سال دوازدهم، شماره‌های ۴۳ و ۴۴، بهار و تابستان ۱۳۹۹
صفحات: ۵۵_۶۹

دو فصلنامه علمی
فناوری اطلاعات و ارتباطات ایران

ارائه‌ی یک روش بهبودیافته مبتنی بر رویکردهای انتشار برچسب و بهینه‌سازی ماژولاریتی برای تشخیص جوامع در شبکه‌های اجتماعی پویا

محمد ستاری* کامران زمانی فر**
*مرکز تحقیقات فناوری اطلاعات در امور سلامت، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران
**گروه مهندسی نرم‌افزار، دانشکده مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان، ایران
تاریخ دریافت: ۱۳۹۸/۱۰/۰۹ تاریخ پذیرش: ۱۳۹۹/۰۴/۱۶
نوع مقاله: پژوهشی

چکیده
تشخیص جوامع در شبکه‌های اجتماعی پویا یکی از مهمترین موضوعات تحقیقاتی است که در سالیان اخیر مورد توجه قرار گرفته است. رویکردهای گوناگونی برای تشخیص جوامع در شبکه‌های اجتماعی در حالت پویا وجود دارد که در بین آنها، رویکرد انتشار برچسب دلیل سادگی و کارایی انتخاب شده است. این رویکرد شامل روش‌های بسیاری است که غالبا تصادفی هستند. از میان این روش‌ها، روش رتبه‌بندی برچسب مبتنی بر زمان یک روش قطعی است. البته مسلما این روش هم با مشکلاتی مواجه است، یکی از مشکلات این است وقتی یک گره می‌خواهد به یک جامعه بپیوندد، ساختار درونی آن جامعه در نظر گرفته نمی‌شود. بنابراین برای حل این مشکل، یک رویکرد حریصانه برای بهبود رویکرد انتشار برچسب استفاده شده است. رویکرد جدید ارائه‌شده و دیگر روش‌های مورد ارزیابی در مجموعه داده‌های واقعی و مصنوعی مورد آزمایش قرار می‌گیرند. نتایج نشان می‌دهد که روش پیشنهادی نسبت به روش‌های دیگر از لحاظ میزان صحت و قابلیت افراز به نسبت بهتر عمل کرده است.

واژگان کلیدی: رتبه‌بندی برچسب مبتنی بر زمان حریصانه، رتبه‌بندی برچسب مبتنی بر زمان، رویکرد انتشار برچسب، تشخیص جوامع، شبکه‌ی اجتماعی پویا

۱. مقدمه

شبکه‌های اجتماعی میلیون‌ها و حتی میلیاردها کاربر را از سراسر جهان در سال‌های اخیر به سوی خود جلب کرده‌اند. برای مثال، شبکه‌ی اجتماعی فیسبوک بیش از یک میلیارد کاربر دارد. یک شبکه‌ی اجتماعی، مجموعه‌ای از گره‌ها و یال ها است گره‌ها معادل با کاربران شبکه هستند و یال‌ها، وابستگی بین گره‌ها مانند دوستی، خویشاوندی، همکاربودن و غیره را مشخص می‌کند. یکی از موضوعات مطرح در زمینه-ی شبکه‌های اجتماعی، مبحث تشخیص جوامع است. به طور کلی یک جامعه در یک شبکه‌ی اجتماعی، مجموعه‌ای از گره-ها است که تراکم ارتباطی بین آنها بالاست. در عین حال تراکم ارتباطی پایینی با گره‌های خارج از جامعه دارند [۱]. تشخیص جوامع نه تنها برای بهبود بازاریابی مفید است، بلکه می‌تواند به بهبود مسیریابی در شبکه کمک کند [۲و۳]. روش‌های بسیاری برای تشخیص جوامع به صورت ایستا در شبکه‌های اجتماعی وجود دارد [۶-۴]، این در حالیست که شبکه‌های اجتماعی ذاتا پویا هستند و مدام در حال تغییر هستند.

رویکردهای مختلفی برای تشخیص جوامع در حالت پویا وجود دارد [۱۰-۷] که در بین آنها، رویکرد

نویسنده مسئول: محمد ستاری
msattarimng.mui@gmail.com

انتشار برچسب[1] به دلیل سادگی و کارایی انتخاب شده است [۶]. در این رویکرد، هر گره یک برچسب به صورت یکتا دارد که آن را با همسایگانش به اشتراک می‌گذارد. در نهایت، برچسب‌های مشابه جوامع را مشخص می‌کنند. این رویکرد با توجه به محلی‌بودن، قابلیت اجرا به صورت موازی را دارد [۱۱]. رویکرد انتشار برچسب، مزایای زیادی دارد، ولی تصادفی بودن این رویکرد می‌تواند منجر به این شود که نتایج آن در اجراهای مختلف، متفاوت شود. یکی از مهمترین روش-هایی که سعی کرده است قابلیت تصادفی‌بودن این رویکرد را به قطعی‌بودن تبدیل کند، روش رتبه‌بندی برچسب مبتنی بر زمان[2] [۱۲] نام دارد که مبتنی بر روش رتبه‌بندی برچسب[3] [۱۳] است.

در روش رتبه‌بندی برچسب مبتنی بر زمان، هر گره در ابتدای یک بردار توزیع برچسب دارد که آن را با گره‌های همسایه‌اش به اشتراک می‌گذارد. در واقع بردار هر گره در طی ۴ گام انتشار،تورم، بریدن و بروزرسانی شرطی دوباره مقداردهی خواهد شد. گام‌های ۱ تا ۴ چندین مرتبه اعمال خواهد شد. در نهایت گره‌هایی که برچسب‌های با بیشترین احتمال‌شان مشابه هستند، تشکیل یک جامعه را می‌دهند. در این روش، معمولا یک گره‌ی که در یک نما به شبکه‌ی اجتماعی اضافه می‌شود به جامعه‌ای می‌پیوندد که بیشترین یال را به آن دارد بدون آنکه ساختار داخلی جوامع را در نظر بگیرد. این مسئله منجر به کاهش دقت تشخیص جوامع در شبکه های اجتماعی در حالت پویا خواهد شد. در روش ارائه‌شده برای حل این مسئله، استفاده از یک رویکرد به نام بهینه‌سازی ماژولاریتی[4] در کنار روش رتبه‌بندی برچسب مبتنی بر زمان پیشنهاد شده است. در این رویکرد از معیاری به نام ماژولاریتی محلی [۱۰] برای در نظر گرفتن ساختار جوامع استفاده شده است. در رویکرد پیشنهادی، در نمای اولیه، روش رتبه‌بندی مبتنی بر زمان استفاده می‌شود و در نماهای بعدی، ترکیبی از روش رتبه‌بندی مبتنی بر زمان و رویکرد حریصانه مبتنی بر بهبود ماژولاریتی محلی[5] استفاده می‌شود. روش رتبه‌بندی مبتنی بر زمان حریصانه به عنوان نام روش پیشنهادی در نظر گرفته شده است.

مجموعه داده‌های مورد استفاده، مجموعه داده‌های واقعی و مصنوعی[6] هستند. مجموعه داده‌های مصنوعی به کمبود جوامع حقیقی در مجموعه داده‌های واقعی استفاده شده‌اند. روش پیشنهادی و دیگر روش‌های مورد ارزیابی براساس معیارهای کیفیت افراز، صحت و سرعت با هم مقایسه می‌شوند.

در بخش دوم کارهای مرتبط بیان می‌شوند. در بخش سوم مقدمات بیان خواهد شد و در بخش چهارم، روش پیشنهادی تشریح خواهد شد. در بخش پنجم، ارزیابی و بخش ششم، نتیجه‌گیری خواهد بود.

۲. کارهای مرتبط

رویکردهای مختلفی برای تشخیص جوامع در شبکه‌های اجتماعی وجود دارد که مهمترین این رویکردهای شامل تئوری بازی، انتشار برچسب و بهینه‌سازی ماژولاریتی است.

۲-۱ تئوری بازی

در رویکرد تئوری بازی [۱۶،۱۵]، هر گره به عنوان یک عامل در نظر گرفته می‌شود که قصد دارد بهره‌وری خود را بهبود دهد. نقطه‌ی ضعف این رویکرد، این است که گره‌هایی که در هر نما وارد شبکه‌ی اجتماعی می‌شوند به نوعی تابع جوامعی قبلی موجود در شبکه‌های اجتماعی هستند و شانس ایجاد جوامع جدید را ندارند.

۲-۲ رویکرد بهینه‌سازی ماژولاریتی

در رویکرد بهینه‌سازی ماژولاریتی [۱۹-۱۷]، هر گره به صورت یک عامل در نظر گرفته می‌شود و هر عملی که گره انجام می‌شود در جهت بهبود ماژولاریتی است. یکی از مشکلات این رویکرد، ایجاد جوامع تک عضوی با تعداد اعضای کم است. روش [19] که به نام C-Blondel شناخته می‌شود. روشی سریعتر از روش لووین [20] بود. دلیل سریعتربودن روش، کاهش اندازه‌ی گراف‌های فشرده‌شده در هر نما بود. این روش بر مبنای این ایده بود که گره‌های و یال‌هایی که در یک نما به گراف اضافه و یا حذف می‌شوند، تاثیر زیادی در نزدیکترین همسایه-های خود دارند. بنابراین تاثیر رویدادهای افزودن و حذف گره و یال در این روش بررسی می‌شد.

۲-۳ رویکرد انتشار برچسب

رویکرد انتشار برچسب، یک رویکرد ساده و کارا است. ایده‌ی روش‌های مبتنی بر انتشار برچسب برای تشخیص جوامع در ابتدا توسط راگاون و همکاران [۶] مطرح شد

[1] Label Propagation
[2] LabelRankT
[3] LabelRank
[4] Modularity Optimization
[5] Local Modularity
[6] Synthetic

به طوری که در آن تشخیص جوامع به صورت مجزا در حالت ایستا در نظر گرفته می‌شد. راگاون اعتقاد داشت که در رویکرد انتشار برچسب، ٪۹۵گره‌ها پس از انجام ۵تکرار، تقریبا برچسب نهایی خود را انتخاب می‌کنند [۶]. لیونگ و همکاران [۲۱] پتانسیل انتشار برچسب را برای شبکه‌های اجتماعی در حالت پویا مطرح کردند. آنها اعتقاد داشتند که با توجه به محلی‌بودن روش انتشار برچسب، این امکان وجود دارد که به صورت پویا انجام شود.

روش‌های مختلفی مبتنی بر انتشار برچسب در حالت پویا وجود دارد. یکی از این روش‌ها، روش رتبه‌بندی برچسب مبتنی بر زمان [۸] است که یک روش قطعی مبتنی بر رویکرد انتشار برچسب است. یکی از نقاط ضعف این روش، در نظر نگرفتن ساختار جوامع در هنگام پیوستن یک گره به جامعه است. هان و همکاران [۲۲] روش انتشار برچسب منطبق را برای تشخیص جوامع در شبکه‌های اجتماعی در حالت پویا ارائه دادند. در این روش، رویکرد انتشار برچسب در نمای اولیه استفاده می‌شود. در نماهای دوم به بعد مشابه روش تشخیص جوامع ذاتا طولی[7] برای هر تغییر به صورت جداگانه عمل می‌شود. یکی از نقاط ضعف این روش، این است که ساختار قبلی جامعه با افزودن یک یال می‌تواند از بین رود.

لیو و همکاران [۲۳] روش انتشار برچسب برجسته‌ی گسترش‌یافته که ارتقایافته‌ی روش انتشار برچسب برجسته [24] بود، برای تشخیص جوامع در شبکه‌های اجتماعی در حالت پویا ارائه دادند. در این روش، مفهومی به نام برچسب برجسته برای هر گره تعریف شده است که نشان‌دهنده‌ی جامعه‌ی اصلی آن گره است. این روش، قابلیت تشخیص جوامع همپوشان و مجزا را با هم فراهم کرده است. در واقع جامعه‌ی اصلی هر گره متناظر با جوامع مجزا هستند. یکی از نقاط ضعف این روش، این است که ممکن است ماکزیمم محلی به عنوان بیشترین بهره‌وری گره در نظر گرفته شود. زینژنگ و همکاران روش دیگری [۲۵] را برای تشخیص جوامع مبتنی بر انتشار برچسب در شبکه‌های اجتماعی پویا ارائه داد. در این روش از الگوریتم ژنتیک در کنار انتشار برچسب جهت تشخیص جوامع در شبکه‌های اجتماعی استفاده شده است. یکی از نقاط ضعف این روش، این است که گره‌های جدیدی که در هر نما وارد می‌شوند معمولا عضو جوامع گره‌های قبلی می‌شوند.

۳. مقدمات

در این قسمت در ابتدا علائمی که استفاده شده است، مختصرا توضیح داده می‌شود و سپس روش رتبه‌بندی برچسب مبتنی بر زمان توضیح داده خواهد شد.

۳-۱ علائم

به طور کلی، یک شبکه توسط یک گراف G(V,E) بیان می‌شود که در آن V مجموعه‌ی رئوس و E مجموعه‌ی یال‌ها است. در این مقاله از جهت یال‌ها صرف‌نظر شده است. جامعه به صورت مجموعه‌ای از گره‌ها در نظر گرفته می‌شود که توسط Com بیان می‌شود. با توجه به اینکه روش ارائه‌شده، جوامع غیر همپوشان یا مجزا را در نظر می‌گیرد، تشخیص جامعه به صورت ماژول‌های مجزا با چگالی درونی بالا و چگالی بیرونی پایین تعریف می‌شود. شبکه‌ی اجتماعی پویا، به صورت مجموعه‌ای از نماها تعریف می‌شود. این نماها در یک مجموعه‌ی داده‌ای واقعی در دوره‌های زمانی مختلف از سیستم گرفته می‌شود، در حالی که در مجموعه‌های داده‌ای مصنوعی تولید می‌شوند. هر نما متناسب با یک گراف است. نماد T، بیان‌کننده‌ی شماره‌ی نما است که از یک شروع می‌شود.

۳-۲ رتبه‌بندی برچسب مبتنی بر زمان

در این روش، فرض می‌شود که تعداد کل گره‌ها برابر با n باشد. در این روش ، رتبه‌بندی برچسب [۱۳] در نمای اول اعمال می‌شود. بدین‌صورت که هر گره یک بردار توزیع برچسب اسپارس به طول تعداد n دارد که هر درایه‌ی این بردار، احتمال عضویت برچسب متناظر با آن درایه را برای گره موردنظر مشخص می‌کند. در ابتدا احتمال عضویت کلیه‌ی برچسب‌ها مساوی و برابر با $\frac{1}{n}$ است. در ادامه برای کلیه‌ی گره‌ها، توزیع برچسب در طی ۴ گام انتشار، تورم، بریدن و بروزرسانی شرطی دوباره مقداردهی خواهد شد [۱۲]. گام‌های ذکر شده به صورت ذیل انجام می‌شود :

[7] Inherently Longitudinal

۱)گام انتشار: در گام انتشار، هر گره ماتریس برچسب خود را برای گره‌های همسایه‌اش ارسال می‌کند. سپس هرگره میانگین ماتریس برچسب‌های ارسال شده توسط همسایگانش را محاسبه کرده و به عنوان ماتریس برچسب جدیدش در نظر می‌گیرد.

۲)گام تورم: در این گام، عملگر تورم اعمال می‌شود. بدین‌صورت که احتمال هر برچسب در ماتریس برچسب جدید بروزرسانی می‌شود. در واقع برای هر برچسب، احتمالش به توان عملگر تورم می‌رسد و بر مجموع احتمال کلیه‌ی برچسب‌ها به توان عملگر تورم تقسیم می‌شود. بدین‌ترتیب احتمال جدید برچسب محاسبه می‌شود. این عمل باعث می‌شود که میزان احتمال برچسب‌های محتمل افزایش یابد و متعاقبا میزان احتمال برچسب‌هایی که محتمل نیستند، کاهش یابد.

۳) گام قطع‌کردن: برچسب‌هایی که احتمالشان کمتر از یک مقدار آستانه است از ماتریس برچسب گره‌ها حذف می‌شوند. این گام، تعداد درایه‌های بردار برچسب را به میزان زیادی کاهش می‌دهد.

۴)گام بروزرسانی شرطی: در گام بروزرسانی شرطی، بروزرسانی‌های ذکرشده در صورتی که ماتریس برچسب گره تفاوت محسوسی با گره‌های همسایه داشته باشد، تغییر خواهد کرد. بنابراین می‌توان گفت که این گام به دنبال حفظ ساختار جوامع است.

گام‌های ۱ تا ۴ چندین مرتبه اعمال خواهد شد. در نهایت گره‌هایی که برچسب‌های با بیشترین احتمال‌شان مشابه هستند، تشکیل یک جامعه را می‌دهند. در نماهای دوم به بعد، برای گره‌هایی که یال‌های متصل به آنها تغییر نکرده‌اند، توزیع برچسب آنها تغییری نمی‌کند. اما برای بقیه‌ی گره‌ها، روش رتبه‌بندی برچسب اعمال می‌شود و توزیع برچسب در طی ۴ فاز انتشار،تورم، قطع‌کردن و به‌روزرسانی شرطی دوباره مقدار-دهی خواهد شد. این مقداردهی دوباره چندین مرتبه اعمال شده و هرکدام از این گره‌های تغییریافته به یکی از جوامع موجود می‌پیوندند. در نهایت گره‌هایی که محتمل‌ترین برچسب‌هایشان مشابه هستند تشکیل یک جامعه را می‌دهند.

۴. روش پیشنهادی

در این مقاله، یک رویکرد حریصانه‌ی جدید به روش رتبه‌بندی برچسب مبتنی بر زمان به نام بهینه‌سازی ماژولاریتی اضافه شده است. روش پیشنهادی به دو بخش ایستا و پویا تقسیم می‌شود. بخش ایستا جوامع را در نمای اولیه تشخیص می‌دهد و بخش پویا جوامع را در نماهای دوم به بعد تشخیص می‌دهد.

۴-۱ بخش ایستا

این بخش شامل اولین نما از گراف شبکه‌ی اجتماعی است که الگوریتم تشخیص جوامع بر روی آن پیاده‌سازی می‌شود. در این بخش، روش رتبه‌بندی برچسب بکار برده می‌شود به گونه‌ای که به هر گره، یک بردار توزیع برچسب اختصاص داده می‌شود. سپس گره‌ها بردار توزیع برچسب خود را به صورت همگام بروز می‌کنند. در نهایت گره‌هایی که محتمل‌ترین برچسب‌هایشان مشابه هستند تشکیل یک جامعه را می‌دهند. بنابراین جوامعی که بدین صورت تشخیص داده شدند به عنوان ورودی به بخش پویا داده خواهند شد.

۴-۲ بخش پویا

این بخش شامل نمای دوم به بعد است. این بخش، یک روشی را ارائه می‌کند که تشخیص جوامع را بهبود دهد. برای رسیدن به این هدف، از یک معیار به نام ماژولاریتی محلی استفاده می‌شود. در ماژولاریتی محلی، فرض می‌شود که یک جامعه به دو بخش مرکزی و مرزی تقسیم می‌شود و هر چه ارتباطات در بخش مرکزی بیشتر و در بخش مرزی کمتر باشد، ماژولاریتی بیشتر خواهد شد. این معیار به صورت ذیل محاسبه می‌شود [۲۶] :

$$Lm = \frac{\frac{\sum_{v\in D} IK_v}{|D|}}{\frac{\sum_{v\in D} EK_v}{|B|}} \qquad (۱)$$

که D یکی از جوامع تشخیص داده شده است و B گره‌های مرزی D را در برمی‌گیرد. گره‌ی می‌تواند گره‌ی مرزی باشد که حداقل یک همسایه خارج از مجموعه‌ی D داشته باشد [26]. نماد IK_v تعداد ارتباطات بین گره v و بقیه‌ی گره‌های موجود در D را بیان می‌کند. نماد EK_v تعداد ارتباطات بین گره‌ی v و گره‌ها در S را بیان می‌کند. مجموعه‌ی S شامل گره‌هایی هست که مجاور گره‌های مجموعه‌ی D هستند، ولی به این مجموعه تعلق ندارند.

روش ارائه‌شده، روش رتبه‌بندی برچسب را با یک رویکرد بهینه‌سازی ماژولاریتی به صورت محلی ترکیب می‌کند. در ابتدا، این روش، یک مجموعه‌ی جوامع کاندیدا برای

هر گره‌ی تغییریافته ایجاد می‌کند. این مجموعه شامل جوامعی است که این گره حداقل یک یال به آنها دارد. در واقع یک گره می‌تواند به یکی از این مجموعه‌ها بپیوندد. اینکه کدامیک را انتخاب کند برحسب ماژولاریتی محلی مشخص می‌شود.

در واقع برای هر یک از جوامع موجود در مجموعه‌ی جوامع کاندیدا، معیار ماژولاریتی محلی محاسبه می‌شود. سپس میزان بهبود یا عدم بهبود ماژولاریتی محلی با افزودن آن گره محاسبه شده و از بین جوامع، جامعه‌ی که بیشترین بهبود را در معیار ماژولاریتی محلی ایجاد کرده است به عنوان جامعه‌ی کاندیدا در نظر گرفته می‌شود. پس از اینکه جامعه‌ی کاندیدا برای گره موردنظر مشخص شد، بردار توزیع جامعه‌ی انتخاب شده در یک ضریب α ضرب شده و بردار توزیع دیگر جوامع در یک ضریب β ضرب می‌شود. مقدار α بزرگتر از یک و مقدار β کوچکتر از یک خواهد بود. استفاده از این ضرایب، شانس رویدادهایی مرتبط با جامعه‌ی منتخب را افزایش می-دهد در واقع شانس رویدادهای جوامع دیگر را کاهش می‌دهد. به طوری که هر چه مقدار α بیشتر از یک باشد، شانس این رویدادها بیشتر خواهد بود. در مقابل هر چه مقدار β به صفر نزدیکتر باشد، شانس رویدادهای جوامع دیگر بیشتر کاهش می‌یابد. در واقع اگر جامعه‌ی منتخب معادل با جامعه‌ی باشند که گره قبلا به آن تعلق داشته، شانس رویداد باقی‌ماندن در جامعه افزایش می‌یابد. در غیراینصورت، شانس رویدادهای تغییر یا توسعه افزایش می‌یابد. در واقع استفاده از ماژولاریتی محلی می‌تواند در مدیریت بهتر موثر باشد. یک بردار توزیع جدید برای هر گره‌ی تغییر یافته به صورتی که در ذیل آمده است، محاسبه می‌شود.

$$\mathrm{p_i} = \frac{\sum_{\mathrm{j}\in \mathrm{neighbour(i)}} coef_j \mathrm{P_j}}{\mathrm{k}} \qquad (۲)$$

که p_i و p_j به ترتیب بردارهای توزیع گره‌های i و j هستند و نماد k، تعداد همسایگان گره‌ی i را مشخص می‌کند.

همچنین ضریب $coef_j$، ضریب گره‌ی i را تعیین می‌کند.

هر یک از گره‌ها، ممکن است چندین ضریب متناظر با گره‌های تغییریافته‌ی متفاوت داشته باشند. به بیان دیگر، برای هر یک از گره‌های تغییریافته، روش رتبه‌بندی برچسب بردار توزیع هر گره را به صورت مجزا محاسبه می‌کند و بردار توزیع گره‌های همسایه بدون تغییر باقی می‌ماند. این روش شامل ۴ گام است :

۱) گام انتشار: در این گام، هر گره ماتریس برچسب خود را برای گره‌های همسایه‌اش می‌فرستد. سپس آن گره میانگین ماتریس برچسب‌های ارسال‌شده توسط همسایگانش را محاسبه کرده و به عنوان ماتریس برچسب جدیدش در نظر می‌گیرد .

۲) گام تورم: در این گام، عملگر تورم اعمال می‌شود. بدین صورت که احتمال هر برچسب در ماتریس احتمال کلیه‌ی برچسب‌ها به توان عملگر تورم تقسیم می‌شود. بدین ترتیب احتمال جدید برچسب محاسبه می‌شود. این عمل باعث می‌شود که شانس انتخاب برچسب‌ها با احتمال بالا، بیشتر افزایش یابد و متعاقبا شانس انتخاب برچسب‌ها با احتمال پایین، بیشتر کاهش یابد.

۳) گام قطع‌کردن: برچسب‌هایی که احتمالشان کمتر از یک مقدار آستانه است از ماتریس برچسب گره‌ها حذف می‌شوند. این گام، تعداد درایه‌های بردار برچسب را به میزان زیادی کاهش می‌دهد.

۴) گام بروزرسانی شرطی: در گام بروزرسانی شرطی، بروزرسانی‌های ذکرشده در صورتی که ماتریس برچسب گره تفاوت محسوسی با گره‌های همسایه داشته باشد، تغییر خواهد کرد. این گام به دلیل حفظ ساختار جوامع است.

گام‌های فوق چندین مرتبه اجرا شده و در نهایت گره‌های که محتمل‌ترین برچسب‌های مشابه هستند، تشکیل یک جامعه را می‌دهند.

۳-۴ الگوریتم روش پیشنهادی

این الگوریتم براساس وجود دو بخش[1] ایستا و پویا برای روش به صورت الگوریتم ۱ و ۲ مشخص می‌شود. در الگوریتم ۱ که شامل بخش ایستای روش است. در ابتدا، بردار توزیع برچسب مقداردهی اولیه می‌شود. سپس گام-های روش رتبه‌بندی برچسب بر روی گره‌ها و یال در اولین نما پیاده‌سازی می‌شود. در آخر، جوامع تشخیص داده شده و به عنوان ورودی نمای دوم در نظر گرفته می‌شوند.

الگوریتم ۲ شامل بخش پویای روش می‌شود و نام آن، رتبه‌بندی برچسب حریصانه است. در این الگوریتم، در ابتدا برای کلیه‌ی گره‌ی تغییر یافته، مجموعه‌های D و S ایجاد می‌شوند. سپس، گره‌هایی که به D اتصال دارند ولی جزیی از D نیستند به عنوان مجموعه‌ی S در نظر گرفته می‌شوند. هر یک از اعضای مجموعه‌ی S به عنوان یک جامعه در نظر گرفته شده و هر یک از اعضای مجموعه‌ی B و D به عنوان یک گره لحاظ می‌شوند. در میان اعضای مجموعه‌ی S، جامعه‌ی با بیشترین افزایش

در ماژولاریتی محلی انتخاب شده و نام آن CCom قرار داده می‌شود. سپس ضرایب گره‌های تغییریافته و همسایگان براساس CCom محاسبه می‌شود. در ادامه، بردار توزیع برچسب برای هر یک از گره‌های تغییریافته براساس ضرایب و عملگرهای رتبه‌بندی برچسب دوباره مقداردهی می‌شوند. بروزرسانی بردار برچسب تا زمانی انجام می‌شود که شرایط متوقف‌شدن بروزرسانی فراهم شود. شرط توقف در بروزرسانی، این است که تعداد گره‌هایی که بردارروزرسانی‌شان تغییر می‌کند از یک آستانه‌ی از پیش‌تعریف‌شده کمتر باشد. آستانه‌ی از پیش‌تعریف‌شده براساس تعداد گره‌های موجود در مجموعه داده‌های مورد ارزیابی تعیین می‌شود. و در نهایت بردار توزیع برچسب‌های تغییر نیافته کپی شده و جوامع استخراج خواهند شد.

الگوریتم ۱ : رتبه بندی برچسب

Algorithm 1 : LabelRank [9]

1. input : snapshot G (V,E) ([0])
2. add selfloop to adjacency matrix W
3. initialize label distribution P
4. **repeat**
5. P′= W*P // Propagate label distribution vector
6. P′= Γ_{in} P′ // Inflate label distribution vector to contract the propagation
7. P′= Φ_r P′ // Remove labels that are below the threshold
8. P′ = Θq (P′, P) // Conditional updating
9. **until** stop criterion satisfied
10. output communities

الگوریتم ۲ : رتبه بندی برچسب حریصانه

Algorithm 2 : GreedyLabelRank

1. input **:** snapshot G (V,E) ([1,2,…,T])
2. **for** t=2:T **do**
tracking the changed nodes in G(t) due to the changes in the edges they attach to since t-1
3. **for each** changed nodes cn
4. Create empty set s,d,B
5. add cn to D and B
6. add all candidate Community of D to S
7. **For all** $Com \in S$ **do**
8. **compute** L'_m
9. **end for**
10. **Find** CCOM with the maximum increasing L'_m
11. **Compute** coefficient of cn and its neighbours based on Com
12. **repeat**
13. $P'_{cn} = \frac{\sum_{jeneigbour(cn)} coeff_j * p_j}{N}$ // compute label distribution vector
$P'_{cn}{}' = \Gamma_{in}$ P // Inflate the node label distribution vector to contract the propagation
15. $P'_{cn} = \Phi_r$ P′ // Remove the node labels that are below the threshold
16. $P'_{cn} = \Theta q$ (P′, P) // Conditional updating for the node label distribution vector
17. **until** stop criterion satisfied
18. **end for**
19. copy label distribution of unchanged nodes
20. **end for**

۴_۴ پیچیدگی زمانی

نماد n را تعداد گره‌ها و نماد m را تعداد یال‌ها در نظر بگیرید. در بخش ایستای روش، قسمت انتشار برچسب بر روی یال‌ها اجرا شده و بقیه‌ی قسمت‌ها بر روی گره‌ها اجرا می‌شوند. بنابراین پیچیدگی زمانی این قسمت متناظر با تعداد یال‌ها و برابر با O(m) است. در بخش دوم که متناظر با بخش پویا است، دو قسمت مهم وجود دارد. قسمت اول که متناظر با محاسبه ی ماژولاریتی محلی است و پیچیدگی زمانی آن بر بر حسب O(n*m) خواهد بود. قسمت دوم که متناظر با روش روش رتبه-بندی برچسب است که بر حسب O(m) خواهد بود لذا با توجه به اینکه t نما داریم، بنابراین پیچیدگی زمانی بر روی کلیه‌ی نماها برابر با O(n*m*t) خواهد بود. با توجه

به اسپارس‌بودن شبکه‌ها، اگر m را متناظر با O(kn) در نظر بگیریم که k، درجه‌ی شبکه‌ی است، آنگاه پیچیدگی زمانی بر حسب $O(n^2*k*t)$ خواهد بود.

۵. ارزیابی

در این بخش، صحت و زمان اجرای روش پیشنهادی و دو روش رتبه‌بندی برچسب مبتنی بر زمان [۱۳]، نسخه‌ی غیر همپوشان روش انتشار برچسب برجسته‌ی گسترش‌یافته [۲۳] ، روش الگوریتم ژنتیک چند هدفه پویا مبتنی بر انتشار برچسب [27] و روش C-Blondel [19] را بر روی مجموعه-های داده‌ای واقعی و مصنوعی مورد ارزیابی قرار خواهد گرفت. در این قسمت در ابتدا مجموعه‌های داده‌ای معرفی می‌شوند و سپس معیارهای ارزیابی بیان خواهند شد. در نهایت، نتایج ارزیابی روش پیشنهادی و روش‌های موجود بر روی شبکه‌ی واقعی و مصنوعی مورد بررسی قرار می‌گیرد.

۵-۱ معرفی مجموعه داده‌های واقعی

مجموعه داده‌های واقعی شامل مجموعه‌های ارجاع مقالات و مسیریابی است که در مورد هر یک از آنها صحبت خواهیم کرد. تفاوت این دو مجموعه در این است که در مجموعه‌های ارجاع مقالات، یگانه رویدادی که برای گره و یال‌ها اتفاق می-افتد، رویداد افزودن گره یا یال است. در حالیکه در مجموعه‌ی داده‌ای مسیریابی، علاوه بر رویداد افزودن، حذف نیز اتفاق می‌اقتد.

۵-۱-۱ مجموعه داده‌های ارجاع مقالات

مجموعه داده‌های ارجاع مقالات شامل مقالات و ارتباط بین آنها است. ارتباط بین دو مقاله بدین صورت شکل می‌گیرد که یک مقاله به مقاله دیگر ارجاع دهد. در واقع ارتباط بین مقالات، یک ارتباط جهت‌دار خواهد بود. نکته‌ای وجود دارد، این است که یک مقاله می‌تواند به مقاله‌ای ارجاع دهد که پیش از آن چاپ شده است.

Arxiv HEP-PH [۲۸]: این یک مجموعه ارجاع مقالات است که شامل مقالاتی است که بین ژانویه ۱۹۹۳ تا آپریل ۲۰۰۳ چاپ شده‌اند. این مجموعه‌ی داده‌ای شامل ۳۴۵۴۶ گره و ۴۲۱۵۷۸ یال است که در آن گره‌ها نشان-دهنده‌ی مقالات و یال‌ها وجود ارجاع بین آنها را نشان می-دهند. به عنوان مثال اگر مقاله‌ی i به مقاله‌ی j ارجاع داده باشد، آنگاه یک یال جهت‌دار بین گره‌ی i و گره‌ی j وجود دارد. این مجموعه داده، مقالاتی که در هر ماه چاپ شده‌اند را به عنوان یک نما در نظر می‌گیرد. لازم به ذکر است اگر یک مقاله درون این مجموعه‌ی داده‌ای به مقاله‌ای خارج از پایگاه داده ارجاع بدهد هیچ یالی بین آنها ایجاد نخواهد شد و گراف مجموعه داده هیچ اطلاعی در این مورد نخواهد داد.

Arxiv HEP-TH [۲۹]: یکی دیگر از مجموعه داده‌های ارجاع مقالات است که مقالات موجود در زمینه-ی تئوری فیزیک را پشتیبانی می‌کند. مشابه مجموعه داده‌ی قبلی، مقالات را معادل گره در گراف شبکه‌ی اجتماعی و ارجاعات بین مقالات را به عنوان یال‌های گراف در نظر می‌گیرد. این مجموعه داده شامل ۲۷۷۷۰ گره و ۳۵۲۸۰۷ یال است. این مجموعه‌ی داده‌ای مقالات ۱۱ سال را از ژانویه ۱۹۹۳ تا آپریل ۲۰۰۳ و مقالات و ارتباطات بین آنها در هر دو ماه را به عنوان یک نما در نظر می‌گیرد. تفاوت این مجموعه داده با مجموعه داده‌ی قبلی در این است که تعداد نماها را نسبت به مجموعه‌ی قبلی کاهش داده است.

۵-۱-۲ مجموعه داده‌های مسیریابی

Autonomous System-Oregon-1 [۲۹]: یک سیستم خودکار است که شامل روترهایی است که بین ۳۱ مارس ۲۰۰۱ تا ۲۶ می ۲۰۰۱ ایجاد شده‌اند. این سیستم شامل ۱۱۱۷۴ گره و ۲۳۴۰۹ یال است که گره‌ها همان روترها و یال‌ها جریان ارتباطی بین آنها را نشان می‌دهند. نکته‌ای که این مجموعه‌ی داده‌ای را نسبت به مجموعه داده‌های دیگر جدا می‌کند، این است که بازه‌ی زمانی کوتاه‌تری را در نظر گرفته است (تقریبا ۲ ماه). بنابراین نماهای کوچکتری در نظر گرفته می‌شوند. در این مجموعه داده، هر هفته به عنوان یک نما در نظر گرفته می‌شود. تمایز این مجموعه داده با مجموعه داده‌ی ارجاع مقالات، این است که گره‌ها و یال‌ها علاوه بر اینکه اضافه می‌شوند، می‌توانند حذف شوند. پس تعداد گره‌ها می‌تواند کم یا زیاد شوند.

Autonomous Systems AS-73 [۲۹]: یک سیستم خودکار است که شامل روترهایی است که بین نوامبر ۱۹۹۷ تا ژانویه ۲۰۰۰ ایجاد شده‌اند. بنابراین بازه-ی زمانی بیشتری را نسبت به مجموعه داده‌ی قبلی در نظر می‌گیرد. این سیستم شامل ۶۴۷۴ گره و ۱۳۸۹۵ یال است که گره‌ها همان روترها و یال‌ها جریان ارتباطی بین آنها را نشان می‌دهند. تعداد نماهای این مجموعه ۷۳۳ نما است که بیشترین تعداد نما را در بین مجموعه

داده‌ها به خود اختصاص داده است. در مقابل کمترین تعداد گره‌ها در بین مجموعه داده‌ها مورد ارزیابی دارد.

مجموعه داده	Autonomous Systems AS-73	Autonomous Systems - Oregon-1
تعداد گره‌ها	6474	11174
تعداد یال‌ها	13895	23409
تعداد نماها	733	9
نوع	جهت‌دار، مبتنی بر زمان، بدون وزن	بدون‌جهت ، مبتنی بر زمان، بدون وزن

جدول ۱. **مجموعه داده‌های ارجاع مقالات**

مجموعه داده	Arxiv HEP-TH	Arxiv HEP-PH
تعداد گره‌ها	27770	34546
تعداد یال‌ها	352807	421578
تعداد نماها	62	124
نوع	جهت‌دار، مبتنی بر زمان ، بدون وزن	جهت‌دار، مبتنی بر زمان ، بدون وزن

جدول ۲. مجموعه داده‌های مسیریابی

۲-۵ مجموعه داده‌های مصنوعی

مجموعه داده‌های مصنوعی مجزا در این مقاله در نظر گرفته شده است که پارامترهای مورد نیاز برای مجموعه داده‌های مصنوعی براساس LFR معرفی شده در [۳۰] مشخص شده است. پارامترهای اصلی برای ایجاد این مجموعه داده در جدول ۳ آمده است. سمبل S معادل با تعداد نماهای مورد نیاز برای ایجاد این گونه از شبکه‌هاست. در جدول ۳، سمبل S، مقدار ۱۵ را گرفته است که نشان‌دهنده‌ی این است که مجموعه داده‌ی مصنوعی شامل ۱۰ نما است. پارامتر دیگری که برای ایجاد مجموعه داده‌های مصنوعی استفاده می‌شود، پارامتر n است که تعداد گره‌ها را نشان می‌دهد. پارامتر n برای مجموعه داده‌ی مصنوعی برابر با ۶۰۰۰ است. پارامتر k، پارامتر بعدی است که میانگین درجات گره‌ها را نشان می‌دهد که برای این مجموعه داده، میانگین درجات ۳۰ است.

سمبل $maxd$ ماکزیمم درجات یک گره را نشان می‌دهد. سمبل $minc$ و $maxc$ به ترتیب مینیمم و ماکزیمم اندازه‌ی یک جامعه را نشان می‌دهد. همچنین

پارامتر	مقدار
S	10
N	6000
k	30
maxd	50
minc	25
maxc	80
O_{nc}	3
μ	0.3

جدول ۳. پارامترهای مرکزی برای ایجاد مجموعه داده‌های مصنوعی

سمبل O_n نشان‌دهنده‌ی تعداد گره‌هایی است که عضو بیش از یک جامعه هستند سمبل O_{nc} تعداد جوامعی که گره‌های فوق‌الذکر به آن تعلق دارد را بیان می‌کند. در جدول ۴ پارامترهای سفارشی آمده است که تعداد هر یک از رویدادها در زوج مرتب رویدادها مشخص شده است.

رویدادها	تعداد
ایجاد و مردن جوامع قبلی	۲۰
گسترش و افول جوامع قبلی	**۲۰**
گسترش و زنده ماندن جوامع قبلی	۲۰
تغییر یک جامعه	۲۰

جدول ۴. **پارامترهای سفارشی استفاده شده برای ایجاد مجموعه داده‌های مصنوعی**

۵_۳ معیارها

به طور کلی دو معیار برای ارزیابی در نظر گرفته می‌شود. معیار اول، صحت و دومین معیار، معیار زمان اجرا است. صحت جامعه نشان می‌دهد که جوامع تشخیص داده‌شده چقدر با جوامع واقعی مطابقت دارند. تعیین صحت جامعه در صورتی که جوامع واقعی وجود داشته باشند، می‌تواند بهتر انجام شود. اگر جوامع حقیقی وجود داشته باشند، می‌توان از معیار اطلاعات متقابل نرمال‌شده استفاده کرد. هر چه شباهت بین جوامع تشخیص داده شده و جوامع واقعی بیشتر باشد، معیار اطلاعات متقابل نرمال‌شده بیشتر خواهد بود. معیار اطلاعات متقابل نرمال‌شده بین صفر تا یک است. بنابراین هر چه این مقدار به یک نزدیک‌تر باشد، جوامع تشخیص‌داده‌شده و جوامع حقیقی به هم شبیه‌تر هستند. این معیار به صورت ذیل تعریف می‌شود [۳۱]:

$$NMI = \frac{-2\sum_{i,j} N_{i,j} log\frac{N_{i,j}N_t}{N_i N_j}}{\sum_i N_i log\frac{N_i}{N_t} + \sum_{i,j} N_j log\frac{N_j}{N_t}} \qquad (۳)$$

که N ماتریس مطابقت جوامع تشخیص داده و جوامع حقیقی است. که هر درایه‌ی آن $N_{i,j}$تعداد گره‌های مشترک بین جوامع تشخیص‌داده‌شده i و جامعه‌ی حقیقی jرا نشان می‌دهد. نمادهای N_i و N_j مجموع درایه‌های هر سطر را مشخص می‌کند.

برای تعیین کیفیت افزار جامعه در صورتی که جوامع حقیقی موجود نباشند، می‌توان از یک معیار به نام معیار ماژولاریتی استفاده کرد. معیار ماژولاریتی مطابقت ساختاری جوامع تشخیص‌داده‌شده را با تعریف جوامع بررسی می‌کند. پیش‌فرض این معیار، این است که جوامع به صورت مجزا هستند و هیچ همپوشانی ندارند. این معیار به صورت ذیل تعریف می‌شود [۳۲]:

$$Q = \frac{1}{2m}\sum_{ij}(A_{ij} - \frac{k_i k_j}{2m})\delta_{ij} \qquad (۴)$$

ماتریس A_{ij} یک ماتریس مجاورت است و m تعداد یال‌ها را مشخص می‌کند. نمادهای k_iو k_j درجه‌ی گره‌های i و jرا بیان می‌کند. همچنین نماد δ_{ij} همان دلتای کرونکر است که اگر دو گره برچسب یکسان داشته باشند، برابر با ۱ خواهد شد، در غیر اینصورت برابر با صفر خواهد بود. دومین معیار، سرعت تشخیص است که متناظر با زمان اجرا است. هر چه زمان اجرا بیشتر باشد، سرعت تشخیص متعاقبا کمتر است.

۵-۴ نتایج بر روی مجموعه داده‌های واقعی

مجموعه‌های داده‌ای واقعی در این مقاله شامل دو مجموعه داده‌ی ارجاع مقالات و مسیریابی است. در این مجموعه‌ها، معیار ماژولاریتی، اطلاعات متقابل نرمال‌شده و زمان اجرا برای ارزیابی روش ارائه‌شده و دیگر روش‌های موجود مورد استفاده قرار گرفته است.

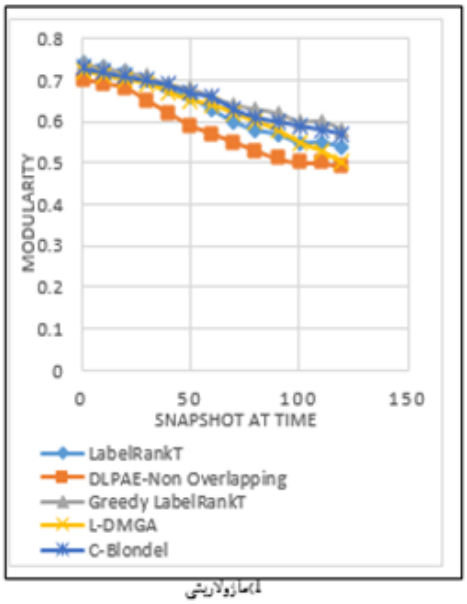
۵-۴-۱ نتایج بر روی مجموعه داده‌های ارجاع مقالات

براساس داده‌ی مربوط به شکل ۱ قسمت الف، مشخص است که روش پیشنهادی در مجموعه داده‌ی Arxiv HP-PH نسبت به دیگر روش‌ها، ماژولاریتی بیشتری دارد. این بهبود ماژولاریتی ناشی از در در نظر گرفتن ساختار داخلی شبکه است. همچنین براساس قسمت ب همین شکل می‌توان گفت که زمان اجرای روش پیشنهادی کمی بیشتر از رتبه‌بندی برچسب مبتنی بر زمان است. این موضوع به دلیل حجم محاسبات ماژولاریتی محلی است.

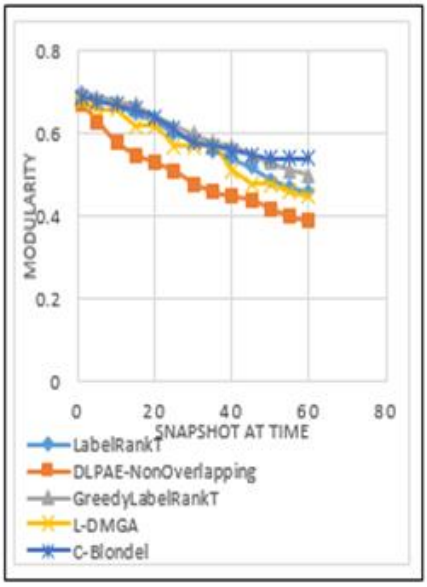
در مجموعه داده‌ی Arxiv HP-TH مشابه مجموعه داده‌ی قبلی، روش پیشنهادی ماژولاریتی بیشتری نسبت به دیگر روش‌های موجود در اکثر نماها دست پیدا کرده است. (شکل ۲ قسمت ۱) فاصله‌ی بین روش پیشنهادی و روش رتبه‌بندی مبتنی بر زمان تقریبا ۱ درصد است. روش C-Blondel عملکرد نزدیکی نسبت به روش پیشنهادی داشته است و حتی در نماهای آخر عملکردش بهتر هم بوده است. زمان اجرای روش پیشنهادی نسبت به دیگر روش‌های موجود خیلی بیشتر نیست (شکل ۲ قسمت ۲).

آنچه که براساس شکل ۱ و ۲ می‌توان گفت، این است که یک روند واضحی از کاهش ماژولاریتی در مجموعه داده‌های ارجاع مقالات وجود دارد. برای هر روش، میزان کاهش در مجموعه داده‌ی Arxiv HEP-PH نسبت به مجموعه داده‌ی Arxiv HEP-TH بیشتر است. با این وجود، زمان اجرای همه‌ی روش‌ها در مجموعه داده‌ی Arxiv HEP-PH بیشتر از Arxiv HEP-TH به طور میانگین است. در هر دو شبکه‌های داده‌ای ارجاع مقالات، روش پیشنهادی نسبت به دیگر روش‌ها، کیفیت افراز بهتری داشته است. این موضوع می‌تواند به دلیل استفاده از ماژولاریتی محلی باشد. بنابراین با توجه به خصوصیت مجموعه داده‌ی ارجاع مقالات، می‌توان نتیجه‌گیری کرد

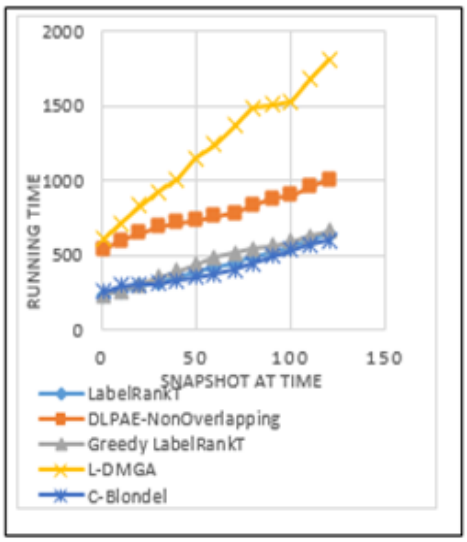
که رویداد اضافه‌کردن گره‌ها، تاثیر نامناسبی بر روی روش پیشنهادی نداشته است.



1)ماژولاریتی

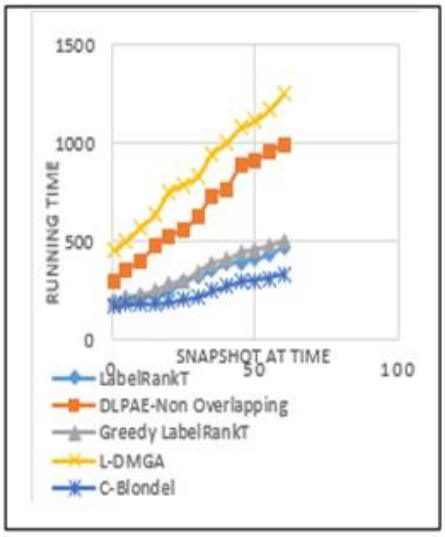


1) ماژولاریتی



2) زمان اجرا

شکل 1.ماژولاریتی و زمان اجرای یدست‌آمده توسط روش پیشنهادی و دیگر روش‌های موجود بر روی مجموعه داده‌ی Arxiv HEP-PH



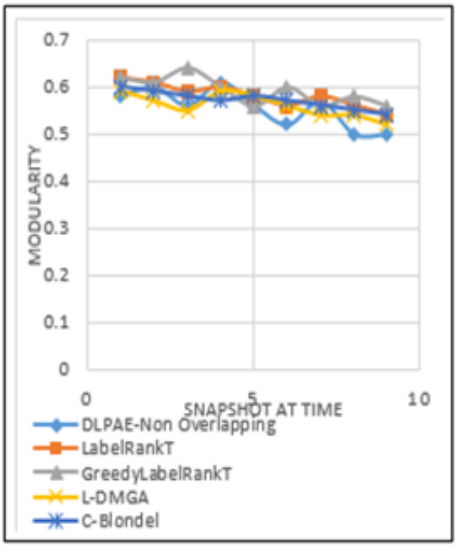
2) زمان اجرا

شکل 2ماژولاریتی و زمان اجرای یدست‌آمده توسط روش پیشنهادی و دیگر روش‌های موجود بر روی مجموعه داده‌ی Arxiv HEP-TH

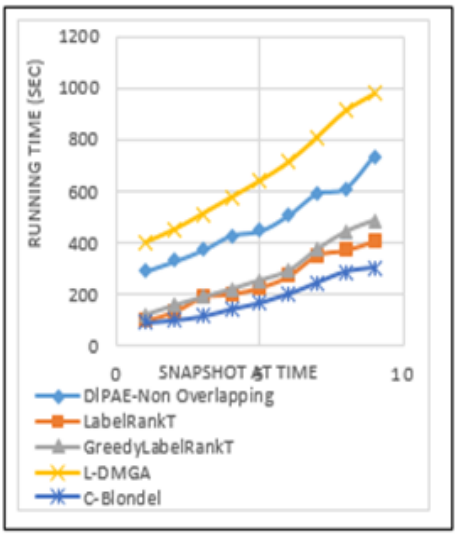
۵-۴-۲. نتایج بر روی مجموعه داده‌های مسیریابی

همانطور که از شکل ۳ قسمت الف مشخص است، هر یک از روش‌های ارائه شده در مجموعه داده‌ی Autonomous System-Oregon-1 به طور نامنظم بالا و پایین می‌روند. در نزدیک نصف نماها، روش پیشنهادی ماژولاریتی بهتری نسبت به روش مبتنی بر زمان دارد. همچنین روش پیشنهادی نسبت به روش رتبه‌بندی برچسب مبتنی بر زمان، نسخه‌ی غیرهمپوشان روش انتشار برچسب برجسته‌ی گسترش‌یافته، روش الگوریتم ژنتیک چند هدفه پویا مبتنی بر انتشار برچسب و روش C-Blondel در اکثر نماها عملکرد بهتری داشته است. همچنین روش پیشنهادی نسبت به نسخه‌ی غیرهمپوشان روش انتشار برچسب برجسته‌ی گسترش‌یافته و روش الگوریتم ژنتیک چند هدفه پویا مبتنی بر انتشار برچسب سریعتر عمل کرده است و نسبت به روش رتبه‌بندی برچسب مبتنی بر زمان و روش C-Blondel، کندتر است (شکل ۳ قسمت **2**).

همچنین با توجه به اینکه در مجموعه داده‌ی Autonomous Systems AS-733 تعداد نماها خیلی بیشتر از دیگر مجموعه‌های داده‌ای و تعداد گره‌ها نیز کمتر است و حجم تغییرات نیز متعاقبا بسیار کم است. بنابراین عملکرد مناسب روش پیشنهادی نسبت به روش‌های موجود در نیمی از نماها بهتر عمل کرده است (شکل ۴ قسمت **1**). به طور کلی در مجموعه‌ی داده‌ای مسیریابی می‌توان گفت که در حالتی که تغییرات بسیار کم باشد، روش پیشنهادی تنها در نیمی از نماها بهتر عمل کرده است. این مسئله به این دلیل است که در حالتی که تغییرات بسیار کم است، نمی‌توان انتظار بهبود زیادی را داشت. زمان اجرای روش پیشنهادی نزدیک به روش رتبه‌بندی برچسب مبتنی بر زمان است (شکل ۴)



ماژولاریتی (1

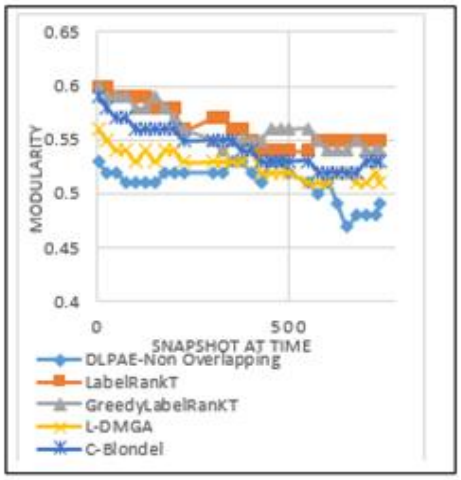


زمان اجرا (2

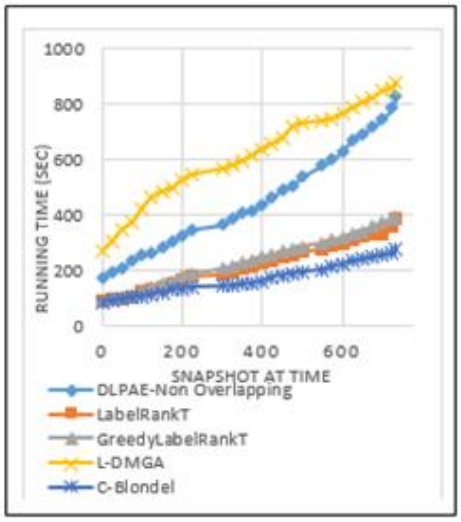
شکل 3. ماژولاریتی و زمان اجرای بدست‌آمده توسط روش پیشنهادی و دیگر روش‌های موجود بر روی مجموعه داده‌ی Autonomous System-Oregon-1

۵-۵ نتایج بر روی مجموعه داده‌های مصنوعی

برای مطالعه‌ی کارایی روش ارائه‌شده، نتایج بر روی ۴ شبکه‌ی پویا پیاده‌سازی شده است. معیار اطلاعات متقابل نرمال‌شده برای کلیه‌ی روش‌ها در این مجموعه‌ی داده‌ای بدست آمده است. آنچه که از شکل ۵ قسمت ۴ قابل توجه است، این است که معیار اطلاعات متقابل نرمال‌شده روش ارائه‌شده و دیگر روش‌ها نزدیک به هم است. در مقابل براساس اشکال ۶ قسمت‌های ۱ و ۲ و ۳ می‌توان گفت که معیار اطلاعات متقابل نرمال‌شده برای روش پیشنهادی بیشتر از روش‌های موجود در اکثر نماها است. براساس موارد فوق می‌توان گفت که روش ارائه‌شده در همه‌ی رویدادها به غیر از رویداد تغییر جامعه نسبت به دیگر روش‌ها بهتر عمل کرده است. این موضوع نقش ماژولاریتی محلی را در بهبود رویدادهای گسترش و ایجاد جوامع را نشان می‌دهد. شکل ۶ هم نشان می‌دهد که زمان اجرا برای روش پیشنهادی به طور میانگین بر روی ۴ مجموعه داده‌ی مصنوعی نسبت به دیگر روش‌های موجود بهتر عمل کرده است. روش پیشنهادی از لحاظ زمان اجرا به طور میانگین نزدیک به روش رتبه‌بندی برچسب مبتنی بر زمان عمل کرده است.

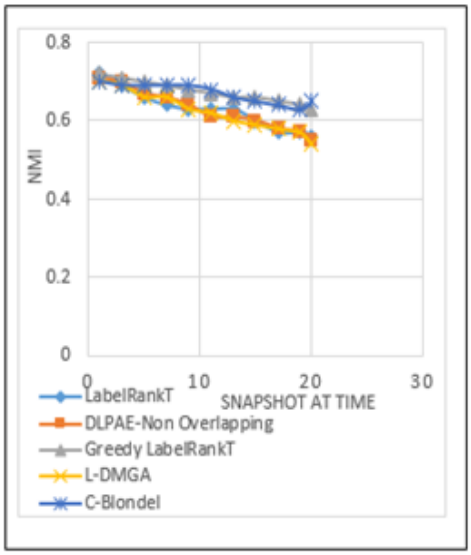


1) ماژولاریتی

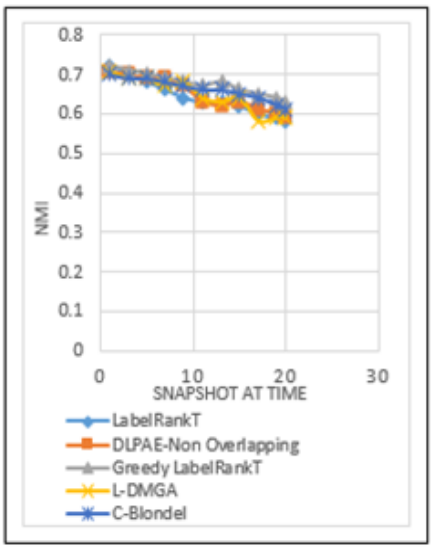


2) زمان اجرا

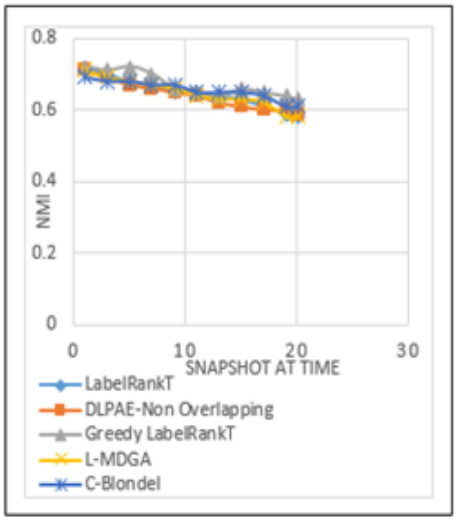
شکل 4. ماژولاریتی و زمان اجرای پدست‌آمده توسط روش پیشنهادی و دیگر روش‌های موجود بر روی مجموعه داده‌ی Autonomous Systems AS-733



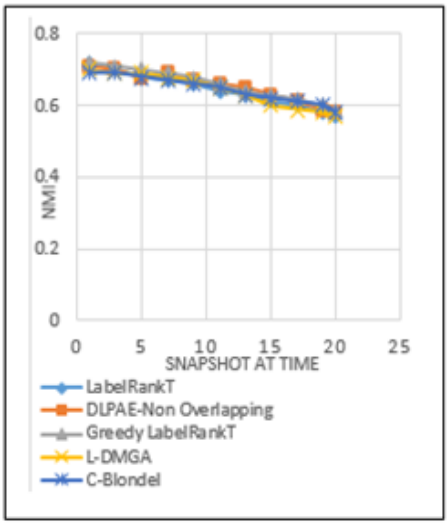
1) ایجاد و مردن جوامع



3) گسترش و زنده ماندن جامعه

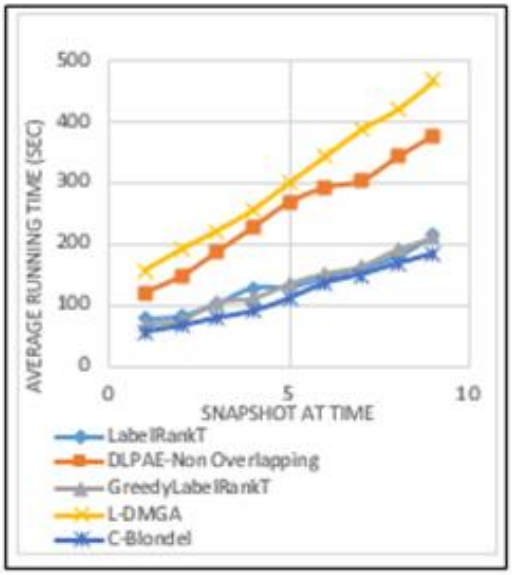


2) گسترش و افول جوامع



4) تغییر یک جامعه

شکل 5. معیار اطلاعات متقابل نرمال‌شده برای روش پیشنهادی و دیگر روش‌های موجود بر روی مجموعه داده-های مصنوعی



شکل 6. میانگین زمان اجرای بدست آمده توسط روش پیشنهادی و دیگر روش‌های موجود بر روی مجموعه داده‌های مصنوعی

۶. نتیجه‌گیری

تشخیص جوامع یکی از موضوعات مهم در آنالیز شبکه‌های اجتماعی است. در این مقاله، یک روش مبتنی بر انتشار برچسب و بهینه‌سازی ماژولاریتی برای تشخیص جوامع به صورت پویا در شبکه‌های اجتماعی ارائه شده است. در روش پیشنهادی، یک معیار حریصانه به روش رتبه‌بندی برچسب مبتنی بر زمان اضافه شده است. در این معیار، هدف بهبود ماژولاریتی محلی است. نتایج نشان می‌دهد که این روش کیفیت افراز جوامع را در شبکه‌ی واقعی و معیار اطلاعات متقابل نرمال‌شده را در شبکه‌های مصنوعی بهبود می‌دهد. علاوه بر این، زمان اجرای روش پیشنهادی کمی بیشتر از روش رتبه‌بندی برچسب مبتنی بر زمان است که فاصله‌ی خیلی محسوسی نیست. به طور کلی، روش پیشنهادی صحت را در مقایسه با دیگر روش‌های موجود در تقریبا همه‌ی مجموعه داده‌های مورد ارزیابی اعم از واقعی و مصنوعی بهبود داده است. تشخیص جوامع به صورت پویا هنوز در ابتدای راه قرار دارد و مسلما در آینده شاهد روش‌های بیشتری در این زمینه خواهیم بود.

مراجع

[۱] M. E. Newman, and M. Girvan, "Finding and evaluating community structure in networks", Physical review E, vol.69 , no.2, p.026113, 2004.

[۲] S. Fortunato, “Community detection in graphs,” *Physics Reports*, pp. 75-174,2010.

[۳] N. P. Nguyen, T. N. Dinh, Y. Shen, and M. T. Thai, “Dynamic social community detection and its applications.”, *PLoS One*, vol. 9, no.4, p. e91431, 2014.

[۴] V. D. Blondel, J. Guillaume, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of Statistical Mechanics: Theory and Experiment*, no. 10, pp. 1–12, 2008.

[۵] S. Soundarajan, and J. E. Hopcroft, "Use of local group information to identify communities in networks". *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 9, no.3, pp 1-35, 2015.

[۶] U. N. Raghavan, R. Albert, and S. Kumara, Near linear time algorithm to detect community structures in large-scale networks, *Physical review E*, vol. 76, no.3, p.036106, 2007.

[۷] M. Sattari, K. Zamanifar, “A spreading activation-based label propagation algorithm for overlapping community detection in dynamic social networks”, *Data & Knowledge Engineering*, vol. 113, pp.155-170, 2018.

[۸] M. Sattari, K. Zamanifar. “A cascade information diffusion based label propagation algorithm for community detection in dynamic social networks”, *Journal of Computational Science*. vol. 25, pp. 122-33, 2018.

[۹] X. Ma, D. Dong, “Evolutionary nonnegative matrix factorization algorithms for community detection in dynamic networks”. IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 5, pp. 1045-58, 2017.

[۱۰] G. Rossetti, L. Pappalardo, Pedreschi D, Giannotti F. “Tiles: an online algorithm for community discovery in dynamic social networks”, *Machine Learning*. vol. 106, no. 8, pp. 1213-41, 2017.

[۱۱] K. Kuzmin, S. Y. Shah, and B. K. Szymanski,"Parallel overlapping community detection with SLPA". *In Social Computing (SocialCom)*, pp. 204-212, 2013.

[۱۲] J. Xie, M. Chen, and B. Szymanski, “LabelrankT: Incremental community detection in dynamic networks via label propagation,” *In Proceedings of the ACM Workshop on Dynamic Networks*

Management and Mining, ACM, pp. 25–32, 2013.

[۱۳] J. Xie, and B. K. Szymanski, "LabelRank: A stabilized label propagation algorithm for community detection in Networks," *In Network Science Workshop (NSW)*, IEEE, pp. 138-143, 2013.

[۱۴] M. Takaffoli, R. Rabbany, and O. R. Za, "Incremental local community identification in dynamic social networks." *In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, IEEE/ACM pp. 90-94, 2013.

[۱۵] H. Alvari, A. Hajibagheri, and G. Sukthankar, "Community detection in dynamic social networks: A game-theoretic approach," *In Advances in Social Networks Analysis and Mining (ASONAM), IEEE/ACM* , pp. 101–107, 2014.

[۱۶] R. I. Lung, C. Chira, and A. Andreica, "Game theory and extremal optimization for community detection in complex dynamic networks". *PloS one*, vol.9, no.2, pp. 891-901, 2014.

[۱۷] M. Takaffoli, R. Rabbany, and O. R. Za, "Incremental local community identification in dynamic social networks." *In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, IEEE/ACM pp. 90-94, 2013.

[۱۸] J. He and D. Chen, "A fast algorithm for community detection in temporal network", *PhysicaA: Statistical Mechanics and its Applications*, vol. 429, pp. 87–94, 2015.

[۱۹] Seifikar, Mahsa, Saeed Farzi, and Masoud Barati. "C-Blondel: An Efficient Louvain-Based Dynamic Community Detection Algorithm." IEEE Transactions on Computational Social Systems 7.2 (2020): 308-318.

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

[۲۱] LXY.Leung, P.Hui, P.Liu , and J.Crowcroft, "Towards real-time community detection in large networks.", *Physical Review E* , vol. 79, no. 6, p.066107, 2009.

[۲۲] J. Han, W. Li, L. Zhao, Z. Su, Y. Zou, and W. Deng, "Community detection in dynamic networks via adaptive label propagation", *PloS one*, vol.12, no.11, p.e0188655,2017.

[۲۳] K. Liu, J. Huang, H. un, M. Wan, Y. Qi, and H. Li, "Label propagation based evolutionary clustering for detecting overlapping and non-overlapping communities in dynamic networks," *Knowledge-Based Systems*, vol. 89, pp. 487–496, 2015.

[۲۴] S. H.L, H. J.B, T. Y.Q., Q.B., and L.Huai-Liang, "Detecting overlapping communities in networks via dominant label propagation," *Chinese Phys.B*, vol. 24, no.1, p. 18703, 2015.

[۲۵] X. Niu, W. Si, and C.Q. Wu, 2017. A label-based evolutionary computing approach to dynamic community detection. *Computer Communications, Vol. 108, pp.110-122, 2017.*

[۲۶] M. Takaffoli, R. Rabbany, and O. R. Za, "Incremental local community identification in dynamic social networks." In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, IEEE/ACM pp. 90-94, 2013.

[۲۷] X. Niu, W. Si, and C.Q. Wu, "A label-based evolutionary computing approach to dynamic community detection", Computer Communications, Vol. 108, pp.110-122, 2017.

[۲۸] J. Gehrke, P. Ginsparg and J. Kleinberg, "Overview of the 2003 KDD Cup", *ACM SIGKDD Explorations Newsletter*, vol. 5, no. 2, pp.149-151, 2003.

[۲۹] J. Leskovec, J. Kleinberg and C. Faloutsos. "Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations.", *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, ACM, pp. 177-187, 2005.

[۳۰] A. Lancichinetti, S. Fortunato, "Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities", *Phys. Rev. E*, vol. 80, no.1, p. 016118, 2009.

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

[۳۲] M. Newman, "modularity and community structure in networks". *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577 –8582, 2002.

An Improved Method for Community Detection in Dynamic Social Networks Based on Label Propagation and Modularity Optimization

Abstract

Community detection in dynamic social networks is one of the most important research topics that has been considered in recent years. There are various approaches to detecting communities in dynamic social networks, among which the label propagation approach has chosen due to simplicity and efficiency. This approach involves many methods that are often random. Among these methods, LabelRankT(Time) is a deterministic method. Of course, this method also has problems, one of the problems is that when a node wants to join a community, the internal structure of that community is not considered. So, for solving this problem, a greedy approach has been used to improve the label publishing approach. The proposed approach and other evaluated methods are tested in real and Synthetic datasets. The results show that the proposed method has performed relatively better than the other methods in terms of accuracy and modularity.

***Keywords*:** GreedyLabelRankT, LabelRankT, Label Propagation Approach, Community Detection, Dynamic Social Network.

نویسنده مسئول: محمد ستاری
msattarimng.mui@gmail.com