

الگوریتم ژنتیک جدید سازگار برای تشخیص ساختار جامعه

چکیده

ساختارهای اجتماعی در شبکه‌ای وجود دارد که دارای پیچیدگی‌های زیستی، اجتماعی، تکنولوژیکی و حاوی اطلاعات مهم باشد. ساختارهای شبکه و جامعه در سیستم‌های کامپیوتری به ترتیب توسط نمودارها و زیرگراف‌ها نمایش داده می‌شود. مسئله تشخیص ساختار جامعه یک مسئله NP سخت است، به ویژه نتایج نهایی بهترین ساختارهای اجتماعی برای شبکه‌های بزرگ پیچیده ناشناخته هستند. در این مقاله، برای حل مسئله تشخیص ساختار جامعه یک الگوریتم مبتنی بر الگوریتم ژنتیک، AGA-net که یکی از تکنیک‌های تکاملی است پیشنهاد شده است. این الگوریتم دارای ویژگی همگرایی سریعی به مقدار بهینه بدون اینکه در بهینه محلی به دام بیافتد است بنابراین توسط پارامترهای جدید پشتیبانی شده است. شبکه دنیای واقعی که اغلب در کارهای گذشته استفاده می‌شود به‌عنوان داده‌های آزمایشی مورد استفاده قرار گرفته و نتایج به دست آمده با 10 الگوریتم متفاوت مقایسه شده است. پس از تجزیه و تحلیل نتایج آزمون مشاهده شده است که الگوریتم پیشنهادی نتایج خوبی در مورد شبکه‌های پیچیده ارائه می‌دهد.

کلیدواژه‌ها: بهینه‌سازی ترکیبی، تشخیص ساختار جامعه، شبکه‌های پیچیده، محاسبات تکاملی، الگوریتم ژنتیک،

ماژولار بودن

1. معرفی

درک شبکه‌ها اطلاعات مهمی درباره استخراج اطلاعات معنی‌دار از سیستم‌های پیچیده ارائه می‌دهد. در بیان معنی‌دار اطلاعات از این شبکه‌ها، اهمیت ساختارهایی که با عنوان ساختارهای اجتماعی نامگذاری شده‌اند، زیاد است. ساختارهای گراف برای ارائه شبکه‌های جهان واقعی استفاده می‌شوند. ساختارهای اجتماعی یا خوشه‌ها می‌تواند به‌عنوان زیرگراف محسوب شود که در ساختارهای گراف به طور جزئی یا کاملاً مستقل از یکدیگر هستند. به‌عنوان مثال، بافت یا اندام‌هایی که در بدن انسان نقش مشابه‌ای دارند به‌عنوان خوشه‌ها در نظر گرفته می‌شوند [1]. تشخیص ساختار جامعه (CSD) برای درک شبکه‌های زیست‌شناسی، اقتصادی، اجتماعی، فن‌آوری و غیره مهم است. این شبکه‌ها می‌توانند شبکه‌های مصنوعی یا دنیای واقعی باشند. در دنیای واقعی شبکه‌ها می‌توانند از نمونه‌هایی مانند شبکه‌های ساختار اقتصادی [2]، شبکه‌های غذایی [3]، شبکه‌های تعامل شیمیایی بین پروتئین‌ها و مولکول‌ها در سلول‌ها [4-6] و شبکه‌های اجتماعی مانند شبکه‌های تعیین دوستی در گروه‌ها، شبکه‌های تحلیل رابطه و شبکه‌های تشخیص حملات تروریستی باشند [7].

اشیاء و اتصالات در شبکه‌ها به ترتیب با گره‌ها و لبه‌ها ارائه می‌شوند. ساختارهای گراف که برای نشان دادن شبکه‌های داده شده مورد استفاده قرار می‌گیرند به‌عنوان ساده‌ترین شکل از شبکه‌های نامنظم نامیده می‌شوند [8]. مسئله کاوش جامعه (CMP) به کشف زیرگراف معنی‌دار در بسیاری از داده‌های شبکه‌های پیچیده اشاره دارد [9]. در این مقاله، بسیاری از داده‌های دنیای واقعی توسط CMP تجزیه و تحلیل شده و نتایج به دست آمده در بخش نتایج تجربی نشان داده شده است.

روش‌های متعددی برای تشخیص ساختارهای جامعه در شبکه‌های پیچیده کشور ایجاد شده است. این روش‌ها براساس بسیاری از ویژگی‌ها که در عین تعریف در مورد به دست آوردن بهترین نتیجه به دست نمی‌آیند، نتایج موفقیت‌آمیزی به همراه داشته‌اند. کارآیی الگوریتم‌های کارهای پیشین در شبکه‌های بزرگ بسیار پایین است. علاوه براین، برای تشخیص جامعه، الگوریتم‌های بسیاری نیاز به دانش قبلی مانند تعداد جوامع دارند. گروه‌بندی مطلوب در شبکه یک

مشکل بسیار دشوار است. بنابراین، مسئله CSD یک مسئله سخت در زمان چندجمله‌های غیردقیق (NP-H) است [10، 11].

معمولا برای حل مسائل پیچیده مانند CSD دو روش متفاوت پیشنهاد می‌شود. این روش‌ها دقیق و (متا) اکتشافی هستند [12]. از این دو روش، روش (متا) اکتشافی می‌تواند راه‌حل‌های راحت‌تری را برای مسائل پیچیده از روش‌های دقیق فراهم آورد. الگوریتم‌هایی مانند ممتیک و ژنتیک با روش‌های متا اکتشافی تحت پوشش قرار می‌گیرند و به عنوان الگوریتم‌های الهام گرفته از زیست‌شناسی شناخته می‌شوند [12]. این الگوریتم‌ها از معیارهای مختلف جامعه بنابه روش‌های خود در مسائل CSD استفاده می‌کنند. رایج‌ترین معیار اندازه‌گیری که اخیرا توسط گیرمان و نیومن استفاده شده است، معیار Q modularity است [13، 14].

الگوریتم شناخته شده‌ای که وجود دارد، الگوریتم تشخیص جامعه Newman-Girman (GN) است [13، 15]. الگوریتم Fast Newman (FN) یک الگوریتم براساس حداکثر مازولار بودن Q [16] است. به‌طورمشابه، الگوریتم دیگری براساس حداکثر مازول Q است که، الگوریتم سریع انحلال (fast unfolding) نامیده می‌شود [17]. علاوه براین الگوریتم‌هایی مانند [18] Eigenvectors, Random Walks [19]، انتشار علامت (LP) [20] روش Spin Glass Type Potts [21] و الگوریتم LTE (گسترش انسجام محلی) [22] در کارهای پیشین استفاده شده است. FN [16]، الگوریتم تشخیص جامعه برای شبکه‌های بزرگ است که توسط Clauseset و همکارانش ارائه شده است [23]، الگوریتم بهینه‌سازی فوق‌العاده [24] و الگوریتم‌های دیگری مانند این‌ها دارای پیچیدگی $O(e^3)$ از نظر پیچیدگی زمانی هستند. در اینجا e به تعداد لبه‌ها اشاره دارد [25].

پیچیدگی زمانی در حجم بزرگ مانند اندازه شبکه افزایش می‌یابد. در شبکه‌های کوچک یا منظم، تشخیص جامعه می‌تواند به راحتی با الگوریتم‌های ارائه شده در بالا انجام شود، اما هنگامی که اندازه شبکه افزایش می‌یابد، الگوریتم‌های موجود از لحاظ عملکرد و موفقیت کافی نیستند. همچنین هنگامی که ورود دانش پیشین به این الگوریتم‌ها اجباری باشد، کشف الگوریتم جدید و کارآمد اجتناب‌ناپذیر است. با توجه به این نیاز، مسائل CSD برای حل با الگوریتم‌هایی مانند الگوریتم ژنتیک، الگوریتم بهینه‌سازی ذرات، الگوریتم بهینه‌سازی کلونی مورچه، الگوریتم‌های ممتیکی مورد

آزمایش قرار گرفته‌اند. در این مقاله الگوریتم ژنتیک که یکی از الگوریتم‌های فوق است ساختار اصلی الگوریتم پیشنهادی را تشکیل می‌دهد. در حال حاضر الگوریتم ژنتیک از نظر محاسبات، پیچیدگی زمان و همگرایی در مسائل NP سخت بسیار موفق عمل کرده است و تقریباً در اکثر مسائل در کارهای پیشین مورد استفاده قرار گرفته است. برای اولین بار Tasgin و همکارانش [26] از الگوریتم ژنتیک در مسئله CSD استفاده کردند. روشی که آنها توسعه دادند با عنوان GATB (یا GATHB [27] نامگذاری شده است) [25]. بعد از آن شروع به استفاده‌ی گسترده از الگوریتم ژنتیک در مسائل CSD شد. به‌طور خاص، برای به‌دست آوردن مقدار مطلوب Q به روشی ارزان‌تر نیاز است که، روش‌های بسیاری با چندین تغییر در روش ژنتیک توسعه یافته، مانند جهش (mutation)، ترکیب (crossover)، انتخاب و غیره این کار را انجام داده‌اند.

Shi و همکارانش برای حل مسئله CSD با استفاده از الگوریتم ژنتیک مبتنی بر الگوریتم GACD [28] تلاش کردند. آنها روش خود را با شبکه‌های دنیای واقعی که در کارهای گذشته بسیار زیادی استفاده شده بود آزمایش کردند و نتایج را با GN [13]، GN Fast [23] و GATB [25] مقایسه کردند. هنگامی که نتایج به دست آمده تجزیه و تحلیل شد، دریافتند که الگوریتم‌های مبتنی بر الگوریتم ژنتیک مانند GACD و GATB برای حل مسائل CSD کاملاً موثر هستند. روش‌های توسعه یافته مبتنی بر الگوریتم ژنتیک مزایای بسیاری دارد. به عنوان مثال GATB [25] دارای پیچیدگی زمانی $O(e)$ است و هیچ دانش قبلی برای CSD نیاز ندارد. در این مقاله، یک رویکرد جدید مبتنی بر الگوریتم ژنتیک ارائه شده است که بانام AGA-net نامگذاری شده است. AGA-net دارای پیچیدگی زمان $O(e)$ است و نیاز به دانش قبلی ندارد. الگوریتم پیشنهاد شده مبتنی بر طراحی سازگاری از الگوریتم ژنتیک برای دستیابی به مناسب‌ترین راه‌حل در زمان کمتری برای مسئله CSD است. AGA-net در شبکه‌های داده شده در بخش 3 مورد آزمایش قرار گرفته و نتایج بدست آمده با برخی از الگوریتم‌های موجود در کارهای پیشین مقایسه شد (بخش 3 را ببینید).

1.1 تشخیص ساختار جامعه

هنگامی که شبکه داده شده با ساختار گراف، ساختار جامعه را به دست آورد می‌تواند به عنوان زیرگرافی محسوب شود که دارای کیفیت یا کمیتی مانند حداکثر ویژگی مشترک در خود، تعداد تعاملات، شباهت‌های موقعیتی و غیره است. گره‌هایی که عناصر این ساختارها هستند باید حداکثر تعامل و خواص مشترک را با گره‌های جامعه خود و تعامل کمتر و خواص مشترک را با گره‌های جامعه دیگر داشته باشند. گروهی از مردم که در محیط اجتماعی دارای ارتباطات قوی در شبکه‌های زیست محیطی هستند با یکدیگر و خوشه‌ای از کامپیوترها حداکثر همکاری و تبادل اطلاعات را دارند. فرض کنید که ساختار گراف $G(V,E)$ نشان‌دهنده یک شبکه نامنظم و بدون وزن باشد. گراف G دارای V گره (رأس) و مجموعه‌ای از لبه‌ها E (یال‌ها) است.

$$V = \{v_i | i = 1, 2, 3, \dots, n\} \text{ and } E = \{e_j | j = 1, 2, 3, \dots, m\}$$

به طوری که i, j, n و m به ترتیب نشان‌دهنده اندیس گره، اندیس لبه یا یال، تعداد گره و تعداد لبه‌ها هستند. فرض کنید ماتریس همجواری (adjacency) را با Adj و با اندازه $n \times n$ تعریف کنیم و فرض کنید ماتریس Adj ارتباط عناصر مجموعه V توسط عناصر از مجموعه E را نشان می‌دهد. ماتریس همجواری توسط معادله (1) تولید می‌شود [23].

$$Adj = \begin{cases} 1 & \text{if } i \text{ and } j \text{ nodes are connected,} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

مدولاسیون Q برای گراف G در معادله (2) داده شده است. تابع برازش توسط نیومن و گیروان در مقاله‌ای با نام Finding and Evaluating Community Structure in Network پیشنهاد شده است [15].

$$Q = \frac{1}{2 \times m} \sum_{ij} \left(Adj_{(i,j)} - \frac{k_i \times k_j}{2 \times m} \right) \times \delta(C_i, C_j) \quad (2)$$

به طوری که Q با عنوان modularity Q نامیده می‌شود و تابع هدف، حداکثر کردن است. $adj_{(i,j)}$ نشان‌دهنده ماتریس همجواری گراف داده شده است. m تعداد کل لبه در شبکه را نشان می‌دهد و توسط معادله (3) محاسبه می‌شود. k_i نشان‌دهنده درجه i^{th} گره، k_j نشان‌دهنده درجه گره j است و مثالی از k_i را می‌توان با معادله (4) محاسبه کرد.

C_i و C_j نشان‌دهنده i^{th} و j^{th} گره جامعه هستند. $\delta(C_i, C_j)$ یک تابع است که نشان‌دهنده‌ی i امین و j امین گره موجود در یک جامعه مشابه است. تابع $\delta(C_i, C_j)$ توسط معادله (5) محاسبه می‌شود.

$$m = \frac{1}{2} \sum_{ij} Adj_{(i,j)} \quad (3)$$

$$k_i = \sum_j Adj_{(i,j)} \quad (4)$$

$$\delta = \begin{cases} 1 & \text{if } C_i = C_j \\ 0 & \text{if } C_i \neq C_j \end{cases} \quad (5)$$

تشخیص ساختار جامعه با توجه به مقدار برآزش با استفاده از Tasgin و همکارانش در سال 2007 بانام تشخیص جامعه در شبکه‌های مجتمع با استفاده از الگوریتم‌های ژنتیک انجام شده است [26]. در مقاله مشخص شده، الگوریتم پیشنهاد شده بانام GATHB [26، 27] نامگذاری شده است. پس از انتشار این مقاله، الگوریتم‌های تکاملی بسیاری برای مسائل CSD اعمال شد. الگوریتم AGA-net که ما پیشنهاد دادیم از تابع هدف یکسانی مانند الگوریتم GATHB استفاده کرده است که در معادله (2) نشان داده شده است.

1.2 الگوریتم ژنتیک

GA ها برای اولین بار توسط جان هلند در دهه 1960 نوشته شده و سپس توسط هلند و دانش‌جویان و همکارانش در دانشگاه میشیگان در دهه 1960 و 1970 توسعه یافته است. هدف هلند این بود که پدیده "سازگاری" را که در طبیعت اتفاق می‌افتد درک کند و روش‌هایی را که مکانیسم‌های سازگاری طبیعی را توسعه می‌دهند، توسعه دهد. کتاب اقتباس از طبیعت و سیستم‌های مصنوعی (هلند، 1975) GA را به عنوان انتزاعی از تکامل بیولوژیکی و چارچوب نظری برای انطباق تحت GA ارائه داده است [29]. GA یک الگوریتم مبتنی بر جمعیت است و می‌تواند بدون نیاز به هیچ مدلی یا دانش قبلی یا مفروضاتی عمل کند. بنابراین این الگوریتم را می‌توان برای مسائل و ساختار کلی با ویژگی خاص اعمال کرد.

2. الگوریتم پیشنهادی

در این مقاله برای حل مشکل CSD، الگوریتم AGA-net پیشنهاد شده است که براساس الگوریتم ژنتیک است. همانطور که هر گره از شبکه در مسئله CSD تعداد محدودی همسایه دارد احتمال اینکه همسایه انتخاب شده دوباره انتخاب شود بسیار بالا است. این وضعیت، جستجو برای بهترین راه حل ممکن است و باعث ورود به چرخه حیرت‌انگیز در شبکه‌های مختلف می‌شود. این مشکل با کمک اپراتورهای ژنتیکی که در بخش 2.4 ارائه شده قابل حل است. بنابراین برای همگرایی سریع به بهترین راه حل، راه‌حل‌های بهتر توسط نخبه‌گرایی انتخاب شده و از طریق مکانیزم ترکیب و جهش مهار می‌شوند. الگوریتم پیشنهاد شده دارای ویژگی همگرایی بهترین مدولار Q بدون اینکه در بهینه محلی قرار داشته باشد است. AGA-net دارای پیچیدگی زمانی خطی است. علاوه بر پارامترهای اساسی و عملگرهای الگوریتم ژنتیک استاندارد، تغییرات خاصی برای مسائل CSD و پارامترهای جدید انجام گرفته است. الگوریتم پیشنهادی با عنوان الگوریتم ژنتیک سازگار (AGA-net) نامگذاری شده است. اصطلاح تطبیقی که در اینجا استفاده می‌شود نشان می‌دهد که هر مکانیسم الگوریتم را می‌توان به تمام شبکه‌ها منطبق کرد. الگوریتم پیشنهاد شده می‌تواند بر روی تمام شبکه‌ها در CSD بدون وابستگی به هر داده داخلی یا خارجی با پارامترهای خاص جدید اعمال شود. مراحل الگوریتم پیشنهاد شده در زیر بخش‌های جداگانه بیان شده است.

2.1 نحوه نمایش ژنتیک

الگوریتم پیشنهادی از ساختار همجواری براساس مکان (LAR) برای نمایش براساس نمودار استفاده می‌کند [30]. هر ژن در کروموزوم دارای دو نوع اطلاعات متفاوت است (communityID و populationID). اطلاعات در مورد آنها در شکل 1 بیان شده است. اولین اطلاعات گره همسایه از گره i^{th} انتخاب شده از بین همسایه‌ها را به‌طور تصادفی ذخیره می‌کند. اطلاعات دوم اطلاعات دانش جامعه (CommunityID) از گره i^{th} برای جوامع تولید شده توسط اولین اطلاعات است. یک نمونه از شبکه با 8 گره در شکل 1 (الف) داده شده است، شکل 1 (ب) یک نمونه از

کروموزوم‌های تولید شده براساس شبکه داده شده و شکل 1 (ج) ساختارهای اجتماعی تولید شده از اطلاعات کروموزوم داده شده را نشان می‌دهد. ساختارهای جامعه به دست آمده در رنگ‌های مختلف داده شده‌اند.

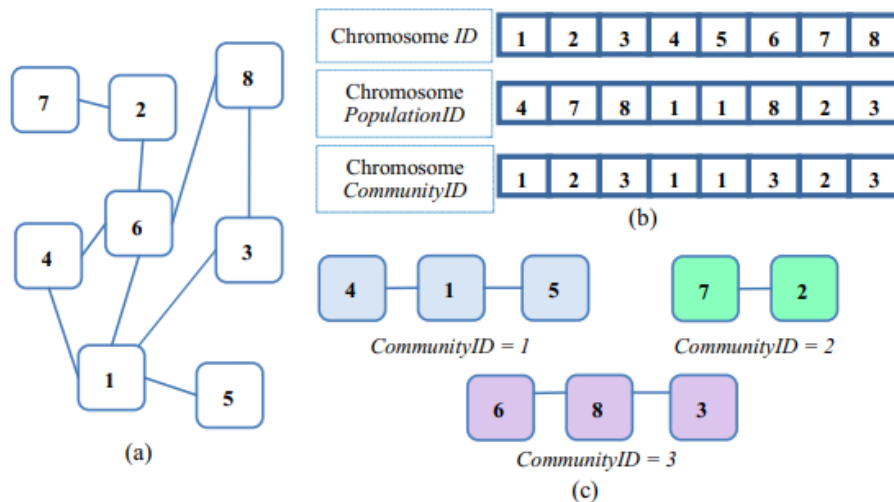


Fig. 1 Examples of a network with 8 node, a chromosome and obtained community structures

کروموزوم داده شده در شکل 1 (ب) توسط 3 آرایه مختلف نشان داده شده است که اطلاعات مربوط به ID، populationID و CommunityID را نگه می‌دارد. آرایه اول شماره توالی گره، آرایه دوم گره‌های همسایه انتخاب شده و آرایه سوم اطلاعات مربوط به جامعه گره را نگه می‌دارد.

2.2 جمعیت اولیه

الگوریتم پیشنهاد شده کروموزوم را مانند اندازه جمعیت در ابتدای فرآیند تولید می‌کند. هر ژن در کروموزوم یک گره را نشان می‌دهد. دومین آرایه از آرایه‌های داده شده در شکل 1 (ب) گره‌ای متناظر با همسایه را براساس ID آن انتخاب می‌کند. پس از آنکه تمام جمعیت تشکیل شد، آرایه 3 با توجه به آرایه 2 که در شکل داده شده است شکل می‌گیرد. آرایه دوم که در شکل 1 (ب) نشان داده شده است کل جامعه را در طول محاسبات Q modularity فراهم می‌کند. درحالی‌که به هنگام تعیین CommunityID ژن در داخل کروموزوم باید با ژن موجود همسایه باشد. براساس این اصل، فضای راه‌حل محدود می‌شود و موجب صرفه‌جویی در زمان می‌شود.

2.3 تابع برازندگی

در این مقاله، Q modularity به عنوان تابع برازندگی استفاده می‌شود. این معیار ابتدا توسط نیومن و گیروان [15] در سال 2004 استفاده شد. این تابع در معادله 2 داده شده است. مسئله CSD را می‌توان به عنوان مسئله بهینه‌سازی ترکیبی با توجه به تابع هدف مشخص شده در نظر گرفت. تابع هدف در بهترین خوشه‌بندی گراف، به حداکثر مقدار Q می‌رسد. مقدار Q در محدوده -1 تا $+1$ متغیر است.

2.4 عملگرهای ژنتیک

نخبه‌کشی (elitism)، انتخاب، ترکیب (crossover) و جهش (mutation) در الگوریتم پیشنهادی مورد استفاده قرار گرفته است. پارامترهای هر عملگر که در این فرایند مورد استفاده قرار گرفته با مسئله CSD برای رسیدن به مناسب‌ترین راه‌حل سازگار است. برخلاف الگوریتم ژنتیک استاندارد جدید، پارامترها در عملگرهای نخبه‌کشی، ترکیب و جهش گنجانده شده‌اند. عملگرها و پارامترهای پیشنهاد شده توسط الگوریتم AGA-net با جزئیات در زیر ارائه شده است.

نخبه‌کشی (elitism): این عملگر در دو مرحله از الگوریتم استفاده می‌شود. در مرحله اول، برای انتقال کروموزوم با نرخ elitismRate (%) انتخاب می‌شود که بهترین مقدار Q در جمعیت را به نسل بعدی به همراه دارد. در مرحله دوم، کروموزوم‌های جدید با مقادیر Q بهتر با کروموزوم بد جایگزین می‌شوند. نرخ elitism از حذف بدترین کروموزوم از خوشه را تضمین می‌کند. این پارامتر با نرخ‌های کوچک برای کاهش تنوع انتخاب کروموزوم مناسب نیست.

انتخاب (selection): فرآیند تولید نسل جدید با انتخاب چرخ رولت (RWS) انجام می‌شود [31]. در الگوریتم پیشنهادی، فرآیند انتخاب براساس روش RWS به شرح زیر انجام می‌شود.

• مقدار برازندگی هر کروموزوم محاسبه شده و مجموع تمام مقادیر برازندگی کروموزومها در جمعیت توسط معادله (6) محاسبه می‌شود.

$$TfitnessQ = \sum_{t=1}^{popSize} fitnessQ_t \quad (6)$$

• احتمال انتخاب هر کروموزوم با معادله (7) محاسبه می‌شود.

$$P_t = fitnessQ_t / TfitnessQ \quad (7)$$

• مجموع تجمعی برای هر کروموزوم محاسبه می‌شود و احتمال تجمع با معادله (8) تعیین می‌شود.

$$Q_t = \sum_{k=1}^t P_k \quad (8)$$

• اعداد تصادفی بین 0 و 1 تولید می‌شود. یک کروموزوم با توجه به محدوده‌ی Q_t اعداد تولید شده انتخاب می‌شود. بنابراین، کروموزومها باید برای انتقال به نسل بعدی باید انتخاب شوند.

ترکیب (crossover): دو پارامتر مختلف مرتبط با این عملگر توسط نامهای نرخ ترکیب (CR) و انتخاب ترکیب (CC) تعریف شده است. پارامترهای CR موجب ترکیب افراد در جمعیت قرار می‌شوند و تعداد تکرار فرآیند مورد بررسی قرار می‌گیرد. سپس پارامتر CC موجب تولید دنباله کنترل تغییر برای جفت کروموزومها تحت فرآیند ترکیب می‌گردد. به طوری که مقدار CC اگر کوچکتر از عدد تصادفی تولید شده باشد، 0 می‌گردد و اگر بزرگتر باشد مقدار 1 را می‌گیرد. فرآیند ترکیب رخ داده در شکل 2 نشان داده شده است.

Chromosome-1	4	7	8	1	1	8	2	3
Chromosome-2	5	6	1	6	1	4	2	6
Change control array	1	0	0	1	1	1	0	1
New chromosome	5	7	8	6	1	4	2	6

Fig. 2 Crossover operation of AGA-net algorithm

جهش (Mutation): در الگوریتم پیشنهاد شده، فرآیند جهش در دو مورد انجام می‌شود، یک نقطه‌ای و چند نقطه‌ای. اولی جهش تک‌نقطه‌ای و دومی جهش چندنقطه‌ای است. همچنین دو پارامتر در روند جهش استفاده شده است. اولین پارامتر نرخ جهش (MR) و پارامتر دوم نرخ جهش چندنقطه‌ای (multiP) است. پارامتر MR با نسبت کوچک انتخاب شده است و تعیین خواهد کرد که آیا کروموزوم ورودی جهش می‌یابد یا نه. پارامتر multiP اجازه می‌دهد تا یکی از گزینه‌های جهش تک‌نقطه‌ای یا چندنقطه‌ای انتخاب شود. اگر مقدار این پارامتر کمتر از مقدار تولید شده به‌طور تصادفی باشد، جهش تک‌نقطه‌ای انتخاب می‌شود، اما اگر برابر یا بزرگتر از مقدار تولید شده به‌طور تصادفی باشد جهش چندنقطه‌ای اعمال می‌شود. نمونه‌های نمایشی نشان‌دهنده این روند در شکل 3 نشان داده شده است. در اینجا، هر ژن انتخاب شده توسط شرایط همسایگی، جهش یافته است (به شکل 1 الف) مراجعه شود).

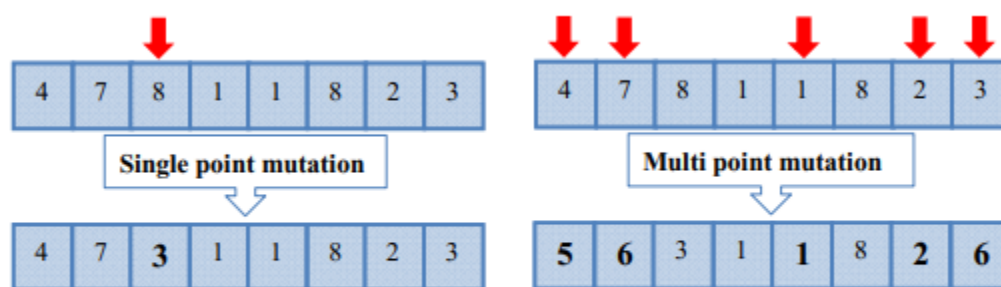


Fig. 3 Single-point and multi-point mutations

3. نتایج تجربی

در این بخش، الگوریتم AGA-net بر روی 5 شبکه واقعی که مربوط به Zachary's Karate Club (Z)، (D) شبکه اجتماعی دلفینها [33]، (A) کالج فوتبال آمریکایی [13]، (B) کتاب‌هایی درباره سیاست ایالات متحده [14] و (C) تعاملات Cattle Protein (IntAct) [34] است آزمایش شده است. این شبکه‌ها به‌صورت غیرمستقیم و بدون وزن سازماندهی شده‌اند. هر گره در شبکه توسط شناسه مشخص می‌شود. برای مثال گره اول شبکه تعاملات Cattle Protein (IntAct) که AATM_BOVIN است [35] با شناسایی 1 مشخص شده است. شبکه‌ها و خواص مورد

استفاده آنها در آزمایشات جدول 1 ارائه شده است. تمام آزمایشات بر روی کامپیوتری انجام شده است که دارای مشخصات زیر است:

سیستم عامل مایکروسافت ویندوز 7 (x64) ، پردازنده اینتل هفت هسته‌ای i7-3632QM (TM) ، 2.20 گیگاهرتز و 4 گیگابایت رم.

Table 1 Networks and their properties used in the experiments

<i>Networks</i> *	<i>Number of Nodes (V)</i>	<i>Number of Edges (E)</i>
[Z] Zachary's Karate Club	34	78
[D] Dolphins Social Network	62	159
[A] American College Football	115	613
[B] Books about US Politics	105	441
[C] Cattle Protein Interactions (IntAct)	268	303

* تعداد گره‌ها و یال‌های داده شده پس از تبدیل شبکه‌ها به گره‌های نامنظم وبدون وزن با کم کردن گره‌های دارای ویژگی‌های خودحلقه به دست آمده است.

الگوریتم 50 بار بر روی شبکه‌های جدول 1 انجام شده است. حداکثر Q ، میانگین Q و انحراف معیار به‌عنوان نتایج آزمون ثبت شده است. تعداد جمعیت مورد استفاده در آزمایش‌ها با توجه به اندازه شبکه متفاوت است. تعداد جمعیتی که برای شبکه‌ها استفاده می‌شود، طبق جدول 1، به ترتیب 20، 30، 100، 80 و 50 است. همچنین برای پارامترهای دیگر مقادیر زیر استفاده شده است؛ نرخ جهش 0.05، نرخ ترکیب 0.8، انتخاب ترکیب 0.5، نرخ جهش 0.2 و نرخ جهش چندنقطه‌ای 0.4. این پارامترها مورد آزمایش قرار گرفته‌اند و تاثیر این مقادیر در الگوریتم از طریق آزمایش و خطا ارزیابی شده است. بهترین نتایج در این زمینه از این مقدار پارامترها برای همه شبکه‌ها بدست آمده است. نتایج تجربی الگوریتم پیشنهاد شده با الگوریتم‌های داده شده در جدول 2 مقایسه شده است. نتایج بدست آمده در جدول 2 ارائه شده است. در جدول، بهترین مقادیر Q به‌صورت پررنگ نشان داده شده و برای سهولت در خواندن، دقت تا سه رقم اعشار نشان داده شده است. به‌عنوان مثال، بهترین مقدار Q از شبکه Z تقریباً 0.419789612097304 بود بنابراین آن را به 0.420 گرد کردیم. متوسط مقادیر Q Modularity و مقادیر انحراف استاندارد برای الگوریتم AGA-net در جدول 3 ارائه شده است.

Table 2 Comparison of the results according to *modularity Q* values

<i>Networks</i> <i>Algorithms</i>	<i>Z</i>	<i>D</i>	<i>A</i>	<i>B</i>
<i>DECD [10]</i>	0.416	-	0.605	-
<i>Eigenvector [19]</i>	0.393	0.491	0.488	0.467
<i>GACD [37]</i>	0.420	0.529	0.604	0.527
<i>GATHB [8]</i>	0.402	0.522	0.551	0.518
<i>GN [13]</i>	0.401	0.519	0.599	0.510
<i>FN [16]</i>	0.381	0.510	0.550	0.502
<i>MA-COM [39]</i>	0.420	0.529	0.605	0.527
<i>MA-Net [8]</i>	0.420	0.529	0.605	0.527
<i>MENSGA [38]</i>	0.420	0.527	0.604	0.526
<i>MOGA-Net [36]</i>	0.416	0.505	0.515	0.518
<i>Proposed method (AGA-net)</i>	0.420	0.529	0.605	0.527

به غیر از شبکه‌های داده شده در جدول 2، حداکثر مقدار Modularity Q برای شبکه‌ای با عنوان تعاملات Cattle Protein 0.720 (IntAct) به دست آمده است. بعد از تجزیه و تحلیل مطالعات حاضر و همانطور که شناخته شده است شبکه (C) پیش از این برای حل مسئله CSD مورد استفاده قرار نگرفته است. بنابراین، این شبکه در لیست مقایسه‌های موجود در جدول 2 قرار نگرفته است.

Table 3 The average *modularity Q* and standard deviation values

<i>Networks</i>	<i>Z</i>	<i>D</i>	<i>A</i>	<i>B</i>	<i>C</i>
<i>Average Q values</i>	0.420	0.528	0.601	0.527	0.719
<i>Standard deviations</i>	± 0	± 0.000790	± 0.002986	± 0.000137	± 0.000017

بهترین مقادیر *modularity Q* در هر اجرا برای همه شبکه‌ها ثبت شده است. با توجه به نتایج به دست آمده، میانگین و بهترین مقدار *Q* یکسان هستند. همچنین سطح پایین انحراف استاندارد نشان می‌دهد که الگوریتم پیشنهاد شده نتایجی سازنده ارائه داده است. به عنوان مثال انحراف معیار شبکه *Z* صفر است بنابراین میانگین و بهترین مقادیر *Q* به طور یکسان 0.420 است. در حالی که بین نتایج به دست آمده برای شبکه‌های دیگر خیلی متفاوت نیست. با مشاهده اعداد تولید شده‌ی جامعه با توجه به بهترین مقادیر *Q* برای شبکه‌های *Z*، *D*، *A*، *B* و *C* به ترتیب مقادیر 4، 5، 10، 5 و 40 بدست آمده‌اند.

همانطور که در جدول 2 مورد بررسی قرار گرفته است، تاکنون الگوریتم AGA-net به بهترین مقدار خود در Q modularity برای چهار شبکه (A, D, Z, و B) که در آن استفاده شده است رسیده است. زمانی که روش مبتنی بر الگوریتم ژنتیک پیشنهاد شده توسط ما با دیگر روش‌های مبتنی بر الگوریتم ژنتیکی داده شده در جدول 2 (GACD و GATHB) مقایسه شده است مشاهده می‌شود که روش پیشنهاد شده نتایج بهتری نسبت به روش‌های دیگر دارد. مهمترین دلیل برای اختلاف دو روش یاد شده و روش پیشنهادی، مکانیسم محدود تولید ژن است در حالی که عملگرهای ترکیب و جهش در روش پیشنهادی، فضا را به راه‌حل مطلوب همگرا می‌کند.

4. نتیجه‌گیری

در این مقاله، مسئله CSD که اغلب در تجزیه و تحلیل شبکه‌های پیچیده مورد استفاده قرار می‌گیرد بحث شده و اطلاعات معنادار از شبکه دنیای واقعی تعیین شده است. برای تست دقت الگوریتم توصیه شده، نتایج به دست آمده با روش‌های پیشرفته در کارهای پیشین مقایسه شده است. در آزمایشات، چهار شبکه اجتماعی و یک شبکه بیولوژیکی استفاده شده است. علاوه بر عملگرهای موجود و پارامترهای الگوریتم ژنتیک استاندارد AGA-net توسط عملگرها و پارامترهای الگوریتم ژنتیک پیشنهادی پشتیبانی می‌شود که در بخش 2.4 ارائه شده است. این اپراتورها و پارامترها سرعت همگرایی بالایی در الگوریتم پیشنهادی با بهترین مقادیر Q دارند. هنگام تجزیه و تحلیل نتایج تجربی مشاهده شده است که الگوریتم AGA-net بهترین مقادیر Q را کنون برای تمام شبکه‌های شناخته شده به دست آورده است. این نتایج نشان می‌دهد که الگوریتم‌های MA-Net [8] و MA-COM [39] نتایج مشابهی با الگوریتم پیشنهاد ارائه می‌کنند. در حالی که الگوریتم پیشنهادی با الگوریتم GACD [37] در 3 شبکه (A, D, Z, و B) نتایج مشابهی به دست آورده است و در مقایسه با شبکه A نتایج بهتری به دست آورده است. به همین ترتیب الگوریتم پیشنهادی با الگوریتم DECD [10] در شبکه A نتایج مشابهی کسب کرده است و نسبت به شبکه Z نتیجه بهتری بدست آورده است. همچنین الگوریتم پیشنهادی در مقایسه با الگوریتم MENSQA [38] در شبکه Z نتایج یکسانی بدست آورده است ولی در شبکه‌های دیگر نتایج بهتری داشته است.

در نتیجه، موفقیت و سازگاری الگوریتم پیشنهاد شده از هر دو جدول مقایسه و جداول استاندارد قابل درک است. علاوه بر این، AGAnet هیچ دانش قبلی‌ای نیاز ندارد و به سرعت کار می‌کند. بنابراین، با استفاده از الگوریتم پیشنهاد شده تقریباً 20 درصد در زمان هر شبکه صرفه‌جویی می‌شود. این الگوریتم به‌گونه‌ای طراحی شده که اندازه گره‌ها و لبه‌های آن را در نظر نمی‌گیرد بنابراین می‌تواند به تمام شبکه‌های مجازی و دنیای واقعی اعمال شود. پس از تجزیه و تحلیل نتایج آزمون مشاهده شد که الگوریتم پیشنهاد شده نتایج خوبی برای تعیین جوامع معنی‌دار از شبکه‌های پیچیده ارائه می‌دهد.

تقدیر و تشکر

این مطالعه توسط پروژه تحقیقات علمی دانشگاه Selcuk و TUBITAK پشتیبانی شده است (شورای تحقیقات علمی و فن‌آوری ترکیه، C-2211 برنامه تحصیلی دکترای اهداف اولویتی، شماره B0314023831649).

References

1. Fortunato, S.: Community Detection in Graphs. *Physics Reports* 486(3), 75–174 (2010)
2. Jackson, M.O.: *Social and Economic Networks*, vol. 3. Princeton University Press, Princeton (2008)
3. Dunne, J.A., Williams, R.J., Martinez, N.D.: Food-Web Structure and Network Theory: The Role of Connectance and Size. *Proceedings of the National Academy of Sciences*. 99(20), 12917–12922 (2002)
4. Gavin, A.-C., et al.: Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440, 631–636 (2006)
5. Krogan, N.J., et al.: Global landscape of protein complexes in the yeast *saccharomyces cerevisiae*. *Nature* 440(7084), 637–643 (2006)
6. Lee, J., Hidden, L.J.: information revealed by optimal community structure from a protein complex bipartite network improves protein function prediction (2013). <http://www.ncbi.nlm.nih.gov/pubmed/23577106>
7. Marcus, S., Moy, M., Coffman, T.: Social network analysis. *Mining Graph Data*, 443–467 (2007)
8. Naeni, L.M., Berretta, R., Moscato, P.: MA-net: a reliable memetic algorithm for community detection by modularity optimization. In: *Proceedings of the 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems*, vol. 1, pp. 311–323. Springer International Publishing (2015)
9. Liu, D., Jin, D., Baquero, C., He, D., Yang, B., Yu, Q.: Genetic Algorithm with a Local Search Strategy for Discovering Communities in Complex Networks. *International Journal of Computational Intelligence Systems* 6(2), 354–369 (2013)
10. Jia, G., et al.: Community detection in social and biological networks using differential evolution. In: *Learning and Intelligent Optimization*, pp. 71–85. Springer Berlin, Heidelberg (2012)
11. Brandes, U., et al.: On Modularity Clustering. *IEEE Transactions on Knowledge and Data Engineering* 20(2), 172–188 (2008)
12. Parpinelli, R.S., Lopes, H.S.: A Computational Ecosystem for Optimization: Review and Perspectives for Future Research. *Memetic Computing* 7(1), 29–41 (2015)
13. Girvan, M., Newman, M.E.J.: Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Sciences* 99(12), 7821–7826 (2002)
14. Newman, M.E.J.: Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences* 103(23), 8577–8582 (2006)
15. Newman, M.E.J., Girvan, M.: Finding and Evaluating Community Structure in Networks. *Physical Review E* 69(2), 026113 (2004)
16. Newman, M.E.J.: Fast Algorithm for Detecting Community Structure in Networks. *Physical Review E* 69(6), 066133 (2004)
17. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008(10), P10008 (2008)
18. Pons, P., Latapy, M.: Computing communities in large networks using random walks. In: *Computer and Information Sciences-ISCIS 2005*, pp. 284–293. Springer Berlin, Heidelberg (2005)
19. Newman, M.E.J.: Finding Community Structure in Networks using The Eigenvectors of Matrices. *Physical Review E* 74(3), 036104 (2006)
20. Raghavan, U.N., Albert, R., Kumara, S.: Near Linear Time Algorithm to Detect Community Structures in Large-Scale Networks. *Physical Review E* 76(3), 036106 (2007)
21. Ronhovde, P., Nussinov, Z.: Local Resolution-Limit-Free Potts Model for Community Detection. *Physical Review E* 81(4), 046114 (2010)
22. Huang, J., Sun, H., Liu, Y., Song, Q., Wenginger, T.: Towards Online Multiresolution Community Detection in Large-Scale Networks. *PLoS one* 6(8), e23829 (2011)
23. Clauset, A., Newman, M.E.J., Moore, C.: Finding Community Structure in Very Large Networks. *Physical Review E* 70(6), 066111 (2004)
24. Duch, J., Arenas, A.: Community Detection in Complex Networks using Extremal Optimization. *Physical Review E* 72(2), 027104 (2005)

25. Tasgin, M., Bingol, A.: Communities detection in complex networks using genetic algorithm. In: Proc. of the European Conference on Complex Systems (ECSS 2006) (2006)
26. Tasgin, M., Herdagdelen, A., Bingol, H.: Community Detection in Complex Networks using Genetic Algorithms. arXiv preprint arXiv:0711.0491 [physics.soc-ph] (2007)
27. Li, J., Song, Y.: Community Detection in Complex Networks using Extended Compact Genetic Algorithm. *Soft Computing* 17(6), 925–937 (2013)
28. Shi, C., Wang, Y., Wu, B., Zhong, C.: A New Genetic Algorithm for Community Detection, pp. 1298–1309. Springer, Heidelberg (2009)
29. Mitchell, M.: Genetic algorithms: An overview. *Complexity* 1(1), 31–39 (1995)
30. Park, Y., Song, M.: A genetic algorithm for clustering problems. In: Proceedings of the Third Annual Conference on Genetic Programming, pp. 568–575 (1998)
31. Sastry, K., Goldberg, D., Kendall, G.: Genetic Algorithms: In: Search methodologies, pp. 97–125. Springer US (2005)
32. Zachary, W.W.: An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research*, 452–473 (1977)
33. Lusseau, D.: The Emergent Properties of a Dolphin Social Network. *Proceedings of the Royal Society of London B: Biological Sciences* 270(Suppl 2), S186–S188 (2003)
34. IntAct Protein-Protein Interaction Network of Cattle. [http://biit.cs.ut.ee/graphweb/exampleInput/Cattle_protein_interactions_\(IntAct\).txt](http://biit.cs.ut.ee/graphweb/exampleInput/Cattle_protein_interactions_(IntAct).txt) (last accessed date June 1, 2015)
35. UniProtKB, Bos taurus (Bovine) - P12344 (AATM_BOVIN). <http://www.uniprot.org/uniprot/P12344> (last accessed date May 15, 2015)
36. Pizzuti, C.: A Multiobjective Genetic Algorithm to Find Communities in Complex Networks. *IEEE Transactions on Evolutionary Computation* 16(3), 418–430 (2012)
37. Shi, C., Yan, Z., Wang, Y., Cai, Y., Wu, B.: A Genetic Algorithm for Detecting Communities in Large-Scale Complex Networks. *Advances in Complex Systems* 13(01), 3–17 (2010)
38. Li, Y., Liu, G., Lao, S.Y.: A Genetic Algorithm for Community Detection in Complex Networks. *Journal of Central South University* 20, 1269–1276 (2013)
39. Gach, O., Hao, J.-K.: A memetic algorithm for community detection in complex networks. In: Coello, C.A.C., Cutello, V., Deb, K., Forrest, S., Nicosia, G., Pavone, M. (eds.) PPSN 2012, Part II. LNCS, vol. 7492, pp. 327–336. Springer, Heidelberg (2012)