Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
Volume 12, Number 3, Autumn 2023, pp. 58-73
Received: 06/13/2023, Revised: 10/08/23, Accepted: 01/01/2024

Discovering Effective Communities in Social Networks Using Fuzzy Semantic Relations Elicitation

Yasser Sadri[1], Saeid Taghavi Afshord[2*], Shahriar Lotfi[3], Vahid Majidnezhad[4]

1- Department of Computer Engineering, Shabestar Branch, Islamic Azad University, Shabestar, Iran.
2*- Department of Computer Engineering, Shabestar Branch, Islamic Azad University, Shabestar, Iran.
3- Department of Computer Science, Faculty of Mathematics, Statistics and Computer Science, University of Tabriz, Tabriz, Iran.
4- Department of Computer Engineering, Shabestar Branch, Islamic Azad University, Shabestar, Iran.

[1] yasser.sadri@iaushab.ac.ir, [2*] sa.taghavi@iau.ac.ir, [3]shahriar_lotfi@tabrizu.ac.ir and [4]vahidmn@iaushab.ac.ir

Corresponding author's address: Saeid Taghavi Afshord, Department of Computer Engineering, Shabestar Branch, Islamic Azad University, East Azerbaijan, Iran.

Abstract- Discovering Communities is a fundamental problem in understanding network performance for social network analysis. Traditional community detection methods merely consider the network topologies. Nonetheless, social media contains valuable data about people's interests, concerns, and sentiments, which is not reflected in the structure. The semantic solutions eventuate to the mislay of precious structural information. Most existing combinatorial methods favor one of the mentioned kinds and have limited performance. This paper introduces a 2-phase way based on fuzzy inferences to determine effective social network communities. A series of real-life and synthetic networks have been used to evaluate the proposed method compared with several relevant algorithms. The experimental results proved that the proposed approach performs better in detecting meaningful communities and is more effective concerning network coherence and node attributes.

Keywords- Social Networks Analysis, Community Detection, Semantic, Fuzzy Inference.

مجله علمی پژوهشی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
جلد ۱۲، شماره ۳، پائیز ۱۴۰۲ ، صفحه ۵۸-۷۳
دریافت: ۱۴۰۲/۰۳/۲۳، بازنگری: ۱۴۰۲/۰۷/۱۶، پذیرش: ۱۴۰۲/۱۰/۱۱

شناسایی جوامع کارا در شبکه‌های اجتماعی با استخراج فازی پیوندهای معنایی

یاسر صدری[۱]، سعید تقوی افشرد *[۲]، شهریار لطفی [۳]، وحید مجید نژاد [۴]
۱- گروه مهندسی کامپیوتر، واحد شبستر، دانشگاه آزاد اسلامی، شبستر، ایران.
*۲- گروه مهندسی کامپیوتر، واحد شبستر، دانشگاه آزاد اسلامی، شبستر، ایران.
۳- گروه علوم کامپیوتر، دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه تبریز، تبریز، ایران.
۴- گروه مهندسی کامپیوتر، واحد شبستر، دانشگاه آزاد اسلامی، شبستر، ایران.
[۱] yasser.sadri@iaushab.ac.ir, [۲*] sa.taghavi@iau.ac.ir, [۳]shahriar_lotfi@tabrizu.ac.ir and [۴]vahidmn@iaushab.ac.ir

* نشانی نویسنده مسئول: سعید تقوی افشرد، شبستر، دانشگاه آزاد اسلامی واحد شبستر، دانشکده فنی و مهندسی، گروه کامپیوتر.

چکیده- استخراج جوامع، از موضوعات چالش‌برانگیز و مهم در حوزه تحلیل شبکه‌های اجتماعی است. مکانیسم تشخیص جوامعی که به طور همزمان نیازهای توپولوژیکی و معنایی را برآورده سازد، اهمیت بیشتری پیدا کرده است. اکثر مطالعات موجود بر روی ساختار شبکه، بدون توجه به ویژگی‌های معنایی، تمرکز دارند. حال اینکه، معنا و محتوای اجتماعی حاوی داده‌های ارزشمندی در مورد علایق، نگرانی‌ها و احساسات افراد است. راه‌حل‌های معنایی محض، منجر به از دست رفتن اطلاعات توپولوژیکی ارزشمند شبکه می‌شوند. روش‌های ادغامی موجود، ساختار و معنا را به صورت بنیادی و اساسی ترکیب نمی‌کنند. اغلب این الگوریتم‌ها به علت ماهیت طراحی، به یکی از روش‌هاي مذکور متمایل بوده و عملکرد محدودی دارند. در این مقاله یک روش مبتنی بر منطق فازی جهت شناسایی جوامع ارائه می‌شود. ابتدا معنا و ساختار در یک شبکه مشترک ترکیب و سپس جوامع استخراج می‌گردند. مجموعه‌ای از آزمایش‌های شبیه‌سازی جهت ارزیابی عملکرد روش پیشنهادی در مقایسه با سایر الگوریتم‌های موجود در سه زمینه ساختاری، معنایی و ادغامی با استفاده از مجموعه داده‌های مصنوعی و طبیعی در اندازه‌های مختلف گره‌ها، یال‌ها، ویژگی‌ها انجام شده‌است. نتایج حاصل نشان می‌دهند، روش پیشنهادی در رابطه با انسجام شبکه و ویژگی‌های گره‌ها به طور موثری عمل می‌کند.

واژه‌های کلیدی: شبکه‌های پیچیده پویا؛ تحلیل شبکه‌های اجتماعی؛ شناسایی جوامع؛ معنا؛ استنتاج فازی.

۱- مقدمه

در سال‌های اخیر، توسعه روزافزون رسانه‌های اجتماعی تأثیر بسزایی بر جنبه‌های مختلف زندگی انسان گذاشته است [۱]. با توجه به تعداد زیاد افراد و اشیاء اجتماعی در کنار حجم عظیم داده در این شبکه‌ها، تحلیل شبکه‌های اجتماعی به عنوان زیرمجموعه مناسبی از سیستم‌های پیچیده‌ی پویا، به موضوعی مهم و چالش‌برانگیز در زمینه‌ی داده‌کاوی تبدیل شده است [۴-۲]. یکی از چالش‌های مهم در این زمینه، شناسایی جوامع است [۵]. تشخیص جوامع می‌تواند به شناسایی گروه‌های کاربری با علایق، پیش‌زمینه‌ها و اهداف مشابه کمک نماید [۶]. از سوی دیگر با تجزیه شبکه به اجتماعات کوچکتر، تحلیل شبکه ساده‌تر می‌گردد. به‌علاوه، شناسایی جوامع می‌تواند در کاربردهای مهمی همچون گسترش اطلاعات، به اشتراک‌گذاری دانش، سیستم‌های توصیه‌گر، تحلیل تمایلات و احساسات و نمایه-سازی رفتاری و روان‌شناختی کاربران و غیره مورد استفاده قرارگیرد [۷, ۸]. در سال‌های اخیر، محتوای به اشتراک گذاشته شده در پلتفرم‌های اجتماعی تحت تأثیر همه‌گیری کووید-۱۹[۱] قرار گرفته است. بسیاری از مردم در مورد نگرانی‌های خود در مورد سرایت، انزوا، فاصله، احتیاط از واکسیناسیون و ترس از بستری شدن صحبت می‌کنند [۹]. کشف جوامع می‌تواند برای تحلیل رفتار افراد استفاده شود و به مطالعه‌ی مناطقی که کووید-۱۹ در آن‌ها با الگوی مشابهی تأثیر گذاشته‌است، کمک می‌کند [۱۰]. علاوه بر آن، تشخیص

جوامع می‌تواند در حوزه‌های اقتصادی و بازاریابی کمک شایانی نماید. کشف دقیق جوامع و تجزیه و تحلیل هر یک از آنها می‌توانند عملکرد طرح‌های تبلیغاتی و صنعت بازاریابی را با تعیین و دسته‌بندی گروه‌های مناسب کاربران در یک شبکه‌ی خاص بهبود بخشند [۱۱]. در این راستا، یک راه موثر برای ارائه توصیه‌هایی سازگار با رفتار و منافع کاربران، استفاده از جنبه‌های ساختار رفتار کاربران و شناسایی جوامع کاربران است. در نهایت اینکه، تصمیمات و نگرش مردم و هم‌چنین روابط اجتماعی آن‌ها می‌تواند تحت تاثیر موقعیت عمومی در شبکه‌های اجتماعی قرار گیرد و از آنجا که افکار عمومی برای جنبه‌های مختلف زندگی انسان ضروری است، شناخت آنها بسیار مهم است [۴, ۱۲].

اجتماع در تعریف سنتی، یک زیرساختار مهمی از شبکه‌های اجتماعی است که در آن گره‌ها دارای ارتباطات قوی داخلی در مقایسه با ارتباطات خارجی می‌باشند (شکل۱-۱-ب). روش‌های تشخیص اجتماع سنتی، صرفا بر روی توپولوژی شبکه تمرکز دارند. بنابراین این روش‌ها عملکرد محدودی دارند، چرا که در شبکه‌های اجتماعی یک رابطه پیچیده مابین توپولوژی و معنا (یکی دیگر از منابع مهم اطلاعاتی) وجود دارد [۱۳]. به عنوان مثال، کاربران ممکن است جوامع مختلفی را به علت منافع و یا دلایل مختلف که بیشتر احتمال دارد در محتوای ارسال شده توسط کاربران منعکس شود (نه ساختار پیوند)، انتخاب کنند [۱۳]. همچنان‌که در شکل ۱-ب مشاهده می‌گردد؛ الگوریتم‌های تشخیص جوامع ساختاری ممکن است چند گره را دریک اجتماع قرار دهند در حالی که گره‌ها به سه موضوع و خصیصه متفاوت تعلق دارند. علاوه بر این، دو کاربر در شبکه اجتماعی ممکن است با وجود عدم ارتباط مستقیم، بدلیل منافع مشابهی که می‌تواند در محتویات کد شود، عضو یک اجتماع باشند [۱۴].

در مقابل، الگوریتم‌هایی وجود دارند که شبکه را بدون توجه به ساختار توپولوژیکی تنها بر مبنای معنا (خصوصیات) افراز می‌کنند. همچنانکه در شکل ۱-ج مشاهده می‌گردد، در واقع جوامع تولید شده منعکس کننده یک اجتماع در معنای سنتی آن نیستند؛ چرا که افراد جامعه که به لحاظ منطقی مشابه هستند ممکن است هیچ-گونه ارتباط صریح با یکدیگر نداشته باشند. چون اغلب افراد، به ویژه در محیط‌های توزیع‌شده، یکدیگر را نمی‌شناسند و هرگز ارتباط برقرار نمی‌کنند [۱۵]. در نتیجه جوامع تولید شده توسط این روش-ها ممکن است شامل کاربران با اتصالات ضعیف باشد [۱۶].

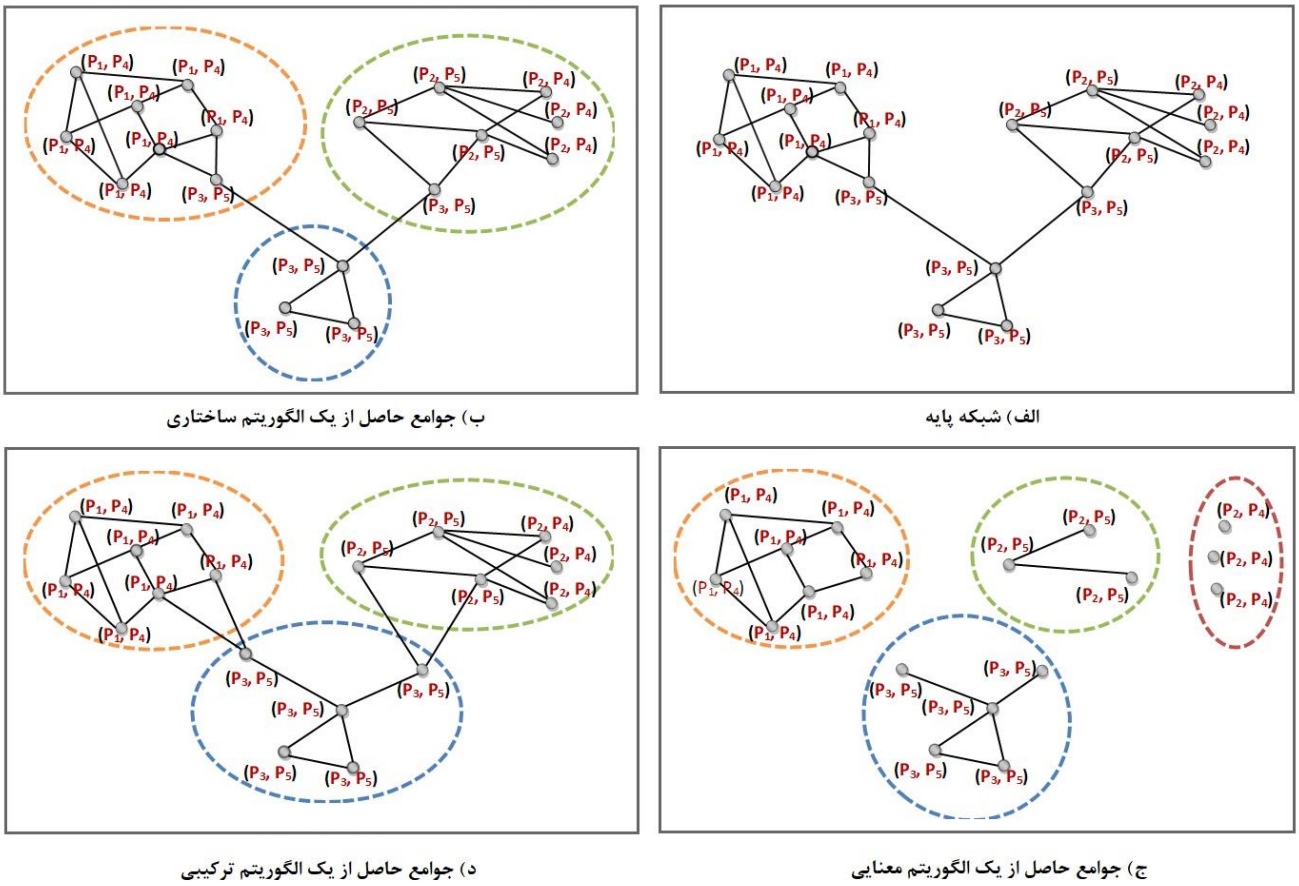
در دنیای واقعی، تعریف اجتماع به مجموعه‌ای از افراد اطلاق می‌شود که با یکدیگر ارتباطات زیادی دارند و دارای خصوصیات معنایی مشابه بوده و به موضوعات مشترکی نیز علاقه‌مند هستند. از این رو، یافتن جوامع معنادار در چنین شبکه‌هایی به موضوعی مهم و چالش‌برانگیز تبدیل شده است. چالش این است که چگونه جوامع شبکه‌های اجتماعی را به گونه‌ای استخراج نماییم که ساختار و محتوا را ترکیب نماید. عیب روش‌های موجود این است که قبل از شناسایی جوامع، راه‌حل اساسی برای این ترکیب ارائه نمی‌دهند. شکل ۱-د یک تشخیص اجتماع ایده‌آل از لحاظ ساختاری و معنایی را نشان می‌دهد که گره‌هایی که از لحاظ ساختاری ارتباط قوی دارند و از لحاظ معنایی دارای خصوصیات مشترک می‌باشند، در یک اجتماع قرار گرفته‌اند. برای دستیابی به این هدف، در این مقاله یک روش تشخیص جوامع با استخراج پیوندهای معنایی فازی در شبکه‌های اجتماعی معرفی می‌شود. روش پیشنهادی روابط مستحکم مابین گره‌ها را بر اساس علایق، احساسات و به طور کلی ویژگی‌های موجودیت‌ها درکنار رفتارها و تعاملات بین آنها جهت ترکیب اساسی ساختار و محتوا کشف می‌کند. در فاز دوم جوامع از شبکه بدست آمده از فاز نخست استخراج می‌گردد. نوآوری‌های این مقاله به شرح زیر خلاصه می‌شوند:

۱. یک چارچوب دو مرحله‌ای جامع برای حل مشکل تشخیص جوامع در شبکه‌های اجتماعی پیشنهاد می‌گردد. در فاز نخست، سیستمی بر اساس منطق فازی طراحی شده است تا توپولوژی شبکه و معنا را در یک پروتکل ساختاری مشترک ترکیب کند. فاز دوم، جوامع را از شبکه غنی شده استخراج می‌نماید.

۲. برای ادغام دو عنصر ناهمگون ساختار و محتوا، تبدیل معنایی به شکل ساختاری ضروری است. در این راستا، یک روش استنتاج فازی کلی برای استخراج پیوندهای معنایی معرفی می‌کنیم. مجموعه متغیرهای کلامی، عضویت‌ها و مراحل استنتاج شامل فازی‌سازی، قواعد استنتاج و غیرفازی‌سازی فرموله می‌شوند.

۳. فاز دوم بر ماجولاریتی جوامع شبکه‌ی حاصل از فاز نخست تمرکز دارد. در این شبکه، یال‌ها علاوه بر ساختار، نشان دهنده‌ی روابط معنایی نیز هستند. بنابراین، یک الگوریتم مناسب انتخاب می‌شود که ضمن افزایش ماجولاریتی، گره‌هایی با ویژگی‌های معنایی مشابه را نیز در یک جامعه قرار می‌دهد.

۴. آزمایش‌های گسترده در شبکه‌های اجتماعی با ویژگی‌های متفاوت از تعامل اجتماعی انسانی، وبلاگ‌نویسی اجتماعی، خرید مشترک، رتبه‌بندی مشترک، شبکه واژگانی و زمینه‌های اشتراک‌گذاری عکس انجام می‌شوند. نتایج حاصل نشان می‌دهند که روش پیشنهادی می‌تواند ساختارهای اجتماعی مطلوبی را کشف و اجتماعات مناسب را با در نظر گرفتن ساختار و معنایی در شبکه‌های اجتماعی به دست آورد. نتایج به دست‌آمده مزیت روش



شکل ۱) نمایی از خروجی انواع روش‌های استخراج جوامع. الف) یک نمونه شبکه اجتماعی فرضی با ویژگی‌های گره‌ها. ب) و ج) جوامعی هستند که به ترتیب با استفاده از الگوریتم‌های ساختاری و معنایی محض استخراج می‌گردند. د) نتایج یک الگوریتم تشخیص جوامع ترکیبی.

ما را از نظر معیارهای ارزیابی در مقایسه با سایر روش‌های تشخیص جوامع شاخص در انواع توپولوژیکی، معنایی و ترکیبی نشان می‌دهد.

ادامه مقاله به شرح زیر سازماندهی می‌شود. بخش دوم یک نمای کلی از کارهای پیشین در الگوریتم‌های تشخیص جوامع ارائه می‌دهد. در بخش سوم، مراحل طراحی و اجرای روش پیشنهادی شرح داده می‌شود. نتایج ارزیابی و شبیه‌سازی در بخش چهارم ارائه می‌شود، که عملکرد روش ما و دیگر الگوریتم‌های تشخیص جوامع مقایسه می‌گردند. نتایج در بخش پنجم مورد بحث قرار می‌گیرند. در نهایت، اظهارات پایانی و کارهای آینده در بخش ششم آورده شده است.

۲- کارهای پیشین

در سال‌های اخیر، کارهای مختلفی در زمینه شناسایی جوامع انجام گرفته و الگوریتم‌ها و ابزارهای متعددی ارائه شده‌است، اما تلاش برای ارائه روش‌های بهتر، کارآمدتر و واقع‌بینانه‌تر هم‌چنان ادامه دارد [۱۷]. اصولا روش‌های ارائه شده به سه دسته کلی ساختاری، معناگرایانه و ادغامی یا ترکیبی تقسیم می‌گردند. هدف از الگوریتم‌های ساختارمحور، فرموله کردن و مشخص نمودن زیرگروه‌های اجتماعی اغلب بر مبنای دو ویژگی ارتباط قوی داخلی و ارتباط ضعیف خارجی است. در یک طبقه‌بندی، روش‌های ساختاری را می‌توان به روش‌های مبتنی بر بخش‌بندی گراف و روش‌های مبتنی بر خوشه‌بندی تقسیم کرد. مشکل عمده این الگوریتم این است که تعداد گروه‌ها باید از پیش تعریف شوند [۱۸]. برخی از این روش‌ها در [۱۹-۲۱] پیشنهاد شده‌اند. خوشه‌بندی، فرآیند گروه‌بندی اقلام مشابه در خوشه‌های یکسان است. الگوریتم‌های ارائه شده در [۲۲-۲۶] برخی از روش‌های شناخته شده هستند. نویسندگان در [۲۷] روش لوین را بسط داده [۲۳] و یک الگوریتم تشخیص جوامع پویای مبتنی بر دانش کسب شده از مراحل قبلی تکامل شبکه ارائه نموده‌اند. الگوریتم تشخیص جوامع چندگانه‌ی آگاه از موتیف در [۲۸] به منظور به حداقل رساندن از دست دادن داده‌ها پیشنهاد شده است. نویسندگان در [۲۹] روش بازیگر-منتقد را بر اساس استراتژی یادگیری تقویتی عمیق با استفاده از بهینه‌سازی محلی تابع چگالی ماجولاریتی برای تشخیص جوامع در شبکه‌های اجتماعی پویا ارائه کرده‌اند. یک روش تشخیص جوامع مبتنی بر یادگیری عمیق موازی در [۳۰] پیشنهاد شده است.

ابتدا، یک روش تقسیم‌بندی برای تقسیم شبکه به چند بخش استفاده می‌شود. در مرحله بعد، این روش با پساانتشار و بهینه‌سازی ازدحام ذرات ادغام می‌شود. فرآیند بهینه‌سازی از طریق پساانتشار در یادگیری عمیق به صورت محلی در هر بخش از شبکه انجام می‌گردد. از بهینه‌سازی ازدحام ذرات برای بهینه‌سازی پساانتشار بخش‌های شبکه استفاده می‌شود. یک روش با رمزگذاری خودکار گذرای عمیق[۲] در [۳۱] برای توصیف شباهت بین رئوس در توپولوژی جهت شناسایی جوامع ساختاری مطالعه شده است.

اخیراً، مدل‌های مختلف معناگرایانه و مبتنی بر خصوصیات گره‌ها جهت استخراج جوامع مورد توجه قرار گرفته است. اساس بیشتر این روش‌ها بر این فرض استوار است که افراد غالباً از طریق نوشتن متن با سایر اعضای شبکه در ارتباط می‌باشند. بنابراین مجموعه‌ای از روش‌های کشف معنا در متن استفاده کرده و افرادی که به موضوعات مشابه علاقه‌مند می‌باشند را در یک اجتماع قرار می‌دهند [۳۲]. الگوریتم‌های کشف معنا، موضوعات را با استفاده از متغیرهای پنهان در اسناد مدل می‌کنند [۳۲]. از معروفترین روش‌های استخراج موضوعات که در بسیاری از رویکردهای شناسایی جوامع معناگرایانه مورد استفاده قرار گرفته‌است، دو الگوریتم تخصیص دریکله پنهان[۳] [۳۳] و تحلیل معناشناسی پنهان[۴] [۳۴] می‌باشند. الگوریتم دریکله پنهان یک روش تولیدی[۵] است که برای پارامترهای پنهان داده شده، به صورت تصادفی، یکسری داده قابل مشاهده تولید می‌کند. اگر این داده‌ها کلماتی باشند که از مستندات بیرون آمده باشند، آنها را در تعدادی موضوع گروه‌بندی می‌کند؛ که به منظور کشف و آشکارسازی ارتباطات مفهومی بین موجودیت‌ها در شبکه به کار می‌رود. خوشه‌بندی مبتنی بر مدل فیلد-ارتباط-موضوع[۶] توسط یو و همکاران [۳۵] برای غلبه بر محدودیت تعیین تعداد جوامع از قبل، ارائه شده است. در این مطالعه اتصالات معنایی وزن‌دار تعریف شده و از محاسبات لازم بر مبنای مدل مذکور برای ارزیابی وزن اتصالات معنایی هر نمونه فیلد استفاده می‌شود. سپس خوشه‌بندی بر اساس اتصالات معنایی وزن‌دار انجام می‌گیرد تا شبکه اجتماعی را در خوشه‌های جداگانه قرار دهد. در الگوریتمی که توسط ابدلباری و الکورانی [۳۶] ارائه شده است، کشف جوامع با استفاده از مدل دو لایه‌ای سازنده محدود شده ماشین بولتزمن انجام می‌شود. این مدل فرض می‌کند که اعضای یک جامعه بر اساس علایق و نگرانی‌های مشترک اعضاء به وجود می‌آیند. همچنین اجازه می‌دهد که اعضا به جوامع مختلف تعلق داشته باشند.

در مرجع [۳۷] مدل نویسنده-گیرنده-موضوع ارائه شده است که روند شناسایی جوامع را به دو مرحله نمونه‌گیری تخصیص دریکله پنهان و تشخیص جامعه، تقسیم‌بندی نموده و خوشه‌بندی جوامع را بکار می‌گیرد. این الگوریتم، هم‌چنین سعی در شناسایی جوامع همپوشان دارد. دای و همکاران [۳۸] تخصیص دریکله پنهان را با فاکتورسازی ماتریس گسترش می‌دهند، به طوری که یادگیری موضوع معنایی و تشخیص جامعه یکدیگر را در طول تخمین پارامتر تقویت می‌کنند. در [۳۹] نمونه‌گیری گیبس پیشنهاد شده توسط ژائو و همکاران [۴۰]، برای نگاشت گره‌های معنایی به فضای معنایی استفاده می‌شود. سپس، یک الگوریتم مبتنی بر بهینه‌سازی ازدحام ذرات[۷] برای تشکیل جوامع معنایی ارائه می‌شود. هاراکاوا و ایواهاشی [۴۱] شبکه را بر اساس شباهت‌های موضوعی بازسازی می‌کنند. سپس با استفاده از روش لوین جوامع را از این شبکه استخراج می‌کنند. روش پیشنهاد شده در [۴۲] جوامع را با بررسی زیرفضاهای خصوصیات شناسایی می‌کند. این روش تشخیص جوامع را به عنوان یک مسئله بهینه‌سازی در نظر می‌گیرد، که هدف آن یادگیری یک ماتریس شباهت جدید برای به حداکثر رساندن هموفیلی است.

هدف از روش‌های ادغامی، ارائه راه‌حلی است که معیارهای ساختاری و معنایی را در شناسایی جوامع همزمان پشتیبانی نمایند. در مطالعه ژائو و همکاران [۸]، یک رویکرد موضوع‌محور با ترکیب خوشه‌بندی اشیاء اجتماعی و تحلیل ارتباطات ارائه شده است. ابتدا یک الگوریتم اصلاح‌شده‌ی خوشه‌بندی k-میانگین، به نام «الگوریتم k-میانگین معیار آنتروپی» برای خوشه‌بندی اشیاء اجتماعی به موضوعات مختلف و یک الگوریتم ماجولاریتی برای خوشه‌بندی زیرفضای اشیاء اجتماعی موجود در موضوعات ارائه شده است. اغلب الگوریتم‌های ترکیبی فعلی توسعه این روش هستند. در مرجع [۴۳]، همانند الگوریتم ارائه شده در [۸]، ابتدا افراد با موضوع‌های مشترک را در یک خوشه قرار می‌دهد که عضویت یک گره در یک موضوع، عددی در بازه‌ی [۱ و۰] در نظر گرفته شده است. سپس با یک روش ساختاری ارتباطات داخل هر خوشه را مجددا بررسی و جوامع درون خوشه‌ای را شناسایی می‌نماید. روش شناسایی جوامع درون خوشه-ای روش لوین است. عارفیان و خیام‌باشی [۴۴] به جای الگوریتم لووین از خوشه‌بندی مارکوف استفاده نموده‌اند. یک تابع بهینه‌سازی برای ترکیب ساختار و خصوصیات گره‌ها در [۱۴] ارائه شده است. برای این منظور، دو نوع اجتماع ساختاری و معنایی به طور جداگانه تعریف و سپس، دو ماتریس مجزا با مقادیر صریح مشخص شده که نشان‌دهنده‌ی عضویت گره‌ها در هر اجتماع ساختاری یا معنایی است. مشکل اصلی، تابع بهینه‌سازی است که ترکیبی خطی از دو تابع ناهمگن ساختاری و معنایی است. نکته‌ی بعدی عدم بهینه‌سازی همزمان ساختار و خصوصیات گره‌هاست. الگوریتم، ابتدا ساختار را ثابت در نظر می‌گیرد و معنا را بهینه و سپس معنا را ثابت و ساختار را بهینه می‌کند. مرجع [۱۳]، تلاش دارد تا جوامع

همپوشان از نظر ویژگی‌های گره‌ها را در شبکه‌های اجتماعی مبتنی بر رتبه‌بندی شناسایی نماید. همجوشی چندلایه مبتنی بر ماتریس برای تشخیص جامعه در شبکه‌های نسبت داده شده در [۴۵] پیشنهاد شده است. این روش، یک روش پارتیشن‌بندی است که بایستی تعداد جوامع از پیش تعیین شده باشند. در [۴۶] از روش تکاملی جهت شناسایی جوامعی که در آن اعضاء دارای روابط توپولوژیکی قوی بوده و ویژگی‌های مشابه را به اشتراک می‌گذارند، استفاده شده است. یک الگوریتم تکاملی تک کروموزومی[۸] با یک عملگر اصلاح معماری در [۴۷] با در نظر گرفتن اتصال گره، تشابه ویژگی برای تشخیص جامعه پیشنهاد شده است. روش [۵]، ابتدا علایق را دسته‌بندی کرده و موضوعات را استخراج می‌کند. سپس، شبکه به خوشه‌های موضعی که در [۸] ارائه شده‌است، تقسیم می‌گردد. در [۵۱]، وابستگی‌های مابین ویژگی‌ها جهت شناسایی جوامع در نظر گرفته شده است. در [۴۹]، نویسندگان یک روش تشخیص جامعه همپوشانی نیمه نظارت شده را ابداع کردند که بر اساس گراف شبکه‌های عصبی است. در مرجع [۴۸]، یک رویکرد با استفاده از گراف شبکه عصبی، در شبکه‌های پرسش و پاسخ اجتماعی پیشنهاد شده است که هدف ترکیب متون منتشر شده در فضای مجازی و کاربران اجتماعی جهت شناسایی جوامع است.

در نهایت، الگوریتم خوشه‌بندی k-means بهبود یافته جهت تشخیص جوامع از شبکه‌ی بدست آمده از گام پیشین استفاده می‌شود. با وجود بهبود نتایج ماجولاریتی به عنوان یک معیار ارزیابی ساختاری نسبت به روش‌های مورد مقایسه، هیچ آزمایشی جهت ارزیابی از دیدگاه معنایی روش فوق ارائه نمی‌گردد. هر چند که خلاء الگوریتم‌های کارآمد مطرح در زمینه افزایش ماجولاریتی در ارزیابی‌ها جهت مقایسه مشهود است. جدول ۱ ویژگی برخی الگوریتم‌های ترکیبی را نشان می‌دهد.

با این حال، الگوریتم‌های شناسایی جوامع موجود تنها برخی ویژگی‌های محدود شبکه را پشتیبانی می‌نمایند و روش‌های ساختاری تنها توپولوژی شبکه را در نظر می گیرند. بنابراین گره‌های داخل یک اجتماع ممکن است ویژگی‌های معنایی و صفات مشترکی

جدول ۱ مقایسه برخی الگوریتم‌های ترکیبی

ویژگی‌ها / الگوریتم	نوع الگوریتم	قطعی/غیرقطعی	همپوشان	معیارهای ارزیابی
وو و همکاران [۴۸]	خوشه‌بندی	قطعی	ناهمپوشان	ماجولاریتی NMI
آکچار و همکاران [۵]	ماجولاریتی محور	قطعی	ناهمپوشان	ماجولاریتی NMI F-score
لو و همکاران [۴۵]	پارتیشن‌بندی	قطعی	ناهمپوشان	RI ARI NMI
پورعباسی و همکاران [۴۷]	خوشه‌بندی	غیرقطعی	ناهمپوشان	NMI شباهت ARI
وانگ و همکاران [۱۴]	خوشه‌بندی	غیرقطعی	ناهمپوشان	NMI شباهت F-score
موضوع‌گرا	خوشه‌بندی – ماجولاریتی	قطعی	همپوشان	ماجولاریتی خلوص F-score
رتبه‌بندی	خوشه‌بندی – ماجولاریتی	قطعی	همپوشان	ماجولاریتی خلوص F-score
هی و همکاران [۴۹]	خوشه‌بندی	غیرقطعی	همپوشان	ONMI F-score
ریحانیان و همکاران [۴۶]	ماجولاریتی محور	غیرقطعی	همپوشان	ماجولاریتی خلوص F-score
لوین توسعه‌یافته [۵۰]	ماجولاریتی محور	قطعی	ناهمپوشان	ماجولاریتی خلوص F-score

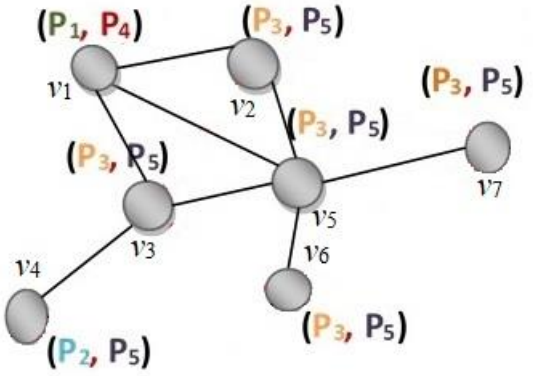
نداشته باشند و در عین حال گره‌های با ویژگی‌های یکسان در اجتماعات مختلف پراکنده گردند. روش‌های معنایی محض، ارتباطات با ارزش ساختاری را نادیده می‌گیرند. گره‌های یک اجتماع ممکن است ارتباطات صریح ساختاری نداشته باشند و جوامعی با گره‌های منفرد داخل اجتماع شکل گیرد. روش‌های ترکیبی موجود، راه‌حل اساسی برای ترکیب توپولوژی و معنایی ارائه نمی‌دهند. آنها اغلب متمایل به یکی از روش‌های ساختاری و یا ترکیبی هستند. غیرقطعی بودن از دیگر معایب اکثر روش‌های مطالعه شده اخیر است. در [۵۰] الگوریتم لوین را جهت پشتیبانی از معنا توسعه داده‌ایم. ولی همچنان ساختار و معنا به صورت بنیادی ترکیب نمی‌گردند. در این مقاله، یک روش قطعی مبتنی بر استنتاج‌های فازی جهت استخراج پیوندهای معنایی و در نتیجه افزایش کارایی تشخیص جوامع با توجه به معیارهای توپولوژیکی و ویژگی‌های گره پیشنهاد می‌نماییم.

۳- روش پیشنهادی

این بخش مکانیسم مبتنی بر منطق فازی را برای استخراج پیوندهای بین گره‌ها با ویژگی‌های مختلف و الزامات معنایی توضیح می‌دهد. ابتدا مساله‌ی مورد تحقیق تعریف، سپس روش تشخیص جوامع پیشنهادی ارائه می‌گردد.

۳-۱- بیان مساله

گراف شبکه اجتماعی، مطابق آنچه در شکل ۲ آورده شده است، به صورت سه‌تایی $G = (V, E, P)$ تعریف می‌گردد. $V = \{v_1, v_2, \dots, v_n\}$ هر گره می‌تواند با چند گره دیگر در ارتباط باشد. $E = \{e_1, e_2, \dots, e_m\}$ یال‌ها، اتصالات و تعاملات مابین گره‌های شبکه را نشان می‌دهد.

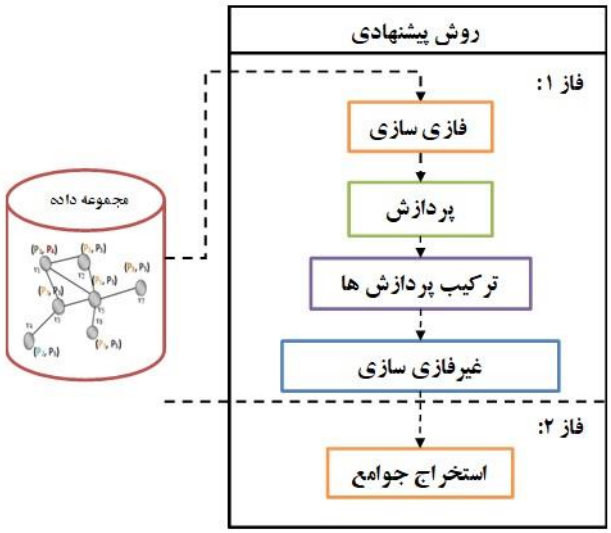


شکل ۲) شبکه اجتماعی با اتصالات ساختاری و ویژگی‌های گره‌ها.

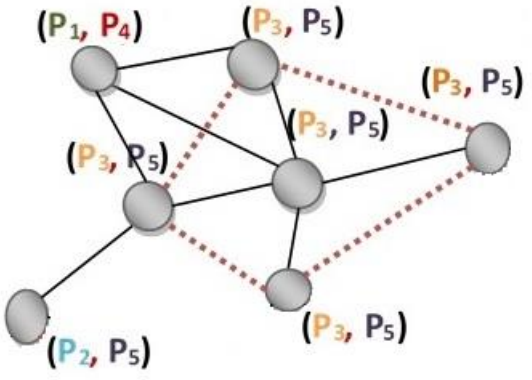
$P = \{p_1, p_2, \dots, p_k\}$ مجموعه موضوعات، زمینه‌ها، خصوصیات و ویژگی‌های گره‌هاست. یک گره مانند v_i می‌تواند دارای چند خصیصه باشد و چند گره می‌توانند دارای یک خصیصه مانند p_j باشند. بنابراین، میان گره‌ها و خصیصه‌ها رابطه‌ی چند به چند حاکم است. فرض می‌کنیم، خصوصیات هر عضو در مرحله تشخیص موضوعات استخراج شده و به هر گره انتساب داده شده است. فرآیند تشخیص جوامع، مجموعه گره‌های V را به q زیرمجموعه (اجتماع) $C = \{c_1, c_2, \dots, c_q\}$ با توجه به ساختار اتصالات E و ویژگی‌های گره P تقسیم می‌کند.

۳-۲- شناسایی جوامع فازی

در این بخش، روش شناسایی جوامع فازی پیشنهادی شرح داده می‌شود. همچنان‌که در شکل ۳ نشان داده شده است، سیستم پیشنهادی از دو فاز تشکیل یافته است؛ پردازش[۹] و پساپردازش[۱۰].



شکل ۳) چارچوب روش پیشنهادی برای تشخیص جوامع



شکل ۴) تشکیل یک پیوند مجازی بین دو گره بر اساس ویژگی‌ها و ساختار شبکه - پیوندهای مجازی با نقطه‌چین نمایش داده شده‌اند.

سیستم در فاز نخست، همانند شکل ۴، پیوندهای موثر بر مبنای ویژگی‌های معنایی و ساختاری را استخراج نموده و سپس در فاز پساپردازش جوامع را شناسایی می‌کند.

۳-۲-۱- پردازش

مرحله‌ی پردازش، سه ماتریس A، H و R را به عنوان ورودی دریافت می‌کند. برای جفت بردار گره $x = (v_i, v_j)$، خروجی $O_2(x)$ از مجموعه $\{0,1\}$ انتخاب می‌گردد. یک و صفر به ترتیب نشان دهنده وجود و یا عدم وجود یال مابین جفت گره (v_i, v_j) است. سیستم‌های فازی، به چهار بخش عمده؛ فازی‌ساز، پایگاه دانش، سیستم استنتاج و غیرفازی‌ساز، تقسیم می‌شوند.

- پردازش معنایی

فازی‌سازی معنایی: جهت استخراج شدت ارتباطات معنایی مابین گره‌ها نخست باید میزان ارتباط گره با خصوصیات و به‌طور کلی درجه‌ی عضویت هر گره در هر خصیصه بدست آید. همچنان‌که پیشتر بیان شد، k متغیر کلامی، به تعداد خصوصیات موجود در گراف برای هر گره v_i در لحظه‌ی شناسایی جوامع تعریف می‌گردد. از این رو، طبق رابطه‌ی ۱، هر خصیصه و ویژگی به عنوان یک متغیر کلامی در نظر گرفته می‌شود. در این گام باید میزان تعلق یک گره v_i به هرکدام از خصوصیات و ویژگی‌ها و یا به عبارت دیگر میزان عضویت هر گره v_i به متغیرهای کلامی تعیین شود.

$$T(V) = \{p_1, p_2, \dots, p_k\} \quad (۱)$$

فازی‌ساز، ورودی‌های غیرفازی x_i را توسط توابع عضویت به یک عدد فازی نگاشت می‌کند. جهت نیل به این هدف، دو رویکرد مختلف می‌توان متصور بود:

۱) به عنوان رویکرد نخست، اعداد فازی می‌توانند مستقیماً از مرحله‌ی شناسایی موضوعات[۱۱] و خصوصیات به سیستم وارد شوند. شناسایی موضوعات، مرحله‌ای قبل از شناسایی جوامع است که از چارچوب این مقاله خارج است.

۲) رویکرد دوم تعیین تابع عضویت است. بدین منظور رابطه‌ی ۲ تعریف می‌گردد. در این رابطه $OA_1(i,r)$ میزان تعلق گره v_i به ویژگی p_r را نشان می‌دهد.

$$OA_1(i,r) = \varphi \cdot normal(A_{i,r}) + (1-\varphi) \cdot \frac{\sum nb_r(v_i)}{\sum nb(v_i)} \quad (۲)$$

تابع $normal$، از رابطه ۳ به دست می‌آید، که عددی در محدوده‌ی [۱و۰] است. $nb(v_i)$، مجموعه همسایگان گره v_i است. $nb_r(v_i)$ همسایگان گره v_i می‌باشند که دارای خصیصه p_r هستند. در رابطه‌ی فوق میزان تاثیر ویژگی همسایگان و یا دوستان یک گره در آن توسط ضریب φ، که عددی مابین [۱و۰] است، تعیین می‌گردد. هر چه ضریب φ به یک نزدیکتر باشد میزان اثرگذاری دوستان یک عضو شبکه کمتر خواهد بود. خروجی این گام را می‌توان به صورت ماتریس OA_1 نمایش داد.

$$normal(A_{i,r}) = \frac{A_{i,r}}{\sum_{l=1}^{K} A_{i,l}} \quad (۳)$$

$$OA_1 = \begin{matrix} \\ V_1 \\ \vdots \\ V_n \end{matrix} \begin{matrix} \begin{matrix} P_1 & \cdots & P_k \end{matrix} \\ \begin{bmatrix} OA_1(1,1) & \cdots & OA_1(1,k) \\ \vdots & \ddots & \vdots \\ OA_1(n,1) & \cdots & OA_1(n,k) \end{bmatrix} \end{matrix} \quad (۴)$$

استنتاج معنایی: جهت ترسیم پیوندهای مجازی، در این مرحله، بر روی ورودی‌های فازی عمل استنتاج انجام می‌گیرد. ایده‌ی اصلی این است که خصوصیات با وابستگی‌های قوی، شباهت گره‌های مربوط به آن‌ها را افزایش می‌دهد. برای ماتریس‌های OA_1 و بردار حالت رئوس $(\overrightarrow{x_i}, \overrightarrow{x_j})$ حاوی k متغیر $\overrightarrow{x_i} = (a_{i,1}, a_{i,2}, \dots, a_{i,k})$ و $\overrightarrow{x_j} = (a_{j,1}, a_{j,2}, \dots, a_{j,k})$، قانون حالت (nST_i, nST_j) به صورت رابطه‌ی ۵ تعیین می‌گردد. در این رابطه $\overrightarrow{x_i}$ ($\overrightarrow{x_j}$) ردیف i (j) در OA_1 است. $a_{i,k}$ میزان عضویت فازی گره v_i در ویژگی k است.

$$if\ \overrightarrow{x_i}\ \ is\ nST_i\ and\ \overrightarrow{x_j}\ \ is\ \ nST_j\ then\ action\ is\ \ ACT_r \quad (۵)$$
$$1 \le i,j \le N\ ,1 \le r \le M$$

$\overrightarrow{x_i}$ میزان عضویت گره در مجموعه‌ی ویژگی‌هاست. nST_i یک حالت گره را با توجه به مجموعه ویژگی‌ها نشان می‌دهد. ACT_r، rامین عمل ممکن که توسط حالت (nST_i, nST_j) انتخاب شده است. ACT بصورت:

$$ACT = \{ACT_r,\ ACT_1: \text{وجود رابطه}\ ,\ ACT_0: \text{عدم وجود رابطه}\}$$

تعریف می‌گردد. هر عمل نشان‌دهنده‌ی قدرت رابطه‌ی معنایی بین هر جفت گره (v_i, v_j) و یک نشان‌دهنده‌ی وجود رابطه میان جفت گره (v_i, v_j) است. سیاست استنتاج فازی در این مقاله مجموع حاصل‌ضرب‌هاست[۱۲]. میزان عضویت گره v_i در مجموعه ویژگی‌ها، در میزان عضویت v_j در مجموعه ویژگی‌ها ضرب می‌گردد. بنابراین، T-normها از رابطه‌ی ۶ بدست می‌آیند. سپس نتایج بدست آمده با استفاده از رابطه ۷ تجمیع می‌گردند.

$$OA_{2.(i,j)}(m) = \overrightarrow{OA_1(i,r)}.\left(\overrightarrow{OA_1(j,r)}\right)^T \quad \forall r, 1 \le r \le K \quad (۶)$$

$$OA_{3.(i,j)} = \sum_{m=1}^{K} OA_{2.(i,j)}(m) \quad (۷)$$

$\left(\overrightarrow{OA_1(J,r)}\right)^T$ ترانهاده‌ی بردار $\overrightarrow{OA_1(J,r)}$ است. معادله‌ی ۶، ماتریس OA_1 را در ماتریس ترانهاده OA_1 ضرب می‌کند. S-norm $OA_{3.(i,j)}$ در معادله ۷، مجموع حاصل ضرب مقادیر عضویت فازی خروجی از معادله ۶ هست. بدین منظور، از ضرب برداری ماتریس استفاده می‌شود.

- پردازش ساختاری

برای ترکیب معنا و ساختار، نخست باید شدت ارتباطات ساختاری بر مبنای درجات فازی بدست آید. در این مقاله، شباهت جاکارد[۱۳] (رابطه ۸ [۵۲]) را برای جفت گره (v_i, v_j) از ماتریس H به عنوان تابع فازی‌ساز بکار می‌بریم. در این رابطه $nb(v_i)$، مجموعه همسایگان گره v_i است.

$$OS1(i,j) = \frac{|nb(v_i) \cap nb(v_j)|}{|nb(v_i) \cup nb(v_j)|} \quad (۸)$$

$$OS2(i,j) = \beta \cdot H + (1-\beta) \cdot OS1(i,j) \quad (۹)$$

ترکیب پردازش‌ها: خروجی OA_3 یک ماتریس n×n است که نشان‌دهنده‌ی وزن یال‌های مجازی مبتنی بر خصوصیات گره‌هاست. از طرفی، $OS2$ نشان‌دهنده‌ی شدت ارتباطات ساختاری با توجه به تعاملات یک گره نسبت به سایر گره‌هاست. ماتریس شبکه را با ترکیب ماتریس‌های $OS2$ و OA_3 بروزرسانی می‌نمائیم. رابطه‌ی ۱۰ نحوه‌ی ترکیب ساختار و معنا را نشان می‌دهد.

$$O1 = \gamma \cdot \ OA_3 + (1-\gamma) \cdot OS2 \quad (۱۰)$$

غیرفازی‌سازی: ورودی‌های الگوریتم فاز پساپردازش اعداد صریح {۰و۱} است. لذا، در این گام نیاز است تا ارتباطات با مقادیر فازی تولید شده را به ارتباطات صریح تبدیل کنیم. بدین منظور ارتباطات قوی را حفظ نموده و ارتباطات ضعیف را حذف می‌نماییم. خروجی نهایی O_2، برای دو گره دلخواه (v_i, v_j) از رابطه‌ی ۱۱ بدست می‌آید. دراین رابطه، TH شدت حدآستانه برای وجود و یا عدم وجود یال است.

$$O_2(i,j) = \begin{cases} 1 & O_1(i,j) \geq TH \\ 0 & O_1(i,j) < TH \end{cases} \quad (۱۱)$$

۳-۲-۲- پسا پردازش

پس از شناسایی پیوندهای مجازی، باید اجتماعات را از گراف غنی-شده O_2 استخراج نمائیم. ایده‌ی این تحقیق، بکارگیری الگوریتمی است که ماجولاریتی (رابطه ۱۲ [۲۴]) را در شناسایی جوامع گراف نهایی O_2 بیشترین مقدار را داشته باشد. چرا که:

- ✓ افزایش ماجولاریتی، سبب شناسایی جوامع ساختاری شبکه که دارای ارتباط قوی مابین خود و ارتباط تنک با گره‌های بیرونی هستند، می‌گردد.
- ✓ در گراف O_2، یال‌ها علاوه بر ارتباطات ساختاری، نشان‌دهنده ارتباطات معنایی نیز می‌باشند. بنابراین، انتخاب الگوریتم مناسب که ماجولاریتی را افزایش دهد، گره‌های با ارتباطات معنایی بالا را در یک اجتماع قرار می‌دهد.

$$Q = \frac{1}{2m}\sum_{uv}\left(H_{uv} - \frac{k_u k_v}{2m}\right)\delta(C_u, C_v) \quad (۱۲)$$

m، تعداد کل یال‌های شبکه، A_{vw} وزن یال بین جفت گره v و w. اگر شبکه بدون وزن باشد، وزن یال یک در نظر گرفته می‌شود. K_v و K_w، درجات گره‌های v و w است. $\delta(C_v, C_w)$، برابر با یک است، اگر گره‌های v و w در یک اجتماع باشند، در غیر این‌صورت مقدار آن صفر است.

بنابراین، الگوریتمی که جوامع با بیشینه ماجولاریتی را شناسایی نماید، می‌تواند برای این گام مطلوب باشد. در عین حال چون جوامع بدست آمده در گام نخست نیمه ماجولار بوده، بنابراین سرعت بالا نسبت به دقت بالا ارجح خواهد بود. از الگوریتم بیشینه‌سازی ماجولاریتی [۲۲] با توجه به سرعت بالا و دقت مطلوب جهت اعمال بر روی گراف O_2 بهره می‌بریم (الگوریتم ۱). البته که هر الگوریتمی که بتواند سطح ماجولاریتی را در شناسایی جوامع بالا ببرد قابل استفاده خواهد بود.

الگوریتم بیشینه‌سازی ماجولاریتی: بدین منظور، یک ماتریس ΔQ تعریف شده است، که هر عضو ΔQ_{ij} در آن مربوط به جفت اجتماع i و j است که حداقل یک یال مابین آن‌ها وجود داشته باشد. مقدار اولیه ΔQ_{ij} از رابطه‌ی ۱۳ و بروزرسانی آن از رابطه‌ی ۱۴ بدست می‌آید [۲۲]. در این رابطه، $a_i = \frac{k_i}{2m}$ است.

$$\Delta Q_{ij} = \begin{cases} \frac{1}{2m}O_2 - \frac{k_i k_j}{2m} & if\ i\ and\ j\ are\ connected \\ 0 & otherwise, \end{cases} \quad (۱۳)$$

$$\Delta Q^*_{ik} = \begin{cases} \Delta Q_{ik} + \Delta Q_{jk}, if\ community\ k\ is\ connected\ to\ i.j \\ \Delta Q_{ik} - 2a_j a_k, if\ community\ k\ is\ connected\ to\ i\ bot\ noto\ j \\ \Delta Q_{jk} - 2a_i a_k, if\ community\ k\ is\ connected\ to\ j\ bot\ not\ \ i \end{cases} \quad (۱۴)$$

الگوریتم ۱: بیشینه‌سازی ماجولاریتی [۲۲]

ورودی: ماتریس O_2

خروجی: مجموعه جوامع C شناسایی شده

۱. **مقداردهی اولیه:** ماتریس‌های ΔQ و a_i طبق رابطه ۱۳
۲. اگر |C|>1
۳. برای هر سطر i در ماتریس ΔQ
۴. یک درخت *max-heap* به نام MH بر اساس بیشترین مقادیر بساز
۵. از ریشه MH، $\{i, j\}$ با بیشترین ΔQ_{ij} را انتخاب کن.
۶. دو اجتماع i و j را ترکیب کن.
۷. ΔQ را بر اساس رابطه‌ی ۱۴ بروزرسانی کن.
۸. سطر j را در ΔQ حذف کن.
۹. ستون j را در ΔQ حذف کن.
۱۰. برو به مرحله ۲.
۱۱. جوامع با بیشترین ماجولاریتی را برگردان.

۴- ارزیابی و نتایج شبیه‌سازی

روش شناسایی جوامع پیشنهادی در متلب[۱۴] [۵۳] بعنوان یکی از مشهورترین و قدرتمندترین زبان‌های شبیه‌سازی، پیاده‌سازی، مورد ارزیابی و با سایر الگوریتم‌های مشهور تحت مجموعه‌ای از پارامترهای شبیه‌سازی، مورد مقایسه قرار می‌گیرد. تمام آزمایشات در سیستمی با پردازنده‌ی اینتل چهار هسته‌ای ۲.۲ گیگاهرتز[۱۵] با حافظه ۸ گیگابایت[۱۶] انجام می‌پذیرد. برای ارزیابی روش پیشنهادی، تنظیم پارامترهای φ و TH، تعریف شده در مرحله پردازش، به عنوان ورودی‌ها ضروری هستند. بر اساس آزمایش‌های متعدد، $\varphi = 0.4$ و $TH = 0.52$ پاسخ بهتری دارند.

۴-۱- معیارهای ارزیابی

معیارهای کیفی مختلفی جهت ارزیابی کیفیت جوامع شناسایی شده ارائه شده‌اند. یکی از مهمترین معیارهای ارزیابی کیفیت ساختاری، ماجولاریتی[۱۷] (Q) است [۸]. خلوص[۱۸] یک معیار مهم برای ارزیابی ویژگی‌های معنایی و کیفی گره‌ها در جوامع است. معیارهای فوق به همراه تابع PurQ_β برای تجزیه و تحلیل روش پیشنهادی و مقایسه آن با الگوریتم‌های موجود استفاده شده است.

۴-۱-۱- ماجولاریتی

جهت مقایسه همزمان عملکرد الگوریتم‌های ساختاری، معنایی و ترکیبی با ویژگی‌های همپوشانی یا ناهمپوشانی جوامع، معادله ۱۵ ارائه شده در [۵۴] را بکار می‌بریم؛ که در آن $\delta(C_u, C_v)$ با $Suv = \sum_{c\in C} a_{u,c}a_{v,c}$ در رابطه‌ی ۱۳ جایگزین شده است. $a_{u,c}$ شدت عضویت گره u در اجتماع c را نشان می‌دهد ($\sum_{c=1}^{|C|} a_{u,c} = 1$). برای جوامع غیرهمپوشان، فقط یک جامعه c برای هر گره با $a_{u,c}$ با مقدار ۱ وجود دارد. بنابراین، معادله ۱۵ به رابطه ۱۲ کاهش می‌یابد [۵۵].

$$Q = \frac{1}{2m}\sum_{c\in C}\sum_{uv}\left(H_{uv} - \frac{k_u k_v}{2m}\right)a_{u,c}a_{v,c} \qquad (۱۵)$$

۴-۱-۲- خلوص

نشان‌دهنده‌ی شباهت معنایی گره‌های داخل یک اجتماع است که از میانگین اختلاف میان میزان علاقه‌مندی کاربران یک اجتماع به موضوعات مختلف بدست می‌آید. هر چه این اختلاف کمتر باشد، گره‌های آن اجتماع از نظر خصوصیات مختلف به یکدیگر شباهت دارند [۸].

$$Purity = \frac{1}{N_{cm}}\sum_{i=1}^{N_{cm}} \underset{1\le j\le K}{max}\left\{\frac{n_{ij}}{n_i}\right\} \qquad (۱۶)$$

N_{cm}، تعداد جوامع حاصل از اجرای الگوریتم است. n_{ij}، تعداد گره‌هایی که دارای ویژگی j بوده و در اجتماع i قرار گرفته است. n_i، تعداد اعضای اجتماع i را نشان می‌دهد. خلوص بالا نشان‌دهنده‌ی جوامع بهتر، از نظر خصوصیات گره‌هاست.

۴-۱-۳- PurQ

یک معیار ارزیابی برگرفته از اندازه‌گیری F[۱۹] که توسط ژائو و همکاران [۸] ارائه شده و در مراجع مختلف به کار گرفته شده است. معیار مناسب جهت بررسی کارآیی الگوریتم‌ها باید هر دو جنبه‌ی ساختار و ویژگی را در نظر بگیرد. این معیار از رابطه‌ی ۱۷ به دست می‌آید و نشان‌دهنده‌ی این است که یک الگوریتم، چه میزان عدالت را مابین ماجولاریتی و خلوص و در نتیجه ساختار و معنا را رعایت می‌نماید. β، یک پارامتر برای تنظیم وزن خلوص و Q است. $\beta \in [0, \infty]$.

$$PurQ_\beta = (1+\beta^2).\frac{Purity.Q}{(\beta^2.Purity + Q)} \qquad (۱۷)$$

$\beta = 1$ به این معنی است که خلوص و ماجولاریتی و در نتیجه خصوصیات و توپولوژی شبکه‌های اجتماعی اهمیت برابر دارند. اگر $1 < \beta < \infty$، ساختار جوامع اهمیت بیشتری نسبت به خصوصیات گره‌ها دارد. برای $0 < \beta < 1$، شباهت خصوصیات اعضای جوامع مهمتر از ساختار آنهاست.

۴-۲- مجموعه‌ی داده

برای مقایسه‌ی عملکرد الگوریتم‌ها، به مجموعه داده‌هایی نیاز داریم که علاوه بر ساختار، معنا را نیز پشتیبانی کنند. شبکه‌های مصنوعی

و دنیای واقعی با ویژگی‌های گره‌ها، در انواع و اندازه‌های مختلف، برای ارزیابی کارایی روش پیشنهادی استفاده شده‌اند.

۴-۲-۱- مجموعه داده‌ی مصنوعی

برای تولید مجموعه داده شامل ویژگی‌ها و جوامع، مولد گراف پیشنهاد شده در [۵۶] استفاده شده است. در شبکه‌های تولید شده، ویژگی‌ها از توزیع نرمال چندمتغیره‌ی kبعدی[۲۰] پیروی می‌کنند. از جوامع مبنا، جهت ارزیابی صحت روش پیشنهادی استفاده می‌گردد. همان‌گونه که در جدول ۲ نمایش داده شده است، پنج شبکه‌ی معیار[۲۱] با تعداد یکسانی از جوامع و ویژگی‌های مبنا تولید شده است. جوامع با تأثیر هموفیلی بین توپولوژی شبکه و ویژگی‌های گره ایجاد می‌گردند. n و m به ترتیب تعداد گره‌ها و یال‌ها هستند. a تعداد صفات است. k و Q تعداد جوامع و ماجولاریتی جوامع مبنا[۲۲] هستند.

۴-۲-۲- مجموعه داده‌ی طبیعی

مجموعه داده فوتبال [۲۱]: این مجموعه داده، شامل شبکه‌ای از بازی‌های فوتبال آمریکایی بین دانشکده‌های مختلف آمریکا در فصل پاییز سال ۲۰۰۰ می‌باشد. تعداد ۱۱۵ گره، نشان‌دهنده‌ی تیم‌ها و ۶۱۳ یال، بازی‌های مابین دو تیم را نشان می‌دهند. این مجموعه دارای ۱۲ ویژگی است. ویژگی‌ها نشان‌دهنده همایش‌هایی هستند که تیم‌ها در آن شرکت داشته‌اند. مقادیر به شرح زیر است: ۰ = «ساحل اقیانوس اطلس[۲۳]»، ۱ = «شرق بزرگ[۲۴]»،

جدول ۲ اطلاعات شبکه‌های مصنوعی مورد استفاده در آزمایشات

شبکه معیار	n	m	A	k
شبکه ۱	۱۴۳	۵۲۷	۲	۵
شبکه ۲	۲۸۶	۱۳۱۷	۲	۵
شبکه ۳	۵۷۱	۳۲۵۱	۲	۵
شبکه ۴	۱۰۰۰	۴۵۲۱	۲	۵
شبکه ۵	۲۰۰۰	۴۲۳۵	۲	۵

۲ = «ده بزرگ[۲۵] »، ۳ = «دوازده بزرگ[۲۶]»، ۴ = «کنفرانس آمریکا[۲۷]»، ۵ = «مستقل‌ها[۲۸]»، ۶ = «آمریکای مرکزی[۲۹]»، ۷ =« کوهستان غرب[۳۰]»، ۸ =«ده آرام[۳۱]»، ۹ =«جنوب شرقی[۳۲]»، ۱۰ =«کمربند خورشید[۳۳]»، ۱۱ =«ورزشگاه غربی[۳۴]».

مجموعه **داده کورا**[۳۵]: این مجموعه، زیرمجموعه‌ای از مجموعه استنادی کورا است که در آن ۲۷۰۸ گره و ۵۴۲۹ یال وجود دارد. هر گره نشان دهنده‌ی یک مقاله و یال‌ها نمایانگر روابط استنادی هستند. تمام نشریات موجود در مجموعه داده‌ها در هفت موضوع تحقیقاتی در زمینه یادگیری ماشین شامل منطق مبتنی بر مورد[۳۶]، الگوریتم‌های ژنتیک، شبکه‌های عصبی، روش‌های احتمالاتی، یادگیری تقویتی، یادگیری قوانین[۳۷] و نظریه طبقه‌بندی می‌شوند [۸]. در این تحقیق از ۴ ویژگی (۲، ۵، ۱۰، ۲۰) استفاده گردید که شبکه‌ی خلاصه شده شامل ۶۰۷ گره و ۷۳۱ پیوند برای تشخیص نتایج است. برای ارزیابی حالت ویژگی‌های وابسته دو ویژگی ۲ و ۱۰ و دو ویژگی ۵ و ۲۰ با یکدیگر پیوند داده شده‌اند.

مجموعه داده کتاب‌های سیاسی[۳۸] [۵۷]: این مجموعه داده، شامل یک شبکه است که در آن گره‌ها (۱۰۵ گره) کتاب‌های مربوط به سیاست‌های ایالات متحده را که توسط کتابفروشی برخط آمازون[۳۹] به فروش می‌رسد، نشان می‌دهد و یال‌ها (۴۴۱ یال)، تعداد خرید همزمان از کتاب‌های مشابه توسط خریداران یکسان است. گره‌ها دارای یکی از سه ویژگی «لیبرال»، «بی‌طرف» یا «محافظه-کار» هستند.

مجموعه داده‌ی فلیکر[۴۰]: یک شبکه اجتماعی به اشتراک‌گذاری عکس است که از تصاویر دارای برچسب[۴۱] ساخته شده است. در سایت فلیکر، کاربران می‌توانند مطالب خود را به اشتراک بگذارند، برچسب‌ها را بارگذاری کرده و در گروه‌های مختلف عضو شوند [۵۸]. این مجموعه شامل ۸۰۵۱۳ گره و ۱۹۵ گروه علایق به عنوان ویژگی گره‌هاست. در این تحقیق، سه گروه (۲ ، ۵ و ۱۰) برای ارزیابی انتخاب شده‌اند. شبکه‌ی هدف دارای ۱۲۸۹ گره با ۹۰۳۳ یال است.

مجموعه داده‌ی وبلاگ‌های سیاسی[۴۲] [۵۹]: این مجموعه داده که در سال ۲۰۰۵ ارائه شد، یک شبکه وبلاگ در مورد مسائل سیاسی ایالات متحده است. در مجموع ۱۴۹۰ وبلاگ و ۱۹۰۹۰ یال میان آنها وجود دارد. هر وبلاگ، ویژگی «محافظه‌کار» و یا «لیبرال»، با مقادیر صفر برای لیبرال، یک برای محافظه کار دارد.

مجموعه داده‌ی اسامی و صفات[۴۳] [۲۴]: این مجموعه داده شامل شبکه عناصر مشترک و اسامی و صفات مجاور در رمان «دیوید کاپ فیلد[۴۴]» چارلز دیکنز است. گره‌ها (۱۱۲ گره) نشان‌دهنده‌ی صفات و اسامی مشترک است که با تعداد بالا در این کتاب وجود دارند. یال‌ها (۴۲۵ یال) جفت کلماتی را که در متن در کنار یکدیگر قرار دارند را متصل می‌نمایند. ویژگی‌های گره‌ها، صفر برای «صفات» و یک برای «اسامی» است.

۴-۳- نتایج شبیه‌سازی

الگوریتم‌های مورد مقایسه همچنان که در جدول ۳ مشاهده می‌گردد، با روش پیشنهادی به سه دسته ساختاری، معنایی و ترکیبی تقسیم می‌گردد.

۴-۳-۱- ارزیابی در شبکه‌های مصنوعی

به منظور راستی‌آزمایی روش پیشنهادی و ارزیابی کارآیی آن، Q را در مقایسه با الگوریتم‌های دیگر ارزیابی می‌نمائیم. جدول ۴ نتایج Q ساختاری حاصل از الگوریتم‌ها را در شبکه‌های معیار جدول ۲ نشان می‌دهند. آزمایشات از نتایج نامطلوب هر دو روش معنایی و ترکیبی حکایت دارند، چرا که شبکه را بر اساس ویژگی‌های گره‌ها تکه‌تکه می‌کنند. الگوریتم‌های ترکیبی به دلیل ملاحظات توپولوژیکی نتایج نسبتاً بهتری نسبت به خوشه‌بندی معنایی دارند.

جدول ۳ الگوریتم‌های مقایسه شده با روش پیشنهادی

الگوریتم	نام و مرجع	نوع
۱	بیشینه‌سازی ماجولاریتی [۲۲]	ساختاری
۲	لوین [۲۳]	
۳	الگوریتم سریع [۲۵]	
۴	خوشه‌بندی معنایی [۶۰]	معنایی
۵	موضوع‌گرا (ژائو و همکاران ۲۰۱۲)	ترکیبی

روش‌های ساختاری نتایج قابل قبولی از خود بجای گذاشته‌اند، اما معنا را در نظر نمی‌گیرند. با این حال، عملکرد مناسبی ندارد. روش پیشنهادی با در نظر گرفتن متعادل ساختار و معنا، از نظر Q عملکرد مطلوبی دارد.

جدول ۴ نشان می‌دهد که الگوریتم‌های ترکیبی و در کنار آنها روش ارائه شده در این مقاله، به دلیل در نظرگرفتن ویژگی‌های گره‌ها در شناسایی جوامع، عملکرد بهتری نسبت به روش‌های ساختاری با خلوص نزدیک به یک دارند. روش‌های معنایی به دلیل این‌که بر اساس ویژگی‌های یکسان جوامع را تقسیم‌بندی می‌نمایند، بیشترین خلوص به میزان یک خواهند داشت. روش پیشنهادی خلوصی بسیار فراتر از الگوریتم‌های بیشینه‌سازی ماجولاریتی و لوین و نزدیک به خلوص روش معنایی دارد.

۴-۳-۲- ارزیابی در شبکه‌های واقعی

جدول ۵ ماجولاریتی جوامع شناسایی شده توسط الگوریتم‌های مختلف در مجموعه داده‌های دنیای واقعی را نشان می‌دهند. همان‌طور که در این جدول گزارش شده است، روش‌های ساختاری در کنار روش ما در مقایسه با الگوریتم معنایی و حتی الگوریتم‌های ترکیبی در ماجولاریتی عملکرد بهتری دارند. الگوریتم‌های ساختاری حریصانه، حامی ماجولاریتی بوده و ویژگی‌های گره را در تشخیص جوامع نادیده می‌گیرند. الگوریتم پیشنهادی عملکرد بهتری نسبت به الگوریتم‌های دیگر در اغلب موارد دارد. این الگوریتم با در نظر گرفتن هدفمند توپولوژی شبکه، جوامع را کشف می‌کند که منجر به ماجولاریتی بالا می‌گردد. در نتایج آزمایشات روش معنایی گاهی اوقات خیلی ضعیف نیستند که ممکن است به این دلیل باشد که گره‌های دارای ویژگی‌های مشترک ارتباط بیشتری با یکدیگر برقرار می‌نمایند. جدول ۶ خلوص الگوریتم‌های مختلف را نشان می‌دهد. این جدول نشان می‌دهد که الگوریتم‌های معنایی و ترکیبی و در کنار آنها روش ارائه شده در این مقاله، به دلیل در نظرگرفتن ویژگی‌های گره‌ها در شناسایی جوامع، عملکرد بهتری نسبت به روش‌های ساختاری با خلوص نزدیک به یک دارند.

جدول ۴ مقایسه مقادیر ماجولاریتی الگوریتم‌ها تحت معیارهای مصنوعی

الگوریتم / شبکه معیار	بیشینه‌سازی ماجولاریتی	لوین	الگوریتم سریع	خوشه‌بندی معنایی	موضوع‌گرا	روش پیشنهادی
۱	۰/۷۳۱۵	۰/۷۲۹۷	۰/۷۳۵۲	۰/۰۴۱۴	۰/۳۲۷۶	۰/۷۳۶۰
۲	۰/۶۴۱۶	۰/۶۳۷۷	۰/۶۴۵۷	-۰/۰۰۷۵	۰/۲۵۶۲	۰/۶۵۲۶
۳	۰/۶۶۳۶	۰/۶۷۲۱	۰/۶۷۵۷	۰/۰۱۵۹	۰/۲۷۷۲	۰/۶۷۶۶
۴	۰/۶۴۰۳	۰/۶۴۰۴	۰/۶۴۳۹	۰/۰۱۱۳	۰/۲۴۱۵	۰/۶۴۴۹
۵	۰/۶۱۸۹	۰/۶۱۲۴	۰/۶۱۶۴	۰/۰۲۰۱	۰/۲۹۳۴	۰/۶۰۵۵

جدول ۴ مقایسه مقادیر خلوص الگوریتم‌ها تحت معیارهای مصنوعی

الگوریتم / شبکه معیار	بیشینه‌سازی ماجولاریتی	لوین	الگوریتم سریع	خوشه‌بندی معنایی	موضوع‌گرا	**روش پیشنهادی**
۱	۰/۸۶۳۸	۰/۸۶۳۸	۰/۸۶۳۸	۱	۱	۰/۹۴۹۱
۲	۰/۹۸۹۴	۰/۹۸۹۴	۰/۸۹۰۴	۱	۱	۰/۹۷۲۱
۳	۰/۷۸۰۲	۰/۷۸۰۲	۰/۷۸۰۲	۱	۱	۰/۹۵۰۳
۴	۰/۸۶۲۱	۰/۸۶۲۱	۰/۸۶۲۱	۱	۱	۰/۹۶۵۲
۵	۰/۸۲۹۵	۰/۸۲۹۵	۰/۸۲۹۵	۱	۱	۰/۹۷۲۱

جدول ۵ مقایسه مقادیر ماجولاریتی الگوریتم‌ها تحت معیارهای واقعی

الگوریتم / شبکه معیار	بیشینه‌سازی ماجولاریتی	لوین	الگوریتم سریع	خوشه‌بندی معنایی	موضوع‌گرا	روش پیشنهادی
فوتبال	۰/۵۶۶۸	۰/۵۸۱۱	۰/۵۷۸۳	۰/۵۵۴۰	۰/۵۵۴۶	۰/۵۸۴۱
کورا	۰/۸۸۷۸	۰/۸۸۹۶	۰/۸۹۱۲	۰/۲۳۰۱	۰/۷۹۸۰	۰/۸۸۴۲
کتاب سیاسی	۰/۴۹۱۶	۰/۴۷۲۷	۰/۴۹۹۳	۰/۴۱۴۹	۰/۲۹۶۹	۰/۵۱۲۳
وبلاگ سیاسی	۰/۴۱۲۱	۰/۴۲۶۲	۰/۴۲۶۷	۰/۴۰۵۲	۰/۴۰۵۲	۰/۴۲۵۱
فلیکر	۰/۳۹۶۳	۰/۴۰۰۳	۰/۴۲۴۸	۰/۲۹۱۸	۰/۲۸۶۳	۰/۴۲۶۶
نام-صفت	۰/۲۶۳۱	۰/۲۵۲۰	۰/۲۵۹۶	-۰/۲۱۱۹	۰/۰۹۱۵	۰/۲۶۱۱

جدول ۶ مقایسه مقادیر خلوص الگوریتم‌ها تحت معیارهای واقعی

الگوریتم / شبکه معیار	بیشینه‌سازی ماجولاریتی	لوین	الگوریتم سریع	خوشه‌بندی معنایی	موضوع‌گرا	روش پیشنهادی
فوتبال	**۰/۸۶۹۷**	**۰/۹۰۸۷**	**۰/۷۸۸۳**	۱	۱	**۰/۹۸۵۵**
کورا	**۰/۹۴۵۶**	**۰/۹۴۳۷**	**۰/۹۴۴۷**	۱	۱	**۰/۹۹۳۲**
کتاب سیاسی	**۰/۷۸۵۲**	**۰/۸۵۴۷**	**۰/۷۳۶۳**	۱	۱	**۰/۹۴۴۱**
وبلاگ سیاسی	**۰/۹۱۹۲**	**۰/۹۷۰۲**	**۰/۸۹۹۲**	۱	۱	**۰/۹۹۸۲**
فلیکر	**۰/۸۳۸۳**	**۰/۸۳۰۰**	**۰/۸۷۰۹**	۱	۱	**۰/۹۹۳۸**
نام-صفت	**۰/۶۷۴۴**	**۰/۶۹۷۳**	**۰/۶۵۲۲**	۱	۱	**۰/۹۸۲۸**

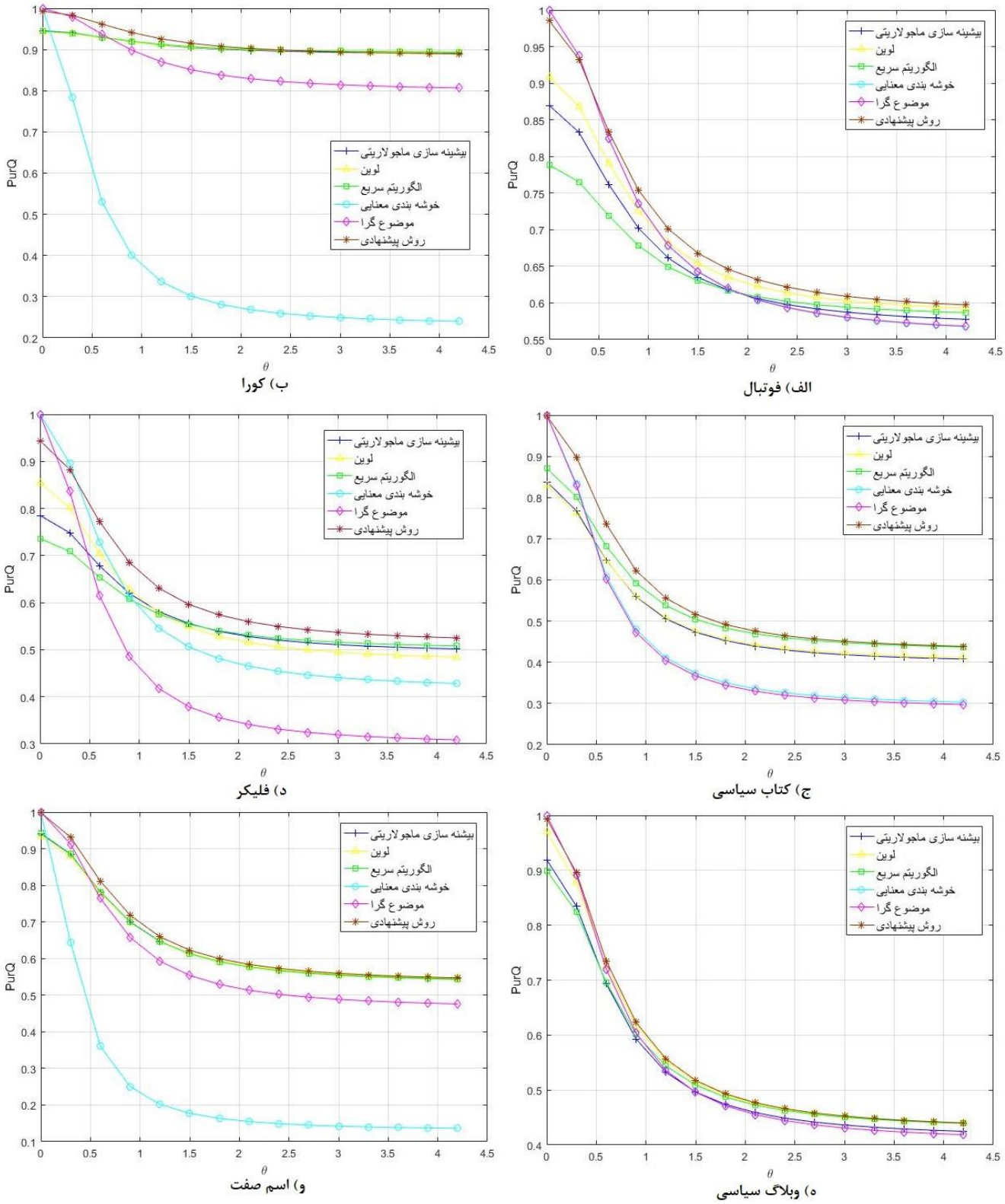
الگوریتم معنایی، به دلیل ماهیت خود، دارای بالاترین خلوص با مقدار یک است. الگوریتم‌های معنایی و ترکیبی از نظر خلوص کمی بهتر از الگوریتم دو فازی این مقاله هستند، زیرا روش‌های معنایی بدون در نظرگرفتن ساختار شبکه و ماجولاریتی، و روش‌های ترکیبی موجود در گام ابتدایی، حریصانه تنها براساس ویژگی‌های گره‌ها خوشه‌بندی می‌کنند.

براساس نتایج ماجولاریتی و خلوص هر یک از الگوریتم‌ها، $PurQ_\theta$ بر اساس این نتایج محاسبه می‌شود تا مشخص گردد در مجموع و با سنجش همزمان ساختار و معنا، کدامیک از الگوریتم‌ها عملکرد بهتری داشته‌اند. شکل ۵ نشان‌دهنده‌ی عملکرد الگوریتم‌های مقایسه شده است. در «$\theta < 1$» ارزش شباهت ویژگی گره‌ها دارای وزن بالاتری است. بنابراین، هر چه از یک به سمت عدد صفر نزدیک می‌شویم، روش‌های معنایی نتایج بهتری خواهند داشت.

اما با نزدیک‌تر شدن به یک به علت افزایش تاثیرات ماجولاریتی، این دو الگوریتم نتیجه‌ی ضعیف‌تری خواهند داشت. روش پیشنهادی به علت اینکه خلوص و ماجولاریتی را هدفمند و کنترل‌شده در تصمیم‌گیری‌ها منظور می‌دارد، از افت Q جلوگیری می‌نماید. «$\theta = 1$» برابری اهمیت خلوص خصوصیات گره‌ها و پیوندهای ساختاری را نشان می‌دهد. در این محدوده، الگوریتمی که بیشترین نتیجه را داراست، عدالت[۴۵] را بهتر مابین توپولوژی و صفات رعایت می‌نماید. روش پیشنهادی بهترین نتیجه را در این محدوده دارد. «$\theta > 1$» نشان‌دهنده‌ی وزن بالاتر ارزش پیوندهای ساختاری است. نتایج به وضوح حاکی از آن هستند که الگوریتم پیشنهادی عملکرد بسیار خوبی در PurQ داشته و از الگوریتم‌های دیگر بهتر عمل می‌کند. زیرا الزامات توپولوژی شبکه و ویژگی‌های گره را در تصمیم‌گیری‌ها هدفمندانه رعایت کرده و در نتیجه تشخیص جامعه را تضمین می‌کند. بنابراین هرچه θ افزایش می‌یابد، نتایج بیشینه‌سازی ماجولاریتی و لوین به الگوریتم پیشنهادی نزدیک‌تر می‌شوند. الگوریتم موضوع‌گرا، در نتیجه ماجولاریتی ناکارآمد داشته و PurQ مناسبی ندارد.

۴-۴- پیچیدگی محاسباتی

تعداد گره‌ها، n و لبه‌ها، m تأثیر اصلی را بر پیچیدگی محاسباتی دارند. تعداد خصوصیات k در مقایسه با n ناچیز است. الگوریتم پیشنهادی، گام پردازش (فاز نخست) را جهت شناسایی پیوندهای مجازی و گام پساپردازش (فاز دوم) را جهت استخراج جوامع بکار می‌گیرد. به دلیل اینکه این دو فاز پشت سر هم اجرا می‌شوند، پیچیدگی نهایی الگوریتم پیشنهادی از مجموع پیچیدگی‌های فاز نخست و فاز دوم بدست می‌آید. در گام پردازش، پیچیدگی زمانی فازی‌ساز $O(nk + mk + m) \approx O(n + m)$ است که در شبکه‌های پراکنده[۴۶] به $O(n)$ کاهش می‌یابد. مرحله‌ی استنتاج



شکل ۵) تحلیل کارآیی $\mathbf{PurQ_{\theta}}$ جوامع حاصل از الگوریتم‌های مختلف در مجموعه داده‌های دنیای واقعی برای $\boldsymbol{\theta}$

فازی از ضرب دو ماتریسی با پیچیدگی $O(mult)$ برای همه در یک استنتاج استفاده می‌کند که برای حالت استقلال ویژگی‌ها در بدترین حالت $O((k+1)n^2)$ خواهد بود. با این حال، الگوریتم‌های ضرب ماتریسی با هزینه‌های زمانی کمتر، پیشتر در منابع مختلف مورد مطالعه قرار گرفته‌اند. در غیرفازی‌سازی، تنها یک مقایسه برای هر یال انجام می‌شود. بنابراین، زمان اجرا برای همه‌ی m یال $O(m)$

است و در نتیجه، پیچیدگی زمانی کل فاز پردازش $O\ (mult + m)$ است. در مرحله پساپردازش، الگوریتم بیشینه‌سازی ماجولاریتی (الگوریتم ۱) استفاده شده است. لذا، پیچیدگی محاسباتی از مرتبه $O(nlog^2n)$ خواهد بود.

۵- بحث و نتیجه‌گیری

به طور کلی، روش‌های موجود را می‌توان به سه دسته راه‌حل‌های ساختاری، معنایی و ترکیبی دسته‌بندی کرد. اکثر مطالعات بدون توجه به ویژگی‌های گره‌ها و محتوای شبکه، متمرکز بر ساختار شبکه هستند. دیدگاه پشت این روش‌ها اغلب بر این فرض استوار است که گره‌هایی با اتصالات قوی، ویژگی‌های یکسانی دارند. اما، در دنیای واقعی، افراد ممکن است با یکدیگر تعامل داشته باشند در حالی که ویژگی‌های متفاوتی دارند. با این حال، نتایج ماجولاریتی بیشتر این الگوریتم‌ها مناسب است، چرا که جوامع بر اساس تعاملات توپولوژیکی گره‌ها شناسایی می‌شوند. روش‌های معنایی منجر به از دست رفتن اطلاعات ارزشمند توپولوژیکی و نتایج ضعیف ماجولاریتی می‌شود. راه‌حل‌های ترکیبی متمایل به یکی از انواع یاد شده هستند.

این مقاله روشی را برای مدیریت ویژگی‌ها در کنار ساختار شبکه برای تشخیص جوامع در شبکه‌های اجتماعی معرفی می‌نماید. روش پیشنهادی ساختار شبکه و ویژگی‌های معنایی را همزمان منظور می‌دارد. در این رویکرد، مساله توپولوژی شبکه و ویژگی‌های گره‌ها در شرایط پویای هر گره و پایگاه دانش سیستم آدرس‌دهی می‌گردد. یک رابطه‌ی پیچیده بین توپولوژی و معنا وجود دارد. این مطالعه ابتدا توپولوژی شبکه و معنا را با معیارهای یکسان در قالب ساختاری جمع و سپس، جوامع را در شبکه‌ی غنی‌شده‌ی هدف بر اساس بیشنه ماجولاریتی استخراج می‌نماید. از منطق فازی برای کشف پیوندهایی استفاده می‌شود که می‌توانند مابین گره‌ها تحت شرایط مختلف برای ترکیب ساختار و معنا بکار گرفته شوند.

از دیدگاه عملکرد ساختاری، نتایج ماجولاریتی برای شبکه‌های مصنوعی و طبیعی اندازه‌گیری شده است. روش‌های ساختاری (الگوریتم‌های لوین و بیشینه‌سازی ماجولاریتی) نتایج قابل قبولی از خود بجای گذاشته‌اند، اما معنا را در نظر نمی‌گیرند. آزمایشات از نتایج نامطلوب هر دو روش معنایی و ترکیبی حکایت دارند چرا که به علت ماهیت، شبکه را بر اساس ویژگی‌های گره‌ها تکه‌تکه می‌کنند. نتایج الگوریتم‌های ترکیبی، به علت تاثیر ماجولاریتی در خوشه‌بندی، بهتر از روش معنایی محض است. نکته‌ی دیگر، نزدیک بودن نتایج لوین و بیشینه‌سازی ماجولاریتی می‌باشد. روش مطالعه شده به دلیل پشتیبانی از ویژگی‌های گره و روابط آنها برای شناسایی جوامع، بسیار بهتر از سایر الگوریتم‌ها عمل می‌کند.

از دیدگاه معنایی، میزان و میانگین خلوص ویژگی‌ها در اجتماعات اندازه‌گیری شده‌اند. قاعدتا، الگوریتم معنایی محض و ترکیبی به جهت ماهیت حریصانه، خلوصی به اندازه یک خواهد داشت، چرا که گره‌های با ویژگی یکسان را در خوشه‌ی یکسان قرار می‌دهد. در کنار آنها، الگوریتم پیشنهادی، به دلیل در نظر گرفتن ویژگی‌های گره‌ها در شناسایی جوامع، عملکرد بهتری نسبت به روش‌های ساختاری با خلوص بسیار نزدیک به مقادیر روش‌های معنایی دارند.

نتایج PurQ کارایی ترکیبی و کلی الگوریتم‌ها را در مجموعه داده‌های مختلف به‌دست آوردیم. نتایج PurQ علاوه بر کارآیی، در θهای مساوی یک، نماینگر عدالت الگوریتم‌ها می‌باشند. الگوریتم‌های موضوع‌گرا و خوشه‌بندی معنایی PurQ مناسبی ندارند. نکته مهم، نتایج غیرقابل قبول روش‌های ترکیبی و معنایی حتی در θهای زیر یک است که اهمیت شباهت گره‌ها بالاتر است. نتایج بسیار خوب PurQ بیانگر آن است که روش پیشنهادی ساختار و معنا را به خوبی ترکیب می‌کند و نتایج خوبی را در ماجولاریتی با حفظ خلوص بر جای می‌گذارد. به‌طورکلی نتایج آزمایشات نشان می‌دهند که روش پیشنهادی در رابطه با انسجام شبکه و ویژگی‌های محتوا به طور موثر عمل می‌کند. از سوی دیگر، با توجه به استخراج قطعی پیوندهای توپولوژیکی و معنایی در شرایط زمانی پویا، سادگی و عمومیت به عنوان یک روش تشخیص جامعه، به نظر می‌رسد که در موارد مشابه از قبیل طراحی توصیه‌کننده‌ها، سیستم‌های تصمیم‌گیر پیش‌بینی لینک در شبکه‌های پیچیده نیز قابل استفاده باشد.

در این مقاله، برای فاز ۲ از الگوریتم بیشنه‌سازی ماجولاریتی برای کشف جوامع، از شبکه‌ی غنی‌شده در فاز نخست استفاده شده است. اگر چه عملکرد این الگوریتم قابل قبول است، مطالعه‌ی الگوریتم‌های کارآمدتر می‌تواند موضوع تحقیقات آینده باشد.

مراجع

[1] Dwivedi, Y.K., et al., *Setting the future of digital and social media marketing research: Perspectives and research propositions.* International Journal of Information Management, 2021. 59: p. 102168.

[2] Kong, X., et al., *Academic social networks: Modeling, analysis, mining and applications.* Journal of Network and Computer Applications, 2019. 132: p. 86-103.

[3] Curiskis, S.A., et al., *An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit.* Information Processing & Management, 2020. 57(2): p. 102034.

[4] Srivastava, D.K. and B. Roychoudhury, *Words are important: A textual content based identity resolution scheme across multiple online social networks.* Knowledge-Based Systems, 2020. 195: p. 105624.

[5] Akachar, E., B. Ouhbi, and B. Frikh, *ACSIMCD: A 2-phase framework for detecting meaningful communities in dynamic social*

Transactions on Computational Social Systems, 2020. 7(2): p. 308-318.
[28] Li, C., et al., *Multiplex network community detection algorithm based on motif awareness.* Knowledge-Based Systems, 2023. 260: p. 110136.
[29] Costa, A.R. and C.G. Ralha, *AC2CD: An actor–critic architecture for community detection in dynamic social networks.* Knowledge-Based Systems, 2023. 261: p. 110202.
[30] Nasser Al-Andoli, M., S. Chiang Tan, and W. Ping Cheah, *Distributed parallel deep learning with a hybrid backpropagation-particle swarm optimization for community detection in large complex networks.* Information Sciences, 2022. 600: p. 94-117.
[31] Xie, Y., et al., *High-performance community detection in social networks using a deep transitive autoencoder.* Information Sciences, 2019. 493: p. 75-90.
[32] Rajendra Prasad, K., M. Mohammed, and R.M. Noorullah, *Correction to: Visual topic models for healthcare data clustering.* Evolutionary Intelligence, 2021. 14(2): p. 563-565.
[33] Blei, D., A. Ng, and M. Jordan, *Latent Dirichlet Allocation*. Vol. 3. 2001. 601-608.
[34] Hastings, P., *Latent Semantic Analysis*. 2004.
[35] Yu, X., J. Yang, and Z.-Q. Xie, *A semantic overlapping community detection algorithm based on field sampling.* Expert Syst. Appl., 2015. 42(1): p. 366–375.
[36] Abdelbary, H.A. and A. El-Korany, *Semantic Topics Modeling Approach for Community Detection.* social networks, 2013. 81.
[37] Xin, Y., et al., *An overlapping semantic community detection algorithm base on the ARTs multiple sampling models.* Expert Systems with Applications, 2015. 42(7): p. 3420-3432.
[38] Dai, T., et al., *Explore semantic topics and author communities for citation recommendation in bipartite bibliographic network.* Journal of Ambient Intelligence and Humanized Computing, 2018. 9(4): p. 957-975.
[39] Han, X., D. Chen, and H. Yang, *A Semantic Community Detection Algorithm Based on Quantizing Progress.* Complexity, 2019. 2019: p. 1-13.
[40] Zhao, F., et al., *A personalized hashtag recommendation approach using LDA-based topic model in microblog environment.* Future Generation Computer Systems, 2016. 65: p. 196-206.
[41] Harakawa, R. and M. Iwahashi, *Ranking of Importance Measures of Tweet Communities: Application to Keyword Extraction From COVID-19 Tweets in Japan.* IEEE Transactions on Computational Social Systems, 2021. 8(4): p. 1030-1041.
[42] Chen, H., et al., *Community detection in subspace of attribute.* Information Sciences, 2022. 602: p. 220-235.
[43] Reihanian, A., B. Minaei-Bidgoli, and H. Alizadeh, *Topic-oriented community detection of rating-based social networks.* Journal of King Saud University - Computer and Information Sciences, 2016. 28(3): p. 303-310.
[44] Arefian, Z. and M.R.K. Bashi, *Scalable Community Detection through Content and Link Analysis in Social Networks.* Journal of Information Systems and Telecommunication (JIST), 2015. 4(12): p. 1.
[45] Luo, S., et al., *Co-association matrix-based multi-layer fusion for community detection in attributed networks.* Entropy, 2019. 21(1): p. 95.
[46] Reihanian, A., M.-R. Feizi-Derakhshi, and H.S. Aghdasi, *An enhanced multi-objective biogeography-based optimization for overlapping community detection in social networks with node attributes.* Information Sciences, 2023. 622: p. 903-929.
[47] Pourabbasi, E., et al., *A new single-chromosome evolutionary algorithm for community detection in complex networks by combining content and structural information.* Expert Systems with Applications, 2021. 186: p. 115854.
[48] Wu, Y., et al., *Heterogeneous question answering community detection based on graph neural network.* Information Sciences, 2023. 621: p. 652-671.
[49] He, C., et al., *Semi-supervised overlapping community detection in attributed graph with graph convolutional autoencoder.* Information Sciences, 2022. 608: p. 1464-1479.
[50] Sadri, Y., et al., *An Extended Louvain Method for Community Detection in Attributed Social Networks.*
[51] Sadri, Y., et al., *Handling topic dependencies alongside topology interactions using fuzzy inferences for discovering communities in social networks.* Expert Systems with Applications, 2022. 208: p. 118188.
networks. Future Generation Computer Systems, 2021. 125: p. 399-420.
[6] Chiu, C.-M., M.-H. Hsu, and E.T.G. Wang, *Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories.* Decision Support Systems, 2006. 42(3): p. 1872-1888.
[7] Wang, S., et al., *Preventing epidemic spreading in networks by community detection and memetic algorithm.* Applied Soft Computing, 2020. 89: p. 106118.
[8] Zhao, Z., et al., *Topic oriented community detection through social objects and link analysis in social networks.* Knowledge-Based Systems, 2012. 26: p. 164-173.
[9] Viviani, M., et al., *Assessing vulnerability to psychological distress during the COVID-19 pandemic through the analysis of microblogging content.* Future Gener Comput Syst, 2021. 125: p. 446-459.
[10] Chaudhary, L. and B. Singh, *Community detection using unsupervised machine learning techniques on COVID-19 dataset.* Social Network Analysis and Mining, 2021. 11(1): p. 28.
[11] Kanavos, A., et al., *Emotional community detection in social networks.* Computers & Electrical Engineering, 2018. 65: p. 449-460.
[12] Wang, W., et al. *Harnessing twitter" big data" for automatic emotion identification*. in *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*. 2012. IEEE.
[13] Qin, M., et al. *Adaptive community detection incorporating topology and content in social networks*. in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. 2017.
[14] Wang, X., et al., *Semantic Community Identification in Large Attribute Networks.* Proceedings of the AAAI Conference on Artificial Intelligence, 2016. 30(1).
[15] Pathak, N., et al., *Social topic models for community extraction.* 2008.
[16] Mahfoudh, A., H. Zardi, and M.A. Haddar, *Detection of dynamic and overlapping communities in social networks.* Int. J. Appl. Eng. Res., 2018. 13(11): p. 9109-9122.
[17] Mittal, S., D. Sengupta, and T. Chakraborty, *Hide and Seek: Outwitting Community Detection Algorithms.* IEEE Transactions on Computational Social Systems, 2021. 8(4): p. 799-808.
[18] Liu, C. and X. Lu, *Analyzing hidden populations online: topic, emotion, and social network of HIV-related users in the largest Chinese online community.* BMC Medical Informatics and Decision Making, 2018. 18(1): p. 2.
[19] Kernighan, B.W. and S. Lin, *An efficient heuristic procedure for partitioning graphs.* The Bell System Technical Journal, 1970. 49(2): p. 291-307.
[20] Al Andoli, M.a.C., Wooi Ping and Tan, Shing Chiang, *Deep learning-based community detection in complex networks with network partitioning and reduction of trainable parameters. Journal of Ambient Intelligence and Humanized Computing (2020). pp. 1-19. ISSN 1868-5137, 1868-5145.* Journal of Ambient Intelligence and Humanized Computing, 2020(1-19).
[21] White, S. and P. Smyth, *A Spectral Clustering Approach To Finding Communities in Graphs*, in *Proceedings of the 2005 SIAM International Conference on Data Mining (SDM)*. 2005, Society for Industrial and Applied Mathematics. p. 274-285.
[22] Clauset, A., M. Newman, and C. Moore, *Finding community structure in very large networks.* Physical review. E, Statistical, nonlinear, and soft matter physics, 2005. 70: p. 066111.
[23] Blondel, V., et al., *Fast Unfolding of Communities in Large Networks.* Journal of Statistical Mechanics Theory and Experiment, 2008. 2008.
[24] Girvan, M. and M. Newman, *Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. Proc. Natl Acad. Sci. USA 99, 7821-7826.* Proceedings of the National Academy of Sciences of the United States of America, 2002. 99: p. 7821-6.
[25] Newman, M. and M. Girvan, *Finding and Evaluating Community Structure in Networks.* Physical review. E, Statistical, nonlinear, and soft matter physics, 2004. 69: p. 026113.
[26] Zhang, H., et al., *Limited Random Walk Algorithm for Big Graph Data Clustering.* Journal of Big Data, 2016. 3.
[27] Seifikar, M., S. Farzi, and M. Barati, *C-Blondel: An Efficient Louvain-Based Dynamic Community Detection Algorithm.* IEEE

[52] Jaccard, P., *The distribution of the flora in the alpine zone. 1.* New phytologist, 1912. 11(2): p. 37-50.
[53] MATLAB. The MathWorks, Inc.,Natick, Massachusetts, United States.: Natick, Massachusetts, United States.
[54] Nepusz, T., et al., *Fuzzy communities and the concept of bridgeness in complex networks.* Physical review. E, Statistical, nonlinear, and soft matter physics, 2008. 77: p. 016107.
[55] Devi, J.C. and E. Poovammal, *An analysis of overlapping community detection algorithms in social networks.* Procedia Computer Science, 2016. 89: p. 349-358.
[56] Largeron, C., et al., *Generating Attributed Networks with Communities.* PLOS ONE, 2015. 10: p. e0122777.
[57] Krebs, V., *Political books network*. 2004.
[58] Wang, D., P. Cui, and W. Zhu, *Structural Deep Network Embedding*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, Association for Computing Machinery: San Francisco, California, USA. p. 1225–1234.
[59] Adamic, L., *The Political Blogosphere and the 2004 U.S. Election: Divided They Blog.* Proceedings of the 3rd International Workshop on Link Discovery, 2005.
[60] Xia, Z. and Z. Bu, *Community detection based on a semantic network.* Knowledge-Based Systems, 2012. 26(Complete): p. 30-39.

پاورقی‌ها:

[1] COVID-19
[2] Deep transitive autoencoding
[3] Latent Dirichlet Allocation (LDA)
[4]Latent semantic analysis (LSA)
[5]Generative model
[6]Link-field-topic (LFT)
[7] Particle swarm optimization
[8] Single-chromosome evolutionary algorithm
[9] Process
[10]Post-Process
[11] Topic Detection
[12] Sum-Product
[13] Jaccard Similarity
[14] MATLAB
[15] Intel Core i3 2.20 GHz
[16] DDR۴
[17] Modularity
[18] Purity
[19] F-measure
[20] k-dimensional multivariate normal distribution
[21] Benchmark
[22] ground truth
[23] Atlantic Coast
[24] Big East
[25] Big Ten
[26] Big Twelve
[27] Conference USA
[28] Independents
[29] Mid-American
[30] Mountain West
[31] Pacific Ten
[32] Southeastern
[33] Sun Belt
[34] Western Athletic
[35] Cora Dataset
[36] Case-based Reasoning
[37] Rule Learning
[38] Political Books Data set
[39] Amazon.com
[40] Flickr (www.flickr.com)
[41] Featuring tag images
[42] Political Blogs Dataset
[43] Adj-Noun Data sets
[44] David Copperfield
[45] Fairness
[46] Spars Networks