Encyclopedia of Digital Transformation
Vol. 3; No. 1; PP. 97-122
Spring 2022
DOI: 10.22034/dtj.2026.428665.1127

MehrAlborz University

A Study on Community Detection in Social Network

Safoura Akhlaghi [1], Mohammad Bagher Menhaj [2], Behrooz Masoumi [3]

Abstract

A social network is a social structure that is composed of various individual or organizational groups. The use of social networks has increased widely in recent years. One of the active and challenging research areas in the discussion of social networks is the issue of community detection. The general meaning of community in a social network is the gathering of a number of agents together in such a way that the members of each community have the most interactions with each other. Humans tend to form groups in these networks based on their similar interests. Such groups are known as communities or clusters. Detecting such a structure gives us an exceptional understanding of the organization and functioning of social networks. One of the most important features in these networks is the existence of community structure. In recent years, many community detection algorithms have been proposed to reveal the structural features and dynamic behaviors of networks. The main objective of this paper is to provide an overview of community detection algorithms and to examine the strengths and weaknesses of each community detection approach, ranging from traditional algorithms to advanced algorithms for overlapping community detection. Algorithms based on dimensionality reduction techniques such as non-negative matrix factorization (NMF) and principal component analysis (PCA) are also considered.

Keywords: *Community detection, static and dynamic networks, community detection algorithms, social network analysis*

1. Department of Computer Engineering and Information Technology, Qazvin Branch, Islamic Azad University, Qazvin, Iran.

2. Faculty member, Amirkabir University of Technology, Tehran, Iran.

3. Faculty member, Qazvin Islamic Azad University, Qazvin, Iran.

Submitted: 2023-12-22
Accepted:2025-11-23
Corresponding Author: Safoura Akhlaghi
Email: safoura.akhlaghi@gmail.com

دانشنامهٔ تحول دیجیتال

دانشگاه مهرالبرز

دورهٔ ۳، شمارهٔ ۱

بهار ۱۴۰۱

صص. ۹۷- ۱۲۲

Encyclopedia of Digital Transformation

DOI: 10.22034/dtj.2026.428665.1127

بررسی تشخیص جوامع در شبکه‌های اجتماعی

صفورا اخلاقی[۱]، محمدباقر منهاج[۲]، بهروز معصومی[۳]

چکیده

شبکه اجتماعی یک ساختار اجتماعی است که از گروه‌های مختلف فردی یا سازمانی تشکیل شده است. استفاده از شبکه‌های اجتماعی در سال‌های اخیر به طور چشمگیری افزایش پیدا کرده است. یکی از حوزه‌های تحقیقاتی فعال و چالش‌برانگیز در زمینه شبکه‌های اجتماعی، موضوع تشخیص جوامع است. به طور کلی، جامعه در یک شبکه اجتماعی به مجموعه‌ای از عامل‌ها یا گره‌ها اطلاق می‌شود که به گونه‌ای در کنار یکدیگر قرار گرفته‌اند که اعضای هر جامعه بیشترین میزان تعامل و ارتباط را با یکدیگر دارند. انسان‌ها تمایل دارند در این شبکه‌ها، گروه‌هایی را بر اساس علایق مشابه خود تشکیل دهند. چنین گروه‌هایی به عنوان جوامع یا خوشه‌ها شناخته می‌شوند. تشخیص چنین ساختاری به ما درک استثنایی از سازمان و عملکرد شبکه‌های اجتماعی می‌دهد. یکی از مهم‌ترین ویژگی در این شبکه‌ها وجود ساختار جامعه است. در سال‌های اخیر، الگوریتم‌های متعددی برای تشخیص جوامع با هدف آشکارسازی ویژگی‌های ساختاری و رفتارهای پویای شبکه‌ها ارائه شده‌اند. هدف اصلی این مقاله، ارائه مروری بر الگوریتم‌های تشخیص جامعه و بررسی نقاط قوت و ضعف هر یک از رویکردها است؛ به‌گونه‌ای که این بررسی طیفی از الگوریتم‌های سنتی تا الگوریتم‌های پیشرفته برای شناسایی جوامع هم‌پوشان را در بر می‌گیرد. الگوریتم‌های مبتنی بر تکنیک‌های کاهش ابعاد مانند فاکتورگیری ماتریس غیر منفی (NMF) و تحلیل مؤلفه‌های اصلی (PCA) نیز مورد توجه قرار گرفته‌اند.

کلمات کلیدی: *تشخیص جامعه، شبکه‌های استاتیک و پویا ، الگوریتم‌های تشخیص جامعه، تحلیل شبکه اجتماعی*

---

۱. گروه مهندسی کامپیوتر و فناوری اطلاعات، واحد قزوین، دانشگاه آزاد اسلامی، قزوین، ایران.

۲. هیات علمی، دانشگاه صنعتی امیرکبیر، تهران، ایران.

۳. هیات علمی، دانشگاه آزاد اسلامی واحد قزوین، قزوین، ایران.

تاریخ دریافت مقاله: ۱۴۰۲/۱۰/۰۱

تاریخ پذیرش نهایی مقاله: ۱۴۰۴/۰۹/۰۲

نویسندهٔ مسئول مقاله: صفورا اخلاقی

Email: safoura.akhlaghi@gmail.com

مقدمه

علم شبکه به مطالعه شبکه‌ها با بهره‌گیری از نظریه‌های ریاضی می‌پردازد و بر تجزیه‌وتحلیل ساختار و توصیف رفتار شبکه‌ها تمرکز دارد. مطالعه شبکه‌ها شاهد تحقیقات قابل‌توجهی با تمرکز بر درک و ارزیابی ویژگی‌های آماری شبکه‌های مقیاس بزرگ بوده است (نیومن[1] ۲۰۰۳). شبکه‌ها معمولا با گراف‌هایی نشان داده می‌شوند که در آن گروهی از گره‌ها (راس) بین آن‌ها (لبه‌ها) ارتباط دارند. نظریه گراف، ریاضیات شبکه‌ها است که برای مدل‌سازی گراف‌ها مورد استفاده قرار می‌گیرد (استام[2] ۲۰۱۴). (ادروس و رنی[3] ۱۹۵۹) گراف‌های تصادفی را معرفی کردند که در آن‌ها احتمال لبه بین دو گره برای هر جفت دیگری یکسان است؛ اما شبکه‌های دنیای واقعی گراف‌ها تصادفی نیستند. یکی از ویژگی‌های مرتبط شبکه‌های دنیای واقعی این است که ساختارهای اجتماعی را می‌توان با استفاده از گراف‌ها به خوبی مدل‌سازی کرد (فورتوناتو[4] ۲۰۱۰). به طور کلی، جامعه یا خوشه به گروهی از گره‌ها اطلاق می‌شود که نسبت به سایر بخش‌های شبکه، وابستگی‌ها و ارتباطات مشابهی با یکدیگر دارند. شناسایی ساختارهای جامعه، گامی به سوی درک ساختارهای مختلف شبکه‌ها با کاربرد در تعدادی از زمینه‌ها مانند: شبکه‌های اجتماعی آنلاین و کلیه علوم فیزیکی و زیستی است. تشخیص جامعه به روش شناسایی گروه‌های رئوس متقابل (یعنی گره‌ها) در یک شبکه بسته به ویژگی‌های ساختاری آن‌ها اشاره دارد (یانگ، مک آلی، لسکووک[5] ۲۰۱۳). الگوریتم‌های متعددی برای تشخیص جوامع با بهره‌گیری از تکنیک‌ها و ابزارهای حوزه‌های مختلفی همچون زیست‌شناسی، فیزیک، علوم اجتماعی، ریاضیات کاربردی و علوم رایانه توسعه یافته‌اند. با این حال، هیچ الگوریتم واحدی برای تشخیص جامعه وجود ندارد که در تمامی انواع شبکه‌ها عملکرد مطلوبی داشته باشد؛ زیرا شبکه‌های پیچیده از تنوع بالایی برخوردارند و از فرایندهای گوناگونی شکل می‌گیرند.

در سال‌های اخیر، حوزه شبکه‌های اجتماعی توجه گسترده‌ای را در میان جوامع پژوهشی به خود جلب کرده است. افراد از حساب‌های کاربری خود در شبکه‌های اجتماعی برای حفظ، تقویت و ایجاد روابط با سایر اعضای جامعه استفاده می‌کنند. شکل‌گیری و گسترش این روابط، توجه پژوهشگران را به توسعه حوزه‌ای با عنوان تحلیل شبکه‌های اجتماعی SNA معطوف کرده است. SNA به عنوان مطالعه و درک روابط و پیوندهای بین دو یا چند فرد در یک شبکه اجتماعی توصیف می‌شود (روما و رومدان[6] ۲۰۱۸). SNA مجموعه‌ای از ابزارها و رویکردها را برای تحلیل بصری روابط شبکه‌های اجتماعی ارائه می‌دهد. این روابط ساختار ساختمان جوامع شبکه‌های اجتماعی است. طبق گفته (فورتوناتو، ۲۰۱۰) یک جامعه را می‌توان به عنوان گروهی از افراد توصیف کرد که احتمالا مسئولیت‌های مشابهی را در شبکه به اشتراک می‌گذارند. جوامع با استفاده از روش‌های تشخیص مختلف که برای چنین اهدافی طراحی شده‌اند، شناسایی می‌شوند. تشخیص جامعه درک ساختار اجتماعی را تسهیل و اطلاعات مربوط به کاربر را در داخل شبکه آشکار می‌کند (الواری، حاجی‌باقری، سوکتانکار[7] ۲۰۱۴). تشخیص جوامع را می‌توان در زمینه‌های مختلف زیست‌شناسی، جامعه‌شناسی و حتی بازاریابی استفاده کرد. زمانی که شرکت‌ها می‌توانند سیستم‌های توصیه‌ای مناسب و کارآمد را برای مشتریان فراهم کنند، به ارتقاء صنعت هم کمک می‌کند. گروه‌بندی افراد با علایق مشابه، جامعه را توسعه می‌دهد. در اوایل دهه ۱۹۵۰، محققان رشته‌های مختلف به طور گسترده بررسی کردند که چگونه روابط و تعاملات کاربران در شبکه‌های اجتماعی می‌تواند به طور مؤثری از متفاوت‌ترین تجزیه‌وتحلیل‌ها، از بازاریابی گرفته تا یافته‌های متخصص، همچنین بسته به نوع خاص شبکه تحلیل شده، پشتیبانی کند. یکی از اولین تعاریف شبکه اجتماعی که اهمیت مرتبط را بهتر نشان می‌دهد توسط میچل

1 Newman
2 Stam
3 Erdos &Renyi
4 Fortunato
5 Yang & McAuley & Leskovec
6 Rhouma& Romdhane
7 Alvari & Hajibagheri & Sukthankar

ارائه شد که آن را به عنوان «مجموعه خاصی از پیوندها در میان مجموعه ای از افراد تعریف شده، با ویژگی اضافی که ویژگی‌های این پیوندها به عنوان یک کل، ممکن است برای تفسیر رفتار اجتماعی افراد استفاده شود.» شبکه اجتماعی همچنین ساختاری اجتماعی است که از افراد (یا سازمان‌ها) به نام «گره‌ها» تشکیل شده است که توسط یک یا چند نوع خاص از وابستگی متقابل، «لبه‌ها» به هم متصل شده‌اند. خویشاوندی، منافع مشترک، مبادله مالی، دانش یا اعتبار. اخیرا، انتشار شبکه‌های اجتماعی آنلاین (OSN) منجر به توسعه یک ابزار ارتباطی جدید شده است که به مردم در سراسر جهان اجازه می‌دهد تا بین آن‌ها تعامل داشته باشند و اطلاعات را منتشرو یا به اشتراک بگذارند.

بنابراین، در طول دو دهه اخیر، مطالعه روابط انسانی با ظهور OSNها کاملا متحول شده است و تجزیه‌وتحلیل شبکه‌های اجتماعی (SNA) با هدف آشکارسازی ترجیحات، رفتار، سلایق، نیازها، علایق کاربران، توجه محققان را به خود جلب کرده است. به عنوان یک نقطه چالش‌برانگیز، افزایش مداوم تعداد افرادی که در این محیط‌ها ثبت‌نام می‌کنند و اطلاعات ناهمگن مشترک، نشان می‌دهد که مدیریت تعاملات بین کاربران بسیار دشوار است و OSNها سیستم‌های بسیار پیچیده‌ای را برای مدیریت نشان می‌دهند (موسکاتو، اسپرلی[1] ۲۰۲۱)

حوزه تشخیص جامعه به سرعت در حال تکامل است و نقش فوق‌العاده‌ای در حل مسائل پیچیده مختلف زندگی واقعی ایفا می‌کند. مطالعات، مقالات مروری و همچنین کتاب‌های مختلفی در مورد کاربردها و روش‌های تشخیص جامعه وجود دارد. اکثر مقالات مروری، الگوریتم‌های تشخیص جامعه را برای نوع خاصی از شبکه‌ها مانند شبکه‌های اجتماعی، شبکه‌های تحمل‌پذیر در برابر تأخیر و شبکه‌های تلفن همراه مطالعه می‌کنند.

پیشینه پژوهش

با توجه به موضوع، مقایسه مقالات مروری در جدول ۱ نشان داده شده است.

جدول ۱: مقایسه مقالات مروری

مقاله مروری	سال	حوزه های مورد تمرکز	حوزه‌هایی که روی آن‌ها تمرکز نشده است
Danon و همکاران	۲۰۰۵	مقایسه الگوریتم‌های رایج شناسایی جوامع در آن زمان از نظر هزینه محاسباتی، بازنگری معیار مدولاریتی برای شناسایی جوامع	روش‌های شناسایی جوامع همپوشان بررسی نشده، مدل‌های تشخیص ناهنجاری لحاظ نشده‌اند
Pons	۲۰۰۷	بحث درباره الگوریتم‌های شناسایی جوامع پویا با استفاده از الگوریتم‌های نوع "راه رفتن تصادفی"	تنها روش‌های تقسیم‌بندی گراف بررسی شدند
Gulbashe و Lehmann	۲۰۰۸	خلاصه روندهای آینده مرتبط با شناسایی جوامع با استفاده از خوشه‌بندی سلسله‌مراتبی	تنها به رویکردهای سلسله‌مراتبی پرداخته شده و بقیه الگوریتم‌های شناسایی جوامع بررسی نشده‌اند
Porter و همکاران	۲۰۰۹	مشکلات حل نشده حوزه کشف جوامع پرداخته، مثال‌های حل شده مربوط به شبکه‌های اجتماعی بزرگ ارائه شده است	فقط رویکردهای تقسیم‌بندی گراف بررسی شده‌اند

1 Moscato & Sperlì

Yang	۲۰۱۰	خلاصه‌سازی مسئله استخراج جوامع شبکه با رویکرد بهینه‌سازی و ابتکاری، گسترش مسئله استخراج داده به جوامع پویا	مدل‌های آگاه به ناهنجاری و معیارهای ارزیابی الگوریتم‌ها بررسی نشده‌اند
Fortunato	۲۰۱۰	بررسی مسائل کلیدی مانند اهمیت خوشه‌بندی، روش‌های تست الگوریتم‌ها، و مقایسه روش‌های ایستا و همپوشان شناسایی جوامع	کاربردهای واقعی اندک، مدل‌های آگاه به ناهنجاری لحاظ نشده‌اند
Papadopoulos و همکاران	۲۰۱۲	بررسی مقیاس‌پذیری الگوریتم‌های شناسایی جوامع در مجموعه داده‌های واقعی، تمرکز بر پیچیدگی و نیازهای حافظه الگوریتم‌های موجود	مدل‌های آگاه به ناهنجاری و کاربردهای عملی این الگوریتم‌ها بررسی نشده‌اند
Xie و همکاران	۲۰۱۳	مرور وضعیت هنر در الگوریتم‌های شناسایی جوامع همپوشان از جمله برخی تکنیک‌های NMF	روش PCA و واریانت‌های آن برای شناسایی جوامع همپوشان بررسی نشده، بحثی درباره کاربردها ارائه نشده است
Malliaros و Vazirgiannis	۲۰۱۳	تقسیم‌بندی توپولوژیک برای الگوریتم‌های شناسایی جوامع ایستا، پویا و همپوشان	مدل‌های آگاه به ناهنجاری لحاظ نشده، تنها به چند کاربرد واقعی اشاره شده است
Chintalapudi و Prasad	۲۰۱۵	مرور الگوریتم‌های شناسایی جوامع مجزا در شبکه‌های هدایت‌شده و بدون جهت	الگوریتم‌های شناسایی جوامع همپوشان و کاربردهای واقعی پوشش داده نشده‌اند

روش پژوهش

شبکه‌های اجتماعی

شبکه‌های اجتماعی در عصر حاضر به بخش بسیار مهم و فراگیر زندگی روزمره ما تبدیل شده‌اند. این اصطلاح "شبکه اجتماعی" اغلب برای اشاره به خدمات آنلاین مختلف مرتبط با یک دسته کلی از موقعیت‌های تعامل اجتماعی و کاری استفاده می‌شود. شبکه اجتماعی چیست و چگونه نمایش داده می‌شود؟ ویژگی‌ها و خواص آن چیست؟

تعریف شبکه‌های اجتماعی

شبکه‌های اجتماعی به عنوان برنامه یا وب‌سایتی تعریف می‌شوند که امکان برقراری ارتباط و تعامل بین کاربران را فراهم می‌کنند. آن‌ها همچنین می‌توانند به عنوان مجموعه‌ای از موجودیت‌ها دیده شوند که با یکدیگر تعامل دارند. پیوندهای شبکه، نشان‌دهنده انواع مختلفی از روابط است که امکان ارتباطات چند طرفه مانند دوستی و پیوندهای وب را فراهم می‌کنند. شبکه‌های اجتماعی به کاربران کمک می‌کنند تا در متون، فیلم ها، تصاویر و ... با یکدیگر ارتباط برقرار کنند (آستون و هو[1] ۲۰۱۴)

1 Aston & Hu

ویژگی‌های شبکه‌های اجتماعی

مطالعه شبکه‌های اجتماعی نشان داد که آن‌ها ویژگی‌های مشترکی مانند تأثیر جهان‌های کوچک، درجه توزیع قانون توان و ساختار جامعه دارند (العتیبی و روما[1] ۲۰۲۱).

- **اثر جهان کوچک**

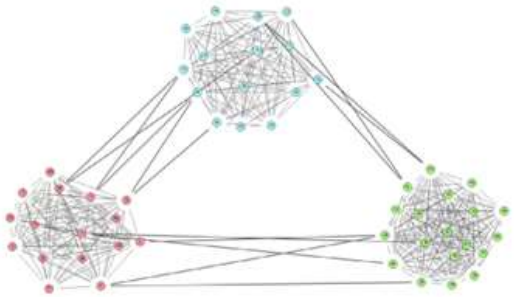
در سال ۱۹۶۷، میلگرام این فرضیه را مطرح کرد که هرکسی با یک زنجیره کوتاه از روابط اجتماعی به هر فردی مرتبط است. به عنوان بررسی این قانون «شش درجه جدایی»، نشان داده شد که طول متوسط یک مسیر بین دو کاربر ۶ است. این قانون این ایده را ثابت کرد که افراد می‌توانند به طور مستقیم یا غیرمستقیم با اتصال از طریق دانش مشترک یکدیگر را بشناسند. این پدیده مطمئنا صحیح به نظر می‌رسد؛ زیرا آمار پارامترهای موجود در گراف‌ها نشان داده است که قطر اکثر شبکه‌های اجتماعی نسبتا کوچک است. قطر به عنوان طولانی‌ترین فاصله در بین تمام کوتاه‌ترین مسیرهای ممکن تعریف می‌شود.

- **درجه توزیع قانون توان**

در شبکه‌های اجتماعی، گره‌های درجه پایین بسیار و گره‌های درجه بالا بسیار کم هستند. در واقع، توزیع گره‌ها از قانون توان $p(k)\ K\ \exp\gamma$ پیروی می‌کند که در آن k درجه و c قانون توان است. در عمل، این توان بین ۲ و ۳ است و نشان‌دهنده نرخ کاهش منحنی درجه است. هر چه c بزرگتر باشد، احتمال به دست آوردن رئوس درجه بالا کمتر است.

- **ساختار جامعه**

در یک شبکه، افراد عمدتا بر پایه علایق، پیشینه، سرگرمی‌های مشترک و ... گروه‌هایی را تشکیل می‌دهند. این گروه‌ها که تحت عنوان جوامع یا خوشه‌ها شناخته می‌شوند، به صورت مجموعه‌ای از گره‌های متراکم تعریف می‌شوند که تعداد پیوندهای داخلی آن‌ها به مراتب بیشتر از تعداد پیوندهای خارجی‌شان با سایر بخش‌های شبکه است. در واقع، میزان شناخت و تعامل افراد در درون جامعه خود، بسیار بیشتر از تعامل با افراد خارج از آن است. (شکل ۱). به همین ترتیب، (جربی و همکاران[2] ۲۰۲۰) در مقاله خود جامعه را به عنوان گروهی تعریف می‌کنند که در آن افراد، حول محور منافع مشترک گروه‌بندی می‌شوند و روابط درون گروه مهم‌تر از روابط خارج از گروه است. از سوی دیگر، (فورتوناتو[3] ۲۰۱۰) در مقاله خود پیشنهاد می‌کند که هیچ معنای مورد توافق گسترده‌ای از اصطلاح «جامعه» وجود ندارد. علاوه بر این، فرآیند کشف و تشخیص این جوامع به عنوان تشخیص جامعه شناخته می‌شود.



شکل ۱: ساختار جامعه در شبکه اجتماعی

1 Alotaibi & Rhouma
2 Djerbi et al
3 Fortunato

الگوریتم‌های تشخیص جامعه

الگوریتم‌های تشخیص جامعه به شدت به توپولوژی (ساختار) شبکه‌ها وابسته هستند؛ چرا که شبکه‌ها می‌توانند به دو صورت ایستا یا پویا باشند. شناسایی جوامع در شبکه‌های ایستا در مقایسه با شبکه‌های پویا با سهولت بیشتری انجام می‌شود. در شبکه‌های ایستا، رویکردهای متعددی برای تشخیص جامعه وجود دارد (گائو و همکاران[1] ۲۰۱۰) که عمدتا بر پایه الگوریتم‌های مبتنی بر بهینه‌سازی استوار هستند؛ این الگوریتم‌ها با هدف دستیابی به یک پاسخ بهینه بر اساس تابع هدف تعریف شده، عمل می‌کنند.

در میان الگوریتم‌های مبتنی بر بهینه‌سازی، رویکردی وجود دارد که با اتکا به یک استراتژی پایین به بالا از خوشه‌بندی مبتنی بر ضرایب همبستگی استفاده می‌کند. (فون لوکزبورگ[2] ۲۰۰۷). حداکثرسازی مدولاریته (نیومن، ۲۰۰۶) و خوشه‌بندی طیفی (آسور و همکاران[3] ۲۰۰۹) به ترتیب، به عنوان اسباب کار برای شناسایی جامعه در شبکه‌های ایستا در نظر گرفته می‌شوند؛ اما بیشتر شبکه‌های دنیای واقعی ماهیت پویا دارند و همچنین مطالعات مختلفی برای درک رفتار تکاملی این شبکه‌های پویا وجود دارد. برخی از پژوهش‌های اخیر نیز بر ردیابی ویژگی‌های وابسته به زمان در جوامع پویا تمرکز کرده‌اند. در گذشته، روش‌هایی مانند خوشه‌بندی طیفی و بیشینه‌سازی مدولاریتی عمدتا بر شناسایی جوامع مجزا در شبکه‌ها متمرکز بودند. با این حال، بسیاری از شبکه‌های دنیای واقعی، مانند شبکه‌های اجتماعی و شبکه‌های زیستی، دارای ساختاری با عضویت‌های چندگانه هستند؛ به این معنا که یک گره می‌تواند به طور هم‌زمان با چندین گروه مختلف ارتباط داشته باشد و در چند جامعه عضویت داشته باشد. با در نظر گرفتن محدودیت در مورد چندین عضویت در جامعه می‌توان راه‌حلی برای این مشکل در نظر گرفت. همچنین، الگوریتم‌هایی نیز پیشنهاد شده‌اند که قادرند هم‌زمان جوامع ناهم‌پیوسته و جوامع هم‌پوشان را در یک شبکه شناسایی کنند. (لیو و همکاران[4] ۲۰۱۴). بر اساس ویژگی‌های ساختاری شبکه‌ها، جوامع را می‌توان به عنوان جوامع مجزا و هم‌پوشانی طبقه‌بندی کرد. در جوامع مجزا، گره ها متعلق به یک جامعه واحد هستند.

جوامع در شبکه می توانند از هم‌گسسته باشند یا هم‌پوشانی داشته باشند:

جوامع از هم گسسته: گره‌ها در جوامع از هم‌گسسته فقط به یک جامعه تعلق دارند. با این حال، این پدیده را نمی‌توان در یک شبکه واقعی مشاهده کرد.

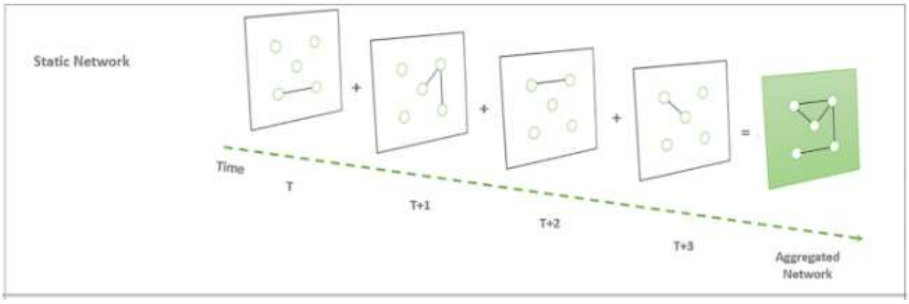
جوامع هم‌پوشانی: هنگام بحث در مورد شبکه‌های واقعی، توجه می‌کنیم که گره‌های خاصی ممکن است به گروه‌ها یا جوامع متعددی تعلق داشته باشند. در این حالت، چیزی خواهیم داشت که یک جامعه هم‌پوشانی نامیده می‌شود.

- **تشخیص جامعه در شبکه‌های استاتیک**

شناسایی جوامع در شبکه‌های اجتماعی به کشف ساختار شبکه کمک می‌کند. علاوه بر این، تجسم کمک ارزشمندی برای درک و تجزیه‌وتحلیل داده‌ها ارائه می‌دهد. در تحقیقات اولیه، اکثر روش‌های پیشنهادی شبکه‌های اجتماعی، شبکه را به عنوان یک شبکه تجمیع شده در نظر می‌گرفتند. در اینجا، تمام فعل و انفعالات مشاهده شده بین موجودیت‌ها در یک دوره زمانی با هم ترکیب می‌شوند تا شبکه ایستا را تشکیل دهند (شکل ۲). چندین روش برای شناسایی جوامع در یک شبکه اجتماعی ثابت ارائه شده است. علاوه بر این، چندین تجدید نظر در این روش‌ها پیشنهاد شده است. (فورتوناتو، ۲۰۱۰) بررسی‌هایی در مورد تشخیص جامعه در گراف‌ها ارائه کرد، به بیش از پنجاه روش مختلف اشاره و رویکرد جدیدی

1 Gao et al
2 Von Luxburg
3 Asur et al
4 Liu et al

را برای شناسایی جوامع بر اساس تعاملات کاربر پیشنهاد کرد. آن‌ها چهار مرحله را برای کشف جوامع دنبال می‌کنند. ابتدا آن‌ها شبکه را تحلیل و ارزیابی کردند تا اطلاعات مربوط به گره‌ها و لبه‌ها را جمع‌آوری کنند. سپس مقادیر عددی موجود در مجموعه داده را بررسی و به مقادیر طبقه‌بندی تبدیل و تعاملات شبکه را به یک مجموعه داده تراکنشی تبدیل کردند. قوانین انجمن (ARL) با استفاده از مجموعه داده تراکنشی مرحله قبل استخراج شد. پس از تعریف این قوانین، آن‌ها را فیلتر و هر گونه افزونگی را حذف و سپس گره‌ها یا کاربران شبکه اجتماعی را بر اساس قوانین ایجاد شده قبلی استخراج کردند، یعنی تمام جوامع ممکن را با تمام ترکیبات ممکن استخراج شد. الگوریتم پیشنهادی تمام جوامع ممکن در شبکه را در مقایسه با بسیاری از الگوریتم‌های دیگر شناسایی شد. تا همین اواخر، پویایی شبکه‌های اجتماعی در فرایند شناسایی جوامع چندان مورد توجه قرار نمی‌گرفت؛ زیرا مطالعه و تحلیل پویایی این شبکه‌ها با دشواری‌های قابل‌توجهی همراه بود. چندین روش تشخیص جامعه عمدتا برای شبکه‌های استاتیکی که در آن هیچ تغییری اتفاق نمی‌افتد، توسعه یافته‌اند. اخیرا تمایل به شناسایی جوامع پویا افزایش یافته و در همین راستا، روش‌های متنوعی در این زمینه توسعه یافته است.



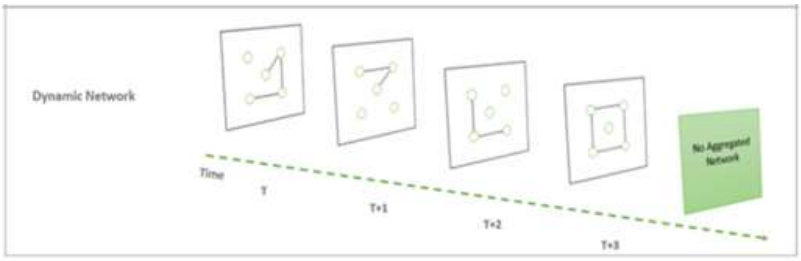
شکل ۲: شبکه استاتیک (منبع: العتیبی و روما، ۲۰۲۱)

- **تشخیص جامعه در شبکه‌های پویا**

تا همین اواخر، به دلیل دشواری مطالعه پویایی شبکه‌های اجتماعی، پویایی آن‌ها در طول تشخیص جوامع در نظر گرفته نمی‌شد. چندین روش تشخیص جامعه عمدتا برای شبکه‌های ایستا که در آن‌ها هیچ تغییری رخ نمی‌دهد، توسعه داده شده‌اند. اخیرا، علاقه به تشخیص جوامع پویا افزایش یافته و انواع مختلفی از روش‌ها توسعه یافته‌اند.

شبکه‌های پویا

به دلیل تمایل افراد به تغییر، پیوسته به شبکه‌ها می‌پیوندند و یا آن‌ها را ترک می‌کنند. به عنوان مثال، کاربران جدید هر روز در فیس‌بوک ثبت‌نام می‌کنند و بسیاری از کاربران دیگر به طور همزمان حساب‌های خود را حذف یا غیرفعال می‌کنند. در لینکدین، کاربران می‌توانند با سایر افراد در شبکه ارتباط برقرار یا قطع کنند. این واقعیت چیزی را ایجاد می‌کند که به عنوان شبکه اجتماعی پویا شناخته می‌شود. بنابراین، شبکه تکامل پیدا می کند. همان‌طور که قبلا ذکر شد، اکثر روش‌های شبکه‌های اجتماعی پیشنهادی، شبکه را به عنوان یک شبکه مجموع در نظر می‌گرفتند. با این حال، در این نوع روش، ماهیت زمانی این‌گونه فعل و انفعالات نادیده گرفته خواهد شد؛ زیرا این تعاملات همچنان در حال تغییر خواهند بود. در این نوع شبکه‌های پویا، (شکل ۳) تعاملات به صورت یک سری شبکه‌های ایستا در نظر گرفته می‌شود.



شکل ۳: شبکه پویا (منبع: العتیبی و روما، ۲۰۲۱)

طبقه‌بندی الگوریتم‌های تشخیص جامعه

موضوع تشخیص جامعه به طور گسترده مورد مطالعه قرار گرفته و الگوریتم‌های متنوعی در این زمینه ارائه شده است. پژوهشگران، الگوریتم‌های تشخیص جامعه را بسته به ابعاد و رویکرد پژوهش خود، به روش‌های گوناگونی طبقه‌بندی کرده‌اند (آلدکوا و مارین[1] ۲۰۱۳). (فورتوناتو، ۲۰۱۰) به طور گسترده الگوریتم‌های تشخیص جامعه را با انطباق با تمام رویکردهای مورد بحث در بررسی‌های دیگر مورد مطالعه قرار داده است؛ اما جدیدترین تکنیک‌های پیشنهادی برای هم‌پوشانی جوامع هنوز باید پوشش داده شوند. بنابراین، در این بررسی، الگوریتم‌های تشخیص جامعه براساس جدول ۲ ارائه شده است.

جدول ۲: مزایا و معایب روش‌های مختلف تشخیص جامعه

دسته	الگوریتم	مزایا	معایب	منابع
الگوریتم‌های سنتی	خوشه‌بندی سلسله مراتبی	نیازی به تعیین تعداد خوشه‌ها از پیش نیست	اگر فرآیند ادغام به‌خوبی انجام نشود، نتایج ممکن است ضعیف باشند	(فورتوناتو، ۲۰۱۰)؛ (شن و همکاران، ۲۰۰۹)؛ (آهان و همکاران، ۲۰۱۰)
	الگوریتم Girvan-Newman (GN)	نیازی به تعیین تعداد خوشه‌ها از پیش نیست	مقیاس‌پذیر نیست و بیش از چند صد گره را پشتیبانی نمی‌کند	گیروان و نیومن، ۲۰۰۲)؛ (نیومن، ۲۰۱۰)
	الگوریتم Tyler and Wilkinson	سریع‌تر از الگوریتم‌های GNکلاسیک است	وجود ابهام در مرحله حذف یال‌ها برای زیرگراف‌های متراکم	تیلر، ۲۰۰۵)؛ (ویلکینسون و هوبرمن، ۲۰۰۴)
	الگوریتم Rattigen؛ (fast variant of GN)	به دلیل استفاده از یک معیار نوین به جای (Edge Betweenness)، سریع است	فقط برای اجتماعات مجزا (غیرهمپوشان) کارایی دارد	(راتیگان، ۲۰۰۷)

1 Aldecoa & Marín

(فون لوکزبورگ، ۲۰۰۷)؛ (دینگ، ۲۰۰۴)	عملکرد ضعیف روی نقاط پرت (Outliers)، سختی در یافتن بهترین تابع هدف	نتایج بسیار خوبی در اشکال پیچیده ارائه می‌دهد	الگوریتم خوشه‌بندی طیفی Spectral Clustering	
(پوتن، ۱۹۹۷)؛ (کرنیگان، ۱۹۷۰)	اندازه خوشه باید از پیش تعیین شود، نتایج ممکن است به دلیل مقداردهی اولیه تصادفی پارتیشن‌ها تصادفی باشند	نسبت به دیگر الگوریتم‌های سنتی مقاوم‌تر است	تقسیم‌بندی گراف مانند الگوریتم Kernighan–Lin)	
(گیمرا، ۲۰۰۵)	وابسته به سرعت همگرایی شبیه‌سازی تبرید	عملکرد خوب به دلیل یافتن راه‌حل بهینه با استفاده از شبیه‌سازی تبرید	شبیه‌سازی تبرید (Guimera and Amaral Algorithm)	الگوریتم‌های مبتنی بر مدولاریتی
(نیومن، ۲۰۰۴)	برای بیش از چند صد گره مقیاس‌پذیر نیست و با روش‌های حریصانه قابل مقایسه نیست	سریع‌تر از الگوریتم GN کلاسیک	الگوریتم GN سریع (Fast GN Algorithm)	
(پوتن، ۱۹۹۷)	نیازمند تعیین اندازه خوشه‌ها از پیش	با استفاده از ساختارهای داده ماتریس پراکنده، الگوریتم GN را به‌طور مؤثر تسریع می‌کند	الگوریتم Clauset	
(گیمرا، ۲۰۰۵)؛ (نیومن، ۲۰۰۴)؛ (بوتچر و پرکوس، ۲۰۰۱)؛ (لیو، ۲۰۱۰)	همان معایب شبیه‌سازی تبرید را دارد ولی با سرعت بیشتر؛ عملکرد خوب در معیارGN	افزایش قابل‌توجه سرعت محاسبه نسبت به شبیه‌سازی تبرید، عملکرد خوب در معیار GN	بهینه‌سازی اکستریم (Extremal Optimization)	
گیروان و نیومن، ۲۰۰۴)؛ (نیومن، ۲۰۰۶)	فقط برای bi-sectioning طیفی (تقسیم به دو بخش) مناسب است و در بیش از دو خوشه شکست می‌خورد	برخلاف تقسیم‌بندی گراف، معیار توقف مشخصی بر اساس تغییرات امتیاز مدولاریتی دارد	بهینه‌سازی طیفی (Spectral Optimization)	
(تاسگین و همکاران، ۲۰۰۷)؛ (زنگ و لیو، ۲۰۱۵)	ممکن است در شبکه‌های تصادفی با مقادیر مدولاریتی بسیار بالا شکست بخورد	نیاز به تعیین از پیش تعداد جوامع ندارد، می‌تواند مسئله بهینه‌سازی چندهدفه کشف اجتماع را حل کند	الگوریتم ژنتیک (Genetic Algorithms)	

الگوریتم‌های پویا (Dynamic Algorithms)	Random Walk (حرکت تصادفی)	می‌تواند همراه با روش‌های تجمیعی (Agglomerative) برای شناسایی اجتماعات استفاده شود.	پرهزینه (برای مثال، کندتر از تکنیک‌های حریصانه است	(هیوز، ۱۹۹۶)؛ (ژو، لیپوسکی، ۲۰۰۴)
	Synchronization (هم‌زمان‌سازی)	دقت بالا و پیچیدگی زمانی پایین	غیرقابل اعتماد در اجتماعات با اندازه متغیر	(بوکالتی و همکاران، ۲۰۰۷)؛ (آریناس و همکاران، ۲۰۰۶)
	Spin Models (مدل‌های اسپین)	دارای پارامترهای قابل تنظیم برای اندازه اجتماعات	بسیار کندتر از الگوریتم حریصانه Clauset، فقط برای شبکه‌های کوچک مقیاس کار می‌کند	(پاپادوپولوس، ۲۰۱۲)؛ (رایکارد بورنهولت، ۲۰۰۴)
Algorithms for Overlapping Communities الگوریتم‌ها برای اجتماعات همپوشان	Fuzzy Detection (Label Propagation) تشخیص فازی - انتشار برچسب	پیچیدگی زمانی پایین و قابلیت تشخیص اجتماع هم‌پوشان	فقط می‌تواند یک اجتماع را شناسایی کند	(گرگوری، ۲۰۰۷)؛ (نپوش و همکاران، ۲۰۰۸)
	Clique Percolation Method (CPM) (روش نفوذ کلیک)	می‌تواند اجتماعات هم‌پوشان شناسایی کند، امکان تشخیص گراف‌های تصادفی از گراف‌های دارای ساختار اجتماع	فقط مناسب زمانی است که زیرگراف‌ها در گراف داده به‌طور کامل به هم متصل باشند	(پالا و همکاران، ۲۰۰۵)؛ (کومپولا و همکاران، ۲۰۰۸)
	Non Negative Matrix Factorization (NMF) با اندازه ماتریس ورودی متغیر	الگوریتم مدرن برای تشخیص اجتماع هم‌پوشان	پرهزینه از لحاظ محاسباتی و مقیاس‌پذیری ضعیف	(روسی و همکاران، ۲۰۱۳)؛ (زارعی، ۲۰۰۹)
	Symmetric Non Negative Matrix Factorization (فاکتورگیری ماتریس غیرمنفی متقارن)	حذف محدودیت‌های اضافی در برآورد	فقط برای گراف‌های بدون جهت	(وانگ و همکاران، ۲۰۱۱)؛ (ژاو و همکاران، ۲۰۱۰)
	Bayesian Nonnegative Matrix Factorization (BNMF) و Initialized-BNMF (IBNMF) (فاکتورگیری ماتریس غیرمنفی بیزی و نسخه اولیه آن)	برطرف کردن ضعف روش بیشینه‌سازی مدولاریتی، محاسبات بسیار کاراتر از NMF	فقط برای شبکه‌های ایستا کاربرد دارد	(پسوراکیس و همکاران، ۲۰۱۱)

(لیو و همکاران، ۲۰۱۶)	تضمین‌کننده کم‌تراکم بودن عوامل نیست	خوشه‌بندی بهتر نسبت به NMF	NMFGR	
(ژانگ و ینگ، ۲۰۱۲)	تعداد اجتماعات باید از قبل تعیین شود	عملکرد بهتر نسبت به روش‌های تجربی NMF	Bounded NMTF (BNMTF)	
(لیو و همکاران، ۲۰۱۶)	داده‌ها می‌توانند به‌طور دلخواه بازمقیاس شوند اگر مؤلفه‌های اصلی با ماتریس همبستگی انتخاب شوند	حل مشکل پیچیدگی زمانی NMF، زمان محاسبات کمتر از الگوریتم‌های کلاسیک	Principal Component Analysis (تحلیل مؤلفه‌های اصلی)	
(هلند و لاین‌هارت، ۱۹۷۶)؛ (یانگ، ۲۰۱۱)	SBM کلاسیک قادر به تشخیص اجتماعات هم‌پوشان نیست	استفاده برای مدل و تحلیل شبکه‌ها مانند شبکه اجتماعی، DBOCD با برطرف کردن ضعف‌های BNMF می‌تواند اجتماعات هم‌پوشان را نیز شناسایی کند	Statistical Inference based methods (SBM) و Bayesian Inference (روش‌های استنباط آماری و استنباط بیزی)	
(شن و همکاران، ۲۰۰۹)؛ (زارعی، ۲۰۰۹)	کیفیت تشخیص اجتماع به خوبی روش‌های بخش‌بندی گره‌ها نیست	روشی نوین که پیوندهای بین گره‌ها را بخش‌بندی می‌کند و می‌تواند اجتماعات هم‌پوشان را هم تشخیص دهد	Line Graph and Link Partitioning گراف خطی و (بخش‌بندی پیوند)	

الگوریتم‌هایی برای جوامع گسسته

الگوریتم‌های زیادی برای تشخیص ساختار جامعه در شبکه‌ها پیشنهاد شده است. بیشتر این الگوریتم‌ها برای کشف جوامع مجزا طراحی شده‌اند که در ادامه به آن‌ها می‌پردازیم.

- **الگوریتم‌های سنتی**

پژوهش در زمینه تشخیص جامعه با استفاده از تحلیل گراف‌ها، در اوایل دهه‌ی ۱۹۷۰ میلادی آغاز شد (فورتوناتو، ۲۰۱۰). از آن زمان تاکنون، الگوریتم‌های متعددی مبتنی بر مفاهیم خوشه‌بندی توسعه یافته‌اند که امروزه تحت عنوان الگوریتم‌های سنتی شناخته می‌شوند. در ادامه، برخی از برجسته‌ترین الگوریتم‌های سنتی که مفاهیم کلیدی را در این حوزه پایه‌گذاری کرده و مسیر را برای توسعه روش‌های پیشرفته بعدی هموار ساختند، معرفی می‌شوند.

- **خوشه‌بندی پارتیشن**

در این الگوریتم، گره‌های شبکه با بیشینه‌سازی یا کمینه‌سازی یک تابع زیان مبتنی بر فاصله K (فرضی) میان آن‌ها، در خوشه از پیش تعیین شده دسته‌بندی می‌شوند. برخی از تکنیک‌های خوشه‌بندی پارتیشن پرکاربرد در زیر فهرست شده‌اند:

Minimum K-clustering: در اینجا تابع هزینه قطر خوشه است که حداکثر فاصله بین دو نقطه از یک خوشه است.

K-clustering sum: شبیه به حداقل K-clustering است با این تفاوت که میانگین فاصله بین جفت نقاط خوشه جایگزین قطر خوشه در تابع هزینه می‌شود.

K-center: در این روش، برای هر خوشه یک مرکز در نظر گرفته می‌شود و بیشترین فاصله هر گره از مرکز خوشه (قطر خوشه) به عنوان تابع هزینه محاسبه می‌شود. سپس مراکز و تخصیص گره‌ها به گونه‌ای تعیین می‌شوند که مقدار بیشینه این فاصله در میان همه خوشه‌ها حداقل شود.

K-median: این روش مشابه الگوریتم K-center است، با این تفاوت که در آن تابع هزینه به جای کمینه‌سازی بیشینه فاصله، بر کمینه‌سازی میانگین فاصله گره‌ها از مراکز خوشه‌ها مبتنی است.

محبوب‌ترین رویکرد برای خوشه‌بندی پارتیشنی، خوشه‌بندی k-means است که تابع زیان مجذور فاصله درون خوشه‌ای را به حداقل می‌رساند. این مسئله k-means را می‌توان به راحتی با الگوریتم لوید حل کرد (لوید[1] ۱۹۸۲). در راستای خوشه‌بندی k-means، یکی دیگر از تکنیک‌های محبوب، خوشه‌بندی k-means فازی است (بزدک[2] ۲۰۱۳). این تکنیک ویژگی یک گره متعلق به دو یا چند خوشه را به طور هم‌زمان در خود جای می‌دهد و کاربردهایی در تشخیص الگو دارد. تعیین تعداد خوشه‌ها در ابتدا محدودیت این الگوریتم است.

- **خوشه‌بندی سلسله مراتبی**

اطلاع از تعداد خوشه‌های موجود در آن بسیار نادر است. یک گراف تقسیم شده است. در چنین مواردی، استراتژی‌های خوشه‌بندی پارتیشن‌بندی ممکن است به دلیل پیش‌فرض‌های غیرمنعطف در مورد اندازه و تعداد خوشه‌ها، در تحلیل گراف‌های خاص عملکرد مطلوبی نداشته باشند. در مقابل، گراف‌های شبکه‌ای غالبا دارای یک ساختار سلسله‌مراتبی هستند که در آن، گره‌ها در قالب سطوح متعددی از خوشه‌ها سازمان‌دهی می‌شوند. در چنین شرایطی، بکارگیری روش‌های خوشه‌بندی سلسله مراتبی کارآمدی بسیار بالایی از خود نشان می‌دهد. تکنیک‌های خوشه‌بندی سلسله‌مراتبی، گزینه‌ای بسیار کارآمد برای تحلیل گراف‌هایی با ساختار چندسطحی و لایه‌ای به شمار می‌روند. (ژانگ و همکاران[3] ۲۰۱۴). در این الگوریتم‌ها، خوشه‌ها به صورت سلسله مراتبی و در قالب یک درخت باینری ادغام می‌شوند که مبنای آن، میزان شباهت میان رئوس است. برخلاف خوشه‌بندی پارتیشن‌بندی که مستلزم تعیین تعداد خوشه‌ها در ابتدای فرآیند است، این رویکرد انعطاف‌پذیری بیشتری دارد. در ادامه، به بررسی مهم‌ترین روش‌های خوشه‌بندی سلسله مراتبی پرداخته خواهد شد.

(i) الگوریتم‌های تجمعی: الگوریتم‌های تجمعی با اتخاذ رویکردی پایین به بالا، به صورت تکرار شونده خوشه‌هایی را که دارای بیشترین میزان شباهت هستند، شناسایی کرده و آن‌ها را در قالب خوشه‌های بزرگ‌تر ادغام می‌کنند (مقبول و بابری[4] ۲۰۰۴). این الگوریتم‌ها فرآیند خوشه‌بندی را با در نظر گرفتن هر گره به عنوان یک خوشه مستقل آغاز کرده و در هر گام، نزدیک‌ترین جفت خوشه‌ها را بر اساس امتیاز شباهت با یکدیگر ادغام می‌کنند. روش‌های خوشه‌بندی با پیوند کامل و خوشه‌بندی مبتنی بر پیوند سلسله‌مراتبی از نمونه‌های شاخص الگوریتم‌های تجمعی به شمار می‌روند. مزیت اصلی این رویکرد، توانایی آن در شناسایی خوشه‌های کوچک و دقیق است که می‌تواند در کشف جوامع محلی بسیار اثربخش باشد؛ با این حال، محدودیت بنیادین این الگوریتم‌ها در برگشت‌ناپذیری مراحل ادغام است؛ به‌گونه‌ای که اگر گره‌ای در مراحل اولیه به‌اشتباه در خوشه‌ای تخصیص پیدا کند، امکان جابه‌جایی یا اصلاح آن در مراحل بعدی وجود نخواهد داشت. مشکل دیگر استراتژی خوشه‌بندی، این است که گره‌های دارای یک همسایه اغلب به عنوان یک خوشه مستقل طبقه‌بندی می‌شوند که در موارد خاص معنا ندارد. خوشه‌بندی سلسله مراتبی تجمعی نمی‌تواند به خوبی مقیاس شود، زمانی که نقاط در فضا تعبیه شده باشند و فاصله به عنوان معیار عدم تشابه در فرآیند خوشه‌بندی استفاده شود.

---

1 Lloyd
2 Bezdek
3 Zhang et al
4 Maqbool & Babri

اگر فاصله به طور ساده تعریف نشده باشد، پیچیدگی سنگین‌تر می شود که O(n2) برای پیوند تک و O(n2 log n) برای اتصال متوسط کامل است.

(ii) الگوریتم‌های تقسیم‌بندی: الگوریتم‌های تقسیم‌بندی، رویکردی بالا به پایین در خوشه‌بندی سلسله مراتبی محسوب می‌شوند (روکس[1] ۲۰۱۵). در این روش، فرآیند با در نظر گرفتن تمامی نمونه‌ها در قالب یک خوشه واحد آغاز شده و سپس با استفاده از الگوریتم‌های خوشه‌بندی مسطح، به زیرخوشه‌های کوچک‌تر تقسیم می‌شود. پس از تشکیل دندروگرام، ساختار جوامع از طریق برش درخت در سطوح مختلف شناسایی می‌شود. در این فرآیند، تعیین دقیق محل برش از اهمیت حیاتی برخوردار است؛ به گونه‌ای که انتخاب نامناسب این نقطه می‌تواند منجر به افت کیفیت و دقت نتایج نهایی شود. در همین راستا (گیروان و نیومن[2] ۲۰۰۲) الگوریتم‌های متعددی را در کلاس روش‌های تقسیم‌بندی برای تشخیص جوامع در شبکه‌ها پیشنهاد کرده‌اند.

یکی از روش‌های آن‌ها مبتنی بر مرکزیت بین لبه است که در آن یال‌هایی که بیشترین مرکزیت را دارند حذف می‌شوند. سپس متریک دوباره محاسبه می‌شود و مرحله حذف لبه تا زمانی که هیچ لبه‌ای باقی نماند تکرار می‌شود. یکی از مزایای این روش این است که نیازی به دانش قبلی در مورد تعداد خوشه‌ها نیست؛ اما این الگوریتم از نظر زمان محاسباتی یعنی O(m2n) پرهزینه بود که برای شبکه‌های مقیاس بزرگ مناسب نیست. (تایلر و همکاران[3] ۲۰۰۵) در الگوریتم GN نوآوری به ارمغان آوردند و به جای محاسبه دقیق بین لبه همه پیوندها، آن را با استفاده از روش تخمین مونت کارلو تقریب زدند. نتیجه بهبود سرعت و هزینه محاسباتی با کاهش دقت بود. یک نوع سریع از الگوریتم‌های GN نیز توسط Rattigan و همکاران پیشنهاد شد (راتیگان و همکاران[4] ۲۰۰۷) که در آن با انجام جستجوهای بین جفت‌های نمونه‌گیری شده از گره‌ها تقریبی شد. او نشان داد که پیچیدگی الگوریتم را می‌توان با اطمینان از دقت در تخمین بین لبه‌ها به O(m) کاهش داد.

- **خوشه‌بندی طیفی**

این شامل تمام الگوریتم‌هایی است که یک گراف را با استفاده از بردارهای ویژه ماتریس داده ورودی به خوشه‌ها تقسیم می‌کنند. مجموعه‌ای از اشیاء را به مجموعه‌ای از نقاط در فضای چند بعدی تبدیل می‌کند که مختصات آن عناصر، بردار ویژه هستند. این تبدیل ویژگی‌های ضمنی مجموعه داده‌های اولیه را نشان می‌دهد و خوشه‌بندی طیفی را می‌توان برای خوشه‌بندی داده‌هایی که نمی‌توان با استفاده مستقیم از k-means با موفقیت انجام داد، استفاده کرد. در حوزه خوشه‌بندی طیفی، اولین مشارکت توسط (دانات و هافمن[5] ۱۹۷۳) پیشنهاد شد که در آن بردارهای ویژه ماتریس مجاورت و مقادیر ویژه ماتریس شباهت برای تقسیم‌بندی گراف استفاده کردند. در همان سال، (فیلدر[6] ۱۹۷۳) تقسیم دوگانه یک گراف را با استفاده از دومین کوچک‌ترین مقدار ویژه ماتریس لاپلاسی (L) که تفاوت ماتریس درجه (D) و ماتریس مجاورت (A) یک گراف است، به دست آورد. خوشه‌بندی طیفی به تکنیک‌های تشخیص جامعه گسترش یافت. (دینگ[7] ۲۰۰۴) در اثر دیگری یک بررسی جامع در مورد تکنیک‌های خوشه‌بندی طیفی نوشته است که می‌توان برای جزئیات بیشتر به آن مراجعه کرد. راندمان خوشه‌بندی طیفی در محیط داده‌های مقیاس بزرگ کاهش پیدا می‌کند؛ زیرا به زمان و فضای زیادی برای ذخیره ماتریس شباهت کامل نیاز دارد و همچنین نیاز به تجزیه ویژه دارد. پیچیدگی زمانی برای تجزیه ویژه و پیچیدگی فضایی برای ذخیره‌سازی ماتریس شباهت به ترتیب حدود $O(n^3)$ و $O(n^2)$ است که در پردازش داده در

1 Roux
2 Girvan & Newman
3 Tyler et al
4 Rattigan et al
5 Donath & Hoffman
6 Fiedler
7 Ding

مقیاس بزرگ غیرقابل قبول است. روش گسترش نیستروم برای حالت خوشه‌بندی داده‌ها در مقیاس بزرگ بسیار مؤثر است؛ زیرا با استفاده از روش‌های تقریب، مشکل توصیف شده هزینه محاسباتی را کاهش می‌دهد. در ابتدا، طراحی تکنیک Nystrom برای حل معادلات انتگرال بود؛ اما بعدا این روش در مسائل یادگیری ماشین هسته برای سرعت بخشیدن به فرآیند تجزیه ویژه استفاده شد (ولیامز و سیگر،[1] ۲۰۰۱). عملکرد خوشه‌بندی با استفاده از روش نیستروم مستقیما به انتخاب نقاط نمونه‌گیری وابسته است، یعنی تعداد نمونه‌های بیشتر منجر به تقریب بهتر می‌شود؛ اما افزایش نرخ نمونه‌گیری نیز در صورت مشکلات کلان داده مطلوب نیست. بنابراین، پسوندهای متعددی در تکنیک‌های Nystrom برای غلبه بر مسئله توصیف شده پیشنهاد شد. (جیا و همکاران،[2] ۲۰۱۷) از تکنیک گسترش نیستروم برای تقریب بردارهای ویژه با نمونه‌برداری از تعدادی نقطه داده با استفاده از روش نمونه‌گیری افزایشی پویا استفاده کرد. روش نمونه‌گیری افزایشی داده‌ها را با توجه به توزیع‌های احتمالی مختلف نمونه‌برداری می‌کند. در اینجا خطای نمونه‌گیری با افزایش زمان نمونه‌برداری کاهش پیدا می‌کند. معمولا دقت خوشه‌بندی برای به دست آوردن کارایی الگوریتم کاهش پیدا می‌کند؛ اما این الگوریتم عملکرد خوشه‌بندی را بهبود بخشیده و تعادل خوبی بین کارایی الگوریتم‌ها و کیفیت نمونه‌گیری تضمین می‌کند.

- **پارتیشن‌بندی گراف**

الگوریتم‌های این کلاس، رئوس را به گروه‌های g تقسیم می‌کنند که اندازه آن‌ها از پیش تعریف شده است تا حداقل تعداد پیوندها بین گروه‌های شناسایی‌شده به دست آید (پوتن،[3] ۱۹۹۷). تعداد رئوس در حال اجرا در میان خوشه‌ها اندازه برش نامیده می‌شود. اگر تعداد خوشه‌ها را از قبل مشخص نکنید و پارتیشنی با حداقل اندازه برش وارد نکنید، خروجی یک راه‌حل بی‌اهمیت خواهد بود. الگوریتم Kernighan-Lin یک رویکرد اکتشافی برای پارتیشن بندی گرافها است و همچنان در ترکیب با تکنیک های دیگر اعمال می شود (کرنیگان و لین،[4] ۱۹۷۰). هدف آن به حداقل رساندن یک عملکرد ارزیابی است که تفاوت پیوندهای درون جامعه و بین جامعه است. انگیزه این رویکرد مشکل بخش‌بندی مدارهای الکترونیکی روی بردها بود که در آن رئوس در بردهای مختلف قرار بود با حداقل تعداد اتصالات به هم مرتبط شوند. در واقع، این یک بهینه‌سازی تابع سود Q است. این تابع را می‌توان به عنوان تفاوت بین تعداد یال‌های داخل ماژول و تعداد یال‌های بین آن‌ها تعریف کرد. علاوه بر رویکرد دوبخشی در تقسیم‌بندی گراف، رویکرد سریع و محبوب دیگری به نام روش تقسیم‌بندی طیفی وجود دارد که بر اساس ویژگی‌های طیف ماتریس لاپلاسی است. روش‌های رایج دیگری برای پارتیشن‌بندی نمودار مانند تقسیم‌بندی ساختار سطح، الگوریتم‌های چندسطحی و الگوریتم‌های هندسی وجود دارد که جزئیات آن‌ها را می‌توان در پژوهش (پوتن، ۱۹۹۷) مشاهده کرد.

- **الگوریتم‌های مبتنی بر مدولاریت**

روش‌های مبتنی بر Modularity تلاش می‌کنند تا modularity را به حداکثر برسانند. این الگوریتم‌ها سعی می‌کنند بهترین مقدار Q را در مسائل تشخیص جامعه NP-hard پیدا کنند که عمدتا با استفاده از برخی ابتکارات انجام می‌شود.

- **بهینه‌سازی افراطی (EO)**

این یک روش جستجوی اکتشافی است که برای تقریب راه‌حل‌های مسائل بهینه‌سازی سخت پیشنهاد شده است (بوتچر و پرکوس،[5] ۲۰۰۱) این روش مبتنی بر پویایی فرآیندهای غیرتعادلی سیستم‌هایی است که بحرانی خود سازمان یافته (SOC) را نشان می‌دهند که در آن راه‌حل‌های مناسب‌تر به طور فعال و بدون تنظیم پارامترها ارائه می‌شوند. این تابع

1 Williams & Seeger
2 Jia et al
3 Pothen
4 Kernighan & Lin
5 Boettcher & Percus

مدولاریت Q را با استفاده از ارزش تناسب الگوریتم های ژنتیک بهینه می‌کند. معیار تناسب یک راس، نسبت بین مدولار بودن راس و درجه آن است. EO از نظر عملکرد معادل Simulated Annealing است؛ اما از نظر محاسباتی کارآمدتر است. مراحل الگوریتمی برای EO به شرح زیر است؛ شبکه داده شده را به دو خوشه با تعداد گره مساوی تقسیم کنید. در هر تکرار، گره‌ای را که کمترین مقدار تناسب را دارد به خوشه‌های دیگر منتقل کنید. با تغییر پارتیشن، مقدار تناسب رئوس را دوباره محاسبه کنید. این روند تا زمانی ادامه پیدا می‌کند که عملکرد Q بهبود پیدا نکند. روش دیگری به نام الگوریتم نیمه نظارت (الگوریتم PCSEO-SS) توسط (لی و همکاران[1] ۲۰۱۴) ارائه شده است که می‌تواند علاوه بر تشخیص دقیق جوامع، مشکل اتصالات نادرست را حل کند. در صورت محدود بودن اطلاعات قبلی، می‌تواند کارآمد باشد.

- **بهینه‌سازی طیفی**

این روش، یک تکنیک بهینه‌سازی مبتنی بر مدولاریت است که از اطلاعات طیفی ماتریس داده‌ها بهره می‌گیرد. تفاوت بارز این رویکرد با الگوریتم‌های خوشه‌بندی طیفی سنتی در این است که روش‌های سنتی فاقد یک تابع هدف مشخص برای بهینه‌سازی (مانند مدولاریت) هستند؛ در حالی که این تکنیک با هدف بیشینه‌سازی مدولاریت، ساختار جوامع را استخراج می‌کند. مقادیر ویژه و بردارهای ویژه ماتریس مدولاریته را می‌توان برای بهینه‌سازی مدولاریت استفاده کرد (چن و همکاران[2] ۲۰۱۴). به عنوان مثال، اگر بردارهای ویژه با دو مقدار ویژه انتخاب شوند، می‌توان یک تقسیم از گراف را در شبکه‌ای از سه خوشه به دست آورد. ماژولاریت روی دو قسمت را می‌توان از طریق دوبخشی طیفی با جایگزینی ماتریس مدولاریته به جای ماتریس لاپلاسی بهینه کرد. نتیجه ممکن است، بیشتر بهبود پیدا کند اگر رئوس از یک جامعه به جامعه دیگر منتقل شوند تا بیشترین افزایش یا کمترین کاهش در مقدار مدولاریته را به دست آورند. این تکنیک همچنین کاربردهایی در الگوریتم‌های حریصانه و بهینه‌سازی افراطی پیدا می‌کند.

- **بهینه‌سازی حریصانه**

(نیومن، ۲۰۰۴) یک روش حریصانه برای به حداکثر رساندن مدولار بودن پیشنهاد کرد. این یک الگوریتم خوشه‌بندی سلسله مراتبی تجمعی است که در آن لبه‌ها به هم متصل می‌شوند تا جوامع بزرگ‌تری را شکل دهند به طوری که مدولاریت افزایش پیدا کند. بهینه‌سازی حریصانه ماژولاریت برای تشکیل جوامع بزرگ با مقدار ضعیف ماکزیمم‌های مدولار انجام می‌شود. (کلاوست و همکاران[3] ۲۰۰۴) نشان دادند که ماتریس eij مورد استفاده نیومن (ماتریس eij کسر یال‌های بین خوشه‌های i و j از پارتیشن در حال اجرا را نشان می‌دهد) برای محاسبه مدولاریت Q، تعداد زیادی عملیات بی‌فایده به دلیل ماتریس مجاورت پراکنده دارد. آن‌ها استفاده از max-heaps را پیشنهاد کردند. یک ساختار داده طراحی شده برای ماتریس‌های پراکنده که با مرتب کردن مجدد داده‌ها در قالب یک درخت دودویی کار می‌کند برای به دست آوردن کارایی در عملکرد الگوریتمی. به‌روزرسانی در این مورد، بسیار سریع‌تر از تکنیک حریصانه نیومن بود. هزینه محاسباتی الگوریتم $O(n \log^2 N)$ بود. این الگوریتم را می‌توان برای یافتن ماکزیمم مدولاریت گراف شبکه بزرگ استفاده کرد. به طور مشابه، یک رویکرد بهتر برای مدولار بودن بهینه در مقایسه با بهینه‌سازی حریصانه نیومن ارائه شده است. در پژوهش (دانون و همکاران[4] ۲۰۰۶)، این ایده بهتر از دستور نیومن عمل می‌کند، زمانی که اندازه جامعه تا حد زیادی متفاوت است. (بلوندل و همکاران[5] ۲۰۰۸) الگوریتم Louvain را پیشنهاد کردند، یک روش اکتشافی مبتنی بر بهینه‌سازی مدولاریته برای تشخیص جامعه در شبکه‌هایی با اندازه بی‌سابقه. از نظر زمان محاسباتی که با تعداد یال‌های گراف، یعنی $O(m)$ خطی است، از الگوریتم نیومن بهتر عمل کرد.

1 Li et al
2 Chen et al
3 Clauset et al
4 Danon et al
5 Blondel et al

- **آنیلینگ شبیه‌سازی شده**

این روش یک روش احتمالاتی برای بهینه‌سازی جهانی است که با به حداکثر رساندن مدولار بودن، جوامع در یک شبکه پیچیده را شناسایی می‌کند. (لیو و همکاران[1] ۲۰۱۰) از simulated annealing با الگوریتم‌های k - means استفاده کرده‌اند. این روش نه تنها جامعه را در یک شبکه پیچیده شناسایی می‌کند، بلکه گره مرکزی هر جامعه را نیز نشان می‌دهد. الگوریتم ارائه شده توسط (گیمرا و آمارال[2] ۲۰۰۵) نیز یک روش بهینه‌سازی مبتنی بر روش تبرید شبیه‌سازی شده برای تنظیم فرآیند جستجوی محلی است. با توجه به SA، الگوریتم ارائه شده توسط (گیمرا و آمارال، ۲۰۰۵) عملکرد خوبی در یافتن یک راه‌حل بهینه جهانی دارد. یکی دیگر از این الگوریتم این است که نیازی به دانش قبلی در مورد تعداد جوامع ندارد.

- **الگوریتم‌های ژنتیک**

الگوریتم‌های ژنتیک (GA) به عنوان زیرمجموعه‌ای از الگوریتم‌های فراابتکاری، تکنیک‌های بهینه‌سازی توانمندی هستند که از فرآیندهای تکامل بیولوژیکی و انتخاب طبیعی الهام گرفته شده‌اند. GA همچنین برای بهینه‌سازی مدولاریت شبکه Q به منظور تشخیص ساختار جامعه یک شبکه استفاده می‌شوند. قبلا از GA ها برای پارتیشن‌بندی یک گراف استفاده می‌شد. (تاسگین و همکاران[3] ۲۰۰۷) برای نخستین‌بار، از الگوریتم‌های ژنتیک جهت شناسایی ساختار جوامع در شبکه‌های پیچیده با تکیه بر بهینه‌سازی معیار مدولاریت بهره بردند. از مزایای بارز این رویکرد در مقایسه با روش‌های کلاسیک، عدم نیاز به تعیین پیش‌فرض تعداد جوامع در ساختار شبکه است. (پیزوتی[4] ۲۰۰۸) یک GA به نام GA-NET ارائه کرد که از مفهوم امتیاز جامعه برای نشان دادن کیفیت پارتیشن‌بندی شبکه‌های اجتماعی استفاده کرد. امتیاز جامعه به حداکثر رساندن پیوندهای داخلی در ساختار جامعه است. هنگامی که فقط همبستگی‌های واقعی همه گره‌ها در هر عملگر در نظر گرفته می‌شد، به طور موثر جستجوی نامعتبر را کاهش داد.

پیش از این، رویکردهای بهینه‌سازی عمدتا از یک شاخص یا تابع هدف واحد جهت سنجش کیفیت جوامع بهره می‌بردند، (گونگ و همکاران[5] ۲۰۱۲) یک الگوریتم تکاملی را بر اساس بهینه‌سازی دو هدف متناقض به نام‌های ارتباط نسبت منفی و برش نسبت معرفی کرد. به طور مشابه، (زنگ و همکاران[6] ۲۰۱۶) الگوریتم‌های تکاملی چند هدفه برای تشخیص جوامع در شبکه‌های اجتماعی پیشنهاد کردند. بررسی‌هایی نیز بر اساس شباهت صورت پذیرفته که قادر به تشخیص جوامع در هر دو جامعه جدا شده و هم‌پوشانی است.

- **الگوریتم‌های پویا**

ما در اینجا سه الگوریتم پویا در رابطه با فرآیندهای در حال اجرا بر روی گراف‌ها، یعنی Random Walk، مدل‌های چرخشی و همگام‌سازی را مورد بحث قرار خواهیم داد.

- **پیاده‌روی تصادفی**

برای ادغام گروه‌های مختلف با استفاده از رویکرد پایین به بالا، می‌توان از یک Random Walk برای شناسایی خوشه‌ها در یک گراف با عبور از گره‌ها به صورت تصادفی استفاده کرد. همه الگوریتم‌های Random Walk را می‌توان برای اعمال بر روی گراف‌های وزنی نیز گسترش داد. مطالعات مختلفی برای تشخیص جامعه با استفاده از Random Walk وجود دارد. در سال ۲۰۰۳، (ژو[7] ۲۰۰۳) فاصله بین یک جفت لبه را با استفاده از Random Walk تعریف کرد. فاصله بین دو راس میانگین تعداد لبه‌هایی است که یک Random Walk برای رسیدن از یک گره به گره دیگر طی می‌کند. لبه‌های

1 Liu et al
2 Guimera & Amaral
3 Tasgin et al
4 Pizzuti
5 Gong et al
6 Zeng et al
7 Zhou

نزدیک احتمالا متعلق به یک جامعه یا خوشه هستند. روشی به نام "Netwalk" پیشنهاد کردند که جوامع را در این Random Walk مغرضانه شناسایی می کند. Netwalk یک روش خوشه‌بندی سلسله مراتبی تجمعی است که در آن شباهت بین رئوس با نزدیکی آن‌ها نشان داده می‌شود.

- **مدل‌های چرخشی**

مدل‌های چرخشی در مکانیک آماری استفاده شده است. مدل Potts رویکرد رایج در این حوزه است. (ریچارت و همکاران[1] ۲۰۰۴) با الهام از ایده خوشه‌بندی ابرپارامغناطیسی داده‌ها، یک روش تشخیص جامعه را پیشنهاد کردند که گراف شبکه را بر روی یک مدل دمای صفر q Potts با تعامل نزدیک‌ترین همسایگان ترسیم می‌کرد. بعدا، (ریچارت و همکاران، ۲۰۰۴) تکنیک‌های اسپین گلاس را معرفی می‌کند که در آن هر رأس منفرد در یک حالت چرخشی فرض می‌شود. این الگوریتم علی‌رغم غیر قطعی بودن، دارای پارامترهای قابل تنظیم در رابطه با اندازه جامعه است.

- **همگام‌سازی**

پدیده همگام‌سازی در سیستم‌های متشکل از واحدهای جفت شده رخ می‌دهد؛ می‌تواند به عنوان رویکردی کارآمد در حل مسائل خوشه‌بندی و تشخیص جوامع نیز به کار گرفته شود. اگر نوسانگرها در رئوس با فاز تصادفی قرار گیرند، در مقایسه با جوامع دیگر، زودتر با جامعه‌ای که در آن حضور دارند، همگام می شوند. اگر زمان تکامل مجاز باشد، خوشه‌ها ممکن است در همگام‌سازی کامل در گراف شناسایی شوند. اولین بار توسط (آریناس و همکاران[2] ۲۰۰۶) نشان داده شد که تشخیص داد مقیاس‌های ساختاری در معرض تکنیک همگام‌سازی، گروه‌هایی از مقادیر ویژه ماتریس لاپلاسی گراف را نشان می‌دهند که به خوشه‌بندی گراف کمک می‌کند. بر اساس اصل همگام سازی، (بوکالتی و همکاران[3] ۲۰۰۷) از یک روش تشخیص جامعه حمایت کرد. این الگوریتم دارای پیچیدگی زمانی O(mn) است و نتایج خوبی را در معیار گیروان نیومن نشان می‌دهد. اشکال اصلی الگوریتم‌های مبتنی بر همگام‌سازی، غیرقابل اعتماد بودن این رویکرد در مورد جوامع با اندازه‌های متفاوت است.

- **الگوریتم‌هایی برای جوامع هم‌پوشانی**

هدف روش‌هایی که تاکنون مورد بحث قرار گرفته است، تشخیص جامعه‌ای از هم گسسته است. در اینجا ما الگوریتم‌هایی را برای جوامع هم‌پوشانی مورد بحث قرار خواهیم داد که می‌توانند برای جوامع مجزا نیز استفاده شوند. (گرگوری[4] ۲۰۰۷) الگوریتم نیومن گیروان خوشه-هم‌پوشانی (CONGA) را برای تشخیص جامعه هم‌پوشانی پیشنهاد کرد. الگوریتم پیشنهادی گونه‌ای از رویکرد خوشه‌بندی سلسله مراتبی سنتی گیروان و نیومن است که با یک تازگی در روش تقسیم راس گسترش یافته است. این الگوریتم پیچیدگی محاسباتی بالا را از الگوریتم GN به ارث برده است و بدترین حالت $O(m^3)$را در یک نمودار پراکنده دارد.

- **گسترش و بهینه‌سازی محلی**

این خط جستجو بر اساس به حداکثر رساندن یک تابع سود محلی است که کیفیت گره‌های بهم پیوسته را نشان می‌دهد. (باومز و همکاران[5] ۲۰۰۵) الگوریتمی برای تشخیص جامعه هم‌پوشانی بر اساس اسکن تکراری (IS) و حذف رتبه (RaRe) پیشنهاد کرد. RaRe گره‌ها را بر اساس معیارهای خاصی رتبه‌بندی می‌کند و گره‌های با رتبه‌بندی بالا حذف می‌شوند تا زمانی که هسته‌های خوشه‌های ناهمگون کوچک تشکیل شوند. این هسته‌ها همچنین به عنوان جوامع بذر برای فرآیند IS نامیده می‌شوند که یک بهینه‌سازی حریصانه را اجرا می‌کند و این جوامع دانه را تا زمانی که یک تابع چگالی بهبود نیابد گسترش می‌دهد. تابع چگالی بهینه سازی را می‌توان به صورت زیر بیان کرد:

---

1 Reichardt
2 Arenas et al
3 Boccaletti et al
4 Gregory
5 Baumes et al

$$F(c)= W^{c}_{in} / (W^{c}_{in}+W^{c}_{out}) \quad (۱)$$

که در آن w_{in} و w_{out} وزن‌های داخلی و خارجی جامعه c هستند . پیچیدگی زمانی این الگوریتم در بدترین حالت $o(n^2)$ است. کیفیت شناسایی جوامع به کیفیت بذر بستگی دارد. به دلیل حذف گره‌ها در طول گسترش، ممکن است اجزای disconnected ایجاد شود.

$$F(c)= W^{c}_{in} / (W^{c}_{in}+W^{c}_{out}) + \lambda ep \quad (۲)$$

که e_p احتمال لبه است. از درجه یک جامعه به عنوان ورودی تابع تناسب به جای وزن جوامع استفاده شده است.

$$F(c)= K^{c}_{in} / (K^{c}_{in}+K^{c}_{out})\alpha \quad (۳)$$

در حالی که k^{c}_{in} و k^{c}_{out} مجموع درجه داخلی و خارجی جامعه c و α پارامتر تفکیک‌پذیری برای کنترل اندازه جوامع هستند. پیچیدگی زمانی LFM $O(n_c s^2)$ است که در آن n_c تعداد جوامع و s اندازه متوسط جوامع است.

- **روش نفوذ CLIQUE (CPM)**

CPM بر پایه مفهوم تشکیل خوشه یا دسته‌ها توسط زیر گراف‌ها است که هر گره به گره‌های دیگر در یک clique متصل است. (پالا و همکاران[1] ۲۰۰۵) پیشنهاد کردند که لبه‌های داخلی با تراکم بالا، خوشه‌ها را تشکیل می‌دهند، در حالی که لبه‌های درون جامعه، خوشه‌ها را شکل نمی‌دهند. اصطلاح k - clique برای بیان یک گراف کامل با k - رئوس استفاده می‌شود. فرض بر این است که شبکه متشکل از k خوشه مجاور با یکدیگر به اشتراک گذاشته می‌شوند. هر clique در یک جامعه است؛ اما خوشه‌ها در جوامع مختلف می‌توانند گره‌ها را به اشتراک بگذارند. این روش اکتشافی در کشف هم‌پوشانی در خوشه‌های گره مفید است. هدف از بسط مسیر بحرانی cpm مورد بررسی قرار گرفت.

گراف‌های وزن‌دار، دو بخشی و جهت‌دار توسط (فارکاس و همکاران[2] ۲۰۰۷) یک آستانه برای وزن‌های clique به عنوان میانگین هندسی وزن تمام لبه‌ها را پیشنهاد کردند. مقدار آستانه تا حدی بزرگ‌تر از مقدار بحرانی انتخاب می‌شود که جامعه k - clique از آن استخراج شود. این روش به کسب ثروتمندترین نوع خوشه‌ها کمک می‌کند. اجرای سریع cpm در (کومپولا و همکاران[3] ۲۰۰۸) با نام الگوریتم نفوذ clique متوالی (scp) توسعه پیدا می‌کند. این روش از یک گراف خالی شروع می‌شود و با قرار دادن یال گراف تحت مطالعه، جوامع k - clique را شناسایی می‌کند. پیچیدگی زمانی این روش به صورت خطی به تعداد خوشه‌ها بستگی دارد؛ اما هنوز هم سریع‌تر از cpm اصلی است. در این روش می‌توان ساختارهای اجتماعی یک گراف را با قرار دادن یال‌ها در یک نقطه به ترتیب نزولی و با اضافه کردن هر یک از آن‌ها پس از اضافه کردن هر یال مشخص کرد. گاهی الگوریتم‌های cpm کل شبکه را پوشش نمی‌دهند، یعنی ممکن است بعضی از گره‌ها بدون توجه به اتصال با گره‌های دیگر به هیچ خوشه‌ای تعلق نداشته باشند. (مایتی و همکاران[4] ۲۰۱۴) یک الگوریتم جدید و cpm را برای اطمینان از اینکه هر نود متعلق به حداقل یک جامعه است پیشنهاد کردند. در ابتدا، جوامع با استفاده از cpm شناسایی می‌شوند و سپس برای گره‌های باقی مانده براساس ضریب تعلق هر گره توسعه داده می‌شوند. ضریب تعلق نشان می‌دهد که گره چگونه به یک جامعه خاص متصل می‌شود. ارزیابی این نسخه گسترش یافته برتری خود را نسبت به cpm کلاسیک ثابت کرد. در نهایت، (چاکرابورتی و همکاران[5] ۲۰۱۳) از cpm استفاده کردند و الگوریتمی

1 Palla et al
2 Farkas et al
3 Kumpula et al
4 Maity et al
5 Chakraborty et al

به نام OverCite را ابداع کردند که می‌تواند جوامع هم‌پوشانی را در شبکه استنادها شناسایی کند که حاوی اطلاعاتی درباره نویسندگان، مقالات و محل برگزاری آن است. به رغم سادگی مفهومی، روش cpm به جای الگوریتم تشخیص جامعه، شبیه الگوریتم تطبیق الگو است؛ زیرا آن‌ها تمایل به پیدا کردن ساختار خاص و محلی در شبکه دارند.

- **گراف خط و پارتیشن‌بندی لینک**

به جای پارتیشن‌بندی گره‌ها در مدل شبکه، پارتیشن‌بندی لینک‌ها در آن شبکه نیز در حوزه تشخیص جامعه مورد توجه قرار گرفته است. اگر پیوندهای متصل به آن به بیش از یک خوشه تعلق داشته باشند، در اینجا یک گره به عنوان هم‌پوشانی نامیده می‌شود. (آن و همکاران[1] ۲۰۱۰) پیوندها را با استفاده از خوشه‌بندی سلسله‌مراتبی شباهت لبه‌ها تقسیم کرد. پیچیدگی زمانی این الگوریتم $O(nk^2_{max})$ است که k_{max} حداکثر درجه گره در شبکه داده شده است. (ایوانز[2] ۲۰۱۰) گراف‌های خطی را برای تشخیص هم‌پوشانی جامعه که در آن دسته‌ها به عنوان گره‌های گراف وزنی عمل می‌کنند، نمودارهای خطی را گسترش داد. هیچ تضمینی برای تشخیص جامعه با کیفیت بالا با استفاده از تکنیک پارتیشن‌بندی لینک وجود ندارد؛ زیرا این الگوریتم‌ها بر تعریف مبهم از جامعه تکیه دارند.

- **الگوریتم های پویا و مبتنی بر عامل**

در الگوریتم انتشار برچسب (LPA) (راغوان[3] ۲۰۰۷) گره‌هایی که برچسب یکسانی دارند، تمایل به تشکیل یک جامعه دارند. هر رأس بازدید می‌شود و بر اساس رای‌گیری همسایگانش برچسبی به آن اختصاص پیدا می‌کند. این تکرار راس تا همگرایی ادامه دارد. LPA به دلیل مفهوم ساده و کارایی محاسباتی آن‌ها که خطی با تعداد یال‌ها در گراف است، یعنی O(m) خطی هستند. LPA را می‌توان برای شناسایی جوامع هم‌پوشانی گسترش داد. (گرگوری، ۲۰۰۷) الگوریتم انتشار هم‌پوشانی جامعه (COPRA) را پیشنهاد کرد که بسط LPA بود. در این الگوریتم، هر گره ضرایبی را از همسایگان خود دریافت می‌کند و آن‌ها را میانگین می‌گیرد تا ضریب تعلق خود را در هر مرحله زمانی به روز کند. برای ماتریس ورودی با اندازه $m \times n$، پیچیدگی زمانی این الگوریتم $O(vm \log(vm /m))$ در هر تکرار است، در حالی که v پارامتری برای کنترل ارتباط یک گره با چندین اجتماع است. (زی و همکاران[4] ۲۰۱۱) LPA را به الگوریتم انتشار برچسب speaker-listener (SLPA) گسترش داد تا جوامع هم‌پوشانی و گره‌های هم‌پوشانی را شناسایی کند. این یک فرآیند انتشار اطلاعات مبتنی بر speaker-listener است. برخلاف LPA که در آن یک گره برچسب را در تکرار قبلی فراموش می‌کند، SLPA برای هر گره حافظه‌ای برای ذخیره برچسب‌های تکرار قبلی فراهم می‌کند. مزیت خوب SLPA این است که از قبل به اطلاعاتی در مورد تعداد جوامع نیاز ندارد. پیچیدگی محاسباتی آن O(tm) است در حالی که t تعداد تکرارهای از پیش تعریف شده و m تعداد یال‌ها است. SLPA را می‌توان به شبکه‌های وزن‌دار و جهت‌دار با گنجاندن قوانین تعامل معروف به SLPAw گسترش داد.

- **تشخیص فازی**

این‌ها روش‌های انتشار برچسب هستند؛ اما همچنین می‌توانند با محاسبه ضریب تعلق یا یک بردار عضویت برای هر راس، جوامع هم‌پوشانی را در شبکه پیدا کنند. ابعاد این فاکتور باید از قبل ارائه شود که اشکال عمده این روش است. (نپوش و همکاران[5] ۲۰۰۸) از روش جامعه فازی برای حل مسئله تشخیص جامعه هم‌پوشانی به عنوان یک بهینه‌سازی محدود غیرخطی استفاده کرد. روش مورد استفاده در این رویکرد، simulated annealing است که تابع هدف را به حداقل رساندند.

1 Ahn et al
2 Evans
3 Raghavan
4 Xie et al
5 Nepusz et al

- **روش‌های استنتاج آماری**

خوشه‌بندی گراف نوعی مسئله استنتاج آماری است. شرح مختصری از برخی از روش‌های مبتنی بر استنتاج برای تشخیص هم‌پوشانی جامعه در گراف‌ها در زیر آورده شده است.

- **مدل بلوک تصادفی (SBM)**

در آمار، SBM یک مدل احتمالی است که نه تنها کاربرد خود را در تشخیص جامعه هم‌پوشانی پیدا می‌کند، بلکه نحوه تشکیل جوامع را نیز توصیف می‌کند. توسط (هالند و لاینهارت[1] ۱۹۷۶) پیشنهاد شد و توسط (فینبرگ و همکاران[2] ۱۹۸۵) اصلاح شد. مدل بلوک یک رویکرد بسیار رایج در مورد تجزیه‌وتحلیل شبکه، به عنوان مثال، تجزیه‌وتحلیل شبکه‌های اجتماعی است. این یک گراف را به دسته یال‌هایی که دارای ویژگی های مشترک هستند، تجزیه می‌کند.

- **روش‌های مبتنی بر استنتاج بیزی**

(یانگ و همکاران[3] ۲۰۱۱) یک مدل بلوک تصادفی پویا (DSBM) را برای مدل‌سازی جوامع و تکامل آن‌ها و روش بیزی را برای تخمین پارامترها پیشنهاد کردند. دو نسخه پیشنهاد شده است، یکی استنتاج آنلاین است که مدل را با زمان تغییر می‌دهد، در حالی که استنتاج آفلاین که مدل را با داده‌های به دست آمده در تمام مراحل زمانی در مقایسه با تحلیل شبکه‌ای که با استنتاج آنلاین سروکار دارد، یاد می‌گیرد. (قربانی و همکاران، ۲۰۱۶) یک آشکارساز جامعه هم‌پوشانی بیزی پویا (DBOCD) را پیشنهاد کردند. فرض بر این است که در هر وقفه زمانی شبکه، اجزای هم‌پوشانی جوامع، از جوامع پیوند به جای جوامع گره عمومی استفاده می‌کنند. این الگوریتم جوامع هم‌پوشانی را استخراج می‌کند و در عین حال ثبات جوامع را در طول زمان حفظ می‌کند. این روش عملکرد بهتری نسبت به روش‌های پویای اخیر برای تشخیص هم‌پوشانی جامعه دارد.

- **رویکردهای تجزیه ماتریس غیر منفی (NMF)**

بسیاری از الگوریتم‌ها، به ویژه روش‌های طیفی، جوامع را از مقادیر ویژه‌ای شناسایی می‌کنند که توضیح معنای فیزیکی آن‌ها با کاربردهای دنیای واقعی بسیار دشوار است. NMF به عنوان ابزاری مفید برای تجزیه‌وتحلیل داده‌ها با قابلیت تفسیر بهبود یافته ظهور کرده است. این به ویژه یک رویکرد اخیر برای کشف ویژگی‌های ساختاری و عملکردی یک شبکه پویا با جوامع هم‌پوشانی است. این یک الگوریتم یادگیری ماشین است که یک ماتریس ویژگی معین را تجزیه می‌کند تا ویژگی‌های یک ساختار معین را کشف کند (روسی و همکاران، ۲۰۱۳)

- **روش‌های مبتنی بر PCA**

(لین و همکاران[4] ۲۰۱۴) روشی مبتنی بر PCA برای تشخیص جوامع هم‌پوشان ارائه دادند. این الگوریتم از PCA برای انتخاب بهترین تعداد بردارهای ویژه استفاد و سپس از ماتریس لاپلاس برای نگاشت گره‌ها در زیرفضای کم‌بعد استفاده می‌کند. این الگوریتم از C- میانگین فازی (FCM) برای آشکار کردن ساختار هم‌پوشان در یک شبکه استفاده می‌کند. در FCM هر رأس دارای یک درجه عضویت مهر و موم شده است که تعلق آن به خوشه‌های مختلف را نشان می‌دهد. PCA برای استخراج اجزای اصلی گره‌های داده شده با بررسی پراکندگی مقادیر ویژه همسایه به منظور انتخاب تعداد بهینه بردارهای ویژه که نقش حیاتی در عملکرد الگوریتم‌های خوشه‌بندی طیفی دارند، استفاده می‌شود. هنگامی که PCA کار خود را انجام می‌دهد، ساختار جوامع هم‌پوشان با استفاده از تحلیل طیفی آشکار می‌شود.

(لی و همکاران، ۲۰۱۶) الگورتیم تشخیص جوامع کلاسیک مبتنی بر k- میانگین را با استفاده از رویکرد PCA بهبود بخشیدند. این الگوریتم در سه مرحله اجرا می‌شود. اول، فاصله بین گره‌های موجود در شبکه را محاسبه می‌کند. فاصله بین گره‌های متعلق به یک جامعه کوچک‌تر از فاصله بین گره‌های متعلق به گروه‌های دیگر است. دوم، گره‌ها در فضای

1 Holland & Leinhardt
2 Fienberg et al
3 Yang et al
4 Lin et al

p بعدی نگاشت می‌شوند. در نهایت، الگوریتم k-means برای آشکارسازی تعداد K جامعه در یک شبکه اعمال می‌شود. (یوان و همکاران[1] ۲۰۱۶) رویکرد دیگری را بر اساس PCA و شاخص عضویت (MI) به نام PCA-MI برای تشخیص جوامع هم‌پوشانی پیشنهاد کردند. PCA ویژگی‌های مربوط به شبکه پیچیده را استخراج می‌کند و سپس MI برای طبقه‌بندی گره‌های متعلق به جوامع مختلف اعمال می‌شود.

نتیجه‌گیری

در حوزه‌ی تحلیل شبکه‌های اجتماعی، گره‌ها (عناصر) به‌گونه‌ای سازمان‌دهی شده‌اند که می‌توان آن‌ها را در قالب گروه‌هایی تحت عنوان جامعه طبقه‌بندی کرد. ویژگی بنیادی این جوامع آن است که عناصر درون یک جامعه، بیشترین میزان شباهت و پیوستگی را با یکدیگر داشته و در مقابل، کمترین شباهت را با اعضای سایر جوامع دارا هستند. با توجه به اهمیت این ساختارها، تاکنون روش‌های متنوعی جهت شناسایی و تشخیص خودکار جوامع در شبکه‌های اجتماعی ارائه شده است. شناسایی جوامع در شبکه‌های اجتماعی از منظر محاسباتی با چالش‌های بنیادینی همراه است؛ چرا که علاوه بر نامعلوم بودن تعداد خوشه‌ها، تراکم درونی جوامع نیز ماهیتی ناهمگون دارد. با توجه به تنوع روش‌های تشخیص جامعه، انتخاب یا طراحی یک الگوریتم بهینه، به طور مستقیم به ویژگی‌های توپولوژیک شبکه و اهداف پژوهش وابسته است. از سوی دیگر، هر الگوریتم از رویکرد و ویژگی‌های متمایزی برخوردار است که منجر به طبقه‌بندی‌های متفاوتی در این حوزه شده است؛ لذا درک دقیق این دسته‌بندی‌ها و شناخت نقاط قوت و ضعف هر دسته، پیش‌شرطی اساسی برای انتخاب الگوریتم مناسب و دستیابی به نتایج معتبر به شمار می‌رود. از این رو در این مقاله به مطالعه الگوریتم‌های تشخیص جامعه در شبکه‌های اجتماعی پرداخته شده است. در حوزه شبکه‌های اجتماعی، تمایز و شناسایی جوامع از پیچیده‌ترین و اساسی‌ترین موضوعات پژوهشی به شمار می‌آید؛ زیرا ساختار درونی شبکه‌ها از روابط پویا، ناهمگون و گاه هم‌پوشان میان گره‌ها تشکیل شده است. روش‌های گوناگونی برای کشف این ساختارها ارائه شده‌اند که هر یک متناسب با نوع شبکه، هدف تحلیل و معیارهای ارزیابی عملکرد، کارایی متفاوتی دارند. در تحلیل تطبیقی میان رویکردهای اصلی شناسایی جامعه می‌توان دریافت که الگوریتم‌های سنتی همچون روش‌های خوشه‌بندی سلسله‌مراتبی و الگوریتم Girvan–Newman از نظر سادگی، شفافیت و درک مفهومی قابل اتکا هستند؛ اما در شبکه‌های با مقیاس بزرگ و پویا به دلیل پیچیدگی محاسباتی و عدم مقیاس‌پذیری عملکرد ضعیفی دارند. این روش‌ها بیشتر برای شبکه‌های ایستا، کوچک و بدون وزن مناسب هستند و تفسیر شهودی بالایی دارند. در مقابل، الگوریتم‌های مبتنی بر مدولاریتی مانند بهینه‌سازی اکستریم، شبیه‌سازی تبرید و نسخه‌های سریع‌تر GN با تعریف تابع هدف صریح، امکان ارزیابی کیفیت جداسازی جوامع را فراهم می‌سازند و در کشف ساختارهای چندسطحی کارآمد هستند. با این حال، این رویکردها ممکن است جوامع کوچک را در ساختارهای بزرگ‌تر ادغام کنند. در شبکه‌های پیچیده‌تر، روش‌های آماری و فاکتورگیری ماتریسی مانند NMF ، BNMF و BNMTF توانستند عملکرد بسیار بهتری در تشخیص اجتماعات هم‌پوشان از خود نشان دهند. این الگوریتم‌ها با مدل‌سازی احتمالاتی و تجزیه ساختار همبستگی داده‌ها، به جای تکیه بر صرف فاصله یا مدولاریتی، قادرند عضویت نسبی گره‌ها در چند جامعه را هم‌زمان لحاظ کنند. در شبکه‌های پویا و دارای وزن، نسخه‌های بیزی یا متقارن NMF نتایج دقیق‌تر و پایاتری ارائه داده‌اند، هرچند در برابر حجم داده‌های بسیار بزرگ نیازمند تنظیم پارامترهای پیچیده و منابع محاسباتی سنگین هستند. همچنین در مقایسه با روش‌های کلاسیک، این دسته رویکرد آماری و ماتریسی از نظر انعطاف‌پذیری در مواجهه با داده‌های نویزی و ساختارهای ضعیف، عملکرد بهتری دارند ولی تفسیر نتایج آن‌ها به‌راحتی روش‌های مبتنی بر گراف نیست. در شبکه‌های با مقیاس بزرگ، الگوریتم‌های مبتنی بر بهینه‌سازی حریصانه مانند Clauset یا Extremal Optimization تعادلی میان دقت و سرعت ایجاد کرده‌اند و به ویژه در ساختارهای بدون وزن یا جهت‌دار ساده کارآمد هستند. با این حال، هر چه پویایی شبکه و هم‌پوشانی جوامع بیشتر شود،

1 Yuan et al

از اثربخشی این روش‌ها کاسته می‌شود. در چنین شرایطی، روش‌های پویا نظیر الگوریتم‌های Random Walk و مدل‌های هم‌زمان‌سازی می‌توانند تغییرات زمانی ارتباطات را مدل کنند و برای تحلیل تحولات اجتماع در شبکه‌های تکاملی سودمند باشند، گرچه نسبت به داده‌های نویزی حساس‌ترند. در یک ارزیابی کلی از عملکرد الگوریتم‌ها بر اساس نوع شبکه می‌توان گفت: در شبکه‌های ایستا و بدون وزن، روش‌های مدولاریتی و تقسیم‌بندی گراف بهترین کارایی را دارند؛ در شبکه‌های پویا، مدل‌های تصادفی و هم‌زمان‌سازی مناسب‌ترند؛ در شبکه‌های جهت‌دار و وزنی، رویکردهای آماری و فاکتورگیری ماتریسی دقت بالاتری نشان می‌دهند؛ در شبکه‌های دارای اجتماعات هم‌پوشان، مدل‌های NMF و روش‌های استنباط بیزی مؤثرتر هستند و در شبکه‌های بسیار بزرگ، ترکیب رویکردهای حریصانه با بهینه‌سازی چندمرحله‌ای مؤثر واقع می‌شود. برآیند این مقایسه نشان می‌دهد که هیچ‌یک از دسته‌های یاد شده به تنهایی پاسخگوی تمام ابعاد مسئله نیست و جهت‌گیری فعلی پژوهش‌ها به سوی مدل‌های ترکیبی است که در آن‌ها ساختارهای گراف با یادگیری عمیق تلفیق می‌شود. به ویژه ترکیب شبکه‌های عصبی گرافی (GNN) با الگوریتم‌های مبتنی بر ماتریس‌های غیرمنفی یا مدل‌های طیفی، زمینه نوینی برای رفع محدودیت‌های فعلی فراهم کرده است. چنین رویکردهایی می‌توانند هم‌زمان با تحلیل وزن، جهت و پویایی شبکه، ساختارهای هم‌پوشان را با دقت بالاتری آشکار کنند. با وجود پیشرفت‌های بسیار، هنوز چالش‌هایی مانند طراحی چارچوب‌های جامع برای شبکه‌های دارای وزن و جهت، بازنمایی داده‌های ناهمگون و تعریف شاخص‌های ارزیابی باقی است. آینده پژوهش در کشف جامعه شبکه‌های اجتماعی در گرو حرکت به سمت مدل‌های هوشمند، تطبیق‌پذیر و میان‌رشته‌ای است که بتوانند مفاهیم نظری علوم اجتماعی را با الگوریتم‌های یادگیری عمیق و ساختارهای گرافی یکپارچه کنند.

محدودیت‌های پژوهش

با گسترش روزافزون شبکه‌های اجتماعی، برخی چالش‌های مطرح شده و برخی دیگر نیاز به بررسی دارند. اگرچه بسیاری از محققان راه‌حل‌هایی را برای شناسایی جوامع در شبکه‌های اجتماعی ثابت ارائه کرده‌اند؛ اما بسیاری دیگر به اندازه کافی مورد مطالعه قرار نگرفته‌اند و نیاز به بررسی دارند. امروزه خوشه‌بندی گراف در شبکه‌های پویا حوزه تحقیقاتی جالبی را نشان می‌دهد و برخی از مشکلات آن تاکنون پوشش داده نشده است.

- شبکه‌های اجتماعی ناهمگن هستند و شامل انواع مختلفی از یال‌ها هستند که نشان‌دهنده رابطه بین دو رأس هستند. بنابراین، جوامع بر روی شبکه‌ها به همراه پیوندهای اجتماعی انواع مختلفی از خوشه‌ها را تولید می‌کنند که دانستن چگونگی استخراج اطلاعات از روابط مختلف در شبکه‌های اجتماعی ناهمگون دشوار است.
- شبکه‌های اجتماعی به طور مداوم در حال توسعه هستند، توانایی عمل به این تغییرات سریع بسیار سخت است. علاوه بر این، رابطه اجتماعی بین هر دو کاربر و توپولوژی شبکه تکامل پیدا می‌کند. تحقیقات در این زمینه هنوز محدود است و بسیاری از مولدهای گراف باید برای طراحی مدل‌های تکاملی که ساختار و معنایی شبکه را ردیابی می‌کنند، تطبیق داده شوند. علاوه بر این، همه مولدهای گراف پویای موجود، گره‌های هم‌پوشانی را در نظر نمی‌گیرند که یک گره ممکن است به بیش از یک جامعه تعلق داشته باشد.
- طیف مختلف وب اجتماعی که شامل بسیاری از پلتفرم‌های رسانه‌های اجتماعی است، منجر به ایجاد چندین حساب برای یک کاربر می‌شود. جالب است که باید این پروفایل‌های کاربر توزیع شده را با هم ادغام کنیم تا بفهمیم چگونه چندین شبکه اجتماعی بر تکامل شبکه‌ها تأثیر می‌گذارد. علاوه بر این، ادغام چنین اطلاعاتی می‌تواند با گرد هم آوردن افراد با علایق مشابه در یک جامعه، به ایجاد یک سیستم توصیه مناسب کمک کند.

شناسایی جوامع در شبکه‌های در حال تکامل یک حوزه پژوهشی امیدوارکننده است. مقالات کمی ارائه شده است که می‌تواند با شبکه‌های پویای بزرگ سروکار داشته باشد. بنابراین، مطالعات بیشتری که شبکه‌های بزرگ را در نظر می‌گیرد، باید انجام شود. خاطر نشان می‌شود که روش‌ها و مطالعات، شبکه‌های پویای هم‌پوشانی جهت‌دار را در نظر می‌گیرد؛

موضوعی که برای تحقیقات آینده حیاتی است، کمیاب است. علاوه بر این، استفاده از سیستم‌های توزیع شده برای متعادل کردن بسیاری از روش‌های فعلی، یک جهت جالب در آینده است.

منابع

Ahn Y.-Y, Bagrow , J. P, Lehmann S. (2010). Link communities reveal multiscale complexity in networks, Nature 466 (7307) 761–764.

Aldecoa R, Marín I. (2013). “Exploring the limits of community detection strategies in complex networks”,Scientific reports.

Alotaibi N, Rhouma D. (2021). “A review on community structures detection in time evolving social networks”. Journal of King Saud University-Computer and Information Sciences.

Alvari H, Hajibagheri A, Sukthankar G. (2014). “Community detection in dynamic social networks: A game-theoretic approach”. In: Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. IEEE Press, pp. 101–107.

Arenas A, Diaz-Guilera A, Pérez-Vicente C. J. (2006). Synchronization reveals topological scales in complex networks, Physical review letters 96 (11) 114102.

Asur S, Parthasarathy S, Ucar D. (2009). “An event-based framework for characterizing the evolutionary behavior of interaction graphs”, ACM Transactions on Knowledge Discovery from Data (TKDD) 3 (4).

Aston N, Hu w. (2014). “Community detection in dynamic social networks”.Commun. Network 6 (02), 124.

Baumes J, Goldberg M. K, Krishnamoorthy M. S, Magdon-Ismail M, Preston N. (2005). “Finding communities by clustering a graph into overlapping subgraphs’., IADIS AC 5,) 97–104.

Bezdek J. C. (2013). “Pattern recognition with fuzzy objective function algorithms”, Springer Science & Business Media.

Blondel V. D, Guillaume J.-L, Lambiotte R, Lefebvre E. (2008). “Fast unfolding of communities in large networks”, Journal of statistical mechanics: theory and experiment 2008 (10).

Boccaletti S, Ivanchenko M, Latora V, Pluchino A, Rapisarda A. (2007). Detecting complex network modularity by dynamical clustering, Physical Review E 75 (4) 045102.

Boettcher S, Percus A. G. (2001). “Optimization with extremal dynamics”, Physical Review Letters 86 (23).

Chakraborty T, Chakraborty A. (2013). “Overcite: Finding overlapping communities in citation network”,in: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ACM, pp. 1124–1131.

Chen M, Kuzmin K, Szymanski B. K. (2014). “Community detection via maximization of modularity and its variants”, IEEE Transactions on Computational Social Systems 1 (1) , 46–65.

Chintalapudi S. R., Prasad M. K. (2015). A survey on community detection algorithms in large scale real world networks, in: Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on, IEEE, pp. 1323–1327.

Clauset A, Newman M. E, Moore C. (2004). “Finding community structure in very large networks”, Physical review E 70 (6).

Danon L, Diaz-Guilera A, Duch J, Arenas A. (2005). Comparing community structure identification, Journal of Statistical Mechanics: Theory and Experiment 2005 (09) P09008.

Danon L, Díaz-Guilera A, Arenas A. (2006). “The effect of size heterogeneity on community identification in complex networks”, Journal of Statistical Mechanics: Theory and Experiment 2006 (11).

Ding C. (2004). “A tutorial on spectral clustering”, in: Talk presented at ICML.(Slides available at http://crd.lbl.gov/~cding/Spectral/).

Djerbi R, Amad M, Imache R. (2020). “A new model for communities’ detection in dynamic social networks inspired from human families”. Int. J. Internet Technol.Secured Trans. 10 (1–2), 24–60.

Donath W. E, Hoffman A. J. (1973). “Lower bounds for the partitioning of graphs”, IBM Journal of Research and Development 17 (5) 420–425.

Erdös P, Rényi A. (1959). “On random graphs”, i, Publicationes Mathematicae (Debrecen) 6, 290–297.

Evans T. S. (2010). “Clique graphs and overlapping communities, Journal of Statistical Mechanics: Theory and Experiment 2010 (12).

Farkas I, Ábel D, Palla G, Vicsek T. (2007). “Weighted network modules”, New Journal of Physics 9 (6).

Fienberg S. E, Meyer M. M, Wasserman S. S. (1985). “Statistical analysis of multiple sociometric relations”, Journal of the american Statistical association 80 (389), 51–67.

Fiedler M. (1973). “Algebraic connectivity of graphs”, Czechoslovak mathematical journal 23 (2), 298–305.
Fortunato S. (2010). “Community detection in graphs”, Physics reports 486 (3), 75–174.
Fortunato S, Hric D. (2016). “Community detection in networks: A user guide”, Physics Reports 659, 1–44.
Gao L, Yang J, Wang H, Zhang H. (2010). “A measure of growth of user community in osns, in: Quality of Service (IWQoS)”, 18th International Workshop on, IEEE, pp. 1–2.
Girvan M, Newman M. E. (2002). “Community structure in social and biological networks”, Proceedings of the national academy of sciences 99 (12), 7821–7826.
Ghorbani M, Rabiee H. R, Khodadadi A. (2016). “Bayesian overlapping community detection in dynamic networks”, arXiv preprint arXiv:1605.02288.
Gong M, Ma L, Zhang Q, Jiao L. (2012). “Community detection in networks by using multiobjective evolutionary algorithm with decomposition”, Physica A: Statistical Mechanics and its Applications 391 (15), 4050–4060.
Gregory S. (2007). “An algorithm to find overlapping community structure in networks”, in: European Conference on Principles of Data Mining and Knowledge Discovery, Springer, pp. 91–102.
Guimera R, Amaral L. A. N. (2005). “Functional cartography of complex metabolic networks”, Nature 433 (7028) , 895–900.
Gulbahce N, Lehmann S. (2008). The art of community detection, BioEssays 30 (10) 934–938.
Hughes B. D. (1996). Random walks and random environments, Clarendon Press Oxford.
Holland P. W, Leinhardt S. (1976). “Local structure in social networks”, Sociological methodology 7 (1) , 1–45.
Jia H, Ding S, Du M. (2017). “A nyström spectral clustering algorithm based on probability incremental sampling”, Soft Computing 21 (19), 5815–5827.
Kernighan B. W, Lin S. (1970). An efficient heuristic procedure for partitioning graphs, Bell system technical journal 49 (2) 291–307.
Kim Y, Jeong H. (2011). Map equation for link communities, Physical Review E 84 (2) , 026110.
Kumpula J. M, Kivelä M, Kaski K, Saramäki J. (2008). Sequential algorithm for fast clique percolation,Physical Review E 78 (2).
Li L, Fan K, Zhangy Z, Xia Z. (2016). Community detection algorithm based on local expansion k-means, Neural Network World 26 (6) 589.
Li L, Du M, Liu G, Hu X, Wu G. (2014). “Extremal optimization-based semi-supervised algorithm with conflict pairwise constraints for community detection”, in: Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on, IEEE, pp. 180–187.
Lin L, Zheng-Min X, Li S.-H, Zhi-Huang S.-N. L. (2014). Detecting overlapping community structure via an improved spread algorithm based on pca, in: International Conference on Computer Science and Software Engineering (CSSE), DEStech Publications, pp. 115–121.
Liu X, Wei Y.-M, Wang J, Wang W.-J, He D.-X, Song Z.-J. (2016). Community detection enhancement using non-negative matrix factorization with graph regularization, International Journal of Modern Physics B 1650130.
Liu J, Liu T. (2010). Detecting community structure in complex networks using simulated annealing with k-means algorithms, Physica A: Statistical Mechanics and its Applications 389 (11) 2300–2309.
Liu C, Liu J, Jiang Z. (2014). “A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks”, IEEE transactions on cybernetics 44 (12), 2274–2287.
Lloyd S. (1982)." Least squares quantization in pcm" ,IEEE transactions on information theory 28 (2),129–137.
Malliaros F. D, Vazirgiannis M. (2013). Clustering and community detection in directed networks: A survey, Physics Reports 533 (4) 95–142.
Maity S, Rath S. K. (2014). “Extended clique percolation method to detect overlapping community structure”,in: Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on, IEEE, pp. 31–37.
Nepusz T, Petróczi A, Négyessy L, Bazsó F. (2008). “Fuzzy communities and the concept of bridgeness in complex networks”, Physical Review E 77 (1).
Newman M. E. (2003). “The structure and function of complex networks”, SIAM review 45 (2), 167–256.
Newman M. E. (2004). Detecting community structure in networks, The European Physical Journal BCondensed Matter and Complex Systems 38 (2) 321–330.
Newman M. E. (2006). Modularity and community structure in networks, Proceedings of the national academy of sciences 103 (23) 8577–8582.

Newman M. E, Girvan M. (2004). Finding and evaluating community structure in networks, Physical review E 69 (2) 026113.
Maqbool O, Babri H. A. (2004). "The weighted combined algorithm: A linkage algorithm for software clustering", in: Software Maintenance and Reengineering, 2004. CSMR 2004. Proceedings. Eighth European Conference on, IEEE, pp. 15–24.
Moscato V, Sperlì G. (2021). " A survey about community detection over On-line Social and Heterogeneous Information Networks" , Knowledge-Based Systems.
Palla G, Derényi I, Farkas I, Vicsek T. (2005). "Uncovering the overlapping community structure of complex networks in nature and society", Nature 435 (7043), 814–818.
Papadopoulos S, Kompatsiaris Y, Vakali A, Spyridonos P. (2012). Community detection in social media, Data Mining and Knowledge Discovery 24 (3) 515–554.
Pizzuti C. (2008). "A genetic algorithm for community detection in social networks", in: PPSN,Springer, , pp. 1081–1090.
Psorakis I, Roberts S, Ebden M, Sheldon B. (2011). Overlapping community detection using Bayesian non-negative matrix factorization, Physical Review E 83 (6) 066114.
Porter M. A, Onnela J.-P, Mucha P. J. (2009). Communities in networks, Notices of the AMS 56 (9) 1082–1097.
Pons P. (2007). Détection de communautés dans les grands graphes de terrain, Ph.D. thesis, Paris 7.
Pothen A. (1997). Graph partitioning algorithms with applications to scientific computing, in: Parallel Numerical Algorithms, Springer, pp. 323–368.
Raghavan U. N, Albert R, Kumara S. (2007). "Near linear time algorithm to detect community structures in large-scale networks", Physical review E 76 (3).
Rattigan M. J, Maier M, Jensen D. (2007). "Graph clustering with network structure indices", in: Proceedings of the 24th international conference on Machine learning, ACM, pp. 783–790.
Reichardt J, Bornholdt S. (2004). Detecting fuzzy community structures in complex networks with a potts model, Physical Review Letters 93 (21) 218701.
Reichardt J, Bornholdt S. (2006). "Statistical mechanics of community detection", Physical Review E 74 (1).
Rhouma. D, Romdhane L.B. (2018). "An efficient multilevel scheme for coarsening large scale social networks". Appl. Intell. 48 (10), 3557–3576, 2018.
Rossi R. A, Gallagher B, Neville J, Henderson K. (2013). Modeling dynamic behavior in large evolving graphs, in: Proceedings of the sixth ACM international conference on Web search and data mining, ACM, pp. 667–676.
Roux M. (2015). "A comparative study of divisive hierarchical clustering algorithms", 2015 .
Shen H, Cheng X, Cai K, Hu M.-B. (2009). Detect overlapping and hierarchical community structure in networks, Physica A: Statistical Mechanics and its Applications 388 (8) 1706–1712.
Stam C. J. (2014). "Modern network science of neurological disorders", Nature Reviews Neuroscience 15 (10), 683–695.
Tasgin M, Herdagdelen A, Bingol H. (2007). Community detection in complex networks using genetic algorithms, arXiv preprint arXiv:0711.0491.
Tyler J. R, Wilkinson D. M, Huberman B. A. (2005). "E-mail as spectroscopy: Automated discovery of community structure within organizations", The Information Society 21 (2), 143–153.
Von Luxburg U. (2007). "A tutorial on spectral clustering, Statistics and computing", 17 (4), 395–416.
Wang F, Li T, Wang X, Zhu S, Ding C. (2011). Community discovery using nonnegative matrix factorization, Data Mining and Knowledge Discovery 22 (3) 493–521.
Williams C. K, Seeger M. (2001). "Using the nyström method to speed up kernel machines", in: Advances in neural information processing systems, pp. 682–688.
Wilkinson D. M, Huberman B. A. (2004). A method for finding communities of related genes, proceedings of the national Academy of sciences 101 (suppl 1) 5241–5248.
Xie J, Szymanski B. K, Liu X. (2011). "Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process", in: Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on, IEEE, pp. 344–349.
Xie J, Kelley S, Szymanski B. K. (2013). Overlapping community detection in networks: The state-of-the-art and comparative study, Acm computing surveys (csur) 45 (4) 43.
Yang B, Liu D, Liu J. (2010). Discovering communities from social networks: Methodologies and applications, in: Handbook of social network technologies and applications, Springer, pp. 331–346.
Yang J, McAuley J, Leskovec J. (2013). "Community detection in networks with node attributes, in: Data Mining (ICDM)", IEEE 13th international conference on, IEEE, pp. 1151–1156.

Yang T, Chi Y, Zhu S, Gong Y, Jin R. (2011). "Detecting communities and their evolutions in dynamic social networks—a bayesian approach", Machine learning 82 (2), 157–189.

Yuan P, Wang W, Song M. (2016). Detecting overlapping community structures with pca technology and member index, in: Proceedings of the 9th EAI International Conference on Mobile Multimedia Communications, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering) , pp. 121–12.

Zarei M, Izadi D, Samani K. A. (2009). Detecting overlapping community structure of networks based on vertex–vertex correlations, Journal of Statistical Mechanics: Theory and Experiment 2009 (11) P11013.

Zeng Y, Liu J. (2015). "Community detection from signed social networks using a multi-objective evolutionary algorithm", in: Proceedings of the 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems, Volume 1, Springer, pp. 259–270.

Zhao K, Zhang S.-W,Pan Q. (2010). Fuzzy analysis for overlapping community structure of complex network, in: 2010 Chinese Control and Decision Conference, IEEE, pp. 3976–3981.

Zhou H. (2003). "Distance, dissimilarity index, and network community structure", Physical review e 67 (6).

Zhang Y, Yeung D.-Y. (2012). Overlapping community detection via bounded nonnegative matrix trifactorization, in: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp. 606–614.

Zhang L, Ye Q, Shao Y, Li C, Gao H. (2014). "An efficient hierarchy algorithm for community detection in complex networks", Mathematical Problems in Engineering.

Zhou H, Lipowsky R. (2004). Network brownian motion: A new method to measure vertex-vertex proximity and to identify communities and subcommunities, in: International conference on computational science, Springer, pp. 1062–1069.