یک الگوریتم مبتنی بر اتوماتای یادگیر سلولی جدید برای شناسایی جوامع همپوشان در شبکه‌های اجتماعی

مصطفی الیاسی[۱]، محمدرضا میبدی[۲]، علیرضا رضوانیان[۳]

[۱] آزمایشگاه محاسبات نرم، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)، m.elyasi@aut.ac.ir

[۲] آزمایشگاه محاسبات نرم، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)، mmeybodi@aut.ac.ir

[۳] آزمایشگاه محاسبات نرم، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)، a.rezvanian@aut.ac.ir

چکیده

در سال‌های اخیر، شبکه‌های اجتماعی آنلاین با ارائه خدمات و اهداف گوناگون طیف وسیعی از کاربران را به خود جذب کرده‌اند. با توجه به حجم انبوه کاربران و وجود گونه‌های اطلاعاتی مختلف در این شبکه‌ها، تحلیل شبکه‌های اجتماعی از اهمیت خاصی برخوردار است. یکی از ویژگی‌های بارز شبکه‌های اجتماعی وجود ساختارهای اجتماعی از روابط و فعالیت‌های کاربران است و کارهای مختلفی توسط محققان برای خوشه‌بندی و به طور اختصاصی‌تر شناسایی جوامع در شبکه‌های اجتماعی صورت گرفته است. در اکثر کارهای انجام شده به طور سنتی ساختار جوامع به طور قطعی تعریف شده است در صورتی که امکان همپوشانی در بین اعضای خوشه‌های شبکه در کاربردهای عملی وجود دارد. در این مقاله با هدف شناسایی جوامع همپوشان یک الگوریتم دو مرحله‌ای با استفاده از اتوماتای یادگیر سلولی پیشنهاد شده است که علاوه بر یافتن ساختارهای جوامع، اعضای همپوشان را نیز شناسایی کند تا به کاربردهای واقعی نزدیکتر باشد. در الگوریتم پیشنهادی برای اولین بار از معیار جدیدی برای محاسبه تعلق نودها به خوشه‌های مختلف استفاده شده است و همچنین در فاز پایانی با استفاده از بهینه کردن معیار پیمانگی سعی در رسیدن به بهترین ساختار را دارد. به منظور ارزیابی الگوریتم پیشنهادی از دادگان استاندارد شبکه‌های اجتماعی واقعی در شبیه‌سازی استفاده شده است و نتایج آزمایشات در مقایسه با سایر الگوریتم‌های معروف شناسایی اجتماعات حاکی از برتری الگوریتم پیشنهادی است.

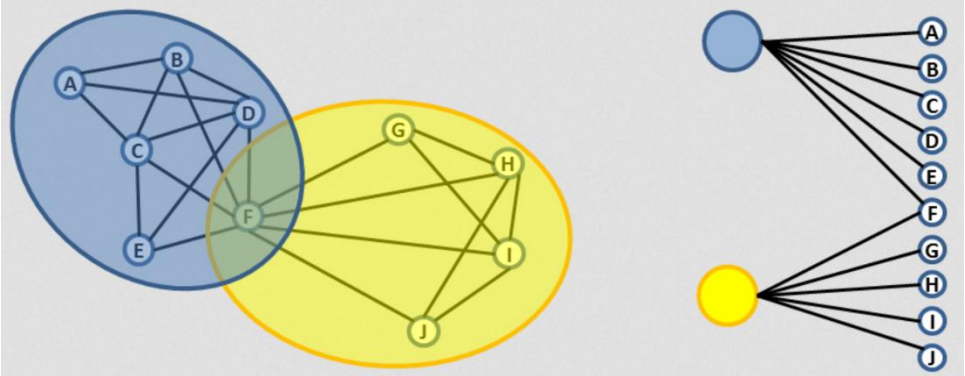
واژه‌های کلیدی

شناسایی اجتماع، جوامع همپوشان، شبکه‌های اجتماعی، ماتریس تعلق گراف، اتوماتای یادگیر سلولی

۱- مقدمه

در سال‌های اخیر، شبکه‌های اجتماعی آنلاین رشد وسیعی در بین رسانه‌های آنلاین داشته و کاربران بسیار زیادی را به خود جذب کرده است. طیف متنوعی از کاربران به طور روزانه فعالیت‌های فراوانی بر روی این شبکه‌ها دارند به طوریکه این شبکه‌ها به منبع اطلاعاتی بالایی از فعالیت‌های کاربران تبدیل شده است و بستر تحقیقاتی مناسبی را برای محققان به وجود آورده است. با توجه به رویکرد فعالیت‌های اجتماعی کاربران در شبکه‌های اجتماعی، ساختارهای اجتماعی از روابط و فعالیت‌های کاربران به وجود می آید که مطالعه ساختاهای اجتماعی کاربران و روابط از اهمیت بالایی در تحلیل شبکه‌های اجتماعی برخوردار است و باعث شده است به عنوان یکی از موضوعات بسیار مهم و مورد توجه در شبکه‌های اجتماعی شناسایی اجتماعات در این شبکه‌ها باشد. در واقع هدف، پیدا کردن گروهی از افراد است که دارای ویژگی‌های مشابهی از نظر فعالیت یا روابط هستند. به طور کلی، شناسایی اجتماع در شبکه‌های اجتماعی همانند خوشه‌بندی در گراف می‌باشد، به این شکل که ویژگی‌های هر گره ارتباطات آن گره با همسایه‌های آن گره می‌باشد. این مسئله از اینجا نشات می‌گیرد که کاربران با فعالیت‌های خود در شبکه توسط افراد تمایل دارند که گروه‌های نزدیک و مشابه ایجاد کنند. یا به عبارت دیگر افراد بیشتر تمایل دارند با اعضای با ویژگی‌های خاص و مشابهی ارتباط داشته باشند. شناسایی گروه‌های منسجم و ارتباطات آنها در شبکه‌های اجتماعی یک مسئله اساسی در تحلیل شبکه‌های اجتماعی است. مدل‌هایی نیز برای کار بر روی این شبکه‌ها مانند مدل پایه‌ی بیز[1][۱]، مدل ادغام

[1] Bayesian principle model



شکل ۱) مثالی از همپوشانی در جوامع یک شبکه [۱۳].

اجتماعات اوسلوم[۲][۲] و مدل ساختمان تصادفی پویا [۳] [۳] که از روش کالمن فیلترینگ استفاده کرده ارائه شده است، این روش که تعمیم یافته مدل سازی بلوکی است از روش‌های آماری می‌باشد و توصیف ساده‌تری از گراف را ارائه می‌دهد که گره‌ها را به گروه‌های هم ارزی تقسیم می‌کند، همچنین مدل‌هایی مانند مدل ترکیب پرتو[۴][۴] و مدل ساختمان تصادفی پویا برای بهبود مدل بیز ارائه شده است و برای شاخص تشخیص کیفیت نیز مدل مقایسه قسمت‌ها[۵] و معیار پیمانگی[۶][۵] ارائه شده است. در خوشه‌بندی یا همان تشخیص اجتماع در شبکه‌های اجتماعی، می‌توان اجتماعات را به شکل‌های مختلف مورد بررسی قرار داد. خوشه‌بندی هم پوشان[۷]، خوشه‌بندی شکننده و خوشه‌بندی فازی [۶] که هر کدام معیارهای سنجش کیفیت خود را دارند.

با توجه به آنکه در دنیای واقعی هر فرد یا گره امکان عضویت در چند گروه را دارد، نگاه جدیدتر و واقعی‌تری به این شبکه‌ها نیز ارائه شده است. به عبارتی هر گره دارای چند ویژگی می‌باشد که امکان عضویت آن را در چند گروه فراهم می‌سازد. به عنوان مثال یک فرد می‌تواند همزمان در دو دانشگاه تدریس کند. پس هم استادی در دانشگاه اول می‌باشد و هم عضو گروه اساتید دانشگاه دوم است. در شکل ۱، یک نمونه از این نوع خوشه‌بندی آورده شده است، این خوشه‌بندی با استفاده از روش شناسایی زیرگراف‌های کامل CPM [۷] و میزان همپوشانی این زیر گراف‌ها انجام شده است. در این روش زیرگراف‌های کامل به عنوان خوشه شناسایی شده و اگر دو زیرگراف کامل دارای گره‌های مشترک باشند، آن دو اجتماع، همپوشان هستند. از دیگر روش‌هایی که تا کنون برای شناسایی اجتماعات همپوشان ارائه شده‌اند، می‌توان از CONGA [۸] نام برد که از میزان مرکزیت یال و میزان مرکزی بودن گره استفاده می‌کند. مورد دیگر CONGO [۹] است که نسخه‌ی بهبود یافته‌ای از CONGA است. روش دیگری که توسط Gregory [۱۰] و همکارانش ارائه شد Copra نام دارد که در آن از انتشار برچسب استفاده می‌شود. پس از آن در سال ۲۰۱۱ Xie[۱۱] و همکاران روش SLPA را ارائه کردند که تعمیم دیگری از انتشار برچسب است. در انتشار برچسب برای شناسایی جوامع مجزا، هر گره صرفاً می‌تواند یک برچسب داشته باشد، اما در SLPA هر گره می‌تواند چند برچسب را نگه دارد. رده دیگر الگوریتم‌هایی که به شناسایی جوامع پرداخته‌اند الگوریتم‌های تکاملی هستند که به صورت تکرار شونده عمل می‌کنند و در هر تکرار به دنبال افزایش کیفیت جوامع کشف شده هستند، مانند الگوریتم GA-Net [۱۲]که از یک تابع شایستگی کارا برای کوچک کردن فضای مسئله و در نهایت پیدا کردن جوامع چگال[۸] استفاده می‌کند.

در این مقاله، با هدف شناسایی جوامع همپوشان یک الگوریتم دو مرحله‌ای با استفاده از اتوماتای یادگیر سلولی پیشنهاد شده است. در الگوریتم پیشنهادی در مرحله اول، از اتوماتای یادگیر برای شناسایی اجتماعات استفاده می‌شود و سپس مقادیر ماتریس تعلق به روز رسانی می‌گردد. در نهایت، الگوریتم پیشنهادی در فاز پایانی با استفاده از بهینه کردن معیار پیمانگی سعی در حذف اجتماعات شناسایی شده‌ی بسیار کوچک و ادغام اجتماعات بزرگتر برای رسیدن به بهترین ساختار را دارد.

در ادامه، در بخش دوم روش‌های شناسایی اجتماع که بصورت تکرار شونده کار می‌کنند معرفی شده است. در بخش سوم یک مرور کلی بر معیارهای سنجش در اجتماع‌های همپوشان ارائه شده است. در بخش چهارم روش پیشنهادی معرفی شده و چگونگی قابلیت‌های آن برای تلفیق با دیگر روش‌های شناسایی اجتماعات مجزا برای تبدیل آنها به الگوریتمی که قابلیت شناسایی اجتماعات همپوشان را نیز داشته باشد، بیان شده است. بخش پنجم به گزارش شبیه‌سازی و بررسی نتایج آزمایشات اختصاص یافته است. با توجه به آزمایشات بررسی نکات قوت و ضعف الگوریتم پرداخته شده است و نتایج آزمایشات در این بخش حاکی از برتری روش پیشنهادی است. در نهایت در بخش ششم این مقاله نتیجه گیری شده است.

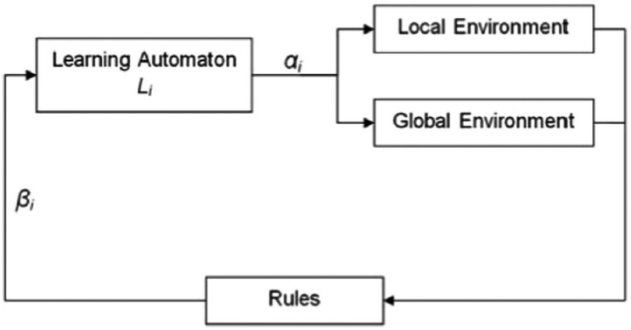
۲- الگوریتم‌های تکرارشونده[۹]

گروهی از الگوریتم‌ها برای رسیدن به جواب در فضای حالت جستجو کرده و به دنبال بهترین جواب می‌گردند. که از جمله‌ی این الگوریتم‌ها،

[2] OSLOM Merge Community
[3] Dynamic stochastic blockmodel(DSBM)
[4] Pareto Optimality
[5] Adjusted Rand Indices (ARI)
[6] Modularity
[7] Overlapping community detection

[8] Dense
[9] Iterative

الگوریتم‌های تکاملی [۱۴] و الگوریتم‌های یادگیری تقویتی[10] [۱۵] می‌باشند. بطور کلی الگوریتم‌هایی که در فضا جستجو می‌کنند و یا دارای قدرت یادگیری هستند به صورت تکرار شونده عمل می‌کنند. این الگوریتم‌ها به این شکل عمل می‌کنند که در یک مرحله یک ساختار از شبکه حدس زده می‌شود و به صورت تصادفی و یا براساس یک تابع شایستگی گره‌ها در اجتماع‌ها قرار می‌گیرند. سپس در مرحله‌ی بعدی سعی می‌شود تا با تغییر عضویت گره‌ها در گروه‌ها به جواب نزدیک شده و یا به ترکیب بهتری دست پیدا کند. این تغییر می‌تواند مانند شکل ۲ با یک بازخورد خارجی کنترل شود و به سمت جواب‌های بهتر حرکت کند. این روال ادامه پیدا می‌کند تا الگوریتم به سمت یک حالت بهینه همگرا شود. این الگوریتم‌ها معمولاً زمان بیشتری برای اجرا نیاز دارند، اما از نظر دقت معمولاً دارای عملکرد بهتری هستند.



شکل ۲) بازخورد از محیط در اتوماتای یادگیر سلولی [۱۵].

اتوماتای یادگیر سلولی یک مدل قدرتمند ریاضی است که از ترکیب اتوماتای سلولی و اتوماتای یادگیر بوجود آمده است، به عبارتی یک اتوماتای سلولی است که در هر خانه آن یک اتوماتای یادگیر قرار دارد. اتوماتای یادگیر سلولی d بعدی را می‌توان گفت که یک چندتایی است به طوریکه:

[۱۶] $CLA = (Z^d, \phi, A, N, F)$

Z^d یک شبکه از d تایی های مرتب از اعداد صحیح می باشد. این شبکه می تواند یک شبکه متناهی، نیمه متناهی یا متناهی باشد.

ϕ یک مجموعه متناهی از حالتها می باشد.

A، یک مجموعه از اتوماتاهای یادگیر (LA) است که هر یک از آنها به یک سلول از اتوماتای سلولی نسبت داده می شود.

$N = \{\overline{x}_1, ..., \overline{x}_{\overline{m}}\}$ یک زیر مجموعه متناهی از Z^d می باشد که بردار همسایگی نامیده می شود.

$F : \underline{\phi}^{\overline{m}} \rightarrow \underline{\beta}$ قانون محلی CLA می باشد به طوریکه $\underline{\beta}$ مجموعه مقادیری است که می تواند به عنوان سیگنال تقویتی پذیرفته شود.

۳- معیارهای ارزیابی

از روش‌های ارزیابی به طور مختصر می‌توان به NMI, Pareto ,ARI و میزان پیمانه‌ای بودن[11] [۵] اشاره کرد. میزان پیمانه‌ای بودن معروف‌ترین معیاری است که در سنجش کیفیت اجتماعات شناسایی شده در مجموعه داده‌های واقعی استفاده می‌شود. در این معیار که توسط Newman و همکارانش ارائه شد (مطابق معادله **(۱)**، هرچقدر میزان چگالی داخل خوشه‌ای نسبت به چگالی بین خوشه‌ای بیشتر باشد مقدار بیشتری را خواهد داشت. این معیار برای خوشه‌بندی مجزا استفاده می‌شود اما Gregory در [۱۷] تعمیمی از این الگوریتم (مطابق معادله **(۳)**) را ارائه داد که قابل استفاده در خوشه‌بندی همپوشان می‌باشد.

$$Q_d = \sum_{i,j \epsilon V} \left[\frac{A_{i,j}}{m} - \frac{P_{i,j}}{m}\right] \delta(c_i, c_j) \qquad (۱)$$

$$P_{i,j} = \frac{k_i^{out} k_j^{in}}{m} \qquad (۲)$$

که در آن V مجموعه راس‌ها، A ماتریس مجاورت، m تعداد یال‌های گراف، P احتمال وجود یال بین دو گره و k درجه هر گره است. همچنین δ درصورت هم خوشه بودن i,j دارای مقدار یک و درغیر اینصورت برابر مقدار صفر خواهد بود.

$$\boldsymbol{Q_{ov} = \frac{1}{m} \sum_{c \epsilon C} \sum_{i,j \epsilon V} [A_{i,j} \beta_{l(i,j),c} - P_{i,j} \beta_{l(i,j),c}{}^{in} \beta_{l(i,j),c}{}^{out}]} \quad (۳)$$

$$\beta_{l(i,j),c} = \frac{1}{(1 + e^{-f(\alpha_{i,c})})(1 + e^{-f(\alpha_{j,c})})} \qquad (۴)$$

که α میزان تعلق گره i به جامعه c را مشخص می‌کند.

۴- روش پیشنهادی

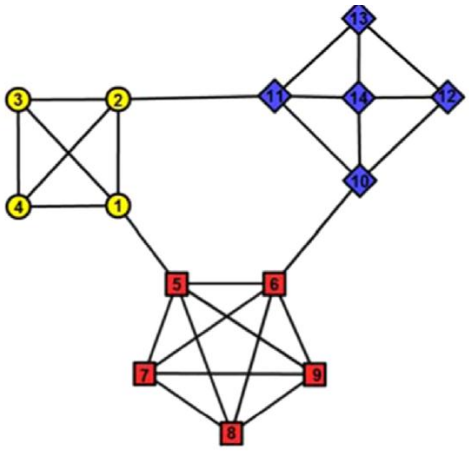
روش پیشنهادی بر مبنای میزان تعلق هر گره به هر جامعه عمل می‌کند. این روش در سه فاز عمل می‌کند، در فاز اول از یک الگوریتم خوشه‌بندی با دقت بالا که بصورت تکرار شونده و غیر همپوشان عمل می‌کند استفاده می‌کنیم. با استفاده از اتوماتای یادگیر سلولی و روشی که در [۱۵] ارایه شده است ابتدا جوامع را بصورت مجزا شناسایی می‌کنیم، در این روش با استفاده از اتوماتای یادگیر سلولی و الگوریتم ساده ی L_{RI} (خطی براساس فقط پاداش) بردار احتمالات اتوماتا به روز رسانی می‌شود. در این به روز رسانی از پارامتر لاندا برای پاداش و الگوریتم CP_{RP} برای محاسبه فعالیتی که استحقاق پاداش را دارد استفاده می‌شود. هر گره یک اتوماتا درنظر گرفته می‌شود و همسایه‌های گره بردار احتمال آن اتوماتا را تشکیل می‌دهند. در هر تکرار هر گره برچسب یکی از همسایه‌های خود را انتخاب کرده و در بردار راه حل ذخیره می‌کند، سپس با یک جستجوی اول سطح، اجتماعات شناسایی شده و در بردار عضویت (membership vector) ذخیره می‌گردند. اگر اجتماعات شناسایی شده باعث افزایش معیار Q_d شوند، بردارهای احتمال گره‌ها بروزرسانی می‌شوند. در شکل ۳ شیوه کار بر روی یک گراف ساده نشان داده شده است.

در فاز دوم که همزمان از خروجی فاز اول استفاده می‌کند به این صورت عمل می‌کنیم که در هر تکرار تعداد یال‌های هر گره در هر اجتماع را با مقدار قبلی آن جمع می‌کنیم. بطور مثال اگر گره i دارای سه یال باشد که دوتا از آنها به خوشه اول تعلق دارند و یکی از آنها به خوشه دوم تعلق دارد،

[10] Reinforcement learning
[11] Modularity

در ستون اول از سطر iام گراف عدد ۲ را به مقدار قبلی آن اضافه می‌کنیم و به ستون دوم از سطر iام عدد ۱ را اضافه می‌کنیم.

الف)



ب)

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Solution Vector	2	1	1	3	6	7	5	9	6	12	14	14	12	11

پ)

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Membership Vector	1	1	1	1	2	2	2	2	2	3	3	3	3	3

شکل ۳) شناسایی اجتماعات مجزا با استفاده از اتوماتای یادگیر [۱۵].
الف)گراف ساده. ب)بردار راه حل. پ) بردار عضویت.

در ابتدا تمام درایه‌های ماتریس صفر است ولی به مرور زمان و در هر تکرار مقداری به آن اضافه می‌شود. در نهایت پس از اتمام خوشه‌بندی و توقف الگوریتم سطرهای ماتریس را نرمال کرده و میزان تعلق هر گره به هر جامعه بدست می‌آید.اگر عدد بدست آمده از حد آستانه λ بیشتر بود گره i به جامعه c تعلق دارد.

$$\lambda = \frac{1}{ov+1} \qquad (۵)$$

در معادله (۵) پارامتر ov که توسط کاربر تنظیم می‌شود مشخص کننده‌ی حداکثر تعداد جامعه‌ای است که یک گره می‌تواند در آن عضو شود. ماتریس گراف شکل ۱ بعد از نرمال سازی در جدول ۱ آورده شده است.

شیوه شناسایی خوشه‌ها در تکرارهای مختلف به این صورت است که ماتریس ما در ابتدا فقط یک ستون دارد یعنی تمام گره‌ها متعلق به یک جامعه هستند ولی با شناسایی اجتماع‌های جدید، ستون‌های جدید به ماتریس اضافه می‌شود. در هر تکرار خوشه‌هایی که شناسایی شده‌اند را با ستون‌های ماتریس مقایسه کرده و اگر نیمی از گره‌های خوشه‌ی جدید با گره‌های موجود در یک ستون ماتریس مشترک بود یعنی این خوشه قبلا موجود بوده است و برابر با همان ستون از ماتریس است ولی درصورتی که با هیچکدام از ستون‌های ماتریس اشتراک (بیش از نیمی از گره‌ها) نداشت یک ستون به ماتریس اضافه می‌کنیم.

جدول ۱) ماتریس تعلق بدست آمده از گراف شکل ۱

گره	جامعه اول	جامعه دوم
A	۱	۰
B	۱	۰
C	۱	۰
D	۱	۰
E	۱	۰
F	۰.۵	۰.۵
G	۰	۱
H	۰	۱
I	۰	۱
J	۰	۱

$ov = 2\ ,\ \lambda = 0.33$

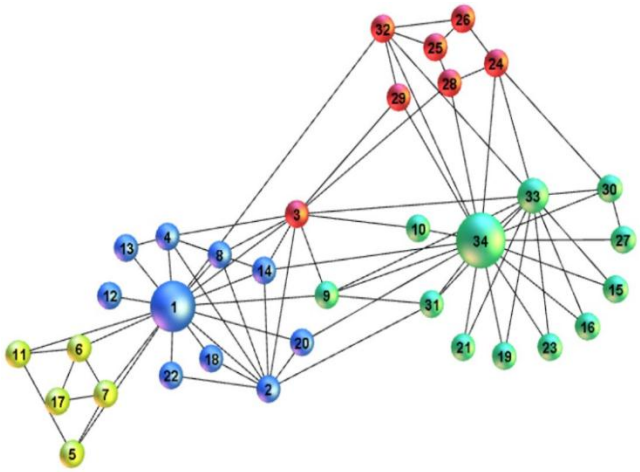
میزان تعلق هر گره به جوامع مختلف را با α نشان می‌دهیم.

$$0 \le \alpha_{i,c} \le 1 \ \forall_i \epsilon V\ , \forall_c \epsilon C \qquad (۶)$$

$$\sum_{c=1}^{|C|} \alpha_{i,c} = 1 \qquad (۷)$$

در شکل ۴ خروجی فاز اول الگوریتم پس از اجرا بر روی مجموعه داده باشگاه کاراته نشان داده شده است. همانطور که مشاهده می‌شود گراف بصورت مجزا خوشه‌بندی شده است. اما برخی گره‌های میانی مانند گره شماره ۳ دارای تعداد یال تقریبا برابر در دو یا چند اجتماع هستند، یعنی می‌توان گفت که به بیش از یک اجتماع تعلق دارند.

خروجی فاز دوم الگوریتم در شکل ۵ نشان داده شده است، هر گره می‌تواند به بیش از یک اجتماع تعلق داشته باشد و همانطور که مشاهده می‌شود اگر میزان تعلق یک گره به یک اجتماع از حد مشخصی بیشتر شود، ما آن گره را در آن اجتماع قرار می‌دهیم. که این میزان تعلق از ماتریس تعلق استخراج می‌شود.



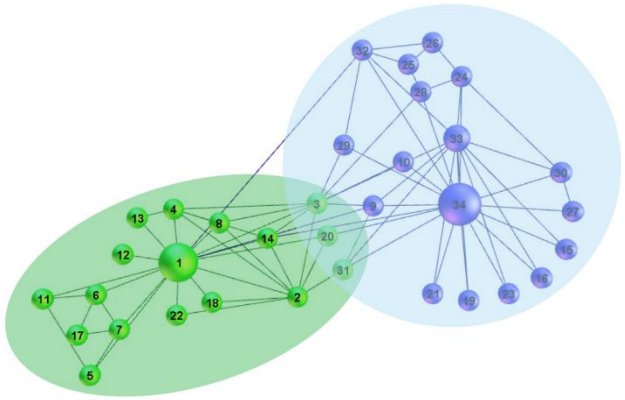
شکل ۴) اجرای فاز اول الگوریتم بر روی مجموعه داده باشگاه کاراته

در فاز سوم، از آنجا که در انتها ممکن است تعدادی جوامع بسیار کوچک تشکیل شده باشند، این الگوریتم درصورت افزایش معیار میزان پیمانه‌ای

جدول ۲) مقایسه نتایج آزمایشات انجام شده با معیار Q_{ov}.

نام شبکه	روش پیشنهادی		SLPA		COPRA		CONGA	Cfinder	LFM	GA-net
	max	std	max	std	max	std				
کاراته	**۰.۷۶**	۰.۰۵	**۰.۷۶**	۰.۲۱	۰.۶۲	۰.۱۸	۰.۵۵	۰.۵۲	۰.۴۲	۰.۴۰
دلفین	**۰.۷۹**	۰.۰۱	۰.۷۸	۰.۰۳	۰.۷۴	۰.۰۴	۰.۷۶	۰.۶۶	۰.۲۸	۰.۵۱
کتاب	**۰.۸۴**	۰.۰۱	۰.۸۳	۰.۰۱	۰.۸۲	۰.۰۵	۰.۸۱	۰.۷۹	۰.۷۴	۰.۵۰
فوتبال	**۰.۷۷**	۰.۰۱	۰.۷۱	۰.۰۱	۰.۷۲	۰.۰۳	۰.۶۴	۰.۶۴	–	۰.۶۰
جاز	۰.۶۶	۰.۰۱	**۰.۷۵**	۰.۰۹	۰.۷۳	۰.۰۵	۰.۵۸	۰.۵۵	–	۰.۴۰
شبکه علوم	**۰.۹۴**	۰.۰۱	۰.۸۶	۰.۰۱	۰.۸۴	۰.۰۲	۰.۹۱	۰.۶۱	۰.۴۶	۰.۵۵
ایمیل	**۰.۶۵**	۰.۰۱	**۰.۶۵**	۰.۰۳	۰.۵۱	۰.۲۲	۰.۶۳	۰.۴۶	۰.۲۵	۰.۳۹

بودن Q_{ov} که در قسمت قبل توضیح داده شد جوامع کوچک را با جوامع بزرگتر ترکیب می‌کند.



شکل ۵) اجرای فاز دوم الگوریتم و شناسایی اجتماعات همپوشان

۵- آزمایشات

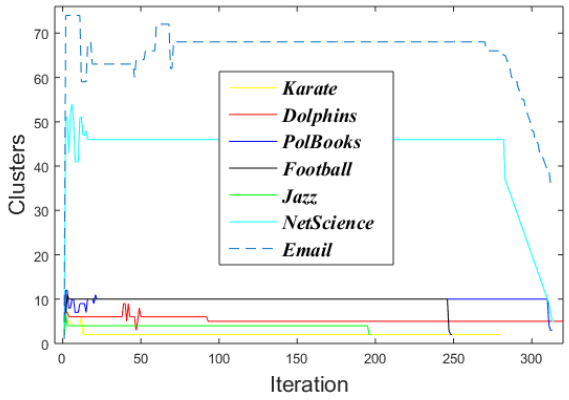
در اینجا آزمایشاتی را بر روی مجموعه داده‌های واقعی انجام داده‌ایم. برای آزمایش از هفت شبکه معروف از مجموعه داده‌های واقعی که مشخصات آن‌ها را در جدول ۳ آورده‌ایم، استفاده کرده‌ایم. و نتایج آن در جدول ۲ آورده شده است. همانطور که مشاهده می‌شود در بیشتر مجموعه داده‌ها این روش بهتر از دیگر الگوریتم‌های مقایسه شده عمل کرده است.

جدول ۳) مشخصات مجموعه داده‌ها[۱۸]

نام شبکه	تعدادگره	تعداد یال	میانگین درجه گره	توضیحات
کاراته	۳۴	۷۸	۴.۵	باشگاه کاراته زاخاری
دلفین	۶۲	۱۶۰	۵.۱	دلفین‌های لوسیوس
کتاب	۱۰۵	۴۴۱	۸.۴	مجموعه کتابهای سیاسی
فوتبال	۱۱۵	۶۱۳	۱۰.۶	باشگاه فوتبال آمریکایی
جاز	۱۹۸	۲۷۴۲	۲۷.۷	شبکه نوازنده‌های جاز
شبکه علوم	۳۷۹	۱۱۳۶	۴.۸	شبکه همکاری دانشمندان
ایمیل	۱۱۳۳	۱۰۹۰۲	۹.۶	شبکه ایمیل‌های آرنا

در همه‌ی الگوریتم‌ها بهترین نتایج با تنظیم دقیق پارامترها ثبت گردیده است. و این درحالی است که انحراف از معیار الگوریتم پیشنهادی از همه کمتر بوده و مقدار آن کمتر از ۰/۰۱ می‌باشد، اما در بقیه الگوریتم‌ها وابسته به نوع مجموعه داده، انحراف از معیاری بین ۰/۰۱ تا ۰/۲ را داشته‌ایم.

در شکل ۶ نمودار تغییرات تعداد اجتماعات شناسایی شده در طول روند اجرای الگوریتم نشان داده شده است. همانطور که مشاهده می‌شود بعد از چند تکرار اولیه تعداد اجتماعات شناسایی می‌شود، و درنهایت در فاز نهایی اجتماعات کوچک در اجتماعات بزرگ ادغام می‌شوند.



شکل ۶) نمودار تغییرات تعداد اجتماعات در طول اجرای برنامه

عملکرد الگوریتم SLPA به شدت وابسته به تنظیم دقیق مقدار r می‌باشد، در الگوریتم COPRA باید مقدار V (حداکثر تعداد جامعه‌ای که یک گره می‌تواند در آن عضو باشد) تنظیم شود، در CONGA باید تعداد جوامع را مشخص کرد و در Cfinder باید اندازه زیرگراف‌های کامل را بدانیم. ولی در روش پیشنهادی تنها یک پارامتر قابل تنظیم λ را داریم که با آزمایشات متعدد انجام شده ،مقدار مناسب برای آن ۰/۳۶ بدست آمده است که می‌توان مقدار آن را ثابت در نظر گرفت. در تمامی آزمایشات آورده شده در جدول ۳ نیز ما λ را همان مقدار ثابت ۰/۳۶ در نظر گرفتیم. پس میبینیم که الگوریتم پیشنهادی نیاز به تنظیم هیچ پارامتری ندارد.

۶- نتیجه گیری و پیشنهادات

این روش می‌تواند به هر الگوریتم شناسایی جوامع مجزا اضافه شده و در خروجی علاوه بر داشتن جوامع مجزا، جوامع همپوشان را نیز شناسایی کند. البته نیاز است تا شیوه کار الگوریتم پایه بصورت تکرار شونده باشد. دیگر مزیت این الگوریتم این است که در بسیاری از الگوریتم‌های شناسایی جوامع همپوشان نیاز به تنظیم پارامترهای زیادی است. مثلا در برخی باید تعداد جوامع را از قبل مشخص کرد، که در این روش نیازی به مشخص کردن تعداد جوامع از قبل نیست. اما اگر ما تعداد جوامع را از قبل داشته باشیم ماتریس ساختار ثابت خواهد داشت و دیگر نیازی به انجام فاز نهایی الگوریتم نیز نخواهد بود.

برای کارهای آتی پیشنهاد می‌شود تا راه حلی برای ترکیب این روش با دیگر الگوریتم‌های شناسایی جوامع که غیر تکرار شونده هستند ارایه شود. و همچنین راه حلی سریعتر برای جایگزینی با فاز نهایی می‌توان ارایه کرد. درضمن انجام آزمایشات بیشتر با محاسبه دیگر معیارهای ارزیابی و انجام آزمایشاتی برروی مجموعه داده‌های مصنوعی قدرت و یا ضعف این الگوریتم را بیشتر نمایان خواهد کرد.

مراجع

[1] K. Deb, "Multi-Objective Optimization," in *Search methodologies*, Springer, 2014, pp. 403–449.

[2] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato, "Finding Statistically Significant Communities in Networks," *PloS One*, vol. 6, no. 4, p. e18961, 2011.

[3] K. S. Xu and A. O. H. Iii, "Dynamic Stochastic Blockmodels: Statistical Models for Time-Evolving Networks," in *Social Computing, Behavioral-Cultural Modeling and Prediction*, A. M. Greenberg, W. G. Kennedy, and N. D. Bos, Eds. Springer Berlin Heidelberg, 2013, pp. 201–210.

[4] K.-J. Hsiao, K. Xu, J. Calder, and A. O. Hero, "Multi-Criteria Anomaly Detection Using Pareto Depth Analysis," in *Advances in Neural Information Processing Systems*, 2012, pp. 845–853.

[5] M. E. Newman and M. Girvan, "Finding and Evaluating Community Structure in Networks," *Phys. Rev. E*, vol. 69, no. 2, p. 026113, 2004.

[6] H. Zanghi, S. Volant, and C. Ambroise, "Clustering Based on Random Graph Model Embedding Vertex Features," *Pattern Recognit. Lett.*, vol. 31, no. 9, pp. 830–836, 2010.

[7] S. Maity and S. K. Rath, "Extended Clique percolation method to detect overlapping community structure," in *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI*, 2014, pp. 31–37.

[8] S. Gregory, "An Algorithm to Find Overlapping Community Structure in Networks," in *Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases*, Berlin, Heidelberg, 2007, pp. 91–102.

[9] S. Gregory, "A Fast Algorithm to Find Overlapping Communities in Networks," in *Machine Learning and Knowledge Discovery in Databases*, W. Daelemans, B. Goethals, and K. Morik, Eds. Springer Berlin Heidelberg, 2008, pp. 408–423.

[10] S. Gregory, "Finding overlapping communities in networks by label propagation," *New J. Phys.*, vol. 12, no. 10, p. 103018, 2010.

[11] J. Xie, B. K. Szymanski, and X. Liu, "SLPA: Uncovering Overlapping Communities in Social Networks via a Speaker-Listener Interaction Dynamic Process," in *2011 IEEE 11th International Conference on Data Mining Workshops*, 2011, pp. 344–349.

[12] C. Pizzuti, "GA-Net: A Genetic Algorithm for Community Detection in Social Networks," in *Parallel Problem Solving from Nature – PPSN X*, G. Rudolph, T. Jansen, N. Beume, S. Lucas, and C. Poloni, Eds. Springer Berlin Heidelberg, 2008, pp. 1081–1090.

[13] R. Ho, "Pragmatic Programming Techniques: Detecting Communities in Social Graph," *Pragmatic Programming Techniques*, 26-Nov-2012. .

[14] Y. Atay and H. Kodaz, "A New Adaptive Genetic Algorithm for Community Structure Detection," in *Intelligent and Evolutionary Systems*, K. Lavangnananda, S. Phon-Amnuaisuk, W. Engchuan, and J. H. Chan, Eds. Springer International Publishing, 2016, pp. 43–55.

[15] Y. Zhao, W. Jiang, S. Li, Y. Ma, G. Su, and X. Lin, "A cellular learning automata based algorithm for detecting community structure in complex networks," *Neurocomputing*, vol. 151, Part 3, pp. 1216–1226, Mar. 2015.

[16] H. Beigy and M. R. Meybodi, "Open synchronous cellular learning automata," *Adv. Complex Syst.*, vol. 10, no. 04, pp. 527–556, 2007.

[17] V. Nicosia, G. Mangioni, V. Carchiolo, and M. Malgeri, "Extending the definition of modularity to directed graphs with overlapping communities," *J. Stat. Mech. Theory Exp.*, vol. 2009, no. 03, p. P03024, 2009.

[18] "Jian Liu's Homepage." [Online]. Available: http://dsec.pku.edu.cn/~jliu/. [Accessed: 06-Aug-2016].