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. 1-16
Received: 06/09/2023, Revised: 08/14/2023, Accepted: 09/26/2023

Community Detection in Complex Networks Using Learning Automata

S. Mehdi Vahidipour[1*] and Zahra Salimiyan [2]
1*- Electrical and Computer Engineer Faculty, University of Kashan, Kashan, Iran
2- Electrical and Computer Engineer Faculty, University of Kashan, Kashan, Iran
[1]vahidipour@kashanu.ac.ir

Corresponding author's address: S. Mehdi Vahidipour, Computer Engineering Department, Faculty of Electrical and Computer Engineering, University of Kashan, Kashan, Iran.

Abstract- Social networks are one of the types of complex networks. Identifying communities in social networks is an effective way to use their information, for which several algorithms have been presented so far. In this paper, novel algorithms are designed, in which a learning automaton is attached to each node; The number of actions of learning automata is fixed and equal to the estimate of the number of network communities. At each step, each of the learning automata chooses an action from its set of actions. Choosing any of these actions means assigning the label of that community to the node. The action chosen by each automaton is evaluated based on the chosen actions of its neighbors ((local attention) and/or communities detected by the entire method (global screening). The result of the evaluation leads to generate rewards or punish signal for the automata. By receiving a reward, the probability of re-choosing the chosen action by the automaton, or the community label, increases, and otherwise, by receiving a fine, the probability of this action decreases. By repeating the algorithm, the optimal action is determined as long as no change occurs in the selected label of the corresponding automata of each node with more iterations, and as a result, the optimal communities are determined as the output of the algorithm. The comparison of the results of the experiments shows the effectiveness of the proposed methods in comparison with the previous methods.

Keywords- Social networks, community discovery, learning automata, similarity measure.

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 ۱-۱۶
دریافت: ۱۴۰۲/۰۳/۱۹، بازنگری: ۱۴۰۲/۰۵/۲۳، پذیرش: ۱۴۰۲/۰۷/۰۴

تشخیص جوامع در شبکه‌های پیچیده با استفاده از آتاماتای یادگیر

سید مهدی وحیدی پور*[۱] و زهرا سلیمیان [۲]
۱*- دانشکده مهندسی برق و کامپیوتر، دانشگاه کاشان، کاشان، ایران.
۲- دانشکده مهندسی برق و کامپیوتر، دانشگاه کاشان، کاشان، ایران.
[1]vahidipour@kashanu.ac.ir

* نشانی نویسنده مسئول: سید مهدی وحیدی‌پور، کاشان، بلوار قطب راوندی، دانشگاه کاشان، دانشکده مهندسی برق و کامپیوتر.

چکیده- شبکه‌های اجتماعی یکی از انواع شبکه‌های پیچیده است. تشخیص جوامع در شبکه‌های اجتماعی روشی مؤثر برای بهره‌گیری از اطلاعات این شبکه‌ها است که تاکنون الگوریتم‌های متعددی برای آن ارائه شده است. در این مقاله روش‌هایی جدید با استفاده از آتوماتاهای یادگیر پیشنهاد شده است که در آنها، یک آتوماتای یادگیر به هر گره شبکه الحاق می‌شود؛ تعداد اقدام آتوماتاهای یادگیر ثابت و برابر با تخمین تعداد جوامع شبکه است. در هر مرحله، هر کدام از آتوماتاهای یادگیر یک اقدام از مجموعه اقدامات خود را انتخاب می‌کند. انتخاب هر یک از این اقدام‌ها به‌منزله‌ی انتساب برچسب آن جامعه به گره است. اقدام انتخاب‌شده توسط هر آتوماتا بر اساس اقدام‌های انتخابی همسایگانش (بررسی محلی) و/یا جوامع تشخیص داده شده توسط کل روش (بررسی سراسری) ارزیابی می‌شود. نتیجه‌ی ارزیابی منجر به صدور پاداش و جریمه برای آتوماتاها می‌شود. با دریافت پاداش احتمال انتخاب مجدد اقدام انتخابی توسط آتوماتا، یا همان برچسب جامعه، افزایش می‌یابد و با دریافت جریمه احتمال این اقدام کاهش می‌یابد. با تکرار الگوریتم، اقدام بهینه مشخص می‌گردد تا آنجا که با تکرارهای بیشتر هیچ تغییری در برچسب انتخابی آتوماتای متناظر هر گره رخ نمی‌دهد و درنتیجه جوامع بهینه به‌عنوان خروجی الگوریتم مشخص می‌گردند. مقایسه نتایج حاصل از آزمایش‌های انجام‌شده، نشان می‌دهد روش‌های پیشنهادی نسبت به برخی روش‌های پیشین عملکرد بهتری را نشان می‌دهد؛ به خصوص بر اساس معیار NMI که یکی از معیارهای رایج در ارزیابی روش‌های تشخیص جامعه است.

واژه‌های کلیدی: شبکه‌های اجتماعی، کشف جوامع، آتوماتای یادگیر، معیار شباهت

۱- مقدمه

بسیاری از پدیده‌های دنیای واقعی توسط شبکه‌های پیچیده مدل‌سازی و تعریف شده‌اند [۱]. شبکه‌های اجتماعی، شبکه‌های اطلاعاتی و شبکه‌های بیولوژیکی از این قبیل شبکه‌های پیچیده هستند. شبکه‌های پیچیده را می‌توان به عنوان گراف با مجموعه‌ای از گره‌ها و اتصالات میان آن‌ها تعریف کرد. برای مثال در شبکه‌های اجتماعی افراد به عنوان گره‌ها و ارتباطات اجتماعی آن‌ها (مانند دوستی‌ها) به عنوان اتصالات میان گره‌ها تعریف می‌شوند [۲].

در این شبکه‌ها، جامعه به مجموعه‌ای از گره‌ها گفته می‌شود که تراکم اتصالات داخلی آن‌ها (گره‌های داخل جامعه) زیاد و تراکم اتصالات متقابل آن‌ها (اتصالات گره‌های یک جامعه با گره‌های جامعه دیگر) کم باشد. مسئله تشخیص جامعه نقش مهمی برای درک ساختار و عملکرد شبکه‌های پیچیده دارد [۳]. برای مثال کشف جامعه در شبکه‌های اجتماعی آنلاین بر اساس ویژگی‌های مشترک کاربران مانند سرگرمی‌ها و علایق می‌تواند خدمات ارائه شده توسط شبکه‌های اجتماعی را بهبود ببخشد [۴].

در سال‌های اخیر، روش‌های متعددی برای حل مسئله تشخیص جامعه در شبکه‌های اجتماعی پیشنهاد شده است. با این حال تحقیقات فورتوناتو و بارتلمی [۵] نشان داد که روش‌های سنتی تشخیص جوامع (مانند بهینه‌سازی ماژولاریتی) قادر به تشخیص جوامع کوچک‌تر از اندازه مشخصی نیستند. از رویکردهایی که برای حل این مشکل ارائه شده است می‌توان به استفاده از آتوماتاهای

یادگیر در تشخیص جامعه اشاره کرد.

آتوماتای یادگیر، یک ماشین تصمیم‌گیری است که دارای مجموعه‌ای از اقدامات است. آتوماتای یادگیر از میان این مجموعه، یکی را انتخاب کرده و به محیطی که با آن تعامل دارد ارسال می‌نماید. بعد از هر بار ارسال اقدام انتخاب شده، محیط به آتوماتای یادگیر پاسخ می‌دهد. با تکرار تعامل میان آتوماتای یادگیر و محیط، اقدام بهینه محیط توسط آتوماتا شناخته می‌شود (به عبارت دیگر آتوماتا با تکرار تعامل با محیط یاد می‌گیرد که کدام اقدام بهینه است) [۶].

در تمامی روش‌های تشخیص جامعه مبتنی بر آتوماتاهای یادگیر که تاکنون ارائه شده است، تعداد اقدامات آتوماتاها، برابر با تعداد همسایگان هر گره متناظر با آتوماتا است. همچنین در هیچ یک از این روش‌ها از معیارهای شباهت میان گره‌ها استفاده نشده است. روش پیشنهادی در این مقاله برای بهبود الگوریتم‌های پیشین از معیارهای شباهت (مانند ژاکارد) استفاده کرده است. همچنین ما در این الگویتم شعاع همسایگی را افزایش دادیم که منجر به بهبود دقت روش تشخیص جامعه می‌شود. از دیگر نوآوری‌های روش پیشنهادی، ثابت در نظر گرفتن تعداد اقدامات آتوماتاها (برابر با تخمین تعداد جوامع هر گراف) است که منجر به امکان پیاده‌سازی این الگوریتم به صورت موازی خواهد شد .

ادامه این مقاله به صورت زیر ارائه شده است. بخش دوم به تعاریف و مفاهیم پایه می‌پردازد. در بخش سوم، مقالات و کارهای صورت گرفته پیشین بررسی شده است. در بخش چهارم نسخه‌های الگوریتم پیشنهادی از ابتدا تا بهترین نسخه طراحی شده و تعمیم‌های آن‌ها و روند کلی کار به‌طور کامل ارائه شده است. در بخش پنجم ابتدا متغیرها و پارامترهای موجود معرفی شده و سپس نتایج به دست آمده از آزمایش‌ها ارائه شده است. در بخش ششم، نتیجه‌گیری کلی و همچنین پیشنهادهایی به منظور توسعه و بهبود چنین روش‌هایی برای کشف جوامع در شبکه‌های پیچیده ارائه شده است.

۲- مفاهیم پایه

در ابتدای این بخش، برای درک مفهوم تشخیص جامعه در شبکه‌های اجتماعی، ابتدا مفهوم شبکه پیچیده بررسی شده و سپس برپایه آن، مفهوم شبکه اجتماعی مطرح شده است. در ادامه با توجه به این مفاهیم تشخیص جامعه و روش‌های آن تشریح شده‌اند. سپس، آتوماتاهای یادگیر و انواع آن معرفی شده‌اند. در انتها معیارهای شباهت و معیارهای اندازه‌گیری کیفیت جوامع استفاده شده در این مقاله تعریف شده است.

۲-۱- تشخیص جامعه[۱] در شبکه‌های پیچیده[۲]

یک شبکه مجموعه‌ای از عناصر (گره‌ها) با پیوندهایی (یال‌ها) است که این عناصر را به هم متصل می‌کنند. هر شبکه را می‌توان با مجموعه‌ی دوتایی G=(V,E) نشان داد که در آن V مجموعه‌ی گره‌های شبکه و E مجموعه یال‌های شبکه است. یک شبکه پیچیده دارای ویژگی‌های توپولوژیکی مشخص است. ویژگی‌هایی که در شبکه‌های ساده از قبیل گراف‌های تصادفی رخ نمی‌دهند ولی اغلب در گراف‌های سیستم‌های دنیای واقعی رخ می‌دهند [۷]. شبکه‌های رایانه‌ای، شبکه‌های تکنولوژی، شبکه‌های مغزی و شبکه‌های اجتماعی از این قبیل شبکه‌ها هستند. دو ویژگی مشترک شبکه‌های پیچیده عبارت است از: ۱) آن‌ها بزرگ هستند، از صدها و هزاران یا بیشتر عنصر (گره) تشکیل شده‌اند و ۲) الگوهای اتصالات آن‌ها نه کاملا منظم و نه کاملاً تصادفی است [۸].

جوامع در شبکه‌ها گروه‌هایی مجزا از گره‌ها هستند که ارتباطات میان گره‌ها در داخل جامعه‌ای که به آن تعلق دارند زیاد و ارتباطات آن‌ها با سایر جوامع کم است. کشف جوامع مهم‌ترین راه برای استخراج و استفاده از اطلاعات شبکه‌ها می‌باشد. هرچند از نظر محاسباتی کشف جوامع در شبکه‌های بزرگ مقیاس غیرقابل‌حل است.

۲-۲- اتوماتای یادگیر

یک آتوماتای یادگیر، نوعی ماشین است که می‌تواند خود را با یک محیط تصادفی تطبیق دهد. آتوماتای یادگیر شامل مجموعه‌ای از اقدام‌ها است که یکی از آن‌ها در هر مرحله انتخاب شده و برای این محیط تصادفی ارسال می‌شود. با پاسخ محیط به اقدام انتخاب شده، آتوماتای یادگیر نحوه عملکرد خود را به‌روزرسانی می‌کند؛ محیط با فرستادن یک سیگنال مثبت یا منفی پاسخ می‌دهد و بر اساس این پاسخ، آتوماتای یادگیر احتمال انجام اقدام‌هایش را به‌روزرسانی می‌کند. این فرآیند تا زمانی ادامه می‌یابد که آتوماتای یادگیر اقدام بهینه خود را از میان مجموعه متناهی اقدام‌هایش تشخیص دهد؛ اقدام بهینه، اقدامی است که محیط با بالاترین احتمال به آن پاسخ مثبت می‌دهد.

یک آتوماتای یادگیر با مجموعه شش‌تایی $\{\phi.\alpha.\beta.A.G.P\}$ تعریف می‌شود که در آن $\alpha = \{\alpha_1.\alpha_2.\dots.\alpha_r\}$ مجموعه اقدام‌های آتوماتا، $\beta = \{\beta_1.\beta_2.\dots.\beta_r\}$ مجموعه ورودی‌های آتوماتا یا بازخوردهای دریافتی از محیط، $\phi(n) = \{\phi_1.\phi_2.\dots.\phi_k\}$ مجموعه وضعیت‌های داخلی آتوماتا در لحظه n ، $A = \phi \times \beta \rightarrow$

ϕ الگوریتم یادگیری، $G = \phi \rightarrow \alpha$ تابعی است که مجموعه ورودی‌ها را به مجموعه خروجی تطبیق می‌دهد و $P = \{P_1.P_2.\ldots.P_r\}$ مجموعه بردارهای احتمالی است که احتمال انتخاب یک اقدام را از مجموعه‌ی اقدامات در هر مرحله تعیین می-کند.

محیط با مجموعه‌ی سه‌تایی $E = \{\alpha.\beta.c\}$ تعریف می‌شود که در آن $\alpha = \{\alpha_1.\alpha_2.\ldots.\alpha_r\}$ مجموعه‌ی ورودی‌های محیط ، $\beta = \{\beta_1.\beta_2.\ldots.\beta_r\}$ مجموعه خروجی‌های محیط و $c = \{c_1.c_2.\ldots.c_r\}$ مجموعه احتمال‌های جریمه می‌باشند. ورودی محیط یکی از r اقدام انتخاب شده‌ی آتوماتاست. خروجی (پاسخ) به هر اقدام i انتخاب شده توسط β_i مشخص می‌شود. در محیطی که β_i یک پاسخ دودویی باشد، $\beta_i = 1$ به عنوان پاسخ نامطلوب یا جریمه و $\beta_i = 0$ به عنوان پاسخ مطلوب یا پاداش در نظر گرفته می‌شوند.

۲-۲-۱- الگوریتم یادگیری

هر آتوماتای یادگیر در هر مرحله اقدامی مانند α_i را انتخاب می‌کند. اگر پاسخ محیط به این اقدام پاداش باشد، احتمال انتخاب مجدد این اقدام (P_i) برای مراحل بعدی افزایش و احتمالات سایر اقدامات کاهش می‌یابد. در غیر این صورت اگر پاسخ محیط برای این اقدام جریمه باشد احتمال این اقدام کاهش و احتمالات سایر اقدامات آتوماتا افزایش می‌یابد. حاصل جمع P_i ها همواره برابر عدد یک است پس تغییرات در میزان احتمالات اقدامات باید به‌گونه‌ای باشد که مجموع احتمالات ثابت بمانند.

اگر a را پارامتر پاداش[۳] و b را پارامتر جریمه[۴] در نظر بگیریم و در مرحله nام اقدام α_i انتخاب شده باشد آنگاه به‌روزرسانی بردار احتمال به شکل زیر خواهد بود. مقدار پارامتر a و پارامتر b همواره در بازه $[0,1]$ می‌باشد. در صورت دریافت سیگنال پاداش از محیط از رابطه (۱) داریم:

$$\begin{aligned} &P_i(n+1) = P_i(n) + a[1 - P_i(n)] \\ &\forall j\ .\ j \neq i \quad P_j(n+1) = (1-a)P_j(n) \end{aligned} \qquad (۱)$$

در صورت دریافت سیگنال جریمه از محیط از رابطه (۲) داریم که در آن r تعداد اقدامات هر آتوماتا می‌باشد.

$$\begin{aligned} &P_i(n+1) = (1-b)P_i(n) \\ &\forall j\ .\ j \neq i \quad P_j(n+1) = \frac{b}{r-1} + (1-b)P_j(n) \end{aligned} \qquad (۲)$$

۲-۲-۲- آتوماتای یادگیر سلولی[۵]

یک آتوماتای سلولی، شبکه‌ای متشکل از سلول‌های یکسان است. هرکدام از این سلول‌ها می‌تواند تعداد محدودی حالت داشته باشد که این حالت‌ها با توجه به قوانین ساده محلی به‌روزرسانی می‌شوند. از آنجاکه فضا، زمان و حالت‌های سیستم در آتوماتای سلولی گسسته هستند پس حالت یک سلول در هر زمان فقط به مرحله قبل همان سلول بستگی دارد. به‌روزرسانی‌های همه‌ی سلول‌ها به‌طور هم‌زمان است بنابراین حالت کل شبکه در مراحل زمانی گسسته پیشرفت می‌کند.

آتوماتای یادگیر سلولی ترکیبی از آتوماتای یادگیر و آتوماتای سلولی است؛ که در آن هر سلول شامل تعدادی آتوماتای یادگیر است. آتوماتای یادگیر سلولی مشابه آتوماتای سلولی دارای قوانین ساده‌ای در محدوده هر سلول است و رفتار هر واحد سلول با توجه به رفتار همسایگان آن و انتخاب بهینه‌ای که توسط آتوماتای یادگیر مشخص می‌شود، تعیین می‌گردد.

۲-۲-۳- آتوماتای یادگیر سلولی نامنظم[۶]

این ماشین، بهبودیافته مدل آتوماتای یادگیر سلولی است. در یک آتوماتای یادگیر سلولی منظم، ساختار منظم و مستطیلی شکل است اما در دنیای واقعی بسیاری از شبکه‌ها، مانند شبکه‌های بی‌سیم ساختاری نامنظم دارند؛ بنابراین در مدل آتوماتای یادگیر سلولی نامنظم، ساختار سلول‌ها نامنظم و گراف بدون جهت است. عملکرد آتوماتاهای یادگیر سلولی نامنظم مشابه با آتوماتای یادگیر سلولی است به‌طور مشابه به هر سلول یک یا چند آتوماتای یادگیر اختصاص داده شده که بر اساس قوانین محلی همان سلول، تعیین می‌شود که عمل انتخاب شده یکی از آتوماتاها در سلول پاداش و یا جریمه دریافت می‌کند. تنها تفاوت این نوع از آتوماتا با آتوماتای یادگیر سلولی در مدل تعریف همسایگان آن‌ها می‌باشد.

۲-۳- معیارهای شباهت[۷]

معیارهای شباهت، معیارهایی هستند که میزان شباهت (ویژگی‌ها یا زمینه‌های مشترک) میان گره‌های شبکه را با توجه به ویژگی‌های ساختاری شبکه محاسبه می‌کنند. معیارهای شباهت زیادی میان گره‌ها در زمینه تشخیص جامعه وجود دارد. در این بخش به‌طور خلاصه سه معیار شباهت ژاکارد[۸]، شاخص آدامیک‌آدار[۹] و همسایگان مشترک[۱۰] بررسی شده‌اند. انتخاب اینکه کدام معیار شباهت برای هر الگوریتم تشخیص جامعه مناسب است تا حد زیادی به اندازه شبکه و فراداده‌های موجود در مورد گره‌ها و تعاملات آن‌ها بستگی دارد.

معیارهای شباهت محلی، شبه محلی یا سراسری هستند. از آنجا که معیارهای شباهت سراسری از نظر هزینه محاسباتی در شبکه‌های بزرگ ایجاد مشکل می‌کنند. در این بخش از این معیارها در مقابل

معیارهای شباهت محلی صرف نظر شده است [۹].

۲-۳-۱- شاخص شباهت ژاکارد

شاخص شباهت ژاکارد معیاری برای محاسبه شباهت مجموعه‌های آماری است. میزان شباهت ژاکارد دو گره i,j از تقسیم اشتراک این دو گره بر اجتماع آن‌ها محاسبه می‌شود و با $J(i,j)$ نمایش داده می‌شود که در رابطه (۳) نشان داده شده است [۹]؛ در این رابطه $NE(i)$ برابر با همسایه‌های گره i می‌باشد.

$$\mathrm{J(i,j)} = \frac{|NE(i) \cap NE(j)|}{|NE(i) \cup NE(j)|} \qquad (۳)$$

۲-۳-۲- شاخص آدامیک‌آدار

شاخص آدامیک‌آدار مربوط به تعداد ویژگی‌های مشترک هر دو مجموعه می‌شود. در مسائل مربوط به تشخیص جوامع ویژگی‌های مشترک همان تعداد همسایه‌های مشترک گره‌ها می‌باشد که با $\mathrm{AA(i,j)}$ نشان داده می‌شود و از طریق رابطه (۴) به دست می‌آید [۱۰]؛ در این رابطه $NE(i)$ برابر با همسایه‌های گره i می‌باشد و گره k همسایه مشترک دو گره i,j است. همچنین $|NE(k)|$ برابر تعداد همسایه‌های گره k می‌باشد.

$$\mathrm{AA(i,j)} = \sum_{k \in NE(i) \cap NE(j)} \frac{1}{\log |NE(k)|} \qquad (۴)$$

۲-۳-۳- همسایگان مشترک

تعداد همسایگان مشترک بین دو گره i,j را اندازه گیری می‌کند. علی‌رغم سادگی این شاخص، در یک شبکه گره‌های با تعداد همسایگان مشترک زیاد با احتمال بیشتری به یک جامعه تعلق دارند و همچنین احتمال به‌وجود آمدن پیوند میان آن‌ها در وجود دارد. همسایگان مشترک با $C(\mathrm{i,j})$ نشان داده می‌شود و از طریق رابطه (۵) محاسبه می‌شود [۱۰]؛ در این رابطه $NE(i)$ برابر با همسایه‌های گره i می‌باشد.

۲-۴- معیارهای ارزیابی

ارزیابی هر روش پیشنهادی نیازمند معیارهای مشخص و قابل‌اطمینان است. از معتبرترین معیارهای ارزیابی در بحث تشخیص جامعه در شبکه‌ها می‌توان به معیارهای ماژولاریتی و NMI اشاره کرد. این دو معیار کمک می‌کنند قدرت روش پیشنهادی این تحقیق نسبت به سایر کارهای مشابه مشخص شود. در ادامه این بخش این دو معیار ارزیابی بررسی می‌شوند.

۲-۴-۱- ماژولاریتی[۱۱]

ماژولاریتی معیاری است که قدرت تقسیم یک شبکه به ماژول‌ها (خوشه‌ها یا جوامع) را اندازه‌گیری می‌کند. شبکه‌های دارای ماژولاریتی بالا دارای اتصالات متراکم درون ماژول‌ها (جوامع) و اتصالات با تراکم کم بین گره‌ها در ماژول‌های مختلف هستند. ماژلاریتی از رابطه (۶) محاسبه می‌شود که در آن m برابر تعداد یال‌های گراف، K_i برابر درجه گره i و A_{ij} ماتریس مجاورت بین دو گره j و i است. اگر برچسب جامعه دو گره i و j یکسان باشد آنگاه مقدار $\delta_{ij} = 1$ و در غیر این صورت $\delta_{ij} = 0$ خواهد بود [۱۱].

$$modu = \frac{1}{2m}\sum_{i,j}(A_{ij} - \frac{K_i K_j}{2m})\delta_{ij} \qquad (۶)$$

۲-۴-۲- معیار NMI

NMI یک معیار مبتنی بر تئوری اطلاعات است که کیفیت ساختار جامعه استخراج شده را نسبت به جامعه واقعی اندازه‌گیری می‌کند. در حقیقت این معیار ساختار جامعه واقعی را به عنوان مبنا در نظر می‌گیرد و از این طریق توانایی الگوریتم تشخیص جامعه را ارزیابی می‌کند. NMI از رابطه (۷) محاسبه می‌شود.

$$\begin{aligned} &NMI(X|Y) = 1 - [H(X|Y) + H(Y|X)]/2 \\ &H(X) = -\sum_{i=1}^{n} p(x_i log x_i) \\ &H(X) = \sum_{i,j} p(x_i, y_i) log \frac{p(y_i)}{p(x_i, y_i)} \end{aligned} \qquad (۷)$$

در این معادله Y, X مجموعه جوامع واقعی و جوامع تشخیص داده شده توسط الگوریتم هستند، x_i, y_i گره‌هایی از مجموعه‌های Y, X و $H(X)$ آنتروپی مجموعه X و همچنین $H(X|Y)$ آنتروپی X به شرط Y است [۱۲]. مقدار معیار NMI همواره در محدوده $[0.1]$ است و هرچه این مقدار بیشتر باشد نشان دهنده مطابقت بیشتر جامعه تشخیص داده شده با جامعه واقعی است.

۳- مرور کارهای گذشته بر پایه آتوماتای یادگیر

درگذشته بسیاری از محققین، در زمینه تشخیص جامعه مطالعات گسترده‌ای انجام داده‌اند. مقالات زیادی این روشها را مرور و دسته‌بندی کرده‌اند. به عنوان مثال در [۱۳]، این روشها به هفت دسته تقسیم شده است. اکثر روشها به تشخیص جامعه غیرهمپوشان اختصاص دارد که در آن‌ها ورودی یک گراف معمولی است (مجموعه‌ای از گره‌ها که نشان دهنده اشخاص است که توسط یال‌ها به هم متصل شده‌اند) و خروجی فهرستی از گروه‌های گره‌ها است که جوامع روی گراف اولیه را نشان می‌دهد (در این روش‌ها

هر شخص فقط متعلق به یک جامعه است). از آنجا که این مقاله بر استفاده از آتوماتای یادگیر در حل مسئله تشخیص جوامع تأکید دارد، روش‌ها و کارهای انجام شده در تشخیص جامعه بر پایه آتوماتای یادگیر مرور می‌شوند. تمامی کارهایی که در ادامه مرور می‌شوند به استفاده از آتوماتای یادگیر در حل مسئله تشخیص جامعه در شبکه یا خوشه‌بندی در گراف می‌پردازند و یا از آن استفاده می‌کنند.

ژائو و همکاران [۱۳] اولین الگوریتم کشف جوامع بر پایه‌ی آتوماتای یادگیر را طراحی کردند. الگوریتم ارائه‌شده بدون نیاز به مقادیر اولیه‌ای، نظیر تعداد جوامع به کشف جوامع از طریق آتوماتای یادگیر در شبکه‌های پیچیده می‌پردازد. این الگوریتم متناظر با شبکه یک ICLA تعریف می‌کند. در اولین مرحله هرکدام از آتوماتاها به‌صورت تصادفی اقدامی را انتخاب و سپس با توجه به بازخوردهای دریافتی از محیط بردار احتمال خود را به‌روزرسانی می‌کنند. دو شرط در این به‌روزرسانی بررسی و منجر به پاداش یا جریمه خواهند شد: ۱) بهبود معیار ماژولاریتی جوامع جدید و ۲) بیشتر بودن ارتباط هر گره با جامعه‌ای که به آن تعلق دارد نسبت به ارتباطات آن با سایر جوامع. اگر این دو شرط برقرار باشند، سیگنال پاداش برای آتوماتا و در غیر این صورت سیگنال جریمه صادر می‌شود. سپس با توجه به بردارهای احتمال جدید، هرکدام از آتوماتاها اقدامی را انتخاب می‌کنند. مراحل فوق تا آنجا تکرار خواهند شد که تغییری در اقدام انتخابی آتوماتاها صورت نگیرد و بردارهای احتمال به حالت پایدار خود برسند. به‌این‌ترتیب جوامع شناسایی می‌شوند. از آنجا که هر گره در این الگوریتم تنها به یک جامعه اختصاص می‌یابد پس کارایی آن تنها در شناسایی جوامع غیرهمپوشان است.

دلیری و همکاران [۱۴] الگوریتمی مبتنی بر آتوماتای یادگیر توزیع شده برای کشف جامعه ارائه کردند. در این الگوریتم به هر گره یک آتوماتای یادگیر الحاق شده که می‌تواند در طول اجرای الگوریتم، فعال یا غیرفعال باشد. در ابتدا تمامی آتوماتاهای یادگیر غیرفعال هستند و با اجرای اولین دور از الگوریتم اولین گره تصادفاً انتخاب و آتوماتای آن فعال می‌شود. سپس از میان گره‌های مجاور خود، گرهی را با توجه به بردار احتمال خود انتخاب می‌کند تا در یک جامعه باشند. اگر تراکم ارتباطات جامعه جدید با افزودن این گره بیشتر از مرحله قبل باشد گره جدید به جامعه تعلق می‌گیرد و آتوماتای آن گره نیز فعال می‌شود. به‌روزرسانی بردار احتمال بر اساس شرط چگالی ارتباطات میان گره‌ها است که اگر بهبودیافته باشد پاداش و در غیر این صورت جریمه خواهد شد. درنهایت این چرخه تا آنجا تکرار خواهد شد که آستانه‌ی تعریف شده برای بردارهای احتمال آتوماتاهای یادگیر ارضا شود و الگوریتم متوقف شود و جوامع بهینه به‌عنوان خروجی مشخص گردند.

غمگسار و همکاران [۱۵] مجدداً الگوریتم جدیدی مطرح کردند که از حالات متفاوت آتوماتای یادگیر برای کشف جامعه استفاده می‌کرد. در این الگوریتم، آتوماتای یادگیر دارای چهار حالت فعال، غیرفعال، شلیک و خاموش است و به هر گره یک آتوماتای یادگیر الحاق شده است. در ابتدا تمامی آتوماتاهای یادگیر غیرفعال هستند، پس اولین مرحله از الگوریتم فعال کردن تعدادی از آتوماتاهای یادگیر است. تعداد اقدام‌های هر آتوماتا برابر با تعداد همسایگان گره متناظر آن است. آتوماتاهای فعال شده به حالت شلیک رفته و از میان همسایگان خود با توجه به بردارهای احتمالشان یکی را انتخاب می‌کنند. در اینجا آتوماتای گره انتخابی به حالت فعال درآمده و آتوماتای قبلی خاموش می‌شود. اگر حداقل نیمی از همسایگان آتوماتای جدید در حالت خاموش یا فعال باشند گره متناظر آن با گره متناظر آتوماتای انتخاب کننده این گره، در یک جامعه قرار می‌گیرند. این روند برای آتوماتاهای فعال شده تکرار می‌شود. برای به‌روزرسانی بردارهای احتمال آتوماتاهای یادگیر از مفاهیم تراکم (متوسط درجات داخلی گره‌های متعلق به جامعه نسبت به متوسط درجه گره‌ها در کل شبکه) و قابلیت تفکیک‌پذیری (نسبت تعداد یال‌های بین هر دو جامعه به متوسط همه‌ی یال‌های درون جوامع) استفاده شده است. در پایان هر تکرار از الگوریتم آتوماتاها به حالت غیرفعال درمی‌آیند. و الگوریتم با پایان یافتن تعداد تکرار مشخص پایان می‌یابد و جوامع بهینه به‌عنوان خروجی الگوریتم مشخص می‌شوند.

دلیری و همکاران الگوریتم بعدی خود را منتشر کردند [۱۶]. در این الگوریتم علاوه بر آتوماتای یادگیر سلولی از امکانات درخت پوشا نیز برای کشف جامعه استفاده‌شده است. در اینجا از تعامل آتوماتاهای یادگیر درخت پوشایی تشکیل می‌شود که نمایانگر جوامع محلی است. ساخت درخت تا آنجا ادامه می‌یابد که یکی از شروط توقف الگوریتم ارضا شود. ازآنجاکه که سیگنال‌های دریافتی از محیط در این الگوریتم به‌صورت محلی و سراسری هستند پس شروط توقف الگوریتم نیز وابسته به رسیدن به آستانه مورد نیاز در هر دو حالت محلی و سراسری می‌باشد.

دلیری و همکاران [۱۷] نیز الگوریتم خود را برای کشف جامعه در شبکه‌های وزن‌دار بر پایه آتوماتای یادگیر مطرح کردند. در این مقاله از طریق تعاملات آتوماتاهای یادگیر و همچنین موازی‌سازی، جوامع بهینه در زمانی مناسب تشخیص داده شده‌اند. این الگوریتم اطلاعات شبکه را به‌عنوان وزن روی آن نشان می‌دهد. هر آتوماتا به تعداد همسایگانش اقدام دارد. با شروع الگوریتم هر آتوماتا اقدامی را

انتخاب می‌کند. اگر دو شرط زیر ارضا شوند اقدام انتخابی پاداش و در غیر این صورت جریمه می‌شود: ۱) بیشتر بودن تعداد اتصالات داخلی جوامع از تعداد اتصالات خارجی آن‌ها و ۲) بیشتر بودن مجموع وزن اتصالات داخلی هر گره در جامعه از مجموع وزن‌های اتصالات خارجی آن. این روند تا آنجا که تغییری در ساختار جوامع ایجاد نشود ادامه می‌یابد و سپس الگوریتم متوقف می‌شود.

فتحی‌نوید ایده استفاده از ICLA در یافتن جامعه را به شبکه‌های چندلایه[۱۲] بسط داده است [۱۸]. در این مقاله نیز به هر گره یک آتوماتای یادگیر الحاق می‌شود. برای بروزرسانی آتوماتای یادگیر از دو سیگنال محلی و سراسری استفاده می‌شود. سیگنال محلی توسط همسایگان گره در لایه مشخص تولید می‌شود ولی سیگنال سراسری از سایر لایه‌های شبکه بدست می‌آید.

استفاده از آتوماتای یادگیر، مفهوم محیط محلی و محیط سراسری در تولید سیگنال به‌روزرسانی برای حل مسئله خوشه‌بندی در مقاله مینوفام و همکارانش نیز دیده می‌شود[۱۹]. اما ایده اصلی این مقاله، روی یادگیری انتقالی[۱۳] است؛ انتقال یادگیری از یک CLA به یک CLA متناظر به صورتی که این انتقال از مبدأ به مقصد تعبیر می‌شود. برای به‌روزرسانی بردارهای احتمال آتوماتاهای یادگیر از محیط محلی و برای مشخص شدن زمان انتقال ازمحیط سراسری استفاده می‌شود. این چارچوب پیشنهادی جدید برای حل مسئله خوشه‌بندی kتایی استفاده شده است.

اما تشخیص جامعه مبتنی بر آتوماتای یادگیر می‌تواند به عنوان یک الگوریتم پایه در سایر مسائل گراف و شبکه کاربرد داشته باشد. به عنوان نمونه [۲۰] به دنبال یافتن k جامعه یا خوشه در گراف است به‌طوری که گره مستقر در مرکز خوشه‌ها را به عنوان مجموعه گره‌های تأثیر گذار انتخاب کند؛ مجموعه‌ای از گره‌ها تأثیرگذارترین هستند اگر بتواند بالاترین تعداد گره‌ها در شبکه را تحت تأثیر قرار دهند؛ به این مسئله، بیشینه سازی انتشار[۱۴] می‌گویند.

۴- روش‌های پیشنهادی

در این بخش، نسخه‌های مختلفی از الگوریتم تشخیص جامعه با استفاده از آتوماتای یادگیر پیشنهاد می‌شوند. در ادامه، دو قسمت ارائه می‌شود. قسمت اول شامل نسخه‌های مختلف الگوریتم پیشنهادی است و قسمت دوم تعمیم‌های الگوریتم پیشنهادی را معرفی می‌کند. این تعمیم‌ها قابلیت اجرا روی تمامِ نسخه‌های مختلف الگوریتم پیشنهادی قسمت اول را دارد. توضیحات هر یک از متغیرهای استفاده شده در روش‌های پیشنهادی در جدول ۱ به‌طور خلاصه آمده است.

جدول ۱: توضیح متغیرهای روش‌های پیشنهادی

نام متغیر	توضیح
LA_i	آتوماتای یادگیر نظیر گره i
v_i	برچسب (جامعه) گره i
$TI[i]$	میانگین شباهت محاسبه شده برای گره i
$AI_t[i]$	حاصل‌جمع مقدار شباهت‌های همسایه‌های گره i در تکرار t ام الگوریتم
$P_{i,t}[v_i]$	احتمال انتخاب برچسب v_i برای گره i توسط LA_i در تکرار t ام الگوریتم
$BI[i]$	متغیر کمکی نگهدارنده مقدار $AI_t[i]$
h	ضریب عددی در گام اول نسخه اول

۴-۱- نسخه‌های مختلف الگوریتم پیشنهادی

در این قسمت دو نسخه الگوریتم پیشنهادی ارائه شده است. نسخه اول، اولین ایده از الگوریتم پیشنهادی است و نسخه بعدی تکامل روش پیشنهادی در نسخه اول را نشان می‌دهد.

۴-۱-۱- نسخه اول

در ابتدا به ازای هر جفت گره همسایه در گراف مقدار معیار شباهت محاسبه می‌شود؛ $sim(i,j)$ مقدار شباهت میان دو گره همسایه i و j. برای محاسبه مقدار $sim(i,j)$ می‌توان از هرکدام از معیارهای شباهت شناخته شده مانند ژاکارد، آدامیک‌آدار یا تخصیص منابع استفاده کرد.

سپس به ازای هر گره i، $TI[i] = \frac{\sum_{j \in N_i} sim(i,j)}{|NE(i)|}$ (میانگین شباهت میان گره‌های همسایه i) محاسبه می‌شود. در این فرمول $NE(i)$ مجموعه همسایگان گره i، $TI[i]$ میانگین شباهت محاسبه شده برای گره i و $sim(i,j)$ مقدار شباهت محاسبه شده میان دو گره i و j است.

در ادامه و بعد از محاسبه مقدار میانگین شباهت برای تمامی گره‌های گراف، به هر گره یک آتوماتای یادگیر انتساب داده می‌شود (آتوماتای LA_i به گره i). تمامی آتوماتاها دارای k اقدام می‌باشد؛ انتخاب اقدام α_e توسط LA_i، بدین معنی است که گره i در جامعه e قرار دارد (بنابراین برچسب گره i، v_i، برابر با e است $v_i = e$). مقدار اولیه بردار احتمال LA_i برابر با $1/k$ در نظر گرفته می‌شود که k در آن برابر با تعداد جوامع است (تعداد جوامع به‌عنوان ورودی به الگوریتم داده می‌شود).

برای تولید سیگنال پاداش/جریمه برای آتوماتایLA_i مقدار $AI[i]$ برای گره نگهداری می‌شود. این مقدار در طول الگوریتم به‌روزرسانی

می‌شود. به‌عنوان مثال، نماد $AI_t[i]$ مقدار مورد نظر در مرحله t ام الگوریتم را نشان می‌دهد. در ابتدای الگوریتم، رابطه $AI_{t=0}[i] = 0, i = 1, \dots, n$ برقرار است.

الگوریتم نسخه اول یک الگوریتم تکراری است. در تکرار t ام از الگوریتم، آتوماتای $LA_i, i = 1, \dots, n$ برای گره i برچسب v_i را انتخاب می‌کند. مراحل نوشته شده در شبه‌کد ۱- قسمت الف برای محاسبه سیگنال تقویتی برای گره i در مرحله t ام و به‌روزرسانی بردار احتمال LA_i انجام می‌شود.

۴-۱-۲- نسخه دوم

در این نسخه هم مانند الگوریتم نسخه اول، معیار شباهت میان گره‌های همسایه محاسبه می‌شود (محاسبه $sim(i,j)$ به ازای تمام گره‌های همسایه i و j). در این نسخه هم برای محاسبه سیگنال پاداش و جریمه به $AI_t[i]$ نیاز است؛ رابطه $AI_{t=0}[i] = 0, i = 1, \dots, n$ برقرار است.

الگوریتم نسخه دوم نیز یک الگوریتم تکراری است. در تکرار t ام از الگوریتم، آتوماتای $LA_i, i = 1, \dots, n$ برچسب گره i را مشخص می‌کند. مراحل مشخص شده در شبه‌کد ۱-قسمت ب برای محاسبه سیگنال تقویتی برای گره i در مرحله t ام و به‌روزرسانی بردار احتمال LA_i انجام می‌شود. بروزرسانی بر اساس روابط ۸ و ۹ انجام می‌شود.

$$\begin{aligned} P_{i,t}[e] &= P_{i,t-1}[e] + a\big(1 - P_{i,t-1}[e]\big), e = v_i \\ P_{i,t}[q] &= (1-a)P_{i,t-1}[q], q \neq v_i \end{aligned} \qquad (۸)$$

$$\begin{aligned} P_{i,t}[e] &= (1-b)P_{i,t-1}[e], e = v_i \\ P_{i,t}[q] &= \left(\frac{b}{(k-1)}\right) + (1-b)P_{i,t-1}[q], q \neq v_i \end{aligned} \qquad (۹)$$

تفاوت دو نسخه پیشنهادی در نحوه محاسبه $AI_t[i]$ است؛ محاسبه حاصل‌جمع مقدار شباهت‌های همسایه‌های گره i در تکرار t ام الگوریتم. به این محاسبات در روابط (۱۰) و (۱۲) اشاره شده است. در هر دو رابطه، به همسایگانی که در یک جامعه مشترک قرار دارند، توجه می‌شود. در نسخه اول تعداد این همسایگان در تصمیم‌گیری تأثیر گذار است اما در نسخه دوم میزان شباهت میان این همسایگان اهمیت بیشتری دارد.

۴-۲- تعمیم‌های الگوریتم

در این قسمت دو تعمیم پیشنهاد شده است. تعمیم اول با افزایش شعاع همسایگی سعی در بهبود نسخه پیشنهادی دوم دارد. تعمیم دوم یک سیگنال سراسری است که می‌تواند به هر دو نسخه پیشنهادی اضافه شود (در دو نسخه پیشنهادی تنها از اطلاعات محلی برای محاسبه سیگنال پاداش و جریمه استفاده شده است).

۴-۲-۱- تعمیم با شعاع دو

در نسخه دوم پیشنهاد شده، برای محاسبه سیگنال تقویتی گره i تنها همسایه‌های این گره (شعاع یک) در نظر گرفته می‌شود اما در این تعمیم شعاع همسایگی، دو می‌باشد. برای رسیدن به این هدف و تعمیم این روش بر نسخه پیشنهادی دوم، باید دو زیر مرحله ۱-۱ و ۱-۲ بعد از گام اول نسخه دوم اجرا شود (شبه‌کد ۱-قسمت ج). بعد از این دو زیر مرحله و تکمیل گام اول، گام‌های بعدی انجام می‌شوند؛ تعیین پاداش یا جریمه و به‌روزرسانی برادر احتمال آتوماتا LA_i.

۴-۲-۲- تعمیم با استفاده از سیگنال سراسری

در تمامی نسخه‌های پیشنهادی قبلی، برای تولید سیگنال تقویتی از اطلاعات همسایگان استفاده شده است؛ به عبارت دیگر سیگنال تقویتی محلی تولید شده و در اختیار آتوماتا قرار می‌گیرد. در تعمیم اول که پیشنهاد شد، همسایه‌های همسایگان یک آتوماتا نیز در تولید این سیگنال محلی تأثیر گذار هستند. این بدان معناست که شعاع همسایگی افزایش یافته است.

اما در تعمیم دوم پیشنهادی، توجه به عملکرد تمامی آتوماتاهای یادگیر در حل مساله تشخیص جوامع در شبکه مورد توجه است. بر این اساس یک سیگنال تقویتی سراسری برای تمامی آتوماتاها تولید می‌شود. آنها می‌توانند از ترکیب سیگنال محلی و سراسری برای بروزرسانی خود استفاده کنند.

تعمیم پیشنهادی که در آن سیگنال سراسری تولید و با سیگنال محلی ترکیب می‌شود در شبه‌کد ۱-قسمت د آمده است. این تعمیم برای دو نسخه پیشنهادی قابل اِعمال است (و حتی نسخه با تعمیم‌یافته شعاع دوم). برای انجام این تعمیم تنها کافی است گام دوم این نسخه‌ها با گام‌های شبه‌کد جابجا شوند.

گام اول) بر اساس رابطه (۱۰) مقدار $AI_t[i]$ برای گره i به‌روز می‌شود:

$$if\ |v_i = v_j|_{j\in N_i} \geq \frac{|N_i|}{2}$$
$$AI_t[i] = AI_{t-1}[i] + TI[i] * h$$
$$\text{Else}\quad AI_t[i] = AI_{t-1}[i] - TI[i] \qquad (۱۰)$$

اگر بیشتر از نیمی از همسایه‌های گره i برچسب مشابه آن را داشته باشند، از یک فرمول و در غیراین صورت فرمول دیگری استفاده می‌شود[۱۵].

گام دوم) حال اگر رابطه (۱۱) برقرار باشد، سیگنال پاداش برای LA_i صادر می‌شود ($\beta = 1$) و در غیر این صورت سیگنال جریمه صادر می‌شود ($\beta = 0$). به‌عبارت‌دیگر، چنانچه مقدار به‌دست‌آمده در مرحله t از متوسط مقادیر کسب‌شده قبلی بیشتر باشد، در این مرحله LA_i تشویق می‌شود.

$$AI_t[i] \geq \frac{\sum_{t=0}^{t-1} AI_t[i]}{t-1} \qquad (۱۱)$$

گام سوم) با دریافت سیگنال پاداش یا جریمه، برادر احتمال آتوماتای $LA_i, i = 1, \dots, n$ بر اساس رابطه (۸) به ازای $\beta = 1$ و رابطه (۹) به ازای $\beta = 0$ به‌روز می‌شود؛ در این روابط a و b پارامترهای پاداش و جریمه هستند، $P_{i,t}[e]$ احتمال انتخاب اقدام α_e است در مرحله t است.

الف. الگوریتم پیشنهادی نسخه اول

گام اول) بر اساس رابطه (۱۲) مقدار $AI_t[i]$ برای گره i به‌روز می‌شود:

$$AI_t[i] = AI_{t-1}[i] + \sum_{j\in N_i, v_i = v_j} sim(i,j) - \sum_{q\in N_i, v_i \neq v_q} sim(i,q) \qquad (۱۲)$$

در این رابطه گره‌های همسایه j با گره i هم برچسب هستند ($v_i = v_j$) و گره‌های همسایه q برچسب متفاوتی با گره i دارند و $sim(i,j)$ مقدار معیار شباهت میان دو گره i و j است.

گام دوم) مشابه با نسخه اول، چنانچه مقدار به‌دست‌آمده در مرحله t از متوسط مقادیر کسب‌شده قبلی بیشتر باشد، رابطه (۱۱)، آتوماتا تشویق و در غیر این صورت جریمه می‌شود.

گام سوم) با دریافت سیگنال پاداش یا جریمه، برادر احتمال آتوماتای $LA_i, i = 1, \dots, n$ بر اساس رابطه (۸) به ازای $\beta = 1$ و رابطه (۹) به ازای $\beta = 0$ به‌روز می‌شود؛ در این روابط a و b پارامترهای پاداش و جریمه هستند، $P_{i,t}[e]$ احتمال انتخاب اقدام α_e است در مرحله t است.

ب. الگوریتم پیشنهادی دوم

۱-۱) در گام اول نسخه دوم مقداری برای $AI_t[i], i = 1, \dots, n$ محاسبه شده است. این مقدار را در متغیر کمکی BI ذخیره می‌کنیم ($BI[i] = AI_t[i]$).

۱-۲) بر اساس رابطه (۱۳) مقدار $AI_t[i]$ برای گره i محاسبه می‌شود:

$$AI_t[i] = BI[i] + \sum_{j\in N_i, v_i = v_j} BI[j] - \sum_{q\in N_i, v_i \neq v_q} BI[q] \qquad (۱۳)$$

گره همسایه j با گره i هم برچسب هستند و گره همسایه q برچسب متفاوتی با گره i دارند.

ج. تعمیم پیشنهادی اول- شعاع دو

گام دوم تعمیم‌یافته:

۲-۱) مقدار ماژولاریتی در هر مرحله به ازای جوامع تشکیل‌شده بر اساس رابطه (۱۴) محاسبه می‌گردد که با $modu[t]$ نشان داده می‌شود.

$$modu[t] = \frac{1}{2m}\sum_{i,j}(A_{ij} - \frac{K_i K_j}{2m})\delta_{ij} \qquad (۱۴)$$

در این رابطه m برابر تعداد یال‌های گراف، K_i برابر درجه گره i و A_{ij} ماتریس مجاورت بین دو گره j و i است. اگر برچسب جامعه دو گره i و j یکسان باشد آنگاه مقدار $\delta_{ij} = 1$ و در غیر این صورت $\delta_{ij} = 0$ خواهد بود.

۲-۲) اگر روابط $AI_t[i] \geq \frac{\sum_{t=0}^{t-1} AI_t[i]}{t-1}$ و $modu[t] \geq modu[t-1]$ برقرار باشند، سیگنال پاداش برای LA_i صادر می‌شود ($\beta = 1$) و در غیر این صورت اگر حداقل یکی از این روابط برقرار نباشد سیگنال جریمه صادر می‌شود ($\beta = 0$).

د. تعمیم پیشنهادی دوم-سیگنال سراسری

شبه‌کد ۱: الگوریتم های پیشنهادی الف) نسخه اول، ب) نسخه دوم، ج) تعمیم پیشنهادی اول و د) تعمیم پیشنهادی دوم

۴-۳- بحث و بهبود روش‌های پیشنهادی

لگوریتم‌های حاصل از نسخه‌ها و تعمیم‌های پیشنهادی، هفت عدد هستند. نسخه‌های اول و دوم هر کدام یک الگوریتم مجزا هستند. ترکیب این دو نسخه با تعمیم‌های پیشنهادی منجر به تولید پنج الگوریتم مجزای دیگر می‌شود.

در ادامه و قبل از ارائه نتایج، چند نکته و سوال را در خصوص روشهای پیشنهادی مطرح کنیم؛ پاسخ برخی از آنها در ادامه آمده است و پاسخ برخی در پیوست الف آمده است تا پیوستگی مقاله حفظ شود:

الف) نقش متغیر $AI_t[i]$ چیست و چطور به یادگیری آتاماتای یادگیر کمک می‌کند؟ (پاسخ در پیوست الف).

ب) آیا استفاده از ماژولاریتی، الگوریتم را با چالش روبرو نمی‌کند؟ (پاسخ در پیوست الف).

ج) ثابت بودن k (تعداد جوامع) در روش پیشنهادی چه نکات مثبتی را به همراه دارد؟ (پاسخ در پیوست الف).

د) پیچیدگی زمانی الگوریتم‌های پیشنهادی چیست؟ (پاسخ در

بخش ۰).

ه) چگونه سرعت همگرایی الگوریتم‌های پیشنهادی را افزایش دهیم؟ (پاسخ در بخش ۰).

۴-۳-۱- پیچیدگی زمانی الگوریتم‌های پیشنهادی

برای رسیدن به پاسخ مشخص و قابل مقایسه با سایر روشهای مشابه، برای آتوماتاهای یادگیر گام‌های زیر را در نظر بگیرید: ۱) انتخاب اقدام، ۲) تولید سیگنال محلی، ۳) تولید سیگنال سراسری و ۴) بروزرسانی بردار احتمالات اقدام‌ها. فرض کنید این گامها به تعداد $maxIter$ تکرار خواهند شد (تعداد تکرار لازم برای حل مسئله). دقت کنید که در روش‌های پیشنهادی، به تعداد گره‌های گراف (n)، آتاماتای یادگیر استفاده شده است. در ادامه، پیچیدگی زمانی هر گام گفته شده برای تمامی آتوماتاهای یادگیر به صورت یکجا محاسبه می‌شود. در نهایت، مجموع هزینه‌های محاسبه شده تمام چهار گام فوق در تعداد تکرار $maxIter$ ضرب خواهد شد.

- **پیچیدگی زمانی گامهای اول و چهارم**: هزینه انتخاب اقدام و بروزرسانی احتمالات اقدام با تعداد اقدام‌های آتوماتاها در ارتباط است. از آنجا که تعداد اقدام در روشهای پیشنهادی ثابت و برابر با k است، هزینه این دو از مرتبه $O(nk)$ خواهد بود.
- **پیچیدگی زمانی گام دوم:** برای محاسبه سیگنال محلی لازم است تا آتوماتای موجود در یک گره با تمامی همسایگان خود ارتباط داشته باشد. در مجموع برای تمامی گره‌های گراف، هزینه محاسبه با تعداد یالهای گراف (m) در ارتباط است: $O(m)$.
- **پیچیدگی زمانی گام سوم:** برای یکبار محاسبه سیگنال سراسری باید از تمام گره‌های گراف استفاده شود. بدین ترتیب هزینه محاسبه از مرتبه $O(n)$ خواهد بود.

بنابراین، هزینه مجموع چهار گام فوق $O(nk)+O(m)+O(n)+O(nk)$ خواهد بود. از آنجا که مقدار k ثابت و کوچک است می‌توان این مجموع را به صورت $O(3n+m)$ خلاصه کرد. در نهایت با در نظر گرفتن تعداد تکرار $maxIter$، پیچیدگی الگوریتم پیشنهادی که از سیگنال سراسری (تعمیم دوم) نیز استفاده می‌کند از مرتبه زیر است:

$$O(maxIter * (3n+m)) \quad (۱۵)$$

حال فرض کنید مانند بسیاری از روشهای مشابه، تعداد اقدام‌های آتوماتا برابر با تعداد همسایگان باشد. در این صورت هزینه الگوریتم چه تغییری می‌کند؟

با این فرض، تغییرات در هزینه گام‌های اول و چهارم اتفاق می‌افتد زیرا این گام‌ها، انتخاب اقدام و بروزرسانی احتمالات اقدام، با تعداد اقدام‌ها در ارتباط هستند. با فرض تعداد اقدام هر آتوماتا برابر با تعداد همسایگان، این هزینه با تعداد یال‌های گراف در ارتباط است. بدین ترتیب هزینه نهایی $O(maxIter * (n+3m))$ خواهد بود. از آنجا که تعداد یال‌ها در بسیاری از گرافهای رایج از تعداد گره‌ها بیشتر است، ایده ثابت بودن تعداد اقدام در آتاماتاهای یادگیر باعث کاهش پیچیدگی می‌شود.

این موضوع را می‌توان از لحاظ اندازه فضای جستجو نیز دنبال کرد. ثابت بودن تعداد اقدام‌های آتوماتا باعث کاهش اندازه فضای جستجوی می‌شود؛ اندازه فضای جستجو برابر با مجموع تعداد اقدام‌های تمامی آتاماتاهای یادگیر است. در روش پیشنهادی این مقدار nk است در حالیکه در روشهای مشابه $2m$ است (در این روش‌ها، هر آتوماتا به تعداد همسایگان اقدام دارد و مجموع تعداد تمام اقدام‌ها برابر با $2m$ خواهد بود). چنانچه مقدار $k \leq \frac{2m}{n}$ باشد، اندازه فضای جستجو برای روش‌های پیشنهادی این مقاله نسبت به روشهای مشابه کمتر است. با توجه به آنکه مقدار k (تعداد جوامع موجود در شبکه) نسبت به $\frac{2m}{n}$ در اکثر گرافهای رایج بسیار کوچکتر است، با روش‌های پیشنهادی فضای جستجوی کوچکتری ایجاد می‌شود.

۴-۳-۲- افزایش سرعت همگرایی

در این بخش یک تغییر در اجرای الگوریتم پیشنهاد می‌شود که سرعت همگرایی روش را افزایش می‌دهد. در این بهبود به‌منظور افزایش سرعت همگرایی آتوماتاهای یادگیر، در به‌روزرسانی بردار احتمال هر آتوماتا:

الف. اگر مقدار احتمال یکی از اقدام‌های آتوماتای یادگیر بالاتر از ۰.۹۵ شود آنگاه احتمال آن اقدام به ۱ و احتمال متناظر با سایر اقدام‌های آتوماتا به صفر تبدیل می‌شود و بردار احتمال این آتوماتا تا انتهای اجرای الگوریتم ثابت می‌ماند (به عبارتی این آتوماتای یادگیر خاموش می‌شود). رابطه (۱۶) این روند را برای اقدام α_e از آتوماتای LA_i نشان می‌دهد که در آن $P_{i,t}[e]$ احتمال انتخاب اقدام α_e در آتوماتای LA_i در لحظه t است و منظور از $\sim e$ سایر اقدام‌ها است.

$$if\ P_{i,t}[e] \geq 0.95\ then\ P_{i,t}[e]=1\ , P_{i,t}[\sim e]=0 \quad (۱۶)$$

ب. اگر مقدار احتمال یکی از اقدام‌ها کمتر از ۰.۰۵ شود آنگاه احتمال آن اقدام صفر شده (این اقدام از مجموعه اقدام‌های آتوماتا

حذف می‌شود) و احتمال آن به‌طور مساوی بین احتمال سایر اقدام‌ها تقسیم می‌شود. رابطه (۱۷) این روند را برای اقدام α_e از آتوماتای LA_i نشان می‌دهد که در آن $P_{i,t}[e]$ احتمال انتخاب اقدام α_e در آتوماتای LA_i در لحظه t است و منظور از $\sim e$ سایر اقدام‌ها است.

$$if\ P_{i,t}[e] \le 0..5\ then\ P_{i,t}[e] = 0\ , P_{i,t}[\sim e] = P_{i,t}[\sim e] + (P_{i,t}[\sim e]/k) \quad (۱۷)$$

جدول ۲: مجموعه داده‌های واقعی مورد استفاده در آزمایش‌ها؛ n به تعداد گره‌های و m تعداد یالهای گراف است.

شماره	نام	n	m	متوسط درجه	k
۱	کاراته	۳۴	۷۸	۲.۳	۲
۲	دلفین	۶۲	۱۵۹	۲.۶	۲ یا ۳*
۳	بینوایان	۷۷	۲۵۴	۳.۳	۱۱
۴	کتب سیاسی	۱۰۵	۴۴۱	۴.۲	۲
۵	فوتبال	۱۱۵	۶۱۳	۵.۳	۱۲
۶	کلمات مجاور	۱۱۲	۴۲۵	۳.۸	۶
7	علم شبکه	۱۵۸۹	۲۷۴۲	۱.۷	۲ یا ۴ یا ۱۱**

* تعداد جوامع در گراف دلفین در مقالات مختلف ثابت نیست. برای مثال در مقاله [۲۱] تعداد واقعی جوامع گراف دلفین برابر با دو در نظر گرفته شده است. در حالی که در مقاله [۲۲] تعداد جوامع این گراف برابر با سه می‌باشد. در این تحقیق تعداد جوامع گراف دلفین در آزمایش‌ها، برابر با دو است.

** مشابه با گراف دلفین گراف علم شبکه نیز در مقالات مختلف با تعداد جوامع مختلف (ده، چهار و یازده) استفاده شده است. در این تحقیق تعداد

۵- آزمایش‌ها

در این بخش هفت آزمایش طراحی شده است تا کارایی نسخه‌های مختلف پیشنهادی و تعمیم‌های آن‌ها نشان داده شوند. اهداف آزمایش‌ها بدین شرح است:

- بهترین حالت الگوریتم پیشنهادی Alg1 (آزمایش اول)
- بهترین نسخه الگوریتم پیشنهادی و یافتن بهترین ترکیب با تعمیم‌های پیشنهادی (آزمایش دوم تا پنجم)
- مقایسه بهترین ترکیب با سایر روشها (آزمایش ششم و هفتم)

مقایسه‌ها بر اساس دو معیار ماژولاریتی و NMI است. مجموعه داده‌های مورد استفاده در این آزمایش‌ها شامل هفت مجموعه داده است که در جدول ۲ مشخصات این مجموعه داده‌ها آمده است؛ شامل تعداد گره‌ها، تعداد یال‌ها، متوسط درجه هر گره و تعداد جوامع.

آزمایش‌های دیگری با استفاده از داده‌های ساختگی نیز در این مقاله انجام شده است. به دلیل حفظ پیوستگی و طولانی نشدن متن اصلی مقاله، توضیح کامل این دسته از آزمایش‌ها در پیوست ب آمده است.

۵-۱- آزمایش اول: انتخاب بهترین ضریب h در نسخه اول

آزمایش اول باهدف انتخاب بهترین مقدار متغیر h در گام اول نسخه اول بر اساس معیار ماژولاریتی روی ۳ شبکه واقعی طراحی شده است. در این آزمایش، مقادیر متفاوتی در بازه $[1,3]$ به عنوان مقدار متغیر h گام اول نسخه اول مقایسه می‌شوند. نتیجه این آزمایش در شکل ۱ آمده است این نتایج میانگین ۱۰ بار اجرای الگوریتم به ازای هر ضریب عددی متفاوت می‌باشد. این شکل نشان می‌دهد که ضریب عددی ۲.۵ بهترین عدد در رابطه با نسخه اول می‌باشد.

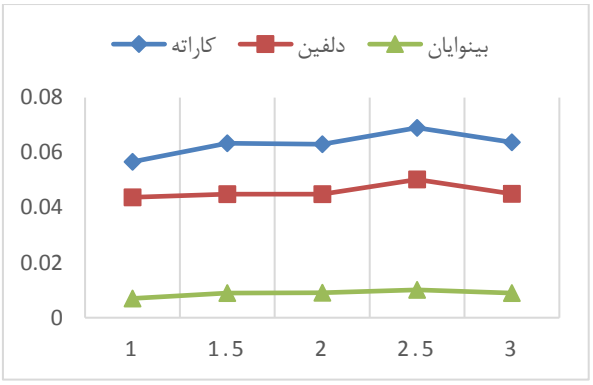
با توجه به شکل ۱ مقادیر عددی کوچک‌تر از ۲.۵ و همچنین مقادیر عددی بزرگ‌تر از ۲.۵ در معیار ماژولاریتی نتایج ضعیف‌تری نسبت به این مقدار دارند. درنتیجه مقدار ۲.۵ به‌عنوان بهترین مقدار در این نسخه در نظر گرفته می‌شود. در آزمایش‌های بعدی این عدد به‌عنوان عددی ثابت در نسخه اول استفاده می‌شود.

۵-۲- آزمایش دوم: مقایسه دو نسخه پیشنهادی بر اساس معیار ماژولاریتی

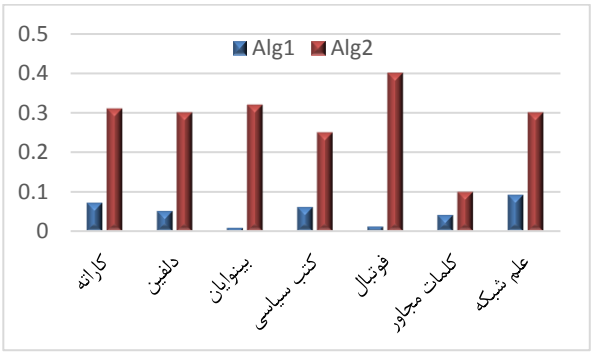
آزمایش دوم با هدف مقایسه دو نسخه پیشنهادی با نام‌های Alg1 و Alg2 بر اساس معیار ماژولاریتی روی ۷ شبکه واقعی طراحی شده است. نتیجه این آزمایش در شکل ۲ آمده است. این شکل نشان می‌دهد که ۱) نسخه اول روی شبکه‌های مختلف پایین‌ترین ماژولاریتی را ایجاد می‌کند و ۲) عملکرد نسخه دوم در تمامی موارد بهتر است. بنابراین در ادامه تنها از نسخه دوم برای مقایسه تعمیم‌های گفته شده استفاده می‌شود.

۵-۳- آزمایش سوم: بررسی تعمیم شعاع دو

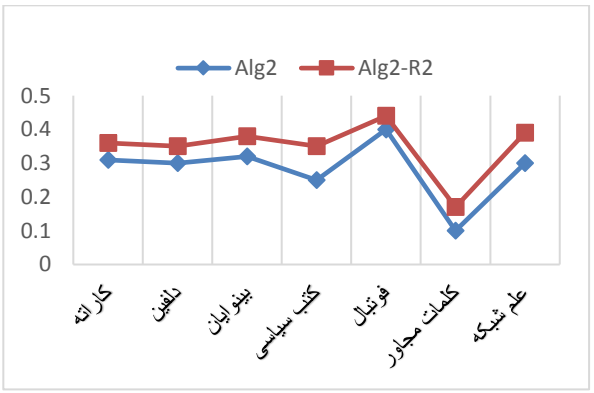
آزمایش سوم با هدف بررسی تعمیم شعاع دو در بهترین نسخه پیشنهادی یعنی Alg2 بر اساس معیار ماژولاریتی روی ۷ مجموعه داده واقعی طراحی شده است. نتیجه این آزمایش در شکل ۳ آمده است. این شکل نشان می‌دهد تعمیم شعاع دوم ماژولاریتی را به شکل قابل توجهی بهبود می‌بخشد. نتایج حاصل از این آزمایش نشان می‌دهد بهره‌گیری از اطلاعات همسایه‌های هر گره تا شعاع دوم می‌تواند بسیار سودمند باشد. در تعمیم و نسخه‌های پیشنهادی به تعمیم دوم بسنده شده است. در ادامه آزمایش‌ها، نسخه دو با تعمیم شعاع دوم (Alg2-R2) به‌عنوان بهترین نسخه پیشنهادی تاکنون استفاده خواهد شد.



شکل ۱: مقایسه ضرایب عددی متفاوت در گام اول نسخه اول بر اساس معیار ماژولاریتی



شکل ۲: مقایسه دو نسخه اصلی بدون تعمیم ها بر اساس معیار ماژولاریتی روی هفت گراف



شکل ۳: مقایسه Alg2 و Alg2-R2 (ترکیب نسخه دوم و تعمیم اول)

۵-۴- آزمایش چهارم: بررسی تعمیم سیگنال سراسری

آزمایش چهارم با هدف بررسی تعمیم سیگنال سراسری در بهترین نسخه پیشنهادی تاکنون (Alg2-R2) بر اساس معیار ماژولاریتی روی سه دیتاست طراحی شده است. نتیجه این آزمایش در شکل ۵ آمده است. برای نمایش بهتر از میانگین تغییرات ماژولاریتی در طول تکرارهای الگوریتم استفاده شده است؛ در هر تکرار میانگین ماژولاریتی تا آن تکرار نشان داده شده است. بدین ترتیب، اعوجاج نمودار از بین می‌رود تا بتوان نتایج گفته شده را به شکل بهتری نمایش داد.

شکل ۴ نشان می‌دهد افزودن تعمیم سیگنال سراسری، منجر به بهبود ماژولاریتی جوامع تشخیص داده می‌شود. همچنین این بهبود در زمان کوتاه‌تر به دست می‌آید. برای مثال در تکرار دویستم از الگوریتم، نتایج Alg2-R2-Global بسیار بهتر از Alg2-R2 است و درنتیجه در زمان کوتاه‌تر، الگوریتم Alg2-R2-Global بهتر عمل می‌کند.

خلاصه میزان بهبود نتایج پس از افزودن هر یک از تعمیم‌های گفته شده را می‌توان در شکل ۵ مشاهده کرد. با توجه به این شکل تعمیم شعاع دوم و تعمیم سیگنال سراسری هر دو منجر به بهبود کیفیت جوامع تشخیص داده شده‌اند. با در نظر گرفتن این نتایج، از این پس در آزمایش‌های بعدی نسخه Alg2-R2-Global به‌عنوان بهترین نسخه پیشنهادی در نظر گرفته می‌شود.

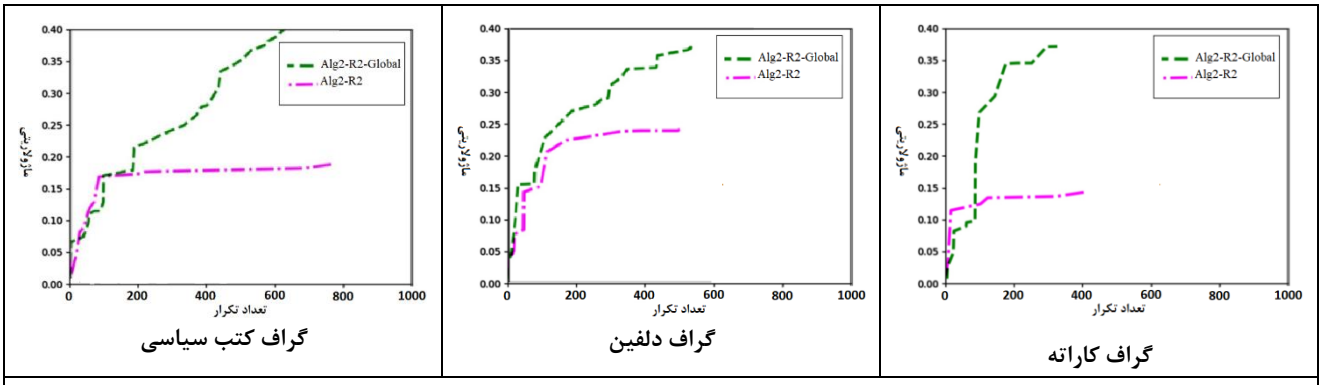
۵-۵- آزمایش پنجم: انتخاب بهترین معیار شباهت

آزمایش پنجم با هدف تعیین بهترین معیار شباهت برای استفاده در بهترین نسخه پیشنهادی تاکنون یعنیAlg2-R2-Global روی ۷ مجموعه داده بر اساس معیار ماژولاریتی طراحی شده است. همان‌طور که در تعریف نسخه‌های پیشنهادی گفته شد از معیارهای شباهت متفاوتی مانند ژاکارد، آدامیک‌آدار و شاخص تخصیص منابع می‌توان استفاده کرد. نتایج این آزمایش در جدول ۳ نمایش داده می‌شود. بر اساس این جدول نتایج حاصل از معیار شباهت ژاکارد و شاخص آدامیک‌آدار بسیار شبیه است و هر دو نسبت به شاخص تخصیص منابع بهتر عمل می‌کنند.

درنتیجه برای تشخیص بهترین معیار شباهت از آزمون-t استفاده شده است. نتایج این آزمون در جدول ۴ آمده است. در این آزمون نتیجه ۱۰ مرتبه اجرای الگوریتم به ازای هرکدام از معیارهای شباهت ژاکارد و شاخص آدامیک‌آدار، برای چهار گراف کاراته، دلفین، بینوایان و کتب سیاسی با یکدیگر مقایسه شده است. نتایج هر چهار آزمون-t نشان می‌دهد که داده‌های حاصل از معیارهای شباهت ژاکارد و شاخص آدامیک‌آدار تفاوت معناداری با یکدیگر ندارند و درنتیجه بالاتر بودن میانگین هرکدام به معنای بهتر بود آن نیست.

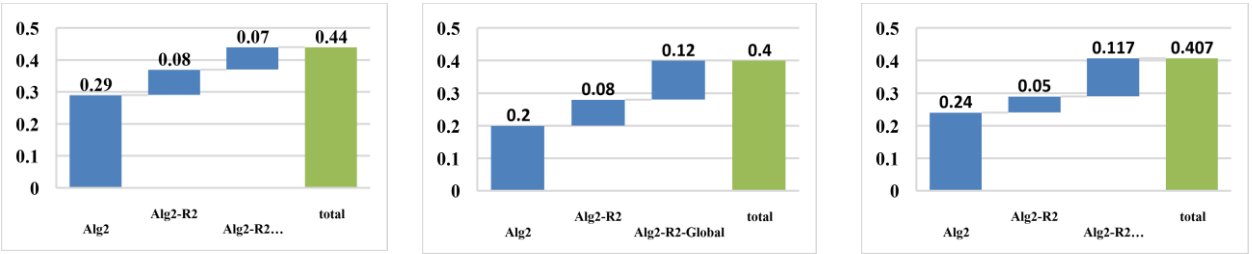
با توجه به نتایج گفته شده هر دو معیار شباهت ژاکارد و شاخص آدامیک‌آدار می‌توانند به‌عنوان بهترین معیار شباهت در نظر گرفته شوند در اینجا از شباهت ژاکارد به‌عنوان معیار شباهت در بهترین نسخه پیشنهادی استفاده می‌شود. با توجه به نتایج حاصل از آزمایش‌های دوم تا پنجم، بهترین نسخه پیشنهادی، نسخه Alg2-R2-Global-J در نظر گرفته می‌شود و در آزمایش‌های بعدی از

این نسخه استفاده خواهد شد.



گراف کاراته | گراف دلفین | گراف کتب سیاسی

شکل ۴: مقایسه ماژولاریتی حاصل از Alg2-R2-Global و Alg2-R2 در گراف‌های کاراته، دلفین و کتب سیاسی



گراف بینوایان | گراف دلفین | گراف کاراته

شکل ۵: نمودار میزان بهبود نتایج ماژولاریتی پس از افزودن هر تعمیم

جدول ۳: نتایج ماژولاریتی حاصل از نسخه Alg2-R2-Global با معیارهای شباهت متفاوت

شبکه	ژاکارد	آدامیک‌آدار	تخصیص منابع
کاراته	0.41	0.37	0.30
دلفین	0.40	0.40	0.29
بینوایان	0.44	0.42	0.35
کتب سیاسی	0.50	0.43	0.33
فوتبال	0.56	0.43	0.30
کلمات مجاور	0.21	0.19	0.14
علم شبکه	0.63	0.64	0.35

جدول ۴: نتایج آزمون t- شباهت ژاکارد (J) و آدامیک‌آدار(AA) ×: اختلاف معناداری ندارند.

	تعداد مشاهدات		مقدار میانگین		مقدار t	two-tail	نتیجه
	J	AA	J	AA			
کاراته	10	10	0.35	0.32	0.99	2.1	×
دلفین	10	10	0.35	0.35	0.001	2.07	×
بینوایان	10	10	0.37	0..0.35	-1.52	2.1	×
کتب سیاسی	10	10	0.37	0.42	-1.35	2.1	×

۵-۶- آزمایش ششم: مقایسه Alg2-R2-Global-J با سایر روش‌ها بر اساس معیار ماژولاریتی

در این آزمایش، بر اساس معیار ماژولاریتی دو سری مقایسه صورت گرفته است: الف) بهترین نسخه پیشنهادی (Alg2-R2-Global-Jaccard) نسبت به سایر روش‌های پیشین روی ۷ شبکه واقعی؛ ب) بهترین نسخه پیشنهادی (Alg2-R2-Global-Jaccard) نسبت به سایر روش‌های پیشین مبتنی بر آتوماتای یادگیر روی ۷ شبکه واقعی.

قسمت الف آزمایش با روشهای سنتی مطرح شده در مراجع [۲۳] و [۲۴] انجام شده است و نتایج آن در جدول ۵ آمده است. قسمت ب آزمایش با روشهای مطرح شده در مراجع [۱۴] و [۱۶] انجام شده است که نتایج آن در جدول ۶ آمده است.

جدول ۵ نشان می‌دهد که روش پیشنهادی نسبت به روش‌های Spectral clustering و Node2vec-SC بهتر عمل کرده است و

نسبت به دو روش دیگر در برخی موارد بهتر و در سایر موارد ضعیف‌تر عمل کرده است. بیشترین ضعف روش پیشنهادی روی گراف علم شبکه است. در مقایسه با سایر گراف‌ها، این شبکه از متوسط درجه کمتری برخوردار است و به عبارتی گراف خلوتی محسوب می‌شود. بدین ترتیب سیگنال محلی مناسبی برای آتوماتاهای یادگیر تولید نمی‌شود و یادگیری مناسبی اتفاق نمی‌افتد. هر چه گراف شلوغتر می‌شود، کارآیی مناسبتری بدست می‌آید.

اما در مقایسه با سایر روشهای مبتنی بر آتوماتای یادگیر، نتایج در جدول ۶ نشان می‌دهد که روش پیشنهادی در مواردی از روش DLACD بهتر عمل کرده است. نسبت به دو روش دیگر در شبکه علم عملکرد ضعیف‌تری دارد که همچنان به نظر می‌رسد در گراف‌های خلوت بدلیل کم بودن همسایه‌های یک گره، اطلاعات قابل قبولی از همسایگان دریافت نمی‌شود و تنها منبع موثر در آموزش سیگنال سراسری است. از آنجا که تمامی آتوماتاها چنین ضعفی دارند، این سیگنال نیز نمی‌تواند تأثیر زیادی داشته باشد.

۵-۷- آزمایش هفتم: مقایسه Alg2-R2-Global-Jaccard با سایر روش‌ها بر اساس معیار NMI

در این بخش با هدف تعیین میزان کارایی بهترین نسخه پیشنهادی (Alg2-R2-Global-J) نسبت به سایر روش‌های پیشین، بر اساس معیار NMI روی ۴ شبکه واقعی، آزمایش طراحی شده است. نتیجه این آزمایش درجدول ۷ آمده است. این جدول نشان می‌دهد که روش پیشنهادی بر اساس معیار NMI بسیار بهتر از روش‌های سنتی عمل می‌کند. روشی که در میان سایر روشها NMI بهتری دارد روش MOCD است که تنها در شبکه فوتبال عملکرد بهتری نسبت به روش پیشنهادی دارد. گراف فوتبال دارای بالاترین مقدار k در میان سایر شبکه‌ها ست. این به آن معناست که تعداد اقدامهای آتوماتا در این شبکه از سایر شبکه‌ها بیشتر است. وقتی تعداد اقدام زیاد باشد، یادگیری نیاز به تعداد تکرار بیشتری دارد؛ زیرا احتمال انجام یک اقدام بسیار کم است و رسیدن احتمال اقدام بهینه به مقدار قطعی (احتمال یک) بسیار طول می‌کشد.

جدول ۵: مقایسه Alg2-R2-Global-J با سایر روش‌های سنتی بر اساس معیار ماژولاریتی

*** مقادیر (-) در جدول نشان دهنده عدم وجود آزمایش یا نتیجه‌ای از روش مورد نظر روی گراف مشخص شده است.**

گراف	CNM [23]	GA-Net [23]	Spectral clustering [24]	Node2vec-SC [24]	Alg2-R2-Global-J
کاراته	0.39	0.41	0.37	0.26	**0.41**
دلفین	0.49	0.47	0.03	0.38	0.40
بینوایان	0.50	0.52	0.30	0.45	0.44
کتب سیاسی	0.50	0.49	0.45	0.49	**0.50**
فوتبال	0.58	0.56	0.45	0.55	0.56
کلمات مجاور	0.29	0.15	0.20	0.20	0.21
علم شبکه	0.95	-	-	-	0.63

جدول ۶: مقایسه Alg2-R2-Global-J با سایر روش‌های بر پایه آتوماتای یادگیر بر اساس معیار ماژولاریتی

گراف	CLA-Net [16]	CLACD [16]	DLACD [14]	Alg2-R2-Global-J
کاراته	0.41	0.40	0.38	0.41
دلفین	0.52	0.52	0.35	0.40
بینوایان	-	-	0.52	0.44
کتب سیاسی	0.52	0.52	0.52	0.50
فوتبال	0.60	0.59	0.58	0.56
کلمات مجاور	-	-	-	0.21
علم شبکه	0.92	0.95	-	0.63

جدول ۷: مقایسه Alg2-R2-Global-J با سایر روش‌ها بر اساس معیار NMI

شبکه	CNM [23]	GA-Net [23]	Infomap [23]	MOCD [23]	Alg2-R2-Global-J
کاراته	0.58	0.53	0.6	0.84	**0.85**
دلفین	0.44	0.29	0.37	0.88	**0.92**
کتب سیاسی	0.51	0.33	0.48	0.58	**0.82**
فوتبال	0.67	0.86	0.92	**0.88**	0.76

۶- نتیجه‌گیری

استفاده از آتوماتای یادگیر در روش‌های کشف جوامع در شبکه‌ها در بسیاری از مقالات دیده شده است. در این مقاله، الگوریتم‌هایی بر پایه آتوماتای یادگیر مطرح شد؛ دو نسخه اصلی به همراه دو تعمیم که در مجموع هفت الگوریتم کشف جامعه پیشنهاد شد. الگوریتمهای ارایه شده دو تفاوت عمده با سایر روشهای مشابه دارند: ۱) استفاده از معیارهای شباهت برای صدور سیگنال‌های پاداش و جریمه و ۲) ثابت در نظر گرفتن تعداد اقدام‌های آتوماتاهای یادگیر (تعداد جوامع هر شبکه). ثابت در نظر گرفتن تعداد اقدام‌ها، منجر به پایین آمده هزینه پیچیدگی و کاهش اندازه فضای جستجو می‌شود. مزایای دیگری برای این ایده در پیوست الف بررسی شده است. الگوریتم‌های پیشنهادی (نسخه‌ها و تعمیم‌ها) با یکدیگر مقایسه شدند تا بهترین نسخه الگوریتم پیشنهادی مشخص شود. سپس نتایج بهترین نسخه پیشنهادی با سایر روش‌های پیشین مورد مقایسه قرار گرفت. نتایج به دست آمده و بررسی دقیق‌تر نتایج از دو معیار ماژولاریتی و NMI کارآمدی الگوریتم را نشان می‌دهد. براساس معیار ماژولاریتی، الگوریتم پیشنهادی در مقایسه با یکی از روش‌های پیشین بهتر عمل کرده و نسبت به سایر روش‌ها عملکرد پایین‌تری دارد. اما بر اساس معیار NMI الگوریتم پیشنهادی از روش‌های سنتی پیشین بهتر عمل کرد.

مراجع

[1] A. Rezvanian, M. Rahmati, and M. R. Meybodi, "Sampling from complex networks using distributed learning automata," *Physica A: Statistical Mechanics and its Applications*, vol. 396. pp. 224–234, 2014.

[2] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47–97, 2002.

[3] R. Rabbany, M. Takaffoli, J. Fagnan, O. R. Zaïane, and R. J. G. B. Campello, "Communities validity: methodical evaluation of community mining algorithms," *Social Network Analysis and Mining*, vol.3, pp.1039-1062, 2013.

[4] B. Krishnamurthy and J. Wang, "On network-aware clustering of Web clients," *Computer Communication Review*, vol.30, pp. 97-110, 2000.

[5] S. Fortunato and M. Barthé, "Resolution limit in community detection," *Proceedings of the National Academy of Sciences of the United States of America*, vol.104, no.1, pp. 36-41, 2007.

[6] M. A. L. Thathachar and P. S. Sastry, "Varieties of learning automata: An overview," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 32, no. 6. pp. 711–722, 2002.

[7] J. Kim and T. Wilhelm, "What is a complex graph?" *Physica A: Statistical Mechanics and its Applications*. vol. 387, no. 11, pp. 2637–2652, 2008.

[8] T. S. Evans, *Complex networks*, Contemporary Physics, vol. 45, no. 6, pp. 455-474, 2004.

[9] S. Forman and B. K. Samanthula, "Secure Similar Document Detection: Optimized Computation Using the Jaccard Coefficient," *IEEE 4th International Conference on Big Data Security on Cloud (BigDataSecurity)*, pp. 1–4, 2018.

[10] C. A. Bliss, M. R. Frank, C. M. Danforth, and P. S. Dodds, "An evolutionary algorithm approach to link prediction in dynamic social networks," *Journal of Computational Science*, vol. 5, no. 5, pp. 750–764, 2014.

[11] S. Fortunato, "Community detection in graphs," *Physics Reports*. vol. 486, no. 3–5, pp. 75–174, 2010.

[12] J. Cheng, M. Leng, L. Li, H. Zhou, and X. Chen, "Active semi-supervised community detection based on must-link and cannot-link constraints," *PLoS ONE*, vol. 9, no. 10. 2014.

[13] M. M. D. Khomami, A. Rezvanian, and M. R. Meybodi, "Distributed learning automata-based algorithm for community detection in complex networks," *International Journal of Modern Physics B*, vol. 30, no. 8. 2016.

[14] M. Ghamgosar, M. M. D. Khomami, N. Bagherpour, and M. Reza, "An extended distributed learning automata based algorithm for solving the community detection problem in social networks," *Iran. Conf. Electr. Eng. ICEE 2017*, pp. 1520–1526, 2017.

[15] M. M. D. Khomami, A. Rezvanian, and M. R. Meybodi, "A new cellular learning automata-based algorithm for community detection in complex social networks", *Journal of Computational Science*. vol. 24, no.1, pp. 413–426, 2018.

[16] M. M. Daliri Khomami, A. Rezvanian, A. M. Saghiri, and M. R. Meybodi, "Utilizing Cellular Learning Automata for Finding Communities in Weighted Networks," *the 6th International Conference on Web Research, ICWR 2020*. pp. 325–329, 2020.

[17] M. M. Daliri Khomami, A. Rezvanian, A. M. Saghiri, and M. R. Meybodi, "Overlapping Community Detection in Social Networks Using Cellular Learning Automata," *the 28th Iran. Conf. Electr. Eng*. pp. 1–6, 2020.

[18] A. Fathinavid, "Multilayer cellular learning automata: A computational model to solve multilayer infrastructure problems with its application in community detection for multilayer networks", *Journal of Computational Science*, vol. 61, p. 101683, 2022.

[19] S. A. H. Minoofam, A. Bastanfard, and M. R. Keyvanpour, "A transfer learning algorithm to improve the convergence rate and accuracy in cellular learning automata", *Nashriyyah-i Muhandisi-i Barq va Muhandisi-i Kampyutar-i Iran*, vol. 88, no. 2, p. 69, 2021.

[20] M. M. D, Khomami, A. Rezvanian, and M. R. Meybodi, "CFIN: A community-based algorithm for finding influential nodes in complex social networks", *Journal of Supercomput*, vol. 77, pp. 2207–2236, 2021.

[21] J. A. Almendral, I. Leyva, D. Li, I. Sendiña-Nadal, S. Havlin, and S. Boccaletti, "Dynamics of overlapping structures in modular networks," Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, vol. 82, no. 1. 2010.

[22] M. Hamann, U. Meyer, M. Penschuck, H. Tran, and D. Wagner, "I/O-efficient generation of massive graphs following the LFR benchmark," ACM Journal of Experimental Algorithmics, vol.23, 2018.

[23] M. E. J. Newman, "Spectral methods for community detection and graph partitioning," Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. vol. 88, no. 4, pp. 042822, 2013.

[24] F. Hu, J. Liu, L. Li, and J. Liang, "Community detection in complex networks using Node2vec with spectral clustering," Phys. A Stat. Mech. its Appl. vol. 545, pp. 123633, 2020.

[25] B. He, L. Gu, and X. D. Zhang, "Nodal domain partition and the number of communities in networks," Journal of Statistical Mechanics: Theory and Experiment, vol. 2012, no. 2. 2012.

[26] P. Wu and L. Pan, "Multi-Objective Community Detection Based on Memetic Algorithm,", *PLoS ONE*, vol.10, no.5 2015.

پیوست الف: بحث مرتبط با روش‌های پیشنهادی

در این پیوست بحث‌های اضافه بر روش‌های پیشنهادی مطرح می‌شود. در ابتدا متغیر $AI_t[i]$ بررسی می‌شود. پس از آن در ارتباط با ماژولاریتی و چالش‌های مرتبط با آن صحبت می‌شود. در انتها دلایل ثابت بودن k (تعداد جوامع) در روش پیشنهادی آورده شده است.

الف-۱: بحث روی متغیر $AI_t[i]$

در نسخه دو و تعمیم‌های گفته شده از متغیر $AI_t[i]$ استفاده شده است. این متغیر برای تولید سیگنال پاداش/جریمه برای آتوماتای LA_i استفاده می‌شود. این متغیر در نسخه‌های الگوریتم پیشنهادی، در طول تکرارهای الگوریتم، حاصل جمعی از مقدارهای مشابهت گره‌های همسایه را نگهداری می‌کند. با نگهداری این مقدارها و محاسبه متوسط آن‌ها (مشابه گام دوم نسخه‌های پیشنهادی) می‌توان بهبود جوامع شناسایی شده در هر مرحله را بررسی کرد. اگر $AI_t[i] \geq \frac{\sum_{t=0}^{t-1} AI_t[i]}{t-1}$ برقرار باشد به این معنی است که در مرحله t ام تعداد همسایه‌های هم برچسب گره i (که احتمالاً میزان مشابهت آن‌ها با گره i نیز زیاد است) بیشتر از مراحل قبلی شده است. درنتیجه با احتمال بیشتری جامعه گره i به‌درستی شناسایی شده است. پس به کمک این متغیر، سیگنال پاداش برای آتوماتای LA_i صادر می‌شود.

الف-۲: بحث پیرامون ماژولاریتی

در فرآیند تشخیص جامعه هر چه مقدار ماژولاریتی بیشتر باشد ساختار جوامع شبکه بهتر هستند. به همین دلیل، تعمیم با استفاده از سیگنال سراسری در روش‌های پیشنهادی مطرح شده است (در این تعمیم یکی از شرط‌های تولید سیگنال پاداش برای آتوماتای LA_i ، بیشتر بودن ماژولاریتی مرحله t ام نسبت به مرحله $(t-1)$ ام است). یعنی اگر ساختار جوامع شناسایی شده در مرحله tام بهبودیافته باشند آنگاه سیگنال پاداش صادر می‌شود. اما نکته موجود در رابطه با محاسبه ماژولاریتی هزینه زمانی آن است. محاسبه ماژولاریتی زمان قابل توجهی از اجرای الگوریتم را به خود اختصاص می‌دهد [۲۵]. با توجه به این موضوع به دو دلیل در روش پیشنهادی از این تعمیم استفاده شده است:

- در مقابل افزایش زمان اجرا الگوریتم، نتایج حاصل از آن به شکل چشمگیری بهبود پیدا می‌کنند (در بخش آزمایش‌ها، آزمایش‌هایی باهدف بررسی تأثیر تعمیم مبتنی بر ماژولاریتی طراحی شده است).
- در سایر مقالات کشف جوامع بر پایه آتوماتای یادگیر مانند مقاله [۱۳] نیز به همین شکل از ماژولاریتی در محاسبه سیگنال پاداش/جریمه استفاده شده است.

الف-۳: دلایل ثابت در نظر گرفتن k (تعداد جوامع)

درروش پیشنهادی (تمامی نسخه‌ها و تعمیم‌ها) تعداد جوامع (k) ثابت است و در ابتدای الگوریتم به‌عنوان ورودی به آن داده می‌شود. سه دلیل برای ثابت بودن k می‌توان ارائه داد:

- مشابه با روش پیشنهادی این تحقیق، در روش‌های خوشه‌بندی k_means نیز k عددی ثابت است و کارایی این روش‌ها مشخص است.
- با ثابت در نظر گرفتن k همچنین می‌توان از این الگوریتم برای حل مسائل طبقه‌بندی گره‌ها۱۶ نیز بهره برد. در این دسته از مسائل تعداد برچسب گره‌ها ثابت است و به هر میزان دو گره به یکدیگر نزدیک باشند برچسب مشابهی دریافت می‌کنند. درنتیجه با توجه به ساختار گفته شده برای الگوریتم پیشنهادی به سادگی می‌توان برای حل این دسته از مسائل نیز از آن استفاده کرد.
- با ثابت در نظر گرفتن k در آینده می‌توان مدل نگاشت-کاهش ارائه کرد که در آن k به‌عنوان کلید است و بدین ترتیب امکان پردازش روی گراف‌های عظیم فراهم شود و بتوان به‌صورت موازی و با سرعت بالا این الگوریتم را برای گراف‌های بسیار بزرگ پیاده‌سازی کرد. پیاده‌سازی این مدل خارج از بحث این مقاله است و در کارهای آینده پیشنهاد می‌شود.

پیوست ب: آزمایش‌ها با داده‌های ساختگی

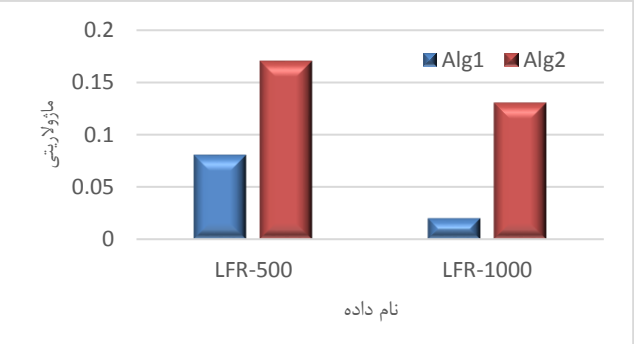
داده‌های ساختگی استفاده شده در آزمایش‌ها شامل دو مجموعه داده است که توسط تابع LFR_benchmark_graph از کتابخانه Networkx.generators.community ساخته شده‌اند. LFR یک مولد گراف است که برای ارزیابی الگوریتم‌های تشخیص جامعه استفاده می‌شود. تابع LFR، گراف‌های آزمایشی (گراف‌های ساختگی شبیه گراف‌های دنیای واقعی) ایجاد می‌کند [۲۶] .

ورودی تابع LFR_benchmark_graph شامل تعداد گره‌ها در گراف ایجاد شده، توزیع توانی برای توزیع درجات گراف ایجاد شده (Tau1)، توزیع توانی برای توزیع اندازه جوامع گراف ایجاد شده (Tau2)، حداقل و حداکثر درجه گره‌ها در گراف ایجاد شده، حداقل و حداکثر اندازه جوامع ایجاد شده در گراف و... است و خروجی آن یک گراف با تعدادی جامعه است. در این قسمت تحقیق دو گراف ساختگی با مشخصاتی که در جدول ب-۱ آمده استفاده شده است.

جدول ب-۱: مجموعه داده‌های ساختگی مورد استفاده در آزمایش‌ها

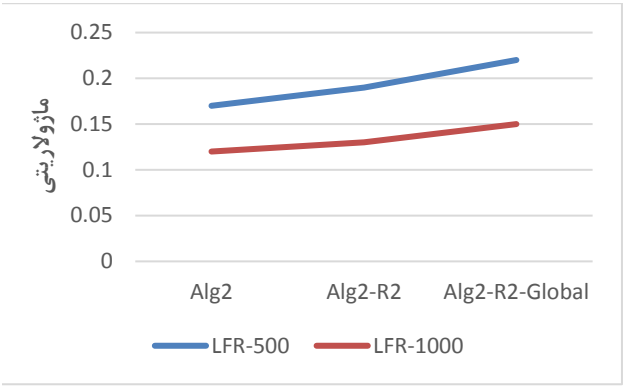
گراف	تعداد گره	Tau1	Tau2	حداقل درجه	حداکثر درجه	حداقل اندازه جوامع	حداکثر اندازه جوامع
۱	500	3	1.1	10	40	150	200
۲	1000	2	1.1	20	50	350	600

آزمایش اول به مقایسه الگوریتم پیشنهادی اول و دوم روی داده ساختگی می‌پردازد که نتیجه آن در شکل ب-۱ آمده است. همانند داده واقعی، کارآیی الگوریتم پیشنهادی دوم از الگوریتم پیشنهادی اول بهتر است.



شکل ب-۱: مقایسه دو الگوریتم پیشنهادی با معیار ماژولاریتی

آزمایش دوم به دنبال یافتن تأثیر دو تعمیم روی داده‌های ساختگی است که در شکل ب-۲ نتایج قابل مشاهده است. تأثیر مثبت هر دو تعمیم در این شکل به وضوح دیده می‌شود که همان نتایج داده‌های واقعی را تأیید می‌کند.



شکل ب-۲: مقایسه ماژولاریتی با اجرای الگوریتمهای Alg2 **(نسخه پایه)، ترکیب با تعمیم شعاع دو** Alg2-R2 **و ترکیب هر دو تعمیم** Alg2-R2-Global

پاورقی‌ها:

1 Community detection
2 Complex networks
3 Reward Parameter
4 Punish Parameter
5 Cellular Learning Automata
6 Irregular Cellular Learning Automata
7 Similarity Measure
8 Jaccard similarity
9 Adamic/Adar Index
10 Common neighbors
11 Modularity
12 Multi-layer network
13 Transfer Learning
14 Influence Maximization
۱۵ مقدار ضریب h در بخش آزمایش‌ها مورد بحث قرار می‌گیرد.
16 Node Classification