

A Novel Multistage Approach to Community Detection in Social Networks Based on Fuzzy DBSCAN Clustering and Reptile Search Optimization

Mostafa. Aliabadi¹, Hamidreza. Ghaffari^{1*}

¹ Department of Computer Engineering, Fe.C., Islamic Azad University, Ferdows, Iran

* Corresponding author email address: hamidghaffary53@iau.ac.ir

Article Info

Article type:

Original Research

How to cite this article:

Aliabadi, M. & Ghaffari, H. (2026). A Novel Multistage Approach to Community Detection in Social Networks Based on Fuzzy DBSCAN Clustering and Reptile Search Optimization. *Decision Science and Intelligent Systems*. 3(1), 1-27.



© 2026 the authors. Published by KMAN Publication Inc. (KMANPUB), Ontario, Canada. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

ABSTRACT

Community detection is a fundamental problem in the analysis of complex and social networks. It aims to identify groups of nodes whose internal connections are denser than their connections with other parts of the network. The primary challenges in this field are the unknown number of communities and the large number of nodes in real-world networks. This study proposes a novel multistage approach that combines the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm with the Reptile Search Algorithm (RSA). In the first stage, the network is initially clustered using DBSCAN and features extracted from the nodes, particularly their neighbor lists. In the second stage, RSA identifies the final communities through three phases: initialization, exploration, and exploitation. The objective function is defined based on two criteria: modularity (Q) and conductance (Cond). The proposed method was evaluated in MATLAB using eight real-world datasets: ca-GrQc, ca-HepPh, ca-HepTh, ca-AstroPh, ca-CondMat, Zachary's Karate Club network, the American college football network, and the dolphin social network. The experimental results demonstrate that the proposed method outperforms existing methods in most cases, achieving higher values for normalized mutual information (NMI) and modularity (Q). **Keywords:** community detection, complex networks, DBSCAN, Reptile Search Algorithm (RSA), modularity, normalized mutual information

Extended Abstract

Introduction

Complex networks provide a powerful analytical framework for representing and examining systems in which numerous entities interact through heterogeneous and often nonlinear relationships. Social networks, scientific collaboration networks, biological networks, communication infrastructures, and organizational relationship networks can all be modeled as graphs in which nodes represent actors or entities and edges represent interactions. One of the most important structural properties of such networks is community structure, defined as the presence of groups of nodes with relatively dense internal connections and comparatively sparse external connections. Accurate community detection can reveal hidden patterns of collective behavior, information diffusion, influence, vulnerability, and coordination within complex systems (Girvan & Newman, 2002; Plantié & Crampes, 2013).

Community detection has substantial theoretical and managerial relevance. Identified communities can support customer segmentation, recommender systems, targeted marketing, fraud detection, criminal-network analysis, organizational communication assessment, and the identification of influential groups. Nevertheless, real-world networks create several computational and methodological difficulties. The number of communities is generally unknown in advance, network size may be extremely large, and observed structures may contain noise, outliers, irregular boundaries, and communities of unequal density. Global approaches can exploit the complete network topology and often produce accurate partitions, but their computational and memory requirements restrict their scalability. Local approaches are faster because they rely primarily on neighborhood information, yet they may overlook the overall organization of the network (Aghaalizadeh et al., 2021; Li et al., 2024).

Several established methods address particular dimensions of this problem. Modularity maximization evaluates whether within-community connections exceed those expected under an appropriate null model and remains one of the most widely used criteria for assessing network partitions (Newman, 2006). The Louvain method improves the computational efficiency of modularity optimization through greedy multilevel aggregation, whereas Infomap detects communities by compressing the description of random walks over a graph (Blondel et al., 2008; Rosvall & Bergstrom, 2008). Density-based clustering is also relevant because DBSCAN can identify irregularly shaped clusters, distinguish noise points, and determine the number of clusters without requiring it to be specified beforehand (Ester et al., 1996). However, the quality of density-based partitions remains sensitive to neighborhood definitions and parameter settings.

Metaheuristic optimization offers a complementary mechanism for refining preliminary communities and escaping locally optimal partitions. The Reptile Search Algorithm models exploration and exploitation behaviors to search complex solution spaces and has demonstrated potential as a general-purpose nature-inspired optimizer (Abualigah et al., 2022). Recent research has therefore emphasized hybrid and multistage strategies that combine local structural information, clustering, learning, and optimization. These strategies can reduce the search space while retaining the capacity to optimize global partition quality (Aliabadi & Ghaffari, 2025; Gmati et al., 2024; Zhang et al., 2023). Accordingly, the present study developed a two-stage community-detection framework combining fuzzy DBSCAN-based initial clustering with Reptile Search Algorithm optimization. The objective was to improve community

identification in both labeled and unlabeled real-world networks by jointly considering modularity and conductance.

Methods and Materials

Social networks were represented as undirected graphs in which vertices denoted users or entities and edges represented interactions. The proposed procedure consisted of two sequential stages. In the first stage, a feature vector was constructed for each node from its structural neighborhood, with particular emphasis on its neighbor list. A fuzzy DBSCAN-based process then generated an initial partition. The neighborhood radius and minimum number of points were used to distinguish core nodes, border nodes, and noise points. This stage estimated the initial number and composition of communities without requiring prior knowledge of the true partition.

In the second stage, the Reptile Search Algorithm refined the preliminary partition. Each candidate solution was encoded as a vector with a length equal to the number of network nodes, and each vector element specified the community label assigned to the corresponding node. The initial population was generated around the structure produced during the first stage rather than through an entirely uninformed random search. The algorithm then proceeded through initialization, exploration, and exploitation. During exploration, long-walk and belly-walk operators promoted coverage of new regions in the search space. During exploitation, hunting-coordination and hunting-cooperation mechanisms concentrated the search around the best candidate partitions.

Candidate solutions were evaluated through a composite fitness function based on modularity and conductance. Modularity measured the strength of within-community connectivity relative to a null expectation, whereas conductance represented the proportion of external connections relative to the internal cohesion of a community. The optimization process therefore sought partitions characterized by high modularity and low conductance. Normalized mutual information was additionally used for datasets with known ground-truth community labels.

The method was implemented in MATLAB R2023b on a system equipped with an Intel Core i7 processor, 8 GB of RAM, and the Windows 8 operating system. Five unlabeled scientific collaboration networks were used: ca-GrQc with 5,242 nodes and 14,496 edges, ca-HepTh with 9,887 nodes and 25,998 edges, ca-HepPh with 12,008 nodes and 118,521 edges, ca-AstroPh with 18,772 nodes and 198,080 edges, and ca-CondMat with 23,133 nodes and 93,497 edges.

Five labeled networks were also examined: the Karate Club network with 34 nodes and 78 edges, the Dolphins network with 62 nodes and 159 edges, the American college football network with 115 nodes and 613 edges, the Polblogs network with 643 nodes and 2,300 edges, and the Yeast network with 1,500 nodes and 1,900 edges. The proposed framework was compared with three benchmark community-detection methods developed by Gharehchopogh, Gmati and colleagues, and Aliabadi and Ghaffari. Each experiment was independently repeated 30 times, and the resulting modularity and normalized mutual information values were recorded.

Findings

For the five unlabeled networks, the proposed method achieved the highest modularity value in every comparison. The modularity values were 0.8154 for ca-GrQc, 0.7710 for ca-HepTh, 0.8216 for ca-HepPh, 0.8883 for ca-AstroPh, and 0.7316 for ca-CondMat. The average modularity across these datasets was 0.8056.

The strongest result among the unlabeled networks was obtained for ca-AstroPh. The proposed method produced a modularity value of 0.8883, compared with 0.7883 for the best-performing benchmark method, representing an absolute improvement of 0.1000. For ca-HepPh, the proposed value of 0.8216 exceeded the best baseline value of 0.7621. Improvements were also observed for ca-GrQc, ca-HepTh, and ca-CondMat, demonstrating that the method remained effective across networks with substantially different numbers of nodes, edges, and average degrees.

The proposed method also produced the highest modularity values for all five labeled networks. The obtained values were 0.9681 for Karate, 0.9016 for Polblogs, 0.7943 for Dolphins, 0.6534 for Football, and 0.7851 for Yeast. Their mean was 0.8205. The largest absolute advantage was observed for the Karate network, for which the proposed method improved modularity from the best baseline value of 0.7868 to 0.9681. A considerable improvement was also recorded for Yeast, where modularity increased from 0.6812 to 0.7851.

Normalized mutual information results similarly favored the proposed method across all labeled datasets. The NMI values were 0.8486 for Karate, 0.7362 for Polblogs, 0.9247 for Dolphins, 0.9025 for Football, and 0.7614 for Yeast, producing an overall mean of 0.8347. The highest NMI was obtained for Dolphins, indicating close correspondence between the detected and actual community structures. The lowest NMI was observed for Polblogs, although the proposed value of 0.7362 remained higher than the best baseline result of 0.7165.

For Dolphins, NMI increased from the best benchmark value of 0.8538 to 0.9247. For Karate, it increased from 0.7805 to 0.8486, and for Yeast, it rose from 0.6923 to 0.7614. The Football result reached 0.9025, compared with the best baseline value of 0.8833. These consistent numerical improvements indicate that the proposed framework not only generated partitions with stronger internal network structure but also more accurately recovered known community memberships.

Discussion and Conclusion

The findings demonstrate that combining density-based preliminary clustering with metaheuristic optimization can provide an effective solution to community detection in complex social networks. The initial DBSCAN-based stage reduced uncertainty regarding the number and approximate composition of communities. It also supplied the optimization stage with structurally meaningful prior information, thereby avoiding the inefficiency associated with beginning the search from completely random partitions.

The Reptile Search Algorithm subsequently improved community boundaries through a balance between global exploration and local exploitation. Exploration mechanisms enabled the algorithm to examine alternative community assignments, whereas exploitation mechanisms refined promising solutions around the best partitions. This balance appears to have reduced premature convergence and improved the algorithm's ability to avoid poor local optima.

The simultaneous use of modularity and conductance was another important feature of the framework. Optimizing modularity alone may produce partitions with dense internal connections but does not always sufficiently penalize excessive external connectivity. Incorporating conductance provided a complementary assessment of community separation. The resulting objective function encouraged communities that were internally cohesive and externally distinguishable.

The method performed consistently across small benchmark networks and much larger scientific collaboration networks. This consistency suggests that the framework is not limited to a particular network scale or density pattern. Its performance on labeled datasets further indicates that increases in modularity

were accompanied by improved agreement with ground-truth structures rather than merely by the production of numerically favorable but substantively inaccurate partitions.

From a managerial perspective, more accurate community detection can improve decision-making in customer analytics, digital marketing, organizational network analysis, recommendation services, risk monitoring, and security management. Identifying cohesive groups enables decision-makers to design differentiated interventions, allocate communication resources more efficiently, detect influential subgroups, and recognize abnormal or potentially threatening interaction patterns.

In conclusion, the proposed fuzzy DBSCAN–RSA framework constitutes an effective multistage approach to community detection. It outperformed the selected benchmark methods across all examined datasets in terms of modularity and normalized mutual information. Future developments should examine alternative structural and behavioral node features, automatic adjustment of DBSCAN parameters, overlapping and dynamic communities, parallel implementation for very large networks, and integration with newer metaheuristic algorithms.

رویکردی نوین چند مرحله‌ای برای تشخیص انجمن در شبکه‌های اجتماعی مبتنی بر خوشه‌بندی فازی DBSCAN و بهینه‌سازی جستجوی خزنده

مصطفی علی آبادی^۱، حمیدرضا غفاری^{۱*}

۱. گروه مهندسی کامپیوتر، واحد فردوس، دانشگاه آزاد اسلامی، فردوس، ایران

*ایمیل نویسنده مسئول: hamidghaffary53@iau.ac.ir

اطلاعات مقاله

چکیده

نوع مقاله

پژوهشی/اصیل

نحوه استناد به این مقاله:

علی آبادی، مصطفی، و غفاری، حمیدرضا. (۱۴۰۵). رویکردی نوین چند مرحله‌ای برای تشخیص انجمن در شبکه‌های اجتماعی مبتنی بر خوشه‌بندی فازی DBSCAN و بهینه‌سازی جستجوی خزنده. علم تصمیم گیری و سیستم های هوشمند، ۳(۱)، ۱-۲۷.



© ۱۴۰۵ تمامی حقوق انتشار این مقاله متعلق به نویسنده است. انتشار این مقاله به صورت دسترسی آزاد مطابق با گواهی (CC BY-NC 4.0) صورت گرفته است.

تشخیص انجمن یکی از مسائل بنیادی در تحلیل شبکه‌های پیچیده و اجتماعی است که هدف آن شناسایی گروه‌هایی از گره‌ها با ارتباطات درونی متراکم‌تر نسبت به سایر بخش‌های شبکه می‌باشد. چالش اصلی در این حوزه، عدم آگاهی از تعداد دقیق جوامع و حجم بالای گره‌ها در شبکه‌های واقعی است. در این پژوهش، یک رویکرد چندمرحله‌ای نوین ارائه می‌شود که از ترکیب الگوریتم خوشه‌بندی مبتنی بر چگالی (DBSCAN) و الگوریتم بهینه‌سازی جستجوی خزنده (RSA) بهره می‌گیرد. در مرحله اول، با استفاده از DBSCAN و ویژگی‌های استخراج شده از گره‌ها به‌ویژه لیست همسایگان، خوشه‌بندی اولیه شبکه انجام می‌شود. در مرحله دوم، الگوریتم RSA با سه فاز مقداردهی اولیه، اکتشاف و بهره‌برداری، جوامع نهایی را شناسایی می‌کند. تابع هدف بر پایه دو معیار ماژولاریتی (Q) و پراکندگی (Cond) تعریف شده است. روش پیشنهادی روی هشت مجموعه داده واقعی شامل ca-GrQc، ca-HepPh، ca-HepTh، ca-AstroPh، ca-CondMat، شبکه کاراته، فوتبال کالج آمریکایی و شبکه دلفین در محیط MATLAB ارزیابی شده است. نتایج آزمایش‌ها نشان می‌دهد که روش پیشنهادی در اکثر موارد عملکرد بهتری نسبت به روش‌های موجود دارد و مقادیر بالاتری از شاخص‌های NMI و ماژولاریتی Q را به دست می‌آورد.

کلیدواژگان: تشخیص انجمن، شبکه‌های پیچیده، DBSCAN، الگوریتم جستجوی خزنده (RSA)، ماژولاریتی، شاخص NMI

۱. مقدمه

در دهه های اخیر، شبکه های پیچیده به ابزاری اساسی برای مدل سازی سیستم های واقعی تبدیل شده اند. شبکه های اجتماعی، شبکه های همکاری علمی، شبکه های زیستی، شبکه های برق و شبکه های وب، همگی نمونه هایی از این سیستم های پیچیده هستند که در آن ها گره ها موجودیت ها و یال ها روابط میان آن ها را نمایش می دهند [۱، ۲]. تحلیل ساختار این شبکه ها اطلاعات ارزشمندی درباره رفتار جمعی، انتشار اطلاعات، و آسیب پذیری های سیستم فراهم می کند [۳].

یکی از مسائل محوری در تحلیل شبکه های پیچیده، **تشخیص انجمن**^۱ است. انجمن به مجموعه ای از گره ها اطلاق می شود که اتصالات درونی آن ها نسبت به اتصالات بیرونی به مراتب متراکم تر است [۴]. شناسایی این ساختارها کاربردهای گسترده ای دارد؛ از جمله: بهبود سیستم های توصیه گر، شناسایی گروه های مجرمانه در شبکه های اجتماعی، تحلیل شبکه های پروتئینی در زیست شناسی، و بهینه سازی مسیریابی در شبکه های ارتباطی [۵، ۶].

روش های تشخیص انجمن را می توان بر اساس نوع اطلاعات مورد استفاده به سه دسته اصلی تقسیم کرد: **روش های مبتنی بر شاخص های محلی**، **روش های مبتنی بر شاخص های سراسری**، و **روش های ترکیبی** [۷]. روش های سراسری با بهره گیری از اطلاعات کل شبکه، دقت بالایی دارند؛ اما پیچیدگی محاسباتی بالایی داشته و برای شبکه های بزرگ مقیاس کارایی لازم را ندارند [۸]. روش های محلی از ساختار همسایگی گره ها بهره می گیرند و سرعت بالاتری دارند، اما به دلیل عدم دسترسی به اطلاعات سراسری، دقت آن ها محدود است [۹]. روش های ترکیبی تلاش می کنند مزایای هر دو دسته را با هم داشته باشند [۱۰].

علی رغم پیشرفت های قابل توجه در این حوزه، چالش های اساسی متعددی همچنان باقی است. مهم ترین این چالش ها عبارتند از: نامشخص بودن تعداد جوامع در شبکه های واقعی، حجم بالای گره ها و یال ها در شبکه های بزرگ مقیاس، حساسیت الگوریتم ها به نویز و داده های پرت، وابستگی شدید به پارامترهای تنظیمی، همگرایی زودرس در الگوریتم های بهینه سازی، و ضعف شاخص های شباهت مورد استفاده [۱۱، ۱۲]. از سوی دیگر، در اکثر پژوهش های پیشین، **انتخاب شاخص های شباهت** به صورت تجربی و بدون توجیه نظری انجام شده است؛ این در حالی است که انتخاب صحیح این شاخص ها تأثیر مستقیمی بر کیفیت تشخیص انجمن دارد [۱۳].

بررسی ادبیات پژوهش نشان می دهد که **روش های چندمرحله ای** که در آن ها ابتدا گره های هسته ای شناسایی شده، سپس با استفاده از شاخص های شباهت و پیش بینی لینک، انجمن ها تشخیص داده می شوند، می توانند بسیاری از این چالش ها را برطرف کنند [۱۴، ۱۵]. با این حال، ترکیب منسجم خوشه بندی مبتنی بر چگالی با الگوریتم های فراابتکاری برای بهینه سازی تعداد و مرزهای جوامع، کمتر مورد توجه قرار گرفته است [۱۶].

در این پژوهش، یک **رویکرد چندمرحله ای نوین** ارائه می شود که از ترکیب الگوریتم **DBSCAN** و الگوریتم بهینه سازی **جستجوی خزنده ((RSA بهره می گیرد [۱۷]. در مرحله اول، DBSCAN با استفاده از لیست همسایگان گره ها، خوشه بندی اولیه را انجام می دهد. در مرحله دوم، RSA با سه فاز مقداردهی اولیه، اکتشاف و بهره برداری، جوامع نهایی را شناسایی و بهینه می کند. تابع هدف بر پایه دو معیار ماژولاریتی (Q) و پراکندگی ($Cond$) تعریف شده است. روش پیشنهادی روی ۸ مجموعه داده استاندارد شامل شبکه های واقعی و مصنوعی ارزیابی شده و نتایج آن با روش های پیشرفته موجود مقایسه گردیده است.**

^۱ Community Detection

ساختار این مقاله به شرح زیر است: بخش ۲ مروری بر پژوهش‌های پیشین ارائه می‌دهد. بخش ۳ روش پیشنهادی را به تفصیل شرح می‌دهد. بخش ۴ به ارزیابی تجربی و مقایسه نتایج اختصاص دارد و بخش ۵ نتیجه‌گیری و پیشنهادهای آتی را در بر می‌گیرد.

۲. پیشینه تحقیق

۲-۱. روش‌های مبتنی بر شاخص‌های محلی

تشخیص انجمن در شبکه‌های پیچیده از دیرباز یکی از چالش‌های اساسی در حوزه تحلیل شبکه بوده است. روش‌های محلی، به عنوان یکی از سه دسته اصلی رویکردهای تشخیص انجمن، بر اساس اطلاعات همسایگی بلافاصله هر گره عمل می‌کنند و نیازی به دانش کلی از ساختار شبکه ندارند [۱]. این ویژگی باعث می‌شود که این روش‌ها از نظر محاسباتی سبک‌تر بوده و برای شبکه‌های بزرگ مقیاس عملی‌تر باشند [۱۶].

از جمله پرکاربردترین شاخص‌های محلی می‌توان به همسایه‌های مشترک^۱ اشاره کرد. این شاخص بر این فرض استوار است که دو گره‌ای که همسایه‌های مشترک بیشتری دارند، احتمال بیشتری برای تعلق به یک انجمن دارند [۴]. به طور رسمی، اگر $\Gamma(u)$ مجموعه همسایگان گره u باشد، شاخص همسایه‌های مشترک بین دو گره u و v به صورت زیر تعریف می‌شود:

$$CN(u, v) = |\Gamma(u) \cap \Gamma(v)| \quad (1)$$

شاخص دیگری که در این دسته مطرح است، وابستگی امتیازی^۲ است که میزان همبستگی آماری بین دو گره را بر اساس ساختار محلی آن‌ها اندازه‌گیری می‌کند [۲]. این شاخص‌ها در سناریوهای پیش‌بینی پیوند و تشخیص انجمن به طور گسترده‌ای به کار رفته‌اند. در حوزه روش‌های محلی، رویکردهای مبتنی بر هسته‌یابی نیز جایگاه ویژه‌ای دارند. در این رویکردها، ابتدا گره‌هایی که ارتباطات داخلی قوی و ارتباطات خارجی کمی دارند به عنوان هسته‌های اولیه شناسایی می‌شوند. سپس هسته‌های کم‌استقلال با هسته‌های همجوار ادغام شده و گره‌های چندعضوی به انجمنی منتقل می‌شوند که استقلال کلی را بیشتر می‌کند [۱۸]. این رویکرد در مقایسه با روش‌های ساده‌تر محلی، کارایی بهتری نشان داده است.

دینگ^۳ و همکاران روشی مبتنی بر توسعه محلی و بازبینی مرز ارائه دادند. در این روش، انجمن‌ها از یک گره هسته‌ای آغاز شده و به تدریج با افزودن گره‌های همسایه گسترش می‌یابند. مرحله بازبینی مرز نیز برای اصلاح گره‌هایی که در مرز دو انجمن قرار دارند به کار می‌رود. این روش توانست پوشش شبکه را بهبود بخشد، اما همچنان به ساختار کلی شبکه توجه کافی ندارد [۱۹].

لی^۴ و همکاران الگوریتمی مبتنی بر امتیازدهی محلی چندعاملی ارائه دادند که در آن هر گره بر اساس معیارهای متنوع محلی امتیازدهی می‌شود. در مرحله بعد، برچسب‌گذاری بر اساس مناسب‌ترین همسایه انجام می‌گیرد و در نهایت گروه‌های کوچک با یکدیگر ادغام می‌شوند تا جوامع نهایی شکل بگیرند. نتایج این روش از نظر دقت و زمان اجرا بهتر از بسیاری از الگوریتم‌های پیشرفته گزارش شده است [۱۶].

ژانگ و همکاران نیز رویکردی مبتنی بر گره هسته و انتشار برچسب لایه‌به‌لایه پیشنهاد دادند. در این روش، گره‌های هسته به عنوان مراکز بالقوه انجمن‌ها شناسایی شده و برچسب‌ها به صورت لایه‌به‌لایه به گره‌های اطراف منتشر می‌شوند. این رویکرد از تأثیر منفی گره‌های کم‌درجه جلوگیری می‌کند و برای شبکه‌های بزرگ مقیاس مناسب است [۳].

¹ Common Neighbors

² Pointwise Mutual Information

³ Ding

⁴ Li

با وجود مزایای ذکرشده، روش های محلی دارای محدودیت های اساسی هستند. مهم ترین چالش این روش ها نبود اطلاعات سراسری است که باعث می شود ساختار کلی شبکه نادیده گرفته شود [۲]. علاوه بر این، این روش ها نسبت به کیفیت معیارهای شباهت حساس بوده و در شبکه هایی با ساختار نامنظم یا نویز بالا، دقت کمتری دارند. این محدودیت ها انگیزه اصلی برای توسعه روش های ترکیبی محلی-سراسری بوده است که در بخش های بعدی مورد بررسی قرار می گیرند.

۲-۲. روش های مبتنی بر شاخص های سراسری

روش های سراسری تشخیص انجمن بر خلاف روش های محلی، از اطلاعات ساختاری کل شبکه بهره می برند. این روش ها به تمام گره ها و روابط میان آن ها دسترسی داشته و با تحلیل جامع ساختار شبکه، به تشخیص انجمن ها می پردازند. این ویژگی باعث می شود که دقت این روش ها در بسیاری از موارد بالاتر از روش های محلی باشد، اما در عوض پیچیدگی محاسباتی بالایی دارند و برای شبکه های بسیار بزرگ مقیاس مناسب نیستند [۱].

۲-۲-۱. خوشه بندی طیفی^۱

یکی از رویکردهای شناخته شده در دسته روش های سراسری، خوشه بندی طیفی است. این روش مسئله تشخیص انجمن را به تجزیه گراف تبدیل کرده و از تحلیل طیفی ماتریس مجاورت یا ماتریس لاپلاسیان گراف برای مشخص کردن بخش های کمینه استفاده می کند. در این رویکرد، مقادیر ویژه و بردارهای ویژه ماتریس گراف محاسبه شده و بر اساس آن ها، گره ها به خوشه های مختلف تخصیص داده می شوند. خوشه بندی طیفی به ویژه در شبکه هایی که ساختار انجمن های آن ها به خوبی تعریف شده است، عملکرد مناسبی دارد، اما محاسبه مقادیر ویژه برای شبکه های بزرگ هزینه بر است [۲۰].

۲-۲-۲. بیشینه سازی پیمانه ای^۲

پیمانه ای بودن^۳ یکی از پرکاربردترین معیارهای کیفیت برای تشخیص انجمن است. این معیار با محاسبه درجه توزیع گره ها، میزان تراکم یال های درون انجمنی را نسبت به یال های بین انجمنی ارزیابی می کند. به طور رسمی، ماژولاریتی Q به صورت زیر تعریف می شود:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (2)$$

که در آن A_{ij} عنصر ماتریس مجاورت، k_i درجه گره i ، m تعداد کل یال ها، و $\delta(c_i, c_j)$ تابع دلتای کرونکر است که اگر گره های i و j در یک انجمن باشند برابر ۱ و در غیر این صورت برابر ۰ است.

روش های مبتنی بر بیشینه سازی پیمانه ای غالباً سلسله مراتبی هستند و در شبکه های بزرگ زمان بر می شوند. بلاندل و همکاران الگوریتم سریعی به نام **Louvain** برای بهینه سازی پیمانه ای ارائه دادند که با استفاده از رویکرد حریصانه و چندمرحله ای، هزینه محاسباتی را کاهش می دهد. این الگوریتم در هر مرحله، گره ها را به انجمن هایی منتقل می کند که بیشترین افزایش در ماژولاریتی را ایجاد می کنند، سپس شبکه را فشرده سازی کرده و فرآیند را تکرار می کند [۲۱].

^۱ Spectral Clustering

^۲ Modularity Maximization

^۳ Modularity

۲-۳-۲. مدل‌های مبتنی بر اطلاعات^۱

راسوال و همکاران یک مدل مبتنی بر اطلاعات برای ساختار انجمن پیشنهاد کردند که در آن مسئله تشخیص انجمن به کدگذاری اطلاعات تبدیل می‌شود. در این رویکرد، هدف یافتن تقسیم‌بندی شبکه است که کوتاه‌ترین توصیف کدگذاری شده از مسیرهای تصادفی روی گراف را ارائه دهد. الگوریتم **Infomap** که بر این اساس طراحی شده، از مفهوم **نظریه اطلاعات** و **آنتروپی شانون** برای شناسایی انجمن‌هایی استفاده می‌کند که جریان اطلاعات درون آن‌ها بیشتر از جریان بین آن‌ها است [۲۲].

روش‌های دیگر سراسری

علاوه بر موارد فوق، روش‌های سراسری دیگری نیز وجود دارند که هر کدام رویکرد خاص خود را دارند:

- **روش‌های مبتنی بر حذف یال‌ها:** مانند الگوریتم **Girvan-Newman** که با حذف تکراری یال‌هایی با بیشترین **betweenness centrality**، شبکه را به انجمن‌های کوچک‌تر تقسیم می‌کند [۲۳].
- **روش‌های احتمالی و بیزین:** که از مدل‌های آماری برای استنتاج ساختار انجمن استفاده می‌کنند.
- **روش‌های مبتنی بر بهینه‌سازی:** که مسئله تشخیص انجمن را به یک مسئله بهینه‌سازی ترکیبی تبدیل کرده و از الگوریتم‌های فراابتکاری برای حل آن استفاده می‌کنند.

آقالیزاده و همکاران الگوریتمی سه‌مرحله‌ای ارائه دادند که با **رتبه‌بندی اهمیت گره‌ها** بر اساس معیارهای سراسری، تشخیص انجمن محلی را در شبکه‌های اجتماعی بهبود می‌بخشد. در این روش، ابتدا گره‌های مهم شناسایی شده و به‌عنوان هسته‌های اولیه انجمن در نظر گرفته می‌شوند. سپس با استفاده از اطلاعات محلی، انجمن‌ها گسترش می‌یابند. نتایج آن‌ها نشان داد که **ترکیب اطلاعات محلی و سراسری** می‌تواند دقت تشخیص را به‌طور قابل‌توجهی افزایش دهد و در عین حال پیچیدگی محاسباتی را نسبت به روش‌های کاملاً سراسری کاهش دهد [۲].

چالش‌های روش‌های سراسری

با وجود دقت بالای روش‌های سراسری، این روش‌ها با چالش‌های اساسی مواجه هستند:

۱. **پیچیدگی محاسباتی بالا:** بسیاری از این روش‌ها دارای پیچیدگی زمانی $O(n^2)$ یا بالاتر هستند که برای شبکه‌های میلیونی غیرعملی است.
۲. **نیاز به حافظه زیاد:** ذخیره‌سازی ماتریس‌های کامل گراف برای شبکه‌های بزرگ مشکل‌ساز است.
۳. **حساسیت به پارامترها:** بسیاری از این روش‌ها نیاز به تنظیم پارامترهایی مانند تعداد انجمن‌ها دارند.
۴. **عدم مقیاس‌پذیری:** برای شبکه‌های پویا و در حال رشد، اجرای مجدد کامل الگوریتم هزینه‌بر است.

این محدودیت‌ها انگیزه اصلی برای توسعه **روش‌های ترکیبی محلی-سراسری** بوده است که سعی می‌کنند مزایای هر دو رویکرد را با هم ترکیب کنند و از ضعف‌های آن‌ها بکاهند [۱۶].

¹ Information-Theoretic Models

۳-۲. روش های مبتنی بر یادگیری عمیق

در سال های اخیر، یادگیری عمیق به عنوان رویکردی نوین در تشخیص انجمن شبکه های پیچیده مطرح شده است. ایده اصلی این رویکردها استفاده از خودرمزگذار^۱ برای کاهش ابعاد فضای ویژگی شبکه و تولید بازنمایی پنهان فشرده از ساختار گراف است. پس از کاهش ابعاد، می توان از روش های خوشه بندی مختلف برای تشخیص انجمن ها در فضای بازنمایی پنهان استفاده کرد. این رویکرد برای شبکه های پیچیده و پرابعد که روش های کلاسیک در آن ها ناکارآمد هستند، مزیت قابل توجهی دارد [۸].

۳-۲-۱. خودرمزگذار پشته ای^۲

ژو^۳ و همکاران یک روش تشخیص انجمن مبتنی بر خودرمزگذار پشته ای را در قالب یک چارچوب خوشه بندی تقویتی ارائه دادند. ویژگی های اصلی این روش عبارتند از:

- استفاده از چهار شاخص شباهت برای استخراج اطلاعات محلی بین گره ها
 - ترکیب یادگیری انتقال و رمزگذار خودکار برای تولید بازنمایی کم بعد از شبکه
 - استفاده از خوشه بندی مبتنی بر فاکتوراسیون ماتریس غیر منفی ((NMF برای استخراج نتایج قابل اعتماد
- خودرمزگذار پشته ای از لایه های متوالی تشکیل می شود که به صورت مجزا آموزش می بینند و یک پارامتر توسعه برای کنترل فرآیند آموزش به کار می رود. این ساختار امکان یادگیری بازنمایی های سلسله مراتبی از ساختار شبکه را فراهم می کند.
- با این حال، این روش دارای محدودیت هایی است: نتایج نهایی به ماتریس شباهت اولیه و خوشه بندی اولیه وابسته هستند و ضعف در مسئله یادگیری انتقال و استفاده از کمینه سازی دامنه هدف به جای انتقال دانش مؤثر، از چالش های اصلی آن محسوب می شود [۸].

۳-۲-۲. یادگیری عمیق با خودرمزگذار پشته ای برای شبکه های اجتماعی

حسینی و گلوی رویکردی مبتنی بر یادگیری عمیق برای تشخیص اجتماع در شبکه های اجتماعی ارائه دادند. روش پیشنهادی آن ها شامل مراحل زیر است:

۱. ساخت ماتریس شباهت جدید از توپولوژی شبکه
 ۲. طراحی خودرمزگذار پشته ای برای کاهش ابعاد
 ۳. اعمال چند الگوریتم خوشه بندی در فضای بازنمایی پنهان
- این روش روی مجموعه داده های متنوعی شامل کاراته، دلفین ها، فوتبال، کتاب های سیاسی، کرا و شهروند ارزیابی شد. نتایج نشان داد که در مجموعه داده فوتبال نسبت به ۱۲ الگوریتم قبلی دقت بالاتری حاصل شده و در سایر مجموعه داده ها نیز بهبود قابل توجهی گزارش شده است [۲۴].

۳-۳-۲. رویکرد ترکیبی kNN و بردار متوسط محلی

یک سناریوی پیشنهادی دیگر در ادبیات، ترکیب خوشه بندی گراف k نزدیک ترین همسایگی ((kNN با بردار متوسط محلی چندتایی در فضای بازنمایی پنهان خودرمزگذار است. در این رویکرد، پس از آموزش خودرمزگذار، بازنمایی پنهان هر گره استخراج شده

^۱ Autoencoder

^۲ Stacked Autoencoder

^۳ Xu

و سپس با ساخت گراف k -نزدیک‌ترین همسایگی در این فضا، انجمن‌ها شناسایی می‌شوند. این روش امکان تشخیص انجمن‌هایی با اشکال دلخواه را فراهم می‌کند که روش‌های مبتنی بر فاصله اقلیدسی ساده از آن ناتوانند.

مزایا و محدودیت‌های روش‌های یادگیری عمیق

روش‌های مبتنی بر یادگیری عمیق در تشخیص انجمن دارای مزایای قابل توجهی هستند:

- کاهش ابعاد مؤثر: توانایی فشرده‌سازی اطلاعات ساختاری شبکه در فضای پنهان کم‌بعد
- استخراج ویژگی خودکار: بدون نیاز به مهندسی دستی ویژگی‌ها
- انعطاف‌پذیری: قابلیت ترکیب با انواع روش‌های خوشه‌بندی
- مناسب برای شبکه‌های بزرگ: عملکرد بهتر نسبت به روش‌های کلاسیک در شبکه‌های پرابعاد

در مقابل، این روش‌ها با چالش‌هایی نیز روبرو هستند، از جمله نیاز به داده‌های آموزشی زیاد، هزینه محاسباتی بالای آموزش، وابستگی به کیفیت ماتریس شباهت اولیه، و تفسیرپذیری پایین مدل‌های یادگیری عمیق.

۴-۲. روش‌های مبتنی بر خوشه‌بندی

۴-۲-۱. خوشه‌بندی کلاسیک و پارتیشنی

روش‌های خوشه‌بندی کلاسیک از نخستین رویکردهایی بودند که برای تشخیص انجمن در شبکه‌های پیچیده به کار گرفته شدند. خوشه‌بندی k -میانگین (k -means) به عنوان شناخته‌شده‌ترین روش پارتیشنی، داده‌ها را به k خوشه تقسیم می‌کند به گونه‌ای که مجموع فاصله هر نقطه از مرکز خوشه متناظرش کمینه شود [۱۳]. فرآیند این الگوریتم به صورت تکراری انجام می‌شود: ابتدا k مرکز اولیه به صورت تصادفی انتخاب می‌شود، سپس هر گره به نزدیک‌ترین مرکز تخصیص می‌یابد و مراکز بر اساس میانگین اعضای هر خوشه به روزرسانی می‌شوند. این فرآیند تا همگرایی ادامه می‌یابد.

با این حال، k -means دارای محدودیت‌های اساسی در تشخیص انجمن شبکه‌های پیچیده است:

- نیاز به تعیین پیشین k : تعداد انجمن‌ها باید از پیش مشخص باشد، در حالی که در اغلب شبکه‌های واقعی این اطلاعات در دسترس نیست.

- فرض کروی بودن خوشه‌ها: این الگوریتم فرض می‌کند خوشه‌ها شکل کروی دارند، در حالی که انجمن‌های شبکه‌های اجتماعی اغلب اشکال نامنظم دارند.

- حساسیت به نقاط پرت: وجود گره‌های پرت^۱ می‌تواند مراکز خوشه را به شدت تحت تأثیر قرار دهد.

- ضعف در شبکه‌های پرابعاد: با افزایش ابعاد فضای ویژگی، کارایی k -means به دلیل پدیده «لعنت ابعاد» کاهش می‌یابد [۱۳].

خوشه‌بندی سلسله‌مراتبی رویکرد دیگری است که یا به صورت تجمیعی^۲ از پایین به بالا یا تقسیمی^۳ از بالا به پایین عمل می‌کند. در رویکرد تجمیعی، هر گره ابتدا یک خوشه مستقل است و در هر مرحله دو خوشه نزدیک‌تر ادغام می‌شوند تا یک درخت سلسله‌مراتبی (دندروگرام) شکل گیرد. این روش نیازی به تعیین پیشین تعداد خوشه‌ها ندارد، اما پیچیدگی زمانی $O(n^2 \log n)$ آن برای شبکه‌های بزرگ مقیاس مناسب نیست [۲۵].

^۱ outlier

^۲ agglomerative

^۳ divisive

۲-۴-۲. خوشه‌بندی مبتنی بر چگالی: DBSCAN

DBSCAN یکی از مهم‌ترین الگوریتم‌های خوشه‌بندی مبتنی بر چگالی است که مزایای قابل توجهی نسبت به روش‌های پارتیشن

دارد [۲۶]. اصل کار این الگوریتم بر پایه دو پارامتر اصلی استوار است:

- ϵ (اپسیلون): شعاع همسایگی هر نقطه
 - $MinPts$: حداقل تعداد نقاط لازم برای تشکیل یک ناحیه پرچگال
- بر اساس این دو پارامتر، گره‌ها به سه دسته تقسیم می‌شوند:
- **گره هسته^۱**: گره‌ای که در شعاع ϵ آن حداقل $MinPts$ گره وجود دارد.
 - **گره مرزی^۲**: گره‌ای که در همسایگی یک گره هسته قرار دارد اما خود هسته نیست.
 - **نقطه پرت^۳**: گره‌ای که نه هسته است و نه مرزی.
- مزایای اصلی DBSCAN در تشخیص انجمن عبارتند از:
- عدم نیاز به تعیین پیشین تعداد انجمن‌ها: این ویژگی برای شبکه‌های واقعی که تعداد جوامع ناشناخته است، بسیار حیاتی است.
 - شناسایی انجمن‌های با شکل دلخواه: برخلاف DBSCAN،-meansk می‌تواند خوشه‌های با هر شکل هندسی را تشخیص دهد.
 - مقاومت در برابر نقاط پرت: گره‌های پرت به‌عنوان نویز شناسایی و از خوشه‌بندی حذف می‌شوند.

در پژوهش حاضر، DBSCAN به‌دلیل حجم بالای گره‌ها و نامشخص بودن تعداد دقیق جوامع به‌عنوان بخشی از روش پیشنهادی DBSCAN+RSA برای خوشه‌بندی اولیه شبکه انتخاب شده است. این الگوریتم روی ویژگی‌های استخراج‌شده از گره‌ها نظیر لیست همسایگان اعمال می‌شود.

۲-۴-۳. فاکتورسازی ماتریس نامنفی (NMF)

تانگ و همکاران روشی مبتنی بر فاکتورسازی ماتریس نامنفی (NMF) با شباهت‌های چندمرحله‌ای برای تشخیص انجمن‌های

پیوندی^۴ ارائه دادند. در این رویکرد، ماتریس مجاورت شبکه A به‌صورت زیر تجزیه می‌شود:

$$A \approx W \cdot H, W \geq 0, H \geq 0 \quad (3)$$

که در آن W ماتریس پایه و H ماتریس ضرایب است. هر ستون از H بازنمایی یک گره در فضای انجمن را نشان می‌دهد. نوآوری اصلی این روش استفاده از شباهت‌های چندمرحله‌ای است که اطلاعات ساختاری در سطوح مختلف همسایگی را در ماتریس شباهت ترکیب می‌کند:

$$S^{(t)} = \alpha S^{(1)} + (1 - \alpha) S^{(1)} \cdot S^{(1)} \quad (4)$$

این رویکرد امکان شناسایی انجمن‌هایی را فراهم می‌کند که گره‌ها می‌توانند به‌طور همزمان به چند انجمن تعلق داشته باشند

(انجمن‌های همپوشان) [۱۰].

¹ Core Point

² Border Point

³ Noise Point

⁴ Link Communities

۲-۴-۴. تشخیص انجمن در شبکه‌های پویا

آینود^۱ و همکاران رویکرد تشخیص چندمرحله‌ای را در شبکه‌های پویا مطرح کردند. در شبکه‌های پویا، ساختار شبکه در طول زمان تغییر می‌کند و روش‌های استاتیک قادر به ردیابی تحولات انجمن‌ها نیستند. رویکرد پیشنهادی آن‌ها با تقسیم بازه زمانی به پنجره‌های متوالی و اعمال الگوریتم تشخیص انجمن در هر پنجره، تغییرات ساختاری را ردیابی می‌کند. این روش برای شبکه‌های اجتماعی که روابط بین کاربران در طول زمان تکامل می‌یابند، کاربرد ویژه‌ای دارد [۲۷].

۲-۴-۵. ترکیب DBSCAN با بهینه‌سازی فراابتکاری

علی آبادی و همکاران رویکرد ترکیبی نوآورانه‌ای ارائه دادند که DBSCAN را با الگوریتم بهینه‌سازی Jaya و روش‌های یادگیری عمیق ادغام می‌کند. در این روش:

۱. ابتدا ویژگی‌های ساختاری گره‌ها با یادگیری عمیق استخراج می‌شود.
 ۲. سپس الگوریتم Jaya پارامترهای بهینه ϵ DBSCAN و $MinPts$ را به صورت خودکار تنظیم می‌کند.
 ۳. در نهایت DBSCAN با پارامترهای بهینه‌شده برای تشخیص انجمن اعمال می‌شود.
- نتایج این روش در مقایسه با روش‌های پیشین رقابتی بوده و نشان می‌دهد که بهینه‌سازی خودکار پارامترهای DBSCAN می‌تواند کیفیت تشخیص انجمن را به طور قابل توجهی بهبود بخشد [۷].

۲-۵. مروری بر مطالعات انجام‌شده در حوزه تشخیص جامعه در شبکه‌های اجتماعی

تشخیص جامعه یکی از مسائل بنیادی و چالش‌برانگیز در تحلیل شبکه‌های پیچیده است. جامعه به مجموعه‌ای از گره‌ها اطلاق می‌شود که چگالی ارتباطی درون گروهی آن‌ها به مراتب بیشتر از ارتباطات برون گروهی است [۲۸، ۲۹]. با توجه به گسترش روزافزون شبکه‌های اجتماعی و حجم بالای داده‌های تولیدشده در این شبکه‌ها، تشخیص جامعه کاربردهای گسترده‌ای در حوزه‌هایی نظیر سیستم‌های توصیه‌گر، پیش‌بینی پیوند، انتشار اطلاعات، امنیت شبکه و بازاریابی یافته است [۲۸، ۲۹].

در این راستا، پژوهشگران رویکردهای متنوعی را برای تشخیص جامعه پیشنهاد داده‌اند. طباطبایی و همکاران (۲۰۲۲) الگوریتمی مبتنی بر انتشار برچسب با در نظر گرفتن تأثیر گره‌های همسایه ارائه کردند. در روش پیشنهادی آن‌ها، گره‌های مهم‌تر با احتمال بیشتری برچسب خود را منتشر می‌کنند و میزان شباهت گره‌ها به پارامتر طول مسیر بین آن‌ها وابسته است. این الگوریتم با تعداد تکرار کمتر و در زمان کوتاه‌تر، جوامع همپوشان را با دقت قابل مقایسه با الگوریتم‌های پیشین شناسایی می‌کند و بر روی شبکه‌های واقعی و مصنوعی با گراف وزن‌دار (با وزن‌دهی جاکارد) و بدون وزن پیاده‌سازی شده است [۳۰].

اسمعیلی و همکاران (۲۰۲۰) الگوریتمی مبتنی بر هوش جمعی مارگاتر (GSO) برای تشخیص جوامع در شبکه‌های اجتماعی پویا معرفی کردند. در این روش، گره‌ها به صورت مارها کدگذاری شده و تابع تناسب برای ارزیابی صحت جامعه‌بندی به کار می‌رود. نتایج نشان داد که الگوریتم پیشنهادی از نظر معیارهای دقت، فراخوانی، F-measure و صحت نسبت به الگوریتم BAT عملکرد بهتری دارد [۳۱].

خفائی و همکاران (۲۰۲۲) روشی برای تشخیص انجمن‌های پایدار در شبکه‌های اجتماعی پویا با استفاده از گره‌های با نفوذ ارائه دادند. در این مطالعه، ویژگی‌های جدیدی بر پایه مرکزیت شبکه معرفی و اهمیت آن‌ها با تئوری مجموعه‌های راف ارزیابی شد. نتایج نشان داد

¹ Aynaud

که افزایش تعداد گره‌های محبوب در یک انجمن، همراه با کاهش ویژگی‌های تراکم، بینابینی و نزدیکی، تأثیر تعداد گره‌های محبوب بر پایداری انجمن را تقویت می‌کند [۳۲].

میرزایی و همکاران (۲۰۲۰) روش ترکیبی تشخیص ناهنجاری مبتنی بر تشخیص انجمن در گراف و انتخاب ویژگی را پیشنهاد دادند. این روش از ناهنجاری به‌عنوان اعضای ناسازگار در انجمن‌ها بهره می‌برد و با الگوریتم تشخیص و ترکیب انجمن‌های مشابه، گره‌های ناهنجار را شناسایی می‌کند. آزمایش‌ها بر روی دو مجموعه داده واقعی، قدرت تشخیص دقیق این روش را در مقایسه با روش‌های پیشرفته تأیید کرد [۳۳].

هوشمندنژاد و نوش‌فر (۱۳۹۵) روش SRank را برای تشخیص انجمن در شبکه‌های اجتماعی بر اساس شباهت ساختاری بین گره‌ها پیشنهاد کردند. این روش هر گره را تنها در یک انجمن قرار می‌دهد و در مقایسه با روش‌های BipartiteRank، SimRank و P-Rank از نظر دقت و زمان اجرا برتری نشان داد [۳۴].

بیات و همکاران (۲۰۲۳) نسخه گسسته‌ای از الگوریتم جستجوی هارمونی را برای کشف جوامع و ساختارهای ارتباطی در شبکه‌های اجتماعی ارائه کردند [۳۵]. آزمایش‌ها بر روی شبکه‌های مصنوعی و طبیعی نشان داد که روش پیشنهادی در مقایسه با روش‌های موجود عملکرد مناسبی دارد. در همین راستا، **شهریاری و همکاران (۱۳۹۵)** نیز الگوریتم جستجوی هارمونی را برای خوشه‌بندی داده‌ها به کار بردند و نشان دادند که این الگوریتم در به‌دست آوردن جواب‌های بهینه کارایی بالایی دارد [۳۶].

علیپور و همکاران (۲۰۲۲) الگوریتمی مبتنی بر یادگیری تقویتی چندعاملی برای تشخیص اجتماع در شبکه‌های پیچیده ارائه کردند. در این رویکرد، هر عامل موجودیتی مستقل با پارامترهای یادگیری متفاوت است و بر اساس همکاری بین عامل‌ها، جستجوی اجتماعات بهینه به‌صورت تکرارشونده انجام می‌شود. ارزیابی بر روی چهار شبکه واقعی و چندین شبکه مصنوعی نشان داد که الگوریتم پیشنهادی از نظر معیارهای Q-ماژولاریتی و NMI به‌ترتیب ۱۲.۳۳٪، ۸۵.۹٪ و بیش از ۲۱٪ بهتر از الگوریتم‌های مقایسه‌شده عمل می‌کند [۳۷].

شریف‌زاده و همکاران (۱۳۹۳) روشی جدید برای شناسایی گره‌های فعال و تأثیرگذار در شبکه‌های اجتماعی با استفاده از شاخص‌های مرکزیت (درجه، بینابینیت، نزدیکی و نیمه‌محلی) ارائه کردند. آن‌ها معیار جدیدی از ترکیب وزن‌دار این شاخص‌ها از طریق الگوریتم ژنتیک پیشنهاد دادند که کارایی و دقت بالاتری نشان داد [۳۸].

در حوزه خوشه‌بندی در شبکه‌های حسگر بی‌سیم، **کریمی و همکاران (۱۴۰۰)** الگوریتم DBCA را ارائه کردند که علاوه بر آگاهی از فاصله، آگاهی از تراکم گره‌ها را نیز لحاظ می‌کند و در شرایط مختلف عملیاتی ۲۵ تا ۴۵ درصد کارآمدتر از الگوریتم‌های پیشین است [۳۹]. **رضایی و همکاران (۱۳۹۵)** نیز الگوریتم مکان‌یابی پویا مبتنی بر خوشه (DCAFP) را برای شبکه‌های موردی متحرک پیشنهاد دادند که با استفاده از خوشه‌بندی محلی، مصرف انرژی و تعداد گره‌های مرده را کاهش می‌دهد [۴۰].

علیخانی و همکاران (۱۳۸۹) روشی پویا برای تشخیص ناهنجاری در شبکه‌های اقتضایی متحرک با پروتکل AODV ارائه کردند که از الگوریتم خوشه‌بندی وزن‌دار WFWC استفاده می‌کند و با به‌روزرسانی نمای عادی، نرخ تشخیص را افزایش و نرخ هشدار نادرست را کاهش می‌دهد [۴۱]. **عثمانی و همکاران (۱۳۸۸)** نیز الگوریتمی مقیاس‌پذیر برای مدیریت مکان گره‌های سیار در پروتکل‌های مسیریابی مبتنی بر مکان ارائه دادند که با بهره‌گیری از هوش گروهی در سطوح بالاتر سلسله‌مراتب، سربار به‌روزرسانی مکان را کاهش می‌دهد [۴۲]. **بهنامیان و همکاران (۱۳۹۷)** نیز الگوریتم ترکیبی جستجوی همسایگی متغیر-شبیه‌سازی تبرید را برای مسئله مکان‌یابی هاب در شبکه حمل‌ونقل چندوجهی پیشنهاد دادند که عملکرد مناسبی در مقایسه با الگوریتم‌های دیگر نشان داد [۴۳].

در مجموع، مرور ادبیات نشان می‌دهد که پژوهش‌های انجام‌شده در حوزه تشخیص جامعه از رویکردهای متنوعی بهره برده‌اند؛ از الگوریتم‌های انتشار برچسب و هوش جمعی گرفته تا یادگیری تقویتی، خوشه‌بندی مبتنی بر تراکم و روش‌های ترکیبی. با این حال، چالش‌هایی نظیر مقیاس‌پذیری در شبکه‌های بزرگ، تشخیص جوامع همپوشان، و پایداری در شبکه‌های پویا همچنان زمینه‌های فعال پژوهشی محسوب می‌شوند که پژوهش حاضر در راستای پاسخ به برخی از آن‌ها طراحی شده است.

۳. روش‌شناسی پژوهش

در این بخش، جزئیات روش پیشنهادی برای تشخیص جوامع در شبکه‌های اجتماعی ارائه می‌شود. رویکرد ما یک مدل ترکیبی دو مرحله‌ای است که در مرحله اول از خوشه‌بندی فازی جهت تخمین تعداد جوامع و در مرحله دوم از الگوریتم بهینه‌سازی جستجوی خزندگان (Crayfish Optimization Algorithm - COA) برای بهبود دقت افراز استفاده می‌کند.

۳-۱. مدل‌سازی سیستم

شبکه‌های اجتماعی به صورت یک گراف بدون جهت $G = \{V, E\}$ مدل می‌شوند که در آن V مجموعه گره‌ها (کاربران) و E مجموعه یال‌ها (تعاملات) است. هدف از تشخیص جامعه، افراز مجموعه V به زیرمجموعه‌های $\{C_1, C_2, \dots, C_k\}$ است به طوری که بیشترین شباهت درون‌خوشه‌ای و کمترین شباهت میان‌خوشه‌ای برقرار باشد.

۳-۲. فاز اول: تشخیص تعداد جوامع با DBSCAN

یکی از چالش‌های اصلی در تشخیص جوامع، عدم اطلاع قبلی از تعداد خوشه‌ها (K) است. در این پژوهش، از الگوریتم DBSCAN برای شناسایی اولیه جوامع استفاده شده است. برتری این روش نسبت به k -means، عدم نیاز به تعیین تعداد خوشه‌ها و توانایی شناسایی نقاط نویز است.

در این مرحله، بردار ویژگی هر گره بر اساس لیست همسایگان استخراج شده و الگوریتم بر اساس دو پارامتر شعاع همسایگی (ϵ) و حداقل نقاط (MinPoints) عمل می‌کند.

۳-۳. فاز دوم: بهینه‌سازی جوامع با الگوریتم جستجوی خزندگان (COA)

پس از تعیین ساختار اولیه، از الگوریتم COA برای یافتن بهترین افراز استفاده می‌شود. فرآیند بهینه‌سازی شامل مراحل زیر است:

۳-۳-۱. مقداردهی اولیه و ساختار راه‌حل

هر راه‌حل (کروموزوم/ذره) به صورت یک بردار نمایش داده می‌شود که طول آن برابر با تعداد گره‌های شبکه است. مقدار هر درایه، برچسب جامعه‌ای است که گره به آن تعلق دارد.

ماتریس جمعیت اولیه X به صورت تصادفی مطابق رابطه (۵) تشکیل می‌شود:

$$X = [x_{1,1} \dots x_{1,n} \ x_{2,1} \dots x_{2,n} \ \vdots \ x_{N,1} \dots x_{N,n}] \quad (5)$$

که در آن هر درایه $x_{i,j}$ با رابطه (۶) مقداردهی می‌شود:

$$x_{i,j} = rand \times (UB - LB) + LB \quad (6)$$

جدول ۱. ساختار یک نمونه راه‌حل در طرح پیشنهادی

شماره گره	۱	۲	۳	۴	۵	۶	۷
برچسب جامعه	۲	۲	۱	۳	۲	۱	۱

۳-۳-۲. مرحله اکتشاف^۱

در این مرحله، خزندگان برای یافتن مناطق جدید در فضای جستجو، دو رفتار «راه رفتن طولانی» و «راه رفتن شکمی» را انجام می دهند. این تغییر رفتار بر اساس زمان تکرار (t) مدیریت می شود:

$$x_{i,j}(t+1) = \{Best_j(t) \times -\eta_{i,j}(t) \times \beta - R_{i,j}(t) \times rand \mid t \leq \frac{T}{4} \mid Best_j(t) \times x_{r_1,j} \times ES(t) \times rand \mid \frac{T}{4} < t \leq \frac{2T}{4} \quad (7)$$

پارامترهای کلیدی در این مرحله عبارتند از:

- اپراتور شکار (η): مطابق رابطه (۸) محاسبه می شود:

$$\eta_{i,j} = Best_j(t) \times P_{i,j} \quad (8)$$

- تابع کاهش (R): برای محدود کردن فضای جستجو:

$$R_{i,j} = \frac{Best_j(t) - x_{r_2,j}}{Best_j(t) + \epsilon} \quad (9)$$

- ضریب تکاملی (ES):

$$ES(t) = 2 \times r_3 \times \left(1 - \frac{t}{T}\right) \quad (10)$$

- درصد اختلاف ($P_{i,j}$):

$$P_{i,j} = \alpha + \frac{x_{i,j} - M(x_i)}{Best_j(t) \times (UB_j - LB_j) + \epsilon} \quad (11)$$

که در آن $M(x_i)$ میانگین موقعیت های جواب i است:

$$M(x_i) = \frac{1}{n} \sum_{j=1}^n x_{i,j} \quad (12)$$

۳-۳-۳. مرحله بهره برداری^۲

در نیمه دوم تکرارها ($t > T/2$)، الگوریتم بر روی یافتن بهینه محلی در نزدیکی بهترین جواب تمرکز می کند. استراتژی های «هماهنگی شکار» و «مشارکت» اعمال می شوند:

$$x_{i,j}(t+1) = \{Best_j(t) \times P_{i,j}(t) \times rand \mid \frac{2T}{4} < t \leq \frac{3T}{4} \mid Best_j(t) \times \eta_{i,j}(t) \times \epsilon - R_{i,j}(t) \times rand \mid \frac{3T}{4} < t \leq T \quad (13)$$

۳-۴. تابع هدف و معیار ارزیابی^۳

برای ارزیابی کیفیت هر افراز پیشنهاد شده توسط الگوریتم COA، از یک تابع هدف ترکیبی شامل «شاخص پراکندگی^۴ و ماژولاریتی^۵

استفاده شده است.

^۱ Exploration Phase

^۲ Exploitation Phase

^۳ Fitness Function

^۴ Conductance

^۵ Modularity

۱. شاخص پراکندگی (C): میزان ارتباطات خروجی یک جامعه نسبت به تراکم داخلی آن را می‌سنجد. هدف، کمینه‌سازی این مقدار

است:

$$C(C_i) = \frac{E(C_i, V \setminus C_i)}{(E(C_i, V), E(V \setminus C_i, V))} \quad (14)$$

۲. ماژولاریتی (M): توانایی تقسیم‌بندی شبکه را می‌سنجد. مقادیر بالاتر نشان‌دهنده ساختار جامعه قوی‌تر است:

$$M = \sum_{k=1}^s \left[\frac{l_k}{L} - \left(\frac{d_k}{2L} \right)^2 \right] \quad (15)$$

۳. تابع برازندگی نهایی^۱: با ترکیب دو معیار فوق، تابع هدف برای کل ساختار جوامع به صورت زیر تعریف می‌شود که الگوریتم سعی

در کمینه‌سازی آن دارد:

$$Fitness(S) = \sum \frac{C(C_i)}{M(C_i)} \quad (16)$$

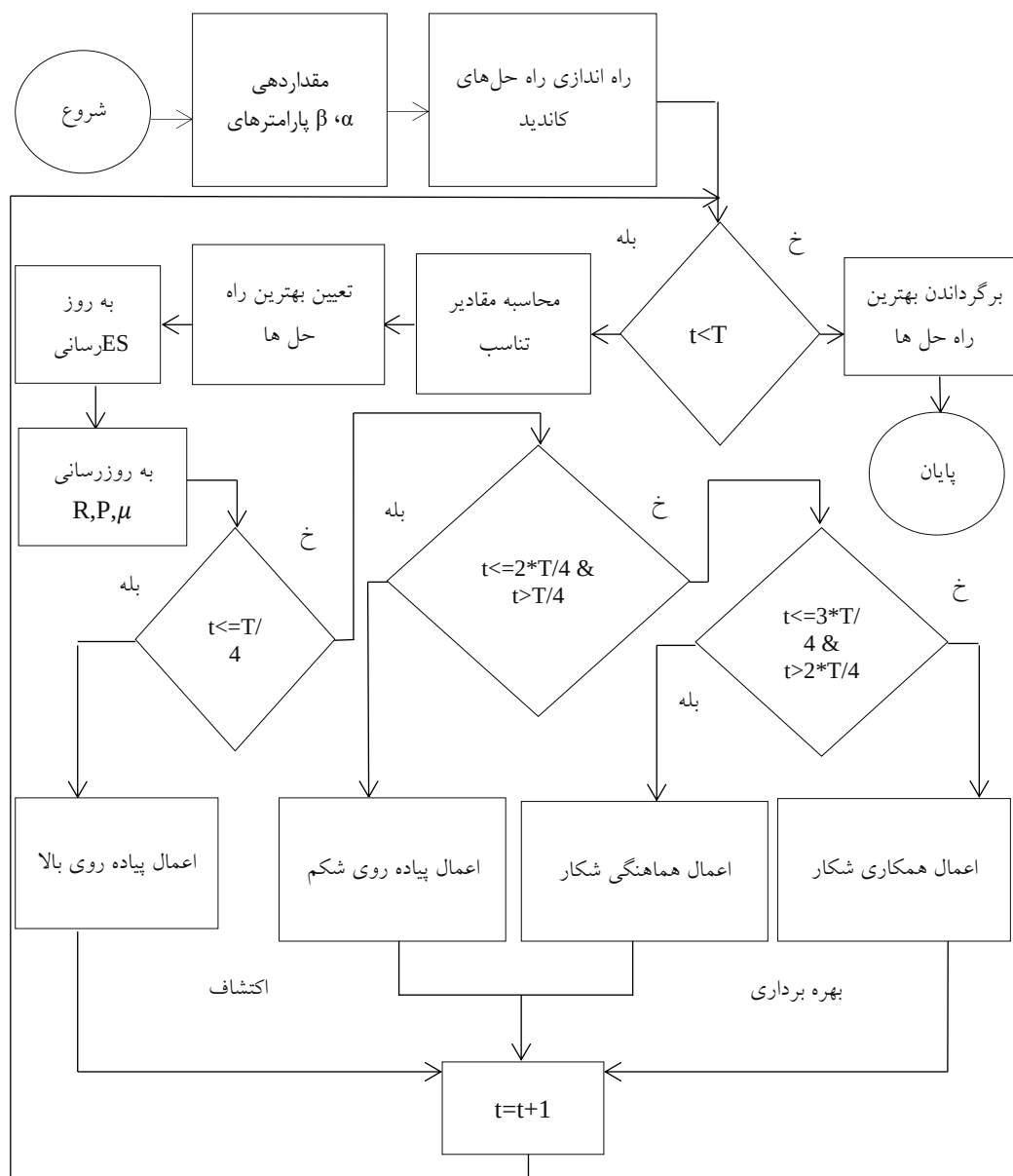
۳-۵. الگوریتم اجرایی و فلوچارت

مراحل پیاده‌سازی روش پیشنهادی در جدول (۲) و شکل (۱) نمایش داده شده است.

جدول ۲. شبه کد الگوریتم جستجوی خزندگان برای حل تابع هدف

- 1: Initialization phase
- 2: Initialize RSA parameters α, β , etc
- 3: Initialize the solutions position randomly. $X: i = 1, \dots, N$.
- 4: while ($t < T$) do
- 5: Calculate the fitness function for the candidate solutions (X).
- 6: Find the Best solution so far.
- 7: Update the ES
- 8: The beginning of the RSA
- 9: for($i = 1$ to N) do
- 10: for($i = 1$ to n) do
- 11: Update the η, R, P and values
- 12: if $\left(t \leq \frac{T}{4} \right)$ then
- 13: $x_{(i,j)}(t+1) = Best_j(t) \times -\eta_{(i,j)}(t) \times \beta - R_{(i,j)}(t) \times rand, \triangleright \{High\ walking\}$
- 14: else if $\left(t \leq 2\frac{T}{4} \text{ and } t > \frac{T}{4} \right)$ then
- 15: $x_{(i,j)}(t+1) = Best_j(t) \times x_{(r_1,j)} \times ES(t) \times rand, \triangleright \{Belly\ walking\}$
- 16: else if $\left(t \leq 3\frac{T}{4} \text{ and } t > 2\frac{T}{4} \right)$ then
- 17: $x_{(i,j)}(t+1) = Best_j(t) \times P_{(i,j)}(t) \times rand, \triangleright \{Hunting\ coordination\}$
- 18: else
- 19: $x_{(i,j)}(t+1) = Best_j(t) - \eta_{(i,j)}(t) \times \epsilon - R_{(i,j)}(t) \times rand, \triangleright \{Hunting\ cooperation\}$
- 20: end if
- 21: end for

¹ Fitness



شکل ۱. فلوچارت الگوریتم جستجوی خزندگان برای حل تابع هدف پیشنهادی

۳-۶. تحلیل پیچیدگی

روش پیشنهادی با بهره‌گیری از DBSCAN در مرحله پیش‌پردازش، فضای جستجوی الگوریتم فراابتکاری را به شدت محدود می‌کند. در حالی که روش‌های سنتی به صورت کاملاً تصادفی جستجو را آغاز می‌کنند، استفاده از دانش اولیه^۱ حاصل از چگالی گره‌ها، سرعت همگرایی COA را افزایش داده و از گرفتار شدن در بهینه‌های محلی ناشی از تعداد نامناسب جوامع جلوگیری می‌کند.

۴. ارزیابی و تجزیه و تحلیل نتایج

در شبکه‌های اجتماعی، گروه‌هایی وجود دارند که در آن‌ها ارتباطات داخلی میان اعضا بسیار بیشتر از ارتباطات با افراد خارج از گروه است. این گروه‌ها که با عنوان «جوامع» شناخته می‌شوند، ممکن است بر پایه علایق مشترک، تعاملات اجتماعی، یا فعالیت‌های مشابه شکل

^۱ Prior Knowledge

گرفته باشند. فرآیند شناسایی این جوامع شامل یافتن گروه‌هایی از کاربران است که ارتباطات داخلی قوی‌تری نسبت به ارتباطات خارجی دارند و به تحلیل ساختار شبکه‌های اجتماعی کمک کرده و دیدگاه‌های ارزشمندی در مورد رفتارها و تعاملات کاربران ارائه می‌دهد.

شناخت جوامع به درک بهتر ساختار شبکه و نحوه ارتباطات میان کاربران منجر می‌شود. شرکت‌ها می‌توانند از این اطلاعات برای طراحی تبلیغات هدفمند و استراتژی‌های بازاریابی بهره گیرند. همچنین، پژوهشگران می‌توانند با استفاده از این داده‌ها رفتارهای اجتماعی را تحلیل کرده و پویایی گروه‌ها را بررسی کنند. در حوزه امنیت و نظارت نیز این اطلاعات می‌تواند برای شناسایی الگوهای مشکوک یا تهدیدات احتمالی به کار گرفته شود. با گسترش شبکه‌های اجتماعی و افزایش حجم داده‌ها، شناسایی جوامع به صورت کارآمد به چالشی بزرگ تبدیل شده است. تنوع تعاملات و ارتباطات در شبکه‌ها، به همراه تغییرات مداوم در ساختار آن‌ها، شناسایی دقیق جوامع را پیچیده‌تر می‌کند. در مجموع، شناسایی جوامع ابزاری کلیدی برای فهم عمیق‌تر از ساختار و تعاملات شبکه‌های اجتماعی است و می‌تواند به تحلیل داده‌ها و تصمیم‌گیری بهتر در حوزه‌های مختلف کمک کند.

۴-۱. مجموعه داده‌های مورد استفاده

برای شبیه‌سازی روش پیشنهادی از نرم‌افزار MATLAB نسخه R2023b استفاده گردید. MATLAB یک محیط نرم‌افزاری پیشرفته برای محاسبات عددی و یک زبان برنامه‌نویسی سطح بالا است که کار با ماتریس‌ها در آن بسیار راحت بوده و به‌طور کلی تمام داده‌ها به صورت ماتریس ذخیره می‌شوند. مشخصات سیستم مورد استفاده برای اجرای این روش در جدول ۳ ارائه شده است.

جدول ۳. مشخصات سخت‌افزاری و نرم‌افزاری زیرساخت محاسباتی

مشخصات	مقدار
CPU	Core i4
RAM	8 GB
OS	Windows 8
MATLAB	R2023b

برای ارزیابی عملکرد طرح پیشنهادی، از جدیدترین روش‌های تشخیص اجتماع شامل روش قره‌چوپوق [۵]، روش Gmati و همکاران [۶] و روش علی‌آبادی و همکاران [۷] استفاده شد. در این بررسی، مجموعه داده‌های مختلف و شناخته‌شده در حوزه شبکه‌های پیچیده به کار گرفته شد.

در ابتدا، از پنج مجموعه داده فاقد برچسب شامل ca-GrQc، ca-HepPh، ca-HepTh، ca-AstroPh و ca-CondMat که توسط گروه شبکه استنفورد توسعه یافته‌اند، استفاده گردید. ویژگی‌های این مجموعه داده‌ها در جدول ۴ ارائه شده است. سپس، برای محاسبه معیار NMI و نیاز به داشتن افراز واقعی گره‌ها در اجتماعات، پنج مجموعه داده دارای برچسب شامل Football، Dolphins، Karate، Polblogs و Yeast مورد استفاده قرار گرفتند. جزئیات مربوط به نوع و اندازه این شبکه‌ها در جدول ۵ آورده شده است. در این جدول، پارامتر n نمایانگر تعداد کل گره‌ها، m نشان‌دهنده تعداد کل یال‌ها و متغیر C معرف میانگین ضریب خوشه‌بندی در هر شبکه است.

جدول ۴. مجموعه داده‌های چندلایه در طرح پیشنهادی

میانگین درجه	تعداد یال‌ها	تعداد گره‌ها	نام شبکه
۵۲/۵	۱۴۴۹۶	۵۲۴۲	ca-GrQc
۲۶/۵	۲۵۹۹۸	۹۸۸۷	ca-HepTh
۷۴/۱۹	۱۱۸۵۲۱	۱۲۰۰۸	ca-HepPh
۱۰/۲۱	۱۹۸۰۸۰	۱۸۷۷۲	ca-AstroPh
۰۸/۸	۹۳۴۹۷	۲۳۱۳۳	ca-CondMat

جدول ۵. مجموعه داده های تک لایه و واقعی در طرح پیشنهادی

نام شبکه	n	m	C
Karate	۳۴	۷۸	۵۸/۰
Dolphins	۶۲	۱۵۹	۳/۰
Football	۱۱۵	۶۱۳	۴/۰
polblogs	۶۴۳	۲۳۰۰	۲۳/۰
yeast	۱۵۰۰	۱۹۰۰	۰۷/۰

در ادامه، توضیح مختصری از هر یک از مجموعه داده های دارای برجسب ارائه می شود:

Karate: این مجموعه داده در دهه ۱۹۷۰ توسط وین زاکری از تعاملات اجتماعی اعضای یک باشگاه کاراته در دانشگاهی در آمریکا گردآوری شد. زاکری طی دو سال رفتار اجتماعی ۳۴ عضو این باشگاه را ثبت کرد و شبکه ای با ۳۴ گره و ۷۸ یال بر اساس روابط دوستی میان اعضا ایجاد کرد. در این شبکه، گره ها نمایانگر افراد و یال ها نشان دهنده روابط دوستی قوی بین آنها هستند.

Dolphins: این شبکه اجتماعی، روابط میان دلفین های با پوزه بطری شکل در آبدرهای در نیوزیلند را نمایش می دهد. گره ها نمایانگر دلفین ها و یال ها نشان دهنده تعاملات مکرر میان آنها هستند. داده های مربوط به این شبکه توسط زیست شناسی به نام David Lusseau طی یک دوره هفت ساله بررسی رفتار دلفین های bottlenose در منطقه Doubtful Sound در نیوزیلند گردآوری شد. این شبکه شامل ۶۲ گره و ۱۵۹ یال است.

Football: این شبکه بازی های فوتبال آمریکایی بین کالج های مختلف را در فصل پاییز سال ۲۰۰۰ بازنمایی می کند.

۴-۲. معیارهای ارزیابی عملکرد

۴-۲-۱. معیار اطلاعات متقابل نرمال شده (NMI)

یکی از معتبرترین معیارها در ارزیابی الگوریتم های تشخیص اجتماع در شبکه های پیچیده، معیار اطلاعات متقابل نرمال شده یا NMI^1 است که در مواردی به کار می رود که ساختار اجتماع واقعی در اختیار باشد. مقدار NMI در بازه [1] قرار می گیرد و مقدار بزرگ تر آن نشان دهنده مطابقت بیشتر اجتماعات شناسایی شده با واقعیت است.

این معیار بر اساس ماتریسی به نام پارامتر میکس^۲ عمل می کند. سطرهای این ماتریس متناظر با اجتماعات واقعی و ستون های آن متناظر با اجتماعات شناسایی شده هستند. عنصر C_{ij} نشان دهنده تعداد گره های مشترک در اجتماع واقعی i و اجتماع شناسایی شده j است. با فرض اینکه A و B دو افراز مختلف از شبکه ای با N گره باشند، C_A و C_B به ترتیب تعداد اجتماعات افرازهای A و B را نشان دهند، و C_i و C_j به ترتیب حاصل جمع عناصر سطر i ام و ستون j ام پارامتر میکس باشند، معیار NMI بر اساس تئوری اطلاعات به صورت رابطه (۱۷) تعریف می شود:

$$NMI(A, B) = \frac{-2 \sum_{i=1}^{C_A} \sum_{j=1}^{C_B} C_{ij} \log \left(\frac{C_{ij} N}{C_i C_j} \right)}{\sum_{i=1}^{C_A} C_i \log \left(\frac{C_i}{N} \right) + \sum_{j=1}^{C_B} C_j \log \left(\frac{C_j}{N} \right)} \quad (17)$$

¹ Normalized Mutual Information

² Confusion Matrix

۴-۲-۲- شاخص پیمانی

شاخص دیگری که در این پژوهش مورد استفاده قرار گرفته، شاخص پیمانی است که فرمول آن پیش‌تر در بخش طرح پیشنهادی به عنوان تابع هدف معرفی گردید. این شاخص کیفیت تقسیم‌بندی شبکه به اجتماعات را بدون نیاز به دانستن ساختار واقعی اجتماعات اندازه‌گیری می‌کند.

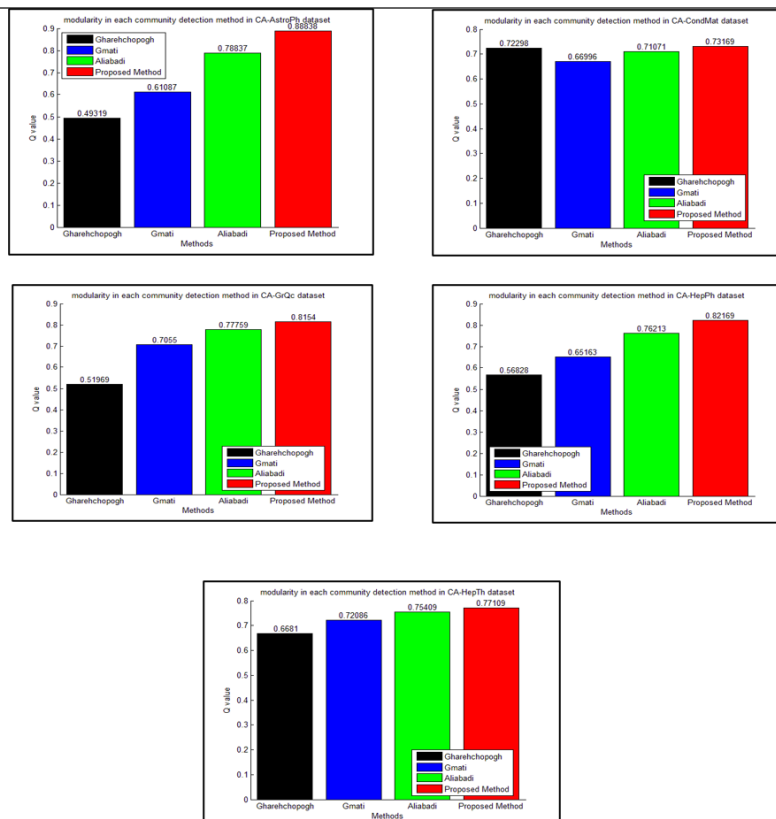
۴-۳. نتایج ارزیابی

۴-۳-۱. ارزیابی شاخص پیمانی روی مجموعه داده‌های فاقد برچسب

عملکرد طرح پیشنهادی با استفاده از رویکردهای پایه و با توجه به شاخص‌های پیمانی و NMI ارزیابی شد. در این بررسی، کارایی روش یک‌بار بدون در نظر گرفتن جامعه اصلی گره‌ها و بار دیگر با لحاظ کردن آن مورد تحلیل قرار گرفت؛ در حالت اول از شاخص پیمانی و در حالت دوم از شاخص NMI بهره‌گیری شد. رویکرد پیشنهادی برای هر مجموعه داده ۳۰ مرتبه اجرا و نتایج ثبت گردید. تحلیل نتایج که در شکل ۲ ارائه شده است نشان می‌دهد بالاترین مقدار میانگین پیمانی مربوط به مجموعه داده CA-AstroPh و پایین‌ترین مقدار آن مربوط به مجموعه داده CA-CondMat است. جزئیات این بررسی در جدول ۶ آمده است.

جدول ۶. ارزیابی شاخص پیمانی طرح پیشنهادی روی مجموعه داده‌های فاقد برچسب

روش	CA-GrQc	CA-HepTh	CA-HepPh	CA-AstroPh	CA-CondMat
Gharehchopogh	۵۱۹۶/۰	۶۶۸۱/۰	۵۶۸۲/۰	۴۹۱۳/۰	۷۲۲۹/۰
Gmati	۷۰۵۵/۰	۷۲۰۸/۰	۶۵۱۶/۰	۶۱۰۸/۰	۶۶۹۹/۰
Aliabadi	۷۷۷۵/۰	۷۵۴۰/۰	۷۶۲۱/۰	۷۸۸۳/۰	۷۱۰۷/۰
Proposed Method	۸۱۵۴/۰	۷۷۱۰/۰	۸۲۱۶/۰	۸۸۸۳/۰	۷۳۱۶/۰



شکل ۲. ارزیابی شاخص مازولاریتی طرح پیشنهادی در مجموعه داده‌های فاقد برچسب

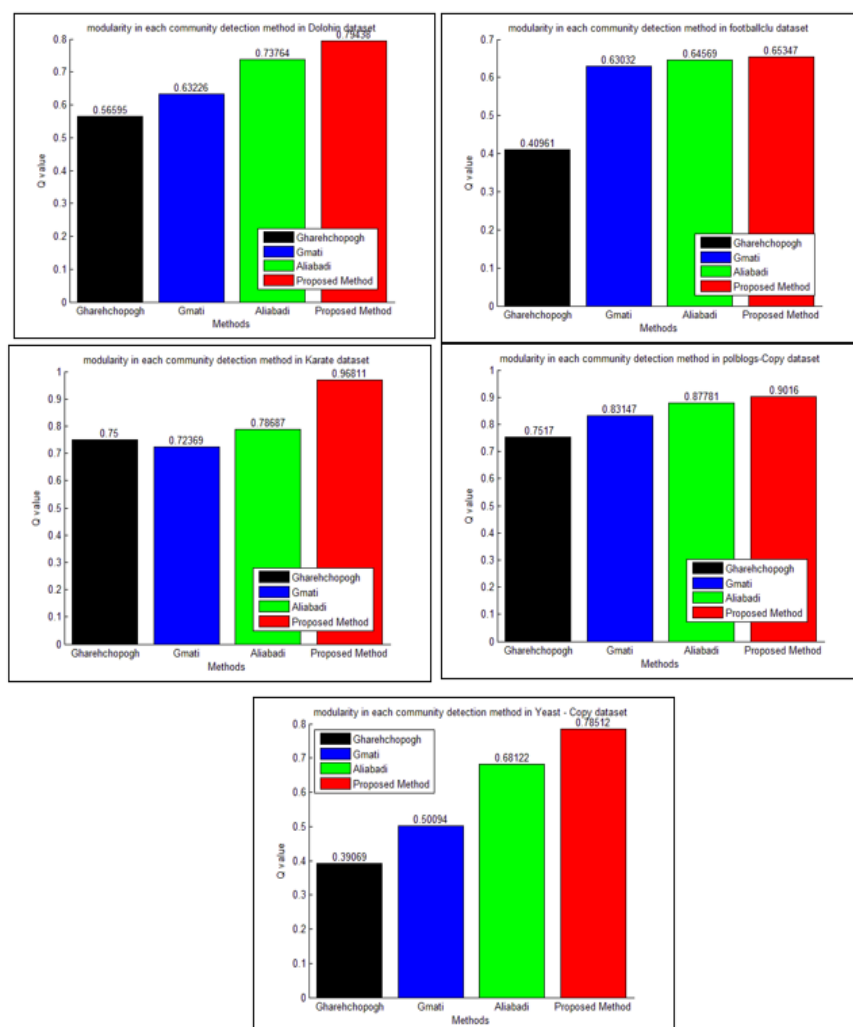
همان طور که در جدول ۶ مشاهده می شود، طرح پیشنهادی در تمامی مجموعه داده ها عملکرد بهتری نسبت به روش های پایه داشته است. بیشترین بهبود در میزان پیمانی مربوط به مجموعه داده CA-AstroPh با افزایش حدود ۱۰ درصد نسبت به بهترین روش پایه بوده و کمترین بهبود برای مجموعه داده CA-HepTh با حدود ۲ درصد ثبت شده است.

۴-۳-۲. ارزیابی شاخص پیمانی روی مجموعه داده های دارای برچسب

در بررسی دیگری، رویکرد پیشنهادی به ازای مجموعه داده های تک لایه دارای برچسب اجتماع (جدول ۵) نیز ۳۰ مرتبه اجرا و نتایج ثبت گردید. نتایج شبیه سازی در جدول ۷ نشان می دهد که طرح پیشنهادی نسبت به روش های پایه بهبود یافته و توانسته است مقدار پیمانی را در تمامی مجموعه داده ها افزایش دهد.

جدول ۷. ارزیابی شاخص پیمانی طرح پیشنهادی

روش	Karate	polblogs	Dolphin	Footbal	Yeast
Gharehchopogh	۷۵/۰	۷۵۱۷/۰	۵۶۵۹/۰	۴۰۹۶/۰	۳۹۰۶/۰
Gmati	۷۲۳۶/۰	۸۳۱۴/۰	۵۳۲۲/۰	۶۳۰۳/۰	۵۰۰۹/۰
Aliabadi	۷۸۶۸/۰	۸۷۷۸/۰	۷۳۷۶/۰	۶۴۵۶/۰	۶۸۱۲/۰
Proposed Method	۹۶۸۱/۰	۹۰۱۶/۰	۷۹۴۳/۰	۶۵۳۴/۰	۷۸۵۱/۰



شکل ۳. ارزیابی شاخص پیمانی طرح پیشنهادی در مجموعه داده های دارای برچسب

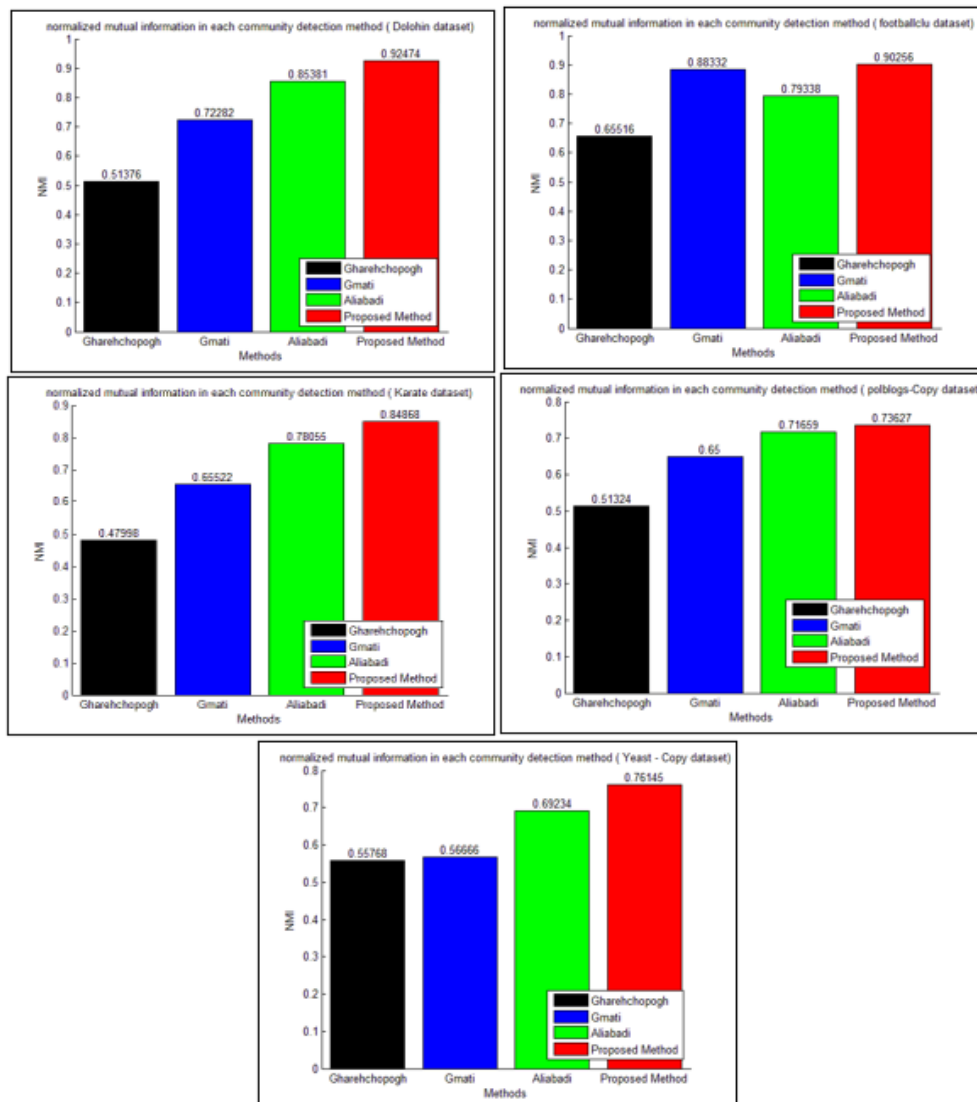
بر اساس نتایج جدول ۷، طرح پیشنهادی در مجموعه داده Karate با مقدار پیمانی ۰.۹۶۸۱ بیشترین بهبود را نسبت به روش‌های پایه نشان می‌دهد. همچنین در مجموعه داده Yeast نیز با مقدار ۰.۷۸۵۱ عملکرد قابل توجهی نسبت به سایر روش‌ها حاصل شده است.

۳-۳-۴. ارزیابی شاخص NMI

برای ارزیابی شاخص NMI، رویکرد پیشنهادی به ازای مجموعه داده‌های تک‌لایه دارای برجسب ۳۰ مرتبه اجرا شده و نتایج ثبت گردید. بررسی نتایج نشان می‌دهد که بالاترین مقدار NMI به صورت میانگین برای مجموعه داده Dolphins و پایین‌ترین مقدار آن برای مجموعه داده Polblogs حاصل شده است. جدول ۸ نتایج این بررسی را نمایش می‌دهد.

جدول ۸. ارزیابی شاخص NMI طرح پیشنهادی

روش	Karate	polblogs	Dolphin	Footbal	Yeast
Gharehchopogh	۴۷۹۹/۰	۵۱۳۲/۰	۵۱۳۷/۰	۶۵۵۱/۰	۵۵۷۶/۰
Gmati	۶۵۵۲/۰	۶۵/۰	۷۲۲۸/۰	۸۸۳۳/۰	۵۶۶۶/۰
Aliabadi	۷۸۰۵/۰	۷۱۶۵/۰	۸۵۳۸/۰	۷۹۳۳/۰	۶۹۲۳/۰
Proposed Method	۸۴۸۶/۰	۷۳۶۲/۰	۹۲۴۷/۰	۹۰۲۵/۰	۷۶۱۴/۰



شکل ۴. ارزیابی شاخص NMI طرح پیشنهادی در مجموعه داده‌های دارای برجسب

۵. نتیجه گیری

در این پژوهش، رویکردی ترکیبی مبتنی بر الگوریتم خوشه‌بندی DBSCAN و روش بهینه‌سازی جستجوی خزنده (RSA) برای شناسایی جوامع در شبکه‌های اجتماعی ارائه شد. در مرحله نخست، به دلیل حجم بالای گره‌ها و عدم آگاهی پیشین از تعداد جوامع، خوشه‌بندی اولیه‌ای بر اساس ویژگی‌های استخراج‌شده از گره‌ها، از جمله لیست همسایگان، با استفاده از روش خوشه‌بندی فازی DBSCAN انجام گرفت. در مرحله دوم، الگوریتم RSA برای بهینه‌سازی تشخیص جوامع به کار گرفته شد. کارایی روش پیشنهادی با استفاده از معیارهای پیمانی (Modularity) و اطلاعات متقابل نرمال‌شده (NMI) بر روی ده مجموعه داده واقعی شامل شبکه‌های فاقد برچسب (ca-ca-GrQc) و (ca-AstroPh، ca-HepTh، HepPh و Polblogs، Football، Dolphins، Karate) و شبکه‌های دارای برچسب (Gmatl و همکاران، و علی‌آبادی و همکاران مقایسه شد.

نتایج ارزیابی نشان داد که روش پیشنهادی در تمامی مجموعه داده‌ها از نظر شاخص پیمانی عملکرد بهتری نسبت به روش‌های پایه داشته است. بیشترین بهبود در مجموعه داده‌های فاقد برچسب مربوط به CA-AstroPh با افزایش حدود ۱۰ درصد نسبت به بهترین روش پایه بود، در حالی که در مجموعه داده‌های دارای برچسب، مجموعه Karate با مقدار پیمانی ۰.۹۶۸۱ برجسته‌ترین بهبود را نشان داد. از منظر شاخص NMI نیز روش پیشنهادی در تمامی مجموعه داده‌های دارای برچسب از دقت بالاتری برخوردار بود؛ بیشترین بهبود به ترتیب در مجموعه داده‌های Football و Dolphins با افزایش حدود ۱۱ و ۷ درصد نسبت به بهترین روش پایه حاصل شد. این نتایج تأیید می‌کنند که رویکرد پیشنهادی توانایی بالاتری در شناسایی ساختار واقعی جوامع شبکه دارد.

در مجموع، ترکیب خوشه‌بندی DBSCAN با الگوریتم بهینه‌سازی RSA به عنوان یک چارچوب مؤثر برای تشخیص جوامع در شبکه‌های اجتماعی پیچیده معرفی می‌شود. در مطالعات آتی، بهره‌گیری از ویژگی‌های نوین استخراج‌شده از تعاملات کاربران، توسعه روش‌های شناسایی مرکزیت گره‌ها، و به‌کارگیری الگوریتم‌های فراابتکاری پیشرفته‌تر نظیر الگوریتم کرکس آفریقایی (AVOA) می‌تواند به ارتقاء بیشتر کیفیت تحلیل و مقیاس‌پذیری سیستم‌های شناسایی جوامع در شبکه‌های بزرگ‌مقیاس منجر شود.

تعارض منافع

در انجام مطالعه حاضر، هیچ‌گونه تضاد منافع وجود ندارد.

مشارکت نویسندگان

در نگارش این مقاله تمامی نویسندگان نقش یکسانی ایفا کردند.

موازین اخلاقی

در انجام این پژوهش تمامی موازین و اصول اخلاقی رعایت گردیده است.

شفافیت داده‌ها

داده‌ها و مآخذ پژوهش حاضر در صورت درخواست از نویسنده مسئول و ضمن رعایت اصول کپی رایت ارسال خواهد شد.

References

1. Plantié, M. and M. Crampes, *Survey on Social Community Detection*, in *Social Media Retrieval*, P. Springer, Editor. 2013, Springer Publishers. p. 65–85.
2. Aghaalizadeh, S., et al., *A three-stage algorithm for local community detection based on the high node importance ranking in social networks*. *Physica A: Statistical Mechanics and its Applications*, 2021. **563**: p. 125420.
3. Zhang, W., R. Shang, and L. Jiao, *Large-scale community detection based on core node and layer-by-layer label propagation*. *Information Sciences*, 2023. **632**: p. 1–18.
4. Ullah, A., et al., *A novel relevance-based information interaction model for community detection in complex networks*. *Expert Systems with Applications*, 2022. **196**: p. 116607.
5. Gharehchopogh, F.S., *An improved Harris Hawks optimization algorithm with multi-strategy for community detection in social network*. *Journal of Bionic Engineering*, 2023. **20**(3): p. 1175–1197.
6. Gmati, H., et al., *A new algorithm for communities detection in social networks with node attributes*. *Journal of Ambient Intelligence and Humanized Computing*, 2024. **15**(2): p. 1779–1791.
7. Aliabadi, M. and H. Ghaffari, *Community detection using Jaya optimization algorithm based on deep learning methods and KNN graph-based clustering*. *Journal of Modelling in Management*, 2025. **20**(3): p. 872–893.
8. Xu, R., et al., *Stacked autoencoder-based community detection method via an ensemble clustering framework*. *Information sciences*, 2020. **526**: p. 151–165.
9. Park, N., et al., *CGC: Contrastive Graph Clustering for Community Detection and Tracking*. 2022.
10. Tang, X., et al., *Link community detection by non-negative matrix factorization with multi-step similarities*. *Modern Physics Letters B*, 2016. **30**(32n33): p. 1650370.
11. Zhang, D., et al., *A novel two-step community detection approach based on community tree and the N-players cooperative game in large-scale social networks*. *Journal of Computational Methods in Science and Engineering*, 2018. **18**(4): p. 1007–1020.
12. Aynaud, T. and J.-L. Guillaume, *Multi-Step Community Detection and Hierarchical Time Segmentation in Evolving Networks*. *Âl' ACM*, 2011.
13. Jain, A.K., M.N. Murty, and P.J. Flynn, *Data clustering: a review*. *ACM computing surveys (CSUR)*, 1999. **31**(3): p. 264–323.
14. Saxena, A., et al., *A review of clustering techniques and developments*. *Neurocomputing*, 2017. **267**: p. 664–681.
15. Liu, Y., et al., *Hierarchical community discovery for multi-stage IP bearer network upgradation*. *Journal of Network and Computer Applications*, 2021. **189**: p. 103151.
16. Li, H., et al., *LMFLS: A new fast local multi-factor node scoring and label selection-based algorithm for community detection*. *Chaos, Solitons & Fractals*, 2024. **185**: p. 115126.
17. Abualigah, L., et al., *Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer*. *Expert Systems with Applications*, 2022. **191**: p. 116158.
18. Bouyer, A. and H. Roghani, *LSMD: A fast and robust local community detection starting from low degree nodes in social networks*. *Future Generation Computer Systems*, 2020. **113**: p. 41–57.
19. Ding, X., J. Zhang, and J. Yang, *A robust two-stage algorithm for local community detection*. *Knowledge-Based Systems*, 2018. **152**: p. 188–199.
20. Newman, M.E., *Modularity and community structure in networks*. *Proc Natl Acad Sci U S A*, 2006. **103**(23): p. 8577–82.
21. Blondel, V., et al., *Fast Unfolding of Communities in Large Networks*. *Journal of Statistical Mechanics Theory and Experiment*, 2008. **2008**.
22. Rosvall, M. and C.T. Bergstrom, *Maps of random walks on complex networks reveal community structure*. *Proc Natl Acad Sci U S A*, 2008. **105**(4): p. 1118–23.
23. Girvan, M. and M.E. Newman, *Community structure in social and biological networks*. *Proc Natl Acad Sci U S A*, 2002. **99**(12): p. 7821–6.
24. حسینی, م. and ا. گلوی, *تشخیص اجتماع در شبکه‌های اجتماعی با رویکرد یادگیری عمیق*. *مطالعات مدیریت کسب و کار هوشمند*, ۲۰۲۳. **12**(44): p. 83–112.
25. Bouguettaya, A., et al., *Efficient agglomerative hierarchical clustering*. *Expert Systems with Applications*, 2015. **42**.
26. Ester, M., et al. *A density-based algorithm for discovering clusters in large spatial databases with noise*. in *kdd*. 1996.
27. Aynaud, T. and J.-L. Guillaume. *Détection de communautés à long terme dans les graphes dynamiques*. in *Journée thématique: Fouille de grands graphes*. 2010.

۲۸. برهمند، ک. and ح. احمدی بنی، یک طبقه بندی برای روش های تشخیص انجمن در شبکه های اجتماعی in، اولین کنفرانس ملی محاسبات توزیعی و پردازش داده های بزرگ، 1394.
۲۹. شفیع موسوی، ع. and م. مزینانی، مروری بر روش های تشخیص جوامع در شبکه های اجتماعی in، دوازدهمین کنفرانس ملی علوم و مهندسی کامپیوتر و فناوری اطلاعات، 1400.
۳۰. طباطبائی، آ. and م. زندی نژاد، الگوریتمی مبتنی بر انتظار برچسب در تشخیص اجتماعات در شبکه های اجتماعی in، هشتمین کنفرانس بین المللی وب پژوهی، 1401.
31. abdar, s.e., *Using garter snakes optimization algorithm for community detection in dynamic social networks*. journal of Information and communication Technology in policing, 2020. 1(2): p. 45–54.
۳۲. خفائی، ط. et al., تشخیص انجمن های پایدار در شبکه های اجتماعی پویا با استفاده از گره های با نفوذ.. 1401.
۳۳. میرزایی، م. and ا. ا. مه آبادی، روش ترکیبی تشخیص ناهنجاری با استفاده از تشخیص انجمن در گراف و انتخاب ویژگی.. 1399.
۳۴. هوشمند نژاد، ت. and م. نوش فر، تشخیص انجمن در گراف شبکه های اجتماعی با استفاده از شباهت ساختاری بین افراد in، اولین کنفرانس بین المللی چشم انداز های نو در مهندسی برق و کامپیوتر، 1395.
35. Deypir, M. and E. Bayat, *Identifying Community Structures in Social Networks using Discrete Harmony Search Algorithm*. Signal and Data Processing, 2023. 20(3): p. 47–60.
36. Shahriari, M., *Proposing a Model for Data Clustering Based on Harmonic Search Algorithm*. Journal of Operational Research In Its Applications (Applied Mathematics)-Lahijan Azad University, 2016. 13(3): p. 1–7.
37. Mir Mohammad, A. and A. Mohsen, *A Multiagent Reinforcement Learning algorithm to solve the Community Detection Problem*. Signal and Data Processing, 2022. 19(1): p. 87–100.
۳۸. فائزه شریف، ز. ک. سمیه، and ب. مرتضی، ارائه روشی جدید برای شناسایی گره های فعال و تاثیر گذار در شبکه های اجتماعی in، همایش ملی مهندسی رایانه و مدیریت فناوری اطلاعات، 1393.
۳۹. کریمی، ح.، الگوریتم خوشه بندی گره های حسگر با توجه به تراکم گره ها در شبکه های حسگر بی سیم.. 1400.
۴۰. رضایی، ع. م. باستانی، and س. آدابی، الگوریتم مکان یابی پویا مبتنی بر خوشه برای مکان یابی گره ها در شبکه موردی متحرک in، دومین کنفرانس بین المللی یافته های نوین پژوهشی در علوم، مهندسی و فناوری، 1395.
۴۱. علیخانی، م. and م. آبادی، روشی مبتنی بر خوشه بندی برای تشخیص ناهنجاری پویا در شبکه های اقتصادی متحرک با پروتکل مسیریابی AODV، in، هفتمین کنفرانس انجمن رمز ایران، 1389.
۴۲. عثمانی، ا. ا. طرقي حقیقت، and ح. پور اکبر، طراحی و ارزیابی یک الگوریتم مقیاس پذیر برای مدیریت مکان گره های سیار در پروتکل های مسیریابی مبتنی بر مکان در شبکه های سیار موردی in، پانزدهمین کنفرانس کامپیوتر سالانه انجمن کامپیوتر ایران، 1388.
43. Behnamian, J. and A.H. Safargholi, *A Hybird Algorithm for Hub Location Problem in Multimodal Logistic Networks*. Journal of Transportation Engineering, 2018. 10(2): p. 335–355.