

یک الگوریتم جدید برای شناسایی جوامع در شبکه‌های اجتماعی با استفاده از اتوماتای یادگیر

مانا هیبت‌اله پورده میرانی^۱، محمدرضا میبیدی^۲، علیرضا رضوانیان^۳

^۱آزمایشگاه محاسبات نرم، دانشکده کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران mana_mirani@aut.ac.ir
^۲آزمایشگاه محاسبات نرم، دانشکده کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران mmeybodi@aut.ac.ir
^۳آزمایشگاه محاسبات نرم، دانشکده کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران a.rezvanian@aut.ac.ir

چکیده

با توجه به ساختار روابط اجتماعی میان کاربران و وجود ساختارهای اجتماعی در شبکه‌های اجتماعی، یکی از مسائل مهم در این شبکه‌ها، شناسایی جوامع به منظور تحلیل ساختاری شبکه اجتماعی است. در مسئله شناسایی جوامع، هدف تقسیم‌بندی شبکه به مجموعه‌ای از زیر گراف‌هاست به گونه‌ای که تراکم ارتباطات داخل جوامع خیلی بالا و ارتباط مابین جوامع خیلی پایین باشد. در این مقاله با استفاده از یک الگوریتم مبتنی بر اتوماتای یادگیر سلولی نامنظم سعی بر شناسایی جوامع در شبکه‌های اجتماعی شده است. در الگوریتم پیشنهادی هر گره از گراف به یک اتوماتان یادگیر مجهز شده و براساس روابط همسایگی گره‌های گراف و روابط کل گره‌های گراف با یکدیگر براساس معیاری جدیدی به نام سیگنیفیکنس، فرآیند پاداش و جریمه به اقدام‌های اتوماتاهای یادگیر اعمال می‌شود. در واقع اساس کار این مقاله بهینه‌سازی معیار عمومی سیگنیفیکنس در کنار یک معیار محلی با استفاده از اتوماتای یادگیر سلولی است تا بتوان جوامع را به درستی تشخیص داد. به منظور ارزیابی کارایی الگوریتم پیشنهادی آزمایشاتی بر روی دادگان استاندارد شبکه‌های اجتماعی ترتیب داده شده‌است که نتایج آزمایشات در مقایسه با الگوریتم‌های متداول شناسایی جوامع حاکی از برتری روش پیشنهادی از نظر دقت و زمان اجرا است.

کلمات کلیدی: شبکه‌های اجتماعی، شناسایی جوامع، اتوماتای یادگیر سلولی، معیار سیگنیفیکنس، بهینه‌سازی.

۱ مقدمه

گسترش یافته و اهمیت استخراج جوامع در این شبکه‌ها به منظور تحلیل شبکه‌های اجتماعی از اهمیت بیشتری پیدا کرده است.

شناسایی جوامع علاوه بر تحلیل ساختاری شبکه‌های اجتماعی در مسائل دیگری همچون دسته‌بندی مشتریان، سیستم‌های پیشنهاد دهنده، برچسب‌زنی رؤس، تحلیل افراد تأثیرگذار شبکه و انتشار اطلاعات نیز کاربرد دارد. این مسئله به علت پیچیدگی، به رغم تلاش‌های مختلفی که در سال‌های اخیر در این زمینه انجام شده، هنوز به صورت کامل حل نشده و جواب رضایت بخشی برای حل این مسئله ارائه نشده است. الگوریتم‌های شناسایی جوامع بسیاری برای تحلیل شبکه‌های اجتماعی ارائه شده‌اند. شناسایی جوامع را می‌توان به صورت یک مسئله‌ی بهینه‌سازی در نظر گرفت، از این منظر روش‌های متعددی مبتنی بر بر بهینه‌سازی معیار مشهور ماژولاریتی برای شناسایی جوامع ارائه شده‌است [۱]. از جمله روش‌های معروف در این زمینه می‌توان به الگوریتم حریصانه‌ی نیومن (FN) [۲] اشاره نمود. در این روش از خوشه‌بندی سلسله مراتبی تجمعی استفاده شده است که گروه‌های متشکل از گره‌ها با موفقیت تشکیل جوامع بزرگتر را می‌دهند اگر و فقط اگر مقدار ماژولاریتی پس از این ادغام افزایش یابد. پس از آن کلاست و همکارانش روش CNM [۳] را ارائه دادند که با استفاده از یک ساختمان داده‌ی پیچیده توانستند با کاهش بار

شبکه اجتماعی، ساختاری اجتماعی است که توسط مجموعه‌ای از افراد (یا سازمان‌ها یا موجودیت‌های اجتماعی دیگر) شکل گرفته است. این مجموعه، به وسیله رابطه‌های اجتماعی مانند مبادله اطلاعات، همکاری، دوستی، خویشاوندی و یا تبادلات مالی مرتبط شده‌اند. واسرمن و همکارانش [۱۷] به عنوان پیش‌تازان تحلیل شبکه‌های اجتماعی برای شبکه‌های اجتماعی نقل کرده‌اند که «شبکه اجتماعی می‌تواند به عنوان سیستم رابطه‌ی اجتماعی که توسط مجموعه‌ای از بازیگران و ارتباطات اجتماعی آن‌ها توصیف می‌شود، تعریف شود». به طور متداول یک شبکه اجتماعی را می‌توان در قالب گراف نمایش داد که دارای مجموعه‌ای از رؤس (گره) و یال (اتصال) می‌باشد بطوریکه رؤس به عنوان بازیگران درون شبکه‌ای و یال‌ها به عنوان روابط بین این بازیگران به حساب می‌آیند. در ساده‌ترین شکل، یک شبکه اجتماعی نگاشتی از تمام یال‌های مربوطه، میان رؤس مورد مطالعه است. تحلیل شبکه‌های اجتماعی رویکردی برای مطالعه فعل و انفعالات تعاملات بشری است و برای بررسی الگوها، ساختار ارتباطی یا سازماندهی شبکه‌های اجتماعی استفاده می‌شود. با گسترش استفاده از روش‌های ارتباط الکترونیکی و آنلاین، شبکه‌های اجتماعی ایجاد شده

محاسباتی ماژولاریتی در الگوریتم نیومن، آن را برای شبکه‌های بزرگ قابل استفاده کنند. شانگ و همکارانش نیز یک الگوریتم ژنتیک بهبودیافته به نام MIGA برای بدست آوردن ماژولاریتی بیشینه ارائه دادند [۴].

فورتوناتو و همکارش در [۵] نشان دادند که گذشته از قابلیت‌های ماژولاریتی، این روش دارای محدودیت‌هایی نیز می‌باشد. یکی از مهم‌ترین محدودیت‌های ماژولاریتی در این زمینه، محدودیت تفکیک‌پذیری^۱ است. مسئله‌ی تفکیک‌پذیری تأثیر بسیار زیادی در مسائل کاربردی دنیای واقعی دارد. زیرا در این گونه مسائل، اجتماع‌ها اندازه‌های گوناگونی دارند و این مشکل باعث می‌شود بسیاری از اجتماعات کوچک، تشخیص داده نشوند. دو راه حل برای غلبه بر محدودیت تفکیک‌پذیری ارائه شده است: اول اینکه متریک‌های کیفیت دیگری در کنار ماژولاریتی مطرح شدند تا در مقیاس‌های مختلف جوامع شناسایی شوند. به‌طور مثال در مقاله‌ی [۶] از معیاری به نام سیگنیفیکنس به جای ماژولاریتی به عنوان تابع هدف در روش لووین استفاده شده است و به نتایج بهتری نسبت به استفاده از ماژولاریتی رسیده است. همچنین پیروزی معیاری به نام "رتبه‌ی جامعه" را معرفی نمود [۷] تا تراکم بالای ارتباطات داخل جوامع و تراکم کم ارتباطات مابین جوامع را تضمین کند. دومین راه حل این مشکل فرموله کردن شناسایی جوامع به عنوان یک مسئله‌ی بهینه‌سازی چند هدفه است. پیروزی در [۸] با استفاده از الگوریتم تکاملی NSGA-II و دو تابع هدف شایستگی جامعه و رتبه جامعه سعی در بدست آوردن جوامع مناسب را دارد. گانگ و همکارانش الگوریتم MOEA/D-Net را ارائه دادند [۹] که سعی در بهینه‌سازی دو تابع هدف در تقابل هم را دارد. در [۱۰] نیز روش CLANet ارائه شده است که با استفاده از یک آتوماتای یادگیر بیشینه‌سازی ماژولاریتی را در کنار یک محدودکننده‌ی محلی انجام داده است.

در این مقاله در ادامه در بخش دوم روش پیشنهادی معرفی شده است. بخش سوم به شبیه‌سازی و گزارش نتایج آزمایشات بر روی شبکه‌های ساختگی اختصاص یافته است و در نهایت در بخش نتیجه گیری این مقاله پایان یافته است.

۲ روش پیشنهادی

در این مقاله، بهینه‌سازی چند هدفه توسط آتوماتای یادگیر سلولی براساس یک معیار مناسب پیشنهاد شده است و روش پیشنهادی به نام SIG-CLA برای شناسایی جوامع سیگنیفیکنت در شبکه‌های اجتماعی نامگذاری شده است. در این روش از یک معیار عمومی به نام سیگنیفیکنس و معیار محلی رفوان [۱۱] برای یادگیری شبکه استفاده شده است. در ادامه در این قسمت، ابتدا ساختار آتوماتای یادگیر سلولی در نظر گرفته شده برای این مسئله شرح داده شده است. سپس معیار سیگنیفیکنس معرفی شده است و بعد از آن جزئیات الگوریتم SIG-CLA برای شناسایی جوامع شرح داده شده است.

۱.۲ ساختار آتوماتای یادگیر سلولی برای مسئله شناسایی جوامع

در این بخش یک روش آتوماتای یادگیر سلولی برای مسئله شناسایی جوامع معرفی می‌شود. ابتدا کل شبکه به عنوان یک آتوماتای یادگیر سلولی نامنظم^۲ مدل می‌شود و ساختار جامعه‌ی بهینه در طی تکامل (یادگیری) آتوماتای یادگیر سلولی شناسایی می‌شود. جهت ساختن این آتوماتای یادگیر سلولی نامنظم، اقدامات ذیل انجام می‌پذیرد. هر گره در شبکه به عنوان یک سلول از آتوماتای یادگیر سلولی در نظر گرفته می‌شود و یک آتوماتان یادگیر به هر سلول نسبت داده می‌شود. حالت^۳ (وضعیت) هر سلول به اقدام جاری منتخب آتوماتان بستگی دارد. ساختار یک آتوماتان یادگیر L_i مربوط به گره‌ی i به صورت سه‌تایی (ρ, β, α) نشان داده می‌شود [۱۳] به طوری که:

$\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in}) = N(i)$ مجموعه اقدامات مربوط به گره i است که $N(i)$ مجموعه گره‌های همسایه‌ی گره i در شبکه می‌باشد.

$\beta_i = \{0, 1\}$ مجموعه مقادیر پاسخ گرفته شده از محیط است که در آن صفر و یک به ترتیب متناظر با پاداش و جریمه می‌باشند.

$\rho_i = (\rho_{i1}, \rho_{i2}, \dots, \rho_{in})$ بردار احتمال اقدامات است که ρ_{ij} احتمال انتخاب اقدام α_{ij} برای آتوماتان یادگیر L_i است.

بردار راه حل با استفاده از وضعیت سلول‌ها در کل شبکه حاصل می‌شود. به عبارتی اقدامات جاری انتخاب شده توسط آتوماتای یادگیر بردار راه حل را تشکیل می‌دهد. بنابراین در تکرار t ام، بردار راه حل به صورت زیر است:

$$S(t) = (\alpha_1(t), \alpha_2(t), \dots, \alpha_n(t))$$

۲.۲ بازنمایی راه حل

ساختار جامعه را می‌توان به صورت بردار عضویت $C = (c_1, c_2, \dots, c_n)$ نشان داد. مشکل اصلی بازنمایی بردار عضویت این است که نیازمند اطلاعات اولیه مربوط به تعداد جوامع در شبکه می‌باشد. لذا در اینجا از بازنمایی همسایگی مبتنی بر مکان^۴ که در [۱۲] ارائه شده است، استفاده می‌شود. در این بازنمایی راه حل با استفاده از یک بردار راه حل $S = (s_1, s_2, \dots, s_n)$ نمایش داده می‌شود که s_i نشان می‌دهد که گره‌ی i و گره‌ی s_i هر دو متعلق به یک جامعه هستند. در واقع بردار راه حل S به جای جوامع، نمایانگر یال‌هاست. از آن‌جا که الگوریتم به دنبال برداری برای جوامع است، نیاز به یک پردازش برای تبدیل بردار راه حل به بردار عضویت است. برای این کار از جستجوی اول سطح یا جستجوی اول عمق که دارای پیچیدگی زمانی خطی هستند استفاده می‌شود.

² Irregular

³ State

⁴ Locus based adjacency

¹ Limit resolution

بعد از این پردازش، بردار راه حل به بردار عضویت برای بازنمایی جوامع تبدیل می‌شود. بنابراین این بازنمایی راه حل دو مزیت مهم دارد که باعث می‌شود در اتوماتای یادگیر از آن استفاده شود. اول اینکه با استفاده از این بازنمایی می‌توان ساختار جامعه‌ی شبکه را به صورت بردار عضویت داشت بدون اینکه لزومی به داشتن اطلاعات اولیه راجع به تعداد جوامع در شبکه باشد. این مسئله از جمله مشکلات یک سری از روش‌های شناسایی جوامع است. دومین مزیت آن خطی بودن این

راه حل است که باعث شده گزینه‌ی مناسبی برای انتخاب آن به عنوان بازنمایی راه حل در اتوماتای یادگیر باشد چرا که پیچیدگی زمانی راه حل باید بهینه باشد تا سرعت الگوریتم به صرفه باشد. در شکل ۱ به عنوان نمونه یک شبکه با ۱۴ گره نشان داده شده است که به سه جامعه بخش‌بندی شده است و برای آن یک بردار راه حل طبق بازنمایی راه حل همسایگی مبتنی بر مکان و بردار عضویت متناظر با آن آورده شده است.



شکل ۱: (الف) یک شبکه با سه جامعه که در آن جوامع با رنگ و شکل متفاوت متمایز شده‌اند. (ب) یک بردار عضویت در بازنمایی مکان مبتنی بر همسایگی برای ساختار جامعه این شبکه ارائه شده است. (پ) بردار عضویت مربوط به ساختار جامعه.

در ادامه معیار سیگنیفیکنس را معرفی می‌کنیم. هدف این مقاله بهینه‌سازی این معیار در کنار یک معیار محلی با استفاده از اتوماتای یادگیر سلولی به منظور یافتن یک ساختار جامعه مناسب است.

۳.۲ معیار سیگنیفیکنس

در روش‌های قبلی از جمله ماژولاریتی همواره پارتیشن‌ها ثابت در نظر گرفته شده‌اند و یال‌ها به صورت تصادفی توزیع شده‌اند و احتمال این که یک پارتیشن ثابت شامل حداقل E یال داخلی شود محاسبه می‌شود که مد نظر ما نیست. ما به دنبال آن هستیم که آیا یک بخش با حداقل E یال داخلی در این گراف تصادفی پیدا می‌شود یا نه که این رویه در اینجا در نظر گرفته شده است. لذا از آنجایی که شناسایی جوامع به دنبال بخش‌های خوب است باید بر روی احتمال پیدا کردن بخش‌های خوب در گراف تصادفی باشیم. احتمال اینکه یک زیرگراف با تراکم q و سایز n_c در گراف تصادفی با سایز n و تراکم p پیدا شود.

$$\Pr(S(n_c, q) \subseteq G(n, p)) = e^{\theta(-\binom{n_c}{2} D(q||p))} \quad (1)$$

که در آن $D(q||p)$ فاصله‌ی بین q و p است و به صورت زیر محاسبه می‌شود:

$$D(q||p) = q \log \frac{q}{p} + (1-q) \log \frac{1-q}{1-p} \quad (2)$$

رابطه‌ی (۱) برای یک بخش (خوشه) است. بنابراین با در نظر گرفتن تمامی بخش‌ها برای بخش‌بندی مورد نظر ضرب تک تک احتمال‌های هر بخش را محاسبه می‌کنیم و خواهیم داشت:

$$\Pr(\sigma) = \prod_c \exp(-\binom{n_c}{2} D(p_c || p)) \quad (3)$$

در نهایت سیگنیفیکنس $S(\sigma)$ ، منفی لگاریتم احتمال این بخش‌بندی است. که در آن p تراکم کل گراف و p_c تراکم زیرگراف c است. n_c نیز تعداد گره‌های زیرگراف c است.

$$S(\sigma) = -\log \Pr(\sigma) = \sum_c \binom{n_c}{2} D(p_c || p) \quad (4)$$

۴.۲ شناسایی جوامع مبتنی بر اتوماتای یادگیر

در اینجا اتوماتای یادگیر سلولی همزمان با^۵ (OSCLA) به کار گرفته شده است. اتوماتای یادگیر در کل شبکه به صورت هم‌زمان فعال می‌شود. هر اتوماتان یادگیر از هر دو محیط محلی و عمومی تاثیر می‌پذیرد. محیط محلی هر اتوماتان یادگیر شامل گره‌های همسایه‌ی آن در اتوماتای یادگیر می‌شود و محیط عمومی شامل تمامی اتوماتای یادگیر در شبکه است. بردار احتمال اقدام اتوماتان یادگیر بر اساس پاسخ دریافتی از محیط و با استفاده از الگوریتم CPRP به‌روز رسانی می‌شود که در ادامه توضیح داده می‌شود. تعاملات بین اتوماتان یادگیر و محیط در شکل ۲ آورده شده است.

$$W_i(t+1) = W_i(t) + (1 - \beta(t)) \quad (6)$$

$$Z_i(t+1) = Z_i(t) + 1$$

$$d_i(t+1) = \frac{W_i(t+1)}{Z_i(t+1)}$$

این رویه تا زمانی که ساختار جامعه‌ی بدست آمده در چندین تکرار متوالی ثابت باقی بماند تکرار می‌شود.

۳ شبیه‌سازی و نتایج آزمایشات

در این بخش، کارایی الگوریتم SIG-CLA با سایر روش‌های شناسایی جوامع مطرح از جمله [۳]CNM، [۸]MOGA، [۷]GA-net، [۱۶]MOCD، [۹]MOEA/D-net، [۱۵]Meme.net و [۱۰]CLA-net مورد مقایسه قرار گرفته است. قابل ذکر است که برای ارزیابی عادلانه تنظیمات پیش‌فرض و یا آنچه در مقاله‌ی نویسندگان ذکر شده بود اعمال شده است و آزمایشات برای حصول اطمینان از نتیجه و عدم اتفاقی بودن، ۳۰ مرتبه تکرار شده‌اند. در پیاده‌سازی SIG-CLA نرخ پاداش ۰,۰۰۵ در نظر گرفته شده است.

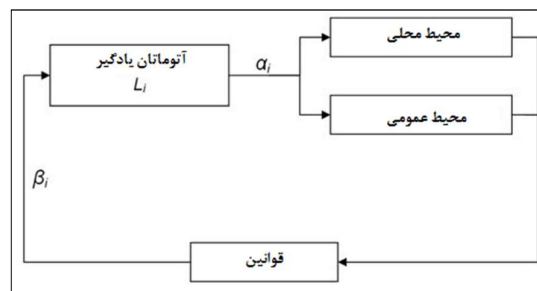
از آن‌جا که الگوریتم‌های مورد مقایسه در اینجا به منظور تشخیص جوامع، بهینه‌سازی را بر روی معیارهای متفاوتی نظیر ماژولاریتی، رتبه جامعه، چگالی و سیگنیفیکنس انجام می‌دهند، نمی‌توان هیچ‌یک از این معیارها را به عنوان معیار ارزیابی برای مقایسه‌ی این روش‌ها استفاده نمود. بنابراین برای مقایسه عملکرد آن‌ها به دادگان برچسب‌دار نیاز است و از آن‌جا که در شبکه‌های واقعی برچسب‌ها موجود نیستند، بنابراین برای مقایسه الگوریتم‌ها، از شبکه‌های مصنوعی LFR [۱۴] استفاده شده و با معیار NMI ارزیابی شده‌اند. در شبکه‌های مصنوعی LFR ویژگی‌های شبکه‌های واقعی از جمله توانی بودن توزیع درجه گره و اندازه جامعه لحاظ شده است و از این نظر ارزیابی با استفاده از این مجموعه دادگان برچسب‌دار مناسب است.

مشخصات شبکه‌های LFR استفاده شده جهت مقایسه در جدول ۱ ذکر شده است. وضوح ساختار جوامع در شبکه‌ی LFR به پارامتر μ بستگی دارد، که نشان دهنده میانگین کسری از اتصالات با دیگر جوامع به ازای هر گره است. به این معنی که هرچه پارامتر μ شبکه آزمایشی کمتر باشد ساختار جامعه‌ی واضح‌تری دارد. برای بررسی بهتر عملکرد روش‌های مورد مقایسه در این آزمایش مقدار پارامتر μ از ۰ تا ۰,۸ با فواصل ۰,۰۵ تغییر داده شده است. در شکل ۳ نتیجه خروجی الگوریتم‌ها ذکر شده است.

جدول ۱: مشخصات شبکه LFR استفاده شده.

اندازه شبکه	پارامترها (T_1 و T_2)	محدوده درجه گره	میانگین درجه گره	محدوده اندازه جامعه	مقدار پارامتر μ
۱۰۰۰	(۱ و ۲)	۰ - ۵۰	۲۰	۲۰ - ۵۰	۰ - ۰,۸

همان‌طور که از شکل ۳ مشخص است زمانی که $\mu < ۰,۱$ است تفاوت قابل توجهی میان الگوریتم‌ها دیده نمی‌شود ولی با افزایش



شکل ۲: تعاملات میان آتوماتای یادگیر باز و محیط

۵.۲ چارچوب الگوریتم SIG-CLA

روند الگوریتم SIG-CLA به این صورت می‌باشد که در هر تکرار، هر آتوماتان یادگیر براساس بردار احتمال اقدام متناظرش یک اقدام انتخاب می‌کند. اقدام‌های آتوماتای یادگیر در کل شبکه، بردار راه‌حل $S(t) = (\alpha_1(t), \alpha_2(t), \dots, \alpha_n(t))$ را تشکیل می‌دهد. برای نمایش ساختار جامعه‌ی شبکه، پردازش روی بردار راه حل انجام می‌شود و $S(t)$ به بردار عضویت $C(t) = (c_1(t), c_2(t), \dots, c_n(t))$ تبدیل می‌شود. در این جا محیط عمومی، سیگنیفیکنس $Sig(t)$ را برای ساختار جامعه‌ی جاری در شبکه محاسبه می‌کند. محیط محلی هر آتوماتان یادگیر جوامع گره‌های همسایه آن را در نظر می‌گیرد. برای آتوماتان یادگیر در هر گره تحت ارضای دو شرط زیر به صورت هم‌زمان، به آن آتوماتان پاداش داده می‌شود:

(۱) سیگنیفیکنس ساختار جامعه‌ی جاری از بهترین مقدار سیگنیفیکنس در تکرارهای قبل کمتر نباشد. یعنی برای دادن پاداش سیگنیفیکنس در نظر گرفته شود بطوری‌که ساختارهای با سیگنیفیکنس بالا شانس دریافت پاداش را داشته باشند.

(۲) گره مورد نظر شرط تعریف جامعه‌ی رقوان را رعایت کند، که مستلزم این است که در آن گره تعداد اتصالات داخلی آن از تعداد اتصالات آن گره به هر جامعه‌ی دیگر کمتر نباشد.

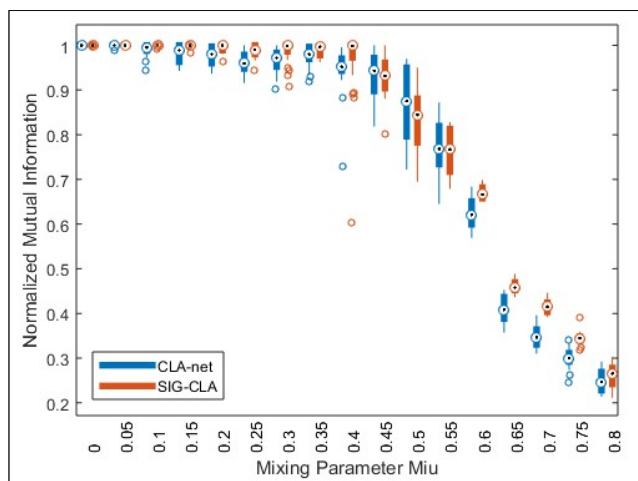
در غیر این صورت آتوماتان یادگیر مذکور از محیط پاسخ جرمه دریافت خواهد کرد. پس از آن الگوریتم CPRP انجام می‌شود. برای هر آتوماتان یادگیر، اقدام بهینه‌ی جاری طبق رابطه‌ی (۵) بدست می‌آید. در این رابطه e_m یک بردار واحد $[0 \dots 1 \dots 0]^T$ است که اقدام بهینه تخمین زده شده جاری را که تخمین پاداش بیشینه نامیده می‌شود، با یک نمایش می‌دهد و بقیه‌ی اقدام‌ها مقدار صفر را دارند.

$$p(t+1) = (1 - \lambda) P(t) + \lambda e_m \quad (5)$$

پس از آن بردار احتمالاتی اقدام‌ها براساس روابط (۶) به روزرسانی می‌شوند. که در این رابطه $Z_i(t)$ تعداد دفعاتی است که آتوماتان اقدام انتخاب شده است و $W_i(t)$ تعداد دفعاتی است که اقدام α_i پاداش گرفته است.

می‌آید. در مجموع با توجه به این منحنی‌ها می‌توان نتیجه گرفت که الگوریتم SIG-CLA در بیشتر مواقع بر روی شبکه‌های معیار عملکرد خوبی دارد.

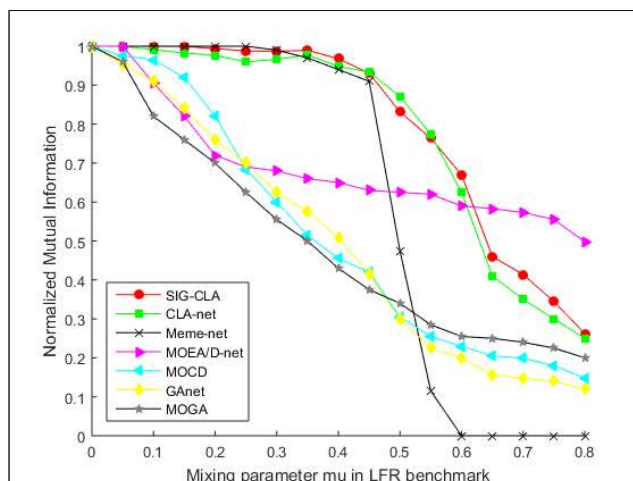
از آنجا که از بین روش‌های مورد مقایسه CLA-net عملکرد بهتری نسبت به سایرین دارد، در آزمایش بعدی، نمایش نمودار شمعی مقایسه‌ای روی توزیع نتایج این روش با الگوریتم پیشنهادی (SIG-CLA) در شکل ۴ نمایش داده شده است. در این نمودار حد بالا و پایین و میانگین نتایج به ازای هر دو روش نشان داده شده‌است. همانطور که از شکل نیز پیداست در هر دو روش پراکندگی جواب‌ها زمانی که پارامتر μ نزدیک به ۰.۵ است زیاد است. چرا که در این شرایط تعداد اتصالات بین گره‌های داخل جوامع و بیرون جوامع تقریباً برابر است و جوامع موجود شرط محلی دو الگوریتم را رعایت نمی‌کنند، لذا باعث شده که در تکرارهای مختلف آزمایش جواب‌های مختلفی به دست آید. به طور کلی عملکرد SIG-CLA بهتر بوده است چرا که پراکندگی جواب‌های آن به ازای هر آزمایش به نسبت کمتر از CLA-net بوده و به جز در حالت $\mu = 0.5$ در بقیه موارد برتری داشته یا یکسان عمل کرده‌است.



شکل ۴: مقایسه‌ی توزیع نتایج دو روش CLA-net و SIG-CLA.

داشته است و NMI بهتری را به دست می‌دهد. لذا می‌توان نتیجه گرفت که این روش به صورت معنی‌دار در تمامی حالات به جز یک مورد نسبت به روش CLA-net و همچنین سایر روش‌های مورد مقایسه برتری داشته است.

مقدار پارامتر μ ، ساختار شبکه پیچیده‌تر شده و کارایی تمامی روش‌ها به جز SIG-CLA، CLA-net، MIGA و Meme-net با شدت زیادی کاهش می‌یابد و این روش‌ها دیگر قادر به تشخیص درست اجتماعات نیستند. وقتی که $\mu < 0.4$ است الگوریتم SIG-CLA همواره بخش‌بندی مناسبی را می‌دهد و $NMI \approx 1$ را به دست می‌دهد. در این بازه روش‌های CLA-net، MIGA و Meme-net نیز نتایج قابل قبولی می‌دهند. به ازای $0.4 < \mu < 0.65$ روش SIG-CLA به وضوح نسبت به تمامی روش‌ها بهتر عمل کرده است (البته به جز در $\mu = 0.5$ که CLA-net نتیجه‌ی بهتری داشته است). برای $\mu < 0.65$ روش SIG-CLA از MOEA/D-net پیشی می‌گیرد ولی روش CLA-net همچنان از سایر روش‌ها بهتر عمل کرده است. لازم به ذکر است که عملکرد MOEA/D-net برای شبکه‌های با مقدار پارامتر μ پایین ($\mu < 0.65$) خوب عمل نکرده است. در صورتی که شبکه‌های اجتماعی واقعی دارای جوامع معنی‌دار هستند به این معنی که اندازه μ در آن‌ها کم است. زمانی که μ بیشتر از ۰.۵ باشد یک سری از جوامع موجود شرط رفوان را ارضا نمی‌کنند و لذا الگوریتم SIG-CLA پاسخ مناسبی از محیط محلی دریافت نمی‌کند در نتیجه کارایی این الگوریتم پایین



شکل ۵: مقایسه‌ی روش SIG-CLA با سایر روش‌ها بر روی شبکه‌های LFR.

همچنین جهت اطمینان از نتایج بدست آمده، از آزمون t^6 استفاده شده است. آزمون t در این ارزیابی با درجه آزادی^۷ ۵۸ و فاصله اطمینان ۹۵ درصد انجام شده است. که به ازای هر آزمایش مشخص می‌کند که آیا SIG-CLA بهتر، بدتر و یا مشابه روش CLA-net عمل کرده است. نتایج این آزمون در جدول ۲ آورده شده است. اختلاف معنی‌دار همراه با برتری با هاشور و تساوی به صورت بی‌رنگ نشان داده شده است. جدول ۲ نشان می‌دهد که روش پیشنهادی SIG-CLA در مقایسه با روش CLA-net در یک مورد بدتر ($\mu = 0.5$) (در ۵ مورد مشابه و در بقیه ۱۱ حالت نسبت به CLA-net برتری

⁶ T-test

⁷ Freedom degree

جدول ۲: مقایسه‌ی روش SIG-CLA با CLA-net از طریق آزمون t.

CLA-net	SIG-CLA	پارامتر μ	CLA-net	SIG-CLA	پارامتر μ
0.9337 ± 0.0488	0.9305 ± 0.0399	0.45	1 ± 0	1 ± 0.002	0
0.8711 ± 0.0779	0.8313 ± 0.0736	0.5	0.9997 ± 0.0019	1 ± 0	0.05
0.7739 ± 0.0589	0.7639 ± 0.0478	0.55	0.9910 ± 0.0122	0.9997 ± 0.0014	0.1
0.6250 ± 0.0343	0.6713 ± 0.0141	0.6	0.9824 ± 0.0194	0.9994 ± 0.0031	0.15
0.4079 ± 0.0261	0.4603 ± 0.0120	0.65	0.9766 ± 0.0215	0.9940 ± 0.0085	0.2
0.3494 ± 0.0219	0.4140 ± 0.0128	0.7	0.9598 ± 0.0214	0.9878 ± 0.0132	0.25
0.3002 ± 0.0195	0.3446 ± 0.0122	0.75	0.9660 ± 0.0252	0.9871 ± 0.0236	0.3
0.2493 ± 0.0241	0.2621 ± 0.0226	0.8	0.9776 ± 0.0212	0.9896 ± 0.0130	0.35
			0.9484 ± 0.0473	0.9681 ± 0.0768	0.4

- [5] Fortunato, Santo, and Marc Barthelemy. "Resolution limit in community detection." *Proceedings of the National Academy of Sciences* 104.1 (2007): 36-41.
- [6] Traag, Vincent A., Gautier Krings, and Paul Van Dooren. "Significant scales in community structure." *Scientific reports* 3 (2013).
- [7] Pizzuti, Clara. "Ga-net: A genetic algorithm for community detection in social networks." *Parallel Problem Solving from Nature-PPSN X*. Springer Berlin Heidelberg, 2008. 1081-1090.
- [8] Pizzuti, Clara. "A multiobjective genetic algorithm to find communities in complex networks." *Evolutionary Computation, IEEE Transactions on* 16.3 (2012): 418-430.
- [9] Gong, Maoguo, et al. "Community detection in networks by using multiobjective evolutionary algorithm with decomposition." *Physica A: Statistical Mechanics and its Applications* 391.15 (2012): 4050-4060.
- [10] Zhao, Yuxin, et al. "A cellular learning automata based algorithm for detecting community structure in complex networks." *Neurocomputing* 151 (2015): 1216-1226.
- [11] Raghavan, Usha Nandini, Réka Albert, and Soundar Kumara. "Near linear time algorithm to detect community structures in large-scale networks." *Physical Review E* 76.3 (2007): 036106.
- [12] Park, YoungJa, and ManSuk Song. "A genetic algorithm for clustering problems." *Proceedings of the third annual conference on genetic programming*. 1998.
- [13] Beigy, Hamid, and Mohammad Reza Meybodi. "A mathematical framework for cellular learning automata." *Advances in Complex Systems* 7.03n04 (2004): 295-319.
- [14] Lancichinetti, Andrea, Santo Fortunato, and Filippo Radicchi. "Benchmark graphs for testing community detection algorithms." *Physical review E* 78.4 (2008): 046110.
- [15] M. Gong, B. Fu, L. Jiao, memetic algorithm for community detection in networks, *Phys. Rev. E* 84(2011)056101.
- [16] C. Shi, Z. Y. Yan, Y. N. Cai, B. Wu, Multi-objective community detection in complex networks, *Appl. Soft Comput.* 12(2012)850-859.
- [17] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, pp. 75-174, 2010.

۴ نتیجه‌گیری و پیشنهادات

در این مقاله یک روش بهینه‌سازی چندهدفه مبتنی بر آتوماتای یادگیر سلولی ارائه شد. در این روش برای یادگیری از دو محیط عمومی و محلی استفاده شد. تعاملات با محیط عمومی تضمین می‌کند که الگوریتم SIG-CLA در جهت یافتن ساختار جامعه‌ای با سیگنیفیکنس بهینه پیش می‌رود. و تعاملات با محیط محلی ساختار جامعه‌ی بدست آمده را به تعریف جامعه‌ی رقوان محدود می‌کند. با استفاده از تعامل با این دو محیط متفاوت SIG-CLA می‌تواند به صورت بهینه به ساختار جامعه‌ی مناسب برسد و دچار مشکل محدودیت تفکیک‌پذیری نشود. از جمله چالش‌های موجود در شبکه‌های اجتماعی می‌توان به توجه محدود محققان به این مسئله که شبکه‌های دنیای واقعی اغلب دارای هم‌پوشانی هستند اشاره نمود چراکه در شبکه‌های واقعی یک فرد ممکن است هم‌زمان عضو جامعه دوستان، همکاران و خانواده باشد. در اینجا روش شناسایی جامعه برای جوامع مجزا از هم انجام شده است. لذا در آینده ما به دنبال راهکاری جهت حل این چالش خواهیم بود.

۵ مراجع

- [1] Newman, Mark EJ, and Michelle Girvan. "Finding and evaluating community structure in networks." *Physical review E* 69.2 (2004): 026113.
- [2] Newman, Mark EJ. "Fast algorithm for detecting community structure in networks." *Physical review E* 69.6 (2004): 066133.
- [3] Clauset, Aaron, Mark EJ Newman, and Cristopher Moore. "Finding community structure in very large networks." *Physical review E* 70.6 (2004): 066111.
- [4] Shang, Ronghua, et al. "Community detection based on modularity and an improved genetic algorithm." *Physica A: Statistical Mechanics and its Applications* 392.5 (2013): 1215-1231.