



Semnan University

Journal of Modeling in Engineering

Journal homepage: <https://modelling.semnan.ac.ir/>

ISSN: 2783-2538



Research Article

A Hybrid Method for Community Detection Based on User Interactions, Topology and Frequent Pattern Mining

Somaye Sayari^a, Ali Harounabadi^{a,*} , Touraj Baniroostam^a

^a Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran.

PAPER INFO

Paper history:

Received: 04 February 2023

Revised: 29 May 2023

Accepted: 17 June 2023

Keywords:

User Interactions,
Community Detection,
Frequent Pattern Mining,
Local Clustering Coefficient,
Social Networks.

ABSTRACT

In recent years, community detection in social networks has become one of the most important research areas. One of the ways to community detection is to use interactions between users. There are different types of interactions in social networks, which, if used together with network topology, improve the precision of community identification. In this paper, a new method based on the combination of user interactions and network topology is proposed to community detection. In the community formation stage, the effective nodes are identified based on eigenvector centrality, and the primary communities around these nodes are formed based on frequent pattern mining. In the community expansion phase, small communities expand using modularity and the degree of interactions among users. To calculate the degree of interaction between users, a new measure based on the local clustering coefficient and interactions between common neighbors is proposed, which improves the accuracy of the degree of user interactions. Analysis of Higgs Twitter and Flickr datasets utilizing internal density metric, NMI and Omega demonstrates that the proposed method outperforms the other five community detection methods.

DOI: <https://doi.org/10.22075/jme.2023.29816.2402>

© 2023 Published by Semnan University Press.

This is an open access article under the CC-BY 4.0 license. (<https://creativecommons.org/licenses/by/4.0/>)

* Corresponding author.

E-mail address: a.harounabadi@iauctb.ac.ir

How to cite this article:

Sayari, S., Harounabadi, A., & Baniroostam, T. (2023). A hybrid method for community detection based on user interactions, topology and frequent pattern mining. *Journal of Modeling in Engineering*, 21(75), 129-145. doi: 10.22075/jme.2023.29816.2402

روشی ترکیبی برای شناسایی جوامع مبتنی بر تعاملات کاربران، توپولوژی شبکه و کاوش الگوی تکرار شونده

سمیه سیاری^۱، علی هارون آبادی^{۲*} و تورج بنی رستم^۳

اطلاعات مقاله	چکیده
دریافت مقاله: ۱۴۰۱/۱۱/۱۵	<p>در سال‌های اخیر، شناسایی جوامع در شبکه‌های اجتماعی به یکی از مهم‌ترین حوزه‌های تحقیقاتی تبدیل شده است. اکثر روش‌های تشخیص جامعه از اطلاعات توپولوژیکی شبکه استفاده می‌کنند. در حالی که انواع مختلفی از تعاملات در شبکه‌های اجتماعی وجود دارد که چنانچه با توپولوژی شبکه ترکیب شود باعث بهبود دقت در شناسایی جوامع می‌شود. در این مقاله، روشی ترکیبی برای شناسایی جوامع، مبتنی بر توپولوژی شبکه، درجه تعامل بهبود یافته کاربران و کاوش الگوی تکرار شونده بر روی تعاملات کاربران پیشنهاد می‌شود. جوامع اولیه، بر اساس مرکزیت بردار ویژه و کاوش الگوی تکرار شونده، حول گره‌های اثرگذار شکل می‌گیرند. جوامع شکل گرفته، مبتنی بر ماژولاریتی و درجه تعامل بهبود یافته کاربران گسترش می‌یابند. در اغلب روش‌ها، تعاملات مستقیم دو کاربر و تعاملات آن‌ها با همسایگان مشترک برای محاسبه درجه تعامل دو کاربر در نظر گرفته می‌شود. در نظر گرفتن تعاملات بین همسایگان مشترک، دقت درجه تعاملات کاربران را بهبود می‌بخشد. در مقاله جاری، برای محاسبه درجه تعامل بین کاربران، معیاری بهبود یافته مبتنی بر ضریب خوشه‌بندی محلی و تعاملات بین همسایگان مشترک ارائه می‌شود. نتایج ارزیابی روی دو مجموعه داده هیگز توپو و فلیکر با استفاده از شاخص‌های NMI، امگا و چگالی داخلی نشان می‌دهد که روش پیشنهادی در مقایسه با پنج روش شناسایی جامعه دیگر عملکرد بهتری دارد.</p>
بازنگری مقاله: ۱۴۰۲/۰۳/۰۸	
پذیرش مقاله: ۱۴۰۲/۰۳/۲۷	
<p>واژگان کلیدی: تعاملات کاربران، شناسایی جوامع، کاوش الگوی تکرار شونده، ضریب خوشه‌بندی محلی، شبکه‌های اجتماعی.</p>	

DOI: <https://doi.org/10.22075/jme.2023.29816.2402>

© 2023 Published by Semnan University Press.

This is an open access article under the CC-BY 4.0 license. (<https://creativecommons.org/licenses/by/4.0/>)

در شبکه‌های اجتماعی است [۲] که به درک و کشف ویژگی‌های مشترک اعضای جامعه کمک می‌کند [۳]. دسته‌بندی افراد در گروه‌های مختلف و بررسی رفتار آن‌ها، تحلیل افراد مؤثر شبکه، یافتن مجموعه‌ای از کاربران با علائق مشابه و ارائه توصیه‌های مناسب به آن‌ها، از مزایای شناسایی جامعه در شبکه‌های اجتماعی است [۴]. اکثر روش‌های تشخیص جامعه، تنها از ارتباطات مبتنی بر توپولوژی گراف، مانند روابط دنبال‌کننده- دنبال‌شونده

۱- مقدمه^۱

در سال‌های اخیر، شبکه‌های اجتماعی محبوبیت زیادی در بین کاربران پیدا کرده‌اند. افراد از طریق این شبکه‌ها می‌توانند دوستان زیادی پیدا کنند، تفکراتشان را به اشتراک بگذارند و یا به تبادل و ترویج ایده‌ها، عقاید و سبک زندگی خود بپردازند. تجزیه و تحلیل داده‌های موجود در این شبکه‌ها می‌تواند منجر به کشف اطلاعات پنهان و روابط ناشناخته شود [۱]. شناسایی جوامع از مهم‌ترین موضوعات

* پست الکترونیک نویسنده مسئول: a.harounabadi@iauctb.ac.ir

۱. دانشجوی دکتری، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، واحد تهران مرکزی، دانشگاه آزاد اسلامی

۲. استادیار، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، واحد تهران مرکزی، دانشگاه آزاد اسلامی

۳. استادیار، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، واحد تهران مرکزی، دانشگاه آزاد اسلامی

استناد به این مقاله:

سیاری، س.، هارون آبادی، علی، و بنی رستم، تورج. (۱۴۰۲). روشی ترکیبی برای شناسایی جوامع مبتنی بر تعاملات کاربران، توپولوژی شبکه و کاوش الگوی تکرار شونده.

مدل سازی در مهندسی، ۲۱(۷۵)، ۱۲۹-۱۴۵. doi: 10.22075/jme.2023.29816.2402

بیشتر باشد؛ احتمال پیوند دو کاربر نیز افزایش می‌یابد [۲۴]. ایده ما در این تحقیق، در نظر گرفتن تعاملات بین همسایگان مشترک برای قرار گرفتن دو کاربر، در یک جامعه مشترک است. به این ترتیب، معیاری بهبود یافته مبتنی بر ضریب خوشه‌بندی محلی برای محاسبه درجه تعامل بین دو کاربر ارائه می‌شود. این معیار، علاوه بر در نظر گرفتن تعامل مستقیم بین دو کاربر و تعامل با همسایگان مشترک، از تعاملات بین همسایگان مشترک نیز بهره می‌برد. از این‌رو، در این مطالعه، روشی ترکیبی برای شناسایی جوامع در شبکه‌های اجتماعی ارائه می‌شود؛ که علاوه بر توپولوژی شبکه از درجه تعامل بهبود یافته کاربران، به‌منظور بهبود کیفیت و دقت جوامع بهره می‌برد. هدف در این روش کشف گره‌های مؤثر در شبکه بر اساس مرکزیت بردار ویژه و تشکیل شبکه اولیه از افرادی است که بیشترین تعامل را با آن‌ها دارند. گره‌های باقی‌مانده، بر اساس معیار ماژولاریتی و معیار پیشنهاد شده درجه تعامل کاربران به جوامع اولیه اختصاص می‌یابند. نوآوری‌های اصلی ما در این مطالعه به‌صورت زیر است:

- پیشنهاد روشی ترکیبی برای شناسایی جوامع که علاوه بر توپولوژی شبکه، از کاوش الگوی تکرارشونده و یک معیار بهبود یافته برای محاسبه درجه تعامل کاربران استفاده می‌کند و منجر به بهبود کیفیت و دقت جوامع می‌شود.
- پیشنهاد معیاری بهبود یافته مبتنی بر ضریب خوشه‌بندی محلی و تعاملات بین همسایگان مشترک برای محاسبه درجه تعامل بین کاربران.
- به‌کارگیری معیاری ترکیبی مبتنی بر چگالی درونی جوامع به‌منظور ارزیابی جوامع کشف شده.

ساختار مقاله حاضر به شرح زیر سازماندهی شده است: بخش دوم، به تشریح کارهای مرتبط با روش‌های موجود در شناسایی جوامع می‌پردازد. بخش سوم، چارچوب روش پیشنهادی و جزئیات آن را بیان می‌کند. بخش چهارم، ارزیابی کار و نتایج به‌دست آمده از تجزیه و تحلیل روش پیشنهادی را ارائه می‌کند. بخش پنجم، به نتیجه‌گیری و پیشنهاد کارهای آتی می‌پردازد.

۲- کارهای مرتبط

با توجه به اهمیت شناسایی جوامع در شبکه‌های اجتماعی،

استفاده می‌کنند و تعاملات کاربران را در نظر نمی‌گیرند [۵ و ۶]. ممکن است بین دو کاربر رابطه دوستی برقرار باشد؛ ولی تعامل کمی با هم داشته باشند [۷]. همچنین ممکن است بین دو کاربر، تعامل اتفاق افتاده باشد؛ درحالی‌که ارتباط دوستی مستقیم نداشته باشند [۸]. تعاملات اجتماعی، نتیجه پیوند میان دو یا چند نفر و جریان اطلاعات میان آن‌ها است [۹]. برای هر کاربر، مجموعه اقدامات مجازی تعریف شده است که به انتخاب خود می‌تواند آن‌ها را انجام دهد. تعاملات بین کاربران با به‌کارگیری این اقدامات شکل می‌گیرند [۱۰]. از آن جمله می‌توان به ارسال پیام، فوتوگ‌ها، اشتراک در فیس‌بوک [۱۱] و پاسخ‌آ، ریتوییت^۳ و اشاره^۴ در توییتر [۱۲] اشاره کرد. در نظر گرفتن این تعاملات به همراه ویژگی‌های ساختاری شبکه باعث بهبود دقت و کسب نتایج معنادارتر در شناسایی جوامع می‌شود. در سال‌های اخیر روش‌هایی ارائه شده‌اند که با ترکیب تعاملات کاربران و توپولوژی شبکه اقدام به شناسایی جوامع کرده‌اند [۱۲، ۱۳]. علاوه بر این، برخی گره‌ها نقش مهم‌تری در انتشار اطلاعات دارند [۱۴]. یکی از روش‌های رایج در شناسایی گره‌های مؤثر، استفاده از مفهوم مرکزیت است [۱۵]. مرکزیت، اهمیت نسبی یک گره در گراف را اندازه‌گیری می‌کند [۱۶]. از انواع مختلف مرکزیت می‌توان به مرکزیت بینابینی [۱۷]، مرکزیت درجه [۱۸] و مرکزیت بردار ویژه [۱۹] اشاره کرد. استفاده از مرکزیت، تنها مبتنی بر ساختار گراف است و ویژگی‌ها، اقدامات و تعاملات گره‌ها را در نظر نمی‌گیرد. یکی دیگر از روش‌های رایج در شناسایی گره‌های مؤثر، کاوش الگوی تکرارشونده بر روی تعاملات و اقدامات کاربران است [۲۰، ۲۱، ۲۲]. الگوی تکرارشونده به الگویی از داده‌ها اشاره دارد که به‌طور متوالی در یک مجموعه داده خاص تکرار می‌شود [۲۳]. می‌توان با اعمال کاوش الگوی تکرارشونده بر روی کاربران، به گروه‌هایی از کاربران دست یافت که یکسری اقدامات و تعاملات یکسان را انجام می‌دهند [۱۳]. از طرفی، اکثر روش‌های مبتنی بر تعاملات کاربران، از دو فاکتور تعاملات مستقیم بین دو کاربر و تعاملات آن‌ها با همسایگان مشترکشان برای محاسبه درجه تعامل بین دو کاربر استفاده می‌کنند [۵، ۷]. این در حالی است که در روش‌های پیش‌بینی پیوند، هر چه روابط بین همسایگان مشترک

² Reply

³ Retweet

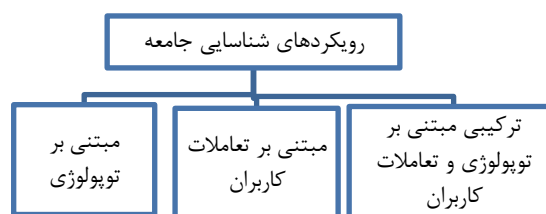
⁴ Reply

پیشنهاد دادند. این الگوریتم از معیار شباهت محلی مبتنی بر ضریب خوشه‌بندی برای شناسایی نزدیک‌ترین همسایگان هر گره استفاده می‌کند. سپس چندین مجموعه از اجزای متصل را با ترکیب جفت‌های مختلف گره به دست می‌آورد و جوامع اولیه تشکیل می‌شوند. در مرحله آخر، استراتژی ادغام جامعه، برای بهینه‌سازی ساختار جامعه اعمال می‌شود. [۳۷] الگوریتم تشخیص جامعه مبتنی بر درخت را در شبکه‌های اجتماعی پویا معرفی کرده است؛ که از دو ویژگی مهم یک شبکه اجتماعی بهره می‌برد: اتصال و نفوذ. این الگوریتم از یک ساختار درختی برای حفظ اطلاعات ساختارهای اجتماعی شبکه در حال تغییر استفاده می‌کند. در [۲] روشی برای تشخیص جامعه با استفاده از اطلاعات محلی و جهانی شبکه پیشنهاد می‌شود. ابتدا، معیارهای شباهت شناسایی می‌شوند و وزن‌های مناسب بر اساس اطلاعات شبکه محلی، به گره‌ها و پیوندهای شبکه اختصاص می‌یابد. در ادامه، معیارهای شباهت برای توسعه جامعه اولیه، بر اساس یک الگوریتم تصادفی بهبود یافته توسط وزن گره‌ها در نظر گرفته می‌شوند. در نهایت، جوامع اولیه ادغام می‌شوند و ساختار جامعه بهینه با استفاده از توابع ارزیابی محاسبه شده بر اساس اطلاعات محلی و جهانی شبکه انتخاب می‌شود.

۲-۲- رویکردهای شناسایی جوامع مبتنی بر تعاملات کاربران

تعاملات بین کاربران در شبکه‌های اجتماعی، زمینه استفاده در تحقیقات زیادی شده‌اند. دو و همکاران [۱۱] روشی را پیشنهاد دادند که در آن ابتدا درجه تعامل بین هر جفت از کاربران و درجه رفتار گروهی آن‌ها محاسبه می‌شود. سپس با استفاده از الگوریتم جمع‌شونده سلسله‌مراتبی بر روی گراف احتمال، جوامع استخراج می‌شوند. نیوین و همکاران [۳۸] یک الگوریتم شناسایی جامعه بر اساس اقدامات کاربران پیشنهاد دادند که شامل سه مرحله است: کشف رهبران، محاسبه وزن بین گره‌ها و اختصاص گره‌ها به جوامع. ابتدا با ایجاد ماتریس تأثیر از روی جدول اقدامات کاربران، رهبران شناسایی می‌شوند [۲۰]. سپس بر اساس تعداد اقدامات مشترک بین دو همسایه، شباهت بین آن‌ها محاسبه می‌شود و گره‌ها به جوامع اختصاص می‌یابند. لو و همکاران [۵] روشی مبتنی بر تعاملات مستقیم و غیرمستقیم کاربران که اثر آبشار گونه بر روابط افراد دارد، پیشنهاد دادند. در این روش ابتدا، تعاملات مستقیم و

مطالعات متعددی بر روی این موضوع متمرکز شده‌اند. بیشتر تحقیقات، افراز و خوشه‌بندی گراف [۲۵ و ۲۶]، تقسیم بین لبه‌ها [۲۷]، بهینه‌سازی مبتنی بر ماژولاریتی [۲۸] و



شکل ۱: دسته‌بندی کارهای مرتبط

[۲۹]، انتشار برچسب [۳۰] و فاکتورسازی ماتریس غیرمنفی [۳۱] بر روی توپولوژی شبکه را به کار گرفته‌اند. در ادامه به بررسی برخی مطالعات انجام شده در زمینه شناسایی جوامع می‌پردازیم. شکل (۱) دسته‌بندی کارهای مرتبط در این مقاله را نشان می‌دهد.

۲-۱- رویکردهای شناسایی جوامع مبتنی بر توپولوژی شبکه

ژاکوبی و کانواتی [۳۲] الگوریتم LICOD را پیشنهاد دادند که ابتدا گره‌های رهبر را در شبکه با استفاده از معیار مرکزیت شناسایی می‌کند. سپس لیست رهبران با گروه‌بندی آن‌ها بر اساس تعداد همسایگان مشترک، کاهش می‌یابد. درجه عضویت گره‌های باقی‌مانده در شبکه نسبت به هر جامعه محاسبه می‌شود و هر گره، به جوامع با رتبه بالاتر تعلق می‌گیرد. لی و همکاران [۳۳] الگوریتم انتشار برچسب مرحله‌به‌مرحله را پیشنهاد دادند که در آن برچسب‌ها بر اساس شباهت منتشر می‌شوند. شباهت هر جفت گره با استفاده از شباهت تخصیص منبع [۳۴] محاسبه می‌شود و برچسب هر گره با برچسب همسایه‌ای که بیشترین شباهت را دارد به‌روز می‌شود. اهاجم و همکاران [۱۸] رویکرد رهبر-جامعه پیشنهاد داده‌اند که ابتدا با استفاده از مرکزیت بردار ویژه، گره‌های رهبر را شناسایی می‌کند. سپس جامعه حول گره رهبر، بر اساس یک تابع شباهت، گسترش می‌یابد. آروی و برودکا [۳۵] روشی ترکیبی پیشنهاد دادند که در آن، گره‌های برتر مشترک در چند معیار مرکزیت، به‌عنوان مجموعه بذر برتر انتخاب می‌شوند. سپس بذرهایی که کمترین فاصله را با هم دارند در یک گروه قرار گرفته و جوامع اولیه شکل می‌گیرند. گره‌های نگاشت نشده به جوامع، به نزدیک‌ترین بذر منتسب می‌شوند. پن و همکاران در [۳۶] الگوریتم ASOCCA

[۱۲] با در نظر گرفتن ویژگی‌های دنبال‌کنندگان مشترک و دوستان مشترک از جنبه ساختاری، هشتگ‌ها از جنبه محتوایی و پاسخ‌ها، ریتوییت‌ها و اشارات از جنبه تعاملی، معیارهای شباهتی را ارائه دادند. سپس با استفاده از الگوریتم انتشار وابستگی به شناسایی جوامع پرداختند. موسوی و همکاران [۱۳] روشی مبتنی بر الگوی تکرار شونده و اقدامات کاربران در شبکه پیشنهاد داده‌اند. ابتدا با به‌کارگیری کاوش الگوی تکرار شونده بر روی جدول فعالیت کاربران، ترتیبی از کاربرانی به دست می‌آید که یکسری اقدامات مشابه را انجام داده‌اند. در صورت تأیید نحوه اتصال گروه‌های به‌دست آمده در گراف اجتماعی، جوامع اولیه شکل می‌گیرند. این اتصال می‌تواند به‌صورت مستقیم و یا اتصال با واسطه باشد. در نهایت، اگر یک گروه، در همسایگی تعداد مشخصی از اعضای یک جامعه قرار گیرد به آن جامعه تعلق می‌گیرد و جوامع گسترش می‌یابند.

جدول ۱ مقایسه کارهای مطالعه‌شده را نشان می‌دهد. اکثر روش‌های پیشین در زمینه شناسایی جوامع در شبکه‌های اجتماعی، تنها مبتنی بر یک جنبه از روابط بین کاربران هستند و استفاده هم‌زمان از توپولوژی شبکه و تعاملات کاربران را در نظر نمی‌گیرند. در روش‌های ترکیبی مطالعه شده نیز، شکل‌گیری اولیه جوامع و گسترش جوامع، تنها مبتنی بر یکی از جنبه‌ها هستند. در برخی کارها هم، تنها از یک نوع تعامل برای شناسایی جامعه استفاده شده است. در برخی دیگر نیز، برای محاسبه درجه تعامل دو کاربر، از تعاملات مستقیم آن‌ها و تعاملات آن‌ها با همسایگان مشترک استفاده شده است. در حالی که در نظر گرفتن تعاملات میان همسایگان مشترک، باعث افزایش دقت در محاسبه درجه تعامل کاربران می‌شود. از این‌رو، در این مقاله، روشی ترکیبی برای شناسایی جوامع پیشنهاد می‌شود؛ که هم در مرحله شکل‌گیری اولیه جوامع و هم در مرحله گسترش جوامع، به‌صورت هم‌زمان از توپولوژی شبکه و تعاملات کاربران بهره می‌برد. همچنین، در مرحله گسترش جوامع، معیاری مبتنی بر ضریب خوشه‌بندی محلی برای محاسبه درجه تعامل کاربران پیشنهاد می‌شود؛ که علاوه بر تعاملات مستقیم و تعاملات با همسایگان مشترک کاربران، تعاملات میان همسایگان مشترک را نیز در نظر می‌گیرد.

غیرمستقیم کاربران بر روی اشیا اجتماعی مانند پست و عکس جمع‌آوری می‌شود. سپس یک گراف رویداد، شامل روابط آبشاری بین تعاملات مستقیم و غیرمستقیم، به‌ازای هر شی اجتماعی ایجاد می‌شود. در هر گراف رویداد، خوشه‌هایی کوچک از گروه‌های کاربری تحت عنوان رویدادهای فرعی استخراج می‌شود. هر رویداد فرعی، به‌صورت یک گره در نظر گرفته می‌شود و یک سوپر گراف شکل می‌گیرد. در آخر یک روش شناسایی جامعه بر روی این سوپر گراف به کار گرفته می‌شود. هر جامعه نشان‌دهنده گروهی از کاربران است که به‌صورت پویا بر روی چندین شی اجتماعی تعامل دارند.

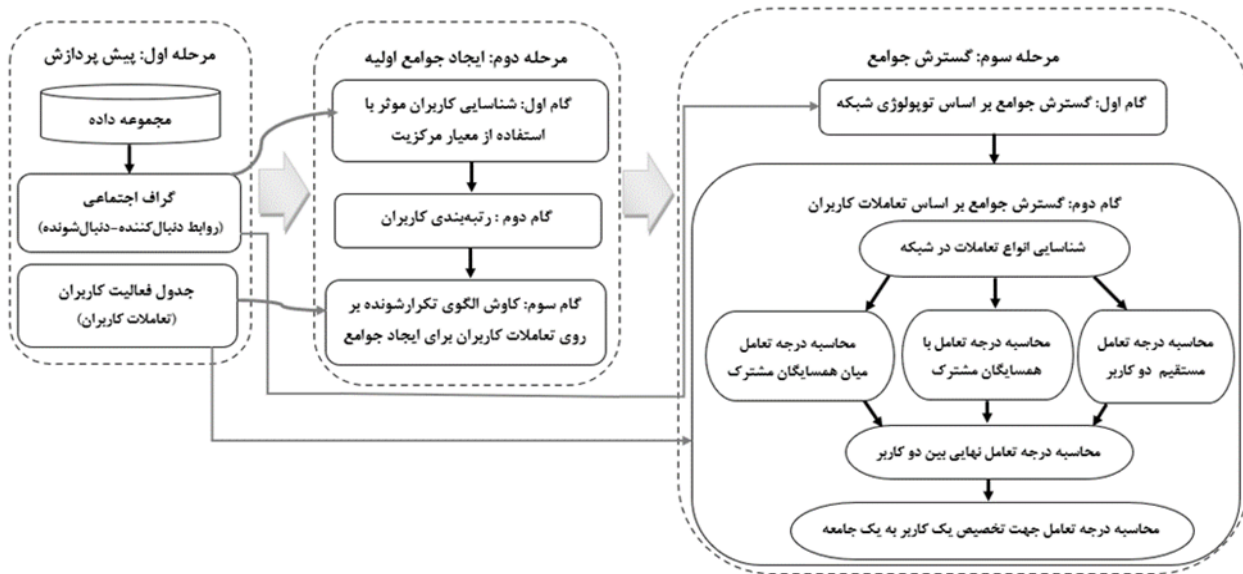
۲-۳- رویکردهای ترکیبی مبتنی بر توپولوژی و تعاملات کاربران

در سال‌های اخیر، با ترکیب توپولوژی شبکه و تعاملات کاربران کارهای زیادی در زمینه تجزیه و تحلیل شبکه‌های اجتماعی انجام شده است. احمد و الکرانی [۸] یک معیار شباهت بین دو کاربر، برای پیش‌بینی پیوند در شبکه‌های اجتماعی پیشنهاد دادند که تعاملات مبتنی بر ریتوییت، اشاره و پاسخ را با یک معیار شباهت مبتنی بر ساختار شبه محلی ترکیب می‌کند. در این روش شباهت دو کاربر متصل و شباهت دو کاربر غیرمتصل بر اساس قدرت اتصال و تعاملات میان کاربران محاسبه می‌شود.

یانگ و همکاران [۳۹] معیار شباهتی مبتنی بر ساختار محلی و فرکانس تعاملات کاربران برای پیش‌بینی پیوند معرفی کردند. آن‌ها معیار آدامیک-آدار را بر اساس فرکانس تعاملات میان کاربران توسعه داده‌اند. پیشنهاد آن‌ها این بوده که کاربران با تمام افراد موجود در لیست دوستانشان تماس ندارند؛ بنابراین روابط دوستی در شبکه‌های اجتماعی نسبت به گراف نمایش داده شده قوی‌تر یا ضعیف‌تر هستند. لذا، تعداد پیام‌های مبادله شده بین هر جفت کاربر به آدامیک-آدار افزوده شده است. لیم و داتا [۴۰] روشی پیشنهاد دادند که در آن ابتدا، مجموعه‌ای از افراد مشهور که علاقه‌مندی‌های یکسانی دارند، انتخاب می‌شوند. سپس دنبال‌کنندگان مشترک بین آن افراد مشهور استخراج می‌شوند. در پایان، با در نظر گرفتن پیوندهای تعاملی دوطرفه میان مجموعه دنبال‌کنندگان مشترک و به‌کارگیری خوشه‌بندی، جوامع استخراج می‌شوند. وتھی و همکاران در

جدول ۱: مقایسه کارهای مطالعه شده

رویکرد	متد	مرجع	توضیح	معايب
توپولوژی شبکه، رهبر-جامعه	LICOD	[۳۲]	شناسایی گره‌های مؤثر با استفاده از معیار مرکزیت و اختصاص گره‌ها به جوامع بر اساس درجه عضویت	هم در مرحله شناسایی گره‌ها و هم در مرحله گسترش جوامع مبتنی بر توپولوژی شبکه است.
توپولوژی	Stepping LPA-S	[۳۳]	استفاده از روش انتشار برچسب مرحله‌به‌مرحله بر اساس معیار شباهت تخصیص منبع	تنها مبتنی بر توپولوژی شبکه است.
توپولوژی، رهبر-جامعه	LCDA	[۱۸]	شناسایی گره‌های مؤثر با استفاده از معیار مرکزیت بردار ویژه و استفاده از شباهت ژاکارد در گسترش جوامع	هم در مرحله شناسایی گره‌های مؤثر و هم در مرحله گسترش جوامع مبتنی بر توپولوژی شبکه است.
توپولوژی	4-S model	[۳۵]	شناسایی گره‌های مؤثر با استفاده از معیار مرکزیت و گسترش جوامع با محاسبه ماتریس فاصله	تنها مبتنی بر توپولوژی شبکه است.
توپولوژی	ASOCCA	[۳۶]	شناسایی جوامع با استفاده از معیار شباهت مبتنی بر ضریب خوشه‌بندی محلی	تنها مبتنی بر توپولوژی شبکه است.
توپولوژی	TCD2	[۳۷]	استفاده از یک ساختار درختی برای حفظ اطلاعات ساختارهای اجتماعی شبکه در حال تغییر	تنها مبتنی بر توپولوژی شبکه است.
توپولوژی	LGICD	[۲]	استفاده از اطلاعات محلی و جهانی شبکه برای شناسایی جوامع	تنها مبتنی بر توپولوژی شبکه است.
تعاملات کاربران	UIBCD	[۱۱]	شناسایی جوامع بر اساس تعاملات مستقیم و تعاملات گروهی کاربران	تعاملات بین همسایگان مشترک کاربران در نظر گرفته نشده است.
تعاملات کاربران	ICDCA	[۵]	شناسایی جوامع بر اساس تعاملات آبخاری بین کاربران	این روش برای محاسبه تعاملات بین کاربران از روش [۱۱] استفاده کرده است.
اقدامات کاربران، رهبر-جامعه	LBCD	[۳۸]	شناسایی گره‌های مؤثر با تشکیل ماتریس تأثیر کاربران و گسترش جوامع بر اساس تعداد اقدامات مشترک بین دو همسایه	این روش تنها از اقدامات مشترک بین کاربران برای گسترش جوامع استفاده کرده است و تعاملات بین کاربران را در نظر نگرفته است.
ترکیبی	HICD	[۴۰]	استخراج دنبال‌کننده‌های مشترک افراد مشهور بر اساس عمل اشاره	تنها از یک نوع تعامل (عمل اشاره) برای شناسایی جوامع استفاده شده است.
ترکیبی	FAUCD	[۲۳]	شناسایی گره‌های مؤثر با استفاده از الگوی تکرار شونده و گسترش جامعه بر اساس k نزدیک‌ترین همسایه	این روش در مرحله شناسایی گره‌های مؤثر از اقدامات کاربران بهره برده است. اما گسترش جوامع تنها مبتنی بر توپولوژی شبکه است.
ترکیبی	LDA model	[۱۲]	شناسایی جوامع بر اساس معیار شباهت مبتنی بر تعاملات کاربران	تعداد تعاملات بین کاربران و همسایگان مشترکشان، در نظر گرفته نشده است.
ترکیبی	LPSUA	[۸]	شباهت مبتنی بر ساختار شبه محلی و تعاملات کاربران در توصیه به کاربران	در محاسبه تعاملات بین کاربران همه فاکتورها محاسبه نشده است.
ترکیبی	FRSIIM	[۳۹]	شباهت مبتنی بر ساختار محلی و تعاملات بین کاربران در توصیه به کاربران	در محاسبه تعاملات بین کاربران همه فاکتورها محاسبه نشده است.



شکل ۲: چارچوب کلی روش پیشنهادی

کاربران نیز شامل رکوردهایی نظیر کاربران تعامل کننده، نوع تعامل و تعداد تعاملات بین کاربران است. تعامل می‌تواند اشاره، ریتوییت و پاسخ در توییتر و یا نظرهای کاربران روی یک تصویر به اشتراک گذاشته شده در فلیکر باشد. هر پیوند در این جدول جهت‌دار است. در این مرحله، با توجه به لزوم حذف داده‌های نویزدار و نامناسب، کاربران و یا تعاملات نامناسب از مجموعه داده حذف می‌شوند [۵].

۳-۲- ایجاد جوامع اولیه

در مرحله دوم، ابتدا کاربران اثرگذار با استفاده از مرکزیت بردار ویژه شناسایی می‌شوند. سپس کاربرانی که با گره‌های با بالاترین مرکزیت بردار ویژه تعامل دارند، استخراج می‌شوند. به این ترتیب جوامع اولیه شکل می‌گیرند.

۳-۲-۱- مرکزیت گره‌ها

در این تحقیق، مرکزیت بردار ویژه برای شناسایی گره‌های مؤثر استفاده می‌شود. این مرکزیت، گسترش یافته مرکزیت درجه است. آن را می‌توان به عنوان تابعی بازگشتی از درجه ارتباطات و قدرت همسایگان یک گره تعریف کرد. یعنی یک گره مهم است؛ در صورتی که به گره‌های اثرگذار دیگر متصل باشد. مرکزیت بردار ویژه به عنوان معیاری برای توصیف میزان ارتباط یک فرد بر اساس روابط مستقیم و غیرمستقیم تعریف می‌شود؛ لذا برای یک گراف G با $|V|$ گره و $|E|$ یال، امتیاز رأس v می‌تواند به صورت مجموع امتیاز همسایگانش در نظر گرفته شود که به صورت رابطه (۱) نوشته می‌شود [۱۷].

$$X_v = \frac{1}{\lambda} \sum_{t \in N(v)} X_t = \frac{1}{\lambda} \sum_{t \in N(v)} a_{v,t} X_t \quad (1)$$

۳-روش پیشنهادی

همان‌طور که در شکل (۲) مشاهده می‌شود، روش پیشنهادی شامل سه مرحله اصلی است و هر مرحله گام‌های مختلفی دارد. در مرحله اول، با انجام پیش‌پردازش بر روی مجموعه داده، گراف اجتماعی و جدول تعاملات کاربران استخراج می‌شوند. در مرحله دوم، ابتدا با استفاده از معیار مرکزیت بردار ویژه، کاربران اثرگذار شناسایی می‌شوند. سپس با به‌کارگیری کاوش الگوی تکرار شونده، کاربرانی که با کاربران اثرگذار بیشترین تعامل را دارند، مشابه روش [۱۳] استخراج می‌شوند و جوامع اولیه شکل می‌گیرند. در مرحله سوم، جوامع اولیه گسترش می‌یابند. ابتدا، کاربران بر اساس معیار ماژولاریتی به جوامع منتسب می‌شوند. باتوجه به انواع مختلف تعاملات موجود در شبکه اجتماعی و بر اساس سه فاکتور تعاملات مستقیم بین دو کاربر، تعاملات با همسایگان مشترک و تعاملات بین همسایگان مشترک، درجه تعامل بین هر کاربر باقی‌مانده در خارج جامعه و هر کاربر داخل جامعه محاسبه می‌شود. سپس درجه تعامل نهایی هر کاربر خارجی نسبت به هر جامعه محاسبه می‌شود و بر این اساس به جامعه‌ای تعلق می‌گیرد. در جوامع حاصل امکان هم‌پوشانی وجود دارد. مراحل روش پیشنهادی به شرح زیر است.

۳-۱- پیش‌پردازش داده‌ها

ابتدا از مجموعه داده، گراف اجتماعی و جدول تعاملات کاربران استخراج می‌شوند. گراف اجتماعی شامل پیوندهای دنبال‌کننده-دنبال‌شونده بین کاربران است. جدول تعاملات

اجرای الگوریتم Eclat پارامترهای حداقل پشتیبان، حداقل ترکیب و حداکثر ترکیب مورد نیاز است. حداقل پشتیبان تعیین می‌کند، دنباله تعامل کنندگان حداقل با چند درصد از گره‌های اثرگذار باید تعامل داشته باشند.

۳-۳-۳- گسترش جوامع

در این مرحله، جوامع حاصل در دو گام گسترش می‌یابند و سایر گره‌های باقی‌مانده در شبکه به عضویت جوامع کوچک درمی‌آیند. در گام اول بر اساس معیار ماژولاریتی، کاربران به جوامع منتسب می‌شوند. در گام دوم، با پیشنهاد معیاری جدید مبتنی بر تعاملات کاربران جوامع گسترش می‌یابند.

۳-۳-۱- گسترش جوامع کوچک بر اساس توپولوژی شبکه به کارگیری معیار ماژولاریتی در شناسایی جوامع منجر به ایجاد جوامع با کیفیت‌تر که گره‌های آن ارتباطات بیشتری نسبت به سایر اعضای شبکه دارند، می‌شود [۴۴]. از این رو، ابتدا هر گره همسایه، با اعضای هر جامعه اولیه را به عنوان عضو آن جامعه در نظر می‌گیریم. سپس میزان تغییرات ماژولاریتی (ΔQ) را محاسبه می‌کنیم. اگر $\Delta Q > 0$ باشد، آن گره به جامعه کوچک منتسب می‌شود. این کار تا سطح L از همسایگی اعضای داخل جامعه، ادامه پیدا می‌کند. از آنجا که گراف اجتماعی ما در این روش، یک گراف جهت‌دار است؛ بنابراین از معیار ماژولاریتی Q_d [۴۵]، طبق رابطه (۳) استفاده شده است:

$$Q_d = \sum_c \left[\frac{L_c}{m} - \left(\frac{k_c^{in} k_c^{out}}{2m} \right)^2 \right] \quad (3)$$

به طوری که m تعداد یال‌های شبکه، L_c تعداد یال‌های داخل جامعه C ، $k_c^{in} k_c^{out}$ مجموع درجه‌های ورودی و درجه‌های خروجی گره‌های جامعه C را نشان می‌دهد. برای درک بهتر این گام، در سطح یک ($L = 1$)، دو جامعه اولیه $C_1 = \{1, 4\}$ و $C_2 = \{8, 12\}$ را در گراف اجتماعی شکل (۳) در نظر بگیرید. در سطح دو ($L = 2$)، تمامی گره‌هایی که به طور مستقیم با گره‌های اولیه در جوامع همسایگی دارند، بررسی می‌شوند. در صورتی که این گره‌ها باعث افزایش ماژولاریتی شوند؛ به جوامع ملحق می‌شوند. بنابراین، گره‌های $\{2, 3, 6, 7, 9, 10\}$ به جامعه C_1 و گره‌های $\{11, 15, 16, 18, 19\}$ به جامعه C_2 می‌پیوندند. در سطح سه ($L = 3$)، گره‌های $\{5, 13\}$ به جامعه C_1 متصل و گره‌های $\{14, 20\}$ به جامعه C_2 متصل می‌شوند.

به طوری که $N(v)$ مجموعه همسایگان گره v ، $a_{v,t} = 1$ اگر بین v و t یال باشد و در غیر این صورت $a_{v,t} = 0$ است. λ یک ثابت است. با کمی تغییر می‌توان رابطه (۱) را به صورت رابطه (۲) نوشت:

$$AX = \lambda X \quad (2)$$

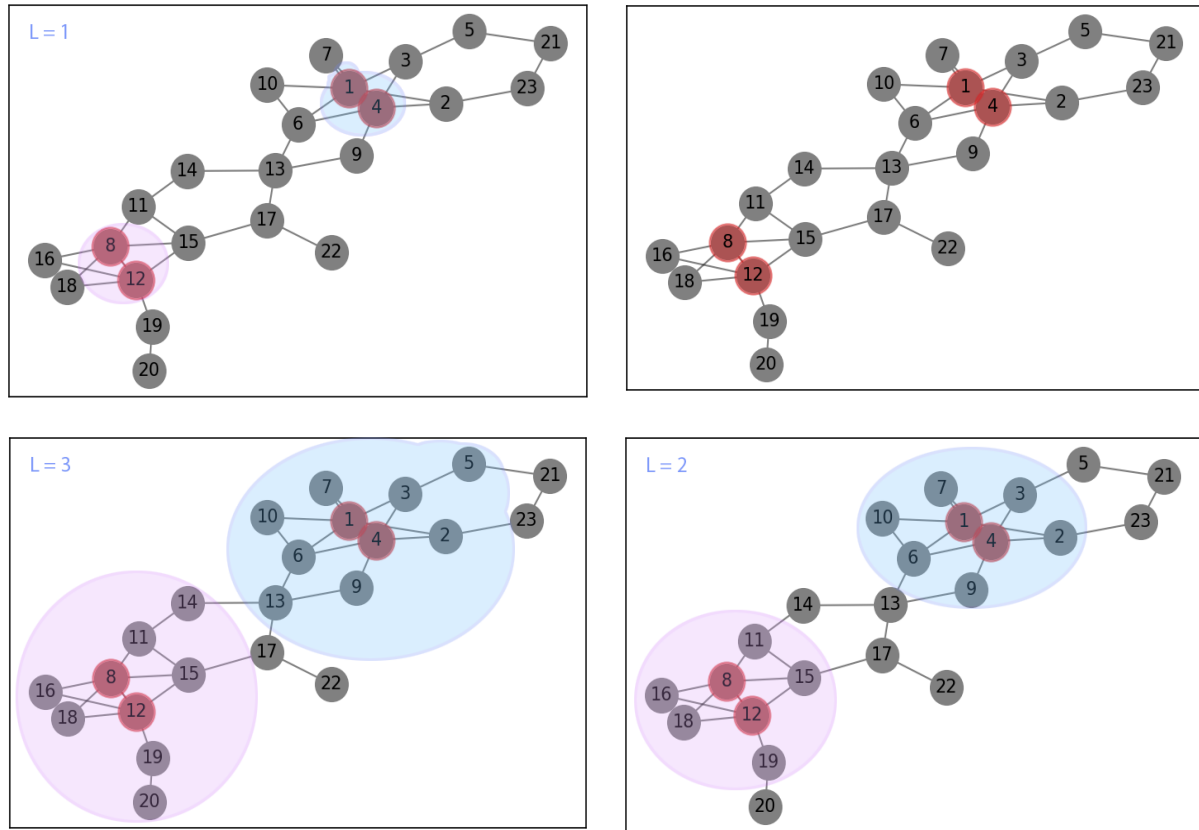
A ماتریس مجاورت گراف، X بردار ویژه و λ یک مقدار ویژه است. بردار ویژه حاصل از بزرگترین مقدار ویژه، امتیاز مرکزیت ویژه گره‌های گراف است [۵]. روش رایج برای محاسبه بردار ویژه روش تکرار توانی است. این روش پس از بیشترین حد تکرار یا زمانی که تغییر در مقادیر بردار ویژه بین دو تکرار، از یک حد آستانه کوچک‌تر باشد، متوقف می‌شود [۴۱]. با توجه به توضیحات مطرح شده، برای شناسایی گره‌های مؤثر، مرکزیت بردار ویژه تمامی کاربران در شبکه محاسبه می‌شود.

۳-۳-۲- رتبه‌بندی گره‌ها

کاربران بر اساس امتیاز مرکزیت بردار ویژه‌شان رتبه‌بندی می‌شوند. کاربرانی که امتیازشان از مقدار آستانه Th بیشتر باشد به عنوان کاربران اثرگذار انتخاب می‌شوند.

۳-۳-۳- کاوش الگوی تکرار شونده

در این گام، کاربرانی که با کاربران اثرگذار منتخب از گام قبل تعامل بالایی دارند، با کاوش الگوی تکرار شونده شناسایی می‌شوند و گروه‌های اولیه از کاربران ایجاد می‌شوند. چند جامعه اولیه با این روش شناسایی می‌شوند. از بین گروه‌های به دست آمده، مجموعه‌هایی که یال ارتباطی بین آن‌ها در گراف اجتماعی موجود باشد به عنوان جوامع اولیه انتخاب می‌شوند. برای این منظور، ابتدا از جدول تعاملات کاربران، کاربرانی که با گره‌های بالاترین مرکزیت بردار ویژه تعامل دارند، استخراج می‌شوند. لازم به ذکر است که گره‌های مؤثری که هیچ کاربری با آن‌ها تعامل نداشته یا تنها یک کاربر تعامل کننده دارند حذف می‌شوند. سپس الگوریتم کاوش الگوی تکرار شونده بر روی جدول تعاملات به دست آمده اجرا می‌شود. در این تحقیق، برای استخراج گروه‌های کاربرانی که بیشترین تعامل را با کاربران اثرگذار دارند از الگوریتم Eclat [۴۲] استفاده شده است. در الگوریتم Eclat، فرمت داده‌ها به صورت عمودی تغییر می‌کند و در مقایسه با الگوریتم Apriori [۴۳] باعث افزایش سرعت در یافتن الگوهای تکراری می‌شود. برای



شکل ۳: انتساب گره‌ها به جوامع بر اساس ماژولاریتی

۳-۲-۳-۲-درجه تعامل مستقیم

باتوجه به اینکه تعاملات بین دو کاربر جهت‌دار است؛ بنابراین تعداد کل تعاملات مستقیم از نوع t بین دو کاربر u و v از رابطه (۴) محاسبه می‌شود [۱۲]:

$$DI_d^t(u, v) = \frac{DI_d^t uv + DI_d^t vu}{Out_u^t + Out_v^t} \quad (4)$$

$DI_d^t uv$ تعداد دفعات تعامل نوع t که مستقیماً u با v داشته است و $DI_d^t vu$ تعداد دفعات تعامل نوع t که مستقیماً v با u داشته است. Out_u^t تعداد دفعات تعامل نوع t که کاربر u با همه کاربران داشته است و Out_v^t تعداد دفعات تعامل نوع t که کاربر v با همه کاربران داشته است.

۳-۲-۳-۳-درجه تعامل با همسایگان مشترک

این معیار بر اساس فرکانس تعاملات گره u و v با همسایگان مشترکشان محاسبه می‌شود؛ بنابراین حتی اگر u و v خودشان با هم تعامل مستقیمی نداشته باشند، می‌توان میزان تعامل غیرمستقیم آن‌ها را از رابطه (۵) محاسبه کرد [۳۷].

۳-۲-۳-۳-گسترش جوامع بر اساس درجه تعامل کاربران

در این گام، معیاری جدید برای محاسبه درجه تعامل بین دو کاربر پیشنهاد می‌شود. سپس با استفاده از معیار پیشنهادی، درجه تعامل بین کاربرانی که به جامعه‌ای تعلق ندارند، و کاربران داخل جامعه محاسبه می‌شود. در پایان، اگر درجه تعامل نهایی یک کاربر آزاد، نسبت به کاربران داخل یک جامعه از یک مقدار آستانه بیشتر شود؛ آن کاربر عضو آن جامعه می‌شود.

۳-۲-۳-۱-انواع تعاملات

باتوجه به اینکه در شبکه‌های اجتماعی، انواع مختلفی از تعاملات بین کاربران وجود دارد، لذا برای تعیین درجه تعامل بین دو کاربر، لازم است تا همه انواع تعاملات در نظر گرفته شوند [۱۰]. فرض کنید $t = \{1, 2, \dots, n\}$ ، تعداد n نوع تعامل را در یک شبکه اجتماعی نشان می‌دهد. آنگاه برای هر نوع تعامل t ، درجه تعامل نهایی بین دو کاربر، یعنی DI_{Int} با در نظر گرفتن فاکتورهای زیر محاسبه می‌شود.

تعاملات نوع t در بین شبکه کاربرانی که هم u و هم v با آنها تعامل داشته‌اند. $|\Gamma(u) \cap \Gamma(v)|$ تعداد کاربرانی است که هم کاربر u و هم کاربر v با آنها تعامل داشتند.

۳-۲-۵-درجه تعامل نهایی بین دو کاربر
باتوجه به رابطه‌های (۴)، (۵) و (۸) درجه تعامل بین دو کاربر u و v به‌ازای یک نوع تعامل t به‌صورت رابطه (۹) محاسبه می‌شود:

$$DI_{int}^t(u, v) = \alpha_1 * DI_d^t(u, v) + \alpha_2 * DI_{in}^t(u, v) + \alpha_3 * DI_{cn}^t(u, v) \quad (9)$$

α_1 ، α_2 و α_3 پارامترهای تعادلی هستند و مقدار آنها بین ۰ و ۱ است، به طوری که $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

درجه تعامل نهایی برای n نوع تعاملات موجود در شبکه از رابطه (۱۰) محاسبه می‌شود:

$$DI_{Intfinal}(u, v) = \frac{n}{\sum_{t=1}^n \frac{1}{DI_{int}^t(u, v)}} \quad (10)$$

۳-۲-۶-اختصاص یک کاربر به یک جامعه
با محاسبه درجه تعامل نهایی بین دو کاربر، درجه تعامل یک کاربر خارج از جامعه با کاربران داخل جامعه از رابطه (۱۱) محاسبه می‌شود:

$$DI(u, C_i) = \frac{\sum_{v \in C_i} DI_{Intfinal}(u, v)}{n_{C_i}} \quad (11)$$

که $\sum_{v \in C_i} DI_{Intfinal}(u, v)$ مجموع درجه تعامل نهایی گره u با تمامی گره‌های داخل جامعه C_i و n_{C_i} تعداد اعضای داخل جامعه C_i را نشان می‌دهد. در صورتی که، مقدار $DI(u, C_i)$ از حد آستانه θ بالاتر باشد؛ $(DI(u, C_i) \geq \theta)$ گره u به جامعه C_i تعلق می‌گیرد.

۴-ارزیابی

۴-۱-مجموعه داده‌ها و تنظیم پارامترها

دو مجموعه داده، برای ارزیابی روش پیشنهادی، در این مقاله به‌کارگرفته شده است: مجموعه داده هیگز توپیتور و مجموعه داده فلیکر.

۴-۱-۱-مجموعه داده هیگز توپیتور

مجموعه داده هیگز توپیتور شامل شبکه روابط اجتماعی دنبال‌کننده-دنبال شونده بین کاربران، شبکه‌های ریتوییت، اشاره و پاسخ کاربران و همچنین جدول فعالیت کاربران است. شناسه‌های کاربری در این مجموعه داده، به‌صورت

$$DI_{in}^t(u, v) = \sum_{z \in \Gamma(u) \cap \Gamma(v)} \frac{DI_d^t(u, z) + DI_d^t(v, z)}{Out_z^t + In_z^t} \quad (5)$$

Z کاربرانی هستند که هم u و هم v با آنها تعامل داشته‌اند (که از رابطه (۴) قابل محاسبه است)، Out_z تعداد کل تعاملات نوع t که Z با همه کاربران داشته و In_z تعداد کل تعاملات نوع t کاربران، با گره Z است.

۳-۲-۳-درجه تعامل میان همسایگان مشترک

همسایه مشترک، پل ارتباطی بین دو کاربر است [۳۶]. منطقی است که فرض کنیم هر چه ارتباطات بین همسایگان مشترک جفت کاربر، قوی‌تر باشد احتمال بیشتری دارد تا آن دو کاربر در یک جامعه قرار گیرند [۴۴]. در [۱۲] تنها تعداد همسایگان مشترک، برای محاسبه فرکانس تعاملات کاربران در نظر گرفته شده است. در [۱۱] و [۵] نیز تنها دو فاکتور تعاملات بین دو کاربر و تعاملات با همسایگان مشترک، برای محاسبه درجه تعامل دو کاربر در نظر گرفته شده است. از طرفی، یکی از روش‌های محاسبه میزان ارتباطات بین همسایگان یک گره، استفاده از مفهوم ضریب خوشه‌بندی است. ضریب خوشه‌بندی محلی برای یک گراف وزن دار از رابطه (۶) به دست می‌آید [۴۶]:

$$wclus_{(i)} = \frac{\sum_{z \neq z' \in N(i)} w_{zz'}}{k_i(k_i - 1)} \quad (6)$$

که $\sum_{z \neq z' \in N(i)} w_{zz'}$ مجموع وزن یال‌های بین همسایگان گره i و k_i درجه گره i است.

در این مقاله، با استفاده از ضریب خوشه‌بندی محلی بین همسایگان مشترک دو گره و میانگین تعداد تعاملات میان همسایگان مشترک رابطه جدیدی برای محاسبه درجه تعامل میان همسایگان مشترک دو کاربر، پیشنهاد شده است:

$$DI_{cn}^t(u, v) = \frac{\sum_{z \neq z' \in \Gamma(u) \cap \Gamma(v)} DI_d^t(z, z')}{|\Gamma(u) \cap \Gamma(v)| * (|\Gamma(u) \cap \Gamma(v)| - 1)} \quad (7)$$

که می‌توانیم آن را به‌صورت ساده به شکل رابطه (۸) بنویسیم:

$$DI_{cn}^t(u, v) = \frac{\sum_{z \neq z' \in \Gamma(u) \cap \Gamma(v)} DI_d^t(z, z')}{|\Gamma(u) \cap \Gamma(v)|^2 (|\Gamma(u) \cap \Gamma(v)| - 1)} \quad (8)$$

به‌طوری‌که $\sum_{z \neq z' \in \Gamma(u) \cap \Gamma(v)} DI_d^t(z, z')$ مجموع تعداد

۲ عضو می‌توانند داشته باشند. در گام ۳-۳-۱، $L=2$ در نظر گرفته شده است. به این ترتیب، گسترش جامعه با استفاده از معیار ماژولاریتی، تا دو سطح از همسایگی، ادامه پیدا می‌کند α_1 ، α_2 و α_3 پارامترهای تعادلی، برای تعیین وزن درجه تعامل مستقیم، درجه تعامل با همسایگان و درجه تعامل بین همسایگان مشترک هستند که به ترتیب ۰.۴، ۰.۳ و ۰.۳ در نظر گرفته شده است؛ به طوری که $\alpha_1 + \alpha_2 + \alpha_3 = 1$. حد آستانه درجه تعامل یک کاربر خارج از جامعه با کاربران داخل جامعه $(DI(u, C_i))$ را نشان می‌دهد که مقدار آن ۰.۵ در نظر گرفته شده است. مقادیر پارامترهای استفاده شده این تحقیق، در جدول ۳ آمده است.

جدول ۲: لیست رتبه‌بندی شده بر اساس مرکزیت بردار ویژه در هیگزتویتر

گره	مقدار مرکزیت بردار ویژه
۱۰۴۷۷۴	۰.۱۰۴۰۸۴۵۶۴
۵۳۸۴۸	۰.۰۹۵۸۰۳۲۰۶
۵۳۸۰۱	۰.۰۹۳۵۱۳۱۷۶
۱۰۴۷۸۵	۰.۰۹۱۹۷۳۱۸۹
۵۳۷۴۱	۰.۰۹۰۵۲۷۵۱۹
۵۳۸۷۶	۰.۰۸۷۲۵۷۳۶۸

جدول ۳: مقادیر پارامترها

مقدار	معادله/الگوریتم	پارامتر
۰.۰۳	مرکزیت بردار ویژه	Th
۰.۰۲	الگوریتم Eclat	min_support
۲	الگوریتم Eclat	min_combination
۲	الگوریتم Eclat	max_combination
۲	سطح همسایگی	L
۰.۴	معادله ۹	α_1
۰.۳	معادله ۹	α_2
۰.۳	معادله ۹	α_3

۴-۲- معیارهای ارزیابی

در این مقاله، معیارهای ارزیابی به کار گرفته شده شامل معیارهای ارزیابی درونی و معیارهای ارزیابی بیرونی هستند. معیارهای ارزیابی درونی، کیفیت جوامع را با توجه به تعریف جامعه بررسی می‌کنند. معیارهای ارزیابی بیرونی، دقت جوامع حاصل را در مقایسه با داده‌های حقیقی بررسی می‌کنند.

ناشناس درآمده‌اند و برای تمامی شبکه‌ها از همین شناسه کاربری استفاده شده است. گراف روابط اجتماعی در این مجموعه داده شامل ۴۵۶۶۲۶ گره و ۱۴۸۵۵۸۴۲ یال و جدول فعالیت کاربران شامل ۵۶۳۰۶۹ رکورد است [۴۶]. این مجموعه داده در [۴۷] در دسترس است.

۴-۱- مجموعه داده فلیکر

مجموعه داده فلیکر، شامل روابط دوستی، رابطه کاربر به گروه عکس‌ها و نظرات کاربران درباره عکس‌ها است. این مجموعه دارای ۲۰۳۷۵۳۸ کاربر و ۲۱۹۰۹۸۶۶۰ رابطه دوستی، ۶۵۵۹۱۷ گروه، ۱۲۶۲۹۷۸ عکس و ۱۴۹۱۳۱۶۴ نظر بر روی عکس است. اسامی کاربران به id نگاشت شده است [۴۸]. مجموعه داده فلیکر در [۴۹] در دسترس است. در مرحله پیش‌پردازش، جدول فعالیت کاربران از روی نظرات روی تصویرها آماده شد. برای این کار، ابتدا رکوردهای کاربران فاقد شناسه کاربری و شناسه عکس‌های به اشتراک گذاشته شده، حذف شدند. با توجه به اینکه تنها تعامل در این مجموعه داده، نظرات کاربران است؛ بنابراین تمامی نظرات حذف شدند. به این ترتیب فقط، شناسه کاربرانی که بر روی یک عکس نظری ثبت کرده بودند، همراه با شناسه مالک تصویر باقی ماند. سپس تعداد دفعاتی که یک کاربر، بر روی تمامی تصویرهای به اشتراک گذاشته شده توسط یک کاربر دیگر، پیام ارسال کرده است، شمارش شد. این مقدار، به عنوان تعداد تعاملات آن کاربر با کاربر مالک عکس‌ها در نظر گرفته شد. به منظور ارزیابی، مجموعه داده‌ها به دو بخش غیرهم‌پوشان تقسیم شدند و ۷۰٪ داده‌ها برای فاز آموزش و ۳۰٪ برای فاز تست استفاده شدند. در فاز آموزش، با استفاده از الگوریتم k -fold، اعتبارسنجی متقابل انجام شد. مقدار $k=10$ در نظر گرفته شد. مقادیر پارامترهای مورد استفاده در روش پیشنهادی، پس از انجام اعتبارسنجی به صورت زیر تنظیم شد: حد آستانه Th برای انتخاب گره‌های اثرگذار در این روش ۰.۰۳ در نظر گرفته شد. بنابراین، گره‌هایی که امتیاز مرکزیت بردار ویژه‌شان از ۰.۰۳ بالاتر است، به عنوان گره‌های اثرگذار انتخاب شدند. جدول ۲، شش گره با بالاترین امتیاز مرکزیت بردار ویژه در مجموعه داده هیگزتویتر را نشان می‌دهد. پارامترهای حداقل پشتیبان، حداقل ترکیب و حداکثر ترکیب در الگوریتم Eclat به ترتیب ۰.۰۲، ۲ و ۲ در نظر گرفته شدند. با این فرض، گروه‌های کاربرانی که حداقل با دو گره اثرگذار یکسان تعامل دارند، استخراج شدند که این گروه‌ها حداکثر

۴-۲-۱- معیارهای ارزیابی درونی

برای ارزیابی روش پیشنهادی، بدون استفاده از داده‌های حقیقی، جوامع شناسایی شده، از نظر تعداد گره‌های داخل جوامع و چگالی یال‌های داخل جوامع ارزیابی شده‌اند. چگالی داخلی یا درونی جامعه C ، کسری از یال‌های درون جامعه به بیشترین تعداد ممکن از یال‌های آن جامعه است و از رابطه (۱۲) محاسبه می‌شود [۳۵].

$$\delta_{int}(C) = \frac{m_c}{\frac{n_c(n_c-1)}{2}} \quad (12)$$

به طوری که m_c تعداد یال‌های داخل جامعه و n_c تعداد گره‌های داخل جامعه است. از آنجا که جوامع کشف شده هم مبتنی بر ساختار و هم تعاملات کاربران هستند؛ بنابراین چگالی داخلی یک جامعه، از دو جنبه روابط ساختاری و روابط تعاملی محاسبه شدند. در این تحقیق، برای محاسبه چگالی داخلی ساختاری و چگالی داخلی تعاملی به وسیله یک معیار، $Score_{SI-int}$ به صورت زیر پیشنهاد شده است:

$$Score_{SI-int} = \frac{(1+\beta^2)(\delta_{S-int} \cdot \delta_{I-int})}{\beta^2 \cdot (\delta_{S-int} + \delta_{I-int})} \quad (13)$$

جایی که δ_{S-int} چگالی داخلی ساختاری و δ_{I-int} چگالی داخلی تعاملی است. $\beta \in (0, \infty)$ پارامتری برای تنظیم چگالی داخلی ساختاری و چگالی داخلی تعاملی است. وقتی $\beta = 1$ بین دو مقدار δ_{S-int} و δ_{I-int} تعادل وجود دارد. اگر $\beta > 1$ ، تاکید بیشتری بر روی δ_{I-int} قرار می‌گیرد.

۴-۲-۲- معیارهای ارزیابی بیرونی

برای ارزیابی مجموعه داده‌ها بر اساس داده‌های حقیقی، دو معیار ارزیابی برای اندازه‌گیری میزان دقت شناسایی جوامع در نظر گرفته شده است: اطلاعات متقابل نرمال هم‌پوشان (NMI_{op}) و شاخص امگا (OI). اطلاعات متقابل نرمال شده هم‌پوشان، گسترش یافته اطلاعات متقابل نرمال شده (NMI) است و در خوشه‌بندی‌های هم‌پوشان مورد استفاده قرار می‌گیرد. اطلاعات متقابل نرمال شده هم‌پوشان دو خوشه‌بندی C و C' به صورت رابطه (۱۴) محاسبه می‌شود [۵۰]:

$$NMI_{op}(X|Y) = \frac{1 - [H(X|Y) + H(Y|X)]}{2} \quad (14)$$

که $NMI_{op}(X|Y) \in [0, 1]$ و $NMI_{op}(X|Y) = 1$

بیشترین انطباق بین خوشه‌بندی‌ها است. $H(X|Y)$ آنتروپی شرطی است. آنتروپی $H(X)$ و آنتروپی شرطی $H(X|Y)$ را می‌توان از توزیع احتمال X و Y مشتق کرد.

شاخص امگا، نسخه هم‌پوشانی شاخص رند تنظیم شده است [۵۰]. این شاخص، جفت گره‌هایی که درون تعداد مشابهی از جوامع تقسیم‌بندی شده‌اند را ارزیابی می‌کند. باتوجه به مجموعه جوامع C شناسایی شده با روش تشخیص جامعه و مجموعه جوامع C' ، شاخص امگا به صورت رابطه (۱۵) محاسبه می‌شود [۵۱]:

$$\omega(C, C') = \frac{\omega_u(C, C') - \omega_e(C, C')}{1 - \omega_e(C, C')} \quad (15)$$

که در آن ω_u شاخص امگای تنظیم نشده است که نشان‌دهنده کسری از جفت‌هایی است که با هم در تعداد مشابهی از جوامع در C و C' رخ می‌دهند. ω_e شاخص امگا مورد انتظار است که نشان‌دهنده مقدار مورد انتظار همان کسر در مدل نول است. برای محاسبه ω_u و ω_e به [۵۱] مراجعه کنید.

۴-۳- روش‌های مقایسه شده

برای ارزیابی، کیفیت جوامع حاصل از روش $UITFCD^5$ ، با پنج روش دیگر مقایسه شده است. روش‌های مقایسه‌ای بر اساس رویکردهای بررسی شده در بخش ۲ انتخاب شده‌اند. این روش‌ها با استفاده از زبان پایتون پیاده‌سازی شده‌اند و یا دارای کدهای آماده در کتابخانه‌های پایتون هستند: (i) $NI-Louvain$ [۵۲]: در این روش، گره‌ها مبتنی بر ساختار، کاهش تراکم شبکه و به حداکثر رساندن ماژولاریتی به جوامع اضافه می‌شوند. (ii) $SLPA$ [۲۵]: این روش مبتنی بر انتشار برجسب گوینده-شنونده، جوامع هم‌پوشان را شناسایی می‌کند. (iii) $LCDA$ [۱۸]: که در آن گره‌های مؤثر با معیار مرکزیت بردار ویژه شناسایی می‌شوند. جوامع، با معیار ژاکارد گسترش می‌یابند. (iv) $FAUCD$ [۱۳]: شکل‌گیری جوامع بر اساس کاوش الگوی تکرار شونده بر روی اقدامات کاربران صورت می‌گیرد. سپس بر اساس توپولوژی شبکه جوامع گسترش می‌یابند. (v) $ICDCA$ [۵]: در این روش از تعاملات مستقیم و غیرمستقیم کاربران که اثر آبخارگونه در روابط افراد دارند، برای شناسایی جوامع استفاده می‌شود.

علت انتخاب روش $SLPA$ ، $NI-Louvain$ و $ICDCA$ برای

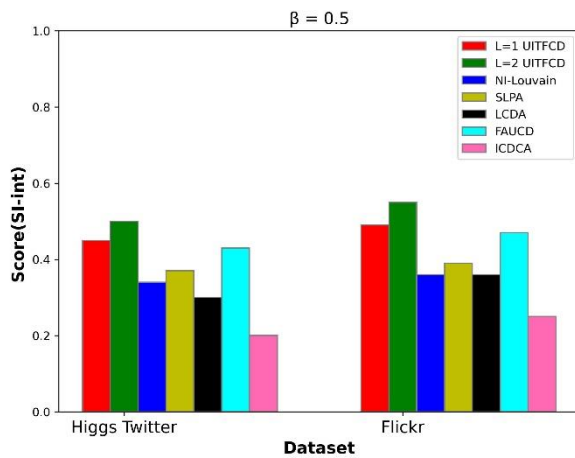
⁵ User Interaction Topology Frequent Pattern Community Detection

ترکیبی است که در آن تشکیل اولیه جوامع بر اساس اقدامات کاربران انجام می‌شود و گسترش جوامع تنها بر اساس توپولوژی شبکه است. در حالی که *UITFCD* ما در هر دو مرحله، از تعاملات کاربران در کنار توپولوژی گراف بهره برده است.

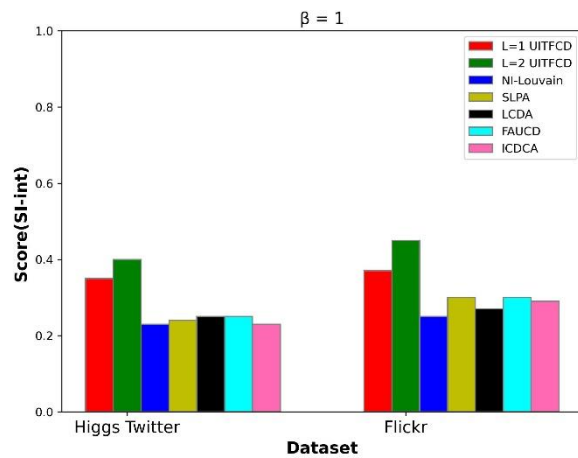
۴-۳-۱- ارزیابی بر اساس چگالی داخلی

شکل (۴) عملکرد روش پیشنهادی را با روش‌های مقایسه‌ای بر اساس معیار $Score_{SI-int}$ پیشنهاد شده در بخش ۴-۲-۱ نشان می‌دهد.

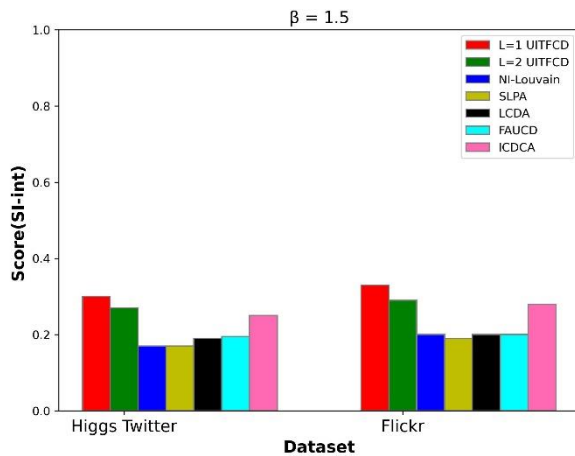
مقایسه با روش پیشنهادی، این است که بررسی شود، چگونه ترکیب تعاملات کاربران و توپولوژی شبکه با یکدیگر بر کشف جوامع مؤثر است و کیفیت جوامع افزایش می‌یابد. در *NI-Louvain* و *SLPA* فقط توپولوژی گراف در شناسایی جوامع در نظر گرفته می‌شود. *ICDCA* نیز، تنها از تعاملات کاربران در شناسایی جوامع استفاده می‌کند. *LCDA* و *FAUCD* مانند *UITFCD* مبتنی بر رویکرد رهبر-دنبال‌کننده هستند. تفاوت در این است که در *LCDA* هم شناسایی گره‌های رهبر و هم گسترش جوامع مبتنی بر توپولوژی شبکه است. *FAUCD* یک روش



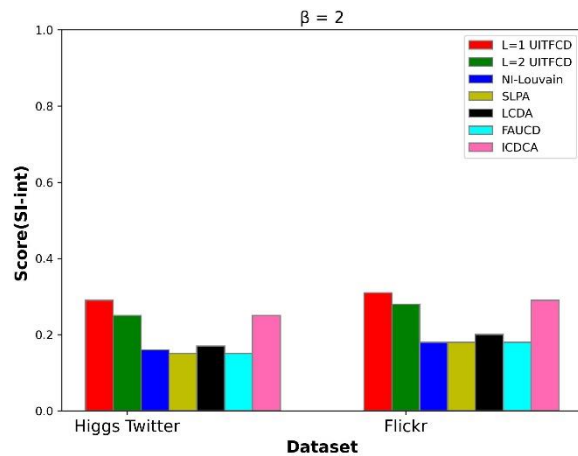
(a)



(b)



(c)



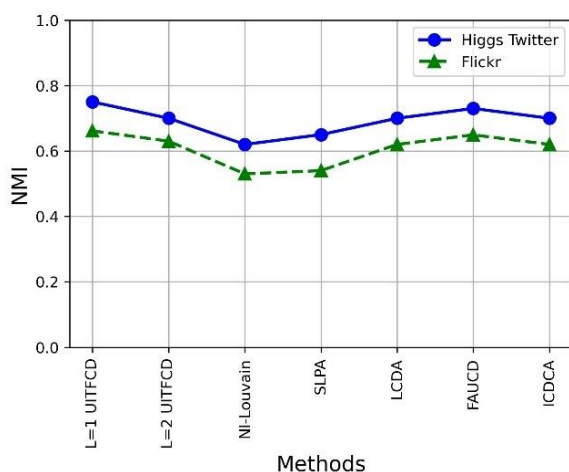
(d)

شکل ۴: ارزیابی بر اساس $Score_{SI-int}$ در مجموعه داده هیگز توئیتر و فلیکر

بر ساختار و مبتنی بر تعامل را بررسی کند. همان‌طور که مشاهده می‌شود، روش ما در همه موارد به $Score_{SI-int}$ بالاتری می‌رسد. برای مقدار $\beta = 0.5$ و مقدار $\beta = 1$ (که تأکید بیشتر بر روی چگالی ساختاری داخلی است)،

هر میله، یک روش تشخیص جامعه را نشان می‌دهد. دو میله اول، مربوط به روش پیشنهادی با مقادیر مختلف L است. برای ارزیابی، β روی مقادیر 0.5، 1، 1.5 و 2 تنظیم شده است؛ تا موارد مختلف مبادله بین چگالی داخلی مبتنی

شکل (۵) مقادیر NMI و Ω Index را در $UITFCD$ و روش‌های دیگر نشان می‌دهد. از نظر هر دو معیار، روش پیشنهادی در هر دو مجموعه داده، در سطح یک همسایگی نسبت به سطح دو، عملکرد بهتری دارد. این نشان می‌دهد، میزان انطباق جوامع حاصل در سطح یک، با جوامع حقیقی بیشتر است. همچنین هم در NMI و هم در Ω Index عملکرد روش ما در سطح یک نسبت به سایر روش‌های دیگر بهتر است. در هر دو مجموعه داده، در سطح دو همسایگی، مقدار NMI ، در روش $UITFCD$ ، از NI - $Louvain$ ، $SLPA$ ، $Louvain$ و $FAUCD$ در مقایسه با $FAUCD$ و $ICDCA$ در هر دو معیار تقریباً یکسان است.



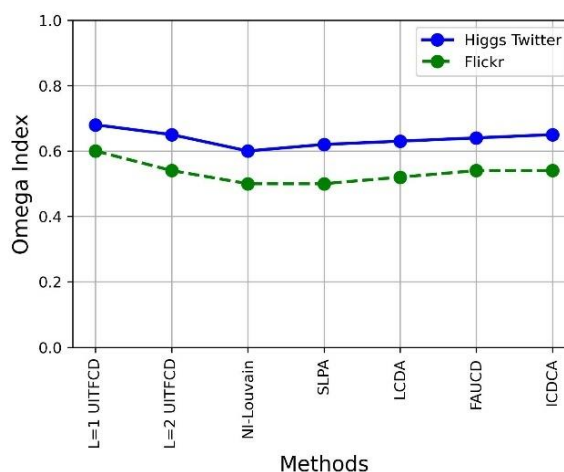
شکل ۵: ارزیابی بر اساس NMI و شاخص امگا در مجموعه داده‌های هیگز توییت و فلیکر

به این ترتیب، میزان دقت در محاسبه درجه تعامل بین کاربران افزایش می‌یابد. ارزیابی روش پیشنهادی بر روی دو مجموعه داده انجام شده است. نتایج با چند روش دیگر مبتنی بر شناسایی جوامع مقایسه شدند. روش $UITFCD$ ، با در نظر گرفتن چند شاخص عملکرد بهتری نسبت به روش‌های دیگر دارد.

پیشنهاد می‌شود در تحقیقات آینده، به منظور ترکیب ایده تحلیل شبکه و داده کاوی از الگوریتم‌های مبتنی بر استخراج الگوهای محلی، نظیر الگوریتم MinerLC برای شکل‌گیری جوامع کوچک محلی استفاده کرد. همچنین می‌توان، روش‌های مبتنی بر محتوا مانند روش توصیف‌گرا را به منظور افزایش تفسیرپذیری جوامع، با روش پیشنهادی ترکیب کرد.

$UITFCD$ در سطح دو عملکرد بهتری نسبت به سطح یک دارد. اما برای مقادیر $\beta = 1.5$ و $\beta = 2$ (که تأکید بر روی چگالی تعاملی است) در سطح یک، عملکرد بهتری مشاهده می‌شود. دلیل آن این است که، بیشتر کاربران در سطح یک، از طریق تعاملات به جوامع منتسب شده‌اند. از طرفی، روش $ICDCA$ فقط مبتنی بر تعاملات مستقیم و تعاملات گروهی کاربران است، در حالی که در $UITFCD$ تعاملات بین همسایگان مشترک نیز در نظر گرفته می‌شود؛ بنابراین در مقادیر $\beta > 1$ نیز روش ما بهتر عمل می‌کند.

۴-۳-۲- ارزیابی بر اساس NMI و Ω Index



۵- نتیجه‌گیری و کارهای آینده

در این مقاله، روشی جدید مبتنی بر ترکیب توپولوژی شبکه و تعاملات بین کاربران، برای شناسایی جوامع پیشنهاد شد. بیشتر روش‌های شناسایی جوامع، فقط یکی از جنبه‌های توپولوژی یا تعاملات کاربران را در نظر گرفته‌اند. در حالی که ترکیب این دو جنبه باعث افزایش دقت در شناسایی جوامع می‌شود. از طرفی، در بیشتر روش‌ها، برای محاسبه درجه تعامل بین دو کاربر، تنها درجه تعامل مستقیم بین آن‌ها و درجه تعاملات آن‌ها با همسایگان مشترکشان در نظر گرفته شده است.

در این تحقیق، معیاری جدید، مبتنی بر تعاملات کاربران پیشنهاد شد که علاوه بر دو فاکتور اشاره شده، تعاملات بین همسایگان مشترک دو کاربر را در نظر می‌گیرد.

مراجع

- [1] Li, XiaoMing, Guangquan Xu, and Minghu Tang, "Community detection for multi-layer social network based on local random walk", *Journal of Visual Communication and Image Representation*. 57 (2018): 91-98.
- [2] Dabaghi-Zarandi, Fahimeh, and Parsa KamaliPour, "Community detection in complex network based on an improved random algorithm using local and global network information", *Journal of Network and Computer Applications*. 206 (2022): 103492.
- [3] Aggarwal, Charu C, *An introduction to social network data analytics*, Springer, 2011.
- [4] Moscato, Vincenzo, and Giancarlo Sperli, "A survey about community detection over On-line Social and Heterogeneous Information Networks", *Knowledge-Based Systems*. 224 (2021): 107112.
- [5] Luo, Linbo, Kexin Liu, Bin Guo, and Jianfeng Ma, "User interaction-oriented community detection based on cascading analysis", *Information Sciences*. 510 (2020): 70-88.
- [6] گلکار، ایمان و مرجان کاندی. "ارائه مدلی برای تخمین میزان برون‌گرایی اعضای شبکه اجتماعی با استفاده از اطلاعات ساختار گراف". نشریه مدل‌سازی در مهندسی ۱۳، ۴۳، (۱۳۹۴): ۹۱-۱۰۶. (inPersian).
- [7] Wilson, Christo, Alessandra Sala, Krishna P. N. Puttaswamy, and Ben Y. Zhao, "Beyond Social Graphs", *ACM Transactions on the Web*. 6 (2012): 1-31.
- [8] Ahmed, Cherry, and Abeer ElKorany, "Enhancing link prediction in Twitter using semantic user attributes". Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, 1155-1161, 2015.
- [9] O'Riordan, Sheila, Joseph Feller, and Tadhg Nagle, "A categorisation framework for a feature-level analysis of social network sites", *Journal of Decision Systems*. 25 (2016): 244-262.
- [10] Moosavi, Seyed Ahmad, and Mehrdad Jalali, "Community detection in online social networks using actions of users". 2014 Iranian Conference on Intelligent Systems (ICIS), 1-7, (IEEE), 2014.
- [11] Dev, Himel, Mohammed Eunos Ali, and Tanzima Hashem, "User interaction based community detection in online social networks". Database Systems for Advanced Applications: 19th International Conference, DASFAA 2014, Bali, Indonesia, April 21-24, 2014. Proceedings, Part II 19, 296-310, (Springer), 2014.
- [12] Vathi, Eleni, Georgios Siolas, Andreas Stafylopatis, Ngoc-Thanh Nguyen, Manuel Núñez, and Bogdan Trawiński, "Mining and categorizing interesting topics in Twitter communities", *Journal of Intelligent & Fuzzy Systems*. 32 (2017): 1265-1275.
- [13] Moosavi, Seyed Ahmad, Mehrdad Jalali, Negin Misaghian, Shahaboddin Shamsirband, and Mohammad Hossein Anisi, "Community detection in social networks using user frequent pattern mining", *Knowledge and Information Systems*. 51 (2017): 159-186.
- [14] Wang, Yang, Zengru Di, and Ying Fan, "Identifying and characterizing nodes important to community structure using the spectrum of the graph", *Public Library of Science*. 6 (2011): e27418.
- [15] Ai, Jun, Tao He, Zhan Su, and Lihui Shang, "Identifying influential nodes in complex networks based on spreading probability", *Chaos, Solitons & Fractals*. 164 (2022): 112627.
- [16] Xia, Yingjie, Xiaolong Ren, Zhengchao Peng, Jianlin Zhang, and Li She, "Effectively identifying the influential spreaders in large-scale social networks", *Multimedia Tools and Applications*. 75 (2016): 8829-8841.
- [17] Hansen, Derek, Ben Shneiderman, and Marc A Smith, *Analyzing Social Media Networks with NodeXL: Insights from a Connected World (Second Edition)*, (Morgan Kaufmann 2020), pp. Chapter 3.
- [18] Ahajjam, Sara, Mohamed El Haddad, and Hassan Badir, "A new scalable leader-community detection approach for community detection in social networks", *Social Networks*. 54 (2018): 41-49.
- [19] Zhong, Lin-Feng, Ming-Sheng Shang, Xiao-Long Chen, and Shi-Ming Cai, "Identifying the influential nodes via eigen-centrality from the differences and similarities of structure", *Physica A: Statistical Mechanics and its Applications*. 510 (2018): 77-82.
- [20] Goyal, Amit, Francesco Bonchi, and Laks VS Lakshmanan, "Discovering leaders from community actions". Proceedings of the 17th ACM conference on Information and knowledge management, 499-508, 2008.

- [21] Lu, Dongyuan, Qiudan Li, and Stephen Shaoyi Liao, "A graph-based action network framework to identify prestigious members through member's prestige evolution", *Decision Support Systems*. 53 (2012): 44-54.
- [22] Bamakan, Seyed Mojtaba Hosseini, Ildar Nurgaliev, and Qiang Qu, "Opinion leader detection: A methodological review", *Expert Systems with Applications*. 115 (2019): 200-222.
- [23] Kolahkaj, Maral, Ali Harounabadi, Alireza Nikravanshalmani, and Rahim Chinipardaz, "A hybrid context-aware approach for e-tourism package recommendation based on asymmetric similarity measurement and sequential pattern mining", *Electronic Commerce Research and Applications*. 42 (2020): 100978.
- [24] Martínez, Víctor, Fernando Berzal, and Juan-Carlos Cubero, "A Survey of Link Prediction in Complex Networks", *ACM Computing Surveys*. 49 (2016): 1-33.
- [۲۵] صالحی، سید محمد مهدی و علی اکبر پویان. "مروری بر روش‌های مدل‌سازی هم‌پوشانی در الگوریتم‌های انجمن‌یابی شبکه‌های اجتماعی". نشریه مدل‌سازی در مهندسی، ۱۷، ۵۶ (۱۳۹۸): ۲۴۷-۲۵۶. (inPersian)
- [26] Paul, Amit, and Animesh Dutta, "Community detection using Local Group Assimilation", *Expert Systems with Applications*. 206 (2022): 117794.
- [27] Tumiran, Siti Aisyah, and Bellie Sivakumar, "Community structure concept for catchment classification: A modularity density-based edge betweenness (MDEB) method", *Ecological Indicators*. 124 (2021): 107346.
- [28] Arab, Mohsen, and Mohsen Afsharchi, "Community detection in social networks using hybrid merging of sub-communities", *Journal of Network and Computer Applications*. 40 (2014): 73-84.
- [29] Dugué, Nicolas, and Anthony Perez, "Direction matters in complex networks: A theoretical and applied study for greedy modularity optimization", *Physica A: Statistical Mechanics and its Applications*. 603 (2022): 127798.
- [30] Berahmand, Kamal, and Asgarali Bouyer, "A Link-Based Similarity for Improving Community Detection Based on Label Propagation Algorithm", *Journal of Systems Science and Complexity*. 32 (2018): 737-758.
- [31] Wu, Wenhui, Sam Kwong, Yu Zhou, Yuheng Jia, and Wei Gao, "Nonnegative matrix factorization with mixed hypergraph regularization for community detection", *Information Sciences*. 435 (2018): 263-281.
- [32] Yakoubi, Zied, and Rushed Kanawati, "LICOD: A Leader-driven algorithm for community detection in complex networks", *Vietnam Journal of Computer Science*. 1 (2014): 241-256.
- [33] Li, Wei, Ce Huang, Miao Wang, and Xi Chen, "Stepping community detection algorithm based on label propagation and similarity", *Physica A: Statistical Mechanics and its Applications*. 472 (2017): 145-155.
- [34] Srilatha, Pulipati, and Ramakrishnan Manjula, "Similarity index based link prediction algorithms in social networks: A survey", *Journal of Telecommunications and Information Technology*. (2016): 87-94.
- [35] Belfin, RV, and Piotr Bródka, "Overlapping community detection using superior seed set selection in social networks", *Computers & Electrical Engineering*. 70 (2018): 1074-1083.
- [36] Pan, Xiaohui, Guiqiong Xu, Bing Wang, and Tao Zhang, "A Novel Community Detection Algorithm Based on Local Similarity of Clustering Coefficient in Social Networks", *IEEE Access*. 7 (2019): 121586-121598.
- [37] Mishra, Sneha, Shashank Sheshar Singh, Shivansh Mishra, and Bhaskar Biswas, "TCD2: Tree-based community detection in dynamic social networks", *Expert Systems with Applications*. 169 (2021): 114493.
- [38] Helal, Nivin A., Rasha M. Ismail, Nagwa L. Badr, and Mostafa G. M. Mostafa, "Leader-based community detection algorithm for social networks", *WIREs Data Mining and Knowledge Discovery*. 7 (2017).
- [39] Yang, Chen, Lei Liu, Li Chen, and Ben Niu, "A novel friend recommendation service based on interaction information mining". 2017 International Conference on Service Systems and Service Management, 1-5, (IEEE), 2017.
- [40] Lim, Kwan Hui, and Amitava Datta, "An interaction-based approach to detecting highly interactive Twitter communities using tweeting links", *Web Intelligence*. 14 (2016): 1-15.
- [41] Bonacich, Phillip, "Power and centrality: A family of measures", *American journal of sociology*. 92 (1987): 1170-1182.
- [42] Zaki, Mohammed Javeed, "Scalable algorithms for association mining", *IEEE transactions on knowledge and data engineering*. 12 (2000): 372-390.
- [43] Telikani, Akbar, Amir H. Gandomi, and Asadollah Shahbahrami, "A survey of evolutionary computation for association rule mining", *Information Sciences*. 524 (2020): 318-352.

- [44] Luo, Wenjian, Nannan Lu, Li Ni, Wenjie Zhu, and Weiping Ding, "Local community detection by the nearest nodes with greater centrality", *Information Sciences*. 517 (2020): 377-392.
- [45] Malliaros, Fragkiskos D, and Michalis Vazirgiannis, "Clustering and community detection in directed networks: A survey", *Physics reports*. 533 (2013): 95-142.
- [46] De Domenico, Manlio, Antonio Lima, Paul Mougel, and Mirco Musolesi, "The anatomy of a scientific rumor", *Scientific reports*. 3 (2013): 2980.
- [47] Platform, Stanford Network Analysis, and (SNAP) higgs-twitter (Accessed February 18, 2023) (2015) <http://snap.stanford.edu/data/higgs-twitter.html>
- [48] Tan, Chenhao, Jie Tang, Jimeng Sun, Quan Lin, and Fengjiao Wang, "Social action tracking via noise tolerant time-varying factor graphs". Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, 1049-1058, 2010.
- [49] (ArnetMiner), AMiner Flickr-large Accessed February 18, 2023 (2006) <https://www.aminer.cn/data-sna#Flickr> large
- [50] Lancichinetti, Andrea, Santo Fortunato, and János Kertész, "Detecting the overlapping and hierarchical community structure in complex networks", *New journal of physics*. 11 (2009): 033015.
- [51] Sun, Peng Gang, Xunlian Wu, Yining Quan, and Qiguang Miao, "Influence percolation method for overlapping community detection", *Physica A: Statistical Mechanics and its Applications*. 596 (2022): 127103.
- [52] Singh, Dipika, and Rakhi Garg, "NI-Louvain: A novel algorithm to detect overlapping communities with influence analysis", *Journal of King Saud University - Computer and Information Sciences*. 34 (2022): 7765-7774.