

مقاله علمی-ترویجی

بهبود خوشه‌بندی شبکه‌های اجتماعی با الگوریتم تکاملی رقابت استعماری و معیار شباهت درونی گره‌های شبکه

محمدامین شایگان^{*}، گروه مهندسی کامپیوتر، واحد شیراز، دانشگاه آزاد اسلامی، شیراز، ایران، shayegan@iaushiraz.ac.ir
 علی حسینی، باشگاه پژوهشگران جوان و نخبگان، واحد شیراز، دانشگاه آزاد اسلامی، شیراز، ایران، a.hosseini@iaushiraz.ac.ir
 سعید صدیقی، باشگاه پژوهشگران جوان و نخبگان، واحد شیراز، دانشگاه آزاد اسلامی، شیراز، ایران، s.sedighi@iaushiraz.ac.ir
 نویسنده مسوول

چکیده

به دلیل رشد روز افزون تمایل افراد به عضویت و استفاده از شبکه‌های اجتماعی، برقراری ارتباط و به اشتراک‌گذاری داده‌های موجود در این شبکه‌ها، مورد توجه علوم مختلف همانند علوم سیاسی، روانشناسی، جامعه‌شناسی، اقتصاد و ... قرار گرفته است. به همین دلیل، محققین اقدام به تشخیص و استخراج روابط بین افراد از داده‌های موجود در این شبکه‌ها، برای ایجاد جوامع دقیق‌تر نموده‌اند. با این حال هنوز روشی موثر جهت شناسایی و استخراج جوامع، بر مبنای داده‌های شبکه‌های اجتماعی ارائه نشده است. در این مقاله، به منظور خوشه‌بندی دقیق‌تر جوامع موجود در یک شبکه اجتماعی، روشی بر پایه استفاده از الگوریتم فرا ابتکاری تکاملی رقابت استعماری و با انتخاب جمعیت اولیه بر اساس معیار خوشه‌بندی مبتنی بر چگالی معرفی شده است. روش پیشنهادی در مقایسه با الگوریتم رقابت استعماری پایه، به‌طور میانگین مقدار ماژولاریتی را ۲۱٫۴۵٪ افزایش داده و جوامع منسجم‌تری را استخراج نموده است.

کلمات کلیدی: الگوریتم رقابت استعماری، الگوریتم‌های تکاملی، خوشه‌بندی گراف، خوشه‌بندی مبتنی بر چگالی، شبکه‌های اجتماعی

Social Network Clustering Enhancement by using Imperial Competitive Evolutionary Algorithm and Inter-Similarity of Network Nodes

— Mohammad Amin Shayegan^{*}, Department of Computer, Faculty of Engineering, Shiraz Branch, Islamic Azad University/ Shiraz-Iran, Shayegan@iaushiraz.ac.ir
 — Ali Hosseini, Young Researchers and Elite Club, Shiraz Branch, Islamic Azad University, Shiraz- Iran/a.hosseini@iaushiraz.ac.ir
 — Saeid Sedighi, Young Researchers and Elite Club, Shiraz Branch, Islamic Azad University, Shiraz- Iran, s.sedighi@iaushiraz.ac.ir
^{*}Corresponding Author

Abstract

Abstract: Due to the growing desire of people to join and use social networks, communication and shar-

ing data in these networks has been considered by various sciences such as political science, psychology, sociology, economics, etc. Hence, researchers have begun to distinguish and extract relationships between individuals from the data contained in these networks, to create more accurate communities. However, there is still no effective method to identify and extract communities based on social media data.

In this article, a method has been proposed for social network accurate clustering by using Imperial Competitive Evolutionary Algorithm (ICEA) and selecting the initial population based on the density-based clustering criterion. The proposed method has improved the result of modularity about 21.45% in average, compared to rival basic ICEA and extracted more denser communities.

Keywords: Imperial Competitive Algorithm, Evolutionary Algorithms, Graph Clustering, density-based clustering, Social Networks

۱- مقدمه

با افزایش استفاده از شبکه‌های اجتماعی^۱ و گرایش افراد به عضویت و به اشتراک گذاری اطلاعات در آن‌ها، ارزیابی و تجزیه و تحلیل داده‌های موجود در این شبکه‌ها، مورد توجه علوم مختلفی همچون جامعه‌شناسی، روانشناسی، اقتصاد، جرم‌شناسی و... قرار گرفته است. چرا که می‌توان از این اطلاعات استخراج شده، در حوزه‌های مختلفی همچون صنعت، اقتصاد، سیاست و یا مسایل امنیتی استفاده نمود.

به منظور استخراج و ارایه راه حلی بهینه برای تجزیه و تحلیل شبکه‌های اجتماعی و یافتن جوامع مختلف موجود در آن‌ها، تا به حال الگوریتم‌های تکاملی^۲ و روش‌های بهینه‌سازی^۳ مختلفی معرفی و استفاده شده است. لیکن مشکلی که امروزه محققین با آن‌ها روبه رو هستند آن است که هنوز یک الگوریتم بهینه که بتواند داده‌های موجود در شبکه‌های اجتماعی را تجزیه، تحلیل و در نهایت جوامع منسجم را تشخیص و استخراج نماید معرفی نشده است [۱].

یکی از ابزارهای مهم جهت نمایش داده‌های حجیم و ارتباطات مابین آن‌ها «گراف» است. در حالت کلی می‌توان از گراف‌ها در نمایش رفتارهای بیولوژیکی پیچیده، سیستم‌های اقتصادی، اجتماعی و تکنولوژیکی استفاده نمود. ساختار شبکه‌های اجتماعی نیز بر پایه خوشه‌بندی^۴ گراف پیاده‌سازی می‌شود. در این راستا، خوشه‌بندی، ابزاری برای کشف ساختار از درون داده‌ها، بدون نیاز به هیچ پیش‌فرضی است. لیکن پایه و اساس خوشه‌بندی سنتی، تشابه اشیاء بوده و روش‌های سنتی خوشه‌بندی را نمی‌توان بر شبکه‌های اجتماعی اعمال نمود، چرا که خوشه‌بندی در شبکه‌های اجتماعی، به صورت پویا و در جهانی واقعی بر اساس ارتباطات موجود بین افراد انجام می‌شود [۱].

در نظریه گراف، خوشه‌بندی داده به خوشه‌بندی گراف، گسترش یافته که این مبثی متفاوت از خوشه‌بندی داده است. معیار خوشه‌بندی در هر خوشه از گراف، بر اساس میزان تراکم لبه‌ها می‌باشد. خوشه‌بندی گراف به دنبال به حداکثر رساندن شباهت پیوندهای داخلی و همزمان به حداقل رساندن تعداد پیشوندهای خارجی است که به‌عنوان پل ارتباطی بین جوامع عمل می‌کنند. با این حال، یافتن این ارتباطات و تخصیص آن‌ها به خوشه‌های مجزا در مقیاس‌های بزرگ داده و اطلاعات حجیم، مبثی دشوار است.

خوشه‌بندی روش‌های متفاوتی دارد که یکی از آن‌ها خوشه‌بندی مبتنی بر چگالی^۵ است. از مهمترین ویژگی‌های الگوریتم خوشه‌بندی مبتنی بر چگالی، توانایی تشخیص خوشه‌ها با اشکال دلخواه، خوشه‌بندی داده‌های همراه با نویز و پیچیدگی زمانی کم می‌باشد [۲]. فرآیند بهینه‌سازی به معنی تغییر دادن ورودی‌ها و خصوصیات یک سیستم به نحوی است که بهترین خروجی یا نتیجه به دست آید.

ورودی‌ها، متغیرهای فرآیند یا تابع مورد بررسی هستند که این نتایج عموماً با نام‌های تابع هدف^۶، تابع هزینه^۷ و یا تابع برازندگی^۸ نامیده می‌شود. خروجی نیز به صورت هزینه، سود و یا برازندگی تعریف می‌شود.

روش‌ها و الگوریتم‌های بهینه‌سازی به دو دسته الگوریتم‌های دقیق و الگوریتم‌های تقریبی، تقسیم‌بندی می‌شود [۳]. الگوریتم‌های دقیق به روش‌هایی گفته می‌شود که بیشتر روش‌های ریاضی جهت یافتن جواب بهینه برای آن‌ها وجود دارد، لیکن در مسایل پیچیده کاربرد ندارند. بر خلاف الگوریتم‌های دقیق، الگوریتم‌های تقریبی قادر به یافتن جواب‌های خوب (نزدیک به بهینه) در زمان کوتاه، برای مسایل بهینه‌سازی هستند. الگوریتم‌های تقریبی خود به دو دسته الگوریتم‌های ابتکاری^۹ و فراابتکاری^{۱۰} تقسیم‌بندی می‌شوند.

الگوریتم‌های فرا ابتکاری، یکی از انواع الگوریتم‌های بهینه‌سازی تقریبی هستند که دارای مکانیزم‌های خروج از بهینه محلی^{۱۱} بوده و قابل استفاده در طیف وسیعی از مسایل هستند. توانایی حل مسایل بسیار پیچیده، از قابلیت‌های دیگر الگوریتم‌های فرا ابتکاری می‌باشد [۴].

در این مقاله سعی شده با استفاده از الگوریتم فرا ابتکاری رقابت استعماری^{۱۲} و ترکیب آن با روشی نوین جهت شناسایی جمعیت اولیه و تابع برازندگی، تشخیص جوامع در شبکه‌های اجتماعی بهینه‌سازی گردد.

بخش‌های آتی این مقاله به شرح زیر می‌باشد. بخش دوم، مروری بر برخی تحقیقات انجام شده در حوزه خوشه‌بندی شبکه‌های اجتماعی است. در بخش سوم، روش پیشنهادی این مقاله جهت بهینه‌سازی تشخیص جوامع در مجموعه داده‌های شبکه‌های اجتماعی، با استفاده از الگوریتم فرا ابتکاری رقابت استعماری به تفصیل معرفی می‌گردد. بخش چهارم نتایج حاصل از روش پیشنهادی معرفی و این نتایج با برخی از نتایج روش‌های رقیب موجود در همین حوزه مقایسه و در پایان در بخش پنجم به مرور، جمع‌بندی و نتیجه‌گیری پرداخته شده است.

۲- ادبیات تحقیق

۲-۱- مبانی نظری

۲-۱-۱- الگوریتم تکاملی رقابت استعماری

این الگوریتم، با تعدادی جمعیت اولیه تصادفی که هر کدام از آن‌ها یک «کشور» نامیده می‌شوند، شروع می‌شود. تعدادی از بهترین عناصر جمعیت (معادل کروموزوم در الگوریتم ژنتیک) به‌عنوان استعمارگر و باقیمانده جمعیت نیز به‌عنوان مستعمره، در نظر گرفته می‌شوند. استعمارگران بسته به قدرتشان، این مستعمرات را با یک روند خاص به سمت خود جذب می‌کنند. قدرت کل هر استعمارگر،

شبهات درون خوشه‌ای و عدم تشابه برون خوشه‌ای را بر روی پایگاه داده‌های ترکیبی زاکاری و فوتبال [۹] و مجموعه داده تماس‌های صوتی کاربران واقعی موبایل [۱۰] ارزیابی کردند. روش پیشنهادی بر اساس الگوریتم خوشه‌بندی تکاملی چند هدفه MOEC، پایه‌گذاری شده و در مقایسه با مدل‌های مشابه دیگر به نتایج بهتری دست یافته است.

سعید و همکارانش [۶] با استفاده از الگوریتم ژنتیک به خوشه‌بندی شبکه‌های اجتماعی پرداختند. آن‌ها به منظور تولید جمعیت اولیه، از معیاری به نام ضریب خوشگی استفاده کردند. بالاترین میزان ماژولاریتی در این پژوهش، برابر با ۰/۹۰۵ بوده است که به نسبت دیگر الگوریتم‌ها، بهبود قابل توجهی داشته است.

شی و همکارانش [۱۱] تحقیقی در خصوص پیش‌بینی ارتباطات کاربران با استفاده از مفهومی به نام فاکتور ماتریس و الگوریتم بهینه‌سازی ذرات^{۱۴} ارائه کردند. آن‌ها نشان دادند که ماتریس فاکتورسازی، یک روش موثر در پیش‌بینی رابطه کاربران است. معیار ارزیابی در این مقاله دقت بوده که برابر با ۸۸٪ به دست آمده است. زامودیو و همکاران [۱۲] تحقیقی برای تشخیص جامعه در شبکه‌های اجتماعی ارائه کردند که از الگوریتم بهینه‌سازی ژنتیک برای کشف اجتماع استفاده می‌کند. بیشترین تشابه درون خوشه‌ای محاسبه شده در روش پیشنهادی، حدوداً ۷۲٪ بوده است.

رستی و همکارانش [۱۳] الگوریتمی برای کشف جامعه در شبکه‌های اجتماعی پویا معرفی کردند. پژوهش آن‌ها بر روی شبکه‌های واقعی اجرا و نتایج آن با استفاده از معیار F-measure با دیگر الگوریتم‌های مشابه مقایسه شده است. مقدار F-measure این روش بر روی مجموع داده Amazon برابر با ۰/۷۸، بر روی مجموع داده DBLP برابر با ۰/۸۰، بر روی مجموع داده یوتیوب برابر با ۰/۶۴ و بر روی مجموع داده‌های LiveJournal برابر با ۰/۷۳ بوده است.

نیومن [۱۴] با بهینه‌سازی پارامتر ماژولار تلاش به تشخیص جامعه در شبکه‌های اجتماعی کرد. این پژوهش بر روی مجموع داده‌های ایمیل، جاز، فوتبال، کتاب، بلاگ‌های سیاسی، دولفین و کاراته اجرا شده است. نتایج حاصل از روش حداکثر سازی ماژولاریتی بر روی مجموع داده‌های فوق، به ترتیب برابر با ۰/۷۸، ۰/۵۹، ۰/۶۷، ۰/۵۹، ۰/۳۶، ۰/۲۷، ۰/۱۹، ۰/۶۳، ۳/۳ بوده است. نتایج این مقاله نشان می‌دهد که حداکثر سازی ماژولاریتی، یک مورد خاص برای بالا بردن احتمال تشخیص جامعه است. در این تحقیق، همه جوامع با خصوصیات مشابه آماری در نظر گرفته شده‌اند.

حسینی و عباسی [۱۵] بر اساس توضیح پیام‌ها، یک روش جدید برای تشخیص جامعه در شبکه‌های اجتماعی ارائه و نتایج را به وسیله پارامتر ماژولاریتی ارزیابی کردند. میزان ماژولاریتی بر روی حجیم‌ترین مجموع داده برابر ۰/۸۵۴ بوده که در مقایسه با روش‌های مشابه، منجر به تولید نتایج مطلوب‌تری شده است.

ژو و همکارانش [۱۶]، الگوریتمی جدید جهت تشخیص جوامع محلی، که روابط اجتماعی و ویژگی‌های موجود در شبکه‌های اجتماعی را به هم متصل می‌کنند، ارائه کردند. تفاوت معیار سنجش تشابه در این مقاله با روش‌های مشابه، ترکیب روابط اجتماعی و ویژگی‌های هر گره، توام است. این در حالی است که متدهای ارائه شده پیشین، تنها از روابط اجتماعی به‌عنوان درجه هر گره در مجموع داده‌های شبکه‌های اجتماعی استفاده می‌کردند. نتایج این الگوریتم با معیار سیلوئت، که تمرکز آن بر میزان کیفیت خوشه‌بندی انجام شده متکی است، برابر با ۰/۴۵ بوده است (هر چه مقدار سیلوئت بالاتر باشد، کیفیت خوشه‌بندی نیز بالاتر است).

یولا و لی [۱۷] به خوشه‌بندی کاربران در شبکه‌های اجتماعی پرداختند. در آن پژوهش از مجموعه داده‌های توپیتور، فوتبال، کاراته و ... استفاده شده است. آنها تعاملات بین کاربران جهت پیش‌بینی رفتار آینده کاربران را به‌عنوان بخش مهمی از تجزیه و تحلیل شبکه‌های اجتماعی در نظر گرفته‌اند. با توجه به اینکه نمی‌توان به صحت همه تعاملات انجام شده در شبکه‌های اجتماعی اعتماد داشت، بنابراین به هر تراکنش، وزن مشخصی از «اعتماد» داده شده است. در این مدل، خوشه‌بندی کاربران بر اساس دو فاکتور «میزان علاقه‌مندی» و «وزن اعتماد» انجام شده است. از نتایج این روش می‌توان به جلوگیری از همگرایی زودرس و اختصاص مناسب گره‌ها به خوشه‌های مناسب اشاره کرد.

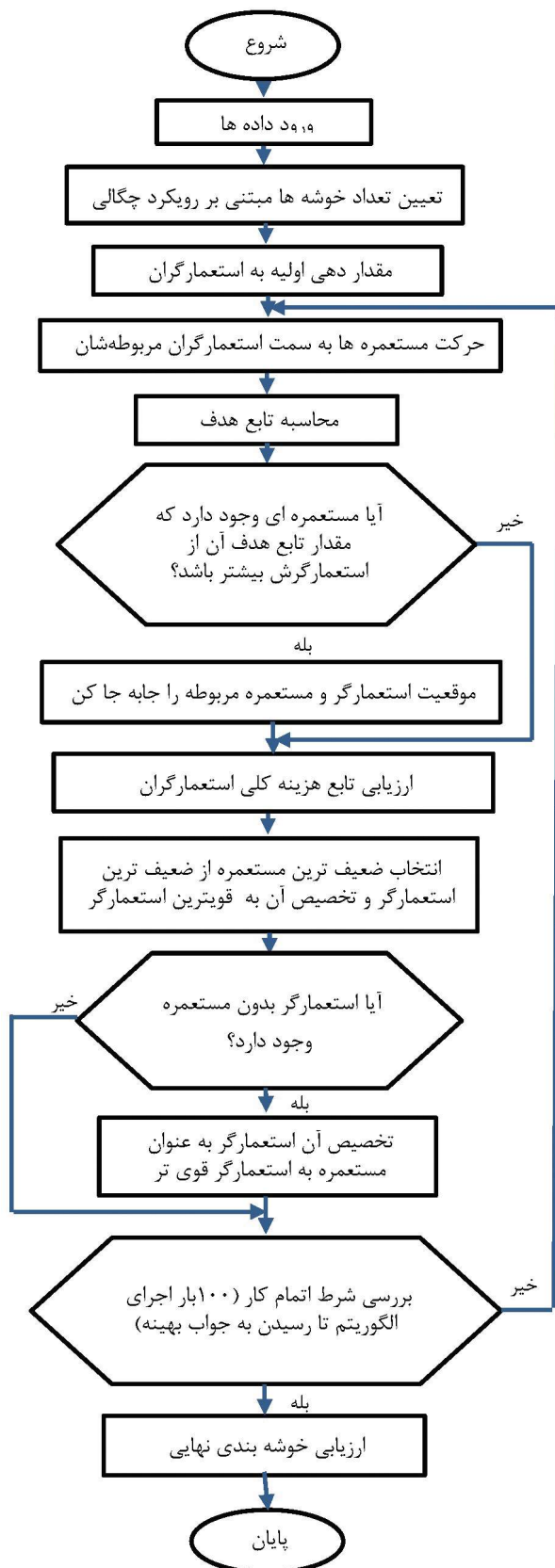
کوغان و همکارانش [۱۸] یک الگوریتم خوشه‌بندی سلسله مراتبی به نام CURE را معرفی کردند. این الگوریتم با مقادیر متفاوت پارامتر فاصله مجاورت، آزمایش شده که بهترین مقدار این پارامتر برابر با ۵ بوده است. نتایج این پژوهش با معیار تشابه برون خوشه‌ای بر روی مجموع داده‌های متفاوت محاسبه شده که در بهترین حالت، این معیار برابر ۰/۰۲۷ شده است. نتایج این پژوهش از نظر زمان اجرا نیز مورد ارزیابی قرار گرفته است که در بهترین حالت در مقایسه با دیگر الگوریتم‌ها، زمان ۰/۸۹ ثانیه بهبود پیدا کرده است.

سانچز و دوارت [۱۹] یک رویکرد فرابینکاری مبتنی بر روش حریمانه، برای تشخیص جوامع در شبکه‌های بزرگ اجتماعی ارائه کرده‌اند. این الگوریتم با استفاده از دو مجموعه داده، توپیتور و فیسبوک آزمایش شده و نتایج به دست آمده برای این دو مجموعه داده، با معیار ماژولاریتی برابر با ۰/۶۷۴۴ و ۰/۵۴۱۶ بوده است.

آزاد و همکارانش [۲۰] تحقیقی در زمینه خوشه‌بندی داده‌های حجیم شبکه اجتماعی با استفاده از الگوریتم مارکوف به صورت موازی ارائه دادند. خوشه‌بندی مارکوف شباهت‌های دنباله‌ای را بر اساس فاصله اقلیدسی در نظر گرفته و گره‌ها را خوشه‌بندی می‌کنند. این پژوهش بر روش شبکه HipMCL با ۷۰ میلیون گره و ۶۸ میلیارد یال اجرا شده است. معیار ارزیابی F-score در این پژوهش برابر با ۱ شده است. زالیک [۲۱] به منظور تشخیص جامعه در شبکه اجتماعی، با استفاده از گره مرکزی و الگوریتم تکاملی Net-Degree پژوهشی انجام داد.

در این پژوهش از یک الگوریتم تشخیص جامعه تکاملی چند هدفه جهت شناسایی جوامع مبتنی بر مرکز استفاده شده است. معیار گره مرکزی برای خوشه‌بندی بهینه و برای بهبود همگرایی الگوریتم تکاملی استفاده می‌شود. نخستین تابع استفاده شده، شمارش گره‌های غیرمتصل با گره مرکزی جامعه است. که برای یک ساختار خوب جامعه، این معیار باید کمینه باشد تا جوامع با دقت جداسازی شوند؛ بنابراین تابع نخست، کیفیت هر جامعه و تابع دوم، کیفیت هر گره را اندازه‌گیری می‌کند. این معیار نیز برای یک ساختار جامعه خوب باید کوچک باشد، چرا که در واقع میزان فاصله درون خوشه‌ای را محاسبه می‌کند. معیار ارزیابی در این پژوهش، ماژولاریتی است که با الگوریتم تک هدفه Net-degree مقایسه شده است. نتایج حاصل به ترتیب برای کیفیت جامعه و کیفیت گره در مجموع داده‌های کاراته، برابر با ۰/۷۴۱ و ۰/۶۹۳ برای مجموع داده دلفین، برابر با ۰/۴۰ و ۰/۴۶۵، برای مجموع داده کتاب، برابر با ۰/۴۸۳ و ۰/۴۶۵، برای مجموع داده جاز، برابر با ۰/۵۳۸ و ۰/۶۳۸ و در مجموع داده فوتبال برابر ۰/۴۵۶ و ۰/۶۴۱ به دست آمده است.

ترشیزی‌نژاد و همکاران [۲۲] با استفاده از الگوریتم رقابت استعماری، اقدام به تعیین اعداد جوامع در مجموع داده‌های یوتیوب و فیسبوک کردند. برای این منظور، آن‌ها در هنگام محاسبه هزینه کل امپراتوری،



شکل ۱: فلوچارت روش پیشنهادی

از ضریب خوشه‌بندی و گشت بسته استفاده کردند. معیار ارزیابی روش آن‌ها معیار اطلاعات متقابل هنجارسازی شده و ماژولاریتی بوده است. نتایج حاصل از آن تحقیق نشان می‌دهد که روش پیشنهادی از نظر معیار اطلاعات متقابل هنجارسازی شده بر روی مجموعه داده یوتیوب، در مقایسه با روش Walktrap به میزان ۰/۰۶۵ درصد و بر روی مجموعه داده فیسبوک در مقایسه با روش COPRA به میزان ۰/۰۲۸ درصد بهبود داشته است.

۳- روش پیشنهادی

با توجه به اهمیت تجزیه و تحلیل داده‌های شبکه‌های اجتماعی و استخراج جوامع منسجم، در این تحقیق رویکردی جدید بر اساس ترکیب روش خوشه‌بندی مبتنی بر چگالی، با الگوریتم رقابت استعماری معرفی شده است. تفاوت عمده این روش با سایر کارهای مشابه، انتخاب جمعیت اولیه بر اساس معیار خوشه‌بندی مبتنی بر چگالی است. یعنی بر خلاف روال طبیعی سایر الگوریتم‌های تکاملی، که مجموع اولیه آن‌ها به تصادف انتخاب می‌شود، روش پیشنهادی با تکیه بر خوشه‌بندی مبتنی بر چگالی و آرایه تابع هدفی که تکیه اصلی آن بر تشابه درون خوشه‌ای است، توانسته خوشه‌های منسجم‌تری را تولید نماید.

در گام نخست ایده پیشنهادی، از الگوریتم مبتنی بر چگالی به منظور تعیین تعداد خوشه و مقداردهی اولیه برای ایجاد کشورها استفاده شده است. حسن استفاده از خوشه‌بندی مبتنی بر چگالی آن است که تعیین تعداد خوشه‌ها به‌عنوان پارامتر ورودی مساله الزامی نیست و در روند خوشه‌بندی، وابسته به نواحی با تراکم بالاتر، خوشه‌ها تشکیل می‌شوند. به این صورت که نواحی با میزان چگالی بالاتر، با در نظر گرفتن میزان فاصله و تعداد خاص همسایگان، به‌عنوان یک خوشه در نظر گرفته می‌شوند. سپس در گام دوم، کیفیت خوشه‌های تولید شده اولیه با استفاده از تابع برازندگی مورد نظر، بررسی و در نهایت جواب تولید شده از طریق معیار ماژولاریتی ارزیابی می‌شود. روند اجرای این فرآیند در شکل (۱) نشان داده شده است.

۳-۱- تولید جمعیت اولیه بر اساس خوشه‌بندی مبتنی بر چگالی

داده‌های آزمایشی در این مقاله، مربوط به هفت شبکه اجتماعی معروف هستند. داده‌های ورودی، به صورت گراف بوده و در قالب ماتریس ذخیره شده‌اند. جدول (۱)، مجموعه داده‌های بکار رفته در این تحقیق را معرفی می‌کند.

جدول ۱: مجموعه داده‌های مورد استفاده در این تحقیق

مجموعه داده	تعداد راس	تعداد ریل
کاراته	۳۴	۷۸
فوتبال	۱۱۵	۶۱۳
دولفین	۶۲	۱۵۹
کتاب‌های سیاسی ایالت متحده آمریکا	۱۰۵	۴۴۱
جاز	۱۹۸	۴۴۱
فیسبوک	۲۸۸۸	۲۹۸۱
توییتر	۲۳۳۷۰	۳۲۸۳۱

۳-۲- تعیین تعداد خوشه‌ها مبتنی بر رویکرد چگالی

در این مرحله، با در نظر گرفتن دو پارامتر تعداد همسایگی گره V_i (که در ابتدا عدد ۵ در نظر گرفته شده و تا نصف تعداد کل گره‌های ورودی افزایش می‌یابد) و فاصله بین گره V_i با همسایگانش (که از

جدول ۲: مقادیر حاصل برای پارامتر ماژولاریتی

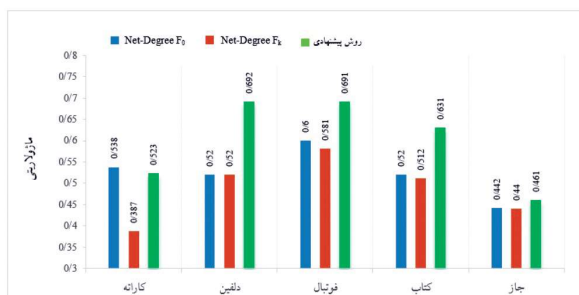
روش پیشنهادی	رقابت استعماری پایه	ژنتیک	کارانه
۵۲٪	۴۸٪	۶۰٪	کارانه
۶۹٪	۵۳٪	۴۴٪	دلفین
۶۹٪	۵۹٪	۶۸٪	فوتبال
۶۳٪	۵۰٪	۴۶٪	کتاب
۴۶٪	۴۱٪	۲۵٪	جاز
۸۴٪	۶۱٪	۶۰٪	توییت
۸۵٪	۷۱٪	۶۴٪	فیسبک

در شکل (۲) مقادیر ماژولاریتی حاصل از الگوریتم CC-GA، LPA و INFOMAP [۶] و مقدار ماژولاریتی روش پیشنهادی بر روی مجموعه داده‌های مختلف نشان داده شده است. مقدار حاصل برای شاخص ماژولاریتی در روش پیشنهادی در مجموع داده‌های کارانه، دلفین، فوتبال، کتاب و فیسبک به ترتیب برابر ۰/۵۲۳، ۰/۶۹۲، ۰/۶۹۱، ۰/۶۳۱، ۰/۸۴۶ بوده که از نتایج سایر الگوریتم‌های معرفی شده در روش رقیب بهتر می‌باشد. (در روش Infomap مقدار ماژولاریتی برای مجموعه داده جاز محاسبه نشده است.)



شکل ۲: مقایسه مقدار ماژولاریتی حاصل از روش پیشنهادی و الگوریتم infomap و LPA، CC-GA

در شکل (۳)، مقدار حاصل از شاخص ماژولاریتی توسط الگوریتم‌های Net-Degree F0 و f_k [21] و مقدار ماژولاریتی روش پیشنهادی نشان داده شده است. مقدار ماژولاریتی محاسبه شده در روش پیشنهادی برای مجموعه داده‌های دلفین، فوتبال، کتاب، جاز، و کارانه، به ترتیب برابر ۰/۶۹۲، ۰/۶۹۱، ۰/۶۳۱، ۰/۴۶۱، ۰/۵۲۳ بوده که برای تمام موارد به جز مجموعه داده کارانه، بهتر از روش‌های رقیب بوده است.



شکل ۳: مقایسه مقدار ماژولاریتی حاصل از روش پیشنهادی و الگوریتم F_k و Net-Degree F0

شکل (۴)، مقدار ماژولاریتی حاصل از الگوریتم Memetic [23] و حاصل از روش پیشنهادی را نمایش می‌دهد. لازم به ذکر است در این مقاله، مقدار ماژولاریتی فقط برای مجموعه داده‌های کارانه، دلفین و

شش فاصله اپسیلونی ۰/۳، ۰/۴، ۰/۵، ۰/۶، ۰/۷ و ۰/۸ استفاده شده است) تعداد خوشه‌ها مشخص می‌گردد. بر این اساس، کشورها که همان جمعیت اولیه بوده و هر یک نحوه‌ای خاص از خوشه‌بندی داده‌های ورودی شبکه‌های اجتماعی را نمایش می‌دهند، به وجود خواهند آمد.

۳-۳- مقدار دهی اولیه به استعمارگران

در این مرحله، به ازاء هر جمعیت اولیه (کشور) تابع هدف محاسبه می‌شود. سپس این کشورها به ترتیب نزولی از بیشترین مقدار تا کمترین مقدار هدف، مرتب شده و ۱۰ کشور اولیه که بیشترین میزان تابع هدف را دارند، به عنوان استعمارگر و مابقی کشورها به عنوان مستعمرات در نظر گرفته می‌شوند.

۳-۴- اجرای الگوریتم رقابت استعماری با تابع هدف پیشنهادی (تمرکز بر شباهت درون خوشه‌ای)

در این مرحله، با توجه به توضیحات بخش ۲-۱، الگوریتم رقابت استعماری برای ایجاد خوشه‌های متراکم مناسب اجرا می‌شود.

۳-۵- شرط توقف- پایان حلقه

پس از انجام مراحل آزمون و خطا در محدوده ۴۰ تا ۲۰۰ بار اجرای الگوریتم، تعداد برنامه جهت رسیدن به جواب بهینه، برابر با ۱۰۰ انتخاب گردید، چرا که در این حالت خوشه‌بندی بهینه توسط الگوریتم ارایه شده، استخراج گردید.

۳-۶- ارزیابی خوشه‌بندی نهایی

در نهایت، نتایج حاصل در روش پیشنهادی با استفاده از معیار ماژولاریتی ارزیابی خواهند شد.

۴- یافته‌ها و نتایج

هدف اصلی این تحقیق، ارایه روشی جهت تولید جوامع منسجم از داده‌های شبکه‌های اجتماعی بوده است. بنابراین در این بخش، به بررسی نتایج حاصل از الگوریتم پیشنهادی و مقایسه آن با الگوریتم‌های CC-GA، LPA و [6] InfoMap، NetDegree F0 و f_k [21] و همچنین Memetic [23] پرداخته شده است.

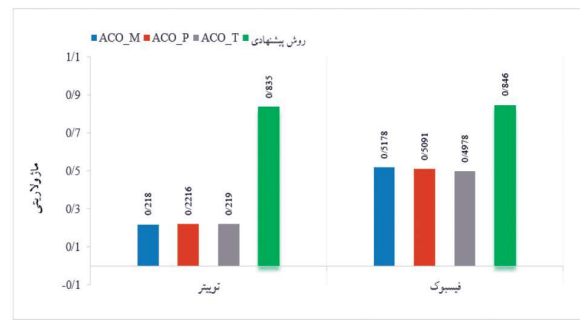
در نخستین آزمایش، به منظور بررسی تاثیر استفاده از الگوریتم رقابت استعماری در تعیین انتخاب جمعیت اولیه بر اساس مقدار خوشه‌بندی مبتنی بر چگالی، در مقایسه با روش‌های خوشه‌بندی کاملاً تصادفی همانند الگوریتم ژنتیک و همچنین رقابت استعماری پایه، عمل خوشه‌بندی برای مجموعه داده‌های موجود در هر سه روش انجام گرفت. جدول (۲) نتایج حاصل برای پارامتر محاسبه شده ماژولاریتی در این آزمایش را نشان می‌دهد. مقادیر بزرگتر به دست آمده برای این پارامتر در روش پیشنهادی در بیشتر موارد، بیانگر کاراتر بودن این روش در مقایسه با الگوریتم‌های فراابتکاری رقیب دیگر است.

پس از آزمایش نخست و مشخص شدن برتری روش پیشنهادی در مقایسه با سایر روش‌های مرسوم استفاده از الگوریتم‌های فراابتکاری، آزمایش‌های دیگری بر روی مجموعه داده‌ها موجود با شرایط مشابه با سایر روش‌های رقیب صورت گرفت. در شکل (۲) تا (۵)، مقادیر حاصل از شاخص ماژولاریتی مربوط به هفت مجموعه داده باشگاه کاراته، زاکاری، دلفین، مدرسه فوتبال امریکایی و کتاب‌های سیاسی ایالت متحده امریکا، فیسبک، توییت و جاز نمایش داده شده است. شاخص ماژولاریتی بیانگر انسجام درونی خوشه‌بندی است و مقدار بیشتر این پارامتر، بیانگر خوشه‌بندی مناسب‌تر روش مربوطه خواهد بود.

فوتبال محاسبه شده است. همان‌طور که در شکل دیده می‌شود، روش پیشنهادی برای مجموعه داده‌های دلفین و فوتبال توانسته میزان انسجام خوشه‌ها را افزایش دهد. لیکن از آنجا که الگوریتم Memetic روشی مبتنی بر رفتارشناسی است، توانسته در مجموع داده کاراته، که داده‌های آن افراد هستند، خوشه‌های منسجم‌تری تولید نماید.



شکل ۴: مقایسه ماژولاریتی حاصل از روش پیشنهادی و الگوریتم Memetic جهت مقایسه دو مجموعه داده فیسبوک و توییتر، که ابعاد بزرگتری نسبت به مجموعه داده‌های قبلی دارند، مقایسه دیگری با تحقیق سانچز و دوارت [۱۹] صورت گرفت. در آن مقاله، مقدار ماژولاریتی حاصل از الگوریتم‌های ACO_T, ACO_P, ACO_M, که اساس کار آن‌ها الگوریتم تکاملی کلونی مورچگان است، برای مجموعه داده فیسبوک و توییتر محاسبه شده است. در شکل (۵) مقادیر حاصل از الگوریتم‌های بالا در مقایسه با روش پیشنهادی مقایسه شده است. در این مقایسه مقدار ماژولاریتی برای روش پیشنهادی در مجموعه داده توییتر برابر با ۰/۸۳۵ و در مجموعه داده فیسبوک برابر با ۰/۸۴۶ بدست آمد که نشان می‌دهد این روش، برای مجموعه داده شبکه‌های اجتماعی با ابعاد بزرگ می‌تواند نتایج تشخیص جوامع را بسیار بهبود بخشد.



شکل ۵: مقایسه مقدار ماژولاریتی حاصل از روش پیشنهادی و الگوریتم‌های ACO_T, ACO_P, ACO_M

۵- مرور و نتیجه‌گیری

در تحلیل شبکه‌های اجتماعی، هر فرد را به صورت رأس یک گراف و روابط اجتماعی وی را به صورت یال نشان می‌دهند. انواع مختلفی از یال‌ها می‌توانند میان رئوس وجود داشته باشند. می‌توان از مفهوم شبکه‌های اجتماعی در بسیاری از سطوح فردی و اجتماعی به منظور شناسایی مسایل و تعیین راه حل آن‌ها، برقراری روابط اجتماعی، اداره امور تشکیلاتی، سیاست‌گذاری و رهنمون‌سازی افراد در مسیر دستیابی به اهداف استفاده نمود. شاید در ابتدا هدف ایجاد شبکه‌های اجتماعی، برقراری ارتباط بین افراد و به اشتراک‌گذاری داده‌های تصویری، متنی و ویدئویی بود. اما در گذر زمان، داده‌های موجود

در آن‌ها، جهت انجام تحقیقات متنوع دیگری نیز مورد استفاده قرار گرفت.

چالش خوشه‌بندی شامل یافتن روش‌های مناسب جهت خوشه‌بندی و همچنین سرعت در اجرای دسته‌بندی بهینه می‌باشد. با در نظر داشتن موارد گفته شده، این پژوهش با استفاده از الگوریتم رقابت استعماری سعی بر بهبود خوشه‌بندی و به دست آوردن بهترین دسته‌ها در یک شبکه اجتماعی را داشته است. استفاده از الگوریتم رقابت استعماری با استفاده از خوشه‌بندی مبتنی بر چگالی جهت تعیین جواب‌های اولیه، روند خوشه‌بندی داده‌های شبکه‌های اجتماعی را به طور میانگین روی مجموعه داده‌های ورودی (به خصوص داده‌های حجیم تر مانند فیسبوک و توییتر) بهبود بخشیده است. این ادعا با محاسبه پارامتر معروف ماژولاریتی بر روی هفت مجموعه داده معروف در حوزه شبکه‌های اجتماعی، به‌عنوان ورودی الگوریتم پیشنهادی و نتایج حاصل از الگوریتم‌های LPA, CC-GA, Infomap, Net-degree تک هدفه و دو هدفه و الگوریتم Memetic اثبات شده است. در برخی مجموعه داده‌ها، مانند مجموعه داده کاراته، الگوریتم Memetic بهتر عمل کرده و آن هم به دلیل رویکرد دو هدفه بودن و رفتارشناسی این الگوریتم است. با بررسی روش پیشنهادی بر روی داده‌های شبکه اجتماعی توییتر، فیسبوک و جاز، در برخی از موارد نتایج بهتر و یا نزدیک به روش‌های دیگر است و می‌توان ادعا نمود با دیگر موارد، به خوبی رقابت کرده و نتایج مناسبی را تولید نموده است.

به‌طور کلی روش پیشنهادی توانسته در مجموع داده فیسبوک، بر اساس شاخص ماژولاریتی خوشه‌بندی را تا دقت ۰/۸۴۶ ارتقا و خوشه‌های منسجم‌تری را تولید نماید. همچنین روش پیشنهادی در مقایسه با الگوریتم رقابت استعماری پایه در هفت مجموعه داده کاراته، دلفین، فوتبال، کتاب، جاز، توییتر و فیسبوک توانسته به طور میانگین ۲۱/۴۵٪ مقدار ماژولاریتی را افزایش داده و جوامع منسجم‌تری را استخراج نموده است. با توجه به نتایج حاصل می‌توان نتیجه گرفت که روش پیشنهادی در جوامع بسیار بزرگ مانند فیسبوک و توییتر بهتر عمل نموده و توانسته خوشه‌بندی بهتری ایجاد کند.

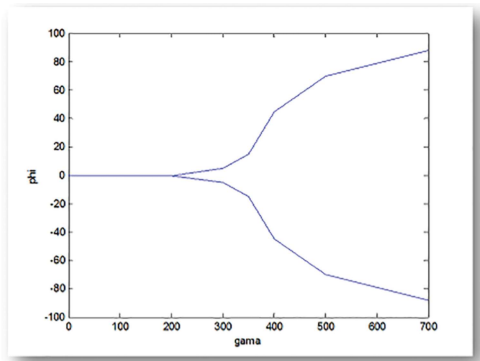
پی‌نوشت‌ها

- 1 Social Networks
- 2 Evolutionary Algorithms
- 3 Optimization
- 4 Clustering
- 5 Density-based Clustering
- 6 Goal Function
- 7 Cost Function
- 8 Fitness Function
- 9 Heuristic Algorithms
- 10 Meta-heuristic Algorithms
- 11 Local Optima
- 12 Imperial Competitive Algorithm
- 13 Modularity
- 14 Particle Swarm Optimization

مراجع

- [۱] مصطفی حسنونند، حسن نادری، فردین ابدالی «رایه مدلی برای خوشه‌بندی برای خوشه‌بندی شبکه‌های اجتماعی یویا»، همایش ملی مهندسی رایانه و مدیریت فناوری اطلاعات، تهران، شرکت علم و صنعت طلوع فرزین، ۱۳۹۳.
- [2] M.H Jajeer and D.Dipankar, "Distributed genetic algorithm to big data clustering", IEEE Symposium series on Computational Intelligence(SSCI), 2016.
- [3] A.K. Jain, M. Narasimha Murty, and P.J Flaynn, "Data clustering: a review", ACM computing surveys(CSUR), 264-323, 1999.
- [۴] آتشیز گرگری، اسماعیل، توسعه الگوریتم بهینه‌سازی اجتماعی و بررسی کارایی آن، پایان

سوپر کریتیکال (فوق بحرانی) می‌باشد. که در آزمایشگاه برای معرفی انشعاب مفید است. در ادامه به‌عنوان پیشنهاد می‌توان برای سیستم ساخته شده در آزمایشگاه، کنترل کننده طراحی کرد یا شناسایی سیستم انجام داد. همچنین برای بررسی بیشتر انشعاب‌ها می‌توان سیستم مهره روی سیم مورب را به‌عنوان موضوع مقالات آتی در نظر گرفت.



شکل ۱۸: دیاگرام انشعاب سیستم حلقه و مهره به‌دست آمده از طریق آزمایشگاه

پی‌نوشت‌ها

- 1 Chaos Theory
- 2 Bifurcation
- 3 saddle-node bifurcation
- 4 Transcritical Bifurcation
- 5 Pitchfork Bifurcation
- 6 supercritical
- 7 subcritical
- 8 HOPF bifurcation

مراجع

- [1] H. Strogatz, S. NONLINEAR DYNAMICS AND CHAOS. New York. 1994.
- [۲] اهاشمی گلپایگانی، سیدمحمد رضا. آشوب و کاربردهای آن در مهندسی. انتشارات دانشگاه صنعتی امیرکبیر. ۱۳۹۸.
- [۳] کلانتری، محمد. سهرابی، سکینه. رشیدیکنعان، حمیدرضا. کرمی، حسین. الگوریتم بهینه‌سازی جستجوی ابرکروی مبتنی بر نظریه آشوب. نشریه مهندسی برق و مهندسی کامپیوتر ایران، سال ۱۶، شماره ۲. ۱۳۹۷. صص ۱۲۱-۱۳۰.
- [۴] به‌نیاسهراب. ضیائی، جاوید. خداوردی زاده، مهدی. سرکوب سیگنال‌های آشوبناک قلبی با استفاده از روش کنترل پویا. لیزر در پزشکی، شماره ۱. ۱۳۹۶. صص ۲۶-۳۱.
- [5] B. Patwari M, R. Manza R, M. Rajput Y, Saswade M, Deshpande N. Personal Identification algorithm based on Retinal Blood Vessels Bifurcation. International Conference on Intelligent Computing Applications. 2014.
- [6] Ludwig, D., Jones, D. D., and Holling, C. S. Qualitative analysis of insect outbreak systems: the spruce budworm and forest. J. Anim. Ecol. 47, 3 15.1978.
- [7] Muratori S, Rinaldi S. Catastrophic bifurcations in a second-order dynamical system with application to acid rain and forest collapse. (13)12, 674-681.1989.
- [8] Liu, F., Mei, Q., Sun, F., Wang, X., O Wang, H. Stability and Neimark-Sacker bifurcation analysis for single gene discrete system with delay. Proceedings of the 37th Chinese Control Conference. 2018.
- [9] Lv, S., Ma, Y., Zhou, X., Gao, Z. Voltage Stability Analysis of Power System Based on Bifurcation Theory. International Conference on Mechatronics and Automation August 5 - 8, Changchun, China. 2018.
- [۱۰] مجیدی، نیما، ولایتی، محمدحسین. بررسی تأثیر مدل‌های بار، پارامترهای سیستم تحریک و محدودیت توان راکتیو تولیدی ژنراتورهای سیستم قدرت در حد دینامیکی پایداری ولتاژ با استفاده از یک آنالیز ترکیبی. مجله مدل سازی در مهندسی. (۲۰)، (بهار ۱۳۸۹). صفحات ۳۹-۵۵.
- [11] Sun, k., Wang, X., Yin, L., Zhu, C. Chaos and bifurcations of the fractional-order unified system. International Workshop on Chaos-Fractal Theory and its Applications. 2010.

- نامه کارشناسی ارشد دانشکده مهندسی برق و کامپیوتر دانشگاه تهران، ۸۸۳۱.
- [5] M.E. Newman, 'modularity and community structure in networks'. Proc. Of the national academy of sciences, 103,23, 8577-8582, 2006.
- [6] A.Said, R.A. Abbasi, O.maqbool, A. Daud and N.R Aljohani, "CC-GA: A clustering coefficient based genetic algorithm for detecting communities in social networks", Applied Soft Computing, 63, pp.59-70,2018.
- [7] A. Biswas and B.Biswas, "Defining quality metrics for graph clustering evaluation", Expert Systems with Applications, Vol.71, pp.1-17,2017.
- [8] A.A Bara'a and H.S Khoder, "A new multi objective evolutionary framework for community mining in dynamic social networks", Swarm and Evolutionary Computation, Vol.31, pp. 90-109,2016.
- [9] M. Girvan and M.E. Newman, "Community structure in social and biological network", Proc. Of the national academy of sciences, 7821-7826,2002.
- [10] D. Greene, D. Doyle and P. Cunningham, "Tracking the evolution of communities in dynamic social networks", Int. Conf. on advances in social networks analysis and mining. IEEE, 2010.
- [11] Z. Shi, W. Chen, L. Yue, J. Han and L. Feng, "User relation predication based on matrix factorization and hybrid particle swarm optimization", Proc. Of the 26th Int. conf. on Wold Wider Web Companion. 2017.
- [12] E. Zamudio, L.S. Berdun and A.A. Amandi, "Social networks and genetic algorithms to choose committees with independent members", Expert Systems with Applications, Vol. 43, pp. 261-270,2016.
- [13] G.Rossetti, L.Papparardo, D.Pedreschi and F. Ginnotti, "Tiles: on online algorithm for community discovery in dynamic social networks", Machine Learning, Vol. 106.8, pp.1213/1241,2017.
- [14] M.E.J. Newman, "Community detection in networks: Modularity optimization and maximum likelihood are equivalent", Phys. Rev. E 94, 052315, Nov 2016.
- [15] S.S Hoseini and S.H. Abbasi, "A new method for community detection in social networks based on message distribution", Int J.I of Computer Science and Network Security, Version 17, pp. 298-308,2017.
- [16] C.Xu, H. Zhang, B Lu and S. Wu, "Local Community Detection Using Social Relations And Topic Features in Social Networks", Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Springer, Cham, pp. 371-383,2017.
- [17] F.Ullah and S. Lee, "Community clustering based on trust modeling weighted by user interests in online social networks", Chaos, Solitons & Fractals, Version 103, pp. 194-204,2017.
- [18] K.M. cogan r and A.R.M Jcob, "A fast DBSCAN clustering algorithm by accelerating neighbor searching using Groups method", Pattern Recognition, Vol. 58, pp. 39-48,2016.
- [19] J.Sanchez-Oro and A. Duarte, "Iterated Greedy algorithm for performing community detection in social networks", Future Generation Computer Systems, 2018.
- [20] A.Azad, G.A. Pavlopoulos, C.A. Ouzounis, N.C. Kyripides and A. Buluc, "HipMCL: a high-performace parallel implementation of the Markov clustering algorithm for large-scale networks", Nucleic acid research, 46.6, e33-e33, 2018.
- [21] K.R. Zalik, "Evolution Algorithm for Community Detection in Social Networks Using Node Centrality", In Intelligent Methods and Big Data in Industrial Applications, Springer, Cham, pp. 73-87,2019.
- [۲۲] فاطمه ترشیزی نژاد، مهرداد جلالی، داوود بهره پور «روش‌های جهت تشخیص جوامع در شبکه‌های اجتماعی مبتنی بر الگوریتم رقابت استعماری»، دومین کنفرانس بین‌المللی فن‌آوری، ارتباطات و دانش، مشهد، ۵۱۰۲ KCTCI، ۴۹۳۱.
- [23] K.R. Zalik and B. Zalik, "Multi-objective evolutionary algorithm using problem-specific generic operations for community detection in networks", Neural Computing and Applications, pp.1-14,2017.
