ban đầu theo bảng 1. Để hiển thị sự ảnh hưởng của khả năng cơ động đối với việc phát hiện cộng đồng, họ đã thiết lập giá trị tối đa của khả năng cơ động là 0, 1/2 và 1 tương ứng, sau đó các giá trị khả năng cơ động được tạo ngẫu nhiên và thêm vào các tập dữ liệu kiểm tra ban đầu cho mỗi người dùng. Ví dụ, khi giá trị tối đa của khả năng cơ động là 0.5, các giá trị khả năng cơ động được tạo ngẫu nhiên cho mỗi người dùng phải thuộc về khoảng [0, 0.5]. Bằng cách thiết lập các tập dữ liệu không có khả năng cơ động, có khả năng cơ động thấp và có khả năng cơ động cao, họ có thể hiểu được tác động tổng quan của khả năng cơ động đối với việc phát hiện cộng đồng.

Ngoài ra, trong khuôn khổ mô hình được đề xuất, các hệ số tác động η ∈ [0, 1] và ξ ∈ [0, 1] được sử dụng để làm cân bằng giữa thuộc tính của cộng đồng và thuộc tính của nút cũng như giữa sự tương đồng cạnh và sự tương đồng quan tâm. Do đó, bằng cách thay đổi các hệ số tác động η ∈ [0, 1] và ξ ∈ [0, 1], chúng ta có thể thu được kết quả tổng quan để hiển thị sự ảnh hưởng của các tham số lên quá trình phát hiện cộng đồng.

Hình 2 và Hình 3 dựa trên mô hình được đề xuất, cho thấy số lượng cộng đồng đã phát hiện và số lượng thành viên của cộng đồng tối đa đã phát hiện

A group of colorful bars

Description automatically generated with medium confidence

Hình 5 Kết quả thực nghiệm 2

Từ Hình 2, chúng ta có thể thấy rằng giá trị của khả năng cơ động càng lớn, số lượng cộng đồng đã phát hiện càng ít. Do đó, khả năng cơ động của nút là một yếu tố rất quan trọng đối với quá trình phát hiện cộng đồng trong các mạng xã hội di động. Khả năng cơ động cao hơn của người dùng làm cho người dùng ít có khả năng tham gia cộng đồng. Ngoài ra, như được thể hiện trong Hình 2, mỗi khi chúng ta thay đổi các hệ số tác động η và ξ từ 0,25 đến 0,75 hoặc từ 0,75 đến 0,25, sự thay đổi của số lượng cộng đồng đã phát hiện rất nhỏ. Mặc dù sự thay đổi của các hệ số tác động η và ξ có ảnh hưởng nhỏ đối với việc phát hiện cộng đồng, điều này cho thấy rằng thuộc tính của cộng đồng, thuộc tính của nút và quan tâm của nút đều có ảnh hưởng và quan trọng đối với kết quả của cộng đồng đã phát hiện. Nói cách khác, kết quả của quá trình phát hiện cộng đồng chủ yếu dựa vào các thuộc tính của cộng đồng tương tự với kết quả của quá trình phát hiện cộng đồng chủ yếu dựa vào các thuộc tính của nút, và kết quả của quá trình phát hiện cộng đồng chủ yếu dựa vào các thuộc tính của cộng đồng và các thuộc tính của nút tương tự với kết quả của quá trình phát hiện cộng đồng chủ yếu dựa vào sở thích của nút. Từ toàn bộ Hình 2, có thể thấy rằng khả năng cơ động của nút đóng vai trò quan trọng trong quá trình phát hiện cộng đồng trong các mạng xã hội di động, và thuộc tính của cộng đồng, thuộc tính của nút và sở thích của nút có độ quan trọng tương tự đối với quá trình phát hiện cộng đồng. Do đó, từ kết quả thử nghiệm, chúng ta biết rằng khả năng cơ động của nút là một yếu tố quan trọng để ảnh hưởng đến tính ổn định của cấu trúc cộng đồng. Khi các cộng đồng được phát hiện nghiêm ngặt hơn, các cộng đồng đã phát hiện trở nên ổn định hơn.

A group of colorful bars

Description automatically generated

Hình 6 Kết quả thực nghiệm 3

Như được thể hiện trong Hình 3, chúng ta có thể thấy rằng giá trị của khẳ năng cơ động càng lớn, số lượng thành viên của cộng đồng có kích thước lớn nhất giảm đi. Nói cách khác, khả năng cơ động của nút là yếu tố rất quan trọng để tạo ra các cộng đồng quy mô lớn trong các mạng xã hội di động. Khi hầu hết người dùng có khả năng cơ động cao, kích thước của các cộng đồng đã phát hiện thường nhỏ hơn. Tương tự như Hình 2, khi các hệ số tác động η và ξ thay đổi, kết quả phát hiện cộng đồng tương tự trong Hình 3. Nói cách khác, các thuộc tính của cộng đồng, thuộc tính của nút và sở thích của nút đều có ảnh hưởng tương tự đối với số lượng thành viên của cộng đồng có kích thước lớn nhất. Từ toàn bộ Hình 3, có thể thấy rằng khả năng cơ động của nút là yếu tố quan trọng trong việc xác định kích thước của cộng đồng đã phát hiện trong các mạng xã hội di động, và các thuộc tính của cộng đồng, thuộc tính của nút và sở thích của nút có độ quan trọng tương tự để tạo ra các cộng đồng quy mô lớn. Tương tự, chúng ta có thể biết rằng khả năng cơ động của nút là yếu tố quan trọng để ảnh hưởng đến tính ổn định của cộng đồng quy mô lớn trong các mạng xã hội di động. Do đó, đề xuất của tác giả cho phương pháp phát hiện cộng đồng có thể làm cho quá trình phát hiện cộng đồng trở nên toàn diện và hợp lý hơn bằng cách giới thiệu thuộc tính của cộng đồng và khẳ năng cơ động của nút. Đề xuất của tác giả có thể giảm khả năng tạo ra các cộng đồng quy mô lớn để cải thiện tính chính xác và đáng tin cậy của quá trình phát hiện cộng đồng. Trong thử nghiệm, tác giả cũng so sánh các phương pháp GN [9] và NM [10] với phương pháp phát hiện cộng đồng của tác giả dựa trên số lượng cộng đồng đã phát hiện và số lượng thành viên của cộng đồng có kích thước lớn nhất. Phương pháp GN là một thuật toán phát hiện cấu trúc cộng đồng dựa trên phân cụm cạnh. Phương pháp NM là một thuật A screenshot of a graph

Description automatically generatedtoán phát hiện cộng đồng dựa trên mô-đun mạng, là một thuật toán tham lam.

Hình 7 Kết quả thực nghiệm 4

Như thể hiện trong Hình 4, phương pháp đề xuất của tác giả có thể phát hiện cộng đồng một cách chính xác dựa trên thuộc tính của cộng đồng và khẳ năng cơ động của nút, số lượng cộng đồng đã phát hiện và số lượng thành viên trong cộng đồng có kích thước lớn nhất tạo ra bởi phương pháp của tác giả đều nhỏ hơn so với GN và NM. Vì vậy, phương pháp của tác giả có thể thực hiện việc phát hiện cộng đồng một cách chính xác và nghiêm ngặt hơn để tăng tính ổn định của cấu trúc cộng đồng. Dựa trên các tập dữ liệu kiểm tra được hiển thị trong Bảng 1, tác giả kiểm tra thời gian chạy của phương pháp của tác giả. Trong bài kiểm tra, tác giả đặt các hệ số tác động η = 0.5 và ξ = 0.5, và đặt giá trị tối đa của khả năng cơ động là 0, 1/2 và 1 tương ứng. Trong các tập dữ liệu, mức độ trung bình của các nút là 5, và số lượng cạnh thay đổi từ 5000 đến 20000. Hình 5 cho thấy kết quả kiểm tra. Từ Hình 5, có thể thấy rằng khi giá trị của khả năng cơ động thuộc cùng một khoảng, thời gian chạy của phương pháp của tác giả tăng theo số lượng cạnh. Hình 5 cũng cho thấy rằng khi số lượng cạnh cố định, nếu giá trị của khả năng cơ động thường lớn hơn, thì thời gian chạy của phương pháp của tác giả trở nên nhỏ hơn. Điều này bởi vì khi hầu hết người dùng có giá trị di động cao hơn, kích thước của các cộng đồng đã phát hiện trở nên nhỏ hơn, do đó phương pháp của tác giả dễ dàng hoàn thành việc phát hiện cộng đồng.

Hơn nữa, chúng ta có thể biết rằng mặc dù thời gian chạy của phương pháp của tác giả tăng theo số lượng cạnh, giá trị khả năng cơ động lớn hơn trong các tập dữ liệu kiểm tra có thể làm giảm sự tăng thời gian được tạo ra với số lượng cạnh. Ví dụ, các thử nghiệm cho thấy thời gian chạy của phương pháp đề xuất của tác giả là khoảng 17 giây khi giá trị tối đa của khả năng cơ động được đặt là 0.5 và số lượng cạnh là khoảng 16000, và thời gian chạy chỉ khoảng 13 giây khi giá trị tối đa của khả năng cơ động được đặt là 1. Do đó, khi các nút đều có khả năng cơ động cao trong các mạng xã hội di động, hiệu suất của phương pháp của tác giả trở nên tốt hơn.

# CHƯƠNG 9: KẾT QUẢ THỰC NGHIỆM

Trong các mạng xã hội di động, nhiều yếu tố có thể ảnh hưởng đến việc phân loại cộng đồng, trong đó độ di động của nút là một yếu tố chính ảnh hưởng đến sự ổn định của cấu trúc cộng đồng. Trong bài báo đề xuất một kế hoạch phát hiện cộng đồng xã hội cho các mạng xã hội di động dựa trên phần mềm xã hội, bao gồm sự tương đồng về thuộc tính xã hội, sự tương đồng về quan tâm của nút và di động của nút.

Kế hoạch được đề xuất chia các thuộc tính xã hội của một người dùng thành các thuộc tính cộng đồng xã hội và thuộc tính nút, và tiếp tục xem xét sự di động của mỗi nút. So với các kế hoạch phát hiện cộng đồng khác, kế hoạch của tác giả có thể phát hiện chính xác các cộng đồng dựa trên thuộc tính cộng đồng và di động của nút. Các thí nghiệm cho thấy số lượng cộng đồng phát hiện được và các thành viên trong cộng đồng có kích thước lớn nhất do kế hoạch của tác giả tạo ra đều nhỏ hơn so với các kế hoạch GN và NM. Ngoài ra, do các nút có độ di động cao trong các mạng xã hội di động, hiệu suất của kế hoạch đề xuất của tác giả tương đối cao hơn.

Các thí nghiệm cho thấy khi giá trị di động trong các bộ dữ liệu thử nghiệm tăng lên, thời gian chạy của kế hoạch đề xuất của tác giả giảm khi số lượng cạnh được cố định. Do đó, kế hoạch của tác giả có thể phát hiện cộng đồng một cách chính xác và hiệu quả hơn để tăng sự ổn định của cấu trúc cộng đồng di động.

Bên cạnh đó, phương pháp đề xuất của tác giả giúp tách biệt các thuộc tính xã hội của người dùng thành các thuộc tính cộng đồng xã hội và thuộc tính của nút, mở rộng khả năng hiểu biết về cộng đồng và quan hệ giữa các thành viên. Tác giả cũng tích hợp sự di động của mỗi nút vào quá trình phân loại cộng đồng, tạo ra sự linh hoạt và ổn định trong cấu trúc cộng đồng di động.

So với các phương pháp khác như GN và NM, kết quả thử nghiệm của tác giả cho thấy số lượng cộng đồng được phát hiện và số thành viên trong cộng đồng lớn nhất có kích thước đều ít hơn. Điều này chứng minh hiệu quả của phương pháp của tác giả trong việc giảm kích thước cộng đồng, giúp tăng khả năng quản lý và theo dõi các nhóm trong mạng xã hội di động.

Đặc biệt, khi giá trị di động tăng trong các tập dữ liệu kiểm thử, thời gian chạy của phương pháp của tác giả giảm, đồng thời đảm bảo độ chính xác của quá trình phân loại cộng đồng. Điều này thể hiện tính linh hoạt và hiệu suất của phương pháp tác giả, làm tăng ổn định của cấu trúc cộng đồng trong môi trường mạng xã hội di động đầy động lực.

Hơn nữa, phương pháp của tác giả không chỉ tập trung vào việc phân loại cộng đồng dựa trên thuộc tính xã hội mà còn kết hợp với sự quan tâm và di động của từng nút trong mạng. Sự tương đồng về quan tâm giúp xác định nhóm người có sở thích và mục tiêu chung, từ đó tạo ra cộng đồng có mối liên kết sâu sắc hơn. Đồng thời, việc tích hợp di động của nút vào quá trình phân loại cộng đồng giúp đảm bảo rằng sự thay đổi về vị trí không gây ảnh hưởng đáng kể đến cấu trúc cộng đồng.

Kết quả thử nghiệm của tác giả cũng cho thấy rằng, so với các phương pháp GN và NM, phương pháp của tác giả không chỉ giảm số lượng cộng đồng mà còn giảm số lượng thành viên trong cộng đồng lớn nhất. Điều này chứng tỏ khả năng tối ưu hóa cấu trúc cộng đồng của phương pháp, giúp người dùng dễ dàng theo dõi và tham gia vào các nhóm có liên quan trong mạng xã hội di động.

Tóm lại, phương pháp phân loại cộng đồng mà tác giả đề xuất không chỉ chú trọng vào thuộc tính xã hội mà còn kết hợp sự quan tâm và di động, từ đó tạo ra một cách tiếp cận toàn diện và hiệu quả. Kết quả thử nghiệm chứng minh tính linh hoạt và khả năng ổn định của phương pháp trong việc phân loại cộng đồng trong môi trường đầy động lực của mạng xã hội di động.

# CHƯƠNG 10: MÔ PHỎNG NGHIÊN CỨU

## 10.1 Mô tả dữ liệu

### 10.1.1 Giới thiệu nguồn dữ liệu

Link Dataset: https://snap.stanford.edu/data/ego-Facebook.html

### 10.1.2 Mô tả dữ liệu

Dataset gồm 4039 nodes, 88234 cạnh với 2 cột dữ liệu. Mỗi hàng đại diện cho một người dùng có mối quan hệ bạn bè với người dùng khác.

Ảnh có chứa văn bản, số, Phông chữ, ảnh chụp màn hình

Mô tả được tạo tự động

Hình 8 Bảng dữ liệu

|  |  |
| --- | --- |
| **Thuộc tính** | **Ý nghĩa** |
| **FromNodeID** | ID người dùng ban đầu |
| **ToNodeID** | ID người dùng có mối quan hệ bạn bè với người dùng ban đầu |

Nguồn trích dẫn : J. McAuley and J. Leskovec. Learning to Discover Social Circles in Ego Networks. NIPS, 2012.

### 10.1.3 Tiền xử lý dữ liệu

Theo như bài báo tác giả thêm tính di động như một thuộc tính của mỗi người dùng, với các giá trị di động được tạo ngẫu nhiên cho mỗi người dùng. Vì vậy chúng ta sẽ tạo thêm một thuộc tính là tính di động và nó sẽ ngẫu nhiên trong khoảng 0, 0.5 và 1.

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ

Mô tả được tạo tự động

Hình 9 Tiền xử lý dữ liệu

Kết quả sau khi thêm thuộc tính di động

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

Hình 10 Kết quả xử lý dữ liệu

## 10.2 Độ đo trung tâm và trực quan hóa

### 10.2.1 Degree Centrality

+ Trực quan hóa kết quả độ đo

Ảnh có chứa màu tím, Nhiều màu sắc, hoa tím

Mô tả được tạo tự động

Hình 11 Trực quan độ đo Degree Centrality

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động+ Top 10 nodes có Degree cao nhất

Hình 12 Top 10 node Degree cao nhất

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động- Tần số và xác suất xuất hiện của từng độ đo Degree

Hình 13 Tần số node Degree xuất hiện nhiều nhất

Nhận xét: Top 3 nodes có các ID: 107,1684 ,1912 có độ đo Degree cao nhất. Hay nói cách khác đây là những ID đi kết bạn nhiều nhất.

### 10.2.2 Betweeness Centrality

+ Trực quan hóa kết quả độ đo

Ảnh có chứa màu tím, hoa tím, Nhiều màu sắc, ảnh chụp màn hình

Mô tả được tạo tự động

Hình 14 Trực quan độ đo Betweeness Centrality

A screenshot of a graph

Description automatically generated+ Top 10 nodes có Betweeness cao nhất

Hình 15 Top 10 node Betweeness cao nhất

- Tần số và xác suất xuất hiện của từng độ đo Betweeness

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

Hình 16 Tần số node Betweeness xuất hiện nhiều nhất

Nhận xét: Node 1684 có độ đo Betweenness Centrality cao nhất tức là người có ID=1684 có tầm quan trọng phổ biến về mức độ kết bạn giữa các kết nối các ID khác trong dữ liệu mạng xã hội Facebook.

### 10.2.3 Closeness Centrality

+ Trực quan hóa kết quả độ đo

Ảnh có chứa màu tím, Nhiều màu sắc, hoa tím, Hoa tử đinh hương

Mô tả được tạo tự động

Hình 17 Trực quan độ đo Closeness Centrality

+ Top 10 nodes có Closeness cao nhất

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

Hình 18 Top 10 node Closeness cao nhất

- Tần số và xác suất xuất hiện của từng độ đo Closeness

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

Hình 19 Tần số node Closeness xuất hiện nhiều nhất

Nhận xét : Có nhiều node có độ đo Closeness Centrality cao bằng 1.0 thể hiện mức độ kết nối cao của một nút với mọi nút khác trong mạng.

## 10.3 Phát hiện cộng đồng và nhận xét

### 10.3.1 Girvan Newman

Ảnh có chứa văn bản, Phần mềm đa phương tiện, Phông chữ, phần mềm

Mô tả được tạo tự động

Kết quả thuật toán Girvan Newman phát hiện được 2 cộng đồng từ tập dữ liệu

Ảnh có chứa màu xanh lá cây, thực vật

Mô tả được tạo tự động+ Group 0: gồm các nodes: {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15…} + Group 1: gồm các nodes: {686, 687, 688, 689, 690, 691, 692, 693, 694…}

Hình 20 Kết quả cộng đồng Girvan Newman

### 10.3.2 Newman Modularity

Ngược lại khi sử dụng Newman Modularity ta phát hiện được 7 cộng đồng trong tập dữ liệu:  
Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ

Mô tả được tạo tự độngẢnh có chứa Nhiều màu sắc, Nghệ thuật phân dạng, tác phẩm nghệ thuật

Mô tả được tạo tự động

Hình 21 Kết quả cộng đồng Newman Modularity

### 10.3.3 Thuật toán của bài báo

Khởi tạo lại đồ thị sao cho giữ nguyên các thuộc tính di động của từng nodes sau đó khởi tạo cộng đồng

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Phông chữ

Mô tả được tạo tự động

Lúc này hàm phát hiện cộng đồng động sẽ nhận 3 tham số đầu vào là đồ thị G, cộng đồng được khởi tạo và eta mặc định là 0.5

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, phần mềm

Mô tả được tạo tự động

Ở hàm này xây dựng như mã giả mà bài báo đã cung cấp đầu tiên Đây là một vòng lặp while với điều kiện là flag khác 1. Thuật toán sẽ tiếp tục chạy trong vòng lặp này cho đến khi flag đạt giá trị 1. Vòng lặp chính qua tất cả các nút và cộng đồng trong đồ thị. Vòng lặp lồng bên trong để tính giá trị CNSi\_z cho mỗi cặp (node\_i, community) như trước. Sau khi vòng lặp while kết thúc, gọi hàm module\_degree\_base\_division để xử lý kết quả và cập nhật giá trị flag và số cộng đồng phát hiện.

+ Kết quả số cộng đồng phát hiện: 5

Ảnh có chứa ảnh chụp màn hình, văn bản, Phần mềm đa phương tiện, Phần mềm đồ họa

Mô tả được tạo tự động

+ Trực quan kết quả cộng đồng được phát hiện:  
Ảnh có chứa hoa

Mô tả được tạo tự động

Hình 22 Kết quả cộng đồng di động

# CHƯƠNG 11: ĐÁNH GIÁ PHƯƠNG PHÁP

## 10.1 Ưu điểm của phương pháp

1. Sử dụng thuộc tính cộng đồng và khả năng cơ động của nút:Hệ thống sử dụng cả thuộc tính của cộng đồng và khả năng cơ động của người dùng để phát hiện cộng đồng, giúp tạo ra các kết quả phát hiện cộng đồng chính xác hơn và ổn định hơn trong mạng xã hội di động.

2. Kiểm soát khả năng cơ động: Hệ thống cho thấy khả năng cơ động của người dùng đóng một vai trò quan trọng trong quá trình phát hiện cộng đồng. Khả năng cơ động cao hơn của người dùng làm cho các cộng đồng có kích thước nhỏ hơn, giúp kiểm soát kích thước của cộng đồng và làm cho phát hiện cộng đồng trở nên ổn định hơn.

3. Thời gian chạy hiệu quả: Hệ thống cho thấy rằng khi khả năng cơ động cao hơn trong các tập dữ liệu kiểm tra, hiệu suất của phương pháp tốt hơn. Điều này có thể giảm thời gian chạy của phương pháp và làm cho việc phát hiện cộng đồng trở nên hiệu quả hơn.

4. So sánh với các phương pháp khác: Hệ thống so sánh phương pháp của mình với các phương pháp khác như GN và NM dựa trên số lượng cộng đồng đã phát hiện và số lượng thành viên của cộng đồng có kích thước lớn nhất. Kết quả cho thấy phương pháp của hệ thống có khả năng phát hiện cộng đồng chính xác hơn.

5. Cải thiện tính chính xác và đáng tin cậy: Hệ thống giới thiệu thuộc tính của cộng đồng và khả năng cơ động của nút để giảm khả năng tạo ra các cộng đồng quy mô lớn, từ đó cải thiện tính chính xác và đáng tin cậy của quá trình phát hiện cộng đồng.

## 10.2 Nhược điểm của phương pháp

Phương pháp sử dụng thuộc tính cộng đồng và khả năng cơ động có thể làm tăng độ phức tạp của thuật toán phát hiện cộng đồng. Việc tích hợp nhiều yếu tố có thể làm cho phương pháp trở nên phức tạp hơn và đòi hỏi sự quản lý kỹ lưỡng.

Hiệu suất của hệ thống có thể phụ thuộc vào cấu hình của mạng xã hội di động cụ thể. Kết quả của hệ thống có thể khác nhau tùy thuộc vào cấu trúc và đặc điểm của mạng xã hội di động cụ thể mà nó được áp dụng.

## 10.3 Kết luận

Phát hiện cộng đồng trong mạng xã hội di động là một quá trình quan trọng để tìm hiểu cấu trúc và chức năng của các đơn vị trong mạng xã hội. Trong môi trường này, có nhiều yếu tố có thể ảnh hưởng đến quá trình phát hiện cộng đồng, và khả năng di chuyển của các thành viên trong mạng xã hội di động đóng một vai trò quan trọng trong việc đảm bảo tính ổn định của cấu trúc cộng đồng.

Các tác giả đã đề xuất một phương pháp phát hiện cộng đồng trong mạng xã hội di động dựa trên sự kết hợp của các yếu tố bao gồm sự tương đồng về thuộc tính xã hội, sự tương đồng về sở thích của người dùng và khả năng di chuyển của họ. Phương pháp của các tác giả phân chia thuộc tính xã hội thành hai phần, gồm thuộc tính của cộng đồng và thuộc tính của từng thành viên, và đồng thời xem xét mức độ di chuyển của từng thành viên.

Kết quả của các tác giả đã cho thấy rằng phương pháp của các tác giả có khả năng phát hiện cộng đồng một cách chính xác dựa trên thuộc tính của cộng đồng và khả năng di chuyển của các thành viên. So với các phương pháp khác như GN và NM, số lượng cộng đồng đã phát hiện và số lượng thành viên trong cộng đồng có kích thước lớn nhất đều thấp hơn. Hơn nữa, khi khả năng di chuyển của thành viên tăng trong các tập dữ liệu kiểm tra, thời gian chạy của phương pháp của các tác giả giảm, làm tăng hiệu suất của phương pháp.

Dựa vào kết quả demo code của Girvan Newman và Newman Modularity, có thể thấy rằng phương pháp đề xuất trong bài báo không chỉ phát hiện số lượng cộng đồng khác nhau so với hai phương pháp truyền thống này mà còn thể hiện sự linh hoạt và hiệu quả trong việc phân loại cộng đồng. Sự kết hợp giữa thuộc tính xã hội và khả năng di chuyển của nút trong mạng xã hội di động đã giúp phương pháp của tác giả đạt được hiệu suất tốt hơn.

Kết quả thực tế chỉ ra rằng, khi áp dụng phương pháp này, số lượng cộng đồng được phát hiện có sự khác biệt so với những kết quả từ Girvan Newman và Newman Modularity. Điều này cho thấy sự độc đáo và độ linh hoạt của phương pháp trong việc thích ứng với đặc điểm cụ thể của mạng xã hội di động.

Trong khi Girvan Newman và Newman Modularity có thể gặp khó khăn khi đối mặt với các tình huống phức tạp và động, phương pháp đề xuất đã cải thiện tính chính xác và độ ổn định của quá trình phát hiện cộng đồng. Sự khả năng điều chỉnh linh hoạt của phương pháp đã giúp nó thích ứng tốt hơn với sự biến động và đa dạng trong môi trường mạng xã hội di động.

Với những điểm mạnh như sử dụng thuộc tính cộng đồng và khả năng cơ động của nút, khả năng kiểm soát kích thước cộng đồng, thời gian chạy hiệu quả, và so sánh tính chính xác với các phương pháp khác, phương pháp này không chỉ là một bước tiến trong nghiên cứu mà còn mang lại giá trị thực tiễn trong việc phân tích cộng đồng trên mạng xã hội di động.

# CHƯƠNG 12: TÀI LIỆU THAM KHẢO

[1] Social Community Detection Scheme Based on Social-Aware in Mobile Social Networks. Ke Gu; Dianxing Liu; Keming Wang <https://ieeexplore.ieee.org/document/8915780>

[2] T. Qiu, B. Chen, A. K. Sangaiah, J. Ma, and R. Huang, "A survey of mobile social networks: Applications, Social Characteristics, and challenges," IEEE Syst. J., vol. 12, no. 4, pp. 3932–3947, Dec. 2018.

[3] K. Zhu, W. Li, X. Fu, and L. Zhang, "Data routing strategies in opportunistic mobile social networks: Taxonomy and open challenges," Comput. Netw., vol. 93, pp. 183–198, Oct. 2015.

[4] R. Fei, K. Yang, and X. Cheng, "A cooperative social and vehicular network and its dynamic bandwidth allocation algorithms," in Proc. IEEE Conf. Compute. Communi. Workshops (INFOCOM WORKSHOPS), Apr. 2011, pp. 63–67.

[5] M. Baqer and A. Kamal, "S-Sensors: Integrating physical world inputs with social networks using wireless sensor networks," in Proc. Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process. (ISSNIP), Dec. 2009, pp. 213–218.

[6] Y. Zhang, E. Pan, L. Song, W. Saad, Z. Dawy, and Z. Han, "Social network enhanced device-to-device communication underlaying cellular networks," in Proc. IEEE/CIC Int. Conf. Communi. China-Workshops (CIC/ICCC), Aug. 2013, pp. 182–186.

[7] L. Maccari, "An analysis of the Linux wireless community network," in Proc. IEEE 9th Int. Conf. Wireless Mobile Comput., Netw. Communi. (WiMob), Oct. 2013, pp. 1–7.

[8] W.-J. Hsu and A. Helmy, "On nodal encounter patterns in wireless LAN traces," IEEE Trans. Mobile Comput., vol. 9, no. 11, pp. 1563–1577, Nov. 2010.

[9] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," Proc. Nat. Acad. Sci. USA, vol. 99, no. 12, pp. 7821–7826, Apr. 2002.

[10] M. E. Newman, "Analysis of weighted networks," Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 70, no. 5, Nov. 2004, Art. no. 056131.

[11] R. Shang, S. Luo, W. Zhang, R. Stolkin, and L. Jiao, "A multiobjective evolutionary algorithm to find community structures based on affinity propagation," Phys. A, Stat. Mech. Appl., vol. 453, pp. 203–227, Jul. 2016.

[12] F. Buccafurri, L. Fotia, G. Lax, and V. Saraswat, "Analysis-preserving protection of user privacy against information leakage of social-network likes," Inf. Sci., vol. 328, pp. 340–358, Jan. 2016.

[13] C. Guo, B. Li, and X. Tian, "Flickr group recommendation using rich social media information," Neurocomputing, vol. 204, pp. 8–16, Sep. 2016.

[14] H. Shen, X. Cheng, K. Cai, and M.-B. Hu, "Detect overlapping and hierarchical community structure in networks," Phys. A, Stat. Mech. Appl., vol. 388, no. 8, pp. 1706–1712, Apr. 2009.

[15] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," J. Stat. Mech., Theory Exp., vol. 2008, no. 10, 2008, Art. no. P10008.

**—HẾT—**