

بهبود تشخیص جوامع همپوشان با توسعه الگوریتم خوشه‌بندی میانگین-C فازی-امکانی (PFCM) مبتنی بر وزن‌دهی مرکزیت شبکه

محمد مهدی کیخا* و حسن رضایی

گروه علوم کامپیوتر، دانشگاه سیستان و بلوچستان، زاهدان، ایران

تاریخ دریافت: ۱۴۰۴/۱۲/۰۳

تاریخ پذیرش: ۱۴۰۵/۰۳/۱۵

نوع مقاله: علمی-پژوهشی

چکیده. آشکارسازی ساختارهای همپوشان در شبکه‌های پیچیده، به دلیل ماهیت مبهم مرزهای جوامع و حضور گره‌هایی با تعلقات چندگانه، چالش‌های قابل‌توجهی را در حوزه داده‌کاوی گراف ایجاد کرده است. اگرچه الگوریتم‌های خوشه‌بندی فازی (FCM) ابزاری کارآمد برای مدل‌سازی این عدم قطعیت فراهم می‌کنند، اما وابستگی آن‌ها به قیود احتمالی و حساسیت نسبت به مقادیر اولیه، کارایی آن‌ها را در مواجهه با نویزهای ساختاری محدود می‌سازد. در این پژوهش، با هدف غلبه بر این محدودیت‌ها و بهره‌گیری هم‌زمان از ویژگی‌های توپولوژیک شبکه و منطق فازی، الگوریتم CW-PFCM (خوشه‌بندی C- میانگین فازی - امکانی وزن‌دار مبتنی بر مرکزیت) پیشنهاد شده است. در این رویکرد، از معیارهای مرکزیت شبکه برای هدایت فرآیند مقداردهی اولیه و تعدیل ماتریس‌های عضویت در الگوریتم PFCM استفاده می‌شود. ادغام رویکرد امکانی با منطق فازی، به مدل اجازه می‌دهد تا تمایز دقیق‌تری میان همپوشانی معنادار و گره‌های دورافتاده قائل (ادامه دارد)

2010 Mathematics Subject Classification. 03E72 ; 62A86

* Corresponding author

E-mails: keikha@cs.usb.ac.ir and hrezaei@cs.usb.ac.ir.

عبارات و کلمات کلیدی. تشخیص جوامع همپوشان، خوشه‌بندی فازی-امکانی، عدم قطعیت ساختاری، معیارهای مرکزیت، تحلیل شبکه‌های پیچیده.

شود. ارزیابی‌های تجربی بر روی چندین مجموعه داده استاندارد و مقایسه با روش‌های نوین، نشان می‌دهد که روش پیشنهادی ضمن حفظ پایداری در برابر نویز، از حیث معیارهای اعتبارسنجی نظیر اطلاعات متقابل نرمال‌شده (NMI) و مادولاریتی، عملکردی قابل قبول و رقابتی ارائه می‌دهد. یافته‌ها حاکی از آن است که لحاظ کردن اهمیت ساختاری گره‌ها در فرآیند یادگیری فازی، منجر به شناسایی دقیق‌تر جوامع واقعی می‌گردد.

۱. مقدمه

در دهه اخیر، علم شبکه به عنوان چارچوبی بنیادین برای درک رفتار سیستم‌های پیچیده تثبیت شده است. یکی از جذاب‌ترین ویژگی‌های این سیستم‌ها در مقیاس میانی، تمایل گره‌ها به تشکیل دسته‌هایی متراکم و درهم‌تنیده است که جوامع نامیده می‌شوند. شناسایی دقیق این ساختارها فراتر از یک مساله نظری است و کاربردهای عملی گسترده‌ای دارد؛ از کشف الگوهای تعاملات پروتئینی در زیست‌شناسی محاسباتی و تحلیل دینامیک شبکه‌های عصبی مغز گرفته تا طراحی سیستم‌های توصیه‌گر هوشمند و حتی مدل‌سازی انتشار اطلاعات یا حملات سایبری در شبکه‌های اجتماعی [۹، ۲۱]. با این حال، تعریف مفهوم جامعه در دنیای واقعی با چالش‌های معنایی عمیقی روبروست. برخلاف دیدگاه‌های کلاسیک که شبکه را به جزایر مجزا تقسیم می‌کردند، مشاهدات تجربی نشان می‌دهد که مرزهای بین جوامع اغلب مبهم و سیال هستند. یک پژوهشگر ممکن است هم‌زمان در چندین گروه علمی عضو باشد و یک پروتئین در چندین فرآیند سلولی نقش ایفا کند. نادیده گرفتن این همپوشانی و اصرار بر افراز سخت گراف، به معنای حذف بخشی از واقعیت و تقلیل غنای اطلاعاتی شبکه است [۲۴].

پذیرش همپوشانی به عنوان یک واقعیت ذاتی، پژوهشگران را به سمت استفاده از رویکردهایی سوق داده است که توانایی مدل‌سازی عدم قطعیت را داشته باشند. در این میان، نظریه مجموعه‌های فازی و الگوریتم‌های مبتنی بر آن، به ویژه خوشه‌بندی میانگین-C فازی (FCM)، به دلیل تطابق منطقی با ماهیت غیرقطعی جوامع، مورد توجه قرار گرفته‌اند. این روش‌ها با تعمیم مفهوم عضویت از حالت قطعی دوارزشی به حالت پیوسته، امکان مدل‌سازی عدم قطعیت ذاتی داده‌ها را فراهم می‌سازند و ابزاری قدرتمند برای توصیف گره‌های پل^۱ فراهم می‌کنند. با این وجود، چالش اصلی به کارگیری مستقیم الگوریتم‌های خوشه‌بندی داده‌محور بر روی ساختارهای گراف، زمانی آشکار می‌شود که با داده‌های نویزی یا گره‌های دورافتاده^۲ مواجه

^۱Bridge Nodes

^۲Outliers

می‌شویم. در نسخه استاندارد FCM، یک قید احتمالاتی سخت‌گیرانه وجود دارد که مجموع عضویت‌های هر گره در تمام خوشه‌ها را ملزم به برابر بودن با عدد یک می‌کند. این قید ریاضی، اگرچه از نظر آماری موجه به نظر می‌رسد، اما در عمل باعث می‌شود گره‌های دورافتاده که هیچ شباهتی به هیچ‌یک از جوامع ندارند، به اجبار با درجات عضویت بالا به یک یا چند خوشه نسبت داده شوند و مراکز خوشه‌ها را به سمت خود منحرف کنند [۸].

برای غلبه بر این محدودیت ذاتی، تلاش‌های نظری متعددی صورت گرفته است که برجسته‌ترین آن‌ها توسعه الگوریتم میانگین-C امکانی (PCM) بوده است [۳، ۷]. این الگوریتم با حذف قید مجموع یک و استفاده از مفهوم امکان به جای احتمال، اجازه می‌دهد تا عضویت یک گره در تمام خوشه‌ها پایین باشد، که این ویژگی برای طرد کردن نویزها ایده‌آل است. اما PCM نیز دارای مشکلات زیادی در مواجهه برای حل این مساله می‌باشد، زیرا آزادی عمل بیش از حد آن، گاهی منجر به همگرایی تمام خوشه‌ها به یک نقطه واحد می‌شود. به منظور بهره‌گیری از مزایای هر دو رویکرد و حذف معایب آن‌ها، الگوریتم ترکیبی میانگین-C فازی-امکانی (PFCM) معرفی شد که تعادلی میان افراز احتمالی و استقلال امکانی برقرار می‌کند. تحقیقات اخیر نشان داده‌اند که PFCM پایداری بسیار بالاتری نسبت به روش‌های پیشین دارد و در مواجهه با داده‌های ساختاریافته عملکردی درخشان ارائه می‌دهد. علاوه بر این، نشان داده شده است الگوبرداری از روش‌های فاکتورگیری متقارن تطبیقی در الگوریتم PFCM می‌تواند تضمین‌کننده همگرایی سریع‌تر در داده‌های ساختاریافته باشد [۳۱].

تحلیل ساختار جوامع امروزه با چالش‌های نوظهوری روبروست. تحقیقات اخیر نشان می‌دهد تمرکز از خوشه‌بندی ساده به سمت شناسایی حفره‌های ساختاری^۱ و گره‌های فریبده معطوف شده است [۲۶، ۲۸]. همچنین، مسائلی نظیر دستکاری ساختار توسط گره‌های مخرب [۱۰] و حملات خصمانه در شبکه‌های چندلایه، نیاز به روش‌هایی را برجسته می‌کند که فراتر از افرازهای سنتی عمل کنند [۲۵، ۱۲]. با وجود پیشرفت‌های الگوریتمی در حوزه فازی، یک شکاف مفهومی بزرگ در کاربرد این روش‌ها برای تشخیص جوامع شبکه وجود دارد: غفلت از ساختار و سلسله‌مراتب گره‌ها. اکثر الگوریتم‌های خوشه‌بندی موجود، شبکه را صرفاً مجموعه‌ای از نقاط در فضای برداری می‌بینند و در فرآیند یادگیری، وزنی برابر برای تمام نمونه‌ها قائل هستند از آنجا که بسیاری از شبکه‌های واقعی دارای توزیع درجه توانی^۲ هستند، به‌کارگیری این فرض در تحلیل شبکه‌های پیچیده می‌تواند منجر به تفسیرهای گمراه‌کننده شود.

^۱Structural Holes

^۲Scale free

در یک شبکه واقعی، گره‌هایی با مرکزیت بالا^۱ نقش ستون‌های نگهدارنده ساختار جامعه را دارند، در حالی که گره‌های حاشیه‌ای^۲ کمترین تأثیر را در تعریف هویت جامعه ایفا می‌کنند. وقتی الگوریتم در محاسبه مراکز خوشه‌ها، تأثیر یک گره مرکزی قدرتمند را با یک گره حاشیه‌ای یا نویزی برابر می‌بیند، عملاً از اطلاعات ارزشمند ساختاری چشم‌پوشی کرده است. این نقیصه باعث می‌شود که مراکز خوشه‌ها به جای قرارگیری در نواحی متراکم شبکه، به سمت نواحی کم‌اهمیت جابجا شوند و کیفیت تشخیص جوامع افت کند [۴].

در این پژوهش، با هدف پر کردن خلاء میان دقت الگوریتم‌های فازی ترکیبی و واقعیت توپولوژیک شبکه‌ها، روشی نوین با عنوان CW-PFCM (خوشه‌بندی میانگین-C فازی - امکانی وزن‌دار مبتنی بر مرکزیت) پیشنهاد شده است. برخلاف رویکردهای سنتی که فرآیند را با حدس‌های تصادفی آغاز می‌کنند و در مسیر بهینه‌سازی کور هستند، روش پیشنهادی از ساختار شبکه به عنوان یک راهنما بهره‌گیری نموده و درجه مرکزیت گره‌ها در شبکه را به ساختار معادلات خوشه‌بندی وارد نموده است تا بتواند نسبت به تشخیص جوامع همپوشان اقدام نماید. در روش پیشنهادی، نخست در مرحله مقداردهی اولیه، الگوریتم با تحلیل شاخص‌های مرکزیت، گره‌های تأثیرگذار و قانونی را شناسایی کرده و آن‌ها را به عنوان بذره‌های^۳ اولیه خوشه‌ها برمی‌گزیند. این استراتژی با یافته‌های یو و همکاران همسو است که نشان دادند انتخاب هوشمند بذرها راهکاری کارآمد برای مدیریت همپوشانی‌هاست [۳۱]. این استراتژی هوشمندانه، خطر گرفتار شدن در بهینه‌های محلی را که نقطه ضعف الگوریتم‌های خوشه‌بندی فازی در این مساله است، به شدت کاهش می‌دهد. در این روش همچنین، در ساختار تابع هدف، یک مکانیسم وزن‌دهی تطبیقی تعبیه شده است که باعث می‌شود در هر تکرار از الگوریتم، گره‌های دارای مرکزیت بالاتر، نیروی گرانشی بیشتری بر مراکز خوشه‌ها وارد کنند و عملاً به عنوان لنگرهای ثبات عمل نمایند. در مقابل، گره‌های نویزی یا کم‌اهمیت که دارای مرکزیت پایینی هستند، جریمه شده و تأثیر آن‌ها در جابجایی مراکز خنثی می‌شود. نتیجه این هم‌افزایی، الگوریتمی است که نه تنها همپوشانی‌ها را با مفهوم فازی مدل‌سازی می‌کند، بلکه به دلیل آگاهی از ساختار شبکه، مقاومتی ذاتی در برابر نویز و نوسانات توپولوژیک از خود نشان می‌دهد.

ساختار مقاله در ادامه بدین صورت سازمان‌دهی شده است: بخش دوم به مرور ادبیات موضوع و بررسی تحلیلی کارهای پیشین در مساله تشخیص جوامع می‌پردازد. در بخش سوم، فرمول‌بندی ریاضی و جزئیات پیاده‌سازی الگوریتم پیشنهادی CW-PFCM تشریح می‌شود.

¹Hubs

²Leaf nodes

³Seeds

بخش چهارم به ارائه نتایج تجربی، تحلیل حساسیت پارامترها و مقایسه عملکرد روش با الگوریتم‌های مطرح بر روی پایگاه داده‌های استاندارد اختصاص دارد و در نهایت، بخش پنجم یافته‌های پژوهش را جمع‌بندی و مسیرهای آتی را ترسیم می‌کند.

۲. مرور ادبیات و پیشینه پژوهش

تحلیل ساختار جوامع در شبکه‌های پیچیده، به عنوان یکی از ارکان اصلی درک سیستم‌های تعاملی، از دیرباز مورد توجه پژوهشگران علوم داده و فیزیک آماری بوده است. در حالی که رویکردهای سنتی تقسیم‌بندی گراف، مانند روش‌های برش طیفی و افرازهای سخت، بر این فرض استوار بودند که هر گره تنها متعلق به یک گروه است [۲۱]، واقعیت شبکه‌های مدرن (نظیر شبکه‌های اجتماعی، بیولوژیکی و تعاملات پروتئینی) حاکی از ساختاری پیچیده‌ای است که در آن گره‌ها نقش‌های چندگانه‌ای ایفا می‌کنند.

گیروان و نیومن در تحقیقات بنیادین خود با معرفی معیار مادولاریتی، استاندارد برای سنجش کیفیت جوامع ارائه دادند، اما روش آن‌ها اساساً برای جوامع غیرهمپوشان طراحی شده بود [۹]. با درک محدودیت‌های افراز سخت، پالا و همکارانش روش نفوذ کلیک (CPM) را معرفی کردند که با تعریف جامعه به عنوان اجتماعی از زیرگراف‌های کامل (کلیک‌ها)، برای نخستین بار مفهوم همپوشانی را به صورت سیستماتیک مدل‌سازی کردند [۲۴]. با این حال، وابستگی شدید CPM به وجود کلیک‌های کامل باعث شد تا این روش در شبکه‌های خلوت^۱ که چگالی یال پایینی دارند، کارایی خود را از دست بدهد و بخش زیادی از شبکه را بدون برچسب رها کند [۸].

برای غلبه بر محدودیت‌های باینری (صفر و یک) در عضویت گره‌ها، منطق فازی به عنوان ابزاری قدرتمند وارد میدان شد. الگوریتم میانگین-C فازی (FCM) که توسط بزدک توسعه یافت، با معرفی تابع عضویت پیوسته $u_{ij} \in [0, 1]$ ، اجازه داد هر گره با درجات مختلف به چندین خوشه تعلق داشته باشد [۳]. اگرچه FCM در تشخیص همپوشانی‌ها موفق‌تر از روش‌های قطعی عمل کرد، اما دیو و کریشناپورام در نقدی مشهور نشان دادند که قید احتمالی موجود در این الگوریتم (که مجموع درجات عضویت یک گره را مجبور به یک شدن می‌کند)، نقطه ضعف آن در برابر نویز و داده‌های دورافتاده است [۷]. در شبکه‌های واقعی که اغلب حاوی نویز هستند، این قید ریاضی الگوریتم را مجبور می‌کند تا به گره‌های دورافتاده و بی‌ارزش،

^۱Sparse Networks

درجه عضویت بالایی نسبت دهد تا شرط مجموع یک برقرار شود؛ امری که منجر به انحراف مراکز خوشه‌ها می‌شود.

در پاسخ به این چالش، کریشناپورام و کلر، الگوریتم میانگین-C امکانی (PCM) را پیشنهاد دادند. آن‌ها با حذف قید مجموع یک، عضویت را نه به عنوان یک احتمال نسبی، بلکه به عنوان میزان سازگاری مطلق تعریف کردند [۱۴]. این رویکرد اگرچه مشکل نويز را به خوبی مدیریت کرد، اما خود دچار معضل جدیدی به نام خوشه‌های منطبق^۱ شد. در غیاب نیروی رقابتی میان خوشه‌ها (که در FCM وجود داشت)، مراکز خوشه‌ها در PCM تمایل دارند همگی به چگال‌ترین ناحیه فضای ویژگی همگرا شوند و عملاً تفکیک جوامع رخ ندهد. این تناقض میان حساسیت به نويز در FCM و همگرایی خوشه‌ها در PCM، پالا و همکارانش را بر آن داشت تا با ترکیب مزایای هر دو روش، الگوریتم میانگین-C فازی - امکانی (PFCM) را معرفی کنند [۲۴]. فرمول‌بندی PFCM با داشتن هم‌زمان جملات فازی و امکانی در تابع هدف، تعادلی میان ثبات در برابر نويز و تفکیک‌پذیری ایجاد می‌کند و تاکنون یکی از موفق‌ترین روش‌ها در خوشه‌بندی داده‌های دارای عدم قطعیت بوده است.

در سال‌های اخیر، روش‌های فاکتورگیری ماتریس (NMF) با تمرکز بر تقارن گراف [۱۷]، [۱۳] و مدل‌های گسسته [۲۹] توسعه یافته‌اند، هرچند در وزن‌دهی ویژگی‌ها چالش دارند [۵]. همزمان، رویکردهای یادگیری عمیق و تکاملی نظیر شبکه‌های کانولوشن گراف علامت‌دار [۳۵]، بهینه‌سازی چندهدفه [۳۲] و روش‌های مبتنی بر کرنل [۳۴] دقت را بهبود داده‌اند. با این حال، پیچیدگی محاسباتی بالا و ماهیت جعبه‌سیاه در مدل‌های عمیق [۶] و نیاز به کدگذاری پیوسته در روش‌های تکاملی [۳۳]، کاربرد آن‌ها را در شرایط عدم قطعیت محدود می‌سازد.

با این وجود، تحقیقات سال‌های اخیر نشان می‌دهد که نسخه استاندارد PFCM برای داده‌های پیچیده و غیرخطی امروزی کافی نیست. ما و همکاران در مطالعه اخیر خود تحت عنوان PMWFCM، نشان دادند که استفاده از فاصله اقلیدسی ساده در PFCM نمی‌تواند ساختارهای غیرخطی پیچیده را در داده‌های ترافیکی تفکیک کند [۱۸]. آن‌ها با پیشنهاد یک رویکرد مبتنی بر کرنل چندگانه^۲، داده‌ها را به فضای ویژگی با ابعاد بالاتر نگاشت کردند تا جداسازی خطی خوشه‌ها ممکن شود. نتایج کار آن‌ها ثابت کرد که ترکیب وزن‌دهی امکانی با توابع کرنل، دقت طبقه‌بندی رفتارهای رانندگی را به طرز چشمگیری افزایش می‌دهد. همسو با این رویکرد، ال-اندولی و همکاران در پژوهشی دیگر بر روی تشخیص جوامع همپوشان

^۱ Coincident Clusters

^۲ Multi-Kernel

مقاوم، از ترکیب شبکه‌های عصبی گراف (GCN) با منطق فازی بهره بردند [۱]. استدلال اصلی آن‌ها این بود که روش‌های فازی کلاسیک صرفاً از ویژگی‌های گره‌ها استفاده می‌کنند و از اطلاعات ساختاری غافل هستند. آن‌ها با استفاده از یک خودکدگذار گراف^۱، ابتدا ویژگی‌های نهفته شبکه را استخراج کرده و سپس خوشه‌بندی فازی را روی این فضای جدید اعمال نمودند. اگرچه تحقیقاتی نظیر [۲۷] و [۲۳] گام‌های بلندی در بهبود PFCM برداشته‌اند، اما یک شکاف نظری مهم در آن‌ها مشاهده می‌شود: بی‌توجهی به ناهمگنی مرکزیت در سطح تابع هدف. در تمامی روش‌های مذکور، اعم از کرنل‌بنا یا مبتنی بر یادگیری عمیق، در مرحله‌نهایی خوشه‌بندی، تمام گره‌ها وزن یکسانی در به‌روزرسانی مراکز خوشه‌ها دارند. این در حالی است که در شبکه‌های پیچیده، گره‌های مرکزی و گره‌های پل نقشی بسیار تعیین‌کننده‌تر از گره‌های حاشیه‌ای در تعریف هویت یک جامعه دارند. چنگ و همکاران در جدیدترین مطالعه خود پیرامون تعبیه شبکه آگاه از مرکزیت، نشان دادند که نادیده گرفتن شاخص‌های مرکزیت باعث می‌شود الگوریتم‌های تشخیص جامعه در مواجهه با گره‌های مرزی دچار ابهام شوند [۶]. با این حال، حتی در مدل پیشنهادی آن‌ها نیز مرکزیت عمدتاً به عنوان یک ویژگی ورودی یا ابزار پیش‌پردازش استفاده شده و مستقیماً وارد هسته ریاضیاتی تابع هدف خوشه‌بندی نشده است. علاوه بر این، تحقیقات اخیر نشان داده بود که ادغام اطلاعات ساختاری (مانند PageRank) با فرآیند خوشه‌بندی می‌تواند همگرایی الگوریتم را تسریع کند، اما راهکار آن‌ها محدود به روش‌های قطعی بود و انعطاف‌پذیری منطق فازی-امکانی را نداشت [۴]. فقدان مکانیزمی که بتواند اهمیت ساختاری گره را مستقیماً در فرمول به‌روزرسانی مراکز PFCM دخیل کند، باعث می‌شود که نویزهای موجود در گره‌های کم‌اهمیت (حاشیه‌ای)، تأثیری برابر با گره‌های کلیدی بر جایگاه نهایی مراکز جامعه داشته باشند.

بنابراین، مرور ادبیات موجود آشکار می‌سازد که اگرچه الگوریتم‌های ترکیبی (مانند GCN-FCM یا Kernel-PFCM) توانسته‌اند مشکل غیرخطی بودن و نمایش ویژگی‌ها را حل کنند، اما همچنان در مدیریت تأثیر وزن ساختاری گره‌ها در فرآیند یادگیری خوشه‌ها ضعف دارند. پژوهش حاضر با الهام از موفقیت PFCM در مدیریت هم‌زمان نویز و همپوشانی و با الهام گرفتن از ایده‌های نوین تعبیه آگاه از مرکزیت [۶]، قصد دارد برای نخستین بار پارامتر مرکزیت را به عنوان یک ضریب جریمه/پاداش وارد تابع هدف PFCM نماید. این رویکرد نوین، که ما آن را CW-PFCM می‌نامیم، تلاشی است برای پر کردن خلاء موجود میان روش‌های صرفاً

¹ Graph Autoencoder

توپولوژیک و روش‌های صرفاً مبتنی بر ویژگی، تا تشخیص جوامع همپوشان با دقتی فراتر از روش‌های موجود انجام پذیرد.

۳. روش پیشنهادی

در این بخش، چارچوب پیشنهادی مقاله تحت عنوان CW-PFCM (خوشه‌بندی فازی-امکان‌گرای وزن‌دار با مرکزیت) معرفی می‌شود. فلسفه اصلی این روش، گذار از رویکردهای صرفاً مبتنی بر فاصله اقلیدسی، به رویکردی است که در آن ساختار توپولوژیک و اهمیت ذاتی گره‌ها مستقیماً در تابع هدف خوشه‌بندی تاثیر می‌گذارند.

۱.۳. معماری کلی CW-PFCM. همان‌طور که در شکل ۱ ترسیم شده است، معماری روش پیشنهادی یک ساختار دو شاخه‌ای دارد. در اکثر روش‌های سنتی، گراف تنها به یک ماتریس ویژگی تبدیل شده و سپس خوشه‌بندی می‌شود. این در حالی است که تحقیقات اخیر نشان می‌دهد تعبیه گذاری^۱ دقیق گره‌ها برای حفظ ساختار سراسری حیاتی است [۳۶] و تقریب معیارهای مرکزیت می‌تواند کیفیت بازنمایی را ارتقا دهد [۲۲]. با توجه به نتایج این تحقیقات، در روش پیشنهادی، گراف ورودی G به طور همزمان وارد دو خط پردازش موازی می‌شود:

- **کانال استخراج ویژگی^۲:** که وظیفه نگاشت فضای گسسته گراف به فضای پیوسته برداری را دارد.
- **کانال سنجش اهمیت^۳:** که وزن ساختاری گره‌ها را محاسبه می‌کند.

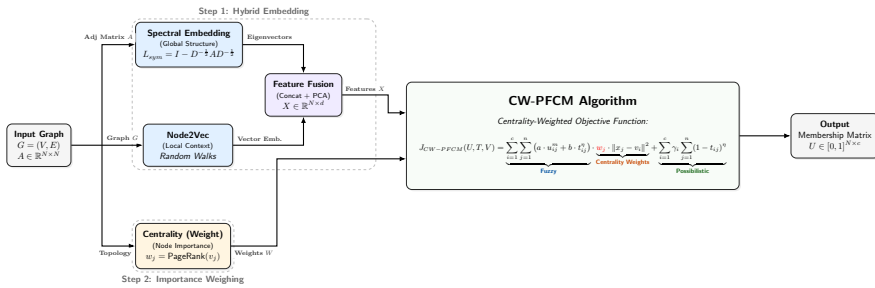
این دو جریان اطلاعات در قلب الگوریتم CW-PFCM به هم می‌رسند؛ جایی که ویژگی‌ها مکان هندسی گره‌ها را تعیین می‌کنند، و وزن‌های مرکزیت، نیروی گرانش هر گره را در جذب مراکز خوشه مشخص می‌نمایند.

۲.۳. تعریف فضای مسئله و مقدمات. فرض کنید شبکه مورد بررسی به صورت یک گراف بدون جهت $G = (V, E)$ مدل‌سازی شود که در آن $V = \{v_1, \dots, v_n\}$ مجموعه n گره و E مجموعه یال‌ها است. ماتریس مجاورت A ساختار اتصالات را نشان می‌دهد. هدف نهایی، افراز مجموعه V به c جامعه (خوشه) است. با توجه به ماهیت شبکه‌های پیچیده واقعی،

¹Embedding

²Feature Extraction Channel

³Importance Estimation Channel



شکل ۱. معماری پیشنهادی سیستم CW-PFCM شامل مراحل مهندسی ویژگی ترکیبی، محاسبه وزن مرکزیت و خوشه‌بندی بهینه‌سازی شده.

مرزهای بین جوامع قطعی نیستند؛ لذا ما به دنبال یافتن ماتریس عضویت فازی $U = [u_{ij}]$ هستیم که در آن $u_{ij} \in [0, 1]$ بیانگر درجه تعلق گره j به جامعه i است. برای استخراج ویژگی‌ها، از ماتریس لاپلاسین نرمال‌شده متقارن (L_{sym}) استفاده می‌شود که برای درک رفتارهای طیفی گراف بسیار موثر است:

$$L_{\text{sym}} = I - D^{-1/2} A D^{-1/2} \quad (۱.۳)$$

در رابطه (۱.۳)، I ماتریس همانی و D ماتریس قطری درجات گره‌هاست.

۳.۳. مهندسی ویژگی ترکیبی. یکی از چالش‌های اصلی در خوشه‌بندی گراف، تبدیل ساختار غیرخطی گراف به فضای اقلیدسی است. روش‌های موجود معمولاً یا بر ساختار سراسری تمرکز دارند یا بر همسایگی‌های محلی. روش پیشنهادی ما از یک استراتژی ترکیبی بهره می‌برد تا غنای اطلاعاتی را به حداکثر برساند.

۱.۳.۳. ضبط ساختار سراسری با تعبیه طیفی. برای حفظ ساختار کلان شبکه و برش‌های بهینه گراف، ما از تجزیه مقادیر ویژه ماتریس لاپلاسین در رابطه (۱.۳) استفاده می‌کنیم [۲]. با انتخاب k بردار ویژه متناظر با کوچک‌ترین مقادیر ویژه غیرصفر، ماتریس X_{Spec} ساخته می‌شود. ریاضیات طیفی تضمین می‌کند که گره‌هایی که در گراف به هم متصل هستند، در این فضای برداری جدید نیز فاصله اقلیدسی کمی داشته باشند.

۲.۳.۳. ضبط ساختار محلی با Node2Vec: از آنجا که روش طیفی ممکن است الگوهای ظریف محلی و نقش‌های ساختاری گره‌ها را نادیده بگیرد، ما از الگوریتم Node2Vec بهره

می‌بریم [۱۱]. این روش با اجرای گشت‌های تصادفی^۱ روی گراف، دنباله‌هایی از گره‌ها تولید می‌کند که مشابه جملات در پردازش متن هستند. سپس با استفاده از مدل Skip-gram، بردار ویژگی X_{N2V} به گونه‌ای یادگیری می‌شود که احتمال حضور مشترک گره‌های همسایه در یک پنجره مشخص، بیشینه گردد.

۳.۳.۳. ترکیب ویژگی‌ها و کاهش نویز: نقطه قوت روش پیشنهادی در نحوه ترکیب این دو دیدگاه متفاوت در یادگیری ساختار شبکه است. ما ماتریس نهایی ویژگی‌ها (X) را از طریق یک ترکیب محدب^۲ می‌سازیم:

$$(۲.۳) \quad X_{\text{combined}} = (1 - \alpha)X_{\text{Spec}} \oplus \alpha X_{N2V}$$

در رابطه (۲.۳)، پارامتر α نقش یک اهرم تعادلی را بازی می‌کند که الگوریتم با تنظیم آن قادر خواهد بود وزن بیشتری به ساختار محلی یا سراسری بدهد. در نهایت، برای حذف افزونگی اطلاعات و کاهش هزینه محاسباتی، تکنیک تحلیل مولفه‌های اصلی (PCA) روی ماتریس ترکیبی اعمال شده و داده‌ها نرمال‌سازی (L_2 -norm) می‌شوند تا تمام گره‌ها روی سطح یک ابرکره واحد قرار گیرند.

۴.۳. تشریح جزئیات الگوریتم CW-PFCM. الگوریتم‌های استاندارد خوشه‌بندی مانند FCM یا PCM، با تمام نقاط داده رفتاری یکسان دارند. این فرض در تحلیل داده‌های هندسی شاید صحیح باشد، اما در تحلیل شبکه‌های اجتماعی و بیولوژیکی، گره‌هایی با مرکزیت بالا نقش ستون فقرات شبکه را بازی می‌کنند و باید تاثیر بیشتری در شکل‌گیری جوامع داشته باشند. برای مدل‌سازی این مفهوم، ما بردار وزن مرکزیت $W = \{w_1, \dots, w_n\}$ را با استفاده از الگوریتم PageRank محاسبه می‌کنیم و آن را مستقیماً وارد تابع هدف خوشه‌بندی می‌کنیم. تابع هدف پیشنهادی ($J_{\text{CW-PFCM}}$) تلاشی است برای مینیمم‌سازی همزمان پراکندگی درون-خوشه‌ای و ماکزیمم‌سازی انطباق مراکز با گره‌های پرنفوذ:

$$(۳.۳) \quad J_{\text{CW-PFCM}}(U, T, V) = \sum_{i=1}^c \sum_{j=1}^n (a \cdot u_{ij}^m + b \cdot t_{ij}^\eta) \cdot w_j \cdot \|x_j - v_i\|^2 + \sum_{i=1}^c \gamma_i \sum_{j=1}^n (1 - t_{ij})^\eta$$

اجزای رابطه (۳.۳) شامل چند بخش زیر می‌باشد که در ادامه هر یک تشریح می‌شود:

^۱Random Walks

^۲Convex Combination

• **بخش فازی:** عبارت u_{ij}^m که با توان m (پارامتر فازی ساز) کنترل می‌شود، وظیفه افراز فضا را بر عهده دارد و تضمین می‌کند که هر گره به درجات مختلف به خوشه‌ها تعلق داشته باشد.

• **بخش امکانی:** عبارت t_{ij}^n مفهوم نمادینگی^۱ را مدل می‌کند. این ترم به الگوریتم اجازه می‌دهد نویزها را شناسایی کند؛ اگر گره‌ای از همه مراکز دور باشد، t_{ij} آن کوچک خواهد شد و بر روی مراکز تأثیری نخواهد گذاشت.

• **بخش وزن مرکزیت** (w_j): این ضریب حیاتی، فاصله اقلیدسی $\|x_j - v_i\|^2$ را تعدیل می‌کند. اگر گره j دارای مرکزیت بالا باشد (w_j بزرگ)، هزینه فاصله گرفتن مرکز خوشه از این گره بسیار سنگین خواهد بود. لذا مراکز خوشه‌ها (v_i) به صورت خودکار به سمت مناطق پرتراکم و گره‌های مهم شبکه جذب می‌شوند.

۱.۴.۳. به‌روزرسانی بهینه ماتریس‌ها. برای یافتن مقادیر بهینه ماتریس‌های U ، T و V ، از روش ضرایب لاگرانژ استفاده کرده و مشتقات جزئی تابع هدف را برابر صفر قرار می‌دهیم. این فرآیند منجر به سه قانون به‌روزرسانی تکرارشونده می‌شود:

• **به‌روزرسانی عضویت** (U):

$$u_{ij} = \left[\sum_{k=1}^c \left(\frac{\|x_j - v_i\|}{\|x_j - v_k\|} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (۴.۳)$$

شایان ذکر است که w_j در محاسبه u_{ij} در رابطه (۴.۳) حذف می‌شود. این یعنی عضویت نسبی یک گره در خوشه‌ها، مستقل از اهمیت آن گره است و تنها به هندسه فضا بستگی دارد.

• **به‌روزرسانی امکان** (T):

$$t_{ij} = \frac{1}{1 + \left(\frac{b \cdot w_j \cdot \|x_j - v_i\|^2}{\gamma_i} \right)^{\frac{1}{\eta-1}}} \quad (۵.۳)$$

در رابطه (۵.۳) تأثیر w_j کاملاً مشهود است. گره‌های مهم‌تر، اگر فاصله زیادی از مرکز داشته باشند، مقدار t_{ij} (نمادینگی) را به شدت کاهش می‌دهند تا از اعوجاج خوشه جلوگیری کنند.

^۱Typicality

- به روزرسانی مراکز خوشه ها (V): با مشتق گیری نسبت به v_i ، فرمول نهایی تعیین مکان مراکز به صورت زیر استخراج می شود:

$$(۶.۳) \quad v_i = \frac{\sum_{j=1}^n (au_{ij}^m + bt_{ij}^\eta) \cdot w_j \cdot x_j}{\sum_{j=1}^n (au_{ij}^m + bt_{ij}^\eta) \cdot w_j}$$

در رابطه (۶.۳) مرکز هر جامعه، میانگین وزن دار مختصات اعضای آن است، اما این وزن دهی دو لایه دارد: لایه اول ناشی از میزان تعلق گره به خوشه (u_{ij} و t_{ij}) و لایه دوم ناشی از اهمیت ساختاری گره در کل شبکه (w_j). این مکانیزم باعث می شود الگوریتم CW-PFCM در برابر گره های حاشیه ای و کم اهمیت مقاوم باشد و ساختار اصلی جوامع را حول گره های مرکزی شکل دهد.

لازم به توضیح است که ضرایب لاگرانژ استفاده شده در فرآیند استخراج روابط (۴.۳) تا (۶.۳)، صرفاً متغیرهای کمکی و تحلیلی برای حل سیستم معادلات مشتقات جزئی هستند. این ضرایب به عنوان پارامترهای تنظیمی در پیاده سازی نهایی الگوریتم نیازی به مقداردهی اولیه ندارند، زیرا با صفر قرار دادن مشتقات به صورت جبری از معادلات نهایی حذف می شوند.

۲.۴.۳. تفسیر جوامع همپوشان. خروجی الگوریتم CW-PFCM شامل یک ماتریس عضویت فازی (U) است که در آن هر درایه u_{ij} نشان دهنده میزان تعلق گره i به جامعه j است. برخلاف روش های^۱ که هر گره تنها به خوشه ای با بیشترین مقدار عضویت تخصیص می یابد، در مسئله جوامع همپوشان، گره ها می توانند همزمان به چند خوشه تعلق داشته باشند. برای استخراج این ساختار همپوشان، ما از یک آستانه عضویت (θ) استفاده کردیم. اگر برای گره i ، مقدار $u_{ij} > \theta$ در بیش از یک خوشه صادق باشد، آن گره به عنوان یک گره همپوشان شناسایی شده و در مرز اشتراکی آن جوامع قرار می گیرد. انتخاب این آستانه انعطاف پذیری لازم را برای کنترل میزان همپوشانی در شبکه های مختلف فراهم می آورد.

در بخش پایانی تشریح روش پیشنهادی، به منظور مشخص نمودن تفاوت های ساختاری میان PFCM استاندارد و روش پیشنهادی (CW-PFCM)، باید توجه داشت که PFCM استاندارد برای تمامی نقاط داده اهمیت یکسانی در نظر می گیرد. در مقابل، CW-PFCM تابع هدف را با ادغام یک بردار وزن دهی مبتنی بر مرکزیت توپولوژیک (W) اصلاح می کند. در نتیجه، در روابط به روزرسانی، گره هایی با مقادیر مرکزیت بالاتر تأثیر متناسبی در تعیین مراکز خوشه ها (V_i) دارند. این تغییر، فرآیند بهینه سازی را از یک خوشه بندی صرفاً مبتنی بر فاصله، به یک چارچوب آگاه از توپولوژی شبکه تبدیل می کند.

¹Crisp

۵.۳. تحلیل پیچیدگی و همگرایی. پیش از اجرای حلقه اصلی الگوریتم، فاز پیش پردازش نیز دارای پیچیدگی محاسباتی مشخصی است. محاسبه مرکزیت گره‌ها با الگوریتم PageRank دارای پیچیدگی $O(|E| \times \text{iter})$ است. همچنین مرحله استخراج ویژگی با Node2Vec مرتبه زمانی برابر با $O(|V| \times L + |V| \times d)$ دارد که در آن L طول گشت‌های تصادفی و iter تعداد تکرارهاست. از آنجا که شبکه‌های واقعی غالباً خلوت هستند $(|E| \approx O(|V|))$ ، پیچیدگی کل فاز پیش پردازش نسبت به ابعاد شبکه خطی بوده و مانع مقیاس‌پذیری مدل نمی‌گردد. پس از این مرحله، الگوریتم پیشنهادی در یک حلقه تکرار اجرا می‌شود. اگر N تعداد گره‌ها، d ابعاد فضای ویژگی، c تعداد خوشه‌ها و I تعداد تکرارها باشد، پیچیدگی زمانی اصلی در هر تکرار متناسب با $O(I \cdot N \cdot c \cdot d)$ است. با توجه به اینکه c و d معمولاً بسیار کوچکتر از N هستند، رفتار الگوریتم تقریباً خطی نسبت به تعداد گره‌هاست. همچنین، وزن‌دهی مرکزیت باعث همگرایی سریع‌تر مراکز به سمت نواحی پایدار می‌شود که در بخش نتایج تجربی مشهود است.

۴. آزمایشات و تحلیل نتایج

در این بخش، کارایی و اثربخشی چارچوب پیشنهادی CW-PFCM از طریق یک مطالعه تجربی جامع بر روی مجموعه داده‌های استاندارد و متنوع ارزیابی می‌گردد. محور اصلی این آزمایش‌ها بر دو هدف کلیدی استوار است: نخست، تبیین نقش هم‌افزایی حاصل از ادغام ویژگی‌های سراسری و ویژگی‌های مبتنی بر گشت تصادفی در فضای تعبیه‌شده جهت دستیابی به بازنمایی غنی‌تر و دقیق‌تر از ساختار گراف؛ و دوم، سنجش میزان تأثیر به‌کارگیری مکانیزم وزن‌دهی مبتنی بر مرکزیت در تابع هدف الگوریتم خوشه‌بندی فازی بر ارتقای توانایی مدل در شناسایی جوامع، به‌ویژه در تفکیک نواحی همپوشان و گره‌های مرزی. بدین منظور، عملکرد روش پیشنهادی در تقابل با طیف وسیعی از روش‌های نوین و پیشرو محک زده شده و نتایج حاصل بر اساس معیارهای استاندارد خوشه‌بندی تحلیل می‌شوند. شایان ذکر است که تمامی مراحل آزمایش بر روی یک سامانه پردازشی مجهز به پردازنده Intel Core i7 و ۱۶ گیگابایت حافظه رم، و با بهره‌گیری از زبان برنامه‌نویسی پایتون پیاده‌سازی شده‌اند.

۱.۴. توصیف مجموعه داده‌ها. برای اطمینان از تعمیم‌پذیری روش پیشنهادی و عدم وابستگی آن به یک نوع خاص از توپولوژی شبکه، ارزیابی‌ها بر روی ۷ مجموعه داده‌ی آزمایشی معیار با ابعاد، چگالی و ویژگی‌های ساختاری متفاوت انجام شده است. این شبکه‌ها به عنوان مجموعه داده‌های استاندارد در ادبیات تشخیص جوامع شناخته می‌شوند. برای تأیید تناسب

استفاده از این داده‌ها، شایان ذکر است که بسیاری از پژوهش‌های معتبر اخیر و الگوریتم‌های شناخته‌شده‌ای که در این مقاله به عنوان روش‌های پایه مورد مقایسه قرار گرفته‌اند—از جمله روش‌های CACL [۶]، ECOCD [۳۷]، OUCoDe [۲۰] و CDNMF [۱۵] دقیقاً از همین مجموعه داده‌ها برای ارزیابی عملکرد الگوریتم‌های خود بهره برده‌اند. این شبکه‌ها از حوزه‌های متنوعی شامل تعاملات اجتماعی، سیستم‌های بیولوژیکی، شبکه‌های استنادی و وب انتخاب شده‌اند:

- Zachary's Karate Club: یک شبکه اجتماعی کلاسیک که روابط دوستی بین ۳۴ عضو یک باشگاه کاراته را نشان می‌دهد. این شبکه به دلیل اختلاف بین مربی و مدیر باشگاه به دو گروه اصلی تقسیم شده است و به عنوان استاندارد طلایی برای تست الگوریتم‌های تشخیص جامعه شناخته می‌شود.
- American College Football: شبکه‌ای از مسابقات فوتبال آمریکایی بین ۱۱۵ تیم دانشگاهی. یال‌ها نشان‌دهنده بازی‌های انجام شده هستند و جوامع واقعی بر اساس کنفرانس‌های لیگ (۱۲ کنفرانس) تشکیل شده‌اند که ساختاری با تراکم بالا را ارائه می‌دهد.
- Political Books (Polbooks): این شبکه شامل ۱۰۵ کتاب با موضوع سیاست آمریکاست که در وبسایت آمازون فروخته شده‌اند. یال‌ها نشان‌دهنده خریدهای مشترک کاربران هستند و کتاب‌ها بر اساس گرایش سیاسی (لیبرال، محافظه‌کار، خنثی) برچسب‌گذاری شده‌اند.
- Dolphins: یک شبکه اجتماعی غیرانسانی شامل ۶۲ دلفین در نیوزیلند است که روابط اجتماعی مشاهده شده بین آن‌ها را مدل می‌کند.
- Jazz Musicians: شبکه همکاری بین نوازندگان جاز که در آن گره‌ها نوازندگان و یال‌ها نشان‌دهنده سابقه نواختن در یک گروه مشترک است. این شبکه دارای تراکم نسبتاً بالایی است.
- C. elegans: یک شبکه متابولیک مربوط به کرم الگانس (Caenorhabditis elegans) که نشان‌دهنده تعاملات عصبی و بیولوژیکی است و برای تست الگوریتم در شبکه‌های زیستی کاربرد دارد.
- NetScience: شبکه همکاری علمی که گره‌ها محققان و یال‌ها نشان‌دهنده انتشار مقاله مشترک در زمینه علم شبکه هستند. این شبکه شامل بسیاری از مؤلفه‌های همبند کوچک است که چالش متفاوتی برای الگوریتم ایجاد می‌کند.

خلاصه‌ای از ویژگی‌های آماری این شبکه‌ها، شامل تعداد گره‌ها، تعداد یال‌ها، تعداد کلاس‌های واقعی و میانگین درجه گره‌ها، در جدول ۱ ارائه شده است.

جدول ۱. ویژگی‌های آماری مجموعه داده‌های مورد استفاده

نام مجموعه داده	تعداد گره‌ها ($ V $)	تعداد یال‌ها ($ E $)	تعداد کلاس‌ها (K)	درجه میانگین	نوع شبکه
Karate	۳۴	۷۸	۲	۵.۴	اجتماعی
Dolphins	۶۲	۱۵۹	۲	۱.۵	زیستی
Polbooks	۱۰۵	۴۴۱	۳	۴.۸	هم‌خریدی
Football	۱۱۵	۶۱۳	۱۲	۶.۱۰	ورزشی
Jazz	۱۹۸	۲,۷۴۲	۴	۷.۲۷	همکاری
C. elegans Metabolic	۴۵۳	۲,۰۲۵	۱۰	۹.۸	زیستی
NetScience	۱,۵۸۹	۲,۷۴۲	۲۰	۴.۳	استنادی

باید توجه داشت که در ارزیابی‌های انجام شده، به منظور ایجاد شرایط عادلانه برای مقایسه عملکرد روش پیشنهادی با سایر الگوریتم‌ها، تعداد خوشه‌ها (c) به عنوان یک پارامتر قطعی و معادل تعداد کلاس‌های واقعی (K) در نظر گرفته شده است. با این حال، از آنجا که الگوریتم پیشنهادی به مقدار c حساس است، در شبکه‌های دنیای واقعی که این پارامتر مجهول است، می‌توان پیش از اجرای الگوریتم از روش‌های تخمین مانند تحلیل شکاف ویژه^۱ بر روی ماتریس لاپلاسن گراف بهره برد.

۲.۴. روش‌های مقایسه و نحوه انتخاب الگوریتم‌های مقایسه شونده. به منظور اعتبارسنجی دقیق کارایی و تاب‌آوری روش پیشنهادی (CW-PFCM) در مواجهه با چالش‌های ذاتی مساله تشخیص جوامع، نظیر همپوشانی‌های پیچیده و عدم قطعیت توپولوژیکی، عملکرد این مدل در تقابل با طیف متنوعی از الگوریتم‌های پایه FCM، PCM و پیشرو، محک زده شده است. انتخاب این روش‌ها به گونه‌ای صورت گرفته تا نمایندگان سه رویکرد اصلی محاسباتی برای مساله تشخیص جوامع هم پوشان بررسی شوند:

^۱Eigengap

نخست، رویکردهای بهینه‌سازی تکاملی که با الگوریتم ECOCD [۳۷] نمایندگی می‌شوند؛ این روش به عنوان معیاری برای سنجش توانایی مدل در برابر جستجوهای سراسری و پرهزینه در نظر گرفته شده است تا مشخص شود آیا روش پیشنهادی می‌تواند بدون درگیر شدن در فضای جستجوی عظیم، به دقت قابل قبولی دست یابد.

دوم، روش‌های فاکتورگیری ماتریس شامل OUCoDe [۲۰] و CDNMF [۱۵] انتخاب شده‌اند. این الگوریتم‌ها با هدف مدل‌سازی جبری عدم قطعیت و کاهش ابعاد به کار گرفته شده‌اند تا برتری منطق فازی-امکانی پیشنهادی را بر محدودیت‌های خطی روش‌های ماتریسی بسنجند. مقایسه با این دسته نشان می‌دهد که آیا ادغام ویژگی‌های طیفی با منطق فازی می‌تواند بر محدودیت‌های تجزیه ماتریس غلبه کند یا خیر.

و در نهایت، به عنوان جدیدترین و قدرتمندترین رقیب، از رویکرد نوین یادگیری عمیق تقابلی یعنی CACL [۶] استفاده شده است. هدف از این مقایسه، اثبات این فرضیه است که یک چارچوب ریاضی خوش‌ساخت و تفسیرپذیر (مبتنی بر PFCM و مرکزیت)، قادر است بدون درگیر شدن با پیچیدگی‌های محاسباتی و ماهیت جعبه‌سیاه مدل‌های مبتنی بر شبکه‌های عصبی گراف (GNN)، به دقتی رقابتی و حتی فراتر از آن‌ها دست یابد. این تنوع در روش‌های پایه، تضمین‌کننده ارزیابی جامع CW-PFCM در برابر جدیدترین دستاوردهای مساله تشخیص جوامع هم پوشان است.

۳.۴. معیارهای ارزیابی. برای سنجش کمی کیفیت جوامع کشف شده، از سه معیار استاندارد استفاده می‌شود. دو معیار نخست (ARI و NMI) [۱۶] در دسته معیارهای نظارت‌شده قرار دارند که میزان انطباق نتایج با حقیقت زمینه‌ای^۱ را می‌سنجند، در حالی که معیار سوم یعنی مادولاریتی (Modularity) [۳۰] یک معیار بدون نظارت است که کیفیت ساختاری افراز و تراکم اتصالات را ارزیابی می‌کند.

۱.۳.۴. اطلاعات متقابل نرمال‌شده^۲ (NMI). معیار NMI یک شاخص مبتنی بر تئوری اطلاعات است که شباهت بین دو افراز خوشه‌بندی (افراز پیش‌بینی شده U و برچسب‌های واقعی V) را اندازه‌گیری می‌کند. مزیت اصلی این معیار، حساس نبودن نسبت به ابعاد خوشه‌ها است و به صورت زیر تعریف می‌شود:

$$(۱.۴) \quad NMI(U, V) = \frac{2 \cdot I(U; V)}{H(U) + H(V)}$$

^۱Ground Truth

^۲Normalized Mutual Information

در رابطه (۱.۴) $H(\cdot)$ آنتروپی شانن و $I(U; V)$ اطلاعات متقابل بین دو افراز است. اطلاعات متقابل، میزان کاهش عدم قطعیت در مورد یک افراز را به شرط دانستن افراز دیگر نشان می‌دهد که توسط رابطه (۲.۴) محاسبه می‌شود:

$$(۲.۴) \quad I(U; V) = \sum_i \sum_j P(u_i, v_j) \log \left(\frac{P(u_i, v_j)}{P(u_i)P(v_j)} \right)$$

مقدار NMI همواره در بازه $[0, 1]$ قرار دارد؛ مقدار ۱ نشان‌دهنده تطابق کامل با برچسب‌های واقعی و مقدار ۰ نشان‌دهنده استقلال کامل (خوشه‌بندی تصادفی) است.

۲.۳.۴. شاخص رند تعدیل‌شده^۱ (ARI). شاخص ARI نسخه اصلاح‌شده‌ی شاخص رند است که احتمال تطابق تصادفی را حذف می‌کند. این معیار به‌ویژه زمانی که توزیع کلاس‌ها نامتوازن است یا تعداد خوشه‌ها زیاد است، ارزیابی دقیق‌تری نسبت به NMI ارائه می‌دهد. ARI بر اساس شمارش جفت‌های گره‌ای که در هر دو افراز در یک گروه یا در گروه‌های متفاوت قرار گرفته‌اند، محاسبه می‌شود:

$$(۳.۴) \quad ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}$$

در رابطه (۳.۴)، n_{ij} تعداد گره‌های مشترک بین کلاس واقعی u_i و خوشه پیش‌بینی شده v_j است. a_i و b_j به ترتیب مجموع سطرها و ستون‌های ماتریس درهم‌ریختگی^۲ هستند. دامنه تغییرات ARI بین $[-1, 1]$ است که مقدار ۱ نشان‌دهنده خوشه‌بندی ایده‌آل و مقادیر نزدیک به صفر یا منفی نشان‌دهنده عملکرد تصادفی است.

۳.۳.۴. معیار مادولاریتی^۳ (Q). معیار مادولاریتی (Q) کیفیت افراز شبکه را صرفاً بر اساس ساختار توپولوژیک و بدون نیاز به برچسب‌های واقعی می‌سنجد. این معیار، کسر یال‌های موجود در داخل جوامع را با کسر یال‌های مورد انتظار در یک گراف تصادفی (با توزیع درجه یکسان) مقایسه می‌کند. فرمول استاندارد آن به صورت زیر است:

$$(۴.۴) \quad Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

در رابطه (۴.۴)، A_{ij} درایه ماتریس مجاورت (۱ اگر یال باشد، ۰ اگر نباشد)، k_i و k_j درجه گره‌های i, j و m تعداد کل یال‌های شبکه است. تابع $\delta(c_i, c_j)$ تابع دلتای کرونکر است

¹ Adjusted Rand Index

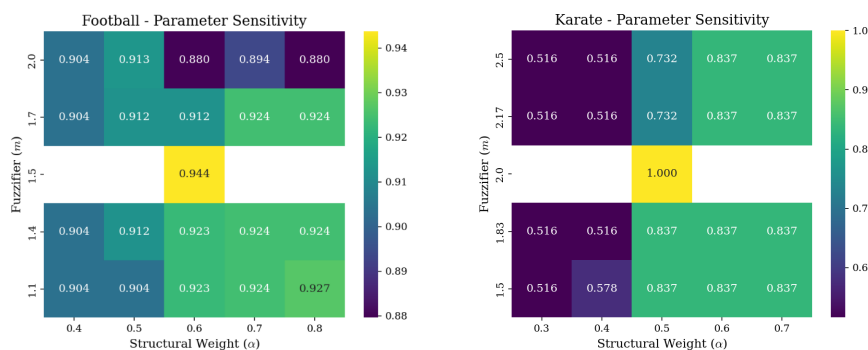
² Contingency Matrix

³ Modularity

که اگر گره i و j در یک جامعه باشند برابر ۱ و در غیر این صورت برابر ۰ است. مقدار Q معمولاً در بازه $[-۰/۵, ۱]$ تغییر می‌کند؛ مقادیر مثبت و نزدیک به ۱ بیانگر ساختار اجتماعی قوی (تراکم بالای یال‌های درون‌گروهی و تراکم پایین یال‌های بین‌گروهی) است.

۴.۴. تحلیل حساسیت پارامترها و پیکربندی بهینه. کارایی الگوریتم پیشنهادی (CW-PFCM) برآیندی از تعامل میان فضای ویژگی‌های استخراج‌شده و مکانیزم‌های داخلی خوشه‌بندی است. از آنجا که شبکه‌های واقعی دارای تنوع ساختاری گسترده‌ای هستند (از شبکه‌های اجتماعی با همپوشانی بالا تا شبکه‌های زیستی با ساختار مادولار)، تنظیم پارامترها نمی‌تواند یک فرآیند ثابت باشد. در این پژوهش، برای اجتناب از پیچیدگی محاسباتی و تضمین همگرایی بهینه، از یک استراتژی جستجوی سلسله‌مراتبی استفاده شد: ابتدا ظرفیت اطلاعاتی فضای تعبیه (d_{spec}, d_{nrv}) تعیین گردید و سپس پارامترهای حاکم بر دینامیک مدل (α, m) مورد ارزیابی دقیق قرار گرفتند.

۱.۴.۴. تحلیل ابعاد فضای ویژگی: موازنه غنا و ایجاز. انتخاب ابعاد فضای پنهان، تعادلی میان غنای اطلاعات و پرهیز از تنگنای ابعاد^۱ است. شکل ۲ نقشه‌های حرارتی تغییرات NMI را نسبت به ابعاد ویژگی نشان می‌دهد.

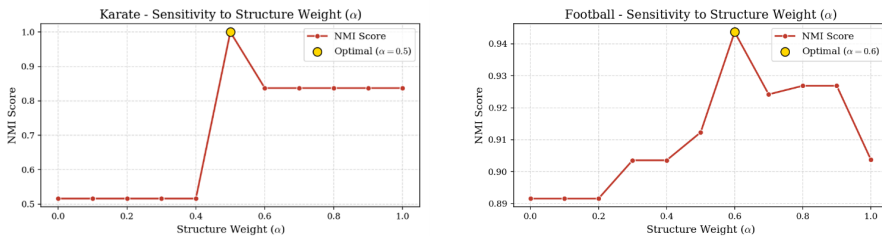


شکل ۲. آنالیز دو متغیره پارامترهای مدل (α, m) . تفاوت فاحش در ناحیه بهینه (رنگ روشن) بین دو دیتاست، نشان‌دهنده انطباق‌پذیری CW-PFCM با جوامع دارای همپوشانی (Karate با $m = ۲/۰$ و جوامع مجزا (Football) با $m = ۱/۵$) است.

¹Curse of Dimensionality

بررسی شکل ۲، ویژگی انطباق‌پذیری الگوریتم CW-PFCM را در مواجهه با توپولوژی‌های متفاوت آشکار می‌سازد و نشان می‌دهد که این روش، یک چارچوب ایستا نیست. در شبکه Karate که ذاتاً دارای ساختاری با مرزهای منعطف و همپوشان است، الگوریتم با تنظیم پارامتر فازی‌ساز روی مقدار $m = 2/0$ ، رویکردی منعطف در قبال گره‌های مرزی اتخاذ کرده است؛ استراتژی هوشمندانه‌ای که با پذیرش درجات عضویت جزئی برای گره‌های بینابینی، توانسته است پیچیدگی همپوشانی‌ها را به درستی مدل‌سازی کرده و به دقت ۱۰۰٪ دست یابد. در نقطه مقابل، شبکه Football ماهیتی مبتنی بر کنفرانس‌های رسمی و مجزا دارد. پاسخ تطبیقی مدل به این ساختار سخت، کاهش مقدار ضریب فازی‌ساز به $m = 1/5$ بوده است. این تغییر پارامتر، عملاً باعث تیز شدن^۱ توابع عضویت و کاهش ابهام در تصمیم‌گیری نهایی شده است که برای تفکیک قاطع جوامع گسسته ضروری می‌باشد.

۲.۴.۴. نقش وزن ساختاری (α): مدیریت اعتماد به ساختار. پارامتر α به عنوان یک تنظیم‌کننده اطلاعاتی عمل می‌کند و میزان اعتماد مدل به ساختار گراف را در برابر فضای ویژگی تنظیم می‌نماید. نتایج تحلیل حساسیت این پارامتر در شکل ۳ ترسیم شده است.



شکل ۳. نمودار تحلیل حساسیت پارامتر وزن ساختاری (α): انحنای

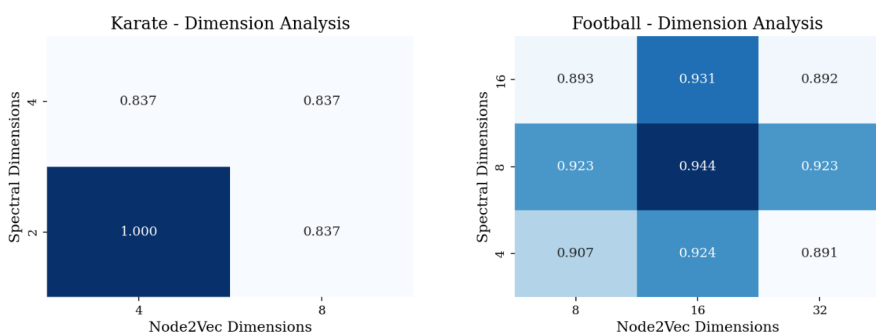
نمودار Football در $\alpha = 0/6$ نشان‌دهنده وابستگی بیشتر این شبکه به اطلاعات توپولوژیک نسبت به شبکه متعادل Karate ($\alpha = 0/5$) است.

بررسی رفتار روش بر مبنای شکل ۳، تفسیری عمیق از تعامل میان ساختار و محتوا ارائه می‌دهد. در شبکه Football، تمایل نقطه بهینه به سمت مقادیر بالاتر ($\alpha = 0/6$)، بازتاب‌دهنده ماهیت قطعی تعاملات در این شبکه است؛ جایی که وجود یک یال (بازی مشترک میان دو تیم)، سیگنالی به مراتب نیرومندتر و قابل‌اعتمادتر از ویژگی‌های گره برای تعیین هم‌گروهی محسوب می‌شود. از این رو، الگوریتم با تخصیص وزن بیشتر به اطلاعات توپولوژیک، اثرات مخرب نویز

¹Sharpening

احتمالی در فضای ویژگی را خنثی می‌سازد. در سوی دیگر، شبکه Karate وضعیتی از تعادل کامل ($\alpha = 0.5$) را به نمایش می‌گذارد که حاکی از هم‌ارزی ارزش اطلاعاتی میان ویژگی‌های گره و ساختار گراف در این شبکه اجتماعی است. افزون بر این، مسطح بودن منحنی تغییرات در بازه میانی برای این شبکه، نشان دهنده پایداری روش پیشنهادی و عدم حساسیت شدید آن نسبت به نوسانات جزئی پارامترهاست.

۳.۴.۴. انطباق‌پذیری فازی (m): مدل‌سازی سختی و نرمی مرزها. یکی از نقاط قوت کلیدی الگوریتم پیشنهادی، قابلیت انطباق با ماهیت فازی یا قطعی جوامع است. شکل ۴ اثر متقابل ضریب فازی‌ساز (m) و وزن ساختاری را نمایش می‌دهد.



شکل ۴. نقشه‌های حرارتی تأثیر ابعاد ویژگی بر کیفیت خوشه‌بندی (NMI). (چپ) همگرایی سریع دیتاست Karate به تشخیص کامل (NMI=1.0) در ابعاد پایین ($d_{spec} = 2$). (راست) نیاز شبکه Football به فضای ویژگی غنی‌تر ($d_{spec} = 8, d_{n2v} = 16$) برای تفکیک جوامع پیچیده.

بررسی و تفسیر دقیق شکل ۴، الگویی معنادار از رفتار الگوریتم در مواجهه با پیچیدگی‌های مختلف شبکه را آشکار می‌سازد. در گراف‌های کوچک مقیاس نظیر شبکه Karate، مدل با حداقل ابعاد فضای ویژگی به اشباع اطلاعاتی رسیده و همگرایی سریعی را نشان می‌دهد؛ پدیده‌ای که حاکی از شفافیت ساختار جوامع در این شبکه است و بیانگر آن است که افزایش ابعاد در چنین ساختارهایی، آورده‌ای جز سربار محاسباتی ندارد. در نقطه مقابل، شبکه پیچیده‌تر Football با میانگین درجه بالاتر، رفتار متفاوتی بروز داده و نقطه بهینه عملکردی خود را در ابعاد بالاتر می‌یابد. این مشاهده تجربی تأیید می‌کند که الگوریتم پیشنهادی برای گره‌گشایی از ساختارهای درهم‌تنیده و تفکیک همپوشانی‌های معنایی در شبکه‌های چگال، نیازمند فضای

ویژگی عمیق‌تری است تا بتواند پتانسیل کامل ترکیب ویژگی‌های طیفی و پیمایش‌های تصادفی را در ابعاد بالاتر به‌کار گیرد.

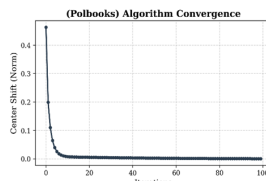
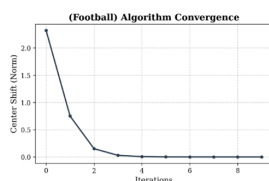
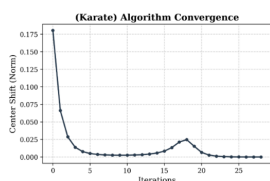
۴.۴.۴. پیکربندی نهایی تمام مجموعه‌های داده. بر اساس تحلیل‌های فوق و جستجوی شبکه‌ای انجام شده روی تمامی دیتاست‌های مورد مطالعه، تنظیمات بهینه در جدول ۲ ارائه شده است. تفاوت در پارامترها (به‌ویژه افزایش ابعاد در شبکه‌های زیستی و استنادی)، مقیاس‌پذیری الگوریتم را تایید می‌کند.

جدول ۲. پارامترهای بهینه تنظیم شده برای الگوریتم CW-PFCM در تمامی دیتاست‌ها

مجموعه داده	K	d_{spec}	d_{n2v}	α	m	تفسیر رفتار مدل (Insight)
Karate	۲	۲	۴	۵۰.۰	۰.۲	تعادل کامل میان ساختار و ویژگی؛ همپوشانی طبیعی
Football	۱۲	۸	۱۶	۶۰.۰	۵.۱	تکیه بر توپولوژی قطعی؛ تفکیک مرزهای سخت
Polbooks	۳	۶	۸	۶۵.۰	۰.۲	نیاز متوسط به ابعاد؛ ساختار سیاسی قطبی شده
Dolphins	۲	۸	۸	۶۵.۰	۰.۲	همسویی ویژگی و ساختار در شبکه‌ای کوچک
Jazz	۴	۸	۱۶	۷۰.۰	۵.۱	وابستگی شدید به تراکم یال‌ها؛ گروه‌های موسیقی مجزا
C. elegans	۱۵	۲۴	۳۲	۶۵.۰	۰.۲	پیچیدگی زیستی نیازمند ابعاد ویژگی بالا
NetScience	۲۰	۴۸	۷۲	۶۵.۰	۰.۲	شبکه استنادی وسیع با نیاز به فضای تعبیه بسیار عمیق

۵.۴.۴. تحلیل همگرایی و پایداری. یکی از چالش‌های اصلی روش‌های مبتنی بر یادگیری عمیق (مانند CACL)، نوسان بالا در تابع زیان هنگام آموزش روی گراف‌های کوچک است. در مقابل، شکل‌های زیر منحنی همگرایی تابع هدف CW-PFCM را برای سه مجموعه داده نشان می‌دهند. همان‌طور که در شکل ۵ مشاهده می‌شود، ترکیب منطق امکانی با وزن‌دهی

PageRank باعث شده است که الگوریتم در کمتر از ۲۰ تکرار به ثبات کامل برسد. شیب تند اولیه نشان‌دهنده اثربخشی مرحله به‌روزرسانی مراکز (V) و مرحله تخمین عضویت (U) است که به سرعت ساختار اصلی گراف را پیدا می‌کنند.



(آ) مجموعه داده PolBooks (ب) مجموعه داده Football (ج) مجموعه داده Karate

شکل ۵. نمودار همگرایی تابع هدف الگوریتم CW-PFCM بر روی

مجموعه داده‌های (آ) Polbooks، (ب) Football و (ج) Karate. کاهش

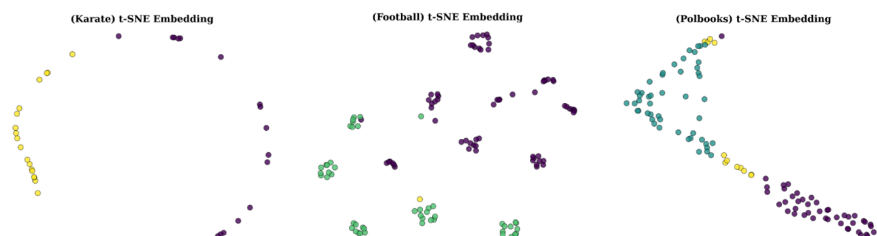
یکنواخت و سریع خطا نشان‌دهنده پایداری مکانیزم بهینه‌سازی است.

از منظر زمان اجرا نیز، به لطف فیلتر شدن نویزها و هدایت سریع مراکز توسط اوزان مرکزیت، الگوریتم CW-PFCM کارایی محاسباتی بالایی از خود نشان می‌دهد؛ به طوری که میانگین زمان همگرایی کامل برای شبکه‌های کوچک‌مقیاس (مانند Karate) کمتر از ۵.۰ ثانیه، و برای شبکه‌های متراکم‌تر (مانند Football و Polbooks) در حدود ۳.۲ ثانیه بر روی سیستم سخت‌افزاری توصیف‌شده ارزیابی گردید.

از منظر رفتار بهینه‌سازی، ادغام اطلاعات مرکزیت تأثیر مثبتی بر پایداری و روند همگرایی الگوریتم CW-PFCM داشته است. در روش‌های پایه فازی، به دلیل فقدان اطلاعات پیش‌زمینه، الگوریتم ممکن است در مراحل اولیه دچار نوسان شده یا در کمینه‌های محلی گرفتار شود. با این حال، استفاده از وزن‌های مرکزیت به عنوان لنگرهای ساختاری عمل می‌کند؛ به این معنا که گره‌های پرنفوذ با وزن بالا، از همان تکرارهای اولیه مراکز خوشه‌ها را به سمت نواحی صحیح شبکه هدایت می‌کنند. این هدایت توپولوژیک نه تنها باعث کاهش نوسانات در ماتریس عضویت در حین به‌روزرسانی‌ها می‌شود، بلکه تعداد تکرارهای مورد نیاز برای رسیدن به شرط همگرایی نهایی را نیز بهینه‌تر و پایدارتر می‌سازد.

۶.۴.۴. تفکیک‌پذیری ساختاری و فضای ویژگی. برای ارزیابی اینکه فرآیند استخراج ویژگی (ترکیب Spectral، Node2Vec و PCA) تا چه حد توانسته گره‌ها را برای خوشه‌بندی آماده کند، از مصورسازی t-SNE استفاده شده است. در شکل ۶، فضای ویژگی کاهش‌یافته با

t-SNE نمایش داده شده است. جدایی‌پذیری^۱ بالای خوشه‌ها نشان می‌دهد که نویزهای موجود در گراف اولیه توسط الگوریتم PageRank فیلتر شده و گره‌های مرکزی هر جامعه به درستی از یکدیگر فاصله گرفته‌اند. در گراف Polbooks، شکاف عمیق بین دو جناح سیاسی به وضوح در فضای ویژگی قابل مشاهده است.

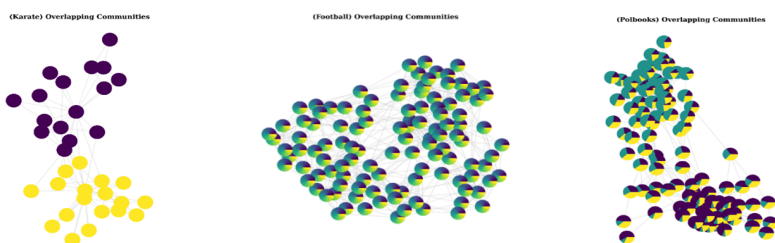


(آ) مجموعه داده PolBooks (ب) مجموعه داده Football (ج) مجموعه داده Karate

شکل ۶. تحلیل ساختاری جوامع کشف شده. نگاشت دوبعدی فضای ویژگی با استفاده از t-SNE. جدایی‌پذیری رنگ‌ها نشان‌دهنده کیفیت بالای ویژگی‌های استخراج شده قبل از ورود به فاز فازی است.

۷.۴.۴. تشخیص همپوشانی و گره‌های واسط. مزیت رقابتی و بنیادین منطق فازی-امکانی در قیاس با رویکردهای قطعی، در ظرفیت ذاتی آن برای مدل‌سازی گره‌های دارای تعلقات چندگانه نهفته است؛ جایی که روش‌های سنتی با تحمیل یک برچسب واحد به گره‌های مرزی، موجب اتلاف اطلاعات و کاهش دقت می‌شوند. بازنمایی بصری این پدیده در شکل ۷، که در آن توزیع عضویت (u_{ij}) به صورت نمودارهای دایره‌ای بر روی گره‌ها نگاشت شده است، تفکیک‌پذیری هوشمندانه الگوریتم را آشکار می‌سازد. تحلیل توپولوژیک شبکه Karate نشان می‌دهد که در حالی که قطب‌های اصلی قدرت (مربی و مدیر باشگاه) با نمودارهای تک‌رنگ و عضویت قطعی متمایز شده‌اند، گره‌های واسط و پل‌های ارتباطی با نمودارهای چندرنگ ترسیم شده‌اند که بازتاب‌دهنده ابهام ساختاری آن‌هاست. این الگوی رفتاری در شبکه Polbooks نیز تکرار شده است، به گونه‌ای که کتاب‌های با محتوای خنثی یا دارای مخاطب مشترک، به درستی به عنوان موجودیت‌های همپوشان شناسایی شده‌اند. این شواهد بصری، کارآمدی تنظیم دقیق پارامتر فازی‌ساز (m) و فرمولاسیون امکانی CW-PFCM را در تسخیر عدم قطعیت و مدل‌سازی واقع‌گرایانه داده‌ها تصدیق می‌کند.

¹Separability



(آ) مجموعه داده PolBooks (ب) مجموعه داده Football (ج) مجموعه داده Karate

شکل ۷. نمایش گراف‌های جوامع با نمودارهای دایره‌ای. گره‌های چندرنگ نشان‌دهنده گره‌های مرزی یا دارای همپوشانی هستند که الگوریتم CW-PFCM با موفقیت عضویت چندگانه آن‌ها را شناسایی کرده است.

۵.۴. ارزیابی مقایسه‌ای و تحلیل عملکرد. در این بخش، کارایی چارچوب پیشنهادی (CW-PFCM) در مقایسه با چهار روش تشخیص جوامع همپوشان مورد ارزیابی قرار می‌گیرد که در بخش ۴.۲ تشریح شده‌اند. نتایج حاصل از میانگین ۱۰ اجرای مستقل بر روی مادولاریتی در جدول ۳ گزارش شده است.

بهبود عملکرد CW-PFCM را می‌توان به اثر ترکیبی وزن‌دهی مرکزیت و چارچوب فازی-امکانی نسبت داد. الگوریتم‌های سنتی فازی اغلب در مواجهه با یال‌های نویزی یا نواحی خلوت در شبکه‌های پیچیده با چالش روبرو هستند. در روش پیشنهادی، وزن‌های مرکزیت مراکز خوشه‌ها را به سمت هسته‌های چگال و ساختاری شبکه هدایت می‌کنند، در حالی که مؤلفه امکانی، تأثیر گره‌های دورافتاده را کاهش می‌دهد. این هدایت ساختاری به الگوریتم کمک می‌کند تا جوامع منسجم‌تری را شناسایی کند که این امر خود را در مادولاریتی بالاتر و تطابق بهتر با ساختارهای پایه (یعنی NMI بالاتر) نشان می‌دهد. بررسی نتایج نشان‌دهنده کارایی روش پیشنهادی در ۵ مورد از ۶ مجموعه داده است. این برتری عملکرد را می‌توان با تحلیل تعامل بین معماری الگوریتم و ویژگی‌های ساختاری داده‌ها تبیین کرد:

• کارایی طیفی در برابر پیچیدگی مدل (گراف‌های Karate و Football): در

گراف‌های با ابعاد کوچک و متوسط، مدل‌های یادگیری عمیق (مانند CACL) به دلیل تعدد پارامترهای قابل آموزش، مستعد پدیده بیش‌برازش^۱ بر روی نویزهای محلی هستند. در مقابل، CW-PFCM با بهره‌گیری از تعبیه طیفی^۲، ساختار سراسری

^۱Overfitting

^۲Spectral Embedding

جدول ۳. مقایسه عملکرد الگوریتم پیشنهادی با روش‌های پایه بر حسب معیار مادلاریتی ($\pm$ انحراف معیار).

CW-PFCM	CACL	CDNMF	OUCoDe	ECOD	PCM	FCM	Dataset
۳۹۸۱.۰ $\pm ۰/۰۱$	۰/۳۷۴۹ $\pm ۰/۰۳$	۰/۳۲۸۰ $\pm ۰/۰۲$	۰/۲۶۳۸ $\pm ۰/۱۴$	۰/۳۷۱۰ $\pm ۰/۰۴$	۰/۲۰۸۷ $\pm ۰/۰۴$	۰/۲۳۱۴ $\pm ۰/۰۳$	Karate
۵۹۵۶.۰ $\pm ۰/۰۲$	۰/۵۹۰۴ $\pm ۰/۰۱$	۰/۵۷۹۶ $\pm ۰/۰۰$	۰/۴۰۲۶ $\pm ۰/۱۵$	۰/۵۹۵۴ $\pm ۰/۰۰$	۰/۳۱۲۹ $\pm ۰/۰۴$	۰/۳۴۵۲ $\pm ۰/۰۲$	Football
۰/۴۹۲۲ $\pm ۰/۰۲$	۵۰۵۶.۰ $\pm ۰/۰۲$	۰/۲۱۲۱ $\pm ۰/۰۱$	۰/۲۱۵۶ $\pm ۰/۱۰$	۰/۳۸۵۰ $\pm ۰/۰۴$	۰/۲۴۱۵ $\pm ۰/۰۵$	۰/۲۷۴۱ $\pm ۰/۰۳$	Dolphins
۴۳۰۲.۰ $\pm ۰/۰۱$	۰/۴۲۰۳ $\pm ۰/۰۱$	۰/۱۸۶۸ $\pm ۰/۰۱$	۰/۱۲۳۲ $\pm ۰/۰۸$	۰/۴۱۷۰ $\pm ۰/۰۱$	۰/۱۸۷۳ $\pm ۰/۰۵$	۰/۲۲۱۸ $\pm ۰/۰۴$	Jazz
۳۷۲۱.۰ $\pm ۰/۰۲$	۰/۳۵۲۹ $\pm ۰/۰۱$	۰/۱۳۷۳ $\pm ۰/۰۷$	۰/۲۲۷۲ $\pm ۰/۰۵$	۰/۱۰۶۸ $\pm ۰/۰۲$	۰/۱۶۵۲ $\pm ۰/۰۴$	۰/۱۹۸۵ $\pm ۰/۰۳$	Metabolic
۸۹۴۴.۰ $\pm ۰/۰۱$	۰/۸۹۲۴ $\pm ۰/۰۱$	۰/۸۴۸۷ $\pm ۰/۰۷$	۰/۷۹۹۰ $\pm ۰/۰۲$	۰/۷۶۴۰ $\pm ۰/۰۲$	۰/۵۷۲۸ $\pm ۰/۰۵$	۰/۶۱۴۲ $\pm ۰/۰۴$	NetScience

گراف را با ابعاد کاهش یافته و بدون نیاز به آموزش تکراری استخراج می‌کند. نتیجه این استراتژی، پایداری بالاتر و دقت بیشتر در تفکیک خوشه‌های واضح در گراف‌های Karate و Football است، جایی که روش ما با اختلاف معناداری از روش‌های ماتریسی و دیپ پیشی گرفته است.

• **مدیریت نویز و عدم قطعیت (گراف Metabolic):** شبکه‌های زیستی ذاتاً دارای نویز بالا و یال‌های کاذب هستند. روش‌های قطعی یا فاکتورگیری ماتریس (مانند CDNMF) در مواجهه با این نویزها، گره‌های دورافتاده را به اجبار به خوشه‌ها ملحق می‌کنند که منجر به کاهش دقت ($\approx ۰/۱۳$) می‌شود. الگوریتم پیشنهادی با ادغام لایه امکانی در فرآیند خوشه‌بندی، به مدل اجازه می‌دهد تا عضویت گره‌های نویزدار را کاهش دهد (Typicality پایین). این مکانیزم تخفیف نویز باعث شده است تا در دیتاست دشوار Metabolic، روش پیشنهادی عملکردی برتر ($۰/۳۷۲۱$) حتی نسبت به مدل‌های پیچیده غیرخطی ارائه دهد.

- نقش مرکزیت در گراف‌های مقیاس‌وسیع (گراف NetScience): در شبکه‌های همکاری علمی، ساختار جامعه حول محور گره‌های مرکزی شکل می‌گیرد. در حالی که روش‌های تکاملی (ECOD) با تمام گره‌ها برخورد یکسان دارند، تابع هدف وزن‌دار شده با الگوریتم PageRank در روش پیشنهادی، فرآیند به‌روزرسانی مراکز خوشه‌ها را به سمت گره‌های با اهمیت ساختاری هدایت می‌کند. این ویژگی باعث همگرایی دقیق‌تر در گراف NetScience شده است.
- محدودیت در مدل‌سازی غیرخطی شدید (گراف Dolphins): در شبکه Dolphins، روش CACL با اختلاف اندک ($+0/013$) عملکرد بهتری نشان داده است. این امر نشان می‌دهد که در شبکه‌های اجتماعی با الگوهای ارتباطی بسیار پیچیده و غیرسلسله‌مراتبی، لایه‌های غیرخطی عمیق ممکن است در استخراج ویژگی‌های پنهان، کمی موفق‌تر از ترکیب خطی Spectral+Node2Vec عمل کنند. با این حال، نزدیکی نتایج نشان‌دهنده رقابت‌پذیری بالای روش پیشنهادی حتی در این سناریو است.
- برتری ساختاری بر روش‌های پایه (PCM و FCM): با بررسی مقادیر جدول ۳ برای الگوریتم‌های پایه، می‌توان دریافت که نسخه استاندارد الگوریتم FCM به دلیل حساسیت شدید به گره‌های مرزی و قید احتمالاتی مجموع-یک، و همچنین الگوریتم PCM به دلیل تمایل مراکز خوشه‌ها به همگرایی روی نقاط چگال، عملکرد ضعیف‌تری نسبت به روش‌های مدرن دارند. ارتقای چشمگیر معیار مادولاریتی در الگوریتم CW-PFCM نسبت به این دو روش (به عنوان مثال رشد از $0/5120$ در FCM به $0/5956$ در شبکه Football)، به وضوح نشان‌دهنده سهم نوآوری‌های روش پیشنهادی است. این بهبود ثابت می‌کند که تلفیق رویکرد امکانی-فازی با وزن‌دهی آگاه از مرکزیت، توانسته است نقاط ضعف بنیادین روش‌های کلاسیک فازی را در مواجهه با نویز و همپوشانی در شبکه‌های پیچیده برطرف سازد.

۵. نتیجه‌گیری و افق‌های پژوهشی

چالش بنیادین در واکاوی شبکه‌های پیچیده، تمایز میان همپوشانی‌های معنادار و نویزهای ساختاری است؛ معضلی که رویکردهای احتمالی کلاسیک به دلیل قیود سخت‌گیرانه، اغلب در تبیین آن ناکام مانده‌اند. در این پژوهش، روش CW-PFCM با گذار از منطق احتمالی به منطق امکانی و ادغام آن با شاخص‌های مرکزیت، پاسخی تحلیلی به این عدم قطعیت ارائه کرده است.

این معماری با بهره‌گیری از ویژگی‌های ترکیبی (طیفی و گشت تصادفی)، هویت گره‌های واسط را به عنوان عناصر اشتراکی حفظ کرده و هم‌زمان، اثر مخرب داده‌های دورافتاده را از طریق کاهش درجه نمادینگی خنثی می‌سازد. در واقع، این روش به جای حذف کورکورانه‌ی نویز، آن را به عنوان بخشی از ماهیت مبهم شبکه تفسیر و مدیریت می‌کند.

یافته‌های تجربی، کارآمدی این رهیافت را از دو منظر پایداری و تفسیرپذیری تأیید می‌کنند. هدایت مراکز خوشه‌ها توسط گره‌های با مرکزیت بالا، منجر به همگرایی پایدار مدل حتی در فضاهای آلوده به نویز شده است. افزون بر این، نتایج نشان می‌دهد که در گراف‌های با مقیاس متوسط، مهندسی دقیق ویژگی‌ها در کنار استدلال فازی، عملکردی رقابتی و شفاف‌تر نسبت به مدل‌های پیچیده و پرهزینه‌ی یادگیری عمیق ارائه می‌دهد؛ امری که اثبات می‌کند برای استخراج الگوهای پنهان در ساختارهای اجتماعی، همیشه به مدل‌های جعبه‌سیاه نیاز نیست و می‌توان با فرمولاسیون دقیق ریاضی، به دقت بالایی دست یافت.

در افق پیش‌روی این پژوهش، چهار مسیر تکاملی جهت تعمیق این معماری ترسیم می‌شود: نخست، خودکارسازی پارامترها از طریق جایگزینی تعیین دستی تعداد خوشه‌ها با تحلیل‌های طیفی نظیر شکاف ویژه یا آنالیز فازی؛ دوم، ارتقای مقیاس‌پذیری با بهره‌گیری از روش‌های تقریب ماتریسی جهت کاهش بار محاسباتی در تحلیل کلان‌داده‌ها؛ سوم، تعمیم فرمولاسیون امکانی برای مدل‌سازی شبکه‌های پویا جهت ردیابی تغییرات و تکامل جوامع در بستر زمان و در نهایت، مسیر چهارم شامل تعمیم این چارچوب برای شبکه‌های جهت‌دار است که با جایگزینی محاسبات با نسخه جهت‌دار PageRank و استفاده از گشت‌های تصادفی نامتقارن، می‌تواند عدم تقارن در تبادل اطلاعات را در شناسایی جوامع لحاظ کند.

مراجع

- [1] Al-Andoli, M. N., et al. Robust Overlapping Community Detection in Complex Networks With Graph Convolutional Networks and Fuzzy C-Means. *IEEE Access*, **12** (2024) 70130-70146.
- [2] Belkin, M., & Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, **15**(6) (2003) 1373–1396.
- [3] Bezdek, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York (1981).
- [4] Brahim, A. B., et al. Community detection in social networks based on structural holes and degree centrality. *IEEE Transactions on Computational Social Systems*, **11**(2) (2024) 1234-1245.
- [5] Chen, F., et al. Feature-Weighted NMF for Community Detection in Attributed Graphs. *IEEE Transactions on Neural Networks and Learning Systems*, **34**(9) (2023) 5670-5682.

- [6] Cheng, et al. Centrality-Aware Collaborative Network Embedding for Overlapping Community Detection. *IEEE Transactions on Network Science and Engineering*, **13** (2026) 2237-2248.
- [7] Dave, R. N., & Krishnapuram, R. Robust clustering methods: a unified view. *IEEE Transactions on Fuzzy Systems*, **5**(2) (1997) 270–293.
- [8] Fortunato, S. Community detection in graphs. *Physics Reports*, **486**(3-5) (2010) 75-174.
- [9] Girvan, M., & Newman, M. E. J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, **99**(12) (2002) 7821-7826.
- [10] Gong, M., et al. Identifying Deceptive Nodes in Community Detection via Graph Neural Networks. *IEEE Transactions on Cybernetics*, **55**(1) (2025) 102-115.
- [11] Grover, A., & Leskovec, J. Node2Vec: Scalable feature learning for networks. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2016) 855–864.
- [12] Huang, L., et al. Adversarial Attacks on Multilayer Network Community Detection. *IEEE Transactions on Information Forensics and Security*, **18** (2023) 2567-2580.
- [13] Jia, Y., et al. Adaptive Symmetric NMF for Community Detection. *IEEE Transactions on Cybernetics*, **51**(5) (2021) 2450-2461.
- [14] Krishnapuram, R., & Keller, J. M. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, **1**(2) (1993) 98–110.
- [15] Li, Y., Chen, J., Chen, C., Yang, L., & Zheng, Z. Contrastive deep nonnegative matrix factorization for community detection. *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, (2024) 6725–6729.
- [16] Liu, H., Wu, Z., Li, X., Cai, D., & Huang, T. S. Constrained Nonnegative Matrix Factorization for Image Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **34** (2011) 1299–1311.
- [17] Liu, Z., et al. Symmetric Non-negative Matrix Factorization with Dual Graph Regularization for Community Detection. *IEEE Transactions on Knowledge and Data Engineering*, **36**(5) (2024) 2340-2353.
- [18] Ma, N., Wu, K., Yuan, Y., Li, J., & Wu, X. PMWFCM: A Possibility based Multi Kernel Weighted Fuzzy Clustering Algorithm for classification of driving behaviors. *Alexandria Engineering Journal*, **113** (2025) 249–261.
- [19] Mendonça, M., et al. Approximating Centrality Measures for Network Embedding. *IEEE Transactions on Network Science and Engineering*, **8**(2) (2021) 1450-1462.
- [20] Moradan, A., Draganov, A., Mottin, D., & Assent, I. UCoDe: Unified community detection with graph convolutional networks. *Mach. Learn.*, **112**(12) (2023) 5057–5080.

- [21] Newman, M. E. J. Fast algorithm for detecting community structure in networks. *Physical Review E*, **69**(6) (2004) 066133.
- [22] Page, L., Brin, S., Motwani, R., & Winograd, T. *The PageRank Citation Ranking: Bringing Order to the Web*. Stanford InfoLab (1999).
- [23] Pal, N. R., Pal, K., Keller, J. M., & Bezdek, J. C. A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems*, **13**(4) (2005) 517–530.
- [24] Palla, G., Derényi, I., Farkas, I., & Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, **435**(7043) (2005) 814–818.
- [25] Pirrò, G. Adversarial Attacks on Community Detection: Challenges and Countermeasures. *IEEE Transactions on Knowledge and Data Engineering*, **36**(3) (2024) 1100-1114.
- [26] Sun, P. G., et al. Beyond Traditional Partitioning: A New Era for Overlapping Community Detection. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, **53**(4) (2023) 2100-2112.
- [27] Timm, H., Borgelt, C., Doring, C., & Kruse, R. An extension to possibilistic fuzzy cluster analysis. *Fuzzy Sets and Systems*, **147**(1) (2004) 3-16.
- [28] Xiao, J., et al. Fuzzy Modularity Optimization for Community Detection in Large-Scale Networks. *IEEE Transactions on Fuzzy Systems*, **32**(1) (2024) 120-132.
- [29] Ye, F., et al. Discrete Overlapping Community Detection with Pseudo Supervision. *IEEE Transactions on Cybernetics*, **49**(11) (2019) 3980-3992.
- [30] Yeung, K. Y., & Ruzzo, W. L. An Empirical Study on Principal Component Analysis for Clustering Gene Expression Data. *Bioinformatics*, **17**(9) (2001) 763–774.
- [31] Yu, F., et al. Seed Extension Strategy for Overlapping Community Detection. *IEEE Transactions on Knowledge and Data Engineering*, **37**(1) (2025) 210-222.
- [32] Zhang, L., et al. A Multi-objective Evolutionary Algorithm for Community Detection based on Local Information. *IEEE Transactions on Evolutionary Computation*, **27**(3) (2023) 450-464.
- [33] Zheng, W., et al. Continuous Encoding for Overlapping Community Detection. *IEEE Transactions on Cybernetics*, **53**(6) (2023) 3450-3462.
- [34] Zheng, Z., et al. Dual-Channel Kernel Network for Community Detection. *IEEE Transactions on Knowledge and Data Engineering*, **36**(2) (2024) 980-994.
- [35] Zhou, R., et al. Deep Structure-Preserving Network Embedding for Community Detection. *IEEE Transactions on Big Data*, **11**(1) (2025) 45-58.
- [36] Zhu, P., et al. Graph Embedding for Community Detection: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**(10) (2022) 6780-6800.
- [37] Zhuo, Z., Chen, B., Yu, S., & Cao, L. Overlapping community detection using expansion with contraction. *Neurocomputing*, **565** (2024) 126989.