مقاله پژوهشی

شناسایی جامعه در شبکه‌های پیچیده با استفاده از درخت پوشای مینیمم و بیشینه‌سازی ماژولاریتی

سندس بهادری و مریم نورائی آباده

چکیده: ماژولاریتی، یکی از ویژگی‌های برجسته شبکه‌های پیچیده است که ساختار این شبکه‌ها را به‌صورت گروه‌های جامعه‌ای تقسیم می‌کند. تاکنون روش‌های زیادی برای شناسایی جوامع در شبکه‌های پیچیده به کار گرفته شده است؛ اما برخی از این روش‌ها بهینه‌سازی‌های محلی دارند که ترتیب پردازش نودها، جواب نهایی را تحت تأثیر قرار می‌دهد. در این مقاله، یک روش جدید برای یافتن جوامع در شبکه‌های پیچیده با استفاده از تقسیم و ادغام پیشنهاد شده که در آن از درخت پوشای کمینه به‌عنوان ابزاری برای تشخیص عدم تشابه بین نودها استفاده می‌شود. در فرایند تقسیم، یال‌هایی که بیشترین عدم تشابه را نشان می‌دهند در درخت پوشای کمینه حذف می‌شوند تا گروه‌های کوچک‌تری از نودهای یک جامعه ایجاد شوند. در فرایند ادغام، هر گروه با گروه همسایه ادغام می‌شود که ترکیب آنها بیشترین افزایش ماژولاریتی را نسبت به گروه‌های همسایه دیگر داشته باشند. نتایج آزمایش‌های انجام‌شده بر روی شبکه‌های واقعی و شبکه‌های ساختگی نشان می‌دهند روش پیشنهادی در این مقاله، دقت مطلوبی برای شناسایی جوامع در شبکه‌های پیچیده دارد.

کلیدواژه: **شبکه‌های پیچیده، جامعه، درخت پوشای مینیمم، افزایش ماژولاریتی.**

۱- مقدمه

امروزه شبکه‌ها به‌صورت گسترده برای مدل‌کردن ارتباطات پیچیده بین افراد یا سازمان‌هایی که به دلایل مختلف از قبیل دوستی، خویشاوندی و غیره به هم وابسته هستند، استفاده می‌شوند. در یک شبکه، ساختار جامعه در میان ویژگی‌های خاص شبکه مانند دنیای کوچک [۱]، مقیاس مستقل [۲] و پیمانه‌ای‌بودن [۳] مهم‌ترین ویژگی است. یک جامعه مجموعه‌ای از گره‌هاست که ارتباط‌های داخلی بیشتری نسبت به ارتباط با گره‌های دیگر دارند. جامعه، کلاستر یا ماژول نیز نامیده می‌شود. شناسایی جامعه‌ها برای درک ساختار، وظایف و تکامل شبکه‌های پیچیده مختلف بسیار مهم است [۴] و [۵] و در کاربردهای گسترده‌ای مانند شناسایی ماژول‌های وظیفه‌ای در شبکه‌های زیستی [۶] تا [۸]، جمع‌آوری صفحات وب مرتبط با یک موضوع خاص [۹] و [۱۰]، گروه‌بندی نویسندگان بر اساس موضوع تحقیق در شبکه‌های همکاری نویسنده‌ها [۱۱] و [۱۲] و ... بسیار مورد توجه قرار گرفته است. در ادامه مقاله، شبکه پیچیده را به‌صورت گراف $G(V,E)$ در نظر گرفته‌ایم که V مجموعه نودها $(|V|=n)$ و E مجموعه یال‌ها $(|E|=m)$ را نشان می‌دهند.

تلاش‌های زیادی برای استخراج جامعه‌ها در شبکه‌های پیچیده صورت گرفته است. روش‌ها و الگوریتم‌های یافتن جامعه در شبکه‌ها، موضوع مورد علاقه محققان علوم مختلف از جمله فیزیک، ریاضی، آمار و علم کامپیوتر است و تاکنون تعداد زیادی الگوریتم برای یافتن جامعه در دهه‌های اخیر ارائه شده است [۱۳] تا [۱۶]. الگوریتم‌های تقسیم‌بندی طیفی[1] [۱۷] و [۱۸]، روش‌های مبتنی بر تقسیم و ترکیب[2] [۴] و [۱۹]، الگوریتم‌های تکاملی [۲۰] و [۲۱] و ماکسیمم‌کردن ماژولاریتی نمونه‌هایی از این الگوریتم‌ها هستند.

روش‌های طیفی بر اساس تحلیل مقادیر ویژه ماتریس‌های مربوط به شبکه هستند که توسط نیومن در [۲۲] بررسی شده‌اند. در روش‌های مبتنی بر تقسیم، سعی بر یافتن یال‌های بین جامعه‌ای و حذف آنهاست. بعد از حذف یال‌های بین جامعه‌ای، آنچه که باقی می‌ماند جامعه‌هاست. الگوریتم پیشگام در این زمینه الگوریتم گرون- نیومن است [۳] که در آن از معیار مرکزیت بینابینی[3] برای شناسایی یال‌های بین جامعه‌ها استفاده می‌شود. این معیار، تعداد مسیرهای کوتاه بین نودها را که از هر یال می‌گذرد تعیین می‌کند. این الگوریتم در بسیاری از شبکه‌ها توانسته است به‌صورت موفقیت‌آمیزی جامعه‌ها را شناسایی کند؛ اما مشکل اصلی این روش پیچیدگی زمانی $O(m^2n)$ است که بسیار بالاست. در متدهای مبتنی بر ترکیب، هر نود به‌عنوان یک جامعه در نظر گرفته شده و سپس بر اساس بعضی معیارهای شباهت، این جامعه‌ها با هم ادغام می‌شوند. یکی از معروف‌ترین الگوریتم‌های این دسته، روش پیشنهادشده توسط گرون- نیومن [۲۲] است که بر اساس مقدار ماژولاریتی است.

ماکسیمم‌نمودن ماژولاریتی، مهم‌ترین روش برای شناسایی جامعه در شبکه‌هاست. این روش بر اساس تابع مفید ماژولاریتی است که کیفیت تقسیم‌بندی شبکه به جامعه‌ها را اندازه‌گیری می‌کند. این تابع را می‌توان بر همه تقسیم‌بندی‌های ممکن شبکه اعمال نمود و آن تقسیم‌بندی را که بیشترین مقدار ماژولاریتی را داشته باشد، برگزید. تعداد تقسیم‌بندی‌های مختلف جامعه به‌صورت نمایی است؛ بنابراین به‌دست آوردن بهینه مطلق امکان‌پذیر نیست. در نتیجه به‌جای پیدانمودن بهینه مطلق از روش‌های بهینه نسبی مانند الگوریتم‌های حریصانه از قبیل [۳]، [۱۹] و [۲۳] تا [۲۵] استفاده می‌شود.

یک الگوریتم مشهور برای یافتن جامعه، Louvain [۲۶] است که در

این مقاله در تاریخ ۲۵ تیر ماه ۱۴۰۲ دریافت و در تاریخ ۱۰ بهمن ماه ۱۴۰۲ بازنگری شد.

سندس بهادری (نویسنده مسئول)، گروه مهندسی کامپیوتر، واحد ایلام، دانشگاه آزاد اسلامی، ایلام، ایران، (email: sondos.bahadori@iau.ac.ir).

مریم نورائی آباده، گروه مهندسی کامپیوتر، واحد آبادان، دانشگاه آزاد اسلامی، آبادان، ایران، (email: ma.nooraei@iau.ac.ir).

1. Spectral Partitioning
2. Divisive and Agglomeration
3. Betweenness

چندین بسته نرم‌افزاری تحلیل شبکه وجود دارد. Louvain یک الگوریتم اکتشافی بر اساس افزایش ماژولاریتی و در عمل یکی از سریع‌ترین الگوریتم‌های شناسایی جامعه است. در این الگوریتم ابتدا هر نود به‌عنوان یک جامعه در نظر گرفته می‌شود؛ سپس هر جامعه با جامعه‌ای که همسایه آن باشد و میزان افزایش ماژولاریتی ترکیب این دو جامعه، بیشتر از جامعه‌های همسایه دیگر باشد ترکیب می‌شوند و این روند تا جایی تکرار می‌شود که ترکیب جامعه‌ها امکان‌پذیر نباشد. همان طور که نویسنده در [۲۶] گفته است، نتیجه نهایی در این الگوریتم به ترتیب پردازش نودها بستگی دارد. برای به‌دست آوردن یک جواب بهینه مناسب، در این مقاله روشی جدید پیشنهاد شده که در آن به‌جای این که هر نود به‌عنوان یک جامعه اولیه در نظر گرفته شود از ساختار درخت پوشای مینیمم به‌عنوان عدم تشابه در بین نودها و سپس از تجزیه این درخت برای به‌دست آوردن جامعه‌های اولیه استفاده می‌شود و نهایتاً این جامعه‌ها بر اساس افزایش ماژولاریتی با جامعه همسایه‌ای که بیشترین افزایش ماژولاریتی را دارد ترکیب شود. این روش پیشنهادی را MST&MM[1] می‌نامیم. نتایج آزمایش‌های مختلف روی شبکه‌های واقعی و شبکه‌های ساختگی نشان می‌دهند که خروجی روش پیشنهادشده بهتر از الگوریتم Louvain و الگوریتم‌های دیگر است.

در این مقاله، یک روش تقسیم و ادغام جدید برای تشخیص جامعه بر اساس درخت پوشای کمینه و ماژولاریتی پیشنهاد می‌شود. درخت پوشای کمینه پس از وزن‌دهی گراف بر اساس حداقل شباهت گره‌های نقطه پایانی برای هر یال ساخته می‌شود. در مرحله بعد و در شرایطی که بهینه‌سازی ماژولاریتی امکان‌پذیر نباشد، این گروه از گره‌ها به‌طور مکرر با استفاده از الگوریتم Louvain ادغام می‌شوند. از این رو مزایای روش پیشنهادی از دو جنبه قابل بررسی است. یکی از مزایای این روش، استفاده از اولین گام بهینه‌سازی گره‌های اولیه با استفاده از درخت پوشای کمینه است که سبب بهبود دقت و زمان اجرا در مقایسه با روش‌های دارای انتخاب تصادفی می‌شود. بهبود دوم در استفاده از معیار ماژولاریتی به‌عنوان معیار کیفیت برای ساختار جوامع است. در بیشتر الگوریتم‌های تشخیص جامعه، ماژولاریتی به‌عنوان معیار کیفیت برای ساختار جامعه انتخاب شده است؛ اما هیچ یک از الگوریتم‌های موجود، گره‌های نامرتبط موجود در زیر جوامع را در نظر نگرفته‌اند. از این کمیت می‌توان برای تشخیص جوامع با دقت بیشتری استفاده کرد.

ادامه مباحث این مقاله به این شرح است: در بخش ۲، روش پیشنهادی و الگوریتم مربوط با جزئیات کامل بررسی شده و در بخش ۳ نیز تحلیل‌ها و نتایج آزمایش‌ها آمده است. در بخش چهارم به بررسی کارهای مشابه پرداخته شده و نهایتاً در بخش ۵ نتیجه‌گیری آمده است.

۲- راهکار پیشنهادشده

یک دسته از تکنیک‌های به‌دست آوردن جامعه‌ها در شبکه‌های پیچیده، تکنیک‌های توسعه محلی و بهینه‌سازی هستند که جواب نهایی در این روش‌ها به ترتیب پردازش نودها بستگی دارد. در این مقاله برای به‌دست آوردن یک جواب بهینه مناسب، روشی جدید بر اساس تقسیم و ادغام برای یافتن جامعه‌ها در شبکه‌های پیچیده پیشنهاد شده که در آن از ساختار درخت پوشای مینیمم به‌عنوان عدم تشابه در بین نودها استفاده می‌شود. در فرایند تقسیم، یال‌هایی که عدم تشابه بیشتری دارند در درخت پوشای مینیمم حذف می‌شوند تا زیرگروه‌های گسسته کوچکی از نودهای یک جامعه ایجاد شوند. در فرایند ادغام، زیرگروه‌ها به‌گونه‌ای با هم ادغام می‌شوند که جامعه‌های اصلی با حداکثرنمودن مقدار ماژولاریتی به‌دست آیند.

فرض کنید $G=(V,E)$ یک شبکه بدون جهت و بدون وزن باشد که $V=\{v_1,v_2,\ldots,v_n\}$ مجموعه نودها و $E=\{e_1,e_2,\ldots,e_m\}$ مجموعه یال‌هاست. هدف این مقاله، پیدانمودن k جامعه $\{c_1,c_2,\ldots,c_k\}$ است که $c_i\neq\emptyset$، $c_i\cap c_j=\emptyset$ و $V=\cup_{i=1}^{k}v_i$. از آنجا که به‌دنبال حذف یال‌های بین رئوسی هستیم که تشابه کمتری دارند برای این منظور از یک معیار عدم تشابه برای برچسب‌زدن یال‌ها استفاده می‌شود [۲۷]. وزن یال‌ها را می‌توان به‌صورت تابع $\omega:E\rightarrow R$ تعریف نمود که مقدار وزن هر یال $e_l=(v_i,v_j)$ نشان‌دهنده عدم تشابه بین دو یال است. عبارت عدم تشابه را می‌توان به‌صورت زیر تعریف نمود

$$\omega(e_l)=\omega(v_i,v_j)=\frac{\left|\Gamma(v_i)\right|+\left|\Gamma(v_j)\right|}{\left|\Gamma(v_i)\cap\Gamma(v_j)\right|} \qquad (۱)$$

که $\Gamma(v_j)$ مجموعه نودهای همسایه نود v_j است. با توجه به (۱) اگر دو رأس تشابه کمتری داشته باشند یال بین آنها وزن بیشتری می‌گیرد و بنابراین احتمال بیشتری برای حذف آن یال هنگام یافتن درخت پوشای مینیمم وجود دارد.

به‌منظور انجام فرایند تقسیم، در گرافی که وزن یال‌های آن با استفاده از تابع $\omega(e_l)$ محاسبه می‌شود ابتدا درخت پوشای مینیمم ایجاد می‌شود. درخت پوشای مینیمم برای گراف $G=(V,E)$، درخت $T=(V,E')$ است که $E'\subset E$ و $W(T)=\sum_{(v_i,v_j)\in T}\omega(v_i,v_j)$ مینیمم باشد که $\left|E'\right|=n-1$ است. در درخت پوشای مینیمم T، مجموعه R_e را انتخاب نموده که R_e مجموعه‌ای با $(n-1)/2$ از یال‌های درخت T با وزن بیشتر است. با حذف‌کردن یال‌های مجموعه R_e از درخت T، در واقع $(n-1)/2$ مؤلفه جدا در T ایجاد می‌گردند و هر یک از این مؤلفه‌ها به‌عنوان یک جامعه اولیه در نظر گرفته می‌شوند؛ در نتیجه تعداد k $(k=(n+1)/2)$ جامعه اولیه ایجاد می‌شود.

بعد از تقسیم درخت پوشای مینیمم به k جامعه، فاز ادغام برای به‌دست آوردن جامعه‌های نهایی ساختار گراف اعمال می‌شود. گام اساسی در فاز ادغام این است که جامعه‌ها بر چه اساسی با هم ادغام شوند. چون هدف، افزایش ماژولاریتی است در فاز ادغام بدین صورت عمل می‌شود که برای جامعه i، تمام جامعه‌های همسایه آن مانند j را در نظر گرفته و مقدار افزایش ماژولاریتی در اثر ادغام جامعه i با هر یک از جامعه‌های همسایه j را محاسبه نموده و جامعه i با جامعه‌ای که مقدار افزایش ماژولاریتی بیشتری دارد ادغام می‌شود. این عمل را آن قدر تکرار نموده تا مقدار ماژولاریتی تغییر نکند. برای محاسبه افزایش مقدار ماژولاریتی از عبارت زیر استفاده می‌شود [۲۶]

$$\Delta Q=[\frac{\sum_{in}+2k_{i,in}}{2m}-(\frac{\sum_{tot}+k_i}{2m})^2]-[\frac{\sum_{in}}{2m}-(\frac{\sum_{tot}}{2m})^2-(\frac{k_i}{2m})^2] \qquad (۲)$$

که $\sum_{in}$ تعداد یال‌های درون جامعه j، $\sum_{tot}$ مجموع درجه نودهای جامعه j، k_i مجموع درجه نودهای جامعه i، $k_{i,in}$ تعداد یال‌های بین دو جامعه i و j و m تعداد یال‌های گراف است.

الگوریتم پیشنهادی را می‌توان به‌صورت شکل ۱ بیان نمود.

1. Minimum Spanning Tree and Modularity Maximization

Procedure MST&MM

Input: Graph G (V, E) a connected graph
output: discrete communities.
1: calculate weight matrix using Eq. (1)
2: find Minimum Spanning Tree (T) in G.
3: remove $(n-1)/2$ edges with highest weight in T.
4: set $(n+1)/2$ remaining subgraph as initial community.
5: calculate modularity(Q).
6: while (Q increases){
7: merge neighbor subgraphs with largest modularity maximization using equation Eq. (2).
8:}
9: return merged subgraphs as communities.

شکل ۱: الگوریتم روش پیشنهادی.

۲-۱ مثال حل‌شده

در شکل ۲- الف گرافی آمده که از آن برای نمایش مراحل اجرای روش پیشنهادی در این مقاله استفاده می‌شود. همان طور که در شکل ۱- الف نشان داده شده است، این گراف دارای دو جامعه جدا از هم است. بعد از اجرای گام ۱ و ۲ از روش پیشنهادی، خروجی برنامه درخت پوشای مینیمم T می‌باشد که در شکل ۲- ب نشان داده شده است. پس از اجرای گام ۳، شکل ۲- ج ایجاد می‌شود که نهایتاً با اجرای گام ۴ جامعه‌های نهایی نشان داده شده در شکل ۲- د حاصل می‌گردد. همان طور که در شکل ۲ نشان داده شده است بر روی گراف شکل الف، ابتدا با استفاده از معیار عدم تشابه، یال‌ها را وزن‌دار و سپس درخت پوشای مینیمم را مانند شکل ب پیدا می‌کنیم. در گام بعدی $(n-1)/2$ از یال‌های با بیشترین برچسب حذف شده و شکل ج حاصل می‌گردد و در گام آخر، زیرجامعه‌های به‌دست آمده با استفاده از افزایش ماژولاریتی با هم ادغام شده تا جوامع نهایی حاصل شوند (مانند شکل د).

۲-۲ تحلیل پیچیدگی زمانی الگوریتم پیشنهادشده

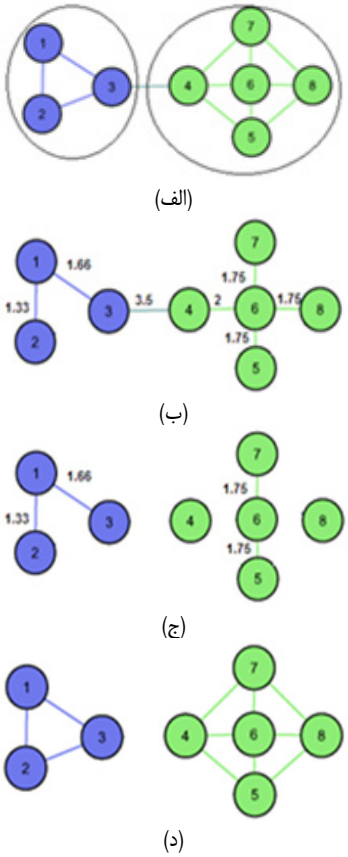
برای تحلیل پیچیدگی زمانی الگوریتم پیشنهادشده باید همه گام‌های اساسی آن را در نظر گرفت. محاسبه ماتریس عدم تشابه دارای پیچیدگی زمانی $O(m)$ و پیداکردن درخت پوشای مینیمم دارای پیچیدگی زمانی $O(m\log n)$ است. حذف مجموعه R_e از مجموعه یال‌های درخت پوشای مینیمم T دارای حداکثر پیچیدگی زمانی $O(n^2)$ است. ادغام زیرگراف‌های به‌دست آمده بر اساس حداکثر مقدار افزایش ماژولاریتی دارای پیچیدگی زمانی $O(kn^2)$ است.

۳- نتایج آزمایش‌ها و تحلیل

۳-۱ معیارهای ارزیابی

معیارهای ارزیابی جامعه‌های به‌دست آمده را می‌توان به دو دسته کلی معیارهای کیفیت و معیارهای دقت دسته‌بندی نمود. از معیارهای کیفیت معمولاً زمانی استفاده می‌شود که زمینه اصلی گراف در دسترس نباشد مانند شبکه‌های دنیای واقعی. ازجمله مهم‌ترین معیارهای کیفیت می‌توان معیار ماژولاریتی و معیار چگالی را نام برد. معیارهای دقت زمانی استفاده می‌شوند که زمینه اصلی گراف موجود باشد و یکی از مهم‌ترین معیارهای دقت معیار NMI است. برای ارزیابی الگوریتم پیشنهادشده در این مقاله از سه معیار ماژولاریتی (Q)، چگالی (D) و NMI استفاده شده که به‌صورت زیر تعریف می‌شوند:

- ماژولاریتی (Q): از این معیار برای ارزیابی روش پیشنهادشده بر روی گراف‌های واقعی استفاده می‌شود. این معیار به‌صورت زیر تعریف می‌گردد



شکل ۲: مراحل اجرای الگوریتم پیشنهادی.

$$Q=\sum_{i=1}^{k}\left[\frac{|m|_i}{|m|}-\left(\frac{d(i)}{2|m|}\right)^2\right] \quad (۳)$$

که $|m|_i$ تعداد یال‌های درون جامعه i، $|m|$ تعداد کل یال‌های گراف و $d(i)$ مجموع درجه نودهای موجود در جامعه i است.

- چگالی (D): این معیار به‌صورت زیر تعریف می‌شود

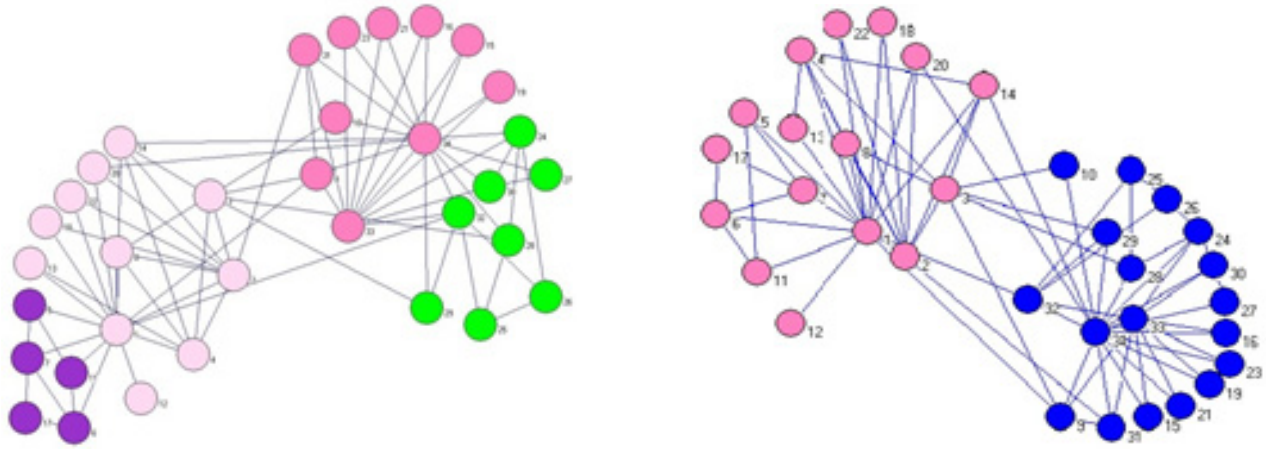
$$D=\frac{2}{|m|}\sum_{c=1}^{r} m_c\frac{m_c-(n_c-1)}{(m_c-2)(m_c-1)} \quad (۴)$$

که m تعداد یال‌ها، m_c تعداد یال‌های درون جامعه‌ای، n_c تعداد نودهای هر جامعه و r تعداد جامعه‌هاست.

- NMI به‌صورت زیر تعریف می‌شود

$$NMI(C,C')=\frac{-2\sum_{i=1}^{k}\sum_{j=1}^{k'} n_{ij}\log\frac{n_{ij}n}{n_i^C n_j^{C'}}}{\sum_{i=1}^{k} n_i^C\log\frac{n_i^C}{n}+\sum_{j=1}^{k'} n_j^{C'}\log\frac{n_j^{C'}}{n}} \quad (۵)$$

که در این عبارت C برچسب جامعه در زمینه اصلی، C' برچسب جامعه به‌دست آمده، k تعداد جامعه‌ها، n_i^C تعداد نودهای زمینه اصلی در جامعه iام، $n_j^{c'}$ تعداد نودهای به‌دست آمده برای جامعه j و n_{ij} تعداد نودهای جامعه i در زمینه اصلی است که برچسب j با استفاده از الگوریتم استفاده‌شده به آنها نسبت داده شده است. اگر نتایج به‌دست آمده با زمینه اصلی یکسان باشد $NMI(C,C')$ ماکسیمم مقدار یعنی یک و در شرایطی که نتایج به‌دست آمده با زمینه اصلی کاملاً متفاوت باشد $NMI(C,C')$ صفر می‌شود. بر اساس مفاهیم تئوری اطلاعات، صورت کسر مطابق با اطلاعات انحصاری بین نتایج جامعه‌های دو پارتیشن و مخرج کسر آنتروپی



(الف) (ب)

شکل ۳: (الف) زمینه اصلی گراف کاراته و (ب) جامعه‌های به‌دست آمده با استفاده از روش پیشنهادی.

جدول ۱: مقایسه معیار ماژولاریتی روش پیشنهادی با روش‌های OSLOM، LOVAIN، LOCAL COMMUNITY DETECTION و MST&MODULATITY بر روی گراف‌های واقعی.

گراف‌های واقعی	الگوریتم				
	MST&MM	[۲۷] MST&Modulatity	[۲۹] Local community detection	[۲۸] Louvain	[۳۰] OSLOM
Football	۰٫۵۰۷۰	۰٫۴۹۳۰۷	۰٫۳۹۳	۰٫۴۹۶۹	۰٫۴۷۷
Karate	۰٫۴۱۱۷	۰٫۴۱۵	۰٫۳۸۱۴	۰٫۴۳۳	۰٫۳۵۹۵
Dolphin	۰٫۵۰۱۴۲	۰٫۴۷۸۹	۰٫۲۹۷۳	۰٫۴۹۶	۰٫۴۳۶
Book	۰٫۵۰۷	۰٫۴۹۴۶	۰٫۴۴۸۵	۰٫۴۹۶	۰٫۵۰۷

جدول ۲: مقایسه معیار DENSITY روش پیشنهادی با روش‌های OSLOM، LOUVAIN، LOCAL COMMUNITY DETECTION و MST&MODULARITY بر روی گراف‌های واقعی.

گراف‌های واقعی	الگوریتم				
	MST&MM	[۲۷] MST&Modularity	[۲۹] Local community detection	[۲۸] Louvain	[۳۰] OSLOM
Football	۰٫۳۱۸۳	۰٫۰۲۴۰	۰٫۰۰۷۶	۰٫۰۱۵۹	۰٫۲۸۲۴
Karate	۰٫۰۷۹۷	۰٫۰۶۵۵	۰٫۰۳۰۱	۰٫۰۵۸۷	۰٫۰۰۱۷
Dolphin	۰٫۰۸۵۶۲	۰٫۰۳۸۷	۰٫۰۱۳۳	۰٫۰۳۶۶	۰٫۰۸۹۲
Book	۰٫۱۶۲۶	۰٫۰۱۳۲	۰٫۰۰۷۰	۰٫۰۱۲۴	۰٫۱۵۳۷

است. هرچه مقدار دو معیار Q و NMI بیشتر باشد کارایی الگوریتم مطلوب‌تر است. در ادامه ابتدا نتایج آزمایش را بر روی گراف‌های واقعی و سپس نتایج آزمایش را بر روی گراف‌های کامپیوتری بررسی می‌کنیم.

۳-۲ شبکه‌های دنیای واقعی

کارایی روش پیشنهادشده در این مقاله از نظر معیار ماژولاریتی و چگالی بر روی گراف‌های واقعی با روش پیشنهادی در [۲۹] و روش‌های Louvain [۲۸] و OSLOM [۳۰] مقایسه شده است. همچنین کارایی روش این مقاله با روش پیشنهادی [۲۷] مقایسه شده است. در این مقاله از درخت پوشای مینیمم برای تجزیه شبکه استفاده شده است؛ اما برای ترکیب، هر بار جامعه‌ها را با هم ترکیب نموده و ترکیبی را که ماژولاریتی بیشتری داشته باشد انتخاب می‌کند. این روش را MST&modulatity می‌نامیم. در ادامه ابتدا گراف‌های واقعی معرفی شده‌اند و نتیجه ارزیابی در جداول ۱ و ۲ آمده‌اند.

۳-۲-۱ فوتبال

شبکه فوتبال[1] یک شبکه معروف برای تعیین جامعه‌هاست [۴]. این شبکه دارای ۱۱۵ تیم است که به ۱۱ باشگاه تقسیم می‌شوند. تیم‌های داخل یک باشگاه، بازی بیشتری با هم دارند. در واقع تیم‌های داخل یک باشگاه، یک جامعه را تشکیل می‌دهند و هر باشگاه یک جامعه است. از این ۱۱۵ تیم، ۴ تیم عضو هیچ باشگاهی نیستند. روش پیشنهادشده در این مقاله برای این دیتاست ۷ جامعه را شناسایی می‌کند. از آنجا که یکی از اهداف این مقاله، تقسیم‌بندی شبکه با افزایش ماژولاریتی است، ماژولاریتی این تقسیم‌بندی ۰٫۵۷۰ است؛ در حالی که ماژولاریتی زمینه اصلی این شبکه ۰٫۵۵۴ است.

۳-۲-۲ باشگاه کاراته

باشگاه کاراته[2] یک شبکه اجتماعی است که ارتباط بین ۳۴ نفر در یک باشگاه دانشگاهی آمریکا را نشان می‌دهد. باشگاه به علت اختلاف بین یکی از مدرسین و مسئول باشگاه به دو جامعه تقسیم شده و زمینه اصلی این گراف در شکل ۳- الف نشان داده شده است. روش پیشنهادشده در این مقاله برای این دیتاست، ۴ جامعه را شناسایی می‌کند که این جامعه‌های شناسایی‌شده در شکل ۳- ب نشان داده شده است. ماژولاریتی این تقسیم‌بندی ۰٫۴۱۱۷ می‌باشد؛ در حالی که ماژولاریتی زمینه اصلی شبکه ۰٫۳۷۱ است.

1. American College Football

2. Zachary's Karate Club

جدول ۳: مشخصات گراف‌های ایجادشده به‌وسیله LFR Benchmark.

Network	n	δ	$\max\delta$	μ	t_1	t_2	$\min c$	$\max c$
n۱۰۰۰	۱۰۰۰	۲۰	۵۰	۰٫۱ −۰٫۸	۱	۲	۲۰	۵۰
n۲۰۰۰	۲۰۰۰	۴۰	۱۰۰	۰٫۱ −۰٫۸	۱	۲	۴۰	۱۰۰

جدول ۴: نتیجه مقایسه مقدار پارامتر NMI بر روی گراف‌های n۱۰۰۰ با تغییر مقدار پارامتر μ برای روش پیشنهادی در این مقاله با روش‌های OSLOM، Louvain، Local community detection و MST&Modularity.

الگوریتم	μ							
	۰٫۸	۰٫۷	۰٫۶	۰٫۵	۰٫۴	۰٫۳	۰٫۲	۰٫۱
OSLOM	۰٫۰۱۵۸	۰٫۰۸۰۲	۰٫۱۹۸۰	۱	۱	۱	۱	۱
Louvain	۰٫۲۶۹	۰٫۳۸۷	۰٫۵۴	۰٫۶۲	۰٫۶۶۷	۰٫۷۱۵	۰٫۷۱۸	۰٫۸۳
Local community detection	۰٫۰۱۵	۰٫۰۳۲	۰٫۰۴۳	۰٫۲۵۵	۰٫۲۹	۰٫۵۵۵	۱	۱
MST&Modularity	۰٫۰۰۳	۰٫۰۰۳	۰٫۰۰۲	۰٫۰۰۶	۰٫۸۷۷	۰٫۹۴۸	۰٫۹۷۵	۰٫۹۸۹
MST&MM	۰٫۲	۰٫۵۵	۰٫۹۳	۱	۱	۱	۱	۱

جدول ۵: نتیجه مقایسه مقدار پارامتر NMI بر روی گراف‌های n۲۰۰۰ با تغییر مقدار پارامتر μ برای روش پیشنهادی در این مقاله با روش‌های OSLOM، Louvain، Local community detection و MST&Modularity.

الگوریتم	μ							
	۰٫۸	۰٫۷	۰٫۶	۰٫۵	۰٫۴	۰٫۳	۰٫۲	۰٫۱
OSLOM	۰٫۲۰۵	۰٫۴۱	۰٫۷۹۸	۰٫۹۰۵	۰٫۹۰۸	۰٫۹۱۲	۰٫۹۴۰	۰٫۹۵۶
Louvain	۰٫۲۳۶	۰٫۳۵۸۹	۰٫۵۳۳۷	۰٫۶۳۷۹	۰٫۷۰۸۷	۰٫۷۶۱	۰٫۷۸۶۹	۰٫۸۲۷
Local community detection	۰	۰	۰	۰	۰	۰	۰٫۶۸۴	۰٫۷۶۷۹
MST&Modularity	۰	۰	۰	۰	۰٫۹۱۴۸	۰٫۹۶	۰٫۹۹۱	۰٫۹۹۷۴
MST&MM	۰٫۲۲۳	۰٫۴۹۸۶	۰٫۸۹۹۸	۰٫۹۵۳۹	۰٫۹۷۶	۰٫۹۸۷۹	۰٫۹۹۳	۰٫۹۹۹۲

۳-۲-۳ شبکه دلفین

این شبکه از مشاهداتی که ۷ سال بر روی ۶۲ دلفین صورت گرفته ایجاد شده است [۲۸]. نودها در این شبکه نشان‌دهنده دلفین‌هاست و نودهایی که بین آنها یال وجود دارد دارای تعاملات بیشتری نسبت به سایر نودها هستند. مطالعات قبلی بر روی این گراف، ۲ و ۴ جامعه را تشخیص داده‌اند. روش پیشنهادشده در این مقاله برای این دیتاست ۵ جامعه را شناسایی می‌کند و ماژولاریتی این تقسیم‌بندی ۰٫۵۱۴۲ است؛ این در حالی است که مقدار این پارامتر برای زمینه اصلی این گراف ۰٫۳۷۳۵ می‌باشد.

۳-۲-٤ شبکه کتاب

شبکه کتاب[1] خرید چندین عنوان کتاب را درباره سیاست‌های آمریکا مدل می‌کند که در طول انتخابات سال ۲۰۰۴ ریاست جمهوری چاپ شده‌اند. این گراف شامل ۱۰۵ کتاب (نود) و ۴۴۱ یال بین کتاب‌های خریداری‌شده با هم است [۲۷]. مقدار ماژولاریتی به‌دست آمده برای این گراف با استفاده از روش پیشنهادی این مقاله ۰٫۵۰۷ است.

۳-۳ گراف‌های ساختگی

۳-۳-۱ گراف‌های ایجادشده با LFR Benchmark

یک دسته از گراف‌های ساختگی که برای ارزیابی کارایی روش پیشنهادی در این مقاله مورد استفاده قرار گرفته‌اند، گراف‌های ساختگی با استفاده از LFR Benchmark [۳۱] است. گراف‌های ایجادشده به‌وسیله LFR از نظر ساختاری و توزیع درجه نودها مشابه گراف‌های واقعی هستند. در دو دسته گراف‌های ایجادشده، تعداد نودها را ۱۰۰۰ و ۲۰۰۰ در نظر گرفتیم؛ این گراف‌ها را به‌ترتیب n۱۰۰۰ و n۲۰۰۰ نامیدیم و مقدار پارامتر μ را که میزان تداخل جامعه‌هاست از ۰٫۱ تا ۰٫۸ تغییر دادیم. پارامتر δ که میانگین درجه نودهاست برای n۱۰۰۰ و n۲۰۰۰ به‌ترتیب ۲۰ و ۴۰ در نظر گرفته شده است. حداقل اندازه هر جامعه را در n۱۰۰۰ مقدار ۲۰ و در n۲۰۰۰ مقدار ۴۰ و حداکثر اندازه هر جامعه را در n۱۰۰۰ مقدار ۵۰ و در n۲۰۰۰ مقدار ۱۰۰ نود در نظر گرفتیم. در جدول ۳ مشخصات گراف‌های ایجادشده آمده است. $\max\delta$ حداکثر درجه هر نود است. $\min c$ حداقل تعداد نودهای هر جامعه و $\max c$ حداکثر تعداد نودهای هر جامعه است. برای هر مقدار پارامتر μ، ۱۰ گراف ایجاد نموده و مقدار پارامتر NMI، میانگین مقدار این پارامتر برای گراف‌های ایجادشده است. از پارامتر NMI برای ارزیابی روش پیشنهادی بر روی گراف‌های LFR Benchmark استفاده شده است. مانند گراف‌های واقعی نتیجه این روش با روش پیشنهادی در [۲۹] و روش‌های Louvain [۲۸]، OSLOM [۳۰] و MST&Modularity [۲۷] مقایسه شده و نتیجه این مقایسه‌ها در جداول ۴ و ۵ آمده است. همان طور که نشان داده شده است، کارایی روش پیشنهادی در این مقاله در همه گراف‌های LFR از سایر روش‌ها به‌جز روش OSLOM بهتر است.

۳-۳-۲ گراف‌های ایجادشده با استفاده از مدل گرون- نیومن

گراف‌هایی که به‌وسیله این مدل تولید می‌شوند به این صورت هستند که دارای ۳۲، ۶۴ و ۱۲۸ نود می‌باشند. رئوس این گراف‌ها به ۴ جامعه با اندازه یکسان به‌ترتیب با اندازه‌های ۸، ۱۶ و ۳۲ نود تقسیم شده‌اند. هر رأس دارای z_{in} یال به رئوسی است که با هم در یک جامعه قرار دارند. همچنین هر رأس دارای z_{out} یال به رئوسی است که عضو جامعه‌های دیگر هستند. در گراف‌های با ۶۴ و ۱۲۸ نود مقادیر z_{in} و z_{out} را

1. Books

جدول ۶: درصدی از نودها که در گراف با ۳۲ نود با تغییر Z_{out} درست خوشه‌بندی شده‌اند.

الگوریتم	Z_{out}			
	۳	۲	۱	۰
OSLOM	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪
Louvain	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪
Local community detection	۱۳٪	۱۶٪	۱۰۰٪	۱۰۰٪
MST&Modularity	۲۸٪	۵۰٪	۱۰۰٪	۱۰۰٪
MST&MM	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪

به‌گونه‌ای تغییر می‌دهیم که $z_{out}+z_{in}=۱۶$ باشد. در گراف با ۳۲ نود، مقادیر z_{in} و z_{out} را به‌گونه‌ای تغییر می‌دهیم که $z_{out}+z_{in}=۷$ باشد. با افزایش مقدار z_{out} از مقادیر کوچک‌تر به مقادیر بیشتر، تشخیص دقیق خوشه‌ها سخت‌تر می‌شود. در جداول ۶ تا ۸ کسری از نودها که به‌صورت صحیح توسط الگوریتم پیشنهادی این مقاله و روش شناسایی جامعه‌ها با استفاده از همسایگی محلی [۲۹] روش‌های Louvain [۲۸]، OSLOM [۳۰] و MST&Modularity [۲۷] خوشه‌بندی شده‌اند به‌صورت تابعی از z_{out} نمایش داده شده است. همان طور که در این جداول نشان داده شده است، عملکرد الگوریتم پیشنهادشده خوب بوده و از هر دو الگوریتم Louvain و MST&Modularity بهتر است.

کارایی روش پیشنهادی بر اساس مقدار معیار NMI در مقایسه با روش شناسایی جامعه‌ها با استفاده از همسایگی محلی [۲۹] روش‌های Louvain [۲۸] و OSLOM [۳۰] برای گراف با ۳۲، ۶۴ و ۱۲۸ نود به ترتیب در جداول ۹ تا ۱۱ آمده است. همان طور که در این شکل‌ها نشان داده شده، کارایی روش پیشنهادی در این مقاله در همه این گراف‌ها از سایر روش‌ها به‌جز روش OSLOM بهتر است.

٤- کارهای مرتبط

در سال‌های اخیر تلاش‌های زیادی برای کشف این موضوع صورت گرفته است. روش گیروان و نیومن [۳] یک رویکرد تشخیص جامعه را بر اساس تعداد کوتاه‌ترین مسیرهایی که از لبه‌ای به نام مرکزیت بین‌المللی عبور می‌کنند، معرفی کرد. دونتی و همکاران [۳] یک رویکرد خوشه‌بندی سلسله‌مراتبی را پیشنهاد کردند که در آن شباهت بین گره‌ها توسط بردارهای ویژه ماتریس لاپلاسی اندازه‌گیری می‌شود. iDBLINK [۳۲] یک الگوریتم خوشه‌بندی پیوند مبتنی بر چگالی افزایشی برای تشخیص جامعه در شبکه‌های پویاست. این الگوریتم تراکم پیوند را در هر شبکه برای هر موردی مانند ایجاد، رشد، ادغام، حذف، انقباض و تقسیم جوامع پیوندها تنظیم می‌کند که می‌تواند ساختار جامعه فعلی را با توجه به ساختار قبلی جامعه به‌سرعت و کارآمد به‌روز کند. InOrder [۳۳] یک الگوریتم تشخیص جامعه مبتنی بر چگالی در شبکه‌های پویاست که شامل دو مرحله است؛ یک فاز آنلاین که در آن توالی پیمایش شبکه حفظ می‌شود و یک فاز آفلاین که در آن جوامع مورد نظر از دنباله استخراج می‌شوند. SCAN [۳۴] خوشه‌ها، هاب‌ها و نقاط پرت را در شبکه‌ها تشخیص می‌دهد. این روش یک گره خاص را با توجه به مشارکت همسایگانش در آن زیرگروه به یک زیرگروه اختصاص می‌دهد. الگوریتم انتشار برچسب مبتنی بر اهمیت گره (NI-LPA) [۳۵] برای شناسایی جوامع همپوشانی در نمودارها پیشنهاد شده و یک مکانیسم استنتاج اطلاعاتی از ویژگی‌های گره‌ها را معرفی می‌کند تا یک فرایند انتشار و فیلترکردن خاص را درک کند. چن و همکاران [۳۶] یک تکنیک انتشار برچسب مبتنی بر آنتروپی اطلاعات را برای شناسایی جوامع در شبکه‌های اجتماعی پیشنهاد کردند. در این رویکرد، ضریب خاصی برای به‌روزرسانی برچسب گره‌ها اعمال می‌شود. رغوان و همکاران [۱۵] یک الگوریتم تشخیص جامعه مبتنی بر انتشار برچسب را معرفی کردند که در آن، یک برچسب منحصربه‌فرد به هر گره اختصاص داده می‌شود و پس از هر مرحله، برچسب گره بر اساس برچسبی که اکثر همسایگان آن در حال حاضر دارند به‌روز می‌شود. GLPA [۳۷]، الگوریتم انتشار برچسب مبتنی بر گراف، جوامع را در دو مرحله تشخیص می‌دهد. ابتدا هر برچسب گره بر اساس شباهت گره بین همسایگان اختصاص داده شده و در مرحله دوم، یک نمودار وزنی از اجزای متصل (یا گره‌های فوق‌العاده) برای شناسایی جوامع نهایی ساخته می‌شود. بلال سعود و همکارش [۳۸] یک رویکرد جدید تشخیص جامعه را بر اساس درخت پوشای حداقل و ماژولاریتی پیشنهاد کردند. این رویکرد، ابتدا زیرگروه‌های کوچک جداشده از یک شبکه را با پارامتر عدم تشابه گره مبتنی بر MST تولید می‌کند. پس از آن، ادغام جامعه در میان زیرگروه‌های مختلف، تنها زمانی انجام می‌شود که افزایش قابل توجهی در ماژولاریتی پس از ترکیب وجود داشته باشد. وو و همکاران [۳۹] یک روش درخت پوشای حداقل دو دور را بر اساس یک ماتریس فاصله برای شناسایی ساختار جامعه در شبکه‌ها پیشنهاد کردند. ژو و همکاران [۴۰] یک رویکرد بهینه‌سازی ماژولاریتی با K-plexes (MOKP) را برای شناسایی جوامع در شبکه‌های پیچیده پیشنهاد کرد. در فاز اول، MOKP با استفاده از k-plexes، دانه‌های جامعه را تولید می‌کند و در فاز دوم، گره‌های باقیمانده بر اساس بهینه‌سازی ماژولاریتی به گروه‌های مربوط اختصاص داده می‌شوند. الگوریتم انتقال حالت گسسته مبتنی بر ماژولاریتی [۴۱] (MDSTA) برای مشکل تشخیص جامعه از عملگرهای رأس و جایگزین جامعه برای جستجوی سراسری استفاده می‌کند. علاوه بر این، یک عملیات متقاطع دوطرفه بر روی افراد نخبه بهینه‌شده انجام می‌شود تا از راه حل‌های بهینه محلی فرار کنند. از سوی دیگر در گذشته نه چندان دور، تعداد زیادی روش تشخیص جامعه پیشنهاد شده که بر تابع نقشه‌برداری مبتنی بر دانش ساختاری محلی یا جهانی مانند LabelRank [۴۲]، EAGLE [۴۳] و الگوریتم Louvain [۴۴] تکیه دارند. یک رویکرد چندمرحله‌ای به نام ma-CODE [۳۳] برای یافتن جوامع بر اساس تداعی‌های ذاتی آنها بدون اینکه از قبل بدانیم چند جامعه وجود دارد استفاده می‌شود. تشخیص جامعه همپوشانی ذاتی در شبکه‌های افزایشی [۴۵] (InOvIn) یک رویکرد مبتنی بر فازی خشن برای تشخیص جوامع غیرمتشابه، همپوشانی و سلسله‌مراتبی (ذاتی) در شبکه‌های در حال تکامل است. در این روش، نویسنده از خوشه‌بندی فازی خشن برای یافتن جوامع همپوشانی استفاده می‌کند. علاوه بر این برای آشکارساختن ساختارهای جامعه ذاتی در داخل یک جامعه، یک تغییر چگالی در طول یک دوره زمانی خاص بر اساس میانگین و انحراف استاندارد محاسبه می‌شود. کارهای مشابه بر اساس تکنیک اصلی روش در جدول ۱۲ آمده است.

۵- نتیجه‌گیری و راهکارهای آتی

از آنجا که در بعضی از تکنیک‌های به‌دست آوردن جامعه‌ها در شبکه‌های پیچیده، جواب نهایی به ترتیب پردازش نودها بستگی دارد، برای به‌دست آوردن یک جواب بهینه مناسب در این مقاله روشی بدون پارامتر برای شناسایی جامعه‌ها در شبکه‌های پیچیده ارائه شده که از معیار عدم تشابه نودها برای تجزیه گراف به زیرگراف‌های مناسبی استفاده می‌کند که نودهای این زیرگراف‌ها دارای بیشترین تشابه هستند. سپس در فرایند ادغام، این زیرگراف‌ها بر اساس افزایش ماژولاریتی با زیرگراف‌های همسایه ترکیب می‌شوند. نتایج اعمال این روش بر روی انواع مختلف

جدول ۷: درصدی از نودها که در گراف با ۶۴ نود با تغییر Z_{out} درست خوشه‌بندی شده‌اند.

الگوریتم	Z_{out}								
	۸	۷	۶	۵	۴	۳	۲	۱	۰
OSLOM	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪
Louvain	۳۲٪	۳۹٪	۳۹٪	۳۹٪	۳۹٪	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪
Local community detection	۲۵٪	۲۵٪	۲۵٪	۲۵٪	۲۵٪	۲۵٪	۸۴٪	۸۴٪	۸۴٪
MST&Modularity	۲۵٪	۲۵٪	۲۵٪	۲۵٪	۲۵٪	۲۵٪	۲۵٪	۲۵٪	۵۰٪
MST&MM	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪	۱۰۰٪

جدول ۸: درصدی از نودها که در گراف با ۱۲۸ نود با تغییر Z_{out} درست خوشه‌بندی شده‌اند.

الگوریتم	Z_{out}								
	۸	۷	۶	۵	۴	۳	۲	۱	۰
OSLOM	۵۳٪	۵۶٪	۵۷٪	۷۴٪	۷۴٪	۸۰٪	۸۰٪	۱۰۰٪	۱۰۰٪
Louvain	۱۲٪	۱۴٪	۱۲٪	۱۴٪	۴۰٪	۸۰٪	۸۰٪	۱۰۰٪	۱۰۰٪
Local community detection	۲۲٪	۲۵٪	۵۵٪	۷۳٪	۷۰٪	۷۶٪	۷۵٪	۱۰۰٪	۱۰۰٪
MST&Modularity	۲۵٪	۲۵٪	۲۵٪	۲۵٪	۲۵٪	۲۵٪	۲۵٪	۵۰٪	۱۰۰٪
MST&MM	۶۰٪	۶۰٪	۶۰٪	۷۴٪	۷۷٪	۸۰٪	۸۰٪	۱۰۰٪	۱۰۰٪

جدول ۹: مقدار پارامتر NMI با تغییر Z_{out} در گراف با ۳۲ نود.

الگوریتم	Z_{out}			
	۳	۲	۱	۰
OSLOM	۱	۱	۱	۱
Louvain	۱	۱	۱	۱
Local community detection	۰٫۵۹۱	۰٫۵۱۲	۱	۱
MST&Modularity	۰٫۵۸	۱	۱	۱
MST&MM	۱	۱	۱	۱

گراف‌های ساختگی و گراف‌های واقعی نشان می‌دهند که این روش در مقایسه با روش‌های دیگر شناسایی جامعه‌ها، کارایی بالایی دارد. این روش به دلیل عدم وابستگی به ترتیب پردازش نودها و همچنین عدم نیاز به تعیین پارامترهای خاص می‌تواند به‌عنوان یک روش مؤثر و مناسب برای شناسایی جامعه‌ها در شبکه‌های پیچیده مورد استفاده قرار گیرد. در آینده می‌توان به بهبود و بهینه‌سازی این روش و همچنین اعمال آن به شبکه‌های بزرگ‌تر و پیچیده‌تر پرداخت. در کارهای آتی با توسعه روش پیشنهادی تلاش خواهیم کرد تا چندین شکل از جوامع مانند جوامع همپوشانی، چندنمایه و جاسازی‌شده را در هر دو حالت همگن و ناهمگن شناسایی کنیم.

مراجع

[1] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-worldl networks," *Nature*, vol. 393, pp. 440-442, 1998.

[2] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509-512, 15 Oct. 1999.

[3] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proc. Natl. Acad. Sci.*, vol. 99, no. 12, pp. 7821-7826, 2002.

[4] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, Article ID: 026113, 2004.

[5] M. E. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167-256, 2003.

[6] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabási, "Hierarchical organization of modularity in metabolic networks," *Science*, vol. 297, no. 5586, pp. 1551-1555, 2002.

[7] D. M. Wilkinson and B. A. Huberman, "A method for finding communities of related genes," in *Proc. Natl. Acad. Sci.* vol. 101, *suppl* 1, pp. 5241-5248, 2004.

[8] R. Guimera and L. A. N. Amaral, "Functional cartography of complex metabolic networks," *Nature* vol. 433, no. 7028, pp. 895-900, 2005.

[9] G. W. Flake, S. Lawrence, C. L. Giles, and F. M. Coetzee, "Self-organization and identification of web communities," *Computer*, vol. 35, no. 3, pp. 66-70, Mar. 2002.

[10] Y. Dourisboure, F. Geraci, and M. Pellegrini, "Extraction and classification of dense communities in the web," in *Proc. of the 16th Int. Conf. on World Wide Web,* pp. 461-470, Banff, Canada, 8-12 May 2007.

[11] A. Perianes-Rodríguez, C. Olmeda-Gómez, and F. Moya-Anegón, "Detecting, identifying and visualizing research groups in co-authorship networks," *Scientometrics*, vol. 82, no. 2, pp. 307-319, 2010.

[12] B. He, Y. Ding, J. Tang, V. Reguramalingam, and J. Bollen, "Mining diversity subgraph in multidisciplinary scientific collaboration networks, a meso perspective," *J. Informetrics*, vol. 7, no. 1, pp. 117-128, Jan. 2013.

[13] M. A. Porter, J. P. Onnela, and P. J. Mucha, "Communities in networks," *Notices of the American Mathematical Society.* vol. 56, no. 9, pp. 1082-1097, 2009.

[14] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, no. 3-5, pp. 75-174, Feb. 2010.

[15] M. E. J. Newman, "Communities, modules and large-scale structure in networks," *Nature Physics*, vol. 8, pp. 25-31, 2012.

[16] M. Coscia, F. Giannotti, and D. Pedreschi, "A classification for community discovery methods in complex networks," *Statistical Analysis and Data Mining*, vol. 4, no. 5, pp. 512-546, Oct. 2011.

[17] Z. Shi, Y. Liu, and J. Liang, "PSO-based community detection in complex networks," in *Proc. 2nd Int. Symp. on Knowledge Acquisition and Modeling, KAM'09,* vol. 3, pp. 114-119, Wuhan, China, 30 Nov.-1 Dec. 2009.

[18] M. E. J. Newman, "Detecting community structure in networks," *the European Physical J. B-Condensed Matter and Complex Systems*, vol. 38, no. 2, pp. 321-330, 2004.

[19] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 70, no. 6, Article ID: 066111, 2004.

[20] C. Shi, Y. Wang, B. Wu, and C. Zhong, "A new genetic algorithm for community detection," *Complex Sciences*, vol. 535, Article ID: 122259, 2009.

[21] C. Pizzuti, "GA-Net: a genetic algorithm for community detection in social networks," in *Proc. of the 10th Parallel Problem Solving from Nature-PPSN X*, pp. 1081-1090, Dortmund, Germany, 13-17 Sept. 2008.

[22] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical Review E*, vol. 74, no. 3, Article ID: 36104, 2006.

جدول ۱۰: مقدار پارامتر NMI با تغییر Z_{out} در گراف با ۶۴ نود.

الگوریتم	Z_{out}								
	۸	۷	۶	۵	۴	۳	۲	۱	۰
OSLOM	۱	۱	۱	۱	۱	۱	۱	۱	۱
Louvain	۰٫۴۸۲	۰٫۶۰۵	۰٫۶۰۵	۰٫۶۰۵	۰٫۶۰۵	۱	۱	۱	۱
Local community detection	۰	۰	۰	۰	۰	۰	۰٫۸۲۷	۰٫۸۲۷	۰٫۸۲۷
MST&Modularity	۰٫۵۵۷	۰٫۵۵۷	۰٫۵۵۷	۰٫۵۵۷	۰٫۵۵۷	۰٫۵۵۷	۰٫۵۵۷	۰٫۸۳۵۷	۱
MST&MM	۱	۱	۱	۱	۱	۱	۱	۱	۱

جدول ۱۱: مقدار پارامتر NMI با تغییر Z_{out} در گراف با ۱۲۸ نود.

الگوریتم	Z_{out}								
	۸	۷	۶	۵	۴	۳	۲	۱	۰
OSLOM	۰٫۷۰	۰٫۷۰	۰٫۷۶	۰٫۸۰	۰٫۸۰	۰٫۸۱	۰٫۸۳	۱	۱
Louvain	۰٫۵۲	۰٫۴۶	۰٫۵۲	۰٫۴۶	۰٫۴۱	۱	۱	۱	۱
Local community detection	۰٫۵۵	۰٫۵۹	۰٫۶۵	۰٫۷۳	۰٫۸۰	۰٫۷۸	۰٫۸۰	۱	۱
MST&Modularity	۰	۰	۰	۰	۰	۰٫۵۵	۰٫۵۵	۰٫۸۲	۱
MST&MM	۰٫۸۱	۰٫۸۱	۰٫۸۱	۰٫۸۷	۰٫۸۷	۰٫۸۴۸	۰٫۸۴۸	۱	۱

جدول ۱۲: کارهای مشابه.

تکنیک مورد استفاده	الگوریتم
Betweenness centrality	Girvan and Newman (۲۰۰۲) [۳]
Nodes similarity	Donetti, *et al.* (۲۰۰۴) [۳۲]
Density-based link clustering	iDBLINK (۲۰۱۶) [۳۳]
Density-based	IncOrder- (۲۰۱۴) [۳۴]
Structural similarity measure	SCAN (۲۰۰۷) [۳۵]
Parallel Louvain method	PLMR (۲۰۱۶) [۳۶]
Label propagation	NI-LPA (۲۰۲۰) [۳۷]
Label propagation with information entropy	Naiyue Chen (۲۰۱۷) [۳۸]
Partitioning approach	PMEP (۲۰۱۶) [۳۹]
Graph-based label propagation	GLPA (۲۰۱۹) [۴۰]
Label propagation	PLPAC (۲۰۱۸) [۴۱]
Louvain method	Hao Lu (۲۰۱۵) [۴۲]
Modularity optimization	MOKP (۲۰۲۰) [۴۳]
Modularity	MDSTA (۲۰۱۹) [۴۴]
Distance dynamics model	Tingqin He (۴۵) [۲۰۱۸]

[23] J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection," *Phys. Rev.,* vol. E 74, no. 1, Article ID: 016110, 2006.

[24] U. Brandes, *et al.*, "On finding graph clusterings with maximum modularity," in *Proc. of the 33rd Int. Workshop on Graph-Theoretic Concepts in Computer Science,* pp. 121-132, Dornburg, Germany, 21-23 Jun. 2007.

[25] P. J. Bickel and A. Chen, "A nonparametric view of network models and Newman-Girvan and other modularities," *Proc. Natl. Acad. Sci.*, vol. 106, pp. 21068-21073, 15 Dec. 2009.

[26] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech.,* vol. 2008, Article ID: P10008, Oct. 2008.

[27] B. Saouda and A. Moussaoui, "Community detection in networks based on minimum spanning tree and modularity," *Physica A*, vol. 460, pp. 230-234, 15 Oct. 2016.

[28] D. Lusseau, "The emergent properties of a dolphin social network," in *Proc. R. Soc. London. Ser. B: Biol. Sci., Suppl. 2*, vol. 270, pp. S86-S188, 7 Nov. 2003.

[29] J. Eustace, X. Wang, and Y. Cui, "Community detection using local neighborhood in complex networks," *Physica A.*, vol. 436, pp. 665-677, 15 Oct. 2015.

[30] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato, "Finding statistically significant communities in networks," *PLoS One*, vol. 6, no. 4, Article ID: e18961, 2011.

[31] A. Lancichinetti and S. Fortunato, "Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities," *Phys. Rev.*, vol. E80, no. 1, Article ID: 016118, Jul. 2009.

[32] F. Meng, *et al.*, "Incremental density-based link clustering algorithm for community detection in dynamic networks," *Mathematical Problems in Engineering*, vol. 2016, no. 1, Article ID: 1873504, 11 pp., Jan. 2016.

[33] K. Nath, R. Shanmugam, and V. Varadaranjan, "ma-CODE: a multi-phase approach on community detection in evolving networks," *Information Sciences*, vol. 569, pp. 326-343, Aug. 2021.

[34] X. Xu, N. Yuruk, Z. Feng, and T. A. Schweiger, "Scan: a structural clustering algorithm for networks," in *Proc. of the 13th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, *KDD'07,* pp. 824-833, San Jose, CA, USA, 12-15 Aug. 2007.

[35] I. B. El Kouni, W. Karoui, and L. B. Romdhane, "Node importance based label propagation algorithm for overlapping community detection in networks," *Expert Systems with Applications*, vol. 162, Article ID: 113020, 30 Dec. 2020.

[36] N. Chen, Y. Liu, H. Chen, and J. Cheng, "Detecting communities in social networks using label propagation with information entropy," *Physica A: Statistical Mechanics and its Applications*, vol. 471, pp. 788-798, 1 Apr. 2017.

[37] G. Yang, W. Zheng, C. Che, and W. Wang, "Graph-based label propagation algorithm for community detection," *International J. of Machine Learning and Cybernetics*, vol. 11, pp. 1319-1329, 2020.

[38] B. Saoud and A. Moussaoui, "Community detection in networks based on minimum spanning tree and modularity," *Physica A: Statistical Mechanics and its Applications*, vol. 460, pp. 230-234, 15 Oct. 2016.

[39] J. Wu, X. Li, L. Jiao, X. Wang, and B. Sun, "Minimum spanning trees for community detection," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 9, pp. 2265-2277, 1 May 2013.

[40] J. Zhu, B. Chen, and Y. Zeng, "Community detection based on modularity and k-plexes," *Information Sciences*, vol. 513, pp. 127-142, Mar. 2020.

[41] X. Zhou, K. Yang, Y. Xie, C. Yang, and T. Huang, "A novel modularity-based discrete state transition algorithm for community detection in networks," *Neurocomputing*, vol. 334, pp. 89-99, 21 Mar. 2019.

[42] J. Xie, M. Chen, and B. K. Szymanski, "LabelrankT: incremental community detection in dynamic networks via label propagation," in *Proc. of the Workshop on Dynamic Networks Management and Mining*, pp. 25-32, New York, NY, USA, 22-27 Jun. 2013.

[43] H. Shen, X. Cheng, K. Cai, and M. B. Hu, "Detect overlapping and hierarchical community structure in networks," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 8, pp. 1706-1712, 15 Apr. 2009.

[44] C. L. Staudt and H. Meyerhenke, "Engineering parallel algorithms for community detection in massive networks," *IEEE Trans. on Parallel and Distributed Systems*, vol. 27, no. 1, pp. 171-184, Jan. 2015.

[45] K. Nath, S. Roy, and S. Nandi, "InOvIn: a fuzzy-rough approach for detecting overlapping communities with intrinsic structures in evolving networks," *Applied Soft Computing*, vol. 89, Article ID: 106096, Apr. 2020.

سندی بهادری تحصیلات مقاطع کارشناسی و کارشناسی ارشد و دکتری خود در رشته سیستم‌های نرم افزاری در سال‌های ۱۳۸۴ و ۱۳۸۷ و ۱۳۹۷ از دانشگاه بوعلی همدان، و دانشگاه آزاد اسلامی به پایان رسانده است و هم‌اکنون استادیار دانشکده مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد ایلام می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: تحلیل شبکه‌های پیچیده، محاسبات نرم و کاربردهای آن در توسعه نرم‌افزارهای بزرگ.

مریم نورائی آباده تحصیلات مقاطع کارشناسی و کارشناسی ارشد و دکتری خود در رشته سیستم‌های نرم افزاری در سال‌های ۱۳۸۴ و ۱۳۸۷ و ۱۳۹۴ از دانشگاه اصفهان، و دانشگاه آزاد اسلامی واحد علوم و تحقیقات به پایان رسانده است و هم‌اکنون استادیار دانشکده مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد آبادان می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: آزمون نرم‌افزار، تحلیل شبکه های پیچیده و ارتقاء توسعه سیستم‌های نرم‌افزاری با هوش مصنوعی.